



# Βιοπληροφορική

# Σύντομη Περιγραφή

1. Εισαγωγή – Τι είναι η βιοπληροφορική.
2. Βασικές έννοιες βιολογίας – Νουκλεοτιδικές αλληλουχίες.
3. Βασικές έννοιες βιολογίας – Πρωτεϊνικές αλληλουχίες.
4. Βιβλιογραφικές βάσεις δεδομένων (Scholar Google, Scopus, PubMed).
5. Βιολογικές βάσεις δεδομένων.
6. Ανάλυση ομοιότητας αλληλουχιών.
7. Πρότυπα, μοτίβα και αποτυπώματα στις αλληλουχίες των βιολογικών μακρομορίων.
8. Βασικές αρχές εξέλιξης.
9. Πρωτεωμική.
10. Ανάλυση των πρωτεϊνικών αλληλεπιδράσεων με τη βιοπληροφορική.
11. Ιατρική Πληροφορική και Βιοπληροφορική.



# Τι είναι Βιοπληροφορική

# Επιστήμη των Υπολογιστών και Βιολογία

- Υπολογιστές και πληροφορικά συστήματα χρησιμοποιούνται για να συγκεντρώσουν, να αναλύσουν και να επιλύσουν βιολογικά προβλήματα.
  - Large databases and data storage.
  - Data mining.
- Βιολογικά συστήματα χρησιμοποιούνται ως πρότυπα με σκοπό τη βελτίωση των υπολογιστών και των πληροφοριακών συστημάτων:
  - Genetic algorithms.
  - Neural networks.

# (Μοριακή) Βιο-πληροφορική

- Η βιοπληροφορική εφαρμόζει «τεχνικές πληροφορικής» (προερχόμενες από επιστημονικά πεδία όπως τα **εφαρμοσμένα μαθηματικά, η πληροφορική και η στατιστική**) για να κατανοήσει και να οργανώσει τις πληροφορίες που συνδέονται με τη **μοριακή βιολογία**, σε μία ευρεία κλίμακα.

Ειδικότερα...

Χρησιμοποιεί υπολογιστικές μεθόδους για να μελετήσει και να οργανώσει τα βιολογικά δεδομένα που υπάρχουν διαθέσιμα.

Συγκεντρώνει, αποθηκεύει και αναλύει βιολογικά δεδομένα.

# Ένα μοριακό αλφάβητο

- Τα περισσότερα **βιολογικά μόρια** είναι **πολυμερή**, διατεταγμένες αλυσίδες απλών μορίων που ονομάζονται **μονομερή**.
- Όλα τα μονομερή ανήκουν στην ίδια γενική κατηγορία, αλλά υπάρχουν διάφοροι τύποι με ξεχωριστά και συγκεκριμένα χαρακτηριστικά.
- Πολλά μονομερή μπορούν να ενωθούν μεταξύ τους και να σχηματίσουν ένα μοναδικό και μεγάλο μακρομόριο.
- Η σειρά τοποθέτησης των μονομερών στο μακρομόριο κωδικοποιεί πληροφορία, όπως ακριβώς τα γράμματα σε ένα αλφάβητο.

# Έννοιες

- **Γονίδιο:** Συγκεκριμένα μέρη του DNA με αποθηκευμένη πληροφορία.
- **Γονιδίωμα:** Το σύνολο των γονιδίων ενός οργανισμού.
- **Πρωτεΐνη:** Βιολογικό μακρομόριο με συγκεκριμένη δομή και λειτουργία. Προκύπτει από το αντίστοιχο γονίδιο.

# Στόχοι Μοριακής Βιολογίας

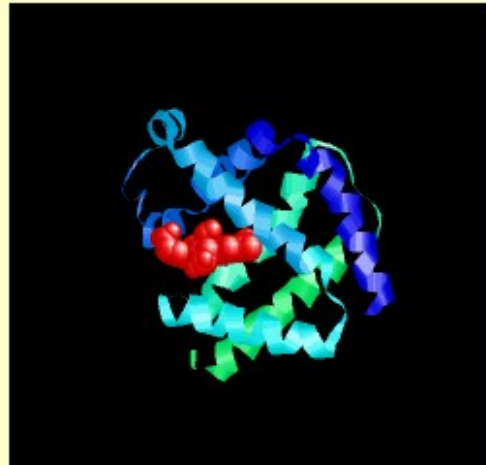
- **Αναγνώριση γονιδίων και καθορισμός των λειτουργιών που ρυθμίζουν:** Ταυτοποίηση γονιδίων, αναγνώριση λειτουργιών και ιδιοτήτων τους.
- **Ακολουθοποίηση και σύγκριση των γονιδιωμάτων διαφορετικών οργανισμών:** Πληροφορίες για την εξελεγκτική πορεία των οργανισμών.
- **Κατανόηση της γονιδιακής έκφρασης:** Κατανόηση βασικών κυτταρικών λειτουργιών, όπως η αναπνοή και ο μεταβολισμός.
- **Κατανόηση γενετικών ασθενειών:** Στόχος η αντιμετώπιση ασθενειών με γενετικό υπόβαθρο.
- Η χρήση εργαλείων Πληροφορικής επιλύει πολλά υπολογιστικά προβλήματα, όπως:



# Κεντρικό παράδειγμα Βιοπληροφορικής

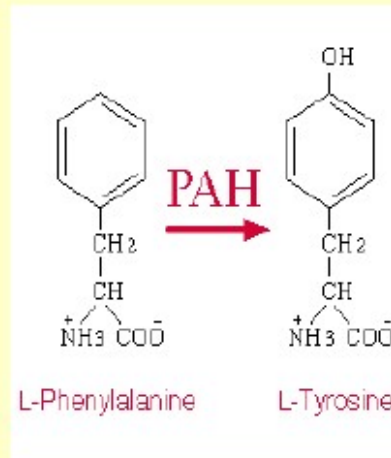
Genetic  
Information

MVHLTPEEKT  
AVNALWGKVN  
VDAVGGEALG  
RLLVVYPWTQ  
RFFESFGDLS  
SPDAVMGNPK  
VKAHGKKVLG  
AFSDGLAHLA  
NLKGTFSQLS  
ELHCDKLHVD  
PENFRLLGNV  
LVCVLARNF  
KEFTPQMCAA  
YQKVVAGVAN  
ALAHKYH

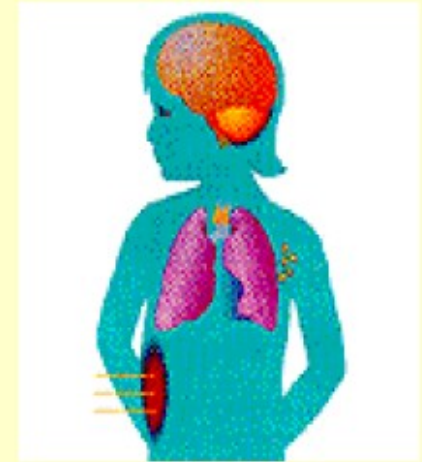


Molecular  
Structure

Biochemical  
Function



Phenotype  
(Symptoms)



# Επίλυση προβλημάτων με την Πληροφορική

1. Ανάγνωση ακολουθιών DNA (Sequencing & sequence assembly).
2. Σύγκριση ακολουθιών DNA (Sequence alignment).
3. Κατηγοριοποίηση πρωτεϊνών.
4. Εξαγωγή πληροφοριών σχετικά με τη δομή και λειτουργία πρωτεϊνών από τις αντίστοιχες ακολουθίες.
5. Μοντελοποίηση ενός κυττάρου για καλύτερη κατανόηση συγκεκριμένων διεργασιών.

# Τομείς έρευνας στη Βιοπληροφορική

1. Αποδοτική οργάνωση των δεδομένων ώστε να είναι δυνατή η αποθήκευση, ανάκτηση και ενημέρωσή τους.
2. Ανάπτυξη εργαλείων που επιτρέπουν την ανάλυση των βιολογικών δεδομένων.
3. Ανάπτυξη εργαλείων που επιτρέπουν την ερμηνεία των αποτελεσμάτων βιολογικής σημασίας.

# Κατηγοριοποίηση/ανάλυση βιολογικών δεδομένων

- Ομαδοποίηση βάσει κοινών ομοιοτήτων:
  - Επαναλαμβανόμενα τμήματα στις ακολουθίες DNA/πρωτεϊνών
  - Παρόμοια δράση πρωτεϊνών
  - Ομοιότητες στην τρισδιάστατη δομή πρωτεϊνών (Protein Data Bank)
- Κύριο ρόλο έχουν οι αλγόριθμοι που χρησιμοποιούνται για την εύρεση ομοιοτήτων μεταξύ μακρομορίων.
- Εφαρμογές ολοκλήρωσης διαφορετικών δεδομένων (data integration) από το χώρο της Βιολογίας:
  - Παράδειγμα: Οι τρισδιάστατες συντεταγμένες μιας πρωτεΐνης είναι περισσότερο χρήσιμες αν συνδυαστούν με πληροφορίες που αφορούν τη λειτουργία των πρωτεϊνών και τις αλληλεπιδράσεις με άλλα μόρια.

# Υλοποίηση και σχεδιασμός υπολογιστών εργαλείων για αυτόματη ανάκτηση γνώσης από βάσεις βιολογικών δεδομένων

- **Αναγκαιότητα:** Επεξεργασία και ανάλυση βιολογικών δεδομένων
- Ανάπτυξη κατάλληλων εργαλείων διαχείρισης και ανάλυσης των αποθηκευμένων πληροφοριών.
- **Πρόβλημα:** Η διαχείριση δεδομένων Μοριακής Βιολογίας παρουσιάζει αυξημένες απαιτήσεις.
- **ΣΤΟΧΟΙ:**
  - Σχεδιασμός και υλοποίηση ενός μοντέλου που να ικανοποιεί τις απαιτήσεις της έρευνας και, κυρίως, την αυτόματη ανάκτηση γνώσης (automated knowledge discovery) από μεγάλο πλήθος πληροφοριών χρησιμοποιώντας τεχνικές ομαδοποίησης των δεδομένων.
  - Σχεδιασμός αποδοτικών τεχνικών και μεθόδων που ανιχνεύουν την ομοιότητα μεταξύ **2D** ή **3D** σχημάτων.

# Τομείς Έρευνας

1. Αποδοτική οργάνωση των δεδομένων ώστε να είναι δυνατή η αποθήκευση, ανάκτηση και ενημέρωσή τους.
2. Ανάπτυξη εργαλείων που επιτρέπουν την ανάλυση των βιολογικών δεδομένων.
3. Ανάπτυξη εργαλείων που επιτρέπουν την ερμηνεία των αποτελεσμάτων βιολογικής σημασίας.

# Ανάλυση ακολουθιών βιολογικών δεδομένων

- **Αντικείμενο:** Διαχείριση και ανάλυση ακολουθιών βιολογικών δεδομένων.
- **Σκοπός:** Αναγνώριση δομικών χαρακτηριστικών κρίσιμων για τη λειτουργία των ζωντανών οργανισμών.
- Κάθε μόριο DNA ή πρωτεΐνης μπορεί να θεωρηθεί ως μια συμβολοσειρά από ένα αλφάβητο:
  - 4 χαρακτήρων (DNA).
  - 20 χαρακτήρων (πρωτεΐνη).
- Περιοδικές επαναλήψεις συμβολοσειρών-μοτίβων (μοτίβο: σύνολο χαρακτήρων που εμφανίζεται παραπάνω από μία φορά σε μία ακολουθία).

# Ανάλυση ακολουθιών βιολογικών δεδομένων

- 2 κατηγορίες προβλημάτων προς λύση:
  - **Ακριβής επανάληψη μοτίβων:** Εντοπισμός των θέσεων σε μία ακολουθία που ένα μοτίβο εμφανίζεται. Δεν επιτρέπεται καμία διαφορά μεταξύ των δύο ακολουθιών.
  - **Προσεγγιστική επανάληψη μοτίβων:** Επιτρέπεται η ύπαρξη διαφορών ανάμεσα στις συγκρινόμενες ακολουθίες, που προκύπτουν από την αντικατάσταση, προσθήκη ή διαγραφή συμβόλων.
- Συχνά, ευθυγραμμίζουμε τις αλληλουχίες (alignment) με σκοπό να τις συγκρίνουμε.
- **Ολική ευθυγράμμιση (Global alignment):** Μελετάμε ολόκληρες τις ακολουθίες.
- **Τοπική ευθυγράμμιση (Local alignment):** Μελετάμε μόνο ένα κομμάτι των αλληλουχιών.



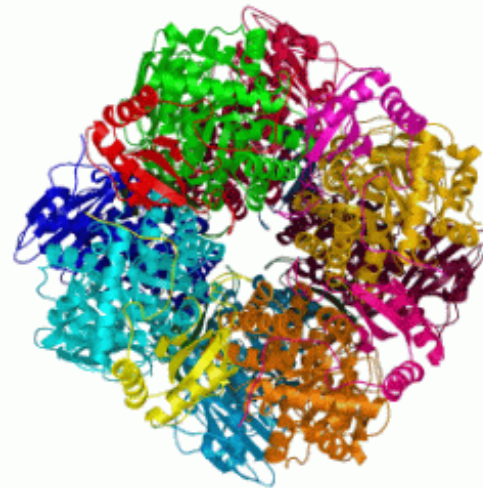
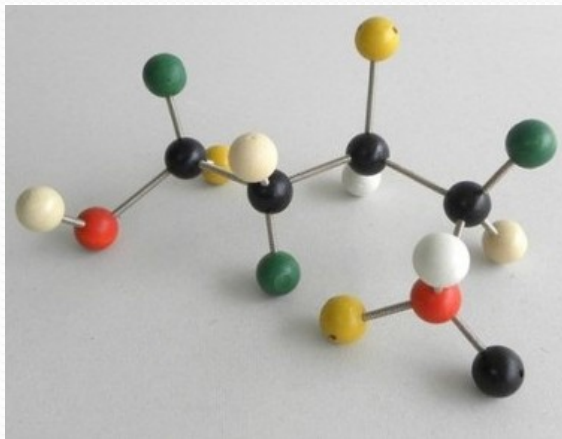
# Τομείς έρευνας

1. Αποδοτική οργάνωση των δεδομένων ώστε να είναι δυνατή η αποθήκευση, ανάκτηση και ενημέρωσή τους.
2. Ανάπτυξη εργαλείων που επιτρέπουν την ανάλυση των βιολογικών δεδομένων.
3. Ανάπτυξη εργαλείων που επιτρέπουν την ερμηνεία των αποτελεσμάτων βιολογικής σημασίας.

# Μοριακή μοντελοποίηση

## Molecular modelling

- Αναπτυσσόμενος επιστημονικός κλάδος που συνδυάζει τις επιστήμες της Βιολογίας και της Πληροφορικής
- **Σκοπός:** Προσπαθεί να μιμηθεί τη συμπεριφορά των μοριακών συστημάτων, βασιζόμενη σε μεγάλο βαθμό στη σχεδίαση μοντέλων μορίων με τη βοήθεια του Η/Υ.



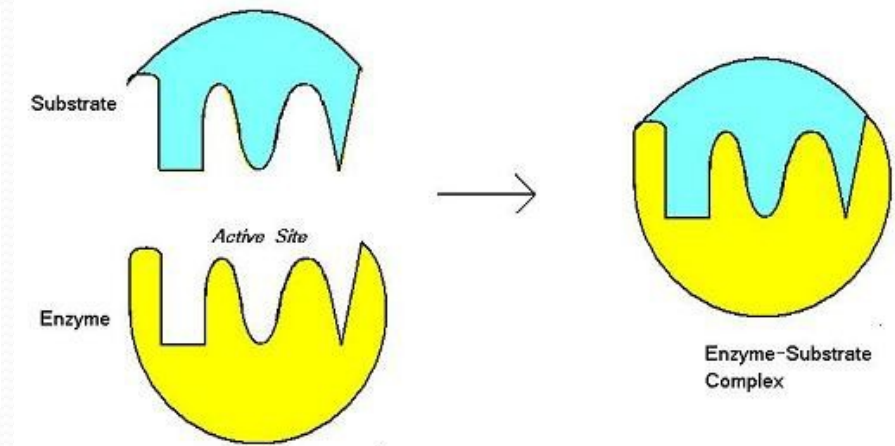
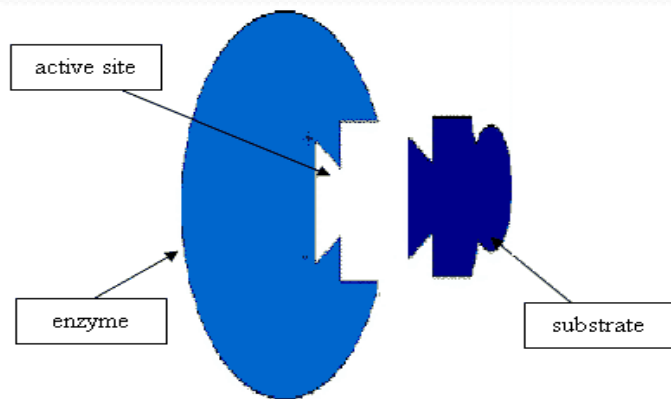
# Μοριακή μοντελοποίηση

## Τρία βασικά βήματα

- 1. Επιλογή του κατάλληλου μοντέλου.** Περιγράφει ικανοποιητικά τις ενδομοριακές και εσωμοριακές συσχετίσεις του μορίου.
- 2. Υπολογισμός της ενεργειακής κατάστασης του συστήματος και ελαχιστοποίηση της.**
- 3. Ανάλυση των παραπάνω υπολογισμών και έλεγχος της τελικής διαμόρφωσης.** Ικανοποίηση όλων των συνθηκών και περιορισμών που ο σχεδιαστής έχει θέσει.

# Ανάλυση πρωτεϊνών

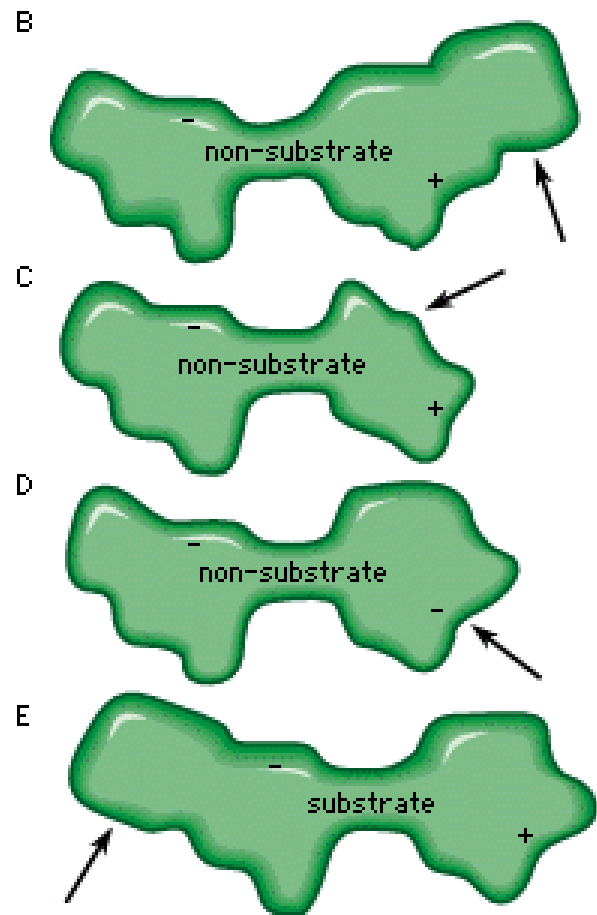
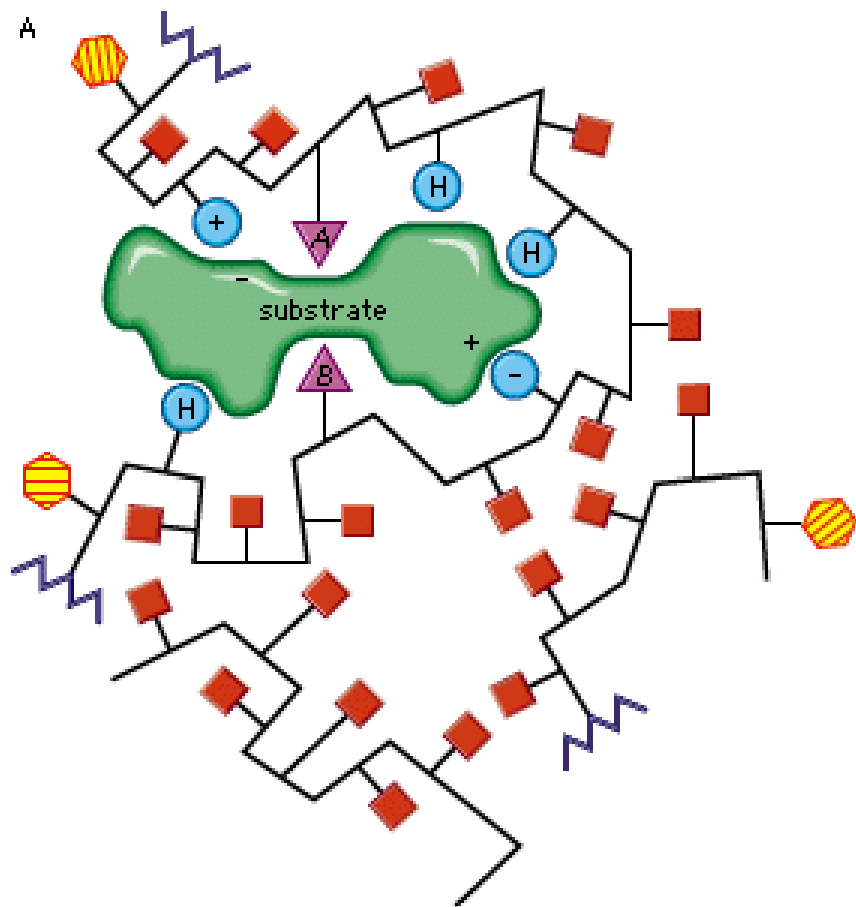
- Οι πρωτεΐνες περιγράφονται από την αμινοξική τους ακολουθία. Ωστόσο, οι ειδικές λειτουργίες τους εξαρτώνται από την τρισδιάστατη δομή τους. Είναι αυτό αρκετό;
- Πολλές πρωτεΐνες-ένζυμα μετατρέπονται σε ενεργά μόρια με βάση μία μικρή περιοχή στην επιφάνειά τους, την ενεργό περιοχή (active site), ή ενεργό κέντρο του ενζύμου.
- Το ενεργό κέντρο έχει γεωμετρικά και φυσικοχημικά χαρακτηριστικά που είναι συμπληρωματικά ενός άλλου μορίου, του υποστρώματος (substrate).



# Ανάλυση πρωτεϊνών

- Ένζυμο και υπόστρωμα για να ενωθούν πρέπει να βρεθούν στην ευνοϊκότερη διαμόρφωση.
- Προκειμένου να εντοπίσουμε την ευνοϊκότερη διαμόρφωση απαιτείται να μελετήσουμε τις επιφάνειες επαφής.
- Οι περισσότεροι αλγόριθμοι βασίζονται στην:
  - Αρχή συμπληρωματικότητας ως προς τη δομή.
  - Αρχή συμπληρωματικότητας ως προς την ηλεκτρονική διαμόρφωση.
- Προκειμένου να βρούμε τα συμπληρωματικά μόρια που ανήκουν σε δύο πρωτεΐνες A & B, έχουμε να επιλύσουμε ένα πρόβλημα ταιριάσματος σε 3 διαστάσεις (**3D matching problem**).

# Ανάλυση πρωτεϊνών



©1997 Encyclopaedia Britannica, Inc.

# Σχετικό πεδίο

## Υπολογιστική Βιολογία

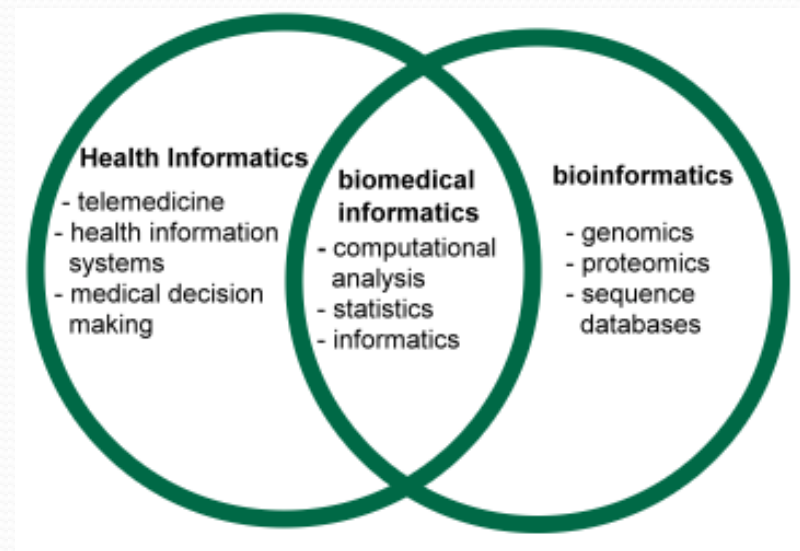
## Computational Biology

- Η μελέτη και η εφαρμογή υπολογιστικών μεθόδων στην κλασική βιολογία.
- Κύριο αντικείμενο μελέτης η εξέλιξη των ειδών (πληθυσμιακή και θεωρητική βιολογία) και όχι η μελέτη σε κυτταρικό και μοριακό επίπεδο.

# Σχετικό πεδίο

## Πληροφορική Υγείας Health Informatics

- Η μελέτη και η εφαρμογή υπολογιστικών μεθόδων για τη βελτίωση της επικοινωνίας, της κατανόησης και της διαχείρισης ιατρικών δεδομένων.
- Κύριο αντικείμενο μελέτης ο χειρισμός/αξιοποίηση της πληροφορίας και όχι η ίδια η πληροφορία.





# Σχετικό πεδίο Chem-informatics

- Η μελέτη και η εφαρμογή υπολογιστικών μεθόδων, σε συνδυασμό με χημικές και βιολογικές τεχνικές, για τον σχεδιασμό και ανάπτυξη φαρμάκων.

# Σχεδιασμός φαρμάκων με τη βοήθεια Η/Υ

## Μεθοδολογία:

1. Αποτελεσματική απεικόνιση των δομών κανονικών και παθολογικών μορίων.
2. Σύγκριση με μόρια-ενεργοποιητές, για την αναγνώριση περιοχών που είναι υπεύθυνες για τη δράση των μορίων.
3. Προσομοίωση της ενεργοποίησης του παθολογικού μορίου από το μόριο-ενεργοποιητή στην οθόνη του Η/Υ.
4. Σχεδίαση και ανάπτυξη μορίων-φαρμάκων με δυνατότητα πρόσδεσης στο παθολογικό μόριο, με σκοπό την αδυναμία ενεργοποίησης του.
5. Επιβεβαίωση προβλέψεων με στοχευμένα βιολογικά πειράματα.

# Σχετικό πεδίο Geno-mics

- Γονιδίωμα ονομάζεται το σύνολο των γονιδίων ενός οργανισμού.
- Ανάλυση και σύγκριση γονιδιωμάτων ενός ή πολλών ειδών μεταξύ τους.
- Το πεδίο έρευνας «Genomics» υπήρχε και πριν από την αποκρυπτογράφηση γονιδιωμάτων διάφορων οργανισμών (π.χ. ανθρώπου, κουνουπιού), αλλά σε πολύ αρχικό στάδιο.

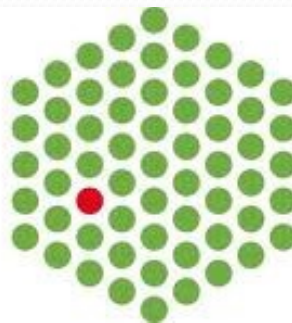
# Σχετικό πεδίο Proteo-mics

- Μελέτη των πρωτεϊνών που παράγονται από γονιδίωμα.
- Τρόποι λειτουργίας και αλληλεπίδρασης των πρωτεϊνών.
- Ενδιαφέρον για την πραγματική κατάσταση των πρωτεϊνών στα κύτταρα και όχι για το σύνολο των πιθανών καταστάσεων που περιγράφεται από το γονιδίωμα.

# Η έρευνα στον τομέα της Βιοπληροφορικής

- **EMBL** – European Molecular Biology Laboratory.
- **NCBI** – National Center for Biothechnology.

EMBL



# Βιβλιογραφικές και βιολογικές βάσεις δεδομένων

- Περιγράφονται οι βιβλιογραφικές βάσεις δεδομένων και οι μηχανές αναζήτησής τους (Scholar Google, Scopus, PubMed).
- Περιγράφονται οι βιολογικές βάσεις δεδομένων και οι μηχανές αναζήτησής τους.
- Ανάλυση της χρησιμότητας των βιολογικών βάσεων δεδομένων στην ενίσχυση των επιστημών για την κατανόηση και ερμηνεία ενός πλήθους βιολογικών φαινομένων.

# Ανάλυση και στοίχιση ακολουθιών

- Πλήθος διαθέσιμων ακολουθιών (κομμάτια DNA ή πρωτεΐνης).
- Αδύνατη ανάλυση με «το χέρι».
- Απαραίτητη η ηλεκτρονική ανάλυση και στοίχιση ακολουθιών.
  - Αλγόριθμοι δυναμικού προγραμματισμού.
  - Ευρετικοί αλγόριθμοι.
- Οι στοιχίσεις μεταξύ ακολουθιών μας δείχνουν ποια κομμάτια είναι μεταβλητά και ποια συντηρούνται.
- Αναγνωρίζονται θέσεις πιθανών μεταλλάξεων, προσθήκης, αφαίρεσης ή και αντικατάστασης μέρους μιας αλληλουχίας.

|                    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|--------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Scarites           | C | T | T | A | G | A | T | C | G | T | A | C | C | A | A | - | - | - | A | A | T | A | T | T | A | C |
| Carenum            | C | T | T | A | G | A | T | C | G | T | A | C | C | A | C | A | - | T | A | C | - | T | T | T | A | C |
| Pasimachus         | A | T | T | A | G | A | T | C | G | T | A | C | C | A | C | T | A | T | A | A | G | T | T | T | A | C |
| Pheropsophus       | C | T | T | A | G | A | T | C | G | T | T | C | C | A | C | - | - | - | A | C | A | T | A | T | A | C |
| Brachinus armiger  | A | T | T | A | G | A | T | C | G | T | A | C | C | A | C | - | - | - | A | T | A | T | A | T | T | C |
| Brachinus hirsutus | A | T | T | A | G | A | T | C | G | T | A | C | C | A | C | - | - | - | A | T | A | T | A | T | A | C |
| Aptinus            | C | T | T | A | G | A | T | C | G | T | A | C | C | A | C | - | - | - | A | C | A | A | T | T | A | C |
| Pseudomorpha       | C | T | T | A | G | A | T | C | G | T | A | C | C | - | - | - | - | - | A | C | A | A | A | T | A | C |



# Φυλογενετική ανάλυση Κατασκευή φυλ. Δέντρων

- **Εξελεκτική βιολογία:**
  - Μελέτη προέλευσης και καταγωγής των ειδών.
  - Αλλαγή των έμβιων όντων με το πέρασμα του χρόνου.
- **Βιοπληροφορική:**
  - Αποκρυπτογραφεί τους ρυθμούς εξέλιξης γονιδίων με την καταγραφή των αλλαγών στο DNA
  - Προβλέπει την έκβαση κάθε συστήματος με το πέρασμα του χρόνου.



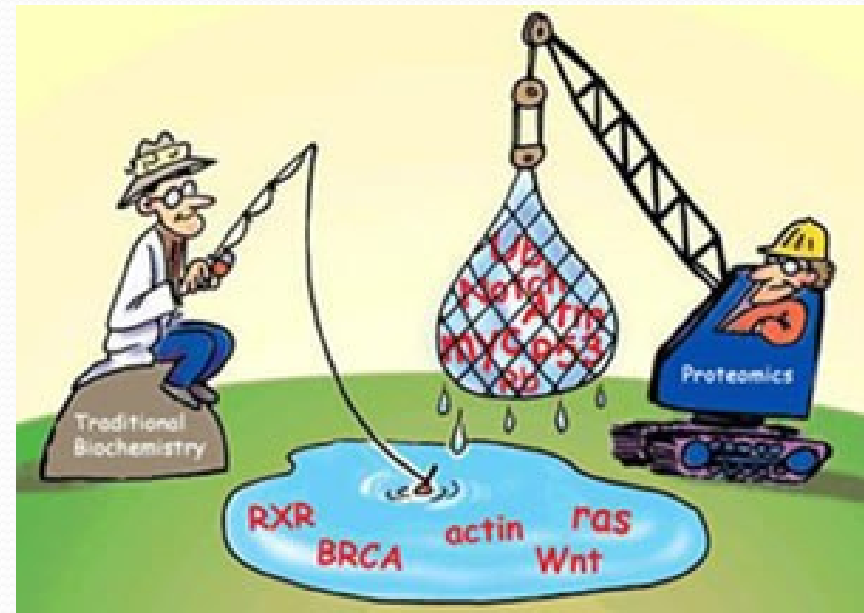
Reproduction rights obtainable from  
[www.CartoonStock.com](http://www.CartoonStock.com)

search ID: mlyn190

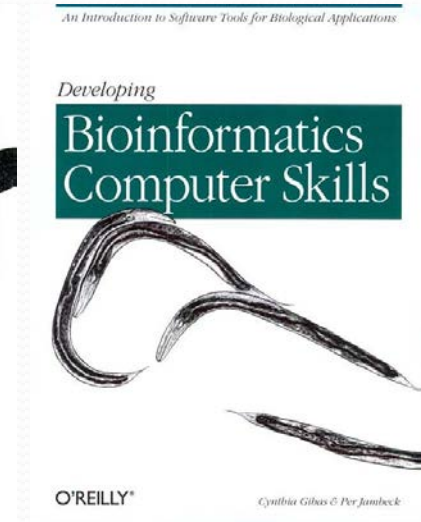
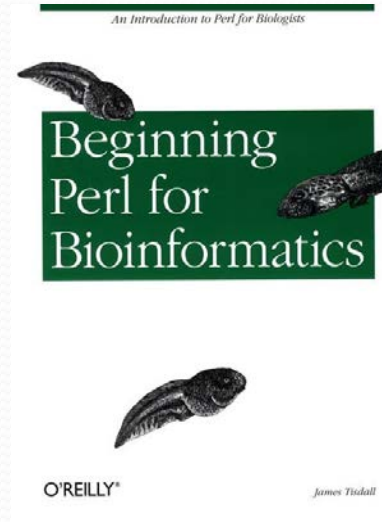
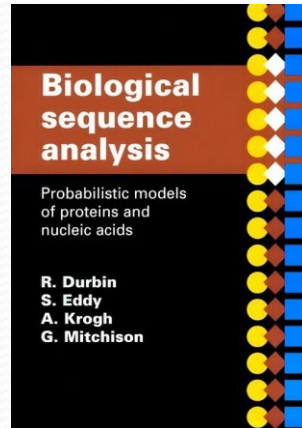
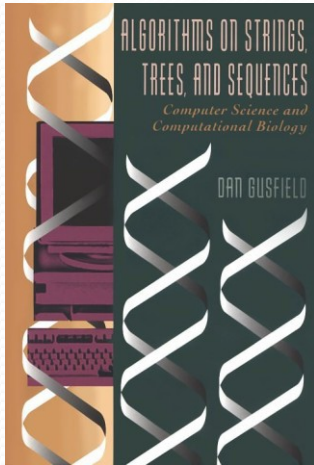


# Φυλογενετική ανάλυση Κατασκευή φυλ. Δέντρων

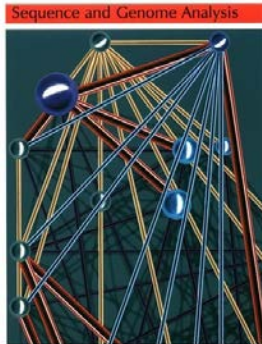
- **Πρωτεϊνική ανάλυση:**
  - Πρόγνωση πρωτεϊνικής δομής από την αλληλουχία τους.
  - Οπτικοποίηση της δομής των πρωτεϊνών.
  - Πρόβλεψη πιθανών αλληλεπιδράσεων μεταξύ των πρωτεϊνών.
- **Πρωτεομική (Proteomics):**
  - Ανάλυση της δομής και λειτουργίας των πρωτεϊνών.



# Βιβλία σχετικά



## Bioinformatics



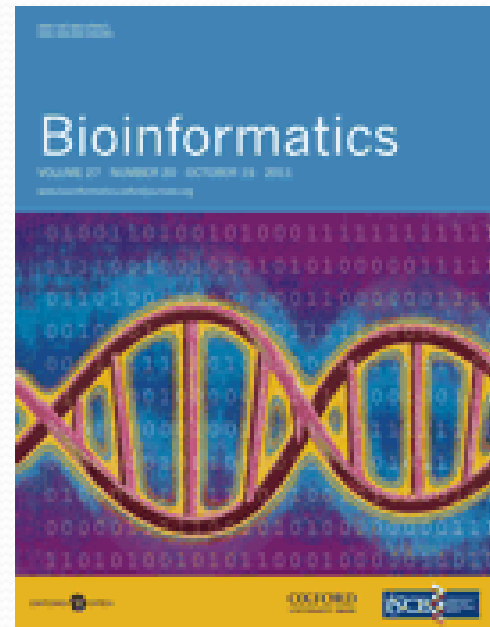
David W. Mount



# Τα κυριότερα επιστημονικά περιοδικά

- Bioinformatics, Oxford University Press.
- IEEE/ACM transactions on Computational Biology and Bioinformatics (TCBB).
- Journal of Computational Biology).

Science, Nature, Nucleic Acid Research, Journal of Molecular Biology, Proceedings of the National Academy of Sciences (PNAS).



# Άλλα χρήσιμα βιβλία

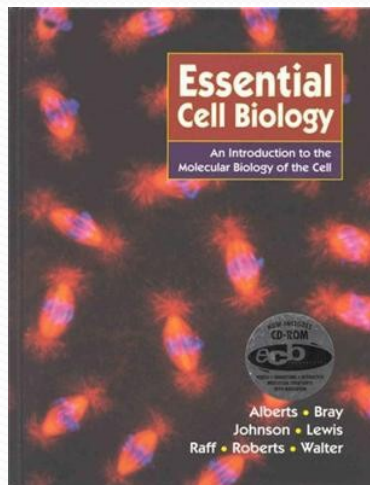
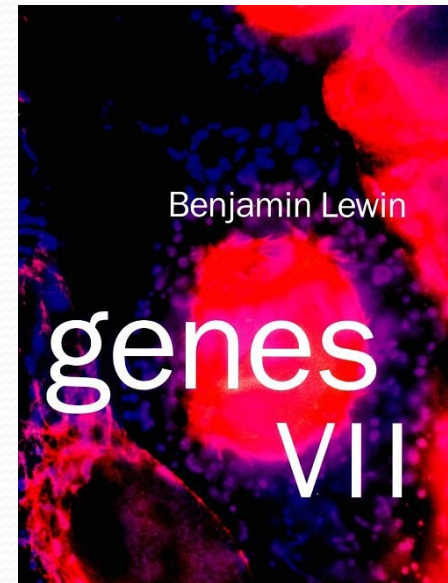
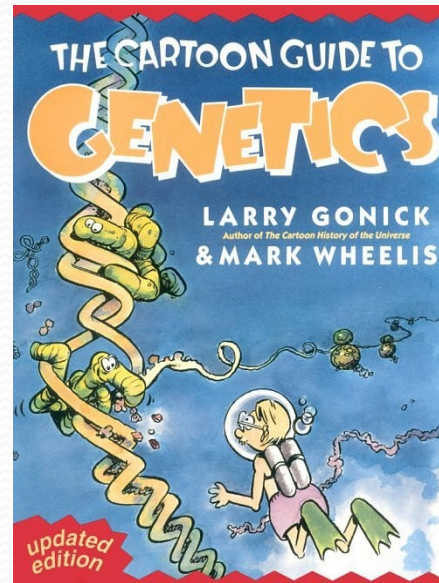
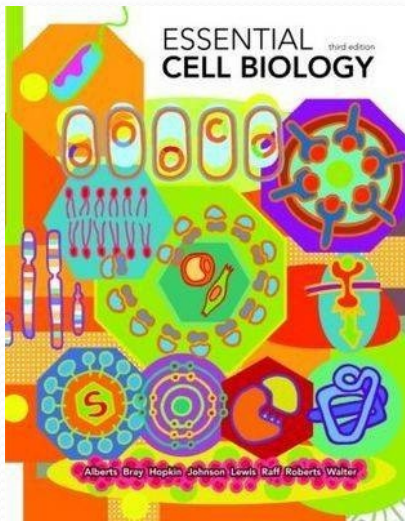


Image Source: <http://www.amazon.com/>

# Θέματα

- **Sequence alignment** (pairwise sequence alignment, spliced alignment and similarity based gene recognition, multiple sequence alignment, parametric sequence alignment).
- **Bioinformatic databases and data mining** (string search in external memory , index structures for approximate matching in sequence databases, algorithms for motif search, data mining in computational biology).
- **Genome scale computational methods** (comparison of long genomic sequence, algorithms and applications, chaining algorithms and applications in comparative genomics, computational analysis of alternative splicing, human genetic linkage analysis, haplotype inference).
- **Phylogenetics** (phylogenetic reconstruction, consensus trees and supertrees, large scale phylogenetic analysis, high performance phylogeny reconstruction).



# Θέματα

- **Microarrays and gene expression analysis** (microarray data: annotation retrieval, storage and communication, computational methods for microarray design, clustering algorithms for gene expression analysis, biclustering algorithms, identifying gene regulatory networks from gene expression data, modeling and analysis of gene networks using feedback control analysis).
- **Computational Structural Biology** (predicting protein structure and supersecondary structure, protein structure prediction with lattice models, proteins structure determination via NMR spectral data, geometric and signal processing of reconstructed 3D maps of molecular complexes, in search of remote homologs, biomolecular modeling using parallel supercomputers).