Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 1. Δομή της Ύλης

Μάθημα 1. Κβαντική θεωρία του ατόμου. Ιοντικός δεσμός

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

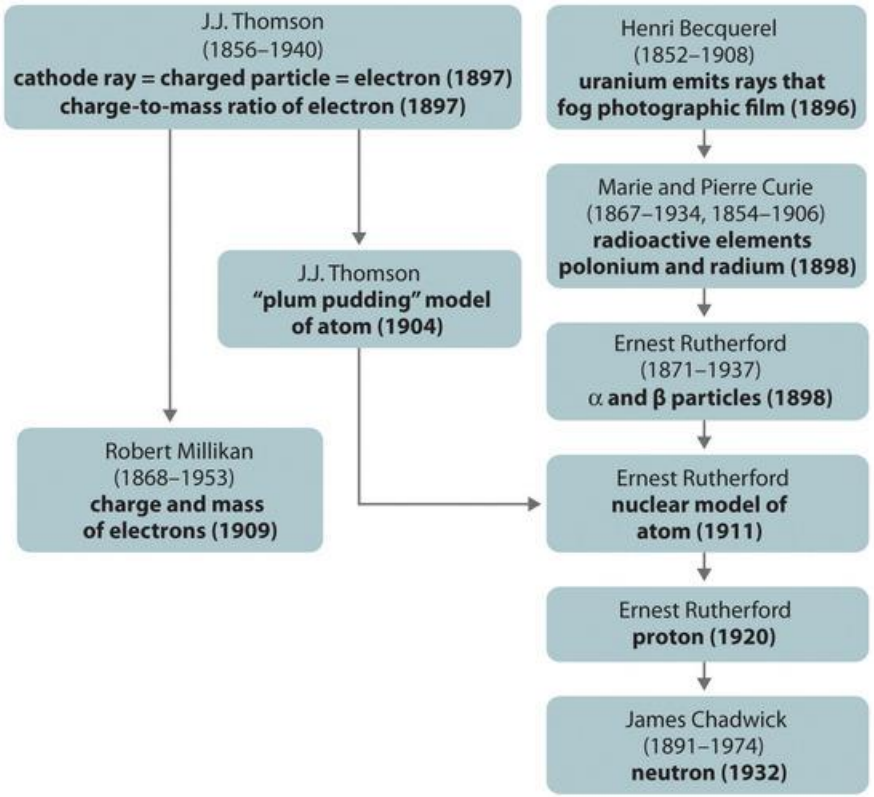
Περιεχόμενα

- Το μοντέλο του Bohr
- Το Κβαντομηχανικό Άτομο
- Κυματοσυναρτήσεις
- Κβαντικοί Αριθμοί
- Ατομικά Τροχιακά
- Ιοντικός Δεσμός

Εισαγωγή

Τα άτομα, τα μικρότερα σωματίδια ενός στοιχείου που παρουσιάζουν τις ιδιότητες αυτού του στοιχείου, αποτελούνται από αρνητικά φορτισμένα ηλεκτρόνια γύρω από έναν κεντρικό πυρήνα που αποτελείται από πιο μαζικά θετικά φορτισμένα πρωτόνια και ηλεκτρικά ουδέτερα νετρόνια. Ραδιενέργεια είναι η εκπομπή ενεργειακών σωματιδίων και ακτίνων (ακτινοβολία) από ορισμένες ουσίες. Τρία σημαντικά είδη ακτινοβολίας είναι τα σωματίδια α (πυρήνες ηλίου), τα σωματίδια β (ηλεκτρόνια που ταξιδεύουν με μεγάλη ταχύτητα) και οι ακτίνες γ.

Μια σύνοψη της ιστορικής ανάπτυξης των μοντέλων των συστατικών και της δομής του ατόμου

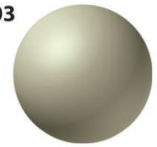


Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

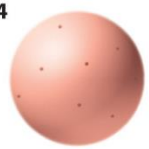
Η Εξέλιξη της Ατομικής Θεωρίας, όπως απεικονίζεται από μοντέλα του ατόμου του Οξυγόνου

1803



Dalton proposes the indivisible unit of an element is the atom.

1904

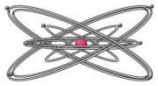


Thomson discovers electrons, believed to reside within a sphere of uniform positive charge (the plum pudding model).

1911

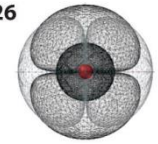
Rutherford demonstrates the existence of a positively charged nucleus that contains nearly all the mass of an atom.

1913



Bohr proposes fixed circular orbits around the nucleus for electrons.

1926



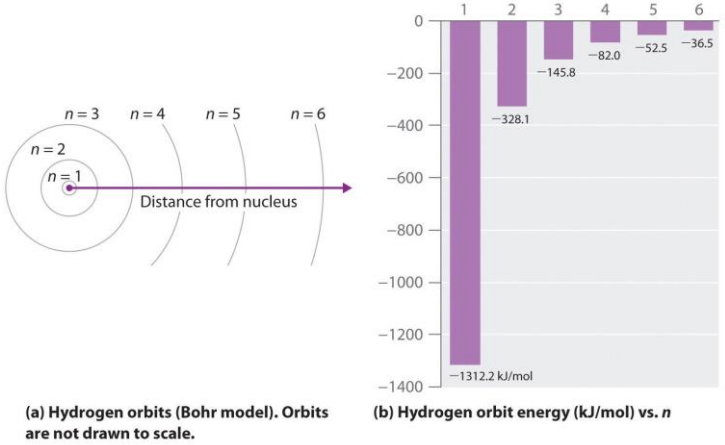
In the current model of the atom, electrons occupy regions of space (orbitals) around the nucleus determined by their energies.

Το μοντέλο του Bohr

Το 1913, ένας Δανός φυσικός, ο Niels Bohr (1885–1962, Βραβείο Νόμπελ Φυσικής, 1922), πρότεινε ένα θεωρητικό μοντέλο για το άτομο υδρογόνου που εξηγούσε το φάσμα εκπομπής του. Το μοντέλο του Bohr απαιτούσε μόνο μία υπόθεση: Το ηλεκτρόνιο κινείται γύρω από τον πυρήνα σε κυκλικές τροχιές που μπορούν να έχουν μόνο ορισμένες επιτρεπόμενες ακτίνες. Το προηγούμενο μοντέλο του ατόμου του Ράδερφορντ είχε επίσης υποθέσει ότι τα ηλεκτρόνια κινούνταν σε κυκλικές τροχιές γύρω από τον πυρήνα και ότι το άτομο συγκρατείται από την ηλεκτροστατική έλξη μεταξύ του θετικά φορτισμένου πυρήνα και του αρνητικά φορτισμένου ηλεκτρονίου. Αν και τώρα γνωρίζουμε ότι η υπόθεση των κυκλικών τροχιών ήταν εσφαλμένη, η διορατικότητα του Bohr ήταν να προτείνει ότι το ηλεκτρόνιο θα μπορούσε να καταλάβει μόνο ορισμένες περιοχές του χώρου.

Χρησιμοποιώντας την κλασική φυσική, ο Niels Bohr έδειξε ότι η ενέργεια ενός ηλεκτρονίου σε μια συγκεκριμένη τροχιά δίνεται από

$$E_n = \frac{-\Re hc}{n^2}$$



(a) Hydrogen orbits (Bohr model). Orbits are not drawn to scale.

(b) Hydrogen orbit energy (kJ/mol) vs. *n*

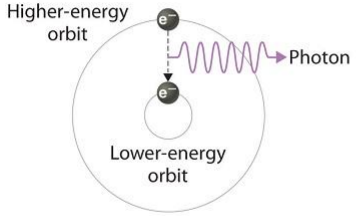
όπου R είναι η σταθερά Rydberg, h η σταθερά του Planck, c είναι η ταχύτητα του φωτός και n είναι ένας θετικός ακέραιος αριθμός που αντιστοιχεί στον αριθμό που έχει εκχωρηθεί στην τροχιά, με n = 1 να αντιστοιχεί στην τροχιά που βρίσκεται πλησιέστερα στον πυρήνα. Σε αυτό το μοντέλο το n = ∞ αντιστοιχεί στο επίπεδο όπου η ενέργεια που συγκρατεί το ηλεκτρόνιο και τον πυρήνα μαζί είναι μηδέν. Σε αυτό το επίπεδο, το ηλεκτρόνιο δεν είναι δεσμευμένο από τον πυρήνα και το άτομο έχει διαχωριστεί σε ένα αρνητικά φορτισμένο (το ηλεκτρόνιο) και ένα θετικά φορτισμένο (τον πυρήνα) ιόν. Σε αυτή την κατάσταση η ακτίνα της τροχιάς είναι επίσης άπειρη.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

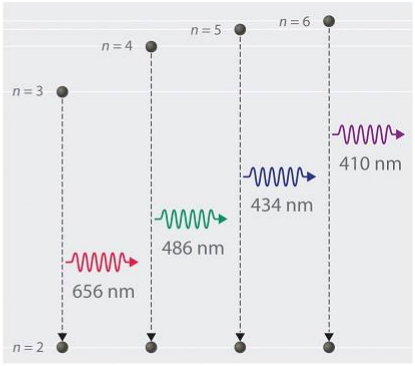
Co-funded by the Erasmus+ Programme of the European Union

Το μοντέλο του Bohr

Καθώς το n μειώνεται, η ενέργεια που συγκρατεί το ηλεκτρόνιο και τον πυρήνα μαζί γίνεται όλο και πιο αρνητική, η ακτίνα της τροχιάς συρρικνώνεται και χρειάζεται περισσότερη ενέργεια για να ιονιστεί το άτομο. Η τροχιά με n = 1 είναι η χαμηλότερη και πιο σφιχτά δεσμευμένη. Το αρνητικό πρόσημο στην εξίσωση δείχνει ότι το ζεύγος ηλεκτρονίου-πυρήνα είναι πιο σφιχτά συνδεδεμένο (δηλαδή σε χαμηλότερη δυναμική ενέργεια) όταν βρίσκονται το ένα κοντά στο άλλο παρά όταν απέχουν πολύ μεταξύ τους. Επειδή ένα άτομο υδρογόνου με το ένα του ηλεκτρόνιο σε αυτή την τροχιά έχει τη χαμηλότερη δυνατή ενέργεια, αυτή είναι η θεμελιώδης κατάσταση (η πιο σταθερή διάταξη ηλεκτρονίων για ένα στοιχείο ή μια ένωση) για ένα άτομο υδρογόνου. Καθώς το n αυξάνεται, η ακτίνα της τροχιάς αυξάνεται. το ηλεκτρόνιο είναι πιο μακριά από το πρωτόνιο, γεγονός που οδηγεί σε μια λιγότερο σταθερή διάταξη με υψηλότερη δυναμική ενέργεια. Ένα άτομο υδρογόνου με ένα ηλεκτρόνιο σε τροχιά με n > 1 είναι επομένως σε διεγερμένη κατάσταση, που ορίζεται ως οποιαδήποτε διάταξη ηλεκτρονίων που είναι υψηλότερη σε ενέργεια από τη θεμελιώδη κατάσταση. Όταν ένα άτομο σε διεγερμένη κατάσταση υφίσταται μια μετάβαση στη θεμελιώδη κατάσταση σε μια διαδικασία που ονομάζεται διάσπαση, χάνει ενέργεια εκπέμποντας ένα φωτόνιο του οποίου η ενέργεια αντιστοιχεί στη διαφορά ενέργειας μεταξύ των δύο καταστάσεων.



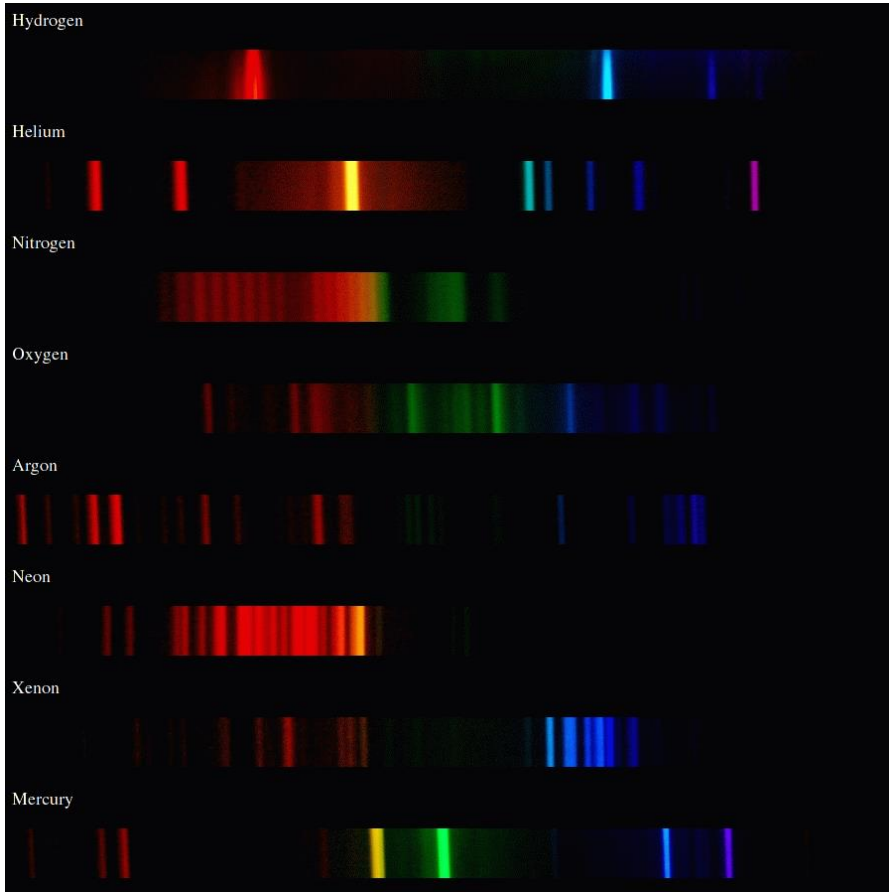
(a) Electronic emission transition



(b) Balmer series transitions

Δυστυχώς, ο Bohr δεν μπορούσε να εξηγήσει γιατί το ηλεκτρόνιο θα έπρεπε να περιοριστεί σε συγκεκριμένες τροχιές. Επίσης, παρά τη μεγάλη δυσκολία, όπως η υπόθεση ότι οι τροχιές θα μπορούσαν να είναι ελλείψεις και όχι κύκλοι, το μοντέλο του δεν μπορούσε να εξηγήσει ποσοτικά τα φάσματα εκπομπής οποιουδήποτε στοιχείου εκτός από το υδρογόνο. Στην πραγματικότητα, το μοντέλο του Bohr λειτούργησε μόνο για είδη που περιείχαν μόνο ένα ηλεκτρόνιο: H, He^{+}, Li^{2+} και ούτω καθεξής.

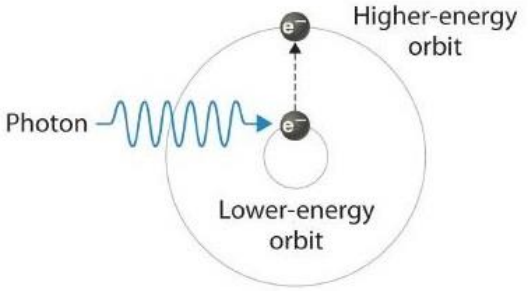
Τα φάσματα ατομικής εκπομπής για διάφορα στοιχεία. Κάθε λεπτή ζώνη σε κάθε φάσμα αντιστοιχεί σε μια μοναδική, μοναδική μετάβαση μεταξύ των ενεργειακών επιπέδων σε ένα άτομο.



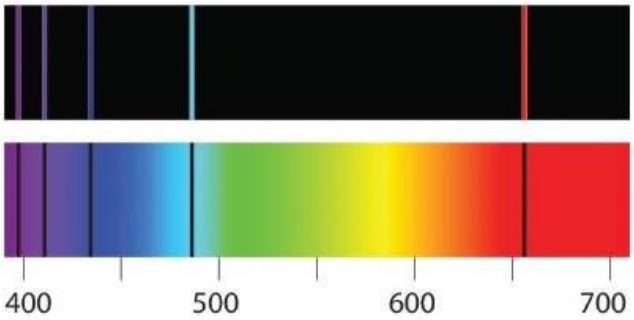
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Οι Ενεργειακές Καταστάσεις του Ατόμου Υδρογόνου

Εάν το λευκό φως περάσει μέσα από ένα δείγμα υδρογόνου, τα άτομα υδρογόνου απορροφούν ενέργεια καθώς ένα ηλεκτρόνιο διεγείρεται σε υψηλότερα επίπεδα ενέργειας (τροχίες με n ≥ 2). Εάν το φως που αναδύεται διέρχεται από ένα πρίσμα, σχηματίζει ένα συνεχές φάσμα με μαύρες γραμμές (που αντιστοιχούν σε κανένα φως που δεν διέρχεται από το δείγμα) στα 656, 468, 434 και 410 nm. Αυτά τα μήκη κύματος αντιστοιχούν στις μεταβάσεις n = 2 έως n = 3, n = 2 έως n = 4, n = 2 έως n = 5 και n = 2 έως n = 6 μεταβάσεις. Επομένως, οποιοδήποτε δεδομένο στοιχείο έχει ένα χαρακτηριστικό φάσμα εκπομπής και ένα χαρακτηριστικό φάσμα απορρόφησης, τα οποία είναι ουσιαστικά συμπληρωματικές εικόνες.



(a) Electronic absorption transition



(b) H_2 emission spectrum (top), H_2 absorption spectrum (bottom)

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Το Κβαντομηχανικό Άτομο

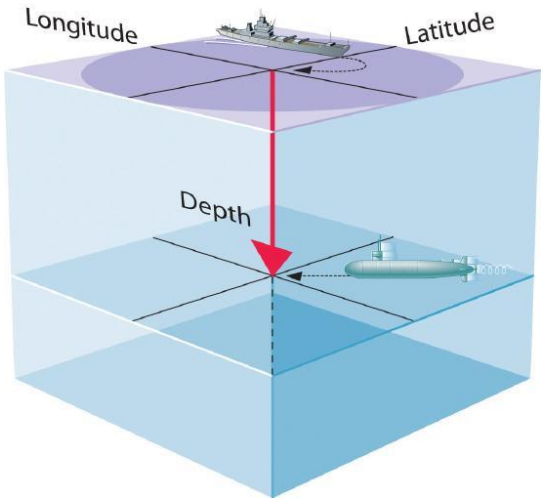
Το παράδοξο που περιγράφεται από την αρχή της αβεβαιότητας του Heisenberg και η κυματοειδής φύση των υποατομικών σωματιδίων όπως το ηλεκτρόνιο κατέστησε αδύνατη τη χρήση των εξισώσεων της κλασικής φυσικής για την περιγραφή της κίνησης των ηλεκτρονίων στα άτομα. Οι επιστήμονες χρειάζονταν μια νέα προσέγγιση που θα λάμβανε υπόψη την κυματική συμπεριφορά του ηλεκτρονίου. Το 1926, ένας Αυστριακός φυσικός, ο Erwin Schrödinger (1887–1961, βραβείο Νόμπελ Φυσικής, 1933), ανέπτυξε την κυματομηχανική, μια μαθηματική τεχνική που περιγράφει τη σχέση μεταξύ της κίνησης ενός σωματιδίου που παρουσιάζει κυματοειδείς ιδιότητες (όπως ένα ηλεκτρόνιο) και τις επιτρεπόμενες ενέργειές του.

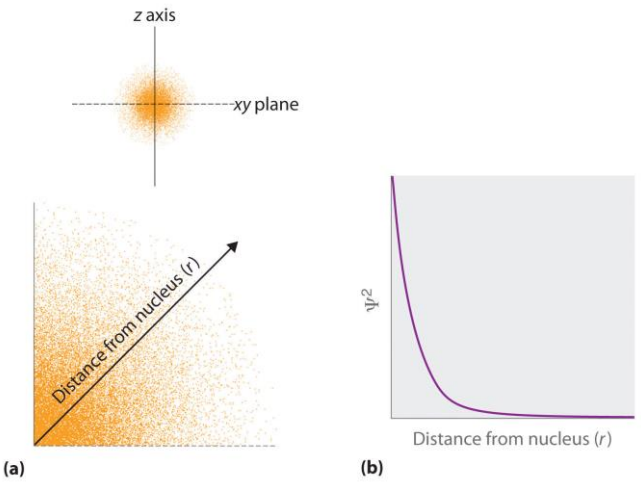
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Κυματοσυναρτήσεις

- Μια κυματοσυνάρτηση χρησιμοποιεί τρεις μεταβλητές για να περιγράψει τη θέση ενός ηλεκτρονίου.
- Το μέγεθος της κυματοσυνάρτησης σε ένα συγκεκριμένο σημείο του χώρου είναι ανάλογο με το πλάτος του κύματος σε αυτό το σημείο
- Το τετράγωνο της κυματοσυνάρτησης σε ένα δεδομένο σημείο είναι ανάλογο με την πιθανότητα να βρεθεί ένα ηλεκτρόνιο σε αυτό το σημείο, γεγονός που οδηγεί σε κατανομή των πιθανοτήτων στο χώρο.
- Η περιγραφή της κατανομής των ηλεκτρονίων ως στάσιμο κύμα οδηγεί σε σύνολα κβαντικών αριθμών που είναι χαρακτηριστικά κάθε κυματοσυνάρτησης.
- Κάθε κυματοσυνάρτηση συνδέεται με μια συγκεκριμένη ενέργεια.





Κβαντικοί Αριθμοί

Ο **κύριος κβαντικός αριθμός (n)** δείχνει τη μέση σχετική απόσταση ενός ηλεκτρονίου από τον πυρήνα. Καθώς το n αυξάνεται για ένα δεδομένο άτομο, τόσο αυξάνεται η μέση απόσταση ενός ηλεκτρονίου από τον πυρήνα.

Ο δεύτερος κβαντικός αριθμός ονομάζεται συχνά **αζιμουθιακός κβαντικός αριθμός (l)** . Η τιμή του l περιγράφει το σχήμα της περιοχής του χώρου που καταλαμβάνει το ηλεκτρόνιο. Οι επιτρεπόμενες τιμές του l εξαρτώνται από την τιμή του n και μπορεί να κυμαίνονται από 0 έως n − 1.

Ο τρίτος κβαντικός αριθμός είναι ο μαγνητικός κβαντικός αριθμός (m_j). Η τιμή του mj περιγράφει τον προσανατολισμό της περιοχής στο χώρο που καταλαμβάνει ένα ηλεκτρόνιο σε σχέση με ένα εφαρμοζόμενο μαγνητικό πεδίο. Οι επιτρεπόμενες τιμές του m_j εξαρτώνται από την τιμή του l: το ml μπορεί να κυμαίνεται από –l έως l σε ακέραια βήματα.

Κάθε κυματοσυνάρτηση με έναν επιτρεπόμενο συνδυασμό τιμών n, l και m_j περιγράφει ένα **ατομικό τροχιακό** , μια συγκεκριμένη χωρική κατανομή για ένα ηλεκτρόνιο. Για ένα δεδομένο σύνολο κβαντικών αριθμών, κάθε κύριο κέλυφος έχει έναν σταθερό αριθμό υποφλοιών και κάθε υποκέλυφος έχει έναν σταθερό αριθμό τροχιακών.

Co-funded by the Erasmus+ Programme of the European Union

Ατομικά Τροχιακά

Ένα τροχιακό είναι η κβαντομηχανική τελειοποίηση της τροχιάς του Bohr. Σε αντίθεση με την αντίληψή του για μια απλή κυκλική τροχιά με σταθερή ακτίνα, τα τροχιακά είναι μαθηματικά προερχόμενες περιοχές του χώρου με διαφορετικές **πιθανότητες** να περιέχουν ένα ηλεκτρόνιο.

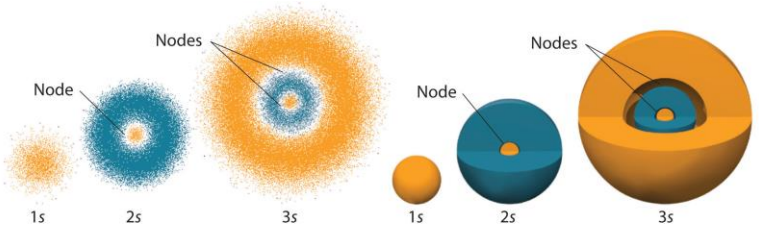
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

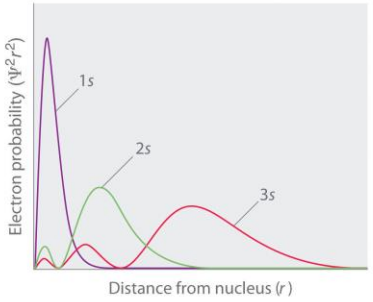
Co-funded by the Erasmus+ Programme of the European Union

s Τροχιακά (l=0)

Τρία πράγματα συμβαίνουν στα τροχιακά s καθώς το n αυξάνεται:

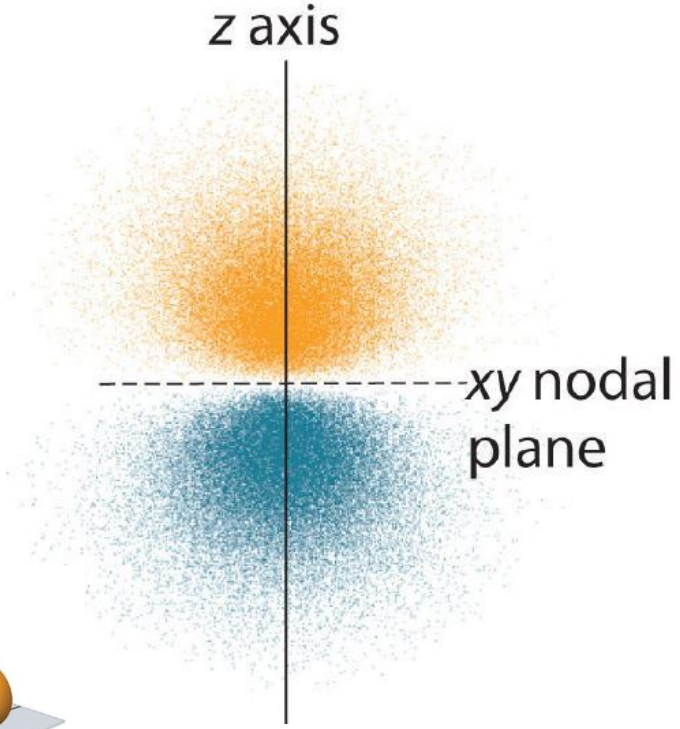
- Γίνονται μεγαλύτερα, εκτείνονται πιο μακριά από τον πυρήνα.
- Περιέχουν περισσότερους κόμβους. Αυτό είναι παρόμοιο με ένα στάσιμο κύμα που έχει περιοχές σημαντικού πλάτους που χωρίζονται από κόμβους, σημεία με μηδενικό πλάτος.
- Για ένα δεδομένο άτομο, τα τροχιακά s γίνονται επίσης υψηλότερα σε ενέργεια καθώς το n αυξάνεται λόγω της αυξημένης απόστασής τους από τον πυρήνα.

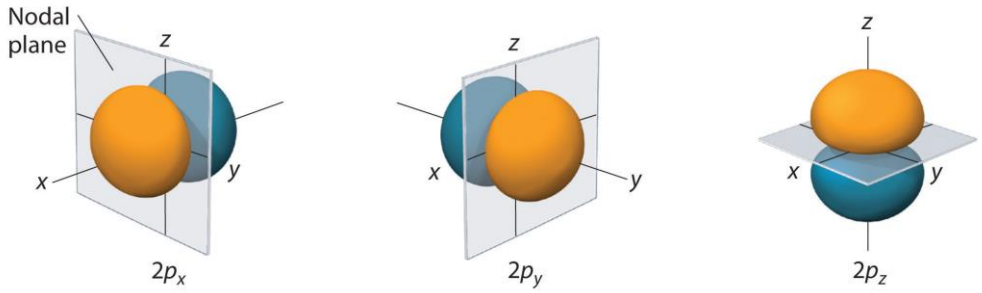




p Τροχιακά (l=1)

Μόνο τα τροχιακά s είναι σφαιρικά συμμετρικά. Καθώς η τιμή του l αυξάνεται, ο αριθμός των τροχιακών σε ένα δεδομένο υποκέλυφος αυξάνεται και τα σχήματα των τροχιακών γίνονται πιο πολύπλοκα. Επειδή το υποκέλυφος 2p έχει l = 1, με τρεις τιμές του m_j (−1, 0 και +1), υπάρχουν τρία τροχιακά 2p.



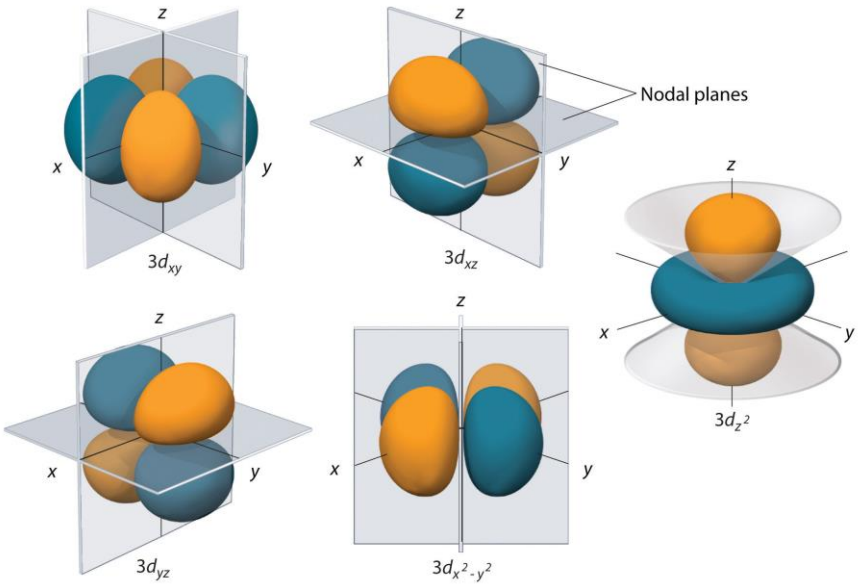


Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

d Τροχιακά (l=2)

Τα υποκελύφη με l = 2 έχουν πέντε d τροχιακά. το πρώτο κύριο κέλυφος που έχει ad subshell αντιστοιχεί σε n = 3. Τα πέντε d τροχιακά έχουν m_j τιμές −2, −1, 0, +1 και +2.



Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Τροχιακές Ενέργειες

Διάγραμμα τροχιακής στάθμης ενέργειας για το άτομο υδρογόνου με ένα μόνο ηλεκτρόνιο. Κάθε κουτί αντιστοιχεί σε ένα τροχιακό. Σημειώστε ότι η διαφορά στην ενέργεια μεταξύ των τροχιακών μειώνεται γρήγορα με την αύξηση των τιμών του n.

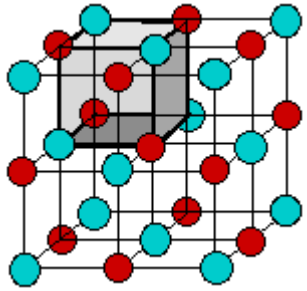
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ιοντικός Δεσμός

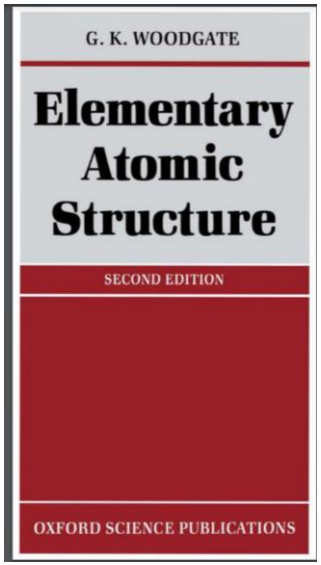
Ο ιοντικός δεσμός σχηματίζεται όταν η «μοιρασιά» είναι τόσο άνιση που ένα ηλεκτρόνιο από το άτομο Α χάνεται εντελώς στο άτομο Β, με αποτέλεσμα ένα ζεύγος ιόντων:

Σύμφωνα με το ιοντικό ηλεκτροστατικό μοντέλο, τα στερεά όπως το NaCl αποτελούνται από θετικά και αρνητικά ιόντα διατεταγμένα σε ένα κρυσταλλικό πλέγμα. Κάθε ιόν έλκεται από γειτονικά ιόντα αντίθετου φορτίου και απωθείται από ιόντα παρόμοιου φορτίου. Αυτός ο συνδυασμός έλξης και απωθήσεων, που ενεργεί προς όλες τις κατευθύνσεις, προκαλεί το ιόν να στερεωθεί σφιχτά στη δική του θέση στο κρυσταλλικό πλέγμα.

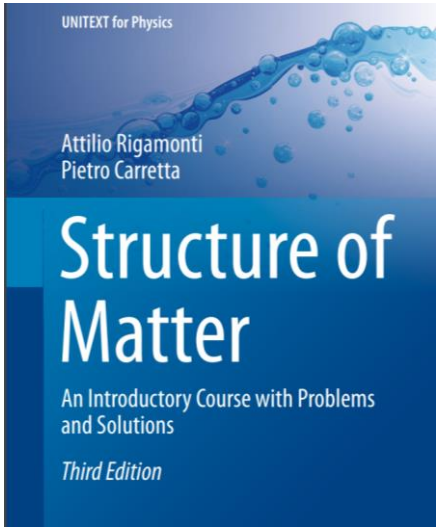


Ερωτήσεις

- Ποια είναι η ιστορική ανάπτυξη των μοντέλων των συστατικών και της δομής του ατόμου;
- Εξηγήστε το Κβαντομηχανικό Άτομο
- Τι είναι οι κυματοσυναρτήσεις;
- Εξηγήστε τον όρο Ατομικά Τροχιακά και δώστε παραδείγματα.



GK Woodgate (2002) Elementary Atomic Structure, Oxford University Press Inc., Νέα Υόρκη



Rigamonti , A., & Carretta , P. (2015). Δομή της Ύλης. UNITEXT for Physics, Springer International Publishing Switzerland

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 1. Δομή της ύλης

Μάθημα 2. Χημικοί δεσμοί σε διατομικά και πολυατομικά μόρια. Γεωμετρική διαμόρφωση μορίων

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Η προσέγγιση Born-Oppenheimer
- Καμπύλες και επιφάνειες δυνητικής ενέργειας
- Δέσμευση σε Πολυατομικά Μόρια
- Παραγωγή μικτών γενών
- Μοριακή γεωμετρία

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Η βάση για την κατανόηση των χημικών δεσμών και των δομών των μορίων είναι η τροχιακή περιγραφή των ηλεκτρονίων της δομής και του σθένους των ατόμων, όπως παρέχεται από την κβαντική μηχανική. Η χρήση της κβαντικής μηχανικής για την πρόβλεψη των μοτίβων χημικών δεσμών, των βέλτιστων γεωμετριών και των φυσικών και χημικών ιδιοτήτων των μορίων είναι ένα μεγάλο και ενεργό πεδίο έρευνας γνωστό ως μοριακή κβαντική μηχανική ή πιο συχνά ως κβαντική χημεία. Οι υπολογισμοί της κβαντικής χημείας επιτρέπουν τον υπολογισμό της γεωμετρίας των μορίων καθώς και για ένα ευρύ φάσμα ιδιοτήτων. Η κβαντική χημεία μπορεί επίσης να χρησιμοποιηθεί με έναν νέο τρόπο, στον οποίο τα ηλεκτρόνια αντιμετωπίζονται με χρήση κβαντομηχανικής, αλλά οι πυρήνες αντιμετωπίζονται ως κλασικά σωματίδια. Χρησιμοποιούμε την κβαντική μηχανική για να υπολογίσουμε τις διαπυρηνικές δυνάμεις, αλλά στη συνέχεια χρησιμοποιούμε αυτές τις δυνάμεις στον Δεύτερο Νόμο του Νεύτωνα για να μελετήσουμε την κίνηση των πυρήνων κατά τη διάρκεια χημικών αντιδράσεων. Αυτό μας δίνει ένα μικροσκοπικό παράθυρο στις συγκεκριμένες κινήσεις, τον πολύπλοκο χορό, που εκτελείται από τους πυρήνες κατά τη διάρκεια μιας απλής ή πολύπλοκης χημικής διαδικασίας.

Η προσέγγιση Born-Oppenheimer

Η προσέγγιση Born-Oppenheimer είναι μια από τις βασικές έννοιες που διέπουν την περιγραφή των κβαντικών καταστάσεων των μορίων. Αυτή η προσέγγιση καθιστά δυνατό τον διαχωρισμό της κίνησης των πυρήνων και της κίνησης των ηλεκτρονίων. Αυτή δεν είναι μια νέα ιδέα για εμάς. Χρησιμοποιήσαμε ήδη αυτήν την προσέγγιση στο μοντέλο σωματιδίων σε κουτί όταν εξηγήσαμε τα ηλεκτρονικά φάσματα απορρόφησης των χρωστικών κυανίνης χωρίς να λάβουμε υπόψη την κίνηση των πυρήνων. Στη συνέχεια συζητήσαμε τη μεταφορική, περιστροφική και δονητική κίνηση των πυρήνων χωρίς να συμπεριλάβουμε την κίνηση των ηλεκτρονίων. Σε αυτό το κεφάλαιο θα εξετάσουμε πιο προσεκτικά τη σημασία και τις συνέπειες αυτής της σημαντικής προσέγγισης. Σημείωση, σε αυτή τη συζήτηση ο πυρήνας αναφέρεται στους ατομικούς πυρήνες ως μέρη μορίων και όχι στην εσωτερική δομή του πυρήνα.

Η προσέγγιση Born-Oppenheimer

Για ένα διατομικό μόριο για παράδειγμα, ο τελεστής Hamiltonian ομαδοποιείται σε τρεις όρους.

$$\hat{H}(r,R) = \hat{T}_{nuc}(R) + \frac{e^2}{4\pi\epsilon_0}\frac{Z_A Z_B}{R} + \hat{H}_{elec}(r,R)$$

$$\hat{T}_{nuc}(R) = -\frac{\hbar^2}{2m_A}\nabla_A^2 - \frac{\hbar^2}{2m_B}\nabla_B^2$$

$$\hat{H}_{elec}(\vec{r},\vec{R}) = \frac{-\hbar^2}{2m}\sum_i \nabla_i^2 + \frac{e^2}{4\pi\epsilon_0}\left(-\sum_i \frac{Z_A}{r_{Ai}} - \sum_i \frac{Z_B}{r_{Bi}} + \frac{1}{2}\sum_i \sum_{j\neq i} \frac{1}{r_{ij}}\right)$$

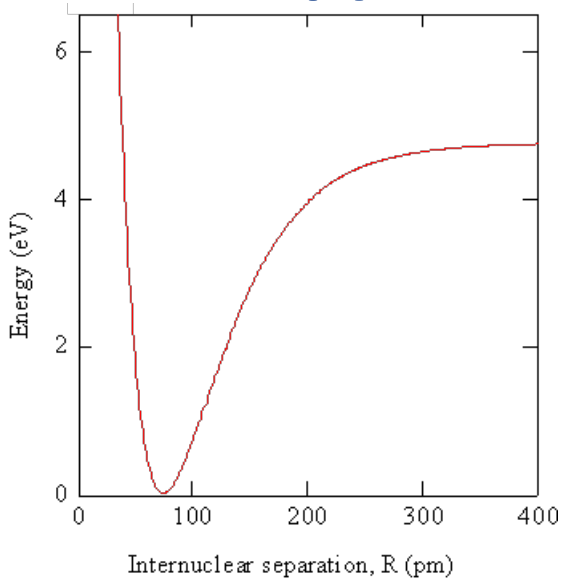
Η προσέγγιση Born-Oppenheimer λέει ότι οι όροι της πυρηνικής κινητικής ενέργειας στο πλήρες Hamiltonian μπορούν να παραμεληθούν κατά την επίλυση των ηλεκτρονικών κυματοσυναρτήσεων και ενεργειών. Κατά συνέπεια, η ηλεκτρονική κυματοσυνάρτηση Το $\phi_e(r,R)$ βρίσκεται ως λύση στην ηλεκτρονική εξίσωση Schrödinger

$$\hat{H}_{elec}(r,R)\varphi_e(r,R) = E_e(R)\varphi_e(r,R)$$

Παρόλο που οι όροι της πυρηνικής κινητικής ενέργειας παραμελούνται, η προσέγγιση Born-Oppenheimer εξακολουθεί να λαμβάνει υπόψη τη διακύμανση στις θέσεις των πυρήνων στον προσδιορισμό της ηλεκτρονικής ενέργειας και την προκύπτουσα ηλεκτρονική κυματοσυνάρτηση εξαρτάται από τις πυρηνικές θέσεις, R. Ως αποτέλεσμα του Προσέγγιση Born-Oppenheimer, η μοριακή κυματοσυνάρτηση μπορεί να γραφτεί ως γινόμενο

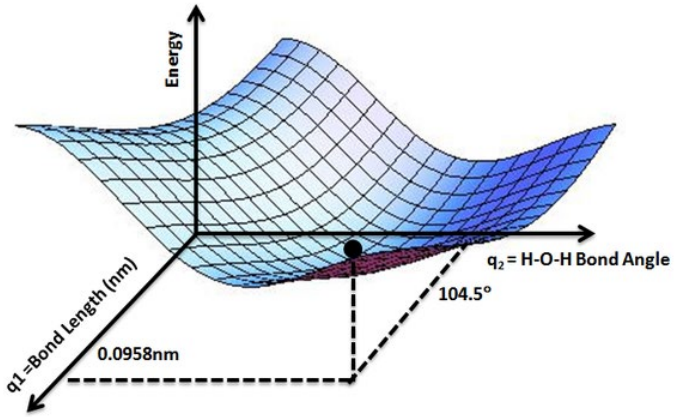
$$\psi_{ne}(r,R) = X_{ne}(R)\varphi_e(r,R)$$

Καμπύλες και επιφάνειες δυνητικής ενέργειας



Η συνάρτηση δυναμικής ενέργειας για ένα διατομικό μόριο. Στην πράξη η ηλεκτρονική εξίσωση Schrödinger λύνεται χρησιμοποιώντας προσεγγίσεις σε συγκεκριμένες τιμές του R για να ληφθούν οι κυματοσυναρτήσεις $\phi_e(r,R)$ και δυνητικές ενέργειες $V_e(R)$

.



Η επιφάνεια δυναμικής ενέργειας για ένα μόριο νερού: Δείχνει το ελάχιστο ενεργειακό που αντιστοιχεί στη βελτιστοποιημένη μοριακή δομή για μήκος δεσμού νερού-OH 0,0958 nm και γωνία δεσμού HOH 104,5°.

Δέσμευση σε Πολυατομικά Μόρια

Βασικά δύο τρόποι προσέγγισης της πολυατομίας:

Πρώτον, είναι η χρήση μετατοπισμένων ΜΟ όπου το e- δεν περιορίζεται σε έναν μόνο δεσμό (περιοχή μεταξύ 2 ατόμων) αλλά μπορεί να περιπλανηθεί σε 3 ή περισσότερα άτομα. Θα χρησιμοποιήσουμε αυτήν την προσέγγιση αργότερα για τη σύνδεση C.

Το δεύτερο είναι να χρησιμοποιήσουμε υβριδισμό ατομικών τροχιακών και στη συνέχεια να τα χρησιμοποιήσουμε για να σχηματίσουμε τοπικούς (συνήθως) δεσμούς.

Ο υβριδισμός συνδυάζει τροχιακά στον ΙΔΙΟ πυρήνα για να σχηματίσει νέα τροχιακά που ονομάζονται υβρίδια. Τα υβρίδια έχουν χαρακτηριστικά και των ατομικών τροχιακών από τα οποία σχηματίζονται.

Παράδειγμα: Το υβρίδιο sp σχηματίζεται συνδυάζοντας (προσθέτοντας) ένα 2s και ένα 2p τροχιακό σε ένα μόνο άτομο.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

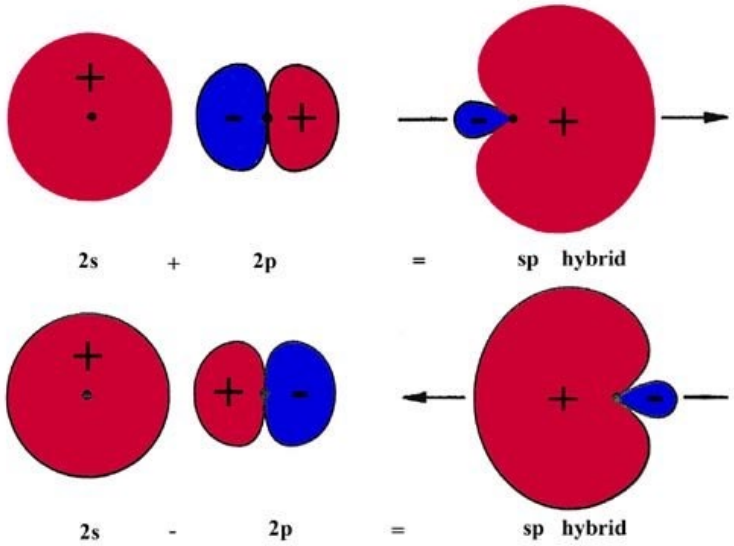
Παραγωγή μικτών γενών

Το μόριο BeH_2 είναι γραμμικό και οι δύο δεσμοί Be-H είναι ισοδύναμοι. Η περιγραφή του δεσμού σθένους του BeH2 αντιπροσώπευε το διπλό σθένος του Be (το οποίο έχει τη διαμόρφωση θεμελιώδους κατάστασης $1s^2 2s^2$) υποθέτοντας ότι η σύνδεση θα συμβεί με μια προωθούμενη διαμόρφωση του Be:

Be* $1s^2$ 2s()2p()

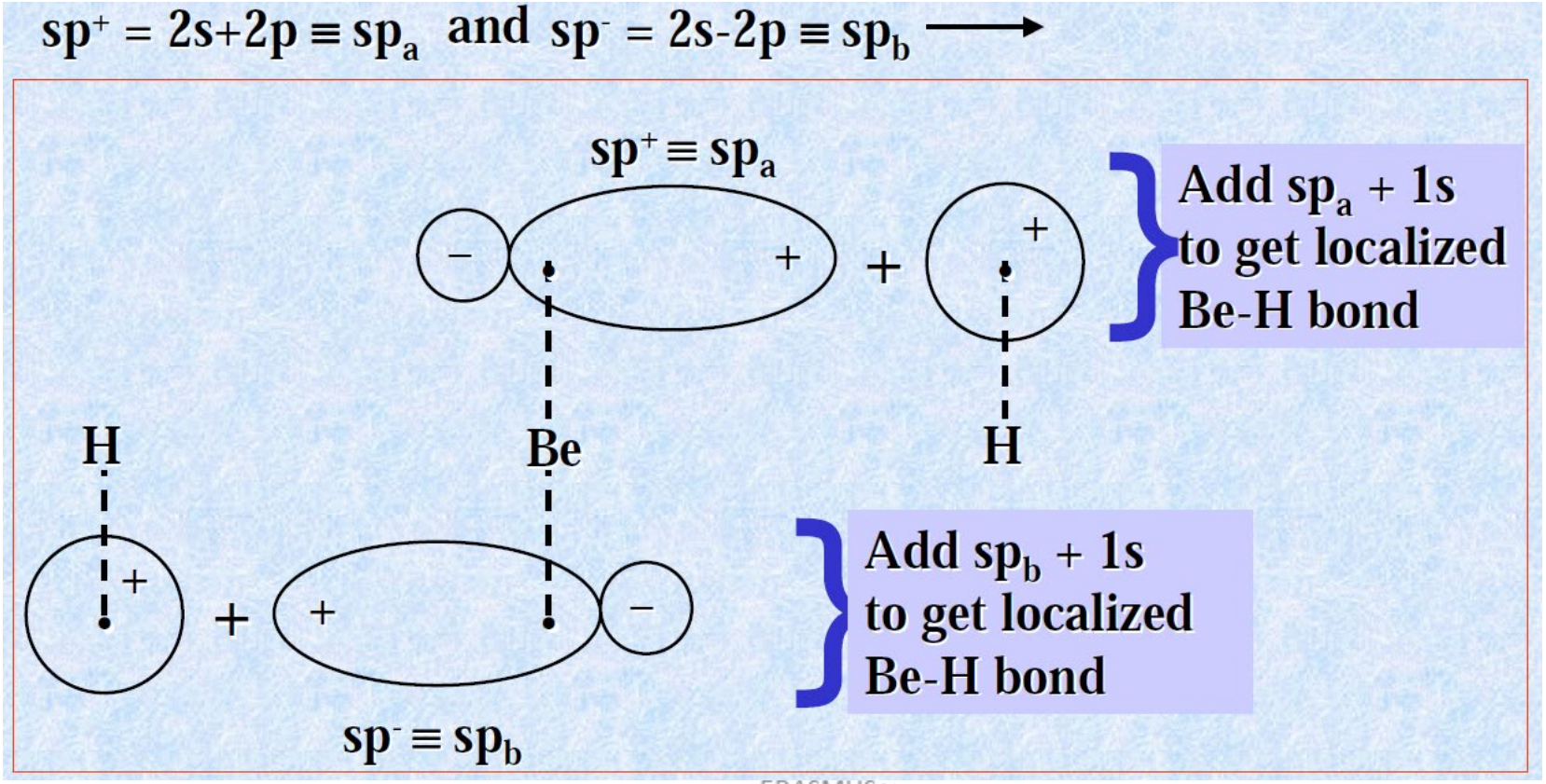
Εκ πρώτης όψεως αυτό υποδηλώνει ότι οι δύο δεσμοί Be-H θα πρέπει να είναι ανόμοιοι και όχι απαραίτητα 180° μεταξύ τους επειδή ο ένας δεσμός προκύπτει από την επικάλυψη με ένα τροχιακό 2s και ο άλλος με ένα τροχιακό 2p στο Be.

Η κατασκευή των υβριδικών τροχιακών ολοκληρώνεται λαμβάνοντας το άθροισμα και τη διαφορά του τροχιακού 2s και ενός από τα τροχιακά 2p, ας πούμε του τροχιακού $2p_x$, και τα δύο τροχιακά είναι επικεντρωμένα στον πυρήνα Be.

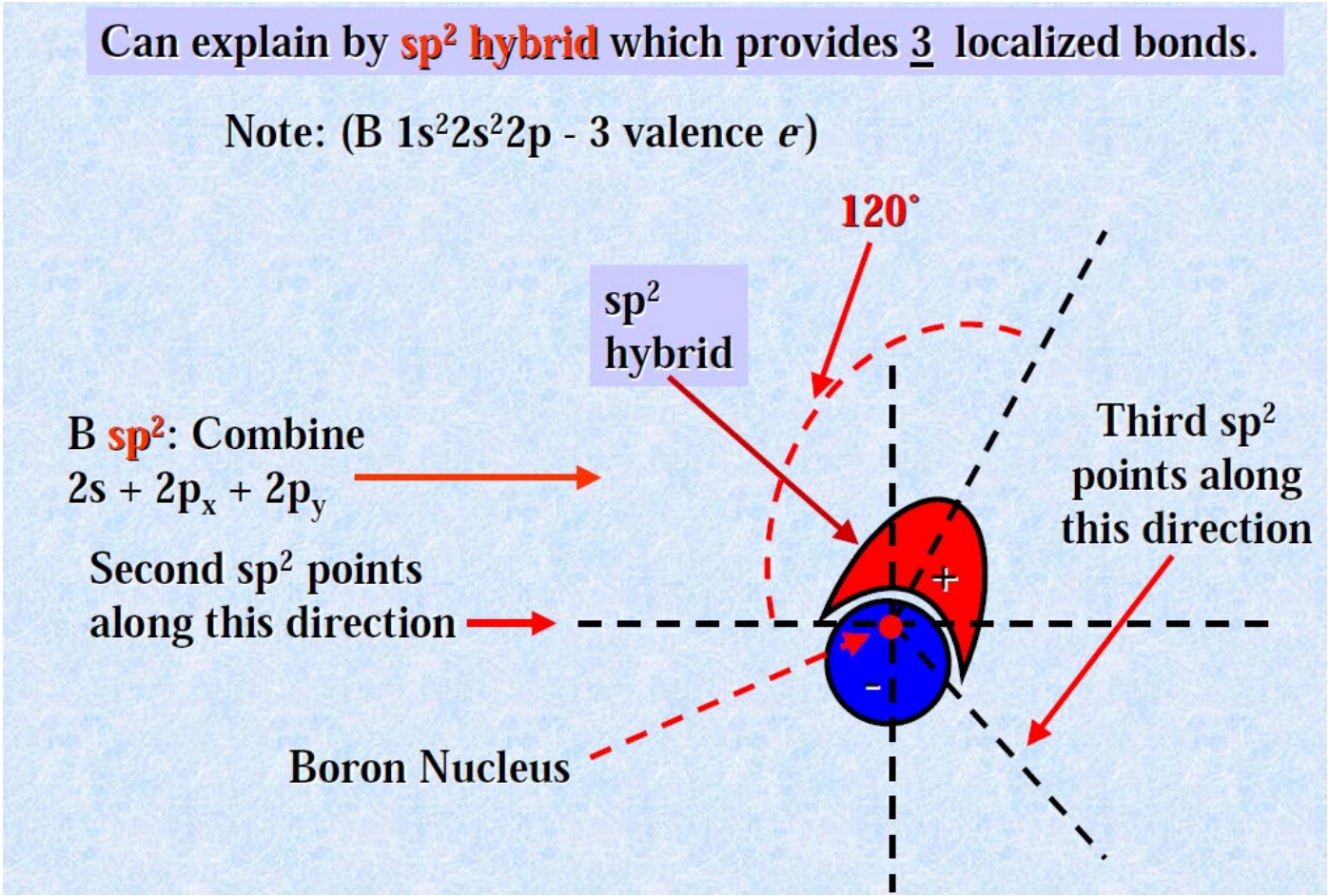


τροχιακά

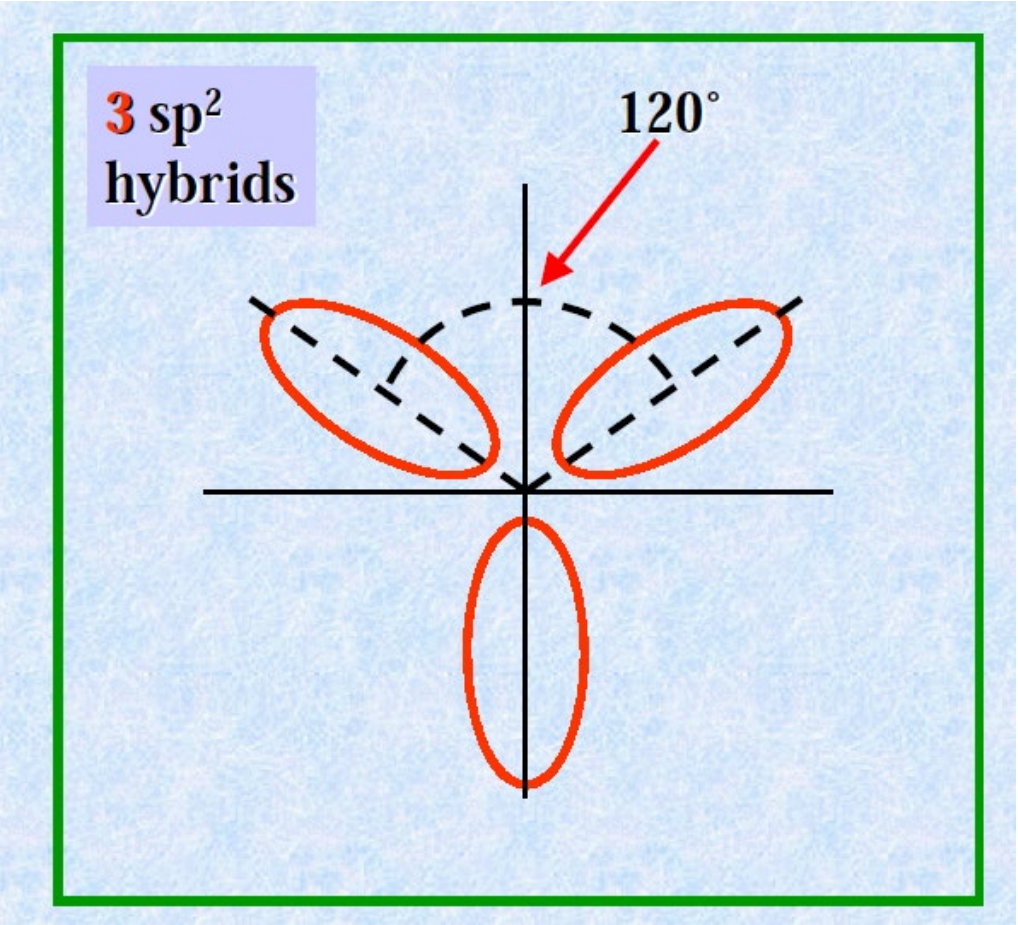
Ξεκινήστε υβριδοποιώντας τα τροχιακά Be 2s, 2p για να δώσετε δύο υβριδικά τροχιακά



Θραύσμα


Can explain by sp² hybrid which provides 3 localized bonds.
Note: (B 1s²2s²2p - 3 valence e⁻)
120°
sp² hybrid
B sp²: Combine 2s + 2p_x + 2p_y
Second sp² points along this direction
Third sp² points along this direction
Boron Nucleus
+
-

Τρία από αυτά σε ένα επίπεδο που δείχνει 120° το ένα ως προς το άλλο

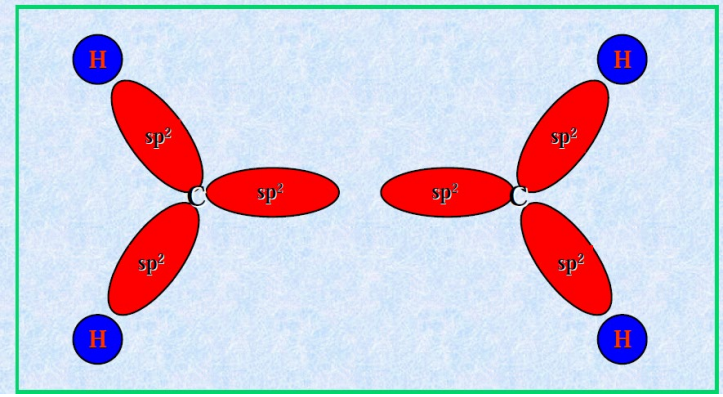


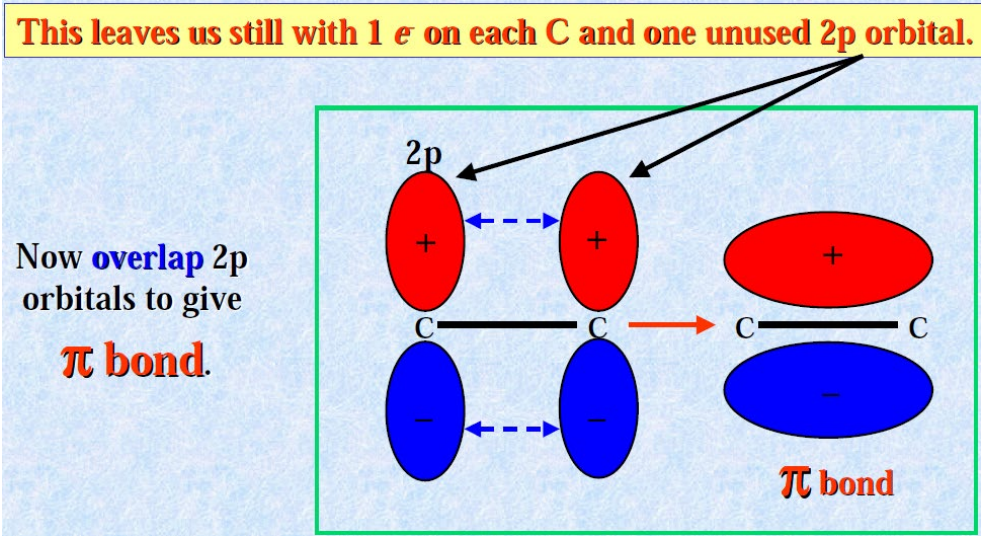
Σύνοψη των αποτελεσμάτων υβριδισμού

Παράδειγμα	Συνημμένες ομάδες στο κέντρο του ατόμου	Υβρίδιο	Γεωμετρία
BeH_2	2	sp	γραμμικό H-Be-H
BH_3	3	sp^2	αεροπλάνο (120º)
CH_4	4	sp^3	τετράεδρο (109º28' γωνίες HCH)

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Ο άνθρακας μπορεί επίσης να παρουσιάσει υβριδισμό sp^2 : C_2H_4 (αιθυλένιο)



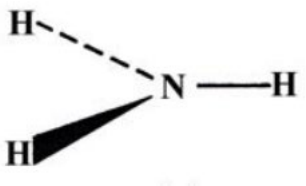


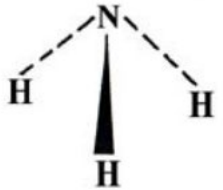
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Μοριακή γεωμετρία

Η θεωρία του σθένους είναι γνωστή ως θεωρία του δεσμού σθένους. Ένα επιπλέον χαρακτηριστικό αυτής της θεωρίας είναι ότι μπορεί να χρησιμοποιηθεί για την πρόβλεψη (ή σε ορισμένες περιπτώσεις, τον εξορθολογισμό) των παρατηρούμενων γεωμετριών των μορίων. Η γεωμετρία ενός μορίου σημαίνει τη σχετική διάταξη των πυρήνων στον τρισδιάστατο χώρο. Για παράδειγμα, αν υποθέσουμε ότι οι δύο δεσμοί OH στο μόριο του νερού είναι παρόμοιοι και επομένως έχουν το ίδιο μήκος, η γωνία που σχηματίζεται από τους δύο δεσμούς OH (η γωνία HOH θα μπορούσε να έχει οποιαδήποτε γωνία από 180° έως κάποια σχετικά μικρή τιμή. Η απαίτηση της απλής θεωρίας μας είναι να προβλέπει σωστά εάν το μόριο του νερού είναι γραμμικό (γωνία δεσμού = 180°) ή λυγισμένο (γωνία δεσμού μικρότερη από 180°) Ή ως άλλο παράδειγμα, θα πρέπει να προβλέψει εάν το μόριο της αμμωνίας είναι επίπεδο ή πυραμιδικό.



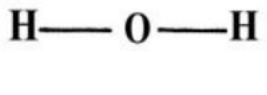


Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

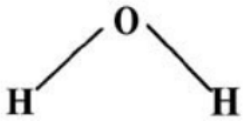
Co-funded by the Erasmus+ Programme of the European Union

Μοριακή γεωμετρία

Η παρατηρούμενη γεωμετρία ενός μορίου είναι αυτή που καθιστά την ενέργεια του συστήματος ελάχιστη. Έτσι θα ευνοηθούν εκείνες οι γεωμετρίες που (i) συγκεντρώνουν τη μεγαλύτερη ποσότητα πυκνότητας φορτίου στην περιοχή δέσμευσης και έτσι δίνουν τους ισχυρότερους μεμονωμένους δεσμούς και (ii) διατηρούν τους πυρήνες όσο το δυνατόν πιο μακριά (σύμφωνα με το (i)) και ως εκ τούτου να μειώσει τις πυρηνικές αντιδράσεις . Εξετάστε ξανά τις δύο πιθανότητες για το μόριο του νερού Είναι σαφές ότι η γραμμική μορφή (α) θα έχει μικρότερη ενέργεια πυρηνικής απώθησης από τα υδρογόνα από την λυγισμένη μορφή (β).



(a)



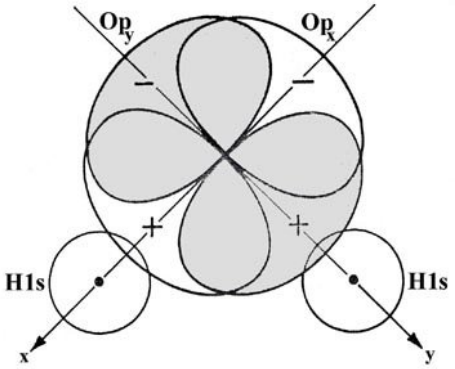
(b)

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Μοριακή γεωμετρία

Εάν η ποσότητα της πυκνότητας ηλεκτρονίων που θα μπορούσε να συγκεντρωθεί στις περιοχές μεταξύ των πυρήνων σε κάθε δεσμό OH (δηλαδή, η ισχύς κάθε δεσμού OH) ήταν ανεξάρτητη από τη γωνία του δεσμού, τότε σαφώς η γραμμική μορφή του μορίου του νερού θα ήταν η πιο σταθερός. Αυτή θα ήταν η κατάσταση εάν όλα τα ατομικά τροχιακά που περιγράφουν τις κινήσεις των ηλεκτρονίων ήταν άκαμπτα σφαιρικά και επικεντρωμένα στους πυρήνες. Αυτό όμως δεν ισχύει. Όπως τονίστηκε νωρίτερα στη συζήτησή μας για τα ατομικά τροχιακά, η κίνηση των ηλεκτρονίων που διαθέτουν γωνιακή ορμή επειδή καταλαμβάνουν τροχιακά με l 1 0 συγκεντρώνεται κατά μήκος ορισμένων αξόνων ή επιπέδων στο διάστημα. Συγκεκριμένα, τα τρία τροχιακά p είναι το μέγιστο κατά μήκος των τριών κάθετων αξόνων στο χώρο. Η θεωρία του δεσμού σθένους του μορίου του νερού περιγράφει τους δύο δεσμούς OH ως προκύπτοντες από την επικάλυψη των τροχιακών H 1s με τα δύο μισογεμάτα τροχιακά 2p του ατόμου οξυγόνου. Δεδομένου ότι τα δύο τροχιακά 2p βρίσκονται σε ορθή γωνία μεταξύ τους, η θεωρία του δεσμού σθένους προβλέπει μια λυγισμένη γεωμετρία για το μόριο του νερού με γωνία δεσμού 90°.

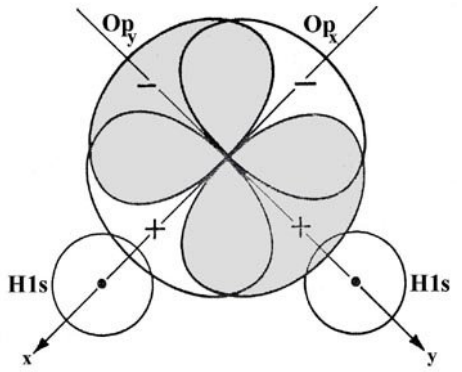


Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Μοριακή γεωμετρία

Η πραγματική γωνία δεσμού στο μόριο του νερού είναι 104,5°. Το άνοιγμα της γωνίας σε μια τιμή μεγαλύτερη από την προβλεπόμενη των 90° μπορεί να ληφθεί υπόψη από την άποψη της μείωσης της απώθησης μεταξύ των πυρήνων του υδρογόνου. Η υπόθεση που κάναμε είναι ότι η μέγιστη ποσότητα πυκνότητας ηλεκτρονίων θα μεταφερθεί στην περιοχή δέσμευσης και ως εκ τούτου θα αποδώσει τον ισχυρότερο δυνατό δεσμό όταν οι πυρήνες υδρογόνου και οξυγόνου βρίσκονται στον άξονα που ορίζεται από την κατεύθυνση του τροχιακού 2p. Για έναν δεδομένο διαπυρηνικό διαχωρισμό, αυτό θα έχει ως αποτέλεσμα τη μέγιστη επικάλυψη των τροχιακών. Επειδή ένα τροχιακό με l 1 0 περιορίζει την κίνηση του ηλεκτρονίου σε ορισμένες προτιμώμενες κατευθύνσεις στο διάστημα, οι γωνίες των δεσμών και η μοριακή γεωμετρία θα καθοριστούν, σε μια πρώτη χονδρική προσέγγιση, από τις διατροχιακές γωνίες.

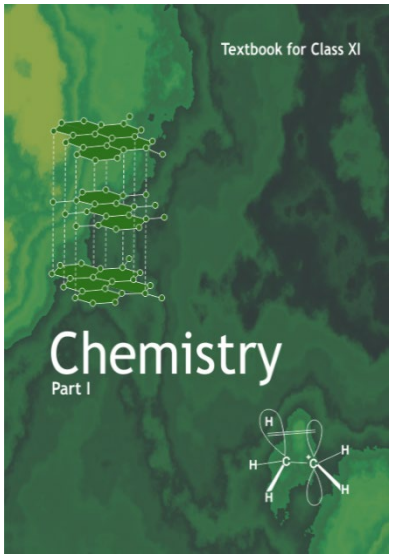


Co-funded by the Erasmus+ Programme of the European Union

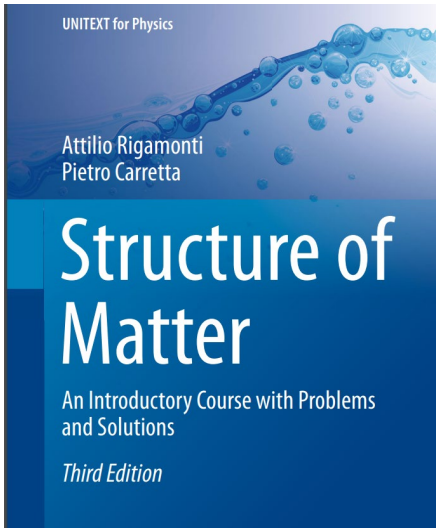
Ερωτήσεις

- Πώς σχηματίζονται οι χημικοί δεσμοί;
- Εξηγήστε τον όρο μοριακά τροχιακά
- Εξηγήστε τον υβριδισμό
- Από τι εξαρτάται η γεωμετρία των μορίων;

Co-funded by the Erasmus+ Programme of the European Union



https://www.neetprep.com/ncert/1591-Chemical-Bonding-Molecular-Structure-Chemical-Bonding-Molecular-Structure--NCERT-Chapter-PDF



Rigamonti , A., & Carretta , P. (2015). Δομή της Ύλης. UNITEXT for Physics, Springer International Publishing Switzerland

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 1. Δομή της Ύλης

Μάθημα 3. Διαμοριακές δυνάμεις. Δεσμοί υδρογόνου, ηλεκτροστατικές και αλληλεπιδράσεις van der Waals

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Δεσμός υδρογόνου
- Υδροφοβικές Αλληλεπιδράσεις
- Δυνάμεις Van der Waals
- Συγκεκριμένες αλληλεπιδράσεις

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Οι διαμοριακές δυνάμεις είναι οι ελκτικές ή απωστικές δυνάμεις μεταξύ των μορίων. Χωρίζονται σε δύο ομάδες. Δυνάμεις μικρής εμβέλειας και μεγάλης εμβέλειας. Δυνάμεις μικρής εμβέλειας συμβαίνουν όταν τα κέντρα των μορίων χωρίζονται από τρία angstroms (10^{-8} cm) ή λιγότερο. Οι δυνάμεις μικρής εμβέλειας τείνουν να είναι απωστικές, ενώ οι δυνάμεις μεγάλης εμβέλειας που δρουν έξω από το εύρος των τριών angstroms είναι ελκτικές. Οι δυνάμεις μεγάλης εμβέλειας είναι επίσης γνωστές ως δυνάμεις Van der Waals. Είναι υπεύθυνες για την επιφανειακή τάση, την τριβή, το ιξώδες και τις διαφορές μεταξύ της πραγματικής συμπεριφοράς των αερίων και αυτής που προβλέπεται από τον νόμο των ιδανικών αερίων. Οι διαμοριακές δυνάμεις είναι υπεύθυνες για τις περισσότερες ιδιότητες όλων των φάσεων. Το ιξώδες, η διάχυση και η επιφανειακή τάση είναι παραδείγματα φυσικών ιδιοτήτων υγρών που εξαρτώνται από διαμοριακές δυνάμεις. Η τάση ατμών, το κρίσιμο σημείο και το σημείο βρασμού είναι παραδείγματα ιδιοτήτων των αερίων. Η τήξη και η εξάχνωση είναι παραδείγματα ιδιοτήτων στερεών που εξαρτώνται από διαμοριακές δυνάμεις.

Co-funded by the Erasmus+ Programme of the European Union

Δεσμός υδρογόνου

Ο δεσμός υδρογόνου είναι ένας ειδικός τύπος έλξης διπόλου-διπόλου που συμβαίνει όταν ένα άτομο υδρογόνου συνδεδεμένο με ένα ισχυρά ηλεκτραρνητικό άτομο υπάρχει κοντά σε ένα άλλο ηλεκτραρνητικό άτομο με ένα μόνο ζεύγος ηλεκτρονίων. Αυτοί οι δεσμοί είναι γενικά ισχυρότεροι από τις συνηθισμένες δυνάμεις διπόλου-διπόλου και διασποράς, αλλά πιο αδύναμοι από τους αληθινούς ομοιοπολικούς και ιοντικούς δεσμούς .

Για να προκύψει ένας δεσμός υδρογόνου πρέπει να υπάρχει και ένας δότης και ένας δέκτης υδρογόνου. Ο δότης σε έναν δεσμό υδρογόνου είναι το άτομο στο οποίο συνδέεται ομοιοπολικά το άτομο υδρογόνου που συμμετέχει στον δεσμό υδρογόνου και είναι συνήθως ένα έντονα ηλεκτραρνητικό άτομο όπως το N, O ή F. Ο δέκτης υδρογόνου είναι το γειτονικό ηλεκτραρνητικό ιόν ή μόριο, και πρέπει να έχει ένα μοναχικό ζεύγος ηλεκτρονίων για να σχηματίσει δεσμό υδρογόνου.

Δεδομένου ότι ο δότης υδρογόνου είναι έντονα ηλεκτραρνητικός, τραβά το ζεύγος ηλεκτρονίων με ομοιοπολικό δεσμό πιο κοντά στον πυρήνα του και μακριά από το άτομο του υδρογόνου. Στη συνέχεια, το άτομο υδρογόνου αφήνεται με ένα μερικό θετικό φορτίο, δημιουργώντας μια έλξη διπόλου-διπόλου μεταξύ του ατόμου υδρογόνου που συνδέεται με τον δότη και του μοναδικού ζεύγους ηλεκτρονίων στον δέκτη.

Co-funded by the Erasmus+ Programme of the European Union

Τύποι δεσμών υδρογόνου

Οι δεσμοί υδρογόνου μπορούν να συμβούν μέσα σε ένα μόνο μόριο, μεταξύ δύο όμοιων μορίων ή μεταξύ δύο διαφορετικών μορίων.

Ενδομοριακοί δεσμοί υδρογόνου: Οι ενδομοριακοί δεσμοί υδρογόνου είναι εκείνοι που εμφανίζονται μέσα σε ένα μόνο μόριο. Αυτό συμβαίνει όταν δύο λειτουργικές ομάδες ενός μορίου μπορούν να σχηματίσουν δεσμούς υδρογόνου μεταξύ τους. Για να συμβεί αυτό, τόσο ένας δότης υδρογόνου όσο και ένας δέκτης πρέπει να είναι παρόντες μέσα σε ένα μόριο και πρέπει να βρίσκονται σε κοντινή απόσταση μεταξύ τους στο μόριο. Για παράδειγμα, ο ενδομοριακός δεσμός υδρογόνου συμβαίνει στην αιθυλενογλυκόλη ($C_2H_4(OH)_2$) μεταξύ των δύο υδροξυλομάδων της λόγω της μοριακής γεωμετρίας.

Διαμοριακοί δεσμοί υδρογόνου: Οι διαμοριακοί δεσμοί υδρογόνου εμφανίζονται μεταξύ χωριστών μορίων μιας ουσίας. Μπορούν να εμφανιστούν μεταξύ οποιουδήποτε αριθμού όμοιων ή μη μορίων, εφόσον υπάρχουν δότες και αποδέκτες υδρογόνου και σε θέσεις στις οποίες μπορούν να αλληλεπιδράσουν. Για παράδειγμα, διαμοριακούς δεσμούς υδρογόνου μπορούν να προκύψουν μεταξύ NH_3 μόνο μόρια, μόνο μεταξύ μορίων H_2O ή μεταξύ NH_3 και μόρια

Co-funded by the Erasmus+ Programme of the European Union

Ιδιότητες και επιδράσεις των δεσμών υδρογόνου

- υψηλότερα σημεία βρασμού
- υψηλότερο ιξώδες
- Ηλεκτραρνητικότητα: Ο δεσμός υδρογόνου δεν μπορεί να συμβεί χωρίς σημαντικές διαφορές ηλεκτραρνητικότητας μεταξύ του υδρογόνου και του ατόμου με το οποίο συνδέεται .
- Μέγεθος ατόμου: Το μέγεθος των δοτών και των αποδεκτών μπορεί επίσης να επηρεάσει την ικανότητα δεσμού υδρογόνου.

Δεσμός υδρογόνου στη φύση

Ο δεσμός υδρογόνου παίζει καθοριστικό ρόλο σε πολλές βιολογικές διεργασίες και μπορεί να ευθύνεται για πολλά φυσικά φαινόμενα, όπως οι ασυνήθιστες ιδιότητες του νερού. Εκτός από το ότι υπάρχει στο νερό, ο δεσμός υδρογόνου είναι επίσης σημαντικός στο σύστημα μεταφοράς νερού των φυτών, στη δομή της δευτερογενούς και τριτογενούς πρωτεΐνης και στο ζεύγος βάσεων DNA.

- Φυτά
- Πρωτεΐνες
- DNA

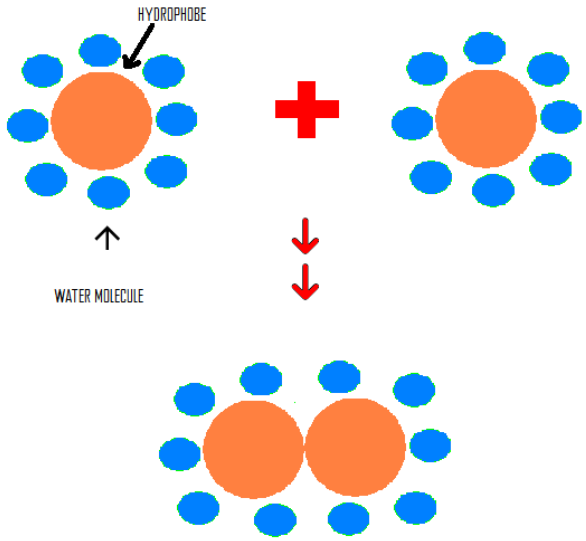
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Υδροφοβικές Αλληλεπιδράσεις

Οι υδρόφοβες αλληλεπιδράσεις περιγράφουν τις σχέσεις μεταξύ νερού και υδρόφοβων (χαμηλά υδατοδιαλυτά μόρια). Τα υδρόφοβα είναι μη πολικά μόρια και συνήθως έχουν μια μακρά αλυσίδα ανθράκων που δεν αλληλεπιδρούν με τα μόρια του νερού. Η ανάμειξη λίπους και νερού είναι ένα καλό παράδειγμα αυτής της συγκεκριμένης αλληλεπίδρασης. Η κοινή παρανόηση είναι ότι το νερό και το λίπος δεν αναμειγνύονται επειδή οι δυνάμεις Van der Waals που δρουν τόσο στα μόρια του νερού όσο και στα μόρια λίπους είναι πολύ αδύναμες. Ωστόσο, αυτό δεν ισχύει. Η συμπεριφορά μιας σταγόνας λίπους στο νερό έχει να κάνει περισσότερο με την ενθαλπία και την εντροπία της αντίδρασης παρά με τις διαμοριακές της δυνάμεις .

Ο Αμερικανός χημικός Walter Kauzmann ανακάλυψε ότι οι μη πολικές ουσίες όπως τα μόρια λίπους τείνουν να συσσωρεύονται μεταξύ τους αντί να κατανέμονται σε ένα υδάτινο μέσο, επειδή αυτό επιτρέπει στα μόρια λίπους να έχουν ελάχιστη επαφή με το νερό.



Ισχύς Υδροφοβικών Αλληλεπιδράσεων

Οι υδρόφοβες αλληλεπιδράσεις είναι σχετικά ισχυρότερες από άλλες ασθενείς διαμοριακές δυνάμεις (δηλαδή, αλληλεπιδράσεις Van der Waals ή δεσμούς υδρογόνου). Η ισχύς των υδροφοβικών αλληλεπιδράσεων εξαρτάται από διάφορους παράγοντες, όπως (κατά σειρά ισχύος επιρροής):

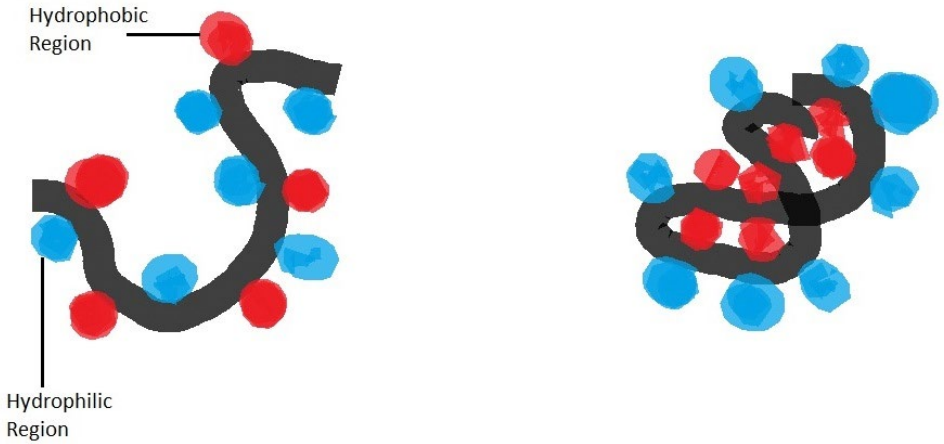
1. Θερμοκρασία: Καθώς η θερμοκρασία αυξάνεται, η ισχύς των υδρόφοβων αλληλεπιδράσεων αυξάνεται επίσης. Ωστόσο, σε ακραίες θερμοκρασίες, οι υδρόφοβες αλληλεπιδράσεις θα μετουσιωθούν.
2. Αριθμός ανθράκων στα υδρόφοβα : Τα μόρια με τον μεγαλύτερο αριθμό ανθράκων θα έχουν τις ισχυρότερες υδρόφοβες αλληλεπιδράσεις.
3. Το σχήμα των υδρόφοβων: Τα αλειφατικά οργανικά μόρια έχουν ισχυρότερες αλληλεπιδράσεις από τις αρωματικές ενώσεις. Οι κλάδοι σε μια ανθρακική αλυσίδα θα μειώσουν την υδρόφοβη επίδραση αυτού του μορίου και η γραμμική αλυσίδα άνθρακα μπορεί να προκαλέσει τη μεγαλύτερη υδρόφοβη αλληλεπίδραση. Αυτό συμβαίνει επειδή τα κλαδιά άνθρακα παράγουν στερικό εμπόδιο, επομένως είναι πιο δύσκολο για δύο υδρόφοβα να έχουν πολύ στενές αλληλεπιδράσεις μεταξύ τους για να ελαχιστοποιηθεί η επαφή τους με το νερό.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Βιολογική σημασία των υδροφοβικών αλληλεπιδράσεων

Οι υδρόφοβες αλληλεπιδράσεις είναι σημαντικές για την αναδίπλωση των πρωτεϊνών. Αυτό είναι σημαντικό για τη διατήρηση μιας πρωτεΐνης σταθερής και βιολογικά ενεργής, επειδή επιτρέπει στην πρωτεΐνη να μειώνεται στην επιφάνεια και να μειώνει τις ανεπιθύμητες αλληλεπιδράσεις με το νερό. Εκτός από τις πρωτεΐνες, υπάρχουν πολλές άλλες βιολογικές ουσίες που βασίζονται σε υδρόφοβες αλληλεπιδράσεις για την επιβίωση και τις λειτουργίες τους, όπως οι μεμβράνες διπλής στοιβάδας φωσφολιπιδίων σε κάθε κύτταρο του σώματός σας!



Co-funded by the Erasmus+ Programme of the European Union

Αλληλεπιδράσεις Van Der Waals

Οι δυνάμεις Van der Waals οδηγούνται από επαγόμενες ηλεκτρικές αλληλεπιδράσεις μεταξύ δύο ή περισσότερων ατόμων ή μορίων που βρίσκονται πολύ κοντά το ένα στο άλλο. Η αλληλεπίδραση Van der Waals είναι η πιο αδύναμη από όλες τις διαμοριακές έλξεις μεταξύ των μορίων. Ωστόσο, με πολλές δυνάμεις Van der Waals που αλληλεπιδρούν μεταξύ δύο αντικειμένων, η αλληλεπίδραση μπορεί να είναι πολύ ισχυρή.

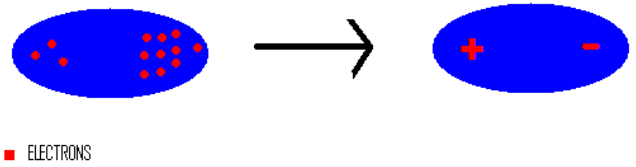
Αιτίες των δυνάμεων Van der Waals

Η Κβαντομηχανική δίνει έμφαση στη σταθερή κίνηση των ηλεκτρονίων σε ένα άτομο μέσω της Εξίσωσης Schrödinger και της Αρχής της Αβεβαιότητας του Heisenberg. Η Αρχή της Αβεβαιότητας του Heisenberg προτείνει ότι η ενέργεια του ηλεκτρονίου δεν είναι ποτέ μηδενική. ως εκ τούτου, κινείται συνεχώς γύρω από την τροχιακή του. Το τετράγωνο της εξίσωσης Schrödinger για ένα σωματίδιο σε ένα κουτί υποδηλώνει ότι είναι πιθανό να βρεθεί το ηλεκτρόνιο (σωματίδιο) οπουδήποτε στο τροχιακό του ατόμου (κουτί).

Αυτές οι δύο σημαντικές πτυχές της Κβαντικής Μηχανικής υποδηλώνουν έντονα ότι τα ηλεκτρόνια κινούνται συνεχώς σε ένα άτομο, επομένως είναι πιθανό να συμβούν δίπολα. Ως δίπολο ορίζονται τα μόρια ή τα άτομα με ίσα και αντίθετα ηλεκτρικά φορτία που χωρίζονται από μια μικρή απόσταση.

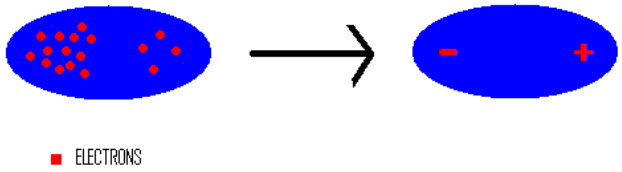
Είναι πιθανό να βρεθούν τα ηλεκτρόνια

σε αυτή την κατάσταση :



Έτσι συμβαίνουν τα αυθόρμητα (ή στιγμιαία) δίπολα. Όταν ομάδες ηλεκτρονίων μετακινούνται στο ένα άκρο του ατόμου, δημιουργείται ένα δίπολο. Αυτές οι ομάδες ηλεκτρονίων κινούνται συνεχώς και έτσι μετακινούνται από το ένα άκρο του ατόμου στο άλλο και πάλι πίσω συνεχώς. Επομένως, η αντίθετη κατάσταση είναι εξίσου πιθανό να συμβεί.

Αντίθετη κατάσταση λόγω διακύμανσης διπόλων:

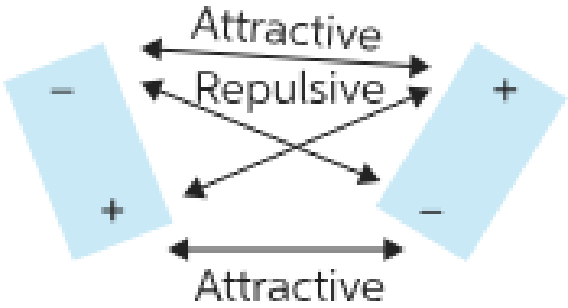


Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ηλεκτροστατικές αλληλεπιδράσεις

Οι διαμοριακές δυνάμεις έλξης στις οποίες έλκονται πλήρη ή μερικά ιοντικά είδη μεταξύ τους ονομάζονται ηλεκτροστατικές αλληλεπιδράσεις. Αυτές οι δυνάμεις έλξης δεν περιλαμβάνουν καμία κατανομή ηλεκτρονίων μεταξύ των ατόμων. Έτσι, ονομάζονται και ως μη ομοιοπολικοί δεσμοί. Ο όρος ηλεκτροστατικές αλληλεπιδράσεις περιλαμβάνει τόσο ελκτικές όσο και απωθητικές δυνάμεις ιοντικά είδη, που σημαίνει ότι ιόντα με αντίθετα φορτία έλκονται μεταξύ τους, ενώ τα παρόμοια φορτία απωθούνται το ένα από το άλλο. Στα υγρά, τα μόρια που συγκρατούνται μεταξύ τους με διαμοριακές αλληλεπιδράσεις είναι συγκριτικά πιο αδύναμα από τις ενδομοριακές αλληλεπιδράσεις εντός πολυατομικών ιόντων και ατόμων μορίων. Αυτά είναι πολύ σημαντικά για την περιγραφή του σχηματισμού διαφορετικών μορίων. Λόγω αλλαγών στις διαμοριακές αλληλεπιδράσεις, οι μεταβάσεις συμβαίνουν μεταξύ της στερεάς και της υγρής ή της υγρής και της αέριας φάσης.



Τύποι ηλεκτροστατικών αλληλεπιδράσεων

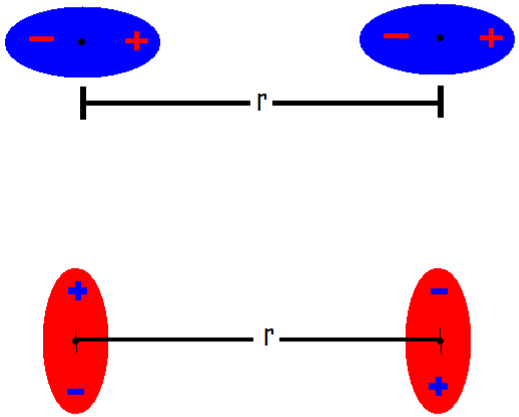
- Οι αλληλεπιδράσεις μεταξύ δύο διπολικών μορίων μέσω του χώρου ονομάζονται **αλληλεπίδραση διπόλου-διπόλου** .
- Το 1930, ο Fritz London, ένας Γερμανός φυσικός δήλωσε ότι υπάρχουν μη μόνιμες διακυμάνσεις κατανομής των ηλεκτρονίων μέσα στα άτομα και τα μη πολικά μόρια, που οδηγούν σε βραχύβιες γρήγορες ροπές διπλού σχηματισμού και παράγουν ελκτικές δυνάμεις γνωστές ως δυνάμεις διασποράς του **Λονδίνου** .
- Ένας ισχυρός τύπος αλληλεπίδρασης διπόλου-διπόλου ονομάζεται **δεσμός υδρογόνου** .
- **Αλληλεπίδραση ιόντων-ιόντων** Αυτοί οι τύποι αλληλεπιδράσεων είναι υπεύθυνοι για τους ιοντικούς δεσμούς.
- **Αλληλεπίδραση ιόντων-διπόλου** Όταν ιονικές ουσίες διαλύονται σε μια πολική ουσία, όπως το νερό και οι αλληλεπιδράσεις ιόντων-διπόλου θα λαμβάνουν χώρα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Αλληλεπίδραση Διπόλου-Διπόλου

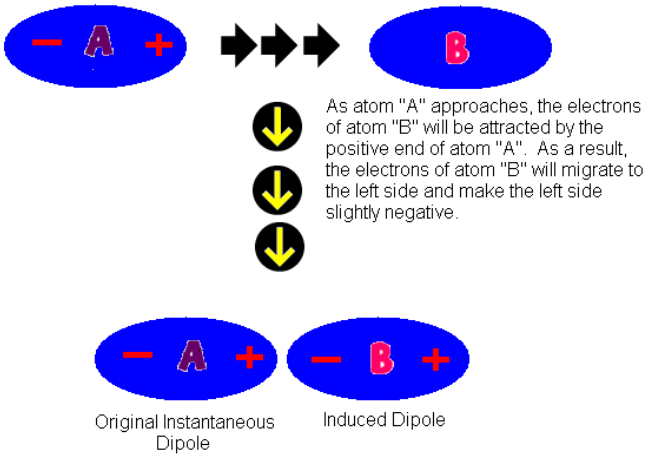
Οι αλληλεπιδράσεις διπόλου-διπόλου συμβαίνουν μεταξύ μορίων που έχουν μόνιμα δίπολα. Αυτά τα μόρια αναφέρονται επίσης ως πολικά μόρια. Το παρακάτω σχήμα δείχνει την ηλεκτροστατική αλληλεπίδραση μεταξύ δύο διπόλων.



Co-funded by the Erasmus+ Programme of the European Union

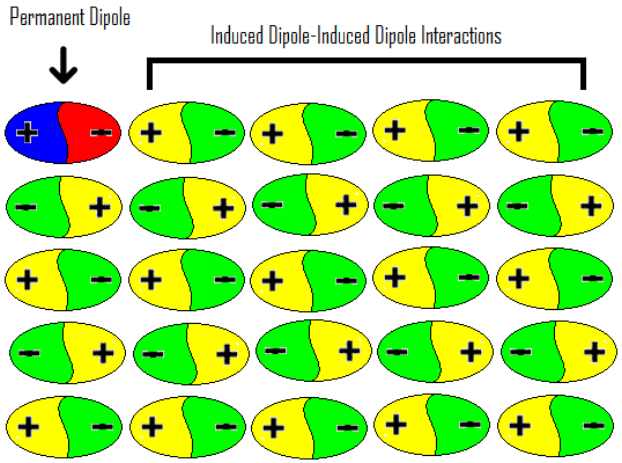
Επαγόμενα δίπολα

Μια επαγόμενη διπολική ροπή είναι μια προσωρινή κατάσταση κατά την οποία ένα ουδέτερο μη πολικό άτομο (δηλ. Ήλιο) υφίσταται διαχωρισμό φορτίων λόγω του περιβάλλοντος. Όταν ένα στιγμιαίο άτομο διπόλου πλησιάζει ένα γειτονικό άτομο, μπορεί να προκαλέσει και αυτό το άτομο να παράγει δίπολα. Το γειτονικό άτομο θεωρείται τότε ότι έχει επαγόμενη διπολική ροπή.



Αυθόρμητη Διπολική Αλληλεπίδραση που προκαλείται από Διπόλο

Οι αυθόρμητες διπολικές αλληλεπιδράσεις που προκαλούνται από τα δίπολα είναι επίσης γνωστές ως διασπορά ή δυνάμεις του Λονδίνου (όνομα από τον Γερμανό φυσικό Fritz London). Είναι μεγάλα δίκτυα διαμοριακών δυνάμεων μεταξύ μη πολικών και μη φορτισμένων μορίων και ατόμων (δηλ. αλκάνια, ευγενή αέρια και αλογόνα). Τα μόρια που έχουν επαγόμενα δίπολα μπορεί επίσης να επάγουν γειτονικά μόρια να έχουν διπολικές ροπές, επομένως μπορεί να υπάρχει ένα μεγάλο δίκτυο επαγόμενων διπολικών αλληλεπιδράσεων διπόλων. Η παρακάτω εικόνα απεικονίζει ένα δίκτυο επαγόμενων διπολικών αλληλεπιδράσεων διπόλων.



Co-funded by the Erasmus+ Programme of the European Union

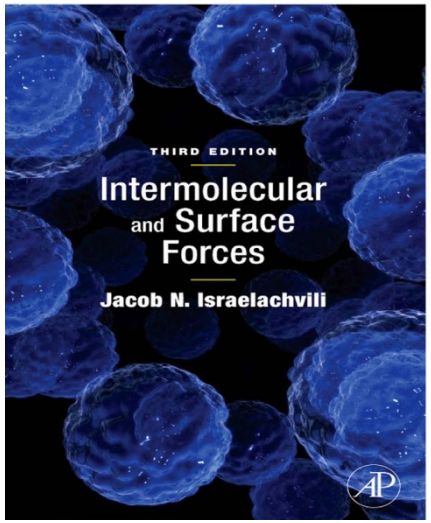
Εφαρμογές Ηλεκτροστατικών αλληλεπιδράσεων

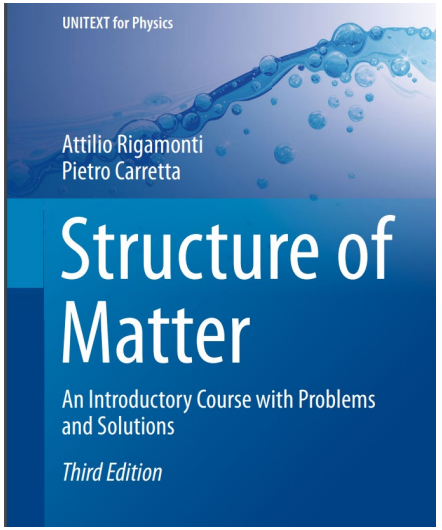
Οι ηλεκτροστατικές αλληλεπιδράσεις γίνονται σημαντικές για τα αέρια σε υψηλές πιέσεις, καθώς αυτά οφείλονται στις παρατηρούμενες αποκλίσεις τους από τον νόμο του ιδανικού αερίου στις υψηλές πιέσεις. Τέτοιες αλληλεπιδράσεις είναι σημαντικές για τη διατήρηση της τρισδιάστατης δομής μεγαλύτερων μορίων. Για παράδειγμα, νουκλεϊκά οξέα και πρωτεΐνες. Σε πολλές βιολογικές διεργασίες, όπου μεγαλύτερα μόρια συνδέονται ειδικά αλλά παροδικά σε ένα, εμπλέκονται τέτοιες αλληλεπιδράσεις. Επιπλέον, αυτές οι αλληλεπιδράσεις επηρεάζουν επίσης βαθιά την κρυσταλλικότητα και το σχεδιασμό των υλικών, γενικά, για τη σύνθεση πολλών οργανικών μορίων και ιδιαίτερα για την αυτοσυναρμολόγηση.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Ποιοι τύποι ηλεκτροστατικών αλληλεπιδράσεων υπάρχουν;
- Εξηγήστε τα χαρακτηριστικά των αλληλεπιδράσεων van der Waals
- Εξηγήστε τη σημασία των δεσμών υδρογόνου στην επιστήμη της ζωής





Jacob N. Israelachvili (2011) Διαμοριακές και επιφανειακές δυνάμεις Elsevier Inc.

Rigamonti , A., & Carretta , P. (2015). Δομή της Ύλης. UNITEXT for Physics, Springer International Publishing Switzerland

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 1. Δομή της ύλης

Μάθημα 4. Κίνηση Brown και κινητική θεωρία

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

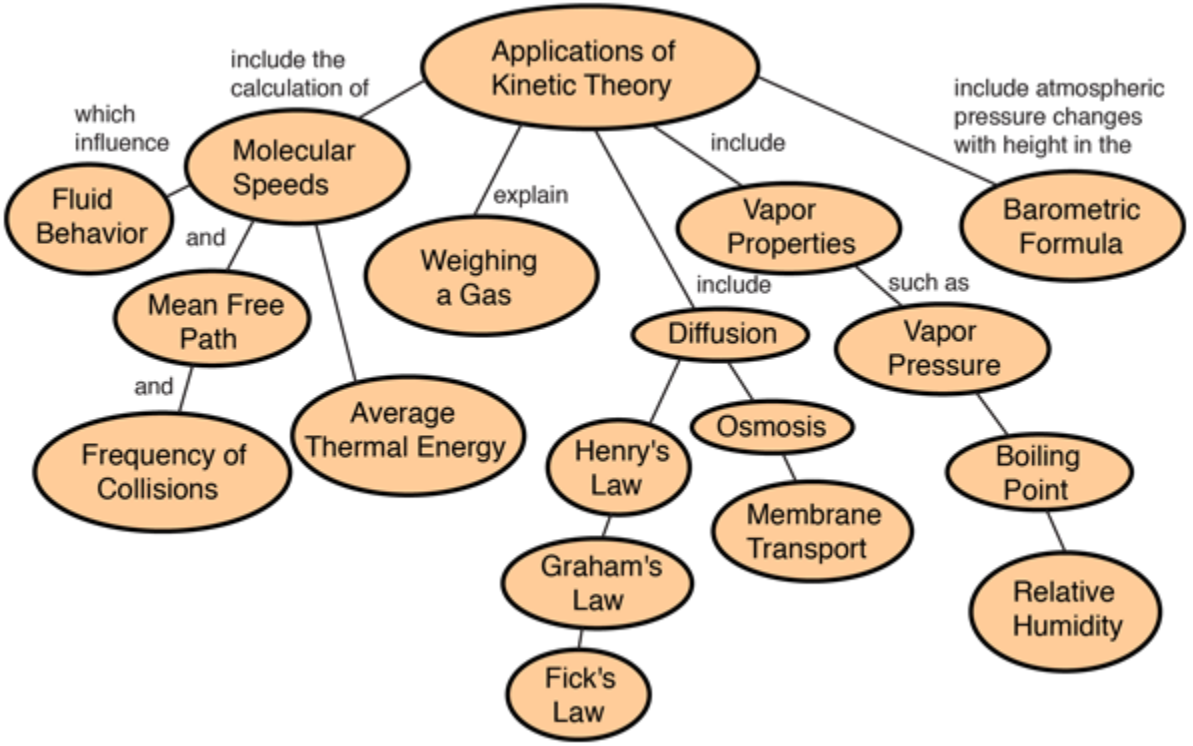
- Εισαγωγή
- Κινητική Θεωρία
- Το Ιδανικό Αέριο
- Brownian Motion
- Η Δομή των Αερίων και των Υγρών

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Οι διαμοριακές δυνάμεις είναι οι ελκτικές ή απωστικές δυνάμεις μεταξύ των μορίων. Χωρίζονται σε δύο ομάδες. δυνάμεις μικρής εμβέλειας και μεγάλης εμβέλειας. Δυνάμεις μικρής εμβέλειας συμβαίνουν όταν τα κέντρα των μορίων χωρίζονται από τρία angstroms (10^{-8} cm) ή λιγότερο. Οι δυνάμεις μικρής εμβέλειας τείνουν να είναι απωστικές, όπου οι δυνάμεις μεγάλης εμβέλειας που δρουν έξω από το εύρος των τριών angstroms είναι ελκυστικές. Οι δυνάμεις μεγάλης εμβέλειας είναι επίσης γνωστές ως δυνάμεις Van der Waals. Είναι υπεύθυνα για την επιφανειακή τάση, την τριβή, το ιξώδες και τις διαφορές μεταξύ της πραγματικής συμπεριφοράς των αερίων και αυτής που προβλέπεται από τον νόμο των ιδανικών αερίων. Οι διαμοριακές δυνάμεις είναι υπεύθυνες για τις περισσότερες ιδιότητες όλων των φάσεων. Το ιξώδες, η διάχυση και η επιφανειακή τάση είναι παραδείγματα φυσικών ιδιοτήτων υγρών που εξαρτώνται από διαμοριακές δυνάμεις. Η τάση ατμών, το κρίσιμο σημείο και το σημείο βρασμού είναι παραδείγματα ιδιοτήτων των αερίων. Η τήξη και η εξάχνωση είναι παραδείγματα ιδιοτήτων στερεών που εξαρτώνται από διαμοριακές δυνάμεις.

Εφαρμογή Υποθέσεων της Κινητικής Θεωρίας



Υποθέσεις της Κινητικής Μοριακής Θεωρίας

Η κινητική θεωρία περιλαμβάνει μια σειρά από υποθέσεις που εστιάζουν στο να μπορούμε να μιλάμε για ένα ιδανικό αέριο.

- Τα μόρια αντιμετωπίζονται ως σημειακά σωματίδια. Συγκεκριμένα, μια συνέπεια αυτού είναι ότι το μέγεθός τους είναι εξαιρετικά μικρό σε σύγκριση με τη μέση απόσταση μεταξύ των σωματιδίων.
- Ο αριθμός των μορίων (N) είναι πολύ μεγάλος, σε βαθμό που δεν είναι δυνατή η παρακολούθηση μεμονωμένων συμπεριφορών σωματιδίων. Αντίθετα, εφαρμόζονται στατιστικές μέθοδοι για την ανάλυση της συμπεριφοράς του συστήματος στο σύνολό του.
- Κάθε μόριο αντιμετωπίζεται ως πανομοιότυπο με οποιοδήποτε άλλο μόριο. Είναι εναλλάξιμα ως προς τις διάφορες ιδιότητές τους. Αυτό βοηθά και πάλι στην υποστήριξη της ιδέας ότι τα μεμονωμένα σωματίδια δεν χρειάζεται να παρακολουθούνται και ότι οι στατιστικές μέθοδοι της θεωρίας είναι επαρκείς για να καταλήξουμε σε συμπεράσματα και προβλέψεις.
- Τα μόρια βρίσκονται σε συνεχή, τυχαία κίνηση. Υπακούουν στους νόμους της κίνησης του Νεύτωνα.
- Οι συγκρούσεις μεταξύ των σωματιδίων και μεταξύ των σωματιδίων και των τοιχωμάτων ενός δοχείου για το αέριο, είναι απόλυτα ελαστικές συγκρούσεις.
- Τα τοιχώματα των δοχείων αερίων αντιμετωπίζονται ως τέλεια άκαμπτα, δεν κινούνται και έχουν απείρως μάζα (σε σύγκριση με τα σωματίδια).

Co-funded by the Erasmus+ Programme of the European Union

Ο Νόμος του Ιδανικού Αερίου

Η κινητική θεωρία των αερίων είναι σημαντική, καθώς το σύνολο των παραδοχών παραπάνω μας οδηγεί να εξαγάγουμε τον νόμο του ιδανικού αερίου, ή την εξίσωση ιδανικού αερίου, που σχετίζεται με την πίεση (p), τον όγκο (V) και τη θερμοκρασία (T), με όρους της σταθεράς του Molar Gas, 8,31 J/ molK (R) και ο αριθμός των mol (n). Η εξίσωση ιδανικού αερίου που προκύπτει είναι:

$$pV = nRT$$

Co-funded by the Erasmus+ Programme of the European Union

Μέση Μοριακή Κινητική Ενέργεια

Η μέση κινητική ενέργεια κάθε μορίου αερίου σε ένα δοχείο μπορεί επίσης να υπολογιστεί χρησιμοποιώντας την παρακάτω εξίσωση:

$$E_k = \frac{1}{2}m(c_{\text{rms}})^2 = \frac{3}{2}kT = \frac{3}{2}\frac{RT}{N_A}$$

- $E_k =$ the **average kinetic energy** in joules (J)
- $m =$ the **mass of each gas molecule** in kilograms (kg)
- $k =$ the **Boltzmann constant** $(= 1.38 \times 10^{-23}\ \text{kgm}^2\text{s}^{-2}\text{K}^{-1})$
- $T =$ the **temperature** in kelvin (K)
- $R =$ the **molar gas constant** $(= 8.31\ \text{kgm}^2\text{s}^{-2}\text{K}^{-1}\text{mol}^{-1})$
- $N_A =$ **Avogadro's number** $(= 6.02 \times 10^{23})$

Co-funded by the Erasmus+ Programme of the European Union

Ρυθμοί διάχυσης και κινητική θεωρία

Διάχυση είναι η διαδικασία κατά την οποία τα μόρια μιας ουσίας αναμειγνύονται με αυτά μιας άλλης ως αποτέλεσμα της κίνησής τους. Οι ουσίες μετακινούνται ελεύθερα από μια περιοχή υψηλής συγκέντρωσης σε μια περιοχή χαμηλής συγκέντρωσης με τον δικό τους ρυθμό. Ο ρυθμός διάχυσης εξαρτάται από τη θερμοκρασία και την πυκνότητα των ουσιών που εμπλέκονται.

Διάχυση

Η διάχυση είναι η διαδικασία κατά την οποία τα μόρια αερίου αναμειγνύονται στενά μεταξύ τους ως αποτέλεσμα της κινητικής ενέργειας των μορίων. Σύμφωνα με το νόμο της διάχυσης του Γκράχαμ, ο χρόνος με τον οποίο διαχέονται τα αέρια είναι αντιστρόφως ανάλογος με την τετραγωνική ρίζα της πυκνότητάς τους, υπό την προϋπόθεση ότι η θερμοκρασία διατηρείται σταθερή. Αυτό γράφεται μαθηματικά ως:

tα1d√t= kd √

t = ρυθμός διάχυσης

d = μοριακή πυκνότητα του αερίου

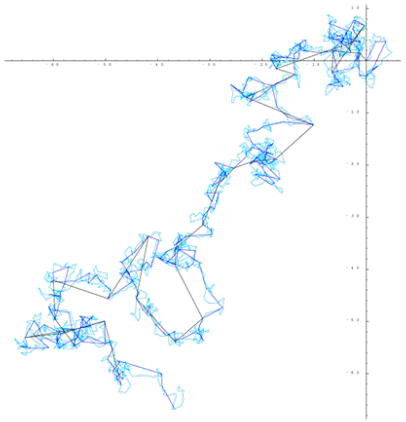
Αυτό σημαίνει ότι, όσο πιο βαρύ είναι το αέριο, τόσο πιο αργή είναι η κίνηση του αερίου.

Αξιώματα Κινητικής Θεωρίας Αερίων

- Τα αέρια αποτελούνται από μεγάλο αριθμό μικροσκοπικών σωματιδίων (άτομα και μόρια). Τέτοια σωματίδια είναι εξαιρετικά μικρά σε σύγκριση με την απόσταση μεταξύ τους. Το μεμονωμένο μέγεθος σωματιδίων ορίζεται ως αμελητέο και η πλειονότητα του όγκου που καταλαμβάνεται από το αέριο είναι κενός χώρος.
- Αυτά τα μόρια φαίνεται να βρίσκονται σε συνεχή τυχαία κίνηση, με αποτέλεσμα να συγκρούονται μεταξύ τους και με τα τοιχώματα του δοχείου. Κάθε φορά που τα μόρια του αερίου συγκρούονται με τα τοιχώματα ενός δοχείου, προσδίδουν ορμή στα τοιχώματα και αυτό έχει ως αποτέλεσμα τη δημιουργία μιας μετρήσιμης δύναμης. Έτσι, αν διαιρέσουμε αυτή τη δύναμη με την περιοχή, παίρνουμε την πίεση.
- Οι συγκρούσεις μεταξύ μορίων και τοιχωμάτων είναι εντελώς ελαστικές. Αυτό σημαίνει ότι όταν τα μόρια συγκρούονται, διατηρούν την κινητική τους ενέργεια. Τα μόρια δεν επιβραδύνουν ποτέ και θα κινούνται πάντα με τον ίδιο ρυθμό.
- Η μέση κινητική ενέργεια των σωματιδίων αερίου έχει επηρεαστεί από τη θερμοκρασία. Δηλαδή, όσο υψηλότερη είναι η θερμοκρασία, τόσο μεγαλύτερη είναι η μέση κινητική ενέργεια του αερίου.
- Εκτός από τη στιγμή που συγκρούονται, τα μόρια δεν έχουν καμία έλξη ή απώθηση το ένα για το άλλο.

Brownian Motion

Η κίνηση Brown είναι η φαινομενικά τυχαία κίνηση σωματιδίων, ατόμων ή μορίων που αναδύεται από τις τυχαίες συγκρούσεις αυτών των σωματιδίων. Η κίνηση Brown μπορεί να παρατηρηθεί καθώς το φως λάμπει μέσα από ένα παράθυρο. Σωματίδια σκόνης ή γύρης μπορούν να φανούν στο φως που επιπλέει στον αέρα, ακολουθώντας κάτι που φαίνεται να είναι τυχαία μοτίβα νευρικότητας. Τα σωματίδια της σκόνης δεν κινούνται μόνα τους, αλλά συγκρούονται με μόρια του αέρα κρατώντας τη σκόνη σε κίνηση.



Co-funded by the Erasmus+ Programme of the European Union

Αιτίες της κίνησης Brown

Προκειμένου να κατανοήσουμε την κίνηση Brown και πώς η κίνηση των σωματιδίων της σκόνης στον αέρα στην πραγματικότητα δεν είναι τόσο τυχαία, πρέπει να γίνουν κατανοητά μερικά πράγματα σχετικά με τα σωματίδια όπως τα άτομα και τα μόρια. Τα άτομα και τα μόρια συμπεριφέρονται με διαφορετικούς τρόπους ανάλογα με την κατάσταση της ύλης τους. Για να κατανοήσουμε την κίνηση Brown, πρέπει πρώτα να κατανοήσουμε τις δομές των αερίων και των υγρών και τον τρόπο με τον οποίο κινούνται τα μόρια ή τα άτομα τους.

Παραδείγματα κίνησης Brown

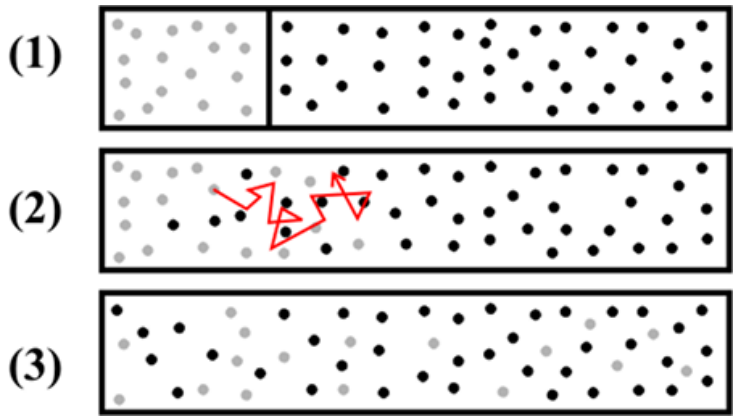
- Η κίνηση των κόκκων γύρης σε ακίνητο νερό.
- Κίνηση κηλίδων σκόνης σε ένα δωμάτιο (αν και επηρεάζεται σε μεγάλο βαθμό από τα ρεύματα αέρα)
- Διάχυση ρύπων στον αέρα.
- Διάχυση ασβεστίου μέσω των οστών.

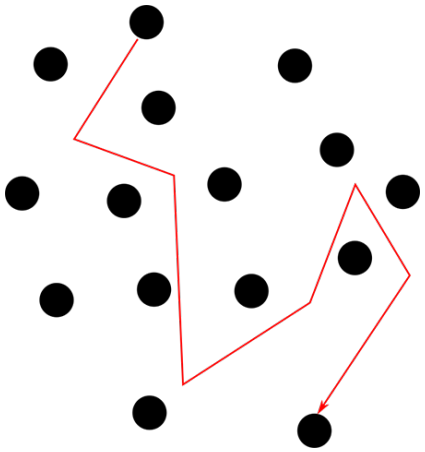
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Η Δομή των Αερίων και των Υγρών

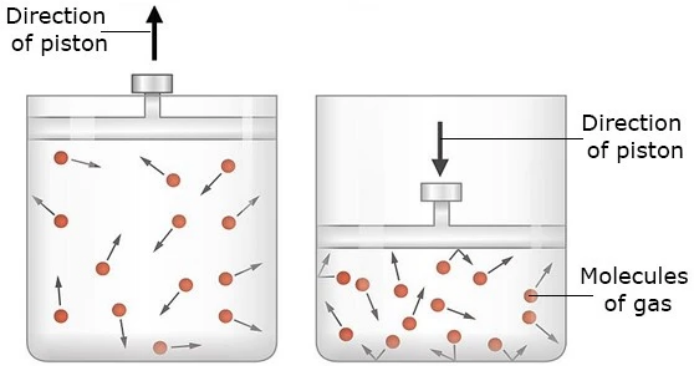
Όλη η ύλη βρίσκεται σε κίνηση σύμφωνα με την κινητική θεωρία. Τα σωματίδια ενός στερεού είναι συμπαγή και δεν κινούνται ελεύθερα το ένα δίπλα στο άλλο. Ωστόσο, τα μόρια ή τα άτομα ενός στερεού εξακολουθούν να δονούνται το οποίο αναδύεται και ανιχνεύεται ως θερμότητα. Αυτή η κινητική ενέργεια ή θερμότητα στα σωματίδια είναι πιο εμφανής και επακόλουθη σε υγρά ή αέρια. Η ύλη στο υγρό και τα αέρια κινούνται πιο ελεύθερα και τείνουν προς την ισορροπία. Δηλαδή, τα άτομα και τα μόρια σε υγρά και αέρια κινούνται προς περιοχές χαμηλότερων συγκεντρώσεων και μακριά από περιοχές υψηλότερων συγκεντρώσεων. Αυτό ονομάζεται διάχυση. Αυτό το φαινόμενο είναι παρατηρήσιμο στα υγρά όταν ένα υγρό διαλύεται σε ένα άλλο, και στα αέρια καθώς ο καπνός εξαπλώνεται για να γεμίσει ένα δωμάτιο.





Πίεση αερίου

Εάν ο όγκος μιας σταθερής μάζας ενός αερίου μειωθεί σε σταθερή θερμοκρασία, τα μόρια θα συμπιεστούν και θα συσκευαστούν πιο στενά μεταξύ τους όπως φαίνεται στο παρακάτω σχήμα. Υπό αυτή την προϋπόθεση, ο χρόνος σύγκρουσης των μορίων με τον εαυτό τους και τα τοιχώματα του δοχείου αυξάνεται, αυξάνοντας έτσι την πίεση του αερίου. Εάν από την άλλη, ο όγκος αυξηθεί ανυψώνοντας το έμβολο σε σταθερή θερμοκρασία, ο χρόνος σύγκρουσης των μορίων θα μειωθεί. Θα χρειαστούν περισσότερο χρόνο για να συγκρουστούν με τα τοιχώματα του δοχείου. Αυτό θα οδηγήσει σε μείωση της πίεσης. Αυτό συμφωνεί με το νόμο του Boyle που δηλώνει ότι «ο όγκος μιας σταθερής μάζας ενός αερίου σε σταθερή θερμοκρασία είναι αντιστρόφως ανάλογος της πίεσής του».



1. Volume of gas increased, pressure of gas decreased at constant temperature

2. Volume of gas reduced, pressure of gas increased at constant temperature

Ωσμωση

Η όσμωση είναι η τάση ενός διαλύτη να περάσει από ένα αραιό διάλυμα, μέσω μιας ημιπερατής μεμβράνης σε ένα συμπυκνωμένο διάλυμα. Η ημιπερατή μεμβράνη είναι μια ουσία όπως το σελοφάν, η περγαμηνή ή φυτικό υλικό που θα επέτρεπε σε ορισμένα μόρια υγρού να διαχυθούν μέσα από αυτά αλλά όχι σε άλλα. Μια τέτοια μεμβράνη μπορεί να επιτρέπει στα μόρια ενός διαλύτη να περάσουν μέσα από αυτήν αλλά όχι σε αυτά μιας διαλυμένης ουσίας.

Co-funded by the Erasmus+ Programme of the European Union

Ιξώδες

Το ιξώδες εξ ορισμού είναι η εσωτερική τριβή μεταξύ στρωμάτων ρευστών σε κίνηση. Τα υγρά που είναι πυκνά χύνονται πιο αργά από αυτά που είναι λιγότερο πυκνά. Πχ το μέλι θα χύνεται πιο αργά από το νερό γιατί είναι πιο πυκνό από την κηροζίνη. Αυτό σημαίνει ότι το μέλι είναι πιο παχύρρευστο υγρό από το νερό. Το ιξώδες μπορεί να αποδειχθεί αν σκεφτούμε ότι ένα ρουλεμάν πέφτει μέσα από ορισμένα υγρά.

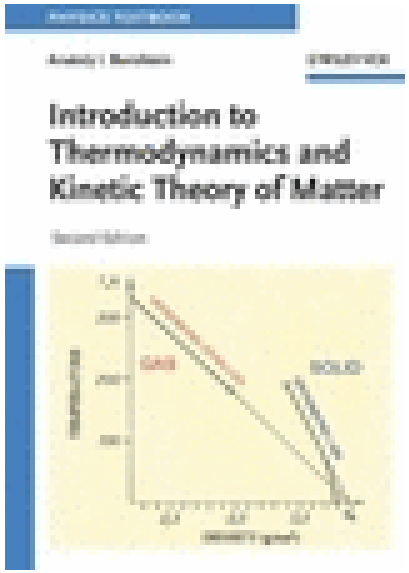
Co-funded by the Erasmus+ Programme of the European Union

Εφαρμογές Επιφανειακής Τάνυσης και Ιξώδους

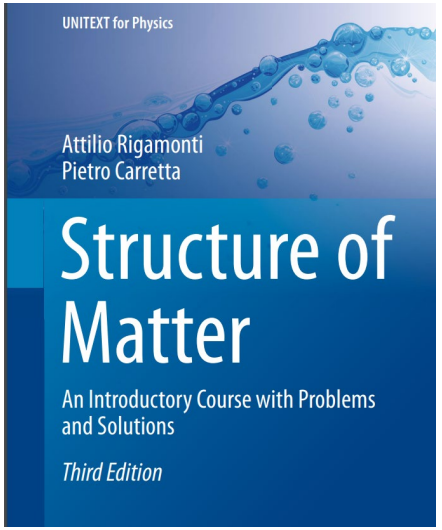
1. Η γνώση της επιφανειακής τάσης εφαρμόζεται σε βιομηχανίες στην κατασκευή ορισμένων υλικών όπως, ομπρέλες, καμβάς, παλτό βροχής και αδιάβροχες σκηνές
2. Είναι δύσκολο να πλένετε τα βρώμικα ρούχα ή τα λιπαρά ρούχα μόνο με νερό. Γι' αυτό χρησιμοποιούμε σαπούνι και απορρυπαντικά για το πλύσιμο. Το σαπούνι και τα απορρυπαντικά εξασθενούν την επιφανειακή τάση του νερού και του επιτρέπουν να επιπλέει μακριά από τη βρωμιά ή το λάδι από το υλικό
3. Τα παχύρρευστα υγρά χρησιμοποιούνται ως λιπαντικά. Παραδείγματα είναι το γράσο και το λάδι κινητήρα.
4. Το ιξώδες εφαρμόζεται στο σχεδιασμό σκαφών, πλοίων και αεροσκαφών.

Ερωτήσεις

- Ποιες είναι οι υποθέσεις της κινητικής μοριακής θεωρίας
- Εξηγήστε την Κινητική Θεωρία των Αερίων
- Εξηγήστε τα Παραδείγματα Κίνησης Brown



AI Burshstein (2005) Εισαγωγή στη Θερμοδυναμική και την Κινητική Θεωρία της Ύλης WILEY-VCH Verlag GmbH & Co. KGaA



Rigamonti , A., & Carretta , P. (2015). Δομή της Ύλης. UNITEXT for Physics, Springer International Publishing Switzerland

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 1. Δομή της ύλης

Μάθημα 5. Από το στερεό στο υγρό

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

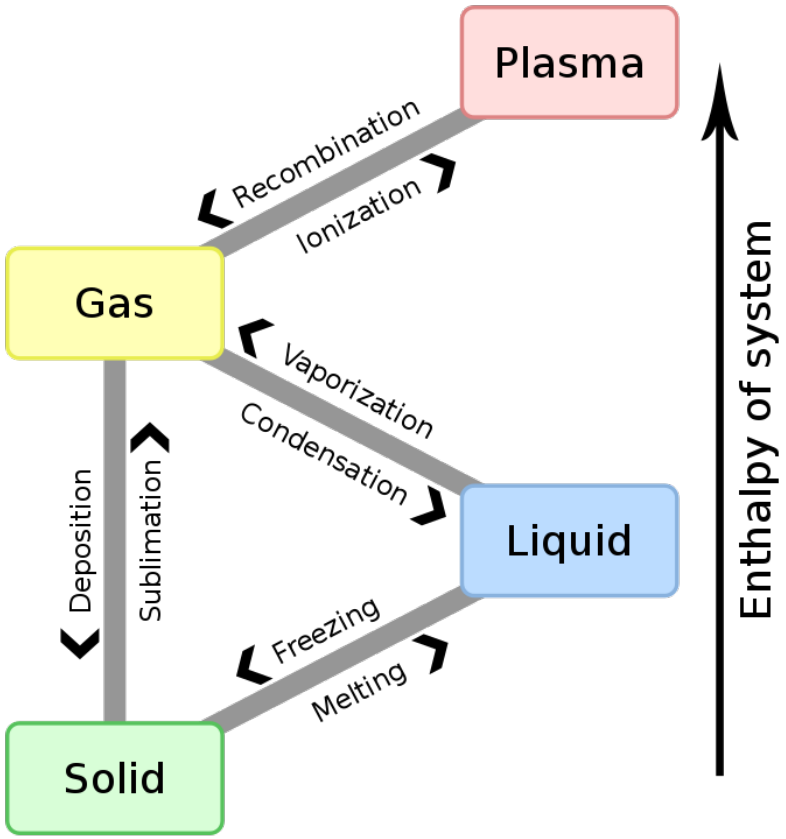
- Εισαγωγή
- Καταστάσεις της ύλης
- Χαρακτηριστικά των Τριών Καταστάσεων της Ύλης
- Ενθαλπία σύντηξης
- Αλλαγή φάσης υγρού αερίου
- Ενθαλπία εξάτμισης
- Καμπύλες θέρμανσης

Εισαγωγή

Οι ουσίες μπορούν να αλλάξουν φάση — συχνά λόγω αλλαγής θερμοκρασίας. Σε χαμηλές θερμοκρασίες, οι περισσότερες ουσίες είναι στερεές. Όσο αυξάνεται η θερμοκρασία, γίνονται υγρά. Σε υψηλότερες ακόμα θερμοκρασίες, γίνονται αέρια.

Co-funded by the Erasmus+ Programme of the European Union

Καταστάσεις της ύλης



Co-funded by the Erasmus+ Programme of the European Union

Η διαδικασία αλλαγής κατάστασης από στερεό σε υγρό ονομάζεται:

(I). Τήξη

(II). Πάγωμα

(III). Βρασμός

(IV). Συμπύκνωση

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Χαρακτηριστικά των Τριών Καταστάσεων της Ύλης

Χαρακτηριστικό γνώρισμα	Στερεός	Υγρό	Αέριο
σχήμα	σαφής	αόριστος	αόριστος
Ενταση ΗΧΟΥ	σαφής	σαφής	αόριστος
σχετική δύναμη διαμοριακής αλληλεπίδρασης	ισχυρός	μέτριος	αδύναμος
σχετικές θέσεις σωματιδίων	σε επαφή και στερεωμένο στη θέση του	σε επαφή αλλά όχι σταθερό	όχι σε επαφή, τυχαίες θέσεις

Co-funded by the Erasmus+ Programme of the European Union

Ενθαλπία σύντηξης

Η διαδικασία του να γίνει ένα στερεό υγρό ονομάζεται τήξη (ένας παλαιότερος όρος που μπορεί να δείτε μερικές φορές είναι σύντηξη). Η αντίθετη διαδικασία, ένα υγρό που γίνεται στερεό, ονομάζεται στερεοποίηση. Για κάθε καθαρή ουσία, η θερμοκρασία στην οποία συμβαίνει τήξη - γνωστή ως σημείο τήξης - είναι χαρακτηριστικό αυτής της ουσίας. Απαιτεί ενέργεια για να λιώσει ένα στερεό σε υγρό. Κάθε καθαρή ουσία έχει ένα ορισμένο ποσό ενέργειας που χρειάζεται για να μετατραπεί από στερεό σε υγρό. Αυτή η ποσότητα ονομάζεται ενθαλπία σύντηξης (ή θερμότητα σύντηξης) της ουσίας, που αντιπροσωπεύεται ως ΔH_{fus}. Η μονάδα ΔH_{fus} είναι kilojoules ανά mole, επομένως πρέπει να γνωρίζουμε την ποσότητα του υλικού για να γνωρίζουμε πόση ενέργεια εμπλέκεται. Το ΔH_{fus} παρουσιάζεται πάντα ως θετικός αριθμός. Ωστόσο, μπορεί να χρησιμοποιηθεί τόσο για τη διαδικασία τήξης όσο και για τη στερεοποίηση, αρκεί να έχετε κατά νου ότι η τήξη είναι πάντα ενδόθερμη (άρα η ΔΗ θα είναι θετική), ενώ η στερεοποίηση είναι πάντα εξώθερμη (άρα η ΔΗ θα είναι αρνητική).

Ενθαλπίες σύντηξης για διάφορες ουσίες

Ουσία (σημείο τήξης)	ΔH_{fus} (kJ/mol)
Νερό (0°C)	6.01
Αλουμίνιο (660°C)	10.7
Βενζόλιο (5,5°C)	9,95
Αιθανόλη (−114,3°C)	5.02
Υδράργυρος (−38,8°C)	2.29

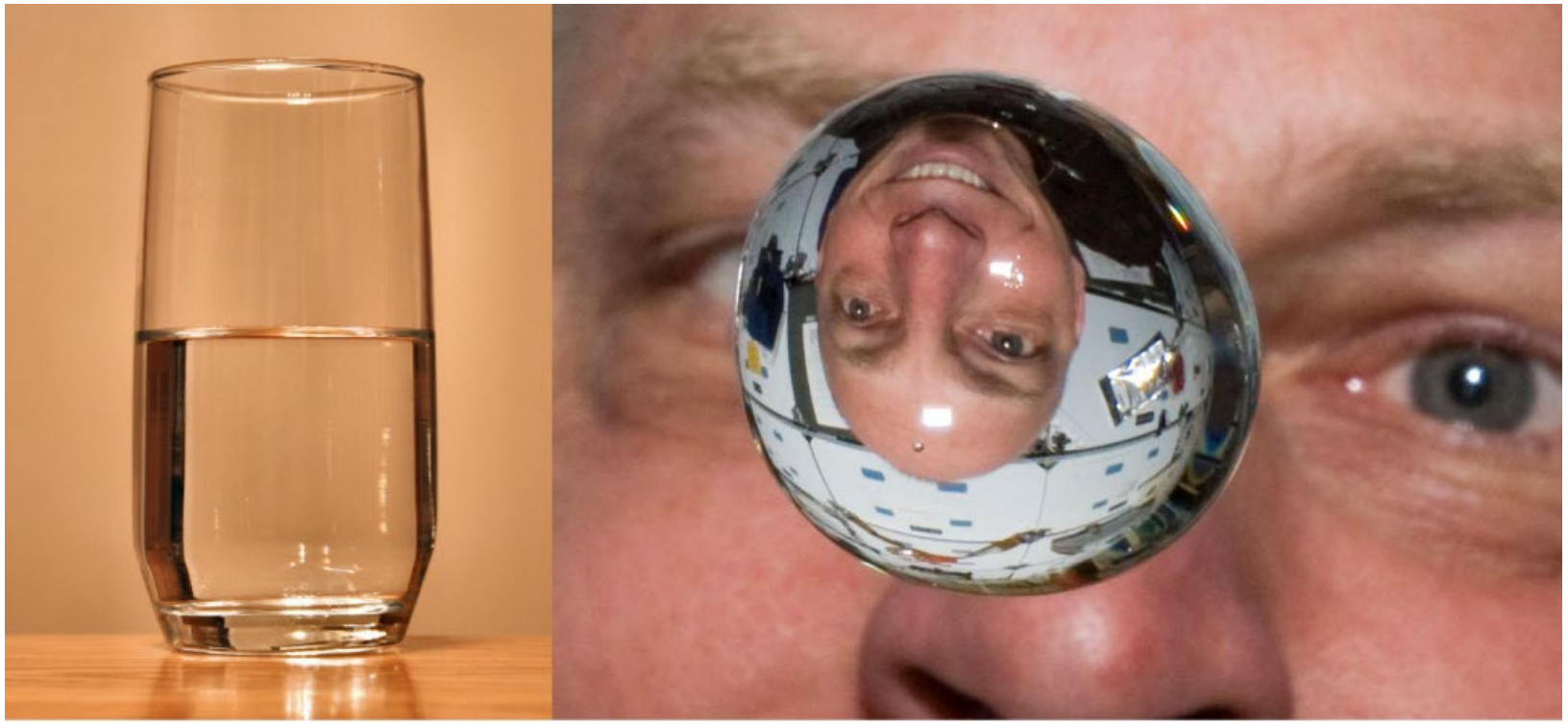
Co-funded by the Erasmus+ Programme of the European Union

Τι συμβαίνει όταν ένα στερεό γίνεται υγρό;

Κατά τη διάρκεια της τήξης, η ενέργεια πηγαίνει αποκλειστικά στην αλλαγή της φάσης μιας ουσίας. δεν μεταβάλλει τη θερμοκρασία μιας ουσίας. Ως εκ τούτου, η τήξη είναι μια ισόθερμη διαδικασία επειδή μια ουσία παραμένει στην ίδια θερμοκρασία. Μόνο όταν λιώσει όλη μια ουσία, οποιαδήποτε επιπλέον ενέργεια πηγαίνει στην αλλαγή της θερμοκρασίας της.

Σε ένα στερεό, μεμονωμένα σωματίδια είναι κολλημένα στη θέση τους επειδή οι διαμοριακές δυνάμεις δεν μπορούν να υπερνικηθούν από την ενέργεια των σωματιδίων. Όταν παρέχεται περισσότερη ενέργεια (π.χ. με την αύξηση της θερμοκρασίας), έρχεται ένα σημείο στο οποίο τα σωματίδια έχουν αρκετή ενέργεια για να κινηθούν αλλά όχι αρκετή ενέργεια για να διαχωριστούν. Αυτή είναι η υγρή φάση: τα σωματίδια εξακολουθούν να βρίσκονται σε επαφή αλλά μπορούν να κινούνται το ένα γύρω από το άλλο. Αυτό εξηγεί γιατί τα υγρά μπορούν να λάβουν το σχήμα των δοχείων τους: τα σωματίδια κινούνται γύρω και, υπό την επίδραση της βαρύτητας, γεμίζουν τον χαμηλότερο δυνατό όγκο (εκτός εάν το υγρό βρίσκεται σε περιβάλλον μηδενικής βαρύτητας .

«Υγρά και Βαρύτητα». (α) Ένα υγρό γεμίζει τον πυθμένα του δοχείου του καθώς έλκεται προς τα κάτω από τη βαρύτητα και τα σωματίδια γλιστρούν το ένα πάνω στο άλλο. (β) Ένα υγρό επιπλέει σε περιβάλλον μηδενικής βαρύτητας. Τα σωματίδια εξακολουθούν να γλιστρούν το ένα πάνω στο άλλο επειδή βρίσκονται στην υγρή φάση, αλλά τώρα δεν υπάρχει βαρύτητα για να τα τραβήξει προς τα κάτω.



(a) (b)

Αλλαγή φάσης υγρού αερίου

Η αλλαγή φάσης μεταξύ υγρού και αερίου έχει κάποιες ομοιότητες με την αλλαγή φάσης μεταξύ στερεού και υγρού. Σε μια ορισμένη θερμοκρασία, τα σωματίδια ενός υγρού έχουν αρκετή ενέργεια για να γίνουν αέριο. Η διαδικασία κατά την οποία ένα υγρό μετατρέπεται σε αέριο ονομάζεται βρασμός (ή εξάτμιση), ενώ η διαδικασία όταν ένα αέριο γίνεται υγρό ονομάζεται συμπύκνωση. Ωστόσο, σε αντίθεση με τη διαδικασία μετατροπής στερεού/υγρού, η διαδικασία μετατροπής υγρού/αερίου επηρεάζεται αισθητά από την περιβάλλουσα πίεση στο υγρό επειδή τα αέρια επηρεάζονται έντονα από την πίεση. Αυτό σημαίνει ότι η θερμοκρασία στην οποία ένα υγρό γίνεται αέριο, το σημείο βρασμού, μπορεί να αλλάξει ανάλογα με την πίεση του περιβάλλοντος. Ως εκ τούτου, ορίζουμε το κανονικό σημείο βρασμού ως τη θερμοκρασία στην οποία ένα υγρό μετατρέπεται σε αέριο όταν η περιβάλλουσα πίεση είναι ακριβώς 1 atm ή 760 torr. Εκτός εάν ορίζεται διαφορετικά, θεωρείται ότι ένα σημείο βρασμού είναι για 1 atm πίεσης.

Ενθαλπία εξάτμισης

Όπως η αλλαγή φάσης στερεάς/υγρής, η αλλαγή φάσης υγρού/αερίου περιλαμβάνει ενέργεια. Η ποσότητα ενέργειας που απαιτείται για τη μετατροπή ενός υγρού σε αέριο ονομάζεται ενθαλπία εξάτμισης (ή θερμότητα εξάτμισης), που αντιπροσωπεύεται ως ΔHvap . Η μονάδα για το ΔHvap είναι επίσης kilojoules ανά mole, επομένως πρέπει να γνωρίζουμε την ποσότητα του υλικού για να γνωρίζουμε πόση ενέργεια εμπλέκεται. Το ΔHvap επίσης καταγράφεται πάντα ως θετικός αριθμός. Μπορεί να χρησιμοποιηθεί τόσο για τη διαδικασία βρασμού όσο και για τη συμπύκνωση, αρκεί να έχετε κατά νου ότι ο βρασμός είναι πάντα ενδόθερμος (άρα το ΔΗ θα είναι θετικό), ενώ η συμπύκνωση είναι πάντα εξώθερμη (άρα το ΔΗ θα είναι αρνητικό).

Co-funded by the Erasmus+ Programme of the European Union

Ενθαλπίες εξάτμισης για διάφορες ουσίες

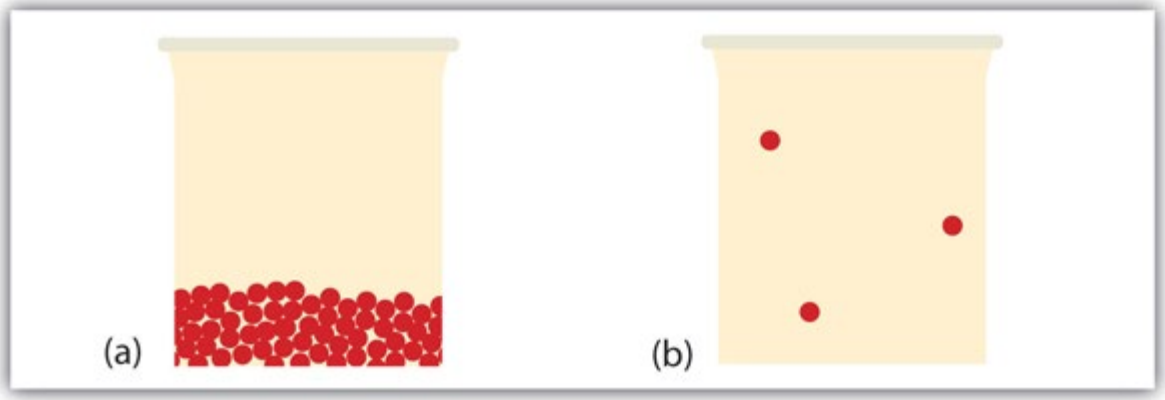
Ουσία (Κανονικό σημείο βρασμού)	ΔH_{vap} (kJ/mol)
Νερό (100°C)	40,68
Βρώμιο (59,5°C)	15.4
Βενζόλιο (80,1°C)	30.8
Αιθανόλη (78,3°C)	38.6
Υδράργυρος (357°C)	59,23

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Τι συμβαίνει όταν ένα υγρό γίνεται αέριο;

Όταν ένα υγρό γίνεται αέριο, τα σωματίδια διαχωρίζονται το ένα από το άλλο, με κάθε σωματίδιο να ακολουθεί τον δικό του δρόμο στο διάστημα. Έτσι τα αέρια τείνουν να γεμίζουν τα δοχεία τους. Πράγματι, στην αέρια φάση το μεγαλύτερο μέρος του όγκου είναι κενός χώρος. μόνο περίπου το ένα χιλιοστό του όγκου καταλαμβάνεται στην πραγματικότητα από την ύλη.



Co-funded by the Erasmus+ Programme of the European Union

Εξάχνωση

Κάτω από ορισμένες συνθήκες, η στερεά φάση μπορεί να μεταβεί απευθείας στην αέρια φάση χωρίς να περάσει από υγρή φάση και ένα αέριο μπορεί να γίνει απευθείας στερεό. Η αλλαγή από στερεό σε αέριο ονομάζεται εξάχνωση, ενώ η αντίστροφη διαδικασία ονομάζεται εναπόθεση. Η εξάχνωση είναι ισοθερμική, όπως και οι άλλες αλλαγές φάσης. Υπάρχει μια μετρήσιμη αλλαγή ενέργειας κατά την εξάχνωση. αυτή η ενεργειακή αλλαγή ονομάζεται ενθαλπία εξάχνωσης, που αναπαρίσταται ως ΔHsub . Η σχέση μεταξύ του ΔHsub και των άλλων αλλαγών ενθαλπίας είναι η εξής:

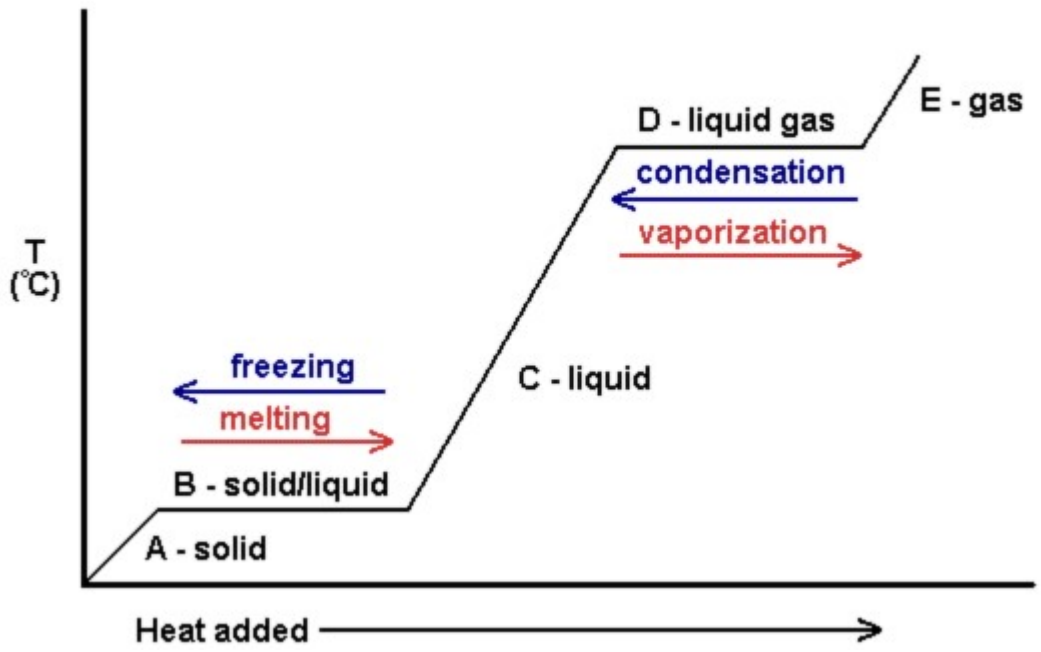
$$\Delta H_{\text{sub}} = \Delta H_{\text{fus}} + \Delta H_{\text{vap}}$$

Ως εκ τούτου, το ΔHsub δεν παρουσιάζεται πάντα σε πίνακα, επειδή μπορεί απλά να υπολογιστεί από τα ΔHfus και ΔHvap

Co-funded by the Erasmus+ Programme of the European Union

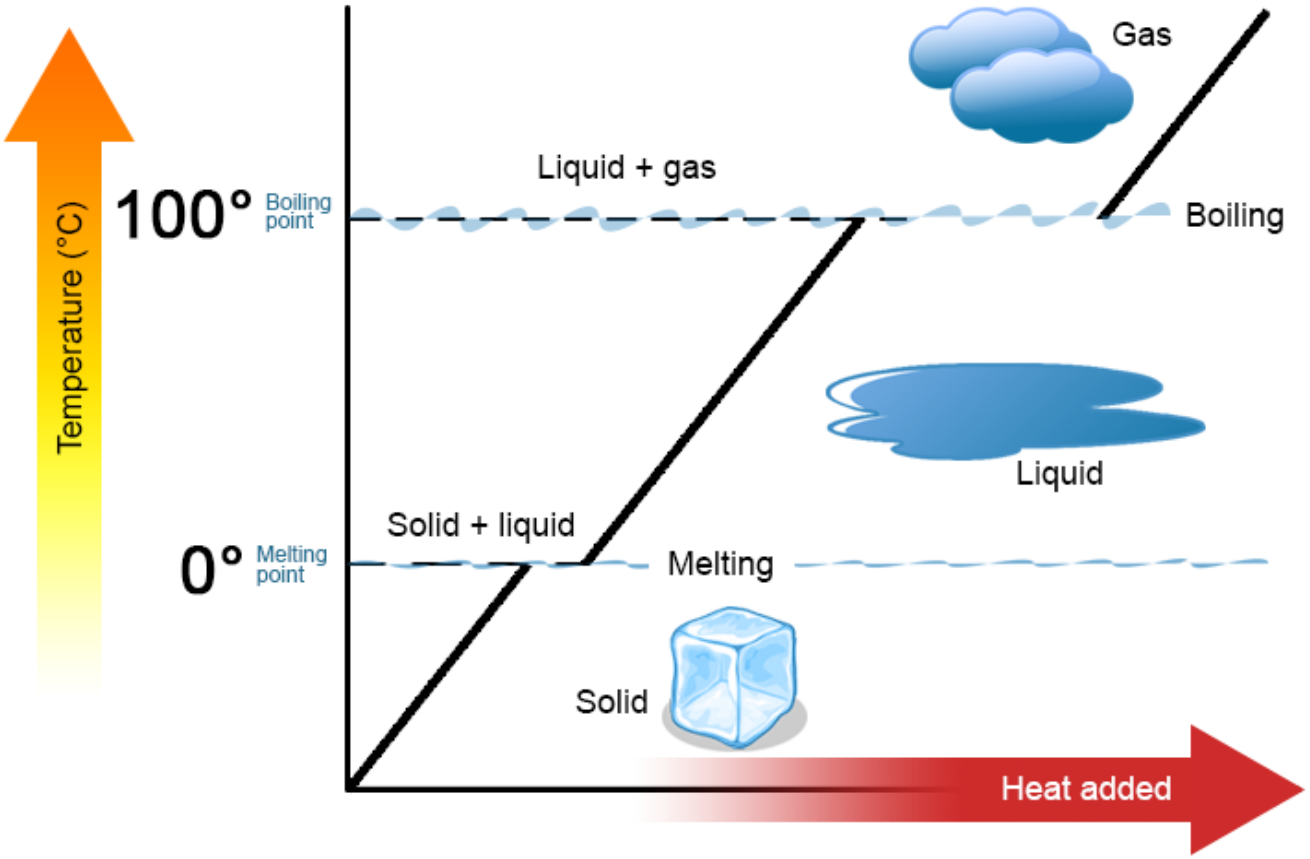
Καμπύλες θέρμανσης

Μια γραφική παράσταση της θερμοκρασίας σε σχέση με την ποσότητα θερμότητας που προστίθεται είναι γνωστή ως καμπύλη θέρμανσης. Αυτά χρησιμοποιούνται συνήθως για να δείξουν οπτικά τη σχέση μεταξύ των αλλαγών φάσης και της ενθαλπίας για μια δεδομένη ουσία.



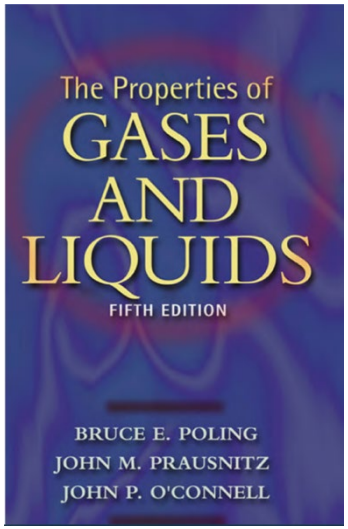
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Καμπύλη θέρμανσης για νερό. Καθώς η θερμότητα προστίθεται στο στερεό νερό

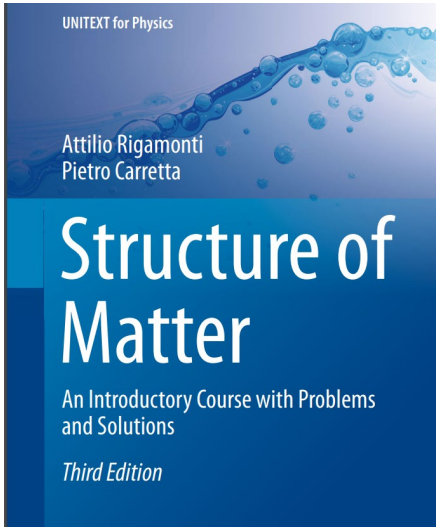


Ερωτήσεις

- Ποια είναι η διαφορά μεταξύ τήξης και στερεοποίησης;
- Ποια είναι η διαφορά μεταξύ βρασμού και συμπύκνωσης;
- Περιγράψτε τις μοριακές αλλαγές όταν ένα στερεό γίνεται υγρό.
- Περιγράψτε τις μοριακές αλλαγές όταν ένα υγρό γίνεται αέριο.



BruceE.Poling , JohnM.Prausnitz , JohnP.O'Connell (2001) Οι ιδιότητες των αερίων και των υγρών
McGraw-Hill



Rigamonti , A., & Carretta , P. (2015). Δομή της Ύλης. UNITEXT for Physics, Springer International Publishing Switzerland

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 2. Βασικές αρχές της μοριακής βιολογίας

Μάθημα 1. Χημικές βάσεις του κυττάρου

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Υδατάνθρακες
- Λιπίδια
- Νουκλεϊκά οξέα
- Πρωτεΐνες

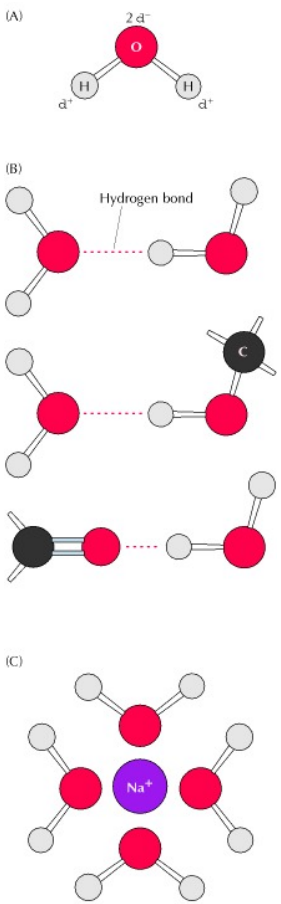
Εισαγωγή

Τα κύτταρα αποτελούνται από νερό, ανόργανα ιόντα και (οργανικά) μόρια που περιέχουν άνθρακα. Το νερό είναι το πιο άφθονο μόριο στα κύτταρα, αντιπροσωπεύοντας το 70% ή περισσότερο της συνολικής κυτταρικής μάζας. Κατά συνέπεια, οι αλληλεπιδράσεις μεταξύ του νερού και των άλλων συστατικών των κυττάρων είναι κεντρικής σημασίας στη βιολογική χημεία. Η κρίσιμη ιδιότητα του νερού από αυτή την άποψη είναι ότι είναι ένα πολικό μόριο, στο οποίο τα άτομα υδρογόνου έχουν ένα ελαφρώς θετικό φορτίο και το οξυγόνο έχει ένα ελαφρώς αρνητικό φορτίο

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

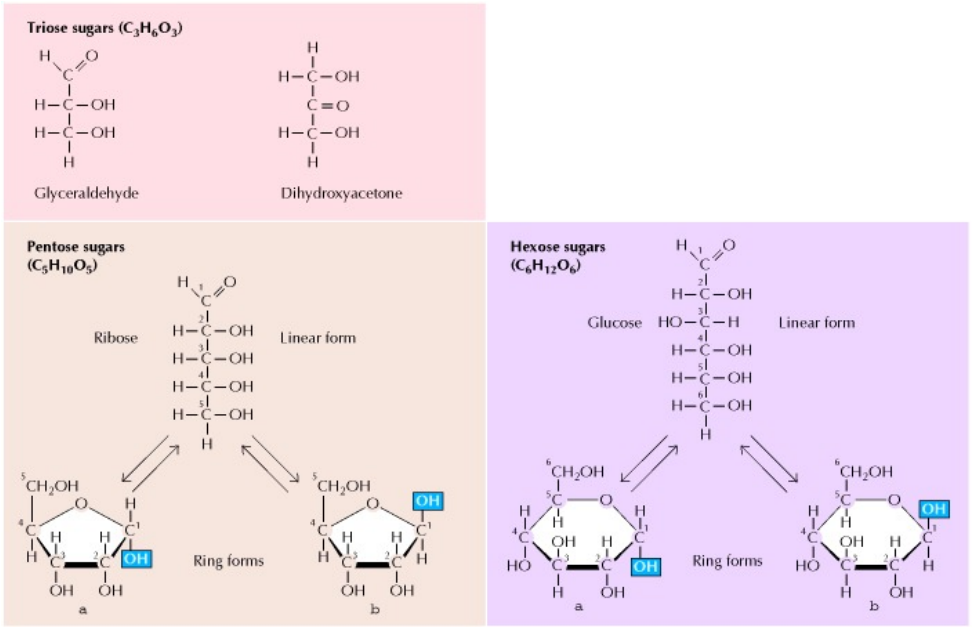
Χαρακτηριστικά του νερού



(A) Το νερό είναι ένα πολικό μόριο, με ένα ελαφρύ αρνητικό φορτίο (δ-) στο άτομο οξυγόνου και ένα ελαφρύ θετικό φορτίο (δ+) στα άτομα του υδρογόνου. Λόγω αυτής της πολικότητας, τα μόρια του νερού μπορούν να σχηματίσουν δεσμούς υδρογόνου (διακεκομμένες γραμμές) είτε μεταξύ τους είτε με άλλα πολικά μόρια (B), εκτός από την αλληλεπίδραση με φορτισμένα ιόντα (C).

Τα ανόργανα ιόντα του κυττάρου, συμπεριλαμβανομένου του νατρίου (Na+), του καλίου (K+), του μαγνησίου (Mg2+), του ασβεστίου (Ca2+), του φωσφορικού (HPO42-), του χλωρίου (Cl-) και του διττανθρακικού (HCO3-), αποτελούν το 1% ή λιγότερο από την κυτταρική μάζα. Αυτά τα ιόντα εμπλέκονται σε μια σειρά από πτυχές του κυτταρικού μεταβολισμού και επομένως παίζουν κρίσιμους ρόλους στη λειτουργία των κυττάρων.

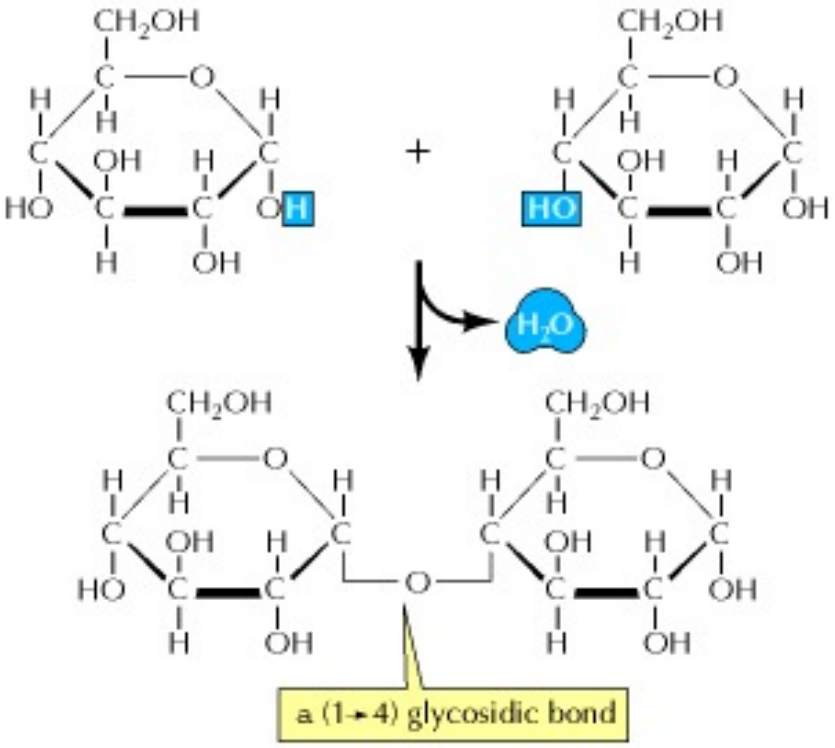
Υδατάνθρακες



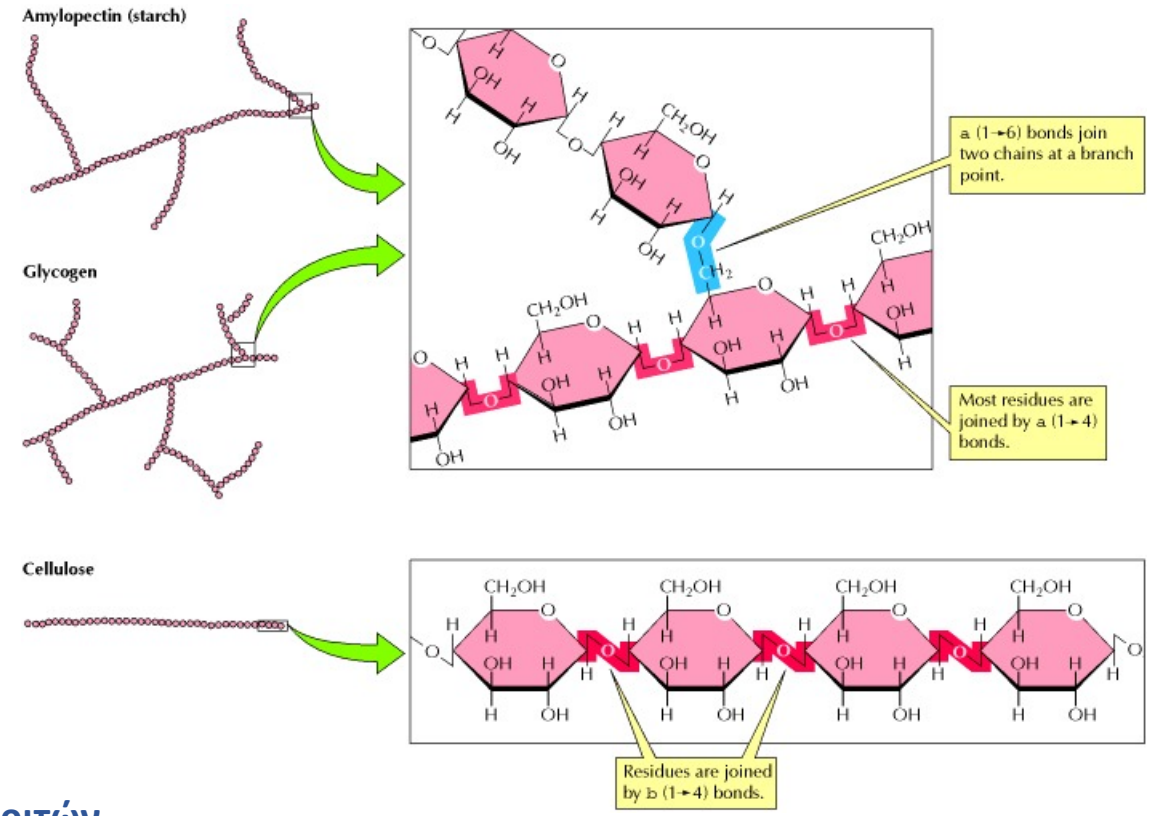
Δομή απλών σακχάρων

Απεικονίζονται αντιπροσωπευτικά σάκχαρα που περιέχουν τρεις, πέντε και έξι άνθρακες (σάκχαρα τριόζης, πεντόζης και εξόζης, αντίστοιχα). Τα σάκχαρα με πέντε ή περισσότερους άνθρακες μπορούν να κυκλοποιηθούν για να σχηματίσουν δακτυλίους, οι οποίοι υπάρχουν σε δύο εναλλακτικές μορφές (α και β) ανάλογα με τη διαμόρφωση του άνθρακα 1.

Co-funded by the Erasmus+ Programme of the European Union



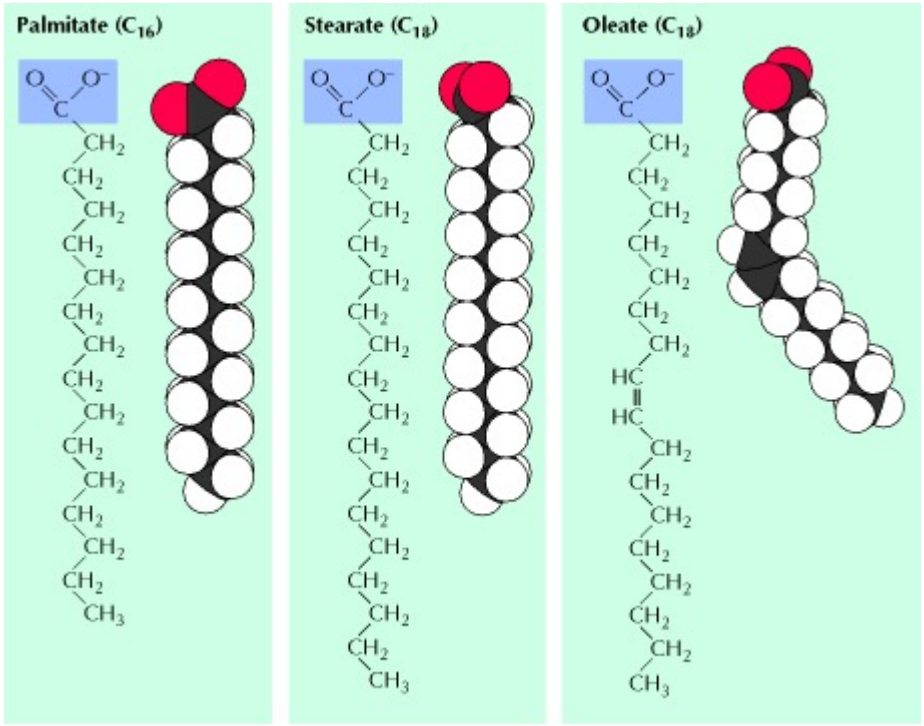
Σχηματισμός γλυκοσιδικού δεσμού
Δύο απλά σάκχαρα ενώνονται με μια αντίδραση αφυδάτωσης (αντίδραση κατά την οποία αφαιρείται το νερό). Στο παράδειγμα που φαίνεται, δύο μόρια γλυκόζης στη διάταξη α ενώνονται με έναν δεσμό μεταξύ των ανθράκων 1 και 4, ο οποίος επομένως ονομάζεται α (1→4) γλυκοσιδικός δεσμός.



Δομή πολυσακχαριτών

Οι πολυσακχαρίτες είναι μακρομόρια που αποτελούνται από εκατοντάδες ή χιλιάδες απλά σάκχαρα. Το γλυκογόνο, το άμυλο και η κυτταρίνη αποτελούνται εξ ολοκλήρου από υπολείμματα γλυκόζης, τα οποία συνδέονται με α (1→4) γλυκοσιδικούς δεσμούς στο γλυκογόνο και το άμυλο αλλά με β (1→4) δεσμούς στην κυτταρίνη. Το γλυκογόνο και μια μορφή αμύλου (αμυλοπηκτίνη) περιέχουν επίσης περιστασιακούς δεσμούς α (1→6), οι οποίοι χρησιμεύουν ως σημεία διακλάδωσης ενώνοντας δύο ξεχωριστές α (1→4) αλυσίδες.

Λιπίδια

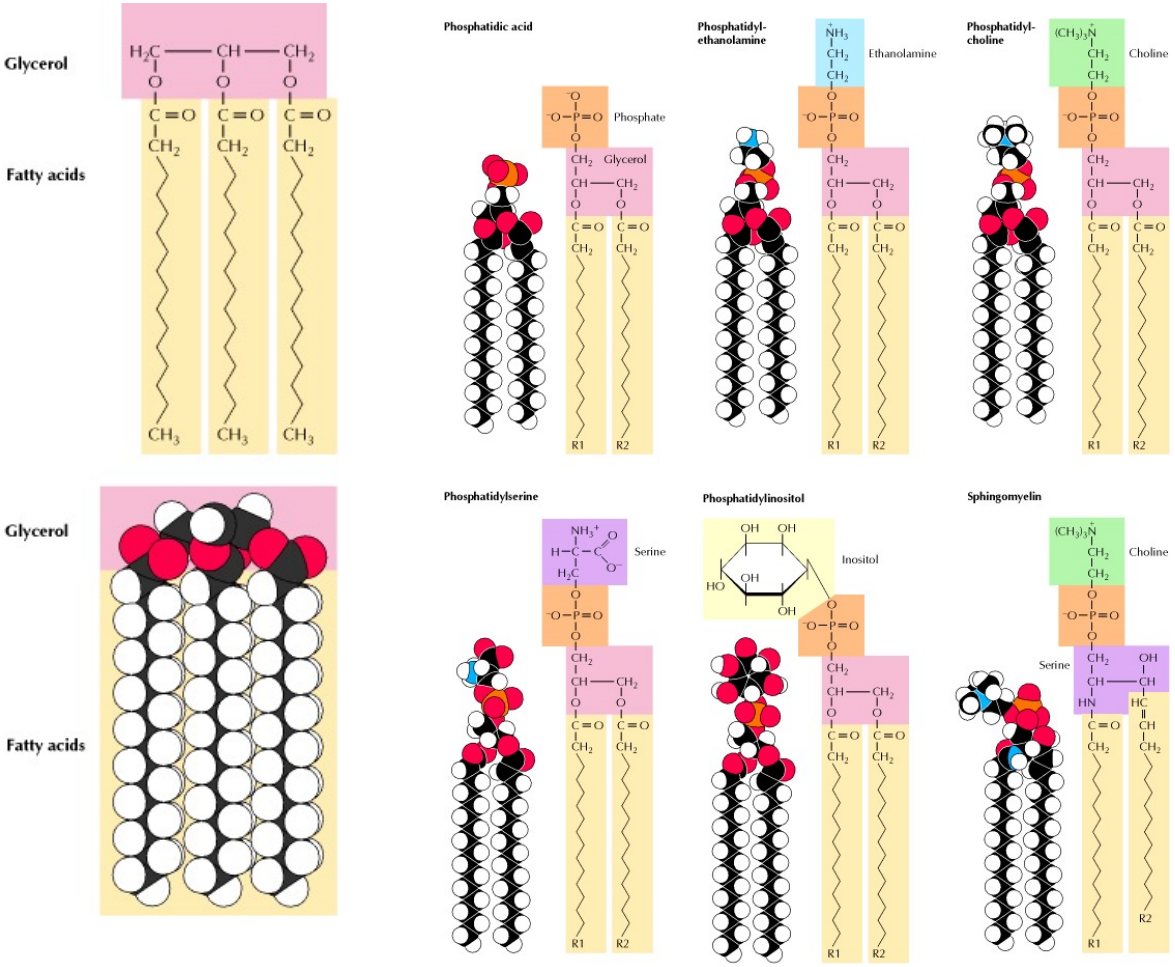


Δομή λιπαρών οξέων

Τα λιπαρά οξέα αποτελούνται από μακριές αλυσίδες υδρογονάνθρακα που καταλήγουν σε μια καρβοξυλική ομάδα (COO-). Το παλμιτικό και το στεατικό είναι κορεσμένα λιπαρά οξέα που αποτελούνται από 16 και 18 άνθρακες, αντίστοιχα. Το ελαϊκό είναι ένα ακόρεστο λιπαρό οξύ 18 ανθράκων που περιέχει διπλό δεσμό μεταξύ των ανθράκων 9 και 10. Σημειώστε ότι ο διπλός δεσμός εισάγει μια στροφή στην αλυσίδα υδρογονάνθρακα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

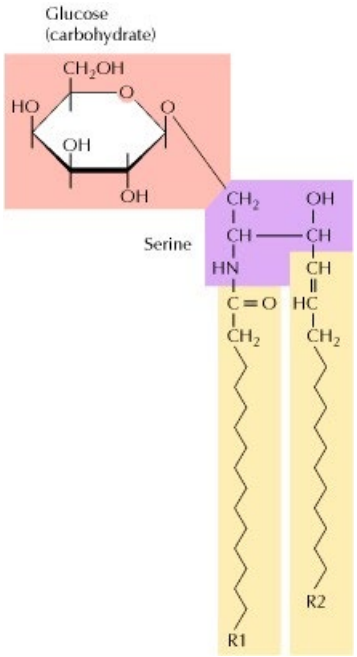


Δομή των φωσφολιπιδίων

Τα φωσφολιπίδια της γλυκερίνης περιέχουν δύο λιπαρά οξέα ενωμένα με τη γλυκερίνη. Τα λιπαρά οξέα μπορεί να είναι διαφορετικά μεταξύ τους και ονομάζονται R1 και R2. Ο τρίτος άνθρακας της γλυκερίνης ενώνεται με μια φωσφορική ομάδα (σχηματίζοντας φωσφατιδικό οξύ), το οποίο με τη σειρά του συχνά ενώνεται με ένα άλλο μικρό πολικό μόριο (σχηματίζοντας φωσφατιδυλαιθανολαμίνη, φωσφατιδυλοχολίνη, φωσφατιδυλοσερίνη ή φωσφατιδυλινοσιτόλη). Στη σφιγγομυελίνη, δύο αλυσίδες υδρογονάνθρακα συνδέονται με μια πολική ομάδα κεφαλής που σχηματίζεται από σερίνη αντί για γλυκερίνη.

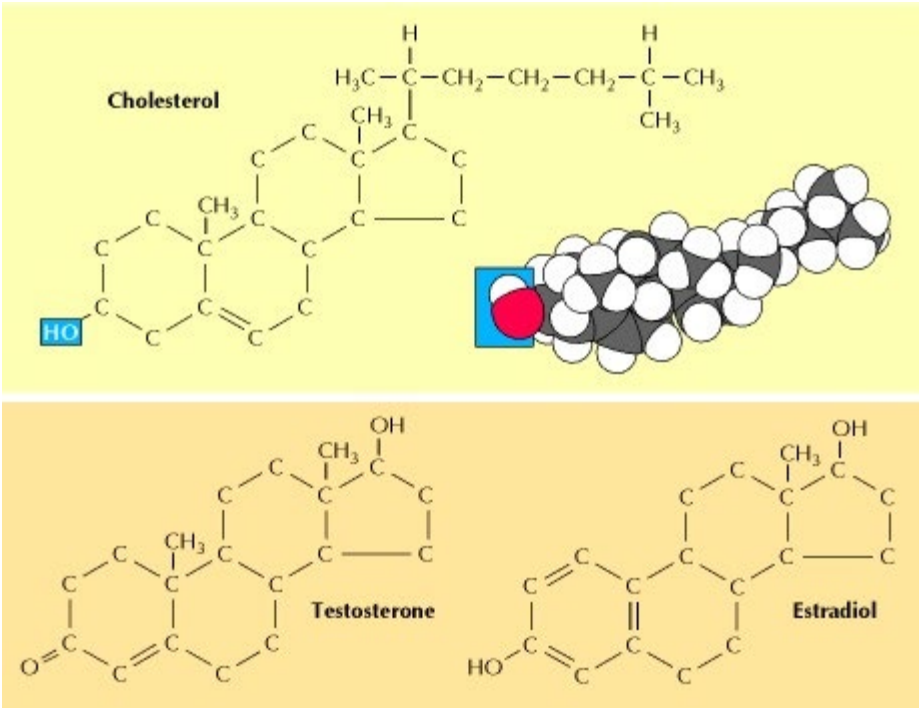
Δομή των τριακυλογλυκερολών

Οι τριακυλογλυκερόλες (λίπη) περιέχουν τρία λιπαρά οξέα ενωμένα με τη γλυκερόλη. Σε αυτό το παράδειγμα, και τα τρία λιπαρά οξέα είναι παλμιτικά, αλλά οι τριακυλογλυκερόλες συχνά περιέχουν ένα μείγμα διαφορετικών λιπαρών οξέων.



Δομή γλυκολιπιδίων

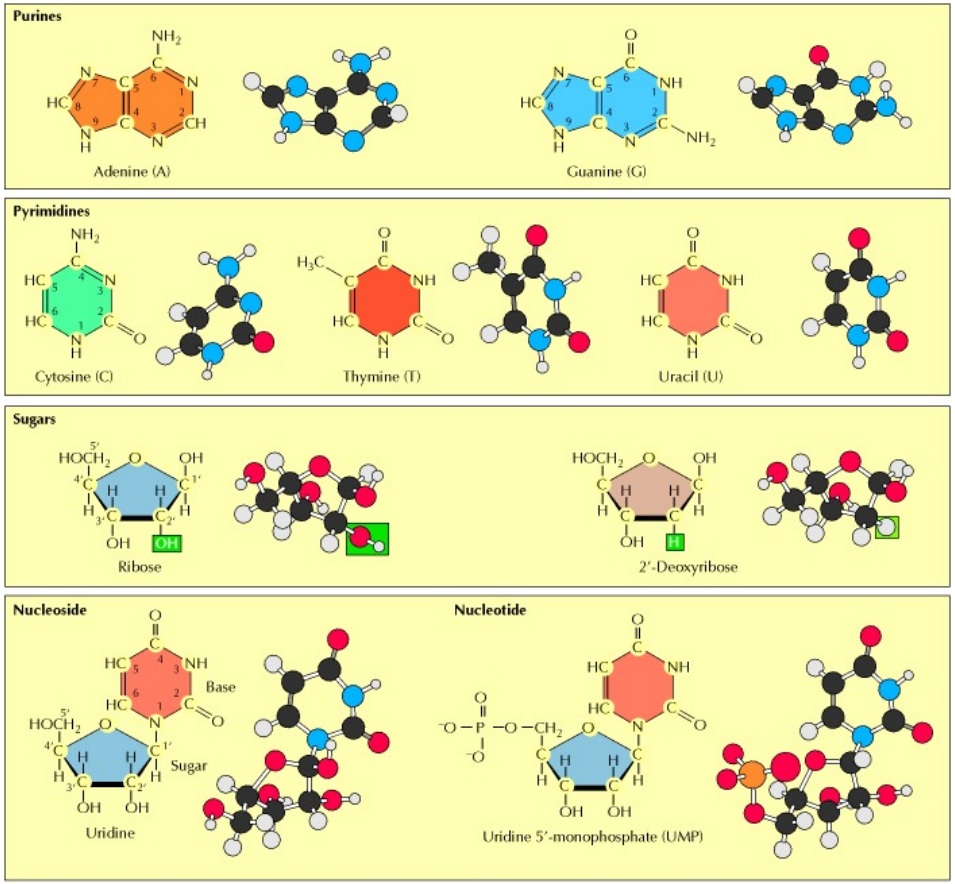
Δύο αλυσίδες υδρογονάνθρακα ενώνονται σε μια ομάδα πολικής κεφαλής που σχηματίζεται από σερίνη και περιέχει υδατάνθρακες (π.χ. γλυκόζη).



Χοληστερόλη και στεροειδείς ορμόνες

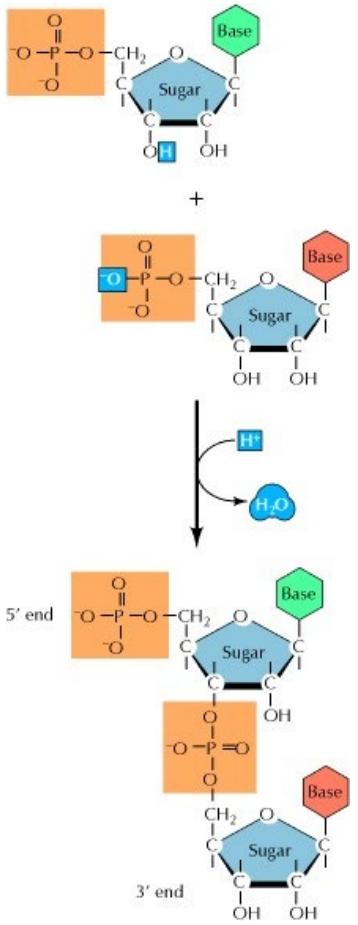
Η χοληστερόλη, ένα σημαντικό συστατικό των κυτταρικών μεμβρανών, είναι ένα αμφιπαθές μόριο λόγω της πολικής υδροξυλικής της ομάδας. Η χοληστερόλη είναι επίσης πρόδρομος των στεροειδών ορμονών, όπως η τεστοστερόνη και η οιστραδιόλη (μια μορφή οιστρογόνου). Τα άτομα υδρογόνου που συνδέονται με τους άνθρακες του δακτυλίου δεν φαίνονται σε αυτό το σχήμα.

Νουκλεϊκά οξέα



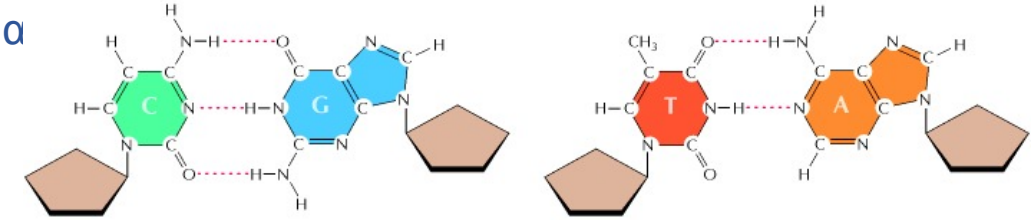
Συστατικά νουκλεϊκών οξέων
Τα νουκλεϊκά οξέα περιέχουν βάσεις πουρίνης και πυριμιδίνης που συνδέονται με φωσφορυλιωμένα σάκχαρα. Μια βάση νουκλεϊκού οξέος που συνδέεται μόνο με ένα σάκχαρο είναι ένας νουκλεοζίτης. Τα νουκλεοτίδια περιέχουν επιπλέον μία ή περισσότερες φωσφορικές ομάδες.

Co-funded by the Erasmus+ Programme of the European Union



Πολυμερισμός νουκλεοτιδίων

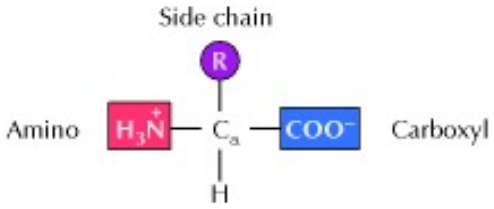
Ένας φωσφοδιεστερικός δεσμός σχηματίζεται μεταξύ της 3' υδροξυλομάδας ενός νουκλεοτιδίου και της 5' φωσφορικής ομάδας ενός άλλου. Μια πολυνουκλεοτιδική αλυσίδα έχει μια αίσθηση κατεύθυνσης, με το ένα άκρο να καταλήγει σε μια φωσφορική ομάδα 5' (το 5' άκρο) και το άλλο σε μια ομάδα



Συμπληρωματικό ζευγάρωμα μεταξύ βάσεων νουκλεϊκού οξέος

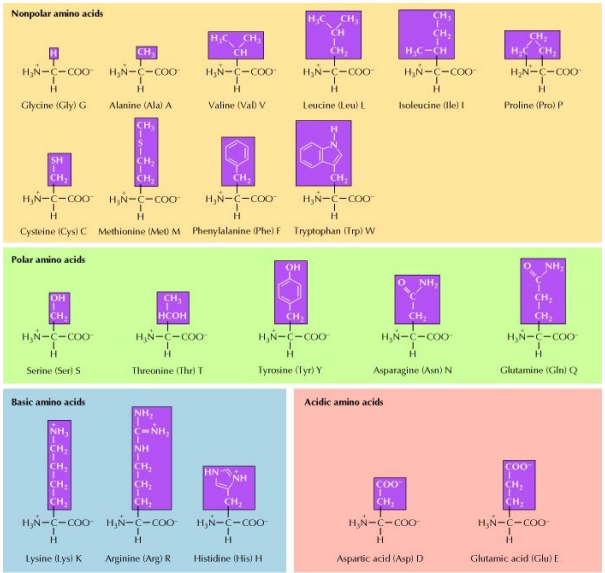
Ο σχηματισμός δεσμών υδρογόνου μεταξύ βάσεων σε αντίθετους κλώνους DNA οδηγεί στο ειδικό ζευγάρωμα της γουανίνης (G) με την κυτοσίνη (C) και της αδενίνης (A) με τη θυμίνη (T).

Πρωτεΐνες



Δομή αμινοξέων

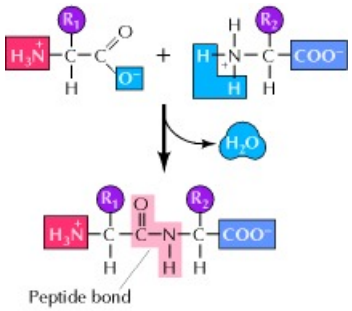
Κάθε αμινοξύ αποτελείται από ένα κεντρικό άτομο άνθρακα (ο άνθρακας α) συνδεδεμένο με ένα άτομο υδρογόνου, μια ομάδα καρβοξυλίου, μια αμινομάδα και μια συγκεκριμένη πλευρική αλυσίδα (που ονομάζεται R). Σε φυσιολογικό pH, τόσο η καρβοξυλική όσο και η αμινομάδα ιονίζονται, όπως φαίνεται.



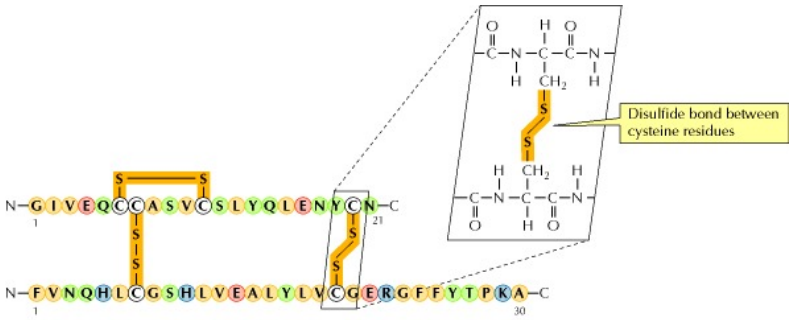
Τα αμινοξέα

Υποδεικνύονται οι συντομογραφίες τριών γραμμάτων και ενός γράμματος για κάθε αμινοξύ. Τα αμινοξέα ομαδοποιούνται σε τέσσερις κατηγορίες ανάλογα με τις ιδιότητες των πλευρικών τους αλυσίδων: μη πολικές, πολικές, βασικές και όξινες.

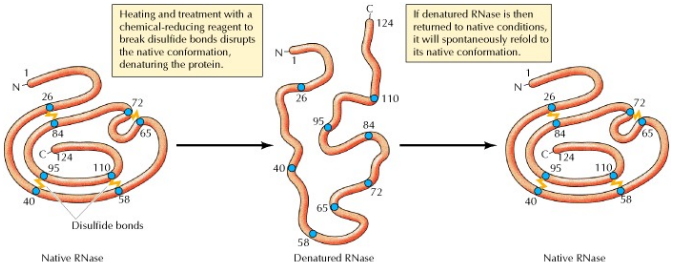
Co-funded by the Erasmus+ Programme of the European Union



Σχηματισμός πεπτιδικού δεσμού
Η καρβοξυλική ομάδα ενός αμινοξέος συνδέεται με την αμινομάδα ενός δευτερολέπτου.

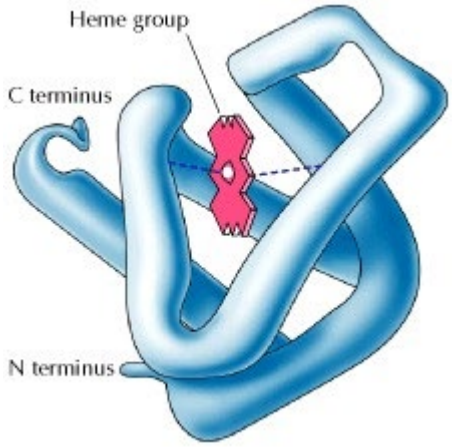


Αλληλουχία αμινοξέων ινσουλίνης
Η ινσουλίνη αποτελείται από δύο πολυπεπτιδικές αλυσίδες, η μία από 21 και η άλλη από 30 αμινοξέα (που υποδεικνύονται εδώ από τους κωδικούς τους με ένα γράμμα). Οι πλευρικές αλυσίδες τριών ζευγών υπολειμμάτων κυστεΐνης ενώνονται με δισουλφιδικούς δεσμούς, δύο από τους οποίους συνδέουν τις πολυπεπτιδικές αλυσίδες.



Πρωτεϊνική μετουσίωση και αναδίπλωση

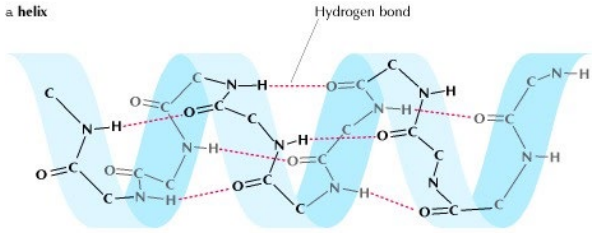
Η ριβονουκλεάση (RNase) είναι μια πρωτεΐνη 124 αμινοξέων (που υποδεικνύονται με αριθμούς). Η πρωτεΐνη κανονικά διπλώνεται στη φυσική της διαμόρφωση, η οποία περιέχει τέσσερις δισουλφιδικούς δεσμούς (που υποδεικνύονται ως ζευγαρωμένοι κύκλοι που αντιπροσωπεύουν τα υπολείμματα κυστεΐνης).

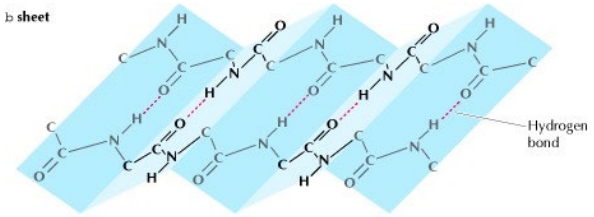


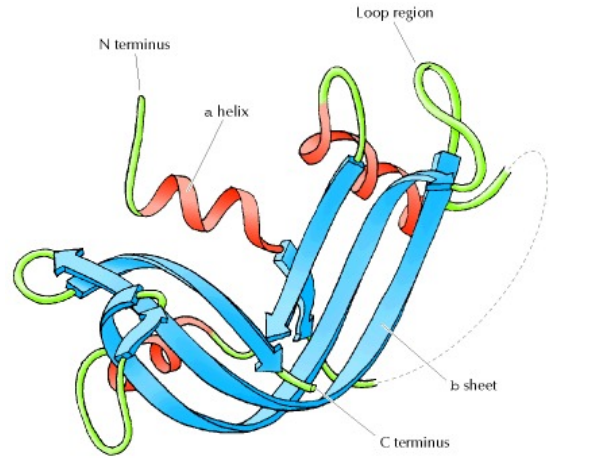
Τρισδιάστατη δομή της μυοσφαιρίνης

Η μυοσφαιρίνη είναι μια πρωτεΐνη 153 αμινοξέων που εμπλέκεται στη μεταφορά οξυγόνου. Η πολυπεπτιδική αλυσίδα διπλώνεται γύρω από μια ομάδα αίμης που χρησιμεύει ως η θέση δέσμευσης οξυγόνου.

Co-funded by the Erasmus+ Programme of the European Union







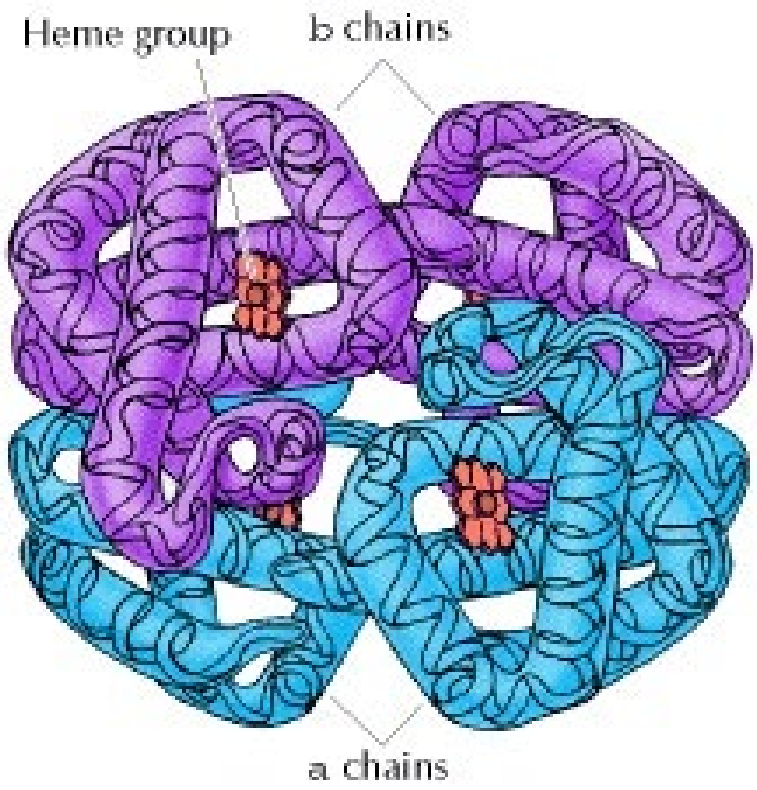
Δευτερεύουσα δομή πρωτεϊνών

Οι πιο συνηθισμένοι τύποι δευτερογενούς δομής είναι η έλικα α και το φύλλο β. Σε μια έλικα α, σχηματίζονται δεσμοί υδρογόνου μεταξύ ομάδων CO και NH πεπτιδικών δεσμών που διαχωρίζονται από τέσσερα υπολείμματα αμινοξέων. Σε ένα φύλλο β, δεσμοί υδρογόνου συνδέουν δύο μέρη μιας πολυπεπτιδικής αλυσίδας που βρίσκονται δίπλα-δίπλα. Οι πλευρικές αλυσίδες αμινοξέων δεν φαίνονται.

Τριτοταγής δομή ριβονουκλεάσης

Περιοχές δευτερευουσών δομών α-έλικας και β-φύλλου, που συνδέονται με περιοχές βρόχου, διπλώνονται στη φυσική διαμόρφωση της πρωτεΐνης. Σε αυτή τη σχηματική αναπαράσταση της πολυπεπτιδικής αλυσίδας ως μοντέλο κορδέλας, οι α έλικες αντιπροσωπεύονται ως σπείρες και τα φύλλα β ως φαρδιά βέλη.

Co-funded by the Erasmus+ Programme of the European Union



Τεταρτοταγής δομή της αιμοσφαιρίνης
Η αιμοσφαιρίνη αποτελείται από τέσσερις πολυπεπτιδικές αλυσίδες, καθεμία από τις οποίες συνδέεται με μια ομάδα αίμης . Οι δύο α αλυσίδες και οι δύο β αλυσίδες είναι πανομοιότυπες.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Ποια είναι τα βασικά μόρια της ζωής;
- Ποια είναι η δομή τους;
- Ποιες είναι οι λειτουργίες τους;

- Cooper GM. The Cell: A Molecular Approach. 2η έκδοση. Σάντερλαντ (MA): Sinauer Associates; 2000. The Molecular Composition of Cells. Διαθέσιμο από: https://www.ncbi.nlm.nih.gov/books/NBK9879/

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 2. Βασικές αρχές της Μοριακής Βιολογίας

Μάθημα 2. Το κύτταρο. Βασικοί τύποι κυττάρων. Κατασκευή κυττάρων. Οργανίδια ευκαρυωτικών κυττάρων. Κυτοσκελετός - συστατικά και δομικές λειτουργίες

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Θεωρία του κυττάρου
- Δομές προκαρυωτικών και ευκαρυωτικών κυττάρων
- Κυτοπλασματική μεμβράνη
- Οργανίδια

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Τα κύτταρα είναι οι βασικές μονάδες όλης της ζωής. Αποτελούν κάθε όργανο κάθε ζώου, φυτού, μύκητα και βακτηρίων. Τα κύτταρα σε ένα σώμα είναι σαν τα δομικά στοιχεία ενός σπιτιού. Έχουν επίσης μια συγκεκριμένη βασική δομή. Τα κύτταρα συνήθως αποτελούνται από:

Την κυτταρική μεμβράνη - αυτή είναι μια λιπιδική διπλοστιβάδα που σηματοδοτεί τα όρια του κυττάρου. Μέσα σε αυτή, μπορούμε να βρούμε τα άλλα δύο βασικά συστατικά του κυττάρου: το DNA και το κυτταρόπλασμα. Όλα τα κύτταρα έχουν κυτταρική ή πλασματική μεμβράνη.

DNA - το DNA περιέχει τις οδηγίες ώστε το κύτταρο να μπορεί να λειτουργήσει. Το γενετικό υλικό μπορεί να προστατεύεται εντός του πυρήνα (ευκαρυωτικά κύτταρα) ή να επιπλέει στο κυτταρόπλασμα (προκαρυωτικά κύτταρα). Τα περισσότερα κύτταρα έχουν DNA, αλλά τα ερυθρά αιμοσφαίρια, για παράδειγμα, δεν έχουν.

Κυτόπλασμα - το κυτταρόπλασμα είναι η παχύρρευστη ουσία εντός της μεμβράνης στην οποία επιπλέουν τα άλλα συστατικά ενός κυττάρου (το DNA/πυρήνας και άλλα οργανίδια).

Θεωρία κυττάρου

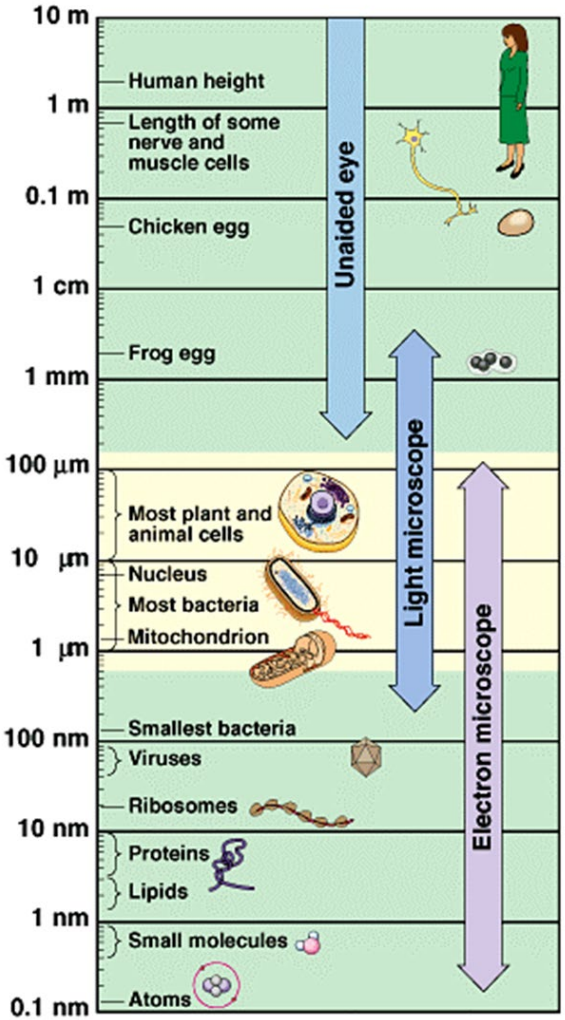
Προτάθηκε από τους Matthais Schleiden και Theodor Schwann το 1839:

-Όλα τα έμβια όντα αποτελούνται από κύτταρα.

-Τα κύτταρα είναι η μικρότερη μονάδα εργασίας όλων των ζωντανών όντων.

-Όλα τα κύτταρα προέρχονται από προϋπάρχοντα κύτταρα μέσω της διαίρεσης των κυττάρων.

Co-funded by the Erasmus+ Programme of the European Union



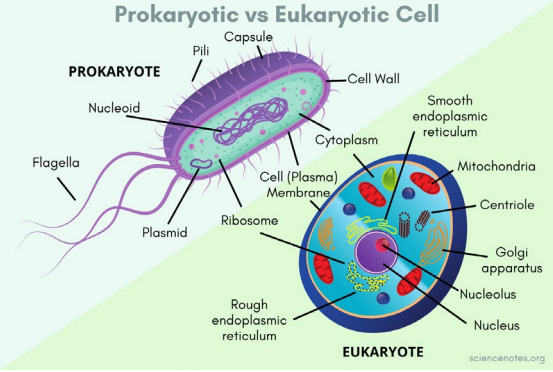
Πώς μελετάμε τα κύτταρα;
-μικροσκόπιο
-κλασματοποίηση κυττάρων

Τα περισσότερα κύτταρα έχουν διάμετρο μεταξύ 1-100 µm, η οποία μπορεί να απεικονιστεί με μικροσκόπιο φωτός.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

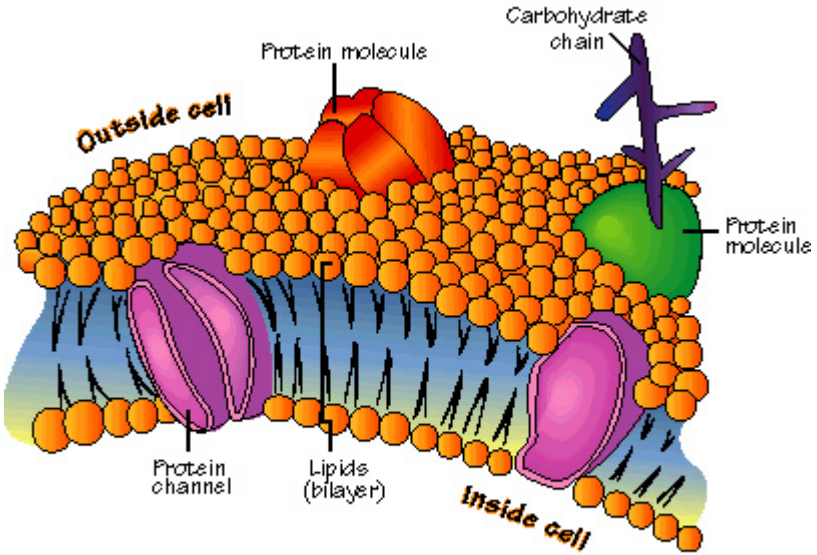
Δομές προκαρυωτικών και ευκαρυωτικών κυττάρων



Ο ορισμός του προκαρυώτου μεταφράζεται χονδρικά ως «χωρίς πυρήνα». Ως εκ τούτου, οι προκαρυώτες δεν έχουν ποτέ πυρήνα. Οι προκαρυώτες είναι συνήθως μονοκύτταροι, πράγμα που σημαίνει ότι τα βακτήρια, για παράδειγμα, αποτελούνται μόνο από ένα μόνο κύτταρο. Υπάρχουν, ωστόσο, εξαιρέσεις σε αυτόν τον κανόνα όπου ο οργανισμός είναι μονοκύτταρος αλλά έχει πυρήνα, επομένως είναι ευκαρυώτης. Η μαγιά είναι ένα παράδειγμα.

Από την άλλη πλευρά, το ευκαρυωτικό μεταφράζεται σε «αληθινός πυρήνας». Αυτό σημαίνει ότι όλοι οι ευκαρυώτες έχουν έναν πυρήνα. Εκτός από τη μαγιά, οι ευκαρυώτες είναι πολυκύτταροι καθώς μπορούν να αποτελούνται από εκατομμύρια κύτταρα. Οι άνθρωποι, για παράδειγμα, είναι ευκαρυώτες, όπως και τα φυτά και τα ζώα. Όσον αφορά τη δομή των κυττάρων, οι ευκαρυώτες και οι προκαρυώτες μοιράζονται ορισμένα χαρακτηριστικά αλλά είναι διαφορετικά σε άλλα. Ο παρακάτω πίνακας δείχνει τις ομοιότητες και τις διαφορές, ενώ μας δίνει επίσης μια γενική επισκόπηση των δομών των κυττάρων που θα συζητήσουμε σε αυτό το άρθρο.

Κυτοπλασματική μεμβράνη



Ονομάζεται επίσης πλασματική μεμβράνη ή κυτταρική μεμβράνη. Η πλασματική μεμβράνη είναι μια ημιπερατή μεμβράνη που χωρίζει το εσωτερικό ενός κυττάρου από το εξωτερικό.

Στα ευκαρυωτικά κύτταρα, η πλασματική μεμβράνη αποτελείται από πρωτεΐνες, υδατάνθρακες και δύο στοιβάδες φωσφολιπιδίων (δηλαδή λιπίδιο με φωσφορική ομάδα). Αυτά τα φωσφολιπίδια διατάσσονται ως εξής:

Οι πολικές, υδρόφιλες (υδατοφιλικές) κεφαλές βλέπουν προς το εξωτερικό και το εσωτερικό του κυττάρου. Αυτές οι κεφαλές αλληλεπιδρούν με το υδατικό περιβάλλον έξω και μέσα σε ένα κύτταρο.

Οι μη πολικές, υδρόφοβες (υδατοαπωθητικές) ουρές τοποθετούνται μεταξύ των κεφαλών και προστατεύονται από το υδατικό περιβάλλον.

Οι επιστήμονες Singer και Nicolson περιέγραψαν τη δομή της διπλοστιβάδας των φωσφολιπιδίων ως το «Ρευστό Μωσαϊκό Μοντέλο». Ο λόγος είναι ότι η διπλή στιβάδα μοιάζει με μωσαϊκό και έχει ημι-ρευστή φύση που επιτρέπει την πλευρική κίνηση των πρωτεϊνών εντός της διπλής στιβάδας.

Λειτουργίες

- Η πλασματική μεμβράνη είναι επιλεκτικά διαπερατή, δηλαδή επιτρέπει τη διέλευση μόνο επιλεγμένων ουσιών.
- Προστατεύει τα κύτταρα από σοκ και τραυματισμούς.
- Η ρευστή φύση της μεμβράνης επιτρέπει την αλληλεπίδραση μορίων εντός της μεμβράνης. Είναι επίσης σημαντικό για την έκκριση, την ανάπτυξη των κυττάρων και τη διαίρεση κ.λπ.
- Επιτρέπει τη μεταφορά μορίων μέσω της μεμβράνης. Αυτή η μεταφορά μπορεί να είναι δύο τύπων:
- Ενεργή μεταφορά – Αυτή η μεταφορά πραγματοποιείται αντίθετα από την βαθμίδα συγκέντρωσης και επομένως απαιτεί ενέργεια. Χρειάζεται επίσης πρωτεΐνες-φορείς και είναι μια εξαιρετικά επιλεκτική διαδικασία.
- Παθητική μεταφορά - Αυτή η μεταφορά συμβαίνει κατά την βαθμίδα συγκέντρωσης και επομένως, δεν απαιτεί ενέργεια. Έτσι, δεν χρειάζεται πρωτεΐνες-φορείς και δεν είναι επιλεκτική.

Co-funded by the Erasmus+ Programme of the European Union

Λειτουργίες

- Η πλασματική μεμβράνη είναι επιλεκτικά διαπερατή, δηλαδή επιτρέπει τη διέλευση μόνο επιλεγμένων ουσιών.
- Προστατεύει τα κύτταρα από σοκ και τραυματισμούς.
- Η ρευστή φύση της μεμβράνης επιτρέπει την αλληλεπίδραση μορίων εντός της μεμβράνης. Είναι επίσης σημαντικό για την έκκριση, την ανάπτυξη των κυττάρων και τη διαίρεση κ.λπ.
- Επιτρέπει τη μεταφορά μορίων μέσω της μεμβράνης. Αυτή η μεταφορά μπορεί να είναι δύο τύπων:
- Ενεργή μεταφορά – Αυτή η μεταφορά πραγματοποιείται σε αντίθεση με την κλίση συγκέντρωσης και επομένως απαιτεί ενέργεια. Χρειάζεται επίσης πρωτεΐνες-φορείς και είναι μια εξαιρετικά επιλεκτική διαδικασία.
- Παθητική μεταφορά - Αυτή η μεταφορά συμβαίνει κατά μήκος της βαθμίδας συγκέντρωσης και επομένως, δεν απαιτεί ενέργεια. Έτσι, δεν χρειάζεται πρωτεΐνες-φορείς και δεν είναι επιλεκτικό.

Co-funded by the Erasmus+ Programme of the European Union

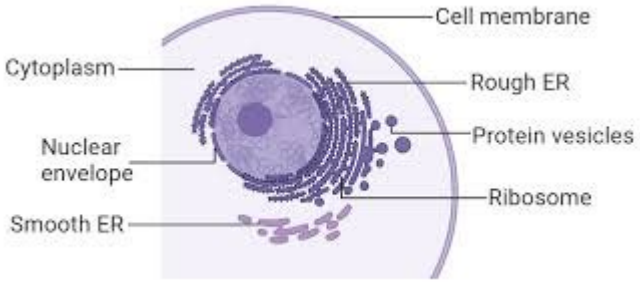
Οργανίδια

Τα οργανίδια είναι δομές μέσα σε ένα κύτταρο που περιβάλλονται από μια μεμβράνη και εκτελούν διαφορετικές λειτουργίες για το κύτταρο. Για παράδειγμα, τα μιτοχόνδρια είναι υπεύθυνα για την παραγωγή ενέργειας για το κύτταρο, ενώ η συσκευή Golgi εμπλέκεται στη διαλογή των πρωτεϊνών, μεταξύ άλλων λειτουργιών.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενδοπλασματικό δίκτυο (ER)



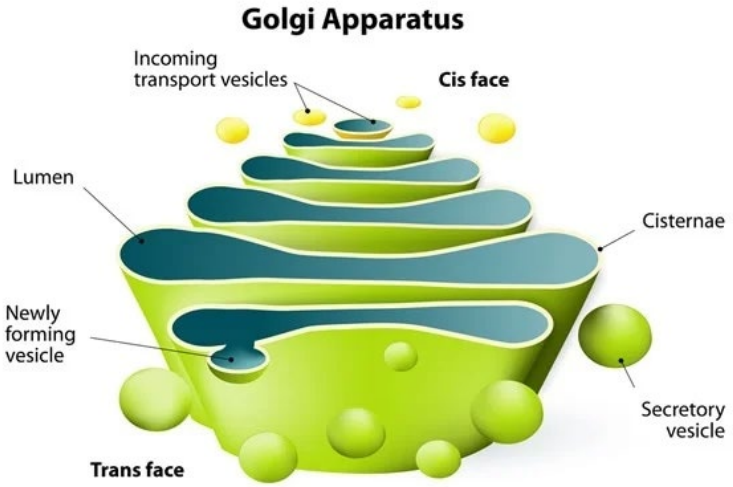
Είναι ένα δίκτυο μικρών, σωληνοειδών κατασκευών. Διαιρεί τον χώρο στο εσωτερικό των ευκαρυωτικών κυττάρων σε δύο μέρη - τον αυλό (μέσα στο ER) και τον εξωαυλικό (κυτταρόπλασμα).

Λειτουργίες

Το SER εμπλέκεται στη σύνθεση λιπιδίων και το RER στη σύνθεση πρωτεϊνών.

Το RER βοηθά στην αναδίπλωση των πρωτεϊνών και τις μεταφέρει στη συσκευή Golgi σε κυστίδια.

Συσκευή Golgi



Πήρε το όνομά του από τον επιστήμονα που το ανακάλυψε, Camillo Golgi. Το Golgi αποτελείται από πολλές επίπεδες δομές σε σχήμα δίσκου που ονομάζονται cisternae. Υπάρχει σε όλα τα ευκαρυωτικά κύτταρα εκτός από τα ανθρώπινα ερυθρά αιμοσφαίρια και τα κύτταρα του κόσκινου των φυτών.

Δομή: Οι δεξαμενές είναι διατεταγμένες παράλληλα και ομόκεντρα κοντά στον πυρήνα ως εξής:

- Cis όψη (όψη σχηματισμού) – Αντικρίζει την πλασματική μεμβράνη και δέχεται εκκρινόμενο υλικό σε κυστίδια.
- Trans όψη (ώριμη όψη) – Κοιτάζει προς τον πυρήνα και απελευθερώνει το λαμβανόμενο υλικό στο κύτταρο.

Λειτουργίες

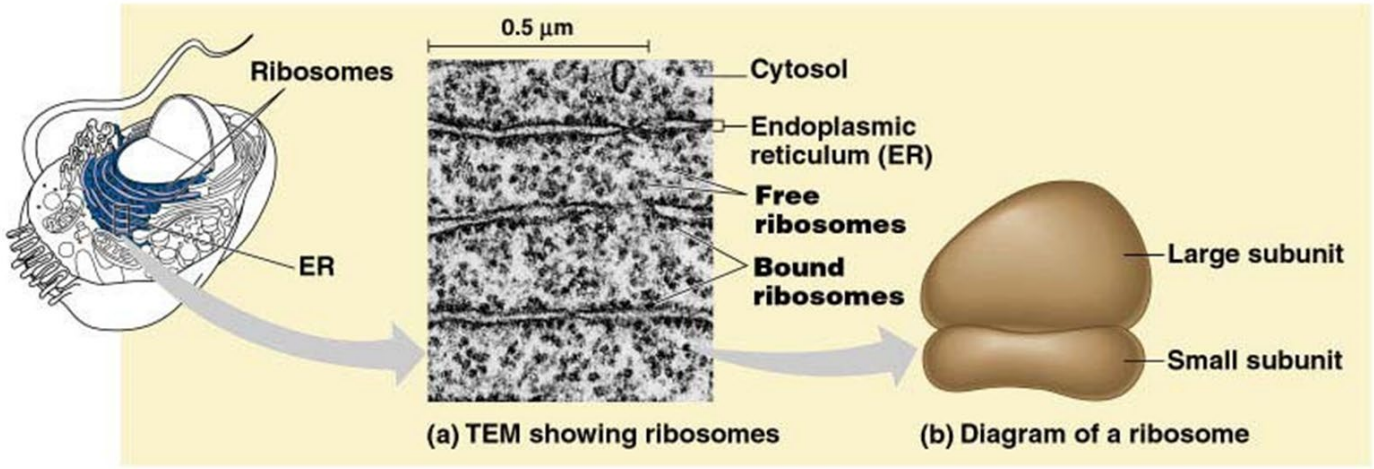
- Ένας σημαντικός χώρος για τη συσκευασία του υλικού μέσα στο κελί.
- Οι πρωτεΐνες τροποποιούνται στο Golgi.
- Σημαντική θέση για το σχηματισμό γλυκολιπιδίων (δηλαδή λιπιδίων με υδατάνθρακες) και γλυκοπρωτεϊνών (δηλαδή πρωτεϊνών με υδατάνθρακες).

Ριβοσώματα

Αυτές οι δομές δεν δεσμεύονται από μια μεμβράνη. Τα ριβοσώματα ονομάζονται επίσης «εργοστάσια πρωτεϊνών» αφού είναι η κύρια περιοχή της πρωτεϊνοσύνθεσης.

Δομή: Αποτελούνται από ριβονουκλεϊκά οξέα και πρωτεΐνες. Τα ευκαρυωτικά ριβοσώματα είναι του τύπου 80S, με 60S (μεγάλη υπομονάδα) και 40S (μικρή υπομονάδα).

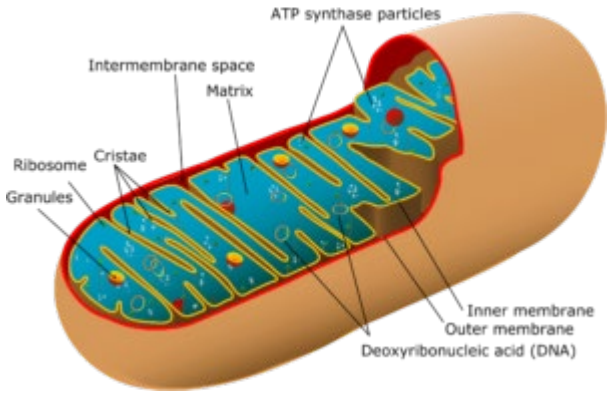
Λειτουργίες: Μια κύρια τοποθεσία για τη σύνθεση πρωτεϊνών και πολυπεπτιδίων (αλυσίδα αμινοξέων).



Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Μιτοχόνδρια



Περιγραφή: Είναι οργανίδια συνδεδεμένα με τη μεμβράνη, γνωστά και ως «ενεργειακά κέντρα του κυττάρου».

Δομή: Έχει δύο μεμβράνες – εξωτερική και εσωτερική. Η εξωτερική μεμβράνη σχηματίζει ένα συνεχές όριο γύρω από τα μιτοχόνδρια. Η εσωτερική μεμβράνη είναι ημιπερατή και χωρίζεται σε πτυχές που ονομάζονται «cristae». Οι μεμβράνες χωρίζουν τον αυλό των μιτοχονδρίων σε εσωτερικό και εξωτερικό διαμέρισμα. Το εσωτερικό διαμέρισμα ονομάζεται μήτρα και το εξωτερικό διαμέρισμα σχηματίζει τον διαμεμβρανικό χώρο.

Λειτουργίες

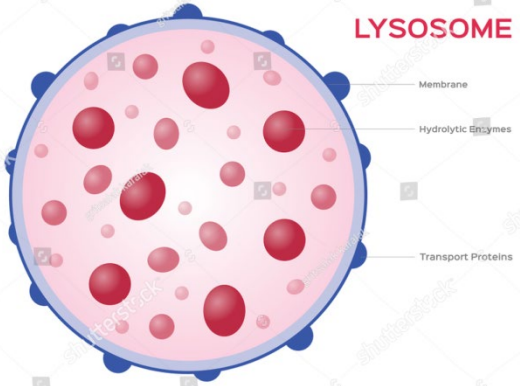
Παράγουν ενέργεια (ATP) και γι' αυτό ονομάζονται «κέντρο ενέργειας του κυττάρου».

Βοηθά στη ρύθμιση του μεταβολισμού των κυττάρων.

Τα μιτοχόνδρια διαθέτουν το δικό τους DNA, RNA και συστατικά που απαιτούνται για τη σύνθεση πρωτεϊνών.

Co-funded by the Erasmus+ Programme of the European Union

Λυσοσώματα



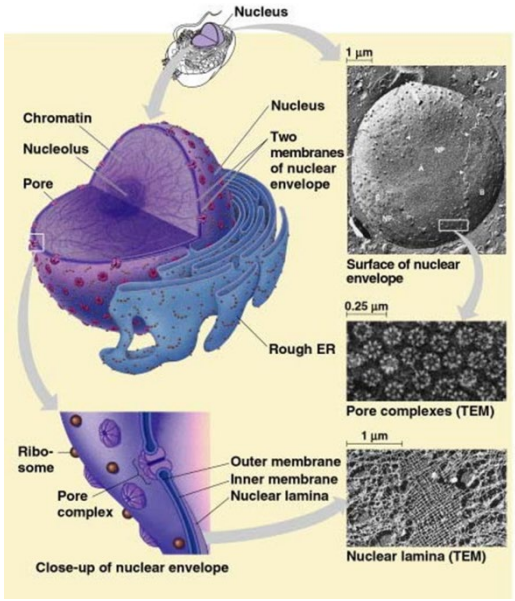
Είναι κυστίδια συνδεδεμένα με μεμβράνη που σχηματίζονται στη συσκευή Golgi. Τα λυσοσώματα ονομάζονται επίσης «αυτοκτονικοί σάκοι» επειδή είναι πλούσια σε υδρολυτικά ένζυμα όπως λιπάσες, πρωτεάσες, υδατάνθρακες κ.λπ. Αυτά τα ένζυμα είναι βέλτιστα ενεργά σε όξινο pH (λιγότερο από 7).

Λειτουργία: Η κύρια λειτουργία των λυσοσωμάτων είναι η πέψη των λιπιδίων, των πρωτεϊνών, των υδατανθράκων και των νουκλεϊκών οξέων.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πυρήνας



Λειτουργίες
Αποθηκεύει γενετικές πληροφορίες (με τη μορφή DNA) που είναι απαραίτητες για την ανάπτυξη και την αναπαραγωγή.
Περιέχει όλες τις απαραίτητες πληροφορίες για τη σύνθεση πρωτεϊνών και τις κυτταρικές λειτουργίες.

Ο πυρήνας είναι το κύριο οργανίδιο ενός κυττάρου. Είναι μια δομή διπλής μεμβράνης με όλες τις γενετικές πληροφορίες. Ως εκ τούτου, ονομάζεται επίσης «εγκέφαλος» ενός κυττάρου. Ο πυρήνας βρίσκεται σε όλα τα ευκαρυωτικά κύτταρα εκτός από τα ανθρώπινα ερυθρά αιμοσφαίρια και τα κοσκινικά κύτταρα των φυτών.

Ένας πυρήνας έχει τα ακόλουθα μέρη:

Πυρηνικός φάκελος – Είναι μια δομή διπλής μεμβράνης που περιβάλλει τον πυρήνα. Η εξωτερική μεμβράνη είναι συνεχής με το ενδοπλασματικό δίκτυο. Η εσωτερική μεμβράνη έχει μικρούς πόρους που ονομάζονται «πυρηνικοί πόροι».

Πυρηνόπλασμα - Είναι το ρευστό υλικό στον πυρήνα που περιέχει τον πυρήνα και τη χρωματίνη.

Πυρήνας – Οι πυρήνες δεν συνδέονται με τη μεμβράνη και είναι ενεργές θέσεις για τη σύνθεση ριβοσωμικού RNA.

Χρωματίνη – Αποτελείται από DNA και πρωτεΐνες που ονομάζονται «ιστόνες». Το DNA είναι οργανωμένο σε χρωμοσώματα. Τα χρωμοσώματα έχουν ορισμένες θέσεις συστολής που ονομάζονται «κεντρομερή». Με βάση τη θέση του κεντρομερούς, μπορούν να χωριστούν ως εξής:

Μετακεντρικό – Με κεντρομερίδιο στο κέντρο και με ίσους βραχίονες χρωμοσωμάτων.

Υπομετακεντρικό - Το Centromere είναι ελαφρώς εκτός κέντρου δημιουργώντας έναν κοντό και έναν μακρύ βραχίονα.

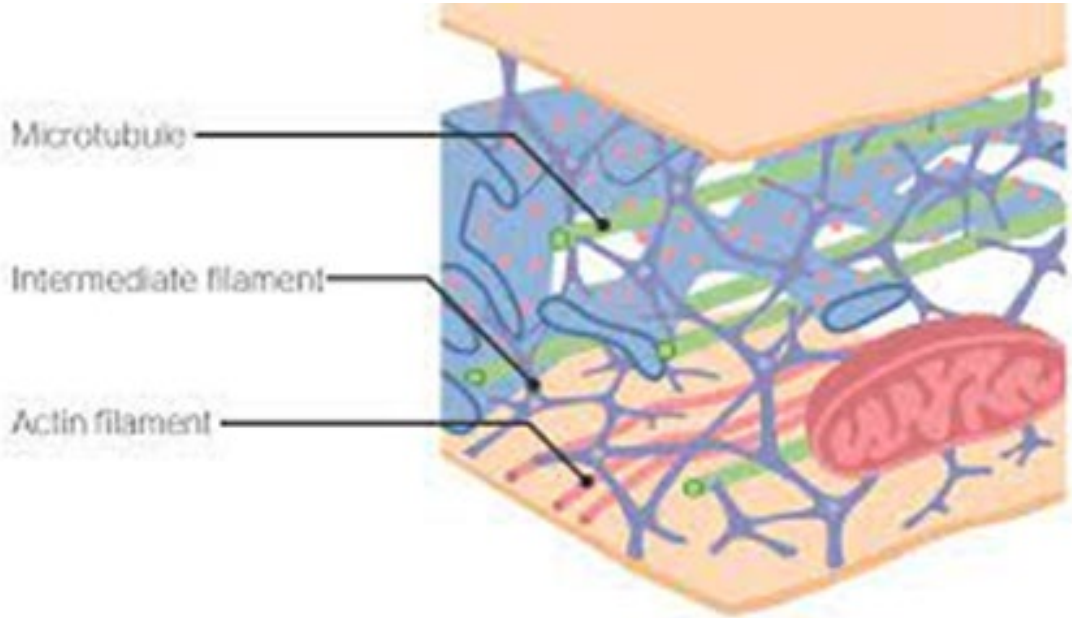
Acrocentric - Το Centromere είναι εξαιρετικά εκτός κέντρου με έναν πολύ μακρύ και έναν πολύ κοντό βραχίονα χρωμοσώματος.

Τελοκεντρικό – Το κεντρομερές τοποθετείται στο ένα άκρο του χρωμοσώματος. Οι άνθρωποι δεν διαθέτουν τελοκεντρικά χρωμοσώματα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Κυτοσκελετός



Κυτοσκελετός

-ένα δίκτυο ινών που εκτείνεται σε όλο το κυτταρόπλασμα

-λειτουργία:

παρέχουν μηχανική αντοχή στο κύτταρο καθορίζουν το σχήμα του κυττάρου

κίνηση (διάφοροι τύποι κινητικότητας κυττάρων) ενδοκυτταρική μεταφορά οργανιδίων

-3 κύριοι τύποι ινών:

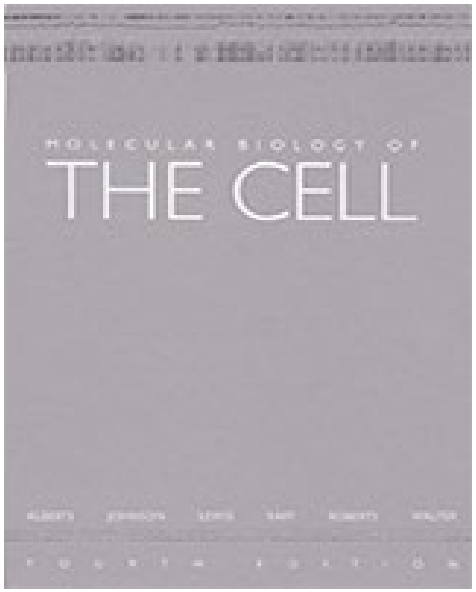
1. μικροσωληνίσκοι: καθορίζουν τις θέσεις των οργανιδίων που περικλείονται από τη μεμβράνη και της ενδοκυτταρικής μεταφοράς

2. μικρονήμα: προσδιορίζει το σχήμα του κυττάρου και είναι απαραίτητο για την κίνηση ολόκληρου του κυττάρου

3. ενδιάμεσο νήμα: παρέχει μηχανική αντοχή και αντοχή στη διατμητική τάση

Ερωτήσεις

- Ποια είναι η διαφορά μεταξύ προκαρυωτικών και ευκαρυωτικών κυττάρων;
- Περιγράψτε τα οργανίδια και ποιες είναι οι λειτουργίες τους;



Μοριακή Βιολογία του Κυττάρου. 4η έκδοση.
Alberts B, Johnson A, Lewis J, et al.
Νέα Υόρκη: Garland Science; 2002.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 2. Βασικές αρχές της Μοριακής Βιολογίας

Μάθημα 3. Λιπίδια. Κατασκευή και λειτουργίες. Βιομεμβράνες και κυτταρική αρχιτεκτονική. Λιπιδική δομή και δομική οργάνωση βιομεμβρανών . Πρωτεϊνικά συστατικά και κύριες λειτουργίες στις βιομεμβράνες . Μεταφορά ιόντων και μικρών μορίων μέσω των κυτταρικών μεμβρανών

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Λιπίδια
- Κατασκευή και λειτουργίες. Βιομεμβράνες και κυτταρική αρχιτεκτονική.
- Πρωτεϊνικά συστατικά και κύριες λειτουργίες στις βιομεμβράνες .
- Μεταφορά ιόντων και μικρών μορίων μέσω των κυτταρικών μεμβρανών

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

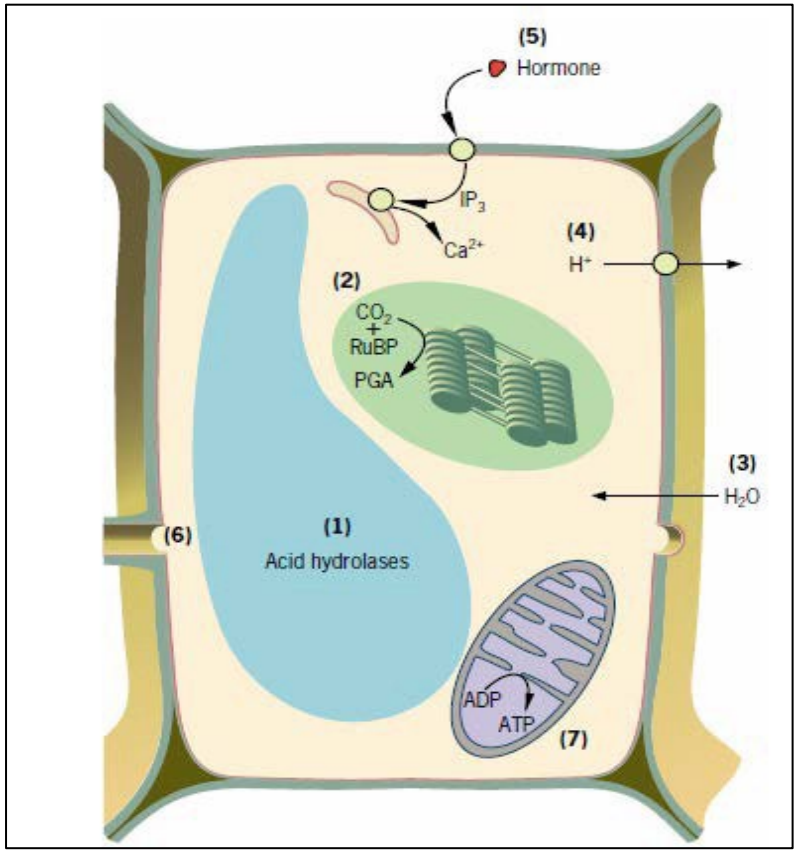
Τα κύτταρα είναι οι βασικές μονάδες όλης της ζωής. Αποτελούν κάθε όργανο κάθε ζώου, φυτού, μύκητα και βακτηρίων. Τα κύτταρα σε ένα σώμα είναι σαν τα δομικά στοιχεία ενός σπιτιού. Έχουν επίσης μια συγκεκριμένη βασική δομή που μοιράζονται τα περισσότερα κελιά. Τα κύτταρα συνήθως αποτελούνται από:

Η κυτταρική μεμβράνη - αυτή είναι μια λιπιδική διπλοστιβάδα που σηματοδοτεί τα όρια του κυττάρου. Μέσα σε αυτό, μπορούμε να βρούμε τα άλλα δύο βασικά συστατικά του κυττάρου: το DNA και το κυτταρόπλασμα. Όλα τα κύτταρα έχουν κυτταρική ή πλασματική μεμβράνη.

DNA - το DNA περιέχει τις οδηγίες ώστε το κύτταρο να μπορεί να λειτουργήσει. Το γενετικό υλικό μπορεί να προστατεύεται εντός του πυρήνα (ευκαρυωτικά κύτταρα) ή να επιπλέει στο κυτταρόπλασμα (προκαρυωτικά κύτταρα). Τα περισσότερα κύτταρα έχουν DNA, αλλά τα ερυθρά αιμοσφαίρια, για παράδειγμα, δεν έχουν.

Κυτόπλασμα - το κυτταρόπλασμα είναι η παχύρρευστη ουσία εντός της πλασματικής μεμβράνης στην οποία επιπλέουν τα άλλα συστατικά ενός κυττάρου (το DNA/πυρήνας και άλλα οργανίδια).

✓ Ολα κύτταρα είναι περικυκλωμένα με μεμβράνες



- Εκλεκτικός εμπόδιο
- **Αλλά επίσης σπουδαίος για:**

1. Διαμερισματικοποίηση
2. Βιοχημική δραστηριότητες
3. Μεταφορά διαλυμένων ουσίες
4. Μεταφορά των ιόντων
5. Σήμα μεταγωγής
6. ΑΛΛΗΛΕΠΙΔΡΑΣΗ Κύτταρο- κύτταρο
7. Ενέργεια μετατροπή

- **Δυναμικές δομές:**

1. Συνεχείς κινήσεις
2. Συνεχές κτίσιμο και υποβάθμιση των συστατικών τους

Co-funded by the Erasmus+ Programme of the European Union

❖ Ολες οι βιολογικός μεμβράνες έχουμ κοινή βασική δομή:

✓πολύ λεπτό στρώμα του λιπίδιου και μόρια πρωτεΐνης συνδεδεμένα με μη ομοιοπολικό δεσμό

✓ δυναμικές και ρευστές δομές

✓ βιοχημική σύνθεση: **λιπίδια, πρωτεΐνες, σάκχαρα**

- μεμβράνες με παρόμοιες λειτουργίες (π.χ από τα ίδια οργανίδια) είναι παρόμοιες σε διαφορετικός κύτταρα
- μεμβράνες με διαφορετικές λειτουργίες (π.χ διαφορετικά οργανίδια) είναι πολύ διαφορετικές στα πλαίσια του ίδιου κύτταρο

✓ **Λιπίδια**

- διπλό δι-στρώμα (πάχος 5-10 nm)
- βασική δομή υγρού

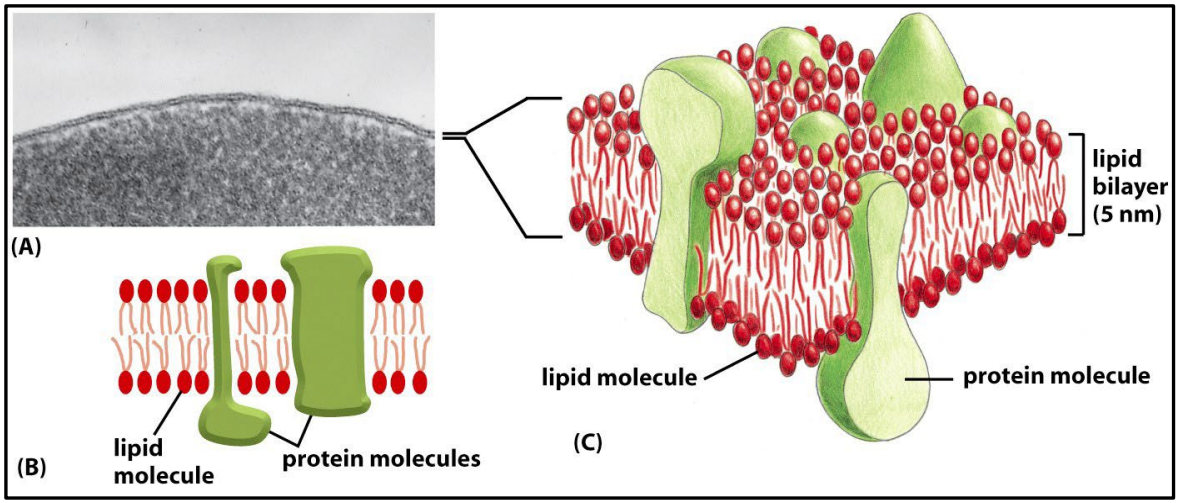
✓ **Πρωτεΐνες**

- εμπλέκονται σε λειτουργίες της μεμβράνης
- μεταφορά, κατάλυση, δομή, υποδοχείς

Α – ΕΜ-φωτογραφία μεμβράνης ερυθροκύτταρου

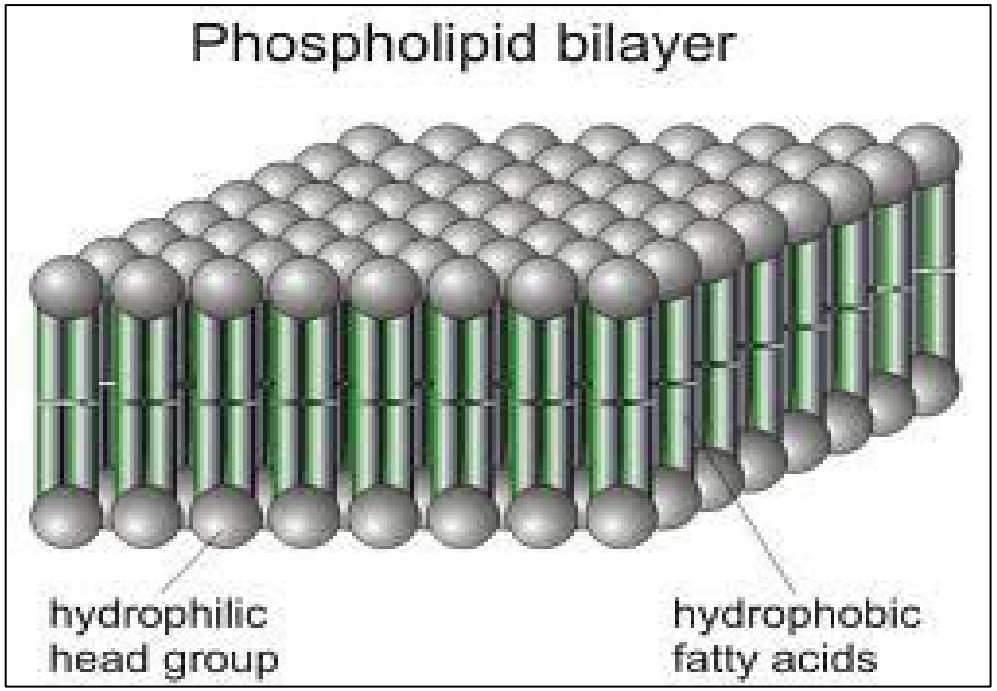
Β – 2D μεμβράνη

Γ – 3D μεμβράνη



Co-funded by the Erasmus+ Programme of the European Union

Λιπίδια Μεμβράνης



- ✓ Αμφιπαθητικά μόρια
- ✓ Αυθόρμητος σχηματισμός του διπλού στρώματος σε υδατικό διάλυμα

✓Φωσφολιπίδια

- φωσφατιδυλοχολίνη
- σφιγγομυελίνη

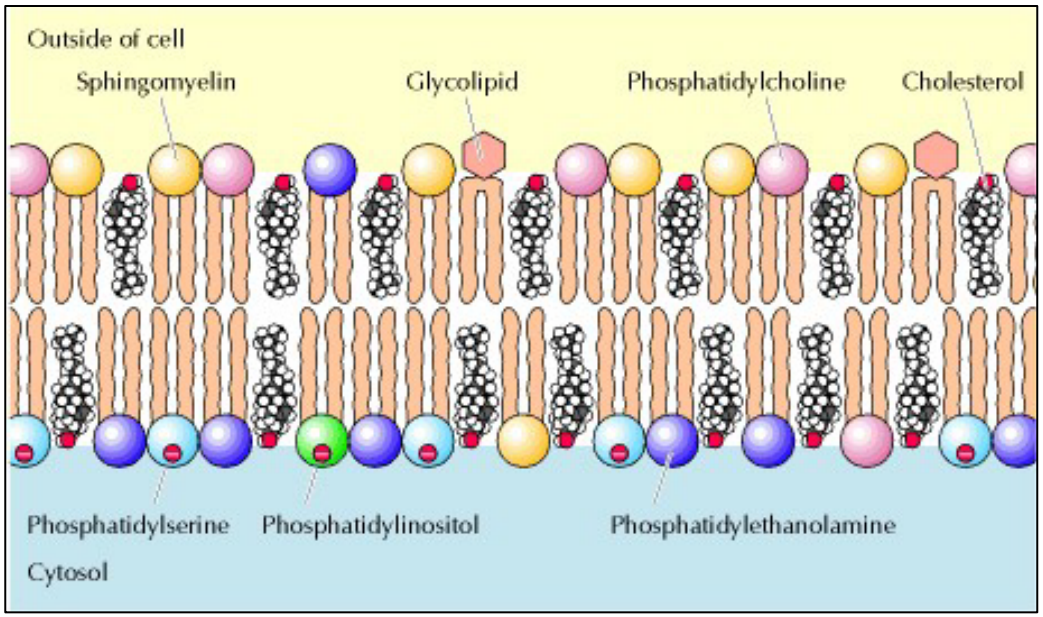
} **εξωτερική (εξωκυτταρική) γλωχίνη**

- φωσφατιδυλαιθανολαμίνη
- φωσφατιδυλοσερίνη
- φωσφατιδυλινοσιτόλη

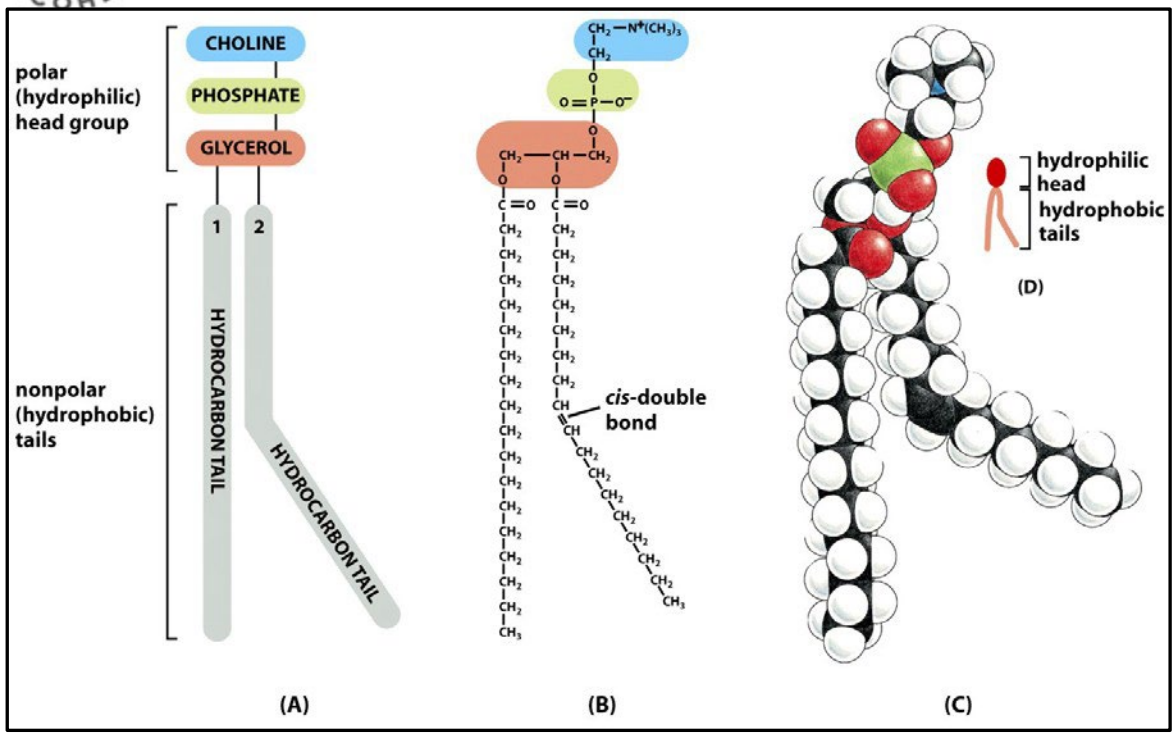
} **εσωτερική (κυτταροπλασματική) γλωχίνη**

✓Χοληστερίνη

✓Γλυκολιπίδια



Co-funded by the Erasmus+ Programme of the European Union



- **Φωσφατιδυλοχολίνη**

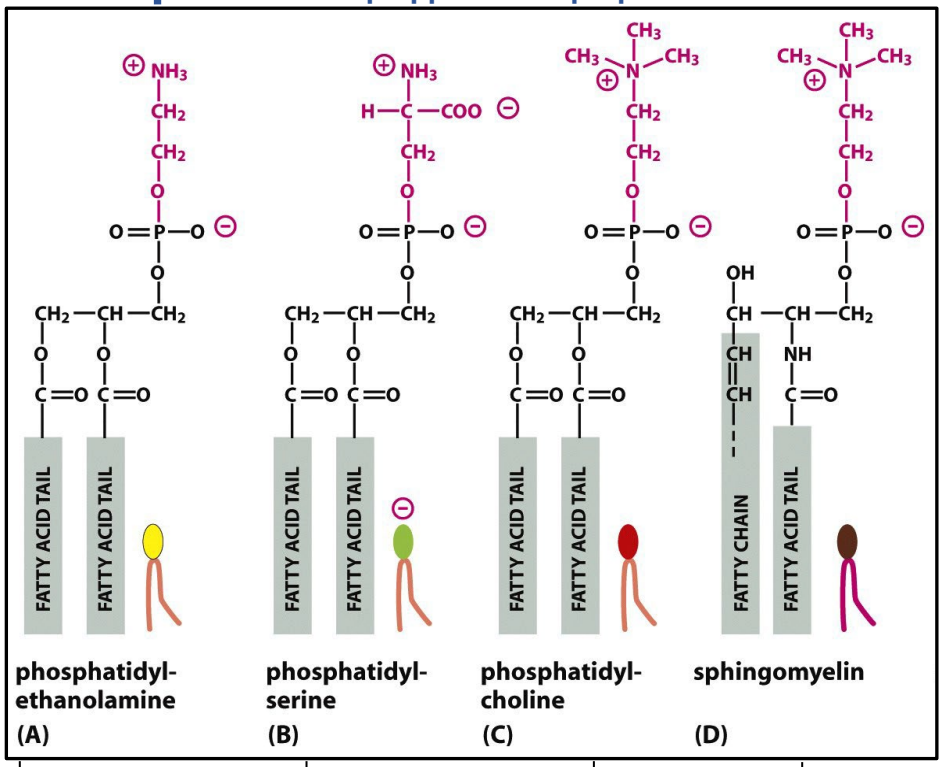
✓ πολική κεφαλή + δύο υδροφοβικές αλυσίδες υδατάνθρακα

✓ ουρές - λιπαρά οξέα (14 – 24 C άτομα)

→ 1η ουρά - όχι διπλούς δεσμούς (**κορεσμένο**)
→ 2η ουρά - 1 ή περισσότεροι cis- διπλοί δεσμοί (**ακόρεστα**)

✓ διαφορές σε μήκη και κορεσμούς → ρευστότητα

- ✓ **Βακτήρια** – ως επί το πλείστον ένας τύπος φωσφολιπίδιου; όχι χοληστερίνη
- ✓ **Ευκαρυώτα** – μίγμα διαφορετικών τύπων φωσφολιπίδιων + χοληστερίνη



εξωτερική γλωχίνη:

→ φωσφατιδυλοχολίνη

→ σφιγγομυελίνη

Εσωτερική γλωχίνη:

→ φωσφατιδυλαιθανολαμίνη

→ φωσφατιδυλοσερίνη *

→ φωσφατιδυλινοσιτόλη

συμπληρωματικός από γλυκερίνη

(φωσφογλυκερίδια)

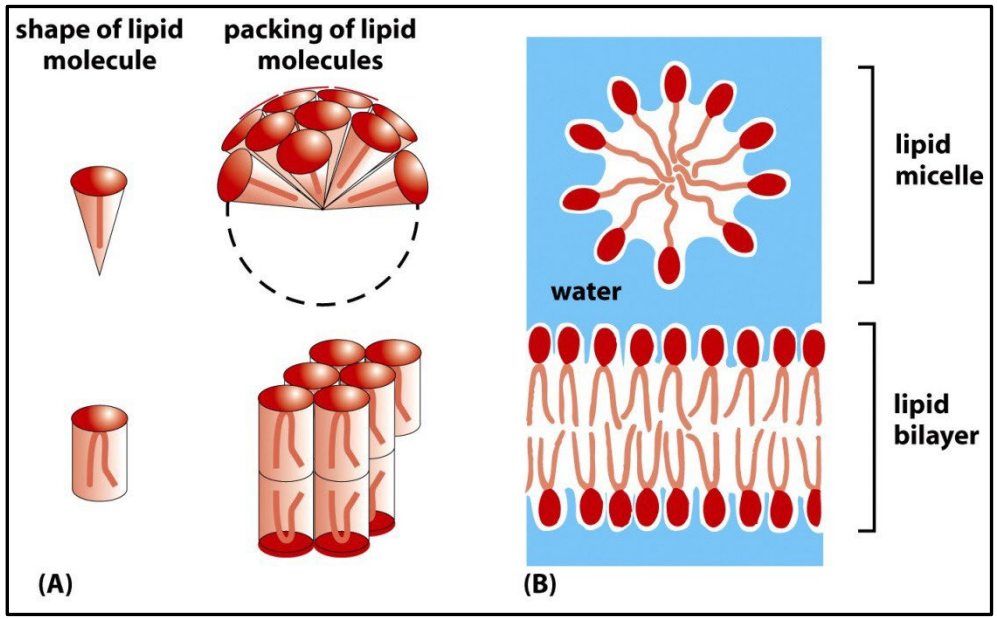
συμπληρωματικός από σφιγγοσίνη

(σφιγγολιπίδιο)

* μετατόπιση προς το εξωτερικό φυλλάδιο
→ απόπτωση

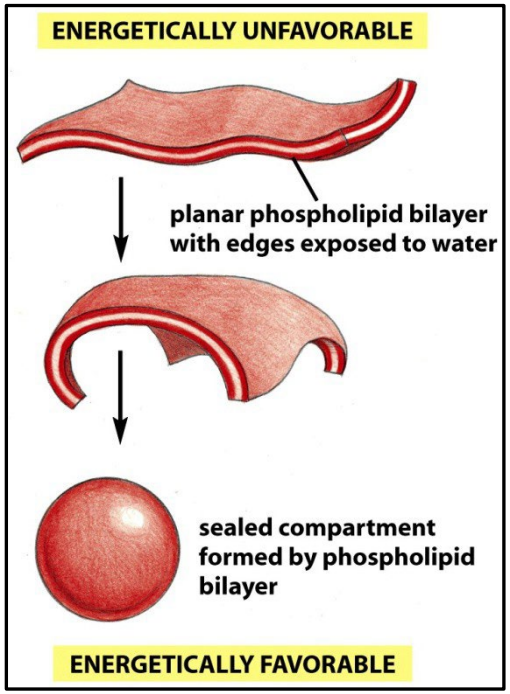
Εικόνα 10-3 *Μοριακή Βιολογία του Κυττάρου* (© Garland Science 2008)

Αυθόρμητος σχηματισμός του διπλού στρώματος λιπιδίου



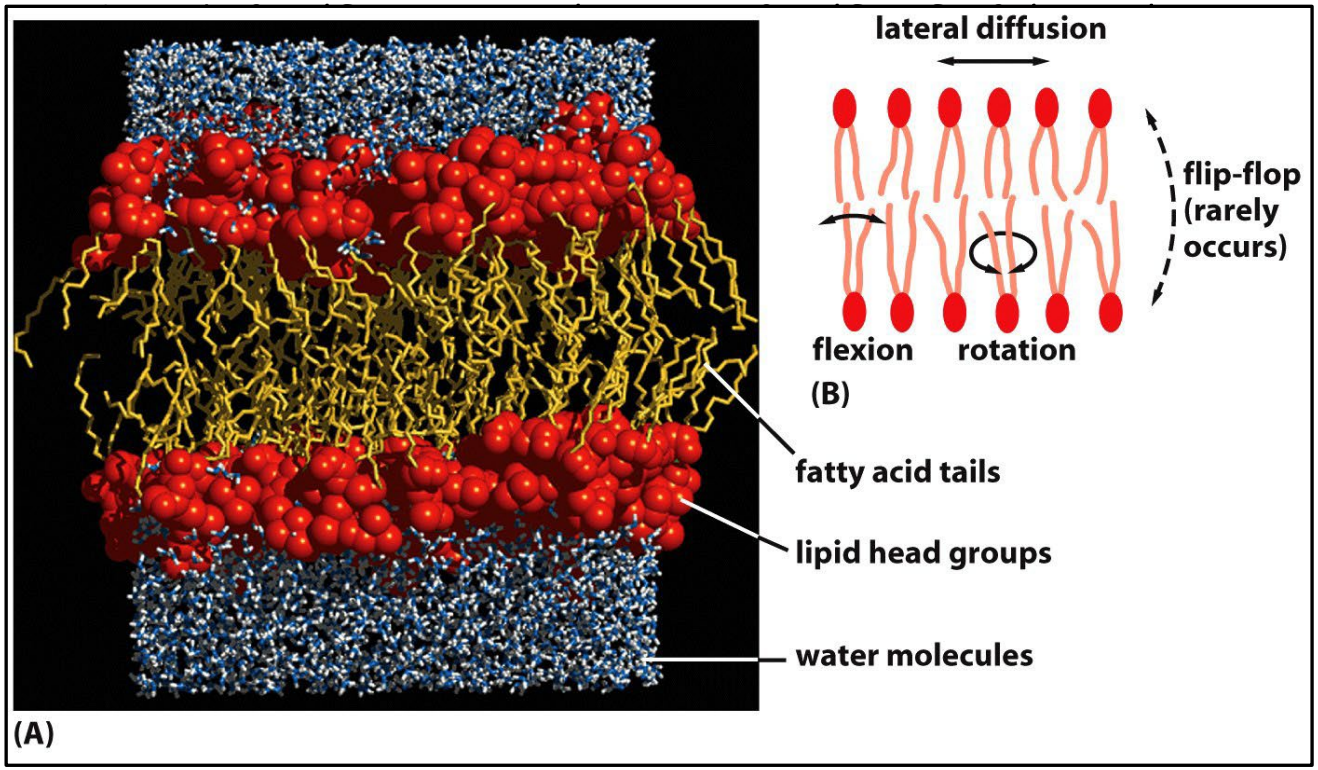
- μία ουρά → **μικκύλια**
- δύο ουρές → **διπλό στρώμα**

✓ **ενεργειακά πλέον ευνοημένη διανομή**



Αυθόρμητο κλείσιμο των λιπιδίων διπλό στρώμα

Co-funded by the Erasmus+ Programme of the European Union



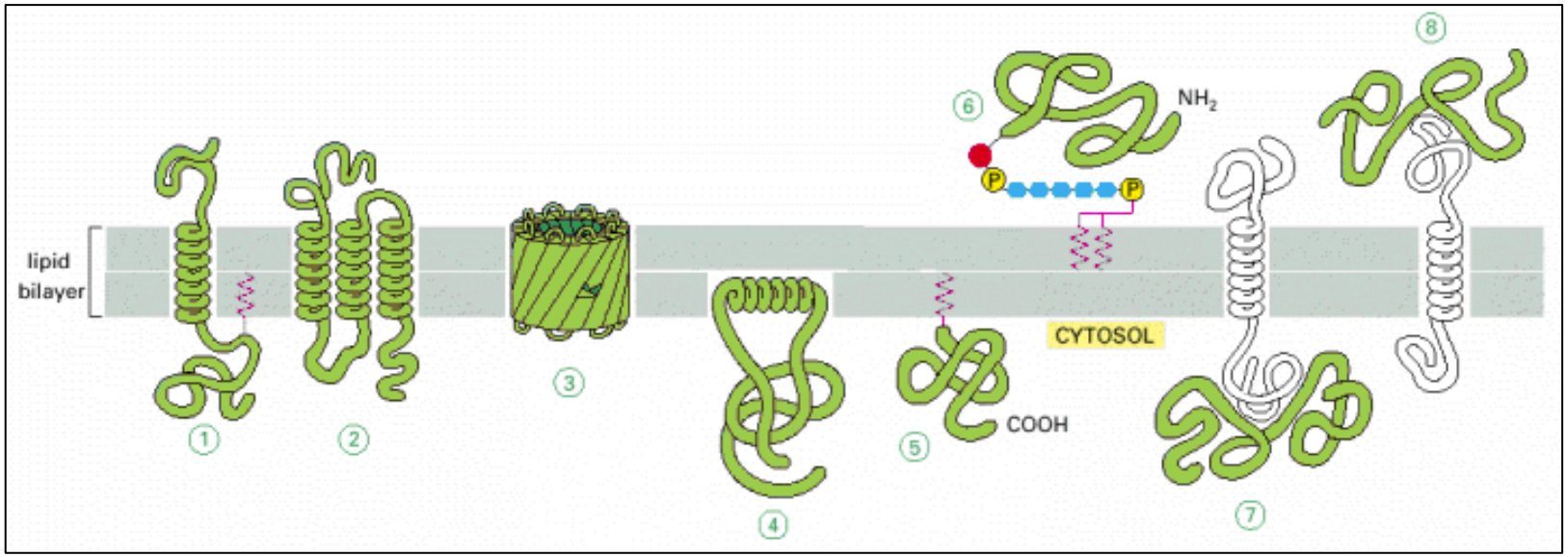
- ✓ **Φλιπ - Φλοπ** – σπάνιος (< 1x ανά μήνα)
- ✓ **Πλευρική διάχυση** – συχνό (~10^{7} ανά δευτ.)
- ✓ **Περιστροφή**
- ✓ **Κάμψη**

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πρωτεΐνες Μεμβράνης

- **Οι πρωτεΐνες μεμβράνης σχετίζονται με τη λιπιδική διπλοστιβάδα με διάφορος τρόπους**



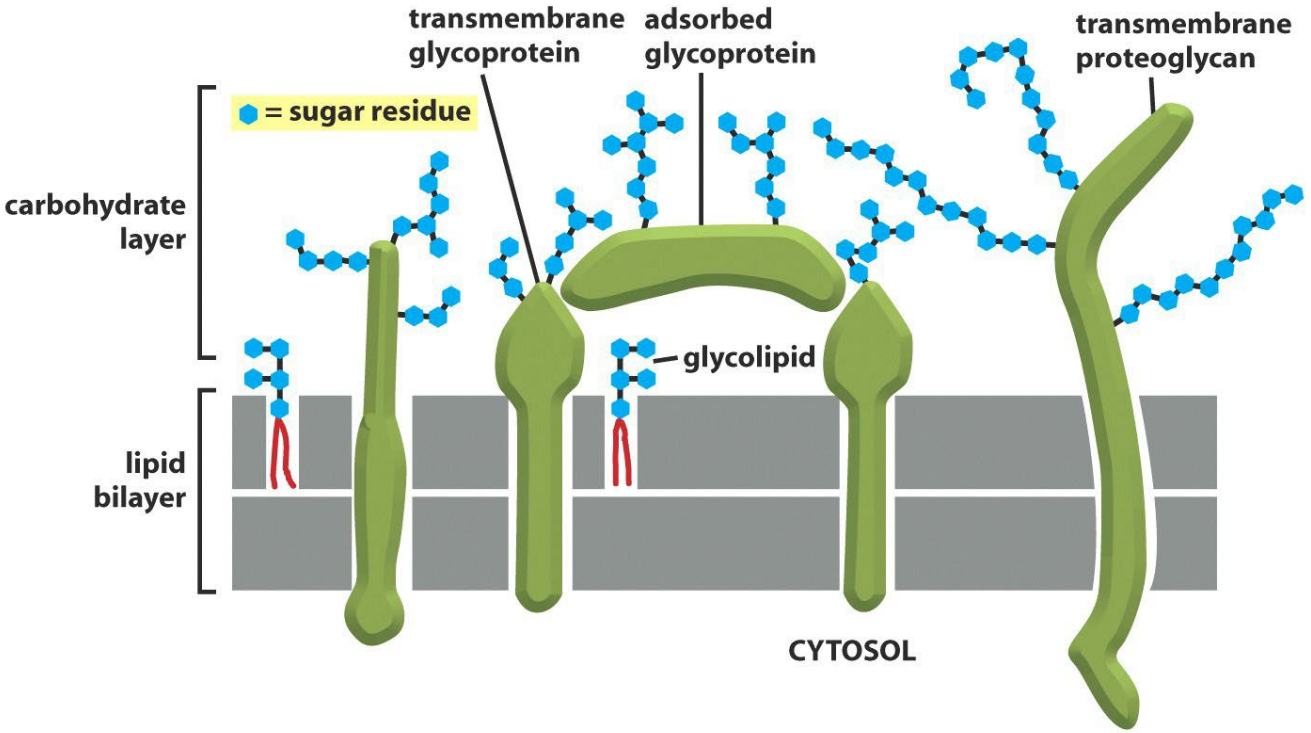
- ✓ **1, 2, 3** – διαμεμβρανικές πρωτεΐνες (αμφιπαθείς)
- ✓ **4, 5, 6** – αγκυρομένες πρωτεΐνες (εκτεθειμένο στο μόνο μια πλευρά)
- ✓ **7, 8** – περιφερειακές πρωτεΐνες (μη ομοιοπολικές αλληλεπιδράσεις με άλλες πρωτεΐνες)

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

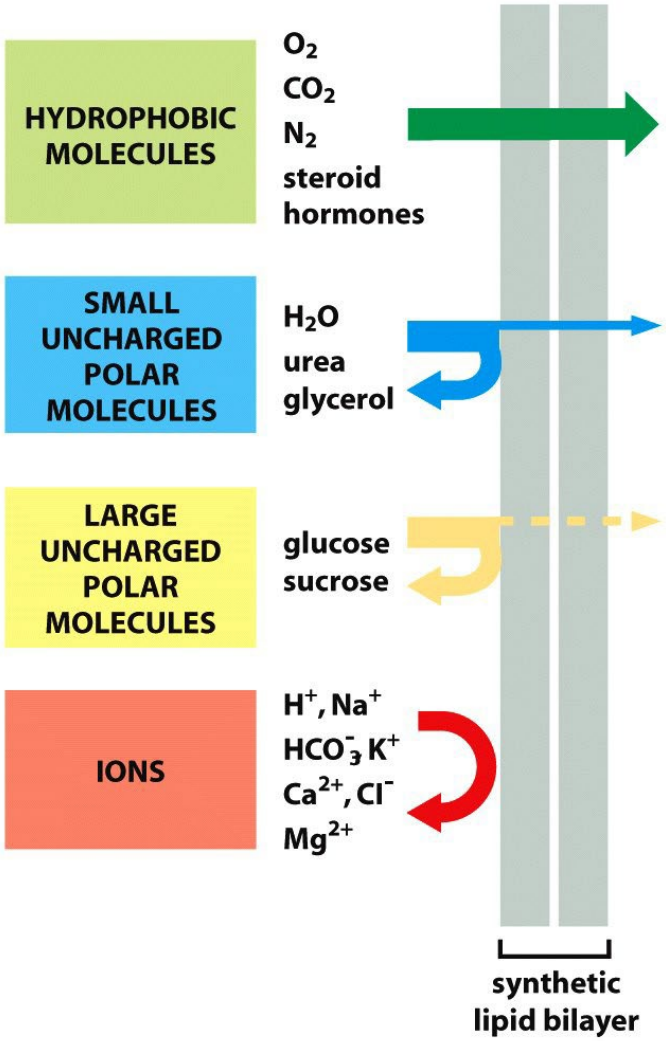
Co-funded by the Erasmus+ Programme of the European Union

Υδατάνθρακες μεμβράνης

- ✓ Ενώνονται με **πρωτεΐνες** ή **λιπίδια**
- ✓ Μόνο σε εξωτερικές γλωχίνες



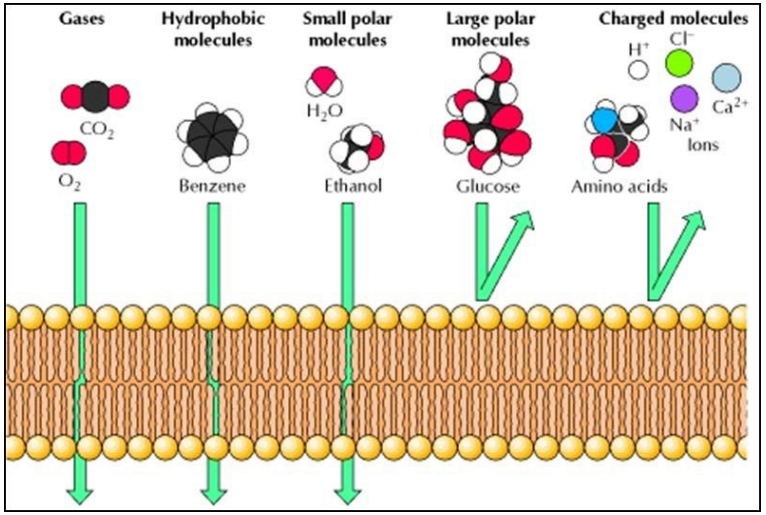
Σχετική διαπερατότητα του διπλού στρώματος συνθετικού λιπιδίου σε διαφορετικά είδη μορίων



- ✓ μικρά αφόρτιστα μόρια μπορούν να διαχέονται ελεύθερα μέσω του διπλού στρώματος φωσφολιπιδίου

- ✓ Το διπλό στρώμα είναι αδιαπέραστο από:
  - μεγαλύτερα πολικα μόρια
  - (τέτοια όπως η γλυκόζη και τα αμινοξέα)
  - ιόντων

✓όσο μικρότερο το μόριο και πιο σημαντικα, όσο λιγότερο αλληλεπιδρά με το νερό, τόσο πιο γρήγορα διαχέεται στο διπλό στρώμα

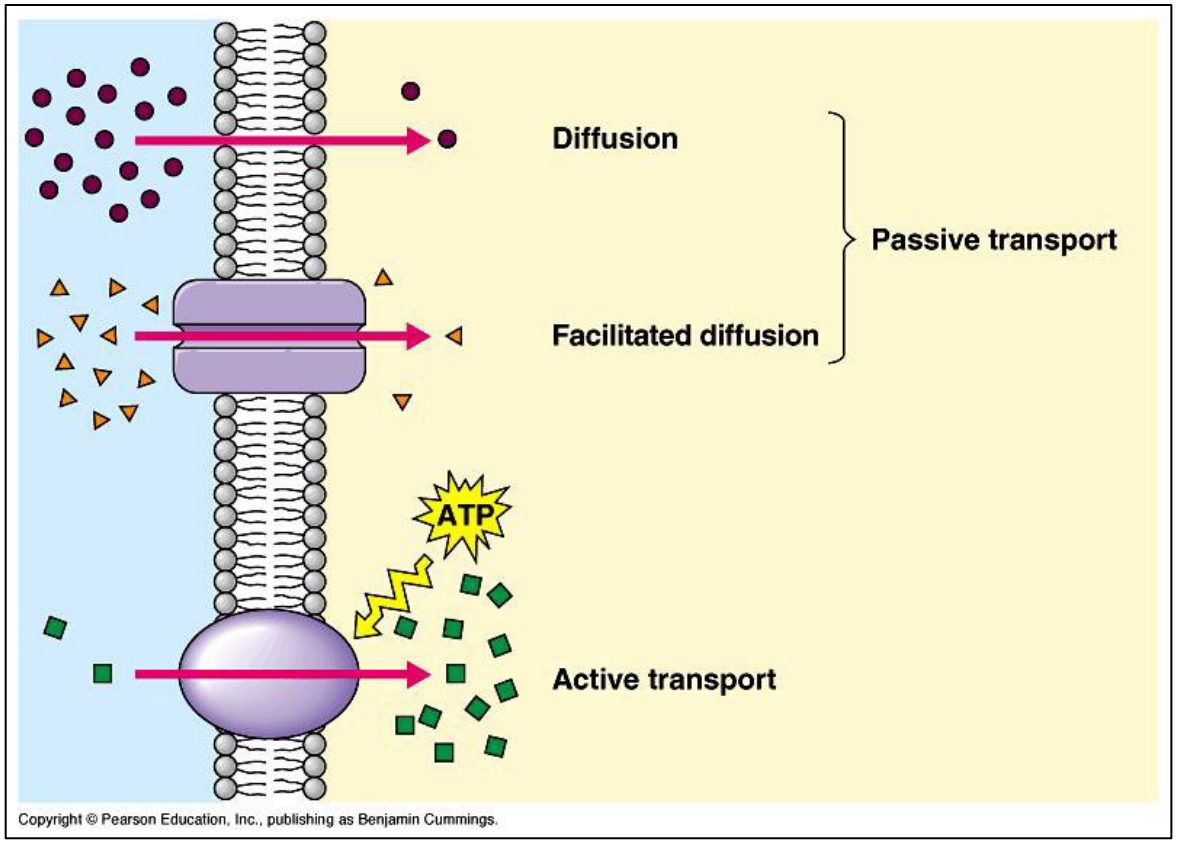


Τρεις τρόποι διάβασης της μεμβράνης από τα μόρια

- Διάχυση
- Υποβοηθούμενη διάχυση

Παθητική μεταφορά – προς μικρότερη συγκέντρωση!

- Ενεργός μεταφορά



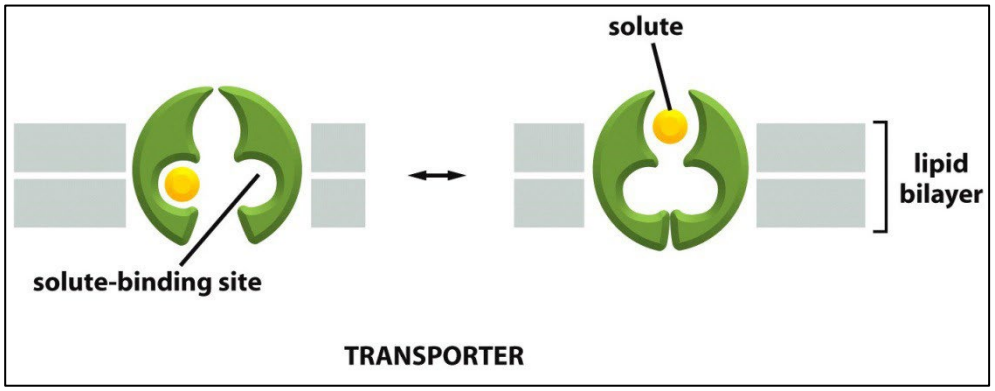
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

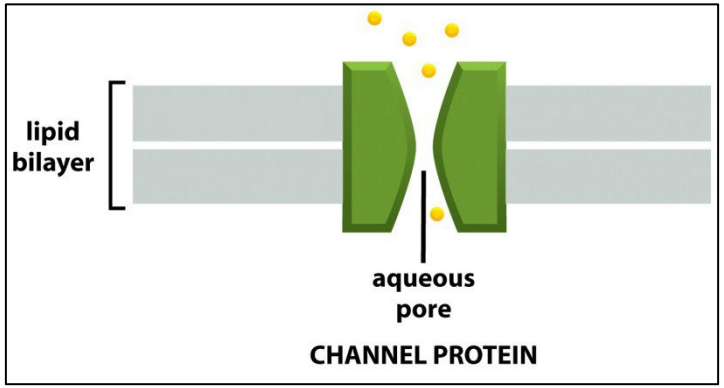
Co-funded by the Erasmus+ Programme of the European Union

Δύο κύρια είδη πρωτεϊνών μεταφοράς μεμβράνης

✓ **Μεταφορείς** → δένεται με το ειδικό διαλυτό προς μεταφορά και υφίσταται μια σειρά από διαμορφωτικές αλλαγές μεταφέροντας το διαμέσου της μεμβράνης

✓ **Κανάλια** → αλληλεπιδρά με το διαλυτό προς μεταφορά πολύ πιο ασθενώς





A - μεταφορέας πρωτεΐνη εναλλάσσεται μεταξύ δύο διαμορφώσεων,
→ διαλυτό – δεσμευτική τοποθεσία είναι διαδοχικά προσβάσιμο από τη μια πλευρά του διπλού στρώματος και μετά από την άλλη

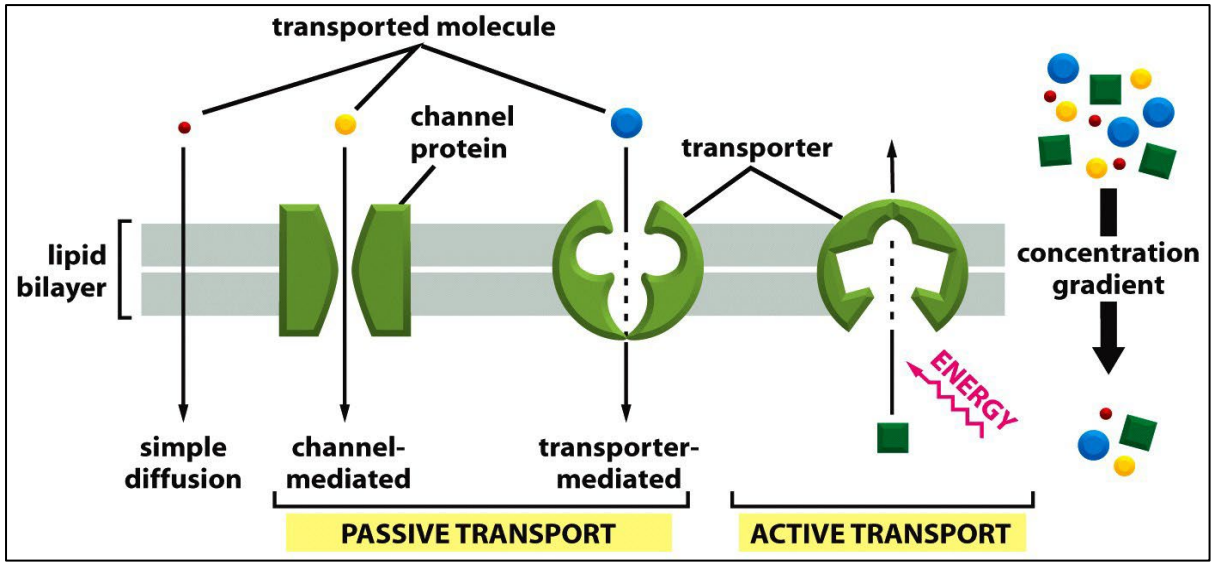
B - Κανάλι πρωτεΐνη σχηματίζει έναν πόρο γεμάτο νερό στο διπλό στρώμα διά μέσου του οποίου συγκεκριμένες διαλυμένες ουσίες μπορούν να διαχυθούν

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Παθητική και ενεργός μεταφορά

- ✓ **Παθητική μεταφορά** – όλες οι πρωτεΐνες - κανάλια και Πολλές μεταφορείς
- ✓ **Ενεργός μεταφορά** – μόνο μεταφορείς. Απαιτεί ενέργεια



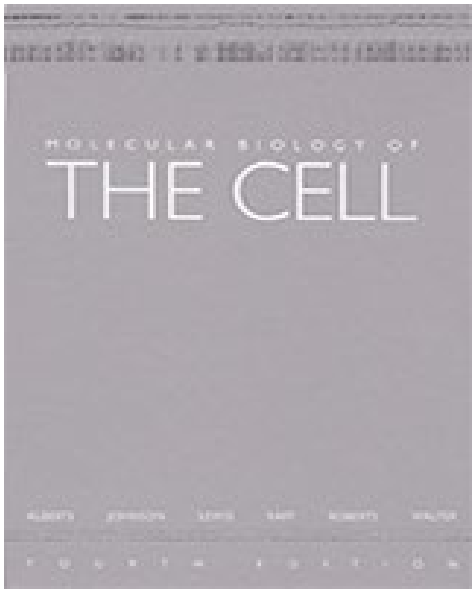
✓ **Παθητική μεταφορά** κατά την ηλεκτροχημική βαθμίδα - λαμβάνει χώρα αυθόρμητα
→ απλή διάχυση διά μέσου του διπλού στρώματος του λιπιδίου
→ επιβοηθούμενη διάχυση διά μέσου καναλιών και παθητικών μεταφορέων

✓ **Ενεργός μεταφορά**
→ απαιτεί ενέργεια μεταβολισμού
→ βοηθούμενη από μεταφορείς που χρησιμοποιούν μεταβολική ενέργεια για να μεταφέρουν το διαλυτό.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Ποιες είναι οι λειτουργίες των λιπιδίων στο κύτταρο;
- Εξηγήστε τις κύριες λειτουργίες στις βιομεμβράνες.
- Εξηγήστε τη μεταφορά ιόντων και μικρών μορίων μέσω των κυτταρικών μεμβρανών.



Μοριακή Βιολογία του Κυττάρου. 4η έκδοση.
Alberts B, Johnson A, Lewis J, et al.
Νέα Υόρκη: Garland Science; 2002.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 2. Βασικές αρχές της Μοριακής Βιολογίας

Μάθημα 4. Κυτταρική ενέργεια. Οξείδωση γλυκόζης και λιπαρών οξέων σε διοξείδιο του άνθρακα

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

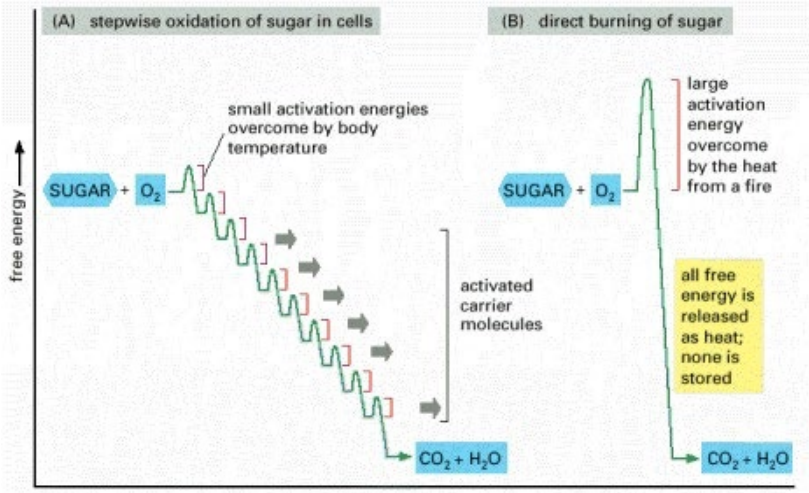
- Εισαγωγή
- Τα μόρια των τροφίμων διασπώνται σε τρία στάδια για να παραχθεί ATP
- Η γλυκόλυση είναι μια κεντρική οδός που παράγει ATP
- Οι ζυμώσεις επιτρέπουν την παραγωγή του ATP απουσία οξυγόνου
- Η γλυκόλυση απεικονίζει πώς τα ένζυμα συνδέουν την οξείδωση με την αποθήκευση ενέργειας
- Τα σάκχαρα και τα λίπη διασπώνται σε ακετυλικό CoA στα μιτοχόνδρια
- Ο κύκλος του κιτρικού οξέος δημιουργεί NADH οξειδώνοντας ομάδες ακετυλίου σε CO_2

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

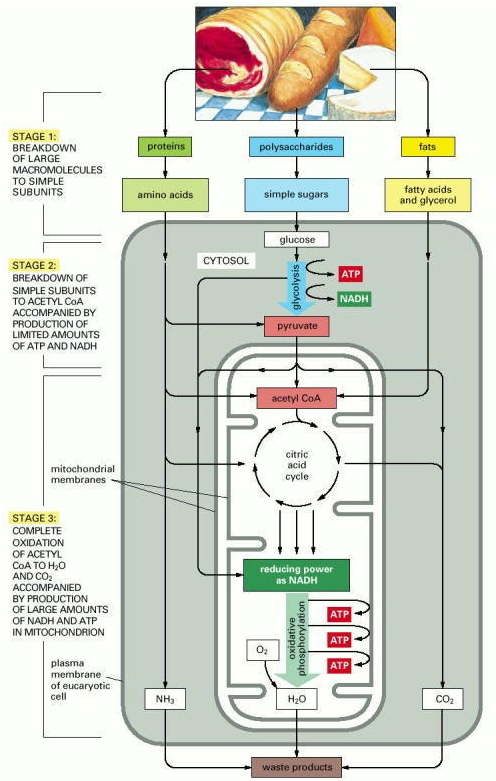
Τα κύτταρα απαιτούν μια συνεχή παροχή ενέργειας για να δημιουργήσουν και να διατηρήσουν τη βιολογική τάξη που τα κρατά ζωντανά. Αυτή η ενέργεια προέρχεται από την ενέργεια του χημικού δεσμού στα μόρια των τροφίμων, τα οποία έτσι χρησιμεύουν ως καύσιμο για τα κύτταρα.

Co-funded by the Erasmus+ Programme of the European Union



Σχηματική αναπαράσταση της ελεγχόμενης σταδιακής οξείδωσης του σακχάρου σε ένα κύτταρο, σε σύγκριση με τη συνηθισμένη καύση
(A) Στο κύτταρο, τα ένζυμα καταλύουν την οξείδωση μέσω μιας σειράς μικρών βημάτων στα οποία η ελεύθερη ενέργεια μεταφέρεται σε πακέτα βολικού μεγέθους σε μόρια-φορείς - πιο συχνά ATP και NADH. Σε κάθε βήμα, ένα ένζυμο ελέγχει την αντίδραση μειώνοντας το ενεργειακό φράγμα ενεργοποίησης που πρέπει να ξεπεραστεί πριν συμβεί η συγκεκριμένη αντίδραση. Η συνολική ελεύθερη ενέργεια που απελευθερώνεται είναι ακριβώς η ίδια στα (A) και (B). Αλλά αν το σάκχαρο οξειδωθεί σε CO2 και H2O σε ένα μόνο βήμα, όπως στο (B), θα απελευθέρωνε μια ποσότητα ενέργειας πολύ μεγαλύτερη από αυτή που θα μπορούσε να δεσμευτεί για χρήσιμους σκοπούς.

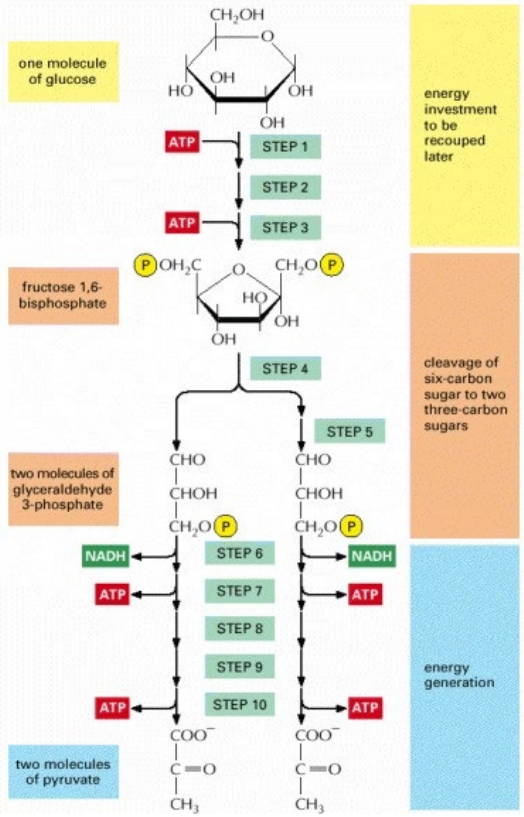
Τα μόρια των τροφίμων διασπώνται σε τρία στάδια για να παραχθεί ATP



Απλοποιημένο διάγραμμα των τριών σταδίων του κυτταρικού μεταβολισμού που οδηγούν από τα τρόφιμα στα απόβλητα σε ζωικά κύτταρα

Αυτή η σειρά αντιδράσεων παράγει ATP, το οποίο στη συνέχεια χρησιμοποιείται για να οδηγήσει τις βιοσυνθετικές αντιδράσεις και άλλες διαδικασίες που απαιτούν ενέργεια στο κύτταρο. Το στάδιο 1 εμφανίζεται έξω από τα κύτταρα. Το στάδιο 2 εμφανίζεται κυρίως στο κυτταρόπλασμα, εκτός από το τελικό στάδιο της μετατροπής του πυροσταφυλικού σε ακετυλομάδες στο ακετυλο CoA, το οποίο εμφανίζεται στα μιτοχόνδρια. Το στάδιο 3 εμφανίζεται στα μιτοχόνδρια.

Η γλυκόλυση είναι μια κεντρική οδός που παράγει ATP



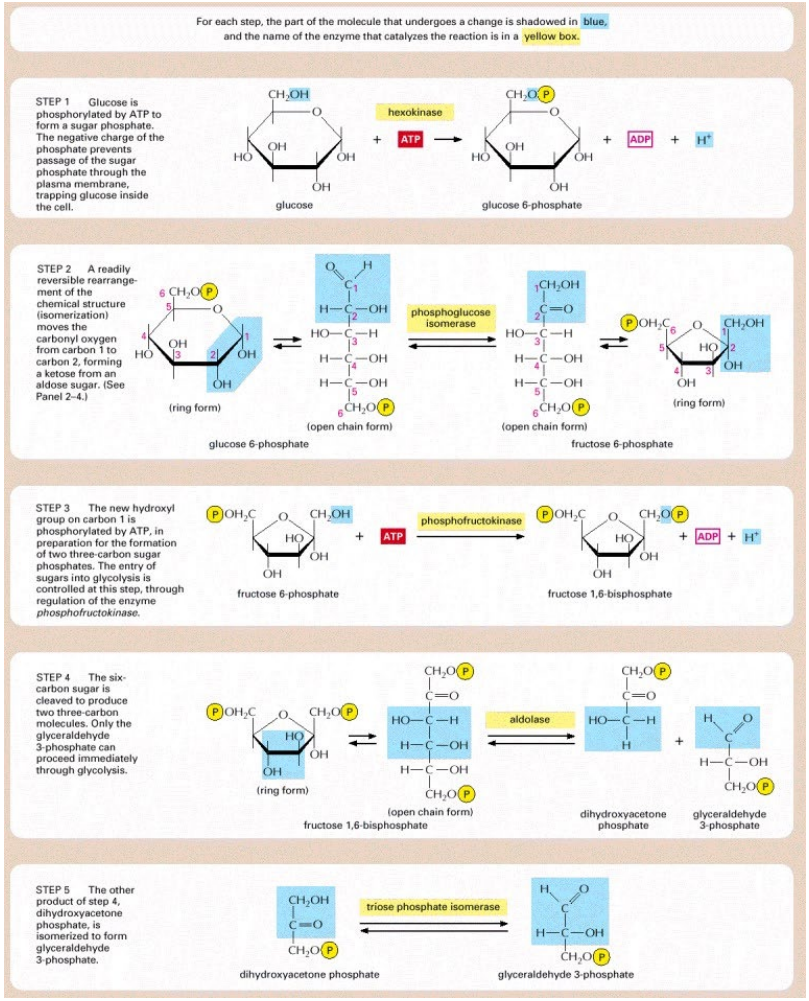
Ένα περίγραμμα της γλυκόλυσης

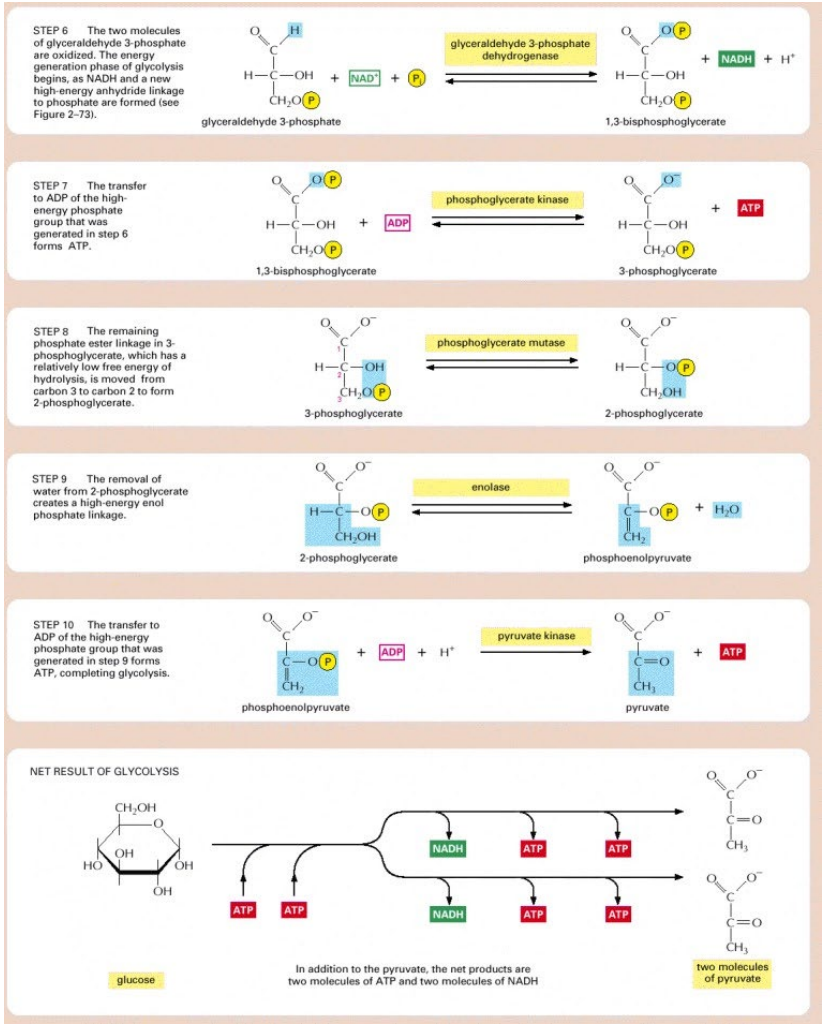
Καθένα από τα 10 βήματα που παρουσιάζονται καταλύεται από διαφορετικό ένζυμο. Σημειώστε ότι το βήμα 4 διασπά ένα σάκχαρο έξι άνθρακα σε δύο σάκχαρα τριών άνθρακα, έτσι ώστε ο αριθμός των μορίων σε κάθε στάδιο μετά από αυτό να διπλασιάζεται. Όπως υποδεικνύεται, το βήμα 6 ξεκινά τη φάση παραγωγής ενέργειας της γλυκόλυσης, η οποία προκαλεί την καθαρή σύνθεση των μορίων ATP και NADH .

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Λεπτομέρειες για τα 10 Βήματα της Γλυκόλυσης



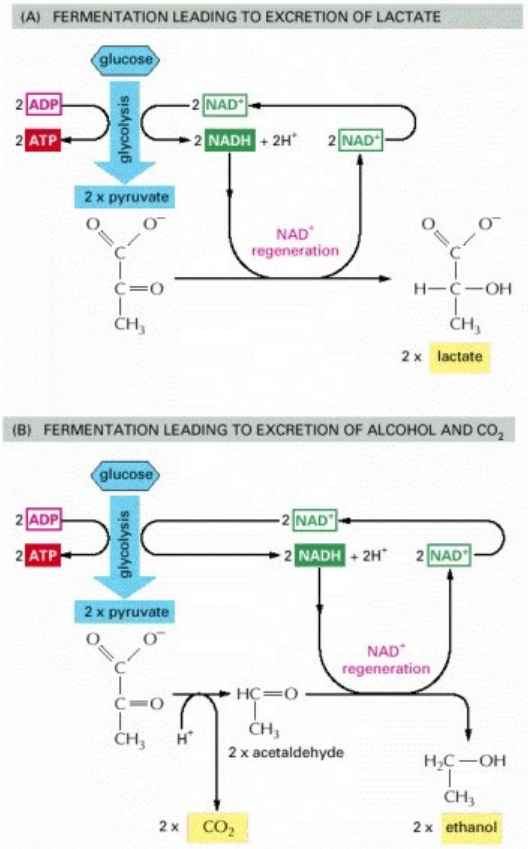


ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

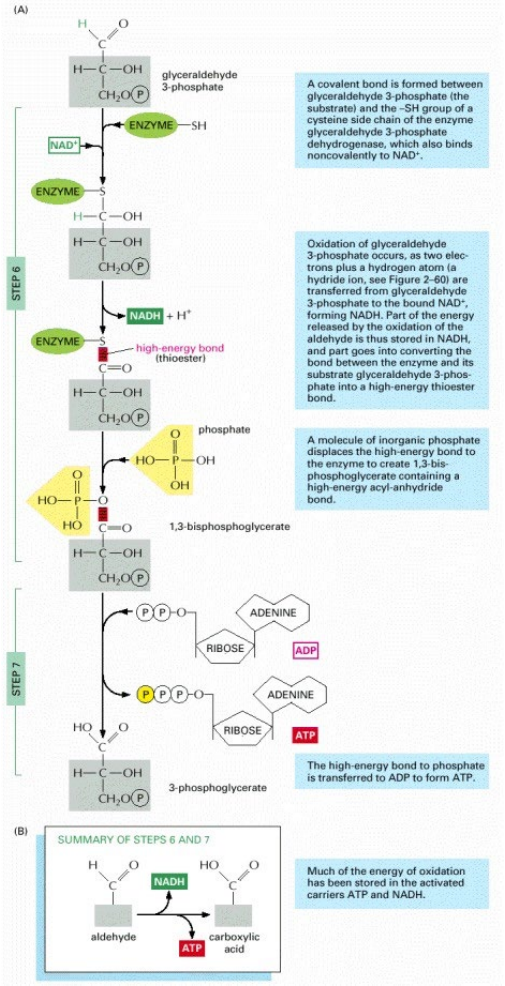
Οι ζυμώσεις επιτρέπουν την παραγωγή του ATP απουσία οξυγόνου



Δύο μονοπάτια για την αναερόβια διάσπαση του πυροσταφυλικού

(A) Όταν υπάρχει ανεπαρκές οξυγόνο, για παράδειγμα, σε ένα μυϊκό κύτταρο που υφίσταται έντονη συστολή, το πυροσταφυλικό που παράγεται από τη γλυκόλυση μετατρέπεται σε γαλακτικό όπως φαίνεται. Αυτή η αντίδραση αναγεννά το NAD+ που καταναλώνεται στο βήμα 6 της γλυκόλυσης, αλλά ολόκληρη η οδός αποδίδει συνολικά πολύ λιγότερη ενέργεια από την πλήρη οξείδωση. (B) Σε ορισμένους οργανισμούς που μπορούν να αναπτυχθούν αναερόβια, όπως οι ζυμομύκητες, το πυροσταφυλικό μετατρέπεται μέσω της ακεταλδεΰδης σε διοξείδιο του άνθρακα και αιθανόλη. Και πάλι, αυτή η οδός αναγεννά το NAD+ από το NADH, όπως απαιτείται για να μπορέσει να συνεχιστεί η γλυκόλυση. Και το (A) και το (B) είναι παραδείγματα ζυμώσεων.

Η γλυκόλυση απεικονίζει πώς τα ένζυμα συνδέουν την οξείδωση με την αποθήκευση ενέργειας

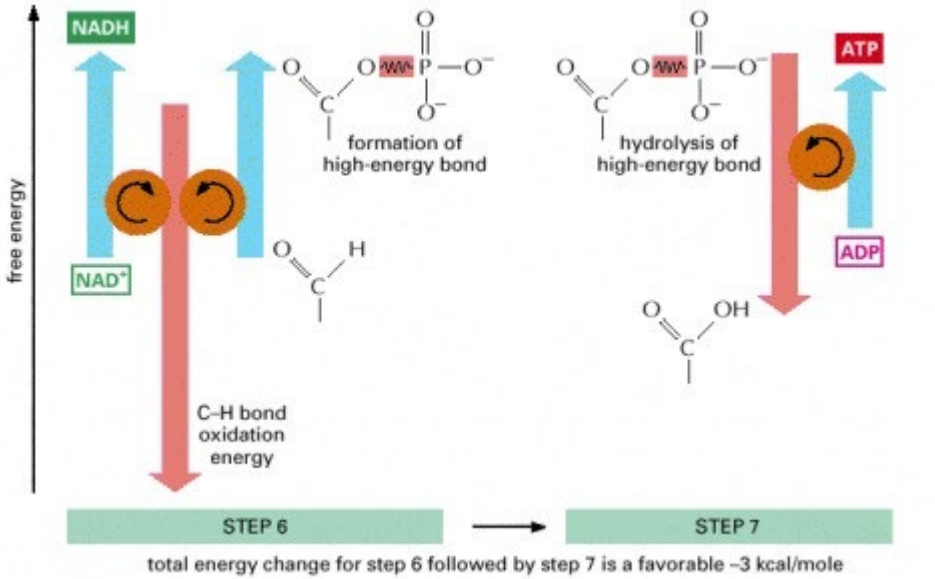


Αποθήκευση ενέργειας στα βήματα 6 και 7 της γλυκόλυσης

Σε αυτά τα στάδια η οξείδωση μιας αλδεΰδης σε ένα καρβοξυλικό οξύ συνδέεται με το σχηματισμό ATP και NADH. (A) Το στάδιο 6 ξεκινά με το σχηματισμό ενός ομοιοπολικού δεσμού μεταξύ του υποστρώματος (3-φωσφορική γλυκεραλδεΰδη) και μιας ομάδας -SH που εκτίθεται στην επιφάνεια του ενζύμου (αφυδρογονάση 3-φωσφορικής γλυκεραλδεΰδης). Στη συνέχεια, το ένζυμο καταλύει τη μεταφορά υδρογόνου (ως ιόν υδριδίου - ένα πρωτόνιο συν δύο ηλεκτρόνια) από τη δεσμευμένη 3-φωσφορική γλυκεραλδεΰδη σε ένα μόριο NAD+. Μέρος της ενέργειας που απελευθερώνεται σε αυτή την οξείδωση χρησιμοποιείται για να σχηματίσει ένα μόριο NADH και μέρος χρησιμοποιείται για τη μετατροπή της αρχικής σύνδεσης μεταξύ του ενζύμου και του υποστρώματος του σε έναν δεσμό θειοεστέρα υψηλής ενέργειας (που φαίνεται με κόκκινο). Στη συνέχεια, ένα μόριο ανόργανου φωσφορικού άλατος εκτοπίζει αυτόν τον δεσμό υψηλής ενέργειας στο ένζυμο, δημιουργώντας έναν δεσμό σακχάρου-φωσφορικού υψηλής ενέργειας (κόκκινο). Σε αυτό το σημείο το ένζυμο όχι μόνο έχει αποθηκεύσει ενέργεια στο NADH, αλλά συνέδεσε επίσης την ενεργειακά ευνοϊκή οξείδωση μιας αλδεΰδης με τον ενεργειακά δυσμενή σχηματισμό ενός φωσφορικού δεσμού υψηλής ενέργειας. Η δεύτερη αντίδραση προκλήθηκε από την πρώτη, ενεργώντας έτσι όπως ο συζεύκτης "τροχού κουπί" στο Σχήμα 2-56.

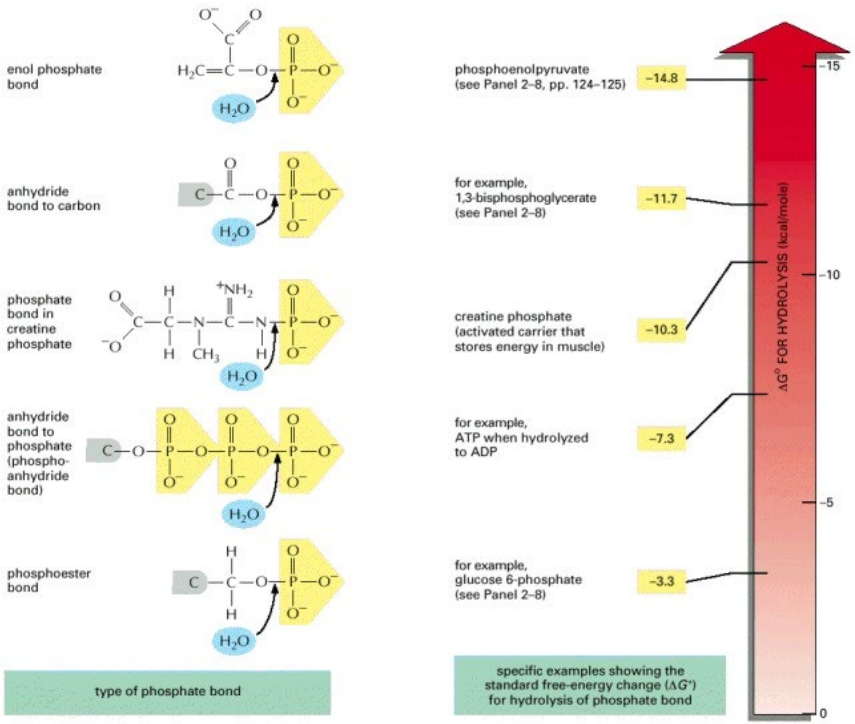
Στο στάδιο αντίδρασης 7, το υψηλής ενέργειας σάκχαρο-φωσφορικό ενδιάμεσο που μόλις παρασκευάστηκε, 1,3-διφωσφογλυκερικό, συνδέεται με ένα δεύτερο ένζυμο, τη φωσφογλυκερική κινάση. Το δραστικό φωσφορικό μεταφέρεται στο ADP, σχηματίζοντας ένα μόριο ATP και αφήνοντας μια ελεύθερη ομάδα καρβοξυλικού οξέος στο οξειδωμένο σάκχαρο. (B) Σύνοψη της συνολικής χημικής μεταβολής που παράγεται από τις αντιδράσεις 6 και 7.

Co-funded by the Erasmus+ Programme of the European Union



Σχηματική άποψη των συζευγμένων αντιδράσεων που σχηματίζουν NADH και ATP στα βήματα 6 και 7 της γλυκόλυσης

Η ενέργεια οξείδωσης του δεσμού CH οδηγεί τον σχηματισμό τόσο του NADH όσο και ενός φωσφορικού δεσμού υψηλής ενέργειας. Η θραύση του δεσμού υψηλής ενέργειας οδηγεί στη συνέχεια στο σχηματισμό ATP.



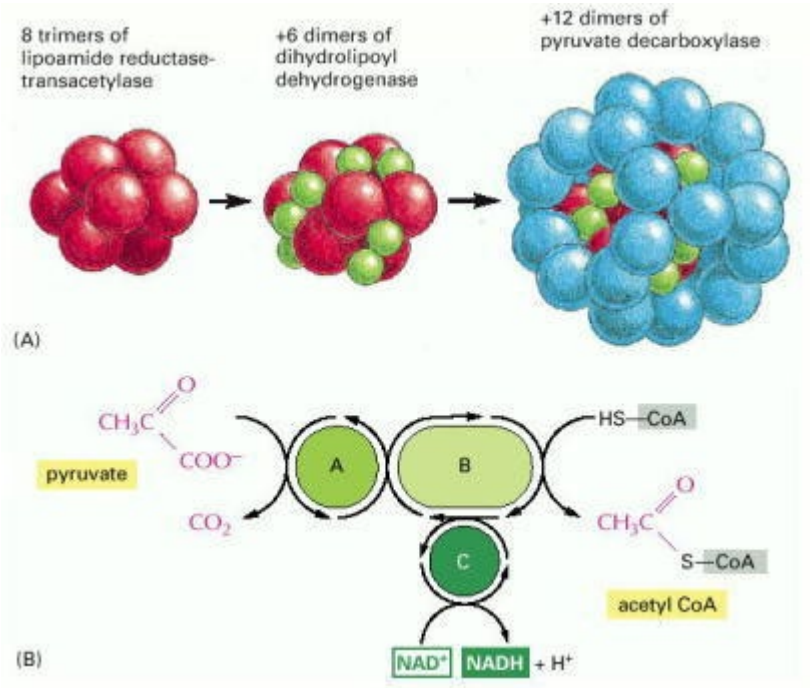
Μερικές ενέργειες φωσφορικών δεσμών

Η μεταφορά μιας φωσφορικής ομάδας από οποιοδήποτε μόριο 1 σε οποιοδήποτε μόριο 2 είναι ενεργειακά ευνοϊκή εάν η τυπική αλλαγή ελεύθερης ενέργειας (ΔG°) για την υδρόλυση του φωσφορικού δεσμού στο μόριο 1 είναι πιο αρνητική από αυτή για την υδρόλυση του φωσφορικού δεσμού στο μόριο 2. Έτσι, για παράδειγμα, μια φωσφορική ομάδα μεταφέρεται εύκολα από το 1,3-διφωσφογλυκερικό στο ADP, σχηματίζοντας ATP. Σημειώστε ότι η αντίδραση υδρόλυσης μπορεί να θεωρηθεί ως η μεταφορά της φωσφορικής ομάδας στο νερό.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

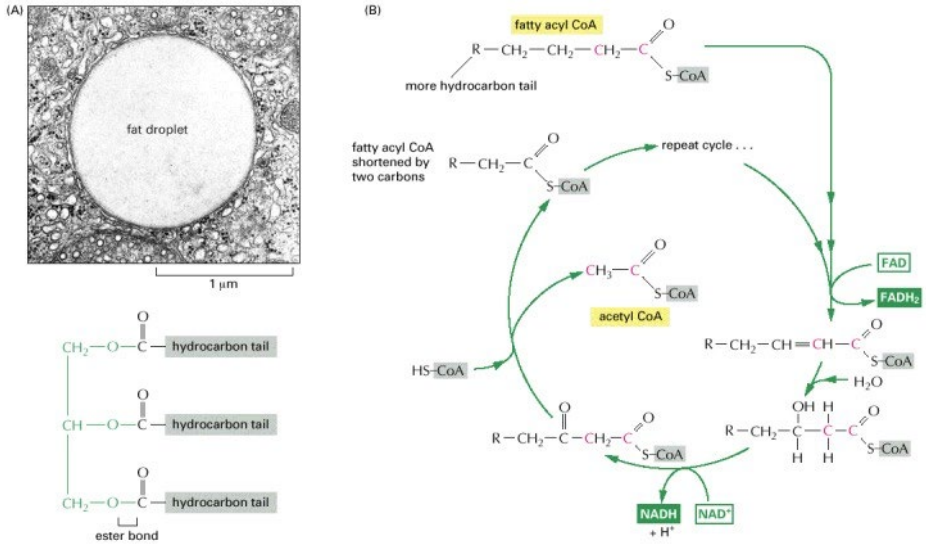
Τα σάκχαρα και τα λίπη διασπώνται σε ακετυλικό CoA στα μιτοχόνδρια



Η οξείδωση του πυροσταφυλικού σε ακετυλ CoA και CO2

(A) Η δομή του συμπλέγματος πυροσταφυλικής αφυδρογονάσης, το οποίο περιέχει 60 πολυπεπτιδικές αλυσίδες. Αυτό είναι ένα παράδειγμα ενός μεγάλου πολυενζυμικού συμπλέγματος στο οποίο τα ενδιάμεσα της αντίδρασης περνούν απευθείας από το ένα ένζυμο στο άλλο. Στα ευκαρυωτικά κύτταρα βρίσκεται στο μιτοχόνδριο. (B) Οι αντιδράσεις που διεξάγονται από το σύμπλοκο πυροσταφυλικής αφυδρογονάσης. Το σύμπλοκο μετατρέπει το πυροσταφυλικό σε ακετύλιο CoA στη μιτοχονδριακή μήτρα. Το NADH παράγεται επίσης σε αυτή την αντίδραση. Τα A, B και C είναι τα τρία ένζυμα πυροσταφυλική αποκαρβοξυλάση, αναγωγάση λιποαμιδίου- τρανσακετυλάση και διυδρολιποϋλ αφυδρογονάση, αντίστοιχα. Αυτά τα ένζυμα απεικονίζονται στο (A). οι δραστηριότητές τους συνδέονται όπως φαίνεται.

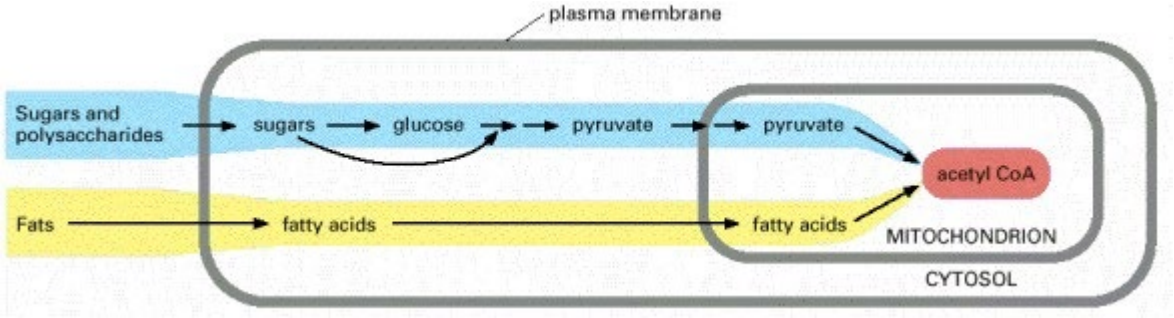
Co-funded by the Erasmus+ Programme of the European Union



Η οξείδωση των λιπαρών οξέων σε ακετυλ-CoA

(A) Ηλεκτρονική μικρογραφία μιας σταγόνας λιπιδίου στο κυτταρόπλασμα (πάνω), και η δομή των λιπών (κάτω). Τα λίπη είναι τριακυλογλυκερίνες. Το τμήμα γλυκερίνης, με το οποίο συνδέονται τρία λιπαρά οξέα μέσω εστερικών δεσμών, εμφανίζεται εδώ με πράσινο χρώμα. Τα λίπη είναι αδιάλυτα στο νερό και σχηματίζουν μεγάλα σταγονίδια λιπιδίων στα εξειδικευμένα λιποκύτταρα (που ονομάζονται λιποκύτταρα) στα οποία αποθηκεύονται. (B) Ο κύκλος οξείδωσης των λιπαρών οξέων. Ο κύκλος καταλύεται από μια σειρά τεσσάρων ενζύμων στο μιτοχόνδριο. Κάθε στροφή του κύκλου συντομεύει την αλυσίδα των λιπαρών οξέων κατά δύο άνθρακες (που φαίνεται με κόκκινο) και δημιουργεί ένα μόριο ακετυλ CoA και ένα μόριο NADH και FADH2 το καθένα.

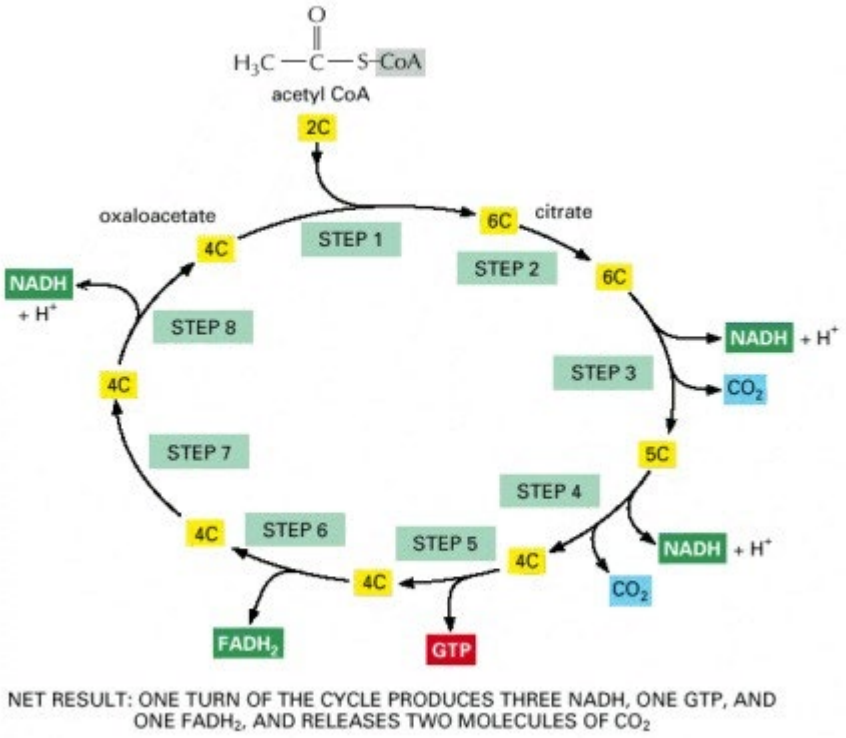
Co-funded by the Erasmus+ Programme of the European Union



Μονοπάτια για την παραγωγή ακετυλ-CoA από σάκχαρα και λίπη

Το μιτοχόνδριο στα ευκαρυωτικά κύτταρα είναι το μέρος όπου παράγεται το ακετυλικό CoA και από τους δύο τύπους κύριων μορίων τροφής. Είναι επομένως το μέρος όπου συμβαίνουν οι περισσότερες από τις αντιδράσεις οξείδωσης του κυττάρου και όπου παράγεται το μεγαλύτερο μέρος του ATP του.

Ο κύκλος του κιτρικού οξέος δημιουργεί NADH οξειδώνοντας ομάδες ακετυλίου σε CO2



NET RESULT: ONE TURN OF THE CYCLE PRODUCES THREE NADH, ONE GTP, AND ONE $FADH_2$, AND RELEASES TWO MOLECULES OF CO_2

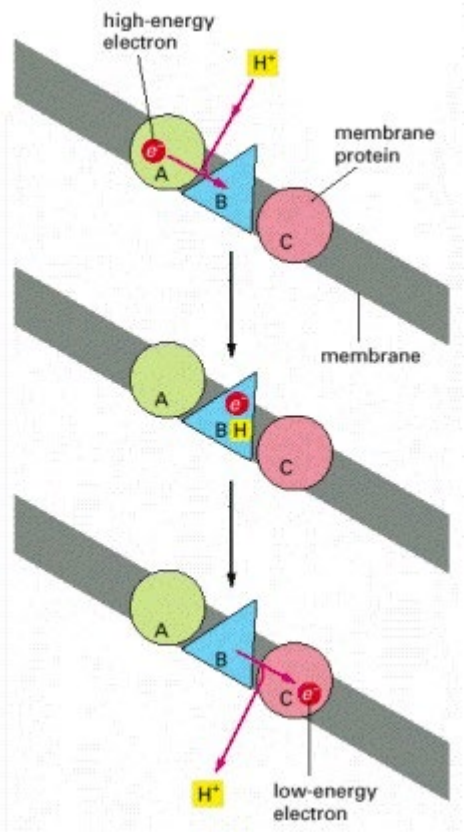
Απλή επισκόπηση του κύκλου του κιτρικού οξέος

Η αντίδραση του ακετυλ CoA με το οξαλοξικό ξεκινά τον κύκλο με την παραγωγή κιτρικού άλατος (κιτρικό οξύ). Σε κάθε στροφή του κύκλου παράγονται δύο μόρια CO2 ως απόβλητα, συν τρία μόρια NADH, ένα μόριο GTP και ένα μόριο FADH2. Ο αριθμός των ατόμων άνθρακα σε κάθε ενδιάμεσο εμφανίζεται σε ένα κίτρινο πλαίσιο.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

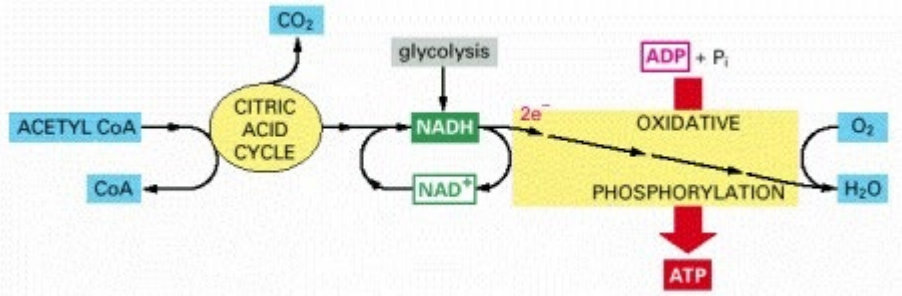
Η μεταφορά ηλεκτρονίων οδηγεί τη σύνθεση της πλειονότητας του ATP στα περισσότερα κύτταρα



Η δημιουργία μιας βαθμίδωσης H+ σε μια μεμβράνη από αντιδράσεις μεταφοράς ηλεκτρονίων

Ένα ηλεκτρόνιο υψηλής ενέργειας (που προέρχεται, για παράδειγμα, από την οξείδωση ενός μεταβολίτη) διέρχεται διαδοχικά από τους φορείς Α, Β και Γ σε μια χαμηλότερη ενεργειακή κατάσταση. Σε αυτό το διάγραμμα ο φορέας Β είναι διατεταγμένος στη μεμβράνη με τέτοιο τρόπο ώστε να παίρνει το H+ από τη μια πλευρά και να το απελευθερώνει στην άλλη καθώς περνά το ηλεκτρόνιο. Το αποτέλεσμα είναι μια κλίση H+. Αυτή η κλίση αντιπροσωπεύει μια μορφή αποθηκευμένης ενέργειας που αξιοποιείται από άλλες μεμβρανικές πρωτεΐνες για να οδηγήσει τον σχηματισμό του ATP.

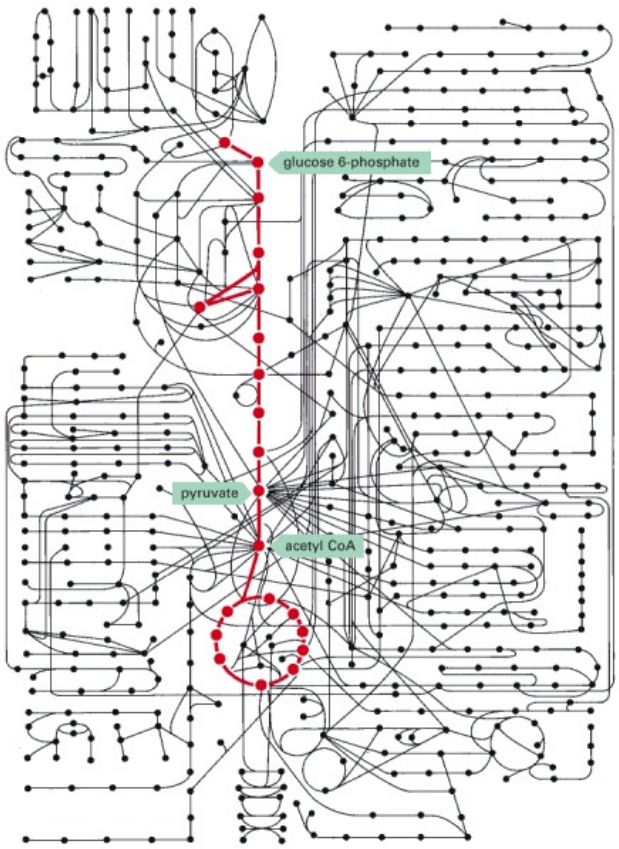
Co-funded by the Erasmus+ Programme of the European Union



Τα τελικά στάδια της οξείδωσης των μορίων των τροφίμων

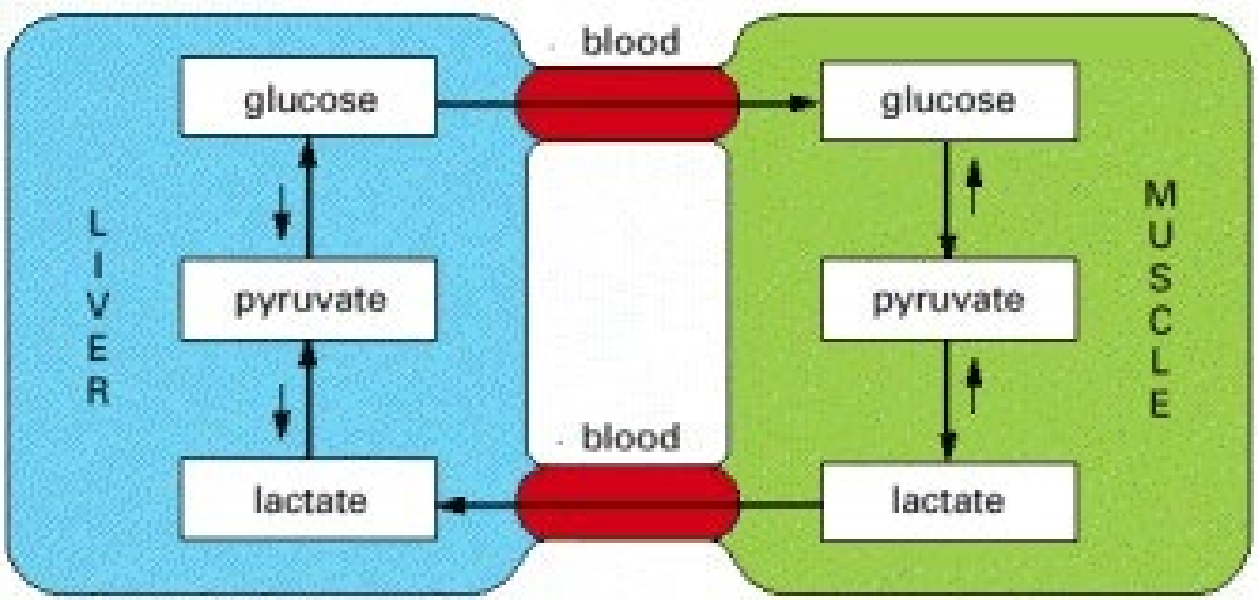
Τα μόρια του NADH και του FADH2 (δεν φαίνεται το FADH2) παράγονται από τον κύκλο του κιτρικού οξέος. Αυτοί οι ενεργοποιημένοι φορείς δωρίζουν ηλεκτρόνια υψηλής ενέργειας που χρησιμοποιούνται τελικά για τη μείωση του αερίου οξυγόνου στο νερό. Ένα μεγάλο μέρος της ενέργειας που απελευθερώνεται κατά τη μεταφορά αυτών των ηλεκτρονίων κατά μήκος μιας αλυσίδας μεταφοράς ηλεκτρονίων στη μιτοχονδριακή εσωτερική μεμβράνη (ή στην πλασματική μεμβράνη των βακτηρίων) αξιοποιείται για να οδηγήσει τη σύνθεση του ATP: εξ ου και η ονομασία οξειδωτική φωσφορυλίωση.

Ο μεταβολισμός είναι οργανωμένος και ρυθμισμένος



Η γλυκόλυση και ο κύκλος του κιτρικού οξέος βρίσκονται στο κέντρο του μεταβολισμού

Περίπου 500 μεταβολικές αντιδράσεις ενός τυπικού κυττάρου παρουσιάζονται σχηματικά με τις αντιδράσεις της γλυκόλυσης και τον κύκλο του κιτρικού οξέος με κόκκινο χρώμα. Άλλες αντιδράσεις είτε οδηγούν σε αυτές τις δύο κεντρικές οδούς - παραδίδοντας μικρά μόρια που καταβολίζονται με παραγωγή ενέργειας - είτε οδηγούν προς τα έξω και έτσι παρέχουν ενώσεις άνθρακα για το σκοπό της βιοσύνθεσης



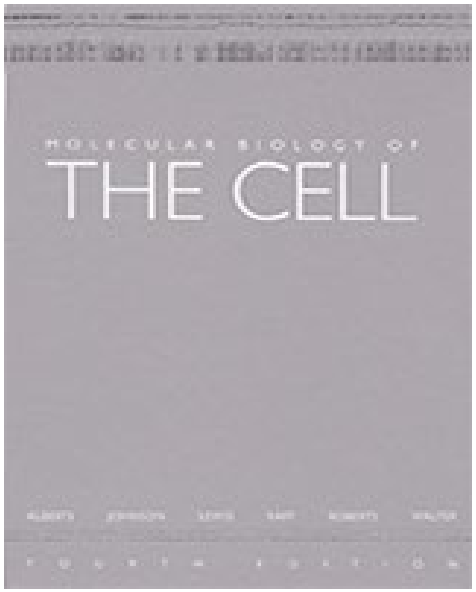
Σχηματική άποψη της μεταβολικής συνεργασίας μεταξύ ηπατικών και μυϊκών κυττάρων

Το κύριο καύσιμο των μυϊκών κυττάρων που συστέλλονται ενεργά είναι η γλυκόζη, μεγάλο μέρος της οποίας παρέχεται από τα ηπατικά κύτταρα. Το γαλακτικό οξύ, το τελικό προϊόν της αναερόβιας διάσπασης της γλυκόζης από τη γλυκόλυση στους μυς, μετατρέπεται ξανά σε γλυκόζη στο ήπαρ με τη διαδικασία της γλυκονεογένεσης.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Ποια είναι η πηγή της κυτταρικής ενέργειας;
- Εξηγήστε τις οδούς παραγωγής ενέργειας.
- Ποιος είναι ο μηχανισμός οξείδωσης της γλυκόζης και των λιπαρών οξέων σε διοξείδιο του άνθρακα;



Μοριακή Βιολογία του Κυττάρου. 4η έκδοση.
Alberts B, Johnson A, Lewis J, et al.
Νέα Υόρκη: Garland Science; 2002.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 2. Βασικές αρχές της Μοριακής Βιολογίας

Μάθημα 5. Σηματοδότηση της κυτταρικής επιφάνειας. Μόρια και υποδοχείς σήματος. Υποδοχείς που συνδέονται με την πρωτεΐνη G. Μονοπάτια σήματος που ελέγχουν τη γονιδιακή δραστηριότητα

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

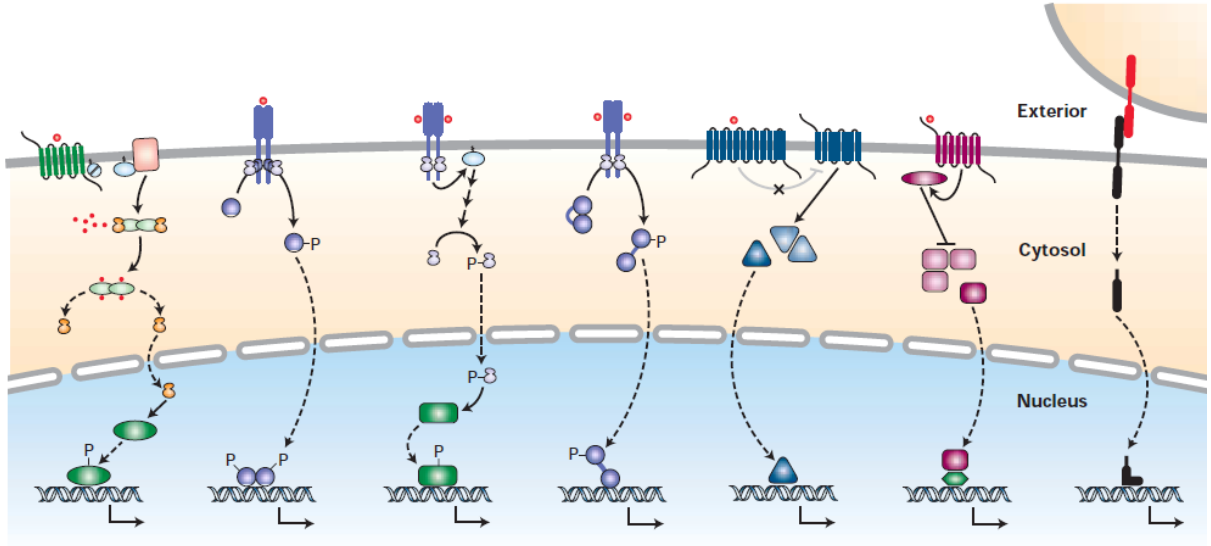
- Εισαγωγή
- Σηματοδότηση της επιφάνειας του κυττάρου
- Μόρια και υποδοχείς σήματος
- Υποδοχείς που συνδέονται με την πρωτεΐνη G
- Μονοπάτια σήματος που ελέγχουν τη γονιδιακή δραστηριότητα

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Σημαντικά στα φυτά και τα ζώα είναι εξωκυτταρικά μόρια σηματοδότησης που λειτουργούν μέσα σε έναν οργανισμό για να ελέγχουν τις μεταβολικές διεργασίες εντός των κυττάρων, την ανάπτυξη και τη διαφοροποίηση των ιστών, τη σύνθεση και έκκριση πρωτεϊνών και τη σύνθεση των ενδοκυτταρικών και εξωκυτταρικών υγρών. Τα γειτονικά κύτταρα συχνά επικοινωνούν με άμεση επαφή κυττάρου-κυττάρου. Τα εξωκυτταρικά μόρια σηματοδότησης συντίθενται και απελευθερώνονται από κύτταρα σηματοδότησης και παράγουν ειδική απόκριση μόνο σε κύτταρα στόχους που έχουν υποδοχείς για τα μόρια σηματοδότησης. Σε πολυκύτταρους οργανισμούς, μια τεράστια ποικιλία χημικών ουσιών, συμπεριλαμβανομένων μικρών μορίων (π.χ. αμινοξέων ή παραγώγων λιπιδίων, ακετυλοχολίνη), πεπτιδίων και πρωτεϊνών, χρησιμοποιούνται σε αυτόν τον τύπο επικοινωνίας κυττάρου σε κύτταρο. Ορισμένα μόρια σηματοδότησης, ειδικά υδρόφοβα μόρια όπως τα στεροειδή, τα ρετινοειδή και η θυροξίνη, διαχέονται αυθόρμητα μέσω της πλασματικής μεμβράνης και συνδέονται με ενδοκυτταρικούς υποδοχείς.

Co-funded by the Erasmus+ Programme of the European Union



G protein-coupled receptors	Cytokine receptors	Receptor tyrosine kinases	TGFβ receptors	Hedgehog (Hh) receptors	Wnt receptors	Notch receptor
Linked to a trimeric G protein that controls the activity of an effector protein (here adenylyl cyclase)	Associated with cytosolic JAK kinases	Cytosolic domain with tyrosine kinase activity	Cytosolic domain with serine/threonine kinase activity	Hh ligand tethered to membrane of signaling cell by cholesterol anchor	Palmitoylated Wnt ligand binds seven transmembrane protein receptor complex	Ligand, Delta, is a transmembrane protein on signaling cell
Activate cytosolic or nuclear transcription factors via several pathways (here one involving protein kinase A)	Activate cytosolic STAT transcription factors by phosphorylation	Activate cytosolic kinases (here MAP kinase) that translocate to the nucleus and activate nuclear transcription factors by phosphorylation	Activate Smad transcription factors in the cytosol by phosphorylation	Control processing of transcription factor by proteolysis; Hh binding causes release from cytosolic complex	Release an activated transcription factor from a multiprotein complex in the cytosol	Cytosolic domain of Notch released by proteolysis acts in association with nuclear transcription factors

Επισκόπηση επτά βασικών κατηγοριών υποδοχέων κυτταρικής επιφάνειας

Σηματοδότηση της επιφάνειας του κυττάρου

Το μόριο σηματοδότησης δρα ως συνδετήρας, ο οποίος δεσμεύεται σε μια δομικά συμπληρωματική θέση στην εξωκυτταρική ή στην περιοχή της μεμβράνης του υποδοχέα. Η δέσμευση ενός προσδέματος στον υποδοχέα του προκαλεί μια διαμορφωτική αλλαγή στην κυτοσολική περιοχή ή περιοχές του υποδοχέα που τελικά επάγει ειδικές κυτταρικές αποκρίσεις. Η συνολική διαδικασία μετατροπής σημάτων σε κυτταρικές αποκρίσεις, καθώς και τα επιμέρους βήματα αυτής της διαδικασίας, ονομάζεται μεταγωγή σήματος. Όπως θα δούμε, οι οδοί μεταγωγής σήματος μπορεί να περιλαμβάνουν σχετικά λίγα ή πολλά στοιχεία.

Co-funded by the Erasmus+ Programme of the European Union

Μόρια σηματοδότησης και υποδοχείς κυτταρικής επιφάνειας

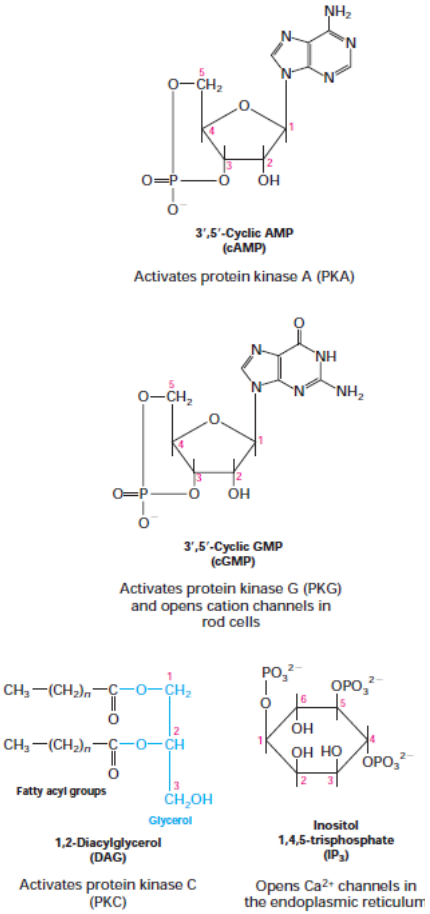
Η επικοινωνία με εξωκυτταρικά σήματα συνήθως περιλαμβάνει τα ακόλουθα βήματα: (1) σύνθεση, (2) απελευθέρωση του μορίου σηματοδότησης από το κύτταρο σηματοδότησης, (3) μεταφορά του σήματος στην κυψέλη στόχο, (4) δέσμευση του σήματος από μια συγκεκριμένη πρωτεΐνη υποδοχέα που οδηγεί στην ενεργοποίησή του, (5) έναρξη μιας ή περισσότερων ενδοκυτταρικών οδών μεταγωγής σήματος από τον ενεργοποιημένο υποδοχέα, (6) συγκεκριμένες αλλαγές στην κυτταρική λειτουργία, μεταβολισμό ή ανάπτυξη, και (7) αφαίρεση του σήματος, το οποίο συχνά τερματίζει την κυτταρική απόκριση.

Co-funded by the Erasmus+ Programme of the European Union

Η συντριπτική πλειονότητα των υποδοχέων ενεργοποιείται με τη δέσμευση εκκρινόμενων ή δεσμευμένων στη μεμβράνη μορίων (π.χ. ορμόνες, αυξητικούς παράγοντες, νευροδιαβιβαστές και φερομόνες).

- Τα μόρια σηματοδότησης στα ζώα λειτουργούν σε διάφορες αποστάσεις
- Οι υποδοχείς ενεργοποιούν έναν περιορισμένο αριθμό μονοπατιών σηματοδότησης

Ενδοκυτταρική μεταγωγή σήματος



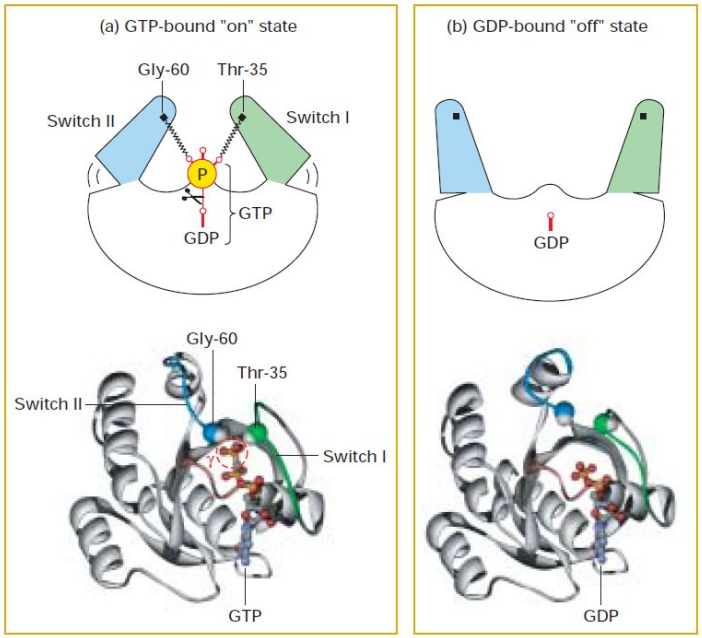
Οι δεύτεροι αγγελιοφόροι μεταφέρουν σήματα από πολλούς υποδοχείς

Η δέσμευση προσδεμάτων («πρώτοι αγγελιοφόροι») σε πολλούς υποδοχείς κυτταρικής επιφάνειας οδηγεί σε βραχύβια αύξηση (ή μείωση) της συγκέντρωσης ορισμένων μικρού μοριακού βάρους ενδοκυτταρικών μορίων σηματοδότησης που ονομάζονται δεύτεροι αγγελιοφόροι. Αυτά τα μόρια περιλαμβάνουν 3',5'-κυκλικό AMP (cAMP), 3',5'-κυκλικό GMP (cGMP), 1,2-διακυλογλυκερόλη (DAG) και 1,4,5-τριφωσφορική ινοσιτόλη (IP3), των οποίων οι δομές φαίνονται στο σχήμα. Άλλοι σημαντικοί δεύτεροι αγγελιοφόροι είναι το Ca^{2-} και διάφορα φωσφολιπίδια ινοσιτόλης, που ονομάζονται επίσης φωσφοϊνοσιτίδια , τα οποία είναι ενσωματωμένα σε κυτταρικές μεμβράνες . Η κύρια άμεση επίδραση ή επιδράσεις κάθε ένωσης υποδεικνύονται κάτω από τον δομικό τύπο της. Το ιόν ασβεστίου (Ca^{2-}) και πολλά φωσφοϊνοσιτίδια που συνδέονται με τη μεμβράνη δρουν επίσης ως δεύτεροι αγγελιοφόροι.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

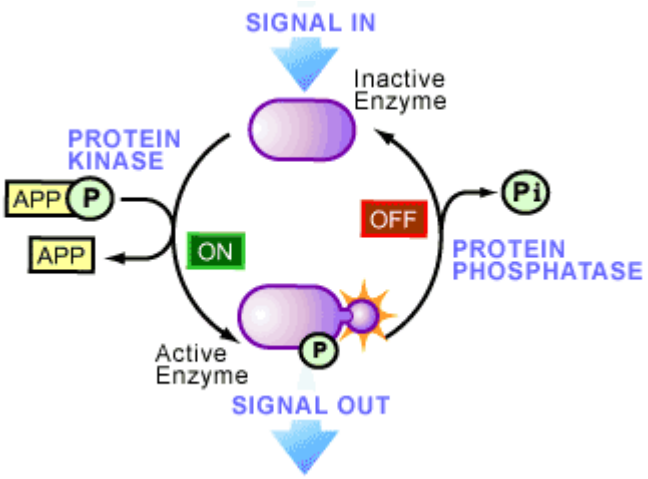
Ενδοκυτταρική μεταγωγή σήματος



Πολλές διατηρημένες ενδοκυτταρικές πρωτεΐνες λειτουργούν στη μεταγωγή σήματος

Οι πρωτεΐνες διακόπτη GTPase είναι πρωτεΐνες που δεσμεύουν νουκλεοτίδια γουανίνης και ενεργοποιούνται όταν συνδέονται με GTP και απενεργοποιούνται όταν συνδέονται με το GDP. Η επαγόμενη από το σήμα μετατροπή της ανενεργής σε ενεργή κατάσταση μεσολαβείται από έναν παράγοντα ανταλλαγής νουκλεοτιδίων γουανίνης (GEF), ο οποίος προκαλεί απελευθέρωση του GDP από την πρωτεΐνη μεταγωγής. Η επακόλουθη δέσμευση της GTP, που ευνοείται από την υψηλή ενδοκυτταρική της συγκέντρωση, προκαλεί μια διαμορφωτική αλλαγή σε δύο τμήματα της πρωτεΐνης, που ονομάζονται διακόπτης I και διακόπτης II, επιτρέποντας στην πρωτεΐνη να συνδεθεί και να ενεργοποιήσει άλλες πρωτεΐνες σηματοδότησης κατάντη.

Ενδοκυτταρική μεταγωγή σήματος



Πολλές διατηρημένες ενδοκυτταρικές πρωτεΐνες λειτουργούν στη μεταγωγή σήματος

Πρωτεϊνικές κινάσες και φωσφατάσες Η ενεργοποίηση όλων των υποδοχέων της κυτταρικής επιφάνειας οδηγεί άμεσα ή έμμεσα σε αλλαγές στη φωσφορυλίωση της πρωτεΐνης μέσω της ενεργοποίησης πρωτεϊνικών κινασών ή πρωτεϊνικών φωσφατάσεων. Τα ζωικά κύτταρα περιέχουν δύο τύπους πρωτεϊνικών κινασών: αυτές που προσθέτουν φωσφορικό άλας στην ομάδα υδροξυλίου στα υπολείμματα τυροσίνης και εκείνες που προσθέτουν φωσφορικό στην ομάδα υδροξυλίου σε υπολείμματα σερίνης ή θρεονίνης (ή και των δύο). Οι φωσφατάσες, οι οποίες αφαιρούν ομάδες φωσφορικών αλάτων, μπορούν να δράσουν σε συνδυασμό με τις κινάσες για να ενεργοποιήσουν ή να απενεργοποιήσουν τη λειτουργία διαφόρων πρωτεϊνών

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

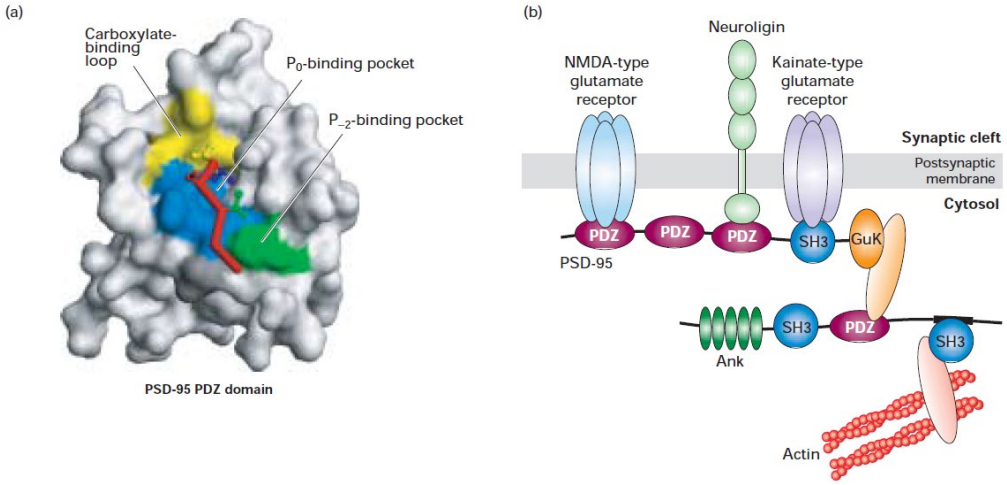
Co-funded by the Erasmus+ Programme of the European Union

Ενδοκυτταρική μεταγωγή σήματος

Ορισμένοι υποδοχείς και πρωτεΐνες μετάδοσης σήματος εντοπίζονται

Ομαδοποίηση πρωτεϊνών μεμβράνης που διαμεσολαβούνται από τομείς προσαρμογής

Συγκέντρωση πρωτεϊνών σε σχεδίες λιπιδίων



Co-funded by the Erasmus+ Programme of the European Union

Ενδοκυτταρική μεταγωγή σήματος

Οι κατάλληλες κυτταρικές αποκρίσεις εξαρτώνται από την αλληλεπίδραση και τη ρύθμιση των μονοπατιών σηματοδότησης

Η ικανότητα των κυττάρων να ανταποκρίνονται κατάλληλα στα εξωκυτταρικά σήματα εξαρτάται επίσης από τη ρύθμιση των ίδιων των οδών σηματοδότησης. Η ευαισθησία ενός κυττάρου σε ένα συγκεκριμένο μόριο σηματοδότησης μπορεί να μειωθεί με ενδοκυττάρωση των υποδοχέων του, μειώνοντας έτσι τον αριθμό στην επιφάνεια του κυττάρου ή τροποποιώντας τη δραστηριότητά τους έτσι ώστε οι υποδοχείς είτε να μην μπορούν να δεσμεύσουν ρίζα ή να σχηματίσουν ένα σύμπλεγμα υποδοχέα-συνδέτη που δεν προκαλεί τη φυσιολογική κυτταρική απόκριση. Αυτή η ρύθμιση της δραστηριότητας του υποδοχέα συχνά προκύπτει από φωσφορυλίωση του υποδοχέα, σύνδεση άλλων πρωτεϊνών σε αυτόν ή και τα δύο.

Co-funded by the Erasmus+ Programme of the European Union

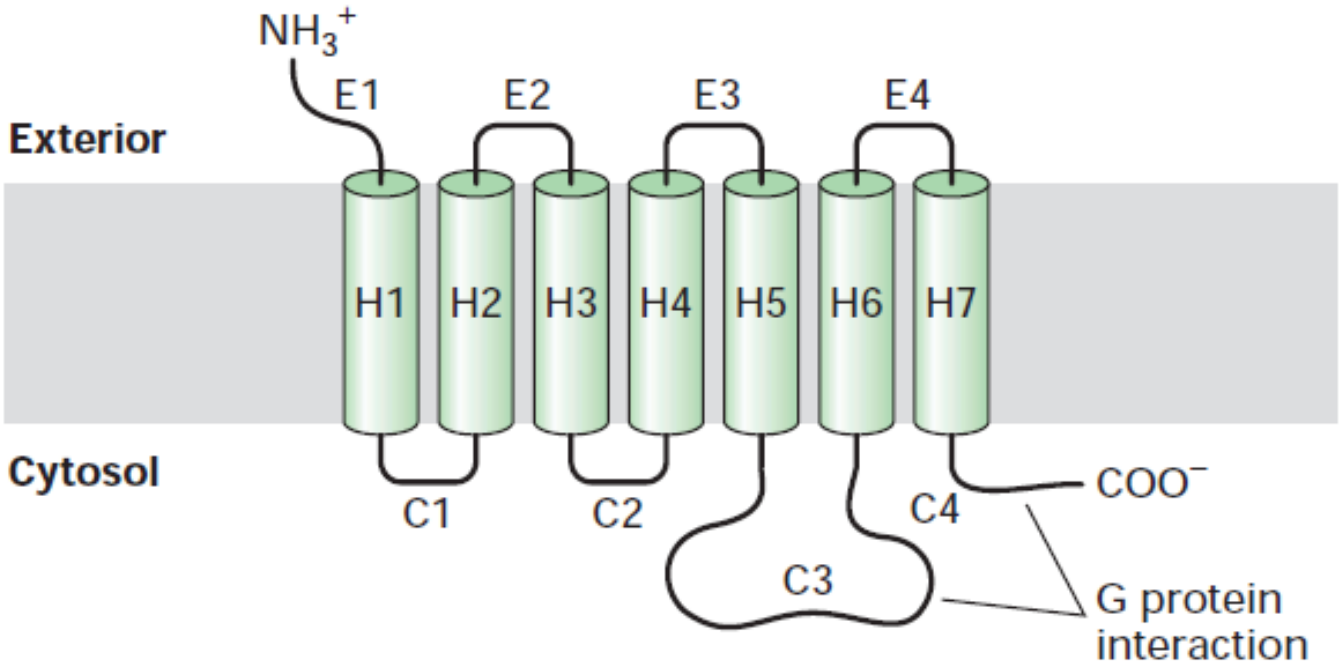
Υποδοχείς συζευγμένοι με πρωτεΐνη G που ενεργοποιούν ή αναστέλλουν την αδενυλυλοκυκλάση

G_α Class	Associated Effector	2nd Messenger	Receptor Examples
$G_{s\alpha}$	Adenylyl cyclase	cAMP (increased)	β-Adrenergic (epinephrine) receptor; receptors for glucagon, serotonin, vasopressin
$G_{i\alpha}$	Adenylyl cyclase K^+ channel ($G_{\beta\gamma}$ activates effector)	cAMP (decreased) Change in membrane potential	α_1-Adrenergic receptor Muscarinic acetylcholine receptor
$G_{olf\alpha}$	Adenylyl cyclase	cAMP (increased)	Odorant receptors in nose
$G_{q\alpha}$	Phospholipase C	IP_3, DAG (increased)	α_2-Adrenergic receptor
$G_{o\alpha}$	Phospholipase C	IP_3, DAG (increased)	Acetylcholine receptor in endothelial cells
$G_{t\alpha}$	cGMP phosphodiesterase	cGMP (decreased)	Rhodopsin (light receptor) in rod cells

*A given G_α subclass may be associated with more than one effector protein. To date, only one major $G_{s\alpha}$ has been identified, but multiple $G_{q\alpha}$ and $G_{i\alpha}$ proteins have been described. Effector proteins commonly are regulated by G_α but in some cases by $G_{\beta\gamma}$ or the combined action of G_α and $G_{\beta\gamma}$. IP_3 = inositol 1,4,5-trisphosphate; DAG = 1,2-diacylglycerol.

SOURCES: See L. Birnbaumer, 1992, *Cell* **71**:1069; Z. Farfel et al., 1999, *New Eng. J. Med.* **340**:1012; and K. Pierce et al., 2002, *Nature Rev. Mol. Cell Biol.* 3:639.

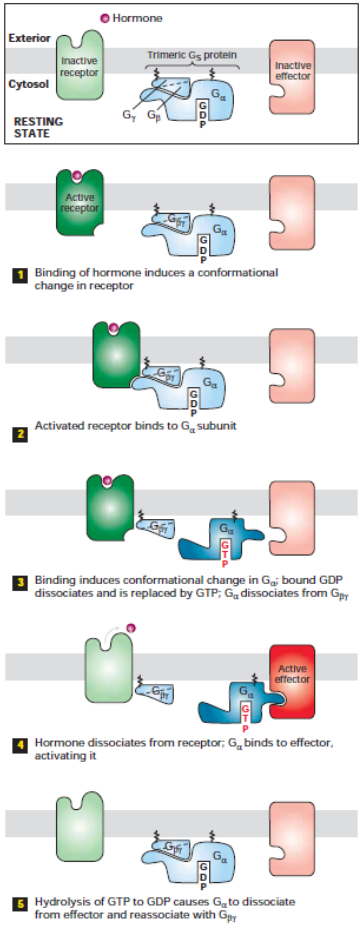
Co-funded by the Erasmus+ Programme of the European Union



Σχηματικό διάγραμμα της γενικής δομής των συζευγμένων με πρωτεΐνη G υποδοχέων.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Η υπομονάδα G α των πρωτεϊνών. G Κύκλοι μεταξύ ενεργών και ανενεργών μορφών

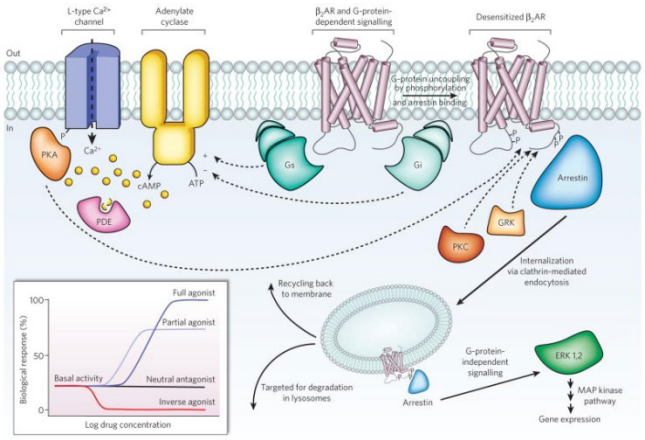


Οι υπομονάδες G α και G βγ των τριμερών πρωτεϊνών G συνδέονται με τη μεμβράνη με ομοιοπολικά συνδεδεμένα μόρια λιπιδίων (μαύρες γραμμές με τρεμούλες). Μετά τη δέσμευση συνδέτη, τη διάσταση της πρωτεΐνης G και την ανταλλαγή GDP με GTP (στάδια 1 – 3), η ελεύθερη G·GTP συνδέεται και ενεργοποιεί μια πρωτεΐνη τελεστή (βήμα 4). Η υδρόλυση του GTP τερματίζει τη σηματοδότηση και οδηγεί σε επανασυναρμολόγηση της τριμερούς μορφής, επιστρέφοντας το σύστημα σε κατάσταση ηρεμίας (βήμα 5). Η δέσμευση ενός άλλου μορίου συνδέτη προκαλεί επανάληψη του κύκλου. Σε ορισμένες οδούς, η πρωτεΐνη τελεστής ενεργοποιείται από την ελεύθερη υπομονάδα G βγ .

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

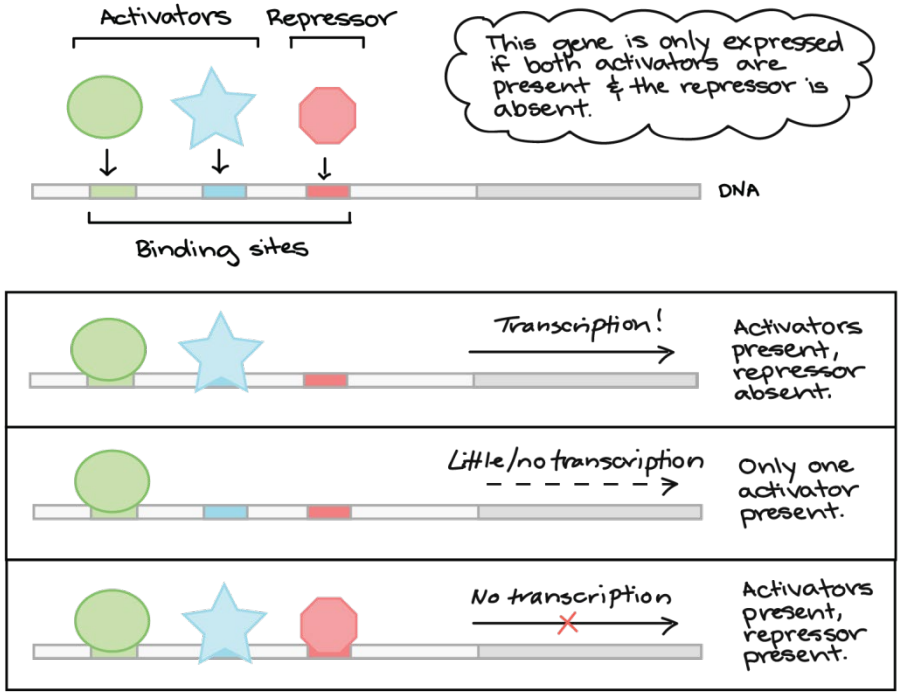
Μετάδοση σήματος σε υποδοχείς συζευγμένους με πρωτεΐνη G



Διαφορετικές οδοί σηματοδότησης που ρυθμίζονται από τον βήτα αδρενεργικό υποδοχέα τύπου 2 (β2AR). Το β2AR μπορεί να ενεργοποιήσει δύο πρωτεΐνες G, Gας και Ga i (μέρος των Gs και Gi ετεροτριμερή , αντίστοιχα), τα οποία ρυθμίζουν διαφορικά την αδενυλική κυκλάση. Η αδενυλική κυκλάση δημιουργεί κυκλικό AMP (cAMP), το οποίο ενεργοποιεί την πρωτεϊνική κινάση A (PKA), μια κινάση που ρυθμίζει τη δραστηριότητα αρκετών κυτταρικών πρωτεϊνών συμπεριλαμβανομένου του καναλιού Ca2+ τύπου L και του β2AR. Τα επίπεδα του δεύτερου αγγελιοφόρου cAMP ρυθμίζονται προς τα κάτω από ειδικές πρωτεΐνες φωσφοδιεστεράσης (PDEs). Η ενεργοποίηση του β2AR οδηγεί επίσης σε φωσφορυλίωση από μια κινάση υποδοχέα συζευγμένη με πρωτεΐνη G (GRK) και επακόλουθη σύζευξη με αρεστίνη . Η αρρεστίνη είναι μια σηματοδοτική και ρυθμιστική πρωτεΐνη που προάγει την ενεργοποίηση εξωκυτταρικών κινασών που ρυθμίζονται με σήμα (ERK), αποτρέπει την ενεργοποίηση των πρωτεϊνών G και προάγει την εσωτερίκευση του υποδοχέα μέσω κοιλοτήτων επικαλυμμένων με κλαθρίνη . PKC, πρωτεϊνική κινάση C. Το ένθετο δείχνει ταξινόμηση της αποτελεσματικότητας του συνδέτη για GPCRs. Πολλοί GPCR επιδεικνύουν βασική, ανεξάρτητη από αγωνιστή δραστηριότητα. Οι αντίστροφοι αγωνιστές αναστέλλουν αυτή τη δράση και οι ουδέτεροι ανταγωνιστές δεν έχουν κανένα αποτέλεσμα. Οι αγωνιστές και οι μερικοί αγωνιστές διεγείρουν βιολογικές αποκρίσεις πάνω από τη βασική δραστηριότητα. Η αποτελεσματικότητα δεν σχετίζεται άμεσα με τη συγγένεια. Για παράδειγμα, ένας μερικός αγωνιστής μπορεί να έχει υψηλότερη συγγένεια για ένα GPCR από έναν πλήρη αγωνιστή.

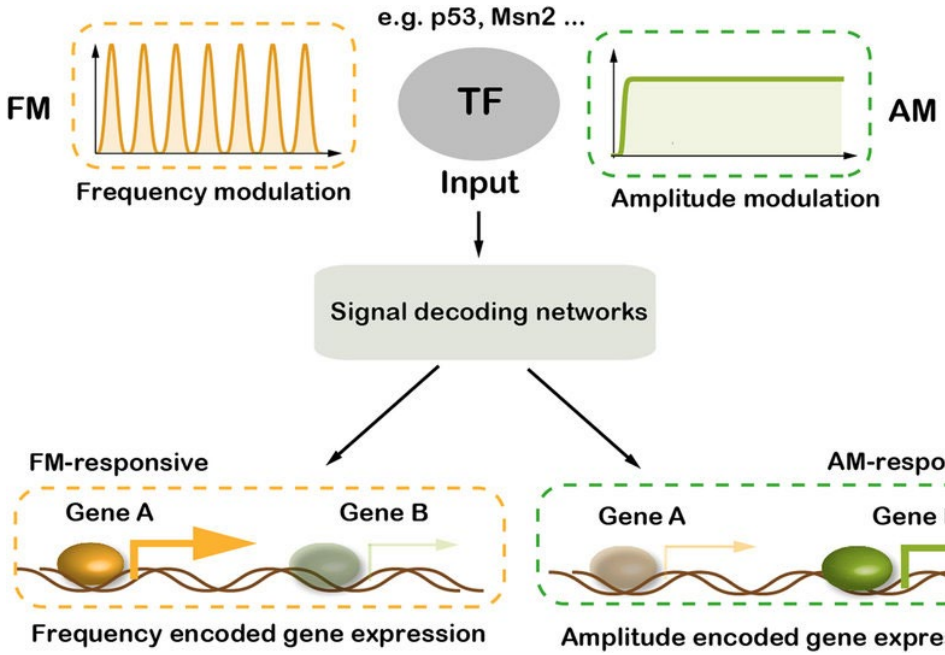
Co-funded by the Erasmus+ Programme of the European Union

Γενικοί και γονιδιακοί παράγοντες μεταγραφής



Τα γονίδια που κωδικοποιούν πρωτεΐνες μεταγράφονται από την RNA πολυμεράση II παρουσία μισής ντουζίνας γενικών μεταγραφικών παραγόντων. Αυτοί οι παράγοντες βοηθούν στην τοποθέτηση της πολυμεράσης κοντά στη θέση έναρξης και επιτρέπουν ένα χαμηλό επίπεδο βασικής μεταγραφής. Αυτοί οι παράγοντες αποτελούνται συχνά από πολλές υπομονάδες και συνολικά το σύμπλοκο έναρξης περιέχει περίπου 30 διαφορετικά πολυπεπτίδια. Συνήθως, κάθε γονίδιο ρυθμίζεται από έναν αριθμό μεταγραφικών παραγόντων που συνδέονται με αλληλουχίες προαγωγέα και ενισχυτή. Για γονίδια που εκφράζονται σε ένα μοναδικό ή μικρό αριθμό τύπων κυττάρων, ένας ή περισσότεροι από αυτούς τους μεταγραφικούς παράγοντες εκφράζονται σε έναν μόνο ή περιορισμένο αριθμό τύπων κυττάρων.

Διαμόρφωση της δραστηριότητας του παράγοντα μεταγραφής

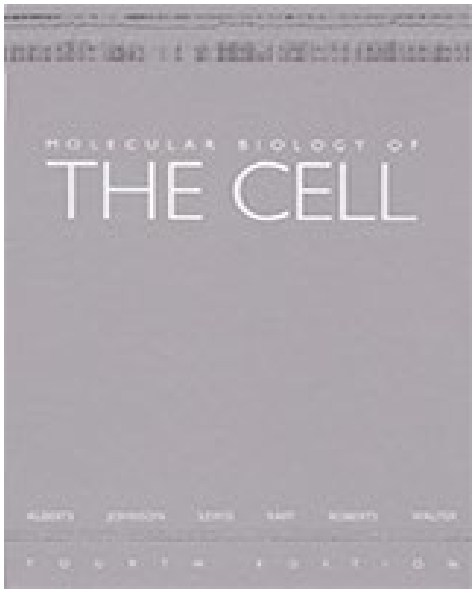


Τα κύτταρα εκτίθενται σε διαλυτά μικρά μόρια, σε πρωτεΐνες στην κυκλοφορία ή σε πρωτεΐνες (γλυκοπρωτεΐνες) στην επιφάνεια γειτονικών κυττάρων. Οι αλλαγές στη συγκέντρωση τέτοιων μορίων γίνονται αισθητές από κύτταρα που φιλοξενούν τους αντίστοιχους υποδοχείς. Θα συζητήσουμε ξεχωριστά τους υποδοχείς ορμονών και βιταμινών και τους υποδοχείς για αυξητικούς παράγοντες και κυτοκίνες. Διαλυτά μόρια όπως στεροειδείς ορμόνες, βιταμίνες κ.λπ., διαχέονται μέσω της μεμβράνης και συνδέονται με τους αντίστοιχους υποδοχείς τους.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε τον μηχανισμό της αλληλεπίδρασης συνδέτη-υποδοχέα.
- Εξηγήστε τον μηχανισμό μεταγωγής σήματος.
- Εξηγήστε τις οδούς σήματος που ελέγχουν τη γονιδιακή δραστηριότητα.



Μοριακή Βιολογία του Κυττάρου. 4η έκδοση.
Alberts B, Johnson A, Lewis J, et al.
Νέα Υόρκη: Garland Science; 2002.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 3. Πρωτεΐνες και ένζυμα

Μάθημα 1. Δομή και ιδιότητες των αμινοξέων

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

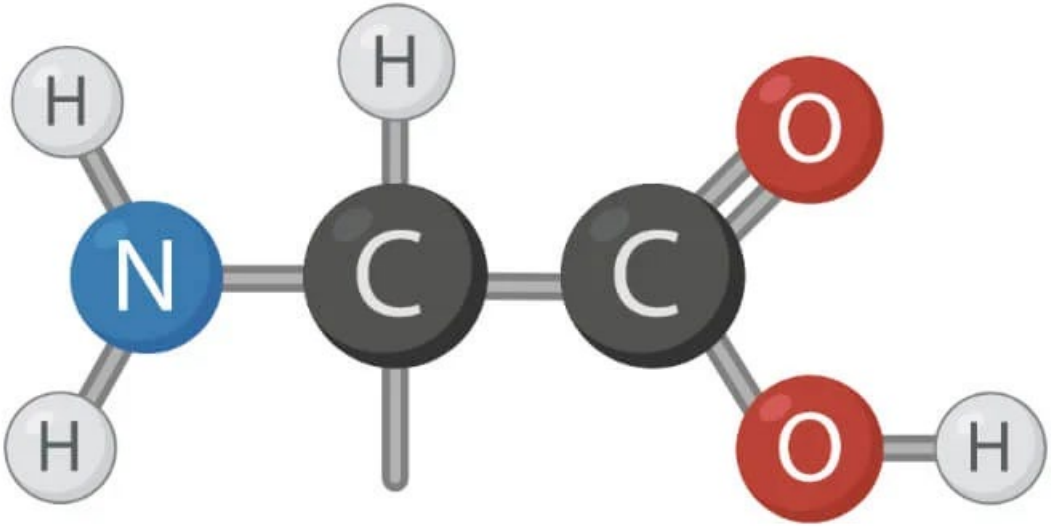
- Εισαγωγή
- Φυσικές ιδιότητες
- Χημικές ιδιότητες
- Δομή αμινοξέων
- Ταξινόμηση αμινοξέων
- Λειτουργίες αμινοξέων

Εισαγωγή

Τα αμινοξέα αποτελούν μια ομάδα ουδέτερων προϊόντων που διακρίνονται σαφώς από άλλες φυσικές ενώσεις χημικά, κυρίως λόγω των αμφολυτικών τους ιδιοτήτων και βιοχημικά, κυρίως λόγω του ρόλου τους ως πρωτεϊνικά συστατικά. Ένα αμινοξύ είναι ένα καρβοξυλικό οξύ που περιέχει μια αλειφατική πρωτοταγή αμινομάδα στη θέση α προς την καρβοξυλική ομάδα και με χαρακτηριστική στερεοχημεία. Οι πρωτεΐνες βιοσυντίθενται από 20 αμινοξέα σε ένα σύστημα που περιλαμβάνει αυστηρό γενετικό έλεγχο. Έτσι, τα αμινοξέα είναι η βασική μονάδα των πρωτεϊνών. Περισσότερα από 300 αμινοξέα βρίσκονται στη φύση, αλλά μόνο 20 αμινοξέα είναι τυπικά και υπάρχουν στην πρωτεΐνη επειδή κωδικοποιούνται από γονίδια. Άλλα αμινοξέα είναι τροποποιημένα αμινοξέα και ονομάζονται μη πρωτεϊνικά αμινοξέα. Μερικά είναι υπολείμματα τροποποιημένα μετά τη σύνθεση μιας πρωτεΐνης με μεταμεταφραστικές τροποποιήσεις. Άλλα είναι αμινοξέα που υπάρχουν σε ζωντανούς οργανισμούς αλλά όχι ως συστατικά πρωτεϊνών.

Co-funded by the Erasmus+ Programme of the European Union

Amino Acids- Properties, Structure, Classification, Functions



Φυσικές ιδιότητες

- Τα αμινοξέα είναι άχρωμα, κρυσταλλικά στερεά.
- Όλα τα αμινοξέα έχουν υψηλό σημείο τήξης μεγαλύτερο από 200°
- Διαλυτότητα: Είναι διαλυτά στο νερό, ελαφρώς διαλυτά στην αλκοόλη και διαλύονται με δυσκολία σε μεθανόλη, αιθανόλη και προπανόλη. Η ομάδα R αμινοξέων και το pH του διαλύτη παίζουν σημαντικό ρόλο στη διαλυτότητα.
- Κατά τη θέρμανση σε υψηλές θερμοκρασίες, αποσυντίθενται.
- Όλα τα αμινοξέα (εκτός από τη γλυκίνη) είναι οπτικά ενεργά.
- Σχηματισμός πεπτιδικού δεσμού: Τα αμινοξέα μπορούν να συνδεθούν με έναν πεπτιδικό δεσμό που περιλαμβάνει τις αμινο και καρβοξυλικές ομάδες τους. Ένας ομοιοπολικός δεσμός που σχηματίστηκε μεταξύ της άλφα-αμινομάδας ενός αμινοξέος και μιας άλφα-καρβοξυλικής ομάδας άλλης που σχηματίζει -CO-NH-σύνδεση. Οι πεπτιδικοί δεσμοί είναι επίπεδοι και εν μέρει ιοντικοί.

Χημικές ιδιότητες

- Διπολική ιδιότητα

Ένα αμφιτεριόν είναι ένα μόριο με λειτουργικές ομάδες, από τις οποίες τουλάχιστον μία έχει θετικό και μία αρνητικό ηλεκτρικό φορτίο. Το καθαρό φορτίο ολόκληρου του μορίου είναι μηδέν. Τα αμινοξέα είναι τα πιο γνωστά παραδείγματα αμφιτεριόντων. Περιέχουν μια ομάδα αμίνης (βασική) και μια καρβοξυλική ομάδα (όξινη). Η ομάδα -NH2 είναι η ισχυρότερη βάση, και έτσι παίρνει το H+ από την ομάδα -COOH για να αφήσει ένα αμφιτεριόν. Το (ουδέτερο) αμφιτεριόν είναι η συνήθης μορφή αμινοξέων που υπάρχουν στο διάλυμα.

- Αμφοτερική ιδιοκτησία

Τα αμινοξέα είναι αμφοτερικής φύσης, δηλαδή δρουν και ως οξέα και ως βάση λόγω των δύο αμινών και καρβοξυλικών ομάδων που υπάρχουν.

Χημικές ιδιότητες

- Δοκιμή νινυδρίνης

Όταν 1 ml διαλύματος Νινυδρίνης προστίθεται σε διάλυμα πρωτεΐνης 1 ml και θερμαίνεται, ο σχηματισμός ιώδους χρώματος υποδηλώνει την παρουσία α-αμινοξέων.

- Ξανθοπρωτεϊκό τεστ

Η ξανθοπρωτεϊκή δοκιμή πραγματοποιείται για την ανίχνευση αρωματικών αμινοξέων (τυροσίνη, τρυπτοφάνη και φαινυλαλανίνη) σε διάλυμα πρωτεΐνης. Η νίτρωση των βενζοειδών ριζών που υπάρχουν στην αλυσίδα αμινοξέων συμβαίνει λόγω αντίδρασης με νιτρικό οξύ, δίνοντας στο διάλυμα κίτρινο χρωματισμό.

Χημικές ιδιότητες

- Αντίδραση με το αντιδραστήριο Sanger

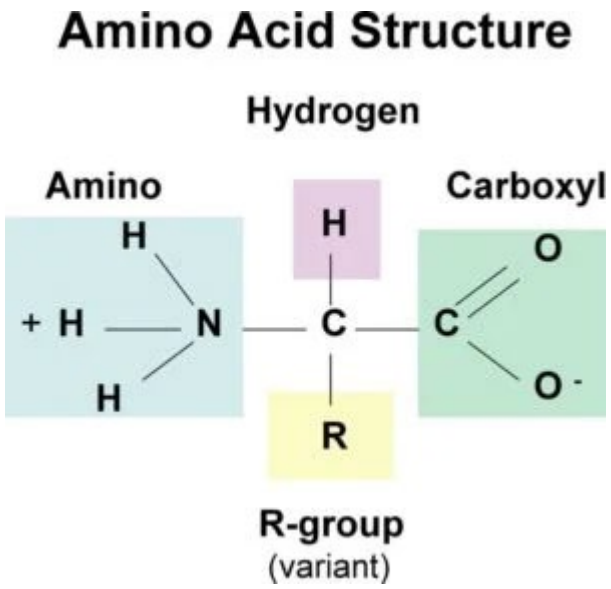
Το αντιδραστήριο Sanger (1-φθορο-2, 4-δινιτροβενζόλιο) αντιδρά με μια ελεύθερη αμινομάδα στην πεπτιδική αλυσίδα σε ένα ήπιο αλκαλικό μέσο υπό ψυχρές συνθήκες.

- Αντίδραση με νιτρώδες οξύ

Το νιτρώδες οξύ αντιδρά με την αμινομάδα για να απελευθερώσει άζωτο και να σχηματίσει το αντίστοιχο υδροξύλιο.

Co-funded by the Erasmus+ Programme of the European Union

Δομή αμινοξέων



Και τα 20 κοινά αμινοξέα είναι άλφα-αμινοξέα. Περιέχουν μια καρβοξυλική ομάδα, μια αμινομάδα και μια πλευρική αλυσίδα (ομάδα R), όλα συνδεδεμένα με τον α-άνθρακα.

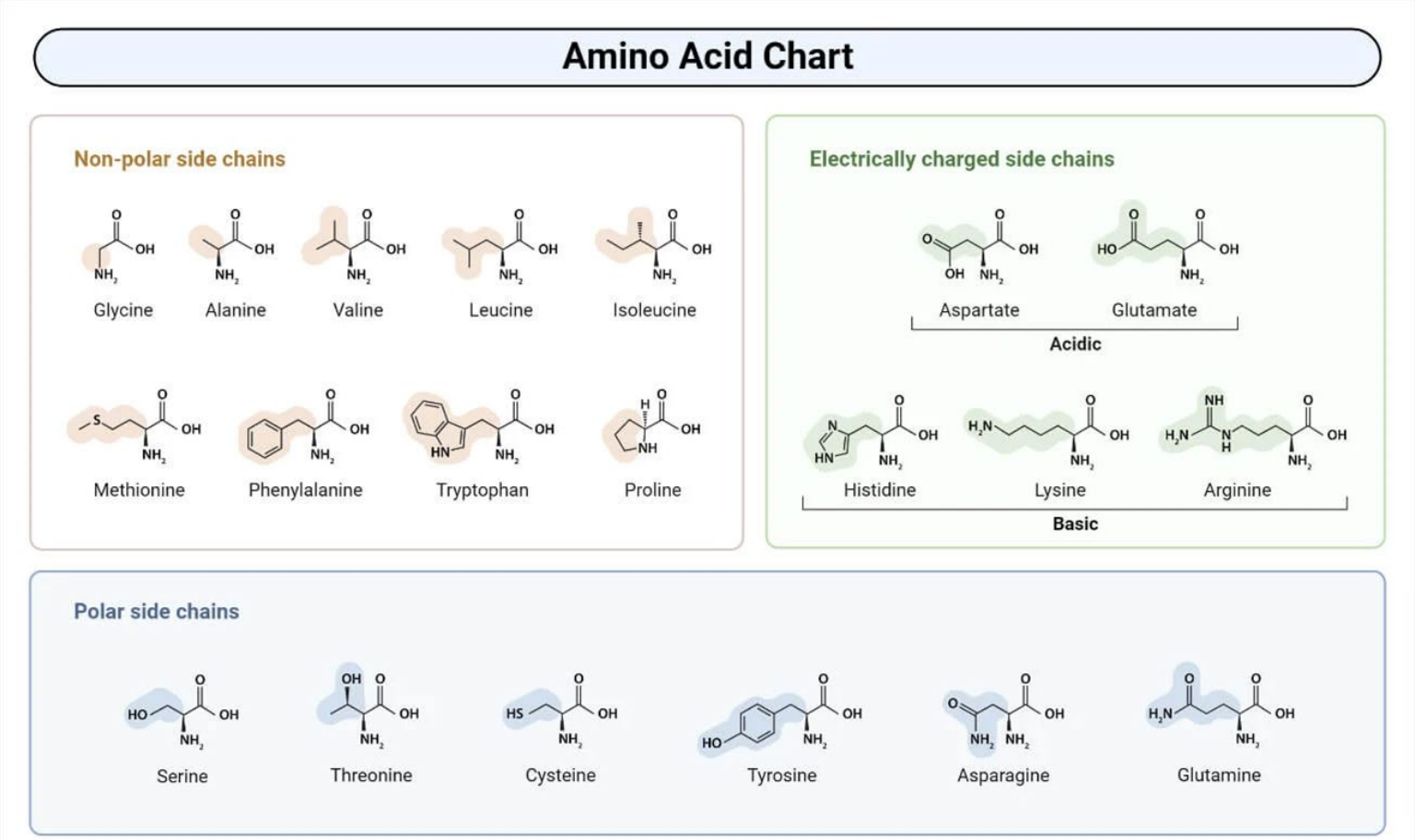
Εξαιρέσεις είναι:

- Γλυκίνη, η οποία δεν έχει πλευρική αλυσίδα. Ο α-άνθρακάς του περιέχει δύο υδρογόνα.
- Προλίνη, στην οποία το άζωτο είναι μέρος ενός δακτυλίου.
- Έτσι, κάθε αμινοξύ έχει μια ομάδα αμίνης στο ένα άκρο και μια ομάδα οξέος στο άλλο, και μια χαρακτηριστική πλευρική αλυσίδα. Η ραχοκοκαλιά είναι ίδια για όλα τα αμινοξέα ενώ η πλευρική αλυσίδα διαφέρει από το ένα αμινοξύ στο άλλο.
- Και τα 20 αμινοξέα εκτός από τη γλυκίνη είναι της L-διάταξης, καθώς για όλα εκτός από ένα αμινοξέα ο α-άνθρακας είναι ασύμμετρος άνθρακας. Επειδή η γλυκίνη δεν περιέχει ασύμμετρο άτομο άνθρακα, δεν είναι οπτικά ενεργή και, επομένως, δεν είναι ούτε D ούτε L.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ταξινόμηση αμινοξέων με βάση την ομάδα R



Ταξινόμηση αμινοξέων με βάση την ομάδα R

1. **Μη πολικά, αλειφατικά αμινοξέα** : Οι ομάδες R σε αυτήν την κατηγορία αμινοξέων είναι μη πολικές και υδρόφοβες. Γλυκίνη, Αλανίνη, Βαλίνη, Λευκίνη, Ισολευκίνη, Μεθειονίνη, Προλίνη.
2. **Αρωματικά αμινοξέα** : Η φαινυλαλανίνη, η τυροσίνη και η τρυπτοφάνη, με τις αρωματικές πλευρικές τους αλυσίδες, είναι σχετικά μη πολικές (υδρόφοβες). Όλοι μπορούν να συμμετέχουν σε υδρόφοβες αλληλεπιδράσεις.
3. **Πολικά, μη φορτισμένα αμινοξέα** : Οι ομάδες R αυτών των αμινοξέων είναι πιο διαλυτές στο νερό ή πιο υδρόφιλες από εκείνες των μη πολικών αμινοξέων, επειδή περιέχουν λειτουργικές ομάδες που σχηματίζουν δεσμούς υδρογόνου με το νερό. Αυτή η κατηγορία αμινοξέων περιλαμβάνει σερίνη, θρεονίνη, κυστεΐνη, ασπαραγίνη και γλουταμίνη.
4. **Όξινα αμινοξέα** : Αμινοξέα στα οποία η ομάδα R είναι όξινη ή αρνητικά φορτισμένη. Γλουταμινικό οξύ και Ασπαρτικό οξύ
5. **Βασικά αμινοξέα** : Αμινοξέα στα οποία η ομάδα R είναι βασική ή θετικά φορτισμένη. Λυσίνη, Αργινίνη, Ιστιδίνη

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ταξινόμηση αμινοξέων με βάση τη διατροφή

Essential	Conditionally Non-Essential	Non-Essential
Histidine	Arginine	Alanine
Isoleucine	Cystine	Asparagine
Leucine	Glutamine	Aspartate
Lysine	Glycine	Glutamate
Methionine	Proline	Serine
Phenylalanine	Tyrosine	
Threonine		
Tryptophan		
Valine		

Ταξινόμηση αμινοξέων με βάση τη διατροφή

Απαραίτητα αμινοξέα (Εννέα)

Εννέα αμινοξέα δεν μπορούν να συντεθούν στο σώμα και, ως εκ τούτου, πρέπει να υπάρχουν στη διατροφή για να συμβεί η πρωτεϊνοσύνθεση.

Αυτά τα απαραίτητα αμινοξέα είναι η ιστιδίνη, η ισολευκίνη, η λευκίνη, η λυσίνη, η μεθειονίνη, η φαινυλαλανίνη, η θρεονίνη, η τρυπτοφάνη και η βαλίνη.

Μη απαραίτητα αμινοξέα (Έντεκα)

Αυτά τα αμινοξέα μπορούν να συντεθούν στο ίδιο το σώμα και ως εκ τούτου δεν χρειάζεται απαραίτητα να αποκτηθούν μέσω της διατροφής.

Αυτά τα μη απαραίτητα αμινοξέα είναι η αργινίνη, η γλουταμίνη, η τυροσίνη, η κυστεΐνη, η γλυκίνη, η προλίνη, η σερίνη, η ορνιθίνη, η αλανίνη, η ασπαραγίνη και το ασπαρτικό.

Ταξινόμηση αμινοξέων με βάση τη μεταβολική μοίρα

Glucogenic amino acids	Glucogenic and ketogenic	Ketogenic amino acids
Alanine, Arginine, Asparagine, Aspartate Asparagine, Cysteine, Methionine Glutamate, Glutamine, Glycine, Histidine Proline, Serine, Threonine,Valine	Tyrosine Isoleucine Phenylalanine Tryptophan	Leucine Lysine

Ταξινόμηση αμινοξέων με βάση τη μεταβολική μοίρα

1. **Γλυκογονικά αμινοξέα** : Αυτά τα αμινοξέα χρησιμεύουν ως πρόδρομοι της γλυκονεογένεσης για το σχηματισμό γλυκόζης. Γλυκίνη, αλανίνη, σερίνη, ασπαρτικό οξύ, ασπαραγίνη, γλουταμικό οξύ, γλουταμίνη, προλίνη, βαλίνη, μεθειονίνη, κυστεΐνη, ιστιδίνη και αργινίνη.
2. **Κετογόνα αμινοξέα** : Αυτά τα αμινοξέα διασπώνται για να σχηματίσουν κετονοσώματα. Λευκίνη και Λυσίνη.
3. **Τόσο γλυκογόνα όσο και κετογονικά αμινοξέα** : Αυτά τα αμινοξέα διασπώνται για να σχηματίσουν πρόδρομες ουσίες τόσο για τα κετονοσώματα όσο και για τη γλυκόζη. Ισολευκίνη, φαινυλαλανίνη, τρυπτοφάνη και τυροσίνη.

Co-funded by the Erasmus+ Programme of the European Union

Λειτουργίες αμινοξέων

1. Συγκεκριμένα, 20 πολύ σημαντικά αμινοξέα είναι ζωτικής σημασίας για τη ζωή καθώς περιέχουν πεπτίδια και πρωτεΐνες και είναι γνωστό ότι αποτελούν τα δομικά στοιχεία για όλα τα έμβια όντα.
2. Η γραμμική αλληλουχία των υπολειμμάτων αμινοξέων σε μια πολυπεπτιδική αλυσίδα καθορίζει την τρισδιάστατη διαμόρφωση μιας πρωτεΐνης και η δομή μιας πρωτεΐνης καθορίζει τη λειτουργία της.
3. Τα αμινοξέα είναι απαραίτητα για τη διατήρηση της υγείας του ανθρώπινου σώματος. Προωθούν σε μεγάλο βαθμό τα:

Παραγωγή ορμονών

- Δομή των μυών
- Η υγιής λειτουργία του ανθρώπινου νευρικού συστήματος
- Η υγεία των ζωτικών οργάνων
- Φυσιολογική κυτταρική δομή

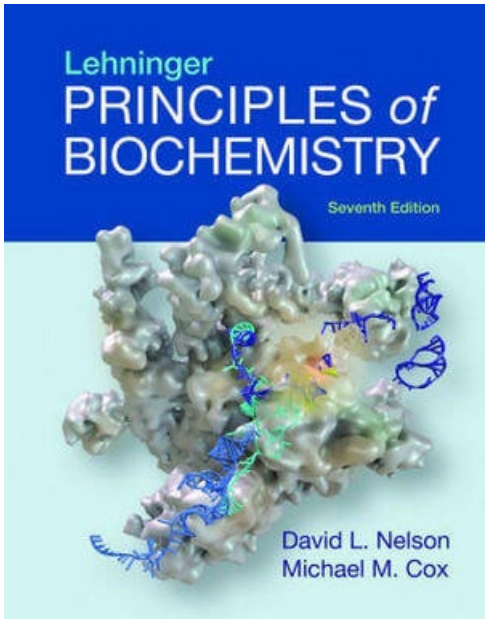
Co-funded by the Erasmus+ Programme of the European Union

Λειτουργίες αμινοξέων

4. Τα αμινοξέα χρησιμοποιούνται από διάφορους ιστούς για τη σύνθεση πρωτεϊνών και για την παραγωγή ενώσεων που περιέχουν άζωτο (π.χ. πουρίνες, αίμη , κρεατίνη , επινεφρίνη) ή οξειδώνονται για την παραγωγή ενέργειας.
5. Η διάσπαση τόσο των διατροφικών πρωτεϊνών όσο και των πρωτεϊνών των ιστών αποδίδει υποστρώματα που περιέχουν άζωτο και σκελετούς άνθρακα.
6. Τα υποστρώματα που περιέχουν άζωτο χρησιμοποιούνται στη βιοσύνθεση πουρινών, πυριμιδινών, νευροδιαβιβαστών, ορμονών, πορφυρινών και μη βασικών αμινοξέων.
7. Οι σκελετοί άνθρακα χρησιμοποιούνται ως πηγή καυσίμου στον κύκλο του κιτρικού οξέος, χρησιμοποιούνται για τη γλυκονεογένεση ή χρησιμοποιούνται στη σύνθεση λιπαρών οξέων.

Ερωτήσεις

- Εξηγήστε τις φυσικές και χημικές ιδιότητες των αμινοξέων.
- Ποιες είναι οι πιο δημοφιλείς ταξινομήσεις αμινοξέων;



Lehninger , AL, Nelson, DL, & Cox, MM (2000). Αρχές βιοχημείας Lehninger . Νέα Υόρκη: Worth Publishers.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 3. Πρωτεΐνες και ένζυμα

Μάθημα 2. Δομή πρωτεϊνών

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

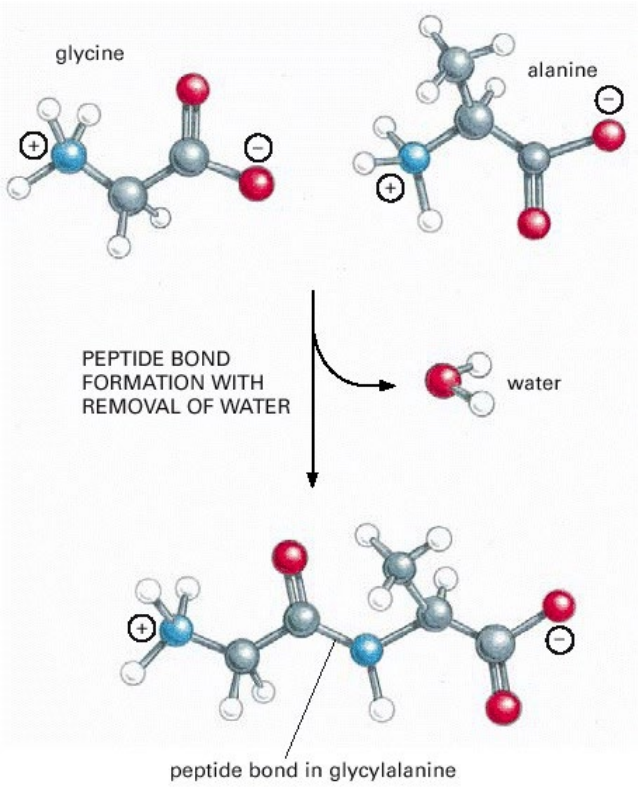
- Εισαγωγή
- Πρωτογενής Δομή
- Δευτερογενής Δομή
- Τριτογενής Δομή
- Τεταρτογενής Δομή

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Από χημική άποψη, οι πρωτεΐνες είναι μακράν τα πιο δομικά πολύπλοκα και λειτουργικά εξελιγμένα μόρια που είναι γνωστά. Αυτό ίσως δεν αποτελεί έκπληξη, μόλις συνειδητοποιήσει κανείς ότι η δομή και η χημεία κάθε πρωτεΐνης έχει αναπτυχθεί και βελτιωθεί σε δισεκατομμύρια χρόνια εξελικτικής ιστορίας. Ξεκινάμε αυτό το κεφάλαιο εξετάζοντας πώς η θέση κάθε αμινοξέος στη μακριά σειρά αμινοξέων που σχηματίζει μια πρωτεΐνη καθορίζει το τρισδιάστατο σχήμα του. Στη συνέχεια, θα χρησιμοποιήσουμε αυτήν την κατανόηση της δομής της πρωτεΐνης σε ατομικό επίπεδο για να περιγράψουμε πώς το ακριβές σχήμα κάθε μορίου πρωτεΐνης καθορίζει τη λειτουργία του σε ένα κύτταρο.

Το σχήμα μιας πρωτεΐνης καθορίζεται από την αλληλουχία αμινοξέων της



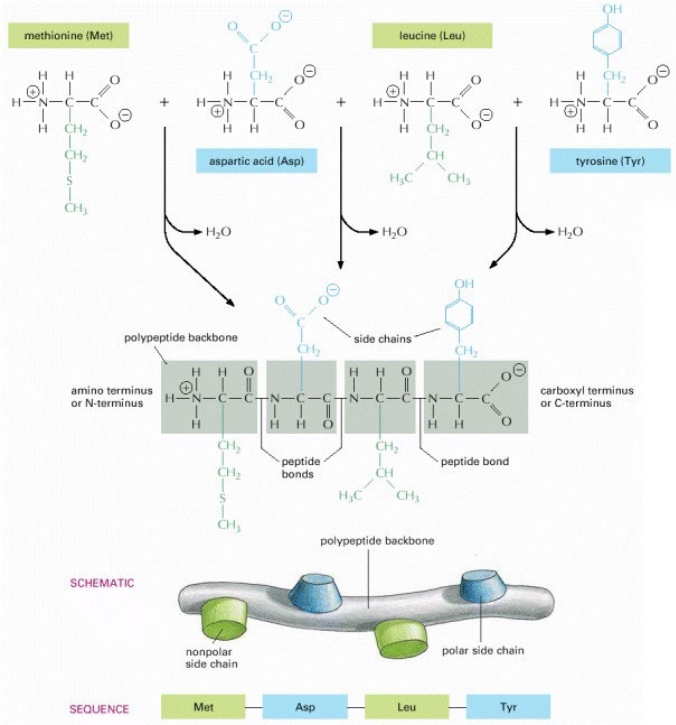
Ένας πεπτιδικός δεσμός

Αυτός ο ομοιοπολικός δεσμός σχηματίζεται όταν το άτομο άνθρακα από την καρβοξυλομάδα ενός αμινοξέος μοιράζεται ηλεκτρόνια με το άτομο αζώτου (μπλε) από την αμινομάδα ενός δεύτερου αμινοξέος. Όπως υποδεικνύεται, ένα μόριο νερού χάνεται σε αυτή την αντίδραση συμπύκνωσης.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Το σχήμα μιας πρωτεΐνης καθορίζεται από την αλληλουχία αμινοξέων της



Τα δομικά συστατικά μιας πρωτεΐνης

Μια πρωτεΐνη αποτελείται από μια πολυπεπτιδική ραχοκοκαλιά με προσαρτημένες πλευρικές αλυσίδες. Κάθε τύπος πρωτεΐνης διαφέρει ως προς την αλληλουχία και τον αριθμό των αμινοξέων. Επομένως, είναι η αλληλουχία των χημικά διαφορετικών πλευρικών αλυσίδων που κάνει κάθε πρωτεΐνη ξεχωριστή. Τα δύο άκρα μιας πολυπεπτιδικής αλυσίδας είναι χημικά διαφορετικά: το άκρο που φέρει την ελεύθερη αμινομάδα (NH3 +, επίσης γραμμένο NH2) είναι το αμινοτελικό άκρο, ή N-άκρο, και αυτό που φέρει την ελεύθερη καρβοξυλική ομάδα (COO–, επίσης γραμμένο COOH) είναι το καρβοξυλικό άκρο ή C-άκρο. Η αλληλουχία αμινοξέων μιας πρωτεΐνης παρουσιάζεται πάντα στην κατεύθυνση N-to-C, διαβάζοντας από αριστερά προς τα δεξιά.

Το σχήμα μιας πρωτεΐνης καθορίζεται από την αλληλουχία αμινοξέων της

AMINO ACID			SIDE CHAIN
Aspartic acid	Asp	D	negative
Glutamic acid	Glu	E	negative
Arginine	Arg	R	positive
Lysine	Lys	K	positive
Histidine	His	H	positive
Asparagine	Asn	N	uncharged polar
Glutamine	Gln	Q	uncharged polar
Serine	Ser	S	uncharged polar
Threonine	Thr	T	uncharged polar
Tyrosine	Tyr	Y	uncharged polar

POLAR AMINO ACIDS

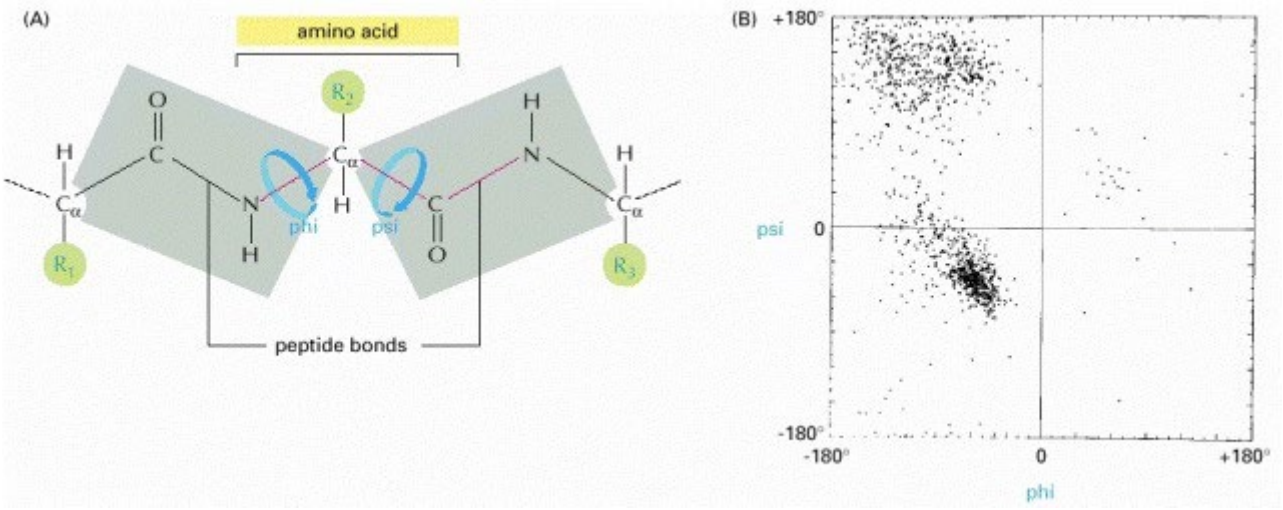
AMINO ACID			SIDE CHAIN
Alanine	Ala	A	nonpolar
Glycine	Gly	G	nonpolar
Valine	Val	V	nonpolar
Leucine	Leu	L	nonpolar
Isoleucine	Ile	I	nonpolar
Proline	Pro	P	nonpolar
Phenylalanine	Phe	F	nonpolar
Methionine	Met	M	nonpolar
Tryptophan	Trp	W	nonpolar
Cysteine	Cys	C	nonpolar

NONPOLAR AMINO ACIDS

Τα 20 αμινοξέα που βρίσκονται στις πρωτεΐνες

Παρατίθενται συντομογραφίες με τρία γράμματα και ένα γράμμα. Όπως φαίνεται, υπάρχουν ίσοι αριθμοί πολικών και μη πολικών πλευρικών αλυσίδων.

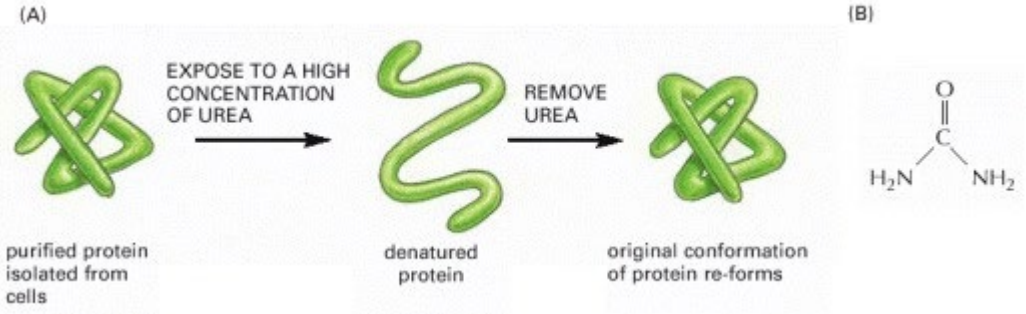
Το σχήμα μιας πρωτεΐνης καθορίζεται από την αλληλουχία αμινοξέων της



Στερικοί περιορισμοί στις γωνίες δεσμού σε μια πολυπεπτιδική αλυσίδα

(A) Κάθε αμινοξύ συνεισφέρει τρεις δεσμούς (κόκκινο) στη ραχοκοκαλιά της αλυσίδας. Ο πεπτιδικός δεσμός είναι επίπεδος (γκρι σκίαση) και δεν επιτρέπει την περιστροφή. Αντίθετα, η περιστροφή μπορεί να συμβεί γύρω από τον δεσμό Ca–C, του οποίου η γωνία περιστροφής ονομάζεται psi (ψ), και για τον δεσμό N–Cα, του οποίου η γωνία περιστροφής ονομάζεται phi (ϕ). Κατά σύμβαση, μια ομάδα R χρησιμοποιείται συχνά για να υποδηλώσει μια πλευρική αλυσίδα αμινοξέων (πράσινοι κύκλοι). (B) Η διαμόρφωση των ατόμων της κύριας αλυσίδας σε μια πρωτεΐνη προσδιορίζεται από ένα ζεύγος γωνιών ϕ και ψ για κάθε αμινοξύ. Λόγω των στερικών συγκρούσεων μεταξύ ατόμων σε κάθε αμινοξύ, τα περισσότερα ζεύγη γωνιών ϕ και ψ δεν συμβαίνουν. Σε αυτό το αποκαλούμενο διάγραμμα Ramachandran, κάθε κουκκίδα αντιπροσωπεύει ένα παρατηρούμενο ζεύγος γωνιών σε μια πρωτεΐνη.

Οι πρωτεΐνες διπλώνουν σε μια διαμόρφωση της χαμηλότερης ενέργειας



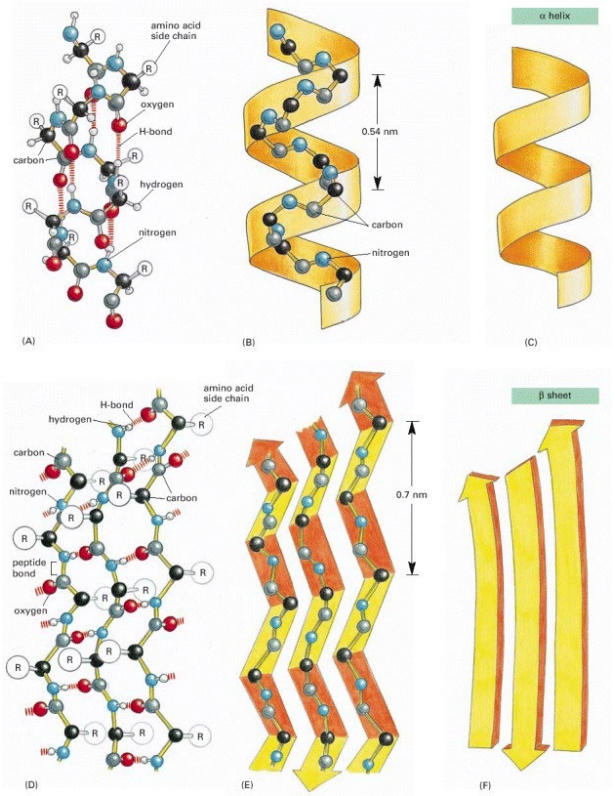
Η αναδίπλωση μιας μετουσιωμένης πρωτεΐνης

(Α) Αυτό το πείραμα καταδεικνύει ότι η διαμόρφωση μιας πρωτεΐνης καθορίζεται αποκλειστικά από την αλληλουχία αμινοξέων της. (Β) Η δομή της ουρίας. Η ουρία είναι πολύ διαλυτή στο νερό και ξεδιπλώνει πρωτεΐνες σε υψηλές συγκεντρώσεις, όπου υπάρχει περίπου ένα μόριο ουρίας για κάθε έξι μόρια νερού.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Η έλικα α και το φύλλο β είναι κοινά αναδιπλούμενα μοτίβα

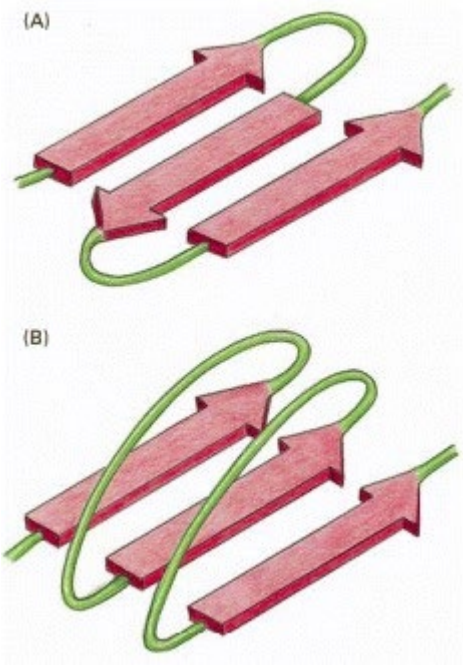


Η κανονική διαμόρφωση του πολυπεπτιδικού σκελετού που παρατηρείται στην α έλικα και στο φύλλο β

(Α, Β και Γ) Η έλικα α. Το N-H κάθε πεπτιδικού δεσμού συνδέεται με υδρογόνο με το C=O ενός γειτονικού πεπτιδικού δεσμού που βρίσκεται τέσσερις πεπτιδικούς δεσμούς μακριά στην ίδια αλυσίδα. (Δ, Ε και ΣΤ) Το φύλλο β. Σε αυτό το παράδειγμα, γειτονικές πεπτιδικές αλυσίδες τρέχουν σε αντίθετες (αντιπαράλληλες) κατευθύνσεις. Οι μεμονωμένες πολυπεπτιδικές αλυσίδες (κλώνοι) σε ένα φύλλο β συγκρατούνται μεταξύ τους με δεσμούς υδρογόνου μεταξύ πεπτιδικών δεσμών σε διαφορετικούς κλώνους και οι πλευρικές αλυσίδες αμινοξέων σε κάθε κλώνο προβάλλουν εναλλάξ πάνω και κάτω από το επίπεδο του φύλλου. Τα (Α) και (Δ) δείχνουν όλα τα άτομα στη ραχοκοκαλιά του πολυπεπτιδίου, αλλά οι πλευρικές αλυσίδες αμινοξέων περικόπτονται και συμβολίζονται με R. Αντίθετα, τα (Β) και (Ε) δείχνουν μόνο τα άτομα ραχοκοκαλιάς, ενώ τα (C) και (F) εμφανίζει τα σύμβολα στενογραφίας που χρησιμοποιούνται για την αναπαράσταση της έλικας α και του φύλλου β σε σχέδια με κορδέλα πρωτεϊνών

Co-funded by the Erasmus+ Programme of the European Union

Η έλικα α και το φύλλο β είναι κοινά αναδιπλούμενα μοτίβα



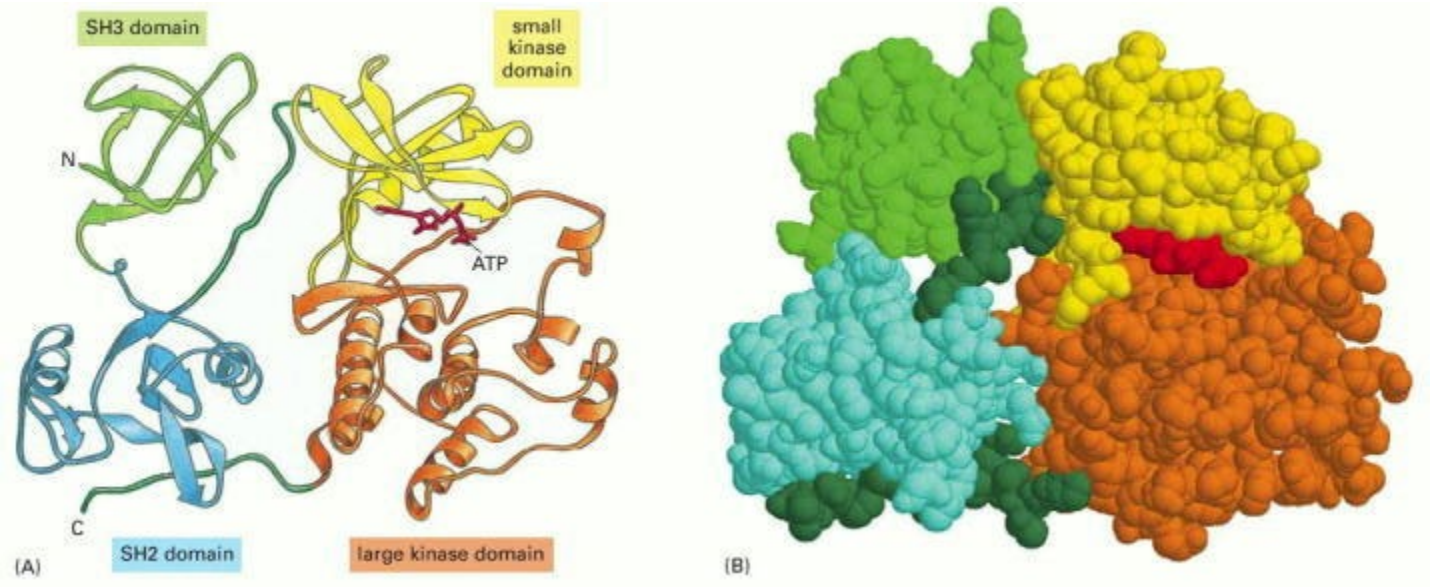
Δύο τύποι δομών φύλλων β

(Α) Ένα αντιπαράλληλο φύλλο β (βλ. Εικόνα 3-9D). (Β) Ένα παράλληλο β φύλλο. Και οι δύο αυτές δομές είναι κοινές στις πρωτεΐνες.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Το Protein Domain είναι μια θεμελιώδης μονάδα οργάνωσης



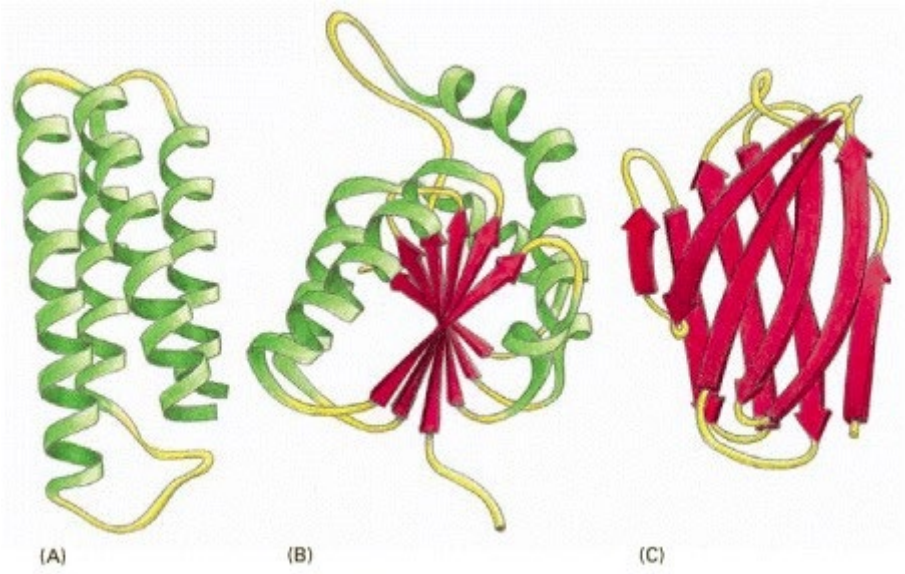
Μια πρωτεΐνη που σχηματίζεται από τέσσερις τομείς

Στην πρωτεΐνη Src που φαίνεται, δύο από τις περιοχές σχηματίζουν ένα ένζυμο πρωτεϊνικής κινάσης, ενώ οι περιοχές SH2 και SH3 εκτελούν ρυθμιστικές λειτουργίες. (A) Μοντέλο κορδέλας, με υπόστρωμα ATP σε κόκκινο χρώμα. (B) Ένα μοντέλο πλήρωσης απόστασης, με υπόστρωμα ATP σε κόκκινο χρώμα. Σημειώστε ότι η θέση που δεσμεύει το ATP είναι τοποθετημένη στη διεπαφή των δύο τομέων που σχηματίζουν την κινάση.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Το Protein Domain είναι μια θεμελιώδης μονάδα οργάνωσης



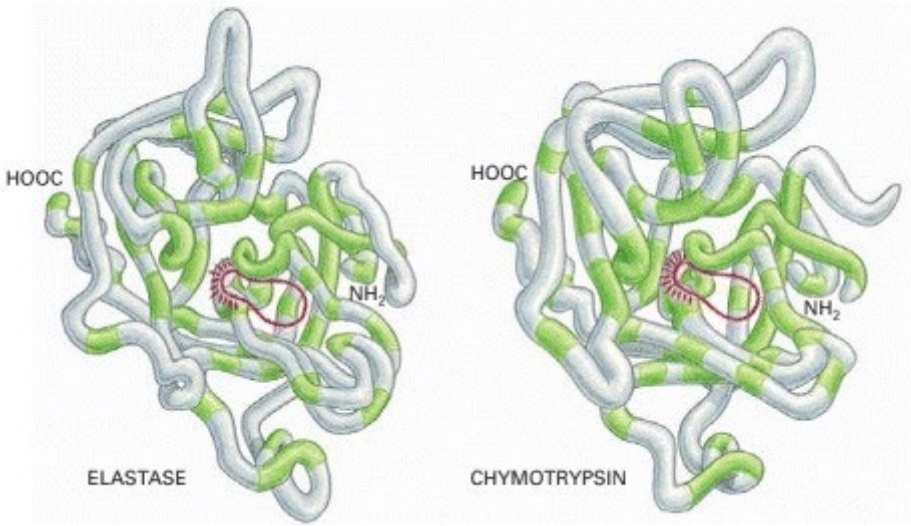
Μοντέλα κορδέλας τριών διαφορετικών πρωτεϊνικών περιοχών

(A) Κυτόχρωμα b562, μια πρωτεΐνη μονού τομέα που εμπλέκεται στη μεταφορά ηλεκτρονίων στα μιτοχόνδρια. Αυτή η πρωτεΐνη αποτελείται σχεδόν εξ ολοκλήρου από α έλικες. (Β) Η περιοχή δέσμευσης NAD του ενζύμου γαλακτική αφυδρογονάση, η οποία αποτελείται από ένα μίγμα ελίκων α και β φύλλων. (Γ) Η μεταβλητή περιοχή μιας ελαφριάς αλυσίδας ανοσοσφαιρίνης (αντισώματος), που αποτελείται από ένα σάντουιτς δύο φύλλων β. Σε αυτά τα παραδείγματα, οι α έλικες φαίνονται με πράσινο χρώμα, ενώ οι κλώνοι που οργανώνονται ως φύλλα β υποδηλώνονται με κόκκινα βέλη.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Οι πρωτεΐνες μπορούν να ταξινομηθούν σε πολλές οικογένειες



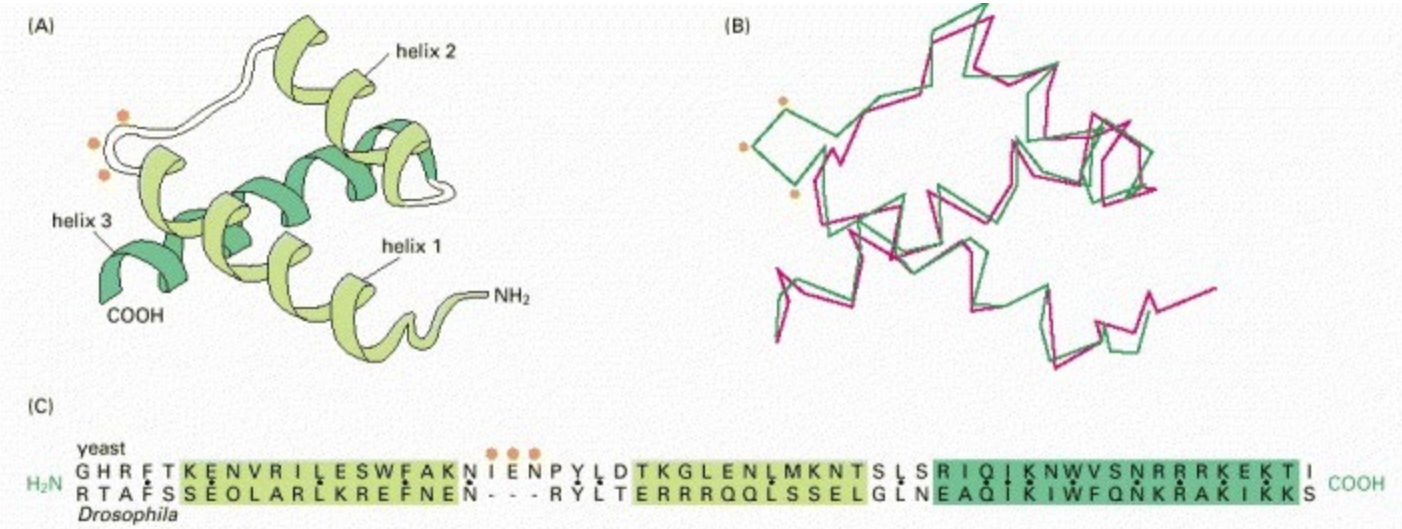
Συγκρίθηκαν οι διαμορφώσεις δύο πρωτεασών σερίνης

Οι διαμορφώσεις της ραχοκοκαλιάς της ελαστάσης και της χυμοθρυψίνης. Αν και μόνο αυτά τα αμινοξέα στην πολυπεπτιδική αλυσίδα που σκιάζονται με πράσινο είναι τα ίδια στις δύο πρωτεΐνες, οι δύο διαμορφώσεις είναι πολύ παρόμοιες σχεδόν παντού. Η ενεργή θέση κάθε ενζύμου είναι κυκλωμένη με κόκκινο χρώμα. Αυτό είναι όπου οι πεπτιδικοί δεσμοί των πρωτεϊνών που χρησιμεύουν ως υποστρώματα συνδέονται και διασπώνται με υδρόλυση. Οι πρωτεάσες σερίνης παίρνουν το όνομά τους από το αμινοξύ σερίνη, του οποίου η πλευρική αλυσίδα είναι μέρος της ενεργού θέσης κάθε ενζύμου και συμμετέχει άμεσα στην αντίδραση διάσπασης.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Οι πρωτεΐνες μπορούν να ταξινομηθούν σε πολλές οικογένειες



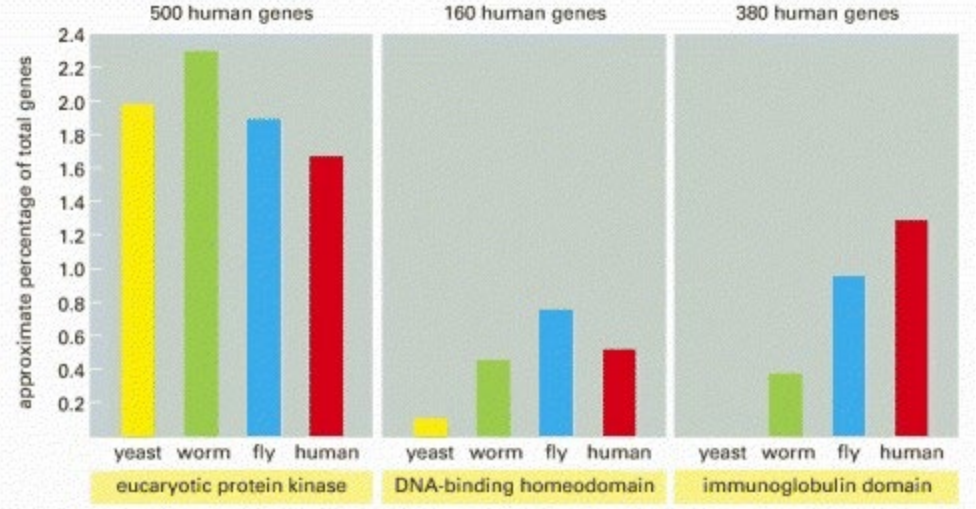
Σύγκριση μιας κατηγορίας τομέων δέσμευσης DNA, που ονομάζονται ομοιοτομείς, σε ένα ζευγάρι πρωτεϊνών από δύο οργανισμούς που χωρίζονται από περισσότερα από ένα δισεκατομμύριο χρόνια εξέλιξης

(Α) Ένα μοντέλο κορδέλας της δομής που είναι κοινή και στις δύο πρωτεΐνες. (Β) Ένα ίχνος των θέσεων α-άνθρακα. Οι τρισδιάστατες δομές που φαίνονται προσδιορίστηκαν με κρυσταλλογραφία ακτίνων Χ για την πρωτεΐνη ζυμομύκητα α2 (πράσινη) και την πρωτεΐνη Drosophila (κόκκινη). (Γ) Σύγκριση αλληλουχιών αμινοξέων για την περιοχή των πρωτεϊνών που παρουσιάζονται στα (Α) και (Β). Οι μαύρες κουκκίδες σηματοδοτούν θέσεις με πανομοιότυπα αμινοξέα. Οι πορτοκαλί κουκκίδες υποδεικνύουν τη θέση ενός ενθέματος τριών αμινοξέων στην πρωτεΐνη α2.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

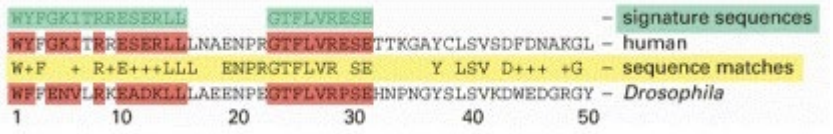
Οι πρωτεΐνες μπορούν να ταξινομηθούν σε πολλές οικογένειες



Ποσοστό συνολικών γονιδίων που περιέχουν ένα ή περισσότερα αντίγραφα του υποδεικνυόμενου τομέα πρωτεΐνης, όπως προέρχονται από πλήρεις αλληλουχίες γονιδιώματος

Σημειώστε ότι ένας από τους τρεις τομείς που επιλέχθηκαν, ο τομέας ανοσοσφαιρίνης, ήταν σχετικά καθυστερημένη προσθήκη και η σχετική αφθονία του έχει αυξηθεί στη γενεαλογία των σπονδυλωτών. Οι εκτιμήσεις του αριθμού των ανθρώπινων γονιδίων είναι κατά προσέγγιση.

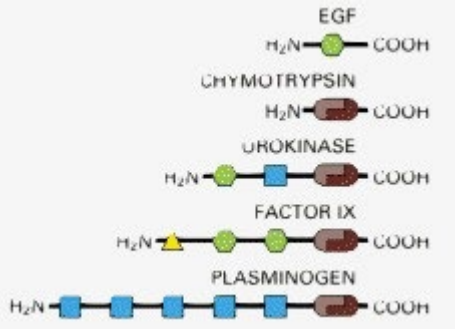
Οι αναζητήσεις ομολογίας ακολουθίας μπορούν να προσδιορίσουν στενούς συγγενείς



Η χρήση σύντομων αλληλουχιών υπογραφής για την εύρεση ομόλογων πρωτεϊνικών περιοχών

Οι δύο σύντομες αλληλουχίες των 15 και 9 αμινοξέων που εμφανίζονται (πράσινο) μπορούν να χρησιμοποιηθούν για την αναζήτηση μεγάλων βάσεων δεδομένων για έναν τομέα πρωτεΐνης που βρίσκεται σε πολλές πρωτεΐνες, τον τομέα SH2. Εδώ, τα πρώτα 50 αμινοξέα της περιοχής SH2 των 100 αμινοξέων συγκρίνονται για την ανθρώπινη πρωτεΐνη και την πρωτεΐνη Drosophila Src . Στη σύγκριση αλληλουχιών που δημιουργείται από υπολογιστή (κίτρινη σειρά), οι ακριβείς αντιστοιχίες μεταξύ των πρωτεϊνών του ανθρώπου και της Drosophila σημειώνονται με τη συντομογραφία ενός γράμματος για το αμινοξύ. Οι θέσεις με ένα παρόμοιο αλλά μη ταυτόσημο αμινοξύ συμβολίζονται με + και οι μη ταιριασμένες είναι κενές. Σε αυτό το διάγραμμα, όπου μία ή και οι δύο πρωτεΐνες περιέχουν ακριβή αντιστοίχιση με μια θέση στις πράσινες αλληλουχίες, και οι δύο ευθυγραμμισμένες αλληλουχίες χρωματίζονται κόκκινο.

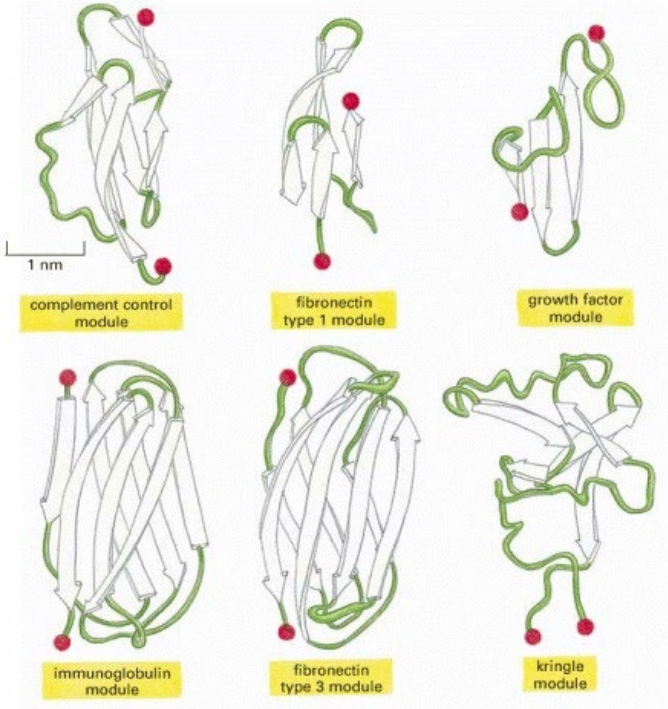
Ορισμένοι τομείς πρωτεΐνης, που ονομάζονται ενότητες, αποτελούν μέρη πολλών διαφορετικών πρωτεϊνών



Μια εκτεταμένη ανακάτεμα των μπλοκ της πρωτεϊνικής αλληλουχίας (πρωτεϊνικοί τομείς) έχει συμβεί κατά την εξέλιξη της πρωτεΐνης. Αυτά τα τμήματα μιας πρωτεΐνης που υποδηλώνονται με το ίδιο σχήμα και χρώμα σε αυτό το διάγραμμα σχετίζονται εξελικτικά. Οι πρωτεάσες σερίνης όπως η χυμοθρυψίνη σχηματίζονται από δύο τομείς (καφέ). Στις τρεις άλλες πρωτεάσες που παρουσιάζονται, οι οποίες είναι εξαιρετικά ρυθμισμένες και πιο εξειδικευμένες, αυτές οι δύο περιοχές πρωτεάσης συνδέονται με έναν ή περισσότερους τομείς ομόλογους με τομείς που βρίσκονται στον επιδερμικό αυξητικό παράγοντα (EGF, πράσινο), με μια πρωτεΐνη που δεσμεύει το ασβέστιο (κίτρινο). ή σε μια περιοχή « kringle » (μπλε) που περιέχει τρεις εσωτερικές δισουλφιδικές γέφυρες.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Ορισμένοι τομείς πρωτεΐνης, που ονομάζονται ενότητες, αποτελούν μέρη πολλών διαφορετικών πρωτεϊνών



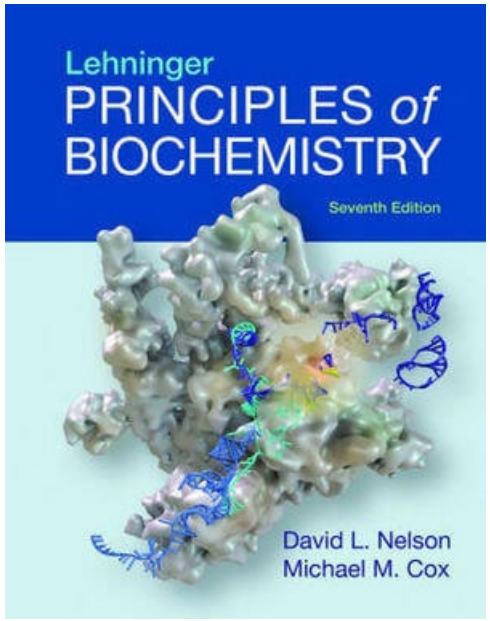
Οι τρισδιάστατες δομές ορισμένων μονάδων πρωτεΐνης

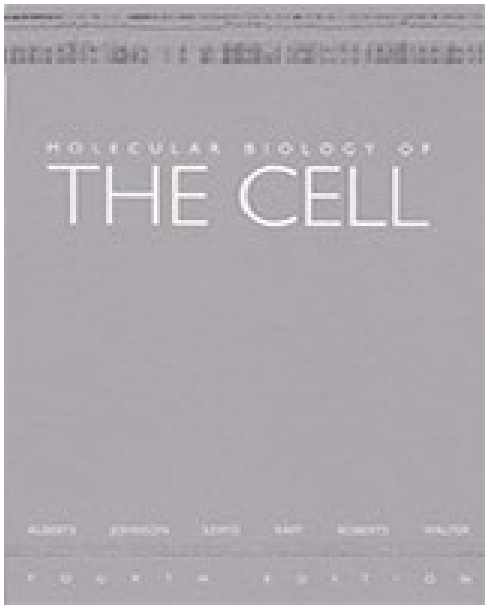
Σε αυτά τα διαγράμματα κορδέλας, οι κλώνοι φύλλων β φαίνονται ως βέλη και τα N- και C-άκρα υποδεικνύονται με κόκκινες σφαίρες

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε τη δομική οργάνωση των πεπτιδίων.
- Ποια είναι τα κύρια χαρακτηριστικά κάθε επιπέδου οργάνωσης των πεπτιδικών μορίων;





Lehninger , AL, Nelson, DL, & Cox, MM (2000). Αρχές βιοχημείας Lehninger . Νέα Υόρκη: Worth Publishers.

Μοριακή Βιολογία του Κυττάρου. 4η έκδοση.
Alberts B, Johnson A, Lewis J, et al. Νέα Υόρκη: Garland Science; 2002.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 3. Πρωτεΐνες και ένζυμα

Μάθημα 3. Χημικοί δεσμοί στο μόριο πρωτεΐνης

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Πεπτιδικοί δεσμοί
- Ιοντικοί δεσμοί
- Δισουλφιδικοί δεσμοί
- Δεσμοί υδρογόνου
- Υδροφοβικοί δεσμοί

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Οι πρωτεΐνες είναι τα πολυμερή των αμινοξέων. Τα αμινοξέα ενώνονται μεταξύ τους με έναν ειδικό τύπο ομοιοπολικού δεσμού (πεπτιδικός δεσμός) για να σχηματίσουν γραμμικές δομές που ονομάζονται πολυπεπτίδια. Τα πολυπεπτίδια στη συνέχεια διπλώνονται σε συγκεκριμένες δομές για να σχηματίσουν τη λειτουργική διαμόρφωση της πρωτεΐνης. Η αναδίπλωση των πρωτεϊνών σε συγκεκριμένα σχήματα και διαμορφώσεις υποβοηθούνται και σταθεροποιούνται από πολλούς τύπους δεσμών σε αυτές. Μερικοί από αυτούς τους δεσμούς είναι ισχυροί δεσμοί ενώ άλλοι είναι αδύναμες αλληλεπιδράσεις. Σημαντικοί τύποι δεσμών που εμπλέκονται στη δομή και τη διαμόρφωση των πρωτεϊνών είναι οι πεπτιδικοί δεσμοί, οι ιοντικοί δεσμοί, οι δισουλφιδικοί δεσμοί, οι δεσμοί υδρογόνου και οι υδρόφοβες αλληλεπιδράσεις.

Τύποι ομολόγων

Υπάρχουν 5 τύποι χημικών δεσμών που παίζουν σημαντικό ρόλο στον καθορισμό και τη σταθεροποίηση της δομής της τρισδιάστατης πρωτεΐνης.

Αυτοί είναι:

1) Πεπτιδικοί δεσμοί

2) Ιοντικοί δεσμοί

3) Δισουλφιδικοί δεσμοί

4) Δεσμοί υδρογόνου

5) Υδροφοβικοί δεσμοί

Co-funded by the Erasmus+ Programme of the European Union

Πεπτιδικοί δεσμοί

- Ορισμός πεπτιδικού δεσμού: ένας ομοιοπολικός δεσμός που σχηματίζεται μεταξύ της καρβοξυλικής ομάδας ενός αμινοξέος και της ομάδας ενός άλλου αμινοξέος
- Ο πεπτιδικός δεσμός είναι ένας ισχυρός ομοιοπολικός δεσμός με υψηλή ενέργεια διάστασης δεσμού.
- Σχηματίζεται από την ένωση δύο υπολειμμάτων αμινοξέων κατά την πρωτεϊνοσύνθεση.
- Η καρβοξυλική ομάδα (-COOH) ενός αμινοξέος συνδυάζεται με την αμινομάδα (-NH2) ενός άλλου αμινοξέος για να σχηματίσει τον πεπτιδικό δεσμό.
- Ο σχηματισμός πεπτιδικού δεσμού είναι ένα παράδειγμα για μια αντίδραση συμπύκνωσης ή απομάκρυνσης.
- Ένα μόριο νερού αποβάλλεται κατά τον σχηματισμό πεπτιδικού δεσμού με την αντίδραση συμπύκνωσης δύο αμινοξέων.
- Η ένωση που προκύπτει μετά το σχηματισμό πεπτιδικού δεσμού ονομάζεται διπεπτίδιο.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

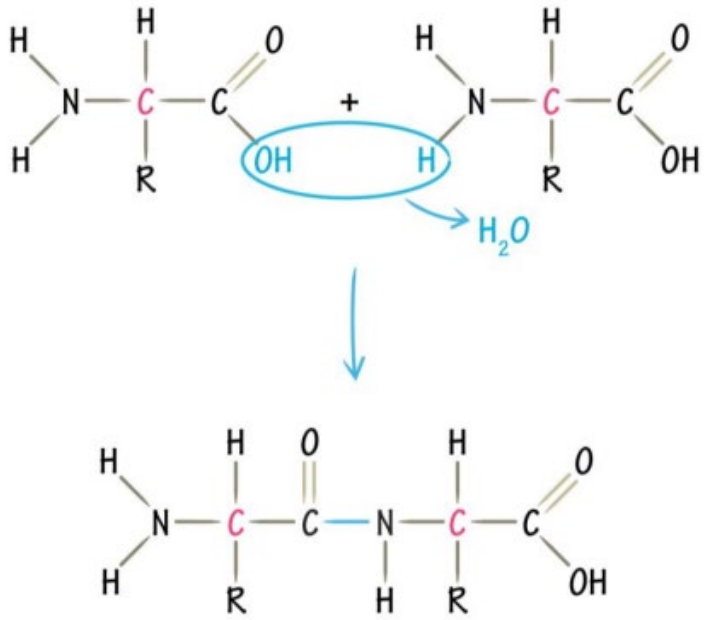
Πεπτιδικοί δεσμοί

- Ένα διπεπτίδιο έχει μια ελεύθερη αμινομάδα στο ένα άκρο και μια καρβοξυλική ομάδα στο άλλο άκρο.
- Η ελεύθερη αμινομάδα ή καρβοξυλική ομάδα ενός διπεπτιδίου μπορεί να σχηματίσει έναν άλλο πεπτιδικό δεσμό με έναν άλλο πεπτιδικό δεσμό με ένα τρίτο αμινοξύ και ούτω καθεξής.
- Πολλά αμινοξέα ενώνονται μεταξύ τους με αυτόν τον τρόπο για να σχηματίσουν ένα πολυπεπτίδιο.
- Ο σχηματισμός πεπτιδικών δεσμών διευκολύνεται από το ένζυμο Πεπτιδυλ τρανσφεράση κατά τη διάρκεια της διαδικασίας μετάφρασης της πρωτεϊνοσύνθεσης.
- Το ένζυμο πεπτιδυλ τρανσφεράση είναι ένα ριβοένζυμο. είναι μέρος του ριβοσωμικού RNA (rRNA) της μεγάλης υπομονάδας του ριβοσώματος.
- Στα προκαρυωτικά το 23S rRNA και στους ευκαρυώτες το 28S rRNA δρα ως ένζυμο πεπτιδυλ τρανσφεράση.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Η πρωτογενής δομή της πρωτεΐνης σταθεροποιείται από τον πεπτιδικό δεσμό



Co-funded by the Erasmus+ Programme of the European Union

Ιοντικός δεσμός

Ορισμός ιοντικού δεσμού: ένας χημικός δεσμός που σχηματίζεται μεταξύ των δύο ιόντων αντίθετων φορτίων.

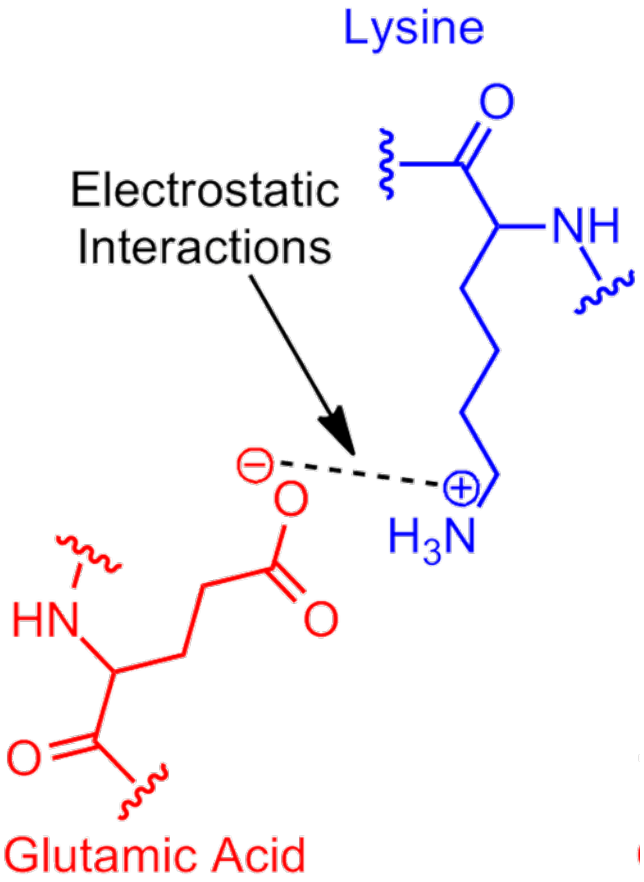
Στις πρωτεΐνες, οι ιοντικοί δεσμοί σχηματίζονται μεταξύ των ιονισμένων όξινων ή βασικών ομάδων αμινοξέων.

Οι ομάδες R (πλευρικές αλυσίδες) ορισμένων αμινοξέων περιέχουν επιπλέον όξινες (-COO-) ή βασικές (-NH3+) ομάδες.

Αυτές οι ομάδες R μπορούν να ιονιστούν για να παράγουν φορτισμένες ομάδες σε συγκεκριμένο pH.

Οι όξινες ομάδες R θα είναι αρνητικά φορτισμένες αφού απελευθερώνουν τα ιόντα H+

Οι βασικές ομάδες R θα είναι θετικά φορτισμένες αφού δέχονται ιόντα H+ από το μέσο



Οι βασικές ομάδες R θα είναι θετικά φορτισμένες αφού δέχονται ιόντα H+ από το μέσο

- Μετά τον ιονισμό της πλευρικής αλυσίδας όπως αναφέρθηκε παραπάνω, τα αμινοξέα στην πρωτεϊνική αλυσίδα μπορούν να έλκονται/απωθούν το ένα το άλλο με βάση τα φορτία τους. {Αυτό ονομάζεται επίσης "Ηλεκτροστατικές Αλληλεπιδράσεις"}
- Η έλξη των αντίθετα φορτισμένων ομάδων R έχει ως αποτέλεσμα το σχηματισμό ιοντικών δεσμών.
- Οι ιοντικοί δεσμοί είναι ασθενείς δεσμοί και είναι πολύ εύθραυστοι σε ένα υδατικό μέσο.
- Ακόμη και μια αλλαγή στο pH μπορεί να διασπάσει τους ιοντικούς δεσμούς.
- Αυτός είναι ο λόγος για τη μετουσίωση των πρωτεϊνών σε όξινο ή βασικό μέσο.
- Οι τριτοταγείς και τεταρτοταγείς δομές των πρωτεϊνών σταθεροποιούνται με ιοντικούς δεσμούς.

Co-funded by the Erasmus+ Programme of the European Union

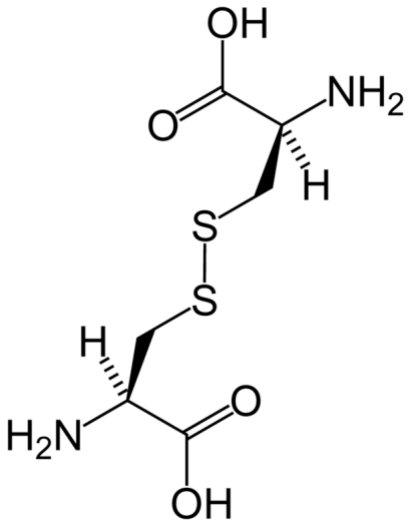
Δισουλφιδικοί δεσμοί

- Ορισμός δισουλφιδικού δεσμού: Ομοιοπολικός δεσμός που σχηματίζεται από δύο ομάδες θειόλης δύο υπολειμμάτων κυστεΐνης σε μια πρωτεΐνη.
- Η κυστεΐνη (Cys ή C, ένα αμινοξύ που περιέχει θείο) περιέχει μια εξαιρετικά δραστική σουλφυδρυλική ομάδα (-SH) στην πλευρική της αλυσίδα (ομάδα R).
- Το σουλφυδρύλιο είναι εξαιρετικά πολικό και εξαιρετικά αντιδραστικό.
- Εάν δύο μόρια μιας κυστεΐνης ευθυγραμμιστούν το ένα δίπλα στο άλλο, οι γειτονικές σουλφυδρυλικές ομάδες μπορούν να οξειδωθούν.
- Αυτή η αντίδραση έχει ως αποτέλεσμα το σχηματισμό μιας μόνιμης ομοιοπολικής σύνδεσης μεταξύ δύο υπολειμμάτων κυστεΐνης που ονομάζεται δισουλφιδικός δεσμός.
- Ο δισουλφιδικός δεσμός στη χημεία των πρωτεϊνών είναι πιο γνωστός ως δισουλφιδική γέφυρα ή δεσμός SS.

Co-funded by the Erasmus+ Programme of the European Union

Δισουλφιδικοί δεσμοί

- Είναι πολύ δυνατοί δεσμοί που δεν σπάνε εύκολα.
- Ένας δισουλφιδικός δεσμός μπορεί να σχηματιστεί μεταξύ των υπολειμμάτων κυστεΐνης της ίδιας πολυπεπτιδικής αλυσίδας ή διαφορετικής πολυπεπτιδικής αλυσίδας μιας λειτουργικής πρωτεΐνης.
- Οι δισουλφιδικοί δεσμοί σταθεροποιούν τις τριτοταγείς δομές της πρωτεΐνης.



Co-funded by the Erasmus+ Programme of the European Union

Δεσμοί υδρογόνου

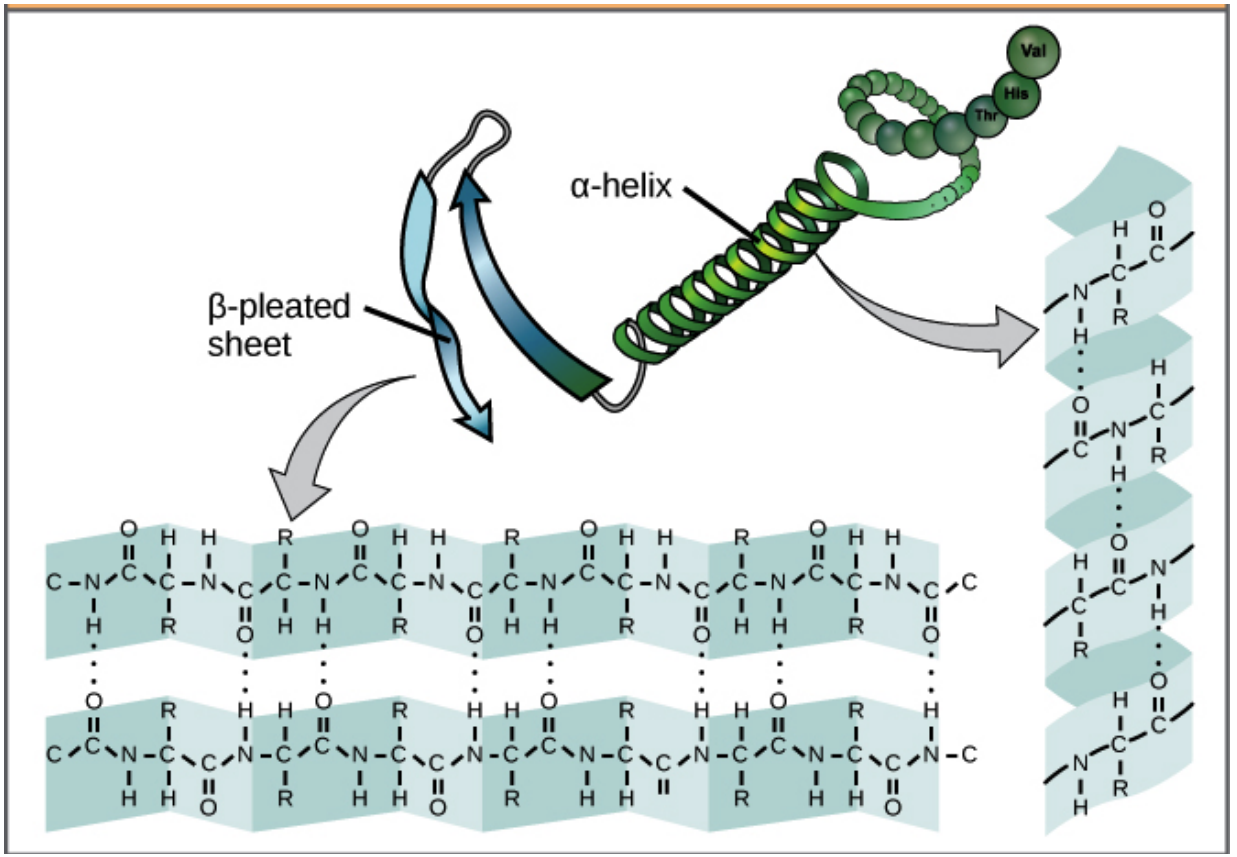
- Ορισμός δεσμού υδρογόνου: Ο δεσμός υδρογόνου είναι μια ηλεκτροστατική έλξη μεταξύ ενός ατόμου υδρογόνου, το οποίο είναι ομοιοπολικά συνδεδεμένο με ένα άτομο υψηλής ηλεκτραρνητικότητας (όπως το οξυγόνο και το άζωτο), με ένα άλλο ηλεκτραρνητικό άτομο των ίδιων ή διαφορετικών μορίων της κοντινής τους γειτονιάς.
- Το υδρογόνο που υπάρχει στην ομάδα-OH του –NH2 των αμινοξέων γίνεται ελαφρώς ηλεκτροθετικό.
- Αυτό οφείλεται στην υψηλή ηλεκτραρνητικότητα του O και του N σε σύγκριση με το υδρογόνο.
- Λόγω της υψηλής ηλεκτραρνητικότητας, το οξυγόνο και το άζωτο προσελκύουν περισσότερο προς το μέρος τους το κοινό ηλεκτρόνιο του υδρογόνου.
- Έτσι το υδρογόνο που συνδέεται με αυτά τα υψηλά ηλεκτραρνητικά άτομα θα λάβει ένα μερικό θετικό φορτίο που ονομάζεται **δ** θετικό ενώ τα ηλεκτραρνητικά άτομα θα πάρουν ένα μερικό αρνητικό φορτίο που ονομάζεται **δ** αρνητικό.

Co-funded by the Erasmus+ Programme of the European Union

Δεσμοί υδρογόνου

- Κατά συνέπεια, το ελαφρώς θετικό Η έλκεται στη συνέχεια προς το γειτονικό ηλεκτραρνητικό οξυγόνο της ομάδας –C=O ή ατόμου αζώτου –NH2.
- Αυτές οι ομάδες –C=O και NH2 εμφανίζονται κατά μήκος της πολυπεπτιδικής αλυσίδας με κανονική σειρά.
- Έτσι ο σχηματισμός δεσμών υδρογόνου δίνει ένα κανονικό σχήμα στην πολυπεπτιδική αλυσίδα όπως οι πλάκες άλφα έλικας και βήτα.
- Οι δεσμοί υδρογόνου είναι πολύ αδύναμοι δεσμοί.
- Η εμφάνιση δεσμών υδρογόνου σε υψηλή συχνότητα συμβάλλει σημαντικά στη μοριακή σταθερότητα των πρωτεϊνών.
- Οι δεσμοί υδρογόνου εμπλέκονται στη σταθεροποίηση της δευτεροταγούς, τριτοταγούς και τεταρτοταγούς δομής των πρωτεϊνών.

Δεσμοί υδρογόνου



Υδροφοβικές Αλληλεπιδράσεις

- Ορισμένες ομάδες R στα αμινοξέα είναι μη πολικές.
- Παράδειγμα: Αλανίνη, βαλίνη, ισολευκίνη, λευκίνη και μεθειονίνη.
- Οι μη πολικές ομάδες R είναι υδρόφοβες και προσπαθούν να μείνουν μακριά από το νερό.
- Σε μια μακρά πολυπεπτιδική αλυσίδα μπορεί να υπάρχουν πολλά τέτοια μη πολικά αμινοξέα τα οποία μπορεί να είναι γειτονικά μεταξύ τους ή να διαχωρίζονται από πολικές ομάδες R.
- Σε ένα υδατικό περιβάλλον (μέσα στο κύτταρο) το γραμμικό πολυπεπτίδιο θα διπλωθεί σε τέτοιο σχήμα που τα υδρόφοβα αμινοξέα έρχονται σε στενή επαφή μεταξύ τους και προσπαθούν να αποκλείσουν το νερό λόγω της υδροφοβικότητας του.
- Με αυτή τη μέθοδο, η πεπτιδική αλυσίδα μιας σφαιρικής πρωτεΐνης θα διπλωθεί σε σφαιρικό σχήμα στο υδατικό περιβάλλον.

Co-funded by the Erasmus+ Programme of the European Union

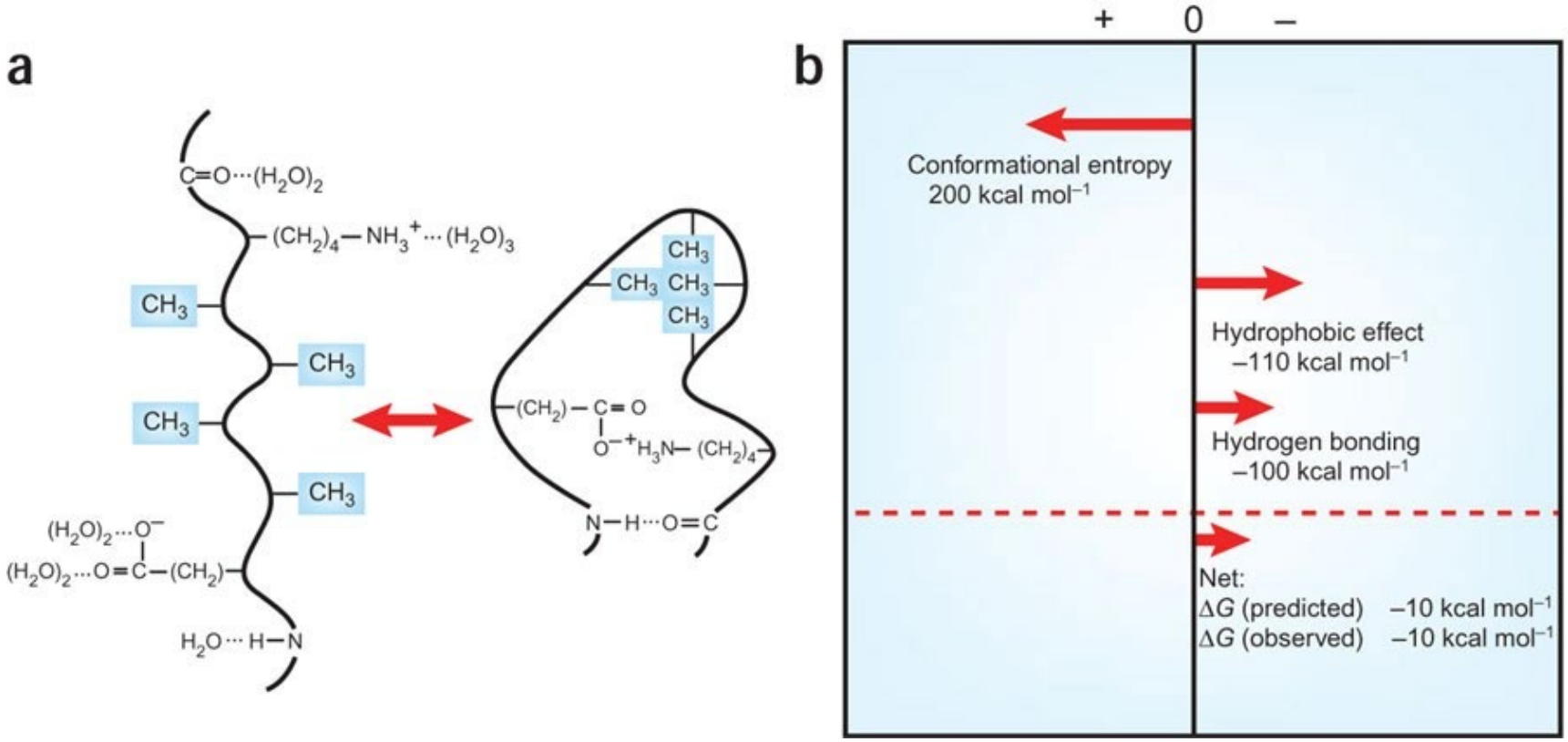
Υδροφοβικές Αλληλεπιδράσεις

- Σε μια διπλωμένη πρωτεΐνη οι υδρόφοβες ομάδες τείνουν να προσανατολίζονται προς την εσωτερική πλευρά της πρωτεΐνης.
- Τα υδρόφιλα υπολείμματα θα σχηματίσουν ένα κέλυφος πάνω από τα υδρόφοβα υπολείμματα.
- Το υδρόφιλο κέλυφος καθιστά την πρωτεΐνη διαλυτή στο υδατικό περιβάλλον.
- Ομοίως, στη μοναδιαία μεμβράνη, οι προσανατολισμοί των πρωτεϊνών της μεμβράνης επηρεάζονται επίσης από τις υδρόφιλες και υδρόφοβες αλληλεπιδράσεις.
- Η υδρόφοβη περιοχή της πρωτεΐνης της μεμβράνης προσανατολίζεται προς το εξωτερικό της μεμβράνης ενώ η υδρόφοβη περιοχή θα προσανατολίζεται προς το εσωτερικό (στο τμήμα λιπιδίων).

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

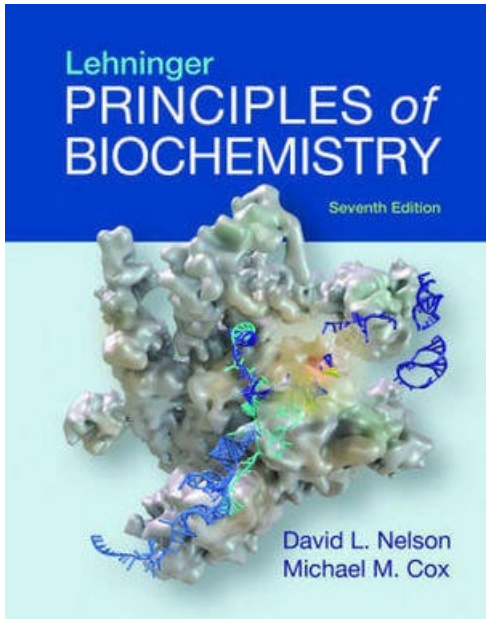
Υδροφοβικές Αλληλεπιδράσεις

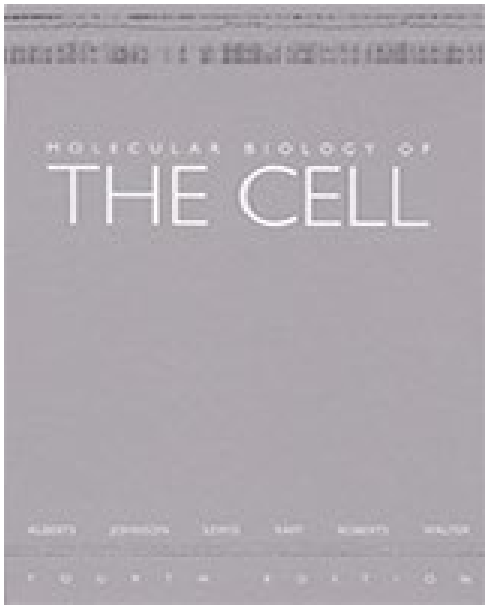


Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε το ρόλο των διαφορετικών αλληλεπιδράσεων του σχηματισμού δομής του μορίου πρωτεΐνης.
- Ποιος τύπος είναι πιο σημαντικός για το σχηματισμό της τριτογενούς δομής;





Lehninger , AL, Nelson, DL, & Cox, MM (2000). Αρχές βιοχημείας Lehninger . Νέα Υόρκη: Worth Publishers.

Μοριακή Βιολογία του Κυττάρου. 4η έκδοση.
Alberts B, Johnson A, Lewis J, et al. Νέα Υόρκη: Garland Science; 2002.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 3. Πρωτεΐνες και ένζυμα

Μάθημα 4. Λειτουργίες πρωτεϊνών

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Δομή
- Ένζυμα
- Ορμόνες
- Μεταφορά
- Αμυνα

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

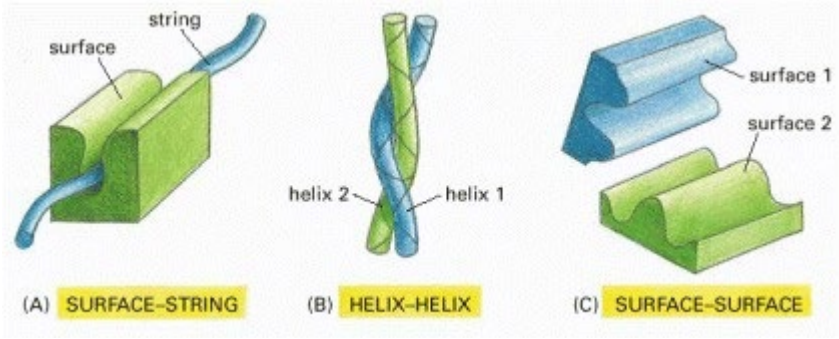
Οι πρωτεΐνες είναι οι «εργάτες» του σώματος και συμμετέχουν σε πολλές σωματικές λειτουργίες. Όπως έχουμε ήδη συζητήσει, οι πρωτεΐνες έρχονται σε όλα τα μεγέθη και σχήματα, και η καθεμία είναι ειδικά δομημένη για τη συγκεκριμένη λειτουργία της. Αυτή η σελίδα περιγράφει μερικές από τις σημαντικές λειτουργίες των πρωτεϊνών. Καθώς τα διαβάζετε, έχετε κατά νου ότι η σύνθεση όλων αυτών των διαφορετικών πρωτεϊνών απαιτεί επαρκείς ποσότητες αμινοξέων. Όπως μπορείτε να φανταστείτε, μια δίαιτα που είναι ανεπαρκής σε πρωτεΐνες και απαραίτητα αμινοξέα μπορεί να βλάψει πολλές από τις λειτουργίες του σώματος.

Τύποι και Λειτουργίες Πρωτεϊνών		
Τύπος	**Παραδείγματα**	**Λειτουργίες**
Δομή	Ακτίνη, μυοσίνη, κολλαγόνο, ελαστίνη, κερατίνη	Δίνει στους ιστούς (οστά, τένοντες, συνδέσμους, χόνδρο, δέρμα, μύες) δύναμη και δομή
Ένζυμα	Αμυλάση, λιπάση, πεψίνη, λακτάση	Χωνεύει τα μακροθρεπτικά συστατικά σε μικρότερα μονομερή που μπορούν να απορροφηθούν. Εκτελεί βήματα στις μεταβολικές οδούς για να επιτρέψει τη χρήση θρεπτικών συστατικών
ορμόνες	Ινσουλίνη, γλυκαγόνη, θυροξίνη	Χημικοί αγγελιοφόροι που ταξιδεύουν στο αίμα και συντονίζουν τις διαδικασίες γύρω από το σώμα
Ισορροπία υγρών και οξέος-βάσης	Λευκωματίνη, αιμοσφαιρίνη	Διατηρεί την κατάλληλη ισορροπία υγρών και pH σε διαφορετικά διαμερίσματα του σώματος
Μεταφορά	Αιμοσφαιρίνη, λευκωματίνη, πρωτεϊνικά κανάλια, πρωτεΐνες φορέα	Μεταφέρουν ουσίες σε όλο το σώμα στο αίμα ή τη λέμφο. Βοηθούν τα μόρια να διασχίζουν τις κυτταρικές μεμβράνες
Αμυνα	Κολλαγόνο, λυσοζύμη, αντισώματα	Προστατεύει το σώμα από ξένα παθογόνα

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

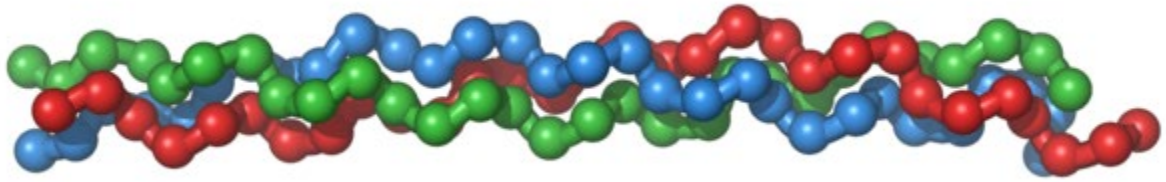
Οι πρωτεΐνες συνδέονται με άλλες πρωτεΐνες μέσω αρκετών τύπων διεπαφών



Τρεις τρόποι με τους οποίους δύο πρωτεΐνες μπορούν να συνδεθούν μεταξύ τους

Δείχνονται μόνο τα αλληλεπιδρώντα μέρη των δύο πρωτεϊνών. (Α) Μια άκαμπτη επιφάνεια σε μια πρωτεΐνη μπορεί να συνδεθεί σε έναν εκτεταμένο βρόχο πολυπεπτιδικής αλυσίδας (μια «χορδή») σε μια δεύτερη πρωτεΐνη. (Β) Δύο α έλικες μπορούν να συνδεθούν μεταξύ τους για να σχηματίσουν ένα τυλιγμένο πηνίο. (Γ) Δύο συμπληρωματικές άκαμπτες επιφάνειες συχνά συνδέουν δύο πρωτεΐνες μεταξύ τους.

ΔΟΜΗ

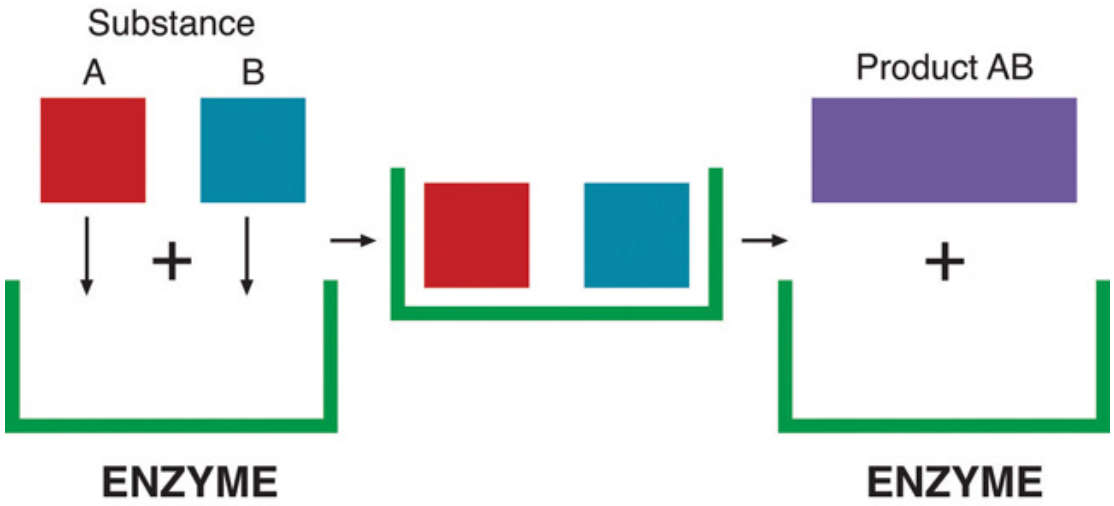


Περισσότερες από εκατό διαφορετικές δομικές πρωτεΐνες έχουν ανακαλυφθεί στο ανθρώπινο σώμα, αλλά η πιο άφθονη με διαφορά είναι το κολλαγόνο, το οποίο αποτελεί περίπου το 6 τοις εκατό του συνολικού σωματικού βάρους. Το κολλαγόνο αποτελεί το 30 τοις εκατό του οστικού ιστού και περιλαμβάνει μεγάλες ποσότητες τενόντων, συνδέσμων, χόνδρων, δέρματος και μυών. Το κολλαγόνο είναι μια ισχυρή, ινώδης πρωτεΐνη που αποτελείται κυρίως από γλυκίνη και αμινοξέα προλίνη. Μέσα στην τεταρτοταγή δομή του, τρεις πρωτεϊνικοί κλώνοι συστρέφονται ο ένας γύρω από τον άλλο σαν σχοινί και στη συνέχεια αυτά τα σχοινιά κολλαγόνου επικαλύπτονται με άλλα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ένζυμα



Ορισμένες πρωτεΐνες λειτουργούν ως ένζυμα. Τα ένζυμα είναι πρωτεΐνες που διεξάγουν συγκεκριμένες χημικές αντιδράσεις. Η δουλειά ενός ενζύμου είναι να παρέχει μια θέση για μια χημική αντίδραση και να μειώνει την ποσότητα της ενέργειας και του χρόνου που χρειάζεται για να συμβεί αυτή η χημική αντίδραση. Αυτός είναι ο λόγος που τα ένζυμα ονομάζονται μερικές φορές καταλύτες. Κατά μέσο όρο, περισσότερες από εκατό χημικές αντιδράσεις συμβαίνουν στα κύτταρα κάθε δευτερόλεπτο και οι περισσότερες από αυτές απαιτούν ένζυμα. Μόνο το ήπαρ περιέχει πάνω από χίλια ενζυμικά συστήματα. Τα ένζυμα είναι συγκεκριμένα και θα χρησιμοποιούν μόνο συγκεκριμένα υποστρώματα που ταιριάζουν στην ενεργή τους θέση, παρόμοια με τον τρόπο που μια κλειδαριά μπορεί να ανοίξει μόνο με ένα συγκεκριμένο κλειδί. Σχεδόν κάθε χημική αντίδραση απαιτεί ένα συγκεκριμένο ένζυμο. Ευτυχώς, ένα ένζυμο μπορεί να εκπληρώσει τον ρόλο του ως καταλύτη ξανά και ξανά, αν και τελικά καταστρέφεται και ξαναχτίζεται. Όλες οι σωματικές λειτουργίες, συμπεριλαμβανομένης της διάσπασης των θρεπτικών συστατικών στο στομάχι και το λεπτό έντερο, η μετατροπή των θρεπτικών ουσιών σε μόρια που μπορεί να χρησιμοποιήσει ένα κύτταρο και η οικοδόμηση όλων των μακρομορίων, συμπεριλαμβανομένης της ίδιας της πρωτεΐνης, περιλαμβάνουν ένζυμα.

Ορμόνες

Οι πρωτεΐνες είναι υπεύθυνες για τη σύνθεση ορμονών. Οι ορμόνες είναι χημικοί αγγελιοφόροι που παράγονται σε ένα μέρος του σώματος και στη συνέχεια μεταφέρονται με το αίμα σε διαφορετικό μέρος του σώματος. Όταν η ορμόνη φτάσει στον ιστό-στόχο/μέρος του σώματος, μεταδίδει ένα μήνυμα για να ξεκινήσει μια συγκεκριμένη αντίδραση ή κυτταρική διαδικασία. Για παράδειγμα, αφού φάτε ένα γεύμα, τα επίπεδα γλυκόζης στο αίμα σας αυξάνονται. Σε απόκριση στην αυξημένη γλυκόζη στο αίμα, το πάγκρεας απελευθερώνει την ορμόνη ινσουλίνη. Η ινσουλίνη λέει στα κύτταρα του σώματος ότι η γλυκόζη είναι διαθέσιμη και να την πάρουν από το αίμα και να την αποθηκεύσουν ή να τη χρησιμοποιήσουν για την παραγωγή ενέργειας ή τη δημιουργία μακρομορίων. Μια κύρια λειτουργία των ορμονών είναι να ενεργοποιούν και να απενεργοποιούν τα ένζυμα, έτσι ορισμένες πρωτεΐνες μπορούν ακόμη και να ρυθμίσουν τη δράση άλλων πρωτεϊνών. Ενώ δεν παράγονται όλες οι ορμόνες από πρωτεΐνες, πολλές από αυτές είναι. Άλλα παραδείγματα ορμονών που παράγονται από πρωτεΐνες περιλαμβάνουν τη γλυκαγόνη, τη μελατονίνη και τη θυρεοειδική ορμόνη.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ορμόνες του θυρεοειδούς αδένα

Η θυρεοσφαιρίνη, οι δραστικές ομάδες της οποίας είναι δύο μόρια της ένωσης θυροξίνης που περιέχει ιώδιο, έχει μοριακό βάρος 670.000. Η θυρεοσφαιρίνη περιέχει επίσης θυροξίνη με δύο και τρία άτομα ιωδίου αντί για τέσσερα και τυροσίνη με ένα και δύο άτομα ιωδίου. Η ένεση της ορμόνης προκαλεί αύξηση του μεταβολισμού. Η έλλειψή του οδηγεί σε επιβράδυνση.

Μια άλλη ορμόνη, η καλσιτονίνη, η οποία μειώνει τα επίπεδα ασβεστίου στο αίμα, εμφανίζεται στον θυρεοειδή αδένα. Οι αλληλουχίες αμινοξέων της καλσιτονίνης από χοίρο, βόειο κρέας και σολομό διαφέρουν από την ανθρώπινη καλσιτονίνη σε ορισμένα αμινοξέα. Όλες, όμως, έχουν τις ημικυστίνες (C) και το προλιναμίδιο (P) στην ίδια θέση.

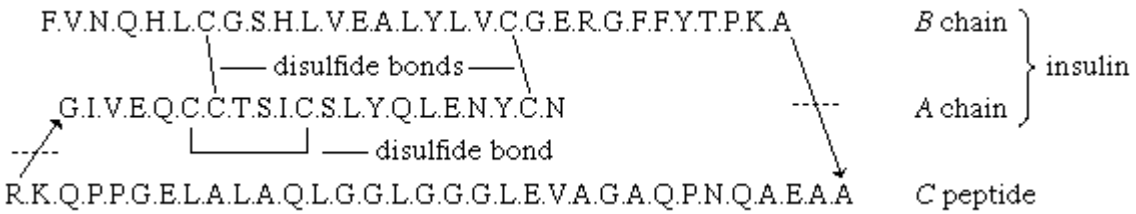
C.S.N.L.S.T.C.V.L.S.A.Y.W.K.D.L.N.N.Y.H.R.F.S.G.M.G.F.G.P.E.T.P(CONH$_2$)

Η αλληλουχία αμινοξέων της ανθρώπινης καλσιτονίνης. Στο αριστερό άκρο η γραμμή αντιπροσωπεύει τον δισουλφιδικό δεσμό. Στο δεξί άκρο (CONH $_2$) υποδηλώνει ότι η C τερματική προλίνη υπάρχει ως προλιναμίδη .

Η παραθορμόνη (παραθορμόνη), που παράγεται σε μικρούς αδένες που είναι ενσωματωμένοι ή βρίσκονται πίσω από τον θυρεοειδή αδένα, είναι απαραίτητη για τη διατήρηση των επιπέδων ασβεστίου στο αίμα. Η μείωση της παραγωγής του οδηγεί σε υπασβεστιαιμία (μείωση των επιπέδων ασβεστίου στην κυκλοφορία του αίματος κάτω από το φυσιολογικό εύρος). Η βόεια παραθορμόνη έχει μοριακό βάρος 8.500. Δεν περιέχει κυστίνη ή κυστεΐνη και είναι πλούσιο σε ασπαρτικό οξύ, γλουταμινικό οξύ ή τα αμίδια τους.

Ορμόνες του παγκρέατος

Αν και η δομή των αμινοξέων της ινσουλίνης είναι γνωστή από το 1949, επαναλαμβανόμενες προσπάθειες σύνθεσής της έδωσαν πολύ φτωχές αποδόσεις λόγω της αποτυχίας των δύο πεπτιδικών αλυσίδων να συνδυαστούν σχηματίζοντας τη σωστή δισουλφιδική γέφυρα. Η ευκολία της βιοσύνθεσης της ινσουλίνης εξηγείται από την ανακάλυψη στο πάγκρεας της προϊνσουλίνης, από την οποία σχηματίζεται η ινσουλίνη. Η απλή πεπτιδική αλυσίδα της προϊνσουλίνης χάνει ένα πεπτίδιο που αποτελείται από 33 αμινοξέα και ονομάζεται πεπτίδιο σύνδεσης, ή πεπτίδιο C, κατά τη μετατροπή του σε ινσουλίνη. Οι δισουλφιδικές γέφυρες της προϊνσουλίνης συνδέουν τις αλυσίδες Α και Β.



Ισοζύγιο υγρών και ηλεκτρολυτών

Η σωστή πρόσληψη πρωτεϊνών δίνει τη δυνατότητα στις βασικές βιολογικές διεργασίες του σώματος να διατηρήσουν το status quo σε ένα μεταβαλλόμενο περιβάλλον. Η ισορροπία υγρών αναφέρεται στη διατήρηση της κατανομής του νερού στο σώμα. Εάν πάρα πολύ νερό στο αίμα μετακινηθεί ξαφνικά σε έναν ιστό, τα αποτελέσματα είναι οίδημα και, ενδεχομένως, κυτταρικός θάνατος. Το νερό ρέει πάντα από μια περιοχή υψηλής συγκέντρωσης σε μια περιοχή χαμηλής συγκέντρωσης. Ως αποτέλεσμα, το νερό κινείται προς περιοχές που έχουν υψηλότερες συγκεντρώσεις άλλων διαλυμένων ουσιών, όπως πρωτεΐνες και γλυκόζη. Οι πρωτεΐνες προσελκύουν υγρά, επομένως για να διατηρηθεί το νερό ομοιόμορφα κατανεμημένο μεταξύ του αίματος και των κυττάρων, οι πρωτεΐνες κυκλοφορούν συνεχώς σε υψηλές συγκεντρώσεις στο αίμα. Όταν η πρόσληψη πρωτεΐνης είναι ελλιπής μπορεί να προκαλέσει οίδημα (πρήξιμο). Η πιο άφθονη πρωτεΐνη στο αίμα είναι η λευκωματίνη. Η παρουσία της λευκωματίνης στο αίμα κάνει τη συγκέντρωση πρωτεΐνης στο αίμα παρόμοια με αυτή στα κύτταρα. Επομένως, η ανταλλαγή υγρών μεταξύ του αίματος και των κυττάρων δεν είναι στα άκρα, αλλά μάλλον ελαχιστοποιείται για να διατηρηθεί το status quo. Οι πρωτεΐνες μεταφοράς (που συζητούνται παρακάτω) στην κυτταρική μεμβράνη βοηθούν στη διατήρηση της σωστής ισορροπίας των ηλεκτρολυτών (όπως το νάτριο και το κάλιο) μέσα και έξω από το κύτταρο.

Ισορροπία οξέος-βάσης (pH).

Η πρωτεΐνη είναι επίσης απαραίτητη για τη διατήρηση της σωστής ισορροπίας του pH (το μέτρο του πόσο όξινη ή βασική είναι μια ουσία) στο αίμα. Η κλίμακα pH κυμαίνεται από 0 (έντονα όξινο) έως 14 (έντονα βασικό/αλκαλικό). Το pH του αίματος διατηρείται μεταξύ 7,35 και 7,45, που είναι ελαφρώς βασικό. Εάν το αίμα γίνει πολύ όξινο (μια κατάσταση γνωστή ως οξέωση) σημαίνει ότι το επίπεδο του υδρογόνου (H+) στο αίμα είναι υπερβολικό. Εάν το αίμα γίνει πολύ βασικό/αλκαλικό (μια κατάσταση γνωστή ως αλκάλωση) σημαίνει ότι το επίπεδο του H+ στο αίμα είναι ανεπαρκές. Ακόμη και μια μικρή αλλαγή στο pH του αίματος μπορεί να επηρεάσει τις λειτουργίες του σώματος. Δύο παραδείγματα αυτού περιλαμβάνουν:

Όταν οι πρωτεΐνες εκτίθενται σε οξέα ή βάσεις, οι πρωτεΐνες αλλάζουν σχήμα και σταματούν να λειτουργούν όπως προβλέπεται. Αυτή η διαδικασία πρωτεϊνών που ξετυλίγονται και χάνουν το σχήμα και τη λειτουργία τους είναι γνωστή ως μετουσίωση. Η μετουσίωση των πρωτεϊνών συμβαίνει επίσης με την έκθεση σε θερμότητα, βαρέα μέταλλα, αλκοόλ και άλλες βλαβερές ουσίες.

Το όξινο αίμα (από κετοξέωση) μπορεί να οδηγήσει σε κώμα ή/και θάνατο σε ακραίες περιπτώσεις.

Το σώμα έχει πολλά συστήματα που διατηρούν το pH του αίματος εντός του φυσιολογικού εύρους για την πρόληψη προβλημάτων. Ορισμένες πρωτεΐνες λειτουργούν ως ρυθμιστικά διαλύματα και απελευθερώνουν υδρογόνο (H+) στο αίμα εάν γίνει πολύ βασικό. Οι πρωτεΐνες μπορούν επίσης να πάρουν υδρογόνο από το αίμα εάν γίνει πολύ όξινο. Με την απελευθέρωση και τη λήψη υδρογόνου όταν χρειάζεται, οι πρωτεΐνες διατηρούν την οξεοβασική ισορροπία και διατηρούν το pH του αίματος σε φυσιολογικά όρια.

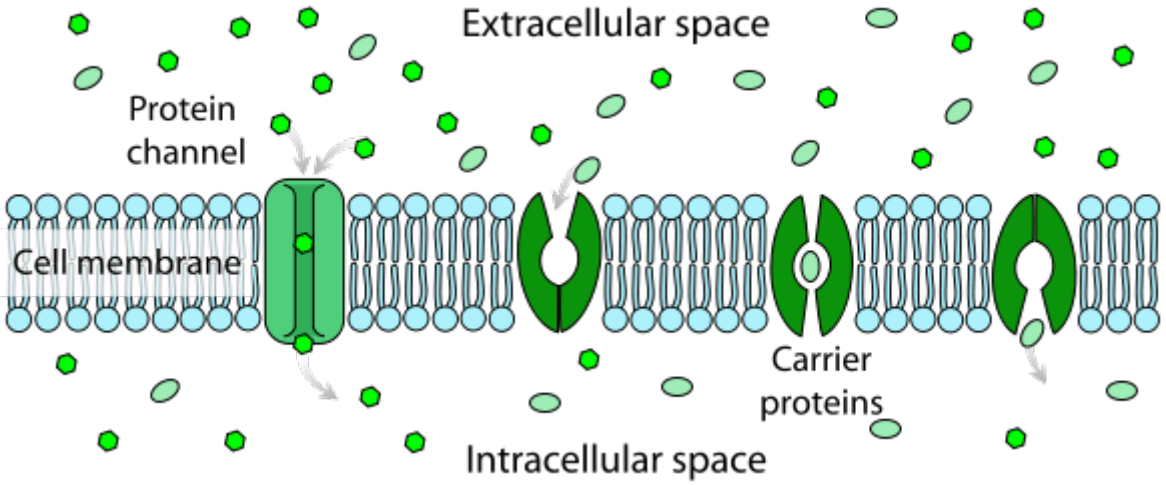
Co-funded by the Erasmus+ Programme of the European Union

Ισορροπία οξέος-βάσης (pH).

- Amino acids in protein buffer systems
- Depend on free and terminal amino acids
- Respond to pH changes by accepting or releasing H^+
- If pH rises:
  - carboxyl group of amino acid dissociates, acting as weak acid, releasing a hydrogen ion
- If pH drops:
  - carboxylate ion and amino group act as weak bases
  - accept H^+
  - form carboxyl group and amino ion

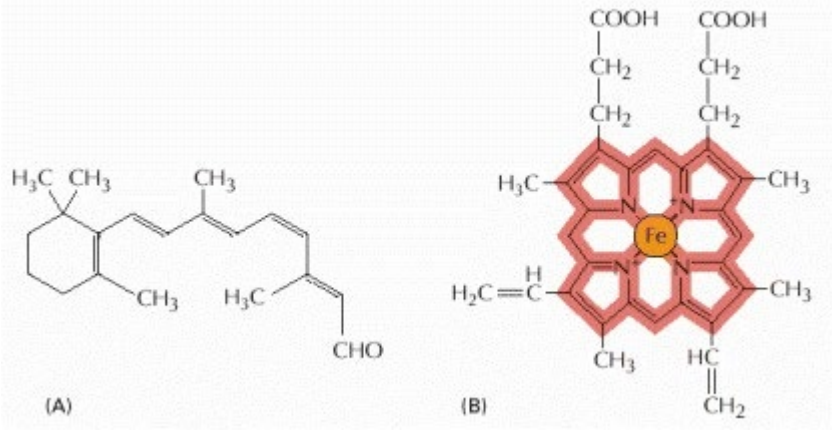
Co-funded by the Erasmus+ Programme of the European Union

Μεταφορά



Οι πρωτεΐνες παίζουν επίσης ρόλο στη μεταφορά θρεπτικών ουσιών. Η μεμβράνη ενός κυττάρου συνήθως δεν είναι διαπερατή από μεγάλα μόρια. Για να μπουν τα απαραίτητα θρεπτικά συστατικά και μόρια στο κύτταρο, υπάρχουν πολλές πρωτεΐνες μεταφοράς στην κυτταρική μεμβράνη. Μερικές από αυτές τις πρωτεΐνες λειτουργούν ως κανάλια που επιτρέπουν σε συγκεκριμένα μόρια να κινούνται μέσα και έξω από τα κύτταρα. Άλλα λειτουργούν ως ταξί μονής κατεύθυνσης και απαιτούν ενέργεια για να λειτουργήσουν.

Μικρά μόρια στενά συνδεδεμένα προσθέτουν επιπλέον λειτουργίες στις πρωτεΐνες

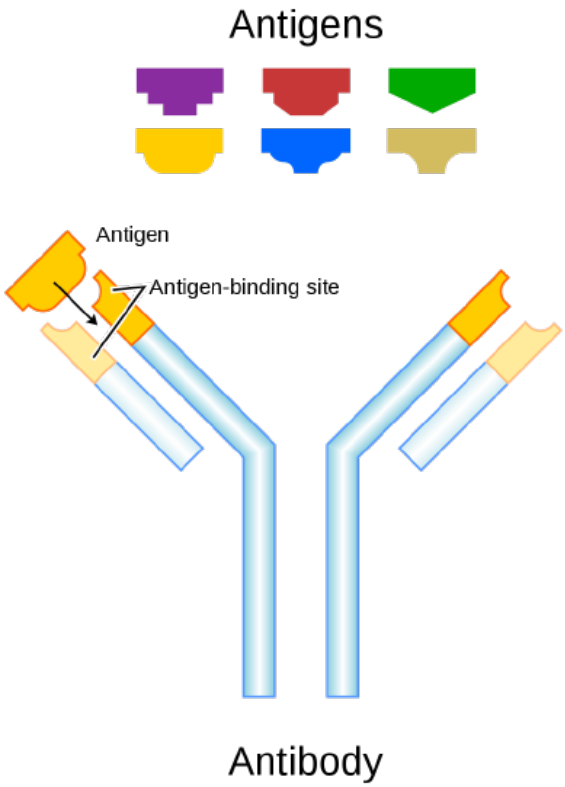


Αμφιβληστροειδής και αίμη

(Α) Η δομή του αμφιβληστροειδούς, του φωτοευαίσθητου μορίου που συνδέεται με τη ροδοψίνη στο μάτι. (Β) Η δομή μιας ομάδας αίμης. Ο δακτύλιος αίμης που περιέχει άνθρακα είναι κόκκινος και το άτομο σιδήρου στο κέντρο του είναι πορτοκαλί. Μια ομάδα αίμης είναι στενά συνδεδεμένη με καθεμία από τις τέσσερις πολυπεπτιδικές αλυσίδες στην αιμοσφαιρίνη, την πρωτεΐνη που μεταφέρει οξυγόνο

Co-funded by the Erasmus+ Programme of the European Union

Αντισώματα



Το ανοσοποιητικό μας σύστημα είναι σχεδιασμένο να επιτίθεται και να καταστρέφει ξένες ουσίες. Όταν μια ξένη ουσία επιτίθεται στο σώμα, το ανοσοποιητικό σύστημα παράγει αντισώματα για να αμυνθεί εναντίον του. Τα αντισώματα είναι ειδικές πρωτεΐνες που αναγνωρίζουν ένα μοναδικό μόριο σε επιβλαβή βακτήρια και ιούς, γνωστό ως αντιγόνο. Τα αντισώματα συνδέονται με το αντιγόνο και το καταστρέφουν. Τα αντισώματα ενεργοποιούν επίσης άλλους παράγοντες στο ανοσοποιητικό σύστημα για να αναζητήσουν και να καταστρέψουν ανεπιθύμητους εισβολείς.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Επούλωση Πληγών, Αναγέννηση Ιστών και Λειτουργία Νεύρων

Οι πρωτεΐνες εμπλέκονται σε όλες τις πτυχές της επούλωσης των πληγών, μια διαδικασία που λαμβάνει χώρα σε τρεις φάσεις: φλεγμονή, πολλαπλασιασμός και αναδιαμόρφωση. Για παράδειγμα, εάν δημιουργηθεί ένα μικρό κόψιμο, το δέρμα σας θα κοκκινίσει και θα φλεγμονή. Η διαδικασία επούλωσης ξεκινά με πρωτεΐνες, όπως η βραδυκινίνη, οι οποίες διαστέλλουν τα αιμοφόρα αγγεία στο σημείο του τραυματισμού. Μια πρόσθετη πρωτεΐνη που ονομάζεται ινώδες βοηθά στη στερέωση των αιμοπεταλίων που σχηματίζουν θρόμβο για να σταματήσει η αιμορραγία. Στη συνέχεια, στη φάση του πολλαπλασιασμού, τα κύτταρα κινούνται μέσα και επιδιορθώνουν τον τραυματισμένο ιστό εγκαθιστώντας νέες ίνες κολλαγόνου (πρωτεΐνης). Οι ίνες κολλαγόνου βοηθούν να τραβήξουν τις άκρες του τραύματος μαζί. Στη φάση της αναδιαμόρφωσης, εναποτίθεται περισσότερο κολλαγόνο, σχηματίζοντας μια ουλή. Ο ουλώδης ιστός είναι μόνο περίπου το 80 τοις εκατό τόσο λειτουργικός όσο ο κανονικός μη τραυματισμένος ιστός. Εάν μια δίαιτα είναι ανεπαρκής σε πρωτεΐνη, η διαδικασία επούλωσης των πληγών επιβραδύνεται σημαντικά.

Επούλωση Πληγών, Αναγέννηση Ιστών και Λειτουργία Νεύρων

Ενώ η επούλωση του τραύματος λαμβάνει χώρα μόνο μετά από τραυματισμό, μια διαφορετική διαδικασία που ονομάζεται αναγέννηση ιστού βρίσκεται σε εξέλιξη στο σώμα. Κατά την αναγέννηση των ιστών δημιουργείται ένα ακριβές δομικό και λειτουργικό αντίγραφο του παλιού ιστού. Τελικά, ο παλιός ιστός αντικαθίσταται με ολοκαίνουργιο, πλήρως λειτουργικό χαρτομάντιλο. Τα κύτταρα διασπώνται συνεχώς, επισκευάζονται και αντικαθίστανται. Όταν οι πρωτεΐνες στα κύτταρα διασπώνται, τα αμινοξέα ανακυκλώνονται σε νέες πρωτεΐνες. Ορισμένα κύτταρα (όπως το δέρμα, τα μαλλιά, τα νύχια και τα εντερικά κύτταρα) έχουν πολύ υψηλό ρυθμό αναγέννησης, ενώ άλλα, (όπως τα μυϊκά κύτταρα της καρδιάς και τα νευρικά κύτταρα) δεν αναγεννώνται σε σημαντικά επίπεδα. Η αναγέννηση των ιστών είναι η δημιουργία νέων κυττάρων (κυτταρική διαίρεση), η οποία απαιτεί πολλές διαφορετικές πρωτεΐνες, συμπεριλαμβανομένων των ενζύμων, των πρωτεϊνών μεταφοράς, των ορμονών και του κολλαγόνου. Τα κύτταρα που καλύπτουν το έντερο αναγεννώνται κάθε τρεις έως πέντε ημέρες. Οι δίαιτες ανεπαρκείς σε πρωτεΐνες βλάπτουν την αναγέννηση των ιστών, προκαλώντας πολλά προβλήματα υγείας, συμπεριλαμβανομένης της διαταραχής της πέψης και της απορρόφησης των θρεπτικών συστατικών.

Τα αμινοξέα μπορούν να χρησιμοποιηθούν για την κατασκευή νευροδιαβιβαστών (π.χ. επινεφρίνη) που μεταδίδουν μηνύματα από το ένα νευρικό κύτταρο στο άλλο.

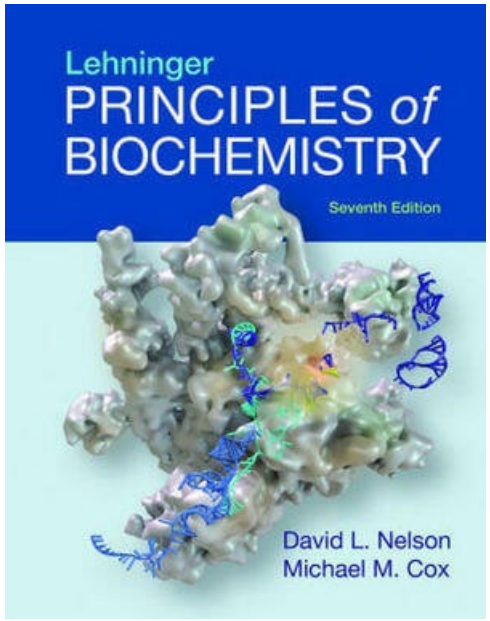
ΠΗΓΗ ΕΝΕΡΓΕΙΑΣ

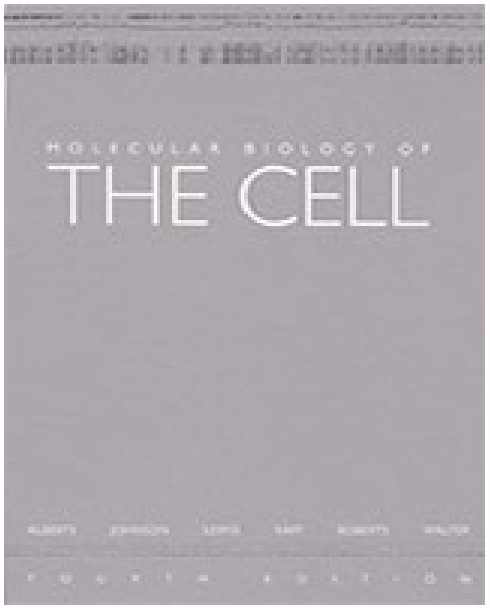
Μερικά από τα αμινοξέα στις πρωτεΐνες μπορούν να αποσυναρμολογηθούν και να χρησιμοποιηθούν για την παραγωγή ενέργειας. Σε υγιείς ανθρώπους, η πρωτεΐνη συμβάλλει ελάχιστα στις ενεργειακές ανάγκες. Εάν η διατροφή ενός ατόμου δεν περιέχει αρκετούς υδατάνθρακες και λίπη, το σώμα του θα χρησιμοποιήσει αμινοξέα για να παράγει ενέργεια. Όταν οι πρωτεΐνες χρειάζονται για ενέργεια, λαμβάνονται από το αίμα και τους ιστούς του σώματος (π.χ. μυς). Για να χρησιμοποιηθούν πρωτεΐνες για ενέργεια, απαιτείται απαμίνωση. Η απαμίνωση είναι μια διαδικασία όπου η ομάδα αμίνης απομακρύνεται από το αμινοξύ και το άζωτο μεταφέρεται στο νεφρό για απέκκριση. Τα υπόλοιπα συστατικά μεταβολίζονται για ενέργεια. Για να προστατεύσουμε τους ιστούς του σώματός μας από τη διάσπαση για ενέργεια, είναι σημαντικό να τρώμε επαρκή ποσότητα λίπους και υδατανθράκων. Είναι επίσης σημαντικό να σημειωθεί ότι το σώμα μας δεν μπορεί να αποθηκεύσει περίσσεια πρωτεΐνης. Η υπερβολική πρόσληψη πρωτεΐνης έχει ως αποτέλεσμα την απέκκριση αζώτου. Τα υπόλοιπα συστατικά χρησιμοποιούνται για ενέργεια ή μετατρέπονται σε λίπος για μελλοντική χρήση.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε τις κύριες λειτουργίες των πρωτεϊνών.
- Δώστε μερικά παραδείγματα δομικών πρωτεϊνών, ορισμένων ορμονών.





Lehninger , AL, Nelson, DL, & Cox, MM (2000). Αρχές βιοχημείας Lehninger . Νέα Υόρκη: Worth Publishers.

Μοριακή Βιολογία του Κυττάρου. 4η έκδοση.
Alberts B, Johnson A, Lewis J, et al. Νέα Υόρκη: Garland Science; 2002.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 3. Πρωτεΐνες και ένζυμα

Μάθημα 5. Μηχανισμός ενζυμικής κατάλυσης

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Ορισμός ενζύμου
- Ομοιοπολική Κατάλυση
- Οξεοβασική Κατάλυση
- Ηλεκτροστατική Κατάλυση
- Υδροφοβικοί δεσμοί

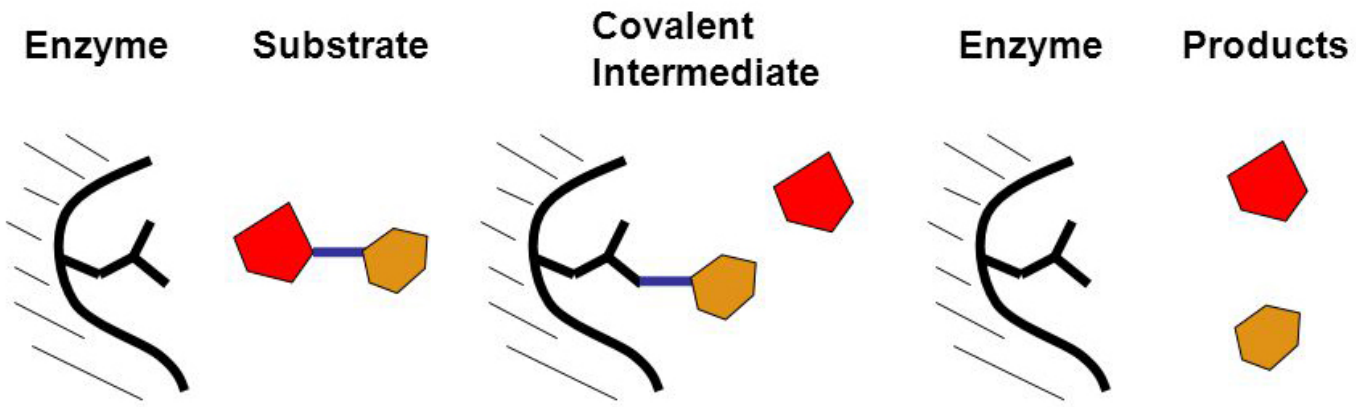
Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Η ενζυματική κατάλυση των αντιδράσεων είναι απαραίτητη για τα ζωντανά συστήματα. Υπό βιολογικά σχετικές συνθήκες, οι μη καταλυόμενες αντιδράσεις τείνουν να είναι αργές - τα περισσότερα βιολογικά μόρια είναι αρκετά σταθερά στο υδατικό περιβάλλον με ουδέτερο pH, ήπιας θερμοκρασίας, μέσα στα κύτταρα. Επιπλέον, πολλές κοινές χημικές διεργασίες είναι δυσμενείς ή απίθανες στο κυτταρικό περιβάλλον, όπως ο παροδικός σχηματισμός ασταθών φορτισμένων ενδιαμέσων ή η σύγκρουση δύο ή περισσότερων μορίων στον ακριβή προσανατολισμό που απαιτείται για την αντίδραση.

Co-funded by the Erasmus+ Programme of the European Union

Ομοιοπολική Κατάλυση

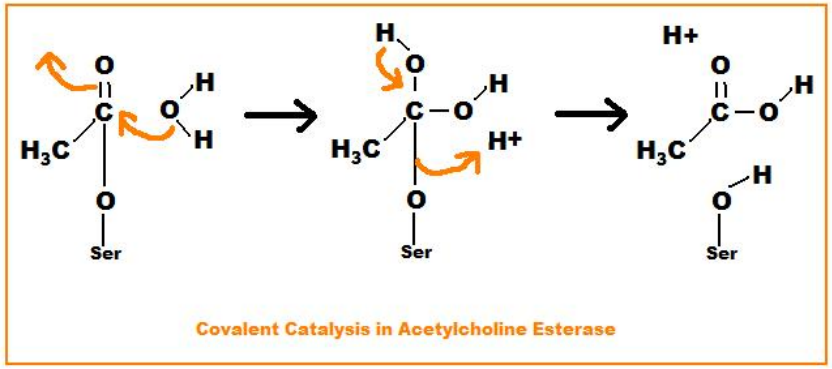


Η ομοιοπολική κατάλυση περιλαμβάνει το σχηματισμό ενός ομοιοπολικού δεσμού μεταξύ του ενζύμου και τουλάχιστον ενός από τα υποστρώματα που εμπλέκονται στην αντίδραση. Συχνά αυτό περιλαμβάνει πυρηνόφιλη κατάλυση που είναι μια υποκατηγορία ομοιοπολικής κατάλυσης. Αρκετές ομάδες αμινοξέων R μπορούν να χρησιμεύσουν ως πυρηνόφιλα και συχνά βρίσκονται στην ενεργό θέση των ενζύμων. Οι πυρηνόφιλες πλευρικές αλυσίδες συχνά ενεργοποιούνται από την αποπρωτονίωση που προκαλείται από γειτονικές πλευρικές αλυσίδες, όπως η ιστιδίνη που μπορεί να λειτουργήσει ως βάση. Εναλλακτικά, το νερό μπορεί επίσης να ενεργοποιήσει το πυρηνόφιλο. Ο ενδιάμεσος σχηματισμός ομοιοπολικού δεσμού μεταξύ του ενζύμου και του υποστρώματος επιτρέπει τη διάσπαση του δεσμού και την απομάκρυνση μιας αποχωρούσας ομάδας

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

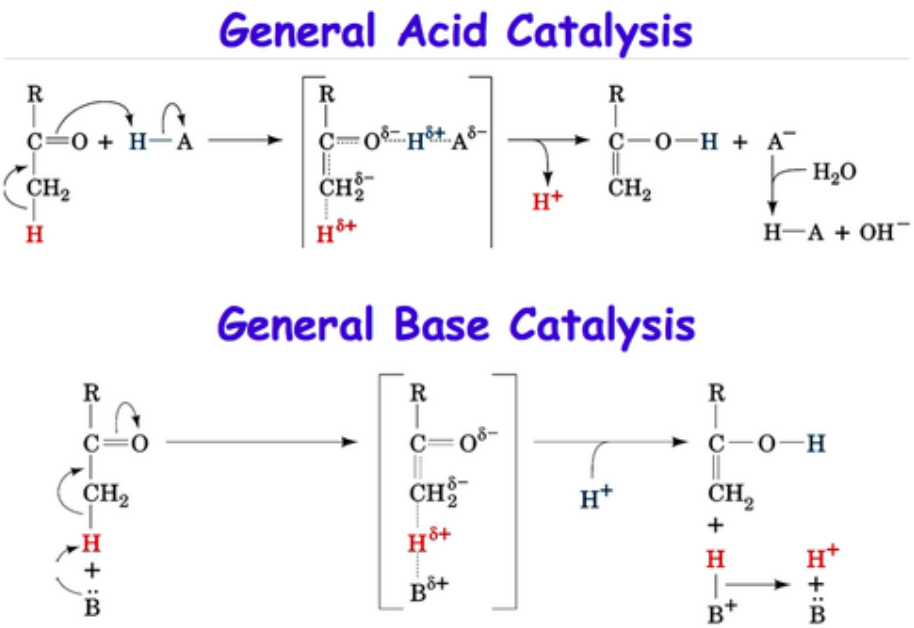
Co-funded by the Erasmus+ Programme of the European Union

Παραδείγματα ενζύμων που συμμετέχουν στην ομοιοπολική κατάλυση



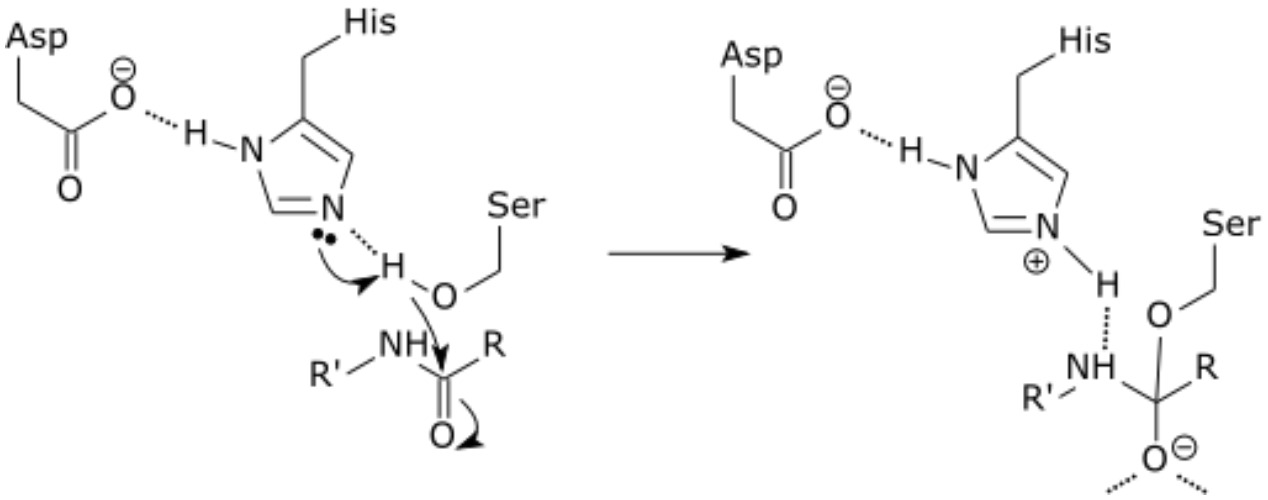
Παραδείγματα ενζύμων που συμμετέχουν στην ομοιοπολική κατάλυση περιλαμβάνουν το πρωτεολυτικό ένζυμο Χυμοτρυψίνη και θρυψίνη στα οποία το πυρηνόφιλο είναι η υδροξυλική ομάδα στη σερίνη. Η χυμοθρυψίνη είναι μια αποικοδομητική πρωτεάση του πεπτικού συστήματος. Καταλύει τη διάσπαση των πεπτιδικών δεσμών που βρίσκονται δίπλα σε μεγάλα αρωματικά ή μη πολικά υπολείμματα. Διασπά τον πεπτιδικό δεσμό στην καρβοξυλική πλευρά της πρωτεΐνης. Η χυμοθρυψίνη έχει τρία κύρια καταλυτικά υπολείμματα που ονομάζονται καταλυτική τριάδα. Αυτά είναι τα His 57, Asp 102 και Ser 195. Κατά την αποπρωτονίωση το υπόλειμμα σερίνης γίνεται ισχυρό πυρηνόφιλο λόγω του αλκοξειδίου του που θα επιτεθεί στον σχετικά μη αντιδραστικό άνθρακα του καρβονυλίου στην πρωτεΐνη. Τυπικά υπολείμματα που χρησιμοποιούνται στην ομοιοπολική κατάλυση είναι τα Lys, His, Cys , Asp, Glu και Ser και μερικά άλλα συνένζυμα.

Οξεοβασική Κατάλυση



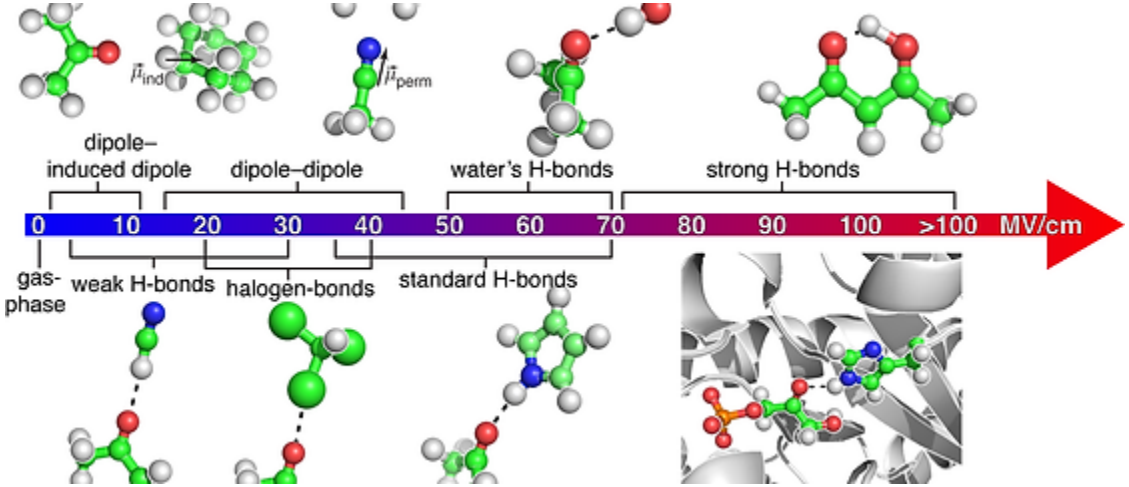
Η κατάλυση οξέος-βάσης εμπλέκεται σε οποιονδήποτε μηχανισμό αντίδρασης που απαιτεί τη μεταφορά ενός πρωτονίου από το ένα μόριο στο άλλο. Είναι πολύ συνηθισμένο να βλέπουμε αυτόν τον μηχανισμό να συνδυάζεται με την Ομοιοπολική Κατάλυση καθώς πολλά πυρηνόφιλα ενεργοποιούνται με την απομάκρυνση ενός πρωτονίου, συμπεριλαμβανομένων των λειτουργικών ομάδων αλκοόλης, θειόλης και αμίνης. Τα ένζυμα που χρησιμοποιούν την κατάλυση οξέος-βάσης μπορούν να υποομαδοποιηθούν περαιτέρω είτε σε συγκεκριμένες αντιδράσεις οξέος-βάσης είτε σε γενικές αντιδράσεις οξέος-βάσης. Η ειδική όξινη ή ειδική κατάλυση βάσης συμβαίνει εάν ένα ιόν υδρονίου (H3O+) ή ένα ιόν υδροξειδίου (OH–), αντίστοιχα, χρησιμοποιούνται απευθείας στον μηχανισμό αντίδρασης και το pH του διαλύματος επηρεάζει τον ρυθμό κατάλυσης. Οι γενικές αντιδράσεις οξέος και γενικής βάσης συμβαίνουν όταν μόρια άλλα από το ιόν υδρονίου (H3O+) ή ένα ιόν υδροξειδίου (OH–) είναι η πηγή δωρεάς ή αποδοχής πρωτονίων. Συνηθέστερα, ένα υπόλειμμα αμινοξέος ενεργού θέσης χρησιμοποιείται για την αποδοχή ή τη δωρεά ενός πρωτονίου εντός του μηχανισμού αντίδρασης. Σε γενικές αντιδράσεις οξέος-βάσης το pH διατηρείται συνήθως σταθερό μέσα σε ένα ρυθμιστικό σύστημα.

Οξεοβασική Κατάλυση



Το αρχικό στάδιο του καταλυτικού μηχανισμού πρωτεάσης σερίνης περιλαμβάνει την ιστιδίνη της ενεργού θέσης που δέχεται ένα πρωτόνιο από το υπόλειμμα σερίνης. Αυτό προετοιμάζει τη σερίνη ως πυρηνόφιλο για να επιτεθεί στον αμιδικό δεσμό του υποστρώματος. Αυτός ο μηχανισμός περιλαμβάνει τη δωρεά ενός πρωτονίου από τη σερίνη (μια βάση, pKa 14) σε ιστιδίνη (ένα οξύ, pKa 6), που έγινε δυνατή λόγω του τοπικού περιβάλλοντος των βάσεων.

Ηλεκτροστατική Κατάλυση

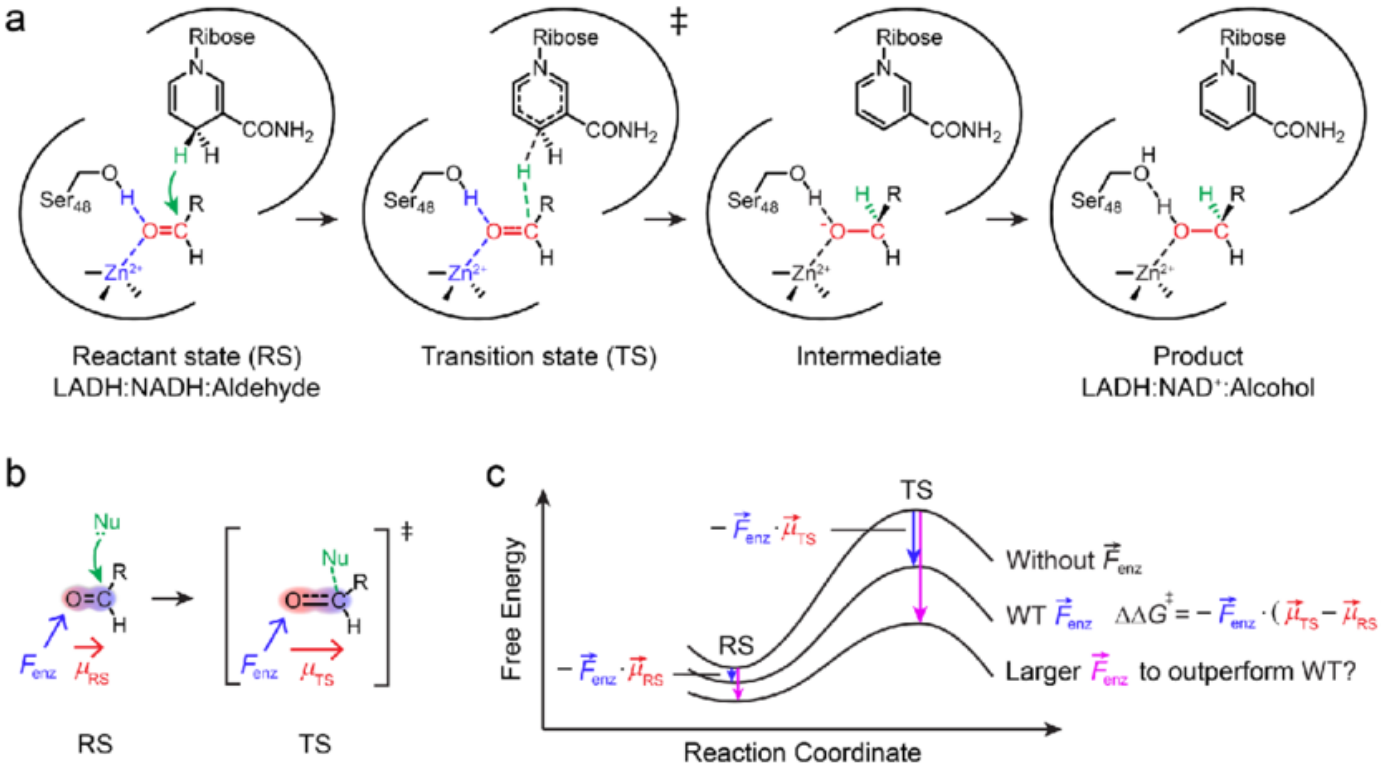


Η ηλεκτροστατική κατάλυση συμβαίνει όταν η ενεργή θέση του ενζύμου σταθεροποιεί τη μεταβατική κατάσταση της αντίδρασης σχηματίζοντας ηλεκτροστατικές αλληλεπιδράσεις με το υπόστρωμα. Οι ηλεκτροστατικές αλληλεπιδράσεις μπορεί να είναι αλληλεπιδράσεις ιοντικής, ιοντικής-διπόλου, διπόλου-διπόλου ή υδρόφοβων. Ο δεσμός υδρογόνου είναι μια από τις πιο κοινές ηλεκτροστατικές αλληλεπιδράσεις που σχηματίζονται στην ενεργό θέση.

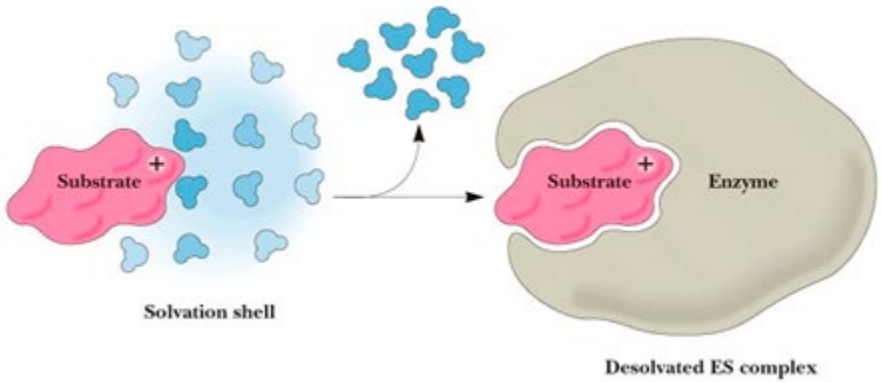
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ηλεκτροστατική Κατάλυση



Διαλυτοποίηση



Οι ενεργές θέσεις των ενζύμων μπορεί να στερούνται νερού και να μιμούνται τα χαρακτηριστικά αντίδρασης της αέριας φάσης. Αυτό μπορεί να αποσταθεροποιήσει την πολωμένη κατάσταση φορτισμένων ομάδων όπως τα οξέα και οι βάσεις. Έτσι, η ουδέτερη μορφή αυτών των τύπων υπολειμμάτων γίνεται η ευνοούμενη κατάσταση. Αυτό οφείλεται σε σημαντικές αλλαγές στο pKa των υπολειμμάτων της ενεργού θέσης εντός του μη πολικού περιβάλλοντος. Αυτό μπορεί να προκαλέσει κανονικά όξινα υπολείμματα όπως το γλουταμικό να αφαιρέσουν ένα πρωτόνιο από την ιστιδίνη και να συμπεριφέρονται ως βάση, για παράδειγμα.

Co-funded by the Erasmus+ Programme of the European Union

Κατάλυση κατά προσέγγιση

Στην κατάλυση κατά προσέγγιση, το ένζυμο ενισχύει τον ρυθμό αντίδρασης δεσμεύοντας με πολλαπλά υποστρώματα και τοποθετώντας τα ευνοϊκά έτσι ώστε η αντίδραση να μπορεί να προχωρήσει. Η σύνδεση με το ένζυμο μειώνει την περιστροφική εντροπία των υποστρωμάτων που διαφορετικά θα επιπλέουν τυχαία ελεύθερα στο διάλυμα και επιτρέπει τη σωστή τοποθέτηση των υποστρωμάτων για την αντίδραση. Η απώλεια της εντροπίας, η οποία δεν είναι ευνοϊκή, αντισταθμίζεται από την ενέργεια δέσμευσης που απελευθερώνεται με την αλληλεπίδραση υποστρώματος-ενζύμου.

Εκτός από τη σωστή τοποθέτηση των υποστρωμάτων ώστε να αλληλεπιδρούν μεταξύ τους, η κατά προσέγγιση μετατρέπει μια αντίδραση που θα ήταν δεύτερης τάξης, με υποστρώματα που επιπλέουν ελεύθερα στο διάλυμα, σε μια αντίδραση πρώτης τάξης, όπου όλα τα υποστρώματα διατηρούνται στη θέση τους. από το ένζυμο και συμπεριφέρονται ως ένα μόνο μόριο. Αυτό μπορεί να βελτιώσει δραματικά τον καταλυτικό ρυθμό της αντίδρασης από 105 σε 107 φορές πιο γρήγορα, ανάλογα με το ενζυμικό σύστημα.

Co-funded by the Erasmus+ Programme of the European Union

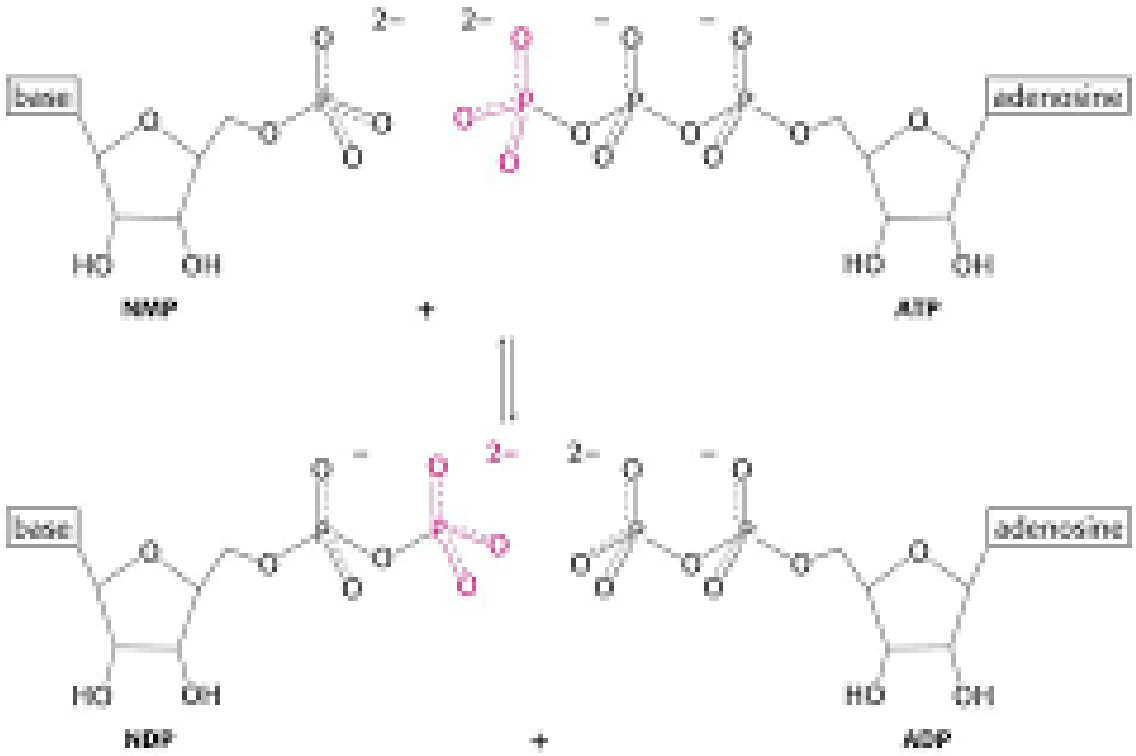
Κατάλυση κατά προσέγγιση

Στην κατάλυση κατά προσέγγιση, το ένζυμο ενισχύει τον ρυθμό αντίδρασης δεσμεύοντας με πολλαπλά υποστρώματα και τοποθετώντας τα ευνοϊκά έτσι ώστε η αντίδραση να μπορεί να προχωρήσει. Η σύνδεση με το ένζυμο μειώνει την περιστροφική εντροπία των υποστρωμάτων που διαφορετικά θα επιπλέουν τυχαία ελεύθερα στο διάλυμα και επιτρέπει τη σωστή τοποθέτηση των υποστρωμάτων για την αντίδραση. Η απώλεια της εντροπίας, η οποία δεν είναι ευνοϊκή, αντισταθμίζεται από την ενέργεια δέσμευσης που απελευθερώνεται με την αλληλεπίδραση υποστρώματος-ενζύμου.

Εκτός από τη σωστή τοποθέτηση των υποστρωμάτων ώστε να αλληλεπιδρούν μεταξύ τους, η κατά προσέγγιση μετατρέπει μια αντίδραση που θα ήταν δεύτερης τάξης, με υποστρώματα που επιπλέουν ελεύθερα στο διάλυμα, σε μια αντίδραση πρώτης τάξης, όπου όλα τα υποστρώματα διατηρούνται στη θέση τους. από το ένζυμο και συμπεριφέρονται ως ένα μόνο μόριο. Αυτό μπορεί να βελτιώσει δραματικά τον καταλυτικό ρυθμό της αντίδρασης από 105 σε 107 φορές πιο γρήγορα, ανάλογα με το ενζυμικό σύστημα.

Co-funded by the Erasmus+ Programme of the European Union

Κατάλυση κατά προσέγγιση



Co-funded by the Erasmus+ Programme of the European Union

Παραμόρφωση στελέχους

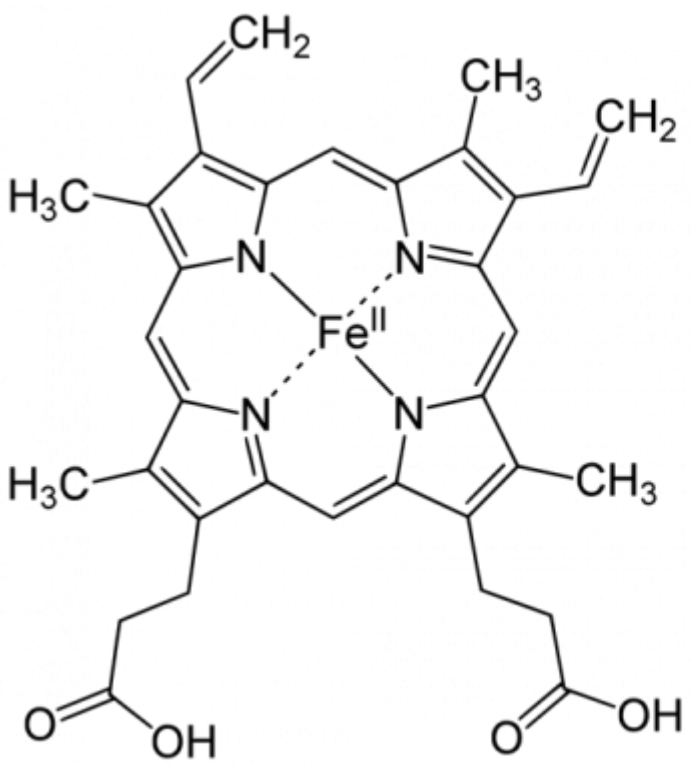
Στην οργανική χημεία, μάθατε ότι ορισμένες δομές όπως οι τριμελείς και τετραμελείς δομές δακτυλίου, όπως τα εποξείδια, ήταν εξαιρετικά αντιδραστικές λόγω της παραμόρφωσης της τάσης που είναι εγγενής στις μη ευνοϊκές γωνίες δεσμού που είναι εγγενείς στον δακτύλιο. Οι ενεργές θέσεις των ενζύμων μπορούν επίσης να χρησιμοποιήσουν την παραμόρφωση του στελέχους μέσα σε ένα δεσμευμένο υπόστρωμα για να αυξήσουν την αντιδραστικότητα του μορίου και να ευνοήσουν το σχηματισμό της μεταβατικής κατάστασης. Πολλά ένζυμα που λειτουργούν με το μοντέλο της επαγόμενης προσαρμογής χρησιμοποιούν επίσης παραμόρφωση παραμόρφωσης εντός του καταλυτικού τους μηχανισμού. Εντός της μη δεσμευμένης κατάστασης παραμένουν σε χαμηλή καταλυτική κατάσταση, ωστόσο η αλληλεπίδραση με το υπόστρωμα επάγει την αποσταθεροποίηση της ενεργού θέσης του ενζύμου ή μπορεί να προκαλέσει καταπόνηση εντός του υποστρώματος προκαλώντας την έναρξη της καταλυτικής δραστηριότητας του ενζύμου.

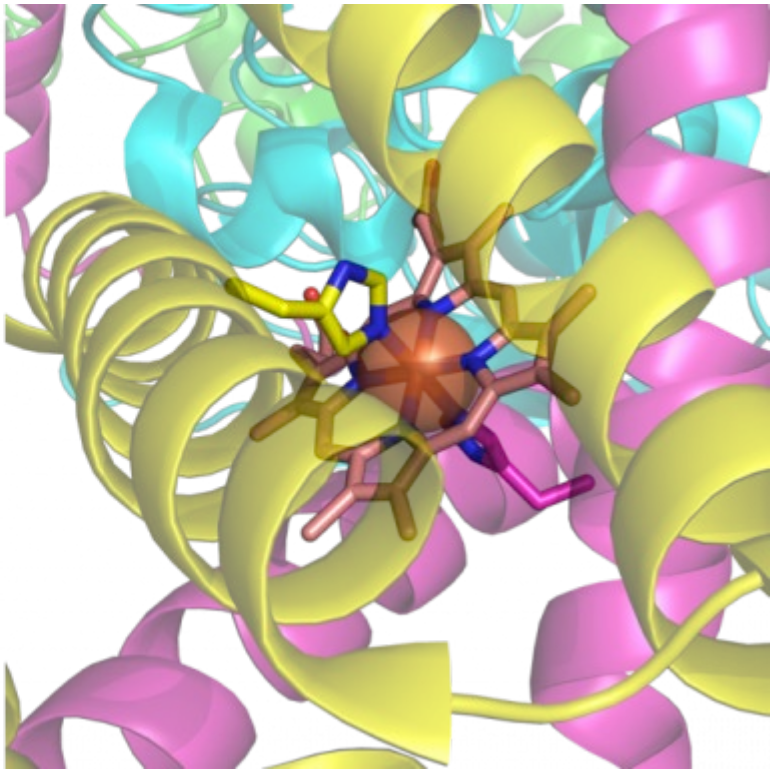
Co-funded by the Erasmus+ Programme of the European Union

Κατάλυση συμπαράγοντα

Οι συμπαράγοντες είναι μόρια που συνδέονται με ένζυμα και απαιτούνται για την καταλυτική δραστηριότητα του ενζύμου. Μπορούν να χωριστούν σε δύο μεγάλες κατηγορίες: μέταλλα και συνένζυμα. Οι συμπαράγοντες μετάλλων που βρίσκονται συνήθως στα ανθρώπινα ένζυμα περιλαμβάνουν: σίδηρο, μαγνήσιο, μαγγάνιο, κοβάλτιο, χαλκό, ψευδάργυρο και μολυβδαίνιο. Τα συνένζυμα είναι μικρά οργανικά μόρια που συχνά προέρχονται από βιταμίνες, οι οποίες είναι απαραίτητα οργανικά θρεπτικά συστατικά που καταναλώνονται στη διατροφή. Τα συνένζυμα μπορούν να συνδεθούν χαλαρά με το ένζυμο και να έχουν την ικανότητα να δεσμεύονται και να απελευθερώνονται από τη δραστική θέση ή μπορεί να συνδέονται στενά και να μην έχουν την ικανότητα να απελευθερώνονται εύκολα από το ένζυμο. Τα συνένζυμα στενής δέσμευσης αναφέρονται ως προσθετικές ομάδες. Τα ένζυμα που δεν συνδέονται ακόμη με έναν απαιτούμενο συμπαράγοντα ονομάζονται αποένζυμα, ενώ τα ένζυμα που συνδέονται με τους απαιτούμενους συμπαράγοντες ονομάζονται ολοένζυμα. Μερικές φορές οργανικά μόρια και μέταλλα συνδυάζονται για να σχηματίσουν συνένζυμα, όπως στην περίπτωση του συμπαράγοντα αίμης. Ο συντονισμός των συμπαραγόντων αίμης με τους ομολόγους τους ενζύμου συχνά περιλαμβάνει ηλεκτροστατικές αλληλεπιδράσεις με υπολείμματα ιστιδίνης όπως φαίνεται στο ένζυμο της ηλεκτρικής αφυδρογονάσης.

Κατάλυση συμπαράγοντα





Co-funded by the Erasmus+ Programme of the European Union

Κατάλυση συμπαράγοντα

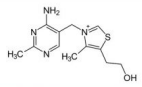
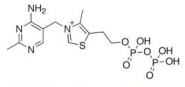
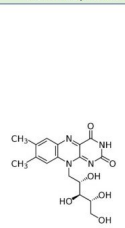
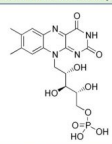
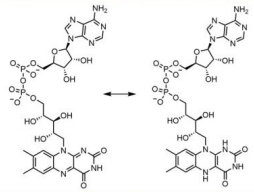
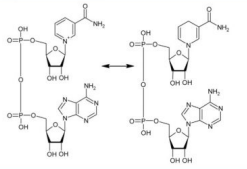
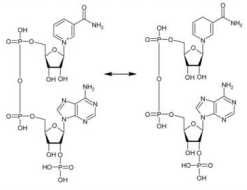
Ως παράδειγμα για το πώς οι βιταμίνες μπορούν να χρησιμοποιηθούν ως συμπαράγοντες, δείχνει τις κοινές βιταμίνες Β και τα συνένζυμα που προέρχονται από τις δομές τους. Πολλές ανεπάρκειες βιταμινών προκαλούν ασθένειες λόγω της αδράνειας των αποενζύμων που δεν μπορούν να λειτουργήσουν χωρίς το σωστά συνδεδεμένο συνένζυμο.

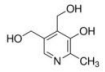
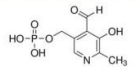
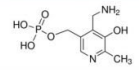
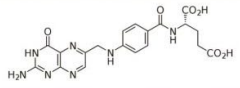
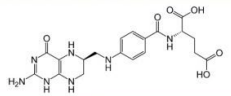
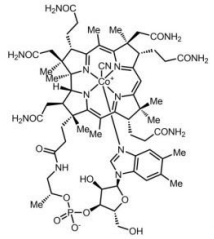
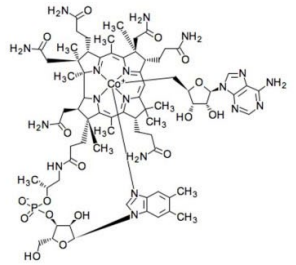
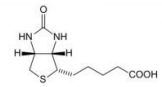
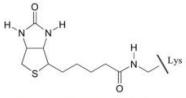
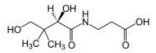
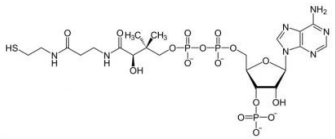
Οι συμπαράγοντες μπορούν να βοηθήσουν στη μεσολάβηση ενζυματικών αντιδράσεων μέσω της χρήσης οποιασδήποτε από τις διαφορετικές καταλυτικές στρατηγικές που αναφέρονται παραπάνω. Μπορούν να χρησιμεύσουν ως πυρηνόφιλα και να μεσολαβήσουν στην ομοιοπολική κατάλυση ή να σχηματίσουν ηλεκτροστατικές αλληλεπιδράσεις με το υπόστρωμα και να σταθεροποιήσουν τη μεταβατική κατάσταση. Μπορούν επίσης να προκαλέσουν παραμόρφωση παραμόρφωσης ή να διευκολύνουν την κατάλυση οξέος-βάσης. Η κατάλυση με τη βοήθεια μετάλλων μπορεί συχνά να χρησιμοποιήσει μηχανισμούς ομολυτικής αντίδρασης που περιλαμβάνουν ριζικά ενδιάμεσα. Αυτό μπορεί να είναι σημαντικό σε αντιδράσεις όπως αυτές που συμβαίνουν στην αλυσίδα μεταφοράς ηλεκτρονίων που απαιτούν την ασφαλή κίνηση μεμονωμένων ηλεκτρονίων.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Απαραίτητες βιταμίνες Β και οι τροποποιημένοι ενζυμικοί συμπαράγοντες τους

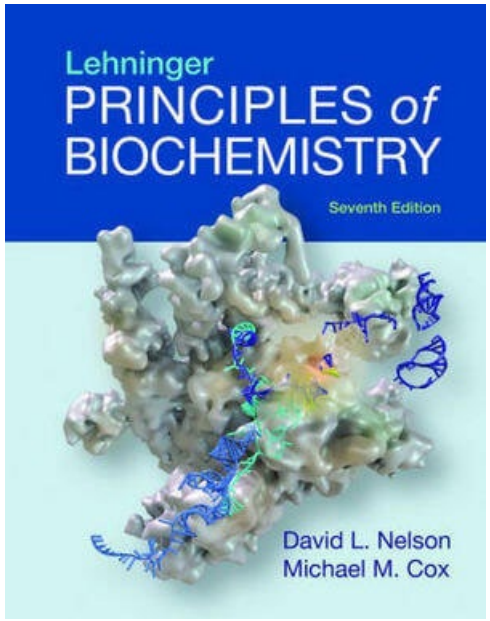
B Vitamins	Modified Enzyme Cofactors
Vitamin B1 (Thymine)	Thymine Diphosphate (TPP)
	
Vitamin B2 (Riboflavin)	Flavin Mononucleotide (FMN)
	
	Flavin Adenine Dinucleotide (FAD ⟷ $FADH_2$)
	
Vitamin B3 (Niacinamide)	Nicotinamide Adenine Dinucleotide (NAD^+ ⟷ NADH)
	
Vitamin B3 (Niacin – carboxylic acid form)	Nicotinamide Dinucleotide Phosphate ($NADP^+$ ⟷ NADPH)
	

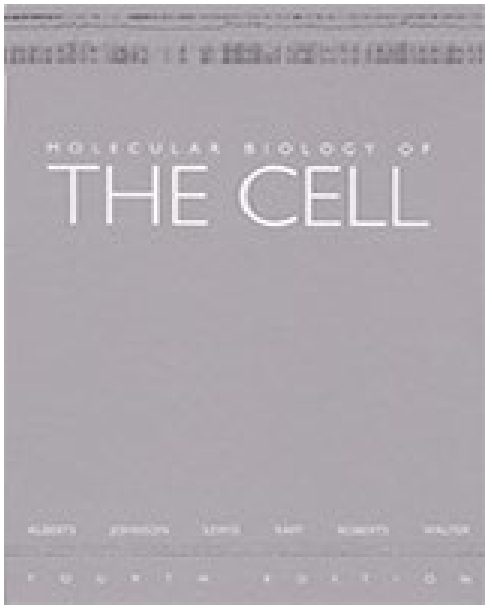
Vitamin B6 (Pyridoxine)	Pyridoxyl 5′-Phosphate
	
	Pyridoxamine 5′-Phosphate
	
Vitamin B9 (Folic Acid)	Tetrahydrofolate
	
Vitamin B12 (Cyanocobalamin)	Adenosylcobalamin
	
Biotin	Biotin-Enzyme Complex
	
Pantothenic Acid	Coenzyme A
	

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Ποιος είναι ο ορισμός του ενζύμου;
- Εξηγήστε τον μηχανισμό των καταλυτικών αντιδράσεων.





Lehninger , AL, Nelson, DL, & Cox, MM (2000). Αρχές βιοχημείας Lehninger . Νέα Υόρκη: Worth Publishers.

Μοριακή Βιολογία του Κυττάρου. 4η έκδοση.
Alberts B, Johnson A, Lewis J, et al. Νέα Υόρκη: Garland Science; 2002.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 3. Πρωτεΐνες και ένζυμα

Μάθημα 6. Τύποι ενζυμικών αντιδράσεων

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Αντιδράσεις οξείδωσης-αναγωγής
- Αντιδράσεις ομαδικής μεταφοράς
- Αντιδράσεις Υδρόλυσης
- Δεσμών Άνθρακα-Άνθρακα
- Αντιδράσεις ισομερισμού
- Αντιδράσεις απολίνωσης

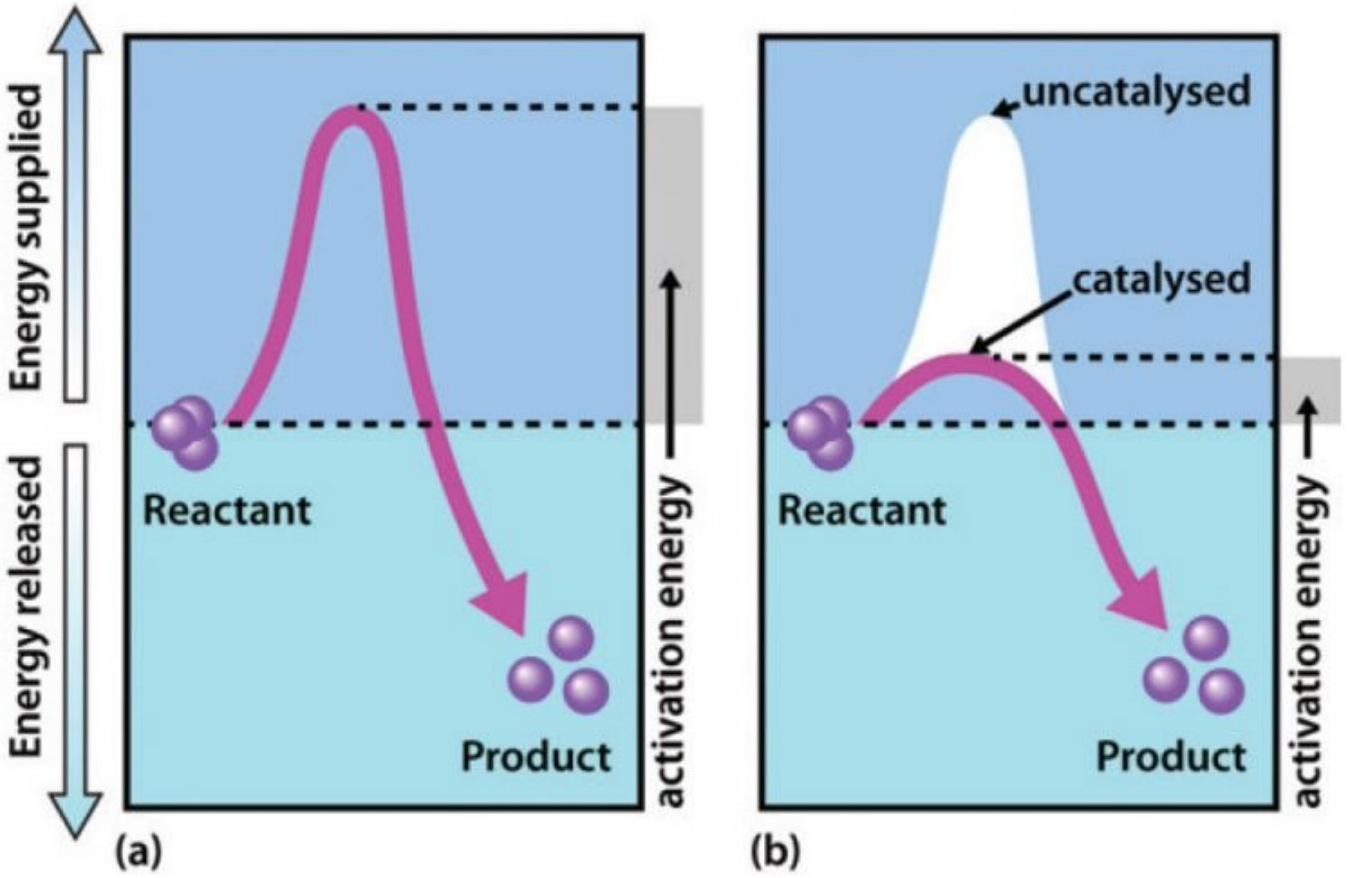
Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Τα ένζυμα είναι βιολογικοί καταλύτες που μειώνουν την ενέργεια ενεργοποίησης που απαιτείται για να προχωρήσει μια αντίδραση προς την κατεύθυνση προς τα εμπρός. Διευκολύνουν το σχηματισμό των ειδών μεταβατικής κατάστασης εντός της αντίδρασης και επιταχύνουν τον ρυθμό της αντίδρασης κατά ένα εκατομμύριο φορές σε σύγκριση με τις μη καταλυόμενες αντιδράσεις. Σημειώστε ότι τα ένζυμα ΔΕΝ μεταβάλλουν το ΔG της αντίδρασης και ΔΕΝ επηρεάζουν τον αυθορμητισμό ή τη θέση ισορροπίας της αντίδρασης. Τα ένζυμα, όπως και άλλοι καταλύτες, επίσης δεν καταναλώνονται κατά τη διάρκεια της αντίδρασης. Τα ένζυμα έχουν υψηλή εξειδίκευση υποστρώματος και μπορούν ακόμη και να εμφανίσουν γεωειδικότητα που οδηγεί στη δημιουργία στερεοειδικών προϊόντων.

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή



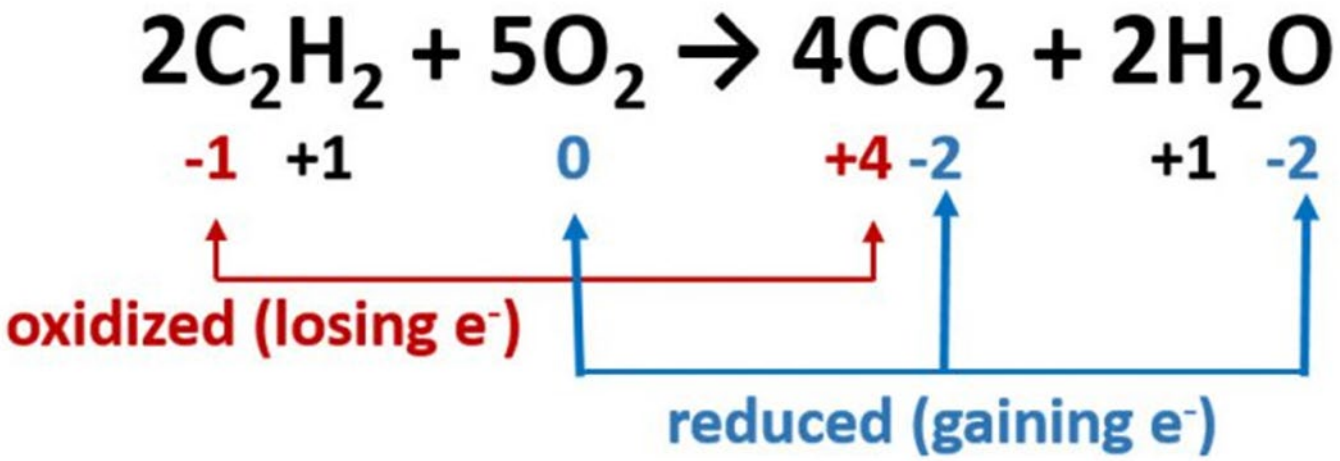
Αντιδράσεις οξείδωσης-αναγωγής

Μια αντίδραση οξείδωσης-αναγωγής (οξειδοαναγωγής) είναι ένας τύπος χημικής αντίδρασης που περιλαμβάνει τη μεταφορά ηλεκτρονίων μεταξύ δύο ατόμων ή ενώσεων. Η ουσία που χάνει τα ηλεκτρόνια λέγεται ότι είναι οξειδωμένη, ενώ η ουσία που κερδίζει τα ηλεκτρόνια λέγεται ότι είναι ανηγμένη. Οι αντιδράσεις οξειδοαναγωγής πρέπει πάντα να συμβαίνουν μαζί. Εάν ένα μόριο οξειδωθεί, τότε ένα άλλο μόριο πρέπει να αναχθεί (δηλ . τα ηλεκτρόνια δεν εμφανίζονται από τον αέρα για να προστεθούν σε μια ένωση, πρέπει πάντα να προέρχονται από κάπου!).

Η αλλαγή στη σύσταση των ηλεκτρονίων μπορεί να αξιολογηθεί στην αλλαγή της κατάστασης (ή του αριθμού) οξείδωσης ενός ατόμου. Επομένως, μια αντίδραση οξείδωσης-αναγωγής είναι οποιαδήποτε χημική αντίδραση στην οποία η κατάσταση οξείδωσης (αριθμός) ενός μορίου, ατόμου ή ιόντος αλλάζει κερδίζοντας ή χάνοντας ένα ηλεκτρόνιο. Θα μάθουμε πώς να αξιολογούμε την κατάσταση οξείδωσης ενός μορίου σε αυτήν την ενότητα. Συνολικά, οι αντιδράσεις οξειδοαναγωγής είναι κοινές και ζωτικής σημασίας για ορισμένες από τις βασικές λειτουργίες της ζωής, συμπεριλαμβανομένης της φωτοσύνθεσης, της αναπνοής, της καύσης και της διάβρωσης ή της σκουριάς.

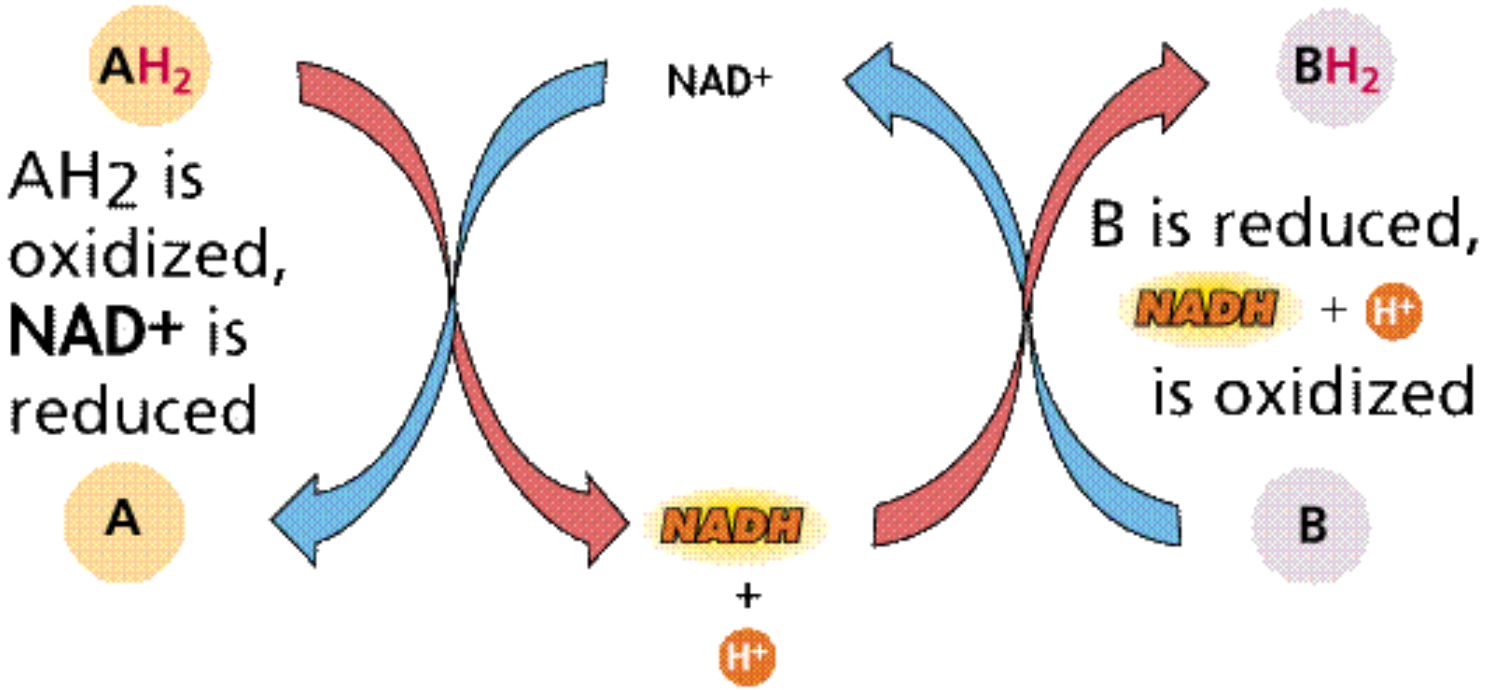
Co-funded by the Erasmus+ Programme of the European Union

Αντιδράσεις οξείδωσης-αναγωγής



Co-funded by the Erasmus+ Programme of the European Union

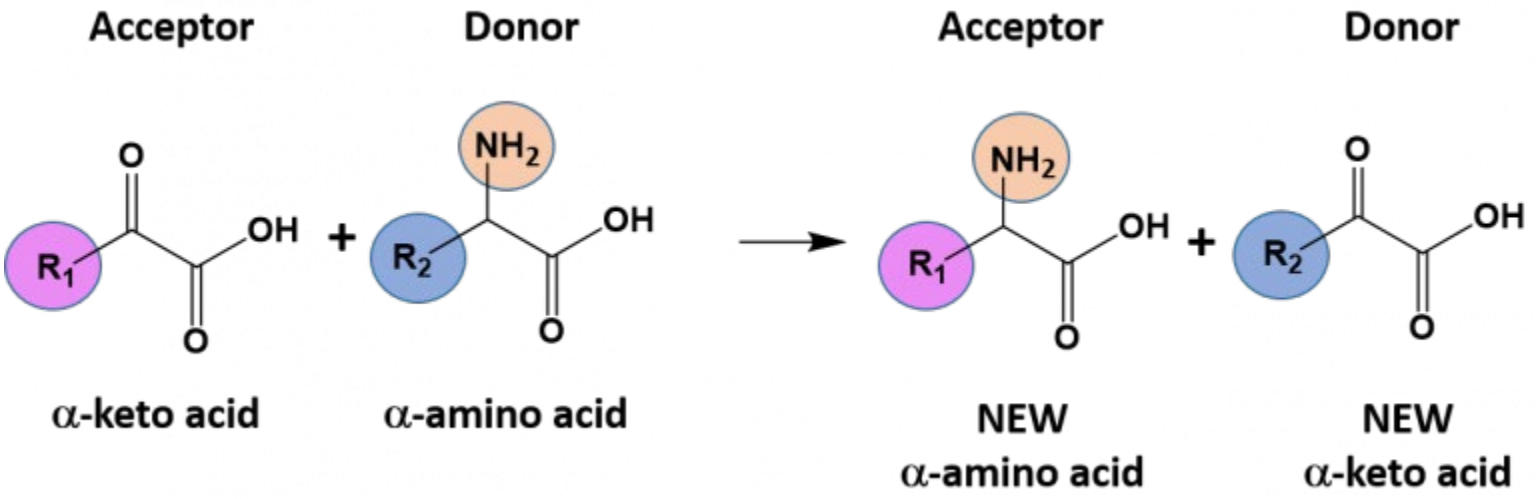
Αντιδράσεις οξείδωσης-αναγωγής



Αντιδράσεις ομαδικής μεταφοράς

Στις αντιδράσεις ομαδικής μεταφοράς, μια λειτουργική ομάδα θα μεταφερθεί από ένα μόριο που χρησιμεύει ως μόριο δότη σε ένα άλλο μόριο που θα είναι το μόριο δέκτη. Η μεταφορά μιας λειτουργικής ομάδας αμίνης από ένα μόριο σε άλλο είναι κοινό παράδειγμα αυτού του τύπου αντίδρασης.

Αντιδράσεις ομαδικής μεταφοράς

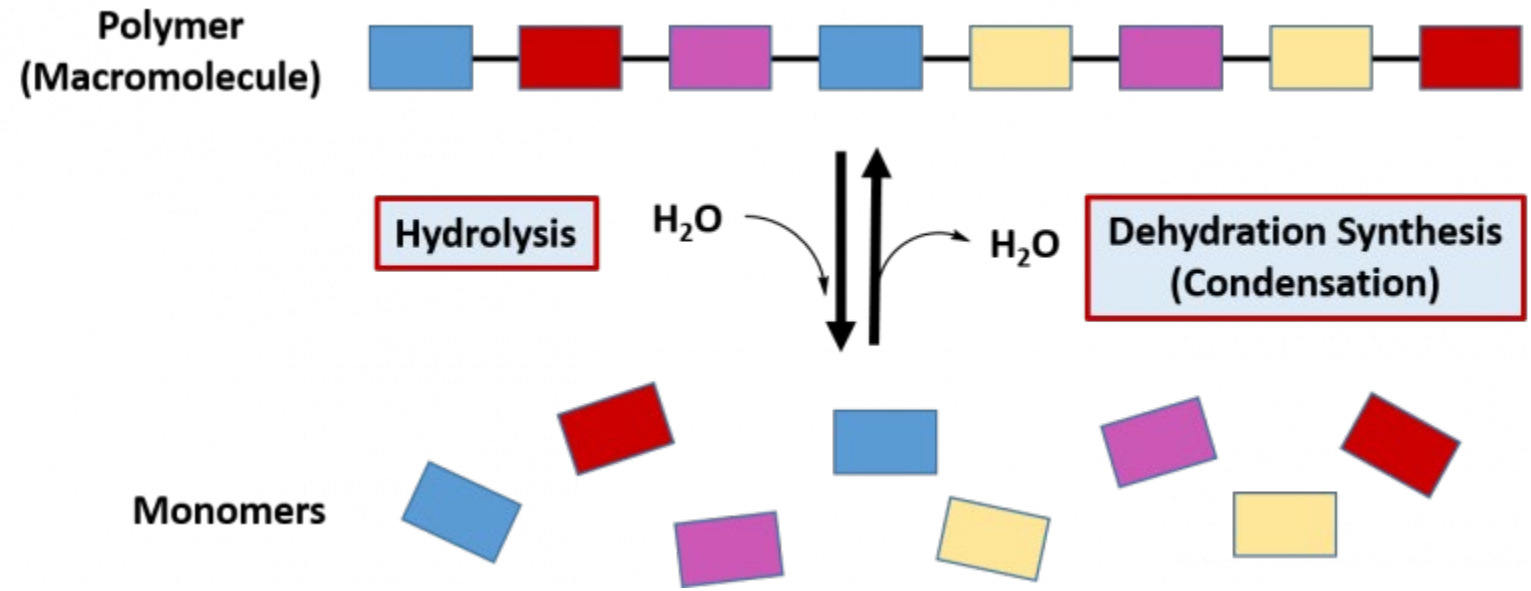


Μεταφορά Λειτουργικής Ομάδας Αμινών. Μια κοινή αντίδραση μεταφοράς ομάδας σε βιολογικά συστήματα είναι αυτή που χρησιμοποιείται για την παραγωγή α-αμινοξέων που μπορούν στη συνέχεια να χρησιμοποιηθούν για τη σύνθεση πρωτεϊνών. Σε αυτή την αντίδραση, ένα α-αμινοξύ χρησιμεύει ως μόριο δότη και ένα α-κετο οξύ (αυτά τα μόρια περιέχουν μια λειτουργική ομάδα καρβοξυλικού οξέος και μια λειτουργική ομάδα κετόνης που διαχωρίζεται από έναν α-άνθρακα) χρησιμεύει ως δέκτης. Στο μόριο δέκτη, το καρβονυλο οξυγόνο αντικαθίσταται με τη λειτουργική ομάδα αμίνης, ενώ στο μόριο δότη, η λειτουργική ομάδα αμίνης αντικαθίσταται από ένα οξυγόνο που σχηματίζει μια νέα λειτουργική ομάδα κετόνης.

Αντιδράσεις Υδρόλυσης

Η ταξινόμηση των αντιδράσεων υδρόλυσης περιλαμβάνει τόσο τις προς τα εμπρός αντιδράσεις που περιλαμβάνουν την προσθήκη νερού σε ένα μόριο για τη διάσπασή του ή την αντίστροφη αντίδραση που περιλαμβάνει την αφαίρεση του νερού για να ενώσει τα μόρια μεταξύ τους, που ονομάζεται σύνθεση αφυδάτωσης (ή συμπύκνωση) (Εικόνα 7.7). Όταν προστίθεται νερό σε ένα μόριο για να το χωρίσει σε δύο μόρια, αυτή η αντίδραση ονομάζεται υδρόλυση. Ο όρος «λύση» σημαίνει το σπάσιμο, και ο όρος «υδρο» αναφέρεται στο νερό. Έτσι, ο όρος υδρόλυση σημαίνει διάσπαση με νερό. Το αντίστροφο αυτής της αντίδρασης περιλαμβάνει την απομάκρυνση του νερού από δύο μόρια για να τα ενώσει μαζί σε ένα μεγαλύτερο μόριο. Δεδομένου ότι τα δύο μόρια χάνουν νερό, αφυδατώνονται. Έτσι, ο σχηματισμός μορίων μέσω της απομάκρυνσης του νερού είναι γνωστός ως σύνθεση αφυδάτωσης. Δεδομένου ότι το νερό είναι επίσης ένα υποπροϊόν αυτών των αντιδράσεων, αναφέρονται επίσης συνήθως ως αντιδράσεις συμπύκνωσης. Όπως είδαμε στο Κεφάλαιο 6, ο σχηματισμός των κύριων κατηγοριών μακρομορίων στο σώμα (πρωτεΐνες, υδατάνθρακες, λιπίδια και νουκλεϊκά οξέα) σχηματίζεται μέσω της σύνθεσης αφυδάτωσης όπου το νερό απομακρύνεται από τα μόρια. Κατά την κανονική πέψη των μορίων της τροφής μας, τα κύρια μακρομόρια διασπώνται στα δομικά στοιχεία τους μέσω της διαδικασίας της υδρόλυσης

Αντιδράσεις Υδρόλυσης

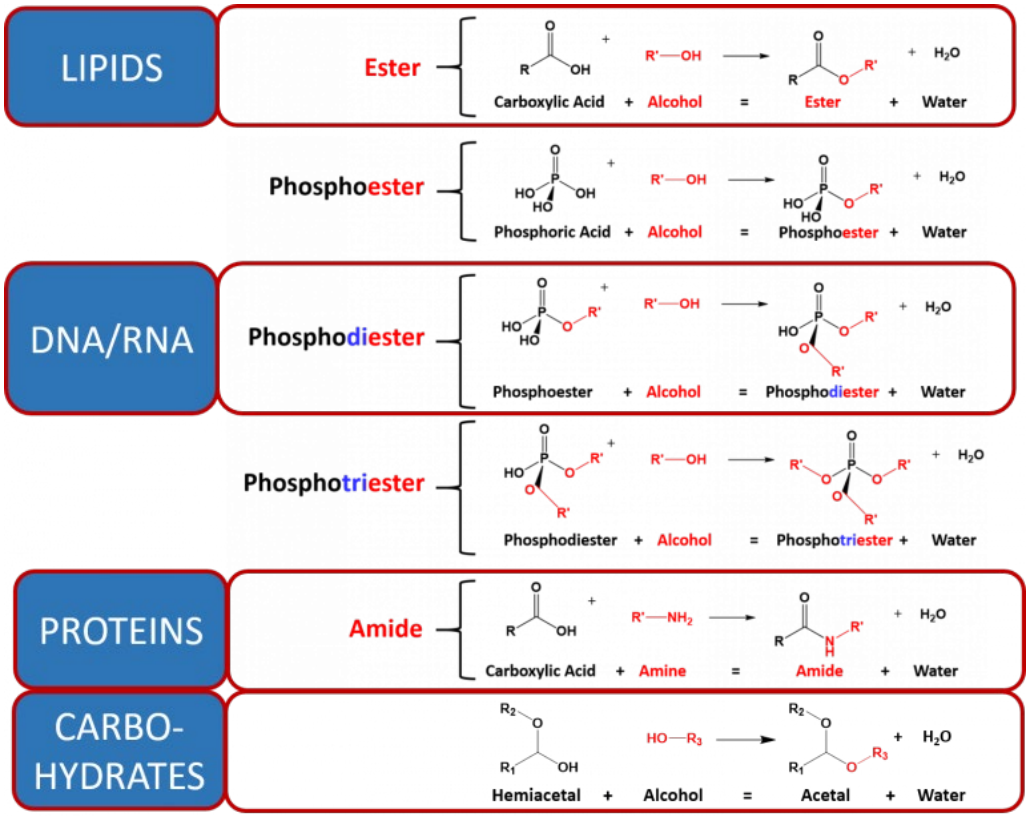


Σύνθεση Υδρόλυσης και Αφυδάτωσης. Οι αντιδράσεις της υδρόλυσης μεσολαβούν στη διάσπαση μεγαλύτερων πολυμερών στα μονομερή δομικά στοιχεία τους με την προσθήκη νερού στα μόρια. Το αντίστροφο της αντίδρασης είναι η σύνθεση αφυδάτωσης, όπου το νερό αφαιρείται από τα δομικά στοιχεία του μονομερούς για να δημιουργηθεί η μεγαλύτερη δομή πολυμερούς.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Αντιδράσεις Υδρόλυσης



Αντιδράσεις σύνθεσης αφυδάτωσης που εμπλέκονται στο σχηματισμό μακρομορίων. Εμφανίζονται οι κύριες οργανικές αντιδράσεις που απαιτούνται για τη βιοσύνθεση λιπιδίων, νουκλεϊκών οξέων (DNA/RNA), πρωτεϊνών και υδατανθράκων. Σημειώστε ότι σε όλες τις αντιδράσεις, υπάρχει μια λειτουργική ομάδα που περιέχει δύο ομάδες απόσυρσης ηλεκτρονίων (το καρβοξυλικό οξύ, το φωσφορικό οξύ και η ημιακετάλη έχουν δύο άτομα οξυγόνου συνδεδεμένα με ένα κεντρικό άτομο άνθρακα ή φωσφόρου). Αυτό σχηματίζει ένα αντιδραστικό μερικώς θετικό κεντρικό άτομο (άνθρακας στην περίπτωση του καρβοξυλικού οξέος και της ημιακετάλης, ή φώσφορος στην περίπτωση του φωσφορικού οξέος) που μπορεί να προσβληθεί από το ηλεκτραρνητικό οξυγόνο ή άζωτο από μια λειτουργική ομάδα αλκοόλης ή αμίνης.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

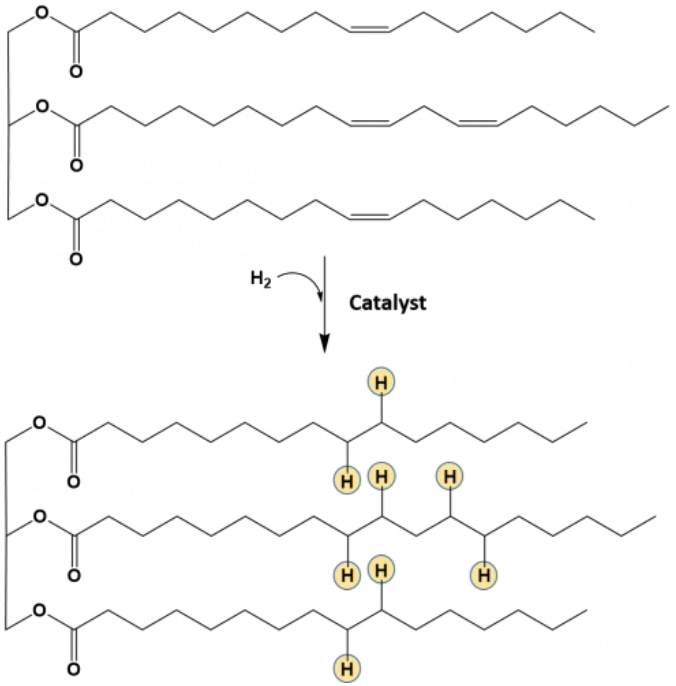
Co-funded by the Erasmus+ Programme of the European Union

Ο Σχηματισμός/Απομάκρυνση των Διπλών Δεσμών Άνθρακα-Άνθρακα

Οι αντιδράσεις που μεσολαβούν στο σχηματισμό και την απομάκρυνση των διπλών δεσμών άνθρακα-άνθρακα είναι επίσης κοινές στα βιολογικά συστήματα και καταλύονται από μια κατηγορία ενζύμων που ονομάζονται λυάσες. Ο σχηματισμός ή η αφαίρεση διπλών δεσμών άνθρακα-άνθρακα χρησιμοποιείται επίσης σε αντιδράσεις συνθετικής οργανικής χημείας για τη δημιουργία επιθυμητών οργανικών μορίων. Ένας από αυτούς τους τύπους αντιδράσεων ονομάζεται αντίδραση υδρογόνωσης, όπου ένα μόριο υδρογόνου (H2) προστίθεται σε έναν διπλό δεσμό CC, μειώνοντάς τον σε έναν απλό δεσμό CC. Εάν αυτό γίνει με χρήση ακόρεστων ελαίων, τα ακόρεστα λίπη μπορούν να μετατραπούν σε κορεσμένα λίπη. Αυτός ο τύπος αντίδρασης γίνεται συνήθως για την παραγωγή μερικώς υδρογονωμένων ελαίων μετατρέποντάς τα από υγρά σε θερμοκρασία δωματίου σε στερεά. Οι μαργαρίνες από φυτικό έλαιο παρασκευάζονται με αυτόν τον τρόπο. Δυστυχώς, ένα παραπροϊόν αυτής της αντίδρασης μπορεί να είναι ο σχηματισμός TAGS που περιέχουν trans διπλούς δεσμούς. Μόλις αναγνωρίστηκαν οι κίνδυνοι για την υγεία από την κατανάλωση τρανς λιπαρών, ο Οργανισμός Τροφίμων και Φαρμάκων (FDA) απαγόρευσε τη συμπερίληψη των τρανς λιπαρών στα τρόφιμα. Αυτή η απαγόρευση τέθηκε σε ισχύ το καλοκαίρι του 2015 και έδωσε στους παρασκευαστές τροφίμων τρία χρόνια για να τους εξαλείψουν από τον εφοδιασμό τροφίμων, με προθεσμία την 18η Ιουνίου 2018.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Ο Σχηματισμός/Απομάκρυνση των Διπλών Δεσμών Άνθρακα-Άνθρακα

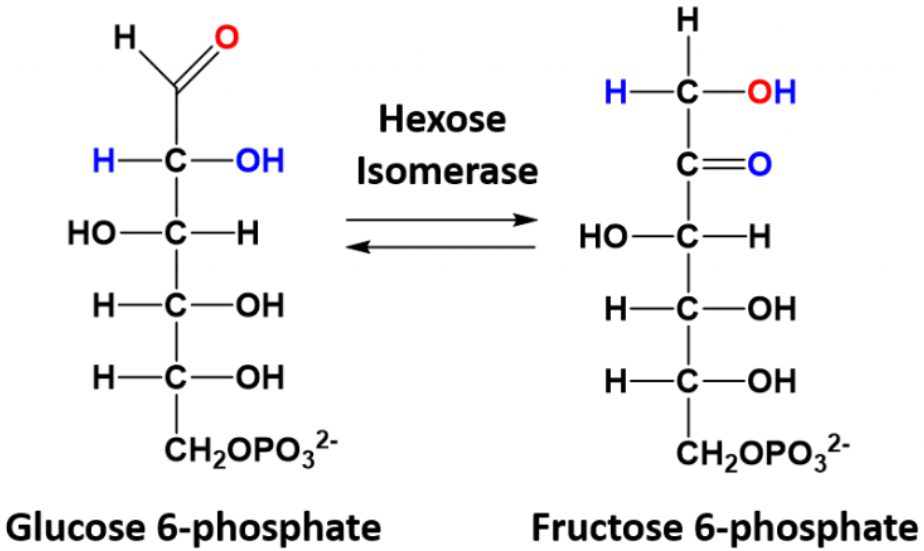


Υδρογόνωση ελαίων για την παραγωγή μαργαρίνης. Τα ακόρεστα έλαια μπορούν μερικώς ή πλήρως να υδρογονωθούν για να παράγουν τα κορεσμένα λιπαρά οξέα για να παράγουν μαργαρίνες που θα παραμείνουν στερεές σε θερμοκρασία δωματίου. Η προσθήκη των νέων ατόμων υδρογόνου για τη δημιουργία των κορεσμένων υδρογονανθράκων εμφανίζεται με κίτρινο χρώμα στο τελικό προϊόν.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Αντιδράσεις ισομερισμού



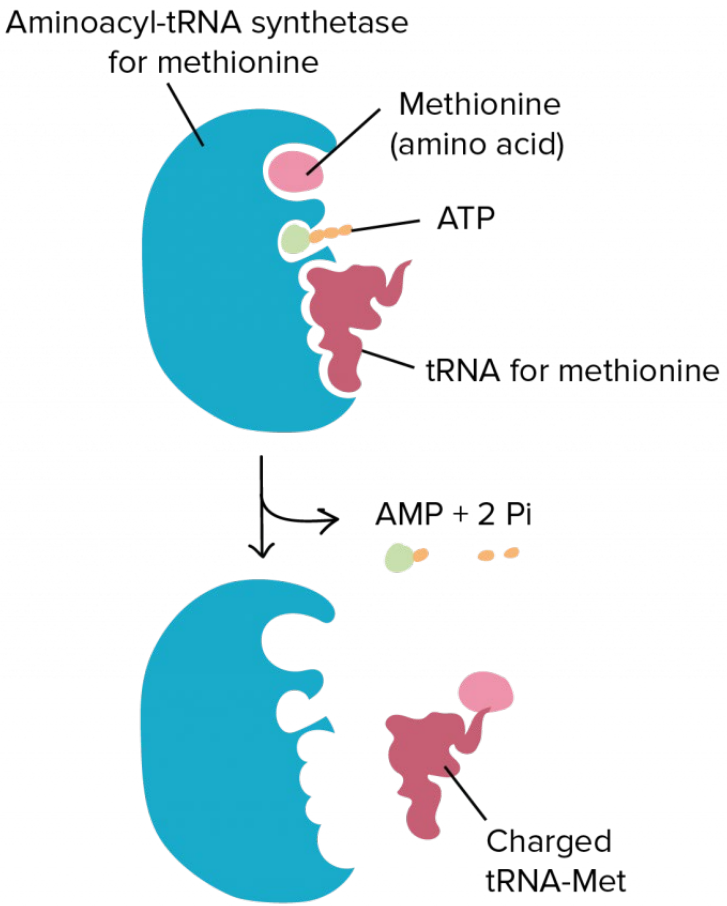
Στις αντιδράσεις ισομερισμού ένα μεμονωμένο μόριο αναδιατάσσεται έτσι ώστε να διατηρεί τον ίδιο μοριακό τύπο αλλά τώρα έχει διαφορετική σειρά σύνδεσης των ατόμων που σχηματίζουν ένα δομικό ή στερεοϊσομερές. Η μετατροπή της 6-φωσφορικής γλυκόζης σε 6-φωσφορική φρουκτόζη είναι ένα καλό παράδειγμα αντίδρασης ισομερισμού

Co-funded by the Erasmus+ Programme of the European Union

Αντιδράσεις απολίνωσης

Οι αντιδράσεις απολίνωσης χρησιμοποιούν την ενέργεια του ATP για να ενώσουν δύο μόρια μεταξύ τους. Ένα παράδειγμα αυτού του είδους αντίδρασης είναι η ένωση του αμινοξέος με το μόριο RNA μεταφοράς (tRNA) κατά τη διάρκεια της πρωτεϊνοσύνθεσης. Κατά τη διάρκεια της πρωτεϊνοσύνθεσης τα μόρια tRNA φέρνουν καθένα από τα αμινοξέα στο ριβόσωμα όπου μπορούν να ενσωματωθούν στη νεοαναπτυσσόμενη πρωτεϊνική αλληλουχία. Για να γίνει αυτό, τα μόρια tRNA πρέπει πρώτα να συνδεθούν με το κατάλληλο αμινοξύ. Υπάρχουν διαθέσιμα συγκεκριμένα ένζυμα που ονομάζονται συνθετάσες αμινοακυλο-tRNA που μεσολαβούν σε αυτήν την αντίδραση. Τα ένζυμα συνθετάσης χρησιμοποιούν την ενέργεια του ATP για να συνδέσουν ομοιοπολικά το αμινοξύ στο μόριο tRNA. Για καθένα από τα 20 αμινοξέα, υπάρχει ένα συγκεκριμένο μόριο tRNA και ένα συγκεκριμένο ένζυμο συνθετάσης που θα εξασφαλίσει τη σωστή σύνδεση του σωστού αμινοξέος με το μόριο του tRNA.

Αντιδράσεις απολίνωσης

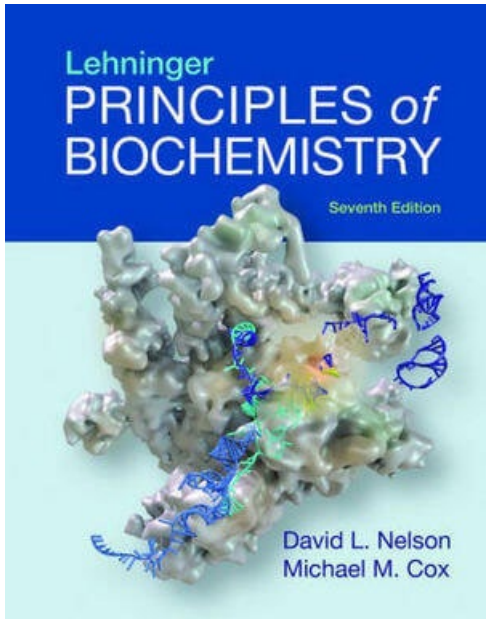


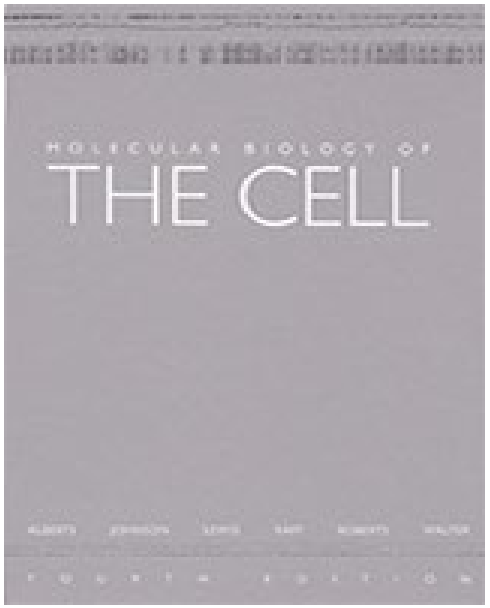
Αντίδραση απολίνωσης Ομοιοπολική σύνδεση μεθειονίνης με το κατάλληλο tRNA. Το ένζυμο αμινο-ακυλο tRNA συνθετάσης για τη μεθειονίνη (εμφανίζεται με μπλε) συνδέει ομοιοπολικά τη μεθειονίνη (ανοιχτό ροζ) με το μόριο tRNA της μεθειονίνης (σκούρο ροζ). Αυτή η αντίδραση απαιτεί την ενέργεια που παρέχεται από τη διάσπαση του μορίου ATP σε AMP, απελευθερώνοντας ενέργεια με τη διάσπαση των φωσφορικών δεσμών σε δύο ανόργανα φωσφορικά ιόντα (2 Pi).

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Ποια είναι τα κύρια χαρακτηριστικά των ενζυμικών αντιδράσεων;
- Εξηγήστε τους διαφορετικούς τύπους καταλυτικών αντιδράσεων.





Lehninger , AL, Nelson, DL, & Cox, MM (2000). Αρχές βιοχημείας Lehninger . Νέα Υόρκη: Worth Publishers.

Μοριακή Βιολογία του Κυττάρου. 4η έκδοση.
Alberts B, Johnson A, Lewis J, et al. Νέα Υόρκη: Garland Science; 2002.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 3. Πρωτεΐνες και ένζυμα

Μάθημα 7. Κινητική ενζυμικών αντιδράσεων

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Σταθερά ισορροπίας
- Παράγοντες που επηρεάζουν τη δραστηριότητα των ενζύμων
- Σταθερά Michaelis

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Η ταχύτητα ή ο ρυθμός της ενζυμικής αντίδρασης εκτιμάται από το ρυθμό αλλαγής του υποστρώματος σε προϊόν ανά μονάδα χρόνου.

Στην πράξη, προσδιορίζεται η αρχική ταχύτητα. Μετά από πολύ χρόνο, η ταχύτητα μπορεί να τείνει να πέσει λόγω μείωσης της συγκέντρωσης του υποστρώματος κάτω από ένα κρίσιμο επίπεδο.

Οι καταλύτες αυξάνουν τον ρυθμό αντίδρασης, αλλά δεν αλλάζουν την ισορροπία.

Co-funded by the Erasmus+ Programme of the European Union

Η ταχύτητα είναι ανάλογη με τη συγκέντρωση των αντιδρώντων μορίων.

A + B → C + D

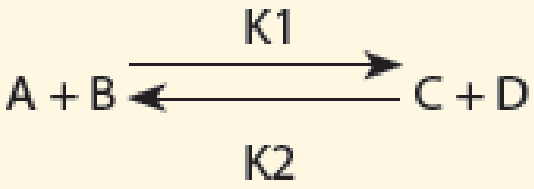
Εάν η συγκέντρωση του Α ή του Β διπλασιαστεί, ο ρυθμός αντίδρασης επίσης διπλασιάζεται. Εάν οι συγκεντρώσεις των Α και Β διπλασιαστούν μαζί, η ταχύτητα γίνεται 4 φορές μεγαλύτερη.

Co-funded by the Erasmus+ Programme of the European Union

Εξαγωγή σταθεράς ισορροπίας

V α [A] [B]

At equilibrium, forward reaction and backward reaction are equal, so that



Forward reaction R1	= K1 [A] [B]
and backward reaction R2	= K2 [C] [D]
At equilibrium, R1	= R2
Or, K1 [A] [B]	= K2 [C] [D]
Or, K1 = [C] [D]	= Keq or
K2 = [A] [B]	Equilibrium constant.

Co-funded by the Erasmus+ Programme of the European Union

1. Η σταθερά ισορροπίας της αντίδρασης είναι ο λόγος σταθερού ρυθμού αντίδρασης προς τα εμπρός και προς τα πίσω.

2. Σε ισορροπία, οι αντιδράσεις προς τα εμπρός και προς τα πίσω είναι ίσες. Η ισορροπία είναι μια δυναμική κατάσταση. Αν και δεν έχουμε καθαρή αλλαγή στις συγκεντρώσεις του υποστρώματος και του προϊόντος, τα μόρια πάντα αλληλομετατρέπονται.

3. Η αριθμητική τιμή της σταθεράς μπορεί να υπολογιστεί με εύρεση των συγκεντρώσεων υποστρωμάτων και προϊόντων.

4. Εάν το Keq είναι μεγαλύτερο από 1, ευνοείται η προς τα εμπρός αντίδραση. Σε τέτοιες περιπτώσεις, η αντίδραση είναι αυθόρμητη και εξώθερμη.

5. Η συγκέντρωση του ενζύμου δεν επηρεάζει το Keq . Η συγκέντρωση του ενζύμου σίγουρα αυξάνει τον ρυθμό αντίδρασης αλλά όχι το Keq ή την τελική κατάσταση. Με άλλα λόγια, το ένζυμο κάνει πιο γρήγορη τη διαδικασία να φτάσουμε στην ισορροπία.

Co-funded by the Erasmus+ Programme of the European Union

Παραγωγή σταθεράς Michaelis (Km)

$$E + S \underset{K2}{\overset{K1}{\rightleftarrows}} E - S \xrightarrow{K3} E + P$$

If concentration of substrate is increased, the forward reaction K1 is increased, and so K3 as well as total velocity is correspondingly enhanced. The three different constants may be made into one equation,

$$Km = \frac{K2 + K3}{K1}$$

Km is called as Michaelis Constant
It is further shown that

$$\text{Velocity } (v) = \frac{Vmax\,[S]}{Km + [S]}$$

When concentration of substrate is made equal to Km, i.e.
When [S] = Km

$$\text{Velocity } (v) = \frac{Vmax\,[S]}{[S] + [S]} = \frac{Vmax\,[S]}{2[S]} = \frac{Vmax}{2}$$

or v = ½ Vmax

Co-funded by the Erasmus+ Programme of the European Union

Παράγοντες που επηρεάζουν τη δραστηριότητα των ενζύμων

1. Συγκέντρωση ενζύμου
2. Συγκέντρωση υποστρώματος
3. Συγκέντρωση προϊόντος
4. Θερμοκρασία
5. Συγκέντρωση ιόντων υδρογόνου (pH)
6. Παρουσία ενεργοποιητών
7. Παρουσία αναστολέων
8. Παρουσία καταστολέα ή αποκατασταλτικό
9. Ομοιοπολική τροποποίηση.

Συγκέντρωση ενζύμου

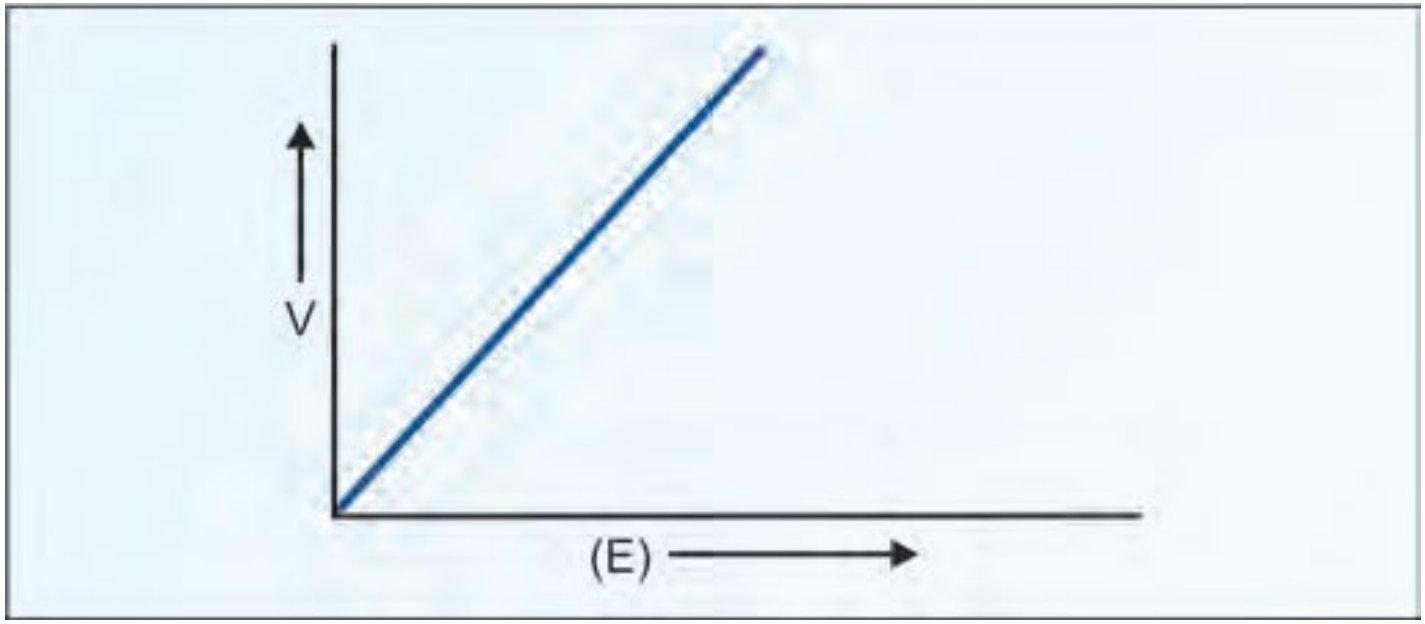
i. Ο ρυθμός μιας αντίδρασης ή η ταχύτητα (V) είναι ανάλογη με τη συγκέντρωση του ενζύμου, όταν υπάρχει επαρκές υπόστρωμα. Η ταχύτητα αντίδρασης αυξάνεται αναλογικά με τη συγκέντρωση του ενζύμου, υπό την προϋπόθεση ότι η συγκέντρωση του υποστρώματος είναι απεριόριστη.

ii. Ως εκ τούτου, αυτή η ιδιότητα χρησιμοποιείται για τον προσδιορισμό του επιπέδου συγκεκριμένου ενζύμου στο πλάσμα, στον ορό ή στους ιστούς.

iii. Ο γνωστός όγκος ορού επωάζεται με υπόστρωμα για καθορισμένο χρόνο, στη συνέχεια η αντίδραση διακόπτεται και το προϊόν ποσοτικοποιείται (μέθοδος τελικού σημείου). Δεδομένου ότι το προϊόν που σχηματίζεται θα είναι ανάλογο με τη συγκέντρωση του ενζύμου, η τελευταία θα μπορούσε να προσδιοριστεί.

Co-funded by the Erasmus+ Programme of the European Union

Συγκέντρωση ενζύμου

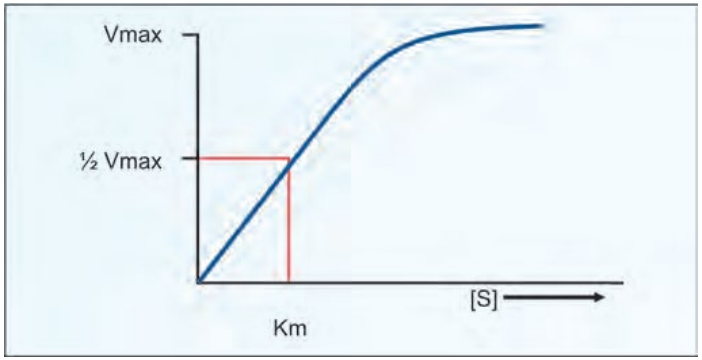


Επίδραση συγκέντρωσης ενζύμου

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

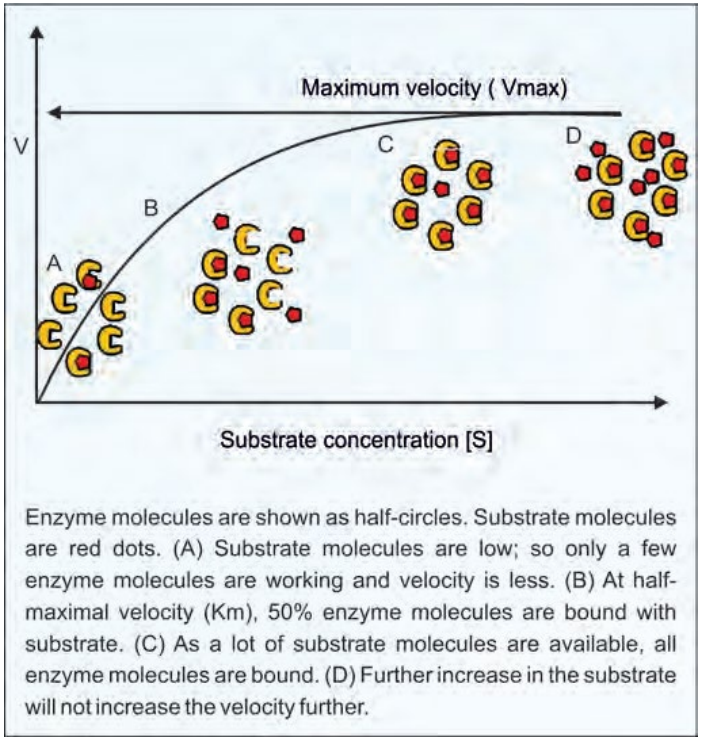
Co-funded by the Erasmus+ Programme of the European Union

Επίδραση της συγκέντρωσης του υποστρώματος



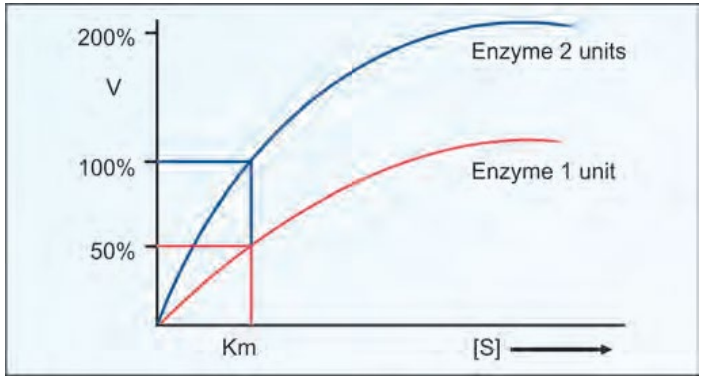
Επίδραση συγκέντρωσης υποστρώματος (καμπύλη κορεσμού υποστρώματος)

Η μέγιστη ταχύτητα που προκύπτει ονομάζεται Vmax



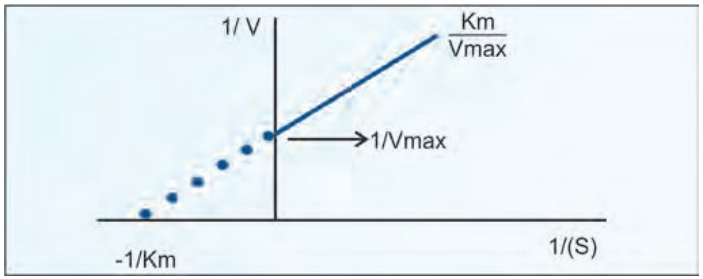
Enzyme molecules are shown as half-circles. Substrate molecules are red dots. (A) Substrate molecules are low; so only a few enzyme molecules are working and velocity is less. (B) At half-maximal velocity (Km), 50% enzyme molecules are bound with substrate. (C) As a lot of substrate molecules are available, all enzyme molecules are bound. (D) Further increase in the substrate will not increase the velocity further.

Μιχαηλής Κωνστάντ



Όσο μικρότερη είναι η αριθμητική τιμή του Km, τόσο η συγγένεια του ενζύμου προς το υπόστρωμα είναι μεγαλύτερη. Για να αναφέρουμε ένα παράδειγμα, η Km γλυκοκινάσης είναι 10 mmol/L και αυτή της εξοκινάσης είναι 0,05 mmol/L. Επομένως, 50% μόρια εξοκινάσης είναι κορεσμένα ακόμη και σε χαμηλότερη συγκέντρωση γλυκόζης. Με άλλα λόγια, η εξοκινάση έχει μεγαλύτερη συγγένεια προς τη γλυκόζη από ό,τι γλυκοκινάση.

Διπλό Αμοιβαίο Οικόπεδο

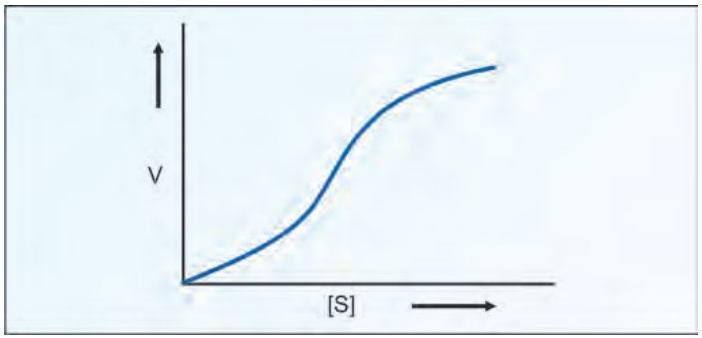


Μερικές φορές δεν είναι πρακτικό να επιτευχθεί υψηλή συγκέντρωση υποστρώματος για την επίτευξη των συνθηκών μέγιστης ταχύτητας. Έτσι, το ½Vmax ή το Km μπορεί να είναι δύσκολο να προσδιοριστεί. Επειτα, τα πειραματικά δεδομένα σε χαμηλότερες συγκεντρώσεις απεικονίζονται αντίστροφα. Η ευθεία που λαμβάνεται με αυτόν τον τρόπο παρεκβάλλεται για να βρούμε το αντίστροφο του Km. Αυτό ονομάζεται διάγραμμα Lineweaver-Burk ή Double Reciprocal Plot που μπορεί να προκύψει από την εξίσωση Michaelis-Menten (Για την εξίσωση Lineweaver-Burk, βλέπε Πλαίσιο 5.9). Αν σχεδιάσουμε το 1/v έναντι του 1/[S], θα δώσει ένα ευθύγραμμο γράφημα. Η τομή στον άξονα Χ είναι μείον 1/Km, από την οποία μπορεί να υπολογιστεί το Km .

Διάγραμμα Dixon

Η ταχύτητα (V) μετριέται σε διάφορες συγκεντρώσεις του αναστολέα (I), όταν η συγκέντρωση του υποστρώματος (S) διατηρείται σταθερή. Χρησιμοποιείται για τον προσδιορισμό σταθερών αναστολής. Μια γραφική παράσταση 1/V έναντι [I] δίνει μια ευθεία γραμμή. Το πείραμα επαναλαμβάνεται σε διαφορετικές συγκεντρώσεις υποστρωμάτων.

Συνεταιριστική Δέσμευση



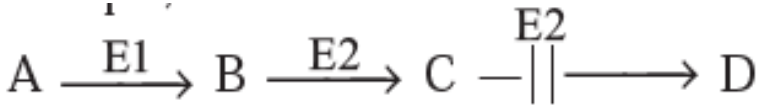
Ορισμένα ένζυμα μπορεί να μην ακολουθούν αυστηρά την κινητική Michaelis-Menten. Όταν το ένζυμο έχει πολλές υπομονάδες, και η δέσμευση του υποστρώματος σε μία μονάδα ενισχύει τη συγγένεια για δέσμευση με άλλες υπομονάδες, προκύπτει μια καμπύλη κορεσμού σε σχήμα σιγμοειδούς. Σε τέτοιες περιπτώσεις, ο προσδιορισμός της τιμής Km, όπως φαίνεται στην προηγούμενη παράγραφο, θα είναι άκυρος. Αντίθετα, χρησιμοποιείται η εξίσωση Hill, που περιγράφηκε αρχικά για την εξήγηση της δέσμευσης οξυγόνου στην αιμοσφαιρίνη.

Co-funded by the Erasmus+ Programme of the European Union

Επίδραση της συγκέντρωσης των προϊόντων

Σε μια αναστρέψιμη αντίδραση, το SDP, όταν επιτευχθεί ισορροπία, σύμφωνα με το νόμο της δράσης μάζας, ο ρυθμός αντίδρασης επιβραδύνεται. Έτσι, όταν η συγκέντρωση του προϊόντος αυξάνεται, η αντίδραση επιβραδύνεται, σταματά ή και αντιστρέφεται. Σε εγγενή λάθη του του μεταβολισμού, ένα ένζυμο μιας μεταβολικής οδού μπλοκάρεται.

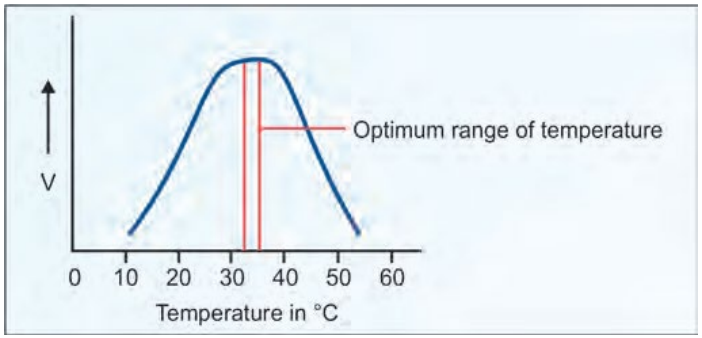
Για παράδειγμα,



Εάν το ένζυμο E3 απουσιάζει, θα συσσωρευτεί το C, το οποίο με τη σειρά του, θα αναστείλει την E2. Κατά συνέπεια, με την πάροδο του χρόνου, ολόκληρο το μονοπάτι είναι φραγμένο.

Co-funded by the Erasmus+ Programme of the European Union

Επίδραση της θερμοκρασίας

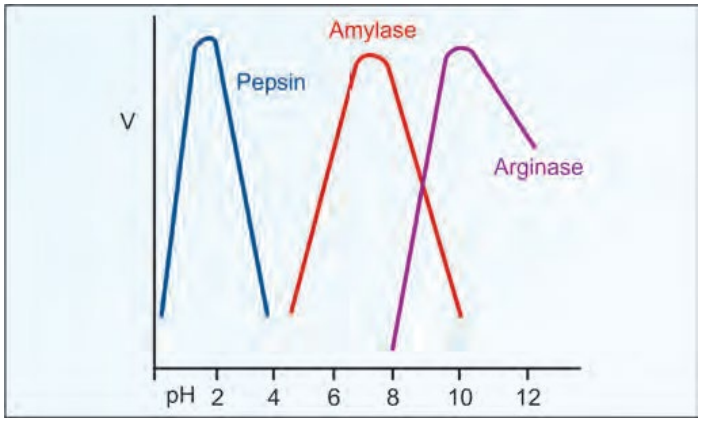


Όταν όμως η θερμοκρασία είναι μεγαλύτερη από 50°C, λαμβάνει χώρα θερμική μετουσίωση και συνακόλουθα απώλεια της τριτογενούς δομής της πρωτεΐνης. Άρα η δραστηριότητα του ενζύμου μειώνεται. Τα περισσότερα ανθρώπινα ένζυμα έχουν βέλτιστη θερμοκρασία γύρω στους 37°C. Ορισμένα βακτήρια που ζουν σε θερμές πηγές θα έχουν ένζυμα με βέλτιστη θερμοκρασία κοντά στους 100°C.

Η ταχύτητα της ενζυμικής αντίδρασης αυξάνεται όταν η θερμοκρασία του μέσου αυξάνεται και φτάνει στο μέγιστο και μετά πέφτει (Καμπύλη σε σχήμα καμπάνας). Η θερμοκρασία στην οποία η μέγιστη ποσότητα του υποστρώματος μετατρέπεται σε προϊόν ανά μονάδα χρόνου ονομάζεται βέλτιστη θερμοκρασία. Καθώς η θερμοκρασία αυξάνεται, περισσότερα μόρια παίρνουν ενέργεια ενεργοποίησης, ή τα μόρια είναι σε αυξημένο ρυθμό κίνησης. Έτσι οι πιθανότητες σύγκρουσής τους είναι αυξημένες και άρα η ταχύτητα της αντίδρασης αυξάνεται. Η θερμοκρασία ο συντελεστής (Q10) είναι ο παράγοντας με τον οποίο ο ρυθμός κατάλυσης αυξάνεται κατά 10°C. Γενικά, ο ρυθμός αντίδρασης των περισσότερων ενζύμων θα διπλασιαστεί κατά άνοδο στους 10°C.

Co-funded by the Erasmus+ Programme of the European Union

Επίδραση του pH

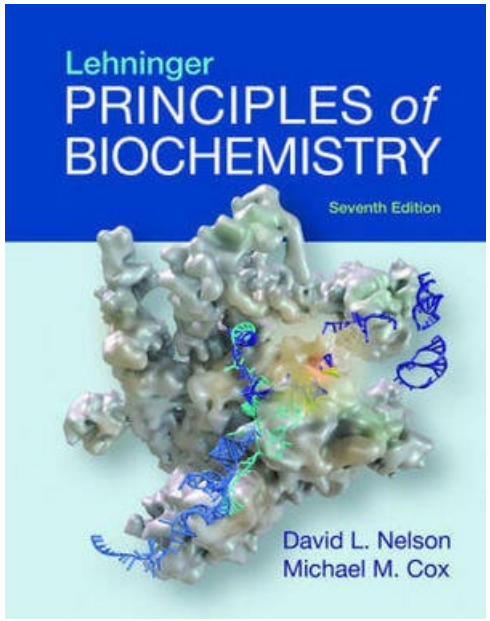


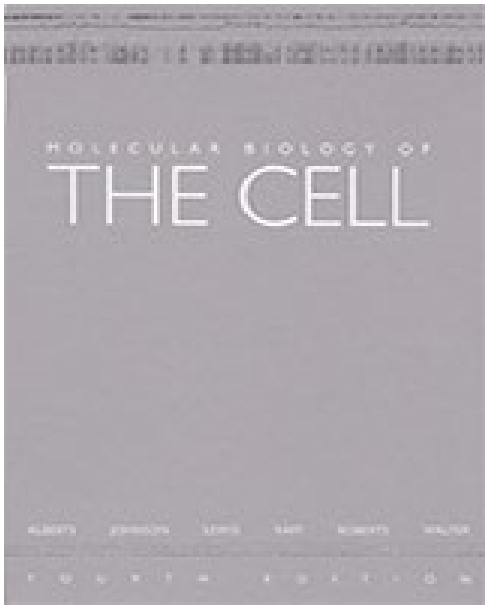
Κάθε ένζυμο έχει ένα βέλτιστο pH, και μακριά απ' αυτό η ταχύτητα θα μειωθεί δραστικά. Το γράφημα δείχνει μια καμπύλη σε σχήμα καμπάνας. Το pH αποφασίζει το φορτίο στα υπολείμματα αμινοξέων στη δραστική θέση. Το καθαρό φορτίο στην ενζυμική πρωτεΐνη θα επηρεάσει το υπόστρωμα δεσμευτική και καταλυτική δραστηριότητα. Το βέλτιστο pH μπορεί να διαφέρει ανάλογα με τη θερμοκρασία, τη συγκέντρωση του υποστρώματος, παρουσία ιόντων κλπ. Συνήθως τα ένζυμα έχουν το βέλτιστο pH μεταξύ 6 και 8. Μερικές σημαντικές εξαιρέσεις είναι η πεψίνη (με βέλτιστο pH 1–2). αλκαλική φωσφατάση (βέλτιστη pH 9-10) και όξινη φωσφατάση (4-5).

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε την κινητική των ενζυματικών αντιδράσεων.
- Ποιες είναι οι σημαντικές σταθερές που περιγράφουν την ενζυματική αντίδραση;
- Τι σημαίνει σταθερά ισορροπίας





Lehninger , AL, Nelson, DL, & Cox, MM (2000). Αρχές βιοχημείας Lehninger. Νέα Υόρκη: Worth Publishers.

Μοριακή Βιολογία του Κυττάρου. 4η έκδοση.
Alberts B, Johnson A, Lewis J, et al. Νέα Υόρκη: Garland Science; 2002.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 3. Πρωτεΐνες και ένζυμα

Μάθημα 8. Ρύθμιση ενζυμικών αντιδράσεων

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Ενζυμική ενεργοποίηση
- Αναστολή ενζύμου
- Αλλοστερικός Κανονισμός

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Τα ένζυμα μπορούν να ρυθμιστούν από άλλα μόρια που είτε αυξάνουν είτε μειώνουν τη δραστηριότητά τους. Τα μόρια που αυξάνουν τη δραστηριότητα ενός ενζύμου ονομάζονται ενεργοποιητές, ενώ τα μόρια που μειώνουν τη δραστηριότητα ενός ενζύμου ονομάζονται αναστολείς.

Υπάρχουν πολλά είδη μορίων που μπλοκάρουν ή προάγουν τη λειτουργία των ενζύμων και που επηρεάζουν τη λειτουργία των ενζύμων με διαφορετικές οδούς.

Ενζυμική ενεργοποίηση

Παρουσία ορισμένων ανόργανων ιόντων, ορισμένα ένζυμα παρουσιάζουν υψηλότερη δραστηριότητα. Έτσι, τα ιόντα χλωρίου ενεργοποιούν την αμυλάση του σιέλου και τα ιόντα ασβεστίου ενεργοποιούν τη λιπάση.

Ενζυμική ενεργοποίηση

Ένας άλλος τύπος ενεργοποίησης είναι η μετατροπή από ανενεργό προένζυμο ή ζυμογόνο σε ενεργό ένζυμο.

i . Με διάσπαση ενός μόνο πεπτιδικού δεσμού και αφαίρεση ενός μικρού πολυπεπτίδιου από τρυψινογόνο, σχηματίζεται η ενεργός θρυψίνη. Αυτό έχει ως αποτέλεσμα την αποκάλυψη του ενεργού κέντρου.

ii. Ομοίως η τρυψίνη ενεργοποιεί το χυμοθρυψινογόνο , έτσι ώστε το πεπτίδιο A (1–13 αμινοξέα), το πεπτίδιο B (16-46) και το πεπτίδιο C (149-245) σχηματίζονται από χυμοθρυψινογόνο . Αυτά τα 3 τμήματα ευθυγραμμίζονται με τρόπο που η ιστιδίνη (57) και το ασπαρτικό (102) και υπολείμματα σερίνης (195) σχηματίζουν την ενεργή θέση.

iii. Όλα τα γαστρεντερικά ένζυμα συντίθενται σε μορφή προ-ενζύμων, και μόνο μετά την έκκριση στο πεπτικό κανάλι, ενεργοποιούνται. Αυτό εμποδίζει την αυτόλυση των κυτταρικών δομικών πρωτεϊνών.

iv. Οι παράγοντες πήξης φαίνονται στο αίμα ως ζυμογόνο μορφή.

v. Παρόμοια ενεργοποίηση της πρόδρομης πρωτεΐνης γίνεται στην περίπτωση συμπληρώματος. Αυτές οι δραστηριότητες χρειάζονται μόνο μερικές φορές. Αλλά όταν χρειάζεται, μεγάλος αριθμός των μορίων πρέπει να παράγονται ακαριαία.

Ανταγωνιστική Αναστολή

- Εδώ τα μόρια αναστολέων ανταγωνίζονται το κανονικά μόρια υποστρώματος για προσκόλληση με το ενεργό σημείο του ενζύμου.

E + S → ES E + P

E + I → EI

- Εφόσον το EI (σύμπλεγμα ενζύμου-αναστολέα) μπορεί να αντιδράσει μόνο για να αναμορφώσει το ένζυμο και τον αναστολέα, ο αριθμός των μορίων ενζύμου που είναι διαθέσιμα για το σχηματισμό ES είναι μειωμένος. Έστω 100 μόρια υποστρώματος και 100 μόρια αναστολέα ανταγωνίζονται 100 μόρια του ενζύμου. Άρα, τα μισά μόρια ενζύμου παγιδεύονται από τον αναστολέα και μόνο τα μισά είναι διαθέσιμα για κατάλυση για να σχηματίσουν το προϊόν.
- iii. Εφόσον η αποτελεσματική συγκέντρωση του ενζύμου μειώνεται, η ταχύτητα της αντίδρασης μειώνεται.

Co-funded by the Erasmus+ Programme of the European Union

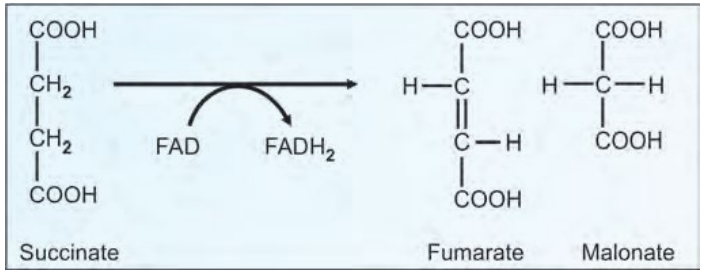
Ανταγωνιστική Αναστολή

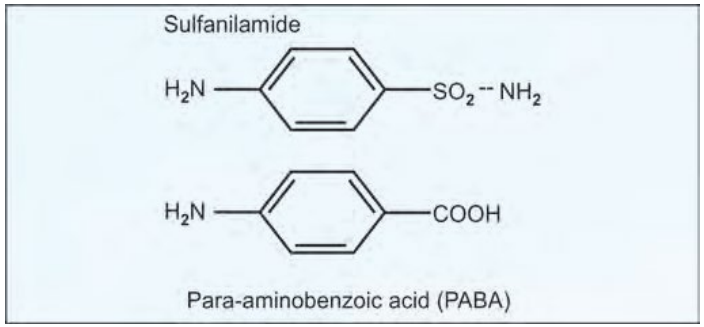
- Στην ανταγωνιστική αναστολή, ο αναστολέας θα είναι δομικό ανάλογο του υποστρώματος. Θα υπάρχει ομοιότητα στην τρισδιάστατη δομή μεταξύ υπόστρωμα (S) και αναστολέα (I). Για παράδειγμα, η αντίδραση της succinate αφυδρογονάσης αναστέλλεται από μηλονικό.
- Η ανταγωνιστική αναστολή είναι συνήθως αναστρέψιμη. Ή αλλιώς, η περίσσεια υποστρώματος καταργεί την αναστολή. Στο προηγούμενο παράδειγμα 100 moles E και 100 moles του I, αν προστεθούν 900 moles S, μόνο το 1/10 από τα μόρια ενζύμου συνδέονται με τον αναστολέα και το 90% δουλεύουν με υπόστρωμα. Έτσι από 50% αναστολή στο πρώτο παράδειγμα, μειώνεται τώρα στο 10% αναστολή.

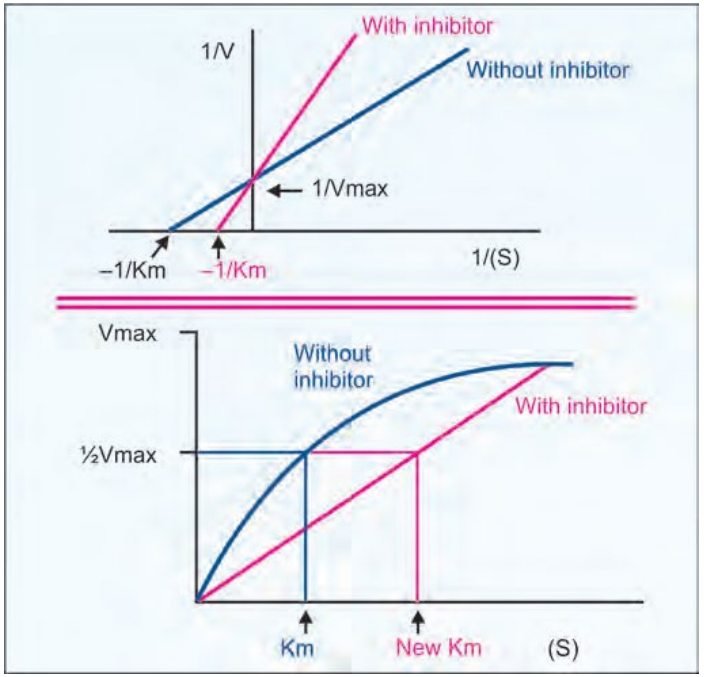
Ανταγωνιστική Αναστολή

- Από τα γραφήματα είναι προφανές ότι στην περίπτωση ανταγωνιστικής αναστολής, το Km αυξάνεται παρουσία ανταγωνιστικού αναστολέα. Άρα ο ανταγωνιστικός αναστολέας προφανώς αυξάνει το Km. Με άλλα λόγια, η συγγένεια του ενζύμου προς το υπόστρωμα προφανώς μειώθηκε παρουσία του αναστολέα.
- Αλλά το Vmax δεν έχει αλλάξει.

Ανταγωνιστική Αναστολή







Μη ανταγωνιστική αναστολή (μη αναστρέψιμη)

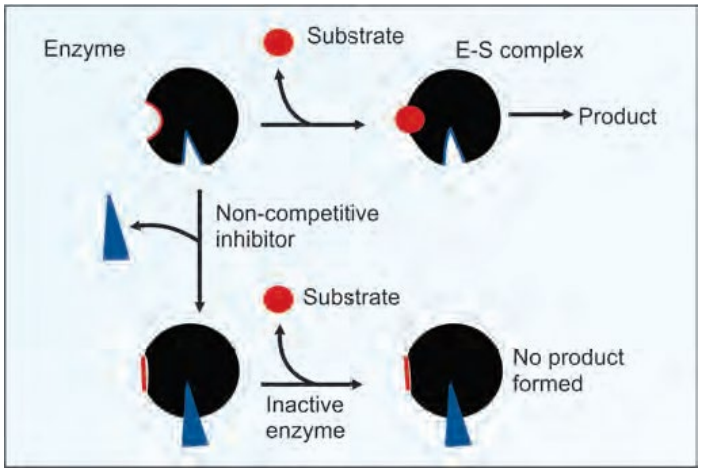
- Μια ποικιλία από δηλητήρια, όπως το ιωδοοξικό, τα βαριά μεταλλικά ιόντα (μόλυβδος, υδράργυρος) και οι οξειδωτικοί παράγοντες δρουν ως μη αναστρέψιμοι, μη ανταγωνιστικοί αναστολείς. Εκεί δεν υπάρχει ανταγωνισμός μεταξύ υποστρώματος και αναστολέα.
- Ο αναστολέας συνήθως συνδέεται σε διαφορετική περιοχή στο ένζυμο, εκτός από τη θέση δέσμευσης του υποστρώματος. Αυτοί οι αναστολείς δεν έχουν καμία δομική ομοιότητα με το υπόστρωμα, αύξηση της συγκέντρωσης του υποστρώματος γενικά δεν ανακουφίζει αυτή την αναστολή. Παραδείγματα είναι:

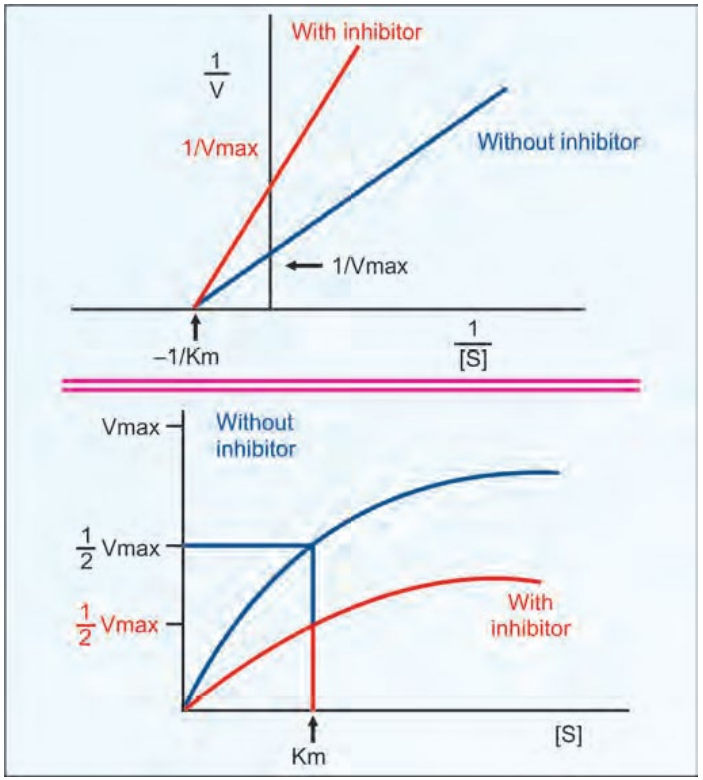
Μη ανταγωνιστική αναστολή (μη αναστρέψιμη)

Παραδείγματα είναι:

- Το κυάνιο αναστέλλει την οξειδάση του κυτοχρώματος.
- Το φθόριο θα αφαιρέσει το μαγνήσιο και το μαγγάνιο και έτσι θα αναστέλλουν το ένζυμο, την ενολάση και κατά συνέπεια τη γλυκόλυση.
- Το ιωδοοξικό θα ανέστειλε ένζυμα που έχουν –SH ομάδα στα ενεργά τους κέντρα.
- Χρησιμοποιείται BAL (British Anti-Lewisite, dimercaprol) ως αντίδοτο για δηλητηρίαση από βαρέα μέταλλα. Αυτά τα μέταλλα δρουν ως ενζυμικά δηλητήρια αντιδρώντας με το Ομάδα SH. Το BAL έχει αρκετές ομάδες SH με τις οποίες τα ιόντα βαρέων μετάλλων μπορούν να αντιδράσουν και ως εκ τούτου τους μειώνονται οι δηλητηριώδεις επιπτώσεις.
- Το ένζυμο ακετυλοχολινεστεράση διασπά την ακετυλοχολίνη για να σχηματίσει οξικό και χολίνη και επομένως τερματίζει τη δράση της ακετυλοχολίνης. Βέβαια χημικές ουσίες π.χ. διισοπροπυλ φθοροφωσφορικά (DFP) δεσμεύεται στην ενεργή θέση, σερίνη του ακετυλοχολινεστεράση. Ως αποτέλεσμα ακετυλοχολίνη συσσωρεύει και υπερδιεγείρει το αυτόνομο νευρικό σύστημα συμπεριλαμβανομένης της καρδιάς, των αιμοφόρων αγγείων και αδένες. Αυτό οδηγεί σε έμετο, σιελόρροια, εφίδρωση, και στη χειρότερη περίπτωση ακόμη και θάνατο. Το DFP σχηματίζει έναν μη αναστρέψιμο ομοιοπολικό δεσμό με την ακετυλοχολινεστεράση, και η δραστηριότητα μπορεί να ανακτηθεί μόνο εάν νέο ένζυμο συντίθεται.

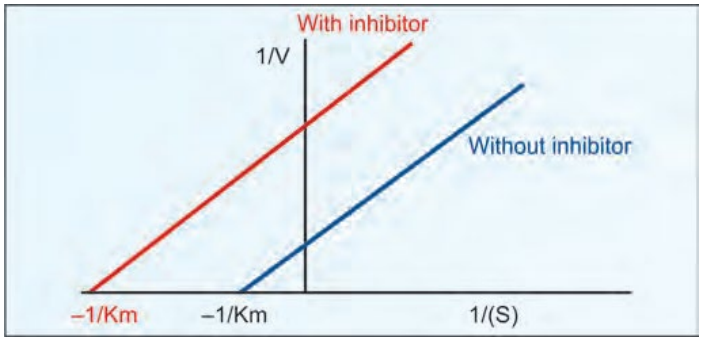
Co-funded by the Erasmus+ Programme of the European Union





Μη ανταγωνιστική αναστολή

Εδώ ο αναστολέας δεν έχει καμία συγγένεια με το ελεύθερο ένζυμο. Ανασταλτικός παράγοντας δεσμεύεται στο σύμπλεγμα ενζύμου-υποστρώματος, αλλά όχι στο ελεύθερο ένζυμο. Σε τέτοιες περιπτώσεις τόσο το Vmax όσο και το Km μειώνονται. Αναστολή της αλκαλικής φωσφατάσης του πλακούντα (ισοένζυμο Regan) από τη φαινυλαλανίνη είναι ένα παράδειγμα μη ανταγωνιστικής αναστολής.



Co-funded by the Erasmus+ Programme of the European Union

Αναστολή αυτοκτονίας

i . Είναι ένας ειδικός τύπος μη αναστρέψιμης αναστολής της ενζυμικής δραστηριότητας. Είναι επίσης γνωστή ως αδρανοποίηση βάσει μηχανισμών. Ο αναστολέας κάνει χρήση του μηχανισμού αντίδρασης του ίδιου του ενζύμου για να το απενεργοποιήσει (απενεργοποίηση βάσει μηχανισμού).

ii. Στην αναστολή της αυτοκτονίας, το δομικό ανάλογο μετατρέπεται σε περισσότερο αποτελεσματικό αναστολέα με τη βοήθεια του ενζύμου που πρόκειται να ανασταλεί. Η ένωση που μοιάζει με υπόστρωμα αρχικά δεσμεύεται με το ένζυμο και τα πρώτα βήματα της διαδρομής καταλύονται.

iii. Αυτό το νέο προϊόν συνδέεται μη αναστρέψιμα με το ένζυμο και αναστέλλει περαιτέρω αντιδράσεις.

Αναστολή αυτοκτονίας

iv. Για παράδειγμα, η αποκαρβοξυλάση ορνιθίνης (ODC) καταλύει το μετατροπή της ορνιθίνης σε πουτρεσκίνη που είναι απαραίτητη για την πολυαμίνη σύνθεση. Όταν το ODC στο τρυπανόσωμα αναστέλλεται ο πολλαπλασιασμός του παρασίτου σταματά. Επομένως αναστολείς του ενζύμου ODC όπως η διφθορομεθυλορνιθίνη (DFMO) έχει βρεθεί ότι είναι αποτελεσματικό κατά της τρυπανοσωμίασης τρυπανοσωμίαση (ασθένεια του ύπνου). Το DFMO είναι αρχικά αδρανές, αλλά δεσμεύεται με το ένζυμο, σχηματίζει μη αναστρέψιμο ομοιοπολικό σύμπλοκο με το συνένζυμο (φωσφορική πυριδοξάλη) και τα υπολείμματα αμινοξέων του το ένζυμο. Στα κύτταρα θηλαστικών, ο ρυθμός ανακύκλωσης του ODC είναι πολύ υψηλό, και έτσι η αναστολή από το DFMO είναι μόνο παροδική. DFMO λοιπόν σκοτώνει τα παράσιτα χωρίς παρενέργειες στον ασθενή.

v. Παρόμοιος μηχανισμός παρατηρείται και στην περίπτωση της αλλοπουρινόλης που οξειδώνεται από την οξειδάση της ξανθίνης σε αλλοξανθίνη που είναι περισσότερο ισχυρός αναστολέας της οξειδάσης της ξανθίνης.

vi. Η αντιφλεγμονώδης δράση της Ασπιρίνης βασίζεται επίσης στην αναχαίτηση αυτοκτονίας. Το αραχιδονικό οξύ μετατρέπεται σε προσταγλανδίνη από το ένζυμο Κυκλο-οξυγενάση. Ακετυλικές ασπιρίνες ένα υπόλειμμα σερίνης στο ενεργό κέντρο της κυκλο-οξυγενάσης, επομένως η σύνθεση των προσταγλανδινών αναστέλλεται και έτσι η φλεγμονή υποχωρεί.

Co-funded by the Erasmus+ Programme of the European Union

Αλλοστερικός Κανονισμός

i. Το αλλοστερικό ένζυμο έχει μία καταλυτική θέση όπου το το υπόστρωμα δεσμεύεται και μια άλλη ξεχωριστή αλλοστερική θέση όπου συνδέεται ο τροποποιητής (allo = άλλο).

ii. Οι αλλοστερικές θέσεις και οι θέσεις δέσμευσης υποστρώματος μπορεί ή όχι να είναι φυσικά δίπλα.

iii. Η δέσμευση του ρυθμιστικού μορίου μπορεί είτε ενισχύουν τη δραστηριότητα του ενζύμου (αλλοστερική ενεργοποίηση), ή αναστέλλουν τη δραστηριότητα του ενζύμου (αλλοστερική αναστολή).

iv. Στην πρώτη περίπτωση, το ρυθμιστικό μόριο είναι γνωστό ως θετικός τροποποιητής και στην τελευταία περίπτωση ως ο αρνητικός τροποποιητής.

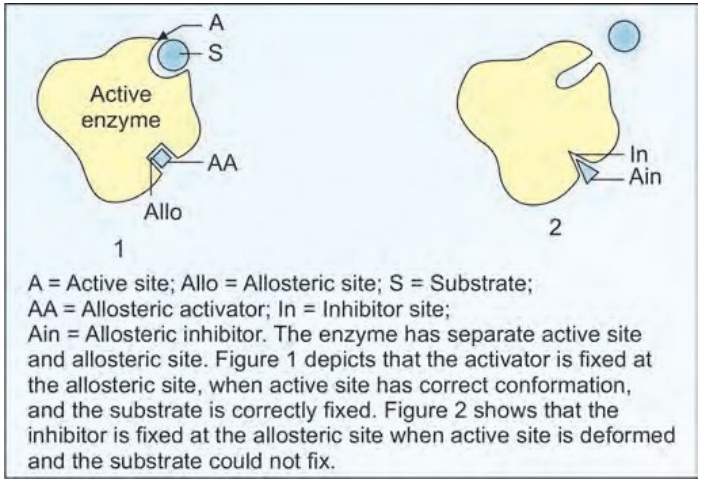
Co-funded by the Erasmus+ Programme of the European Union

Αλλοστερικός Κανονισμός

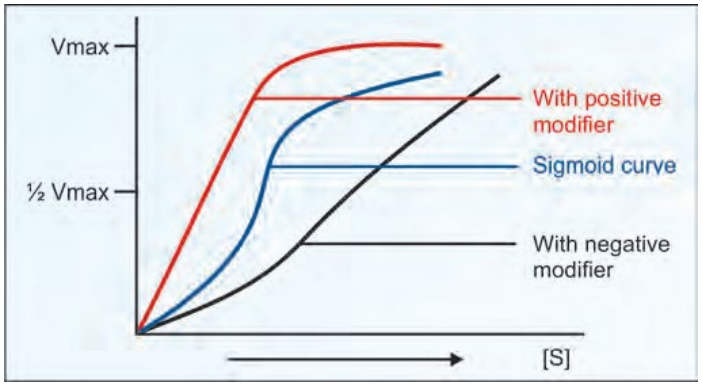
v. Η πρόσδεση του υποστρώματος σε μια από τις υπομονάδες του ένζυμο μπορεί να ενισχύσει τη σύνδεση του υποστρώματος από άλλες υπομονάδες. Αυτό το αποτέλεσμα λέγεται ότι είναι η θετική συνεργασία. Εάν η σύνδεση του υποστρώματος σε μια από τις υπομονάδες μειώνει την απληστία της δέσμευσης του υποστρώματος από άλλα τοποθεσίες, το αποτέλεσμα ονομάζεται αρνητική συνεργασία.

vi. Στις περισσότερες περιπτώσεις παρατηρείται συνδυασμός με αποτέλεσμα μια καμπύλη σε σχήμα σιγμοειδούς.

Αλλοστερικός Κανονισμός

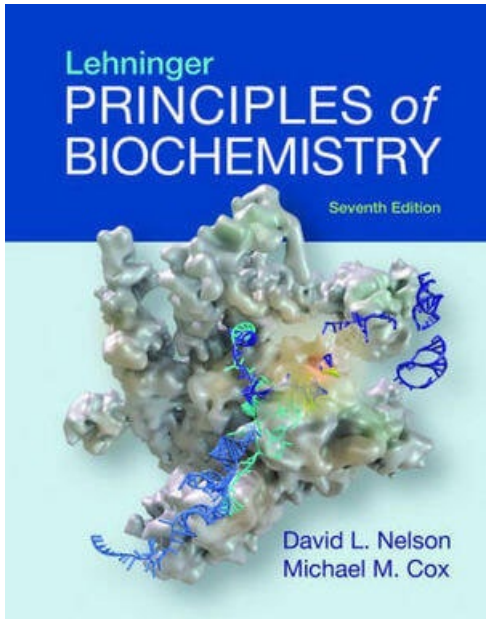


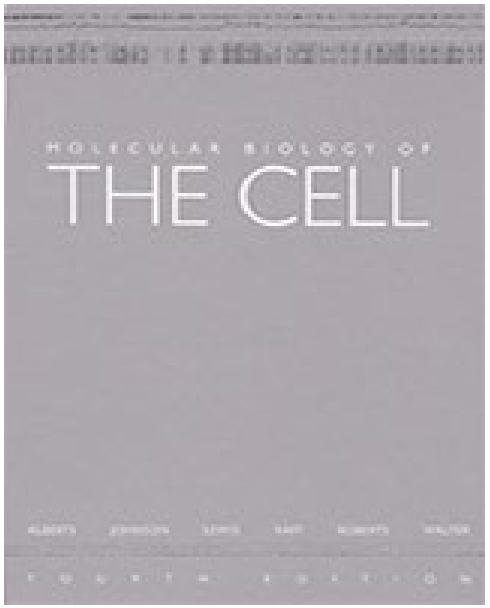
A = Active site; Allo = Allosteric site; S = Substrate; AA = Allosteric activator; In = Inhibitor site; Ain = Allosteric inhibitor. The enzyme has separate active site and allosteric site. Figure 1 depicts that the activator is fixed at the allosteric site, when active site has correct conformation, and the substrate is correctly fixed. Figure 2 shows that the inhibitor is fixed at the allosteric site when active site is deformed and the substrate could not fix.



Ερωτήσεις

- Τι είδους ρυθμίσεις για τη δράση των ενζύμων υπάρχουν;
- Εξηγήστε τον μηχανισμό ενεργοποίησης.
- Εξηγήστε τον μηχανισμό αναστολής.
- Εξηγήστε τον αλλοστερισμό .





Lehninger , AL, Nelson, DL, & Cox, MM (2000). Αρχές βιοχημείας Lehninger . Νέα Υόρκη: Worth Publishers.

Μοριακή Βιολογία του Κυττάρου. 4η έκδοση.
Alberts B, Johnson A, Lewis J, et al. Νέα Υόρκη: Garland Science; 2002.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 3. Πρωτεΐνες και ένζυμα

Μάθημα 9. Ταξινόμηση ενζύμων

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Ταξινόμηση IUBMB

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Όταν οι πρωτο-απομονώθηκαν ορισμένα ένζυμα, δόθηκαν περίεργα ονόματα. Μερικά από αυτά, όπως Pepsin, Trypsin, Chymotrypsin κ.λπ. χρησιμοποιούνται ακόμα. Αργότερα, συμφωνήθηκε να ονομαστούν τα ένζυμα προσθέτοντας το επίθημα "-άση " στο υπόστρωμα. Έτσι, το ένζυμο Λακτάση δρα στο υπόστρωμα λακτόζη και σχηματίζονται τα προϊόντα γλυκόζη και γαλακτόζη. Τα ένζυμα που υδρολύουν το άμυλο (αμυλόζη) ονομάζονται αμυλάσες. Αυτές που αφυδρογονώνουν τα υποστρώματα ονομάζονται αφυδρογονάσες. Αυτά είναι γνωστά ως τα ασήμαντα ονόματα των ενζύμων.

Σύστημα ταξινόμησης IUBMB

Η Διεθνής Ένωση Βιοχημείας και Μοριακής Βιολογίας (IUBMB) το 1964, (τροποποιήθηκε το 1972 και το 1978), πρότεινε το σύστημα ονοματολογίας των ενζύμων IUBMB. Είναι περίπλοκο και δυσκίνητο. αλλά σαφές. Σύμφωνα με αυτό το σύστημα, το όνομα ξεκινά με EC (κατηγορία ενζύμων) ακολουθούμενο από 4 ψηφία.

Το πρώτο ψηφίο αντιπροσωπεύει την τάξη

Το δεύτερο ψηφίο αντιπροσωπεύει την υποκλάση

Το τρίτο ψηφίο είναι η υπο-υποκατηγορία ή υποομάδα

Το τέταρτο ψηφίο δίνει τον αριθμό του συγκεκριμένου ενζύμου στη λίστα.

Τα ένζυμα ομαδοποιούνται στις ακόλουθες έξι κύριες τάξεις (κλάσεις)

Ταξινόμηση ενζύμων

Κλάση 1 : Οξειδοαναγωγάσες: Μεταφορά υδρογόνου ή προσθήκη οξυγόνου. π.χ. γαλακτική αφυδρογονάση (NAD); Γλυκόζη-6-φωσφορική αφυδρογονάση (NADP); Ηλεκτρική αφυδρογονάση (FAD); διοξυγενάσες.

Κλάση 2: Τρανσφεράσες: Μεταφορά ομάδων εκτός του υδρογόνου. Παράδειγμα, Αμινοτρανσφεράση. (Υποκατηγορία: Κινάση, μεταφορά φωσφορυλομάδας από ATP, π.χ. Εξοκινάση).

Κλάση 3: Υδρολάσες: Διαχωρίστε τον δεσμό και προσθέστε νερό. π.χ. εστεράση ακετυλοχολίνης. Τρυψίνη.

Κλάση 4 : Λυάσες: Κόψτε χωρίς προσθήκη νερού, π.χ. HMG CoA λυάση; ATP Κιτρική λυάση. (Υποκατηγορία: Υδρατάση, προσθέστε νερό σε διπλό δεσμό).

Κλάση 5: Ισομεράσες: Ενδομοριακές μεταφορές. Περιλαμβάνουν ρακεμάσες και επιμεράσες. Παράδειγμα, ισομεράση φωσφορικής τριόζης.

Κλάση 6: Λιγκάσες: εξαρτώμενη από το ATP συμπύκνωση δύο μορίων, π.χ. Ακετυλο CoA καρβοξυλάση. Συνθετάση γλουταμίνης; PRPP συνθετάση.

Κατηγορία 1: Οξειδορεδουκτάσες

Αυτή η ομάδα ενζύμων θα καταλύσει την οξείδωση ενός υποστρώματος με ταυτόχρονη αναγωγή ενός άλλου υποστρώματος ή συνενζύμου. Αυτό μπορεί να αναπαρασταθεί ως

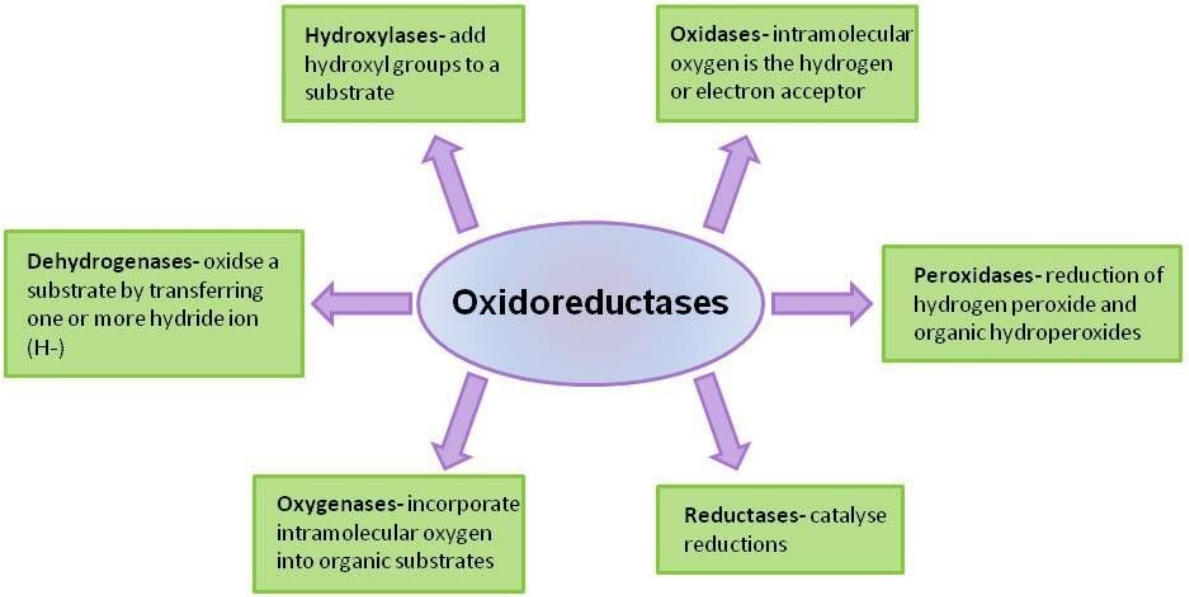
$AH_2 + B \rightarrow A + BH_2$

για παράδειγμα,

Αλκοόλη + $NAD^+ \rightarrow$ Αλδεϋδη + NADH + H^+

Το ένζυμο είναι η αφυδρογονάση της αλκοόλης. Το όνομα IUB είναι Alcohol-NAD-oxidoreductase. Ο κωδικός αριθμός είναι EC.1.1.1.1. Οι οξειδορεδουκτάσες μπορούν επίσης να οξειδώσουν υποστρώματα προσθέτοντας οξυγόνο, π.χ. οξειδάσες, οξυγενάσες και αφυδρογονάσες

Κατηγορία 1: Οξειδορεδουκτάσες



Co-funded by the Erasmus+ Programme of the European Union

Κλάση 2: Τρανφεράσες

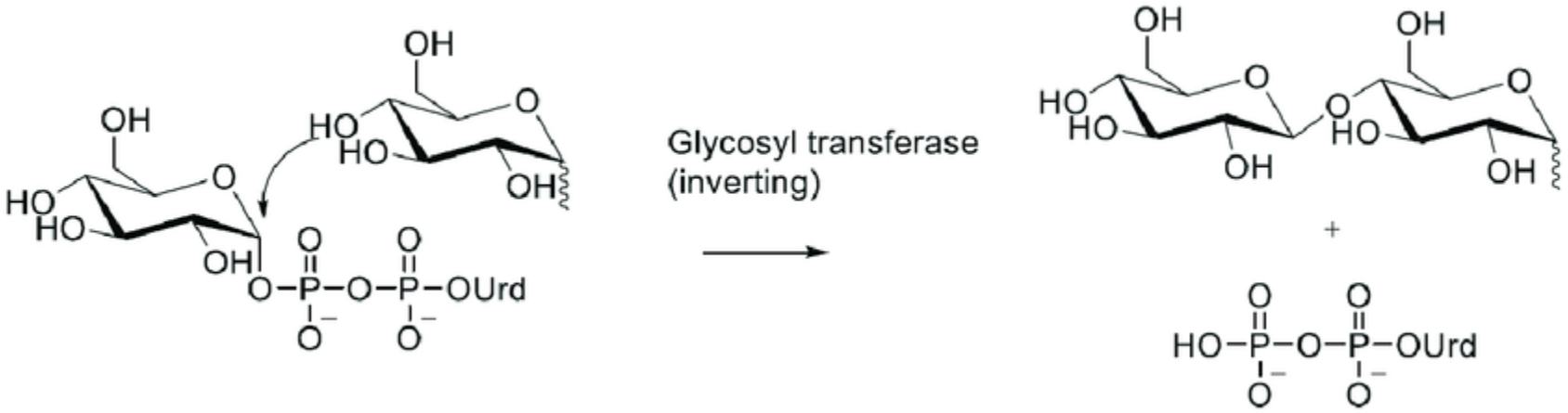
Αυτή η κατηγορία ενζύμων μεταφέρει μια ομάδα (εκτός από το υδρογόνο) από το υπόστρωμα σε ένα άλλο υπόστρωμα. Αυτό μπορεί να αναπαρασταθεί ως

AR + B → A + BR, Για παράδειγμα,

Εξόζη + ATP → Εξόζη-6-φωσφορική + ADP

Το όνομα του ενζύμου είναι εξοκινάση και η συστηματική ονομασία είναι ATP-Εξόζη-6-φωσφορική-τρανσφεράση

Κλάση 2: Τρανφεράσες



Co-funded by the Erasmus+ Programme of the European Union

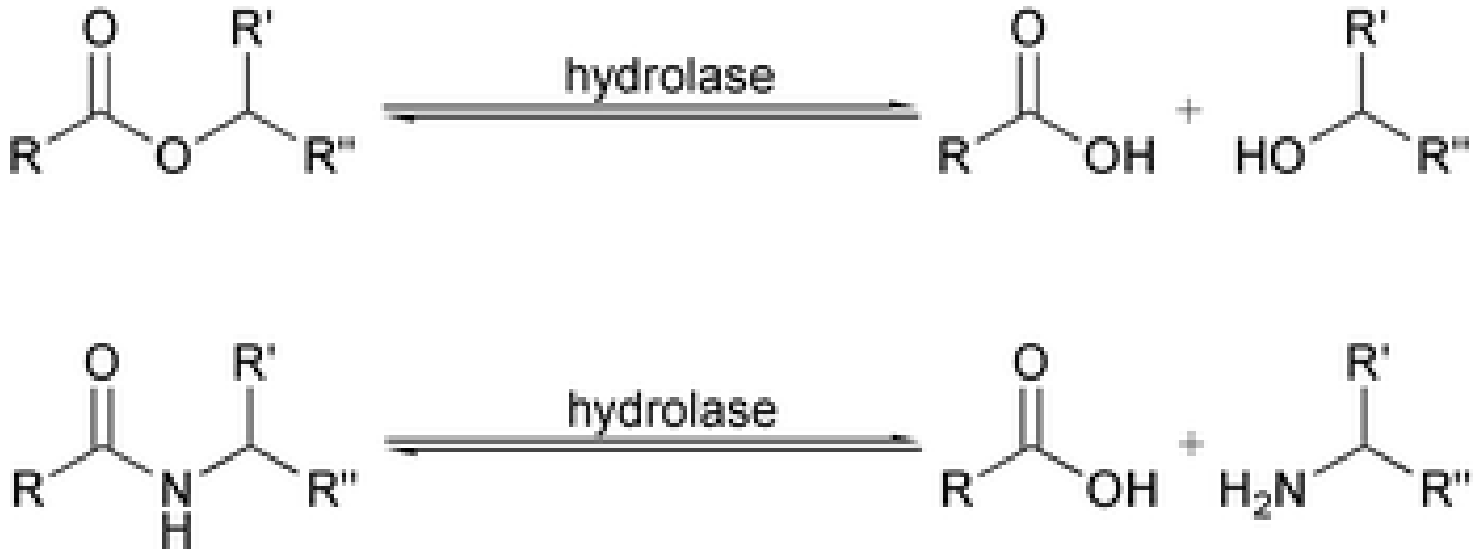
Κλάση 3: Υδρολάσες

Αυτή η κατηγορία ενζύμων μπορεί να υδρολύσει εστερικούς, αιθερικούς, πεπτιδικούς ή γλυκοσιδικούς δεσμούς προσθέτοντας νερό και στη συνέχεια σπάζοντας τον δεσμό.

Ακετυλοχολίνη + H_2O → Χολίνη + οξική

Το ένζυμο είναι η εστεράση της ακετυλοχολίνης ή η υδρολάση της ακετυλοχολίνης (συστηματική). Όλα τα πεπτικά ένζυμα είναι υδρολάσες.

Κλάση 3: Υδρολάσες

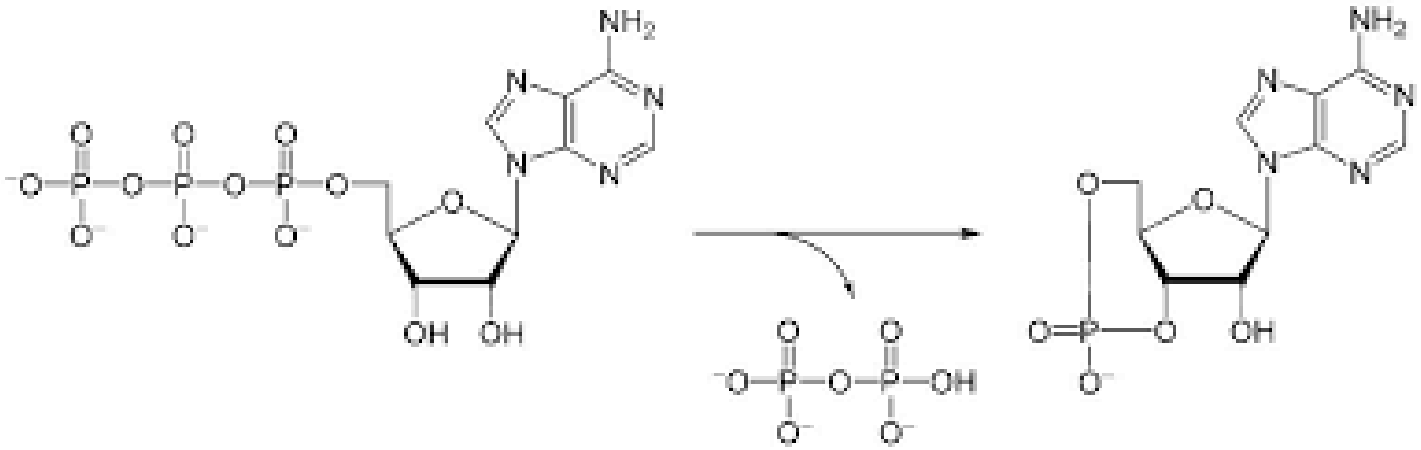


Κλάση 4: Λυάσες

Αυτά τα ένζυμα μπορούν να αφαιρέσουν ομάδες από υποστρώματα ή να σπάσουν δεσμούς με μηχανισμούς διαφορετικούς από την υδρόλυση. Για παράδειγμα,

Φρουκτόζη-1,6-διφωσφορική → Γλυκεραλδεΰδη-3-φωσφορική + διυδροξυ ακετόνη φωσφορική

Κλάση 4: Λυάσες



Κλάση 5: Ισομεράσες

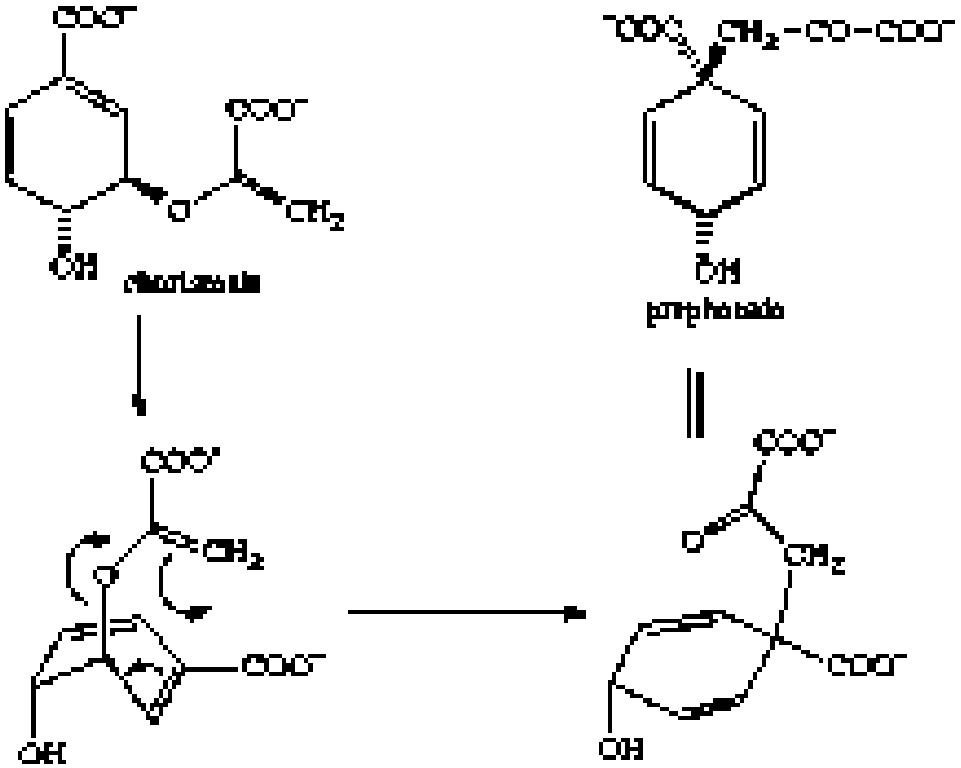
Αυτά τα ένζυμα μπορούν να παράγουν οπτικά, γεωμετρικά ή ισομερή θέσης υποστρωμάτων. Ρακεμάσες, επιμεράσες, ισομεράσες cistrans είναι παραδείγματα.

Γλυκεραλδεϋδη-3-φωσφορική → Διυδροξυ ακετόνη φωσφορικό άλας

Το ένζυμο είναι η τριόζη φωσφορική ισομεράση.

Co-funded by the Erasmus+ Programme of the European Union

Κλάση 5: Ισομεράσες



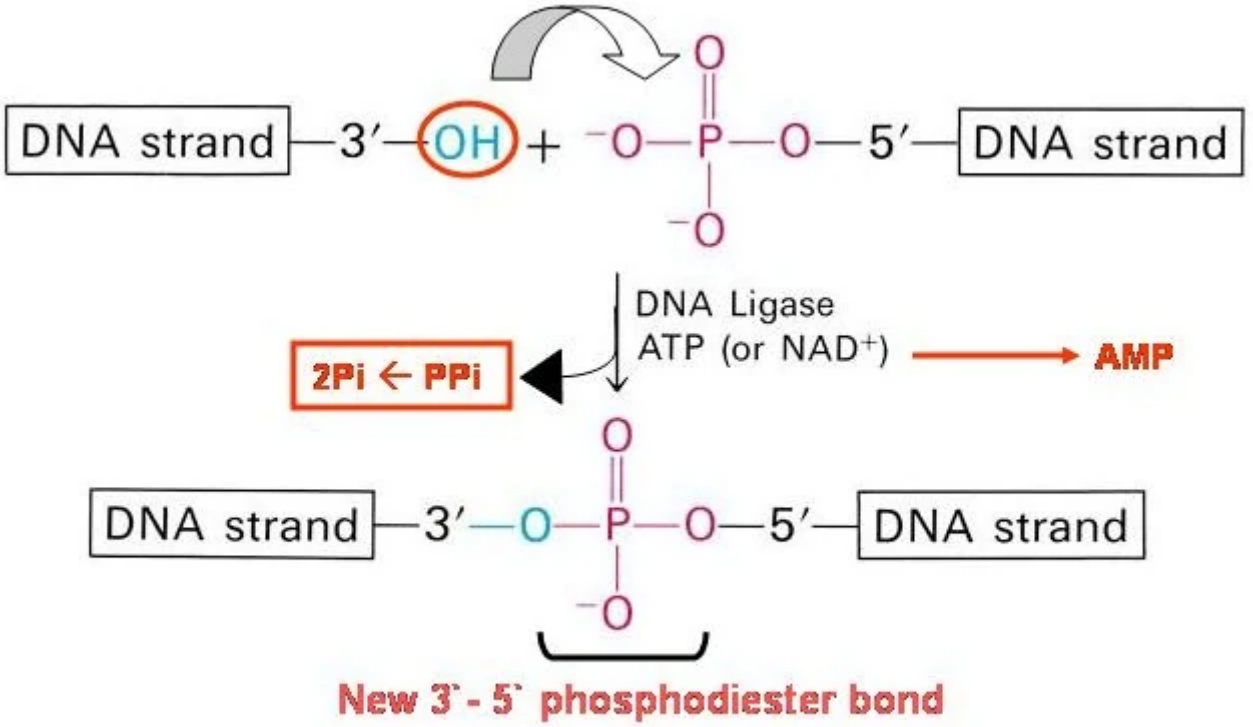
Κατηγορία 6: Λιγκάσες

Αυτά τα ένζυμα συνδέουν δύο υποστρώματα μεταξύ τους, συνήθως με την ταυτόχρονη υδρόλυση του ATP, (Λατινικά, Ligare = δέσμευση). Για παράδειγμα,

Ακετυλ CoA + CO2 + ATP → Malonyl CoA + ADP +Pi

Το ένζυμο είναι η καρβοξυλάση της ακετυλ CoA.

Κατηγορία 6: Λιγκάσες



Co-funded by the Erasmus+ Programme of the European Union

Η συνθετάση και η συνθάση είναι διαφορετικές

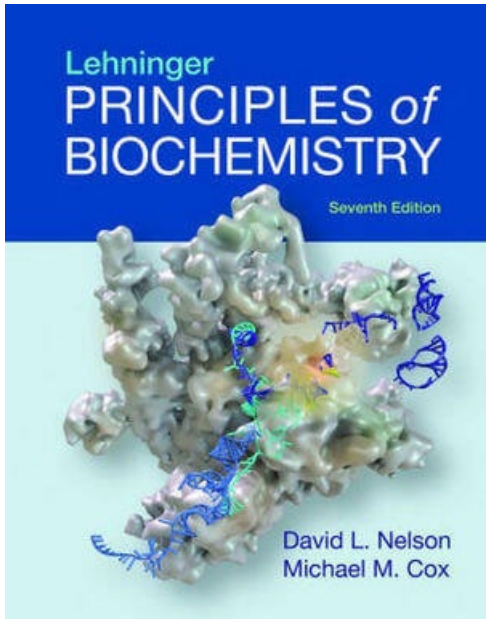
Οι συνθετάσες είναι ένζυμα εξαρτώμενα από το ATP που καταλύουν βιοσυνθετικές αντιδράσεις. ανήκουν στις Λιγκάσες (κατηγορία 6). Παραδείγματα είναι η καρβαμοϋλοφωσφορική συνθετάση. Συνθετάση ηλεκτρικής αργινίνης ; Συνθετάση PRPP και συνθετάση γλουταμίνης.

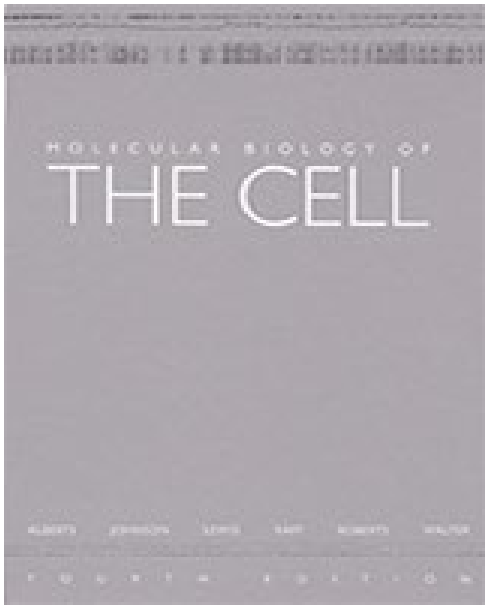
Οι συνθάσες είναι ένζυμα που καταλύουν βιοσυνθετικές αντιδράσεις. αλλά δεν απαιτούν απευθείας ATP. Ανήκουν σε άλλες τάξεις εκτός των Λιγκασών. Παραδείγματα είναι η συνθάση του γλυκογόνου και η συνθάση ALA.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Ποιο είναι το κύριο κριτήριο με το οποίο ταξινομούνται τα ένζυμα;
- Δώστε μερικά παραδείγματα από κάθε κατηγορία ενζύμων και εξηγήστε τη δράση εκεί.





Lehninger , AL, Nelson, DL, & Cox, MM (2000). Αρχές βιοχημείας Lehninger . Νέα Υόρκη: Worth Publishers.

Μοριακή Βιολογία του Κυττάρου. 4η έκδοση.
Alberts B, Johnson A, Lewis J, et al. Νέα Υόρκη: Garland Science; 2002.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 4. Βιοφυσικά προβλήματα

Μάθημα 1. Θερμοδυναμική

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Περιεχόμενα

- Εισαγωγή
- Βασικές Έννοιες και Ορισμοί
- Οι ιδιότητες μιας καθαρής ουσίας
- Εργασία και θερμότητα
- Ο Πρώτος Νόμος της Θερμοδυναμικής
- Ο δεύτερος και ο τρίτος νόμος της θερμοδυναμικής

Εισαγωγή

Η θερμοδυναμική είναι η επιστήμη που ασχολείται με τη θερμότητα και το έργο και αυτές τις ιδιότητες των ουσιών που έχουν σχέση με τη θερμότητα και το έργο. Όπως όλες οι επιστήμες, η βάση της θερμοδυναμικής είναι η πειραματική παρατήρηση.

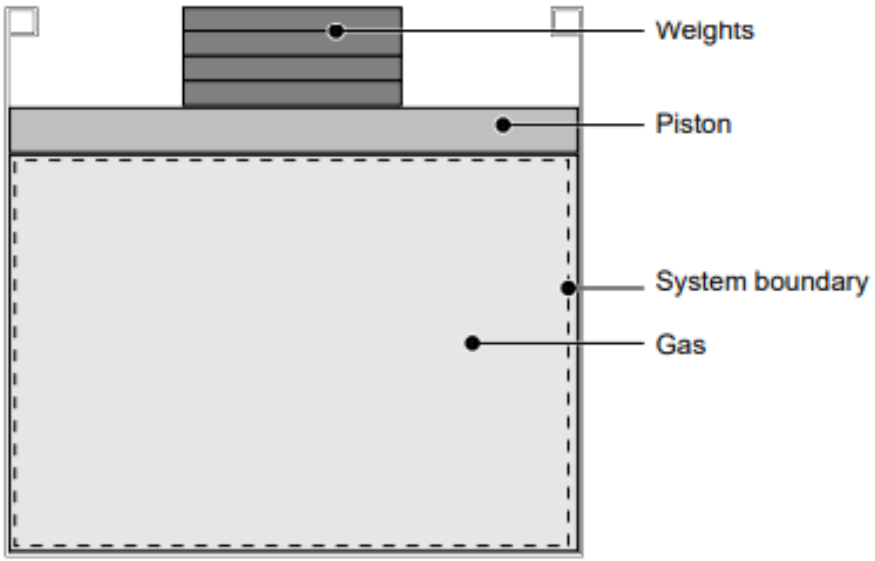
Στη συνέχεια, θα παρουσιάσουμε τους βασικούς θερμοδυναμικούς νόμους και θα τους εφαρμόσουμε σε μια σειρά αντιπροσωπευτικών παραδειγμάτων.

Βασικές Έννοιες και Ορισμοί

Θερμοδυναμικό σύστημα

Ένα **θερμοδυναμικό σύστημα** ορίζεται ως μια ποσότητα ύλης σταθερής μάζας και ταυτότητας στην οποία επικεντρώνεται η προσοχή για μελέτη. Οτιδήποτε εξωτερικό του συστήματος είναι το περιβάλλον και το σύστημα διαχωρίζεται από το περιβάλλον με τα όρια του συστήματος. Αυτά τα όρια μπορεί να είναι είτε κινητά είτε σταθερά.

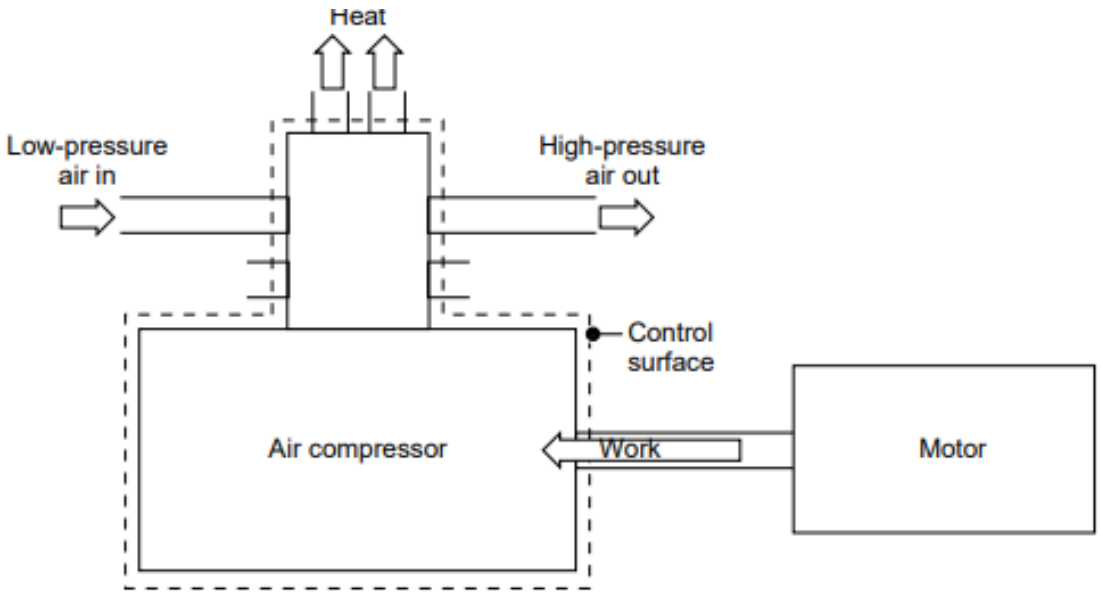
Example of a thermodynamic system.



Βασικές Έννοιες και Ορισμοί

Έλεγχος έντασης

Η μάζα, καθώς και η θερμότητα και το έργο (και η ορμή), μπορούν να ρέουν στην επιφάνεια ελέγχου



Βασικές Έννοιες και Ορισμοί

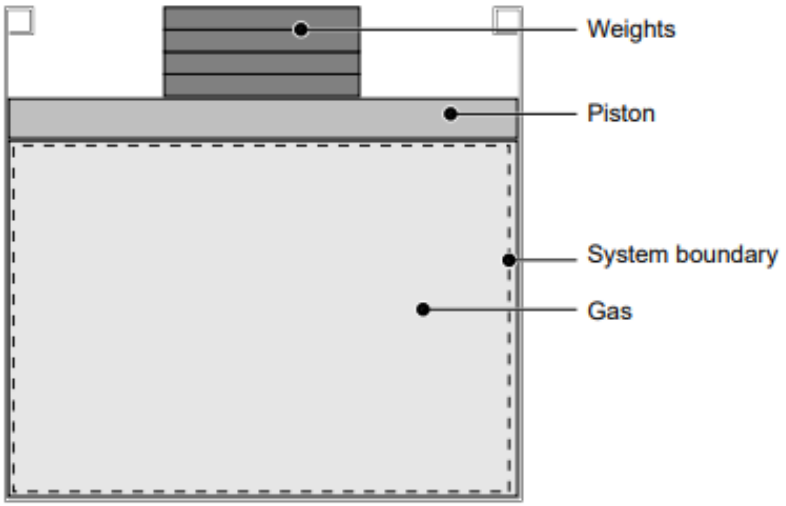
Ιδιότητες και κατάσταση μιας ουσίας

Μια δεδομένη μάζα νερού μπορεί να υπάρχει σε διάφορες μορφές. Εάν είναι υγρό, μπορεί να γίνει ατμός όταν θερμαίνεται ή στερεό όταν ψύχεται. Έτσι, μιλάμε για τις διάφορες Φάσεις μιας ουσίας. Ως φάση ορίζεται μια ποσότητα ύλης που είναι ομοιογενής σε όλη την έκταση. Όταν υπάρχουν περισσότερες από μία φάσεις, οι φάσεις διαχωρίζονται μεταξύ τους με τα όρια των φάσεων. Σε κάθε φάση η ουσία μπορεί να υπάρχει σε διάφορες πιέσεις και θερμοκρασίες ή, για να χρησιμοποιήσουμε τον θερμοδυναμικό όρο, σε διάφορες καταστάσεις. Η κατάσταση μπορεί να προσδιοριστεί ή να περιγραφεί από ορισμένες παρατηρήσιμες, μακροσκοπικές ιδιότητες. Μερικές γνωστές είναι η θερμοκρασία, η πίεση και η πυκνότητα.

Βασικές Έννοιες και Ορισμοί

Διαδικασία και κύκλοι

Κάθε φορά που αλλάζει μία ή περισσότερες από τις ιδιότητες ενός συστήματος, έχει συμβεί αλλαγή κατάστασης. Η διαδρομή της διαδοχής των καταστάσεων από την οποία διέρχεται το σύστημα ονομάζεται διαδικασία. Όταν ένα σύστημα σε μια δεδομένη αρχική κατάσταση περνά από μια σειρά από διαφορετικές αλλαγές κατάστασης ή διεργασιών και τελικά επιστρέφει στην αρχική του κατάσταση, το σύστημα έχει υποστεί έναν κύκλο. Επομένως, στο τέλος ενός κύκλου, όλες οι ιδιότητες έχουν την ίδια τιμή που είχαν στην αρχή. Ο ατμός (νερό) που κυκλοφορεί μέσω ενός ατμοηλεκτρικού σταθμού (όπως η συμβατική πλευρά ενός πυρηνικού αντιδραστήρα) υφίσταται έναν κύκλο.



Βασικές Έννοιες και Ορισμοί

Ενέργεια

Μία από τις πολύ σημαντικές έννοιες σε μια μελέτη της θερμοδυναμικής είναι η έννοια της ενέργειας. Η ενέργεια είναι μια θεμελιώδης έννοια, όπως η μάζα ή η δύναμη και, όπως συμβαίνει συχνά με τέτοιες έννοιες, είναι πολύ δύσκολο να οριστεί. Η ενέργεια ορίζεται ως η ικανότητα να παράγει ένα αποτέλεσμα. Είναι σημαντικό να σημειωθεί ότι η ενέργεια μπορεί να αποθηκευτεί μέσα σε ένα σύστημα και μπορεί να μεταφερθεί (ως θερμότητα, για παράδειγμα) από το ένα σύστημα στο άλλο.

Co-funded by the Erasmus+ Programme of the European Union

Βασικές Έννοιες και Ορισμοί

Ειδικός Όγκος

Ο ειδικός όγκος μιας ουσίας ορίζεται ως ο όγκος ανά μονάδα μάζας και δίνεται με το σύμβολο v.

$$v = \delta V / \delta m [m^3/kg]$$

Η πυκνότητα μιας ουσίας (ρ) ορίζεται ως η μάζα ανά μονάδα όγκου, και επομένως είναι η αντίστροφη του ειδικού όγκου.

$$\rho = 1/v [kg/m^3]$$

Co-funded by the Erasmus+ Programme of the European Union

Βασικές Έννοιες και Ορισμοί

Πίεση

Η πίεση ορίζεται ως η κάθετη στην επιφάνεια συνιστώσα της δύναμης ανά μονάδα επιφάνειας.

$$p = \frac{\delta \Gamma_n}{\delta A} \quad \left[\frac{N}{m^2} \text{ or } Pa\right]$$

Βασικές Έννοιες και Ορισμοί

Θερμοκρασία

Αν και η θερμοκρασία είναι μια οικεία ιδιότητα, ο ακριβής ορισμός της είναι δύσκολος. Έτσι, ορίζουμε την ισότητα των θερμοκρασιών. Δύο σώματα έχουν ίσες θερμοκρασίες εάν, όταν βρίσκονται σε θερμική επικοινωνία, δεν υπάρχει αλλαγή στην παρατηρήσιμη ιδιότητα.

Βασικές Έννοιες και Ορισμοί

Ο μηδενικός νόμος της θερμοδυναμικής

Ο μηδενικός νόμος της θερμοδυναμικής δηλώνει ότι όταν δύο σώματα έχουν ίσες θερμοκρασίες με ένα τρίτο σώμα, αυτά με τη σειρά τους έχουν ίσες θερμοκρασίες μεταξύ τους.

Οι ιδιότητες μιας καθαρής ουσίας

Η καθαρή ουσία

Καθαρή ουσία είναι αυτή που έχει ομοιογενή και αμετάβλητη χημική σύσταση. Μπορεί να υπάρχει σε περισσότερες από μία φάσεις, αλλά η χημική σύσταση είναι ίδια σε όλες τις φάσεις. Έτσι, το υγρό νερό, ένα μείγμα υγρού νερού και υδρατμών (ατμός) και ένα μείγμα πάγου και υγρού νερού είναι όλα καθαρές ουσίες. Κάθε φάση έχει την ίδια χημική σύσταση.

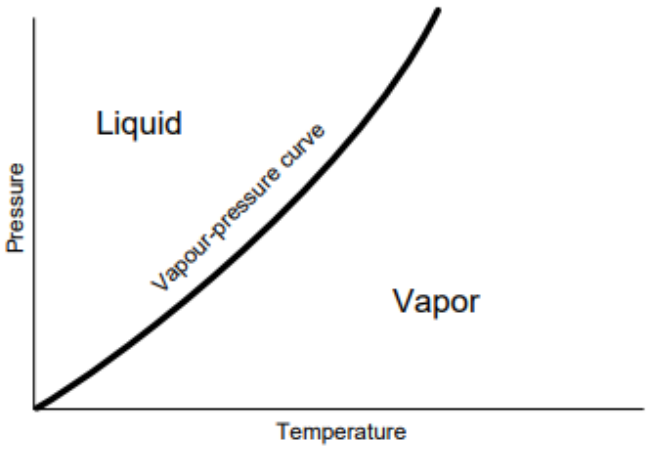
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Οι ιδιότητες μιας καθαρής ουσίας

Ο όρος **Θερμοκρασία Κορεσμού** δηλώνει τη θερμοκρασία στην οποία λαμβάνει χώρα η εξάτμιση σε μια δεδομένη πίεση. Αυτή η πίεση ονομάζεται **Πίεση Κορεσμού** για τη δεδομένη θερμοκρασία. Έτσι, για νερό στους 99,6 °C η πίεση κορεσμού είναι 0,1 MPa και για νερό σε 0,1 MPa η θερμοκρασία κορεσμού είναι 99,6 °C. Για μια καθαρή ουσία υπάρχει μια σαφής σχέση μεταξύ της πίεσης κορεσμού και της θερμοκρασίας κορεσμού.

Vapour-pressure curve of a pure substance.



Οι ιδιότητες μιας καθαρής ουσίας

Εξίσωση κατάστασης για τη φάση ατμού μιας απλής συμπιεστής ουσίας

Από πειραματικές παρατηρήσεις έχει διαπιστωθεί ότι η συμπεριφορά PvT των αερίων σε χαμηλή πυκνότητα δίνεται στενά από την ακόλουθη εξίσωση κατάστασης:

$$\mathbf{P\,v = RT}$$

όπου

$$\mathbf{R = 8{,}3144\ kN\ m/kmol\ K}$$

είναι η παγκόσμια σταθερά αερίων.

Εργασία και θερμότητα

Ορισμός της εργασίας

Το έργο ορίζεται ως μια δύναμη F που ενεργεί μέσω μιας μετατόπισης x, και η μετατόπιση είναι προς την κατεύθυνση της δύναμης.

$$W = \int_1^2 F dx$$

Η Θερμοδυναμική ορίζει το έργο ως εξής: το έργο γίνεται από ένα σύστημα εάν η μόνη επίδραση στο περιβάλλον (κάτι που είναι εξωτερικό του συστήματος) θα μπορούσε να είναι η ανύψωση ενός βάρους. Το έργο που γίνεται από ένα σύστημα θεωρείται θετικό και το έργο που γίνεται σε ένα σύστημα θεωρείται αρνητικό

Co-funded by the Erasmus+ Programme of the European Union

Εργασία και θερμότητα

Ορισμός της θερμότητας

Η θερμότητα ορίζεται ως η μορφή ενέργειας που μεταφέρεται μεταξύ των ορίων ενός συστήματος σε μια δεδομένη θερμοκρασία σε ένα άλλο σύστημα (ή το περιβάλλον) σε χαμηλότερη θερμοκρασία λόγω της διαφοράς θερμοκρασίας μεταξύ των δύο συστημάτων. Μια άλλη πτυχή αυτού του ορισμού της θερμότητας είναι ότι ένα σώμα δεν περιέχει ποτέ θερμότητα. Αντίθετα, η θερμότητα μπορεί να αναγνωριστεί μόνο όταν διασχίζει τα όρια. Έτσι, η θερμότητα είναι ένα παροδικό φαινόμενο.

Co-funded by the Erasmus+ Programme of the European Union

Πρώτος νόμος της θερμοδυναμικής

The change in internal energy of a system is equal to the heat added to the system minus the work done by the system.

$$\Delta U = Q - W$$

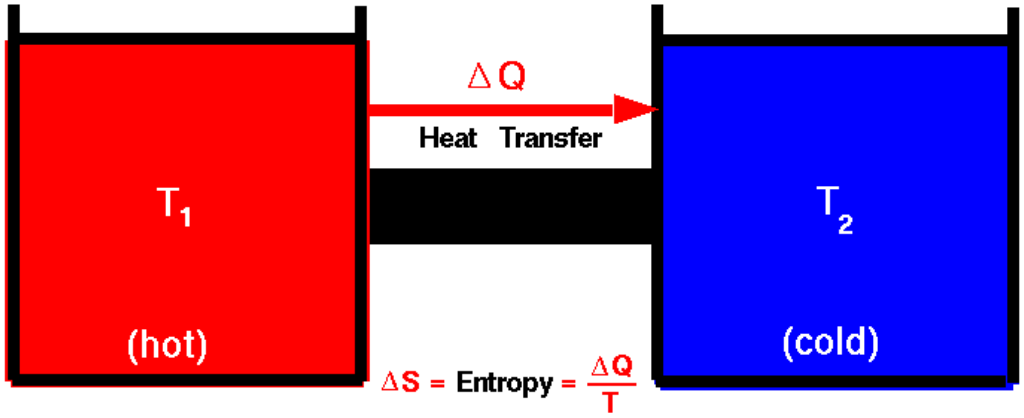
Change in internal energy

Heat added to the system

Work done by the system

Co-funded by the Erasmus+ Programme of the European Union

Δεύτερος νόμος της θερμοδυναμικής



There exists a useful thermodynamic variable called entropy (S). A natural process that starts in one equilibrium state and ends in another will go in the direction that causes the entropy of the system plus the environment to increase for an irreversible process and to remain constant for a reversible process.

$S_f = S_i$ (reversible) $S_f > S_i$ (irreversible)

Ενέργεια Gibbs

- Μια διαδικασία είναι αυθόρμητη εάν η συνολική μεταβολή της εντροπίας για το σύστημα είναι θετική.

Άρα η συνολική μεταβολή της εντροπίας, ΔS_{tot}, σε σταθερή θερμοκρασία και πίεση δίνεται από:

$$\Delta S_{tot} = \Delta S - \frac{\Delta H}{T}$$

$$\Delta S_{sur} = -\frac{\Delta H}{T}$$

Πολλαπλασιάζοντας και τις δύο πλευρές με −Τ π

$$-T\Delta S_{tot} = -T\Delta S + \Delta H$$

Εφόσον η θερμοκρασία είναι πάντα θετικός αριθμός, η αντίδραση είναι αυθόρμητη εάν ο όρος $-T\Delta S_{tot}$ είναι αρνητικός. Εάν η διαδικασία βρίσκεται σε ισορροπία τότε αυτός ο όρος είναι ίσος με μηδέν. Το γινόμενο της θερμοκρασίας και της εντροπίας έχει μονάδες ενέργειας και σχετίζεται με την ποσότητα ενέργειας που είναι διαθέσιμη για την εκτέλεση εργασίας. Αυτός ο όρος, $-T\Delta S_{tot}$, ονομάζεται συνήθως διαφορά ενέργειας Gibbs, ΔG, και γράφεται ως:

$$\Delta G = \Delta H - T\Delta S$$

- **εάν το ΔG είναι θετικό τότε η αντίδραση είναι δυσμενής και η αρχική κατάσταση ευνοείται,**
- **αν το ΔG είναι μηδέν η αντίδραση βρίσκεται σε ισορροπία, και**
- **μόνο αν το ΔG είναι αρνητικό η αντίδραση θα συμβεί αυθόρμητα.**

Σχέση μεταξύ της ενέργειας Gibbs και της σταθεράς ισορροπίας

- Για κάθε δεδομένη αντίδραση A ↔ B με σταθερά ισορροπίας K, η τιμή της σταθεράς ισορροπίας μπορεί να γραφτεί ως προς τη μεταβολή της ενέργειας Gibbs:

$$K = e^{-\Delta G/kT}$$

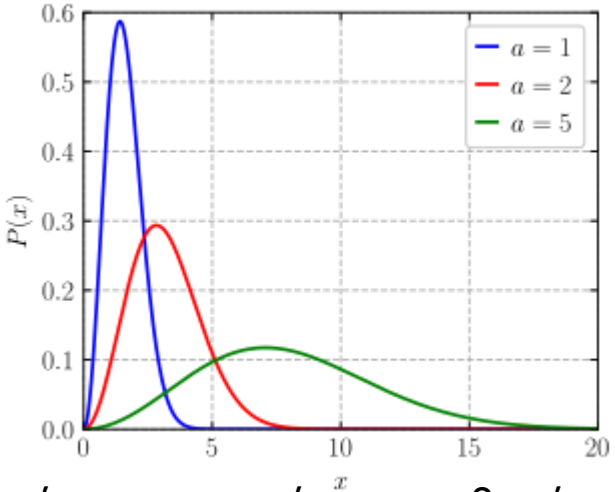
Relationships among the equilibrium constant, K, Gibbs energy change, ΔG, and direction of a chemical reaction*.

K	ΔG	Direction
>1.0	Negative	Proceeds forward
1.0	Zero	At equilibrium
<1.0	Positive	Proceeds in reverse

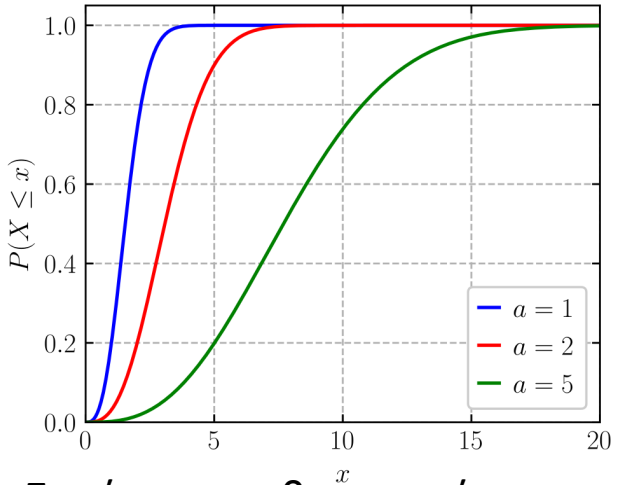
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Η κατανομή Boltzmann και η στατιστική θερμοδυναμική

Στη στατιστική μηχανική και τα μαθηματικά, μια κατανομή Boltzmann (ονομάζεται επίσης κατανομή Gibbs) είναι μια κατανομή πιθανότητας ή μέτρο πιθανότητας που δίνει την πιθανότητα ένα σύστημα να βρίσκεται σε μια συγκεκριμένη κατάσταση ως συνάρτηση της ενέργειας αυτής της κατάστασης και της θερμοκρασίας του συστήματος.



Συνάρτηση πυκνότητας πιθανότητας



Συνάρτηση αθροιστικής κατανομής

Co-funded by the Erasmus+ Programme of the European Union

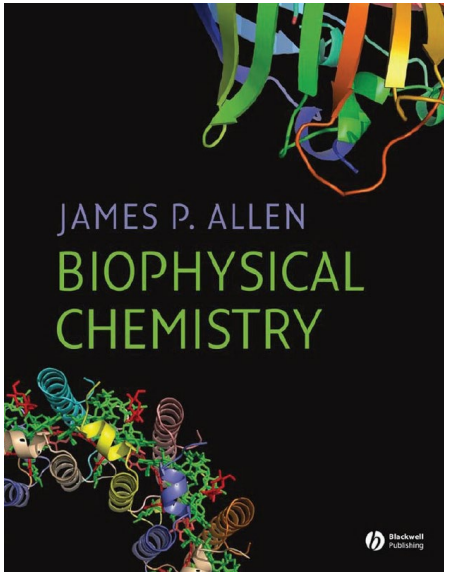
Ερωτήσεις

- Εξηγήστε την κατάσταση ισορροπίας. Συζητήστε επίσης τη θερμική, χημική και μηχανική ισορροπία με κατάλληλα παραδείγματα
- Εξηγήστε τον Μηδενικό νόμο της θερμοδυναμικής
- Εξηγήστε τον Πρώτο νόμο της θερμοδυναμικής
- Εξηγήστε τον Δεύτερο νόμο της θερμοδυναμικής

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

βιβλιογραφικές αναφορές



Βιοφυσική Χημεία,
Τζέιμς Π. Άλεν
2008 από την Blackwell Publishing

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 4. Βιοφυσικά προβλήματα

Μάθημα 2. Κινητική

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Τάξη και μοριακότητα
- Κινητική αντίδρασης πρώτης και δεύτερης τάξης
- Ψευδομονομοριακή αντίδραση
- Εξίσωση Arrhenius

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Η χημική κινητική είναι ο κλάδος της φυσικής χημείας που ασχολείται με τη μελέτη της ταχύτητας των χημικών αντιδράσεων. Τέτοιες μελέτες μας επιτρέπουν επίσης να κατανοήσουμε τον μηχανισμό με τον οποίο συμβαίνει η αντίδραση. Έτσι, στη χημική κινητική μπορούμε επίσης να προσδιορίσουμε το ρυθμό της χημικής αντίδρασης. Από κινητική άποψη οι αντιδράσεις ταξινομούνται σε δύο ομάδες: α) ομοιογενείς αντιδράσεις που συμβαίνουν εξ ολοκλήρου σε μία φάση β) ετερογενείς αντιδράσεις όπου ο μετασχηματισμός λαμβάνει χώρα στην επιφάνεια ενός καταλύτη ή στα τοιχώματα ενός δοχείου.

Ρυθμός αντίδρασης

Ο ρυθμός αντίδρασης, δηλαδή η ταχύτητα μιας αντίδρασης είναι η ποσότητα μιας χημικής αλλαγής που συμβαίνει ανά μονάδα χρόνου. Ο ρυθμός γενικά εκφράζεται ως η μείωση της συγκέντρωσης ενός αντιδρώντος ή ως η αύξηση στη συγκέντρωση του προϊόντος. Εάν C η συγκέντρωση ενός αντιδραστηρίου ανά πάσα στιγμή t είναι, ο ρυθμός είναι – dC /dt ή εάν η συγκέντρωση ενός προϊόντος είναι x σε οποιαδήποτε στιγμή t, ο ρυθμός θα είναι dx/dt .

Co-funded by the Erasmus+ Programme of the European Union

Ρυθμός αντίδρασης

Παράγοντες που επηρεάζουν τον ρυθμό αντίδρασης Ο ρυθμός μιας χημικής αντίδρασης επηρεάζεται από τους ακόλουθους παράγοντες

(i) Θερμοκρασία

(ii) Συγκέντρωση των αντιδρώντων

(iii) Φύση των αντιδρώντων

(iv) Καταλύτες

(v) Ακτινοβολία

Ρυθμός αντίδρασης

(i) **Θερμοκρασία** Στις περισσότερες περιπτώσεις, ο ρυθμός μιας αντίδρασης σε μια ομοιογενή αντίδραση διπλασιάζεται ή τριπλασιάζεται περίπου με μια αύξηση της θερμοκρασίας μόνο κατά 100 C. Σε ορισμένες περιπτώσεις η αύξηση των ρυθμών αντίδρασης είναι ακόμη μεγαλύτερη.

(ii) **Συγκέντρωση των αντιδρώντων** Σε σταθερή θερμοκρασία και απουσία καταλύτη, ο ρυθμός της δεδομένης αντίδρασης αυξάνεται με αυξημένη συγκέντρωση των αντιδρώντων. Με την αύξηση της συγκέντρωσης του αντιδρώντος ο αριθμός των μορίων ανά μονάδα όγκου αυξάνεται, επομένως αυξάνεται η συχνότητα σύγκρουσης, η οποία τελικά προκαλεί αυξημένο ρυθμό αντίδρασης.

(iii) **Φύση των αντιδρώντων** Μια χημική αντίδραση περιλαμβάνει την αναδιάταξη των ατόμων μεταξύ των μορίων που αντιδρούν στο προϊόν. Οι παλιοί δεσμοί σπάνε και δημιουργούνται νέοι δεσμοί. Κατά συνέπεια, η φύση και η ισχύς των δεσμών στα αντιδρώντα μόρια επηρεάζουν σε μεγάλο βαθμό τον ρυθμό μετατροπής του σε προϊόντα. Η αντίδραση στην οποία περιλαμβάνει μικρότερη αναδιάταξη δεσμού προχωρά πολύ πιο γρήγορα από εκείνες που περιλαμβάνουν μεγαλύτερη αναδιάταξη δεσμού.

Ρυθμός αντίδρασης

(iv) **Καταλύτες** Ο ρυθμός μιας χημικής αντίδρασης αυξάνεται παρουσία ενός καταλύτη που ενίσχυσε τελικά την ταχύτητα μιας χημικής αντίδρασης.

(v) **Ακτινοβολία** Ο ρυθμός ενός αριθμού χημικών αντιδράσεων αυξάνεται όταν ακτινοβολίες συγκεκριμένου μήκους κύματος απορροφώνται από τα αντιδρώντα μόρια. Τέτοιες αντιδράσεις ονομάζονται φωτοχημικές αντιδράσεις. Για παράδειγμα, το χλώριο μπορεί να αναμιχθεί με ασφάλεια με υδρογόνο στο σκοτάδι, καθώς η αντίδραση μεταξύ των δύο είναι πολύ αργή. Ωστόσο, όταν το μείγμα εκτίθεται στο φως, η αντίδραση είναι εκρηκτική.

Co-funded by the Erasmus+ Programme of the European Union

Τάξη αντίδρασης

Η τάξη είναι ο αριθμός των όρων συγκέντρωσης από τους οποίους εξαρτώνται οι ρυθμοί αντίδρασης. Έτσι, εάν ο ρυθμός μιας αντίδρασης εξαρτάται από την πρώτη δύναμη της συγκέντρωσης του αντιδρώντος, δηλαδή Ρυθμός = KC1, η αντίδραση λέγεται ότι είναι **πρώτης τάξης** . Όταν ο ρυθμός είναι ανάλογος με το προϊόν δύο συγκεντρώσεων αντιδρώντων ή το τετράγωνο της συγκέντρωσης ενός αντιδρώντος, η αντίδραση είναι **δεύτερης τάξης** .

Co-funded by the Erasmus+ Programme of the European Union

Τάξη αντίδρασης

Η μοριακή ικανότητα μιας αντίδρασης ορίζεται ως ο αριθμός των μορίων ή των ατόμων που συμμετέχουν στη διαδικασία μιας χημικής αλλαγής. Η αντίδραση λέγεται ότι είναι **μονομοριακή** , **διμοριακή** , **τριμοριακή** σύμφωνα με το ότι ένα, δύο ή τρία μόρια εμπλέκονται στη διαδικασία μιας χημικής αλλαγής. Ο όρος μονομοριακή χρησιμοποιήθηκε για όλες τις αντιδράσεις πρώτης τάξης, ο όρος διμοριακή για αντιδράσεις 2ης τάξης κ.λπ.

Co-funded by the Erasmus+ Programme of the European Union

Κινητική πρώτης τάξης

Μια αντίδραση πρώτης τάξης αναπαρίσταται ως

$$X \rightarrow Y$$

όπου X είναι το αντιδρόν και Y το προϊόν. Ο ρυθμός της αντίδρασης θα είναι ευθέως ανάλογος με τη συγκέντρωση, π.

(1.1) $- dc / dt = KC$

όπου C είναι η συγκέντρωση του αντιδρώντος σε οποιαδήποτε στιγμή t και το K είναι μια σταθερά, που ονομάζεται σταθερά ταχύτητας ή ειδικός ρυθμός αντίδρασης. Έτσι $- dc / C = K\, dt$

Κινητική Δεύτερης Τάξης

Μια αντίδραση θα είναι δεύτερης τάξης όταν η ταχύτητα αντίδρασης εξαρτάται από το γινόμενο δύο συγκεντρώσεων. Οι αντιδράσεις δεύτερης τάξης είναι δύο τύπων

(i) Ο ρυθμός είναι ανάλογος με το τετράγωνο της ίδιας συγκέντρωσης αντιδρώντος

(ii) Ο ρυθμός είναι ανάλογος με το γινόμενο της συγκέντρωσης των δύο αντιδρώντων

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ψευδομονομοριακές Αντιδράσεις

Υπάρχει ένας αριθμός αντιδράσεων, οι οποίες ακολουθούν την κινητική πρώτης τάξης, αν και περισσότερα από ένα είδη αντιδρώντων εμπλέκονται στην αντίδραση. Συνηθισμένα παραδείγματα είναι η αναστροφή ζάχαρης από ζαχαροκάλαμο ή η υδρόλυση ενός εστέρα σε όξινο μέσο.

$$C_{12}H_{22}O_{11} + H_2O = 2C_6H_{12}O_6$$

$$CH_3COOC_2H_5 + H_2O = CH_3COOH + C_2H_5OH$$

Αυτές είναι πραγματικά αντιδράσεις δεύτερης τάξης. Συμμετέχουν δύο ουσίες νερό και ζαχαροκάλαμο, ή νερό και εστέρας, αλλά ο ρυθμός αντίδρασης παρατηρείται πειραματικά ότι εξαρτάται μόνο από τη συγκέντρωση του σακχάρου ή του εστέρα από ζαχαροκάλαμο.

Co-funded by the Erasmus+ Programme of the European Union

Καταλυτικές Αντιδράσεις

Ορισμός Κατάλυση είναι η διαδικασία αύξησης του ρυθμού της αντίδρασης μέσω μιας ξένης ουσίας που παραμένει αμετάβλητη σε μάζα και χημική σύσταση.
Ταξινόμηση

i) **Ομογενής Κατάλυση** Ομογενής Κατάλυση είναι όπου ο καταλύτης και το αντιδρόν αποτελούν μια ενιαία φάση.

ii) **Ετερογενής Κατάλυση** Ετερογενής Κατάλυση είναι όπου ο καταλύτης και το αντιδρόν αποτελούν χωριστή φάση.

Co-funded by the Erasmus+ Programme of the European Union

Καταλυτικές Αντιδράσεις

Χαρακτηριστικά

i) Ο καταλύτης παραμένει αμετάβλητος σε μάζα και χημική σύσταση στο τέλος της αντίδρασης.

ii) Μια πολύ μικρή ποσότητα ενός καταλύτη μπορεί να έχει σημαντική επίδραση στην ταχύτητα μιας αντίδρασης.

iii) Ένας καταλύτης δεν μπορεί να ξεκινήσει μια αντίδραση αλλά αυξάνει μόνο την ταχύτητά του. Ο καταλύτης παρέχει πραγματικά μια εναλλακτική διαδρομή για τον μετασχηματισμό στον οποίο η απαιτούμενη ενέργεια ενεργοποίησης είναι μικρότερη. Η μικρότερη ενέργεια ενεργοποίησης επιτρέπει μεγαλύτερες ποσότητες αντίδρασης σε δεδομένο χρόνο.

iv) Ο καταλύτης δεν επηρεάζει την τελική κατάσταση ισορροπίας.

Co-funded by the Erasmus+ Programme of the European Union

Θερμοκρασία και Ρυθμοί Αντίδρασης

Η Θερμοκρασία της Εξίσωσης Arrhenius έχει βαθιά επίδραση στην ταχύτητα της αντίδρασης. Στις ομοιογενείς θερμικές αντιδράσεις, για κάθε αύξηση δέκα βαθμών της θερμοκρασίας, η ταχύτητα της αντίδρασης διπλασιάζεται ή τριπλασιάζεται. Ο λόγος ονομάζεται **συντελεστής θερμοκρασίας** .

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Θερμοκρασία και Ρυθμοί Αντίδρασης

Ο Arrhenius (1889) έδειξε ότι η σταθερά ταχύτητας (K) μιας χημικής διεργασίας αυξάνεται εκθετικά με τη θερμοκρασία για μεγάλο αριθμό αντιδράσεων. Παρατηρήθηκε ότι η γραφική παράσταση του logK έναντι της θερμοκρασίας δίνει μια γραμμική σχέση.

Πρότεινε λοιπόν εμπειρικά τη σχέση ως

$d\ln K / dT = E / RT2$ ή

$K = Ae^{-E/RT}$ (A = σταθερά)

Ο Α ονομάζεται συντελεστής συχνότητας ή προεκθετικός παράγοντας.

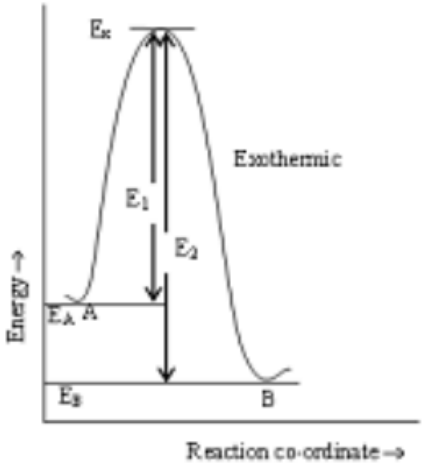
Ε ονομάζεται ενέργεια ενεργοποίησης της αντίδρασης.

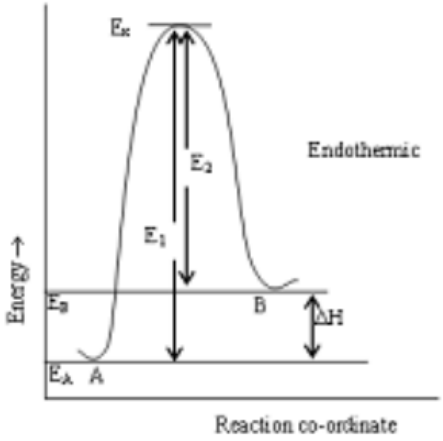
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενέργεια ενεργοποίησης μιας αντίδρασης: η σημασία της

Η μέση ενέργεια του αντιδρώντος αντιπροσωπεύεται από ΕΑ και αυτή του προκύπτοντος με ΕΒ. Αλλά εάν το ΕΑ είναι μεγαλύτερο από το ΕΒ, το αντιδρών Α δεν θα μετατραπεί ευθέως στο προϊόν Β. Υπάρχει ένα ελάχιστο επίπεδο ενέργειας για την αντίδραση που υποδηλώνεται με Εx στο οποίο πρέπει να ανυψωθεί το μόριο του αντιδραστηρίου για να μπορέσει να υποστεί τη χημική αλλαγή. Η περίσσεια ή η πρόσθετη ενέργεια (ΕΧ – ΕΑ) που πρέπει να αποκτήσει το αντιδρών για να υποστεί μετασχηματισμό είναι η ενέργεια ενεργοποίησης Ε1.



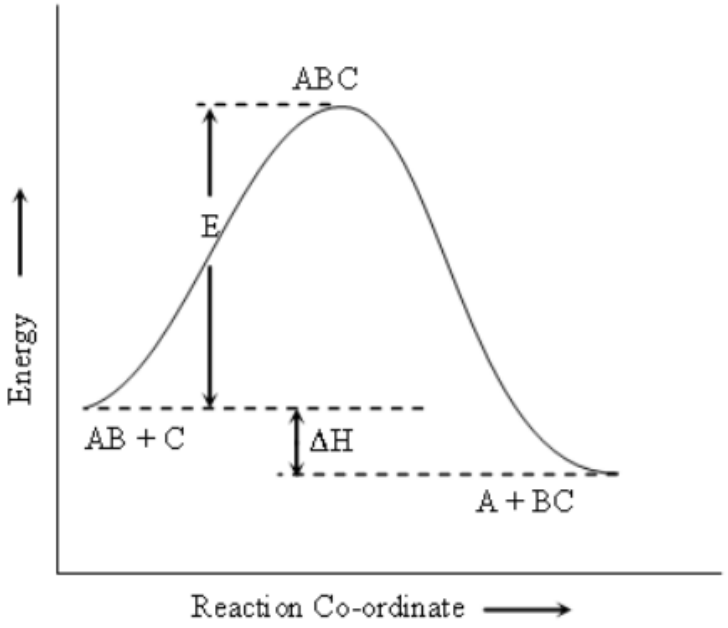


Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Θεωρία μεταβατικής κατάστασης

Σύμφωνα με αυτή τη θεωρία, τα δύο αντιδραστήρια, τα AB και C σχηματίζουν πρώτα ένα σύμπλοκο μετάπτωσης A——B——C το οποίο στη συνέχεια αποσυντίθεται σε A και BC. Η ενέργεια που απαιτείται για να ωθήσει το C να πλησιάσει το B για να σχηματίσει αυτή τη μεταβατική κατάσταση ή το ενεργοποιημένο σύμπλεγμα είναι η ενέργεια της ενεργοποίησης. Ο παρατηρούμενος ρυθμός αντίδρασης θα ήταν ίσος με τον καθαρό ρυθμό σχηματισμού του ενεργοποιημένου συμπλόκου.



Co-funded by the Erasmus+ Programme of the European Union

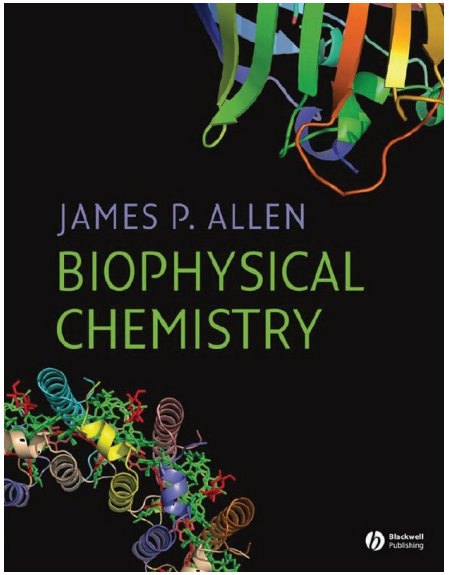
Ερωτήσεις

- Ποιος είναι ο ρυθμός μιας αντίδρασης;
- Τι είναι η σταθερά ρυθμού;
- Ποια είναι η τάξη μιας αντίδρασης; Δώστε ένα παράδειγμα αντίδρασης κλασματικής τάξης.
- Τι είναι η μοριακότητα μιας αντίδρασης;
- Ποια είναι η διαφορά μεταξύ τάξης και μοριακού χαρακτήρα μιας χημικής αντίδρασης;

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

βιβλιογραφικές αναφορές



Βιοφυσική Χημεία,
Τζέιμς Π. Άλεν
2008 από την Blackwell Publishing

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 4. Βιοφυσικά προβλήματα

Μάθημα 3. Κβαντομηχανική

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Κλασικές έννοιες
- Αρχές κβαντικής θεωρίας
- Η εξίσωση Schrodinger
- Ερμηνεία Born

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Για να κατανοήσουμε τις ιδιότητες των κυτταρικών μεμβρανών, των πρωτεϊνών και των νουκλεϊκών οξέων, είναι απαραίτητο να γνωρίζουμε πώς τα μόρια αυτά, που σχηματίζουν τα βιολογικά συστατικά, ανταποκρίνονται στις διαφορετικές αλληλεπιδράσεις τους με το περιβάλλον.

Κλασικές έννοιες

Το κλασικό σωματίδιο περιγράφεται από έναν αριθμό παραμέτρων: μάζα, m, θέση, r, ταχύτητα, u, και φορτίο, q. Για απλότητα, αυτά θεωρούνται ως απλές μεταβλητές παρά ως διανύσματα που χρειάζονται για να περιγράψουν τρισδιάστατη κίνηση. Η κινητική ενέργεια του σωματιδίου, KE, και η συνολική ενέργεια, E, σχετίζονται με αυτές τις παραμέτρους και τη δυναμική ενέργεια, V, από:

$$KE = \frac{1}{2}mv^2$$

$$E = KE + V$$

Σύμφωνα με την κλασική μηχανική, η ενέργεια του συστήματος διατηρείται πάντα. Διατηρείται επίσης η γραμμική ορμή, p, η οποία ισούται με:

$$p = mv$$

Co-funded by the Erasmus+ Programme of the European Union

Κλασικές έννοιες

Διάφορες πιθανές αλληλεπιδράσεις μεταξύ σωματιδίων, όπως η βαρύτητα ή οι ηλεκτροστατικές αλληλεπιδράσεις, περιγράφονται από καλά καθορισμένες σχέσεις που περιλαμβάνουν τις παραμέτρους που περιγράφουν τα σωματίδια. Για παράδειγμα, το ηλεκτροστατική δύναμη, F, μεταξύ δύο φορτισμένων σωματιδίων, q1 και q2, διαχωρισμένα από την απόσταση, r12, δίνεται από:

$$F = \frac{1}{4\pi\varepsilon_o}\frac{q_1 q_2}{r_{12}^2}$$

όπου ε_0 είναι μια σταθερά, η διαπερατότητα κενού.

Κλασικές έννοιες

Σύμφωνα με την κλασική μηχανική, όταν όλες οι αλληλεπιδράσεις και οι αρχικές συνθήκες έχουν καθιερωθεί, τότε μπορεί να προβλεφθεί η χρονική εξέλιξη του συστήματος για κάθε στιγμή χρησιμοποιώντας την αρχή της διατήρησης της ενέργειας, η οποία αναφέρει ότι η συνολική ενέργεια του συστήματος παραμένει πάντα η ίδια, και χρησιμοποιώντας τους νόμους της μηχανικής, όπως η σχέση μεταξύ δύναμης, F, μάζας, m, και θέσης, x:

$$F = ma = m\frac{\mathrm{d}^2x}{\mathrm{d}t^2}$$

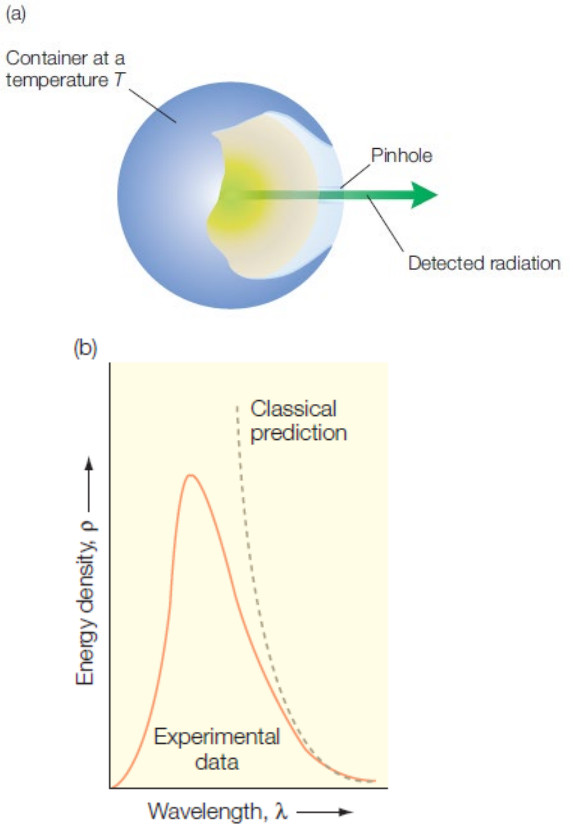
όπου α είναι μια σταθερά.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ακτινοβολία μελανού σώματος

Όταν τα αντικείμενα θερμαίνονται εκπέμπουν ακτινοβολία. Το είδος της ακτινοβολίας που εκπέμπεται εξαρτάται από τη θερμοκρασία. Για παράδειγμα, όταν θερμαίνετε κάτι το χρώμα μπορεί να είναι κόκκινο ή μπλε ή ακόμα και λευκό. Στις περισσότερες περιπτώσεις, το χρώμα εκπέμπεται σε διαφορετικά σημεία του αντικειμένου, θα ποικίλλει λόγω διαφορών στη θερμοκρασία. Για έναν ιδανικό πομπό που ονομάζεται μελανό σώμα, η εκπεμπόμενη ακτινοβολία είναι σε θερμική ισορροπία με το αντικείμενο, με αποτέλεσμα την ομοιόμορφη εκπομπή ακτινοβολίας. Ένα μελανό σώμα μπορεί να διαμορφωθεί ως μια σφαίρα στην οποία το φως που εκπέμπεται από τους εσωτερικούς τοίχους παγιδεύεται – δηλαδή απορροφάται και εκπέμπεται εκ νέου – εκτός από ένα μικρό τμήμα που μπορεί να διαφύγει μέσα από μια τρύπα .



Co-funded by the Erasmus+ Programme of the European Union

Ακτινοβολία μαύρου σώματος

Η εξάρτηση του ρ από το μήκος κύματος καθορίστηκε για πολλά υλικά και βρέθηκε να ακολουθεί πάντα την ίδια γενική εξάρτηση. Η εξάρτηση της κατανομής ενέργειας μπορεί να μοντελοποιηθεί χρησιμοποιώντας στατιστικά επιχειρήματα για την κλασική θερμοδυναμική. Υποθέτοντας ότι η ακτινοβολούμενη ενέργεια ακολουθεί την κλασική εξάρτηση από το πλάτος και είναι ανεξάρτητη από το μήκος κύματος, αποτυγχάνει για μικρά μήκη κύματος. Αντί να προβλέπει πως η πυκνότητα θα πέσει στο μηδέν, η θεωρία προβλέπει τη λεγόμενη υπεριώδη καταστροφή: ότι η ενεργειακή πυκνότητα θα γίνει άπειρη:

$$\rho = \frac{8\pi kT}{\lambda^4}$$

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ακτινοβολία μαύρου σώματος

Χρησιμοποιώντας στατιστικά επιχειρήματα αλλά με διαφορετική εξάρτηση για την ενέργεια, ο Planck εξήγαγε μια νέα εξάρτηση για την ενεργειακή πυκνότητα που συμφωνούσε με τα πειραματικά δεδομένα:

$$\rho = \frac{8\pi hc}{\lambda^5}\left(\frac{1}{e^{hc/\lambda kT} - 1}\right)$$

Σε μεγάλα μήκη κύματος, αυτή η εξίσωση συμφωνεί με την κλασική πρόβλεψη. Σε μεγάλα μήκη κύματος ο εκθετικός όρος είναι πολύ μικρός και η εκθετική μπορεί να γραφτεί περίπου ως:

$$e^{x} - 1 \approx x \text{ όταν } x << 1$$

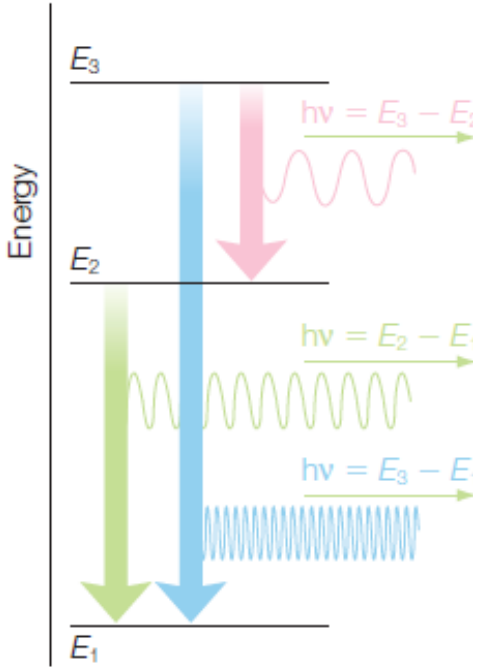
Χρησιμοποιώντας αυτή την προσέγγιση, η ενεργειακή πυκνότητα μπορεί να γραφτεί ως:

$$\rho \approx \frac{8\pi hc}{\lambda^5}\left(\frac{hc}{\lambda kT}\right)^{-1} = \frac{8\pi kT}{\lambda^4}$$

Co-funded by the Erasmus+ Programme of the European Union

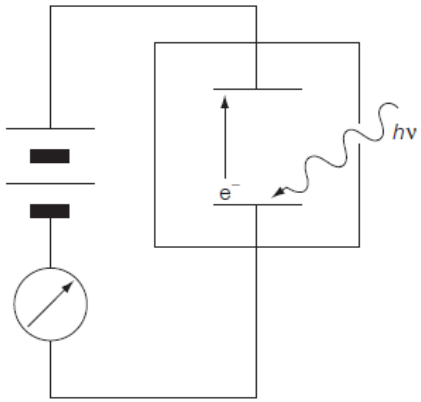
Φωτοηλεκτρικό φαινόμενο

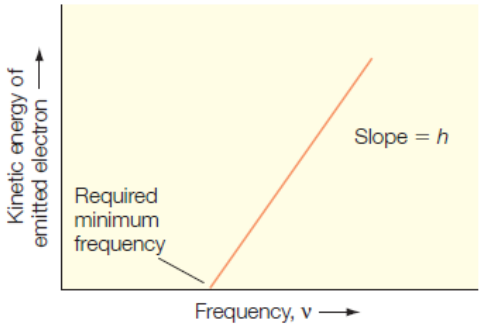
Όταν φως ορισμένων μηκών κύματος προσπίπτει σε μεταλλική επιφάνεια, ένα ηλεκτρόνιο εκτοξεύεται από το μέταλλο στο φωτοηλεκτρικό φαινόμενο. Κλασσικώς, η κινητική ενέργεια του εκτινασσόμενου ηλεκτρονίου θα πρέπει να σχετίζεται με την ένταση του φωτός, ή το πλάτος στο τετράγωνο, και είναι ανεξάρτητη από τη συχνότητα. Πειραματικά είναι δυνατή η μέτρηση αυτής της ιδιότητας για διαφορετικά μέταλλα και διαφορετικές συνθήκες φωτισμού (Εικόνα 9.3). Διαπιστώνεται ότι δεν εκτοξεύονται ηλεκτρόνια ανεξάρτητα από την ένταση αν η συχνότητα φωτός είναι κάτω από μια συγκεκριμένη τιμή που είναι χαρακτηριστική του μετάλλου. Για φως πάνω από αυτή την κρίσιμη συχνότητα, η κινητική ενέργεια εξαρτάται γραμμικά από τη συχνότητα και η ένταση αλλάζει μόνο τον αριθμό των ηλεκτρονίων που εκτινάσσονται.



Ατομικά φάσματα

Όταν τα αντικείμενα θερμαίνονται εκπέμπουν φως που είναι χαρακτηριστικό για το καθένα στοιχείο. Πειραματικά, το εκπεμπόμενο φως παρατηρείται ότι δεν είναι συνεχές αλλά διακριτό. Αυτό το χαρακτηριστικό υποδηλώνει ότι η ενέργεια που σχετίζεται με τα άτομα είναι διακριτή ή κβαντισμένη. Στην κλασική φυσική, δεν υπάρχει λόγος για αυτή την ιδιότητα της κβαντισμένης εκπομπής φωτός.



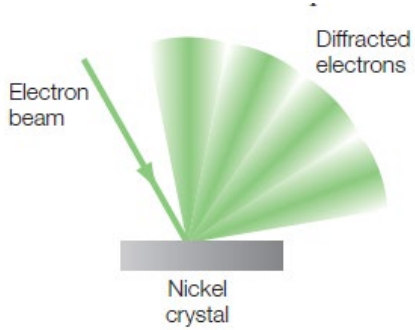


Αρχές κβαντικής θεωρίας

Δυαδικότητα κύματος-σωματιδίου

Κλασικά, τα σωματίδια και τα κύματα έχουν διαφορετικές παραμέτρους:

- Σωματίδια: μάζα m, θέση r, ταχύτητα u , φορτίο q
- Κύματα: μήκος κύματος λ, συχνότητα ν, ταχύτητα u, πλάτος A



Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Αρχές κβαντικής θεωρίας

Η εξίσωση του Σρέντινγκερ

Κλασικά, η ενέργεια ενός σωματιδίου δίνεται από το άθροισμα της κινητικής και δυναμικής ενέργειας:

$$E = \frac{1}{2}mv^2 + V(x) = \frac{p^2}{2m} + V(x)$$

Στην κβαντομηχανική αυτή η έκφραση τροποποιείται με την εισαγωγή των τελεστών, όπως προβλέπεται στον Πίνακα. Σε μερικούς από αυτούς τους τελεστές εμπλέκονται τα ħ και i , τα οποία ορίζονται ως:

$$\hbar = \frac{h}{2\pi} \quad \text{and} \quad i = \sqrt{-1}$$

Physical variables and the corresponding quantum operators.

Variable	Operator
X	x
V	V
p_x	$-i\hbar\frac{\partial}{\partial x}$
t	t
E	$i\hbar\frac{\partial}{\partial t}$

$$E\psi(r) = -\frac{\hbar^2}{2m}\nabla^2\psi(r) + V(r)\psi(r)$$

Co-funded by the Erasmus+ Programme of the European Union

Αρχές κβαντικής θεωρίας

Ερμηνεία Born

Η φυσική ερμηνεία της κβαντικής μηχανικής, ιδιαίτερα η ερμηνεία της κυματοσυνάρτησης, αναπτύχθηκε από πολλούς επιστήμονες, με πιο αξιοσημείωτο τον Max Born. Δεδομένου ότι όλα τα σωματίδια είναι επίσης κύματα, τα σωματίδια κατανέμονται πάντα στο χώρο. Η κυματοσυνάρτηση στην εξίσωση του Schrödinger δεν έχει άμεση φυσική σημασία και μπορεί να είναι μια μιγαδική συνάρτηση παρά μια πραγματική συνάρτηση. Στην ερμηνεία Born, η πιθανότητα εύρεσης οποιουδήποτε σωματιδίου σε μια συγκεκριμένη θέση δεν δίνεται από την ίδια την κυματοσυνάρτηση αλλά μάλλον από το τετράγωνο της κυματοσυνάρτησης. Αυτές οι ιδέες οδήγησαν σε πολλά θεμελιώδη αξιώματα της κβαντικής μηχανικής, ως εξής:

Αρχές κβαντικής θεωρίας

Ερμηνεία Born

1. Ένα σωματίδιο δεν βρίσκεται ποτέ σε μια συγκεκριμένη θέση, αλλά έχει μόνο μια πιθανότητα να βρίσκεται εκεί. Η πιθανότητα εύρεσης ενός σωματιδίου σε μια συγκεκριμένη θέση δίνεται από το τετράγωνο της κυματοσυνάρτησης επί τον όγκο dτ ως:

$$\psi^*(r)\psi(r)\,d\tau$$

όπου $d\tau = dx\,dy\,dz = r^2 \sin\theta\,dr\,d\vartheta\,d\phi$

Σε αυτή την εξίσωση το ψ* είναι το μιγαδικό συζυγές. Δεδομένου ότι οποιαδήποτε κυματοσυνάρτηση ψ μπορεί να γραφτεί με όρους δύο πραγματικών συναρτήσεων, A και B, το συζυγές μπορεί να οριστεί ως:

$$\psi = A + iB \text{ και } \psi^* = A - iB$$

Η πιθανότητα να βρεθεί ένα σωματίδιο μέσα σε έναν όγκο V είναι τότε:

$$\int_V \psi^*(r)\psi(r)\,d\tau$$

όπου ψ*(x) είναι το μιγαδικό συζυγές της κυματοσυνάρτησης

Αρχές κβαντικής θεωρίας

Ερμηνεία Born

2. Η κυματοσυνάρτηση μπορεί να είναι μιγαδική συνάρτηση αλλά η πιθανότητα είναι πάντα πραγματική, αφού:

$$\psi(x) = A(x) + iB(x)$$

$$\psi^*(x) = A(x) - iB(x)$$

οπότε $\psi^*(x)\psi(x) = [A(x) + iB(x)][A(x) - iB(x)] = A^2(x) + B^2(x) \geq 0$

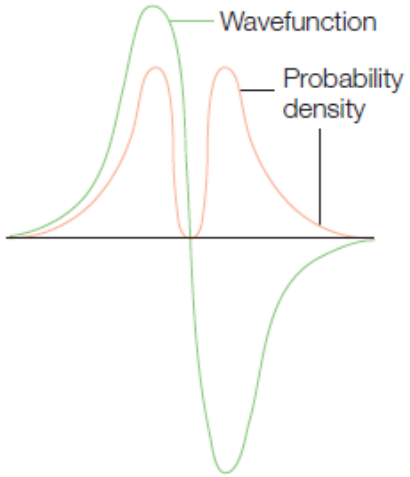
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Αρχές κβαντικής θεωρίας

Ερμηνεία Born

3. Η πιθανότητα είναι πάντα θετικός και πραγματικός αριθμός. Η συνολική πιθανότητα να βρεθεί ένα αντικείμενο οπουδήποτε στο διάστημα πρέπει να είναι ίση με μονάδα. Το άθροισμα όλων των πιθανοτήτων γράφεται μαθηματικά ως ολοκλήρωμα της πιθανότητας. Επομένως, το ολοκλήρωμα της πιθανότητας σε όλο τον χώρο πρέπει να είναι ίσο με μονάδα:



$$1 = \int_0^{\infty} \psi^*(r)\psi(r)\,\mathrm{d}\tau$$

Co-funded by the Erasmus+ Programme of the European Union

Αρχές κβαντικής θεωρίας

Ερμηνεία Born

4. Τα σωματίδια δεν έχουν συγκεκριμένη θέση ή ορμή αλλά μάλλον υπάρχει μια κατανομή τιμών που αντικατοπτρίζουν την κατανομή του σωματιδίου. Η φυσική ποσότητα είναι η μέση τιμή ή η προσδοκία. Κάθε φυσικό παρατηρήσιμο p έχει έναν συσχετισμένο τελεστή και ο μέσος όρος ή η προσδοκώμενη τιμή του παρατηρήσιμου δίνεται από:

$$\langle p \rangle = \int \psi^*(r)\hat{p}\psi(r)\,\mathrm{d}\tau$$

όπου V είναι ο τελεστής.

Για παράδειγμα, οι μέσες τιμές της θέσης και της ορμής ενός σωματιδίου μπορούν να υπολογιστούν αντικαθιστώντας τους τελεστές για τη θέση r και x συνιστώσα της ορμής:

$$\langle r \rangle = \int \psi^*(r) r \psi(r)\,\mathrm{d}\tau$$

$$\langle x^2 \rangle = \int \psi^*(x) x^2 \psi(x)\,\mathrm{d}x$$

$$\langle p_x \rangle = \int \psi^*(r)\left(-i\hbar\frac{\partial}{\partial x}\right)\psi(r)\,\mathrm{d}\tau$$

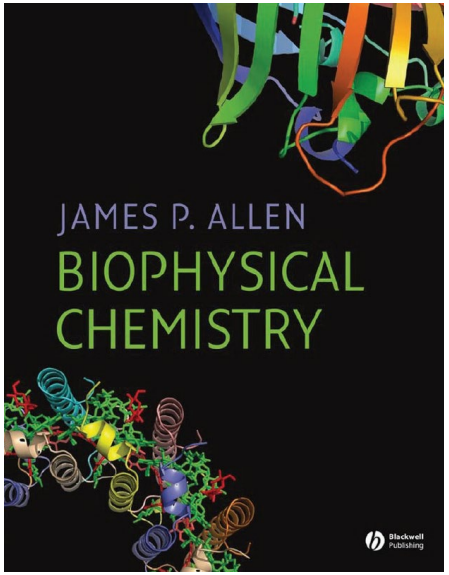
Ερωτήσεις

- Γιατί η παρατηρούμενη ακτινοβολία μελανού σώματος έρχεται σε σύγκρουση με την κλασική φυσική;
- Γιατί η παρουσία διακριτών γραμμών στα φάσματα εκπομπής των ατόμων υποστηρίζει τις ιδέες της κβαντικής θεωρίας;
- Γιατί χρειάζεται μια ελάχιστη συχνότητα για να προκληθεί το φωτοηλεκτρικό φαινόμενο;
- Ποια είναι η πιθανότητα να βρεθεί ένα σωματίδιο μέσα σε έναν όγκο V_0;

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

βιβλιογραφικές αναφορές



Βιοφυσική Χημεία,
Τζέιμς Π. Άλεν
2008 από την Blackwell Publishing

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 4. Βιοφυσικά προβλήματα

Μάθημα 4. Φασματοσκοπία

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Υπέρυθρη φασματοσκοπία
- Φασματομετρία μάζας
- Φασματοσκοπία πυρηνικού μαγνητικού συντονισμού
- Φασματοσκοπία υπεριώδους

Co-funded by the Erasmus+ Programme of the European Union

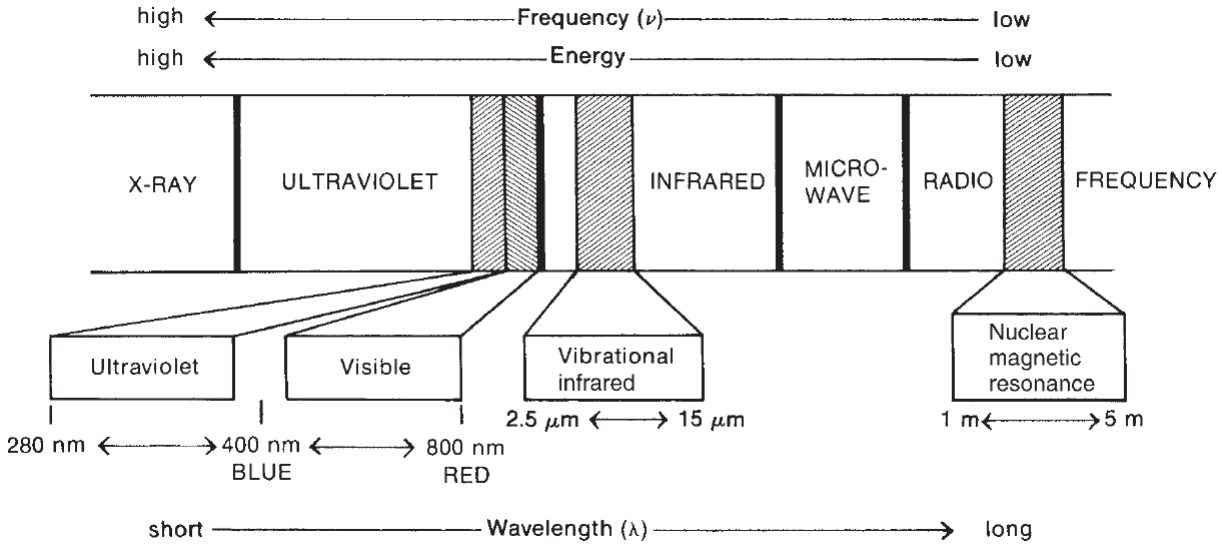
Εισαγωγή

Η φασματοσκοπία είναι βασικά ένα πειραματικό θέμα και ασχολείται με την απορρόφηση, εκπομπή ή σκέδαση ηλεκτρομαγνητικής ακτινοβολίας από άτομα ή μόρια. Η ηλεκτρομαγνητική ακτινοβολία καλύπτει ένα ευρύ φάσμα μήκους κύματος, από ραδιοκύματα έως ακτίνες γ , και τα άτομα ή τα μόρια μπορεί να βρίσκονται στην αέρια, υγρή ή στερεή φάση ή, μεγάλης σημασίας στη χημεία της επιφάνειας, να προσροφηθούν σε μια στερεά επιφάνεια.

Οι πειραματικές μέθοδοι φασματοσκοπίας ξεκίνησαν στην πιο προσιτή ορατή περιοχή του ηλεκτρομαγνητικού φάσματος όπου το μάτι μπορούσε να χρησιμοποιηθεί ως ανιχνευτής.

Υπέρυθρη φασματοσκοπία

Σχεδόν κάθε ένωση που έχει ομοιοπολικούς δεσμούς, είτε οργανική είτε ανόργανη, απορροφά διάφορες συχνότητες ηλεκτρομαγνητικής ακτινοβολίας στην υπέρυθρη περιοχή του ηλεκτρομαγνητικού φάσματος. Αυτή η περιοχή βρίσκεται σε μήκη κύματος μεγαλύτερα από αυτά που σχετίζονται με το ορατό φως, τα οποία κυμαίνονται από περίπου 400 έως 800 nm (1 nm = 10^{-9} m), αλλά βρίσκεται σε μήκη κύματος μικρότερα από αυτά που σχετίζονται με τα μικροκύματα, τα οποία είναι μεγαλύτερα από 1 mm. Για χημικούς σκοπούς, μας ενδιαφέρει το δονητικό τμήμα της υπέρυθρης περιοχής. Περιλαμβάνει ακτινοβολία με μήκη κύματος (l) μεταξύ 2,5 μm και 25 μm (1μm = 10^{-6} m). Αν και η πιο σωστή τεχνικά μονάδα για το μήκος κύματος στην υπέρυθρη περιοχή του φάσματος είναι το μικρόμετρο (μm), θα δείτε συχνά το μικρό (μ) να χρησιμοποιείται στα υπέρυθρα φάσματα. Το σχήμα απεικονίζει τη σχέση της υπέρυθρης περιοχής με άλλες που περιλαμβάνονται στο ηλεκτρομαγνητικό φάσμα.



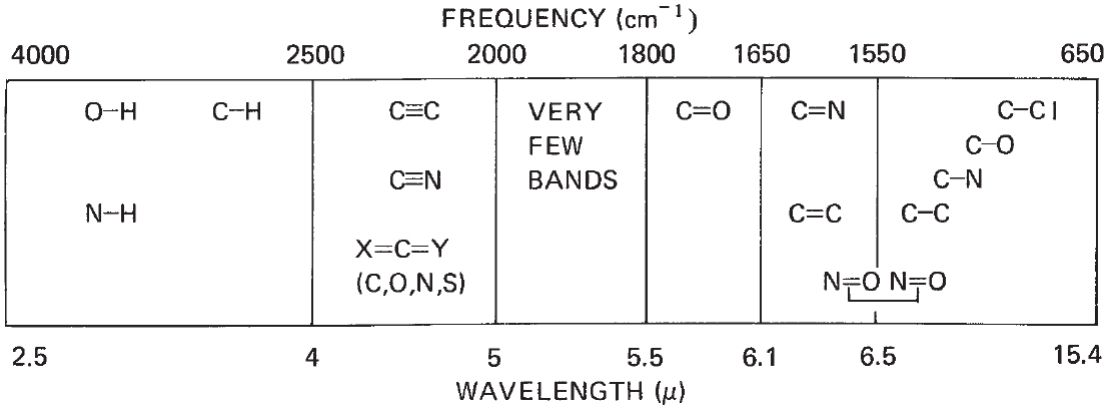
Υπέρυθρη φασματοσκοπία

Η απορρόφηση της υπέρυθρης ακτινοβολίας αντιστοιχεί σε αλλαγές ενέργειας της τάξης των 8 έως 40 kJ/mole. Η ακτινοβολία σε αυτό το ενεργειακό εύρος αντιστοιχεί στο εύρος που περιλαμβάνει τις συχνότητες δόνησης τάνυσης και κάμψης των δεσμών στα περισσότερα ομοιοπολικά μόρια. Στη διαδικασία απορρόφησης, οι συχνότητες εκείνες της υπέρυθρης ακτινοβολίας που ταιριάζουν με τις φυσικές συχνότητες δόνησης του εν λόγω μορίου απορροφώνται και η ενέργεια που απορροφάται χρησιμεύει για την αύξηση του εύρους των δονητικών κινήσεων των δεσμών στο μόριο. Σημειώστε, ωστόσο, ότι δεν είναι όλοι οι δεσμοί σε ένα μόριο ικανοί να απορροφούν υπέρυθρη ενέργεια, ακόμα κι αν η συχνότητα της ακτινοβολίας ταιριάζει ακριβώς με αυτή της κίνησης του δεσμού. Μόνο εκείνοι οι δεσμοί που έχουν διπολική ροπή που μεταβάλλεται σε συνάρτηση με το χρόνο είναι ικανοί να απορροφούν την υπέρυθρη ακτινοβολία. Οι συμμετρικοί δεσμοί, όπως αυτοί του H2 ή του Cl2, δεν απορροφούν την υπέρυθρη ακτινοβολία. Ένας δεσμός πρέπει να παρουσιάζει ένα ηλεκτρικό δίπολο που αλλάζει με την ίδια συχνότητα με την εισερχόμενη ακτινοβολία για να μεταφερθεί ενέργεια.

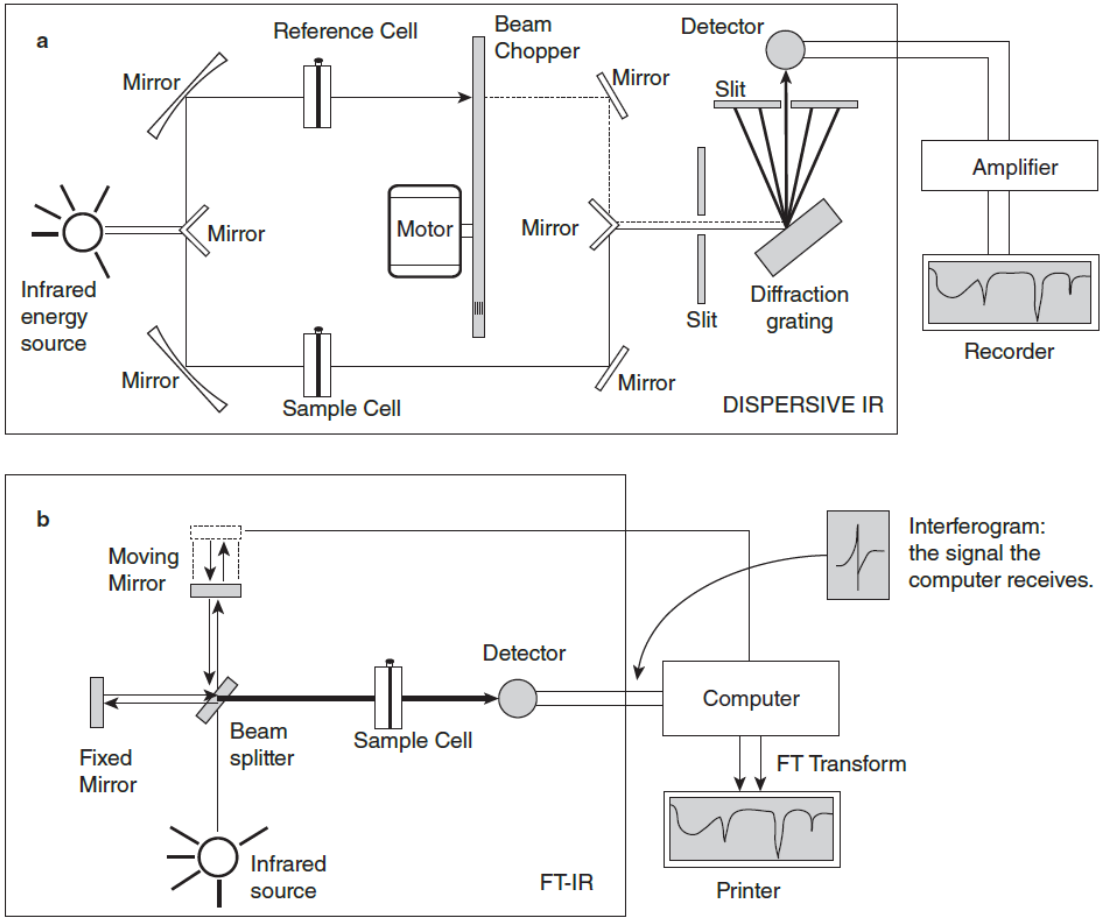
Region of Spectrum	Energy Transitions
X-rays	Bond breaking
Ultraviolet/visible	Electronic
Infrared	Vibrational
Microwave	Rotational
Radiofrequencies	Nuclear spin (nuclear magnetic resonance) Electronic spin (electron spin resonance)

Χρήσεις του υπέρυθρου φάσματος

- Το υπέρυθρο φάσμα μπορεί να χρησιμοποιηθεί για μόρια όπως ένα δακτυλικό αποτύπωμα μπορεί να χρησιμοποιηθεί για τον άνθρωπο.
- Η πιο σημαντική χρήση του υπέρυθρου φάσματος είναι ο προσδιορισμός δομικών πληροφοριών για ένα μόριο.



Σχηματικά διαγράμματα (α) φασματοφωτομέτρων διασποράς και (β) υπέρυθρου μετασχηματισμού Fourier

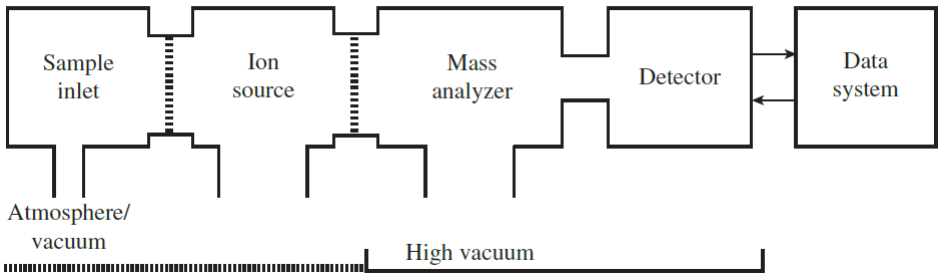


Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

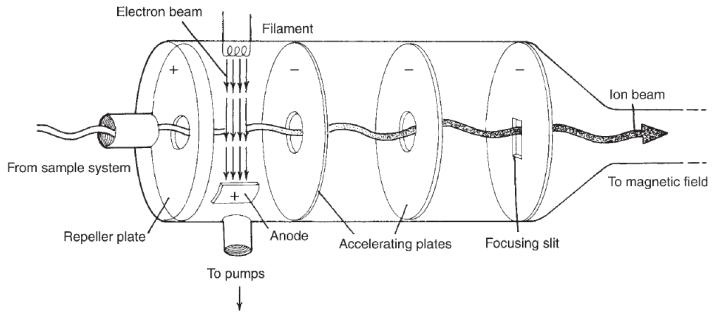
Φασματομετρία μάζας

Στην απλούστερη μορφή του, το φασματόμετρο μάζας έχει πέντε συστατικά. Το πρώτο συστατικό του φασματόμετρου μάζας είναι η είσοδος του δείγματος, η οποία φέρνει το δείγμα από το εργαστηριακό περιβάλλον (1 atm) στη χαμηλότερη πίεση του φασματόμετρου μάζας. Οι πιέσεις μέσα στο φασματόμετρο μάζας κυμαίνονται από μερικά χιλιοστά υδραργύρου σε μια πηγή χημικού ιονισμού έως μερικά μικρόμετρα υδραργύρου στις περιοχές αναλυτή μάζας και ανιχνευτή του οργάνου. Η είσοδος του δείγματος οδηγεί στην πηγή ιόντων, όπου τα μόρια του δείγματος μετατρέπονται σε ιόντα αέριας φάσης. Πρόσφατα αναπτύχθηκαν ορισμένα όργανα που συνδυάζουν την είσοδο δείγματος και την πηγή ιόντων υπό συνθήκες περιβάλλοντος, απλοποιώντας έτσι πολύ την προετοιμασία του δείγματος. Στη συνέχεια, τα ιόντα επιταχύνονται από ένα ηλεκτρομαγνητικό πεδίο. Στη συνέχεια, ο αναλυτής μάζας διαχωρίζει τα ιόντα του δείγματος με βάση την αναλογία μάζας προς φορτίο (m/z). Στη συνέχεια, τα ιόντα μετρώνται από τον ανιχνευτή και το σήμα καταγράφεται και επεξεργάζεται από το σύστημα δεδομένων, συνήθως έναν προσωπικό υπολογιστή (PC). Η έξοδος από το σύστημα δεδομένων είναι το φάσμα μάζας - ένα γράφημα του αριθμού των ιόντων που ανιχνεύονται ως συνάρτηση του λόγου m/z τους.



Μέθοδοι ιοντισμού

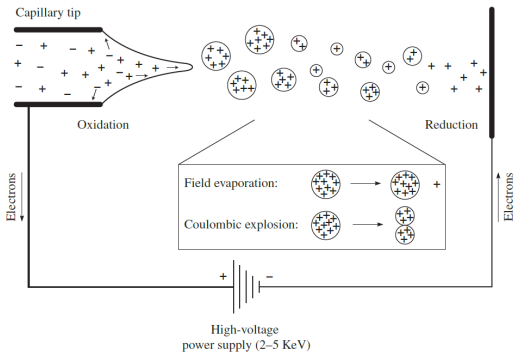
Ιονισμός ηλεκτρονίων (EI)

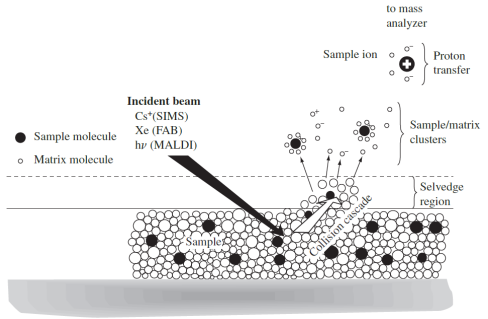


Χημικός ιονισμός (CI)

Τεχνικές ιονισμού εκρόφησης (SIMS, FAB και MALDI)

Ιονισμός ηλεκτροψεκασμού (ESI)





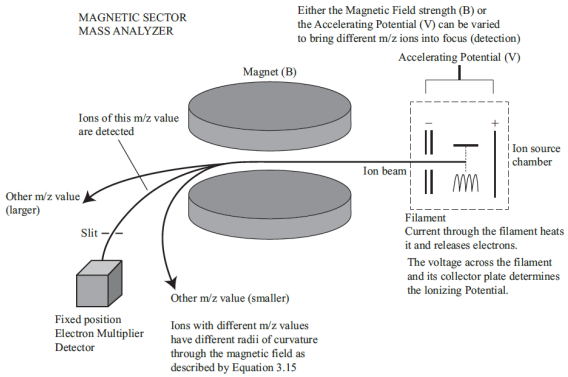
Μαζική ανάλυση

Ο αναλυτής μάζας Magnetic Sector

Αναλυτές μάζας διπλής εστίασης

Τετραπολικοί αναλυτές μάζας

Αναλυτές μάζας χρόνου πτήσης



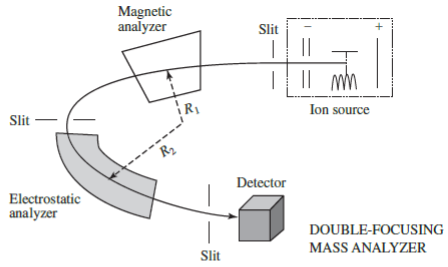
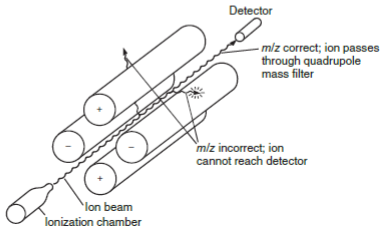
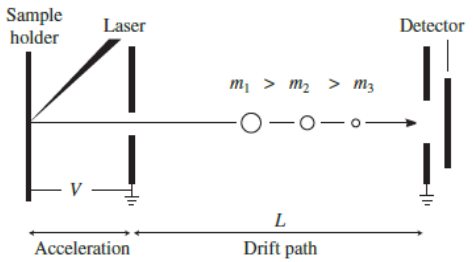
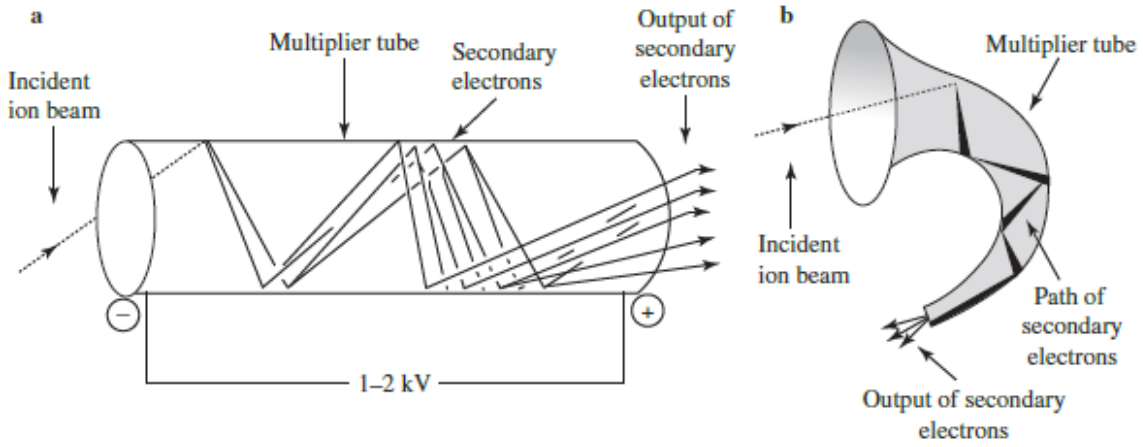


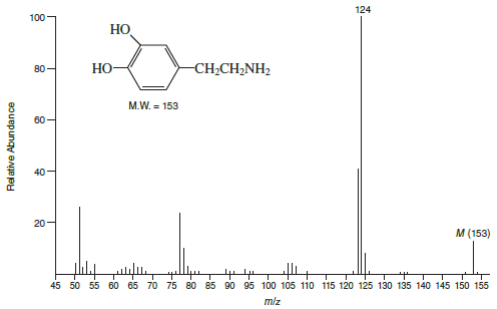
FIGURE 3.12 Schematic of a double-focusing mass analyzer.





Ανίχνευση και ποσοτικοποίηση: το φάσμα μάζας





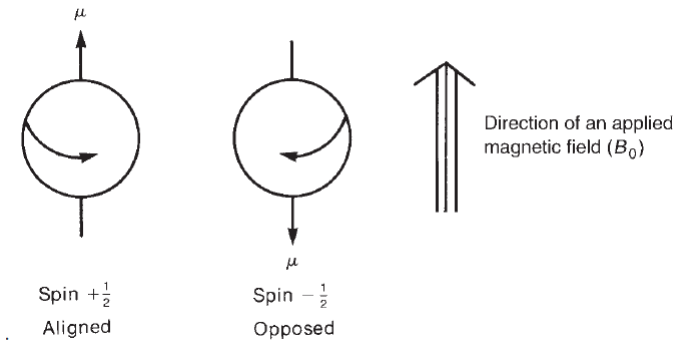
Φασματοσκοπία πυρηνικού μαγνητικού συντονισμού

Ο πυρηνικός μαγνητικός συντονισμός (NMR) είναι μια φασματοσκοπική μέθοδος που είναι ακόμη πιο σημαντική για τον οργανικό χημικό από την υπέρυθρη φασματοσκοπία. Πολλοί πυρήνες μπορούν να μελετηθούν με τεχνικές NMR, αλλά το υδρογόνο και ο άνθρακας είναι πιο συχνά διαθέσιμα. Ενώ η υπέρυθρη φασματοσκοπία (IR) αποκαλύπτει τους τύπους των λειτουργικών ομάδων που υπάρχουν σε ένα μόριο, το NMR δίνει πληροφορίες σχετικά με τον αριθμό των μαγνητικά διακριτών ατόμων του τύπου που μελετάται. Όταν μελετώνται πυρήνες υδρογόνου (πρωτόνια), για παράδειγμα, μπορεί κανείς να προσδιορίσει τον αριθμό καθενός από τους διαφορετικούς τύπους πυρήνων υδρογόνου καθώς και να λάβει πληροφορίες σχετικά με τη φύση του άμεσου περιβάλλοντος κάθε τύπου. Παρόμοιες πληροφορίες μπορούν να προσδιοριστούν για τους πυρήνες άνθρακα. Ο συνδυασμός δεδομένων IR και NMR είναι συχνά επαρκής για να προσδιοριστεί πλήρως η δομή ενός άγνωστου μορίου.

Element	$^{1}_{1}H$	$^{2}_{1}H$	$^{12}_{6}C$	$^{13}_{6}C$	$^{14}_{7}N$	$^{16}_{8}O$	$^{17}_{8}O$	$^{19}_{9}F$	$^{31}_{15}P$	$^{35}_{17}Cl$
Nuclear spin quantum number	$\frac{1}{2}$	1	0	$\frac{1}{2}$	1	0	$\frac{5}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{2}$
Number of spin states	2	3	0	2	3	0	6	2	2	4

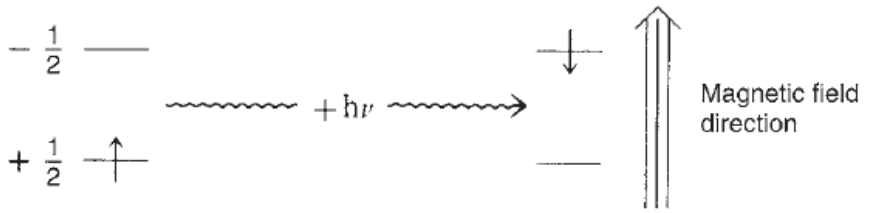
Πυρηνικές μαγνητικές ροπές

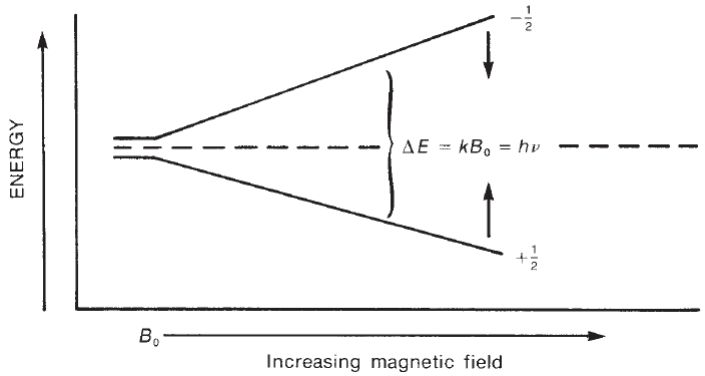
Οι καταστάσεις σπιν δεν έχουν ισοδύναμη ενέργεια σε ένα εφαρμοζόμενο μαγνητικό πεδίο επειδή ο πυρήνας είναι ένα φορτισμένο σωματίδιο και κάθε κινούμενο φορτίο δημιουργεί ένα δικό του μαγνητικό πεδίο. Έτσι, ο πυρήνας έχει μια μαγνητική ροπή m που δημιουργείται από το φορτίο και το σπιν του. Ένας πυρήνας υδρογόνου μπορεί να έχει δεξιόστροφη ή αριστερόστροφη στροφορμή και οι πυρηνικές μαγνητικές ροπές (μ) στις δύο περιπτώσεις είναι στραμμένες προς αντίθετες κατευθύνσεις.



Απορρόφηση ενέργειας

Το φαινόμενο πυρηνικού μαγνητικού συντονισμού συμβαίνει όταν πυρήνες ευθυγραμμισμένοι με ένα εφαρμοζόμενο πεδίο προκαλούνται να απορροφήσουν ενέργεια και να αλλάξουν τον προσανατολισμό του σπιν τους σε σχέση με το εφαρμοζόμενο πεδίο. Η απορρόφηση ενέργειας είναι μια κβαντοποιημένη διαδικασία και η ενέργεια που απορροφάται πρέπει να ισούται με τη διαφορά ενέργειας μεταξύ των δύο εμπλεκόμενων καταστάσεων.





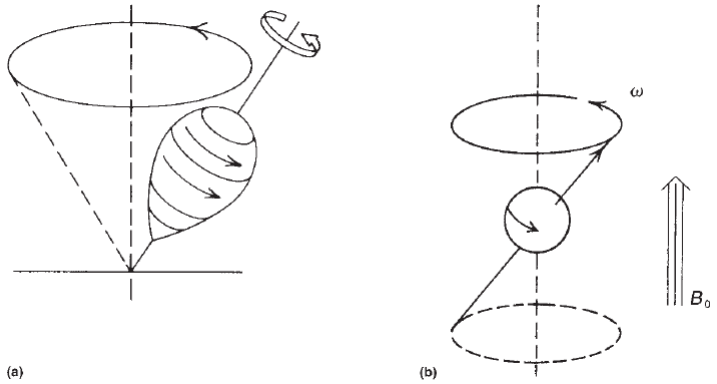
Συχνότητες και εντάσεις πεδίου στις οποίες επιλεγμένοι πυρήνες έχουν τους πυρηνικούς συντονισμούς τους

Isotope	Natural Abundance (%)	Field Strength, B_0 (Tesla[a])	Frequency, ν (MHz)	Magnetogyric Ratio, γ (radians/Tesla)
^{1}H	99.98	1.00	42.6	267.53
		1.41	60.0	
		2.35	100.0	
		4.70	200.0	
		7.05	300.0	
^{2}H	0.0156	1.00	6.5	41.1
^{13}C	1.108	1.00	10.7	67.28
		1.41	15.1	
		2.35	25.0	
		4.70	50.0	
		7.05	75.0	
^{19}F	100.0	1.00	40.0	251.7
^{31}P	100.0	1.00	17.2	108.3

[a] 1 Tesla = 10.000 Gauss.

Ο μηχανισμός απορρόφησης (συντονισμός)

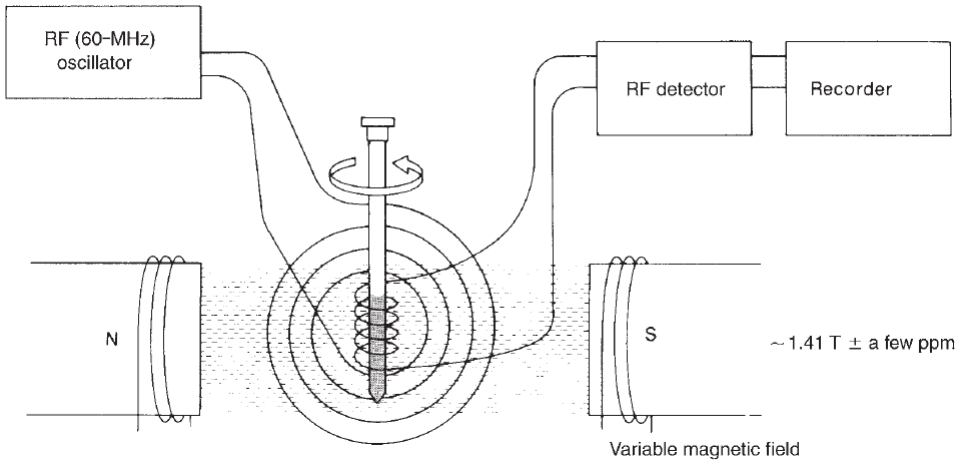
Για να κατανοήσουμε τη φύση μιας μετάβασης πυρηνικής περιστροφής, είναι χρήσιμη η αναλογία της σβούρας ενός παιδιού. Τα πρωτόνια απορροφούν ενέργεια επειδή αρχίζουν να προχωρούν σε ένα εφαρμοζόμενο μαγνητικό πεδίο. Το φαινόμενο της μετάπτωσης είναι παρόμοιο με αυτό της σβούρας. Λόγω της επιρροής του βαρυτικού πεδίου της γης, η κορυφή αρχίζει να «ταλαντεύεται» ή να προχωρά γύρω από τον άξονά της. Ένας περιστρεφόμενος πυρήνας συμπεριφέρεται με παρόμοιο τρόπο υπό την επίδραση ενός εφαρμοζόμενου μαγνητικού πεδίου.



Το φασματόμετρο πυρηνικού μαγνητικού συντονισμού

Το όργανο Continuous-Wave (CW).

Το όργανο παλμικού μετασχηματισμού Fourier (FT).



Φασματοσκοπία υπεριώδους

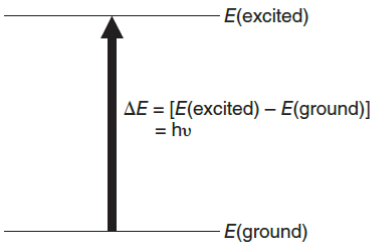
Τα περισσότερα οργανικά μόρια και λειτουργικές ομάδες είναι διαφανή στα τμήματα του ηλεκτρομαγνητικού φάσματος που ονομάζουμε περιοχές υπεριώδους (UV) και ορατές (VIS)—δηλαδή, στις περιοχές όπου τα μήκη κύματος κυμαίνονται από 190 nm έως 800 nm. Κατά συνέπεια, η φασματοσκοπία απορρόφησης είναι περιορισμένης χρησιμότητας σε αυτό το εύρος μηκών κύματος. Ωστόσο, σε ορισμένες περιπτώσεις μπορούμε να αντλήσουμε χρήσιμες πληροφορίες από αυτές τις περιοχές του φάσματος. Αυτές οι πληροφορίες, όταν συνδυάζονται με τις λεπτομέρειες που παρέχονται από τα φάσματα υπέρυθρου και πυρηνικού μαγνητικού συντονισμού (NMR), μπορούν να οδηγήσουν σε πολύτιμες δομικές προτάσεις.

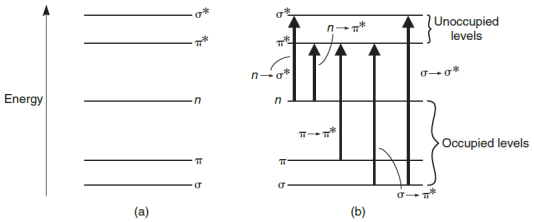
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

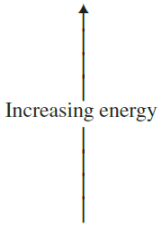
Co-funded by the Erasmus+ Programme of the European Union

Η φύση των ηλεκτρονικών διεγέρσεων

Όταν η συνεχής ακτινοβολία διέρχεται από ένα διαφανές υλικό, ένα μέρος της ακτινοβολίας μπορεί να απορροφηθεί. Εάν συμβεί αυτό, η υπολειπόμενη ακτινοβολία, όταν διέρχεται από ένα πρίσμα, παράγει ένα φάσμα με κενά σε αυτό, που ονομάζεται φάσμα απορρόφησης. Ως αποτέλεσμα της απορρόφησης ενέργειας, τα άτομα ή τα μόρια περνούν από μια κατάσταση χαμηλής ενέργειας (την αρχική ή βασική κατάσταση) σε μια κατάσταση υψηλότερης ενέργειας (τη διεγερμένη κατάσταση). Το σχήμα απεικονίζει αυτή τη διαδικασία διέγερσης, η οποία είναι κβαντισμένη. Η ηλεκτρομαγνητική ακτινοβολία που απορροφάται έχει ενέργεια ακριβώς ίση με τη διαφορά ενέργειας μεταξύ της διεγερμένης και της βασικής κατάστασης.





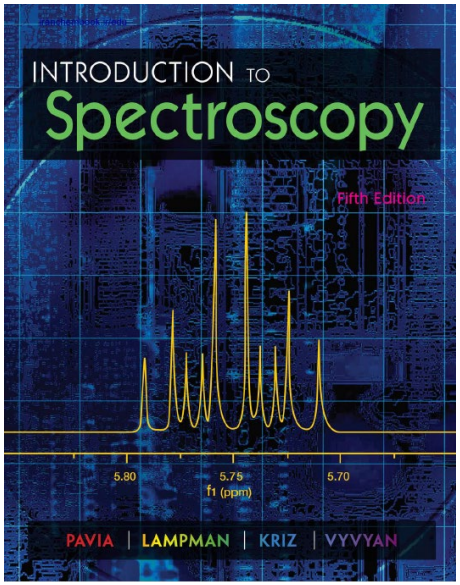


$\sigma \longrightarrow \sigma^*$	In alkanes
$\sigma \longrightarrow \pi^*$	In carbonyl compounds
$\pi \longrightarrow \pi^*$	In alkenes, carbonyl compounds, alkynes, azo compounds, and so on
$n \longrightarrow \sigma^*$	In oxygen, nitrogen, sulfur, and halogen compounds
$n \longrightarrow \pi^*$	In carbonyl compounds

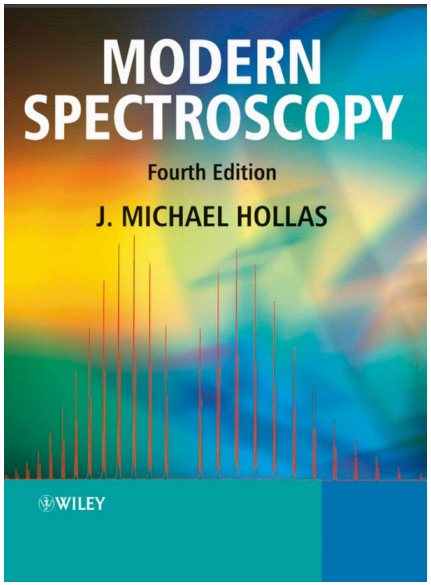
Ερωτήσεις

- Γιατί η παρατηρούμενη ακτινοβολία μελανού σώματος έρχεται σε σύγκρουση με την κλασική φυσική;
- Γιατί η παρουσία διακριτών γραμμών στα φάσματα εκπομπής των ατόμων υποστηρίζει τις ιδέες της κβαντικής θεωρίας;
- Γιατί χρειάζεται μια ελάχιστη συχνότητα για να προκληθεί το φωτοηλεκτρικό φαινόμενο;
- Ποια είναι η πιθανότητα να βρεθεί ένα σωματίδιο μέσα σε έναν όγκο V_0;

βιβλιογραφικές αναφορές



Εισαγωγή στη Φασματοσκοπία,
Πέμπτη Έκδοση
Donald L. Pavia, Gary M. Lampman,
George S. Kriz και James R. Vyvyan
2015, 2009 Cengage Learning



Σύγχρονη φασματοσκοπία
Τέταρτη Έκδοση
J. Michael Hollas
2004 από την John Wiley & Sons Ltd, The Atrium, Southern Gate

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 1. Μοριακή βάση της ζωής

Θέμα 4. Βιοφυσικά προβλήματα

Μάθημα 5. Κατανόηση των βιολογικών συστημάτων χρησιμοποιώντας τη φυσική χημεία

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Μεταφορά σήματος
- Δυναμικά μεμβράνης, μεταφορείς και κανάλια
- Μοριακή απεικόνιση
- Φωτοσύνθεση

Co-funded by the Erasmus+ Programme of the European Union

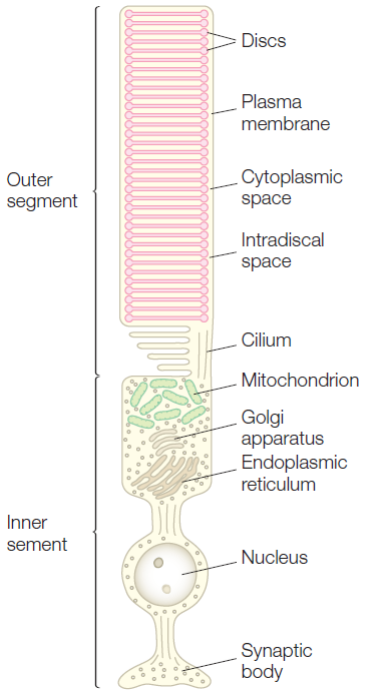
Εισαγωγή

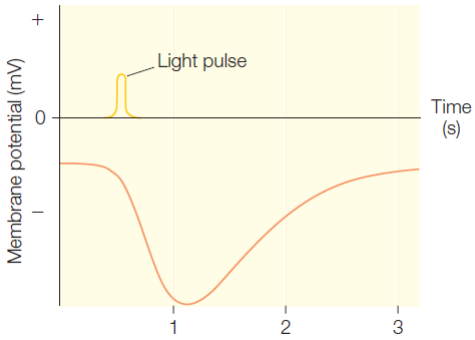
Η βιοφυσική χημεία είναι ένα διεπιστημονικό πεδίο, που συνδυάζει στοιχεία της βιολογίας, της χημείας και της φυσικής. Αποσκοπεί στη συλλογή και ανάλυση ποσοτικών δεδομένων για προγνωστικά φυσικά μοντέλα βιολογικών φαινομένων σε μοριακό επίπεδο. Χρησιμοποιείται επίσης για να περιγράψει τεχνικές που μελετούν τις φυσικές ιδιότητες σημαντικών βιολογικών μορίων σε χημικό επίπεδο.

Co-funded by the Erasmus+ Programme of the European Union

Μεταφορά σήματος

Η οπτική διαδικασία μπορεί να χωριστεί σε τέσσερα βήματα: αναγνώριση, μετατροπή, ενίσχυση και επεξεργασία. Το βήμα αναγνώρισης στην όραση είναι ασυνήθιστο για τη βιολογική μεταγωγή σήματος, καθώς περιλαμβάνει την απορρόφηση φωτός από μια χρωστική ουσία που είναι θαμμένη μέσα σε μια πρωτεΐνη που ονομάζεται ροδοψίνη και όχι τη δέσμευση ενός μορίου σήματος σε έναν υποδοχέα πρωτεΐνης, όπως βρίσκεται σε άλλες διαδικασίες μετάδοσης σήματος. Μετά την αναγνώριση, υπάρχει μια μετατροπή σήματος, η οποία απαιτεί μια δομική αλλαγή της πρωτεΐνης και ενός σχετικού μορίου που ονομάζεται αμφιβληστροειδής ως απόκριση στην απορρόφηση φωτός. Μόλις το σήμα ενεργοποιήσει τη ροδοψίνη, ενισχύεται κατά πολλές τάξεις μεγέθους, επιτρέποντας στο σήμα να επεξεργαστεί σε αλλαγή στο δυναμικό της μεμβράνης και κατά συνέπεια σε σήμα προς τον εγκέφαλο.



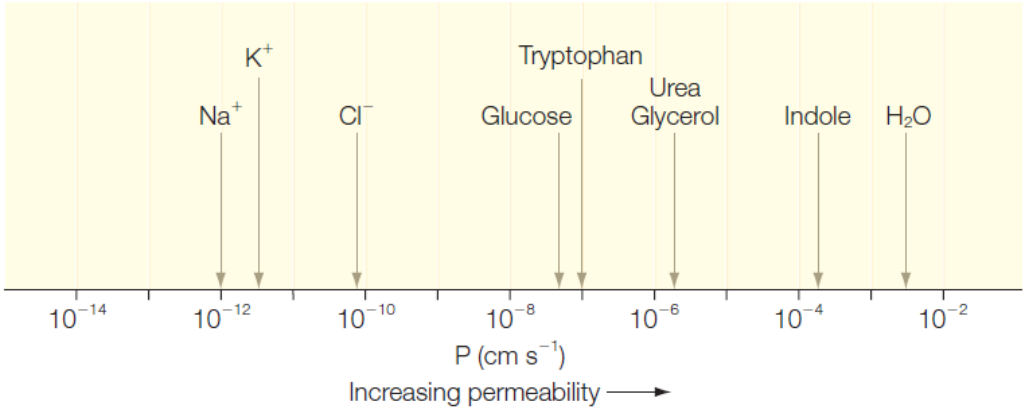


Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Δυναμικό μεμβράνης

Η κυτταρική μεμβράνη είναι μια λιπιδική διπλοστιβάδα. Η σύνθεση των κυτταρικών μεμβρανών ποικίλλει μεταξύ των τύπων κυττάρων και των συνθηκών ανάπτυξης, αλλά γενικά η κύρια συστατικό είναι τα φωσφολιπίδια. Τα φωσφολιπίδια είναι αμφιπαθή μόρια. έχουν δηλαδή και μια ομάδα πολικής κεφαλής και ένα υδρόφοβο λιπαρό οξύ τμήμα. Λόγω του κεντρικού υδρόφοβου πυρήνα της διπλής στιβάδας, η κυτταρική μεμβράνη είναι σε μεγάλο βαθμό αδιαπέραστη από ιόντα και πολικά μόρια. Αντίθετα, το νερό διασχίζει τις μεμβράνες πολύ πιο εύκολα. Η διαπερατότητα των μικρών μορίων κυμαίνεται σε πολλές τάξεις μεγέθους και συσχετίζεται με τη διαλυτότητα κάθε μορίου σε μη πολικούς διαλύτες σε σχέση με το νερό. Το κύτταρο κάνει χρήση αυτής της ικανότητας ελέγχου οι σχετικές συγκεντρώσεις ιόντων για τη δημιουργία βαθμίδων ιόντων κατά μήκος της κυτταρικής μεμβράνης για πολλές μεταβολικές διεργασίες.

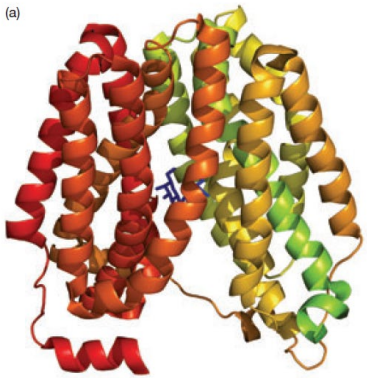


Ενέργεια της μεταφοράς μέσω των μεμβρανών

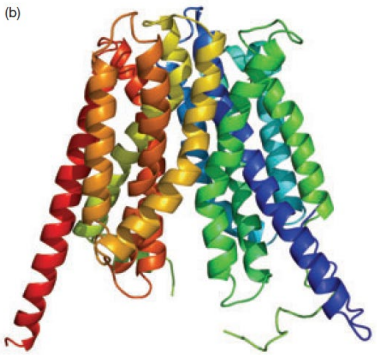
Η μεταφορά των ιόντων εξαρτάται από την αλλαγή της ελεύθερης ενέργειας του μεταφερόμενου ιόντος. Η ενεργή μεταφορά απαιτεί συνδεδεμένη εισροή ελεύθερης ενέργειας. Αυτή η διαδικασία είναι ενεργειακά δυσμενής και δεν θα συμβεί αυθόρμητα. Η μεταφορά θα μπορούσε να συζευχθεί με την υδρόλυση ATP αλλά εναλλακτικά θα μπορούσε να συζευχθεί με τη μεταφορά ενός άλλου μορίου μέσω της μεμβράνης. Αυτό συμβαίνει συχνά στις κυτταρικές μεμβράνες, σε πρωτεΐνες γνωστές ως αντιμεταφορείς, οι οποίες μπορούν να μεταφέρουν ταυτόχρονα δύο διαφορετικά μόρια σε αντίθετες κατευθύνσεις.

Μεταφορείς

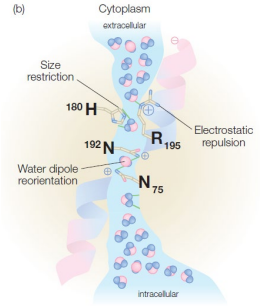
Οι πρωτεΐνες μεταφοράς μεσολαβούν επιλεκτικά στη διέλευση των μορίων η μεμβράνη, η οποία κατά τα άλλα είναι αδιαπέραστη. Περισσότεροι από 360 μεταφορείς έχουν εντοπιστεί οικογένειες, αναδεικνύοντας τον κρίσιμο ρόλο του διεργασίες μεταφοράς στα κύτταρα.



Lacy



GlpT



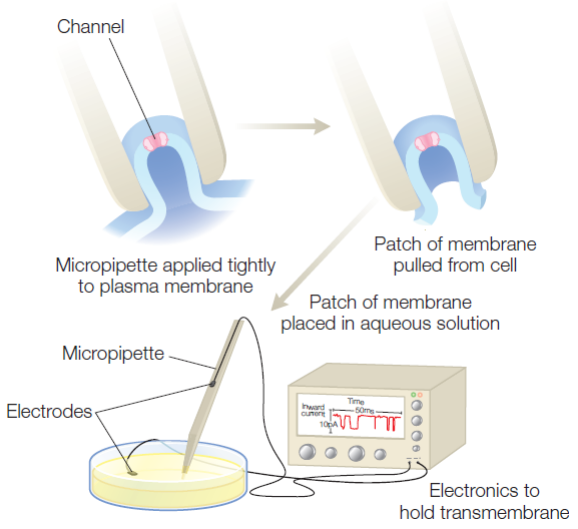
Οι ακουαπορίνες. (α) Η τρισδιάστατη δομή της ακουαπορίνης. (β) Μία σχηματική αναπαράσταση του καναλιού νερού της ακουαπορίνης.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Κανάλια ιόντων

Οι κυτταρικές μεμβράνες διαθέτουν κανάλια ιόντων που είναι πρωτεΐνες σχεδιασμένες να μεταφέρουν συγκεκριμένα ιόντα κατά μήκος της κυτταρικής μεμβράνης. Διακρίνονται τα κανάλια ιόντων από τους μεταφορείς ιόντων με ορισμένα χαρακτηριστικά. Τα κανάλια μπορούν να μεταφέρουν ιόντα σε σημαντικά ταχύτερο ρυθμό από τους μεταφορείς. Επίσης, ο ρυθμός μεταφοράς ιόντων σε ένα κανάλι είναι συνήθως κλειστός. Δηλαδή, το κανάλι είναι ανοιχτό ή κλειστό ανάλογα με έναν άλλο παράγοντα. Οι παράγοντες ελέγχου είναι η δέσμευση ενός συγκεκριμένου συνδετήρα για κανάλια που καλύπτονται από συνδέτη και η μεμβράνη δυναμικού για κανάλια με πύλη τάσης.



Η τεχνική patch-clamp για μέτρηση ενός μεμονωμένου διαύλου ιόντων.

Co-funded by the Erasmus+ Programme of the European Union

Μοριακή απεικόνιση

Ο φθορισμός έχει χρησιμοποιηθεί για πολλά χρόνια για την οπτικοποίηση κυτταρικών συστατικών. Αρχικά, μικρές οργανικές βαφές προσκολλήθηκαν μέσω της χρήσης αντισωμάτων και με τη χρήση φθοροφόρων που αναγνώριζαν οργανίδια. Πιο πρόσφατα, οι φθορίζουσες πρωτεΐνες έδωσαν την ευκαιρία ανίχνευσης έκφρασης γονιδίων, διακίνησης πρωτεϊνών και αποκρίσεις σε σήματα. Η Φασματοσκοπία με χρήση φθορισμού έχει επεκταθεί γρήγορα καθώς αυτά τα εργαλεία έχουν γίνει περισσότερο κοινός τόπος, επιτρέποντας στους επιστήμονες να στοχεύουν συγκεκριμένες τοποθεσίες ακόμη και στο μεμονωμένο μοριακό επίπεδο.

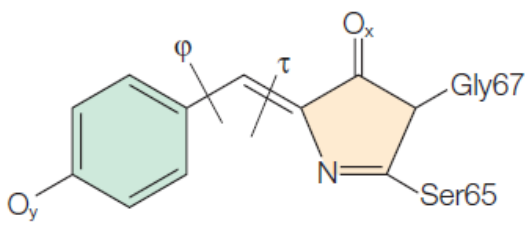
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πράσινη φθορίζουσα πρωτεΐνη

Η πράσινη φθορίζουσα πρωτεΐνη (GFP) ανακαλύφθηκε ως συντροφική πρωτεΐνη στην ακουορίνη, μια χημειοφωταύγεια πρωτεΐνη από τη μέδουσα Aequorea, και βρέθηκε ότι αποτελείται από ένα μόνο πολυπεπτίδιο με 238 αμινοξύ υπολείμματα. Το GFP εκπέμπει πράσινο φθορισμό όταν ακτινοβολείται με υπεριώδη ακτινοβολία. Η σωστή έκφραση του χρωμοφόρου του GFP βρέθηκε ότι απαιτεί μοριακό οξυγόνο αλλά να είναι ανεξάρτητο από οποιοδήποτε ένζυμο. Βιοσύνθεση του χρωμοφόρου βρέθηκε να αποτελεί εξαίρεση σε σύγκριση με τις περισσότερες πρωτεΐνες που περιέχουν χρωμοφόρο καθώς η πρωτεΐνη βρέθηκε ότι καταλύει το σύνθεση της χρωστικής από την πολυπεπτιδική αλυσίδα.

Στις περισσότερες περιπτώσεις, το γονίδιο για το GFP μπορεί να συνδεθεί με ένα γονίδιο του ενδιαφέρον και η προκύπτουσα πρωτεΐνη θα είναι φθορίζουσα, επιτρέποντας παρακολούθηση γονιδίων με χρήση μοριακής βιολογίας.



Δομή του σχηματίστηκε χρωμοφόρο του GFP μετά την κυκλοποίηση του αμινοξέος υπολείμματα οξέος Ser-65, Tyr-66, και Gly-67.

Απεικόνιση σε οργανισμούς

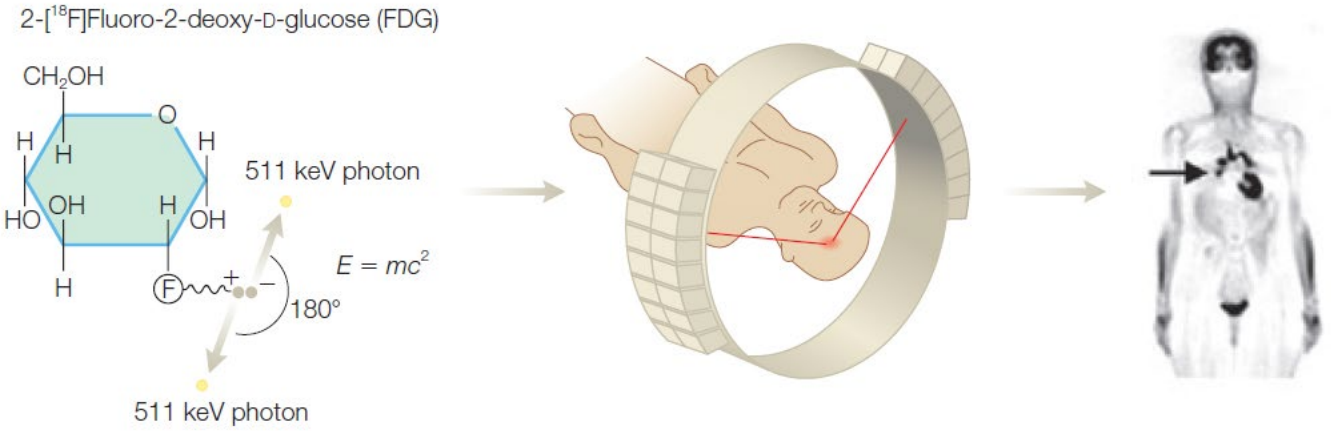
Πολλές από τις φασματοσκοπικές τεχνικές, όπως η μαγνητική τομογραφία και ο φθορισμός, χρησιμοποιούνται σε κλινικές ρυθμίσεις για οπτικοποίηση σε μεταβολικές διεργασίες στο ανθρώπινο σώμα. Μια άλλη τεχνική απεικόνισης, Η τομογραφία εκπομπής ποζιτρονίων (PET), εξετάζεται όλο και περισσότερο για θεραπεία ασθενών στην εξέταση μοριακών διεργασιών και την αποτυχία τους σε ασθένεια. Σε αυτή την τεχνική, μια συγκεκριμένη διαδικασία είναι στόχος και κατάλληλη σχεδιασμένος ανιχνευτής. Οι ανιχνευτές που ανιχνεύονται στο PET είναι φάρμακα ή ανάλογα που έχουν επισημανθεί με ραδιοϊσότοπα. Οι μετρήσεις είναι ευαίσθητες και πολύ χαμηλές δόσεις μπορούν να εφαρμοστούν για να ελαχιστοποιηθούν οι απρόβλεπτες παρενέργειες. Όπως υποδηλώνει το όνομα, αυτή η τεχνική βασίζεται στην τομογραφία, που σημαίνει ότι η προκύπτουσα πληροφορία μετριέται σε τρεις διαστάσεις έτσι ώστε να μπορεί να χαρτογραφηθεί στο σώμα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ραδιενεργή διάσπαση

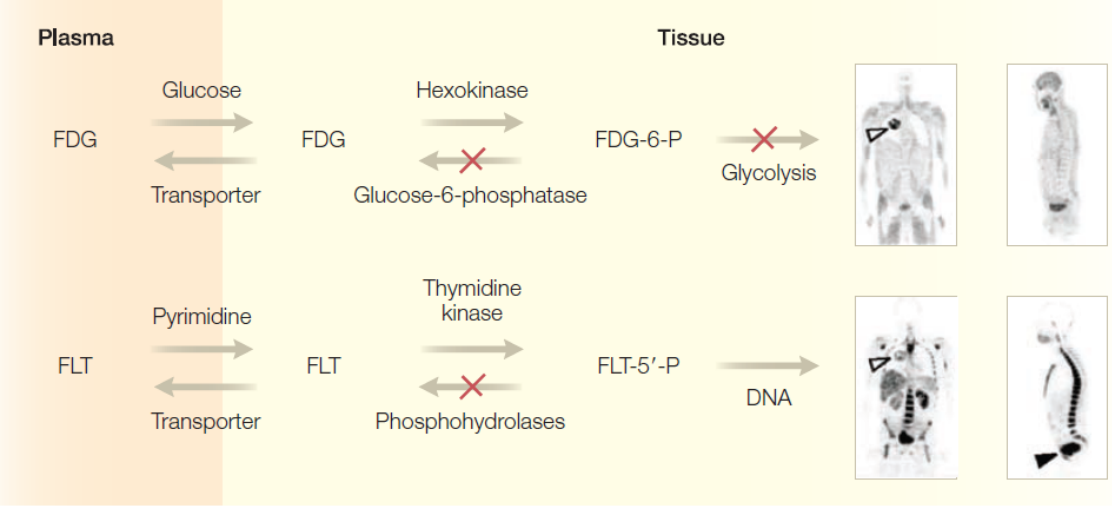
Υπάρχουν πολλές διαφορετικές ραδιενεργές διεργασίες στις οποίες ορισμένα ισότοπα στοιχεία μπορούν να υποστούν. Για να κατανοήσουμε αυτές τις διαδικασίες είναι απαραίτητο να παρακολουθείτε τα συγκεκριμένα ισότοπα που εμπλέκονται. Για τις διαδικασίες αποσύνθεσης κάθε ισότοπο συμβολίζεται με τον μαζικό αριθμό, ο οποίος ισούται με τον αριθμό των πρωτονίων και των νετρονίων, και ο ατομικός αριθμός, που ισούται με τον αριθμό πρωτονίων.



Co-funded by the Erasmus+ Programme of the European Union

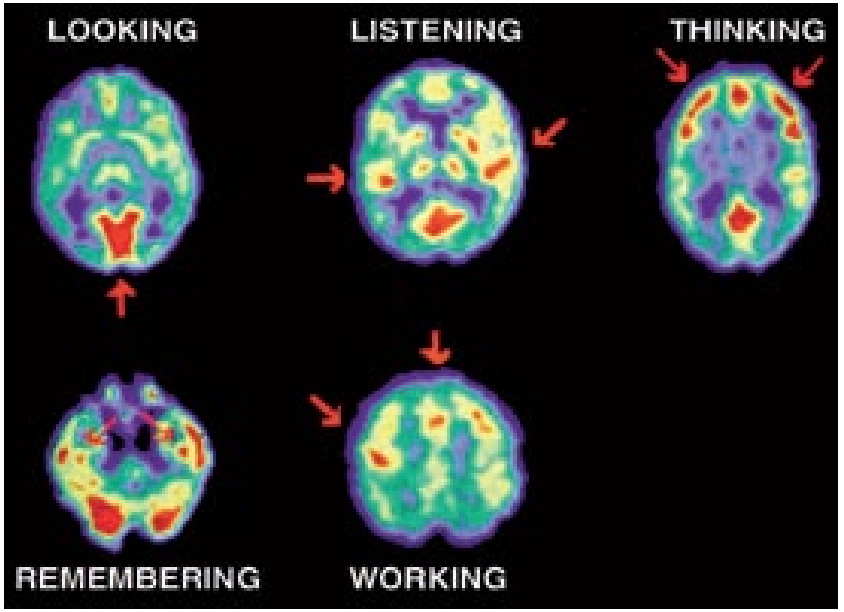
PET

Η τεχνική του PET είναι μια τεχνική απεικόνισης που κάνει χρήση της ποζιτρονεκπομπής ισότοπα ως ανιχνευτές συγκεκριμένων βιοχημικών διεργασιών *in vivo* . Στο PET, το άτομο εγχέεται με ένα μόριο που περιέχει μια ετικέτα με ισότοπο που θα εκπέμπει ποζιτρόνια .



Co-funded by the Erasmus+ Programme of the European Union

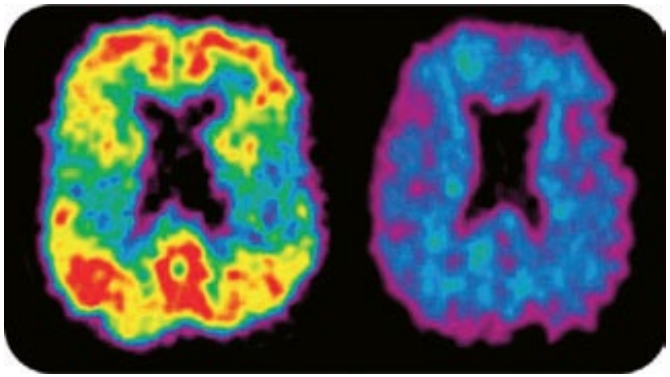
Εκτός από αυτές τις κλινικές χρήσεις, το PET μπορεί να είναι χρησιμοποιείται για τον χαρακτηρισμό των αποκρίσεων διέγερσης του εγκεφάλου. Για παράδειγμα, συγκεκριμένες περιοχές του ο εγκέφαλος έχει υψηλό μεταβολισμό γλυκόζης ως απόκριση σε διαφορετικά αισθητηριακά ερεθίσματα. Το PET παρέχει μια σχετικά καλοήθη προσέγγιση για τη μελέτη της εγκεφαλικής δραστηριότητας και την εξέταση για πιθανές θεραπείες. Για παράδειγμα, με σύνδεση ένα γονίδιο αναφοράς PET σε ένα πιθανό θεραπευτικό γονίδιο, είναι δυνατό να εντοπιστεί πού το θεραπευτικό γονίδιο μεταφράζεται και συσχετίζεται το γονίδιο δραστηριότητας με την φυσιολογική απόκριση.



Co-funded by the Erasmus+ Programme of the European Union

Νόσος Πάρκινσον

Η νόσος του Πάρκινσον είναι η δεύτερη πιο συχνή νευροεκφυλιστική νόσος μετά τη νόσο του Αλτσχάιμερ. Το πρώτο προειδοποιητικό σημάδι είναι η ανάπτυξη τρόμου στο ένα χέρι. Καθώς η ασθένεια εξελίσσεται, η εκούσια κίνηση γίνεται πιο αργά με ορισμένους ασθενείς να παρουσιάζουν προσωρινή απώλεια της ικανότητας για κίνηση.



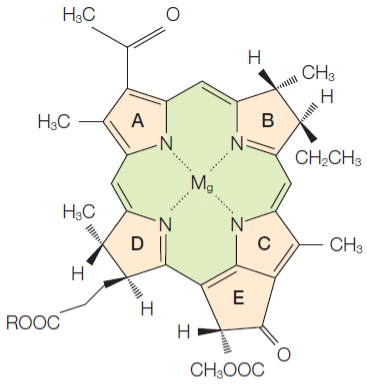
Σε σύγκριση με ένα υγιές άτομο (δεξιά), ένας ασθενής με νόσο του Πάρκινσον (αριστερά) θα έχει επιδείνωση της δραστηριότητας του μεταφορέα ντοπαμίνης, που είναι είναι εμφανές στις σαρώσεις PET.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

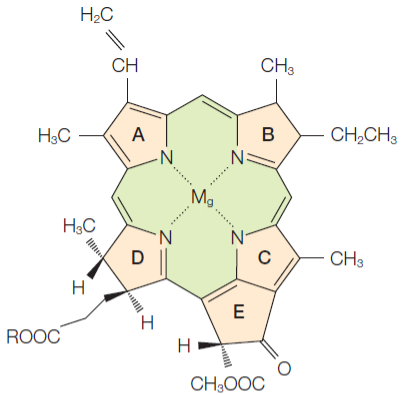
Co-funded by the Erasmus+ Programme of the European Union

Φωτοσύνθεση

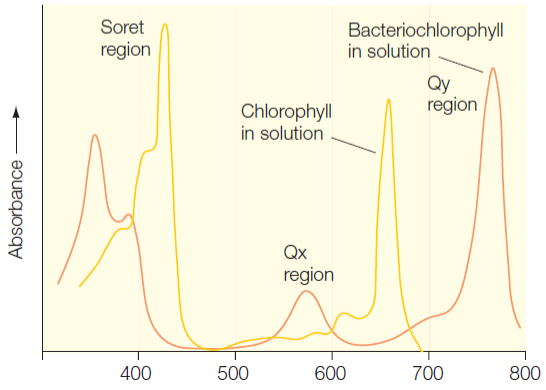
Η φωτοσύνθεση είναι η βιολογική διαδικασία με την οποία η ενέργεια του ήλιου μετατρέπεται σε ενώσεις πλούσιες σε ενέργεια που χρησιμοποιούνται για την κινητοποίηση των κυτταρικών διαδικασιών.



Bacteriochlorophyll *a*



Chlorophyll *a*



Co-funded by the Erasmus+ Programme of the European Union

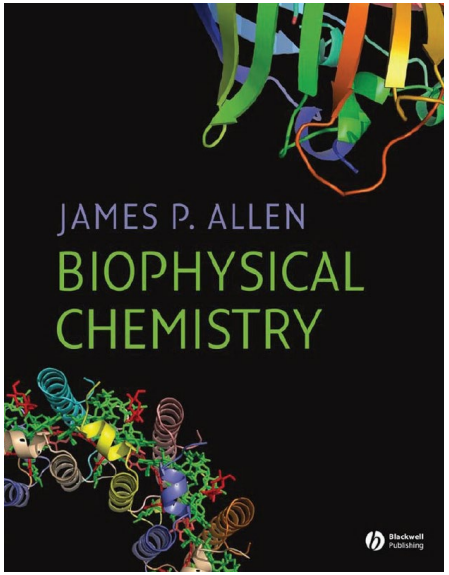
Ερωτήσεις

- Η διπλοστοιβάδα λιπιδίων είναι διαπερατή σε ποιο τύπο μικρών μορίων;
- Ποιος είναι ο βιολογικός ρόλος των μεταφορέων;
- Πώς μπορεί να χρησιμοποιηθεί το PET για την παρακολούθηση του μεταβολισμού της γλυκόζης;
- Πώς μπορεί να χρησιμοποιηθεί το PET με τη νόσο του Αλτσχάιμερ και τις σχετικές ασθένειες;

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

βιβλιογραφικές αναφορές



Βιοφυσική Χημεία,
Τζέιμς Π. Άλεν
2008 από την Blackwell Publishing

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 2. Γενετική

Θέμα 1. Μοριακή Γενετική

Μάθημα 1. DNA: Το γενετικό υλικό, Πειραματική απόδειξη της γενετικής, Λειτουργία DNA, Γενετικός Ρόλος του DNA στους Βακτηριοφάγους

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Πειραματική Απόδειξη της Γενετικής Λειτουργίας του DNA
- Γενετικός ρόλος του DNA σε βακτηριοφάγους

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Κάθε είδος ζωντανού οργανισμού έχει ένα μοναδικό σύνολο κληρονομικών χαρακτηριστικών που το κάνει διαφορετικό από τα άλλα είδη. Κάθε είδος έχει το δικό του αναπτυξιακό σχέδιο, που συχνά περιγράφεται ως ένα είδος «σχεδίου» για την οικοδόμηση του οργανισμού, το οποίο κωδικοποιείται στα μόρια DNA που υπάρχουν στα κύτταρά του. Αυτό το αναπτυξιακό σχέδιο καθορίζει τα χαρακτηριστικά που κληρονομούνται.

Co-funded by the Erasmus+ Programme of the European Union

Το ότι ο πυρήνας του κυττάρου παίζει βασικό ρόλο στην κληρονομικότητα αναγνωρίστηκε στη δεκαετία του 1870 από την παρατήρηση ότι οι πυρήνες των ανδρικών και θηλυκών αναπαραγωγικών κυττάρων υφίστανται σύντηξη κατά τη διαδικασία της γονιμοποίησης. Λίγο αργότερα, **τα χρωμοσώματα** παρατηρήθηκαν για πρώτη φορά μέσα στον πυρήνα ως αντικείμενα που μοιάζουν με νήματα που γίνονται ορατά στο μικροσκόπιο φωτός όταν το κύτταρο χρωματίζεται με ορισμένες βαφές. Τα χρωμοσώματα βρέθηκαν να παρουσιάζουν μια χαρακτηριστική συμπεριφορά «διαίρεσης» στην οποία κάθε θυγατρικό κύτταρο που σχηματίζεται από κυτταρική διαίρεση λαμβάνει ένα πανομοιότυπο συμπλήρωμα χρωμοσωμάτων

Co-funded by the Erasmus+ Programme of the European Union

Περαιτέρω στοιχεία για τη σημασία των χρωμοσωμάτων παρασχέθηκαν από την παρατήρηση ότι, ενώ ο αριθμός των χρωμοσωμάτων σε κάθε κύτταρο μπορεί να διαφέρει μεταξύ των βιολογικών ειδών, ο αριθμός των χρωμοσωμάτων είναι σχεδόν πάντα σταθερός στα κύτταρα οποιουδήποτε συγκεκριμένου είδους. Αυτά τα χαρακτηριστικά των χρωμοσωμάτων ήταν καλά κατανοητά από το 1900 περίπου, και έκαναν να φαίνεται πιθανό ότι τα χρωμοσώματα ήταν οι φορείς των γονιδίων.

Co-funded by the Erasmus+ Programme of the European Union

Μικροσκοπικές μελέτες με ειδικές χρώσεις έδειξαν ότι το DNA υπάρχει στα χρωμοσώματα. Τα χρωμοσώματα περιέχουν επίσης διάφορους τύπους πρωτεϊνών, αλλά η ποσότητα και τα είδη των χρωμοσωμικών πρωτεϊνών διαφέρουν πολύ από τον έναν τύπο κυττάρου στον άλλο, ενώ η ποσότητα του DNA ανά κύτταρο είναι σταθερή. Επιπλέον, σχεδόν όλο το DNA που υπάρχει σε κύτταρα ανώτερων οργανισμών υπάρχει στα χρωμοσώματα. Αυτά τα επιχειρήματα για το DNA ως γενετικό υλικό δεν ήταν πειστικά, ωστόσο, επειδή οι ακατέργαστες χημικές αναλύσεις είχαν υποδείξει (λανθασμένα, όπως αποδείχθηκε) ότι το DNA δεν έχει τη χημική ποικιλομορφία που απαιτείται σε μια γενετική ουσία.

Co-funded by the Erasmus+ Programme of the European Union

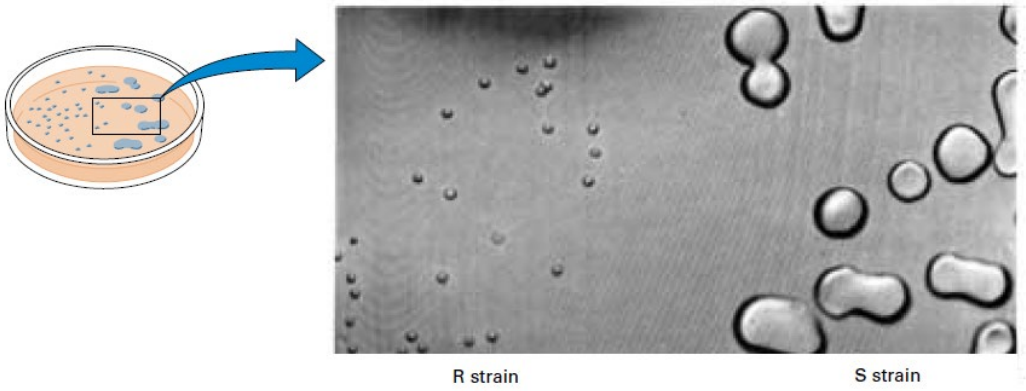
Ο προτιμώμενος υποψήφιος για το γενετικό υλικό ήταν η πρωτεΐνη, επειδή οι πρωτεΐνες ήταν γνωστό ότι ήταν μια εξαιρετικά ποικιλόμορφη συλλογή μορίων. Ως εκ τούτου, οι πρωτεΐνες έγιναν ευρέως αποδεκτές ως το γενετικό υλικό και το DNA θεωρήθηκε ότι λειτουργεί απλώς ως το δομικό πλαίσιο των χρωμοσωμάτων. Τα πειράματα που περιγράφονται παρακάτω κατέδειξαν τελικά ότι το DNA είναι το γενετικό υλικό.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πειραματική Απόδειξη της Γενετικής Λειτουργίας του DNA

Αποικίες τραχιών (R, οι μικρές αποικίες) και λείων (S, οι μεγάλες αποικίες) στελεχών Streptococcus pneumoniae. Οι αποικίες S είναι μεγαλύτερες λόγω της ζελατινώδους κάψουλας στα κύτταρα S.



Ένα σημαντικό πρώτο βήμα έγινε από τον Frederick Griffith το 1928 όταν έδειξε ότι ένα φυσικό χαρακτηριστικό μπορεί να περάσει από το ένα κύτταρο στο άλλο. Εργαζόταν με δύο στελέχη του βακτηρίου Streptococcus pneumoniae που ταυτοποιήθηκαν ως S και R.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πειραματική Απόδειξη της Γενετικής Λειτουργίας του DNA

Όταν ένα βακτηριακό κύτταρο αναπτύσσεται σε στερεό μέσο, υφίσταται επαναλαμβανόμενες κυτταρικές διαιρέσεις για να σχηματίσει μια ορατή συστάδα κυττάρων που ονομάζεται αποικία. Ο τύπος S του S. pneumoniae συνθέτει μια ζελατινώδη κάψουλα που αποτελείται από σύνθετους υδατάνθρακες (πολυσακχαρίτη). Η περιβάλλουσα κάψουλα κάνει κάθε αποικία μεγάλη και της δίνει μια αστραφτερή ή λεία (S) εμφάνιση. Αυτή η κάψουλα επιτρέπει επίσης στο βακτήριο να προκαλέσει πνευμονία προστατεύοντάς το από τους αμυντικούς μηχανισμούς ενός μολυσμένου ζώου.

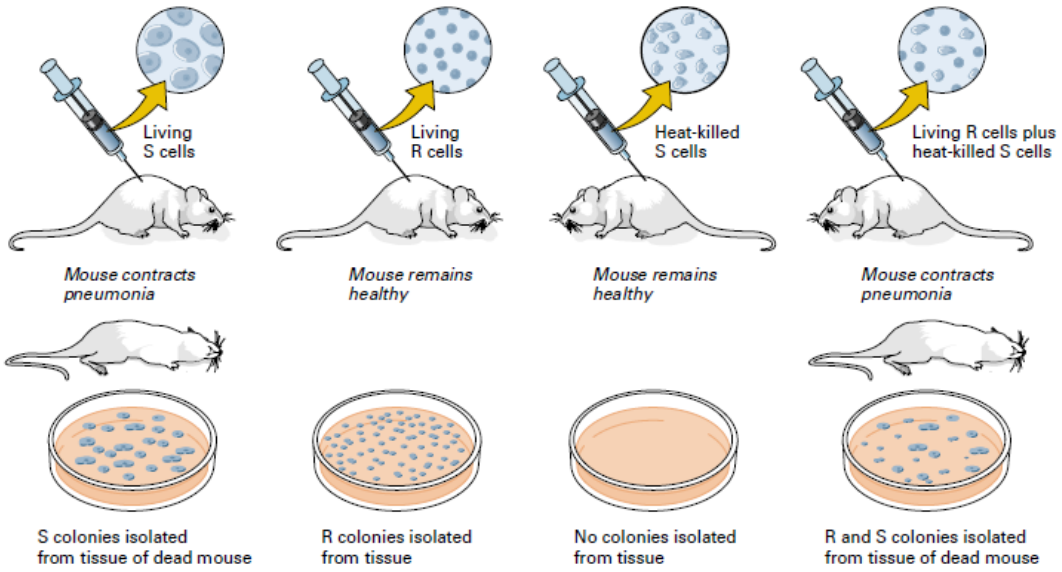
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πειραματική Απόδειξη της Γενετικής Λειτουργίας του DNA

Τα στελέχη R του S. pneumoniae δεν μπορούν να συνθέσουν τον καψικό πολυσακχαρίτη. σχηματίζουν μικρές αποικίες που έχουν τραχιά (R) επιφάνεια. Αυτό το στέλεχος του βακτηρίου δεν προκαλεί πνευμονία, γιατί χωρίς την κάψουλα τα βακτήρια αδρανοποιούνται από το ανοσοποιητικό σύστημα του ξενιστή. Και οι δύο τύποι βακτηρίων «αναπαράγονται αληθινά» με την έννοια ότι οι απόγονοι που σχηματίζονται από την κυτταρική διαίρεση έχουν τον καψικό τύπο του γονέα, είτε S είτε R.

Co-funded by the Erasmus+ Programme of the European Union



Το πείραμα του Γκρίφιθ που δείχνει τον βακτηριακό μετασχηματισμό. Ένα ποντίκι παραμένει υγιές εάν εγχυθεί είτε το μη λοιμώδες στέλεχος R του S. pneumoniae ή θραύσματα κυττάρων που θανατώνονται από τη θερμότητα του συνήθως λοιμογόνου στελέχους S. Τα κύτταρα R παρουσία κυττάρων S που θανατώνονται από τη θερμότητα μετασχηματίζονται στο λοιμογόνο στέλεχος S, προκαλώντας πνευμονία στο ποντίκι.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πειραματική Απόδειξη της Γενετικής Λειτουργίας του DNA

Τα ποντίκια στα οποία έγινε ένεση με ζωντανά κύτταρα S παθαίνουν πνευμονία. Τα ποντίκια στα οποία έγινε ένεση είτε με ζωντανά κύτταρα R είτε με κύτταρα S που σκοτώθηκαν από τη θερμότητα παραμένουν υγιή. Εδώ είναι το κρίσιμο εύρημα του Γκρίφιθ:

ποντίκια στα οποία γίνεται ένεση με ένα μείγμα ζωντανών κυττάρων R και κυττάρων S που σκοτώνονται από τη θερμότητα προσβάλλονται από την ασθένεια — συχνά πεθαίνουν από πνευμονία.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πειραματική Απόδειξη της Γενετικής Λειτουργίας του DNA

Ο μετασχηματισμός στον στρεπτόκοκκο ανακαλύφθηκε αρχικά το 1928, αλλά μόλις το 1944 αναγνωρίστηκε η χημική ουσία που είναι υπεύθυνη για την αλλαγή των κυττάρων R σε κύτταρα S. Σε ένα πείραμα ορόσημο, οι Oswald Avery, Colin MacLeod και Maclyn McCarty έδειξαν ότι η ουσία που προκάλεσε τον μετασχηματισμό των κυττάρων R σε κύτταρα S ήταν το DNA. Κάνοντας αυτά τα πειράματα, έπρεπε πρώτα να αναπτύξουν χημικές διαδικασίες για την απομόνωση σχεδόν καθαρού DNA από κύτταρα, κάτι που δεν είχε γίνει ποτέ πριν.

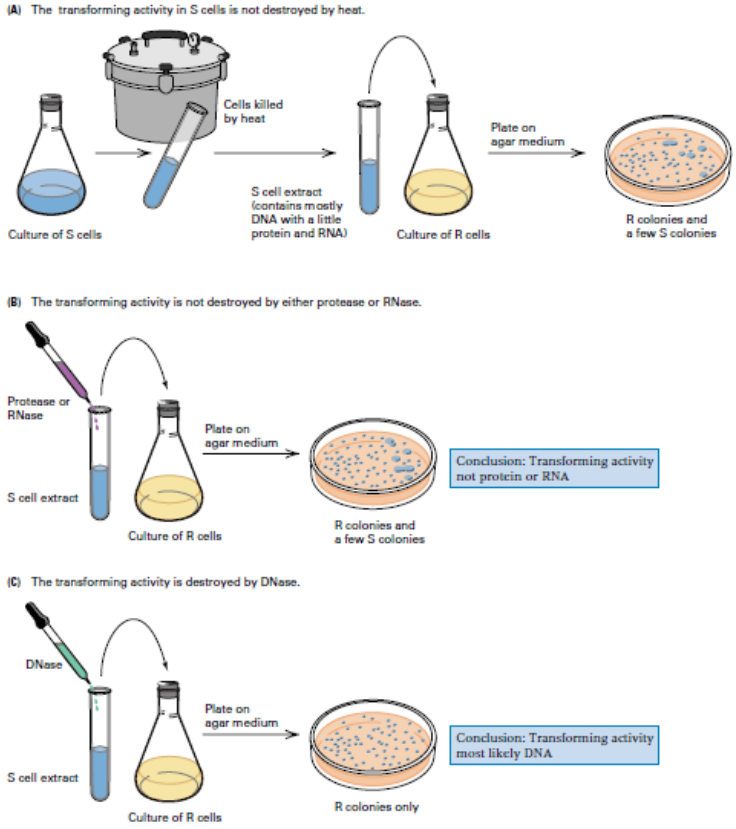
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πειραματική Απόδειξη της Γενετικής Λειτουργίας του DNA

Όταν πρόσθεσαν DNA που απομονώθηκε από κύτταρα S σε αναπτυσσόμενες καλλιέργειες κυττάρων R, παρατήρησαν μετασχηματισμό: Παρήχθησαν μερικά κύτταρα κυττάρων τύπου S. Αν και τα παρασκευάσματα DNA περιείχαν ίχνη πρωτεΐνης και RNA (ριβονουκλεϊκό οξύ, ένα άφθονο κυτταρικό μακρομόριο που σχετίζεται χημικά με το DNA), η δραστηριότητα μετασχηματισμού δεν αλλοιώθηκε από θεραπείες που κατέστρεψαν είτε την πρωτεΐνη είτε το RNA. Ωστόσο, οι θεραπείες που κατέστρεψαν το DNA εξάλειψαν τη μεταμορφωτική δραστηριότητα

Co-funded by the Erasmus+ Programme of the European Union

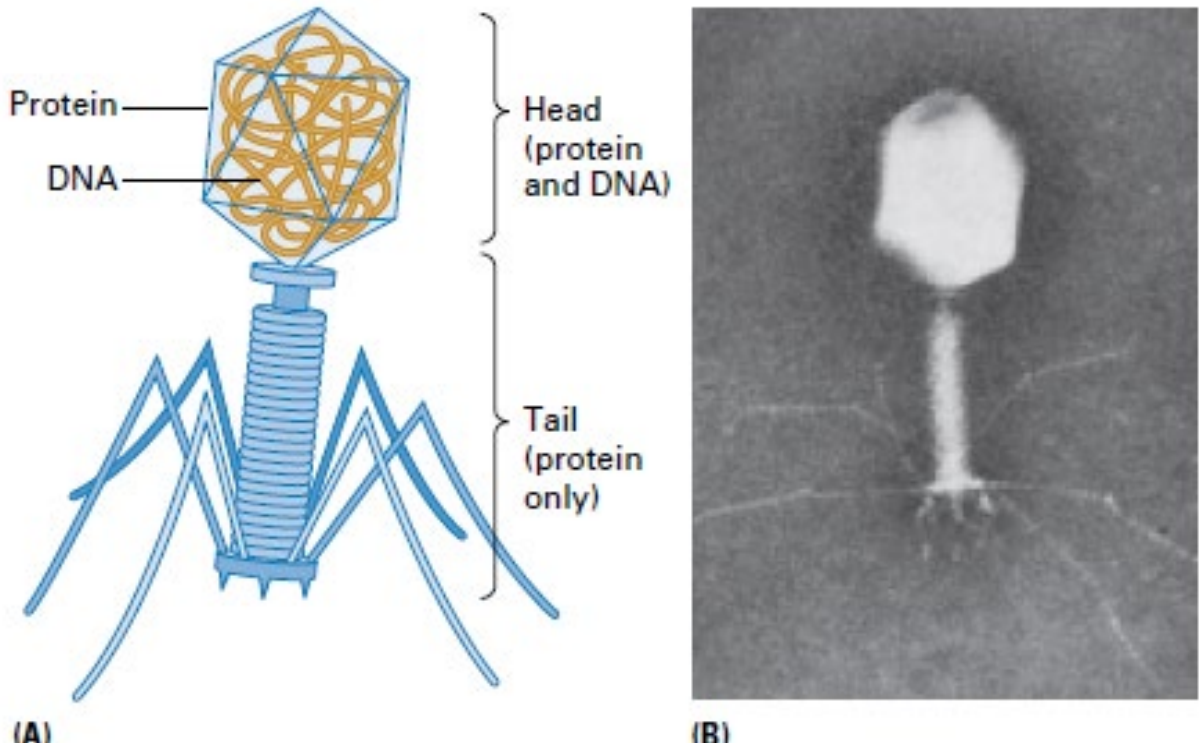


Ένα διάγραμμα του πειράματος Avery–MacLeod–McCarty που έδειξε ότι το DNA είναι το ενεργό υλικό στον βακτηριακό μετασχηματισμό. (Α) Το καθαρισμένο DNA που εξάγεται από κύτταρα S που έχουν σκοτωθεί από τη θερμότητα μπορεί να μετατρέψει μερικά ζωντανά κύτταρα R σε κύτταρα S, αλλά το υλικό μπορεί να εξακολουθεί να περιέχει μη ανιχνεύσιμα ίχνη πρωτεΐνης και/ή RNA. (Β) Η δραστηριότητα μετασχηματισμού δεν καταστρέφεται ούτε από πρωτεάση ούτε από RNase. (Γ) Η μεταμορφωτική δραστηριότητα καταστρέφεται από την DNase και έτσι πιθανότατα αποτελείται από DNA.

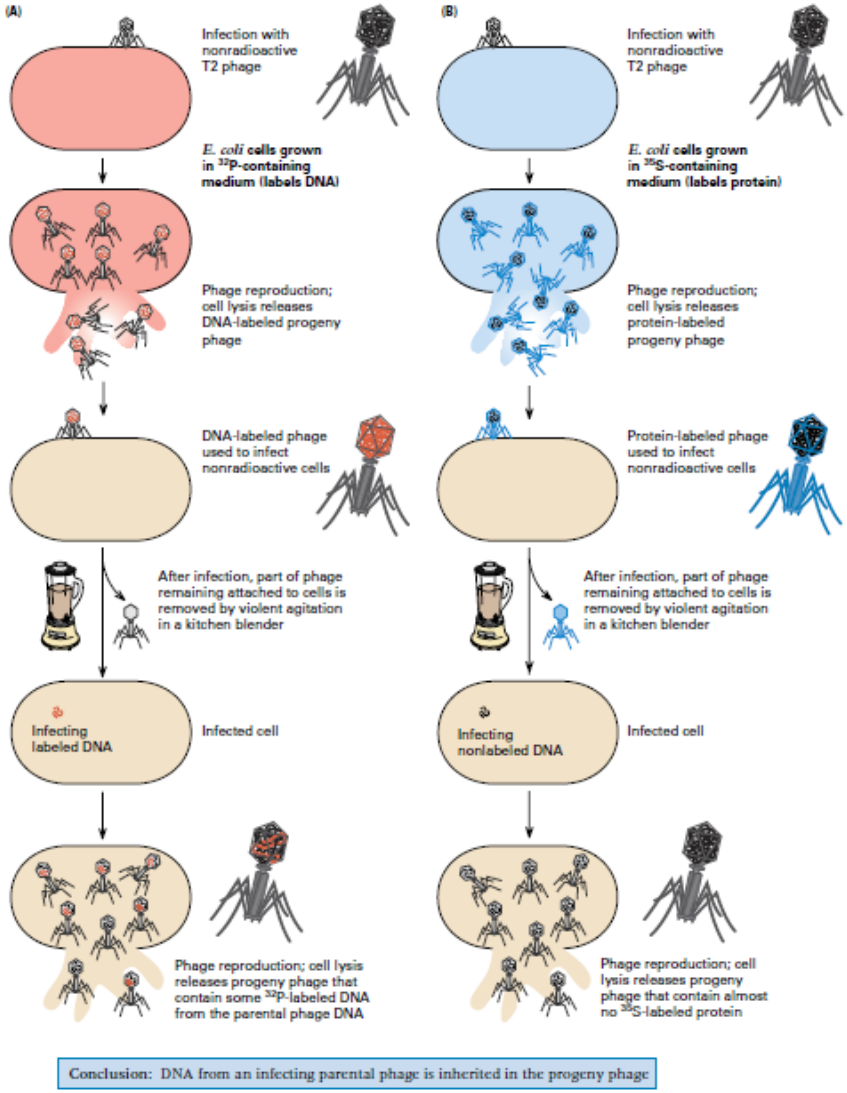
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Γενετικός ρόλος του DNA σε βακτηριοφάγους



Α) Σχέδιο E. coli φάγου T2, που δείχνει διάφορα συστατικά. Το DNA περιορίζεται στο εσωτερικό του κεφαλιού. Β) Μια ηλεκτρονική μικρογραφία του φάγου T4, ενός στενά συνδεδεμένου φάγου.



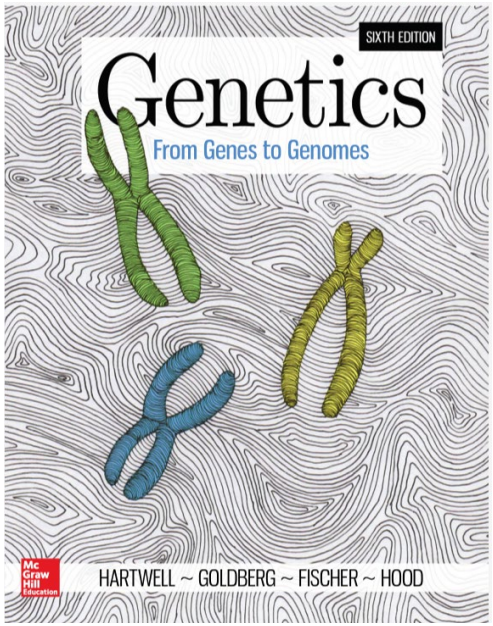
Το πείραμα Hershey–Chase («μπλέντερ») που δείχνει ότι το DNA, όχι η πρωτεΐνη, είναι υπεύθυνο για την κατεύθυνση της αναπαραγωγής του φάγου T2 σε μολυσμένα κύτταρα E. coli. (A) Το ραδιενεργό DNA μεταδίδεται στους απογόνους φάγους σε σημαντικές ποσότητες. (B) Η ραδιενεργή πρωτεΐνη μεταδίδεται στους απογόνους φάγους σε αμελητέες ποσότητες.

Co-funded by the Erasmus+ Programme of the European Union

Επιθεώρηση των ερωτήσεων

- Ποιος είναι ο ρόλος του DNA;
- Πώς διατηρείται το γενετικό υλικό στο κύτταρο;
- Πώς αποδεικνύουν τον ρόλο του DNA;

Co-funded by the Erasmus+ Programme of the European Union



Leland Hartwell , Michael Goldberg, Janice Fischer, Lee Hood, Charles F. Aquadro , Bruce Bejcek . (2015) Γενετική: από τα γονίδια στα γονιδιώματα. McGraw-Hill Education, Νέα Υόρκη, Η.Π.Α

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 2. Γενετική

Θέμα 1. Μοριακή Γενετική

Μάθημα 2. Δομή DNA: Η διπλή έλικα

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Η Δομή του DNA
- Πώς το DNA είναι διατεταγμένο στο κύτταρο

Co-funded by the Erasmus+ Programme of the European Union

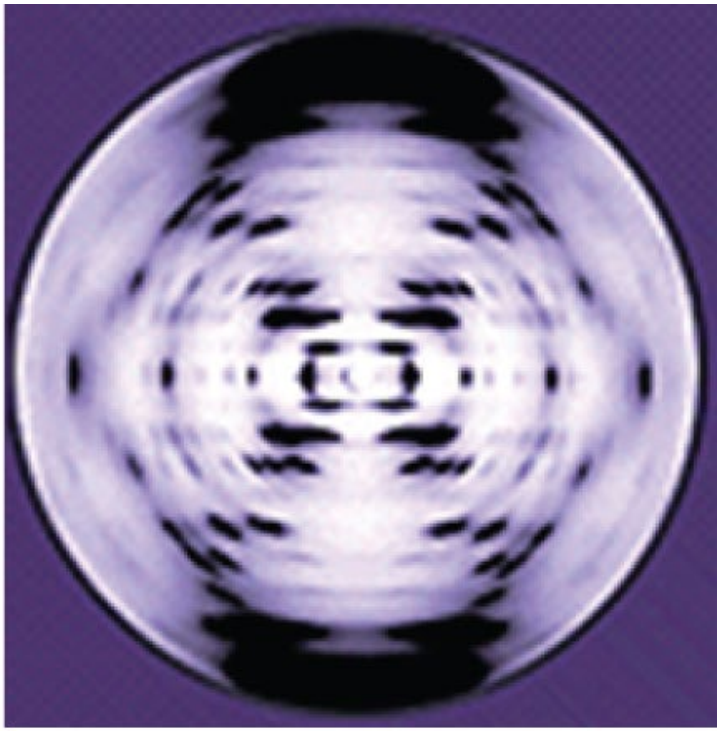
Εισαγωγή

Στη δεκαετία του 1950, ο Φράνσις Κρικ και ο Τζέιμς Γουάτσον εργάστηκαν μαζί στο Πανεπιστήμιο του Κέιμπριτζ, στην Αγγλία, για να καθορίσουν τη δομή του DNA. Άλλοι επιστήμονες, όπως ο Linus Pauling και ο Maurice Wilkins, εξερευνούσαν επίσης ενεργά αυτό το πεδίο. Ο Pauling είχε ανακαλύψει τη δευτερογενή δομή των πρωτεϊνών χρησιμοποιώντας κρυσταλλογραφία ακτίνων Χ. Η κρυσταλλογραφία ακτίνων Χ είναι μια μέθοδος για τη διερεύνηση της μοριακής δομής με την παρατήρηση των μοτίβων που σχηματίζονται από τις ακτίνες Χ που εκτοξεύονται μέσω ενός κρυστάλλου της ουσίας. Τα σχέδια δίνουν σημαντικές πληροφορίες για τη δομή του μορίου που μας ενδιαφέρει. Στο εργαστήριο του Wilkins, η ερευνήτρια Rosalind Franklin χρησιμοποιούσε κρυσταλλογραφία ακτίνων Χ για να κατανοήσει τη δομή του DNA. Ο Watson και ο Crick μπόρεσαν να συνθέσουν το παζλ του μορίου DNA χρησιμοποιώντας τα δεδομένα του Franklin.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

(a)



(b)

Οι πρωτοπόροι επιστήμονες (α) James Watson και Francis Crick απεικονίζονται εδώ με τον Αμερικανό γενετιστή Maclyn McCarty. Η επιστήμονας Rosalind Franklin ανακάλυψε (β) το μοτίβο περίθλασης ακτίνων Χ του DNA, το οποίο βοήθησε να διαλευκανθεί η δομή της διπλής έλικας του. (α: τροποποίηση της εργασίας της Marjorie McCarty· β: τροποποίηση της εργασίας από το NIH

Co-funded by the Erasmus+ Programme of the European Union

Οι Watson και Crick είχαν επίσης βασικές πληροφορίες διαθέσιμες από άλλους ερευνητές, όπως τους κανόνες του Chargaff. Ο Chargaff είχε δείξει ότι από τα τέσσερα είδη μονομερών (νουκλεοτίδια) που υπάρχουν σε ένα μόριο DNA, δύο τύποι υπήρχαν πάντα σε ίσες ποσότητες και οι υπόλοιποι δύο τύποι ήταν επίσης πάντα παρόντες σε ίσες ποσότητες. Αυτό σήμαινε ότι ήταν πάντα ζευγάρια με κάποιο τρόπο. Το 1962, οι James Watson, Francis Crick και Maurice Wilkins τιμήθηκαν με το Νόμπελ Ιατρικής για το έργο τους στον προσδιορισμό της δομής του DNA.

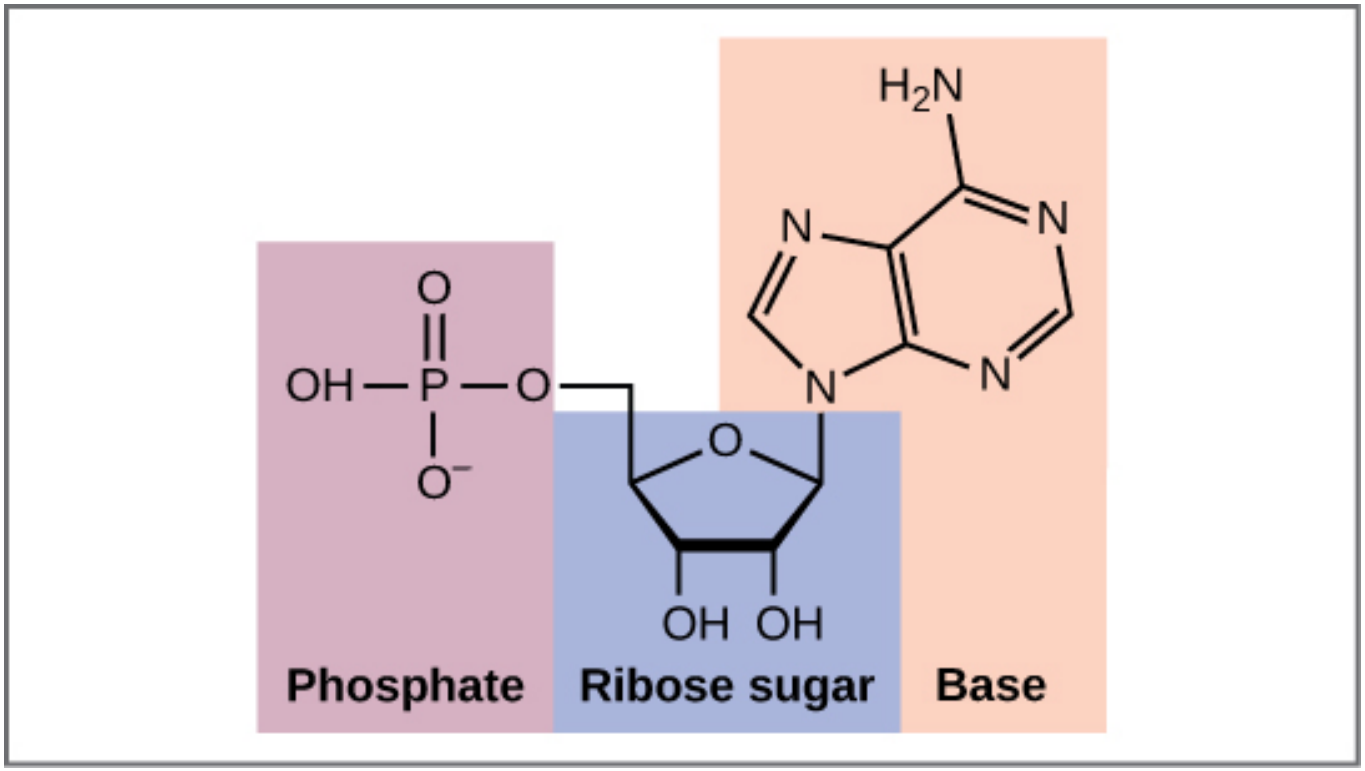
Η Δομή του DNA

Υπάρχουν δύο τύποι νουκλεϊκών οξέων, το δεοξυριβονουκλεϊκό οξύ (DNA) και το ριβονουκλεϊκό οξύ (RNA). Τα δομικά στοιχεία του DNA είναι νουκλεοτίδια, τα οποία αποτελούνται από τρία μέρη: μια δεοξυριβόζη (σάκχαρο 5 άνθρακα), μια φωσφορική ομάδα και μια αζωτούχα βάση. Υπάρχουν τέσσερις τύποι αζωτούχων βάσεων στο DNA. Η αδενίνη (A) και η γουανίνη (G) είναι πουρίνες διπλού δακτυλίου και η κυτοσίνη (C) και η θυμίνη (T) είναι μικρότερες πυριμιδίνες ενός δακτυλίου. Το νουκλεοτίδιο ονομάζεται σύμφωνα με την αζωτούχα βάση που περιέχει.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

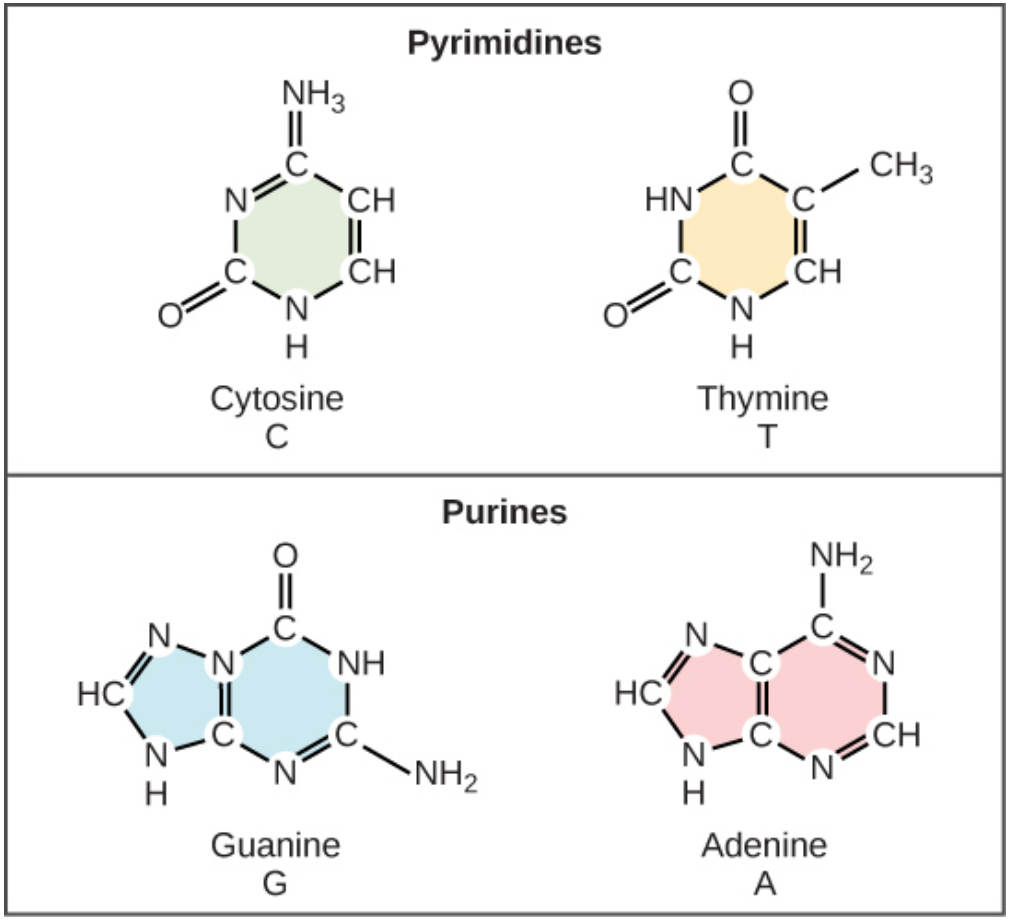
Η Δομή του DNA



Κάθε νουκλεοτίδιο DNA αποτελείται από ένα σάκχαρο, μια φωσφορική ομάδα και μια βάση.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Δομή



Η κυτοσίνη και η θυμίνη είναι πυριμιδίνες. Η γουανίνη και η αδενίνη είναι πουρίνες

Co-funded by the Erasmus+ Programme of the European Union

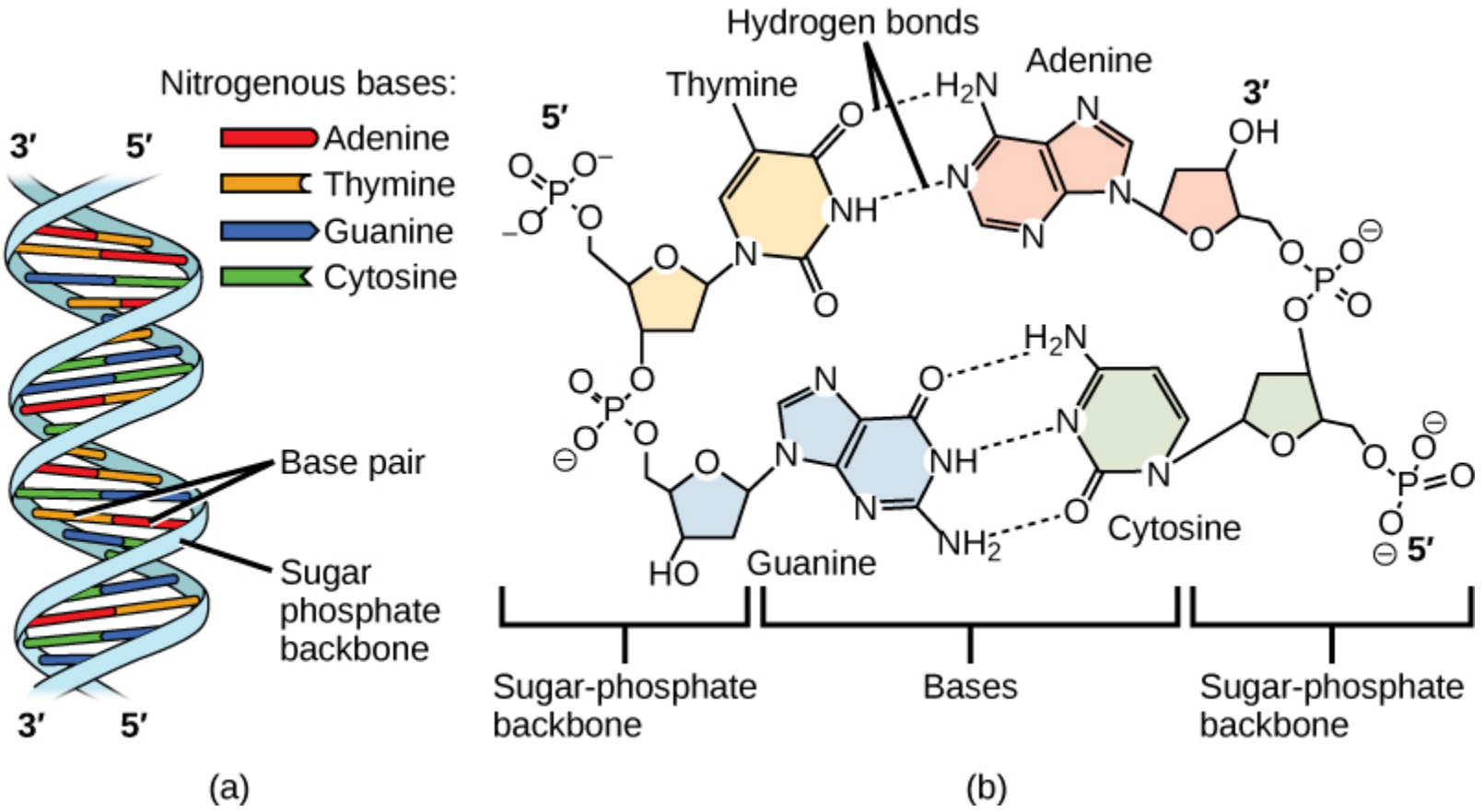
Η φωσφορική ομάδα ενός νουκλεοτιδίου συνδέεται ομοιοπολικά με το μόριο σακχάρου του επόμενου νουκλεοτιδίου και ούτω καθεξής, σχηματίζοντας ένα μακρύ πολυμερές μονομερών νουκλεοτιδίων. Οι ομάδες σακχάρου-φωσφορικού παρατάσσονται σε μια «ραχοκοκαλιά» για κάθε κλώνο DNA και οι βάσεις νουκλεοτιδίων προεξέχουν από αυτήν τη ραχοκοκαλιά. Τα άτομα άνθρακα του σακχάρου με πέντε άνθρακα αριθμούνται δεξιόστροφα από το οξυγόνο ως 1', 2', 3', 4' και 5' (1' διαβάζεται ως "ένας πρώτος"). Η φωσφορική ομάδα συνδέεται με τον άνθρακα 5' ενός νουκλεοτιδίου και τον άνθρακα 3' του επόμενου νουκλεοτιδίου. Στη φυσική του κατάσταση, κάθε μόριο DNA στην πραγματικότητα αποτελείται από δύο μονούς κλώνους που συγκρατούνται μαζί κατά το μήκος τους με δεσμούς υδρογόνου μεταξύ των βάσεων.

Co-funded by the Erasmus+ Programme of the European Union

Οι Watson και Crick πρότειναν ότι το DNA αποτελείται από δύο κλώνους που είναι στριμμένοι ο ένας γύρω από τον άλλο για να σχηματίσουν μια δεξιόστροφη έλικα, που ονομάζεται διπλή έλικα. Το ζευγάρωμα βάσεων λαμβάνει χώρα μεταξύ μιας πουρίνης και της πυριμιδίνης: συγκεκριμένα, τα ζεύγη A με το T και τα ζεύγη G με το C. Με άλλα λόγια, η αδενίνη και η θυμίνη είναι συμπληρωματικά ζεύγη βάσεων και η κυτοσίνη και η γουανίνη είναι επίσης συμπληρωματικά ζεύγη βάσεων. Αυτή είναι η βάση για τον κανόνα του Chargaff. Λόγω της συμπληρωματικότητας τους, υπάρχει τόση αδενίνη όση θυμίνη σε ένα μόριο DNA και τόση γουανίνη όσο και η κυτοσίνη.

Co-funded by the Erasmus+ Programme of the European Union

Η αδενίνη και η θυμίνη συνδέονται με δύο δεσμούς υδρογόνου και η κυτοσίνη και η γουανίνη συνδέονται με τρεις δεσμούς υδρογόνου. Τα δύο σκέλη είναι αντιπαράλληλα στη φύση τους. Δηλαδή, ο ένας κλώνος θα έχει τον άνθρακα 3′ της ζάχαρης στην “ανοδική” θέση, ενώ ο άλλος κλώνος θα έχει τον άνθρακα 5′ στην ανοδική θέση. Η διάμετρος της διπλής έλικας του DNA είναι ομοιόμορφη καθ' όλη τη διάρκεια, επειδή μια πουρίνη (δύο δακτύλιοι) ζευγαρώνει πάντα με μια πυριμιδίνη (ένας δακτύλιος) και τα συνδυασμένα μήκη τους είναι πάντα ίσα.



Το DNA (α) σχηματίζει μια δίκλωνη έλικα και (β) η αδενίνη ζεύγη με τη θυμίνη και τα ζεύγη κυτοσίνης με τη γουανίνη. (α: τροποποίηση του έργου των Jerome Walker, Dennis Myts)

Co-funded by the Erasmus+ Programme of the European Union

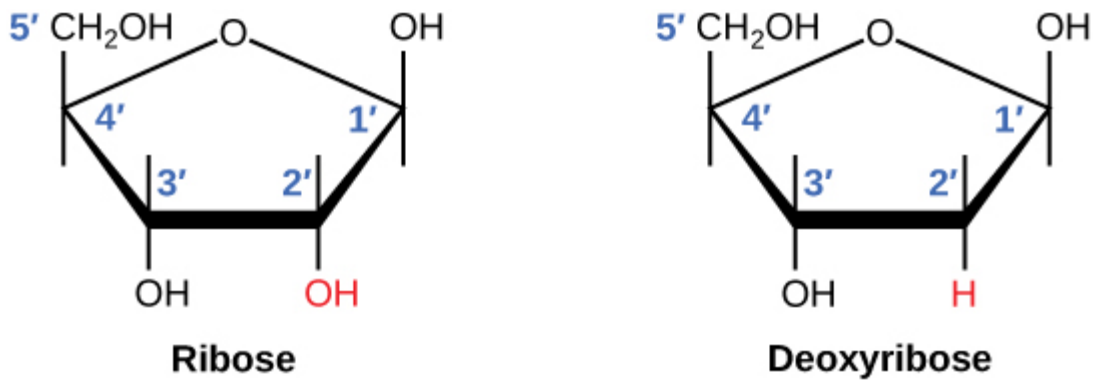
Η Δομή του RNA

Υπάρχει ένα δεύτερο νουκλεϊκό οξύ σε όλα τα κύτταρα που ονομάζεται ριβονουκλεϊκό οξύ ή RNA. Όπως το DNA, το RNA είναι ένα πολυμερές νουκλεοτιδίων. Κάθε ένα από τα νουκλεοτίδια στο RNA αποτελείται από μια αζωτούχα βάση, ένα σάκχαρο πέντε άνθρακα και μια φωσφορική ομάδα. Στην περίπτωση του RNA, το σάκχαρο πέντε άνθρακα είναι ριβόζη, όχι δεοξυριβόζη. Η ριβόζη έχει μια ομάδα υδροξυλίου στο 2' άνθρακα, σε αντίθεση με τη δεοξυριβόζη, η οποία έχει μόνο ένα άτομο υδρογόνου.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Η Δομή του RNA



Η διαφορά μεταξύ της ριβόζης που βρίσκεται στο RNA και της δεοξυριβόζης που βρίσκεται στο DNA είναι ότι η ριβόζη έχει μια ομάδα υδροξυλίου στον άνθρακα 2'.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Η Δομή του RNA

Τα νουκλεοτίδια RNA περιέχουν τις αζωτούχες βάσεις αδενίνη, κυτοσίνη και γουανίνη. Ωστόσο, δεν περιέχουν θυμίνη, η οποία αντικαθίσταται από ουρακίλη, που συμβολίζεται με ένα "U". Το RNA υπάρχει ως μονόκλωνο μόριο παρά ως δίκλωνη έλικα. Οι μοριακοί βιολόγοι έχουν ονομάσει διάφορα είδη RNA με βάση τη λειτουργία τους. Αυτά περιλαμβάνουν αγγελιαφόρο RNA (mRNA), RNA μεταφοράς (tRNA) και ριβοσωμικό RNA (rRNA)—μόρια που εμπλέκονται στην παραγωγή πρωτεϊνών από τον κώδικα DNA.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

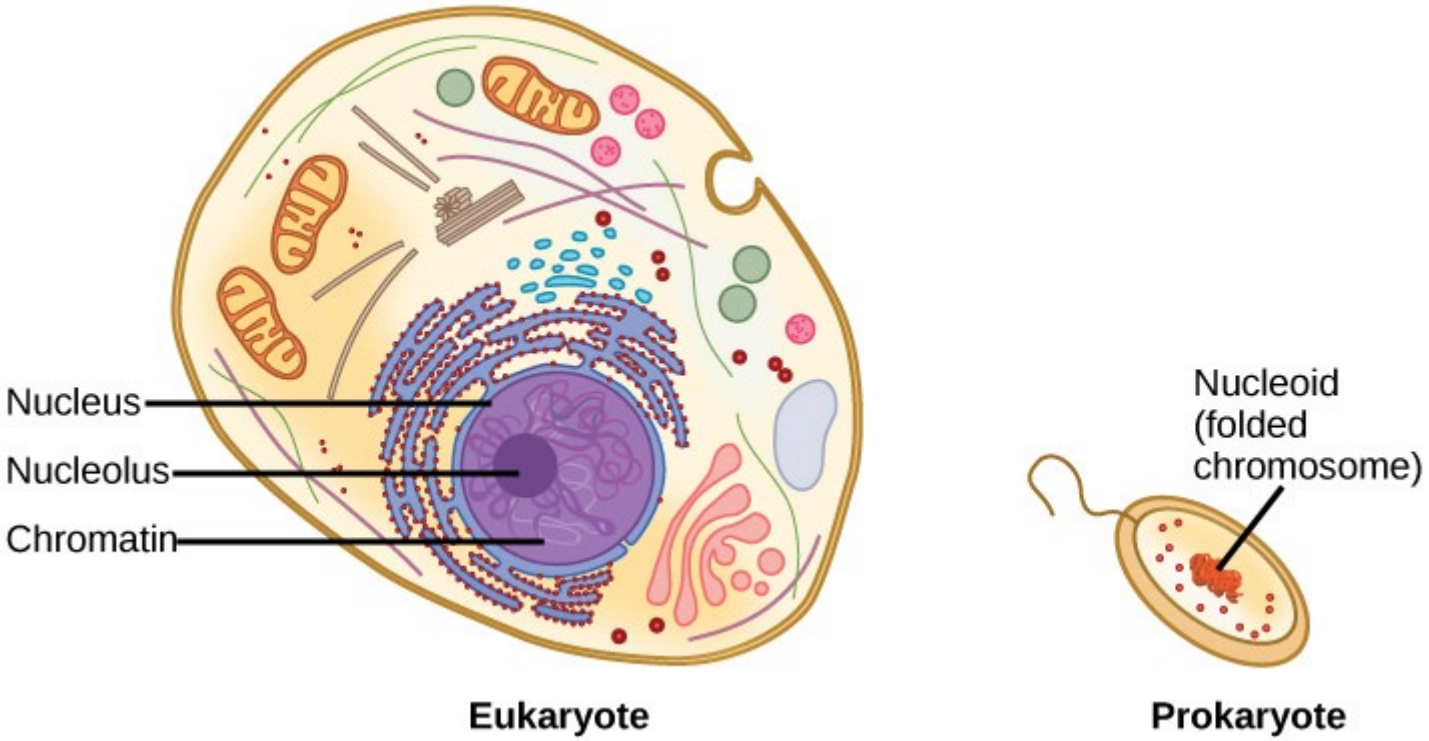
Πώς το DNA είναι διατεταγμένο στο κύτταρο

Το DNA είναι ένα λειτουργικό μόριο. πρέπει να αναπαραχθεί όταν ένα κύτταρο είναι έτοιμο να διαιρεθεί και πρέπει να «διαβαστεί» για να παραχθούν τα μόρια, όπως οι πρωτεΐνες, για να εκτελεστούν οι λειτουργίες του κυττάρου. Για το λόγο αυτό, το DNA προστατεύεται και συσκευάζεται με πολύ συγκεκριμένους τρόπους. Επιπλέον, τα μόρια DNA μπορεί να είναι πολύ μακριά. Τεντωμένα από άκρη σε άκρη, τα μόρια DNA σε ένα μόνο ανθρώπινο κύτταρο θα έφταναν σε μήκος περίπου 2 μέτρων. Έτσι, το DNA για ένα κύτταρο πρέπει να συσκευάζεται με πολύ τακτοποιημένο τρόπο ώστε να ταιριάζει και να λειτουργεί μέσα σε μια δομή (το κύτταρο) που δεν είναι ορατή με γυμνό μάτι. Τα χρωμοσώματα των προκαρυωτικών είναι πολύ πιο απλά από αυτά των ευκαρυωτικών σε πολλά από τα χαρακτηριστικά τους. Τα περισσότερα προκαρυωτικά περιέχουν ένα μόνο, κυκλικό χρωμόσωμα που βρίσκεται σε μια περιοχή στο κυτταρόπλασμα που ονομάζεται νουκλεοειδές.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

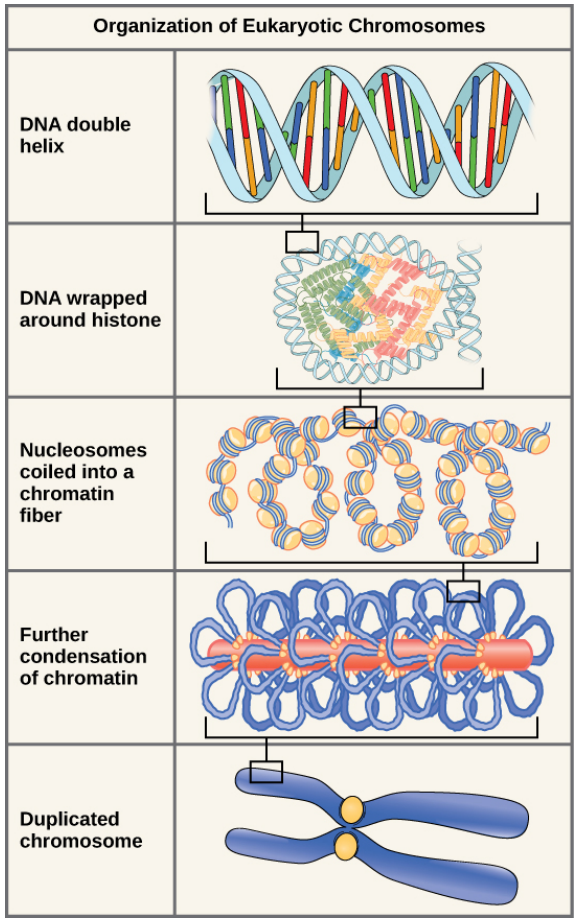
Πώς το DNA είναι διατεταγμένο στο κύτταρο



Ένας ευκαρυώτης περιέχει έναν καλά καθορισμένο πυρήνα, ενώ στους προκαρυώτες, το χρωμόσωμα βρίσκεται στο κυτταρόπλασμα σε μια περιοχή που ονομάζεται νουκλεοειδές.

Co-funded by the Erasmus+ Programme of the European Union

Το μέγεθος του γονιδιώματος σε ένα από τα πιο καλά μελετημένα προκαρυωτικά, το Escherichia coli, είναι 4,6 εκατομμύρια ζεύγη βάσεων, τα οποία θα επεκτείνονταν σε απόσταση περίπου 1,6 mm εάν τεντώνονταν. Πώς ταιριάζει λοιπόν αυτό μέσα σε ένα μικρό βακτηριακό κύτταρο; Το DNA είναι στριμμένο πέρα από τη διπλή έλικα σε αυτό που είναι γνωστό ως υπερέλιξη. Μερικές πρωτεΐνες είναι γνωστό ότι εμπλέκονται στην υπερέλιξη. Άλλες πρωτεΐνες και ένζυμα βοηθούν στη διατήρηση της υπερτυλιγμένης δομής.



Οι ευκαρυώτες, των οποίων τα χρωμοσώματα αποτελούνται από ένα γραμμικό μόριο DNA, χρησιμοποιούν έναν διαφορετικό τύπο στρατηγικής συσκευασίας για να χωρέσουν το DNA τους μέσα στον πυρήνα. Στο πιο βασικό επίπεδο, το DNA τυλίγεται γύρω από πρωτεΐνες γνωστές ως ιστόνες για να σχηματίσει δομές που ονομάζονται νουκλεοσώματα. Το DNA είναι τυλιγμένο σφιχτά γύρω από τον πυρήνα ιστόνης. Αυτό το νουκλεόσωμα συνδέεται με το επόμενο με μια μικρή αλυσίδα DNA που είναι απαλλαγμένη από ιστόνες. Αυτό είναι επίσης γνωστό ως δομή "χάντρες σε μια χορδή". τα νουκλεοσώματα είναι τα «σφαιρίδια» και τα μικρά μήκη του DNA μεταξύ τους είναι η «χορδή». Τα νουκλεοσώματα, με το DNA τους τυλιγμένο γύρω τους, στοιβάζονται συμπαγώς το ένα πάνω στο άλλο για να σχηματίσουν μια ίνα πλάτους 30 nm. Αυτή η ίνα τυλίγεται περαιτέρω σε μια παχύτερη και πιο συμπαγή δομή. Στο στάδιο της μετάφασης της μίτωσης, όταν τα χρωμοσώματα είναι ευθυγραμμισμένα στο κέντρο του κυττάρου, τα χρωμοσώματα είναι στο πιο συμπαγές τους. Έχουν πλάτος περίπου 700 nm και βρίσκονται σε συνδυασμό με πρωτεΐνες ικριώματος.

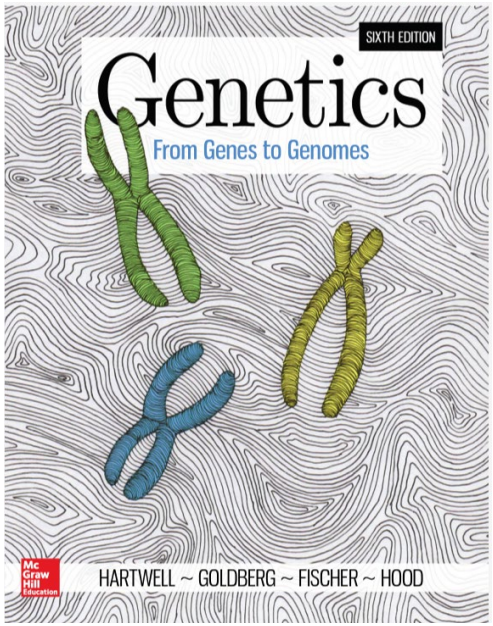
Αυτά τα σχήματα απεικονίζουν τη συμπίεση του ευκαρυωτικού χρωμοσώματος.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε τη δομή του DNA.
- Πώς οργανώνεται το γενετικό υλικό στον πυρήνα;

Co-funded by the Erasmus+ Programme of the European Union



Leland Hartwell , Michael Goldberg, Janice Fischer, Lee Hood, Charles F. Aquadro , Bruce Bejcek . (2015) Γενετική: από τα γονίδια στα γονιδιώματα. McGraw-Hill Education, Νέα Υόρκη, Η.Π.Α

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 2. Γενετική

Θέμα 1. Μοριακή Γενετική

Μάθημα 3. Μια επισκόπηση της αντιγραφής του DNA

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Η βασική ιδέα
- DNA πολυμεράση
- Αντιγραφή DNA σε ευκαρυώτες

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Η αντιγραφή του DNA, ή η αντιγραφή του DNA ενός κυττάρου, δεν είναι απλή υπόθεση! Υπάρχουν περίπου 3 δισεκατομμύρια ζεύγη βάσεων DNA στο γονιδίωμά σας, τα οποία πρέπει να αντιγραφούν με ακρίβεια όταν ένα από τα τρισεκατομμύρια κύτταρά σας διαιρείται.

Οι βασικοί μηχανισμοί αντιγραφής του DNA είναι παρόμοιοι μεταξύ των οργανισμών. Σε αυτό το άρθρο, θα επικεντρωθούμε στην αντιγραφή του DNA καθώς λαμβάνει χώρα στο βακτήριο E. coli, αλλά οι μηχανισμοί αντιγραφής είναι παρόμοιοι στους ανθρώπους και σε άλλους ευκαρυώτες.

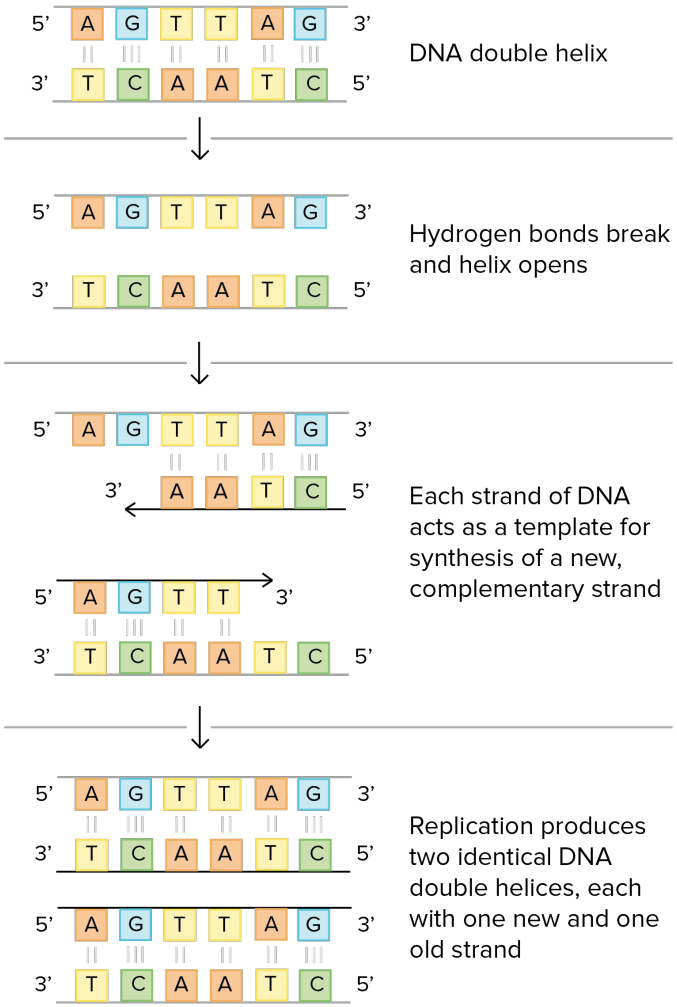
Ας ρίξουμε μια ματιά στις πρωτεΐνες και τα ένζυμα που πραγματοποιούν αντιγραφή, βλέποντας πώς συνεργάζονται για να εξασφαλίσουν ακριβή και πλήρη αντιγραφή του DNA.

Η βασική ιδέα

Η αντιγραφή του DNA είναι ημισυντηρητική, που σημαίνει ότι κάθε κλώνος στη διπλή έλικα του DNA λειτουργεί ως πρότυπο για τη σύνθεση ενός νέου, συμπληρωματικού κλώνου.

Αυτή η διαδικασία μας οδηγεί από ένα αρχικό μόριο σε δύο «παιδιά» μόρια, με κάθε νεοσχηματισμένη διπλή έλικα να περιέχει ένα νέο και ένα παλιό σκέλος.

Co-funded by the Erasmus+ Programme of the European Union

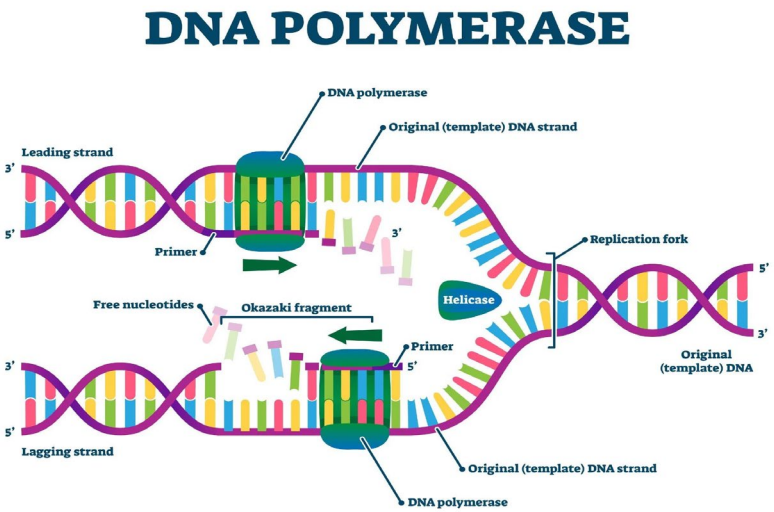


Κατά μία έννοια, αυτό είναι το μόνο που υπάρχει για την αντιγραφή του DNA! Αλλά αυτό που είναι στην πραγματικότητα πιο ενδιαφέρον για αυτή τη διαδικασία είναι πώς εκτελείται σε ένα κύτταρο.

Τα κύτταρα πρέπει να αντιγράψουν το DNA τους πολύ γρήγορα και με πολύ λίγα λάθη (ή να αντιμετωπίσουν κινδύνους, όπως ο καρκίνος). Για να το κάνουν αυτό, χρησιμοποιούν μια ποικιλία ενζύμων και πρωτεϊνών, τα οποία συνεργάζονται για να διασφαλίσουν ότι η αντιγραφή του DNA εκτελείται ομαλά και με ακρίβεια.

DNA πολυμεράση

Ένα από τα βασικά μόρια στην αντιγραφή του DNA είναι το ένζυμο DNA πολυμεράση. Οι πολυμεράσες DNA είναι υπεύθυνες για τη σύνθεση του DNA: προσθέτουν νουκλεοτίδια ένα προς ένα στην αναπτυσσόμενη αλυσίδα DNA, ενσωματώνοντας μόνο εκείνα που είναι συμπληρωματικά με το εκμαγείο.



Co-funded by the Erasmus+ Programme of the European Union

DNA πολυμεράση

Ακολουθούν ορισμένα βασικά χαρακτηριστικά των πολυμερασών DNA:

- Χρειάζονται πάντα ένα πρότυπο
- Μπορούν να προσθέσουν νουκλεοτίδια μόνο στο 3' άκρο ενός κλώνου DNA
- Δεν μπορούν να ξεκινήσουν να φτιάχνουν μια αλυσίδα DNA από το μηδέν, αλλά απαιτούν μια προϋπάρχουσα αλυσίδα ή ένα μικρό τμήμα νουκλεοτιδίων που ονομάζεται εκκινητής
- Διορθώνουν ή ελέγχουν τη δουλειά τους, αφαιρώντας τη συντριπτική πλειοψηφία των «λάθος» νουκλεοτιδίων που προστίθενται κατά λάθος στην αλυσίδα

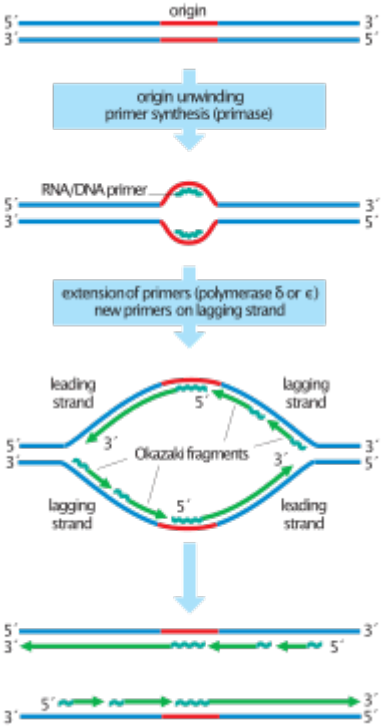
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Έναρξη αντιγραφής DNA

Η αντιγραφή ξεκινά πάντα σε συγκεκριμένες θέσεις στο DNA, οι οποίες ονομάζονται αφετηρίες αντιγραφής και αναγνωρίζονται από την αλληλουχία τους.

Οι εξειδικευμένες πρωτεΐνες αναγνωρίζουν την προέλευση, δεσμεύονται σε αυτή τη θέση και ανοίγουν το DNA. Καθώς το DNA ανοίγει, σχηματίζονται δύο δομές σε σχήμα Υ που ονομάζονται πιρούνια αντιγραφής, οι οποίες μαζί αποτελούν αυτό που ονομάζεται φυσαλίδα αντιγραφής. Οι διχάλες αναπαραγωγής θα κινούνται προς αντίθετες κατευθύνσεις καθώς προχωρά η αναπαραγωγή.



Co-funded by the Erasmus+ Programme of the European Union

Αστάρια και primase

Οι πολυμεράσες DNA μπορούν να προσθέσουν νουκλεοτίδια μόνο στο 3' άκρο ενός υπάρχοντος κλώνου DNA. (Χρησιμοποιούν την ελεύθερη ομάδα -OH που βρίσκεται στο 3' άκρο ως «άγκιστρο», προσθέτοντας ένα νουκλεοτίδιο σε αυτήν την ομάδα στην αντίδραση πολυμερισμού.) Πώς, λοιπόν, η DNA πολυμεράση προσθέτει το πρώτο νουκλεοτίδιο σε μια νέα διχάλα αντιγραφής;

Μόνο του, δεν γίνεται! Το πρόβλημα λύνεται με τη βοήθεια ενός ενζύμου που ονομάζεται πριμάση. Η πριμάση κάνει έναν εκκινητή RNA, ή ένα σύντομο τμήμα νουκλεϊκού οξέος συμπληρωματικό προς το εκμαγείο, το οποίο παρέχει ένα άκρο 3' για τη δράση της DNA πολυμεράσης. Ένας τυπικός εκκινητής έχει μήκος περίπου πέντε έως δέκα νουκλεοτίδια. Ο εκκινητής ξεκινά τη σύνθεση DNA.

Μόλις τοποθετηθεί ο εκκινητής RNA, η πολυμεράση DNA τον "επεκτείνει", προσθέτοντας νουκλεοτίδια ένα προς ένα για να δημιουργήσει έναν νέο κλώνο DNA που είναι συμπληρωματικός του κλώνου του εκμαγείου.

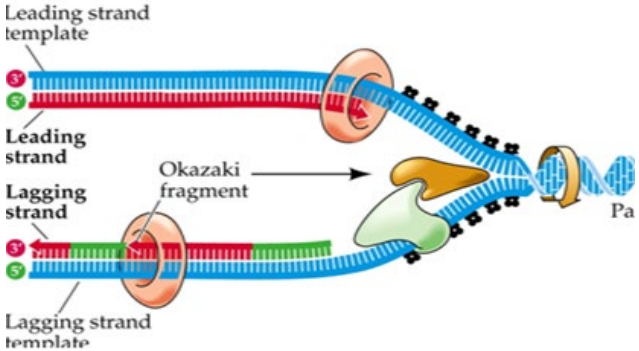
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Προπορευόμενα και υστερούντα σκέλη

Οι πολυμεράσες DNA μπορούν να παράγουν DNA μόνο στην κατεύθυνση 5' προς 3' και αυτό δημιουργεί πρόβλημα κατά την αντιγραφή. Μια διπλή έλικα DNA είναι πάντα αντιπαράλληλη. Με άλλα λόγια, το ένα σκέλος τρέχει προς την κατεύθυνση 5' προς 3', ενώ το άλλο προς την κατεύθυνση 3' προς 5'. Αυτό καθιστά απαραίτητο τα δύο νέα σκέλη, τα οποία είναι επίσης αντιπαράλληλα με τα πρότυπά τους, να κατασκευαστούν με ελαφρώς διαφορετικούς τρόπους.

Ένα νέο σκέλος, που εκτείνεται 5' έως 3' προς το πιρούνι αναπαραγωγής, είναι το εύκολο. Αυτός ο κλώνος κατασκευάζεται συνεχώς, επειδή η DNA πολυμεράση κινείται προς την ίδια κατεύθυνση με τη διχάλα αντιγραφής. Αυτός ο συνεχώς συντιθέμενος κλώνος ονομάζεται οδηγός κλώνος.



Προπορευόμενα και υστερούντα σκέλη

Το άλλο νέο σκέλος, το οποίο απέχει 5' έως 3' από το πιρούνι, είναι πιο δύσκολο. Αυτός ο κλώνος κατασκευάζεται σε θραύσματα επειδή, καθώς το πιρούνι κινείται προς τα εμπρός, η πολυμεράση DNA (η οποία απομακρύνεται από το πιρούνι) πρέπει να αποκολληθεί και να προσκολληθεί ξανά στο DNA που μόλις εκτέθηκε. Αυτό το δύσκολο σκέλος, το οποίο κατασκευάζεται σε θραύσματα, ονομάζεται υστερούντα σκέλη.

Τα μικρά θραύσματα ονομάζονται θραύσματα Okazaki, που ονομάστηκαν από τον Ιάπωνα επιστήμονα που τα ανακάλυψε. Ο οδηγός κλώνος μπορεί να επεκταθεί μόνο από έναν εκκινητή, ενώ ο υστερούμενος κλώνος χρειάζεται ένα νέο εκκινητή για κάθε ένα από τα μικρά θραύσματα Okazaki.

Το συνεργείο συντήρησης και καθαρισμού

Κάποιες άλλες πρωτεΐνες και ένζυμα, επιπλέον των κυριότερων παραπάνω, χρειάζονται για να διατηρηθεί ομαλά η αντιγραφή του DNA. Το ένα είναι μια πρωτεΐνη που ονομάζεται συρόμενος σφιγκτήρας, ο οποίος συγκρατεί τα μόρια της DNA πολυμεράσης III στη θέση τους καθώς συνθέτουν το DNA. Ο συρόμενος σφιγκτήρας είναι μια πρωτεΐνη σε σχήμα δακτυλίου και κρατά την πολυμεράση DNA του υστερούντος κλώνου να μην επιπλέει όταν ξαναρχίζει σε ένα νέο θραύσμα Okazaki.

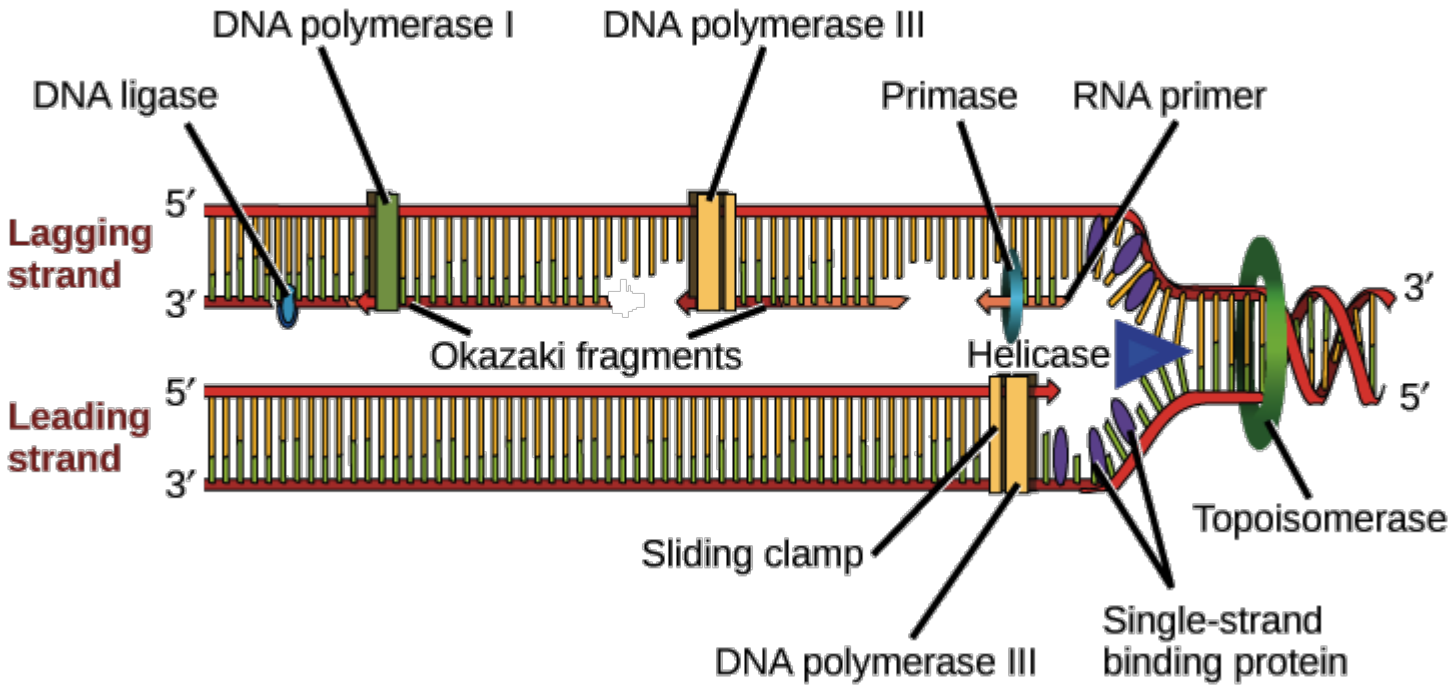
Co-funded by the Erasmus+ Programme of the European Union

Το συνεργείο συντήρησης και καθαρισμού

Η τοποϊσομεράση παίζει επίσης σημαντικό ρόλο συντήρησης κατά την αντιγραφή του DNA. Αυτό το ένζυμο εμποδίζει τη διπλή έλικα του DNA μπροστά από το πιρούνι αντιγραφής να τυλιχτεί πολύ σφιχτά καθώς το DNA ανοίγει. Ενεργεί κάνοντας προσωρινές εγκοπές στην έλικα για να απελευθερώσει την τάση, στη συνέχεια σφραγίζοντας τις εγκοπές για να αποφευχθεί η μόνιμη ζημιά.

Τέλος, υπάρχει μια μικρή εργασία καθαρισμού που πρέπει να κάνουμε εάν θέλουμε DNA που δεν περιέχει RNA ή κενά. Οι εκκινητές RNA αφαιρούνται και αντικαθίστανται από DNA μέσω της δραστικότητας της DNA πολυμεράσης I, της άλλης πολυμεράσης που εμπλέκεται στην αντιγραφή. Οι εγκοπές που παραμένουν μετά την αντικατάσταση των εκκινητών σφραγίζονται από το ένζυμο DNA λιγάση.

Co-funded by the Erasmus+ Programme of the European Union



Co-funded by the Erasmus+ Programme of the European Union

- Η ελικάση ανοίγει το DNA στη διχάλα αντιγραφής.
- Οι μονόκλωνες πρωτεΐνες δέσμευσης επικαλύπτουν το DNA γύρω από τη διχάλα αντιγραφής για να αποτρέψουν την επανατύλιξη του DNA.
- Η τοποϊσομεράση λειτουργεί στην περιοχή μπροστά από τη διχάλα αντιγραφής για να αποτρέψει την υπερέλιξη.
- Το Primase συνθέτει εκκινητές RNA συμπληρωματικά προς τον κλώνο DNA.
- Η DNA πολυμεράση III επεκτείνει τους εκκινητές, προσθέτοντας στο 3' άκρο, για να δημιουργήσει το μεγαλύτερο μέρος του νέου DNA.
- Οι εκκινητές RNA αφαιρούνται και αντικαθίστανται με DNA από DNA πολυμεράση I.
- Τα κενά μεταξύ των θραυσμάτων DNA σφραγίζονται με DNA λιγάση.

Co-funded by the Erasmus+ Programme of the European Union

Αντιγραφή DNA σε ευκαρυώτες

Τα βασικά στοιχεία της αντιγραφής του DNA είναι παρόμοια μεταξύ των βακτηρίων και των ευκαρυωτών όπως ο άνθρωπος, αλλά υπάρχουν επίσης ορισμένες διαφορές:

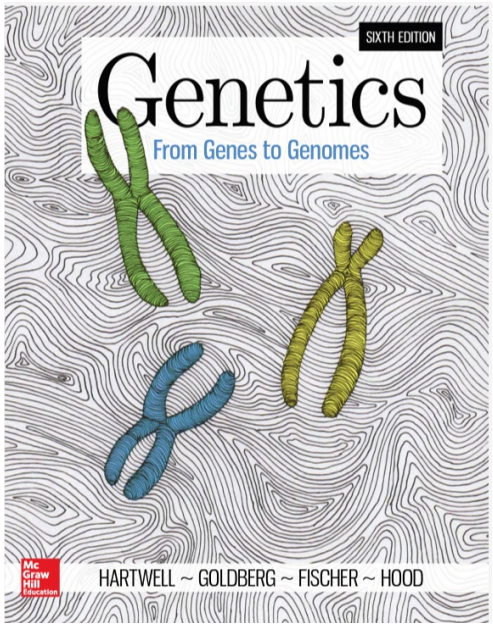
- Οι ευκαρυώτες έχουν συνήθως πολλαπλά γραμμικά χρωμοσώματα, το καθένα με πολλαπλές απαρχές αντιγραφής. Οι άνθρωποι μπορούν να έχουν έως και 100.000 απαρχές αντιγραφής
- Τα περισσότερα από τα ένζυμα E. coli έχουν αντίστοιχα στην αντιγραφή του ευκαρυωτικού DNA, αλλά ένα μόνο ένζυμο στο E. coli μπορεί να αντιπροσωπεύεται από πολλαπλά ένζυμα στους ευκαρυώτες. Για παράδειγμα, υπάρχουν πέντε ανθρώπινες πολυμεράσες DNA με σημαντικούς ρόλους στην αντιγραφή
- Τα περισσότερα ευκαρυωτικά χρωμοσώματα είναι γραμμικά. Εξαιτίας του τρόπου με τον οποίο κατασκευάζεται η καθυστερημένη αλυσίδα, χάνεται μέρος του DNA από τα άκρα των γραμμικών χρωμοσωμάτων (τα τελομερή) σε κάθε γύρο αντιγραφής

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε τα κύρια στάδια αντιγραφής του DNA.
- Καταγράψτε τα βασικά μόρια (ένζυμα) στη διαδικασία αντιγραφής του DNA και εξηγήστε το ρόλο τους.

Co-funded by the Erasmus+ Programme of the European Union



Leland Hartwell , Michael Goldberg, Janice Fischer, Lee Hood, Charles F. Aquadro , Bruce Bejcek . (2015) Γενετική: από τα γονίδια στα γονιδιώματα. McGraw-Hill Education, Νέα Υόρκη, Η.Π.Α

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 2. Γενετική
Θέμα 1. Μοριακή Γενετική
Μάθημα 4. Γονίδια και Πρωτεΐνες

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Τα γονίδια κωδικοποιούν πρωτεΐνες
- Μετάλλαξη
- Κληρονομικές μεταλλάξεις
- ΦΥΣΙΚΗ ΕΠΙΛΟΓΗ

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Αν ο κώδικας θεωρείται ως μια σειρά από γράμματα σε ένα φύλλο χαρτιού, τότε τα γονίδια αποτελούνται από διακριτές λέξεις που σχηματίζουν προτάσεις και παραγράφους που δίνουν νόημα στο σχέδιο των γραμμάτων. Αυτό που δημιουργείται από τους πολύπλοκους και διαφορετικούς κώδικες DNA είναι η πρωτεΐνη, μια κατηγορία μακρομορίων που φέρει τις περισσότερες από τις δραστηριότητες στο κελί.

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Το περιεχόμενο πληροφοριών του DNA βρίσκεται στην αλληλουχία των βάσεων του. Οι τέσσερις βάσεις σε κάθε αλυσίδα είναι σαν τα γράμματα ενός αλφαβήτου. μπορεί να διαδέχονται το ένα το άλλο με οποιαδήποτε σειρά, και διαφορετικές ακολουθίες γράφουν διαφορετικές «λέξεις». Κάθε λέξη έχει τη δική της σημασία, δηλαδή τη δική της επίδραση στον φαινότυπο. Το AGTCAT, για παράδειγμα, σημαίνει ένα πράγμα, ενώ το CTAGGT σημαίνει ένα άλλο. Αν και το DNA έχει μόνο τέσσερα διαφορετικά γράμματα, ή δομικά στοιχεία, η δυνατότητα για διαφορετικούς συνδυασμούς και επομένως διαφορετικά σύνολα πληροφοριών σε μια μακρά αλυσίδα νουκλεοτιδίων είναι εκπληκτική. Μερικά ανθρώπινα χρωμοσώματα, για παράδειγμα, αποτελούνται από αλυσίδες μήκους 250 εκατομμυρίων νουκλεοτιδίων. Επειδή οι διαφορετικές βάσεις μπορεί να διαδέχονται η μία την άλλη με οποιαδήποτε σειρά, τέτοιες αλυσίδες θα μπορούσαν να περιέχουν οποιαδήποτε από $4^{250.000.000}$ (που μεταφράζεται σε 1 ακολουθούμενο από 150.515.000 μηδενικά) πιθανές αλληλουχίες νουκλεοτιδίων.

Τα γονίδια κωδικοποιούν πρωτεΐνες

Από την εκ νέου ανακάλυψη του έργου του Mendel το 1900, ο ορισμός του γονιδίου έχει εξελιχθεί από μια αφηρημένη μονάδα κληρονομικότητας σε μια απτή μοριακή οντότητα ικανή για αντιγραφή, μεταγραφή, μετάφραση και μετάλλαξη. Τα γονίδια αποτελούνται από DNA και είναι γραμμικά διατεταγμένα στα χρωμοσώματα. Ορισμένα γονίδια κωδικοποιούν δομικά και ρυθμιστικά RNA. Υπάρχουν αυξανόμενα στοιχεία από την έρευνα που καταγράφει το προφίλ του μεταγραφώματος των κυττάρων (το πλήρες σύνολο όλων των μεταγραφών RNA που υπάρχουν σε ένα κύτταρο) ότι αυτές μπορεί να είναι οι μεγαλύτερες κατηγορίες RNA που παράγονται από ευκαρυωτικά κύτταρα, υπερβαίνοντας κατά πολύ τα αγγελιαφόρα RNA που κωδικοποιούν πρωτεΐνες (mRNA). αλλά τα 20.000 γονίδια που κωδικοποιούν πρωτεΐνες που βρίσκονται συνήθως σε ζωικά κύτταρα και τα 30.000 γονίδια που κωδικοποιούν πρωτεΐνες που βρίσκονται συνήθως στα φυτικά κύτταρα, έχουν ωστόσο τεράστιες επιπτώσεις στην κυτταρική λειτουργία.

Τα γονίδια που κωδικοποιούν πρωτεΐνες προσδιορίζουν τις αλληλουχίες των αμινοξέων, τα οποία είναι τα δομικά στοιχεία των πρωτεϊνών. Με τη σειρά τους, οι πρωτεΐνες είναι υπεύθυνες για την ενορχήστρωση σχεδόν κάθε λειτουργίας του κυττάρου. Τόσο τα γονίδια που κωδικοποιούν πρωτεΐνες όσο και οι πρωτεΐνες που είναι τα γονιδιακά τους προϊόντα είναι απολύτως απαραίτητα για τη ζωή όπως τη γνωρίζουμε.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Τα γονίδια κωδικοποιούν πρωτεΐνες

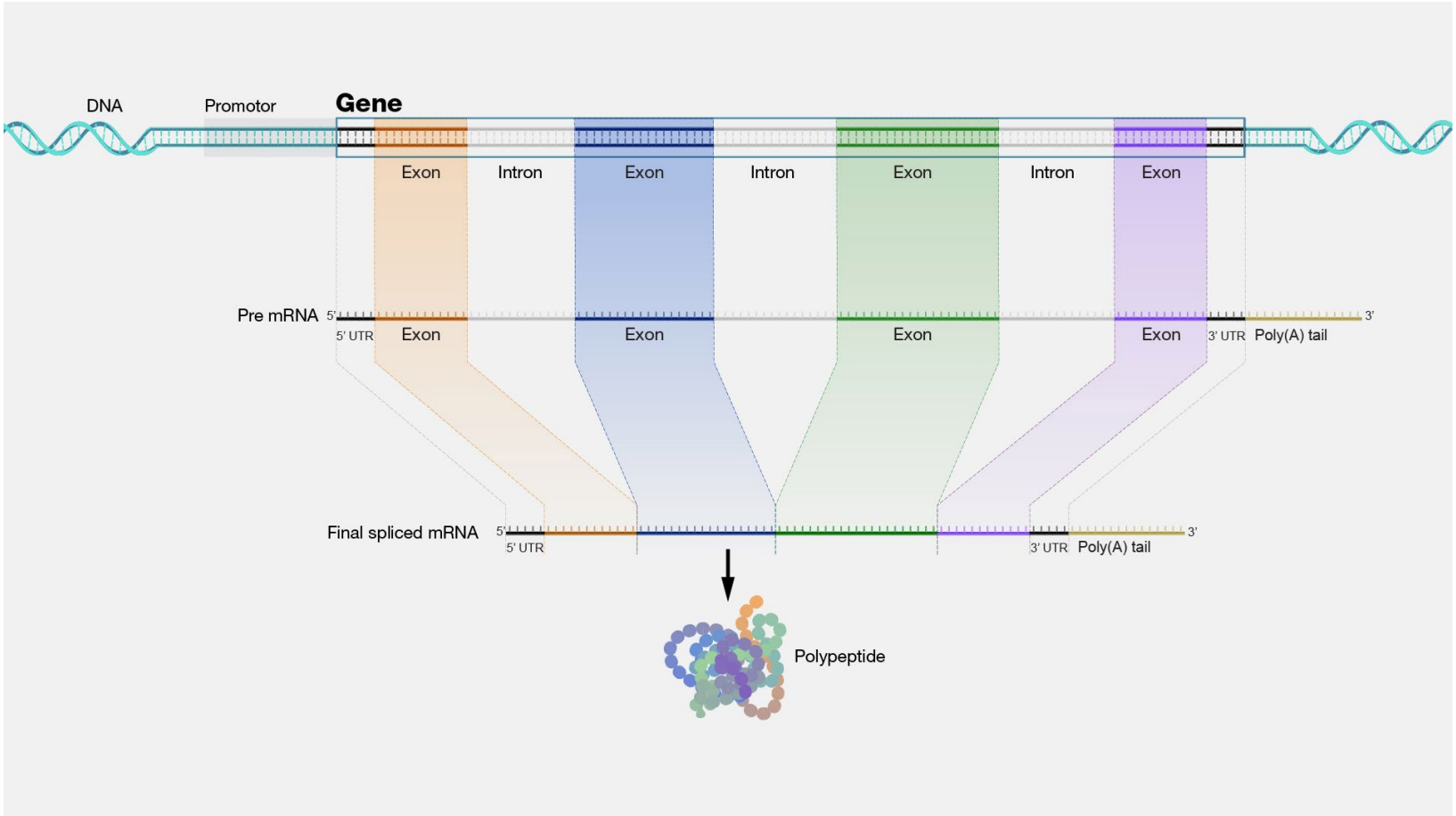
Η αντιγραφή, η μεταγραφή και η μετάφραση είναι οι τρεις κύριες διαδικασίες που χρησιμοποιούνται από όλα τα κύτταρα για να διατηρήσουν τις γενετικές τους πληροφορίες και να μετατρέψουν τις γενετικές πληροφορίες που κωδικοποιούνται στο DNA σε γονιδιακά προϊόντα, τα οποία είναι είτε RNA είτε πρωτεΐνες, ανάλογα με το γονίδιο. Στα ευκαρυωτικά κύτταρα, ή εκείνα τα κύτταρα που έχουν πυρήνα, η αντιγραφή και η μεταγραφή λαμβάνουν χώρα εντός του πυρήνα ενώ η μετάφραση λαμβάνει χώρα έξω από τον πυρήνα στο κυτταρόπλασμα. Στα προκαρυωτικά κύτταρα, ή σε εκείνα τα κύτταρα που δεν έχουν πυρήνα, και οι τρεις διεργασίες συμβαίνουν στο κυτταρόπλασμα.

Η αντιγραφή είναι η βάση της βιολογικής κληρονομικότητας. Αντιγράφει το DNA ενός κυττάρου. Το ένζυμο πολυμεράση DNA αντιγράφει ένα μονογονικό μόριο δίκλωνου DNA σε δύο θυγατρικά δίκλωνα μόρια DNA. Η μεταγραφή δημιουργεί RNA από DNA. Το ένζυμο RNA πολυμεράση δημιουργεί ένα μόριο RNA που είναι συμπληρωματικό σε ένα τμήμα DNA που κωδικοποιεί γονίδιο. Η μετάφραση δημιουργεί πρωτεΐνη από mRNA. Το ριβόσωμα δημιουργεί μια πολυπεπτιδική αλυσίδα αμινοξέων χρησιμοποιώντας mRNA ως εκμαγείο. Η πολυπεπτιδική αλυσίδα διπλώνει για να γίνει πρωτεΐνη.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

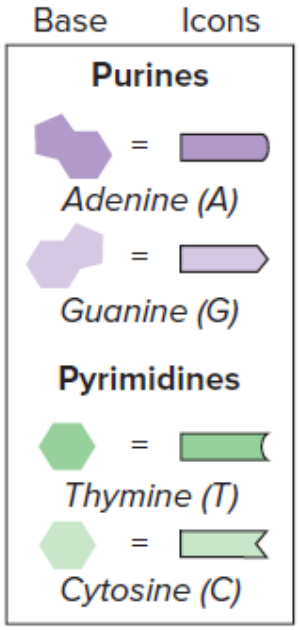
Τα γονίδια κωδικοποιούν πρωτεΐνες



Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Οι περισσότερες γενετικές πληροφορίες διαβάζονται από μη τυλιγμένες αλυσίδες DNA

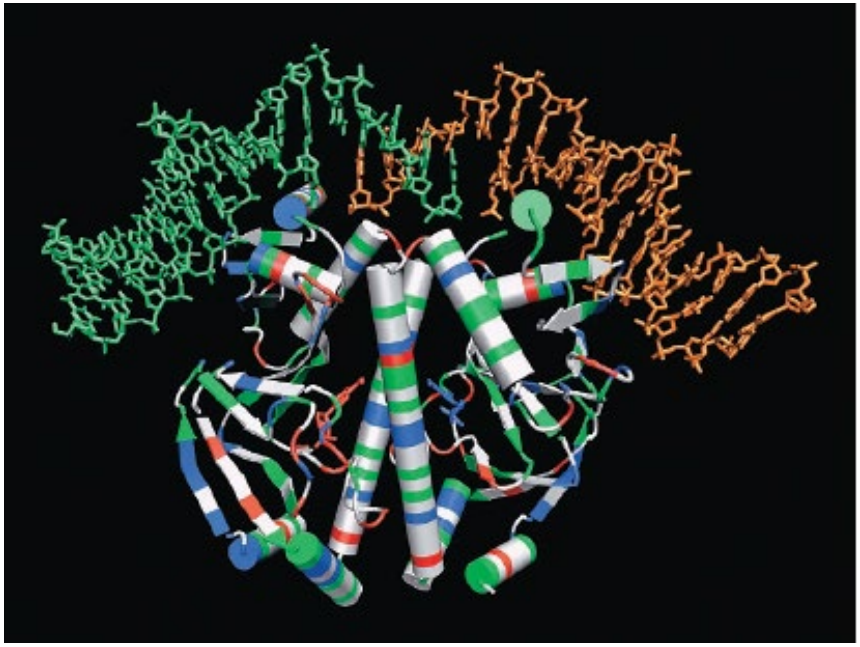


Το ξετύλιγμα ενός μορίου DNA εκθέτει μια ενιαία ακολουθία βάσεων σε καθέναν από τους δύο κλώνους. Οι πρωτεΐνες διαβάζουν τις πληροφορίες σε έναν μόνο κλώνο DNA συνθέτοντας ένα τμήμα RNA (μια διαδικασία που ονομάζεται μεταγραφή) ή DNA (μια διαδικασία που ονομάζεται αντιγραφή) συμπληρωματικά σε μια συγκεκριμένη αλληλουχία.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ορισμένες γενετικές πληροφορίες είναι προσβάσιμες χωρίς να ξετυλίγεται το DNA



Ορισμένες πρωτεΐνες μπορούν να αναγνωρίσουν και να δεσμευτούν σε συγκεκριμένες αλληλουχίες ζευγών βάσεων σε δίκλωνο DNA. Οι πληροφορίες αυτές προκύπτουν κυρίως από διαφορές μεταξύ των τεσσάρων βάσεων που εμφανίζονται στις μείζονες και τις ελάσσονες αυλακώσεις. Εντός των αυλακώσεων, ορισμένα άτομα στην περιφέρεια των βάσεων είναι εκτεθειμένα, και ιδιαίτερα στην κύρια αυλάκωση, αυτά τα άτομα μπορεί να έχουν χωρικά μοτίβα που παρέχουν χημικές πληροφορίες. Οι πρωτεΐνες μπορούν να έχουν πρόσβαση σε αυτές τις πληροφορίες για να ανιχνεύσουν την αλληλουχία βάσης σε μια έκταση DNA χωρίς να αποσυναρμολογήσουν τη διπλή έλικα. Οι ειδικές για την αλληλουχία πρωτεΐνες δέσμευσης DNA περιλαμβάνουν μεταγραφικούς παράγοντες που ενεργοποιούν και απενεργοποιούν τα γονίδια καθώς και βακτηριακά περιοριστικά ένζυμα που κόβουν το DNA σε συγκεκριμένες θέσεις.

Co-funded by the Erasmus+ Programme of the European Union

Σε ορισμένους ιούς, το RNA είναι η αποθήκη γενετικών πληροφοριών

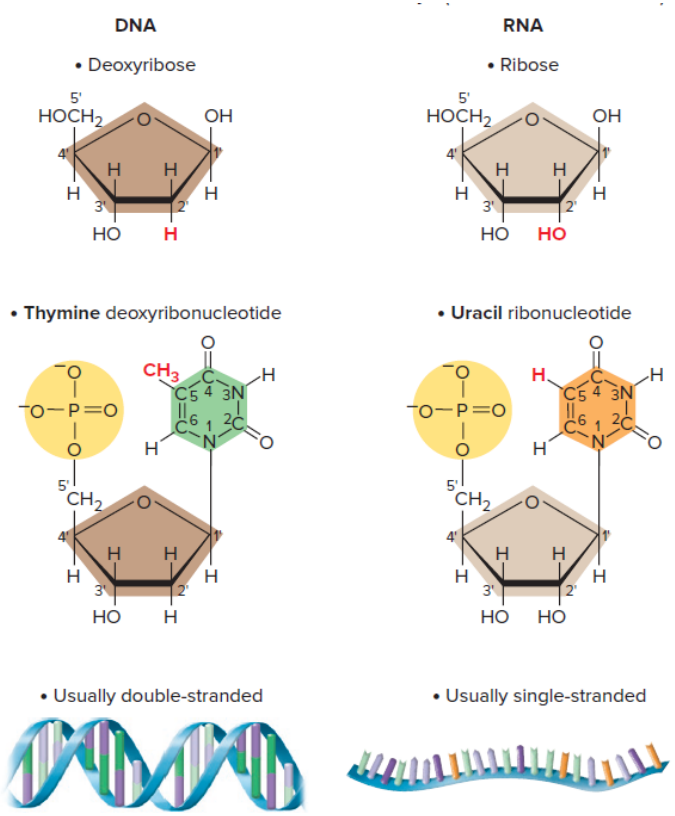
Το DNA μεταφέρει τη γενετική πληροφορία σε όλες τις κυτταρικές μορφές ζωής και σε πολλούς ιούς. Προκαρυώτες όπως τα βακτήρια Escherichia coli μεταφέρουν το DNA τους σε ένα δίκλωνο, ομοιοπολικά κλειστό κυκλικό χρωμόσωμα. Τα ευκαρυωτικά κύτταρα συσκευάζουν το DNA τους σε δίκλωνα γραμμικά χρωμοσώματα. Οι ιοί του DNA το μεταφέρουν σε μικρά μόρια που είναι μονόκλωνα ή δίκλωνα, κυκλικά ή γραμμικά.

Αντίθετα, οι ρετροϊοί, που περιλαμβάνουν αυτούς που προκαλούν πολιομυελίτιδα και AIDS, χρησιμοποιούν ριβονουκλεϊκό οξύ ή RNA ως γενετικό τους υλικό.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

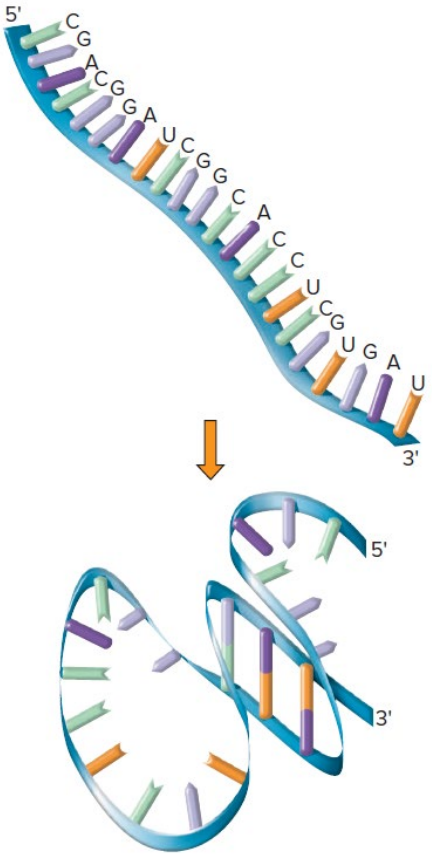
Διαφορές στη χημική δομή του DNA και του RNA



Τρεις σημαντικές χημικές διαφορές διακρίνουν το RNA από το DNA. Πρώτον, το RNA παίρνει το όνομά του από τη ζάχαρη ριβόζη, την οποία ενσωματώνει αντί της δεοξυριβόζης που βρίσκεται στο DNA. Δεύτερον, το RNA περιέχει τη βάση ουρακίλη (U) αντί για τη βάση θυμίνη (T). Το U, όπως και το T, ζεύγη βάσεων με το A. Τέλος, τα περισσότερα μόρια RNA είναι μονόκλωνα και περιέχουν πολύ λιγότερα νουκλεοτίδια από τα πολύ μακριά μόρια DNA που βρίσκονται στα πυρηνικά χρωμοσώματα.

Co-funded by the Erasmus+ Programme of the European Union

Διαφορές στη χημική δομή του DNA και του RNA



Μέσα σε ένα μονόκλωνο μόριο RNA, η αναδίπλωση μπορεί να φέρει σε επαφή δύο αντίθετα προσανατολισμένες περιοχές που φέρουν συμπληρωματικές αλληλουχίες νουκλεοτιδίων για να σχηματίσουν μια μικρή, ζευγών βάσεων τέντωμα εντός του μορίου. Αυτό σημαίνει ότι, σε σύγκριση με το σχετικά απλό, διπλό ελικοειδές σχήμα ενός μορίου DNA, πολλά RNA έχουν μια περίπλοκη δομή μικρών δίκλωνων τμημάτων διάσπαρτα με μονόκλωνους βρόχους. Το RNA έχει την ίδια ικανότητα με το DNA να μεταφέρει πληροφορίες στην αλληλουχία των βάσεων του, αλλά είναι πολύ λιγότερο σταθερό από το DNA. Εκτός από το ότι χρησιμεύει ως γενετικό υλικό για μια σειρά ιών, το RNA εκπληρώνει αρκετές ζωτικές λειτουργίες σε όλα τα κύτταρα. Για παράδειγμα, συμμετέχει στη γονιδιακή έκφραση και στη σύνθεση πρωτεϊνών, όπως παρουσιάζεται στο Κεφάλαιο 8. Το RNA παίζει επίσης έναν εκπληκτικά σημαντικό ρόλο στην αντιγραφή του DNA, τον οποίο τώρα περιγράφουμε.

Co-funded by the Erasmus+ Programme of the European Union

- Τα γονίδια περιέχονται στα χρωμοσώματα, τα οποία βρίσκονται στον πυρήνα του κυττάρου.
- Ένα χρωμόσωμα περιέχει εκατοντάδες έως χιλιάδες γονίδια.
- Κάθε φυσιολογικό ανθρώπινο κύτταρο περιέχει 23 ζεύγη χρωμοσωμάτων, για ένα σύνολο 46 χρωμοσωμάτων.
- Ένα χαρακτηριστικό είναι οποιοδήποτε χαρακτηριστικό που καθορίζεται από το γονίδιο και συχνά καθορίζεται από περισσότερα από ένα γονίδια.
- Ορισμένα χαρακτηριστικά προκαλούνται από μεταλλαγμένα γονίδια που κληρονομούνται ή που είναι το αποτέλεσμα μιας νέας γονιδιακής μετάλλαξης.

Co-funded by the Erasmus+ Programme of the European Union

Οι πρωτεΐνες είναι ίσως η πιο σημαντική κατηγορία υλικού στο σώμα. Οι πρωτεΐνες δεν είναι μόνο δομικά στοιχεία για τους μύες, τους συνδετικούς ιστούς, το δέρμα και άλλες δομές. Χρειάζονται επίσης για την παραγωγή ενζύμων. Τα ένζυμα είναι σύνθετες πρωτεΐνες που ελέγχουν και πραγματοποιούν σχεδόν όλες τις χημικές διεργασίες και αντιδράσεις μέσα στο σώμα. Το σώμα παράγει χιλιάδες διαφορετικά ένζυμα. Έτσι, ολόκληρη η δομή και η λειτουργία του σώματος διέπεται από τους τύπους και τις ποσότητες πρωτεϊνών που συνθέτει το σώμα. Η πρωτεϊνοσύνθεση ελέγχεται από γονίδια, τα οποία περιέχονται στα χρωμοσώματα.

Ο **γονότυπος** (ή γονιδίωμα) είναι ο μοναδικός συνδυασμός γονιδίων ή γενετικής σύνθεσης ενός ατόμου. Έτσι, ο γονότυπος είναι ένα πλήρες σύνολο οδηγιών σχετικά με το πώς το σώμα αυτού του ατόμου συνθέτει πρωτεΐνες και επομένως πώς υποτίθεται ότι αυτό το σώμα είναι χτισμένο και λειτουργεί.

Ο **φαινότυπος** είναι η πραγματική δομή και λειτουργία του σώματος ενός ατόμου. Ο φαινότυπος είναι ο τρόπος με τον οποίο εκδηλώνεται ο γονότυπος σε ένα άτομο - δεν μπορούν να εκτελεστούν (ή να εκφραστούν) όλες οι οδηγίες στον γονότυπο. Το αν και πώς εκφράζεται ένα γονίδιο καθορίζεται από μια πολύπλοκη αλληλεπίδραση πολλαπλών παραγόντων, συμπεριλαμβανομένου του γονότυπου, της γονιδιακής έκφρασης, των περιβαλλοντικών παραγόντων (συμπεριλαμβανομένων των ασθενειών και της διατροφής) και άλλων παραγόντων, ορισμένοι από τους οποίους είναι άγνωστοι.

Ο **καρυότυπος** είναι μια εικόνα του πλήρους συνόλου των χρωμοσωμάτων στα κύτταρα ενός ατόμου.

Μετάλλαξη

Για την αποφυγή λαθών κατά τη διάρκεια της αναπαραγωγής, τα κύτταρα έχουν μια λειτουργία "διόρθωσης" για να διασφαλιστεί ότι οι βάσεις έχουν αντιστοιχιστεί σωστά. Υπάρχουν επίσης χημικοί μηχανισμοί για την επιδιόρθωση του DNA που δεν αντιγράφηκε σωστά. Ωστόσο, λόγω των δισεκατομμυρίων ζευγών βάσεων που εμπλέκονται και της πολυπλοκότητας της διαδικασίας πρωτεϊνοσύνθεσης, μπορεί να συμβούν λάθη. Τέτοια λάθη μπορεί να συμβούν για πολλούς λόγους (συμπεριλαμβανομένης της έκθεσης σε ακτινοβολία, φάρμακα ή ιούς) ή χωρίς προφανή λόγο. Μικρές παραλλαγές στο DNA είναι πολύ συχνές και εμφανίζονται στους περισσότερους ανθρώπους. Οι περισσότερες παραλλαγές δεν επηρεάζουν τα επόμενα αντίγραφα του γονιδίου. Τα λάθη που επαναλαμβάνονται σε επόμενα αντίγραφα ονομάζονται μεταλλάξεις.

Κληρονομικές μεταλλάξεις

Κληρονομικές μεταλλάξεις είναι αυτές που μπορεί να μεταδοθούν στους απογόνους. Οι μεταλλάξεις μπορούν να κληρονομηθούν μόνο όταν επηρεάζουν τα αναπαραγωγικά κύτταρα (σπέρμα ή ωάριο). Οι μεταλλάξεις που δεν επηρεάζουν τα αναπαραγωγικά κύτταρα επηρεάζουν τους απογόνους του μεταλλαγμένου κυττάρου (για παράδειγμα, γίνονται καρκίνος) αλλά δεν μεταβιβάζονται στους απογόνους.

Οι μεταλλάξεις μπορεί να είναι μοναδικές για ένα άτομο ή οικογένεια και οι περισσότερες επιβλαβείς μεταλλάξεις είναι σπάνιες. Οι μεταλλάξεις που γίνονται τόσο συχνές που επηρεάζουν περισσότερο από το 1% ενός πληθυσμού ονομάζονται πολυμορφισμοί (για παράδειγμα, οι τύποι αίματος του ανθρώπου A, B, AB και O). Οι περισσότεροι πολυμορφισμοί έχουν μικρή ή καθόλου επίδραση στον φαινότυπο (την πραγματική δομή και λειτουργία του σώματος ενός ατόμου).

Οι μεταλλάξεις μπορεί να περιλαμβάνουν μικρά ή μεγάλα τμήματα DNA. Ανάλογα με το μέγεθος και τη θέση της, η μετάλλαξη μπορεί να μην έχει εμφανή επίδραση ή μπορεί να αλλάξει την αλληλουχία αμινοξέων σε μια πρωτεΐνη ή να μειώσει την ποσότητα της πρωτεΐνης που παράγεται. Εάν η πρωτεΐνη έχει διαφορετική αλληλουχία αμινοξέων, μπορεί να λειτουργήσει διαφορετικά ή και καθόλου. Μια πρωτεΐνη που απουσιάζει ή δεν λειτουργεί είναι συχνά επιβλαβής ή θανατηφόρα. Για παράδειγμα, στη φαινυλκετονουρία, μια μετάλλαξη έχει ως αποτέλεσμα την ανεπάρκεια ή την απουσία του ενζύμου υδροξυλάση της φαινυλαλανίνης. Αυτή η ανεπάρκεια επιτρέπει στο αμινοξύ φαινυλαλανίνη (που απορροφάται από τη διατροφή) να συσσωρεύεται στο σώμα, προκαλώντας τελικά σοβαρή διανοητική αναπηρία.

Σε σπάνιες περιπτώσεις, μια μετάλλαξη εισάγει μια πλεονεκτική αλλαγή. Για παράδειγμα, στην περίπτωση του δρεπανοκυτταρικού γονιδίου, όταν ένα άτομο κληρονομήσει δύο αντίγραφα του ανώμαλου γονιδίου, το άτομο θα αναπτύξει δρεπανοκυτταρική αναιμία. Ωστόσο, όταν ένα άτομο κληρονομεί μόνο ένα αντίγραφο του δρεπανοκυτταρικού γονιδίου (που ονομάζεται φορέας), το άτομο αναπτύσσει κάποια προστασία έναντι της ελονοσίας (λοίμωξη του αίματος). Αν και η προστασία από την ελονοσία μπορεί να βοηθήσει έναν φορέα να επιβιώσει, η δρεπανοκυτταρική αναιμία (σε ένα άτομο που έχει δύο αντίγραφα του γονιδίου) προκαλεί συμπτώματα και επιπλοκές που μπορεί να μειώσουν τη διάρκεια ζωής.

Co-funded by the Erasmus+ Programme of the European Union

ΦΥΣΙΚΗ ΕΠΙΛΟΓΗ

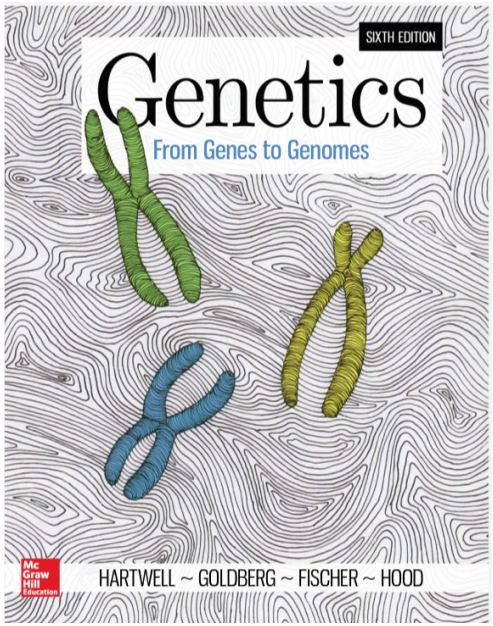
Η φυσική επιλογή αναφέρεται στην έννοια ότι οι μεταλλάξεις που επηρεάζουν την επιβίωση σε ένα δεδομένο περιβάλλον είναι λιγότερο πιθανό να μεταδοθούν στους απογόνους (και επομένως γίνονται λιγότερο συχνές στον πληθυσμό), ενώ οι μεταλλάξεις που βελτιώνουν την επιβίωση σταδιακά γίνονται πιο συχνές. Έτσι, οι ευεργετικές μεταλλάξεις, αν και αρχικά σπάνιες, τελικά γίνονται κοινές. Οι αργές αλλαγές που συμβαίνουν με την πάροδο του χρόνου που προκαλούνται από μεταλλάξεις και τη φυσική επιλογή σε έναν διασταυρούμενο πληθυσμό συλλογικά ονομάζονται εξέλιξη.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Πού αποθηκεύονται οι γενετικές πληροφορίες;
- Εξηγήστε τις διαφορές στη χημική δομή του DNA και του RNA

Co-funded by the Erasmus+ Programme of the European Union



Leland Hartwell, Michael Goldberg, Janice Fischer, Lee Hood, Charles F. Aquadro , Bruce Bejcek . (2015) Γενετική: από τα γονίδια στα γονιδιώματα. McGraw-Hill Education, Νέα Υόρκη, Η.Π.Α

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 2. Γενετική
Θέμα 1. Μοριακή Γενετική
Μάθημα 5. Έκφραση γονιδίων

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Το Κεντρικό Δόγμα
- Μεταγραφή
- Μετάφραση
- Ο Γενετικός Κώδικας

Co-funded by the Erasmus+ Programme of the European Union

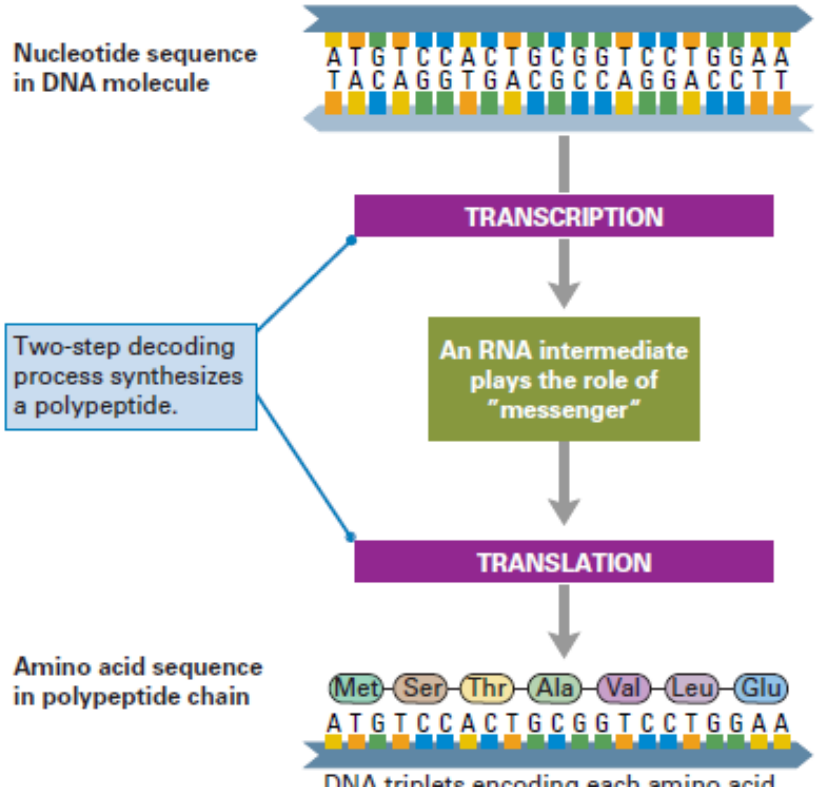
Εισαγωγή

Ο Watson και ο Crick είχαν δίκιο στην πρόταση ότι η γενετική πληροφορία στο DNA περιέχεται στην ακολουθία των βάσεων κατά τρόπο ανάλογο με τα γράμματα που είναι τυπωμένα σε μια λωρίδα χαρτιού. Σε μια περιοχή του DNA που κατευθύνει τη σύνθεση μιας πρωτεΐνης, ο γενετικός κώδικας για την πρωτεΐνη περιέχεται μόνο σε ένα σκέλος, και αποκωδικοποιείται με γραμμική σειρά. Μια τυπική πρωτεΐνη αποτελείται από ένα ή περισσότερα πολυπεπτίδια αλυσίδες? κάθε πολυπεπτιδική αλυσίδα αποτελείται από μια γραμμική αλληλουχία συνδεδεμένων αμινοξέων από άκρη σε άκρη. Για παράδειγμα, το ένζυμο ΠΑΥ αποτελείται από τέσσερα πανομοιότυπα πολυπεπτιδικές αλυσίδες, καθεμία από 452 αμινοξέα σε μήκος. Στην αποκωδικοποίηση του DNA, το καθένα η διαδοχική «κωδική λέξη» στο DNA καθορίζει το επόμενο αμινοξύ που θα προστεθεί στο πολυπεπτιδική αλυσίδα όπως κατασκευάζεται. Η ποσότητα DNA που απαιτείται για την κωδικοποίηση της πολυπεπτιδικής αλυσίδας της ΠΑΥ είναι επομένως 452 3 1356 ζεύγη νουκλεοτιδίων. Το συνολικό γονίδιο είναι πολύ μεγαλύτερο—περίπου 90.000 ζεύγη νουκλεοτιδίων. Μόνο το 1,5 τοις εκατό των γονιδίων είναι αφιερωμένο στην κωδικοποίηση των αμινοξέων. Το μη κωδικοποιητικό μέρος περιλαμβάνει ορισμένες ακολουθίες που ελέγχουν τη δραστηριότητα των γονιδίων, αλλά δεν είναι γνωστό πόσο από το γονίδιο εμπλέκεται στη ρύθμιση.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Το Κεντρικό Δόγμα



Υπάρχουν 20 διαφορετικά αμινοξέα. Μόνο τέσσερις βάσεις κωδικοποιούν αυτά τα 20 αμινοξέα, με κάθε «λέξη» στον γενετικό κώδικα να αποτελείται τριών παρακείμενων βάσεων. Για παράδειγμα, η βασική αλληλουχία ATG προσδιορίζει το αμινο όξινη μεθειονίνη (Met), TCC ορίζει σερίνη (Ser), το ACT καθορίζει τη θρεονίνη (Thr) και Το GCG προσδιορίζει την αλανίνη (Ala). Υπάρχουν 64 πιθανοί συνδυασμοί τριών βάσεων αλλά μόνο 20 αμινοξέα γιατί κάποιοι συνδυασμοί κωδικός για το ίδιο αμινοξύ. Για παράδειγμα, TCT, TCC, TCA, TCG, AGT και AGC όλοι οι κωδικοί για σερίνη (Ser) και CTT, CTC, CTA, CTG, TTA και TTG όλοι κωδικοποιούν τη λευκίνη (Leu). Ενα παράδειγμα της σχέσης μεταξύ των αλληλουχία βάσεων σε διπλό DNA και το αλληλουχία αμινοξέων της αντίστοιχης πρωτεΐνη. Το συγκεκριμένο Το διπλό DNA είναι η ανθρώπινη αλληλουχία που κωδικοποιεί για τα πρώτα επτά αμινοξέα στο πολυπεπτιδική αλυσίδα PAH .

Το DNA κωδικοποιεί την πρωτεΐνη όχι άμεσα αλλά έμμεσα μέσα από τις διαδικασίες της μεταγραφής και της μετάφρασης. Η έμμεση διαδρομή μεταφοράς πληροφοριών,

DNA → RNA → Πρωτεΐνη

είναι γνωστή ως το κεντρικό δόγμα της μοριακής γενεσιολογίας. Ο όρος δόγμα σημαίνει «σύνολο πεποιθήσεων»· χρονολογείται από την εποχή που η ιδέα προβάλλεται πρώτα ως θεωρία. Από τότε το «δόγμα» έχει επιβεβαιωθεί πειραματικά αλλά ο όρος επιμένει. Η κύρια έννοια στο κεντρικό δόγμα είναι ότι το DNA δεν κωδικοποιεί άμεσα την πρωτεΐνη αλλά μάλλον δρα μέσω ενός ενδιάμεσου μορίου που λέγεται ριβονουκλεϊκό οξύ (RNA). Η δομή του RNA είναι παρόμοια, αλλά όχι ταυτόσημη με, αυτήν του DNA. Υπάρχει διαφορά στο σάκχαρο (το RNA περιέχει το σάκχαρο ριβόζη αντί για δεοξυριβόζη), είναι συνήθως το RNA μονόκλωνο (όχι διπλό) και περιέχει τη βάση ουρακίλη (U) αντί για θυμίνη (T), η οποία υπάρχει στο DNA. Στην πραγματικότητα, συμμετέχουν τρεις τύποι RNA στη σύνθεση πρωτεϊνών:

Co-funded by the Erasmus+ Programme of the European Union

• Ένα μόριο αγγελιοφόρου RNA (mRNA), που μεταφέρει τη γενετική πληροφορία από DNA και χρησιμοποιείται ως πρότυπο για σύνθεση πολυπεπτίδιου. Στα περισσότερα μόρια mRNA, υπάρχει μεγάλη αναλογία νουκλεοτίδιων που στην πραγματικότητα κωδικοποιεί τα αμινοξέα. Για παράδειγμα, το mRNA για την PAH είναι 2400 νουκλεοτίδια σε μήκος και κωδικοί για ένα πολυπεπτίδιο 452 αμινοξέα; σε αυτή την περίπτωση, πάνω από 50% του μήκους του mRNA κωδικοποιεί για αμινοξέα.

• Διάφοροι τύποι ριβοσωμικού RNA (rRNA), που είναι κύρια συστατικά του κυτταρικού σωματίδια που ονομάζονται ριβοσώματα στα οποία λαμβάνει χώρα σύνθεση πολυπεπτιδίου.

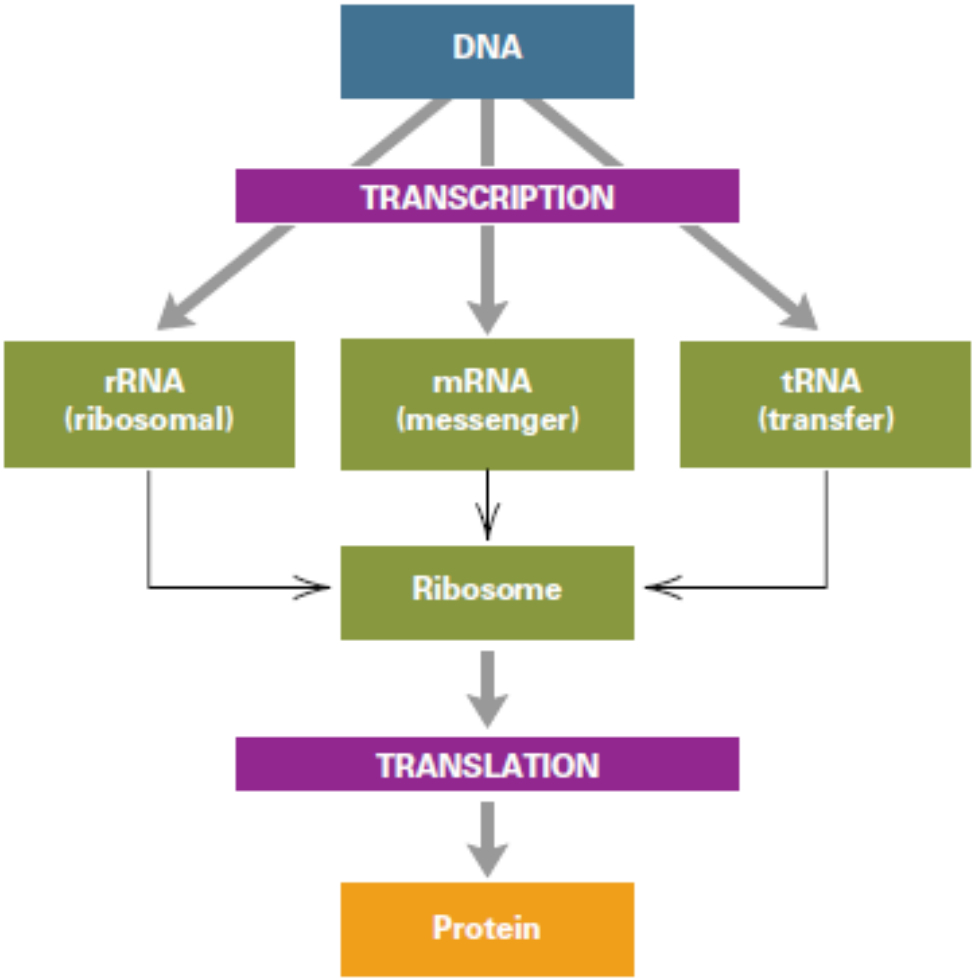
• Ένα σύνολο μορίων μεταφοράς RNA (tRNA), καθένα από τα οποία φέρει ένα συγκεκριμένο αμινοξύ καθώς και μια περιοχή αναγνώρισης τριών βάσεων ότι ζεύγη βάσης με μια ομάδα τριών γειτονικών βάσεις στο mRNA. Καθώς κάθε tRNA συμμετέχει σε μετάφραση γίνεται το αμινοξύ του η τερματική υπομονάδα προστέθηκε στο μήκος του την αναπτυσσόμενη πολυπεπτιδική αλυσίδα. Το tRNA που φέρει μεθειονίνη συμβολίζεται tRNAMet , αυτό που φέρει σερίνη συμβολίζεται tRNASer , και ούτω καθεξής.

Co-funded by the Erasmus+ Programme of the European Union

Το κεντρικό δόγμα είναι η θεμελιώδης αρχή της μοριακής γενετικής γιατί συνοψίζει τον τρόπο με τον οποίο υπάρχουν οι γενετικές πληροφορίες. Το DNA εκφράζεται σε αλληλουχία αμινοξέων σε πολυπεπτιδική αλυσίδα:

Η αλληλουχία των νουκλεοτιδίων σε ένα γονίδιο προσδιορίζει την αλληλουχία των νουκλεοτιδίων σε ένα μόριο του αγγελιοφόρος RNA; με τη σειρά της, η ακολουθία των νουκλεοτίδιων στο αγγελιοφόρο RNA καθορίζει την αλληλουχία των αμινοξέων στο πολυπεπτίδιο.

Co-funded by the Erasmus+ Programme of the European Union



Δεδομένης μιας διαδικασίας τόσο εννοιολογικά απλής όσο η κωδικοποίηση του DNA για πρωτεΐνη, τι μπορεί να εξηγήσει την πρόσθετη πολυπλοκότητα των ενδιάμεσων RNA; Ένας πιθανός λόγος είναι ότι ένα ενδιάμεσο RNA δίνει ένα άλλο επίπεδο για έλεγχο, για παράδειγμα, αποικοδομώντας το mRNA για μια πρωτεΐνη που δεν χρειάζεται. Ένας άλλος πιθανός λόγος μπορεί να είναι ιστορικός. Η δομή του RNA είναι μοναδική στο ότι έχει τόσο ένα πληροφοριακό περιεχόμενο που υπάρχει στην αλληλουχία βάσεων του όσο και μια πολύπλοκη, διπλωμένη τρισδιάστατη δομή που προικίζει ορισμένα μόρια RNA με καταλυτικές δραστηριότητες. Πολλοί επιστήμονες πιστεύουν ότι στις πρώτες μορφές ζωής, το RNA χρησίμευε τόσο για γενετικές πληροφορίες όσο και για κατάλυση. Καθώς προχωρούσε η εξέλιξη, ο πληροφοριακός ρόλος μεταφέρθηκε στο DNA και ο καταλυτικός στην πρωτεΐνη. Ωστόσο, το RNA κλειδώθηκε στην κεντρική του θέση ως ενδιάμεσος στις διαδικασίες μεταφοράς πληροφοριών και πρωτεϊνικής σύνθεσης.

Co-funded by the Erasmus+ Programme of the European Union

Αυτή η υπόθεση υπονοεί ότι τη συμμετοχή του RNA στη σύνθεση πρωτεϊνών είναι ένα λείψανο από τα πρώτα στάδια της εξέλιξης- ένα «μοριακό απολίθωμα». Η υπόθεση υποστηρίζεται από ποικίλες παρατηρήσεις. Για παράδειγμα,

(1) Η αντιγραφή του DNA απαιτεί ένα μόριο RNA για να ξεκινήσει

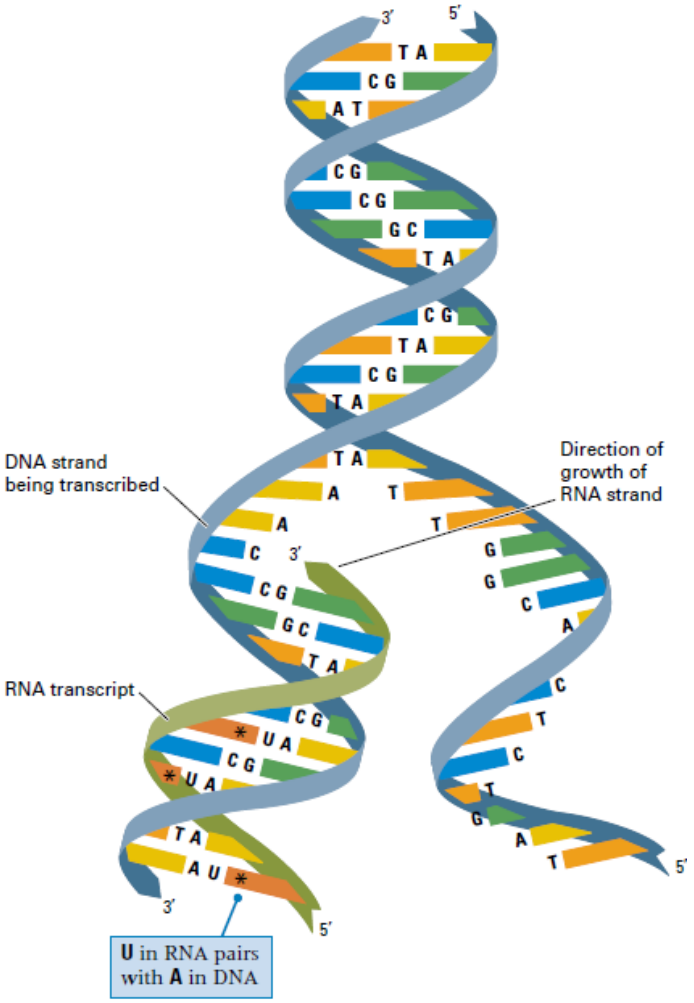
(2) ένα μόριο RNA είναι απαραίτητο στη σύνθεση των άκρων των χρωμοσωμάτων

(3) κάποια RNA μόρια δρουν για να καταλύουν βασικές αντιδράσεις στην πρωτεϊνική σύνθεση

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Μεταγραφή



Ο τρόπος με τον οποίο υπάρχουν οι γενετικές πληροφορίες μεταφέρονται από το DNA στο RNA: Το DNA ανοίγει, και ένα από τα σκέλη χρησιμοποιείται ως πρότυπο για τη σύνθεση ενός συμπληρωματικού κλώνου RNA. Η διαδικασία παρασκευής ενός κλώνου RNA από ένα πρότυπο DNA λέγεται **μεταγραφή** και το μόριο RNA που προκύπτει είναι η μεταγραφή.

Η ακολουθία βάσης στο RNA είναι συμπληρωματική (ζευγαρωμα Watson–Crick) σε αυτό το πρότυπο DNA, εκτός από το U (το οποίο κάνει ζεύγη με το A) που υπάρχει στο RNA στη θέση του T. Κάθε κλώνος RNA έχει πολικότητα—άκρο 5' και 3' τέλος — και, όπως στη σύνθεση του DNA, νουκλεοτίδια προστίθενται μόνο στο 3' άκρο του αυξανόμενου κλώνου RNA. Εξ ου και το 5' τέλος του RNA συντίθεται πρώτα και η μεταγραφή προχωρά κατά μήκος του προτύπου κλώνου DNA στην κατεύθυνση 3'-προς-5'. Καθε γονίδιο περιλαμβάνει αλληλουχίες νουκλεοτιδίων που ξεκινούν και τερματίζουν τη μεταγραφή.

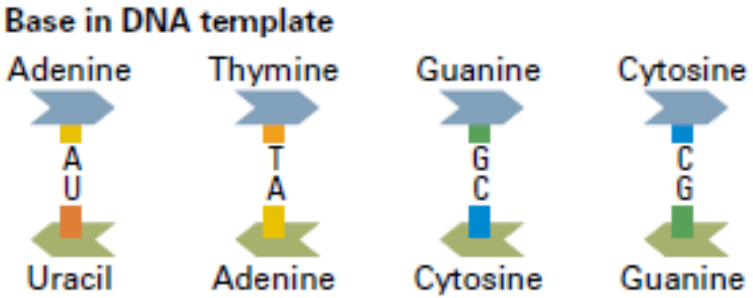
Co-funded by the Erasmus+ Programme of the European Union

Το μεταγράφημα RNA που παράγεται από οποιοδήποτε γονίδιο ξεκινά από τη θέση έναρξης στον κλώνο του εκμαγείου, ο οποίος βρίσκεται «ανοδικά» από την περιοχή κωδικοποίησης αμινοξέων και τελειώνει στη θέση τερματισμού, η οποία βρίσκεται «κατάντη» από το αμινοξύ που κωδικοποιεί περιοχή. Για οποιοδήποτε γονίδιο, το μήκος της μεταγραφής RNA είναι πολύ μικρότερο από το μήκος του DNA στο χρωμόσωμα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

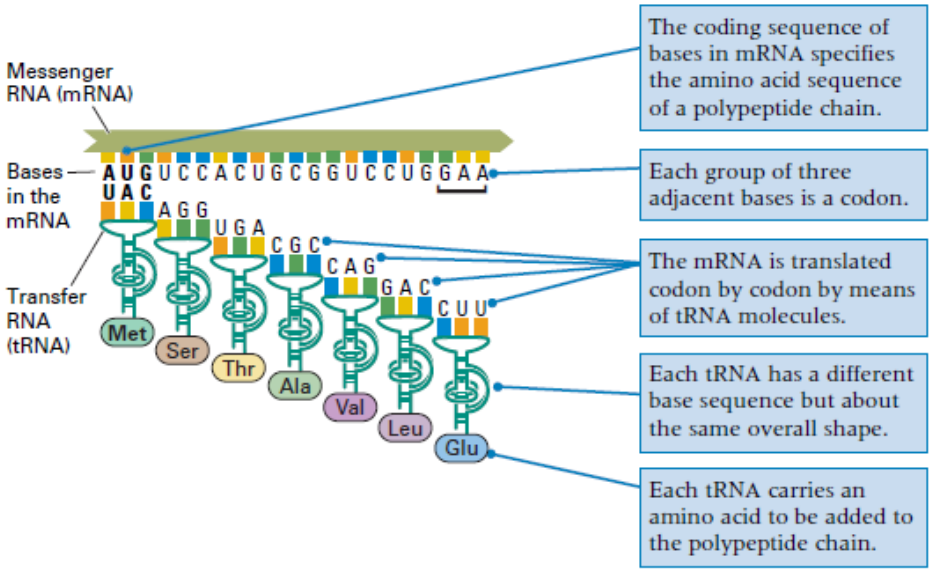
Μετάφραση



Το σωστό αμινοξύ συνδέεται με το άλλο άκρο του tRNA, και όταν το tRNA έρχεται σε σειρά, το αμινοξύ στο οποίο είναι επισυνάπτεται γίνεται η πιο πρόσφατη προσθήκη στο αναπτυσσόμενο άκρο του πολυπεπτιδίου αλυσίδα.

Η σύνθεση ενός πολυπεπτιδίου υπό την κατεύθυνση ενός μορίου mRNA είναι γνωστό ως μετάφραση. Αν και η ακολουθία των βάσεων στο mRNA κωδικοποιεί την αλληλουχία των αμινοξέων σε ένα πολυπεπτίδιο, τα μόρια που στην πραγματικότητα κάνουν τη «μετάφραση» είναι τα μόρια tRNA. Το μόριο mRNA μεταφράζεται σε μη επικαλυπτόμενες ομάδες του τριών βάσεων που ονομάζονται κωδικόνια. Για κάθε κωδικόνιο στο mRNA που προσδιορίζει ένα αμινοξύ, υπάρχει ένα μόριο tRNA που περιέχει συμπληρωματική ομάδα τριών γειτονικών βάσεων που μπορούν να ζευγαρώσουν με αυτές σε αυτό το κωδικόνιο.

Co-funded by the Erasmus+ Programme of the European Union

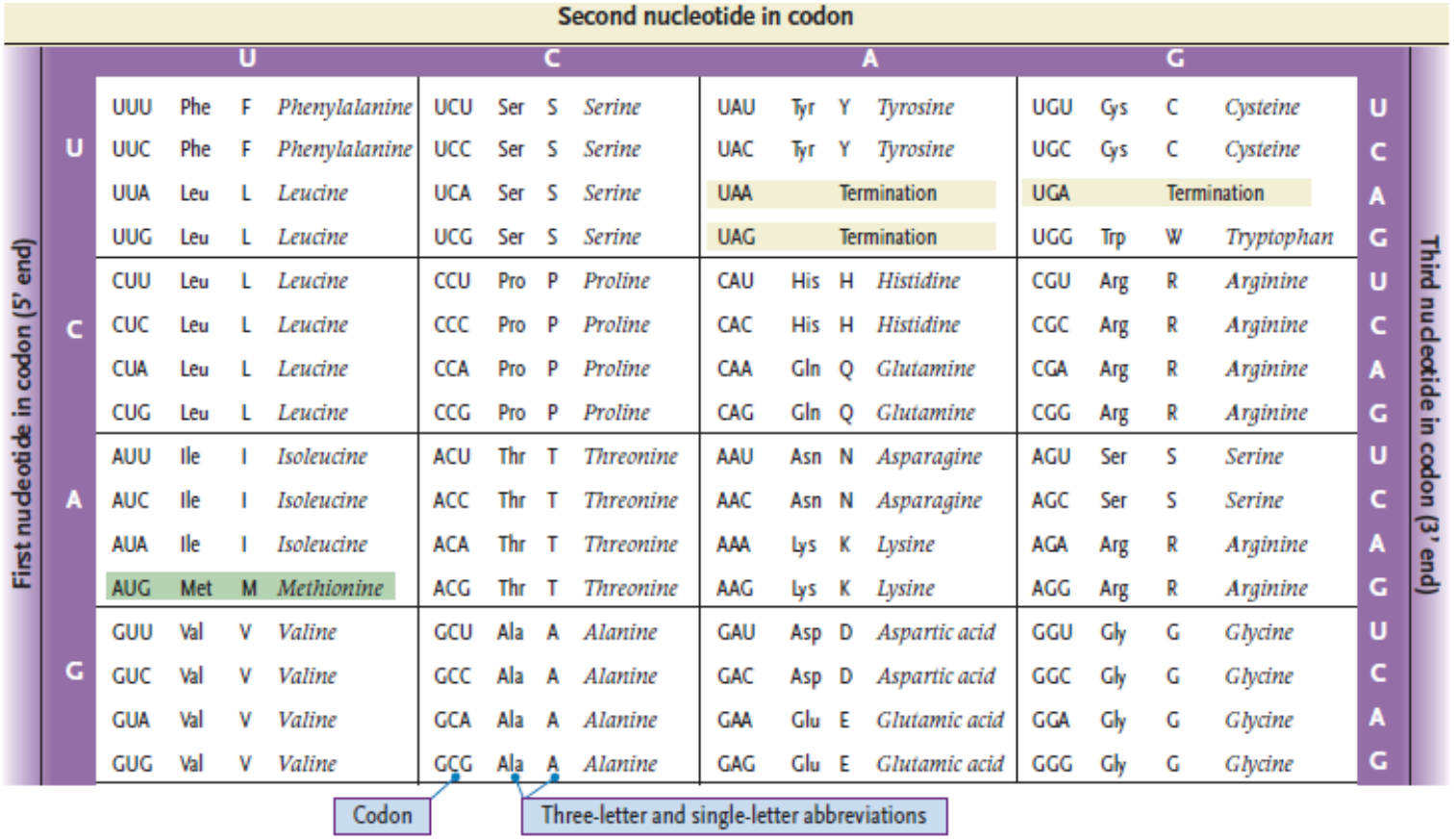


Ο ρόλος του tRNA στη μετάφραση και μπορεί να περιγραφεί ως ακολουθεί:

Το mRNA διαβάζεται κωδικόνιο με κωδικόνιο. Κάθε κωδικόνιο που προσδιορίζει ένα αμινοξύ ταιριάζει με μια συμπληρωματική ομάδα τριών γειτονικών βάσεων σε ένα μόνο μόριο tRNA. Το ένα άκρο του tRNA είναι συνδεδεμένο με το σωστό αμινοξύ, οπότε το σωστό αμινοξύ έρχεται σε ευθεία.

Co-funded by the Erasmus+ Programme of the European Union

Ο Γενετικός Κώδικας



First nucleotide in codon (5' end)	Second nucleotide in codon: U				Second nucleotide in codon: C				Second nucleotide in codon: A				Second nucleotide in codon: G				Third nucleotide in codon (3' end)
U	UUU	Phe	F	*Phenylalanine*	UCU	Ser	S	*Serine*	UAU	Tyr	Y	*Tyrosine*	UGU	Cys	C	*Cysteine*	U
	UUC	Phe	F	*Phenylalanine*	UCC	Ser	S	*Serine*	UAC	Tyr	Y	*Tyrosine*	UGC	Cys	C	*Cysteine*	C
	UUA	Leu	L	*Leucine*	UCA	Ser	S	*Serine*	UAA	Termination			UGA	Termination			A
	UUG	Leu	L	*Leucine*	UCG	Ser	S	*Serine*	UAG	Termination			UGG	Trp	W	*Tryptophan*	G
C	CUU	Leu	L	*Leucine*	CCU	Pro	P	*Proline*	CAU	His	H	*Histidine*	CGU	Arg	R	*Arginine*	U
	CUC	Leu	L	*Leucine*	CCC	Pro	P	*Proline*	CAC	His	H	*Histidine*	CGC	Arg	R	*Arginine*	C
	CUA	Leu	L	*Leucine*	CCA	Pro	P	*Proline*	CAA	Gln	Q	*Glutamine*	CGA	Arg	R	*Arginine*	A
	CUG	Leu	L	*Leucine*	CCG	Pro	P	*Proline*	CAG	Gln	Q	*Glutamine*	CGG	Arg	R	*Arginine*	G
A	AUU	Ile	I	*Isoleucine*	ACU	Thr	T	*Threonine*	AAU	Asn	N	*Asparagine*	AGU	Ser	S	*Serine*	U
	AUC	Ile	I	*Isoleucine*	ACC	Thr	T	*Threonine*	AAC	Asn	N	*Asparagine*	AGC	Ser	S	*Serine*	C
	AUA	Ile	I	*Isoleucine*	ACA	Thr	T	*Threonine*	AAA	Lys	K	*Lysine*	AGA	Arg	R	*Arginine*	A
	AUG	Met	M	*Methionine*	ACG	Thr	T	*Threonine*	AAG	Lys	K	*Lysine*	AGG	Arg	R	*Arginine*	G
G	GUU	Val	V	*Valine*	GCU	Ala	A	*Alanine*	GAU	Asp	D	*Aspartic acid*	GGU	Gly	G	*Glycine*	U
	GUC	Val	V	*Valine*	GCC	Ala	A	*Alanine*	GAC	Asp	D	*Aspartic acid*	GGC	Gly	G	*Glycine*	C
	GUA	Val	V	*Valine*	GCA	Ala	A	*Alanine*	GAA	Glu	E	*Glutamic acid*	GGA	Gly	G	*Glycine*	A
	GUG	Val	V	*Valine*	GCG	Ala	A	*Alanine*	GAG	Glu	E	*Glutamic acid*	GGG	Gly	G	*Glycine*	G

Codon

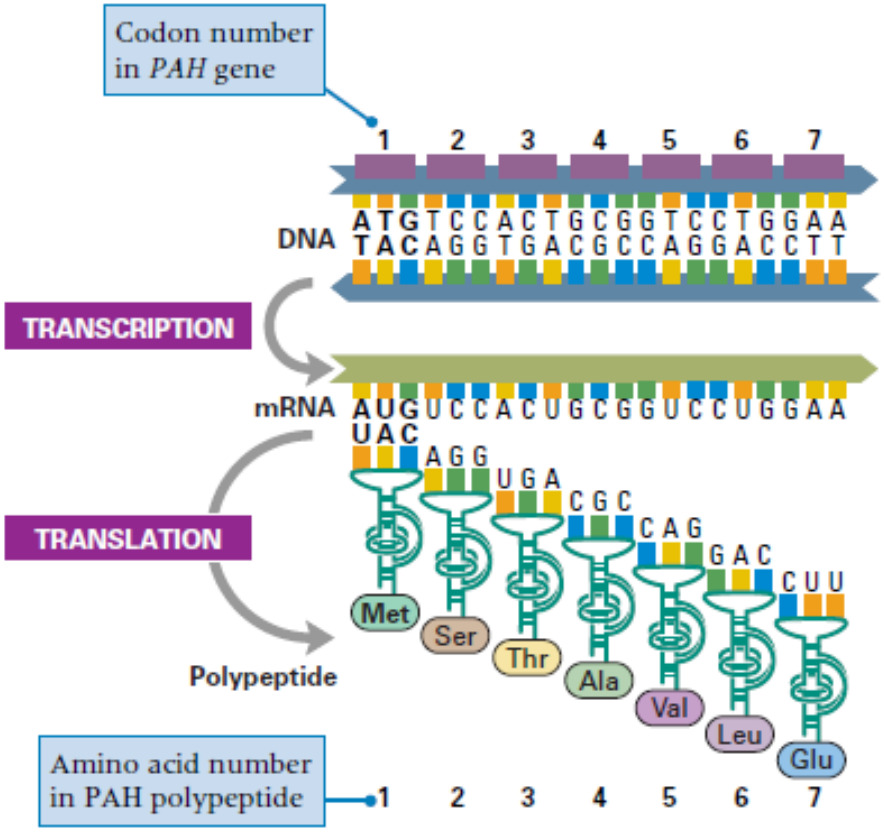
Three-letter and single-letter abbreviations

Εκτός από τα 61 κωδικόνια που κωδικοποιούν μόνο αμινοξέα, υπάρχουν τέσσερα κωδικόνια που έχουν εξειδικευμένες λειτουργίες:

• Το κωδικόνιο AUG, το οποίο καθορίζει το Met (μεθειονίνη), είναι επίσης το κωδικόνιο «έναρξης» για τη σύνθεση πολυπεπτιδίων. Η τοποθέτηση ενός tRNAMet δεσμευμένου στο AUG είναι ένα από τα πρώτα βήματα στην έναρξη της σύνθεσης πολυπεπτιδίου, επομένως όλες οι πολυπεπτιδικές αλυσίδες ξεκινούν με Met. (Πολλά πολυπεπτίδια έχουν αποκόψει το αρχικό Met μετά την ολοκλήρωση της μετάφρασης.) Στους περισσότερους οργανισμούς, το tRNAMet που χρησιμοποιείται για την έναρξη της μετάφρασης είναι το ίδιο tRNAMet που χρησιμοποιείται για τον προσδιορισμό της μεθειονίνης σε εσωτερικές θέσεις σε μια πολυπεπτιδική αλυσίδα.

• Τα κωδικόνια UAA, UAG και UGA είναι το καθένα μια «στάση» που καθορίζει τον τερματισμό της μετάφρασης και οδηγεί στην απελευθέρωση της ολοκληρωμένης πολυπεπτιδικής αλυσίδας από το ριβόσωμα. Αυτά τα κωδικόνια δεν έχουν μόρια tRNA που τα αναγνωρίζουν, αλλά αναγνωρίζονται από πρωτεϊνικούς παράγοντες που τερματίζουν τη μετάφραση.

Co-funded by the Erasmus+ Programme of the European Union



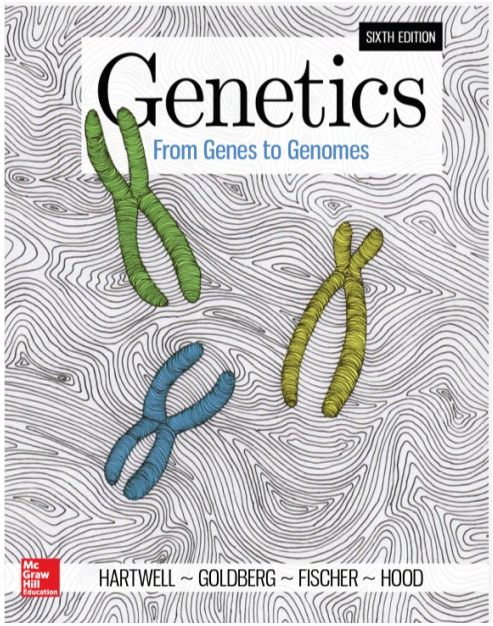
Το κεντρικό δόγμα σε δράση. Το DNA που κωδικοποιεί την ΠΑΥ χρησιμεύει ως πρότυπο για την παραγωγή ενός αγγελιοφόρου RNA, και το mRNA χρησιμεύει για τον προσδιορισμό της αλληλουχίας των αμινοξέων στην PAH πολυπεπτιδική αλυσίδα μέσω αλληλεπιδράσεων με το ριβόσωμα και τα tRNA μόρια.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε τη διαδικασία της γονιδιακής έκφρασης.
- Εξηγήστε τη διαδικασία της πρωτεϊνικής σύνθεσης σε ένα ζωντανό κύτταρο.

Co-funded by the Erasmus+ Programme of the European Union



Leland Hartwell, Michael Goldberg, Janice Fischer, Lee Hood, Charles F. Aquadro , Bruce Bejcek . (2015) Γενετική: από τα γονίδια στα γονιδιώματα. McGraw-Hill Education, Νέα Υόρκη, Η.Π.Α

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 2. Γενετική
Θέμα 1. Μοριακή Γενετική
Μάθημα 6. Μετάλλαξη, Αναδίπλωση Πρωτεϊνών και Σταθερότητα

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Τύποι μεταλλάξεων DNA και ο αντίκτυπός τους
- Τύποι αλλαγών στο DNA
- Αναδίπλωση και σταθερότητα πρωτεΐνης

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Ο όρος **μετάλλαξη** αναφέρεται σε οποιαδήποτε κληρονομήσιμη αλλαγή σε ένα γονίδιο (ή, γενικότερα, στο γενετικό υλικό) ή στη διαδικασία με την οποία λαμβάνει χώρα μια τέτοια αλλαγή. Ένας τύπος μετάλλαξης οδηγεί σε αλλαγή στην αλληλουχία των βάσεων στο DNA. Η αλλαγή μπορεί να είναι απλή, όπως η αντικατάσταση ενός ζεύγους βάσεων σε ένα μόριο διπλής όψης για ένα διαφορετικό ζεύγος βάσεων. Για παράδειγμα, ένα ζεύγος CG σε ένα μόριο διπλής όψης μπορεί να μεταλλαχθεί σε TA, AT ή GC. Η αλλαγή στην αλληλουχία βάσεων μπορεί επίσης να είναι πιο περίπλοκη, όπως η διαγραφή ή η προσθήκη ζευγών βάσεων. Οι γενετιστές χρησιμοποιούν επίσης τον όρο **μεταλλαγμένος** , ο οποίος αναφέρεται στο αποτέλεσμα μιας μετάλλαξης. Μια μετάλλαξη αποδίδει ένα μεταλλαγμένο γονίδιο, το οποίο με τη σειρά του παράγει ένα μεταλλαγμένο mRNA, μια μεταλλαγμένη πρωτεΐνη και τελικά έναν μεταλλαγμένο οργανισμό που παρουσιάζει τα αποτελέσματα της μετάλλαξης - για παράδειγμα, ένα εγγενές σφάλμα μεταβολισμού.

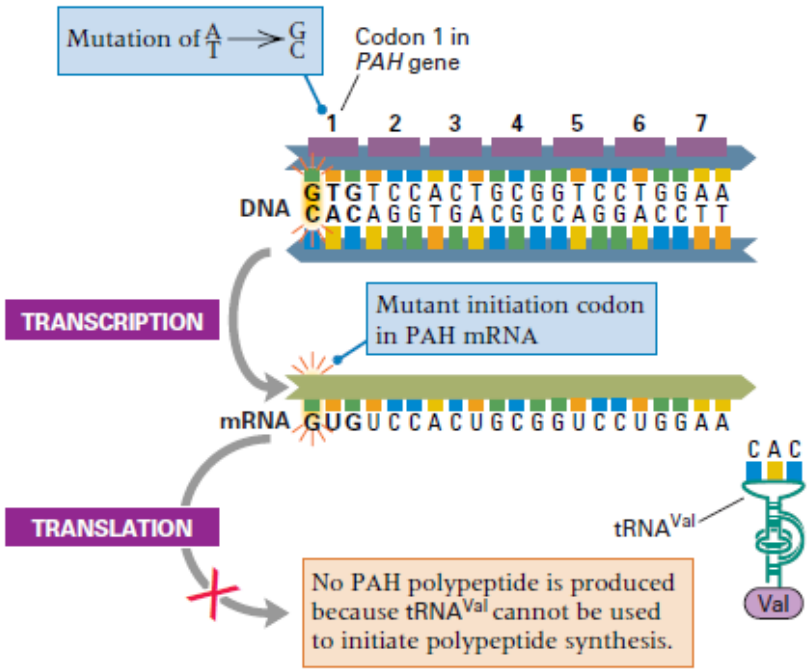
Co-funded by the Erasmus+ Programme of the European Union

Πότε συμβαίνουν γενετικές μεταλλάξεις;

Γενετικές μεταλλάξεις συμβαίνουν κατά την κυτταρική διαίρεση όταν τα κύτταρα διαιρούνται και αναπαράγονται. Υπάρχουν δύο τύποι κυτταρικής διαίρεσης:

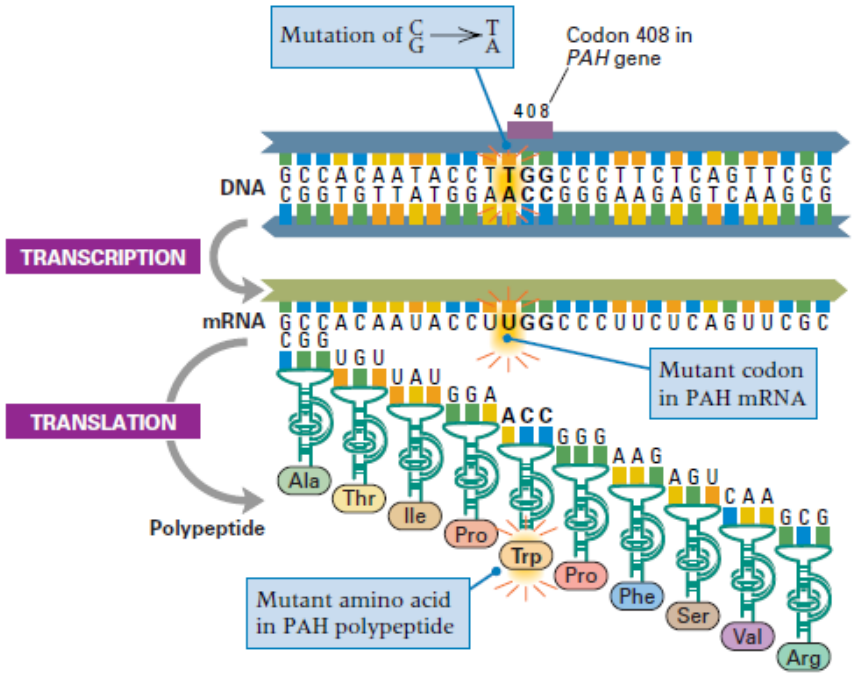
- **Μίτωση** : Η διαδικασία δημιουργίας νέων κυττάρων για το σώμα σας. Κατά τη διάρκεια της μίτωσης, τα γονίδιά σας καθοδηγούν τα κύτταρά σας να χωριστούν στα δύο δημιουργώντας ένα αντίγραφο των χρωμοσωμάτων σας.
- **Μείωση** : Η διαδικασία παραγωγής ωαρίων και σπερματοζωαρίων για την επόμενη γενιά. Κατά τη διάρκεια της μείωσης, τα χρωμοσώματα αντιγράφονται με τη μισή ποσότητα χρωμοσωμάτων από το πρωτότυπο (από 46 σε 23). Έτσι μπορείτε να πάρετε το γενετικό σας υλικό εξίσου από κάθε γονέα.

Co-funded by the Erasmus+ Programme of the European Union



Το DNA από ασθενείς από όλο τον κόσμο που πάσχουν από φαινυλκετονουρία έχει μελετηθεί για να καθοριστεί ποιοι τύποι μεταλλάξεων ευθύνονται για το εγγενές σφάλμα. Υπάρχει μεγάλη ποικιλία τύπων μεταλλαγμένων. Περισσότερες από 400 διαφορετικές μεταλλάξεις έχουν περιγραφεί στο γονίδιο της ΠΑΥ. Σε ορισμένες περιπτώσεις λείπει μέρος του γονιδίου, επομένως απουσιάζει η γενετική πληροφορία για τη δημιουργία ενός πλήρους ενζύμου PAH. Σε άλλες περιπτώσεις το γενετικό ελάττωμα είναι πιο λεπτό, αλλά το αποτέλεσμα εξακολουθεί να είναι είτε η αποτυχία παραγωγής μιας πρωτεΐνης PAH είτε η παραγωγή μιας πρωτεΐνης PAH που είναι ανενεργή.

Co-funded by the Erasmus+ Programme of the European Union



Μία μετάλλαξη PAH που είναι αρκετά κοινή ονομάζεται R408W, που σημαίνει ότι το κωδικόνιο 408 στην πολυπεπτιδική αλυσίδα PAH έχει αλλάξει από ένα που κωδικοποιεί την αργινίνη (R) σε ένα που κωδικοποιεί την τρυπτοφάνη (W). Αυτή η μετάλλαξη είναι μία από τις τέσσερις πιο κοινές μεταξύ των Ευρωπαίων Καυκάσιων με PKU.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Τι είναι οι γενετικές διαταραχές;

Μια γενετική διαταραχή είναι μια κατάσταση που προκαλείται από αλλαγές στο γονιδίωμά σας ή στο γενετικό υλικό που υπάρχει σε έναν άνθρωπο. Περιλαμβάνει το DNA, τα γονίδια και τα χρωμοσώματά σας. Διάφοροι παράγοντες προκαλούν γενετικές παθήσεις, όπως:

- Μετάλλαξη ενός γονιδίου (μονογενής).
- Μετάλλαξη πολλαπλών γονιδίων (πολυπαραγοντική κληρονομικότητα).
- Μετάλλαξη ενός ή περισσότερων χρωμοσωμάτων.
- Περιβαλλοντικοί παράγοντες (χημική έκθεση, ακτίνες UV) που αλλάζουν τη γενετική σας σύνθεση.

Μπορείτε να κληρονομήσετε τη γενετική πάθηση από τους γονείς σας (αν πρόκειται για DNA γεννητικών κυττάρων στο σπέρμα ή το ωάριο) ή η γενετική κατάσταση μπορεί να συμβεί τυχαία, χωρίς να έχετε ιστορικό της γενετικής πάθησης στην οικογένειά σας.

Co-funded by the Erasmus+ Programme of the European Union

Ποιες είναι οι κοινές γενετικές διαταραχές;

Υπάρχουν χιλιάδες γενετικές καταστάσεις. Μερικές από τις πιο κοινές γενετικές παθήσεις είναι:

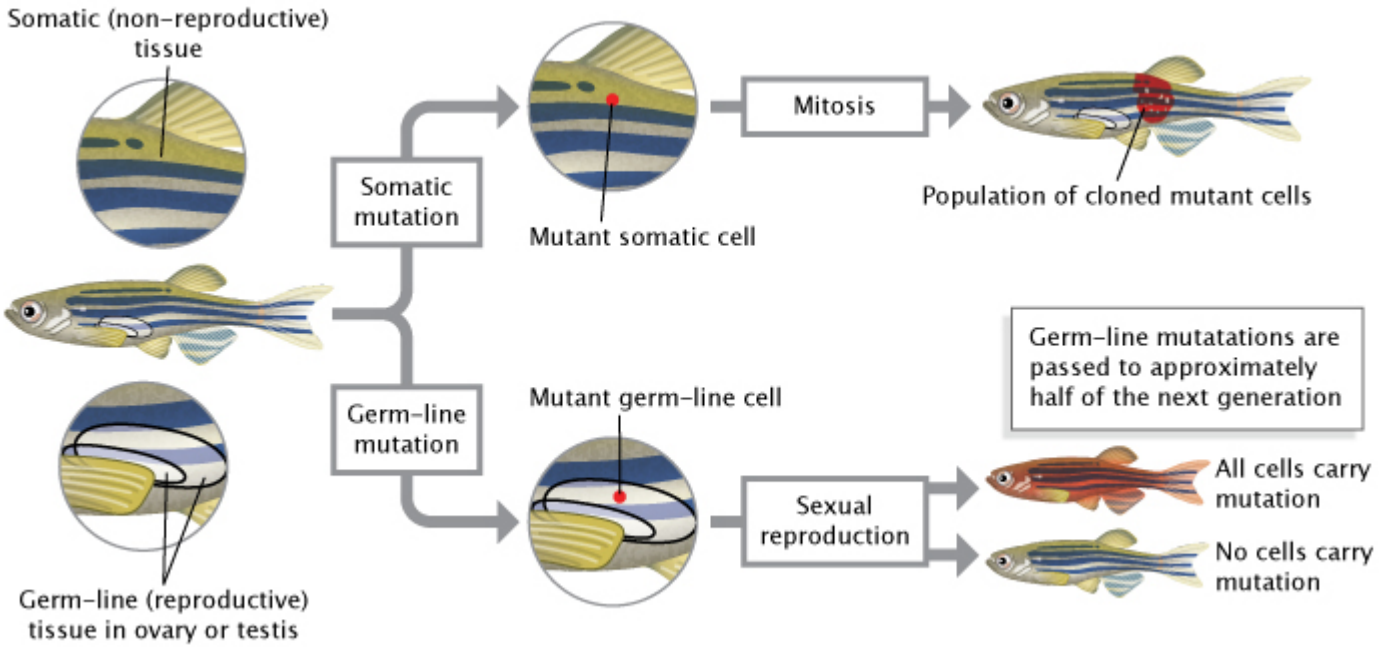
- Η ασθένεια Αλτσχάϊμερ.
- Μερικοί καρκίνοι.
- Κυστική ίνωση.
- Σύνδρομο Down.
- Δρεπανοκυτταρική αναιμία.

Co-funded by the Erasmus+ Programme of the European Union

Η σχέση μεταξύ μεταλλάξεων και πολυμορφισμών

Ενώ μια μετάλλαξη ορίζεται ως οποιαδήποτε αλλαγή στην αλληλουχία του DNA, οι βιολόγοι χρησιμοποιούν τον όρο "πολυμορφισμός ενός νουκλεοτιδίου" (SNP) για να αναφερθούν σε μια μόνο αλλαγή ζεύγους βάσεων που είναι κοινή στον πληθυσμό. Συγκεκριμένα, πολυμορφισμός είναι οποιαδήποτε γενετική τοποθεσία στην οποία βρίσκονται τουλάχιστον δύο διαφορετικές αλληλουχίες, με κάθε αλληλουχία να υπάρχει σε τουλάχιστον 1% του πληθυσμού. Σημειώστε ότι ο όρος "πολυμορφισμός" χρησιμοποιείται γενικά για να αναφέρεται σε μια φυσιολογική παραλλαγή ή σε μια παραλλαγή που δεν προκαλεί άμεσα ασθένεια. Επιπλέον, το όριο επικράτησης τουλάχιστον 1% για μια παραλλαγή που πρέπει να ταξινομηθεί ως πολυμορφισμός είναι κάπως αυθαίρετο. Εάν η συχνότητα είναι χαμηλότερη από αυτή, το αλληλόμορφο τυπικά θεωρείται ως μετάλλαξη.

Τύποι αλλαγών στο DNA



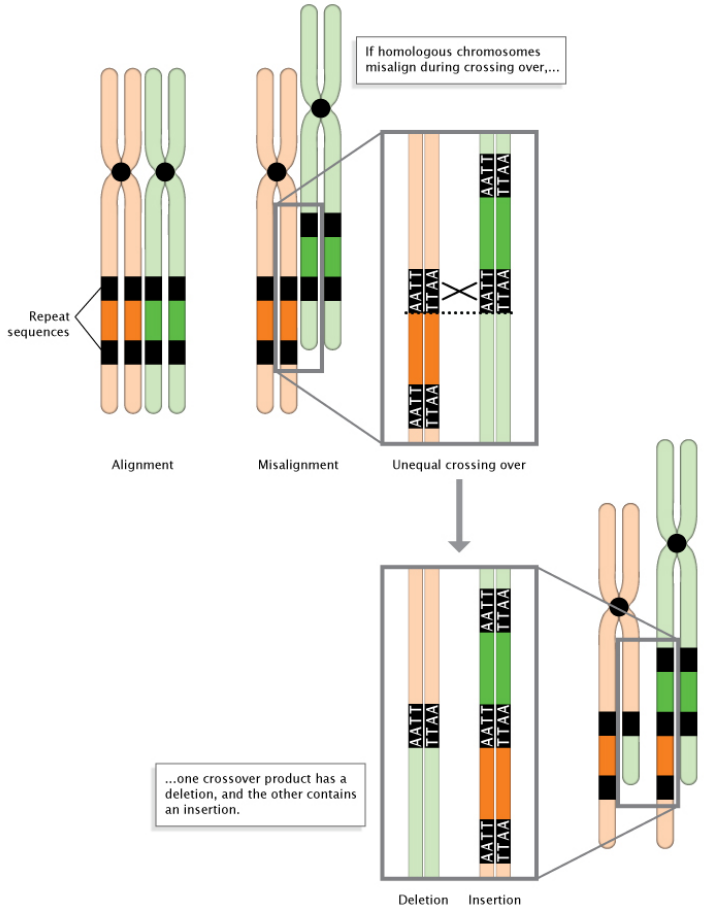
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Τύποι μεταλλάξεων DNA και ο αντίκτυπός τους

Class of Mutation	Type of Mutation	Description	Human Disease(s) Linked to This Mutation
Point mutation	Substitution	One base is incorrectly added during replication and replaces the pair in the corresponding position on the complementary strand	Sickle-cell anemia
	Insertion	One or more extra nucleotides are inserted into replicating DNA, often resulting in a frameshift	One form of beta-thalassemia
	Deletion	One or more nucleotides is "skipped" during replication or otherwise excised, often resulting in a frameshift	Cystic fibrosis
Chromosomal mutation	Inversion	One region of a chromosome is flipped and reinserted	Opitz-Kaveggia syndrome
	Deletion	A region of a chromosome is lost, resulting in the absence of all the genes in that area	Cri du chat syndrome
	Duplication	A region of a chromosome is repeated, resulting in an increase in dosage from the genes in that region	Some cancers
	Translocation	A region from one chromosome is aberrantly attached to another chromosome	One form of leukemia
Copy number variation	Gene amplification	The number of tandem copies of a locus is increased	Some breast cancers
	Expanding trinucleotide repeat	The normal number of repeated trinucleotide sequences is expanded	Fragile X syndrome, Huntington's disease

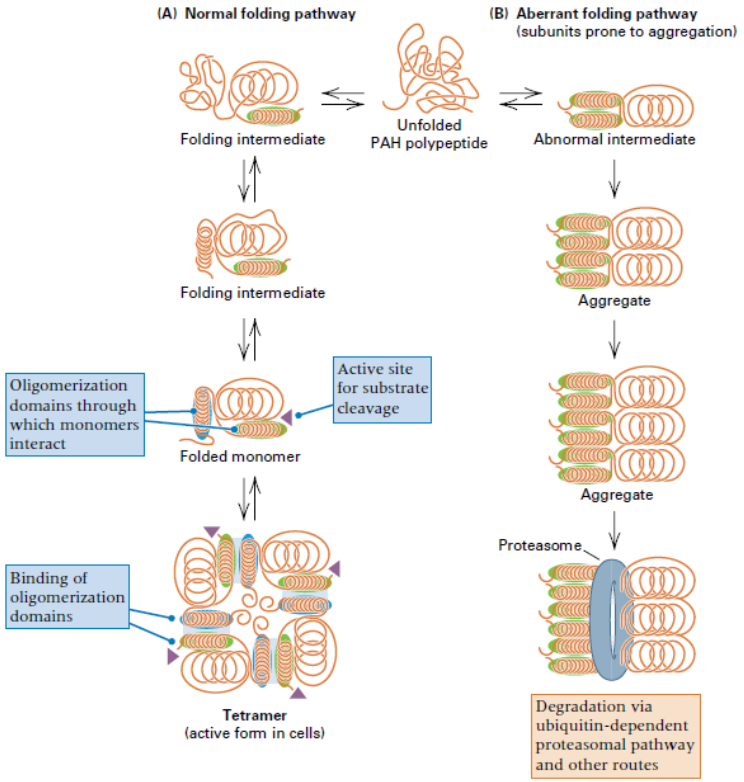
Άνιση διασταύρωση κατά τη διάρκεια της μείωσης



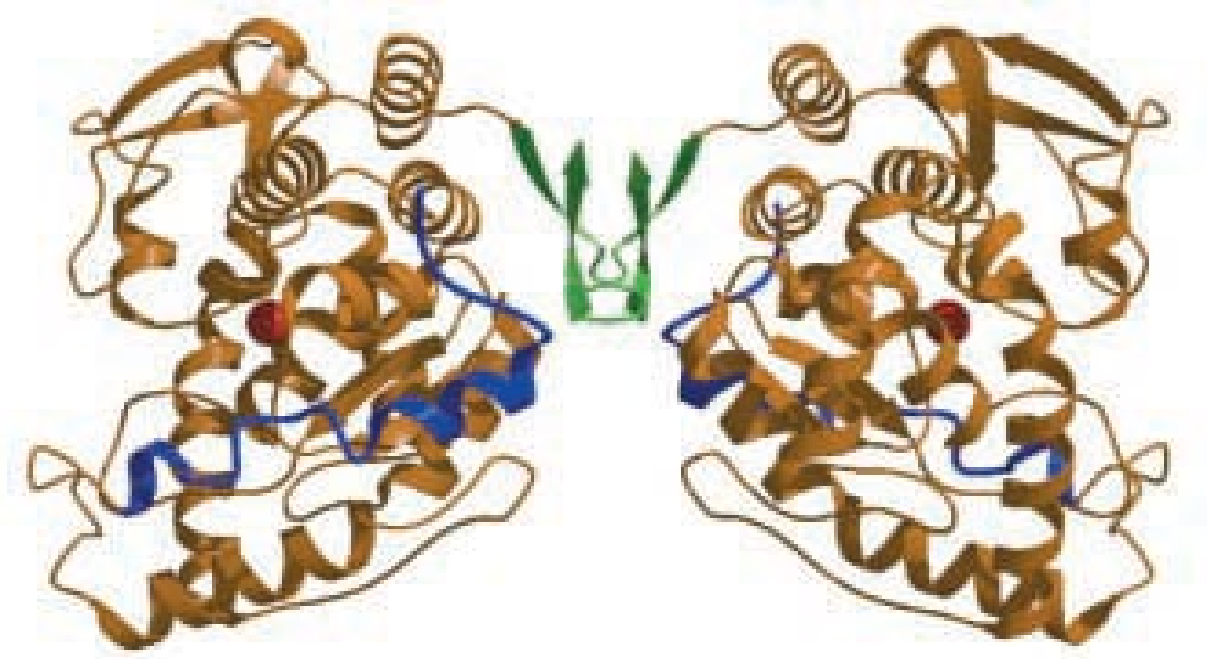
Πώς συμβαίνουν οι μεταλλάξεις

- Μεταλλάξεις και Περιβάλλον
- Μεταλλάξεις που προκαλούνται από χημικές ουσίες
- Αυθόρμητες Μεταλλάξεις
- Σφάλματα κατά την αντιγραφή του DNA

Αναδίπλωση και σταθερότητα πρωτεΐνης



Ορισμένες αντικαταστάσεις αμινοξέων διαταράσσουν την ικανότητα μιας πρωτεΐνης να αναδιπλώνεται σωστά. (A) Η κανονική αναδίπλωση σε υδροξυλάση φαινυλαλανίνης σχηματίζει το ενεργό τετραμερές. (B) Η μη φυσιολογική αναδίπλωση μιας μεταλλαγμένης πολυπεπτιδικής αλυσίδας έχει ως αποτέλεσμα το σχηματισμό πολυπεπτιδικών συσσωματωμάτων, τα οποία διασπώνται προοδευτικά στα συστατικά αμινοξέα μέσω μιας οδού



Τρισδιάστατη δομή δύο από τις τέσσερις υπομονάδες στο ενεργό τετραμερές της υδροξυλάσης φαινυλαλανίνης. Οι αλυσίδες των αμινοξέων αντιπροσωπεύονται ως μια ακολουθία από μπούκλες, βρόχους και επίπεδα βέλη, τα οποία αντιπροσωπεύουν διαφορετικούς τύπους τοπικής δομής. Η περιοχή ολιγομερισμού (σε αυτή την περίπτωση, τετραμερισμός) εμφανίζεται με πράσινο και η καταλυτική περιοχή που περιέχει την ενεργό θέση εμφανίζεται με χρυσό

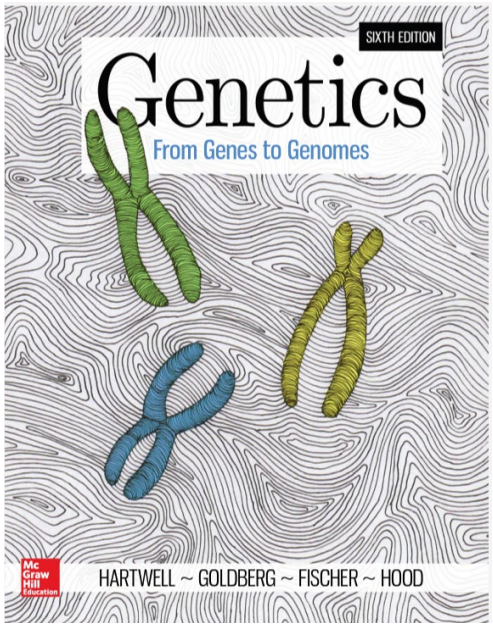
Εικόνα 1

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Πότε συμβαίνουν γενετικές μεταλλάξεις;
- Τι είναι οι γενετικές διαταραχές;
- Πώς συμβαίνουν οι μεταλλάξεις;

Co-funded by the Erasmus+ Programme of the European Union



Leland Hartwell, Michael Goldberg, Janice Fischer, Lee Hood, Charles F. Aquadro , Bruce Bejcek . (2015) Γενετική: από τα γονίδια στα γονιδιώματα. McGraw-Hill Education, Νέα Υόρκη, Η.Π.Α

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 2. Γενετική

Θέμα 2. Θεωρητική Μοντελοποίηση στη Γονιδιωματική

Μάθημα 1. Θεωρητική Μοντελοποίηση στη Γονιδιωματική

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Με γνώμονα τα δεδομένα: Βιοπληροφορική
- Ανάλυση ισορροπίας ροής
- Συνήθεις διαφορικές εξισώσεις
- Στοχαστική

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Ο στόχος της έρευνας στη θεωρητική μοντελοποίηση είναι να κατανοήσει τις εσωτερικές λειτουργίες των βιολογικών διεργασιών σε ένα θεμελιώδες επίπεδο, και από αυτό να διατυπώσει ακριβείς προβλέψεις που οδηγούν σε σχεδιασμός πειραμάτων και εφαρμογών. Ως εκ τούτου, η θεωρητική μοντελοποίηση τείνει να συμπληρώνει την πειραματική έρευνα προσφέροντας νέες προοπτικές. Γνωστό και ως μοντελοποίηση, αυτό το ζωντανό θέμα μπορεί να περιλαμβάνει προσομοίωση βιολογικών συστημάτων *in silico* , ανάλυση μεγάλων ποσοτήτων πειραματικών δεδομένων και δημιουργία αναλυτικών θεωριών για την πρόβλεψη μετρήσιμων φαινομένων.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Με γνώμονα τα δεδομένα: Βιοπληροφορική

Η ποικιλομορφία των ζωντανών συστημάτων είναι μεγάλη. Από το 2011, ο αριθμός των διαφορετικών ειδών ευκαρυωτών και μόνο εκτιμήθηκε ότι ήταν περίπου (8,7 ± 1,3) εκατομμύρια. Αυτά, με τη σειρά τους, μπορεί το καθένα να περιέχει χιλιάδες γονίδια στο DNA τους. Ως αποτέλεσμα αυτής της ποικιλομορφίας, όταν κάποιος μελετά ζωντανά συστήματα, καταλήγει συχνά με τεράστιες ποσότητες δεδομένων. Η ανάλυσή τους και η ανάκτηση χρήσιμων πληροφοριών έχει γίνει ένα πεδίο από μόνο του: Βιοπληροφορική.

Co-funded by the Erasmus+ Programme of the European Union

Με γνώμονα τα δεδομένα: Βιοπληροφορική

Η βιοπληροφορική εφαρμόζεται σε ένα ευρύ φάσμα τομέων. Μερικά από τα πιο σημαντικά είναι τα ακόλουθα:

- Φυλογενετική: η μελέτη των εξελικτικών σχέσεων μεταξύ γονιδίων και οργανισμών.
- Omics: ο χαρακτηρισμός και η ποσοτικοποίηση των βιομορίων.
- Βιολογία συστημάτων: η εξερεύνηση βιολογικών συστημάτων με τη μελέτη των αλληλεπιδράσεων μεταξύ των συστατικών τους.
- Λειτουργική επισημείωση: ο χαρακτηρισμός των γονιδίων και η λειτουργία τους.
- Πρόβλεψη δομής πρωτεΐνης και μοντελοποίηση ομολογίας: η μελέτη της σχέσης μεταξύ της σύνθεσης της αλυσίδας αμινοξέων και της προκύπτουσας τρισδιάστατης διαμόρφωσης.
- Ευθυγράμμιση αλληλουχίας: η ανάλυση ομοιοτήτων μεταξύ δύο ή περισσότερων αλληλουχιών DNA, RNA ή πρωτεΐνης.

Co-funded by the Erasmus+ Programme of the European Union

Ανάλυση ισορροπίας ροής

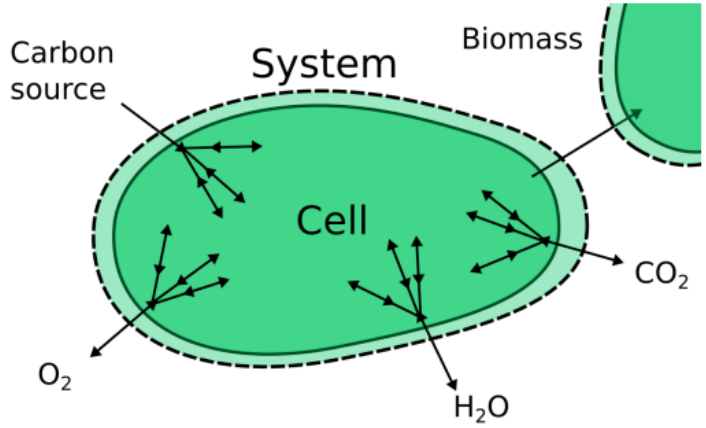
Μόλις υπάρχουν αρκετά δεδομένα σε ένα σύστημα, μπορεί να κατασκευαστεί ένα μοντέλο που να περιγράφει τις ιδιότητές του. Ωστόσο, τα διαθέσιμα δεδομένα μπορεί μερικές φορές να μην είναι ακόμη επαρκή για τη διατύπωσή τους με δυναμικό τρόπο. Ένα κοινό παράδειγμα για μια εναλλακτική σε τέτοιες δυναμικές προσεγγίσεις, στον τομέα των μεταβολικών μοντέλων κλίμακας γονιδιώματος, είναι η λεγόμενη μέθοδος ανάλυσης ισορροπίας ροής.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ανάλυση ισορροπίας ροής

Τα μοντέλα κλίμακας γονιδιώματος προορίζονται να προσομοιώσουν μια ολόκληρη μεταβολική οδό ή ακόμα και έναν ολόκληρο οργανισμό. Ως αποτέλεσμα, περιλαμβάνουν πολλούς τύπους συζευγμένων χημικών αντιδράσεων και μεταβολιτών. Αυτές περιλαμβάνουν τις λεγόμενες αντιδράσεις ανταλλαγής, οι οποίες αντιπροσωπεύουν την εισαγωγή ή την εξαγωγή μεταβολιτών, καθώς και αντιδράσεις στο εσωτερικό του συστήματος, οι οποίες παράγουν ή καταναλώνουν τους μεταβολίτες ανταλλαγής και άλλα ενδιάμεσα.



Σκίτσο ενός συστήματος κλίμακας γονιδιώματος ενός βιολογικού κυττάρου. Τα βέλη που προέρχονται από ή οδηγούν έξω από το κύτταρο αντιπροσωπεύουν αντιδράσεις ανταλλαγής, ενώ τα βέλη, τα οποία βρίσκονται εξ ολοκλήρου μέσα στο κύτταρο, αντιστοιχούν σε ενδοκυτταρικές αντιδράσεις.

Co-funded by the Erasmus+ Programme of the European Union

Ανάλυση ισορροπίας ροής

Για την εκτέλεση δυναμικών προσομοιώσεων με συστήματα αυτής της κλίμακας και αυτού του επιπέδου λεπτομέρειας, θα συναντούσε κανείς δύο ουσιαστικές δυσκολίες:

- Πρώτον, υπολογιστικά μιλώντας, η επίλυση ενός μερικώς συζευγμένου συστήματος χιλιάδων εξισώσεων δυναμικής αντίδρασης είναι μάλλον δαπανηρή.
- Δεύτερον, όλες οι εξισώσεις απαιτούν εμπειρικές παραμέτρους, όπως οι ρυθμοί αντίδρασης, για πολλές από τις οποίες δεν υπάρχουν διαθέσιμα πειραματικά δεδομένα.

Η προσέγγιση ανάλυσης ισοζυγίου ροής παρακάμπτει και τα δύο αυτά προβλήματα κάνοντας δύο υποθέσεις: συνθήκες σταθερής κατάστασης και βελτιστοποίηση των ροών προς έναν στόχο. Κατά συνέπεια, αυτή η διατύπωση προορίζεται να υποθέσει ένα στατικό σύστημα και όχι να προσομοιώσει τη δυναμική.

Συνήθεις διαφορικές εξισώσεις

Οι δυναμικοί ντετερμινιστικοί αλγόριθμοι ανήκουν στην κατηγορία εκείνων που προσομοιώνουν τη χρονική εξέλιξη ενός συστήματος. Συχνά λειτουργούν με βάση μία ή περισσότερες συνηθισμένες διαφορικές εξισώσεις (ODEs). Η επίλυση των εξισώσεων ενός τέτοιου μοντέλου παρέχει, συνήθως, μια συνέχεια λύσεων, οι οποίες διακρίνονται μέσω της αποκλειστικής τους εξάρτησης από τις αρχικές συνθήκες. Μόλις καθοριστούν αυτές οι αρχικές συνθήκες, η δυναμική ορίζεται πλήρως και μπορεί να βρεθεί μια μοναδική λύση. Συγκεκριμένα, αυτό σημαίνει ότι, εξαιρουμένης της ψηφιοποίησης και των αριθμητικών σφαλμάτων, οποιαδήποτε προσομοίωση που εκτελείται με τις ίδιες παραμέτρους έναρξης έχει ακριβώς το ίδιο αποτέλεσμα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

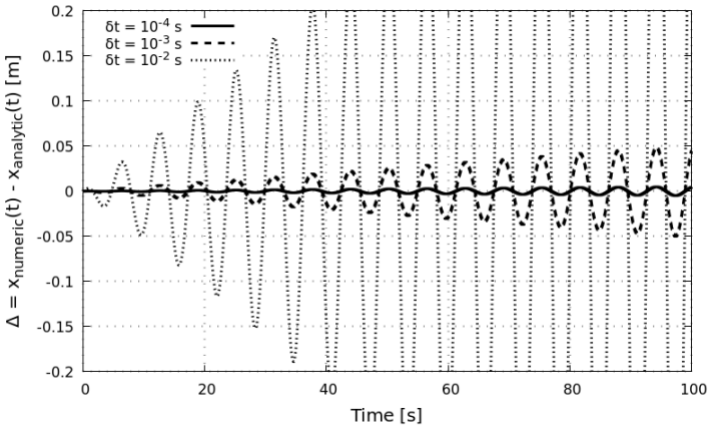
Co-funded by the Erasmus+ Programme of the European Union

Συνήθεις διαφορικές εξισώσεις

Εάν είναι γνωστές οι αρχικές συνθήκες, η χρονική εξέλιξη μπορεί, καταρχήν, να προσομοιωθεί με χαμηλή υπολογιστική προσπάθεια. Η αντιστάθμιση έγκειται στην ακρίβεια και τη σταθερότητα. Κατά τη διάρκεια κάθε χρονικού βήματος, η προσέγγιση της πεπερασμένης διαφοράς οδηγεί σε σφάλμα της τάξης δt^2. Καθώς αυτό συσσωρεύεται, η αριθμητική λύση αποκλίνει όλο και περισσότερο από την αναλυτική. Αυτό φαίνεται πολύ καλά όταν συγκρίνουμε αριθμητικές λύσεις της απλής αρμονικής εξίσωσης ταλαντωτή: $md^2x/dt^2 = -kx$, με την αναλυτική της λύση:

$$x(t) = A \cdot \cos(\sqrt{\frac{k}{m}} \cdot t + \phi),$$

όπου A = 1, k = 1 και m = 1 σε αυτό το παράδειγμα.



Διαφορά Δ μεταξύ της αναλυτικής λύσης και της αριθμητικής λύσης του απλού αρμονικού ταλαντωτή που δημιουργείται μέσω του αλγόριθμου Euler. Κάθε καμπύλη αντιπροσωπεύει ένα διαφορετικό μέγεθος του χρονικού βήματος ολοκλήρωσης.

Στοχαστική

Ορισμένες διαδικασίες, για παράδειγμα η κύλιση μιας μήτρας ή η ραδιενεργή διάσπαση ενός ατόμου, δεν μπορούν να περιγραφούν με ακρίβεια με ντετερμινιστικά μοντέλα, επειδή περιέχουν κάποια εγγενή τυχαιότητα. Ωστόσο, παρόλο που η ακριβής δυναμική του συστήματος δεν μπορεί να μοντελοποιηθεί, η μέση συμπεριφορά του συχνά παρουσιάζει ενδιαφέρον. Μπορεί επίσης να είναι ότι το υπολογιστικό κόστος της χρήσης μιας προσέγγισης που βασίζεται σε ODE είναι πολύ υψηλό. Αυτά τα σημεία υποστηρίζουν τη χρήση στοχαστικών μοντέλων και δύο κοινών τύπων αλγορίθμων, του Monte Carlo και του Gillespie.

Αλγόριθμοι Μόντε Κάρλο

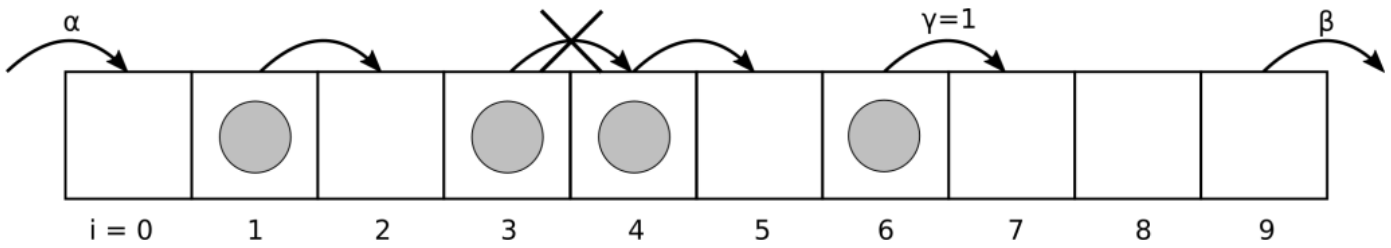
Οι αλγόριθμοι του Μόντε Κάρλο είναι μερικοί από τους πρώτους στοχαστικούς αλγόριθμους και ίσως οι πιο συνηθισμένοι.

Η συγκεκριμένη μέθοδος ποικίλλει, αλλά το κύριο σημείο είναι η σταδιακή δειγματοληψία τυχαίων τιμών από μια κατανομή καταστάσεων συστήματος. Για να το θέσουμε αυτό με λιγότερο αφηρημένους όρους, εξετάστε το πρόβλημα της εύρεσης του ελάχιστου αριθμού μιας συνάρτησης. Ένας απλός αλγόριθμος Monte Carlo θα απαιτούσε τυχαία επιλογή και σύγκριση τιμών συναρτήσεων από ένα δεδομένο διάστημα. Εάν σε κάθε βήμα το ελάχιστο ορίζεται ως το μικρότερο από την προηγούμενη και την τρέχουσα τιμή συνάρτησης, επιτυγχάνεται μια τιμή κοντά στο πραγματικό ελάχιστο μετά από επαρκή αριθμό βημάτων. Ένας πιο εκλεπτυσμένος αλγόριθμος Monte Carlo μπορεί επίσης να περιλαμβάνει προκαταλήψεις προς ορισμένες ιδιότητες του συστήματος.

Ο αλγόριθμος Gillespie

Ο αλγόριθμος Gillespie μπορεί να θεωρηθεί ως επέκταση των αλγορίθμων του Μόντε Κάρλο. Ωστόσο, περιέχει δύο βασικές διαφορές.

Το πρώτο είναι το γεγονός ότι σε κάθε βήμα ενημερώνεται η λίστα των αλλαγών (ονομάζονται επίσης αντιδράσεις), οι οποίες είναι επί του παρόντος δυνατές εντός των κανόνων και των περιορισμών του μοντέλου. Στη συνέχεια, η νέα αλλαγή που πρέπει να γίνει κατά τη διάρκεια του βήματος επιλέγεται τυχαία σε αυτήν τη λίστα. Κατά συνέπεια, όλες οι επιλεγμένες αλλαγές μπορούν, σίγουρα, να πραγματοποιηθούν. Αυτό είναι διαφορετικό από την κατάσταση στους κανονικούς αλγόριθμους Monte Carlo, όπου η τυχαία επιλεγμένη αλλαγή επιχειρείται αλλά δεν είναι πάντα εφικτή. Ένας τρόπος για να οπτικοποιήσετε τη διαφορά είναι να εξετάσετε τη λεγόμενη Διαδικασία Πλήρους Ασύμμετρης Απλής Εξαίρεσης (TASEP).



Κατά την προσομοίωση της συμπεριφοράς του συστήματος χρησιμοποιώντας έναν αλγόριθμο Monte Carlo, υπάρχουν δύο επιλογές. Στον απλούστερο αλγόριθμο, γίνεται δειγματοληψία μιας τοποθεσίας πλέγματος από τη λίστα όλων των τοποθεσιών σε κάθε βήμα. Στη συνέχεια, ελέγχεται εάν η τοποθεσία περιέχει ένα σωματίδιο και τέλος, εάν το σωματίδιο μπορεί να κινηθεί. Καθώς μπορεί να υπάρχουν πολλές κενές θέσεις πλέγματος, αυτό μπορεί να οδηγήσει σε μεγάλο αριθμό βημάτων που απορρίπτονται. Κατά την εφαρμογή του δεύτερου, πιο εκλεπτυσμένου αλγόριθμου, η δειγματοληψία γίνεται μόνο από τη λίστα των σωματιδίων στο πλέγμα. Αυτό μειώνει τον αριθμό των βημάτων που δεν μπορούν να πραγματοποιηθούν, επειδή δεν γίνεται πλέον δειγματοληψία κενών τοποθεσιών πλέγματος. Ωστόσο, είναι ακόμα πιθανό το σωματίδιο που λαμβάνεται δείγμα να μην μπορεί να κινηθεί, οπότε το βήμα δεν είναι επίσης αποδεκτό.

Όταν χρησιμοποιείτε τον αλγόριθμο Gillespie από την άλλη πλευρά, η δειγματοληψία γίνεται, όχι από τη λίστα σωματιδίων ή δικτυακών τοποθεσιών, αλλά από τη λίστα με τις τρέχουσες πιθανές κινήσεις. Με αυτόν τον τρόπο, οποιαδήποτε αλλαγή γίνεται κατά τη διάρκεια ενός βήματος είναι πάντα δυνατή και επομένως είναι πάντα αποδεκτή. Αυτή η διαφορά στη στρατηγική δειγματοληψίας μπορεί να οδηγήσει σε δραστικές βελτιώσεις όσον αφορά την υπολογιστική προσπάθεια. Χρησιμοποιώντας το παράδειγμα από το Σχήμα 10, ο απλός αλγόριθμος Monte Carlo θα οδηγήσει σε δειγματοληψία από μια λίστα δέκα τοποθεσιών, επτά από τις οποίες θα απορριφθούν. Ο πιο εκλεπτυσμένος αλγόριθμος θα οδηγούσε σε δειγματοληψία από μια λίστα τεσσάρων σωματιδίων, ένα από τα οποία θα απορριφόταν. Τέλος, στον αλγόριθμο Gillespie, η δειγματοληψία θα γινόταν από μια λίστα πέντε κινήσεων, οι οποίες θα ήταν όλες αποδεκτές.

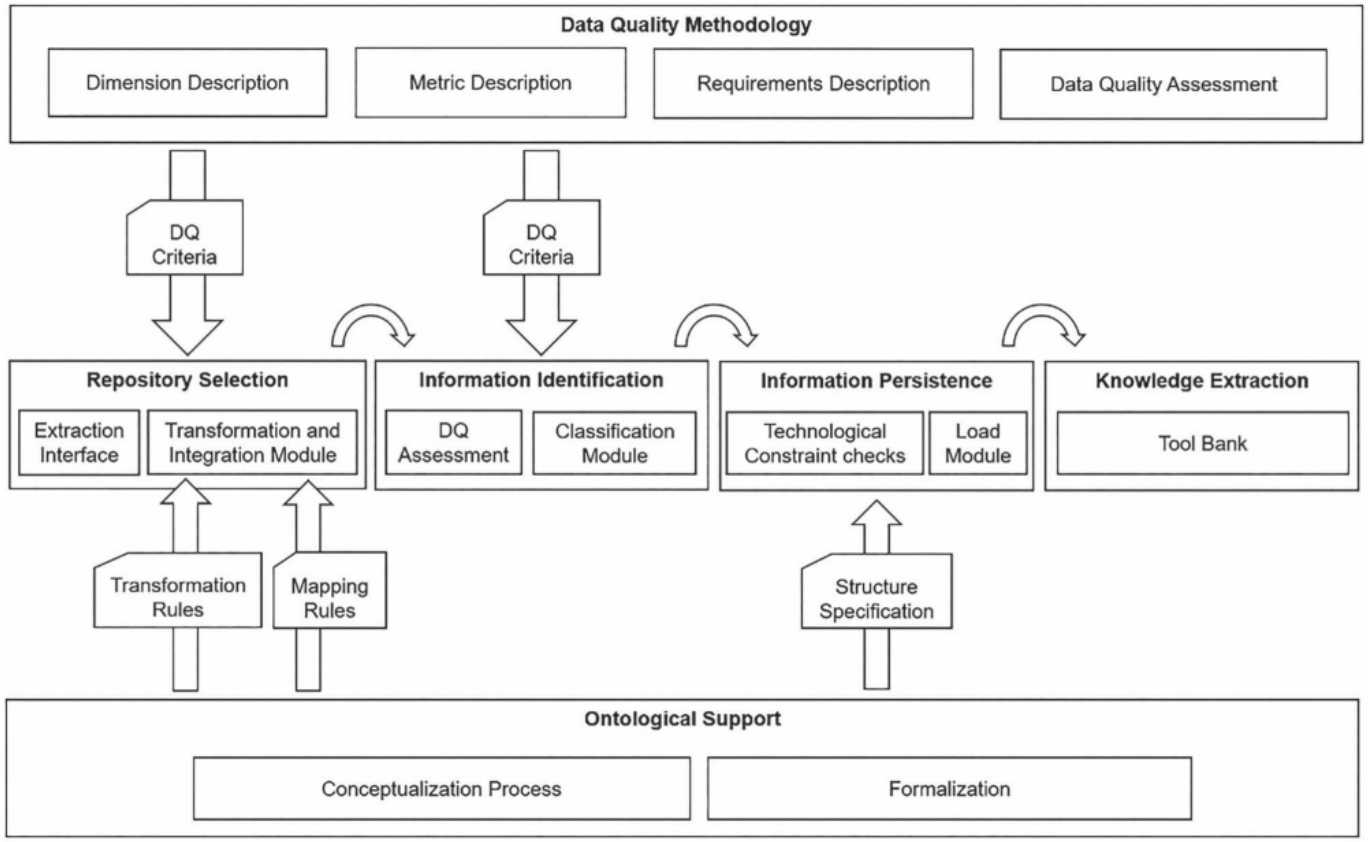
Η δεύτερη, και ακόμη πιο σημαντική διαφορά, έγκειται στον υπολογισμό του «πραγματικού» χρόνου. Σε αντίθεση με τους κανονικούς αλγόριθμους Monte Carlo, στους οποίους κάθε βήμα διαρκεί μία μονάδα χρόνου, όταν χρησιμοποιείται ο αλγόριθμος Gillespie, σε κάθε βήμα υπολογίζεται όχι μόνο η αλλαγή στο σύστημα, αλλά και ο χρόνος που χρειάζεται. Αυτή η τιμή προσδιορίζεται τυχαία, αλλά το διάστημα από το οποίο δειγματοληπτείται καθορίζεται χρησιμοποιώντας γνωστές παραμέτρους που συνδέονται με τη δυναμική του συστήματος. Το αποτέλεσμα είναι ότι, κατά μέσο όρο, ο «πραγματικός» χρόνος μέσα στην προσομοίωση αντιστοιχεί στον χρόνο που περνά στο βιολογικό σύστημα. Αυτό είναι ένα σημαντικό πλεονέκτημα, καθώς βελτιώνει τη συγκρισιμότητα μεταξύ της προσομοίωσης και των αντίστοιχων πειραματικών δεδομένων.

Co-funded by the Erasmus+ Programme of the European Union

Εννοιολογική μοντελοποίηση

Με την πρόοδο της τεχνολογίας γονιδιωματικής αλληλουχίας, μεγάλος όγκος δεδομένων είναι δημόσια διαθέσιμος για την ερευνητική κοινότητα για την εξαγωγή σημαντικών και αξιόπιστων συσχετισμών μεταξύ των γονιδίων κινδύνου και των μηχανισμών της νόσου. Ωστόσο, αυτή η εκθετική η ανάπτυξη των δεδομένων κατανέμεται σε πάνω από χιλιάδες ετερογενή αποθετήρια, που αντιπροσωπεύονται σε πολλαπλές μορφές και με διαφορετικές επίπεδα ποιότητας που εμποδίζει τη διαφοροποίηση των κλινικά έγκυρων σχέσεων από εκείνες που είναι λιγότερο καλά διατηρημένες και που θα μπορούσε να οδηγήσει σε λάθος διάγνωση. Υπάρχει ένα εννοιολογικό μοντέλο που παίζει βασικό ρόλο στην αποτελεσματική διαχείριση γονιδιωματικών δεδομένων. Αυτά τα δεδομένα πρέπει να είναι προσβάσιμα, ενημερωτικά και αρκετά αξιόπιστα ώστε να εξάγουν πολύτιμες γνώσεις στο πλαίσιο της ταυτοποίησης στοιχείων που υποστηρίζουν τη σχέση μεταξύ παραλλαγών DNA και ασθένειας.

Πλαίσιο για τη διαχείριση γονιδιωματικών πληροφοριών

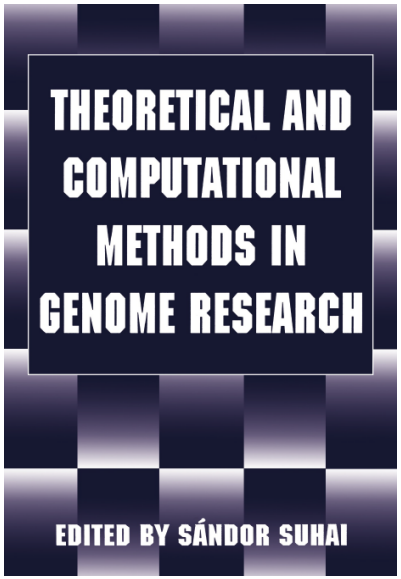


Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε τις μεθόδους μοντελοποίησης στη γονιδιωματική.
- Εξηγήστε το πλαίσιο για τη διαχείριση των γονιδιωματικών πληροφοριών.

Co-funded by the Erasmus+ Programme of the European Union



Sandor Suhai (1997) Θεωρητική και Υπολογιστική Έρευνα Γονιδιώματος, Springer Science+Business Media Νέα Υόρκη

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 2. Θέμα Γενετικής

2. Θεωρητική Μοντελοποίηση στη Γονιδιωματική

Μάθημα 2. Συστηματική ανάλυση των λειτουργιών των γονιδίων από τη βάση δεδομένων Metabolic Pathway

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Κυότο Εγκυκλοπαίδεια Γονιδίων και Γονιδιωμάτων (KEGG)
- Ανακατασκευή μονοπατιών ειδικών για τον οργανισμό
- Σύγκριση βιολογικών δικτύων
- Βάσεις δεδομένων μεταβολικών μονοπατιών

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Τα έργα αλληλουχίας γονιδιώματος διαφορετικών οργανισμών παράγουν γρήγορα καταλόγους γονιδίων και γονιδιακών προϊόντων. Το επόμενο προφανές βήμα είναι να κατανοήσουμε τις λειτουργικές επιπτώσεις, δηλαδή να αποκρυπτογραφήσουμε τόσο πειραματικά όσο και υπολογιστικά πότε, πού και πώς λειτουργούν τα γονίδια και τα μόρια στους ζωντανούς οργανισμούς. Στην πραγματικότητα, οι γνώσεις μας σχετικά με τη λειτουργία των γονιδίων και των μορίων επεκτείνονται επίσης γρήγορα λόγω της προόδου των πειραματικών τεχνολογιών σε ευρείες περιοχές της μοριακής και κυτταρικής βιολογίας. Προκειμένου να γίνει πλήρης χρήση των πληροφοριών που λαμβάνονται από τα έργα γονιδιώματος, είναι σημαντικό αυτά τα λειτουργικά δεδομένα να είναι κατάλληλα ηλεκτρονικά.

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

	LIGAND	KEGG	SwissProt	PIR	PDB
Oxidoreductase (EC1)	899	360	303	327	59
Transferase (EC2)	1002	306	332	319	47
Hydrolase (EC3)	992	172	428	370	107
Lyase (EC4)	332	160	129	126	21
Isomerase (EC5)	145	66	54	54	13
Ligase (EC6)	119	59	68	61	13
Total	3489	1123	1314	1257	260

LIGAND: enzymatic reactions
KEGG: metabolic pathways
SwissProt and PIR: amino acid sequences
PDB: 3D structures

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

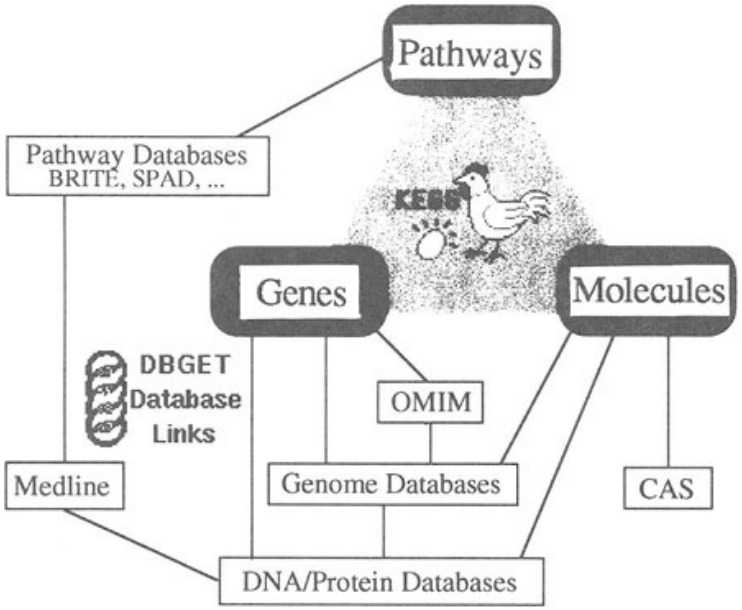
Κυότο Εγκυκλοπαίδεια Γονιδίων και Γονιδιωμάτων (KEGG)

Το KEGG (Kyoto Encyclopedia of Genes and Genomes) είναι η προσπάθειά μας να μηχανογραφήσουμε γνωστά μονοπάτια και να τα συσχετίσουμε με καταλόγους γονιδίων. Το KEGG έχει κατασκευαστεί για χρήση στον Παγκόσμιο Ιστό (WWW) και μια προκαταρκτική έκδοση διατίθεται δημόσια μέσω του διακομιστή GenomeNet WWW. Το KEGG αποτελείται από τρεις τύπους δεδομένων:

Γονίδια - ιεραρχικά δεδομένα κειμένου

Μόρια - ιεραρχικά δεδομένα κειμένου

Διαδρομές - δεδομένα γραφικών



Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Επαγωγική Βάση Δεδομένων για Μεταβολικά Μονοπάτια

Προκειμένου να υπολογιστούν πιθανές διαδρομές από μια δεδομένη λίστα ενζύμων, είναι δυνατό να οργανωθεί μια απαγωγική βάση δεδομένων χρησιμοποιώντας το CORAL. Ας υποθέσουμε ότι το ένζυμο Ε καταλύει μια χημική αντίδραση με το υπόστρωμα Χ και το προϊόν Υ.

αντίδραση (E, X, Y).

Όταν η μετατροπή της ένωσης Χ στην ένωση Υ είναι μια διαδικασία πολλαπλών σταδίων που αποτελείται από έναν αριθμό ενζύμων, η ενζυματική οδός αντιπροσωπεύεται από:

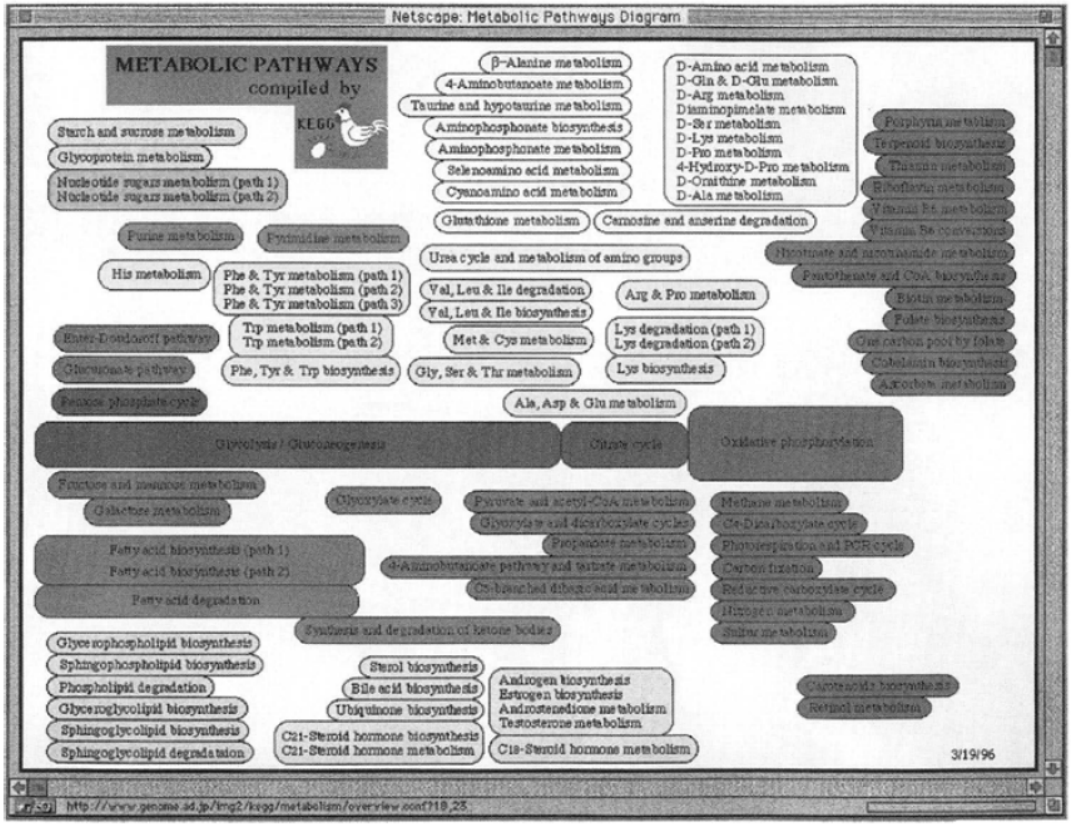
διαδρομή(X, Y, [E]) ← αντίδραση (E, X, Y).

διαδρομή(X, Y, [EI ELI) ← αντίδραση (E, X, Z), path(Z, Y, EL).

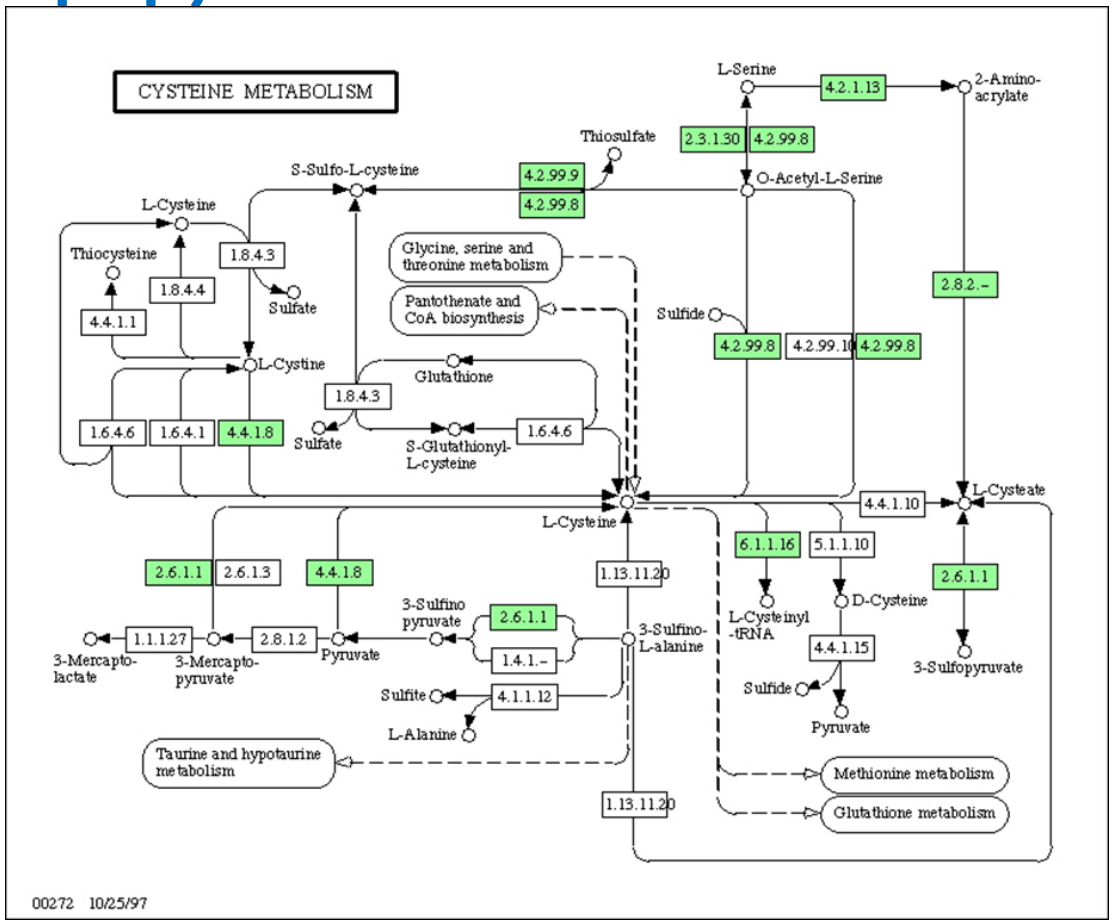
Με δεδομένο έναν κατάλογο γονιδίων ενζύμων, μπορούμε να υπολογίσουμε πιθανές οδούς που καταλύονται από τα γονιδιακά προϊόντα. Έτσι, για παράδειγμα, η ορθότητα της ταυτοποίησης γονιδίου στο έργο προσδιορισμού αλληλουχίας γονιδιώματος μπορεί να ελεγχθεί σε σχέση με τον βαθμό πληρότητας των παραγόμενων οδών. Είναι επίσης δυνατή η σύγκριση διαγραμμάτων μονοπατιών και η αναζήτηση για τοπικές ομοιότητες, εκτός από την ανάλυση της αλληλουχίας και των δομικών ομοιοτήτων των ενζύμων στα διαγράμματα μονοπατιών.

Co-funded by the Erasmus+ Programme of the European Union

Καταχώρηση για επιλογή διαγραμμάτων μεταβολικής οδού στο KEGG.



Ένα παράδειγμα του διαγράμματος διαδρομής KEGG

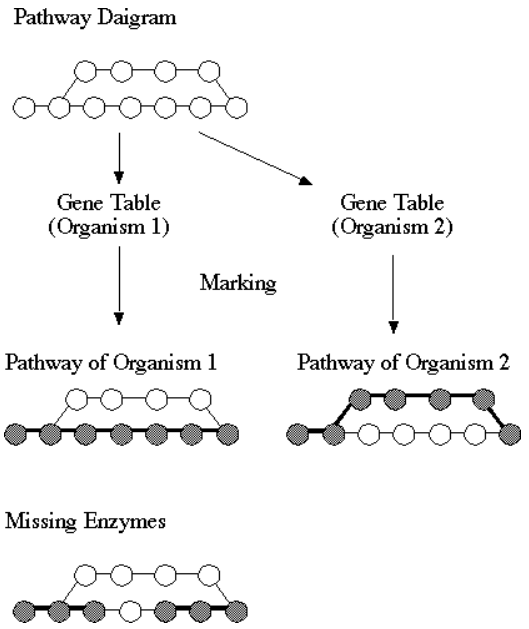


Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ανακατασκευή μονοπατιών ειδικών για τον οργανισμό

Δημιουργείται αυτόματα μια οδός ειδική για τον οργανισμό αντιστοιχίζοντας το διάγραμμα διαδρομής αναφοράς και τον κατάλογο γονιδίων σύμφωνα με τον αριθμό EC. Όταν το γονίδιο για ένα ένζυμο υπάρχει στον κατάλογο γονιδίων, το πλαίσιο που αντιπροσωπεύει το αντίστοιχο ένζυμο επισημαίνεται με χρώμα στην οδό. Η διαδοχική εμφάνιση των χρωματιστών πλαισίων θα θεωρηθεί τότε μια οδός συγκεκριμένη για τον οργανισμό. Για να είναι επιτυχής αυτή η διαδικασία, το διάγραμμα αναφοράς θα πρέπει να περιέχει όλες τις γνωστές εναλλακτικές οδούς αντίδρασης και όχι μόνο τη συναίνεση. Στην πραγματικότητα, μαθαίνουμε από τα πλήρη γονιδιώματα ότι διαφορετικοί οργανισμοί έχουν διαφορετικά σύνολα μεταβολικών οδών που αντικατοπτρίζουν το περιβάλλον διαβίωσης και τη στρατηγική της προσαρμογής. Η αυτόματη ανακατασκευή των μονοπατιών που σχετίζονται με τον οργανισμό είναι ένα πρώτο βήμα προς τη δημιουργία ενός καλά επαληθευμένου συνόλου διαγραμμάτων αναφοράς.



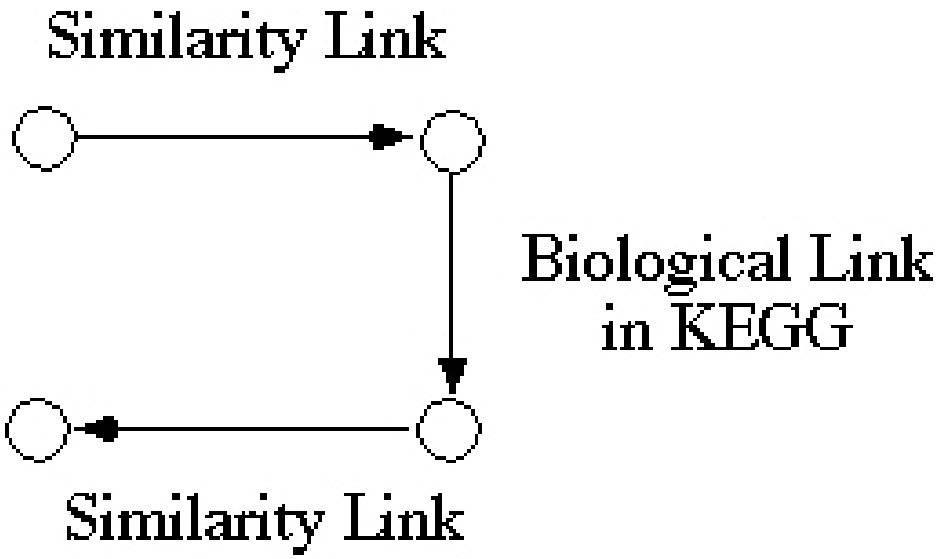
Μια σχηματική απεικόνιση των μονοπατιών δημιουργίας συγκεκριμένων οργανισμών

Υπολογιστικές διαδρομές συλλογισμού

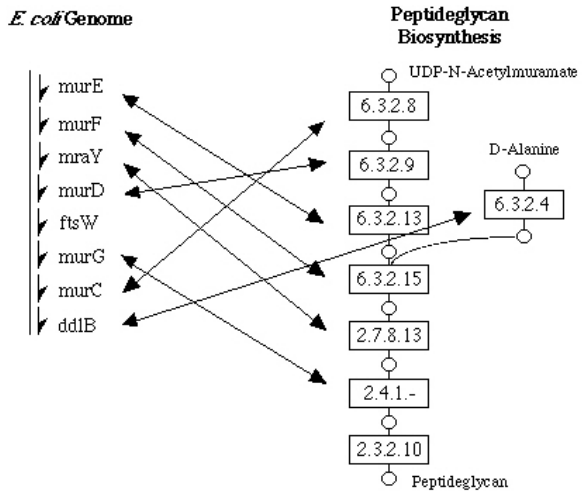
Τύπος	Παράδειγμα
πραγματικός σύνδεσμος	παραπομπές σε βάσεις δεδομένων δεσμούς μεταξύ γονιδίων και λειτουργιών
σύνδεσμος ομοιότητας	ομολογία ακολουθίας (ορθολογία / παραλογία) Τρισδιάστατη ομοιότητα και συμπληρωματικότητα
βιολογικός σύνδεσμος	σχέσεις υποστρώματος-προϊόντος σε ενζυμικές αντιδράσεις αλληλεπιδρώντα μόρια σε ένα κύτταρο γειτονικά γονίδια στο γονιδίωμα

Co-funded by the Erasmus+ Programme of the European Union

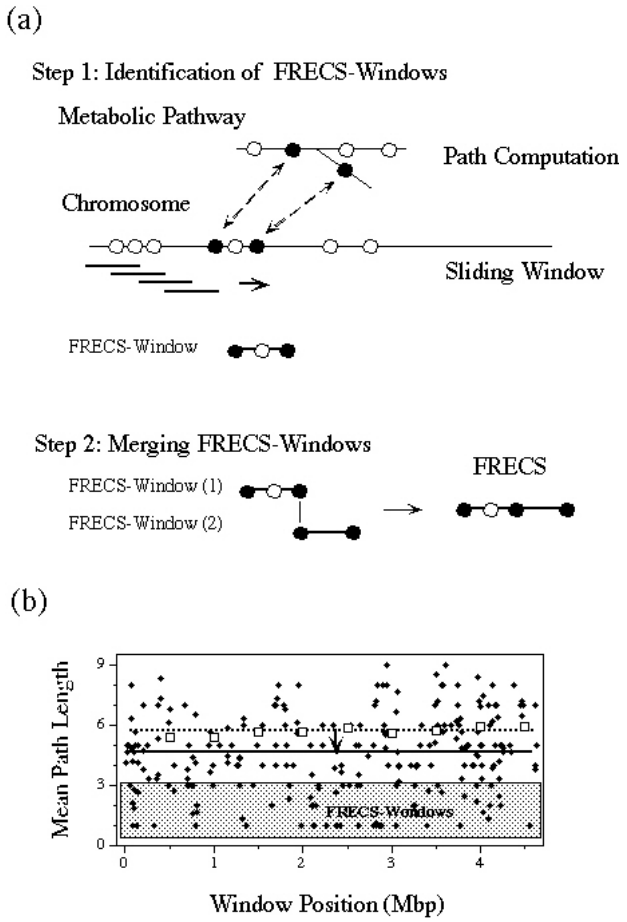
Ένα παράδειγμα των μονοπατιών του συλλογισμού που μπορούν να υπολογιστούν με συνδυασμό διαφόρων ειδών συνδέσμων.



Οι χρωμοσωμικές θέσεις των γονιδίων είναι επίσης ενημερωτικές για την αποσαφήνιση της ρύθμισης ή της εξέλιξης του μεταβολισμού σε ζυμομύκητες και βακτήρια. Στο KEGG η θέση και η σειρά των ενζυμικών γονιδίων στο γονιδίωμα μπορούν να εξεταστούν γραφικά, καθώς και η θέση και η σειρά των ενζύμων στο μονοπάτι. Το παράδειγμα στο σχήμα δείχνει ότι ένα σύνολο ενζύμων που κωδικοποιούνται σε κοντινές θέσεις κατά μήκος του γονιδιώματος του E. coli σχηματίζει ένα μπλοκ μονοπατιών αντίδρασης στη μεταβολική οδό, η οποία είναι στην πραγματικότητα μια γενική τάση όπως περιγράφεται παρακάτω. Αν και το χαρακτηριστικό δεν παρέχεται ακόμη, η συσχέτιση θέσης στο γονιδίωμα, η οποία αντιπροσωπεύεται επίσης από δυαδικές σχέσεις, θα συμπεριληφθεί στην αυτόματη βιολογική διαδικασία συλλογισμού στο KEGG.



Σύγκριση βιολογικών δικτύων



(α) Ένα λειτουργικά σχετιζόμενο τμήμα κωδικοποίησης ενζύμων (FRECS) ορίζεται από τα ζεύγη ενζύμων που εμφανίζονται μέσα σε ένα συρόμενο παράθυρο στο γονιδίωμα και που διαχωρίζονται από δεδομένα μήκη διαδρομής στη μεταβολική οδό (βλ. κείμενο). (β) Το μέσο μήκος διαδρομής των ζευγών ενζύμων απεικονίζεται στη θέση του παραθύρου για δύο διαφορετικά μεγέθη παραθύρου: 1 Mbp (λευκά πλαίσια) και 10 kbp (μαύρα κουτιά). Οι διακεκομμένες και συμπαγείς γραμμές δείχνουν τα μέσα επίπεδα του μήκους διαδρομής για παράθυρα 1 Mbp και 10 kbp , αντίστοιχα, και τοποθετείται ένα βέλος για να υποδείξει τη διαφορά των επιπέδων. Το FRECS αντιστοιχεί, στην περίπτωση αυτή, στο μήκος διαδρομής 3 ή λιγότερο

Βάσεις δεδομένων μεταβολικών μονοπατιών

Database/content	Reactions	Compounds	Pathways	Proteins	Version
KEGG	10307	17787	474838	6836	80.2
MetaCyc	15162	13585	2884	17090	20.5
Reactome	78927	88597	22072	120935	59
PMN	74013	2509	13117	279301	11.0

Αν και ποικίλλουν ως προς το εύρος τους, οι βάσεις δεδομένων μεταβολισμού θα πρέπει να είναι σε θέση να παρέχουν αξιόπιστες πληροφορίες για τέσσερα από τα βασικά συστατικά των μεταβολικών δικτύων. Αυτά περιλαμβάνουν βιοχημικές αντιδράσεις, ένζυμα που καταλύουν τις αντιδράσεις, μονοπάτια και μεταβολίτες. Οι μεταβολίτες θα πρέπει να αντιπροσωπεύονται στις κατάλληλες φορτισμένες ή ουδέτερες καταστάσεις στις βάσεις δεδομένων. Ομοίως, οι βάσεις δεδομένων θα πρέπει να ενσωματώνουν πληροφορίες σχετικά με την κατεύθυνση της αντίδρασης, το στοιχειακό ισοζύγιο και το φορτίο και το διαμέρισμα στο οποίο ανήκουν η αντίδραση και οι ενώσεις.

Reactome

Το Reactome είναι δημόσια προσβάσιμη, ανοιχτού κώδικα, μη αυτόματα επιμελημένη και αξιολογημένη βάση δεδομένων ανθρώπινων οδών. Ο πρωταρχικός στόχος του Reactome είναι να παρέχει μοριακές λεπτομέρειες μεταγωγής σήματος, μεταφοράς, μεταβολισμού, αντιγραφής DNA και άλλων κυτταρικών διεργασιών ως διατεταγμένο δίκτυο μοριακών μετασχηματισμών του Homo sapiens. Η τελευταία έκδοση περιλαμβάνει οργανισμούς από άλλους τομείς της ζωής. Οι χρήστες μπορούν να εκμεταλλευτούν τρία σημαντικά χαρακτηριστικά του Reactome . Μπορούν να περιηγηθούν, να οπτικοποιήσουν και να αναλύσουν αντιδράσεις, μεταβολίτες και ένζυμα σε κάθε μονοπάτι και να κατεβάσουν τις πληροφορίες σε διάφορες μορφές.

Co-funded by the Erasmus+ Programme of the European Union

MetaCyc

Το MetaCyc είναι μια εξαιρετικά επιμελημένη βάση δεδομένων με πειραματικά επικυρωμένα μεταβολικά μονοπάτια από όλους τους τομείς της ζωής. Περιέχει μονοπάτια που προέρχονται από εκτενώς μεγάλο αριθμό πρωτογενούς βιβλιογραφίας. Το MetaCyc στοχεύει να χρησιμεύσει ως μια γενική βάση δεδομένων αναφοράς για το μεταβολισμό. Ένα σημαντικό χαρακτηριστικό του MetaCyc είναι το εργαλείο διαδρομής του που μπορεί να χρησιμοποιηθεί για την υπολογιστική πρόβλεψη μοντέλων μεταβολικού δικτύου οποιουδήποτε οργανισμού από ένα γονιδίωμα αλληλουχίας.

Co-funded by the Erasmus+ Programme of the European Union

PMN

Το δίκτυο μεταβολισμού φυτών είναι μια βάση δεδομένων μεταβολικών οδών που φιλοξενεί μία βάση δεδομένων αναφοράς (PlantCyc) και 22 βάσεις δεδομένων για συγκεκριμένα είδη. Όλοι τους εστιάζουν ιδιαίτερα στις μεταβολικές οδούς των φυτών. Στο κέντρο του PMN βρίσκεται το PlantCyc , το οποίο είναι μια βάση δεδομένων αναφοράς μεταβολικής οδού που περιέχει περισσότερα από 900 μονοπάτια, τα καταλυτικά ένζυμα και τα γονίδιά τους. Επιπλέον, το PlantCyc περιέχει ενώσεις από περισσότερα από 350 είδη φυτών. Η πηγή δεδομένων στο PlantCyc καλύπτει μονοπάτια από πειραματικά επικυρωμένη βιβλιογραφία και την επιμέλεια του PMN και των συνεργατών του.

Σύγκριση των βάσεων δεδομένων με βάση πρόσθετες συναρτήσεις

Name	Tools	Function
Reactome	Pathway browser	• A tool for visualizing and interacting with Reactome biological pathways
	Analyze data	• Merges pathway identifier mapping, over-representation, and expression analysis tools into a single tabbed data analysis portal with integrated visualization and summary features, which can accept a gene, protein or small molecule list, or an expression dataset
	Species comparison	• Allows users to compare pathways between human and any of the other species inferred from Reactome by orthology
	Reactome FI network	• Cytoscape plugin designed to find network patterns related to cancer and other types of diseases
MetaCyc	Pathway tools	• Development of organism-specific databases • Metabolic reconstruction and modeling • Scientific visualization, web publishing • Visual analysis of gene expression and metabolomics datasets • Computational inferences • Comparative genome and pathway analyses • Analysis of biological networks
PMN	E2P2	• Enzyme function annotation software. Predicts metabolic enzymes in a sequenced genome
	SAVI	• Pathway validation software. Processes predicted metabolic pathways using pathway metadata such as taxonomic distribution and key reactions and makes decisions about which pathways to keep, remove, or subject to manual validation
KEGG	PlantClusterFinder	• A pipeline to predict metabolic gene clusters from plant genomes
	KegHier	• Java application for browsing BRITE hierarchy files
	KegArray	• Java application for microarray data analysis
	KegDraw	• Java application for drawing compound and glycan structures

Co-funded by the Erasmus+ Programme of the European Union

Αποθετήρια μοντέλων

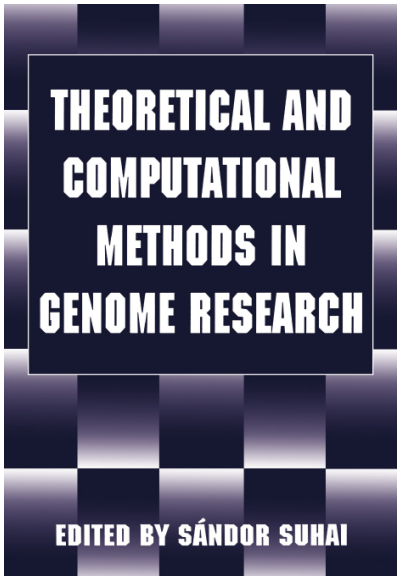
Δεδομένου ότι τα μοντέλα προέρχονται από διαφορετικές ερευνητικές ομάδες ή άτομα, μία από τις τρέχουσες προκλήσεις που σχετίζονται με τα μοντέλα μεταβολικών δικτύων είναι οι ασυνέπειες στην αναπαράσταση των συστατικών τους όπως η αντίδραση ή τα ονόματα των ενώσεων και οι τύποι και τα σύμβολα που χρησιμοποιούνται για τον προσδιορισμό των κυτταρικών διαμερισμάτων. Ως εκ τούτου, τα διαδικτυακά αποθετήρια μοντέλων θα πρέπει να υπερβαίνουν το να χρησιμεύουν απλώς ως αποθετήριο σε πλατφόρμες συστηματικής ανάλυσης και τυποποίησης αυτών των μοντέλων, οι προσπάθειες του BiGG και του MetaNetX.org μπορούν να αναφερθούν από αυτή την άποψη. Επιπλέον, τα αποθετήρια μοντέλων αναμένεται να παρέχουν συνοπτικές περιγραφές των μοντέλων, όπως συνθήκες προσομοίωσης, που μπορεί να περιλαμβάνουν την κατάσταση του μέσου ανάπτυξης in-silico, τις ενεργειακές απαιτήσεις, τη λειτουργία αντικειμενικού και τις πληροφορίες συμβατότητας αρχείων. Έχοντας αυτό υπόψη, η ακόλουθη ενότητα θα αξιολογήσει μερικές από τις πιο κοινές βάσεις δεδομένων μοντέλων μεταβολικών δικτύων.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε πώς θα μπορούσαν να ανακατασκευαστούν τα μονοπάτια που σχετίζονται με τον οργανισμό.
- Δώστε μερικά παραδείγματα βάσεων δεδομένων μεταβολικής οδού και ποιες πληροφορίες θα μπορούσαν να ληφθούν από αυτές.

Co-funded by the Erasmus+ Programme of the European Union



Sandor Suhai (1997) Θεωρητική και Υπολογιστική Έρευνα Γονιδιώματος, Springer Science+Business Media Νέα Υόρκη

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 2. Γενετική

Θέμα 2. Θεωρητική Μοντελοποίηση στη Γονιδιωματική

Μάθημα 3. Δυναμική πολυμερών του DNA, της χρωματίνης και των χρωμοσωμάτων

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Το γονιδίωμα ως ελαστική αλυσίδα
- Brownian Dynamics και προσομοιώσεις Monte-Carlo του Superhelical DNA
- Το μοντέλο της χρωματίνης

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Η χωρική οργάνωση του DNA στον ευκαρυωτικό πυρήνα παίζει εξέχοντα ρόλο στον προσδιορισμό της γονιδιακής δραστηριότητας. Στο επίπεδο της τοπικής δομής του DNA (περίπου 100-1000 bp), η δράση των μεταγραφικών παραγόντων συχνά μεσολαβείται από τον βρόχο DNA μεταξύ της ρυθμιστικής θέσης και του προαγωγέα, και η δομή του παρεμβαλλόμενου DNA έχει αποφασιστική επίδραση στην ισχύ της αλληλεπίδρασης μεταξύ του μεταγραφικού παράγοντα και του συμπλέγματος έναρξης πολυμεράσης RNA. Σε υψηλότερο επίπεδο οργάνωσης, έχει περιγραφεί ένας μεγάλος αριθμός περιπτώσεων όπου η έκφραση ενός γονιδίου καθορίζεται από τη συσκευασία του DNA του σε μια περισσότερο ή λιγότερο συμπαγή δομή χρωματίνης.

Co-funded by the Erasmus+ Programme of the European Union

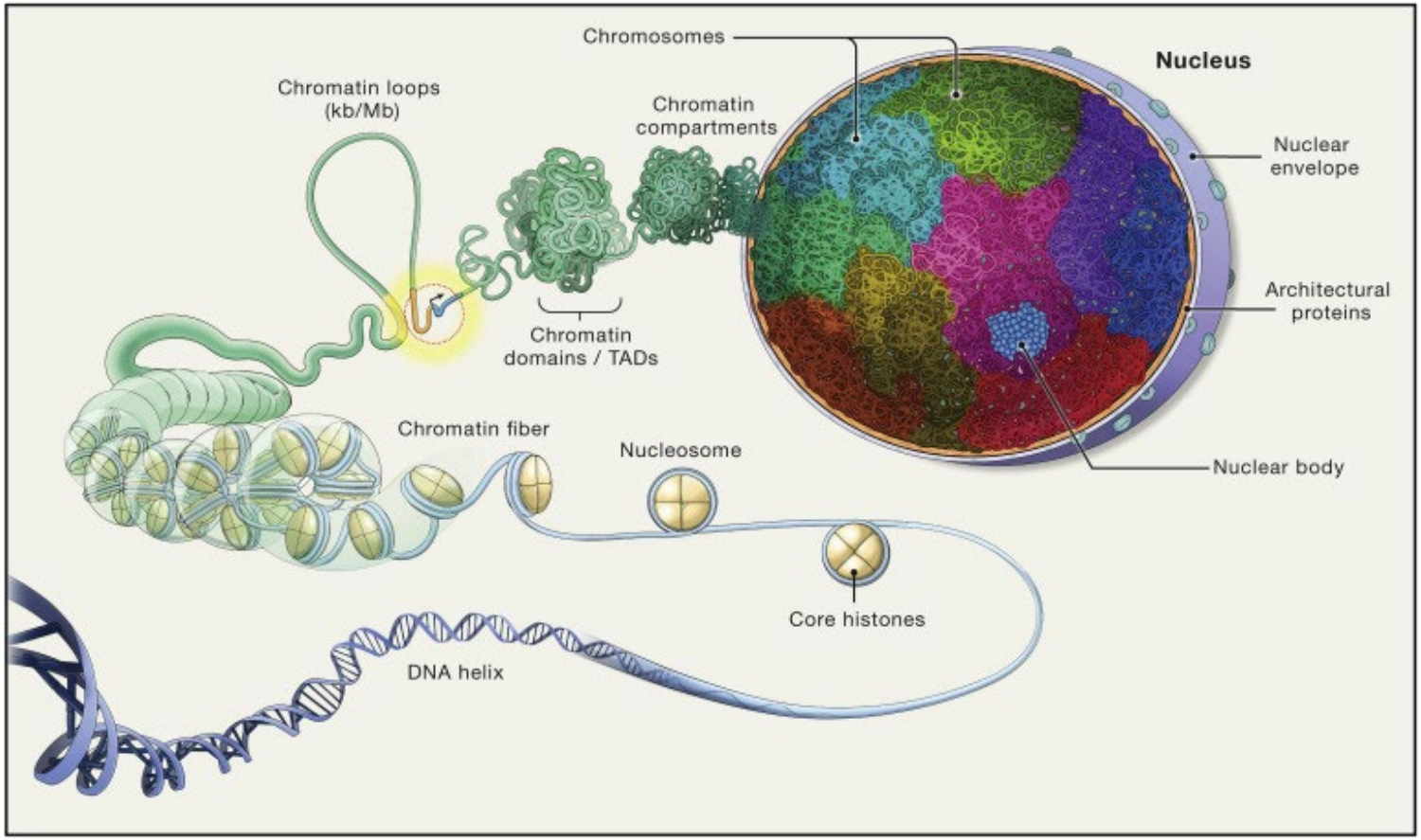
Το γονιδίωμα ως ελαστική αλυσίδα

Η χονδροειδής άποψη του γονιδιώματος ως ελαστικού νήματος είναι επαρκής εφόσον ενδιαφέρεται για ερωτήματα όπως η κινητική προσέγγισης απομακρυσμένων κομματιών DNA, οι δομικές διακυμάνσεις ενός υπερπηνίου ή μιας ίνας χρωματίνης ή η τρισδιάστατη οργάνωση ενός ολόκληρου χρωμοσώματος. Τέτοιες διεργασίες συμβαίνουν σε κλίμακες μήκους πολύ μεγαλύτερες από τις διατομικές αποστάσεις, τυπικά > 100 m, και σε κλίμακες χρόνου πολύ πιο αργές από τις μοριακές δονήσεις, τυπικά > I ns. Έτσι, η ατομική λεπτομέρεια μπορεί να παραμεληθεί με ασφάλεια.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Το γονιδίωμα ως ελαστική αλυσίδα



Co-funded by the Erasmus+ Programme of the European Union

Το γονιδίωμα ως ελαστική αλυσίδα

Για την περιγραφή των κινήσεων μεγάλης κλίμακας ενός μακρού μορίου DNA (πολλές εκατοντάδες ή χιλιάδες ζεύγη βάσεων), η προσέγγιση του ελαστικού νήματος είναι επίσης απαραίτητη: οι υπολογισμοί ατομικού επιπέδου σε ένα τόσο μεγάλο σύστημα, συμπεριλαμβανομένων των μορίων νερού διαλύτη, είναι αδιανόητοι με τον σύγχρονο υπολογιστικό εξοπλισμό.

Το γονιδίωμα ως ελαστική αλυσίδα

Για την περιγραφή των κινήσεων μεγάλης κλίμακας ενός μακρού μορίου DNA (πολλές εκατοντάδες ή χιλιάδες ζεύγη βάσεων), η προσέγγιση του ελαστικού νήματος είναι επίσης απαραίτητη: οι υπολογισμοί ατομικού επιπέδου σε ένα τόσο μεγάλο σύστημα, συμπεριλαμβανομένων των μορίων νερού διαλύτη, είναι αδιανόητοι με το σημερινό υπολογιστικό εξοπλισμό.

- Ελαστικές Παράμετροι
- Τοπολογικοί Περιορισμοί
- Ηλεκτροστατικές Αλληλεπιδράσεις
- Υδροδυναμικές Αλληλεπιδράσεις

Brownian Dynamics and Monte-Carlo Simulations of Superhelical DNA

Για τον υπολογισμό της δυναμικής του DNA, πρέπει να δημιουργηθούν εξισώσεις κίνησης για την τμηματοποιημένη αλυσίδα DNA χρησιμοποιώντας τα δυναμικά ενδομοριακής αλληλεπίδρασης που περιγράφονται παραπάνω και συμπεριλαμβανομένης της θερμικής κίνησης μέσω μιας τυχαίας δύναμης. Αυτή είναι η μέθοδος Brownian Dynamics (BD) την οποία πολλές ομάδες εφάρμοσαν στην ερμηνεία πειραματικών δεδομένων από την εκπόλωση φθορισμού, τη δυναμική σκέδαση φωτός ή την τριπλή ανισοτροπία. Το κυκλικό DNA πρόσφατα μοντελοποιήθηκε επίσης με μια προσέγγιση BD.

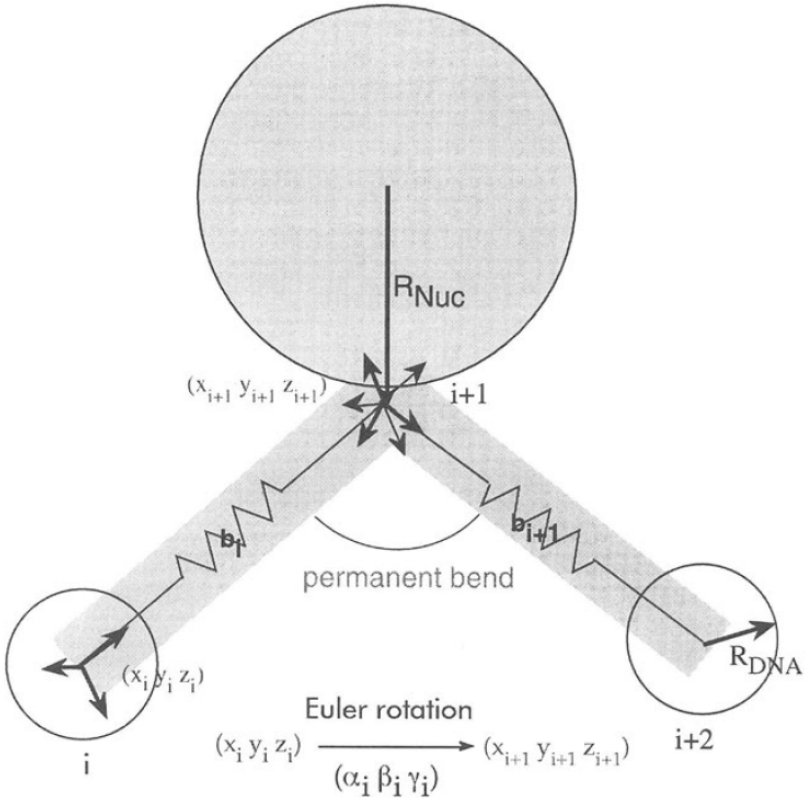
Το μοντέλο της χρωματίνης

Πρόσφατες αναθεωρήσεις των υφιστάμενων δομικών δεδομένων για τις ίνες χρωματίνης υποδηλώνουν μια τοπική διάταξη «ζιγκ-ζαγκ» των νουκλεοσωμάτων που αναδιπλώνεται σε μια δομή υψηλότερης τάξης τυχαίου περιπάτου. Η τυχαιότητα μπορεί να προκύψει είτε μέσω του μη ομοιόμορφου μήκους του συνδέτη DNA, προκαλώντας συστροφή του ενός νουκλεοσώματος σε σχέση με το άλλο, είτε λόγω θερμικών διακυμάνσεων.

Με βάση αυτές τις ιδέες, μπορεί κανείς να αναπτύξει ένα μοντέλο BD για τη δομή της χρωματίνης μέσω κάποιων απλών επεκτάσεων του μοντέλου DNA ελαστικής αλυσίδας.

Co-funded by the Erasmus+ Programme of the European Union

Το μοντέλο της χρωματίνης



Μοντέλο βρόχου RG των χρωμοσωμάτων ενδιάμεσης φάσης

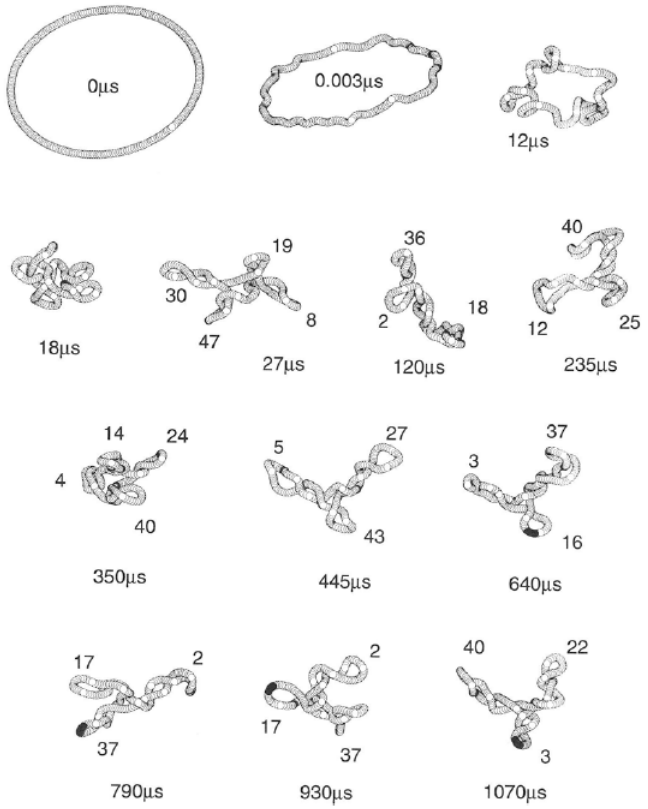
Το βασικό δομικό στοιχείο του μοντέλου των χρωμοσωμάτων είναι η ίνα χρωματίνης. Αν και η σαφής δομή των ινών χρωματίνης στους ανθρώπινους πυρήνες ενδιάμεσης φάσης είναι ακόμη υπό συζήτηση, οι μέσες γεωμετρικές ιδιότητές τους σε κλίμακες μήκους πάνω από μερικές χιλιάδες ζεύγη βάσεων μπορούν να περιγραφούν από ένα πολυμερές που αποτελείται από άκαμπτα τμήματα ορισμένου μήκους (μήκος Kuhn) και εξαιρούμενο όγκο αλληλεπίδραση που αντιπροσωπεύει τη διάμετρο της ίνας χρωματίνης. Με κίνητρο τις διαφορετικές ιδιότητες των ζωνών, π.χ. πυκνότητα γονιδίων ή χρόνο αντιγραφής, θεωρείται ότι υπάρχει διαφορετική οργάνωση των ζωνών χρωμοσωμάτων και στη μεσοφάση. Κάθε «ενεργοποιημένη» ζώνη (π.χ. μια ζώνη R πλούσια σε γονίδια) σχηματίζει έναν ενιαίο γιγάντιο βρόχο στην περιοχή 3-10 Mbp , ενώ η χρωματίνη εντός των «απενεργοποιημένων» ζωνών (π.χ. ζώνες G ή C) διπλώνεται σε αρκετούς μικρούς βρόχους περίπου 100 kbp το καθένα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

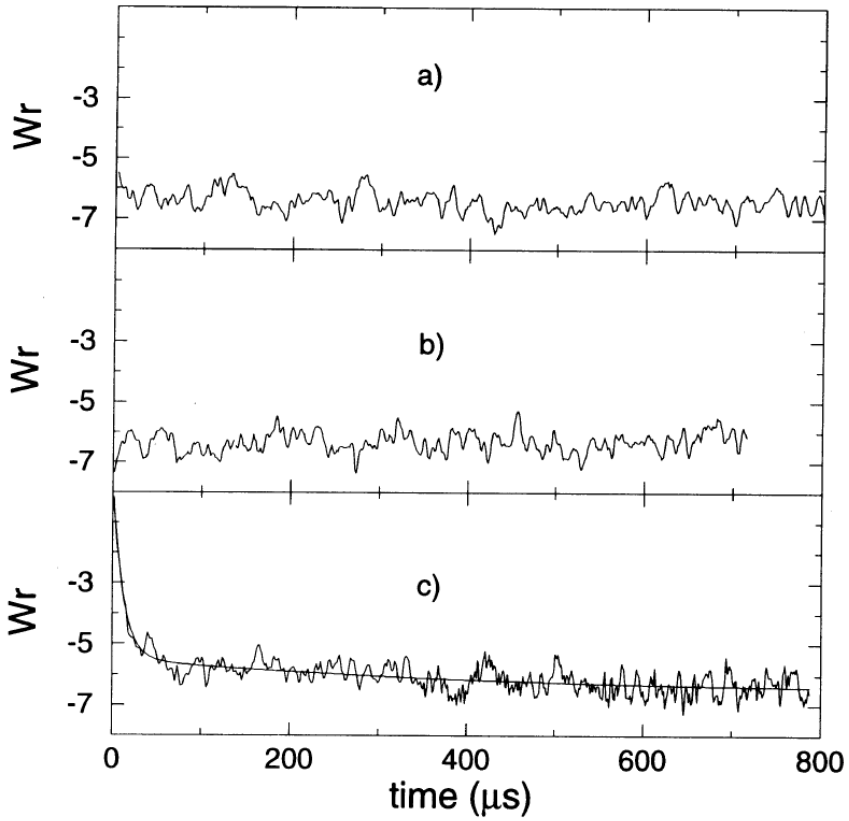
Co-funded by the Erasmus+ Programme of the European Union

Δομή και Δυναμική Υπερελικοειδούς DNA

Διαμορφώσεις από την τροχιά υπερέλιξης του ani 870 bp DNA με Δ Lk = -10 ξεκινώντας από επίπεδο κύκλο. Οι χρόνοι προσομοίωσης εκτυπώνονται δίπλα σε κάθε διαμόρφωση. Οι αριθμοί δηλώνουν τις θέσεις των τελικών βρόχων. Σε t = 450 μS μια μόνιμη κάμψη 100° εισήχθη στη θέση 15 (τμήματα σημειωμένα με μαύρο χρώμα).

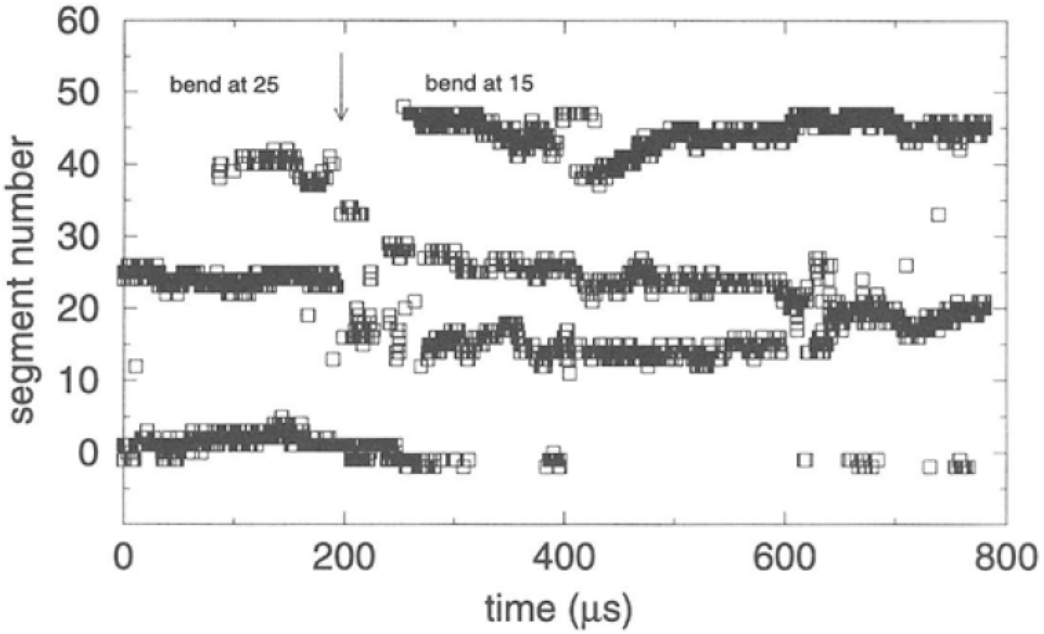


Κινητική συστροφής για τροχιές που ξεκινούν από έναν επίπεδο κύκλο διαμόρφωσης που δημιουργήθηκε από το Μόντε Κάρλο (a), μια ευθύγραμμη διατυλιγμένη δομή (b) και έναν επίπεδο κύκλο (c). Η ευθεία στο c) είναι διπλή εκθετική προσαρμογή στην κινητική καμπύλη.



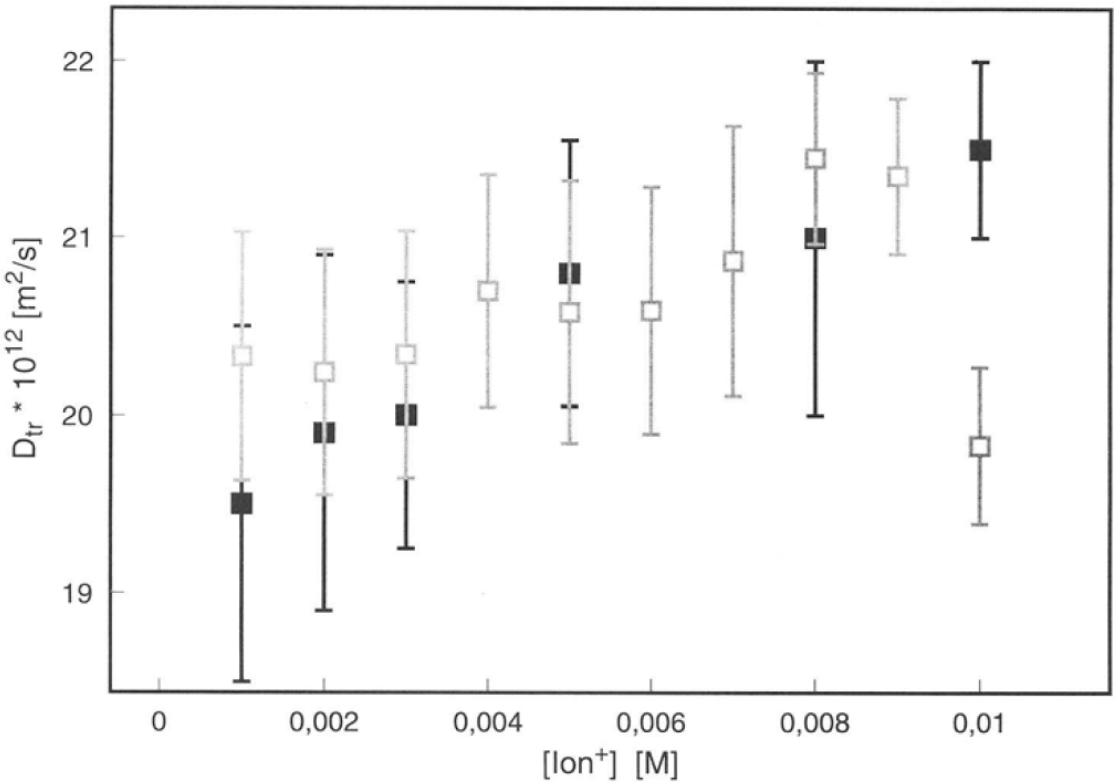
Co-funded by the Erasmus+ Programme of the European Union

Μοντελοποίηση χρωματίνης



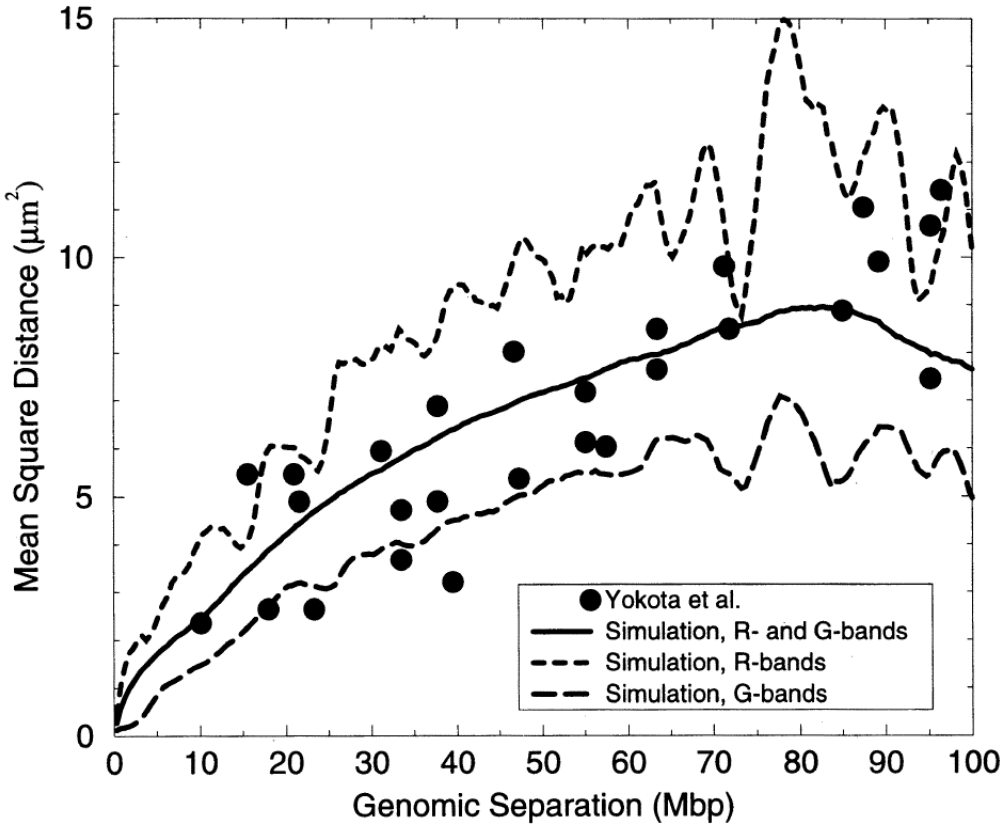
Θέσεις τελικού βρόχου σε σχέση με το χρόνο για μια τροχιά που ξεκινά από μια διατυλιγμένη υπερέλικα DNA 1870 bp με Δ Lk = -10 και μια μόνιμη κάμψη εισήχθη στη θέση 25. Σε t=200 μS, η κάμψη μετατοπίστηκε στη θέση 15

Οργάνωση χρωμοσωμάτων

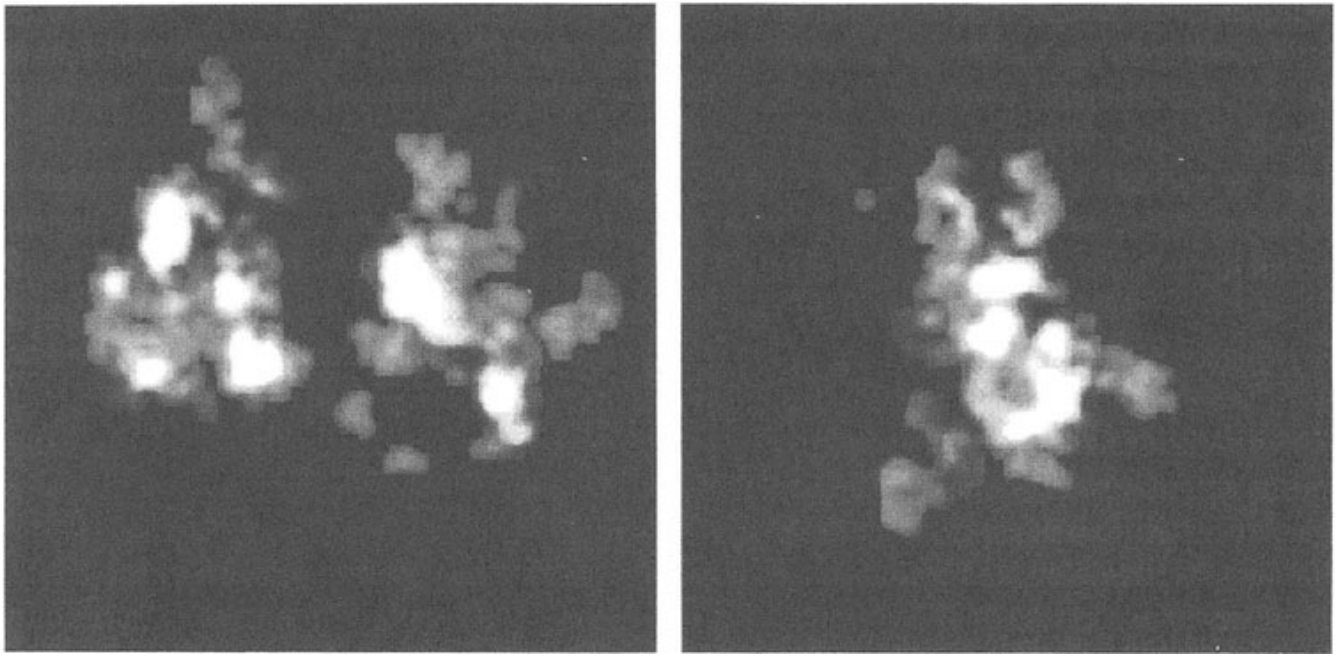


Οι συντελεστές διάχυσης των δινουκλεοσωμάτων όπως μετρήθηκαν από τους Yao et al. ως συνάρτηση της ιοντικής ισχύος (■). (□) Το D υπολογίζεται από προσομοιωμένη τροχιά.

Χωρικές αποστάσεις γενετικών δεικτών ως συνάρτηση του διαχωρισμού τους σε Mbp σε δισδιάστατες προβολές πυρήνων μεσοφάσεως. (•) πειραματικά δεδομένα από Yokota et al ., συμπαγής γραμμή: δεδομένα μοντέλου, κατά μέσο όρο σε ζώνες R και G , διακεκομμένες γραμμές: μόνο ζώνες R (επάνω), μόνο ζώνες G (κάτω).



Co-funded by the Erasmus+ Programme of the European Union



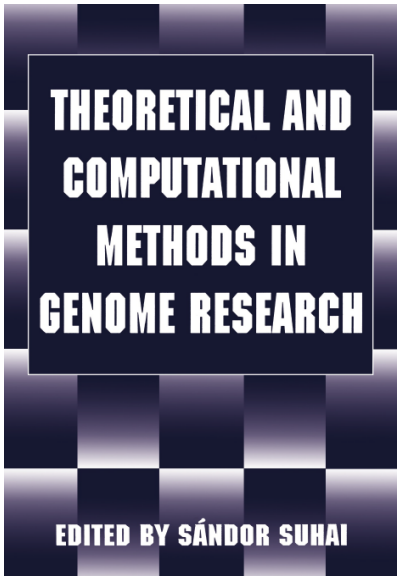
Προσομοιωμένη τρισδιάστατη ομοεστιακή τομή των βραχιόνων p- (αριστερά) και q- (δεξιά) του ανθρώπινου χρωμοσώματος 3. Ο βραχίονας p εμφανίζεται σε δύο μέρη, καθώς το κεντρικό τμήμα βρίσκεται πίσω από το επίπεδο της εικόνας. Ο χωρισμός των δύο βραχιόνων
οι περιοχές που προβλέπονται από το μοντέλο είναι σαφώς ορατές.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε το μοντέλο του γονιδιώματος ως ελαστική αλυσίδα.
- Εξηγήστε το μοντέλο της χρωματίνης.

Co-funded by the Erasmus+ Programme of the European Union



Sandor Suhai (1997) Θεωρητική και Υπολογιστική Έρευνα Γονιδιώματος, Springer Science+Business Media Νέα Υόρκη

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 2. Γενετική

Θέμα 2. Θεωρητική Μοντελοποίηση στη Γονιδιωματική

Μάθημα 4. Μοτίβα ακολουθίας Διαγνωστικά Δομής και Λειτουργίας

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Μοτίβα Διαγνωστικά Δομής
- Κρυμμένο μοντέλο Markov
- Συνδυασμένα μοτίβα
- Αξιολόγηση Μοτίβου

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Υπάρχει επί του παρόντος ένας τεράστιος πλούτος δεδομένων αλληλουχίας αμινοξέων από πάρα πολλά διαφορετικά γονίδια και διαφορετικούς οργανισμούς. Αναμένεται ότι αυτός ο πλούτος θα συνεχίσει να αυξάνεται. Αυτά τα δεδομένα αντιπροσωπεύουν μεγάλο μέρος της τρέχουσας ακατέργαστης γνώσης μας για τα βιολογικά συστήματα στη γη. Ωστόσο, η ικανότητά μας να εκμεταλλευόμαστε αυτά τα δεδομένα εξακολουθεί να είναι περιορισμένη: πρώτον, από την αδυναμία μας να προβλέψουμε τις βιοχημικές λειτουργίες μιας πρωτεΐνης ή τη δομή που κωδικοποιεί αυτές τις λειτουργίες απευθείας από τη γνώση της αλληλουχίας αμινοξέων της. και δεύτερον, λόγω της περιορισμένης ικανότητάς μας να πραγματοποιούμε άμεσο πειραματικό προσδιορισμό είτε της συνάρτησης είτε της δομής. Είναι ενδιαφέρον ότι αυτή τη στιγμή είναι ευκολότερο να προβλεφθεί η λειτουργία παρά η δομή. Αυτό ισχύει ακόμα κι αν η γνώση της διπλωμένης δομής είναι απαραίτητη για την κατανόηση των λεπτομερών στερεοχημικών περιορισμών και της χημείας της πλευρικής αλυσίδας που καθορίζουν τη(τις) συνάρτηση(ες). Μόνο ένα μικρό κλάσμα γονιδίων είχε βιοχημικές ή κυτταρικές λειτουργίες που προσδιορίστηκαν άμεσα με πείραμα.

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Μέσω της αναγνώρισης ομοιοτήτων αλληλουχίας ή κοινών μοτίβων στοιχείων αλληλουχίας, μπορέσαμε να συμπεράνουμε τη λειτουργία δεκάδων χιλιάδων επιπλέον γονιδίων. Αυτό έχει αποδειχθεί μια ισχυρή προσέγγιση. Υποδεικνύοντας μια πιθανή συνάρτηση για μια πρόσφατα καθορισμένη ακολουθία, μόνο μερικές πειραματικές δοκιμές πρέπει να πραγματοποιηθούν για επιβεβαίωση. Σε ορισμένες περιπτώσεις, η απλή πρόταση που παρέχεται από ένα κοινό σύνολο κοινών στοιχείων δημιουργεί προφανή εικόνα σε προηγούμενες πειραματικές παρατηρήσεις καθιστώντας την επικύρωση εντελώς προφανής. Έτσι, η αναγνώριση της βιολογικά σημαντικής ομοιότητας αλληλουχίας έχει γίνει ο λίθος της Ροζέτας της σύγχρονης μοριακής βιολογίας. Τα εργαλεία που χρησιμοποιούνται εξαρτώνται από μεθοδολογίες που προέρχονται από τα μαθηματικά, την επιστήμη των υπολογιστών και τη γλωσσολογία καθώς και από τη φυσική, τη χημεία και την υπόλοιπη βιολογία. Τέλος, πιστεύεται ότι θα αναπτυχθούν ανάλογες προσεγγίσεις για την ακριβή πρόβλεψη της δομής της πρωτεΐνης.

Μοτίβα Διαγνωστικά Δομής

Τα πιο προφανή και απλά δομικά μοτίβα είναι αυτά που σχετίζονται με δευτερεύουσες δομές. Για παράδειγμα, ένα διάγραμμα υδροφοβικότητας κατά μήκος μιας αλληλουχίας πρωτεΐνης που δείχνει μια σαφή περίοδο 3,6, 2,0 ή μηδέν είναι ένα μοτίβο ενδεικτικό επιφανειακών αμφιπαθών ελίκων, επιφανειακών κλώνων ή μιας πλήρως θαμμένης περιοχής (συμπεριλαμβανομένων των διαμεμβρανικών τμημάτων). Τέτοια μοτίβα μπορούν να εκφραστούν ως κανονικές εκφράσεις (το χαμηλότερο σκαλί της επίσημης γλωσσικής ιεραρχίας του Chomsky) ή περιβαλλοντικά προφίλ. Το τελευταίο είναι μια αλληλουχία φορέων προτίμησης αμινοξέων, ένας που σχετίζεται με κάθε δομική θέση με μοτίβο.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Μοτίβα Διαγνωστικά Δομής

Consensus patterns

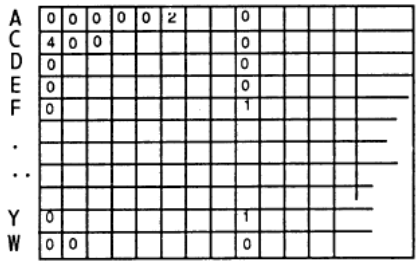
Only the conserved residues are represented:

```
CLVKWVYPFLWIDSK
CLVRWMYPKLPIKSK
CLIKWAYPYLWIESK
CLIKWAYPFLWIDSK
---------------
CL--W-YP-L-I-SK
```

Frequency Profiles

Matrix containing frequencies of occurance for each amino acid at each residue:

```
CLVKWVYPFLWIDSK
CLVRWMYPKLPIKSK
CLIKWAYPYLWIESK
CLIKWAYPFLWIDSK
```



Normally converted to a matrix of log likelihood ratios.

Regular Expressions

Using minimal amino acid covering or probable substitution classes:

```
CLVKWVYPFLWIDSK
CLVRWMYPKLPIKSK
CLIKWAYPYLWIESK
CLIKWAYPFLWIDSK
---------------
CLanWfYPxLxIrSK
```

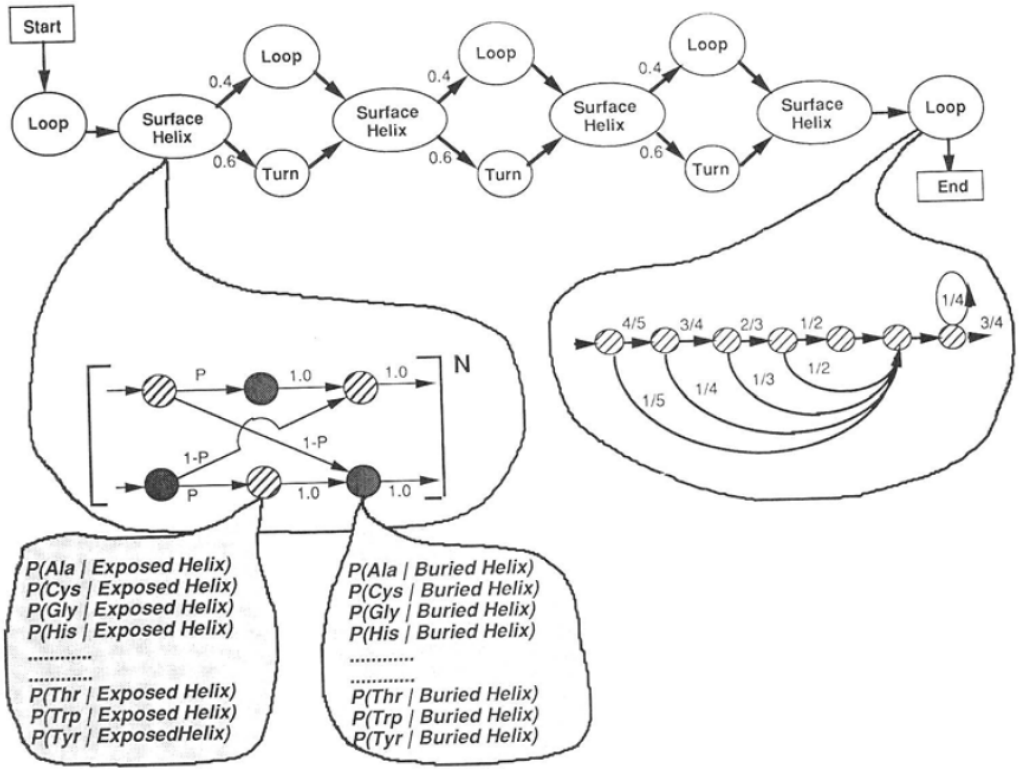
Converted to a UNIX *grep*-style regular expression statement

```
CL[ILV][KR]W[IVLMFWYCA]YP.L.I[DENQKRH]SK
```

Μοτίβα Διαγνωστικά Δομής

Ωστόσο, τα δομικά μοτίβα μπορεί να είναι πολύ πιο περίπλοκα. Μπορούν να είναι ιεραρχικά, αποτελούμενα από ένα μοτίβο δευτερευουσών δομών (ενδεικτικό μιας συγκεκριμένης κατηγορίας αναδίπλωσης 3-D) καθεμία από τις οποίες αποτελείται, επί της ουσίας, από ένα σχέδιο περιβαλλόντων θέσης (με σχετιζόμενες προτιμήσεις αμινοξέων ή/και υδροφοβικότητας). Για παράδειγμα, η κατηγορία των απλών παράλληλων δεσμίδων τεσσάρων έλικας θα έχει ένα μοτίβο που αποτελείται από τέσσερις σχεδόν ίσου μήκους αμφιπαθικές έλικες που χωρίζονται από τρία ή περισσότερα αμινοξέα συμβατά με βρόχο ή στομ. Κάθε έλικα in tum θα αντιπροσωπεύεται από ένα σχέδιο θαμμένων και εκτεθειμένων θέσεων, καθεμία από τις οποίες θα συσχετίζεται με έναν φορέα προτιμήσεων αμινοξέων συμβατών με την έλικα. Εδώ χρειάζεται γενικά να ανέβει κανείς στην ιεραρχία του Chomsky σε "γραμματικές ελεύθερες περιβάλλοντος" για να μπορέσει να επιβάλει τον "μη τοπικό" περιορισμό παρόμοιων μηκών μεταξύ των τεσσάρων έλικων. Αυτό μπορεί να γίνει χρησιμοποιώντας Μοντέλα Διακριτών Καταστάσεων ή Κρυφές Αλυσίδες Markov.

Μια ιεραρχική γραφική αναπαράσταση ενός μοντέλου διακριτής κατάστασης μιας παράλληλης δέσμης τεσσάρων ελίκων



Co-funded by the Erasmus+ Programme of the European Union

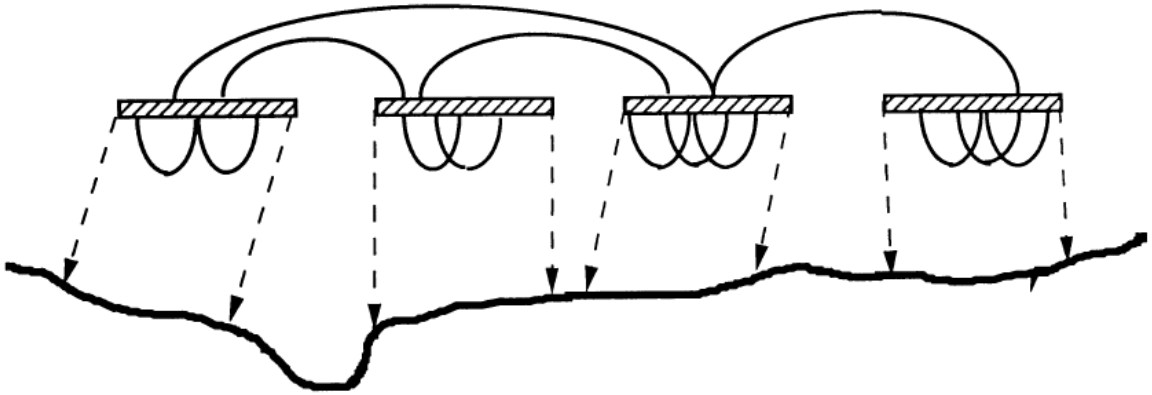
Μοτίβα Διαγνωστικά Δομής

Τα μοντέλα που χρησιμοποιούνται για την προσέγγιση "Threading" στο πρόβλημα της αντίστροφης αναδίπλωσης πρωτεΐνης αντιπροσωπεύονται συχνά από ακόμη πιο πολύπλοκα μοτίβα. Εκεί, εκτός από το ότι περιέχουν μοτίβα περιβαλλόντων θέσης αλληλουχίας και τις σχετικές προτιμήσεις αμινοξέων, προστίθενται περιορισμοί ανά ζεύγη ή προτιμήσεις επαφής. Αυτά μπορούν να αντιπροσωπεύονται από ένα μοτίβο προτιμήσεων αμινοξέων κατά ζεύγη που σχετίζονται με μια μήτρα γειτνίασης (ή γράφημα επαφής) μεταξύ ορισμένων υποομάδων θέσεων. Συνήθως υπάρχει μια βαθμολογία ψευδοενέργειας που σχετίζεται με κάθε ζεύγος υπονοούμενων επαφών για οποιοδήποτε συγκεκριμένο σπείρωμα ή ευθυγράμμιση ακολουθίας με δομικό μοντέλο. Εδώ κάποιος αναγκάζεται σε μια αναπαράσταση χωρίς πλαίσιο όπως τα τυχαία πεδία Markov.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Μια γραφική αναπαράσταση ενός πιθανού συνόλου επαφών κατά ζεύγη που θα συμπεριληφθούν σε ένα μοντέλο σπειρώματος



Co-funded by the Erasmus+ Programme of the European Union

Διαγνωστικά μοτίβων λειτουργίας

Όπως στην απλούστερη περίπτωση ενός δομικού σχεδίου, κανονικές εκφράσεις" ή προφίλ χρησιμοποιούνται για να αναπαραστήσουν τα πρότυπα των διατηρημένων αμινοξέων που είναι κοινά σε μια δεδομένη οικογένεια λειτουργικών πρωτεϊνών. Θα πρέπει να είναι προφανές ότι τέτοια μοτίβα συνήθως περιλαμβάνουν πολλά δομικά στοιχεία, καθώς οι ίδιες λειτουργίες είναι σχεδόν πάντα κωδικοποιούνται σε παρόμοιες δομές. Επιπλέον, περιλαμβάνονται κοινά στοιχεία μοτίβου, μόνο επειδή δεν έχει ακόμη παρατηρηθεί η επιτρεπόμενη παραλλαγή. Και τα δύο είναι απλώς το αναμενόμενο αποτέλεσμα της συντηρητικής/ευκαιριακής φύσης της εξέλιξης. Ενώ είναι γενικά δύσκολο να διαχωριστούν αυτές οι κατηγορίες στοιχείων, είναι συχνά δυνατό να προσδιοριστούν με σαφήνεια πειραματικά εκείνα τα στοιχεία μοτίβου που είναι απαραίτητα για μια «ενεργή τοποθεσία». μοντελοποίηση ότι αυτά μπορούν να περιοριστούν στο ελάχιστο.

Co-funded by the Erasmus+ Programme of the European Union

Διαγνωστικά μοτίβων λειτουργίας

Η ευθυγράμμιση μοτίβων είναι ιδιαίτερα χρήσιμη αφού όχι μόνο προτείνει μια συνολική συνάρτηση, αλλά και πληροφορίες για το ποια μέρη μιας νέας ακολουθίας είναι πιο πιθανό να είναι κρίσιμα για τη λειτουργία της. Ωστόσο, υπάρχουν περιορισμοί στις περισσότερες τρέχουσες υλοποιήσεις. Καμία από τις τρέχουσες μεθόδους αναγνώρισης προτύπων δεν ασχολείται με πειστικό τρόπο με οικογένειες πρωτεϊνών που μοιράζονται ετερόλογα σύνολα λειτουργικών και/ή δομικών τομέων. Οι τρέχουσες μέθοδοι παράγουν μοτίβα κοινών στοιχείων ενώ εξαλείφουν τα στοιχεία ετερόλογης ακολουθίας. Αυτή είναι η δύναμή τους όταν ασχολούνται με πρωτεΐνες ενός τομέα, γιατί είναι οι εξαιρετικά μεταβλητές περιοχές μεταξύ ομόλογων αλληλουχιών που μπορούν να οδηγήσουν σε τυχαίες και παραπλανητικές πλήρεις αντιστοιχίσεις αναζήτησης βάσης δεδομένων. Ωστόσο, στις ετερόλογες οικογένειες πρωτεϊνών πολλαπλών περιοχών, ολόκληροι και σημαντικοί τομείς μπορεί επίσης να απορριφθούν ή να μειωθούν σε σχεδόν πλήρη εκφυλισμό στη δημιουργία ενός ενιαίου προτύπου κοινού για όλους.

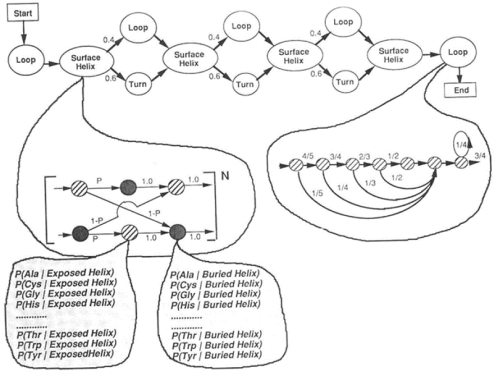
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Κρυμμένο μοντέλο Markov

Έχει πραγματοποιηθεί σημαντική έρευνα στον τομέα της αναγνώρισης προτύπων ομιλίας χρησιμοποιώντας μοντέλα Hidden Markov ή «διακριτικής κατάστασης». Ενώ υπήρξαν περιορισμένες εφαρμογές στην ανάλυση αλληλουχίας πρωτεϊνών, τα προβλήματα είναι εννοιολογικά πολύ παρόμοια. Ο λόγος είναι μια γραμμική ακολουθία στοιχείων. Οι ερμηνείες του προτύπου του είναι εξαιρετικά περίπλοκες και εξαρτώνται πολύ από το πλαίσιο, δεν διαφέρουν από την κωδικοποίηση δομής και λειτουργίας της αλληλουχίας πρωτεΐνης. Τέτοια μοντέλα μπορούν να θεωρηθούν ως γεννήτριες σημάτων ή ως γεννήτριες αλληλουχιών αμινοξέων.

Καθένας από τους N κόμβους αντιπροσωπεύει μια μοντελοποιημένη κατάσταση. Τα συνδετικά τόξα αντιπροσωπεύουν τις επιτρεπόμενες μεταβάσεις μεταξύ των καταστάσεων και την πιθανότητα τους. Κάθε κατάσταση μοντέλου περιλαμβάνει ένα υποσύνολο «κρυφών» καταστάσεων εξόδου. Κάθε κατάσταση εξόδου ορίζει μια κατανομή σε κάποιο αλφάβητο συμβόλων εξόδου, ή στην περίπτωσή μας τα αμινοξέα. Τα μοτίβα που αναπαριστώνται από το κρυφό μοντέλο Markov αποτελούνται επομένως από έναν πίνακα μετάβασης N επί N από κατάσταση σε κατάσταση και ένα σύνολο συναρτήσεων κατανομής κατάστασης εξόδου.



Co-funded by the Erasmus+ Programme of the European Union

Συνδυασμένα μοτίβα

Σε εκείνες τις περιπτώσεις όπου ένα μοτίβο λειτουργικής αλληλουχίας πιστεύεται ότι περιέχει κυρίως μόνο εκείνα τα στοιχεία αλληλουχίας που κωδικοποιούν άμεσα τη συνάρτηση αντί να κωδικοποιούν έμμεσα την κοινή δομή σε αυτή τη λειτουργική οικογένεια, τα μοτίβα είναι συχνά πολύ σύντομα ή ανεπαρκούς εξειδίκευσης για να είναι διαγνωστικά. Ένα κλασικό παράδειγμα είναι το μοτίβο των υπολειμμάτων που σχετίζονται με την καταλυτική τριάδα των πρωτεασών σερίνης. Έτσι, δεδομένου ότι όλες οι γνωστές πρωτεάσες σερίνης βρέθηκαν να είναι όλα βήτα (ακατέργαστα τυλιγμένα διπλά) φύλλα, ένα μοτίβο που αποτελείται από την εξαιρετικά ευαίσθητη, αλλά μη ειδική καταλυτική τριάδα "ενσωματωμένη" σε ένα γενικό σύνολο βήτα δομικό σχέδιο μπορεί να έχει βελτιωθεί σημαντικά η εξειδίκευση χωρίς μειωμένη ευαισθησία.

Αξιολόγηση Μοτίβου

Τουλάχιστον, δύο ποσότητες θα πρέπει να αναφέρονται για οποιοδήποτε μοτίβο: η ευαισθησία και η ειδικότητά του (αντανακλούν σφάλματα τύπου I και τύπου II).

Ευαισθησία = TP(True Positives)/{TP + FN(False Negatives)}

και

Ειδικότητα (επιλεκτικότητα) = TN (True Negatives)/{TN + FP(False Positives)}

Co-funded by the Erasmus+ Programme of the European Union

Εδώ TP είναι ο αριθμός των καθοριστικών (ή θετικών) αντιστοιχίσεων σετ σετ και FP είναι ο αριθμός των αντιστοιχιών σετ ελέγχου (ή αρνητικών). Η ευαισθησία ενός σχεδίου (π.χ., διαγνωστικά για ψευδάργυρα δάχτυλα) αντανακλά την πιθανότητα ότι όλα τα περιστατικά (π.χ. όλα τα «πραγματικά» δάχτυλα ψευδάργυρου) έχουν εντοπιστεί και έτσι θα βρεθούν. Αντικατοπτρίζει πόσο ευαίσθητο είναι το μοτίβο ως ανιχνευτής (π.χ. ψευδαργύρου δακτύλων). Η ειδικότητα είναι ένα μέτρο της πιθανότητας να εντοπιστούν σωστά οι αρνητικές περιπτώσεις. Η διαγνωστική ικανότητα ενός μοτίβου μπορεί εύκολα να συνοψιστεί σε έναν κανονικό πίνακα αλήθειας ή απρόβλεπτου δύο προς δύο, δεδομένου ενός θετικού και αρνητικού συνόλου ελέγχου.

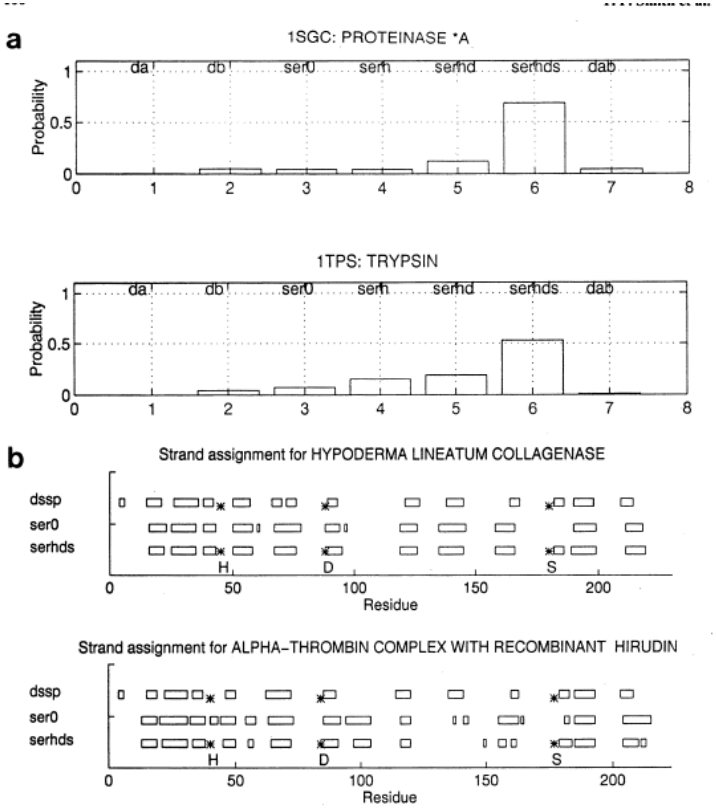
Co-funded by the Erasmus+ Programme of the European Union

Αξιολόγηση Μοτίβου

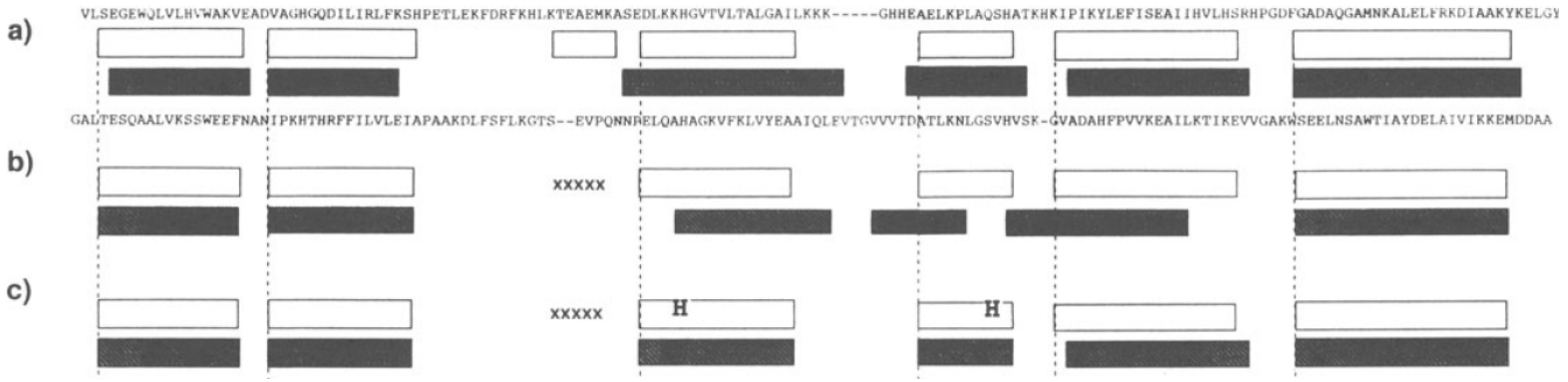
Η βιβλιογραφία της επιστήμης των υπολογιστών και της στατιστικής αναγνώρισης προτύπων είναι γεμάτη από συζητήσεις για την εκτίμηση της συμμετοχής στην τάξη χρησιμοποιώντας την Μπεϋζιανή προσέγγιση. Επομένως, μπορεί να είναι χρήσιμο να προσδιοριστεί η σχέση μεταξύ αυτών των συζητήσεων και της ευαισθησίας και της ιδιαιτερότητας. Έστω Pi ένα μοτίβο και f_j η πρωτεϊνική δομή ή λειτουργική τάξη. Στη συνέχεια, μπορεί κανείς να ορίσει τη σχέση μεταξύ των υπό όρους πιθανοτήτων από τον ορισμό της κοινής πιθανότητας. Αυτή είναι γνωστή ως σχέση Bayes:

$$P(f_j \mid p_i) = \frac{P(p_j \mid f_j) * P(f_j)}{P(p_i)}$$

(α) Σχετικές πιθανότητες επτά διαφορετικών Μοντέλων Διακριτών Καταστάσεων που έχουν δημιουργήσει δύο διαφορετικές αλληλουχίες αμινοξέων πρωτεάσης σερίνης.
(β) Δύο παραδείγματα αυξημένης ακρίβειας πρόβλεψης δευτερεύουσας δομής με αυξανόμενη εξειδίκευση ακολουθίας σε συνδυασμό με ένα μοντέλο διακριτικής κατάστασης beta.



Αλληλουχία λεγιμοσφαιρίνης περασμένη μέσω του δομικού μοντέλου μυοσφαιρίνης. (Α) Ευθυγράμμιση αναφοράς. (Β) Η σύντομη έλικα "D" αφαιρέθηκε, δεν υπάρχουν περιορισμοί ενεργής τοποθεσίας. (Γ) Η μικρή έλικα "D" αφαιρέθηκε, απαιτώντας στοιχεία μοντέλου στις θέσεις E7 και F8 να καταληφθούν από την ιστιδίνη, Η, από την αλληλουχία.

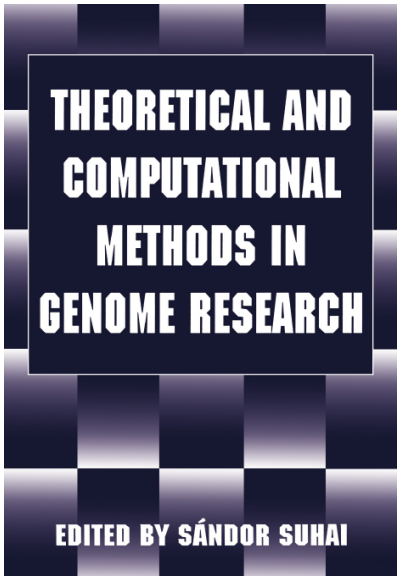


Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε τα διαγνωστικά πρότυπα της δομής.
- Εξηγήστε τα διαγνωστικά πρότυπα της λειτουργίας.
- Πώς θα μπορούσαν να αξιολογηθούν;

Co-funded by the Erasmus+ Programme of the European Union



Sandor Suhai (1997) Θεωρητική και Υπολογιστική Έρευνα Γονιδιώματος, Springer Science+Business Media Νέα Υόρκη

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 2. Γενετική

Θέμα 2. Θεωρητική Μοντελοποίηση στη Γονιδιωματική

Μάθημα 5. Αναγνώριση λειτουργικών τομέων σε βιολογικές ακολουθίες

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Πίνακες βάρους
- Ευθυγραμμίσεις που απαιτούν κενά
- Διανυσματικές ερμηνείες
- Διακριτικοί πίνακες

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

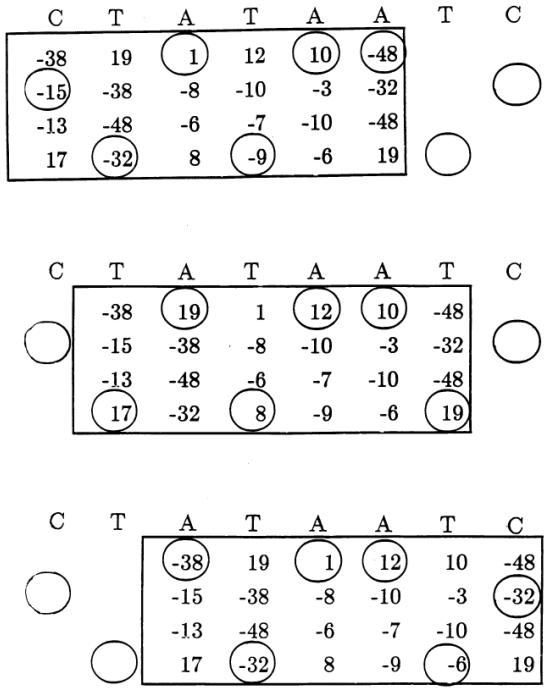
Οι πίνακες βάρους είναι γενικές αναπαραστάσεις της εξειδίκευσης της αλληλουχίας που παρουσιάζεται από τις κοινές τοποθεσίες. Αρχικά αναπτύχθηκαν για να αντιπροσωπεύουν ρυθμιστικές θέσεις στο DNA (ή RNA), όπως θέσεις δέσμευσης ριβοσώματος, προαγωγείς, θέσεις ματίσματος και άλλα. Έχουν επίσης υπάρξει επεκτάσεις για να αντιπροσωπεύει πρωτεϊνικά μοτίβα, και χρησιμοποιούνται στη βάση δεδομένων BLOCKS του conserved περιοχές σε οικογένειες πρωτεϊνών. Αυτές προσεγγίζουν όλες τις χρησιμοποιούμενες τοποθεσίες σταθερού μεγέθους. Η ευθυγράμμιση των τοποθεσιών δεν απαιτούσε την προσθήκη κενών. Οι Gribskov et al. έδειξαν πώς η ίδια βασική ιδέα μπορεί να επεκταθεί σε ευθυγραμμίσεις που περιλαμβάνουν κενά, σε αυτό που λέγεται "προφίλ." Αυτή η επέκταση απαιτεί δυναμικό προγραμματισμό για την εύρεση των βέλτιστων ευθυγραμμίσεων μεταξύ του πίνακα και της ακολουθίας. Πιο πρόσφατα Hidden Markov Models (HMM) έχουν χρησιμοποιηθεί για τον καθορισμό των μοτίβων σε διατηρημένες ακολουθίες. Αν και κλασσικό βάρος Οι πίνακες, τα προφίλ και τα HMM διαφέρουν σε πολλές λεπτομέρειες, μπορούν επίσης να θεωρηθούν ως παραλλαγές ενός κοινού θέματος, όπως θα περιγραφεί εν συντομία παρακάτω.

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή στους πίνακες βάρους

Η ιδέα προέρχεται από τη σκέψη ότι κάθε θέση σε έναν τόπο συνεισφέρει κάτι στη δραστηριότητά του. Προφανώς η βάση συναίνεσης σε μια θέση θα συνεισέφερε τα μέγιστα στη δραστηριότητα, αλλά οποιαδήποτε μεμονωμένη αλλαγή μακριά από τη συναίνεση είναι πιθανό να μειώσει κάπως τη δραστηριότητα και όχι να την εξαλείψει εντελώς. Αυτό, φυσικά, θα εξαρτηθεί από τον τύπο της τοποθεσίας. Ορισμένες θέσεις δέσμευσης πρωτεϊνών μπορεί να είναι πολύ συγκεκριμένες, όπως οι θέσεις περιορισμού, έτσι ώστε τυχόν αλλαγές από τη βάση συναίνεσης θα μειώσουν αποτελεσματικά τη δραστηριότητα πέρα από την ανίχνευση. Αλλά για τοπους όπως οι προωθητές, αυτό δεν φαίνεται να ισχύει.

Co-funded by the Erasmus+ Programme of the European Union



Χρησιμοποιώντας έναν πίνακα για τη βαθμολογία θέσεων σε μια ακολουθία. Σε κάθε θέση του πίνακα το σκορ είναι το άθροισμα από τα στοιχεία μήτρας που αντιστοιχούν στην ακολουθία σε αυτή τη θέση (τα κυκλωμένα στοιχεία).

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή στους πίνακες βάρους

Η μήτρα βάρους έχει τιμές για όλες τις πιθανές βάσεις σε κάθε θέση εντός της πιθανής θέσης, ενώ η ακολουθία έχει συγκεκριμένες βάσεις σε κάθε θέση. Οι τιμές του πίνακα βάρους για αυτές τις βάσεις αθροίζονται για να ληφθεί η συνολική βαθμολογία για τη συγκεκριμένη τοποθεσία. Στην περίπτωση μιας πρωτεΐνης που απλώς συνδέεται με το DNA, οι τιμές της μήτρας βάρους μπορούν να θεωρηθούν ως ενέργειες μερικής δέσμευσης που συνεισφέρονται από κάθε βάση που αλληλεπιδρά με την πρωτεΐνη και το άθροισμα αυτών είναι η συνολική ενέργεια δέσμευσης για την πρωτεΐνη ιστοσελίδα.

Co-funded by the Erasmus+ Programme of the European Union

A	0	1	0	1	1	0
C	0	0	0	0	0	0
G	0	0	0	0	0	0
T	1	0	1	0	0	1

A	0	10	8	10	10	0
C	1	0	1	1	3	0
G	1	0	1	1	1	0
T	10	0	10	1	1	10

A	2	95	26	59	51	1
C	9	2	14	13	20	3
G	10	1	16	15	13	0
T	79	3	44	13	17	96

A	-38	19	1	12	10	-48
C	-15	-38	-8	-10	-3	-32
G	-13	-48	-6	-7	-10	-48
T	17	-32	8	-9	-6	19

Διαφορετικές μήτρες που αντιπροσωπεύουν την περιοχή -10 των προαγωγέων E. coli. Ο κορυφαίος πίνακας είναι απλός συναινετική ακολουθία, ενώ κάθε ακόλουθος πίνακας προσθέτει πρόσθετες πληροφορίες σχετικά με τα μοτίβα που παρατηρούνται στο -10 περιφέρειες.

Co-funded by the Erasmus+ Programme of the European Union

AAAA	2.63
AAAC	2.63
AAAG	2.63
AAAT	0.46
AACA	1.79
AACC	1.79
AACG	1.79
AACT	-0.38
.	
.	
.	
AGCG	-0.25
AGCT	-2.42
AGGA	0.59
.	
.	
.	
CTTT	-0.37
GAAA	4.01
.	
.	
.	
TTTG	2.63
TTTT	0.46

A	-0.55	+1.38	+0.42	+1.38
C	0	+0.55	-0.42	+1.38
G	+0.83	-0.66	+0.42	+1.38
T	+0.83	+0.42	0	-0.79

Η αριστερή στήλη αντιπροσωπεύει την ενέργεια δέσμευσης και στις 256 αλληλουχίες 4 μήκους (μόνο μερικές από αυτές είναι απεικονίζεται). Ο πίνακας στα δεξιά παρέχει τις ίδιες πληροφορίες επειδή κάθε τιμή στη λίστα στα αριστερά μπορεί να είναι υπολογίζεται χρησιμοποιώντας αυτόν τον πίνακα.

Co-funded by the Erasmus+ Programme of the European Union

Ευθυγραμμίσεις που απαιτούν κενά

Οι πίνακες βάρους μπορούν να επεκταθούν για να περιλαμβάνουν ευθυγραμμίσεις με κενά, όπως στα προφίλ. Σε αυτήν την περίπτωση, ο πίνακας περιέχει επιπλέον σειρές για να υποδείξει τη σχετική βαθμολογία με κενό στην ευθυγράμμιση. Τυπικά θα υπάρχουν βαθμολογίες τόσο για την εισαγωγή του ένα κενό και μια διαφορετική βαθμολογία για την επέκταση μιας διαφοράς, όπως είναι ευρέως αποδεκτό ότι εκείνες δύο γεγονότα έχουν διαφορετικές πιθανότητες. Στις μεθόδους HMM, θα υπάρχουν διαφορετικές βαθμολογίες για την εισαγωγή παρεμβολών παρά για διαγραφές, αλλά έχει αποδειχθεί πώς και τα δύο Τα HMM και τα προφίλ μπορούν να περιγραφούν στο ίδιο συντακτικό πλαίσιο. Η μέθοδος εκπαίδευσης σχετίζεται με την Προσδοκία – Μεγιστοποίηση Μέθοδοι (EM) που χρησιμοποιήθηκαν προηγουμένως για ungapped ευθυγραμμίσεις| και ευθυγραμμίσεις με περιορισμένα κενά. Άλλες μέθοδοι είναι επίσης ικανές να βρουν κενά ευθυγραμμίσεις και τον προσδιορισμό των αναπαραστάσεων πινάκων αυτών των ευθυγραμμίσεων ταυτόχρονα.

Co-funded by the Erasmus+ Programme of the European Union

A.

A	9	214	63	142	118	8
C	22	7	26	31	52	13
G	18	2	29	38	29	5
T	193	19	124	31	43	216

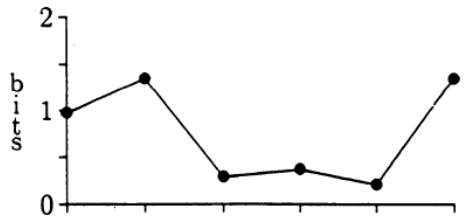
B.

A	0.04	0.88	0.26	0.59	0.49	0.03
C	0.09	0.03	0.11	0.13	0.22	0.05
G	0.07	0.01	0.12	0.16	0.12	0.02
T	0.80	0.08	0.51	0.13	0.18	0.89

C.

A	-2.76	1.82	0.06	1.23	0.96	-2.92
C	-1.46	-3.11	-1.22	-1.00	-0.22	-2.21
G	-1.76	-5.00	-1.06	-0.67	-1.06	-3.58
T	1.67	-1.66	1.04	-1.00	-0.49	1.84

D.



Προσδιορισμός της μήτρας ειδικότητας και του περιεχομένου πληροφοριών των -10 περιοχών των υποκινητών. Η πάνω μήτρα δείχνει την εμφάνιση κάθε βάσης σε κάθε θέση σε μια συλλογή 242 προαγωγέων. Ο επόμενος πίνακας μετατρέπει αυτούς τους αριθμούς στο κλάσμα κάθε βάσης σε κάθε θέση. Ο τρίτος πίνακας είναι ο λόγος λογαριθμικής πιθανότητας των παρατηρούμενων συχνοτήτων διαιρεμένος με την αναμενόμενη, χρησιμοποιώντας 0,25 ως την αναμενόμενη εμφάνιση για οποιαδήποτε βάση. (Ο λογάριθμος λαμβάνεται στη βάση 2.)

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Διανυσματικές ερμηνείες

Συχνά είναι χρήσιμο να σκεφτόμαστε τους πίνακες βάρους και τις ακολουθίες με όρους «χώρου ακολουθίας». Ο ίδιος ο πίνακας βάρους είναι ένα σημείο, ή διάνυσμα σε αυτό το σημείο, στο χώρο. Και, φυσικά, κάθε ακολουθία είναι επίσης ένα διάνυσμα σε αυτόν τον χώρο. Ωστόσο, πρέπει να σκεφτούμε τον χώρο όπως ορίζεται από τη μήτρα βάρους, να έχει διαστάσεις όχι μόνο για όλες τις πιθανές βάσεις σε κάθε θέση, αλλά και τις δυνατότητες κενών, επεκτάσεων κενών και μη ανεξάρτητων θέσεων. Προκειμένου να προσδιοριστεί το διάνυσμα για μια συγκεκριμένη ακολουθία (ή θέση) σε αυτό το διάστημα, είναι πρώτα απαραίτητο να ευθυγραμμιστεί με τη μήτρα. Η βαθμολογία για οποιαδήποτε ακολουθία, δεδομένης μιας μήτρας, είναι τότε μόνο το γινόμενο τελείας του διανύσματος για τον πίνακα με το διάνυσμα για την ακολουθία. Το διάνυσμα για την ακολουθία έχει απλώς ένα 1 για κάθε βάση που εμφανίζεται σε κάθε θέση (που αντιστοιχεί στους κύκλους κάτω από την ακολουθία) και ένα 0 για όλες τις βάσεις που δεν εμφανίζονται. Αυτό χρησιμοποιείται συνήθως σε στατιστικές μεθόδους και είναι γνωστό ως "εικονική κωδικοποίηση", όπου οι μεταβλητές που υπάρχουν για ένα συγκεκριμένο παράδειγμα δίνονται είναι και οι μεταβλητές που δεν υπάρχουν δίνονται 0 s. Οι ευθυγραμμισμένες ακολουθίες με κενά εξακολουθούν να κωδικοποιούνται με τον ίδιο τρόπο, αλλά τώρα τα 1 μπορεί να αναφέρονται σε θέσεις κενών ή επεκτάσεις κενών, αντί απλώς σε υπολείμματα σε θέσεις. Τότε η βαθμολογία οποιασδήποτε ευθυγραμμισμένης ακολουθίας (που ονομάζεται Sc (S_i) για την ακολουθία S_i) με τον πίνακα είναι απλώς το άθροισμα των μεταβλητών που εμφανίζονται και είναι απλώς το γινόμενο τελείας της ακολουθίας με τον πίνακα βάρους:

$$Sc(S_i) = \vec{W} \cdot \vec{S_i} = |W||S_i| \cos \alpha_i$$

Co-funded by the Erasmus+ Programme of the European Union

Διακριτικοί πίνακες

Κάποιος μπορεί να θέλει να χρησιμοποιήσει μια μήτρα βάρους με κάπως διαφορετικό τρόπο. Αντί να πάρουμε απλώς μια συλλογή από γνωστές τοποθεσίες και να προσπαθήσουμε να τις αναπαραστήσουμε αποτελεσματικά, ή ακόμα και να προσπαθήσουμε να δημιουργήσουμε έναν πίνακα με καλές ποσοτικές προβλέψεις της δραστηριότητας του ιστότοπου, μπορεί κανείς να έχει στόχο να βρει έναν πίνακα που είναι βέλτιστα αποτελεσματικός στη διάκριση πραγματικών τοποθεσιών από άλλους ακολουθίες, μη τοποθεσίες. Για παράδειγμα, εάν κάποιος έχει μια συλλογή από τοποθεσίες που είναι γνωστό ότι είναι λειτουργικές και ένα άλλο σύνολο ακολουθιών που είναι γνωστό ότι δεν είναι λειτουργικές τοποθεσίες, θα μπορούσε να προσπαθήσει να βρει έναν πίνακα που διακρίνει τα δύο σύνολα δίνοντας σε κάθε τοποθεσία βαθμολογία μεγαλύτερη από οποιαδήποτε μη τοποθεσία. Στην πραγματικότητα, αυτή ήταν η πρώτη εφαρμογή πινάκων βάρους, που χρησιμοποίησαν ένα απλό νευρωνικό δίκτυο που ονομάζεται "perceptron" για τον προσδιορισμό των στοιχείων της μήτρας.

Διακριτικοί πίνακες

Ένας δεδομένος πίνακας θα καθορίσει μια βαθμολογία για οποιαδήποτε συγκεκριμένη ακολουθία. Εάν οι θέσεις είναι θέσεις δέσμευσης για κάποια πρωτεΐνη, αυτή η προσέγγιση σχετίζεται άμεσα με την εύρεση των θέσεων με μέγιστη ειδική ενέργεια δέσμευσης για την πρωτεΐνη. Η μέθοδος λειτουργεί με τον υπολογισμό της «συνάρτησης κατανομής» ή της συνολικής πιθανότητας δέσμευσης, σε όλες τις πιθανές θέσεις στη συλλογή (όλο το γονιδίωμα για μια πρωτεΐνη που δεσμεύει το DNA). Τότε ο στόχος είναι να βρεθεί η μήτρα βάρους που μεγιστοποιεί

$$U = \frac{1}{N}\sum_{i=1}^{N} \vec{W} \cdot \vec{S}_i - \ln Y$$

όπου το άθροισμα είναι σε όλες τις N τοποθεσίες και το Y είναι η μέση πιθανότητα (δηλαδή, e dot -product) σε ολόκληρη τη συλλογή των ακολουθιών, τόσο τοποθεσιών όσο και μη.

Διακριτικοί πίνακες

Οι τρέχουσες προσεγγίσεις χρησιμοποιούν μια ποικιλία μεθόδων, από τη σύγκριση μιας νέας ακολουθίας με μεμονωμένες ακολουθίες στη βάση δεδομένων, όπως με BLAST ή FASTA, έως τη σύγκριση της νέας ακολουθίας με οικογενειακές αναπαραστάσεις όπως προφίλ, HMM, BLOCKS ή μοτίβα Prosite . Ως αρχική δοκιμή συγκρίθηκε η ικανότητα των βελτιστοποιημένων πινάκων να διακρίνουν τα μέλη της οικογένειας με αυτή της βάσης δεδομένων BLOCKS. Αυτή η βάση δεδομένων προέρχεται από τις οικογένειες πρωτεϊνών που περιγράφονται στο Prosite , αλλά αντιπροσωπεύει τις πιο συντηρημένες, χωρίς διάκενο περιοχές ως πίνακες βάρους που μπορούν να χρησιμοποιηθούν για τη βαθμολογία νέων αλληλουχιών και τον προσδιορισμό του εάν είναι μέλη της οικογένειας ή όχι. Προηγουμένως οι μήτρες βάρους είχαν καθοριστεί αποκλειστικά από τα μέλη της οικογένειας, χωρίς να ληφθούν υπόψη οι αλληλουχίες των πρωτεϊνών που δεν είναι μέλη της οικογένειας. Έχει δοθεί μεγάλη έμφαση στον τρόπο στάθμισης των αλληλουχιών των μελών της οικογένειας, έτσι ώστε να αντισταθμίζεται η μεροληψία στην αναπαράσταση στην τρέχουσα βάση δεδομένων και να αντισταθμίζονται τα περιορισμένα μεγέθη δειγμάτων.

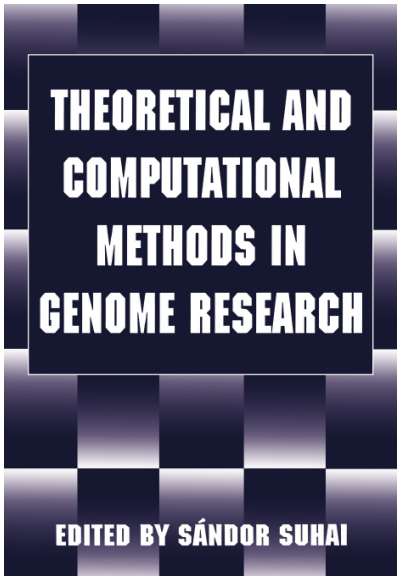
Η κατανομή των βαθμολογιών για όλες τις ακολουθίες στο SwissProt 29 χρησιμοποιώντας τη βελτιστοποιημένη μήτρα για το Homeodomain BLOCK. Οι κορυφές και των δύο κατανομών εκτείνονται πέρα από την κορυφή του σχήματος και δεν φαίνονται.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε το ρόλο των matrices για την αναγνώριση λειτουργικού τομέα.
- Τι είναι η "συνάρτηση διαμερίσματος" ή η συνολική πιθανότητα δέσμευσης και πώς θα μπορούσε να υπολογιστεί ;

Co-funded by the Erasmus+ Programme of the European Union



Sandor Suhai (1997) Θεωρητική και Υπολογιστική Έρευνα Γονιδιώματος, Springer Science+Business Media Νέα Υόρκη

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 2. Γενετική

Θέμα 2. Θεωρητική Μοντελοποίηση στη Γονιδιωματική

Μάθημα 6. Προσομοιώσεις Αλληλεπιδράσεων Πρωτεϊνών-DNA

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Προγράμματα ανίχνευσης ή διαδικτυακοί διακομιστές που μπορούν να ανιχνεύσουν ενεργούς τόπους
- Στρατηγικές στην τοποθέτηση συνδέτη
- Άκαμπτο αμάξωμα σύνδεσης
- Ευέλικτη βάση σύνδεσης

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Η μοριακή αναγνώριση είναι ένα ουσιαστικό συστατικό σε όλες σχεδόν τις βιομοριακές διεργασίες, ειδικά σε διαδικασίες που σχετίζονται με τη μεταγραφή και τη μετάφραση του γενετικού υλικού. Μεγάλη πρόοδος έχει σημειωθεί τα τελευταία χρόνια ως προς τον χαρακτηρισμό αρκετών τέτοιων συστημάτων με δομικούς όρους, παρέχοντας μια εικόνα για το θεμελιώδες ζήτημα της δομικής βάσης για αλληλεπιδράσεις και δεσμεύσεις που εξαρτώνται από την ακολουθία. Συγκεκριμένα, μπορεί κανείς να εντοπίσει ορισμένες αρχές αναγνώρισης και δομικής οργάνωσης μέσα στις οικογένειες των παραγόντων μεταγραφής.

Η προσομοίωση μοριακής δυναμικής (MD) παρέχει μια πολύ λεπτομερή, δομική και δυναμική περιγραφή των βιομοριακών συστημάτων. αυτό το επίπεδο λεπτομέρειας, το οποίο είναι πολύ δύσκολο να αποκτηθεί με άλλα μέσα, είναι πολύ πολύτιμο για μια ενδελεχή κατανόηση της λεπτής ισορροπίας μεταξύ των ανταγωνιστικών αλληλεπιδράσεων που εμπλέκονται στις διαδικασίες μοριακής αναγνώρισης. Από μια σύγκριση υπολογισμένων ενεργειών αλληλεπίδρασης (ενθαλπίες) σε σύμπλοκα υποστρώματος:πρωτεΐνης , με υπολογισμένες τιμές ελεύθερης ενέργειας καθώς και με πειραματικά δεδομένα, είναι ξεκάθαρο ότι οι απλές, διαισθητικές εικασίες για το αποτέλεσμα των πειραμάτων μετάλλαξης σε πολύπλοκα συστήματα είναι αναξιόπιστες. Οι επιδράσεις των ελαφρών δομικών αλλαγών, της αλληλεπίδρασης με διαλύτες και ιόντα και τα εντροπικά φαινόμενα είναι πολύ δύσκολο να μαντέψουν. Ως εκ τούτου, είναι απαραίτητες πιο ακριβείς μέθοδοι, όπως η διαταραχή ελεύθερης ενέργειας ή οι υπολογισμοί του δυναμικού της μέσης δύναμης. Αν και ορισμένες πτυχές αυτών των συστημάτων μπορεί επίσης να απαιτούν συνδυασμένους υπολογισμούς ενέργειας μοριακής μηχανικής/κβαντομηχανικής, οι διαδικασίες μη ομοιοπολικής δέσμευσης που βρίσκονται στο επίκεντρο αυτής της έκθεσης έχουν μελετηθεί χρησιμοποιώντας την κλασική μηχανική και τις εμπειρικές ενεργειακές συναρτήσεις.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Προσομοιώσεις Συστημάτων Πρωτεϊνών-DNA

Στη βιβλιογραφία έχουν αναφερθεί μερικές δυναμικές προσομοιώσεις συστημάτων πρωτεΐνης-DNA. Η κεφαλή καταστολέα lac σε σύμπλεγμα με τον χειριστή της προσομοιώθηκε χρησιμοποιώντας περιορισμούς απόστασης από μελέτες NMR, με ένα κομμάτι απεριόριστης τροχιάς 50ps. Από αυτή την προσομοίωση, οι ειδικές επαφές μεταξύ της πρωτεΐνης και του DNA βρέθηκαν να σχηματίζονται κυρίως από μη πολικές επαφές και σε μικρότερο βαθμό μέσω δεσμών υδρογόνου που διαμεσολαβούνται από το νερό. Η περιοχή δέσμευσης DNA του υποδοχέα γλυκοκορτικοειδών (GRDBD) σε σύμπλοκο με DNA έχει αποτελέσει αντικείμενο μελετών από τρεις ομάδες που χρησιμοποιούν ελαφρώς διαφορετικές προσεγγίσεις. Ο Harris et al διεξήγαγε μια προσομοίωση 1ns ενός μοντέλου του διμερούς GRDBD που είναι συνδεδεμένο με το DNA με ένα κέλυφος νερού 10 Å που περιβάλλει το σύστημα (που αποδίδει συνολικά 9625 άτομα) και οι Bishop & Schulten έχουν αναφέρει δύο προσομοιώσεις 90ps του διμερούς GRDBD δεσμεύεται σε δύο αλληλουχίες DNA με διαφορετικές αποστάσεις μεταξύ των μισών θέσεων αναγνώρισης, χρησιμοποιώντας μια ελλειψοειδή γεωμετρία για το περιβάλλον νερό (περίπου 13600 άτομα συνολικά). Οι προσομοιώσεις του συστήματος GRDBD του Luisi et al χρησιμοποίησαν τις κρυσταλλογραφικές συντεταγμένες για το σύμπλοκο σε μια σφαίρα νερού ακτίνας 42 Å (περίπου 31000 άτομα συνολικά), η οποία προσομοιώθηκε για 200ps. Συμπλέγματα μονομερών GRDBD συνδεδεμένων σε διάφορες αλληλουχίες DNA ημιθέσης προσομοιώθηκαν για 200ps και χρησιμοποιήθηκαν επίσης σε υπολογισμούς διαταραχών ελεύθερης ενέργειας των σχετικών συγγένειων του GRDBD για αυτές τις αλληλουχίες DNA.

Προσομοιώσεις άλλων συμπλεγμάτων νουκλεϊκών οξέων

Η θερμοδυναμική των αλληλεπιδράσεων νουκλεϊκού οξέος με πρωτεΐνες και μικρούς συνδέτες έχει επίσης προσομοιωθεί με αποτελέσματα σε καλή συμφωνία με το πείραμα.

DNA- Σύμπλοκο

Το μικρό πεπτιδικό φάρμακο νετροπσίνη συνδέεται κατά προτίμηση με τη δευτερεύουσα αύλακα των περιοχών AT του DNA, οι οποίες ευνοούνται από 4,4 kcal/mol έναντι των περιοχών GC, μια διαφορά που έχει αποδοθεί στην απώθηση της αμινομάδας που προεξέχει στη δευτερεύουσα αύλακα ενός GC . ζεύγος, αλλά το οποίο απουσιάζει σε ένα ζεύγος AT. Έχουμε εφαρμόσει τη μέθοδο θερμοδυναμικού κύκλου-διαταραχής για να υπολογίσουμε αυτή τη μεταβολή στην ελεύθερη ενέργεια δέσμευσης με πολύ ενθαρρυντικά αποτελέσματα. Λάβαμε μια τιμή για το ΔΔ G στην περιοχή από 3,4 έως 4,0 kcal/mol με εκτιμώμενο σφάλμα 1-2 kcal/mol. Τα αποτελέσματα δεν φαίνεται να εξαρτώνται από το μέγεθος του συστήματος ή το μήκος των προσομοιώσεων εντός των ορίων που έχουμε δοκιμάσει: 1Q-20ps σειρές είτε ενός DNA εξαμερούς διπλής όψης στο νερό (1200 άτομα) είτε ενός DNA οκταμερούς διπλής όψης στο νερό (2600 άτομα).

Co-funded by the Erasmus+ Programme of the European Union

Ριβονουκλεάση T_1

Η ριβονουκλεάση T_1 είναι ένα μικρό ένζυμο που διασπά επιλεκτικά το μονόκλωνο RNA στην 3' πλευρά των γουανινών και του οποίου η τρισδιάστατη δομή είναι γνωστή με και χωρίς δεσμευμένο υπόστρωμα. Αυτό καθιστά το ένζυμο ιδανικό για τη μελέτη των αλληλεπιδράσεων των πρωτεϊνονουκλεϊκών οξέων. Η εργασία μας σε αυτό το σύστημα έχει επικεντρωθεί μέχρι στιγμής σε αλλαγές στη δομή του ενζύμου άγριου τύπου, ιδιαίτερα στην ενεργό θέση του, κατά τη δέσμευση του αναστολέα 2'GMP. Οι αρχικές πειραματικές μελέτες που χρησιμοποιούσαν φθορισμό με ανάλυση χρόνου έδειξαν την παρουσία μιας κίνησης σε picosecond της Trp59 (της μόνης τρυπτοφάνης στο ένζυμο) η οποία μεταβλήθηκε από τη δέσμευση του 2'GMP. Αυτό παρατηρήθηκε επίσης σε μια σειρά προσομοιώσεων μοριακής δυναμικής των ίδιων συστημάτων, σε νερό καθώς και σε κενό.

Co-funded by the Erasmus+ Programme of the European Union

Ριβονουκλεάση T1

Από μια σειρά προσομοιώσεων ελεύθερης ενέργειας που συγκρίνουν τη δέσμευση της 2-αμινοπουρίνης έναντι της 3'GMP στο ένζυμο βρέθηκε ότι η 2-αμινοπουρίνη έχει μια λιγότερο ευνοϊκή αλληλεπίδραση (κατά -5 kcal/mol) η οποία είναι λογική. η επίδραση της αντισταθμιστικής μετάλλαξης Glu46-Gln46 φαίνεται κυρίως να είναι μείωση της συγγένειας για GMP, χωρίς σχεδόν καμία άμεση επίδραση στη δέσμευση της 2-αμινοπουρίνης.

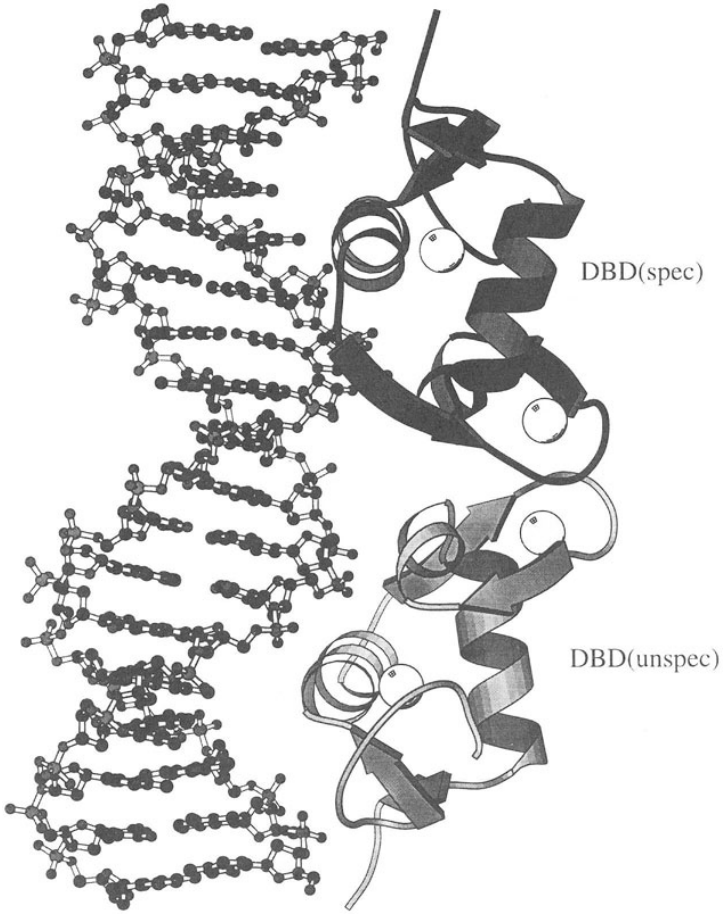
Co-funded by the Erasmus+ Programme of the European Union

Υποδοχείς πυρηνικών ορμονών

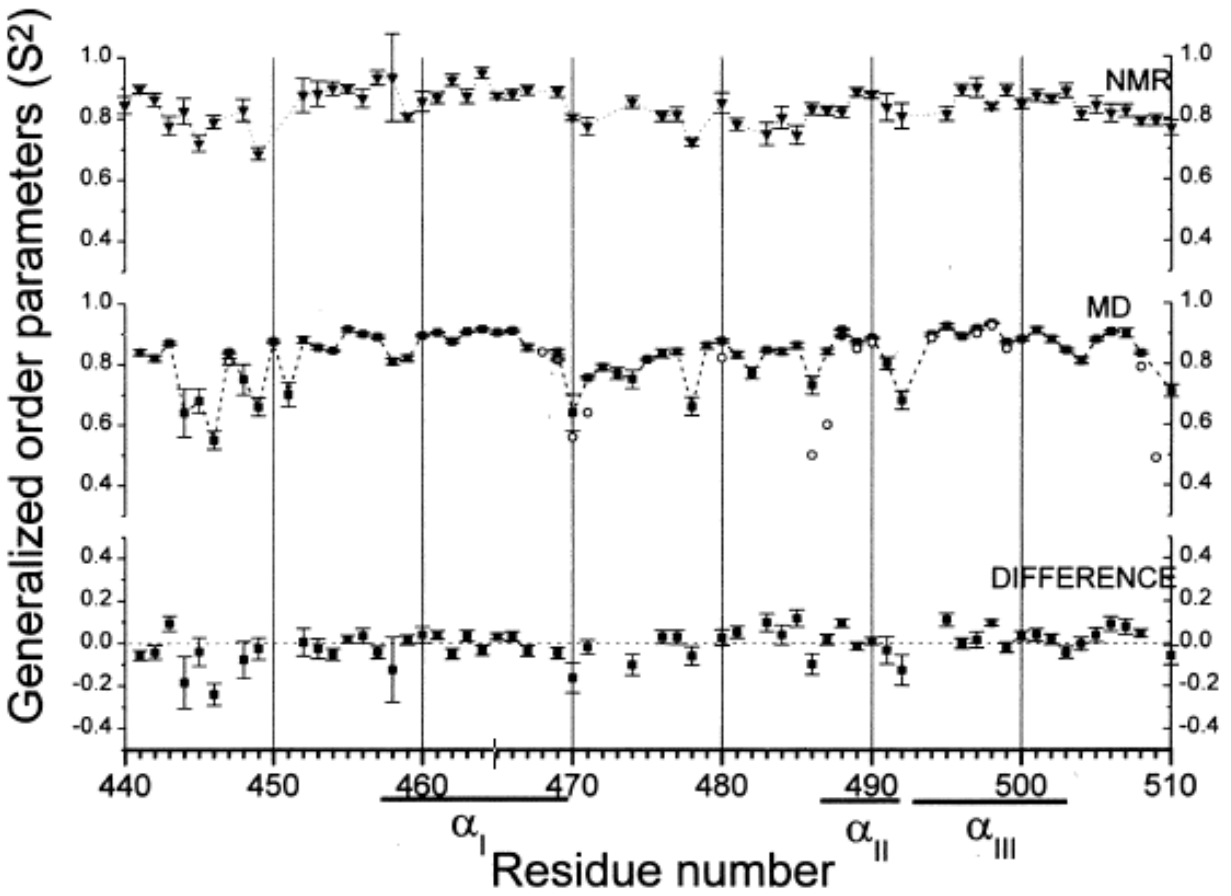
Οι μεταγραφικοί παράγοντες που ενεργοποιούνται από ορμόνες όπως τα γλυκοκορτικοειδή, τα οιστρογόνα και τα ρετινοειδή, αποτελούν τη μεγάλη οικογένεια των πυρηνικών ορμονικών υποδοχέων. Αυτοί οι υποδοχείς φαίνεται να έχουν κοινή εξελικτική προέλευση και να έχουν εξελιχθεί σε τρεις διακριτές υποοικογένειες, μία από τις οποίες είναι η ομάδα υποδοχέων στεροειδών ορμονών. Οι υποδοχείς είναι πρωτεΐνες πολλαπλών τομέων, με τις διαφορετικές λειτουργίες που απαιτούνται εντός του συστήματος μεταγωγής σήματος που εκτελούνται από διαφορετικές περιοχές και συνδέονται με τα στοιχεία απόκρισης του DNA σε διάφορους διμερείς συνδυασμούς. Ορισμένες από τις κύριες περιοχές περιέχουν τις λειτουργίες σύνδεσης συνδέτη, σύνδεσης DNA και μετενεργοποίησης.

Υποδοχείς πυρηνικών ορμονών

Υπάρχουν δύο οικογένειες στοιχείων απόκρισης για τους υποδοχείς στεροειδών, τα στοιχεία απόκρισης γλυκοκορτικοειδούς (GRE) με την ημιθέση αλληλουχία TGTTCT και τα στοιχεία απόκρισης οιστρογόνου (ERE) με την αλληλουχία ημιθέσης TGACCT. Οι κρυσταλλογραφικές δομές έχουν προσδιοριστεί για συστήματα και των δύο ειδών.

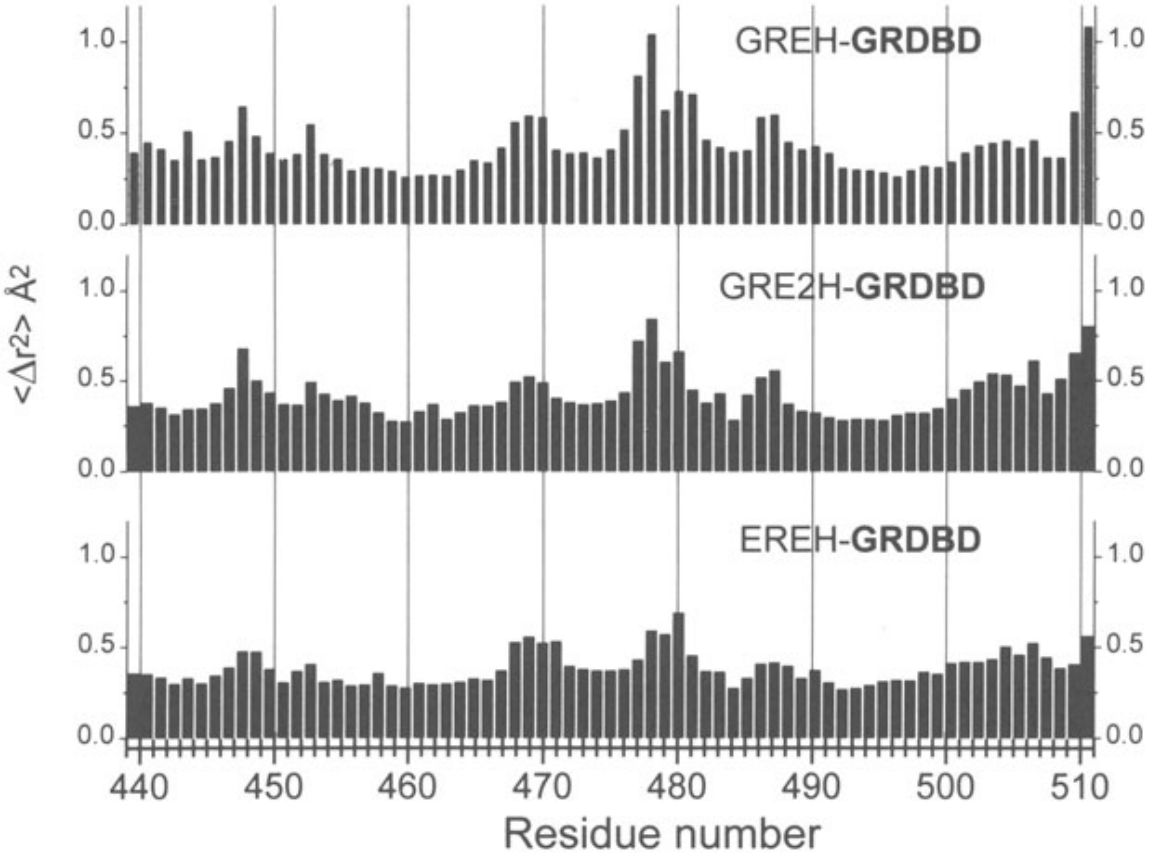


Σχέδιο του συμπλέγματος GRE:GRDBD



Παράμετροι σειράς έναντι υπολειμμάτων στο μονομερές GRDBD

Μέσες τετραγωνικές ατομικές διακυμάνσεις σε διαφορετικά σύμπλοκα GRDBD και στοιχεία απόκρισης DNA



Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Αλληλεπιδράσεις πρωτεΐνης-DNA

Συνολικά η δομή του συμπλέγματος διατηρείται, μια μικρή κάμψη στο DNA γίνεται ελαφρώς πιο έντονη λόγω κάπως στενότερων επαφών μεταξύ πρωτεΐνης και DNA και παρατηρούνται επίσης ορισμένες τοπικές παραμορφώσεις στο DBD. Υπάρχει μία περιοχή του DBD όπου οι δομές NMR και ακτίνων X διαφέρουν και στις προσομοιώσεις του μονομερούς DBD σε διάλυμα, ξεκινώντας από έναν από αυτούς τους διαμορφωτές, η προσομοιωμένη δομή παραμένει κοντά στην αρχική της διαμόρφωση, υποδεικνύοντας ότι και οι δύο διαμορφωτές βρίσκονται στα τοπικά ελάχιστα. Τα αποτελέσματα στη δυναμική της ραχοκοκαλιάς συσχετίζονται καλά με τα δεδομένα χαλάρωσης NMR από ^{15}N-σημασμένες πρωτεΐνες, και στις περιπτώσεις με σημαντικές αποκλίσεις, μια ανάλυση των πραγματικών κινήσεων όπως βρέθηκαν στις προσομοιωμένες τροχιές αποκάλυψε ότι ήταν γενικά το αποτέλεσμα σπάνιων διεργασιών, π.χ. Ο δεσμός υδρογόνου NH θα πηδούσε μπρος-πίσω μεταξύ διαφορετικών ομάδων αποδοχής δεσμών H, προκαλώντας ανεπαρκή δειγματοληψία.

Co-funded by the Erasmus+ Programme of the European Union

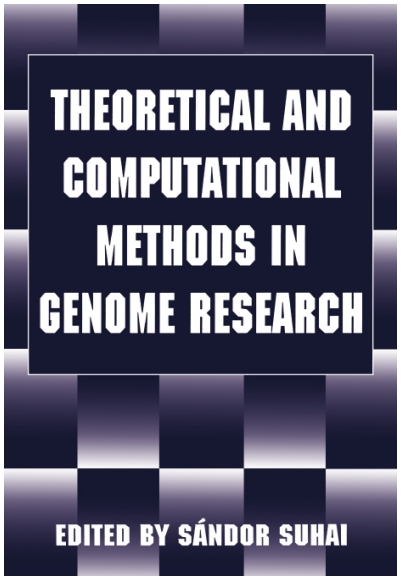
Οι προσομοιώσεις συστημάτων πρωτεΐνης:DNA είναι εφικτές σε χρονική κλίμακα νανοδευτερόλεπτου και αποδίδουν δομικά, δυναμικά και θερμοδυναμικά αποτελέσματα που συμφωνούν καλά με τα διαθέσιμα πειραματικά δεδομένα. Επιπλέον, βρίσκουμε λεπτομέρειες σχετικά με τη δυναμική στις διεπαφές επαφής όσον αφορά τους άμεσους δεσμούς υδρογόνου καθώς και τους δεσμούς υδρογόνου που διαμεσολαβούνται από το νερό που δεν φαίνονται εύκολα σε πειράματα. Θα μπορούσαμε επίσης να κάνουμε κάποιες προτάσεις σχετικά με τους υποκείμενους λόγους για τις διαφορές στις σταθερότητες και τη συνεργασία του σχηματισμού συμπλόκου στις περιπτώσεις GRDBD:GREH και GRDBD:EREH, όπως η ακαμψία του GRDBD όταν δεσμεύεται στο EREH.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Δώστε μερικά παραδείγματα προσομοιώσεων συστημάτων πρωτεΐνης-DNA.
- Πώς μπορούν να γίνουν οι προσομοιώσεις συστημάτων πρωτεΐνης-DNA;

Co-funded by the Erasmus+ Programme of the European Union



Sandor Suhai (1997) Θεωρητική και Υπολογιστική Έρευνα Γονιδιώματος, Springer Science+Business Media Νέα Υόρκη

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 2. Γενετική

Θέμα 3. Υπολογιστική Γονιδιωματική

Μάθημα 1. Αποκωδικοποίηση DNA

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Πλατφόρμες αλληλουχίας
- Αλληλουχία Ολόκληρου Γονιδιώματος και Αλληλουχία Ολόκληρου Μεταγραφώματος

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Η χρήση υπολογιστικών μεθοδολογιών για την ανάλυση βιολογικών δεδομένων δεν είναι πρόσφατη. Ωστόσο, με τη μείωση του κόστους της αλληλουχίας DNA που σχετίζεται με την αύξηση του όγκου των γονιδιωματικών δεδομένων που παράγονται από τις πλατφόρμες προσδιορισμού αλληλουχίας, έχει καταστεί απαραίτητο να χρησιμοποιηθούν υπολογιστικές προσεγγίσεις για το χειρισμό και την εξαγωγή περισσότερων πληροφοριών από τα δεδομένα πλήρων γονιδιωμάτων ή/και μεταγραφήματα χρησιμοποιώντας εργαλεία βιοπληροφορικής.

Co-funded by the Erasmus+ Programme of the European Union

Πλατφόρμες αλληλουχίας

Η αποκωδικοποίηση του DNA σε βιολογικά δείγματα έχει γίνει ένα ουσιαστικό βήμα σε μια ποικιλία ερευνητικών εφαρμογών. Με την πρόοδο των τεχνολογιών, η αλληλουχία του DNA έγινε πιο γρήγορα και ταυτοποιήθηκε με μεγαλύτερη ακρίβεια όσον αφορά τη γενετική σύνθεση και την οργάνωση, η οποία είναι θεμελιώδης πληροφορία για την κατανόηση βιολογικών διεργασιών, εκτός από την καθοδήγηση μεταγονιδιωματικών μελετών όπως η μεταγραφική.

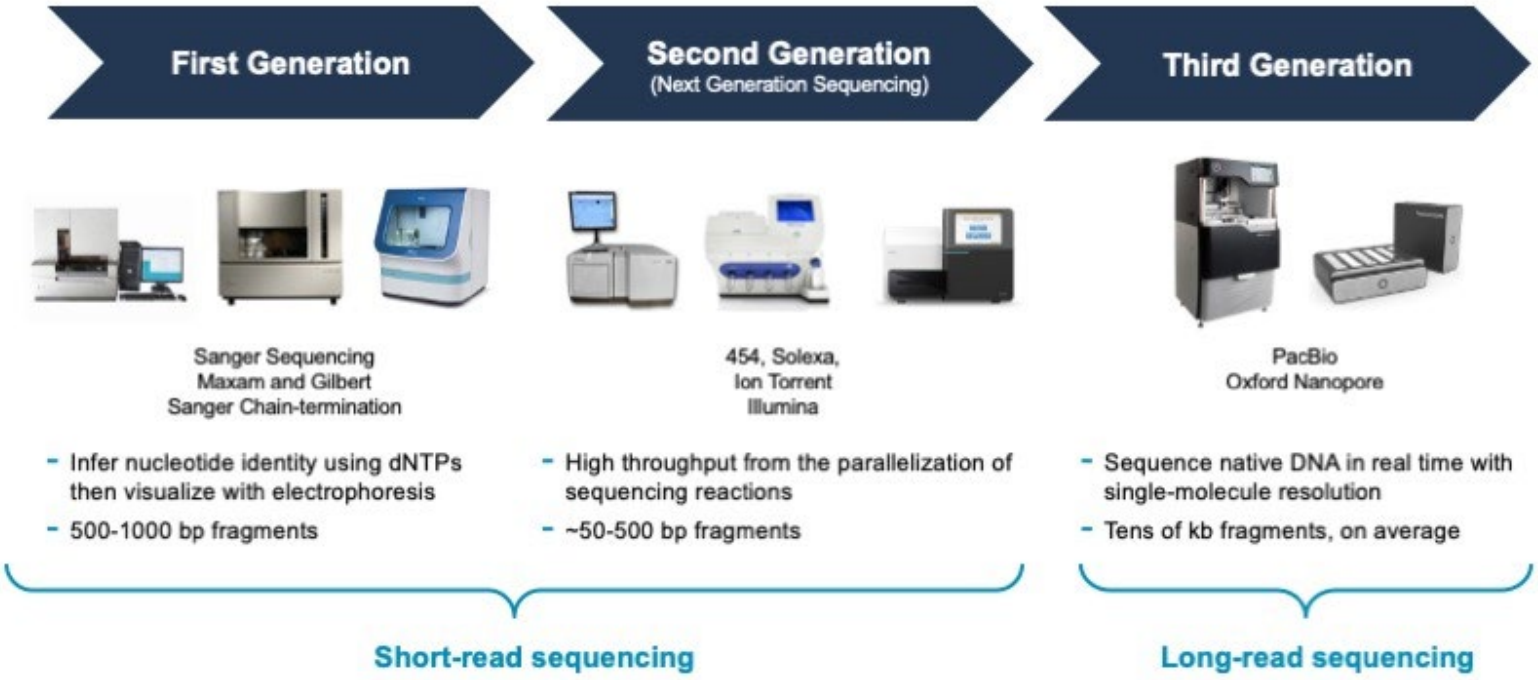
Αλληλουχία επόμενης γενιάς (NGS)

Η αλληλουχία επόμενης γενιάς (NGS), επίσης γνωστή ως αλληλουχία υψηλής απόδοσης, είναι μια ισχυρή τεχνολογία που έχει φέρει επανάσταση στον τομέα της γονιδιωματικής, της μεταγραφομηχανικής, ακόμη και της πρωτεϊνομικής. Επιτρέπει τον γρήγορο και οικονομικά αποδοτικό προσδιορισμό αλληλουχιών DNA και RNA και την ανίχνευση τροποποιήσεων βάσεων, επιτρέποντας την ανάλυση ολόκληρων γονιδιωμάτων, μεταγραφωμάτων και μεταγονιδιωμάτων σε άνευ προηγουμένου ανάλυση και βάθος. Η ικανότητα δημιουργίας τεράστιων ποσοτήτων δεδομένων αλληλουχίας σε σύντομο χρονικό διάστημα έχει φέρει επανάσταση σε πολλούς τομείς της έρευνας και έχει οδηγήσει σε πολυάριθμες ανακαλύψεις: λεπτομερής σχολιασμός των γονιδιωμάτων, καθώς και αναγνώριση γενετικών παραλλαγών που σχετίζονται με ασθένειες, βαθύτερη κατανόηση της γονιδιακής ρύθμισης , κυτταρική διαφοροποίηση και λειτουργικά μονοπάτια και, στη γεωργία, βελτιωμένη αναπαραγωγή καλλιεργειών και διαχείριση παρασίτων.

Αλληλουχία επόμενης γενιάς (NGS)

Σήμερα, υπάρχουν πολλές διαθέσιμες πλατφόρμες αλληλουχίας, καθεμία με τα μοναδικά της χαρακτηριστικά που παρέχουν εναλλακτικές επιλογές για μηχανισμούς αλληλουχίας, μήκη ανάγνωσης, χρόνους εκτέλεσης, ευκολία στη χρήση και επεκτασιμότητα. Οι πλατφόρμες πρώτης γενιάς βασίζονται στην κλασική μέθοδο τερματισμού της αλυσίδας, που στην καθομιλουμένη αναφέρεται ως αλληλουχία Sanger. Οι πλατφόρμες NGS δεύτερης γενιάς βασίζονται στον προσδιορισμό αλληλουχίας με σύνθεση ή σε ανίχνευση αλλαγής pH και δημιουργούν σύντομες αναγνώσεις. Αυτές οι πλατφόρμες περιλαμβάνουν συσκευές Ion Torrent, sequencers από την Illumina, καθώς και από την BGI Group. Οι πλατφόρμες NGS τρίτης γενιάς μπορούν να παράγουν μεγάλες αναγνώσεις, μέσω αλληλουχίας μεμονωμένων μορίων σε πραγματικό χρόνο ή νανοπόρου. Στη συνέχεια, θα εξετάσουμε τις πιο χρησιμοποιούμενες τεχνολογίες αλληλουχίας και τις διαφορετικές αρχές τους. Η κατανόηση των χαρακτηριστικών καθενός από αυτά είναι απαραίτητη για να γνωρίζετε ποιο είναι το καλύτερο κατά το σχεδιασμό των σπουδών σας.

Co-funded by the Erasmus+ Programme of the European Union



Η εξέλιξη των εργαλείων προσδιορισμού αλληλουχίας DNA

Co-funded by the Erasmus+ Programme of the European Union

Διαδικασίες αλληλουχίας

- Η μέθοδος τερματισμού αλυσίδας: Αλληλουχία Sanger
- Αλληλουχία ημιαγωγών
- Αλληλουχία με σύνθεση: ενίσχυση γέφυρας
- Συνδυαστική σύνθεση ανιχνευτή-αγκύρωσης: αλληλουχία νανοσφαιρών
- Αλληλουχία ενός μορίου σε πραγματικό χρόνο (SMRT)
- Αλληλουχία νανοπόρων

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

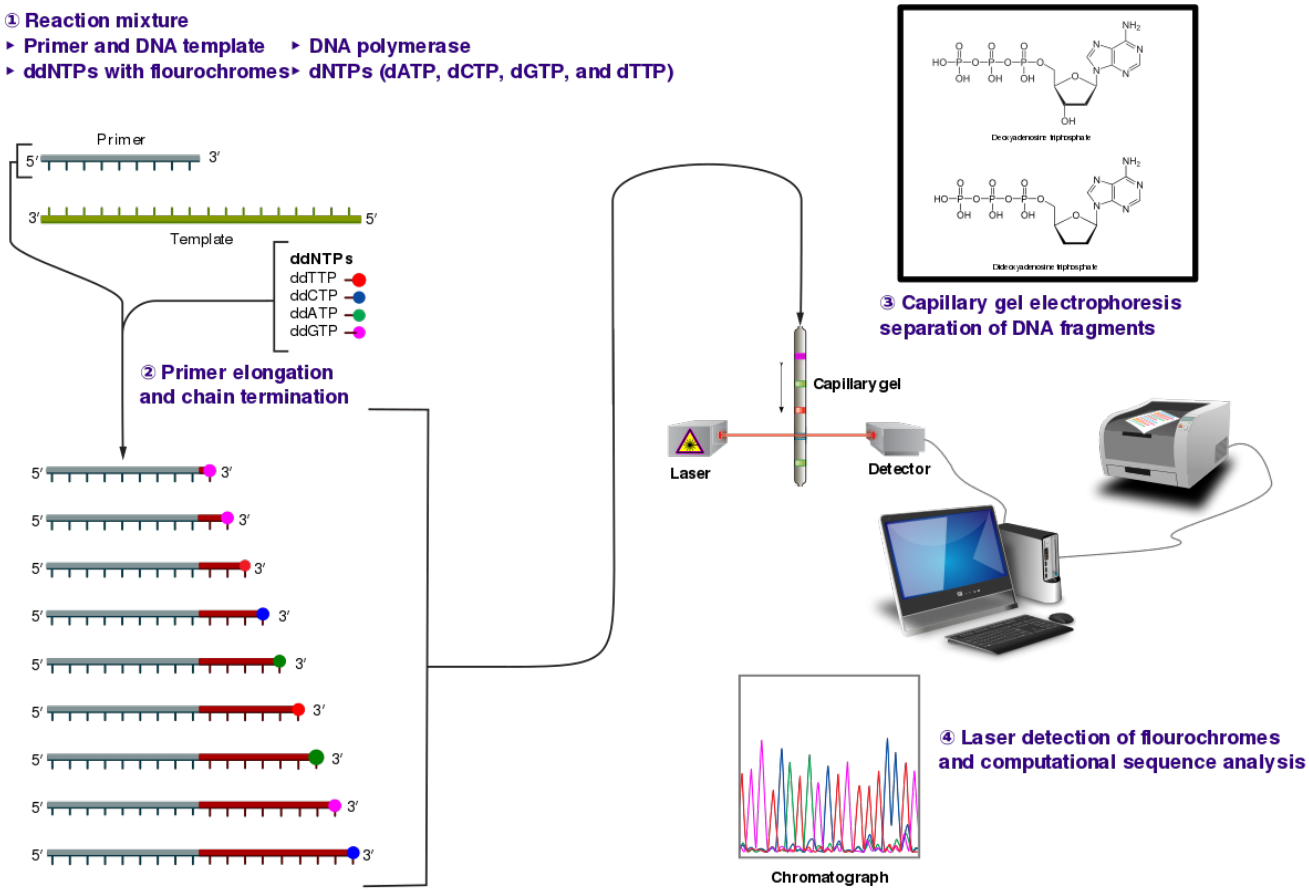
Co-funded by the Erasmus+ Programme of the European Union

Η μέθοδος τερματισμού αλυσίδας: Αλληλουχία Sanger

Η δομή του DNA, που περιγράφεται από τους James Watson και Francis Crick το 1953, έδωσε τη δυνατότητα σε άλλους ερευνητές να εφαρμόσουν μεθοδολογίες προσδιορισμού αλληλουχίας για τον προσδιορισμό της αλληλουχίας νουκλεοτιδίων των νουκλεϊκών οξέων, που ονομάζονται τεχνολογίες αλληλούχισης πρώτης γενιάς. Σε αυτήν την πρώτη γενιά, οι αρχικές προσπάθειες επικεντρώθηκαν στον προσδιορισμό της αλληλουχίας RNA, δημιουργώντας το 1965 από τους Robert Holley et al., την πρώτη πλήρη αλληλουχία νουκλεϊκού οξέος, την αλανίνη tRNA από το Saccharomyces cerevisiae. Εκείνη την εποχή, αρκετοί ερευνητές άρχισαν να προσαρμόζουν τις μεθόδους τους για την αλληλουχία του DNA. Από τα μέσα της δεκαετίας του 1970, ο προσδιορισμός της αλληλουχίας DNA χρησιμοποιήθηκε μέσω της μεθόδου Maxam-Gilbert και της μεθόδου "συν και πλην". Ωστόσο, η ανακάλυψη που άλλαξε τη διαδικασία του τρόπου με τον οποίο αναλύθηκε η αλληλουχία του DNA ήρθε μόλις το 1977, όταν δημοσιεύτηκε η διδεοξυ τεχνική «τερματισμού αλυσίδας» του Sanger.

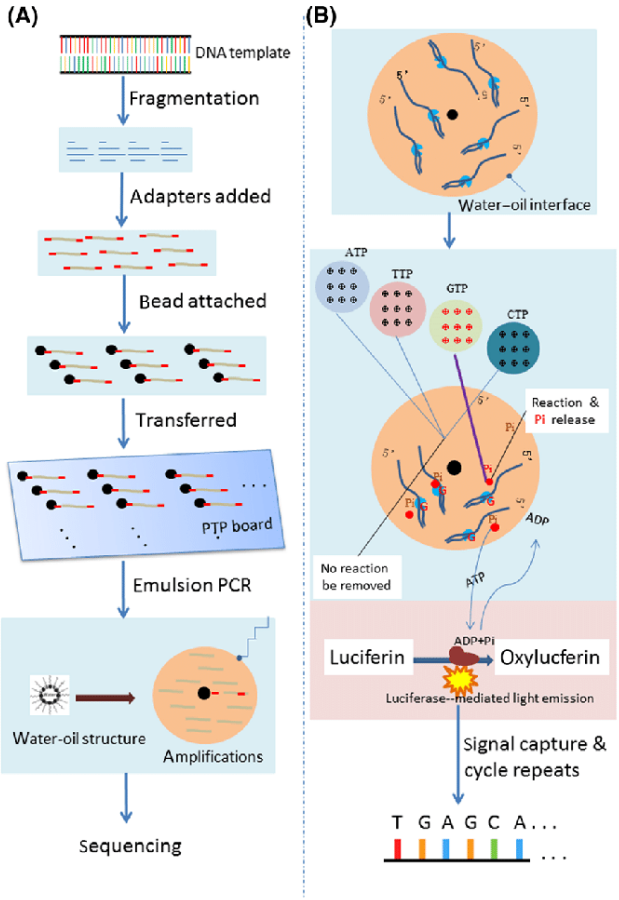
Η μέθοδος Sanger έχει υποστεί πολυάριθμες αλλαγές, όπως η ανάπτυξη ημιαυτόματων sequencers με ηλεκτροφόρηση σε τριχοειδή αγγεία γεμάτα με gel, και ένα σύστημα ανίχνευσης που χρησιμοποιεί ομοεστιακό φθορισμό διεγερμένο από λέιζερ, το οποίο έφερε πλεονεκτήματα όπως: μείωση του χειρισμού τοξικών χημικών ουσιών, εφαρμογή δειγμάτων με ηλεκτροέγχυση και ταυτόχρονη ηλεκτροφόρηση με έως και 384 ανεξάρτητα τριχοειδή αγγεία, για τη δημιουργία θραυσμάτων περίπου 750 ζευγών βάσεων. Το 1986, το εργαστήριο του Leroy Hood στο Caltech (Καλιφόρνια—ΗΠΑ), μαζί με την Applied Biosystems, κυκλοφόρησαν τον πρώτο ημιαυτόματο sequencer, βασισμένο στη μέθοδο Sanger.

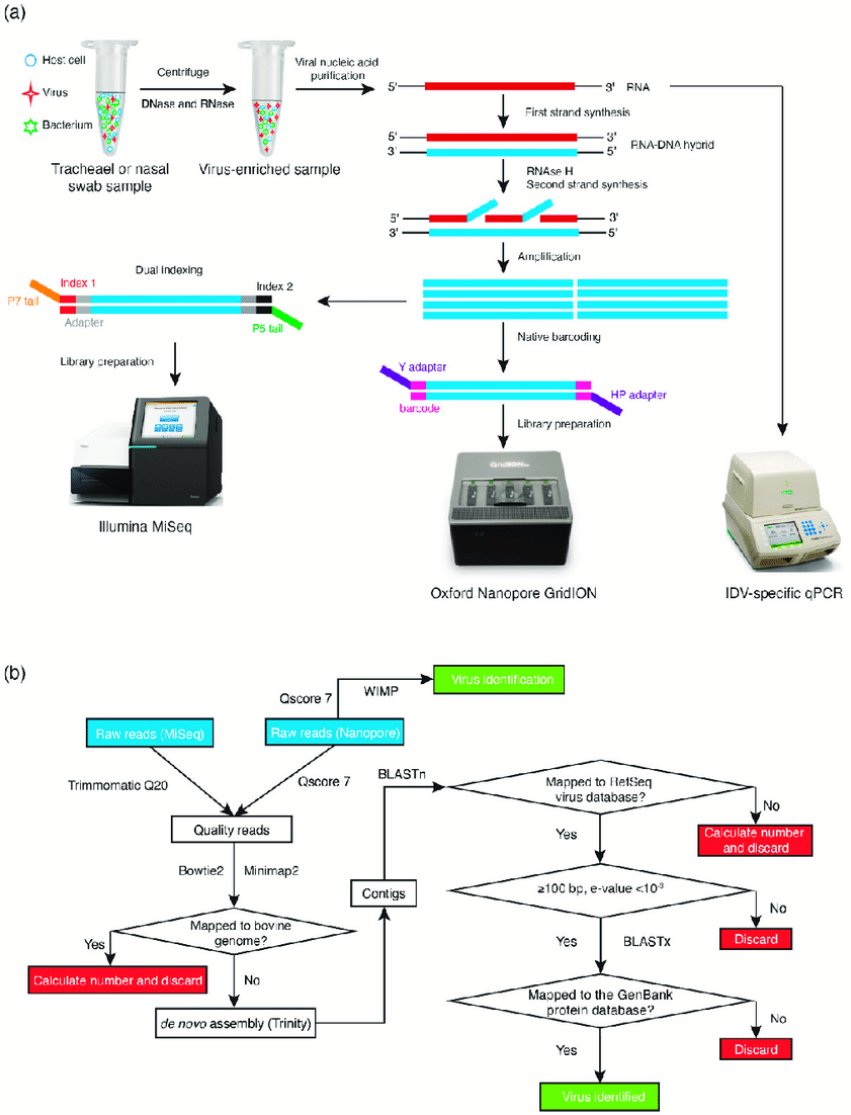
Ακολουθία Sanger



Ταυτόχρονα, με τις προσπάθειες προσδιορισμού αλληλουχίας του Sanger, αναπτύχθηκε μια μέθοδος φωταύγειας για τη μέτρηση της σύνθεσης πυροφωσφορικών. Το Pyrosequencing, όπως ονομάστηκε αυτή η τεχνική, έλαβε στη συνέχεια άδεια στην 454 Life Sciences, μια εταιρεία βιοτεχνολογίας που ιδρύθηκε από τον Jonathan Rothberg. Αυτή η εταιρεία εξελίχθηκε στην πρώτη μεγάλη εμπορική τεχνολογία «επόμενης γενιάς αλληλουχίας» (NGS). Το 2004, η πρώτη μηχανή αλληλουχίας υψηλής απόδοσης (HTS) που ήταν μαζικά διαθέσιμη στο κοινό ήταν η μηχανή 454 GS/20 Roche, η οποία πρόσφερε αύξηση στον αριθμό των αναγνώσεων (έως 100 bp), καθώς και βελτιωμένη ποιότητα δεδομένων . Ο μεγαλύτερος αριθμός αντιδράσεων, που δημιουργήθηκαν σε παράλληλη αλληλουχία σε μικρομετρική κλίμακα, συχνά κατέστη δυνατός ως αποτέλεσμα βελτιώσεων στη μικροκατασκευή και εικόνων υψηλής ανάλυσης, ήταν το σημείο που καθόρισε τη δεύτερη γενιά αλληλουχίας DNA.

Σχηματική αναπαράσταση πυροαλληλουχίας Roche 454.



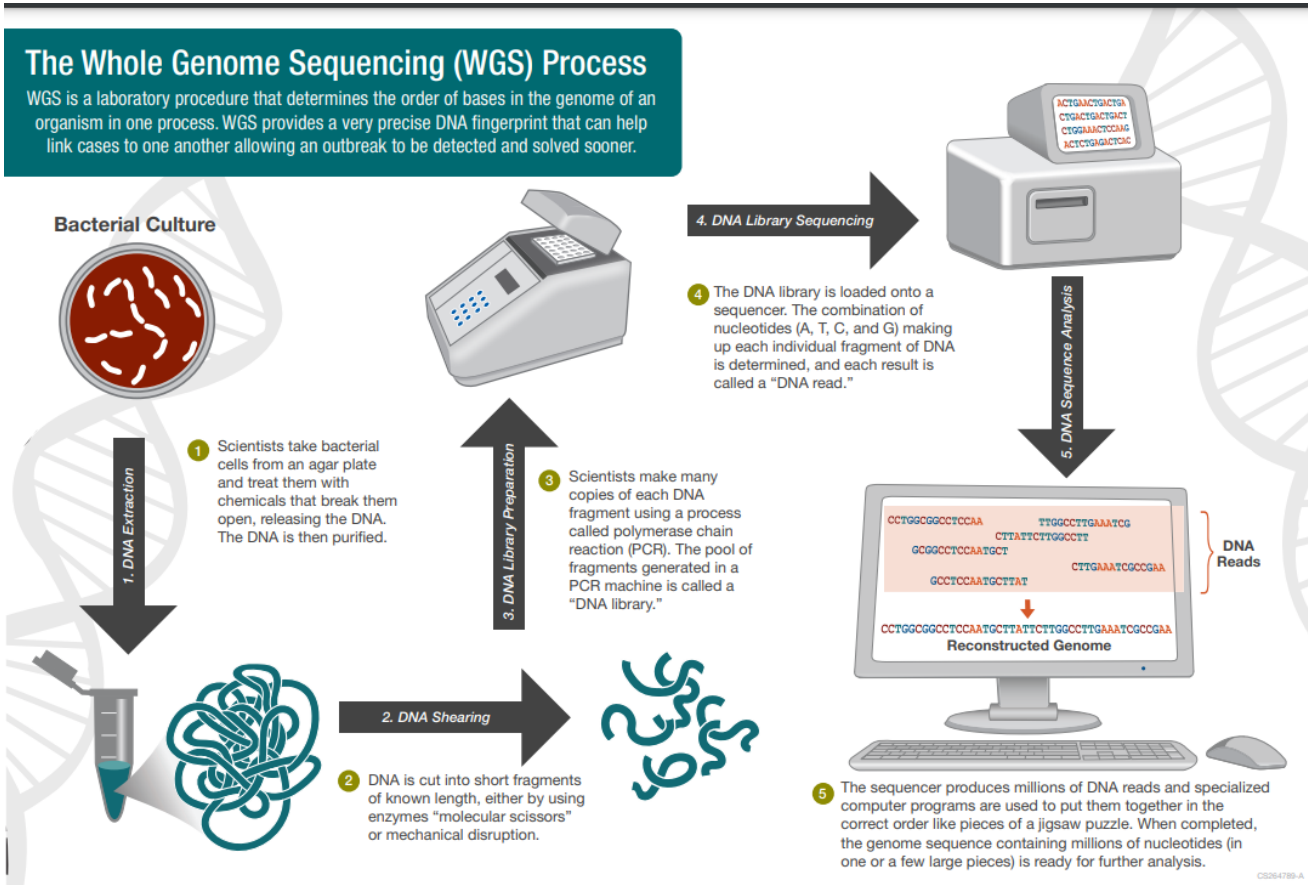


Στις αρχές του 2012, ανακοινώθηκε η πρώτη πλατφόρμα αλληλούχισης νανοπόρου από την Oxford Nanopore, παρουσιάζοντας δύο κύριες εκδόσεις των sequencers: GridION και MinION , ικανές να παράγουν μεγάλες ποσότητες δεδομένων, με απλή προετοιμασία δείγματος που έχει ως αποτέλεσμα μεγάλες αναγνώσεις με χαμηλό κόστος. Το MinION είναι μια μικρή, φορητή συσκευή, ικανή να αναλύει την αλληλουχία 30 Gb DNA, ενώ το GridION μπορεί να δημιουργήσει έως και 150 Gb δεδομένων που μεταδίδονται σε πραγματικό χρόνο για άμεση ανάλυση. Μια άλλη τεχνολογία που λανσαρίστηκε από την ίδια εταιρεία ήταν η PromethION , η οποία μπορεί να δημιουργήσει έως και 8 Tb δεδομένων.

Αλληλουχία Ολόκληρου Γονιδιώματος και Αλληλουχία Ολόκληρου Μεταγραφώματος

Η βελτίωση των τεχνολογιών NGS κατέστησε δυνατή την πραγματοποίηση αλληλουχίας γονιδιώματος και την ολοκλήρωση έργων μεταγραφής σε μεγάλη κλίμακα. Η ανάλυση των γονιδιωμάτων και η πλήρης μεταγραφή επέτρεψαν την ταυτοποίηση της γονιδιακής λειτουργίας εντός του βιολογικού πλαισίου. Για παράδειγμα, με τον προσδιορισμό της αλληλουχίας ενός γονιδιώματος, κατέστη δυνατή η αναγνώριση γονιδίων που γενικά δεν λειτουργούν ανεξάρτητα και οι λειτουργίες τους δεν ελέγχονται απευθείας από τον προαγωγέα, αλλά από πολλά άλλα ρυθμιστικά στοιχεία, όπως ενισχυτές, στοιχεία απόκρισης και σιγαστήρες.

Αλληλουχία Ολόκληρου Γονιδιώματος



Co-funded by the Erasmus+ Programme of the European Union

Επισκόπηση των πιο κοινών πλατφορμών αλληλουχίας

Platform	Advantages	Drawbacks	Recommended applications
Sanger	• Costs (low target number) • Established workflow • Simple data analysis	• Sensitivity • Scalability • Sample input requirements	TRS, validation of NGS data
Ion Torrent	• Costs • Speed	• Short length reads • Accuracy	TRS, metagenomics
Illumina	• Sensitivity • Amount of generated data with same DNA • High throughput	• Costs for low target numbers • Short length reads	WGS, WES, TRS, RNAseq, epigenomics, metagenomics
BGI Group	• Accuracy • No optical duplicates	• Short length reads	WGS, WES, TRS
Pacific Biosciences	• Long reads • High accuracy with CCS mode • Direct detection of epigenetic modifications	• Costs • Large amounts of starting material • Error rate with CLR mode	WGS, TRS, RNAseq
Oxford Nanopore Technologies	• Very long reads • Direct sequencing of RNA • Detection of RNA modifications	• Costs • Error rate • Large amounts of starting material	WGS*, TRS, RNAseq, epigenomics, metagenomics

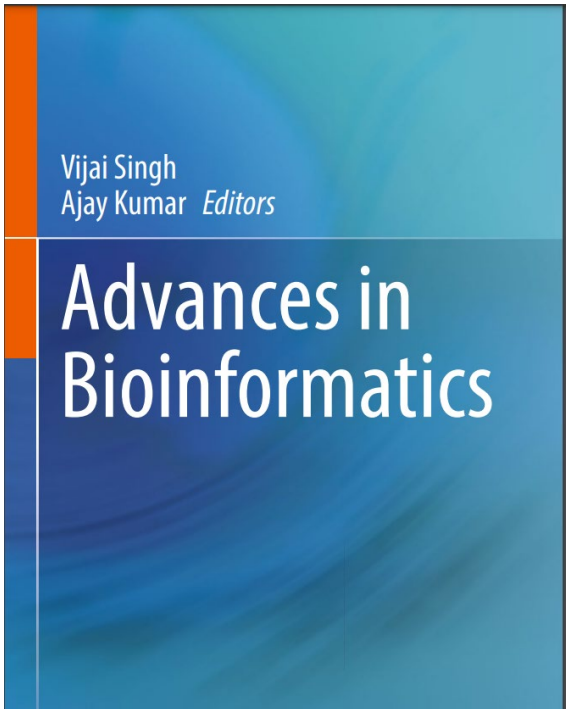
*Small whole genome sequences.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε την αλληλουχία επόμενης γενιάς.
- Ποιες διαδικασίες αλληλουχίας υπάρχουν;

Co-funded by the Erasmus+ Programme of the European Union



Vijai Singh, Ajay Kumar (2021) Advances in Bioinformatics Springer Nature Singapore Pte Ltd

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 2. Γενετική

Θέμα 3. Υπολογιστική Γονιδιωματική

Μάθημα 2. Στοίχιση ακολουθιών

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Ευθυγράμμιση Βιολογικής Αλληλουχίας
- Ζεύγη στοίχισης ακολουθιών και δυναμικός προγραμματισμός
- Ευθυγράμμιση πολλαπλών ακολουθιών (MSA)

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Η σύγκριση των βιολογικών αλληλουχιών μας επιτρέπει να αντιμετωπίσουμε τις διαφορές μεταξύ οργανισμών και ειδών σε επίπεδο γονιδίου. Η συγκριτική γονιδιωματική, ένας κλάδος της επιστήμης που χρησιμοποιεί εξαντλητικά τεχνικές βιοπληροφορικής για την παρακολούθηση γονιδίων σε διάφορα είδη και τη μελέτη των ομοιοτήτων και των διαφορών τους, χρησιμοποιεί αυτές τις μελέτες για να συμπεράνει τα λειτουργικά και δομικά χαρακτηριστικά των πρωτεϊνών που ανακαλύφθηκαν πρόσφατα ή υπάρχουν. Η ανάλυση των βιολογικών ακολουθιών δεν διαφέρει πολύ από τις τεχνικές που χρησιμοποιούνται για τη σύγκριση χορδών και κειμένων και, ως εκ τούτου, η έννοια της ευθυγράμμισης γίνεται πολύ σημαντική. Οι αλληλουχίες που εξελίσσονται σε είδη και κλάδες μέσω μεταλλάξεων περιλαμβάνουν εισαγωγές, διαγραφές (indels) και ασυμβατότητες. Κατά τη σύγκριση δύο βιολογικών αλληλουχιών, δημιουργείται μια ευθυγράμμιση για να απεικονιστούν οι διαφορές μεταξύ των αλληλουχιών σε κάθε θέση.

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Η ευθυγράμμιση αλληλουχίας είναι ένα από τα κύρια καθήκοντα της βιοπληροφορικής. Συνίσταται στην ευθυγράμμιση μιας ακολουθίας ερωτημάτων με μια ακολουθία αναφοράς, η οποία συνήθως βρίσκεται σε μια δημόσια βάση δεδομένων ακολουθιών, με στόχο να καθοριστεί εάν έχουν στατιστικά σημαντικές αντιστοιχίες μεταξύ τους. Διαφέρει από το κλασικό υπολογιστικό πρόβλημα της ακριβούς αντιστοίχισης συμβολοσειρών, όπου υπάρχει ενδιαφέρον για την εύρεση ακριβών αντιστοιχιών. Η στοίχιση συμβολοσειρών είναι ένα πρόβλημα αντιστοίχισης συμβολοσειρών κατά προσέγγιση ή αντιστοίχισης συμβολοσειρών που επιτρέπει σφάλματα. Το πρόβλημα, στην πιο γενική του μορφή, είναι να βρείτε σε ένα κείμενο (ή ακολουθία χαρακτήρων) τη θέση όπου εμφανίζεται ένα συγκεκριμένο μοτίβο, επιτρέποντας έναν περιορισμένο αριθμό σφαλμάτων στις αντιστοιχίες. Η απόσταση μεταξύ των δύο ακολουθιών ορίζεται ως η ελάχιστη ακολουθία πράξεων που απαιτούνται για τη μετατροπή της μίας στην άλλη. Όσον αφορά την πιθανότητα, ένα κόστος αποδίδεται στις πράξεις, έτσι ώστε οι πιο πιθανές πράξεις να κοστίζουν λιγότερο. Στόχος είναι η ελαχιστοποίηση του συνολικού κόστους. Τελικά, ο τελικός στόχος της ευθυγράμμισης αλληλουχίας είναι να προσδιοριστεί η ομοιότητα μεταξύ τμημάτων του γονιδιωματικού κώδικα. Μεταξύ των γνωστών εφαρμογών αυτού του τύπου εργασίας, μπορούμε να αναφέρουμε την ανακάλυψη γονιδίων, την πρόβλεψη της λειτουργίας και τη συναρμολόγηση της αλληλουχίας του γονιδιώματος.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ευθυγράμμιση Βιολογικής Αλληλουχίας

Μια στοίχιση μεταξύ δύο χορδών είναι απλώς η αντιστοίχιση ζευγών μεταξύ των γραμμάτων σε κάθε συμβολοσειρά. Η ευθυγράμμιση των αλληλουχιών νουκλεοτιδίων ή αμινοξέων είναι σε θέση να αντικατοπτρίζει την εξελικτική σχέση μεταξύ δύο ή περισσότερων ομόλογων αλληλουχιών που μοιράζονται έναν κοινό πρόγονο. Εάν το ίδιο γράμμα υπάρχει και στις δύο ακολουθίες, η θέση διατηρήθηκε στην εξέλιξη. Εάν τα γράμματα είναι διαφορετικά, τότε είναι δυνατόν να συμπεράνουμε ότι οι δύο χορδές προέρχονται από ένα προγονικό γράμμα (το οποίο μπορεί να είναι ένα από τα δύο ή κανένα). Ωστόσο, αλληλουχίες που είναι ομόλογες μπορεί να έχουν διαφορετικά μήκη, τα οποία μπορούν εν μέρει να εξηγηθούν με παρεμβολές ή διαγραφές στις αλληλουχίες. Με αυτόν τον τρόπο, ένα γράμμα ή ένα τμήμα γραμμάτων μπορεί να συνδυαστεί με παύλες στην άλλη ακολουθία για να υποδηλώσει αυτήν την εισαγωγή ή εξαίρεση.

```
A T T G C A T C  A A G C T A T A  A A T T G C A A
A T G - A C - C  A A - G - A T A  A A - - C C A A
```

Ζεύγη στοίχισης ακολουθιών και δυναμικός προγραμματισμός

Η στοίχιση κατά ζεύγη συνίσταται στη σύγκριση δύο αλληλουχιών μεταξύ τους για να βρεθεί η καλύτερη δυνατή στοίχιση μεταξύ τους. Η διαδικασία περιλαμβάνει ένα σύστημα βαθμολόγησης για κάθε θέση όπου υπάρχει αντιστοιχία, αναντιστοιχία και indels. Δεδομένου ότι οι αντιστοιχίσεις προτιμώνται από τις διαγραφές, λαμβάνουν συνήθως τις υψηλότερες βαθμολογίες και οι χαμηλότερες βαθμολογίες εκχωρούνται σε εισαγωγές. Η ομοιότητα μεταξύ δύο ακολουθιών είναι αντιστρόφως ανάλογη με τον αριθμό των αναντιστοιχιών και των indels στη στοίχιση. Αναπτύχθηκαν διαφορετικά μοντέλα βαθμολόγησης με βάση τη στατιστικά σχετική συχνότητα του ενός αμινοξέος να γίνεται άλλο.

Υπάρχουν δύο τύποι στοίχισης για ανάλυση ακολουθίας σε ζεύγη με βάση τη μέθοδο δυναμικού προγραμματισμού: Καθολική και Τοπική στοίχιση.

Παγκόσμια ευθυγράμμιση

Ονομάζεται επίσης ευθυγράμμιση από άκρο σε άκρο. Η ιδέα πίσω από τη μέθοδο είναι να προσπαθήσουμε να ευθυγραμμίσουμε όλα τα υπολείμματα σε κάθε ακολουθία. Αυτή η προσέγγιση είναι χρήσιμη όταν οι αλληλουχίες που συγκρίνονται είναι παρόμοιες και έχουν κατά προσέγγιση μέγεθος. Οι Needleman και Wunsch ήταν οι πρώτοι που παρουσίασαν έναν αλγόριθμο ικανό να βρει τη συνολική ευθυγράμμιση μεταξύ δύο αλληλουχιών αμινοξέων. Ο αλγόριθμος βασίζεται σε δυναμικό προγραμματισμό και επιτυγχάνει τη συνολική ευθυγράμμιση δύο ακολουθιών. Ο αλγόριθμος καλύπτει τρία κύρια βήματα: αρχικοποίηση, υπολογισμό και ανίχνευση. Αρχικοποιείται ένας πίνακας διαστάσεων i , j, όπου i και j είναι το μήκος των δύο σειρών σε σύγκριση. Στη συνέχεια, υπολογίζεται η υψηλότερη βαθμολογία F(i,j) για κάθε σύγκριση σε κάθε θέση,

$$F(i,j) = \max\{F(i-1,j-1) + s(Xi,Yi), F(i-1,j) - d, F(i,j-1) - d\},$$

όπου s(Xi, Yi) είναι το σκορ αντιστοίχισης/αναντιστοιχίας και d είναι η ποινή για διαγραφή.

Co-funded by the Erasmus+ Programme of the European Union

Μετά τον υπολογισμό της μέγιστης βαθμολογίας για κάθε θέση στον πίνακα, το ίχνος προς τα πίσω ξεκινά από το τελευταίο κελί (κάτω δεξιά) του πίνακα. Σε κάθε βήμα, μετακινείται από το τρέχον κελί σε αυτό από το οποίο προήλθε η τρέχουσα τιμή κελιού. Μια αντιστοίχιση ή αναντιστοιχία εκχωρείται εάν η μέγιστη βαθμολογία προήλθε από ένα διαγώνιο κελί. Μια εισαγωγή/διαγραφή εκχωρείται εάν η βαθμολογία προήλθε από το επάνω ή το αριστερό κελί. Μετά την ολοκλήρωση του ίχνους πίσω, υπάρχουν δύο ακολουθίες ευθυγραμμισμένες από άκρο σε άκρο με μια βέλτιστη βαθμολογία ευθυγράμμισης.

Co-funded by the Erasmus+ Programme of the European Union

Τοπική ευθυγράμμιση

Αυτός ο τύπος διάταξης είναι πιο χρήσιμος για διαφορετικές ακολουθίες που πιθανώς περιέχουν περιοχές ομοιότητας στο ευρύτερο πλαίσιο της ακολουθίας. Οι Smith και Waterman (Smith et al. 1981) εισήγαγαν έναν διαφορετικό αλγόριθμο για τη βαθμολόγηση των ομοιοτήτων προκειμένου να βρουν τις βέλτιστες υποακολουθίες τοπικής ευθυγράμμισης, ακόμη και με το κόστος της συνολικής βαθμολογίας. Ο αλγόριθμος επιτυγχάνει τοπική ευθυγράμμιση των χορδών και είναι αρκετά παρόμοιος με τη μέθοδο Needleman–Wunsch. Η τοπική στοίχιση μπορεί να χρησιμοποιηθεί σε περιπτώσεις όπου θέλετε να ευθυγραμμίσετε μικρότερες υποσυμβολοσειρές από δύο ακολουθίες. Στο βιολογικό πλαίσιο, μια τέτοια κατάσταση μπορεί να προκύψει κατά την αναζήτηση ενός τομέα ή μοτίβου μέσα σε μεγαλύτερες ακολουθίες. Ο αλγόριθμος περιλαμβάνει τα ίδια βήματα με το Needleman–Wunsch, ωστόσο, με δύο κύριες διαφορές. Ο υπολογισμός της μέγιστης βαθμολογίας περιλαμβάνει επίσης μια επιλογή 0:

$$F(i,j) = \max\{0, F(i-1,j-1) + s(Xi,Yi), F(i-1,j) - d, F(i,j-1) - d\}.$$

		M	V	S	S	D
	0	-2	-4	-6	-8	-10
M	-2	2	0	-2	-4	-6
V	-4	0	4	2	0	-2
S	-6	-2	2	6	4	2
D	-8	-4	0	4	5	6

Alignment 1					
	M	V	S	S	D
	M	V	S	-	D

Alignment 2					
	M	V	S	S	D
	M	V	-	S	D

Needleman–Wunsch matrix. Ο υπολογισμός χρησιμοποιεί βαθμολογίες: +2 για αγώνα, -1 για αναντιστοιχία και
-2 για κενό. Τα βέλη δείχνουν το κελί μήτρας από όπου δημιουργείται η τιμή. Τα κελιά με τιμές με κόκκινο χρώμα δείχνουν το ίχνος προς τα πίσω που δημιουργεί τη στοίχιση

Η ανάθεση του 0 ως μέγιστη βαθμολογία αντιστοιχεί στην αρχή μιας νέας στοίχισης. Αυτό επιτρέπει στις ευθυγραμμίσεις να τελειώνουν οπουδήποτε στον πίνακα. Το ίχνος πίσω, επομένως, ξεκινά από την υψηλότερη τιμή του F(i , j) στον πίνακα και τελειώνει εκεί που βρίσκει το 0.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

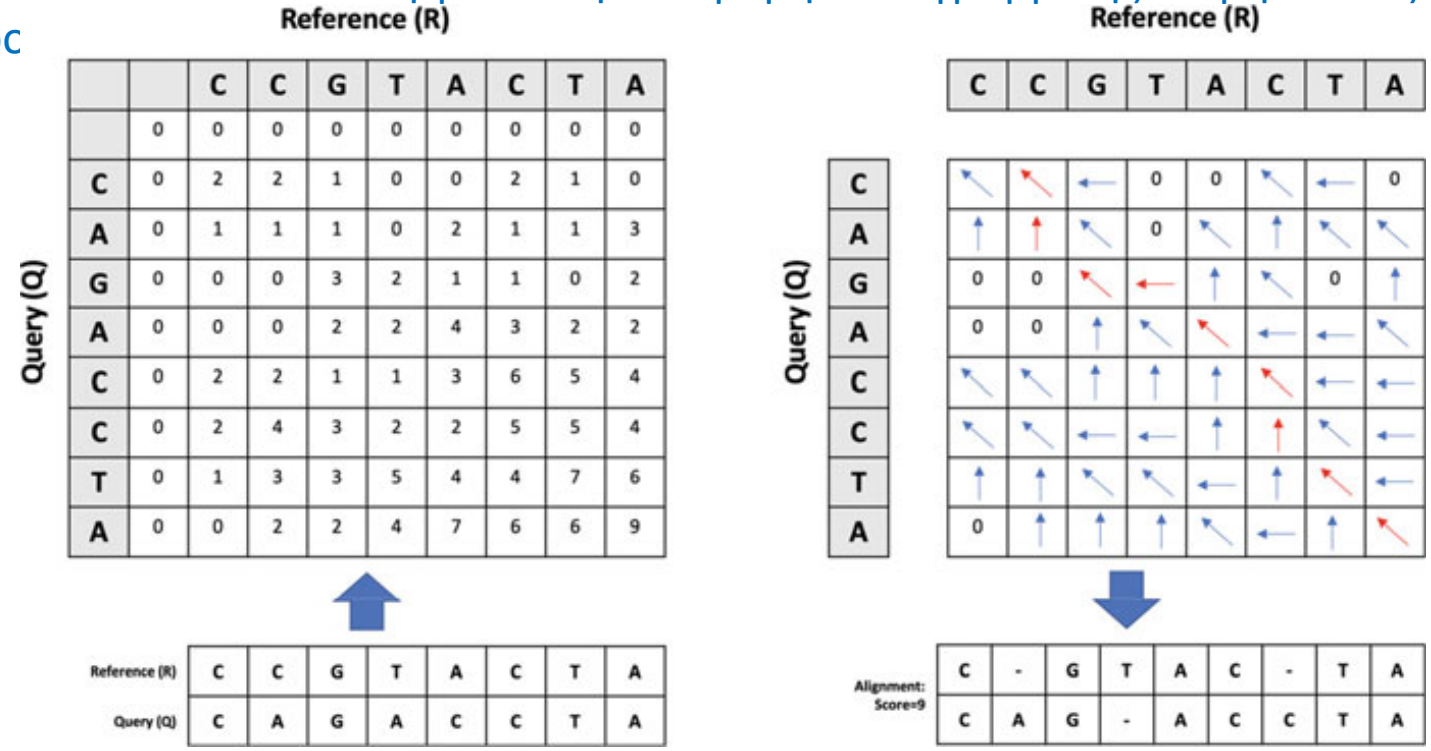
Co-funded by the Erasmus+ Programme of the European Union

Ευθυγράμμιση πολλαπλών ακολουθιών (MSA)

Όταν πρόκειται για ανάλυση βιολογικών αλληλουχιών, μία από τις μεγαλύτερες προκλήσεις είναι η αποκωδικοποίηση του μεγάλου αριθμού και μήκους των αλληλουχιών. Οι βιολογικές βάσεις δεδομένων αποθηκεύουν μια τεράστια ποσότητα πρωτεϊνών και αλληλουχιών DNA και συγκεντρώνουν περισσότερες από 100 εκατομμύρια αλληλουχίες, από τα πιο ξεχωριστά είδη της φύσης. Παρόλο που οι μέθοδοι ευθυγράμμισης που βασίζονται σε δυναμικό προγραμματισμό είναι αρκετά ακριβείς και μπορούν να επιτύχουν καλές ευθυγραμμίσεις με βάση τις βαθμολογίες, είναι αργές και μη πρακτικές για αυτές τις βάσεις δεδομένων με εκατομμύρια ακολουθίες. Η χρονική πολυπλοκότητα των αλγορίθμων δυναμικού προγραμματισμού είναι O(mn), δηλαδή το γινόμενο των μηκών της ακολουθίας. Ως αρχικός τρόπος προσπάθειας βελτίωσης της ταχύτητας σύγκρισης μεταξύ ακολουθιών, δημιουργήθηκαν ευρετικοί αλγόριθμοι όπως BLAST, BLAT και FASTA. Στην ίδια κατεύθυνση, αλγόριθμοι όπως LSCluster , Usearch , Vsearch , Diamond και Ghostx έχουν προταθεί για να προσπαθήσουν να βελτιώσουν την αποτελεσματικότητα αναζήτησης λόγω ομοιότητας.

Γενικά, αυτοί οι αλγόριθμοι αναζητούν ακριβείς αντιστοιχίσεις και επεκτείνουν την ευθυγράμμιση αυτών των αντιστοιχιών, προσπαθώντας να εκτιμήσουν την ιδανική ευθυγράμμιση βαθμολογίας. Έτσι, ευρετικοί αλγόριθμοι με προσεγγίσεις κατά προσέγγιση αντιστοιχίας προσπαθούν να λύσουν τη στοίχιση πολλαπλών ακολουθιών βρίσκοντας ομοιότητες μεταξύ τους, όπως στην περίπτωση της οικογένειας λογισμικού CLUSTAL, η οποία χρησιμοποιεί τον προοδευτικό αλγόριθμο των Feng και Doolittle (1987).

Πίνακας Smith–Waterman με ποινή γραμμικού χάσματος. Ο υπολογισμός χρησιμοποιεί βαθμολογίες: +2 για, -1 για αναντιστοιχία και 1 για κενό. Ο αριστερός πίνακας αντιπροσωπεύει τις ακολουθίες εισόδου και ο δεξιός πίνακας αντιπροσωπεύει τις ακολουθίες είναι ευθυγραμμίσεις. Ο αριστερός πίνακας είναι ο αντίστοιχος (n + 1) επί (m + 1) πίνακας βαθμολογίας. Η δεξιά μήτρα είναι η μήτρα του trace back, με κόκκινα βέλη που υποδεικνύουν τη βέλτιστη διαδρομή ευθυγράμμισης. Ο μηδενικός δείκτης αντιπρς



Co-funded by the Erasmus+ Programme of the European Union

Το BLAST (Βασικό Εργαλείο Αναζήτησης Τοπικής Στοίχισης) είναι ένα λογισμικό που βασίζεται στην ιδέα ότι η καλύτερη στοίχιση ακολουθίας βαθμολόγησης θα πρέπει να περιέχει τον μεγαλύτερο αριθμό πανομοιότυπων αντιστοιχιών ή υπο-ευθυγραμμίσεων με υψηλή βαθμολογία. Ο αλγόριθμος λειτουργεί εκτελώντας τα ακόλουθα βήματα:

1. Μείωση της ακολουθίας ερωτημάτων σε μικρές υποακολουθίες που ονομάζονται seeds.

2. Αναζήτηση για αυτούς τους σπόρους σε ολόκληρη τη βάση δεδομένων αναζητώντας ακριβείς αντιστοιχίσεις.

3. Επέκταση του μεγέθους των ακριβών αντιστοιχιών σε μια ευθυγράμμιση χωρίς κενό έως ότου επιτευχθεί μια μέγιστη επέκταση βαθμολογίας.

Η χρήση των σπόρων για την πρώτη αναζήτηση ακριβών αντιστοιχίσεων αυξάνει σημαντικά ολόκληρη τη διαδικασία αναζήτησης και η ευθυγράμμιση χωρίς κενά χάνει μόνο ένα μικρό σύνολο σημαντικών αντιστοιχίσεων. Η ακρίβεια και η ευαισθησία του BLAST το έχουν καταστήσει έναν από τους πιο ευρέως χρησιμοποιούμενους αλγόριθμους αναζήτησης στον βιολογικό κόσμο. Μια παραλλαγή του BLAST που ονομάζεται Position-Specific-Iterative BLAST (PSI-BLAST) επεκτείνει τον βασικό αλγόριθμο BLAST. Το PSI-BLAST εκτελεί πολλές επαναλήψεις BLAST και χρησιμοποιεί τις επιτυχίες που βρίσκονται σε μία επανάληψη ως ερώτημα για την επόμενη επανάληψη. Αν και το PSI-BLAST ανταποκρίνεται πιο αργά στον μεγάλο αριθμό των απαιτούμενων υπολογισμών, θεωρείται αξιόπιστο εργαλείο για την εύρεση μακρινών σχέσεων ομολογίας.

Co-funded by the Erasmus+ Programme of the European Union

Αν και το BLAST και το PSI-BLAST εξακολουθούν να χρησιμοποιούνται ευρέως, ορισμένες μέθοδοι που αναπτύχθηκαν πρόσφατα προσφέρουν αποτελέσματα με μεγαλύτερη ακρίβεια και ευαισθησία. Τα κρυφά μοντέλα Markov (HMM) έχουν χρησιμοποιηθεί αποτελεσματικά σε πολλές εφαρμογές για την κατανόηση και την εξερεύνηση βιολογικών δεδομένων. Ένα παράδειγμα είναι η γρήγορη αναζήτηση ακολουθίας HMM-HMM (HHblits) (Remmert et al. 2012). Το εργαλείο μπορεί να χρησιμοποιηθεί ως εναλλακτική λύση στο BLAST και το PSI-BLAST και είναι 50–100 φορές πιο ευαίσθητο. Αυτή η υψηλή ευαισθησία του εργαλείου μπορεί να αποδοθεί στον αλγόριθμο που βασίζεται στη σύγκριση των αναπαραστάσεων HMM των ακολουθιών. Αν και οι ευθυγραμμίσεις προφίλ-προφίλ ή HMM-HMM είναι πολύ αργές λόγω υπολογισμών, το προφίλτρο HHblits μειώνει την απαιτούμενη κλίμακα ευθυγράμμισης από εκατομμύρια σε χιλιάδες, αυξάνοντας σημαντικά την ταχύτητά του. Το HHblits αντιπροσωπεύει κάθε ακολουθία στη βάση δεδομένων ως προφίλ HMM. Αυτή η προεπεξεργασία μειώνει τον αριθμό των συγκρίσεων HMM για αναζήτηση ομοιότητας, επιλέγοντας μόνο εκείνες τις αλληλουχίες στόχου όπου υπάρχει η υψηλότερη στοίχιση χωρίς κενό. Στο τέλος, μια στοίχιση Smith–Waterman δείχνει μια σημαντική τιμή E.

Co-funded by the Erasmus+ Programme of the European Union

Υπάρχει ένα άλλο σύνολο μεθόδων που χρησιμοποιούνται για την εκτέλεση της ευθυγράμμισης πολλαπλών ακολουθιών (MSA), ενώ μειώνοντας τα εγγενή σφάλματα σε προοδευτικές μεθόδους, ονομάζονται επαναληπτικές . Αυτές οι κατηγορίες λειτουργούν με παρόμοιο τρόπο με τις προοδευτικές μεθόδους, αλλά ευθυγραμμίζουν εκ νέου τις αρχικές ακολουθίες επανειλημμένα καθώς και προσθέτουν νέες ακολουθίες στο αναπτυσσόμενο MSA. Ένας πολύ χρησιμοποιούμενος αλγόριθμος που βασίζεται στην επανάληψη ονομάζεται MUSCLE (Multiple Sequence Alignment by Log-Expectation) και βελτιώνει την απόδοση των προοδευτικών μεθόδων μέσω μιας πιο ακριβούς μέτρησης απόστασης για την αξιολόγηση της σχέσης μεταξύ δύο ακολουθιών.

Αυτές οι μέθοδοι εκμεταλλεύονται τον χώρο αναζήτησης σε έναν λογαριασμό για να βρουν ένα σχεδόν βέλτιστο αποτέλεσμα. Συνήθως, τα Metaheuristics κατηγοριοποιούνται σε:

1. Μέθοδοι βασισμένες σε μία λύση: αυτή η κατηγορία περιλαμβάνει έναν αλγόριθμο τοπικής αναζήτησης που ασχολείται κυρίως με την τροποποίηση και τη βελτίωση μεμονωμένων πιθανών λύσεων. Η μοναδική μέθοδος που βασίζεται σε λύσεις περιλαμβάνει την Αναζήτηση Tabu (TS),

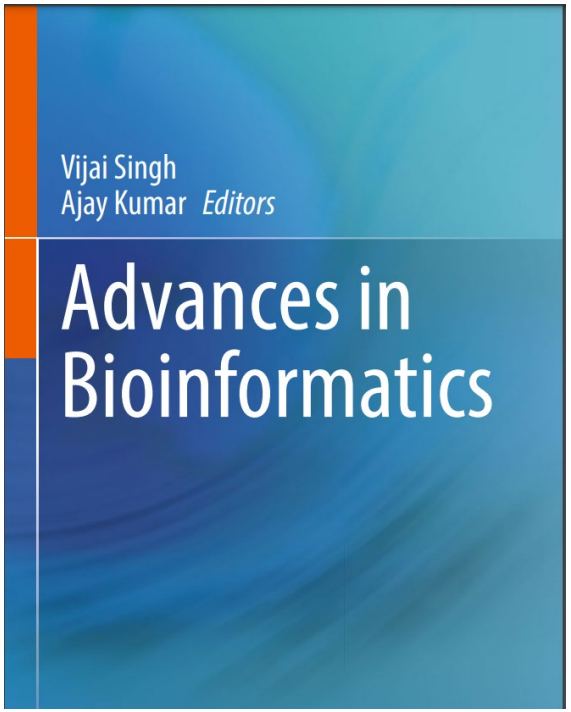
2. Προσομοίωση ανόπτησης (SA), Αναζήτηση μεταβλητής γειτονιάς (VNS), Επαναλαμβανόμενη τοπική αναζήτηση (ILS), αλλά δεν περιορίζεται μόνο σε αυτές τις μεθόδους. 2. Μέθοδοι που βασίζονται στον πληθυσμό: Η μεταευρετική με βάση τον πληθυσμό ξεκινά τη διαδικασία της με τον αρχικό πληθυσμό και συνεχίζει να επαναλαμβάνει έως ότου ικανοποιηθούν τα κριτήρια διακοπής. Αυτές οι μέθοδοι είναι απομιμήσεις ή εμπνευσμένες από φυσικά φαινόμενα. Χρησιμοποιούν βιο-εμπνευσμένους τελεστές όπως επιλογή, διασταύρωση, μετάλλαξη, για να δημιουργήσουν τη δεξαμενή απογόνων από τον προηγούμενο πληθυσμό. Αυτή είναι η κύρια διαφορά μεταξύ των μεθόδων που βασίζονται σε μεταευρετική μεμονωμένη λύση και των μεθόδων που βασίζονται στον πληθυσμό. Οι μεταευρετικές μέθοδοι που βασίζονται σε πληθυσμό περιλαμβάνουν εξελικτικούς αλγόριθμους (EAs), γενετικό αλγόριθμο (GA), διαφορική εξέλιξη (DE), βελτιστοποίηση σμήνος σωματιδίων (PSO), βελτιστοποίηση αποικίας μυρμηγκιών (ACO), βελτιστοποιητή αναζήτησης ομάδας (GSO), τεχνητό ανοσοποιητικό σύστημα (AIS).), και τα λοιπά.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε τους δύο τύπους ευθυγραμμίσεων για ανάλυση ακολουθίας σε ζεύγη με βάση τη μέθοδο δυναμικού προγραμματισμού.
- Τι είναι χαρακτηριστικό για τη στοίχιση πολλαπλών ακολουθιών;

Co-funded by the Erasmus+ Programme of the European Union



Vijai Singh, Ajay Kumar (2021) Προόδους στη Βιοπληροφορική Springer Nature Singapore Pte Ltd

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 2. Γενετική

Θέμα 3. Υπολογιστική Γονιδιωματική

Μάθημα 3. Συναρμολόγηση και σχολιασμός γονιδιώματος

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

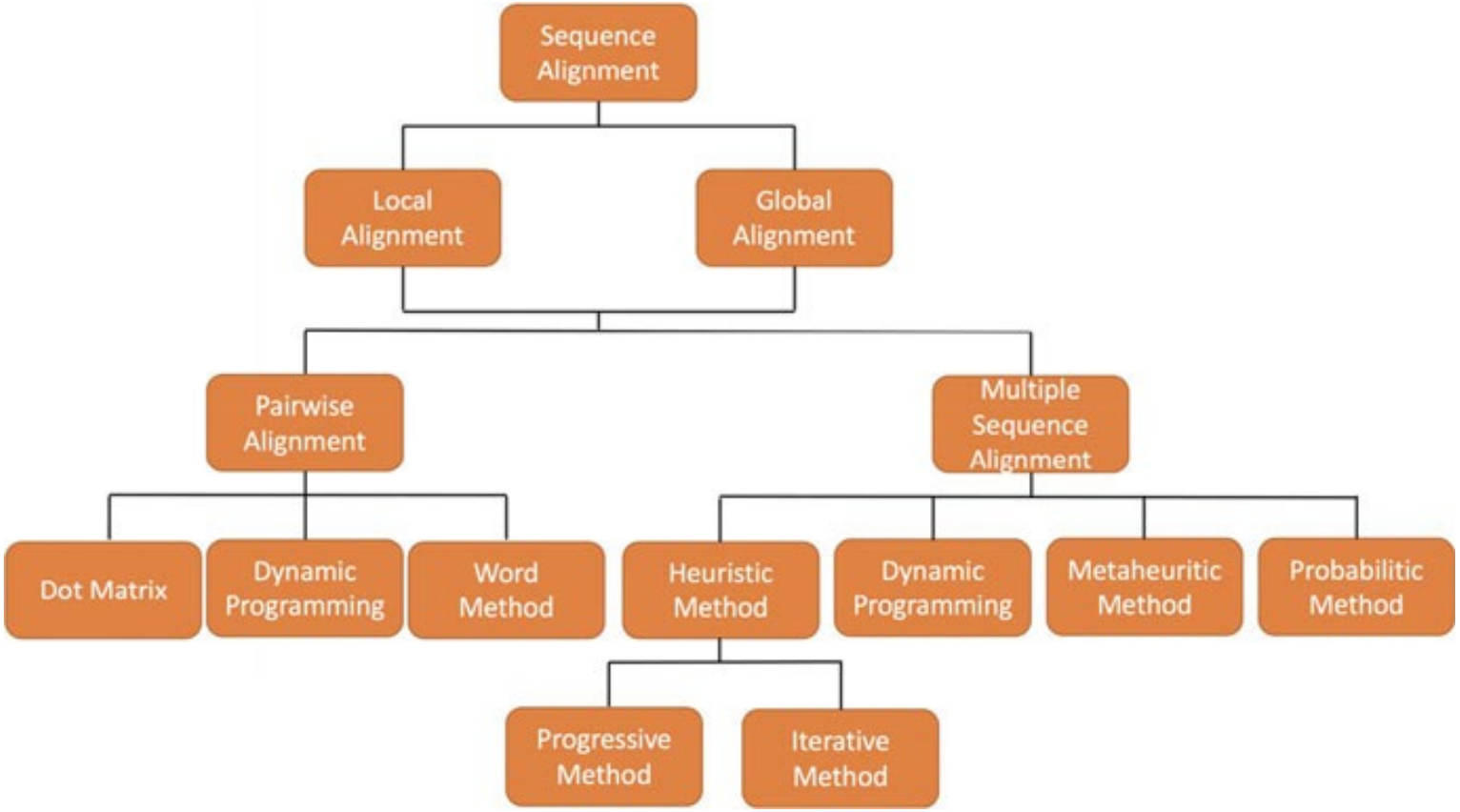
- Εισαγωγή
- Συνέλευση βάσει αναφοράς
- Συνέλευση De Novo
- Υβριδική συναρμολόγηση
- Γονιδιακή Πρόβλεψη και Σχολιασμός

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Από τα δεδομένα που δημιουργούνται από τις τεχνολογίες NGS, έχουν προκύψει αρκετές νέες εφαρμογές, όπως η μελέτη μικροβιακών κοινοτήτων, η ανακάλυψη δομικών παραλλαγών στα γονιδιώματα και η ανάλυση της δομής και της έκφρασης των γονιδίων. Λόγω του μικρού μήκους των ακολουθιών που δημιουργούνται από τις πιο κοινές πλατφόρμες NGS, πολλές από αυτές τις αναλύσεις ξεκινούν με την υπολογιστική διαδικασία συναρμολόγησης ακολουθιών, η οποία συνίσταται στην ομαδοποίηση των παραγόμενων θραυσμάτων με βάση την ταυτότητα βάσης τους. Υπάρχουν δύο γενικές προσεγγίσεις για τη συναρμολόγηση θραυσμάτων NGS: προσεγγίσεις που βασίζονται σε αναφορά και προσεγγίσεις de novo. Στη διάταξη που βασίζεται σε αναφορά, ένα γονιδίωμα αναφοράς του ίδιου οργανισμού ή σχετικού είδους χρησιμοποιείται ως οδηγός για την ευθυγράμμιση των αναγνώσεων. Αυτή είναι, σε πολλές περιπτώσεις, η ανάλυση των επαναλαμβανόμενων δεδομένων.

Διαφορετικές μέθοδοι για τη στοίχιση ακολουθιών



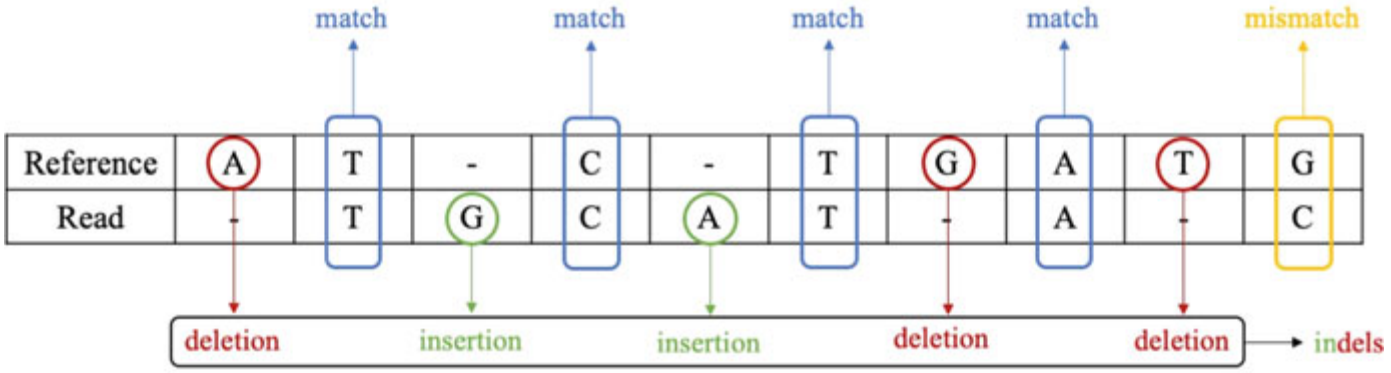
Co-funded by the Erasmus+ Programme of the European Union

Συνέλευση βάσει αναφοράς

Το συγκρότημα που βασίζεται σε αναφορά απαιτεί λιγότερο υπολογιστικό κόστος σε σύγκριση με την de novo προσέγγιση. Είναι μια τεχνική για τον εντοπισμό των διαφορών μεταξύ των αναγνώσεων που λαμβάνονται σε έναν προσδιορισμό αλληλουχίας σε σύγκριση με ένα προηγουμένως διαθέσιμο γονιδίωμα αναφοράς, που χρησιμοποιείται παραδοσιακά στον επαναπροσδιορισμό αλληλουχίας, αλλά δεν περιορίζεται σε αυτό.

Στη διάταξη που βασίζεται σε αναφορά, κάθε ανάγνωση συγκρίνεται με την ακολουθία αναφοράς, βάση προς βάση, προσπαθώντας να χαρτογραφήσει αυτές τις βάσεις. Σε αυτό το σενάριο, υπάρχουν τέσσερις πιθανότητες: αντιστοιχίσεις, αναντιστοιχίες και εισαγωγές και διαγραφές (indels). Μια αντιστοίχιση συμβαίνει όταν μια βάση αντιστοιχίζεται στην ακολουθία αναφοράς. μια αναντιστοιχία συμβαίνει όταν μια βάση δεν χαρτογραφείται έναντι του γονιδιώματος αναφοράς. Μια εισαγωγή συμβαίνει όταν μια βάση είναι παρούσα στην ακολουθία ανάγνωσης αλλά όχι στην ακολουθία αναφοράς, ενώ μια διαγραφή συμβαίνει όταν η βάση είναι παρούσα στην ακολουθία αναφοράς αλλά όχι στην ακολουθία ανάγνωσης. Ο συνδυασμός εισαγωγών και διαγραφών ονομάζεται indels. Για να αξιολογηθεί πόσο καλή είναι η στοίχιση, μετράται ο αριθμός των αντιστοιχιών και στη συνέχεια διαιρείται με το μέγεθος της ακολουθίας. Αυτή η διαίρεση αντιπροσωπεύει το ποσοστό της ταυτότητας.

Reference sequence: ATCTGATG (n = 8)
Read sequence: TGCATAC (n = 7)



matches = 4 mismatch = 1 insertions = 2 deletions = 3 percentage of identity = 4/7 ~ 51%

Λόγω της υψηλής κάλυψης της αλληλουχίας που παρέχεται από τις πλατφόρμες αλληλουχίας επόμενης γενιάς, η ανάλυση παραλλαγών στο γονιδίωμα όπως τα SNP και τα SNV βασίζεται σε αναγνώσεις χαρτογράφησης έναντι ενός γονιδιώματος αναφοράς, όπου η στοίχιση επιτυγχάνει μια ελάχιστη βαθμολογία, που συνήθως αντιπροσωπεύεται από το ποσό αντιστοιχίσεις, αναντιστοιχίες, indels και/ή ποσοστό ταυτότητας.

Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμοι και εργαλεία χαρτογράφησης

Η στοίχιση στο σχήμα δείχνει δύο πολύ απλές και μικρές ακολουθίες για την επίδειξη απλών εννοιών. Στην πραγματικότητα, λαμβάνοντας υπόψη το μέγεθος των αλληλουχιών γονιδιώματος αναφοράς και την ποσότητα των δεδομένων που παράγονται από τις τεχνολογίες NGS, απαιτείται κάποια ισχυρή μεθοδολογία. Η απόδοση χαρτογράφησης των αναγνώσεων σε σχέση με την ακολουθία αναφοράς, με υψηλή ακρίβεια, είναι καθοριστική για την ποιότητα της κατάντη ανάλυσης.

Υπάρχουν πάνω από 50 διαφορετικοί αλγόριθμοι χαρτογράφησης. Τα περισσότερα από αυτά απαιτούν ειδικές δομές δεδομένων, δείκτες, κατασκευασμένους για αναγνώσεις ακολουθιών και της ακολουθίας αναφοράς. Με βάση τον τρόπο με τον οποίο αυτοί οι αλγόριθμοι χρησιμοποιούν τους δείκτες τους, είναι δυνατό να ομαδοποιηθούν σε δύο κατηγορίες: αλγόριθμοι που βασίζονται σε πίνακες κατακερματισμού και αλγόριθμοι βασισμένοι σε μετασχηματισμούς Burrows–Wheeler (BWT).

Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμοι και εργαλεία χαρτογράφησης

Οι αλγόριθμοι που βασίζονται σε πίνακες κατακερματισμού ομαδοποιούνται σε τύπους: αυτοί που κατακερματίζουν το γονιδίωμα και αυτοί που κατακερματίζουν τις αναγνώσεις. Η κύρια ιδέα και για τους δύο τύπους είναι η κατασκευή ενός πίνακα κατακερματισμού για υποακολουθίες αναγνώσεων και γονιδιωμάτων. Το κλειδί κατακερματισμού για κάθε καταχώρηση στον πίνακα κατακερματισμού είναι μια υποακολουθία και η τιμή για αυτό το κλειδί είναι μια λίστα με τις συντεταγμένες όπου μπορεί να εντοπιστεί αυτή η υποακολουθία. Παραδείγματα αλγορίθμων που βασίζονται σε πίνακες κατακερματισμού: GSNAP, FANGS και MAQ.

Οι αλγόριθμοι που βασίζονται σε BWT είναι πολύ αποτελεσματικοί στην ευρετηρίαση δεδομένων και στη διατήρηση μικρής χρήσης μνήμης όταν εκτελείται μια αναζήτηση. Τα τρέχοντα εργαλεία που βασίζονται σε BWT χρησιμοποιούν μια τροποποιημένη έκδοση του αλγορίθμου BWT που χρησιμοποιεί έναν διαφορετικό τύπο δομής δεδομένων, που ονομάζεται δείκτης FM, Created by Ferragina and Manzini (2000). Ο μετασχηματισμός των γονιδιωμάτων σε δείκτη FM βελτιώνει την απόδοση αναζήτησης, βελτιώνοντας τον αλγόριθμο στο σύνολό του. Λόγω της αποτελεσματικότητάς του, οι αλγόριθμοι που βασίζονται σε BWT έγιναν οι πιο χρησιμοποιούμενοι στις εφαρμογές χαρτογράφησης. Σε αυτό το πλαίσιο, δύο λογισμικά ξεχωρίζουν: το BWA και το Bowtie.

Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμοι και εργαλεία χαρτογράφησης

Το Bowtie ξεκινά με την κατασκευή του δείκτη FM για την ακολουθία αναφοράς και στη συνέχεια χρησιμοποιεί μια τροποποιημένη έκδοση του αλγόριθμου χαρτογράφησης Ferragina και Manzini για να εντοπίσει τη θέση της ευθυγράμμισης. Επί του παρόντος, υπάρχουν δύο εκδόσεις παπιγιόν, το Bowtie και το Bowtie 2. Το Bowtie 2 αναπτύχθηκε κυρίως για να χειρίζεται αναγνώσεις μεγαλύτερες από 50 ζεύγη βάσεων, ενώ η πρώτη έκδοση παπιγιόν περιλαμβάνει μόνο ακολουθίες έως και 35 ζεύγη βάσεων.

Το BWA μοιάζει πολύ με το Bowtie. Χρησιμοποιεί επίσης μια τροποποιημένη έκδοση του αλγόριθμου χαρτογράφησης Ferragina and Manzini (2000) για να βρει ακριβείς αντιστοιχίσεις. Για να χειριστεί ανακριβείς αντιστοιχίσεις, το BWA αναζητά αντιστοιχίσεις μεταξύ υποακολουθιών της ακολουθίας αναφοράς λαμβάνοντας υπόψη μια ορισμένη απόσταση. Γενικά, το Bowtie είναι το καλύτερο για τις περισσότερες αναλύσεις, ενώ το BWA έχει καλύτερη απόδοση για μεγαλύτερες αναγνώσεις.

Co-funded by the Erasmus+ Programme of the European Union

Συνέλευση De Novo

Η de novo προσέγγιση συναρμολόγησης βασίζεται στην επικάλυψη των αναγνώσεων ή μέρους αυτών με ένα άλλο. Αυτή η στρατηγική είναι χρήσιμη σε άγνωστα γονιδιώματα: νέα στελέχη ή είδη, και είναι σε θέση να αντιπροσωπεύει περιοχές που δεν μπορούν να αναγνωριστούν με συγκρότημα αναφοράς λόγω της απουσίας τους στο γονιδίωμα αναφοράς.

Για να βελτιωθεί η ακρίβεια της συναρμολόγησης de novo, συνιστάται ιδιαίτερα η αφαίρεση των βάσεων χαμηλής ποιότητας (μετρική Phred) από τα άκρα των αναγνώσεων και η ανάγνωση με χαμηλές βαθμολογίες ποιότητας για την αποφυγή αστοχίας συναρμολόγησης . Μετά την κυκλοφορία των πλατφορμών αλληλουχίας επόμενης γενιάς, η πρόκληση έγινε να ομαδοποιηθούν σύντομες αναγνώσεις (<30 bp) με βάση την ταυτότητά τους για την παραγωγή μεγάλων ακολουθιών (contigs), και με τη σειρά τους τα contigs μπορούν να παραγγελθούν και να προσανατολιστούν για τη δημιουργία ικριωμάτων.

Co-funded by the Erasmus+ Programme of the European Union

Συνέλευση De Novo

Μία από τις κύριες δυσκολίες είναι η συναρμολόγηση επαναλαμβανόμενων περιοχών, μεγαλύτερες από το μήκος αναγνώσεων, του γονιδιώματος και ορισμένες βάσεις ή περιοχές δεν μπορούν να αναπαρασταθούν στο συναρμολογημένο γονιδίωμα, αυτές οι περιοχές ονομάζονται κενά και συνήθως αντιπροσωπεύονται από το N.

Στην πραγματικότητα, το μήκος αναγνώσεων αυξάνεται και η χρήση ζευγαρωμένων βιβλιοθηκών είναι χρήσιμη για την αντιμετώπιση ορισμένων επαναλαμβανόμενων περιοχών για τον προκαρυωτικό, αλλά παραμένει πραγματικό πρόβλημα κατά τη συναρμολόγηση των ευκαρυωτών λόγω των μεγαλύτερων επαναλαμβανόμενων περιοχών.

Οι κύριες στρατηγικές που χρησιμοποιούνται για τη συναρμολόγηση γονιδιωμάτων είναι το Γράφημα Eulerian De Bruijn , το Γράφημα Hamiltonian De Bruijn , το String graph και το Overlap-Layout-Consensus (OLC) που εφαρμόζονται με ορισμένες διαφορές από πολλούς συναρμολογητές, όπως ALLPATHS, Velvet, ABySS , SOAPdenovo , και SPAdes .

Co-funded by the Erasmus+ Programme of the European Union

Συνέλευση De Novo

Πριν από την εκτέλεση οποιασδήποτε προσέγγισης, ο χρήστης μπορεί να υιοθετήσει ορισμένες στρατηγικές για τη διόρθωση σφαλμάτων στις αναγνώσεις, οι περισσότερες από αυτές βασίζονται στη συχνότητα των αναγνώσεων ή k- mers για να ορίσουν τις σίγουρες και εσφαλμένες ακολουθίες.

Η ποιότητα των αποτελεσμάτων που λαμβάνονται από τη διαδικασία συναρμολόγησης του γονιδιώματος μπορεί να αξιολογηθεί με βάση το μήκος, την ποσότητα των βάσεων και το contig που δημιουργείται, και πόσο μεγάλες είναι οι παραγόμενες αλληλουχίες, για να διερευνηθούν τα αποτελέσματα οι κοινές μετρήσεις που πρέπει να αξιολογηθούν είναι: N50—Η Η τιμή N50 σημαίνει ότι το 50% των βάσεων που παράγονται από τη διαδικασία συναρμολόγησης αποτελούν μέρος των contigs με μήκος μεγαλύτερο ή ίσο με την τιμή N50. NG50—το ίδιο για το N50 αλλά το ποσοστό των βάσεων στο γονιδίωμα αναφοράς επιλεγμένο. L50—ο αριθμός των contigs που χρησιμοποιούνται για την επίτευξη της τιμής N50. Θα πρέπει επίσης να αξιολογηθούν και άλλες μετρήσεις, όπως ο αριθμός των παραγόμενων βάσεων, ο αριθμός των contigs και των ασυναρμολογήσεων όταν είναι διαθέσιμο ένα γονιδίωμα αναφοράς για έλεγχο με λογισμικό όπως το Quast.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Υβριδική συναρμολόγηση

Έχουν αναπτυχθεί νέες υβριδικές στρατηγικές για την αξιοποίηση κάθε τύπου συναρμολόγησης. Μεταξύ αυτών, είναι δυνατό να επισημανθούν τεχνικές που συνδυάζουν αναγνώσεις και συγκροτήματα από διαφορετικές τεχνολογίες αλληλουχίας και διαφορετικούς αλγόριθμους συναρμολόγησης που μπορούν να εφαρμοστούν σε διάφορες εργασίες, όπως συγκροτήματα de novo, διόρθωση σφαλμάτων αλληλουχίας και βελτίωση της ποιότητας ακολουθίας. Αυτός ο τύπος υβριδικής συναρμολόγησης χρησιμοποιεί αναγνώσεις από διαφορετικούς αλληλουχιστές για την ανακατασκευή του γονιδιώματος, κυρίως χρησιμοποιώντας μεθόδους που βασίζονται σε επικάλυψη-διάταξη-συναίνεση. Μια άλλη προσέγγιση υβριδικής συναρμολόγησης εμφανίζεται όταν χρησιμοποιούνται διαφορετικοί συναρμολογητές. Αντί να εκτελεί συναρμολόγηση από αναγνώσεις, αυτό το είδος υβριδικής στρατηγικής, επίσης γνωστό ως μετα-συναρμολόγηση, χρησιμοποιεί συγκροτήματα που δημιουργούνται από διαφορετικούς συναρμολογητές, συνδυάζοντας τα αποτελέσματα (συναρμολογητές και/ή ικριώματα) που παράγονται από αυτά τα εργαλεία για την παραγωγή μιας νέας ακολουθίας. Ωστόσο, οι έννοιες του υβριδικού συναρμολογητή και της υβριδικής συναρμολόγησης δεν πρέπει να συγχέονται. Όταν πρόκειται για assemblers, το "υβριδικό" αναφέρεται στην ικανότητα ενός assembler να λειτουργεί με σύντομες και μεγάλες αναγνώσεις, ενώ σε σχέση με τη διαδικασία συναρμολόγησης, το "υβριδικό" αναφέρεται στη χρήση περισσότερων του ενός τύπων στρατηγικής συναρμολόγησης (DBG/ OLC), sequencer (ανεξάρτητα από τον τύπο ανάγνωσης) ή σύνολο δεδομένων εισόδου (read/mount).

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Γονιδιακή Πρόβλεψη και Σχολιασμός

Μετά την ολοκλήρωση της διαδικασίας συναρμολόγησης του γονιδιώματος, έχοντας ένα ολόκληρο γονιδίωμα ή ένα προσχέδιο, η αναγνώριση των Open Read Frames (ORFs), η αλληλουχία μεταξύ του κωδικονίου έναρξης και λήξης, είναι το επόμενο βήμα, ακολουθούμενο από σχολιασμό γονιδίου: η διαδικασία λήψης μεταδεδομένων αφορά το γενετικό προϊόν για κάθε ORF που προσδιορίζεται. Ο σχολιασμός του γονιδίου συχνά βασίζεται σε βιολογική βάση δεδομένων που δείχνει τη λειτουργία, τα προϊόντα και τις διαδικασίες που μπορεί να εμπλέκεται το γονίδιο πέρα από άλλες πληροφορίες. Ορισμένες μέθοδοι που υιοθετούνται για την πρόβλεψη γονιδίων βασίζονται σε ένα σύνολο δεδομένων εκπαίδευσης, επομένως όταν αυτό το σύνολο γονιδίων που επιλέγεται για εκπαίδευση δεν είναι καλό, μπορεί να οδηγήσει σε άσχημα αποτελέσματα. Για την εκπαιδευτική εργασία, τα περισσότερα προγράμματα σήμερα χρησιμοποιούν μοντέλα Markov (HMM—Hidden Markov Models ή IMM—Interpolated Markov Models) (για παράδειγμα, SNAP; GlimmerHMM ; GeneMark ; GlimmerIMM) για αυτήν την εκπαίδευση, όπου τα γονίδια μοντελοποιούνται με τα μοντέλα Markov που χρησιμοποιήστε μια σειρά καταστάσεων για να αναπαραστήσετε μια γενική δομή των γονιδίων. Η εκπαίδευση δεδομένων για προγράμματα πρόβλεψης γονιδίων και σχολιασμού επιλέγεται συχνά τυχαία από ένα υποσύνολο γονιδίων υψηλής ποιότητας που αντιπροσωπεύουν ιδανικά την παραλλαγή που βρίσκεται σε ένα γονιδίωμα. Όταν τα προγράμματα για γονιδιακή πρόβλεψη και σχολιασμό εκπαιδεύονται σε ένα υποσύνολο γονιδίων χόρτου με τυχαίο περιεχόμενο GC, ουσιαστικά εκπαιδεύονται σε δύο κατηγορίες γονιδίων ταυτόχρονα και αυτό μπορεί να έχει ως αποτέλεσμα κακή απόδοση όταν τα γονίδια προβλέπονται σε νέες αλληλουχίες γονιδιώματος .

Γονιδιακή Πρόβλεψη και Σχολιασμός

Στην πραγματικότητα, η αλληλουχία του ανθρώπινου γονιδιώματος μπορεί να γίνει για λιγότερο από χίλια δολάρια. Λόγω αυτής της μείωσης στην τιμή αλληλουχίας, υπήρξε πρόοδος στους αλγόριθμους συναρμολόγησης και ευθυγράμμισης. Ως αποτέλεσμα, η απόκτηση ενός προσχέδιο συναρμολόγησης υψηλής ποιότητας έγινε εφικτός στόχος για τα περισσότερα έργα γονιδιώματος. Αυτό προκάλεσε το σημείο συμφόρησης στις γονιδιωματικές μελέτες να αλλάξει εστίαση, ο σχολιασμός του γονιδιώματος έχει γίνει μια πρόκληση λόγω της δυσκολίας συλλογής ή πρόβλεψης πρωτεϊνών, κυρίως για μεγάλα γονιδιώματα, που απαιτούν άλλες πηγές δεδομένων, όπως RNA-Seq και βάσεις δεδομένων για εκπαίδευση, βελτιστοποίηση. και να διαμορφώσετε τα εργαλεία σχολιασμού γονιδίων.

Η μη αυτόματη επιμέλεια μπορεί να χρησιμοποιηθεί για τη βελτίωση της ποιότητας του σχολιασμού γονιδίων για την περιγραφή των Όρων Γονιδιακής Οντολογίας (GO) ή των γονιδιακών προϊόντων που βασίζονται στη βάση δεδομένων βιολογικών σχολιασμών, όπως το Blast2GO και το GoFeat , τα οποία χρησιμοποιούν σχολιασμένα γονίδια και τη δομική τους ομοιότητα για τη λήψη νέων πληροφοριών και γνώσεις, που βασίζονται πάντα σε υπολογιστικές προσεγγίσεις που αντιπροσωπεύουν τους περισσότερους σχολιασμούς που βρίσκονται στις βιολογικές βάσεις δεδομένων.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Γονιδιακή Πρόβλεψη και Σχολιασμός

Η ακρίβεια του σχολιασμού των γονιδίων είναι απαραίτητη για την επόμενη ανάλυση για την αξιολόγηση των γονιδίων που βρέθηκαν και της σχέσης τους στον οργανισμό, η οποία θα οδηγήσει σε ανακαλύψεις σχετικά με λειτουργίες και φαινοτύπους που μπορούν να συσχετιστούν με τον οργανισμό σε πολλές εφαρμογές, όπως αλληλεπιδράσεις παθογόνου-ξενιστή και αντιβιοτικά αντίσταση.

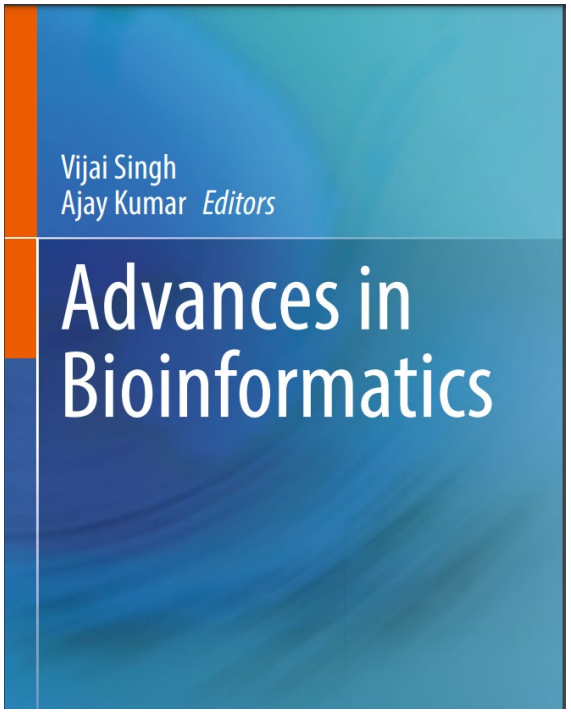
Ένας από τους κύριους περιορισμούς στην εργασία του γονιδιωματικού σχολιασμού προέρχεται από την ανάγκη για βάση δεδομένων με σχολιασμούς που έχουν ήδη γίνει. Υπάρχουν ορισμένοι τομείς της βιολογίας που έχουν μελετηθεί περισσότερο και επομένως έχουν περισσότερα δεδομένα (πλήρη δεδομένα, καλύτερα περιγραφόμενα και μερικές φορές επιμελημένα) για γνωστές διαδικασίες, πέρα από τον αριθμό των βάσεων δεδομένων για συγκεκριμένη ανάλυση που δεν είναι ενσωματωμένες σε μεγάλες βάσεις δεδομένων όπως η Genbank , DDBJ και EBI. Σήμερα με την εξέλιξη των προγραμμάτων σχολιασμού, τα περισσότερα από αυτά είναι πλέον αυτοματοποιημένα, για παράδειγμα, RAST, PATRIC. Αυτοί οι αγωγοί έχουν βασικά δύο καθήκοντα: αναζήτηση μοτίβων που προσδιορίζουν το γονίδιο του είδους (π.χ. ESTs—Expressed Sequence Tag, πρωτεΐνες, RNA-Seq) και χαρακτηρισμό αυτών των μοτίβων σε μια βάση δεδομένων (π.χ. Interpro , Uniprot , Pfam χρησιμοποιώντας Blast ή Diamond .

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Ποιοι τύποι συναρμολόγησης γονιδιώματος υπάρχουν;
- Εξηγήστε το καθένα από αυτά.

Co-funded by the Erasmus+ Programme of the European Union



Vijai Singh, Ajay Kumar (2021) Advances in Bioinformatics Springer Nature Singapore Pte Ltd

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 2. Γενετική

Θέμα 3. Υπολογιστική Γονιδιωματική

Μάθημα 4. Δίκτυα Βιολογικής Αλληλεπίδρασης

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Ιδιότητες Βιολογικού Δικτύου
- Τύποι Βιολογικών Δικτύων

✓Μεταβολικά Δίκτυα

✓Δίκτυο σηματοδότησης

✓Δίκτυο γονιδιακής ρύθμισης

✓Δίκτυο αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης

✓Δίκτυο Βιολογικής Συνέκφρασης

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

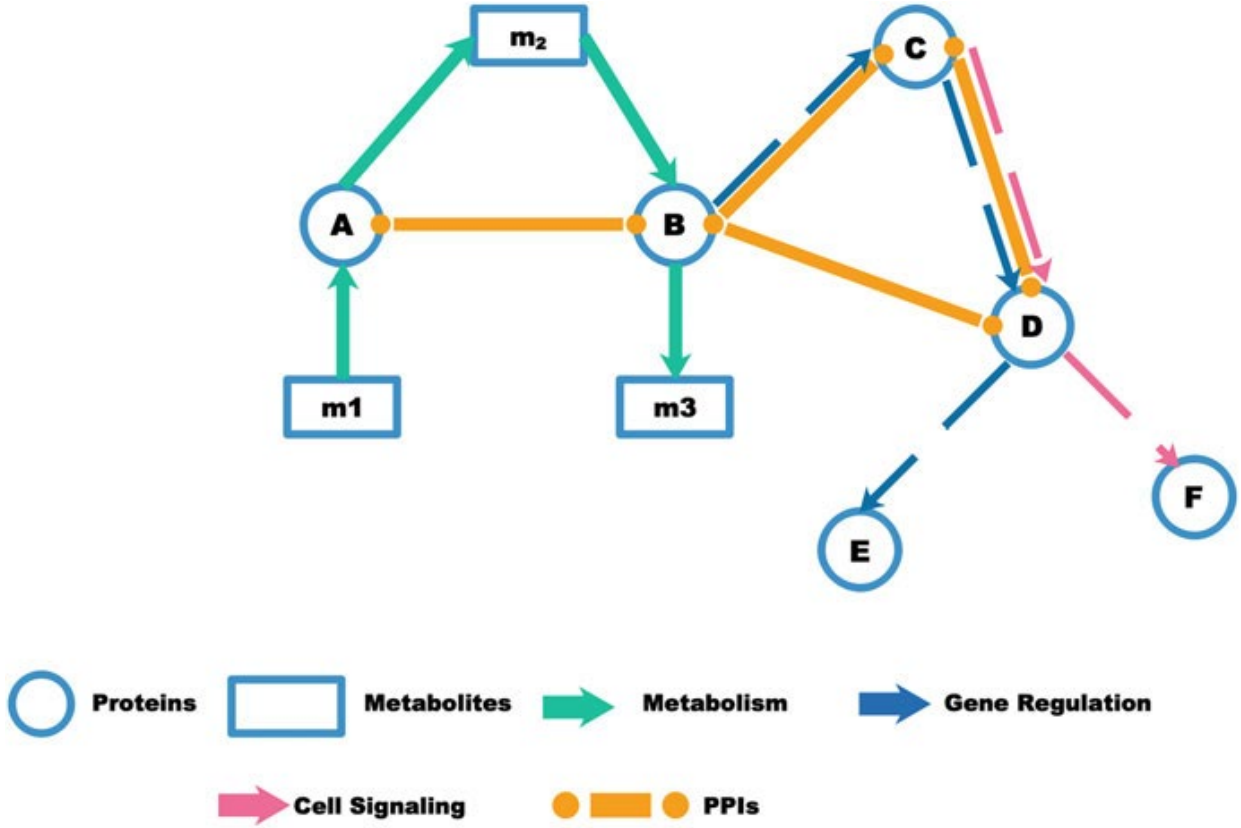
Τα βιολογικά δίκτυα χρησιμοποιούνται σε διάφορες βιολογικές επιστήμες, όπως η μελέτη της αλληλεπίδρασης, η μελέτη καρκίνου, η πρόβλεψη φαρμάκων, η ανάλυση μεταγονιδιώματος, η πρωτεομική ανάλυση, οι μοριακές αλληλεπιδράσεις και οι αλληλεπιδράσεις κυττάρων, μεταξύ άλλων τομέων.

Ένα βιολογικό δίκτυο μπορεί να οριστεί ως μια συλλογή μονάδων (βιομόρια), που δυνητικά αλληλεπιδρούν ως σύστημα. Με άλλα λόγια, ένα δίκτυο βιολογικών αλληλεπιδράσεων μπορεί να αναπαρασταθεί μια αφαίρεση των αλληλεπιδράσεων που λαμβάνονται μέσω μαθηματικών ή υπολογιστικών μοντέλων, όπου αναπαρίσταται ένα ομοιόμορφο σύνολο κόμβων που συνδέονται με ένα ομοιόμορφο σύνολο ακμών που μπορούν να κατευθυνθούν ή να μην κατευθυνθούν. Σε αυτόν τον τύπο δικτύου, οι κόμβοι μπορούν να αντιπροσωπεύουν βιομόρια (γονίδια, πρωτεΐνες, νευρώνες, οργανισμούς, κύτταρα, μεταξύ άλλων) και τα όρια συνήθως αντιπροσωπεύουν σχέσεις και αλληλεπιδράσεις (βιοχημικές, μεταγραφικές, ροή ενέργειας, ρύθμιση, συνέκφραση, μεταβολικές, μεταξύ άλλων).

Η μελέτη και η ανάλυση των δικτύων είναι μέρος της βιολογίας του δικτύου. Αυτό το παράδειγμα μας επιτρέπει να κατανοήσουμε τις πολύπλοκες αλληλεπιδράσεις των βιομορίων μέσα στα κύτταρα αναπαριστάνοντας και αναλύοντας βιολογικά συστήματα μέσω εργαλείων και μεθόδων που προέρχονται από τη θεωρία γραφημάτων, τα μαθηματικά, τη φυσική, τη στατιστική, τη μηχανική μάθηση και άλλα.

Τα συμπεράσματα των βιολογικών δικτύων που χρησιμοποιούν δεδομένα NGS επιτρέπουν τη λήψη σχετικών πληροφοριών σχετικά με τις διαδικασίες έκφρασης και ρύθμισης μέσα στους οργανισμούς. Τα δίκτυα βιολογικών αλληλεπιδράσεων μπορούν να κατασκευαστούν χρησιμοποιώντας διαφορετικές μεθόδους αντίστροφης μηχανικής που χρησιμοποιούν δεδομένα υψηλής και χαμηλής απόδοσης, καθώς και στατιστικές, μαθηματικές και υπολογιστικές τεχνικές που επιτρέπουν την ανακατασκευή του τρόπου με τον οποίο τα στοιχεία των βιολογικών δικτύων ενσωματώνονται ως σύστημα.

Τύποι βιολογικών αλληλεπιδράσεων που μπορούν να αναπαρασταθούν από δίκτυα



Τα βιολογικά δίκτυα έχουν χρησιμοποιηθεί για τη μελέτη των διαδικασιών μεταγραφής-ρύθμισης στο Escherichia Coli, όπου έχει αποδειχθεί ότι τα βασικά μοριακά στοιχεία συμβάλλουν στην εξειδίκευση της δυναμικής των παγκόσμιων αποκρίσεων, γεγονός που επιτρέπει στα βακτήρια να έχουν μια πιο ισχυρή και ταχύτερη απόκριση σε διαδικασίες και περιβαλλοντικές πινακίδες. Χρησιμοποιούνται επίσης για τη μελέτη δικτύων μεταβολικών αλληλεπιδράσεων. μελέτη αλληλεπιδράσεων ξενιστή-παθογόνου, ανακάλυψη νέων μετρήσεων, αναγνώριση βιοδεικτών. ταυτοποίηση γονιδίων που εμπλέκονται σε συγκεκριμένες διαδικασίες κυτταρικού κύκλου· ταυτοποίηση γονιδίων που σχετίζονται με ασθένειες κρίσιμες βιολογικές διεργασίες. Τα δίκτυα έχουν χρησιμοποιηθεί για τη δημιουργία μοντέλων της σχέσης μεταξύ στοιχείων συνόλων βιολογικών δεδομένων, καθώς και για την ανάλυση του σχηματισμού χρωματίνης μέσα στα κύτταρα. την αναγνώριση των μεταβολικών οδών που σχετίζονται με τη γενετική ρύθμιση· καθώς και να μοντελοποιήσουν τις αλληλεπιδράσεις πρωτεΐνης-πρωτεΐνης που λαμβάνουν χώρα εντός των οργανισμών.

Τα δίκτυα βιολογικών συμπερασμάτων μπορούν να ταξινομηθούν σε τέσσερις τύπους: πρωτεϊνική πρωτεΐνη , δίκτυα γονιδιακής ρύθμισης, μεταβολικά δίκτυα, δίκτυα σηματοδότησης και δίκτυα συνέκφρασης.

Ιδιότητες Βιολογικού Δικτύου

Τα δίκτυα βιολογικών αλληλεπιδράσεων έχουν συγκεκριμένες αρχιτεκτονικές και ιδιότητες που επιτρέπουν την ανάλυση και την ερμηνεία της πολυπλοκότητας των αλληλεπιδράσεων που υπάρχουν μέσα στους διαφορετικούς τομείς και στοιχεία που υπάρχουν σε αυτά. Μερικά από τα στοιχεία και τις ιδιότητες που αποτελούν μέρος των δικτύων είναι τα εξής:

• Κόμβος: είναι ένα μεμονωμένο στοιχείο εντός του δικτύου.

• Edge: αντιπροσωπεύει τις αλληλεπιδράσεις και τις διασυνδέσεις μεταξύ κόμβων.

• Components: είναι μια ομάδα ή ομάδες κόμβων που συνδέονται αμοιβαία.

• Βαθμός/συνδεσιμότητα: ο αριθμός των άκρων που συνδέονται μεταξύ ενός εστιακού κόμβου και

οι άλλοι κόμβοι?

• Πυκνότητα δικτύου: περιγράφει το τμήμα των πιθανών συνδέσεων σε ένα δίκτυο που είναι πραγματικές συνδέσεις. Μια πιθανή σύνδεση είναι μια σύνδεση που μπορεί να υπάρξει μεταξύ δύο κόμβων, ανεξάρτητα από το αν υπάρχουν ή όχι.

• Μεταξύ: είναι η μέτρηση που μετρά πώς ένας κόμβος βρίσκεται στη διαδρομή μεταξύ των άλλων κόμβων. Οι κόμβοι με υψηλή κεντρικότητα μπορεί να έχουν ισχυρή επιρροή λόγω του ελέγχου τους στη μετάδοση πληροφοριών εντός του δικτύου.

Ιδιότητες Βιολογικού Δικτύου

• Κοντότητα: είναι το μέτρο της συντομότερης διαδρομής μεταξύ ενός κόμβου και των άλλων κόμβων εντός του δικτύου.

• Συντελεστής ομαδοποίησης: είναι ένα μέτρο της αναλογίας των γειτόνων που επιτυγχάνονται μέσω ενός κόμβου στους άλλους γείτονες. Αυτή η μέτρηση δείχνει τον βαθμό στον οποίο οι κόμβοι σε ένα δίκτυο τείνουν να ομαδοποιηθούν. Αυτή η μέτρηση επιτρέπει τη μέτρηση της συνοχής του δικτύου.

• Βαθμός κατανομής: είναι η κατανομή των συχνοτήτων των βαθμών των κόμβων μεμονωμένα για ένα ολόκληρο δίκτυο.

• Ενότητες ή ομαδοποίηση: ένα σύνολο πυκνά διασυνδεδεμένων κόμβων εντός του δικτύου.

• Μοτίβο: είναι μικρά υποδίκτυα ή μοτίβα που υπερεκπροσωπούνται στατιστικά μέσα στο δίκτυο.

• Κλίκα: αποτελείται από ένα πλήρως συνδεδεμένο υποδίκτυο σε ένα δεδομένο δίκτυο.

• Κατευθυνόμενο γράφημα: οι κόμβοι σε ένα κατευθυνόμενο γράφημα συνδέονται με μια ασύμμετρη σχέση, όπως η αρπαγή.

• Μη κατευθυνόμενο γράφημα: οι κόμβοι σε ένα μη κατευθυνόμενο γράφημα συνδέονται με μια συμμετρική σχέση, όπως οι φυσικές αλληλεπιδράσεις.

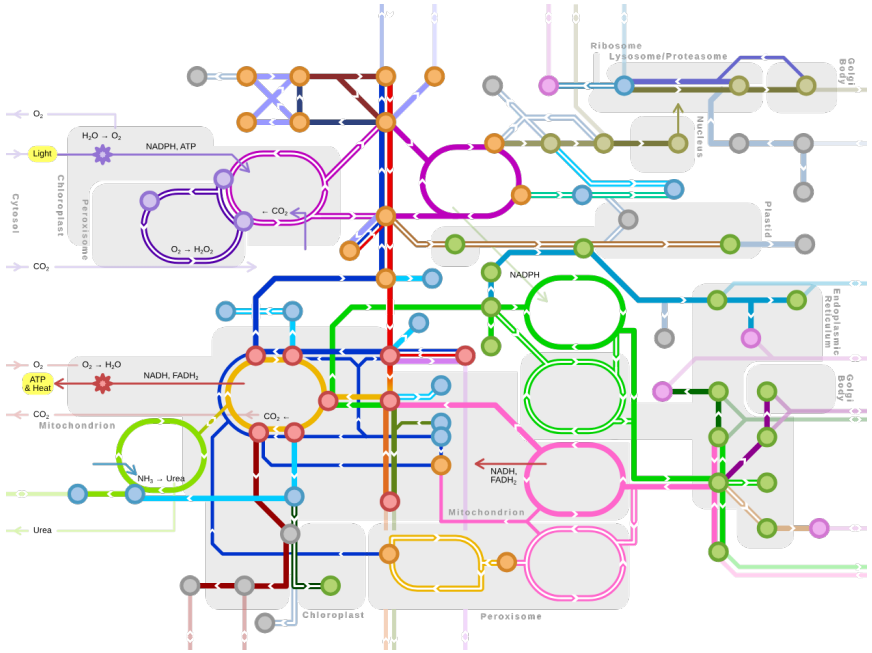
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Τύποι Βιολογικών Δικτύων

Μεταβολικά Δίκτυα

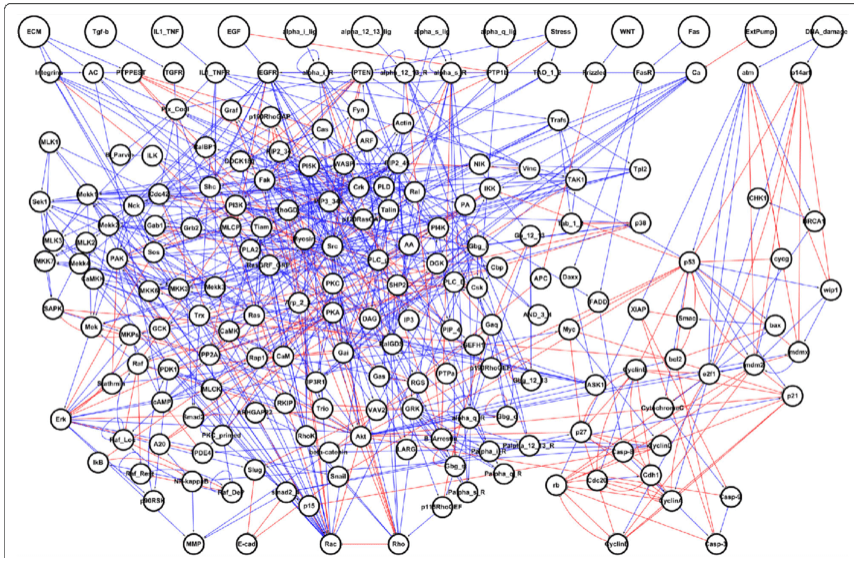
Αυτός ο τύπος δικτύου επιτρέπει τον σχολιασμό των γονιδίων και των μεταβολικών, προσδιορίζοντας στοιχεία των σχέσεων, της δομής και της δυναμικής των μεταβολικών δικτύων. Αυτό το δίκτυο συμπεραίνει την ενζυμική λειτουργία μιας συγκεκριμένης πρωτεΐνης ή αναδομεί τις μεταβολικές οδούς στις οποίες συμμετέχει.



Τύποι Βιολογικών Δικτύων

Δίκτυο σηματοδότησης

Επιτρέπει την αναπαράσταση αφαιρέσεων μοριακών αλληλεπιδράσεων και χημικών τροποποιήσεων που δρουν σε μια αλυσίδα για να μεταφέρουν ερεθίσματα (ορμόνες, παθογόνα, θρεπτικά συστατικά) που ανιχνεύονται από τους υποδοχείς της κυτταρικής μεμβράνης στον πυρήνα του κυττάρου, για να συντονίσουν την έναρξη των κατάλληλων μεταβολικών και γενετικών αποκρίσεων. Για την ανακατασκευή αυτού του τύπου δικτύου, έχουν χρησιμοποιηθεί τεχνικές όπως το γενετικό νοκ άουτ, οι οποίες επιτρέπουν τη μελέτη των διαφορετικών αποκρίσεων των οργανισμών σε αυτό το ερέθισμα.

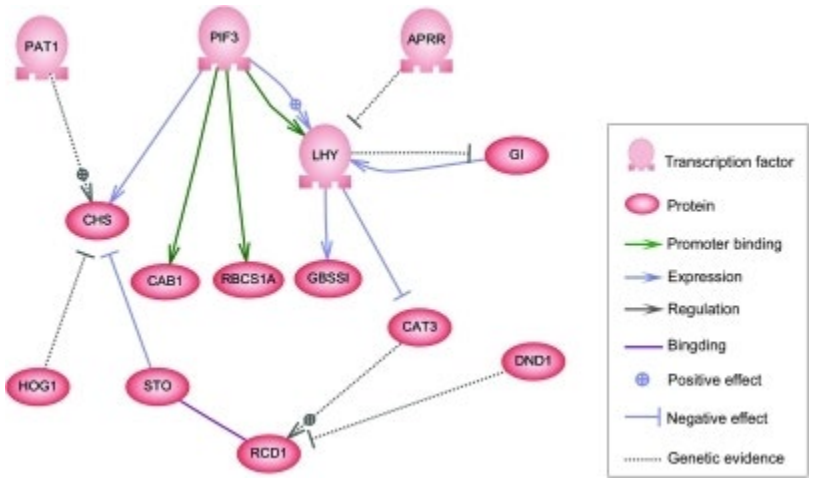


Τύποι Βιολογικών Δικτύων

Δίκτυο γονιδιακής ρύθμισης

Γνωστό και ως δίκτυο ρύθμισης της μεταγραφής, τα δίκτυα αντιπροσωπεύουν τις περιστασιακές αλληλεπιδράσεις μεταξύ μεταγραφικών παραγόντων και γονιδίων. Συνήθως αναπαρίστανται ως ένα κατευθυνόμενο γράφημα, του οποίου η κατεύθυνση ορίζεται από την έκφραση των γονιδίων. Οι διασυνδέσεις μεταξύ των γονιδίων μπορεί να αντιπροσωπεύουν βιοχημικές διεργασίες όπως η αντίδραση, ο μετασχηματισμός, η αλληλεπίδραση, η ενεργοποίηση ή η αναστολή.

Τα δίκτυα γονιδιακής ρύθμισης παρουσιάζουν το σύνολο των αλληλεπιδράσεων γονιδίων ενεργοποίησης και αναστολής εντός των κυττάρων. Στη διαδικασία της μεταγραφής συμμετέχουν αρκετοί μεταγραφικοί παράγοντες και γονιδιακά προϊόντα, ρυθμίζοντας, άμεσα ή έμμεσα, άλλα βιομόρια εντός του γονιδιώματος, μέσω ρυθμιστικών αλυσίδων. Από την άλλη πλευρά, μέσα σε αυτή τη διαδικασία μπορούν επίσης να δημιουργηθούν βρόχοι ανατροφοδότησης, ρυθμίζοντας αρνητικά (μη ρυθμιζόμενη) ή θετικά (υπορυθμισμένη) την παραγωγή γονιδιακού προϊόντος.



Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμοι και εργαλεία χαρτογράφησης

Δίκτυο αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης

Τα δίκτυα αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης (PPIs) αποτελούνται από πρωτεΐνες και τις αλληλεπιδράσεις τους. Σε αυτό το δίκτυο, οι κόμβοι αντιπροσωπεύουν τις πρωτεΐνες και οι άκρες αντιστοιχούν στις αλληλεπιδράσεις μεταξύ των πρωτεϊνών. Οι πρωτεΐνες οργανώνονται σε διαφορετικά υποτιθέμενα σύμπλοκα, καθένα από τα οποία εκτελεί μια συγκεκριμένη εργασία ή διαδικασία εντός των κυττάρων. Μια αλληλεπίδραση πρωτεΐνης-πρωτεΐνης συμβαίνει όταν δύο ή περισσότερες πρωτεΐνες ενώνονται προσωρινά για να τροποποιήσουν η μία την άλλη, να πυροδοτήσουν μεταγωγή σήματος ή να εκτελέσουν συγκεκριμένες βιολογικές λειτουργίες για παρατεταμένο χρονικό διάστημα.

Η κατασκευή και η ανάλυση των δικτύων PPI επιτρέπουν τον εντοπισμό συμπλεγμάτων πρωτεϊνών, γεγονός που επιτρέπει τη μελέτη και την κατανόηση των μηχανισμών που ρυθμίζουν τη ζωή, εξηγώντας το σήμα της εξελικτικής ορθολογίας , την πρόβλεψη βιολογικών λειτουργιών αχαρακτήριστων πρωτεϊνών και ανιχνεύσεις στόχων φαρμάκων για συγκεκριμένες ασθένειες. Μία από τις πιο χρησιμοποιούμενες τεχνικές για την ανίχνευση συμπλεγμάτων πρωτεϊνών είναι οι τεχνικές ομαδοποίησης, οι οποίες επιτρέπουν την ταυτοποίηση ομάδων πρωτεϊνών που μοιράζονται ομοιότητες ή κοινές περιοχές.

Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμοι και εργαλεία χαρτογράφησης

Δίκτυο αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης

Τα δίκτυα PPI επιτρέπουν την αναπαράσταση των σχέσεων που εμφανίζονται μεταξύ των πρωτεϊνών μέσα στα κύτταρα, δηλαδή την αλληλεπίδραση των οργανισμών που βρίσκονται υπό μελέτη. Η χρήση τεχνικών ανίχνευσης αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης, όπως ο καθαρισμός συνάφειας υψηλής απόδοσης σε συνδυασμό με φασματομετρία μάζας και τον προσδιορισμό δύο υβριδίων ζυμομύκητα και τους αλγόριθμους πρόβλεψης PPI, κατέστησαν δυνατή την κατασκευή και τη μελέτη πιο πολύπλοκων και ολοκληρωμένων αλληλεπιδράσεων. Παρά τις προόδους, είναι σημαντικό να σημειωθεί ότι η τρέχουσα γνώση σχετικά με την αλληλεπίδραση είναι ελλιπής και θορυβώδης. Οι τεχνικές που χρησιμοποιούνται έχουν περιορισμούς ως προς τον αριθμό των γνήσιων φυσιολογικών αλληλεπιδράσεων και παρουσιάζουν ορισμένα ψευδώς θετικά και ψευδώς αρνητικά.

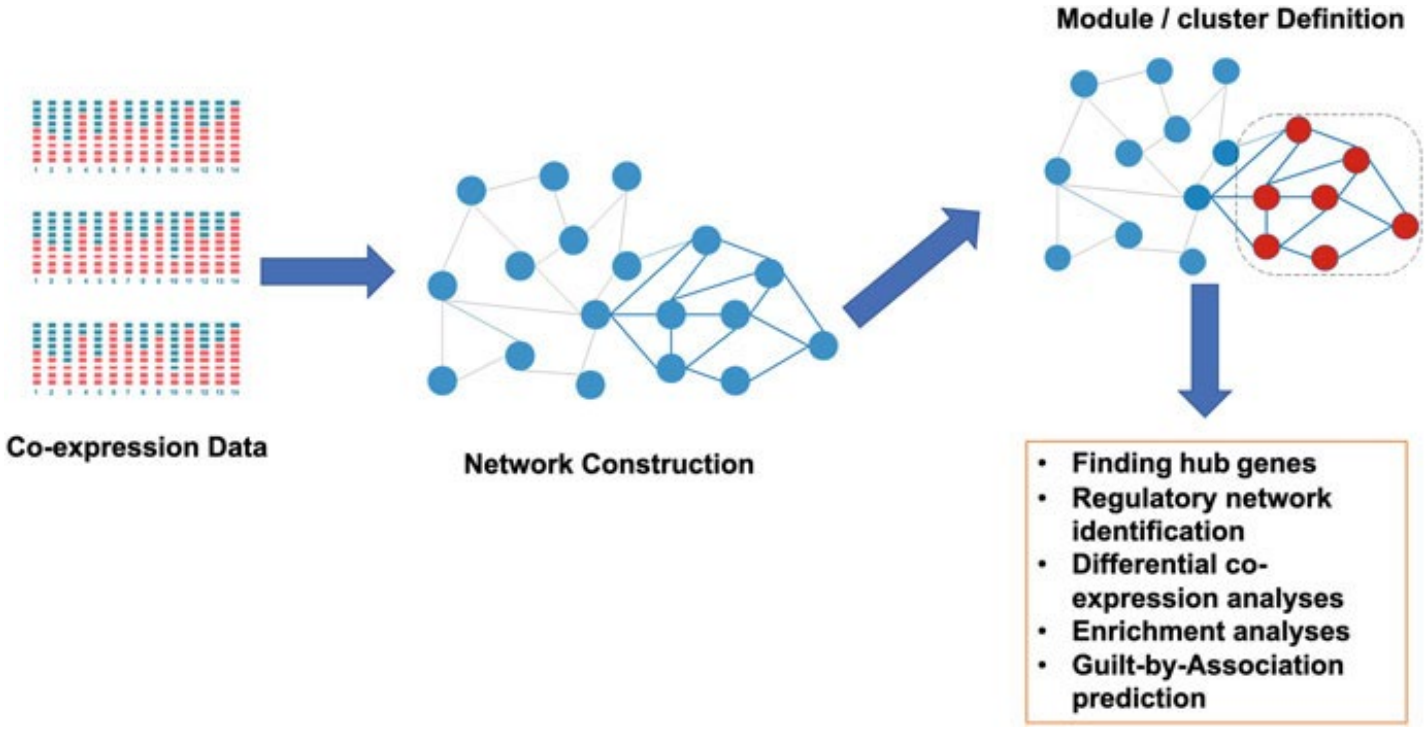
Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμοι και εργαλεία χαρτογράφησης

Δίκτυο Βιολογικής Συνέκφρασης

Αυτά τα δίκτυα επιτρέπουν την εξερεύνηση των υφιστάμενων συσχετισμών και αλληλεπιδράσεων μεταγραφής-μεταγραφής, όπου τα γονίδια είναι γενικά διασυνδεδεμένα αφού υπάρχει μια συσχέτιση συνέκφρασης μεταξύ τους. Αυτό το δίκτυο είναι κατασκευασμένο από δεδομένα γονιδιακής έκφρασης και αναπαρίσταται ως ένα μη κατευθυνόμενο δίκτυο.

Τα δίκτυα συνέκφρασης (CEN) μπορούν να χρησιμοποιηθούν για διάφορους σκοπούς, όπως ο εντοπισμός γονιδίων με τις πιο σημαντικές επιρροές μέσα στα δίκτυα, η ιεράρχηση των υποψήφιων γονιδίων ασθενειών, ο λειτουργικός σχολιασμός των γονιδίων και ο εντοπισμός ρυθμιστικών γονιδίων μέσα στα δίκτυα.



Ροή εργασιών για τη δημιουργία και την ανάλυση ενός δικτύου συν-έκφρασης. Το σχήμα δείχνει τα διάφορα είδη αναλύσεων που μπορούν να πραγματοποιηθούν με αυτόν τον τύπο δικτύου.

Τα δίκτυα συνέκφρασης δημιουργούνται με ανάλυση της ομοιότητας των προφίλ γονιδιακής έκφρασης, χρησιμοποιώντας τεχνικές όπως ο συντελεστής συσχέτισης, οι μετρήσεις απόστασης και οι αναπτυγμένοι αλγόριθμοι που βασίζονται σε στατιστικές μετρήσεις. Οι διασυνδέσεις μεταξύ των γονιδίων προσδιορίζονται χρησιμοποιώντας μια αποκοπή που επιτρέπει τη δόμηση των διασυνδέσεων εντός των δικτύων. Η χρήση αυτών των τύπων ελέγχου μέσω αποκοπής επιτρέπει στο δίκτυο να αναπαραστήσει περίπλοκες διαδικασίες και μοτίβα μέσα στους οργανισμούς.

Τα CEN έχουν ιδιότητες όπως η μεταβατικότητα, επιτρέπουν την αναγνώριση πυκνών κοινοτήτων γονιδίων μέσα στα δίκτυα, τα οποία υποδεικνύουν ότι τα γονίδια μέλη αυτών των κοινοτήτων σχετίζονται λειτουργικά. μέσα στο σύμπλεγμα, υπάρχουν κόμβοι (γονίδια) που παρουσιάζουν υψηλό βαθμό διασύνδεσης, αυτοί ονομάζονται κεντρικά γονίδια (hubs), αυτός ο τύπος γονιδίων είναι γενικά πιο σημαντικός για τον καθορισμό της λειτουργικότητας του δικτύου, εκτός από το ότι μπορεί να εξηγήσει την καλύτερη λειτουργία κάθε ενότητας. Μια άλλη ιδιότητα των CEN είναι η ιδιοκτησία ελεύθερης κλίμακας. Αυτό το είδος δικτύου έχει έναν βαθμό κατανομής που ακολουθεί την κατανομή του νόμου ισχύος, στην οποία οι περισσότεροι κόμβοι έχουν χαμηλούς βαθμούς, σε αντίθεση με την ύπαρξη μερικών κόμβων με υψηλούς βαθμούς, που δείχνει την πιθανότητα ένας κόμβος να συνδεθεί με έναν άλλο ευθέως ανάλογη του βαθμού της. Λόγω αυτής της ιδιότητας, είναι δυνατό να εντοπιστεί ένα μικρό σύνολο κεντρικών γονιδίων (hubs) και ένα ευρύ σύνολο γονιδίων με λίγες αλληλεπιδράσεις. Αυτό το χαρακτηριστικό επιτρέπει στα δίκτυα να είναι πιο ισχυρά.

Co-funded by the Erasmus+ Programme of the European Union

Τα δίκτυα συνέκφρασης έχουν χρησιμοποιηθεί για την κατανόηση των σχέσεων μεταξύ της έκφρασης των γονιδίων και της μελέτης διαφορετικών φαινομένων και αλληλεπιδράσεων μεταξύ γονιδίων. Τα CEN χρησιμοποιήθηκαν για τον εντοπισμό γονιδίων που σχετίζονται με τη σύνθεση και τον μεταβολισμό της φαινβενδαζόλης και της μεγλουμίνης της φλουνιξίνης στο συκώτι χοίρων. αυτή η έρευνα βρήκε οκτώ γονιδιακές ενότητες που έδειξαν υψηλή σχέση με το επίπεδο των μεταγραφών που σχετίζονται με το μεταβολισμό αυτών των φαρμάκων.

Συναρπαστική έρευνα που χρησιμοποίησε δίκτυα συνέκφρασης διεξήχθη από τους Shaik και Ramakrishna. Αυτή η μελέτη παρουσιάζει τα κοινά γονίδια για τις αντιδράσεις στο νερό και το βακτηριακό στρες που υπάρχουν στο ρύζι και στο Arabidopsis. Η ομάδα κατάφερε να εντοπίσει αρκετές κοινές γονιδιακές ενότητες που έδειξαν υψηλή συνέκφραση και συγκεκριμένους κόμβους που σχετίζονται με αυτά τα στρες.

Τα CEN χρησιμοποιήθηκαν για τη μελέτη των γονιδίων που εμπλέκονται στην ανάπτυξη των σκελετών και της μυϊκής μάζας των ποντικών για τη μυοστατίνη. Στη μελέτη που αναπτύχθηκε από τον Yang, οι ερευνητές κατασκεύασαν δίκτυα συνέκφρασης χρησιμοποιώντας δεδομένα μικροσυστοιχιών, τα οποία τους επέτρεψαν να μελετήσουν βιολογικές διεργασίες και μεταβολικές οδούς που σχετίζονται με την ανάπτυξη μυών και σκελετών σε άγρια ποντίκια. Αυτή η μελέτη επέτρεψε να επιβεβαιωθούν και να εντοπιστούν νέοι μεταγραφικοί ρυθμιστές.

Co-funded by the Erasmus+ Programme of the European Union

Το CEN μπορεί να χρησιμοποιηθεί για την ανίχνευση βιοδεικτών. Η έρευνα που αναπτύχθηκε από την ομάδα των Zhao και Li (2019), μελέτησε τον σακχαρώδη διαβήτη κύησης και κατάφερε να εντοπίσει δέκα πιθανούς βιοδείκτες που βοηθούν στη διάγνωση και τη θεραπεία αυτής της νόσου μέσω δικτύων συνέκφρασης.

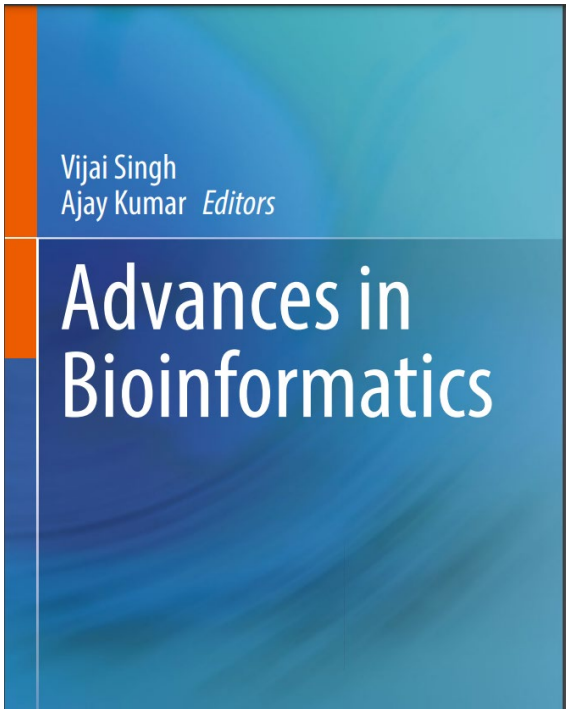
Στην έρευνα των Yuan et al. (2018), αναλύθηκαν βιοδείκτες για τη διάγνωση καρκινώματος του φλοιού των επινεφριδίων. Στο πλαίσιο αυτής της μελέτης, ανέλυσαν 12 κεντρικά γονίδια (hubs) εντός των δικτύων που έδειξαν συσχέτιση που σχετίζεται με την πρόγνωση και την πρόοδο της νόσου. Μια άλλη ομάδα χρησιμοποίησε δίκτυα συνέκφρασης για να ανιχνεύσει υποψήφια ρυθμιστικά γονίδια που παρουσιάζουν διαφορική έκφραση και που συμβάλλουν στην εξάπλωση της Salmonella enterica στους χοίρους.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Δώστε μερικά παραδείγματα βιολογικών αλληλεπιδράσεων που μπορούν να αναπαρασταθούν από δίκτυα.
- Περιγράψτε τους αλγόριθμους και τα εργαλεία χαρτογράφησης.

Co-funded by the Erasmus+ Programme of the European Union



Vijai Singh, Ajay Kumar (2021) Προόδους στη Βιοπληροφορική Springer Nature Singapore Pte Ltd

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

Θέμα 1. Εισαγωγή στη Βιοπληροφορική

Μάθημα 1. Ιστορία της Βιοπληροφορικής

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Οι καταβολές
- Η υποβοηθούμενη από υπολογιστή γενεαλογία της ζωής
- Μετατόπιση παραδείγματος από την ανάλυση πρωτεΐνης στην ανάλυση DNA
- Παράλληλες εξελίξεις στη βιολογία και την επιστήμη των υπολογιστών
- Γονιδιωματική, δομική βιοπληροφορική και η λεωφόρος πληροφοριών

Co-funded by the Erasmus+ Programme of the European Union

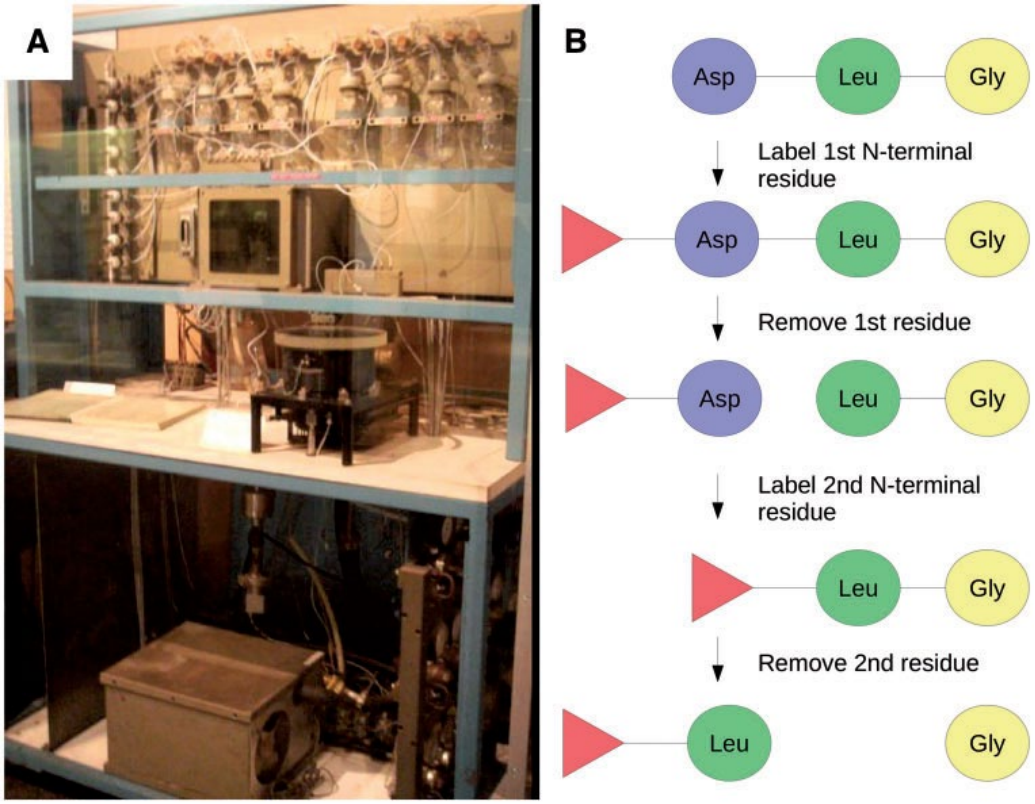
Εισαγωγή

Οι υπολογιστές και το εξειδικευμένο λογισμικό έχουν γίνει ουσιαστικό μέρος της εργαλειοθήκης του βιολόγου. Είτε για ανάλυση ρουτίνας DNA ή αλληλουχίας πρωτεϊνών είτε για ανάλυση σημαντικών πληροφοριών σε τεράστια σύνολα βιολογικών δεδομένων μεγέθους gigabyte, σχεδόν όλα τα σύγχρονα ερευνητικά έργα στη βιολογία απαιτούν, σε κάποιο βαθμό, τη χρήση υπολογιστών. Αυτό ισχύει ιδιαίτερα από την εμφάνιση της αλληλουχίας επόμενης γενιάς (NGS) που άλλαξε θεμελιωδώς τους τρόπους της πληθυσμιακής γενετικής, της ποσοτικής γενετικής, της μοριακής συστηματικής, της μικροβιακής οικολογίας και πολλών άλλων ερευνητικών πεδίων.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

1950–1970: Οι καταβολές



Αυτοματοποιημένη αλληλουχία πεπτιδίων Edman. (Α) Ένας από τους πρώτους αυτοματοποιημένους προσδιοριστές αλληλουχίας πεπτιδίων, σχεδιασμένος από τον William J. Dreyer. (Β) Αλληλουχία Edman: το πρώτο Ν τερματικό αμινοξύ μιας πεπτιδικής αλυσίδας επισημαίνεται με φαινυλισοθειοκυανικό (PITC, κόκκινο τρίγωνο) και στη συνέχεια διασπάται με μείωση του pH. Με την επανάληψη αυτής της διαδικασίας, μπορεί κανείς να προσδιορίσει μια αλληλουχία πεπτιδίου, ένα Ν-τερματικό αμινοξύ τη φορά.

Co-funded by the Erasmus+ Programme of the European Union

Η ανάλυση πρωτεϊνών ήταν το σημείο εκκίνησης

Στα τέλη της δεκαετίας του 1950, εκτός από τις σημαντικές προόδους στον προσδιορισμό των πρωτεϊνικών δομών μέσω της κρυσταλλογραφίας [9], δημοσιεύτηκε η πρώτη αλληλουχία (δηλ. διάταξη αλυσίδας αμινοξέων) μιας πρωτεΐνης, της ινσουλίνης. Αυτό το μεγάλο άλμα διευκόλυνε τη συζήτηση σχετικά με τη διάταξη της πολυπεπτιδικής αλυσίδας των πρωτεϊνών. Επιπλέον, ενθάρρυνε την ανάπτυξη πιο αποτελεσματικών μεθόδων για τη λήψη αλληλουχιών πρωτεϊνών. Η μέθοδος αποικοδόμησης Edman προέκυψε ως μια απλή μέθοδος που επέτρεπε τον προσδιορισμό της αλληλουχίας των πρωτεϊνών, ένα αμινοξύ τη φορά ξεκινώντας από το N-άκρο. Σε συνδυασμό με την αυτοματοποίηση, περισσότερες από 15 διαφορετικές οικογένειες πρωτεϊνών αναλύθηκαν σε αλληλουχία τα επόμενα 10 χρόνια.

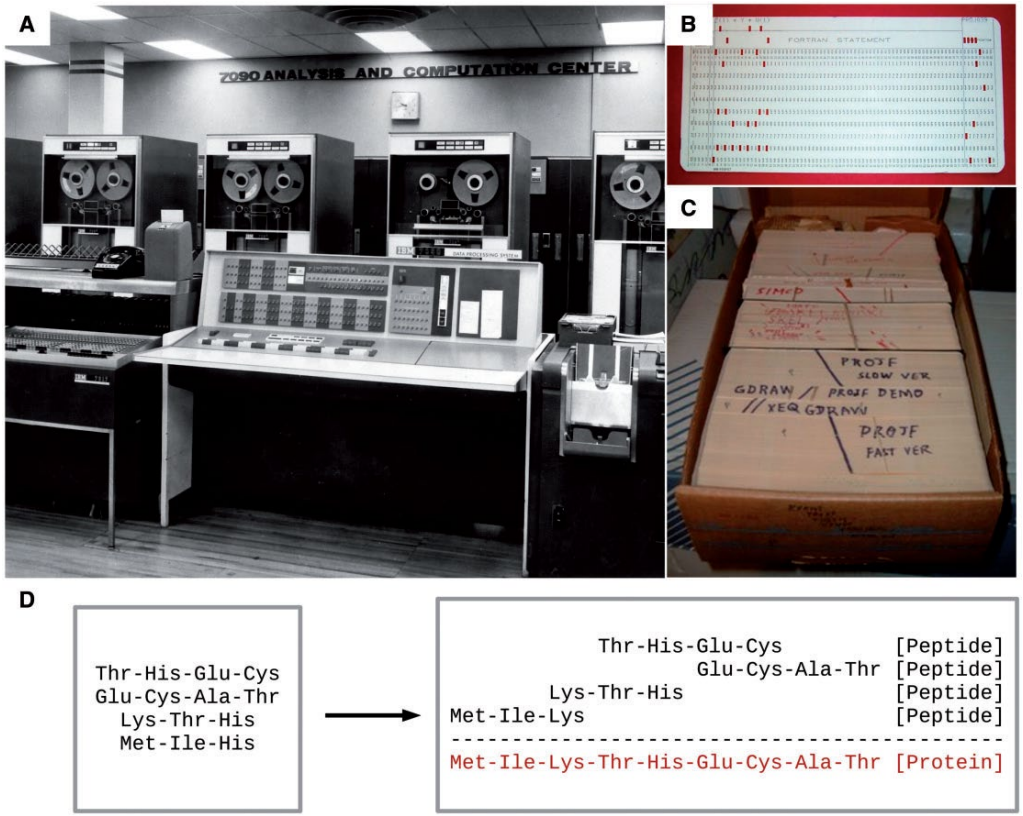
Ένα σημαντικό πρόβλημα με τον προσδιορισμό της αλληλουχίας Edman ήταν η απόκτηση μεγάλων αλληλουχιών πρωτεΐνης και όχι η αλληλούχιση μιας πρωτεΐνης από μόνη της, αλλά μάλλον η συναρμολόγηση ολόκληρης της αλληλουχίας πρωτεΐνης από εκατοντάδες μικρές αλληλουχίες πεπτιδίου Edman.

Dayhoff : ο πρώτος βιοπληροφορικός

Η Margaret Dayhoff (1925–1983) ήταν Αμερικανίδα φυσικοχημικός που πρωτοστάτησε στην εφαρμογή υπολογιστικών μεθόδων στον τομέα της βιοχημείας. Η συμβολή της Dayhoff σε αυτόν τον τομέα είναι τόσο σημαντική που ο David J. Lipman, πρώην διευθυντής του Εθνικού Κέντρου Βιοτεχνολογικών Πληροφοριών (NCBI), την αποκάλεσε «η μητέρα και ο πατέρας της βιοπληροφορικής».

άρχισε να συνεργάζεται με τον Robert S. Ledley , έναν φυσικό που επίσης προσπάθησε να φέρει υπολογιστικούς πόρους σε βιοϊατρικά προβλήματα. Από το 1958 έως το 1962, και οι δύο συνδύασαν την τεχνογνωσία τους και ανέπτυξαν το COMPROTEIN, «ένα πλήρες πρόγραμμα υπολογιστή για το IBM 7090» σχεδιασμένο για τον προσδιορισμό της πρωτογενούς δομής πρωτεΐνης χρησιμοποιώντας δεδομένα αλληλουχίας πεπτιδίων Edman. Αυτό το λογισμικό, εξ ολοκλήρου κωδικοποιημένο σε FORTRAN σε κάρτες διάτρησης, είναι η πρώτη εμφάνιση αυτού που θα ονομάζαμε σήμερα de novo assembler ακολουθίας.

Dayhoff : ο πρώτος βιοπληροφορικός



COMPROTEIN, το πρώτο λογισμικό βιοπληροφορικής. (Α) Ένας κεντρικός υπολογιστής IBM 7090, για τον οποίο η COMPROTEIN κατασκευάστηκε για να λειτουργεί. (Β) Μια κάρτα διάτρησης που περιέχει μία γραμμή κώδικα FORTRAN (η γλώσσα με την οποία γράφτηκε το COMPROTEIN). (Γ) Ο πηγαίος κώδικας ενός ολόκληρου προγράμματος σε κάρτες διάτρησης. (Δ) Μια απλοποιημένη επισκόπηση της εισόδου της COMPROTEIN (δηλ. πεπτιδικές αλληλουχίες Edman) και της εξόδου (μια συναινετική αλληλουχία πρωτεΐνης).

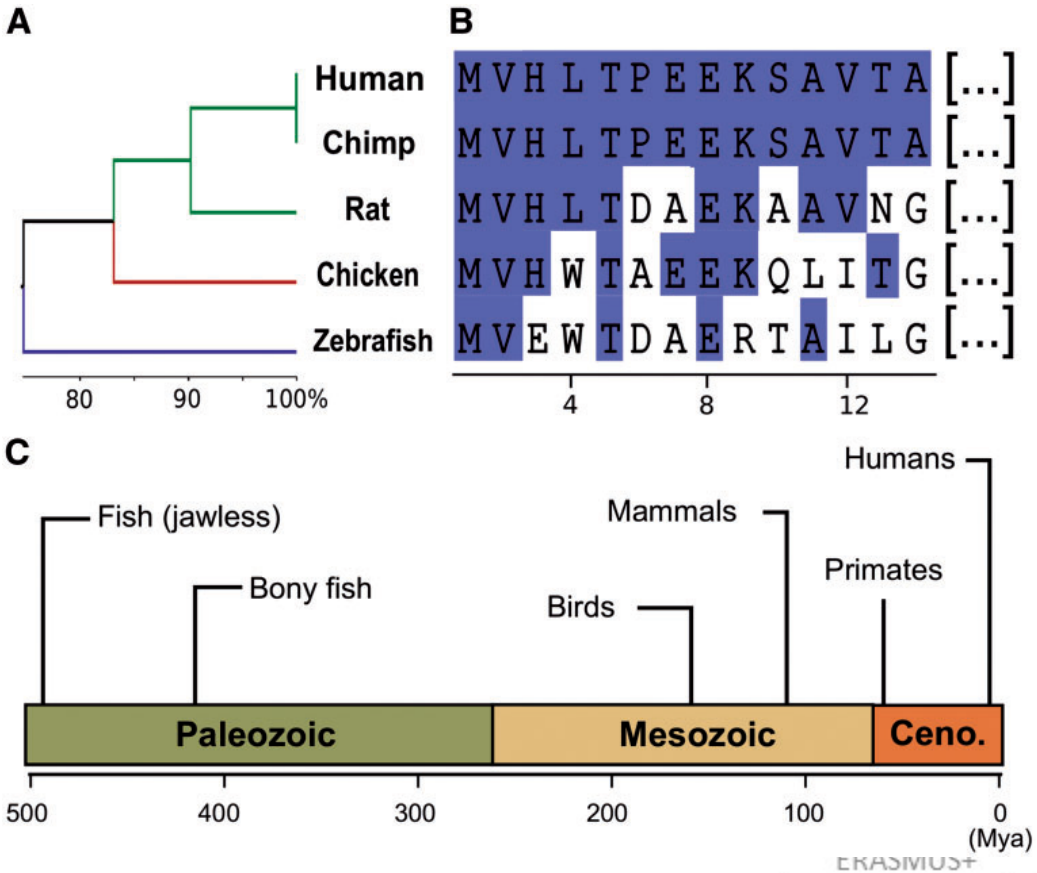
Η υποβοηθούμενη από υπολογιστή γενεαλογία της ζωής

Ο Emile Zuckerkandl και ο Linus Pauling αποχώρησαν από αυτό το παράδειγμα ερευνώντας τις βιομοριακές αλληλουχίες ως «φορείς πληροφοριών». Ακριβώς όπως οι λέξεις είναι σειρές γραμμάτων των οποίων η συγκεκριμένη διάταξη μεταδίδει νόημα, η μοριακή λειτουργία (δηλαδή το νόημα) μιας πρωτεΐνης προκύπτει από το πώς τα αμινοξέα της είναι διατεταγμένα για να σχηματίσουν μια «λέξη». Και οι δύο παρατήρησαν ότι οι ορθολογικές πρωτεΐνες από οργανισμούς σπονδυλωτών, όπως η αιμοσφαιρίνη, έδειξαν έναν βαθμό ομοιότητας πολύ υψηλό σε μεγάλο χρόνο εξέλιξης για να είναι αποτέλεσμα είτε τυχαίας είτε συγκλίνουσας εξέλιξης (Ibid). Η ίδια η έννοια της ορθολογίας ορίστηκε το 1970 από τον Walter M. Fitch για να περιγράψει την ομολογία που προέκυψε από ένα συμβάν ειδογένεσης

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Η υποβοηθούμενη από υπολογιστή γενεαλογία της ζωής



Η ανομοιότητα αλληλουχίας μεταξύ ορθολογικών πρωτεϊνών από οργανισμούς-μοντέλους συσχετίζεται με την εξελικτική τους ιστορία, όπως αποδεικνύεται από το αρχείο απολιθωμάτων. (Α) Μέση απόσταση δέντρου υπομονάδας αιμοσφαιρίνης βήτα-1 (HBB-1) από άνθρωπο (Homo sapiens), χιμπατζή (Pan troglodytes), αρουραίο (Rattus norvegicus), κοτόπουλο (Gallus gallus) και ζέβρα (Danio rerio) . (Β) Όψη ευθυγράμμισης των πρώτων 14 υπολειμμάτων αμινοξέων του HBB-1 σε σύγκριση με το (Α) (τα υπολείμματα που επισημαίνονται με μπλε είναι πανομοιότυπα με την ανθρώπινη αλληλουχία HBB-1). (Γ) Χρονοδιάγραμμα των πρώιμων απολιθωμάτων που βρέθηκαν για διαφορετικά υδρόβια και χερσαία ζώα

Co-funded by the Erasmus+ Programme of the European Union

Ένα μαθηματικό πλαίσιο για υποκαταστάσεις αμινοξέων

Το 1978, οι Dayhoff , Schwartz και Orcutt συνέβαλαν σε ένα άλλο ορόσημο στη βιοπληροφορική αναπτύσσοντας το πρώτο πιθανολογικό μοντέλο υποκαταστάσεων αμινοξέων. Αυτό το μοντέλο, που ολοκληρώθηκε 8 χρόνια μετά την έναρξή του, βασίστηκε στην παρατήρηση 1572 αποδεκτών μεταλλάξεων σημείων (PAMs) στα φυλογενετικά δέντρα 71 οικογενειών πρωτεϊνών που μοιράζονται πάνω από 85% ταυτότητα. Το αποτέλεσμα ήταν μια μήτρα ασύμμετρης υποκατάστασης 20 x 20 που περιείχε τιμές πιθανότητας με βάση τις παρατηρούμενες μεταλλάξεις κάθε αμινοξέος (δηλαδή την πιθανότητα ότι κάθε αμινοξύ θα αλλάξει σε ένα δεδομένο μικρό εξελικτικό διάστημα). Ενώ η αρχή του (δηλ. ο ελάχιστος αριθμός αλλαγών) χρησιμοποιήθηκε προηγουμένως για την ποσοτικοποίηση της εξελικτικής απόστασης στις φυλογενετικές ανακατασκευές, η μήτρα PAM εισήγαγε την αντικατάσταση ως τη μέτρηση της εξελικτικής αλλαγής.

Co-funded by the Erasmus+ Programme of the European Union

Ένα μαθηματικό πλαίσιο για υποκαταστάσεις αμινοξέων

10^4 P[a]		Ala A	Arg R	Asn N	Asp D	Cys C	Gln Q	...	Val V
Ala	A	9867	2	9	10	3	8	...	18
Arg	R	1	9913	1	0	1	10	...	1
Asn	N	4	1	9822	36	0	4	...	1
Asp	D	6	0	42	9859	0	6	...	1
Cys	C	1	1	0	0	9973	0	...	2
Gln	Q	3	9	4	5	0	9876	...	1
...	...	...	...	...	...	...	...	...	...
Val	V	13	2	1	1	3	2	...	9901

[a]Each numeric value represents the probability that an amino acid from the *i*-th column be substituted by an amino acid in the *j*-th row (multiplied by 10 000).

1970–1980: Μετατόπιση παραδείγματος από την ανάλυση πρωτεΐνης στην ανάλυση DNA

Αποκρυπτογράφηση της γλώσσας του DNA: ο γενετικός κώδικας

Οι προδιαγραφές για κάθε ζωντανό ον (ακριβέστερα, οι «πρωτεΐνες» του) κωδικοποιούνται στις συγκεκριμένες νουκλεοτιδικές διατάξεις του μορίου του DNA. Αυτή η άποψη επισημοποιήθηκε στην υπόθεση της αλληλουχίας του Francis Crick (που ονομάζεται επίσης σήμερα «Κεντρικό Δόγμα»), στην οποία υπέθεσε ότι οι αλληλουχίες RNA, μεταγραμμένες από το DNA, καθορίζουν την αλληλουχία αμινοξέων των πρωτεϊνών που κωδικοποιούν. Με τη σειρά της, η αλληλουχία αμινοξέων καθορίζει την τρισδιάστατη δομή της πρωτεΐνης.

Επομένως, αν μπορούσε κανείς να καταλάβει πώς το κύτταρο μεταφράζει τη «γλώσσα DNA» σε πολυπεπτιδικές αλληλουχίες, θα μπορούσε να προβλέψει την πρωτογενή δομή οποιασδήποτε πρωτεΐνης που παράγεται από έναν οργανισμό «διαβάζοντας το DNA του». Μέχρι το 1968, αποκρυπτογραφήθηκαν και τα 64 κωδικόνια του γενετικού κώδικα [35]. Το DNA ήταν πλέον «αναγνώσιμο» και αυτό το πρωτοποριακό επίτευγμα απαιτούσε απλούς και οικονομικούς τρόπους απόκτησης αλληλουχιών DNA.

Οικονομική ανάγνωση του DNA

Η δυνατότητα λήψης αλληλουχιών DNA από έναν οργανισμό έχει πολλά πλεονεκτήματα όσον αφορά την παροχή πληροφοριών. Ενώ οι πρωτεΐνες πρέπει να καθαρίζονται μεμονωμένα για να προσδιοριστούν η αλληλουχία τους, ολόκληρο το γονιδίωμα ενός οργανισμού μπορεί θεωρητικά να προέρχεται από ένα μόνο εκχύλισμα γονιδιωματικού DNA. Από αυτήν την αλληλουχία DNA ολόκληρου του γονιδιώματος, μπορεί κανείς να προβλέψει την πρωτογενή δομή όλων των πρωτεϊνών που εκφράζονται από έναν οργανισμό μέσω μετάφρασης των γονιδίων που υπάρχουν στην αλληλουχία. Αν και η αρχή μπορεί να φαίνεται απλή, η χειροκίνητη εξαγωγή πληροφοριών από αλληλουχίες DNA περιλαμβάνει τα εξής:

1. συγκρίσεις (π.χ. εύρεση ομολογίας μεταξύ αλληλουχιών από διαφορετικούς οργανισμούς).

2. υπολογισμοί (π.χ. κατασκευή ενός φυλογενετικού δέντρου πολλαπλών ορθολόγων πρωτεϊνών χρησιμοποιώντας τη μήτρα PAM1).

3. και αντιστοίχιση προτύπων (π.χ. εύρεση ανοιχτών πλαισίων ανάγνωσης σε μια ακολουθία DNA).

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Χρήση αλληλουχιών DNA στο φυλογενετικό συμπέρασμα

Ο Felsenstein ήταν ο πρώτος που ανέπτυξε μια μέθοδο μέγιστης πιθανότητας (ML) για να συμπεράνει τα φυλογενετικά δέντρα από αλληλουχίες DNA. Σε αντίθεση με τις μεθόδους παρρησίας, οι οποίες ανακατασκευάζουν ένα εξελικτικό δέντρο χρησιμοποιώντας τον ελάχιστο αριθμό αλλαγών, η εκτίμηση ML «περιλαμβάνει την εύρεση αυτού του εξελικτικού δέντρου που αποδίδει την υψηλότερη πιθανότητα εξέλιξης των παρατηρούμενων δεδομένων». Μετά την εργασία του Felsenstein στη μοριακή φυλογένεση, αναπτύχθηκαν αρκετά εργαλεία βιοπληροφορικής που χρησιμοποιούν ML και εξακολουθούν να αναπτύσσονται ευρέως, όπως και οι νέες στατιστικές μέθοδοι για την αξιολόγηση της ευρωστίας των κόμβων. Αυτό ενέπνευσε ακόμη και τη δεκαετία του 1990, τη χρήση των Bayesian στατιστικών στη μοριακή φυλογένεση, οι οποίες εξακολουθούν να χρησιμοποιούνται ευρέως στη βιολογία.

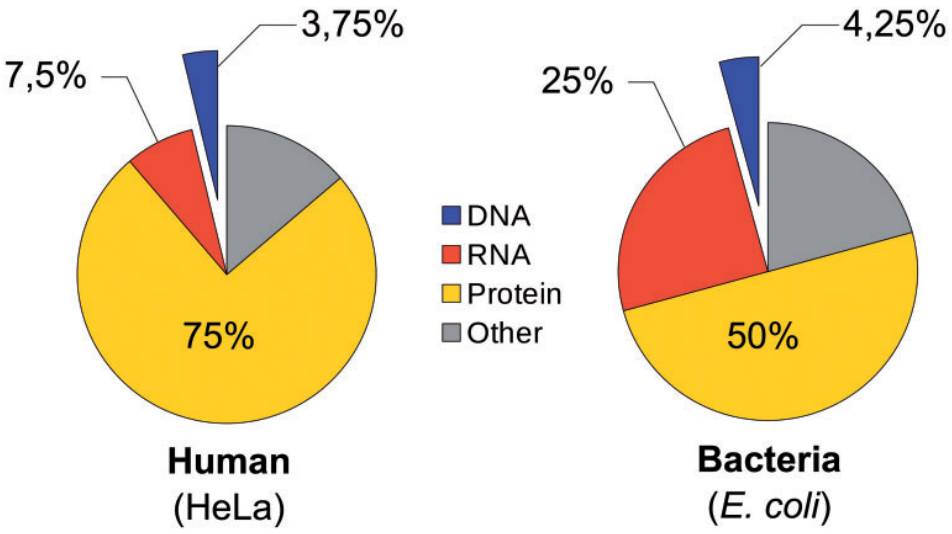
1980–1990: Παράλληλες εξελίξεις στη βιολογία και την επιστήμη των υπολογιστών

Μοριακές μέθοδοι στόχευσης και ενίσχυσης συγκεκριμένων γονιδίων

Τα γονίδια, σε αντίθεση με τις πρωτεΐνες και τα RNA, δεν μπορούν να κλασματωθούν βιοχημικά και στη συνέχεια να αναλυθούν μεμονωμένα την αλληλουχία τους, επειδή όλα βρίσκονται συνεχόμενα σε μια χούφτα μόρια DNA ανά κύτταρο. Επιπλέον, τα γονίδια υπάρχουν συνήθως σε ένα ή λίγα αντίγραφα ανά κύτταρο. Τα γονίδια είναι επομένως τάξεις μεγέθους λιγότερο άφθονα από τα προϊόντα που κωδικοποιούν.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

1980–1990: Παράλληλες εξελίξεις στη βιολογία και την επιστήμη των υπολογιστών



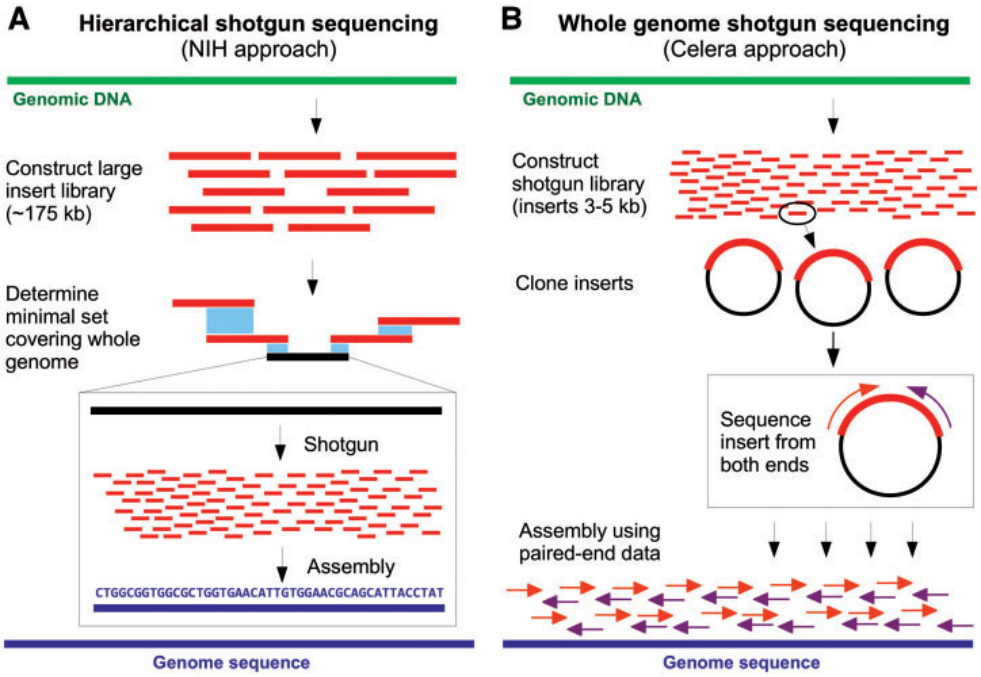
Το DNA είναι το λιγότερο άφθονο μακρομοριακό κυτταρικό συστατικό που μπορεί να προσδιοριστεί η αλληλουχία. Τα ποσοστά (%) αντιπροσωπεύουν την αφθονία κάθε συστατικού σε σχέση με το συνολικό ξηρό βάρος των κυττάρων. Δεδομένα από τη Βιβλιοθήκη Τεχνικών Πόρων

1990–2000: Γονιδιωματική, δομική βιοπληροφορική και η λεωφόρος πληροφοριών

	Fortran[a]	C	R	Java
First appeared	1957	1972	1993	1995
Typical use	Algorithmics, calculations, programming modules for other applications	Optimized command-line tools	Statistical analysis, data visualization	Graphical user interfaces, data visualization, network analysis
Notable fields of application	Biochemistry, Structural Bioinformatics	Various	Metagenomics, Transcriptomics, Systems Biology	Genomics, Proteomics, Systems Biology
Specialized bioinformatics repository?	None	None	Bioconductor, [73], since 2002	BioJava [74], since 2002
Example software or packages	Clustal [32, 33], WHAT IF [75]	MUSCLE [76], PhyloBayes [77]	edgeR [78], phyloseq [79]	Jalview [80], Jemboss [81], Cytoscape [82]

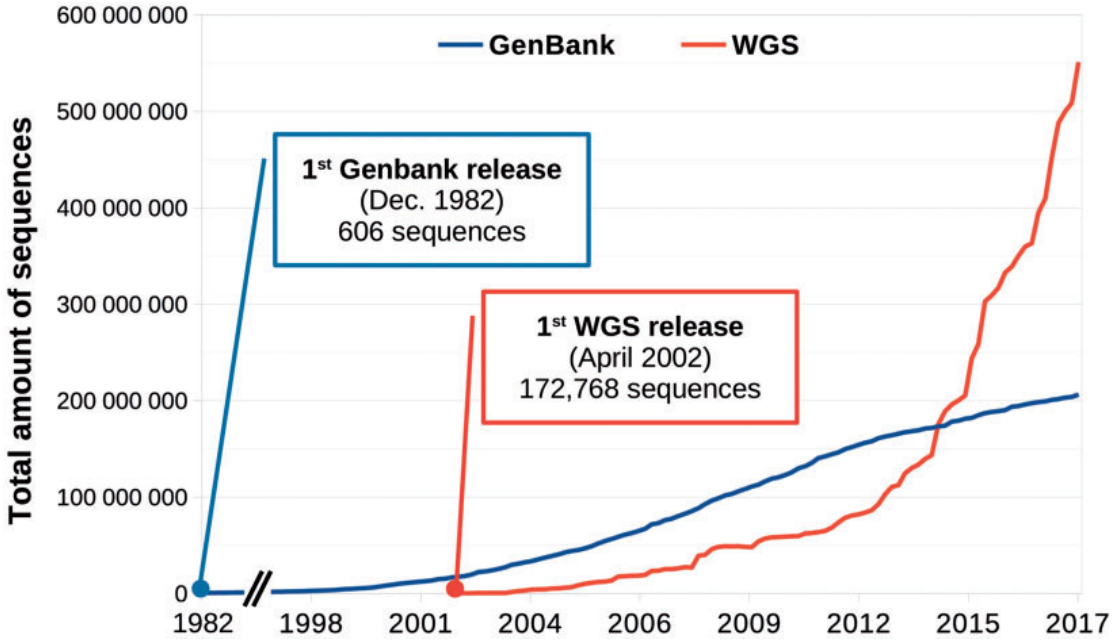
[a]Even though the earliest bioinformatics software were written in Fortran, it is seldom used to code standalone programs nowadays. It is rather used to code modules for other programs and programming languages (such as C and R mentioned here).

Αξιοσημείωτες γλώσσες προγραμματισμού



Ιεραρχική αλληλουχία κυνηγετικών όπλων έναντι αλληλουχίας κυνηγετικών όπλων ολόκληρου του γονιδιώματος. Και οι δύο προσεγγίσεις αντιστοίχως έδειξαν τη μεθοδολογική αντιπαλότητα μεταξύ των προσπαθειών του δημόσιου (NIH, A) και του ιδιωτικού (Celera, B) για την αλληλουχία του ανθρώπινου γονιδιώματος. Ενώ η ομάδα του NIH πίστευε ότι η αλληλουχία κυνηγετικών όπλων ολόκληρου του γονιδιώματος (WGS) ήταν τεχνικά αδύνατη για γονιδιώματα μεγέθους gigabase , η ομάδα Celera του Venter πίστευε ότι όχι μόνο αυτή η προσέγγιση ήταν εφικτή, αλλά ότι θα μπορούσε επίσης να ξεπεράσει το υλικοτεχνικό βάρος της ιεραρχικής αλληλουχίας κυνηγετικών όπλων, με την προϋπόθεση ότι είναι διαθέσιμοι αποδοτικοί αλγόριθμοι συναρμολόγησης και επαρκής υπολογιστική ισχύς. Λόγω της μερικής χρήσης των δεδομένων NIH σε συγκροτήματα Celera, η πραγματική σκοπιμότητα της αλληλουχίας WGS για το ανθρώπινο γονιδίωμα έχει συζητηθεί έντονα και από τις δύο πλευρές

2000–2010: Βιοπληροφορική υψηλής απόδοσης

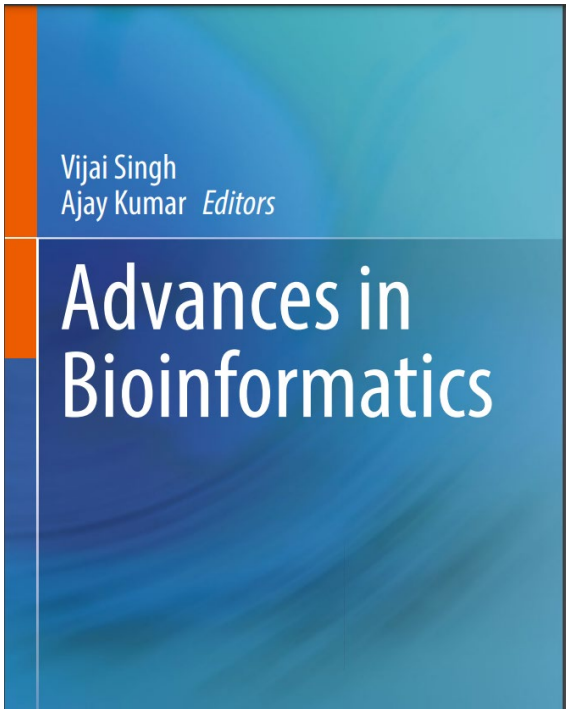


Συνολικός αριθμός ακολουθιών στις βάσεις δεδομένων NCBI GenBank και WGS (Whole Genome Shotgun) με την πάροδο του χρόνου. Ο αριθμός των πρόχειρων/ατελών γονιδιωμάτων έχει ξεπεράσει τον αριθμό των πλήρους γονιδιωματικών αλληλουχιών τον Ιούνιο του 2014 και οι φωτογραφίες συνεχίζονται με εκθετικό ρυθμό. Πηγή: https://www.ncbi.nlm.nih.gov/genbank/statistics/

Ερωτήσεις

- Επισημάνετε τις κύριες περιόδους ανάπτυξης της βιοπληροφορικής.
- Τι είναι χαρακτηριστικό για καθμία από αυτές;

Co-funded by the Erasmus+ Programme of the European Union



Vijai Singh, Ajay Kumar (2021) Προόδους στη Βιοπληροφορική Springer Nature Singapore Pte Ltd

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

Θέμα 1. Εισαγωγή στη Βιοπληροφορική

Μάθημα 2. Στόχοι Βιοπληροφορικής

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Στοιχεία βιολογίας
- Τι είναι η βιοπληροφορική;
- Επέκταση της έννοιας της βιοπληροφορικής
- Σχετικά με ορισμένα προβλήματα στη βιοπληροφορική

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

«Οι δύο τεχνολογίες που θα διαμορφώσουν τον επόμενο αιώνα είναι η βιοτεχνολογία και η τεχνολογία της πληροφορίας»

Μπιλ Γκέϊτς

«Οι δύο τεχνολογίες που θα έχουν τον μεγαλύτερο αντίκτυπο η μία στην άλλη στη νέα χιλιετία είναι η βιοτεχνολογία και η τεχνολογία πληροφοριών»

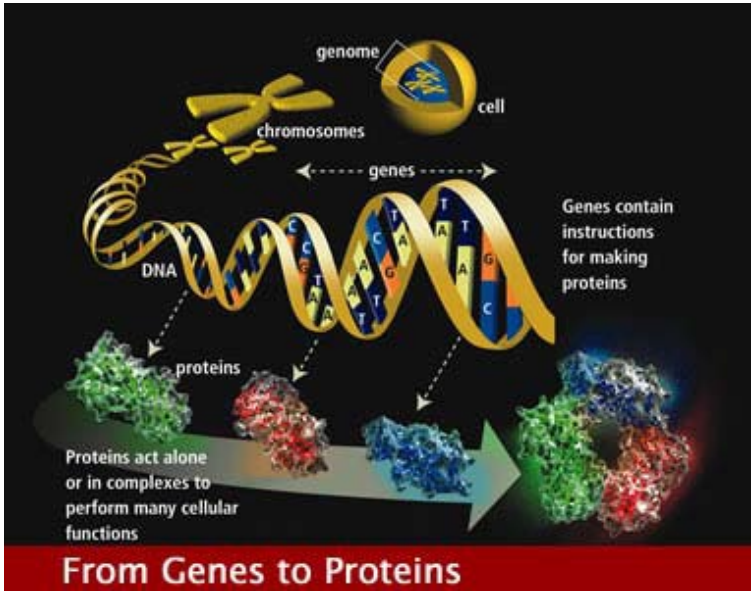
Martina McGloughlin

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Βασική μοριακή βιολογία

Τα περισσότερα από τα 100 δισεκατομμύρια κύτταρα στο ανθρώπινο σώμα περιέχουν ένα αντίγραφο ολόκληρου του ανθρώπινου γονιδιώματος (όλες τις γενετικές πληροφορίες που είναι απαραίτητες για την οικοδόμηση ενός ανθρώπου).



Ο πυρήνας του κυττάρου περιέχει έξι πόδια DNA συσκευασμένο σε 23 ζεύγη χρωμοσωμάτων. Ο καθένας μας κληρονομεί ένα σύνολο 23 χρωμοσωμάτων από τη μητέρα μας και ένα άλλο από τον πατέρα μας. Το DNA περιέχει τον κώδικα για το σώμα (γονίδια) που διέπει όλες τις πτυχές της κυτταρικής ανάπτυξης και κληρονομικότητας.

Η πρωτεΐνη, που αποτελείται από αμινοξέα, είναι απαραίτητα συστατικά όλων των οργάνων και των χημικών δραστηριοτήτων.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Μόρια της ζωής

1. Μικρά μόρια
2. Πρωτεΐνες
3. DNA
4. RNA

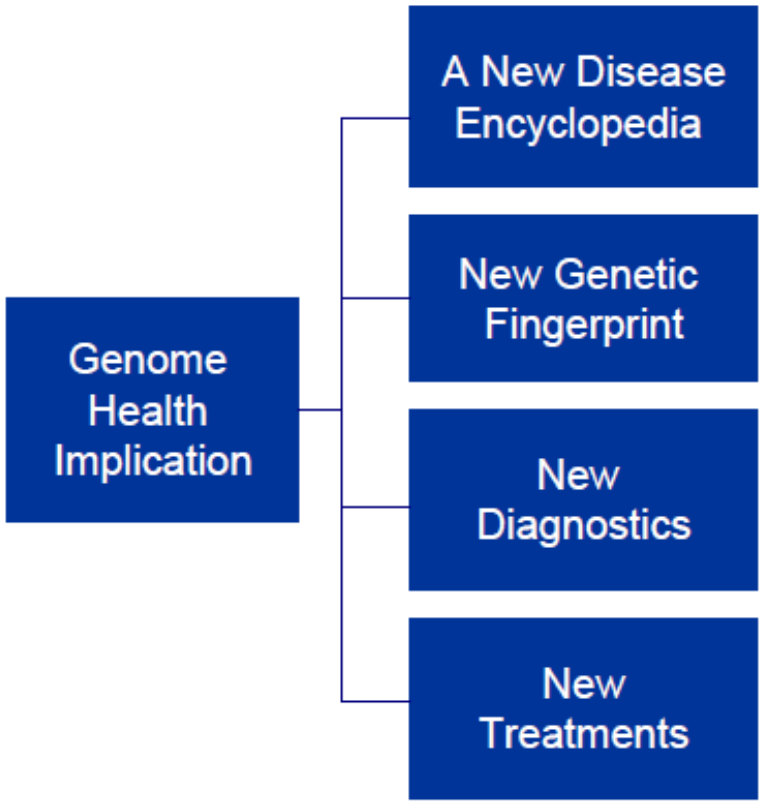
Βιολογικά μακρομόρια

Πρόβλημα πρόβλεψης γονιδίων

Πρόβλεψη γονιδίων: Είναι μια ενδιαφέρουσα ερώτηση: Δεδομένης της αλληλουχίας του γονιδιωματικού DNA, μπορούμε να πούμε που είναι τα γονίδια;

Organism	The number of predicted genes	Part of the genome that encodes proteins (exons)
E.Coli (bacteria)	5000	90%
Yeast	6000	70%
Worm	18,000	27%
Fly	14,000	20%
Weed	25,500	20%
Human	30,000	< 5%

Έργο ανθρώπινου γονιδιώματος



Στόχος

- προσδιορίσει όλα τα περίπου 30.000 γονίδια στο ανθρώπινο DNA,
- προσδιορίστε τις ακολουθίες από τα 3 δισεκατομμύρια ζεύγη χημικών βάσεων που συνθέτουν το ανθρώπινο DNA,
- αποθηκεύστε αυτές τις πληροφορίες σε βάσεις δεδομένων,
- βελτίωση των εργαλείων για την ανάλυση δεδομένων,
- μεταφορά σχετικών τεχνολογιών σε τον ιδιωτικό τομέα και
- αντιμετώπιση των ηθικών, νομικών και κοινωνικά θέματα (ELSI) που μπορεί προκύπτουν από το έργο.

Co-funded by the Erasmus+ Programme of the European Union

Τι είναι η βιοπληροφορική;

Βιο: Μοριακή Βιολογία
Πληροφορική: Επιστήμη των υπολογιστών

Βιοπληροφορική: Επίλυση προβλημάτων που προκύπτουν από τη βιολογία χρησιμοποιώντας μεθοδολογία από επιστήμη των υπολογιστών.

Συνώνυμα: Υπολογιστική βιολογία,
Υπολογιστική μοριακή βιολογία,
Βιουπολογιστική

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

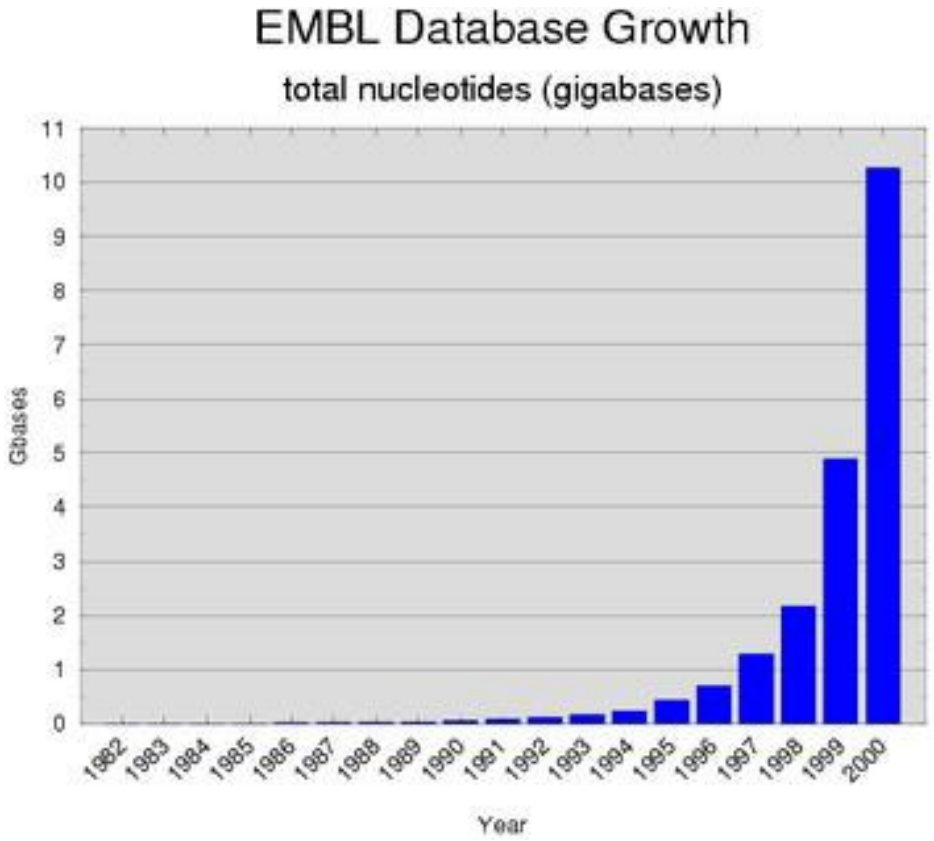
Co-funded by the Erasmus+ Programme of the European Union

Αλλαγή παραδείγματος στη Βιολογία

The new paradigm, now emerging, is that all the 'genes' will be known (in the sense of being resident in databases available electronically), and that the starting point of a biological investigation will be theoretical. An individual scientist will begin with a **theoretical conjecture**, only **then turning to experiment** to follow or test that hypothesis.

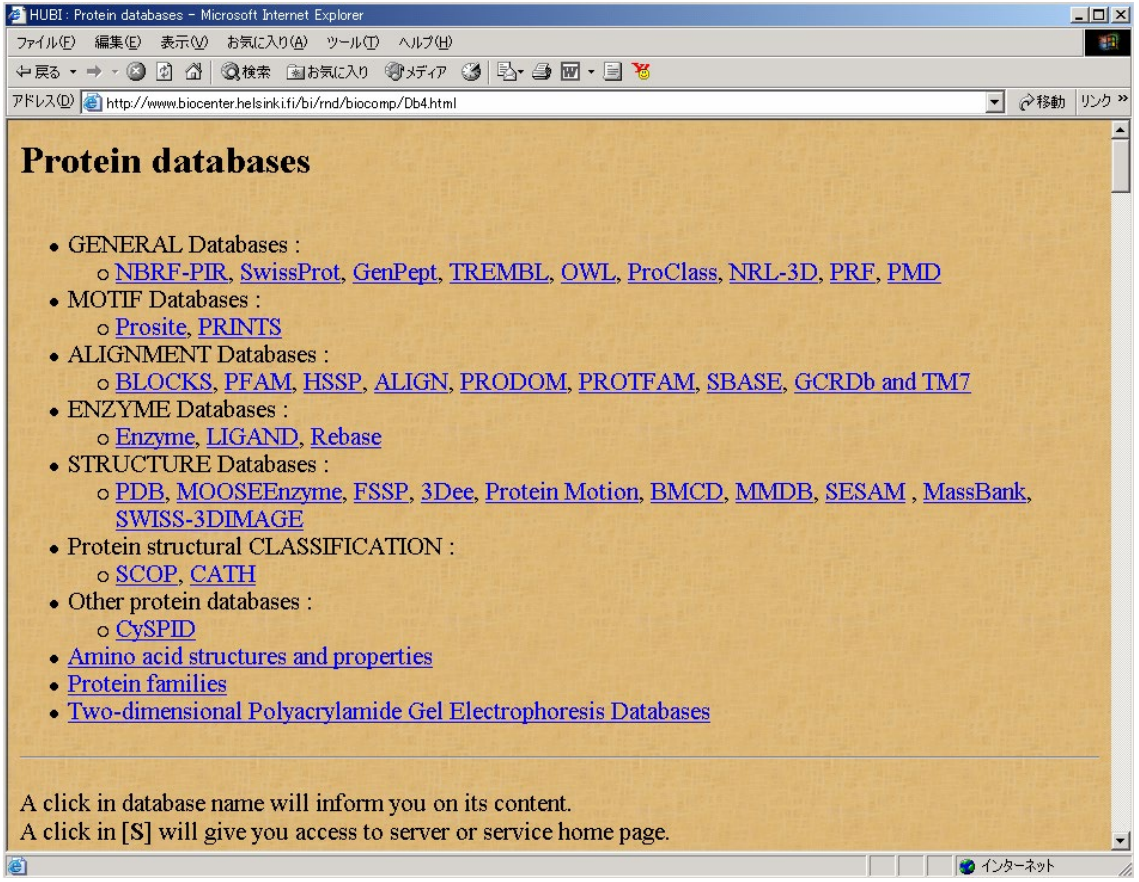
To use [the] flood of knowledge, which will pour across the computer networks of the world, biologists not only must become **computer literate**, but also **change their approach** to the problem of understanding life.

Ζεύγη βάσεων στη GenBank



10.267.507.282 βάσεις σε 9.092.760 εγγραφές.

Δημόσιες βάσεις δεδομένων



Επέκταση της έννοιας της βιοπληροφορικής

- **Genomics**
  - Functional genomics
  - Structural genomics
- **Proteomics:** large scale analysis of the proteins of an organism
- **Pharmacogenomics:** developing new drugs that will target a particular disease
- **Microarray (genome chip):** DNA chip, protein chip

The identification and functional characterization of genes.

The study of gene expression at the protein level, by the identification and characterization of proteins present in a biological sample.

The use of genetic information to predict the safety, toxicity and/or efficacy of drugs in individual patients or groups of patients.

a new technology aims to monitor the whole genome on a single chip so that researchers can have a better picture of the interactions among thousands of genes simultaneously

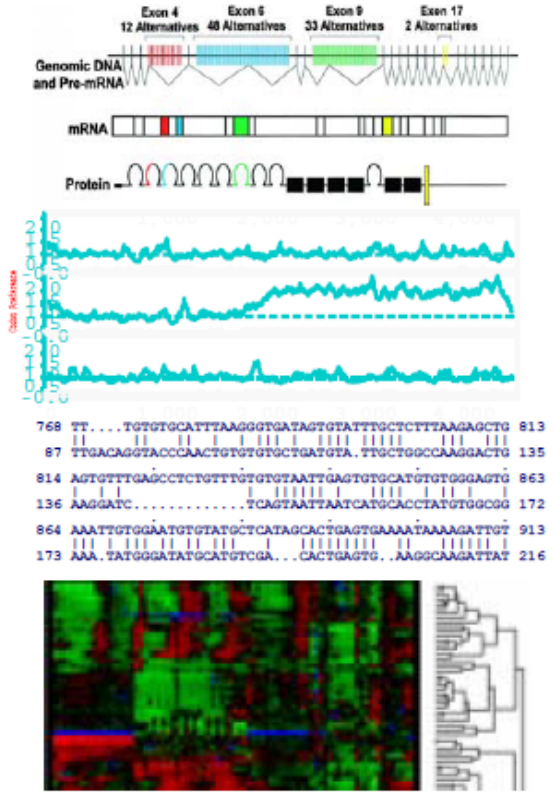
Προβλήματα στη Βιοπληροφορική

Structure analysis
- Protein structure comparison
- Protein structure prediction
- RNA structure modeling

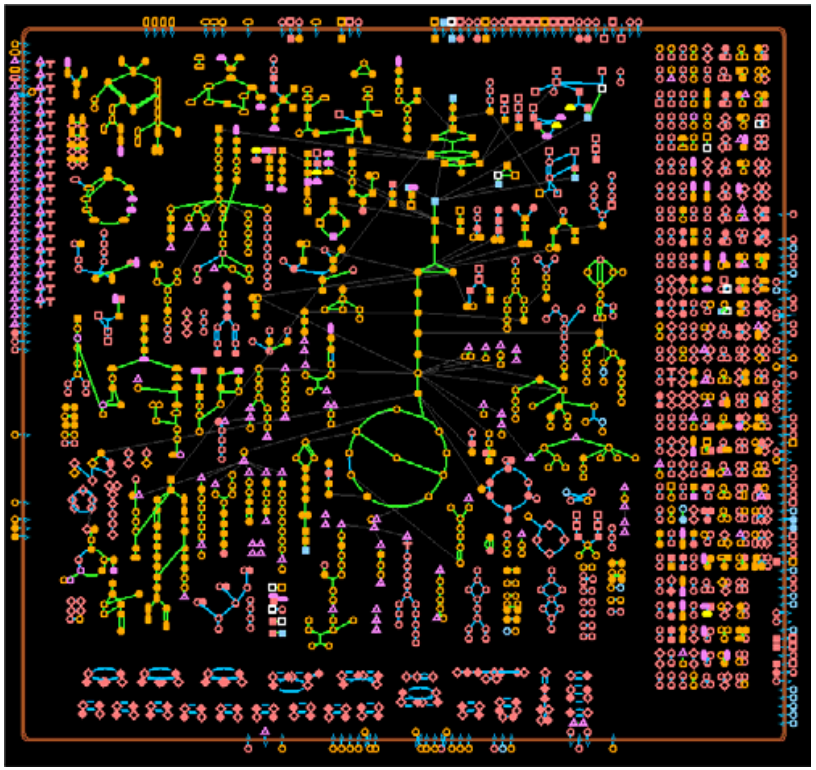
Pathway analysis
- Metabolic pathway
- Regulatory networks

Sequence analysis
- Sequence alignment
- Structure and function prediction
- Gene finding

Expression analysis
- Gene expression analysis
- Gene clustering

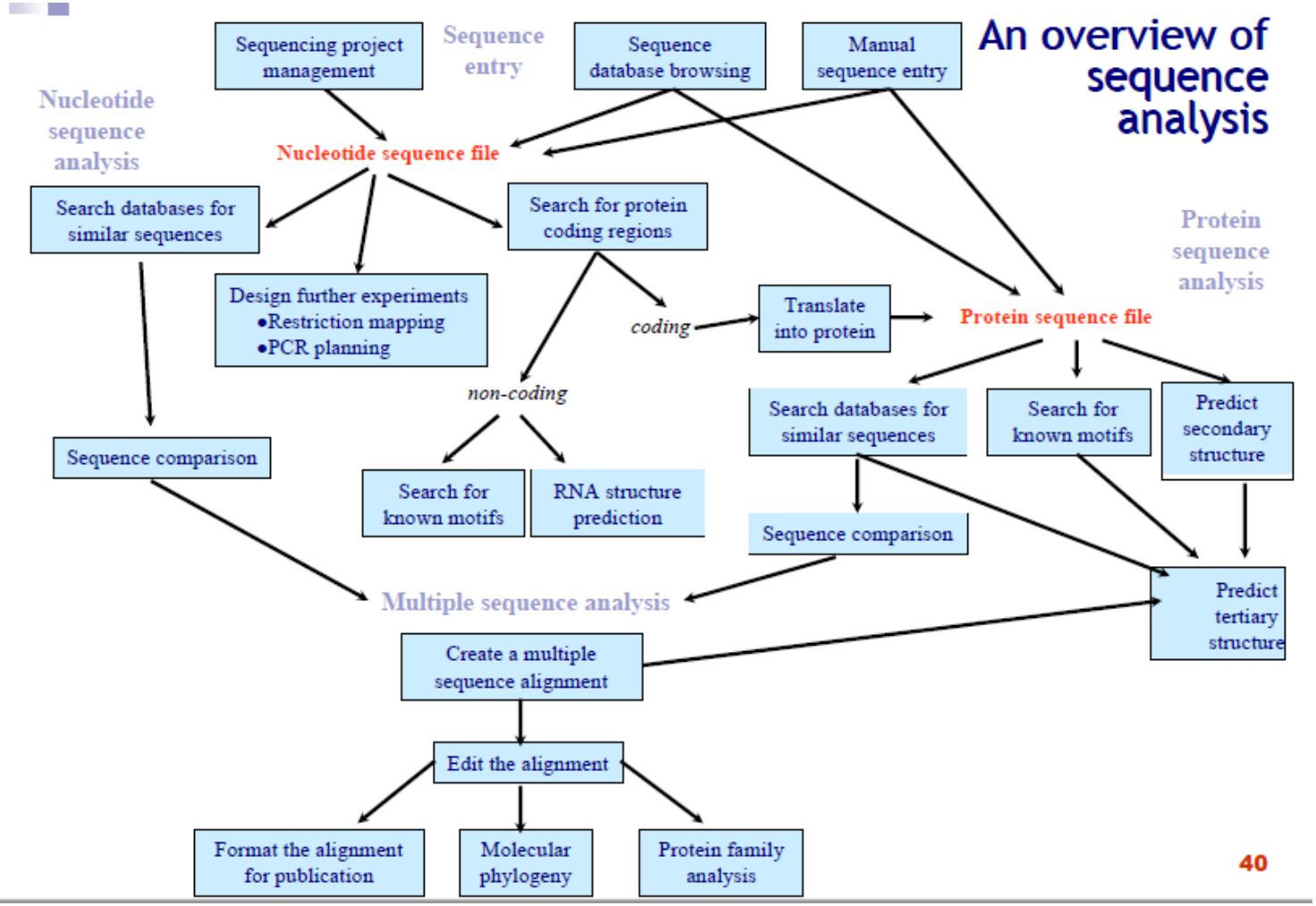


Ανάλυση διαδρομής

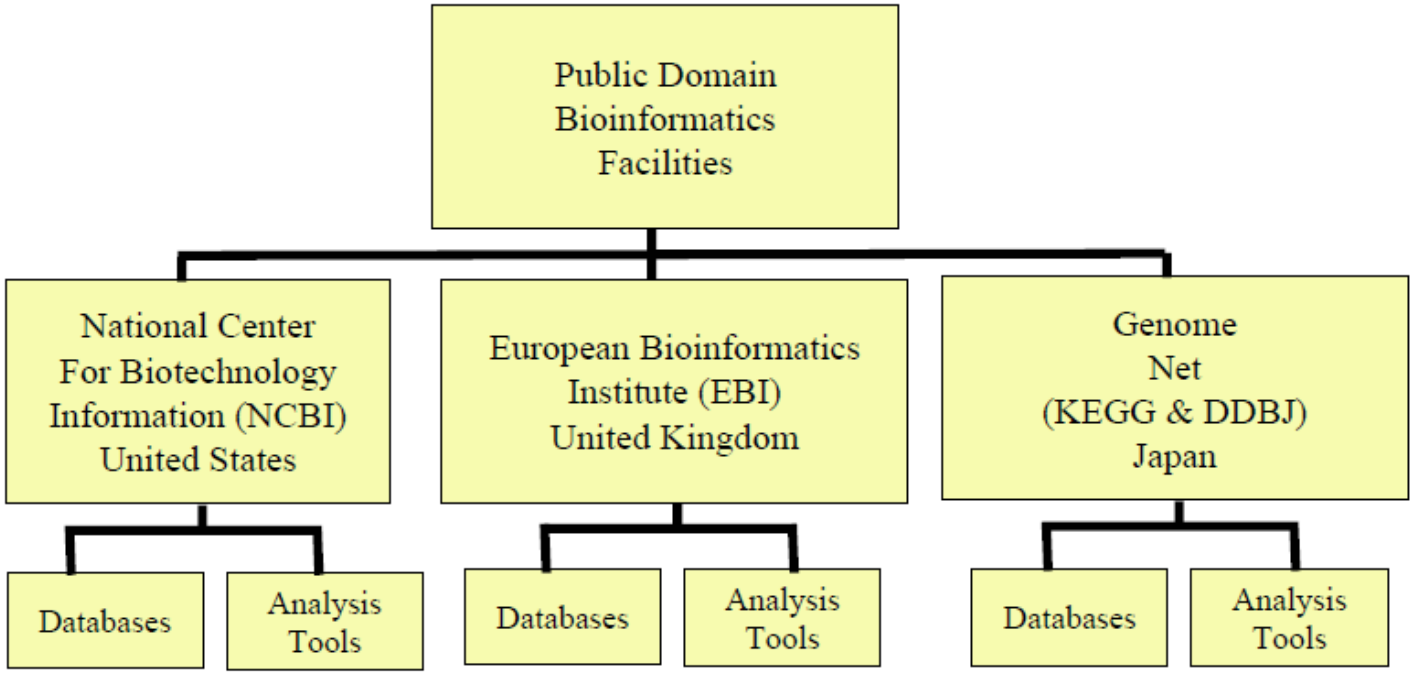


- Σε μια χημική αντίδραση αλληλομετατρέπονται χημικές ενώσεις
- Ένα ένζυμο είναι μια πρωτεΐνη που επιταχύνει τη χημική αντίδραση
- Ένα μονοπάτι είναι ένα συνδεδεμένο σύνολο αντιδράσεων

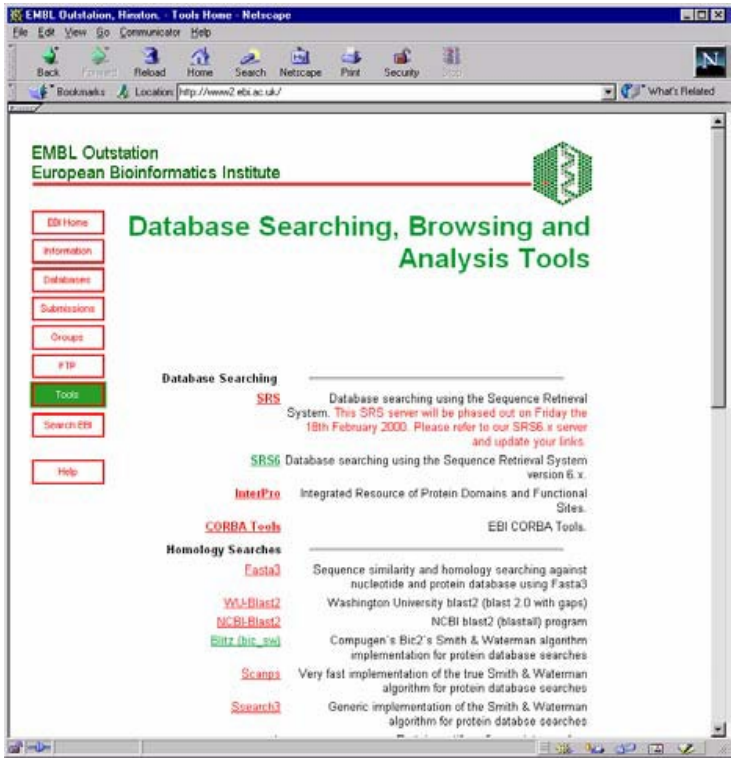
Co-funded by the Erasmus+ Programme of the European Union



Διακομιστές



Εργαλεία Ανάλυσης



Το EBI διατηρεί εκδόσεις των βασικών εργαλείων αναζήτησης και ανάλυσης ακολουθιών του δημόσιου τομέα (public domain), π.χ. Υλοποιήσεις FASTA, CLUSTALW, BLAST και Smith & Waterman.

Co-funded by the Erasmus+ Programme of the European Union

Προκλήσεις στη Βιοπληροφορική

Η βιοπληροφορική απαιτεί:

- Πρόσβαση σε πολλαπλούς κατανεμημένους πόρους
- Χρειάζεται οι πληροφορίες να είναι ενημερωμένες
- Ελάχιστος πλεονασμός δεδομένων
- Στιβαρές εφαρμογές
- Επεκτάσιμες εφαρμογές
  - Μονολιθική εφαρμογή έναντι εξαρτημάτων
- Φορητό λογισμικό

Co-funded by the Erasmus+ Programme of the European Union

Προκλήσεις στη Βιοπληροφορική

Έκρηξη πληροφοριών

- Ανάγκη για ταχύτερη, αυτοματοποιημένη ανάλυση για την επεξεργασία μεγάλων ποσοτήτων των δεδομένων
- Ανάγκη για ενοποίηση μεταξύ διαφορετικών τύπων πληροφοριών (αλληλουχίες, βιβλιογραφία, σχολιασμοί, επίπεδα πρωτεΐνης, επίπεδα RNA κλπ ...)
- Ανάγκη για πιο «έξυπνο» λογισμικό για τον εντοπισμό ενδιαφέροντος σχέσεις σε πολύ μεγάλα σύνολα δεδομένων

Έλλειψη «βιοπληροφορικών»

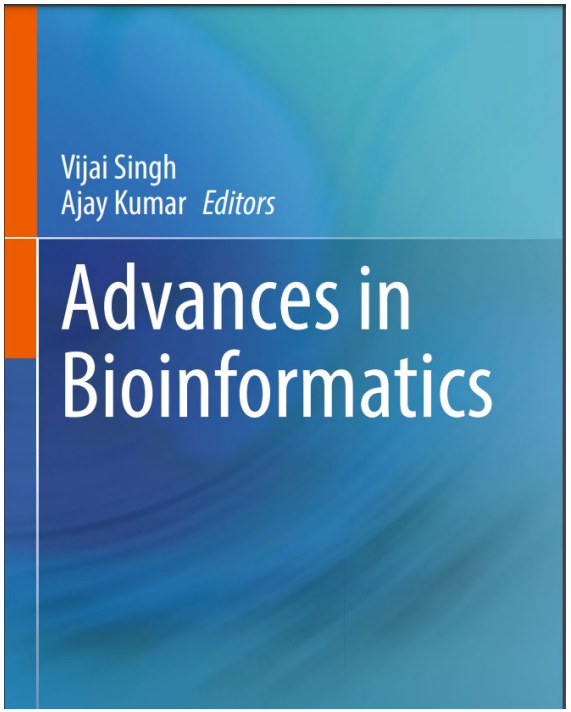
- Το λογισμικό πρέπει να είναι ευκολότερο στην πρόσβαση, τη χρήση και την κατανόηση
- Οι βιολόγοι πρέπει να μάθουν για το λογισμικό, τους περιορισμούς του, και πώς να ερμηνεύσετε τα αποτελέσματά του

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Ποιοι είναι οι κύριοι στόχοι της βιοπληροφορικής;
- Δώστε μερικά παραδείγματα των προκλήσεων στη βιοπληροφορική.

Co-funded by the Erasmus+ Programme of the European Union



Vijai Singh, Ajay Kumar (2021) Προόδους στη Βιοπληροφορική Springer Nature Singapore Pte Ltd

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

Θέμα 1. Εισαγωγή στη Βιοπληροφορική

Μάθημα 3. Συνιστώσες της Βιοπληροφορικής

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

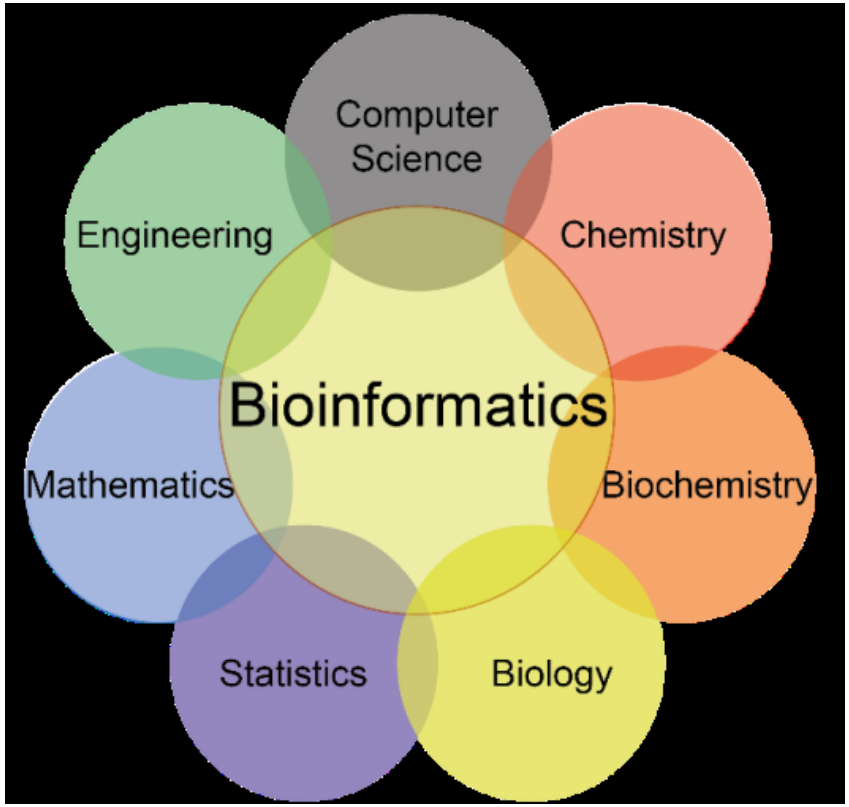
Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Δεδομένα
- Βάσεις δεδομένων
- Διάφορες κατηγορίες Αναλύσεων

Co-funded by the Erasmus+ Programme of the European Union

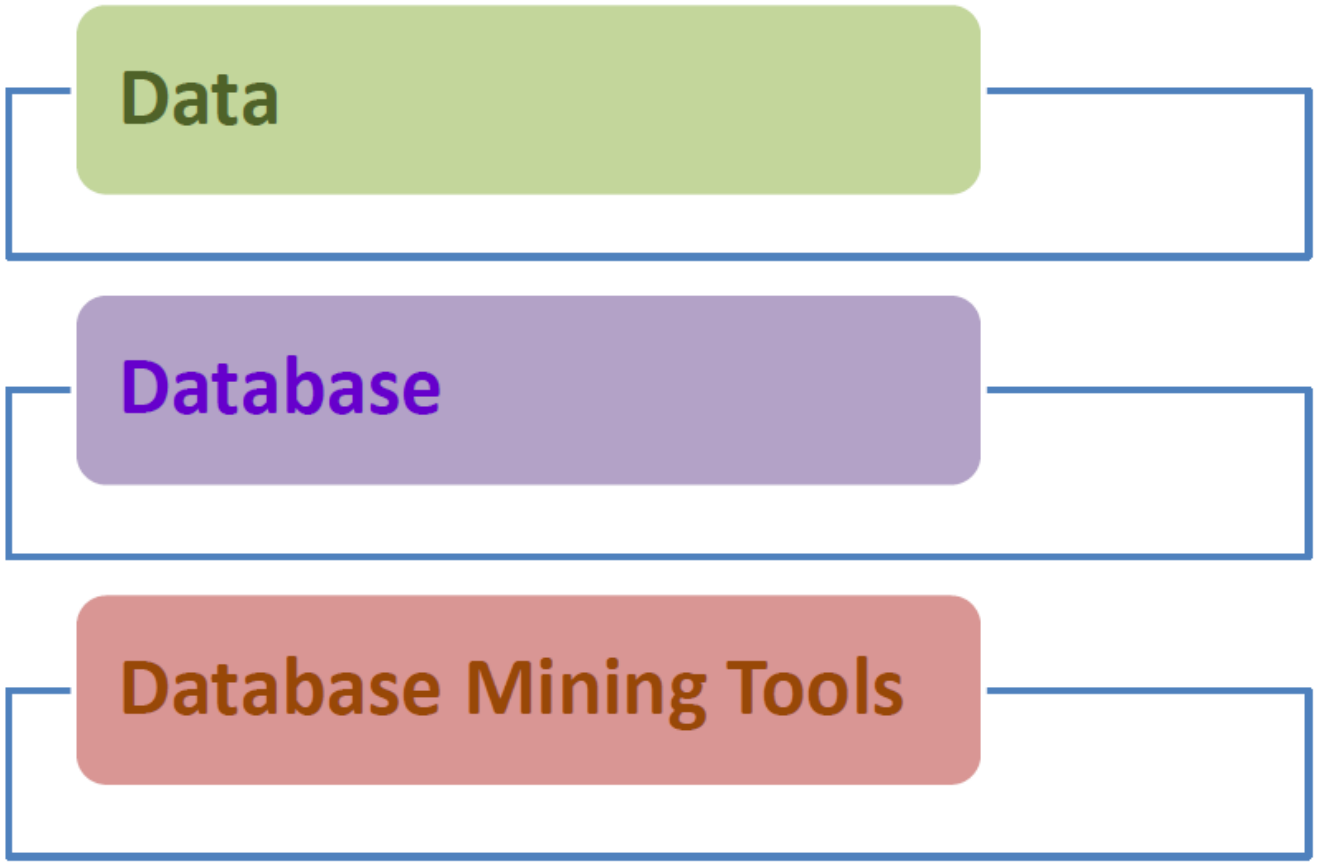
Εισαγωγή



Η βιοπληροφορική είναι ένας κλάδος της επιστήμης που ενσωματώνει την επιστήμη των υπολογιστών, τα μαθηματικά και τη στατιστική, τη χημεία και τη μηχανική για ανάλυση, εξερεύνηση, ολοκλήρωση και εκμετάλλευση δεδομένων βιολογικών επιστημών, στην Έρευνα και Ανάπτυξη

Η βιοπληροφορική ασχολείται με την αποθήκευση, ανάκτηση, ανάλυση και ερμηνεία βιολογικών δεδομένων χρησιμοποιώντας λογισμικό και εργαλεία που βασίζονται σε υπολογιστή

Co-funded by the Erasmus+ Programme of the European Union



Co-funded by the Erasmus+ Programme of the European Union

Δεδομένα

- **Αλληλουχίες νουκλεϊκών οξέων**
  - Ακατέργαστες αλληλουχίες DNA
  - Ετικέτες γονιδιωματικής ακολουθίας (GSTs)
  - αλληλουχίες cDNA
  - Ετικέτες εκφρασμένης ακολουθίας (ESTs)
  - Οργανικές αλληλουχίες DNA
  - Αλληλουχίες RNA
- **Αλληλουχίες πρωτεϊνών**
- **Πρωτεϊνικές δομές**
- **Μεταβολικά μονοπάτια**
- **Εικόνες τζελ**
- **Βιβλιογραφία**

Co-funded by the Erasmus+ Programme of the European Union

Βάσεις δεδομένων

Μια βάση δεδομένων είναι μια τεράστια συλλογή δεδομένων που σχετίζονται με ένα συγκεκριμένο θέμα π.χ. αλληλουχία νουκλεοτιδίων, αλληλουχία πρωτεϊνών κ.λπ. σε ένα ηλεκτρονικό περιβάλλον

- **Είναι η καρδιά της βιοπληροφορικής**
- **Μηχανογραφημένη αποθήκη δεδομένων (αρχεία).**
- **Επιτρέπει την εξαγωγή καθορισμένων εγγραφών**
- **Επιτρέπει την προσθήκη, αλλαγή, αφαίρεση και συγχώνευση εγγραφών**
- **Χρησιμοποιεί τυποποιημένες μορφές**

Co-funded by the Erasmus+ Programme of the European Union

Βάσεις δεδομένων: Τύποι

- Βάσεις δεδομένων ακολουθίας
- Δομικές Βάσεις Δεδομένων
- Βάσεις δεδομένων ενζύμων
- Βάσεις δεδομένων μικροσυστοιχίας
- Κλινική Βάση Δεδομένων
- Βάσεις δεδομένων διαδρομής
- Βάσεις Χημικών Δεδομένων
- Ενσωματωμένες βάσεις δεδομένων
- Βιβλιογραφικές Βάσεις Δεδομένων

Βάσεις Δεδομένων Ακολουθιών Νουκλεοτιδίων

NCBI GenBank: www.ncbi.nlm.nih.gov/GenBank)

EMBL: www.ebi.ac.uk/embl

DDBJ: www.ddbj.nig.ac.jp)

Οι 3 βάσεις δεδομένων ενημερώνονται και ανταλλάσσονται σε καθημερινή βάση και οι αριθμοί προσχώρησης είναι συνεπείς.

Δεν υπάρχει νομικός περιορισμός στη χρήση αυτών των βάσεων δεδομένων. Ωστόσο, υπάρχουν κάποιες κατοχυρωμένες με δίπλωμα ευρεσιτεχνίας ακολουθίες στη βάση δεδομένων.

The International Nucleotide Sequence Database Collaboration (INSD)

Co-funded by the Erasmus+ Programme of the European Union

Βάσεις δεδομένων πρωτεϊνών

Βάση δεδομένων αλληλουχίας πρωτεϊνών

- Λειτουργεί ως αποθήκη ακατέργαστων δεδομένων: δύο τύποι
- Πρωταρχικός
- Δευτερεύων

Βάση δεδομένων δομής πρωτεΐνης

Co-funded by the Erasmus+ Programme of the European Union

Πρωτογενείς βάσεις δεδομένων

Ομάδες SWISS PROT στο Swiss Institute of Bioinformatics (

Σημειώνει τις ακολουθίες

Περιγράφει τις λειτουργίες της πρωτεΐνης

Οι δομές τομέα του

Οι τροποποιήσεις των μεταφράσεων του

Παρέχει υψηλό επίπεδο σχολιασμού

Ελάχιστο επίπεδο πλεονασμού

Υψηλό επίπεδο ενοποίησης με άλλες βάσεις δεδομένων

ΤΡΕΜΒΛ

Συμπληρώματα με επισημειώσεις του SWISS PROT που περιέχει όλες τις μεταφράσεις των καταχωρήσεων νουκλεοτιδίων EMBL που δεν έχουν ακόμη ενσωματωθεί στο SWISS PROT.

PIR Protein Information Resource, τμήμα της NBRF στις ΗΠΑ

Συνεργάστηκε με το Κέντρο Πληροφοριών του Μονάχου για Αλληλουχίες Πρωτεϊνών (και Ιαπωνική Διεθνής Βάση Δεδομένων Αλληλουχιών Πρωτεϊνών (

Αναζήτηση για καταχωρήσεις

Ομοιότητα ακολουθίας

Το PIR παράγει επίσης MRL 3D db αλληλουχιών που εξάγονται από δομές 3D στο PDB)

Δευτερεύουσες βάσεις δεδομένων

Δευτερεύουσα db μεταγλωττίζει και φιλτράρει δεδομένα αλληλουχίας από διαφορετικά πρωτεύοντα db Αυτά τα db περιέχουν πληροφορίες που προέρχονται από αλληλουχίες πρωτεϊνών και βοηθούν τον χρήστη να προσδιορίσει εάν μια νέα αλληλουχία ανήκει σε μια γνωστή οικογένεια πρωτεϊνών

PROSITE

db μοτίβων και προφίλ σύντομης αλληλουχίας πρωτεϊνών που χαρακτηρίζουν βιολογικά σημαντικές θέσεις σε πρωτεΐνες

Βασίζεται σε κανονικές εκφράσεις που περιγράφουν χαρακτηριστικές αλληλουχίες συγκεκριμένων οικογενειών και περιοχών πρωτεΐνης

Είναι μέρος του SWISS PROT και διατηρείται με τον ίδιο τρόπο

PRINTS

Το PRINTS παρέχει μια συλλογή αποτυπωμάτων πρωτεΐνης (ομάδες διατηρημένων μοτίβων που χαρακτηρίζουν μια οικογένεια πρωτεϊνών)

Τώρα έχει μια σχεσιακή έκδοση, έκδοση "PRINTS S"

BLOCKS

Μοτίβα χωρίς κενά σε ευθυγραμμισμένες οικογένειες πρωτεϊνών που ορίζονται από το PROSITE, που βρέθηκαν με αναζήτηση μοτίβων και αλγόριθμους στατιστικής δειγματοληψίας Αυτόματος προσδιορισμός διατηρημένων τμημάτων χωρίς διάκενα

Pfam

Db οικογενειών πρωτεϊνών που ορίζονται ως τομείς

Για κάθε τομέα, περιέχει μια πολλαπλή στοίχιση ενός συνόλου καθοριστικών ακολουθιών και τις άλλες ακολουθίες σε SWISS PROT και TrEMBL που μπορούν να αντιστοιχιστούν στη στοίχιση

Co-funded by the Erasmus+ Programme of the European Union

Πρωτεϊνική Δομική Βάση Δεδομένων

PDB (Protein Data Bank)

Κύρια db τρισδιάστατων δομών βιολογικών μακρομορίων (που προσδιορίζονται με κρυσταλλογραφία ακτίνων X και NMR)

Οι καταχωρήσεις PDB περιέχουν τις ατομικές συντεταγμένες και ορισμένες δομικές παραμέτρους που συνδέονται με τα άτομα ή υπολογίζονται από τις δομές (δευτερεύουσα δομή)

Το PDB παρέχει πρωτεύον αρχείο και των 3D δομών για μακρομόρια όπως πρωτεΐνες, DNA, RNA και διάφορα σύμπλοκα

SCOP (Δομική Ταξινόμηση Πρωτεϊνών):

Η Db ξεκίνησε με στόχο να ταξινομήσει τις δομές πρωτεΐνης 3D σε ένα ιεραρχικό σχήμα δομικών τάξεων

Βασίζεται σε δεδομένα σε ένα πρωτεύον db , αλλά προσθέτει πληροφορίες μέσω ανάλυσης και οργάνωσης (όπως ταξινόμηση δομών 3D σε ιεραρχικό σχήμα πτυχών, υπεροικογενειών και οικογενειών)

CATH (αρχιτεκτονική, τοπολογία, ομόλογη σούπερ οικογένεια)

Η CATH εκτελεί ιεραρχική ταξινόμηση δομών πρωτεϊνικών τομέων Συγκεντρώνει πρωτεΐνες σε τέσσερα κύρια δομικά επίπεδα

Co-funded by the Erasmus+ Programme of the European Union

Εργαλεία εξόρυξης βάσεων δεδομένων (Εργαλεία ανάλυσης)

Analysis Tool	Function
BLAST (NCBI, USA)	Used to analyse sequence information and detect homologous sequences
ENTREZ (NCBI, USA)	Used to access literature (abstracts), sequence and structure db
DNAPLOT (EBI, UK)	Sequence alignment tool
LOCUS LINK (NCBI, USA)	Assessing information on homologous genes
LIGAND (GenomNet, Japan)	A chemical db, allows search for a combination of enzymes and links to all publically accessible db.
BRITE (GenomNet, Japan)	Biomolecular relations information transmission and expression db; links to all publically accessible db.
TAXONOMY BROWSER (NCBI, USA)	Taxonomic classification of various species as well as genetic information
STRUCTURE	It support Molecular Modelling Database (MMDB) and software tools for structure analysis

Διάφορες κατηγορίες Αναλύσεων

Ανάλυση αλληλουχίας μεμονωμένου γονιδίου (πρωτεΐνης).

Ομοιότητα με άλλα γνωστά γονίδια

Φυλογενετικά δέντρα; εξελικτικές σχέσεις

Προσδιορισμός καλά καθορισμένων τομέων στην ακολουθία

Χαρακτηριστικά αλληλουχίας (φυσικές ιδιότητες, θέσεις σύνδεσης, θέσεις τροποποίησης)

Πρόβλεψη υποκυτταρικού εντοπισμού

Πρόβλεψη πρωτεϊνικών δευτερογενών και τριτογενών δομών

Διάφορες κατηγορίες Αναλύσεων

Ανάλυση ολόκληρων γονιδιωμάτων

Θέση των γονιδίων variuos στα χρωμοσώματα, συσχέτιση με λειτουργία ή εξέλιξη

Επέκταση/διπλασιασμός οικογενειών γονιδίων

Ποιες οικογένειες γονιδίων υπάρχουν, ποιες λείπουν;

Παρουσία ή απουσία βιοχημικών οδών

Προσδιορισμός ενζύμων «που λείπουν».

Γεγονότα μεγάλης κλίμακας στην εξέλιξη των οργανισμών

Διάφορες κατηγορίες Αναλύσεων

Ανάλυση γονιδίων και γονιδιωμάτων σε σχέση με τη λειτουργία (Functional Annotation)

Transcriptomics Ανάλυση έκφρασης δεδομένων μικροσυστοιχίας (mRNA/αναλύσεις μεταγραφής)

Ποιοτικές και ποσοτικές αναλύσεις πρωτεϊνών Proteomics, ομοιοπολικές τροποποιήσεις

Σύγκριση και ανάλυση βιοχημικών οδών

Διαγραφή ή μεταλλαγμένοι γονότυποι έναντι φαινοτύπων

Προσδιορισμός βασικών γονιδίων ή γονιδίων που εμπλέκονται σε συγκεκριμένες διαδικασίες

Co-funded by the Erasmus+ Programme of the European Union

Διάφορες κατηγορίες Αναλύσεων

Συγκριτική γονιδιωματική

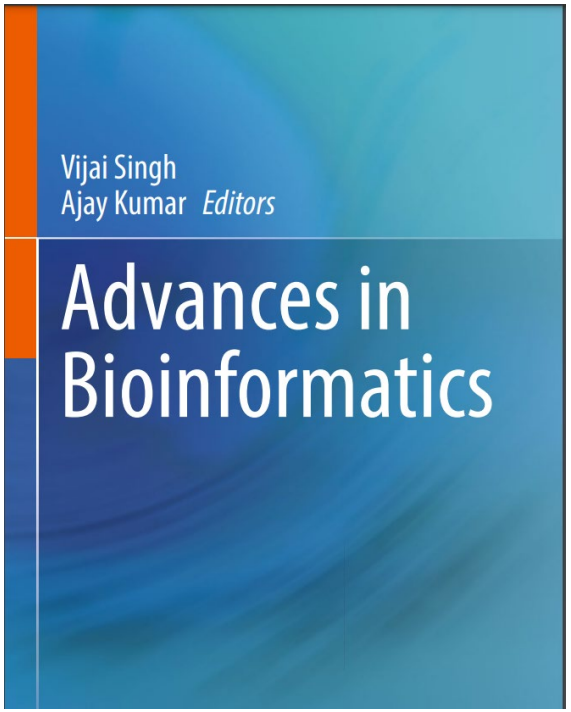
Προσδιορισμός μοναδικών στόχων συγκεκριμένων παθογόνων για το σχεδιασμό νέων φαρμάκων

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Ποια είναι τα συστατικά στοιχεία της βιοπληροφορικής;
- Δώστε μερικά παραδείγματα δεδομένων, βάσης δεδομένων , εργαλείων εξόρυξης βάσεων δεδομένων.

Co-funded by the Erasmus+ Programme of the European Union



Vijai Singh, Ajay Kumar (2021) Προόδους στη Βιοπληροφορική Springer Nature Singapore Pte Ltd

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

Θέμα 2. Πρόβλημα σύνδεσης

Μάθημα 1. Πού να βρεθούν δομές πρωτεΐνης και προσδέματος

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Βάση δεδομένων ΠΣΠ
- Μοντελοποίηση ομολογίας
- Οι δομές των προσδεμάτων
- Βάσεις δεδομένων μικρών μοριακών δομών

Co-funded by the Erasmus+ Programme of the European Union

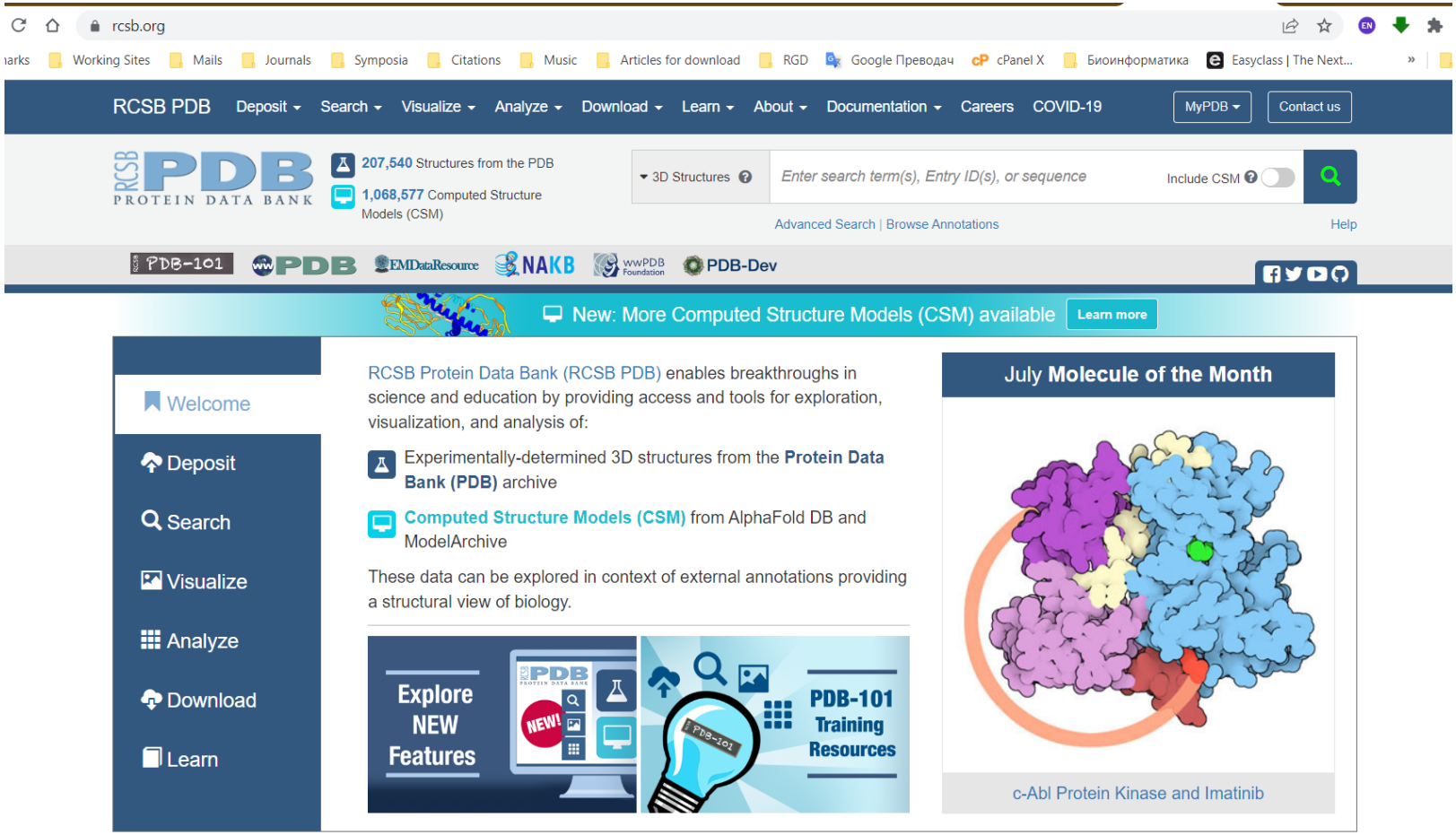
Εισαγωγή

Οι μοριακές αναγνωρίσεις συμπεριλαμβανομένων των αλληλεπιδράσεων ενζύμου-υποστρώματος, φαρμάκου-πρωτεΐνης, φαρμάκου-νουκλεϊκού οξέος, πρωτεΐνης-νουκλεϊκού οξέος και πρωτεΐνης-πρωτεΐνης παίζουν σημαντικό ρόλο σε πολλές βιολογικές διεργασίες όπως η μεταγωγή σήματος, η κυτταρική ρύθμιση και άλλες μακρομοριακές συνθέσεις. Επομένως, ο προσδιορισμός του τρόπου δέσμευσης και της συγγένειας μεταξύ των συστατικών μορίων στη μοριακή αναγνώριση είναι ζωτικής σημασίας για την κατανόηση των μηχανισμών αλληλεπίδρασης και για το σχεδιασμό θεραπευτικών παρεμβάσεων. Λόγω των δυσκολιών και του οικονομικού κόστους των πειραματικών μεθόδων για τον προσδιορισμό των δομών των συμπλόκων, οι υπολογιστικές μέθοδοι όπως η μοριακή σύνδεση είναι επιθυμητές για την πρόβλεψη υποτιθέμενων τρόπων σύνδεσης και συγγένειες.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Βάση δεδομένων ΠΣΠ



Co-funded by the Erasmus+ Programme of the European Union

Βάση δεδομένων ΠΣΠ

Καθορίσει είτε έχει κατατεθεί ο αντίστοιχος στόχος είτε όχι.

- **Ένας τρόπος για να γίνει αυτό είναι να προσδιοριστεί εάν υπάρχει πρωτεΐνη στο ΠΣΠ με μια ακολουθία που είναι 100% ίδια με αυτήν του πρωτεϊνικού στόχου.**
- **Εάν η πρωτεΐνη στόχος δεν έχει ακόμη κατατεθεί στο ΠΣΠ αλλά η βάση δεδομένων περιέχει μια πρωτεΐνη με παρόμοια αλληλουχία, τότε η μοντελοποίηση ομολογίας μπορεί να χρησιμοποιηθεί για την πρόβλεψη της τρισδιάστατης δομής της πρώτης πρωτεΐνη.**

Co-funded by the Erasmus+ Programme of the European Union

Μοντελοποίηση Ομολογίας

Τα προγράμματα ή οι διακομιστές ιστού επιτρέπουν στους χρήστες να δημιουργούν αυτά τα μοντέλα ελέγχοντας αυστηρά τις τιμές πολλών παραμέτρων, ορισμένες βάσεις δεδομένων πρωτεϊνικών δομών που μοντελοποιούνται με ομολογία έχουν δημιουργηθεί αυτόματα από αυτά τα προγράμματα ή διακομιστές ιστού χρησιμοποιώντας προεπιλεγμένες τιμές παραμέτρων κατά τη δημιουργία του μοντέλου ομολογίας.

Παραδείγματα τέτοιων βάσεων δεδομένων είναι η ModBase (http://modbase.compbio.ucsf.edu/modbase-cgi/search_form.cgi) και η SWISS-MODEL

ModBase: Database of Comparative Protein Structure Models

Sali Lab Home | ModWeb | ModLoop | ModBase | ModEval | PCSS | FoXS | IMP | ModPipe

ModBase Home | ModBase Datasets for User:Anonymous | User Login | Help | News | Contact | Current Datasets

- General Information
- Statistics and Genome Datasets
- News
- Project Pages
- Authors and Acknowledgements
- Publications
- Downloads
- Related Resources

Please address inquiries to: modbase@salilab.org

MODBASE is *no longer under active development*, and contains theoretically calculated models, not experimentally determined structures. The models may contain *significant* errors.

ModBase Search

ModBase is a database of comparative protein structure models, calculated by our modeling pipeline ModPipe.

Search

Search type: Model(Default) | Display type: Model Detail (graphical)

All available datasets are selected

To include the academic (comprehensive) dataset, go to Current Datasets!

Search by properties

Property: Database Accession Number — A8K5B1

Organism: ALL — or

Advanced search

Search

Model Details Example

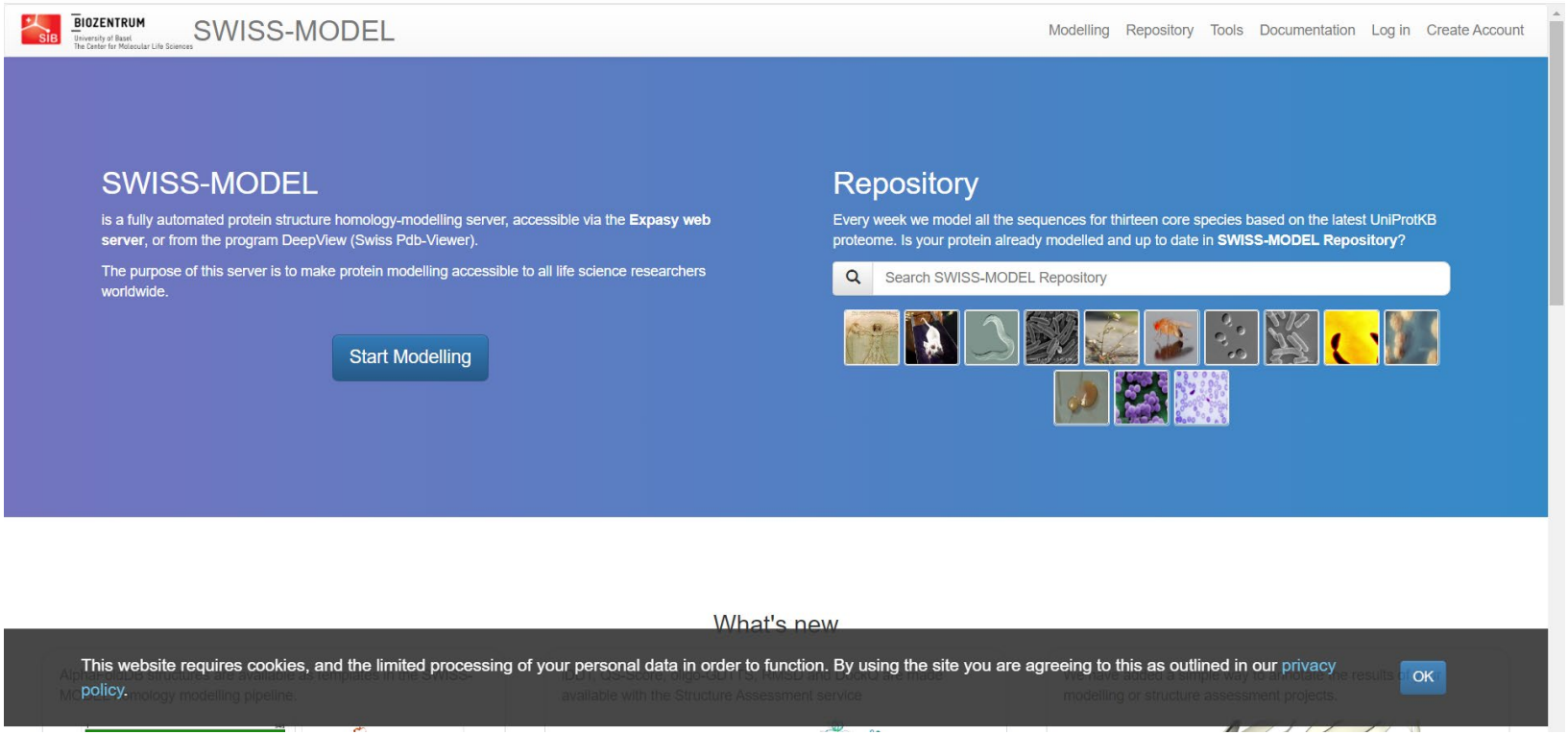
Database ID: Q12321

This page makes all models and model details for one sequence available.

Overview Example

Dengue Virus

Models for several sequences are displayed on the **Sequence Overview** page.


BIOZENTRUM
University of Basel
The Center for Molecular Life Sciences
SWISS-MODEL
Modelling
Repository
Tools
Documentation
Log in
Create Account
SWISS-MODEL
is a fully automated protein structure homology-modelling server, accessible via the Expasy web server, or from the program DeepView (Swiss Pdb-Viewer).
The purpose of this server is to make protein modelling accessible to all life science researchers worldwide.
Start Modelling
Repository
Every week we model all the sequences for thirteen core species based on the latest UniProtKB proteome. Is your protein already modelled and up to date in SWISS-MODEL Repository?
Search SWISS-MODEL Repository
What's new
This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our privacy policy.
OK

Co-funded by the Erasmus+ Programme of the European Union

Βάσεις δεδομένων πρωτεϊνών

Αποθετήριο (http://swissmodel.expasy.org/repository/) του οποίου τα μοντέλα είναι εύκολα προσβάσιμα από συνδέσμους εντός των κατάλληλων Εγγραφών UniProt Knowledgebase. Σε αυτό το σημείο, αξίζει να επισημανθεί ότι πριν από μερικά χρόνια, η προσάρτηση σε ενεργές πρωτεϊνικές θέσεις που δημιουργήθηκαν με ομολογία αλληλουχίας θεωρούνταν πηγή αναξιόπιστων αποτελεσμάτων. Ωστόσο, η βελτίωση των μεθόδων μοντελοποίησης ομολογίας έχει αλλάξει ριζικά αυτή την άποψη και έχουν πλέον χρησιμοποιηθεί με επιτυχία σε πειράματα σύνδεσης.

Οι δομές των προσδεμάτων

Υπάρχουν δύο κύριες πηγές δομών συνδέτη.

- **Η πρώτη πηγή είναι προγράμματα υπολογιστών που χρησιμοποιούν γραφικά εργαλεία ή γλώσσες όπως το SMILES (δηλ. Απλοποιημένο Σύστημα Εισόδου Γραμμής Μοριακής Εισόδου) για τη δημιουργία αυτών των δομών.**
- **Η δεύτερη πηγή είναι βάσεις δεδομένων με, μερικές φορές, ενώσεις που μπορούν να αγοραστούν.**

Co-funded by the Erasmus+ Programme of the European Union

SMILES

Παραδείγματα τέτοιων προγραμμάτων είναι:

(a) ChemDraw UltraTM (CambridgeSoft Corporation, Cambridge MA, ΗΠΑ, http://www.cambridgesoft.com/);

(b) KnowItAll ChemWindow TM (Bio-Rad Laboratories Inc., Hercules CA, USA, http://www.bio-rad.com/);

(c) ISIS/ DrawTM (MDL Information Systems, Inc., San Ramon CA, USA, http:// www.mdli.com/); και

(d) ACD/ ChemSketchTM (Advanced Chemistry Development, Inc., Τορόντο, Καναδάς, http://www.acdlabs.com/).

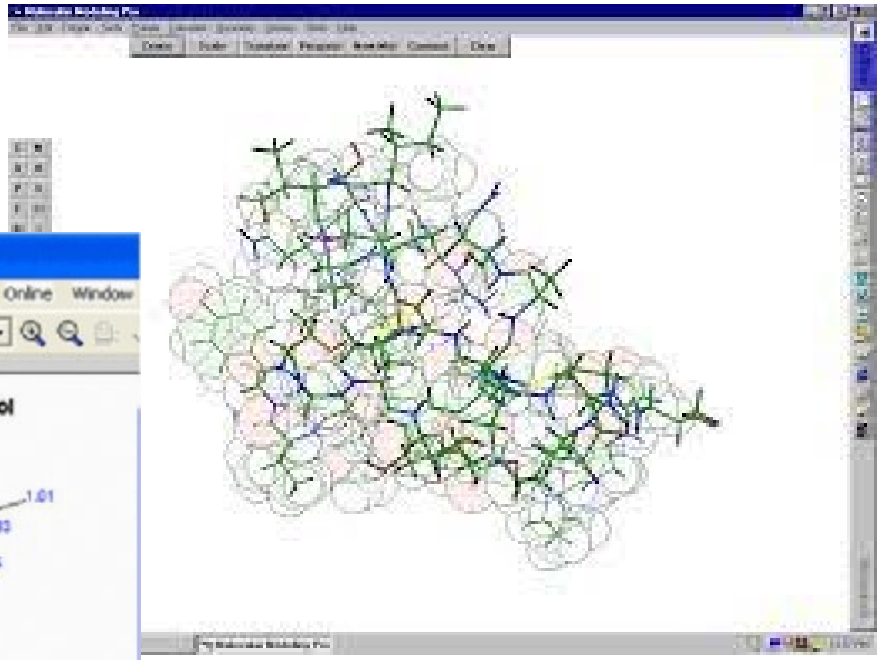
Αυτά τα προγράμματα είναι χρήσιμα μόνο εάν το σύνολο των προσδεμάτων είναι περιορισμένο ή εάν τα μόρια δεν έχουν προηγουμένως αναφερθεί.

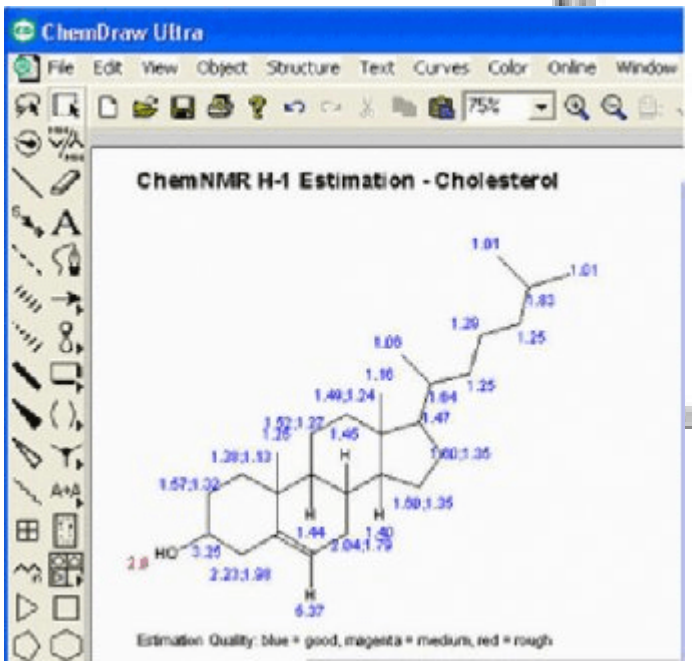
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

ChemDraw Ultra

ChemDraw®

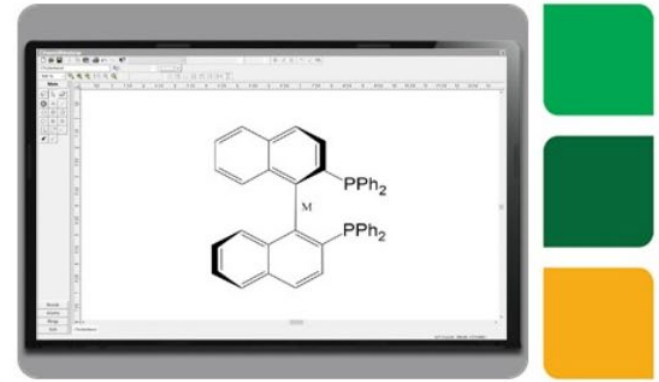




Co-funded by the Erasmus+ Programme of the European Union

KnowItAll ChemWindow

ChemWindow Chemical Structure Drawing Software



ISIS/DRAW

ISIS/Draw

ISIS/Draw is a chemically intelligent drawing program that understands the fundamentals of chemistry such as valence limits, bond angles, and aromatic ring systems. This understanding enables users to create chemical graphics intuitively, which can be inserted into documents or used to build and search databases of 2D and 3D molecules, polymers, and reactions.

Create Publication Quality Graphics

With ISIS/Draw, scientists can graphically represent almost any kind of chemical or polymer structure for use in reports, publications, and presentations.

- **Sketch large biomolecules with ease and speed:** Draw complicated structures including DNA, RNA, and amino acid sequences using special shortcut tools.
- **Create many types of polymer structures**: including homopolymers; source-based polymers; structure-based polymers; copolymers; crosslinked polymers; branched polymers; graft polymers; and statistical, random, and unspecified copolymers.
- **Control the display of your structures:** Scale atom labels, bonds, and other chemical attributes proportionally to make sketches easier to read.
- **Speed drawing using templates:** Create structures by piecing together predrawn templates of commonly used structures. Design your own template structures to reuse later.

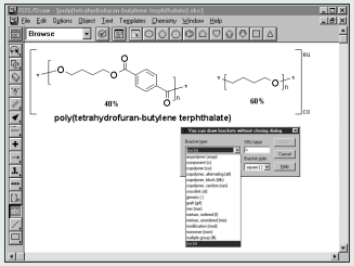


Figure 1: *When drawing polymers in ISIS/Draw, the bracket tool presents the user with a full range of polymer types from which to choose. Data such as weight percentages can be attached to polymer components and searched on later as part of the database.*

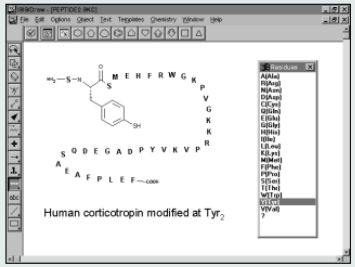
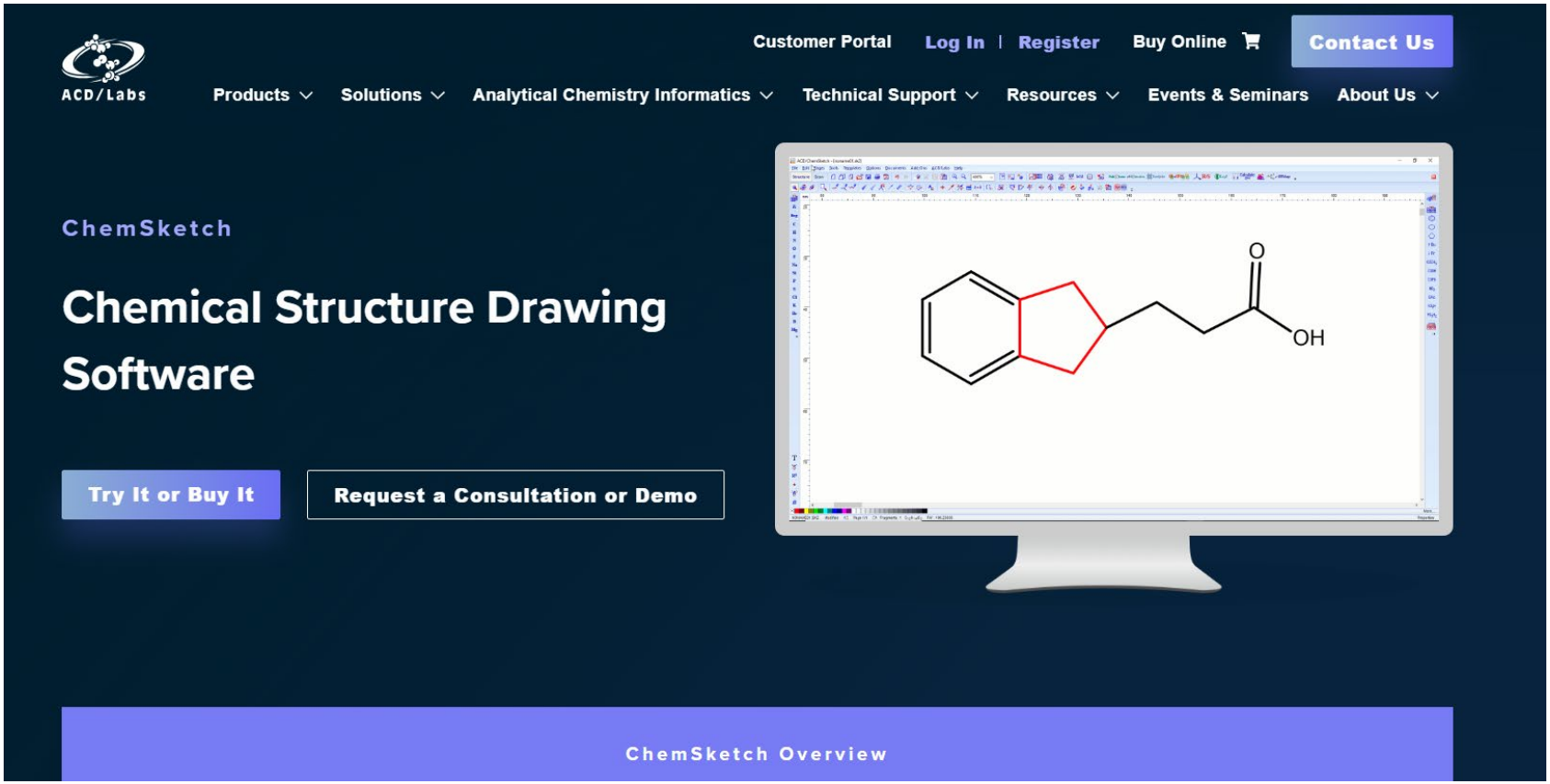


Figure 2: *Flexible tools in ISIS/Draw make it easy to draw peptides and other large molecules. The Sequence Tool gives you two options: Type in the amino acid abbreviations to build the structure, or simply click on the appropriate amino acid residues in the Residue List. To complete the drawing, orient the structure using the Sequence Shape Tool.*

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

ACD/ ChemSketch



Βάσεις δεδομένων μικρών μοριακών δομών

Παραδείγματα τέτοιων βάσεων δεδομένων χημικών είναι:

(a) το Available Chemicals DirectoryTM (http://www.mdl.com/products/experiment/available_chem_dir), ο οποίος περιέχει 571000 μοναδικές χημικές ουσίες που μπορούν να αγοραστούν από περισσότερους από 719 προμηθευτές σε όλο τον κόσμο.

(b) τον κατάλογο ενώσεων διαλογής MDL® (πρώην ACD-SCTM; http://www.mdl.com/products/experiment/screening_compounds/_) όπου μπορούν να βρεθούν 3,4 εκατομμύρια δομές από 46 διεθνείς προμηθευτές χημικών.

(c) τη Βιβλιοθήκη iResearchTM (http://www.chemnavigator.com/cnc/products/iRL.asp), η οποία περιέχει πάνω από 37,2 εκατομμύρια χημικές δομές (περίπου 20,6 εκατομμύρια μοναδικές) από περισσότερους από 252 προμηθευτές χημείας.

(d) τη βάση δεδομένων του Εθνικού Ινστιτούτου Καρκίνου (http://129.43.27.140/ncidb2/download.html) με 250251 ενώσεις.

Βάσεις δεδομένων μικρών μοριακών δομών

Παραδείγματα τέτοιων βάσεων δεδομένων χημικών είναι:

e) η βάση δεδομένων Ligand.Info (http://Ligand.info/), η οποία είναι μια συλλογή από διάφορες δημόσιες βάσεις δεδομένων μικρών μορίων [π.χ. ChemBank (http://chembank.broad.harvard .edu/) , ChemPDB (http://www.ebi.ac.uk/msd-srv/msdchem/cgi-bin/cgi.pl), KEGG (http:// www. genome.ad.jp/ligand/), NCI (http : / / dtp . nci . nih . gov /) , AK os G mb H (http : / / ww w . akosg mb h . eu /) , A sinex L td (http://www.asinex.com/), και TimTec (http://www.timtec.net/)] και περιέχει 1159274 καταχωρήσεις.

f) (στ) τη βάση δεδομένων GDB (δηλαδή την Generated Database of Chemical Space of SmallMolecules;http ://www.dcb.unibe.ch/groups/reymond/), η οποία περιέχει 26,4 εκατομμύρια ενώσεις, η συντριπτική πλειονότητα των οποίων δεν έχει ποτέ συντεθεί ; και

g) (ζ) τη βάση δεδομένων ZINC (http://blaster.docking.org/zinc/), η οποία περιέχει πάνω από 4,6 εκατομμύρια ενώσεις σε έτοιμη μορφή από τους καταλόγους 46 διαφορετικών προμηθευτών χημικών.

Βάσεις δεδομένων μικρών μοριακών δομών

Κάθε μόριο στη βιβλιοθήκη είναι έτοιμο για λήψη και χρήση από διάφορα δημοφιλή προγράμματα σύνδεσης και σχολιάζεται επίσης με τις ακόλουθες ιδιότητες:

(a) Μοριακό βάρος;

(b) πωλητής και αρχικός αριθμός καταλόγου·

(c) Υπολογισμένο LogP ;

(d) αριθμός περιστρεφόμενων δεσμών·

(e) αριθμός δοτών δεσμών υδρογόνου·

(f) αριθμός αποδεκτών δεσμών υδρογόνου·

(g) αριθμός χειρόμορφων κέντρων.

(h) αριθμός χειρόμορφων διπλών δεσμών (ισομέρεια E/Z).

(i) πολικό και πολικό ενέργεια αποδιάλυσης (σε kcal/mol).

(j) καθαρή χρέωση?

(k) αριθμός άκαμπτων θραυσμάτων· και

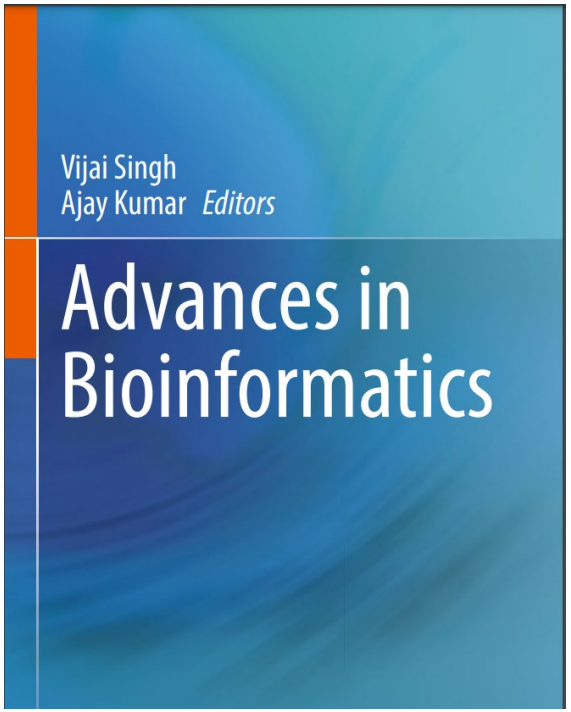
(l) λειτουργία ή δραστηριότητα (όταν είναι διαθέσιμη).

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Δώστε μερικά παραδείγματα των βάσεων δεδομένων πρωτεϊνών.
- Δώστε μερικά παραδείγματα βάσεων δεδομένων μικρών μοριακών δομών.
- Δώστε ένα παράδειγμα λογισμικού που χρησιμοποιείται για κτιριακές κατασκευές.

Co-funded by the Erasmus+ Programme of the European Union



Vijai Singh, Ajay Kumar (2021) Προόδους στη Βιοπληροφορική Springer Nature Singapore Pte Ltd

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Εργαλεία βιοπληροφορικής

Θέμα 2. Πρόβλημα σύνδεσης

Μάθημα 2. Πού μπορούν να προσδεθούν οι σύνδεσμοι

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Προγράμματα σύνδεσης

Εισαγωγή

Στη σύγχρονη ανακάλυψη φαρμάκων, πρωτεΐνη-συνδέτης ή πρωτεΐνη-πρωτεΐνη Η σύνδεση παίζει σημαντικό ρόλο στην πρόβλεψη του προσανατολισμού του συνδέτη όταν συνδέεται με έναν υποδοχέα πρωτεΐνης ή ένα ένζυμο χρησιμοποιώντας σχήμα και ηλεκτροστατικές αλληλεπιδράσεις για την ποσοτικοποίηση του. ο Οι αλληλεπιδράσεις van der Waals παίζουν επίσης σημαντικό ρόλο, σε προσθήκη στις κουλομβικές αλληλεπιδράσεις και το σχηματισμό υδρογόνου δεσμούς. Το άθροισμα όλων αυτών των αλληλεπιδράσεων υπολογίζεται κατά προσέγγιση από μια βαθμολογία σύνδεσης, η οποία αντιπροσωπεύει τη δυνατότητα δέσμευσης. Στα πιο απλά συστήματα άκαμπτου σώματος, γίνεται αναζήτηση του συνδετήρα σε έναν εξαδιάστατο περιστροφικό ή μεταφορικό χώρο για να χωρέσει τη θέση δέσμευσης, η οποία μπορεί να χρησιμεύσει ως ένωση οδηγός για το φάρμακο σχέδιο .

Co-funded by the Erasmus+ Programme of the European Union

AutoDock

Το AutoDock είναι μια σουίτα αυτοματοποιημένων εργαλείων σύνδεσης. Έχει σχεδιαστεί για να προβλέπει πώς μικρά μόρια, όπως υποστρώματα ή υποψήφια φάρμακα, συνδέονται με έναν υποδοχέα γνωστής τρισδιάστατης δομής. Με τα χρόνια, έχει τροποποιηθεί και βελτιωθεί για να προσθέσει νέες λειτουργίες και έχουν αναπτυχθεί πολλαπλοί κινητήρες.

Το AutoDock έχει εφαρμογές σε:

- **Κρυσταλλογραφία ακτίνων Χ;**
- **Σχεδιασμός φαρμάκων με βάση τη δομ**
- **βελτιστοποίηση μολύβδου?**
- **εικονική προβολή (HTS);**
- **συνδυαστικός σχεδιασμός βιβλιοθήκης.**
- **σύνδεση πρωτεΐνης-πρωτεΐνης?**
- **μελέτες χημικών μηχανισμών**

DOCK

Το DOCK 6 είναι γραμμένο σε C++ και χωρίζεται λειτουργικά σε ανεξάρτητα στοιχεία, επιτρέποντας υψηλό βαθμό ευελιξίας προγράμματος. Τα βοηθητικά προγράμματα είναι γραμμένα σε C και Fortran 77. Παρέχουμε τον πηγαίο κώδικα για όλα τα προγράμματα. Η σουίτα προγραμμάτων DOCK έχει μέτριες απαιτήσεις χώρου στο δίσκο και μνήμης.

Τα νέα χαρακτηριστικά του DOCK 6 περιλαμβάνουν: γενετικούς αλγόριθμους και de novo σχεδιασμό για αναζήτηση συνδέτη με βάση θραύσματα. πρόσθετες επιλογές βαθμολόγησης κατά την ελαχιστοποίηση. DOCK 3.5 βαθμολόγηση-συμπεριλαμβανομένης της ηλεκτροστατικής Delphi, των διορθώσεων διαμορφωτικής εντροπίας συνδέτη, της αποδιάλυσης προσδέματος, της αποδιάλυσης υποδοχέα . Βαθμολογία επίλυσης Hawkins-Cramer- Truhlar GB/SA με προαιρετικό έλεγχο αλατιού. και AMBER βαθμολόγησης-συμπεριλαμβανομένης της ευελιξίας του υποδοχέα, η πλήρης λειτουργία βαθμολόγησης της μοριακής μηχανικής AMBER με άρρητο διαλύτη, ελαχιστοποίηση συζευγμένης κλίσης και δυνατότητες προσομοίωσης μοριακής δυναμικής. Επειδή το DOCK 6 είναι μια επέκταση του DOCK 5, περιλαμβάνει επίσης όλες τις προηγούμενες λειτουργίες.

Co-funded by the Erasmus+ Programme of the European Union

GOLD

Το GOLD είναι το επικυρωμένο, διαμορφώσιμο λογισμικό σύνδεσης πρωτεΐνης-συνδέτη για την ανακάλυψη φαρμάκων από ειδικούς. Για εικονικό έλεγχο έως και βελτιστοποίηση μολύβδου.

Προσαρμόστε με περιορισμούς για να καθοδηγήσετε τα αποτελέσματα προς γνωστά χαρακτηριστικά ή συμπεριφορές και να αξιολογήσετε τον αντίκτυπο των μορίων του νερού στη βάση σύνδεσης. Ή χρησιμοποιήστε τον οδηγό για γρήγορα αποτελέσματα σύνδεσης πρωτεΐνης-συνδέτη.

Οι εκτενώς επικυρωμένες λειτουργίες βαθμολόγησης στο GOLD εμπιστεύονται οι επιστήμονες παγκοσμίως για την πρόβλεψη πόζας και τον εικονικό έλεγχο – με τα αποτελέσματα να φαίνονται σε όλη τη βιβλιογραφία.

Co-funded by the Erasmus+ Programme of the European Union

Glide

Το Glide προσφέρει το πλήρες εύρος επιλογών ταχύτητας έναντι ακρίβειας, από τη λειτουργία HTVS (εικονική διαλογή υψηλής απόδοσης) για αποτελεσματικό εμπλουτισμό εκατομμυρίων βιβλιοθηκών σύνθετων συστατικών, έως τη λειτουργία SP (τυπική ακρίβεια) για αξιόπιστη σύνδεση δεκάδων έως εκατοντάδων χιλιάδων υποκαταστατών με υψηλή ακρίβεια, στη λειτουργία XP (έξτρα ακρίβειας) όπου η περαιτέρω εξάλειψη των ψευδώς θετικών επιτυγχάνεται με εκτενέστερη δειγματοληψία και προηγμένη βαθμολόγηση, με αποτέλεσμα ακόμη υψηλότερο εμπλουτισμό.

Co-funded by the Erasmus+ Programme of the European Union

SeeSAR

Το SeeSAR είναι μια τρισδιάστατη πλατφόρμα μοντελοποίησης επιτραπέζιων υπολογιστών. Διευκολύνει τη διαδραστική, οπτική, σύνθετη ιεράρχηση - καθώς και την εξέλιξη των σύνθετων. Οι εργασίες σχεδίασης που βασίζονται στη δομή υποστηρίζουν ιδανικά μια βελτιστοποίηση πολλαπλών παραμέτρων για τη μεγιστοποίηση της πιθανότητας επιτυχίας και όχι μόνο τη συγγένεια.

Co-funded by the Erasmus+ Programme of the European Union

Molegro Virtual Docker

Το Molegro Virtual Docker είναι μια ολοκληρωμένη πλατφόρμα για την πρόβλεψη αλληλεπιδράσεων πρωτεΐνης - συνδέτη. Το Molegro Virtual Docker χειρίζεται όλες τις πτυχές της διαδικασίας σύνδεσης, από την προετοιμασία των μορίων έως τον προσδιορισμό των πιθανών θέσεων δέσμευσης της πρωτεΐνης στόχου και την πρόβλεψη των τρόπων σύνδεσης των προσδεμάτων.

Το Molegro Virtual Docker παρέχει

- **Υψηλή ακρίβεια σύνδεσης: ο κινητήρας σύνδεσης έχει αποδειχθεί ότι εντοπίζει σωστά τους τρόπους σύνδεσης με υψηλή ακρίβεια. Το Molegro Virtual Docker έχει αποδειχθεί ότι ξεπερνά τα άλλα προγράμματα σύνδεσης όσον αφορά την αναγνώριση των σωστών τρόπων σύνδεσης.**
- **Εύκολο στη χρήση διεπαφή: οι ενσωματωμένοι οδηγοί επιτρέπουν στο χρήστη να ρυθμίζει και να εκτελεί εύκολα εργασίες σύνδεσης. Παρέχονται προηγμένα εργαλεία οπτικοποίησης και ανάλυσης για την εξέταση των αλληλεπιδράσεων συνδέτη-υποδοχέα και τη λεπτομερή ρύθμιση των λύσεων σύνδεσης που βρέθηκαν.**
- **Cross-platform: υποστηρίζεται σε Linux, Windows και Mac, επιτρέποντας εύκολη διαλειτουργικότητα μεταξύ πλατφορμών.**

Co-funded by the Erasmus+ Programme of the European Union

MPI-Vina

Το MPI-Vina είναι ένας παραλληλισμός ανοιχτού κώδικα του AutoDock Vina, ο οποίος μειώνει μαζικά τον χρόνο εικονικής προβολής χρησιμοποιώντας συμπλέγματα υπολογιστών ή δίκτυο υπολογιστών. Αναπτύχθηκε με βάση το MPI και προορίζεται για περιβάλλον κατανεμημένης μνήμης. Ο στόχος του MPI-Vina είναι να μειώσει το συνολικό χρόνο του συνόλου δεδομένων προσδέματος διαλογής. Το AutoDock Vina είναι το κύριο πρόγραμμα σύνδεσης του MPI-Vina.

Στο MPI-Vina, η σύνδεση πρωτεΐνης-συνδέτη κατανέμεται σε διαφορετικούς κόμβους ενός υπολογιστικού συμπλέγματος όπου κάθε κόμβος εκτελεί σύνδεση ενός μεμονωμένου συνδετήρα έναντι ενός υποδοχέα στόχου. Η κατανομή του συνόλου υποκαταστατών επιτρέπει τη σύνδεση πολλαπλών προσδεμάτων σε πολλούς κόμβους ταυτόχρονα.

Το MPI-Vina μειώνει δραματικά τον συνολικό χρόνο της εικονικής προβολής σε σχέση με την παραδοσιακή προσέγγιση εικονικής προβολής χρησιμοποιώντας παράλληλη επεξεργασία τελευταίας τεχνολογίας. Μειώνει επίσης το μέγεθος και την πολυπλοκότητα του προβλήματος του προσυμπτωματικού ελέγχου και εστιάζει στην ανακάλυψη φαρμάκων και τις προσπάθειες βελτιστοποίησης για τους πιο υποσχόμενους πελάτες.

Co-funded by the Erasmus+ Programme of the European Union

OEDocking

Το OEDocking είναι μια σουίτα καλά επικυρωμένων εργαλείων μοριακής σύνδεσης και ροών εργασίας, το καθένα ειδικά σχεδιασμένο για να αντιμετωπίζει τη δική του μοναδική πτυχή της αλληλεπίδρασης πρωτεΐνης-συνδέτη. Συγκεκριμένα, διαθέτει POSIT για ενημερωμένη πρόβλεψη πόζας καθώς και FRED και HYBRID ως συμπληρωματικά εργαλεία για εικονικό έλεγχο.

Co-funded by the Erasmus+ Programme of the European Union

rDock

Το rDock είναι ένα γρήγορο και ευέλικτο πρόγραμμα σύνδεσης ανοιχτού κώδικα που μπορεί να χρησιμοποιηθεί για τη σύνδεση μικρών μορίων έναντι πρωτεϊνών και νουκλεϊκών οξέων. Έχει σχεδιαστεί για καμπάνιες εικονικής προβολής υψηλής απόδοσης (HTVS) και μελέτες πρόβλεψης Binding Mode.

Το rDock είναι κυρίως γραμμένο σε C++ και τα βοηθητικά σενάρια και τα προγράμματα είναι γραμμένα σε γλώσσες C++, perl ή python.

Το πλήρες πακέτο λογισμικού rDock απαιτεί λιγότερο από 50 MB χώρου στον σκληρό δίσκο και είναι μεταγλωττίσιμο (μόνο αυτή τη στιγμή) σε όλους τους υπολογιστές Linux.

Co-funded by the Erasmus+ Programme of the European Union

Lead Finder

Το λογισμικό Lead Finder είναι μια εφαρμογή υπολογιστικής χημείας για τη μοντελοποίηση της αλληλεπίδρασης πρωτεΐνης-συνδέτη. Το Lead Finder μπορεί να χρησιμοποιηθεί σε μελέτες μοριακής σύνδεσης και για την ποσοτική αξιολόγηση της δέσμευσης συνδέτη και της βιολογικής δραστηριότητας.

Το Lead Finder εισάγει τρεις λειτουργίες βαθμολόγησης βελτιστοποιημένες για την ακριβή πρόβλεψη τρισδιάστατων θέσεων προσδέματος, ενέργειας δέσμευσης πρωτεΐνης-συνδέτη και ενεργών και ανενεργών ενώσεων με σειρά κατάταξης σε πειράματα εικονικής διαλογής. Το Lead Finder έχει σχεδιαστεί για να ικανοποιεί τις ανάγκες υπολογιστικών και φαρμακευτικών χημικών που συμμετέχουν στη διαδικασία ανακάλυψης φαρμάκων, φαρμακολόγων και τοξικολόγων που εμπλέκονται στη μοντελοποίηση και αξιολόγηση των ιδιοτήτων ADMET στο silico και βιοχημικών και ενζυμολόγων που εργάζονται για την εξειδίκευση ενζύμου και τον ορθολογικό σχεδιασμό ενζύμων.

Co-funded by the Erasmus+ Programme of the European Union

ADAM

Το ADAM είναι ένα αυτοματοποιημένο εργαλείο σύνδεσης, το οποίο μπορεί να προβλέψει τη σταθερή λειτουργία δέσμευσης του εύκαμπτου μορίου συνδέτη με το μακρομόριο στόχο. Είναι σε θέση να κατασκευάζει ενεργειακά ευνοϊκά μοντέλα σύνδεσης σε υψηλές ταχύτητες, εξερευνώντας τον διαμορφωτικό χώρο του εύκαμπτου συνδετήρα συνεχώς και εξαντλητικά. Η ακρίβεια σύνδεσης και η αξιοπιστία στον πραγματικό σχεδιασμό του φαρμάκου είναι το κορυφαίο επίπεδο στον κόσμο (ακόμη και σε περιπτώσεις υποτιθέμενης κίνησης επαγόμενης προσαρμογής της πρωτεΐνης στόχου). Επιπλέον, ο εικονικός έλεγχος υψηλής απόδοσης μπορεί να πραγματοποιηθεί εύκολα με τη χρήση του ADAM. Σε πολλά εσωτερικά έργα και χρηματοδοτούμενες έρευνες για την ανάπτυξη φαρμάκων, ο εικονικός έλεγχος ADAM έχει ήδη σημειώσει επιτυχίες. Το ADAM θα πρέπει να συμβάλει σε μεγάλο βαθμό σε μια ποικιλία ερευνητικών τομέων, συμπεριλαμβανομένης της μελέτης QSAR και της ανάλυσης της αλληλεπίδρασης πρωτεΐνης-συνδέτη, καθώς και στον σχεδιασμό και την ανακάλυψη φαρμάκων.

Co-funded by the Erasmus+ Programme of the European Union

PLANTS

Στη διαδικασία ανάπτυξης φαρμάκου, η πρόβλεψη της πολύπλοκης δομής ενός μικρού συνδετήρα με μια πρωτεΐνη, το λεγόμενο πρόβλημα σύνδεσης πρωτεΐνης-συνδέτη, χρησιμοποιείται εκτενώς στην εικονική διαλογή μεγάλων βάσεων δεδομένων και στη βελτιστοποίηση μολύβδου. Δεδομένης μιας δομής πρωτεΐνης, μιας δομής συνδέτη και μιας συνάρτησης βαθμολόγησης, ο στόχος είναι να βρεθεί μια διαμόρφωση συνδέτη χαμηλής ενέργειας στη θέση δέσμευσης της πρωτεΐνης που αντιστοιχεί στο συνολικό ελάχιστο της συνάρτησης βαθμολόγησης. Ο αλγόριθμος σύνδεσης PLANTS βασίζεται σε μια κατηγορία αλγορίθμων στοχαστικής βελτιστοποίησης που ονομάζεται βελτιστοποίηση αποικίας μυρμηγκιών (ACO). Το ACO εμπνέεται από τη συμπεριφορά των πραγματικών μυρμηγκιών που βρίσκουν το συντομότερο μονοπάτι ανάμεσα στη φωλιά τους και μια πηγή τροφής. Τα μυρμήγκια χρησιμοποιούν έμμεση επικοινωνία με τη μορφή μονοπατιών φερομόνης που σηματοδοτούν διαδρομές μεταξύ της φωλιάς και μιας πηγής τροφής. Στην περίπτωση σύνδεσης πρωτεΐνης-συνδέτη, χρησιμοποιείται μια τεχνητή αποικία μυρμηγκιών για να βρεθεί μια ελάχιστη ενεργειακή διαμόρφωση του συνδέτη στη θέση δέσμευσης. Αυτά τα μυρμήγκια χρησιμοποιούνται για να μιμηθούν τη συμπεριφορά πραγματικών μυρμηγκιών και να επισημάνουν τις διαμορφώσεις συνδέτη χαμηλής ενέργειας με ίχνη φερομόνης. Οι πληροφορίες ίχνης τεχνητής φερομόνης τροποποιούνται σε επόμενες επαναλήψεις για να δημιουργήσουν διαμορφώσεις χαμηλής ενέργειας με μεγαλύτερη πιθανότητα.

Co-funded by the Erasmus+ Programme of the European Union

SwissDock

Το SwissDock βασίζεται στο λογισμικό σύνδεσης EADock DSS, του οποίου ο αλγόριθμος αποτελείται από τα ακόλουθα βήματα:

- **πολλοί τρόποι σύνδεσης δημιουργούνται είτε σε ένα κουτί (τοπική σύνδεση) είτε κοντά σε όλες τις κοιλότητες στόχου (τυφλή σύνδεση).**
- **Ταυτόχρονα, οι ενέργειές τους CHARMM υπολογίζονται σε ένα πλέγμα.**
- **Οι τρόποι σύνδεσης με τις πιο ευνοϊκές ενέργειες αξιολογούνται με FACTS και ομαδοποιούνται.**
- **Τα πιο ευνοϊκά συμπλέγματα μπορούν να οπτικοποιηθούν στο διαδίκτυο και να τα κατεβάσετε στον υπολογιστή σας.**

Co-funded by the Erasmus+ Programme of the European Union

PatchDock

Το PatchDock είναι ένας αλγόριθμος για μοριακή σύνδεση. Η είσοδος είναι δύο μόρια οποιουδήποτε τύπου: πρωτεΐνες, DNA, πεπτίδια, φάρμακα. Η έξοδος είναι μια λίστα πιθανών συμπλεγμάτων ταξινομημένων με κριτήρια συμπληρωματικότητας σχήματος.

PatchDock είναι εμπνευσμένος από τεχνικές αναγνώρισης αντικειμένων και τμηματοποίησης εικόνας που χρησιμοποιούνται στο Computer Vision. Η σύνδεση μπορεί να συγκριθεί με τη συναρμολόγηση ενός παζλ. Όταν λύνουμε το παζλ προσπαθούμε να ταιριάξουμε δύο κομμάτια επιλέγοντας ένα κομμάτι και αναζητώντας το συμπληρωματικό. Επικεντρωνόμαστε στα μοτίβα που είναι μοναδικά για το στοιχείο του παζλ και αναζητούμε τα ταιριαστά μοτίβα στα υπόλοιπα κομμάτια. Το PatchDock χρησιμοποιεί παρόμοια τεχνική. Λαμβάνοντας υπόψη δύο μόρια, οι επιφάνειές τους χωρίζονται σε μπαλώματα ανάλογα με το σχήμα της επιφάνειας. Αυτά τα μπαλώματα αντιστοιχούν σε μοτίβα που διακρίνουν οπτικά τα κομμάτια του παζλ. Μόλις εντοπιστούν τα μπαλώματα, μπορούν να τοποθετηθούν σε υπέρθεση χρησιμοποιώντας αλγόριθμους αντιστοίχισης σχήματος. Ο αλγόριθμος έχει τρία κύρια στάδια:

Co-funded by the Erasmus+ Programme of the European Union

DockingServer

Ο DockingServer προσφέρει μια διαδικτυακή, εύχρηστη διεπαφή που χειρίζεται όλες τις πτυχές της μοριακής σύνδεσης από τη ρύθμιση συνδέτη και πρωτεΐνης.

Ενώ η φιλική προς τον χρήστη διεπαφή επιτρέπει τον υπολογισμό σύνδεσης και την αξιολόγηση των αποτελεσμάτων από ερευνητές που προέρχονται από όλα τα πεδία της βιοχημείας, ο DockingServer παρέχει επίσης πλήρη έλεγχο στη ρύθμιση συγκεκριμένων παραμέτρων ρύθμισης συνδέτη και πρωτεΐνης και υπολογισμούς σύνδεσης για πιο προχωρημένους χρήστες.

Η εφαρμογή μπορεί να χρησιμοποιηθεί για σύνδεση και ανάλυση μεμονωμένων προσδεμάτων καθώς και για σύνδεση βιβλιοθηκών συνδετών υψηλής απόδοσης σε πρωτεΐνες στόχευσης.

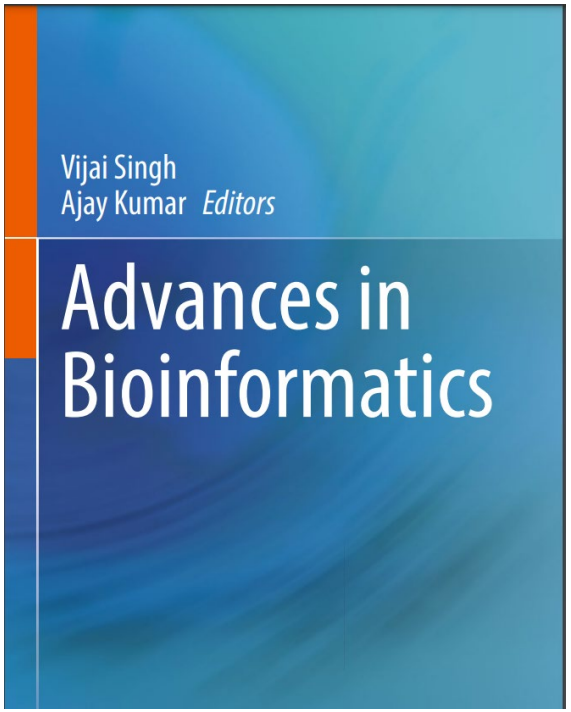
Ο DockingServer ενσωματώνει έναν αριθμό λογισμικού υπολογιστικής χημείας που στοχεύει ειδικά στον σωστό υπολογισμό των παραμέτρων που απαιτούνται σε διαφορετικά στάδια της διαδικασίας σύνδεσης, π.χ. βελτιστοποίηση ακριβούς γεωμετρίας συνδετήρα, ελαχιστοποίηση ενέργειας, υπολογισμό φορτίου, υπολογισμό σύνδεσης και αναπαράσταση συμπλόκου πρωτεΐνης-συνδέτη. Έτσι, η χρήση του DockingServer επιτρέπει στο χρήστη να πραγματοποιεί εξαιρετικά αποτελεσματικούς και ισχυρούς υπολογισμούς σύνδεσης ενσωματώνοντας έναν αριθμό δημοφιλών λογισμικού που χρησιμοποιούνται στη χημεία του πυριτίου σε μια ολοκληρωμένη υπηρεσία Ιστού.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Δώστε παραδείγματα προγραμμάτων σύνδεσης.

Co-funded by the Erasmus+ Programme of the European Union



Vijai Singh, Ajay Kumar (2021) Προόδους στη Βιοπληροφορική Springer Nature Singapore Pte Ltd

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

Θέμα 2. Πρόβλημα σύνδεσης

Μάθημα 3. Πώς λειτουργεί ένα πρόγραμμα σύνδεσης

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Αλγόριθμοι

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Η ικανότητα παραγωγής ενός μεγάλου και διαφορετικού συνόλου θέσεων συνδέτη (δηλ. υποψήφιες συντεταγμένες για διαμορφώσεις, θέσεις και προσανατολισμούς του συνδέτη εντός του συμπλέγματος πρωτεΐνης-συνδέτη) αποτελεί προϋπόθεση για να είναι χρήσιμο ένα εργαλείο σύνδεσης.

Αλγόριθμοι

Υπάρχουν δύο κύριοι τύποι αλγορίθμων που επιτρέπουν στα προγράμματα σύνδεσης να αναζητήσουν τον διαμορφωτικό χώρο του συνδέτη προκειμένου να βρουν τις θέσεις του:

(a) συστηματικές ή κατευθυνόμενες προσεγγίσεις·

(β) τυχαίες ή στοχαστικές μέθοδοι.

Μέθοδοι προσομοίωσης όπως η μοριακή δυναμική και η κβαντομηχανική είναι επίσης σε θέση να το κάνουν αυτό, αλλά, προς το παρόν, είναι πολύ δαπανηρές υπολογιστικά για να εφαρμοστούν κατά τη διάρκεια του VLS.

Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμοι αναζήτησης

Υπάρχουν τρεις υποτύποι συστηματικών ή κατευθυνόμενων αλγορίθμων αναζήτησης:

(α) μέθοδοι διαμορφωτικής αναζήτησης·

β) μέθοδοι κατακερματισμού ή σταδιακής κατασκευής· και

(γ) μέθοδοι βάσης δεδομένων.

Όλοι έχουν κοινό ότι οι αλγόριθμοι προσπαθούν να εξερευνήσουν όλους τους βαθμούς ελευθερίας του συνδέτη. ο τρόπος όμως που πραγματοποιούν την έρευνα είναι διαφορετικός. Από αυτή την άποψη, οι αλγόριθμοι διαμορφωτικής αναζήτησης προσπαθούν να λάβουν όλες τις πιθανές διαμορφώσεις συνδέτη υποβάλλοντας όλους τους δεσμούς συνδέτη που μπορούν να περιστραφούν σε στροφή 360º χρησιμοποιώντας μια σταθερή αύξηση.

Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμοι αναζήτησης

Ένα από τα κύρια προβλήματα αυτής της μεθοδολογίας είναι ότι ο αριθμός των διαμορφώσεων συνδέτη που μπορούν να δημιουργηθούν αυξάνεται εκθετικά με τον αριθμό των περιστρεφόμενων δεσμών και, επομένως, η εφαρμογή του στην καθαρότερη μορφή του είναι πολύ περιορισμένη (δηλ. συνήθως αρκετοί περιορισμοί και περιορισμοί στο πρόσδεμα απαιτούνται για τη μείωση του προβλήματος της συνδυαστικής έκρηξης).

Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμοι αναζήτησης

Από την άλλη πλευρά, μέθοδοι κατακερματισμού ή σταδιακής κατασκευής χρησιμοποιούνται επί του παρόντος από προγράμματα σύνδεσης όπως τα eHiTS ®, LUDITM (http://www.accelrys.com/products/insight/sbd_modules.html), FlexXTM (http://www.biosolveit.de/FlexX/), DOCK (http://dock.compbio.ucsf.edu/) και Hammerhead για αναζήτηση διαθέσιμων διαμορφώσεων συνδέτη.

Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμοι αναζήτησης

Σε αυτές τις μεθόδους, οι συνδέτες αναπτύσσονται σταδιακά στη θέση δέσμευσης είτε διαιρώντας το πρόσδεμα σε πολλά άκαμπτα θραύσματα, αγκυροβολώντας τα και τελικά προσπαθώντας να ξαναχτίσουν τη δομή του συνδέτη ενώνοντας τα άκαμπτα θραύσματα με τα εύκαμπτα τμήματα που τα ενώνουν (αυτό είναι που ονομάζεται προσέγγιση τοποθέτησης και σύνδεσης) ή με διαίρεση του συνδέτη σε ένα άκαμπτο θραύσμα πυρήνα που πρώτα στερεώνεται και στη συνέχεια προστίθενται διαδοχικά τα υπόλοιπα τμήματα συνδέτη (αυτή είναι η λεγόμενη σταδιακή προσέγγιση).

Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμοι αναζήτησης

Ο τελευταίος υποτύπος αλγορίθμων συστηματικής αναζήτησης είναι οι μέθοδοι βάσης δεδομένων που χρησιμοποιούν βιβλιοθήκες προδημιουργημένων διαμορφώσεων (τα λεγόμενα διαμορφωτικά σύνολα) που στη συνέχεια υπόκεινται σε ένα άκαμπτο σώμα σύνδεσης. Ένα παράδειγμα προγράμματος σύνδεσης που χρησιμοποιεί αυτή τη μεθοδολογία βάσης δεδομένων για να αντιμετωπίσει το ζήτημα της ευελιξίας του συνδετήρα είναι το FLOG.

Αλγόριθμοι τυχαίας ή στοχαστικής αναζήτησης

Υπάρχουν τρεις υποτύποι τυχαίων ή στοχαστικών αλγορίθμων αναζήτησης:

(a) Μέθοδοι Monte Carlo (MC).

(b) Μέθοδοι γενετικού αλγόριθμου (GA). και

(c) tabu (TS).

Όλοι προσπαθούν να δειγματίσουν τον διαμορφωτικό χώρο εκτελώντας μια τυχαία διαμορφωτική αλλαγή στο πρόσδεμα που ακολουθείται από αποδοχή ή απόρριψη του προκύπτοντος διαμορφωτή χρησιμοποιώντας μια προκαθορισμένη συνάρτηση πιθανότητας.

Αλγόριθμοι τυχαίας ή στοχαστικής αναζήτησης

Εάν η διαμόρφωση του συνδέτη γίνει τελικά αποδεκτή, χρησιμοποιείται ως το σημείο εκκίνησης για μια νέα τυχαία διαμορφωτική αλλαγή. Ο κύριος περιορισμός τέτοιων μεθόδων είναι ότι η στάση που ταιριάζει με την πειραματική διαμόρφωση του συνδέτη στο σύμπλεγμα μπορεί να μην επιτευχθεί και, επομένως, δεν θα αξιολογηθεί από την αντίστοιχη συνάρτηση βαθμολόγησης.

Co-funded by the Erasmus+ Programme of the European Union

Μέθοδοι MC

Στις μεθόδους MC, ο συνδέτης τοποθετείται τυχαία στη θέση δέσμευσης του υποδοχέα, βαθμολογείται και δημιουργείται μια νέα διαμόρφωση από τυχαίες αλλαγές που εφαρμόζονται σε:

(α) οι περιστρεφόμενοι δεσμοί του συνδετήρα· και

(β) τη θέση του συνδέτη (δηλαδή ο συνδετήρας περιστρέφεται τυχαία και μεταφράζεται).

Μετά από κάθε αλλαγή, ο συνδέτης τυπικά ελαχιστοποιείται και βαθμολογείται. Στη συνέχεια, εάν η νέα λύση έχει καλύτερη βαθμολογία από την προηγούμενη, γίνεται αμέσως αποδεκτή. Από την άλλη πλευρά, εάν η τελευταία διαμόρφωση δεν είναι ένα νέο ελάχιστο, εφαρμόζεται μια συνάρτηση πιθανότητας που βασίζεται σε Boltzmann. Εάν η στάση επιτύχει σε αυτόν τον έλεγχο συνάρτησης πιθανότητας, γίνεται αποδεκτή. αν όχι, απορρίπτεται (αυτό λέγεται κριτήριο Μητρόπολης).

Co-funded by the Erasmus+ Programme of the European Union

Μέθοδοι MC

Προγράμματα σύνδεσης που μπορούν να αντιμετωπίσουν διαμορφωτικές αναζητήσεις συνδετήρων χρησιμοποιώντας αλγόριθμους βασισμένους σε MC είναι το AutoDock (http://autodock.scripps.edu/) (στην πραγματικότητα, ήταν το πρώτο πρόγραμμα σύνδεσης στο οποίο εφαρμόστηκε ένας αλγόριθμος ανόπτησης προσομοίωσης MC), PRODOCK , ICMTM (http://www.molsoft.com/docking.html), MCDOCK, DockVisionTM (http://www.dockvision.com/) και QXP.

Μέθοδοι γενετικού αλγόριθμου (GA).

Τα GA χρησιμοποιούν έννοιες που προέρχονται από τη γενετική και τη θεωρία της βιολογικής εξέλιξης για να εξερευνήσουν τον διαμορφωτικό χώρο των προσδεμάτων. Σε αντίθεση με τις μεθόδους MC, τα GA ξεκινούν από έναν αρχικό πληθυσμό διαφορετικών διαμορφώσεων του συνδέτη και το καθένα ορίζεται από ένα σύνολο μεταβλητών κατάστασης ή γονιδίων που περιγράφουν τη μετάφραση, τον προσανατολισμό και τη διαμόρφωση του συνδέτη σε σχέση με τον υποδοχέα. Το πλήρες σύνολο αυτών των γονιδίων συνδέτη είναι ο γονότυπος, ενώ οι προκύπτουσες ατομικές συντεταγμένες είναι ο φαινότυπος.

Co-funded by the Erasmus+ Programme of the European Union

Μέθοδοι γενετικού αλγόριθμου (GA).

Προγράμματα σύνδεσης που μπορούν να χρησιμοποιήσουν αλγόριθμους GA ή τύπου GA για την αντιμετώπιση διαμορφωτικών αναζητήσεων προσδέματος είναι GOLD (http://www.ccdc.cam.ac.uk/products/life_sciences/gold/), AutoDock v3.0 και v4.0 , DIVALI, MVDTM και DARWIN. Ο τελευταίος υποτύπος αλγορίθμων τυχαίας αναζήτησης είναι οι μέθοδοι TS που λειτουργούν επιβάλλοντας περιορισμούς που αποτρέπουν την επίσκεψη σε ήδη εξερευνημένες περιοχές του διαμορφωτικού χώρου συνδέτη και, ως εκ τούτου, ευνοούν την ανάλυση νέων διαμορφώσεων. Για να γίνει αυτό, όταν μια νέα διαμόρφωση προσδέματος είναι διαθέσιμη, υπολογίζεται η μέση τετραγωνική απόκλιση ρίζας του (δηλ. RMSD) σε σχέση με τις διαμορφώσεις που επισκεφτήκατε προηγουμένως.

Co-funded by the Erasmus+ Programme of the European Union

Μέθοδοι γενετικού αλγόριθμου (GA).

Όταν ολοκληρωθεί αυτός ο υπολογισμός, το χαμηλότερο RMSD συγκρίνεται με μια ορισμένη τιμή κατωφλίου και, εάν είναι υψηλότερη, τότε η αναλυόμενη διαμόρφωση για το πρόσδεμα γίνεται αποδεκτή και οι συντεταγμένες του αποθηκεύονται και χρησιμοποιούνται για την αποδοχή ή την απόρριψη νέων διαμορφώσεων. Τα PRO_LEADS και MVDTM είναι τα πιο δημοφιλή

προγράμματα σύνδεσης που μπορούν να χρησιμοποιήσουν έναν αλγόριθμο TS.

Co-funded by the Erasmus+ Programme of the European Union

tabu (TS).

Tabu (TS) είναι μια μεταευρετική μέθοδος αναζήτησης που χρησιμοποιεί μεθόδους τοπικής αναζήτησης που χρησιμοποιούνται για μαθηματική βελτιστοποίηση. Δημιουργήθηκε από τον Fred W. Glover το 1986 και επισημοποιήθηκε το 1989.

Οι τοπικές (γειτονικές) αναζητήσεις λαμβάνουν μια πιθανή λύση σε ένα πρόβλημα και ελέγχουν τους άμεσους γείτονές του (δηλαδή, λύσεις που είναι παρόμοιες εκτός από πολύ λίγες μικρές λεπτομέρειες) με την ελπίδα να βρουν μια βελτιωμένη λύση. Οι τοπικές μέθοδοι αναζήτησης έχουν την τάση να κολλάνε σε μη βέλτιστες περιοχές ή σε οροπέδια όπου πολλές λύσεις είναι εξίσου κατάλληλες.

tabu (TS).

Tabu βελτιώνει την απόδοση της τοπικής αναζήτησης χαλαρώνοντας τον βασικό της κανόνα. Πρώτον, σε κάθε βήμα μπορούν να γίνουν αποδεκτές κινήσεις επιδείνωσης εάν δεν υπάρχει διαθέσιμη βελτιωτική κίνηση (όπως όταν η αναζήτηση έχει κολλήσει σε ένα αυστηρό τοπικό ελάχιστο). Επιπλέον, εισάγονται απαγορεύσεις (εφεξής ο όρος tabu) για να αποθαρρύνουν την αναζήτηση να επιστρέψει σε λύσεις που είχαν επισκεφτεί προηγουμένως.

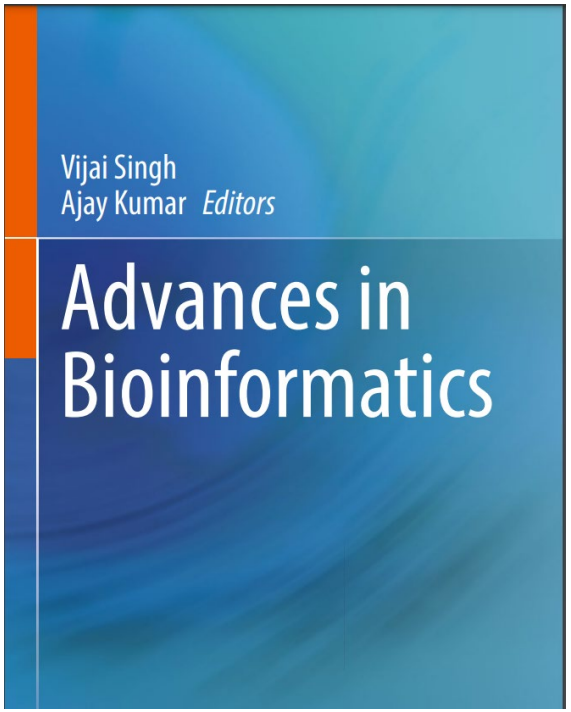
Η υλοποίηση της αναζήτησης tabu χρησιμοποιεί δομές μνήμης που περιγράφουν τις επισκέψεις λύσεις ή σύνολα κανόνων που παρέχονται από τον χρήστη.[2] Εάν μια πιθανή λύση έχει προηγουμένως επισκεφτεί εντός μιας ορισμένης βραχυπρόθεσμης περιόδου ή εάν έχει παραβιάσει έναν κανόνα, επισημαίνεται ως " tabu " (απαγορευμένη) έτσι ώστε ο αλγόριθμος να μην εξετάζει αυτή τη δυνατότητα επανειλημμένα.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε τον τρόπο λειτουργίας του προγράμματος σύνδεσης;
- Ποιοι είναι οι πιο δημοφιλείς αλγόριθμοι για σύνδεση;

Co-funded by the Erasmus+ Programme of the European Union



Vijai Singh, Ajay Kumar (2021) Προόδους στη Βιοπληροφορική Springer Nature Singapore Pte Ltd

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

Θέμα 2. Πρόβλημα σύνδεσης

Μάθημα 4. Οι συναρτήσεις βαθμολόγησης

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Ισορροπία μεταξύ ταχύτητας και ακρίβειας
- Επιβολή συναρτήσεων βαθμολόγησης βάσει πεδίου
- Εμπειρικές συναρτήσεις βαθμολόγησης
- Λειτουργίες βαθμολόγησης βασισμένες στη γνώση

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Μόλις προβλεφθούν οι υποψήφιες διαμορφώσεις για τον συνδέτη στο σύμπλοκο, η συγγένεια δέσμευσής τους για τον υποδοχέα πρέπει να βαθμολογηθεί. Αυτό γίνεται μέσω μιας συνάρτησης βαθμολόγησης που αξιολογεί τα αποτελέσματα αναζήτησης και στη συνέχεια δίνει, ιδανικά, την υψηλότερη βαθμολογία στη σωστή πόζα. Στην πραγματικότητα, εάν ο αλγόριθμος αναζήτησης μπορεί να βρει τη σωστή στάση αλλά η συνάρτηση βαθμολόγησης δεν μπορεί να την αναγνωρίσει, το πρόγραμμα θα κάνει μια άκυρη και άχρηστη πρόταση στον επιστήμονα. Επομένως, ο ρόλος της συνάρτησης βαθμολόγησης είναι κρίσιμος σε κάθε πρωτόκολλο σύνδεσης.

Ισορροπία μεταξύ ταχύτητας και ακρίβειας

Η έλλειψη κατάλληλης συνάρτησης βαθμολόγησης, τόσο από άποψη ταχύτητας όσο και από άποψη ακρίβειας, είναι το σημαντικότερο εμπόδιο στο docking. Μερικές πρόσφατες μελέτες έχουν δείξει ότι η απόδοση της συνάρτησης βαθμολόγησης σε μια συγκεκριμένη πρωτεΐνη-στόχο εξαρτάται σε μεγάλο βαθμό από την περίπτωση. Επομένως, οποιαδήποτε μελέτη σύνδεσης θα πρέπει να ξεκινά είτε με μια αντικειμενική αξιολόγηση των διαθέσιμων λειτουργιών βαθμολόγησης στην πρωτεΐνη-στόχο έτσι ώστε να μπορεί να επιλεγεί η καταλληλότερη είτε με μια συνάρτηση βαθμολόγησης που έχει ρυθμιστεί σύμφωνα με τα χαρακτηριστικά της θέσης δέσμευσης.

Co-funded by the Erasmus+ Programme of the European Union

Επιβολή συναρτήσεων βαθμολόγησης βάσει πεδίου

Οι συναρτήσεις βαθμολόγησης που βασίζονται στο πεδίο δύναμης είναι παρόμοιες με τις εμπειρικές συναρτήσεις επειδή και οι δύο προβλέπουν την ελεύθερη ενέργεια δέσμευσης ενός συμπλόκου πρωτεΐνης-συνδέτη προσθέτοντας μεμονωμένες συνεισφορές από διαφορετικούς τύπους αλληλεπιδράσεων. Ωστόσο, οι όροι αλληλεπίδρασης του πρώτου προέρχονται από τη θεωρητική φυσική που διέπει τα πεδία δύναμης της μοριακής μηχανικής σε αντίθεση με τις πειραματικές συγγένειες που χρησιμοποιούνται για την εξαγωγή της δεύτερης. Έτσι, οι συναρτήσεις βαθμολόγησης δυναμικού πεδίου χρησιμοποιούν συναρτήσεις ενέργειας από την κλασική μοριακή μηχανική και, γενικά, ποσοτικοποιούν το άθροισμα δύο ενεργειών:

(a) την ενέργεια αλληλεπίδρασης μεταξύ του υποδοχέα και του συνδέτη. και

(b) την εσωτερική ενέργεια του συνδέτη.

Co-funded by the Erasmus+ Programme of the European Union

Η δεσμευτική ελεύθερη ενέργεια

Η ελεύθερη ενέργεια δέσμευσης της διαμοριακής αλληλεπίδρασης συχνά προσεγγίζεται ως το αποτέλεσμα του αθροίσματος ενός όρου ενέργειας van der Waals (που περιγράφεται μέσω της συνάρτησης δυναμικού Lennard-Jones) και ενός όρου ηλεκτροστατικής ενέργειας (που περιγράφεται μέσω μιας κουλομβικής διατύπωσης με μια διηλεκτρική συνάρτηση που εξαρτάται από την απόσταση που μειώνει τη συμβολή από τις αλληλεπιδράσεις φορτίου-φόρτισης). Επιπλέον, το AutoDock v3.0 και v4.0 και το G-Score εξετάζουν επίσης έναν όρο δεσμού υδρογόνου (αν και με διαφορετικές λειτουργικές μορφές) σε μια προσπάθεια να αυξηθεί η δυνατότητα συγκεκριμένης μοριακής αναγνώρισης.

Η εσωτερική ενέργεια του συνδέτη

Οι συναρτήσεις που περιγράφουν την εσωτερική ενέργεια του συνδέτη είναι συνήθως πολύ παρόμοιες με την ενέργεια αλληλεπίδρασης πρωτεΐνης-προσδέματος επειδή περιέχουν επίσης van der Waals ή/και ηλεκτροστατικές συνεισφορές (αν και το AutoDock - και προαιρετικά το GoldScore - θεωρεί επίσης έναν όρο δεσμού υδρογόνου και το G- Η βαθμολογία περιλαμβάνει έναν όρο στρεπτικής εντροπίας).

Περιορισμοί

Οι αναγκαστικές συναρτήσεις βαθμολόγησης βάσει πεδίου έχουν τους ακόλουθους περιορισμούς:

(α) έχουν δυσκολίες να εξετάσουν τους όρους διαλυτότητας και εντροπίας και είτε τους αγνοούν είτε τους αντιμετωπίζουν με πρόχειρο τρόπο (επειδή αρχικά είχαν διαμορφωθεί για να μοντελοποιήσουν τις συνεισφορές της ενθαλπικής αέριας φάσης στη δομή και την ενέργεια).

(β) οι μη δεσμευμένες αλληλεπιδράσεις μπορούν να αντιμετωπιστούν μόνο με την εισαγωγή, με περισσότερο ή λιγότερο αυθαίρετο τρόπο, αποστάσεις αποκοπής που περιπλέκουν την ακριβή αντιμετώπιση των μακροπρόθεσμων επιπτώσεων που εμπλέκονται στη δέσμευση

(γ) οι πολικές αλληλεπιδράσεις υπερτονίζονται. και

(δ) ο υπολογισμός των ατομικών μερικών φορτίων βασίζεται σε γρήγορες αλλά ανακριβείς μεθόδους που βασίζονται σε δείκτες ηλεκταρνητικότητας αντί για μεθόδους κβαντομηχανικής όπως αυτές που χρησιμοποιούνται στην ανάπτυξη πεδίου δύναμης.

Οι πιο συχνά χρησιμοποιούμενες συναρτήσεις βαθμολόγησης πεδίου δύναμης

οι λειτουργίες βαθμολόγησης δυναμικού πεδίου που χρησιμοποιούνται πιο συχνά από προγράμματα σύνδεσης είναι D-Score, G-Score (με βάση το πεδίο δύναμης Tripos), GoldScore , αυτές που χρησιμοποιούνται από το AutoDock v3.0 και v4.0 [με βάση το AMBER (Υποβοηθούμενη κατασκευή μοντέλου και βελτίωση ενέργειας) πεδίο δύναμης αλλά τροποποιημένο με εμπειρικές παραμέτρους], η πλήρης συνάρτηση βαθμολόγησης μοριακής μηχανικής AMBER με σιωπηρό διαλύτη που χρησιμοποιείται από το DOCK v6.1 και αυτόν που χρησιμοποιείται από τη σύνδεση μεμονωμένου συνδετήρα στο DockVision . Προφανώς, πιο αυστηρά πεδία μοριακής μηχανικής δύναμης όπως AMBER, CHARMM (Χημεία στο Μακρομοριακή Μηχανική του HARvard), GROMOS (GROningen Το σύστημα μοριακής προσομοίωσης) και το OPLS (Βελτιστοποιημένα Δυναμικά για Προσομοιώσεις Υγρών) μπορούν επίσης να χρησιμοποιηθούν, αν και σε βάρος μιας τεράστιας αύξησης στο υπολογιστικό κόστος.

Co-funded by the Erasmus+ Programme of the European Union

Εμπειρικές συναρτήσεις βαθμολόγησης

Οι εμπειρικές συναρτήσεις βαθμολόγησης βασίζονται στην ιδέα ότι η ενέργεια δέσμευσης (δηλ. ΔG_{bind}) μπορεί να ληφθεί προσθέτοντας αρκετούς μεμονωμένους και μη συσχετισμένους όρους και στην πρωτοποριακή εργασία του Hans-Joachim Böhm στην Κεντρική Έρευνα BASF AG στο Ludwigshafen (Γερμανία). Πολλοί από τους όρους στις εμπειρικές συναρτήσεις βαθμολόγησης έχουν ισοδυναμίες στις συναρτήσεις βαθμολόγησης πεδίου δύναμης, αλλά οι πρώτοι είναι συνήθως απλούστεροι σε μορφή από τους δεύτερους. Οι εμπειρικές συναρτήσεις βαθμολόγησης περιέχουν όρους που αντιπροσωπεύουν την καθαρή συνεισφορά διαφορετικών τύπων μη δεσμευμένων αλληλεπιδράσεων (π.χ. δεσμοί υδρογόνου, ιοντικές και υδρόφοβες αλληλεπιδράσεις με τον υποδοχέα) στη συνολική ενέργεια δέσμευσης.

Co-funded by the Erasmus+ Programme of the European Union

Εμπειρικές συναρτήσεις βαθμολόγησης

Ορισμένες από αυτές τις συναρτήσεις βαθμολόγησης περιέχουν επίσης όρους που, σε κάποιο βαθμό, εξηγούν:

(a) μη ενθαλπικές συνεισφορές όπως ο αποκαλούμενος όρος ρότορα, ο οποίος προσεγγίζει τις ποινές εντροπίας κατά τη δέσμευση από ένα σταθμισμένο άθροισμα του αριθμού των περιστρεφόμενων δεσμών σε συνδετήρες. και

(b) επιλύσεις και επιδράσεις αποδιάλυσης .

Επιπλέον, οι όροι που αντιστοιχούν στη συνεισφορά ενός συγκεκριμένου τύπου μη δεσμευμένης αλληλεπίδρασης μπορεί να είναι συναρτήσεις (όπως ο επιπλέον πρόσθετος όρος για αρωματικές αλληλεπιδράσεις στο F-Score) ή να υλοποιούνται ανάλογα με τη συνάρτηση βαθμολόγησης.

Co-funded by the Erasmus+ Programme of the European Union

Εμπειρικές συναρτήσεις βαθμολόγησης

Η κύρια δύναμη των εμπειρικών συναρτήσεων είναι ότι οι όροι τους είναι συχνά απλοί στον υπολογισμό. Έτσι, πειραματικά δεδομένα όπως ενέργειες δέσμευσης ή ένα σύνολο συμπλεγμάτων υποδοχέα ακτίνων Χ-προσδέματος χρησιμοποιούνται ως είσοδος για μια ανάλυση παλινδρόμησης που ταιριάζει καλύτερα στη συνάρτηση βαθμολόγησης στα πειραματικά δεδομένα.

Co-funded by the Erasmus+ Programme of the European Union

Εμπειρικές συναρτήσεις βαθμολόγησης

Τα βασικά τους μειονεκτήματα όμως είναι ότι:

(a) εξαρτώνται σε μεγάλο βαθμό από τα πειραματικά δεδομένα που χρησιμοποιούνται στη διαδικασία παραμετροποίησης και, ως εκ τούτου, οι όροι από διαφορετικές συναρτήσεις βαθμολόγησης δεν μπορούν εύκολα να ανασυνδυαστούν σε μια νέα συνάρτηση βαθμολόγησης. και

(b) Δεν είναι σαφές εάν είναι σε θέση να προβλέψουν τη συγγένεια δέσμευσης των προσδεμάτων που είναι δομικά διαφορετικοί από αυτούς που χρησιμοποιούνται στο σετ εκπαίδευσης.

Co-funded by the Erasmus+ Programme of the European Union

Εμπειρικές συναρτήσεις βαθμολόγησης

Παραδείγματα εμπειρικών συναρτήσεων βαθμολόγησης είναι η συνάρτηση βαθμολόγησης του Böhm (η οποία συνοδεύεται από LUDI, F-Score, ChemScore , SCORE, Fresno, X-SCORE, PLP και Slide. Η συνάρτηση βαθμολόγησης που χρησιμοποιείται στο Hammerhead και το Surflex βασίζεται ουσιαστικά στην προσέγγιση του Böhm , ενώ η εμπειρική Οι βασισμένες συναρτήσεις βαθμολογίας που χρησιμοποιούνται από το GOLD και το Glide (http://www.schrodinger.com/) είναι τροποποιήσεις της συνάρτησης ChemScore του Eldridge. Το FlexX , από την άλλη πλευρά, εφαρμόζει τροποποιήσεις και των δύο συναρτήσεων Böhm και Eldridge. Τέλος, το " Η συνάρτηση βαθμολόγησης MolDock Score" που χρησιμοποιείται από το MVDTM προέρχεται από το PLP αλλά προσθέτει έναν νέο όρο δεσμών υδρογόνου και χρησιμοποιεί νέα σχήματα φόρτισης.Έχει υποστηριχθεί ότι οι εμπειρικές συναρτήσεις βαθμολόγησης είναι αυτές που δίνουν καλύτερα αποτελέσματα.

Co-funded by the Erasmus+ Programme of the European Union

Λειτουργίες βαθμολόγησης βασισμένες στη γνώση

Οι συναρτήσεις βαθμολόγησης που βασίζονται στη γνώση βασίζονται στη γνώση των προτιμήσεων επαφής και βασίζονται στην ιδέα της κλασικής στατιστικής φυσικής ότι οι παρατηρούμενες κατανομές γεωμετριών μπορούν να χρησιμοποιηθούν για να συναχθεί το δυναμικό που προκαλεί την παρατηρούμενη κατανομή. Προτάθηκαν για πρώτη φορά για τη μελέτη της αναδίπλωσης των πρωτεϊνών, αλλά έχουν επίσης χρησιμοποιηθεί στη σύνδεση με πρωτεϊνικούς συνδέτες και για την πρόβλεψη πρωτεϊνικών δομών και συμπλεγμάτων πρωτεΐνης-πρωτεΐνης. Στο πεδίο σύνδεσης πρωτεΐνης-συνδέτη, χρησιμοποιούνται για τη βαθμολογία θέσεων δέσμευσης συνδέτη μέσω σχετικά απλών δυναμικών ζεύγους ατομικής αλληλεπίδρασης που δημιουργούνται από στατιστικές αναλύσεις πειραματικά προσδιορισμένων δομών πρωτεΐνης-συνδέτη.

Λειτουργίες βαθμολόγησης βασισμένες στη γνώση

Οι συναρτήσεις βαθμολόγησης γνώσης προσπαθούν να εξάγουν κανόνες συλλαμβάνοντας τέτοιες πληροφορίες όπως η συχνότητα εμφάνισης και η μη εμφάνιση (δηλ. αρνητικά δεδομένα) διαφορετικών επαφών ζεύγους ατόμου-ατόμου και άλλες τυπικές αλληλεπιδράσεις στις πειραματικές δομές συμπλεγμάτων πρωτεΐνης-συνδέτη, επειδή υποτίθεται ότι :

(a) οι διατομικές αποστάσεις που εμφανίζονται συχνότερα από κάποια μέση τιμή θα πρέπει να αντιπροσωπεύουν ευνοϊκές επαφές και αντίστροφα. και

(b) η παρατηρούμενη κατανομή των αποστάσεων μεταξύ ζευγών διαφορετικών τύπων ατόμων αντανακλά την ενέργεια αλληλεπίδρασής τους.

Στην πράξη, μεγάλα σετ εκπαίδευσης δομών πρωτεΐνης-προσδέματος αναλύονται για να παρέχουν σύνολα συναρτήσεων κατανομής που στη συνέχεια μετατρέπονται σε σύνολα δυναμικών ζευγών ατόμων μέσω του αντίστροφου νόμου Boltzmann (ο οποίος παρέχει μια ενεργειακή τιμή για μια δεδομένη κατάσταση με βάση τα παρατηρούμενα πιθανότητες).

Λειτουργίες βαθμολόγησης βασισμένες στη γνώση

Ανάμεσα στα δυνατά σημεία αυτών των λειτουργιών είναι:

(a) Οι απλές ατομικές τους αλληλεπιδράσεις-δυναμικό ζεύγους έχουν χαμηλό υπολογιστικό κόστος, πράγμα που σημαίνει ότι οι μεγάλες σύνθετες βάσεις δεδομένων μπορούν να ελεγχθούν αποτελεσματικά. και

(b) ότι δεν απαιτούν να προκύψουν πειραματικές συγγένειες δέσμευσης.

Co-funded by the Erasmus+ Programme of the European Union

Λειτουργίες βαθμολόγησης βασισμένες στη γνώση

Αντίθετα, ορισμένοι περιορισμοί είναι ότι:

(a) είναι δύσκολο να προβλεφθεί η συμπεριφορά τους (δηλ. θα πρέπει να χρησιμοποιούνται για VLS μόνο όταν είναι διαθέσιμες αρκετές πληροφορίες συνδέτη και δομής για την επικύρωση της ρύθμισης του υπολογισμού).

(b) Έχουν σχεδιαστεί για να αναπαράγουν πειραματικές δομές αντί για δεσμευτικές ενέργειες.

(c) χρησιμοποιούνται για τον εντοπισμό δολωμάτων ή σε συνδυασμό με άλλες λειτουργίες βαθμολόγησης κατά τη διάρκεια του VLS αλλά όχι κατά τη φάση βελτιστοποίησης. και

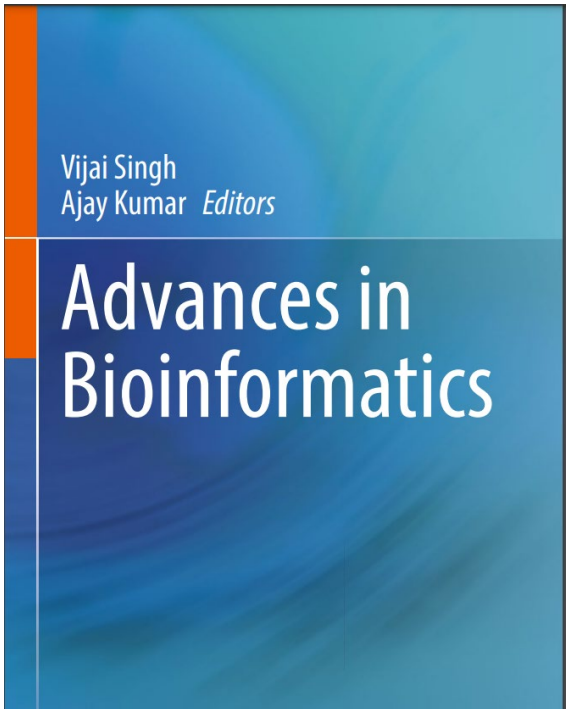
(d) Δεν είναι γενικά ή αρκετά ακριβή λόγω του περιορισμένου αριθμού αλληλεπιδράσεων που μπορούν να συναχθούν από τις κρυσταλλικές δομές και την ανεπαρκή περιγραφή των απωστικών δυνάμεων.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Ποιοι είναι οι τύποι των λειτουργιών βαθμολόγησης που χρησιμοποιούνται στο λογισμικό σύνδεσης;
- Δώστε μερικά παραδείγματα και εξηγήστε τα χαρακτηριστικά τους.

Co-funded by the Erasmus+ Programme of the European Union



Vijai Singh, Ajay Kumar (2021) Προόδους στη Βιοπληροφορική Springer Nature Singapore Pte Ltd

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

Θέμα 2. Πρόβλημα σύνδεσης

Μάθημα 5. Η σημασία της εξέτασης της ευελιξίας του υποδοχέα κατά τη σύνδεση

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Διαμορφωτική μεταβλητότητα
- Αλλοστερία
- Σύνδεση με εύκαμπτες πλαϊνές αλυσίδες
- Επαγόμενη προσαρμογή βάσης

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Η σύνδεση εφαρμόζεται για πρόβλεψη ο τρόπος δέσμευσης και η συγγένεια των μικρών μορίων. Όταν η βάση σύνδεσης εφαρμόζεται σε μεγάλο σύνθετη βάση δεδομένων, αναφέρεται ως εικονική προβολή (VS) και συχνά επικεντρώνεται στο πρόβλεψη νέων προσδεμάτων με υψηλή συγγένεια για ένα μόριο στόχο για τον εμπλουτισμό του συνόλου των ενώσεων για πειραματική δοκιμή. Ο σχεδιασμός de novo εκτελείται με σκοπό την πρόβλεψη νέων ενώσεις στον νέο χημικό χώρο. Τεχνικές χαρτογράφησης θραυσμάτων που χρησιμοποιούν λειτουργικές Συχνά εφαρμόζονται ομάδες σε θέσεις δέσμευσης ανιχνευτών για αυτόν τον τύπο προσέγγισης.

Διαμορφωτική μεταβλητότητα

Η κατανόησή μας για τη δέσμευση υποδοχέα-συνδέτη έχει προχωρήσει σημαντικά από την αρχική μοντέλο κλειδώματος και κλειδιού που προτάθηκε από τον Fischer (1890). Πρώιμες πειραματικές μελέτες έδειξαν ότι η Η πράξη δέσμευσης συνδέτη επηρεάζει τη διαμόρφωση της πρωτεΐνης, που αναφέρεται ως διαμορφωτική επαγωγή ή επαγόμενη προσαρμογή. Ένα άλλο μοντέλο σύνδεσης συνδέτη είναι διαμορφωτική επιλογή, όπου ο συνδέτης επιλέγει έναν εταίρο δέσμευσης μεταξύ τους διαθέσιμες καταστάσεις στο διαμορφωτικό σύνολο, μετατοπίζοντας έτσι την κατανομή του πληθυσμού.

Διαμορφωτική μεταβλητότητα

οι συντεταγμένες για πρωτεΐνη που έχουν ληφθεί με περίθλαση ακτίνων Χ και αυτή είναι διαθέσιμα από το ΠΣΠ αντιστοιχούν στους μέσους όρους των συντεταγμένες αυτής της πρωτεΐνης:

(1) στα κελιά μονάδας που αποτελούν μέρος του διαθλασμένου κρυστάλλου. και

(2) κατά το χρόνο που η περίθλαση πραγματοποιείται το πείραμα.

Παρόλα αυτά, οι γραμμές που περιέχουν συντεταγμένες ατόμων στα αρχεία PDB περιέχουν επίσης α τιμή παραμέτρου (δηλαδή ο συντελεστής θερμοκρασίας ή ο συντελεστής B) που μπορεί να θεωρηθεί ως μέτρο του πόσο το αντίστοιχο το άτομο ταλαντώνεται ή δονείται γύρω από την καθορισμένη θέση με τις αντίστοιχες συντεταγμένες (δηλαδή όσο μεγαλύτερη είναι η τιμή του ο παράγοντας B, τόσο λιγότερο ακριβείς είναι οι συντεταγμένες του ατόμου).

Co-funded by the Erasmus+ Programme of the European Union

Αλλοστερία

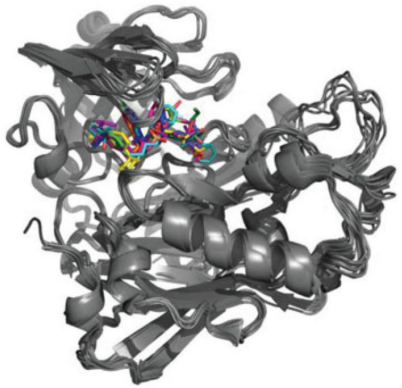
Η ευελιξία πρωτεΐνης-προσδέματος κατά τη δέσμευση είναι κρίσιμη για τη σωστή κατανόηση του αλλοστερικού κανονισμός λειτουργίας. Ο Nussinov και οι συνεργάτες του υπέθεσαν ότι οι περισσότερες πρωτεΐνες υπάρχουν σε ένα σύνολο καταστάσεις, και έτσι οι περισσότερες πρωτεΐνες έχουν τη δυνατότητα αλλοστερίας . Με βάση την πειραματική βιβλιογραφία, απέδειξαν ότι η δέσμευση ενός αλλοστερικού ο συνδέτης μετατόπισε τον πληθυσμό των διαμορφωτικών υποκαταστημάτων , επηρεάζοντας έτσι την ικανότητα του άλλους συνδέτες για τη σύνδεση μιας εναλλακτικής θέσης.

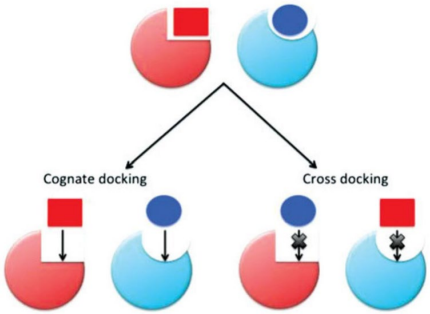
Co-funded by the Erasmus+ Programme of the European Union

Αλλοστερία

Η αντοχή στα φάρμακα είναι ένα αυξανόμενο πρόβλημα που απαιτεί νέες προσεγγίσεις στη φαρμακευτική θεραπεία. Η διερεύνηση του αλλοστερικού ελέγχου σε πρωτεϊνικούς στόχους παρέχει συναρπαστικές ευκαιρίες για εύρεση νέων τρόπους δράσης και ως εκ τούτου, υπερνίκηση της αναδυόμενης αντίστασης και ανάπτυξη κοκτέιλ από φάρμακα για τη βελτίωση της θεραπείας.

Το πρόβλημα του cross-docking





Η έρευνα για την ευελιξία των πρωτεϊνών και την αλλοστερία έχει υποστηρίξει τη σημασία του που αντιπροσωπεύουν πολλαπλές καταστάσεις σε δεσμευτικές μελέτες. Οι Mobley & Dill (2009) σημείωσαν ότι Η ελεύθερη ενέργεια δέσμευσης (ΔGbind) και η εντροπία επηρεάζονται από το σχήμα και το πλάτος του συνόλου διαμορφωτικό τοπίο, παρά μια ενιαία άκαμπτη στάση. Οι Murray et al. (1999) εξέτασε προσεγγίσεις για τη χρήση άκαμπτων υποδοχέων σε μελέτες σύνδεσης. Όταν οι συγγραφείς προσπάθησαν να ελλιμενιστούν ένας γνωστός συνδέτης σε μια δομή πρωτεΐνης επιλύθηκε παρουσία διαφορετικού συνδετήρα (αναφέρεται ως cross-docking), διαπίστωσαν ότι η ενεργή θέση ήταν προκατειλημμένη προς τον εγγενή συνδέτη. Μια ποικιλία διαφορών στην επιφάνεια της θέσης δέσμευσης αναγνωρίστηκε για το ίδια πρωτεΐνη λυμένη με διαφορετικούς συνδέτες ή απουσία συνδέτη.

Co-funded by the Erasmus+ Programme of the European Union

Το πρόβλημα του cross-docking

Μια υποκείμενη υπόθεση στις προσπάθειες άκαμπτης σύνδεσης είναι ότι η σύνθετη κατάσταση είναι η χαμηλότερη κατάσταση ελεύθερης ενέργειας. Ωστόσο, η δεσμευμένη σε συνδέτη κατάσταση δεν είναι πάντα η χαμηλότερη ενεργειακή κατάσταση για είτε η πρωτεΐνη είτε ο συνδέτης.

Co-funded by the Erasmus+ Programme of the European Union

Πηγές δομικών πληροφοριών

Η χρήση ενός συνόλου παρόμοιων διαμορφώσεων δημιουργεί μόνο μια πεπερασμένη ποσότητα πληροφοριών πιθανούς δεσμευτικούς εταίρους. Θα πρέπει να διερευνηθεί πρόσθετος χώρος υποδοχέα, ειδικά σε σενάρια όπου παρουσιάζουν ενδιαφέρον εναλλακτικοί τρόποι σύνδεσης. Κρυσταλλογραφία ακτίνων Χ, NMR δομές, μοντέλα ομολογίας, ανάλυση κανονικού τρόπου (NMA) και μοριακή μηχανική Οι προσομοιώσεις είναι όλες πιθανές πηγές δομικής ποικιλομορφίας. Η πρόοδος της τεχνολογίας έχει επέτρεψε σημαντικές αυξήσεις στις διαθέσιμες διαρθρωτικές πληροφορίες για έναν τεράστιο αριθμό στόχους υποδοχέων.

Co-funded by the Erasmus+ Programme of the European Union

Πηγές δομικών πληροφοριών

Αρκετές ερευνητικές ομάδες, καθώς και χιλιάδες των χρηστών υπολογιστών, έχουν αφιερώσει την ισχύ του υπολογιστή τους στην πρόβλεψη κατανεμημένης δομής πρωτεΐνες με Folding@HOME , POEM@HOME και Rosetta@HOME .

Η μαλακή σύνδεση επιτρέπει περιορισμένη ανακάλυψη παρατηρήθηκε νέος χώρος διαμόρφωσης κατά τη σύνδεση με εύκαμπτα δείγματα πλευρικής αλυσίδας διαμορφωτικός χώρος. Πιο πρόσφατες τεχνικές ενσωματώνουν σύνολα δομών που επιτρέπουν για μεγαλύτερο μέτρο εγγενούς ευελιξίας ή για την αποκάλυψη νέων διαμορφώσεων. Βασιζόμενοι σε πολλές πρωτεϊνικές διαμορφώσεις, νέος χημικός χώρος μπορεί να εξερευνηθεί, αλλοστερικές τοποθεσίες μπορούν να γίνουν ανακαλύφθηκαν και μπορούν να δημιουργηθούν πιο ακριβείς διαμορφώσεις σύνδεσης. Η αποδεκτή μέτρηση για μια καλά συνδεδεμένη στάση είναι ≤2,0 Å RMSD από την πειραματικά προσδιορισμένη δομή.

Co-funded by the Erasmus+ Programme of the European Union

Μαλακή βάση σύνδεσης

Η ευελιξία πρωτεΐνης μπορεί να ενσωματωθεί σε ένα μοντέλο θέσης δέσμευσης πριν από την έναρξη της σύνδεσης ή μπορεί να επιτραπεί η μετα-προσάρτηση μέσω της τελειοποίησης του δεσμευμένου συμπλέγματος. Η σύνδεση μπορεί να πραγματοποιηθεί σε μια μέση δομή ή σύνολο, σε ένα σύνολο διαμορφώσεων υποδοχέα (διαμορφώσεις πολλαπλών υποδοχέων (MRC)) ή με διαμορφώσεις που δημιουργούνται «εν κινήσει» (επαγόμενη προσαρμογή σύνδεσης (IFD)).

Μέθοδοι χαλάρωσης

Η βελτιστοποίηση του συμπλέγματος αγκυροβόλησης είναι μια άλλη απλή προσέγγιση που προσαρμόζει την ευελιξία των πρωτεϊνών μοντελοποιώντας εφέ επαγόμενης προσαρμογής. Η ενσωμάτωση της ευελιξίας πρωτεΐνης στο «οπίσθιο άκρο» μπορεί να γίνει μόνο όταν η τεχνική σύνδεσης βασίστηκε σε μια δομή με όλα τα άτομα. δεν μπορεί να εκτελεστεί σε ένα πλέγμα ή σε οποιαδήποτε άλλη κατάσταση όπου η πρωτεΐνη δεν αναπαρίσταται ρητά. Οι προσομοιώσεις Monte Carlo (MC) ή MD είναι μια δημοφιλής επιλογή για βελτίωση, επειδή επιτρέπουν τη βελτιστοποίηση των θέσεων σύνδεσης, τη διερεύνηση των επιδράσεων του διαλύτη, την εξέταση της κινητικής σταθερότητας και την πρόβλεψη του ΔGbind από συναρτήσεις βαθμολόγησης που βασίζονται στη φυσική. Η βελτίωση εκτελείται συχνά ως τελευταίο βήμα σε πολλές από τις προσεγγίσεις σύνδεσης.

Co-funded by the Erasmus+ Programme of the European Union

Σύνδεση με εύκαμπτες πλαϊνές αλυσίδες

Οι πρώτες προσπάθειες για την επιτόπια σύνδεση επικεντρώθηκαν στις πλευρικές αλυσίδες, με τη χρήση μιας βιβλιοθήκης περιστροφικού που βασίζεται σε διεδρικές γωνίες ραχοκοκαλιάς για την περιγραφή της ευελιξίας των πρωτεϊνών.

Η κίνηση της πλευρικής αλυσίδας κατά τη δέσμευση συνδέτη έχει επίσης προσεγγιστεί μέσω μεθόδων δειγματοληψίας συνεχούς. Τεχνικές που εφαρμόζουν κάποια μορφή συνεχούς δειγματοληψίας στην κίνηση πλευρικής αλυσίδας είναι το AutoDock 4.0, Dynasite /GOLD, FlexX , ICM/MC, Mining Minima, Skelgen και SLIDE.

Το να επιτρέπεται η ευελιξία της πλευρικής αλυσίδας απαιτεί λιγότερους πόρους από τις μεθόδους πλήρους ευελιξίας και επέτρεψε κάποια διαμορφωτική μεταβλητότητα μέσω της εξερεύνησης προσανατολισμών χαμηλής ενέργειας των πλευρικών αλυσίδων. Ωστόσο, η ενσωμάτωση της ευελιξίας της πλευρικής αλυσίδας απέτυχε σε περιπτώσεις πρωτεϊνών με μεγάλης κλίμακας αναδιατάξεις άρθρωσης ή βρόχου ή ακόμη και μετακίνηση πλευρικών αλυσίδων που εξαρτάται από τη ραχοκοκαλιά, καμία από τις οποίες δεν μπορεί να ληφθεί υπόψη με διακριτή ή συνεχή δειγματοληψία

Co-funded by the Erasmus+ Programme of the European Union

Επαγόμενη προσαρμογή βάσης

Σημαντική πρόοδος έχει σημειωθεί από την αρχική εφαρμογή των εύκαμπτων πλευρικών αλυσίδων στο docking. Η δειγματοληψία κίνησης των πλευρικών αλυσίδων εν κινήσει αυξάνει τον δυνητικό ενεργειακό χώρο, αλλά μπορεί να παραβλέπει τις παγκόσμιες διαμορφωτικές αλλαγές. Το να επιτρέπονται αλλαγές διαμόρφωσης κατά τη σύνδεση μπορεί να είναι μια πολύ ακριβής τεχνική για τη μοντελοποίηση δεσμευμένων θέσεων συμπλεγμάτων πρωτεΐνης-συνδέτη. Το IFD είναι σημαντικό επειδή μπορεί να επιτρέψει στην προσομοίωση σύνδεσης να αναζητήσει νέο διαμορφωτικό χώρο. Ωστόσο, αυτή η δειγματοληψία βαθμών ελευθερίας υποδοχέα και συνδέτη είναι επίσης αρκετά υπολογιστικά εντατική, γεγονός που περιορίζει την εφαρμογή της σε μεγάλης κλίμακας μελέτες εικονικής διαλογής.

Co-funded by the Erasmus+ Programme of the European Union

Σύνδεση συνόλου

Η σύνδεση συνόλου διαφέρει από την IFD στο ότι η ευελιξία πρωτεΐνης λαμβάνεται υπόψη πριν από την πραγματική σύνδεση. Υπάρχουν δύο διαφορετικοί τύποι μεθόδων για την αναπαράσταση της ευελιξίας του υποδοχέα κατά τη σύνδεση: σύνολα που βασίζονται σε πλέγμα και σύνολα που βασίζονται σε δομή. Συχνά, εναλλακτικές διαμορφώσεις πρωτεΐνης αναπαρίστανται σε ένα πλέγμα δυναμικού δύο σωμάτων, επιτρέποντας γρήγορες, φθηνές προσομοιώσεις σύνδεσης. Η ευελιξία μπορεί επίσης να ενσωματωθεί σε προβλέψεις δέσμευσης μέσω της διαδοχικής πρόσδεσης σε δομές σε ένα διαμορφωτικό σύνολο ή σύνδεσης σε μια δομή υποδοχέα κατά μέσο όρο/ενωμένη. Λόγω του χρόνου που απαιτείται για την αρχική δημιουργία συνόλου και τις επαναλαμβανόμενες διαδρομές σύνδεσης σε ένα σύνολο στατικών δομών, η διαδοχική σύνδεση είναι συνήθως η πιο εντατική υπολογιστικά προσέγγιση. Ωστόσο, αποφεύγει την ανακάλυψη συμπλεγμάτων υποδοχέα-συνδέτη που είναι φυσικά αδύνατα (παράδοξοι συνδετήρες), τα οποία μερικές φορές φαίνονται στα αποτελέσματα από τη σύνδεση σε μια μέση δομή ή πλέγμα.

Co-funded by the Erasmus+ Programme of the European Union

Σύνολα που βασίζονται σε πλέγμα

Η πρώτη μέθοδος σύνδεσης για χρήση σύνθετου πλέγματος εκτελέστηκε μέσω του DOCK3.5 προκειμένου να αξιολογηθεί ο αντίκτυπος της αναπαράστασης της μεταβλητότητας της διαμόρφωσης ως δομικού συνόλου στα αποτελέσματα της πόζας σύνδεσης.

Η ενσωμάτωση της ευελιξίας πρωτεΐνης μέσω σύνδεσης σε ένα πλέγμα αλληλεπίδρασης που αντιπροσωπεύει το σύνολο των υποδοχέων είναι μια κοινή προσέγγιση για το SBDD και πρόσθετες μέθοδοι είναι αυτές που βασίζονται σε AutoDock 3.0, Flog, GRID/CPCA, IFREDA, ISCD και σε σύνολα συναινετικών δομών.

Co-funded by the Erasmus+ Programme of the European Union

Σύνολα με βάση τη δομή

Η μεθοδολογία ταυτόχρονης αναζήτησης πολλαπλών αντιγράφων (MCSS) ήταν η πιο πρώιμη υπολογιστική προσέγγιση για τη χαρτογράφηση θέσεων δέσμευσης με ανιχνευτές λειτουργικών ομάδων.

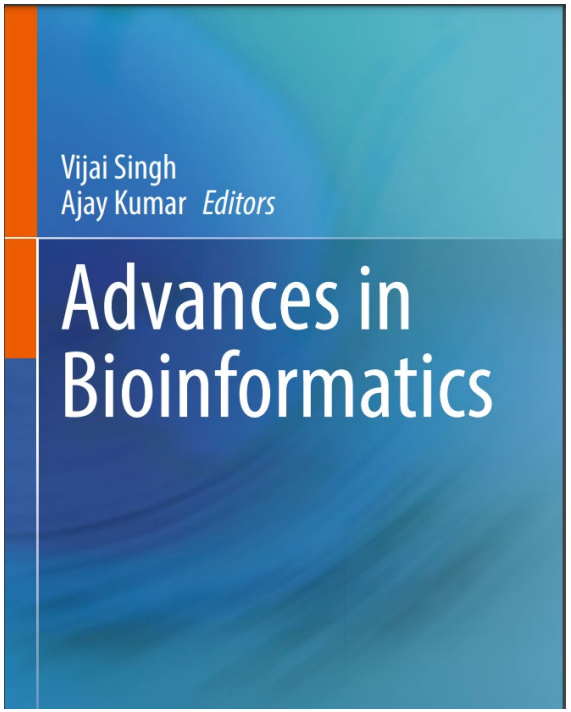
Από την ποικιλία των τεχνικών που έχουν αναπτυχθεί για τη βελτίωση των προβλέψεων δέσμευσης για σύμπλοκα πρωτεΐνης-συνδέτη, η χρήση ενός διαμορφωτικού συνόλου είναι ίσως η πιο κοινή προσέγγιση.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Πόσο σημαντικό είναι να λαμβάνεται υπόψη η ευελιξία του υποδοχέα κατά τη σύνδεση;
- Ποιοι τύποι διαδικασιών σύνδεσης υπάρχουν σύμφωνα με την ευελιξία των προσδεμάτων και των πρωτεϊνών;

Co-funded by the Erasmus+ Programme of the European Union



Vijai Singh, Ajay Kumar (2021) Προόδους στη Βιοπληροφορική Springer Nature Singapore Pte Ltd

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

Θέμα 2. Πρόβλημα σύνδεσης

Μάθημα 6. Κύρια χαρακτηριστικά επιλεγμένου λογισμικού σύνδεσης

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Παραδείγματα λογισμικού σύνδεσης

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Η μοριακή σύνδεση έχει γίνει ένα σημαντικό κοινό συστατικό της εργαλειοθήκης ανακάλυψης φαρμάκων και οι σχετικές φθηνές συνέπειες και η αντιληπτή απλότητα χρήσης της έχουν τονώσει μια ολοένα αυξανόμενη δημοτικότητα στις ακαδημαϊκές κοινότητες. Ο στόχος της σύνδεσης συνδέτη-πρωτεΐνης είναι να προβλέψει τον κυρίαρχο τρόπο(ους) δέσμευσης ενός συνδέτη με μια πρωτεΐνη γνωστής τρισδιάστατης δομής. Οι επιτυχημένες μέθοδοι σύνδεσης αναζητούν αποτελεσματικά χώρους υψηλών διαστάσεων και χρησιμοποιούν μια συνάρτηση βαθμολόγησης που κατατάσσει σωστά τις υποψήφιες βάσεις σύνδεσης.

eHiTS ® (ηλεκτρονικός έλεγχος υψηλής απόδοσης, http://www.simbiosys.ca/ehits/

Το eHiTS ® αναπτύσσεται από τη SymBioSys Inc. στο Τορόντο (Καναδάς) και τα κύρια πλεονεκτήματά του είναι:

(a) Είναι εύκολο στη χρήση?

(b) αποδίδει πολύ καλά (είναι γρήγορο και ακριβές). και

(c) έχει πολλά αυτοματοποιημένα χαρακτηριστικά που απλοποιούν τη ροή εργασιών σχεδιασμού φαρμάκων και παρέχουν καινοτόμες λύσεις σε κοινά προβλήματα σύνδεσης (π.χ. κατάσταση πρωτονίωσης του ζεύγους συνδέτη/υποδοχέα, εξαντλητική αναζήτηση των στάσεων του συνδέτη, επιτάχυνση των υπολογισμών στο VLS Αυτόματη αναγνώριση πιθανών θέσεων δέσμευσης, ικανότητα προσαρμογής της συνάρτησης βαθμολόγησης στα χαρακτηριστικά της θέσης δέσμευσης του υποδοχέα, κ.λπ.).

eHiTS ® (ηλεκτρονικός έλεγχος υψηλής απόδοσης, http://www.simbiosys.ca/ehits/

Το eHiTS ® δεν είναι τέλειο και έχει επίσης ορισμένους περιορισμούς όπως:

(a) η ευελιξία του υποδοχέα περιορίζεται στις περιστροφές των ομάδων -OH ή -NH3+ από ορισμένα αμινοξέα (δηλ. Ser, Thr , Tyr και Lys).

(b) Όλα τα συστήματα δακτυλίου στους συνδετήρες θεωρούνται άκαμπτα και επομένως, οι διαμορφώσεις τους δεν αλλάζουν κατά τη σύνδεση (επομένως, για μια πλήρη δειγματοληψία διαμόρφωσης είναι απαραίτητο να χρησιμοποιηθούν πολλαπλοί διαμορφωτές δακτυλίου). και

(c) Δεν μπορούν να επιβληθούν περιορισμοί που βασίζονται στη γνώση στη βάση σύνδεσης (π.χ. ένα συγκεκριμένο άτομο συνδέτη δεν μπορεί να εξαναγκαστεί να βρίσκεται σε μια συγκεκριμένη θέση στις στάσεις, ορισμένες αλληλεπιδράσεις δεν μπορούν να αποτραπούν από την εμφάνιση κ.λπ.).

GOLD ™ (Genetic Optimization for Ligand Docking; http://www.ccdc.cam.ac.uk/products/life_sciences/gold/

Το GOLD ™ αναπτύχθηκε αρχικά από το Πανεπιστήμιο του Σέφιλντ (http://www.shef.ac.uk/), την GlaxoSmithKline plc (http://www.gsk.com/) και το CCDC (Cambridge Crystallographic Data Centre, http: //www.ccdc.cam.ac.uk/). Επί του παρόντος, η CCDC αναπτύσσει και διατηρεί το GOLD ™ σε ένα συνεχιζόμενο έργο με την Astex Therapeutics (http://www.astex-therapeutics.com/) και τη Syngenta (http:// www.syngenta.com).

GOLD ™ (Genetic Optimization for Ligand Docking; http://www.ccdc.cam.ac.uk/products/life_sciences/gold/

Τα κύρια πλεονεκτήματα του GOLD είναι ότι: (α) μπορεί να συμπεριληφθεί στους υπολογισμούς κάποια ευελιξία της ραχοκοκαλιάς και της πλευρικής αλυσίδας (προς το παρόν, έως και δέκα υπολείμματα που ορίζονται από τον χρήστη). (β) το πρόγραμμα μπορεί να χρησιμοποιήσει λειτουργίες βαθμολόγησης που ορίζονται από το χρήστη και όχι μόνο τις δύο προεπιλεγμένες που παρέχονται με το GOLD ™ (και ακόμη και να τις τροποποιήσει). (γ) οι ενεργειακές συναρτήσεις βασίζονται εν μέρει σε διαμορφωτικές και μη δεσμευμένες πληροφορίες επαφής που προέρχονται από τη βάση δεδομένων του Cambridge Structural Database (http://www.ccdc.cam.ac.uk/products/csd/)· (δ) μια ποικιλία επιλογών περιορισμού μπορεί να οριστεί για τη βάση σύνδεσης. (ε) κρυσταλλογραφικά μόρια νερού στη θέση δέσμευσης του συνδέτη μπορούν να ληφθούν υπόψη κατά τη σύνδεση. (στ) είναι βελτιστοποιημένο για παράλληλη εκτέλεση σε δίκτυα επεξεργαστών και μια κατανεμημένη έκδοση του GOLD ™ είναι διαθέσιμη για χρήση σε εμπορικά συστήματα δικτύου. (ζ) μπορεί να χειριστεί αυτόματα άτομα μετάλλου εάν έχουν ρυθμιστεί σωστά στο αρχείο εισαγωγής πρωτεΐνης. και (η) Τα αποτελέσματα VHTS μπορούν να αναλυθούν και να υποβληθούν σε επεξεργασία μετά την εύκολη επεξεργασία με τα συνοδευτικά προγράμματα SILVER ™ ή GoldMine ™ .

GOLD ™ (Genetic Optimization for Ligand Docking; http://www.ccdc.cam.ac.uk/products/life_sciences/gold/

Οι κύριοι περιορισμοί του GOLD είναι ότι: (α) χρησιμοποιεί μια τυχαία μέθοδο (δηλ. μια GA) αντί για μια συστηματική για να αναζητήσει τις θέσεις του συνδέτη στο σύμπλεγμα πρωτεΐνης-συνδέτη. (β) οι προεπιλεγμένες συναρτήσεις βαθμολόγησης έχουν βελτιστοποιηθεί για να προβλέψουν τις θέσεις του συνδέτη και όχι τις συγγένειες δέσμευσης. (γ) έχουν αναφερθεί συστηματικά προβλήματα στην κατάταξη πολύ πολικών προσδεμάτων και γενικών προσδεμάτων σε μεγάλες κοιλότητες. (δ) η πρωτεΐνη και ο συνδέτης απαιτούν προηγούμενη ρύθμιση από τον χρήστη που περιλαμβάνει, για παράδειγμα, την προσθήκη όλων των ατόμων υδρογόνου. και (ε) Το GOLD ™ δεν μεταβάλλει τις καταστάσεις ταυτομερούς ή ιονισμού κατά τη σύνδεση και, επομένως, εάν ο χρήστης δεν είναι σίγουρος για τη σωστή ταυτομερή κατάσταση ενός συγκεκριμένου υπολείμματος, τότε πρέπει να εκτελεστούν ξεχωριστές δοκιμές GOLD

Molegro Virtual Docker (Molecule & Allegro Virtual Docker; http://www.molegro.com/

Το MVDTM (παλαιότερα ονομαζόταν MolDock) αναπτύσσεται από τη Molegro ApS (Άαρχους, Δανία).

Co-funded by the Erasmus+ Programme of the European Union

Molegro Virtual Docker (Molecule & Allegro Virtual Docker; http://www.molegro.com/

Τα κύρια πλεονεκτήματα της MVD είναι ότι: (α) μπορεί να ρυθμίσει αυτόματα τις δομές εισόδου εκχωρώντας φορτίσεις, εντολές δεσμών και υβριδισμό και προσθέτοντας υδρογόνα. (β) είναι σε θέση να προβλέψει αυτόματα πιθανές θέσεις δέσμευσης στον υποδοχέα. (γ) μπορεί να αντιμετωπίσει την ευελιξία της πλευρικής αλυσίδας του υποδοχέα λαμβάνοντας υπόψη τις επαγόμενες αλληλεπιδράσεις προσαρμογής. (δ) είναι σε θέση να προσδεθεί σε προυπολογισμένα ενεργειακά δίκτυα (πράγμα που επιταχύνει τους υπολογισμούς). (ε) μπορεί να αντιμετωπίσει περιορισμούς που καθορίζονται από το χρήστη κατά τη σύνδεση. (στ) μπορεί να επωφεληθεί από τη χρήση προτύπων (δηλαδή φαρμακοφόρων) κατά την ελλιμενοποίηση· και (ζ) είναι σε θέση να διανέμει τους υπολογισμούς σε πολλούς υπολογιστές.

Molegro Virtual Docker (Molecule & Allegro Virtual Docker; http://www.molegro.com/

Οι κύριοι περιορισμοί της MVD είναι ότι: (α) η στοχαστική φύση της μηχανής σύνδεσης συνεπάγεται ότι μπορεί να χρειαστούν περισσότερες από μία διαδρομές σύνδεσης για τον προσδιορισμό του σωστού τρόπου σύνδεσης. και (β) η κατάσταση πρωτονίωσης του ζεύγους συνδέτη/υποδοχέα δεν αναζητείται εξαντλητικά και χρησιμοποιούνται προκαθορισμένες καταστάσεις πρωτονίωσης κατά τη σύνδεση.

AutoDock (http://autodock.scripps.edu).

Το AutoDock (παλαιότερα ονομαζόταν AutoDoq δεδομένου ότι χρησιμοποιούσε τεταρτημόρια για την εκτέλεση περιστροφών) ήταν το πρώτο πακέτο σύνδεσης για τη μοντελοποίηση του συνδετήρα με πλήρη ευελιξία διαμόρφωσης. Αναπτύσσεται από τον καθηγητή Arthur J. Olson στο Τμήμα Μοριακής Βιολογίας του Ερευνητικού Ινστιτούτου Scripps στη La Jolla (Καλιφόρνια, ΗΠΑ) και, επί του παρόντος, είναι το λογισμικό σύνδεσης με τις περισσότερες αναφορές στη βιβλιογραφία. Το πακέτο αποτελείται από δύο προγράμματα (δηλ. AutoGrid και AutoDock), τα οποία εφαρμόζονται διαδοχικά. Το AutoGrid χρησιμοποιείται αρχικά για τον υπολογισμό της μη ομοιοπολικής ενέργειας αλληλεπίδρασης μεταξύ του άκαμπτου τμήματος του υποδοχέα και ενός ατόμου ανιχνευτή που βρίσκεται στα διάφορα σημεία πλέγματος του πλέγματος (όπου τα άτομα ανιχνευτή είναι εκείνα που υπάρχουν στους συνδετήρες που θα να προσαρτηθεί στον υποδοχέα).

AutoDock (http://autodock.scripps.edu).

του AutoDock είναι ότι: (α) μπορεί να αντιμετωπίσει την ευελιξία του υποδοχέα κατά τη σύνδεση. (β) μπορεί να χρησιμοποιηθεί σε blind-docking όταν η ακριβής θέση της θέσης δέσμευσης του συνδετήρα είναι άγνωστη. (γ) χρησιμοποιεί προ-υπολογισμένους χάρτες πλέγματος σε μια θέση δέσμευσης προκειμένου να προσδένει συνδετήρες με την επακόλουθη εξοικονόμηση υπολογιστικού χρόνου. (δ) οι χάρτες ατομικού πλέγματος μπορούν επίσης να χρησιμοποιηθούν ως οδηγός για το σχεδιασμό νέων συνδετών που συνδέονται με μεγαλύτερη συγγένεια στον υποδοχέα. (ε) έχει μια συνάρτηση βαθμολόγησης ελεύθερης ενέργειας που βασίζεται σε ανάλυση γραμμικής παλινδρόμησης, στο πεδίο δύναμης AMBER και σε ένα μεγάλο σύνολο συμπλεγμάτων πρωτεΐνης-προσδέματος με γνωστές σταθερές αναστολής . και (στ) παρέχει καλές συσχετίσεις μεταξύ των προβλεπόμενων σταθερών αναστολής και των πειραματικών.

AutoDock (http://autodock.scripps.edu).

του AutoDock είναι ότι: (α) η πρωτεΐνη και ο συνδέτης πρέπει να ρυθμιστούν πριν χρησιμοποιηθούν από το AutoGrid / AutoDock [π.χ. πρέπει να προστεθούν υδρογόνα και μερικά φορτία, τα μη πολικά υδρογόνα και τα φορτία τους πρέπει να συγχωνευθούν με μητρικό άτομο άνθρακα, περιστρεφόμενοι δεσμοί στον συνδέτη πρέπει να δημιουργηθούν κ.λπ. Το μεγαλύτερο μέρος αυτής της εργασίας μπορεί να γίνει με το AutoDockTools (http://autodock.scripps.edu/downloads/resources/adt)] . (β) οι διαφορετικοί αλγόριθμοι αναζήτησης πόζας είναι στοχαστικοί και όχι συστηματικοί. και (γ) η κατάσταση πρωτονίωσης του ζεύγους συνδέτη/υποδοχέα δεν αναζητείται εξαντλητικά και χρησιμοποιούνται προκαθορισμένες καταστάσεις πρωτονίωσης κατά τη σύνδεση.

Co-funded by the Erasmus+ Programme of the European Union

Glide ™ (Grid-based Ligand Docking with Energetics; http://www.schrodinger.com/

Το Glide ™ αναπτύσσεται από τη Schrödinger (Πόρτλαντ, Όρεγκον, ΗΠΑ). Τα κύρια πλεονεκτήματά του είναι ότι: (α) όχι μόνο θεωρεί ότι ο υποδοχέας είναι εύκαμπτος αλλά επίσης ότι ο υποδοχέας μπορεί να αλλάξει τη διαμόρφωσή του κατά τη δέσμευση του συνδέτη λόγω ενός μηχανισμού επαγόμενης προσαρμογής. (β) μπορεί να χρησιμοποιήσει περιορισμούς που καθορίζονται από το χρήστη για τον περιορισμό ενός ατόμου συνδέτη ώστε να βρίσκεται μέσα σε μια συγκεκριμένη περιοχή που ορίζεται σε σχέση με τα χαρακτηριστικά του υποδοχέα που είναι υπεύθυνα για τη δέσμευση του συνδέτη. (γ) μπορεί να χρησιμοποιηθεί μαζί με δύο ακόμη προϊόντα της Schrödinger (π.χ. Liaison ™ ή Qsite ™) για να ληφθούν πιο ακριβείς ενέργειες δέσμευσης για ζεύγη συνδέτη-υποδοχέα. (δ) μπορεί να εξετάσει μόρια νερού στην ενεργό θέση του υποδοχέα κατά τη διάρκεια της σύνδεσης ή όχι. και (ε) έχει επιλογές παράλληλης επεξεργασίας ή διανομής.

Glide ™ (Grid-based Ligand Docking with Energetics; http://www.schrodinger.com/

Οι κύριοι περιορισμοί του Glide είναι ότι: (α) Οι υποδοχείς PDB πρέπει να προετοιμαστούν χειροκίνητα (δηλαδή πρέπει να εκχωρηθούν οι τύποι ατόμων και οι εντολές δεσμών, να διορθωθούν οι καταστάσεις φορτίου και πρωτονίου, να επαναπροσανατολιστούν οι πλευρικές αλυσίδες εάν είναι απαραίτητο, οι ελλιπείς πλευρικές αλυσίδες στην ενεργή θέση να έχουν να ξαναχτιστεί, και να ανακουφιστούν οι στερικές συγκρούσεις). και (β) η κατάσταση πρωτονίωσης του ζεύγους συνδέτη/υποδοχέα δεν αναζητείται εξαντλητικά και πρέπει να χρησιμοποιούνται προκαθορισμένες καταστάσεις πρωτονίωσης κατά τη σύνδεση.

Ποιο λογισμικό σύνδεσης δίνει καλύτερα αποτελέσματα;

Η σύγκριση της προγνωστικής ισχύος και της απόδοσης των προγραμμάτων σύνδεσης πρωτεϊνών-συνδέτη είναι επί του παρόντος ένας ενεργός τομέας έρευνας. Αυτή η προγνωστική ισχύς συνήθως αξιολογείται μελετώντας την απόδοση του μελετημένου λογισμικού σε σχέση με άλλα προγράμματα σύνδεσης σε: (1) αναλύσεις επανασύνδεσης στις οποίες τα κρυσταλλογραφικά σύμπλοκα πρωτεΐνης-προσδέματος διαχωρίζονται και ο προκύπτων συνδέτης προσδένεται στην προκύπτουσα πρωτεΐνη. (2) μελέτες εμπλουτισμού στις οποίες το λογισμικό πρέπει να διακρίνει μεταξύ αληθινών συνδετικών και δολωμάτων για έναν στόχο πρωτεΐνης σε μελέτες εικονικής διαλογής. και (3) πρόβλεψη δεσμευτικών ελεύθερων ενεργειών από τη πόζα με το καλύτερο σκορ. Ωστόσο, η σύγκριση της απόδοσης των προγραμμάτων σύνδεσης είναι εκπληκτικά πολύπλοκη και πολλοί παράγοντες πρέπει να ληφθούν υπόψη πριν επιλεγεί το καλύτερο. Δεν έχει βρεθεί ακόμη πρόγραμμα που να προσφέρει ισχυρές και ακριβείς λύσεις γενικής εφαρμογής στα περισσότερα από τα διάφορα προβλήματα σύνδεσης.

Ποιο λογισμικό σύνδεσης δίνει καλύτερα αποτελέσματα;

(a) όλα τα προγράμματα πρέπει να επικυρώνονται με το ίδιο σύνολο συμπλεγμάτων πρωτεϊνών (δηλ. το λεγόμενο σετ δοκιμής).

(b) όλα τα προγράμματα πρέπει να αναζητούν τον ίδιο τρισδιάστατο χώρο για τη στάση του συνδέτη (αυτό είναι δύσκολο γιατί δεν ορίζουν όλα τα προγράμματα με τον ίδιο τρόπο τον χώρο αναζήτησης όπου θα προσδιοριστεί η σύνδεση του συνδετήρα και αυτό είναι σημαντικό για την επιτυχία της σύνδεσης.

(c) η ορθότητα των κορυφαίων στάσεων πρέπει να συγκρίνεται προσεκτικά με τις πειραματικές λύσεις και να μην βασίζεται σε τυφλά μέτρα RMSD.

(d) όλοι οι υποκαταστάτες πρέπει να παρασκευάζονται με τον ίδιο τρόπο. και

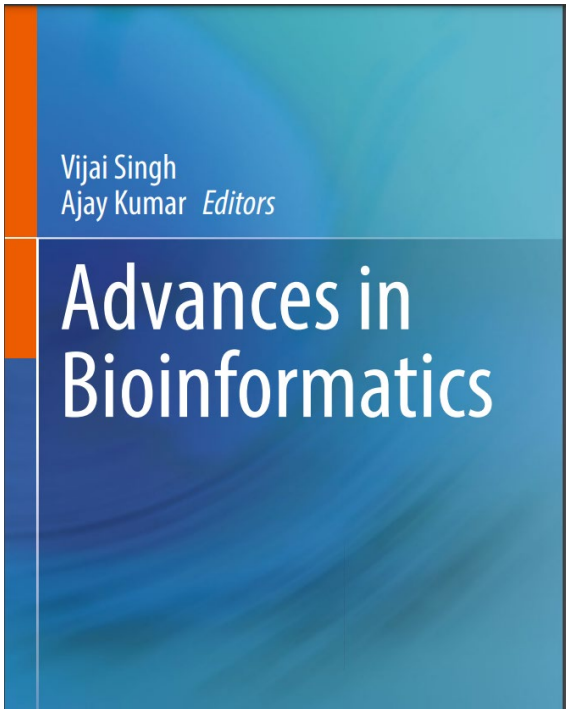
(e) ο χρόνος που επιτρέπεται για τη σύνδεση ενός συνδετήρα πρέπει να είναι ο ίδιος για όλα τα προγράμματα (επειδή ένα πρόγραμμα μπορεί να είναι καλύτερο από ένα άλλο όταν χρησιμοποιούνται γρήγορες ρυθμίσεις αλλά χειρότερο με αργές ρυθμίσεις).

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Δώστε μερικά παραδείγματα λογισμικού σύνδεσης.

Co-funded by the Erasmus+ Programme of the European Union



Vijai Singh, Ajay Kumar (2021) Προόδους στη Βιοπληροφορική Springer Nature Singapore Pte Ltd

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

Θέμα 3. Πρόβλημα αναδίπλωσης πρωτεΐνης

Μάθημα 1. Αναδίπλωση πρωτεΐνης

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

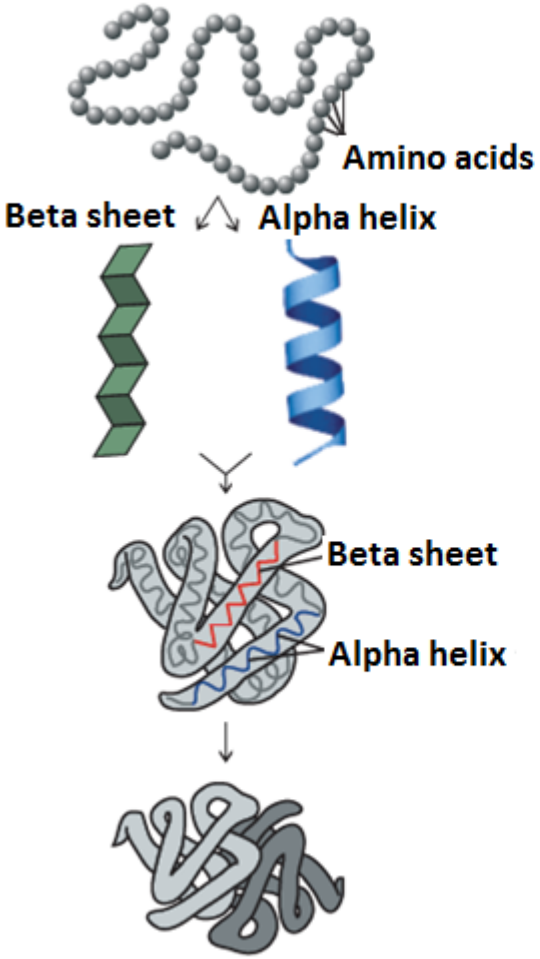
Τριτογενής δομή

- Οι πρωτεΐνες παίρνουν μια βασική θέση μεταξύ των ουσιών που σχετίζονται με τη ζωή όπως τη γνωρίζουμε.
- Οι πρωτεΐνες έχουν μια πολύπλοκη χωρική δομή και εκτελούν μια ποικιλία βιολογικών λειτουργιών - από τυπικά δομικές, προστατευτικές, μεταφορικές έως καταλυτικές και ρυθμιστικές.
- Οι λειτουργίες μιας πρωτεΐνης εξαρτώνται από την τριτοταγή δομή της, η οποία με τη σειρά της εξαρτάται από την πρωτογενή δομή της πρωτεΐνης.
- Η πρόβλεψη της τριτογενούς δομής των πρωτεϊνών από την αλληλουχία αμινοξέων τους είναι ένα από τα σημαντικότερα προβλήματα στον τομέα της βιοπληροφορικής.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Τριτογενής δομή



Primary protein structure
- amino acids sequense

Secondary protein structure
- occurs when a sequence of amino acids is linked by hydrogen bonds

Tertiary protein structure
- occurs when there is an interaction between the alpha helix and the beta sheet

Quaternary protein structure
- a protein containing more than one amino acid chain

Ομόλογη μοντελοποίηση

- Κατασκευή ομόλογου μοντέλου.
  - τεχνικές πρόβλεψης τριτογενούς πρωτεϊνικής δομής.
- Αναζητήστε δομές που έχουν παρόμοιες πτυχές χωρίς εμφανή ομοιότητα στην αλληλουχία πρωτεΐνης.
  - Λαμβάνει την αλληλουχία ερωτήματος (πρωτεΐνη με άγνωστη δομή) και την εκτελεί μέσω των συντεταγμένων της πρωτεΐνης στόχου (πρωτεΐνη με γνωστή δομή), με κρυσταλλογραφία ακτίνων Χ ή MRI (απεικόνιση μαγνητικού συντονισμού).
  - Η ακολουθία κινείται από θέση σε θέση μέσω της δομής, υπακούοντας σε κάποιους προκαθορισμένους φυσικούς περιορισμούς.
    - Τα μήκη των δευτερευόντων δομικών στοιχείων και των κυκλικών περιοχών μπορούν να καθοριστούν ή να μεταβληθούν σε ένα εύρος.

Ομόλογη μοντελοποίηση

- Σε κάθε σύγκριση αλληλουχίας-δομής προσδιορίζονται ζεύγη υδρόφοβων αλληλεπιδράσεων μεταξύ μη τοπικών υπολειμμάτων.
  - Αυτοί οι θερμοδυναμικοί υπολογισμοί χρησιμοποιούνται για τον προσδιορισμό της πιο ενεργειακά ευνοϊκής και διαμορφωτικά σταθερής ευθυγράμμισης της διεκδικούμενης ακολουθίας σύμφωνα με τη δομή στόχο.
- Τα προγράμματα μοντελοποίησης ομολογίας είναι υπολογιστικά εντατικά και απαιτούν ισχυρούς υπολογιστές.
  - Απαιτείται γνώση εξειδικευμένων γλωσσών Η/Υ.
  - Η μέθοδος είναι χρήσιμη όταν οι μέθοδοι πρόβλεψης βασισμένες σε αλληλουχία πρωτεϊνών αποτυγχάνουν να προσδιορίσουν μια κατάλληλη δομή.

Co-funded by the Erasmus+ Programme of the European Union

Ομόλογη μοντελοποίηση

- Οι τεχνικές μοντελοποίησης ομολογίας που χρησιμοποιούν το μοντέλο νήματος είναι πολύ ισχυρά εργαλεία.
  - Οι απαιτήσεις όσον αφορά τόσο το υλικό όσο και την τεχνογνωσία μπορεί να αποδειχθούν εμπόδιο για τους περισσότερους βιολόγους.
- Υπάρχουν εύχρηστα προγράμματα που δίνουν στον μέσο βιολόγο μια καλή πρώτη προσέγγιση για συγκριτική μοντελοποίηση πρωτεϊνών (WHAT-IF και LOOK, παρέχουν προηγμένες δυνατότητες).
  - SWISS-MODEL - ένα πρόγραμμα που εκτελεί αυτοματοποιημένες συγκρίσεις ακολουθιών σε δύο βήματα

Ομόλογη μοντελοποίηση *SWISS-MODEL*

1. Θα πρέπει να καθοριστεί εάν η ακολουθία μπορεί να μοντελοποιηθεί καθόλου.
   - Το SWISS-MODEL συγκρίνει την υποβληθείσα ακολουθία με την κρυσταλλογραφική βάση δεδομένων ExpPdb και επιχειρεί να τη μοντελοποιήσει μόνο εάν υπάρχει ομόλογο της ζητούμενης ακολουθίας στο ExPdb .
   - Επιλέγει δομές προτύπου με τουλάχιστον 25% ταυτότητα αλληλουχίας σε μια περιοχή με περισσότερα από 20 υπολείμματα.
2. Λαμβάνοντας υπόψη μία ή περισσότερες κατάλληλες καταχωρήσεις στο ExpPdb , κατασκευάζονται ατομικά μοντέλα και πραγματοποιείται ελαχιστοποίηση ενέργειας για τη δημιουργία του καλύτερου μοντέλου.
   - Οι συντεταγμένες του ατόμου για το μοντέλο που προκύπτει καθώς και οι δομικές ευθυγραμμίσεις επιστρέφονται ως μήνυμα ηλεκτρονικού ταχυδρομείου.

Co-funded by the Erasmus+ Programme of the European Union

Ομόλογη μοντελοποίηση *SWISS-MODEL*

```
TARGET      1          QRRQ RTHFTSQQLQ QLEATFQRNR YPDMSTRERI AVWTNLTEAR
11PJL       0        KQRRS RTTFSASQLD ELERAFERTQ YPDIYTREEL AQRTNLTEAR
21PJL       1          QRRS RTTFSASQLD ELERAFERTQ YPDIYTREEL AQRTNLTEAR
11B72       1   ARTFDWMKVL RTNFTTRQLT ELEKEFHFNK YLSRARRVEI AATLELNETQ
22HDD       1          KRP RTAFSSEQLA RLKREFNENR YLTERRRQQL SSELGLNEAQ
12HOA       0        MRKRG RQTYTRYQTL ELEKEFHFNR YLTRRRRIEI AHALSLTERQ
                           .  *  ..  *    *   *     *     * ..  .    * * .

TARGET
11PJL                        hhhhhh hhhhhhhhh     hhhhhhhh hhhhh hhhh
21PJL                        hhhhhh hhhhhhhhh     hhhhhhhh hhhhh hhhh
11B72                        hhhhhh hhhhhhhhh     hhhhhhhh hhhhh hhhh
22HDD                        hhhhhh hhhhhhhhh     hhhhhhhh hhhhh hhhh
12HOA                        hhhhhh hhhhhhhhh     hhhhhhhh hhhh  hhhh
```

Η αλληλουχία εισόδου είναι η περιοχή ομοιοπεδίου της ανθρώπινης πρωτεΐνης PITX2

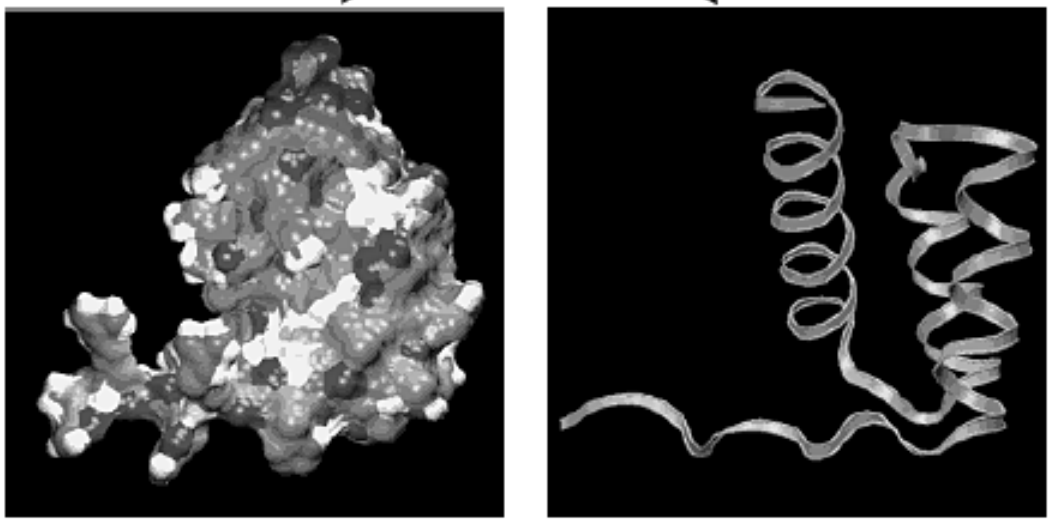
```
ATOM      1  H1  GLN     1       9.226 107.177  13.966  1.00 99.00
ATOM      2  H2  GLN     1      10.769 107.671  13.751  1.00 99.00
ATOM      3  N   GLN     1       9.824 107.785  13.444  1.00 25.00
ATOM      4  H3  GLN     1       9.549 108.738  13.592  1.00 99.00
ATOM      5  CA  GLN     1       9.728 107.473  11.999  1.00 25.00
ATOM      6  CB  GLN     1       8.265 107.520  11.538  1.00 25.00
ATOM      7  CG  GLN     1       7.468 106.270  11.932  1.00 25.00
ATOM      8  CD  GLN     1       8.001 104.970  11.312  1.00 25.00
ATOM      9  OE1 GLN     1       8.748 104.928  10.343  1.00 25.00
ATOM     10  NE2 GLN     1       7.629 103.853  11.899  1.00 25.00
ATOM     11  HE21GLN     1       7.979 103.008  11.502  1.00 99.00
ATOM     12  HE22GLN     1       7.015 103.860  12.683  1.00 99.00
```

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ομόλογη μοντελοποίηση
SWISS-MODEL

```
ATOM      1  H1  GLN     1       9.226 107.177  13.966  1.00 99.00
ATOM      2  H2  GLN     1      10.769 107.671  13.751  1.00 99.00
ATOM      3  N   GLN     1       9.824 107.785  13.444  1.00 25.00
ATOM      4  H3  GLN     1       9.549 108.738  13.592  1.00 99.00
ATOM      5  CA  GLN     1       9.728 107.473  11.999  1.00 25.00
ATOM      6  CB  GLN     1       8.265 107.520  11.538  1.00 25.00
ATOM      7  CG  GLN     1       7.468 106.270  11.932  1.00 25.00
ATOM      8  CD  GLN     1       8.001 104.970  11.312  1.00 25.00
ATOM      9  OE1 GLN     1       8.748 104.928  10.343  1.00 25.00
ATOM     10  NE2 GLN     1       7.629 103.853  11.899  1.00 25.00
ATOM     11  HE21GLN     1       7.979 103.008  11.502  1.00 99.00
ATOM     12  HE22GLN     1       7.015 103.860  12.683  1.00 99.00
```



Co-funded by the Erasmus+ Programme of the European Union

Ομόλογη μοντελοποίηση *UCLA*

- Μια άλλη αυτοματοποιημένη μέθοδος για την πρόβλεψη της αναδίπλωσης πρωτεΐνης περιλαμβάνει:
  1. Εύρεση αντιστοιχίσεων που βασίζονται σε ακολουθία για να εκχωρήσετε ένα πιθανό πάσο στην ζητούμενη ακολουθία.
  2. Πρόσθετη πρόβλεψη δευτερευουσών δομικών πληροφοριών για την ζητούμενη ακολουθία.
- Η σωστή αντιστοίχιση πτυχών εξαρτάται από τα βαθμολογημένα αποτελέσματα που λαμβάνονται για την ζητούμενη ακολουθία.
  - Αυτά τα αποτελέσματα βασίζονται στη συμβατότητα της ζητούμενης ακολουθίας με καθεμία από τις δομές σε μια βιβλιοθήκη τρισδιάστατων δομών στόχου.

Co-funded by the Erasmus+ Programme of the European Union

Ομόλογη μοντελοποίηση *UCLA*

- Περαιτέρω πρόβλεψη της δευτερεύουσας δομής βελτιώνει την κατανομή αναδίπλωσης κατά περίπου 25%.
- Τα δεδομένα εισόδου για αυτήν τη μέθοδο είναι μια αλληλουχία πρωτεΐνης που υποβάλλεται μέσω της διεπαφής ιστού.
- Ανοίγει μια ιστοσελίδα που περιέχει τα αποτελέσματα.
  - Τα αποτελέσματα αποθηκεύονται φυσικά στον διακομιστή UCLA για μελλοντική αναφορά.

Co-funded by the Erasmus+ Programme of the European Union

Ομόλογη μοντελοποίηση
DALI

- Συγκρίνει δομές με κατασκευές.
- Ο αλγόριθμος αναζητά παρόμοια μοτίβα επαφής μεταξύ δύο πρωτεϊνών, εκτελεί μια βελτιστοποίηση και επιστρέφει το καλύτερο σύνολο λύσεων για να ευθυγραμμίσει τις δομές των πρωτεϊνών ερωτήματος (Holm and Sander, 1993).
  - Τα κενά μπορεί να είναι οποιουδήποτε μήκους και επιτρέπουν εναλλακτικές συνδέσεις μεταξύ ευθυγραμμισμένων τμημάτων, διευκολύνοντας έτσι τον εντοπισμό συγκεκριμένων περιοχών που είναι παρόμοιες σε δύο διαφορετικές πρωτεΐνες, ακόμα κι αν οι πρωτεΐνες είναι γενικά διαφορετικές.

Co-funded by the Erasmus+ Programme of the European Union

Ομόλογη μοντελοποίηση *DALI*

- Η διαδικτυακή διεπαφή DALI πραγματοποίησε την ανάλυση σε δύο σύνολα συντεταγμένων, χρησιμοποιώντας ένα σύνολο συντεταγμένων που παρέχεται από τον χρήστη σε μορφή PDB.

- Εάν οι ζητούμενες πρωτεΐνες είναι διαθέσιμες στο PDB, τότε οι προηγουμένως υπολογισμένοι δομικοί γείτονές τους μπορούν να βρεθούν με πρόσβαση στη βάση δεδομένων FSSP
  - Το FSSP περιέχει δομικά ευθυγραμμισμένες οικογένειες αναδιπλωμένων πρωτεϊνών (Holm and Sander, 1994) που ελήφθησαν από μια σύγκριση όλων κατά όλων των καταχωρήσεων PDB.

Ομόλογη μοντελοποίηση *TOPITS*

- TOPITS (Rost , 1995)
  1. Δημιουργεί μια βάση δεδομένων με δυνατότητα αναζήτησης μετατρέποντας την τρισδιάστατη δομή των πρωτεϊνών PDB σε μονοδιάστατες «χορδές» δευτερογενούς δομής.
  2. Η δευτερεύουσα δομή και η προσβασιμότητα του διαλύτη για την ζητούμενη ακολουθία προσδιορίζονται με τη μέθοδο PHD (τα αποτελέσματα αποθηκεύονται ως μονοδιάστατη συμβολοσειρά).
  3. Τόσο η συμβολοσειρά ερωτήματος όσο και η συμβολοσειρά στόχου ευθυγραμμίζονται με δυναμικό προγραμματισμό για την πρόβλεψη της δομής της πρωτεΐνης ερωτήματος.
- Τα αποτελέσματα επιστρέφονται ως λίστα κατάταξης που δείχνει τη βέλτιστη στοίχιση της ζητούμενης ακολουθίας έναντι της δομής στόχου.
  - Περιέχει επίσης μια βαθμολογία πιθανότητας (Z-score) για την ακρίβεια της πρόβλεψης.

Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

- Ένας μεγάλος αριθμός επιστημόνων προσπαθούν να αναπτύξουν μαθηματικά μοντέλα και αλγόριθμους για να λύσουν το λεγόμενο «Πρόβλημα Αναδίπλωσης Πρωτεϊνών».
  - Να προβλέψει την τρισδιάστατη δομή των πρωτεϊνών από πληροφορίες βασικής αλληλουχίας αμινοξέων.
- Τα τελευταία χρόνια, οι επιστήμονες έχουν κάνει μεγάλα βήματα στην ανάπτυξη μεθόδων για την πρόβλεψη της τριτογενούς δομής των πρωτεϊνών χρησιμοποιώντας υπολογιστή.
- Οι τρέχουσες μέθοδοι που βασίζονται στη βιοπληροφορική είναι σε θέση να προβλέψουν την αναδίπλωση μικρών σφαιρικών πρωτεϊνών, οι οποίες είναι ομόλογες πρωτεΐνες με γνωστή δομή.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

Protein Structure Prediction

Christian Anfinsen, 1961:

denatured RNase refolds into functional state (in vitro)

$\Rightarrow$ no external folding machinery

$\Rightarrow$ Anfinsen's dogma/thermodynamic hypthesis:
all information about native structure is in the sequence
(at least for small globular proteins)

native structure = minimum of the free energy

- unique
- stable
- kinetically accessible

S.Will, 18.417, Fall 2011

Πρωτεΐνη πτυσσόμενος πρόβλημα

- Το πρόβλημα αναδίπλωσης πρωτεΐνης μπορεί να χωριστεί σε τρία διακριτά υποπροβλήματα :
  1. **αναδιπλούμενος κώδικας** - η ερώτηση *"Πώς προέρχεται η φυσική δομή της πρωτεΐνης από τις ενδοατομικές δυνάμεις που δρουν στην αλληλουχία αμινοξέων;"* είναι θερμοδυναμική.
     - Η αλληλουχία πρωτεΐνης πρέπει να έχει μια μοναδικά διπλωμένη διαμόρφωση που να είναι σταθερή και συνεπής με τη φυσική δομή
  2. **Πρόβλεψη δομής πρωτεΐνης -** ένα υπολογιστικό πρόβλημα του τρόπου πρόβλεψης της φυσικής δομής μιας πρωτεΐνης από την αλληλουχία αμινοξέων της.
  3. **ρυθμός αναδίπλωσης (παράδοξο του Levitall) -** πώς μπορεί η πρωτεΐνη να διπλωθεί γρήγορα.
- ***Έχει αποδειχθεί ότι ακόμη και σε απλοποιημένα μοντέλα, το πρόβλημα αναδίπλωσης πρωτεΐνης σε 2D και 3D είναι NP-complete.***

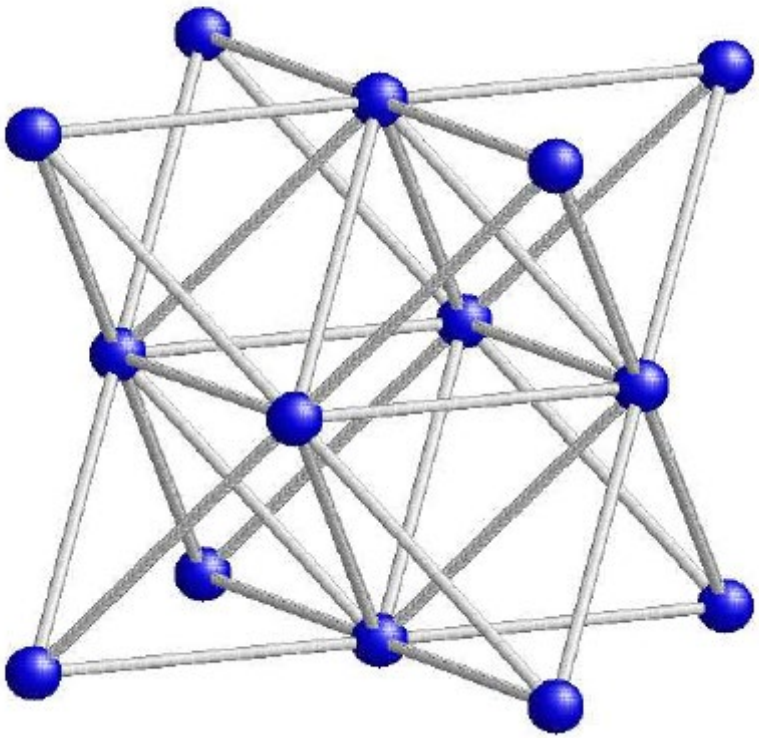
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

Μοντέλο εκτός πλέγματος

$$FCC = \{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{Z}^3 \mid x + y + z \text{ even} \}$$

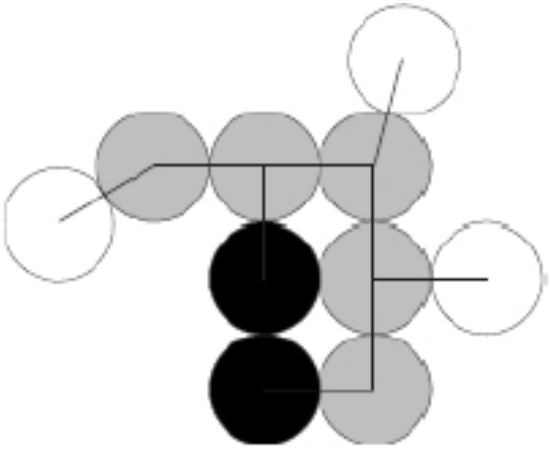


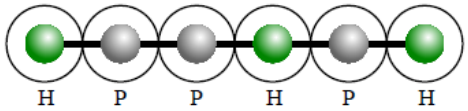
- Ένα από τα σημαντικά ερωτήματα που σχετίζονται με τη μελέτη των μοντέλων αναδίπλωσης πρωτεϊνών είναι εάν οι αλγόριθμοι για μοντέλα πλέγματος μπορούν να γενικευθούν σε αλγόριθμους για μοντέλα εκτός δικτύου .
  - Το 1611. Ο Κέπλερ διατύπωσε μια υπόθεση που δηλώνει ότι η υψηλότερη πυκνότητα συσσώρευσης πανομοιότυπων σφαιρών επιτυγχάνεται με τη διάταξη των σφαιρών σε ένα πλέγμα Face-Center-Cubic (FCC).

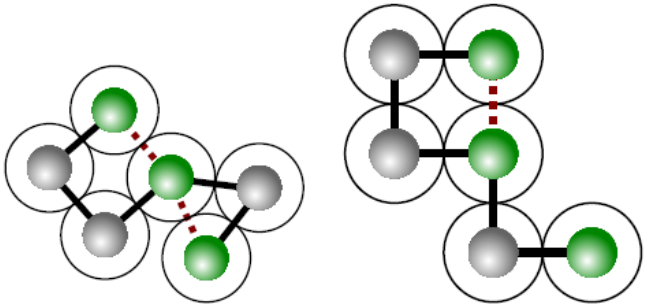
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

Μοντέλο εκτός πλέγματος







- Το 1997, ο Thomas Hales έδειξε ότι η πιο πυκνή διάταξη αμινοξέων εκτός πλέγματος θα μπορούσε να γίνει στο πλέγμα FCC.
- Το 1997, οι Hart και Istrail παρουσιάζουν ένα μοντέλο εκτός πλέγματος που ονομάζεται *μοντέλο HP με πλευρική αλυσίδα εφαπτομένης-σφαιρικής.*
  - γειτονικά μόρια βάσης και πλευρικής αλυσίδας αντιπροσωπεύονται από εφαπτόμενες σφαίρες ίδιου μεγέθους στον τρισδιάστατο χώρο.
  - Οι πλευρικές αλυσίδες συμβολίζονται ως υδρόφοβες (H) ή υδρόφιλο (P) , και η διαμορφωτική ενέργεια είναι ο αριθμός των HH εφαπτόμενες σφαίρες.

Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

Μοντέλο εκτός πλέγματος

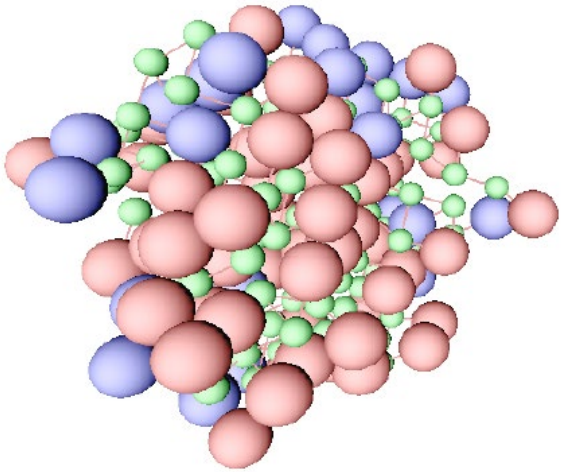
- Ο κατά προσέγγιση αλγόριθμος για το μοντέλο πλευρικής αλυσίδας στο πλέγμα FCC μπορεί να εισαχθεί στο πλαίσιο εκτός πλέγματος και έτσι να παρέχει εγγύηση απόδοσης εκτός πλέγματος που πλησιάζει τη βέλτιστη (86% της βέλτιστης).
- Ο αριθμός των γειτόνων στο πλέγμα FCC και ο μέγιστος αριθμός πιθανών γειτόνων στο χώρο είναι ο ίδιος. Αυτό υποδηλώνει ότι οι αλγόριθμοι πλέγματος FCC έχουν καλή απόδοση εκτός πλέγματος .
- Τα παρακάτω σχήματα δείχνουν την αναδίπλωση σε ένα 2D εξαγωνικό (τριγωνικό) πλέγμα, ένα 3D κυβικό πλέγμα γραμμικής αλυσίδας και ένα πλέγμα FCC πλευρικής αλυσίδας 3D.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

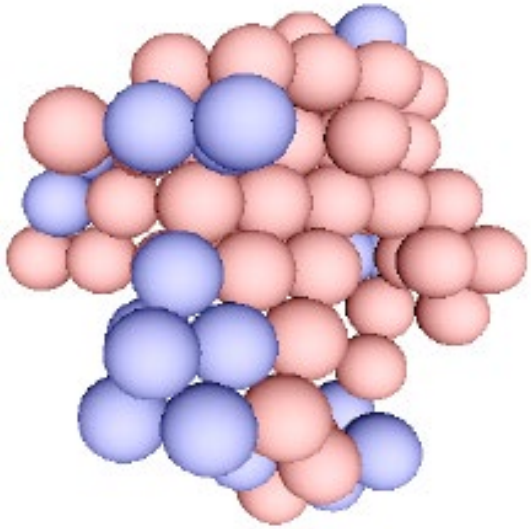
Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

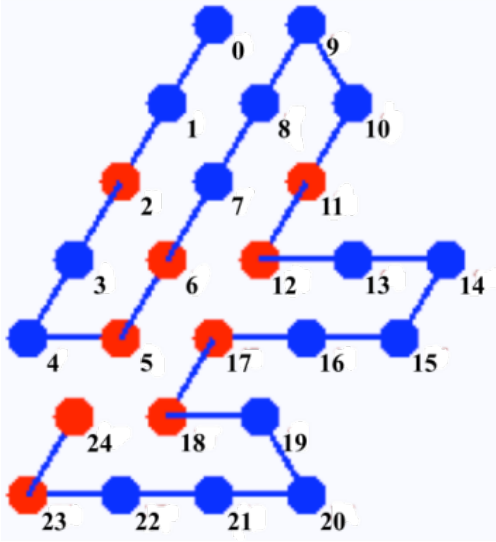
Μοντέλο εκτός πλέγματος



Πτυσσόμενο μοντέλο πλευρικής αλυσίδας από πλέγμα FCC



Διπλώνοντας μια πρωτεΐνη μήκους 85 σε ένα τρισδιάστατο πλέγμα FCC

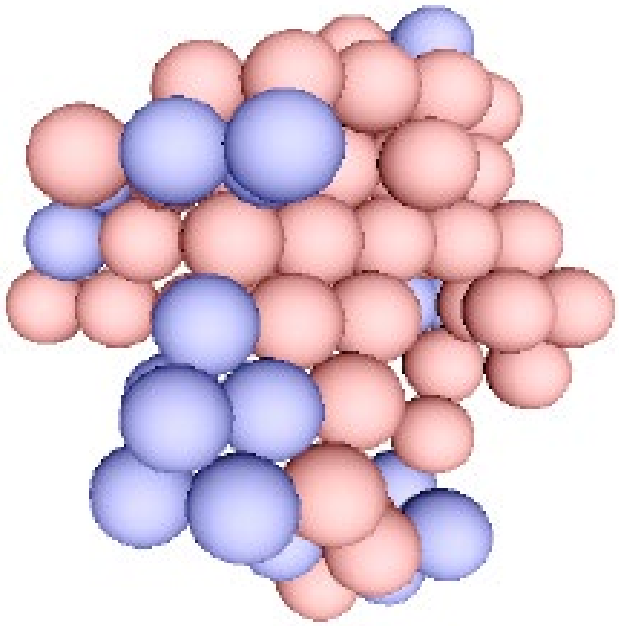


Αναδίπλωση σε 2D εξαγωνικό (τριγωνικό) πλέγμα για πρωτεΐνη μήκους 25

Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

Μοντέλα πλέγματος



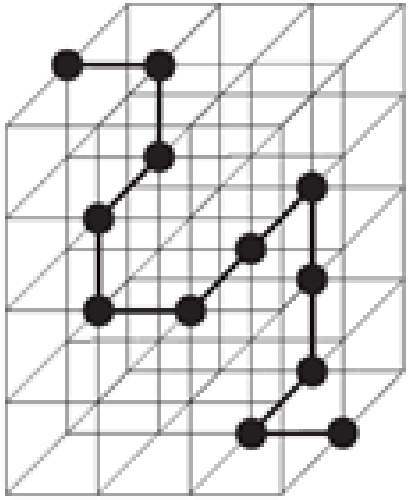
- Υπάρχουν δύο τύποι μοντέλων πλέγματος:
  - Ο πρώτος τύπος περιγράφει τις βασικές φυσικές δυνάμεις που εμπλέκονται στη διαδικασία αναδίπλωσης των πρωτεϊνών.
  - Ο δεύτερος τύπος σχεδιάστηκε για να δημιουργήσει μια ρεαλιστική αναδίπλωση πραγματικών πρωτεϊνών και να παραμετροποιήσει τη χρήση πρωτεϊνών ως προτύπων για υπάρχουσες δομές.
- Στον τρισδιάστατο χώρο, το πλέγμα Face-Center-Cubic (FCC) είναι πολύ σημαντικό.
  - Στο πλέγμα FCC, μπορεί να επιτευχθεί η πιο πυκνή συσκευασία πανομοιότυπων σφαιρών σε τρισδιάστατο χώρο.

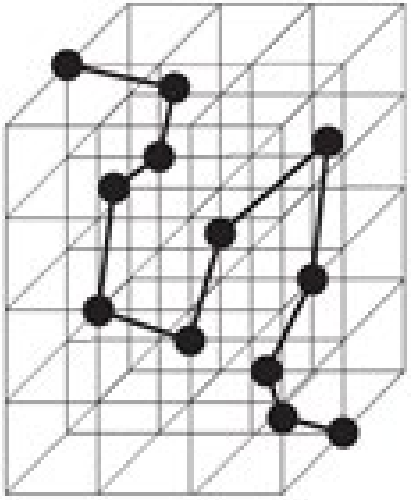
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

Μοντέλα πλέγματος

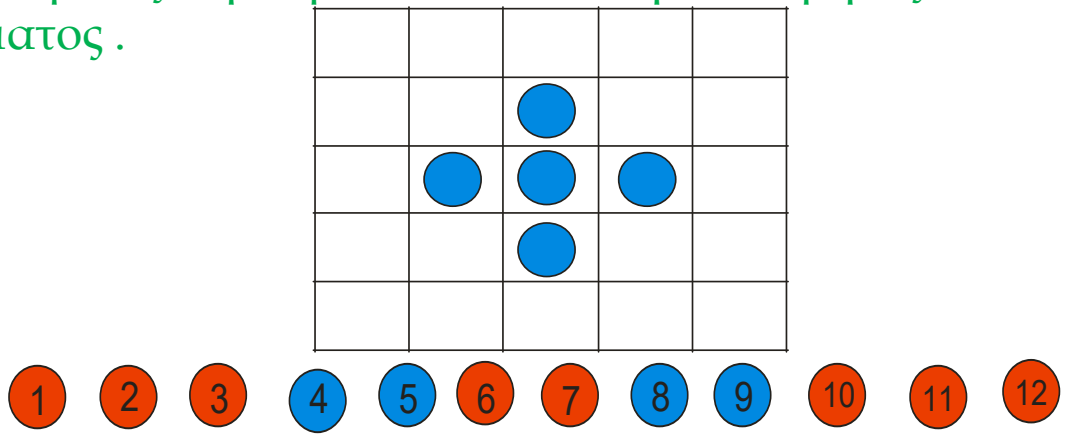




- Το μοντέλο πλέγματος για το πρόβλημα αναδίπλωσης πρωτεΐνης τοποθετεί κάθε αμινοξύ σε ένα σημείο σε ένα 2D τεταρτημόριο ή ένα 3D κυβικό πλέγμα
  - Τα γειτονικά αμινοξέα στην πρωτεϊνική αλληλουχία πρέπει να καταλαμβάνουν γειτονικά σημεία στο πλέγμα προκειμένου να διατηρηθεί η συνδεσιμότητα της πρωτεΐνης.
- Το μοντέλο του πλέγματος δίνει :
  - ίσα μεγέθη αμινοξέων ;
  - ίσα μήκη δεσμών μεταξύ όλων των ζευγών αμινοξέων .
  - ορθογώνια ή γραμμική γωνία, και για τα τρία διαδοχικά αμινοξέα .

Πρόβλημα αναδίπλωσης πρωτεΐνης
Μοντέλα πλέγματος – ορισμός προβλήματος

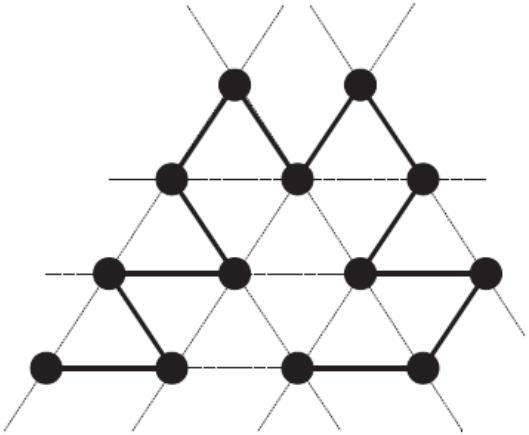
- Ο σκοπός είναι να τοποθετηθούν όλα τα αμινοξέα της πρωτεΐνης σε ένα πλέγμα 2D και 3D ως :
  - Τα παρακείμενα αμινοξέα στην πρωτεϊνική αλληλουχία πρέπει να καταλαμβάνουν γειτονικά κύτταρα στο πλέγμα .
  - Μπορούμε να βάλουμε μόνο ένα αμινοξύ σε κάθε κύτταρο πλέγματος .
  - Κάθε αμινοξύ μπορεί να καταλάβει ακριβώς ένα κύτταρο του πλέγματος .



Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης
Μοντέλα πλέγματος – ορισμός προβλήματος

- Λόγω της ιδιότητας του μοντέλου τετραγωνικού ή κυβικού πλέγματος, δύο αμινοξέα που απέχουν ίσο μήκος μεταξύ τους δεν μπορούν να είναι γειτονικά, κάτι που δεν ισχύει απαραίτητα στον πραγματικό κόσμο.
  - Ως εκ τούτου, έχουν εισαχθεί πολλά διαφορετικά μοντέλα πλέγματος, όπως το μοντέλο τριγωνικού πλέγματος και το μοντέλο διαγώνιου πλέγματος.



Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

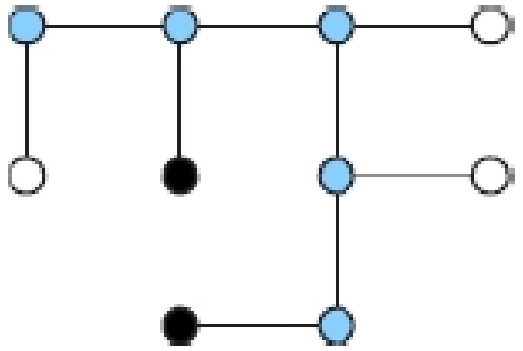
Μοντέλα πλέγματος

Γραμμή-αλυσίδα

- Hart και Istrail καθώς και ο αλγόριθμος του Newman είναι κατά προσέγγιση αλγόριθμοι για μοντέλα αναδίπλωσης πρωτεϊνών γραμμικής αλυσίδας.
- Επίσης, οι Hart και Eastrail αναπτύσσουν έναν βασικό αλγόριθμο προσέγγισης ως γενική μέθοδο για την αναδίπλωση πρωτεϊνών σε μοντέλα HP.
  - Αυτός ο αλγόριθμος μπορεί να αναφέρεται σε μια μεγάλη κατηγορία μοντέλων πλέγματος.

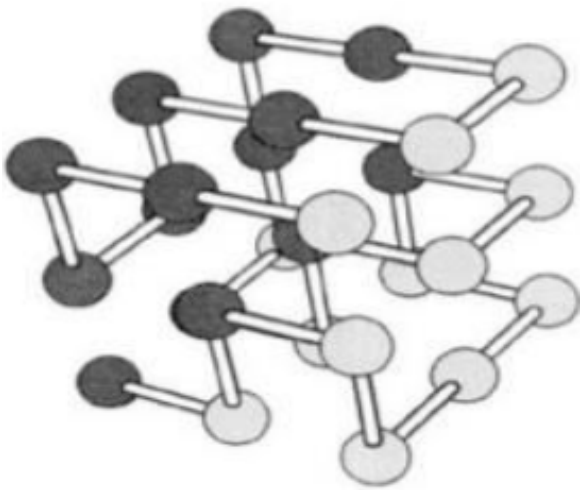
Πλευρική αλυσίδα

- Προκειμένου να βελτιωθεί η ακρίβεια των μεθόδων πρόβλεψης πρωτεϊνικής δομής, εκτεταμένα μοντέλα θεωρούν την πρωτεΐνη ως έναν πυρήνα που σχηματίζεται από πολλαπλούς πεπτιδικούς δεσμούς που συνδέονται με πλευρικές αλυσίδες.



Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης μοντέλο HP Lattice



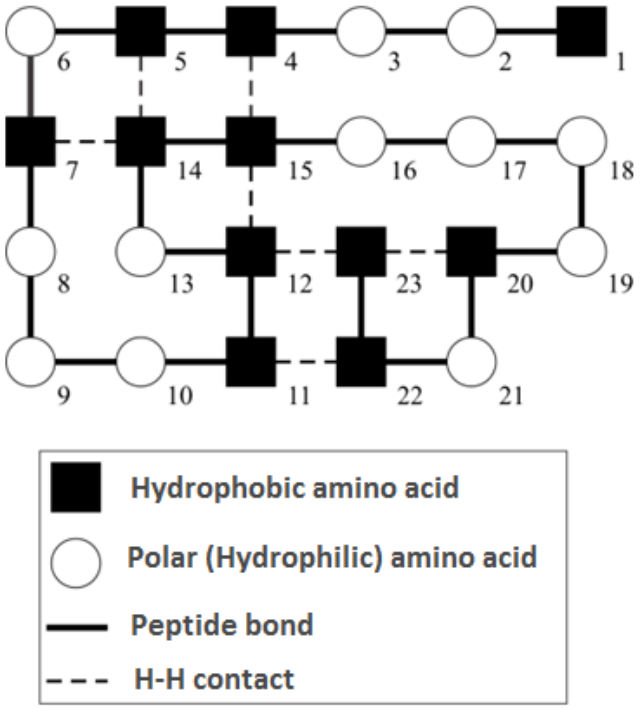
- Το 1985, ο Ken Dill πρότεινε το μοντέλο Υδροφοβικό-Υδρόφιλο (HP)
  - Το μοντέλο HP βασίζεται στην παρατήρηση ότι οι υδρόφοβοι δεσμοί μεταξύ αμινοξέων οδηγούν τις διαδικασίες αναδίπλωσης των πρωτεϊνών.
- Η υδρόφοβη δράση των αμινοξέων παίρνει ένα σημαντικό μέρος της συνολικής ενεργειακής συνάρτησης, επομένως αυτή είναι η πιο σημαντική δύναμη στον προσδιορισμό της δομής της πρωτεΐνης.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης μοντέλα HP Lattice

Ονομα	Σύμβολο	Ταξινόμηση	Ονομα	Σύμβολο	Ταξινόμηση
Αλανίνη	ΕΝΑ	Υδροφόβος	Λευκίνη	μεγάλο	Υδροφόβος
Αργινίνη	R	Πολικός	Λυσίνη	κ	Πολικός
Ασπαραγίνη	Ν	Πολικός	Μεθειονίνη	Μ	Υδροφόβος
Ασπαρτικό Οξύ	ρε	Πολικός	Φαινυλαλανίνη	φά	Υδροφόβος
Κυστεΐνη	ντο	Πολικός	Προλίνη	Π	Υδροφόβος
Γλουταμινικό Οξύ	μι	Πολικός	Σερίνη	μικρό	Πολικός
Γλουταμίνη	Q	Πολικός	Θρεονίνη	Τ	Πολικός
Γλυκίνη	σολ	Πολικός	Τρυπτοφάνη	W	Υδροφόβος
Ιστιδίνη	Η	Πολικός	Τυροσίνη	Υ	Πολικός
Ισολευκίνη	Εγώ	Υδροφόβος	Βαλίν	V	Υδροφόβος

Πρόβλημα αναδίπλωσης πρωτεΐνης μοντέλα HP Lattice



- Το μοντέλο πλέγματος HP για δίπλωμα πρωτεΐνης μπορεί να περιγραφεί ως εξής:

Μεγιστοποίηση :

Ο αριθμός των επαφών HH .

Θέμα προς :

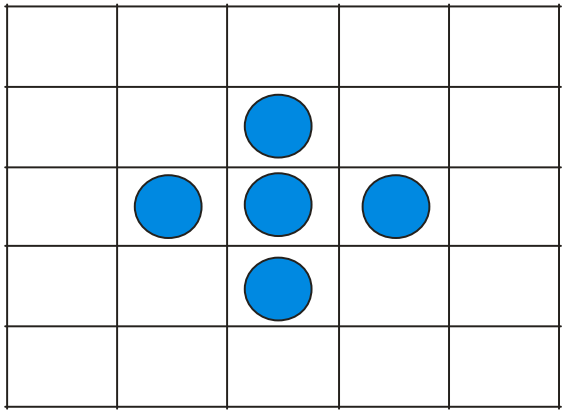
1. (*Εργασία*) Κάθε αμινοξύ μπορεί να καταλάβει ακριβώς ένα κύτταρο στο πλέγμα .
2. (*Μη επικαλυπτόμενο*) Μπορούμε να βάλουμε μόνο ένα αμινοξύ σε κάθε κύτταρο πλέγματος .
3. (*Συνδεσιμότητα*) Τα αμινοξέα που είναι γειτονικά στην πρωτεϊνική αλληλουχία πρέπει να καταλαμβάνουν γειτονικά κύτταρα στο πλέγμα .

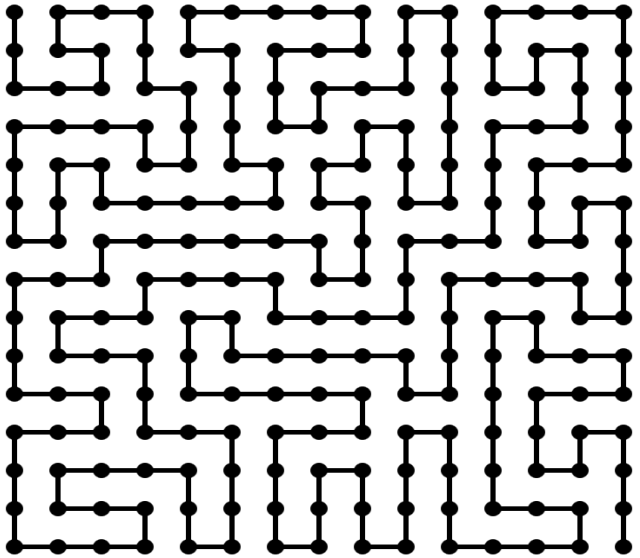
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης μοντέλα HP Lattice

- Μια αυτο-αποφυγή βόλτα είναι μια διαδρομή σε ένα γράφημα από ένα σημείο (κορυφή) σε ένα άλλο που δεν διέρχεται από το ίδιο σημείο (κορυφή) περισσότερες από μία φορές.





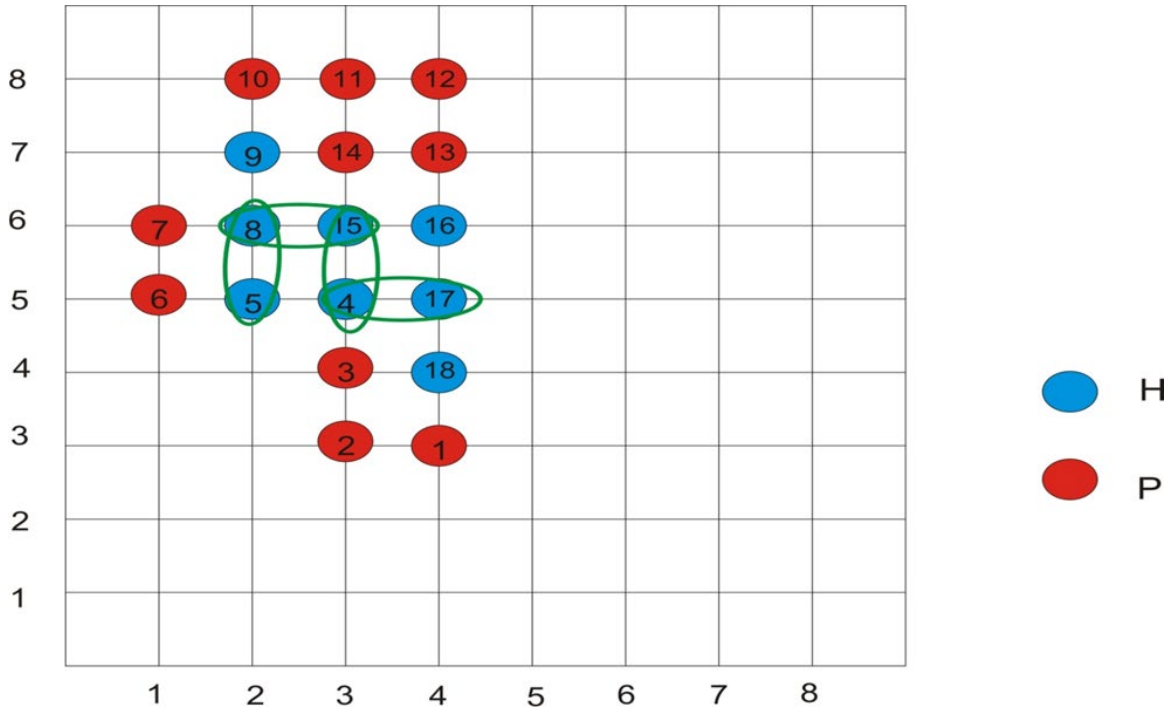
Πρόβλημα αναδίπλωσης πρωτεΐνης μοντέλα HP Lattice

- Ο αριθμός των αυτοαποφευγόμενων μονοπατιών μεταξύ δύο διαγώνιων κορυφών σε ένα πλέγμα μεγέθους *n* x *n* είναι:

N x *N*	**Броя пътища**
1	2
2	12
3	184
4	184
5	1262816
6	575780564
7	789360053252
8	3266598486981642
9	41044208702632496804
10	1568758030464750013214100
11	182413291514248049241470885236
12	645280393432700189633571851584821 18
13	69450664761521361664274701548907358996488
14	227449714676812739631826459327989863387613323440
15	22667455688626727463745673967130989348663248854083190 28
16	68745445609149931587631563132489232824587945968099457285419306
17	63448146112379639713102975407955244004494439868664806936463693878553 36
18	17821128408420651298933849466523252751678380657047676559314524746058266927825 32
19	15233449717048799930807428103192296908994542553232945557760298667373550605928775692558 44
20	10^{97}
21	10^{107}

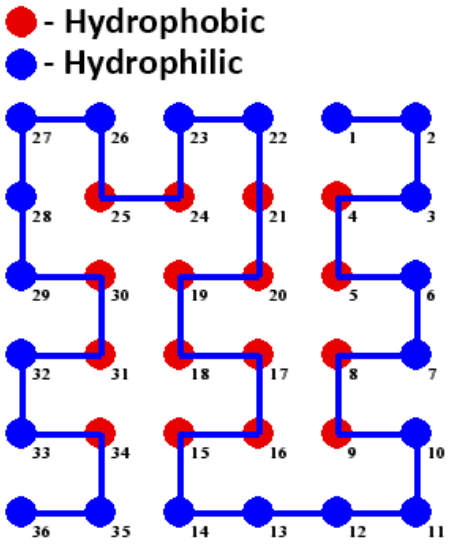
Πρόβλημα αναδίπλωσης πρωτεΐνης μοντέλα HP Lattice

- Ο κύριος σκοπός αυτού του μοντέλου είναι να μεγιστοποιήσει τον αριθμό των επαφών μεταξύ των υδρόφοβων αμινοξέων.

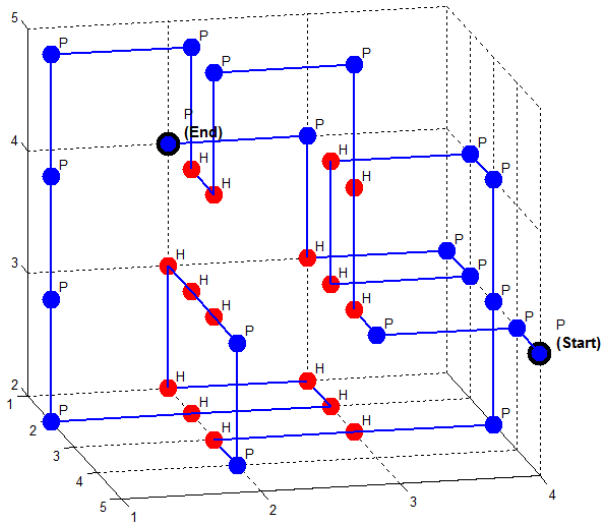


Πρόβλημα αναδίπλωσης πρωτεΐνης μοντέλα HP Lattice

- Η βέλτιστη διαμόρφωση κατά την αναδίπλωση μιας πρωτεΐνης στο μοντέλο HP είναι η διαμόρφωση που έχει τον μέγιστο αριθμό επαφών μεταξύ των υδρόφοβων αμινοξέων, γεγονός που δίνει τη χαμηλότερη ενεργειακή αξία της πρωτεΐνης.



14 επαφές



18 επαφές

Co-funded by the Erasmus+ Programme of the European Union

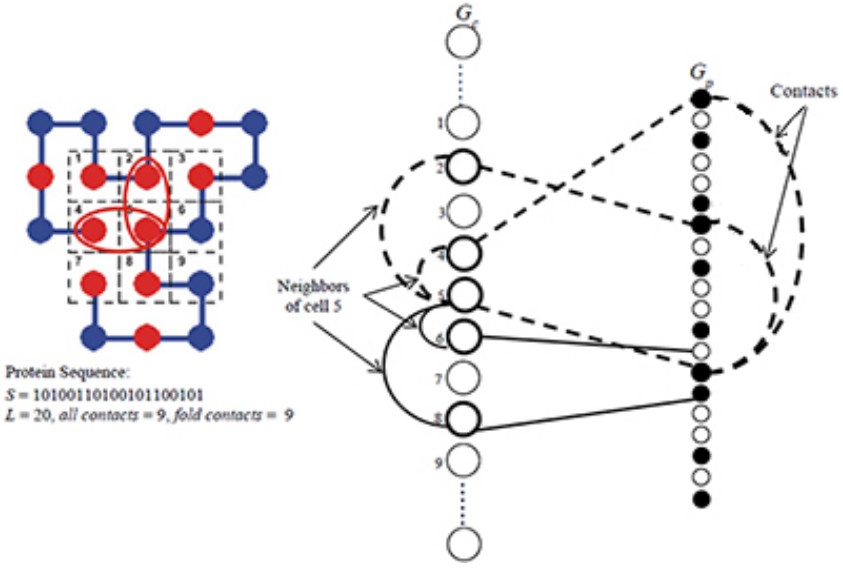
Πρόβλημα αναδίπλωσης πρωτεΐνης
Αλγόριθμοι για την επίλυση του προβλήματος

- Οι συνδυαστικές μέθοδοι που χρησιμοποιούνται για την επίλυση του προβλήματος της αναδίπλωσης των πρωτεϊνών είναι οι κατά προσέγγιση αλγόριθμοι και οι αλγόριθμοι συνδυαστικής βελτιστοποίησης.
  - Οι κατά προσέγγιση αλγόριθμοι είναι αλγόριθμοι πολυωνυμικού χρόνου που, για κάθε αλληλουχία πρωτεΐνης, δημιουργούν μια αναδίπλωση για την οποία ο αριθμός των επαφών είναι σχεδόν βέλτιστος.
  - Οι αλγόριθμοι συνδυαστικής βελτιστοποίησης είναι εκθετικοί αλγόριθμοι που, για ορισμένες αλληλουχίες πρωτεϊνών, δημιουργούν μια αναδίπλωση για την οποία ο αριθμός των επαφών μπορεί να αποδειχθεί βέλτιστος.
- Οι ευρετικοί αλγόριθμοι είναι η τρίτη κατηγορία συνδυαστικών μεθόδων.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

Αλγόριθμοι για την επίλυση του προβλήματος



- Βασικά στοιχεία για την επίλυση του προβλήματος είναι :
  - Ένα υποπλέγμα ;
  - Ακολουθία HP .
- Ένα πρόβλημα γραφήματος που μπορεί να λυθεί, είτε με βελτιστοποίηση ακεραίων είτε με τη θεωρία γραφημάτων .
- Στη δισδιάστατη περίπτωση το υποπλέγμα είναι ένα τετράγωνο και στην περίπτωση 3Δ είναι ένας κύβος με κορυφές χρωματισμένες μαύρο (το σύνολο V_b) και λευκό (το σύνολο V_w).

Πρόβλημα αναδίπλωσης πρωτεΐνης
Αλγόριθμοι για την επίλυση του προβλήματος σε 2D

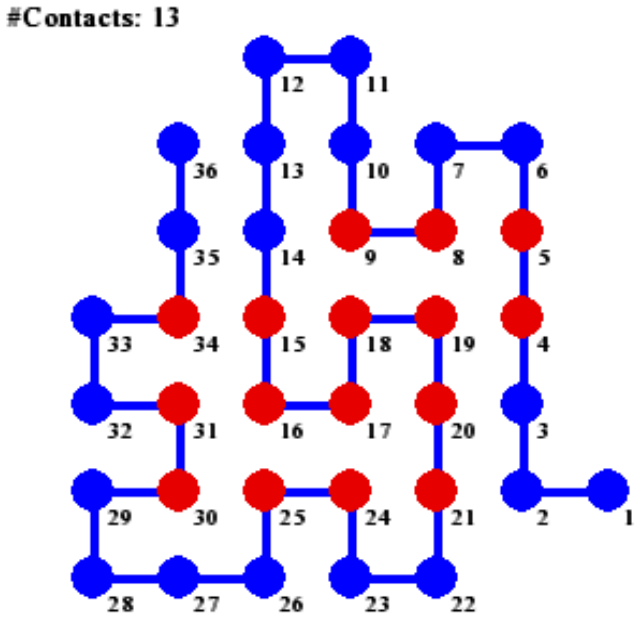
- Χωρίζουμε την ακολουθία S σε τμήματα προκαθορισμένου μεγέθους.
- Παίρνουμε το *i* - ο μέρος του S και δημιουργούμε όλες τις πιθανές πτυχές. Επιλέγουμε την πτυχή με τον μέγιστο αριθμό επαφών και την τοποθετούμε στο τετράγωνο πλέγμα.
- Προσθέστε *i* +1 μέρος του S στην ήδη ληφθείσα πτυχή, βρείτε όλες τις πιθανές πτυχές σε σχέση με την ήδη επιλεγμένη πτυχή.
- Από τις νέες πτυχές που αποκτήθηκαν, επιλέξτε αυτή με τον μέγιστο αριθμό επαφών και τοποθετήστε την στο τετράγωνο πλέγμα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

Αλγόριθμοι για την επίλυση του προβλήματος σε 2D

- Ο αλγόριθμος σταματά σε δύο περιπτώσεις :
  - Όλες οι πιθανές πτυχές έχουν εξαντληθεί.
  - Βρήκαμε ένα δίπλωμα στο οποίο ο αριθμός των επαφών είναι μέγιστος ($C\_2D\ (S)$=2 λεπτά{$|ODD|, |EVEN|$}).



Αναδίπλωση πρωτεΐνης με μήκος 36 αμινοξέων, κατασκευασμένο από ευρετικό αλγόριθμο

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

Αλγόριθμοι για την επίλυση του προβλήματος σε 2D

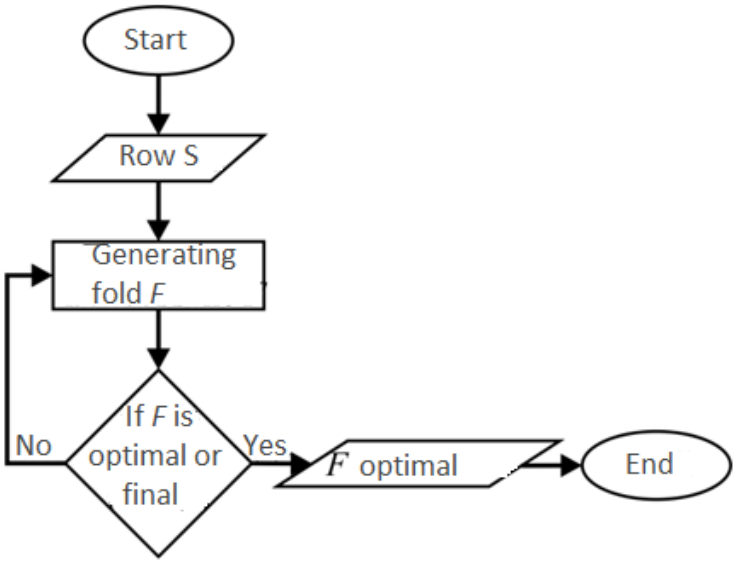
Algorithm GREEDYLIKE:

```
1:  for i in {1, 2, ..., k} do
2:      rec := 0
3:      PATHFINDER(i, 0)
4:  end for
def PATHFINDER(i, m)    / Generates all substrings with a defined length
5:  if m > len(Si) then
6:      if cont(Si,m) > rec then          / * cont() counts the contacts for
7:          rec := cont(Si,m)               the obtained conformation. */
8:            Saves Si,m
9:      end if
10:     return
11: end if
12: for a in {s, r, l} do
13:     st[m] := a
14:     if check(Si,m) then          / * check() checks whether the resulting
15:         PATHFINDER(i, m+1)         conformation is a self-avoiding path. */
16:     end if
17: end for
```

Πρόβλημα αναδίπλωσης πρωτεΐνης

Αλγόριθμοι για την επίλυση του προβλήματος σε 2D

- Κάθε πτυχή ξεκινά με 2 αμινοξέα στερεωμένα στο κέντρο ενός τετράγωνου πλέγματος.
- Από κάθε σταθερό κελί μπορείτε να πάτε αριστερά (-1), δεξιά (+1), κάτω (-N) και πάνω (+N).



Πρόβλημα αναδίπλωσης πρωτεΐνης

Αλγόριθμοι για την επίλυση του προβλήματος σε 3D

- Τα κύρια συστατικά των αλγορίθμων που περιγράφονται παρακάτω είναι:
  - Δημιουργία του μοντέλου - δημιουργήστε ένα . *lp* αρχείο για το μοντέλο
  - Επίλυση MIP (Mixed Integer Programming) ως CPLEX ή GUROBI
- Η κοινή είσοδος είναι ένα κυβικό υποπλέγμα $d_1 \geq X \geq 0$, $d_2 \geq y \geq 0$, $d_3 \geq z \geq 0$ συνδεδεμένο με το μήκος n της ακολουθίας HP εισόδου S . Για τους ευρετικούς αλγόριθμους **HE-1** και **HE-2** , διαιρούμε *το* S σε τμήματα προκαθορισμένου μήκους, i . μι. , $S = S_1 \cup S_2 \cup \ldots \cup S_k$, $S_i \cap S_{i+1} = \varnothing$.
- Ακριβής Αλγόριθμος - **ΕΑ** :

  Αν *μήκος* (S) < t **έπειτα**

  Βήμα 1: Δημιουργήστε το μοντέλο.

  Βήμα 2: Λύστε το μοντέλο χρησιμοποιώντας το MIP Solver.

  Τέλος αν

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

Αλγόριθμοι για την επίλυση του προβλήματος σε 3D

- Ευρετικός Αλγόριθμος-1 – **HE-1**

Για *Εγώ* = 1, 2, …, *k* **κάνω:**

Αν *Εγώ* = 1 **τότε**

Δημιουργούμε και λύνουμε το μοντέλο για την είσοδο S_i και $z = i$ (2D υποπλέγμα)

Αλλού:

L et $v_j^* = 1$for $j \in J$ στη βέλτιστη λύση για S_{i-1}.

Βήμα 1: Δημιουργήστε το μοντέλο για την είσοδο S_i και $z = i$ (2D υποπλέγμα). Προσθήκη $\sum v_j$ στη συνάρτηση προορισμού (1). Για να διατηρήσουμε τη συνδεσιμότητα προσθέτουμε $x_{s1} = 1$ στο σύνολο περιορισμών, όπου s είναι ο αριθμός της γραμμής (στο ευθυγραμμισμένο γράφημα ή απλώς οι συντεταγμένες) για το τελευταίο σύμβολο του S_{i-1} στη βέλτιστη λύση.

Βήμα 2: Λύστε το μοντέλο .

Τέλος εαν

Τέλος για

Πρόβλημα αναδίπλωσης πρωτεΐνης

– –

- Ευρετικός Αλγόριθμος-2 – **HE-2**

Για *Εγώ* = 1, 2, …, *k* **κάνω:**

Αν *Εγώ* = 1 **τότε**

Δημιουργήστε και λύστε το μοντέλο για την είσοδο S_i.

Αλλού:

Βήμα 1: Δημιουργήστε το μοντέλο για την είσοδο $T_i = S_1 + S_2 + \ldots + S_i$. Για κάθε $x_{ik} = 1$ για τη βέλτιστη λύση για την είσοδο T_{i-1} προσθέστε (στο σύνολο των περιορισμών για την είσοδο T_i) $x_{ik} = 1$.

Βήμα 2: Λύστε το μοντέλο .

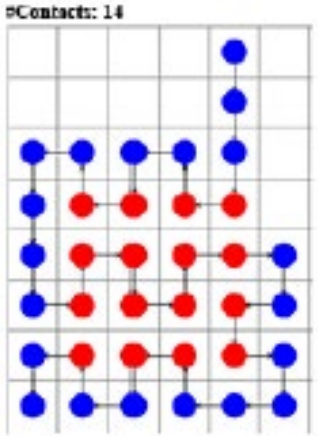
Τέλος εαν

Τέλος για

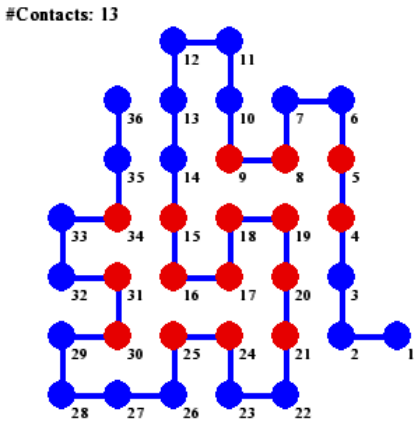
Πρόβλημα αναδίπλωσης πρωτεΐνης

Αποτελέσματα

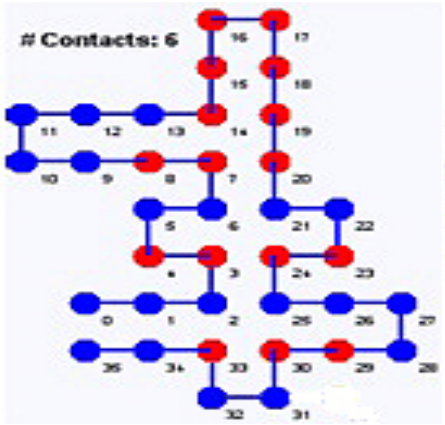
- Για ένα δείγμα δοκιμής, χρησιμοποιούμε μια μικρή ανθρώπινη πρωτεΐνη με 36 αμινοξέα. Η αλληλουχία αμινοξέων για αυτήν την πρωτεΐνη είναι PPPHHPPHHPPPPPHHHHHHHHPPHHPPPPHHPPHPP:



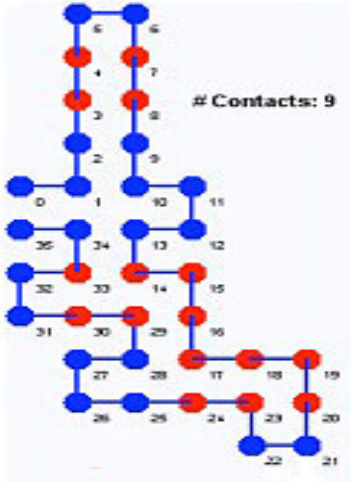
Βέλτιστη λύση



Ευρετικός αλγόριθμος



Αλγόριθμος



Αλγόριθμος Newman

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

Αποτελέσματα

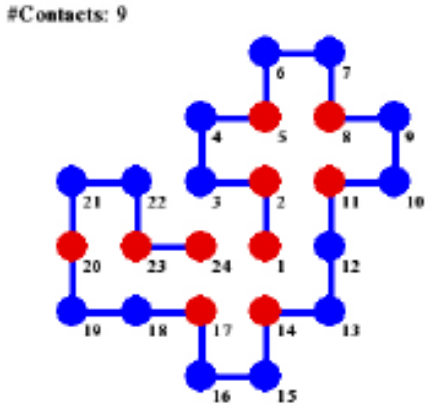
- Έχουν εφαρμοστεί υπολογιστικά πειράματα με χρήση πρωτεϊνών διαφορετικού μήκους.

Αριθμός	HP Sequence	Μήκος
1	HPHP P H H PHP P HPH H P P HPH	20
2	HHPPHPPHPPHPPHPPHPPHPPHH	24
3	P P HP P H H P PPP H H P PPP H H P PPP H H	25
4	PPPHHPPHHPPPPPHHHHHHHHPPHHPPPPPHHPPHPP	36
5	PPHHHPHHHPPPHHPHHPPHPHHHHPHPPHHHHHHPHPHHPPHHP	4 6
6	PPHPPHHPPHHPPPPPHHHHHHHHHHPPPPPHHPPHHPPHPPHHHHH	48
7	HHPHPHPHPHHHHPHPPPHPPPHPPPPPHPPPHPPPHHHHPHPHPHPHP H	50
8	$P(PH_3)_2H_5P_3H_{10}PHP_3H_{12}P_4H_6PH_2PHP$	60
9	$H_{12}PHPHP_2H_2P_2H_2P_2HP_2H_2P_2H_2P_2HP_2H_2P_2H_2P_2HPHPH_{12}$	64
10	$PH_2P_5H_2P_2H_2PHP_2HP_7HP_3H_2PH_2P_6HP_2HPHP_2HP_5H_3P_4H_2PH_2P_5H_2P_4H_4PHP_8H_5P_2HP_2$	102

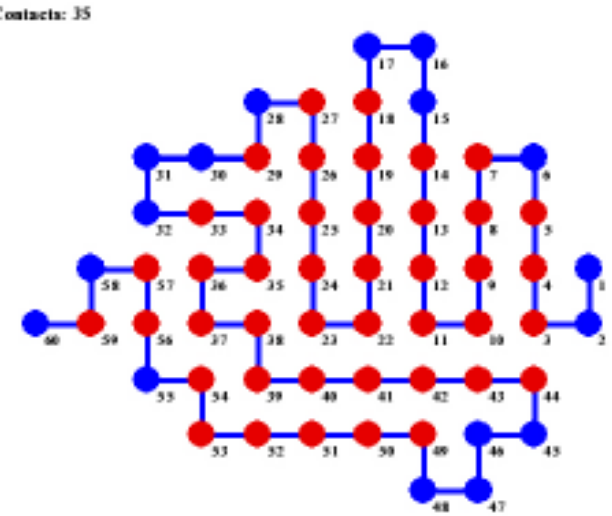
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

Αποτελέσματα



- Αναδίπλωση πρωτεΐνης μήκους 24. Ο αριθμός των επαφών είναι 9:
  HHPPHPPHPPHPPHPPHPPHPPHH



- Δίπλωμα πρωτεΐνης μήκους 60
  Ο αριθμός των επαφών είναι 35:
  $P(PH_3)_2H_5P_3H_{10}PHP_3H_{12}P_4H_6PH_2P$

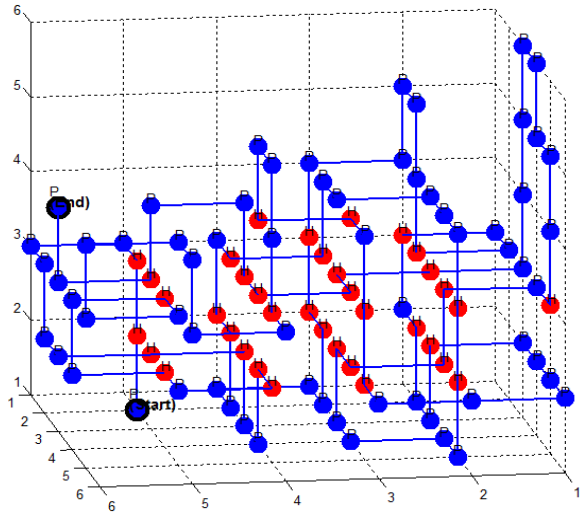
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

Αποτελέσματα

- Το παρακάτω σχήμα δείχνει τη λύση για μια πρωτεΐνη μήκους 102 αμινοξέων σε αναλυτική μορφή, όπου τα $x/\bar{x}$, $y/\bar{y}$ ή $z/\bar{z}$ αντιπροσωπεύουν συν/πλην ένα στην αντίστοιχη συντεταγμένη του προηγούμενου αμινοξέος και σε γραφικό μορφή.



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
P	H	H	P	P	P	P	P	H	H	P	P	H	H	P	H	P	P	H	P	P	P	P	P	P	P	H	P	P	P
	z	z	$\bar{y}$	x	y	$\bar{z}$	y	$\bar{x}$	x	$\bar{z}$	y	z	y	x	$\bar{y}$	x	z	$\bar{x}$	y	x	z	$\bar{x}$	y	$\bar{z}$	$\bar{x}$	$\bar{y}$	z	z	$\bar{y}$

31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
H	H	P	H	H	P	P	P	P	P	P	H	P	P	H	P	H	P	P	H	P	P	P	P	P	H	H	H	P	P
$\bar{z}$	x	y	$\bar{z}$	$\bar{z}$	y	$\bar{x}$	$\bar{x}$	$\bar{y}$	$\bar{y}$	$\bar{y}$	x	$\bar{z}$	$\bar{y}$	z	$\bar{y}$	z	$\bar{x}$	y	x	z	z	$\bar{y}$	$\bar{z}$	x	$\bar{z}$	$\bar{z}$	y	z	$\bar{y}$

61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
P	P	H	H	P	H	H	P	P	P	P	P	H	H	P	P	P	P	H	H	H	H	P	H	P	P	P	P	P	P
x	y	z	$\bar{y}$	z	y	$\bar{x}$	z	y	$\bar{x}$	y	y	$\bar{z}$	$\bar{z}$	$\bar{z}$	$\bar{y}$	x	$\bar{y}$	z	y	$\bar{x}$	z	$\bar{x}$	y	z	z	$\bar{y}$	z	$\bar{y}$	z

91	92	93	94	95	96	97	98	99	100	101	102
P	P	H	H	H	H	H	P	P	H	P	P
$\bar{z}$	$\bar{z}$	x	x	y	x	$\bar{y}$	z	x	$\bar{z}$	x	z

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

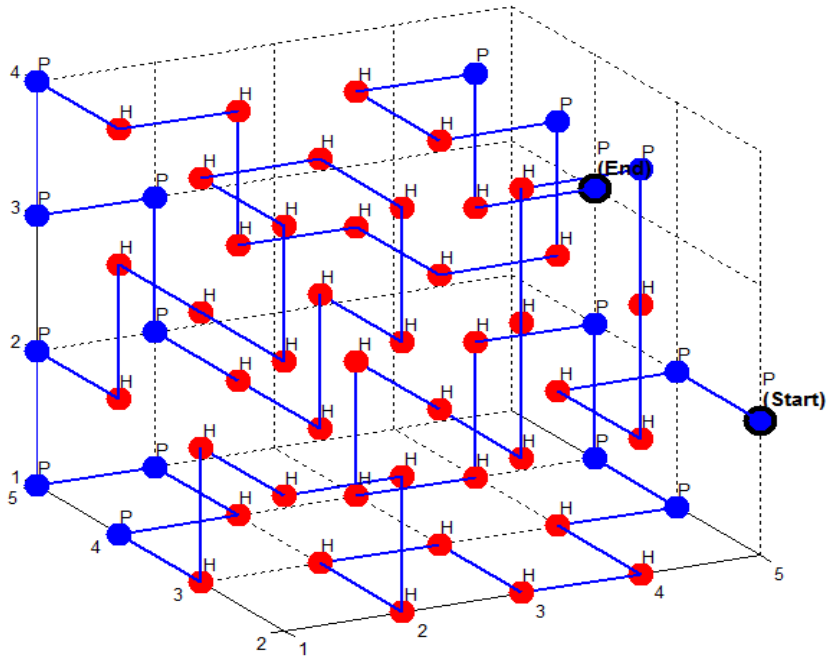
Co-funded by the Erasmus+ Programme of the European Union

Πρόβλημα αναδίπλωσης πρωτεΐνης

Αποτελέσματα

60 αμινοξέα

PPHHHHHHHHHHPPHHHHHHHHHHHPHPPPHHHHHH
HHHHHHHPPPPHHHHHHHPHHPHP



Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Σας ευχαριστώ για την προσοχή σας !

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

Θέμα 3. Πρόβλημα αναδίπλωσης πρωτεϊνών

Μάθημα 2. Στοίχιση ακολουθίας

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Στοίχιση κατά ζεύγη ακολουθίας

- Σε αυτή τη διάλεξη θα δείξουμε μερικές από τις μεθόδους για την ευθυγράμμιση δύο αλληλουχιών πρωτεΐνης, DNA ή RNA:
  - Είναι δύο αλληλουχίες ομόλογες;
  - Έχουν οι δεδομένες ακολουθίες κοινές περιοχές ή περιοχές;
  - Προσδιορίστε την ακριβή θέση κοινών χαρακτηριστικών, όπως δισουλφιδικές γέφυρες ή καταλυτικά ενεργές θέσεις.
  - Σύγκριση δύο γονιδίων ή/και προϊόντων τους.
- Συνήθως, οι βιολόγοι χρησιμοποιούν τεχνικές ευθυγράμμισης Pairwise για πρωτεΐνη, DNA ή RNA για να βελτιώσουν τα αποτελέσματα που λαμβάνονται από μια συγκεκριμένη αναζήτηση βάσης δεδομένων και να εκτελέσουν λεπτομερή ανάλυση.

Στοίχιση κατά ζεύγη ακολουθίας

- Το πιο λεπτό σημείο είναι η επιλογή των ίδιων των ακολουθιών για σύγκριση .
- Μην χρησιμοποιείτε μεθόδους στοίχισης κατά ζεύγη για να βρείτε μια ακολουθία που μπορεί να είναι ομόλογη με μια ακολουθία που έχετε ήδη.
  - Τα προγράμματα αναζήτησης βάσεων δεδομένων απλώς συγκρίνουν την ακολουθία που ζητήθηκε με όλες τις ακολουθίες στη βάση δεδομένων, επομένως δεν υπάρχει πραγματική ανάγκη για περαιτέρω συγκρίσεις ανά ζεύγη.
  - Προγράμματα όπως οι βάσεις δεδομένων αναζήτησης BLAST που χρησιμοποιούν μεθόδους που είναι βελτιστοποιημένες για ταχύτητα αναζήτησης αντί για ακρίβεια ευθυγράμμισης.

Co-funded by the Erasmus+ Programme of the European Union

Επιλογή κατάλληλων ακολουθιών

- Ο λόγος για την ευθυγράμμιση δύο αλληλουχιών είναι η υπόθεση ότι αυτές οι αλληλουχίες είναι ομόλογες και έχουν κοινή προέλευση.
  - Συχνά έχουν παρόμοιες τρισδιάστατες δομές και λειτουργίες.
- Επιλέξτε ακολουθίες σύμφωνα με τα ακόλουθα (συντηρητικά) κριτήρια:
  - Αλληλουχία DNA: Τουλάχιστον 70% ταυτότητα κατά μήκος περισσότερων από 100 βάσεων μεταξύ της αλληλουχίας που ζητήθηκε και της αναζήτησης ή παρουσία τιμής E μικρότερη από 10^{-4}.
  - Αλληλουχία Πρωτεϊνών: Πάνω από 25% ταυτότητα σε μήκος περισσότερων από 100 αμινοξέων μεταξύ της ζητούμενης και της αναζητούμενης αλληλουχίας ή παρουσία τιμής E μικρότερη από 10^{-4}.

Επιλογή κατάλληλων ακολουθιών

- Η αναζήτηση μιας βάσης δεδομένων είναι σαν να αναζητάτε "νέα λίστες διαμερισμάτων" ☺
- Η επιλογή μιας ακολουθίας από ένα αποτέλεσμα BLAST είναι σαν να κάνετε έναν συμβιβασμό μεταξύ νέων πληροφοριών και ασφάλειας.
  - Το επιθυμητό αντικείμενο πρέπει να είναι μια ευρέως σχολιασμένη ακολουθία Swiss-Prot που ικανοποιεί το ερώτημα με τιμή E πολύ μικρότερη από 10^{-4} .
  - Μια δυνατότητα είναι να επιλέξετε τις ακολουθίες σας από το αποτέλεσμα που λαμβάνεται από τη βάση δεδομένων.
  - Μια άλλη δυνατότητα είναι να επιλέξετε δύο ακολουθίες χρησιμοποιώντας πειραματικά κριτήρια , π.χ. να έχουν την ίδια συνάρτηση ή άλλες παρόμοιες ιδιότητες.
    - Για το σκοπό αυτό, μπορείτε να χρησιμοποιήσετε το Σύστημα Ανάκτησης Ακολουθιών (SRS), το οποίο σας επιτρέπει να αναγνωρίζετε τις ακολουθίες με λέξεις-κλειδιά.

Co-funded by the Erasmus+ Programme of the European Union

Επιλογή κατάλληλων ακολουθιών

- Εάν έχετε μόνο μία ακολουθία, είναι δυνατό (και μερικές φορές συνιστάται) να κάνετε την ευθυγράμμιση με την ίδια την ακολουθία.
  - Οδηγεί στην ανακάλυψη ενδιαφέρουσες ιδιότητες:
    - Επαναλαμβανόμενοι τομείς.
    - Περιοχές με επαναλαμβανόμενο μικρό μοτίβο (χαμηλό επίπεδο πολυπλοκότητας).
    - Παλίνδρομα - θραύσματα DNA που επαναλαμβάνονται σε αντίθετες κατευθύνσεις .
    - Πιθανές δευτερογενείς δομές στο RNA.

Επιλογή κατάλληλης μεθόδου ευθυγράμμισης

- Ένα από τα πιο δύσκολα καθήκοντα για τους βιολόγους είναι η ευθυγράμμιση

Method name	Situation
Dot plot	General survey of your sequence. Repeat detection. Finding long insertion/deletions. Extraction of parts of sequences to perform multiple alingment.
Local alignment	Comparison of sequences with partial homology. Comparing parts of a sequence. Making quality alinements. Performing residual versus residual type analyses.
Global alignment	Comparing two sequences along their entire length. Identification of long insertions/deletions. Data Verification. Identify each mutation in the sequence.

Επιλογή κατάλληλης μεθόδου ευθυγράμμισης

- Το Dot plot είναι πολύ ισχυρά εργαλεία που δίνουν μια στιγμιαία συνολική εικόνα σε μία μόνο στοίχιση.
  - Ένα ιδανικό εργαλείο για να αποφασίσετε το επόμενο βήμα στην ανάλυσή σας.
  - Το διάγραμμα με κουκκίδες δεν επαρκεί για μια πιο ακριβή μελέτη.
  - Η γραφική παράσταση κουκκίδων δεν πραγματοποιεί πραγματικά ευθυγράμμιση.
  - Δίνει γενικές οδηγίες.
  - Δεν δείχνει ξεκάθαρα πώς τα αμινοξέα ή τα νουκλεοτίδια αντιστοιχούν μεταξύ τους μεταξύ των δύο αλληλουχιών.
    - Για να εξαχθούν αυτές οι πληροφορίες, πρέπει να γίνει μια ευθυγράμμιση.

Επιλογή κατάλληλης μεθόδου ευθυγράμμισης

- Υπάρχουν δύο τύποι ευθυγράμμισης:
  - Καθολική στοίχιση - οι δύο αλληλουχίες συγκρίνονται σε όλο το μήκος τους.
  - Τοπική στοίχιση - συγκρίνονται μόνο τα πιο παρόμοια μέρη των δύο ακολουθιών .
- Μια καθολική στοίχιση εκτελείται μόνο εάν οι δύο ακολουθίες συνδέονται σε όλο το μήκος τους και εάν δεν περιέχουν πολύ μεγάλες εισαγωγές ή διαγραφές.
- Εάν δεν ξέρετε πώς να επιλέξετε μεταξύ καθολικής και τοπικής στοίχισης, είναι πάντα καλύτερο να χρησιμοποιείτε τοπικές μεθόδους.
- Όταν συγκρίνετε δύο ακολουθίες, ξεκινήστε με μια γραφική παράσταση Dot, συγκρίνοντας κάθε ακολουθία με τον εαυτό της.

Στοίχιση με κουκκίδες

- **Παράδειγμα 1.** Φανταστείτε ότι θέλετε να συγκρίνετε τις ακόλουθες δύο ακολουθίες:
  - Ακολουθία 1: THEFA S TCAT
  - Ακολουθία 2: THEFATCAT
- Για να φτιάξετε ένα διάγραμμα με κουκκίδες, πάρτε ένα κομμάτι χαρτί, σχεδιάστε ένα πλέγμα και:
  1. Γράψτε τα επιμέρους γράμματα της ακολουθίας 1 το ένα πάνω από το άλλο (σε στήλη).
  2. Γράψτε την ακολουθία 2 από αριστερά προς τα δεξιά (σε σειρά).

Στοίχιση με κουκκίδες

- **Παράδειγμα 1.**
  3. Εξετάστε κάθε κελί ένα προς ένα και διαγράψτε το εάν η επάνω ακολουθία και η αντίστοιχη στη σειρά περιέχουν το ίδιο σύμβολο.

  4. Νίκες με τη μεγαλύτερη διαγώνιο! - Εμφανίζει τμήματα ομοιότητας μεταξύ των δύο ακολουθιών.

Co-funded by the Erasmus+ Programme of the European Union

Στοίχιση με κουκκίδες

- Οι τεχνικές με κουκκίδες σχετίζονται στενά με τις μεθόδους συρόμενου παραθύρου.
  - Το παράθυρο είναι μέρος μιας ακολουθίας που συγκρίνεται με μέρος μιας άλλης ακολουθίας παρόμοιου μεγέθους.
- Η γραφική παράσταση με κουκκίδες έχει δύο βασικά πλεονεκτήματα:
  - Είναι εύκολο να εφαρμοστεί και δεν απαιτείται βιολογική υπόθεση.
  - Είναι δύσκολο να κάνεις λάθος.
- Η ερμηνεία της γραφικής παράστασης κουκκίδων βασίζεται στην καλύτερη συσκευή ανίχνευσης, ακριβώς στα μάτια.

Στοίχιση

- Το Dotlet είναι πολύ φιλικό γιατί:
  - Εύχρηστος;
  - Δωρεάν ;
  - Δεν απαιτείται εγκατάσταση .
  - Λειτουργεί σε (σχεδόν) οποιονδήποτε υπολογιστή έχει πρόσβαση στο Διαδίκτυο.
- Το Dotlet είναι ιδανικό για αλληλουχίες με μήκος μικρότερο από 10.000 αμινοξέα ή νουκλεοτίδια .
  - Εάν θέλετε να περιηγηθείτε σε μεγαλύτερες ακολουθίες χρησιμοποιώντας διαδικτυακά εργαλεία, τότε θα χρειαστεί να χρησιμοποιήσετε το Dnadot , το οποίο είναι ένα εργαλείο σχεδιασμένο για γραφική παράσταση κουκκίδων μεταξύ μεγάλων ακολουθιών.
- Εάν οι ακολουθίες σας είναι μεγαλύτερες από 100.000 χαρακτήρες, δεν υπάρχει κατάλληλο διαδικτυακό εργαλείο διαθέσιμο.

Co-funded by the Erasmus+ Programme of the European Union

Στοίχιση

- Διαφορετικά προγράμματα Dot plot

Name	Used for:	Range:	URL:	Platform:
Dotlet	proteins, DNA	10000	http://www.ch.embnet.org/	всички (Java)
Dnadot	proteins, DNA	100000	http://arbl.cvmbs.colostate.edu/molkit/dnadot/	всички (Java)
Dotter	proteins, DNA	100000	http://arbl.cvmbs.colostate.edu/molkit/dnadot/	Unix, Linux, Windows
Dottup	Whole genomes, DNA	> 100000	http://emboss.sourceforge.net/	Unix, Linux

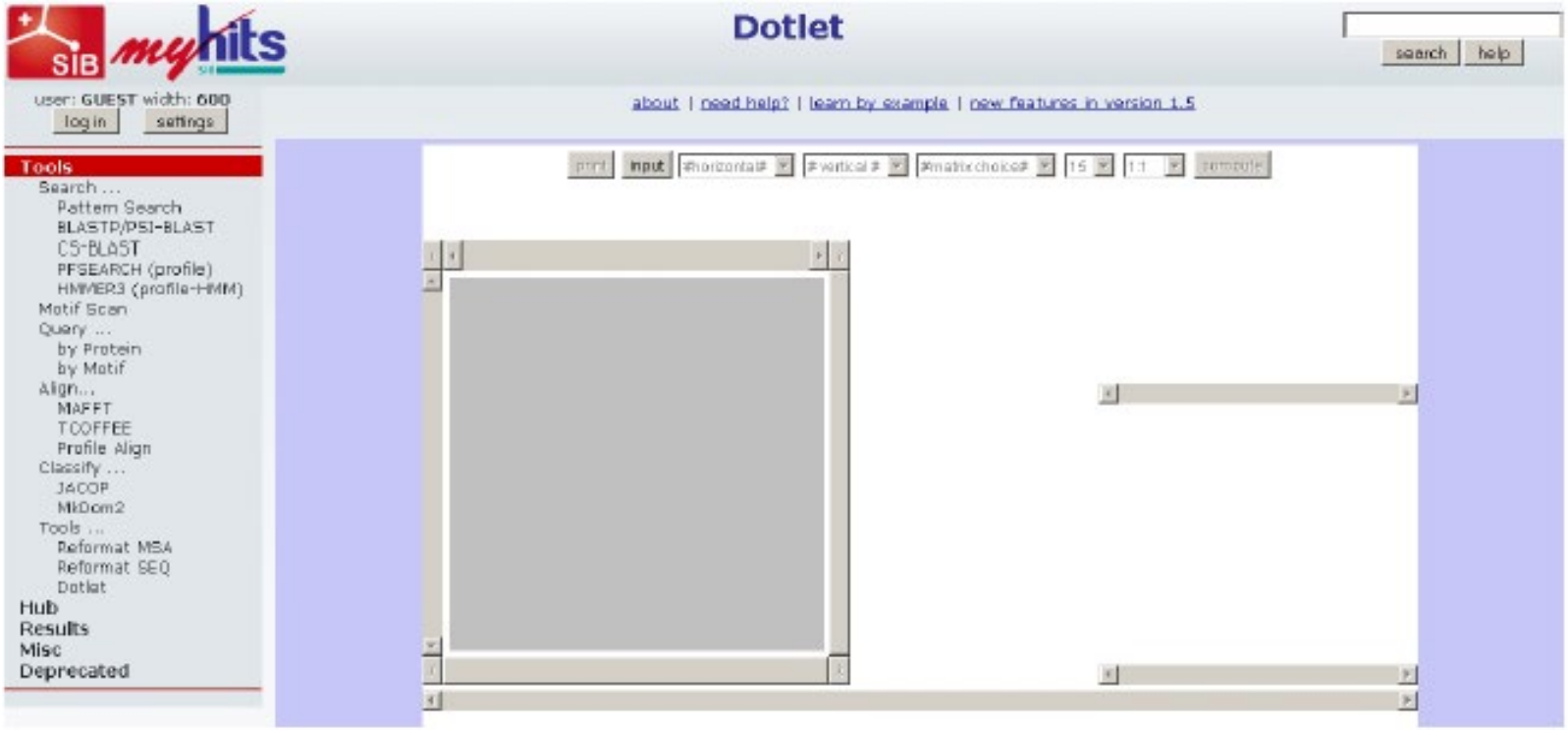
Co-funded by the Erasmus+ Programme of the European Union

Στοίχιση

- Dotlet - μια μικροεφαρμογή Java στην οποία μπορείτε να αποκτήσετε πρόσβαση στον διακομιστή EMBnet (Ευρωπαϊκό Δίκτυο Μοριακής Βιολογίας) .
  - Το Applet είναι ένα μικρό πρόγραμμα που φορτώνεται αυτόματα στο πρόγραμμα περιήγησής σας και στη συνέχεια εκτελείται στον υπολογιστή σας.
  - Δεν χρειάζεται να εγκαταστήσετε τίποτα αρκεί να έχετε Java.
- Το Dotlet δεν είναι ένα πολύπλοκο πρόγραμμα, αλλά έχει πολλές λειτουργίες.
- Αλλάξτε τις παραμέτρους μέχρι να έχετε το επιθυμητό αποτέλεσμα.
  - Βασικός κανόνας όταν χρησιμοποιείτε το Dotlet
  - Θέλετε να βρείτε ένα αποτέλεσμα, θα πρέπει να αλλάξετε διάφορες παραμέτρους μέχρι να το αποκτήσετε.

Στοίχιση

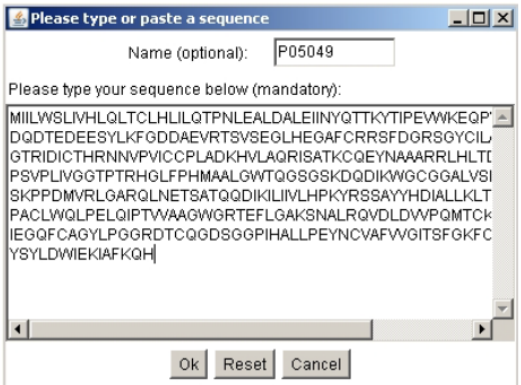
1. Starting Dotlet - Εισαγάγετε την ακόλουθη διεύθυνση στο πρόγραμμα περιήγησής σας:



Στοίχιση

2. Εισαγωγή αλληλουχίας στο Dotlet - συγκρίνονται δύο αλληλουχίες που περιέχουν μια συγκεκριμένη περιοχή κινάσης:
   1. Ανοίξτε ένα νέο παράθυρο προγράμματος περιήγησης και πληκτρολογήστε την παρακάτω διεύθυνση : http://www.uniprot.org/uniprot/P05049.fasta
   2. Αντιγράψτε τη σειρά που εμφανίζεται στο δεύτερο πρόγραμμα περιήγησης, ανοίξτε το πρώτο πρόγραμμα περιήγησης, κάντε κλικ στο κουμπί ***Εισαγωγή*** , επικολλήστε τη σειρά σε ένα παράθυρο και κάντε κλικ στο κουμπί ***OK*** .

Η σειρά ορίζεται σε ακατέργαστη μορφή



Μπορείτε να καθορίσετε ένα όνομα ακολουθίας

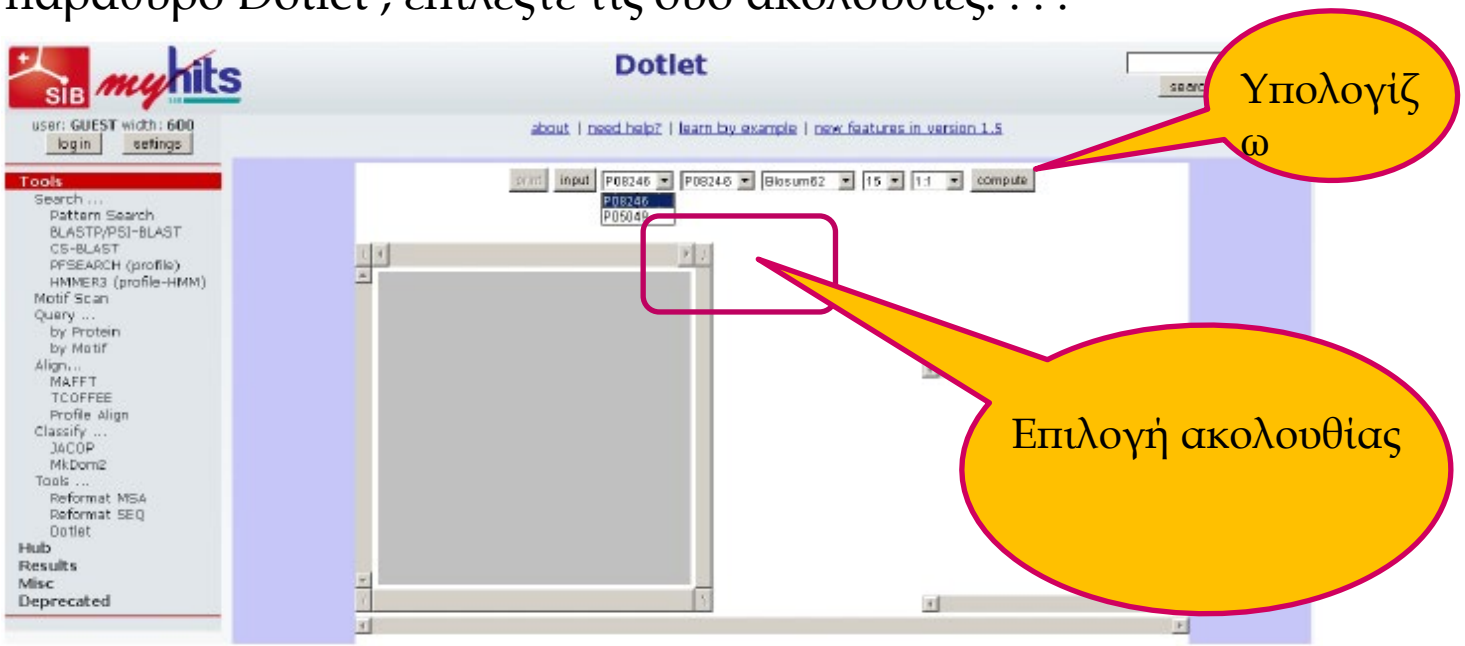
Co-funded by the Erasmus+ Programme of the European Union

Στοίχιση

2. Εισαγωγή αλληλουχίας στο Dotlet - συγκρίνονται δύο αλληλουχίες που περιέχουν μια συγκεκριμένη περιοχή κινάσης:
    3. Φόρτωση http://www.uniprot.org/uniprot/P08246.fasta στο δεύτερο παράθυρο που ανοίξατε στο βήμα 1.
    4. Κάντε κλικ στο κουμπί ***Εισαγωγή*** , αντιγράψτε την ακολουθία και επικολλήστε την στο παράθυρο Dotlet .
    5. Κάντε κλικ στο κουμπί ***OK*** στο παράθυρο.
    6. Στο παράθυρο Dotlet , επιλέξτε τις δύο ακολουθίες που θέλετε να συγκρίνετε χρησιμοποιώντας τις δύο αναπτυσσόμενες λίστες που βρίσκονται κοντά στο κουμπί ***Εισαγωγή*** .

Στοίχιση

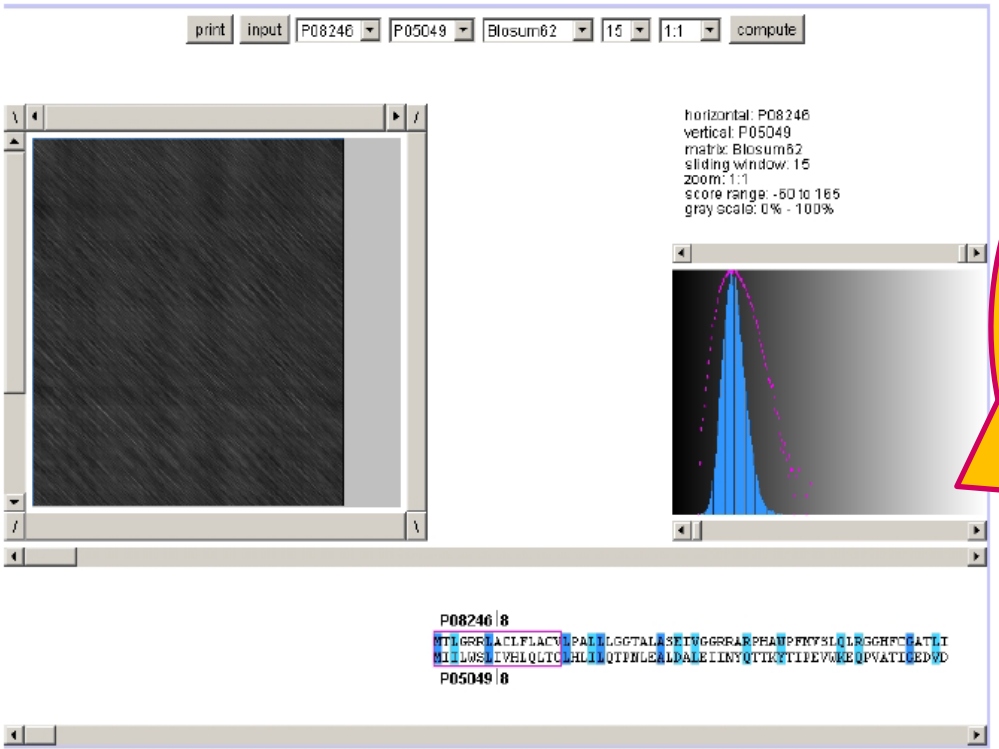
2. Εισαγωγή αλληλουχίας στο Dotlet - συγκρίνονται δύο αλληλουχίες που περιέχουν μια συγκεκριμένη περιοχή κινάσης:
   7. Στο παράθυρο Dotlet , επιλέξτε τις δύο ακολουθίες. . . .



8. Στο παράθυρο Dotlet , κάντε κλικ στο ***Υπολογισμός*** κουμπί που βρίσκεται στην άκρη δεξιά.

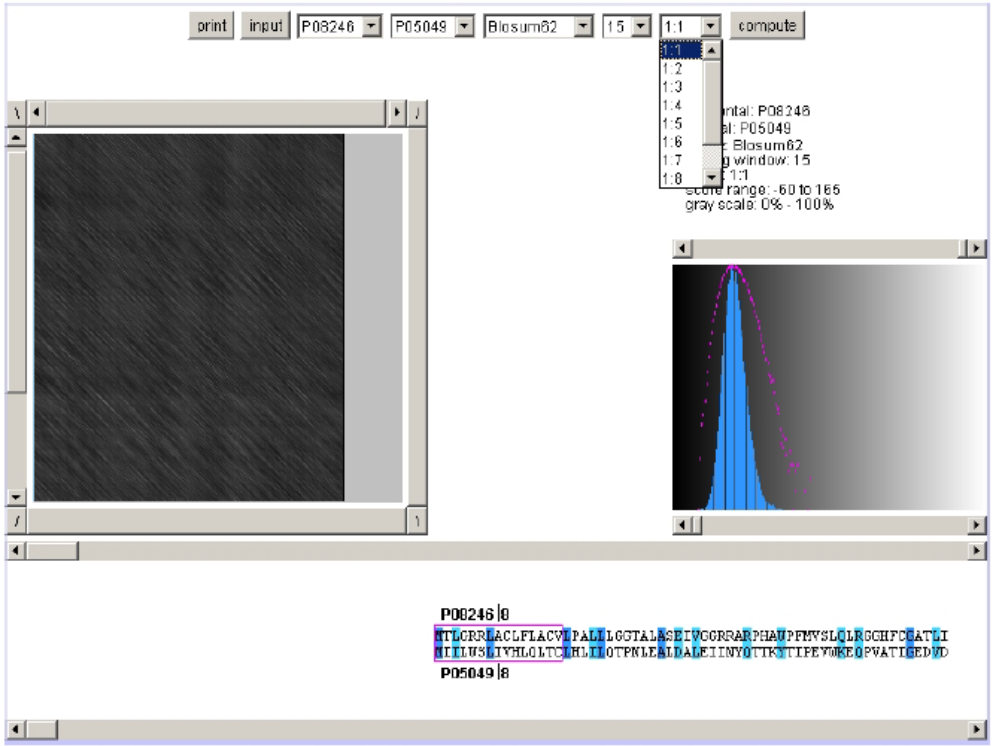
Στοίχιση

2. Εισαγωγή αλληλουχίας στο Dotlet - συγκρίνονται δύο αλληλουχίες που περιέχουν μια συγκεκριμένη περιοχή κινάσης:
   - Το αποτέλεσμα μοιάζει με αυτό



Παράθυρο κατωφλίου - δείχνει την ευαισθησία του γραφήματος Dot plot

Λεπτομερής ρύθμιση



- Οι οδηγίες που περιγράφονται ακολουθούν μια προσέγγιση πάνω-κάτω.

Προσδιορισμός πολλαπλασιαστή κλίμακας - Από προεπιλογή, το Dotlet έχει πολλαπλασιαστή κλίμακας 1:1 (1 pixel αντιστοιχεί σε 1 υπόλοιπο).

1. Όταν επιλέγετε την κλίμακα από το αναπτυσσόμενο μενού στα αριστερά του κουμπιού ***Υπολογισμός*** , επιλέξτε 1:2
2. Κάντε κλικ στο ***Υπολογισμός*** κουμπί .

Λεπτομερής ρύθμιση

2. *Επιλογή μεγέθους παραθύρου* - Όταν το Dotlet συγκρίνει δύο αμινοξέα ή δύο νουκλεοτίδια, συγκρίνει επίσης τα αμινοξέα ή τα νουκλεοτίδια σε άμεση γειτνίαση. Το μέγεθος παραθύρου ελέγχει το μέγεθος αυτών των περιοχών :
   1. Επιλέξτε 51 από το αναπτυσσόμενο μενού ***Παράθυρο (βρίσκεται στα αριστερά του πλαισίου κλιμάκωσης).***
   2. Κάντε κλικ στο ***Υπολογισμός*** κουμπί .
      - Τα μακριά παράθυρα κάνουν πιο ξεκάθαρη την γραφική παράσταση της κουκκίδας.
      - Το μέγεθος του παραθύρου πρέπει να είναι εντός του εύρους μεγεθών των στοιχείων που αναζητάτε.
      - Τα μικρότερα παράθυρα είναι πιο ευαίσθητα αλλά περιέχουν αρκετό θόρυβο.
      - Ξεκινήστε με ένα μεγάλο παράθυρο και κλιμακώστε μέχρι να εμφανιστεί το σήμα που αναζητάτε.

Λεπτομερής ρύθμιση

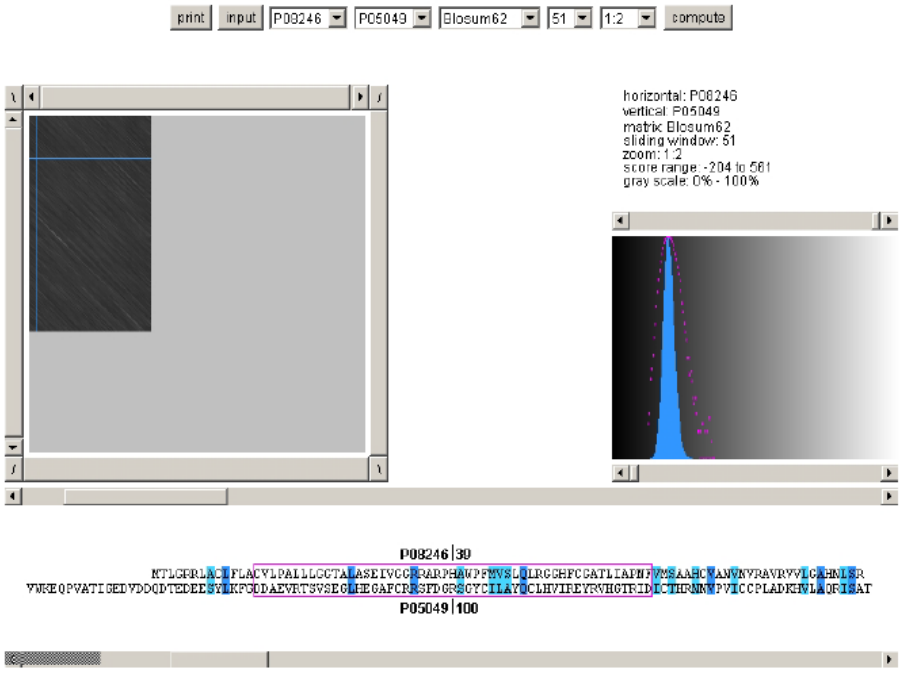
3. *Προσαρμογή κατωφλίου* - Το όριο είναι μια τιμή που καθορίζει το χρώμα κάθε κουκκίδας στο παράθυρο γραφικής παράστασης κουκκίδων. Οι πόντοι των οποίων η βαθμολογία είναι πάνω από το όριο είναι λευκοί και αυτοί με σκορ κάτω από το όριο είναι μαύροι:

   – Ο οριζόντιος άξονας του παραθύρου κατωφλίου αντιπροσωπεύει το αποτέλεσμα. Τα χαμηλά είναι αριστερά, τα ψηλά είναι δεξιά.

   – Η μεγάλη κορυφή στην καμπύλη υποδεικνύει το πιο κοινό αποτέλεσμα για δύο υπολείμματα που επιλέχθηκαν τυχαία.

   – Μπορείτε να φανταστείτε το κατώφλι ως δύο κάθετες γραμμές που περνούν από το πάνω και το κάτω μέρος του δείκτη. Η θέση του ελέγχει την εμφάνιση του παραθύρου γραφικής παράστασης κουκκίδων:
      - Οι βαθμοί με σκορ κάτω από το κατώτερο όριο είναι μαύροι.
      - Οι πόντοι με σκορ πάνω από το ανώτερο όριο είναι με λευκό χρώμα.
      - Οι βαθμοί με σκορ μεταξύ των δύο ορίων είναι γκρι.

Λεπτομερής ρύθμιση



3. *Προσαρμογή ορίου:*

1. Μετακινήστε τον κάτω κέρσορα του παραθύρου κατωφλίου τέρμα προς τα δεξιά - Αυτό κάνει το διάγραμμα κουκκίδων στα αριστερά εντελώς μαύρο.
2. Πάρτε τη δεξιά πλευρά του παραθύρου του κατωφλίου στη μέση και σύρετέ το αργά προς τα αριστερά, κρατώντας πατημένο το δεξί κουμπί του ποντικιού.
   - Όταν ο επάνω και ο κάτω κέρσορας βρίσκονται στην κορυφή, οι κουκκίδες στο διάγραμμα κουκκίδων είναι είτε μαύρες είτε λευκές.

Λεπτομερής ρύθμιση

4. *Αποθήκευση σχεδίου κουκκίδων :*

1. Πατήστε το κουμπί *PrntScrn* στο πληκτρολόγιό σας (συνήθως βρίσκεται στην επάνω δεξιά γωνία).

2. Ξεκινήστε την παρουσίαση/ζωγραφική του Power Point.

3. Χρησιμοποιήστε τις συντομεύσεις πληκτρολογίου Ctrl+X και Ctrl+V για να πάρετε και να επικολλήσετε τις πληροφορίες στην παρουσίαση.

Co-funded by the Erasmus+ Programme of the European Union

Ερμηνεία αποτελεσμάτων

- Οι δύο πρωτεΐνες που εξετάζονται στο παράδειγμα (P05049 (πρωτεάση σερίνης) και P08246) σχετίζονται σε μεγάλο βαθμό από απόσταση.
  - Εάν εκτελέσετε ένα BLAST μεταξύ τους, θα λάβετε έναν αγώνα E-value 10-4.
- Η διεξαγωγή μιας γραφικής παράστασης κουκκίδων είναι κατάλληλη για να ελέγξετε εάν δύο ακολουθίες είναι ομόλογες σε μια συγκεκριμένη περιοχή.
- Ο εντοπισμός διατηρημένων περιοχών μεταξύ δύο πρωτεϊνών (ακόμα και αν δεν γνωρίζετε τίποτα για τις λειτουργίες τους) είναι μια καλή προσέγγιση για να βρείτε μια συγκεκριμένη περιοχή υπεύθυνη για μια δεδομένη λειτουργία.

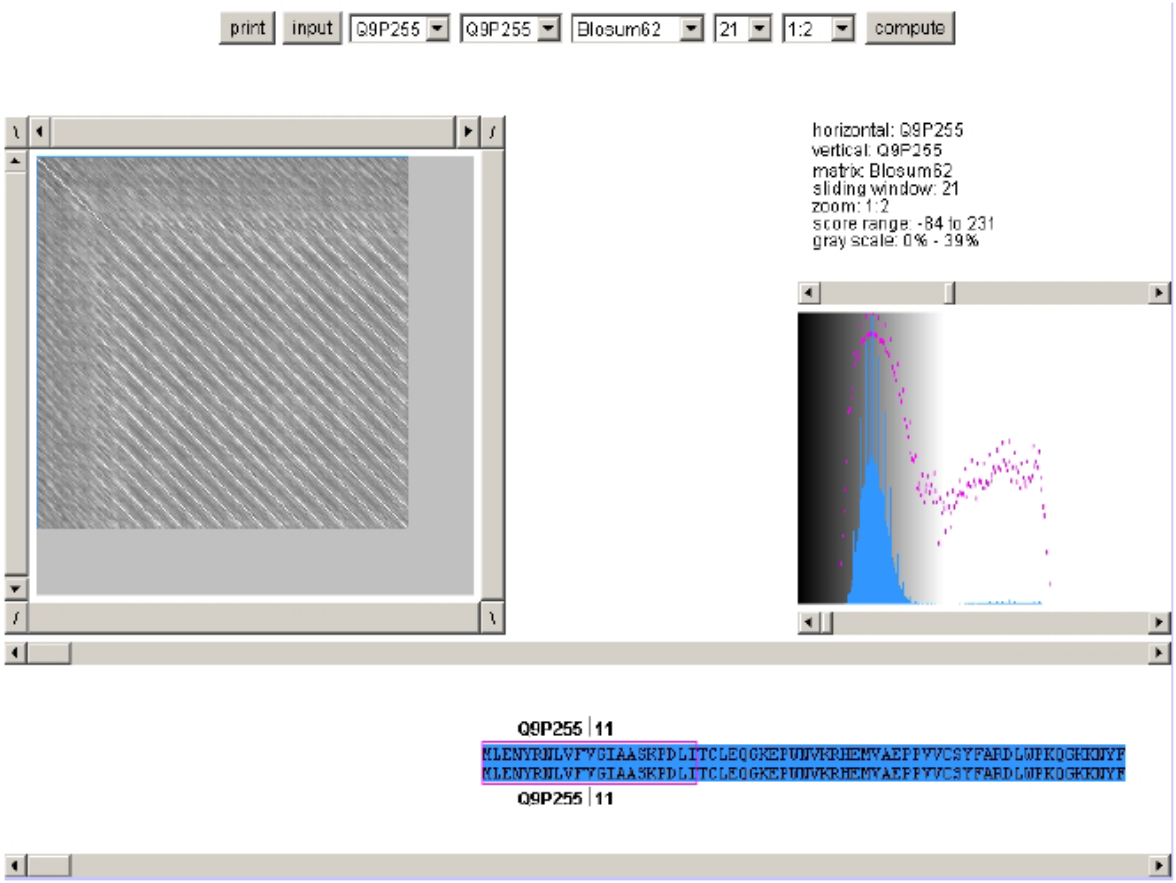
Co-funded by the Erasmus+ Programme of the European Union

Βιολογική ανάλυση με χρήση Dot plot

- Το διάγραμμα κουκκίδων επιτρέπει την αναπαράσταση σημαντικών βιολογικών σχέσεων.
- ***Καθορισμός διαδοχικών επαναλήψεων*** **-** Οι πρωτεΐνες πολύ συχνά περιέχουν σύντομους τομείς πολλαπλών επαναλήψεων.
  - Ο εσωτερικός διπλασιασμός είναι ένα εργαλείο που χρησιμοποιείται συχνά από την εξέλιξη για τη δημιουργία νέων πρωτεϊνών ή πρωτεϊνών που λειτουργούν πιο αποτελεσματικά.

- Η γραφική παράσταση με τελείες είναι ο καλύτερος τρόπος για τον εντοπισμό επαναλαμβανόμενων τομέων.

Βιολογική ανάλυση με χρήση Dot plot



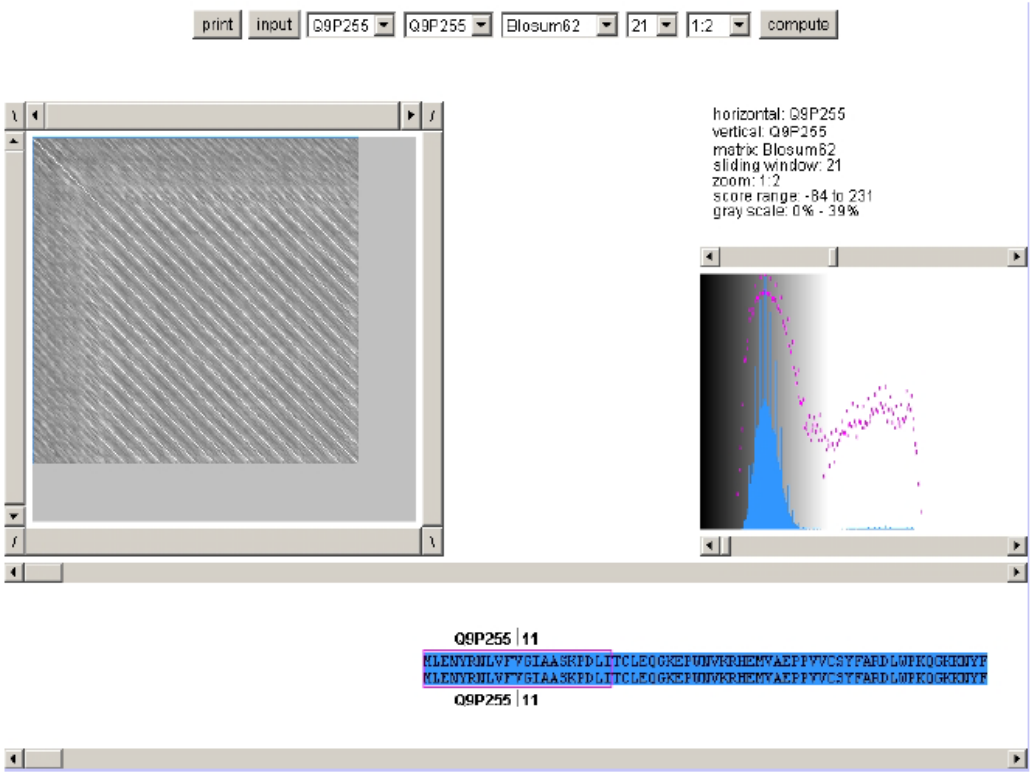
Η YE73 (ανθρώπινη πρωτεΐνη) είναι ένας πιθανός παράγοντας μεταγραφής (που εμπλέκεται στη μεταγραφή του RNA) που περιέχει 13 τομείς με επαναλαμβανόμενο *δακτύλιο Zn*

Βιολογική ανάλυση με χρήση Dot plot

- Για να αναλύσετε το YE73, ακολουθήστε τα εξής βήματα:
  1. Φορτώστε την ακόλουθη διεύθυνση στο πρόγραμμα περιήγησής σας : http://www.uniprot.org/uniprot/Q9P255.fasta
  2. Εισαγάγετε την ακολουθία στο Dotlet .
  3. Ορίστε τις ακόλουθες ρυθμίσεις παραθύρου και κατωφλίου:
     1. Ορισμός μεγέθους παραθύρου 21 ,
     2. Κλίμακα 1:2,
     3. 39% για την κλίμακα του γκρι, για τον επάνω και τον κάτω κέρσορα του παραθύρου κατωφλίου.

- Η κύρια διαγώνιος αντιπροσωπεύει την ακολουθία απέναντι από την ίδια.
- Οι επαναλήψεις εμφανίζονται ως μεγάλες συνεχείς διαγώνιοι πάνω και κάτω από την κύρια διαγώνιο.
- Οι διαγώνιοι είναι ομοιόμορφα κατανεμημένες.
- Οι διαγώνιοι περιορίζονται από ένα τετράγωνο σχήμα.

Βιολογική ανάλυση με χρήση Dot plot

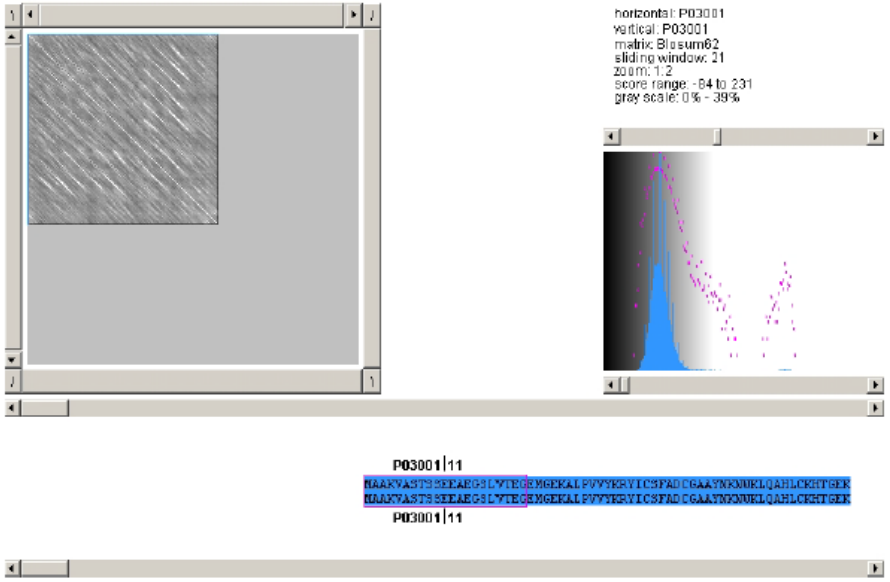


- Τρία συμπεράσματα :
  - Ο αριθμός των επαναλήψεων είναι ίσος με τον αριθμό των διαγωνίων κάτω (ή πάνω) από την κύρια διαγώνιο (συμπεριλαμβανομένης της ίδιας της κύριας διαγωνίου).
  - Η απόσταση μεταξύ δύο γειτονικών διαγωνίων αντιπροσωπεύει το μέγεθος της επανάληψης.
  - Η συντομότερη διαγώνιος δίνει τις συντεταγμένες μιας μοναδικής μονάδας επανάληψης.

Βιολογική ανάλυση με χρήση Dot plot

- Οι πρωτεΐνες που περιέχουν επαναλήψεις που σχετίζονται με απόσταση παράγουν πιο πολύπλοκα μοτίβα σημείων.
- Το Tf3a είναι ένας παράγοντας μεταγραφής που περιέχει επίσης διαδοχικούς τομείς *Zn-ring* :

http: print input P03001 P03001 Blosum62 21 1:2 compute .fasta



Βιολογική ανάλυση με χρήση Dot plot

- Οι διαδοχικές περιοχές δεν είναι η μόνη μορφή επαναλαμβανόμενων περιοχών σε πρωτεΐνες και DNA.
  - Η περιγραφή των επαναλήψεων είναι η ίδια για διαφορετικές μορφές επαναλαμβανόμενων τομέων (πρωτεΐνες/DNA): Μια διαγώνιος που βρίσκεται εκτός της κύριας διαγώνιου όταν ευθυγραμμίζετε μια ακολουθία με τον εαυτό της.
- Εάν βρείτε έναν επαναλαμβανόμενο τομέα με άγνωστη λειτουργία και καμία ομοιότητα με τις πρωτεΐνες που έχουν ήδη χαρακτηριστεί, τότε θα πρέπει να κάνετε τα εξής:
  1. Εξάγετε κάθε επαναλαμβανόμενη μονάδα.
  2. Κάνετε στοίχιση πολλαπλών τομέων.
  3. Προσδιορίστε συντηρητικές τοποθεσίες στον τομέα.
  4. Συμπεριλάβετε τον τομέα σας στο μοντέλο ή στο προφίλ PROSITE.
  5. Αναζητήστε το Swiss- Prot για να ελέγξετε αν αυτό το μοτίβο σχετίζεται με ορισμένα χαρακτηριστικά.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Εύρεση περιοχών χαμηλής πολυπλοκότητας σε πρωτεΐνες

- ***Χαμηλός περιφέρειες πολυπλοκότητας*** – θραύσματα που περιέχουν ορισμένα αμινοξέα πιο συχνά από το αναμενόμενο στις κανονικές πρωτεΐνες.
  - Για παράδειγμα, ένα τμήμα 100 αμινοξέων που περιέχει 45 προλίνες είναι σίγουρα ένα θραύσμα χαμηλής πολυπλοκότητας.
  - Συχνά έχουν βιολογικές λειτουργίες όπως αλληλεπίδραση πρωτεΐνης-πρωτεΐνης (φερμουάρ λευκίνης) ή μη ειδική σύνδεση DNA/RNA (περιοχές πλούσιες σε ARG).

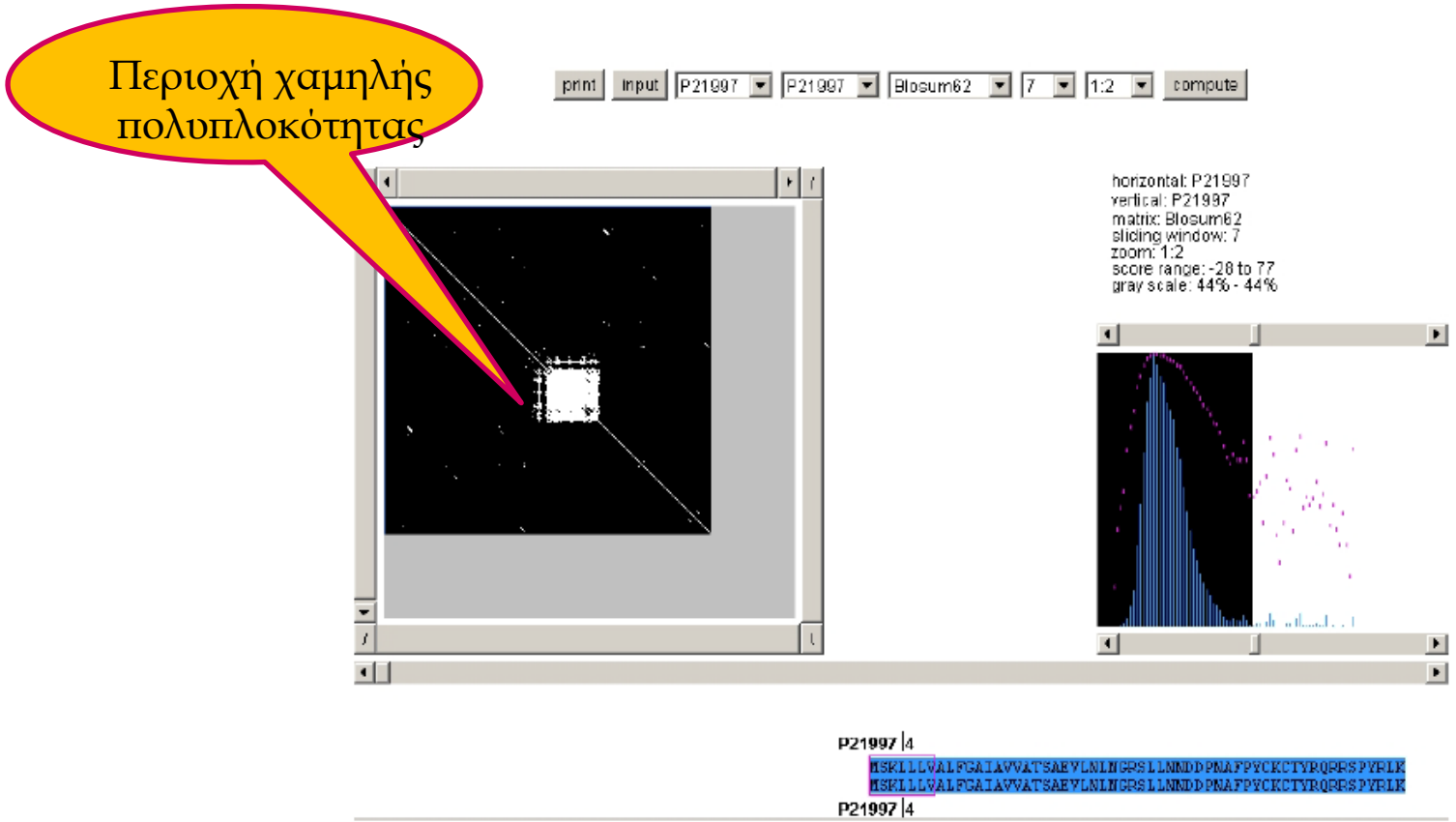
- Κατά την ευθυγράμμιση μιας ακολουθίας με τον εαυτό της, οι περιοχές χαμηλής πολυπλοκότητας εμφανίζονται ως τετράγωνα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Εύρεση περιοχών χαμηλής πολυπλοκότητας σε πρωτεΐνες

- Προσδιορισμός τμημάτων χαμηλής πολυπλοκότητας με το Dotlet



Co-funded by the Erasmus+ Programme of the European Union

Εύρεση περιοχών χαμηλής πολυπλοκότητας σε πρωτεΐνες

- Για να προσδιορίσετε περιοχές χαμηλής πολυπλοκότητας, ακολουθήστε τα εξής βήματα:

1. Φορτώστε την ακόλουθη διεύθυνση στο πρόγραμμα περιήγησής σας:

http://www.uniprot.org/uniprot/P21997.fasta

   - Το P21997 είναι ο αριθμός προσχώρησης της θειικής γλυκοπρωτεΐνης 185.

2. Εισαγάγετε την ακολουθία στο Dotlet .

3. Ορίστε τις ακόλουθες ρυθμίσεις :
   - Μέγεθος παραθύρου 7;
   - Κλίμακα 1:2;
   - 44% για την κλίμακα του γκρι, για τον επάνω και τον κάτω δρομέα του παραθύρου κατωφλίου.

Αναλύοντας τα νουκλεϊκά οξέα με Dotlet

- Το διάγραμμα κουκκίδων είναι πολύ κατάλληλο για χαρτογράφηση γονιδίων.
  - Το Dotlet μπορεί να είναι αρκετά αργό για εφαρμογές που απαιτούν τακτική ευθυγράμμιση αλληλουχιών με μήκος μεγαλύτερο από 10.000 υπολείμματα.
- Κατά την ανάλυση αλληλουχιών νουκλεοτιδίων, το Dotlet περιέχει δύο χρήσιμα χαρακτηριστικά:
  - Όταν ευθυγραμμίζεται μια αλληλουχία πρωτεΐνης με μια αλληλουχία νουκλεοτιδίων, το Dotlet μεταφράζει αυτόματα την αλληλουχία νουκλεοτιδίων και στα τρία πιθανά πλαίσια ανάγνωσης.
  - Όταν ευθυγραμμίζει μια αλληλουχία νουκλεοτιδίων με τον εαυτό της, το Dotlet αντικαθιστά αυτόματα μια από τις αλληλουχίες με τη συμπληρωματική της αλληλουχία.
- Ηλεκτρονικός οδηγός :

  https://myhits.sib.swiss/util/dotlet/doc/dotlet_help.html

Εκτέλεση τοπικής ευθυγράμμισης

- Δύο τύποι ευθυγράμμισης :
  - **Καθολικά -** οι δύο ακολουθίες είναι ευθυγραμμισμένες σε όλο το μήκος τους
  - **Τοπικό -** το πρόγραμμα συγκρίνει μόνο τα πιο παρόμοια μέρη των δύο ακολουθιών και αγνοεί τα υπόλοιπα.
- Εάν έχετε μια γραφική παράσταση με κουκκίδες που δείχνει ότι οι δύο ακολουθίες σας συνδέονται σε όλο το μήκος τους, δεν χρειάζεται να κάνετε τοπική στοίχιση , μπορείτε να μεταβείτε απευθείας σε μια καθολική .
- Δύο κύριοι λόγοι για να αναλύσετε τις ακολουθίες σας χρησιμοποιώντας τοπική στοίχιση :
  1. Συγκρίνετε δύο απομακρυσμένα συγγενείς αλληλουχίες που μοιράζονται μόνο λίγους μη συνεχόμενους τομείς/περιοχές.
  2. Να αναλύσει επαναλαμβανόμενα στοιχεία μέσα στην ίδια σειρά.

Εκτέλεση τοπικής ευθυγράμμισης

- τοπικές μέθοδοι στοίχισης κάνουν αυτό που υπονοεί το όνομά τους: ορίζονται δύο ακολουθίες, με αποτέλεσμα να επιστρέφει μια στοίχιση των πιο όμοιων τμημάτων αυτών των ακολουθιών.
  - Αφαιρεί αυτόματα περιοχές από αμινοξέα και νουκλεοτίδια που δεν μπορούν να συγκριθούν .
- Θεωρητικά, η τοπική ευθυγράμμιση αντιστοιχεί σε μία από τις διαγώνιες που εμφανίζονται σε ένα διάγραμμα με τελείες.
- Στην πράξη, μπορεί να δείτε ένα σήμα γραφικής κουκκίδας που δεν αντιστοιχεί σε καμία τοπική στοίχιση.
- Περιστασιακά, τα τοπικά προγράμματα ευθυγράμμισης αναφέρουν στοίχιση που δεν εμφανίζονται στο διάγραμμα με κουκκίδες.

Co-funded by the Erasmus+ Programme of the European Union

Κατάλληλη επιλογή τοπικής ευθυγράμμισης

- Δύο βασικές μέθοδοι για τη διεξαγωγή τοπικής ευθυγράμμισης :
  - Γρήγορη, ευρετική μέθοδος ενσωματωμένη στο πρόγραμμα BLAST
  - Πιο αργό και πιο ακριβές, το πρόγραμμα Lalign .

Ability	BLAST	Lalign
Speed	Very fast	Slower
Sequences lenght	Very long sequences	Shorter sequences
Result	E-value	E-value
Alignments	Reports the best one	Reports the best ten (or more)
Sequence type	Best with DNA	Best with proteins

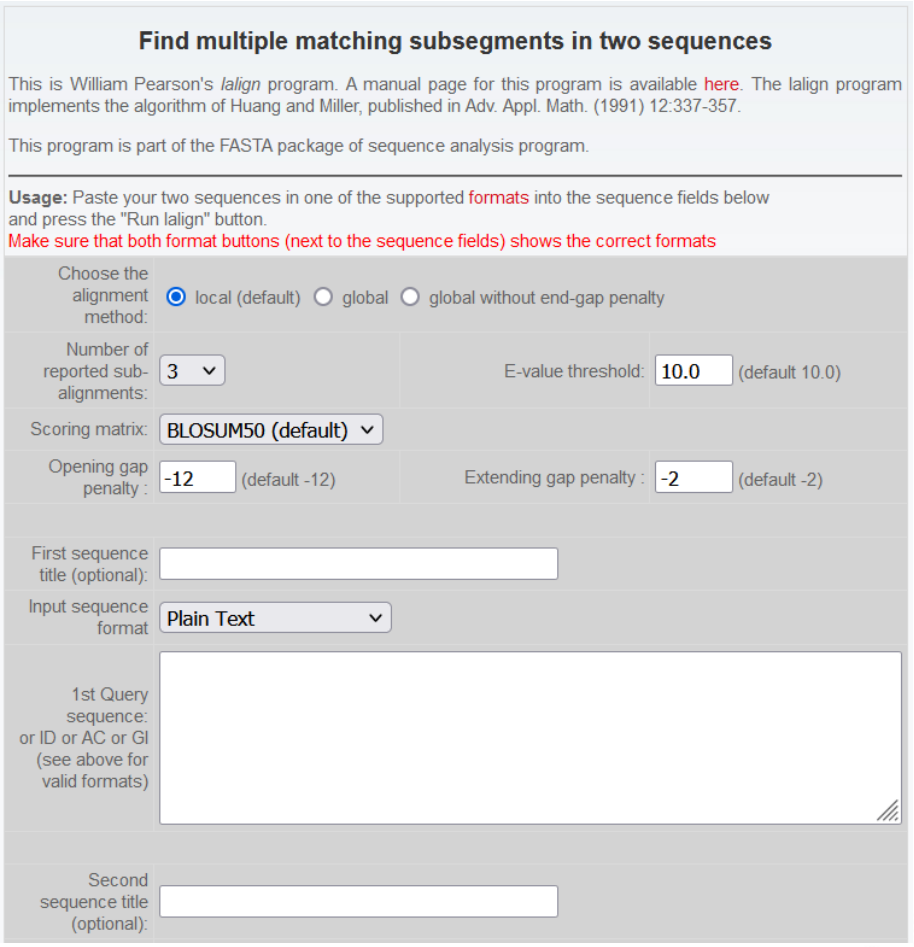
Co-funded by the Erasmus+ Programme of the European Union

Κατάλληλη επιλογή τοπικής ευθυγράμμισης

- Το BLAST μπορεί επίσης να χρησιμοποιηθεί για αναζήτηση σε μια δεδομένη βάση δεδομένων.
  - BLAST που συγκρίνει δύο αλληλουχίες (bl2seq).
  - Έχει προσαρμοστεί έτσι ώστε να μπορείτε να το περιορίσετε σε δύο μόνο ακολουθίες.
  - Δεν δημιουργεί στοίχιση εκτός από τα αποτελέσματα BLAST που λαμβάνετε όταν κάνετε αναζήτηση σε μια δεδομένη βάση δεδομένων.

https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE%20TYPE=BlastHome

Χρησιμοποιώντας το Laign για να βρείτε δέκα καλύτερες τοπικές ευθυγραμμίσεις



Find multiple matching subsegments in two sequences

This is William Pearson's *lalign* program. A manual page for this program is available here. The lalign program implements the algorithm of Huang and Miller, published in Adv. Appl. Math. (1991) 12:337-357.

This program is part of the FASTA package of sequence analysis program.

Usage: Paste your two sequences in one of the supported formats into the sequence fields below and press the "Run lalign" button.
Make sure that both format buttons (next to the sequence fields) shows the correct formats

Choose the alignment method: local (default) global global without end-gap penalty

Number of reported sub-alignments: 3 E-value threshold: 10.0 (default 10.0)

Scoring matrix: BLOSUM50 (default)

Opening gap penalty : -12 (default -12) Extending gap penalty : -2 (default -2)

First sequence title (optional):

Input sequence format: Plain Text

1st Query sequence: or ID or AC or GI (see above for valid formats)

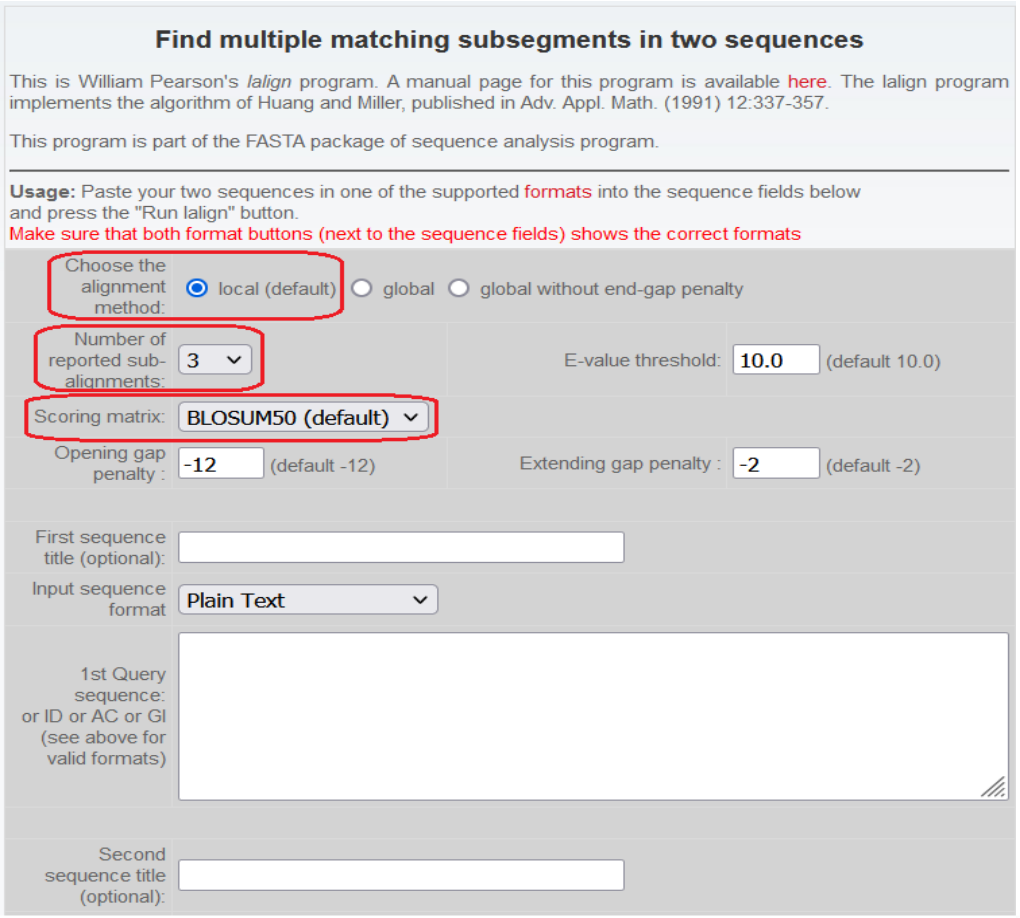
Second sequence title (optional):

- Το Lalign είναι μια ιδανική προσθήκη στο BLAST.
  - Πιο αργό, αλλά πιο ακριβές, και ως αποτέλεσμα δίνει όσες τοπικές ευθυγραμμίσεις θέλετε (καλύτερη, επόμενη, κ.λπ. μέχρι τον αριθμό που θέλατε και καθορίσατε).
  - Κατάλληλο για ανάλυση σύνθετων πρωτεϊνών με πολλές επαναλήψεις.

Χρησιμοποιώντας το Laign για να βρείτε δέκα καλύτερες τοπικές ευθυγραμμίσεις

- **Παράδειγμα 2.** Εξάγετε τοπικές ευθυγραμμίσεις από δύο απομακρυσμένα σχετιζόμενες αλληλουχίες που περιέχουν τομείς πρωτεασών σερίνης:
  1. Φορτώστε την ακόλουθη διεύθυνση στο πρόγραμμα περιήγησής σας: https://embnet.vital-it.ch/software/LALIGN_form.html
  2. Επιλέξτε ***Τοπικό*** από τις επιλογές για τις μεθόδους στοίχισης.
  3. Επιλέξτε τον αριθμό των υπο-ευθυγραμμίσεων που εμφανίζονται:
     - Πρέπει να επιλέξετε έναν αριθμό που αντιστοιχεί περίπου στον αριθμό των διαγωνίων που παρατηρήσατε στο διάγραμμα με τελείες.
  4. Επιλέξτε έναν πίνακα αξιολόγησης - ελέγχει την τιμή των μεταλλάξεων όταν το Lalign συγκρίνει αλληλουχίες.

Χρησιμοποιώντας το Laign για να βρείτε δέκα καλύτερες τοπικές ευθυγραμμίσεις



- **Παράδειγμα 2.** Εξάγετε τοπικές ευθυγραμμίσεις από δύο σχετικές απομακρυσμένες αλληλουχίες που περιέχουν τομείς πρωτεϊνάσης σερίνης.

Co-funded by the Erasmus+ Programme of the European Union

Χρησιμοποιώντας το Laign για να βρείτε δέκα καλύτερες τοπικές ευθυγραμμίσεις

- **Παράδειγμα 2.** Εξάγετε τοπικές ευθυγραμμίσεις από δύο απομακρυσμένα συγγενείς αλληλουχίες που περιέχουν τομείς πρωτεϊνάσης σερίνης:
  5. Ορισμός μεγέθους για ***άνοιγμα χάσμα επιλογή ποινής*** - ορίζει τη δυνατότητα ανοίγματος μιας "τρύπας" σε μία από τις ακολουθίες κατά την ευθυγράμμιση.
  6. Καθορίστε μια τιμή για ***την επέκταση χάσμα ποινή*** - Η τιμή για ***την επέκταση χάσμα ποινή*** + η τιμή για ***το άνοιγμα χάσμα ποινή*** χαρακτηρίζουν το χάσμα ευθυγράμμισης ακολουθίας.
     - Η τιμή για την ***επέκταση χάσμα ποινή*** πρέπει να είναι τουλάχιστον δέκα φορές χαμηλότερη από την τιμή για το ***άνοιγμα χάσμα ποινή*** .
  7. Επιλέξτε ***Swiss- Prot ID ή AC*** από την ***Είσοδο*** Αναπτυσσόμενο μενού ***Μορφή ακολουθίας .***

Χρησιμοποιώντας το Laign για να βρείτε δέκα καλύτερες τοπικές ευθυγραμμίσεις

- **Παράδειγμα 2.** Εξάγετε τοπικές ευθυγραμμίσεις από δύο απομακρυσμένα συγγενείς αλληλουχίες που περιέχουν τομείς πρωτεϊνάσης σερίνης:
  8. Εισαγάγετε ***P05049*** στο πρώτο πεδίο .
  9. Επιλέξτε ***Swiss-Prot ID ή AC*** από την ***Είσοδο*** Αναπτυσσόμενο μενού ***Μορφή ακολουθίας .***
  10. Εισαγάγετε ***P08246*** στο δεύτερο πεδίο .
  11. Κάντε κλικ στο ***Τρέξιμο Laign*** κουμπί .
  12. Αποθήκευση αποτελεσμάτων - Για να αποθηκεύσετε το αποτέλεσμα σε ένα αρχείο, χρησιμοποιήστε το ***Αρχείο→Αποθήκευση ως*** εντολές ως επιλογή στο πρόγραμμα περιήγησής σας.

Ερμηνεία των αποτελεσμάτων

- Το Lalign δίνει ως αποτέλεσμα τον αριθμό των ευθυγραμμίσεων που έχετε καθορίσει, ταξινομημένα με βάση τη βαθμολογία τους.

Readseq version 2.1.30 (12-May-2010) Readseq version 2.1.30 (12-May-2010)

lalign output for P05049 vs. P08246

[EMBnet-Server] Date: Tue Oct 5 3:46:51 2021

```
# bin/lalign36 -E 10.0 -f -12 -g -2 29292.1.seq 29292.2.seq -J -K 5
LALIGN finds non-overlapping local alignments
 version 36.3.5e Nov, 2012(preload8)
Please cite:
 X. Huang and W. Miller (1991) Adv. Appl. Math. 12:373-381

Query: 29292.1.seq
  1>>>sp|P05049|SNAK_DROME (snk)RecName: Full=Serine protease snake; EC=3.4.21.-; Flags: Precursor;[Drosophila melanogast - 435 aa
Library: 29292.2.seq
      267 residues in     1 sequences

Statistics: (shuffled [500]) MLE statistics: Lambda= 0.1852;  K=0.03785
 statistics sampled from 1 (1) to 500 sequences
Threshold: E() < 10 score: 33
Algorithm: Smith-Waterman (SSE2, Michael Farrar 2006) (7.2 Nov 2010)
Parameters: BL50 matrix (15:-5), open/ext: -12/-2
 Scan time:  0.010

>>sp|P08246|ELNE_HUMAN (ELANE)RecName: Full=Neutrophil e  (267 aa)
 Waterman-Eggert score: 181;  53.1 bits; E(1) <  1.2e-11
25.5% identity (53.0% similar) in 251 aa overlap (186-430:30-245)

         190       200       210       220       230       240
sp|P05 IVGGTPTRHGLFPHMAALGWTQGSGSKDQDIKWGCGGALVSELYVLTAAHCATSGSKPPD
       :::: .:   .: :..:    :          ::..:..  .:..::::... .
sp|P08 IVGGRRARPHAWPFMVSLQLRGGHF---------CGATLIAPNFVMSAAHCVANVNVRAV
      30        40        50                60        70        80
```

Ερμηνεία των αποτελεσμάτων

- Laign είναι σε μορφή τύπου BLAST, διατεταγμένες σύμφωνα με τις τιμές E τους.
- Τα ακόλουθα χαρακτηριστικά βρίσκονται στην πρώτη γραμμή:
  - ***Ποσοστό ταυτότητας -*** Το κλάσμα των πανομοιότυπων υπολειμμάτων σε σύγκριση μεταξύ τους.
  - ***Μήκος τοπικής στοίχισης (Επικάλυψη) -*** Αυτό είναι το συνολικό μήκος της τοπικής στοίχισης.
  - ***Βαθμολογία (Σύγκριση Βαθμολογίας)*** - Όσο μεγαλύτερη είναι αυτή η τιμή, τόσο καλύτερη είναι η στοίχιση.
  - ***E-value -*** δείχνει πόσες φορές κατά τύχη θα περίμενε κανείς να βρει μια τόσο καλή στοίχιση για δύο ακολουθίες.
    - Σημειώστε ότι η τιμή E σε αυτή τη μέθοδο είναι πολύ μικρότερης σημασίας από ό,τι στο BLAST, που σχετίζεται με την αναζήτηση μιας βάσης δεδομένων.
    - Μια καλή τιμή E πρέπει να είναι κάτω από 10^{-4}

Ερμηνεία των αποτελεσμάτων

- Η ίδια η στοίχιση περιέχει τρεις τύπους πληροφοριών
  - ***Δείκτης υπολειμμάτων -*** βρίσκεται στη σειρά πάνω από την ακολουθία.
  - ***Η ίδια η στοίχιση*** - τα κενά στην ακολουθία αντιπροσωπεύονται με παύλες.
  - ***Ταυτότητα και ομοιότητα -*** η σειρά μεταξύ των δύο αλληλουχιών που συγκρίνονται.
    - ***Το σύμβολο (_) σημαίνει ταυτότητα -*** Η ταυτότητα μπορεί να παρατηρηθεί κατά τη σύγκριση αλληλουχιών πρωτεΐνης και DNA
    - ***(.) σημαίνει ομοιότητα -*** μπορεί να υπάρχει μόνο στην πρωτεϊνική ευθυγράμμιση.
    - Ομοιότητα δίνει το κλάσμα των αμινοξέων που έχουν παρόμοια φυσικοχημικά χαρακτηριστικά.
- Δύο υπολείμματα είναι παρόμοια όταν η βαθμολογία αντικατάστασής τους είναι μεγαλύτερη από 0.

Ερμηνεία των αποτελεσμάτων

- Για να αξιολογήσετε την ποιότητα της ευθυγράμμισής σας, δώστε προσοχή στο καθορισμένο μέγεθος.
  - Χαμηλή πολυπλοκότητα θραυσμάτων, χαμηλά επίπεδα επανάληψης σε αλληλουχίες πρωτεϊνών, επαναλαμβανόμενα μοντέλα υδρόφοβων/υδρόφιλων υπολειμμάτων κ.λπ.
- Τα προγράμματα βιοπληροφορικής δεν είναι τέλεια. Μπορούν να κάνουν λάθη και να δώσουν λανθασμένα αποτελέσματα.
  - Η πιο ανώτερη στοίχιση που δημιουργεί το Lalign απεικονίζει καλά αυτόν τον περιορισμό. Λόγω των πληροφοριών που παρέχει το Dot plot, εμπιστευόμαστε αυτήν την ευθυγράμμιση.
  - Οι δύο σεκάνς είναι τόσο μακρινές και κάπως παρόμοιες που πρέπει να είμαστε αρκετά προσεκτικοί στις ερμηνείες που κάνουμε.

Εκτέλεση καθολικής ευθυγράμμισης

- Μια συνολική ευθυγράμμιση είναι αυτό που υποδηλώνει το όνομα - μια ευθυγράμμιση κάθε αμινοξέος ή νουκλεοτιδίου στις αλληλουχίες σας.

- Σημαντικό σε πολλαπλές γραμμές.

- Οι καθολικές ευθυγραμμίσεις δεν είναι χρήσιμες για την ανίχνευση ομοιοτήτων μεταξύ δύο ακολουθιών, καθώς η στατιστική μέθοδος εκτίμησης E-value δεν ισχύει για αυτές.

- Σε μια παγκόσμια ευθυγράμμιση, δεν υπάρχουν αμινοξέα ή νουκλεοτίδια που εξαφανίζονται μυστηριωδώς.

Εκτέλεση καθολικής ευθυγράμμισης

- Τρεις λόγοι για να προτιμήσετε να χρησιμοποιήσετε μια παγκόσμια συμμαχία:
  - Έλεγχος για μικρές διαφορές μεταξύ δύο ακολουθιών .
    - Αυτό μπορεί να συμβεί με ακολουθίες που έχετε χειριστεί και πιθανώς έχετε αλλάξει. Η παγκόσμια κλίμακα είναι ο καλύτερος τρόπος εντοπισμού πιθανών προβλημάτων.
  - Ανάλυση πολυμορφισμών (π.χ. SNPs) μεταξύ στενά συγγενών αλληλουχιών.
  - Σύγκριση δύο αλληλουχιών που επικαλύπτονται εν μέρει.
- Η χρήση του Lalign για τη διεξαγωγή μιας καθολικής ευθυγράμμισης είναι σαν να διεξάγετε μια τοπική ευθυγράμμιση.
  - Το μόνο πρόσθετο πράγμα που πρέπει να κάνετε είναι να κάνετε κλικ στο ***Global without End-Gap Penalty***.

Ευθυγράμμιση πρωτεΐνης με αλληλουχίες DNA

- Τα εργαλεία ευθυγράμμισης που εξετάσαμε είναι αποτελεσματικά μόνο όταν συγκρίνονται αλληλουχίες παρόμοιου τύπου - πρωτεΐνη με πρωτεΐνη ή DNA με DNA.
- Μερικές φορές, είναι απαραίτητο να συγκρίνουμε μια πρωτεΐνη με ένα θραύσμα DNA (για παράδειγμα, το αρχικό της γονίδιο).
  - Ινστιτούτο Παστέρ (http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms)
  - Δρ. Ο Peer Bork και η ομάδα του στο Ευρωπαϊκό Εργαστήριο Μοριακής Βιολογίας στη Χαϊδελβέργη (http://www.bork.embl.de/pal2nal/).
- Και οι δύο διακομιστές απαιτούν να έχετε εκ των προτέρων τις αλληλουχίες πρωτεΐνης και DNA.
  - Protogene (http://www.tcoffee.org/) εξάγει αυτόματα την αλληλουχία DNA που αντιστοιχεί σε μια δεδομένη πρωτεΐνη .

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Σας ευχαριστώ για την προσοχή σας !

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

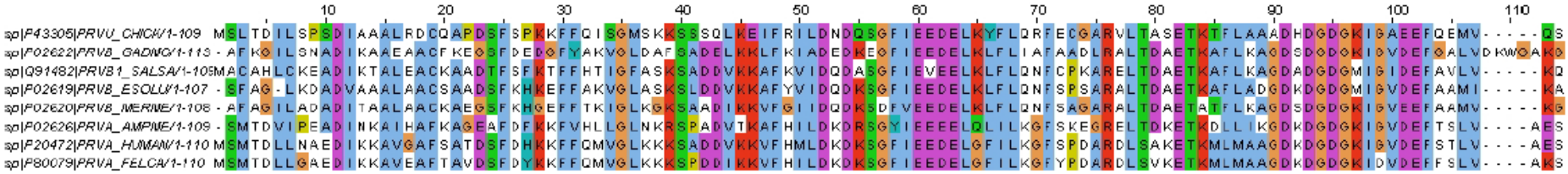
Θέμα 3. Πρόβλημα αναδίπλωσης πρωτεΐνης

Μάθημα 3. Στοίχιση πολλαπλών ακολουθιών

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Πολλαπλή ευθυγράμμιση

- Ιδανικό για τη μελέτη μιας οικογένειας ακολουθιών με κοινό πρόγονο.



  - Οι μέθοδοι πολλαπλών ακολουθιών δεν λειτουργούν καλά για τη συναρμολόγηση ακολουθίας σε έργα αλληλουχίας.
  - Δεν είναι κατάλληλο για πολλαπλάσια σύντομων, μερικώς επικαλυπτόμενων αλληλουχιών ή εάν θέλετε να μετατρέψετε EST (Ετικέτες εκφρασμένης ακολουθίας) σε αλληλουχία γονιδίου.

Πολλαπλή ευθυγράμμιση

- Χρησιμοποιούνται εξειδικευμένα εργαλεία για τη συναρμολόγηση ακολουθίας, όπως:
  - Phred και Phrap (http://www.phrap.org/)
  - Ανοχή σε σφάλματα Συνθεσάιζερ (http://genome.cs.mtu.edu/Tools.html).
- Δεν απαιτείται η εκτέλεση μιας πολλαπλής στοίχισης όταν η ακολουθία που μας ενδιαφέρει δεν έχει ομολογία με καμία ακολουθία στη βάση δεδομένων.
  - Μπορείτε να προσπαθήσετε να βρείτε πολλές ακολουθίες χρησιμοποιώντας λειτουργικά κριτήρια ή αναζήτηση BLAST.
- Μπορείτε να δημιουργήσετε μια σωστή πολλαπλή στοίχιση μόνο εάν υπάρχουν κατάλληλες ακολουθίες.

Co-funded by the Erasmus+ Programme of the European Union

Πολλαπλή ευθυγράμμιση

- Η βασική ιδέα πίσω από μια πολλαπλή ευθυγράμμιση είναι η τοποθέτηση αμινοξέων ή νουκλεοτιδίων στην ίδια στήλη με βάση ομοιότητες σύμφωνα με κάποιο κριτήριο.

- Τρία βασικά κριτήρια για την κατασκευή πολλαπλών ευθυγραμμίσεων

Κριτήριο	Εννοια
Δομική ομοιότητα	Αμινο οξέα με το ίδιο ρόλος σε κάθε δομή είναι σε ένα στήλη . Ακρίβεια τοποθέτησης τα προγράμματα είναι τα μόνα που χρησιμοποιούν αυτό το κριτήριο.
Λειτουργική ομοιότητα	Αμινο οξέα ή νουκλεοτίδια με την ίδια λειτουργία είναι σε ένα στήλη . Δεν υπάρχει _ αυτόματο πρόγραμμα που ρητά χρήσεις αυτό το κριτήριο, αλλά εάν οι πληροφορίες είναι διαθέσιμες , εσείς μπορώ δύναμη μερικοί προγράμματα για να το ακολουθήσετε , ή εσείς μπορώ επεξεργασία τα δικα σου ευθυγραμμία χειροκίνητα .
Ομοιότητα ακολουθίας	Αμινο οξέα σε ένα στήλη είναι αυτές που δίνουν μια στοίχιση με μέγιστη ομοιότητα . Πλέον προγράμματα χρήση αλληλουχία ομοιότητα γιατί είναι το πιο εύκολο κριτήριο. Όταν οι ακολουθίες είναι κοντά σχετίζεται με δικα τους δομική και λειτουργική οι ομοιότητες είναι ισοδύναμες με τη σειρά ομοιότητα

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πολλαπλή ευθυγράμμιση

Εφαρμογή	Διαδικασία
Παρέκταση	Μια καλή πολλαπλή ευθυγράμμιση μπορεί να σας πείσει εάν μια άγνωστη αλληλουχία είναι πραγματικό μέλος μιας οικογένειας πρωτεϊνών.
P hylogenetic ανάλυση	Εάν επιλέξετε σωστά τις αλληλουχίες για μια δεδομένη πολλαπλή στοίχιση, μπορείτε να ξαναγράψετε το ιστορικό αυτών των πρωτεϊνών.
Ταυτοποίηση κινήτρων	Με την ανίχνευση Πολλά διατηρημένο θέσεις , μπορείτε να ορίσετε μια περιοχή που είναι κοινή για μια συγκεκριμένη περιοχή χαρακτηριστικό .
Αναγνώριση	Είναι δυνατό μια πολλαπλή ευθυγράμμιση να γίνει ένα προφίλ που περιγράφει μια οικογένεια πρωτεΐνης ή έναν τομέα πρωτεΐνης (PSSM). Μπορείτε να χρησιμοποιήσετε αυτό το προφίλ για να σαρώσετε βάσεις δεδομένων για νέα μέλη της οικογένειας.
Ρυθμιστικά στοιχεία DNA	Μπορείτε να μετατρέψετε μια πολλαπλή ευθυγράμμιση DNA σε μια μήτρα παράγοντα μεταγραφής και να σαρώσετε άλλες αλληλουχίες DNA για δυνητικά παρόμοιες θέσεις δέσμευσης.
Πρόβλεψη δομών	Μια καλή πολλαπλή ευθυγράμμιση μπορεί να σας δώσει μια τέλεια πρόβλεψη της δευτερογενούς δομής, τόσο της πρωτεΐνης όσο και του RNA. Μερικές φορές μπορεί να βοηθήσει στην κατασκευή ενός τρισδιάστατου μοντέλου.
Ανάλυση SNP	Ένας αριθμός αλληλόμορφων γονιδίων έχουν συχνά διαφορετικές αλληλουχίες αμινοξέων. Η πολλαπλή ευθυγράμμιση μπορεί να βοηθήσει στην πρόβλεψη του εάν οι μη συνώνυμοι πολυμορφισμοί σε μεμονωμένα νουκλεοτίδια είναι ωφέλιμοι ή επιβλαβείς.
Ανάλυση PCR	Ένα πολλαπλάσιο ευθυγραμμία μπορώ βοήθεια ταυτίζεσαι λιγότερο _ εκφυλισμένος μέρη μιας πρωτεΐνης οικογένεια , η οποία θα επιτρέπουν την ανακάλυψη νέων _ μέλη αυτής της οικογένειας με PCR (πολυμεράση αλυσίδα αντίδραση). Ή τουλάχιστον φτιάξε τα δικα σου αστάρια πιο βέλτιστη.

- Βασικές εφαρμογές πολλαπλής ευθυγράμμισης.
  - Ανεξάρτητα από την επιλογή ή το κίνητρό σας, θα πρέπει να έχετε κατά νου ότι ο λόγος για τη χρήση πολλαπλής ευθυγράμμισης είναι ότι δίνει μια στιγμιαία εικόνα των δυνάμεων που διαμορφώνουν την εξέλιξη !

Co-funded by the Erasmus+ Programme of the European Union

Πολλαπλή ευθυγράμμιση

- Μια πιθανή υπόθεση σε πολλαπλή ευθυγράμμιση:
  - Οι διατηρημένες θέσεις (στήλες όπου όλες οι αλληλουχίες περιέχουν το ίδιο αμινοξύ ή νουκλεοτίδιο) είναι πιο σημαντικές για τη λειτουργία από τις μη διατηρημένες θέσεις (στήλες όπου οι αλληλουχίες περιέχουν διαφορετικά αμινοξέα ή νουκλεοτίδια).
- Η πολλαπλή ευθυγράμμιση είναι ένας πολύ καλός τρόπος για να παρουσιάσετε τα αποτελέσματά σας.
  - Σας επιτρέπει να τοποθετείτε πολλές πληροφορίες σε ένα μόνο μοντέλο και να εντοπίζετε εύκολα ασυνέπειες ή πιθανά προβλήματα.

Co-funded by the Erasmus+ Programme of the European Union

Επιλογή κατάλληλων ακολουθιών

Πρόβλημα	Λύση
Πρωτεΐνες ή DNA	Χρησιμοποιήστε πρωτεΐνες όταν είναι δυνατόν. Εσείς μπορώ μετατρέπω τους πίσω στο DNA μετά εκτελώντας πολλαπλούς ευθυγραμμία
Πολλαπλές ακολουθίες	Ξεκινήστε με 5-10 ακολουθίες, αποφύγετε την ευθυγράμμιση με περισσότερες από 30.
Πολλαπλές διαφορετικές ακολουθίες	Συχνά υπάρχει πρόβλημα με αλληλουχίες που είναι λιγότερο από 30 τοις εκατό πανομοιότυπες με περισσότερες από τις μισές άλλες αλληλουχίες της ομάδας.
Πανομοιότυπες σειρές	Δεν υπάρχει ποτέ καμία χρησιμότητα για αυτούς. Εάν έχετε λόγο να εργαστείτε με τέτοιες ακολουθίες, εξαιρέστε αυτές που είναι πανομοιότυπες περισσότερο από 90% από άλλες ακολουθίες της ομάδας.
Ημιτελείς ακολουθίες	Τα πολλαπλά προγράμματα ευθυγράμμισης προτιμούν ακολουθίες που έχουν σχετικά ομοιόμορφο μήκος. Συχνά έχουν πρόβλημα να συγκρίνουν ολόκληρες ακολουθίες και ημιτελή θραύσματα.
Πολλαπλοί επαναλαμβανόμενοι τομείς	Ακολουθίες με πολλαπλά αλλεπάλληλος τομείς συνήθως αιτία προβλήματα για πολλά ευθυγραμμία προγράμματα (ειδικά αν ο αριθμός των τομέων είναι διαφορετικός). Σε αυτήν την περίπτωση , ίσως είναι καλύτερο να εξαγάγετε τους τομείς με το Dotlet ή Ευθυγραμμίστε και κάντε πολλαπλούς

- Επιλέξτε προσεκτικά τις ακολουθίες που θέλετε να ευθυγραμμίσετε.

Co-funded by the Erasmus+ Programme of the European Union

Χρησιμοποιώντας αλληλουχίες DNA ή πρωτεϊνών

- Εάν ενδιαφέρεστε για μη κωδικοποιητικές αλληλουχίες, προφανώς δεν έχετε άλλη επιλογή από το να χρησιμοποιήσετε το DNA.
  - Οι μη κωδικοποιητικές αλληλουχίες DNA είναι δύσκολο να ευθυγραμμιστούν!
- Μπορείτε να χρησιμοποιήσετε τοπικές μεθόδους πολλαπλής ευθυγράμμισης, όπως το δείγμα Gibbs ή το Pratt.
- Οι μέθοδοι πολλαπλής ευθυγράμμισης είναι πιο κατάλληλες για πρωτεΐνες.
- Στη φυλογενετική ανάλυση πολλαπλών κωδικοποιητικών αλληλουχιών DNA:
  1. Μετατρέψτε τις αλληλουχίες του DNA σας σε πρωτεΐνες.
  2. Εκτελέστε πολλαπλή ευθυγράμμιση αυτών των πρωτεϊνών.
  3. Μετατρέψτε τις πρωτεΐνες που χρησιμοποιούνται στην πολλαπλή ευθυγράμμιση πίσω σε DNA χρησιμοποιώντας pal2nal

  (http://www.bork.embl.de/pal2nal/) ή Protogene (http://www.tcoffee.org/).

Επιλογή του σωστού αριθμού ακολουθιών

- Δεν υπάρχει απόλυτη απάντηση.
- Ξεκινήστε με έναν σχετικά μικρό αριθμό ακολουθιών.
- Εάν ξεκινήσετε με εκατοντάδες ακολουθίες, μπορεί να αντιμετωπίσετε προβλήματα αμέσως:
  - Ο υπολογισμός μιας μεγάλης ευθυγράμμισης είναι δύσκολος .
  - Η δημιουργία μιας μεγάλης συμμαχίας είναι επίσης δύσκολη.
  - Η εμφάνιση μεγάλης ευθυγράμμισης είναι δύσκολη.
  - Η χρήση μιας μεγάλης ευθυγράμμισης είναι αναποτελεσματική.
  - Η πραγματοποίηση μιας ακριβούς και μεγάλης ευθυγράμμισης είναι σχεδόν αδύνατη.
- Ξεκινήστε με λιγότερα και σταδιακά αυξήστε τον αριθμό των ακολουθιών με τις οποίες εκτελείτε πολλαπλές ευθυγραμμίσεις μέχρι να συμπεριλάβετε όλες τις ακολουθίες που σας ενδιαφέρουν.

Συμβιβασμός μεταξύ ομοιότητας και νέων πληροφοριών

- Αν νομίζετε ότι πολύ παρόμοιες ακολουθίες δίνουν πολύ καλές ευθυγραμμίσεις, έχετε δίκιο!
- Η εργασία με ακολουθίες που είναι πολύ απομακρυσμένες σε σχέση δεν είναι επίσης κατάλληλη.
- Δύο πράγματα που δεν αρέσουν πραγματικά σε πολλά προγράμματα ευθυγράμμισης:
  - Ακολουθίες που είναι πολύ διαφορετικές από τις υπόλοιπες στην ομάδα.
  - Ακολουθίες που απαιτούν μεγάλες εισαγωγές/διαγραφές για σωστή ευθυγράμμιση.
- Οι ακολουθίες θα πρέπει να σχετίζονται όσο το δυνατόν περισσότερο χωρίς να απαιτούνται πάρα πολλά διαλείμματα κατά τη διεξαγωγή μιας ευθυγράμμισης.

Συμβιβασμός μεταξύ ομοιότητας και νέων πληροφοριών

- Βασικά βήματα στην επιλογή ακολουθίας:
  1. Επιλέξτε πολλές ακολουθίες.
  2. Εκτελέστε πολλαπλή ευθυγράμμιση χρησιμοποιώντας έναν από τους διακομιστές.
  3. Εάν η ευθυγράμμισή σας φαίνεται καλή, κρατήστε τις ακολουθίες.
  4. Εάν η ευθυγράμμισή σας είναι δύσκολο να ερμηνευτεί:
     - Κοιτάξτε τις ακολουθίες με περισσότερες λεπτομέρειες - προσπαθήστε να αφαιρέσετε τις «προσβλητικές» ακολουθίες, αυτές που είναι πιο απομακρυσμένες σε σχέση ή αυτές που προκαλούν μεγάλες εισαγωγές/διαγραφές.
     - Εκτελέστε ξανά την ευθυγράμμιση με λιγότερες ακολουθίες.
     - Αποθηκεύστε το αρχικό σύνολο ακολουθιών προτού λάβετε ικανοποιητικά, εύκολα ερμηνευτικά αποτελέσματα.

Ρύθμιση του ονόματος της ακολουθίας

- Τα πολλαπλά προγράμματα στοίχισης δεν έχουν τυπικούς τρόπους χειρισμού ονομάτων ακολουθιών:
  - Μην χρησιμοποιείτε κενές θέσεις όταν ονομάζετε ακολουθίες.
  - Μη χρησιμοποιείτε ειδικούς χαρακτήρες. Συνεχίστε να χρησιμοποιείτε κανονικά γράμματα, αριθμούς και υπογράμμιση (_) για να αντικαταστήσετε τα κενά. Αποφύγετε όλους τους άλλους χαρακτήρες, ειδικά αυτούς που είναι πιο ελκυστικοί για ειδικές ακολουθίες (όπως @, #, _, L, κ.λπ.).
  - Μη χρησιμοποιείτε ονόματα μεγαλύτερα από 15 χαρακτήρες.
  - Μην δίνετε το ίδιο όνομα σε δύο διαφορετικές ακολουθίες που περιέχονται σε ένα σύνολο.

Ρύθμιση του ονόματος της ακολουθίας

- Υπάρχουν δύο τύποι ακολουθιών που μπορείτε να ενσωματώσετε σε μια ενιαία πολλαπλή στοίχιση:
  - Χαρακτηρισμένες/Σχολιασμένες ακολουθίες - Πρόκειται για ακολουθίες για τις οποίες υπάρχουν καλοί σχολιασμοί και πειραματικά δεδομένα για την υποστήριξή τους.
  - Άγνωστες ακολουθίες - Αυτή η κατηγορία περιλαμβάνει τις ακολουθίες σας καθώς και ακολουθίες από διαφορετικές βάσεις δεδομένων.
- Χρησιμοποιώντας το BLAST, μπορείτε να αναζητήσετε σε μια βάση δεδομένων ακολουθίες που είναι ομόλογες (παρόμοιες) με το ερώτημά σας.

Address	Opportunities
http://www.expasy.ch/tools/blast/	Extract whole sequences, export sequences to FESTA format, request sequences to ClustalW, Tcoffee or MAFFT
http://blast.ncbi.nlm.nih.gov/	Extract whole sequences, extract fragments, export sequences in FESTA format, request sequences to ClustalW
http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-page+srsq2+-noSession	Заявяване на секвенции към ClustalW

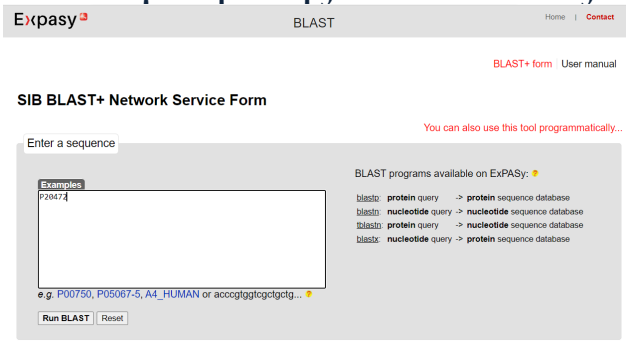
Επιλογή ακολουθιών μέσω του διακομιστή

- Μπορείτε να χρησιμοποιήσετε το ExPASy για λήψη αλληλουχιών πρωτεΐνης.
  - Εάν ενδιαφέρεστε να συλλέξετε αλληλουχίες DNA, χρησιμοποιήστε το NCBI, BLAST (http://blast.ncbi.nlm.nih.gov).
  - Ο διακομιστής BLAST στην τοποθεσία PBIL είναι πολύ παρόμοιος με το ExpASy .

1. Εισαγάγετε την ακόλουθη διεύθυνση στο πρόγραμμα περιήγησής σας:

http://web.expasy.org/blast/

2. Εισαγάγετε τον αριθμό αναγνώρισης ακολουθίας ***P20472*** .



Επιλογή ακολουθιών μέσω του διακομιστή

- Μπορείτε να χρησιμοποιήσετε το ExPASy για λήψη αλληλουχιών πρωτεΐνης.

  Διατηρήστε τις προεπιλεγμένες επιλογές (***Ολοκληρωμένη βάση δεδομένων***) για την αναπτυσσόμενη λίστα **UniProt Knowledgebase (UniProtKB) .**

  3. Μεταβείτε στην καρτέλα **Επιλογή επιλογών** και ορίστε τον αριθμό των καλύτερων ακολουθιών που εμφανίζονται σε 1000.
  4. Στην ίδια καρτέλα, ορίστε τον αριθμό των καλύτερων γραμμών που εμφανίζονται σε 1000.
  5. Κάντε κλικ στο κουμπί **Εκτέλεση BLAST** .
  6. Επιλέξτε τις ακολουθίες που θέλετε.
     - Δεν υπάρχει απόλυτος κανόνας κατά την επιλογή ακολουθιών, αλλά μπορείτε να χρησιμοποιήσετε τις ακόλουθες οδηγίες:

Co-funded by the Erasmus+ Programme of the European Union

Επιλογή ακολουθιών μέσω του διακομιστή

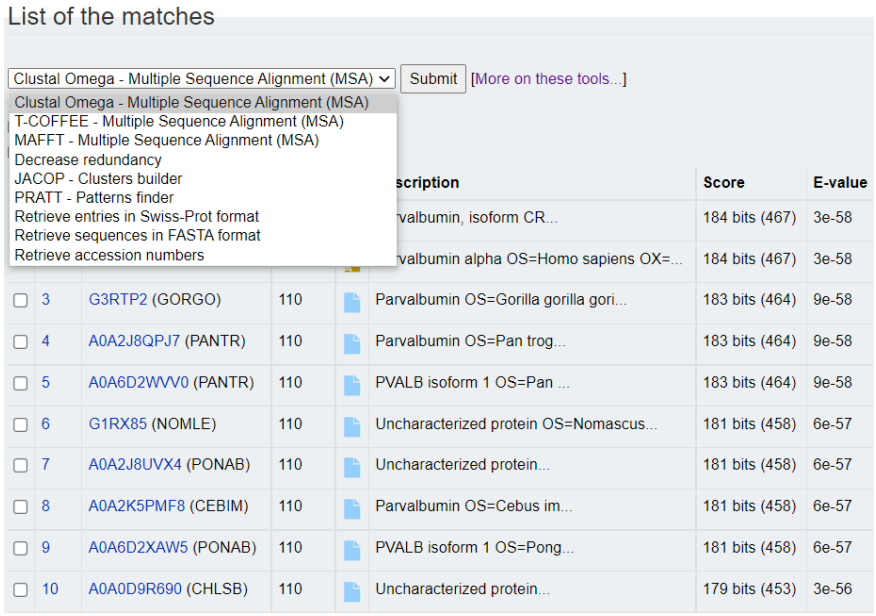
- Μπορείτε να χρησιμοποιήσετε το ExPASy για λήψη μόνο πρωτεϊνικών αλληλουχιών.
  7. Κάντε κύλιση προς τα κάτω στη σελίδα για να επιλέξετε τις ακολουθίες που θέλετε.
     - Δεν υπάρχει απόλυτος κανόνας κατά την επιλογή ακολουθιών, αλλά μπορείτε να χρησιμοποιήσετε τις ακόλουθες οδηγίες:
       - Επιλέξτε την επάνω ακολουθία.
       - Στην πρώτη ανάλυση, θα πρέπει να επιλέξετε δέκα ακολουθίες ή λιγότερες. Στην ιδανική περίπτωση, οι δέκα ακολουθίες θα πρέπει να έχουν τιμή E μεταξύ 10^{-40} και 10^{-5}.
       - Πριν επιλέξετε μια ακολουθία, ελέγξτε ότι είναι παρόμοια σε όλο το μήκος του ερωτήματός σας.

Επιλογή ακολουθιών μέσω του διακομιστή

- Μπορείτε να χρησιμοποιήσετε το ExPASy για λήψη μόνο πρωτεϊνικών αλληλουχιών.
  8. Από το αναπτυσσόμενο μενού Αποστολή **επιλεγμένων ακολουθιών , επιλέξτε τη μέθοδο με την οποία θέλετε να εξαγάγετε τις επιλεγμένες ακολουθίες:**
     - ***FASTA*** – Δημιουργεί ένα αρχείο που περιέχει την ακολουθία σας σε μορφή FASTA.
     - ***ClustalW*** , ***Tcoffee*** και ***MAFFT*** – Αυτά είναι πακέτα πολλαπλής ευθυγράμμισης που εκτελούνται στον διακομιστή EMBnet .
     - ***Μειώστε τις πλεονάζουσες επαναλήψεις -*** Αυτή η επιλογή θα εξαγάγει τις πιο σημαντικές ακολουθίες από ένα σύνολο δεδομένων - μια καλή επιλογή εάν είναι διαθέσιμες πάρα πολλές ακολουθίες και δεν ξέρετε ποιες να επιλέξετε.
     - ***Pratt*** – Σαρώνει για διατηρημένα κίνητρα με τη δηλωμένη ακολουθία χωρίς να τα περιλαμβάνει στη στοίχιση.

Επιλογή ακολουθιών μέσω του διακομιστή

- Μπορείτε να χρησιμοποιήσετε το ExPASy για λήψη μόνο πρωτεϊνικών αλληλουχιών.
  8. Από το αναπτυσσόμενο μενού Αποστολή **επιλεγμένων ακολουθιών , επιλέξτε τη μέθοδο με την οποία θέλετε να εξαγάγετε τις επιλεγμένες ακολουθίες**



Συλλέγοντας μια διάσημη συλλογή από ελβετικές ακολουθίες

- Εάν γνωρίζετε ήδη το όνομα ή τον αριθμό πρόσβασης καθεμιάς από τις ακολουθίες που θέλετε να συμπεριλάβετε στην πολλαπλή στοίχιση και εάν είναι σε Swiss- Prot ή TrEMBL , μπορείτε να αποκτήσετε πρόσβαση απευθείας σε αυτές χρησιμοποιώντας ειδικά διαδικτυακά εργαλεία ExpASy .
  1. Εισαγάγετε την ακόλουθη διεύθυνση στο πρόγραμμα περιήγησής σας: http://www.uniprot.org/batch/
  2. Εισαγάγετε τους αριθμούς πρόσβασης για τις επιθυμητές ακολουθίες στο παράθυρο Παροχή των αναγνωριστικών σας .
  3. Επιλέξτε τις επιθυμητές ακολουθίες, κάντε κλικ στο κουμπί Λήψη , επιλέξτε ***FASTA*** στην αναπτυσσόμενη λίστα **Μορφοποίηση και κάντε κλικ στο** κουμπί **Μετάβαση .**
  4. Αποθηκεύστε τα αποτελέσματα στον υπολογιστή σας.

Επιλογή κατάλληλης μεθόδου πολλαπλής ευθυγράμμισης

- Η κατασκευή μιας καλής πολλαπλής ευθυγράμμισης απαιτεί εξάσκηση.
- ClustalW (η τελευταία έκδοση ονομάζεται ClustalOmega) - το πιο ευρέως χρησιμοποιούμενο πακέτο πολλαπλής ευθυγράμμισης.
- Tcoffee - ένα από τα τελευταία προγράμματα πολλαπλής ευθυγράμμισης που δημιουργήθηκαν:
  - Συνδυασμός ακολουθιών και δομών ,
  - Αξιολόγηση μιας ευθυγράμμισης ,
  - Συγχώνευση πολλών εναλλακτικών πολλαπλών ευθυγραμμίσεων σε ένα μόνο αποτέλεσμα .

Co-funded by the Erasmus+ Programme of the European Union

Επιλογή κατάλληλης μεθόδου πολλαπλής ευθυγράμμισης

- ClustalOmega / ClustalW - το πιο συχνά χρησιμοποιούμενο πρόγραμμα για την εκτέλεση πολλαπλών ευθυγραμμίσεων .
- Το ClustalOmega / ClustalW χρησιμοποιεί αυξητικές μεθόδους για τη σύναψη της ευθυγράμμισης.
  - Αντί να συγκρίνει όλες τις ακολουθίες ταυτόχρονα, τις προσθέτει μία προς μία.
- Υπάρχουν πολλοί διακομιστές ClustalOmega / ClustalW .
- Το ClustalOmega / ClustalW είναι ένα τυπικό παράδειγμα προγράμματος που μπορεί να παράγει ένα καλό αποτέλεσμα με τις προεπιλεγμένες ρυθμίσεις του .
  - Αλλάξτε τις ρυθμίσεις μόνο εάν θέλετε να αλλάξετε τη μορφή εξόδου του προγράμματος .

Επιλογή κατάλληλης μεθόδου πολλαπλής ευθυγράμμισης

- Πώς λειτουργεί το ClustalOmega / ClustalW ;
  - Η μέθοδος που χρησιμοποιεί το πρόγραμμα είναι γρήγορη και απλή - προοδευτική ευθυγράμμιση.
  - Αρχικά, όλες οι ακολουθίες συγκρίνονται 2 επί 2 και το πρόγραμμα τις τακτοποιεί παρόμοια.
    - Αυτή η ομαδοποίηση μοιάζει με φυλογενετικό δέντρο, ονομάζεται **δέντρο οδηγών** και αποθηκεύεται σε ένα αρχείο με την επέκταση **. dnd .**

- Το *W* στο ClustalW προέρχεται από *τα Βάρη* - σε κάθε ακολουθία στη στοίχιση δίνεται ένα "βάρος" ανάλογο με τις νέες πληροφορίες που συνεισφέρει.

Εκτελέστε τον διακομιστή EBI

- Πριν ξεκινήσετε να εργάζεστε με το ClustalOmega / ClustalW , πρέπει να συλλέξετε όλες τις ακολουθίες με τις οποίες θέλετε να εργαστείτε.
  - Ο πιο βολικός τρόπος είναι να εισάγετε τις ακολουθίες σε μορφή FASTA, αλλά το ClustalOmega / ClustalW δέχεται άλλες μορφές, συμπεριλαμβανομένων των Swiss- Prot και PIR, καθώς και ήδη ευθυγραμμισμένες ακολουθίες στις πιο κοινές μορφές μετά από πολλαπλή στοίχιση.
- Εάν θέλετε να χρησιμοποιήσετε αποτελεσματικά το ClustalOmega / ClustalW , έχετε υπόψη σας τα εξής:
  - Εάν καθορίσετε μια ομάδα ακολουθιών που έχουν ήδη περάσει πολλαπλή στοίχιση, το ClustalOmega / ClustalW δεν αφαιρεί τα υπάρχοντα "κενά".
  - Η σειρά των ακολουθιών στο αρχείο που υποβάλλετε για επεξεργασία μπορεί μερικές φορές να επηρεάσει την τελική στοίχιση.

Co-funded by the Erasmus+ Programme of the European Union

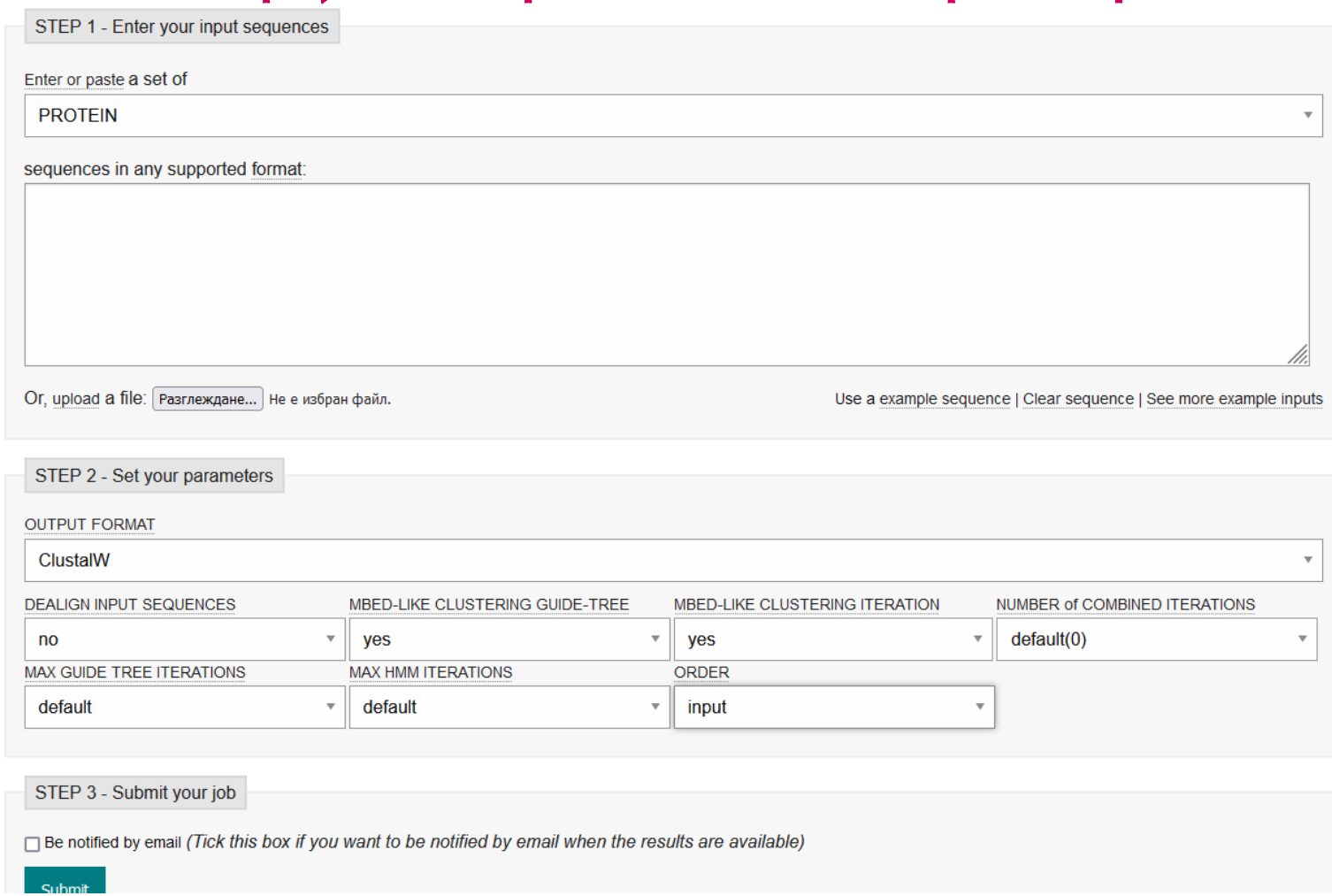
Εργασία με τον διακομιστή

1. Μεταβείτε στη σελίδα διακομιστή EBI ClustalOmega : https://www.ebi.ac.uk/Tools/msa/clustalo/
2. Εισαγάγετε τις ακολουθίες που έχετε συγκεντρώσει στο κατάλληλο πεδίο ή μεταφορτώστε το αρχείο που τις περιέχει στο ***πεδίο Μεταφόρτωση αρχείου*** .
3. Επιλέξτε μια ***μορφή εξόδου*** .
   - Ο πιο ασφαλής τρόπος είναι να χρησιμοποιήσετε το " ***Aln Without Numbers*** "
   - Για να επιλέξετε το καταλληλότερο για εσάς, ΜΗΝ επανεκκινήσετε ολόκληρη την ευθυγράμμιση!
   - Μπορείτε εύκολα να διαμορφώσετε ξανά τις ευθυγραμμίσεις σας χρησιμοποιώντας ένα από τα διαδικτυακά εργαλεία αναδιαμόρφωσης (π.χ. EMBOSS Seqret).

Εργασία με τον διακομιστή

4. Επιλέξτε την ***Εισαγωγή*** από ***την έξοδο Σειρά*** μενού .
   - Εάν επιλέξετε την επιλογή ***Εισαγωγή*** , το αρχείο ευθυγράμμισης ClustalOmega / ClustalW εμφανίζει τις ακολουθίες με την αρχική σειρά, i . μι. τη σειρά με την οποία εισήχθησαν.
   - Εάν επιλέξετε την επιλογή ***Στοίχιση*** , οι ακολουθίες εμφανίζονται με τη σειρά που τις τακτοποίησε το λογισμικό όταν δημιούργησε το δέντρο οδηγών - οι στενά συνδεδεμένες ακολουθίες βρίσκονται κοντά μεταξύ τους.
   - Είναι καλύτερο να τακτοποιήσετε εκ των προτέρων τις ακολουθίες σας με τον πιο κατατοπιστικό τρόπο στο αρχείο που θα υποβάλετε και, στη συνέχεια, επιλέξτε την επιλογή ***Εισαγωγή .***
5. Κάντε κλικ στην ***επιλογή Υποβολή*** κουμπί.

Εργασία με τον διακομιστή



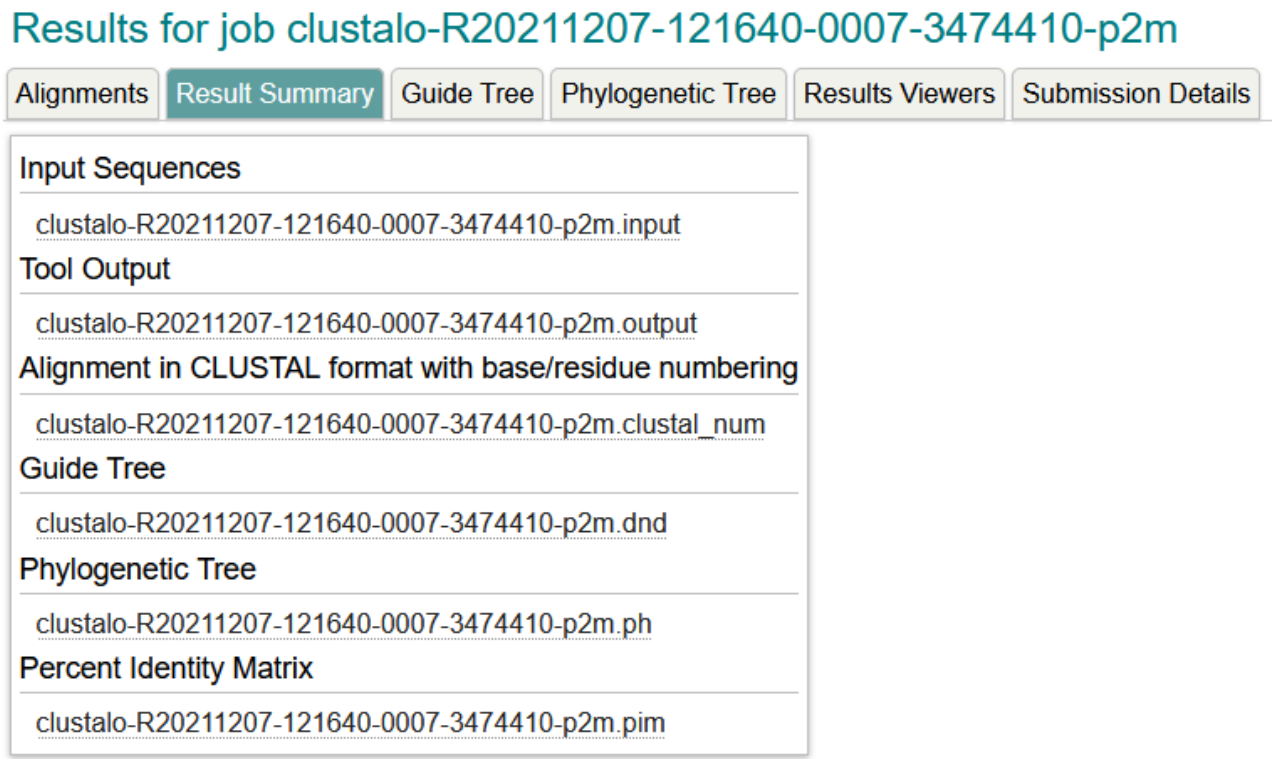
Co-funded by the Erasmus+ Programme of the European Union

Εργασία με τον διακομιστή

6. Αποθηκεύστε τα αποτελέσματά σας - Τα αποτελέσματα εμφανίζονται σε έναν πίνακα με πολλές ενότητες με υπερσυνδέσμους. Κάνοντας κλικ πάνω τους ανοίγει:
   - ***Αρχείο εξόδου*** - Βαθμολογίες Pairwise.
   - ***Στοίχιση*** - πολλαπλή στοίχιση - μπορείτε να κάνετε προεπισκόπηση αυτής της ενότητας στην οθόνη ή να την αποθηκεύσετε ως αρχείο κειμένου.
   - ***Αρχείο δέντρου οδηγού*** - Οι περισσότεροι διακομιστές Web ClustalOmega / ClustalW (συμπεριλαμβανομένου αυτού στο EBI) εμφανίζουν το δέντρο οδηγών ως *. dnd* αρχείο.
   - ***Το αρχείο εισόδου σας*** - το αρχείο με τις ακολουθίες που εισαγάγατε στο ClustalW .

Εργασία με τον διακομιστή

- Σελίδα αποτελεσμάτων



Co-funded by the Erasmus+ Programme of the European Union

Αλλαγή στις παραμέτρους

Parameter	Effect
Substitution matrix	These matrices estimate the value of mutations in the allignment. If you select a category such as PAM or BLOSUM, ClustalW automatically selects the most adapted index. It is difficult to predict the effect of changing the matrix and there is no ideal matrix. If your sequences have high identity, changing the matrix will have no effect. If the alignment you get is difficult to interpret, it may be useful to change the matrix from BLOSUM to PAM. These matrices measure the probability of the bases changing/mutating over the course of evolution, thus providing an estimate when comparing them between sequences in an alignment.
Gap-opening penalty (GOP)	This parameter evaluates the opening of "gaps"/sequence breaks in the alignment. The larger the set value, the harder it is to insert a "gap" into your alignment. The GOP is applying once for opening each gap. If you change it, it won't have much effect, because ClustalW re-sets these values automatically.
Gap-extension penalty (GEP)	This parameter controls the size of the gaps in the alignment. It is not possible to predict the optimal GOP/GEP pair, but for each protein family, for example, one exists and can only be determined by experience.

- Τρεις παραμέτρους που μπορείτε να αλλάξετε στη γραμμή σας
  - *S αντικατάσταση μήτρες*
    ,
  - *G ap-άνοιγμα ποινικές ρήτρες*
  - *G ap-επέκταση ποινικές ρήτρες*

Αλλαγή στις παραμέτρους

- Ποτέ μην αλλάζετε παραμέτρους μόνο και μόνο για να κάνετε τα ClustalOmega / ClustalW να παράγουν την ευθυγράμμιση που γνωρίζετε ότι είναι σωστή!
  - Στη συνέχεια, μπορείτε να το επεξεργαστείτε με ένα από τα προγράμματα για αυτόν τον σκοπό.
- Οι αλλαγές παραμέτρων έχουν νόημα μόνο εάν θέλετε να κατανοήσετε πώς μικρές αλλαγές μπορούν να βελτιώσουν τη συνολική εμφάνιση της ευθυγράμμισής σας ως μπλοκ ή ασθενώς διατηρημένες θέσεις.

Co-funded by the Erasmus+ Programme of the European Union

Σύγκριση αλληλουχιών και δομών με το Tcoffee

- Το Tcoffee είναι μια σύγχρονη μέθοδος για πολλαπλή ευθυγράμμιση ακολουθιών.
  - Χρησιμοποιεί μια αρχή κάπως παρόμοια με αυτή του ClustalOmega , αλλά δίνει πολύ πιο ακριβείς ευθυγραμμίσεις σε βάρος ενός ελαφρώς μεγαλύτερου χρόνου εκτέλεσης.
- Το Tcoffee κάνει προοδευτική ευθυγράμμιση, αλλά συγκρίνει τμήματα από ένα ολόκληρο σύνολο ακολουθιών.
- Εκτός από την ακρίβεια, άλλα μεγάλα πλεονεκτήματα του Tcoffee είναι:
  - Η ικανότητά του να συγκρίνει ακολουθίες και δομές (EXPRESSO).
  - την ικανότητα αξιολόγησης της ακρίβειας της ευθυγράμμισης (CORE).
  - τη δυνατότητα συνδυασμού πολλών εναλλακτικών πολλαπλών ευθυγραμμίσεων σε μία (Mcoffee).

Co-funded by the Erasmus+ Programme of the European Union

Σύγκριση αλληλουχιών και δομών με το Tcoffee

- Προγράμματα διαθέσιμα στο www.tcoffee.org .

Program	With it you can
TCOFFEE	make multiple alinements with Tcoffee
CORE	evaluate the reliability of an existing multiple alignment
MCOFFEE	run any multiple alignment analyses and combine all results into one final alignment
EXPRESSO	include any available structural information in your alignment. If there are known structures, it makes the best alignment of sequences.

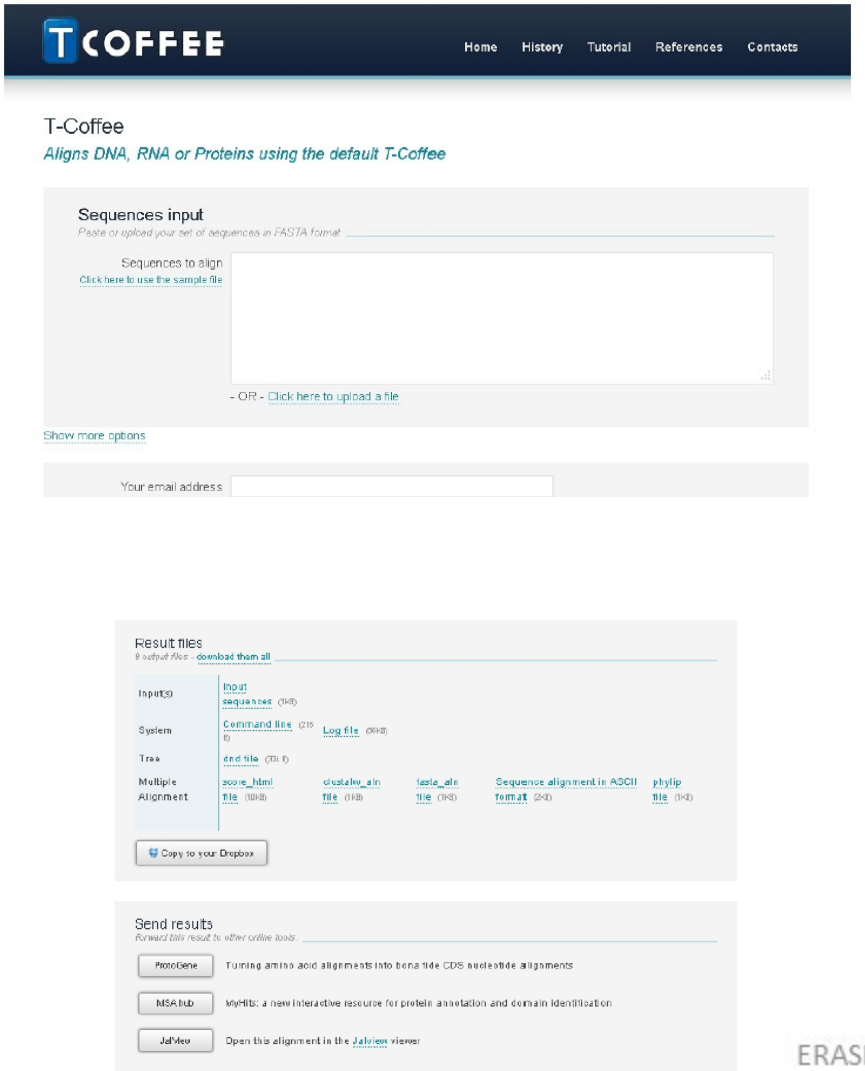
Πολλαπλές ευθυγραμμίσεις με Tcoffee

- Για να κάνετε πολλές ευθυγραμμίσεις στον συνηθισμένο (κανονικό) διακομιστή Tcoffee , θα χρειαστεί μόνο να τοποθετήσετε τις επιθυμητές ακολουθίες στο κατάλληλο παράθυρο.
  1. Μεταβείτε στην κύρια σελίδα του διακομιστή Tcoffee : www.tcoffee.org .
  2. Πατήστε το κουμπί T-Coffee.
  3. Τοποθετήστε τις ακολουθίες στο παράθυρο *Ακολουθίες για ευθυγράμμιση* .
  4. Κάντε κλικ στο *Υποβολή* .
  5. Ρίξτε μια ματιά στα αποτελέσματά σας.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πολλαπλές ευθυγραμμίσεις με Tcoffee



Συνδυασμός ακολουθιών και δομών με το EXPRESSO

- Το EXPRESSO είναι η πιο πρόσφατη εφεύρεση του Tcoffee , αντικαθιστώντας το 3D-Coffee (παλιά έκδοση).
  - Όταν εκτελείτε το Expresso, το πρόγραμμα χρησιμοποιεί το BLAST για να πραγματοποιήσει αναζήτηση στη βάση δεδομένων PDB (βάση δεδομένων δομής πρωτεΐνης) για δομές των οποίων οι ακολουθίες είναι παρόμοιες με αυτές που έχουν χρησιμοποιηθεί.
  - Στη συνέχεια χρησιμοποιεί αυτές τις δομές για να υποστηρίξει την ευθυγράμμιση.
  - Το EXPRESSO είναι πιο αργό από το Tcoffee , αλλά αν βρει αρκετές δομές, παράγει τις πιο ακριβείς ευθυγραμμίσεις των ακολουθιών που είναι γνωστές επί του παρόντος.
  - Η EXPRESSO συγκρίνει δομές με το SAP (πρόγραμμα Taylor & Orengo) και ακολουθίες καθώς και δομές χρησιμοποιώντας FUGUE (λογισμικό που αναπτύχθηκε στο Πανεπιστήμιο του Κέιμπριτζ).
  - Για να εκκινήσετε το EXPRESSO, κάντε κλικ στο κουμπί EXPRESSO

  τη σελίδα www.tcoffee.org .

Συνδυασμός ακολουθιών και δομών με το EXPRESSO

- Στην ενότητα αποτελεσμάτων, αναζητήστε ένα αρχείο που ονομάζεται template_list .
  - Περιέχει μια λίστα με όλες τις δομές που το πρόγραμμα μπόρεσε να συσχετίσει με τις ακολουθίες που έχετε υποβάλει.
  - Εάν αυτό το αρχείο είναι κενό, τότε καμία δομή δεν ταιριάζει με τις ακολουθίες που χρησιμοποιήθηκαν. Τότε η στοίχιση EXPRESSO μόλις εκτέλεσε μια τυπική ευθυγράμμιση με το Tcoffee .
- Το Tcoffee μπορεί να αξιολογήσει την ποιότητα της ευθυγράμμισής σας για να σας βοηθήσει να κατανοήσετε ποια μέρη του μπορείτε να εμπιστευτείτε.
- Αυτές οι αξιολογήσεις είναι μόνο εμπειρικές - δεν υποκαθιστούν την τιμή Ε. Ωστόσο, είναι χρήσιμο να γνωρίζετε ότι όλες οι βάσεις ή τα ΑΑ που εμφανίζονται σε κίτρινο/πορτοκαλί/κόκκινο φόντο έχουν δείκτη πάνω από 5 (έως 10) - δηλαδή πάνω από 80% πιθανότητα να έχουν συγκριθεί σωστά.

Co-funded by the Erasmus+ Programme of the European Union

Επεξεργασία μεγάλων συνόλων δεδομένων με MUSCLE

- Το MUSCLE είναι ένα νέο μέλος της οικογένειας των προγραμμάτων πολλαπλής ευθυγράμμισης, αλλά είναι πολύ αποτελεσματικό λογισμικό, τόσο γρήγορο όσο και υψηλής ποιότητας.
- Το MUSCLE είναι ιδανικό αν θέλετε να κάνετε μια ευθυγράμμιση πολλών εκατοντάδων ακολουθιών ταυτόχρονα.
- Μπορείτε να τα εκτελέσετε από πολλούς διακομιστές, συμπεριλαμβανομένης μιας σελίδας στη διεύθυνση www.drive5.commuscle .
- Η εργασία με το MUSCLE είναι πολύ απλή - το μόνο που έχετε να κάνετε είναι να επικολλήσετε τις ακολουθίες σας στο κατάλληλο παράθυρο.

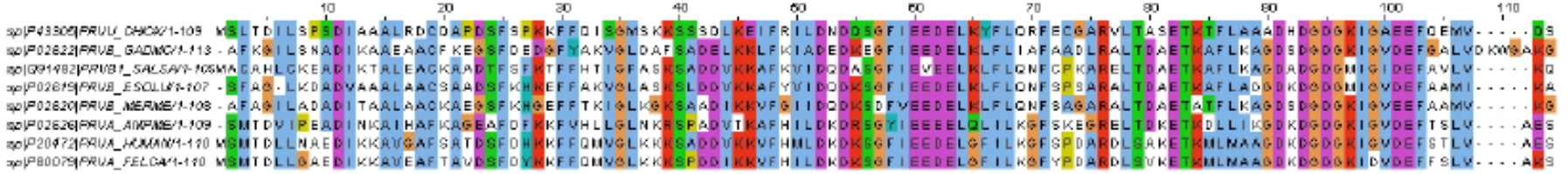
Ερμηνεύοντας την πολλαπλή ευθυγράμμισή σας

- Οι ηλεκτρονικές τιμές, οι οποίες υποδεικνύουν πόσο αξιόπιστη είναι μια αναζήτηση ομοιότητας, δεν υπάρχουν ακόμη για πολλαπλές ευθυγραμμίσεις, επομένως η αξιολόγηση της ορθότητάς τους είναι συχνά θέμα εμπειρίας και εικασιών.

- Οι ευθυγραμμίσεις των αλληλουχιών DNA είναι οι πιο δύσκολες στην ερμηνεία.
  - Ένα μπλοκ DNA είναι ενημερωτικό μόνο εάν περιέχει μια ομάδα πολλών πανομοιότυπων στηλών.

Αναγνώριση καλών τμημάτων σε ευθυγράμμιση πρωτεΐνης

- Στις πρωτεϊνικές δομές, υπάρχουν πιο «μαλακά» και ταχέως μεταβαλλόμενα τμήματα - «βρόχοι» - που βρίσκονται στην επιφάνεια και συνδέουν «σκληρότερες» περιοχές (με διαφορετικές λειτουργίες) που αλλάζουν πιο αργά από τους βρόχους.
  - Έτσι, σε μια πολλαπλή ευθυγράμμιση, θα αναμένονταν ωραία μπλοκ χωρίς κενά που αντιστοιχούν σε σημαντικές δομές, καθώς και περισσότερες μεταβλητές περιοχές που περιέχουν κενά, αντικαταστάσεις κ.λπ. - που αντιστοιχούν σε "βρόχους".



Αναγνώριση καλών τμημάτων σε ευθυγράμμιση πρωτεΐνης

- Πώς να μάθετε εάν ένα μπλοκ είναι καλό;
  - Όταν κοιτάζετε μια στοίχιση που έγινε με ClustalOmega , MUSCLE ή Tcoffee , θα παρατηρήσετε ότι η τελευταία γραμμή περιέχει περίεργους χαρακτήρες - (*), (:) ή (.):
    - (*) υποδεικνύει μια πλήρως διατηρημένη στήλη.
    - (:) τα δύο σημεία υποδηλώνουν μια στήλη στην οποία όλα τα υπολείμματα AA έχουν περίπου το ίδιο μέγεθος και ιδιότητες υδροφιλικότητας/υδροφοβίας.
    - (.) η τελεία υποδηλώνει μια στήλη στην οποία το μέγεθος και οι ιδιότητες υδροφιλικότητας/υδροφοβίας έχουν διατηρηθεί στην εξέλιξη
- Για παράδειγμα, ένα μέσο καλό μπλοκ είναι μια συμπαγής μονάδα τουλάχιστον 10-30 AA χωρίς ασυνέχειες, που έχει τουλάχιστον 1-3 (*), μερικές ακόμη (:) που βρίσκονται κοντά στους αστερίσκους και επίσης μερικές κουκκίδες διάσπαρτες παντού .

Co-funded by the Erasmus+ Programme of the European Union

Αναγνώριση καλών τμημάτων σε ευθυγράμμιση πρωτεΐνης

- Η μαγεία των πολλαπλών ευθυγραμμίσεων είναι ότι 4 ή 5 διατηρημένες θέσεις από 50+ συγκριτικά ΑΑ μπορεί να είναι αρκετές για να λάβετε ένα αληθινό σήμα.
  - Αυτό είναι κάτω από 10% ταυτότητα!
- Ένα άλλο κριτήριο για μια χρήσιμη πολλαπλή ευθυγράμμιση είναι να γνωρίζετε ποιοι τύποι αμινοξέων θα πρέπει να περιμένετε να διατηρηθούν.
- Οι διατηρημένες στήλες σε μια πολλαπλή στοίχιση έχουν σημασία μόνο όταν οι γύρω στήλες δεν διατηρούνται.
  - Ο παρακάτω πίνακας παραθέτει τις πιο κοινές ιδιότητες που σχετίζονται με ορισμένες διατηρημένες στήλες που θα συναντήσετε στη στοίχισή σας.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

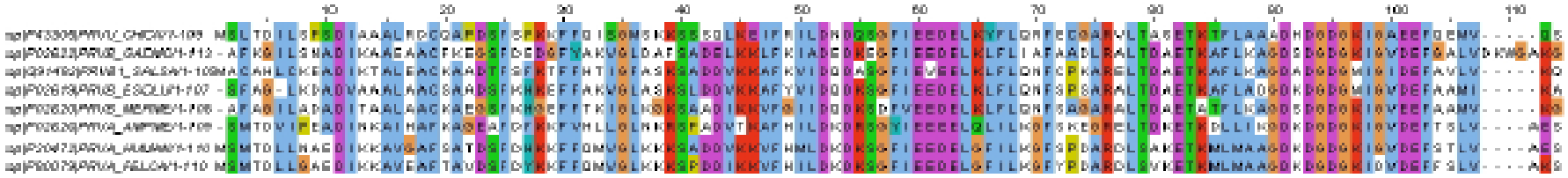
Co-funded by the Erasmus+ Programme of the European Union

Αναγνώριση καλών τμημάτων σε ευθυγράμμιση πρωτεΐνης

Amino-acid	Characteristic
W,Y, F	Conservative columns with tryptophan meet very often. It has a large hydrophobic residue that is ofetdeep in the core of proteins and is very important for thier stability. When tryptophan mutates (which is rare), it is usually replaced by another hydrophobic AA, such as phenylalanine or tyrosine. Conservative columns with aromatic AAs are the most common signs to recognise protein domain.
G, P	Consrevative columns of glycine or proline are common in multiple alignment. These two AAs often coincide with the ends of well-structured beta chains or alpha coils.
C	Cysteine is known for its ability to make C-C (disulfide) bridges. Conservative columns of cysteines are commonly found and usually indicate such bridges, and when at a certain distance give a good cue for the recognition of protein domains and folds.
H, S	Histidine and serine are often participants in catalytic sites, especially of proteases. These AAs are good candidates for active site participants.
K, R, D, E	Charged AA lysine, arginine, aspartic acid and glutamic acid are often involved in ligand binding. Highly conserved columns involving them may also signify a salt bridge in the core of the protein.
L	Leucines are rarely very conservative unless involved in protein-protein interactions, e. g. "leucine zipper"

Αναγνώριση καλών περιοχών σε ευθυγράμμιση πρωτεΐνης

- Η πολλαπλή ευθυγράμμιση υποδεικνύει σημαντικές θέσεις στις πρωτεΐνες, τα διατηρημένα αμινοξέα - αμινοξέα που δεν μεταλλάσσονται ακόμη και σε απομακρυσμένα συγγενείς πρωτεΐνες.



- Μια πιθανή στρατηγική για να δείξουμε τους εξελικτικούς περιορισμούς στην πρωτεΐνη μας είναι να συμπεριλάβουμε στην ευθυγράμμιση ακριβώς εκείνες τις αλληλουχίες που η έξοδος αναζήτησης BLAST ως οριακά χτυπά στην αναζήτηση ομολογίας.

Co-funded by the Erasmus+ Programme of the European Union

Αναγνώριση καλών περιοχών σε ευθυγράμμιση πρωτεΐνης

1. Μεταβείτε στη διεύθυνση http://www.uniprot.org/batch/
2. Στο Παρέχετε τα αναγνωριστικά σας παράθυρο, εισαγάγετε τους ακόλουθους αριθμούς πρόσβασης ακολουθίας:
   - ***P20472, P80079, P02626, P02619, P43305, Q91482, P02620, P02622, P02586*** .
3. Κάντε κλικ στο κουμπί Υποβολή .
4. Κατεβάστε τις ακολουθίες σε μορφή FASTA (αρχείο fasta)
5. Ανοίξτε το αρχείο FASTA και αντίγραφο .

Co-funded by the Erasmus+ Programme of the European Union

Αναγνώριση καλών περιοχών σε ευθυγράμμιση πρωτεΐνης

- Όταν συγκεντρώσετε όλες τις ακολουθίες, είναι ώρα για ευθυγράμμιση ξανά.

1. διακομιστή ClustalW της EBI :

http://www.ebi.ac.uk/Tools/msa/clustal o /

2. Τοποθετήστε τις συλλεγμένες ακολουθίες στην Ακολουθία παράθυρο.

3. Επιλέξτε τη μορφή εξόδου που θέλετε.

4. Επιλέξτε Input από το μενού Output Order.

5. Πατήστε το κουμπί Υποβολή.

6. Ασφαλίστε τα αποτελέσματά σας .

Αναγνώριση καλών περιοχών σε ευθυγράμμιση πρωτεΐνης

- Πολλαπλή ευθυγράμμιση πρωτεϊνών μεγάλων αποστάσεων.

```
CLUSTAL O(1.2.4) multiple sequence alignment

sp|P20472|PRVA_HUMAN      ------------------------------------------MSMTDLLNAE-DIKKAV 16
sp|P80079|PRVA_FELCA      ------------------------------------------MSMTDLLGAE-DIKKAV 16
sp|P02626|PRVA_AMPME      ------------------------------------------SMTDVIPEA-DINKAI 15
sp|P02619|PRVB_ESOLU      ------------------------------------------SFAG-LKDA-DVAAAL 14
sp|P43305|PRVU_CHICK      ------------------------------------------MSLTDILSPS-DIAAAL 16
sp|Q91482|PRVB1_SALSA     ------------------------------------------MACAHLCKEA-DIKTAL 16
sp|P02620|PRVB_MERME      ------------------------------------------AFAGILADA-DITAAL 15
sp|P02622|PRVB_GADMC      ------------------------------------------AFKGILSNA-DIKAAE 15
sp|P02586|TNNC2_RABIT     MTDQQAEARSYLSEEMIAEFKAAFDMFDADGGGDISVKELGTVMRMLGQTPTKEELDAII 60
                                                                     ::

sp|P20472|PRVA_HUMAN      G--AFSATDSFDHKKFFQMV------GLKKKSADDVKKVFHMLDKDKSGFIEEDELGFIL 68
sp|P80079|PRVA_FELCA      E--AFTAVDSFDYKKFFQMV------GLKKKSPDDIKKVFHILDKDKSGFIEEDELGFIL 68
sp|P02626|PRVA_AMPME      H--AFKAGEAFDFKKFVHLL------GLNKRSPADVTKAFHILDKDRSGYIEEEELQLIL 67
sp|P02619|PRVB_ESOLU      A--ACSAADSFKHKEFFAKV------GLASKSLDDVKKAFYVIDQDKSGFIEEDELKLFL 66
sp|P43305|PRVU_CHICK      R--DCQAPDSFSPKKFFQIS------GMSKKSSSQLKEIFRILDNDQSGFIEEDELKYFL 68
sp|Q91482|PRVB1_SALSA     E--ACKAADTFSFKTFFHTI------GFASKSADDVKKAFKVIDQDASGFIEVEELKLFL 68
sp|P02620|PRVB_MERME      A--ACKAEGSFKHGEFFTKI------GLKGKSAADIKKVFGIIDQDKSDFVEEDELKLFL 67
sp|P02622|PRVB_GADMC      A--ACFKEGSFDEDGFYAKV------GLDAFSADELKKLFKIADEDKEGFIEEDELKLFL 67
sp|P02586|TNNC2_RABIT     EEVDEDGSGTIDFEEFLVMMVRQMKEDAKGKSEEELAECFRIFDRNADGYIDAEELAEI- 119
                                ::.   *          .    *  :: : * : *.: ..::: :**  :

sp|P20472|PRVA_HUMAN      KGFSPDARDLSAKETKMLMAAGDKDGDGKIGVDEFSTLVAES----       110
sp|P80079|PRVA_FELCA      KGFYPDARDLSVKETKMLMAAGDKDGDGKIDVDEFFSLVAKS----       110
sp|P02626|PRVA_AMPME      KGFSKEGRELTDKETKDLLIKGDKDGDGKIGVDEFTSLVAES----       109
sp|P02619|PRVB_ESOLU      QNFSPSARALTDAETKAFLADGDKDGDGMIGVDEFAAMIKA-----       107
sp|P43305|PRVU_CHICK      QRFECGARVLTASETKTFLAAADHDGDGKIGAEEFQEMVQS-----       109
sp|Q91482|PRVB1_SALSA     QNFCPKARELTDAETKAFLKAGDADGDGMIGIDEFAVLVKQ-----       109
sp|P02620|PRVB_MERME      QNFSAGARALTDAETATFLKAGDSDGDGKIGVEEFAAMVKG-----       108
sp|P02622|PRVB_GADMC      IAFAADLRALTDAETKAFLKAGDSDGDGKIGVDEFGALVDKWGAKG       113
sp|P02586|TNNC2_RABIT     --FRASGEHVTDEEIESLMKDGDKNNDGRIDFDEFLKMMEGVQ---       160
                            *    . ::  *   ::  .* :.** *. :**  ::
```

Αναγνώριση καλών περιοχών σε ευθυγράμμιση πρωτεΐνης

- Σε αυτό το σημείο, θα ήταν καλή ιδέα να προσθέσετε μια άλλη ακολουθία μεγάλης εμβέλειας στην ευθυγράμμιση που προκύπτει.
  - Ο στόχος είναι να επαληθευτεί ότι αυτές οι μικρές εξαιρετικά διατηρημένες περιοχές είναι πράγματι διατηρημένες σε όλη την οικογένεια πρωτεϊνών.
- Για να το κάνετε αυτό, πρέπει να περάσετε ξανά από τους δύο προηγούμενους αλγόριθμους, αυτή τη φορά χρησιμοποιώντας μια ακόμη ακολουθία:
  - P20472, P80079, P02626, P02619, P43305, Q91482, P02620, P02622, P02586, P19123

Αναγνώριση καλών περιοχών σε ευθυγράμμιση πρωτεΐνης

- Συμπερίληψη ακόμη πιο μακρινών ακολουθιών στην πολλαπλή στοίχιση

```
CLUSTAL O(1.2.4) multiple sequence alignment

sp|P02586|TNNC2_RABIT      MTDQQAEARSYLSEEMIAEFKAAFDMFDAD-GGGDISVKELGTVMRMLGQTPTKEELDAI 59
sp|P19123|TNNC1_MOUSE      MDDIYKAAVEQLTEEQKNEFKAAFDIFVLGAEDGCISTKELGKVMRMLGQNPTPEELQEM 60
sp|P02626|PRVA_AMPME       ----------------------------------------------SMTDVIPEA-DINKA 14
sp|P20472|PRVA_HUMAN       ----------------------------------------------MSMTDLLNAE-DIKKA 15
sp|P80079|PRVA_FELCA       ----------------------------------------------MSMTDLLGAE-DIKKA 15
sp|P43305|PRVU_CHICK       ----------------------------------------------MSLTDILSPS-DIAAA 15
sp|P02622|PRVB_GADMC       ----------------------------------------------AFKGILSNA-DIKAA 14
sp|Q91482|PRVB1_SALSA      ----------------------------------------------MACAHLCKEA-DIKTA 15
sp|P02619|PRVB_ESOLU       ----------------------------------------------SFAG-LKDA-DVAAA 13
sp|P02620|PRVB_MERME       ----------------------------------------------AFAGILADA-DITAA 14
                                                                            ::

sp|P02586|TNNC2_RABIT      IEEVDEDGSGTIDFEEFLVMMVRQMKEDAKGKSEEELAECFRIFDRNADGYIDAEELAEI 119
sp|P19123|TNNC1_MOUSE      IDEVDEDGSGTVDFDEFLVMMVRCMKDDSKGKSEEELSDLFRMFDKNADGYIDLDELKMM 120
sp|P02626|PRVA_AMPME       IH--AFKAGEAFDFKKFVHLL------GLNKRSPADVTKAFHILDKDRSGYIEEEELQLI 66
sp|P20472|PRVA_HUMAN       VG--AFSATDSFDHKKFFQMV------GLKKKSADDVKKVFHMLDKDKSGFIEEDELGFI 67
sp|P80079|PRVA_FELCA       VE--AFTAVDSFDYKKFFQMV------GLKKKSPDDIKKVFHILDKDKSGFIEEDELGFI 67
sp|P43305|PRVU_CHICK       LR--DCQAPDSFSPKKFFQIS------GMSKKSSSQLKEIFRILDNDQSGFIEEDELKYF 67
sp|P02622|PRVB_GADMC       EA--ACFKEGSFDEDGFYAKV------GLDAFSADELKKLFKIADEDKEGFIEEDELKLF 66
sp|Q91482|PRVB1_SALSA      LE--ACKAADTFSFKTFFHTI------GFASKSADDVKKAFKVIDQDASGFIEVEELKLF 67
sp|P02619|PRVB_ESOLU       LA--ACSAADSFKHKEFFAKV------GLASKSLDDVKKAFYVIDQDKSGFIEEDELKLF 65
sp|P02620|PRVB_MERME       LA--ACKAEGSFKHGEFFTKI------GLKGKSAADIKKVFGIIDQDKSDFVEEDELKLF 66
                                  :..   *          .    *  :: . * : *.: ..::: :**  :

sp|P02586|TNNC2_RABIT      FRA---SGEHVTDEEIESLMKDGDKNNDGRIDFDEFLKMMEGVQ---      160
sp|P19123|TNNC1_MOUSE      LQA---TGETITEDDIEELMKDGDKNNDGRIDYDEFLEFMKGVE---      161
sp|P02626|PRVA_AMPME       LKGFSKEGRELTDKETKDLLIKGDKDGDGKIGVDEFTSLVAES----      109
sp|P20472|PRVA_HUMAN       LKGFSPDARDLSAKETKMLMAAGDKDGDGKIGVDEFSTLVAES----      110
sp|P80079|PRVA_FELCA       LKGFYPDARDLSVKETKMLMAAGDKDGDGKIDVDEFFSLVAKS----      110
sp|P43305|PRVU_CHICK       LQRFECGARVLTASETKTFLAAADHDGDGKIGAEEFQEMVQS-----      109
sp|P02622|PRVB_GADMC       LIAFAADLRALTDAETKAFLKAGDSDGDGKIGVDEFGALVDKWGAKG      113
sp|Q91482|PRVB1_SALSA      LQNFCPKARELTDAETKAFLKAGDADGDGMIGIDEFAVLVKQ-----      109
sp|P02619|PRVB_ESOLU       LQNFSPSARALTDAETKAFLADGDKDGDGMIGVDEFAAMIKA-----      107
sp|P02620|PRVB_MERME       LQNFSAGARALTDAETATFLKAGDSDGDGKIGVEEFAAMVKG-----      108
                           :       . ::  :   ::  .* :.** *. :**  ::
```

Co-funded by the Erasmus+ Programme of the European Union

Πόροι Διαδικτύου για πολλαπλή ευθυγράμμιση

- Μεγάλος αριθμός διαδικτυακών πόρων για πολλαπλή ευθυγράμμιση.
- Όταν επιλέγετε με ποιο να συνεργαστείτε, λάβετε υπόψη τις ακόλουθες συμβουλές:
  - Χρησιμοποιήστε αξιόπιστους πόρους και κάνετε πάντα κάποιες απλές δοκιμές για να δείτε αν κάνουν αυτό που περιμένετε να κάνουν.
  - Ποτέ μην εμπιστεύεστε τυφλά πόρους στους οποίους δεν έχετε κανέναν έλεγχο.
  - Εάν θέλετε να κάνετε πολλές ευθυγραμμίσεις ή για κάποιο λόγο δεν θέλετε να "αναρτήσετε" τις ακολουθίες σας στο Διαδίκτυο, θα χρειαστεί να εγκαταστήσετε τέτοια προγράμματα τοπικά στον υπολογιστή σας.

Co-funded by the Erasmus+ Programme of the European Union

Ανακαλύψτε τη μέθοδο που προτιμάτε για πολλαπλή ευθυγράμμιση

- Ο παρακάτω πίνακας παραθέτει τις διευθύνσεις πολλών διακομιστών ευθυγράμμισης.

Method	Description/ Advantage	Web page
Tcoffee	Exact combination of sequences and structures	www.tcoffee.org http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi http://www.ebi.ac.uk/Tools/msa/tcoffee/
MUSCLE	Fast and accurate method	www.drive5.com/muscle/ http://www.ebi.ac.uk/Tools/msa/muscle/
Kalign	Fast method	http://msa.sbc.su.se/cgi-bin/msa.cgi
MAFFT	Fast and accurate method using - Fast Fourier Transforms	http://align.bmr.kyushu-u.ac.jp/mafft/software/
Dialign	Ideal for sequences with local homology	http://bibiserv.techfak.uni-bielefeld.de/dialign/

Σύγκριση αλληλουχιών στις οποίες δεν μπορεί να γίνει ευθυγράμμιση

- Μερικές φορές είναι απαραίτητο να συγκρίνουμε αλληλουχίες που δεν έχουν απαραιτήτως κοινή προέλευση ή σχετίζονται τόσο μακριά που είναι δύσκολο να αναγνωριστούν ως ομόλογα.
- Σε τέτοιες καταστάσεις, μην περιμένετε θαύματα από τη βιοπληροφορική. Στις περισσότερες περιπτώσεις, η πρώτη σας εντύπωση ότι δεν υπάρχει τίποτα να βγείτε από αυτές τις ακολουθίες αποδεικνύεται σωστή.
  - Δειγματολήπτης Gibbs, εξετάζει πολλαπλές ακολουθίες ταυτόχρονα για σύντομα, μερικώς διατηρημένα τμήματα χωρίς διακοπές.
  - Pratt, αναζητά ευέλικτα μοντέλα - μια ειδική κατηγορία τμημάτων που μπορεί να περιέχουν σπασίματα και διατηρούνται μόνο σε ορισμένες θέσεις.

Πώς να κάνετε πολλαπλές τοπικές ευθυγραμμίσεις με το δείγμα Gibbs

- Ο δειγματολήπτης Gibbs είναι μια στοχαστική μέθοδος:
  - πρώτα ανακατεύει τις ακολουθίες,
  - στη συνέχεια τα συγκρίνει τυχαία μέχρι να εμφανιστεί ένα καλό αποτέλεσμα και, στη συνέχεια, συνεχίζει να ανακατεύει για να βελτιώσει περαιτέρω αυτό το αποτέλεσμα.
- Οι στοχαστικές μέθοδοι είναι οι πιο ισχυρές στη βιοπληροφορική μέχρι στιγμής.
  - Για παράδειγμα, εάν αναλύσετε το ίδιο σύνολο ακολουθιών δύο φορές με το δειγματολήπτη Gibbs, ενδέχεται να μην λάβετε ακριβώς τα ίδια αποτελέσματα.
- Ωστόσο, αν υποθέσετε ελαφρώς πιο ασυνήθιστη λογική, μπορεί να διαπιστώσετε ότι ο δειγματολήπτης Gibbs παρέχει πολύ καλές λύσεις για εξαιρετικά πολύπλοκα προβλήματα.

http://ccmbweb.ccv.brown.edu/gibbs/gibbs.html

Αναζήτηση συντηρημένων κινήτρων

- Το δείγμα Gibbs είναι χρήσιμο μόνο εάν οι ακολουθίες που αναζητάτε έχουν το ίδιο μήκος, όπως συμβαίνει με τους τομείς Helix-Turn-Helix.
- Εάν οι αλληλουχίες σας περιέχουν ένα παρόμοιο μοτίβο αλλά δεν συνδέονται ή είναι πολύ ασθενώς συνδεδεμένες, ο μόνος τρόπος για να τις αναλύσετε είναι να χρησιμοποιήσετε ένα μοτίβο εύρεσης προτύπων.
- Υπάρχουν πολλά εργαλεία για την ανίχνευση τέτοιων κοινών κακοδιατηρημένων μοτίβων.
  - Ένα από τα πιο ισχυρά είναι το Pratt επειδή επιτρέπει κάποια ευελιξία του χώρου μεταξύ διατηρημένων θέσεων. Μπορείτε επίσης να χρησιμοποιήσετε TEIRESIAS, MEME ή SMILE.

Co-funded by the Erasmus+ Programme of the European Union

Πόροι για αναζήτηση μοτίβων και μοτίβων

- Ο παρακάτω πίνακας παραθέτει μερικούς από τους πιο συχνά χρησιμοποιούμενους διαδικτυακούς πόρους για την αναζήτηση κοινών μοτίβων σε ακολουθίες που σχετίζονται πολύ κακά για να έχουν πολλαπλές στοίχιση.

Method	Address
Gibbs Sampler	http://mobyle.pasteur.fr/cgi-bin/portal.py http://ccmbweb.ccv.brown.edu/gibbs/gibbs.html
Pratt	http://www.ebi.ac.uk/Tools/pfa/pratt/
eMotif	http://motif.stanford.edu/projects.html
MEME	http://meme.sdsc.edu/meme/
TEIRESIAS	http://cbcsrv.watson.ibm.com/Tspd.html
Improbizer	http://users.soe.ucsc.edu/~kent/improbizer/improbizer.html

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ευχαριστώ !

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Εργαλεία Βιοπληροφορικής

Θέμα 4. Τεχνητά νευρωνικά δίκτυα

Μάθημα 1. Γενικές έννοιες ML και ANN

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Τεχνητά Νευρωνικά Δίκτυα

Μάθημα 1

Γενική εισαγωγή

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή στο AI, ML και DL

Τεχνητή νοημοσύνη (ΤΝ) :

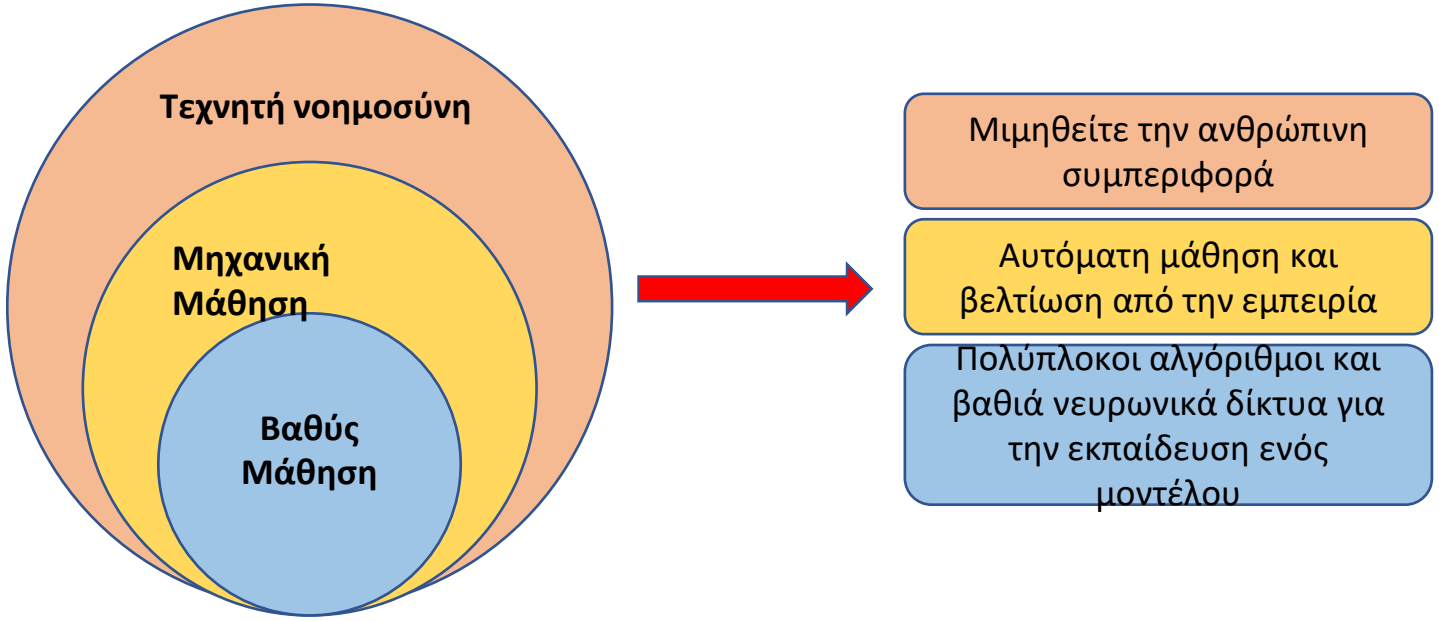
- Η ευρύτερη έννοια ότι οι μηχανές είναι σε θέση να εκτελούν εργασίες με τρόπο που θα θεωρούσαμε «έξυπνο» ή «έξυπνο».

Μηχανική Μάθηση (ΜΜ) :

- Ένα υποσύνολο τεχνητής νοημοσύνης που επιτρέπει στους υπολογιστές να μαθαίνουν από δεδομένα, αντί να προγραμματίζονται ρητά για να εκτελούν μια εργασία

Βαθιά Μάθηση (ΒΜ) :

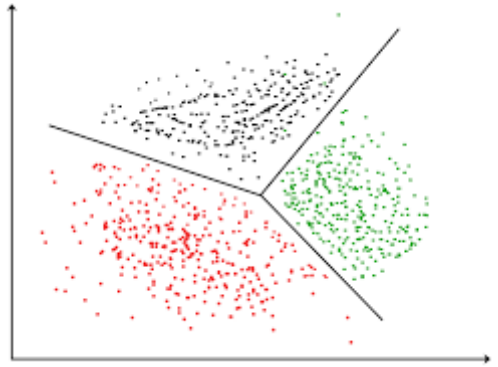
- Ένα υποσύνολο της ML που χρησιμοποιεί νευρωνικά δίκτυα με πολλά επίπεδα (άρα "βαθύ").

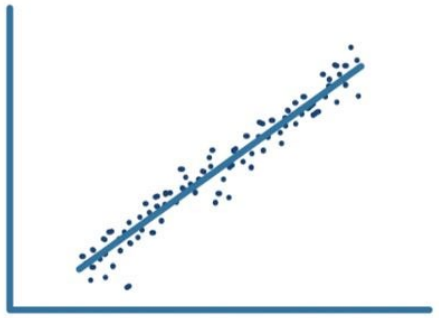


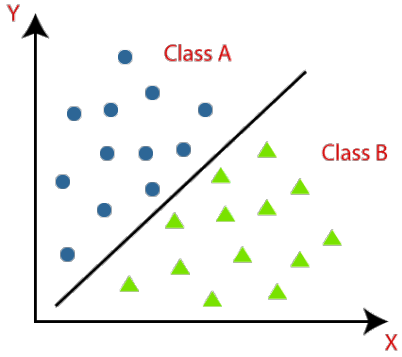
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

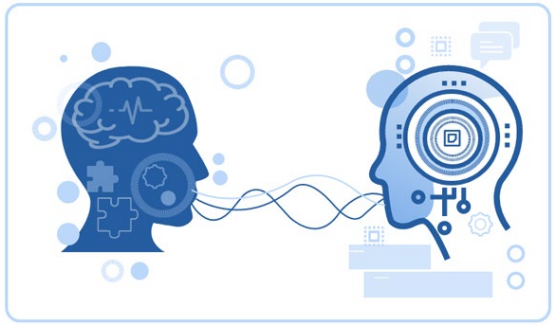
Προβλήματα που αντιμετωπίζονται από τη Μηχανική Μάθηση

- Οπισθοδρόμηση
- Ταξινόμηση
- Αναγνώριση εικόνας
- Φυσική γλώσσα
- Ομαδοποίηση
- Ενισχυτική Μάθηση
- Ανίχνευση ανωμαλιών
- Συστήματα Συστάσεων









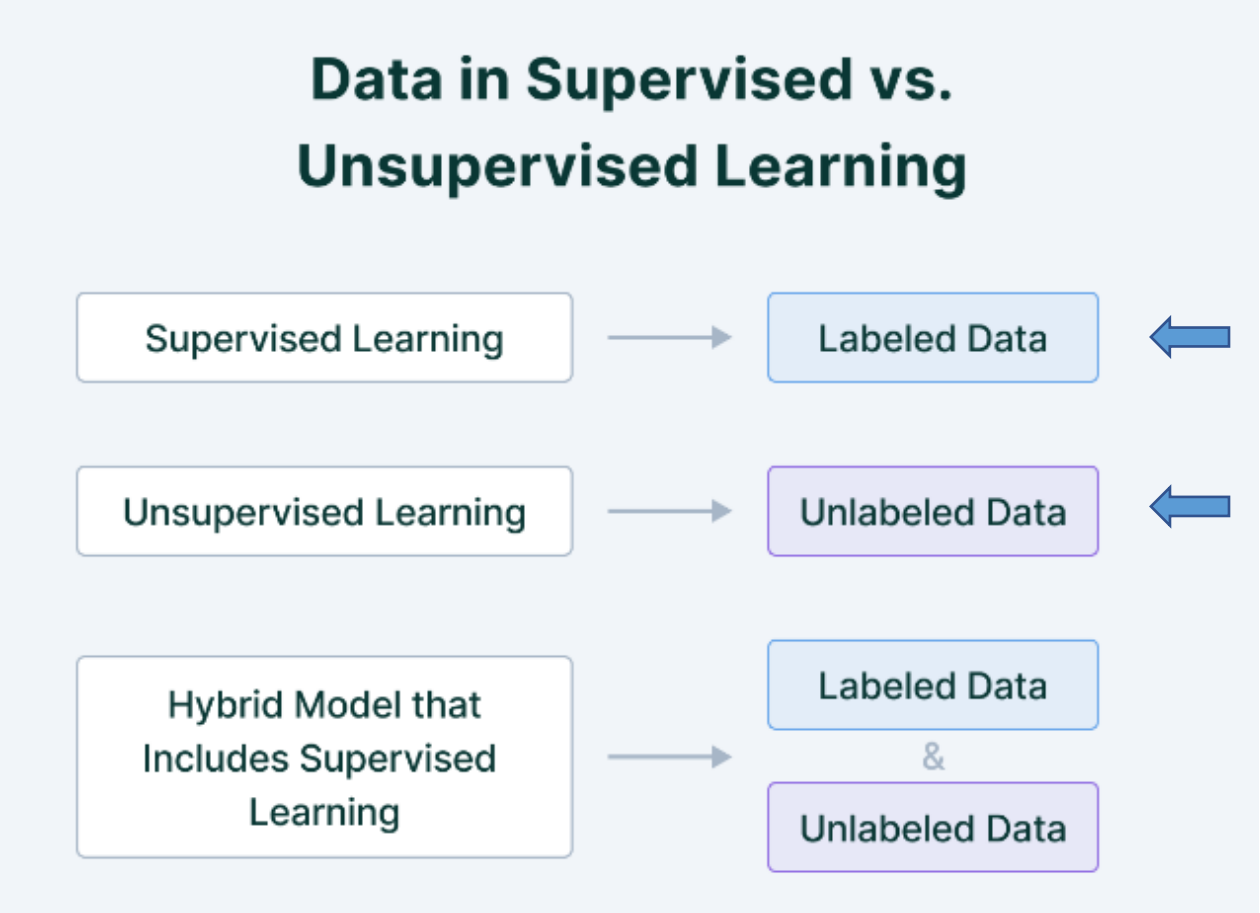
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Εποπτευόμενη / Μη εποπτευόμενη μάθηση

Βασικές διαφορές :

1. **Τύπος δεδομένων** : Το εποπτευόμενο χρησιμοποιεί δεδομένα με ετικέτα, ενώ το μη εποπτευόμενο χρησιμοποιεί δεδομένα χωρίς ετικέτα.
2. **Στόχος** : Η εποπτευόμενη στοχεύει στην πρόβλεψη των αποτελεσμάτων, ενώ η μη εποπτευόμενη στοχεύει στην εύρεση μοτίβων ή ομαδοποιήσεων.
3. **Βρόχος σχολίων** : Το εποπτευόμενο έχει βρόχο ανάδρασης (διόρθωση βάσει ετικέτας), ενώ το μη εποπτευόμενο όχι.



Παλινδρόμηση, Ταξινόμηση

Ομαδοποίηση, Μείωση Διαστάσεων

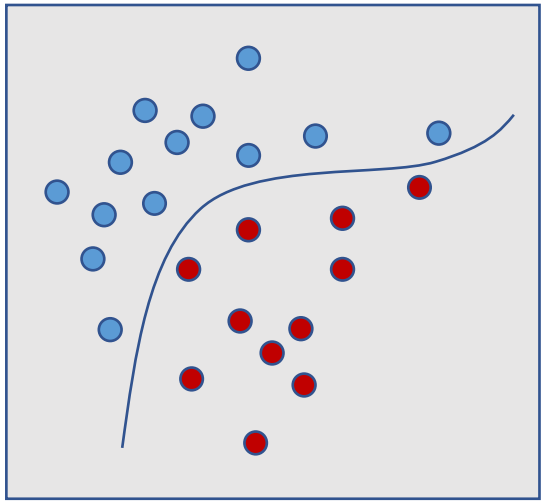
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

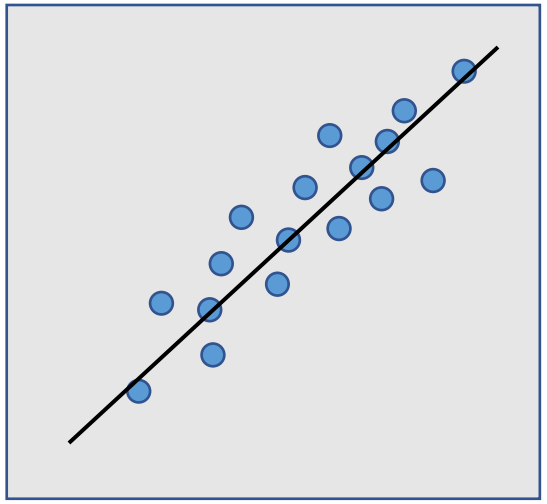
Παλινδρόμηση και Ταξινόμηση

Βασικές διαφορές :

1. **Τύπος εξόδου** : Η ταξινόμηση προβλέπει κατηγορίες, ενώ η παλινδρόμηση προβλέπει αριθμητικές τιμές
2. **Μετρικές αξιολόγησης** : Η ταξινόμηση χρησιμοποιεί ακρίβεια, ακρίβεια, ανάκληση κ.λπ., ενώ η παλινδρόμηση χρησιμοποιεί μέσο τετράγωνο σφάλμα, R-τετράγωνο κ.λπ.



Ταξινόμηση σημείων δεδομένων σε διαφορετικές κατηγορίες



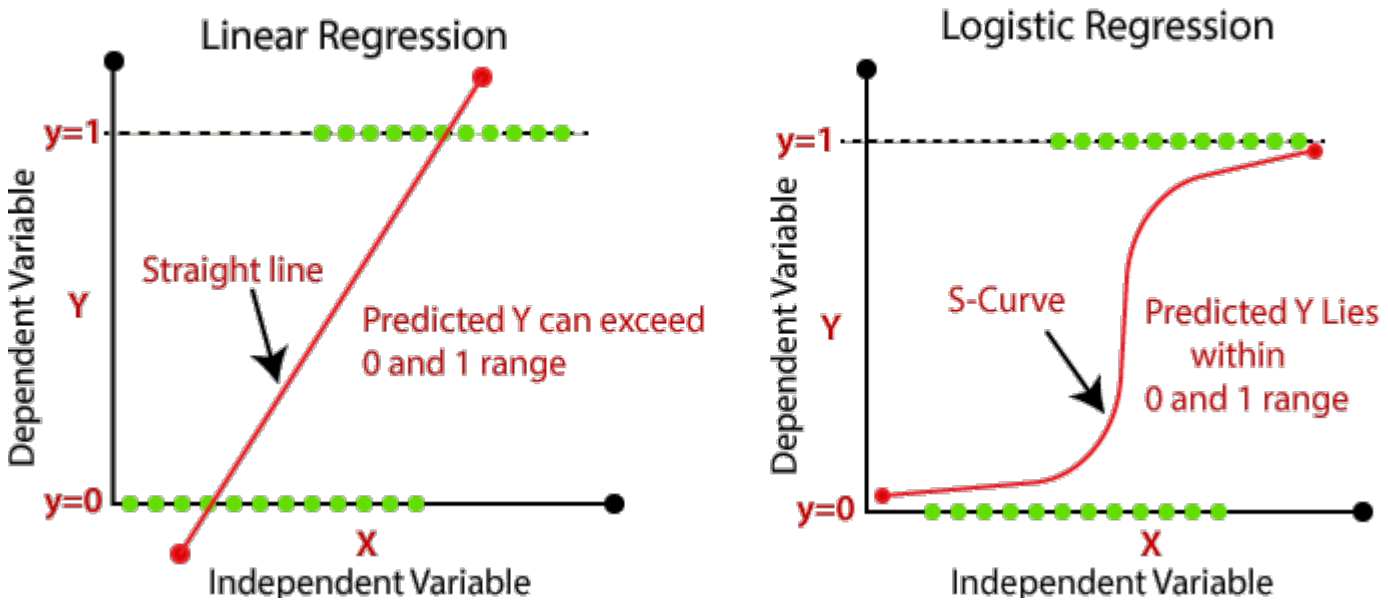
Προβλέψτε την τιμή της εξαρτημένης μεταβλητής δεδομένης μιας σειράς ανεξάρτητων μεταβλητών

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

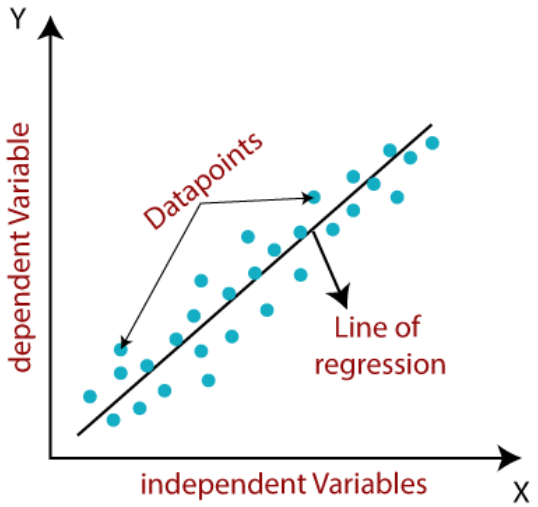
Γραμμική και λογιστική παλινδρόμηση (1)

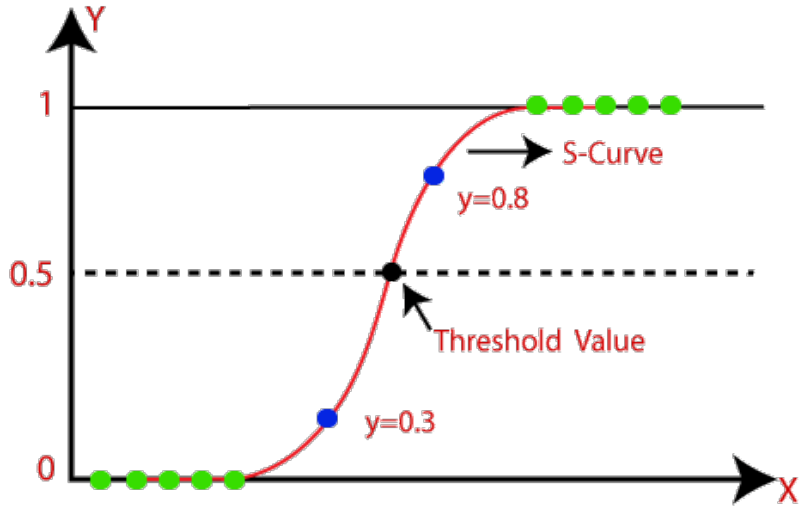
- *Η γραμμική παλινδρόμηση* είναι μια γραμμική προσέγγιση για τη μοντελοποίηση της σχέσης μεταξύ μιας απόκρισης και μιας ή περισσότερων επεξηγηματικών μεταβλητών
- *Η λογιστική παλινδρόμηση* εκτιμά την πιθανότητα να συμβεί ένα συμβάν, με βάση ένα δεδομένο σύνολο ανεξάρτητων μεταβλητών



Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Γραμμική και λογιστική παλινδρόμηση (2)





$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \cdots$$

$$\log\left(\frac{y}{1-y}\right) = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \cdots$$

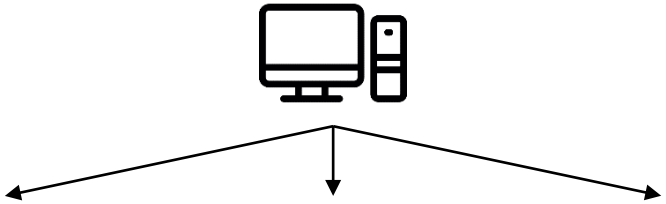
Find coefficients $a_i, i = 1, 2, \ldots$ that best describe the problem

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Αναγνώριση εικόνας

Ταξινόμηση



ποδήλατο Γάτα φράουλα

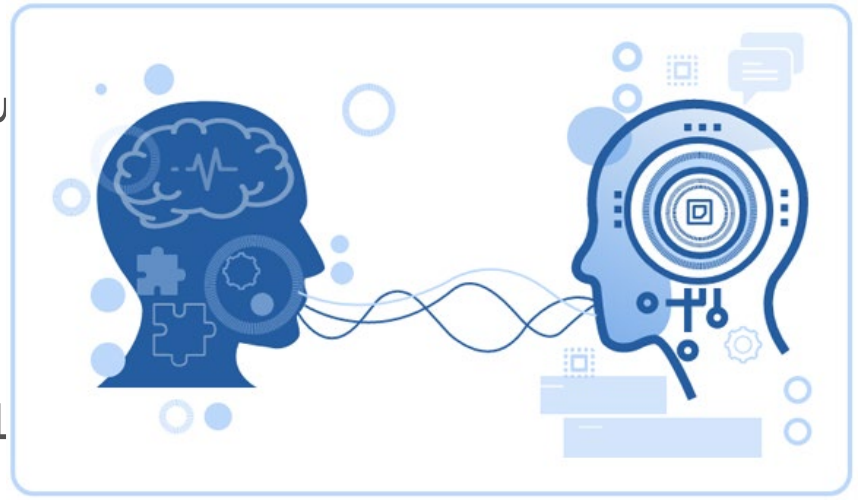
Ανίχνευση

Επεξεργασία Φυσικής Γλώσσας (NLP)

Αναγνώριση ομιλίας, Γραμματική προσθήκη ετικετών, Αποσαφήνιση της έννοιας λέξης, Ανάλυση συναισθημάτων, Δημιουργία φυσικής γλώσσας

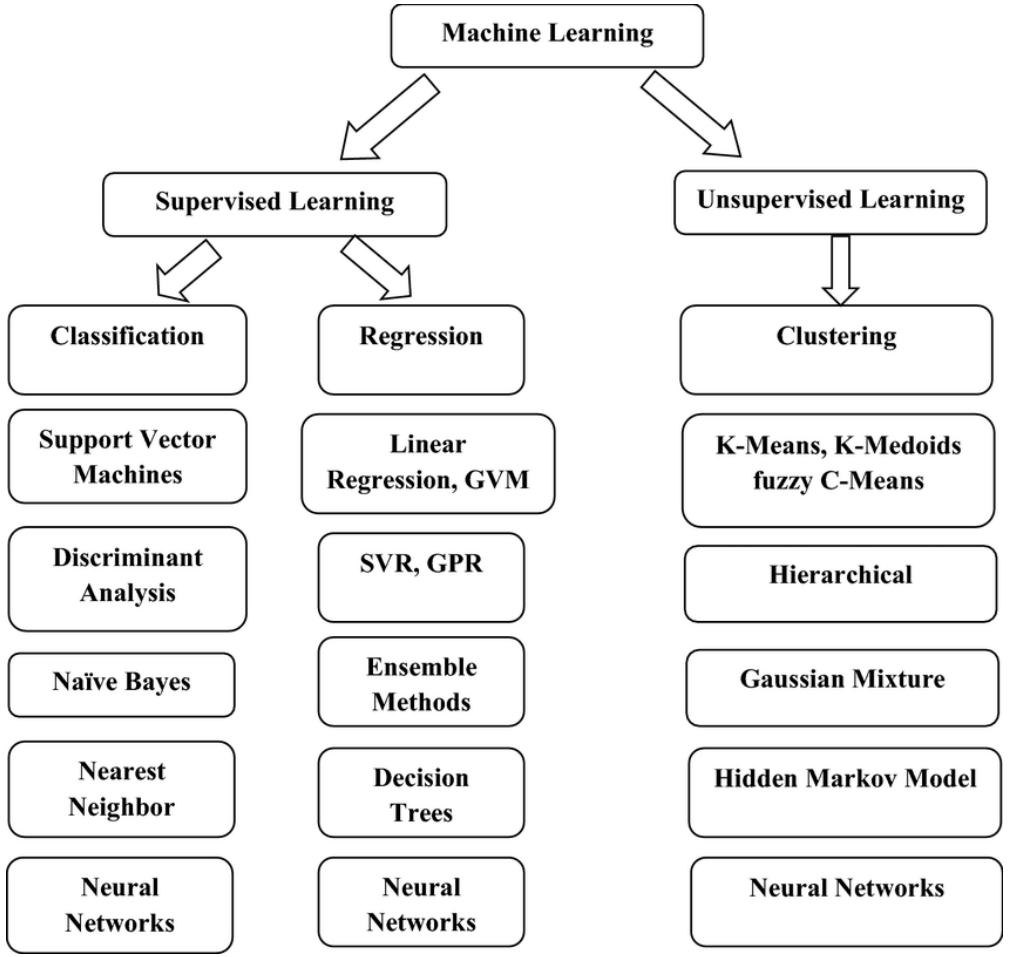
Κοινές εργασίες:

- Ανίχνευση ανεπιθύμητων μηνυμάτω
- Μηχανική μετάφραση
- Εικονικοί πράκτορες και chatbots
- Ανάλυση συναισθήματος μέσων κοι
- Σύνοψη κειμένου



Co-funded by the Erasmus+ Programme of the European Union

Μοντέλα Μηχανικής Μάθησης

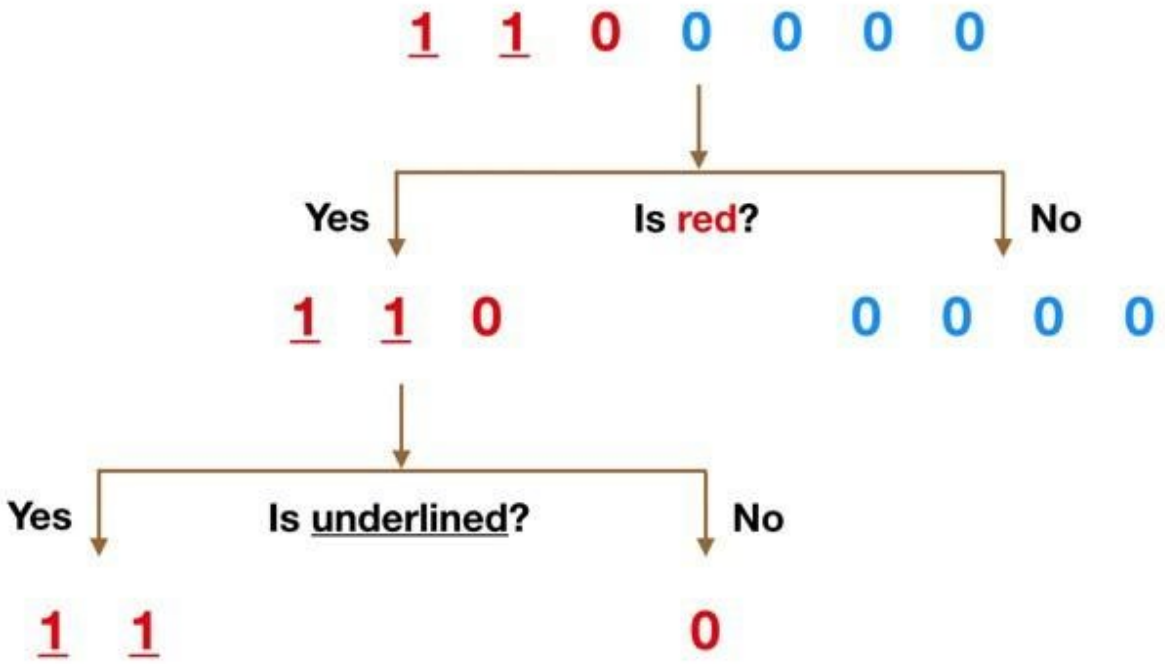


Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

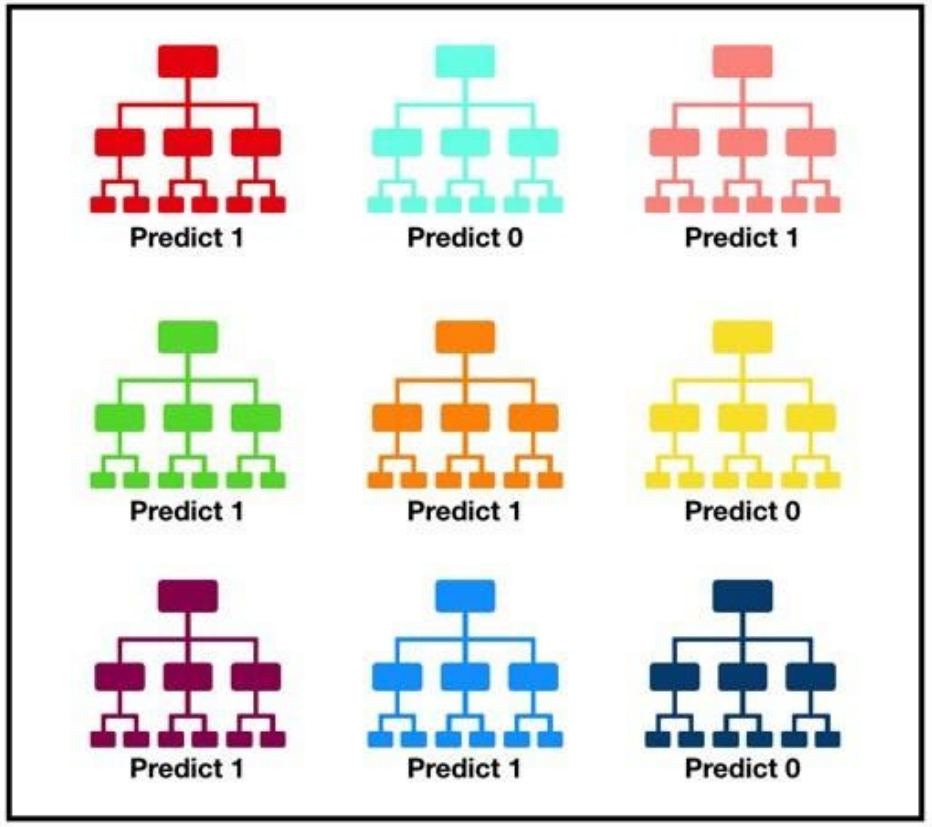
Co-funded by the Erasmus+ Programme of the European Union

Δέντρα απόφασης – Τυχαία δάση

Δέντρο απόφασης



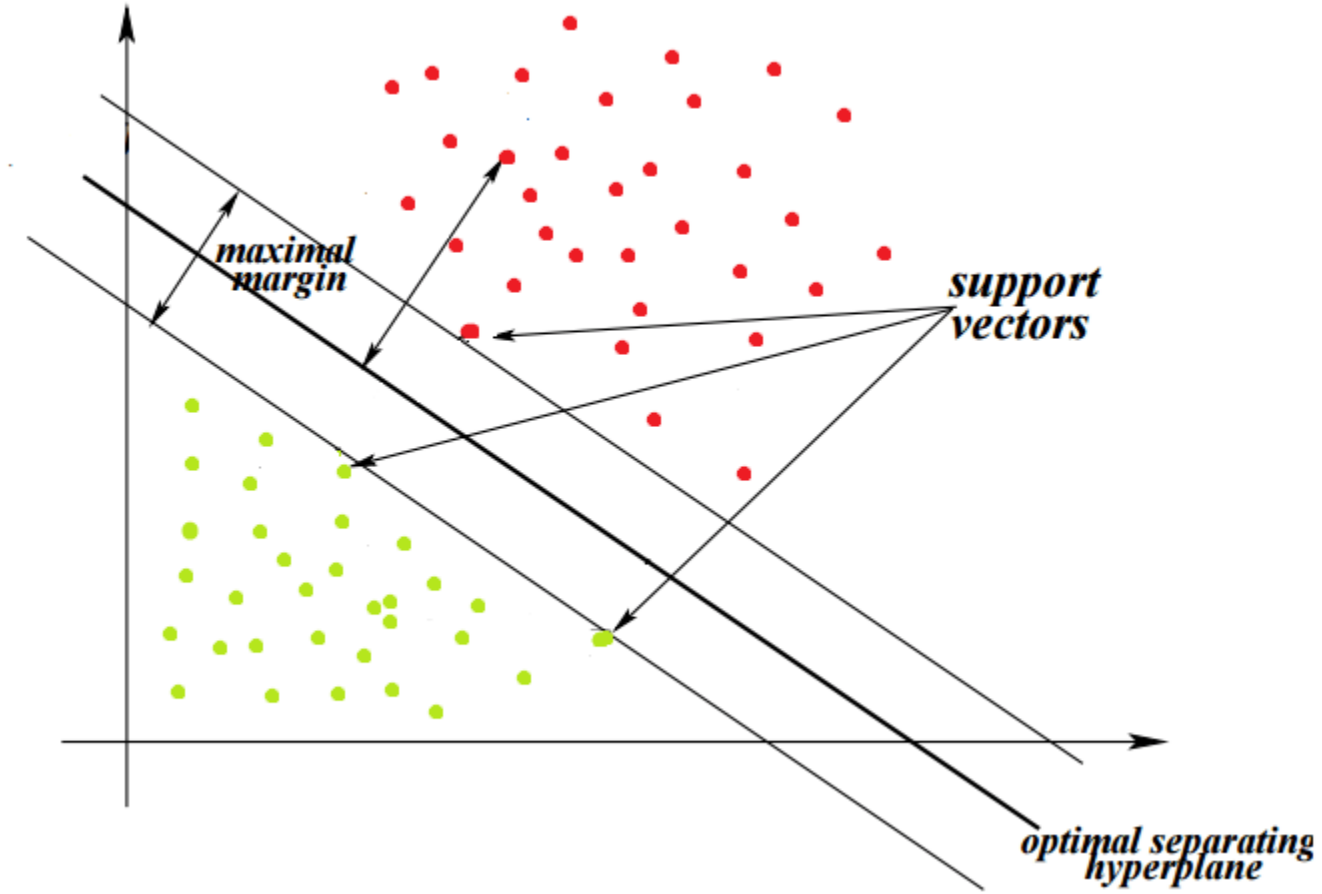
Random Forrest



Tally: Six 1s and Three 0s
Prediction: 1

Co-funded by the Erasmus+ Programme of the European Union

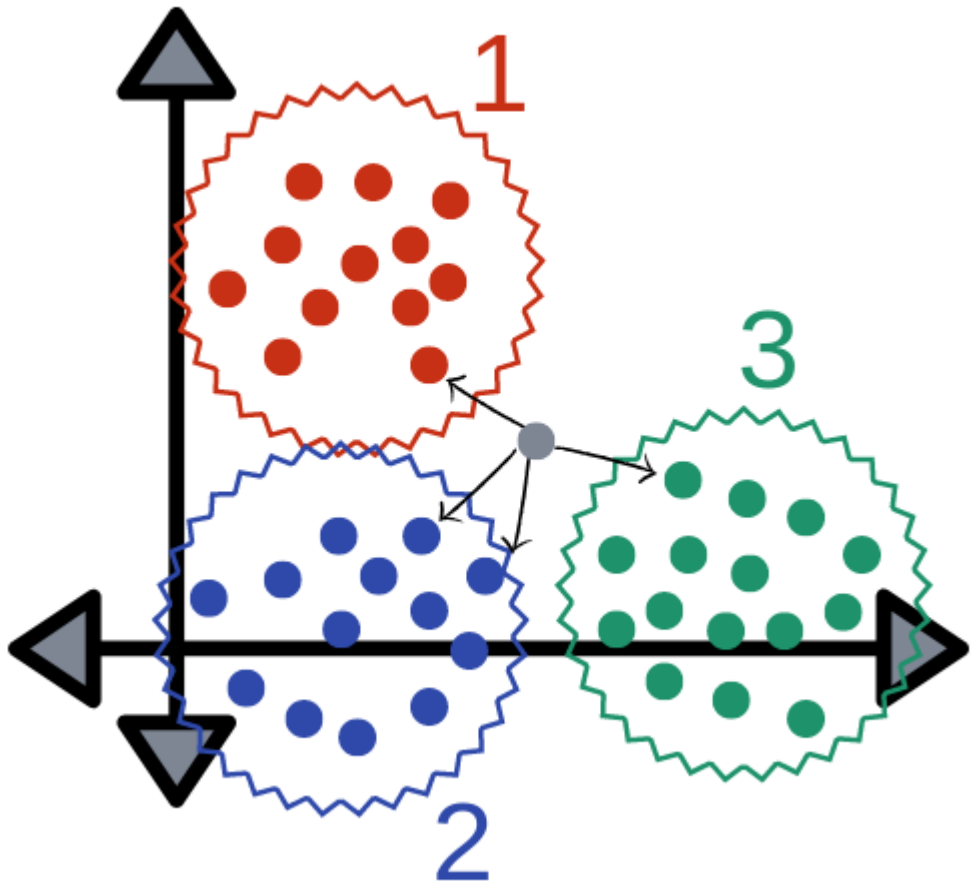
Support Vector Machines (SVM)



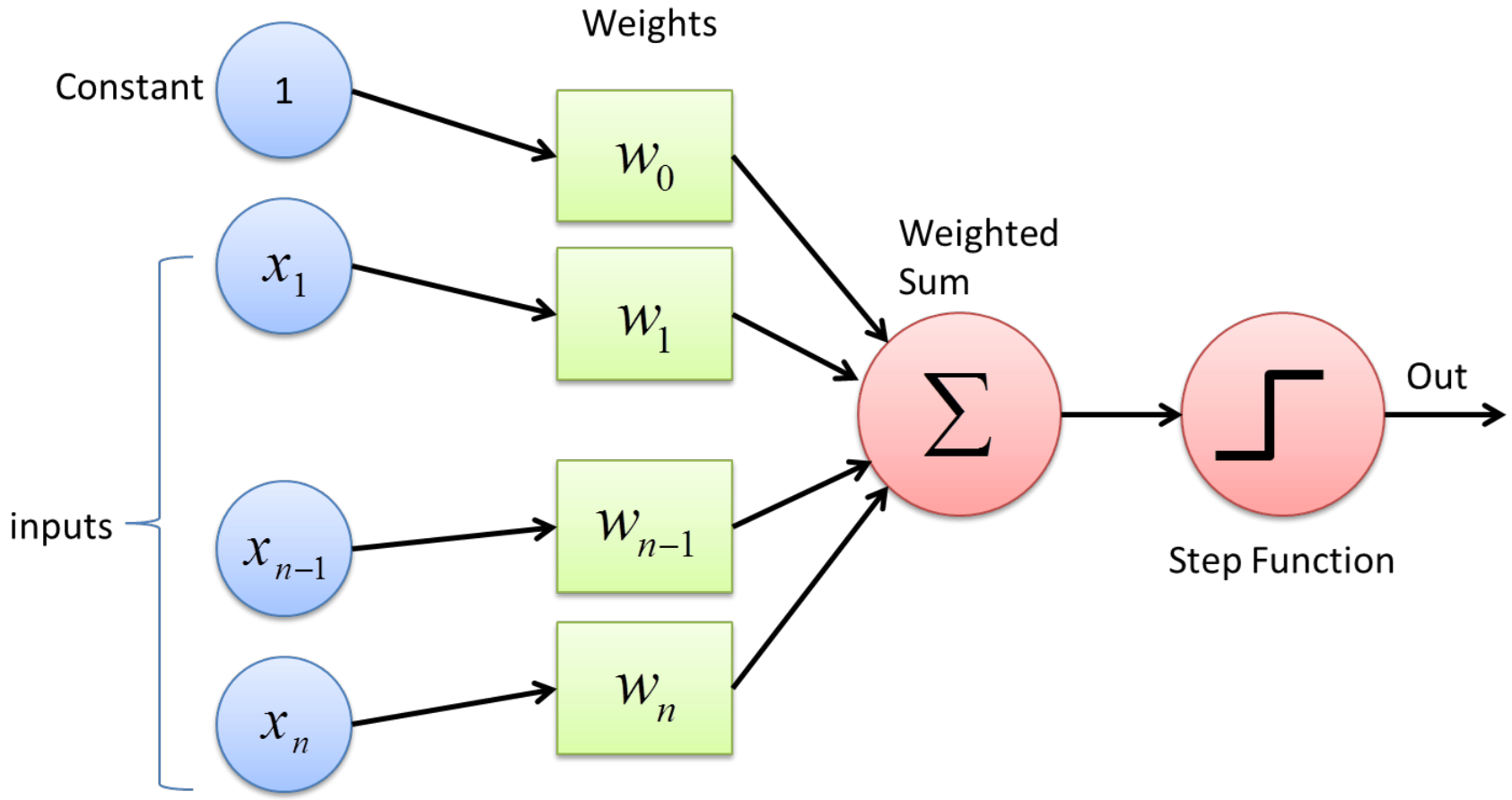
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

k Πλησιέστεροι γείτονες (kNN)

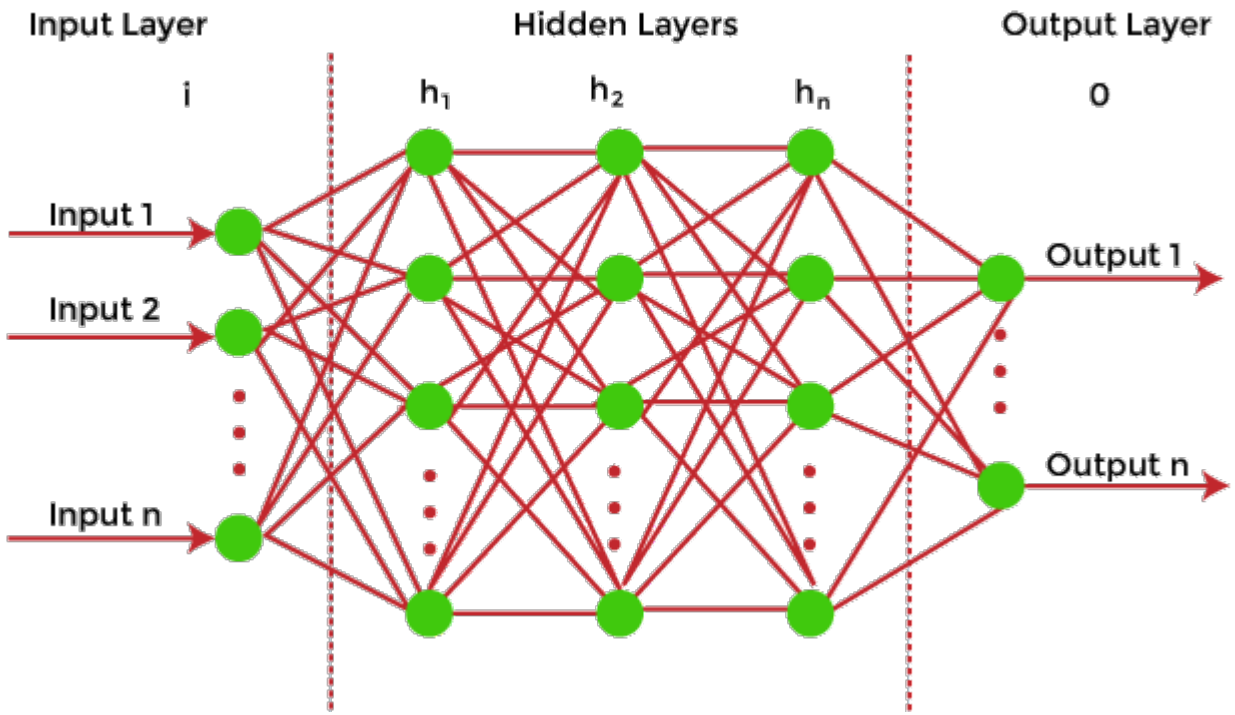


Perceptron


Weights
Constant
1
w_0
x_1
w_1
Weighted Sum
Σ
Out
inputs
x_{n-1}
w_{n-1}
x_n
w_n
Step Function

Co-funded by the Erasmus+ Programme of the European Union

Πολυστρωματικά Perceptrons (MLP)



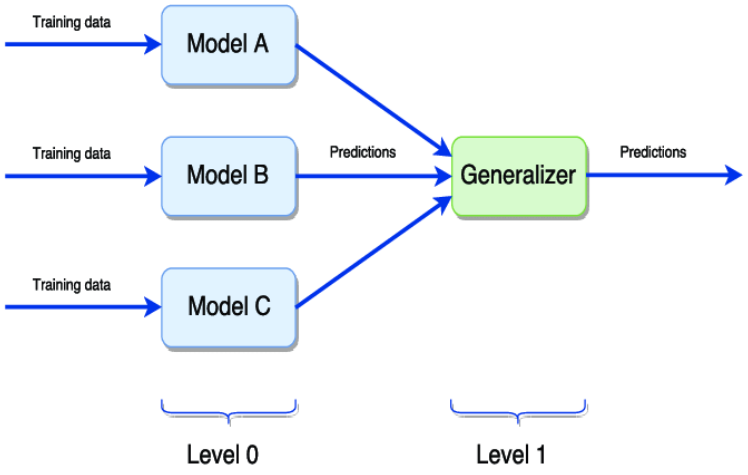
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Μέθοδοι συνόλου στη Μηχανική Μάθηση

Τύποι μεθόδων συνόλου :

1. **Bagging (Συγκέντρωση Bootstrap)** :
   - Χρησιμοποιεί διαφορετικά υποσύνολα δεδομένων εκπαίδευσης για την εκπαίδευση πολλαπλών μοντέλων.
   - Βρίσκει τον μέσο όρο του αποτελέσματος (για παλινδρόμηση) ή λαμβάνει πλειοψηφία (για ταξινόμηση).
   - Παράδειγμα: Τυχαία δάση.
2. **Ενίσχυση** :
   - Εκπαιδεύει τα μοντέλα διαδοχικά, όπου κάθε μοντέλο διορθώνει τα λάθη του προκατόχου του.
   - Εστιάζει σε περιπτώσεις που είχαν ταξινομηθεί εσφαλμένα από προηγούμενα μοντέλα.
   - Παραδείγματα: AdaBoost, Gradient Boosting, XGBoost.
3. **Στοίβαξη** :
   - Συνδυάζει πολλαπλά μοντέλα ταξινόμησης ή παλινδρόμησης μέσω ενός μετα-ταξινομητή ή ενός μετα-αναδρομητή.
   - Τα μοντέλα βασικού επιπέδου εκπαιδεύονται σε ένα πλήρες σετ εκπαίδευσης, ενώ το μετα-μοντέλο εκπαιδεύεται στις εξόδους του μοντέλου βασικού επιπέδου ως χαρακτηριστικά.

- **Ορισμός** : Οι μέθοδοι συνόλου συνδυάζουν πολλαπλά μοντέλα μηχανικής εκμάθησης για να παράγουν ένα βέλτιστο μοντέλο πρόβλεψης.
- **Γιατί να χρησιμοποιήσετε Ensemble;** :
  - Βελτιώστε την ακρίβεια.
  - Μειώστε την υπερπροσαρμογή.
  - Βελτιώστε την ευρωστία του μοντέλου.



Συμπέρασμα και Πόροι

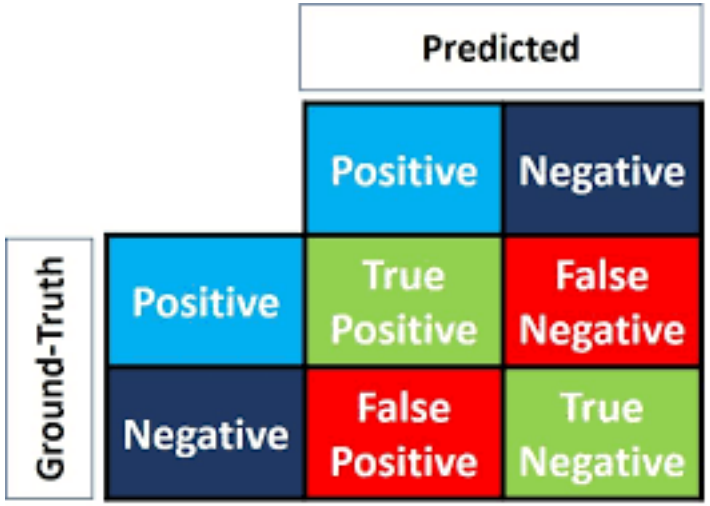
- **Ανακεφαλαίωση** :
  1. Εξερευνήστε τις θεμελιώδεις έννοιες του AI, ML και DL.
  2. Εμβαθύνοντας στα βασικά προβλήματα που στοχεύει να λύσει η ML.
  3. Μη συσκευασμένες βασικές τεχνικές ML και οι εφαρμογές τους.
- **Γιατί έχει σημασία η ML** :
  1. Η μηχανική μάθηση βρίσκεται στο επίκεντρο πολλών σύγχρονων τεχνολογικών εξελίξεων.
  2. Οι εφαρμογές του καλύπτουν διάφορους τομείς, από την υγειονομική περίθαλψη έως τη χρηματοδότηση, καθιστώντας το μια κρίσιμη δεξιότητα για το μέλλον.
- **Πόροι για περαιτέρω μάθηση** :
  1. **Βιβλία** :
     - «Αναγνώριση προτύπων και μηχανική μάθηση» του Christopher Bishop.
     - «The Hundred-Page Machine Learning Book» του Andriy Burkov.
  2. **Διαδικτυακά Μαθήματα** :
     - Coursera: "Machine Learning" του Andrew Ng.
     - Udacity: «Εισαγωγή στη Μηχανική Μάθηση».
  3. **Ιστοσελίδες** :
     - Προς Επιστήμη Δεδομένων
     - Μάθηση Μηχανικής

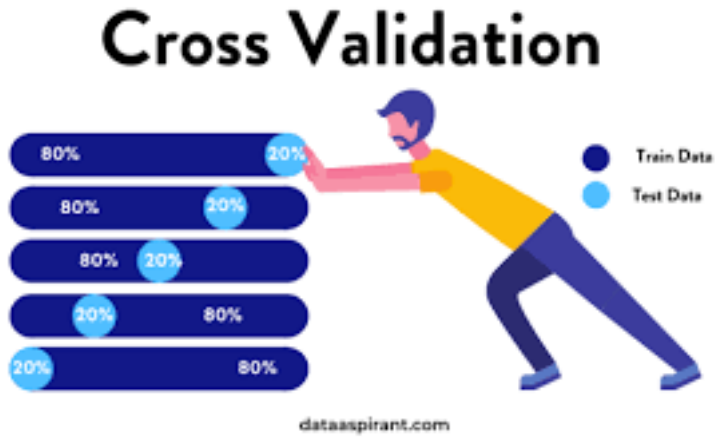
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Επικύρωση, Μετρήσεις και Μηχανική Χαρακτηριστικών στο AI

Εισαγωγή στην Αξιολόγηση και Βελτιστοποίηση Μοντέλων





Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή στην Αξιολόγηση και Βελτιστοποίηση Μοντέλων

1. **Αξιολόγηση μοντέλου:**
   - Η διαδικασία προσδιορισμού του πόσο καλά αποδίδει ένα μοντέλο μηχανικής μάθησης σε αόρατα δεδομένα.
   - γενίκευσης του μοντέλου .
2. **Βελτιστοποίηση μοντέλου:**
   - Βελτιώστε το μοντέλο για να βελτιώσετε την απόδοσή του.
   - Περιλαμβάνει τεχνικές όπως συντονισμός υπερπαραμέτρων, επιλογή χαρακτηριστικών και πολλά άλλα.
3. **Συνάφεια με τα τεχνητά νευρωνικά δίκτυα (ANN):**
   - Τα ANN είναι ισχυρά, αλλά μπορούν εύκολα να υπερπροσαρμόσουν ή να υπολειτουργήσουν χωρίς σωστή αξιολόγηση και βελτιστοποίηση.
   - Η σωστή αξιολόγηση διασφαλίζει ότι το μοντέλο νευρωνικών δικτύων είναι ισχυρό και αξιόπιστο .

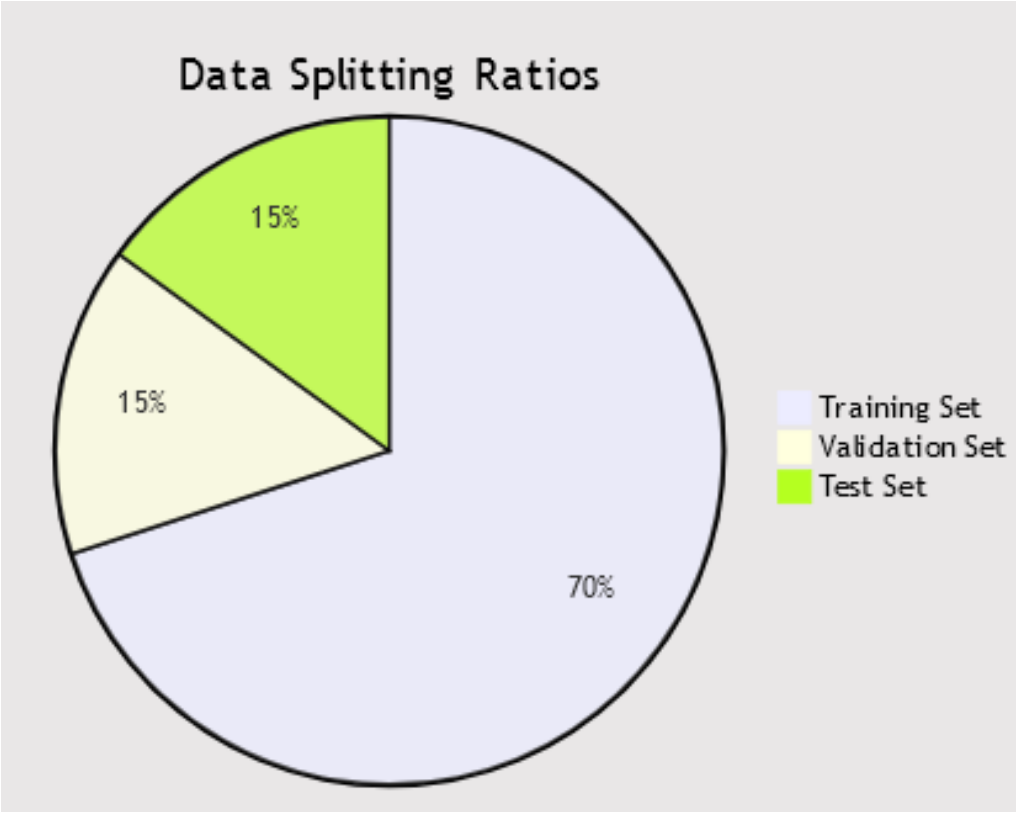
Ακριβώς όπως η κατασκευή ενός μοντέλου είναι ζωτικής σημασίας, η αξιολόγηση της απόδοσής του και η βελτιστοποίησή του για καλύτερα αποτελέσματα είναι εξίσου σημαντική, ειδικά στο πλαίσιο σύνθετων μοντέλων όπως τα ANN.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Διαχωρισμός δεδομένων

1. **Σημασία του διαχωρισμού δεδομένων:**
   - Διασφαλίζει ότι το μοντέλο μπορεί να γενικευτεί καλά σε νέα, αόρατα δεδομένα.
   - Βοηθά στην αξιολόγηση της απόδοσης του μοντέλου
2. **Σετ προπόνησης:**
   - Χρησιμοποιείται για την εκπαίδευση του μοντέλου μηχανικής εκμάθησης.
   - Συνήθως το μεγαλύτερο μέρος του συνόλου δεδομένων.
3. **Σετ επικύρωσης:**
   - Χρησιμοποιείται για τη ρύθμιση υπερπαραμέτρων και τη λήψη αποφάσεων σχετικά με την αρχιτεκτονική του μοντέλου.
   - Βοηθά στην πρόληψη της υπερβολικής προσαρμογής κατά τη φάση της προπόνησης.
4. **Σετ δοκιμής:**
   - Χρησιμοποιείται για την αξιολόγηση της τελικής απόδοσης του μοντέλου μετά την εκπαίδευση και την επικύρωση.
   - Δεν πρέπει ποτέ να χρησιμοποιείται κατά τη φάση της εκπαίδευσης για να διασφαλιστεί η αμερόληπτη αξιολόγηση.



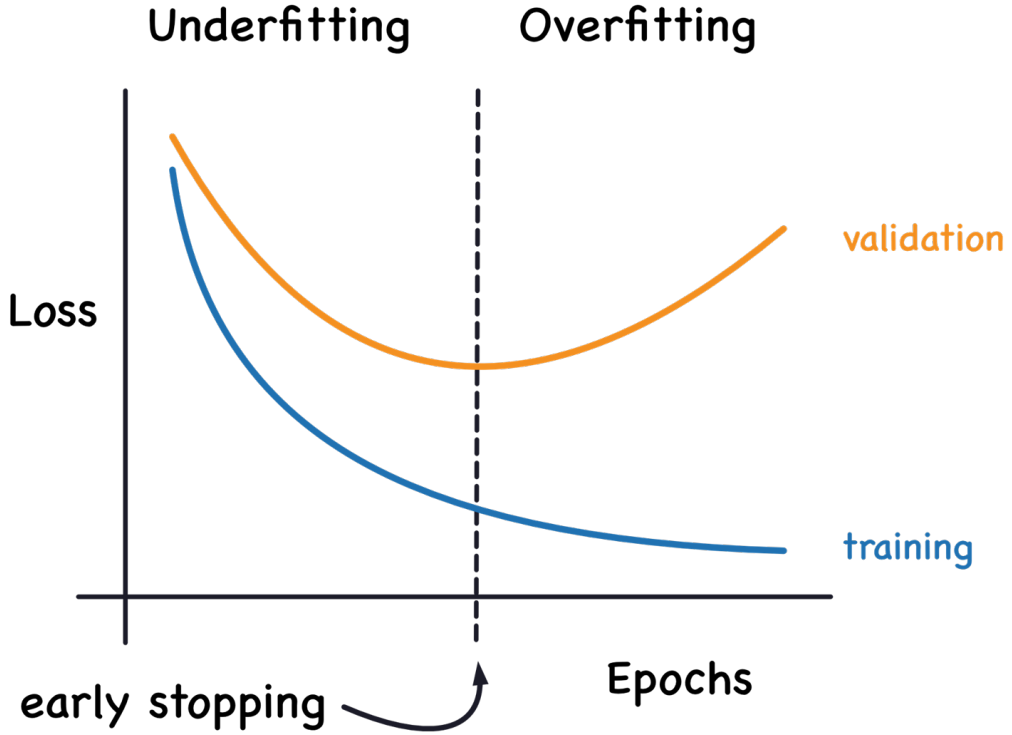
Ο σωστός διαχωρισμός δεδομένων είναι θεμελιώδης για τη δημιουργία ενός αξιόπιστου και ισχυρού μοντέλου μηχανικής εκμάθησης. Διασφαλίζει ότι έχουμε μια δίκαιη αξιολόγηση των δυνατοτήτων και των αδυναμιών του μοντέλου μας

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Κατανόηση των συνόλων επικύρωσης και δοκιμής

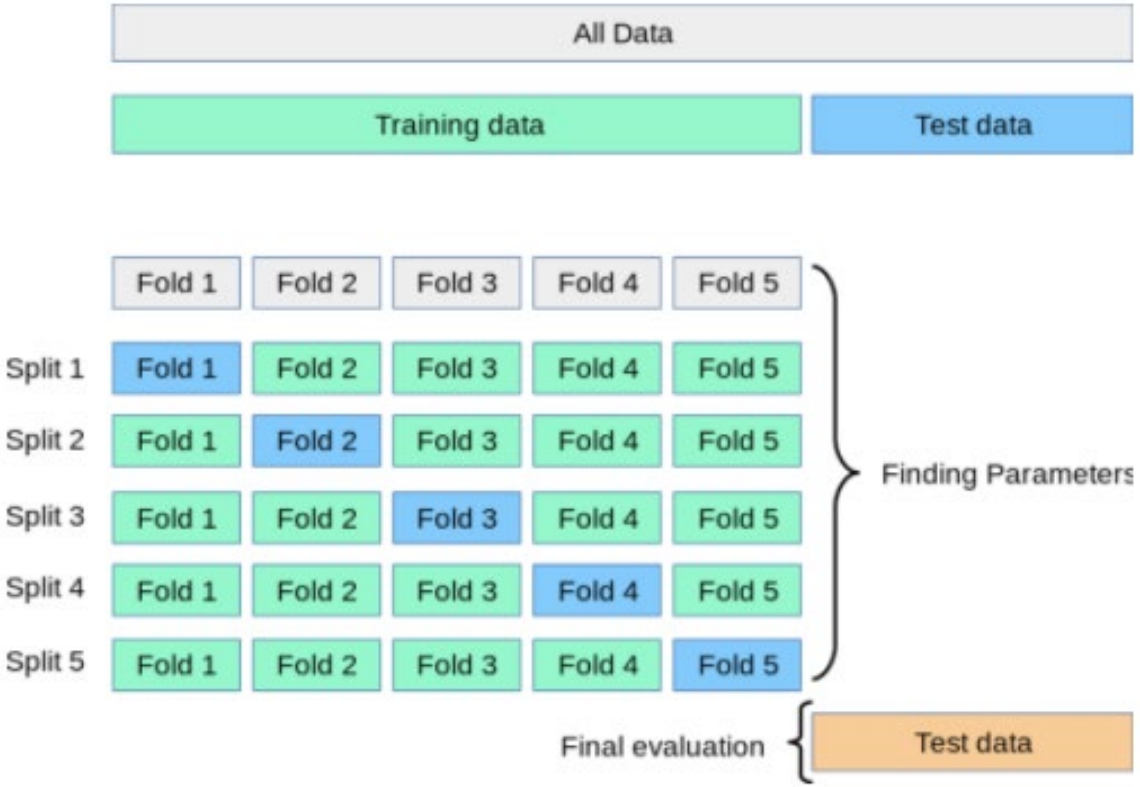
1. **Σετ επικύρωσης:**
   - Ορισμός : Ένα υποσύνολο του συνόλου δεδομένων που χρησιμοποιείται για την αξιολόγηση της απόδοσης του μοντέλου κατά τη διάρκεια της εκπαίδευσης και τη λήψη αποφάσεων σχετικά με την αρχιτεκτονική του.
   - Σκοπός : Βοηθά στον συντονισμό των παραμέτρων του μοντέλου, στην επιλογή χαρακτηριστικών και στην αποφυγή της υπερβολικής προσαρμογής.
2. **Σετ δοκιμής:**
   - Ορισμός : Ένα υποσύνολο του συνόλου δεδομένων που χρησιμοποιείται για την αξιολόγηση της απόδοσης του μοντέλου αφού έχει εκπαιδευτεί και επικυρωθεί.
   - Σκοπός : Παρέχει μια αμερόληπτη αξιολόγηση της τελικής απόδοσης του μοντέλου σε αόρατα δεδομένα.
3. **Υπερπροσαρμογή έναντι Υποπροσαρμογής:**
   - Υπερπροσαρμογή : Όταν το μοντέλο αποδίδει εξαιρετικά καλά στα δεδομένα εκπαίδευσης αλλά κακώς σε νέα, αόρατα δεδομένα.
   - Υποπροσαρμογή : Όταν το μοντέλο έχει κακή απόδοση τόσο στα δεδομένα εκπαίδευσης όσο και στα νέα δεδομένα.
   - Σημασία : Η σωστή χρήση των συνόλων επικύρωσης και δοκιμών βοηθά στον εντοπισμό και την πρόληψη τόσο της υπερπροσαρμογής όσο και της υποσυναρμολόγησης.



Η επικύρωση και τα σύνολα δοκιμών διαδραματίζουν κρίσιμο ρόλο στη διασφάλιση της ευρωστίας και της αξιοπιστίας των μοντέλων μηχανικής μάθησης. Βοηθούν στην επίτευξη ισορροπίας μεταξύ της πολυπλοκότητας του μοντέλου και της απόδοσης.

Διασταυρωμένη επικύρωση

1. **Τι είναι η Cross Validation;**
   - Ορισμός : Μια τεχνική που χρησιμοποιείται για την αξιολόγηση της απόδοσης των μοντέλων μηχανικής μάθησης διαχωρίζοντας το αρχικό δείγμα σε ένα σύνολο εκπαίδευσης για την εκπαίδευση του μοντέλου και σε ένα σύνολο επικύρωσης για την αξιολόγησή του.
   - Σκοπός : Μείωση της υπερπροσαρμογής και απόκτηση ενός πιο γενικευμένου μοντέλου.
2. **K-Fold Cross Validation:**
   - Επεξήγηση : Το αρχικό δείγμα χωρίζεται τυχαία σε υποδείγματα ίσου μεγέθους «K». Από τα υποδείγματα «K», ένα μεμονωμένο υποδείγμα διατηρείται ως δεδομένα επικύρωσης και τα υπόλοιπα υποδείγματα «K-1» χρησιμοποιούνται ως δεδομένα εκπαίδευσης. Η διαδικασία επαναλαμβάνεται «K» φορές.
3. **Οφέλη της διασταυρούμενης επικύρωσης:**
   - Πιο αξιόπιστη εκτίμηση της απόδοσης του μοντέλου.
   - Χρησιμοποιεί ολόκληρο το σύνολο δεδομένων τόσο για εκπαίδευση όσο και για επικύρωση.
   - Βοηθά στον εντοπισμό των καλύτερων υπερπαραμέτρων μοντέλου.



Η Cross Validation είναι μια ισχυρή μέθοδος για την αξιολόγηση της απόδοσης των μοντέλων μηχανικής μάθησης, διασφαλίζοντας ότι το μοντέλο είναι καλά εκπαιδευμένο και **γενικευμένο** , και όχι απλώς **απομνημόνευση** των δεδομένων εκπαίδευσης.

Οφέλη από διασταυρούμενη επικύρωση

1. **Αξιολόγηση ισχυρού μοντέλου:**
   - Παρέχει μια πιο ολοκληρωμένη αξιολόγηση της απόδοσης ενός μοντέλου από ό,τι μια ενιαία διαίρεση δοκιμής αμαξοστοιχίας.
   - Βοηθά στην κατανόηση της συμπεριφοράς του μοντέλου σε διαφορετικά υποσύνολα δεδομένων.
2. **Μετριάζει την υπερπροσαρμογή:**
   - Με την εκπαίδευση σε διαφορετικά υποσύνολα και την επικύρωση σε διαφορετικά σύνολα, μειώνονται οι πιθανότητες το μοντέλο να απομνημονεύει συγκεκριμένα μοτίβα δεδομένων.
   - Διασφαλίζει ότι το μοντέλο γενικεύεται καλά σε αόρατα δεδομένα.
3. **Βέλτιστη χρήση δεδομένων:**
   - Κάθε σημείο δεδομένων πρέπει να βρίσκεται τόσο σε σετ εκπαίδευσης όσο και σε σετ επικύρωσης.
   - Ιδιαίτερα ωφέλιμο όταν το μέγεθος δεδομένων είναι περιορισμένο.
4. **Ρύθμιση υπερπαραμέτρων:**
   - Βοηθά στον εντοπισμό του καλύτερου συνόλου υπερπαραμέτρων για το μοντέλο.
   - Μπορεί να συνδυαστεί με αναζήτηση πλέγματος ή τυχαία αναζήτηση για βέλτιστες παραμέτρους μοντέλου.

Η Cross Validation δεν είναι απλώς μια τεχνική για την αξιολόγηση του μοντέλου. είναι ένα ισχυρό εργαλείο που διασφαλίζει τη δημιουργία ενός ισχυρού, καλά γενικευμένου και βελτιστοποιημένου μοντέλου μηχανικής εκμάθησης.

Co-funded by the Erasmus+ Programme of the European Union

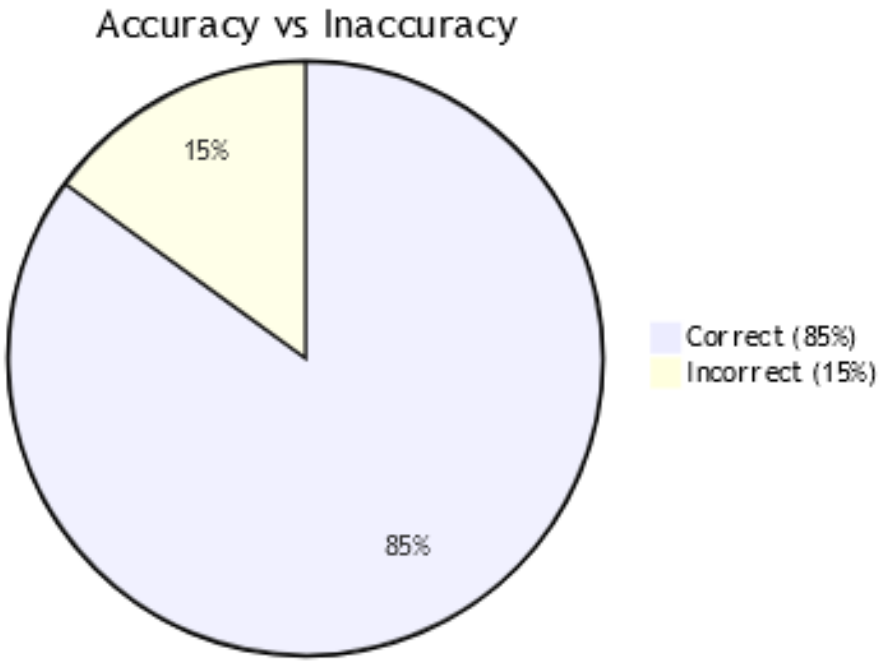
Μετρικές Εισαγωγή

1. **Γιατί Metrics;**
   - Σημασία: Οι μετρικές παρέχουν ποσοτικά μέτρα για την αξιολόγηση της απόδοσης των μοντέλων μηχανικής εκμάθησης.
   - Ρόλος: Καθοδηγούν τη διαδικασία βελτιστοποίησης και βοηθούν στην επιλογή μοντέλου.
2. **Τύποι μετρήσεων:**
   - **Μετρήσεις ταξινόμησης:** Χρησιμοποιούνται για μοντέλα που προβλέπουν κατηγορικά αποτελέσματα. Παραδείγματα περιλαμβάνουν Ακρίβεια, Ακρίβεια, Ανάκληση, βαθμολογία F1 κ.λπ.
   - **Μετρήσεις παλινδρόμησης:** Χρησιμοποιούνται για μοντέλα που προβλέπουν συνεχή αποτελέσματα. Παραδείγματα περιλαμβάνουν το μέσο απόλυτο σφάλμα (MAE), το μέσο τετράγωνο σφάλμα (MSE), το R-τετράγωνο κ.λπ.
3. **Επιλέγοντας τη σωστή μέτρηση:**
   - Εξαρτάται από το πρόβλημα που αντιμετωπίζει: Για παράδειγμα, στις ιατρικές διαγνώσεις, τα ψευδώς αρνητικά μπορεί να είναι πιο δαπανηρά από τα ψευδώς θετικά.
   - Αντισταθμίσεις εξισορρόπησης: Ακρίβεια έναντι ανάκλησης, Μεροληψία έναντι διακύμανσης.

Οι μετρήσεις αξιολόγησης μοντέλων είναι ζωτικής σημασίας στον αγωγό μηχανικής μάθησης. Παρέχουν πληροφορίες για τα δυνατά και αδύνατα σημεία του μοντέλου, καθοδηγώντας τις βελτιώσεις και διασφαλίζοντας ότι το μοντέλο ευθυγραμμίζεται με τους συγκεκριμένους στόχους της εργασίας.

Ακρίβεια

1. **Ορισμός Ακρίβειας:**
   - Τύπος: Ακρίβεια = (Αριθμός σωστών προβλέψεων)/(Συνολικός αριθμός προβλέψεων)
   - Ερμηνεία: Αναλογία των συνολικών προβλέψεων που είναι σωστές.
2. **Πότε να χρησιμοποιήσετε την ακρίβεια:**
   - Κατάλληλο για ισορροπημένα σύνολα δεδομένων όπου ο αριθμός των δειγμάτων σε κάθε τάξη είναι περίπου ο ίδιος.
   - Παρέχει ένα γρήγορο στιγμιότυπο της συνολικής απόδοσης του μοντέλου.
3. **Περιορισμοί Ακρίβειας:**
   - Μπορεί να είναι παραπλανητικό για μη ισορροπημένα σύνολα δεδομένων.
   - Δεν παρέχει πληροφορίες για τους τύπους σφαλμάτων που έγιναν (ψευδώς θετικά έναντι ψευδώς αρνητικών).



Η ακρίβεια είναι μια απλή και ευρέως χρησιμοποιούμενη μέτρηση. Ωστόσο, είναι σημαντικό να γνωρίζετε τους περιορισμούς του και να βεβαιωθείτε ότι είναι η σωστή μέτρηση για την εκάστοτε εργασία.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ευαισθησία (Ανάκληση) και Ειδικότητα

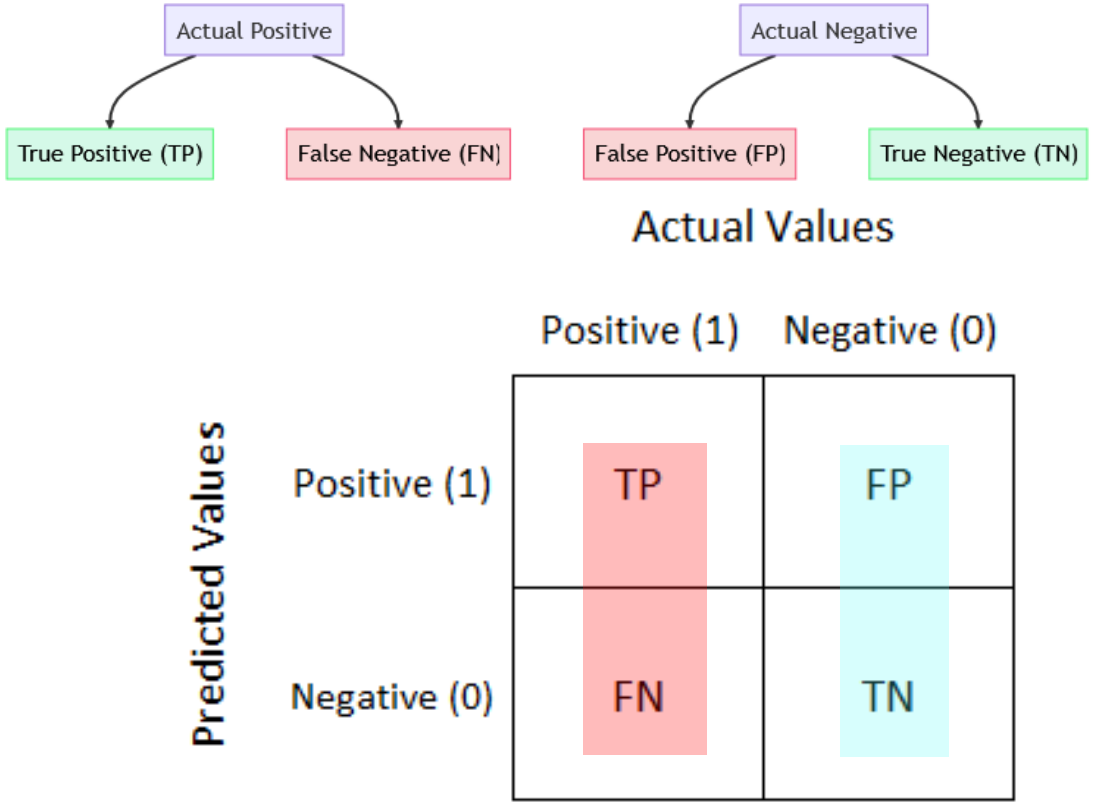
1. **Ορισμοί:**
   - **Ευαισθησία (True Positive Rate):** Η ικανότητα ενός μοντέλου να εντοπίζει σωστά τις θετικές περιπτώσεις μεταξύ όλων των πραγματικών θετικών περιπτώσεων.

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

   - **Ειδικότητα (True Negative Rate):** Η ικανότητα ενός μοντέλου να εντοπίζει σωστά τις αρνητικές περιπτώσεις μεταξύ όλων των πραγματικών αρνητικών περιπτώσεων.

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positives}$$

2. **Σημασια:**
   - **Ιατρικές εξετάσεις:** Η ευαισθησία είναι ζωτικής σημασίας για να διασφαλιστεί ότι δεν χάνονται ασθένειες (**ελαχιστοποίηση των ψευδώς αρνητικών**), ενώ η Ειδικότητα διασφαλίζει ότι τα υγιή άτομα δεν διαγιγνώσκονται λανθασμένα (**ελαχιστοποιώντας τα ψευδώς θετικά**).
   - **Ανισόρροπα σύνολα δεδομένων:** Σε σύνολα δεδομένων όπου η μία κλάση υπερτερεί σημαντικά της άλλης, η ευαισθησία και η ειδικότητα παρέχουν μια σαφέστερη εικόνα της απόδοσης του μοντέλου από την Ακρίβεια.



Η ευαισθησία και η ειδικότητα είναι βασικές μετρήσεις, ειδικά σε σενάρια οπου οι συνεπειες της εσφαλμενης ταξινομησης ειναι σημαντικές. Παρέχουν μια πιο λεπτή κατανόηση της απόδοσης ενός μοντέλου, ειδικά στο πλαίσιο προβλημάτων δυαδικής ταξινόμησης.

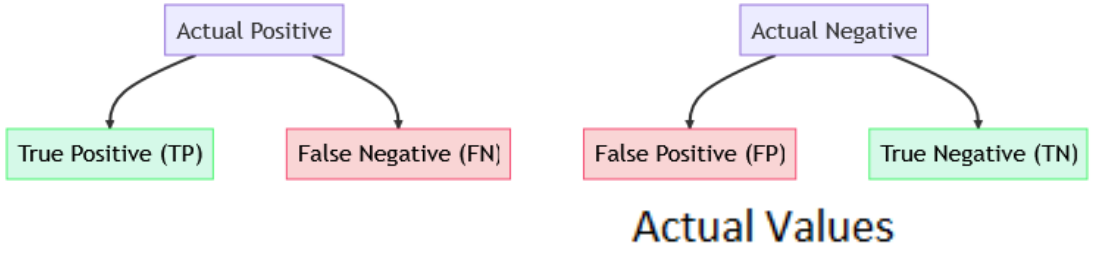
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ακρίβεια και ανάκληση

Ορισμοί:

- **Ακρίβεια:** Από όλες τις προβλεπόμενες θετικές περιπτώσεις, πόσες ήταν πραγματικά θετικές; Τύπος: Ακρίβεια = Αληθινά θετικά / (αληθινά θετικά + ψευδώς θετικά)

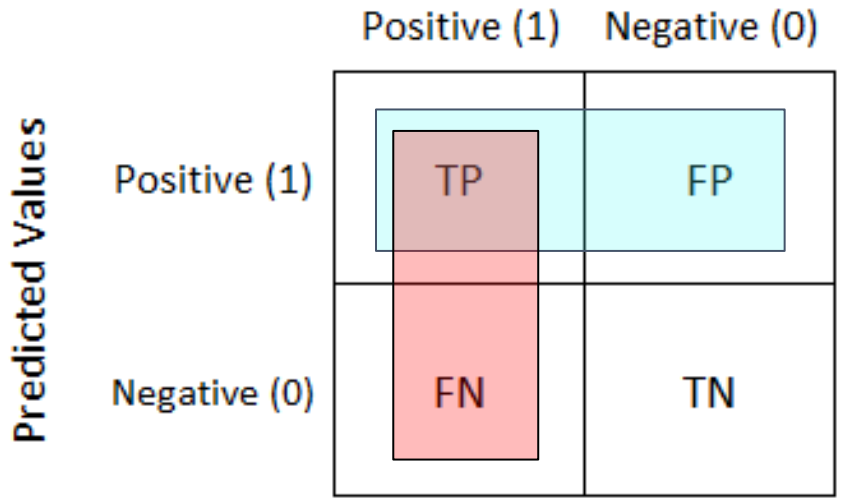


Πράξη εξισορρόπησης:

- Η ακρίβεια και η ανάκληση (ευαισθησία) σχετίζονται συχνά αντιστρόφως. Η βελτίωση του ενός μπορεί να μειώσει το άλλο.
- Η επιλογή μεταξύ ιεράρχησης της Ακρίβειας ή της Ανάκλησης εξαρτάται από το συγκεκριμένο πρόβλημα και τις επιπτώσεις του.

Περιπτώσεις χρήσης:

- **Υψηλή ακρίβεια:** Όταν το κόστος των ψευδών θετικών είναι υψηλό. Π.χ., ο εντοπισμός ανεπιθύμητων μηνυμάτων ηλεκτρονικού ταχυδρομείου όπου τα νόμιμα μηνύματα ηλεκτρονικού ταχυδρομείου που ταξινομούνται ως ανεπιθύμητα μπορεί να είναι προβληματική.
- **Υψηλή ανάκληση:** Όταν το κόστος των ψευδών αρνητικών είναι υψηλό. Π.χ., διάγνωση ασθενειών όπου η απώλεια ενός πραγματικού θετικού κρούσματος μπορεί να έχει σοβαρές συνέπειες.



Η ακρίβεια και η ανάκληση είναι κρίσιμες μετρήσεις που παρέχουν πληροφορίες για την αποοοση των μοντελων ταξινομησης, ειοικα σε μη ισορροπημένα σύνολα δεδομένων. Η κατανόηση των ορισμών, των ανταλλαγών και των συνεπειών τους είναι απαραίτητη για τη δημιουργία και την αξιολόγηση ισχυρών μοντέλων.

F1-Βαθμολογία

1. **Ορισμός της βαθμ**
   - Τύπος: $F_1 = 2 * \frac{Precision * Recall}{Precision + Recall}$
   - Ερμηνεία: Αρμονικός μέσος όρος Ακρίβειας και Ανάκλησης, που παρέχει μια ισορροπία μεταξύ των δύο.
2. **Γιατί F1-Score;**
   - Σημασία σε μη ισορροπημένα σύνολα δεδομένων: Όταν μια κλάση υπερτερεί σημαντικά της άλλης, η Ακρίβεια μπορεί να είναι παραπλανητική. Το F1-Score παρέχει μια πιο ισορροπημένη μέτρηση.
   - Συνδυάζει τόσο τα ψευδώς θετικά όσο και τα ψευδώς αρνητικά σε μια ενιαία μέτρηση.
3. **Πότε να χρησιμοποιήσετε το F1-Score:**
   - Όταν τόσο τα ψευδώς θετικά όσο και τα ψευδώς αρνητικά είναι κρίσιμα.
   - Όταν υπάρχει άνιση κατανομή της τάξης.

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (1)$$

$$precision = \frac{TP}{TP + FP} \quad (2)$$

$$recall = \frac{TP}{TP + FN} \quad (3)$$

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Μείωση Διαστάσεων

1. **Τι είναι η μείωση διαστάσεων;**
   - Ορισμός: Η διαδικασία μείωσης του αριθμού των υπό εξέταση τυχαίων μεταβλητών με τη λήψη ενός συνόλου κύριων μεταβλητών.
   - Σκοπός: Απλοποιήστε τα δεδομένα χωρίς να χάσετε πολλές πληροφορίες, διευκολύνοντας την οπτικοποίηση, την ανάλυση και την επεξεργασία.
2. **Γιατί μείωση διαστάσεων;**
   - **Οπτικοποίηση δεδομένων:** Η μείωση των διαστάσεων (π.χ. από εκατοντάδες σε 2 ή 3) μπορεί να βοηθήσει στην οπτικοποίηση μοτίβων δεδομένων.
   - **Επιτάχυνση αλγορίθμων:** Λιγότερα δεδομένα σημαίνει ταχύτερους χρόνους εκπαίδευσης για μοντέλα μηχανικής εκμάθησης.
   - **Αποφύγετε την υπερβολική τοποθέτηση:** Η μείωση των περιττών χαρακτηριστικών μπορεί να βοηθήσει στην αποφυγή υπερβολικής τοποθέτησης μοντέλων.
3. **Κοινές τεχνικές:**
   - **PCA (Principal Component Analysis):** Γραμμική τεχνική που βρίσκει τους άξονες στα δεδομένα που μεγιστοποιούν τη διακύμανση.
   - **t-SNE (t-Distributed Stochastic Neighbor Embedding):** Μη γραμμική τεχνική κατάλληλη για οπτικοποίηση δεδομένων υψηλών διαστάσεων σε 2D ή 3D.
   - **LDA (Linear Discriminant Analysis):** Εποπτευόμενη μέθοδος που χρησιμοποιείται για την εύρεση ενός γραμμικού συνδυασμού χαρακτηριστικών που διαχωρίζει καλύτερα δύο ή περισσότερες κλάσεις.

Το Dimensionality Reduction είναι ένα ισχυρό εργαλείο στην επιστήμη των δεδομένων και τη μηχανική μάθηση, που βοηθά στην οπτικοποίηση δεδομένων, στην επιτάχυνση των αλγορίθμων και στην πρόληψη της υπερπροσαρμογής. Είναι σημαντικό να επιλέξετε τη σωστή τεχνική με βάση τη φύση και τον σκοπό των δεδομένων.

Επιλογή χαρακτηριστικών

1. **Τι είναι η επιλογή χαρακτηριστικών;**
   - Ορισμός: Η διαδικασία επιλογής ενός υποσυνόλου σχετικών χαρακτηριστικών (μεταβλητές, προγνωστικοί παράγοντες) για χρήση στην κατασκευή μοντέλων.
   - Σκοπός: Βελτίωση της απόδοσης του μοντέλου, μείωση της υπερβολικής προσαρμογής και ενίσχυση της ερμηνείας του μοντέλου.
2. **Γιατί Επιλογή Χαρακτηριστικών;**
   - **Βελτιωμένη απόδοση:** Με την κατάργηση άσχετων ή περιττών λειτουργιών, τα μοντέλα μπορούν συχνά να έχουν καλύτερη απόδοση.
   - **Ταχύτερη εκπαίδευση:** Λιγότερες δυνατότητες σημαίνουν μειωμένο υπολογιστικό κόστος και ταχύτερους χρόνους εκπαίδευσης.
   - **Βελτιωμένη ερμηνευτικότητα:** Ένα απλούστερο μοντέλο με λιγότερα χαρακτηριστικά είναι πιο εύκολο να κατανοηθεί και να ερμηνευτεί.
3. **Κοινές τεχνικές:**
   - **Μέθοδοι φίλτρου:** Με βάση τις εγγενείς ιδιότητες των δεδομένων. Τα παραδείγματα περιλαμβάνουν βαθμολογίες συντελεστών συσχέτισης και τεστ χ-τετράγωνο.
   - **Μέθοδοι περιτυλίγματος:** Χρησιμοποιήστε ένα υποσύνολο χαρακτηριστικών και εκπαιδεύστε ένα μοντέλο χρησιμοποιώντας τις. Στα παραδείγματα περιλαμβάνονται η επιλογή προς τα εμπρός χαρακτηριστικών, η εξάλειψη χαρακτηριστικών προς τα πίσω και η αναδρομική εξάλειψη χαρακτηριστικών.
   - **Ενσωματωμένες μέθοδοι:** Αλγόριθμοι που έχουν ενσωματωμένες μεθόδους επιλογής χαρακτηριστικών. Παραδείγματα περιλαμβάνουν τα δέντρα LASSO και αποφάσεων.

Η επιλογή χαρακτηριστικών είναι ένα κρίσιμο βήμα στη γραμμή μηχανικής εκμάθησης. Όχι μόνο βελτιώνει την απόδοση του μοντέλου, αλλά διασφαλίζει επίσης ότι τα μοντέλα είναι αποτελεσματικά και ερμηνεύσιμα. Η επιλογή της μεθόδου επιλογής χαρακτηριστικών θα πρέπει να ευθυγραμμίζεται με τη φύση των δεδομένων και τους συγκεκριμένους στόχους της εργασίας.

Co-funded by the Erasmus+ Programme of the European Union

Οφέλη μείωσης διαστάσεων & επιλογής χαρακτηριστικών

1. **Βελτιωμένη απόδοση μοντέλου:**
   - Η κατάργηση άσχετων ή περιττών δεδομένων μπορεί να οδηγήσει σε καλύτερη ακρίβεια και γενίκευση του μοντέλου.
   - Βοηθά στη μείωση του θορύβου στα δεδομένα.
2. **Αποδοτικότητα & Ταχύτητα:**
   - Μειωμένο υπολογιστικό κόστος λόγω λιγότερων σημείων δεδομένων και χαρακτηριστικών.
   - Ταχύτεροι χρόνοι εκπαίδευσης και πρόβλεψης για μοντέλα μηχανικής εκμάθησης.
3. **Βελτιωμένη οπτικοποίηση δεδομένων:**
   - Τα δεδομένα χαμηλότερης διάστασης μπορούν εύκολα να οπτικοποιηθούν, βοηθώντας στην κατανόηση προτύπων και σχέσεων.
   - Βοηθά στον εντοπισμό συμπλεγμάτων ή ομάδων στα δεδομένα.
4. **Μειωμένος κίνδυνος υπερβολικής τοποθέτησης:**
   - Τα απλοποιημένα μοντέλα με λιγότερα χαρακτηριστικά είναι λιγότερο πιθανό να υπερπροσαρμόσουν στα δεδομένα εκπαίδευσης.
   - Διασφαλίζει ότι το μοντέλο καταγράφει τα υποκείμενα μοτίβα και όχι τον θόρυβο.
5. **Βελτιωμένη ερμηνευτικότητα:**
   - Τα πιο απλά μοντέλα είναι ευκολότερα κατανοητά και εξηγούνται.
   - Σημαντικό για εφαρμογές όπου η διαφάνεια του μοντέλου είναι ζωτικής σημασίας, όπως σε ιατρικούς ή χρηματοοικονομικούς τομείς.

Τόσο η μείωση διαστάσεων όσο και η επιλογή χαρακτηριστικών παίζουν καθοριστικό ρόλο στη φάση προεπεξεργασίας δεδομένων. Διασφαλίζουν ότι τα μοντέλα μηχανικής μάθησης δεν είναι μόνο ακριβή αλλά και αποτελεσματικά, ερμηνεύσιμα και γενικεύσιμα.

Ενσωμάτωση με νευρωνικά δίκτυα

1. **Προεπεξεργασία για νευρωνικά δίκτυα:**
   - Τα νευρωνικά δίκτυα επωφελούνται συχνά από απλοποιημένα και τυποποιημένα δεδομένα εισόδου.
   - Η μείωση διαστάσεων και η επιλογή χαρακτηριστικών μπορούν να χρησιμεύσουν ως βήματα προεπεξεργασίας πριν από την τροφοδοσία δεδομένων σε ένα νευρωνικό δίκτυο.
2. **Βελτιωμένη απόδοση προπόνησης:**
   - Οι μειωμένες διαστάσεις και χαρακτηριστικά δεδομένων μπορούν να οδηγήσουν σε ταχύτερη σύγκλιση κατά τη διάρκεια της εκπαίδευσης.
   - Βοηθά στην αποτελεσματική χρήση των υπολογιστικών πόρων, ειδικά σε μοντέλα βαθιάς μάθησης με πολλές παραμέτρους.
3. **Βελτιωμένη γενίκευση:**
   - Τα απλοποιημένα δεδομένα μπορούν να μειώσουν τον κίνδυνο υπερπροσαρμογής στα νευρωνικά δίκτυα.
   - Διασφαλίζει ότι το δίκτυο μαθαίνει σημαντικά μοτίβα αντί για θόρυβο.
4. **Χαρακτηριστικά στρώματα εξαγωγής:**
   - Ορισμένες αρχιτεκτονικές νευρωνικών δικτύων, ειδικά τα συνελικτικά νευρωνικά δίκτυα (CNN), εκτελούν εγγενώς εξαγωγή χαρακτηριστικών.
   - Ο συνδυασμός της παραδοσιακής επιλογής χαρακτηριστικών με την εξαγωγή χαρακτηριστικών που βασίζεται σε νευρώνες μπορεί να οδηγήσει σε ισχυρά μοντέλα.
5. **Ενσωματώσεις και λανθάνουσες αναπαραστάσεις:**
   - Τα νευρωνικά δίκτυα, ειδικά οι αυτοκωδικοποιητές, μπορούν να χρησιμοποιηθούν για μείωση διαστάσεων μαθαίνοντας συμπαγείς λανθάνουσες αναπαραστάσεις των δεδομένων.
   - Αυτές οι ενσωματώσεις μπορούν να καταγράψουν μη γραμμικές σχέσεις στα δεδομένα.

Η ενσωμάτωση της μείωσης διαστάσεων και της επιλογής χαρακτηριστικών με τα νευρωνικά δίκτυα μπορεί να οδηγήσει σε αποτελεσματικά, ακριβή και στιβαρά μοντέλα. Η συνέργεια μεταξύ των παραδοσιακών τεχνικών απλοποίησης δεδομένων και των νευρωνικών αρχιτεκτονικών μπορεί να ξεκλειδώσει ισχυρές δυνατότητες μοντελοποίησης.

Περίληψη

1. **Κατανόηση της πολυπλοκότητας δεδομένων:**
   - Υπογράμμισε τη σημασία της κατανόησης της πολυπλοκότητας και των διαστάσεων των δεδομένων μας.
   - Αναγνώρισε ότι δεν είναι όλα τα χαρακτηριστικά ή οι διαστάσεις εξίσου ενημερωτικά.
2. **Μείωση διαστάσεων και επιλογή χαρακτηριστικών:**
   - Τόνισε τη σημασία της απλοποίησης των δεδομένων μέσω της μείωσης των διαστάσεων και της επιλογής σχετικών χαρακτηριστικών.
   - Συζήτησε διάφορες τεχνικές και τις εφαρμογές τους.
3. **Μετρήσεις και αξιολόγηση μοντέλου:**
   - Επανεξέτασε τη σημασία διαφόρων μετρήσεων όπως η ακρίβεια, η ευαισθησία, η ειδικότητα και η βαθμολογία F1.
   - Τόνισε την ανάγκη για συνολική αξιολόγηση του μοντέλου.
4. **Ενοποίηση νευρωνικών δικτύων:**
   - Συζήτησε τη συνέργεια μεταξύ τεχνικών απλοποίησης δεδομένων και αρχιτεκτονικών νευρωνικών δικτύων.
   - Διερεύνησε τις δυνατότητες συνδυασμού παραδοσιακών μεθόδων με βαθιά μάθηση για ισχυρή μοντελοποίηση.
5. **Ο δρόμος μπροστά:**
   - Ενθάρρυνε τη συνεχή εξερεύνηση και προσαρμογή τεχνικών που βασίζονται σε συγκεκριμένα δεδομένα και απαιτήσεις προβλημάτων.
   - Τόνισε τη δυναμική φύση του τομέα της τεχνητής νοημοσύνης και της μηχανικής μάθησης, προτρέποντας την ανάγκη για δια βίου μάθηση.

Το ταξίδι μέσω της μείωσης διαστάσεων, της επιλογής χαρακτηριστικών, των μετρήσεων και της ενσωμάτωσης νευρωνικών δικτύων μάς εξόπλισε με μια ολοκληρωμένη εργαλειοθήκη για τεχνητή νοημοσύνη και μηχανική μάθηση. Καθώς προχωράμε, είναι απαραίτητο να προσαρμόζουμε, να πειραματιζόμαστε και να βελτιώνουμε συνεχώς τις προσεγγίσεις μας για να παραμείνουμε στην πρώτη γραμμή αυτού του συνεχώς εξελισσόμενου τομέα.

Συζήτηση

Προτεινόμενα βιβλία:

- «Αναγνώριση προτύπων και μηχανική μάθηση» του Christopher M. Bishop
- "The Elements of Statistical Learning" των Trevor Hastie, Robert Tibshirani και Jerome Friedman
- «Deep Learning» των Ian Goodfellow, Yoshua Bengio και Aaron Courville

Διαδικτυακοί πόροι:

- Coursera: Machine Learning από τον Andrew Ng
- edX: Αρχές Μηχανικής Μάθησης
- Fast.ai: Πρακτική βαθιά μάθηση για κωδικοποιητές

Ερευνητικές εργασίες:

- "Μερικά χρήσιμα πράγματα που πρέπει να γνωρίζετε για τη μηχανική μάθηση" του Pedro Domingos
- "Κατανοώντας τη Μηχανική Μάθηση: Από τη Θεωρία στους Αλγόριθμους" των Shai Shalev-Shwartz και Shai Ben-David

Το ταξίδι κατανόησης και εξοικείωσης με την τεχνητή νοημοσύνη και τη μηχανική μάθηση είναι συνεχές. Υπάρχουν πάντα περισσότερα για να μάθετε, να εξερευνήσετε και να συζητήσετε. Συμμετέχετε σε συζητήσεις, βουτήξτε σε περαιτέρω αναγνώσεις και μείνετε ενημερωμένοι για να παραμείνετε στην αιχμή αυτού του συναρπαστικού τομέα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Εργαλεία Βιοπληροφορικής

Θέμα 4. Τεχνητά νευρωνικά δίκτυα

Μάθημα 2. Perceptron

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

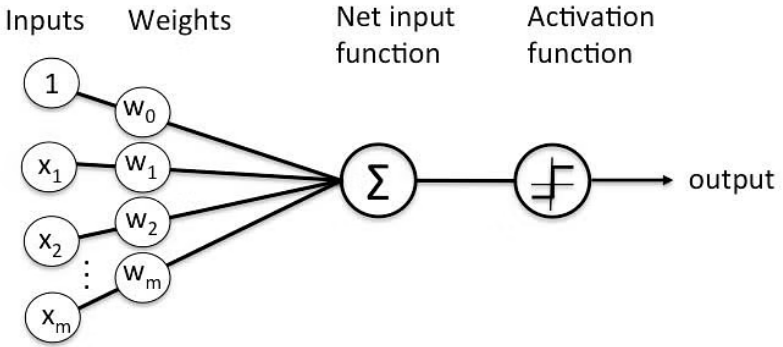
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή στο Perceptron

Καλώς ήρθατε στον κόσμο των Perceptrons!

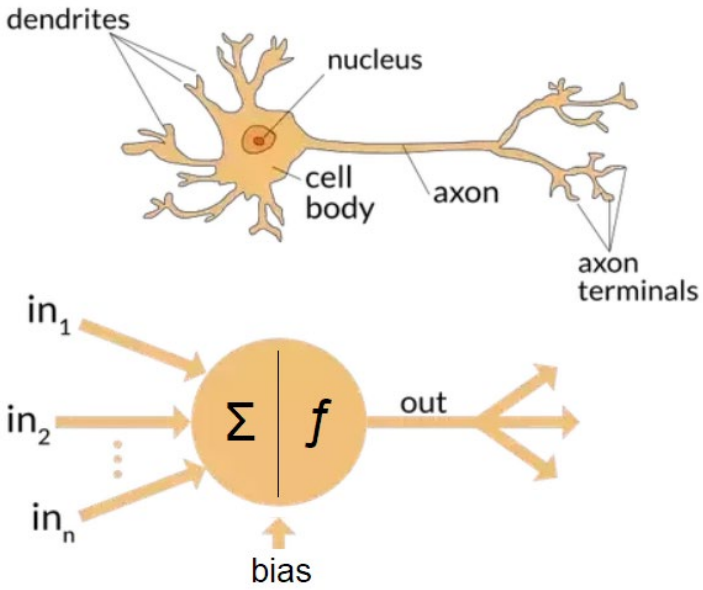
- **Ορισμός:** Το perceptron είναι μια από τις απλούστερες αρχιτεκτονικές τεχνητών νευρωνικών δικτύων, που χρησιμεύει ως θεμελιώδες δομικό στοιχείο για πιο πολύπλοκα νευρωνικά δίκτυα.
- **Ιστορικό πλαίσιο:** Εισήχθη στα τέλη της δεκαετίας του 1950 από τον Frank Rosenblatt, το perceptron εμπνεύστηκε από την επεξεργασία πληροφοριών ενός και μόνο βιολογικού νευρώνα.
- **Σημασία:**
  - Λειτουργεί ως δυαδικός ταξινομητής, λαμβάνοντας αποφάσεις με βάση γραμμικούς προγνωστικούς παράγοντες.
  - Θέτει τα θεμέλια για πολυεπίπεδα νευρωνικά δίκτυα και βαθιά μάθηση.
  - Παρέχει πληροφορίες για τις βασικές αρχές της μηχανικής μάθησης, όπως η εκπαίδευση, τα βάρη και η προκατάληψη.



Το perceptron είναι μια θεμελιώδης έννοια στα νευρωνικά δίκτυα και τη μηχανική μάθηση. Η απλότητά του προσφέρει σαφήνεια, καθιστώντας το ιδανικό σημείο εκκίνησης για την κατανόηση πιο περίπλοκων νευρωνικών αρχιτεκτονικών.

Έμπνευση - Βιολογικός Νευρώνας

Ο σχεδιασμός της φύσης του βιολογικού νευρώνα, με την περίπλοκη δομή του και την εξελιγμένη ικανότητα επεξεργασίας πληροφοριών, χρησιμεύει ως η θεμελιώδης έμπνευση για το perceptron στα τεχνητά νευρωνικά δίκτυα.



Co-funded by the Erasmus+ Programme of the European Union

Βιολογικός Νευρώνας

- **Τι είναι ο Νευρώνας;**
  - Θεμελιώδης μονάδα του εγκεφάλου και του νευρικού συστήματος.
  - Είναι υπεύθυνος για τη λήψη, την επεξεργασία και τη μετάδοση πληροφοριών μέσω ηλεκτρικών και χημικών σημάτων.
- **Συστατικά ενός νευρώνα:**
  - **Δενδρίτες:** Δομές που μοιάζουν με κλάδους που λαμβάνουν σήματα από άλλους νευρώνες.
  - **Κυτταρικό σώμα (Soma):** Περιέχει τον πυρήνα και επεξεργάζεται τα λαμβανόμενα σήματα.
  - **Άξονας:** Μια μακρά προβολή που μεταδίδει σήματα σε άλλους νευρώνες ή μύες.
  - **Συνάψεις (άκρα άξονα):** Διασταυρώσεις όπου ο νευρώνας συνδέεται με άλλους νευρώνες ή κύτταρα, διευκολύνοντας τη μετάδοση σήματος.

- **Μηχανισμός Εργασίας Νευρώνα:**
  - Λαμβάνει σήματα μέσω των δενδριτών.
  - Επεξεργάζεται τα σήματα στο κυτταρικό σώμα.
  - Εάν το επεξεργασμένο σήμα υπερβεί ένα ορισμένο όριο, ο νευρώνας πυροδοτείται, στέλνοντας το σήμα κάτω από τον άξονα σε άλλους νευρώνες ή κύτταρα.
- **Σχέση με το Perceptron:**
  - Το perceptron μιμείται τη λειτουργία ενός βιολογικού νευρώνα.
  - Οι είσοδοι σε ένα perceptron είναι ανάλογες με τους δενδρίτες, τα βάρη μοιάζουν με συναπτικές δυνάμεις και η συνάρτηση ενεργοποίησης εμπνέεται από τον μηχανισμό πυροδότησης του νευρώνα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Γεφύρωση Βιολογίας και Υπολογισμού

Η έμπνευση:

- Ο σχεδιασμός του perceptron είναι άμεσα εμπνευσμένος από τη δομή και τη λειτουργία του βιολογικού νευρώνα.
- Και τα δύο συστήματα επεξεργάζονται εισόδους, ζυγίζουν τη σημασία τους και παράγουν μια έξοδο με βάση ένα όριο.

Βασικές αναλογίες:

- **Δενδρίτες ↔ Είσοδοι:** Ακριβώς όπως οι δενδρίτες λαμβάνουν σήματα σε έναν νευρώνα, τα perceptron λαμβάνουν πολλαπλές τιμές εισόδου.
- **Συναπτικές δυνάμεις ↔ Βάρη:** Στους νευρώνες, οι συνάψεις έχουν ποικίλες δυνάμεις. Ομοίως, τα perceptron αποδίδουν διαφορετικά βάρη στις εισόδους.
- **Πυροδότηση νευρώνων ↔ Λειτουργία ενεργοποίησης:** Ένας νευρώνας ενεργοποιείται εάν το σήμα εισόδου ξεπεράσει ένα όριο. Σε ένα perceptron, η συνάρτηση ενεργοποίησης καθορίζει εάν η συνδυασμένη είσοδος υπερβαίνει μια ορισμένη τιμή για να παράγει μια έξοδο.

Διαφορές:

- Ενώ οι βιολογικοί νευρώνες χρησιμοποιούν ηλεκτροχημικά σήματα, τα perceptron χρησιμοποιούν μαθηματικές συναρτήσεις.
- Ο σχεδιασμός του perceptron είναι μια απλοποιημένη αφαίρεση του βιολογικού νευρώνα, εστιάζοντας στις βασικές αρχές της επεξεργασίας εισόδου και της παραγωγής εξόδου.

Εξέλιξη της έννοιας:

- Το perceptron έθεσε τις βάσεις για πιο προηγμένες αρχιτεκτονικές νευρωνικών δικτύων.
- Τα σύγχρονα νευρωνικά δίκτυα αποτελούνται από πολλαπλά στρώματα διασυνδεδεμένων perceptrons (ή νευρώνων), επιτρέποντας την πολύπλοκη επεξεργασία δεδομένων και την αναγνώριση προτύπων.

Το perceptron, ως θεμελιώδης έννοια στα τεχνητά νευρωνικά δίκτυα, περικλείει όμορφα την ουσία του βιολογικού νευρώνα. Αυτή η συνέργεια μεταξύ βιολογίας και υπολογισμού έχει ανοίξει το δρόμο για τις εξελίξεις που βλέπουμε στη σύγχρονη τεχνητή νοημοσύνη και τη βαθιά μάθηση.

Ανατομία ενός Perceptron

Βασικά συστατικά:

- **Είσοδοι (x1, x2, ... xn):** Τιμές που λαμβάνονται από το perceptron, ανάλογες με τα σήματα που λαμβάνονται από τους δενδρίτες ενός νευρώνα.
- **Βάρη (w1, w2, ... wn):** Εκχωρούνται σε κάθε είσοδο, προσδιορίζοντας τη σημασία ή την επιρροή αυτής της εισόδου. Συγκρίσιμο με τις συναπτικές δυνάμεις σε έναν νευρώνα.
- **Προκατάληψη (b):** Μια σταθερή τιμή που προστίθεται στο σταθμισμένο άθροισμα των εισόδων, που επιτρέπει στο perceptron να προσαρμόζει την έξοδό του ανεξάρτητα από τις εισόδους του.
- **Συνάρτηση ενεργοποίησης (f):** Μια μαθηματική συνάρτηση που επεξεργάζεται το σταθμισμένο άθροισμα των εισόδων και της πόλωσης για να παράγει την έξοδο του perceptron. Καθορίζει αν το perceptron «πυροδοτεί» ή όχι.

Η δομή του perceptron, αν και απλή, ενσωματώνει τις βασικές αρχές της νευρωνικής επεξεργασίας. Κάθε στοιχείο παίζει καθοριστικό ρόλο στον προσδιορισμό της εξόδου του perceptron, καθιστώντας το ένα θεμελιώδες δομικό στοιχείο για πιο σύνθετες αρχιτεκτονικές νευρωνικών δικτύων.

Co-funded by the Erasmus+ Programme of the European Union

Το Perceptron σε δράση

- **Μηχανισμός Εργασίας:**
  a. Το perceptron υπολογίζει το σταθμισμένο άθροισμα των εισόδων του και προσθέτει την προκατάληψη.
  b. Αυτό το άθροισμα περνά στη συνέχεια μέσω της συνάρτησης ενεργοποίησης για να παραχθεί η έξοδος.
- **Μαθηματική παράσταση:**
  a. Έξοδος = f(Σ (βάρος * είσοδος) + προκατάληψη)
  b. Για παράδειγμα, με μια συνάρτηση βήματος ως συνάρτηση ενεργοποίησης, το perceptron βγάζει 1 εάν το άθροισμα υπερβαίνει ένα όριο, και 0 διαφορετικά.

Η διαδικασία λήψης αποφάσεων του perceptron είναι μια αρμονική αλληλεπίδραση ζύγισης εισόδων, άθροισης τιμών και εφαρμογής μιας συνάρτησης ενεργοποίησης. Αυτή η συστηματική προσέγγιση του επιτρέπει να ταξινομεί τις εισροές με βάση τα μαθημένα πρότυπα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Λειτουργία ενεργοποίησης

Τι είναι η λειτουργία ενεργοποίησης;

- Μια μαθηματική εξίσωση που καθορίζει την έξοδο ενός perceptron ή ενός νευρώνα.
- Εισάγει τη μη γραμμικότητα στην έξοδο ενός νευρώνα, επιτρέποντας στα νευρωνικά δίκτυα να μαθαίνουν από τα σφάλματα και να κάνουν προσαρμογές, κάτι που είναι απαραίτητο για την εκμάθηση πολύπλοκων προτύπων.

Ρόλος σε ένα Perceptron:

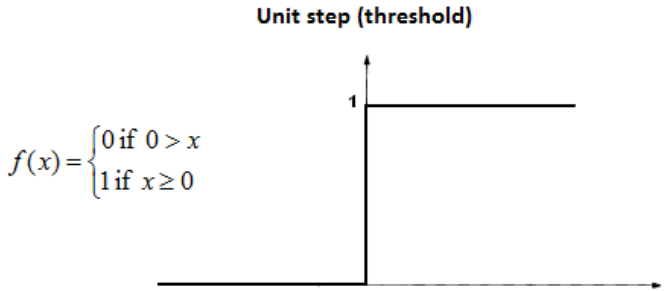
- Αποφασίζει εάν ένας νευρώνας πρέπει να ενεργοποιηθεί ή όχι, υπολογίζοντας το σταθμισμένο άθροισμα και προσθέτοντας περαιτέρω προκατάληψη.
- Βοηθά στην κανονικοποίηση της εξόδου, διασφαλίζοντας ότι είναι σε μορφή κατάλληλη για περαιτέρω επίπεδα ή ως τελική έξοδο.

Γιατί είναι σημαντικό?

- Χωρίς μια συνάρτηση ενεργοποίησης, ένα perceptron θα ήταν απλώς ένα μοντέλο γραμμικής παλινδρόμησης, περιορίζοντας την ικανότητά του να μαθαίνει και να εκτελεί πιο σύνθετες εργασίες.
- Επιτρέπει στα perceptron να συλλαμβάνουν μη γραμμικά μοτίβα στα δεδομένα.

Λειτουργία Βήματος:

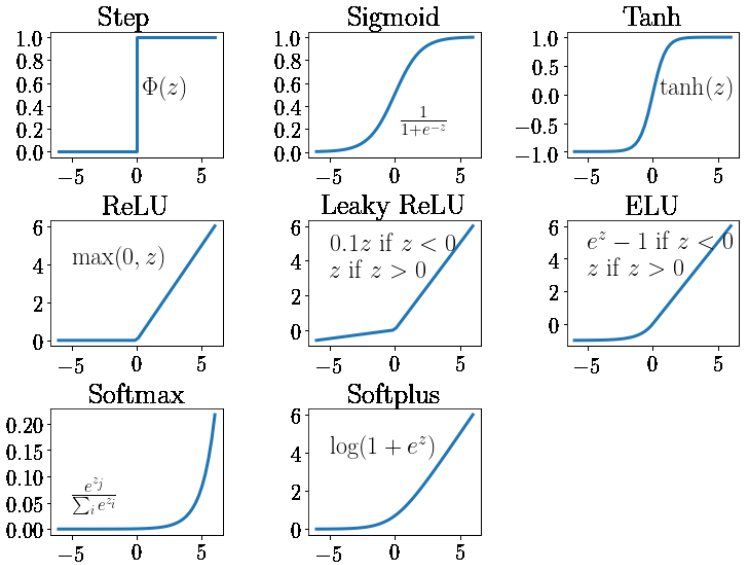
- Η απλούστερη μορφή λειτουργίας ενεργοποίησης, που παράγει μια δυαδική έξοδο με βάση ένα όριο.
- Εάν η τιμή εισόδου ξεπεράσει το όριο, η συνάρτηση εξάγει ένα 1 (ή «πυροδοτεί»)



Οι συναρτήσεις ενεργοποίησης παίζουν καθοριστικό ρόλο στη διαμόρφωση της εξόδου ενός perceptron. Εισάγουν την απαραίτητη **μη γραμμικότητα** , επιτρέποντας στο perceptron να μάθει και να προσαρμοστεί. Η συνάρτηση βήματος, ως βασικό παράδειγμα, προσφέρει μια ματιά στον κόσμο των λειτουργιών ενεργοποίησης.

Κοινές Λειτουργίες ενεργοποίησης

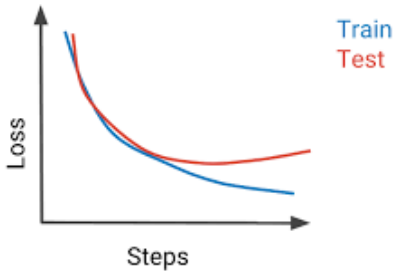
- **Βηματική:**
  - Παράγει μια δυαδική έξοδο με βάση ένα όριο.
  - Εάν η τιμή εισόδου ξεπεράσει το όριο, η συνάρτηση βγάζει 1. Διαφορετικά, βγάζει 0.
- **Σιγμοειδής συνάρτηση:**
  - Εξάγει τιμές μεταξύ 0 και 1.
  - Ομαλή κλίση, αποτρέποντας τα ξαφνικά άλματα στις τιμές εξόδου.
- **ReLU (Διορθωμένη Γραμμική Μονάδα):**
  - Επιτρέπει στις θετικές τιμές να περάσουν αμετάβλητες, ενώ οι αρνητικές τιμές ορίζονται στο μηδέν.
  - Χρησιμοποιείται συνήθως σε κρυφά επίπεδα νευρωνικών δικτύων.
- **Tanh (Υπερβολική Εφαπτομένη):**
  - Εξάγει τιμές μεταξύ -1 και 1.
  - Παρόμοιο με το σιγμοειδές αλλά μπορεί να δώσει αρνητικές τιμές.
- **Softmax:**
  - Χρησιμοποιείται στο επίπεδο εξόδου των προβλημάτων ταξινόμησης.
  - Μετατρέπει ένα διάνυσμα αριθμών σε διάνυσμα πιθανοτήτων.



Διαφορετικές λειτουργίες ενεργοποίησης εξυπηρετούν διαφορετικούς σκοπούς και επιλέγονται με βάση τις συγκεκριμένες απαιτήσεις ενός μοντέλου νευρωνικού δικτύου. Τα διαφορετικά χαρακτηριστικά τους επιτρέπουν στα νευρωνικά δίκτυα να καταγράφουν ένα ευρύ φάσμα μοτίβων και πολυπλοκοτήτων στα δεδομένα.

Ο ρόλος των συναρτήσεων σφάλματος (απώλειας).

- **Τι είναι η συνάρτηση σφάλματος (απώλειας);**
  - Μια μαθηματική εξίσωση που ποσοτικοποιεί πόσο καλά ευθυγραμμίζεται η πρόβλεψη του μοντέλου με τα πραγματικά δεδομένα.
  - Μετρά τη διαφορά μεταξύ της προβλεπόμενης παραγωγής και της πραγματικής παραγωγής.
- **Σκοπός στα νευρωνικά δίκτυα:**
  - Παρέχει μια μέτρηση για βελτιστοποίηση κατά τη διάρκεια της προπόνησης.
  - Ο στόχος της εκπαίδευσης είναι να ελαχιστοποιηθεί αυτό το σφάλμα, οδηγώντας καλύτερες προβλέψεις μοντέλων.
- **Γιατί είναι σημαντικό?**
  - Καθοδηγεί την προσαρμογή των παραμέτρων του μοντέλου (όπως βάρη και προκαταλήψεις).
  - Μια καλά επιλεγμένη λειτουργία απώλειας μπορεί να επιταχύνει τη διαδικασία εκπαίδευσης και να βελτιώσει την ακρίβεια του μοντέλου.
- **Κοινοί τύποι:**
  - **Μέσο τετράγωνο σφάλμα (MSE):** Χρησιμοποιείται για προβλήματα παλινδρόμησης. Υπολογίζει το τετράγωνο της διαφοράς μεταξύ προβλεπόμενων και πραγματικών τιμών.
  - **Απώλεια Cross-Entropy:** Χρησιμοποιείται για προβλήματα ταξινόμησης. Μετρά τη διαφορά μεταξύ δύο κατανομών πιθανοτήτων.



Οι συναρτήσεις σφάλματος (απώλειας) διαδραματίζουν κρίσιμο ρόλο στην εκπαίδευση των νευρωνικών δικτύων. Παρέχουν ένα μέτρο για το πόσο μακριά απέχουν οι προβλέψεις από τις πραγματικές τιμές, καθοδηγώντας τις προσαρμογές που έγιναν στις παραμέτρους του μοντέλου για βελτιωμένη απόδοση.

Κοινές Λειτουργίες ενεργοποίησης

- **Μέσο τετράγωνο σφάλμα (MSE):**
  - **Χρήση:** Χρησιμοποιείται συνήθως για προβλήματα παλινδρόμησης.
  - **Χαρακτηριστικά:** Τιμωρεί περισσότερο τα μεγάλα λάθη λόγω του τετραγωνισμού των διαφορών.
- **Απώλεια Cross-Entropy (Απώλεια καταγραφής):**
  - **Χρήση:** Ιδανικό για προβλήματα ταξινόμησης δυαδικών και πολλαπλών τάξεων.
  - **Χαρακτηριστικά:** Μετρά την απόδοση ενός μοντέλου ταξινόμησης του οποίου η έξοδος είναι μια τιμή πιθανότητας μεταξύ 0 και 1.
- **Απώλεια άρθρωσης (Μέγιστη απώλεια περιθωρίου):**
  - **Χρήση:** Χρησιμοποιείται για την ταξινόμηση "μέγιστου περιθωρίου", κυρίως για μηχανές διανυσμάτων υποστήριξης.
  - **Χαρακτηριστικά:** Προορίζεται για χρήση με δυαδική ταξινόμηση όπου οι τιμές στόχου είναι στο σύνολο {-1, 1}.
- **Απώλεια Huber:**
  - **Χρήση:** Χρήσιμο όταν υπάρχουν ακραίες τιμές, αλλά είναι επιθυμητό ένα τετράγωνο σφάλμα.
  - **Χαρακτηριστικά:** Συνδυάζει ιδιότητες του MSE και σημαίνει απόλυτο σφάλμα.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

$$L(y,t) = -\sum_i t_i \ln y_i$$

$$\ell(y) = \max(0, 1 - t \cdot y)$$

$$L_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{for } |a| \le \delta, \\ \delta \cdot \left(|a| - \frac{1}{2}\delta\right), & \text{otherwise.} \end{cases}$$

Διαφορετικές λειτουργίες ενεργοποίησης εξυπηρετούν διαφορετικούς σκοπούς και επιλέγονται με βάση τις συγκεκριμένες απαιτήσεις ενός μοντέλου νευρωνικού δικτύου. Τα διαφορετικά χαρακτηριστικά τους επιτρέπουν στα νευρωνικά δίκτυα να καταγράφουν ένα ευρύ φάσμα μοτίβων και πολυπλοκοτήτων στα δεδομένα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ανοίγοντας το μονοπάτι προς τη μάθηση: Ο κανόνας του Δέλτα

Τι είναι ο κανόνας Δέλτα;

- Γνωστός και ως κανόνας εκμάθησης Widrow-Hoff: Ένας κανόνας εκμάθησης βαθμίδωσης που χρησιμοποιείται για τη ρύθμιση των βαρών ενός perceptron.

Σκοπός στα νευρωνικά δίκτυα:

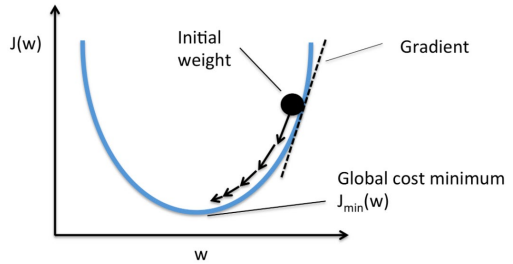
- Παρέχει έναν μηχανισμό για την ελαχιστοποίηση του σφάλματος του perceptron ρυθμίζοντας τα βάρη του.
- Βοηθά το μοντέλο να συγκλίνει στη βέλτιστη λύση μειώνοντας επαναληπτικά το σφάλμα.

Πώς λειτουργεί?

- Τα βάρη ρυθμίζονται προς την κατεύθυνση που μειώνει το σφάλμα.
- Η μεταβολή στα βάρη είναι ανάλογη με την αρνητική κλίση του σφάλματος ως προς τα βάρη.

$$\Delta w = \alpha(t-o)x$$

- Οπου:
  - **Δw** είναι η αλλαγή βάρους.
  - **α** είναι ο ρυθμός μάθησης.
  - **t** είναι η έξοδος στόχος.
  - **o** είναι η πραγματική έξοδος.



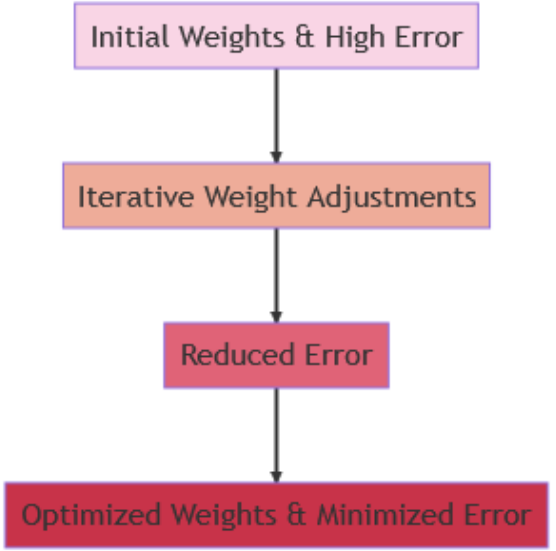
Ο κανόνας Delta είναι ένας θεμελιώδης αλγόριθμος εκμάθησης για τα perceptron, ο οποίος καθοδηγεί τις προσαρμογές που γίνονται στα βάρη για την ελαχιστοποίηση του σφάλματος. Ακολουθώντας τη διαβάθμιση του σφάλματος, διασφαλίζει ότι το perceptron κινείται προς καλύτερη απόδοση με κάθε επανάληψη.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Δέλτα κανόνας σε δράση

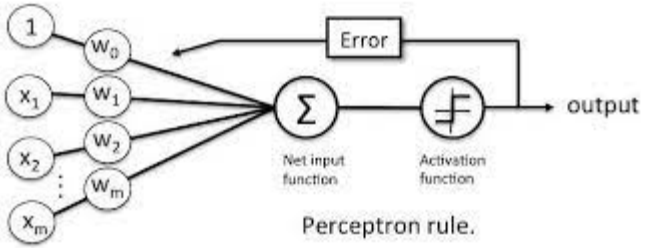
- **Επαναληπτική μάθηση:**
  - Ο κανόνας Delta εφαρμόζεται επαναληπτικά, προσαρμόζοντας τα βάρη μετά από κάθε παράδειγμα προπόνησης.
  - Με την πάροδο του χρόνου, αυτές οι προσαρμογές οδηγούν σε ένα perceptron που κάνει πιο ακριβείς προβλέψεις.
- **Ρύθμιση βάρους:**
  - Εάν το perceptron παράγει ένα σφάλμα (διαφορά μεταξύ προβλεπόμενης και πραγματικής εξόδου), τα βάρη προσαρμόζονται χρησιμοποιώντας τον κανόνα Delta.
  - Το μέγεθος και η κατεύθυνση της ρύθμισης εξαρτώνται από το σφάλμα και τις τιμές εισόδου.
- **Ποσοστό μάθησης (α):**
  - Μια υπερπαράμετρος που καθορίζει το μέγεθος του βήματος κατά τη ρύθμιση του βάρους.
  - Ένας υψηλός ρυθμός μάθησης μπορεί να υπερβεί τη βέλτιστη λύση, ενώ ένας χαμηλός ρυθμός μάθησης μπορεί να συγκλίνει αργά.
- **Σύγκλιση:**
  - Με τον σωστό ρυθμό εκμάθησης και επαρκείς επαναλήψεις, ο κανόνας Delta διασφαλίζει ότι το perceptron συγκλίνει σε μια λύση όπου το σφάλμα ελαχιστοποιείται.

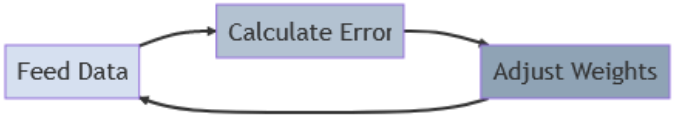


Η πρακτική εφαρμογή του κανόνα Delta δείχνει τη δύναμή του στη βελτίωση της απόδοσης ενός perceptron. Μέσω επαναληπτικών προσαρμογών, κατευθύνει το perceptron προς μια βελτιστοποιημένη κατάσταση, διασφαλίζοντας ότι μαθαίνει και προσαρμόζεται αποτελεσματικά στα δεδομένα εκπαίδευσης.

Εκπαίδευση ενός Perceptron

- **Δεδομένα τροφοδοσίας:**
  - Το perceptron παρουσιάζεται με παραδείγματα εκπαίδευσης, τα οποία αποτελούνται από εισόδους σε συνδυασμό με σωστές εξόδους.
  - Κάθε είσοδος πολλαπλασιάζεται με ένα βάρος και τα αποτελέσματα συνοψίζονται.
- **Σφάλμα υπολογισμού:**
  - Μόλις το perceptron παράγει μια έξοδο, συγκρίνεται με τη σωστή έξοδο για να προσδιοριστεί το σφάλμα.
  - Το σφάλμα είναι απλώς η διαφορά μεταξύ της προβλεπόμενης παραγωγής και της πραγματικής παραγωγής.
- **Ρύθμιση βαρών:**
  - Με βάση το υπολογισμένο σφάλμα, ρυθμίζονται τα βάρη του perceptron.
  - Ο κανόνας Delta εφαρμόζεται για να καθοδηγήσει αυτές τις προσαρμογές, διασφαλίζοντας ότι το perceptron κινείται προς καλύτερη απόδοση.
  - Η διαδικασία επαναλαμβάνεται για πολλές επαναλήψεις ή μέχρι να ελαχιστοποιηθεί το σφάλμα σε ένα αποδεκτό επίπεδο.





Η εκπαίδευση ενός perceptron είναι μια επαναληπτική διαδικασία που περιλαμβάνει την παρουσίαση δεδομένων, την αξιολόγηση της απόδοσης και τη βελτίωση του μοντέλου. Μέσω συνεχών προσαρμογών, το perceptron μαθαίνει να κάνει ακριβείς προβλέψεις, επιδεικνύοντας τη δύναμη της μηχανικής μάθησης.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Σύγκλιση Perceptron

Καθορισμένη σύγκλιση:

- Το σημείο στο οποίο σταθεροποιούνται τα βάρη του perceptron και το σφάλμα δεν μειώνεται πλέον σημαντικά.
- Υποδεικνύει ότι το perceptron έχει μάθει τα μοτίβα στα δεδομένα εκπαίδευσης στο μέγιστο των δυνατοτήτων του.

Παράγοντες που επηρεάζουν τη σύγκλιση:

- **Ρυθμός εκμάθησης:** Καθορίζει το μέγεθος του βήματος κατά τις προσαρμογές βάρους. Ένας κατάλληλος ρυθμός μάθησης εξασφαλίζει σταθερή σύγκλιση.
- **Δεδομένα εκπαίδευσης:** Η ποιότητα και η ποσότητα των δεδομένων μπορεί να επηρεάσει το πόσο γρήγορα και αποτελεσματικά συγκλίνει ένα perceptron.
- **Αρχικά βάρη:** Τα αρχικά βάρη μπορούν να επηρεάσουν τη διαδρομή και την ταχύτητα της σύγκλισης.

Εγγυημένη σύγκλιση:

- Για **γραμμικά διαχωρίσιμα** δεδομένα, το perceptron είναι εγγυημένο ότι συγκλίνει.
- Ο αριθμός των απαιτούμενων επαναλήψεων μπορεί να ποικίλλει ανάλογα με τους παράγοντες που αναφέρονται παραπάνω.

Προκλήσεις:

- Για μη γραμμικά διαχωρίσιμα δεδομένα, ένα perceptron **δεν μπορεί ποτέ** να συγκλίνει χρησιμοποιώντας τυπικούς αλγόριθμους εκπαίδευσης.
- Αυτός ο περιορισμός οδήγησε στην ανάπτυξη πιο προηγμένων αρχιτεκτονικών νευρωνικών δικτύων.

Η σύγκλιση είναι μια κρίσιμη έννοια στην εκπαίδευση του perceptron. Σημαίνει το σημείο όπου το perceptron έχει μάθει αποτελεσματικά από τα δεδομένα εκπαίδευσης. Η κατανόηση των παραγόντων που επηρεάζουν τη σύγκλιση μπορεί να βοηθήσει στο σχεδιασμό καλύτερων αλγορίθμων εκπαίδευσης και στη διασφάλιση αποτελεσματικής μάθησης.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Εφαρμογές Perceptrons

Δυαδική ταξινόμηση:

- Τα perceptron είναι εγγενώς δυαδικοί ταξινομητές, καθιστώντας τα κατάλληλα για εργασίες όπως η ανίχνευση ανεπιθύμητων μηνυμάτων, η ανάλυση συναισθημάτων και πολλά άλλα.

Αναγνώριση εικόνας:

- Οι πρώτες εφαρμογές των perceptrons ήταν στην αναγνώριση απλών μοτίβων σε εικόνες, όπως η αναγνώριση σχημάτων ή χαρακτήρων.

Γραμμικά όρια απόφασης:

- Για σύνολα δεδομένων που έχουν γραμμικό όριο απόφασης, τα perceptron μπορούν να είναι ένα αποτελεσματικό εργαλείο. Για παράδειγμα, ταξινόμηση σημείων δεδομένων με βάση δύο γραμμικά διαχωρισμένα χαρακτηριστικά.

Γραμμικής παλινδρόμησης:

- Με τις κατάλληλες συναρτήσεις ενεργοποίησης, τα perceptrons μπορούν να χρησιμοποιηθούν για την πρόβλεψη συνεχών τιμών, καθιστώντας τα κατάλληλα για απλές εργασίες γραμμικής παλινδρόμησης.

Συστήματα Ελέγχου:

- Τα Perceptrons μπορούν να χρησιμοποιηθούν σε προσαρμοστικά συστήματα ελέγχου όπου το σύστημα μαθαίνει να ελέγχει μια λειτουργία με βάση την είσοδο που λαμβάνει.

Θεμέλια νευρωνικών δικτύων:

- Εκτός από τους περιορισμούς τους, έθεσαν τις βάσεις για πολυεπίπεδα νευρωνικά δίκτυα και μοντέλα βαθιάς μάθησης που μπορούν να χειριστούν πιο περίπλοκες εργασίες.

Ενώ τα perceptron έχουν τους περιορισμούς τους, έχουν βρει εφαρμογές σε διάφορους τομείς λόγω της απλότητας και της αποτελεσματικότητάς τους. Η κατανόηση αυτών των εφαρμογών παρέχει πληροφορίες για την ευελιξία αυτού του θεμελιώδους μοντέλου μηχανικής μάθησης.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Εργαλεία Βιοπληροφορικής

Θέμα 4. Τεχνητά νευρωνικά δίκτυα

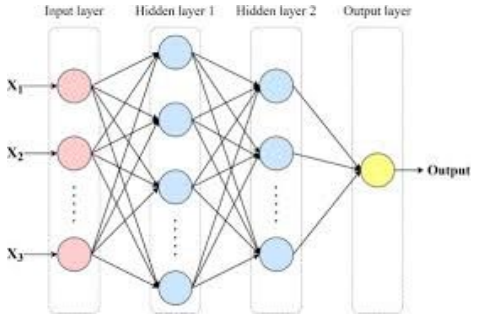
Μάθημα 3. Πολυστρωματικό Perceptron - Backpropagation

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Εισαγωγή στο MLP

Παρουσιάζοντας το Perceptron πολλαπλών στρωμάτων (MLP)

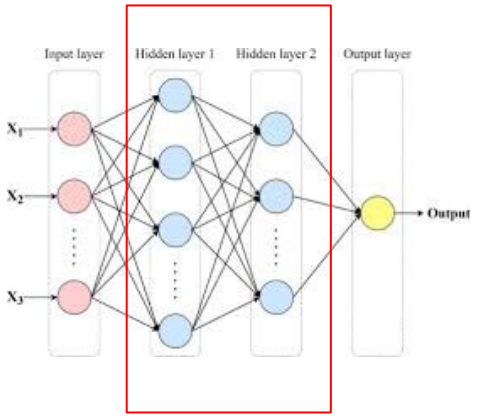
- **Τι είναι το MLP;**
  - Ένας τύπος τεχνητού νευρωνικού δικτύου.
  - Περιλαμβάνει πολλαπλά στρώματα κόμβων ή νευρώνων.
  - Μπορεί να μοντελοποιήσει πολύπλοκες, μη γραμμικές σχέσεις.
- **Γιατί είναι σημαντικό το MLP;**
  - Λειτουργεί ως βάση για βαθιά μάθηση.
  - Ικανός να χειριστεί ένα ευρύ φάσμα εργασιών, από ταξινόμηση έως παλινδρόμηση.
  - Προσφέρει ευελιξία στο σχεδιασμό με ποικίλα βάθη και πλάτη.
- **Βασική Δομή:**
  - **Επίπεδο εισόδου:** Το αρχικό επίπεδο που λαμβάνει δεδομένα.
  - **Κρυφά επίπεδα:** Ενδιάμεσα επίπεδα που επεξεργάζονται δεδομένα. Ο αριθμός μπορεί να ποικίλλει.
  - **Επίπεδο εξόδου:** Παράγει την τελική πρόβλεψη ή ταξινόμηση.



Το Multi-Layer Perceptron είναι μια θεμελιώδης ιδέα στα νευρωνικά δίκτυα και τη βαθιά μάθηση. Η πολυεπίπεδη δομή του του επιτρέπει να μαθαίνει και να αναπαριστά περίπλοκα μοτίβα στα δεδομένα, καθιστώντας το ένα ευέλικτο εργαλείο στην εργαλειοθήκη AI.

Κρυφά Επίπεδα

- **Τι είναι τα κρυφά επίπεδα;**
  - Επίπεδα που βρίσκονται μεταξύ των επιπέδων εισόδου και εξόδου.
  - Αποτελείται από πολλαπλούς νευρώνες ή κόμβους.
  - Υπεύθυνος για την εξαγωγή και την επεξεργασία χαρακτηριστικών από δεδομένα εισόδου.
- **Βάθος έναντι πλάτους σε MLP:**
  - **Βάθος:** Αναφέρεται στον αριθμό των κρυφών επιπέδων. Περισσότερα στρώματα μπορούν να αποτυπώσουν πολύπλοκα μοτίβα, αλλά ενδέχεται να υπάρχει κίνδυνος υπερβολικής προσαρμογής.
  - **Width:** Αναφέρεται στον αριθμό των νευρώνων σε κάθε κρυφό στρώμα. Ένα ευρύτερο επίπεδο μπορεί να αντιπροσωπεύει περισσότερα χαρακτηριστικά αλλά μπορεί να αυξήσει το υπολογιστικό κόστος.
- **Σημασία των κρυφών επιπέδων:**
  - Επιτρέπουν στο MLP να μοντελοποιεί μη γραμμικές σχέσεις.
  - Κάθε επίπεδο μπορεί να συλλάβει διαφορετικά επίπεδα αφαίρεσης, από βασικά έως πολύπλοκα χαρακτηριστικά.
  - Λειτουργεί ως χώρος μετασχηματισμού, μετατρέποντας τα δεδομένα εισόδου σε μια μορφή που διευκολύνει το επίπεδο εξόδου να κάνει προβλέψεις.



Τα κρυφά επίπεδα είναι η καρδιά ενός MLP, επιτρέποντάς του να μαθαίνει περίπλοκα μοτίβα και σχέσεις στα δεδομένα. Ο σχεδιασμός τους, ως προς το βάθος και το πλάτος, παίζει καθοριστικό ρόλο στην απόδοση και την αποδοτικότητα του δικτύου.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Λειτουργίες ενεργοποίησης

Τι είναι η λειτουργία ενεργοποίησης;

- Μια μαθηματική συνάρτηση που εφαρμόζεται στην έξοδο ενός νευρώνα.
- Καθορίζει την έξοδο του νευρώνα με βάση την είσοδο του.
- Εισάγει μη γραμμικότητα στο δίκτυο.

Γιατί είναι σημαντικό?

- Χωρίς συναρτήσεις ενεργοποίησης, τα MLP θα ήταν ισοδύναμα με τα μοντέλα γραμμικής παλινδρόμησης, περιορίζοντας την ικανότητά τους να μοντελοποιούν περίπλοκες σχέσεις.
- Επιτρέπουν στους νευρώνες να λαμβάνουν αποφάσεις, επιτρέποντας στο δίκτυο να μαθαίνει από τα λάθη και να κάνει προσαρμογές.

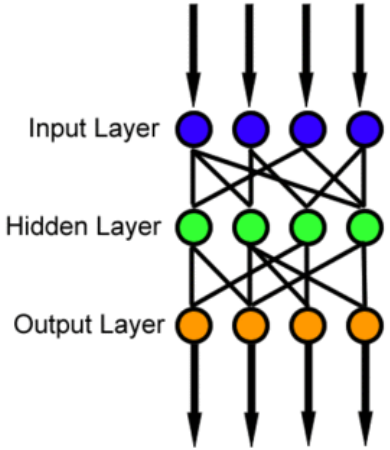
Κοινές λειτουργίες ενεργοποίησης για MLP:

- **Βήμα Συνάρτηση:** Δυαδική έξοδος με βάση ένα όριο.
- **Σιγμοειδές:** Παράγει εξόδους μεταξύ 0 και 1, που χρησιμοποιούνται συχνά στη δυαδική ταξινόμηση.
- **ReLU (Rectified Linear Unit):** Επιτρέπει στις θετικές τιμές να περάσουν αμετάβλητες, ενώ ορίζει τις αρνητικές τιμές στο μηδέν. Χρησιμοποιείται συνήθως σε κρυφά επίπεδα.

Οι συναρτήσεις ενεργοποίησης είναι ζωτικής σημασίας για την εισαγωγή της μη γραμμικότητας στα νευρωνικά δίκτυα, επιτρέποντάς τους να συλλαμβάνουν και να μοντελοποιούν περίπλοκα μοτίβα στα δεδομένα. Η επιλογή της σωστής λειτουργίας ενεργοποίησης μπορεί να επηρεάσει σημαντικά την απόδοση ενός MLP.

Μηχανισμός προώθησης τροφοδοσίας

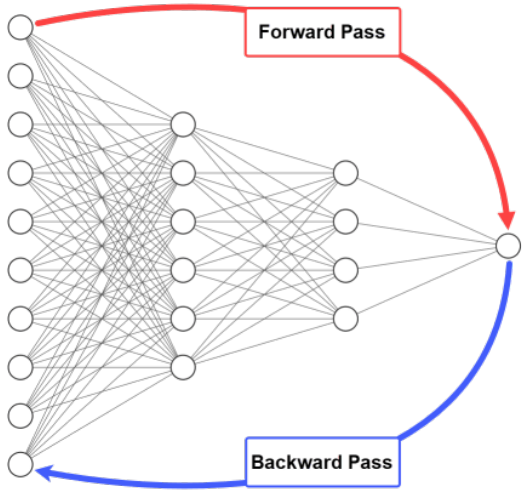
- **Τι είναι το Feed Forward;**
  a. Η αρχική φάση όπου τα δεδομένα εισόδου διαβιβάζονται μέσω του δικτύου.
  b. Τα δεδομένα ρέουν από το επίπεδο εισόδου, μέσω κρυφών επιπέδων, στο επίπεδο εξόδου.
  c. Κάθε νευρώνας επεξεργάζεται τα δεδομένα, εφαρμόζει τη συνάρτηση ενεργοποίησης και τα περνά στο επόμενο επίπεδο.
- **Βήματα στο Feed Forward:**
  a. **Λήψη εισόδου:** Οι νευρώνες στο στρώμα εισόδου λαμβάνουν τα δεδομένα.
  b. **Σταθμισμένο άθροισμα:** Κάθε νευρώνας στο κρυφό στρώμα υπολογίζει ένα σταθμισμένο άθροισμα των εισόδων του.
  c. **Ενεργοποίηση:** Η συνάρτηση ενεργοποίησης εφαρμόζεται στο σταθμισμένο άθροισμα για την παραγωγή της εξόδου του νευρώνα.
- **Ρόλος στην Πρόβλεψη:**
  a. Το τελικό αποτέλεσμα της διαδικασίας προώθησης τροφοδοσίας είναι η πρόβλεψη του δικτύου.
  b. Για εργασίες ταξινόμησης, αυτό θα μπορούσε να είναι μια ετικέτα κλάσης ή κατανομή πιθανότητας σε κλάσεις.
  c. Για εργασίες παλινδρόμησης, θα μπορούσε να είναι μια συνεχής τιμή.



Ο μηχανισμός προώθησης τροφοδοσίας είναι η ουσία του τρόπου με τον οποίο ένας MLP κάνει προβλέψεις. Είναι μια συστηματική διαδικασία όπου τα δεδομένα μετασχηματίζονται και επεξεργάζονται στρώμα-στρώμα, με αποκορύφωμα την τελική πρόβλεψη του δικτύου.

Οπισθοδιάδοση

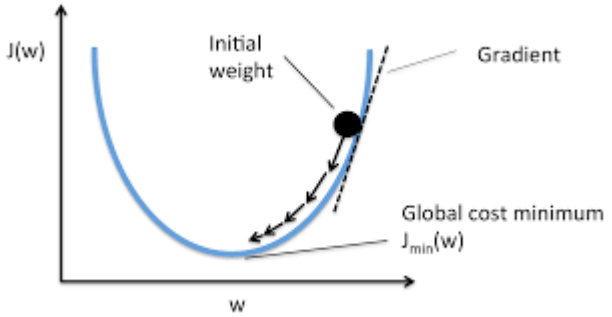
- **Τι είναι η οπισθοδιάδοση;**
  a. Η διαδικασία με την οποία ένας MLP μαθαίνει και προσαρμόζει τα βάρη του.
  b. Με βάση το σφάλμα μεταξύ της προβλεπόμενης εξόδου και του πραγματικού στόχου.
  c. Περιλαμβάνει τον υπολογισμό της κλίσης του σφάλματος σε σχέση με κάθε βάρος.
- **Γιατί είναι κρίσιμο;**
  a. Επιτρέπει στο δίκτυο να ελαχιστοποιεί το σφάλμα στις προβλέψεις του.
  b. Προσαρμόζοντας τα βάρη στη σωστή κατεύθυνση, το δίκτυο βελτιώνει την απόδοσή του με την πάροδο του χρόνου.
  c. Είναι ο πρωταρχικός μηχανισμός που επιτρέπει στα νευρωνικά δίκτυα να μαθαίνουν από δεδομένα.
- **Βασικά βήματα:**
  a. **Υπολογισμός σφαλμάτων:** Προσδιορίστε τη διαφορά μεταξύ της προβλεπόμενης εξόδου και του πραγματικού στόχου.
  b. **Υπολογισμός κλίσης:** Υπολογίστε πόσο συνέβαλε κάθε βάρος στο σφάλμα.
  c. **Προσαρμογή βάρους:** Ενημερώστε τα βάρη στην αντίθετη κατεύθυνση της κλίσης για να μειώσετε το σφάλμα.



Ο μηχανισμός προώθησης τροφοδοσίας είναι η ουσία του τρόπου με τον οποίο ένας MLP κάνει προβλέψεις. Είναι μια συστηματική διαδικασία όπου τα δεδομένα μετασχηματίζονται και επεξεργάζονται στρώμα-στρώμα, με αποκορύφωμα την τελική πρόβλεψη του δικτύου.

Gradient Descent - Η μηχανή βελτιστοποίησης

- **Τι είναι το Gradient Descent;**
  - Ένας αλγόριθμος βελτιστοποίησης που χρησιμοποιείται για την ελαχιστοποίηση του σφάλματος στα νευρωνικά δίκτυα.
  - Ρυθμίζει τα βάρη του δικτύου προς την κατεύθυνση της πιο απότομης μείωσης του σφάλματος.
  - Επαναληπτικά βελτιώνει τα βάρη έως ότου το σφάλμα φτάσει σε μια ελάχιστη τιμή.
- **Πώς λειτουργεί?**
  - **Υπολογισμός κλίσης:** Υπολογίστε τη διαβάθμιση του σφάλματος σε σχέση με κάθε βάρος. Αυτό δείχνει την κατεύθυνση και το μέγεθος της αλλαγής που απαιτείται.
  - **Ενημέρωση βαρών:** Προσαρμόστε τα βάρη κατά ένα κλάσμα (ρυθμός εκμάθησης) της διαβάθμισης.
  - **Επανάληψη:** Επαναλάβετε τη διαδικασία έως ότου το σφάλμα σταματήσει να μειώνεται ή μειώνεται πολύ αργά.
- **Ποσοστό εκμάθησης:**
  - Μια υπερπαράμετρος που καθορίζει το μέγεθος του βήματος κατά τις ενημερώσεις βάρους.
  - Πολύ μεγάλο: Μπορεί να υπερβεί το ελάχιστο και να προκαλέσει ταλαντώσεις.
  - Πολύ μικρό: Η σύγκλιση μπορεί να είναι πολύ αργή.



Το Gradient Descent είναι η κινητήρια δύναμη πίσω από τη διαδικασία μάθησης στα MLP. Με την πλοήγηση στο τοπίο των σφαλμάτων και τις στρατηγικές προσαρμογές βάρους, διασφαλίζει ότι το δίκτυο συγκλίνει σε μια βέλτιστη λύση.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Τα Μαθηματικά Πίσω από τη Μάθηση

1. **Weight Update Equation:**
   - $W_{new} = W_{old} - \alpha \times \nabla E$
     - W_{new}: Updated weight
     - W_{old}: Previous weight
     - α: Learning rate
     - ∇E: Gradient of the error
2. **Error Gradient:**
   - For a given weight w and error E, the gradient is given by:
     - $\nabla E_w = \frac{\partial E}{\partial w}$
   - Represents how much the error changes with a small change in weight.
3. **Chain Rule & Back Propagation:**
   - The gradient is often computed using the chain rule of calculus.
   - This allows the error from the output layer to be propagated backward through the network, adjusting weights in all layers.

$$\frac{\partial C}{\partial w_{jk}^l} = \frac{\partial C}{\partial z_j^l}\frac{\partial z_j^l}{\partial w_{jk}^l} \qquad chain\ rule$$

$$z_j^l = \sum_{k=1}^{m} w_{jk}^l a_k^{l-1} + b_j^l \qquad by\ definition$$

$$m\ -\ number\ of\ neurons\ in\ l-1\ layer$$

$$\frac{\partial z_j^l}{\partial w_{jk}^l} = a_k^{l-1} \qquad by\ differentiation\ (calculating\ derivative)$$

$$\frac{\partial C}{\partial w_{jk}^l} = \frac{\partial C}{\partial z_j^l} a_k^{l-1} \qquad final\ value$$

Η διαδικασία ενημέρωσης βάρους στα MLP διέπεται από μαθηματικές εξισώσεις. Αυτές οι εξισώσεις διασφαλίζουν ότι τα βάρη προσαρμόζονται με τρόπο που μειώνει συστηματικά το σφάλμα, οδηγώντας τη διαδικασία εκμάθησης.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Παρτίδες & Μίνι-Παρτίδες

Τι είναι οι παρτίδες;

- Ομαδοποίηση δεδομένων εκπαίδευσης που χρησιμοποιούνται για την ενημέρωση βαρών δικτύου.
- Τύποι:
  - **Batch Gradient Descent:** Χρησιμοποιεί ολόκληρο το σύνολο δεδομένων για τον υπολογισμό της διαβάθμισης και την ενημέρωση των βαρών μία φορά ανά εποχή.
  - **Στοχαστική κάθοδος κλίσης (SGD):** Χρησιμοποιεί ένα σημείο δεδομένων τη φορά για την ενημέρωση των βαρών.
  - **Mini-Batch Gradient Descent:** Επιβεβαιώνει μια ισορροπία χρησιμοποιώντας ένα υποσύνολο του συνόλου δεδομένων για κάθε ενημέρωση.

Γιατί Mini-Batches;

- Ταχύτερη σύγκλιση: Οι πιο συχνές ενημερώσεις βάρους μπορούν να οδηγήσουν σε ταχύτερη μάθηση.
- Αποδοτικότητα μνήμης: Κατάλληλο για μεγάλα σύνολα δεδομένων που δεν χωρούν στη μνήμη.
- Μπορεί να ξεφύγει από τα τοπικά ελάχιστα λόγω του θορύβου στην εκτίμηση της κλίσης.

Επιλέγοντας το σωστό μέγεθος παρτίδας:

- Πολύ μικρό: Μπορεί να οδηγήσει σε θορυβώδεις ενημερώσεις βάρους.
- Πολύ μεγάλο: Μπορεί να είναι υπολογιστικά ακριβό και μπορεί να κολλήσει στα τοπικά ελάχιστα.
- Ένα τυπικό εύρος είναι μεταξύ 32 και 512, αλλά εξαρτάται από το πρόβλημα.

Η επιλογή του μεγέθους της παρτίδας μπορεί να επηρεάσει σημαντικά την αποδοτικότητα και την αποτελεσματικότητα της μαθησιακής διαδικασίας. Η βαθμίδα κάθοδος μίνι παρτίδας παρέχει συχνά μια καλή ισορροπία μεταξύ ταχύτητας και σταθερότητας.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Παρτίδες εναντίον Μίνι-Παρτίδες εναντίον Στοχαστικών

Κάθοδος διαβάθμισης παρτίδας:

- **Ορισμός:** Χρησιμοποιεί ολόκληρο το σύνολο δεδομένων για τον υπολογισμό της κλίσης.
- **Πλεονεκτήματα:** Σταθερή σύγκλιση, σταθερή κατεύθυνση κλίσης.
- **Μειονεκτήματα:** Μπορεί να είναι αργή, ειδικά για μεγάλα σύνολα δεδομένων. Εντατική μνήμη. Μπορεί να κολλήσει στα τοπικά ελάχιστα.
- **Περίπτωση χρήσης:** Ιδανικό για μικρότερα σύνολα δεδομένων ή όταν είναι εγγυημένη μια κυρτή επιφάνεια σφάλματος.

Στοχαστική κλίση κάθοδος (SGD):

- **Ορισμός:** Χρησιμοποιεί ένα σημείο δεδομένων τη φορά για ενημερώσεις βάρους.
- **Πλεονεκτήματα:** Ταχύτερη σύγκλιση για μεγάλα σύνολα δεδομένων. Η εγγενής τυχαιότητα μπορεί να ξεφύγει από τα τοπικά ελάχιστα.
- **Μειονεκτήματα:** Θορυβώδης σύγκλιση λόγω συχνών ενημερώσεων. Πιο επιρρεπείς σε ταλαντώσεις.
- **Περίπτωση χρήσης:** Όταν οι υπολογιστικοί πόροι είναι περιορισμένοι ή όταν το σύνολο δεδομένων έχει πολύ πλεονασμό.

Mini-Batch Gradient Descent:

- **Ορισμός:** Μια μέση λύση μεταξύ Batch και SGD.
- **Πλεονεκτήματα:** Εξισορροπεί την ταχύτητα και τη σταθερότητα. Κατάλληλο για παράλληλη επεξεργασία σε υλικό όπως GPU.
- **Μειονεκτήματα:** Απαιτεί προσεκτική επιλογή μεγέθους παρτίδας. Ενδέχεται να εξακολουθεί να αντιμετωπίζει προκλήσεις τόσο των μεθόδων SGD όσο και των μεθόδων παρτίδας.
- **Περίπτωση χρήσης:** Οι περισσότερες εφαρμογές πραγματικού κόσμου, ειδικά η βαθιά εκμάθηση, λόγω της ισορροπίας των πλεονεκτημάτων τους.

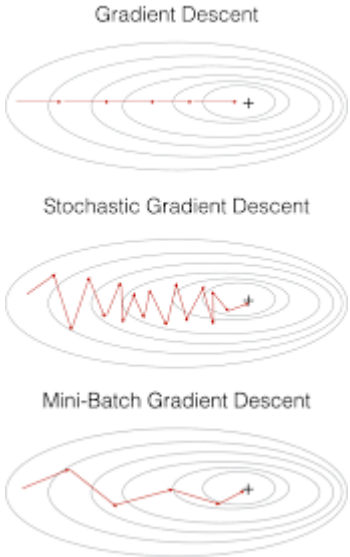
Πρόσθετες εκτιμήσεις:

- **Ποσοστό εκμάθησης:** Κρίσιμο για όλες τις μεθόδους. Οι τεχνικές προσαρμοστικού ρυθμού μάθησης μπορούν να βοηθήσουν στη βελτιστοποίηση της σύγκλισης.
- **Εποχές εναντίον επαναλήψεων:** Η εποχή είναι ένα πλήρες πέρασμα προς τα εμπρός και προς τα πίσω όλων των παραδειγμάτων εκπαίδευσης. Μια επανάληψη είναι ένα πέρασμα μιας μόνο παρτίδας. Το SGD και το Mini-Batch απαιτούν περισσότερες επαναλήψεις ανά εποχή.

Η επιλογή μεταξύ Batch, SGD και Mini-Batch εξαρτάται από το μέγεθος δεδομένων, τους περιορισμούς μνήμης και τις επιθυμητές ιδιότητες σύγκλισης. Συχνά, το Mini-Batch προσφέρει μια καλή αντιστάθμιση μεταξύ ταχύτητας και σταθερότητας.

Στοχαστική Κάθοδος Κλίσης

- **Τι είναι το SGD;**
  a. Μια παραλλαγή κλίσης κατάβασης όπου τα βάρη ενημερώνονται μετά από κάθε παράδειγμα προπόνησης.
  b. Εισάγει την τυχαιότητα στη διαδικασία εκμάθησης, καθώς οι ενημερώσεις βασίζονται σε μεμονωμένα σημεία δεδομένων.
- **Πλεονεκτήματα του SGD:**
  a. **Ταχύτερη σύγκλιση:** Οι συχνές ενημερώσεις μπορούν να οδηγήσουν σε ταχύτερη μάθηση.
  b. **Escape Local Minima:** Η θορυβώδης φύση των ενημερώσεων μπορεί να βοηθήσει στην αποφυγή των ρηχών τοπικών ελάχιστων.
  c. **Αποδοτικότητα μνήμης:** Δεν απαιτεί ολόκληρο το σύνολο δεδομένων στη μνήμη ταυτόχρονα.
- **Προκλήσεις & Λύσεις:**
  a. **Θορυβώδεις ενημερώσεις:** Μπορεί να οδηγήσει σε ταλαντώσεις στη μάθηση. Λύση: Χρησιμοποιήστε έναν μειωμένο ρυθμό μάθησης ή ορμή.
  b. **Ευαίσθητο στο ρυθμό μάθησης:** Απαιτεί προσεκτικό συντονισμό. Λύση: Τεχνικές προσαρμοστικού ρυθμού μάθησης όπως το Adagrad, το RMSprop ή το Adam.



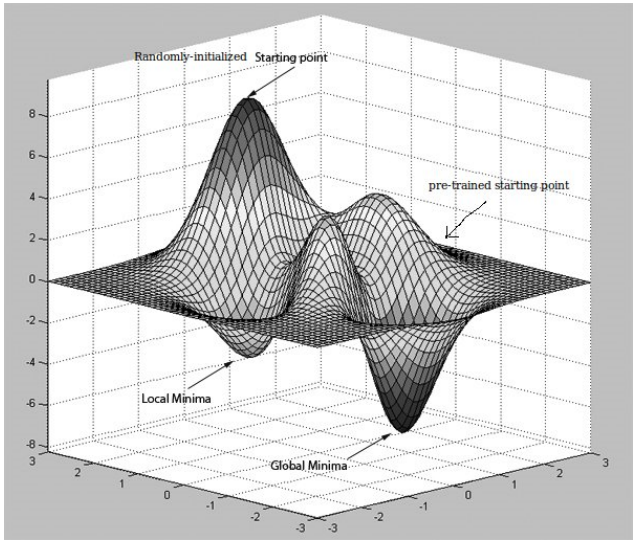
Το SGD προσφέρει μια δυναμική προσέγγιση στη μάθηση, με συχνές ενημερώσεις που βασίζονται σε μεμονωμένα σημεία δεδομένων. Ενώ εισάγει την τυχαιότητα, με τις σωστές τεχνικές, μπορεί να οδηγήσει σε αποτελεσματική και αποτελεσματική μάθηση.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Σφάλμα Επιφάνεια & Τοπίο βελτιστοποίησης

- **Τι είναι η επιφάνεια σφάλματος;**
  a. Μια γραφική αναπαράσταση του σφάλματος ως συνάρτηση των βαρών του δικτύου. Οπτικοποιεί το τοπίο βελτιστοποίησης στο οποίο πλοηγούνται οι αλγόριθμοι ντεγκραντέ.
- **Χαρακτηριστικά της επιφάνειας σφάλματος:**
  a. **Τοπικά ελάχιστα:** Σημεία όπου το σφάλμα είναι χαμηλότερο από τα γειτονικά σημεία αλλά όχι το χαμηλότερο συνολικά.
  b. **Καθολικό ελάχιστο:** Το σημείο με το χαμηλότερο σφάλμα σε ολόκληρη την επιφάνεια.
  c. **Σημεία σέλας:** Τοποθεσίες όπου το σφάλμα είναι επίπεδο σε ορισμένες κατευθύνσεις αλλά με κλίση σε άλλες.
- **Προκλήσεις στη Βελτιστοποίηση:**
  a. Κολλήστε στα τοπικά ελάχιστα ή σημεία σέλας.
  b. Ταλαντώσεις ή αργή σύγκλιση λόγω του σχήματος της επιφάνειας.
- **Ενίσχυση βελτιστοποίησης:**
  a. **Ορμή:** Βοηθά τον αλγόριθμο να κινείται πιο γρήγορα σε επίπεδες περιοχές και να αποφεύγει να κολλήσει.
  b. **Προσαρμοστικοί ρυθμοί μάθησης:** Προσαρμόζει το ρυθμό εκμάθησης με βάση το τοπίο, π.χ. Adam ή RMSprop.



Η επιφάνεια σφάλματος παρέχει πληροφορίες για τις προκλήσεις που αντιμετωπίζονται κατά τη βελτιστοποίηση. Κατανοώντας τα χαρακτηριστικά του, μπορούμε να χρησιμοποιήσουμε τεχνικές για να το πλοηγηθούμε πιο αποτελεσματικά και να επιτύχουμε καλύτερα μαθησιακά αποτελέσματα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πρακτικές ρυθμίσεις βάρους

Ο νόμος εξισορρόπησης:

- Η προσαρμογή των βαρών είναι μια ισορροπία μεταξύ της γρήγορης εκμάθησης και της μη υπέρβασης των βέλτιστων τιμών.
- Σημασία του ρυθμού εκμάθησης: Το πολύ υψηλό μπορεί να προκαλέσει απόκλιση, το πολύ χαμηλό μπορεί να προκαλέσει αργή σύγκλιση.

Ξεπερνώντας τα οροπέδια:

- Κατά τη διάρκεια της εκπαίδευσης, το μοντέλο μπορεί να συναντήσει περιοχές όπου το σφάλμα δεν αλλάζει πολύ (πλατώ).
- Τεχνικές όπως η ορμή μπορούν να βοηθήσουν στην προώθηση αυτών των περιοχών.

Αποφυγή ταλαντώσεων:

- Εάν τα βάρη συνεχίζουν να ταλαντώνονται εμπρός και πίσω χωρίς να καθιζάνουν, αυτό δείχνει υψηλό ρυθμό μάθησης.
- Οι τεχνικές προσαρμοστικού ρυθμού μάθησης μπορούν να βοηθήσουν στη μείωση των ταλαντώσεων.

Τεχνικές τακτοποίησης:

- Μέθοδοι όπως η τακτοποίηση L1 και L2 μπορούν να αποτρέψουν τα βάρη από το να γίνουν πολύ μεγάλα, γεγονός που μπορεί να οδηγήσει σε υπερβολική προσαρμογή.
- Προσθέτουν μια ποινή στη συνάρτηση απώλειας με βάση το μέγεθος των βαρών.

Αρχικοποίηση βάρους:

- Τα αρχικά βάρη μπορούν να επηρεάσουν τη διαδικασία προπόνησης.
- Τεχνικές όπως η αρχικοποίηση Xavier και He μπορούν να ορίσουν τα βάρη με τρόπο που να βοηθά την ταχύτερη σύγκλιση.

Η προσαρμογή των βαρών στα νευρωνικά δίκτυα δεν αφορά μόνο την τήρηση μαθηματικών κανόνων. Πρόκειται για την κατανόηση των προκλήσεων και τη χρήση στρατηγικών για την αποτελεσματική πλοήγησή τους.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Προηγμένες Τεχνικές Βελτιστοποίησης

Ορμή:

- **Έννοια:** Ενσωματώνει την κατεύθυνση της προηγούμενης κλίσης για να επιταχύνει τη σύγκλιση.
- **Οφέλη:** Βοηθά στην ταχύτερη σύγκλιση και διαφυγή των τοπικών ελάχιστων.

Adagrad (Adaptive Gradient Algorithm):

- **Έννοια:** Προσαρμόζει τους ρυθμούς εκμάθησης όλων των παραμέτρων του μοντέλου, κλιμακώνοντάς τες αντιστρόφως ανάλογες με την τετραγωνική ρίζα του αθροίσματος των ιστορικών τετραγωνικών τιμών.
- **Πλεονέκτημα:** Επιτρέπει διαφορετικούς ρυθμούς εκμάθησης για κάθε παράμετρο.

RMSprop (Μέση τετραγωνική διάδοση ρίζας):

- **Έννοια:** Τροποποιεί το Adagrad για να χρησιμοποιεί έναν κινητό μέσο όρο τετραγωνικών κλίσεων.
- **Οφέλη:** Αντιμετωπίζει τα ριζικά μειωμένα ποσοστά μάθησης του Adagrad.

Adam (Προσαρμοστική Εκτίμηση Στιγμών):

- **Έννοια:** Συνδυάζει ιδέες από το Momentum και το RMSprop. Διατηρεί έναν εκθετικά μειωμένο μέσο όρο των προηγούμενων κλίσεων.
- **Όφελος:** Εξισορροπεί τα οφέλη τόσο του Momentum όσο και του RMSprop.

Nesterov Accelerated Gradient (NAG):

- **Έννοια:** Μια προσαρμογή στην ορμή όπου η κλίση υπολογίζεται μετά το τρέχον βήμα ορμής.
- **Όφελος:** Παρέχει μια πιο ακριβή εκτίμηση της κλίσης.

Οι προηγμένες τεχνικές βελτιστοποίησης παρέχουν εκλεπτυσμένους τρόπους πλοήγησης στο τοπίο σφαλμάτων, διασφαλίζοντας ταχύτερη και πιο σταθερή σύγκλιση σε πολύπλοκα σενάρια.

Co-funded by the Erasmus+ Programme of the European Union

Συμπέρασμα & Μελλοντική Προοπτική

- **Ανακεφαλαιώνουμε:**
  - **Η ισχύς του MLP:** Τα πολυστρωματικά perceptron προσφέρουν έναν βαθύτερο και πιο περίπλοκο τρόπο μοντελοποίησης σύνθετων σχέσεων σε δεδομένα σε σύγκριση με τα perceptron μονής στρώσης.
  - **Βασικές έννοιες:** Τα κρυφά επίπεδα, οι συναρτήσεις ενεργοποίησης, οι συναρτήσεις σφάλματος και οι προηγμένες τεχνικές βελτιστοποίησης αποτελούν τον πυρήνα των MLP.
- **Μελλοντικές προοπτικές:**
  - **Deep Learning:** Τα MLP είναι θεμελιώδη για τα βαθιά νευρωνικά δίκτυα, τα οποία φέρνουν επανάσταση σε τομείς όπως η όραση υπολογιστή, η επεξεργασία φυσικής γλώσσας και πολλά άλλα.
  - **Προσαρμογή:** Καθώς η έρευνα προχωρά, θα δούμε πιο προσαρμοσμένες λειτουργίες ενεργοποίησης, τεχνικές βελτιστοποίησης και αρχιτεκτονικές για συγκεκριμένες εργασίες.
  - **Προκλήσεις:** Αν και τα MLP είναι ισχυρά, έρχονται με προκλήσεις όπως η υπερβολική προσαρμογή, οι διαβαθμίσεις που εξαφανίζονται και οι υπολογιστικές απαιτήσεις. Η συνεχιζόμενη έρευνα στοχεύει στον μετριασμό αυτών.
- **Συνεχής Μάθηση:**
  - **Μείνετε ενημερωμένοι:** Το πεδίο των νευρωνικών δικτύων εξελίσσεται ταχέως. Η συνεχής μάθηση και η ενημέρωση με τις πιο πρόσφατες έρευνες είναι ζωτικής σημασίας.
  - **Πρακτική πρακτική:** Η θεωρητική γνώση είναι ζωτικής σημασίας, αλλά η πρακτική εφαρμογή και ο πειραματισμός θα ενισχύσουν την κατανόηση.

Τα MLP άνοιξαν το δρόμο για την επανάσταση της βαθιάς μάθησης. Η αποδοχή των εννοιών τους και η κατανόηση των περιπλοκών τους είναι ζωτικής σημασίας για όποιον καταδύεται στον κόσμο των τεχνητών νευρωνικών δικτύων.

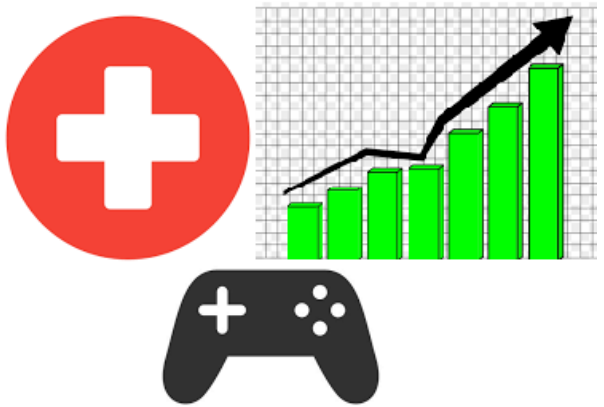
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πολυστρωματικά Perceptrons

Γιορτάζοντας την επιτυχία των MLPs

- **Επισκόπηση ταξιδιού:**
  - Από την ίδρυσή τους, τα MLP ήταν στην πρώτη γραμμή πολλών ανακαλύψεων στην τεχνητή νοημοσύνη. Αυτό το μάθημα γιορτάζει τα επιτεύγματά τους και τις κρίσιμες στιγμές τους.
- **Γιατί να γιορτάσουμε τα MLP;**
  - Έχουν ξεπεράσει σημαντικές προκλήσεις, όπως το πρόβλημα XOR, θέτοντας τις βάσεις για πιο προηγμένα νευρωνικά δίκτυα.
  - Η προσαρμοστικότητά τους έχει οδηγήσει σε επιτυχίες σε διάφορους τομείς, από την υγειονομική περίθαλψη μέχρι τη χρηματοδότηση.
- **Μια ματιά μπροστά:**
  - Ενώ θα ξεκινήσουμε με τον θρίαμβο του XOR, θα εξερευνήσουμε επίσης σύγχρονες επιτυχίες, δείχνοντας τη διαρκή συνάφεια των MLP.



Η ιστορία των MLP είναι μια ιστορία ανθεκτικότητας, προσαρμοστικότητας και συνεχούς εξέλιξης. Ελάτε μαζί μας καθώς ταξιδεύουμε στις αξιοσημείωτες επιτυχίες τους.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Το πρόβλημα XOR - Σημείο καμπής

Η πρόκληση XOR:

- Το πρόβλημα XOR (αποκλειστικό OR) ήταν μια σημαντική πρόκληση για τα perceptron μονής στρώσης. Αντιπροσωπεύει ένα σενάριο όπου τα δεδομένα δεν διαχωρίζονται γραμμικά, που σημαίνει ότι μια ευθεία γραμμή δεν μπορεί να διαχωρίσει τις κλάσεις.
- Παράδειγμα: Για δυαδικές εισόδους (0 ή 1), το XOR βγάζει 1 μόνο όταν οι είσοδοι είναι διαφορετικές.

Θρίαμβος του MLP:

- Ενώ τα μονοστρωματικά perceptron απέτυχαν να λύσουν το πρόβλημα XOR, η εισαγωγή ενός κρυφού στρώματος στα MLPs το κατέστησε δυνατό.
- Αυτή η επιτυχία έδειξε τη δύναμη του βάθους στα νευρωνικά δίκτυα, ανοίγοντας το δρόμο για πιο σύνθετες αρχιτεκτονικές.

Σημασία:

- Η ικανότητα επίλυσης του προβλήματος XOR ήταν μια **σημαντική στιγμή** στην τεχνητή νοημοσύνη. Παρουσίασε τις δυνατότητες των νευρωνικών δικτύων για την αντιμετώπιση μη γραμμικών προβλημάτων.
- Τόνισε επίσης τους περιορισμούς των απλούστερων μοντέλων και την ανάγκη για πιο εξελιγμένες αρχιτεκτονικές.

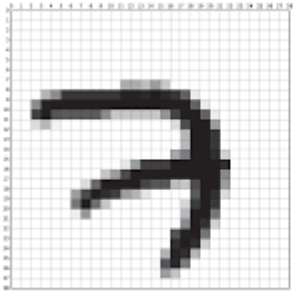
Το πρόβλημα XOR χρησίμευσε τόσο ως πρόκληση όσο και ως επικύρωση για τα νευρωνικά δίκτυα. Η επιτυχία του MLP στην επίλυσή του σηματοδότησε μια σημαντική πρόοδο στον τομέα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

MLP στην Αναγνώριση Εικόνας

- **Πρώιμες ημέρες αναγνώρισης εικόνας:**
  a. Πριν από την εποχή της βαθιάς μάθησης, τα MLP ήταν μεταξύ των πρωταρχικών εργαλείων που χρησιμοποιούνταν για εργασίες αναγνώρισης εικόνων.
  b. Χρησιμοποιήθηκαν για να αναγνωρίσουν χειρόγραφα ψηφία, βασικά σχήματα και άλλα απλά οπτικά μοτίβα.
- **Πώς λειτουργούν τα MLP στην Αναγνώριση Εικόνων:**
  a. Οι εικόνες αναλύονται σε τιμές pixel, οι οποίες χρησιμεύουν ως είσοδοι στο MLP.
  b. Το δίκτυο μαθαίνει να αναγνωρίζει μοτίβα και χαρακτηριστικά, όπως άκρες και υφές, μέσω των επιπέδων του.
- **Επιτεύγματα και ορόσημα:**
  a. Τα MLP πέτυχαν αξιοσημείωτη επιτυχία σε σύνολα δεδομένων όπως το MNIST (χειρόγραφα ψηφία).
  b. Έθεσαν τα θεμέλια για πιο προηγμένες αρχιτεκτονικές όπως τα Συνελικτικά Νευρωνικά Δίκτυα (CNN) που κυριαρχούν στον τομέα σήμερα.



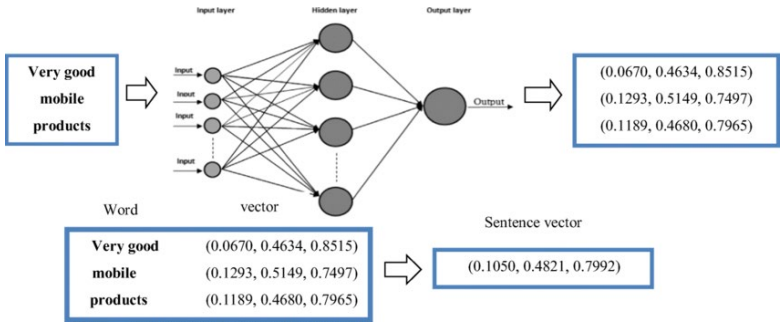
(a) MNIST sample belonging to the digit '7'.

(b) 100 samples from the MNIST training set.

Ενώ τα σύγχρονα μοντέλα βαθιάς μάθησης έχουν ξεπεράσει τα MLP στην αναγνώριση εικόνων, το θεμελιώδες έργο που επιτελείται από τους MLP σε αυτόν τον τομέα είναι αναμφισβήτητο και άνοιξε το δρόμο για μελλοντικές εξελίξεις.

MLP στην Επεξεργασία Φυσικής Γλώσσας (NLP)

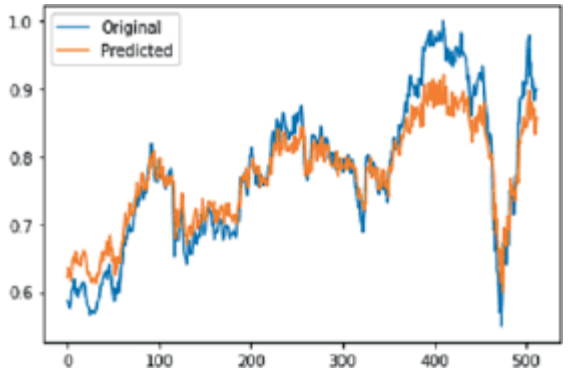
- **NLP - Μια σύντομη επισκόπηση:**
  a. Η Επεξεργασία Φυσικής Γλώσσας περιλαμβάνει το να κάνουμε τους υπολογιστές να κατανοούν, να ερμηνεύουν και να δημιουργούν ανθρώπινη γλώσσα.
  b. Είναι ένα πολύπλοκο έργο λόγω των αποχρώσεων, των ιδιωμάτων και των δομών στις γλώσσες.
- **MLP σε πρώιμες εργασίες NLP:**
  a. Πριν από την εμφάνιση μοντέλων όπως οι μετασχηματιστές, τα MLP χρησιμοποιούνταν για βασικές εργασίες NLP.
  b. Χρησιμοποιήθηκαν για ταξινόμηση κειμένων, ανάλυση συναισθημάτων και βασική μοντελοποίηση γλώσσας.
- **Πώς τα MLP επεξεργάζονται το κείμενο:**
  a. Τα δεδομένα κειμένου μετατρέπονται σε αριθμητικά διανύσματα χρησιμοποιώντας τεχνικές όπως το TF-IDF ή ενσωματώσεις λέξεων.
  b. Αυτά τα διανύσματα χρησιμεύουν ως είσοδοι στο MLP, το οποίο στη συνέχεια μαθαίνει μοτίβα και σχέσεις μεταξύ λέξεων και φράσεων.



Τα MLP έπαιξαν θεμελιώδη ρόλο στις πρώιμες εργασίες NLP, συμβάλλοντας στη γεφύρωση του χάσματος μεταξύ της ανθρώπινης γλώσσας και της μηχανικής κατανόησης, θέτοντας τις βάσεις για πιο προηγμένα μοντέλα NLP

MLPs στη χρηματοοικονομική πρόβλεψη

- **Οικονομικές προβλέψεις - Η πρόκληση:**
  - Η πρόβλεψη των κινήσεων της χρηματιστηριακής αγοράς, των συναλλαγματικών ισοτιμιών και άλλων χρηματοοικονομικών μετρήσεων είναι εγγενώς πολύπλοκη λόγω πολλών παραγόντων που επηρεάζουν.
  - Οι ακριβείς προβλέψεις μπορούν να οδηγήσουν σε σημαντικά οικονομικά οφέλη, καθιστώντας την περιζήτητη εφαρμογή της τεχνητής νοημοσύνης.
- **Ο ρόλος των MLP:**
  - Τα MLP έχουν χρησιμοποιηθεί για τη μοντελοποίηση και την πρόβλεψη δεδομένων οικονομικών χρονοσειρών.
  - Αποτυπώνουν μοτίβα, τάσεις και σχέσεις σε ιστορικά δεδομένα για να κάνουν μελλοντικές προβλέψεις.
- **Χαρακτηριστικά & Επεξεργασία Δεδομένων:**
  - Τα οικονομικά δεδομένα, όπως οι τιμές των μετοχών, οι όγκοι συναλλαγών και οι τεχνικοί δείκτες, τροφοδοτούνται στα MLP.
  - Η κανονικοποίηση και η προεπεξεργασία δεδομένων είναι ζωτικής σημασίας για να διασφαλιστεί ότι το δίκτυο μαθαίνει αποτελεσματικά από τις τάσεις του παρελθόντος.



Ενώ οι σύγχρονοι αλγόριθμοι και μοντέλα έχουν βελτιώσει περαιτέρω τις οικονομικές προβλέψεις, τα MLP έθεσαν τις βάσεις, επιδεικνύοντας τις δυνατότητες των νευρωνικών δικτύων στην πρόβλεψη πολύπλοκων οικονομικών μετρήσεων.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

MLP στη Διαγνωστική Υγείας

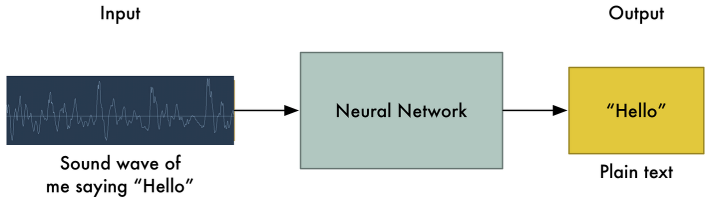
- **Διαγνωστικά υγειονομικής περίθαλψης - Η σημασία:**
  - Η ακριβής και έγκαιρη διάγνωση είναι ζωτικής σημασίας για την υγειονομική περίθαλψη, επηρεάζοντας τις αποφάσεις θεραπείας και τα αποτελέσματα των ασθενών.
  - Η τεχνητή νοημοσύνη έχει τη δυνατότητα να βοηθήσει τους επαγγελματίες υγείας, διασφαλίζοντας καλύτερη φροντίδα των ασθενών.
- **MLP στην Ιατρική Απεικόνιση και Διαγνωστική:**
  - Πριν από την εμφάνιση πιο προηγμένων μοντέλων βαθιάς μάθησης, τα MLP χρησιμοποιούνταν για την ανάλυση ιατρικών εικόνων όπως ακτινογραφίες, μαγνητικές τομογραφίες και αξονικές τομογραφίες.
  - Βοήθησαν στον εντοπισμό ανωμαλιών, όγκων και άλλων ιατρικών καταστάσεων.
- **Δεδομένα & Χαρακτηριστικά:**
  - Οι ιατρικές εικόνες υποβάλλονται σε επεξεργασία και μετατρέπονται σε μορφή κατάλληλη για MLP.
  - Χαρακτηριστικά όπως άκρες, υφές και μοτίβα εξάγονται και χρησιμοποιούνται για διάγνωση.

Τα MLP έχουν παίξει καθοριστικό ρόλο στην έγκαιρη ιατρική διάγνωση με χρήση τεχνητής νοημοσύνης, ενισχύοντας την ακρίβεια και την αποτελεσματικότητα της ανίχνευσης ασθενειών και της φροντίδας των ασθενών.

MLP στην Αναγνώριση Φωνής

- **Αναγνώριση φωνής - Η Επανάσταση:**
  - Η ικανότητα αναγνώρισης και ερμηνείας της ανθρώπινης φωνής έχει φέρει επανάσταση στις διεπαφές χρήστη, οδηγώντας σε βοηθούς φωνής, συστήματα που ενεργοποιούνται με φωνή και πολλά άλλα.
  - Περιλαμβάνει την κατανόηση μοτίβων ομιλίας, τόνους και αποχρώσεις.
- **MLP στην πρώιμη αναγνώριση φωνής:**
  - Πριν ξεκινήσουν τα μοντέλα βαθιάς μάθησης όπως τα RNN και τα LSTM, τα MLP ήταν μεταξύ των εργαλείων που χρησιμοποιήθηκαν για βασικές εργασίες αναγνώρισης φωνής.
  - Χρησιμοποιήθηκαν για τον εντοπισμό συγκεκριμένων φωνητικών εντολών, την αναγνώριση των ηχείων και πολλά άλλα.

- **Επεξεργασία δεδομένων και χαρακτηριστικά:**
  - Τα φωνητικά δεδομένα μετατρέπονται σε φασματογράμματα ή άλλες αναπαραστάσεις κατάλληλες για MLP.
  - Το δίκτυο μαθαίνει από τα μοτίβα συχνότητας, τις διακυμάνσεις του τόνου και άλλα χαρακτηριστικά φωνής.



Τα MLP έπαιξαν καθοριστικό ρόλο στα πρώτα στάδια της τεχνολογίας αναγνώρισης φωνής, θέτοντας τα θεμέλια για τα εξελιγμένα συστήματα ενεργοποίησης φωνής που έχουμε σήμερα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

MLP σε παιχνίδια και προσομοιώσεις

- **Παιχνίδια & Προσομοιώσεις - The Digital Playground:**
  - Τα βιντεοπαιχνίδια και οι προσομοιώσεις προσφέρουν καθηλωτικές εμπειρίες, απαιτώντας συχνά την τεχνητή νοημοσύνη για τη δημιουργία ρεαλιστικών και δυναμικών περιβαλλόντων.
  - Οι χαρακτήρες που βασίζονται σε AI (NPC) και τα γεγονότα ενισχύουν το παιχνίδι και την αφοσίωση των χρηστών.
- **MLP στο Game AI:**
  - Στις πρώτες μέρες της τεχνητής νοημοσύνης του παιχνιδιού, τα MLP χρησιμοποιήθηκαν για να οδηγήσουν τη συμπεριφορά των NPC, κάνοντάς τα να αντιδρούν στις ενέργειες των παικτών.
  - Βοήθησαν στην εξεύρεση μονοπατιών, στη λήψη αποφάσεων και στη δημιουργία προσαρμοστικών προκλήσεων για τους παίκτες.
- **Δεδομένα & Μάθηση:**
  - Τα δεδομένα παιχνιδιού, όπως οι ενέργειες παίκτη, οι καταστάσεις παιχνιδιού και οι μεταβλητές περιβάλλοντος, είναι είσοδοι για το MLP.
  - Το δίκτυο μαθαίνει να προβλέπει τη συμπεριφορά των παικτών και να προσαρμόζει ανάλογα το περιβάλλον του παιχνιδιού.

Τα MLP έχουν συμβάλει στην εξέλιξη της τεχνητής νοημοσύνης παιχνιδιών, κάνοντας τους εικονικούς κόσμους πιο διαδραστικούς, απρόβλεπτους και ελκυστικούς.

MLP στο Customer Analytics

- **Αναλύσεις πελατών - The Business Edge:**
  - Η κατανόηση της συμπεριφοράς, των προτιμήσεων και των προτύπων των πελατών είναι ζωτικής σημασίας για τις επιχειρήσεις να προσαρμόσουν τις προσφορές τους και να βελτιώσουν την ικανοποίηση των πελατών.
  - Τα προγνωστικά αναλυτικά στοιχεία μπορούν να προβλέψουν τις ανάγκες των πελατών, ενισχύοντας τις στρατηγικές μάρκετινγκ και πωλήσεων.
- **Ο ρόλος των MLP στην πρώιμη ανάλυση πελατών:**
  - Τα MLP χρησιμοποιήθηκαν για την ανάλυση δεδομένων πελατών, την τμηματοποίηση αγορών και την πρόβλεψη της αγοραστικής συμπεριφοράς.
  - Βοήθησαν τις επιχειρήσεις να εντοπίσουν πιθανούς πελάτες υψηλής αξίας, ποσοστά απόρριψης και πολλά άλλα.
- **Πληροφορίες και δυνατότητες δεδομένων:**
  - Τα δεδομένα από τις πωλήσεις, τις αλληλεπιδράσεις με τους πελάτες, τα σχόλια και τη διαδικτυακή συμπεριφορά επεξεργάζονται από το MLP.
  - Το δίκτυο προσδιορίζει μοτίβα, τάσεις και τμήματα, παρέχοντας χρήσιμες πληροφορίες για τις επιχειρήσεις.

Αναλύοντας τεράστιες ποσότητες δεδομένων πελατών, τα MLP παρείχαν στις επιχειρήσεις πολύτιμες πληροφορίες, διαμορφώνοντας εκστρατείες μάρκετινγκ, εξελίξεις προϊόντων και στρατηγικές εξυπηρέτησης πελατών.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Προκλήσεις που ξεπερνιούνται από MLP

Υπερβολική πολυπλοκότητα και ταίριασμα:

- **Πρόκληση:** Οι MLP, με την ικανότητά τους να χωρούν σύνθετα δεδομένα, συχνά απομνημόνευαν δεδομένα εκπαίδευσης, οδηγώντας σε κακή γενίκευση των αόρατων δεδομένων.
- **Λύση:** Εισήχθησαν τεχνικές όπως η εγκατάλειψη, η τακτοποίηση και η πρόωρη διακοπή. Αυτές οι μέθοδοι εμποδίζουν το δίκτυο να γίνει πολύ περίπλοκο και βοηθούν στην επίτευξη καλύτερης γενίκευσης.

Διαβαθμίσεις εξαφάνισης και έκρηξης:

- **Πρόκληση:** Κατά τη διάρκεια της οπίσθιας διάδοσης, οι κλίσεις μπορεί να γίνουν πολύ μικρές (εξαφανίζονται) ή πολύ μεγάλες (εκρήγνυνται), κάνοντας την προπόνηση ασταθή ή στάσιμη.
- **Λύση:** Τεχνικές αρχικοποίησης, όπως η προετοιμασία He ή Xavier, και λειτουργίες ενεργοποίησης όπως το ReLU, βοήθησαν να μετριαστούν αυτά τα ζητήματα.

Τοπικά ελάχιστα:

- **Πρόκληση:** Οι MLP θα μπορούσαν να κολλήσουν στα τοπικά ελάχιστα κατά τη διάρκεια της προπόνησης, εμποδίζοντάς τους να φτάσουν στην καλύτερη δυνατή λύση.
- **Λύση:** Τα προγράμματα βελτιστοποίησης όπως η ορμή και οι προσαρμοστικοί ρυθμοί μάθησης βοήθησαν στην πιο αποτελεσματική πλοήγηση στην επιφάνεια του σφάλματος.

Υπολογιστικές Προκλήσεις:

- **Πρόκληση:** Η εκπαίδευση μεγάλων MLP απαιτούσε σημαντικούς υπολογιστικούς πόρους.
- **Λύση:** Οι επιταχυντές υλικού όπως οι GPU και οι κατανεμημένες τεχνικές εκπαίδευσης κατέστησαν εφικτή την εκπαίδευση μεγαλύτερων και πιο περίπλοκων MLP.

Παρά τις πολυάριθμες προκλήσεις, εμφανίστηκαν καινοτομίες και τεχνικές για τη βελτίωση της εκπαίδευσης και της απόδοσης των MLP, ενισχύοντας τη θέση τους στο τοπίο της τεχνητής νοημοσύνης.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

The Modern Era - Deep Learning

Εξέλιξη σε βαθιά μάθηση:

- Ενώ τα MLP έθεσαν τα θεμέλια, η εισαγωγή των βαθιάς νευρωνικών δικτύων (DNN) σηματοδότησε ένα σημαντικό άλμα στις δυνατότητες της τεχνητής νοημοσύνης.
- Τα DNN αποτελούνται από πολλαπλά κρυφά επίπεδα, επιτρέποντάς τους να μοντελοποιούν ακόμη πιο πολύπλοκες σχέσεις στα δεδομένα.

Προόδους στο Υλικό:

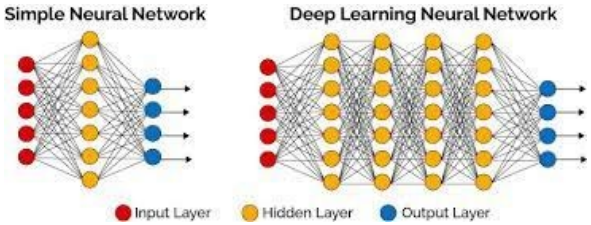
- Η άνοδος των GPU και των TPU διευκόλυνε την εκπαίδευση βαθύτερων δικτύων, διαχειρίζοντας αποτελεσματικά εκατομμύρια παραμέτρους.
- Οι δυνατότητες παράλληλης επεξεργασίας κατέστησαν εφικτή την εκπαίδευση δικτύων σε τεράστια σύνολα δεδομένων.

Σημαντικές ανακαλύψεις σε διάφορους τομείς:

- Η βαθιά εκμάθηση έχει επιτύχει αποτελέσματα αιχμής σε τομείς όπως η αναγνώριση εικόνων, η επεξεργασία φυσικής γλώσσας και η ενισχυτική μάθηση.
- Καινοτομίες όπως τα συνελικτικά νευρωνικά δίκτυα (CNN) και τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) προέκυψαν από τις θεμελιώδεις έννοιες των MLP.

Προκλήσεις & Διαρκής Έρευνα:

- Παρά τη δύναμή τους, τα DNN έρχονται με προκλήσεις όπως η ερμηνευτικότητα και το υπολογιστικό κόστος.
- Η έρευνα σε τομείς όπως η μάθηση μεταφοράς, οι μηχανισμοί προσοχής και οι μετασχηματιστές συνεχίζει να ωθεί τα όρια του δυνατού.



Οι αρχές των MLP άνοιξαν το δρόμο για την επανάσταση βαθιάς μάθησης, μεταμορφώνοντας το τοπίο της τεχνητής νοημοσύνης και οδηγώντας σε ανακαλύψεις σε πολλούς τομείς.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ο δρόμος μπροστά για MLP και Deep Learning

Επέκταση Τομέων:

- Καθώς η τεχνολογία προχωρά, η εφαρμογή των MLP και η βαθιά μάθηση θα επεκταθεί σε νεότερους τομείς όπως ο κβαντικός υπολογισμός, η νευρομορφική μηχανική και πολλά άλλα.
- Η ενσωμάτωση με άλλες τεχνικές τεχνητής νοημοσύνης, όπως η ενισχυτική μάθηση και τα παραγωγικά μοντέλα, θα οδηγήσει σε πιο ολιστικές λύσεις.

Αντιμετώπιση τρεχουσών προκλήσεων:

- Η έρευνα συνεχίζεται για να καταστήσει τα μοντέλα βαθιάς μάθησης πιο ερμηνεύσιμα, ενεργειακά αποδοτικά και πιο ανθεκτικά σε αντίπαλες επιθέσεις.
- Καταβάλλονται προσπάθειες για τη μείωση της εξάρτησης των μοντέλων από δεδομένα, καθιστώντας τα πιο γενικεύσιμα με λιγότερα παραδείγματα.

Ηθική και υπεύθυνη τεχνητή νοημοσύνη:

- Καθώς η τεχνητή νοημοσύνη ενσωματώνεται περισσότερο στην καθημερινή ζωή, δίνεται όλο και μεγαλύτερη έμφαση στη δημιουργία μοντέλων που είναι ηθικά, αμερόληπτα και διαφανή.
- Οι κανονισμοί και οι κατευθυντήριες γραμμές θα διαδραματίσουν κρίσιμο ρόλο στη διασφάλιση της τεχνητής νοημοσύνης προς όφελος όλης της ανθρωπότητας.

Συνεργατική τεχνητή νοημοσύνη:

- Στο μέλλον θα δούμε τους ανθρώπους και την τεχνητή νοημοσύνη να συνεργάζονται πιο στενά, με μοντέλα που αυξάνουν τις ανθρώπινες ικανότητες αντί να τις αντικαθιστούν.
- Έμφαση στα συστήματα ανθρώπινης σύνδεσης, όπου η τεχνητή νοημοσύνη παρέχει πληροφορίες και οι άνθρωποι λαμβάνουν τις τελικές αποφάσεις.

Το ταξίδι των MLP και της βαθιάς μάθησης απέχει πολύ από το να έχει τελειώσει. Καθώς κοιτάζουμε προς το μέλλον, οι δυνατότητες για ανάπτυξη, καινοτομία και θετικό αντίκτυπο είναι τεράστιες.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Εργαλεία Βιοπληροφορικής

Θέμα 4. Τεχνητά νευρωνικά δίκτυα

Μάθημα 4. Deep Learning with Python

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή στη Deep Learning στην Python

Απελευθερώνοντας τη δύναμη των νευρωνικών δικτύων

Τι είναι το Deep Learning;

- Ένα υποσύνολο μηχανικής μάθησης που χρησιμοποιεί νευρωνικά δίκτυα με πολλά επίπεδα (άρα "βαθιά") για την ανάλυση διαφόρων παραγόντων δεδομένων.
- Δυνατότητα αυτόματης εκμάθησης αναπαραστάσεων δεδομένων χωρίς χειροκίνητη εξαγωγή χαρακτηριστικών.

Ο ρόλος της Python στη βαθιά μάθηση:

- Η Python έχει αναδειχθεί ως η κορυφαία γλώσσα για τη βαθιά μάθηση λόγω της **απλότητας** , **της αναγνωσιμότητας** και **των εκτεταμένων βιβλιοθηκών της** που είναι προσαρμοσμένες για την επιστήμη δεδομένων και την τεχνητή νοημοσύνη.
- Το τεράστιο οικοσύστημα της Python, σε συνδυασμό με τη δυναμική κοινότητά του, έχει οδηγήσει στην ανάπτυξη ισχυρών πλαισίων όπως το TensorFlow, το Keras και το PyTorch.

Επισκόπηση μαθήματος:

- Βουτήξτε στις βασικές έννοιες της βαθιάς μάθησης χρησιμοποιώντας Python.
- Εξερευνήστε δημοφιλή πλαίσια, χειρισμό δεδομένων, οπτικοποίηση, αξιολόγηση μοντέλων και ανάπτυξη.

Με την ευελιξία της Python και τη δύναμη της βαθιάς μάθησης, μπορούμε να λύσουμε σύνθετα προβλήματα, που κυμαίνονται από την αναγνώριση εικόνας έως την επεξεργασία φυσικής γλώσσας και όχι μόνο.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Η κυριαρχία της Python στη βαθιά μάθηση

Απλότητα και αναγνωσιμότητα:

- Η σύνταξη της Python είναι σαφής και διαισθητική, καθιστώντας την προσβάσιμη για αρχάριους και αποτελεσματική για τους ειδικούς.
- Επιτρέπει την ταχεία δημιουργία πρωτοτύπων και επαναληπτική ανάπτυξη.

Εκτεταμένες βιβλιοθήκες:

- Η Python διαθέτει ένα πλούσιο οικοσύστημα βιβλιοθηκών προσαρμοσμένων για την επιστήμη δεδομένων, τη μηχανική μάθηση και τη βαθιά μάθηση (π.χ. **TensorFlow, Keras, Pandas, NumPy**).
- Αυτές οι βιβλιοθήκες απλοποιούν πολύπλοκες εργασίες, από την προεπεξεργασία δεδομένων έως την εκπαίδευση μοντέλων.

Υποστηρικτική Κοινότητα:

- Μια δυναμική και αναπτυσσόμενη κοινότητα που συμβάλλει σε έργα ανοιχτού κώδικα, μοιράζεται γνώσεις και παρέχει υποστήριξη μέσω φόρουμ όπως το *Stack Overflow* .
- Τακτικές ενημερώσεις, σεμινάρια και τεκμηρίωση για πλαίσια βαθιάς μάθησης .

Διαλειτουργικότητα:

- Η Python ενσωματώνεται απρόσκοπτα με άλλες γλώσσες όπως η C, η C++ και η Java, επιτρέποντας τη βελτιστοποίηση της απόδοσης και τη χρήση του παλαιού κώδικα.
- Προσφέρει δεσίματα σε άλλες πλατφόρμες και εργαλεία, ενισχύοντας τη χρησιμότητά του.

Πλατφόρμα Agnostic:

- Η Python είναι cross-platform, δηλαδή τρέχει σε διάφορα λειτουργικά συστήματα χωρίς τροποποίηση.
- Διευκολύνει την ανάπτυξη και την κλιμάκωση εφαρμογών βαθιάς μάθησης.

Ο συνδυασμός της απλότητας, των ισχυρών βιβλιοθηκών, της υποστήριξης της κοινότητας και της ευελιξίας της Python την καθιστά την κορυφαία επιλογή για προσπάθειες βαθιάς μάθησης.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ρύθμιση του περιβάλλοντος

Εγκατάσταση Python:

- Βεβαιωθείτε ότι έχετε εγκαταστήσει την πιο πρόσφατη έκδοση της Python.
- Χρησιμοποιήστε επίσημες διανομές Python ή εργαλεία όπως το Anaconda για μια ολοκληρωμένη ρύθμιση της επιστήμης δεδομένων.

Εικονικά περιβάλλοντα:

- Χρησιμοποιήστε το venv ή το conda για να δημιουργήσετε **απομονωμένα περιβάλλοντα** για διαφορετικά έργα.
- Βοηθά στη διαχείριση των εξαρτήσεων και στην αποφυγή διενέξεων εκδόσεων .

Βασικές βιβλιοθήκες και πλαίσια:

- Εγκαταστήστε βασικές βιβλιοθήκες: **NumPy, Pandas, Matplotlib** για χειρισμό και οπτικοποίηση δεδομένων.
- Πλαίσια βαθιάς μάθησης: **TensorFlow, Keras, PyTorch** . Επιλέξτε με βάση τις ανάγκες του έργου.

Ολοκληρωμένα Περιβάλλοντα Ανάπτυξης (IDEs):

- Εργαλεία όπως το Jupyter Notebook, το PyCharm ή το Visual Studio Code βελτιώνουν την αποτελεσματικότητα της κωδικοποίησης.
- **Το Jupyter** είναι ιδιαίτερα δημοφιλές για διαδραστική ανάλυση δεδομένων και εκπαίδευση μοντέλων.

Ρύθμιση GPU (προαιρετικό αλλά συνιστάται):

- Για εντατικούς υπολογισμούς, η ρύθμιση μιας GPU μπορεί να μειώσει δραστικά τους χρόνους εκπαίδευσης.
- Βιβλιοθήκες όπως το CUDA και το cuDNN βοηθούν στην αξιοποίηση των GPU της NVIDIA με το TensorFlow ή το PyTorch.

Δοκιμή της εγκατάστασης:

- Εκτελέστε ένα απλό σενάριο βαθιάς εκμάθησης ή σημειωματάριο για να διασφαλίσετε ότι όλες οι εγκαταστάσεις λειτουργούν απρόσκοπτα. Επαληθεύστε τη χρήση της GPU εάν έχει ρυθμιστεί.

Ένα καλά διαμορφωμένο περιβάλλον είναι ζωτικής σημασίας για την αποτελεσματική ανάπτυξη βαθιάς μάθησης. Βεβαιωθείτε ότι όλα τα εργαλεία είναι σωστά εγκατεστημένα και ενσωματωμένα.

Εισαγωγή δεδομένων σε Python

1. **Πηγές Δεδομένων:**
   - Τοπικά αρχεία (βάσεις δεδομένων CSV, Excel, SQL).
   - Πηγές Ιστού (API, web scraping).
   - Cloud storage (AWS S3, Google Cloud Storage).

2. **Χρήση Pandas για εισαγωγή δεδομένων:**
   - **CSV:** `data = pd.read_csv('path_to_file.csv')`
   - **Προέχω:** `data = pd.read_excel('path_to_file.xlsx')`
   - **SQL:**

```
import sqlite3

conn = sqlite3.connect('database.db')

data = pd.read_sql_query('SELECT * FROMtable_name', conn)
```

1. **Προετοιμασία δεδομένων για διαχωρισμό:**
   - ➔ **Χαρακτηριστικά (X):** `X = data.drop('target_column', axis=1)`
   - ➔ **Στόχος (y):** `y = δεδομένα['target_column']`

Διαχωρισμός δεδομένων - Σύνολα προπονήσεων και δοκιμών

1. **Γιατί Σπλιτ;**
   - Εκπαιδεύστε το μοντέλο σε ένα υποσύνολο (**σετ εκπαίδευσης**).
   - Αξιολογήστε την απόδοση του μοντέλου σε ένα άλλο υποσύνολο (**σετ δοκιμής**).

2. **Χρησιμοποιώντας** το train_test_split **από το Scikit-Learn:**
   python

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split( X, y, test_size= 0,2 )
```

3. **Θεωρήσεις:**
   - **Διαστρωμάτωση:** Διασφάλιση ότι η αμαξοστοιχία και τα σετ δοκιμών έχουν παρόμοιες αναλογίες ετικετών κλάσης.
   - **Τυχαία κατάσταση:** Για αναπαραγωγιμότητα των αποτελεσμάτων.
   - **Μέγεθος δοκιμής:** Συνήθως μεταξύ 20% - 30% των δεδομένων.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή στο TensorFlow

- **Τι είναι το TensorFlow;**
  - Ένα πλαίσιο βαθιάς μάθησης ανοιχτού κώδικα που αναπτύχθηκε από την Google Brain.
  - Επιτρέπει τη δημιουργία, την εκπαίδευση και την ανάπτυξη μοντέλων μηχανικής εκμάθησης με ευκολία.
- **Βασικά χαρακτηριστικά:**
  - **Υπολογιστικά γραφήματα:** Αντιπροσωπεύει τους υπολογισμούς ως κατευθυνόμενα γραφήματα, επιτρέποντας την αποτελεσματική παράλληλη επεξεργασία.
  - **Λειτουργίες τανυστή:** Χειρίζεται πολυδιάστατους πίνακες δεδομένων (τανυστές) για πράξεις όπως πολλαπλασιασμούς πινάκων.
  - **Αυτόματη διαφοροποίηση:** Υπολογίζει αυτόματα τις κλίσεις, απλοποιώντας τις εργασίες βελτιστοποίησης.
- **Ευελιξία και επεκτασιμότητα:**
  - Υποστηρίζει ένα ευρύ φάσμα αρχιτεκτονικών νευρωνικών δικτύων.
  - Κλιμακώνεται απρόσκοπτα από μία μόνο CPU σε πολλές GPU, ακόμη και TPU.
- **TensorBoard:**
  - Ένα εργαλείο οπτικοποίησης που συνοδεύεται από TensorFlow.
  - Επιτρέπει την παρακολούθηση της εκπαίδευσης μοντέλων, την οπτικοποίηση υπολογιστικών γραφημάτων και την ανάλυση μετρήσεων.
- **Ενσωμάτωση με Keras:**
  - Το TensorFlow 2.x ενσωμάτωσε το Keras ως το API υψηλού επιπέδου του, κάνοντας τη δημιουργία μοντέλων πιο διαισθητική.
  - Προσφέρει προκαθορισμένα επίπεδα, βελτιστοποιητές και λειτουργίες απώλειας.
- **Οικοσύστημα και Κοινότητα:**
  - Ένα τεράστιο οικοσύστημα εργαλείων, βιβλιοθηκών και επεκτάσεων (π.χ. TF Lite, TF.js, TF Hub).
  - Ισχυρή υποστήριξη της κοινότητας με τακτικές ενημερώσεις, σεμινάρια και πόρους.

TensorFlow

Το TensorFlow παρέχει μια ολοκληρωμένη πλατφόρμα για τη δημιουργία και την ανάπτυξη εξελιγμένων μοντέλων μηχανικής μάθησης, που υποστηρίζονται από ένα ισχυρό οικοσύστημα και κοινότητα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Δημιουργία απλού μοντέλου στο TensorFlow (1)

```
import tensorflow as tf
from tensorflow import keras

model = keras.Sequential([
    keras.layers.Dense(128, activation='relu', input_shape=(784,)),
    keras.layers.Dense(10, activation='softmax')
])

model.compile(optimizer='adam',
              loss='sparse_categorical_crossentropy',
              metrics=['accuracy'])
model.fit(train_data, train_labels, epochs=10)
```

- Εισαγάγετε το TensorFlow και άλλες απαραίτητες βιβλιοθήκες.
- Καθορισμός του μοντέλου: Χρησιμοποιήστε το Sequential API για στοίβαξη επιπέδων.
- Μεταγλώττιση του μοντέλου: Καθορίστε το εργαλείο βελτιστοποίησης, τη συνάρτηση απώλειας και τις μετρήσεις.
- Εκπαιδεύστε το μοντέλο

Το TensorFlow παρέχει μια ολοκληρωμένη πλατφόρμα για τη δημιουργία και την ανάπτυξη εξελιγμένων μοντέλων μηχανικής μάθησης, που υποστηρίζονται από ένα ισχυρό οικοσύστημα και κοινότητα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Δημιουργία απλού μοντέλου στο TensorFlow (2)

Αξιολόγηση και Πρόβλεψη:

```
test_loss, test_acc = model.evaluate(test_data, test_labels)
predictions = model.predict(new_data)
```

Για λόγους απλότητας, παραλείψαμε την αρχική φάση, όπου τα δεδομένα εισάγονται και στη συνέχεια χωρίζονται σε σύνολο εκπαίδευσης και σετ δοκιμής .

- Αξιολογήστε την απόδοση του μοντέλου σε δεδομένα δοκιμής
- Κάνε προβλέψεις .

```
model.save('model_path.h5')
new_model = tf.keras.models.load_model('model_path.h5')
```

Αποθηκεύστε ή φορτώστε το μοντέλο για μελλοντική χρήση ή ανάπτυξη

Μια συνοπτική περιγραφή της οικοδόμησης ενός βασικού μοντέλου νευρωνικών δικτύων στο TensorFlow, από τον καθορισμό της αρχιτεκτονικής έως την εκπαίδευση και την αξιολόγηση. Τα αποσπάσματα κώδικα προσφέρουν μια πρακτική ματιά στη διαδικασία.

Εισαγωγή στον Κέρα

Τι είναι το Keras;

- Ένα API νευρωνικών δικτύων υψηλού επιπέδου, γραμμένο σε Python.
- Σχεδιασμένο για να επιτρέπει γρήγορο πειραματισμό με βαθιά νευρωνικά δίκτυα.
- Εκτελείται πάνω από το TensorFlow, το Theano ή το CNTK.

Γιατί Κεράς;

- **Απλότητα:** Διαισθητικό και φιλικό προς το χρήστη API.
- **Ευελιξία:** Υποστηρίζει τόσο συνελικτικά δίκτυα όσο και επαναλαμβανόμενα δίκτυα, καθώς και συνδυασμούς και των δύο.
- **Modularity:** Τα μοντέλα κατασκευάζονται χρησιμοποιώντας διαμορφώσιμα δομικά στοιχεία (στρώσεις, βελτιστοποιητές, συναρτήσεις ενεργοποίησης).
- **Ενσωμάτωση:** Ενσωματώνεται απρόσκοπτα με το TensorFlow, επιτρέποντας προηγμένες λειτουργίες και βελτιστοποιήσεις.

Βασικά χαρακτηριστικά:

- **Προεπεξεργασμένα σύνολα δεδομένων:** Έρχεται με ενσωματωμένα σύνολα δεδομένων όπως CIFAR-10, MNIST κ.λπ.
- **Προεκπαιδευμένα μοντέλα:** Πρόσβαση σε μοντέλα όπως το VGG, το ResNet και άλλα για μεταφορά εκμάθησης.
- **Επανακλήσεις:** Σύνολο λειτουργιών που εφαρμόζονται κατά τη διάρκεια της εκπαίδευσης για την παρακολούθηση της προόδου ή την ενεργοποίηση ενεργειών.

Κατασκευάζοντας ένα απλό μοντέλο στον Κέρα (1)

1. **Sequential Model:**
   - The simplest type of model, a linear stack of layers.

python Copy code

```
model = keras.Sequential()
```

2. **Adding Layers:**
   - Easily add layers using the `add()` method.

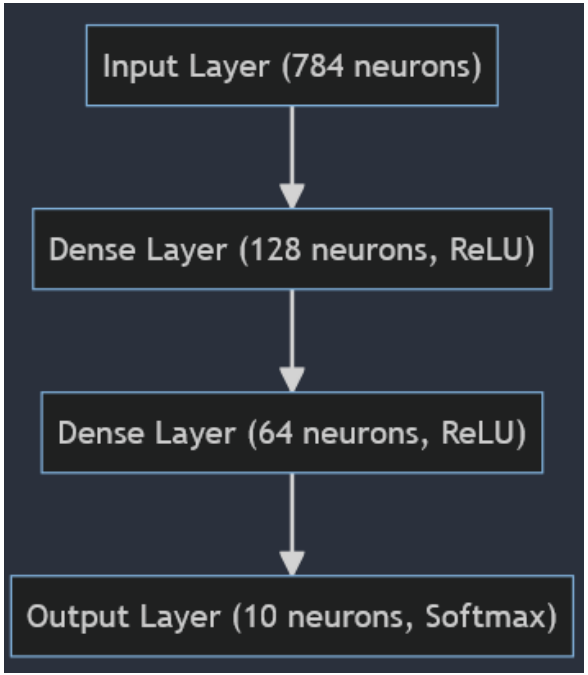
python Copy code

```
model.add(Dense(128, activation='relu', input_shape=(784,)))
model.add(Dense(64, activation='relu'))
model.add(Dense(10, activation='softmax'))
```

3. **Configuring the Learning Process:**
   - Once the model is defined, configure its learning process with `compile()`.

python Copy code

```
model.compile(optimizer='adam',
              loss='categorical_crossentropy',
              metrics=['accuracy'])
```



Co-funded by the Erasmus+ Programme of the European Union

Κατασκευάζοντας ένα απλό μοντέλο στον Κέρα (2)

4. **Training the Model:**
   - Use the `fit()` method to train the model on data.

```python
model.fit(x_train, y_train, epochs=10, batch_size=32)
```

5. **Evaluation and Prediction:**
   - Assess the model's performance and make predictions.

```python
loss, accuracy = model.evaluate(x_test, y_test)
predictions = model.predict(new_data)
```

Οπτικοποίηση με Matplotlib και Seaborn

1. **Γιατί Οπτικοποίηση;**
   - Κατανοήστε τη διανομή δεδομένων.
   - Προσδιορίστε μοτίβα και ακραίες τιμές.
   - Λάβετε τεκμηριωμένες αποφάσεις κατά την κατασκευή του μοντέλου.
2. **Matplotlib:**
   - Βιβλιοθήκη βασικής σχεδίασης σε Python.
   - Ευέλικτο και παρέχει ένα βασικό επίπεδο για πολλές άλλες βιβλιοθήκες οπτικοποίησης.
   - Παράδειγμα:

```
import matplotlib.pyplot as plt
plt.plot( x, y )
plt.xlabel(' άξονας X ')
plt.ylabel(' άξονας Y ')
plt.title(' Samp Plot ')
plt.show()
```

3. **Seaborn:**
   - Χτισμένο στην κορυφή του Matplotlib.
   - Παρέχει μια διεπαφή υψηλού επιπέδου για τη σχεδίαση ελκυστικών στατιστικών γραφικών.
   - Έρχεται με πολλά ενσωματωμένα θέματα και χρωματικές παλέτες για να κάνει τα σχέδια πιο ελκυστικά οπτικά.
   - Παράδειγμα:

```
import seaborn as sns
sns.histplot( data , bins= 30 , kde=True
)
```

Αξιολόγηση μοντέλου (1)

Γιατί Αξιολόγηση Μοντέλου;

- Εξασφαλίστε τη γενίκευση του μοντέλου σε μη ορατά δεδομένα.
- Συγκρίνετε την απόδοση διαφορετικών μοντέλων.
- Προσδιορίστε τομείς βελτίωσης.

Κοινές μετρήσεις ταξινόμησης:

- **Ακρίβεια:** Συνολική ορθότητα του μοντέλου.
- **Precision:** Ορθότητα θετικών προβλέψεων.
- **Ανάκληση (Ευαισθησία):** Δυνατότητα ανίχνευσης όλων των θετικών περιπτώσεων.
- **F1-Score:** Αρμονική μέση ακρίβεια και ανάκληση.

Μήτρα σύγχυσης:

Ένας πίνακας που χρησιμοποιείται για την αξιολόγηση της απόδοσης ενός μοντέλου ταξινόμησης. Δείχνει αληθινά θετικά, αληθινά αρνητικά, ψευδώς θετικά και ψευδώς αρνητικά.

		Predicted: Positive	Predicted: Negative
Ground-Truth	Positive	True Positive	False Negative
Ground-Truth	Negative	False Positive	True Negative

Η σωστή αξιολόγηση του μοντέλου είναι ζωτικής σημασίας για την κατανόηση των δυνατών και των αδυναμιών του, για την καθοδήγηση περαιτέρω βελτίωσης και για τη διασφάλιση της αξιοπιστίας του σε εφαρμογές του πραγματικού κόσμου.

Αξιολόγηση μοντέλου (2)

Συνήθεις μετρήσεις παλινδρόμησης:

- **Μέσο απόλυτο σφάλμα (MAE):** Μέσος όρος απόλυτων διαφορών μεταξύ προβλέψεων και πραγματικών τιμών.
- **Μέσο τετράγωνο σφάλμα (MSE):** Μέσος όρος τετραγωνικών διαφορών.
- **R-τετράγωνο:** Αναλογία διακύμανσης που εξηγείται από το μοντέλο.

Μετρήσεις παλινδρόμησης στο Scikit-learn:

```
from το sklearn.metrics import mean_absolute_error, mean_squared_error,
r2_score
y_pred = model.predict( X_test )
print(" MAE :", mean_absolute_error( y_test, y_pred ))
print(" MSE :", mean_squared_error( y_test, y_pred ))
print(" R2 Score: ", r2_score( y_test, y_pred ))
```

Η Python, με το πλούσιο οικοσύστημα βιβλιοθηκών της, προσφέρει έναν απλό και αποτελεσματικό τρόπο αξιολόγησης μοντέλων μηχανικής μάθησης και βαθιάς μάθησης.

Αξιολόγηση μοντέλου (3)

Μετρήσεις ταξινόμησης στο Scikit-learn:

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
y_pred = model.predict( X_test )
print(" Ακρίβεια: ", accuracy_score( y_test, y_pred ))
print(" Precision: ", precision_score( y_test, y_pred ))
print(" Ανάκληση: ", recall_score( y_test, y_pred ))
print(" F1-Score: ", f1_score( y_test, y_pred ))
```

Αξιολόγηση μοντέλων Keras:

```
loss, accuracy = model.evaluate( X_test, y_test )
print (" Απώλεια: ", loss )
print(" Ακρίβεια: ", accuracy )
```

Η Python, με το πλούσιο οικοσύστημα βιβλιοθηκών της, προσφέρει έναν απλό και αποτελεσματικό τρόπο αξιολόγησης μοντέλων μηχανικής μάθησης και βαθιάς μάθησης.

Αποθήκευση και φόρτωση (1)

Γιατί να αποθηκεύσετε και να φορτώσετε μοντέλα;

- **Αποδοτικότητα:** Αποφύγετε την επανεκπαίδευση από την αρχή.
- **Ανάπτυξη:** Χρησιμοποιήστε εκπαιδευμένα μοντέλα σε περιβάλλοντα παραγωγής.
- **Αναπαραγωγιμότητα:** Εξασφαλίστε συνεπή αποτελέσματα σε διαφορετικές πλατφόρμες και χρόνους.

Αποθήκευση και φόρτωση με το Scikit-learn:

- **Pickle:**

```
Import pickle
# Αποθηκεύσετε
with open(' model.pkl ', 'wb') as file:
pickle.dump ( model , file )
# Φορτώνω
with open(' model.pkl ', 'rb') ως file :
    loaded_model = pickle.load( file )
```

- **Joblib (συνιστάται για μεγαλύτερα μοντέλα) :**

```
from joblib import dump, load
# Αποθηκεύσετε
dump( model , ' model.joblib ')
# Φορτώνω
loaded_model = load(' model.joblib ')
```

Η αποθήκευση και η φόρτωση μοντέλων είναι ζωτικής σημασίας για πρακτικές ροές εργασίας μηχανικής εκμάθησης, επιτρέποντας απρόσκοπτες μεταβάσεις μεταξύ των φάσεων εκπαίδευσης, αξιολόγησης και ανάπτυξης.

Αποθήκευση και φόρτωση (2)

Αποθήκευση και φόρτωση με Keras:

```
# Αποθηκεύστε ολόκληρο το μοντέλο (αρχιτεκτονική, βάρη,
# βελτιστοποιητής)
model.save('model.h5')

# Φορτώστε το μοντέλο
from κεράς.models import load_model
loaded_model = load_model('model.h5')
```

- Διασφάλιση συμβατότητας : Τα μοντέλα που είναι αποθηκευμένα σε μια έκδοση μιας βιβλιοθήκης ενδέχεται να μην είναι συμβατά με μια άλλη έκδοση.
- Ασφάλεια : Να είστε προσεκτικοί όταν φορτώνετε μοντέλα από μη αξιόπιστες πηγές

Όταν φορτώνετε το μοντέλο χρησιμοποιώντας τις αντίστοιχες μεθόδους, θα είναι στην **ακριβή κατάσταση** που ήταν κατά την αποθήκευση, συμπεριλαμβανομένης της αρχιτεκτονικής του και των εκπαιδευμένων βαρών .

Αυτό διασφαλίζει ότι μπορείτε να χρησιμοποιήσετε αμέσως το φορτωμένο μοντέλο για προβλέψεις, περαιτέρω εκπαίδευση ή αξιολόγηση χωρίς να χρειάζεται να το εκπαιδεύσετε ξανά.

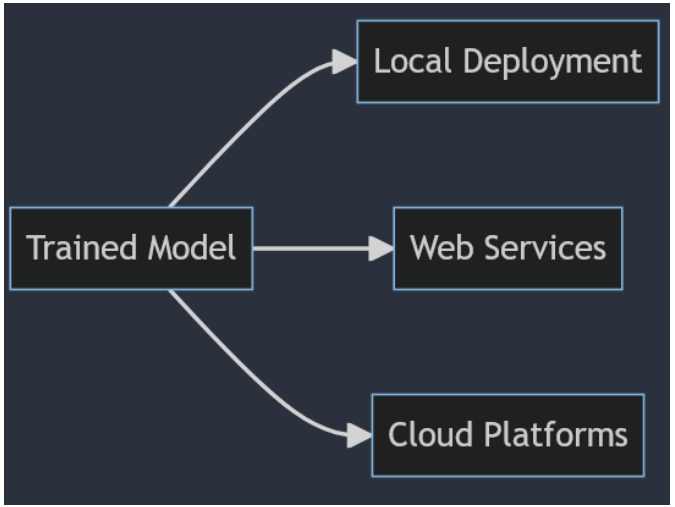
Η αποθήκευση και η φόρτωση μοντέλων είναι ζωτικής σημασίας για πρακτικές ροές εργασίας μηχανικής εκμάθησης, επιτρέποντας απρόσκοπτες μεταβάσεις μεταξύ των φάσεων εκπαίδευσης, αξιολόγησης και ανάπτυξης.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ανάπτυξη μοντέλων Deep Learning

1. **Τι είναι η ανάπτυξη;**
   - Μετάβαση ενός εκπαιδευμένου μοντέλου σε μια εφαρμογή όπου μπορεί να λάβει νέα δεδομένα και να κάνει προβλέψεις σε πραγματικό χρόνο.
2. **Γιατί είναι σημαντική η ανάπτυξη;**
   - **Χρήση:** Η πραγματική αξία ενός εκπαιδευμένου μοντέλου γίνεται αντιληπτή όταν χρησιμοποιείται σε πραγματικές εφαρμογές.
   - **Προσβασιμότητα:** Επιτρέπει στους τελικούς χρήστες, που μπορεί να μην είναι εξοικειωμένοι με την ML, να επωφεληθούν από τις δυνατότητες του μοντέλου.
3. **Επιλογές ανάπτυξης στην Python:**
   - **Τοπική ανάπτυξη:** Χρήση του μοντέλου σε τοπικές εφαρμογές ή σενάρια.
   - **Υπηρεσίες Ιστού:** Ανάπτυξη του μοντέλου ως API χρησιμοποιώντας πλαίσια όπως το Flask ή το FastAPI.
   - **Πλατφόρμες Cloud:** Αξιοποιήστε πλατφόρμες όπως το AWS SageMaker, το Google AI Platform ή το Azure Machine Learning για επεκτάσιμες αναπτύξεις.

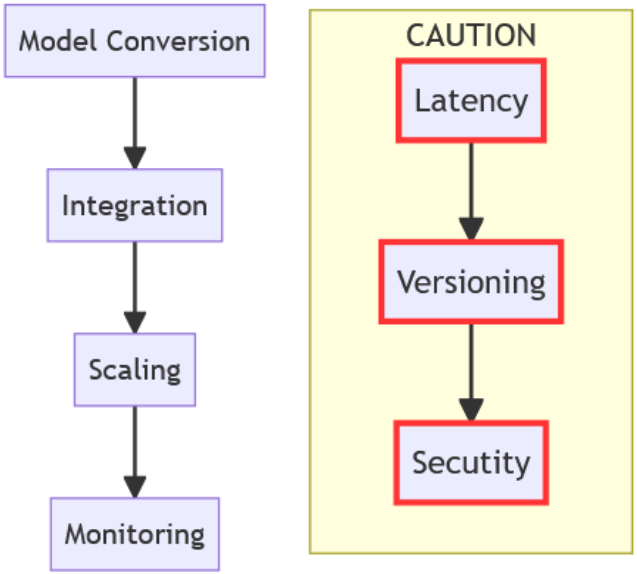


Η ανάπτυξη γεφυρώνει το χάσμα μεταξύ της ανάπτυξης μοντέλων και του αντίκτυπου στον πραγματικό κόσμο, καθιστώντας την μια κρίσιμη φάση στον κύκλο ζωής της μηχανικής μάθησης.

Βήματα και προκλήσεις ανάπτυξης

Βήματα για την ανάπτυξη:

1. **Μετατροπή μοντέλου:** Μετατρέψτε το μοντέλο σε μορφή κατάλληλη για ανάπτυξη (π.χ. TensorFlow Serving, ONNX).
2. **Ενσωμάτωση:** Ενσωματώστε το μοντέλο στην επιθυμητή εφαρμογή ή πλατφόρμα.
3. **Κλιμάκωση:** Βεβαιωθείτε ότι η ανάπτυξη μπορεί να χειριστεί τον αναμενόμενο αριθμό αιτημάτων.
4. **Παρακολούθηση:** Παρακολουθήστε την απόδοση και την υγεία του μοντέλου στο αναπτυσσόμενο περιβάλλον.



Προκλήσεις στην ανάπτυξη:

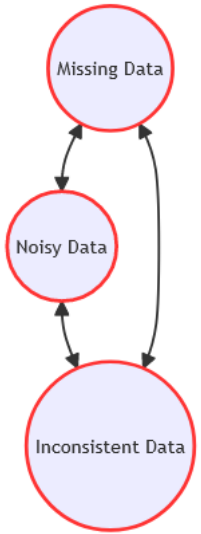
- **Latency:** Εξασφάλιση προβλέψεων σε πραγματικό χρόνο με ελάχιστη καθυστέρηση.
- **Εκδόσεις:** Διαχείριση ενημερώσεων στο μοντέλο χωρίς διακοπή της υπηρεσίας.
- **Ασφάλεια:** Προστασία του μοντέλου και των δεδομένων από πιθανές απειλές.

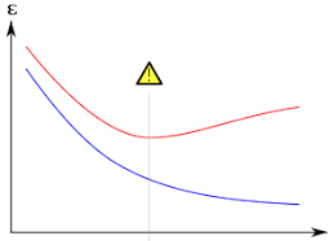
Η ανάπτυξη γεφυρώνει το χάσμα μεταξύ της ανάπτυξης μοντέλων και του αντίκτυπου στον πραγματικό κόσμο, καθιστώντας την μια κρίσιμη φάση στον κύκλο ζωής της μηχανικής μάθησης.

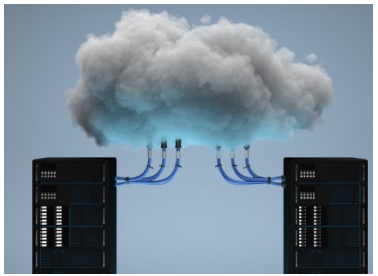
Κοινές προκλήσεις στη βαθιά μάθηση

Κοινές Προκλήσεις

1. **Υπερπροσαρμογή:** Όταν ένα μοντέλο έχει καλή απόδοση στα δεδομένα προπόνησης αλλά κακή σε αόρατα δεδομένα.
2. **Ποιότητα δεδομένων:** Τα ανακριβή ή ασυνεπή δεδομένα μπορεί να οδηγήσουν σε κακή απόδοση του μοντέλου.
3. **Υπολογιστικοί Πόροι:** Τα μοντέλα βαθιάς μάθησης, ειδικά τα πολύπλοκα, απαιτούν σημαντική υπολογιστική ισχύ και μνήμη.







Η υπερβολική προσαρμογή, η κακή ποιότητα δεδομένων και οι υπολογιστικές απαιτήσεις είναι πρωταρχικές προκλήσεις στη βαθιά μάθηση. Η διασφάλιση της ακεραιότητας των δεδομένων και η ύπαρξη επαρκών υπολογιστικών πόρων είναι ζωτικής σημασίας για την επιτυχία του μοντέλου.

Βέλτιστες πρακτικές για Deep Learning στην Python

1. **Τακτοποίηση:** Τεχνικές όπως η εγκατάλειψη ή η τακτοποίηση L1/L2 μπορούν να βοηθήσουν στην αποφυγή υπερβολικής προσαρμογής.
2. **Επαύξηση δεδομένων:** Βελτίωση του συνόλου δεδομένων εκπαίδευσης με την εφαρμογή τυχαίων τροποποιήσεων. Ιδιαίτερα χρήσιμο για δεδομένα εικόνας.
3. **Πρόωρη διακοπή:** Παρακολουθήστε την απόδοση επικύρωσης και σταματήστε την προπόνηση μόλις αρχίσει να υποβαθμίζεται.
4. **Χρησιμοποιήστε προεκπαιδευμένα μοντέλα:** Η αξιοποίηση μοντέλων που έχουν ήδη εκπαιδευτεί σε μεγάλα σύνολα δεδομένων μπορεί να εξοικονομήσει χρόνο και πόρους.

Τεχνικές όπως η τακτοποίηση, η αύξηση δεδομένων και η πρόωρη διακοπή βελτιώνουν την ευρωστία και την αποτελεσματικότητα του μοντέλου. Η χρήση προεκπαιδευμένων μοντέλων μπορεί να επιταχύνει την εκπαίδευση και να βελτιώσει την απόδοση, ειδικά με περιορισμένα δεδομένα.

Συμπέρασμα και Πόροι

1. **Συμπέρασμα:**
   - Το Deep Learning στην Python προσφέρει μια ισχυρή εργαλειοθήκη για τη δημιουργία εξελιγμένων μοντέλων.
   - Ενώ υπάρχουν προκλήσεις, η τήρηση βέλτιστων πρακτικών διασφαλίζει ισχυρά και αποτελεσματικά μοντέλα.
   - Η συνεχής μάθηση και η ενημέρωση με τις πιο πρόσφατες τεχνικές είναι το κλειδί για την επιτυχία στον συνεχώς εξελισσόμενο τομέα της βαθιάς μάθησης.
2. **Πηγές για περαιτέρω μάθηση:**
   - **Βιβλία:**
     i. «Deep Learning» των Ian Goodfellow, Yoshua Bengio και Aaron Courville.
     ii. «Python Deep Learning» των Ivan Vasilev και Daniel Slater.
   - **Διαδικτυακά Μαθήματα:**
     i. Coursera "Deep Learning Specialization" του Andrew Ng.
     ii. Το "Deep Learning Nanodegree" του Udacity.
   - **Ιστότοποι:**
     i. Επίσημη Τεκμηρίωση TensorFlow.
     ii. Επίσημη Τεκμηρίωση Keras.
     iii. Εγχειρίδιο Python Data Science Handbook από τον Jake VanderPlas (διαθέσιμο στο διαδίκτυο).
3. **Κοινότητες και φόρουμ:**
   - Υπερχείλιση στοίβας (ετικέτες Deep Learning).
   - Reddit r/MachineLearning.
   - Φόρουμ κοινότητας TensorFlow.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Εργαλεία Βιοπληροφορικής

Θέμα 4. Τεχνητά νευρωνικά δίκτυα

Μάθημα 5. Συνελικτικά νευρωνικά δίκτυα

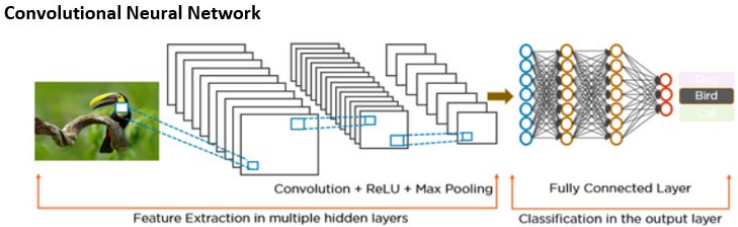
ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή στα Συνελικτικά Νευρωνικά Δίκτυα
(CNN)

- **Ορισμός:**
  - Ένας τύπος βαθιού νευρωνικού δικτύου που έχει σχεδιαστεί για να αναγνωρίζει μοτίβα σε δεδομένα, ειδικά οπτικά δεδομένα.
- **Βασικά χαρακτηριστικά:**
  - Δομή ιεραρχικού στρώματος.
  - Δυνατότητα αυτόματης και προσαρμοστικής εκμάθησης χωρικών ιεραρχιών χαρακτηριστικών.
- **Σημασια:**
  - Επανάσταση στον τομέα της όρασης υπολογιστών.
  - Εξουσιοδοτεί την αναγνώριση εικόνων και βίντεο, την αναγνώριση προσώπου και πολλές εφαρμογές πραγματικού κόσμου.



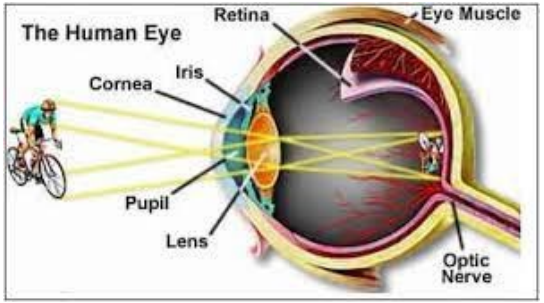
Τα CNN είναι εξειδικευμένα νευρωνικά δίκτυα που έχουν αποδειχθεί εξαιρετικά αποτελεσματικά στην αναγνώριση και ταξινόμηση οπτικών μοτίβων στα δεδομένα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

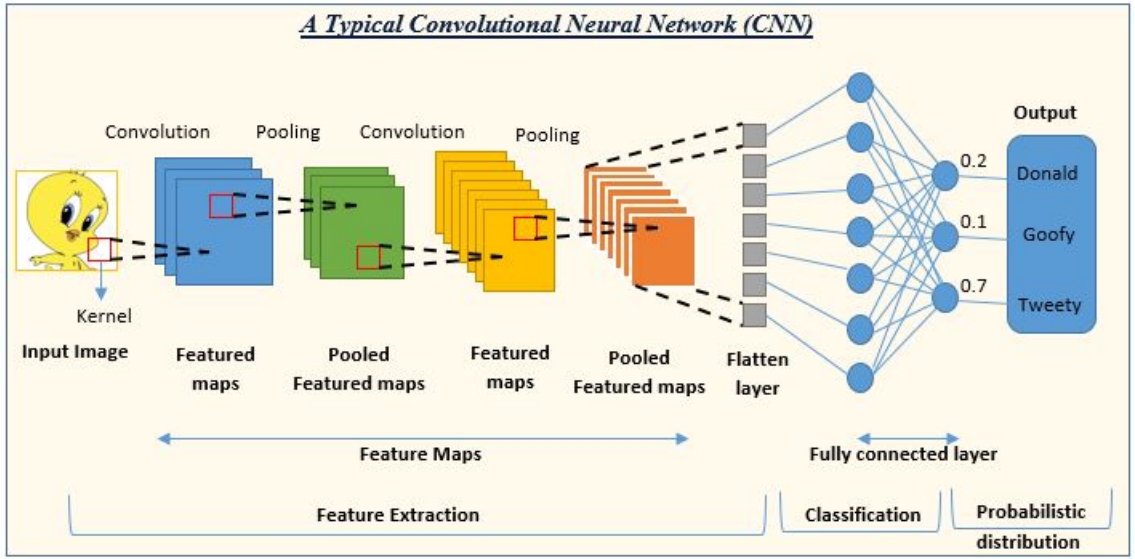
Έμπνευση - Ανθρώπινο Οπτικό Σύστημα

- **Το μάτι ως κάμερα:**
  - Το ανθρώπινο μάτι συλλαμβάνει το φως και το επεξεργάζεται, παρόμοια με το πώς μια φωτογραφική μηχανή συλλαμβάνει μια εικόνα.
- **Επεξεργασία σε επίπεδα:**
  - Οι οπτικές πληροφορίες επεξεργάζονται σε στάδια, από την ανίχνευση απλών χαρακτηριστικών όπως ακμές έως την αναγνώριση πολύπλοκων αντικειμένων.
- **Νευρώνες Εξειδικευμένοι στην Όραση:**
  - Διαφορετικοί νευρώνες στον οπτικό φλοιό ανταποκρίνονται σε διαφορετικά οπτικά ερεθίσματα, όπως χρώματα, σχήματα και κινήσεις.
- **Ιεραρχική Επεξεργασία:**
  - Τα χαμηλότερα επίπεδα ανιχνεύουν απλά χαρακτηριστικά, ενώ τα υψηλότερα επίπεδα συνδυάζουν αυτά τα χαρακτηριστικά για να αναγνωρίσουν πιο περίπλοκα μοτίβα.



Ο σχεδιασμός των CNN είναι εμπνευσμένος από την ιεραρχική και πολυεπίπεδη επεξεργασία του ανθρώπινου οπτικού συστήματος, όπου διαφορετικά στρώματα νευρώνων επεξεργάζονται διαφορετικές πτυχές οπτικών πληροφοριών.

Βασική δομή ενός CNN (1)



Τα CNN αποτελούνται από μια σειρά επιπέδων, το καθένα σχεδιασμένο να επεξεργάζεται και να μετασχηματίζει τα δεδομένα εισόδου με συγκεκριμένο τρόπο, οδηγώντας τελικά σε ακριβή ταξινόμηση εικόνων.

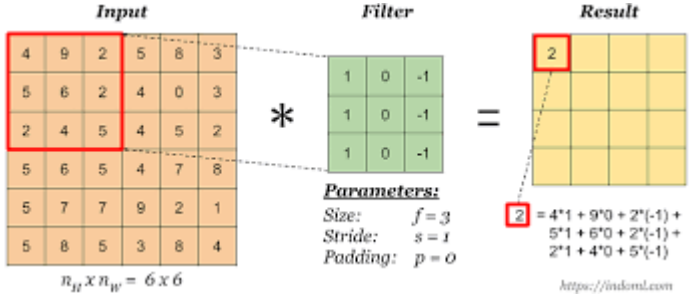
Βασική δομή ενός CNN (2)

- **Επίπεδο εισόδου:**
  - Λαμβάνει τις μη επεξεργασμένες τιμές pixel της εικόνας.
- **Συνελικτικά στρώματα:**
  - Εφαρμόστε συνελικτικά φίλτρα στα δεδομένα εισόδου για να εντοπίσετε χαρακτηριστικά όπως άκρες, υφές και μοτίβα.
  - Κάθε φίλτρο εστιάζει σε ένα συγκεκριμένο χαρακτηριστικό και η έξοδος του είναι ένας χάρτης χαρακτηριστικών.
- **Επίπεδα συγκέντρωσης (υποδειγματοληψίας):**
  - Μειώστε τις χωρικές διαστάσεις των χαρτών χαρακτηριστικών, καθιστώντας το δίκτυο λιγότερο υπολογιστικά εντατικό.
  - Συνήθεις μέθοδοι: Μέγιστη συγκέντρωση, Μέση συγκέντρωση.
- **Πλήρως συνδεδεμένα επίπεδα:**
  - Μετά από πολλά επίπεδα συνέλιξης και συγκέντρωσης, τα δεδομένα ισοπεδώνονται και διέρχονται από ένα ή περισσότερα πλήρως συνδεδεμένα στρώματα.
  - Αυτά τα επίπεδα εκτελούν την τελική ταξινόμηση με βάση τα εντοπισμένα χαρακτηριστικά.
- **Επίπεδο εξόδου:**
  - Παράγει το τελικό αποτέλεσμα ταξινόμησης, χρησιμοποιώντας συχνά μια συνάρτηση ενεργοποίησης softmax για να δώσει πιθανότητες για κάθε τάξη.

Τα CNN αποτελούνται από μια σειρά επιπέδων, το καθένα σχεδιασμένο να επεξεργάζεται και να μετασχηματίζει τα δεδομένα εισόδου με συγκεκριμένο τρόπο, οδηγώντας τελικά σε ακριβή ταξινόμηση εικόνων.

Convolutional Layer - Εισαγωγή

- **Ο ρόλος του συνελικτικού επιπέδου:**
  - Η κύρια λειτουργία είναι να ανιχνεύει τοπικά μοτίβα, υφές, άκρες και συγκεκριμένα χαρακτηριστικά από την εικόνα εισόδου.
  - Λειτουργεί σε μια μικρή, τοπική περιοχή των δεδομένων εισόδου (ένα πεδίο υποδοχής), αλλά εκτείνεται σε όλο το βάθος του όγκου εισόδου.
- **Φίλτρα/Πυρήνες:**
  - Μικροί, μαθησιακοί πίνακες βάρους.
  - Σύρετε πάνω από τα δεδομένα εισόδου (όπως ένα συρόμενο παράθυρο) για να δημιουργήσετε έναν χάρτη χαρακτηριστικών.
  - Κάθε φίλτρο είναι υπεύθυνο για τον εντοπισμό ενός συγκεκριμένου είδους χαρακτηριστικού.
- **Χαρακτηριστικοί Χάρτες:**
  - Προκύπτουν χάρτες ενεργοποίησης 2D που δείχνουν πού στην εικόνα εντοπίστηκαν τα χαρακτηριστικά του φίλτρου.
  - Οι υψηλότερες τιμές στον χάρτη χαρακτηριστικών υποδεικνύουν την παρουσία του επιθυμητού χαρακτηριστικού στην αντίστοιχη θέση της εισόδου.



Συνέλιξη σε δράση: Χρήση φίλτρου για τον εντοπισμό χαρακτηριστικών και την παραγωγή ενός χάρτη χαρακτηριστικών.

Το συνελικτικό επίπεδο χρησιμοποιεί φίλτρα για να σαρώσει την εικόνα εισόδου και να δημιουργήσει χάρτες χαρακτηριστικών, καταγράφοντας βασικά μοτίβα και λεπτομέρειες.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

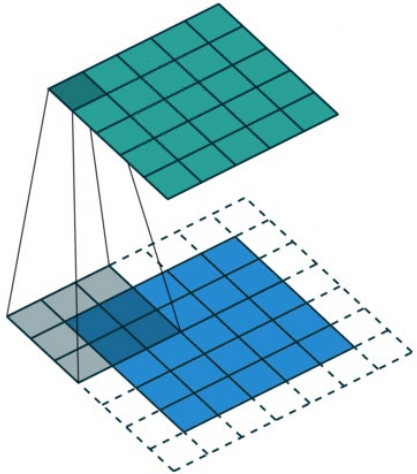
Η Λειτουργία Συνέλιξης

Δρασκελιά:

- Ο αριθμός των εικονοστοιχείων με τα οποία το φίλτρο ολισθαίνει πάνω από τη μήτρα εισόδου.
- Ένας διασκελισμός 1 σημαίνει ότι το φίλτρο μετακινείται ένα εικονοστοιχείο κάθε φορά, ενώ ένας διασκελισμός 2 σημαίνει ότι μεταπηδά δύο εικονοστοιχεία.
- Επηρεάζει το μέγεθος του χάρτη χαρακτηριστικών εξόδου.

Υλικό γεμίσματος:

- Προσθήκη επιπλέον pixel γύρω από την εικόνα εισόδου.
- Βοηθά στη διατήρηση των χωρικών διαστάσεων της εισόδου.
- Δύο τύποι: Έγκυρος (χωρίς γέμιση) και Ίδιος (επένδυση για να διατηρείται το μέγεθος εξόδου ίδιο με την είσοδο).



Η συνέλιξη περιλαμβάνει την ολίσθηση ενός φίλτρου πάνω από την είσοδο, με τον βηματισμό που καθορίζει το μέγεθος του βήματος και την επένδυση που διασφαλίζει τη διατήρηση των χωρικών διαστάσεων.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

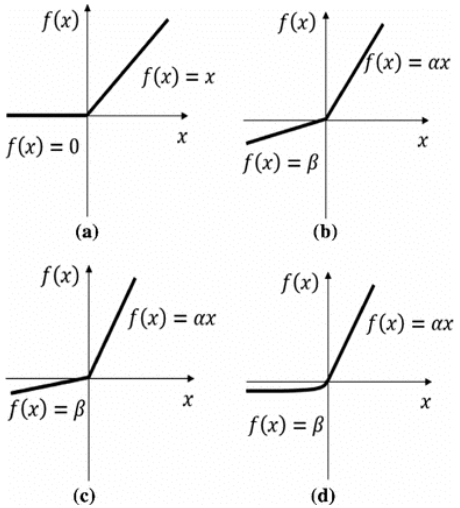
Λειτουργίες ενεργοποίησης σε CNN

ReLU (Διορθωμένη Γραμμική Μονάδα):

- Μαθηματική έκφραση: f(x) = max(0,x)
- Πλεονεκτήματα: Βοηθά στον μετριασμό του προβλήματος της κλίσης που εξαφανίζεται, είναι υπολογιστικά αποτελεσματικό.

Παραλλαγές του ReLU:

- **Leaky ReLU:** Επιτρέπει μια μικρή κλίση για αρνητικές τιμές.
- **Παραμετρικό ReLU (PReLU):** Ένας τύπος ReLU με διαρροή όπου μαθαίνεται η κλίση για τις αρνητικές τιμές κατά τη διάρκεια της προπόνησης.
- **Εκθετική Γραμμική Μονάδα (ELU):** Προσπαθεί να κάνει τις μέσες ενεργοποιήσεις πιο κοντά στο μηδέν, γεγονός που επιταχύνει τη μάθηση.



Οι συναρτήσεις ενεργοποίησης, ειδικά το ReLU και οι παραλλαγές του, διαδραματίζουν κεντρικό ρόλο στα CNN εισάγοντας τη μη γραμμικότητα, επιτρέποντας στο δίκτυο να καταγράφει περίπλοκα μοτίβα στα δεδομένα.

Στρώμα συγκέντρωσης

Σκοπός του στρώματος συγκέντρωσης:

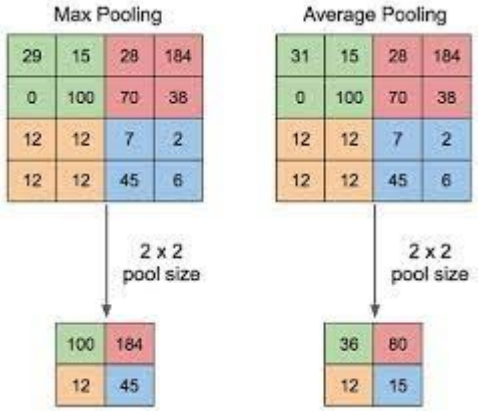
- Μειώνει τις χωρικές διαστάσεις (πλάτος & ύψος) του όγκου εισόδου.
- Βοηθά στη μείωση της υπολογιστικής πολυπλοκότητας.
- Παρέχει μια μορφή μεταφραστικής μεταβλητότητας στο μοντέλο.

Τύποι συγκέντρωσης:

- **Max Pooling:** Λαμβάνει τη μέγιστη τιμή από ένα τμήμα της εισόδου.
- **Μέση συγκέντρωση:** Υπολογίζει τη μέση τιμή από ένα τμήμα της εισόδου.
- **Min Pooling:** Λαμβάνει την ελάχιστη τιμή από ένα τμήμα της εισόδου.

Μέγεθος διασκελισμού και παραθύρου:

- **Διασκελισμός:** Το μέγεθος του βήματος ή ο αριθμός των εικονοστοιχείων που μετατοπίστηκαν κατά τη μετακίνηση του παραθύρου ομαδοποίησης.
- **Μέγεθος παραθύρου:** Το μέγεθος της λειτουργίας συγκέντρωσης (π.χ. 2x2, 3x3).



Τα επίπεδα συγκέντρωσης απλοποιούν τις πληροφορίες, διατηρώντας μόνο τα πιο σημαντικά χαρακτηριστικά για περαιτέρω επεξεργασία

Πλήρως συνδεδεμένο επίπεδο: Γεφύρωση δυνατοτήτων με προβλέψεις

Ο ρόλος του πλήρως συνδεδεμένου επιπέδου:

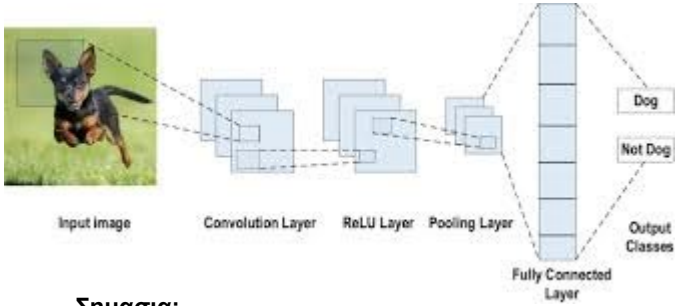
- Συνδέει κάθε νευρώνα από το προηγούμενο στρώμα με τους νευρώνες του τρέχοντος στρώματος.
- Λειτουργεί ως ταξινομητής πάνω από τα εξαγόμενα χαρακτηριστικά από προηγούμενα επίπεδα.

Χάρτες χαρακτηριστικών ισοπέδωσης:

- Πριν από τη σύνδεση στο πλήρως συνδεδεμένο στρώμα, οι χάρτες χαρακτηριστικών 2D ή 3D ισοπεδώνονται σε ένα ενιαίο μεγάλο συνεχές γραμμικό διάνυσμα.
- Αυτό προετοιμάζει τα δεδομένα για την παραδοσιακή επεξεργασία νευρωνικών δικτύων.

Κάνοντας προβλέψεις:

- Το πλήρως συνδεδεμένο επίπεδο χρησιμοποιεί τους ισοπεδωμένους χάρτες χαρακτηριστικών για να κάνει προβλέψεις σχετικά με τα δεδομένα εισόδου.
- Συχνά, το τελικό πλήρως συνδεδεμένο στρώμα έχει νευρώνες ίσους με τον αριθμό των κλάσεων στο πρόβλημα ταξινόμησης.



Σημασια:

- Συνδυάζει όλα τα χαρακτηριστικά γνωρίσματα από προηγούμενα επίπεδα για να προβλέψει την τελική κατηγορία της εικόνας εισόδου.
- Παρέχει τη δυνατότητα εκμάθησης μη γραμμικών συνδυασμών χαρακτηριστικών υψηλού επιπέδου.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

CNN με Python (Keras)

1. Εισαγάγετε τις απαραίτητες βιβλιοθήκες:

```
from keras.models import Sequential
from keras.layers import Conv2D, MaxPooling2D, Flatten, Dense
```

2. Αρχικοποιήστε το CNN:

```
model = Sequential()
```

3. Προσθέστε συνελικτικό επίπεδο :

```
model.add( Conv2D( 32 , ( 3, 3 ), input_shape =( 64, 64, 3 ), activation
=' relu '))
```

4. Προσθήκη επιπέδου συγκέντρωσης:

```
model.add ( MaxPooling2D( pool_size =( 2, 2 )))
```

5. Ισοπεδώστε και συνδέστε πλήρως:

```
model.add(Flatten())
model.add(Dense(units= 128 , activation=' relu '))
model.add(Dense(units= 1 , activation=' sigmoid '))
```

6. Σύνταξη του CNN:

```
model.compile ( optimizer =' adam', loss =
'binary_crossentropy ', metrics =[ 'accuracy' ])
```

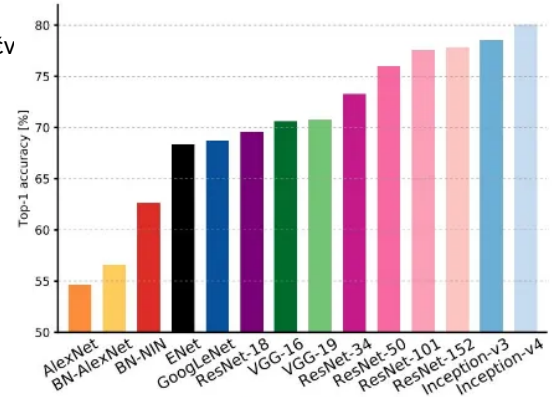
7. Εκπαιδεύστε το μοντέλο:

```
model.fit ( training_set , epochs = 25 ,
validation_data = test_set )
```

Το Keras παρέχει μια φιλική προς τον χρήστη και αρθρωτή πλατφόρμα για τη δημιουργία, την εκπαίδευση και την αξιολόγηση μοντέλων βαθιάς μάθησης, καθιστώντας το μια δημοφιλή επιλογή για την υλοποίηση CNN στην Python.

Αρχιτεκτονικές του CNN

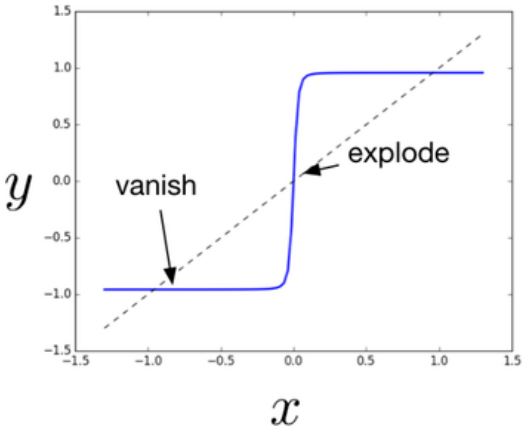
1. **LeNet (1998):**
   - Αναπτύχθηκε από τον Yann LeCun.
   - Χρησιμοποιείται κυρίως για εργασίες αναγνώρισης ψηφίων.
   - Αποτελείται από 2 συνελικτικά στρώματα που ακολουθούνται από 2 πλήρως συνδεδεμένα στρώματα.
2. **AlexNet (2012):**
   - Αναπτύχθηκε από τους Alex Krizhevsky, Ilya Sutskever και Geoffrey Hinton.
   - Νικητής του ImageNet Large Scale Visual Recognition Challenge (ILSVRC) το 2012.
   - Εισήγαγε τη χρήση των ενεργοποιήσεων ReLU και είχε 5 συνελικτικά στρώματα.
3. **VGG (2014):**
   - Αναπτύχθηκε από την ομάδα Visual Geometry στην Οξφόρδη.
   - Γνωστό για την απλότητα και το βάθος του, με διαμορφώσεις που έχουν 16 έως 19 στρώσεις.
   - Χρησιμοποιεί μικρά συνελικτικά φίλτρα 3x3 σε όλο το δίκτυο.
4. **ResNet (2015):**
   - Αναπτύχθηκε από τον Kaiming He και την ομάδα.
   - Εισήγαγε τα "υπολειπόμενα μπλοκ" για την καταπολέμηση του εξαφανιζόμενου προβλήματος της κλίσης στα βαθιά δίκτυα.
   - Νικητής του ILSVRC 2015 με αρχιτεκτονικές που φτάνουν τα 152 επίπεδα.



Οι αρχιτεκτονικές του CNN έχουν εξελιχθεί σημαντικά με τα χρόνια, με κάθε νέα αρχιτεκτονική να εισάγει νέες τεχνικές και δομές για τη βελτίωση της απόδοσης και την αντιμετώπιση των προκλήσεων στη βαθιά μάθηση.

Εκπαίδευση CNN

- **Πίσω διάδοση σε CNN:**
  - Στα CNN, οι διαβαθμίσεις ρέουν προς τα πίσω από την έξοδο σε όλα τα επίπεδα.
  - Τα βάρη κάθε φίλτρου ενημερώνονται για να ελαχιστοποιηθεί η απώλεια.
- **Προκλήσεις στην προπόνηση:**
  - **Διαβάθμιση εξαφάνισης:** Καθώς οι διαβαθμίσεις μεταφέρονται πίσω από βαθιά δίκτυα, μπορεί να γίνουν πολύ μικρές, με αποτέλεσμα τα βάρr να ενημερώνονται πολύ αργά και να καθυστερούν ενδεχομένως η προπόνηση.
  - **Exploding Gradient:** Αντίθετα, οι κλίσεις μπορούν να αυξηθούν εκθετικά σε βαθιά δίκτυα, προκαλώντας μεγάλες ενημερώσεις βάρους και ασταθή προπόνηση.
- **Λύσεις:**
  - Χρήση του ReLU ή των παραλλαγών του για τον μετριασμό των κλίσεων που εξαφανίζονται.
  - Ομαλοποίηση παρτίδας για σταθεροποίηση των ενεργοποιήσεων.
  - Απόκομμα ντεγκραντέ για να αποτρέψετε το να γίνουν πολύ μεγάλες διαβαθμίσεις.



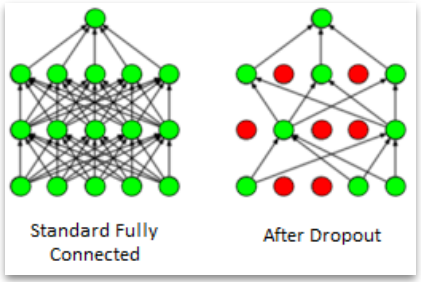
Η εκπαίδευση των CNN απαιτεί προσεκτική διαχείριση των κλίσεων. Ενώ η backpropagation είναι ένα ισχυρό εργαλείο, προκλήσεις όπως η εξαφάνιση και η έκρηξη προβλημάτων κλίσης απαιτούν πρόσθετες τεχνικές και προβληματισμούς.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Τεχνικές τακτοποίησης σε CNN

- **Εγκατάλειψη:**
  - Μια τεχνική όπου οι τυχαία επιλεγμένοι νευρώνες αγνοούνται κατά τη διάρκεια της προπόνησης.
  - Βοηθά στην αποφυγή της υπερπροσαρμογής διασφαλίζοντας ότι το δίκτυο παραμένει ισχυρό και ότι δεν βασίζεται πολύ σε κανέναν μεμονωμένο νευρώνα.
  - Συνήθως εφαρμόζεται στα πλήρως συνδεδεμένα στρώματα.
- **Κανονικοποίηση παρτίδας:**
  - Κανονικοποιεί τις ενεργοποιήσεις ενός δεδομένου όγκου εισόδου πριν τον περάσει μέσα από το επίπεδο.
  - Βοηθά στη σταθεροποίηση της μάθησης και στην άμβλυνση του προβλήματος της διαβάθμισης που εξαφανίζεται/εκρήγνυται.
  - Συχνά χρησιμοποιείται μετά από συνελικτικά και πλήρως συνδεδεμένα στρώματα.
- **Αύξηση δεδομένων:**
  - Διαδικασία τεχνητής αύξησης του μεγέθους του συνόλου εκπαίδευσης.
  - Συνήθεις τεχνικές: περιστροφή, ζουμ, αναστροφή και περικοπή εικόνων.
  - Βοηθά στη βελτίωση της γενίκευσης του μοντέλου και στην πρόληψη της υπερβολικής προσαρμογής.



$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

$$\sigma^2_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_{\mathcal{B}})^2 \qquad \text{// mini-batch variance}$$

$$\widehat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma^2_{\mathcal{B}} + \epsilon}} \qquad \text{// normalize}$$

$$y_i \leftarrow \gamma \widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$

Οι τεχνικές τακτοποίησης όπως η εγκατάλειψη, η ομαλοποίηση παρτίδας και η αύξηση δεδομένων είναι απαραίτητα εργαλεία στην εργαλειοθήκη βαθιάς μάθησης, διασφαλίζοντας ότι τα CNN είναι ακριβή και ισχυρά

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Εκμάθηση μεταφοράς

Ορισμός της Μεταγραφικής Μάθησης:

- Αξιοποίηση της γνώσης από έναν τομέα για τη βελτίωση της απόδοσης σε έναν άλλο σχετικό τομέα.

Οφέλη της Εκμάθησης Μεταβίβασης:

- **Ταχύτητα:** Ταχύτερη εκπαίδευση καθώς το μοντέλο έχει ήδη εκπαιδευτεί σε παρόμοια εργασία.
- **Αποδοτικότητα δεδομένων:** Χρήσιμο όταν υπάρχουν περιορισμένα δεδομένα με ετικέτα για τη συγκεκριμένη εργασία.
- **Απόδοση:** Μπορεί να επιτύχει καλύτερη ακρίβεια, ειδικά για εργασίες με περιορισμένα δεδομένα.

Χρήση προεκπαιδευμένων μοντέλων:

- Ως αφετηρία μπορούν να χρησιμοποιηθούν μοντέλα που εκπαιδεύονται σε μεγάλα σύνολα δεδομένων.
- Τα στρώματα από αυτά τα μοντέλα μπορούν να παγώσουν για να διατηρήσουν τα μαθημένα τους μοτίβα.

Βελτιστοποίηση:

- Ρύθμιση των βαρών του προεκπαιδευμένου μοντέλου συνεχίζοντας την οπίσθια διάδοση.
- Μπορεί να γίνει σε ολόκληρο το μοντέλο ή μόνο σε συγκεκριμένα επίπεδα, ανάλογα με την εργασία και τη διαθεσιμότητα δεδομένων.

Συνήθη προεκπαιδευμένα μοντέλα:

- VGG16, VGG19, ResNet, Inception, MobileNet, κ.λπ.

Εφαρμογές:

- Ταξινόμηση εικόνων, ανίχνευση αντικειμένων, τμηματοποίηση σε τομείς όπου τα δεδομένα ενδέχεται να είναι σπάνια.

Η μάθηση με μεταφορά μας επιτρέπει να αξιοποιούμε προεκπαιδευμένα μοντέλα σε νέες εργασίες, μειώνοντας σημαντικά τον χρόνο εκπαίδευσης και τις απαιτήσεις δεδομένων. Με τη λεπτή ρύθμιση αυτών των μοντέλων, μπορούμε να επιτύχουμε υψηλή απόδοση ακόμη και με περιορισμένα δεδομένα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Εφαρμογές CNN

Ταξινόμηση εικόνων:

- Αντιστοίχιση ετικέτας σε μια εικόνα από ένα σύνολο προκαθορισμένων κατηγοριών.
- Παράδειγμα: Διάκριση μεταξύ εικόνων γατών και σκύλων.

Ανίχνευση αντικειμένου:

- Αναγνώριση αντικειμένων μέσα στις εικόνες και παροχή ενός πλαισίου οριοθέτησης γύρω τους.
- Παράδειγμα: Ανίχνευση και επισήμανση πολλαπλών αντικειμένων όπως αυτοκίνητα, πεζοί και φανάρια σε μια σκηνή του δρόμου.

Δημιουργία εικόνων:

- Δημιουργία νέων εικόνων που μοιάζουν με ένα σύνολο δεδομένων εικόνων.
- Συχνά χρησιμοποιούνται τεχνικές όπως τα Generative Adversarial Networks (GAN).

Περιπτώσεις χρήσης στον πραγματικό κόσμο :

- **Ιατρική Απεικόνιση:** Ανίχνευση όγκων, ανωμαλιών ή ασθενειών σε ακτινογραφίες, μαγνητικές τομογραφίες κ.λπ.
- **Αυτόνομα Οχήματα:** Αναγνώριση πεζών, άλλων οχημάτων και σημάτων κυκλοφορίας για την ασφαλή πλοήγηση.
- **Αναγνώριση προσώπου:** Αναγνώριση ή επαλήθευση ενός ατόμου από μια εικόνα.
- **Επαυξημένη πραγματικότητα:** Επικάλυψη ψηφιακών πληροφοριών στον πραγματικό κόσμο

Τα συνελικτικά νευρωνικά δίκτυα έχουν φέρει επανάσταση στις εργασίες επεξεργασίας εικόνας, από τη βασική ταξινόμηση εικόνων έως πολύπλοκες εφαρμογές του πραγματικού κόσμου, όπως η ιατρική απεικόνιση και η αυτόνομη οδήγηση.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Προκλήσεις στα CNN

1. Ανάγκη για μεγάλα σύνολα δεδομένων:

- **Δεδομένα εκπαίδευσης:** Τα CNN, ειδικά οι βαθύτερες αρχιτεκτονικές, απαιτούν τεράστιες ποσότητες δεδομένων με ετικέτα για να εκπαιδεύονται αποτελεσματικά. Χωρίς επαρκή δεδομένα, μπορούν εύκολα να υπερφορτωθούν.
- **Ποικιλομορφία δεδομένων:** Δεν έχει να κάνει μόνο με την ποσότητα. Η ποιότητα και η ποικιλομορφία των δεδομένων εκπαίδευσης παίζουν καθοριστικό ρόλο στην ικανότητα γενίκευσης του μοντέλου.

2. Υπολογιστικές απαιτήσεις:

- **Απαιτήσεις υλικού:** Η εκπαίδευση CNN, ειδικά οι αρχιτεκτονικές αιχμής, απαιτεί ισχυρές GPU ή TPU. Αυτό μπορεί να αποτελέσει εμπόδιο για μεμονωμένους ερευνητές ή μικρούς οργανισμούς.
- **Χρόνος εκπαίδευσης:** Ακόμη και με ισχυρό υλικό, η εκπαίδευση ενός CNN σε ένα μεγάλο σύνολο δεδομένων μπορεί να διαρκέσει ημέρες ή και εβδομάδες.
- **Πολυπλοκότητα μοντέλου:** Τα βαθύτερα δίκτυα με εκατομμύρια παραμέτρους μπορεί να είναι υπολογιστικά εντατικά, επηρεάζοντας την ανάπτυξη σε εφαρμογές σε πραγματικό χρόνο.

Ενώ τα CNN προσφέρουν αξιοσημείωτη απόδοση σε εργασίες που σχετίζονται με την εικόνα, η ανάγκη τους για εκτεταμένα δεδομένα και υπολογιστική ισχύ μπορεί να δημιουργήσει προκλήσεις, ειδικά για όσους δεν έχουν πρόσβαση σε πόρους υψηλής τεχνολογίας. Είναι σημαντικό να εξισορροπηθεί η πολυπλοκότητα του μοντέλου με τους διαθέσιμους πόρους και τη συγκεκριμένη εργασία.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Το μέλλον των CNN

1. Εξέλιξη Αρχιτεκτονικών:

- **Αποτελεσματικά δίκτυα:** Καθώς οι υπολογιστικοί πόροι γίνονται μέλημα, γίνεται μια στροφή προς το σχεδιασμό αποδοτικών αρχιτεκτονικών που παρέχουν καλή απόδοση με λιγότερες παραμέτρους.
- **Δίκτυα καψουλών:** Προτείνονται ως εναλλακτική λύση στα CNN, στοχεύουν στην αναγνώριση μοτίβων σε μια εικόνα ανεξάρτητα από τον χωρικό τους προσανατολισμό.

2. Πέρα από τα δεδομένα εικόνας:

- **Ανάλυση βίντεο:** CNN σε συνδυασμό με RNN ή 3D συνελίξεις για ταξινόμηση βίντεο και ανίχνευση ανωμαλιών.
- **Πολυτροπική μάθηση:** Ενσωμάτωση δεδομένων από διαφορετικές πηγές (π.χ. εικόνα και κείμενο) για πλουσιότερες αναπαραστάσεις.

3. Ηθική και δίκαιη τεχνητή νοημοσύνη:

- **Μετριασμός μεροληψίας:** Διασφάλιση ότι τα CNN εκπαιδεύονται σε διαφορετικά σύνολα δεδομένων για την αποφυγή μεροληψιών και τη διασφάλιση της δικαιοσύνης.
- **Επεξηγησιμότητα:** Κάνοντας τις αποφάσεις του CNN πιο ερμηνεύσιμες για τους χρήστες, ειδικά σε κρίσιμες εφαρμογές όπως η υγειονομική περίθαλψη.

4. Ενσωμάτωση με επαυξημένη πραγματικότητα (AR) και εικονική πραγματικότητα (VR):

- **Επεξεργασία σε πραγματικό χρόνο:** Χρήση CNN για αναγνώριση εικόνας και σκηνής σε πραγματικό χρόνο σε εφαρμογές AR/VR.
- **Διαδραστική μάθηση:** Εκπαίδευση CNN σε ένα διαδραστικό περιβάλλον όπου μπορούν να μάθουν από τα σχόλια των χρηστών.

Το μέλλον των CNN δεν αφορά μόνο τη βελτίωση της ακρίβειας, αλλά και τη βελτίωση της αποτελεσματικότητας, της ερμηνείας και της ηθικής τους. Καθώς η τεχνολογία εξελίσσεται, τα CNN θα βρίσκουν εφαρμογές σε τομείς που δεν έχουμε καν φανταστεί ακόμα, καθιστώντας τα έναν ακρογωνιαίο λίθο του μέλλοντος της τεχνητής νοημοσύνης.

Συμπέρασμα και Πόροι

- **Πρωτοποριακή τεχνολογία:** Τα CNN έχουν φέρει επανάσταση στον τομέα της όρασης υπολογιστών, καθιστώντας δυνατές εργασίες όπως ταξινόμηση εικόνων, ανίχνευση αντικειμένων, ακόμη και δημιουργία εικόνων.
- **Συνεχής εξέλιξη:** Όπως συμβαίνει με όλες τις τεχνολογίες, τα CNN συνεχίζουν να εξελίσσονται, με τους ερευνητές να πιέζουν συνεχώς τα όρια για να βελτιώσουν την απόδοση και την αποδοτικότητα.
- **Πανταχού παρούσα Εφαρμογή:** Από την υγειονομική περίθαλψη μέχρι την ψυχαγωγία, τα CNN έχουν βρει εφαρμογές σχεδόν σε κάθε τομέα, επιδεικνύοντας την ευελιξία τους.

- **Βιβλία:**
  a. «Deep Learning» των Ian Goodfellow, Yoshua Bengio και Aaron Courville.
  b. «Neural Networks and Deep Learning» του Michael Nielsen.
- **Διαδικτυακά Μαθήματα:**
  a. Coursera: Convolutional Neural Networks του Andrew Ng
  b. Udacity: Deep Learning Nanodegree
- **Ιστότοποι:**
  a. Distill: Οπτικοποίηση χαρακτηριστικών
  b. CS231n: Συνελικτικά νευρωνικά δίκτυα για οπτική αναγνώριση

Τα CNN αποτελούν απόδειξη της δύναμης της βαθιάς μάθησης. Καθώς συνεχίζουμε να εκμεταλλευόμαστε αυτή τη δύναμη, είναι σημαντικό να παραμένετε ενημερωμένοι με τις πιο πρόσφατες εξελίξεις και τις βέλτιστες πρακτικές στον τομέα. Οι παρεχόμενοι πόροι είναι ένα εξαιρετικό σημείο εκκίνησης για όποιον θέλει να εμβαθύνει στον κόσμο των CNN.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

Θέμα 5. Αλγόριθμοι στη Βιοπληροφορική

Μάθημα 1. Βιολογικά Δίκτυα

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

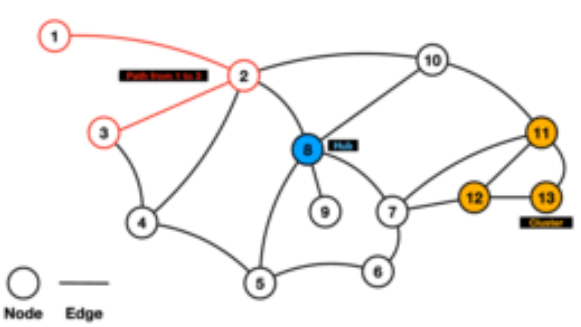
Co-funded by the Erasmus+ Programme of the European Union

Βιολογικά Δίκτυα

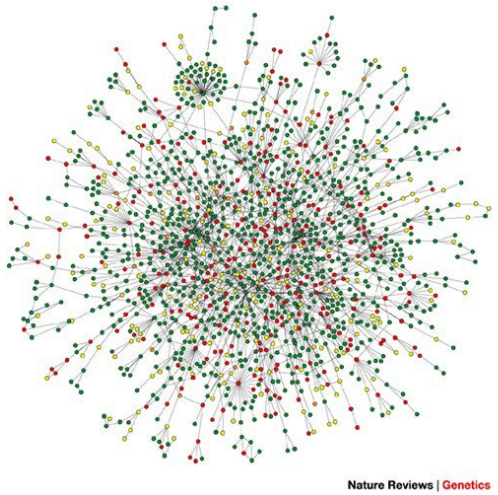
- Τα βιολογικά δίκτυα είναι ένας τρόπος αναπαράστασης και ανάλυσης πολύπλοκων βιολογικών συστημάτων, όπως οι μοριακές αλληλεπιδράσεις, τα μονοπάτια σηματοδότησης και τα μεταβολικά μονοπάτια.

- Ένα βιολογικό δίκτυο είναι μια συλλογή κόμβων (ή κορυφών) και ακμών (ή συνδέσμων) που τους συνδέουν.

- Οι κόμβοι αντιπροσωπεύουν βιολογικές οντότητες, όπως γονίδια, πρωτεΐνες, μεταβολίτες ή κύτταρα, ενώ οι ακμές αντιπροσωπεύουν τις σχέσεις ή τις αλληλεπιδράσεις μεταξύ τους.



- Διακρίνονται μεταξύ
  - Δίκτυα στο κύτταρο
  - Δίκτυα εκτός κυττάρου



- Αυτά τα δίκτυα μπορούν να αναπαρασταθούν μαθηματικά ως γραφήματα και μπορούν να αναλυθούν χρησιμοποιώντας τη θεωρία γραφημάτων και τις τεχνικές επιστήμης των δικτύων.
- Η κατανόηση της δομής και της δυναμικής των βιολογικών δικτύων είναι ζωτικής σημασίας για την κατανόηση πολύπλοκων βιολογικών φαινομένων και για την ανάπτυξη νέων θεραπευτικών στρατηγικών.

Co-funded by the Erasmus+ Programme of the European Union

Βιολογικά δίκτυα για τη βιοπληροφορική

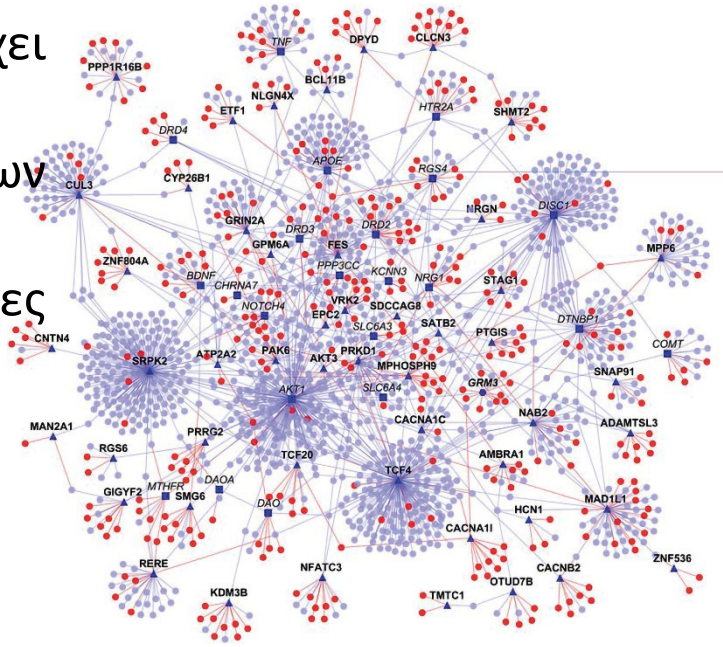
- Η μελέτη των βιολογικών δικτύων έχει γίνει το κύριο επίκεντρο της έρευνας στη βιοπληροφορική τα τελευταία χρόνια.
- Μπορεί να οδηγήσει σε
  - καλύτερη κατανόηση των βιολογικών συστημάτων
  - ταυτοποίηση βιοδεικτών ασθένειας
  - και ανάπτυξη νέων φαρμάκων.
  - Προσδιορισμός βασικών κόμβων, κρίσιμων για το δίκτυο, και τη λειτουργία του (πιθανοί στόχοι) .

- Συνολικά, η χρήση βιολογικών δικτύων στη βιοπληροφορική έχει φέρει επανάσταση στον τρόπο με τον οποίο μελετάμε και κατανοούμε πολύπλοκα βιολογικά συστήματα, και υπόσχεται πολλά για την προώθηση των γνώσεών μας για τους υποκείμενους μηχανισμούς της ζωής.

Πρωτεΐνη – Αλληλεπίδραση Πρωτεϊνών

- Τα δίκτυα αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης (PPIN) είναι γραφήματα που αντιπροσωπεύουν τις φυσικές αλληλεπιδράσεις μεταξύ πρωτεϊνών σε ένα κύτταρο.

- Η ανάλυση των PPIN μπορεί να παρέχει πληροφορίες
  - στους ρόλους των μεμονωμένων πρωτεϊνών
  - αλληλεπιδράσεις τους στις διαδικασίες της νόσου.



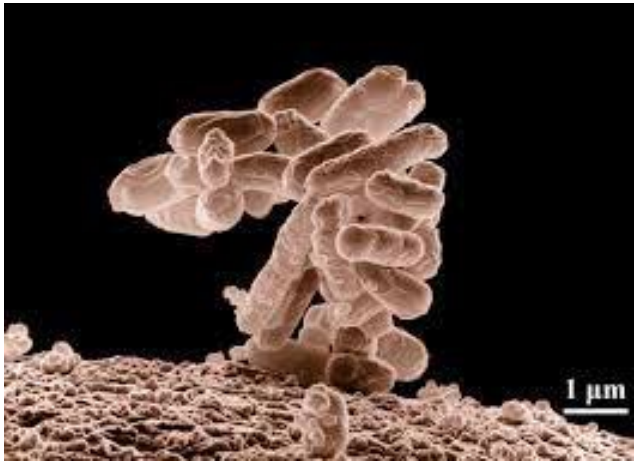
Αλγόριθμοι για πρωτεϊνικά δίκτυα

- Οι αλγόριθμοι που εφαρμόζονται στα PPIN μπορούν να παρέχουν πολύτιμες πληροφορίες για τους μηχανισμούς ασθενειών και τους πιθανούς στόχους φαρμάκων
- Βήματα κατασκευής ενός τέτοιου αλγορίθμου
  - Κατασκευή PPIN
  - Προσδιορισμός μονάδων δικτύου ή συστάδων πρωτεϊνών υψηλής διασύνδεσης
  - Ενσωμάτωση πληροφοριών φαρμάκου-στόχου και δεδομένων γονιδιακής έκφρασης
  - Προτεραιότητα των πιθανών στόχων φαρμάκων με βάση την κεντρική τους θέση στο δίκτυο και τη συσχέτιση με οδούς που σχετίζονται με ασθένειες

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Μεταβολικά Δίκτυα

- Τα μεταβολικά δίκτυα είναι ένας τύπος βιολογικού δικτύου που αντιπροσωπεύει το διασυνδεδεμένο σύνολο μεταβολικών αντιδράσεων μέσα σε ένα κύτταρο ή οργανισμό.
- Κατευθυνόμενα γραφήματα
  - Οι κόμβοι αντιστοιχούν σε μεταβολίτες
  - Τα άκρα αντιστοιχούν σε αντιδράσεις μεταξύ μεταβολιτών.
- Παραδείγματα
  - Escherichia coli
  - Saccharomyces cerevisiae



Σημασία Μεταβολικών δικτύων

- Σημαντικό για την κατανόηση των πολύπλοκων αλληλεπιδράσεων μεταξύ μεταβολιτών και ενζύμων που διέπουν τις μεταβολικές διεργασίες εντός των κυττάρων.

- Πρόβλεψη της συμπεριφοράς των μεταβολικών συστημάτων υπό διαφορετικές συνθήκες

- Προσδιορισμός πιθανών στόχων φαρμάκων για μεταβολικές ασθένειες.

- Ωστόσο, η ανάλυση των μεταβολικών δικτύων είναι προκλητική λόγω της πολυπλοκότητάς τους και της ανάγκης να ενσωματωθούν δεδομένα από πολλαπλές πηγές, συμπεριλαμβανομένης της κινητικής των ενζύμων, της γονιδιακής έκφρασης και των δεδομένων μεταβολισμού.

Ομαδοποίηση σε μεταβολικά δίκτυα

- Οι αλγόριθμοι ομαδοποίησης μπορούν να εφαρμοστούν για τον εντοπισμό ομάδων μεταβολιτών ή ενζύμων που σχετίζονται λειτουργικά και ενδέχεται να εμπλέκονται σε συγκεκριμένες βιολογικές οδούς.

- Συνολικά, οι αλγόριθμοι ομαδοποίησης παρέχουν ένα ισχυρό εργαλείο για την ανάλυση των μεταβολικών δικτύων

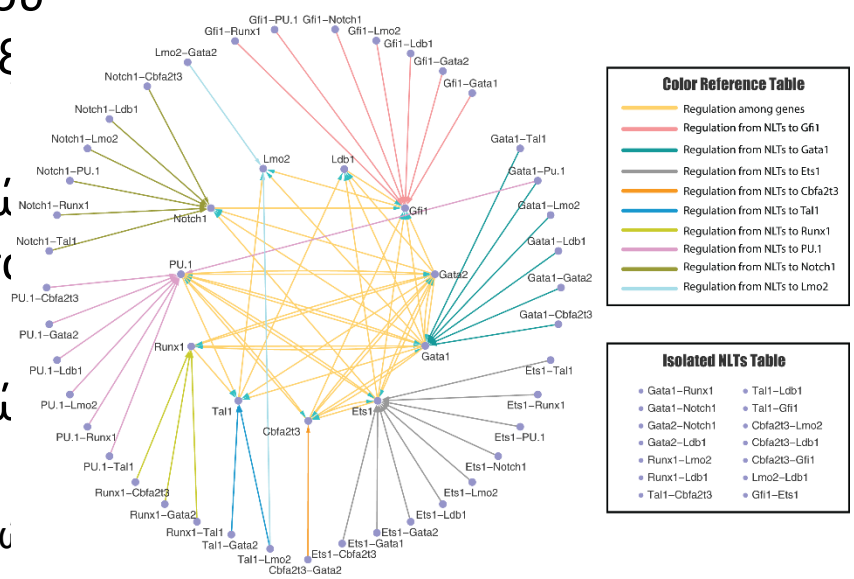
  - Προσδιορισμός λειτουργικά σχετικών μεταβολιτών και ενζύμων

  - παρέχοντας γνώσεις για βασικές μεταβολικές οδούς και πιθανούς στόχους φαρμάκων.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Ρυθμιστικά δίκτυα γονιδίων

- Τα γονιδιακά ρυθμιστικά δίκτυα είναι ένας τύπος βιολογικού δικτύου που μοντελοποιεί τις αλληλεπιδράσεις μεταξ των γονιδίων και των ρυθμιστών τους.
- Χρησιμοποιείται για να κατανοήσει πώ ρυθμίζονται τα γονίδια και πώς ελέγχετ η έκφρασή τους.
- Παραδείγματα γονιδιακών ρυθμιστικώ δικτύων
  - αλληλεπιδράσεις μεταξύ των μεταγραφικ παραγόντων και των γονιδίων-στόχων τους
  - αλληλεπιδράσεις μεταξύ των microRNA και των mRNA-στόχων τους.



Σημασία των ρυθμιστικών δικτύων γονιδίων

- Κρίσιμος ρόλος στην κατανόηση της ρύθμισης της γονιδιακής έκφρασης σε διάφορες βιολογικές διεργασίες.

- Είναι απαραίτητα για την κατανόηση των μοριακών μηχανισμών που αποτελούν τη βάση της ανάπτυξης, της διαφοροποίησης και της ασθένειας.

- Εντοπίζω βασικά ρυθμιστικά γονίδια και μονοπάτια, πιθανούς στόχους φαρμάκων και βιοδείκτες για τη διάγνωση και τη θεραπεία ασθενειών.

- Αναγνώριση γενετικών παραλλαγών που σχετίζονται με την ευαισθησία στη νόσο και την απόκριση στα φάρμακα, καθώς και πρόβλεψη των επιπτώσεων μεταλλάξεων και περιβαλλοντικών παραγόντων στη γονιδιακή έκφραση.

Ρυθμιστικά δίκτυα στη βιοπληροφορική

- Αλγόριθμοι ομαδοποίησης έχουν επίσης εφαρμοστεί σε ρυθμιστικά δίκτυα γονιδίων για τον εντοπισμό ομάδων γονιδίων με παρόμοιες λειτουργίες.

- Ένα παράδειγμα αυτού είναι η χρήση ιεραρχικής ομαδοποίησης σε δεδομένα γονιδιακής έκφρασης για τον εντοπισμό συν-ρυθμιζόμενων γονιδίων.

- Οι ομάδες θα μπορούσαν να χρησιμοποιηθούν για την πρόβλεψη της λειτουργίας γονιδίων που δεν είχαν χαρακτηριστεί προηγουμένως.

- Αυτή η προσέγγιση έχει τη δυνατότητα να επιταχύνει την ανακάλυψη νέων ρυθμιστικών μηχανισμών και θεραπευτικού στόχου

Co-funded by the Erasmus+ Programme of the European Union

Ασκηση

- Ορίστε τα βιολογικά δίκτυα
- Ονομάστε 3 κατηγορίες βιολογικών δικτύων
- Πώς μπορούν να εφαρμοστούν αυτά τα είδη δικτύων στη βιοπληροφορική; Ονομάστε μια από τις αιτήσεις τους
- Δώστε ένα παράδειγμα δικτύου για κάθε κατηγορία

Co-funded by the Erasmus+ Programme of the European Union

Περίληψη

- Τα βιολογικά δίκτυα είναι ένας τρόπος αναπαράστασης και ανάλυσης πολύπλοκων βιολογικών συστημάτων, όπως οι μοριακές αλληλεπιδράσεις, τα μονοπάτια σηματοδότησης και τα μεταβολικά μονοπάτια.
- Είναι μια συλλογή από κόμβους (ή κορυφές) και ακμές (ή συνδέσμους) που τους συνδέουν.
- Παραδείγματα βιολογικών δικτύων
  - Πρωτεΐνη-Πρωτεΐνη
  - Μεταβολικός
  - Ρυθμιστικό Γονιδίων
- Κρίσιμο για την έρευνα

Co-funded by the Erasmus+ Programme of the European Union

Δίκτυα ασθενειών

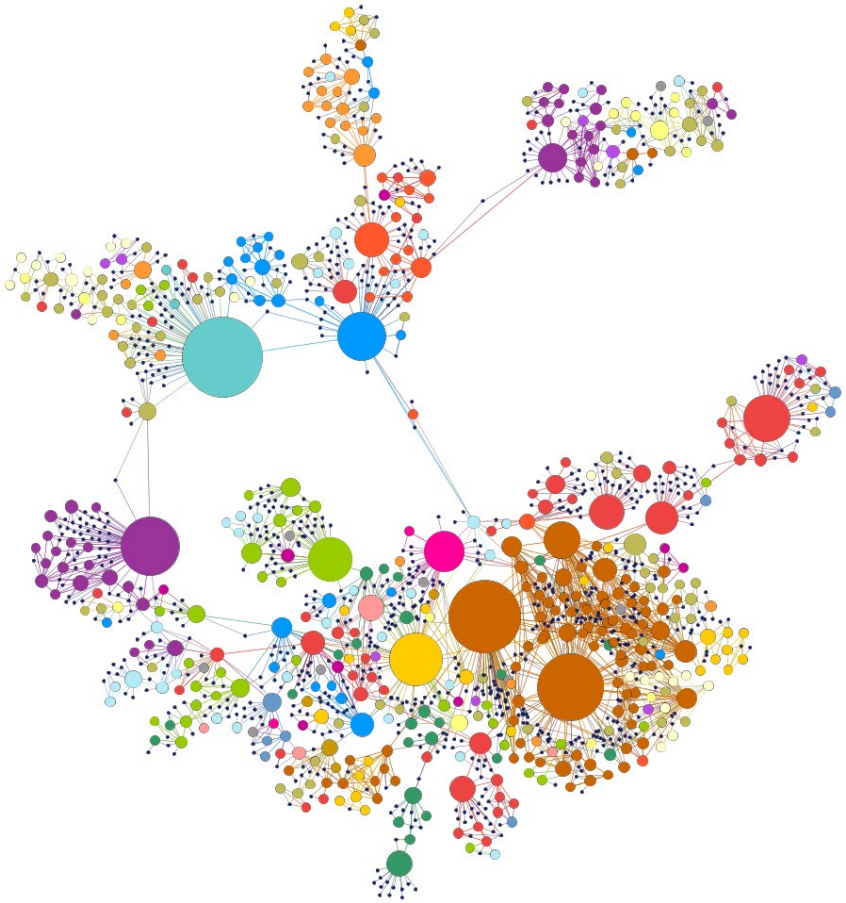
- Τα δίκτυα ασθενειών είναι γραφικές αναπαραστάσεις του
  - οι περίπλοκες σχέσεις μεταξύ των ασθενειών
  - τα γενετικά τους συστατικά
  - και τους μοριακούς μηχανισμούς τους.

- Κατασκευάζονται με την ενσωμάτωση διαφόρων τύπων βιολογικών δεδομένων, όπως π.χ
  - γονιδιακή έκφραση
  - αλληλεπιδράσεις πρωτεΐνης-πρωτεΐνης
  - και μονοπάτια.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

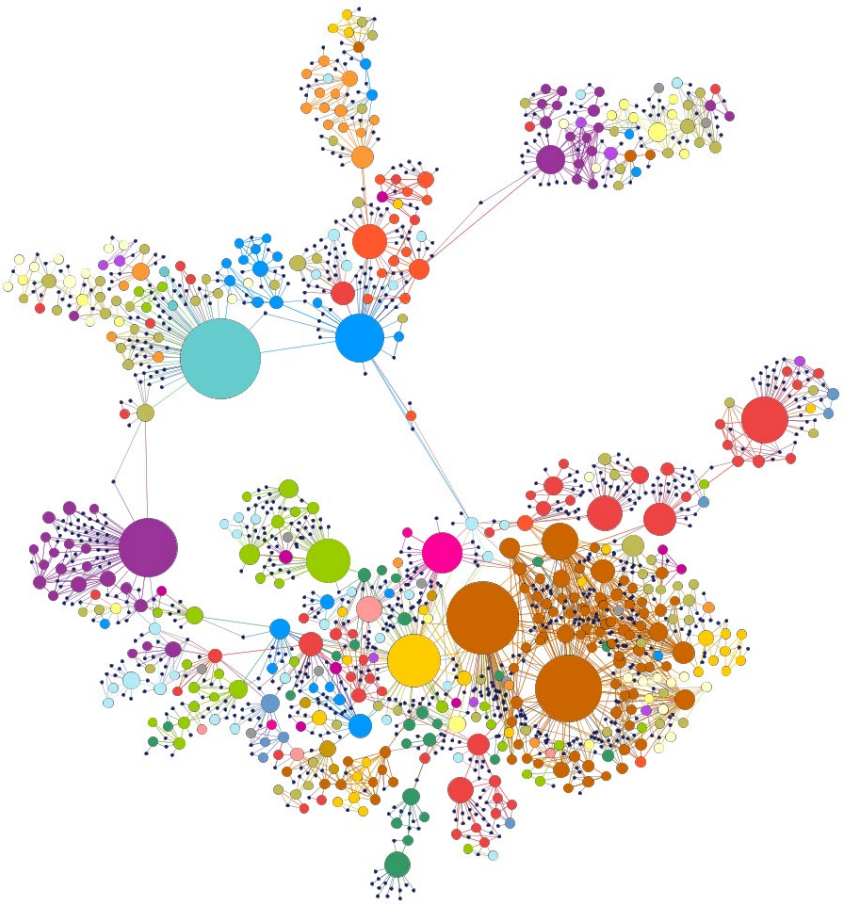
Δίκτυο Ανθρώπινων Ασθενειών

- Οι Ν ωδές αντιπροσωπεύουν ασθένειες
- D συνδέονται μεταξύ τους εάν μοιράζονται τουλάχιστον ένα γονίδιο στο οποίο οι μεταλλάξεις σχετίζονται και με τις δύο ασθένειες
- Σε αυτό το παράδειγμα
  - ορατά ομαδοποιούνται σύμφωνα με τις κύριες κατηγορίες ασθενειών



Δίκτυα ασθενειών στη Βιοπληροφορική

- Η ανάλυση δικτύου αποκαλύπτ ομάδες ασθενειών με παρόμοιοι μοριακούς μηχανισμού υποδεικνύοντας πιθανούς κοινοί στόχους φαρμάκων και στρατηγικε θεραπείας.
- Βασικές ασθένειες και γονίδια πο προσδιορίζονται μέσω ανάλυσι δικτύου διερευνώνται περαιτέρω σ πειράματα σε υγρό εργαστήρι οδηγώντας στην ανακάλυψη ενο νέου θεραπευτικού στόχου για μι ασθένεια που δεν θεραπεύτηκ προηγουμένως.



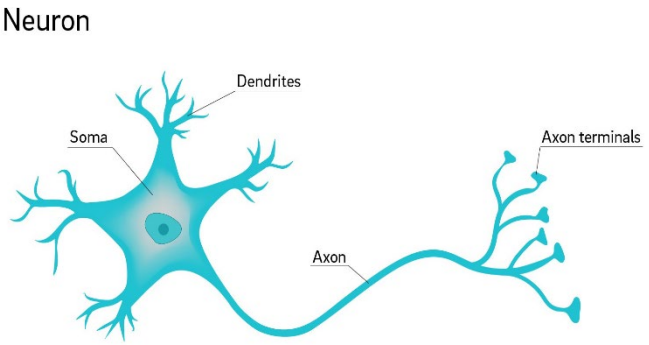
Βιολογικά Νευρωνικά Δίκτυα

- Τα βιολογικά νευρωνικά δίκτυα αναφέρονται στα δίκτυα των διασυνδεδεμένων νευρώνων που υπάρχουν στο νευρικό σύστημα των ζωντανών οργανισμών.
- Αυτά τα δίκτυα παίζουν καθοριστικό ρόλο
  - στην επεξεργασία πληροφοριών,
  - ελεγκτική συμπεριφορά,
  - ρύθμιση των σωματικών λειτουργιών.

- Ο εγκέφαλος είναι το πιο γνωστό παράδειγμα βιολογικού νευρωνικού δικτύου, που αποτελείται από δισεκατομμύρια νευρώνες και τρισεκατομμύρια συνδέσεις.
- Άλλα παραδείγματα περιλαμβάνουν τα νευρωνικά δίκτυα που βρέθηκαν
  - στον νωτιαίο μυελό,
  - τον αμφιβληστροειδή του ματιού,
  - και το εντερικό νευρικό σύστημα στο έντερο.

Νευρωνικά Δίκτυα στη Βιοπληροφορική

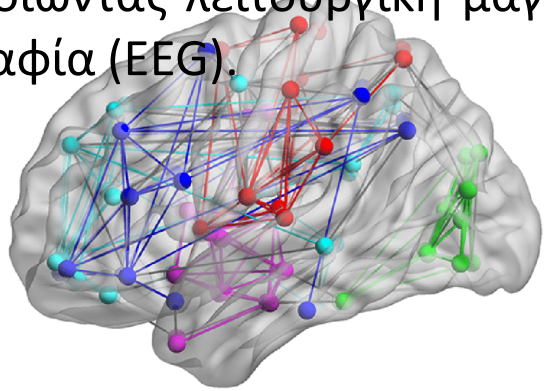
- Η μελέτη των βιολογικών νευρωνικών δικτύων μπορεί να μας βοηθήσει να κατανοήσουμε πώς λειτουργεί ο εγκέφαλος και πώς επεξεργάζεται τις πληροφορίες.
- Αυτή η γνώση μπορεί να οδηγήσει στην ανάπτυξη πιο προηγμένων αλγορίθμων μηχανικής μάθησης και τεχνητής νοημοσύνης.
- Οι ερευνητές μελετούν επίσης τα νευρωνικά δίκτυα για να αποκτήσουν γνώσεις σχετικά με τις νευρολογικές διαταραχές και να αναπτύξουν θεραπείες για αυτές.



Co-funded by the Erasmus+ Programme of the European Union

Δίκτυα εγκεφάλου

- Τα δίκτυα εγκεφάλου αναφέρονται στα πολύπλοκα δίκτυα των νευρώνων και τις συνδέσεις τους στον εγκέφαλο που αποτελούν τη βάση της γνωστικής λειτουργίας και συμπεριφοράς.
- Δομικά δίκτυα: αυτές είναι οι φυσικές συνδέσεις μεταξύ νευρώνων και περιοχών του εγκεφάλου. Μπορούν να μελετηθούν με τη χρήση μεθόδων απεικόνισης τανυστή διάχυσης (DTI) ή τρακτογραφίας.
- Λειτουργικά δίκτυα: αυτά είναι τα πρότυπα συντονισμένης δραστηριότητας μεταξύ των περιοχών του εγκεφάλου. Μπορούν να μελετηθούν χρησιμοποιώντας λειτουργική μαγνητική τομογραφία (fMRI) ή ηλεκτροεγκεφαλογραφία (EEG).



Δίκτυα εγκεφάλου στη Βιοπληροφορική

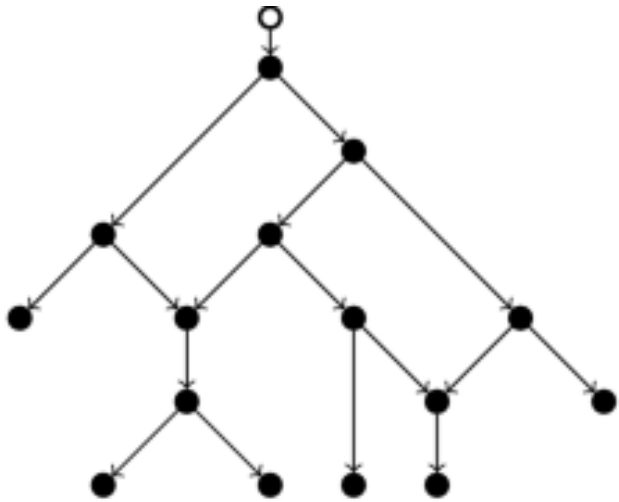
- Τα δίκτυα εγκεφάλου αποτελούν βασικό επίκεντρο της νευροπληροφορικής , η οποία είναι η μελέτη της οργάνωσης και της λειτουργίας του εγκεφάλου σε επίπεδο συστημάτων.
- Η κατανόηση της δομής και της λειτουργίας των εγκεφαλικών δικτύων είναι ζωτικής σημασίας για την κατανόηση του τρόπου λειτουργίας του εγκεφάλου και για την ανάπτυξη θεραπειών για νευρολογικές και ψυχιατρικές διαταραχές.
- Οι μέθοδοι ανάλυσης δικτύου μπορούν να εφαρμοστούν σε δίκτυα εγκεφάλου για τον εντοπισμό προτύπων συνδεσιμότητας που σχετίζονται με συγκεκριμένες γνωστικές ή συμπεριφορικές λειτουργίες.

- Μία από τις κύριες εφαρμογές είναι η ταξινόμηση διαφορετικών εγκεφαλικών καταστάσεων
- Σε πολλές περιπτώσεις, οι μετρήσεις των εγκεφαλικών δικτύων που παράγονται από άτομα με την ίδια πάθηση (επιληψία, AD, ADHD) διαφέρουν από τα υγιή άτομα
- Αφού εκπαιδεύσουμε ένα μεγάλο σύνολο δεδομένων, μπορούμε να συμπεράνουμε από το δίκτυο ενός ατόμου εάν είναι άρρωστο ή όχι.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Φυλογενετικά Δίκτυα

- Τύπος βιολογικού δικτύου που αντιπροσωπεύει εξελικτικές σχέσεις μεταξύ ειδών ή γονιδίων που δεν μπορούν να συλληφθούν με παραδοσιακές προσεγγίσεις που βασίζονται σε δέντρα.
- Χρησιμοποιούνται για τη μελέτη εξελικτικών γεγονότων όπως ο υβριδισμός, η οριζόντια μεταφορά γονιδίων και ο ανασυνδυασμός που δημιουργούν δίκτυα με δικτυωτή δομή.



Σημασία των φυλογενετικών Δικτύων

- Παρέχουν μια πιο ακριβή αναπαράσταση της εξελικτικής ιστορίας των ειδών ή των γονιδίων.
- Οι παραδοσιακές προσεγγίσεις που βασίζονται σε δέντρα υποθέτουν ότι οι εξελικτικές σχέσεις είναι αυστηρά διχασμένες, αλλά αυτό δεν συμβαίνει πάντα.
- Μπορεί να βοηθήσει τους ερευνητές να αποκτήσουν μια πληρέστερη κατανόηση της εξέλιξης των βιολογικών συστημάτων.

Co-funded by the Erasmus+ Programme of the European Union

Ασκηση

- Ονομάστε τις 4 κατηγορίες βιολογικών δικτύων που αναφέρονται στο μάθημα
- Πώς μπορούν να εφαρμοστούν αυτά τα είδη δικτύων στη βιοπληροφορική; Ονομάστε μία από τις αιτήσεις τους
- Δώστε ένα παράδειγμα δικτύου για κάθε κατηγορία

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Περίληψη

- Σε αυτό το μάθημα εξερευνήσαμε ορισμένους τύπους βιολογικών δικτύων και τη σημασία τους στη βιοπληροφορική. Αυτά τα δίκτυα είναι
- Δίκτυα ασθενειών
  - Πολύπλοκες σχέσεις μεταξύ ασθενειών
- Βιολογικά Νευρωνικά Δίκτυα
  - Συνδέσεις μεταξύ νευρώνων
- Δίκτυα εγκεφάλου
  - Συνδέσεις μεταξύ περιοχών του εγκεφάλου/νευρώνων
- Φυλογενετικά δίκτυα
  - Εξελικτικά δέντρα

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

Θέμα 5. Αλγόριθμοι στη Βιοπληροφορική

Μάθημα 2. Αλγόριθμοι

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

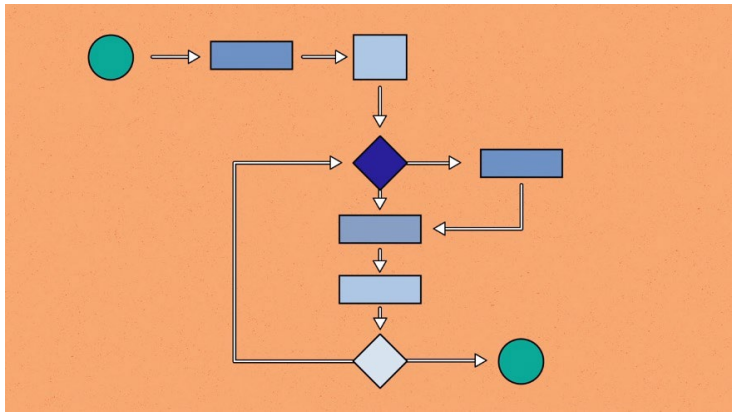
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμοι

- Οι αλγόριθμοι είναι διαδικασίες για την επίλυση προβλημάτων ή την εκτέλεση εργασιών.
- Είναι απαραίτητα στην επιστήμη των υπολογιστών και σε συναφείς τομείς, καθώς παρέχουν έναν συστηματικό τρόπο επίλυσης πολύπλοκων προβλημάτων.
- Στη βιοπληροφορική, οι αλγόριθμοι χρησιμοποιούνται για την ανάλυση βιολογικών δεδομένων, τη μοντελοποίηση βιολογικών συστημάτων και την προσομοίωση πολύπλοκων βιολογικών διεργασιών.

- Παραδείγματα αλγορίθμων για τη βιοπληροφορική
  - αλγόριθμοι ευθυγράμμισης αλληλουχιών για σύγκριση αλληλουχιών DNA και πρωτεϊνών,
  - αλγόριθμοι ανάλυσης δικτύου για τη μελέτη βιολογικών δικτύων
  - Αλγόριθμοι μηχανικής μάθησης για την πρόβλεψη βιολογικών αποτελεσμάτων

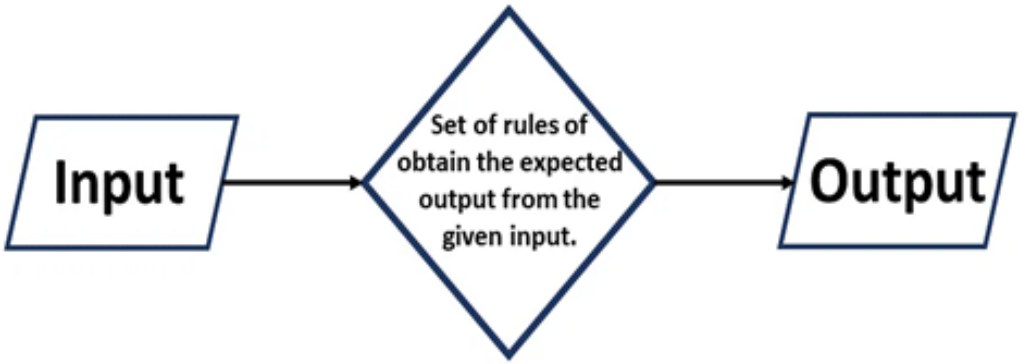


Στόχος του μαθήματος

- Στόχος του μαθήματος είναι να εισαγάγει τους φοιτητές στους αλγόριθμους και τις εφαρμογές τους στη βιοπληροφορική.
- Θα καλύψουμε διαφορετικούς τύπους αλγορίθμων,
- τις ιδιότητές τους
- περίπλοκο
- και πώς μπορούν να εφαρμοστούν για την επίλυση βιολογικών προβλημάτων.

Αλγόριθμοι, Ορισμός

- Οι αλγόριθμοι είναι ένα σύνολο από καλά καθορισμένες οδηγίες για την επίλυση ενός προβλήματος ή την εκτέλεση μιας εργασίας.
- Αποτελούνται από μια σειρά βημάτων που μετατρέπουν τις εισόδους σε εξόδους.



Algorithm

Co-funded by the Erasmus+ Programme of the European Union

Βήματα

- Είσοδοι: Τα δεδομένα ή οι πληροφορίες στις οποίες λειτουργεί ο αλγόριθμος.
- Έξοδοι: Το αποτέλεσμα του αλγορίθμου.
- Δομές ελέγχου: Οι μηχανισμοί λήψης αποφάσεων που ελέγχουν τη ροή του αλγορίθμου, όπως εντολές υπό όρους και βρόχοι.
- Λειτουργίες: Οι ενέργειες που εκτελεί ο αλγόριθμος στα δεδομένα εισόδου για να τα μετατρέψει σε δεδομένα εξόδου

Co-funded by the Erasmus+ Programme of the European Union

Εφαρμογές

- Οι αλγόριθμοι χρησιμοποιούνται για την επίλυση προβλημάτων στην επιστήμη των υπολογιστών, τη μηχανική, τα μαθηματικά και άλλους τομείς.
- Χρησιμοποιούνται για την αυτοματοποίηση εργασιών, τη βελτιστοποίηση διαδικασιών και τη λήψη αποφάσεων με βάση δεδομένα.
- Παραδείγματα
  - Αλγόριθμοι αναζήτησης για την εύρεση πληροφοριών σε μια βάση δεδομένων ή στο διαδίκτυο.
  - Αλγόριθμοι ταξινόμησης για την οργάνωση δεδομένων με συγκεκριμένη σειρά.
  - Αλγόριθμοι κρυπτογράφησης για την ασφαλή μετάδοση δεδομένων.
  - Αλγόριθμοι βελτιστοποίησης για την ελαχιστοποίηση του κόστους ή τη μεγιστοποίηση της απόδοσης.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Παράδειγμα

- Φτιάχνοντας καφέ
- Είσοδοι:
  - Καφές, φίλτρο, καφετιέρα, κούπα
- Έξοδοι
  - Κούπα καφέ
- Βήματα
  - Τοποθετήστε τον καφέ στην καφετιέρα
  - Τοποθετήστε το φίλτρο
  - Ενεργοποιήστε το μηχάνημα
  - Όταν ο καφές είναι έτοιμος, σβήστε τον
  - Τοποθετήστε το σε μια κούπα

Ανάλυση Αλγορίθμων

- Είναι σημαντικό να αναλύονται οι αλγόριθμοι για να διασφαλίζεται ότι είναι αποτελεσματικοί και επεκτάσιμοι για μεγάλες εισόδους όσον αφορά την πολυπλοκότητα του χρόνου και του χώρου.

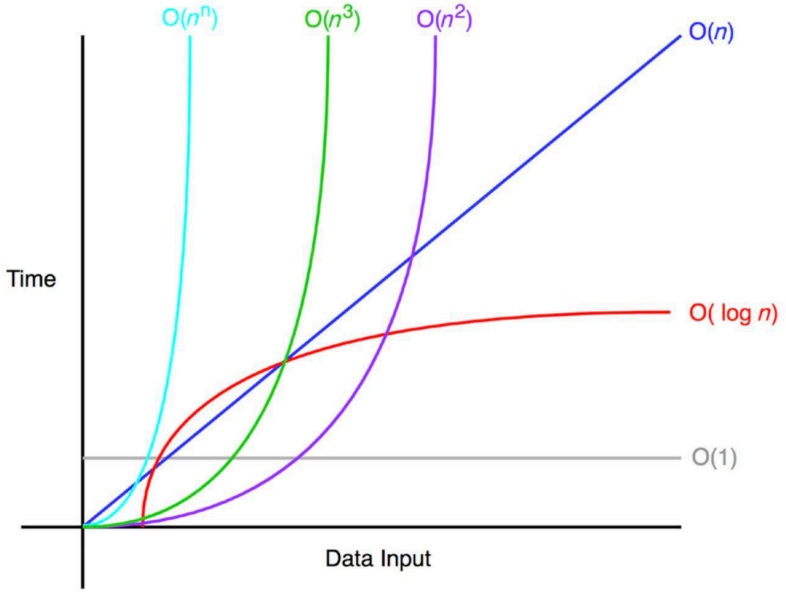
- Η χρονική πολυπλοκότητα αναφέρεται στον χρόνο που χρειάζεται ένας αλγόριθμος για να εκτελεστεί καθώς αυξάνεται το μέγεθος εισόδου.

- Η πολυπλοκότητα χώρου αναφέρεται στην ποσότητα μνήμης που χρησιμοποιεί ένας αλγόριθμος καθώς αυξάνεται το μέγεθος εισόδου.

Big – O σημειογραφία

- Ένας τρόπος ανάλυσης αλγορίθμων είναι η χρήση σημειογραφίας Big-O, η οποία εκφράζει το ανώτερο όριο του χρόνου εκτέλεσης ή του χώρου χρήσης ενός αλγορίθμου.

- Οι κοινές χρονικές πολυπλοκότητες και οι σημειώσεις τους Big-O περιλαμβάνουν:
  - Σταθερός χρόνος: O(1)
  - Γραμμικός χρόνος: O(n)
  - Τετραγωνικός χρόνος: O(n’
  - Εκθετικός χρόνος: O(2^n)

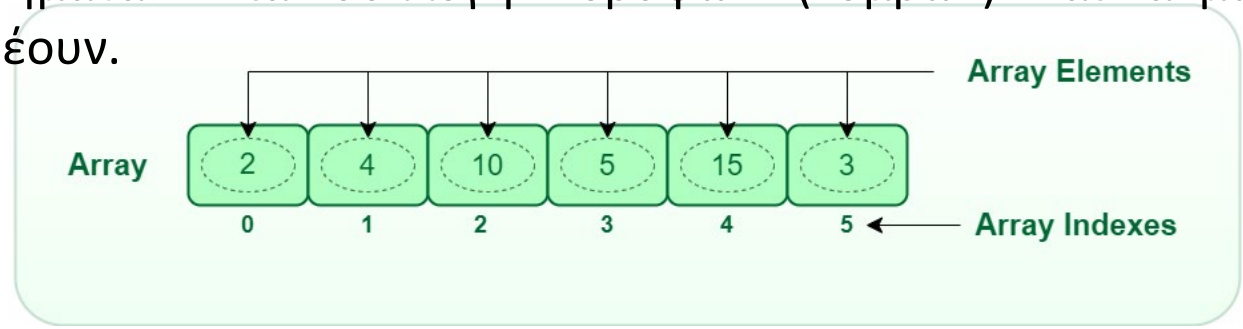


Τεχνικές Σχεδιασμού Αλγορίθμων

- Divide and Conquer: διάσπαση ενός προβλήματος σε υποπροβλήματα, επίλυση τους αναδρομικά και συνδυασμός λύσεων.
  - Συγχώνευση ταξινόμησης
- Δυναμικός προγραμματισμός: διάσπαση ενός προβλήματος σε μικρότερα υποπροβλήματα και επίλυση κάθε υποπροβλήματος μόνο μία φορά.
  - Παράδειγμα: Ακολουθία Fibonacci.
- Greedy Algorithms: Μια τεχνική που κάνει την τοπικά βέλτιστη επιλογή σε κάθε βήμα με την ελπίδα να βρεθεί ένα παγκόσμιο βέλτιστο.
  - Κωδικοποίηση Huffman.
- Backtracking: αναζήτηση λύσης δημιουργώντας υποψηφίους σταδιακά και απορρίπτοντάς τους μόλις δεν μπορούν να οδηγήσουν σε μια έγκυρη λύση.
  - Παράδειγμα: Επίλυση Sudoku.

ΔΟΜΕΣ ΔΕΔΟΜΕΝΩΝ

- Δομές δεδομένων: Πίνακες, συνδεδεμένες λίστες, στοίβες, ουρές, δέντρα και γραφήματα.
- Πίνακες: Μια συλλογή στοιχείων του ίδιου τύπου που αποθηκεύονται σε συνεχόμενες θέσεις μνήμης.
- Συνδεδεμένες λίστες: Μια συλλογή στοιχείων, καθένα από τα οποία περιέχει μια αναφορά στο επόμενο στοιχείο της λίστας.
- Στοίβες: Μια συλλογή στοιχείων που υποστηρίζει δύο λειτουργίες: push (προσθήκη στοιχείου στην κορυφή της στοίβας) και pop (αφαίρεση του επάνω στοιχείου από τη στοίβα).
- Γραφήματα: Μια συλλογή κορυφών (κόμβων) και ακμών που τις συνδέουν.

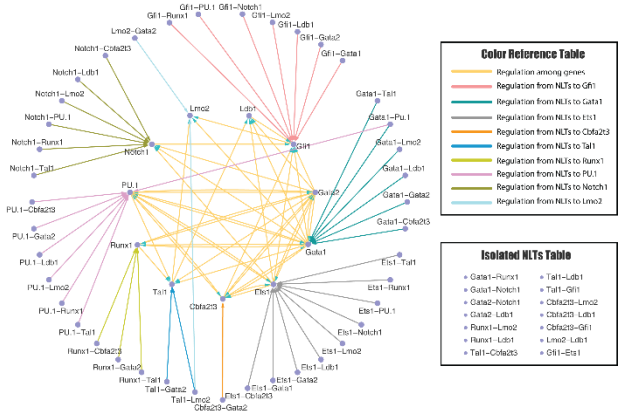


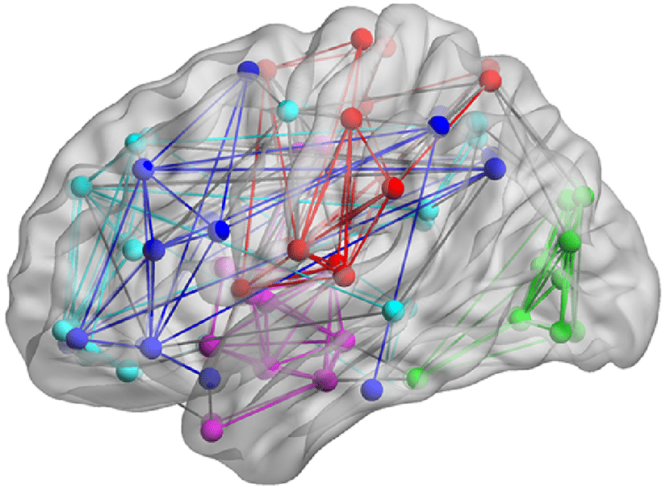
Σύνδεσμος για Βιοπληροφορική

- Επισκόπηση του τρόπου με τον οποίο χρησιμοποιούνται οι αλγόριθμοι στη βιοπληροφορική και την υπολογιστική βιολογία.
  - Χρήση αλγορίθμων για να
    - Απεικονίζω,
    - Μια ανάλυση,
    - ερμηνεύω

  βιολογικά δεδομένα.
  - Σημασία αλγοριθμικών τεχνικών για την αναγνώριση προτύπων σε βιολογικά δεδομένα.

- Βιολογικά δεδομένα ως δίκτυα:
  - Επεξήγηση του τρόπου με τον οποίο τα βιολογικά δεδομένα μπορούν να μοντελοποιηθούν ως δίκτυα.
  - Παραδείγματα βιολογικών δικτύων όπως δίκτυα αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης, γονιδιακά ρυθμιστικά δίκτυα, μεταβολικά δίκτυα, δίκτυα εγκεφάλου.





Co-funded by the Erasmus+ Programme of the European Union

- Αλγοριθμική Ανάλυση Βιολογικών Δικτύων:
  - Επισκόπηση του τρόπου με τον οποίο οι αλγόριθμοι μπορούν να χρησιμοποιηθούν για την ανάλυση βιολογικών δικτύων.
  - Παραδείγματα αλγορίθμων που χρησιμοποιούνται στην ανάλυση δικτύου, όπως αλγόριθμοι ομαδοποίησης, θεωρητική μοντελοποίηση δικτύου, ευθυγράμμιση δικτύου.
- Πάρε μακριά:
  - Οι αλγόριθμοι διαδραματίζουν κρίσιμο ρόλο στην ανάλυση βιολογικών δεδομένων, ειδικά στο πλαίσιο των βιολογικών δικτύων. Η κατανόηση του πώς μπορούν να χρησιμοποιηθούν οι αλγόριθμοι στη βιοπληροφορική μπορεί να οδηγήσει σε νέες ιδέες και ανακαλύψεις στο πεδίο.

Co-funded by the Erasmus+ Programme of the European Union

Ασκηση

- Ασκηση 1

Ας υποθέσουμε μια πραγματική διαδικασία. Κάνεις μια πρόποση. Αναφέρετε τις εισόδους, τις εξόδους και τα βήματα που χρειάζονται.

- Άσκηση 2

Δημιουργήστε το δικό σας παράδειγμα και ακολουθήστε την ίδια διαδικασία.

Co-funded by the Erasmus+ Programme of the European Union

Περίληψη

- Οι αλγόριθμοι είναι καθορισμένες διαδικασίες, σημαντικές στην επιστήμη των υπολογιστών και σε συναφείς τομείς
- Ένας αλγόριθμος έχει εισόδους, εξόδους, δομές ελέγχου και λειτουργίες
- Σχεδιασμός Τεχνικής
  - Διαίρει και βασίλευε
  - Δυναμικός Προγραμματισμός
  - Άπληστοι Αλγόριθμοι
- Εφαρμογές στη Βιοπληροφορική, όπου μας βοηθούν οι αλγόριθμοι
  - Οραματιστείτε, αναλύστε και ερμηνεύστε τα βιολογικά δεδομένα
  - Δικτυακή ανάλυση βιολογικών δικτύων

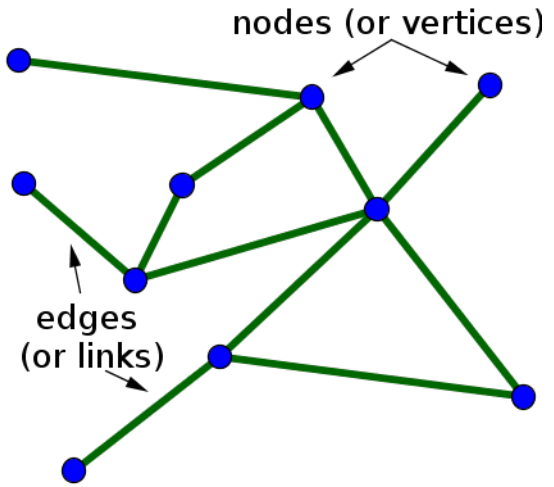
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμοι Γράφων και Βιοπληροφορική

- Οι αλγόριθμοι γραφημάτων διαδραματίζουν κρίσιμο ρόλο στη βιοπληροφορική, καθώς τα βιολογικά δεδομένα μπορούν συχνά να αναπαρασταθούν ως γραφήματα

- Οι αλγόριθμοι γραφημάτων είναι μια συλλογή αλγορίθμων που έχουν σχεδιαστεί για την επίλυση προβλημάτων σε γραφήματα

- Γράφοι
  - Μαθηματικές δομές που αποτελούνται από κόμβους και κορυφές
  - Αναπαριστά την αλληλεπίδραση μεταξύ των κόμβων



- Στη βιοπληροφορική, οι γράφοι μπορούν να χρησιμοποιηθούν για να αναπαραστήσουν μια ποικιλία βιολογικών δεδομένων, όπως τα δίκτυα αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης, τα ρυθμιστικά δίκτυα γονιδίων και τα μεταβολικά μονοπάτια.

- Αναλύοντας αυτά τα βιολογικά δίκτυα χρησιμοποιώντας αλγόριθμους γραφημάτων, οι ερευνητές μπορούν να αποκτήσουν γνώσεις για τις υποκείμενες βιολογικές διεργασίες και αλληλεπιδράσεις.

- Στις επόμενες διαφάνειες, θα εξερευνήσουμε μερικούς κοινούς αλγόριθμους γραφημάτων που χρησιμοποιούνται στη βιοπληροφορική και τις εφαρμογές τους.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

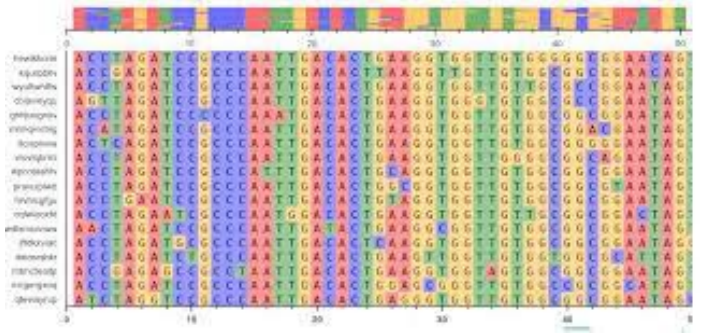
Αλγόριθμοι Δυναμικού Προγραμματισμού

- Ο δυναμικός προγραμματισμός είναι μια τεχνική για την επίλυση προβλημάτων με
  - αναλύοντάς τα σε μικρότερα υποπροβλήματα και
  - την επίλυσή τους ανεξάρτητα,
  - στη συνέχεια συνδυάζοντας τις λύσεις τους για την επίλυση του αρχικού προβλήματος. Μπορεί να είναι ιδιαίτερα χρήσιμο όταν τα υποπροβλήματα επικαλύπτονται.

- Στη βιοπληροφορική, οι αλγόριθμοι δυναμικού προγραμματισμού μπορούν να χρησιμοποιηθούν για εργασίες όπως η στοίχιση ακολουθιών

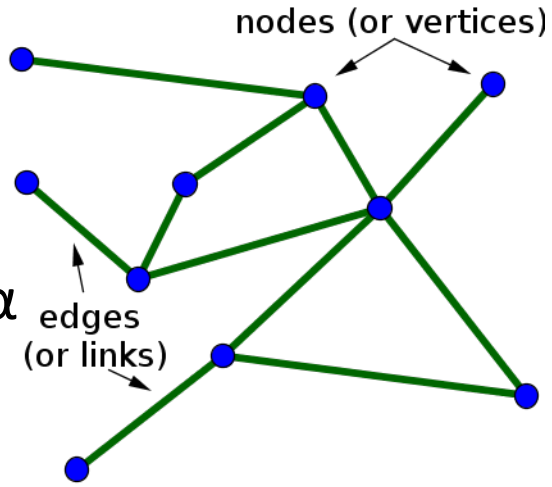
- όπου ο στόχος είναι να ευθυγραμμιστούν δύο ή περισσότερες αλληλουχίες DNA, RNA ή πρωτεΐνης

- για τον εντοπισμό περιοχών ομοιότητας.



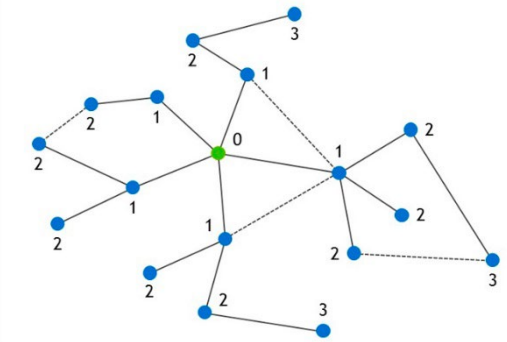
Παραδείγματα σε γραφήματα

- Ο δυναμικός προγραμματισμός είναι μια ισχυρή αλγοριθμική τεχνική που μπορεί να χρησιμοποιηθεί για την επίλυση πολύπλοκων προβλημάτων βελτιστοποίησης
- Συχνά εφαρμόζεται σε προβλήματα γραφημάτων που προκύπτουν στην ανάλυση βιολογικών δεδομένων.
- Παραδείγματα δυναμικού προγραμματισμού σε γραφήματα:
- Η εύρεση του
  - Το πιο σύντομο
  - Τα μεγαλύτερα μονοπάτια
- Ανάμεσα σε δύο κορυφές σε ένα γράφημα



Παράδειγμα

- Αυτοί οι αλγόριθμοι λειτουργούν με τον υπολογισμό της απόστασης μεταξύ των κόμβων σε ένα γράφημα,
- Η απόσταση ορίζεται ως ο αριθμός των άκρων κατά μήκος της διαδρομής μεταξύ των κόμβων.
- Ο αλγόριθμος της συντομότερης διαδρομής βρίσκει τη διαδρομή με την ελάχιστη απόσταση
- Ο αλγόριθμος μακρύτερης διαδρομής βρίσκει τη διαδρομή με τη μέγιστη απόσταση.



Ο αλγόριθμος του Dijkstra

- Δημιουργήστε ένα σύνολο κόμβων χωρίς επίσκεψη και ορίστε την απόσταση του κόμβου έναρξης στο 0.
- Ενώ υπάρχουν κόμβοι που δεν έχουν επισκεφτεί, επιλέξτε τον κόμβο με τη μικρότερη απόσταση και σημειώστε τον ως επισκέψιμο.
- Για κάθε γείτονα του επιλεγμένου κόμβου, υπολογίστε την απόσταση από τον κόμβο έναρξης μέσω του επιλεγμένου κόμβου.
- Εάν η υπολογισμένη απόσταση είναι μικρότερη από την τρέχουσα απόσταση του γείτονα, ενημερώστε την απόσταση.
- Επαναλάβετε τα βήματα 2-4 έως ότου γίνει επίσκεψη στον τελικό κόμβο ή δεν υπάρχουν άλλοι κόμβοι που δεν έχουν επισκεφτεί.
- Η απόσταση του τερματικού κόμβου είναι η συντομότερη διαδρομή από τον κόμβο έναρξης.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

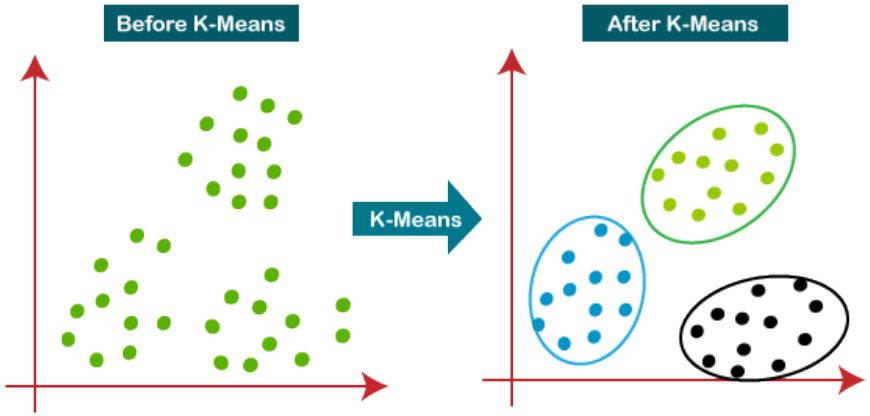
Τυχαιοποιημένοι αλγόριθμοι

- Οι τυχαιοποιημένοι αλγόριθμοι είναι αλγόριθμοι που χρησιμοποιούν τυχαίους αριθμούς για την επίλυση υπολογιστικών προβλημάτων.
- Αυτοί οι αλγόριθμοι είναι χρήσιμοι σε καταστάσεις όπου τα δεδομένα εισόδου είναι πολύ μεγάλα ή πολύ περίπλοκα για να τα χειριστούν αποτελεσματικά οι ντετερμινιστικοί αλγόριθμοι.
- Στη βιοπληροφορική, τυχαιοποιημένοι αλγόριθμοι μπορούν να χρησιμοποιηθούν για εργασίες όπως η ομαδοποίηση και η ταξινόμηση.
- Γρήγορο και αποδοτικό στη μνήμη.
- λιγότερο προβλέψιμο και πιο δύσκολο στην ερμηνεία από αυτά των ντετερμινιστικών αλγορίθμων.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

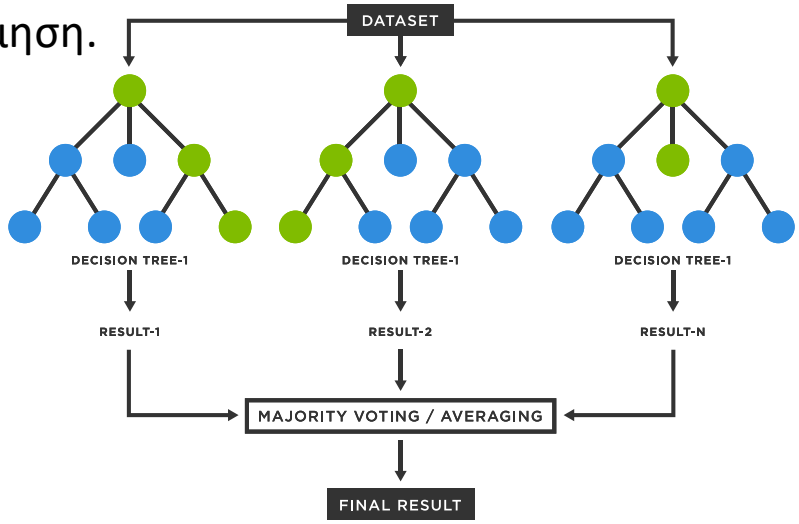
K-Means Clustering

- Ένα παράδειγμα ενός τυχαιοποιημένου αλγορίθμου στη βιοπληροφορική είναι η ομαδοποίηση k-means.
- Αυτός ο αλγόριθμος χρησιμοποιείται για να χωρίσει ένα σύνολο σημείων δεδομένων σε k συμπλέγματα, όπου κάθε σημείο δεδομένων ανήκει στο σύμπλεγμα με τον πλησιέστερο μέσο όρο.
- Ο αλγόριθμος λειτουργεί επιλέγοντας τυχαία k αρχικά κεντροειδή συμπλέγματος, εκχωρώντας κάθε σημείο δεδομένων στο πλησιέστερο κέντρο,
- Στη συνέχεια, ενημερώνονται τα κεντροειδή με βάση τον μέσο όρο των σημείων που έχουν εκχωρηθεί σε κάθε σύμπλεγμα .



Τυχαίο Δάσος

- Ένα άλλο παράδειγμα ενός τυχαιοποιημένου αλγορίθμου στη βιοπληροφορική είναι ο αλγόριθμος Random Forest για ταξινόμηση.
- Αυτός ο αλγόριθμος χρησιμοποιεί ένα σύνολο δέντρων αποφάσεων, το καθένα εκπαιδευμένο σε ένα τυχαία επιλεγμένο υποσύνολο δεδομένων, για να ταξινομήσει νέα σημεία δεδομένων.
- Ο αλγόριθμος συνδυάζει τις προβλέψεις των μεμονωμένων δέντρων για να κάνει μια τελική ταξινόμηση.



Αλγόριθμοι προσέγγισης

- Οι αλγόριθμοι προσέγγισης είναι αλγόριθμοι που βρίσκουν μια κατά προσέγγιση λύση σε ένα πρόβλημα που είναι δύσκολο ή αδύνατο να λυθεί ακριβώς.

- Στη βιοπληροφορική, αυτοί οι αλγόριθμοι χρησιμοποιούνται για την επίλυση προβλημάτων βελτιστοποίησης που σχετίζονται με την πρόβλεψη της δομής των πρωτεϊνών, την πρόβλεψη της γονιδιακής λειτουργίας και άλλους τομείς.

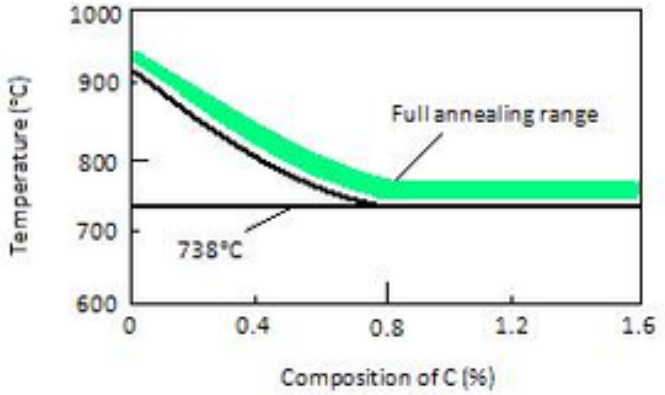
- Συχνά θυσιάζει την ακρίβεια για την ταχύτητα.

- Οι αλγόριθμοι προσέγγισης μπορούν να παρέχουν μια σχεδόν βέλτιστη λύση χρησιμοποιώντας μια ευρετική προσέγγιση για την εξερεύνηση ενός υποσυνόλου των πιθανών διαμορφώσεων.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Προσομοίωση ανόπτησης

- Ένα άλλο παράδειγμα είναι ο αλγόριθμος προσομοίωσης ανόπτησης
- R andomized αλγόριθμος που μπορεί να χρησιμοποιηθεί για την εύρεση σχεδόν βέλτιστων λύσεων σε πολύπλοκα προβλήματα βελτιστοποίησης.
- Αυτός ο αλγόριθμος είναι εμπνευσμένος από τη διαδικασία της ανόπτησης στη μεταλλουργία
- όπου ένα μέταλλο
  - θερμός
  - στη συνέχεια ψύχθηκε αργά
- να αυξήσει τη δύναμη και την αντοχή τ



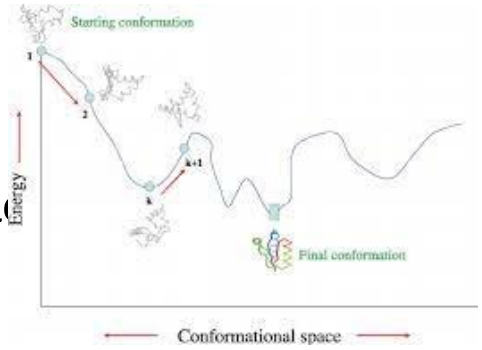
Προσομοίωση ανόπτησης

- Μπορεί να εφαρμοστεί σε ένα ευρύ φάσμα προβλημάτων βελτιστοποίησης στη βιοπληροφορική, όπως η πρόβλεψη δομής πρωτεϊνών, η αλληλουχία DNA και η συναρμολόγηση γονιδιώματος.

- Στην πρόβλεψη της δομής της πρωτεΐνης, η προσομοίωση ανόπτησης χρησιμοποιείται συχνά για να βρεθεί η διαμόρφωση της πρωτεΐνης που έχει τη χαμηλότερη ενέργεια.

- Ο αλγόριθμος ξεκινά με μια τυχαία διαμόρφωση

- αναζητά επαναληπτικά για καλύτερες διαμορφώσει



Co-funded by the Erasmus+ Programme of the European Union

Βήματα

- Διατύπωση του προβλήματος: Ορίστε το πρόβλημα γραφήματος που θέλετε να λύσετε, όπως η ομαδοποίηση γραφημάτων.
- Ορισμός της αντικειμενικής συνάρτησης: Δημιουργήστε μια αντικειμενική συνάρτηση που ποσοτικοποιεί την ποιότητα μιας λύσης στο πρόβλημα του γραφήματος
- Αρχικοποίηση της λύσης: Ξεκινήστε με μια αρχική λύση, η οποία θα μπορούσε να είναι τυχαία
- Ορίστε τη γειτονιά: Ορίστε ένα σύνολο πιθανών λύσεων, που μπορούν να προσεγγιστούν από την τρέχουσα λύση. Σε ένα πλαίσιο γραφήματος, αυτό θα μπορούσε να περιλαμβάνει την τροποποίηση της εκχώρησης τιμών σε κόμβους ή ακμές γραφημάτων.
- Αποδοχή ή απόρριψη νέων λύσεων: Συγκρίνετε την ποιότητα της νέας λύσης με την τρέχουσα λύση. Εάν η νέα λύση βελτιώνει την τιμή της αντικειμενικής συνάρτησης, αποδεχτείτε την ως τη νέα τρέχουσα λύση. Εάν είναι χειρότερο, αποδεχτείτε το με μια ορισμένη πιθανότητα βάσει του προγράμματος προσομοίωσης ανόπτησης
- Εκτελέστε επαναλήψεις: Εξερευνήστε επαναληπτικά το χώρο της λύσης δημιουργώντας νέες λύσεις από την τρέχουσα λύση. Σε κάθε επανάληψη, επιλέξτε μια τυχαία γειτονική λύση και αξιολογήστε την ποιότητά της χρησιμοποιώντας τη συνάρτηση αντικειμενικού.

Co-funded by the Erasmus+ Programme of the European Union

Ασκηση

- Συνδυάστε όσα μάθατε και
  - Ορίστε τις κατηγορίες των αλγορίθμων που αναφέρονται στο μάθημα
  - Ονομάστε μια ομοιότητα ή μια διαφορά μεταξύ τους
  - Επιλέξτε ένα παράδειγμα αλγορίθμου και αναφέρετε τα βήματά του

Περίληψη

- Οι αλγόριθμοι είναι απαραίτητα εργαλεία της βιοπληροφορικής
- Οι αλγόριθμοι γραφημάτων είναι σημαντικά εργαλεία που μας βοηθούν να κατανοήσουμε και να οπτικοποιήσουμε τη δομή των βιολογικών οντοτήτων
- Κατηγορίες αλγορίθμων
  - Δυναμικός Προγραμματισμός (ανάλυση σε μικρότερα προβλήματα)
    - Τα συντομότερα μονοπάτια
    - Τα μακρύτερα μονοπάτια
  - Τυχαία (χρήση τυχαίων αριθμών για επίλυση)
    - K-means
    - Τυχαίο Δάσος
  - Προσέγγιση (Βρείτε κατά προσέγγιση λύσεις)
    - Προσομοίωση ανόπτησης

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

Θέμα 5. Αλγόριθμοι στη Βιοπληροφορική

Μάθημα 3. Ιδιότητες δικτύου

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Μοντελοποίηση Δικτύου

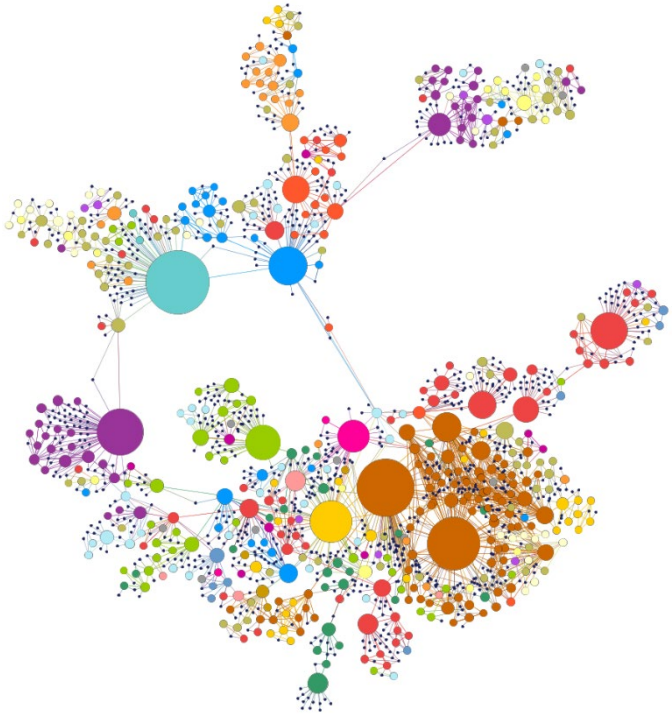
- Η μοντελοποίηση δικτύου είναι η μελέτη πολύπλοκων συστημάτων που αποτελούνται από διασυνδεδεμένα στοιχεία, που αναπαρίστανται ως κόμβοι και ακμές.

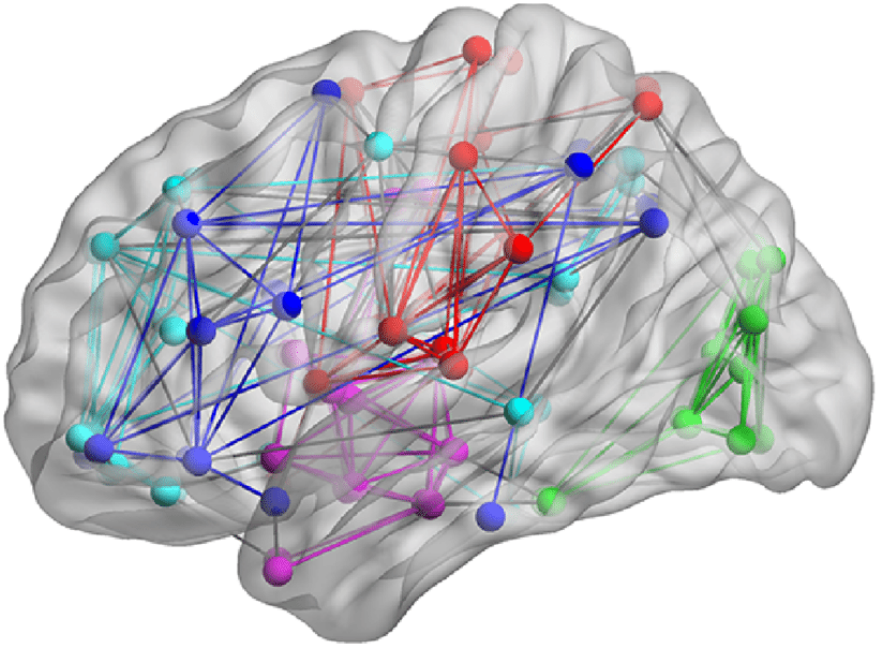
- Ένα σημαντικό εργαλείο στη βιοπληροφορική επειδή τα βιολογικά συστήματα είναι εγγενώς πολύπλοκα και αλληλένδετα.

- Με τη μοντελοποίηση των βιολογικών συστημάτων ως δικτύων, οι ερευνητές μπορούν να αποκτήσουν γνώσεις για τη δομή, τη λειτουργία και την εξέλιξή τους.

- Σε αυτό το μάθημα, θα εισαγάγουμε βασικές έννοιες στη μοντελοποίηση δικτύου

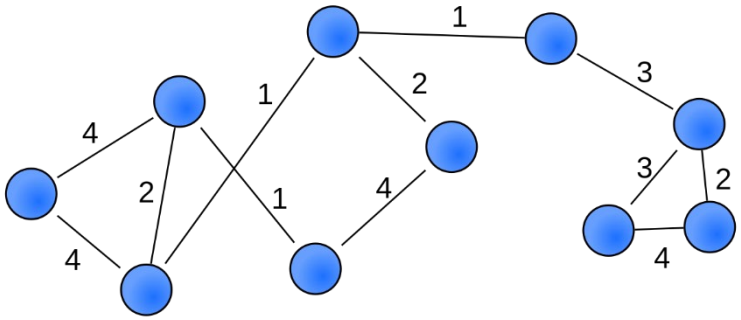
- Ένα δίκτυο αποτελείται από δύο βασικά στοιχεία: κόμβους και ακμές.
- Οι κόμβοι αντιπροσωπεύουν μεμονωμένες οντότητες στο δίκτυο, ενώ οι ακμές αντιπροσωπεύουν τις σχέσεις μεταξύ των κόμβων.





Σταθμισμένα δίκτυα

- Σε ένα σταθμισμένο δίκτυο, σε κάθε άκρο εκχωρείται ένα βάρος ή μια τιμή που αντιπροσωπεύει τη δύναμη ή τη σημασία της σύνδεσης μεταξύ δύο κόμβων.
- Τα σταθμισμένα δίκτυα χρησιμοποιούνται για την καταγραφή της ισχύος των αλληλεπιδράσεων μεταξύ κόμβων σε διάφορα βιολογικά συστήματα.
- Ακριβέστερη αναπαράσταση βιολογικών συστημάτων από τα μη σταθμισμένα δίκτυα, καθώς λαμβάνουν υπόψη την ισχύ των αλληλεπιδράσεων μεταξύ των κόμβων.
- Παράδειγμα
  - Δίκτυα αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης (PPIN): Στα PPIN, το βάρος μιας άκρης μπορεί να αντιπροσωπεύει τη βαθμολογία εμπιστοσύνης της αλληλεπίδρασης μεταξύ δύο πρωτεϊνών.



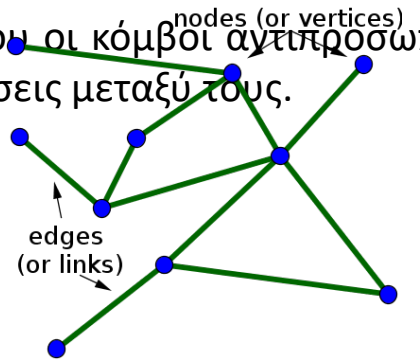
Μη σταθμισμένα δίκτυα

- Τα μη σταθμισμένα δίκτυα αναφέρονται σε δίκτυα όπου όλες οι ακμές έχουν την ίδια σημασία και δεν τους εκχωρείται κανένα βάρος.
- Σε αντίθεση με τα σταθμισμένα δίκτυα, τα μη σταθμισμένα δίκτυα καταγράφουν μόνο την παρουσία ή την απουσία συνδέσεων μεταξύ κόμβων.
- Παραδείγματα
  - Μεταβολικά δίκτυα: Αυτά τα δίκτυα αντιπροσωπεύουν τις μεταβολικές οδούς μέσα σε ένα κύτταρο και συχνά μοντελοποιούνται ως μη σταθμισμένα δίκτυα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

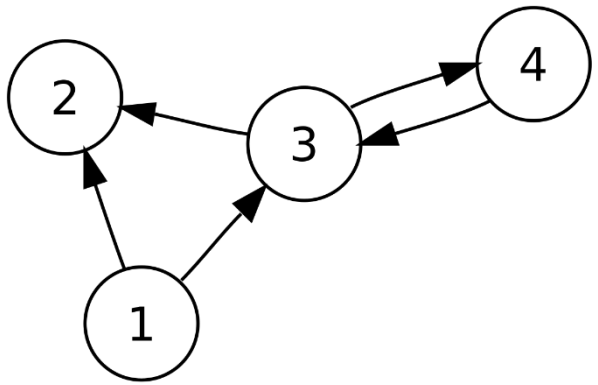
Μη κατευθυνόμενα δίκτυα

- Σε ένα μη κατευθυνόμενο δίκτυο, οι άκρες δεν έχουν κατεύθυνση ή συγκεκριμένο προσανατολισμό. Απλώς αντιπροσωπεύουν μια σύνδεση μεταξύ δύο κόμβων.
- Αυτό σημαίνει ότι εάν ο κόμβος Α είναι συνδεδεμένος με τον κόμβο Β, τότε ο κόμβος Β συνδέεται επίσης με τον κόμβο Α.
- Παραδείγματα
  - Κοινωνικά δίκτυα: Δίκτυα φιλίας, όπου δύο άτομα που θεωρούν τους εαυτούς τους φίλους συνδέονται με ένα μη κατευθυνόμενο άκρο.
  - Βιολογικά δίκτυα: Μεταβολικά δίκτυα, όπου οι κόμβοι αντιπροσωπεύουν μεταβολίτες και οι ακμές αντιπροσωπεύουν τις αντιδράσεις μεταξύ τους.



Κατευθυνόμενα Δίκτυα

- Ένα κατευθυνόμενο δίκτυο είναι ένας τύπος δικτύου στο οποίο οι ακμές έχουν κατεύθυνση και αντιπροσωπεύουν μια μονόδρομη σχέση μεταξύ των κόμβων.
- Παραδείγματα:
  - Κοινωνικά δίκτυα: Twitter, όπου οι χρήστες μπορούν να ακολουθούν άλλους, αλλά αυτοί οι άλλοι μπορεί να μην τους ακολουθούν απαραίτητα
  - Ρυθμιστικά δίκτυα γονιδίων: όπου ένα γονίδιο μπορεί να ρυθμίσει την έκφραση ενός άλλου γονιδίου, αλλά όχι το αντίστροφο



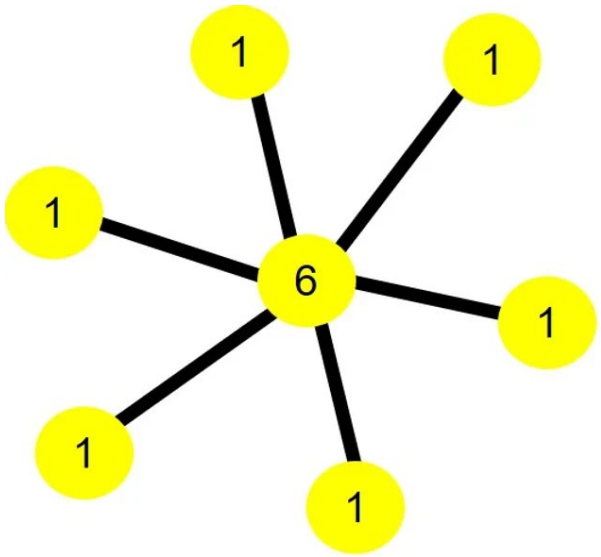
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Βαθμός (degree) κόμβου

- Ο βαθμός κόμβου είναι ένα μέτρο του αριθμού των συνδέσεων που έχει ένας κόμβος σε ένα δίκτυο.
- Υπολογίζεται ως ο αριθμός των ακμών που συνδέονται με έναν κόμβο, γνωστό και ως βαθμός του.

Υπολογισμός του βαθμού κεντρικότητας αυτού του δικτύου:

- Οι περιφερειακοί κόμβοι έχουν βαθμό 1
- Το κεντρικό 6

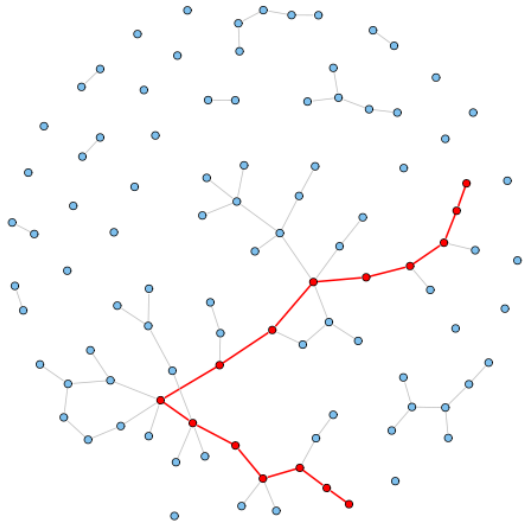


- Ο βαθμός είναι μια ευρέως χρησιμοποιούμενη μέτρηση στην ανάλυση δικτύου που μετρά τον αριθμό των συνδέσεων που έχει ένας κόμβος με άλλους κόμβους του δικτύου.
- Είναι σημαντικό γιατί παρέχει πληροφορίες για τη σχετική σημασία των κόμβων σε ένα δίκτυο.
- Οι κόμβοι με υψηλό βαθμό θεωρούνται πιο σημαντικοί στο δίκτυο, καθώς έχουν μεγάλο αριθμό συνδέσεων με άλλους κόμβους

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

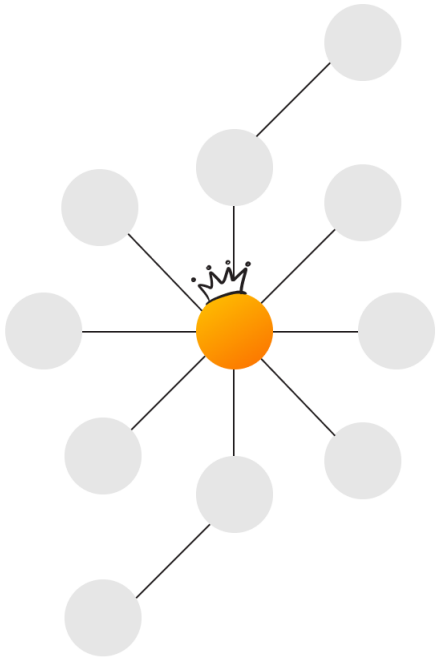
Το μικρότερο μήκος διαδρομής

- Το μικρότερο μήκος διαδρομής μεταξύ δύο κόμβων σε ένα δίκτυο είναι ο ελάχιστος αριθμός ακμών που πρέπει να διασχιστούν για να μεταβείτε από τον έναν κόμβο στον άλλο.
- Το μήκος της συντομότερης διαδρομής είναι ένα σημαντικό μέτρο της αποδοτικότητας και της ευρωστίας του δικτύου.
- Μπορεί να βοηθήσει στον εντοπισμό βασικών κόμβων ή ακμών που είναι σημαντικοί για τη συνολική συνδεσιμότητα του δικτύου.



Ενδιάμεση (betweenness) Κεντρικότητα

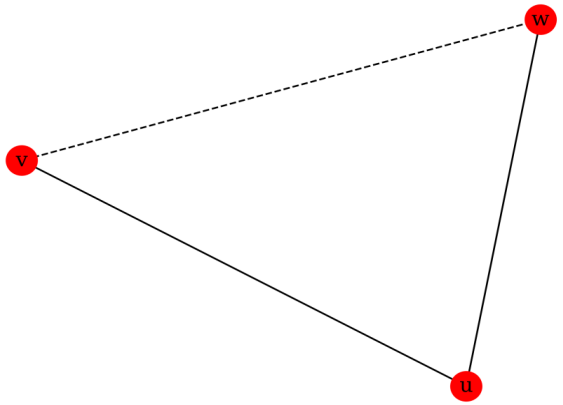
- Η ενδιάμεση κεντρικότητα μετράει πόσο συχνά ένας κόμβος βρίσκεται στη συντομότερη διαδρομή μεταξύ άλλων κόμβων στο δίκτυο.
- Ένας κόμβος με υψηλή ενδιάμεση κεντρικότητα είναι σημαντικός για τη διατήρηση της ροής πληροφοριών στο δίκτυο.
- Μπορεί επίσης να βοηθήσει στον εντοπισμό βασικών κόμβων που, εάν αφαιρεθούν, θα μπορούσαν να επηρεάσουν σημαντικά τη συνολική δομή και λειτουργία του δικτύου.



A node with a high degree centrality score

Συντελεστής ομαδοποίησης

- Ο συντελεστής ομαδοποίησης μετρά τον βαθμό στον οποίο οι κόμβοι σε ένα δίκτυο τείνουν να συγκεντρώνονται μεταξύ τους. Ορίζεται ως το κλάσμα των ζευγών των γειτόνων ενός κόμβου που είναι επίσης γείτονες μεταξύ τους.
- Ο συντελεστής ομαδοποίησης ενός κόμβου υπολογίζεται διαιρώντας τον αριθμό των ακμών μεταξύ των γειτόνων του κόμβου με τον συνολικό αριθμό των πιθανών ακμών μεταξύ τους.
- Οι υψηλοί συντελεστές ομαδοποίησης μπορούν να υποδεικνύουν την παρουσία στενά συνδεδεμένων ομάδων ή κοινοτήτων μέσα σε ένα δίκτυο, ενώ οι χαμηλοί συντελεστές ομαδοποίησης μπορεί να υποδηλώνουν μια πιο τυχαία ή αποδιοργανωμένη δομή



Θεωρητική μοντελοποίηση δικτύου

- Ενώ πολλά δίκτυα στη βιοπληροφορική βασίζονται σε δεδομένα του πραγματικού κόσμου
- Ωστόσο, τα θεωρητικά μοντέλα παρέχουν ένα χρήσιμο πλαίσιο για την κατανόηση της συμπεριφοράς και των ιδιοτήτων του δικτύου.
- Μιμούνται τις ιδιότητες των πραγματικών δικτύων
- Υπάρχουν πολλοί διαφορετικοί τύποι θεωρητικών μοντέλων δικτύου, το καθένα με το δικό του σύνολο υποθέσεων και χαρακτηριστικών.

Co-funded by the Erasmus+ Programme of the European Union

Σημασια

- Κατανόηση των βιολογικών συστημάτων: Τα δίκτυα παρέχουν ένα ισχυρό εργαλείο για την κατανόηση πολύπλοκων βιολογικών συστημάτων, συμπεριλαμβανομένων των δικτύων γενετικών, μεταβολικών και ασθενειών.
- Πρόβλεψη συμπεριφοράς δικτύου: Αναλύοντας τις ιδιότητες του δικτύου και προσομοιώνοντας θεωρητικά μοντέλα δικτύων, οι ερευνητές μπορούν να προβλέψουν πώς θα συμπεριφέρονται τα βιολογικά δίκτυα υπό διαφορετικές συνθήκες.
- Προσδιορισμός βασικών εξαρτημάτων: Οι ιδιότητες του δικτύου όπως η κεντρικότητα βαθμών, η μεταξύ τους κεντρική θέση και ο συντελεστής ομαδοποίησης μπορούν να χρησιμοποιηθούν για τον εντοπισμό βασικών στοιχείων εντός βιολογικών δικτύων, όπως σημαντικά γονίδια ή πρωτεΐνες.
- Προώθηση της Ιατρικής Ακριβείας: Η μοντελοποίηση δικτύου μπορεί να βοηθήσει στον εντοπισμό δικτύων ασθενειών ειδικά για κάθε άτομο, τα οποία μπορούν να χρησιμοποιηθούν για την ανάπτυξη εξατομικευμένων σχεδίων θεραπείας για ασθενείς με βάση τους μοναδικούς γενετικούς και περιβαλλοντικούς παράγοντες τους.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

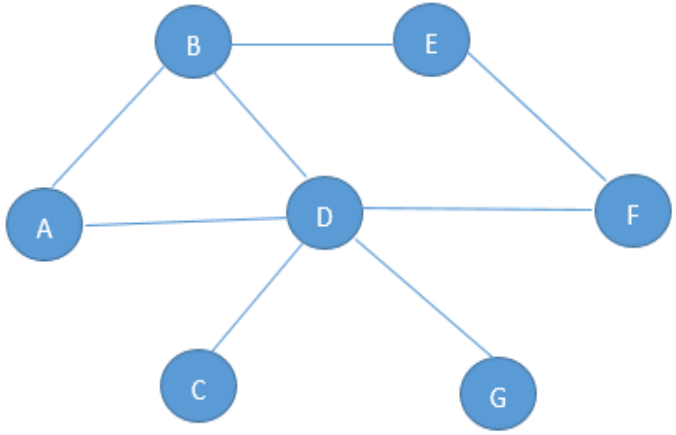
Ασκηση

- Ασκηση 1

Ορίστε τις ιδιότητες δικτύου που αναφέρονται σε αυτό το μάθημα. Τι σημαίνουν οι υψηλές/χαμηλές τιμές αυτών των ιδιοτήτων για αυτό το δίκτυο;

- Άσκηση 2

Δεδομένου του δικτύου στην εικόνα,



υπολογίστε:

- Ο βαθμός των κόμβων

Α, Ε, Δ.

- Το μικρότερο μήκος διαδρομής μεταξύ:

Οι κόμβοι (A,G), (D,E)

Co-funded by the Erasmus+ Programme of the European Union

Περίληψη

- Σε αυτό το μάθημα έχουμε εισαγάγει τις βασικές έννοιες του δικτύου
  - Κατευθυνόμενα Δίκτυα
  - Μη κατευθυνόμενα δίκτυα
  - Σταθμισμένα Δίκτυα
  - Μη σταθμισμένα δίκτυα
- Έχουμε ορίσει ιδιότητες δικτύου
  - Κεντρικότητα πτυχίου
  - Κεντρικότητα μεταξύ
  - Μήκος Συντομότερων Διαδρομών
  - Συντελεστής Ομαδοποίησης
- Θεωρητική μοντελοποίηση δικτύου
- Σημασία της μοντελοποίησης δικτύου στη Βιοπληροφορική

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

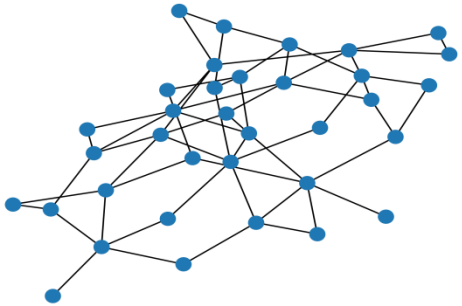
Θέμα 5. Αλγόριθμοι στη Βιοπληροφορική

Μάθημα 4. Τυχαία δίκτυα

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Μοντέλα Δικτύων

- Επιστήμη δικτύου: Αναπαράγει ιδιότητες πραγματικών δικτύων.
- Η απλούστερη τοπολογία δικτύου είναι το τυχαίο δίκτυο.
- Οι σύνδεσμοι μεταξύ των κόμβων επιλέγονται τυχαία.
- Κάθε αλληλεπίδραση έχει την ίδια πιθανότητα ύπαρξης.

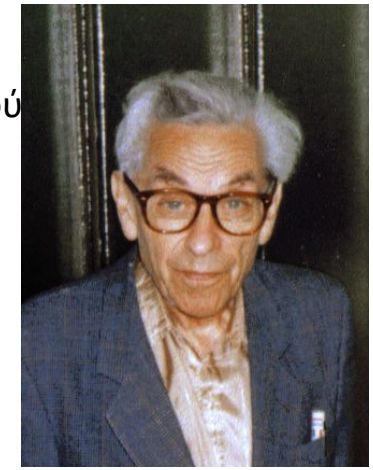


Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Erdős - Rényi Model

- Μοντέλο G(N, L).
  - Ν κόμβοι με συνδέσμους L που τοποθετούνται τυχαία (που χρησιμοποιού και Rényi)
- G(N, p) Μοντέλο
  - Ν κόμβοι και κάθε ακμή έχουν την ίδια πιθανότητα p να υπάρχουν
- Τα τυχαία γραφήματα ονομάζονται Erdős - Rényi Networks για συνεισφορά τους
- Για την κατασκευή του δικτύου
  - Ξεκινήστε με Ν κόμβους
  - Δημιουργήστε έναν τυχαίο αριθμό για ένα ζεύγος κόμβων.
  - Εάν ο αριθμός είναι μικρότερος από το p, συνδέστε τους συνδέσμους.
  - Επαναλάβετε για κάθε ζεύγος κόμβων στο δίκτυο



Φιλαράκος Erdős

Αλφρεδος Rényi

Co-funded by the Erasmus+ Programme of the European Union

G(N, p) Μοντέλο

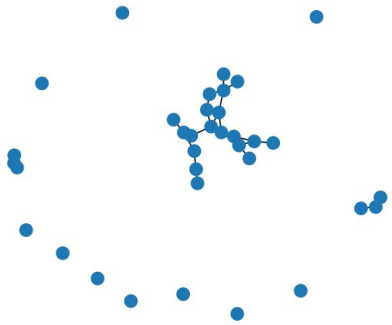
- Όλα τα ζεύγη κόμβων είναι $L_{\max} = \binom{N}{2} = \frac{N(N-1)}{2}$
- Αναμενόμενος αριθμός συνδέσμων: $< L >= p\frac{N(N-1)}{2}$
- Μέσος βαθμός: $< k >= \frac{2 < L >}{N} = p(N-1)$

Όπου k_i είναι ο αριθμός των κόμβων, με τον κόμβο i συνδέεται

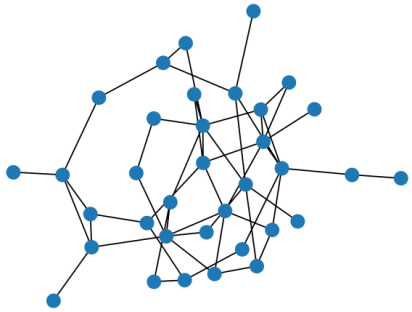
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

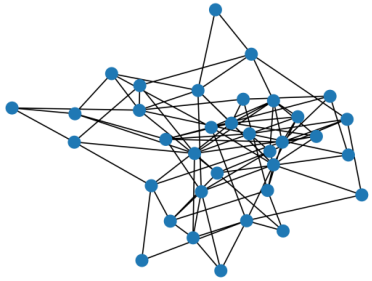
Πιθανότητα σύνδεσης



p = 0,05



p = 0,1



p = 0,15

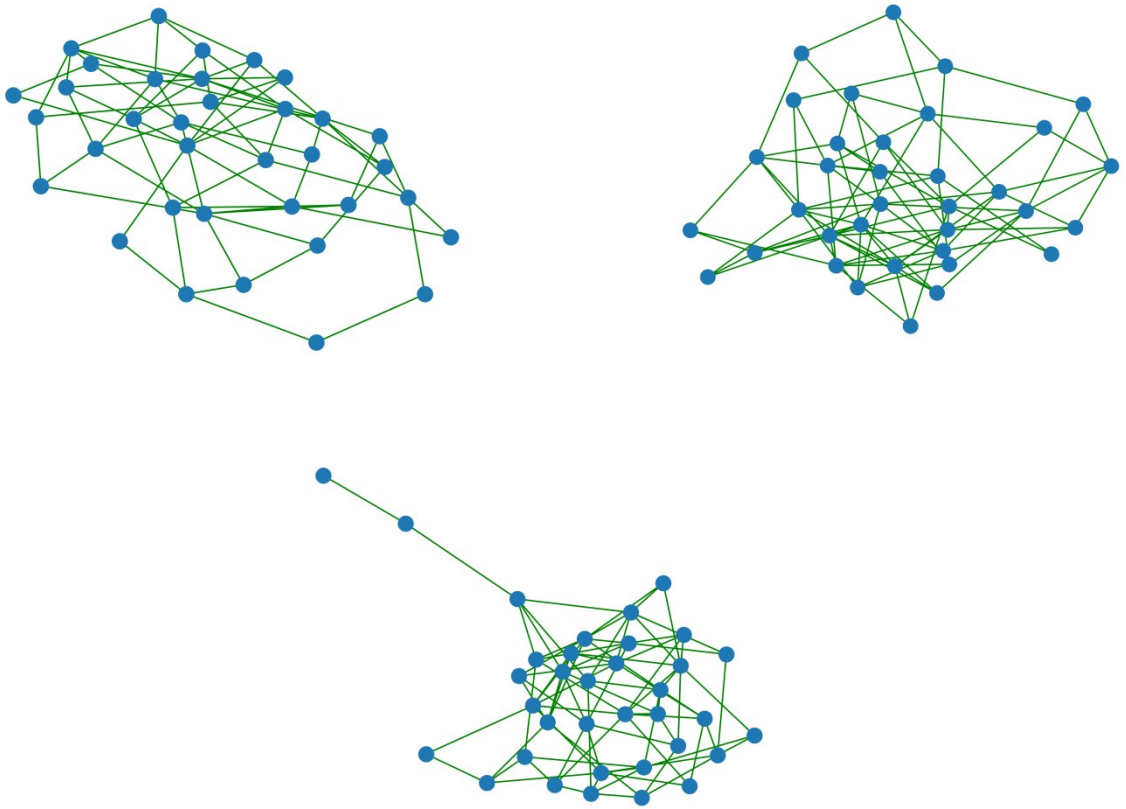
Στοχαστικό μοντέλο

- Η πιθανότητα p είναι η παράμετρος που ελέγχει την πυκνότητα του δικτύου
- Ωστόσο, το G(N, p) είναι ένα στοχαστικό μοντέλο, που δημιουργείται από ψευδοτυχαίους αριθμούς
- Ο σπόρος παίζει σημαντικό ρόλο
- Παρακάτω παρουσιάζουμε 3 δίκτυα με N=35, p = 0,15 και διαφορετικούς σπόρους

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

N=35, ρ=0,15



Co-funded by the Erasmus+ Programme of the European Union

Τυχαιότητα

- p και N είναι τα ίδια
- Ωστόσο, αυτά τα γραφήματα έχουν διαφορετικό αριθμό συνολικών συνδέσμων
- Παρόμοια αλλά όχι ακριβώς ίδια δομή
- Αν ο πρώτος ψευδοτυχαίος αριθμός (σπόρος) και η γεννήτρια τυχαίων αριθμών ήταν ίδια, θα είχαμε ακριβώς το ίδιο δίκτυο σε κάθε υλοποίηση

Κατανομή βαθμών

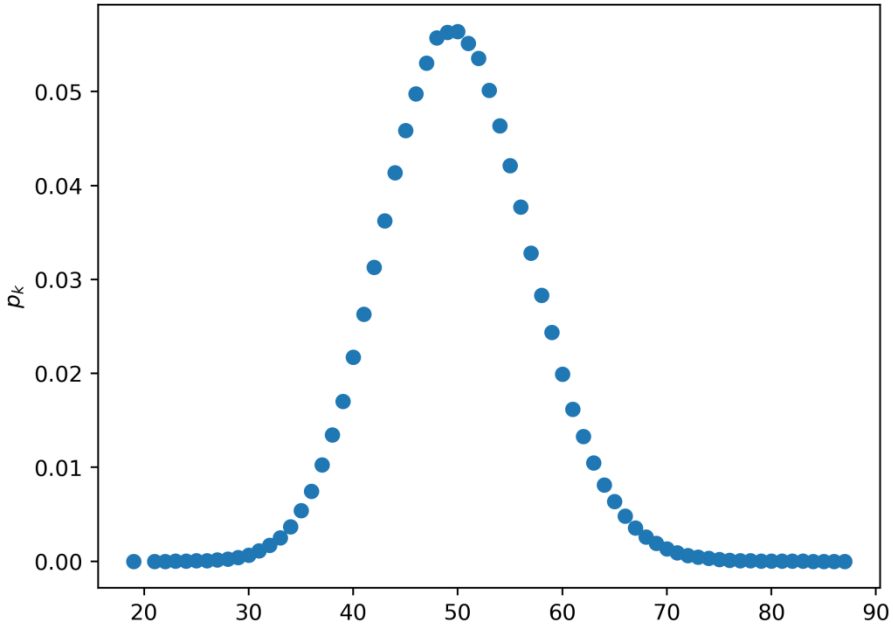
- Πιθανότητα να υπάρχουν k σύνδεσμοι: p^k
- Πιθανότητα να λείπουν οι υπόλοιποι σύνδεσμοι: $(1-p)^{N-1-k}$
- τρόποι επιλογής k άλλες κορυφές. $\binom{N-1}{k}$
- Πιθανότητα σύνδεσης με k άλλους κόμβους:

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

- Διωνυμική κατανομή

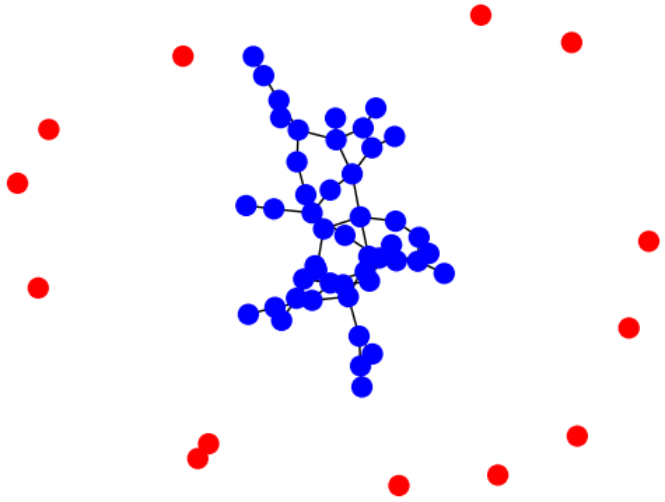
Co-funded by the Erasmus+ Programme of the European Union

Κατανομή βαθμών



Κατανομή βαθμών για ένα τυχαίο δίκτυο με N = 10000, p = 0,005, με μέσο όρο για 100 πραγματοποιήσεις. Η μορφή της διανομής είναι διωνυμική

Γιγαντιαίο συστατικό

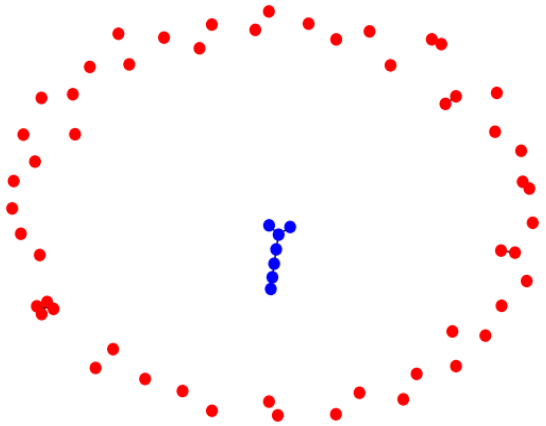


- Όταν το p είναι αρκετά μεγάλο, ένα μεγάλο μέρος του δικτύου είναι συνδεδεμένο

- Αυτό το κλάσμα ονομάζεται Giant Component (Οι μπλε κόμβοι στην εικόνα)

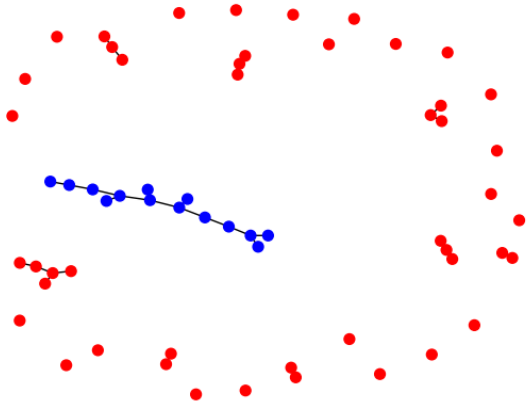
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Συντελεστής ομαδοποίησης



•<k> < 1

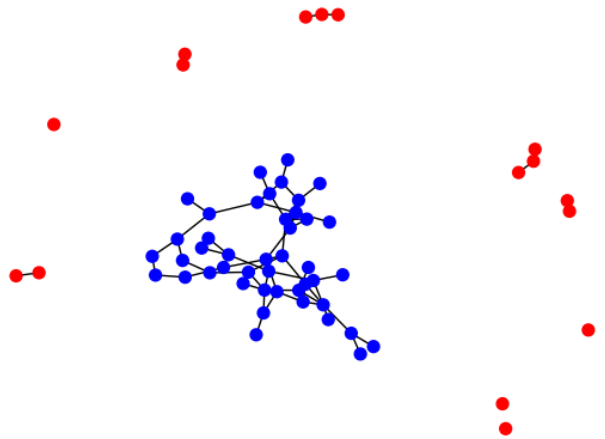
•Χωρίς γιγάντιο στοιχείο



•<k> = 1

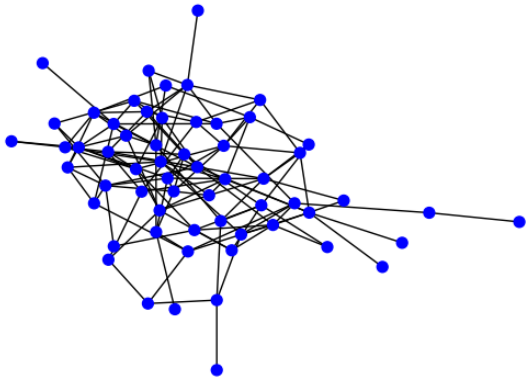
•Χωρίς γιγάντιο στοιχείο

Θεωρητική μοντελοποίηση δικτύου



•<k> > 2

•Ενιαίο γιγάντιο εξάρτημα



•<k> >> ln (N)

•Ενιαίο γιγάντιο εξάρτημα

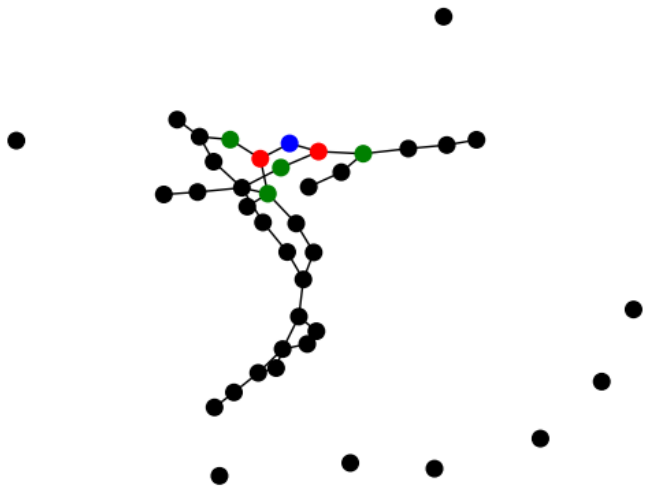
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Μήκη διαδρομής

•Ένας τυχαίος κόμβος (μπλε κόμβος) έχει κατ μέσο όρο

- <k> γείτονες στην απόσταση 1 (κόκκινο κόμβοι)
- <k> ^{2} γείτονες στην απόσταση 2 (πράσινο κόμβοι)
- <k> $^{\delta}$ γείτονες σε απόσταση d



N = 40, ρ = 0,051

•Το μικρότερο μήκος διαδρομής

Μια διαδρομή μεταξύ δύο κόμβων με τον ελάχιστο αριθμό ακμών

Co-funded by the Erasmus+ Programme of the European Union

Μικρός Κόσμος

- Μέση απόσταση <d>

  Μέση απόσταση μεταξύ όλων των ζευγών κόμβων

- Διάμετρος δικτύου $d_{μέγ}$

  Η μικρότερη απόσταση μεταξύ των πιο απομακρυσμένων κόμβων

- Τα <d> και d_{max} είναι μικρά στα δίκτυα Erdős - Rényi
  - Μικρά παγκόσμια φαινόμενα

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ασκηση 1

- Δημιουργήστε ένα δίκτυο Erdős–Rényi με N = 10000 κόμβους. Εφαρμόστε τον κανόνα ότι μεταξύ δύο κόμβων υπάρχει πιθανότητα p = 1/6 να υπάρχει σύνδεση.
- Βρείτε τον αριθμό των k συνδέσεων κάθε κόμβου. Θα είναι ένας πίνακας με N = 1000 τιμές ακεραίων.
- Να σχηματίσετε μια γραφική παράσταση της κατανομής του k, P(k), σε συνάρτηση με το k και να υπολογίσετε τη μέση τιμή του k. Τα αποτελέσματα θα πρέπει να είναι κατά μέσο όρο 1000 προσομοιώσεις.
- Κάντε το ίδιο για ένα δίκτυο με N = 100000 κόμβους.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Άσκηση 2

- Δημιουργήστε δίκτυα Erdős–Rényi με N=10000 κόμβους ακολουθώντας την ίδια διαδικασία όπως στην άσκηση 1, για <k> = 1, 3, 15.
- Η πιθανότητα p θα υπολογιστεί ως συνάρτηση του <k> χρησιμοποιώντας τις σχέσεις που παρουσιάζονται στο powerpoint.
- Για κάθε <k> δημιουργήστε 1000 υλοποιήσεις. Για κάθε υλοποίηση υπολογίστε τον αριθμό των κόμβων που ανήκουν στο γιγάντιο συστατικό.
- Σχεδιάστε την κατανομή του σχετικού μεγέθους του μεγαλύτερου συμπλέγματος (κόμβοι που ανήκουν στη γιγαντιαία συνιστώσα ως κλάσμα στο σύνολο των κόμβων) για κάθε <k>.

Co-funded by the Erasmus+ Programme of the European Union

Περίληψη

- Σε αυτό το μάθημα εισαγάγαμε το μοντέλο τυχαίου δικτύου και επικεντρωθήκαμε στις ιδιότητές του
- Οι ιδιότητες που αναλύθηκαν ήταν:
  - Κατανομή πτυχίων
  - Συντελεστής ομαδοποίησης
  - Γιγαντιαίο συστατικό
  - Μήκη διαδρομής
- Μικρός κόσμος

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

Θέμα 5. Αλγόριθμοι στη Βιοπληροφορική

Μάθημα 5. Δίκτυα μικρού κόσμου

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Το πείραμα του Μίλγκραμ

- Τα πραγματικά κοινωνικά δίκτυα έχουν μικρό μέσο μήκος διαδρομής.
- Για να το επιβεβαιώσει, ο ψυχολόγος Stanley Milgram πραγματοποίησε ένα πείραμα.
- Έστειλε 296 πακέτα σε τυχαία επιλεγμένα άτομα από την Ομάχα

Στάνλεϊ Μίλγκραμ

(φωτογραφία από τη Wikipedia)

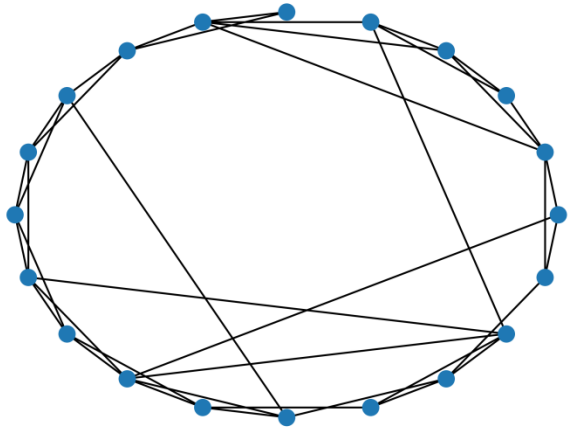
Το πείραμα του Μίλγκραμ

- Κάθε άτομο είχε πληροφορίες για έναν στόχο (φίλος του Μίλγκραμ)
- Κάθε άτομο έπρεπε να στείλει τα πακέτα σε κάποιον που:
  - Αυτοί ήξεραν
  - Νόμιζαν ότι ήταν κοντά στον στόχο
- Ο στόχος έλαβε 64/296 πακέτα
- Μέσο μήκος διαδρομής για τα ληφθέντα πακέτα: 5.5

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Μικρά Παγκόσμια Δίκτυα

- Δίκτυα με τις ακόλουθες ιδιότητες
  - Ιδιαίτερα συγκεντρωμένη
  - Μικρά μέσα μήκη διαδρομής
- Εφαρμόζονται σε
  - Κοινωνιολογία
  - Επιστήμες της Γης
  - Χρήση υπολογιστή
  - Δίκτυα εγκεφάλου



Συντελεστής Ομαδοποίησης

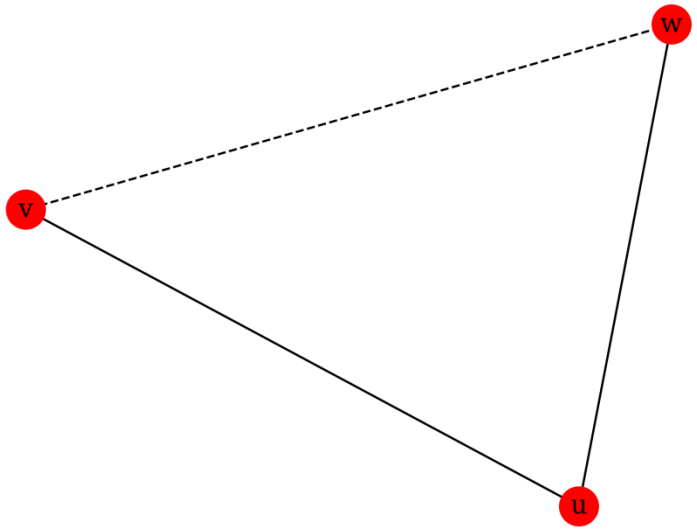
•το u είναι φίλος του v και ο w είναι φίλος σου. Είναι και ο va φίλος του w;

•Αν το v είναι φίλος του w η τριάδα vuw είναι κλειστή

•**Συντελεστής ομαδοποίησης**

C = (αριθμός μηκών κλειστής διαδρομής 2)/(αριθμός μηκών διαδρομής 2)

•Ποσοτικοποιεί τις κλειστές τριάδες



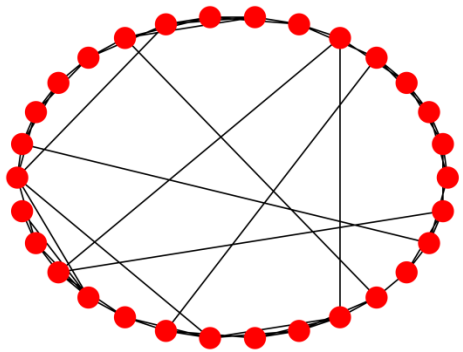
Συντελεστής Ομαδοποίησης

- Ας υποθέσουμε ότι έχουμε ένα τυχαίο δίκτυο με κάθε κόμβο να έχει τον ίδιο αριθμό συνδέσεων c.
- Ας υποθέσουμε ότι το u είναι φίλος του v και ο w είναι φίλος του w.
- Η πιθανότητα ο w να είναι φίλος του u είναι c/N.
- Μια τυπική τιμή για το c/N θα ήταν $3*10^{-4}$ (Film actors, Newman: Networks, μια εισαγωγή).
- Ωστόσο, πολύ μεγαλύτερες τιμές έχουν παρατηρηθεί (για παράδειγμα σε κοινωνικά ή τεχνολογικά δίκτυα).
- Ένας φίλος ενός φίλου είναι πιο πιθανό να είναι και φίλος μου.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Γραφήματα Watts Strogatz

- Το πιο γνωστό μοντέλο που δημιουργεί γραφήματα μικρού κόσμου
- Παρουσιάστηκε από τους Duncan Watts και Steve Strogatz (Watts and Strogatz 1998)
- **Μοντέλο** :
  - Ξεκινήστε με N κόμβους, καθένας από τους οποίους συνδέεται με k γείτονες
  - Συνδέστε ξαν ής με πιθανότητα p

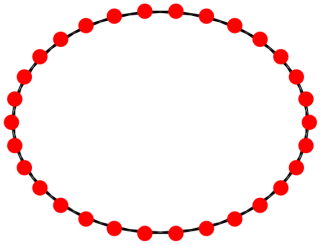


N=30, <k>=4, p=0,2

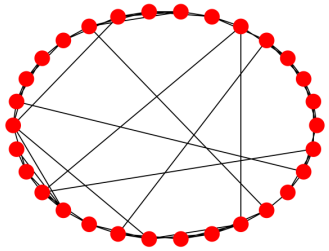
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

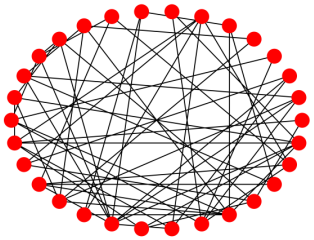
Παραλλαγή του p



p=0



p=0,2



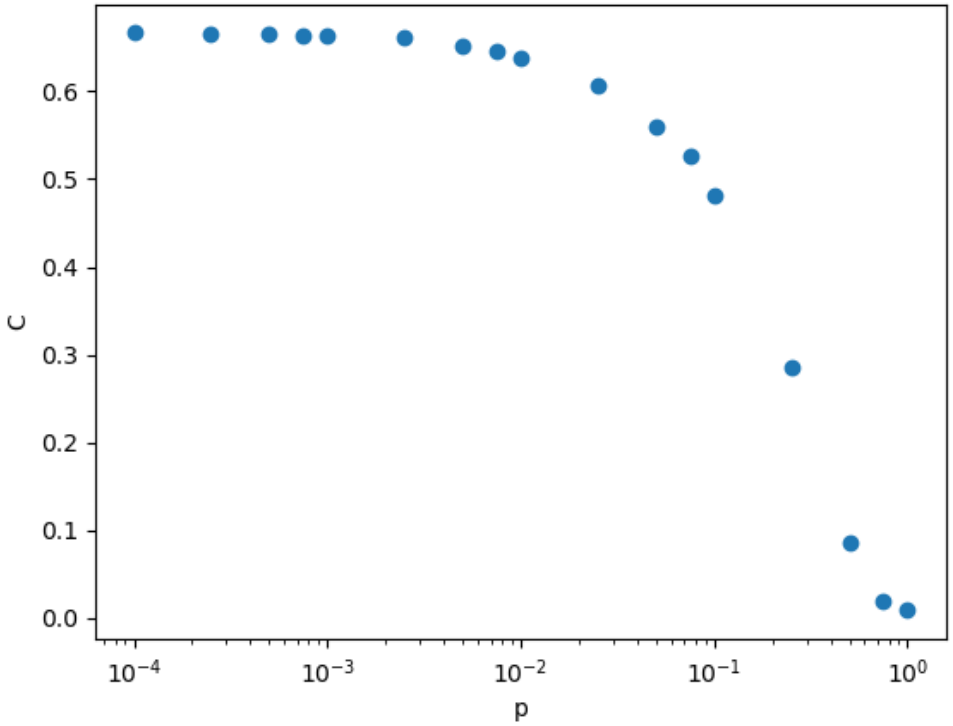
p=1

Co-funded by the Erasmus+ Programme of the European Union

Τυχαιότητα

- Το p ορίζει την τυχαιότητα.
- Στο p=0, βλέπουμε ένα απόλυτα συμμετρικό γράφημα. Κάθε κόμβος συνδέεται με k πιο κοντινούς γείτονες.
- Όταν το p αυξάνεται, η τυχαιότητα αυξάνεται
- Στο p = 1 έχουμε ένα τυχαίο γράφημα

Υψηλή ομαδοποίηση



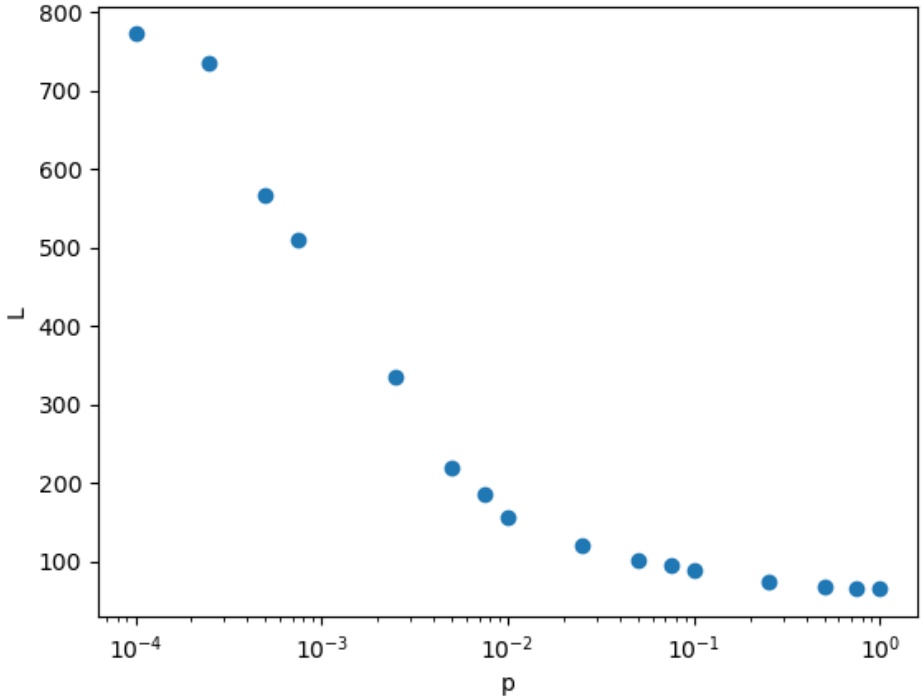
N = 1000, <k> = 10

•Παρατηρούμε μεγάλες τιμές του συντελεστή ομαδοποίησης για ένα ευρύ φάσμα τιμών πιθανότητας επανακαλωδίωσης

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Χαμηλό μέσο μήκος μικρότερου μονοπατιού



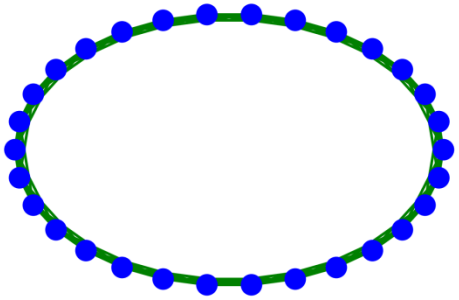
•Παρατηρούμε μικρές τιμές της μέσης συντομότερης διαδρομής για ένα ευρύ φάσμα τιμών της πιθανότητας επανακαλωδίωσης

N = 1000, <k> = 10

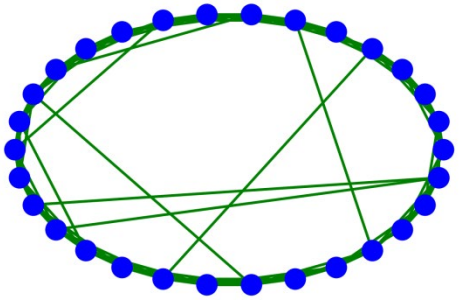
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

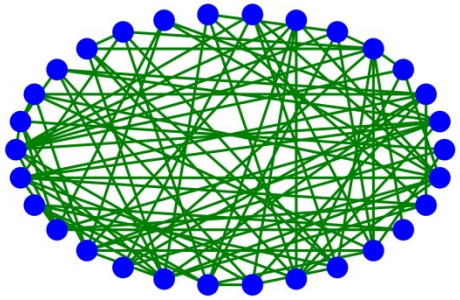
Αύξηση p



p = 0
Υψηλή ομαδοποίηση
Υψηλή διάμετρος

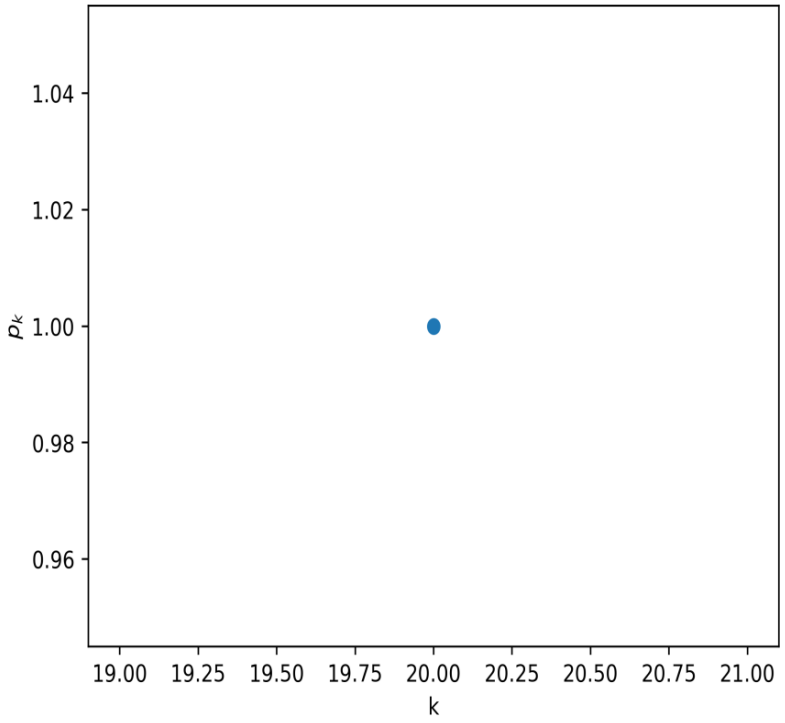


p = 0,1
Υψηλή ομαδοποίηση
Χαμηλή διάμετρος



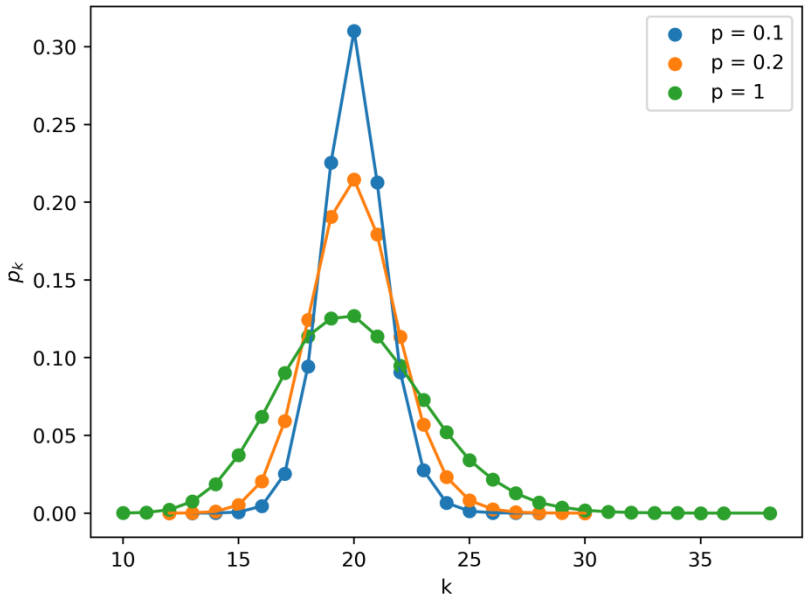
p = 1
Χαμηλή ομαδοποίηση
Χαμηλή διάμετρος

Κατανομή βαθμών, p=0



•Όταν p=0, όλοι οι κόμβοι έχουν ακριβώς k συνδέσεις. Η κατανομή πτυχίων είναι ένα σημείο.

Κατανομή βαθμών



- Όταν το p αυξάνεται, η κατανομή διευρύνεται. Υπάρχει μεγαλύτερη διακύμανση στις τιμές των συνδέσεων που έχει κάθε κόμβος. (Τυχαία)

- Όταν p=1, το δίκτυο είναι ένα τυχαίο δίκτυο

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ασκηση 1

- Δημιουργήστε δίκτυα Small World Watts Strogatz με N = 1000.
- Αρχικά, κάθε κόμβος του δικτύου θα έχει ακριβώς k συνδέσεις και στη συνέχεια οι ακμές θα καλωδιωθούν εκ νέου με πιθανότητα ίση με p.
- Προετοιμάστε τις ακόλουθες 2 γραφικές παραστάσεις: Σχεδιάστε τη συνάρτηση κατανομής πιθανότητας PDF, P(k) ως συνάρτηση του k, διατηρώντας σταθερή την πιθανότητα επανακαλωδίωσης στην πρώτη γραφική παράσταση (p = 0,4) και σταθερή <k> στη δεύτερη (<k> = 10).
- Αναμένονται 3 καμπύλες και στα δύο οικόπεδα. Στην πρώτη σχεδιάστε μια καμπύλη για κάθε <k> ίση με <k> = 6, 10, 16.
- Στη δεύτερη γραφική παράσταση, ετοιμάστε 3 καμπύλες, μία για κάθε πιθανότητα επανακαλωδίωσης ίση με p = 0,1, 0,2, 0,4. Τα αποτελέσματα θα πρέπει να είναι κατά μέσο όρο 1000 προσομοιώσεις.

Co-funded by the Erasmus+ Programme of the European Union

Άσκηση 2

- Δημιουργήστε δίκτυα Watts Strogatz με N=10000, <k> = 6, 10, 15.
- Για κάθε <k> υπολογίστε τον Συντελεστή Ομαδοποίησης και τη Συντομότερη Μέση Διαδρομή ως συνάρτηση της πιθανότητας επανακαλωδίωσης p,
- Το εύρος του p θα είναι 0 έως 1 με βήμα 0,1.
- Τα αποτελέσματα θα είναι μέσοι όροι 1000 πραγματοποιήσεων. Συγκρίνετε τις ποσότητες για τις διαφορετικές τιμές του <k>.

Co-funded by the Erasmus+ Programme of the European Union

Περίληψη

- Τα δίκτυα του πραγματικού κόσμου έχουν
  - Υψηλή ομαδοποίηση
  - Μικρές μέσες διαδρομές

Όπως τα δίκτυα του μικρού κόσμου. Ωστόσο, τα δίκτυα του Μικρού Κόσμου παράγουν:

- Μη ρεαλιστική κατανομή βαθμών
  - Τα δίκτυα του πραγματικού κόσμου έχουν πολλούς κόμβους με μικρό βαθμό αλλά έχουν επίσης κόμβους (κόμβους με πολύ υψηλό βαθμό σε σύγκριση με τον μέσο βαθμό)

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Εργαλεία βιοπληροφορικής

Θέμα 5. Αλγόριθμοι στη Βιοπληροφορική

Μάθημα 6. Δίκτυα ελεύθερα κλίμακας

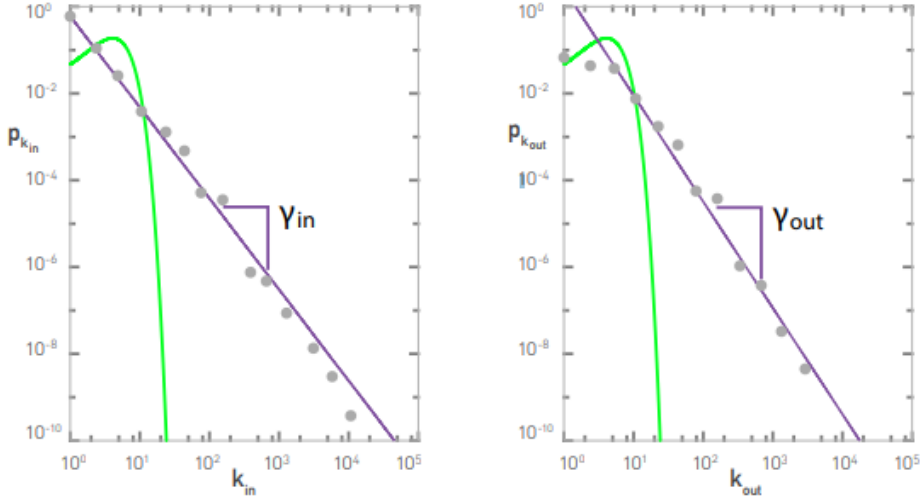
ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

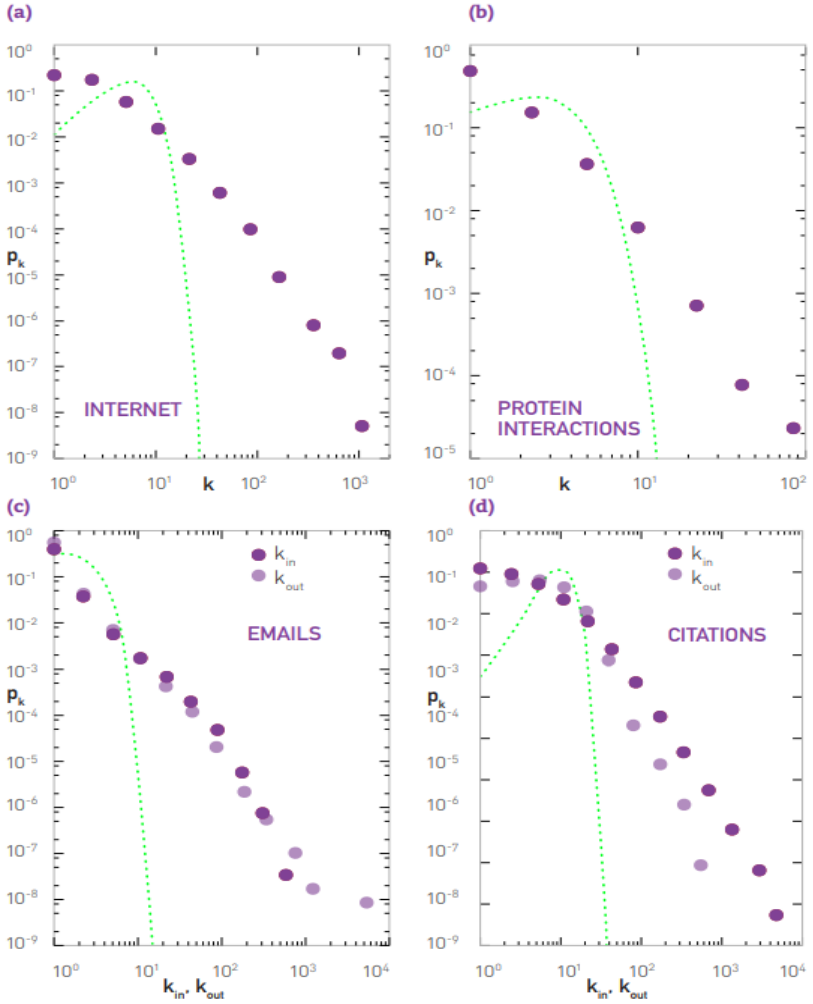
Παγκόσμιος Ιστός

- Ο παγκόσμιος ιστός είναι ένα κατευθυνόμενο δίκτυο
- Κάθε κόμβος έχει ένα βαθμό $k_{\text{μέσα}}$ (οι σύνδεσμοι προς τον κόμβο) και $k_{\text{έξω}}$ (οι σύνδεσμοι προς τα έξω του κόμβου)
- Η κατανομή βαθμών του διαφέρει από τις διανομές των δικτύων ER και Small World, έχει μορφή νόμου δύναμης και είναι αντιπροσωπευτικό πολλών δικτύων



Πηγή: Barabási , Network Science

Πραγματικά Δίκτυα



Πολλά δίκτυα Real έχουν κατανομή βαθμών νόμου δύναμης

Πηγή: Barabási , Network Science

Διακριτή επισημοποίηση

- Ένας κόμβος έχει ακριβώς k βαθμό όπου k = 0, 1, 2, 3… Η κατανομή βαθμών είναι ίση με

$$p_k = Ck^{-\gamma}$$

- Το p_k είναι η συνάρτηση πυκνότητας πιθανότητας:

$$\sum_{k=1}^{\infty} p_k = 1$$

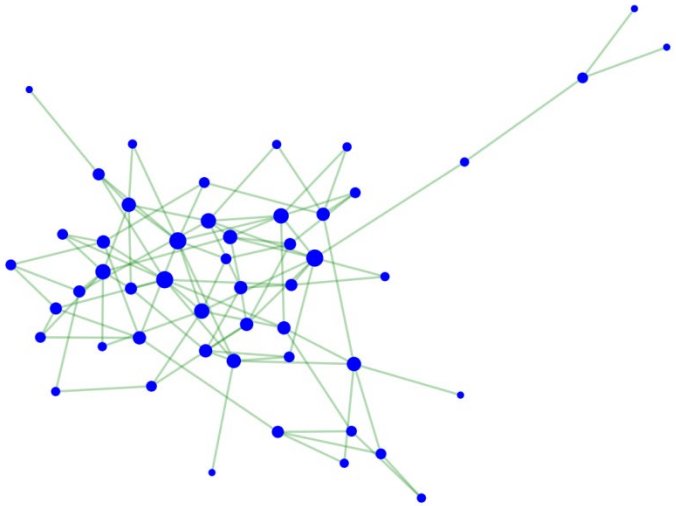
Ετσι

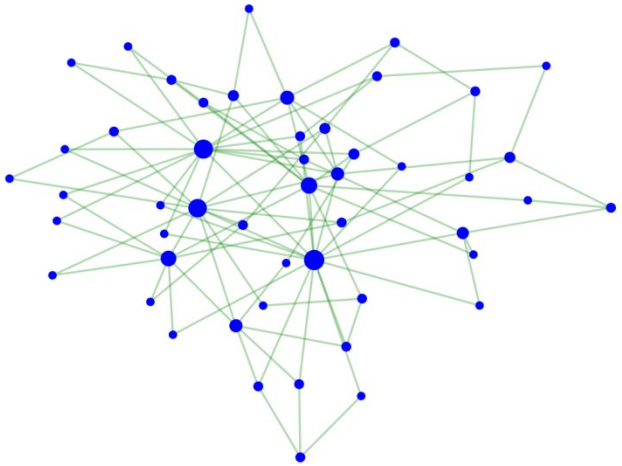
$$C = \frac{1}{\sum_{k=1}^{\infty} p_k}$$

Και

$$p_k = \frac{1}{\sum_{k=1}^{\infty} p_k} k^{-\gamma}$$

Τυχαίο vs Scale Free



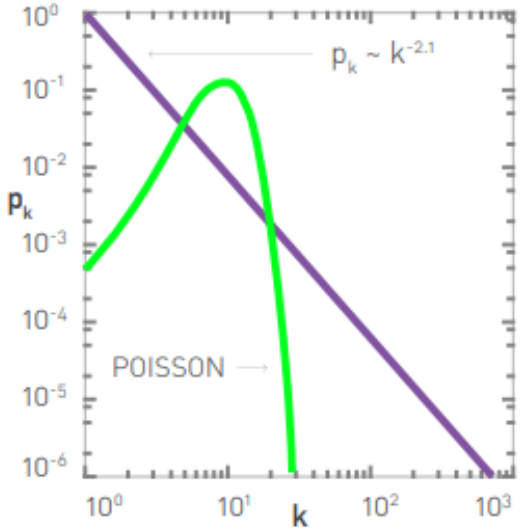


Τυχαίο Δίκτυο
•Δεν υπάρχει μεγάλη διαφορά στο k μεταξύ των κόμβων

Δίκτυο χωρίς κλίμακα
•Λίγοι κόμβοι με πολλές συνδέσεις (hubs)
•Πολλοί κόμβοι με μικρό βαθμό

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Νόμος Poisson εναντίον Power



Pk ως συνάρτηση του k για κατανομή Poisson και νόμου ισχύος. <k> = 11 και για τα δύο. (πηγή Barabási : Network Science)

- Και για τις δύο κατανομές το <k> είναι ίσο με 11
- Για την κατανομή του νόμου δύναμης (Δίκτυα χωρίς κλίμακα) οι περισσότεροι κόμβοι έχουν μικρό k και λίγοι κόμβοι έχουν υψηλό k
- Για την κατανομή Poisson (Τυχαία Δίκτυα) οι περισσότεροι κόμβοι έχουν παρόμοια k.

Barabási - Albert Model

Albert- László Μπαραμπάσι

Ρέκα Άλμπερτ

- Το Preferential Attachment, έχει εισαχθεί σε πολλούς επιστημονικούς τομείς για να εξηγήσει διάφορες διανομές από το 1923.

- Το 1999, ο επιστήμονας δικτύου László Barabási & Réka Albert εισήγαγαν ένα μοντέλο δικτύου

- Αυτό το μοντέλο βασίζεται σε 2 ιδιότητες, το Preferential Attachment και την Growth

- Δημιουργεί Scale Free Networks και είναι ένα από τα πιο γνωστά μοντέλα δικτύου.

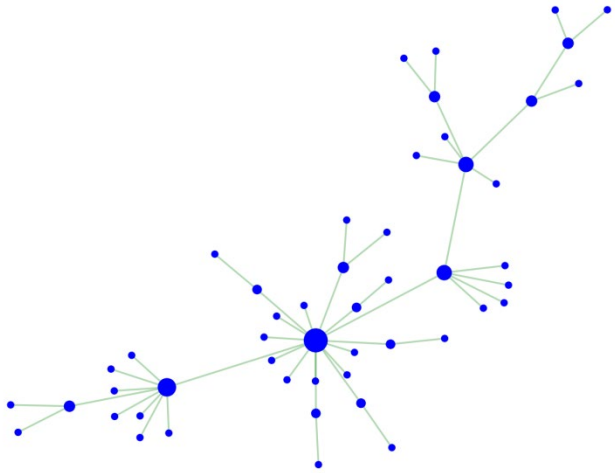
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Ανάπτυξη & Προνομιακή Προσκόλληση

δημιουργείται ένα δίκτυο m $_{0\ \text{κόμβων}}$

Ανάπτυξη

•Νέοι κόμβοι εισάγονται στο δίκτυο. Κάθε νέος κόμβος δημιουργεί m συνδέσεις.



Προτιμητική σύνδεση

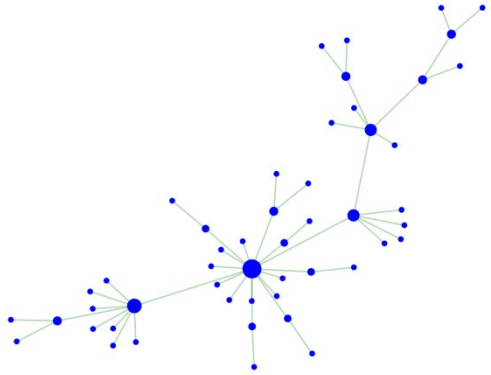
•Κάθε νέος κόμβος έχει μεγαλύτερη πιθανότητα να συνδεθεί με έναν κόμβο με υψηλό βαθμό. Η πιθανότητα ο κόμβος j να συνδεθεί με έναν κόμβο i είναι ίση με

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

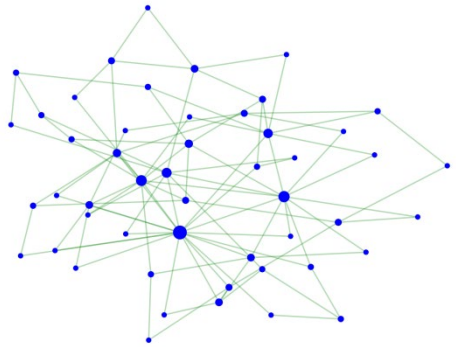
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

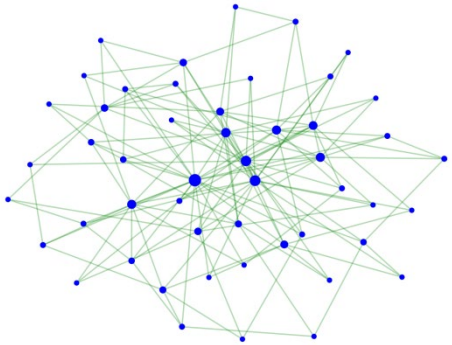
Δίκτυα BA με N=50



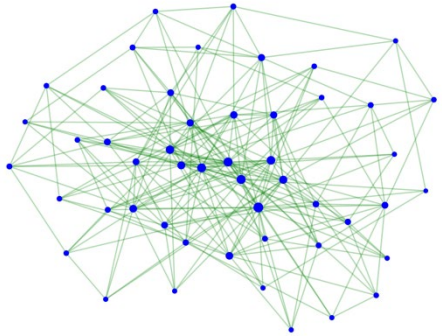
m=1



m=2



m=3

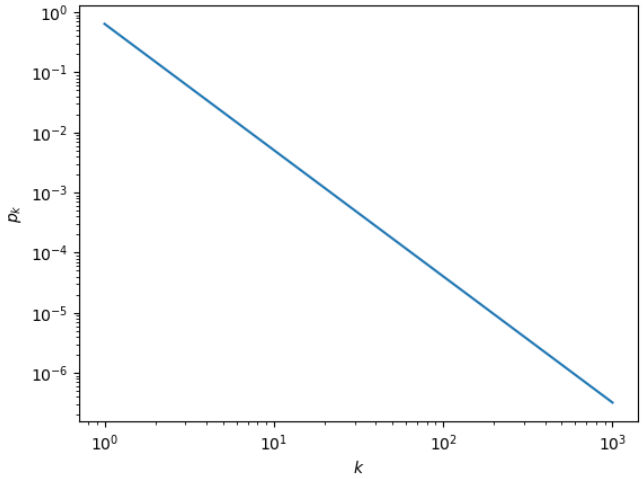


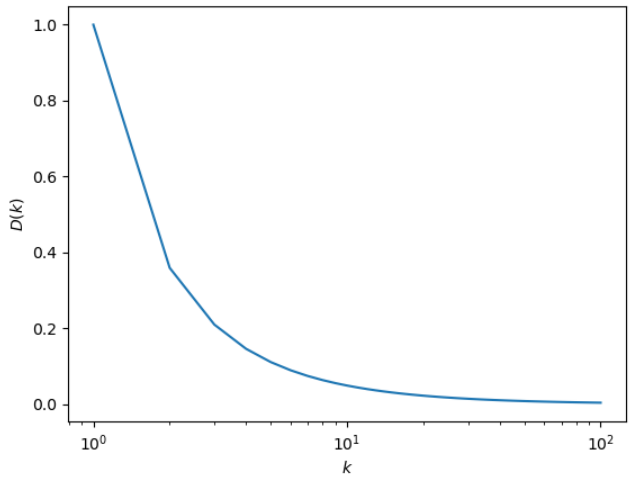
m=10

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Δημιουργήστε k ακολουθία





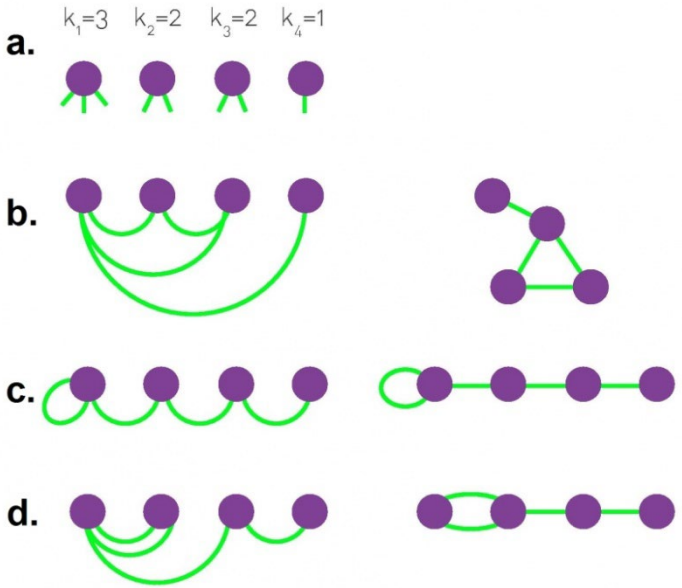
- Υπολογίστε το p_k ως συνάρτηση του k
- Υπολογίστε D(k), όπου

$$D(k) \sim \sum_{k' \geq k} p_{k'}$$

- Δημιουργήστε N τυχαίους αριθμούς μεταξύ 0 και 1
- Χρησιμοποιήστε το διάγραμμα του D(k) για να αντιστοιχίσετε κάθε τυχαίο αριθμό σε ένα k. $k = D^{-1}(r)$

όπου r είναι ο τυχαίος αριθμός

Μοντέλο διαμόρφωσης



Μοντέλο διαμόρφωσης. Πηγή: Barabási , Network Science

•Γνωρίζοντας την κατανομή βαθμών του δικτύου, εκχωρούμε το k του κάθε κόμβου

•Από την ακολουθία βαθμών δημιουργούνται τυχαία ακμές με πιθανότητα

$$p_{ij} = \frac{k_i k_j}{2L - 1}$$

•Επιτρέπονται οι πολλαπλοί σύνδεσμοι και οι αυτο-βρόχοι.

Free scale

- n- η moment της κατανομής βαθμών

$$< k^n >= \sum_{k_{\min}}^{\infty} k^n p_k = \int_{k_{\min}}^{\infty} k^n p(k) dk$$

Για δίκτυα χωρίς κλίμακα

$$< k^n >= C \frac{k_{\max}^{n-\gamma+1} - k_{\min}^{n-\gamma+1}}{n-\gamma+1}$$

- n=1, <k> είναι ο μέσος βαθμός
- n=2, με $<k^2>$ υπολογίζουμε τη διακύμανση $\sigma^2 = < k^2 > - <k>^2$
- Στο όριο
    - Αν $n-1+\gamma \leq 0$
        - Όλες οι moments είναι πεπερασμένες

Free scale

- Αν n-1+ γ > 0
  - moments μεγαλύτερες από γ – 1 αποκλίνουν

- Ο βαθμός δικτύων χωρίς κλίμακα είναι περίπου 2-3
  - Η πρώτη ροπή (moment) είναι πεπερασμένη
  - Η δεύτερη ροπή αποκλίνει

συμπέρασμα

- Τα δίκτυα χωρίς κλίμακα δεν διαθέτουν κλίμακα.
- Αν επιλέξουμε τυχαία έναν κόμβο, δεν ξέρουμε τι να περιμένουμε. Θα μπορούσε να έχει εξαιρετικά μεγάλο βαθμό ή εξαιρετικά χαμηλό βαθμό.

Συμπεριφορά συστήματος ως συνάρτηση του γ

•Ένα ενδιαφέρον ερώτημα είναι ποια θα ήταν η συμπεριφορά του συστήματος σε συνάρτηση με το γ

•γ ≤ 2 (Ανώμαλο καθεστώς)
- ○Η πρώτη ροπή διαφέρει
- ○Η 2η ροπή αποκλίνει
- ○Το k_{max} αυξάνεται ταχύτερα από το δίκτυο (Περίεργες προβλέψεις)

•2 < γ < 3
- ○Η πρώτη ροπή είναι πεπερασμένη
- ○Η δεύτερη ροπή αποκλίνει
- ○ $k_{\max} \sim N^{\frac{1}{\gamma-1}}$

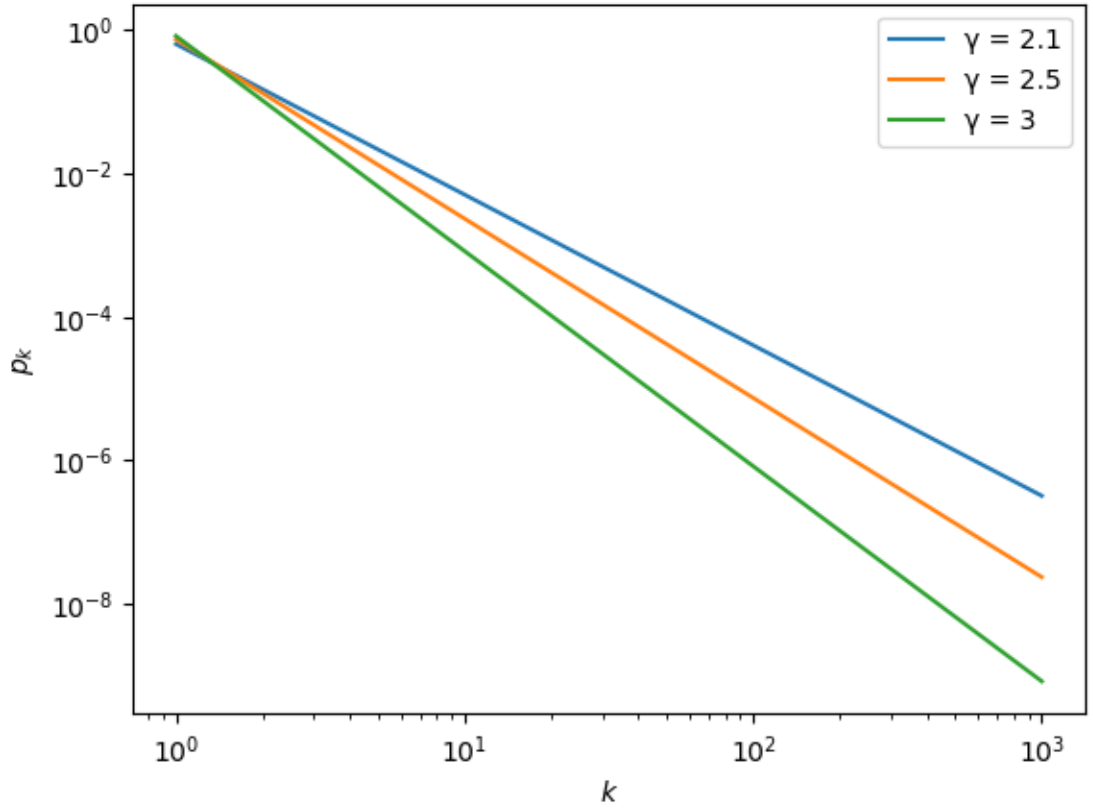
- γ>3 (τυχαίο καθεστώς)
  - 1η ροπή πεπερασμένη
  - 2η ροπή πεπερασμένη

Ίδια συμπεριφορά με ένα τυχαίο δίκτυο

Ανώμαλο καθεστώς	**2**	Ελεύθερη κλίμακα καθεστώς	**3**	τυχαίο καθεστώς	**γ**
•Το <k> αποκλίνει •<k^2> αποκλίνει		•<k> πεπερασμένο •<k^2> αποκλίνει		•<k> πεπερασμένο •<k^2> πεπερασμένο	

Co-funded by the Erasmus+ Programme of the European Union

Παραλλαγή του γ



Co-funded by the Erasmus+ Programme of the European Union

Ασκηση

- Δημιουργήστε ένα δίκτυο Barabási-Albert με N = 1000 κόμβους και m=2,4,10.
- Στο ίδιο σχήμα σχεδιάστε τις τρεις κατανομές βαθμών, μία για κάθε τιμή του m.
- Βρείτε τη συνδεσιμότητα του μεγαλύτερου κόμβου του δικτύου. Ποια είναι τα συμπεράσματά σας;
- Υπολογίστε την 1η και τη 2η ροπή χρησιμοποιώντας τις σχέσεις που αναφέρονται στο powerpoint
- Τα αποτελέσματα πρέπει να είναι κατά μέσο όρο 1000 προσομοιώσεων.

Co-funded by the Erasmus+ Programme of the European Union

Περίληψη

- Σε αυτό το μάθημα έχουμε εισαγάγει δίκτυα χωρίς κλίμακα
- Εξετάσαμε:
- Barabasi Albert.
- Ένα μοντέλο που βασίζεται στις ακόλουθες ιδέες
  - Ανάπτυξη
  - Προνομιακή κατάσχεση
- Μοντέλο διαμόρφωσης
  - Ένα τυχαίο δίκτυο, με κατανομή βαθμών (θα μπορούσε να είναι χωρίς κλίμακα)
- Τα δίκτυα χωρίς κλίμακα δεν διαθέτουν κλίμακα

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

Θέμα 5. Αλγόριθμοι στη Βιοπληροφορική

Μάθημα 7. Ομαδοποίηση

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

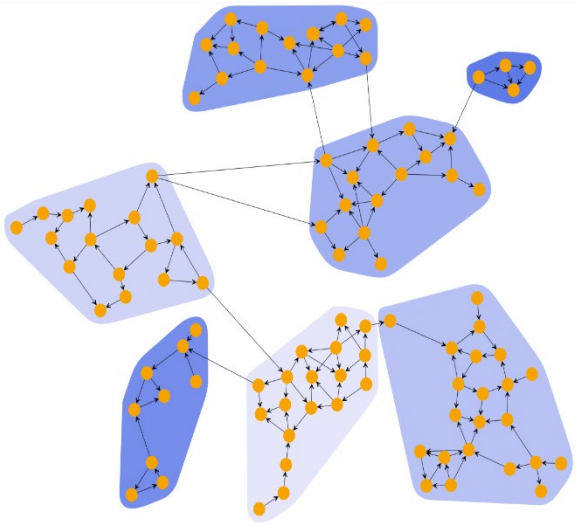
Co-funded by the Erasmus+ Programme of the European Union

Ομαδοποίηση

- Η ομαδοποίηση είναι η διαδικασία ομαδοποίησης παρόμοιων αντικειμένων με βάση ορισμένα κριτήρια.

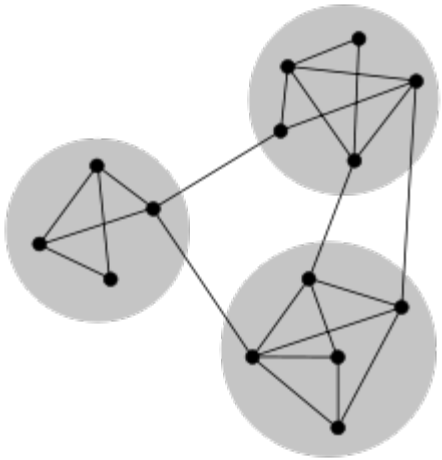
- Η ομαδοποίηση είναι μια θεμελιώδης τεχνική στην ανάλυση δικτύων και έχει εφαρμογές σε διάφορους τομείς όπως η βιολογία, οι κοινωνικές επιστήμες και η επιστήμη των υπολογιστών.

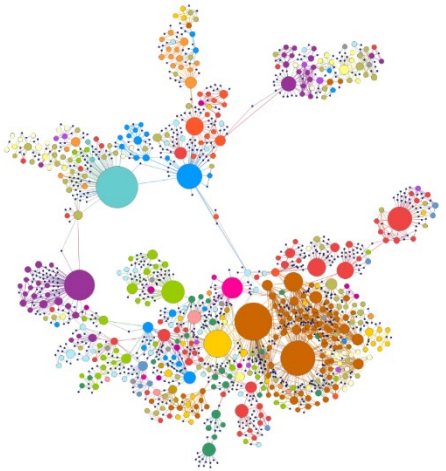
- Σε αυτή την παρουσίαση, θα συζητήσουμε διαφορετικούς τύπους αλγορίθμων ομαδοποίησης και τις εφαρμογές τους στην ανάλυση δικτύου.



Στόχος της ομαδοποίησης

- Η ομαδοποίηση μπορεί να χρησιμοποιηθεί για την κατανόηση της δομής και της λειτουργίας ενός δικτύου και μπορεί να αποκαλύψει σημαντικές πληροφορίες σχετικά με τις σχέσεις μεταξύ των κόμβων
- Για τον εντοπισμό κοινοτήτων ή λειτουργικών μονάδων εντός του δικτύου, όπου οι κόμβοι μέσα σε μια κοινότητα είναι πιο πυκνά συνδεδεμένοι μεταξύ τους παρά με κόμβους σε άλλες κοινότητες

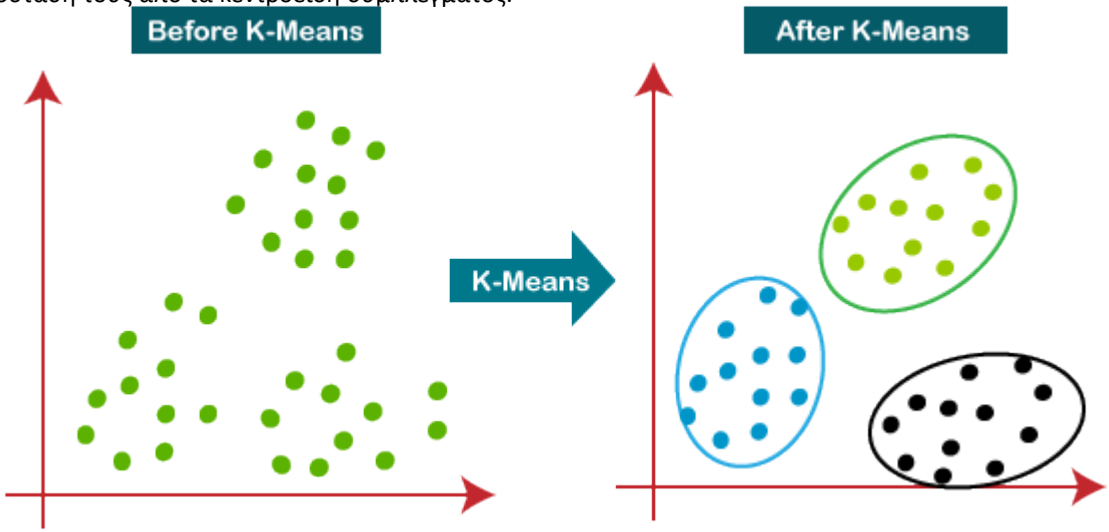




Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Τύποι Ομαδοποίησης

- Α. Συγκεντρώστε κόμβους με παρόμοιες ιδιότητες (όπως βαθμός)
  - Ιεραρχική Ομαδοποίηση:
    - Μια μέθοδος ομαδοποίησης που δημιουργεί μια δομή που μοιάζει με δέντρο συγχωνεύοντας επαναληπτικά μικρότερα συμπλέγματα σε μεγαλύτερα.
  - Ομαδοποίηση K-Means:
    - Μια μέθοδος ομαδοποίησης βάσει διαμερισμάτων που χωρίζει τους κόμβους σε έναν προκαθορισμένο αριθμό συστάδων με βάση την απόστασή τους από τα κεντροειδή συμπλέγματος.



Co-funded by the Erasmus+ Programme of the European Union

Κοινότητες

- Β. Για τον εντοπισμό κοινοτήτων ή λειτουργικών μονάδων εντός του δικτύου, όπου οι κόμβοι μέσα σε μια κοινότητα είναι πιο πυκνά συνδεδεμένοι μεταξύ τους παρά με κόμβους σε άλλες κοινότητες
  - Ομαδοποίηση βάσει αρθρώσεων:
    - Ένας αλγόριθμος ανίχνευσης κοινότητας που στοχεύει στη μεγιστοποίηση της βαθμολογίας σπονδυλωτότητας ενός δικτύου διαχωρίζοντας τους κόμβους σε μη επικαλυπτόμενες κοινότητες.
- Υπάρχουν διάφοροι τύποι αλγορίθμων ομαδοποίησης, ο καθένας με τα δικά του δυνατά και αδύνατα σημεία.

Ιεραρχική Ομαδοποίηση

- Η ιεραρχική ομαδοποίηση είναι μια μέθοδος ομαδοποίησης που δημιουργεί μια ιεραρχία συστάδων που μοιάζει με δέντρο.
- Δύο κύριοι τύποι ιεραρχικής ομαδοποίησης:
- Συσσωματωτική ομαδοποίηση
  - ξεκινά με κάθε σημείο δεδομένων ως το δικό του σύμπλεγμα και τα συγχωνεύει επαναληπτικά.
- Διχαστική ομαδοποίηση
  - ξεκινά με όλα τα σημεία δεδομένων ως ένα σύμπλεγμα και τα χωρίζει αναδρομικά.

Co-funded by the Erasmus+ Programme of the European Union

Παράδειγμα

- Σκεφτείτε ένα δίκτυο κοινωνικών αλληλεπιδράσεων μεταξύ των ανθρώπων.
- Ο αλγόριθμος ιεραρχικής ομαδοποίησης θα ομαδοποιούσε τους ανθρώπους με βάση τον βαθμό αλληλεπίδρασής τους, με τις πιο στενά συνδεδεμένες ομάδες να σχηματίζουν συστάδες σε υψηλότερα επίπεδα της ιεραρχίας.
- Η προκύπτουσα ιεραρχία θα μπορούσε να απεικονιστεί ως δενδρογράφημα, με τα φύλλα να αντιπροσωπεύουν μεμονωμένους κόμβους και τους κλάδους να αντιπροσωπεύουν τη συγχώνευση συστάδων.

Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμος

- Βήμα 1: Ξεκινήστε αντιστοιχίζοντας κάθε κόμβο στο δικό του σύμπλεγμα.
- Βήμα 2: Υπολογίστε την απόσταση μεταξύ κάθε ζεύγους συστάδων χρησιμοποιώντας μια μετρική απόστασης όπως η Ευκλείδεια απόσταση ή η απόσταση συσχέτισης.
- Βήμα 3: Συγχωνεύστε τα δύο πλησιέστερα συμπλέγματα σε ένα ενιαίο σύμπλεγμα.
- Βήμα 4: Υπολογίστε ξανά τις αποστάσεις μεταξύ αυτού του νέου συμπλέγματος και όλων των άλλων συστάδων.
- Βήμα 5: Επαναλάβετε τα βήματα 3 και 4 έως ότου όλοι οι κόμβοι είναι σε ένα ενιαίο σύμπλεγμα ή μέχρι να ικανοποιηθεί ένα κριτήριο διακοπής.
- Βήμα 6: Δημιουργήστε ένα δενδρογράφημα για να απεικονίσετε την ιεραρχία των συστάδων.

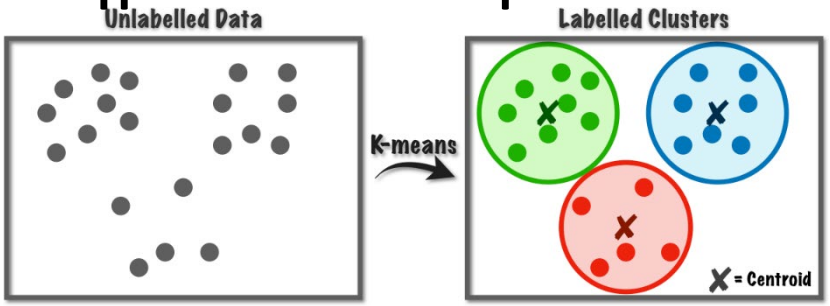
K-means ομαδοποίηση

- Μια τεχνική μηχανικής μάθησης χωρίς επίβλεψη που χρησιμοποιείται για ανάλυση ομαδοποίησης
- Χωρίζει τα δεδομένα σε ομάδες K βάσει της ομοιότητάς τους
- Κάθε σύμπλεγμα έχει το δικό του κέντρο, το οποίο αντιπροσωπεύει το κέντρο του συμπλέγματος

Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμος

- Επιλέξτε τον αριθμό των συμπλεγμάτων (k) για να ομαδοποιήσετε τα δεδομένα.
- Επιλέξτε τυχαία k σημεία δεδομένων που θα χρησιμεύσουν ως αρχικά κεντροειδή για κάθε σύμπλεγμα.
- Αντιστοιχίστε κάθε σημείο δεδομένων στο πλησιέστερο κέντρο.

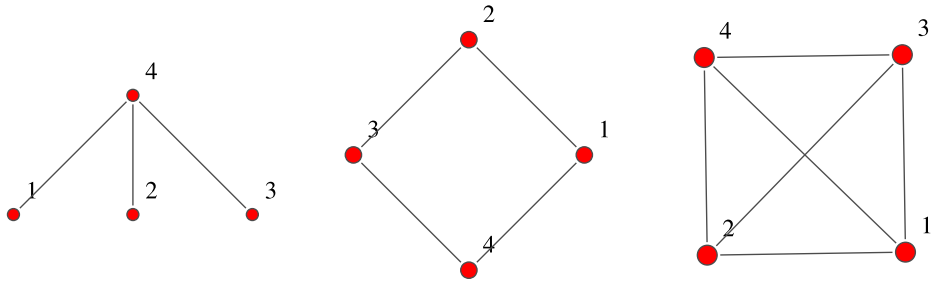


Αλγόριθμος

- Υπολογίστε ξανά τα κεντροειδή ως τον μέσο όρο όλων των σημείων δεδομένων που του έχουν εκχωρηθεί.
- Επαναλάβετε τα βήματα 3 και 4 έως ότου τα κεντροειδή να μην μετακινούνται πλέον ή να επιτευχθεί ο μέγιστος αριθμός επαναλήψεων.
- Τα συμπλέγματα που προκύπτουν βασίζονται στις τελικές κεντροειδείς θέσεις.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Παράρτημα: Πίνακας γειτνίασης



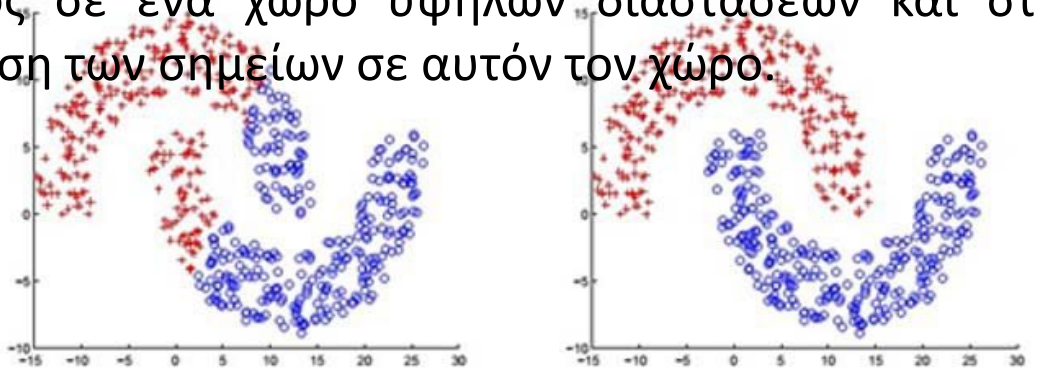
$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

- Matrix, που αντιπροσωπεύει τα άκρα του δικτύου
- Ένα $_{ij}$ = 1 εάν υπάρχει σύνδεση μεταξύ των κόμβων i , j
- A $_{ij}$ = 0 εάν οι κόμβοι i , j δεν είναι συνδεδεμένοι

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Φασματική ομαδοποίηση

- Η φασματική ομαδοποίηση είναι μια τεχνική κατάτμησης γραφήματος που περιλαμβάνει τη χρήση των ιδιοτιμών και των ιδιοδιανυσμάτων του πίνακα Λαπλασιανής ενός γραφήματος για τη διαίρεση του γραφήματος σε συστάδες.
- Ο πίνακας Λαπλασίανης ενός γραφήματος ορίζεται ως η διαφορά μεταξύ του πίνακα βαθμών και του πίνακα γειτνίασης.
- Η φασματική ομαδοποίηση περιλαμβάνει την ενσωμάτωση του γραφήματος σε ένα χώρο υψηλών διαστάσεων και στη συνέχεια τη ομαδοποίηση των σημείων σε αυτόν τον χώρο.



(a) K-means (b) Spectral Clustering

Co-funded by the Erasmus+ Programme of the European Union

Βήματα

- Κατασκευάστε τον πίνακα γειτνίασης ή τον πίνακα Λαπλασιανής του γραφήματος/δικτύου.
- Υπολογίστε τις ιδιοτιμές και τα ιδιοδιανύσματα του πίνακα.
- Επιλέξτε ένα υποσύνολο των ιδιοδιανυσμάτων που αντιστοιχούν στις k μεγαλύτερες ιδιοτιμές.
- Σχηματίστε έναν πίνακα με τα επιλεγμένα ιδιοδιανύσματα ως στήλες.
- Εφαρμόστε τον αλγόριθμο ομαδοποίησης (π.χ. k-means) στις σειρές του νέου πίνακα.
- Αντιστοιχίστε κάθε κόμβο στο σύμπλεγμα στο οποίο ανήκει με βάση το αποτέλεσμα της ομαδοποίησης.

Co-funded by the Erasmus+ Programme of the European Union

Ασκηση

1. Λάβετε ένα σύνολο δεδομένων που περιέχει βιολογικά δεδομένα που είναι κατάλληλα για ανάλυση ομαδοποίησης, όπως δεδομένα γονιδιακής έκφρασης, δεδομένα αλληλεπίδρασης πρωτεϊνών ή δεδομένα γονιδιωματικής αλληλουχίας. (https://biodbnet-abcc.ncifcrf.gov/)
2. Εφαρμόστε τον αλγόριθμο ιεραρχικής ομαδοποίησης στο σύνολο δεδομένων και οπτικοποιήστε το δενδρόγραμμα που προκύπτει.
3. Εφαρμόστε τον αλγόριθμο ομαδοποίησης k-means στο ίδιο σύνολο δεδομένων
4. Αξιολογήστε και συγκρίνετε τα αποτελέσματα της ομαδοποίησης που προέκυψαν από την ιεραρχική ομαδοποίηση και την ομαδοποίηση k-means. Αναλύστε τις ομοιότητες και τις διαφορές μεταξύ των δύο προσεγγίσεων και συζητήστε τα δυνατά και τα όριά τους.

Co-funded by the Erasmus+ Programme of the European Union

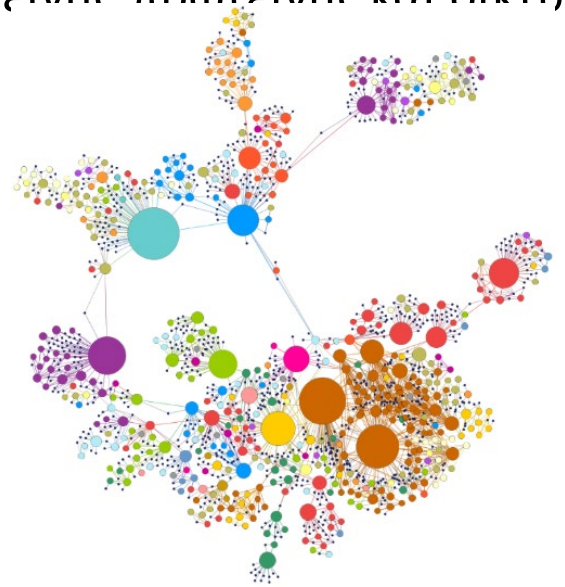
Περίληψη

- Ορισμός ομαδοποίησης
- Διαφορετικοί τύποι ομαδοποίησης και ο αλγόριθμός τους
  - k-means
    - Μια μέθοδος ομαδοποίησης βάσει διαμερισμάτων που χωρίζει τους κόμβους σε έναν προκαθορισμένο αριθμό συστάδων με βάση την απόστασή τους από τα κεντροειδή συμπλέγματος.
  - Ιεραρχικός
    - Μια μέθοδος ομαδοποίησης που δημιουργεί μια δομή που μοιάζει με δέντρο συγχωνεύοντας επαναληπτικά μικρότερα συμπλέγματα σε μεγαλύτερα.
  - Φασματική ομαδοποίηση
    - Μια μέθοδος ομαδοποίησης που βασίζεται σε γράφημα που χρησιμοποιεί τις ιδιοτιμές και τα ιδιοδιανύσματα του πίνακα Laplacian του δικτύου σε κόμβους κατάτμησης.

Co-funded by the Erasmus+ Programme of the European Union

Κοινότητες

- Μια κοινότητα ορίζεται ως μια ομάδα κόμβων που είναι πυκνά συνδεδεμένοι εντός της ομάδας και αραιά συνδεδεμένοι με κόμβους εκτός της ομάδας.
- Αυτή η προσέγγιση έχει αποδειχθεί αποτελεσματική στον εντοπισμό λειτουργικών μονάδων σε βιολογικά δίκτυα, όπως δίκτυα αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης και δίκτυα συνέκφρασης γονιδίων.



Co-funded by the Erasmus+ Programme of the European Union

Αρθρωτότητα

- Το modularity είναι ένα μέτρο που ποσοτικοποιεί τον βαθμό στον οποίο ένα δίκτυο μπορεί να χωριστεί σε μη επικαλυπτόμενες κοινότητες ή ενότητες. Ορίζεται μαθηματικά ως:

$$Q = \frac{1}{2m}\sum_{i,j}(A_{ij} - \frac{k_i k_j}{2m}) \cdot \delta_{c_i c_j}$$

- m: συνολικός αριθμός άκρων
- Βάρος της ακμής με τους κόμβους i και j (0 ή 1 σε μη κατευθυνόμενα δίκτυα)
- k_i είναι ο βαθμός κεντρικότητας του κόμβου I
- δ είναι η συνάρτηση δέλτα Kronecker (1 αν οι κοινότητες c_i και c_j είναι ίσες, 0 διαφορετικά)

Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμος Louvain

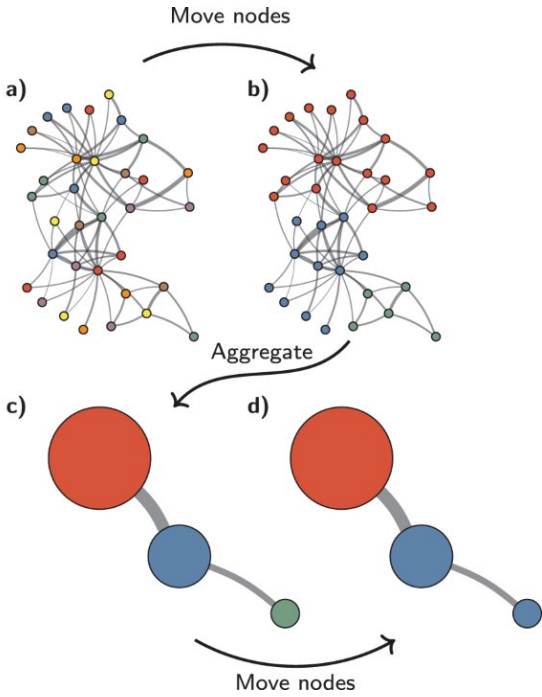
- Παρακάτω παρουσιάζουμε τα βήματα του ενός
- Αρχικοποίηση:
  - Αντιστοιχίστε κάθε κόμβο στη δική του κοινότητα
  - Υπολογίστε τη βαθμολογία αρθρωτής Q του αρχικού διαμερίσματος

Level 1

- Επανάληψη:
  - Για κάθε κόμβο i , υπολογίστε το κέρδος σε αρθρωτότητα μετακινώντας τον σε κάθε κοινότητα των γειτόνων του
  - Επιλέξτε την κοινότητα που μεγιστοποιεί το κέρδος αρθρωτότητας για τον κόμβο i και μετακινήστε την εκεί

Level 2

  - Επαναλάβετε αυτή τη διαδικασία για όλους τους κόμβους έως ότου δεν μπορούν να επιτευχθούν άλλα κέρδη σπονδυλωτών



- Συσσωμάτωση:
  - Συμβιβάστε κάθε κοινότητα σε έναν μόνο κόμβο, διατηρώντας το συνολικό βάρος των ακμών μεταξύ των κόμβων σε διαφορετικές κοινότητες
  - Υπολογίστε ξανά τη βαθμολογία αρθρωτής Q για τη νέα κατάτμηση κοινοτήτων

- Επαναλάβετε τα βήματα 2-3 έως ότου δεν μπορούν να επιτευχθούν άλλα κέρδη σπονδυλωτής ή να ικανοποιηθεί ένα κριτήριο διακοπής

Σύγκριση μεταξύ αλγορίθμων

- Η ιεραρχική ομαδοποίηση είναι ευέλικτη και μπορεί να χειριστεί διαφορετικούς τύπους δεδομένων, αλλά μπορεί να είναι υπολογιστικά εντατική και ευαίσθητη στο θόρυβο.

- Η ομαδοποίηση K-means είναι γρήγορη και αποτελεσματική, αλλά απαιτεί προηγούμενη γνώση του αριθμού των συστάδων και μπορεί να είναι ευαίσθητη στις αρχικές συνθήκες.

- Η φασματική ομαδοποίηση μπορεί να χειριστεί μη γραμμικές σχέσεις και δεν περιορίζεται από το σχήμα των συστάδων, αλλά απαιτεί προσεκτική επιλογή του αριθμού των συστάδων και μπορεί να είναι ευαίσθητη στην επιλογή του μέτρου ομοιότητας.

- Η ομαδοποίηση βάσει σπονδυλωτών είναι κατάλληλη για τον εντοπισμό κοινοτήτων σε δίκτυα , αλλά μπορεί να είναι υπολογιστικά εντατική για μεγάλα δίκτυα.

- Άλλοι αλγόριθμοι ομαδοποίησης, όπως η ομαδοποίηση με βάση την πυκνότητα και η ασαφής ομαδοποίηση, έχουν επίσης τα δικά τους πλεονεκτήματα και αδυναμίες.

- Η επιλογή του κατάλληλου αλγόριθμου ομαδοποίησης εξαρτάται από το ερευνητικό ερώτημα, τον τύπο και το μέγεθος των δεδομένων και τους διαθέσιμους υπολογιστικούς πόρους.

Ομαδοποίηση στη Βιοπληροφορική

- Οι αλγόριθμοι ομαδοποίησης χρησιμοποιούνται ευρέως στη βιοπληροφορική για την ανάλυση μεγάλων βιολογικών δικτύων, όπως:
  - Δίκτυα αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης
  - Δίκτυα συνέκφρασης γονιδίων
  - Μεταβολικά δίκτυα
  - Ρυθμιστικά δίκτυα
- Η ομαδοποίηση μπορεί να βοηθήσει στον εντοπισμό λειτουργικά σχετιζόμενων ομάδων γονιδίων ή πρωτεϊνών, οι οποίες μπορούν να παρέχουν πληροφορίες για βιολογικές διεργασίες και μηχανισμούς ασθενειών.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πρόγνωση

- Η ανάλυση ομαδοποίησης μας επιτρέπει να διαστρωματοποιήσουμε τους ασθενείς με βάση τα μοριακά τους προφίλ ή τα χαρακτηριστικά της νόσου.
- Εξετάζοντας παρόμοιους ασθενείς μέσα σε ομάδες, μπορούμε να προβλέψουμε
  - εξέλιξη της νόσου
  - ανταπόκριση στη θεραπεία
  - αποτελέσματα ασθενών.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Ερευνα για τον καρκίνο

- Στο πλαίσιο της έρευνας για τον καρκίνο, μπορούμε να εφαρμόσουμε αλγόριθμους ομαδοποίησης σε δίκτυα αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης για να εντοπίσουμε πιθανούς θεραπευτικούς στόχους.
- Εστιάζοντας σε συγκεκριμένες ομάδες ή ενότητες εντός του δικτύου, μπορούμε να αποκαλύψουμε πρωτεΐνες που παίζουν κρίσιμους ρόλους στην εξέλιξη του καρκίνου και μπορούν να χρησιμεύσουν ως αποτελεσματικοί στόχοι για την ανάπτυξη φαρμάκων.

Louvain στα μεταβολικά δίκτυα

- Εφαρμόζοντας τον αλγόριθμο Louvain σε ένα μεταβολικό δίκτυο, μπορούμε να αναγνωρίσουμε κοινότητες ή ενότητες μεταβολιτών που εμπλέκονται σε μια συγκεκριμένη μεταβολική οδό.
- Αυτό μας δίνει τη δυνατότητα να εντοπίσουμε βασικούς μεταβολίτες που παίζουν κρίσιμους ρόλους στην οδό και μπορεί να χρησιμεύσουν ως πιθανοί στόχοι για την ανακάλυψη φαρμάκων και τις θεραπευτικές παρεμβάσεις.

Co-funded by the Erasmus+ Programme of the European Union

Βασικοί μεταβολίτες

- Αναλύοντας την τοπολογία του μεταβολικού δικτύου, μπορούμε να αναγνωρίσουμε μεταβολίτες που έχουν υψηλούς βαθμούς (αριθμός συνδέσεων) ή υψηλή μεταξύ τους κεντρικότητα (δρώντας ως κρίσιμα ενδιάμεσα μεταξύ πολλαπλών μονοπατιών).
- Αυτοί οι βασικοί μεταβολίτες είναι σημαντικοί επειδή εμπλέκονται σε κρίσιμες βιοχημικές αντιδράσεις και έχουν τη δυνατότητα να επηρεάσουν τη συνολική λειτουργία της μεταβολικής οδού.
- Η στόχευση αυτών των μεταβολιτών μπορεί να έχει σημαντικό αντίκτυπο στη δραστηριότητα της οδού, καθιστώντας τους ελκυστικούς υποψήφιους για ανακάλυψη φαρμάκων.

Co-funded by the Erasmus+ Programme of the European Union

Ασκηση

1. Για το ίδιο σύνολο δεδομένων της άσκησης στο μάθημα 9
2. Εφαρμόστε τον αλγόριθμο Louvain
3. Αξιολογήστε και συγκρίνετε τα αποτελέσματα της ομαδοποίησης που προέκυψαν από τους αλγόριθμους ομαδοποίησης. Αναλύστε τις ομοιότητες και τις διαφορές μεταξύ των αλγορίθμων
4. Δημιουργήστε ένα ιστόγραμμα του αριθμού των κοινοτήτων που δημιουργήθηκαν

Co-funded by the Erasmus+ Programme of the European Union

Περίληψη

- Σε αυτό το μάθημα επικεντρωθήκαμε σε αλγόριθμους ομαδοποίησης που χωρίζουν τις κοινότητες του γραφήματος
- Αλγόριθμος Louvain
  - Διαμερισμός του δικτύου σε κοινότητες, μεγιστοποίηση της αρθρωτής δομής
- Συγκρίναμε τους αλγόριθμους ομαδοποίησης
- Παραδείγματα εφαρμογών ομαδοποίησης στη βιοπληροφορική
  - Πρόγνωση
  - Ερευνα για τον καρκίνο
  - Εύρεση των βασικών στοιχείων των δικτύων

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 3. Βιοπληροφορικά εργαλεία

Θέμα 5. Αλγόριθμοι στη Βιοπληροφορική

Μάθημα 8. Ευθυγράμμιση Δικτύου

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Αλγόριθμοι στη Βιοπληροφορική

Μάθημα 11

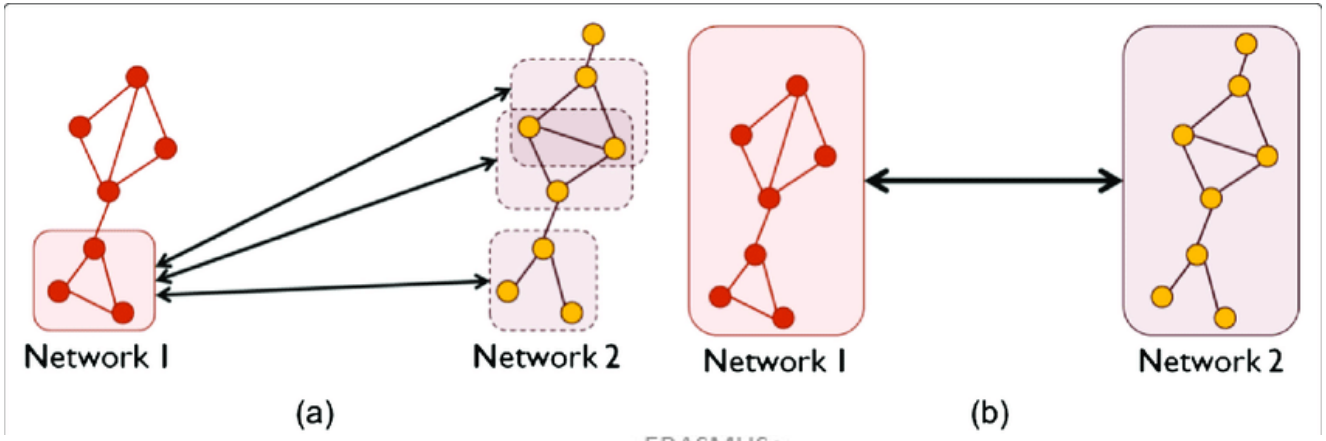
Ευθυγράμμιση δικτύου

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Ευθυγράμμιση

- Η ευθυγράμμιση δικτύου είναι η διαδικασία αναγνώρισης παρόμοιων δομών σε δύο ή περισσότερα δίκτυα.
- Περιλαμβάνει την εύρεση μιας χαρτογράφησης μεταξύ των κόμβων στα δίκτυα που διατηρεί τη δομική ομοιότητα.
- Στη βιοπληροφορική, η ευθυγράμμιση δικτύου χρησιμοποιείται για τη σύγκριση βιολογικών δικτύων, για την αναγνώριση
  - λειτουργικές ομοιότητες
  - και διαφορές.



Στόχος της ευθυγράμμισης

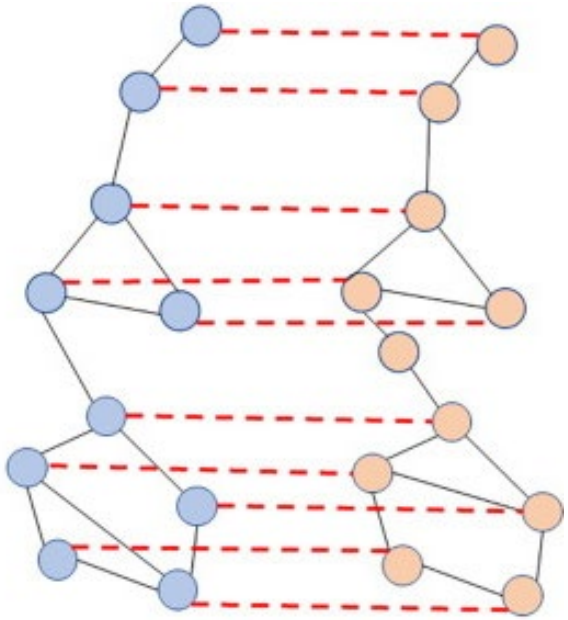
- F εύρεση μιας αντιστοιχίας μεταξύ των κόμβων και των άκρων δύο ή περισσότερων δικτύων.
- Ο στόχος της ευθυγράμμισης δικτύου είναι ο εντοπισμός παρόμοιων ή ομόλογων κόμβων και ακμών σε διαφορετικά δίκτυα.
- Επιτρέπει στους ερευνητές να αξιοποιήσουν τις διαθέσιμες λειτουργικές πληροφορίες σε ένα δίκτυο για να συμπεράνουν ή να προβλέψουν τις λειτουργίες των κόμβων στο άλλο δίκτυο

Ορισμός του προβλήματος

- Ο στόχος του προβλήματος ευθυγράμμισης βιολογικών δικτύων είναι η ομαδοποίηση κόμβων σε διαφορετικά δίκτυα με βάση
  - η βιολογική (αλληλουχία) ομοιότητά τους
  - και τα πρότυπα αλληλεπίδρασης των γειτονικών κοινοτήτων τους (δηλ. ομοιότητα τοπολογίας).
- Βρείτε μια αλληλογραφία ένα προς ένα (ή πολλά-προς-πολλά).

Παγκόσμια ευθυγράμμιση

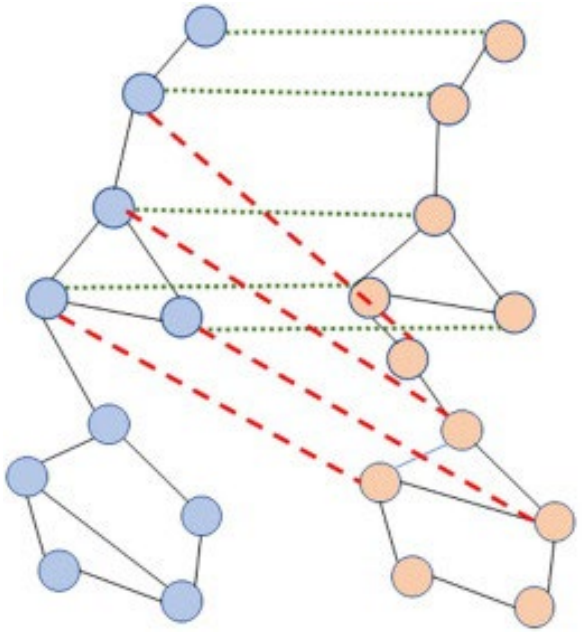
- Η συνολική ευθυγράμμιση δικτύου στοχεύει στην εύρεση μιας συνολικής βέλτιστης χαρτογράφησης κόμβων μεταξύ δύο δικτύων.
- Στην καθολική ευθυγράμμιση δικτύου, ολόκληρα δίκτυα συγκρίνονται και ευθυγραμμίζονται, αντί για τοπικούς υπογράφους.
- Η συνολική ευθυγράμμιση δικτύου μπορεί να διατυπωθεί ως πρόβλημα βελτιστοποίησης, όπου ο στόχος είναι να βρεθεί μια αντιστοίχιση κόμβου που μεγιστοποιεί μια προκαθορισμένη αντικειμενική συνάρτηση.



(b) Global network alignment

Τοπική ευθυγράμμιση

- Εντοπίζω περιοχές υψηλής ομοιότητας μεταξύ δύο δικτύων αντί να ευθυγραμμίζω ολόκληρα τα δίκτυα
- P είναι ιδιαίτερα χρήσιμο για τη σύγκριση δικτύων με διαφορετικές δομές ή μεγέθη
- Ο στόχος είναι να εντοπιστούν διατηρημένες υποδομές μεταξύ των δύο δικτύων, όπως μοτίβα ή λειτουργικές μονάδες
- S et ευθυγραμμισμένων περιοχών ή υποδικτύων, αντί για αντιστοίχιση ένα προς ένα μεταξύ κόμβων

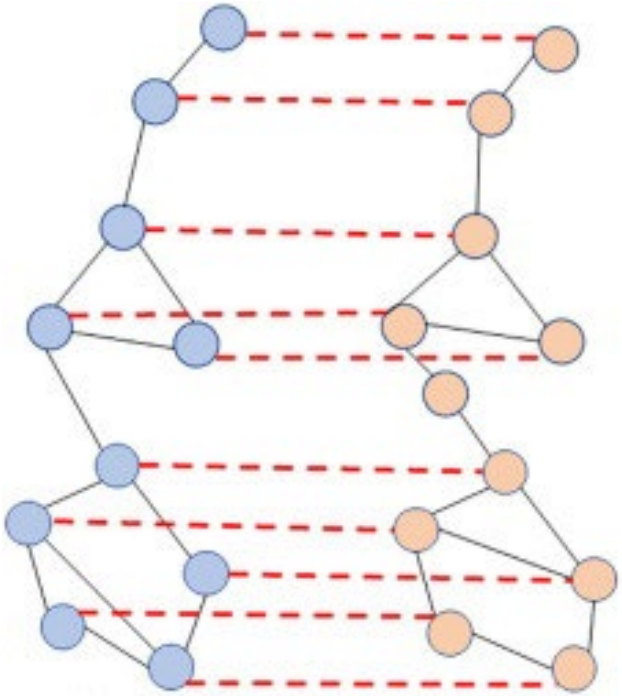


(a) Local network alignment

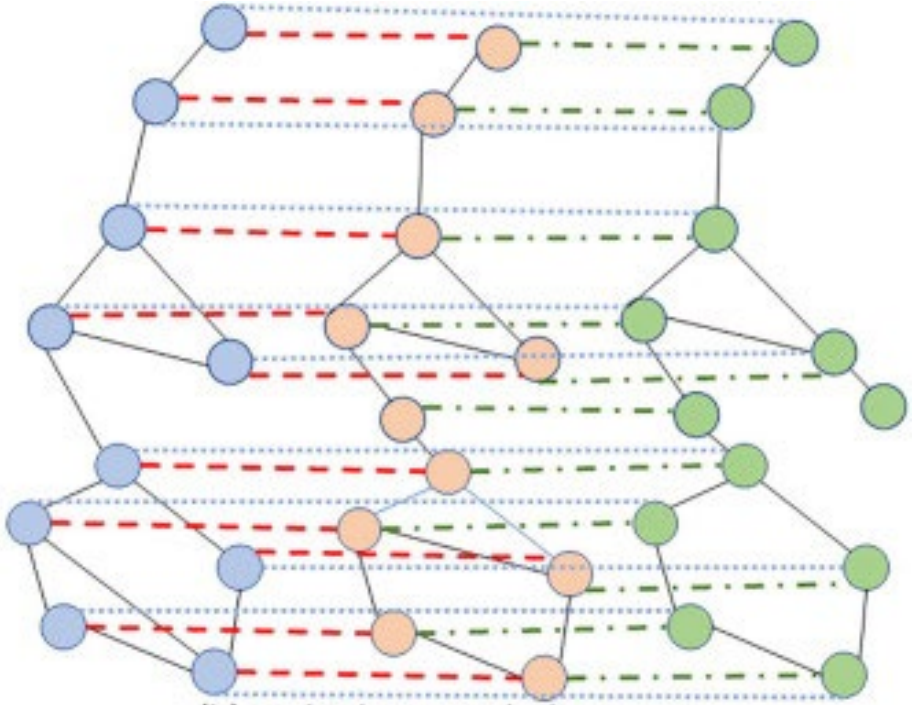
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Pairwise vs Multiple



(a) Pairwise network alignment

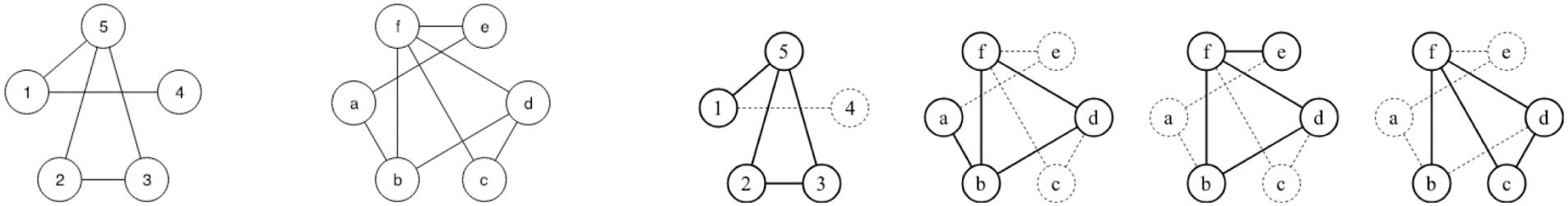


(b) Multiple network alignment

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Μ μέγιστο κοινό υπογράφημα

- Το μέγιστο κοινό υπογράφημα δύο γραφημάτων είναι το μεγαλύτερο δυνατό κοινό υπογράφημα, δηλαδή το κοινό υπογράφημα με όσο το δυνατόν περισσότερες κορυφές.



Επιπτώσεις στη Βιοπληροφορική

- Η ευθυγράμμιση δικτύου είναι ένα ισχυρό εργαλείο για την ανάλυση των βιολογικών δικτύων και τη σύγκριση των δομών τους.
- Παροχή γνώσεων για τις εξελικτικές σχέσεις και τους μηχανισμούς της μοριακής λειτουργίας.
- Με την ευθυγράμμιση δικτύων για συγκεκριμένες ασθένειες, η ευθυγράμμιση δικτύου μπορεί να βοηθήσει στον εντοπισμό νέων θεραπευτικών στόχων και στην πρόβλεψη των αποτελεσμάτων της νόσου.
- Η ευθυγράμμιση δικτύου μπορεί επίσης να χρησιμοποιηθεί για την ενοποίηση δεδομένων multi-omics και τη δημιουργία πιο ολοκληρωμένων μοντέλων δικτύου για την κατανόηση πολύπλοκων βιολογικών διεργασιών.
- Έτσι, η ευθυγράμμιση δικτύου είναι μια ουσιαστική τεχνική στη βιοπληροφορική για την αποκρυπτογράφηση των βιολογικών δικτύων και των λειτουργιών τους.

Co-funded by the Erasmus+ Programme of the European Union

Μέτρα ομοιότητας

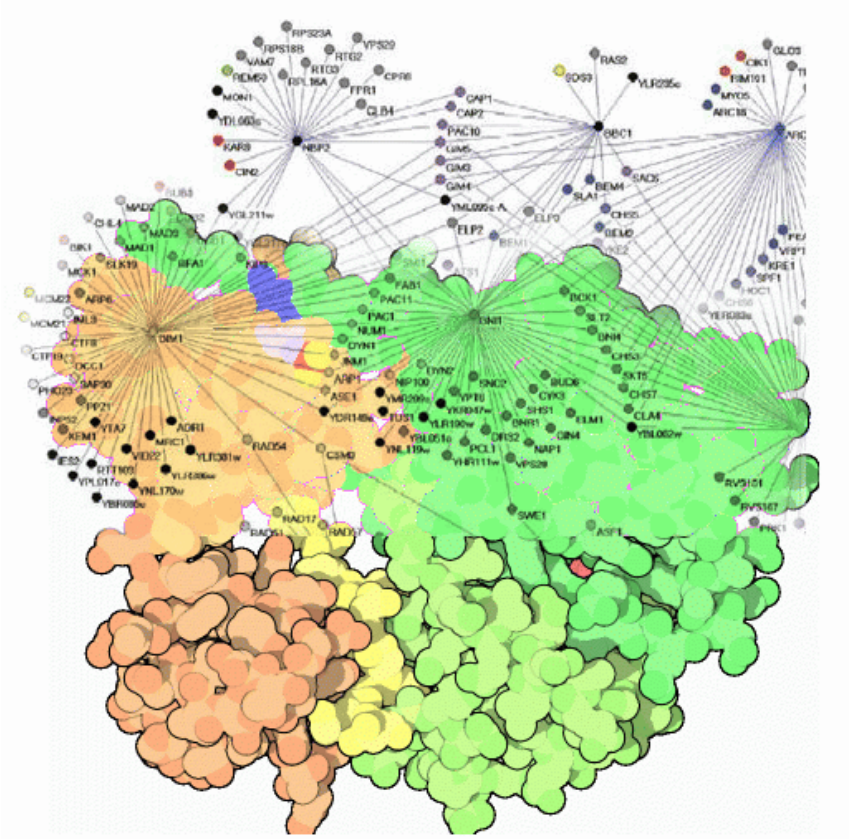
- Τα μέτρα ομοιότητας είναι ζωτικής σημασίας για την αξιολόγηση της ποιότητας των ευθυγραμμίσεων δικτύου
- Βάσει κόμβου
  - Οι κόμβοι με παρόμοια χαρακτηριστικά (π.χ. βαθμός, κεντρική θέση ή λειτουργικός σχολιασμός) ευθυγραμμίζονται μεταξύ τους
- Edged Based
  - Οι άκρες που συνδέουν παρόμοιους κόμβους ευθυγραμμίζονται μεταξύ τους
  - Τα μέτρα που βασίζονται σε ακμές περιλαμβάνουν τοπολογική επικάλυψη, διατήρηση ακμών και κατανομή βαθμών γραφήματος
- Βασισμένο σε δίκτυο
  - Οι συνολικές ιδιότητες δικτύου χρησιμοποιούνται για τη μέτρηση της ομοιότητας μεταξύ των δικτύων
  - Τα μέτρα που βασίζονται σε δίκτυο περιλαμβάνουν τη φασματική ομοιότητα, τη συσχέτιση γραφημάτων και την απόκλιση δικτύου

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

IsoRank

- Ο παγκόσμιος αλγόριθμος ευθυγράμμισης δικτύου που προτείνεται από τους Singh et al.
- Η γενική ιδέα στην κατασκευή του πίνακα R είναι ότι δύο κόμβοι i και j ταιριάζουν καλά εάν οι γείτονές τους ταιριάζουν επίσης καλά



Co-funded by the Erasmus+ Programme of the European Union

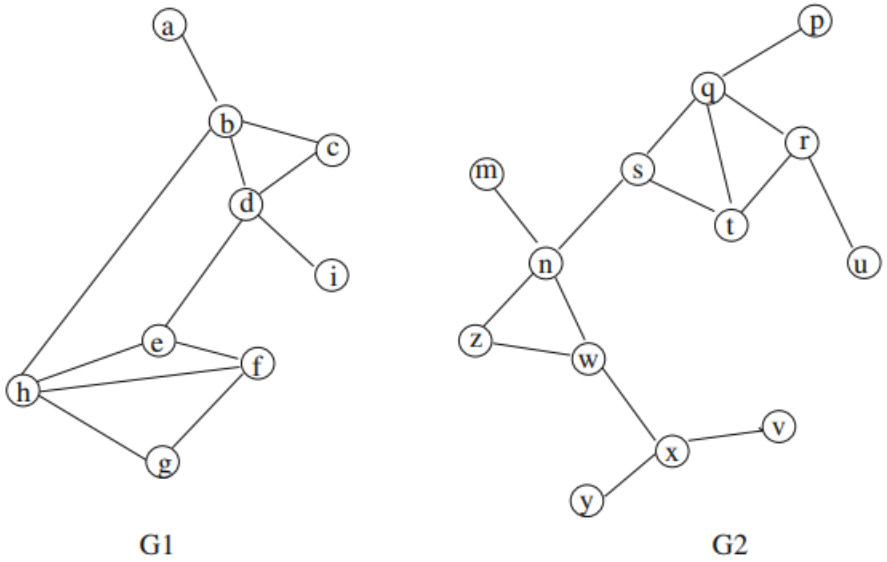
IsoRank

- Πρώτον, υπολογίζει ένα σκορ ομοιότητας για κάθε ζεύγος κόμβων στα δύο δίκτυα και το χρησιμοποιεί για να εντοπίσει τους καλύτερους κόμβους που ταιριάζουν.
- Στο δεύτερο βήμα, εξάγονται κόμβοι που ταιριάζουν πολύ από το R για να βρεθεί η στοίχιση
- Μπορεί να χειριστεί δίκτυα με διαφορετικά μεγέθη και τοπολογίες

Co-funded by the Erasmus+ Programme of the European Union

Ασκηση

- Βρείτε το M μέγιστο κοινό υπογράφημα αυτών των γραφημάτων



G1 G2

Co-funded by the Erasmus+ Programme of the European Union

Περίληψη

- Η ευθυγράμμιση δικτύου είναι ένα ισχυρό εργαλείο για την ανάλυση των βιολογικών δικτύων και τη σύγκριση των δομών τους.
- Διαδικασία αναγνώρισης παρόμοιων δομών σε δύο ή περισσότερα δίκτυα.
- Περιλαμβάνει την εύρεση μιας χαρτογράφησης μεταξύ των κόμβων στα δίκτυα που διατηρεί τη δομική ομοιότητα.
- Τύποι ευθυγράμμισης
  - Τοπικό vs Παγκόσμιο vs Μερικό
  - Pairwise vs Multiple
- Παράδειγμα αλγόριθμου για καθολική στοίχιση
  - IsoRank

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 4. Μαθήματα επιλογής

Μάθημα 1. Εισαγωγή στη Bio Java

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Java για Βιοπληροφορική;

- Cross platform σημαίνει ανάπτυξη σε μια πλατφόρμα ανάπτυξη σε οποιαδήποτε.
- Ευρέως αποδεκτό βιομηχανικό πρότυπο.
- Πολλές βιβλιοθήκες υποστήριξης για σύγχρονες τεχνολογίες (XML, WebServices, JDBC).
- Κλιμακώνεται καλά από μικρά σε βιομηχανικά προγράμματα μεγέθους επιχειρήσεων.

Java για Βιοπληροφορική;

- Αντικειμενοστραφής.
- Ταχεία ανάπτυξη λόγω
  - Πολύ αυστηροί τύποι
  - Απλή σαφής σύνταξη
  - Εξαίρεση χειρισμός και ανάκτηση
  - Σταυρός πλατφόρμα
  - Εκτεταμένη βιβλιοθήκη τάξης
  - Επαναχρησιμοποίηση κώδικα

Τι είναι το BioJava ;

- Μια συλλογή αντικειμένων Java που αντιπροσωπεύουν και χειρίζονται βιολογικά δεδομένα
- Όχι ένα πρόγραμμα, μάλλον μια βιβλιοθήκη προγραμματισμού
- Ανοιχτού κώδικα (LGPL) ανοιχτό για κάθε ανάπτυξη, ακόμη και εμπορική. Όχι «κολλώδης» ή «ιογενής».

Τι είναι το BioJava ;

- Συλλογή αντικειμένων που βοηθούν την έρευνα στη βιοπληροφορική
- Ξεκίνησε στο EBI/Sanger το 1998 από τους Matthew Pocock και Thomas Down
- 25+ προγραμματιστές έχουν συνεισφέρει (5 πυρήνες)

Co-funded by the Erasmus+ Programme of the European Union

Τι είναι το BioJava ;

- Η BioJava έχει αποκτήσει 1100+ κλάσεις, 130.000+ γραμμές κώδικα.
- Χρησιμοποιεί έλεγχο έκδοσης CVS, δοκιμή JUnit και εκδόσεις ANT.
- Τώρα έχει ένα αρκετά σταθερό API.
- 76 πακέτα!

Πού είναι η BioJava

- Αρχική Σελίδα
- www.biojava.org
- BioJava in Anger
- http://www.biojava.org/docs/bj_in_anger/
- Λίστες αλληλογραφίας
- biojava-l@biojava.org
- biojava-dev@biojava.org
- Νυχτερινά Build
- http://www.derkholm.net/autobuild/

Co-funded by the Erasmus+ Programme of the European Union

Απόκτηση BioJava

- Κατεβάστε
  - http://www.biojava.org/download/
  - Λάβετε δυαδικά αρχεία, πηγή και έγγραφα
- biojava-live (απαιτείται βιογραφικό)
  - cvs -d :pserver:cvs@cvs.open-bio.org:/home/repository/biojava login
  - Ο κωδικός είναι "cvs"
  - cvs -d :pserver:cvs@cvs.open-bio.org:/home/repository/biojava checkout biojava-live
  - ενημέρωση -Pd

Σύνταξη biojava -live

- Απαιτεί το εργαλείο κατασκευής ANT
  - http://jakarta.apache.org/ant/
- Το εργαλείο ANT θα χρησιμοποιήσει το build.xml για να
  - Τακτοποιήστε τον πηγαίο κώδικα
  - Μεταγλώττιση πηγής
  - Δημιουργία αρχείου jar
  - Δημιουργήστε έγγραφα Java
  - Δημιουργία επιδείξεων
  - Δημιουργία και εκτέλεση δοκιμών
  - Αλλαγή σε biojava-live. τύπου μυρμήγκι
- Η δοκιμή μονάδας απαιτεί JUnit
  - http://junit.sourceforge.net/

Ρύθμιση του BioJava

- Τοποθετήστε τα ακόλουθα αρχεία JAR στη διαδρομή της τάξης σας:
- biojava.jar
- bytecode-0.92.jar
- commons-cli.jar
- commons-collections-2.1.jar
- commons-dbcp-1.1.jar
- commons-pool-1.1.jar

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Object Orient Patterns και BioJava Design

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

BioJava Design

- Χρησιμοποιεί κάποιες εύλογα «προχωρημένες» έννοιες
  - Σχεδιασμός μέσω διασύνδεσης
  - Προστατευμένοι ή ιδιωτικοί κατασκευαστές
  - Εργοστασιακές τάξεις και μέθοδοι
  - Flyweight/ Αντικείμενα Singleton

Διεπαφές Απόκρυψη υλοποίησης

- Στο BioJava υπάρχουν αρκετές υλοποιήσεις της διεπαφής διανομής.
- Οποιαδήποτε μπορεί να επιστραφεί νόμιμα με μια μέθοδο που επιστρέφει μια διανομή (η μέθοδος επιστροφής μπορεί ακόμη και να επιστρέψει διαφορετικές ανάλογα με την περίπτωση).
- Οποιοδήποτε μπορεί να χρησιμοποιηθεί νομικά ως επιχείρημα σε μια μέθοδο που απαιτεί Κατανομή.
- Όλα είναι εγγυημένα ότι περιέχουν ένα ελάχιστο σύνολο κοινών μεθόδων.

Flyweight και Singleton αντικείμενα

- Ένα Singleton είναι μια κλάση με μόνο μία παρουσία και μόνο ένα σημείο πρόσβασης.
- Ένα Singleton θα χρειαστεί έναν ιδιωτικό κατασκευαστή και μπορεί να είναι στατικό (π.χ. AlphabetManager).
- Ένα αντικείμενο Flyweight χρησιμοποιεί κοινή χρήση για να υποστηρίξει αποτελεσματικά μεγάλους αριθμούς λεπτόκοκκων αντικειμένων.
- Για παράδειγμα, στο BioJava υπάρχει μόνο μία περίπτωση του συμβόλου DNA "A". Μια ακολουθία των A είναι στην πραγματικότητα απλώς μια λίστα δεικτών σε αυτό το αντικείμενο.

Εργοστασιακές και Στατικές μέθοδοι

- Μερικές φορές είναι χρήσιμο να εμποδίζετε έναν χρήστη να κατασκευάσει απευθείας ένα αντικείμενο μέσω ενός κατασκευαστή.
  - Αν η κατασκευή είναι πολύπλοκη.
  - Εάν η επιλογή της βέλτιστης υλοποίησης αφεθεί καλύτερα στον προγραμματιστή του API.
  - Εάν οι σημαντικοί πόροι προστατεύονται καλύτερα από τους τελικούς χρήστες, π.χ. Singletons/ Flyweights.
- Αντί να δημιουργηθεί το αντικείμενο μέσω του κατασκευαστή του, χρησιμοποιείται μια στατική μέθοδος ή ένα αντικείμενο Factory

Co-funded by the Erasmus+ Programme of the European Union

Παραδείγματα

- Στατική μέθοδος:
  - `FiniteAlphabet dna = DNATools.getDNA();`
- Στατικό πεδίο:
  - `DistributionFactory df = DistributionFactory.DEFAULT;`
- Εργοστασιακή μέθοδος:
  - `Κατανομή d = df.createDistribution(dna);`

Δύο επίπεδα BioJava

- Προγραμματισμός τύπου μακροεντολής
  - Κλάσεις εργαλείων (SeqIOTools, DistributionTools κ.λπ.).
  - Στατικές μέθοδοι για κοινές εργασίες.
- Πλήρης προγραμματισμός
  - Πολλές προσαρμογές και «plug and play» είναι δυνατές.
  - Περισσότερη έκθεση στις αιχμηρές άκρες του API. Λιγότερη τεκμηρίωση.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Αλφάβητα, σύμβολα και ακολουθίες

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Σύμβολα

- Στο BioJava το υπόλειμμα DNA "A" είναι ένα αντικείμενο.
- Στο Bioperl το "A" θα ήταν ένα String.
- Το αντικείμενο "A" είναι μέρος της ακολουθίας και όχι της ακολουθίας.
- Το «A» από το DNA δεν είναι ίσο με το «A» από το RNA ή το «A» από την πρωτεΐνη.

Co-funded by the Erasmus+ Programme of the European Union

Γιατί όχι Strings;

- DNA A != RNA A != Πρωτεΐνη A
- Για συμβολοσειρές "A".equals("A");
- Το αλφάβητο DNA περιέχει επίσης K,Y,W,S,R,M,B,D,G,V,N

Co-funded by the Erasmus+ Programme of the European Union

Γιατί όχι Strings;

- Το αντικείμενο Y περιέχει C και T, η συμβολοσειρά "Y" δεν περιέχει τίποτα
- Τα μεταφραστικά HashMaps με συμβολοσειρές είναι ελαττωματικά.
  - Το Biojava GGN μεταφράζεται σε GLY
  - Συμβολοσειρά GGN χάρτες σε null
- Ένας πλήρως περιττός πίνακας μετάφρασης String to String HashMap απαιτεί 4096 κλειδιά!

Τα σύμβολα είναι Κανονικά

- DNATools.a() == DNATools.a();
  - Υπάρχει μόνο μία περίπτωση του 'a'
- DNATools.a().equals(DNATools.a());
- ProteinTools.a() != DNATools.a();
- Ακόμα και σε Remote JVM's!
  - Κατά τη σειριοποίηση, η ευρετηρίαση Alphabet είναι παροδική και «επανασυνδέεται» μέσω των μεθόδων readResolve().

Co-funded by the Erasmus+ Programme of the European Union

Αλφάβητα

- Ένα σύνολο συμβόλων
- Τα αλφάβητα μπορεί να είναι άπειρα
  - DoubleAlphabet, IntegerAlphabet
- Ορισμένα αλφάβητα έχουν πεπερασμένο αριθμό συμβόλων
  - DNA, RNA κ.λπ
- Διεπαφές Alphabet και FiniteAlphabet

org.biojava.bio.Alphabet

boolean **περιέχει** (Symbol s)
Επιστρέφει αν αυτό το Αλφάβητο περιέχει ή όχι το σύμβολο.

Λίστα **getAlphabets** ()
Επιστροφή ταξινομημένης λίστας με τα αλφάβητα που συνθέτουν ένα σύνθετο αλφάβητο.

Σύμβολο **getAmbiguity** (java.util.Set syms)
Λάβετε ένα σύμβολο που αντιπροσωπεύει το σύνολο των συμβόλων στα syms.

Σύμβολο **getGapSymbol** ()
Λάβετε το σύμβολο ασάφειας 'gap' που είναι πιο κατάλληλο για αυτό το αλφάβητο

Συμβολοσειρά **getName** ()
Λάβετε το όνομα του αλφαβήτου.

Σύμβολο **getSymbol** (java.util.List rl)
Λάβετε ένα σύμβολο από το Αλφάβητο που αντιστοιχεί στην καθορισμένη ταξινομημένη λίστα συμβόλων.

SymbolTokenization **getTokenization** (java.lang.Όνομα συμβολοσειράς)
Λάβετε μια SymbolTokenization με όνομα.

void **validate** (Symbol s)
Πραγματοποιεί μια προκαθορισμένη IllegalSymbolException εάν το σύμβολο δεν περιλαμβάνεται σε αυτό το Αλφάβητο.

org.biojava.bio.FiniteAlphabet

- Εκτός από τις προηγούμενες μεθόδους

void **addSymbol** (Symbol s)
Προσθέτει ένα σύμβολο σε αυτό το Αλφάβητο

Iterator **iterator** ()
Ανακτήστε ένα Iterator πάνω από τα σύμβολα σε αυτό το αλφάβητο.

void **removeSymbol** (Symbol s)
Αφαιρέστε ένα σύμβολο από αυτό το αλφάβητο.

size int ()
Ο αριθμός των συμβόλων στο αλφάβητο.

Τα προεπιλεγμένα αλφάβητα

- DNA (a,c,g,t)
- RNA (a,c,g,u)
- ΠΡΩΤΕΪΝΗ (όλα τα αμινοξέα συμπεριλαμβανομένου του «Sel»)
- PROTEIN-TERM (όλες οι ΠΡΩΤΕΪΝΕΣ συν “ * ”)
- ΔΟΜΗ (σύμβολα δομής ΠΣΠ)
- Αλφάβητο όλων των ακεραίων (Άπειρο αλφάβητο)
  - Μπορεί να δημιουργήσει SubInteger Alphabets
- Αλφάβητο όλων των διπλών (Άπειρο αλφάβητο)

Λήψη των κοινών αλφαβήτων

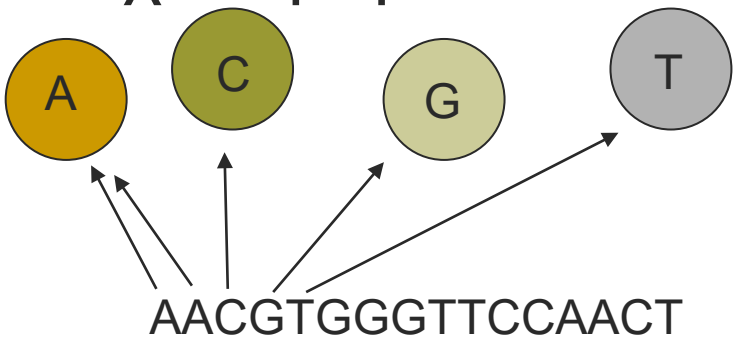
```
εισαγωγή org.biojava.bio.symbol.*;
      εισαγωγή java.util.*;
      εισαγωγή org.biojava.bio.seq.*;

      δημόσιο κλάση AlphabetΠαράδειγμα {
        δημόσιο στατικός void main(String[] args) {
     Alphabet dna, rna, prot;

          //πάρτε το αλφάβητο του DNA με το όνομα
     dna = AlphabetManager.alphabetForName( "DNA" );

          //πάρτε το αλφάβητο του RNA με το όνομα
     rna = AlphabetManager.alphabetForName( "RNA" );

          //πάρτε το αλφάβητο της πρωτεΐνης με το όνομα
     prot = AlphabetManager.alphabetForName( "PROTEIN" );
          //λάβετε το πρωτεϊνικό αλφάβητο που περιλαμβάνει το * τερματισμό
     Σύμβολο
     prot = AlphabetManager.alphabetForName( "ΠΡΟΤΕΙΝΗ-ΟΡΟΣ" );

          //πάρτε τα ίδια αλφάβητα από τις κλάσεις Εργαλεία
     dna = DNATools.getDNA();
     rna = RNATools.getRNA();
     prot = ProteinTools.getAlphabet();
          //ή αυτό με το σύμβολο *
     prot = ProteinTools.getTAlphabet();

     }
     }
```

Οι λίστες συμβόλων αποτελούνται από σύμβολα

- org.biojava.bio.symbol.SymbolList
- Μια ακολουθία συμβόλων από το ίδιο αλφάβητο.
- Χρησιμοποιεί βιολογικές συντεταγμένες από 1 έως μήκος
  - cf Συμβολοσειρά από 0 έως μήκος-1

Αυτό δεν σπαταλά τη μνήμη;

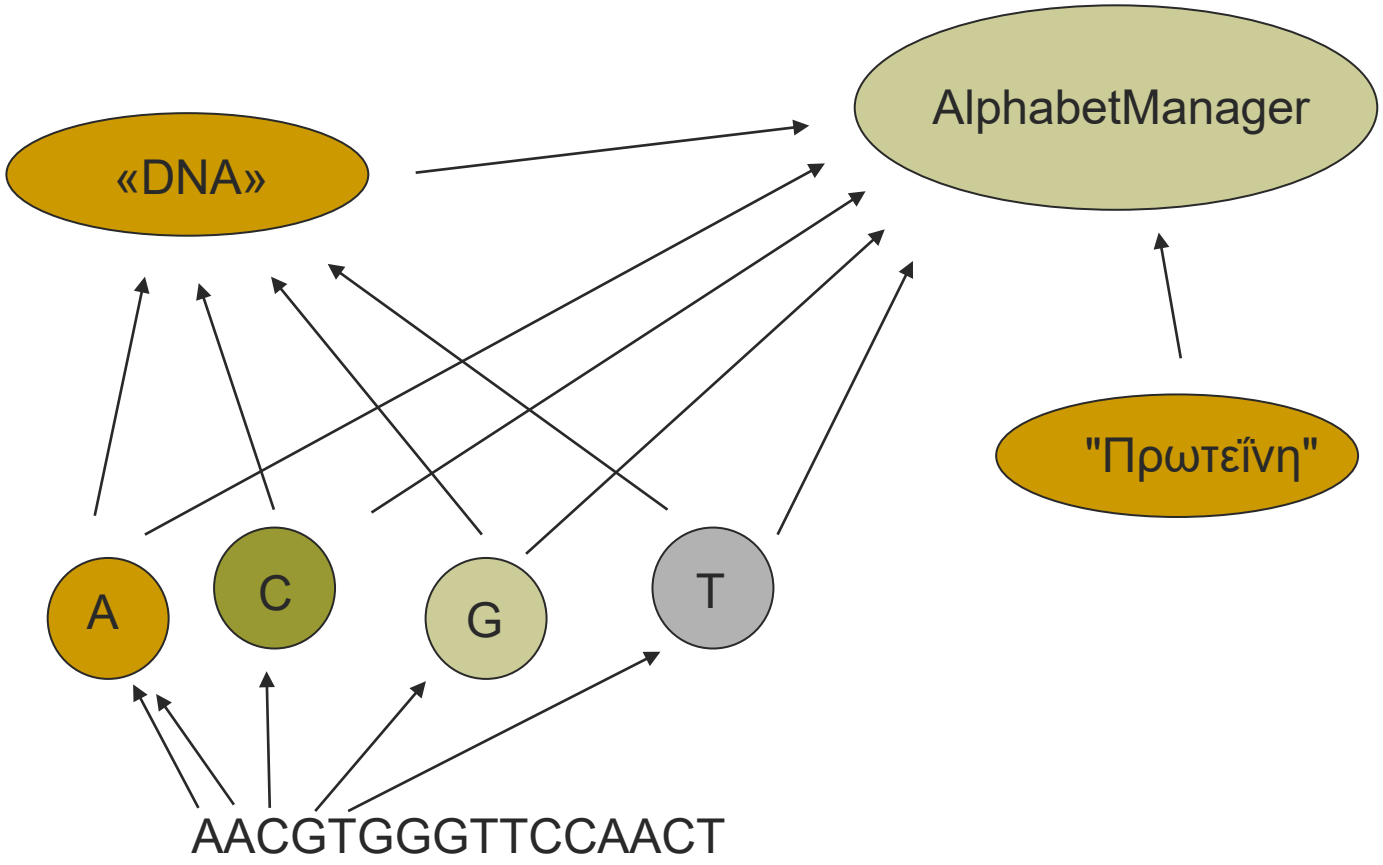
- Μια λίστα συμβόλων δεν είναι στην πραγματικότητα μια λίστα αντικειμένων συμβόλων.
- Μάλλον μια λίστα αναφορών αντικειμένων.
- Ακόμα λίγο πιο βαρύ από ένα κάρβουνο[] αλλά όχι σοβαρό.



Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Η μεγαλύτερη εικόνα



Η διεπαφή SymboList

void **edit** (Επεξεργασία επεξεργασίας)
Εφαρμόστε μια επεξεργασία στη Λίστα συμβόλων όπως καθορίζεται από το αντικείμενο επεξεργασίας.

Αλφάβητο **getAlphabet** ()
Το αλφάβητο ότι αυτή η Λίστα Συμβόλων έχει τελειώσει.

Iterator **iterator** ()
Ένας Iterator για όλα τα σύμβολα σε αυτήν τη Λίστα Συμβόλων.

int **length** ()
Ο αριθμός των συμβόλων σε αυτήν τη Λίστα συμβόλων.

Συμβολοσειρά **seqString** ()
Συμπληρώστε αυτή τη λίστα συμβόλων.

Λίστα συμβόλων **Υπολίστα** (int start, int end)
Επιστρέψτε μια νέα Λίστα Συμβόλων για τα σύμβολα που ξεκινούν μέχρι το τέλος.

String **subStr** (int start, int end)
Επιστρέφετε μια περιοχή αυτής της λίστας συμβόλων ως συμβολοσειρά.

Σύμβολο **symbolAt** (int index)
Επιστρέψτε το σύμβολο στο index, μετρώντας από το 1.

List **toList** ()
Επιστρέφει μια λίστα συμβόλων.

Συμβολοσειρά στη λίστα συμβόλων

```
εισαγωγή org.biojava.bio.seq.*
εισαγωγή org.biojava.bio.symbol.*;

δημόσιο κλάση StringToSymbolList {
δημόσιο στατικός void main(String[] args) {

      δοκιμάστε {
        //δημιουργήστε μια λίστα συμβόλων DNA από μια λίστα συμβόλων
      συμβολοσειράς dna = DNATools.createDNA( "atcggtcggctta" );

        //δημιουργώ μια λίστα συμβόλων RNA από μια συμβολοσειρά
      SymbolList rna = RNATools.createRNA( "auugccuacauaggc" );

        //create a Protein Symbollist from a String
      SymbolList aa = ProteinTools.createProtein( "AGFAVENDSA" );
      }
      catch (IllegalSymbolException ex) {
        //αυτό θα συμβεί εάν χρησιμοποιήσετε έναν χαρακτήρα σε μία από τις
      συμβολοσειρές σας
        //δεν είναι αποδεκτός χαρακτήρας IUB για αυτό το σύμβολο.
      π.χ.printStackTrace();
      }
}
}
```

SymboList σε συμβολοσειρά

```
εισαγωγή org.biojava.bio.symbol.*;

δημόσιο κλάση SymboListToString {

δημόσιο στατικός void main(String[] args) {
    SymbolList sl = null ;

    //κωδικοποιήστε εδώ για στιγμιότυπο sl

    //μετατροπή sl σε συμβολοσειρά
Συμβολοσειρά s = sl.seqString();
}
}
```

Η διεπαφή ακολουθίας

- Μια ακολουθία είναι μια λίστα συμβόλων με περισσότερες πληροφορίες.
- Εκτός από το Annotatable και το SymboList:

Συμβολοσειρά **getName** ()
Το όνομα αυτής της ακολουθίας.

String **getURN** ()
Ένα Uniform Resource Identifier (URI) που προσδιορίζει την ακολουθία που αντιπροσωπεύεται από αυτό το αντικείμενο.

- Επίσης υλοποιεί το FeatureHolder που επιτρέπει την προσθήκη αντικειμένων χαρακτηριστικών.

Δημιουργήστε γρήγορα μια Ακολουθία

```
εισαγωγή org.biojava.bio.seq.*;
      εισαγωγή org.biojava.bio.symbol.*;

      δημόσιο κλάση StringToSequence {
        δημόσιο στατικός void main(String[] args) {

          δοκιμάστε {
            //δημιουργία ακολουθίας DNA με το όνομα dna_1
     Αλληλουχία dna = DNATools.createDNASequence( "atgctg" , "dna_1" );

            //δημιουργήστε μια ακολουθία RNA με το όνομα rna_1
     Αλληλουχία rna = RNATools.createRNASequence( "augcug" , "rna_1" );

            //δημιουργήστε μια ακολουθία πρωτεΐνης με το όνομα prot_1
     Sequence prot = ProteinTools.createProteinSequence( "AFHS" , "prot_1" );
     }
          catch (IllegalSymbolException ex) {
            //μια εξαίρεση δημιουργείται εάν χρησιμοποιείτε ένα σύμβολο που δεν
     είναι IUB
     ex.printStackTrace();
     }
     }
     _
```

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πιο σύνθετα σύμβολα και αλφάβητα

Co-funded by the Erasmus+ Programme of the European Union

Σύμβολα αμφισημίας

- Τα διφορούμενα ή ασαφή δεδομένα είναι μια πραγματικότητα, ειδικά με την αλληλουχία.
- Τα ίχνη DNA μπορούν να περιέχουν σύμβολα όπως n, r, w, v, h, k, y, n κ.λπ.
- Στο BioJava DNA τα σύμβολα a, c, g, t είναι AtomicSymbols.
- Διφορούμενα σύμβολα όπως το y είναι BasisSymbols.

Co-funded by the Erasmus+ Programme of the European Union

Σύμβολα βάσης

- Ένα BasisSymbol μπορεί να αναπαρασταθεί ως μια λίστα με ένα ή περισσότερα σύμβολα.
- Το BasisSymbol επεκτείνει το σύμβολο.
- Τα σύμβολα αμφισημίας είναι πάντα σύμβολα βάσης
- getSymbols() Η λίστα των συμβόλων από τα οποία αποτελείται αυτό το σύμβολο.

Co-funded by the Erasmus+ Programme of the European Union

Ατομικά Σύμβολα

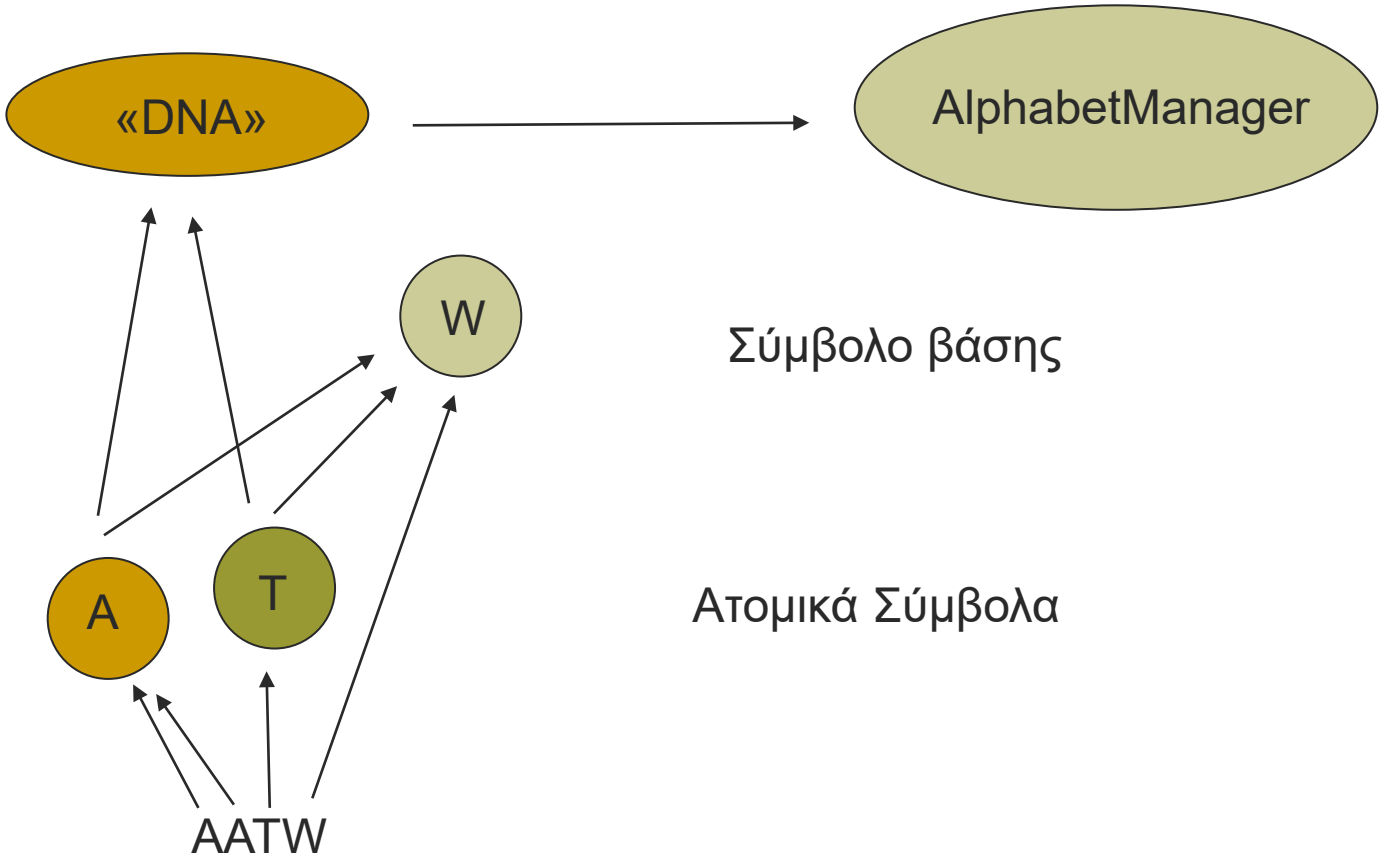
- Τα AtomicSymbols δεν είναι διφορούμενα.
- Δεν μπορούν να χωριστούν περαιτέρω σε Σύμβολα που είναι *έγκυρα μέλη του μητρικού Αλφαβήτου* .
- Στην περίπτωση των σύνθετων αλφαβήτων μπορούν να χωριστούν σε έγκυρα σύμβολα από συστατικά αλφάβητα.

Ατομικά Σύμβολα

- Η διεπαφή AtomicSymbol επεκτείνει το BasisSymbol αλλά δεν προσθέτει νέες μεθόδους μόνο συμβάσεις συμπεριφοράς.
- Οι παρουσίες AtomicSymbol εγγυώνται ότι η getMatches() επιστρέφει ένα Αλφάβητο που περιέχει ακριβώς αυτό το Σύμβολο και κάθε στοιχείο της Λίστας που επιστρέφεται από την getSymbols() είναι επίσης ατομικό.

Co-funded by the Erasmus+ Programme of the European Union

Ατομική και Βάση



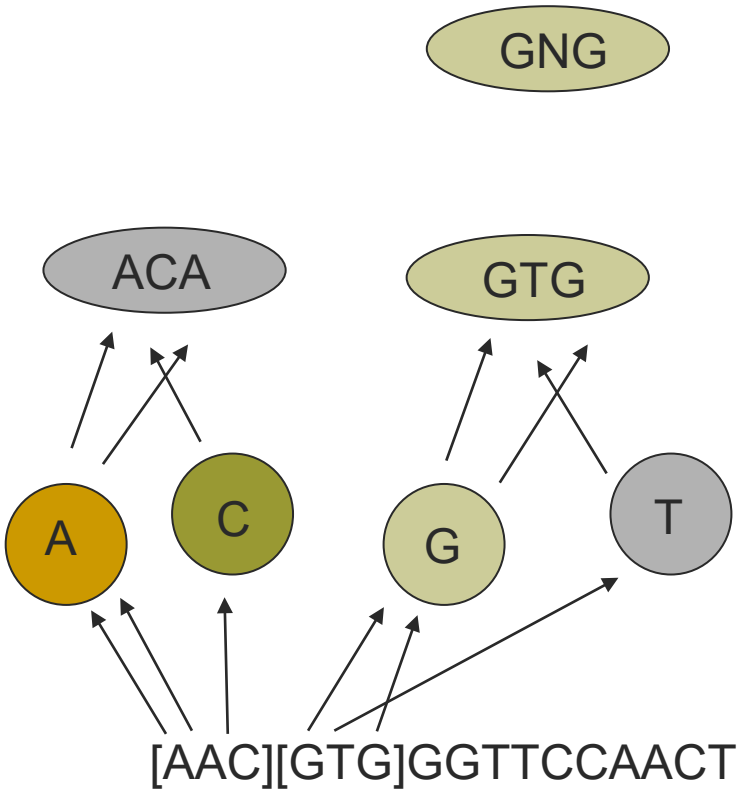
Μεταφραστική αμφισημία

- Η BioJava χειρίζεται τη μετάφραση της ασάφειας πολύ ομαλά.
- DNA 'n' = [a,c,g,t]
- Μεταγράφει σε RNA 'n' [a,c,g,u]
- Το ggn μεταφράζεται σε Gly
- Το agn μεταφράζεται σε [Ser, Arg]
- Οι περισσότερες ασάφειες πρωτεΐνης δεν έχουν «κουπόνι» και εκτυπώνονται ως «X»

Αλφάβητα CrossProduct

- Ένα CrossProductAlphabet είναι ένας συνδυασμός δύο ή περισσότερων αλφαβήτων.
- Οποιοσδήποτε τύπος CrossProductAlphabet είναι δυνατός
- Διμερή (DNA x DNA)
- Κωδόνιο (DNA x DNA x DNA)
- Υπό όρους ((DNA x DNA) x DNA)
- Μικτό ((DNA x DNA x DNA) x ΠΡΩΤΕΪΝΗ)

Πεπερασμένα και Σύνθετα Άλφα



(DNA x DNA x DNA)
Σύμβολο βάσης

(DNA x DNA x DNA)
Ατομικά Σύμβολα

Ατομικά σύμβολα DNA

Co-funded by the Erasmus+ Programme of the European Union

Σε τι χρησιμεύουν;

- Σύμβολα κωδικονίων (DNA x DNA x DNA).
- Πολλές κλάσεις ανάλυσης όπως το Count και η Distribution χρησιμοποιούν το Symbol ως όρισμα. Ένα εξάμερο μπορεί να είναι ένα AtomicSymbol.
- Το Phred είναι DNA x Ακέραιος
- 1ης και ανώτερης τάξης χρησιμοποιούν CrossProductAlphabets ·

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Πώς μπορώ να φτιάξω ένα CrossProductAlphabet ;

```
εισαγωγή java.util.*;
εισαγωγή org.biojava.bio.seq.*;
εισαγωγή org.biojava.bio.symbol.*;

δημόσιο class CrossProduct {
  δημόσιο στατικός void main(String[] args) {

      //δημιουργία CrossProductAlphabet από λίστα λίστας
    l = Collections.nCopies( 3 , DNATools.getDNA());
    Κωδικό αλφαβήτου = AlphabetManager.getCrossProductAlphabet(l);

      //πάρτε το ίδιο αλφάβητο με το όνομα
    Alphabet codon2 =
    AlphabetManager.generateCrossProductAlphaFromName(
"(DNA x DNA x DNA)"
)

      //δείχνετε ότι τα δύο αλφάβητα είναι κανονικά
    System.out.println(codon == codon2);
    }
    }
```

Δημιουργία προβολών τριπλής σε μια λίστα συμβόλων

```
εισαγωγή org.biojava.bio.seq.*;
      εισαγωγή org.biojava.bio.symbol.*;

       δημόσιο κλάση CodonView {
         δημόσιο στατικός void main(String[] args) {
           δοκιμάστε {
             //δημιουργία λίστας συμβόλων DNA SymbolList
      dna = DNATools.createDNA( "atgcccgcgtaa" );
      System.out.println( "Μήκος dna " + dna.length());

             //πάρτε μια προβολή κωδικονίου (μέγεθος παραθύρου τριών)
      Κωδόνια SymbolList = SymbolListViews.windowedSymbolList(dna, 3 );
      System.out.println( "Μήκος κωδικονίων " + codons.length());

             //get a Triplet View
      SymbolList triplets = SymbolListViews.orderNSymbolList(dna, 3 );
      System.out.println( "Μήκος τριπλών " + triplets.length());
      }
           catch (Exception ex) {
      ex.printStackTrace();
      }
      }
      _
```

Λήψη συμβόλου για ένα κωδικόνιο

```
εισαγωγή java.util.*;

      εισαγωγή org.biojava.bio.seq.*;
      εισαγωγή org.biojava.bio.symbol.*;

      δημόσιο κλάση MakeATG {
        δημόσιο στατικός void main(String[] args) {
          //δημιουργία CrossProductAlphabet από λίστα λίστας
     l = Collections.nCopies( 3 , DNATools.getDNA());
     Κωδικό αλφαβήτου = AlphabetManager.getCrossProductAlphabet(l);

          //πάρτε το κωδικόνιο από atg
     List syms = new ArrayList( 3 );
     syms.add(DNATools.a());
     syms.add(DNATools.t());
     syms.add(DNATools.g());

     Σύμβολο atg = null ;
          δοκιμάστε {
     atg = codon.getSymbol(syms);
     }
          catch (IllegalSymbolException ex) {
            //χρησιμοποιημένο σύμβολο από το Alphabet που δεν αποτελεί στοιχείο
     του κωδικονίου
     ex.printStackTrace();
     }
     System.out.println( "Όνομα atg: " + atg.getName());
     }
     }
```

Σπάζοντας ένα κωδικόνιο στα μέρη του

```
εισαγωγή java.util.*;
      εισαγωγή org.biojava.bio.seq.*;
      εισαγωγή org.biojava.bio.symbol.*;

      δημόσιο class BreakingComponents {
        δημόσιο στατικός void main(String[] args) {
          //δημιουργήστε τη λίστα αλφαβήτου 'κωδικού'
     l = Collections.nCopies( 3 , DNATools.getDNA());
     Alphabet alpha = AlphabetManager.getCrossProductAlphabet(l);

          //πάρτε το πρώτο σύμβολο στο αλφάβητο
     Iterator iter = ((FiniteAlphabet)alpha).iterator();
     Κωδόνιο AtomicSymbol = (AtomicSymbol)iter.next();
     System.out.print(codon.getName()+ " αποτελείται από: " );

          // σπάσε το σε λίστα τα συστατικά του
     Λίστα συμβόλων = codon.getSymbols();
          for ( int i = 0 ; i < symbols.size(); i++){
            if (i != 0 )
     System.out.print( ", " );
     Symbol sym = (Symbol)symbols.get(i);
     System.out.print(sym.getName());
     }
     }

     _
```

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Λειτουργίες βασικής ακολουθίας

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Λήψη ενός τμήματος μιας λίστας συμβόλων

- symbolAt(int i)
  - Επιστρέφει ένα σύμβολο
- Υπολίστα (int min, int max)
  - Επιστρέφει μια λίστα συμβόλων
- δευτερεύουσα συμβολοσειρά (int min, int max)
  - Επιστρέφει την υποενότητα με διακριτικό σε μια συμβολοσειρά

Μεταγραφή

- Στο BioJava οι αλληλουχίες DNA και οι αλληλουχίες RNA είναι από διαφορετικά αλφάβητα. Για να μετατρέψετε μεταξύ τους:

```
//δημιουργήστε μια λίστα συμβόλων DNA
SymbolList dna = DNATools.createDNA( "atgccgaatcgtaa" );

//μετατρέψτε το σε RNA
SymbolList rna = DNATools.toRNA(dna);

// για να αποδείξω ότι λειτούργησε
System.out.println(rna.seqString()); //augccgaaucguaa

//βιολογική μεταγραφή (δηλαδή αντιγραφή και αντίστροφη
    αλυσίδα)
rna = DNATools.transcribeToRNA(dna); //5' atgccgaatcgtaa 3'
System.out.println(rna.seqString()); //5' uuacgauucggcau 3'
```

Αντίστροφο συμπλήρωμα

```
εισαγωγή org.biojava.bio.symbol.*;
 εισαγωγή org.biojava.bio.seq.*;

 δημόσιο κλάση ReverseCompiment {
   δημόσιο στατικός void main(String[] args) ρίχνει Exception{
SymbolList forward = DNATools.createDNA( "atcgctagcgatcg" );

    //δύο βήματα
SymbolList reverse = SymboListViews.reverse(forward);
SymboList revc1 = DNATools.complement(reverse);

    //one step
SymbolList revc2 = DNATools.reverseComplement(forward);

    //τεστ για ισοδυναμία
System.out.println(revc1.equals(revc2));
}
}
```

Co-funded by the Erasmus+ Programme of the European Union

Μετάφραση

- Το RNATools περιέχει τον «Καθολικό» Πίνακα Μετάφρασης RNA σε πρωτεΐνη.

- Η τυπική διαδικασία είναι η μεταγραφή του DNA σε RNA και στη συνέχεια η μετάφραση.

Παράδειγμα μετάφρασης

```
εισαγωγή org.biojava.bio.symbol.*;
 εισαγωγή org.biojava.bio.seq.*;

 δημόσιο τάξη Μετάφραση {

   δημόσιο στατικός void main(String[] args) {
     δοκιμάστε {
       //δημιουργία λίστας συμβόλων DNA SymbolList
symL = DNATools.createDNA( "atggccattgaatga" );

       //μεταγραφή σε RNA
symL = RNATools.toRNA(symL);

       //translate σε πρωτεΐνη
symL = RNATools.translate(symL);

       //αποδείξει ότι λειτούργησε
System.out.println(symL.seqString());
}
     catch (Exception ex) {
ex.printStackTrace()
}
}
}
```

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ακολουθία Ι/Ο

Μην γράψετε ποτέ άλλο Parser

- Αν μπορείτε να το αποφύγετε!
- Υποστηρίζει BioJava
  - Genbank, GenPept, RefSeq, EMBL, SwissProt, PDB, Fasta, ABI, LocusLink, Unigene (απαιτείται Java 1.4)
  - GAME, AGABE
  - Blast, Fasta, HMMER (μοντέλα και αποτελέσματα), BlastXML, MEME, Phred
  - OBDA, BioIndex, BioSQL, DAS, GFF, XFF
  - Ensembl (με συσκευασία biojava-ensembl)
- StAX/ Τιμή ετικέτας
- RMI και Serialization

Co-funded by the Erasmus+ Programme of the European Union

Απλή I/O

- Οι περισσότερες από τις απλούστερες λειτουργίες εισόδου/εξόδου της BioJava περικλείονται άνετα πίσω από στατικές μεθόδους από την κλάση SeqIOTools.
- Το SeqIOTools μπορεί να διαβάζει και να γράφει:
  - Fasta (πρωτεΐνη ή DNA)
  - EMBL
  - GenBank (επίπεδο αρχείο και XML)
  - SwissProt
  - GenPept
  - ΓΧΣ (πρωτεΐνη ή DNA)
  - Fasta Ευθυγραμμίσεις

Co-funded by the Erasmus+ Programme of the European Union

Μέθοδοι ανάγνωσης SeqIOTools

```
SequenceIterator i = SeqIOTools.readGenbank(br);
SequenceIterator i = SeqIOTools.readGenpept(br);
SequenceIterator i = SeqIOTools.readSwissprot(br);
SequenceIterator i = SeqIOTools.readEmbl(br);
```

και τα λοιπά…

```
SequenceIterator i = (SequenceIterator)
    SeqIOTools.fileToBiojava("fasta", "dna", br);
```

```
Στοίχιση a =
(Στοίχιση) SeqIOTools.fileToBiojava("MSF", "rna", br);
```

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Χαρακτηριστικά, Τοποθεσίες, Σχολιασμοί

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Χαρακτηριστικά και σχολιασμοί

- Τα δεδομένα αλληλουχίας συχνά συνοδεύονται από πρόσθετες πληροφορίες σχετικά με τις διάφορες ιδιότητες της ακολουθίας (Genbank, SwissProt κ.λπ.).
- Το BioJava διαιρεί αυτές τις πληροφορίες σε καθολικές ιδιότητες (Σχολιασμοί) και Τοπικές ιδιότητες (Δυνατότητες).

Co-funded by the Erasmus+ Programme of the European Union

Σημειώσιμος

- Το Annotatable είναι μια διεπαφή "mix-in" που υποδεικνύει ότι το αντικείμενο υλοποίησης περιέχει ένα αντικείμενο Annotation.
- Ορίζει μία μέθοδο.
  - Σχολιασμός getAnnotation();

Co-funded by the Erasmus+ Programme of the European Union

Σχολιασμοί

- org.biojava.bio.Σχολιασμός
- Οι σχολιασμοί χρησιμοποιούνται για καθολικές ιδιότητες.
- Είδος, Αριθμός Πρόσβασης, xrefs, ημερομηνία, δημοσίευση.
- Χάρτες κλειδιού – αξίας.
- Το κλειδί και η τιμή είναι αντικείμενα αλλά σχεδόν πάντα είναι συμβολοσειρές.
- Annotation.EMPTY_ANNOTATION
  - τάξη στατικής ευκολίας
  - καλός κάτοχος θέσης, αποφεύγει τις εξαιρέσεις μηδενικού δείκτη
  - αμετάβλητος

Co-funded by the Erasmus+ Programme of the European Union

Annotation API

map **asMap** ()
Επιστρέψτε έναν χάρτη που περιέχει τα ίδια κλειδιά/τιμές με αυτόν τον σχολιασμό.

boolean **containProperty** (java.lang.Key αντικειμένου)
Επιστρέφει εάν έχει οριστεί η ιδιότητα εκεί.

Αντικείμενο **getProperty** (java.lang.key αντικειμένου)
Ανακτήστε την τιμή μιας ιδιότητας ανά κλειδί.

Πλήκτρα ρύθμισης ()
Λάβετε ένα σύνολο αντικειμένων κλειδιών.

void **removeProperty** (java.lang.key αντικειμένου)
Διαγραφή ιδιότητας

void **setProperty** (java.lang.key αντικειμένου, java.lang.Τιμή αντικειμένου)
Ορίστε την τιμή μιας ιδιότητας.

Co-funded by the Erasmus+ Programme of the European Union

FeatureHolder

- Το FeatureHolder είναι μια άλλη διεπαφή "mix-in" που επιτρέπει στο αντικείμενο υλοποίησης να κρατά Features.
- Η ακολουθία υλοποιεί το FeatureHolder.
- Οι δυνατότητες δημιουργούνται από FeatureHolders.
- Τα FeatureHolders μπορούν να φιλτραριστούν.

Μέθοδοι FeatureHolder

boolean **containFeature** (Χαρακτηριστικό στ)
Ελέγξτε εάν η δυνατότητα υπάρχει σε αυτήν τη θήκη.

int **countFeatures** ()
Μετρήστε πόσα χαρακτηριστικά περιέχονται.

χαρακτηριστικό **createFeature** (Feature.Template ft)
Δημιουργήστε ένα νέο Feature και προσθέστε το σε αυτό το FeatureHolder.

Χαρακτηριστικά Iterator ()
Επανάληψη πάνω από τα χαρακτηριστικά με καμία καλά καθορισμένη σειρά.

FeatureHolder **φίλτρο** (Φίλτρο FeatureFilter)
Αναζητήστε αυτό το σύνολο δυνατοτήτων χρησιμοποιώντας ένα παρεχόμενο FeatureFilter.

FeatureHolder **φίλτρο** (FeatureFilter fc, boolean recurse)
Επιστρέψτε ένα νέο FeatureHolder που περιέχει όλα τα παιδιά αυτού που πέρασαν το φίλτρο fc.

Φίλτρο χαρακτηριστικών **getSchema** ()
Επιστρέψτε ένα φίλτρο σχήματος για αυτό το FeatureHolder.

void **removeFeature** (Feature f)
Αφαιρέστε ένα χαρακτηριστικό από αυτό το FeatureHolder.

Τα χαρακτηριστικά είναι σχολιάσιμα

- Τα χαρακτηριστικά υλοποιούν το Annotatable
  - Μπορεί να κρατήσει έναν σχολιασμό
  - Καθολικοί σχολιασμοί ενός χαρακτηριστικού
    - /Note:
    - /db_xref: κ.λπ

Τα χαρακτηριστικά μπορεί να είναι ένθετα

- Τα χαρακτηριστικά υλοποιούν το FeatureHolder!
  - Επομένως, τα Features ενδέχεται να περιέχουν ένθετα χαρακτηριστικά
  - cf Το μενού AWT είναι ένα στοιχείο μενού
  - π.χ. Ένα γονίδιο έχει εξόνια και εσώνια
  - Το φιλτράρισμα μπορεί να είναι αναδρομικό
  - Ένα χαρακτηριστικό *δεν μπορεί* να συγκρατηθεί (άμεσα ή έμμεσα)

Co-funded by the Erasmus+ Programme of the European Union

Τοποθεσία API

- Οι θέσεις είναι αντικείμενα που καθορίζουν ένα ελάχιστο και μέγιστο όριο σε μια περιοχή ακολουθίας.
- Περιέχει μερικές χρήσιμες μεθόδους, ιδιαίτερα getMin() και getMax().
- Πολλές μέθοδοι έχουν καταργηθεί και έχουν πλέον εκχωρηθεί στο LocationTools.
- Το LocationTools είναι το καλύτερο μέρος για τη λήψη νέων παρουσιών μιας Τοποθεσίας.
- PointLocation, RangeLocation, CircularLocation, CompoundLocation.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Εργαλεία τοποθεσίας

static boolean **areEqual** (Location locA, Location locB)
Επιστρέφετε εάν δύο θέσεις είναι ίσες.

Το στατικό boolean **περιέχει** (Location locA, Location locB)
Επιστρέφει αληθές εάν όλοι οι δείκτες στο locB περιέχονται επίσης από το locA.

στατική Τοποθεσία **flip** (Location loc, int len)
Αναστρέφει μια τοποθεσία σε σχέση με ένα μήκος.

στατική Τοποθεσία **διασταύρωση** (Τοποθεσία locA, Τοποθεσία locB)
Επιστρέψτε τη διασταύρωση δύο τοποθεσιών.

στατική κυκλική θέση **makeCircularLocation** (int min, int max, int seqLength)
Μια απλή μέθοδος για τη δημιουργία ενός RangeLocation τυλιγμένο σε μια CircularLocation

στατική Τοποθεσία **makeLocation** (int min, int max)
Επιστρέψτε μια συνεχόμενη τοποθεσία από το min στο max.

στατικές δυαδικές **επικαλύψεις** (Τοποθεσία locA, τοποθεσία locB)
Καθορίζει εάν οι τοποθεσίες επικαλύπτονται ή όχι.

στατική Τοποθεσία **αφαίρεση** (Θέση x, Θέση y)
Αφαιρέστε μια τοποθεσία από μια άλλη.

στατική Τοποθεσία **ένωση** (java.util.Collection locs)
Η ένωση n-way μιας συλλογής τοποθεσιών.static

Τοποθεσία **ένωση** (Location locA, Location locB)
Επιστρέψτε την ένωση δύο τοποθεσιών.

Παράδειγμα τοποθεσίας

```
εισαγωγή org.biojava.bio.symbol.*;
 εισαγωγή org.biojava.bio.seq.*;

 δημόσιο κλάση SpecifyRange {
   δημόσιο στατικός void main(String[] args) {
     δοκιμάστε {
       //δημιουργία RangeLocation καθορίζοντας τα υπολείμματα 3-8
Location loc = LocationTools.makeLocation( 3 , 8 );
       //εκτύπωση της τοποθεσίας
System.out.println( "Τοποθεσία: " +loc.toString());

       //δημιουργία λίστας συμβόλων
SymbolList sl = RNATools.createRNA( "gcagcuaggcggaaggagc" );
System.out.println( "SymbolList: " +sl.seqString());

       //λάβετε τη Λίστα Συμβόλων που καθορίζεται από τη
Λίστα Συμβόλων Τοποθεσίας sym = loc.symbols(sl);
System.out.println( "Symbols specified by Location: " +sym.seqString());
}
     catch (IllegalSymbolException ex) {
       //παράνομο σύμβολο που χρησιμοποιείται για τη δημιουργία του sl
ex.printStackTrace();
}
}
_
```

Λειτουργίες φιλτραρίσματος

- Τα FeatureHolders έχουν μια μέθοδο φιλτραρίσματος που δέχεται ένα FeatureFilter ως όρισμα.
- Οι δυνατότητες που γίνονται αποδεκτές από το FeatureFilter επιστρέφονται ως νέο FeatureHolder.
- Το φιλτράρισμα μπορεί να γίνει αναδρομικά, έτσι ώστε οι ένθετες λειτουργίες να υπόκεινται στο ίδιο FeatureFilter .

Φίλτρα χαρακτηριστικών

- Το FeatureFilter είναι μια διεπαφή που καθορίζει μία μέθοδο.
  - boolean αποδοχή (Δυνατότητα f)
- Υπάρχουν 26 υλοποιήσεις του FeatureFilter στο BioJava που είναι διαθέσιμες ως εσωτερικές κατηγορίες της διεπαφής FeatureFilter.
- Οι πιο συχνά χρησιμοποιούμενες είναι οι ByType, BySource, StrandFilter, OverlapsLocation, ContainedByLocation.
- Επίσης φίλτρα λογικής λογικής: Και, Ή, Όχι

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ανάλυση και Διανομές

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Διανομές και μετρήσεις

- Οι διεπαφές Distribution και Count προέρχονται από το πακέτο org.biojava.bio.dist.
- Οι μετρήσεις είναι χάρτες από AtomicSymbols έως μετρήσεις.
- Οι διανομές είναι χάρτες από σύμβολα έως συχνότητες.

Co-funded by the Erasmus+ Programme of the European Union

Διανομές

- Οι διανομές είναι κεντρικές για την ανάλυση
- Χάρτης σύμβολα στις συχνότητες
- Μπορεί να προπονηθεί ή να ρυθμιστεί βάρη
- Χρησιμοποιείται πολύ σε πακέτο dp (δυναμικός προγραμματισμός).
  - Μεταβάσεις και εκπομπές HMM
- Πολλές υλοποιήσεις που χρησιμοποιούνται συχνά είναι:
  - SimpleDistribution
  - ΠαραγγελίαΔιανομή
  - Ομοιόμορφη Κατανομή

Co-funded by the Erasmus+ Programme of the European Union

API διανομής

Αλφάβητο **getAlphabet** ()

Το αλφάβητο από το οποίο αυτό το φάσμα εκπέμπει σύμβολα.

Διανομή **getNullModel** ()

Ανακτήστε το null model Distribution που αναγνωρίζει αυτή η Distribution.

διπλό **βάρος** (σύμβολο)

Επιστρέψτε την πιθανότητα ότι το σύμβολο s εκπέμπεται από αυτό το φάσμα.

void **registerWithTrainer** (DistributionTrainerContext dtc)

Καταχωρίστε αυτήν τη διανομή με ένα εκπαιδευτικό πλαίσιο.

Σύμβολο **δείγμαΣύμβολο** ()

Δείτε ένα σύμβολο από την κατανομή πιθανοτήτων αυτής της κατάστασης.

void **setNullModel** (Distribution nullDist)

Ορίστε το μηδενικό μοντέλο διανομής που αναγνωρίζει αυτή η διανομή.

void **setWeight** (Σύμβολο s, διπλό w)

Ορίστε την πιθανότητα ή τις πιθανότητες να εκπέμπεται το σύμβολο s από αυτήν την κατάσταση.

Co-funded by the Erasmus+ Programme of the European Union

Distribution Factory

- Γενικά μια διανομή δημιουργείται χρησιμοποιώντας ένα DistributionFactory.
- Η διεπαφή DistributionFactory περιέχει μια στατική εσωτερική κλάση που ονομάζεται DEFAULT που υλοποιεί το DistributionFactory

- DistributionFactory df = DistributionFactory.DEFAULT;
- Διανομή d = df.createDistribution(dna.getAlphabet());

Co-funded by the Erasmus+ Programme of the European Union

Εκπαίδευση Διανομής

- Οι κατανομές μπορούν να εκπαιδευτούν σε παρατηρούμενες ακολουθίες χρησιμοποιώντας ένα DistributionTrainerContext.
- Μία ή περισσότερες Διανομές μπορούν να καταχωρηθούν με τον κωδικό DTC.
  - //καταχωρίστε τις Διανομές με τον εκπαιδευτή
    dtc.registerDistribution(dnaDist);

DistributionTrainerContext

- Ένα DistributionTrainer εκχωρείται σε κάθε καταχωρημένη διανομή από το DTC.
- Εάν απαιτείται ασυνήθιστη συμπεριφορά εκπαίδευσης, μπορείτε να καταχωρίσετε το δικό σας DistributionTrainer ταυτόχρονα.
- Το dtc μπορεί επίσης να προσθέσει ψευδομετρήσεις εάν χρειάζεται.
- Οι ασάφειες αντιμετωπίζονται αυτόματα.
  - Οι μετρήσεις χωρίζονται σύμφωνα με το μηδενικό μοντέλο.

Παράδειγμα εκπαίδευσης

```
//δημιουργία λίστας συμβόλων DNA SymbolList
dna = DNATools.createDNA( "atcgctagcgtyagcntatsggca" );

//πάρτε ένα DistributionTrainerContext
DistributionTrainerContext dtc = new SimpleDistributionTrainerContext();

//make the Distribution
Distribution dnaDist =
DistributionFactory.DEFAULT.createDistribution(dna.getAlphabet());

//καταχωρίστε τη διανομή με τον εκπαιδευτή
dtc.registerDistribution(dnaDist);

for ( int j = 1 ; j <= dna.length(); j++){
dtc.addCount(dnaDist, dna.symbolAt(j), 1.0 );
}

//train the Distribution
dtc.train();
```

setWeight () Παράδειγμα

```
FiniteAlphabet a = DNATools.getDNA();
Distribution d =
    DistributionFactory.DEFAULT.createDistribution(a);
//ορίστε το βάρος κάθε συμβόλου
d.setWeight(DNATools.a(), 0,3 );
d.setWeight(DNATools.c(), 0.2 );
d.setWeight(DNATools.g(), 0,2 );
d.setWeight(DNATools.t(), 0,3 );
```

Co-funded by the Erasmus+ Programme of the European Union

Εργαλεία διανομής

- Το DistributionTools διατηρεί στατικές μεθόδους για τη δημιουργία και τον χειρισμό Διανομών.
- Οι εργασίες περιλαμβάνουν:
  - Ίσα φάσματα εκπομπής;
  - Shannon Entropy, πληροφορίες, KL Distance.
  - Δημιουργήστε μεροληπτικές ακολουθίες.
  - Κάντε μια κατανομή[] από μια στοίχιση (κάθε διανομή αντιπροσωπεύει μια θέση σε μια ευθυγράμμιση.
  - Μέσος όρος δύο ή περισσότερων διανομών.
  - Τυχαιοποιήστε μια κατανομή.
  - Κάντε μια διανομή από μια καταμέτρηση.

Σειριοποίηση Διανομών

- Οι διανομές είναι σειριοποιήσιμες
  - Γράψτε και διαβάστε από το Binary
  - RMI
- XMLDistributionWriter
  - Γράψτε οποιαδήποτε διανομή σε μια ροή σε μορφή XML.
- XMLDistributionReader
  - SAXParser
  - Διαβάστε οποιαδήποτε διανομή από μια ροή XML

Co-funded by the Erasmus+ Programme of the European Union

Έξοδος XML

```
<?xml version="1.0" ?>
< Τύπος διανομής =" Διανομή ">
   < αλφάβητο όνομα =" DNA " />
   < βάρος sym =" αδενίνη " prob =" 0.32178516910737204 " />
   < βάρος sym =" κυτοσίνη " prob =" 0.0459619929395364 " />
   < βάρος sym =" γουανίνη " prob =" 0.1405504188012911 " />
   < βάρος sym =" θυμίνη " prob =" 0.491702419097383 2 " />
</ Διανομή >
```

Co-funded by the Erasmus+ Programme of the European Union

Τι άλλο??

- Δυναμικός Προγραμματισμός (HMM)
- Βιβλιογραφία
- Ευθυγραμμίσεις
- Ανάλυση Blast και Fasta

Co-funded by the Erasmus+ Programme of the European Union

Τι άλλο??

- Υποστήριξη BioSQL
- Στοιχεία GUI
- Χρωματογραφήματα
- Μοριακή Βιολογία (pI, μάζα, ένζυμα περιορισμού)
- Μοριακή Δομή

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 4. Μαθήματα επιλογής

Μάθημα 2. Εισαγωγή στο BioPython

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Λειτουργικότητα και εργαλεία Biopython

- Η δυνατότητα ανάλυσης αρχείων βιοπληροφορικής σε δομές δεδομένων που μπορούν να χρησιμοποιηθούν από την Python
- Υποστηρίξτε τις ακόλουθες μορφές:
  - Έξοδος έκρηξης
  - Clustalw
  - ΦΑΣΤΑ
  - PubMed και Medline
  - Αρχεία ExpASy
  - SCOP
  - SwissProt
  - ΠΣΠ
- Τα αρχεία στις υποστηριζόμενες μορφές μπορούν να επαναληφθούν σε εγγραφή ανά εγγραφή ή να ευρετηριαστούν και να προσπελαστούν μέσω μιας διεπαφής λεξικού

Λειτουργικότητα και εργαλεία Biopython

- Κώδικας για την αντιμετώπιση προορισμών on-line βιοπληροφορικής (NCBI, ExPASy)
- Διεπαφή σε κοινά προγράμματα βιοπληροφορικής (Blast, ClustalW)
- Μια ακολουθία obj που ασχολείται με seqs, seq IDs, seq features
- Εργαλεία για πράξεις σε ακολουθίες
- Εργαλεία για την αντιμετώπιση ευθυγραμμίσεων
- Εργαλεία για τη διαχείριση των πρωτεϊνικών δομών
- Εργαλεία για την εκτέλεση εφαρμογών

Co-funded by the Erasmus+ Programme of the European Union

Το όνομα της ενότητας Biopython είναι **`Bio`**

Πρέπει να γίνει λήψη και εγκατάσταση
(**`http://biopython.org/wiki/Download`**)

Πρέπει πρώτα να εγκαταστήσετε **`το numpy`**

```
>>>import bio
```

Co-funded by the Erasmus+ Programme of the European Union

Πρόγραμμα

- **Εισαγωγή στο Biopython**
  - Αντικείμενα ακολουθίας (I)
  - Ακολουθία Εγγραφή αντικειμένων (I)
  - Πρωτεϊνικές δομές (ενότητα PDB) (II)
- **Εργασία με αλληλουχίες DNA και πρωτεϊνών**
  - Μεταγραφή και Μετάφραση
- **Εξαγωγή πληροφοριών από βιολογικούς πόρους**
  - Ανάλυση αρχείων Swiss- Prot (I)
  - Ανάλυση εξόδου BLAST (I)
  - Πρόσβαση στις βάσεις δεδομένων Entrez του NCBI (II)
  - Ανάλυση εγγραφών Medline (II)
- **Εκτέλεση εξωτερικών εφαρμογών (π.χ. BLAST) τοπικά και από σενάριο**
  - Εκτέλεση BLAST μέσω Διαδικτύου
  - Εκτέλεση BLAST τοπικά
- **Εργασία με μοτίβα**
  - Ανάλυση εγγραφών PROSITE
  - Ανάλυση εγγραφών τεκμηρίωσης PROSITE

Εισαγωγή στο Biopython (I)

- Σειρά αντικειμένων
- Ακολουθία Εγγραφή αντικειμένων

Αντικείμενο ακολουθίας

- Seq αντικείμενα έναντι συμβολοσειρών Python:
  - Έχουν διαφορετικές μεθόδους
  - Το αντικείμενο Seq έχει το χαρακτηριστικό **`αλφάβητο`** (βιολογική σημασία του Seq)

```
>>> import bio
>>> from Bio.Seq import Seq
>>> my_seq = Seq("AGTACACTGGT")
>>> my_seq
Seq('AGTACACTGGT', Alphabet())
>>> print my_seq
Seq('AGTACACTGGT', Alphabet())
>>> my_seq.alphabet
Alphabet()
>>>
```

Co-funded by the Erasmus+ Programme of the European Union

Το χαρακτηριστικό

- Τα αλφάβητα ορίζονται στη μονάδα **`Bio.Alphabet`**
- Θα χρησιμοποιήσουμε τα αλφάβητα IUPAC (http://www.chem.qmw.ac.uk/iupac)
- **`Το Bio.Alphabet.IUPAC`** παρέχει ορισμούς για DNA, RNA και πρωτεΐνες + παρέχει επέκταση και προσαρμογή βασικών ορισμών:
  - **`IUPACProtein (Πρότυπο IUPAC AA)`**
  - **`ExtendedIUPACProtein (+ σεληνοκυστεΐνη, X, κ.λπ.)`**
  - **`IUPACUAmbiguousDNA (βασικά γράμματα GATC)`**
  - **`IUPACambiguousDNA (+ γράμματα ασάφειας)`**
  - **`ExtendedIUPACDNA (+ τροποποιημένες βάσεις)`**
  - **`IUPAOμοσύμφορο RNA`**
  - **`IUPACambiguousRNA`**

Το χαρακτηριστικό αλφάβητο

```
>>> import bio
>>> from Bio.Seq import Seq
>>> από την εισαγωγή IUPAC της Bio.Alphabet
>>> my_seq = Seq("AGTACACTGGT", IUPAC.unambiguous_dna)
>>> my_seq
Seq('AGTACACTGGT', IUPACUNAmbiguousDNA())
>>> my_seq.alphabet
IUPACambiguousDNA()
>>> my_seq = Seq("AGTACACTGGT", IUPAC.protein)
>>> my_seq
Seq('AGTACACTGGT', IUPACProtein())
>>> my_seq.alphabet
IUPACProtein ()
>>>
```

Οι ακολουθίες λειτουργούν σαν χορδές

```
>>> my_seq = Seq("AGTAACCCTTAGCACTGGT", IUPAC.unambiguous_dna )
>>> για ευρετήριο, γράμμα στο enumerate( my_seq ):
... ευρετήριο εκτύπωσης, γράμμα
...
0 A
1 Γ
2 T
3 A
4 A
5 Γ
... κλπ
>>> print len ( my_seq )
19
>>> print my_seq [0]
ENA
>>> print my_seq [2:10]
Seq('TAACCCTT', IUPACProtein ())
>>> my_seq.count ('A')
5
>>> 100*float( my_seq.count ('C')+ my_seq.count ('G'))/ len ( my_seq
47.368421052631582
```

Μετατρέψτε τα αντικείμενα Seq σε χορδές

Μπορεί να χρειαστείτε τη συμβολοσειρά απλής ακολουθίας (π.χ. για εγγραφή σε ένα αρχείο ή για εισαγωγή σε μια βάση δεδομένων)

```
>>> my_seq = Seq("AGTAACCCTTAGCACTGGT", IUPAC.unambiguous_dna)
>>>>>> οδός (my_seq)
'AGTAACCCTTAGCACTGGT'
>>> print my_seq
AGTAACCCTTAGCACTGGT
>>> fasta_format_string = ">DNA_id\n %s \n" % my_seq
>>> print fasta_format_string
>DNA_id
AGTAACCCTTAGCACTGGT
# Biopython 1.44 ή παλαιότερο
>>>my_seq.tostring()
'AGTAACCCTTAGCACTGGT'
```

Συνδεόμενες ακολουθίες

Δεν μπορείτε να προσθέσετε αλληλουχίες με ασύμβατα αλφάβητα (αλληλουχία πρωτεϊνών και αλληλουχία DNA)

```
>>> dna_seq = Seq("AGTAACCCTTAGCACTGGT", IUPAC.unambiguous_dna)
>>> protein_seq = Seq("KSMKPPRTHLIMHWIIL", IUPAC.IUPACProtein())
>>> protein_seq + dna_seq
Traceback (πιο πρόσφατη τελευταία κλήση):
Αρχείο "<stdin>", γραμμή 1, σε ?
Αρχείο "/home/abarbato/biopython-1.53/build/lib.linux-x86_64-
2.4/Bio/Seq.py", γραμμή 216, σε __add__
raise TypeError("Μη συμβατά αλφάβητα %s και %s" \
TypeError: Μη συμβατά αλφάβητα IUPACProtein() και
IUPACUNambiguousDNA()
```

ΑΛΛΑ, αν δώσετε γενικό αλφάβητο στο dna_seq και στο protein_seq :

```
>>> from Bio.Alphabet import generic_alphabet
>>> dna_seq.alphabet = γενικό_αλφάβητο
>>> protein_seq.alphabet = γενικό_αλφάβητο
>>> protein_seq + dna_seq
Seq('KSMKPPRTHLIMHWIILAGTAACCCTTAGCACTGGT', Alphabet())
```

Αλλαγή υπόθεσης

Τα αντικείμενα Seq έχουν μεθόδους **`upper()`** και **`low()`.**

```
>>> from Bio.Alphabet import generic_dna
>>> dna_seq = Seq("acgtACGT", generic_dna)
>>> dna_seq.upper()
Seq('ACGTACGT', DNAAlphabet())
>>> dna_seq.lower()
Seq('acgtacgt', DNAAlphabet())
>>>
```

Σημειώστε ότι τα αλφάβητα IUPAC είναι μόνο για κεφαλαία

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Αλληλουχίες νουκλεοτιδίων και (αντίστροφα) συμπληρώματα

Τα αντικείμενα Seq έχουν μεθόδους **`upper()`** και **`low()`.**

```
>>> from Bio.Seq import Seq
>>> από την εισαγωγή IUPAC της Bio.Alphabet
>>> dna_seq = Seq("AGTAACCCTTAGCACTGGT", IUPAC.unambiguous_dna)
>>> dna_seq. complement ()
Seq('TCATTGGGAATCGTGACCA', IUPACUAmbiguousDNA())
>>> dna_seq. reverse_complement ()
Seq('ACCAGTGCTAAGGGTTACT', IUPACUAmbiguousDNA())
```

Σημειώστε ότι αυτές οι επεμβάσεις δεν επιτρέπονται με πρωτε
αλφάβητα

Co-funded by the Erasmus+ Programme of the European Union

Μεταγραφή

DNA coding strand (aka Crick strand, strand +1)

5’ ATGGCCATTGTAATGGGCCGCTGAAAGGGTGCCCGATAG 3’

3’ TACCGGTAACATTACCCGGCGACTTTCCCACGGGCTATC 5’

DNA template strand (aka Watson strand, strand −1)

Transcription

5’ AUGGCCAUUGUAAUGGGCCGCUGAAAGGGUGCCCGAUAG 3’

Single stranded messenger RNA

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Μεταγραφή

```
>>> coding_dna =
Seq("ATGGCCATTGTAATGGGCCGCTGAAAGGGTGCCCGATAG",
IUPAC.unambiguous_dna )
>>> template_dna = coding_dna. reverse_complement ()
>>> template_dna
Seq('CTATCGGGCACCCTTTCAGCGGCCCATTACAATGGCCAT',
IUPACUAmbiguousDNA ())
>>> messenger_rna = coding_dna. μεταγραφή ()
>>> messenger_rna
Seq('AUGGCCAUUGUAAUGGGCCGCUGAAAGGGUGCCCGAUAG',
IUPACUAmbiguousRNA ())
```

Σημείωση: το μόνο που κάνει είναι ένας διακόπτης T --> U και ρυθμίστε το αλφάβητο.

Το αντικείμενο Seq περιλαμβάνει επίσης μια μέθοδο αναδρομικής μεταγραφής:

```
>>> messenger_rna. back_transscribe ()
Sequence('ATGGCCATTGTAATGGGCCGCTGAAAGGGTGCCCGATAG',
IUPACUAmbiguousDNA())
```

Co-funded by the Erasmus+ Programme of the European Union

Μετάφραση

```
>>> from Bio.Seq import Seq
>>> από την εισαγωγή IUPAC της Bio.Alphabet
>>> messenger_rna = Seq('AUGGCCAUUGUAAUGGGCCGCUGAAAGGGUGCCCGAUAG',
IUPAC.unambiguous_rna)
>>> messenger_rna. μετάφραση ()
Seq('MAIVMGR*KGAR*', HasStopCodon(IUPACProtein(), '*'))
>>>
```

Μπορείτε επίσης να μεταφράσετε απευθείας από την κωδικοποιητική αλληλουχία DNA

```
>>> from Bio.Seq import Seq
>>> από την εισαγωγή IUPAC
>>> coding_dna = Seq("ATGGCCATTGTAATGGGCCGCTGAAAGGGTGCCCGATAG",
IUPAC.unambiguous_dna)
>>> coding_dna. μετάφραση ()
Seq('MAIVMGR*KGAR*', HasStopCodon(IUPACProtein(), '*'))
>>>
```

Μετάφραση με διαφορετικούς πίνακες μετάφρασης

Οι πίνακες μετάφρασης που διατίθενται στο Biopython βασίζονται σε αυτούς του NCBI.

Από προεπιλογή, η μετάφραση θα χρησιμοποιεί τον τυπικό γενετικό κώδικα (αναγνωριστικό πίνακα NCBI 1)

Εάν ασχολείστε με μιτοχονδριακές αλληλουχίες:

```
>>> coding_dna.translate(table="Σπονδυλωτό Μιτοχόνδριο")
Seq('MAIVMGRWKGAR*', HasStopCodon(IUPACProtein(), '*'))
>>> coding_dna.translate(πίνακας=2)
Seq('MAIVMGRWKGAR*', HasStopCodon(IUPACProtein(), '*'))
```

Εάν θέλετε να μεταφράσετε τα νουκλεοτίδια μέχρι το πρώτο στοπ στο πλαίσιο και μετά να σταματήσετε (όπως συμβαίνει στη φύση):

```
>>> coding_dna.translate ( to_stop = True)
Seq('MAIVMGR', IUPACProtein ())
>>> coding_dna.translate ( πίνακας=2 , to_stop = True)
Seq('MAIVMGRWKGAR', IUPACProtein ())
```

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πίνακες μετάφρασης

Οι πίνακες μετάφρασης που διατίθενται στο Biopython βασίζονται σε αυτούς του NCBI.`http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi`

Από προεπιλογή, η μετάφραση θα χρησιμοποιεί τον τυπικό γενετικό κώδικα (αναγνωριστικό πίνακα NCBI 1)

```
>>> import from Bio.Data CodonTable
>>> standard_table =
CodonTable.unambiguous_dna_by_name["Standard"]
>>> mito_table =
CodonTable.unambiguous_dna_by_name["Σπονδυλωτό Μιτοχόνδριο"]
#Χρήση των αναγνωριστικών πινάκων NCB:
>>>standard_table = CodonTable.unambiguous_dna_by_id[1]
>>> mito_table = CodonTable.unambiguous_dna_by_id[2]
```

Co-funded by the Erasmus+ Programme of the European Union

Πίνακες μετάφρασης

```
>>>print standard_table
Πίνακας 1 Πρότυπο, SGC0

| T | C | A | G |
--+---------+---------+-----------------------
T | TTT F | TCT S | TAT Y | TGT C | T
T | TTC F | TCC S | TAC Y | TGC C | ντο
T | TTA L | TCA S | Στάση TAA| TGA Stop| ENA
T | TTG L(s)| TCG S | TAG Stop| TGG W | σολ
--+---------+---------+-----------------------
C | CTT L | CCT P | CAT H | CGT R | T
C | CTC L | CCC P | CAC H | CGC R | ντο
C | CTA L | CCA P | CAA Q | CGA R | ENA
C | CTG L(s)| CCG P | CAG Q | CGG R | σολ
--+---------+---------+-----------------------
A | ATT I | ΠΡΑΞΗ T | AAT N | AGT S | T
A | ATC I | ACC T | AAC N | AGC S | ντο
A | ATA I | ACA T | AAA K | AGA R | ENA
A | ATG M(s)| ACG T | ΑΑΓ K | AGG R | σολ
--+---------+---------+-----------------------
G | GTT V | GCT A | GAT D | GGT G | T
G | GTC V | GCC A | GAC D | GGC G | ντο
G | GTA V | GCA A | GAA E | ΓΤΑ Γ | ENA
G | GTG V | GCG A | GAG E | GGG G | σολ
--+---------+---------+-----------------------
```

Co-funded by the Erasmus+ Programme of the European Union

Πίνακες μετάφρασης

```
>>> print mito_table
Πίνακας 2 Μιτοχόνδριο Σπονδυλωτών, SGC1

| T | C | A | G |
--+---------+---------+----------------------
T | TTT F | TCT S | TAT Y | TGT C | T
T | TTC F | TCC S | TAC Y | TGC C | ντο
T | TTA L | TCA S | Στάση TAA| TGA W | ENA
T | TTG L | TCG S | TAG Stop| TGG W | σολ
--+---------+---------+----------------------
C | CTT L | CCT P | CAT H | CGT R | T
C | CTC L | CCC P | CAC H | CGC R | ντο
C | CTA L | CCA P | CAA Q | CGA R | ENA
C | CTG L | CCG P | CAG Q | CGG R | σολ
--+---------+---------+----------------------
A | ATT I(s)| ΠΡΑΞΗ T | AAT N | AGT S | T
A | ATC I(ες)| ACC T | AAC N | AGC S | ντο
A | ATA M(s)| ACA T | AAA K | AGA Stop| ENA
A | ATG M(s)| ACG T | ΑΑΓ K | AGG Stop| σολ
--+---------+---------+----------------------
G | GTT V | GCT A | GAT D | GGT G | T
G | GTC V | GCC A | GAC D | GGC G | ντο
G | GTA V | GCA A | GAA E | ΓΤΑ Γ | ENA
G | GTG V(s)| GCG A | GAG E | GGG G | σολ
--+---------+---------+----------------------
```

Αντικείμενα

Όπως οι συμβολοσειρές Python, τα αντικείμενα Seq είναι αμετάβλητα

```
>>> from Bio.Seq import Seq
>>> από την εισαγωγή IUPAC
>>> my_seq =
Seq('CGCGCGGGTTTATGATGACCCAAATATAGAGGGCACAC',
IUPAC.unambiguous_dna)
>>> my_seq[5] = 'A'
Traceback (πιο πρόσφατη τελευταία κλήση):
Αρχείο "<stdin>", γραμμή 1, σε ?
TypeError: το αντικείμενο δεν υποστηρίζει την
εκχώρηση στοιχείων
>>>
```

Ωστόσο, μπορείτε να το μετατρέψετε σε μια μεταβλητή ακολουθία (ένα αντικείμενο **MutableSeq**)

```
>>> mutable_seq = my_seq.tomutable()
>>> mutable_seq
MutableSeq('CGCGCGGGTTTATGATGACCCAAATATAGAGGGCACAC',
IUPACUAmbiguousDNA())
```

Αντικείμενα MutableSeq

Μπορείτε να δημιουργήσετε ένα μεταβλητό αντικείμενο απευθείας

```
>>> από το Bio.Seq εισαγωγής Seq
>>> από την εισαγωγή IUPAC
>>> mutable_seq = MutableSeq ('CGCGC G
GGTTTATGATGACCCAAATATAGAGGGCACAC',
IUPAC.unambiguous_dna)
>>> mutable_seq[5] = 'A'
>>> mutable_seq
MutableSeq('CGCGC A
GGTTTATGATGACCCAAATATAGAGGGCACAC',
 IUPACUAmbiguousDNA())
```

Ένα αντικείμενο **MutableSeq** μπορεί εύκολα να μετατραπεί σε ακολουθία μόνο για ανάγνωση:

```
>>> new_seq = mutable_seq.toseq()
>>> new_seq
Seq('CGCGCAGGTTTATGATGACCCAAATATAGAGGGCACAC',
IUPACUAσαφής DNA())
```

Co-funded by the Erasmus+ Programme of the European Union

Ακολουθία Εγγραφή αντικειμένων

Η κλάση **`SeqRecord`** ορίζεται στη μονάδα **`Bio.SeqRecord`**

Αυτή η κλάση επιτρέπει σε χαρακτηριστικά υψηλότερου επιπέδου, όπως αναγνωριστικά και χαρακτηριστικά, να συσχετιστούν με μια ακολουθία

```
>>> from Bio.SeqRecord import SeqRecord
>>> help (SeqRecord)
```

class SeqRecord (__ ενσωματωμένο __.object)
Ένα αντικείμενο SeqRecord περιέχει μια ακολουθία και πληροφορίες σχετικά με αυτό.

Κύρια χαρακτηριστικά:

id - Αναγνωριστικό, όπως μια ετικέτα τοποθεσίας (συμβολοσειρά)
seq - Η ίδια η ακολουθία (αντικείμενο Seq ή παρόμοιο)

Πρόσθετα χαρακτηριστικά:

όνομα - Όνομα ακολουθίας, π.χ. όνομα γονιδίου (συμβολοσειρά)
περιγραφή - Πρόσθετο κείμενο (συμβολοσειρά)
dbxrefs - Λίστα διασταυρούμενων αναφορών db (λίστα συμβολοσειρών)
χαρακτηριστικά - Οποιεσδήποτε (υπο)καθορισμένες δυνατότητες (λίστα αντικειμένων SeqFeature)
σχολιασμοί - Περισσότερες πληροφορίες για ολόκληρη τη σειρά (λεξικό)

Οι περισσότερες καταχωρήσεις είναι συμβολοσειρές ή λίστες συμβολοσειρών.

επιστολές_σχολιασμοί –
Σχολιασμός ανά γράμμα/σύμβολο (περιορισμένο λεξικό). Αυτό κρατά ακολουθίες Python (λίστες, συμβολοσειρές ή πλειάδες) των οποίων το μήκος
ταιριάζει με αυτό της σειράς. Μια τυπική χρήση θα ήταν να κρατάτε ένα
λίστα ακεραίων αριθμών που αντιπροσωπεύουν βαθμολογίες ποιότητας ακολουθίας, ή α
συμβολοσειρά που αντιπροσωπεύει τη δευτερεύουσα δομή.

Συνήθως θα χρησιμοποιείτε **το Bio.SeqIO** για να διαβάζετε σε ακολουθίες από αρχεία ως Αντικείμενα **SeqRecord.** Ωστόσο, μπορεί να θέλετε να δημιουργήσετε το δικό σας **SeqRecord** αντικείμενα άμεσα:

```
>>> from Bio.Seq import Seq
>>> from Bio.SeqRecord import SeqRecord
>>> TMP = Seq('MKQHKAMIVALIVICITAVVAALVTRKDLCEVHIRTGQTEVAVF')
>>> TMP_r = SeqRecord (TMP)
>>> TMP_r.id
"<άγνωστο αναγνωριστικό>"
>>> TMP_r.id = 'YP_025292.1'
>>> TMP_r.description = 'τοξική πρωτεΐνη μεμβράνης'
>>> εκτύπωση TMP_r
ID: YP_025292.1
Όνομα: <άγνωστο όνομα>
Περιγραφή: τοξική πρωτεΐνη μεμβράνης
Αριθμός χαρακτηριστικών: 0
Seq('MKQHKAMIVALIVICITAVVAALVTRKDLCEVHIRTGQTEVAVF', Alphabet())
>>> TMP_r.seq
Seq('MKQHKAMIVALIVICITAVVAALVTRKDLCEVHIRTGQTEVAVF', Alphabet())
```

Μπορείτε επίσης να δημιουργήσετε τα δικά σας αντικείμενα **SeqRecord** ως εξής:

```
>>> from Bio.Seq import Seq
>>> from Bio.SeqRecord import SeqRecord
>>> import IUPAC
>>> record
SeqRecord(seq=Seq('MKQHKAMIVALIVICITAVVAALVTRKDLCEVHIRTGQ
TEVAVF', IUPACProtein()), id='YP_025292.1', name='HokC',
περιγραφή='τοξική πρωτεΐνη μεμβράνης', dbxrefs=[]
>>> print record
ID: YP_025292.1
Όνομα: HokC
Περιγραφή: τοξική πρωτεΐνη μεμβράνης
Αριθμός χαρακτηριστικών: 0
Seq('MKQHKAMIVALIVICITAVVAALVTRKDLCEVHIRTGQTEVAVF',
IUPACProtein())
>>>
```

Η μέθοδος `format()`.

Επιστρέφει μια συμβολοσειρά που περιέχει το καλώδιο σας μορφοποιημένο χρησιμοποιώντας μία από τις μορφές αρχείων εξόδου που υποστηρίζονται από το **`Bio.SeqIO`**

```
>>> from Bio.Seq import Seq
>>> from Bio.SeqRecord import SeqRecord
>>> from Bio.Alphabet import generic_protein
>>> rec =
SeqRecord(Seq("MGSNKSKPKDASQRRRSLEPSENVHGAGGAFPASQTPSKPASADGHRGPSA
AFVPPAAEPKLFGGFNSSDTVTSPQRAGALAGGVTTFVALYDYESRTETDLSFKGERLQIVNTRKW
TGWLAIDSLTE),inTRKWDVDVREGTE,inTRKWDVDVREGTE = "P05480", περιγραφή
= "SRC_MOUSE Νευρωνικό πρωτο-ογκογονίδιο κινάσης τυροσίνης-
πρωτεΐνης Src: MY TEST")
>>> print rec. μορφή ("fasta")
>P05480 SRC_MOUSE Νευρωνικό πρωτο-ογκογονίδιο τυροσινο-πρωτεϊνική
κινάση Src: Η ΔΟΚΙΜΗ ΜΟΥ
MGSNKSKPKDASQRRRSLEPSENVHGAGGAFPASQTPSKPASADGHRGPSAAFVPPAAEP
KLFGGFNSSDTVTSPQRAGALAGGVTTFVALYDYESRTETDLSFKKGERLQIVNTRKVD
VREGDWWLAHSLSTGQTGYIPS
```

Co-funded by the Erasmus+ Programme of the European Union

```
Ακολουθία 1 "ACTGGGAGCTAGC"
Seq2 "TTGATCGATCGATCG"
Seq3 "GTGTAGCTGCT"
```

ΑΡΧΕΙΟ ΕΙΣΑΓΩΓΗΣ

```
F = open ("input.txt")
For line in F:
<ανάλυση γραμμής>
<get seq id>
<λήψη περιγραφής>
<get sequence>
<λάβετε άλλες πληροφορίες>
from Bio.Seq import Seq
from Bio.SeqRecord import SeqRecord
from Bio.Alphabet import generic_protein
rec = SeqRecord(Seq(<ακολουθία>, αλφάβητο),
id = <seq_id>, περιγραφή = <description>)
Format_rec = rec.format("fasta")
Out.write(Format_rec)
```

SCRIPT.py

ΑΡΧΕΙΟ ΕΞΟΔΟΥ

```
>P05480 SRC_MOUSE Νευρωνικό πρωτο-ογκογονίδιο τυροσινο-πρωτεϊνική
κινάση Src: Η ΔΟΚΙΜΗ ΜΟΥ
MGSNKSKPKDASQRRRSLEPSENVHGAGGAFPASQTPSKPASADGHRGPSAAFVPPAAEP
KLFGGFNSSDTVTSPQRAGALAGGVTTFVALYDYESRTETDLSFKKGERLQIVNTRKVD
```

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πρωτεϊνικές Δομές

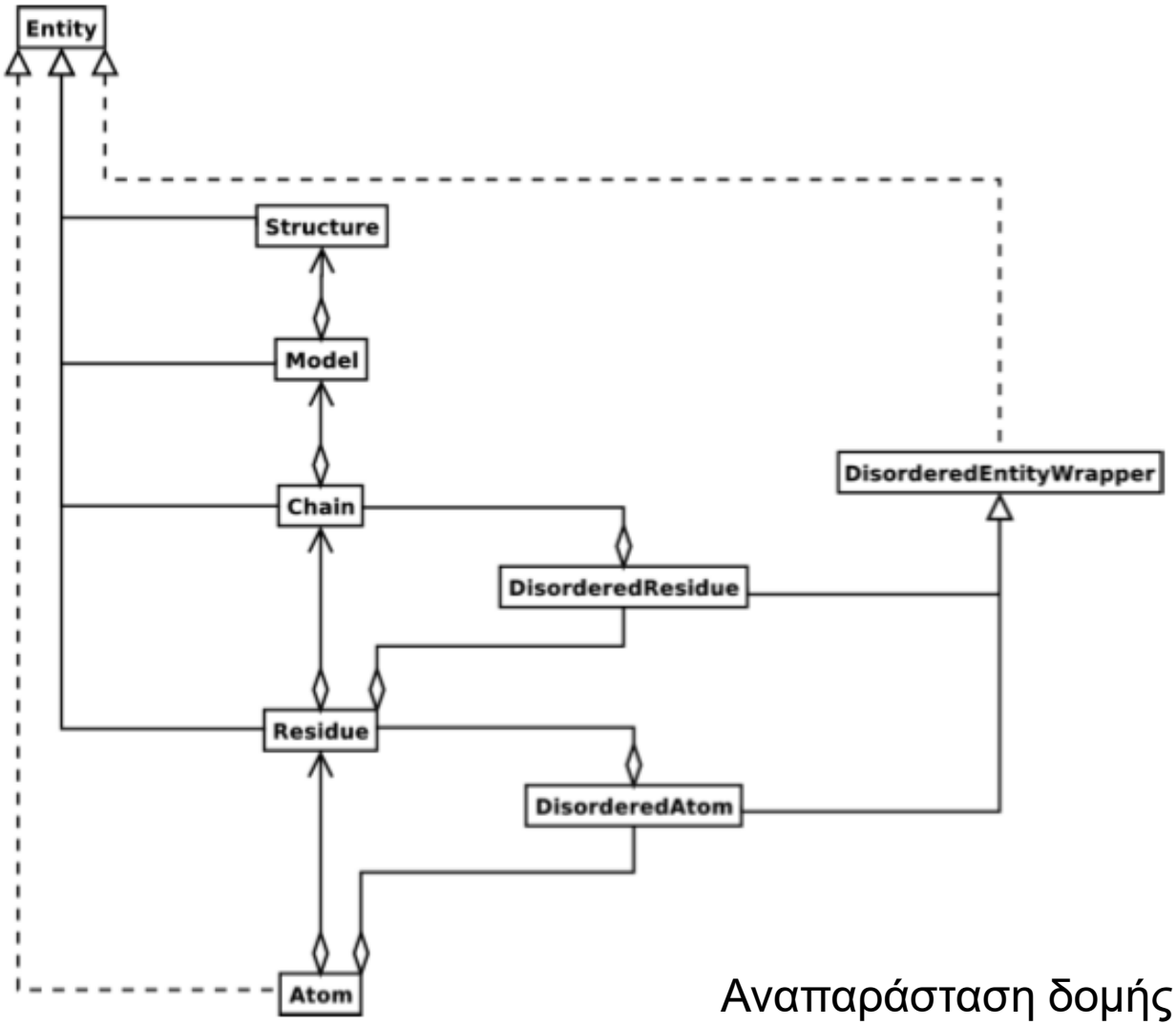
ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Πρωτεϊνικές δομές (μονάδα PDB)

S δομή
M μοντέλο
C hain
R υπόλειμμα
A tom



Αναπαράσταση δομής

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

```
ATOM 1 N PRO A 134 33.747 46.582 69.532 1.00 56.70 N
ATOM 2 CA PRO A 134 34.692 45.947 70.456 1.00 55.76 C
ATOM 3 C PRO A 134 35.040 44.558 69.953 1.00 55.49 C
ATOM 4 O PRO A 134 34.894 43.546 70.722 1.00 55.72 O
ATOM 5 CB PRO A 134 33.839 45.768 71.707 1.00 56.31 C
ATOM 6 CG PRO A 134 32.384 45.405 71.086 1.00 56.74 C
ATOM 7 CD PRO A 134 32.401 45.963 69.637 1.00 56.69 C
ATOM 8 N ALA A 135 35.457 44.461 68.683 1.00 53.77 N
ATOM 9 CA ALA A 135 35.892 43.141 68.150 1.00 52.13 C
ATOM 10 C ALA A 135 37.099 42.612 68.967 1.00 50.42 C
ATOM 11 O AΛA A 135 37.255 41.387 69.178 1.00 51.42 O
ATOM 12 CB ALA A 135 36.196 43.216 66.650 1.00 51.84 C
ATOM 13 N SER A 136 37.939 43.538 69.439 1.00 48.55 N
ATOM 14 CA SER A 136 38.987 43.178 70.385 1.00 46.48 C
ATOM 15 C SER A 136 38.789 43.888 71.746 1.00 45.81 C
ATOM 16 O SER A 136 38.568 45.056 71.764 1.00 44.23 O
ATOM 17 CB SER A 136 40.330 43.486 69.763 1.00 46.19 C
ATOM 18 OG SER A 136 41.344 43.236 70.671 1.00 42.35 O
...
ATOM 1885 N ALA B 135 38.255 -6.602 103.702 1.00 44.10 N
ATOM 1886 CA ALA B 135 37.061 -6.004 102.994 1.00 44.16 C
ATOM 1887 C ALA B 135 37.338 -5.794 101.533 1.00 44.32 C
ATOM 1888 O ALA B 135 36.433 -5.890 100.696 1.00 45.88 O
ATOM 1889 CB ALA B 135 36.673 -4.681 103.600 1.00 43.74 C
ATOM 1890 N SER B 136 38.566 -5.418 101.235 1.00 43.63 N
ATOM 1891 CA SER B 136 38.999 -5.265 99.848 1.00 42.10 C
ATOM 1892 C SER B 136 40.151 -6.245 99.680 1.00 42.68 C
ATOM 1893 O SER B 136 41.069 -6.309 100.494 1.00 43.55 O
ATOM 1894 CB SER B 136 39.485 -3.837 99.612 1.00 41.77 C
```

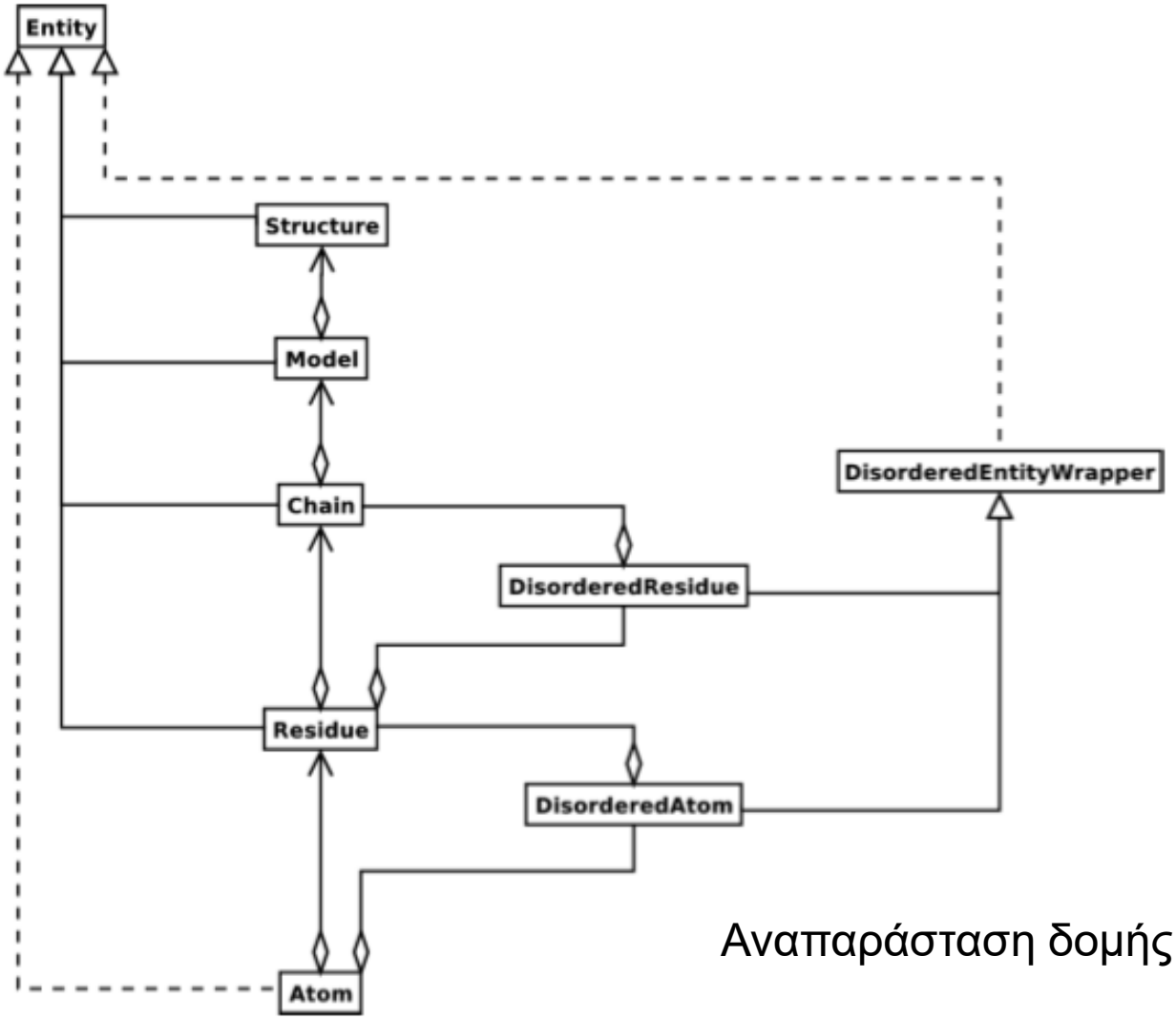
ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή στο Biopython (II)

Πρωτεϊνικές δομές (μονάδα PDB)

S δομή
M μοντέλο
C hain
R υπόλειμμα
A tom



Αναπαράσταση δομής

Πρωτεϊνικές δομές (μονάδα PDB)

Αυτή η δομή δεδομένων είναι απαραίτητη για μια καλή ερμηνεία των δεδομένων που υπάρχουν σε ένα αρχείο που περιγράφει τη δομή (PDB ή MMCIF)

Εάν η ιεραρχία δεν μπορεί να αναπαραστήσει το περιεχόμενο ενός αρχείου δομής, είναι αρκετά βέβαιο ότι το αρχείο περιέχει ένα σφάλμα ή τουλάχιστον δεν περιγράφει τη δομή με σαφήνεια.

Εάν δεν μπορεί να δημιουργηθεί ένα SMCRA, υπάρχει λόγος να υποπτευόμαστε ότι υπάρχει πρόβλημα

```
child_entity = parent_entity[child_id]
```

Μοντέλο
Αλυσίδα
Υπόλειμμα
Ατομο

Δομή
Μοντέλο
Αλυσίδα
Υπόλειμμα

Μοντέλο = 0
Αλυσίδα = A
Residue = 'SER'
Atom = 'CA'

```
child_list = parent_entity.get_list()
```

Δομή
Μοντέλο
Αλυσίδα
Υπόλειμμα

Μοντέλο
Αλυσίδα
Υπόλειμμα
Ατομο

```
parent_entity = child_entity.get_parent()
```

Co-funded by the Erasmus+ Programme of the European Union

```
full_id = residue.get_full_id()
print full_id
("1H8A", 0, "A", ("", 10, "A")
```

Η δομή με αναγνωριστικό "1H8A"
Το μοντέλο με id 0
Η αλυσίδα με id "A"
The Residue with id (“ ”, 10, “A”)

```
#Λάβετε το αναγνωριστικό της οντότητας
entity.get_id()
#Ελέγξτε αν υπάρχει παιδί με
δεδομένο αναγνωριστικό
entity.has_id( entity_id )
#Λάβετε αριθμό παιδιών
nr_chikdren = len(entity)
```

Αντικείμενο δομής

Βρίσκεται στην κορυφή της ιεραρχίας. Το αναγνωριστικό του είναι μια συμβολοσειρά που δίνεται από τον χρήστη. Περιέχει έναν αριθμό παιδιών Μοντέλων .

Αντικείμενο μοντέλου

Το id του είναι ένας ακέραιος, ο οποίος προέρχεται από τη θέση του μοντέλου στο αναλυμένο αρχείο. Αποθηκεύει μια λίστα με τα παιδιά της αλυσίδας .

Αντικείμενο αλυσίδας

Το αναγνωριστικό του είναι το αναγνωριστικό αλυσίδας στο αρχείο δομής και μπορεί να είναι οποιαδήποτε συμβολοσειρά. Κάθε αλυσίδα σε ένα μοντέλο έχει ένα μοναδικό αναγνωριστικό. Αποθηκεύει μια λίστα με τα παιδιά υπολειμμάτων .

Αντικείμενο υπολειμμάτων

Περιέχει μια συμβολοσειρά με όνομα υπολειμμάτων (“ALA”). Το αναγνωριστικό του αποτελείται από 3 μέρη: `(ετεροπεδίο, αναγνωριστικό ακολουθίας, κωδικός εισαγωγής)`. Αποθηκεύει ένα σύνολο παιδιών Atom .

Αντικείμενο ατόμου

Αποθηκεύει τα δεδομένα που σχετίζονται με ένα άτομο και δεν έχει παιδιά . Το αναγνωριστικό του είναι το όνομα του ατόμου ("CA"). Τα αντικείμενά του μπορούν να χρησιμοποιηθούν για την εξαγωγή συγκεκριμένων πληροφοριών.

Co-funded by the Erasmus+ Programme of the European Union

```
From Bio.PDB.PDBParser import PDBParser
p = PDBParser(ALLOW = 1)
structure_id = '2H8L'
filename = '2H8L.pdb'
s = p.get_structure(structure_id, filename)
```

```
#print p.get_header()
#print p.get_trailer()
first_model = s[0]
Print type(first_model)
```

```
chain_A = first_model["A"]
print chain_A
```

```
#print chain_A.child_dict.keys()
για k στο chain_A.child_dict.keys():
print chain_A[k], chain_A[k].get_id(), chain_A[k].get_resname()
chain print_A[k].get_list()
#print chain_A.get_list()
```

```
res266 = chain_A[266]
#print dir(res266)
εκτύπωση res266.get_id()
εκτύπωση res266.get_resname()
εκτύπωση res266.get_segid()
for atom in res266.get_list():
  print atom.get_name(),atom.get_id(),atom.get_coord(),atom.get_bfactor()
```

```
From Bio.PDB.PDBParser import PDBParser
from math import sqrt

Απόσταση αποκορύφωσης (συνορμές1, κοόρντες2):
dx = (συντεταγμένες1[0] - συντεταγμένες2[0])**2
        dy = (συντεταγμένες1[1] - συντεταγμένες2[1])**2
        dz = (συντεταγμένες1[2] - συντεταγμένες2[2])**2
RMSD = sqrt(dx + dy + dz )
επιστροφή RMSD

p = PDBParser (ΕΠΙΤΡΕΠΟ = 1)
structure_id = '2H8L'
όνομα αρχείου = '2H8L.pdb'
s = p.get_structure ( structure_id , όνομα αρχείου)
first_model = s[0]
chain_A = first_model ["A"]
chain_B = first_model ["B"]
για το resA στην αλυσίδα_A. get_list () :
εάν resA. has_id ( "CA" ) :
                caA = resA ["CA"]
                coordsA = caA. get_coord ()
#print caA , coordsA
για resB στην αλυσίδα_B. get_list () :
if resB.has_id ("CA"):
                                caB = resB ["CA"]
                                coordsB = caB. get_coord ()
D = απόσταση ( συντεταγμένεςA,συντεταγμένεςB )
αν D < 6: εκτύπωση resA.resname , resA.get_id ()[1],\
          resB.resname , resB.get_id ()[1], D
```

Εξαγωγή πληροφοριών από βιολογικούς πόρους: ανάλυση αρχείων Swiss-Prot, αρχείων PDB, εγγραφών ENSEMBLE, αρχείων εξόδου έκρηξης κ.λπ.

- Ακολουθία I/O
  - Ανάλυση ή ανάγνωση ακολουθιών
  - Σύνταξη αρχείων ακολουθίας

`Bio.SeqIO`

Μια απλή διεπαφή για εργασία με διάφορες μορφές αρχείων με ομοιόμορφο τρόπο

```
>>>from Bio import SeqIO
>>>help (SeqIO)
```

`Bio.SeqIO.parse()`

Διαβάζει σε σειρά δεδομένα ως αντικείμενα SeqRecord.
Αναμένει δύο επιχειρήματα.

- Μια *λαβή* για την ανάγνωση της φόρμας δεδομένων. Μπορεί να είναι:
  - ένα αρχείο που άνοιξε για ανάγνωση
  - την έξοδο από ένα πρόγραμμα γραμμής εντολών
  - δεδομένα που έχουν ληφθεί από το διαδίκτυο

- Μια συμβολοσειρά με πεζά γράμματα που καθορίζει τη μορφή ακολουθίας (δείτε http://biopython.org/wiki/SeqIO για μια πλήρη λίστα των υποστηριζόμενων μορφών).

Το αντικείμενο που επιστρέφεται από `το Bio.SeqIO` είναι ένας επαναλήπτης που επιστρέφει αντικείμενα `SeqRecord`

Co-funded by the Erasmus+ Programme of the European Union

```
>>> from Bio import SeqIO
>>> handle = open ("P05480.fasta")
>>> for seq_rec in SeqIO.parse (handle, "fasta"):
... print seq_rec.id
... print repr(seq_rec.seq)
... print len (seq_rec)
...
sp|P05480|SRC_MOUSE
Seq('MGSNKSKPKDASQRRRSLERGPSA...ENL', SingleLetterAlphabet())
541
>>> handle.close()
```

```
>>> for seq_rec in SeqIO.parse (handle, "genbank"):
... print seq_rec.id
... print repr(seq_rec.seq)
... Print len(seq_rec)
...
U49845.1
Seq('GATCCTCCATATACAACGGTACGGAA...ATC', IUPACA διφορούμενο DNA()
5028
>>> handle.close()
```

Γονιδιωματικό DNA Candida albicans, χρωμόσωμα 7, πλήρης αλληλουχία

```
>>> from Bio import SeqIO
>>> handle = open ("AP006852.gbk")
>>> για seq_rec στο SeqIO.parse(handle, "genbank"):
... εκτύπωση seq_rec.id
... εκτύπωση repr(seq_rec.seq)
... φακός εκτύπωσης (seq_rec)
...
AP006852.1
Seq('CCACTGTCCAATACCCCAACAGGAAT...TGT', IUPACA διφορούμενο DNA())
949626
>>>
```

Χρήση της κατανόησης λίστας:

```
>>>handle = open("AP006852.gbk")
>>>identifiers=[seq_rec.id για seq_rec στο SeqIO.parse(handle"genbank")]
>>>handle.close()
>>>αναγνωριστικά
['AP006852.1']
>>>
```

Εδώ το κάνουμε χρησιμοποιώντας το αρχείο **sprot_prot**

```
>>> from Bio import SeqIO
>>> handle = open("sprot_prot.fasta")
>>> ids = [seq_rec.id για seq_rec στο SeqIO.parse(handle,"fasta")]
>>> αναγνωριστικά
['sp|P24928|RPB1_HUMAN', 'sp|Q9NVU0|RPC5_HUMAN',
'sp|Q9BUI4|RPC3_HUMAN', 'sp|Q9BUI4|RPC3_HUMAN',
'sp|Q9NW08|RPC2_HUMAN', 'sp|Q9H1D9|RPC6_HUMAN',
'sp|P19387|RPB3_HUMAN', 'sp|O11CspHUMAN', 'sp|O11CspHUMAN'|
|P52435|RPB11_HUMAN', 'sp|O15318|RPC7_HUMAN',
'sp|P62487|RPB7_HUMAN', 'sp|O15514|RPB4_HUMAN',
'sp|Q9GZS1|RPA49_HUMAN', 'sp'|P39_SP' |RPC8_HUMAN',
'sp|O95602|RPA1_HUMAN', 'sp|Q9Y2Y1|RPC10_HUMAN',
'sp|Q9H9Y6|RPA2_HUMAN', 'sp|P78527|PRKDC_HUMAN',
'sp|O15160|RPAC1_HUMAN', 'sp|O15160|RPAC1_HUMAN', |RPAP1_HUMAN']
>>> ]
```

Επανάληψη πάνω από τις εγγραφές σε ένα αρχείο ακολ

Αντί να χρησιμοποιήσετε έναν βρόχο for, μπορείτε επίσης να χρησιμοποιήσετε τη μέθοδο **`next()`** ενός επαναλήπτη για να κάνετε βήμα μέσα από τις εγγραφές

```
>>> handle = open("sprot_prot.fasta")
>>> rec_iter = SeqIO.parse(handle, "fasta")
>>> rec_1 = rec_iter. Επόμενο()
>>> rec_1
SeqRecord(seq=Αλληλ. _Ανθρώπινο DNA κατευθυνόμενο Υπομονάδα RNA
πολυμεράσης II RPB1 OS=Homo sapiens GN=POLR2A PE=1 SV=2',
dbxrefs=[])
>>> rec_2 = rec_iter. Επόμενο()
>>> rec_2
SeqRecord(seq=Seq('MANEEDDPVVQEIDVYLAKSLAEKLYLFQYPVRPASMTYDDIPHLS
AKIKPKQQ...VQS', SingleLetterAlphabet()),
id='sp|Q9NVU0|RPC5_HUMAN', name='sp|Q9NVCUP' _Ανθρώπινο DNA
κατευθυνόμενο Υπομονάδα RNA πολυμεράσης III RPC5 OS=Homo sapiens
GN=POLR3E PE=1 SV=1', dbxrefs=[])
>>> handle.close()
```

Bio.SeqIO.read()

Co-funded by the Erasmus+ Programme of the European Union

Εάν το αρχείο σας έχει μία και μόνο μία εγγραφή (π.χ. ένα αρχείο GenBank για ένα μεμονωμένο χρωμόσωμα), τότε χρησιμοποιήστε το **`Bio.SeqIO.read()`** .

Αυτό θα ελέγξει ότι δεν υπάρχουν επιπλέον μη αναμενόμενες εγγραφές

```
>>> rec_iter = SeqIO.parse (open("1293613.gbk"), "genbank")
>>> rec = rec_iter. Επόμενο()
>>> εκτύπωση rec
Αναγνωριστικό: U49845.1
Όνομα: SCU49845
Περιγραφή: Γονίδιο TCP1-βήτα Saccharomyces cerevisiae, μερικά cds και γονίδια Axl2p (AXL2) και Rev7p
(REV7), πλήρη cds.
Αριθμός χαρακτηριστικών: 6
/sequence_version=1
/source=Saccharomyces cerevisiae (μαγιά αρτοποιίας)
/taxonomy=['Eukaryota', 'Mungi', 'Ascomycota', 'Saccharomycotina', 'Saccharomycetes',
'Saccharomycetales', 'Saccharomycetaceae', 'Saccharomyces']
/keywords=['']
/references=[Αναφορά(title='Κλωνοποίηση και αλληλουχία του REV7, ενός γονιδίου του οποίου η λειτουργία
απαιτείται για μεταλλαξογένεση που προκαλείται από βλάβη στο DNA στο Saccharomyces cerevisiae', ...),
Αναφορά(title='Η επιλογή αξονικών θέσεων ανάπτυξης σε ζυμομύκητες απαιτεί Axl2p, μια νέα γλυκοπρωτεΐνη
μεμβράνης πλάσματος', ...), Αναφορά(τίτλος='Άμεση υποβολή', ...)]
/accessions=['U49845']
/data_file_division=PLN
/ημερομηνία=21-IOYN-1999
/organism=Saccharomyces cerevisiae
/gi=1293613
Seq('GATCCTCCATATACAACGGTATCTCCACCTCAGGTTTAGATCTCAACAACGGAA...ATC', IUPACA διφορούμενοDNA())
```

Ακολουθήστε τα αρχεία ως λίστες

```
>>> from Bio import SeqIO
>>> handle = open("ncbi_gene.fasta")
>>> records = list (SeqIO. ανάλυση (handle, "fasta"))
>>> >>> records[-1]
SeqRecord(seq=Seq('gggggggggggggatcactctctttcagtaacctcaac...ccc',
SingleLetterAlphabet()), id='A10421', name='A10421', περιγραφή='A10421 Συνθετική
ακολουθία νουκλεοτιδίων που λαμβάνεται από ανθρώπινη αλληλουχία p-281 από το ανθρώπινο
γονίδιο p5IL-IL . : Τοποθεσία: 1..1000', dbxrefs=[])
```

Ακολουθήστε τα αρχεία ως λεξικά

```
>>> handle = open("ncbi_gene.fasta")
>>> records = SeqIO. to_dict (SeqIO.parse(handle, "fasta"))
>>> handle.close()
>>> records.keys()
['M69013', 'M69012', 'AJ580952', 'J03005', 'J03004', 'L13858', 'L04510', 'M94539',
'M19650', 'A10420', 'A10420', '2906, 'AJ' A06662', 'S62035', 'M57424', 'M90035',
'A06280', 'X95521', 'X95520', 'M28269', 'S50017', 'L13857', 'AJ23', 'AJ3803', 'AJ3801,' ,
'AB020593', 'M17219', 'DQ854814', 'M27543', 'X62025', 'M90043', 'L22075', 'X56614',
'M90027']
>>> seq_record = εγγραφές['X95521']
'X95521 M.musculus mRNA για κυκλική νουκλεοτιδική φωσφοδιεστεράση : Τοποθεσία:1..1000'
```

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ανάλυση ακολουθιών από το δίχτυ

Οι λαβές δεν προέρχονται πάντα από αρχεία

Ανάλυση αρχείων GenBank από το δίχτυ

```
>>>from Bio import Entrez
>>>from Bio import SeqIO
>>>handle = Entrez.efetch (db="nucleotide",rettype="fasta",id="6273291")
>>>seq_record = SeqIO.read(handle,”fasta”)
>>>handle.close()
>>>seq_record.description
```

Ανάλυση της ακολουθίας SwissProt από το δίκτυο

```
>>>from Bio import ExpASy
>>>from Bio import SeqIO
>>>handle = ExPASy.get_sprot_raw ("6273291")
>>>seq_record = SeqIO.read(handle,"swiss")
>>>handle.close()
>>>print seq_record.id
>>>print seq_record.name
>>>print seq_record.description
```

Ευρετηρίαση πραγματικά μεγάλων αρχείων

`Το Bio.SeqIO.index()` επιστρέφει ένα λεξικό χωρίς να διατηρεί τα πάντα στη μνήμη.

Λειτουργεί καλά ακόμα και για εκατομμύρια ακολουθίες

Το κύριο μειονέκτημα είναι η μικρότερη ευελιξία: είναι μόνο για ανάγνωση

```
>>> from Bio import SeqIO
>>> recs_dict = SeqIO.index ("ncbi_gene.fasta", "fasta")
>>> len(recs_dict)
34`
>>> recs_dict.keys()
['M69013', 'M69012', 'AJ580952', 'J03005', 'J03004', 'L13858', 'L04510', 'M94539',
'M19650', 'A10420', 'A10420', '2906, 'AJ' A06662', 'S62035', 'M57424', 'M90035', 'A06280',
'X95521', 'X95520', 'M28269', 'S50017', 'L13857', 'AJ23', 'AJ3803', 'AJ3801,' ,
'AB020593', 'M17219', 'DQ854814', 'M27543', 'X62025', 'M90043', 'L22075', 'X56614',
'M90027']
>>> print recs_dict['M57424']
ID: M57424
Όνομα: M57424
Περιγραφή: M57424 Γονίδιο μετατοπιστή-2 (ANT-2) νουκλεοτιδίου ανθρώπινης αδενίνης, πλήρη
cds. : Τοποθεσία: 1..1000
Αριθμός χαρακτηριστικών: 0
Seq('gagctctggaatagaatacagtagaggcatgctcaaagagagtagcagatg...agc', SingleLetterAlphabet())
```

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Σύνταξη αρχείων ακολουθίας

Bio.SeqIO.write()

Αυτή η συνάρτηση παίρνει τρία ορίσματα:
1. μερικά αντικείμενα SeqRecord
2. μια λαβή για να γράψετε
3. μια μορφή ακολουθίας

```
from Bio.Seq import Seq
from Bio.SeqRecors import SeqRecord
from Bio.Alphabet import generic_protein
Rec1 = SqRecord(Seq("ACCA…",generic_protein), id="1", περιγραφή="")
Rec1 = SqRecord(Seq("CDRFAA",generic_protein), id="2", description="")
Rec1 = SqRecord(Seq("GRKLM",generic_protein), id="3", περιγραφή="")
My_records = [Rec1, Rec2, Rec3]
from Bio import SeqIO
handle = open ("MySeqs.fas","w")
SeqIO.write(My_records, handle, "fasta")
handle.close()
```

Μετατροπή μεταξύ μορφών αρχείων ακολουθίας

Μπορούμε να κάνουμε μετατροπή αρχείων συνδυάζοντας **`Bio.SeqIO.parse()`** και **`Bio.SeqIO.write()`**

```
from Bio import SeqIO
>>> In_handle = open ("AP006852.gbk", "r")
>>> Out_handle = open ("AP006852.fasta", "w")
>>> records = SeqIO.parse (In_handle, "genbank")
>>> count = SeqIO.write (records, Out_handle, "fasta")
>>> count
1
>>>
>>> In_handle.close()
>>> Out_handle.close()
```

Εκτέλεση εξωτερικών εφαρμογών (π.χ. BLAST) τοπικά και από σενάριο

Εκτέλεση BLAST τοπικά

```
blastProgram -p Όνομα προγράμματος -d Βάση δεδομένων -i QueryFile -o BLASTreportOutputFile
```

Χρησιμοποιώντας τη γραμμή εντολών του κελύφους

```
$blastall -p blastp -d nr -i P05480.fasta -o Blastall.out
```

Από script

Χρησιμοποιώντας το os.system()

```
>>>import os
>>>S = "blastall -p blastp -d nr -i P05480.fasta -o Blastall.out"
>>>os.system(S)
```

Χρήση Biopython

```
>>>από το Bio.Blast.Εισαγωγή του BlastallCommandline του Applications
>>>cline = BlastallCommandline(program ="blastp", >>>infile="P05480.fasta",
database="nr", output="Blastall.out" )
>>> τυπωμένη ύλη
>>>blastall -d nr -i P05480.fasta -o Blastall.out -p blastp
>>>os.system(str(cline))
```

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Εκτέλεση BLAST μέσω Διαδικτύου

Bio.Blast.NCBIWWW

```
>>> from import Bio.Blast import NCBIWWW
>>> result_handle = NCBIWWW.qblast("blastn","nr","8332116")
>>> save_file = open("qblast_blastn.out", "w")
>>> save_file.write(result_handle.read())
>>> save_file.close()
>>> result_handle.close()
```

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Ανάλυση της εξόδου BLAST

Μπορείτε να λάβετε έξοδο BLAST σε XML με διάφορους τρόπους: για ο αναλυτής δεν έχει σημασία πώς δημιουργήθηκε η έξοδος αρκεί να είναι σε μορφή XML

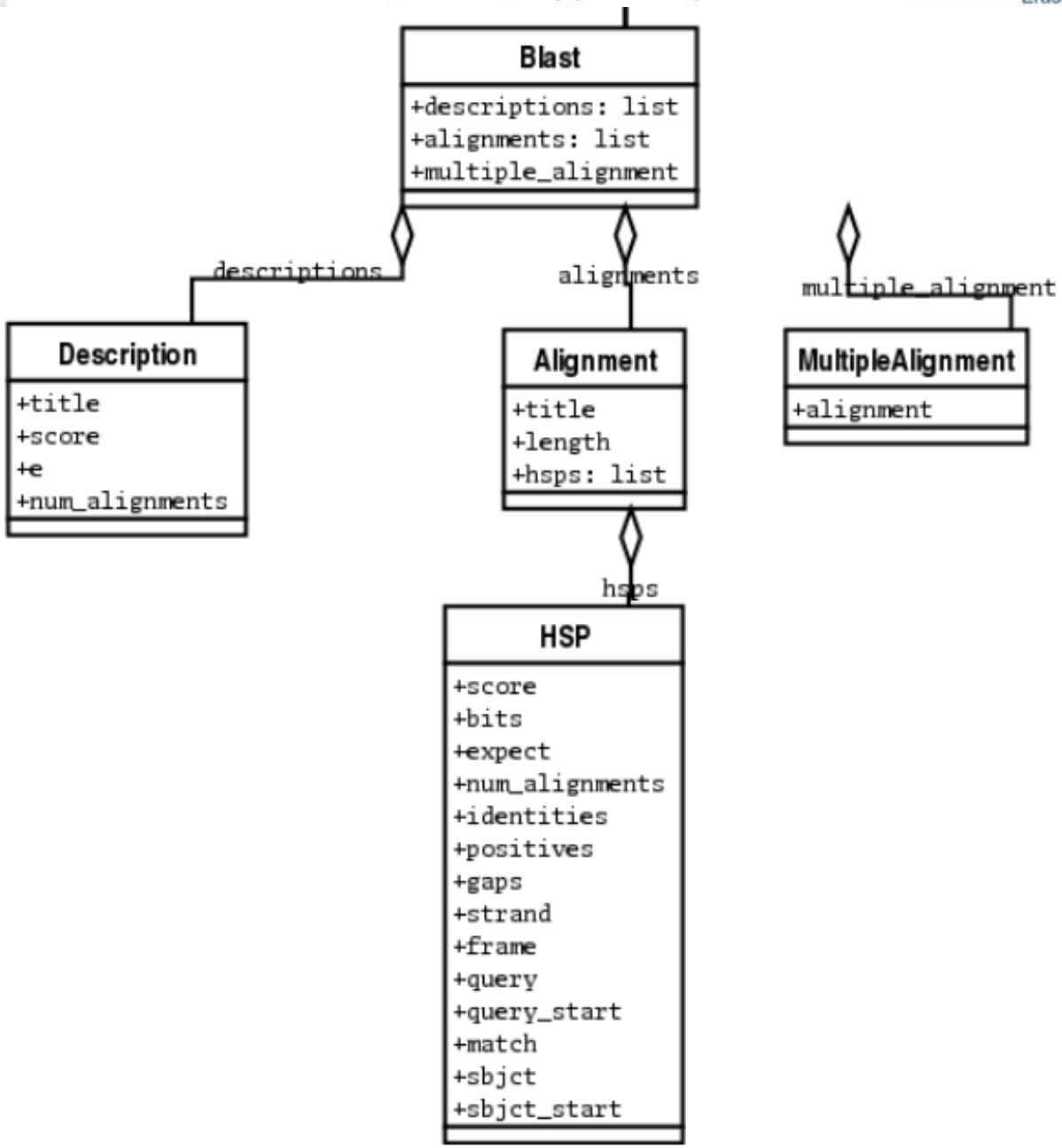
- Χρησιμοποιήστε το Biopython για να εκτελέσετε το BLAST μέσω Διαδικτύου
- Χρησιμοποιήστε το Biopython για να εκτελέσετε τοπικά το BLAST
- Κάντε την αναζήτηση BLAST μόνοι σας στον ιστότοπο του NCBI μέσω του ιστού σας πρόγραμμα περιήγησης και, στη συνέχεια, αποθηκεύστε τα αποτελέσματα (επιλέξτε μορφή XML για την έξοδο)
- Εκτελέστε το BLAST τοπικά χωρίς να χρησιμοποιήσετε το Biopython και αποθηκεύστε το αποτέλεσμα σε ένα αρχείο
(επιλέξτε μορφή XML για το αρχείο εξόδου)

Ανάλυση της εξόδου BLAST

Bio.Blast.NCBIXML

```
>>> result_handle = open(" qblast_blastn.out ")
>>> from Bio.Blast import NCBIXML
#Εάν περιμένετε ένα μόνο αποτέλεσμα BLAST
>>> blast_record = NCBIXML.read ( result_handle )

# Εάν έχετε πολλά αποτελέσματα
>>> blast_records = NCBIXML.parse ( result_handle )
```

Co-funded by the Erasmus+ Programme



Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 4. Θέματα Επιλογής

Μάθημα 3. Επιχειρησιακή Έρευνα

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Η Φύση και η Ιστορία της Επιχειρησιακής Έρευνας

- Το αντικείμενο, η επιχειρησιακή έρευνα ή η επιστήμη της διαχείρισης (παρόλο που μπορεί να υπάρχουν φιλοσοφικές διαφορές, χρησιμοποιούμε τους δύο όρους εναλλακτικά), έχει οριστεί από πολλούς ερευνητές στο πεδίο.
- Οι ορισμοί κυμαίνονται από "μια επιστημονική προσέγγιση στη λήψη αποφάσεων", στη "χρήση ποσοτικών εργαλείων για συστήματα που προέρχονται από την πραγματική ζωή", "επιστημονική λήψη αποφάσεων" και άλλα.
- Στα μέσα της δεκαετίας του 1970, η Επιχειρησιακή Ερευνητική Εταιρεία της Αμερικής (τότε μία από τις δύο μεγάλες επαγγελματικές εταιρείες στον τομέα) όρισε το θέμα ως εξής:

«Η Επιχειρησιακή Έρευνα ασχολείται με την επιστημονική απόφαση για τον καλύτερο σχεδιασμό και τη λειτουργία συστημάτων ανθρώπου-μηχανής συνήθως υπό συνθήκες που απαιτούν την κατανομή σπάνιων πόρων».

Εισαγωγή

- **Η Επιχειρησιακή Έρευνα** είναι Τέχνη και Επιστήμη
- Είχε τις πρώιμες ρίζες του στον Β' Παγκόσμιο Πόλεμο και ανθίζει στις επιχειρήσεις και τη βιομηχανία με τη βοήθεια υπολογιστών
- Οι κύριοι τομείς εφαρμογών της Επιχειρησιακής Έρευνας περιλαμβάνουν την πρόβλεψη, τον προγραμματισμό παραγωγής, τον έλεγχο αποθεμάτων, τον προϋπολογισμό κεφαλαίου και τη μεταφορά.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Τι είναι η Επιχειρησιακή Έρευνα;

Λειτουργίες
Οι δραστηριότητες που πραγματοποιούνται σε έναν οργανισμό.
Ερευνα
Η διαδικασία παρατήρησης και δοκιμής που χαρακτηρίζεται από την επιστημονική μέθοδο. Κατάσταση, δήλωση προβλήματος, κατασκευή μοντέλου, επικύρωση, πειραματισμός, υποψήφιες λύσεις.

Η Επιχειρησιακή Έρευνα είναι μια ποσοτική προσέγγιση στη λήψη αποφάσεων που βασίζεται στην επιστημονική μέθοδο επίλυσης προβλημάτων.

Τι είναι η Επιχειρησιακή Έρευνα;

- **<u>Επιχειρησιακή έρευνα</u>** είναι η επιστημονική προσέγγιση για την εκτέλεση της λήψης αποφάσεων, η οποία αποτελείται από :
  - Η τέχνη των ***μαθηματικών*** σύνθετων καταστάσεων
  - Η επιστήμη της ανάπτυξης ***τεχνικών λύσης*** χρησιμοποιείται για την επίλυση αυτών των μοντέλων
  - Η ικανότητα αποτελεσματικής ***επικοινωνίας*** τα αποτελέσματα στον υπεύθυνο λήψης αποφάσεων

Co-funded by the Erasmus+ Programme of the European Union

Μοντέλα Επιχειρησιακής Έρευνας

Ντετερμινιστικά Μοντέλα Στοχαστικά Μοντέλα

• Γραμμικός προγραμματισμός • Αλυσίδες Markov διακριτού χρόνου

• Βελτιστοποίηση δικτύου • Αλυσίδες Markov Continuous-Time

• Προγραμματισμός ακέραιων αριθμών • Θεωρία ουράς (γραμμές αναμονής)

• Μη γραμμικός προγραμματισμός • Ανάλυση αποφάσεων

•Μοντέλα απογραφής Θεωρία παιγνίων

Προσομοίωση

Co-funded by the Erasmus+ Programme of the European Union

Ντετερμινιστικά έναντι Στοχαστικών Μοντέλων

Ντετερμινιστικά μοντέλα
Ας υποθέσουμε ότι όλα τα δεδομένα είναι γνωστά με βεβαιότητα

Στοχαστικά μοντέλα
αντιπροσωπεύουν ρητά αβέβαια δεδομένα μέσω τυχαίες μεταβλητές ή στοχαστικές διαδικασίες.

Τα ντετερμινιστικά μοντέλα περιλαμβάνουν βελτιστοποίηση

Στοχαστικά μοντέλα
χαρακτηρίζουν / εκτιμούν την απόδοση του συστήματος.

Ανάπτυξη Μοντέλου

- Τα μοντέλα είναι αναπαραστάσεις πραγματικών αντικειμένων ή καταστάσεων.
- Τρεις μορφές μοντέλων είναι τα εικονικά, τα αναλογικά και τα μαθηματικά.
  - Τα εικονικά μοντέλα είναι φυσικά αντίγραφα (βαθμωτές αναπαραστάσεις) πραγματικών αντικειμένων.
  - Τα αναλογικά μοντέλα έχουν φυσική μορφή, αλλά δεν μοιάζουν φυσικά με το αντικείμενο που μοντελοποιείται.
  - Τα μαθηματικά μοντέλα αντιπροσωπεύουν προβλήματα του πραγματικού κόσμου μέσω ενός συστήματος μαθηματικών τύπων και εκφράσεων που βασίζονται σε βασικές υποθέσεις, εκτιμήσεις ή στατιστικές αναλύσεις.

Πλεονεκτήματα των μοντέλων

- Γενικά, ο πειραματισμός με μοντέλα (σε σύγκριση με τον πειραματισμό με την πραγματική κατάσταση):
  - απαιτεί λιγότερο χρόνο
  - είναι λιγότερο ακριβό
  - εμπεριέχει λιγότερο κίνδυνο

Μαθηματικά Μοντέλα

- Κατά την επιλογή ενός κατάλληλου μαθηματικού μοντέλου πρέπει να λαμβάνονται υπόψη το κόστος/ωφέλεια.
- Συχνά ένα λιγότερο περίπλοκο (και ίσως λιγότερο ακριβές) μοντέλο είναι πιο κατάλληλο από ένα πιο περίπλοκο και ακριβές λόγω κόστους και ευκολίας λύσης.

Μαθηματικά Μοντέλα

- Συσχέτιση μεταβλητών απόφασης (ελεγχόμενες είσοδοι) με σταθερές ή μεταβλητές παραμέτρους (μη ελεγχόμενες εισόδους).
- Συχνά επιδιώκετε να μεγιστοποιήσετε ή να ελαχιστοποιήσετε κάποια αντικειμενική συνάρτηση που υπόκειται σε περιορισμούς.
- Λέγεται ότι είναι στοχαστική εάν κάποια από τις μη ελεγχόμενες εισόδους (παράμετροι) υπόκειται σε παραλλαγή (τυχαία), διαφορετικά λέγεται ότι είναι ντετερμινιστική .
- Γενικά, τα στοχαστικά μοντέλα είναι πιο δύσκολο να αναλυθούν.
- Οι τιμές των μεταβλητών απόφασης που παρέχουν την καλύτερη μαθηματικά έξοδο αναφέρονται ως η βέλτιστη λύση για το μοντέλο.

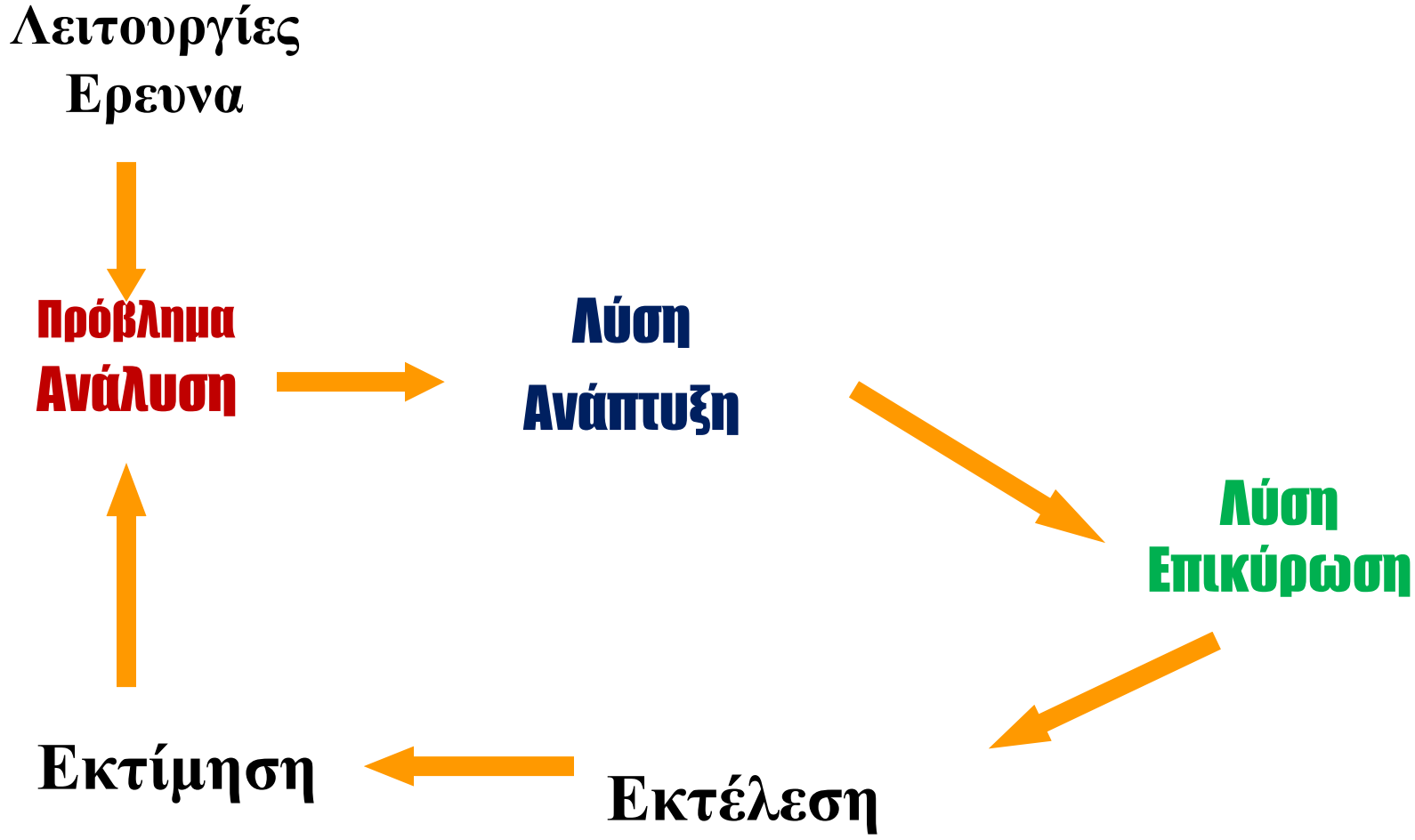
Μοντέλο Λύσης

- Περιλαμβάνει τον προσδιορισμό των τιμών των μεταβλητών απόφασης που παρέχουν το «καλύτερο» αποτέλεσμα για το μοντέλο.
- Μια προσέγγιση είναι η δοκιμή και το σφάλμα.
  - μπορεί να μην παρέχει την καλύτερη λύση
  - αναποτελεσματικό (απαιτούνται πολυάριθμοι υπολογισμοί)
- Ειδικές διαδικασίες επίλυσης έχουν αναπτυχθεί για συγκεκριμένα μαθηματικά μοντέλα.
  - ορισμένα μικρά μοντέλα/προβλήματα μπορούν να λυθούν με υπολογισμούς με το χέρι
  - Οι περισσότερες πρακτικές εφαρμογές απαιτούν τη χρήση υπολογιστή

Δοκιμή και επικύρωση μοντέλου

- Συχνά, η καλοσύνη/ακρίβεια ενός μοντέλου δεν μπορεί να αξιολογηθεί έως ότου δημιουργηθούν λύσεις.
- Μικρά προβλήματα δοκιμής με γνωστές ή τουλάχιστον αναμενόμενες λύσεις μπορούν να χρησιμοποιηθούν για τη δοκιμή και την επικύρωση του μοντέλου.
- Εάν το μοντέλο παράγει αναμενόμενες λύσεις:
  - χρησιμοποιήστε το μοντέλο για το πρόβλημα πλήρους κλίμακας.
- Εάν εντοπιστούν ανακρίβειες ή πιθανές ελλείψεις που είναι εγγενείς στο μοντέλο, προβείτε σε διορθωτικές ενέργειες όπως:
  - συλλογή πιο ακριβών δεδομένων εισαγωγής
  - τροποποίηση του μοντέλου

Επιχειρησιακή Έρευνα – Μεθοδολογία



Προσέγγιση Επιχειρησιακής Έρευνας

- **Φάση -Ι ΑΝΑΛΥΣΗ ΠΡΟΒΛΗΜΑΤΟΣ**
  - **Ορίστε την Επιχειρησιακή Έρευνα**
  - **Αναλύστε το πρόβλημα και χωρίστε το σε μικρότερες μονάδες**
  - **Καθορισμός ερευνητικών προτεραιοτήτων**

Προσέγγιση Επιχειρησιακής Έρευνας

- **Φάση-ΙΙ ΑΝΑΠΤΥΞΗ ΛΥΣΗΣ**
  - **Προσδιορίστε τους στόχους λύσης**
  - **Προσδιορίστε μεταβλητές απόφασης και ορίστε περιορισμούς στη λύση**
  - **Προσδιορίστε ή κατασκευάστε ένα κατάλληλο μοντέλο για την ανάπτυξη λύσεων**
  - **Προσδιορίστε και λάβετε τα απαιτούμενα δεδομένα**
  - **Ανάπτυξη λύσεων χρησιμοποιώντας αναλυτικό μοντέλο**

Προσέγγιση Επιχειρησιακής Έρευνας

- **Φάση-ΙΙΙ ΕΠΙΚΥΡΩΣΗ ΛΥΣΗΣ**
  - **Δοκιμή πεδίου σχεδιασμού**
  - **Εφαρμογή δοκιμής πεδίου**
  - **Αξιολογήστε την προτεινόμενη λύση**
  - **τροποποιήστε εάν είναι απαραίτητο**
  - **Ενσωματώστε τη λύση με το μεγαλύτερο σύστημα**

Φάση -Ι, Ανάλυση Προβλήματος

1. Ορίστε το λειτουργικό πρόβλημα

Συζήτηση του προβλήματος με όλους τους ενδιαφερόμενους

Προσδιορισμός υποσυστημάτων

Αναπτύξτε στόχους για υποσυστήματα

επίτευξη συναίνεσης για τους στόχους του υποσυστήματος

Φάση -Ι, Ανάλυση Προβλήματος

2. Αναλύστε το πρόβλημα και χωρίστε σε μικρότερες μονάδες

Το λειτουργικό πρόβλημα αναλύεται στο λεπτομέρεια.

Το πρόβλημα χωρίζεται σε μικρότερα μονάδες.

Αυτό παρουσιάζεται γραφικά.

Το γράφημα δείχνει τις εισόδους, τις εξόδους και τους εξωτερικούς παράγοντες.

Co-funded by the Erasmus+ Programme of the European Union

Φάση -Ι, Ανάλυση Προβλήματος

3. Καθορισμός ερευνητικών προτεραιοτήτων

Πρώτα κοιτάξτε το σύστημα ως σύνολο και μετά ως ένα σύμπλεγμα μεγάλων υποσυστημάτων.

Στη συνέχεια, ρωτήστε πού θα πρέπει να περιμένουμε ότι οι τροποποιήσεις θα έχουν τον μεγαλύτερο θετικό αντίκτυπο στα αποτελέσματα.

Φάση -ΙΙ, Ανάπτυξη Λύσης

1. Προσδιορίστε τους στόχους λύσης

Οι στόχοι λύσης είναι δηλώσεις των επιθυμητών λύσεων στο λειτουργικό πρόβλημα.

Αυτά συνήθως δηλώνονται ως προς την αποδοτικότητα του συστήματος που είναι η διατήρηση των εισροών και η μεγιστοποίηση των εκροών.

Σταθερές είσοδοι με μέγιστες εξόδους.

Ελάχιστες είσοδοι με μέγιστες εξόδους.

Φάση -ΙΙ, Ανάπτυξη Λύσεων

2. Προσδιορίστε μεταβλητές απόφασης και ορίστε περιορισμούς στη λύση

- **Οι μεταβλητές απόφασης είναι παράγοντες που παίζουν ρόλο στον καθορισμό του τρόπου λειτουργίας ενός συστήματος και είναι επίσης υπό τον έλεγχο του διαχειριστή συστήματος.**
  - **Ποσότητα πακέτων ORS**
  - **Μέθοδοι εκπαίδευσης**
- **Οι παράγοντες που περιορίζουν το πρακτικό εύρος μιας ή περισσότερων μεταβλητών απόφασης ονομάζονται περιορισμοί.**
  - **Κοινωνικο-οικονομικό, εθνικότητα, επάγγελμα, μορφωτικό επίπεδο, καιρός, έδαφος**

Co-funded by the Erasmus+ Programme of the European Union

Φάση -ΙΙ, Ανάπτυξη Λύσεων

3. Προσδιορίστε ή κατασκευάστε ένα κατάλληλο μοντέλο για την ανάπτυξη λύσεων

Η κατασκευή μοντέλων είναι η ουσία της επιχειρησιακής έρευνας

Ένα μοντέλο είναι η απλοποιημένη αναπαράσταση του πραγματικού κόσμου

- **Κατανομή των πόρων**
- **Δικτύωση**
- **Ανάλυση κόστους**

Co-funded by the Erasmus+ Programme of the European Union

Φάση-ΙΙΙ ΕΠΙΚΥΡΩΣΗ ΛΥΣΗΣ

1. Δοκιμή πεδίου σχεδιασμού

Πόσο πιστευτό είναι το αποτέλεσμα της δοκιμής πεδίου .

Πειραματικός

Μη πειραματικό

Οιονεί πειραματικό

Co-funded by the Erasmus+ Programme of the European Union

Φάση-ΙΙΙ ΕΠΙΚΥΡΩΣΗ ΛΥΣΗΣ

2. Εφαρμογή της δοκιμής πεδίου

- Προσδιορισμός των απαιτούμενων πόρων
- Ανάπτυξη σχεδίου διαχείρισης
- Ανάπτυξη εκπαιδευτικού υλικού
- Ρυθμίσεις για την απόκτηση ανθρώπινων και άλλων πόρων που απαιτούνται
- Σχεδιασμός πληροφοριακού συστήματος

Co-funded by the Erasmus+ Programme of the European Union

Φάση-ΙΙΙ ΕΠΙΚΥΡΩΣΗ ΛΥΣΗΣ

3. Αξιολόγηση, τροποποίηση και ολοκλήρωση των λύσεων

- **Συνεχής αξιολόγηση για την εξασφάλιση βελτιώσεων των εισροών**

- **Ενσωμάτωση της λύσης στο σύστημα**

Παράδειγμα: Ponderosa Development Corp.

Υπάλληλος	**Μηνιαίος μισθός**
Πρόεδρος	$10.000
VP, Ανάπτυξη	6.000
Αντιπρόεδρος, Μάρκετινγκ	4.500
Project Manager	5.500
Ελεγκτής	4.000
Office Manager	3.000
Ρεσεψιονίστ	2.000

Παράδειγμα: Ponderosa Development Corp.

- **Ερώτηση** :

Προσδιορίστε όλα τα κόστη και υποδηλώστε το οριακό κόστος και τα οριακά έσοδα για κάθε σπίτι.

- **Απάντηση** :

Οι μηνιαίοι μισθοί ανέρχονται σε 35.000 $ και το μηνιαίο κόστος μίσθωσης γραφείου και προμήθειας συνολικά άλλα 5.000 $. Αυτά τα 40.000 $ είναι ένα μηνιαίο πάγιο κόστος.

Το συνολικό κόστος γης, υλικών, εργασίας και προμήθειας πωλήσεων ανά σπίτι, 105.000 $, είναι το οριακό κόστος για ένα σπίτι.

Η τιμή πώλησης των 115.000 $ είναι τα οριακά έσοδα ανά σπίτι.

Παράδειγμα: Ponderosa Development Corp.

- **Ερώτηση** :

Γράψτε τη συνάρτηση μηνιαίου κόστους $c(x)$, τη συνάρτηση εσόδων $r(x)$ και τη συνάρτηση κέρδους $p(x)$.

- **Απάντηση** :

$\gamma(x)$ = μεταβλητό κόστος + σταθερό κόστος = $105.000\,x + 40.000$

$r(x) = 115.000\,x$

$p(x) = r(x) - c(x) = 10.000\,x - 40.000$

Παράδειγμα: Ponderosa Development Corp.

- **Ερώτηση** :

Ποιο είναι το νεκρό σημείο για τις μηνιαίες πωλήσεις των κατοικιών;

- **Απάντηση** :

$r(x) = c(x)$ ή $115.000x = 105.000x + 40.000$

Επίλυση, $x = 4$.

- **Ερώτηση** :

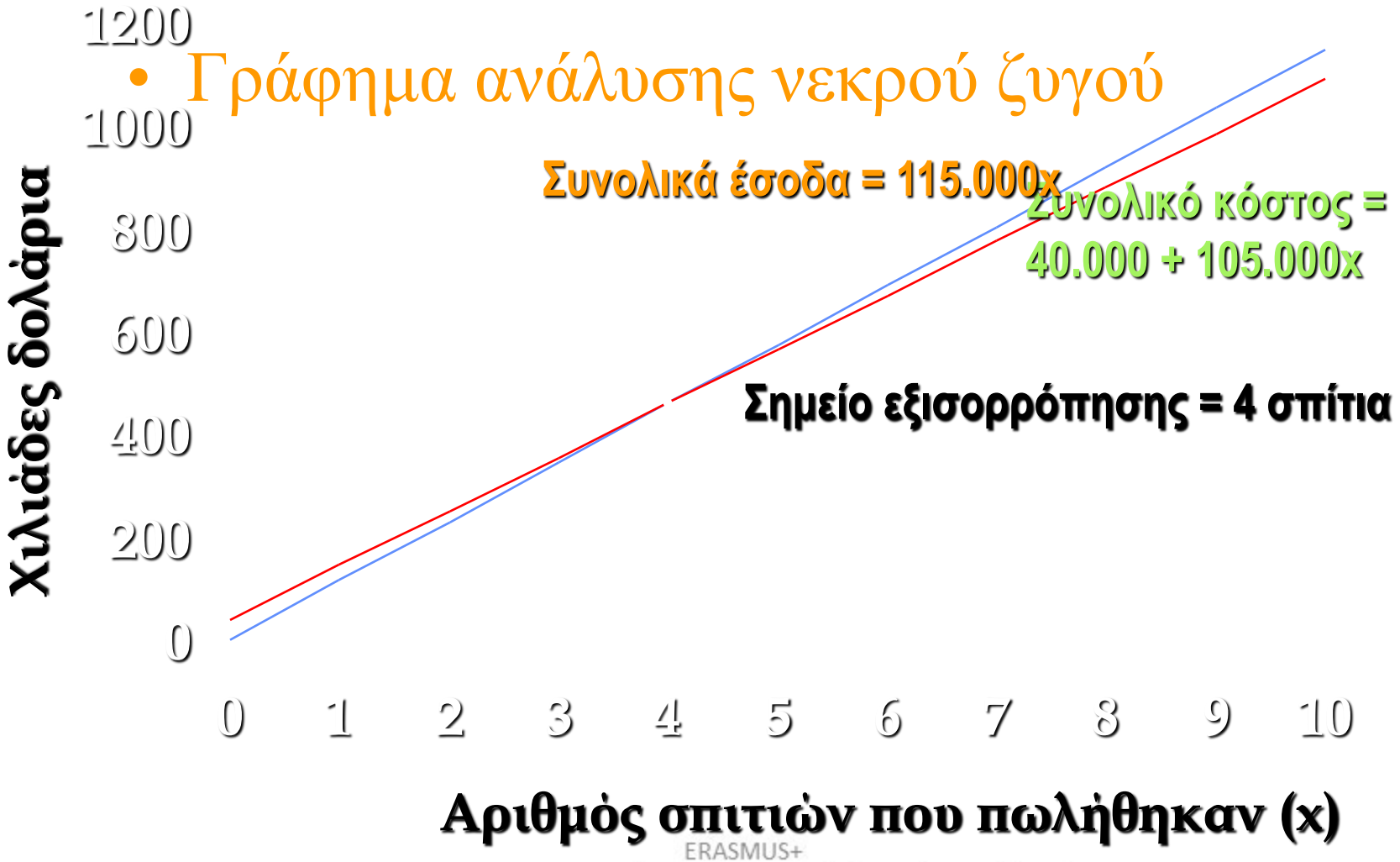
Ποιο είναι το μηνιαίο κέρδος αν χτιστούν και πουληθούν 12 σπίτια το μήνα;

- **Απάντηση** :

$p(12) = 10.000(12) - 40.000 = 80.000$ $ μηνιαίο κέρδος

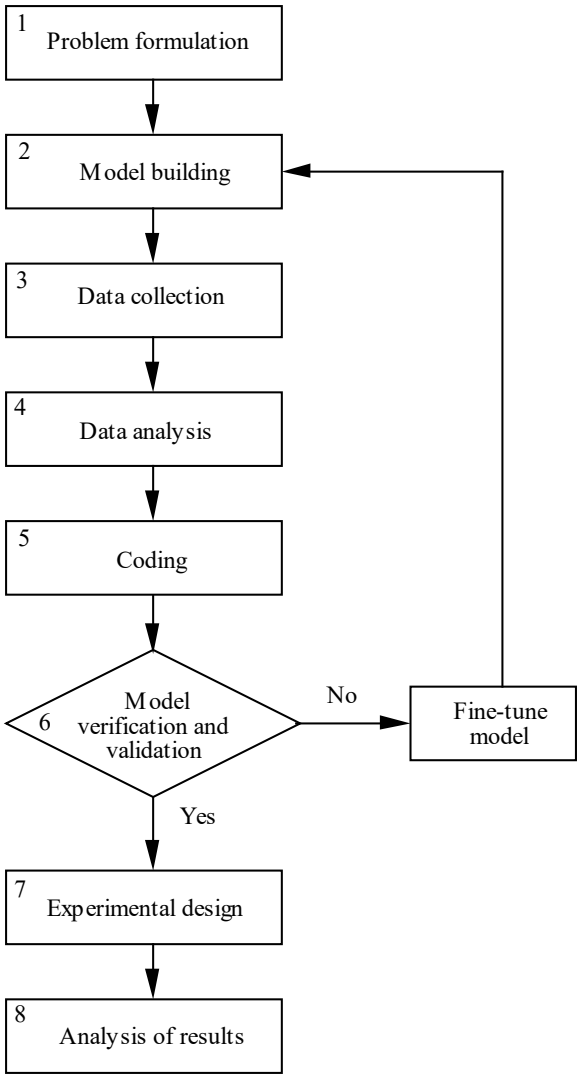
Παράδειγμα: Ponderosa Development Corp.

- Γράφημα ανάλυσης νεκρού ζυγού



Co-funded by the Erasmus+ Programme of the European Union

Βήματα στη μελέτη Η



Co-funded by the Erasmus+ Programme of the European Union

Περιοχές Εφαρμογής

- Στρατηγικό σχεδιασμό
- Διαχείριση εφοδιαστικής αλυσίδας
- Διαχείριση τιμών και εσόδων
- Logistics και τοποθεσία τοποθεσίας
- Βελτιστοποίηση
- Έρευνα μάρκετινγκ

Co-funded by the Erasmus+ Programme of the European Union

Τομείς Εφαρμογών

- Χρονοδρομολόγηση
- Διαχείριση χαρτοφυλακίου
- Ανάλυση αποθέματος
- Πρόβλεψη
- Ανάλυση πωλήσεων
- Δημοπρασία
- Ανάλυση κινδύνου

Παραδείγματα

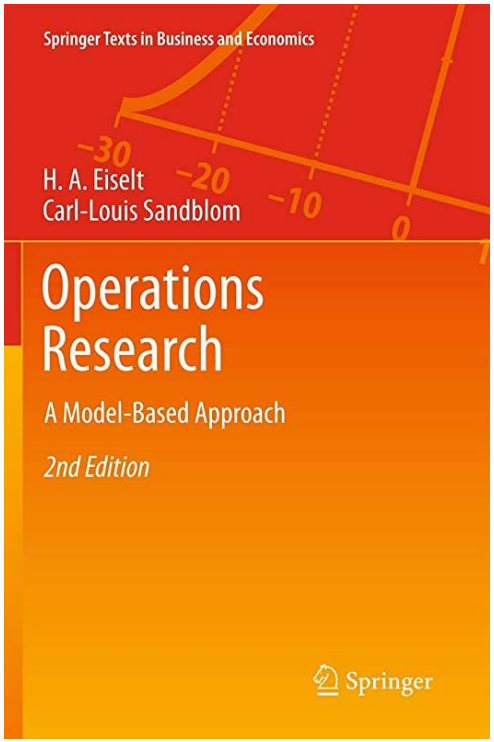
- Η British Telecom χρησιμοποίησε το OR για να προγραμματίσει το εργατικό δυναμικό για περισσότερους από 40.000 αρχειακούς μηχανικούς. Το σύστημα εξοικονομούσε 150 εκατομμύρια δολάρια ετησίως από το 1997 έως το 2000. Το εργατικό δυναμικό προβλέπεται να εξοικονομήσει 250 εκατομμύρια δολάρια.
- Η Sears χρησιμοποιεί OR για να δημιουργήσει ένα σύστημα δρομολόγησης και προγραμματισμού οχημάτων που θα λειτουργεί πιο αποτελεσματικά τον στόλο παράδοσης και εξυπηρέτησης στο σπίτι -- 42 εκατομμύρια δολάρια σε ετήσια εξοικονόμηση
- Η UPS χρησιμοποιεί το OR για να επανασχεδιάσει το δίκτυο παράδοσης μιας νύχτας, εξοικονομώντας 87 εκατομμύρια δολάρια από το 2000 έως το 2002. Αναμένονται άλλα 189 εκατομμύρια δολάρια την επόμενη δεκαετία.
- Η USPS χρησιμοποιεί OR για να προγραμματίσει τον εξοπλισμό και το εργατικό δυναμικό στα κέντρα επεξεργασίας και διανομής αλληλογραφίας της. Μπορεί να επιτευχθεί εκτιμώμενη εξοικονόμηση σε 500 εκατομμύρια δολάρια.

Ερωτήσεις και ασκήσεις:

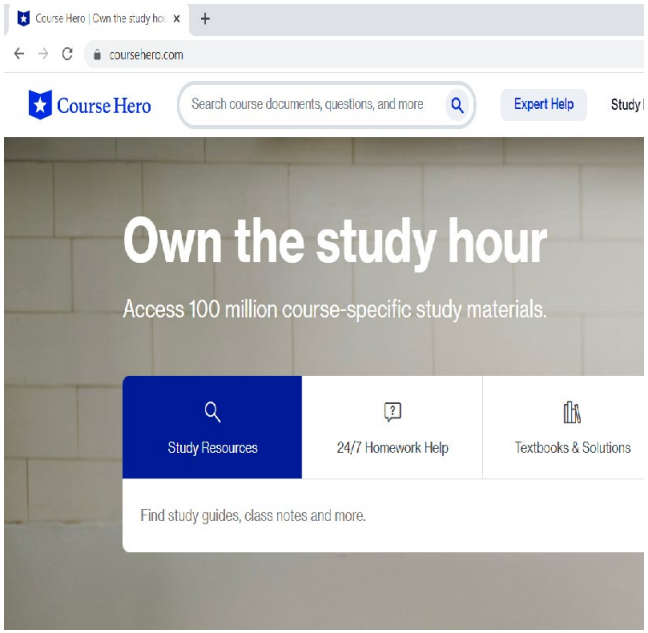
1. Τι είναι η Επιχειρησιακή Έρευνα και ποιοι είναι οι κύριοι στόχοι της;

2. Ποια είναι τα κύρια βήματα που περιλαμβάνονται στη διαδικασία Επιχειρησιακής Έρευνας;

3. Ποιες είναι οι πιο κοινές μέθοδοι που χρησιμοποιούνται στην Επιχειρησιακή Έρευνα;



HA Eiselt και Carl-Louis Sandblom. (2012) Επιχειρησιακή Έρευνα. A Model-Based Approach, Second Edition, Springer-Verlag Berlin Heidelberg.



https://www.coursehero.com/

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 4. Μαθήματα επιλογής

Μάθημα 4. Σύγχρονες Μέθοδοι στη Βιολογία Υπολογιστών

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Ανάλυση Αλληλουχίας
- Πρόβλεψη Δομής
- Γενετική επιδημιολογία και στατιστική γενετική
- Λειτουργική γονιδιωματική
- Φαρμακογονιδιωματική και εξατομικευμένη ιατρική

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Η υπολογιστική βιολογία και η βιοπληροφορική είναι ένα διεπιστημονικό πεδίο που αναπτύσσει και εφαρμόζει υπολογιστικές μεθόδους για την ανάλυση μεγάλων συλλογών βιολογικών δεδομένων, όπως γενετικές αλληλουχίες, πληθυσμούς κυττάρων ή δείγματα πρωτεϊνών, για να κάνει νέες προβλέψεις ή να ανακαλύψει νέα βιολογία. Οι υπολογιστικές μέθοδοι που χρησιμοποιούνται περιλαμβάνουν αναλυτικές μεθόδους, μαθηματική μοντελοποίηση και προσομοίωση.

Co-funded by the Erasmus+ Programme of the European Union

Ανάλυση Αλληλουχίας

Ανίχνευση κανονικοτήτων - Αυτές οι μέθοδοι συνήθως βασίζονται στο συμπέρασμα των μοτίβων ή κανονικοτήτων που αναζητούνται από ένα σύνολο γνωστών δεδομένων, π.χ. εκπαιδεύοντας ένα κρυφό μοντέλο Markov όπως αναφέρεται από τον Krogh. τότε οι εμφανίσεις αυτών των μοτίβων ή κανονικοτήτων μπορούν να ανιχνευθούν για να ληφθούν πληροφορίες για μια νέα ακολουθία.

Σύγκριση αλληλουχίας – Οι ομόλογες αλληλουχίες μπορούν να συγκριθούν απλώς για να ληφθούν πληροφορίες σχετικά με το πόσο απομακρυσμένα σχετίζονται ή για να προσπαθήσουμε να συμπεράνουμε εξελικτικούς περιορισμούς, π.χ. ένα μόριο RNA.

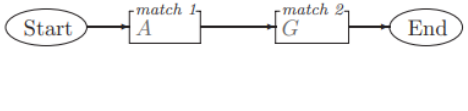
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

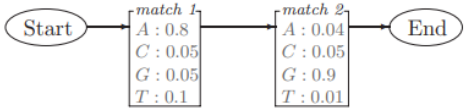
Ανάλυση Αλληλουχίας

Μοντέλα απόστασης

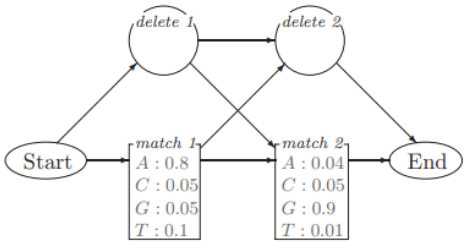
Κρυμμένα μοντέλα Markov



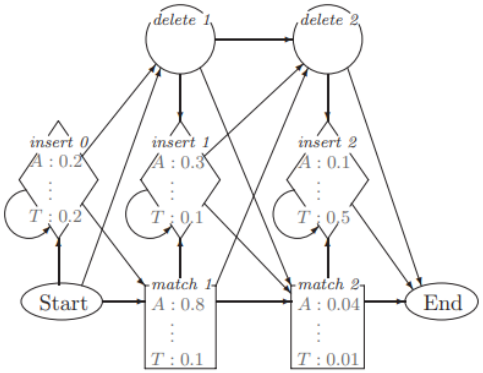
(a) A hidden Markov model for the sequence AG.



(b) A hidden Markov model for a sequence of two characters allowing substitutions.



(c) A hidden Markov model for an ancestral sequence of two characters allowing substitutions and deletions.



(d) A profile hidden Markov model with two ancestral positions.

Co-funded by the Erasmus+ Programme of the European Union

Πρόβλεψη Δομής

Ο σκοπός όλου του γενετικού υλικού που αποθηκεύεται σε όλους τους ζωντανούς οργανισμούς δεν είναι να δώσει στους βιολόγους χρήσιμες πληροφορίες για τη σύγκριση των οργανισμών. Μάλλον το γενετικό υλικό είναι το προσχέδιο για τα βιοχημικά μόρια που χρειάζονται οι οργανισμοί στις διαδικασίες που τους κάνουν να ζουν. Το γενετικό υλικό αποθηκεύεται στο DNA και μεταφράζεται σε RNA ή πρωτεΐνες, όλους τους τύπους μορίων που μας επιτρέπουν να το δούμε ως αλληλουχίες.

Πρόβλεψη Δομής

Η δευτερογενής δομή ενός βιομορίου αποτελείται από τοπικά δομικά στοιχεία της πλήρους τριτοταγούς δομής του μορίου. Για τις πρωτεΐνες αυτές οι τοπικές πληροφορίες αποτελούνται από τμήματα διαδοχικών αμινοξέων που σχηματίζουν κανονικά δομικά μοτίβα. Για το RNA είναι ένα σύνολο ζευγών βάσεων, που δεν βρίσκονται απαραίτητα κοντά το ένα στο άλλο, που σχηματίζουν ισχυρές αλληλεπιδράσεις. Μια σημαντική χρήση για πληροφορίες δευτερεύουσας δομής είναι η μείωση του αριθμού των βαθμών ελευθερίας κατά την προσπάθεια πρόβλεψης της πλήρους τριτογενούς δομής, αλλά και ως βοήθεια στην ευθυγράμμιση ακολουθιών και για την εξαγωγή λειτουργικών πληροφοριών βρίσκει χρήση.

Co-funded by the Erasmus+ Programme of the European Union

Πρόβλεψη Δομής

Στις πρωτεϊνικές δομές παρατηρούνται μικρά τμήματα διαδοχικών αμινοξέων που σχηματίζουν πολύ κανονικά δομικά μοτίβα που ονομάζονται α-έλικες και β-κλώνοι. Η α-έλικα είναι μια σπείρα περιέλιξης που μοιάζει με τιρμπουσόν και σταθεροποιείται με το σχηματισμό δεσμών υδρογόνου μεταξύ γειτονικών στροφών. Οι β-κλώνοι είναι τεντωμένα τμήματα της αλληλουχίας αμινοξέων που σταθεροποιούνται σχηματίζοντας δεσμούς υδρογόνου σε γειτονικούς β-κλώνους που εκτείνονται παράλληλα ή αντιπαράλληλα με τον κλώνο στη δομή αλλά μπορεί να βρίσκονται πολύ μακριά στην αλληλουχία. Υπολογιστικά η πρόβλεψη δευτερογενούς δομής πρωτεΐνης είναι συνήθως στη σφαίρα της αναγνώρισης προτύπων, π.χ. από νευρωνικά δίκτυα ή κρυφά μοντέλα Markov.

Co-funded by the Erasmus+ Programme of the European Union

Πρόβλεψη Δομής

Στην τριτοταγή δομή ενός μορίου RNA το πιο χαρακτηριστικό γνώρισμα είναι η στοίβαξη των ζευγών βάσεων. Ένα μόριο RNA, ακριβώς όπως ένα μόριο DNA, αποτελείται από μια ραχοκοκαλιά σακχάρου/φωσφορικού με (σχεδόν) τις ίδιες τέσσερις βάσεις αμίνης ως πιθανές πλευρικές αλυσίδες. Δεν αποτελεί λοιπόν έκπληξη το γεγονός ότι οι συμπληρωματικές βάσεις σε ένα μόριο RNA σχηματίζουν ισχυρές αλληλεπιδράσεις με δεσμούς υδρογόνου παρόμοιες με αυτές που βρίσκονται στη διπλή έλικα του DNA. Πρέπει να αναφερθεί ότι οι τέσσερις βάσεις του RNA μπορούν να σχηματίσουν άλλους τύπους δεσμών υδρογόνου από τους κλασσικούς δεσμούς Watson-Crick και ως μόριο RNA σχηματίζει μια δομή αναδιπλώνοντας τον εαυτό του και όχι σχηματίζοντας ζεύγη βάσεων με ένα μόριο με ένα συμπληρωματικό αλληλουχία βάσεων ως DNA, άλλα ζεύγη βάσεων από τα κλασικά ζεύγη βάσεων Watson-Crick παρατηρούνται σε δομές RNA. Ωστόσο, μόνο τα ζεύγη βάσεων G, U ή τα ζεύγη βάσεων ταλάντωσης είναι τόσο συχνά που χαρακτηρίζονται ως τυπικό ζεύγος βάσεων RNA.

Co-funded by the Erasmus+ Programme of the European Union

Πρόβλεψη Δομής

Η δυνατότητα πρόβλεψης της τριτοταγούς δομής μιας πρωτεΐνης (ή ενός μορίου RNA) είναι πολύ μεγαλύτερης σημασίας από την πρόβλεψη της δευτερογενούς δομής. Πρώτα απ 'όλα, από την πλήρη τριτογενή δομή είναι ασήμαντο να αφαιρεθεί η δευτερεύουσα δομή. Δεύτερον, η τριτογενής δομή είναι πολύ πιο ενημερωτική για την εξαγωγή λειτουργικών ιδιοτήτων. Θα πρέπει να αναφερθεί, ωστόσο, ότι μια πρωτεΐνη δεν είναι ένα εντελώς άκαμπτο μόριο, αλλά έχει μια κάπως δυναμική δομή που δονείται γύρω από μια ισορροπία γνωστή ως φυσική κατάσταση. Μια αλλαγή στο περιβάλλον της πρωτεΐνης θα τροποποιήσει συχνά αυτή την ισορροπία, μια τροποποίηση που είναι σημαντική κατά τις καταλυτικές διεργασίες στις οποίες συμμετέχει η πρωτεΐνη. Έτσι η τεταρτοταγής δομή είναι το πιο σημαντικό λειτουργικά δομικό επίπεδο.

Co-funded by the Erasmus+ Programme of the European Union

Γενετική επιδημιολογία και στατιστική γενετική

Ερευνητές στους τομείς της στατιστικής γενετικής και της γενετικής επιδημιολογίας αναπτύσσουν νέες μεθόδους και βελτιώνουν τις υπάρχουσες ροές εργασίας για τον εντοπισμό γενετικών παραγόντων κινδύνου για πολύπλοκα κλινικά χαρακτηριστικά. Το αποτέλεσμα αυτών των προσπαθειών είναι η βαθύτερη κατανόηση της αιτιολογίας της νόσου.

Co-funded by the Erasmus+ Programme of the European Union

Λειτουργική γονιδιωματική

Ο λειτουργικός σχολιασμός των γονιδιωμάτων βελτιώνει την κατανόηση των γενετικών μηχανισμών που οδηγούν σε κίνδυνο ασθένειας, συμπτώματα ή πρόγνωση και είναι σημαντικός για την ερμηνεία του πλήθους των προσωπικών γονιδιωματικών παραλλαγών σε σχέση με διάφορες ασθένειες και καρκίνο. Μια τέτοια ερμηνεία είναι μια από τις πιο σημαντικές προκλήσεις για τη βιοϊατρική ανακάλυψη και την εξατομικευμένη ιατρική. Επειδή οι περισσότερες γενετικές παραλλαγές εμφανίζονται σε μη κωδικοποιητικές περιοχές του γονιδιώματος, τα αποτελέσματά τους παραμένουν ελάχιστα κατανοητά.

Φαρμακογονιδιωματική και εξατομικευμένη ιατρική

Αναπτύσσονται μεθοδολογίες και λογισμικό βιοπληροφορικής για την ανάκριση πολύπλοκων δεδομένων αλληλουχίας και την εφαρμογή τεχνικών βιολογίας συστημάτων για απάντηση σε διαγνωστικά και φαρμακογονιδιωματικά ερωτήματα, συμβάλλοντας έτσι στην υπέρβαση των ορίων κατανόησης των γενετικών παραγόντων πίσω από μια μεγάλη ποικιλία ανθρώπινων ασθενειών, συμπεριλαμβανομένου του καρκίνου.

Τα προσαρμοσμένα σχέδια θεραπείας που βασίζονται σε γενετικές εκτροπές ενδέχεται να επηρεάσουν τις πρωτεϊνικές αλληλεπιδράσεις των ασθενών, την κυτταρική συμπεριφορά, τη λειτουργία οργάνων και τη σωματική υγεία. Η επινόηση τέτοιων σχεδίων απαιτεί βαθιά κατανόηση της ανάλυσης αλληλουχίας, της φαρμακογονιδιωματικής και της βιολογίας συστημάτων. Οι ειδικοί του Τμήματος Υπολογιστικής Βιολογίας συνεργάζονται με ερευνητές και κλινικούς γιατρούς για να ανακαλύψουν γονιδιωματικές παραλλαγές που απαιτούν προσαρμοσμένα σχέδια θεραπείας και να τους υποστηρίξουν στη δημιουργία γενετικά προσαρμοσμένων σχεδίων φροντίδας.

Co-funded by the Erasmus+ Programme of the European Union

Υπολογιστική και στατιστική μεταγονιδιωματική

Η συλλογή των μικροοργανισμών που βρίσκονται στο ανθρώπινο σώμα (το ανθρώπινο μικροβίωμα) παίζει ζωτικό ρόλο στην υγεία και τις ασθένειες. Το ανθρώπινο μικροβίωμα μελετάται με τη βοήθεια τεχνολογιών μεταγονιδιωματικής αλληλουχίας, χρησιμοποιώντας είτε στοχευμένη είτε κυνηγετική προσέγγιση.

Αναπτύσσονται και εφαρμόζονται υπολογιστικά και στατιστικά εργαλεία στην ανάλυση δεδομένων αλληλουχίας μεταγονιδιωματικών. Αυτοί οι σημαντικοί πόροι υποστηρίζουν την έρευνα μικροβιώματος που εστιάζεται σε ασθένειες σε όλο το ίδρυμα και βοηθούν τους κλινικούς ερευνητές να κατανοήσουν τον ρόλο του ανθρώπινου μικροβιώματος στην ευαισθησία, την έναρξη, την εξέλιξη και την ανταπόκριση στη θεραπεία της νόσου και τελικά να ενσωματώσουν δεδομένα μικροβιώματος στην εξατομικευμένη ιατρική.

Co-funded by the Erasmus+ Programme of the European Union

Γονιδιωματική και εξέλιξη του καρκίνου

Ένας νέος ορίζοντας στη μελέτη του καρκίνου είναι ο τρόπος με τον οποίο οι επιστήμονες ταξινομούν, ποσοτικοποιούν και δίνουν νόημα στην αλλαγή των γονιδιωματικών υπογραφών εντός και μεταξύ των όγκων με την πάροδο του χρόνου. Η χωρική και χρονική εξέλιξη των καρκινικών κυττάρων διερευνάται μέσω της ανάλυσης μεμονωμένων κυττάρων. Αυτές οι μελέτες θα επιτρέψουν στους επιστήμονες και τους κλινικούς γιατρούς να κατανοήσουν καλύτερα την εξέλιξη του καρκίνου οποιουδήποτε ασθενούς και να αναπτύξουν πιο αποτελεσματικές θεραπείες. Ο τομέας της γονιδιωματικής και της εξέλιξης του καρκίνου έχει τη δυνατότητα να φέρει επανάσταση στην κατανόηση του καρκίνου και να μεταμορφώσει τη φροντίδα των ασθενών.

Co-funded by the Erasmus+ Programme of the European Union

Μονοκυτταρική γονιδιωματική

Η αλληλουχία ενός κυττάρου παρέχει μια άνευ προηγουμένου ανάλυση δραστηριότητας σε κυτταρικό επίπεδο. Οι κλινικοί και οι βασικοί ερευνητές χρησιμοποιούν όλο και περισσότερο την αλληλουχία γονιδιώματος μονοκυττάρου, μεθυλώματος και μεταγραφώματος για να κατανοήσουν καλύτερα τους μηχανισμούς της νόσου. Οι ερευνητές του τμήματος δραστηριοποιούνται τόσο στην εφαρμοσμένη όσο και στη μεθοδολογική έρευνα της μονοκυτταρικής γονιδιωματικής. Προσπαθούν να καθιερώσουν τις βέλτιστες πρακτικές για την ανάλυση δεδομένων αλληλουχίας ενός κυττάρου και να αναπτύξουν υπολογιστικές αγωγούς για τον εξορθολογισμό της ανάλυσης. Χρησιμοποιώντας τις πρόσφατες εξελίξεις στις στατιστικές υψηλών διαστάσεων και τη βαθιά μάθηση, αναπτύσσουν ποσοτικά εργαλεία για την κάλυψη των μεταγενέστερων αναλυτικών αναγκών, συμπεριλαμβανομένης της ανάλυσης διαφορικής αφθονίας και έκφρασης, της ολοκλήρωσης πολλαπλών ωμικών και της μοντελοποίησης πρόβλεψης.

Πρωτεομική και μεταβολομική

Το πρωτείωμα και το μεταβολίωμα μαζί είναι οι κύριοι μεσολαβητές μεταξύ του γονότυπου ενός βιολογικού συστήματος και του φαινοτύπου του. Είναι επίσης βασικοί φορείς της απόκρισης ενός συστήματος σε εξωτερικά ερεθίσματα. Ως εκ τούτου, η μελέτη πρωτεϊνών και μεταβολιτών και η δυναμική τους είναι το κλειδί για την κατανόηση ενός βιολογικού συστήματος. Για το σκοπό αυτό, η πρωτεϊνομική και η μεταβολομική προσφέρουν μια σειρά εργαλείων «ωμικής» υψηλής απόδοσης για τη μέτρηση της έκφρασης και της τροποποίησης των πρωτεϊνών και των μεταβολιτών που υπάρχουν σε ένα σύστημα. Τα μέλη ΔΕΠ Υπολογιστικής Βιολογίας είναι ειδικοί στην ανάλυση πρωτεομικών και μεταβολομικών δεδομένων για να βρουν μόρια και μονοπάτια που εκφράζονται διαφορικά ή τροποποιούνται διαφορικά μεταξύ πειραματικών συνθηκών ή ομάδων ασθενών. Η σχολή μας έχει εξειδίκευση στη χρήση νέων αλγορίθμων για την ανάλυση πρωτεωμικών και μεταβολικών δεδομένων για την αντιμετώπιση μιας μεγάλης ποικιλίας ερωτήσεων στους τομείς της κλινικής διάγνωσης, της φαινομενικής του καρκίνου, του ανθεκτικού στην ινσουλίνη διαβήτη, της καρδιαγγειακής υγείας και άλλων οξειών ή χρόνιων φυσιολογικών παθήσεων

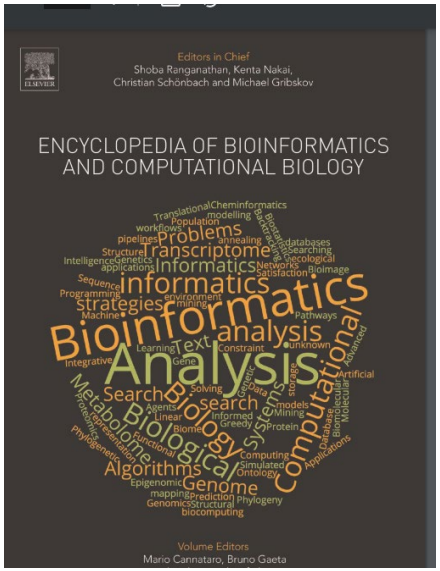
Ψηφιακή παθολογία

Οι κύριοι περιορισμοί της MVD είναι ότι: (α) η στοχαστική φύση της μηχανής σύνδεσης συνεπάγεται ότι μπορεί να χρειαστούν περισσότερες από μία διαδρομές σύνδεσης για τον προσδιορισμό του σωστού τρόπου σύνδεσης. και (β) η κατάσταση πρωτονίωσης του ζεύγους συνδέτη/υποδοχέα δεν αναζητείται εξαντλητικά και χρησιμοποιούνται προκαθορισμένες καταστάσεις πρωτονίωσης κατά τη σύνδεση.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Ποιο είναι το αντικείμενο της υπολογιστικής βιολογίας;
- Ποιοι είναι οι τομείς εφαρμογής των μεθόδων υπολογιστών στη βιολογία;



Shoba Ranganathan, Michael Gribskov , Kenta Nakai , Christian Schönbach (2019) ENCYCLOPEDIA OF BIOINFORMATICS AND COMPUTATIONAL BIOLOGY, Elsevier, Άμστερνταμ, Ολλανδία

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 4. Μαθήματα Επιλογής

Μάθημα 5. Στατιστική Ανάλυση

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Γιατί να μελετήσετε στατιστικά;

1. Τα δεδομένα υπάρχουν παντού
2. Οι στατιστικές τεχνικές χρησιμοποιούνται για τη λήψη πολλών αποφάσεων που επηρεάζουν τη ζωή μας
3. Ανεξάρτητα από την καριέρα σας, θα πάρετε επαγγελματικές αποφάσεις που περιλαμβάνουν δεδομένα. Η κατανόηση των στατιστικών μεθόδων θα σας βοηθήσει να πάρετε αυτές τις αποφάσεις αποτελεσματικά

Εφαρμογές στατιστικών εννοιών στον επιχειρηματικό κόσμο

- **Οικονομικά** – συσχέτιση και παλινδρόμηση, αριθμοί δεικτών, ανάλυση χρονοσειρών
- **Μάρκετινγκ** – έλεγχος υποθέσεων, τεστ chi-square, μη παραμετρικές στατιστικές
- **Προσωπικό** – έλεγχος υποθέσεων, τεστ chi-square, μη παραμετρικοί έλεγχοι
- **Λειτουργική διαχείριση** – έλεγχος υποθέσεων, εκτίμηση, ανάλυση διασποράς, ανάλυση χρονοσειρών

Co-funded by the Erasmus+ Programme of the European Union

Εφαρμογές στατιστικών εννοιών στον επιχειρηματικό κόσμο

- **Η στατιστική ανάλυση διαδραματίζει κρίσιμο ρόλο στη βιοπληροφορική, η οποία είναι το πεδίο που εφαρμόζει υπολογιστικές τεχνικές για την ανάλυση βιολογικών δεδομένων.**
- **Με την αυξανόμενη διαθεσιμότητα δεδομένων υψηλής απόδοσης από διάφορα βιολογικά πειράματα, η στατιστική ανάλυση έχει γίνει ένα θεμελιώδες εργαλείο για την κατανόηση βιολογικών φαινομένων και την πραγματοποίηση ανακαλύψεων.**

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

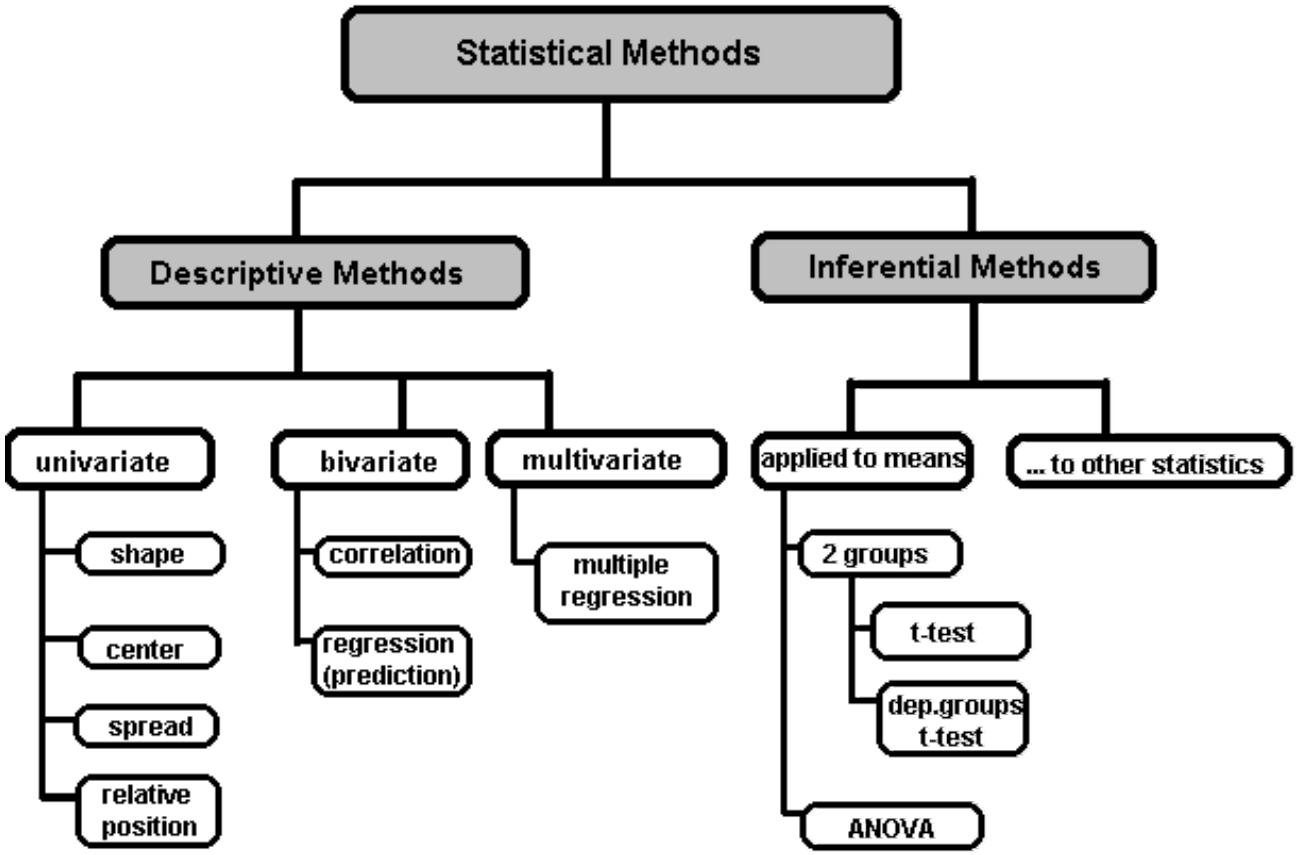
Βασικά στοιχεία της Στατιστικής

Ορισμός : Επιστήμη συλλογής, παρουσίασης, ανάλυσης και λογικής ερμηνείας δεδομένων.

Η στατιστική παρουσιάζει μια αυστηρή επιστημονική μέθοδο για την απόκτηση εικόνας στα δεδομένα. Για παράδειγμα, ας υποθέσουμε ότι μετράμε το βάρος 100 ασθενών σε μια μελέτη. Με τόσες πολλές μετρήσεις, η απλή εξέταση των δεδομένων αποτυγχάνει να παρέχει έναν ενημερωτικό λογαριασμό. Ωστόσο, τα στατιστικά στοιχεία μπορούν να δώσουν μια στιγμιαία συνολική εικόνα των δεδομένων με βάση τη γραφική παρουσίαση ή την αριθμητική περίληψη, ανεξάρτητα από τον αριθμό των σημείων δεδομένων. Εκτός από τη σύνοψη των δεδομένων, ένα άλλο σημαντικό καθήκον της στατιστικής είναι η εξαγωγή συμπερασμάτων και η πρόβλεψη των σχέσεων των μεταβλητών.

Co-funded by the Erasmus+ Programme of the European Union

Μια Ταξονομία της Στατιστικής



Στατιστική Περιγραφή Δεδομένων

- Η στατιστική περιγράφει ένα αριθμητικό σύνολο δεδομένων από το
  - **Κέντρο**
  - **Μεταβλητότητα**
  - **Σχήμα**
- Η στατιστική περιγράφει ένα κατηγορηματικό σύνολο δεδομένων από
  - Συχνότητα, ποσοστό ή αναλογία κάθε κατηγορίας

Μερικοί ορισμοί

Μεταβλητή - οποιοδήποτε χαρακτηριστικό ενός ατόμου ή μιας οντότητας. Μια μεταβλητή μπορεί να λάβει διαφορετικές τιμές για διαφορετικά άτομα. Οι μεταβλητές μπορεί να είναι *κατηγορικές* ή *ποσοτικές* .

- **Ονομαστική** - Κατηγορικές μεταβλητές χωρίς εγγενή σειρά ή σειρά κατάταξης, όπως ονόματα ή κλάσεις (π.χ. φύλο). Η τιμή μπορεί να είναι αριθμητική, αλλά χωρίς αριθμητική τιμή (π.χ., I, II, III). Η μόνη λειτουργία που μπορεί να εφαρμοστεί στις Ονομαστικές μεταβλητές είναι η απαρίθμηση.
- **Τακτική** - Μεταβλητές με εγγενή κατάταξη ή τάξη, π.χ. ήπια, μέτρια, σοβαρή. Μπορεί να συγκριθεί για ισότητα, ή μεγαλύτερη ή μικρότερη, αλλά όχι *πόσο* μεγαλύτερη ή μικρότερη.
- **Διάστημα** - Οι τιμές της μεταβλητής ταξινομούνται όπως στην Τακτική, και επιπλέον, οι διαφορές μεταξύ των τιμών είναι σημαντικές, ωστόσο, η κλίμακα δεν είναι απολύτως αγκυρωμένη. Οι ημερολογιακές ημερομηνίες και οι θερμοκρασίες στην κλίμακα Φαρενάιτ είναι παραδείγματα. Η πρόσθεση και η αφαίρεση, αλλά όχι ο πολλαπλασιασμός και η διαίρεση είναι πράξεις με νόημα.
- **Ratio** - Μεταβλητές με όλες τις ιδιότητες του Interval συν ένα απόλυτο, μη αυθαίρετο σημείο μηδέν, π.χ. ηλικία, βάρος, θερμοκρασία (Kelvin). Η πρόσθεση, η αφαίρεση, ο πολλαπλασιασμός και η διαίρεση είναι όλες πράξεις με νόημα.

Μερικοί ορισμοί

Κατανομή - (μιας μεταβλητής) μας λέει ποιες τιμές παίρνει η μεταβλητή και πόσο συχνά παίρνει αυτές τις τιμές.

- Unimodal - έχοντας μια ενιαία κορυφή
- Bimodal - έχοντας δύο διακριτές κορυφές
- Συμμετρικό - το αριστερό και το δεξί μισό είναι κατοπτρικά είδωλα.

Κατανομή συχνότητας

Εξετάστε ένα σύνολο δεδομένων 26 παιδιών ηλικίας 1-6 ετών. Στη συνέχεια, η κατανομή συχνότητας της μεταβλητής «ηλικίας» μπορεί να καταγραφεί ως εξής:

Συχνότητα Κατανομής Ηλικίας

Ηλικία	1	2	3	4	5	6
Συχνότητα	5	3	7	5	4	2

Ομαδοποιημένη κατανομή συχνότητας ηλικιών:

Ηλικιακή ομάδα	1-2	3-4	5-6
Συχνότητα	8	12	6

Αθροιστική Συχνότητα

Αθροιστική συχνότητα δεδομένων στην προηγούμενη σελίδα

Ηλικία	1	2	3	4	5	6
Συχνότητα	5	3	7	5	4	2
Αθροιστική Συχνότητα	5	8	15	20	24	26

Ηλικιακή ομάδα	1-2	3-4	5-6
Συχνότητα	8	12	6
Αθροιστική Συχνότητα	8	20	26

Co-funded by the Erasmus+ Programme of the European Union

Παρουσίαση Δεδομένων

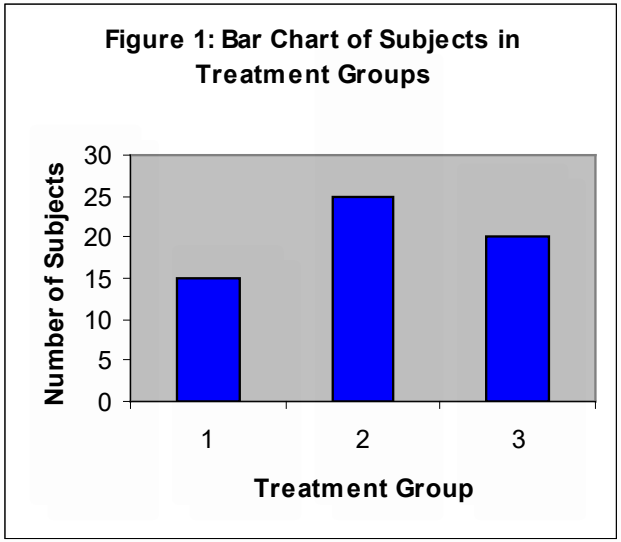
Δύο είδη στατιστικής παρουσίασης δεδομένων - γραφική και αριθμητική .

Γραφική παρουσίαση : Αναζητούμε το συνολικό μοτίβο και για εντυπωσιακές αποκλίσεις από αυτό το μοτίβο. Πάνω από όλα τα μοτίβα συνήθως περιγράφονται από το σχήμα, το κέντρο και την εξάπλωση των δεδομένων. Μια μεμονωμένη τιμή που πέφτει έξω από το συνολικό μοτίβο ονομάζεται ***ακραία τιμή*** .
Το διάγραμμα ράβδων και τα γραφήματα πίτας χρησιμοποιούνται για κατηγορικές μεταβλητές.

Το ιστόγραμμα, το στέλεχος και το φύλλο και το Box-plot χρησιμοποιούνται για την αριθμητική μεταβλητή.

Παρουσίαση Δεδομένων – Κατηγορική Μεταβλητή

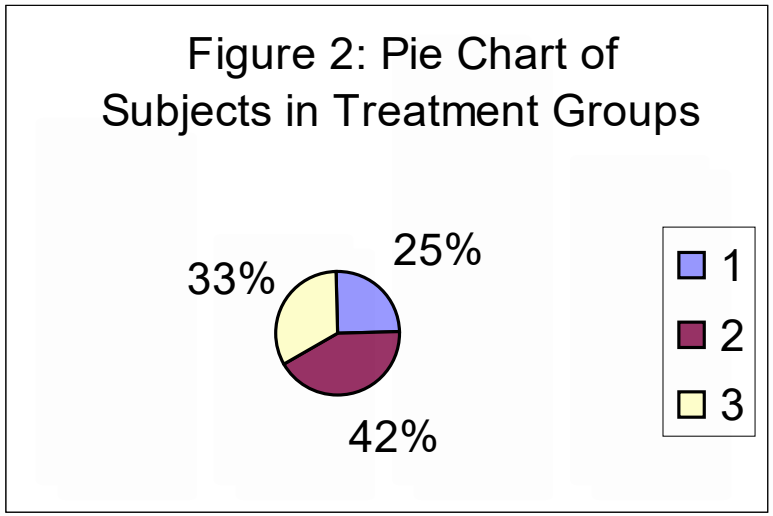
Διάγραμμα ράβδων : Παραθέτει τις κατηγορίες και παρουσιάζει το ποσοστό ή τον αριθμό των ατόμων που εμπίπτουν σε κάθε κατηγορία.



Θεραπεία Ομάδα	Συχνότητα	Ποσοστό	Τοις εκατό (%)
1	15	(15/60)=0,25	25.0
2	25	(25/60)=0,333	41.7
3	20	(20/60)=0,417	33.3
Σύνολο	60	1.00	100

Παρουσίαση Δεδομένων – Κατηγορική Μεταβλητή

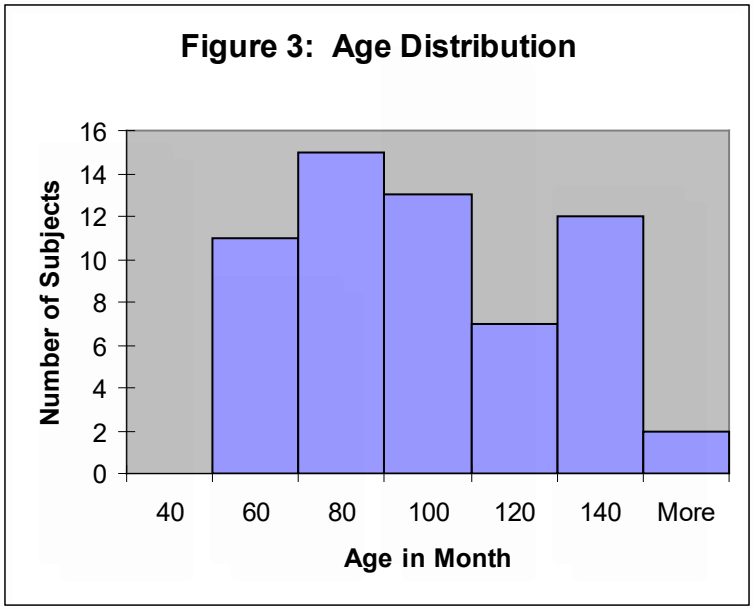
Διάγραμμα πίτας : Παραθέτει τις κατηγορίες και παρουσιάζει το ποσοστό ή τον αριθμό των ατόμων που εμπίπτουν σε κάθε κατηγορία.



Θεραπεία Ομάδα	Συχνότητα	Ποσοστό	Τοις εκατό (%)
1	15	(15/60)=0,25	25.0
2	25	(25/60)=0,333	41.7
3	20	(20/60)=0,417	33.3
Σύνολο	60	1.00	100

Γραφική παρουσίαση – Αριθμητική μεταβλητή

Ιστόγραμμα : Το συνολικό μοτίβο μπορεί να περιγραφεί από το **σχήμα** , **το κέντρο** και **την εξάπλωσή του** . Η ακόλουθη ηλικιακή κατανομή είναι **σωστά λοξή** . Το **κέντρο** βρίσκεται μεταξύ **80 και 100** . **Χωρίς ακραίες τιμές** .



Μέσος	90,41666667
Τυπικό σφάλμα	3,902649518
Διάμεσος	84
Τρόπος	84
Τυπική απόκλιση	30,22979318
Δείγμα Διακύμανσης	913.8403955
Κύρτωση	-1,183899591
Λοξοτροπία	0,389872725
Εύρος	95
Ελάχιστο	48
Ανώτατο όριο	143
Αθροισμα	5425
Πλήθος	60

Αριθμητική Παρουσίαση

Μια θεμελιώδης έννοια στις συνοπτικές στατιστικές είναι αυτή της *κεντρικής τιμής* για ένα σύνολο παρατηρήσεων και ο βαθμός στον οποίο η κεντρική τιμή χαρακτηρίζει ολόκληρο το σύνολο δεδομένων. Οι μετρήσεις της κεντρικής τιμής, όπως ο μέσος όρος ή η διάμεσος, πρέπει να συνδυάζονται με μέτρα διασποράς δεδομένων (π.χ. μέση απόσταση από τη μέση τιμή) για να δείξουν πόσο καλά η κεντρική τιμή χαρακτηρίζει τα δεδομένα στο σύνολό τους.

Για να κατανοήσουμε πόσο καλά μια κεντρική τιμή χαρακτηρίζει ένα σύνολο παρατηρήσεων, ας εξετάσουμε τα ακόλουθα δύο σύνολα δεδομένων:

A: 30, 50, 70
B: 40, 50, 60

Ο μέσος όρος και των δύο συνόλων δεδομένων είναι 50. Όμως, η απόσταση των παρατηρήσεων από τον μέσο όρο στο σύνολο δεδομένων Α είναι μεγαλύτερη από ό,τι στο σύνολο δεδομένων Β. Έτσι, ο μέσος όρος του συνόλου δεδομένων Β είναι μια καλύτερη αναπαράσταση του συνόλου δεδομένων από είναι η περίπτωση για το σύνολο Α.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Περιγραφική Στατιστική
Μέθοδοι Κεντρικής Μέτρησης

Η κεντρική μέτρηση είναι μια συνοπτική μέτρηση του συνολικού επιπέδου ενός συνόλου δεδομένων

Οι μέθοδοι που χρησιμοποιούνται συνήθως είναι **ο μέσος όρος** , **ο διάμεσος** , **ο τρόπος** , **ο γεωμετρικός μέσος όρος** κ.λπ.

Μέσος όρος : Αθροίζοντας όλη την παρατήρηση και διαιρώντας με τον αριθμό των παρατηρήσεων. Ο μέσος όρος των 20, 30, 40 είναι (20+30+40)/3 = 30.

Notation : Let $x_1, x_2, \ldots x_n$ are n observations of a variable x. Then the mean of this variable,

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

Περιγραφική Στατιστική
Μέθοδοι Κεντρικής Μέτρησης

Διάμεσος : Η μεσαία τιμή σε μια διατεταγμένη ακολουθία παρατηρήσεων. Δηλαδή, για να βρούμε τη διάμεσο πρέπει να παραγγείλουμε το σύνολο δεδομένων και μετά να βρούμε τη μεσαία τιμή. Σε περίπτωση ζυγού αριθμού παρατηρήσεων, ο μέσος όρος των δύο μεσαίων περισσότερων τιμών είναι η διάμεσος. Για παράδειγμα, για να βρούμε τη διάμεσο του {9, 3, 6, 7, 5}, ταξινομούμε πρώτα τα δεδομένα δίνοντας {3, 5, 6, 7, 9} και μετά επιλέγουμε τη μεσαία τιμή 6. Εάν ο αριθμός των παρατηρήσεων είναι άρτιος, π.χ., {9, 3, 6, 7, 5, 2}, τότε η διάμεσος είναι ο μέσος όρος των δύο μεσαίων τιμών από την ταξινομημένη ακολουθία, σε αυτήν την περίπτωση, (5 + 6) / 2 = 5,5.

Λειτουργία : Η τιμή που παρατηρείται πιο συχνά. Ο τρόπος λειτουργίας είναι απροσδιόριστος για ακολουθίες στις οποίες δεν επαναλαμβάνεται καμία παρατήρηση.

Περιγραφική Στατιστική
Μέση ή Διάμεσος

Ο διάμεσος είναι λιγότερο ευαίσθητος σε ακραίες τιμές (ακραίες βαθμολογίες) από τον μέσο όρο και επομένως καλύτερο μέτρο από το μέσο όρο για πολύ λοξές κατανομές, π.χ. οικογενειακό εισόδημα.

Για παράδειγμα, ο μέσος όρος των 20, 30, 40 και 990 είναι (20+30+40+990)/4 =270. Η διάμεσος αυτών των τεσσάρων παρατηρήσεων είναι (30+40)/2 =35. Εδώ 3 από τις 4 παρατηρήσεις βρίσκονται μεταξύ 20-40.

Έτσι, ο μέσος όρος 270 αποτυγχάνει πραγματικά να δώσει μια ρεαλιστική εικόνα του μεγαλύτερου μέρους των δεδομένων. Επηρεάζεται από την ακραία τιμή 990.

Περιγραφική Στατιστική
Μέθοδοι Μέτρησης Μεταβλητότητας

Η μεταβλητότητα (ή η διασπορά) μετρά την ποσότητα της διασποράς σε ένα σύνολο δεδομένων.

Μέθοδοι που χρησιμοποιούνται συνήθως : *εύρος* , *διακύμανση* , *τυπική απόκλιση* , *διατεταρτημόριο εύρος* , *συντελεστής διακύμανσης* κ.λπ.

Εύρος : Η διαφορά μεταξύ των μεγαλύτερων και των μικρότερων παρατηρήσεων. Το εύρος των 10, 5, 2, 100 είναι (100-2)=98. Είναι ένα χονδροειδές μέτρο μεταβλητότητας.

Περιγραφική Στατιστική
Μέθοδοι Μέτρησης Μεταβλητότητας

Διακύμανση : Η διακύμανση ενός συνόλου παρατηρήσεων είναι ο μέσος όρος των τετραγώνων των αποκλίσεων των παρατηρήσεων από τον μέσο όρο τους. Στα σύμβολα, η διακύμανση των n παρατηρήσεων $x_1, x_2, \ldots x_n$ είναι

$$S^2 = \frac{(x_1 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n-1}$$

Διακύμανση 5, 7, 3; Ο μέσος όρος είναι (5+7+3)/3 = 5 και η διακύμανση είναι

$$\frac{(5-5)^2 + (3-5)^2 + (7-5)^2}{3-1} = 4$$

Τυπική απόκλιση : Τετραγωνική ρίζα της διακύμανσης. Η τυπική απόκλιση του παραπάνω παραδείγματος είναι 2.

Διαφορές ή Σχέσεις;

Εξετάζω διαφορές μεταξύ των ομάδων

↓

T-test, ANOVA

Εξετάζω σχέσεις μεταξύ μεταβλητών

↓

Συσχέτιση, Ανάλυση Παλινδρόμησης

Παράδειγμα συσχέτισης

Υπάρχει συσχέτιση μεταξύ:

- IQ παιδιών και IQ γονέων
- Βαθμός κοινωνικής εμπιστοσύνης και αριθμός μελών σε εθελοντικό σωματείο ;
- Αστική ανάπτυξη και παραβιάσεις της ποιότητας του αέρα;
- Χρηματοδότηση GRA και αριθμός δημοσιεύσεων από Ph.D. Φοιτητές
- Αριθμός περιπολίας αστυνομίας και αριθμός εγκλημάτων
- Βαθμός στις εξετάσεις και ώρα στις εξετάσεις

Συσχέτιση

- Συντελεστής συσχέτισης: στατιστικός δείκτης του βαθμού στον οποίο συσχετίζονται ή σχετίζονται δύο μεταβλητές.

- Μπορούμε να προσδιορίσουμε εάν μια μεταβλητή σχετίζεται με μια άλλη βλέποντας εάν βαθμολογείται στις δύο μεταβλητές εάν διαφέρουν μαζί .

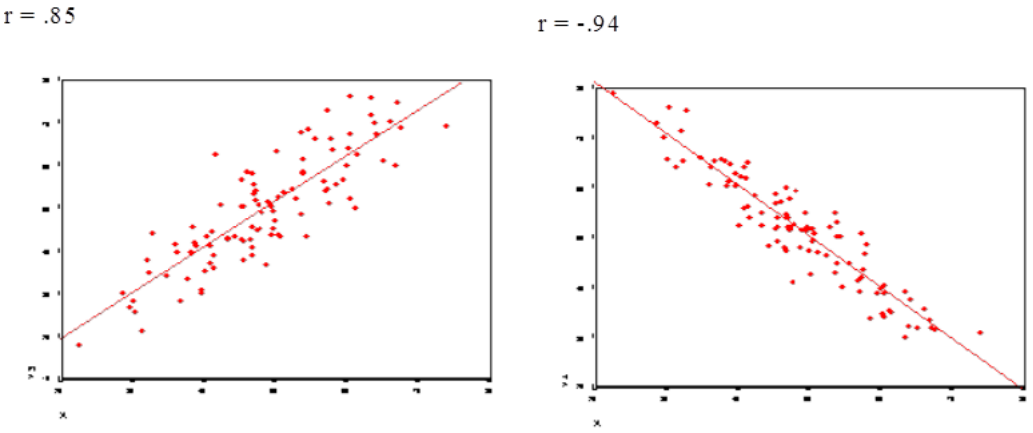
Διάγραμμα διασποράς

- Η σχέση μεταξύ οποιωνδήποτε δύο μεταβλητών μπορεί να απεικονιστεί γραφικά σε έναν άξονα x και y.
- Κάθε θέμα i_1 έχει ($x_{1,}$ y_1). Όταν σχεδιάζεται η βαθμολογία s για ένα ολόκληρο δείγμα, το αποτέλεσμα ονομάζεται **διάγραμμα διασποράς** .

Κατεύθυνση της σχέσης

Οι μεταβλητές μπορούν να συσχετιστούν θετικά ή αρνητικά.

- Θετική συσχέτιση : Μια τιμή μιας μεταβλητής αύξησης, τιμή μιας άλλης μεταβλητής αύξησης.
- Αρνητική συσχέτιση : Μια τιμή μιας μεταβλητής αύξησης, τιμή μιας άλλης μεταβλητής μείωσης.



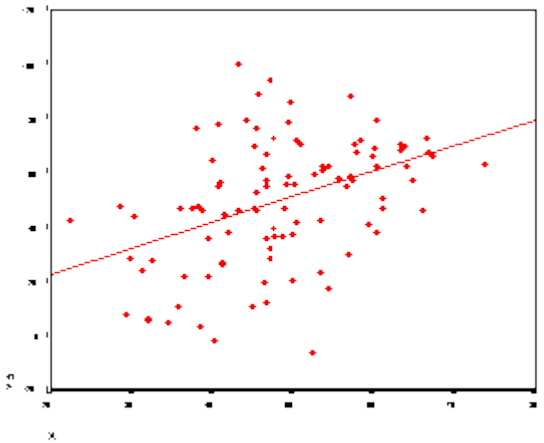
Δύναμη της σχέσης

Το μέγεθος της συσχέτισης :

- Υποδεικνύεται από την αριθμητική του τιμή
- αγνοώντας το σημάδι
- εκφράζει την ισχύ της γραμμικής σχέσης μεταξύ των μεταβλητών.

r = 1,00

r = .42



r =.85

r =.17

Ανάλυση Γραμμικής Παλινδρόμησης

- Η ανάλυση παλινδρόμησης χρησιμοποιείται για την πρόβλεψη της τιμής μιας μεταβλητής (η ***εξαρτημένη μεταβλητή***) με βάση άλλες μεταβλητές (οι ***ανεξάρτητες μεταβλητές***).

- Εξαρτημένη μεταβλητή: συμβολίζεται με **Y**
- Ανεξάρτητες μεταβλητές: συμβολίζονται $X_1, X_2, \ldots, X_k$
- **Αν έχουμε μόνο ΜΙΑ ανεξάρτητη μεταβλητή, το μοντέλο είναι**

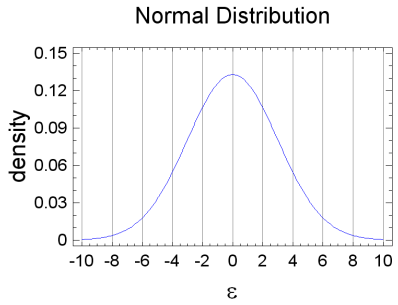
$$y = \beta_0 + \beta_1 x + \varepsilon$$

- η οποία αναφέρεται ως απλή γραμμική παλινδρόμηση. Θα μας ενδιέφερε να εκτιμήσουμε τα β_0 και β_1 από τα δεδομένα που συλλέγουμε.

Ανάλυση Γραμμικής Παλινδρόμησης

$$y = \beta_0 + \beta_1 x + \varepsilon$$

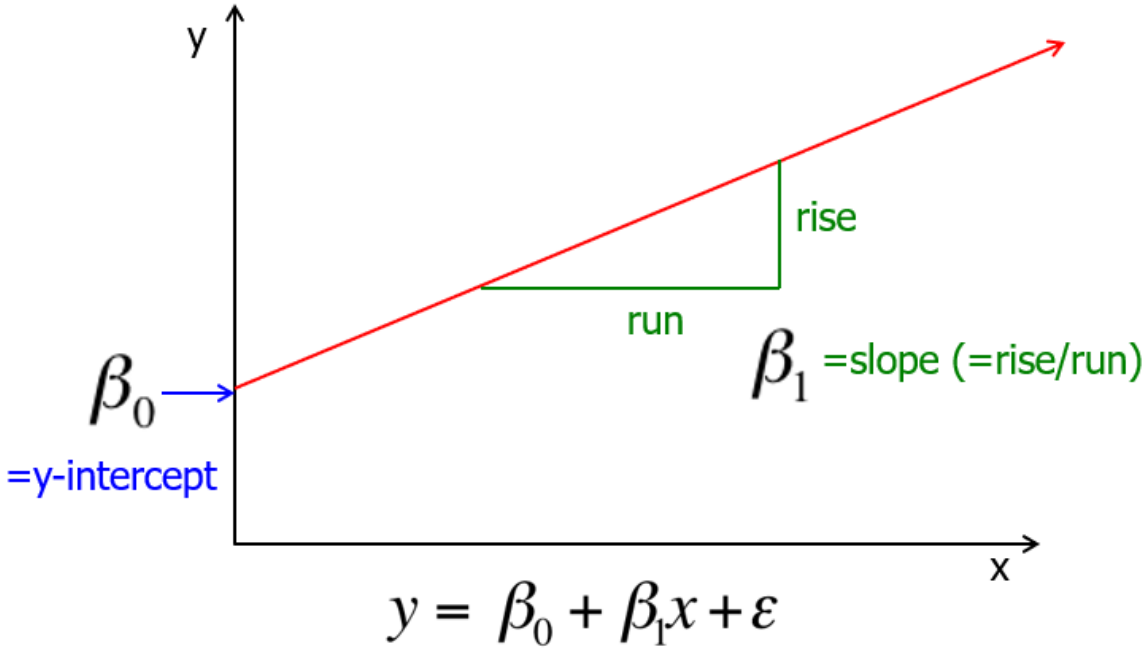
- Μεταβλητές:
- X = Ανεξάρτητη Μεταβλητή (παρέχουμε αυτό)
- Y = Εξαρτημένη μεταβλητή (παρατηρούμε αυτό)
- Παράμετροι:
- β_0 = Y-Intercept
- β_1 = Κλίση
- ε ~ Κανονική Τυχαία Μεταβλητή ($\mu_\varepsilon = 0$, $\sigma_\varepsilon = ???$) [Θόρυβος]



Απλό μοντέλο γραμμικής παλινδρόμησης

Meaning of β_0 and β_1

$\beta_1 > 0$ [positive slope] $\beta_1 < 0$ [negative slope]



Πολλαπλή παλινδρόμηση

- Η πολλαπλή παλινδρόμηση είναι μια επέκταση μιας απλής γραμμικής παλινδρόμησης.
- Στην πολλαπλή παλινδρόμηση, μια εξαρτημένη μεταβλητή προβλέπεται από περισσότερες από μία ανεξάρτητες μεταβλητές
- $Y = a + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$

Τύποι Στατιστικού Λογισμικού

Υπάρχουν πολλά λογισμικά για την εκτέλεση στατιστικής ανάλυσης και οπτικοποίησης δεδομένων. Κάποιοι από αυτούς είναι:

SAS (System for Statistical Analysis);

R;

MATLAB;

Minitab;

Stata;

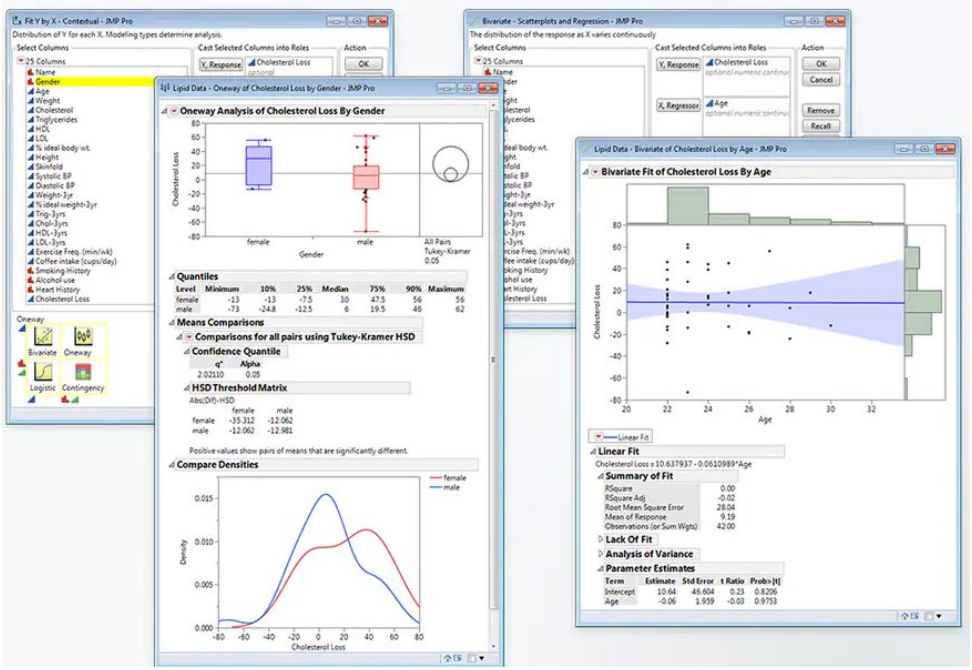
SPSS;

GraphPad Prism;

Στατιστική;

MS Excel:

…και τα λοιπά.



συμπέρασμα

- Η στατιστική ανάλυση είναι ένα ισχυρό εργαλείο για την κατανόηση δεδομένων και τη λήψη τεκμηριωμένων αποφάσεων.
- Χρησιμοποιείται σε πολλά πεδία και μπορεί να εφαρμοστεί σε μεγάλη ποικιλία τύπων δεδομένων και καταστάσεων.
- Κατανοώντας τις βασικές έννοιες της στατιστικής ανάλυσης, μπορείτε να ερμηνεύσετε και να αναλύσετε καλύτερα τα δεδομένα και να λάβετε πιο τεκμηριωμένες αποφάσεις.
- Η στατιστική ανάλυση είναι ένα κρίσιμο συστατικό της έρευνας στη βιοπληροφορική, που βοηθά στην εξαγωγή ουσιαστικών γνώσεων από μεγάλα και πολύπλοκα βιολογικά σύνολα δεδομένων.

Co-funded by the Erasmus+ Programme of the European Union

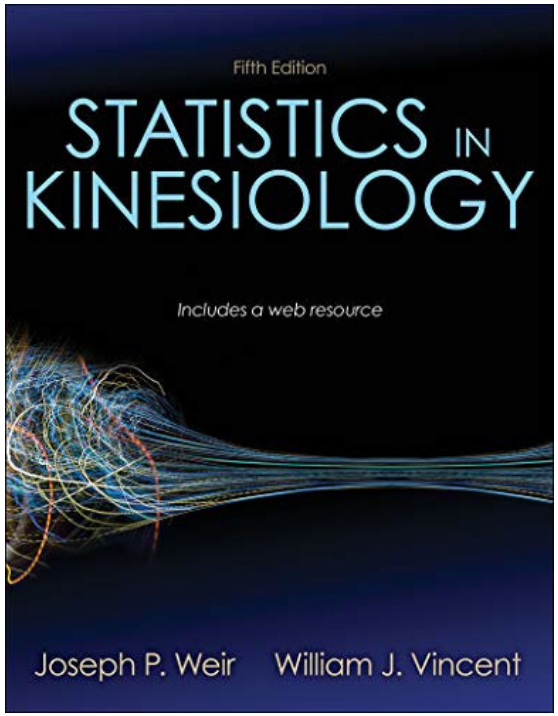
Ερωτήσεις και ασκήσεις:

1. Τι είναι η στατιστική ανάλυση;

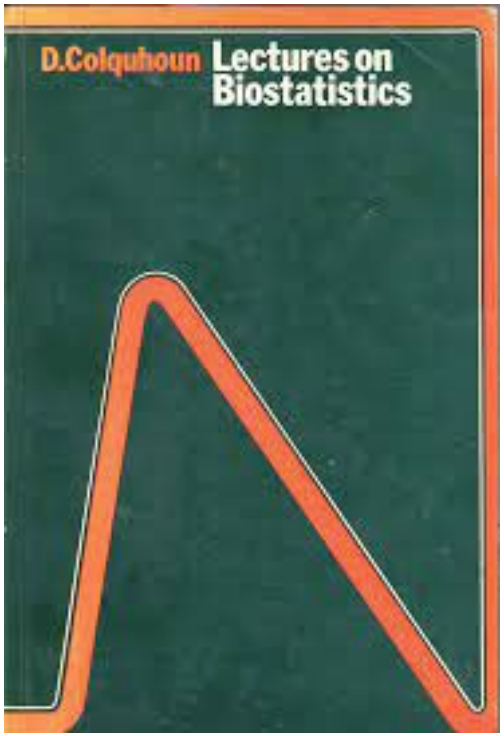
2. Ποιοι είναι οι δύο κύριοι τύποι δεδομένων;

3. Πώς μπορούν να χρησιμοποιηθούν περιγραφικές στατιστικές για τη σύνοψη δεδομένων;

4. Τι είναι η συσχέτιση και πώς χρησιμοποιείται στη στατιστική ανάλυση;



P. Weir και William J. Vincent. (2020). Statistics in Kinesiology, 5th Edition, Human Kinetics..



Colquhoun, D. (1971). Διαλέξεις για βιοστατιστική; μια εισαγωγή στη στατιστική με εφαρμογές στη βιολογία και την ιατρική. Οξφόρδη: Clarendon Press.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 4. Μαθήματα Επιλογής

Μάθημα 6. Μοριακή Μοντελοποίηση Υπολογιστή

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Κβαντική μηχανική
- Μοριακή Μηχανική
- Πεδία Δύναμης
- Διαδικασίες Ελαχιστοποίησης Ενέργειας
- Διαμορφωτική Ανάλυση ενός Πεπτιδίου: Πρόβλημα πολλαπλών ελαχίστων
- Σχέδιο

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Ο ορισμός που είναι επί του παρόντος αποδεκτός για το τι είναι η μοριακή μοντελοποίηση μπορεί να δηλωθεί ως εξής: «μοριακή μοντελοποίηση είναι οτιδήποτε απαιτεί τη χρήση υπολογιστή για τη ζωγραφική, περιγραφή ή αξιολόγηση οποιασδήποτε πτυχής των ιδιοτήτων της δομής ενός μορίου» (Pensak, 1989). Οι μέθοδοι που χρησιμοποιούνται στον χώρο της μοριακής μοντελοποίησης αφορούν την αυτόματη δημιουργία δομών, την ανάλυση τρισδιάστατων (3D) βάσεων δεδομένων, την κατασκευή πρωτεϊνικών μοντέλων με τεχνικές που βασίζονται στην ομολογία αλληλουχίας, την ανάλυση ποικιλομορφίας, τη σύνδεση προσδεμάτων ή τις μεθόδους συνεχούς.

Co-funded by the Erasmus+ Programme of the European Union

Κβαντική μηχανική

Υπάρχουν δύο διαφορετικές προσεγγίσεις για τον υπολογισμό της ενέργειας ενός μορίου. Πρώτον, η κβαντική μηχανική, μια διαδικασία που βασίζεται στις πρώτες αρχές. Σε αυτή την προσέγγιση, οι πυρήνες είναι διατεταγμένοι στο χώρο και τα αντίστοιχα ηλεκτρόνια απλώνονται σε όλο το σύστημα σε μια συνεχή ηλεκτρονική πυκνότητα και υπολογίζονται λύνοντας την εξίσωση Schroedinger . Για τα βιομόρια αυτή η διαδικασία μπορεί να γίνει εντός της προσέγγισης Born-Oppenheimer και για τους περισσότερους σκοπούς το αυτοσυνεπές πεδίο Hartree-Fock είναι η καταλληλότερη διαδικασία για τον υπολογισμό της ηλεκτρονικής πυκνότητας και της ενέργειας του συστήματος.

Μοριακή Μηχανική

Όταν οι χημικές αντιδράσεις δεν χρειάζεται να προσομοιωθούν, η κλασική μηχανική μπορεί να περιγράψει τη συμπεριφορά ενός βιομοριακού συστήματος. Αυτό το μαθηματικό μοντέλο είναι γνωστό ως μοριακή μηχανική και μπορεί να χρησιμοποιηθεί για τον υπολογισμό της ενέργειας συστημάτων που περιέχουν μεγάλο αριθμό ατόμων, όπως μόρια ή πολύπλοκα συστήματα βιοχημικού και βιοϊατρικού ενδιαφέροντος. Σε αντίθεση με την κβαντική μηχανική, η μοριακή μηχανική αγνοεί τα ηλεκτρόνια και υπολογίζει την ενέργεια ενός συστήματος μόνο ως συνάρτηση των πυρηνικών θέσεων. Στη συνέχεια, είναι δυνατό να ληφθεί υπόψη με σιωπηρό τρόπο το ηλεκτρονικό στοιχείο του συστήματος με επαρκή παραμετροποίηση της συνάρτησης δυναμικής ενέργειας. Το σύνολο των εξισώσεων και των παραμέτρων που ορίζουν τη δυναμική επιφάνεια ενός μορίου ονομάζεται πεδίο δύναμης.

Co-funded by the Erasmus+ Programme of the European Union

Πεδία Δύναμης

Στη μοριακή μηχανική τα ηλεκτρόνια και οι πυρήνες των ατόμων δεν περιλαμβάνονται ρητά στους υπολογισμούς. Η μοριακή μηχανική θεωρεί ένα μόριο ως μια συλλογή μαζών που αλληλεπιδρούν η μία με την άλλη μέσω αρμονικών δυνάμεων. Έτσι, τα άτομα στα μόρια αντιμετωπίζονται ως μπάλα διαφορετικών μεγεθών και γεύσεων που ενώνονται μεταξύ τους με ελατήρια μεταβλητής ισχύος και αποστάσεων ισορροπίας (δεσμοί). Αυτή η απλοποίηση επιτρέπει τη χρήση της μοριακής μηχανικής ως ένα γρήγορο υπολογιστικό μοντέλο που μπορεί να εφαρμοστεί σε μόρια οποιουδήποτε μεγέθους.

Co-funded by the Erasmus+ Programme of the European Union

Πεδία Δύναμης

Κατά τη διάρκεια ενός υπολογισμού, η συνολική ενέργεια ελαχιστοποιείται σε σχέση με τις ατομικές συντεταγμένες και αποτελείται από ένα άθροισμα διαφορετικών συνεισφορών που υπολογίζουν τις αποκλίσεις από την ισορροπία των μηκών, των γωνιών και των στρέψεων των δεσμών συν τις μη δεσμευμένες αλληλεπιδράσεις:

$$E_{tot} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{elec} + \dots$$

όπου E_{tot} είναι η συνολική ενέργεια του μορίου, E_{str} είναι ο όρος ενέργειας τάνυσης δεσμού, E_{bend} είναι ο όρος ενέργειας κάμψης γωνίας, E_{tors} είναι ο όρος στρεπτικής ενέργειας, E_{vdw} είναι ο ενεργειακός όρος van der Waals και Elec είναι $_o$ όρος ηλεκτροστατικής ενέργειας.

Co-funded by the Erasmus+ Programme of the European Union

Διαδικασίες Ελαχιστοποίησης Ενέργειας

Οι μέθοδοι ελαχιστοποίησης ενέργειας μπορούν να χωριστούν σε διαφορετικές κατηγορίες ανάλογα με τη σειρά του παραγώγου που χρησιμοποιείται για τον εντοπισμό ενός ελάχιστου στην ενεργειακή επιφάνεια. Οι μέθοδοι μηδενικής τάξης είναι εκείνες που χρησιμοποιούν μόνο τη συνάρτηση ενέργειας για τον εντοπισμό περιοχών χαμηλής ενέργειας μέσω μιας διαδικασίας αναζήτησης δικτύου. Η πιο γνωστή μέθοδος αυτού του είδους είναι η μέθοδος SIMPLEX. Στις τεχνικές πρώτης παραγώγου, υπάρχουν αρκετές διαδικασίες όπως η μέθοδος πιο απότομης κατάβασης ή η μέθοδος συζυγούς κλίσης που χρησιμοποιούν τη διαβάθμιση της συνάρτησης. Μέθοδοι δεύτερης παραγώγου, όπως ο αλγόριθμος Newton-Raphson, χρησιμοποιούν το hessian για να εντοπίσουν ελάχιστα. Στην παρούσα μελέτη έχουν χρησιμοποιηθεί μόνο μέθοδοι πρώτης παραγώγου και θα περιγραφούν εν συντομία.

Μέθοδος απότομης κατάβασης

Στη μέθοδο της πιο απότομης καθόδου, ο ελαχιστοποιητής υπολογίζει αριθμητικά την πρώτη παράγωγο της συνάρτησης ενέργειας για να βρει ένα ελάχιστο. Η ενέργεια υπολογίζεται για την αρχική γεωμετρία και μετά πάλι αφού ένα από τα άτομα έχει μετακινηθεί σε μια μικρή αύξηση σε μία από τις κατευθύνσεις του συστήματος συντεταγμένων. Αυτή η διαδικασία επαναλαμβάνεται για όλα τα άτομα που τελικά μετακινούνται σε νέα θέση κατηφορικά στην ενεργειακή επιφάνεια. Η διαδικασία σταματά όταν εκπληρωθεί μια προκαθορισμένη συνθήκη κατωφλίου. Η διαδικασία βελτιστοποίησης είναι αργή κοντά στο ελάχιστο, και κατά συνέπεια, η μέθοδος της πιο απότομης καθόδου χρησιμοποιείται συχνά για κατασκευές μακριά από την ελάχιστη ως πρώτη, πρόχειρη και εισαγωγική εκτέλεση που ακολουθείται από μια επακόλουθη ελαχιστοποίηση χρησιμοποιώντας έναν πιο προηγμένο αλγόριθμο όπως η συζευγμένη κλίση.

Co-funded by the Erasmus+ Programme of the European Union

Μέθοδος συζυγούς κλίσης

Ο αλγόριθμος συζευγμένης κλίσης συσσωρεύει τις πληροφορίες σχετικά με τη συνάρτηση από τη μια επανάληψη στην επόμενη. Με αυτή τη διαδικασία μπορεί να αποφευχθεί το αντίστροφο της προόδου που σημειώθηκε σε προηγούμενη επανάληψη. Για κάθε βήμα ελαχιστοποίησης η κλίση υπολογίζεται και χρησιμοποιείται ως πρόσθετη πληροφορία για τον υπολογισμό του νέου διανύσματος κατεύθυνσης της διαδικασίας ελαχιστοποίησης. Έτσι, κάθε διαδοχικό βήμα βελτιώνει την κατεύθυνση προς το ελάχιστο. Η υπολογιστική προσπάθεια και οι απαιτήσεις αποθήκευσης είναι μεγαλύτερες από ό,τι για την πιο απότομη κάθοδο, αλλά οι συζευγμένες κλίσεις είναι η μέθοδος επιλογής για μεγαλύτερα συστήματα. Το μεγαλύτερο συνολικό υπολογιστικό κόστος και ο μεγαλύτερος χρόνος ανά επανάληψη αντισταθμίζεται περισσότερο από την αποτελεσματικότερη σύγκλιση στο ελάχιστο που επιτυγχάνεται από τις συζευγμένες κλίσεις.

Χρήση γομώσεων και διαλυτών

Οι υπολογισμοί της μοριακής μηχανικής πραγματοποιούνται παραδοσιακά σε συνθήκες κενού ε = 1. Η διερεύνηση μορίων που περιέχουν φορτία και δίπολα απαιτεί ωστόσο την εξέταση των επιδράσεων του διαλύτη. Διαφορετικά οι διαμορφώσεις επηρεάζονται από ισχυρές ηλεκτροστατικές αλληλεπιδράσεις. Τα πεδία δύναμης προσπαθούν να μεγιστοποιήσουν την ελκυστική ηλεκτροστατική αλληλεπίδραση, με αποτέλεσμα ενεργειακά έντονα προτιμώμενες αλλά μη ρεαλιστικές διαμορφώσεις χαμηλής ενέργειας του μορίου. Αυτό μπορεί να αποφευχθεί χρησιμοποιώντας την αντίστοιχη διηλεκτρική σταθερά διαλύτη. Για παράδειγμα, στο νερό το ε είναι 80.

Η ισχύς της ηλεκτροστατικής αλληλεπίδρασης μειώνεται αργά με το r-1. Επομένως, σε ορισμένες περιπτώσεις η διηλεκτρική σταθερά επιλέγεται να εξαρτάται από την απόσταση προκειμένου να μειώνεται ταχύτερα, αποφεύγοντας την ανάγκη να θεωρηθούν τα άτομα μακριά το ένα από το άλλο, προσομοιώνοντας την επίδραση της μετατόπισης των μορίων του διαλύτη κατά την προσέγγιση ενός προσδέματος μόριο σε μια μακρομοριακή επιφάνεια.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Διαμορφωτική Ανάλυση ενός Πεπτιδίου: Πρόβλημα πολλαπλών ελαχίστων

Η διαμορφωτική ανάλυση είναι ο χαρακτηρισμός των δομών που μπορεί να υιοθετήσει ένα μόριο και πώς αυτές επηρεάζουν τις ιδιότητές του. Βασικό συστατικό μιας διαμορφωτικής ανάλυσης είναι η διαμορφωτική αναζήτηση, αντικείμενο της οποίας είναι ο εντοπισμός των προτιμώμενων διαμορφώσεων ενός μορίου, δηλαδή εκείνων των διαμορφώσεων που καθορίζουν τη συμπεριφορά του. Αυτό συνήθως απαιτεί τον χαρακτηρισμό των διαμορφώσεων που είναι ελάχιστες στην επιφάνεια δυναμικής ενέργειας. Για ένα πεπτίδιο, λόγω της υψηλής διαμορφωτικής ευελιξίας του στο διάλυμα, υπάρχει τόσο μεγάλος αριθμός ελάχιστων στην ενεργειακή επιφάνεια που δεν είναι πρακτικό να χαρακτηριστούν όλα. Αυτό είναι γνωστό ως πρόβλημα πολλαπλών ελάχιστων και είναι η κύρια δυσκολία να χαρακτηριστεί δομικά ένα πεπτίδιο. Συγκεκριμένα, τα περισσότερα από τα πεπτίδια υπάρχουν σε φυσιολογικές συνθήκες ως ένα μείγμα εναλλάξιμων διαμορφώσεων με παρόμοιες ενέργειες που κατοικούνται σύμφωνα με την κατανομή Boltzmann.

Αξιολόγηση της Βιοδραστικής Διαμόρφωσης

Οι υπολογιστικές μέθοδοι που χρησιμοποιούνται για την εξερεύνηση ενός πεπτιδικού διαμορφωτικού χώρου αποτελούν μια εναλλακτική διαδικασία σε σχέση με τις πειραματικές προσεγγίσεις για τον προσδιορισμό της βιοδραστικής διαμόρφωση. Οι υπολογιστικές μέθοδοι είναι πολύ χρήσιμες όταν μελετώνται ποικίλες δομές-δραστηριότητα διαφορετικά ανάλογα του άγριου πεπτιδίου είναι διαθέσιμα. Μία από τις βασικές υποθέσεις έμμεσα μέθοδοι είναι ότι η βιοδραστική διαμόρφωση είναι μία από τις θερμοδυναμικά προσβάσιμες διαμορφώσεις του πεπτιδίου. Από ενεργειακή άποψη, είναι μια λογική υπόθεση να θεωρήστε ότι η βιοδραστική διαμόρφωση είναι αυτή που παρουσιάζει υψηλή θερμοδυναμική πιθανότητα να υπάρχουν σε λύση. Όσο πιο πυκνοκατοικημένη είναι αυτή η διαμόρφωση, τόσο μικρότερη είναι η απώλεια εντροπίας διαμόρφωσης θα συμβεί στο βήμα δέσμευσης. Αυτό υποστηρίζεται από το γεγονός ότι είναι διαμορφωτικά περιορισμένοι Τα ανάλογα εμφανίζουν υψηλότερη συγγένεια για τον υποδοχέα από τα γραμμικά πεπτίδια. Αυτό πιθανότατα οφείλεται στην μειωμένο σύνολο διαμορφώσεων που είναι προσβάσιμες σε αυτούς και στη μειωμένη απώλεια εντροπίας στο συγκρότημα σχηματισμός.

Co-funded by the Erasmus+ Programme of the European Union

Folding Studies on Peptides

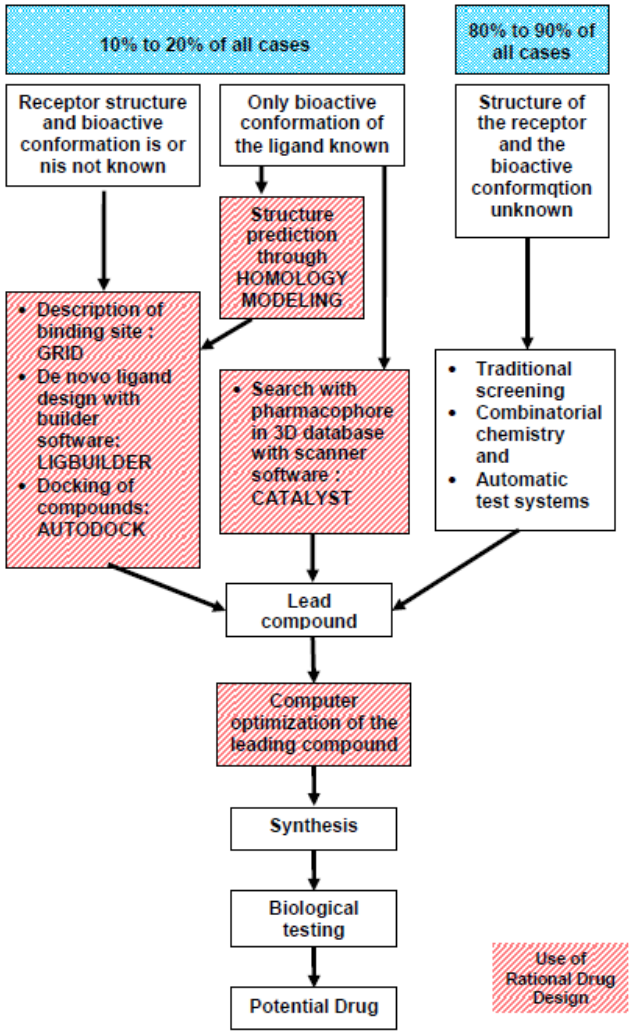
Η συμπεριφορά του συνόλου είναι δύσκολο να αξιολογηθεί. Παραδοσιακά η τάση ήταν η αξιολόγηση προγνωστική επιτυχία χρησιμοποιώντας την απόκλιση ριζικού μέσου τετραγώνου (RMSD) από μια δομή ακτίνων Χ. Αυτό είναι μια συνεχής κλιμακωτή τιμή και έτσι, ο ορισμός των εγγενών και των ξεδιπλωμένων καταστάσεων γίνεται α υποκειμενικό θέμα. Duan et al. (1998) ταξινόμησε τις δομές με έναν αλγόριθμο ομαδοποίησης και χρησιμοποίησε αυτή η ταξινόμηση για να μελετήσει τις μεταβάσεις προκειμένου να παρακολουθήσει την εξέλιξη του αναδιπλούμενου γεγονότος. Η μέθοδος είναι κατάλληλη για την αξιολόγηση της εξέλιξης του διαδικασία αναδίπλωσης και η μελέτη των μεταβάσεων μεταξύ διαφορετικών ομάδων διαμορφώσεων. Επιπλέον, είναι δυνατό να γνωρίζουμε πόσο μακριά έχουμε προχωρήσει στη διαδικασία αναδίπλωσης κοιτάζοντας το ρυθμός εμφάνισης νέων διαμορφώσεων, καθώς όσο μεγαλύτερη είναι η τροχιά τόσο λιγότερες είναι οι νέες διαμορφώσεις εμφανίζομαι. Η μέθοδος είναι κατάλληλη για πεπτίδια και πρωτεΐνες οποιουδήποτε μεγέθους και θα μπορούσε εύκολα να αυτοματοποιηθεί για τη μελέτη παράλληλων διαφορετικών πεπτιδίων ή πρωτεϊνών.

Σχεδιασμός φαρμάκων με τη βοήθεια υπολογιστή

Ο σχεδιασμός φαρμάκων με τη βοήθεια υπολογιστή, που συχνά ονομάζεται σχεδιασμός με βάση τη δομή περιλαμβάνει τη χρήση του βιοχημικές πληροφορίες της αλληλεπίδρασης συνδέτη-υποδοχέα προκειμένου να υποβληθούν οι βελτιώσεις συνδέτη. Για Για παράδειγμα, εάν γνωρίζουμε τη θέση δέσμευσης, η στερική συμπληρωματικότητα του συνδέτη θα μπορούσε να βελτιωθεί αυξάνει τη συγγένεια για τον υποδοχέα του. Πράγματι, χρησιμοποιώντας την κρυσταλλική δομή του συμπλέγματος μπορούμε περιοχές στόχου του συνδέτη που ταιριάζουν ελάχιστα εντός της ενεργού θέσης και υποθέτουν χημικές τροποποιήσεις που μειώνουν το ενεργειακό δυναμικό καθιστώντας πιο αρνητικούς τους όρους Van der Waals, επομένως βελτίωση της συμπληρωματικότητας με τον υποδοχέα. Με παρόμοιο τρόπο, λειτουργικές ομάδες στον συνδέτη μπορεί να αλλάξει προκειμένου να αυξηθεί η ηλεκτροστατική συμπληρωματικότητα με τον υποδοχέα.

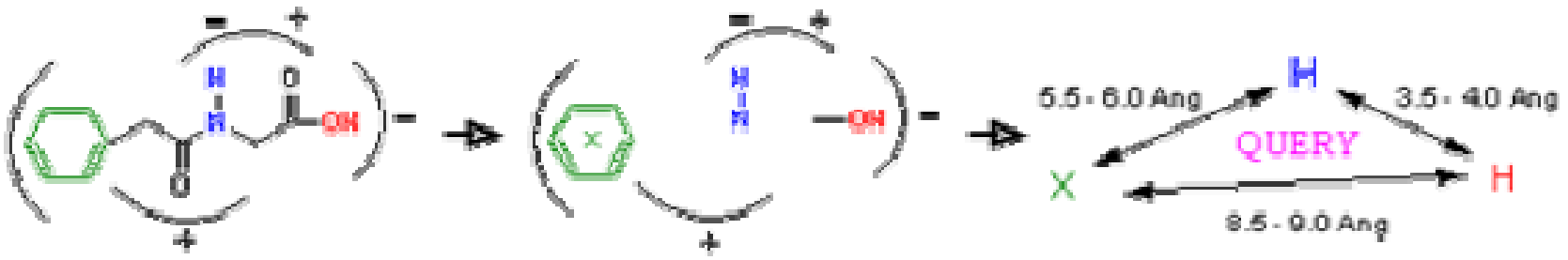
Σχεδιασμός φαρμάκων με τη βοήθεια υπολογιστή

Επισκόπηση διαφορετικών στρατηγικών που χρησιμοποιούνται για την αναζήτηση νέων πιθανών φαρμάκων.



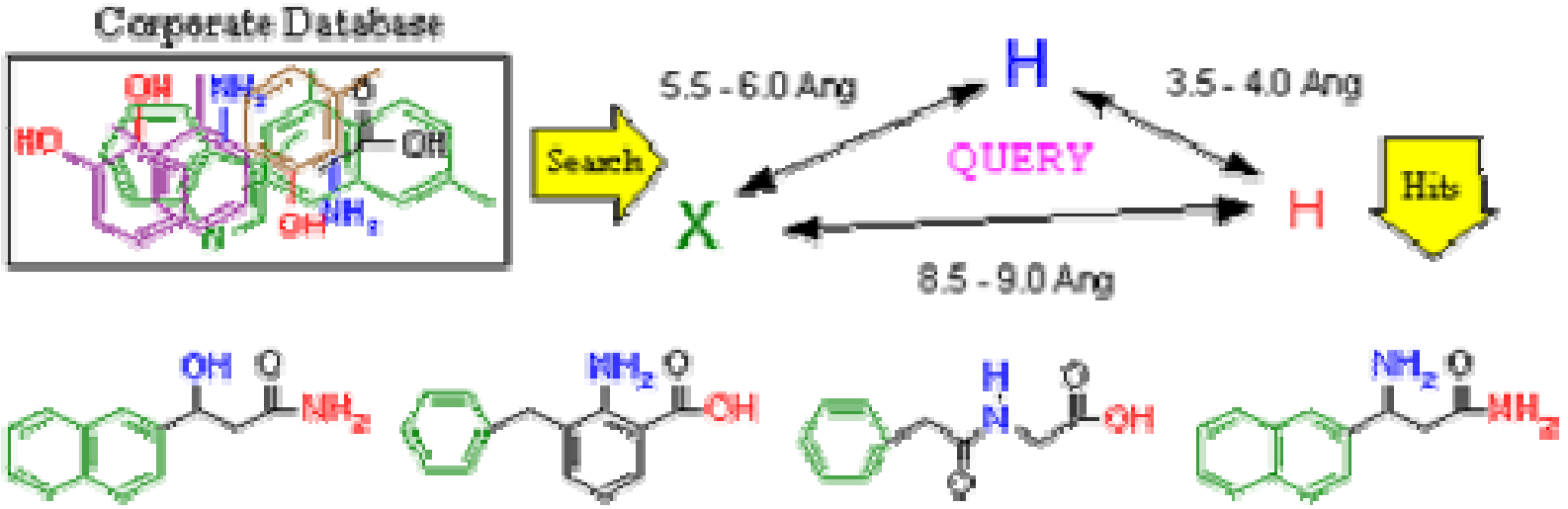
Co-funded by the Erasmus+ Programme of the European Union

Ορισμός φαρμακοφόρου από μια γνωστή βιοδραστική διαμόρφωση



Co-funded by the Erasmus+ Programme of the European Union

Τα αποτελέσματα προέκυψαν σε μια τρισδιάστατη αναζήτηση βάσης δεδομένων



Ερωτήσεις

- Ποιες είναι οι αρχές της κβαντικής μηχανικής και της μοριακής μηχανικής;
- Ποιες είναι οι διαδικασίες ελαχιστοποίησης της ενέργειας;
- Πώς μπορούν να αναλυθούν οι πρωτεϊνικές διαμορφώσεις;
- σχεδιασμό φαρμάκων

Leach A. (2001) Μοριακή μοντελοποίηση: αρχές και εφαρμογές, Pearson Education, Essex, Αγγλία

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 4. Μαθήματα επιλογής

Μάθημα 7. Εργαστήριο Υπολογιστών Βιοπληροφορικής

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Μοντελοποίηση Ομολογίας
- Σύνδεση

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Ένα παράδειγμα εργασίας σε εργαστήριο ηλεκτρονικών υπολογιστών βιοπληροφορικής παρουσιάζεται εδώ.

Παρουσιάζεται η εργασία με δύο λογισμικά για μοντελοποίηση ομολογίας και σύνδεση.

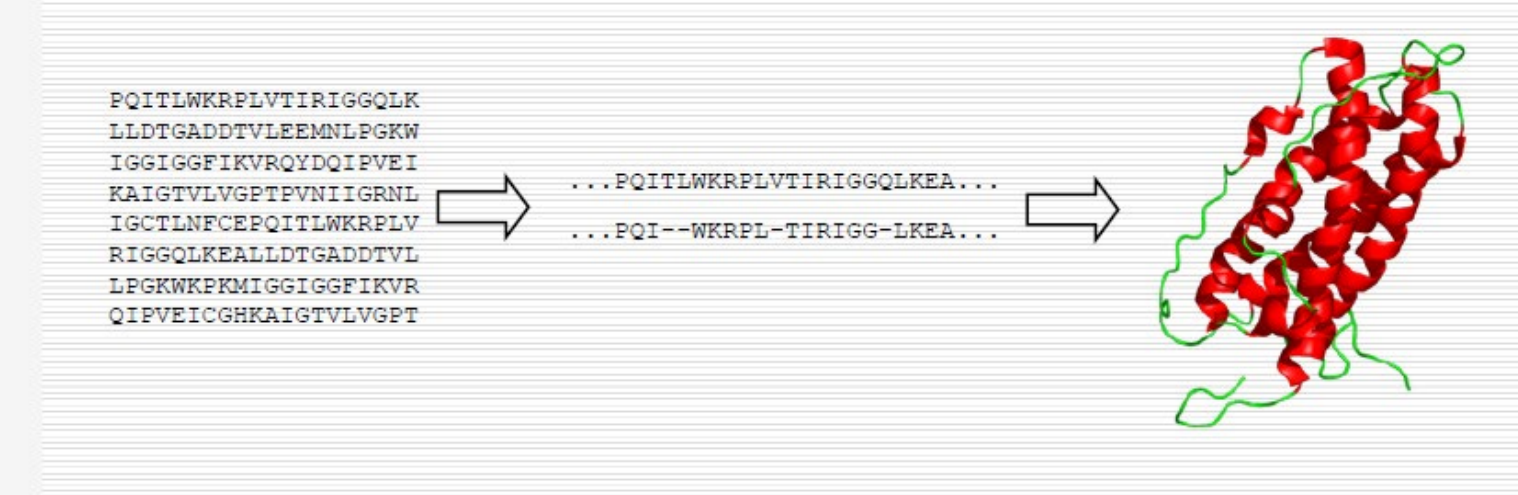
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

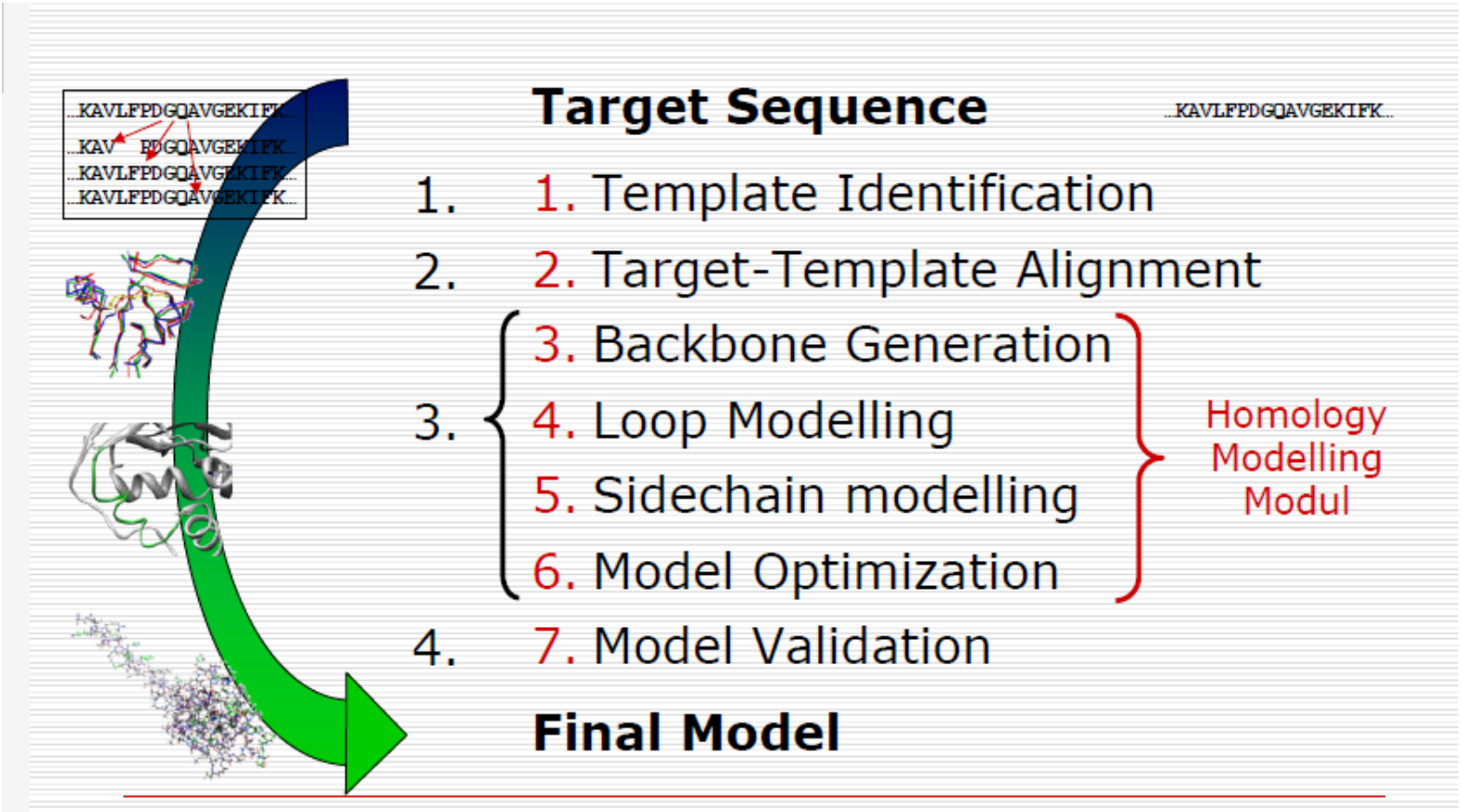
Μοντελοποίηση Ομολογίας

Εξελικτικό: Ομολογία (Συγκριτική) Μοντελοποίηση

- Οι πρωτεΐνες που σχετίζονται με την εξέλιξη διατηρούν την ίδια πτυχή παρά τη συσσώρευση διαφορετικών μεταλλάξεων
- δομική απόκλιση πιο αργή από την απόκλιση ακολουθίας
- αναζήτηση για ομόλογες πρωτεΐνες / αναλογικές πτυχές

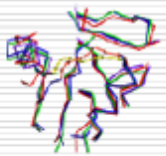


Μοντελοποίηση Ομολογίας



Μοντελοποίηση ομολογίας ΜΟΕ

1. Initial Partial Geometry Specification



- copy all-heavy atom coordinates if similar residue
- copy only backbone coordinates if dissimilar residue
- leave as indel if target proline and template not

2. Insertions and Deletions („indels")
   - missing residues in loops (insertion in target)
   - connecting chain (deletion in target)

   →using fragments of high-resolution chains or using a segment search if no complete fragment found

3. Loop selection and sidechain packing

- collection of independent models
- loops modelled first (analyzed using energey function)
- sidechain modelled using a rotamer library

Co-funded by the Erasmus+ Programme of the European Union

Μοντελοποίηση ομολογίας MOE

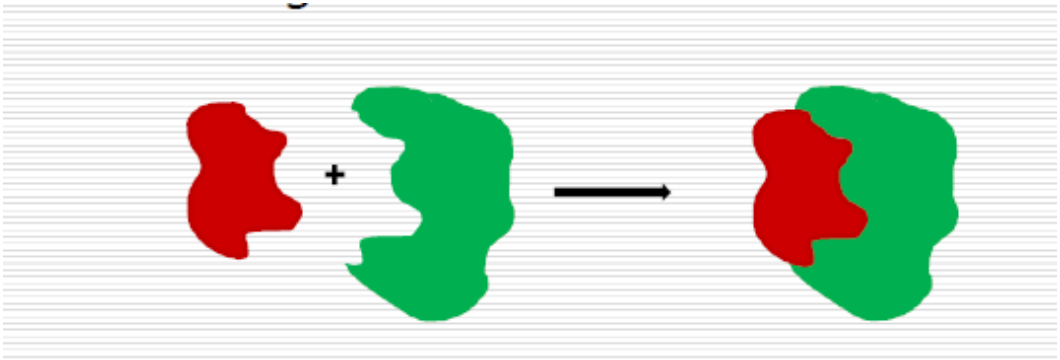
4. Final Model Selection and Refinement

→ final model based on best-scoring intermediate

- RMSD to average position of all models
- electrostatic solvation energy
- residue packing quality function (solvent accessibility, H-bond contacts, ration polar vs non-polar)
- effective atomic contact energy

Inspection using Protein Geometry

Μοριακή Αναγνώριση



Αλληλεπίδραση πρωτεΐνης-προσδέματος

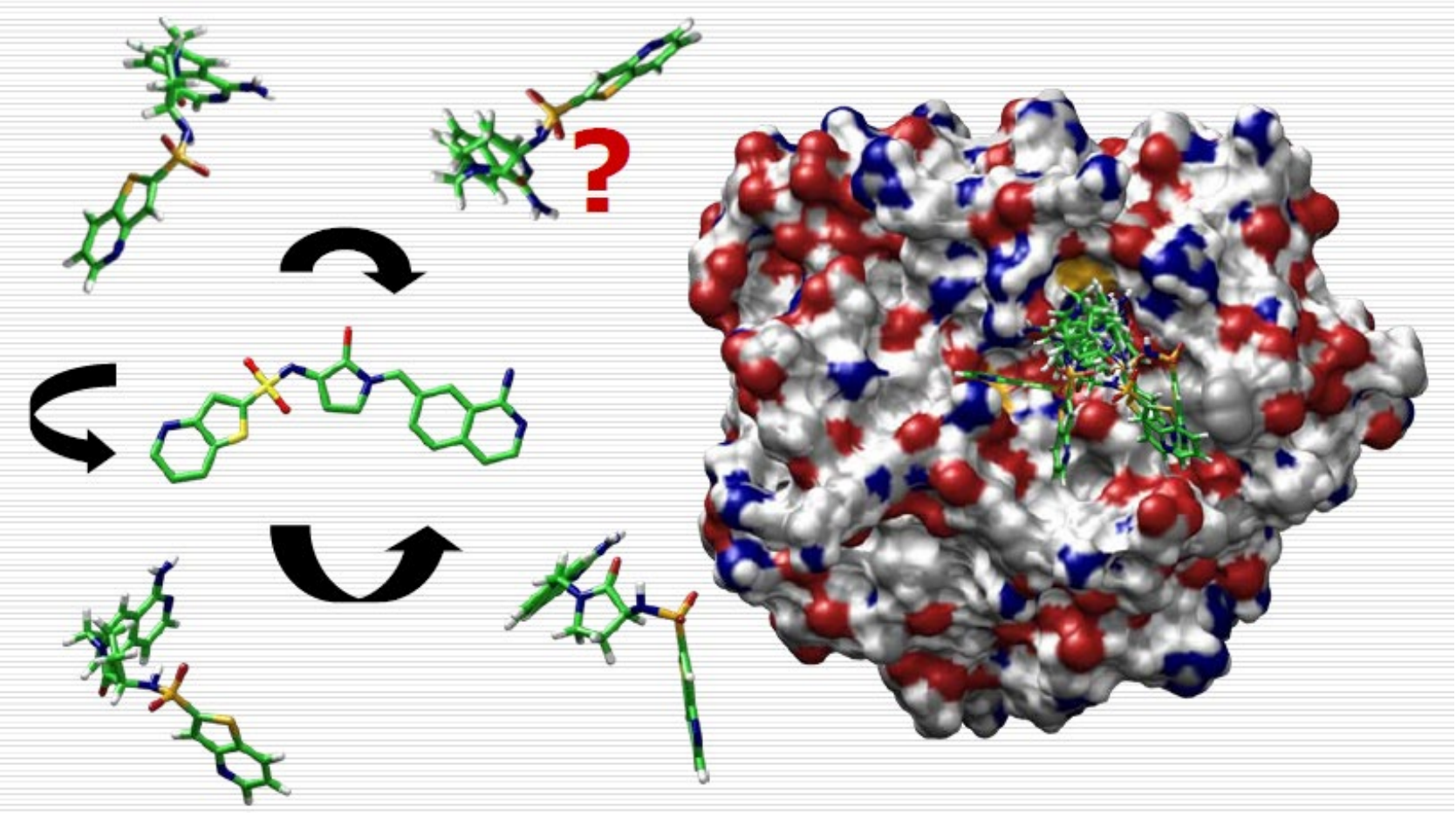
- Δομή του συμπλέγματος πρωτεΐνης-προσδέματος;
- Ισχύς της αλληλεπίδρασης πρωτεΐνης-προσδέματος;

Χρησιμοποιείται σε πρόβλεψη λειτουργίας δέσμευσης και εικονική προβολή

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

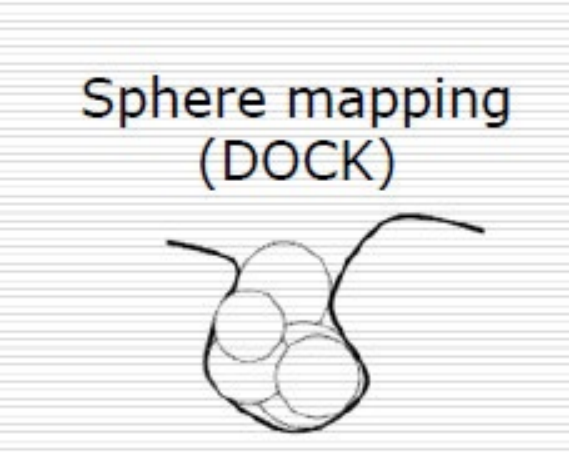
Docking: 3D Puzzle



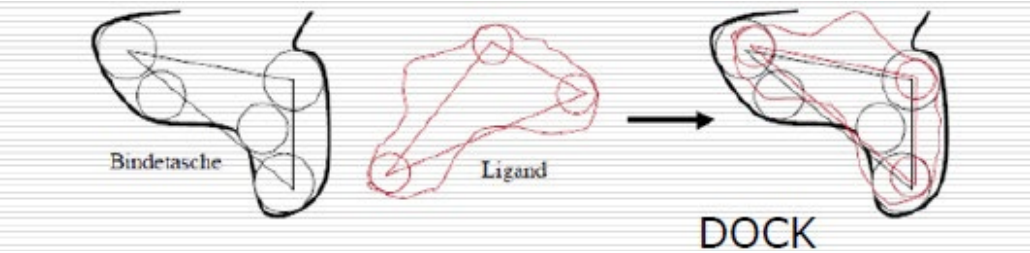
ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Σύνδεση

Βασική έννοια: σχήμα και αλληλεπίδραση συμπληρωματικά

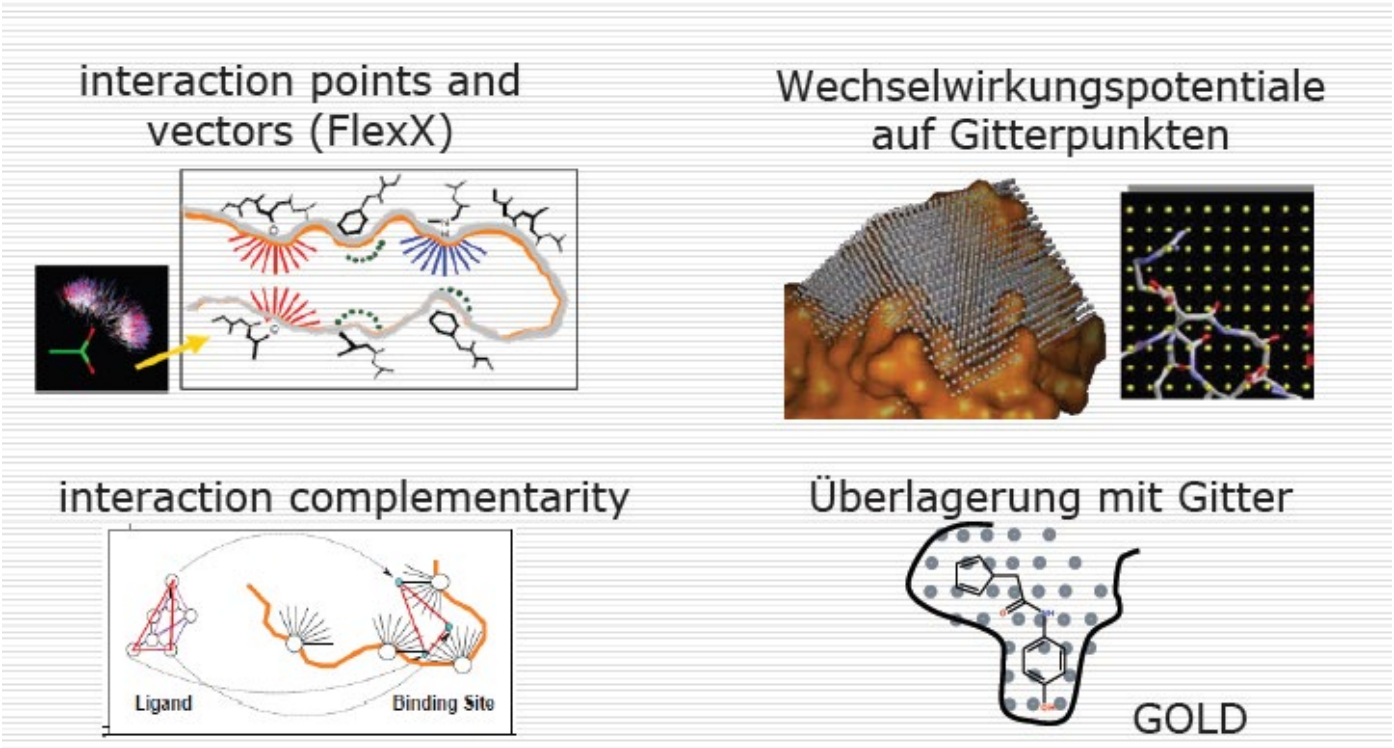


Σύγκριση απόστασης μεταξύ των κέντρων σφαιρών



Σύνδεση

Βασική έννοια: συμπληρωματικότητα σχήματος και αλληλεπίδρασης



Co-funded by the Erasmus+ Programme of the European Union

Αξιολόγηση

Ποια είναι η ακριβής «πόζα»;

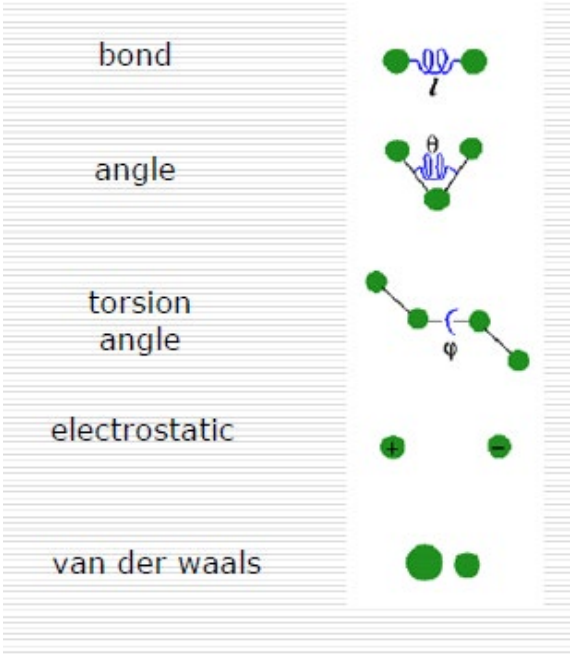
- Υπόθεση: η σύνθετη δομή του πειράματος αντιστοιχεί στην υψηλότερη συγγένεια
- Αναζήτηση για συνδυασμό πρωτεϊνικών προσδέματος με υψηλότερη συγγένεια

Αξιολόγηση

Αξιολόγηση με χρήση της συνάρτησης βαθμολόγησης ενθαλπίας ελεύθερης δέσμευσης ΔG

Πεδία Δύναμης Μοριακής Μηχανικής

- άτομα ως φορτισμένες σφαίρες
- δεσμεύει ως ελατήριο
- χωρίς ηλεκτρόνια



Αξιολόγηση

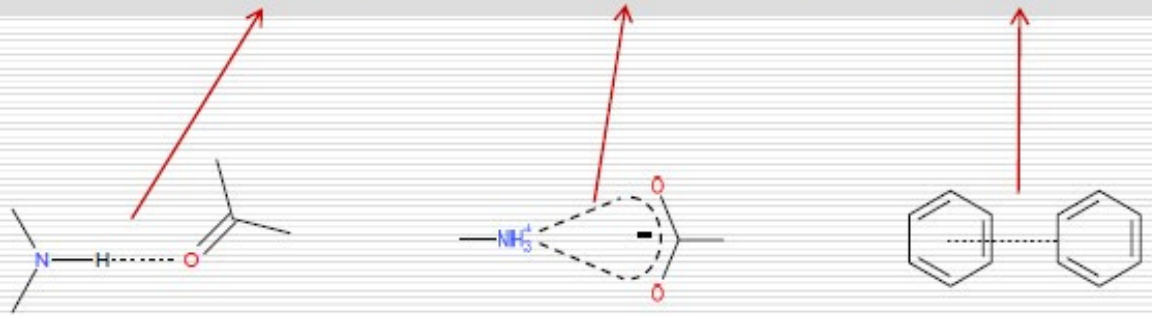
Αξιολόγηση με χρήση της συνάρτησης βαθμολόγησης

ενθαλπίας ελεύθερης δέσμευσης ΔG

Εμπειρικές συναρτήσεις (βάσει παλινδρόμησης)

→ παραγοντοποίηση του ΔG σε όρους απλής αλληλεπίδρασης

$$\Delta G_{bind} = \Delta G_0 + \Delta G_{hb} \Sigma_{H\text{-}bonds} f(\Delta r, \Delta\alpha) + \Delta G_{ionic} \Sigma_{ionic\ int.} f(\Delta r, \Delta\alpha) + \Delta G_{lipo} A_{lipo} + \Delta G_{rot} N_{rot}$$

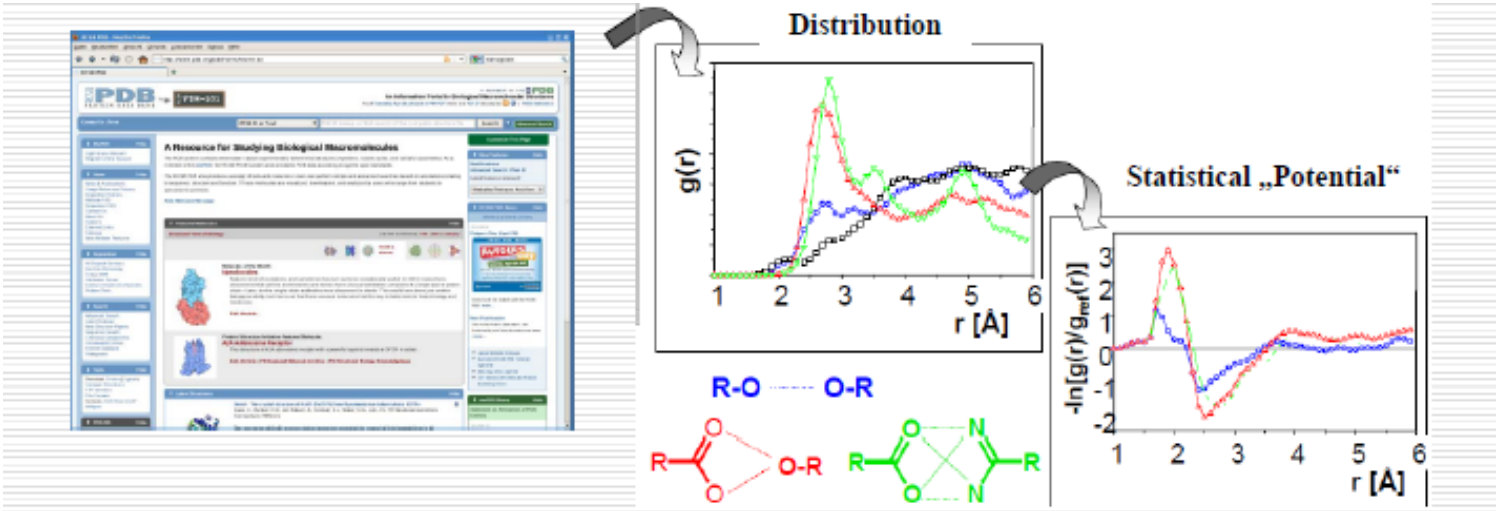


Αξιολόγηση

Αξιολόγηση με χρήση της συνάρτησης βαθμολόγησης ενθαλπίας ελεύθερης δέσμευσης ΔG

Λειτουργίες βαθμολόγησης με βάση τη γνώση

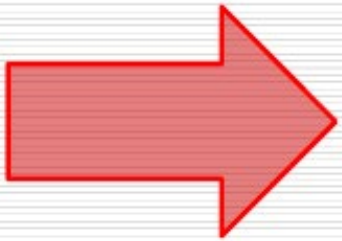
→ με βάση στατιστική παρατήρηση διαμοριακών στενών επαφών σε μεγάλες τρισδιάστατες βάσεις δεδομένων



Ρύθμιση σύνδεσης

Protein Setup

- Add Hydrogens
- Delete Waters
- Delete Ligand



Done in MOE

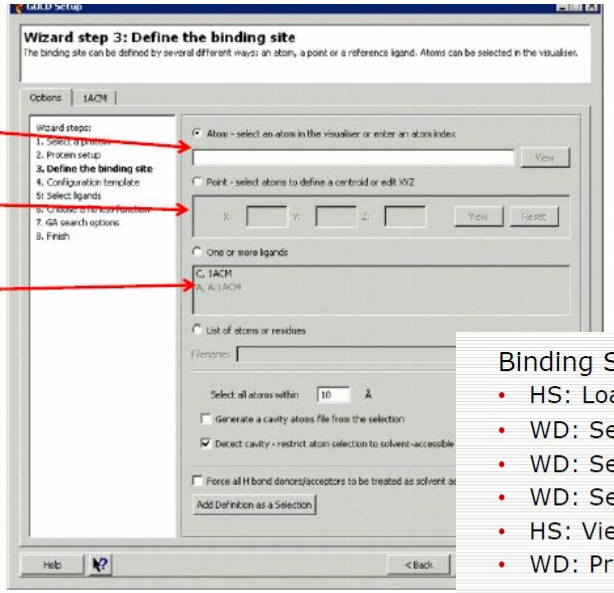
- HERMES: File | Open: ***1acm_pa1.mol2***
- WIZARD: Press [Next]

Ορισμός δεσμευτικής τοποθεσίας

Based on a specific atom

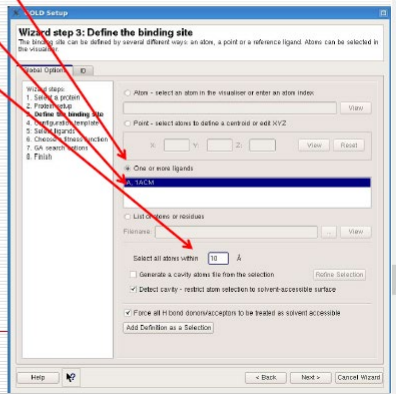
Based on a point

Based on a reference ligand in the binding site

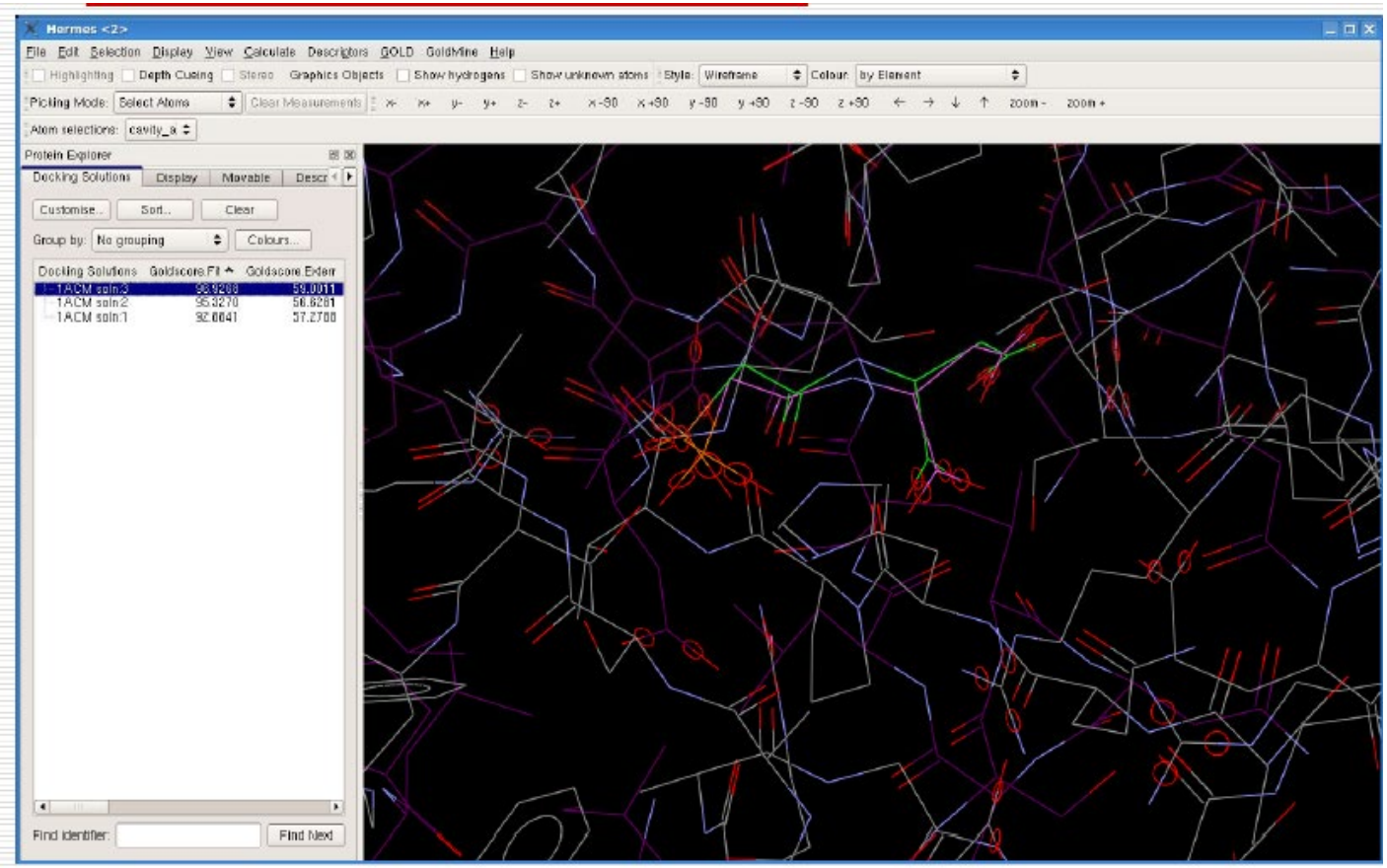


Binding Site Definition

- HS: Load `1acm_pal.mol2`
- WD: Select [One Or more ligands]
- WD: Select [A, 1ACM]
- WD: Set B.S. Radius to 10Å
- HS: View Binding Site
- WD: Press [Next]

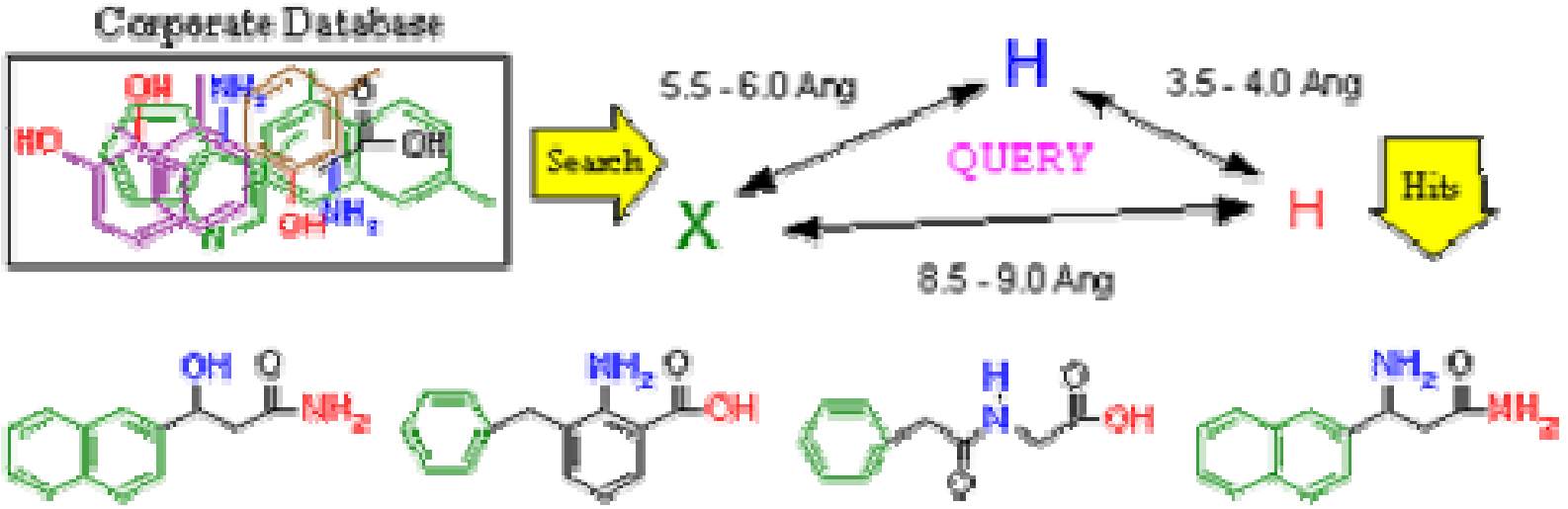


Προβολή λύσεων



Co-funded by the Erasmus+ Programme of the European Union

Τα αποτελέσματα προέκυψαν σε μια τρισδιάστατη αναζήτηση βάσης δεδομένων



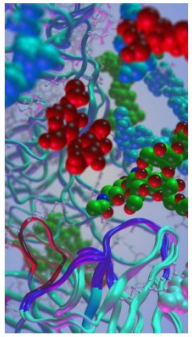
Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Εξηγήστε τα βήματα της μοντελοποίησης ομολογίας.
- Εξηγήστε τα βήματα της προετοιμασίας της πρωτεΐνης και των προσδεμάτων για σύνδεση
- Πώς μπορούμε να αναλύσουμε τα αποτελέσματα που έχουμε;

MOE

Molecular Operating Environment



- Molecular Modeling and Simulations
- Protein Modeling and Bioinformatics
- Structure/Fragment-Based Drug Design
- Pharmacophore Modeling
- Cheminformatics
- Development Environment

GOLD User Guide

GOLD User Guide
Conditions of Use
Introduction and Overview
Getting Started
Overview of the GOLD Interface
Using the GOLD Docking Wizard
Setting Up the Protein(s)
Essential Steps
Specifying the Protein File or Files
Protonation and Tautomeric States
Adding Hydrogen Atoms to the Protein Using Program Defaults
Applying Protonation Rules
Flipping Asn and Gln Residues
Specifying Histidine Tautomers
Deleting Ligands, Cofactors and Metal Ions
Water Molecules
Methodology For Handling Waters
Specifying Waters
Defining the Binding Site
Overview
Defining a Binding Site from an Atom
Defining a Binding Site from a Point
Defining a Binding Site from a Reference Ligand or Cofactor
Defining a Binding Site from a List of Atoms or Residues
Cavity Detection
Solvent Accessibility
Rotatable O-H and NH_3 Groups
Docking into a Rigid Protein
Metal Ions
Preparing a Protein Input File which Contains a Metal Ion

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 4. Μαθήματα Επιλογής

Μάθημα 8. Ποσοτική Φαρμακολογία

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Co-funded by the Erasmus+ Programme of the European Union

Περιεχόμενα

- Εισαγωγή
- Δείκτες Τοξικότητας και Φαρμακολογικής Δραστηριότητας
- Σχέσεις μεταξύ δεικτών: Από τους πρακτικούς περιγραφείς στην εμπειρική σχέση
- Σχέσεις Δομής-Δραστηριότητας και Προσεγγίσεις In Silico

Co-funded by the Erasmus+ Programme of the European Union

Εισαγωγή

Οι κλίμακες τοξικότητας και φαρμακολογικής δραστηριότητας των μορίων, ιδίως των τοξικών, των ξενοβιοτικών, των φαρμάκων, των θρεπτικών ουσιών κ.λπ., περιγράφονται με πολλαπλούς δείκτες και η πιο δημοφιλής είναι η διάμεση θανατηφόρα δόση (LD50). Σε μοριακό επίπεδο, οι αναστρέψιμες σταθερές αναστολής ή δέσμευσης παρέχουν μοναδικές πληροφορίες για την πιθανή δραστηριότητα των μορίων. Τονίζεται το σημαντικό πρόβλημα σχετικά με τη σημασία του IC50 για μη αναστρέψιμους συνδέτες/αναστολείς. Οι ορισμοί και οι αρχές για τον προσδιορισμό αυτών των ποσοτικών παραμέτρων παρουσιάζονται εν συντομία σε αυτό το άρθρο. Ιδιαίτερη προσοχή δίνεται στις σχέσεις μεταξύ αυτών των δεικτών. Τέλος, εξετάζονται εν συντομία διαφορετικές προσεγγίσεις που επιτρέπουν τη σύνδεση των φαρμακολογικών και τοξικολογικών ιδιοτήτων των μορίων όσον αφορά τις μοριακές αλληλεπιδράσεις (ή τις χημικές αντιδράσεις) με τους βιολογικούς τους στόχους. Επισημαίνονται οι πειραματικές τάσεις για μελλοντική εξέταση υψηλής απόδοσης ενεργών μορίων

Δείκτες Τοξικότητας και Φαρμακολογικής Δραστηριότητας

- Η μισή-μέγιστη αναστολτική συγκέντρωση (IC_{50})
- Θανατηφόρα δόση-50 (LD_{50})

Co-funded by the Erasmus+ Programme of the European Union

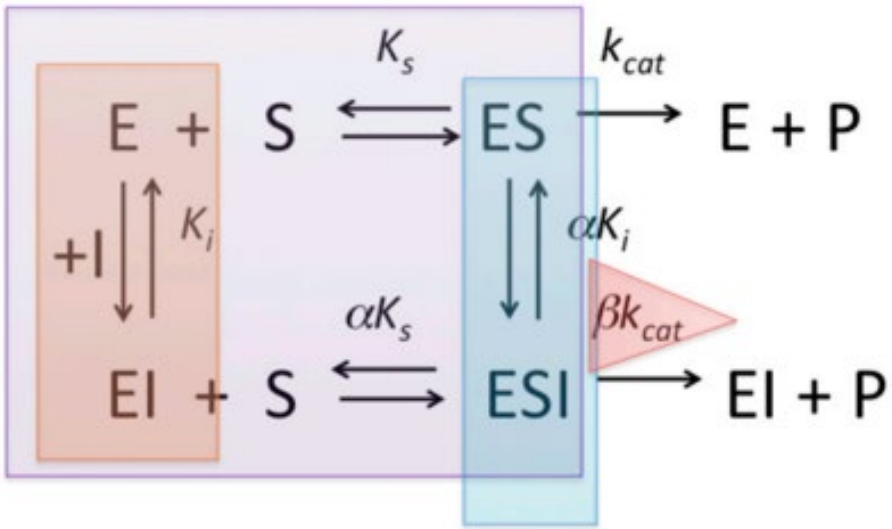
Η μισή-μέγιστη αναστολτική συγκέντρωση (IC_{50})

Στο επίπεδο στόχο, το IC50 $_{αντιστοιχεί}$ στη συγκέντρωση του αναστολέα που προκαλεί 50% αναστολή της δραστηριότητας του ενζύμου ή του υποδοχέα. Οι καμπύλες δόσης-απόκρισης για την αναστολή του ενζύμου ή τη δέσμευση του υποδοχέα μπορούν να προσδιοριστούν γραφικά με τη γραφική παράσταση της υπολειπόμενης κλασματικής δραστηριότητας ενός ενζύμου (v_i/v_0) ή της κλασματικής κατάληψης (B) ενός υποδοχέα ως συνάρτηση της συγκέντρωσης αναστολέα (ή συνδέτη). .

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Η μισή-μέγιστη ανασταλτική συγκέντρωση (IC_{50})



Ελάχιστοι μηχανισμοί για αναστρέψιμη αναστολή της ενζυμικής κατάλυσης. Η αναστολή μπορεί να είναι καθαρά ανταγωνιστική, ο αναστολέας I ανταγωνίζεται το υπόστρωμα S (α→∞, πορτοκαλί κουτί), μη ανταγωνιστικό (α ≠ 1, μωβ κουτί), μεικτό (α = 1, μωβ κουτί) ή μη ανταγωνιστικό, δεσμεύεται σε σύμπλοκο ES (Ki →∞; α Ki « Ki, κυανό κουτί). Εάν η αναστολή είναι μερική, β < 1 (ροζ τρίγωνο)

Co-funded by the Erasmus+ Programme of the European Union

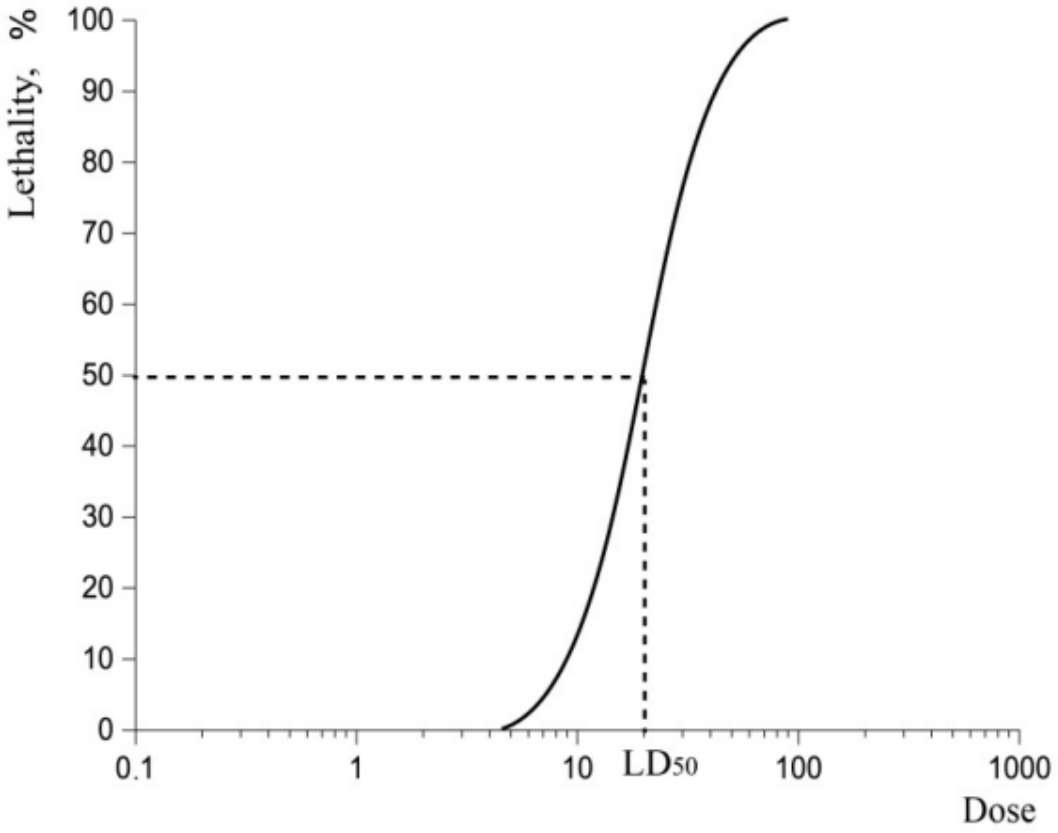
Η μισή-μέγιστη αναστολτική συγκέντρωση (IC_{50})

Type of Inhibitor		IC_{50}	IC_{50} vs. $[S]/K_M$
Competitive	$K_i\left(1+\frac{[S]}{K_M}\right)$	if $[S] << K_M$, $K_i = IC_{50}$	Linear ascending plot
Non-competitive	K_i	$\alpha = 1$	Linear horizontal plot
Mixed type	$\frac{[S]+K_M}{\left(\frac{K_M}{K_i}+\frac{[S]}{\alpha K_i}\right)}$	if $\alpha = 1$, $K_i = IC_{50}$	$\alpha > 1$: curvilinear ascending plot $\alpha < 1$: curvilinear descending plot
Uncompetitive	$\alpha K_i\left(1+\frac{K_M}{[S]}\right)$	if $[S] >> K_M$, $\alpha K_i \approx IC_{50}$	Curvilinear descending plot

Σχέσεις μεταξύ αναστρέψιμων αναστολέων και IC_{50}.

Co-funded by the Erasmus+ Programme of the European Union

Θανατηφόρα δόση-50 (LD_{50})



Co-funded by the Erasmus+ Programme of the European Union

Θανατηφόρα δόση-50 (LD_{50})

Όταν μιλάμε για επιδράσεις στις φυσιολογικές διεργασίες σε ολόκληρους οργανισμούς, και όχι μόνο για αναστολή/ενεργοποίηση ενζύμων και υποδοχέων, χρησιμοποιείται η έννοια της διάμεσης θανατηφόρας δόσης (LD50). Το LD50 είναι η δόση που σκοτώνει το 50 τοις εκατό μιας ομάδας ζώων.

Σχέσεις μεταξύ δεικτών: Από τους πρακτικούς περιγραφείς στην εμπειρική σχέση

- Σχέσεις μεταξύ IC_{50}, της μέσης αποτελεσματικής δόσης (ED_{50}) και LD_{50}
  - Σχέση μεταξύ LD50 και IC50
  - Θεραπευτικός Δείκτης (TI)
  - Θεραπευτικό Παράθυρο
- Σχέσεις Δομής-Δραστηριότητας και Προσεγγίσεις In Silico

Co-funded by the Erasmus+ Programme of the European Union

Σχέση μεταξύ LD_{50} και IC_{50}

$$LD_{50}\,(\mathrm{mg/kg}) = 0.372 \log IC_{50}\,(\mathrm{microg/mL}) + 2.024$$

Για την ελαχιστοποίηση του αριθμού των ζώων που χρησιμοποιούνται για μελέτες οξείας τοξικότητας, έχουν καθιερωθεί εμπειρικές σχέσεις μεταξύ IC_{50} και LD_{50}. Είναι δυνατό να γίνει μια υπόθεση για τις τιμές LD_{50} με βάση τα δεδομένα $IC_{50\ in\ vitro}$. Για παράδειγμα, η Διυπηρεσιακή Συντονιστική Επιτροπή για την επικύρωση εναλλακτικών μεθόδων πρότεινε έναν εμπειρικό τύπο για τους αρουραίους

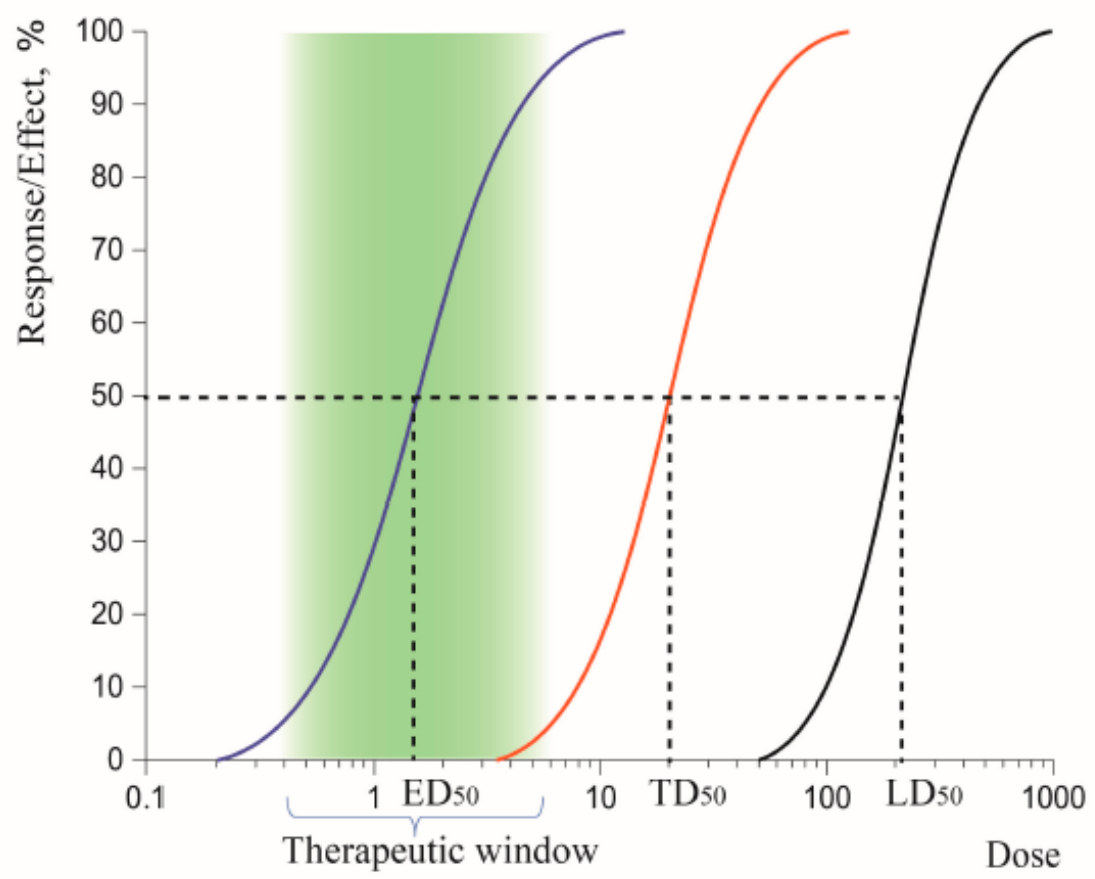
Θεραπευτικός Δείκτης (TI)

Παρά το γεγονός ότι το LD_{50} είναι ένα άχρηστο τελικό σημείο επειδή δεν μπορεί να προεκταθεί απευθείας στον άνθρωπο, χρησιμοποιείται στον προσδιορισμό του θεραπευτικού δείκτη (TI). Το TI είναι η αναλογία του LD_{50} προς το ED_{50}

$$TI = LD_{50} / ED_{50}$$

Co-funded by the Erasmus+ Programme of the European Union

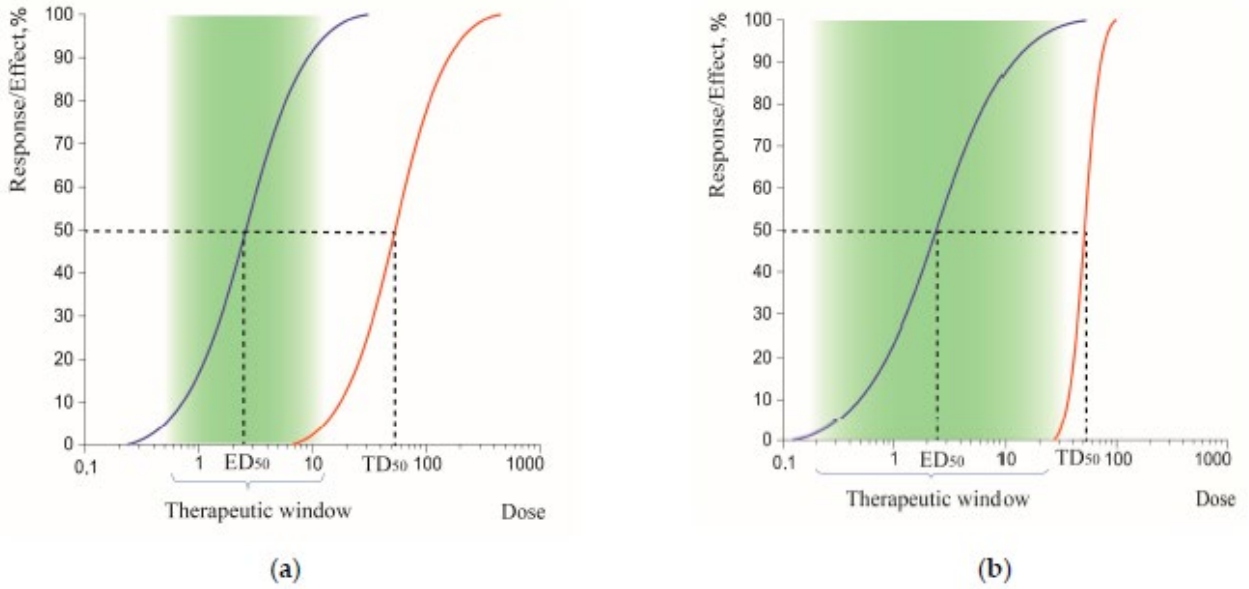
Θεραπευτικός Δείκτης (TI)



Θεραπευτικό Παράθυρο

Το θεραπευτικό παράθυρο αντιπροσωπεύει το εύρος των δόσεων όπου ένα φάρμακο είναι αποτελεσματικό χωρίς ανεπιθύμητες ιατρογενείς επιδράσεις. Ενώ ένα εγχειρίδιο φαρμακολογίας γράφει ότι «το TI είναι ένα μέτρο της ασφάλειας ενός φαρμάκου, επειδή μια μεγαλύτερη τιμή υποδηλώνει ένα μεγάλο περιθώριο μεταξύ των δόσεων που είναι αποτελεσματικές και των δόσεων που είναι τοξικές», ο Yartsev παρατήρησε ότι υπάρχουν πιθανότητες το TI να παραμένει το ίδιο αλλά θεραπευτικό παράθυρο μπορεί να αλλάξει, επομένως θα μπορούσε να υπάρχει μια πιο λογική έννοια από το TI

Θεραπευτικό Παράθυρο



Στενότερα (α) και ευρύτερα (β) θεραπευτικά παράθυρα (πράσινοι χώροι) με το ίδιο TI

Σχέσεις Δομής-Δραστηριότητας και Προσεγγίσεις In Silico

Σε μοριακό επίπεδο, οι πρώτες απόπειρες του Hansch για συσχετίσεις μεταξύ φυσικοχημικών παραμέτρων (πολικότητα, υδροφοβικότητα, συντονισμός, ισομερισμός και στερικότητα ...) βιοδραστικών μορίων και βιοχημικών διεργασιών οδήγησαν στα θεμέλια του QSAR (ποσοτικές σχέσεις δομής-δραστικότητας (QSTR) και QSTR) ποσοτικές σχέσεις δομής-τοξικότητας) στη δεκαετία του 1970.

Σχέσεις Δομής-Δραστηριότητας και Προσεγγίσεις In Silico

Αυτή η προσέγγιση επιτρέπει την ποσοτική περιγραφή των φαρμακολογικά και τοξικολογικά ενεργών μορίων από την άποψη των μοριακών αλληλεπιδράσεων. Για το σκοπό αυτό, δύο σημαντικές μεθοδολογικές πτυχές πρέπει να ληφθούν υπόψη: οι μοριακές περιγραφές των φαρμακολογικών ή/και τοξικολογικών επιδράσεων και ο σχεδιασμός αλγορίθμων και οι υπολογιστικές προσεγγίσεις που θα χρησιμοποιηθούν για την κατασκευή μοντέλων QSAR. Έτσι, εκτός από τους πρώτους μοριακούς περιγραφείς που εισήγαγε ο Hansch και οι πρωτοπόροι στο πεδίο, πρόσφατα εισήχθησαν νέες παράμετροι πρόβλεψης για την πρόβλεψη της τοξικότητας ή της φαρμακολογικής δραστηριότητας των μορίων. Για παράδειγμα, η μέθοδος συμβολής θραυσμάτων επιφάνειας ατόμου (ASFC) που βασίζεται στο σίγμα (σ) επιφανειακές επιφάνειες των ατόμων χρησιμοποιήθηκε με επιτυχία για την αξιολόγηση της τοξικότητας των ιοντικών υγρών και ακόμη και για την αποκάλυψη διαφορών τοξικότητας μεταξύ ισομερών έναντι μιας κυτταρικής γραμμής λευχαιμίας αρουραίου (IPC- 81)

Σχέσεις Δομής-Δραστηριότητας και Προσεγγίσεις In Silico

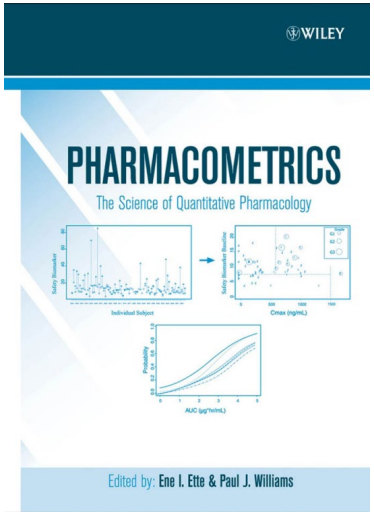
Ένας αιώνας έρευνας, μελετών σε ζώα και εναλλακτικών μεθόδων για τη διερεύνηση της τοξικότητας και της φαρμακολογικής αποτελεσματικότητας των χημικών ουσιών επέτρεψαν τον ορισμό αξιόπιστων παραμέτρων για την ποσοτικοποίηση της οξείας τοξικότητας των ξενοβιοτικών και την ασφαλή χρήση των φαρμάκων. Νέες προσεγγίσεις, που βρίσκονται ακόμη σε εξέλιξη, που συνδυάζουν QSAR, μικρορευστικές συσκευές που χρησιμοποιούν καλλιέργειες κυττάρων 2D και 3D, νέα μοντέλα ζώων και κυττάρων και τεχνολογίες έκφρασης ενός γονιδίου επιτρέπουν τον έλεγχο υψηλής απόδοσης τεράστιων βιβλιοθηκών χημικών ουσιών και την εξόρυξη νέων ενώσεων φαρμακολογικού ενδιαφέροντος ή ανίχνευση ανεπιθύμητων τοξικών ενώσεων στο περιβάλλον.

Co-funded by the Erasmus+ Programme of the European Union

Ερωτήσεις

- Ποιοι είναι οι δείκτες τοξικότητας και φαρμακολογικής δραστηριότητας;
- Εξηγήστε τις σχέσεις μεταξύ των δεικτών: από πρακτικούς περιγραφείς έως εμπειρικές σχέσεις

Co-funded by the Erasmus+ Programme of the European Union



Ene I. Ette, Paul J. Williams (2007)
Pharmacometrics : the science of quantitative pharmacology, John Wiley & Sons, Inc., Hoboken, New Jersey

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 4. Μαθήματα επιλογής

Μάθημα 9. Γλώσσες Σεναρίου (Script)

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Γλώσσες σεναρίου

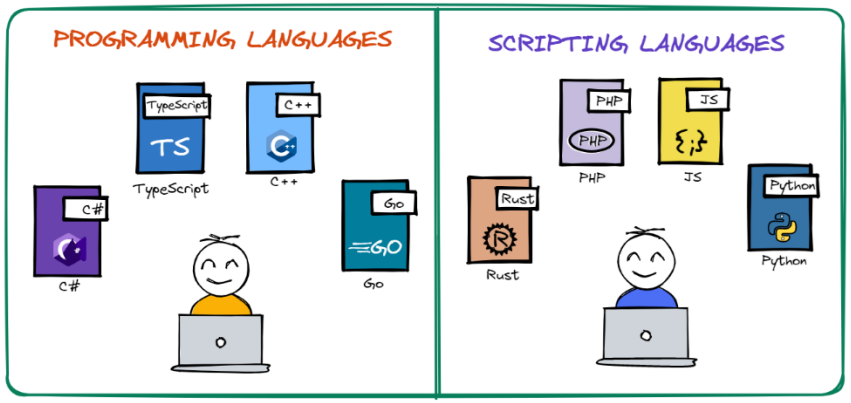
Μάθημα 1

Γλώσσες σεναρίου

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

ΣΦΑΙΡΙΚΗ ΕΙΚΟΝΑ

- Ο σκοπός αυτού του μαθήματος είναι να εξερευνήσει τον κόσμο των γλωσσών scripting και να εμβαθύνει στη σημασία τους στον σύγχρονο προγραμματισμό και αυτοματισμό.
- Σημασια
  - απλοποίηση και εξορθολογισμός διαφόρων εργασιών στον τομέα του προγραμματισμού και της ανάπτυξης λογισμικού
  - προσφέρουν μια ισορροπία μεταξύ της ισχύος των παραδοσιακών γλωσσών προγραμματισμού και της απλότητας του αυτοματισμού
  - Σύνταξη αποτελεσματικού και συνοπτικού κώδικα



Αυτοματοποίηση

- Οι γλώσσες δέσμης ενεργειών είναι εξειδικευμένες γλώσσες προγραμματισμού που έχουν σχεδιαστεί για
  - αυτοματοποίηση
  - την εκτέλεση
  - απλοποίηση συγκεκριμένων εργασιών ή διαδικασιών.
- Οι γλώσσες δέσμης ενεργειών είναι καθοριστικές για την αυτοματοποίηση επαναλαμβανόμενων και πολύπλοκων εργασιών.
- Παρέχουν μια υψηλού επιπέδου, φιλική προς το χρήστη διεπαφή για τον έλεγχο λογισμικού, συστημάτων και εφαρμογών.

Απλουστεύσεις εργασιών

- Οι γλώσσες δέσμης ενεργειών απλοποιούν την κωδικοποίηση προσφέροντας ένα σύνολο προκατασκευασμένων λειτουργιών και βιβλιοθηκών.
- Αφαιρούν πολλές λειτουργίες χαμηλού επιπέδου, μειώνοντας την ανάγκη για εκτεταμένη χειροκίνητη κωδικοποίηση.
- Αυτή η απλοποίηση επιταχύνει τη διαδικασία ανάπτυξης και ελαχιστοποιεί την πιθανότητα σφαλμάτων.

Χαρακτηριστικά

- **Ερμηνεύεται**

Οι περισσότερες γλώσσες δέσμης ενεργειών ερμηνεύονται, επιτρέποντας την άμεση εκτέλεση κώδικα χωρίς μεταγλώττιση.

- **Δυναμική πληκτρολόγηση**

Συχνά χρησιμοποιούν δυναμική πληκτρολόγηση, επιτρέποντας στις μεταβλητές να αλλάζουν τύπους δεδομένων κατά τη διάρκεια του χρόνου εκτέλεσης.

- **Ευκαμψία**

Οι γλώσσες σεναρίου δίνουν προτεραιότητα στην ευκολία χρήσης και την προσαρμοστικότητα έναντι των αυστηρών κανόνων.

Πυτηον

- **Περιγραφή**

Η Python είναι μια ευέλικτη, υψηλού επιπέδου γλώσσα δέσμης ενεργειών γνωστή για την αναγνωσιμότητα και την απλότητά της.

- **Συνήθεις περιπτώσεις χρήσης** :
  - Ανάπτυξη Ιστού (Django, Flask).
  - Ανάλυση και οπτικοποίηση δεδομένων (Pandas, Matplotlib).
  - Μηχανική μάθηση και AI (TensorFlow, PyTorch).
  - Εργασίες αυτοματισμού και δέσμης ενεργειών.

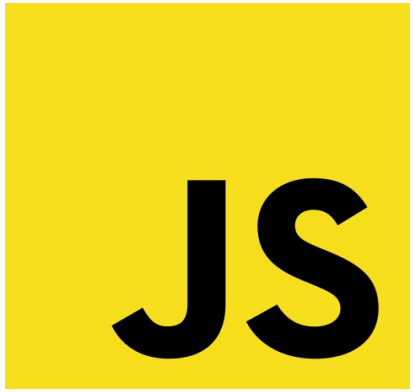
Co-funded by the Erasmus+ Programme of the European Union

JavaScript

- **Περιγραφή** :

Η JavaScript είναι κυρίως μια γλώσσα δέσμης ενεργειών από την πλευρά του πελάτη που χρησιμοποιείται για την ανάπτυξη Ιστού.

- **Συνήθεις περιπτώσεις χρήσης** :
  - Ανάπτυξη web front-end (React, Angular, Vue.js).
  - Δημιουργία διαδραστικών διαδικτυακών εφαρμογών.
  - δέσμες ενεργειών από την πλευρά του διακομιστή (Node.js).



Ρουμπίνι

- **Εισαγωγή** : Η Ruby ξεκίνησε στην Ιαπωνία στα μέσα της δεκαετίας του 1990, και σχεδιάστηκε με έμφαση στην απλότητα και την παραγωγικότητα.

- **Ruby on Rails** : Ένα δημοφιλές πλαίσιο εφαρμογών ιστού γραμμένο σε Ruby. Δίνει έμφαση στη σύμβαση έναντι της διαμόρφωσης και την αρχή DRY (Don't Repeat Yourself).

- **Εφαρμογές** : Εφαρμογές Ιστού, βοηθητικά προγράμματα συστήματος, διακομιστές ιστού και άλλα.

Perl

- **Εισαγωγή** : Γνωστό ως «ελβετικό μαχαίρι» των γλωσσών σεναρίου, η Perl αναπτύχθηκε αρχικά για επεξεργασία κειμένου.

- **Κανονικές εκφράσεις** : Η Perl είναι γνωστή για τις ισχυρές της δυνατότητες επεξεργασίας κειμένου και κανονικής έκφρασης.

- **Εφαρμογές** : Ανάπτυξη Ιστού, βιοπληροφορική, διαχείριση συστήματος και πολλά άλλα.

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Shell Scripting - Bash

- **Εισαγωγή** : Σενάρια γραμμένα για το περιβάλλον φλοιού ενός λειτουργικού συστήματος.

- **Bash** : Η πιο δημοφιλής γλώσσα δέσμης ενεργειών φλοιού, ειδικά σε περιβάλλοντα Linux.

- **Εφαρμογές** : Αυτοματοποίηση εργασιών συστήματος, εκτέλεση προγράμματος και επεξεργασία κατά παρτίδες.

Lua

- **Εισαγωγή** : Μια ελαφριά γλώσσα σεναρίου σχεδιασμένη για ενσωματωμένα συστήματα.
- **Ανάπτυξη παιχνιδιών** : Το Lua χρησιμοποιείται συχνά στη βιομηχανία τυχερών παιχνιδιών, ειδικά για τη δημιουργία σεναρίων παιχνιδιών.
- **Εφαρμογές** : Διακομιστές Ιστού, παιχνίδια και ενσωματωμένα συστήματα.

Πλεονεκτήματα των γλωσσών δέσμης ενεργειών

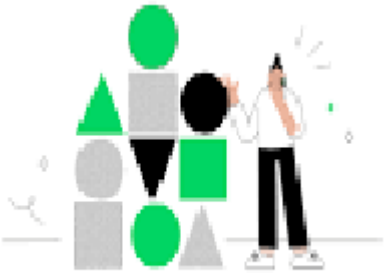
- **Ταχεία ανάπτυξη** : Οι γλώσσες δέσμης ενεργειών επιτρέπουν τη γρήγορη δημιουργία πρωτοτύπων και ανάπτυξη.

- **Cross-Platform** : Πολλές γλώσσες σεναρίου είναι ανεξάρτητες από την πλατφόρμα.

- **Επεκτασιμότητα** : Μπορούν εύκολα να ενσωματωθούν με άλλες γλώσσες και τεχνολογίες.

Περιορισμοί και Θεωρήσεις

- **Απόδοση** : Μετά την ερμηνεία, οι γλώσσες δέσμης ενεργειών ενδέχεται να μην είναι τόσο γρήγορες όσο οι μεταγλωττισμένες γλώσσες.



- **Ασφάλεια** : Τα σενάρια μπορεί να είναι ευάλωτα εάν δεν γράφονται με γνώμονα την ασφάλεια.

- **Βέλτιστες πρακτικές** : Είναι σημαντικό να ακολουθείτε τις βέλτιστες πρακτικές για να διασφαλίσετε αποτελεσματικά και ασφαλή σενάρια.

Συμπέρασμα και Πόροι

- **Ανακεφαλαίωση** : Σύντομη επισκόπηση της σημασίας των γλωσσών δέσμης ενεργειών στη σύγχρονη πληροφορική.

- **Περαιτέρω ανάγνωση** : Εξερευνήστε κάθε γλώσσα σεναρίου σε βάθος.

Πόροι :

- W3Schools
- Mozilla Developer Network (MDN)
- Επίσημη τεκμηρίωση Python
- Επίσημη Τεκμηρίωση Ruby
- Τεκμηρίωση Perl

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Ενότητα 4. Μαθήματα επιλογής

Μάθημα 10. Μοντέλα τεχνητών νευρωνικών δικτύων

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Μοντέλα Τεχνητών Νευρωνικών Δικτύων

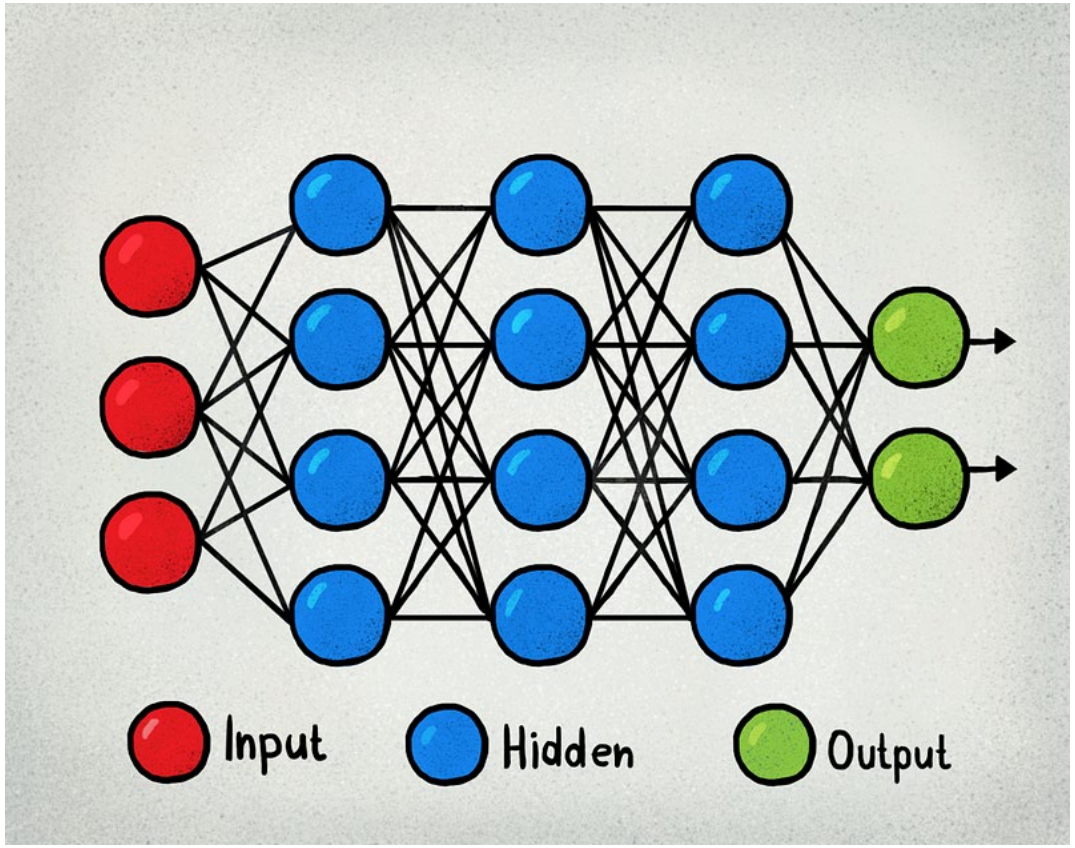
Από τα perceptron στη βαθιά μάθηση

ERASMUS+
Key Action KA2 - Cooperation for innovation and the exchange of good practices
Action Type KA226 - Partnerships for Digital Education Readiness

Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131
Co-funded by the Erasmus+ Programme of the European Union

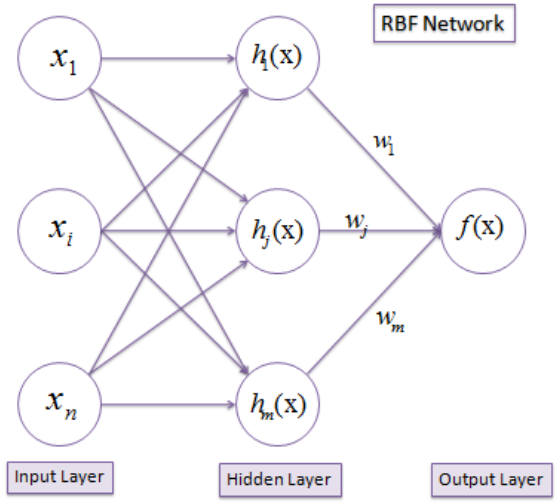
Εισαγωγή στα μοντέλα ANN

- **Εξέλιξη των νευρωνικών δικτύων:** Μια σύντομη ιστορία του πώς τα τεχνητά νευρωνικά δίκτυα έχουν εξελιχθεί με την πάροδο του χρόνου, από απλά perceptrons έως πολύπλοκες αρχιτεκτονικές βαθιάς μάθησης.
- **Θεμελιώδεις Έννοιες:** Επισήμανση των βασικών αρχών και ιδεών που στηρίζουν τα νευρωνικά δίκτυα, όπως οι νευρώνες, τα βάρη, οι προκαταλήψεις και οι συναρτήσεις ενεργοποίησης.
- **Πρόοδος από τα Perceptrons στη Deep Learning:** Μια γρήγορη επισκόπηση του πώς τα πρώιμα μοντέλα όπως τα perceptron έθεσαν τα θεμέλια για πιο προηγμένες αρχιτεκτονικές όπως τα συνελικτικά νευρωνικά δίκτυα (CNN) και τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNNs).



Μοντέλα Θεμελιωδών Νευρωνικών Δικτύων

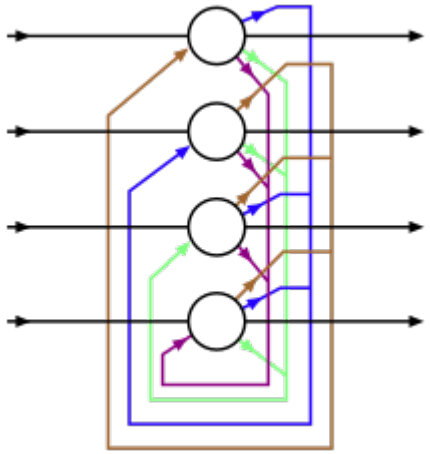
- **Perceptron:** Η απλούστερη μορφή νευρωνικού δικτύου, ικανή για δυαδική ταξινόμηση. Παρουσιάστηκε από τον Frank Rosenblatt στα τέλη της δεκαετίας του 1950.
- **Multi-Layer Perceptron (MLP):** Μια επέκταση του perceptron, που αποτελείται από ένα στρώμα εισόδου, ένα ή περισσότερα κρυφά στρώματα και ένα στρώμα εξόδου. Μπορεί να μοντελοποιήσει πιο σύνθετες σχέσεις.



$$f(\mathrm{x}) = \sum_{j=1}^{m} w_j h_j(\mathrm{x})$$

$$h(x) = \exp\left(-\frac{(x-c)^2}{r^2}\right)$$

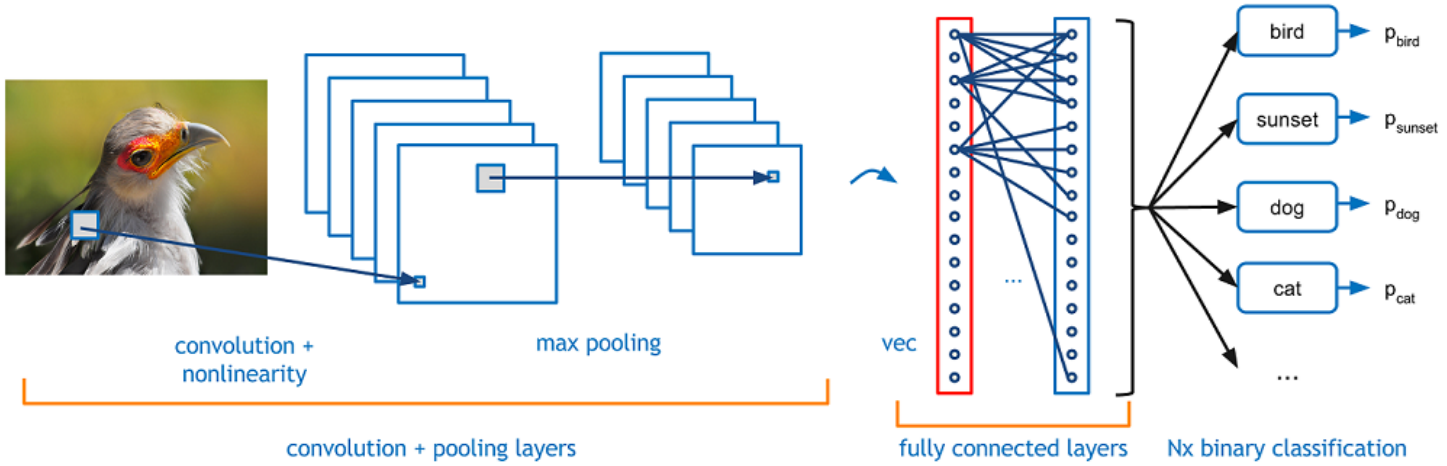
- **Δίκτυο συνάρτησης ακτινικής βάσης (RBFN):** Χρησιμοποιεί συναρτήσεις ακτινικής βάσης ως συναρτήσεις ενεργοποίησης, που χρησιμοποιούνται κυρίως για προσέγγιση και παρεμβολή συναρτήσεων.
- **Hopfield Network:** Ένα επαναλαμβανόμενο νευρωνικό δίκτυο που χρησιμεύει ως συσχετιστικό σύστημα μνήμης. Παρουσιάστηκε από τον John Hopfield στις αρχές της δεκαετίας του 1980.



Συνελικτικά νευρωνικά δίκτυα (1)

Τι είναι τα Συνελικτικά Νευρωνικά Δίκτυα;

- **Έμπνευση:** Σχεδιασμένο σύμφωνα με το ανθρώπινο οπτικό σύστημα, συγκεκριμένα τον οπτικό φλοιό.
- **Μοναδικό χαρακτηριστικό:** Χρησιμοποιεί συνελικτικά επίπεδα για αυτόματη και προσαρμοστική εκμάθηση χωρικών ιεραρχιών χαρακτηριστικών από εικόνες εισόδου.
- **Εφαρμογή:** Χρησιμοποιείται κυρίως σε εργασίες αναγνώρισης εικόνας και βίντεο, αλλά και σε οποιαδήποτε δεδομένα με χωρικές δομές.
- **Δομή:** Αποτελείται από συνελικτικά στρώματα, στρώματα συγκέντρωσης και πλήρως συνδεδεμένα στρώματα.



Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Συνελικτικά νευρωνικά δίκτυα (2)

Ψάχνοντας βαθύτερα στα CNN

- **Convolutional Layer:** Εφαρμόζει τη λειτουργία συνέλιξης στην είσοδο, καταγράφοντας τοπικά μοτίβα.
- **Pooling Layer:** Μειώνει τις χωρικές διαστάσεις, δίνοντας έμφαση στα κυρίαρχα χαρακτηριστικά.
- **Πλήρως συνδεδεμένο επίπεδο:** Λαμβάνει τα χαρακτηριστικά υψηλού επιπέδου από προηγούμενα επίπεδα και τα μετατρέπει στην τελική έξοδο.
- **Λειτουργίες ενεργοποίησης:** Εισαγάγετε τη μη γραμμικότητα, με το ReLU (Rectified Linear Unit) να είναι το πιο κοινό στα CNN.
- **Strides & Padding:** Τεχνικές για τον έλεγχο των χωρικών διαστάσεων μετά τη συνέλιξη.

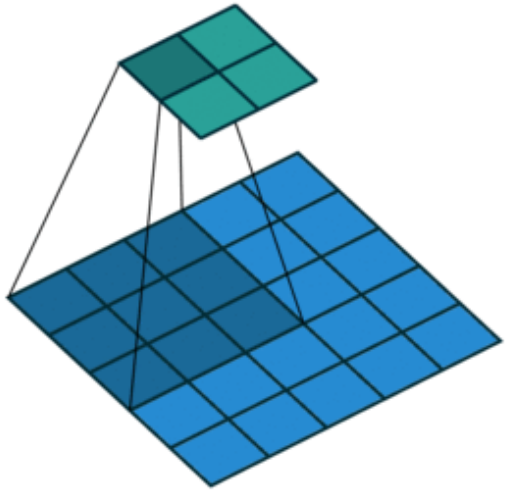
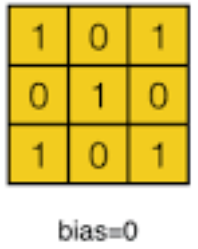
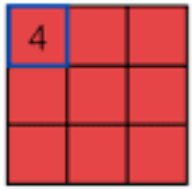


image 5*5



filter 3*3

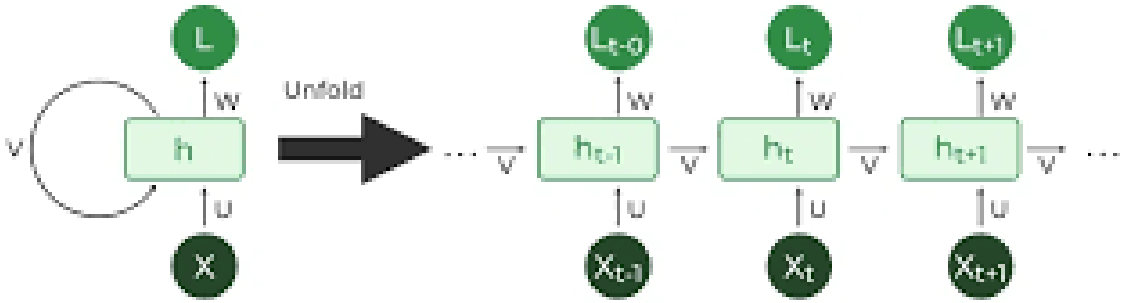


feature map 3*3

Επαναλαμβανόμενα νευρωνικά δίκτυα (1)

Τι είναι τα RNN;

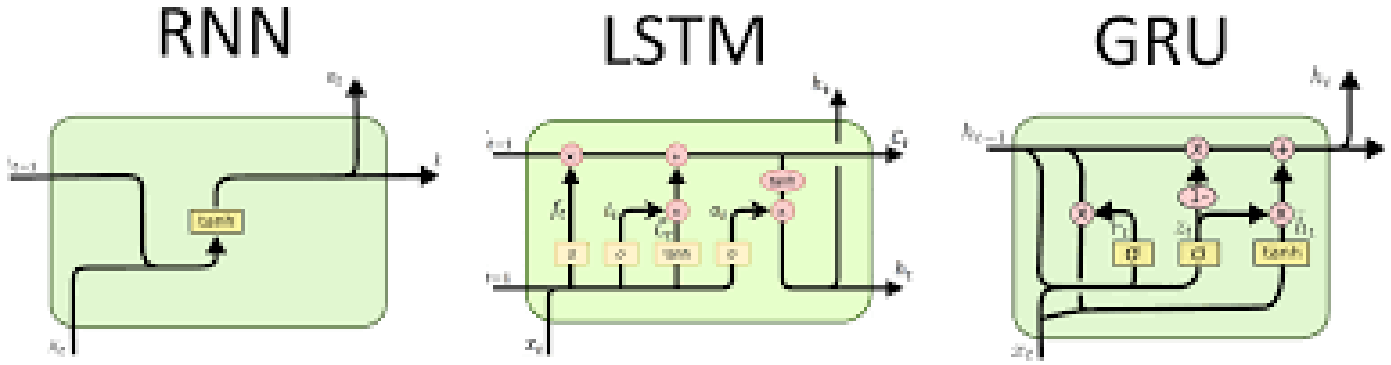
- **Ορισμός:** Νευρωνικά δίκτυα με βρόχους, που επιτρέπουν στις πληροφορίες να παραμένουν.
- **Μοναδικό χαρακτηριστικό:** Διαθέτει μνήμη που καταγράφει πληροφορίες σχετικά με όσα έχουν υπολογιστεί μέχρι τώρα.
- **Εφαρμογή:** Χρησιμοποιείται κυρίως σε εργασίες πρόβλεψης ακολουθίας όπως η πρόβλεψη χρονοσειρών, η επεξεργασία φυσικής γλώσσας και η αναγνώριση ομιλίας.
- **Πρόκληση:** Εξαφάνιση και έκρηξη προβλημάτων κλίσης, καθιστώντας δύσκολη την εκπαίδευσή τους σε μεγάλες ακολουθίες.



Επαναλαμβανόμενα νευρωνικά δίκτυα (2)

Βασικά εξαρτήματα και παραλλαγές των RNN

- **Βασικό RNN:** Το παραδοσιακό μοντέλο, παλεύει με μακροχρόνιες εξαρτήσεις.
- **Μακροπρόθεσμη Μνήμη (LSTM):** Ένα ειδικό είδος RNN, ικανό να μάθει μακροπρόθεσμες εξαρτήσεις. Χρησιμοποιεί πύλες για τον έλεγχο της ροής των πληροφοριών.
- **Gated Recurrent Units (GRU):** Μια απλοποιημένη έκδοση του LSTM, με λιγότερες πύλες και παραμέτρους.
- **Αμφίδρομα RNN:** Επεξεργάζονται ακολουθίες και από τις δύο κατευθύνσεις (από παρελθόν προς το μέλλον και αντίστροφα).
- **Μοντέλα ακολουθίας σε ακολουθία:** Χρησιμοποιούνται σε εργασίες όπως η αυτόματη μετάφραση, όπου οι ακολουθίες εισόδου και εξόδου μπορεί να έχουν διαφορετικά μήκη.



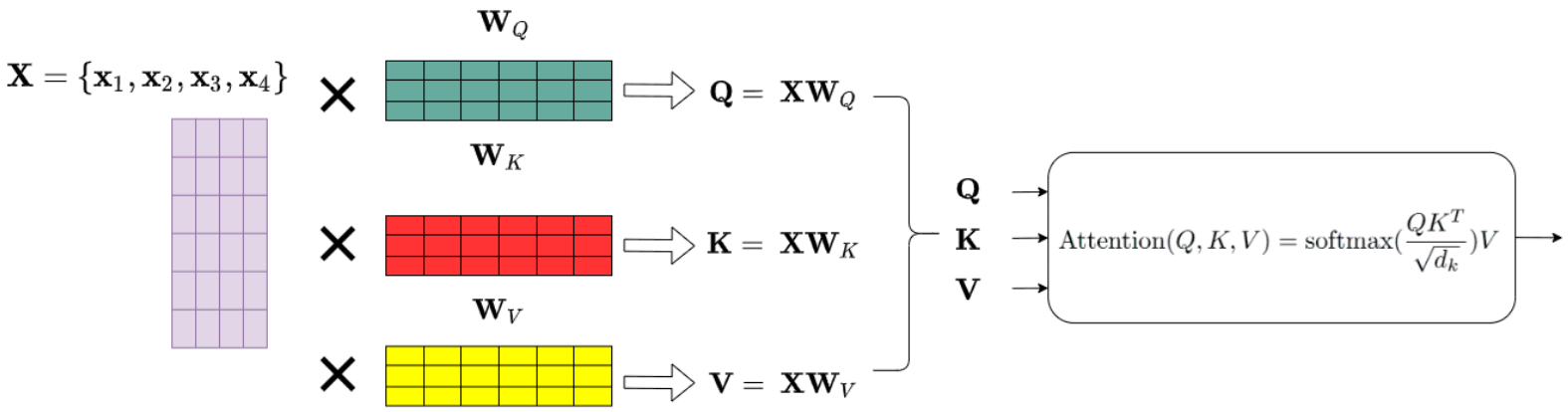
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Προηγμένα μοντέλα και παραλλαγές

Μετασχηματιστές και Μηχανισμοί Προσοχής

- **Ορισμός:** Μια αρχιτεκτονική μοντέλου σχεδιασμένη για το χειρισμό διαδοχικών δεδομένων, καθιστώντας την ιδανική για εργασίες όπως η μετάφραση και η σύνοψη κειμένου.
- **Μηχανισμός Προσοχής:** Επιτρέπει στο μοντέλο να εστιάζει σε διαφορετικά μέρη των δεδομένων εισόδου, παρόμοια με τον τρόπο με τον οποίο οι άνθρωποι δίνουν προσοχή σε συγκεκριμένα μέρη εισόδου.
- **Αυτοπροσοχή:** Το μοντέλο αποδίδει διαφορετικούς βαθμούς προσοχής σε διαφορετικές λέξεις, επιτρέποντάς του να ερμηνεύει το πλαίσιο πιο αποτελεσματικά.
- **Εφαρμογές:** Οι μετασχηματιστές είναι η ραχοκοκαλιά μοντέλων όπως το BERT, το GPT και το T5.



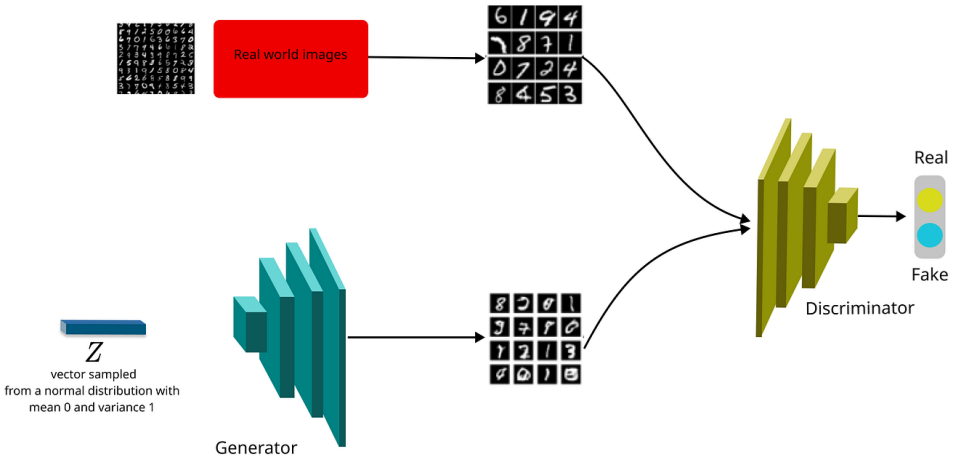
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Προηγμένα μοντέλα και παραλλαγές

Δίκτυα Generative Adversarial Networks (GAN)

- **Ορισμός:** Περιλαμβάνει δύο δίκτυα, μια γεννήτρια και μια συσκευή διάκρισης, που εκπαιδεύονται μαζί.
- **Γεννήτρια:** Προσπαθεί να παράγει δεδομένα.
- **Διακριτικός:** Προσπαθεί να διακρίνει μεταξύ πραγματικών και πλαστών δεδομένων.
- **Εκπαίδευση:** Εκπαιδεύονται παράλληλα έως ότου η γεννήτρια παράγει δεδομένα που ο χρήστης που διακρίνει δεν μπορεί να διακρίνει από τα πραγματικά δεδομένα.
- **Εφαρμογές:** Παραγωγή εικόνας, μεταφορά στυλ, αύξηση δεδομένων.



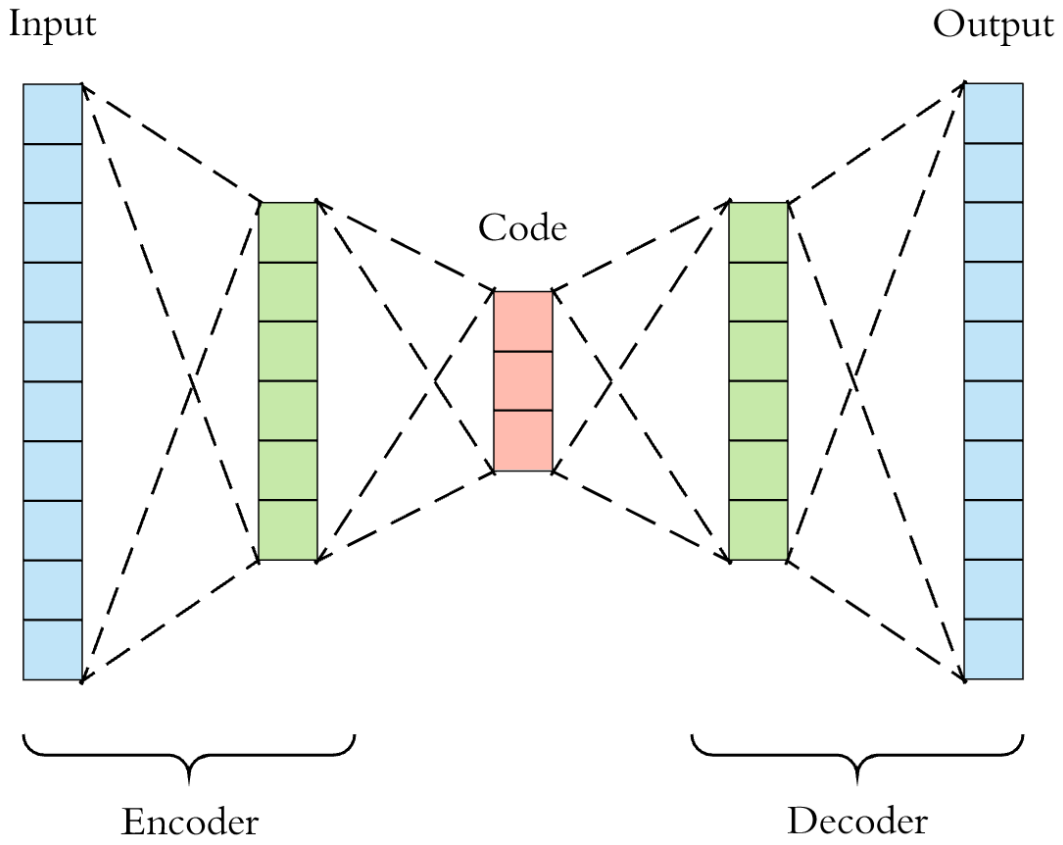
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Προηγμένα μοντέλα και παραλλαγές

Αυτοκωδικοποιητές

- **Ορισμός:** Ένας τύπος νευρωνικού δικτύου που χρησιμοποιείται για εργασίες μάθησης χωρίς επίβλεψη, κυρίως συμπίεση και παραγωγή δεδομένων.
- **Δομή:** Περιλαμβάνει δύο κύρια μέρη:
  - **Κωδικοποιητής:** Συμπιέζει την είσοδο σε μια συμπαγή, λανθάνουσα αναπαράσταση.
  - **Αποκωδικοποιητής:** Ανακατασκευάζει τα δεδομένα εισόδου από αυτήν τη συμπαγή αναπαράσταση.
- **Παραλλαγές:**
  - **Denoising Autoencoder:** Χρησιμοποιείται για την αφαίρεση του θορύβου από τα δεδομένα.
  - **Variational Autoencoder (VAE):** Δημιουργεί νέα δεδομένα παρόμοια με τα δεδομένα εισόδου.
- **Εφαρμογές:** Συμπίεση δεδομένων, ανίχνευση ανωμαλιών, αποθορυβοποίηση εικόνας και εργασίες δημιουργίας.



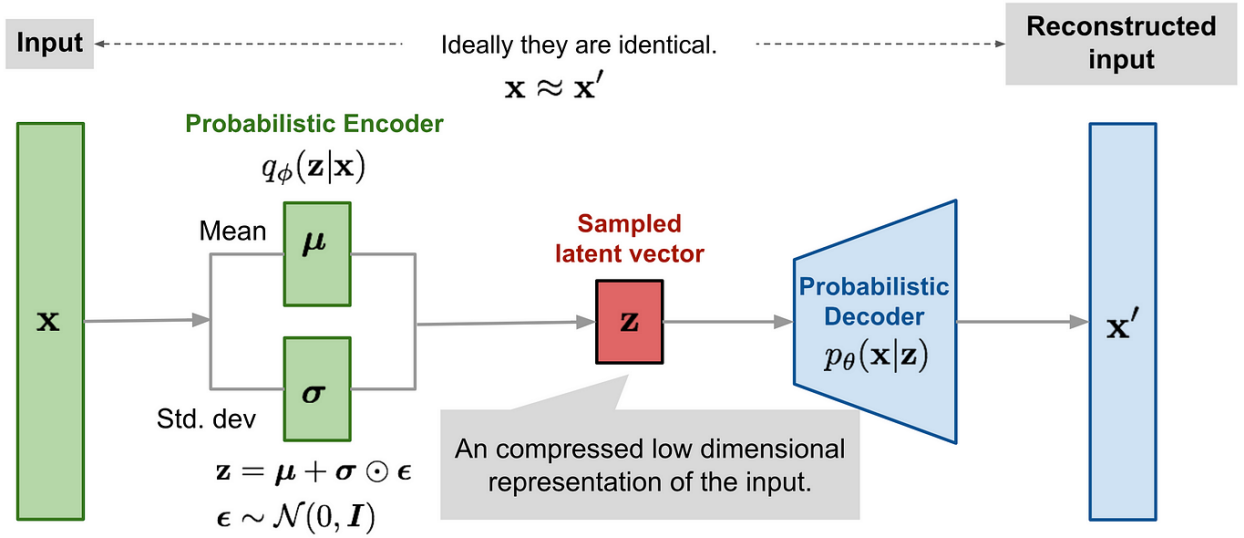
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Προηγμένα μοντέλα και παραλλαγές

Variational Autoencoders (VAEs)

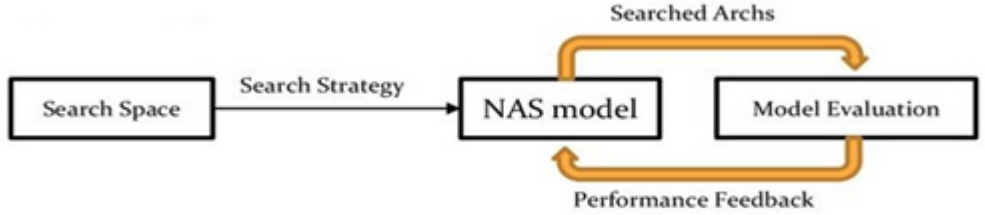
- **Ορισμός:** Ένας τύπος αυτόματου κωδικοποιητή που προσθέτει πιθανοτικούς κωδικοποιητές και αποκωδικοποιητές.
- **Λειτουργικότητα:** Αντί να συμπιέζουν τα δεδομένα εισόδου σε έναν σταθερό «κώδικα» στον λανθάνοντα χώρο του κωδικοποιητή, τα VAE μετατρέπουν τα δεδομένα σε μια διανομή πιθανών κωδικών.
- **Απώλεια ανακατασκευής:** Μετρά πόσο καλά κάνει ο αποκωδικοποιητής στην ανακατασκευή των δεδομένων εισόδου.
- **Τακτοποίηση:** Διασφαλίζει ότι αυτές οι διανομές κατανέμονται κατά προσέγγιση κανονικά.
- **Εφαρμογές:** Διαγραφή θορύβου εικόνας,

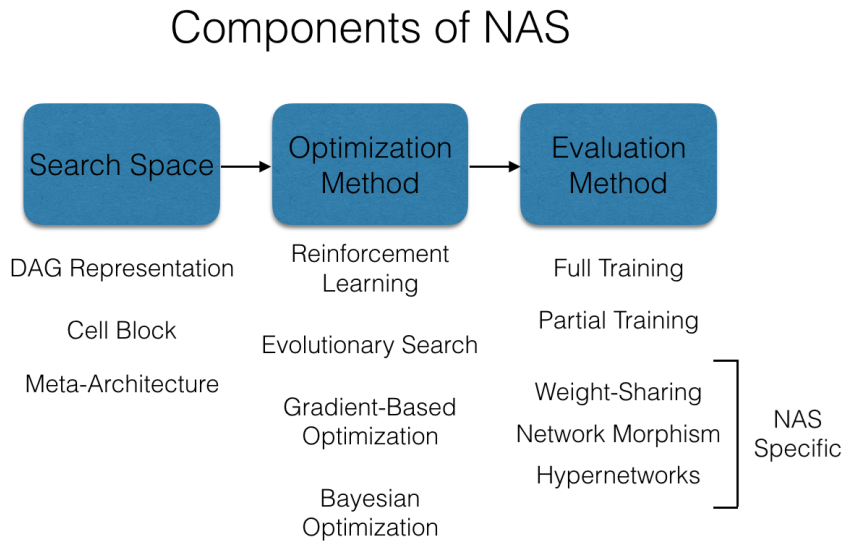


Προηγμένα μοντέλα και παραλλαγές

Αναζήτηση Νευρωνικής Αρχιτεκτονικής (NAS)

- **Ορισμός:** Μια αυτοματοποιημένη διαδικασία για την εύρεση της καλύτερης αρχιτεκτονικής νευρωνικών δικτύων για ένα δεδομένο σύνολο δεδομένων και εργασία.
- **Search Space:** Καθορίζει το σύνολο όλων των πιθανών αρχιτεκτονικών που μπορούν να αναζητηθούν.
- **Στρατηγική:** Χρησιμοποιεί ενισχυτική μάθηση, εξελικτικούς αλγόριθμους ή Bayesian βελτιστοποίηση για την αναζήτηση βέλτιστων αρχιτεκτονικών.
- **Πλεονεκτήματα:** Μπορεί να ανακαλύψει νέες αρχιτεκτονικές που ξεπερνούν τις χειροποίητες αρχιτεκτονικές.
- **Εφαρμογές:** Χρησιμοποιείται σε εργασίες όπου η βέλτιστη αρχιτεκτονική δεν είναι γνωστή, όπως νέα σύνολα δεδομένων ή εργασίες.





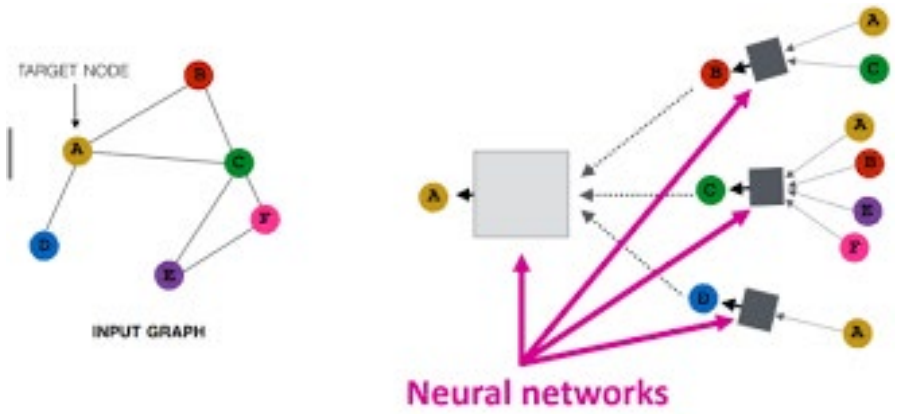
Tackling COVID-19 – Challenges Aiming at Bettering the Higher Education Quality (TACOHEQ)
Project Number: 2020-1-BG01-KA226-HE-095131

Co-funded by the Erasmus+ Programme of the European Union

Εξειδικευμένα μοντέλα νευρωνικών δικτύων

νευρωνικά δίκτυα γράφων (GNN)

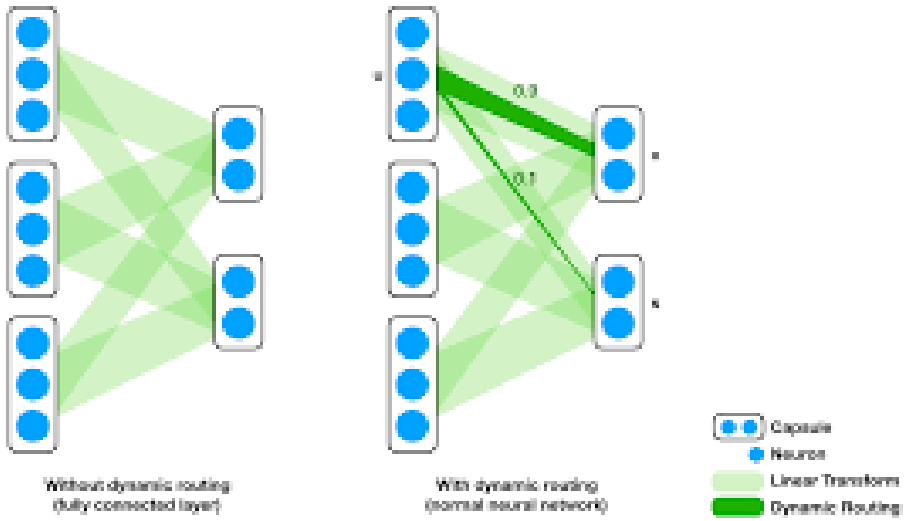
- **Ορισμός:** Νευρωνικά δίκτυα σχεδιασμένα να λειτουργούν απευθείας με δεδομένα δομημένα ως γραφήματα.
- **Κανόνας διάδοσης:** Οι κόμβοι συγκεντρώνουν πληροφορίες από τους γείτονές τους.
- **Εφαρμογές:**
  - **Ανάλυση κοινωνικών δικτύων:** Κατανόηση κοινωνικών συμπεριφορών και συνδέσεων.
  - **Συστήματα συστάσεων:** Προτάσεις προϊόντων με βάση γραφήματα αλληλεπίδρασης χρήστη-στοιχείου.
  - **Μοριακή Χημεία:** Πρόβλεψη μοριακών ιδιοτήτων.



Εξειδικευμένα μοντέλα νευρωνικών δικτύων

Δίκτυα καψουλών (CapsNets)

- **Ορισμός:** Αρχιτεκτονική νευρωνικών δικτύων σχεδιασμένη για να ξεπερνά τις αδυναμίες των CNN στη διατήρηση των χωρικών ιεραρχιών μεταξύ των χαρακτηριστικών.
- **Δυναμική δρομολόγηση:** Διασφαλίζει ότι η έξοδος μιας κάψουλας αποστέλλεται σε έναν κατάλληλο γονέα στο παραπάνω επίπεδο.
- **Πλεονεκτήματα:** Καλύτερη διατήρηση των χωρικών σχέσεων, πιο ισχυρή σε αντίπαλες επιθέσεις και απαιτεί λιγότερα δεδομένα για εκπαίδευση.
- **Εφαρμογές:** Ταξινόμηση εικόνων, αναγνώριση προσώπου και ανίχνευση αντικειμένων.



Προκλήσεις και Μελλοντικές Κατευθύνσεις

Προκλήσεις:

- **Ερμηνευσιμότητα Μοντέλου:** Κατανόηση γιατί τα νευρωνικά δίκτυα παίρνουν ορισμένες αποφάσεις.
- **Υπερπροσαρμογή:** Τα πολύπλοκα μοντέλα μπορεί να έχουν εξαιρετικά καλή απόδοση σε δεδομένα εκπαίδευσης, αλλά να αποτυγχάνουν σε νέα, αόρατα δεδομένα.
- **Υπολογιστικό κόστος:** Η εκπαίδευση προηγμένων μοντέλων απαιτεί σημαντικούς υπολογιστικούς πόρους.
- **Αντιθετικές επιθέσεις:** Μικρές, σκόπιμα δημιουργημένες αλλαγές στην είσοδο μπορούν να παραπλανήσουν τα νευρωνικά δίκτυα.

Μελλοντικές κατευθύνσεις:

- **Neural Architecture Search (NAS):** Αυτοματοποιημένες μέθοδοι για την εύρεση της καλύτερης αρχιτεκτονικής δικτύου.
- **Κβαντικά νευρωνικά δίκτυα:** Διερεύνηση των δυνατοτήτων του κβαντικού υπολογισμού στη βαθιά μάθηση.
- **Ενεργειακά αποδοτικά μοντέλα:** Κάνοντας τα μοντέλα πιο αποδοτικά για ανάπτυξη σε συσκευές άκρων.

Συμπέρασμα και Πόροι

- **Ανακεφαλαίωση:** Εξερευνήσαμε ένα ευρύ φάσμα μοντέλων νευρωνικών δικτύων, από βασικές αρχιτεκτονικές έως τις πιο πρόσφατες καινοτομίες, κατανοώντας τα δυνατά σημεία, τις εφαρμογές και τις προκλήσεις τους.

- **Συνεχής Εξέλιξη:** Ο τομέας των νευρωνικών δικτύων είναι δυναμικός, με συνεχείς προόδους και ανακαλύψεις.

Πόροι για περαιτέρω εξερεύνηση:

- **Βιβλία:** "Deep Learning" των Goodfellow, Bengio και Courville.
- **Διαδικτυακά Μαθήματα:** Coursera's "Neural Networks and Deep Learning" του Andrew Ng.
- **Ιστότοποι:** Neural Network Zoo - Ένας οπτικός οδηγός για διάφορες αρχιτεκτονικές.
- **Ερευνητικά Περιοδικά:** Journal of Artificial Neural Networks, Google Scholar (για τις τελευταίες εργασίες).

Stay Curious: Το ταξίδι στα νευρωνικά δίκτυα είναι ατελείωτο. Συνεχίστε να εξερευνάτε, να πειραματίζεστε και να μαθαίνετε!