

# Βιοπληροφορική

Διάλεξη 1η :

Οργάνωση, στόχος, εξετάσεις,  
βαθμολογία, σύγγραμμα, σχέδιο μαθήματος,  
εισαγωγή (ορισμοί, θέση στη σύγχρονη βιολογία).

# Οργάνωση μαθήματος

---

## Διαλέξεις (12 ?)

- Η παρακολούθηση των διαλέξεων είναι αληθώς προαιρετική.
- Τα slides των διαλέξεων είναι διαθέσιμα μέσω δικτύου [ως PDF (portable document format) αρχεία].
- Η κεντρική ιστοσελίδα του μαθήματος είναι διαθέσιμη μέσω e-Class :

<http://eclass.duth.gr/>

# Οργάνωση μαθήματος

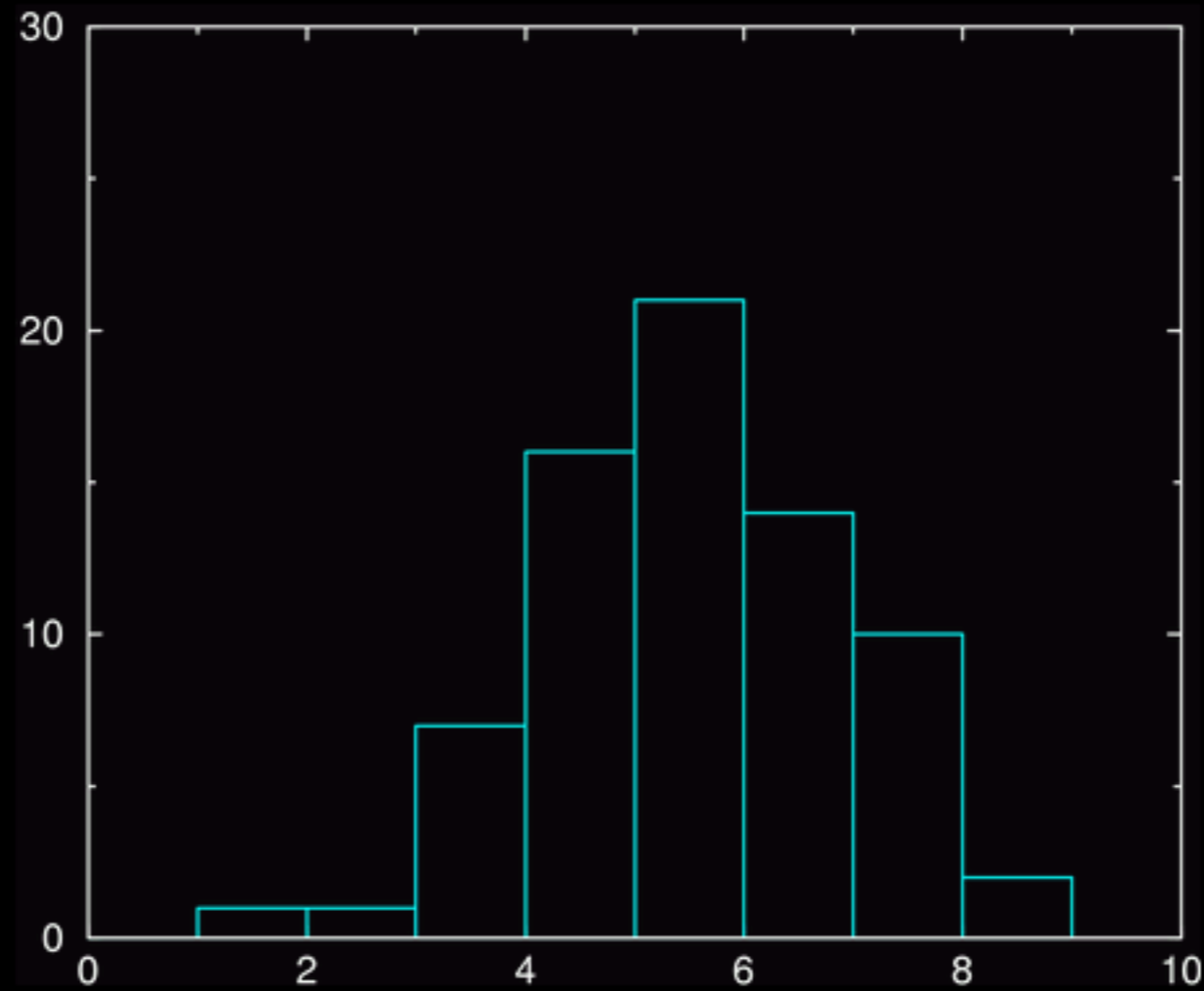
---

Εργαστηριακές ασκήσεις (2-3 ?)

- Η συμμετοχή στις εργαστηριακές ασκήσεις είναι αληθώς υποχρεωτική.
- Λόγω της σχεδόν πλήρους απουσίας δικτύου και την (δυσ)αναλογία φοιτητών/υπολογιστών (400/15) οι ασκήσεις θα είναι λίγες.
- Λόγω της (δυσ)αναλογίας φοιτητών/προσωπικού απουσίες δεν επιτρέπονται και επαναληπτικά εργαστήρια δεν προβλέπονται.

# Στόχος μαθήματος

Να το περάσετε.



# Βαθμολογία

---

Εργαστήρια : 30%

Εξεταστική : 70% του τελικού βαθμού.

- Για να περάσετε το μάθημα θα πρέπει (α) να γράψετε κατ'ελάχιστο 5 στις εξετάσεις Ιανουαρίου (ασχέτως του βαθμού του εργαστηρίου), (β) ο τελικός βαθμός του μαθήματος (Εργαστήρια+Εξεταστική) θα πρέπει επίσης να είναι κατ'ελάχιστο 5.

- Το εργαστήριο δεν θα έχει ξεχωριστές εξετάσεις και ο βαθμός του προκύπτει από την (γραπτή) απόδοσή σας στις επιμέρους ασκήσεις.

# Βαθμολογία (Σενάρια)

Εργαστήριο : 10/10

Εξεταστική : 4/10 → Σεπτέμβριος

Εργαστήριο : 10/10

Εξεταστική : 5/10 → 6.5 (7)

Εργαστήριο : 0/10 (απών/απούσα)

Εξεταστική : 5/10 → Σεπτέμβριος

Εργαστήριο : 0/10 (απών/απούσα)

Εξεταστική : 6/10 → Σεπτέμβριος

Εργαστήριο : 0/10 (απών/απούσα)

Εξεταστική : 7/10 → 5

# Εξετάσεις

---

~30 θέματα τύπου multiple choice σε περίπου 15 διαφορετικά set. Όλα τα sets θα έχουν τις ίδιες ερωτήσεις και πιθανές απαντήσεις μόνο που η σειρά τους (τόσο των ερωτήσεων όσο και των πιθανών απαντήσεων) θα είναι τυχαία.

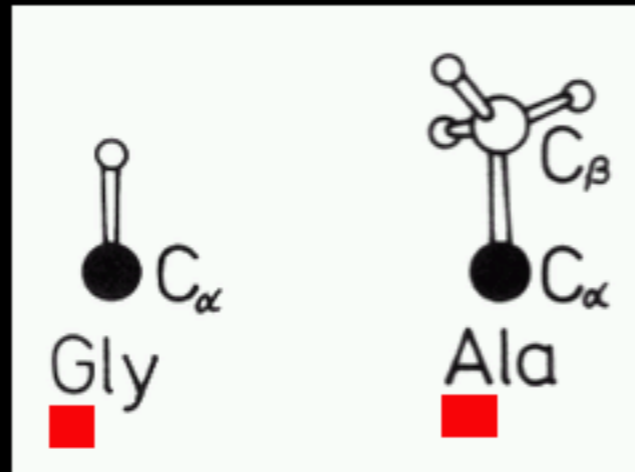
# Εξετάσεις

Έστω δυο αλληλουχίες, οι QKHVRTLYRDF και NSKRMSKYHDI, και ένας πίνακας βαθμολόγησης τέτοιος ώστε (α) εάν δυο κατάλοιπα είναι ίδια, η τιμή του πίνακα είναι 4, (β) εάν δύο κατάλοιπα είναι φορτισμένα (συμπεριλαμβανομένης και της ιστιδίνης στα φορτισμένα κατάλοιπα) η τιμή του πίνακα είναι 3, και, (γ) μηδέν σε κάθε άλλη περίπτωση. Εάν το gap penalty έχει τιμή 1, ποιά είναι η βαθμολογία της βέλτιστης στοίχισης αυτών των δυο αλληλουχιών κατά Needleman & Wunsch ;

1. 20
2. 21
3. Καμία από αυτές.
4. 17
5. 19

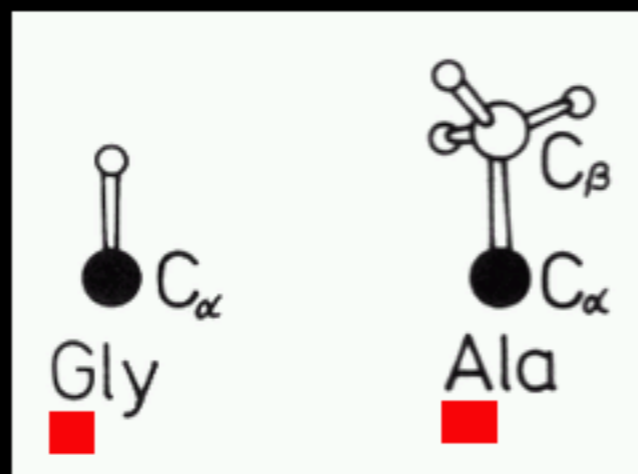


# Gly (G), Ala (A)



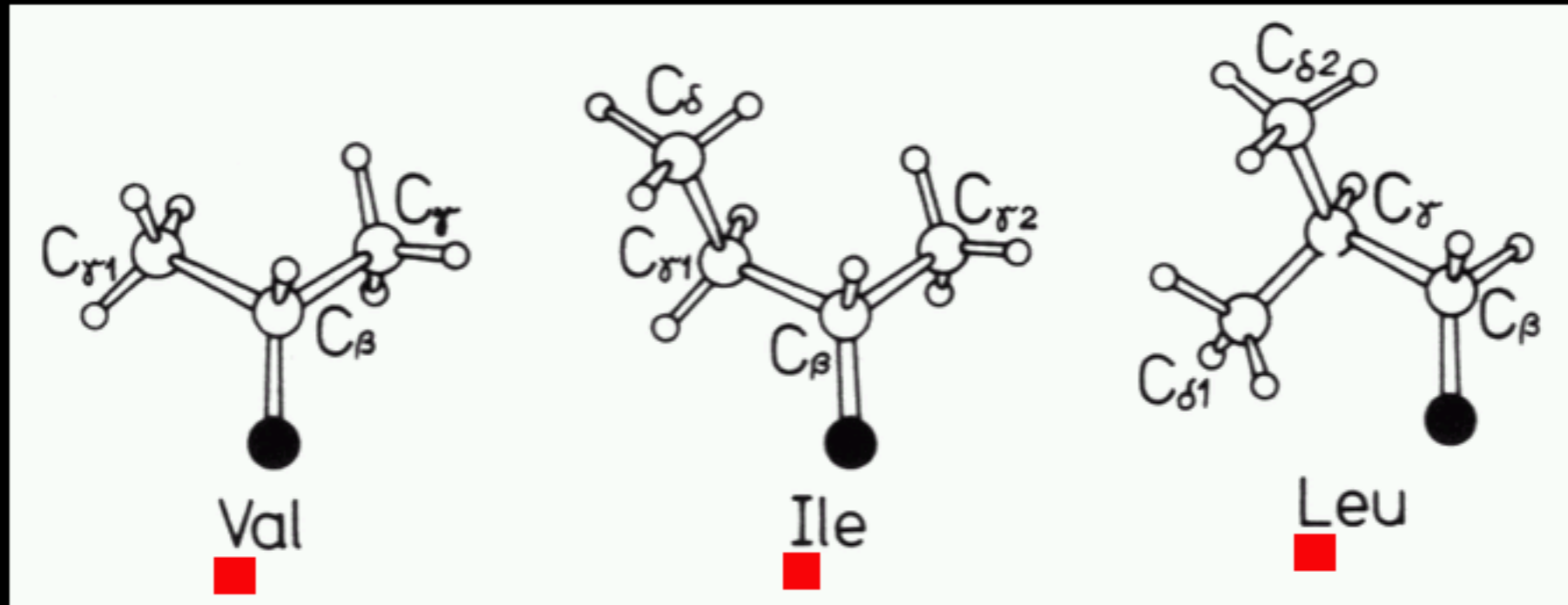
Η γλυκίνη είναι το μοναδικό αμινοξύ χωρίς πλευρική ομάδα (μόνο ένα άτομο υδρογόνου). Η απουσία πλευρικής ομάδας (και συνεπώς η απουσία στεreoχημικών παρεμποδίσεων με την κυρίως αλυσίδα) της επιτρέπουν να υιοθετεί διαμορφώσεις οι οποίες είναι ενεργειακά ανέφικτες για τα υπόλοιπα αμινοξέα (π.χ. απότομες μεταβολές στη διεύθυνση της πολυπεπτιδικής αλυσίδας ('σπασίματα', kinks)). Για το λόγο αυτό οι εξελικτικά συντηρημένες γλυκίνες είναι αρκετά συχνές.

# Gly (G), Ala (A)



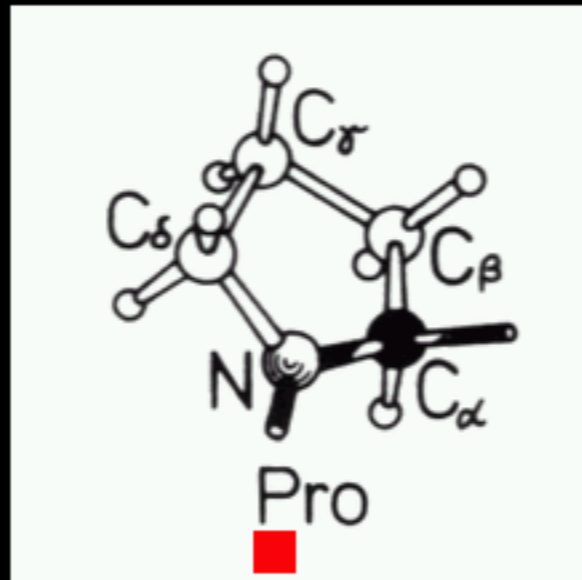
Η αλανίνη έχει τη δεύτερη μικρότερη πλευρική ομάδα, μια μεθυλομάδα. Είναι ένα μη πολικό κατάλοιπο το οποίο παρ'όλα αυτά απαντάται αρκετά συχνά και στην επιφάνεια των πρωτεϊνών. Η συχνότητα της Ala στις πρωτεϊνικές αλληλουχίες είναι αρκετά υψηλή, ίσως λόγω του μικρού ενεργειακού κόστους σύνθεσης της.

# Val (V), Ile (I), Leu (L)



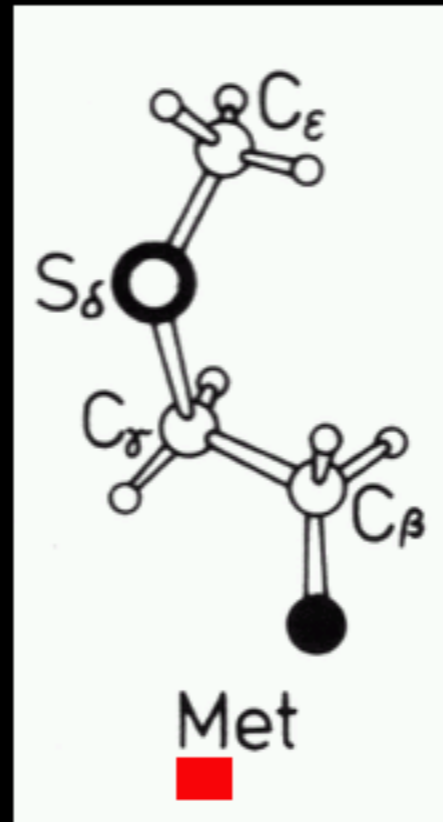
Αυτά είναι και τα τρία υδρόφοβα αμινοξέα τα οποία χαρακτηρίζονται από την διακλαδιζόμενη πλευρική τους αλυσίδα. Η ύπαρξη διακλάδωσης (ιδιαίτερα για τις Val και Ile που διακλαδίζονται στο C $\beta$ ) περιορίζει αισθητά τις προτιμώμενες διαμορφώσεις τους (γνωστές ως rotamers).

# Pro (P)



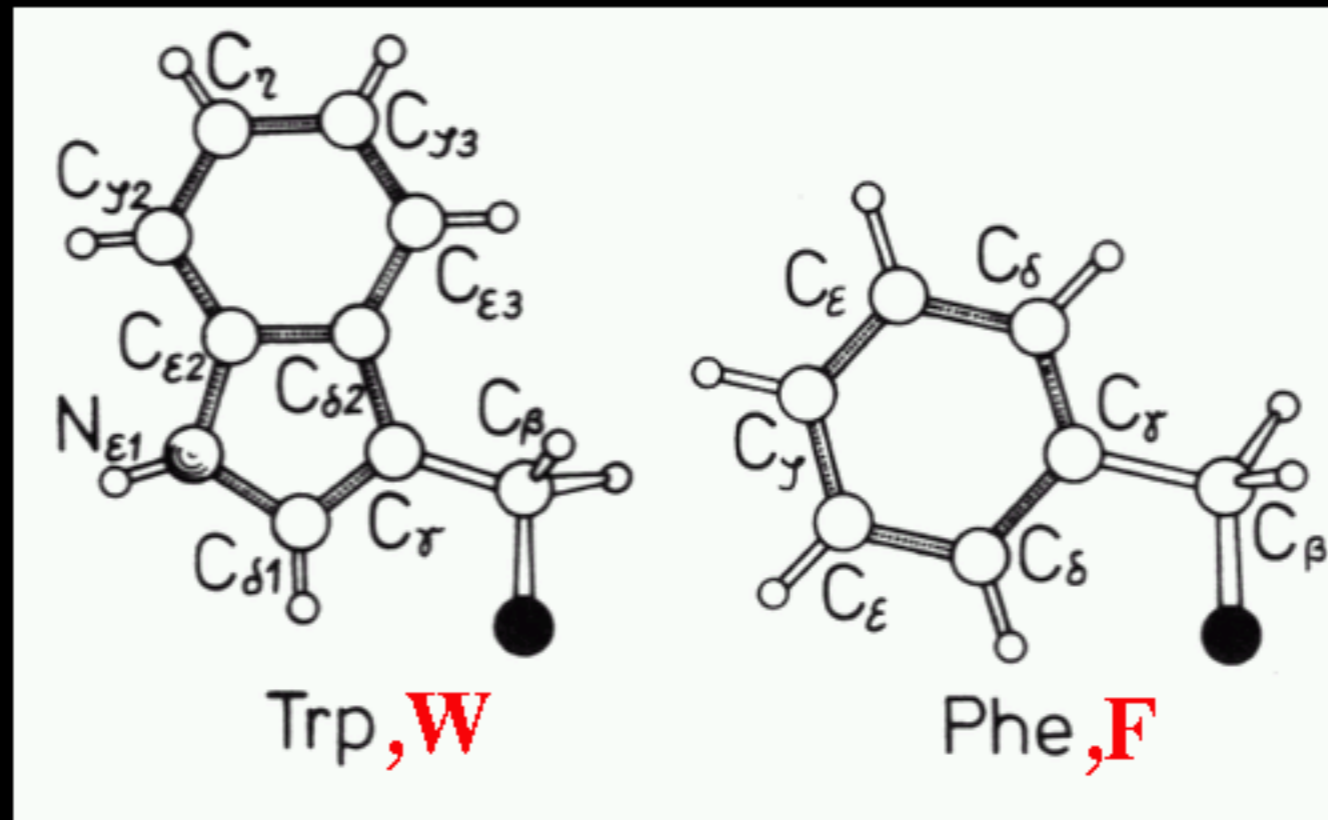
Η προλίνη είναι ένα ιδιαίτερο αμινοξύ : Η πλευρική ομάδα επιστρέφει και συνδέεται ομοιοπολικά με το σκελετό της πρωτεΐνης. Η πιο άμεση συνέπεια είναι ότι αυτή η πλευρική ομάδα είναι πρακτικά ακίνητη (ο μοναδικός βαθμός ελευθερίας βρίσκεται στη διαμόρφωση του δακτυλίου). Η δεύτερη συνέπεια είναι ότι η δίεδρη γωνία ανάμεσα στο NH και το Cα περιορίζεται σε ένα πολύ μικρό εύρος ενεργειακά αποδεκτών τιμών.

# Met (M)



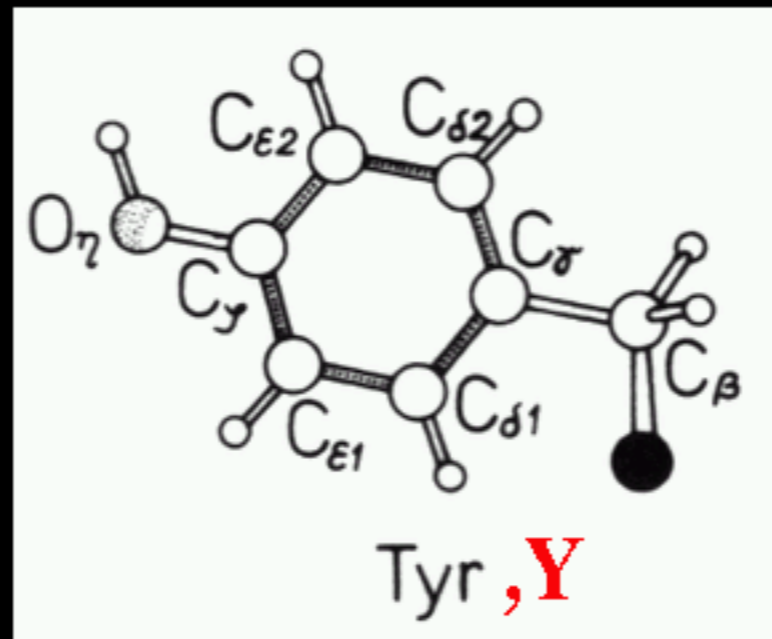
Η μεθειονίνη έχει μια μάλλον ευκίνητη πλευρική με ένα άτομο θείου σε θειο-αιθερικό δεσμό. Η διαφορά ηλεκτραρνητικότητας ανάμεσα στο θείο και τους άνθρακες οδηγεί στην δημιουργία διπολικής ροπής. Ως επί το πλείστον απαντάται στο εσωτερικό των πρωτεϊνικών δομών.

# Trp (W), Phe (F)



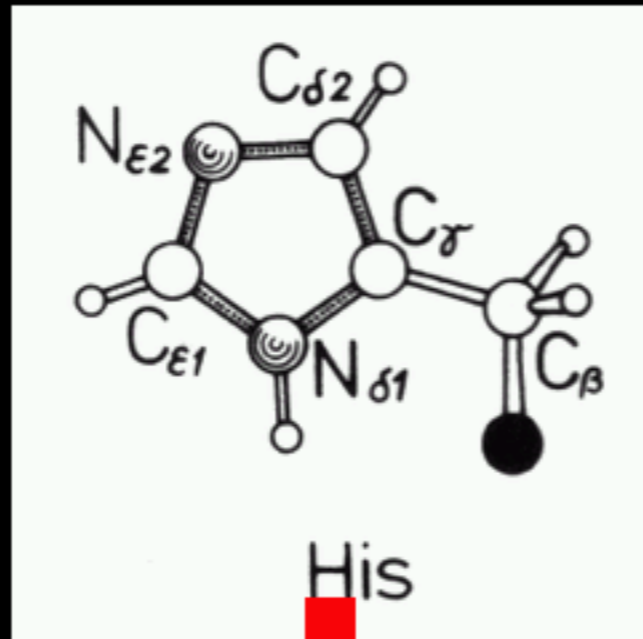
Η τρυπτοφάνη και η φαινυλαλανίνη είναι από τα πλέον υδρόφοβα αμινοξέα και ως εκ τούτου απαντώνται κατά κύριο λόγο στο εσωτερικό των πρωτεϊνών, αν και η πλευρική ομάδα της τρυπτοφάνης είναι ελαφρά πολική λόγω του ετεροκυκλικού ινδολικού δακτυλίου.

# Tyr (Y)



Η τυροσίνη έχει την ιδιαιτερότητα ότι αν και έχει αρωματικό δακτύλιο, έχει επίσης μια υδροξυλομάδα (με pK 10.1) η οποία την κατατάσσει στα πολικά αμινοξέα (και μάλιστα με δυνητικά ισχυρούς υδρογονικούς δεσμούς). Όπως τα περισσότερα αφόρτιστα αλλά πολικά αμινοξέα, η τυροσίνη μπορεί να βρεθεί τόσο στο εσωτερικό όσο και στην επιφάνεια των πρωτεϊνών.

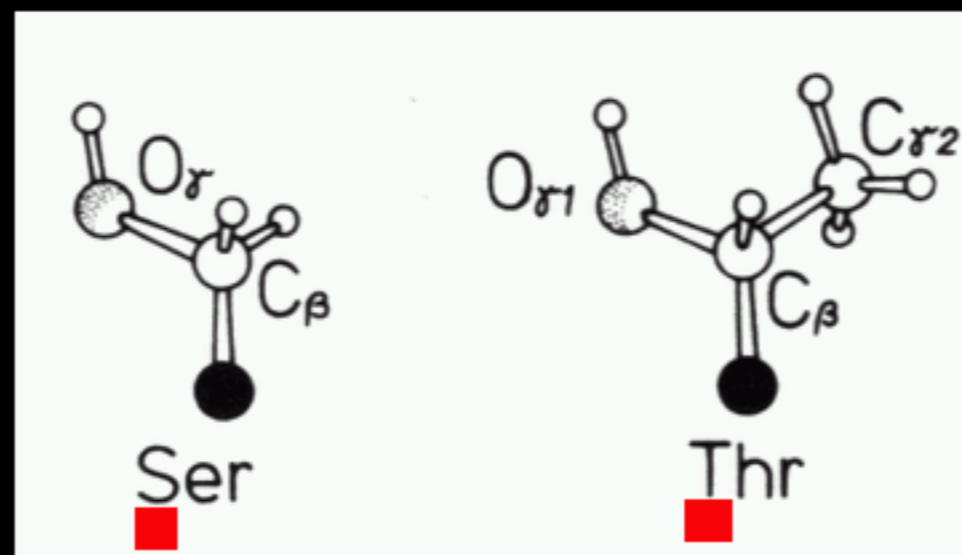
# His (H)



Η ιστοιδίνη έχει ένα ετεροκυκλικό αρωματικό δακτύλιο με μία χαρακτηριστική ιδιαιτερότητα : το  $pK$  της είναι 6, το οποίο σημαίνει ότι στα φυσιολογικά (ενδοκυτταρικά)  $pH$ , οι ιστοιδίνες μπορούν να είναι είτε αφόρτιστες ( $pH > 6$ ) ή φορτισμένες ( $pH < 6$ ). Επειδή, δε, η ακριβής τιμή του  $pK$  εξαρτάται (και μπορεί να ρυθμιστεί) από το περιβάλλον της ιστοιδίνης, αυτό είναι ένα από τα συνηθέστερα αμινοξέα σε ενζυμικά καταλυτικά κέντρα.

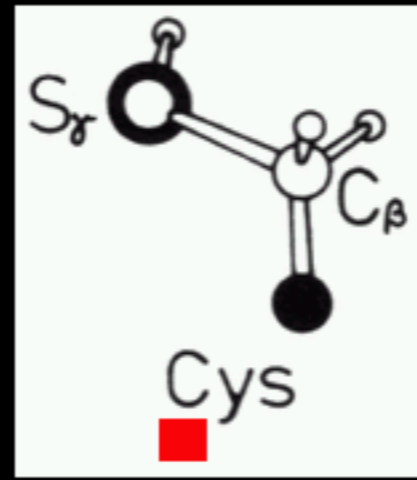


# Ser (S), Thr (T)



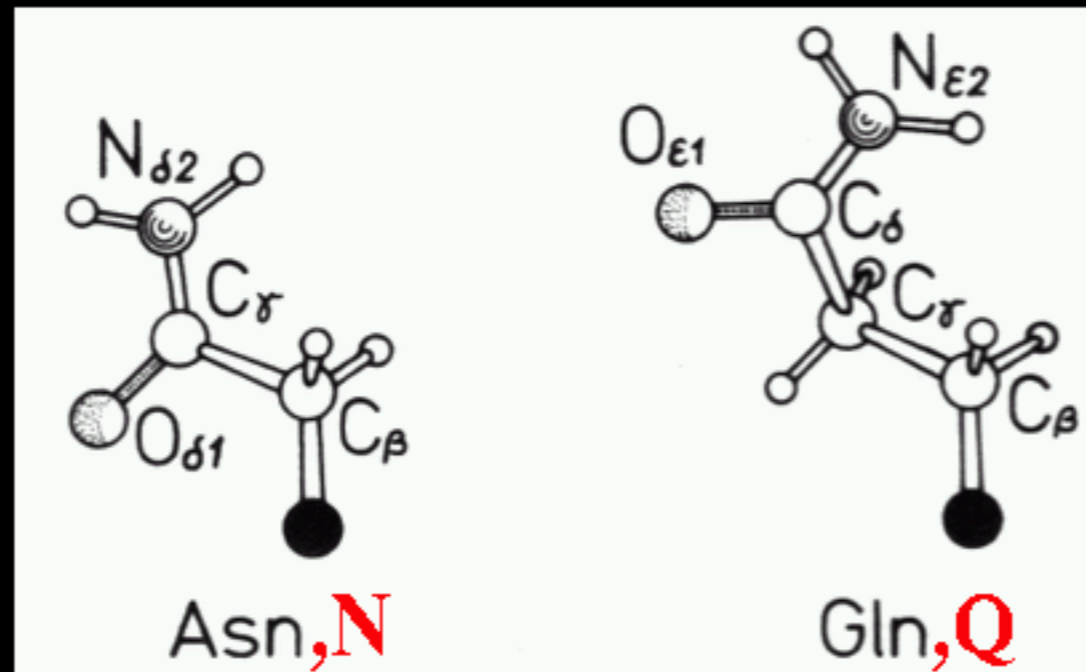
Η σερίνη και η θρεονίνη είναι τυπικά αφόρτιστα-πολικά κατάλοιπα. Έχουν και οι δύο υδροξυλομάδες οι οποίες μπορούν να συμμετάσχουν στο σχηματισμό υδρογονικών δεσμών. Το C $\beta$  της θρεονίνης είναι ασύμμετρο (και όπως συνήθως, μόνο το ένα ισομερές απαντάται σε φυσικές πρωτεΐνες).

# Cys (C)



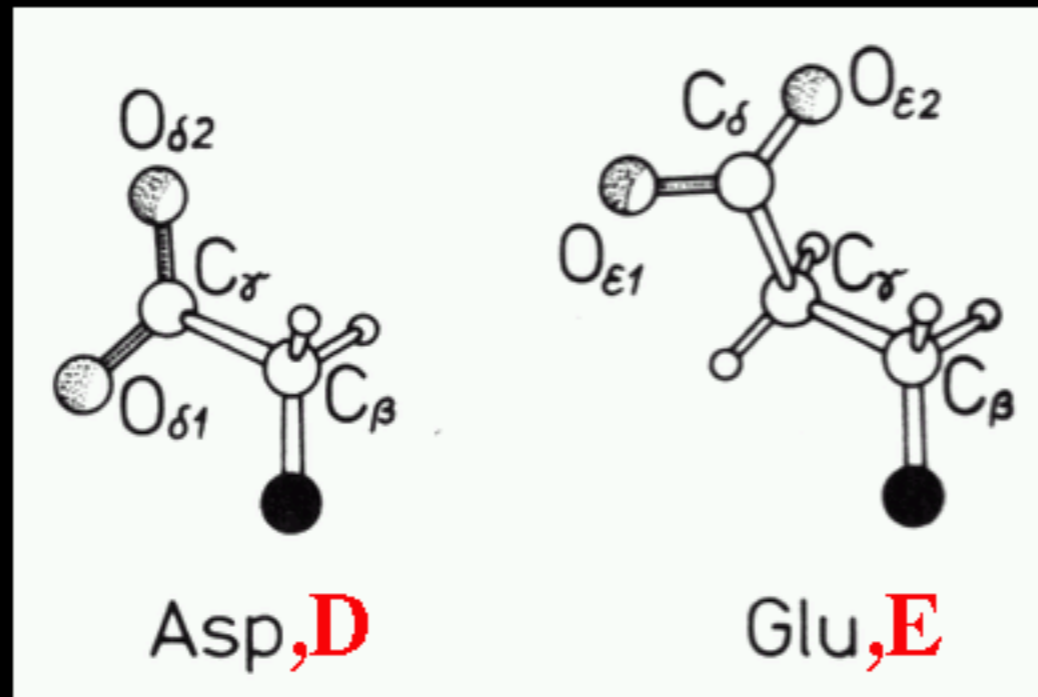
Η κυστεΐνη είναι η μοναδική πλευρική ομάδα η οποία (μέσω της οξείδωσης της σε κυστίνη) μπορεί να δημιουργήσει ομοιοπολικούς δεσμούς οι οποίοι είναι ειδικοί και χαρακτηριστικοί για μια συγκεκριμένη πρωτεϊνική δομή. Οι δισουλφιδικές γέφυρες σταθεροποιούν την τρισδιάστατη δομή και είναι ιδιαίτερα συνηθισμένες σε εξωκυτταρικές (εκκρινόμενες) πρωτεΐνες.

# Asn (N), Gln (Q)



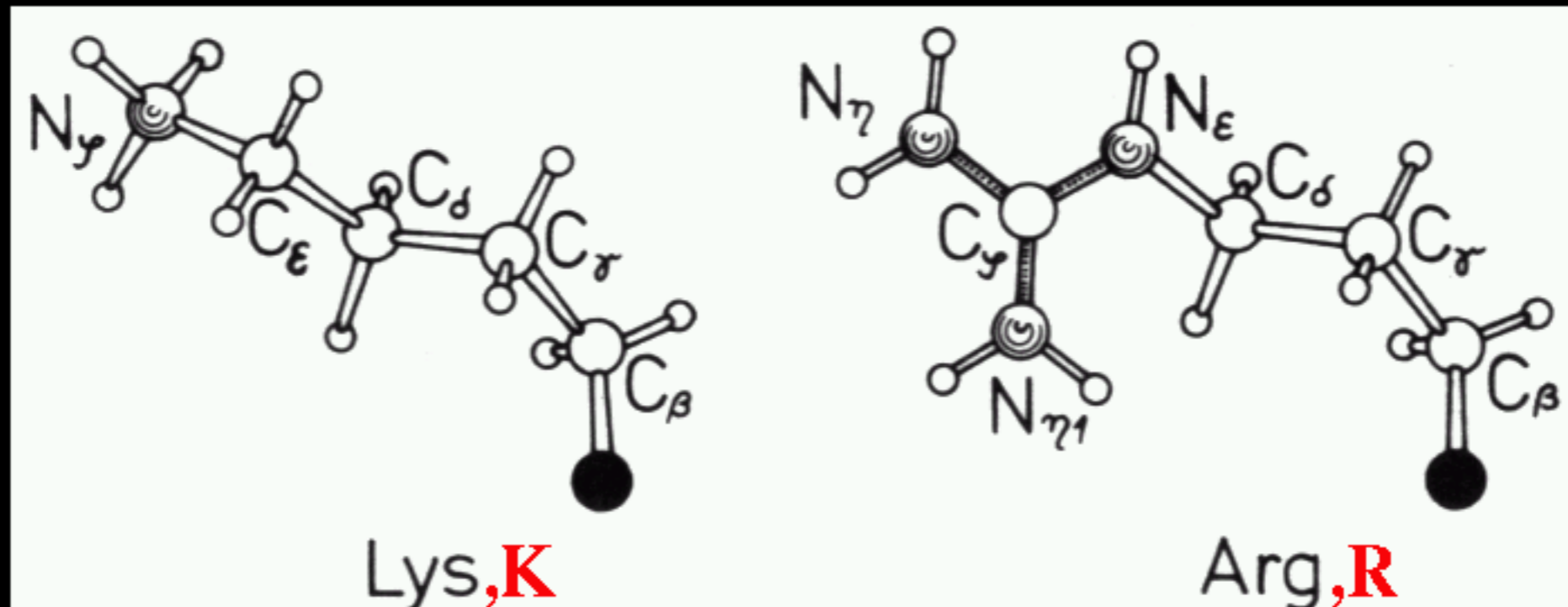
Η ασπαραγίνη και η γλουταμίνη έχουν την ιδιαιτερότητα ότι μπορούν ταυτόχρονα να είναι δότες (αμιδική ομάδα) και δέκτες (καρβονυλική ομάδα) υδρογονοδεσμών. Είναι σχετικά ευκίνητες πλευρικές ομάδες (ιδιαίτερα της γλουταμίνης) οι οποίες όταν βρίσκονται στο εσωτερικό πρωτεϊνών σχεδόν πάντα εμπλέκονται στο σχηματισμό δεσμών υδρογόνου.

# Asp (D), Glu (E)



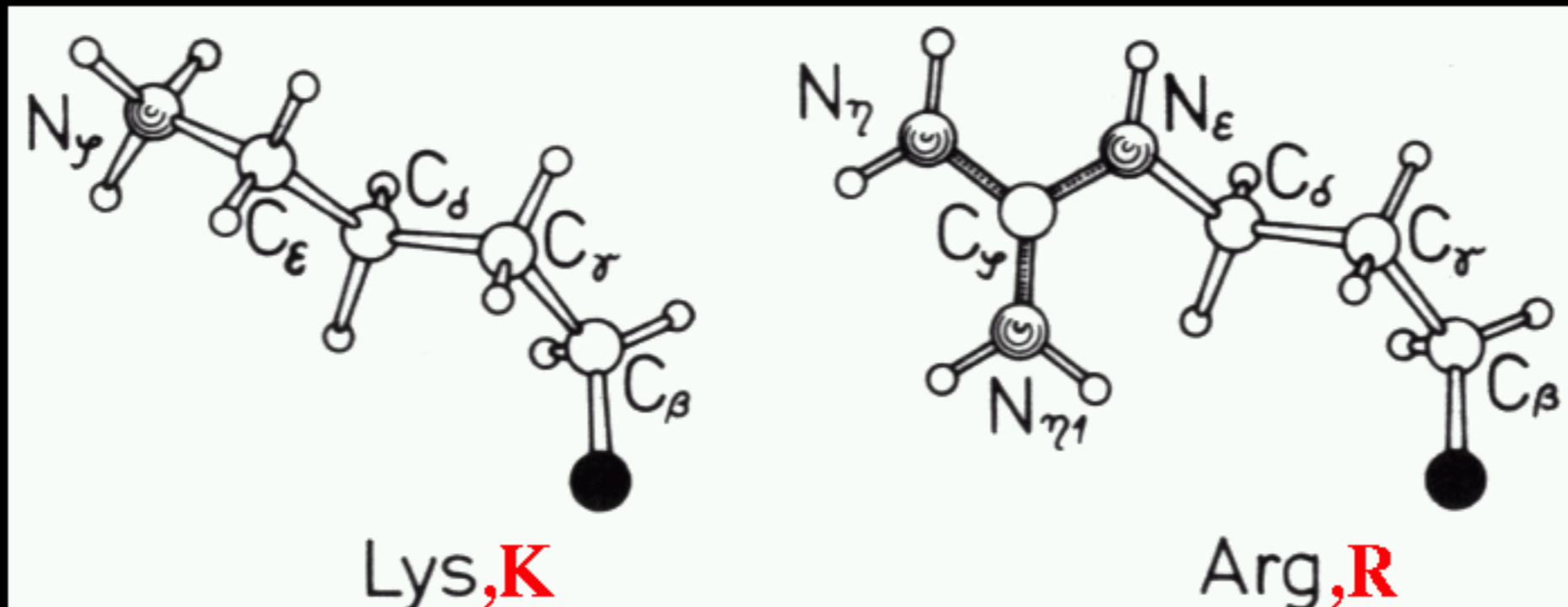
Το ασπαρτικό οξύ και το γλουταμικό οξύ είναι αρνητικά φορτισμένα αμινοξέα (σε φυσιολογικά pH : τα pK είναι 3.9 και 4.3 αντίστοιχα). Η σχετικά μικρή πλευρική ομάδα του ασπαρτικού μειώνει το εύρος των πιθανών διαμορφώσεων του και το κάνει πιο συνηθισμένο σε ενεργά κέντρα ενζύμων.

# Lys (K), Arg (R)



Η λυσίνη και η αργινίνη είναι θετικά φορτισμένα αμινοξέα (σε φυσιολογικά pH : τα pK είναι 10.5 και 12.5 αντίστοιχα). Είναι από τις πλέον ευκίνητες επιφανειακές πλευρικές ομάδες των πρωτεϊνών, σε τέτοιο βαθμό ώστε συνηθέστατα να μην υπάρχει πληροφορία για τη διαμόρφωση τους (τουλάχιστον από κρυσταλλογραφικά πειράματα).

# Lys (K), Arg (R)



Παρ'όλα αυτά, μερικές φορές μπορούν να λάβουν μέρος σε γέφυρες άλατος (στο εσωτερικό των πρωτεϊνών) ή να είναι τμήματα του καταλυτικού κέντρου ενζύμων (για παράδειγμα, στις οψίνες το χρωμοφόρο συνδέεται ομοιοπολικά με μια λυσίνη).

# Βιβλίο

---

"Βιοπληροφορική"

Editors : Baxevanis, A.D. & Ouellette, B.F.F.

Επιμέλεια Ελληνικής Έκδοσης :

Μουνδριανάκης, Ε.Ν. &  
Χαμόδρακας, Σ.Ι.

# Σύνοψη διαλέξεων

- Εφαρμογές υπολογιστών στη βιολογία, ορισμοί.
- Η Βιοπληροφορική ως εργαλείο και ερευνητικό πεδίο. Αλγόριθμοι, προγράμματα, η σημασία του διαδικτύου (client-server model).
  - Βάσεις δεδομένων : Δομή και αναζήτηση πληροφοριών, οι πλέον γνωστές βάσεις δεδομένων.
  - Στοίχιση δύο αλληλουχιών : Σχολαστικοί αλγόριθμοι : Needleman & Wunsch, Smith & Waterman, Πίνακες βαθμολόγησης (PAM, BLOSUM).
  - Στοίχιση δύο αλληλουχιών : Ευρεστικοί αλγόριθμοι : οι αλγόριθμοι που χρησιμοποιούνται στα προγράμματα BLAST και FASTA.



# Σύνοψη διαλέξεων

- Στοίχιση πολλών αλληλουχιών : Προβλήματα, αλγόριθμοι και ευρέως χρησιμοποιούμενα προγράμματα
  - Φυλογενετικά δένδρα. Ορισμοί, μορφές δένδρων. Αλγόριθμοι για την δημιουργία δένδρων μέσω στοιχίσεων αλληλουχιών.
  - Πρωτεϊνικά μοτίβα : προσδιορισμός, αναζήτηση, βάσεις δεδομένων και εργαλεία αναζήτησης.
  - Πρόβλεψη ανοικτών πλαισίων ανάγνωσης.
  - Πρόβλεψη μεταγραφικών ρυθμιστικών στοιχείων.
  - Expressed Sequence Tags : ιδιαιτερότητες και προβλήματα.
  - Λειτουργική γενομική και γονιδιακή έκφραση : microarrays. Προβλήματα, αλγόριθμοι, προγράμματα.

# Σύνοψη διαλέξεων

---

- Εφαρμογές στη Δομική Βιολογία : πρόβλεψη δευτεροταγούς δομής πρωτεϊνών και RNA, πρόβλεψη διαμεμβρανικών τμημάτων, homology modeling, αναγνώριση μοτίβου αναδίπλωσης, ab initio πρόβλεψη πρωτεϊνικής δομής.

# Σύνοψη εργαστηρίων

---

"Έκανα click στο link που είπατε αλλά έχει περάσει μισή ώρα και δεν έχει συμβεί τίποτα. Μήπως χάλασε ο υπολογιστής ;"

Προβλήματα :

- Απουσία δικτύου.
- Απουσία ικανού αριθμού υπολογιστών.
- Απουσία βοηθητικού προσωπικού.
- Απουσία χρόνου.

Λύση : "distance learning"

# Αίθουσα υπολογιστών

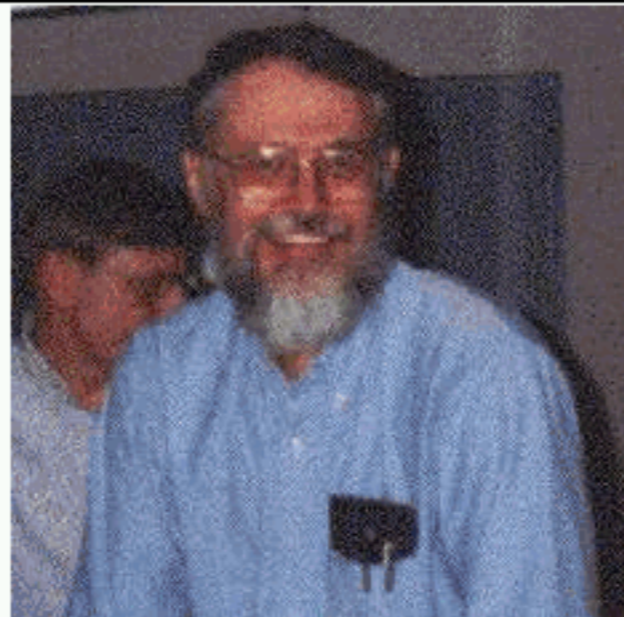
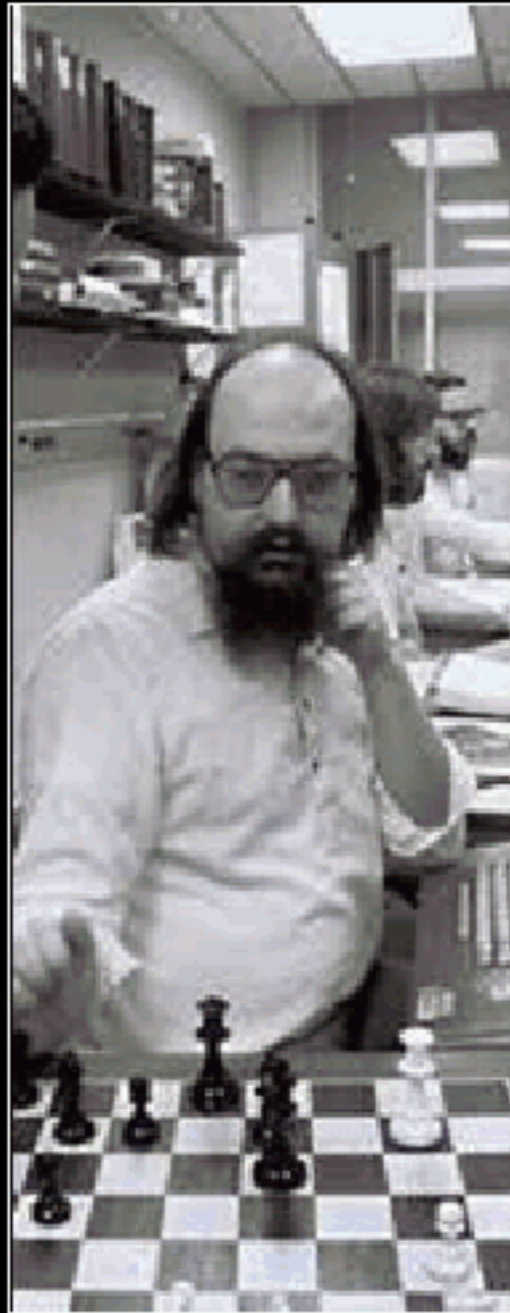
---

Είναι ανοικτή για ελεύθερη χρήση από τους φοιτητές του τμήματος, αλλά υπό όρους :

- Κανονισμός εργαστηρίου.
- Διαθεσιμότητα χειριστών.
- Διαθεσιμότητα γραμματείας.

Επιπλέον, θα πρέπει να είστε έτοιμοι/μες να χρησιμοποιήσετε ένα POSIX λειτουργικό σύστημα (linux).

# Unix vs. Windows ?



# Βιοπληροφορική: ορισμοί

---

Υπολογιστικές μέθοδοι και τεχνικές οι οποίες είναι δημοσιεύσιμες σε διεθνή επιστημονικά περιοδικά και τα αποτελέσματα από την εφαρμογή των οποίων εμπίπτουν στο γνωστικό αντικείμενο της βιολογίας.

Ο ορισμός αυτός δεν κάνει διάκριση ανάμεσα στη βιοπληροφορική και την υπολογιστική βιολογία, και δίνει βαρύτητα στο καθαρά ερευνητικό κομμάτι της βιοπληροφορικής.

# Βιοπληροφορική: ορισμοί (NIH)

Η έρευνα, ανάπτυξη ή εφαρμογή υπολογιστικών εργαλείων και προσεγγίσεων για την επέκταση της χρήσεως των βιολογικών δεδομένων καθώς και για τη συλλογή, αποθήκευση, οργάνωση, αρχειοθέτηση, ανάλυση και οπτικοποίηση αυτών των δεδομένων.

Υπολογιστική βιολογία : η ανάπτυξη και εφαρμογή αναλυτικών και θεωρητικών μεθόδων, μαθηματικών μοντέλων και τεχνικών υπολογιστικών προσομοιώσεων για τη μελέτη βιολογικών συστημάτων.

# Θέση στη μοριακή βιολογία

## 1. Προσδιορισμός αλληλουχίας DNA

- Χαρτογράφηση
- Fragment assembly

## 2. Ανάλυση αλληλουχίας DNA

- Εύρεση γονιδίων, κωδικοποιούσας περιοχής
- Πολλαπλή στοίχιση αλληλουχιών
- Φυλογενετικά δένδρα

## 3. Ανάλυση πρωτεϊνικής αλληλουχίας

- Στοιχίσεις
- Εξελικτικές σχέσεις
- Πρόβλεψη δευτεροταγούς δομής
- Πρόβλεψη τριτοταγούς δομής ?



# Θέση στη μοριακή βιολογία

---

4. Ανάλυση αλληλεπιδράσεων

- Μεταξύ πρωτεϊνών
- Μεταξύ πρωτεϊνών και ligands.

5. Ανάλυση έκφρασης πρωτεϊνών και γονιδίων

- Microarrays

6. Εύρεση και ανάλυση πιθανών φαρμάκων ?

# ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ ή ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ ;

- Για τη μεγάλη πλειοψηφία των βιολόγων, η βιοπληροφορική είναι εργαλείο δουλειάς και αυτό που προέχει είναι η ικανότητα τους να κρίνουν τη βιολογική σημασία των αποτελεσμάτων που προκύπτουν από την εφαρμογή των υπολογιστικών εργαλείων.
- Για τη μεγάλη πλειοψηφία των εμπλεκόμενων μη βιολόγων (μαθηματικοί, επιστήμης υπολογιστών, ...), η βιοπληροφορική είναι το καθ'αυτό ερευνητικό αντικείμενο (και όχι μέσο προς σκοπό).

# ΒΙΟπληροφορική λοιπόν, αλλά ...

---

Η κρίση των αποτελεσμάτων από την εφαρμογή υπολογιστικών εργαλείων, απουσία γνώσης των μεθόδων (αλγορίθμων) που κρύβονται πίσω από αυτά τα εργαλεία, είναι δύσκολη.

Η αλληλεπίδραση με ένα υπολογιστικό εργαλείο (π.χ. εισαγωγή ή αλλαγή παραμέτρων), απουσία γνώσης του πως δουλεύει αυτό το εργαλείο, είναι αδύνατη.

# ΒΙΟπληροφορική λοιπόν, αλλά ...

Επιπλέον, η προσήλωση στην "πρακτική" εφαρμογή, δημιουργεί την ψευδαίσθηση της ουσιαστικής γνώσης, όταν στην πραγματικότητα εκπαιδεύει μόνο την τεχνική δεξιότητα (μετρούμενη σε CPR, "clicks per minute").

Άρα, το ζητούμενο του μαθήματος ΔΕΝ είναι να εκπαιδευτείτε ως χρήστες όσων το δυνατόν περισσότερων προγραμμάτων και εργαλείων βιοπληροφορικής υπάρχουν διαθέσιμα, αλλά η σε βάθος περιγραφή μερικών από τους 'εκ των ουκ άνευ' αλγόριθμους και των εφαρμογών τους στη λύση βιολογικών προβλημάτων.

# Αλγόριθμοι

---

Ορισμός : Αλγόριθμος ενός προβλήματος είναι η αναλυτική περιγραφή της ακολουθίας των βημάτων που απαιτούνται για την επίλυση του προβλήματος.

Πιο απλά είναι η αναλυτική περιγραφή της μεθόδου επίλυσης ενός προβλήματος.

Ειδικά για τον προγραμματισμό υπολογιστών, ο αλγόριθμος είναι μια περιγραφή της λογικής του προγράμματος.

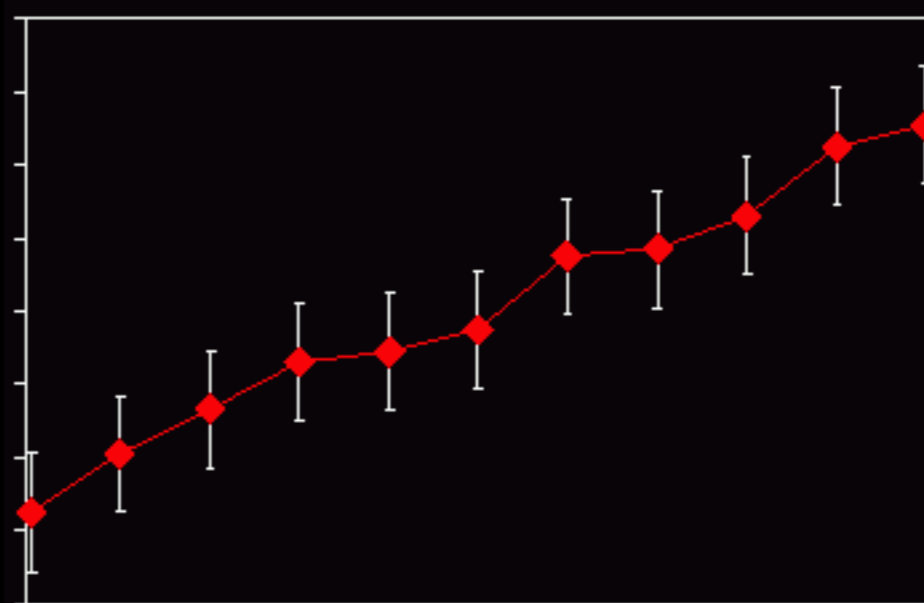
# Αλγόριθμοι : παραδείγματα

## Η μέθοδος Bradford :

- \* Είναι μια φασματοσκοπική μέθοδος προσδιορισμού πρωτεϊνικής συγκέντρωσης.
- \* Στηρίζεται στην αλλαγή του χρώματος μιας χρωστικής (Coomassie Blue G) όταν υπάρχουν μόρια πρωτεΐνης στο διάλυμα.
- \* Για τη χρήση της μεθόδου κατασκευάζεται πρότυπη καμπύλη με τη βοήθεια πρωτεϊνικών διαλυμάτων γνωστής συγκέντρωσης : για το κάθε ένα από αυτά μετράμε την οπτική πυκνότητα (στα 595nm).

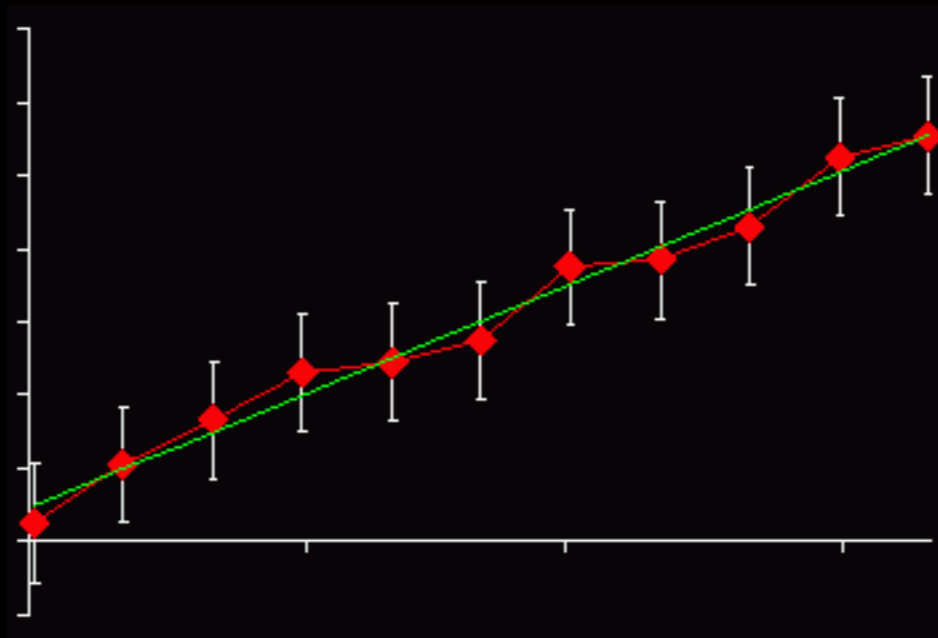
# Αλγόριθμοι : παραδείγματα

- \* Για χαμηλές πρωτεϊνικές συγκεντρώσεις τα σημεία της πρότυπης καμπύλης αναμένεται να ανήκουν σε μια ευθεία γραμμή.
- \* Αλλά εάν στα αλήθεια κάνετε το πείραμα, είναι σχεδόν βέβαιο ότι τα σημεία δεν θα ανήκουν σε μια ευθεία.



# Αλγόριθμοι : παραδείγματα

- \* Για χαμηλές πρωτεϊνικές συγκεντρώσεις τα σημεία της πρότυπης καμπύλης αναμένεται να ανήκουν σε μια ευθεία γραμμή.
- \* Αλλά εάν στα αλήθεια κάνετε το πείραμα, είναι σχεδόν βέβαιο ότι τα σημεία δεν θα ανήκουν σε μια ευθεία.





# Αλγόριθμοι : παραδείγματα

---

## Λύση :

Η μέθοδος των ελαχίστων τετραγώνων (least squares method) :

- Υπέθεσε ότι τα σφάλματα στις τιμές της πρωτεϊνικής συγκέντρωσης στα πρότυπα διαλύματα είναι τόσο μικρές ώστε να μπορούν να θεωρηθούν αμελητέες.

- Υπέθεσε ότι όλες οι μετρούμενες τιμές οπτικής πυκνότητας έχουν παρόμοιες τιμές σφαλμάτων.

# Αλγόριθμοι : παραδείγματα

---

Τότε :

Η ευθεία που θέλουμε είναι αυτή για την οποία το άθροισμα των τετραγώνων των διαφορών ανάμεσα στις παρατηρούμενες και υπολογιζόμενες τιμές της οπτικής πυκνότητας των διαλυμάτων ελαχιστοποιείται.

# Αλγόριθμοι : παραδείγματα

Δηλαδή : Έστω μια ευθεία με εξίσωση  $y=ax+b$  όπου το  $x$  είναι η πρωτεύουσα συγκέντρωση των πρότυπων διαλυμάτων και  $y$  είναι η μετρούμενη οπτική πυκνότητα. Για κάθε τιμή του  $x$  μπορούμε να βρούμε την τιμή του  $y$  που θα περιμέναμε εάν όντως αυτή η εξίσωση ίσχυε. Συνεπώς για κάθε τέτοια ευθεία, μπορούμε να υπολογίσουμε τις διαφορές ανάμεσα στις παρατηρούμενες και τις υπολογιζόμενες τιμές των  $y$  για όλα τα  $x$ .

# Αλγόριθμοι : παραδείγματα

Άρα, αυτό που θέλουμε είναι να ελαχιστοποιήσουμε  
το :

$$\Delta = \sum_i [y_i - y_{calc}]^2$$

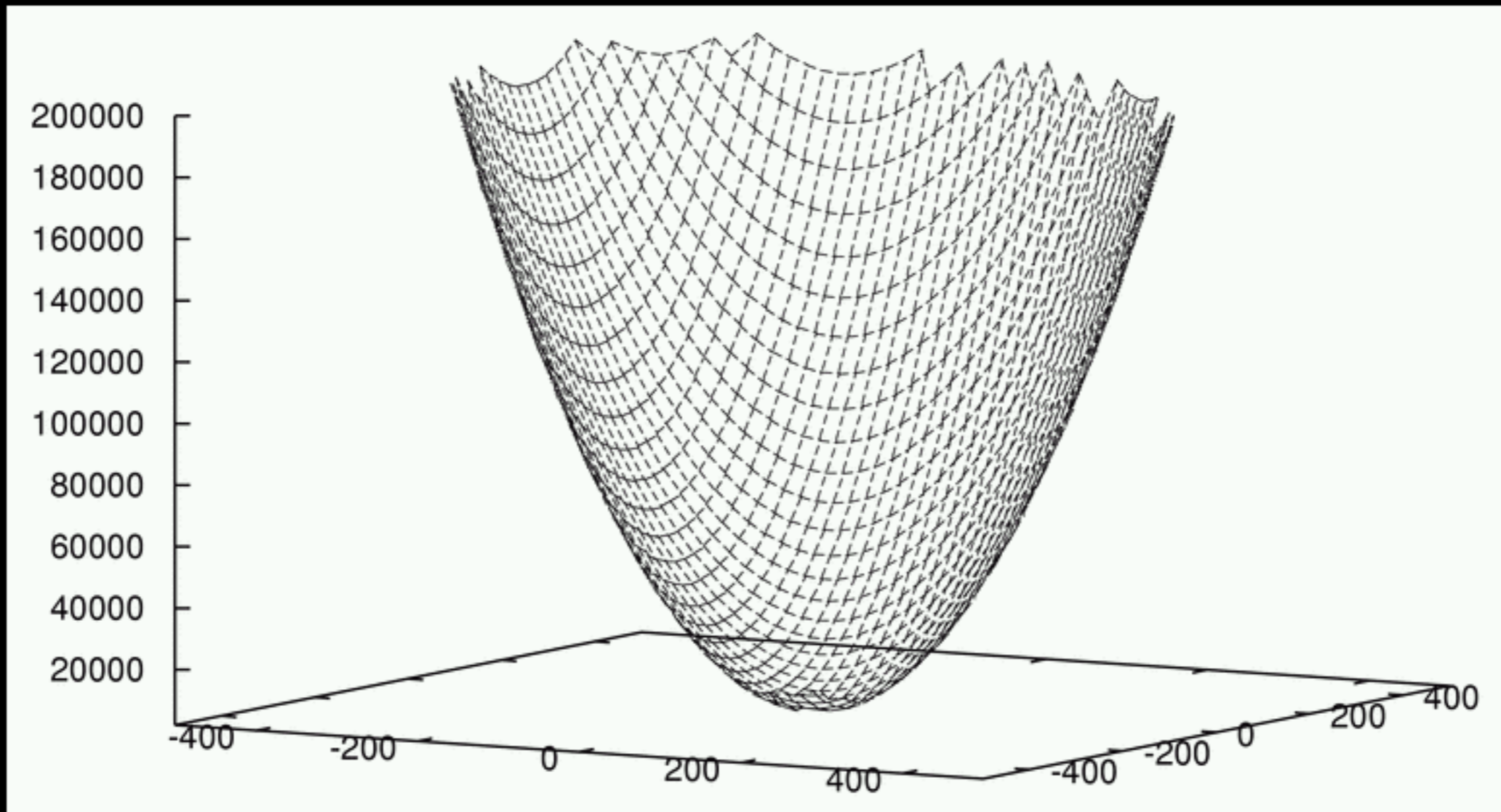
$$\Delta = \sum_i [y_i - (ax_i + b)]^2$$

...

ως συνάρτηση των συντελεστών  $a$  και  $b$  της ευθείας  
 $y=ax+b$ .

# Αλγόριθμοι : παραδείγματα

Αυτή είναι μια δευτεροβάθμια και στις δύο παραμέτρους



# Αλγόριθμοι : παραδείγματα

Η λύση της οποίας είναι

$$a = \frac{N \sum yx - \sum x \sum y}{N \sum x^2 - (\sum x)^2}$$

$$b = \frac{\sum y \sum x^2 - \sum xy \sum x}{N \sum x^2 - (\sum x)^2}$$

αλλά αυτό ΔΕΝ είναι αλγόριθμος.

# Αλγόριθμοι : παραδείγματα

Πλησιέστερα στο τι εννοούμε "αλγόριθμο" είναι το :

- \* Διάβασε τα δεδομένα  $x[i]$ ,  $y[i]$  ( $N$  ζεύγη τιμών).
- \* Υπολόγισε τις τιμές των αθροισμάτων  $\Sigma x$ ,  $\Sigma y$ ,  $\Sigma xy$ , και  $\Sigma x^2$ .
- \* Υπολόγισε τα  $a$  και  $b$ .

Αλλά και αυτό απέχει από το να είναι αρκετά αναλυτικό (από υπολογιστική άποψη). Καλύτερο είναι το εξής ...

# Αλγόριθμοι : παραδείγματα

1. Αρχικοποίησε τις τιμές των αθροισμάτων  $\Sigma x$ ,  $\Sigma y$ ,  $\Sigma xy$ , και  $\Sigma x^2$  στο μηδέν.
2. Διάβασε την τιμή του  $N$  (αριθμός σημείων)
3. Εάν  $N > 1$  συνέχισε, αλλιώς γκρίνιαξε.
4. Για κάθε ζεύγος τιμών, από το πρώτο μέχρι το  $N$ -οστό, κάνε τα παρακάτω :
  - Διάβασε τα  $x$ ,  $y$
  - Πρόσθεσε στην τιμή του  $\Sigma x$  την τιμή του  $x$
  - Πρόσθεσε στην τιμή του  $\Sigma x^2$  την τιμή του  $x^2$
  - Πρόσθεσε στην τιμή του  $\Sigma y$  την τιμή του  $y$
  - Πρόσθεσε στην τιμή του  $\Sigma xy$  την τιμή του γινομένου  $x \cdot y$ .
  - Επανάλαβε για το επόμενο ζεύγος τιμών
5. Υπολόγισε τα  $a$  και  $b$
6. Τύπωσε τις τιμές των  $a$  και  $b$

Κωδικοποιείται ως πρόγραμμα υπολογιστών



# Αλγόριθμοι : παραδείγματα

Εύρεση της θέσης ενός μοτίβου σε μια αλληλουχία :

Έστω μια αλληλουχία χαρακτήρων  $D[n]$  (π.χ. μια αλληλουχία DNA ή πρωτεΐνης) και μια μικρότερη σε μήκος υπακολουθία χαρακτήρων  $S[m]$  (με  $m \leq n$ ).

Το πρόβλημα είναι να προσδιορίσουμε (εάν υπάρχει) τη θέση (ή τις θέσεις) της αλληλουχίας  $D$  στην οποία υπάρχει η υπακολουθία  $S$ .

# Αλγόριθμοι : παραδείγματα

Υπάρχει, και εάν ναι, που, η υπακολουθία "FLARE"  
στην κάτωθι αλληλουχία :

```
MGLPPLLSLP SNSAPRSLGR VETPPEVVDF MVSLAEAPRG GRVLEPACAH  
GPFLRAFREA HGTGYRFVGV EIDPKALDLP PWAEGILADF LLWEPGEAFD  
LILGNPPYGI VGEASKYPIH VFKAVKDLYK KAFSTWKGY NLYGAFLEKA  
VRLKPGGVL VVVVPATWLV LEDFALLREF LAREGKTSVY YLGEVFPQKK  
VSAVVIRFQK SGKGLSLWDT QESESGFTPI LWAEYPHWEG EIIRFETEET  
RKLEISGMPL GDLFHIRFAA RSPEFKKHPA VRKEPGPGLV PVLTRNLKP  
GWVDYEKNHS GLWMPKERAK ELRDFYATPH LVVAHTKGTR VVAAWDERAY  
PWREEFHLLP KEGVRLDPSS LVQWLNSEAM QKHVRTLYRD FVPHLTRML  
ERLPVRREYG FHTSPESARN F
```

# Αλγόριθμοι : παραδείγματα

- Για κάθε θέση  $(i)$  της αλληλουχίας  $D$  :
  - Για κάθε θέση  $(k)$  της αλληλουχίας  $S$  :
    - Ισχύει  $D[i+k] == S(k)$  για κάθε  $(k)$  ?
      - \* Εάν ναι : η θέση  $(i)$  είναι λύση
      - \* Εάν όχι : αύξησε το  $(i)$  κατά 1 και επανέλαβε (μέχρι το τέλος της αλληλουχίας  $D$ ).

# Αλγόριθμοι : παραδείγματα

---

```
for ( i=0 ; i < μήκος(D) ; i++ )
{
    for ( k=0 ; k < μήκος(S) ; k++ )
        if ( D[i+k] != S[k] )
            break;

    if ( k == μήκος(S) )
        τύπωσε "Λύση για τη θέση (i)"
}
```

# Προγράμματα υπολογιστών

---

Ορισμός : Ένα πρόγραμμα είναι ένας αλγόριθμος κωδικοποιημένος σε μια γλώσσα προγραμματισμού.

Γλώσσα προγραμματισμού είναι ένα σύνολο εντολών κατανοητών από ηλεκτρονικό υπολογιστή. Υπάρχουν πολλές τέτοιες γλώσσες (π.χ. assembly, C, C++, FORTRAN, Perl, ...), κάθε μια με τα δικά της ιδιότυπα χαρακτηριστικά.

Διάκριση ανάμεσα σε κώδικα μηχανής (εκτελέσιμο πρόγραμμα) και το ίδιο πρόγραμμα γραμμένο σε κάποια υψηλού επιπέδου γλώσσα προγραμματισμού.

# Βιοπληροφορική & προγράμματα

---

Για πολλά χρόνια το κυρίαρχο μοντέλο ήταν το "get the program & run", δηλαδή, ο κάθε χρήστης ήταν υποχρεωμένος να βρίσκει και να εγκαθιστά τα διάφορα προγράμματα στον υπολογιστή του.

Η έλευση του δικτύου άλλαξε άρδην το κυρίαρχο μοντέλο : the client-server model.

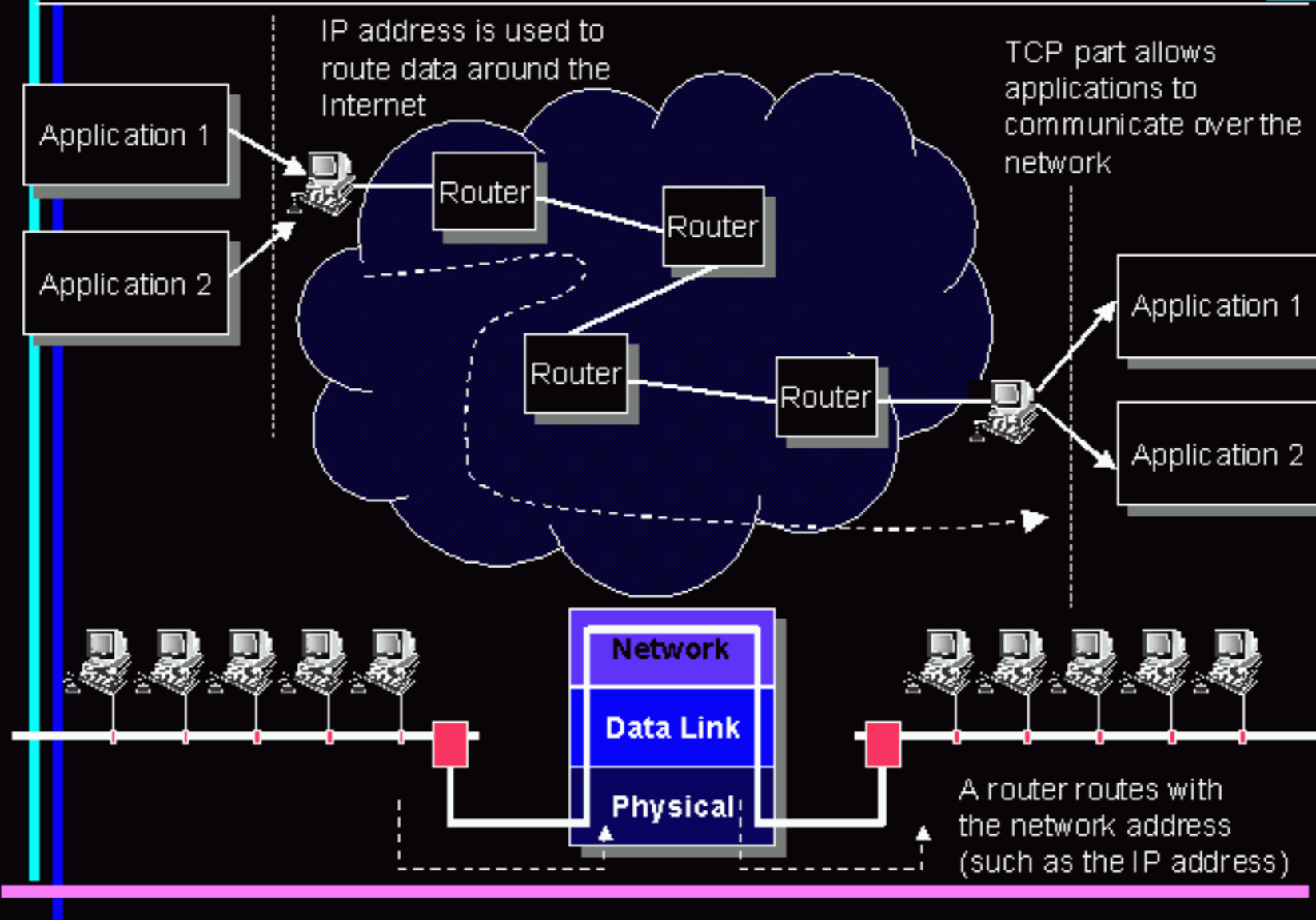
# The client-server model

---

Το πρόγραμμα αυτό καθ'αυτό είναι απροσπέλαστο και εκτελείται (αδιαφανώς για τον χρήστη) στο υπολογιστικό σύστημα αυτού που παρέχει την υπηρεσία. Ο χρήστης βλέπει μόνο ένα user interface (π.χ. μια φόρμα σε κάποια ιστοσελίδα), και τα αποτελέσματα από την εφαρμογή του προγράμματος. Ο φορέας για τη ροή πληροφορίας ανάμεσα στο χρήστη και το πρόγραμμα είναι το δίκτυο.

# Παρένθεση : TCP/IP

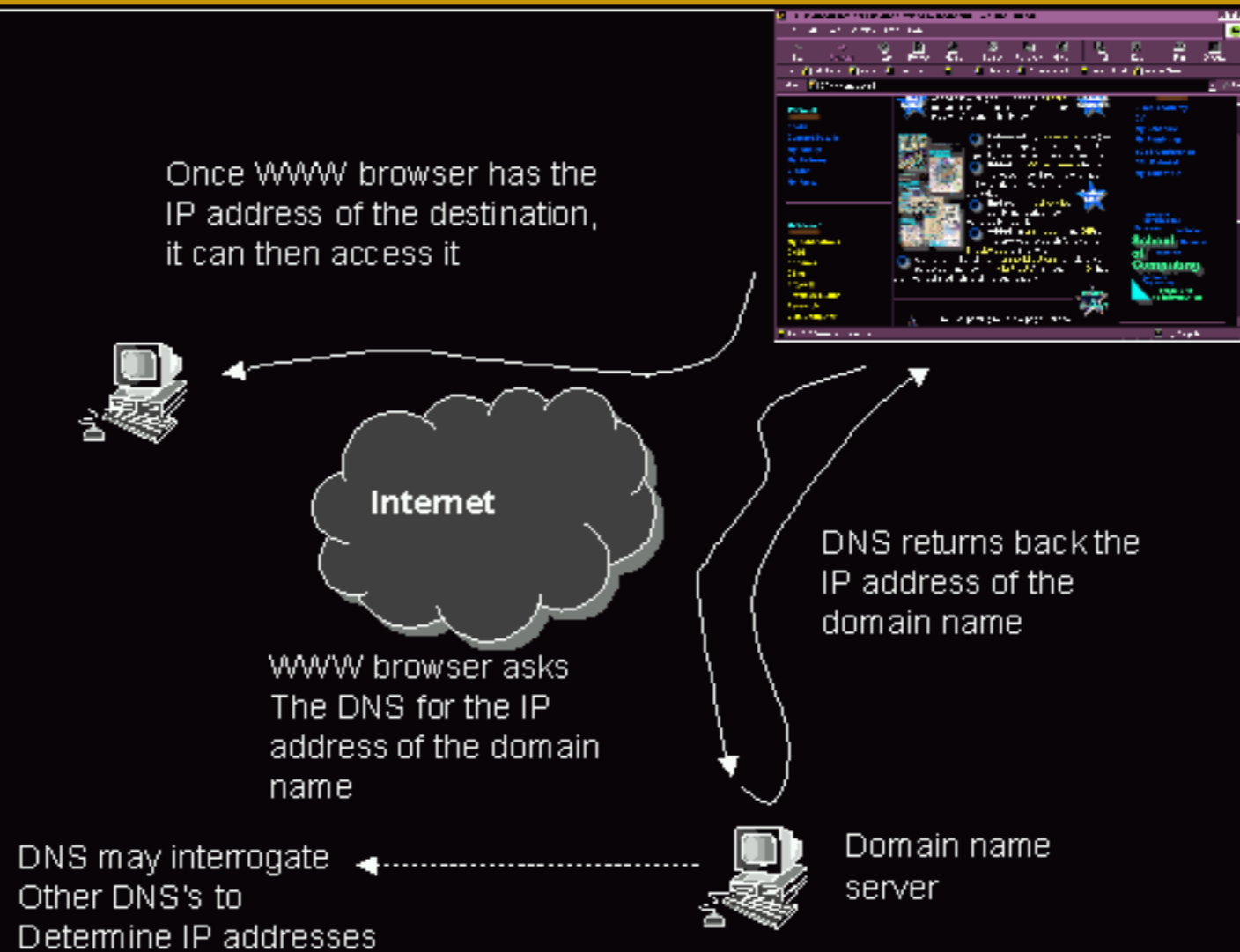
## 5.2 IP and TCP





# Παρένθεση : TCP/IP

## 5.16 Domain name server



# The client-server model

---

## Πλεονεκτήματα :

- Χρονοβόροι υπολογισμοί γίνονται εφικτοί για όλους μέσω της δημιουργίας λίγων αλλά ισχυρών εξυπηρετητών (servers).
- Υπολογισμοί που χρησιμοποιούν μεγάλες βάσεις δεδομένων γίνονται εφικτοί για όλους χωρίς να απαιτείται η (περιττή) δημιουργία πολλαπλών αντιγράφων των ίδιων δεδομένων.
- Οι τελικοί χρήστες αποφεύγουν όλες τις τεχνικές δυσκολίες εγκατάστασης και συντήρησης των προγραμμάτων, καθώς και τα έξοδα υλικού (hardware) που πιθανό να απαιτούνταν.

# The client-server model

---

## Πλεονεκτήματα :

- Οι τελικοί χρήστες αποφεύγουν όλες τις τεχνικές δυσκολίες εγκατάστασης και συντήρησης των βάσεων δεδομένων, καθώς και τα έξοδα υλικού (hardware) που πιθανό να απαιτούνταν.
- "Ασφαλέστερο" περιβάλλον εργασίας για τους νέους χρήστες μέσω καλά σχεδιασμένων ιστοσελίδων (online help, forms with default values, ...).

# The client-server model

---

## Μειονεκτήματα :

- No internet ? No party.
- Αδυναμία των έμπειρων χρηστών να ελέγξουν ή και να τροποποιήσουν τα προγράμματα (μη ελεύθερο λογισμικό).
- Εξάρτηση από τους παρέχοντες την υπηρεσία.
- Πιθανότητα έξωθεν ελέγχου και παρακολούθησης των πεπραγμένων ενός χρήστη.

# The client-server model

---

## Μειονεκτήματα :

- No internet ? No party.
- Αδυναμία των έμπειρων χρηστών να ελέγξουν ή και να τροποποιήσουν τα προγράμματα (μη ελεύθερο λογισμικό).
- Εξάρτηση από τους παρέχοντες την υπηρεσία.
- Πιθανότητα έξωθεν ελέγχου και παρακολούθησης των πεπραγμένων ενός χρήστη.