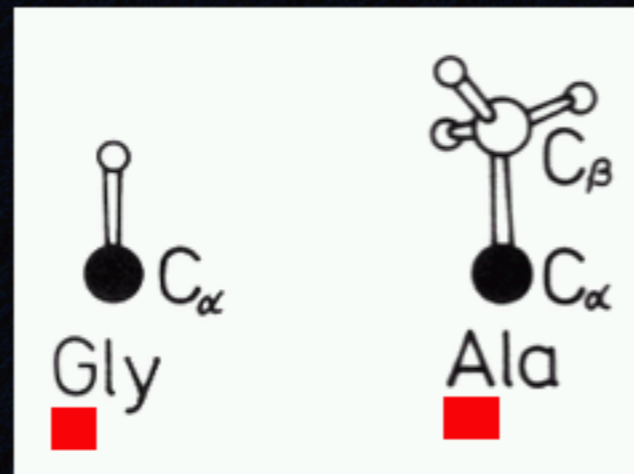


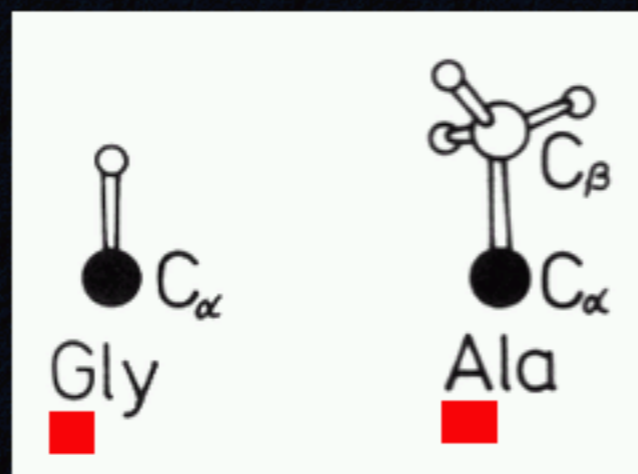
Βιοπληροφορική

Gly (G), Ala (A)



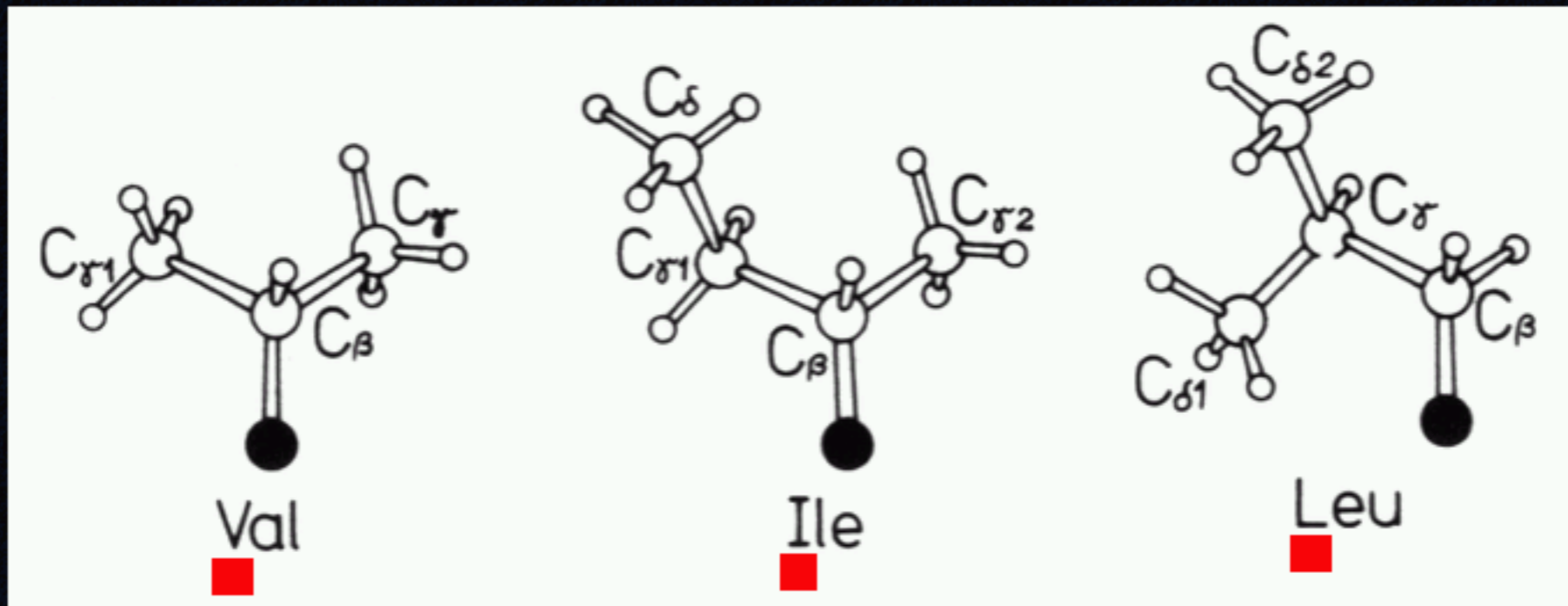
Η γλυκίνη είναι το μοναδικό αμινοξύ χωρίς πλευρική ομάδα (μόνο ένα άτομο υδρογόνου). Η απουσία πλευρικής ομάδας (και συνεπώς η απουσία στεreoχημικών παρεμποδίσεων με την κυρίως αλυσίδα) της επιτρέπουν να υιοθετεί διαμορφώσεις οι οποίες είναι ενεργειακά ανέφικτες για τα υπόλοιπα αμινοξέα (π.χ. απότομες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

Gly (G), Ala (A)



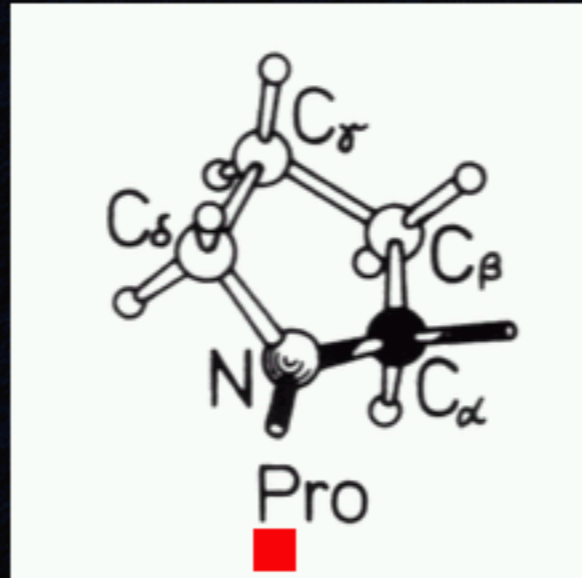
Η αλανίνη έχει τη δεύτερη μικρότερη πλευρική ομάδα, μια μεθυλομάδα. Είναι ένα μη πολικό κατάλοιπο το οποίο παρ'όλα αυτά απαντάται αρκετά συχνά και στην επιφάνεια των πρωτεϊνών. Η συχνότητα της Ala στις πρωτεϊνικές αλληλουχίες είναι αρκετά υψηλή, ίσως λόγω του μικρού ενεργειακού κόστους σύνθεσης της.

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



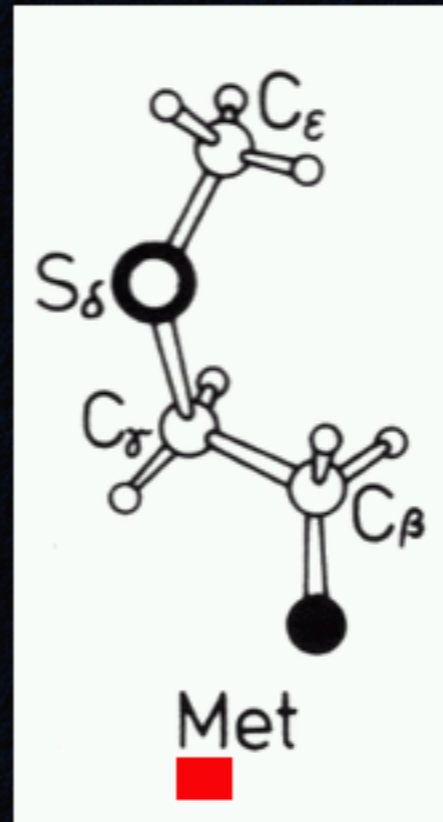
Αυτά είναι και τα τρία υδρόφοβα αμινοξέα τα οποία χαρακτηρίζονται από την διακλαδιζόμενη πλευρική τους αλυσίδα. Η ύπαρξη διακλάδωσης (ιδιαίτερα για τις Val και Ile που διακλαδίζονται στο C β) περιορίζει αισθητά τις προτιμώμενες διαμορφώσεις τους (γνωστές ως rotamers).

Pro (P)



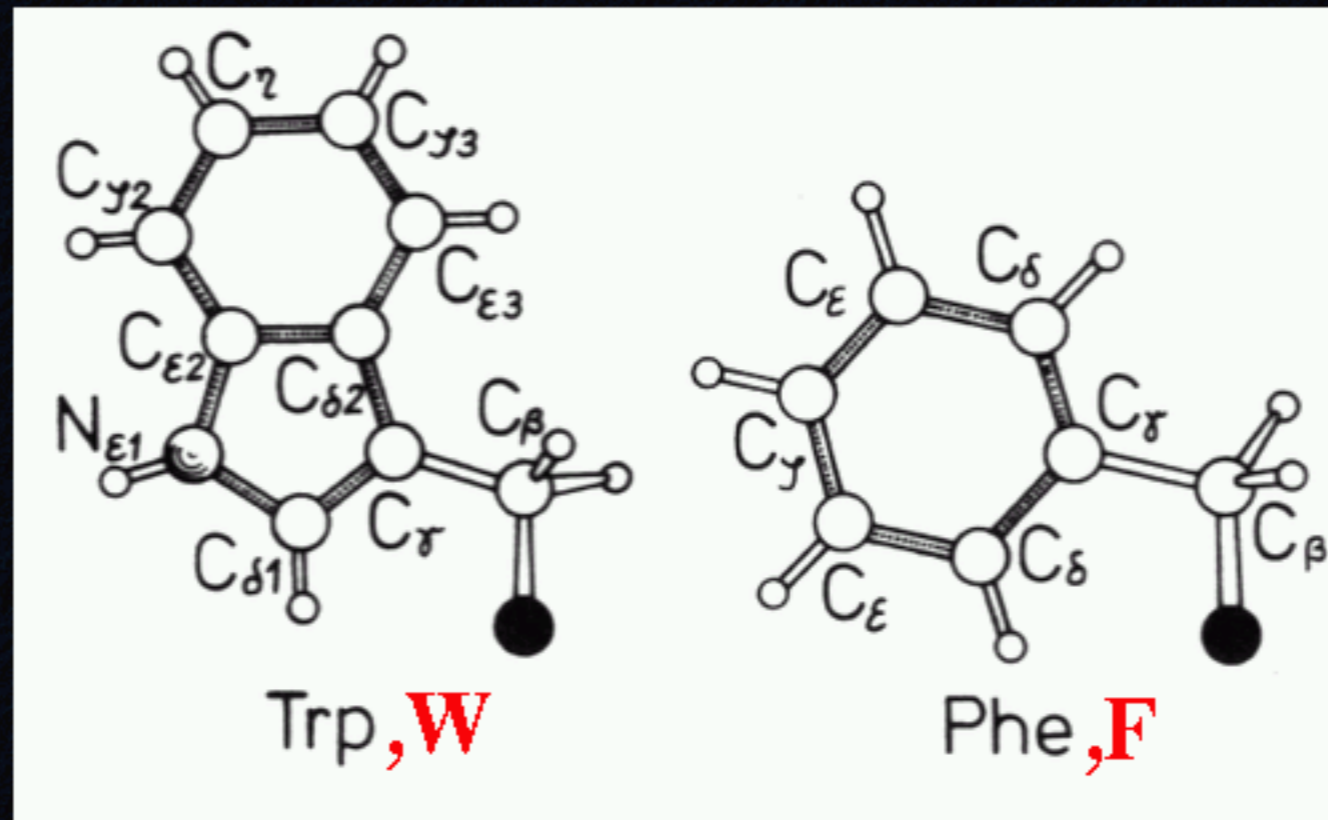
Η προλίνη είναι ένα ιδιαίτερο αμινοξύ : Η πλευρική ομάδα επιστρέφει και συνδέεται ομοιοπολικά με το σκελετό της πρωτεΐνης. Η πιο άμεση συνέπεια είναι ότι αυτή η πλευρική ομάδα είναι πρακτικά ακίνητη (ο μοναδικός βαθμός ελευθερίας βρίσκεται στη διαμόρφωση του δακτυλίου). Η δεύτερη συνέπεια είναι ότι η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

Met (M)



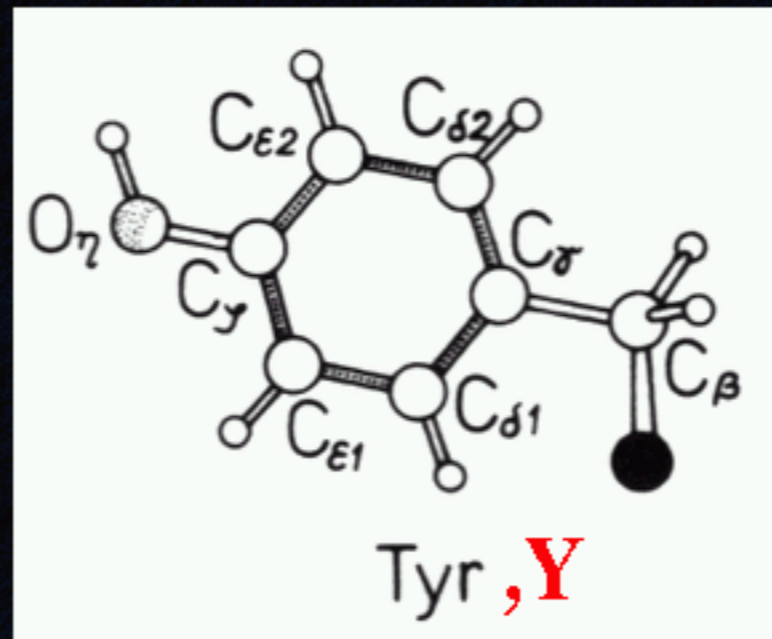
Η μεθειονίνη έχει μια μάλλον ευκίνητη πλευρική με ένα άτομο θείου σε θειο-αιθερικό δεσμό. Η διαφορά ηλεκτραρνητικότητας ανάμεσα στο θείο και τους άνθρακες οδηγεί στην δημιουργία διπολικής ροπής. Ως επί το πλείστον απαντάται στο εσωτερικό των πρωτεϊνικών δομών.

Trp (W), Phe (F)



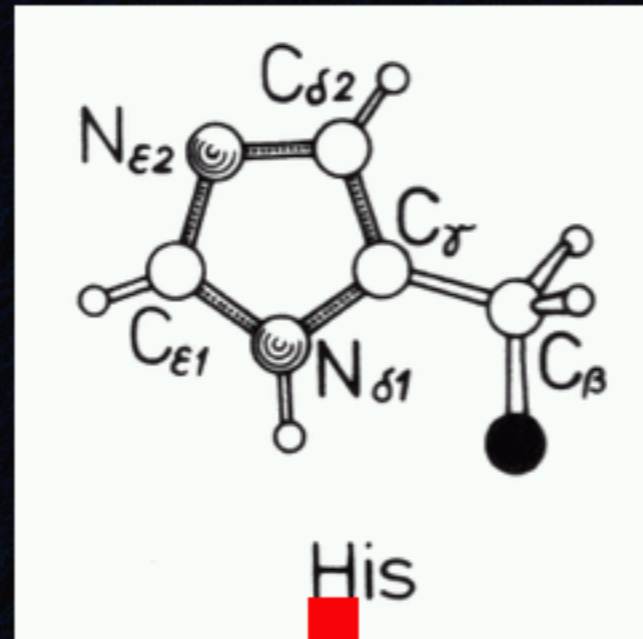
Η τρυπτοφάνη και η φαινυλαλανίνη είναι από τα πλέον υδρόφοβα αμινοξέα και ως εκ τούτου απαντώνται κατά κύριο λόγο στο εσωτερικό των πρωτεϊνών, αν και η πλευρική ομάδα της τρυπτοφάνης είναι ελαφρά πολική λόγω του ετεροκυκλικού ινδολικού δακτυλίου.

Tyr (Y)



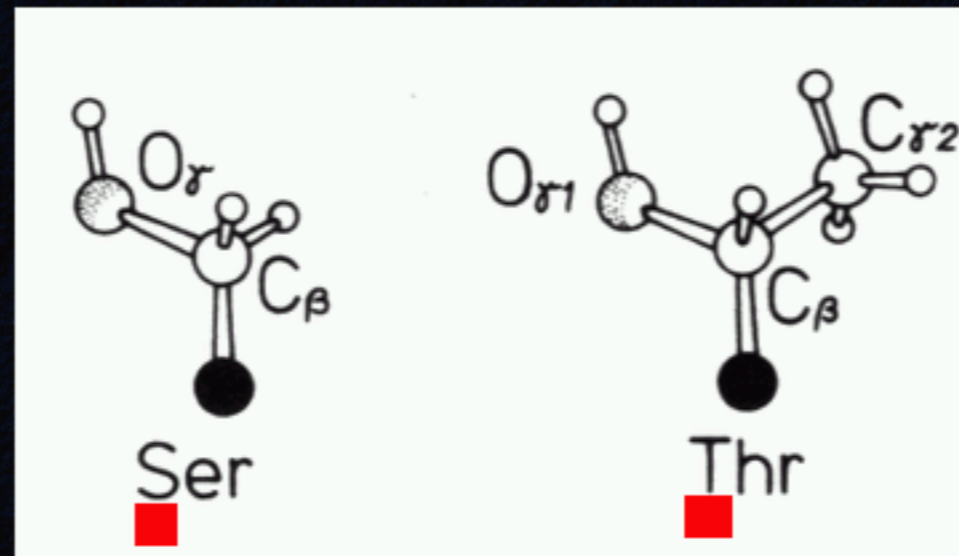
Η τυροσίνη έχει την ιδιαιτερότητα ότι αν και έχει αρωματικό δακτύλιο, έχει επίσης μια υδροξυλομάδα (με pK 10.1) η οποία την κατατάσσει στα πολικά αμινοξέα (και μάλιστα με δυνητικά ισχυρούς υδρογονικούς δεσμούς). Όπως τα περισσότερα αφόρτιστα αλλά πολικά αμινοξέα, η τυροσίνη μπορεί να βρεθεί τόσο στο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

His (H)



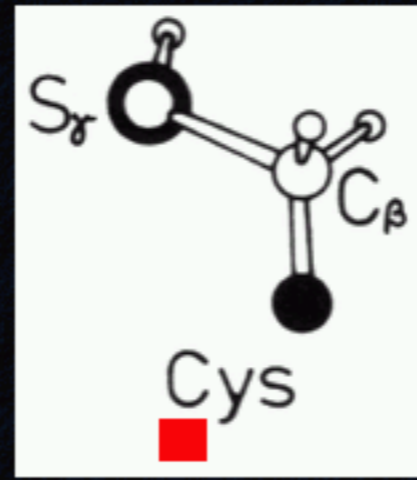
Η ιστοιδίνη έχει ένα ετεροκυκλικό αρωματικό δακτύλιο με μία χαρακτηριστική ιδιαιτερότητα : το pK της είναι 6, το οποίο σημαίνει ότι στα φυσιολογικά (ενδοκυτταρικά) pH , οι ιστοιδίνες μπορούν να είναι είτε αφόρτιστες ($pH > 6$) ή φορτισμένες ($pH < 6$). Επειδή, δε, η ακριβής τιμή του pK εξαρτάται (και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

Ser (S), Thr (T)



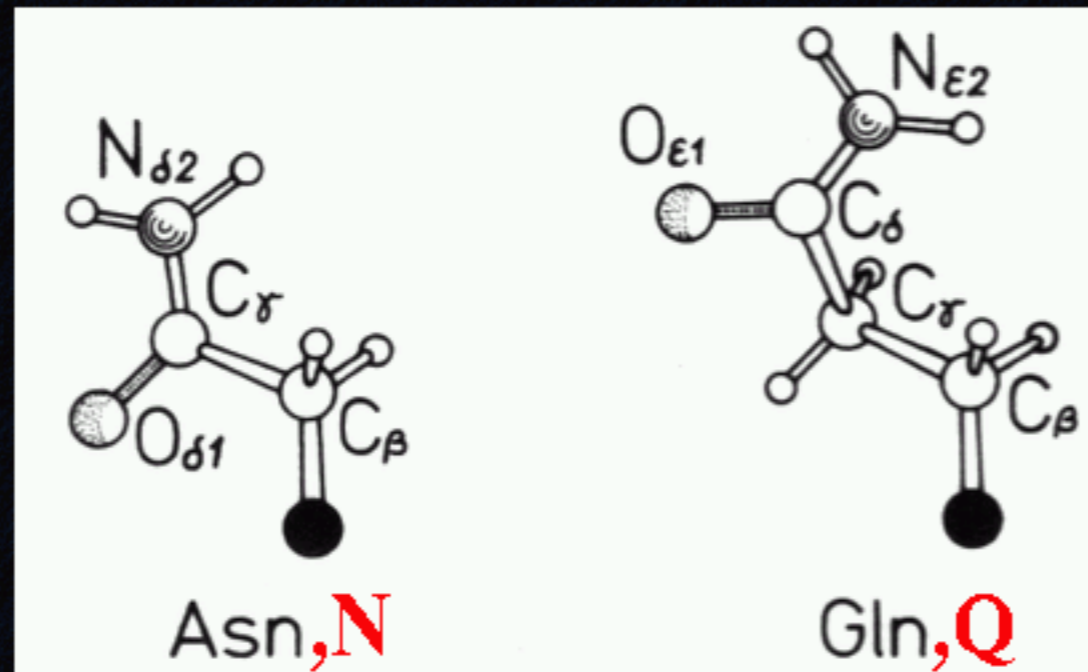
Η σερίνη και η θρεονίνη είναι τυπικά αφόρτιστα-πολικά κατάλοιπα. Έχουν και οι δύο υδροξυλομάδες οι οποίες μπορούν να συμμετάσχουν στο σχηματισμό υδρογονικών δεσμών. Το C β της θρεονίνης είναι ασύμμετρο (και όπως συνήθως, μόνο το ένα ισομερές απαντάται σε φυσικές πρωτεΐνες).

Cys (C)



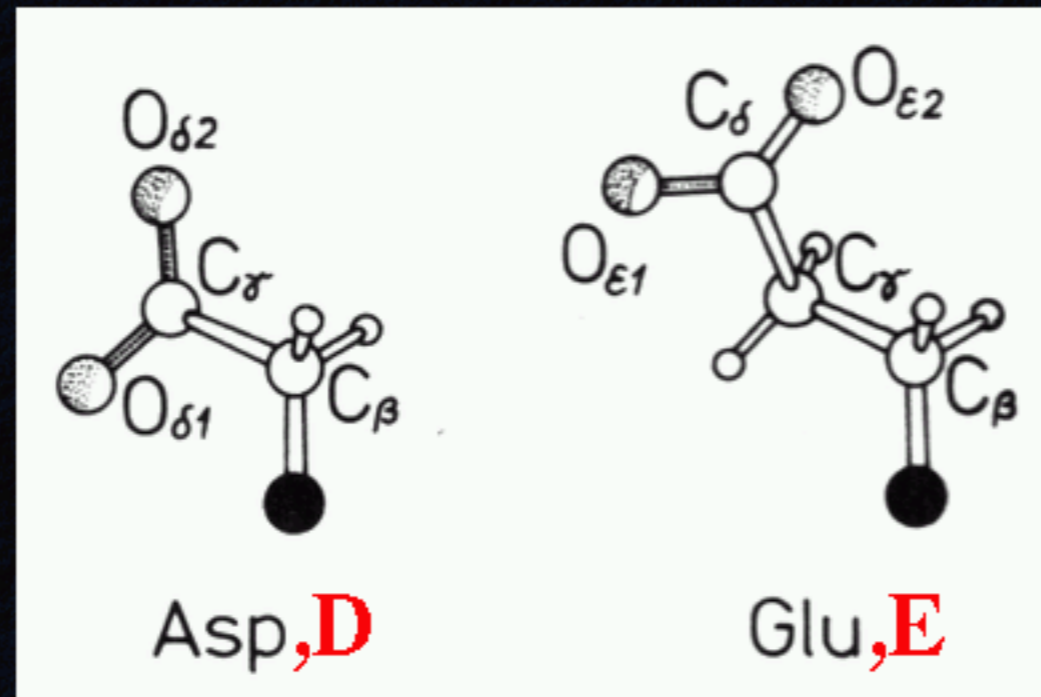
Η κυστεΐνη είναι η μοναδική πλευρική ομάδα η οποία (μέσω της οξείδωσης της σε κυστίνη) μπορεί να δημιουργήσει ομοιοπολικούς δεσμούς οι οποίοι είναι ειδικοί και χαρακτηριστικοί για μια συγκεκριμένη πρωτεϊνική δομή. Οι δισουλφιδικές γέφυρες σταθεροποιούν την τρισδιάστατη δομή και είναι ιδιαίτερα συνηθισμένες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

Asn (N), Gln (Q)



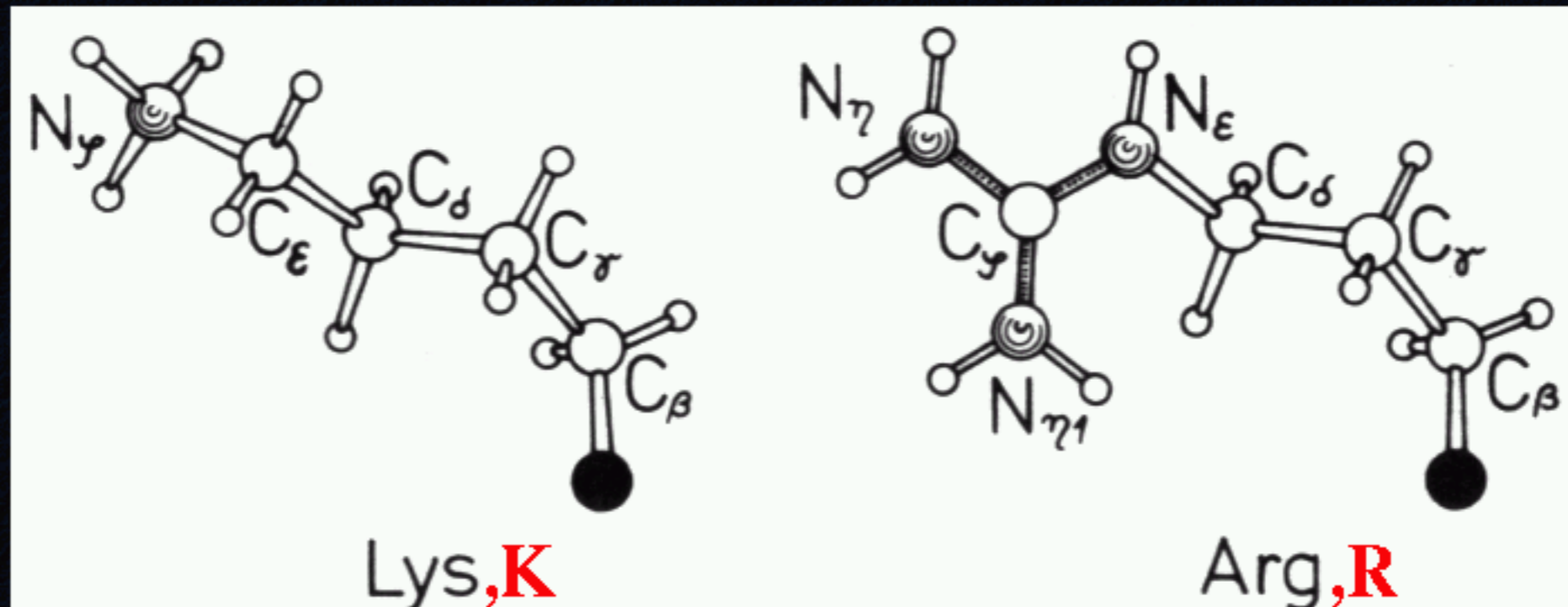
Η ασπαραγίνη και η γλουταμίνη έχουν την ιδιαιτερότητα ότι μπορούν ταυτόχρονα να είναι δότες (αμιδική ομάδα) και δέκτες (καρβονυλική ομάδα) υδρογονοδεσμών. Είναι σχετικά ευκίνητες πλευρικές ομάδες (ιδιαίτερα της γλουταμίνης) οι οποίες όταν βρίσκονται στο εσωτερικό πρωτεϊνών σχεδόν πάντα εμπλέκονται στο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

Asp (D), Glu (E)



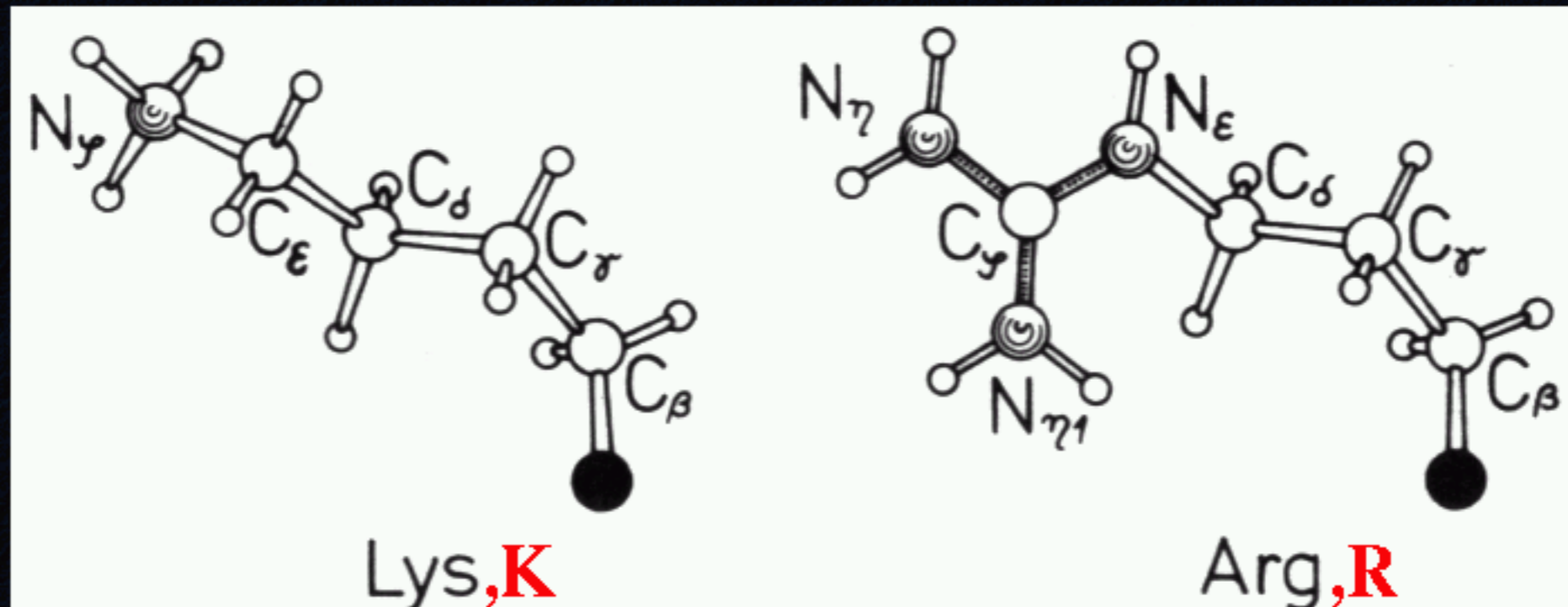
Το ασπαρτικό οξύ και το γλουταμικό οξύ είναι αρνητικά φορτισμένα αμινοξέα (σε φυσιολογικά pH : τα pK είναι 3.9 και 4.3 αντίστοιχα). Η σχετικά μικρή πλευρική ομάδα του ασπαρτικού μειώνει το εύρος των πιθανών διαμορφώσεων του και το κάνει πιο συνηθισμένο σε ενεργά κέντρα ενζύμων.

Lys (K), Arg (R)



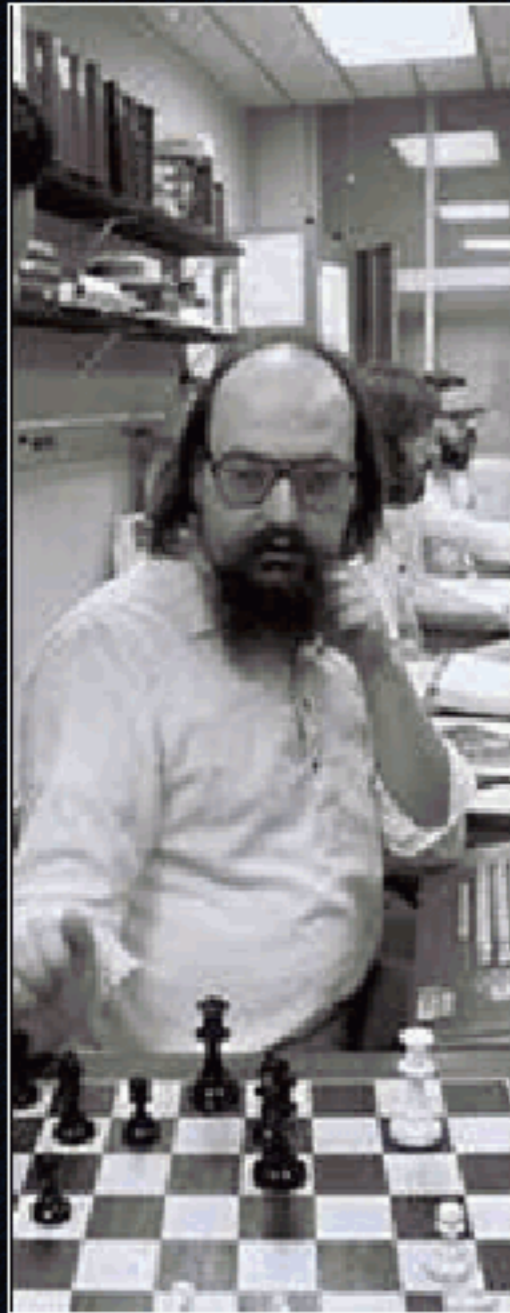
Η λυσίνη και η αργινίνη είναι θετικά φορτισμένα αμινοξέα (σε φυσιολογικά pH : τα pK είναι 10.5 και 12.5 αντίστοιχα). Είναι από τις πλέον ευκίνητες επιφανειακές πλευρικές ομάδες των πρωτεϊνών, σε τέτοιο βαθμό ώστε συνηθέστατα να μην υπάρχει πληροφορία για τη διαμόρφωση τους (τουλάχιστον από κρυσταλλογραφικά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

Unix vs. Windows ?



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 ?

Βιοπληροφορική 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

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 ιδιοτυπικά 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.

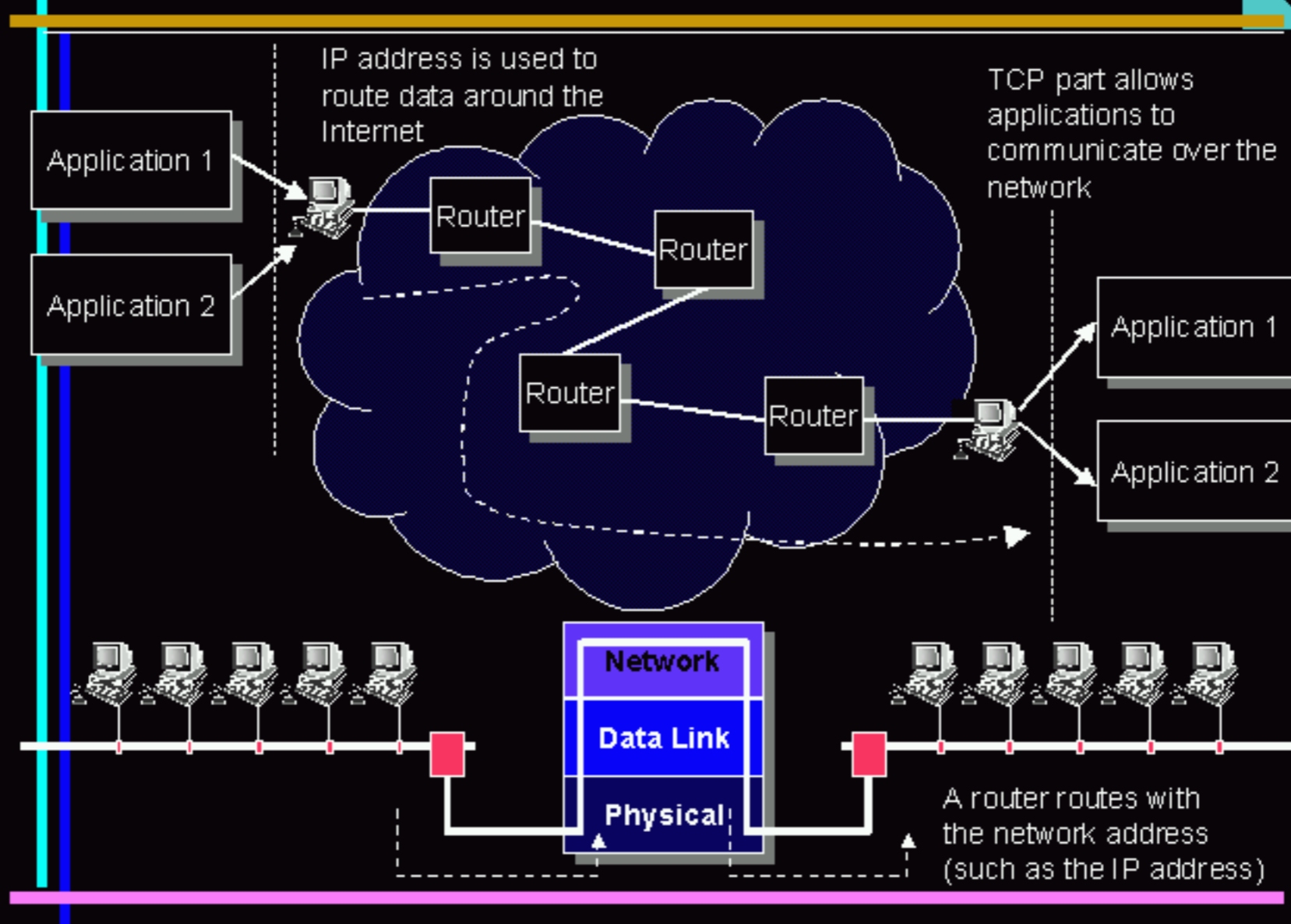
Η έλευση του δικτύου άλλαξε άρδην το κυρίαρχο μοντέλο : the client-server model.

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.

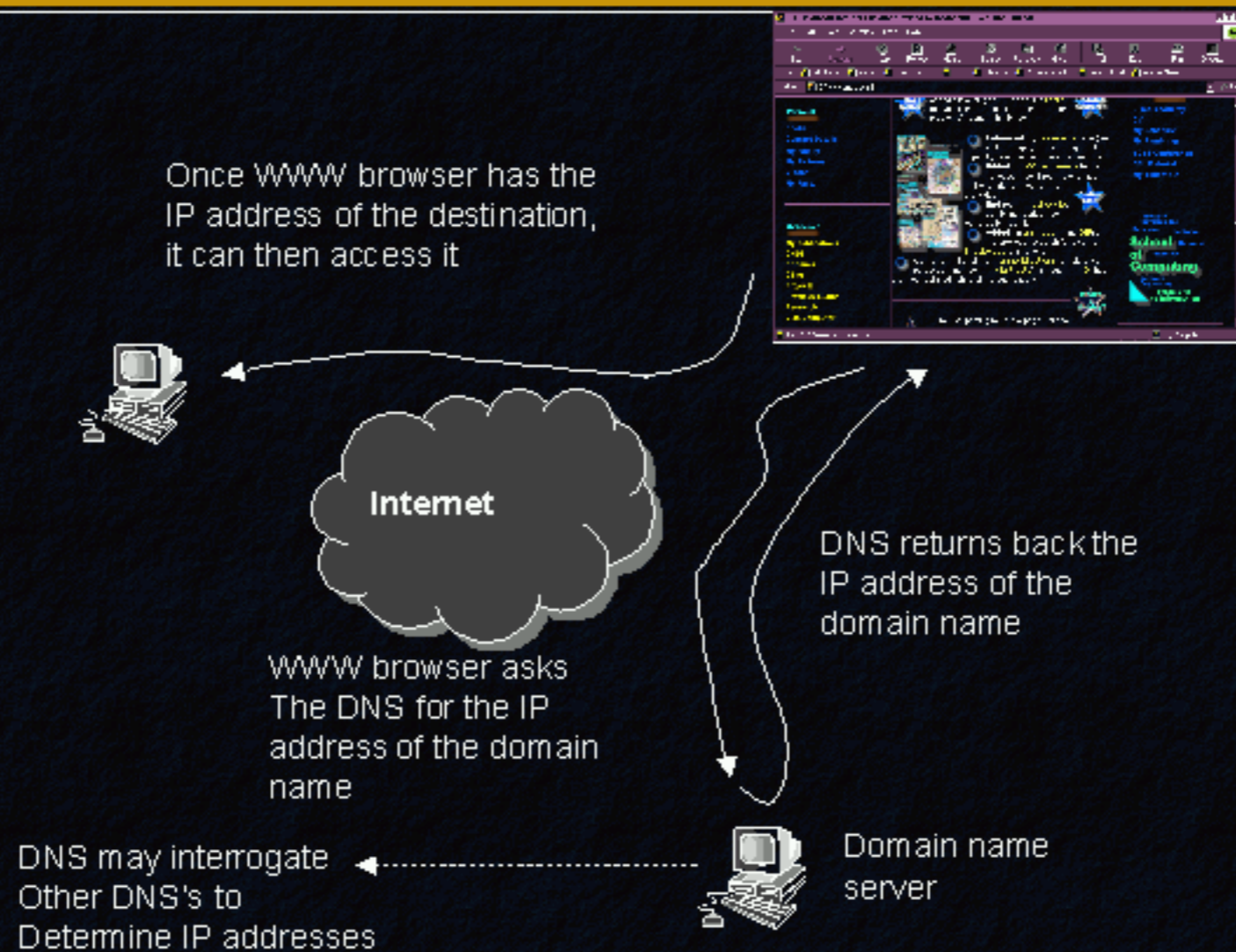
Παρένθεση : TCP/IP

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.

Βάσεις δεδομένων

Βάσεις δεδομένων

Ανάλογα με τον τρόπο αποθήκευσης των δεδομένων μπορεί να είναι απλές συλλογές αρχείων (flat-file), σχεσιακές (relational) ή και αντικειμενοστραφείς (object-oriented databases).

Αυτή η διάκριση αφορά τους μηχανισμούς αποθήκευσης και διαχείρισης των δεδομένων και όχι τον τύπο των βιολογικών δεδομένων που περιέχουν (αλληλουχίες, μοτίβα, δομές, ...).

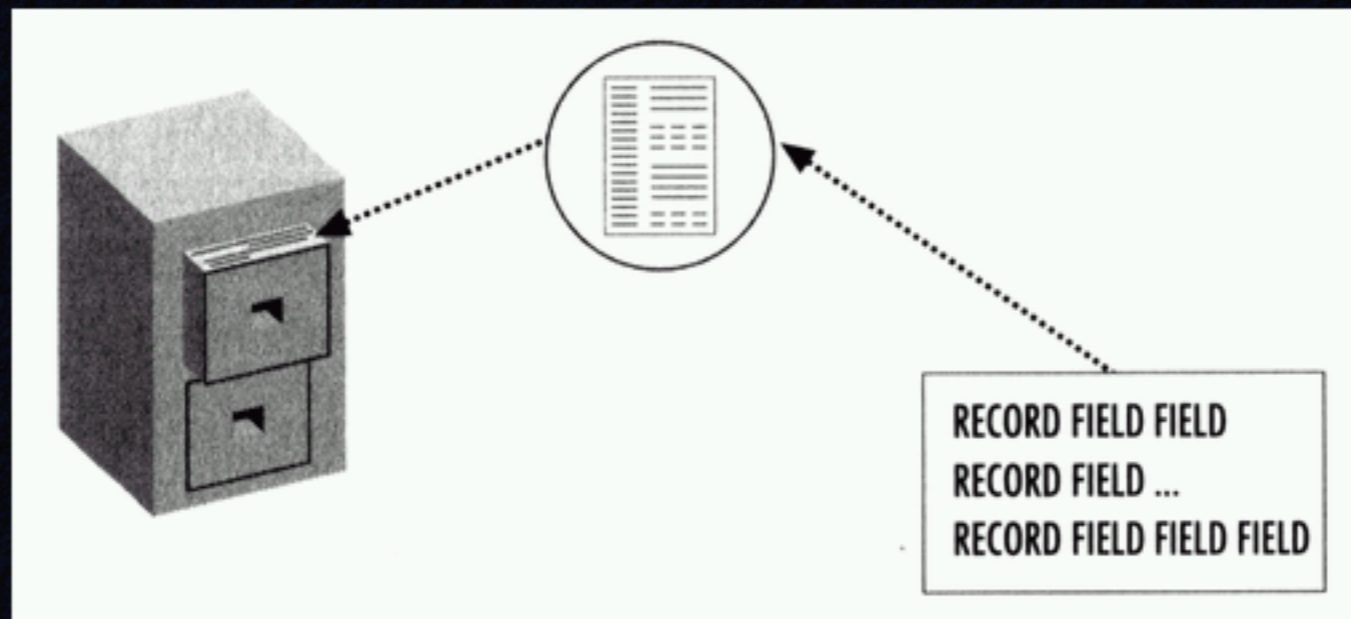
Η μεγάλη πλειοψηφία από τις γνωστότερες βάσεις δεδομένων είναι (ή παρουσιάζονται στον τελικό χρήστη σαν να είναι) του τύπου flat-file (π.χ. GenBank, Swiss-Prot, PDB, Prosite, ...).

Βάσεις τύπου flat-file

Πρόκειται για οργανωμένες συλλογές αρχείων με καθορισμένη (για κάθε αρχείο) μορφή (format).

Βάση => Αρχεία => Εγγραφές => Πεδία

Η αναζήτηση πληροφοριών σε αυτού του τύπου τις βάσεις πραγματοποιείται μέσω της χρήσης δεικτών.



Βάσεις τύπου flat-file

Παράδειγμα PDB καταχώρησης.

```
JRNL      AUTH      S.KLIMASAUSKAS,S.KUMAR,R.J.ROBERTS,X.CHENG      1MHT
JRNL      TITL      HHAI METHYLTRANSFERASE FLIPS ITS TARGET BASE OUT OF      1MHT
JRNL      TITL 2    THE DNA HELIX      1MHT
JRNL      REF       CELL(CAMBRIDGE,MASS.)      V. 76 357 1994      1MHT
JRNL      REFN      ASTM CELLB5 US ISSN 0092-8674      0998      1MHT
REMARK    1      1MHT
REMARK    1 REFERENCE 1      1MHT
REMARK    1 AUTH      X.CHENG,S.KUMAR,J.POSFAI,J.W.PFLUGRATH,R.J.ROBERTS      1MHT
REMARK    1 TITL      CRYSTAL STRUCTURE OF THE HHAI DNA METHYLTRANSFERASE      1MHT
REMARK    1 TITL 2    COMPLEXED WITH S-ADENOSYL-L-METHIONINE      1MHT
REMARK    1 REF       CELL(CAMBRIDGE,MASS.)      V. 74 299 1993      1MHT
REMARK    1 REFN      ASTM CELLB5 US ISSN 0092-8674      0998      1MHT
.....
ATOM      1  N  MET A  1      -31.051  33.174  72.182  1.00 18.75      1MHT
ATOM      2  CA MET A  1      -29.948  32.944  71.220  1.00 19.58      1MHT
ATOM      3  C  MET A  1      -30.533  33.023  69.837  1.00 22.57      1MHT
ATOM      4  O  MET A  1      -31.738  33.171  69.691  1.00 32.38      1MHT
ATOM      5  CB MET A  1      -29.249  31.598  71.463  1.00 17.84      1MHT
```


Σχισιακές βάσεις δεδομένων

Σε αυτές, τα δεδομένα είναι οργανωμένα σε πίνακες. Κάθε πίνακας περιέχει πληροφορία για ένα αυτοτελές τμήμα της καταχώρησης (π.χ. ατομικές θέσεις, βιβλιογραφικές αναφορές, ...). Οι πίνακες με την σειρά τους είναι οργανωμένοι σε γραμμές (το αντίστοιχο των εγγραφών για τις flat-file) και οι γραμμές σε πεδία. Κάθε πεδίο πρέπει να περιέχει μία διακριτή πληροφορία (δεν μπορεί, για παράδειγμα, να είναι ένας κατάλογος από ονόματα).

Σχεσιακές βάσεις δεδομένων

mmCIF καταχώρηση : Πίνακας αναφορών

```
loop_
_citation.id
_citation.title
_citation.journal_abbrev
_citation.journal_volume
_citation.page_first
_citation.page_last
_citation.year
_citation.journal_id_ASTM
_citation.country
_citation.journal_id_ISSN
_citation.journal_id_CSD
_citation.book_publisher
_citation.pdbx_database_id_PubMed
primary 'HhaI methyltransferase flips its target base out of the DNA helix.'
Cell          76 357 369 1994 CELLB5 US 0092-8674 0998 ? 8293469
1
;Crystal Structure of the HhaI DNA Methyltransferase Complexed with SAM
;
'Cell (Cambridge,Mass.)' 74 299 ?    1993 CELLB5 US 0092-8674 0998 ? ?
```


Σχεσιακές βάσεις δεδομένων

mmCIF καταχώρηση : Πίνακας συγγραφέων

```
loop_
_citation_author.citation_id
_citation_author.name
primary 'Klimasauskas, S.'
primary 'Kumar, S.'
primary 'Roberts, R.J.'
primary 'Cheng, X.'
1       'Cheng, X.'
1       'Kumar, S.'
1       'Posfai, J.'
1       'Pflugrath, J.W.'
1       'Roberts, R.J.'
```


Σχισιακές βάσεις δεδομένων

mmCIF καταχώρηση : Πίνακας ατομικών θέσεων

```
loop_
_atom_site.group_PDB
_atom_site.id
_atom_site.type_symbol
_atom_site.label_atom_id
_atom_site.label_alt_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_entity_id
_atom_site.label_seq_id
_atom_site.pdbx_PDB_ins_code
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
ATOM      1      P P      . G      A 1 1      ? -22.481 25.283 100.660 1.00 56.36
ATOM      2      O OP1    . G      A 1 1      ? -21.792 25.069 102.003 1.00 56.38
```


Πρωτοταγείς βάσεις

Ανάλογα με τον τύπο των δεδομένων, οι βάσεις διακρίνονται σε πρωτοταγείς και δευτεροταγείς. Οι πρωτοταγείς βάσεις περιέχουν την πειραματικά προσδιορισμένη πληροφορία, για παράδειγμα τις αλληλουχίες νουκλεϊκών οξέων και πρωτεϊνών. Οι πλέον γνωστές πρωτοταγείς βάσεις νουκλεϊκών οξέων είναι οι : EMBL (European Molecular Biology Laboratory, Ευρώπη), GenBank (NCBI, Αμερική) και DDBJ (Ιαπωνία). Οι πλέον γνωστές πρωτοταγείς βάσεις πρωτεϊνικών αλληλουχιών είναι η SWISS-PROT, η PIR (Protein Information Resource), η TrEMBL (Translated EMBL), ...

Σύνθετες πρωτοταγείς βάσεις

Η πληθώρα των πρωτοταγών βάσεων δεδομένων, μαζί με την ετερογένεια του τρόπου αποθήκευσης των πληροφοριών σε αυτές, οδήγησε στην ανάγκη δημιουργίας σύνθετων βάσεων δεδομένων. Αυτές (όπως μαρτυρεί και το όνομα τους) είναι βάσεις οι οποίες χρησιμοποιούν πληροφορία από πολλές άλλες πρωτοταγείς βάσεις, την φιλτράρουν (για να αφαιρεθούν πολλαπλές παρουσίες των ίδιων δεδομένων) και την παρουσιάζουν στον τελικό χρήστη με έναν ενιαίο τρόπο. Παραδείγματα τέτοιων είναι οι : OWL(SWISS-PROT, PIR, GenBank, NRL-3D), NRDB, UniProt (SWISS-PROT+TrEMBL).

Δευτεροταγείς βάσεις

Οι δευτεροταγείς βάσεις περιέχουν πληροφορία η οποία προήλθε από την ανάλυση των πρωτοταγών βάσεων. Παραδείγματα τέτοιων βάσεων για π.χ. την εύρεση μοτίβων (με βάση πολλαπλές στοιχίσεις) είναι :

PROSITE	SWISS-PROT	Patterns
Profiles	SWISS-PROT	Weighted Matrices
PRINTS	OWL	Fingerprints
BLOCKS	PROSITE/PRINTS	Aligned motifs

Όπως και για τις πρωτοταγείς, έτσι και για τις δευτεροταγείς υπάρχουν σύνθετες βάσεις (π.χ. ProWeb).

Βάσεις, πολλές βάσεις ...

Data Banks Available at EMBL-Heldelberg

03-Oct-2003 16:35

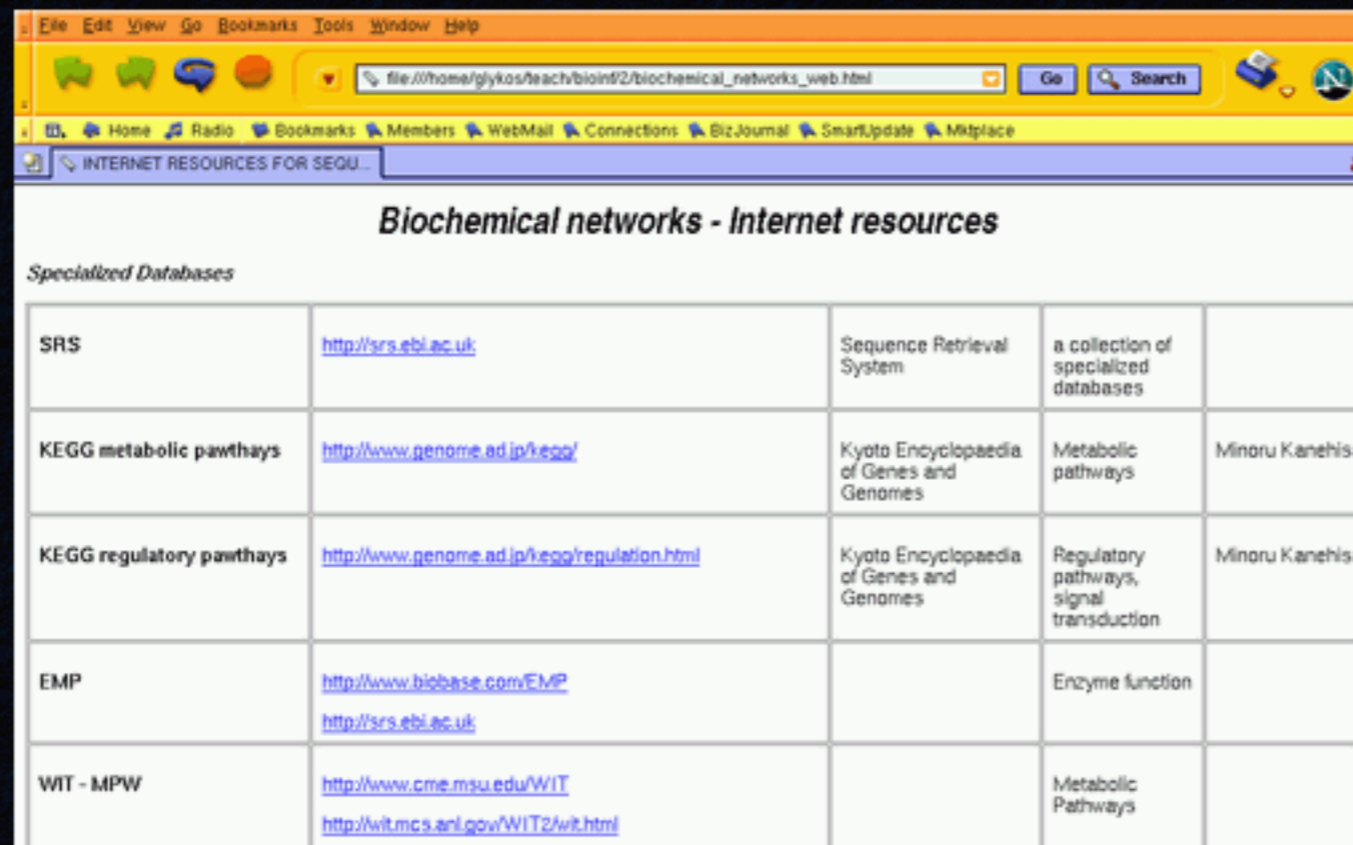
Data Bank	Release	No Entries	Indexing Date	Group	Availability
SWISSPROT		122564	04-Apr-2003	Sequence	ok
SWISSNEW		51725	11-Sep-2003	Sequence	ok
NRDB		1324752	20-Sep-2003	Sequence	ok
SWALL		1241995	11-Sep-2003	Sequence	ok
TREMBLNEW		232563	20-Sep-2003	Sequence	ok
TREMBL		1533261	21-Sep-2003	Sequence	ok
SPTREMBL		944868	09-Jul-2003	Sequence	ok
SPTREMBLNEW		195509	15-Sep-2003	Sequence	ok
REMTREMBL		90492	09-Jul-2003	Sequence	ok
PIR		283346	09-Sep-2003	Sequence	ok
WORMPEP		19538	01-Oct-2003	Sequence	ok
DROSOPHILA		14100	05-Apr-2003	Sequence	ok
EMBLNEW		2664786	10-Jun-2003	Sequence	ok
EMBL		9246940	17-Sep-2003	Sequence	ok
EMBLEST		18001332	19-Sep-2003	Sequence	ok
GENBANK		8497982	24-Jul-2003	Sequence	ok

Δηλαδή, πολλές βάσεις.

GENBANKEST	17026746	26-Jul-2003	Sequence	ok
REFSEQP	19804	11-Sep-2003	Sequence	ok
SUBTILIST	1	07-Apr-2003	Sequence	ok
PROSITE	1649	11-Sep-2003	SeqRelated	ok
PROSITEDOC	1213	11-Sep-2003	SeqRelated	ok
BLOCKS	4034	07-Apr-2003	SeqRelated	ok
EPD	1375	07-Apr-2003	SeqRelated	ok
ENZYME	4173	21-Aug-2003	SeqRelated	ok
PRINTS	865	07-Apr-2003	SeqRelated	ok
TFSITE	4342	07-Apr-2003	TransFac	ok
TFFACTOR	1799	07-Apr-2003	TransFac	ok
TFCELL	816	07-Apr-2003	TransFac	ok
TFCLASS	27	07-Apr-2003	TransFac	ok
TFMATRIX	246	07-Apr-2003	TransFac	ok
TFGENE	1035	07-Apr-2003	TransFac	ok
PDB	21838	29-Jul-2003	Protein3DStruct	ok
DSSP	20140	29-Jul-2003	Protein3DStruct	ok
HSSP	19838	29-Jul-2003	Protein3DStruct	ok
PDBFINDER	22448	15-Sep-2003	Protein3DStruct	ok
NRL3D	6063	07-Apr-2003	Protein3DStruct	ok

Δηλαδή, πολλές βάσεις.

FLYGENES	7556	07-Apr-2003	Genome	ok
FLYREFS	0	07-Apr-2003	Genome	ok
OMIM	12322	07-Apr-2003	Mutations	ok
REPTILIA	8134	21-Aug-2003	Others	ok



Biochemical networks - Internet resources

Specialized Databases

SRS	http://srs.ebi.ac.uk	Sequence Retrieval System	a collection of specialized databases	
KEGG metabolic pathways	http://www.genome.ad.jp/kegg/	Kyoto Encyclopaedia of Genes and Genomes	Metabolic pathways	Minoru Kanehisa
KEGG regulatory pathways	http://www.genome.ad.jp/kegg/regulation.html	Kyoto Encyclopaedia of Genes and Genomes	Regulatory pathways, signal transduction	Minoru Kanehisa
EMP	http://www.biobase.com/EMP http://srs.ebi.ac.uk		Enzyme function	
WIT - MPW	http://www.cme.msu.edu/WIT http://wit.mcs.anl.gov/WIT2/wit.html		Metabolic Pathways	

Βάσεις, αυξανόμενες βάσεις ...

Data Banks Available at EMBL-Heldelberg

03-Oct-2003 16:35

Data Bank	Release	No Entries	Indexing Date	Group	Availability
SWISSPROT		122564	04-Apr-2003	Sequence	ok
SWISSNEW		51725	11-Sep-2003	Sequence	ok
NRDB		1324752	20-Sep-2003	Sequence	ok
SWALL		1241995	11-Sep-2003	Sequence	ok
TREMBLNEW		232563	20-Sep-2003	Sequence	ok
TREMBL		1533261	21-Sep-2003	Sequence	ok
SPTREMBL		944868	09-Jul-2003	Sequence	ok
SPTREMBLNEW		195509	15-Sep-2003	Sequence	ok
REMTREMBL		90492	09-Jul-2003	Sequence	ok
PIR		283346	09-Sep-2003	Sequence	ok
WORMPEP		19538	01-Oct-2003	Sequence	ok
DROSOPHILA		14100	05-Apr-2003	Sequence	ok
EMBLNEW		2664786	10-Jun-2003	Sequence	ok
EMBL		9246940	17-Sep-2003	Sequence	ok
EMBLEST		18001332	19-Sep-2003	Sequence	ok
GENBANK		8497982	24-Jul-2003	Sequence	ok

Βάσεις, αυξανόμενες βάσεις ...

Data Banks Available at EMBL-Heldelberg

26-Oct-2005 10:35

Data Bank	Release No	Entries	Indexing Date	Group	Availability
SWISSPROT		163235	10-Jun-2005	Sequence	ok
SWISSNEW		75960	26-Oct-2005	Sequence	ok
NRDB		2116230	11-Oct-2005	Sequence	ok
SWALL		2612174	6-Oct-2005	Sequence	ok
TREMBLNEW		35463	09-Oct-2005	Sequence	ok
TREMBL		2576118	04-Oct-2005	Sequence	ok
SPTREMBL		1449374	6-Jun-2005	Sequence	ok
SPTREMBLNEW		143140	17-Jun-2005	Sequence	ok
REMTREMBL		92182	20-Jun-2005	Sequence	ok
PIR		283416	6-Jun-2005	Sequence	ok
WORMPEP		19538	16-Jun-2005	Sequence	ok
DROSOPHILA		14100	16-Jun-2005	Sequence	ok
EMBLNEW		3016836	19-Aug-2005	Sequence	ok
EMBL		18544391	19-Sep-2005	Sequence	ok
EMBLEST		28770984	18-Sep-2005	Sequence	ok
GENBANK		17810569	13-Jul-2005	Sequence	ok

Βάσεις, αυξανόμενες βάσεις ...

SWISSPROT	122564	04-Apr-2003	Sequence	ok
SWISSPROT	163235	10-Jun-2005	Sequence	ok
SWISSPROT	399749	14-Oct-2008	Sequence	ok
SWISSPROT	540958	16-Oct-2013	Sequence	ok

Παραδείγματα καταχωρήσεων

GenBank

LOCUS AF111785 5925 bp mRNA PRI 01-SEP-1999
DEFINITION Homo sapiens myosin heavy chain IIx/d mRNA, complete cds.
ACCESSION AF111785
VERSION AF111785.1 GI:4808814
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5925)
AUTHORS Weiss,A., McDonough,D., Wertman,B., Acakpo-Satchivi,L.,
Montgomery,K., Kucherlapati,R., Leinwand,L. and Krauter,K.
TITLE Organization of human and mouse skeletal myosin heavy chain gene
clusters is highly conserved
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (6), 2958-2963 (1999)
MEDLINE 99178997
PUBMED 10077619
REFERENCE 2 (bases 1 to 5925)
AUTHORS Weiss,A., Schiaffino,S. and Leinwand,L.A.
TITLE Comparative sequence analysis of the complete human sarcomeric
myosin heavy chain family: implications for functional diversity
JOURNAL J. Mol. Biol. 290 (1), 61-75 (1999)
MEDLINE 99318869
PUBMED 10388558
REFERENCE 3 (bases 1 to 5925)
AUTHORS Weiss,A. and Leinwand,L.A.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1998) MCDB, University of Colorado at Boulder,
Campus Box 0347, Boulder, Colorado 80309-0347, USA

Παραδείγματα : GenBank

```
FEATURES                     Location/Qualifiers
    source                     1..5925
                                /organism="Homo sapiens"
                                /mol_type="mRNA"
                                /db_xref="taxon:9606"
                                /chromosome="17"
                                /map="17p13.1"
                                /tissue_type="skeletal muscle"
    CDS                         1..5820
                                /note="MyHC"
                                /codon_start=1
                                /product="myosin heavy chain IIx/d"
                                /protein_id="AAD29951.1"
                                /db_xref="GI:4808815"
                                /translation="MSSDSEMAIFGEAAPFLRKSERERIEAQNKPFDAKTSVFVVDPK
                                ESFVKATVQSREGGKVTAKTEAGATVTVKDDQVFPMPKPKYDKIEDMAMMTHLHEPAV
                                LYNLKERYAAWMIYTYSGLFCVTVNPKWLPVYNAEVVTAIRGKKRQEAPPHIFSISD
                                .....
                                GLRKHHERKVVELTYQTEEDRKNILRLQDLVDKQLQAKVKS YKRQAEAEAEQSNVNLSKF
                                RRIQHELEAEERADIAESQVNKLRVKSREVHTKI ISEE"
```

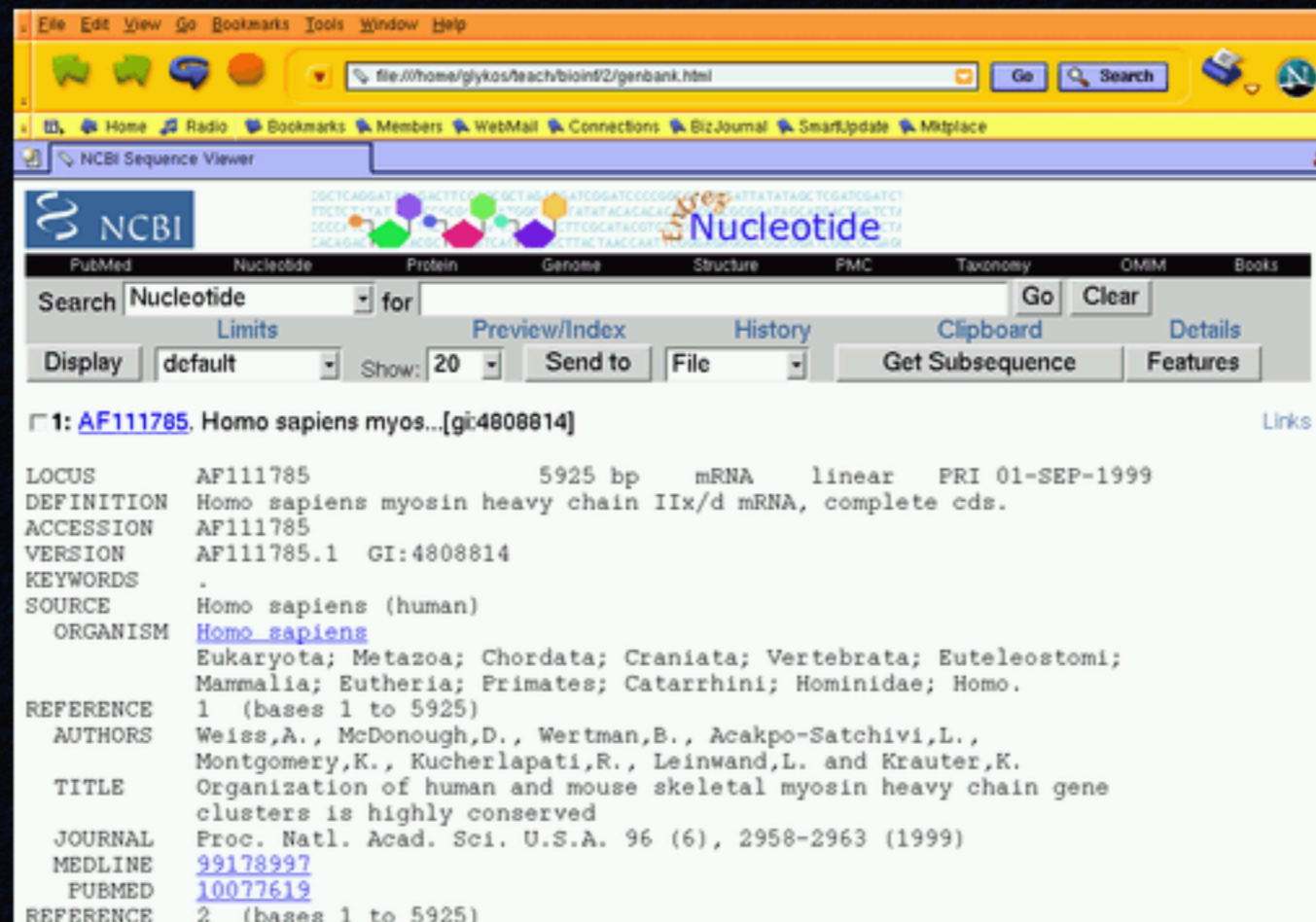
BASE COUNT 1890 a 1300 c 1613 g 1122 t

ORIGIN

```
    1 atgagttctg actctgagat ggccatthtt ggggaggctg ctcctttcct ccgaaagtct
   61 gaaagggagc gaattgaagc ccagaacaag ccttttgatg ccaagacatc agtctttgtg
  121 gtggacccta aggagtcctt tgtgaaagca acagtgcaga gcagggaagg ggggaagggtg
  181 acagctaaga ccgaagctgg agctactgta acagtgaaag atgaccaagt cttcccatg
  .....
 5821 tttatctaac tgctgaaagg tgaccaaga aatgcacaaa atgtgaaaat ctttgtcact
 5881 ccattttgta cttatgactt ttggagataa aaaatttatc tgcca
```


Παραδείγματα : GenBank

Η εμφάνιση της ίδιας καταχώρησης μέσω ενός browser δεν είναι σημαντικά διαφορετική [με την εξαίρεση της ύπαρξης ενεργών συνδέσεων προς άλλες βάσεις δεδομένων].



The screenshot displays the NCBI Sequence Viewer interface in a web browser. The browser's address bar shows the URL: file://home/glykos/teach/bioinf2/genbank.html. The page title is "NCBI Sequence Viewer". The NCBI logo is visible at the top left. A navigation bar includes links for PubMed, Nucleotide, Protein, Genome, Structure, PMC, Taxonomy, OMIM, and Books. The search bar contains "Nucleotide" and "for". Below the search bar are options for "Limits", "Preview/Index", "History", "Clipboard", and "Details". The "Display" dropdown is set to "default", and "Show" is set to "20". There are buttons for "Send to" (set to "File"), "Get Subsequence", and "Features". The main content area shows the entry for **AF111785**, Homo sapiens myos...[gi:4808814]. The entry details include:

- LOCUS: AF111785 5925 bp mRNA linear PRI 01-SEP-1999
- DEFINITION: Homo sapiens myosin heavy chain IIx/d mRNA, complete cds.
- ACCESSION: AF111785
- VERSION: AF111785.1 GI:4808814
- KEYWORDS: .
- SOURCE: Homo sapiens (human)
- ORGANISM: [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
- REFERENCE: 1 (bases 1 to 5925)
- AUTHORS: Weiss,A., McDonough,D., Wertman,B., Acakpo-Satchivi,L., Montgomery,K., Kucherlapati,R., Leinwand,L. and Krauter,K.
- TITLE: Organization of human and mouse skeletal myosin heavy chain gene clusters is highly conserved
- JOURNAL: Proc. Natl. Acad. Sci. U.S.A. 96 (6), 2958-2963 (1999)
- MEDLINE: [99178997](#)
- PUBMED: [10077619](#)
- REFERENCE: 2 (bases 1 to 5925)

Παραδείγματα : GenBank

LOCUS AF111785 5925 bp mRNA PRI 01-SEP-1999

- Κωδικός καταχώρησης (AF111785)
- Μήκος αλληλουχίας σε ζεύγη βάσεων (5925 bp)
- Τύπος μορίου (DNA, RNA)
- Κωδικός τμήματος της βάσης (EST για Expressed Sequence Tags, HTG για High Throughput Genome sequences, ...)
- Ημερομηνία δημοσιοποίησης της αλληλουχίας (01-SEP-1999)

Παραδείγματα : GenBank

DEFINITION Homo sapiens myosin heavy chain IIx/d mRNA, complete cds.

Σύνοψη της βιολογίας της καταχώρησης. Το συντακτικό της για mRNA είναι :

- Γένος είδος (Homo sapiens)
- Όνομα προϊόντος, σύμβολο γονιδίου (myosin heavy chain IIx/d)
- Ακολουθείται από "mRNA, complete cds.", όπου το cds σημαίνει "coding sequence".

Παραδείγματα : GenBank

ACCESSION AF111785
VERSION AF111785.1 GI:4808814

Χαρακτηριστικός κωδικός της καταχώρησης (AF111785)
και έκδοση της (AF111785.1 GI:4808814).

Παραδείγματα : GenBank

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini;
Hominidae; Homo.

Οργανισμός προέλευσης και ταξινόμική του.

Παραδείγματα : GenBank

REFERENCE 1 (bases 1 to 5925)
AUTHORS Weiss,A., McDonough,D., Wertman,B., Acakpo-Satchivi,L.,
Montgomery,K., Kucherlapati,R., Leinwand,L. and Krauter,K.
TITLE Organization of human and mouse skeletal myosin heavy
chain gene clusters is highly conserved
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (6), 2958-2963 (1999)
MEDLINE 99178997
PUBMED 10077619

Αναφορά για την καταχώρηση με συγγραφείς, τίτλο άρθρου, περιοδικό (με τόμο, τεύχος, σελίδες και έτος). Οι κωδικοί "MEDLINE" και "PUBMED" είναι συνδέσεις (για το εν λόγω άρθρο) στις αντίστοιχες βιβλιογραφικές βάσεις δεδομένων.

Παραδείγματα : GenBank

```
FEATURES                     Location/Qualifiers
    source                     1..5925
                               /organism="Homo sapiens"
                               .....
    CDS                         1..5820
                               /note="MyHC"
                               .....
```

Η λέξη-κλειδί "FEATURES" σηματοδοτεί την έναρξη ενός πίνακα με περιγραφές χαρακτηριστικών της καταχώρησης. Ο πίνακας αυτός χωρίζεται σε ενότητες [όπως "source", "CDS" (αρχικά του coding sequence), "Gene", "RNA"]. Κάθε μια από αυτές τις ενότητες χωρίζεται με τη σειρά της σε υπο-ενότητες που σηματοδοτούνται από λέξεις-κλειδιά [όπως "/organism=", "/mol_type=", "/chromosome=", "/codon_start=", ...].

Παραδείγματα : GenBank

FEATURES

source

Location/Qualifiers

1..5925

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/chromosome="17"

/map="17p13.1"

/tissue_type="skeletal muscle"

Χαρακτηριστικά της πηγής της καταχώρησης (οργανισμός, τύπος μορίου, χρωμόσωμα, χαρτογραφική θέση, τύπος ιστού). Η καταχώρηση "/db_xref="taxon:9606" είναι ένα παράδειγμα σύνδεσης με μια άλλη βάση δεδομένων, σε αυτή την περίπτωση, ταξινομική.

Παραδείγματα : GenBank

```
CDS          1..5820
             /note="MyHC"
             /codon_start=1
             /product="myosin heavy chain IIx/d"
             /protein_id="AAD29951.1"
             /db_xref="GI:4808815"
             /translation="MSSDSEMAIFG.....
             .....
             ....AEERADIAESQVNKLRVKSREVHTKIISEE"
```

Χαρακτηριστικά κωδικοποιούσας αλληλουχίας. Δίδονται θέση (1..5820), μια σημείωση ("MyHC" για Myosin Heavy Chain), πλαίσιο ανάγνωσης (codon_start=1), όνομα προϊόντος (myosin heavy chain IIx/d), κωδικός αναγνώρισης του προϊόντος (protein_id="AAD29951.1"), μια σύνδεση για το προϊόν (db_xref="GI:4808815") και η αλληλουχία του προϊόντος (translation="MSS...EE").

Παραδείγματα : GenBank

```
BASE COUNT      1890 a      1300 c      1613 g      1122 t
ORIGIN
      1 atgagttctg actctgagat ggccattttt .....
      .....
     5881 .... cttatgactt ttggagataa aaaatttatc tgcca
//
```

Η καθ'αυτό καταχώρηση. Δίδονται οι συχνότητες των βάσεων (1890-A, 1300-C, 1613-G, 1122-T), και η αλληλουχία το τέλος της οποίας σηματοδοτείται από το σύμβολο "//". Η λέξη-κλειδί "ORIGIN" (σε αυτό το παράδειγμα κενή) χρησιμοποιείται για να σημειωθεί η θέση στο γονιδίωμα της αλληλουχίας που ακολουθεί.

Παραδείγματα καταχωρήσεων

Swiss-Prot

```
ID   GRAA_HUMAN      STANDARD;          PRT;   262 AA.
AC   P12544;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Granzyme A precursor (EC 3.4.21.78) (Cytotoxic T-lymphocyte proteinase
DE   1) (Hanukkah factor) (H factor) (HF) (Granzyme 1) (CTL tryptase)
DE   (Fragmentin 1).
GN   GZMA OR CTLA3 OR HFSP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=T-cell;
RX   MEDLINE=88125000; PubMed=3257574;
RA   Gershenfeld H.K., Hershberger R.J., Shows T.B., Weissman I.L.;
RT   "Cloning and chromosomal assignment of a human cDNA encoding a T
RT   cell- and natural killer cell-specific trypsin-like serine
RT   protease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:1184-1188(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
```


Παραδείγματα : Swiss-Prot

RN [3]
RP SEQUENCE OF 1-23 FROM N.A.
RA Goralski T.J., Krensky A.M.;
RT "The upstream region of the human granzyme A locus contains both
RT positive and negative transcriptional regulatory elements.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.

RN [4]
RP SEQUENCE OF 29-53.
RX MEDLINE=88330824; PubMed=3047119;
RA Poe M., Bennett C.D., Biddison W.E., Blake J.T., Norton G.P.,
RA Rodkey J.A., Sigal N.H., Turner R.V., Wu J.K., Zweerink H.J.;
RT "Human cytotoxic lymphocyte tryptase. Its purification from granules
RT and the characterization of inhibitor and substrate specificity.";
RL J. Biol. Chem. 263:13215-13222(1988).

RN [5]
RP SEQUENCE OF 29-40, AND CHARACTERIZATION.
RX MEDLINE=89009866; PubMed=3262682;
RA Hameed A., Lowrey D.M., Lichtenheld M., Podack E.R.;
RT "Characterization of three serine esterases isolated from human IL-2
RT activated killer cells.";
RL J. Immunol. 141:3142-3147(1988).

RN [6]
RP SEQUENCE OF 29-39, AND CHARACTERIZATION.
RX MEDLINE=89035468; PubMed=3263427;
RA Kraehenbuhl O., Rey C., Jenne D.E., Lanzavecchia A., Groscurth P.,
RA Carrel S., Tschopp J.;
RT "Characterization of granzymes A and B isolated from granules of
RT cloned human cytotoxic T lymphocytes.";
RL J. Immunol. 141:3471-3477(1988).

Παραδείγματα : Swiss-Prot

```
RN [7]
RP 3D-STRUCTURE MODELING.
RX MEDLINE=89184501; PubMed=3237717;
RA Murphy M.E.P., Moulton J., Bleackley R.C., Gershenfeld H.,
RA Weissman I.L., James M.N.G.;
RT "Comparative molecular model building of two serine proteinases from
RT cytotoxic T lymphocytes.";
RL Proteins 4:190-204(1988).
CC -!- FUNCTION: This enzyme is necessary for target cell lysis in cell-
CC mediated immune responses. It cleaves after Lys or Arg. May be
CC involved in apoptosis.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins, including fibronectin,
CC type IV collagen and nucleolin. Preferential cleavage: Arg-|-Xaa,
CC Lys-|-Xaa &gt;&gt; Phe-|-Xaa in small molecule substrates.
CC -!- SUBUNIT: Homodimer; disulfide-linked.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic granules.
CC -!- SIMILARITY: Belongs to peptidase family S1. Granzyme subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```


Παραδείγματα : Swiss-Prot

```
DR   EMBL; M18737; AAA52647.1; -.
DR   EMBL; BC015739; AAH15739.1; -.
DR   EMBL; U40006; AAD00009.1; -.
DR   PIR; A28943; A28943.
DR   PIR; A30525; A30525.
DR   PIR; A30526; A30526.
DR   PIR; A31372; A31372.
DR   PDB; 1HF1; 15-OCT-94.
DR   MEROPS; S01.135; -.
DR   Genew; HGNC:4708; GZMA.
DR   MIM; 140050; -.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00089; trypsin; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Zymogen; Signal; T-cell; Cytolysis;
KW   Apoptosis; 3D-structure.
```


Παραδείγματα : Swiss-Prot

```
FT SIGNAL 1 26
FT PROPEP 27 28 ACTIVATION PEPTIDE.
FT CHAIN 29 262 GRANZYME A.
FT ACT_SITE 69 69 CHARGE RELAY SYSTEM (BY SIMILARITY) .
FT ACT_SITE 114 114 CHARGE RELAY SYSTEM (BY SIMILARITY) .
FT ACT_SITE 212 212 CHARGE RELAY SYSTEM (BY SIMILARITY) .
FT DISULFID 54 70 BY SIMILARITY.
FT DISULFID 148 218 BY SIMILARITY.
FT DISULFID 179 197 BY SIMILARITY.
FT DISULFID 208 234 BY SIMILARITY.
FT CARBOHYD 170 170 N-LINKED (GLCNAC...) (POTENTIAL) .
FT STRAND 30 30
FT STRAND 33 34
FT TURN 37 38
FT TURN 41 42
.....
FT TURN 234 235
FT TURN 237 238
FT STRAND 241 245
FT TURN 246 249
FT HELIX 252 260
SQ SEQUENCE 262 AA; 28968 MW; DA87363A0D92BAF4 CRC64;
MRNSYRFLAS SLSVVVSLLL IPEDVCEKII GGNEVTPHSR PYMVLLSLDR KTICAGALIA
KDWVLTAABC NLNKRQVIL GAHSITREEP TKQIMLVKKE FPYPCYDPAT REGDLKLLQL
TEKAKINKYV TILHLPKKGD DVKPGTMCQV AGWGRTHNSA SWSDTLREVN ITIIDRKVCN
DRNHYNFNPV IGMNMVCAGS LRGGRDSCNG DSGSPLLCEG VFRGVTSFGL ENKCGDPRGP
GVYILLSKKH LNWIIMTIKG AV
```

//

Παραδείγματα : Swiss-Prot

ID GRAA_HUMAN STANDARD; PRT; 262 AA.
AC P12544;

Περιγραφικός κωδικός καταχώρησης, κατάσταση καταχώρησης (PRELIMINARY ή STANDARD), τύπος μορίου (PRT για PRoTein), αριθμός αμινοξικών καταλοίπων (262 Amino-Acids).

Το AC είναι ο σειριακός αριθμός καταχώρησης.

Παραδείγματα : Swiss-Prot

```
DT      01-OCT-1989 (Rel. 12, Created)
DT      01-OCT-1989 (Rel. 12, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
```

Ημερομηνίες δημιουργίας και μεταβολών της καταχώρησης.

Παραδείγματα : Swiss-Prot

DE Granzyme A precursor (EC 3.4.21.78) (Cytotoxic T-lymphocyte
DE proteinase 1) (Hanukkah factor) (H factor) (HF) (Granzyme 1)
DE (CTL tryptase) (Fragmentin 1).

Συνοπτική περιγραφή της καταχώρησης. Συνήθως αρκεί για την πλήρη ταυτοποίηση της πρωτεΐνης.

Παραδείγματα : Swiss-Prot

GN GZMA OR CTLA3 OR HFSP.

Το όνομα του γονιδίου που κωδικοποιεί για την πρωτεΐνη που περιγράφεται στην καταχώρηση (GN για Gene Name).

Παραδείγματα : Swiss-Prot

```
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI_TaxID=9606;
```

Οργανισμός προέλευσης, ταξινόμική του, σύνδεση με εξωτερική βάση ταξινόμικών δεδομένων (Organism Species, Organism Classification, Organism taxonomy).

Παραδείγματα : Swiss-Prot

```
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=T-cell;
RX      MEDLINE=88125000; PubMed=3257574;
RA      Gershenfeld H.K., Hershberger R.J., Shows T.B., Weissman I.L.;
RT      "Cloning and chromosomal assignment of a human cDNA encoding a T
RT      cell- and natural killer cell-specific trypsin-like serine
RT      protease.";
RL      Proc. Natl. Acad. Sci. U.S.A. 85:1184-1188(1988).
```

Πεδία βιβλιογραφικών αναφορών. Δίδονται αύξων αριθμός αναφοράς (RN, Reference Number), η εργασία που εκτελέστηκε (RP, Reference Position), σχόλια (RC, Reference Comment), συνδέσεις με βιβλιογραφικές βάσεις (RX, Reference cross-reference), συγγραφείς (RA, Reference Author), τίτλος (RT, Reference Title), περιοδικό (RL, Reference Location).

Παραδείγματα : Swiss-Prot

```
RN      [7]
RP      3D-STRUCTURE MODELING.
RX      MEDLINE=89184501; PubMed=3237717;
RA      Murphy M.E.P., Moulton J., Bleackley R.C., Gershenfeld H.,
RA      Weissman I.L., James M.N.G.;
RT      "Comparative molecular model building of two serine proteinases
RT      from cytotoxic T lymphocytes.";
RL      Proteins 4:190-204(1988).
```

Πεδία βιβλιογραφικών αναφορών. Δίδονται αύξων αριθμός αναφοράς (RN, Reference Number), η εργασία που εκτελέστηκε (RP, Reference Position), σχόλια (RC, Reference Comment), συνδέσεις με βιβλιογραφικές βάσεις (RX, Reference cross-reference), συγγραφείς (RA, Reference Author), τίτλος (RT, Reference Title), περιοδικό (RL, Reference Location).

Παραδείγματα : Swiss-Prot

```
CC  -!- FUNCTION: This enzyme is necessary for target cell lysis in
CC      cell-mediated immune responses. It cleaves after Lys or Arg.
CC      May be involved in apoptosis.
CC  -!- CATALYTIC ACTIVITY: Hydrolysis of proteins, including
CC      fibronectin, type IV collagen and nucleolin. Preferential
CC      cleavage: Arg-|-Xaa, Lys-|-Xaa >> Phe-|-Xaa in small
CC      molecule substrates.
CC  -!- SUBUNIT: Homodimer; disulfide-linked.
CC  -!- SUBCELLULAR LOCATION: Cytoplasmic granules.
CC  -!- SIMILARITY: Belongs to peptidase family S1.
CC      Granzyme subfamily.
```

Σχόλια για την καταχώρηση. Αυτά είναι οργανωμένα σε θέματα όπως ALLERGEN, ALTERNATIVE PRODUCTS, BIOTECHNOLOGY, CATALYTIC ACTIVITY, COFACTOR, DISEASE, DOMAIN, FUNCTION, INDUCTION, ...

Παραδείγματα : Swiss-Prot

```
DR   EMBL; M18737; AAA52647.1; -.
DR   EMBL; BC015739; AAH15739.1; -.
DR   EMBL; U40006; AAD00009.1; -.
DR   PIR; A28943; A28943.
DR   PIR; A30525; A30525.
DR   PIR; A30526; A30526.
DR   PIR; A31372; A31372.
DR   PDB; 1HF1; 15-OCT-94.
DR   MEROPS; S01.135; -.
DR   Genew; HGNC:4708; GZMA.
DR   MIM; 140050; -.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00089; trypsin; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
```

Αναφορές σε εξωτερικές βάσεις δεδομένων. Δίδονται η κωδική ονομασία βάσης και οι κωδικοί καταχώρησης.

Παραδείγματα : Swiss-Prot

KW Hydrolase; Serine protease; Zymogen; Signal; T-cell; Cytolysis;
KW Apoptosis; 3D-structure.

Λέξεις-κλειδιά για την καταχώρηση. Αυτές επιλέγονται από έναν προϋπάρχοντα κατάλογο πιθανών λέξεων-κλειδιών (ορίζονται από την Swiss-Prot).

Παραδείγματα : Swiss-Prot

FT	SIGNAL	1	26	
FT	PROPEP	27	28	ACTIVATION PEPTIDE.
FT	CHAIN	29	262	GRANZYME A.
FT	ACT_SITE	69	69	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	114	114	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	212	212	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	DISULFID	54	70	BY SIMILARITY.
FT	DISULFID	148	218	BY SIMILARITY.
FT	DISULFID	179	197	BY SIMILARITY.
FT	DISULFID	208	234	BY SIMILARITY.
FT	CARBOHYD	170	170	N-LINKED (GLCNAC...) (POTENTIAL).
FT	STRAND	30	30	
FT	STRAND	33	34	

Πίνακας με χαρακτηριστικά γνωρίσματα της αλληλουχίας όπως : μετά-μεταφραστικές τροποποιήσεις, κατάλοιπα ενεργού κέντρου, στοιχεία δευτεροταγούς δομής, ...

Παραδείγματα : Swiss-Prot

```
SQ SEQUENCE 262 AA; 28968 MW; DA87363A0D92BAF4 CRC64;  
MRNSYRFLAS SLSVVVSLLL IPEDVCEKII GGNEVTPHSR PYMVLLSLDR KTICAGALIA  
KDWVLTAHC NLNKRSQLVIL GAHSITREEP TKQIMLVKKE FPYPCYDPAT REGDLKLLQL  
TEKAKINKYV TILHLPKKGD DVKPGTMCQV AGWGRTHNSA SWSDTLREVN ITIIDRKVCN  
DRNHYNFNPV IGMNMVCAGS LRGGRDSCNG DSGSPLLCEG VFRGVTSFGL ENKCGDPRGP  
GVYILLSKKH LNWIIMTIKG AV
```

//

Η καθ'αυτό αλληλουχία. Η πρώτη γραμμή δίδει :
αριθμό καταλοίπων, μοριακό βάρος (σε D) και,
τέλος, το αποτύπωμα της αλληλουχίας (CRC για
Cyclic Redundancy Check) για λόγους επιβεβαίωσης
της ορθότητας της αλληλουχίας.

Παραδείγματα : Swiss-Prot

Ενδιάμεσοι εξυπηρετητές όπως το SRS ή Entrez, διαμορφώνουν τις καταχωρήσεις ώστε να είναι αναγνώσιμες (σε απλά Αγγλικά) ...

The screenshot shows a web browser window displaying the Swiss-Prot entry for GRAA_HUMAN (P12544). The browser's address bar shows the URL: file://home/glykos/teach/bioinf2/swiss-nice.html. The page title is "Text Entry | SwissEntry | NiceProt". The main content area is titled "Entry 1 of 2 from Query 1" and includes a "Reset" button and a "Next Entry" link. The entry information is displayed in a table format with the following details:

General information	
Entry name	GRAA_HUMAN
Accession number	P12544
Created	Rel. 12, 1-OCT-1989
Sequence update	Rel. 12, 1-OCT-1989
Annotation update	Rel. 42, 15-SEP-2003

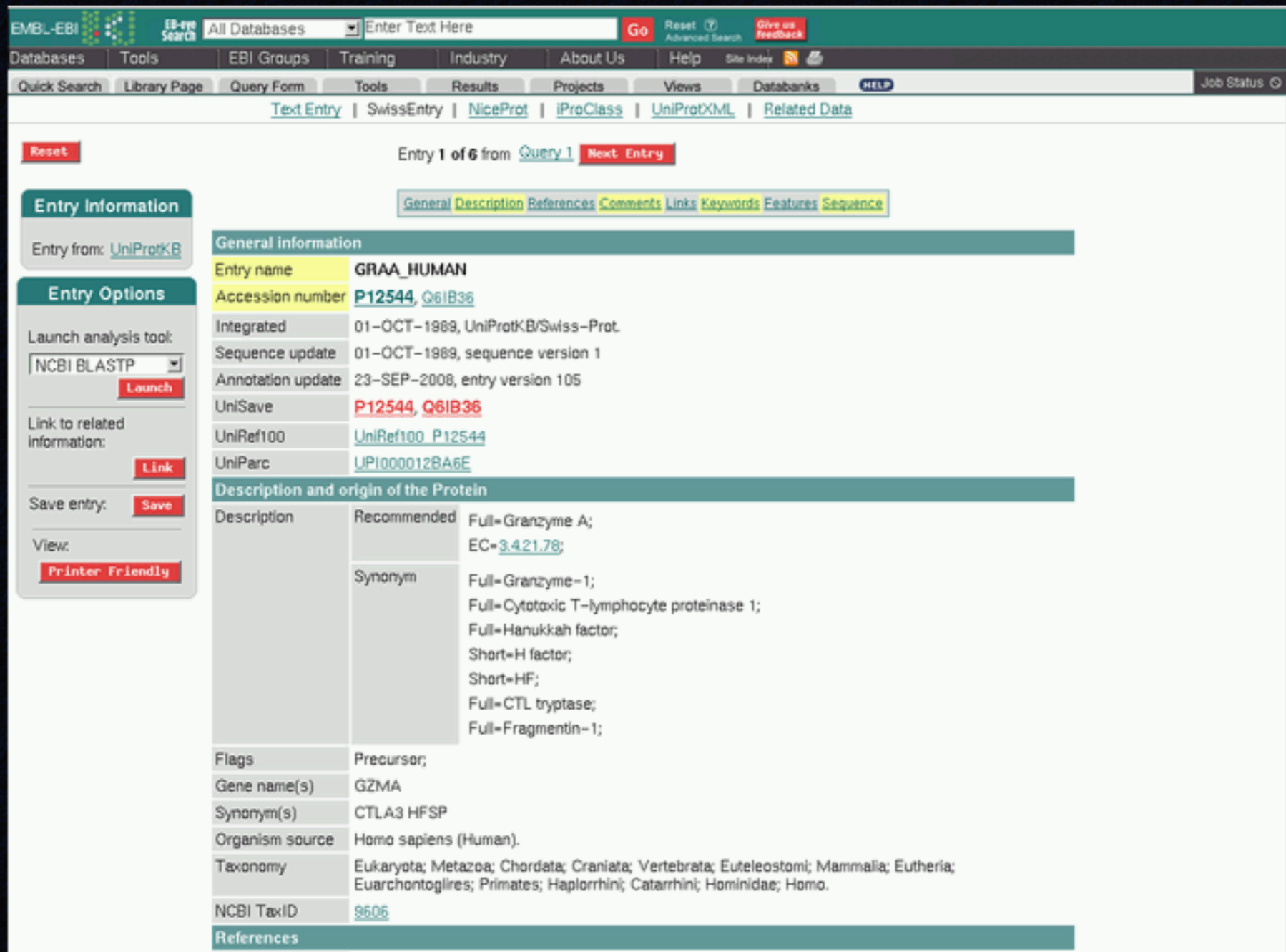
Below the general information, there is a section for "Description and origin of the Protein" with the following details:

Description and origin of the Protein	
Description	Granzyme A precursor (EC 3.4.21.78) (Cytotoxic T-lymphocyte protein factor) (HF) (Granzyme 1) (CTL tryptase) (Fragmentin 1).
Gene name(s)	GZMA OR CTLA3 OR HFSP.
Organism source	Homo sapiens (Human).
Taxonomy	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Primates; Hominidae; Homo.
NCBI TaxID	9606

The page also includes a "References" section at the bottom. On the left side, there are "Entry Options" including a "Launch analysis tool" dropdown menu set to "BlastP", a "Launch" button, a "Link to related information" link, a "Save entry" button, and a "View" dropdown menu with a "Printer friendly" button.

Παραδείγματα : Swiss-Prot

... και όχι μόνο (δείτε active links για βιβλιογραφικές αναφορές).



EMBL-EBI EBI Search All Databases Enter Text Here Go Reset Advanced Search Give us Feedback

Databases Tools EBI Groups Training Industry About Us Help Site Index

Quick Search Library Page Query Form Tools Results Projects Views Databanks HELP Job Status

[Text Entry](#) | [SwissEntry](#) | [NiceProt](#) | [iProClass](#) | [UniProtXML](#) | [Related Data](#)

[Reset](#) Entry 1 of 6 from [Query 1](#) [Next Entry](#)

[General](#) [Description](#) [References](#) [Comments](#) [Links](#) [Keywords](#) [Features](#) [Sequence](#)

Entry Information

Entry from: [UniProtKB](#)

Entry Options

Launch analysis tool:
NCBI BLASTP [Launch](#)

Link to related information: [Link](#)

Save entry: [Save](#)

View: [Printer Friendly](#)

General information

Entry name	GRAA_HUMAN
Accession number	P12544 , Q6IB36
Integrated	01-OCT-1989, UniProtKB/Swiss-Prot
Sequence update	01-OCT-1989, sequence version 1
Annotation update	23-SEP-2008, entry version 105
UniSave	P12544 , Q6IB36
UniRef100	UniRef100_P12544
UniParc	UPI000012BA6E

Description and origin of the Protein

Description	Recommended	Full=Granzyme A; EC= 3.4.21.78 ;
	Synonym	Full=Granzyme-1; Full=Cytotoxic T-lymphocyte proteinase 1; Full=Hanukkah factor; Short=H factor; Short=HF; Full=CTL tryptase; Full=Fragmentin-1;

Flags: Precursor;

Gene name(s): GZMA

Synonym(s): CTLA3 HFSP

Organism source: Homo sapiens (Human).

Taxonomy: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

NCBI TaxID: [9606](#)

References

Παραδείγματα καταχωρήσεων

Prosite

Η Prosite είναι μια βάση πρωτεϊνικών μοτίβων. Πρόκειται για μια δευτερογενή βάση δεδομένων, πρωτογενής πηγή της οποίας είναι η Swiss-Prot. Είναι μια flat-file βάση, με την ιδιαιτερότητα ότι σε κάθε καταχώρηση αντιστοιχούν δύο αρχεία. Το πρώτο αρχείο περιέχει μια αναλυτική περιγραφή του μοτίβου, μαζί με λεπτομέρειες για τη βιολογική του σημασία και σχετικές βιβλιογραφικές αναφορές. Το δεύτερο αρχείο περιέχει το καθ'αυτό μοτίβο.

Παραδείγματα : Prosite (doc)

```
{PDOC00211}
```

```
{PS00238; OPSIN}
```

```
{BEGIN}
```

```
*****
```

```
* Visual pigments (opsins) retinal binding site *
```

```
*****
```

Visual pigments [1,2] are the light-absorbing molecules that mediate vision. They consist of an apoprotein, opsin, covalently linked to the chromophore cis-retinal. Vision is effected through the absorption of a photon by cis-retinal which is isomerized to trans-retinal. This isomerization leads to a change of conformation of the protein. Opsins are integral membrane proteins with seven transmembrane regions that belong to family 1 of G-protein coupled receptors (see <PDOC00210>).

In vertebrates four different pigments are generally found. Rod cells, which mediate vision in dim light, contain the pigment rhodopsin. Cone cells, which function in bright light, are responsible for color vision and contain three or more color pigments (for example, in mammals: red, blue and green).

In *Drosophila*, the eye is composed of 800 facets or ommatidia. Each ommatidium contains eight photoreceptor cells (R1-R8): the R1 to R6 cells are outer cells, R7 and R8 inner cells. Each of the three types of cells (R1-R6, R7 and R8) expresses a specific opsin.

Παραδείγματα : Prosite (doc)

Proteins evolutionary related to opsins include:

- Squid retinochrome, also known as retinal photoisomerase, which converts various isomers of retinal into 11-cis retinal.
- Mammalian opsin 3 (Encephalopsin) that may play a role in encephalic photoreception.
- Mammalian opsin 4 (Melanopsin) that may mediate regulation of circadian rhythms and acute suppression of pineal melatonin.
- Mammalian retinal pigment epithelium (RPE) RGR [3], a protein that may also act in retinal isomerization.

The attachment site for retinal in the above proteins is a conserved lysine residue in the middle of the seventh transmembrane helix. The pattern we developed includes this residue.

-Consensus pattern: [LIVMFWAC]-[PSGAC]-x(3)-[SAC]-K-[STALIMR]-[GSACPNV]-
[STACP]-x(2)-[DENF]-[AP]-x(2)-[IY]
[K is the retinal binding site]

-Sequences known to belong to this class detected by the pattern: ALL.

-Other sequence(s) detected in Swiss-Prot: 2.

-Last update: December 2001 / Pattern and text revised.

[1] Applebury M.L., Hargrave P.A.
Vision Res. 26:1881-1895(1986).

[2] Fryxell K.J., Meyerowitz E.M.
J. Mol. Evol. 33:367-378(1991).

[3] Shen D., Jiang M., Hao W., Tao L., Salazar M., Fong H.K.W.
Biochemistry 33:13117-13125(1994).

{END}

Παραδείγματα : Prosite (patterns)

```
ID OPSIN; PATTERN.
AC PS00238;
DT APR-1990 (CREATED); DEC-2001 (DATA UPDATE); DEC-2001 (INFO UPDATE).
DE Visual pigments (opsins) retinal binding site.
PA [LIVMFWAC]-[PSGAC]-x(3)-[SAC]-K-[STALIMR]-[GSACPNV]-[STACP]-x(2)-[DENF]-
PA [AP]-x(2)-[IY].
NR /RELEASE=41.25,134803;
NR /TOTAL=193(192); /POSITIVE=189(188); /UNKNOWN=0(0); /FALSE_POS=4(4);
NR /FALSE_NEG=1; /PARTIAL=4;
CC /TAXO-RANGE=??E??; /MAX-REPEAT=1;
CC /SITE=5,retinal;
DR Q9H1Y3, OPN3_HUMAN, T; Q9WUK7, OPN3_MOUSE, T; Q9UHM6, OPN4_HUMAN, T;
DR Q9QXZ9, OPN4_MOUSE, T; P22269, OPS1_CALVI, T; P06002, OPS1_DROME, T;
DR P28678, OPS1_DROPS, T; Q25157, OPS1_HEMSA, T; P35360, OPS1_LIMPO, T;
DR O15973, OPS1_PATYE, T; Q94741, OPS1_SCHGR, T; P08099, OPS2_DROME, T;
.....
DR O14718, OPSX_HUMAN, T; O35214, OPSX_MOUSE, T; P23820, REIS_TODPA, T;
DR P47803, RGR_BOVIN, T; P47804, RGR_HUMAN, T;
DR P17645, OPS3_DROVI, P; O18911, OPSG_ODOVI, P; O18914, OPSR_CANFA, P;
DR O18912, OPSR_HORSE, P;
DR Q9Z2B3, RGR_MOUSE, N;
DR Q9CL24, OADB_PASMU, F; P22056, POLS_ONNVG, F; Q99NF8, RP17_MOUSE, F;
DR P09009, TERM_BPPRD, F;
3D 1BOJ; 1BOK; 1F88; 1HZX; 1JFP; 1KPN; 1KPW; 1KPX; 1LN6;
DO PDOC00211;
//
```


Παραδείγματα : Prosite (patterns)

```
ID    OPSIN; PATTERN.  
AC    PS00238;  
DT    APR-1990 (CREATED); DEC-2001 (DATA UPDATE); DEC-2001 (INFO  
UPDATE) .
```

Χαρακτηριστικό όνομα για την πρωτεϊνική οικογένεια (ID), κωδικός καταχώρησης (AC, accession number), και ημερομηνίες κατάθεσης και τροποποιήσεων.

Παραδείγματα : Prosite (patterns)

DE Visual pigments (opsins) retinal binding site.

Σύντομη περιγραφή της οικογένειας (εκτενής περιγραφή στο doc αρχείο που αντιστοιχεί στο μοτίβο).

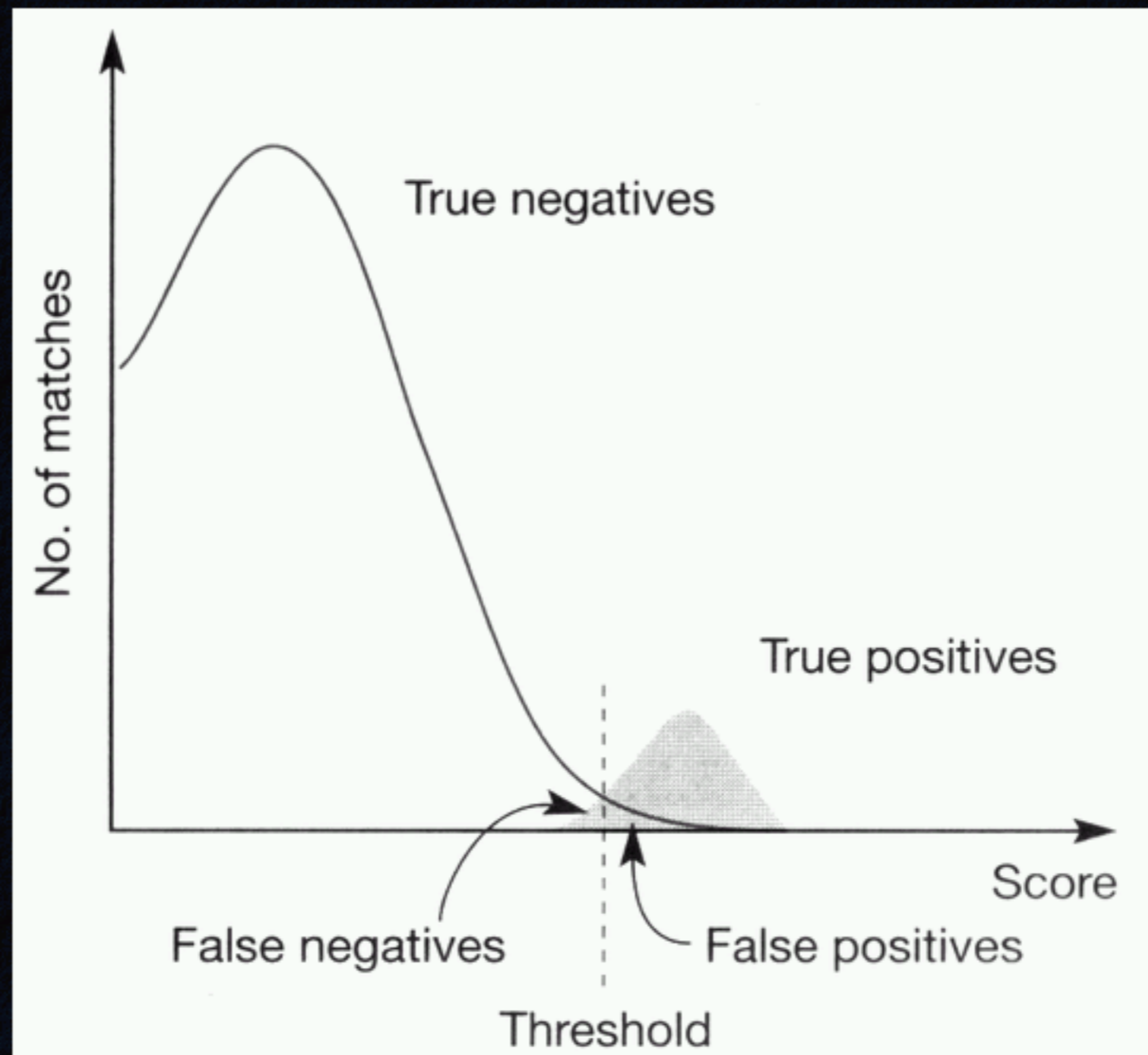
Παραδείγματα : Prosite (patterns)

PA [LIVMFWAC] - [PSGAC] - x(3) - [SAC] - K - [STALIMR] - [GSACPNV] -
PA [STACP] - x(2) - [DENF] - [AP] - x(2) - [IY] .

Το μοτίβο. Τα κεφαλαία γράμματα αντιστοιχούν σε αμινοξέα (ένα γράμμα - ένα αμινοξικό κατάλοιπο) με βάση τον κώδικα του ενός γράμματος. Οι αγκύλες περικλείουν αμινοξέα αποδεκτά για την θέση στην οποία αντιστοιχούν. Τα "x" σημαίνει ότι οποιοδήποτε αμινοξύ μπορεί να βρίσκεται στην εν λόγω θέση [x(N) σημαίνει ότι N διαδοχικά αμινοξέα είναι αόριστα].

Παραδείγματα : Prosite (patterns)

```
NR /RELEASE=41.25,134803;  
NR /TOTAL=193 ...  
NR /FALSE_POS=4 ...  
NR /FALSE_NEG=1 ...
```



Παραδείγματα : Prosite (patterns)

```
CC /TAXO-RANGE=??E??; /MAX-REPEAT=1;
```

```
CC /SITE=5,retinal;
```

Συστηματική πρωτεϊνικής οικογένειας (E για Eucaryotic), μέγιστος αριθμός επαναλήψεων του μοτίβου, πληροφορία για λειτουργικά κέντρα της οικογένειας (στη συγκεκριμένη περίπτωση πρόκειται για μία λυσίνη στην οποία προσδένεται το χρωμόφορο).

Παραδείγματα : Prosite (patterns)

```
DR Q9H1Y3, OPN3_HUMAN,T; Q9WUK7, OPN3_MOUSE,T; Q9UHM6, OPN4_HUMAN,T;  
DR Q9QXZ9, OPN4_MOUSE,T; P22269, OPS1_CALVI,T; P06002, OPS1_DROME,T;
```

Κωδικοί καταχώρησης της Swiss-Prot (accession, ID),
και αποτέλεσμα εφαρμογής του μοτίβου :

T : αληθώς θετικό (True positive)

F : ψευδώς θετικό (False positive)

N : ψευδώς αρνητικό (false Negative)

P : πιθανώς θετικό ('Possible' match), κυρίως
για την περίπτωση ελλειπών αλληλουχιών.

Παραδείγματα : Prosite (patterns)

```
3D 1BOJ; 1BOK; 1F88; 1HZX; 1JFP; 1KPN; 1KPW; 1KPX; 1LN6;  
DO PDOC00211;
```

Αναφορές στην PDB (βάση δομών) για μέλη της οικογένειας, και κωδικός καταχώρησης του αρχείου που περιέχει την αναλυτική περιγραφή της οικογένειας (DO για document).

Παραδείγματα : Prosite

Ενδιάμεσοι εξυπηρετητές όπως το SRS ή Entrez, διαμορφώνουν τις καταχωρήσεις ώστε να είναι αναγνώσιμες (σε απλά Αγγλικά) ...

File Edit View Go Bookmarks Tools Window Help

file://home/glykos/teach/bioinf2/prosite-niceview.html

Home Radio Bookmarks Members WebMail Connections BizJournal SmartUpdate Mktplace

NiceSite View of PROSITE: PDOC00...

ExpPASy Home page Site Map Search ExpPASy Contact us PROSITE

Hosted by NCSC US Mirror sites: Bolivia Canada China Korea Switzerland Taiwan

Search PROSITE for Go Clear

NiceSite View of PROSITE: [PDOC00211](#) (documentation)

Visual pigments (opsins) retinal binding site

PROSITE cross-reference(s)

PS00238; OPSIN	Retrieve an alignment of Swiss-Prot true positive hits: [Clustal format, color, condensed view] [Clustal format, color] [Clustal format, plain text] [Fasta format]
--------------------------------	--

Documentation

Visual pigments [1,2] are the light-absorbing molecules that mediate vision. They consist of an apoprotein, opsin, covalently linked to the chromophore cis-retinal. Vision is effected through the absorption of a photon by cis-retinal which is isomerized to trans-retinal. This isomerization leads to a change of conformation of the protein. Opsins are integral membrane proteins with seven transmembrane regions that belong to family 1 of G-protein coupled receptors (see <[PDOC00210](#)>).

Παραδείγματα : Prosite

... και όχι μόνο ...

ExPASy Home page Site Map Search ExPASy Contact us PROSITE Swiss-Prot

Search for

Home ScanProsite ProRule Documents Downloads Links Funding

prosite Sequence logo for Visual pigments (opsins) retinal binding site (PS00238)

PS00238 / #=223

Position	Conserved Residues
1	L, P, W
2	P, G, A, S
3	A, F, Y, S
4	F, Y, S
5	F, S
6	F, K, S
7	K
8	S, A, T, A
9	S, A, N, T
10	A, S, T, V
11	A, V, I
12	Y, I
13	Y, P, D
14	P
15	V, M, V
16	V
17	Y

Number of UniProtKB/Swiss-Prot true positive hits used to build the logo:223 [\[More details\]](#)

ExPASy Home page Site Map Search ExPASy Contact us PROSITE Swiss-Prot

Hosted by CBR Canada | Mirror sites: [Australia](#) | [Brazil](#) | [China](#) | [Korea](#) | [Switzerland](#)

Βάσεις δομικής βιολογίας

Στη δομική βιολογία τα πράγματα είναι απλούστερα : υπάρχει μία μόνο πρωτοταγής βάση δομών επιπέδου ατομικών μοντέλων, η PDB (Protein Data Bank). Τα προϊόντα της ανάλυσης και ταξινόμησης αυτών των δομών εμπεριέχονται σε μία πληθώρα δευτεροταγών βάσεων.

Protein Data Bank (PDB)

Είναι η μοναδική πρωτοταγής βάση δεδομένων της δομικής βιολογίας. Είναι φαινομενικά μια flat-file βάση, αν και στην πραγματικότητα είναι σχεσιακή. Στις επόμενες διαφάνειες θα παρουσιαστεί η μορφή μιας καταχώρησης της.

Η διεύθυνση της βάσης είναι :

<http://www.pdb.org/>

PDB : παράδειγμα

```
HEADER      TRANSCRIPTION REGULATION          16-JAN-99      1B6Q
TITLE      ALANINE 31 PROLINE MUTANT OF ROP PROTEIN
COMPND     MOL_ID: 1;
COMPND     2 MOLECULE: ROP;
COMPND     3 CHAIN: NULL;
COMPND     4 SYNONYM: ROM;
COMPND     5 ENGINEERED: YES;
COMPND     6 MUTATION: A31P;
COMPND     7 BIOLOGICAL_UNIT: HOMODIMER
SOURCE     MOL_ID: 1;
SOURCE     2 ORGANISM_SCIENTIFIC: ESCHERICHIA COLI;
SOURCE     3 CELLULAR_LOCATION: CYTOPLASM;
SOURCE     4 PLASMID: COLE1;
SOURCE     5 GENE: ROP;
SOURCE     6 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE     7 EXPRESSION_SYSTEM_STRAIN: 71/72 (71/18 PLUS PCI857);
SOURCE     8 EXPRESSION_SYSTEM_PLASMID: PEX43
```


PDB : παράδειγμα

```
KEYWDS      TRANSCRIPTION REGULATION
EXPDTA      X-RAY DIFFRACTION
AUTHOR      N. GLYKOS , G. CESARENI , M. KOKKINIDIS
REVDAT      1   09-JUL-99 1B6Q   0
JRNL        AUTH      N. M. GLYKOS , G. CESARENI , M. KOKKINIDIS
JRNL        TITL      PROTEIN PLASTICITY TO THE EXTREME: CHANGING THE
JRNL        TITL 2    TOPOLOGY OF A 4-ALPHA-HELICAL BUNDLE WITH A SINGLE
JRNL        TITL 3    AMINO-ACID SUBSTITUTION
JRNL        REF       STRUCTURE (LONDON)                V.   7   597 1999
JRNL        REFN      ASTM STRUE6  UK ISSN 0969-2126                2005
REMARK      1
REMARK      1 REFERENCE 1
REMARK      1 TITL     MEANINGFUL REFINEMENT OF POLY-ALANINE MODELS USING
REMARK      1 TITL 2   RIGID-BODY SIMULATED ANNEALING : APPLICATION TO
REMARK      1 TITL 3   THE STRUCTURE DETERMINATION OF THE A31P ROP MUTANT
REMARK      1 REF      ACTA CRYSTALLOGR., SECT.D          V.  55  1301 1999
REMARK      1 REFN     ASTM ABCRE6  DK ISSN 0907-4449                0766
```


PDB : παράδειγμα

```
REMARK 2 RESOLUTION. 1.80 ANGSTROMS.
REMARK 3
REMARK 3 REFINEMENT.
REMARK 3 PROGRAM : X-PLOR 3.851
REMARK 3 AUTHORS : BRUNGER
REMARK 3
REMARK 3 DATA USED IN REFINEMENT.
REMARK 3 RESOLUTION RANGE HIGH (ANGSTROMS) : 1.80
REMARK 3 RESOLUTION RANGE LOW (ANGSTROMS) : 40.825
REMARK 3 DATA CUTOFF (SIGMA (F)) : 0.0
REMARK 3 DATA CUTOFF HIGH (ABS (F)) : NULL
REMARK 3 DATA CUTOFF LOW (ABS (F)) : NULL
REMARK 3 COMPLETENESS (WORKING+TEST) (%) : 99.9
REMARK 3 NUMBER OF REFLECTIONS : 5103
REMARK 3
REMARK 3 FIT TO DATA USED IN REFINEMENT.
REMARK 3 CROSS-VALIDATION METHOD : THROUGHOUT
REMARK 3 FREE R VALUE TEST SET SELECTION : RANDOM
REMARK 3 R VALUE (WORKING SET) : 0.189
REMARK 3 FREE R VALUE : 0.240
REMARK 3 FREE R VALUE TEST SET SIZE (%) : 5.0
REMARK 3 FREE R VALUE TEST SET COUNT : 286
```


PDB : παράδειγμα

```
REMARK 3 FIT IN THE HIGHEST RESOLUTION BIN.
REMARK 3 TOTAL NUMBER OF BINS USED : 8
REMARK 3 BIN RESOLUTION RANGE HIGH (A) : 1.80
REMARK 3 BIN RESOLUTION RANGE LOW (A) : 1.88
REMARK 3 BIN COMPLETENESS (WORKING+TEST) (%) : 99.9
REMARK 3 REFLECTIONS IN BIN (WORKING SET) : 588
REMARK 3 BIN R VALUE (WORKING SET) : 0.332
REMARK 3 BIN FREE R VALUE : 0.400
REMARK 3 BIN FREE R VALUE TEST SET SIZE (%) : 5.0
REMARK 3 BIN FREE R VALUE TEST SET COUNT : 31
REMARK 3 ESTIMATED ERROR OF BIN FREE R VALUE : NULL
REMARK 3
REMARK 3 NUMBER OF NON-HYDROGEN ATOMS USED IN REFINEMENT.
REMARK 3 PROTEIN ATOMS : 489
REMARK 3 NUCLEIC ACID ATOMS : 0
REMARK 3 HETEROGEN ATOMS : 0
REMARK 3 SOLVENT ATOMS : 56
REMARK 3
```


PDB : παράδειγμα

```
REMARK      3  B VALUES.
REMARK      3  FROM WILSON PLOT              (A**2) : 28.9
REMARK      3  MEAN B VALUE              (OVERALL, A**2) : 26.0
REMARK      3  OVERALL ANISOTROPIC B VALUE.
REMARK      3
REMARK      3  ESTIMATED COORDINATE ERROR.
REMARK      3  ESD FROM LUZZATI PLOT        (A) : 0.23
REMARK      3  ESD FROM SIGMAA            (A) : NULL
REMARK      3  LOW RESOLUTION CUTOFF      (A) : 40.825
REMARK      3
REMARK      3  CROSS-VALIDATED ESTIMATED COORDINATE ERROR.
REMARK      3  ESD FROM C-V LUZZATI PLOT    (A) : NULL
REMARK      3  ESD FROM C-V SIGMAA          (A) : NULL
REMARK      3
REMARK      3  RMS DEVIATIONS FROM IDEAL VALUES.
REMARK      3  BOND LENGTHS                    (A) : 0.007
REMARK      3  BOND ANGLES                    (DEGREES) : 0.909
REMARK      3  DIHEDRAL ANGLES                (DEGREES) : NULL
REMARK      3  IMPROPER ANGLES                  (DEGREES) : NULL
```


PDB : παράδειγμα

```
REMARK 200 EXPERIMENTAL DETAILS
REMARK 200 EXPERIMENT TYPE           : X-RAY DIFFRACTION
REMARK 200 DATE OF DATA COLLECTION  : NULL
REMARK 200 TEMPERATURE                (KELVIN) : 293
REMARK 200 PH                         : 4.3
REMARK 200 NUMBER OF CRYSTALS USED    : 1
REMARK 200
REMARK 200 SYNCHROTRON                (Y/N)   : N
REMARK 200 RADIATION SOURCE           : NULL
REMARK 200 BEAMLINE                   : NULL
REMARK 200 X-RAY GENERATOR MODEL      : SEALED TUBE
REMARK 200 MONOCHROMATIC OR LAUE      (M/L)   : M
REMARK 200 WAVELENGTH OR RANGE        (A)     : 1.5418
REMARK 200 MONOCHROMATOR              : GRAPHITE MONOCHROMATOR
REMARK 200 OPTICS                     : MONOCHROMATOR
REMARK 200
```


PDB : παράδειγμα

```
REMARK 200 DETECTOR TYPE : DIFFRACTOMETER
REMARK 200 DETECTOR MANUFACTURER : ENRAF-NONIUS
REMARK 200 INTENSITY-INTEGRATION SOFTWARE : CAD4 SOFTWARE
REMARK 200 DATA SCALING SOFTWARE : CAD4 SOFTWARE
REMARK 200
REMARK 200 NUMBER OF UNIQUE REFLECTIONS : 5103
REMARK 200 RESOLUTION RANGE HIGH (A) : 1.80
REMARK 200 RESOLUTION RANGE LOW (A) : 40.825
REMARK 200 REJECTION CRITERIA (SIGMA(I)) : NONE
REMARK 200
```


PDB : παράδειγμα

```
REMARK 200 OVERALL.  
REMARK 200 COMPLETENESS FOR RANGE (%) : 99.9  
REMARK 200 DATA REDUNDANCY : 1.0  
REMARK 200 R MERGE (I) : NULL  
REMARK 200 R SYM (I) : NONE  
REMARK 200 <I/SIGMA(I)> FOR THE DATA SET : 12.4  
REMARK 200  
REMARK 200 IN THE HIGHEST RESOLUTION SHELL.  
REMARK 200 HIGHEST RESOLUTION SHELL, RANGE HIGH (A) : 1.80  
REMARK 200 HIGHEST RESOLUTION SHELL, RANGE LOW (A) : 1.86  
REMARK 200 COMPLETENESS FOR SHELL (%) : 99.6  
REMARK 200 DATA REDUNDANCY IN SHELL : 1.0  
REMARK 200 R MERGE FOR SHELL (I) : NULL  
REMARK 200 R SYM FOR SHELL (I) : NONE  
REMARK 200 <I/SIGMA(I)> FOR SHELL : 2.48  
REMARK 200
```


PDB : παράδειγμα

```
REMARK 350 GENERATING THE BIOMOLECULE
REMARK 350 COORDINATES FOR A COMPLETE MULTIMER REPRESENTING THE KNOWN
REMARK 350 BIOLOGICALLY SIGNIFICANT OLIGOMERIZATION STATE OF THE
REMARK 350 MOLECULE CAN BE GENERATED BY APPLYING BIOMT TRANSFORMATIONS
REMARK 350 GIVEN BELOW. BOTH NON-CRYSTALLOGRAPHIC AND
REMARK 350 CRYSTALLOGRAPHIC OPERATIONS ARE GIVEN.
REMARK 350
REMARK 350 APPLY THE FOLLOWING TO CHAINS: NULL
REMARK 350 BIOMT1 1 -1.000000 0.000000 0.000000 0.000000
REMARK 350 BIOMT2 1 0.000000 1.000000 0.000000 0.000000
REMARK 350 BIOMT3 1 0.000000 0.000000 -1.000000 40.70000
```


PDB : παράδειγμα

```
REMARK 999 SEQUENCE
REMARK 999 1B6Q          SWS          P03051          57 -          63 NOT IN ATOMS LIST
DBREF  1B6Q           1          56  SWS          P03051  ROP_ECOLI           1          56
SEQADV 1B6Q  PRO           31  SWS  P03051  ALA           31 ENGINEERED MUTATION
SEQRES  1           63  MET THR LYS GLN GLU LYS THR ALA LEU ASN MET ALA ARG
SEQRES  2           63  PHE ILE ARG SER GLN THR LEU THR LEU LEU GLU LYS LEU
SEQRES  3           63  ASN GLU LEU ASP PRO ASP GLU GLN ALA ASP ILE CYS GLU
SEQRES  4           63  SER LEU HIS ASP HIS ALA ASP GLU LEU TYR ARG SER CYS
SEQRES  5           63  LEU ALA ARG PHE GLY ASP ASP GLY GLU ASN LEU
FORMUL  2  HOH      *56(H2 O1)
HELIX   1     1  LYS           3  GLU           28  1                               26
HELIX   2     2  PRO           31  ALA           54  1                               24
CRYST1  30.400  42.100  81.400  90.00  90.00  90.00  C 2 2 21  8
ORIGX1  1.000000  0.000000  0.000000  0.000000
ORIGX2  0.000000  1.000000  0.000000  0.000000
ORIGX3  0.000000  0.000000  1.000000  0.000000
SCALE1  0.032895  0.000000  0.000000  0.000000
SCALE2  0.000000  0.023753  0.000000  0.000000
SCALE3  0.000000  0.000000  0.012285  0.000000
```


PDB : παράδειγμα

ATOM	1	N	MET	1	-2.053	13.510	-6.199	1.00	47.14
ATOM	2	CA	MET	1	-1.894	13.110	-4.767	1.00	49.45
ATOM	3	C	MET	1	-0.688	12.186	-4.582	1.00	47.94
ATOM	4	O	MET	1	-0.774	10.982	-4.841	1.00	51.90
ATOM	5	CB	MET	1	-3.163	12.402	-4.276	1.00	51.84
ATOM	6	CG	MET	1	-3.059	11.814	-2.871	1.00	57.35
ATOM	7	SD	MET	1	-4.121	12.669	-1.683	1.00	62.43
ATOM	8	CE	MET	1	-2.938	13.041	-0.373	1.00	61.79
ATOM	9	N	THR	2	0.434	12.748	-4.134	1.00	42.11
ATOM	10	CA	THR	2	1.636	11.954	-3.914	1.00	34.48
ATOM	11	C	THR	2	1.551	11.168	-2.615	1.00	32.97
ATOM	12	O	THR	2	0.726	11.447	-1.749	1.00	32.39

Άλλες βάσεις : PIR

Η PIR (Protein Information Resource) είναι μία πρωτοταγής βάση πρωτεϊνικών αλληλουχιών η οποία ανάλογα με με την ποιότητα των δεδομένων και την πληρότητα σχολιασμού τους (annotation), χωρίζεται σε τέσσερα τμήματα, PIR1 μέχρι και PIR4. Η PIR1 περιέχει πλήρως ταξινομημένες και σχολιασμένες καταχωρήσεις. Η PIR2 περιέχει καταχωρήσεις που βρίσκονται σε προκαταρκτικό στάδιο και δεν έχουν εξεταστεί ενδελεχώς. Η PIR3 περιέχει καταχωρήσεις που δεν έχουν εξεταστεί καθόλου. Η PIR4 περιέχει υποθετικές αλληλουχίες που προκύπτουν από τη (θεωρητική) μετάφραση DNA αλληλουχιών.

Άλλες βάσεις : TrEMBL

Η TrEMBL (Translated EMBL) είναι μία πρωτοταγής βάση πρωτεϊνικών αλληλουχιών η οποία περιέχει μεταφράσεις όλων των κωδικοποιουσών αλληλουχιών (CDS, coding sequences) της EMBL. Αποτελείται από δύο τμήματα : Το SP-TrEMBL (Swiss-Prot TrEMBL) περιέχει αλληλουχίες οι οποίες είναι υπό ένταξη στην Swiss-Prot. Το REM-TrEMBL περιέχει αλληλουχίες οι οποίες δεν πρόκειται να ενταχθεί στη Swiss-Prot και περιλαμβάνει ανοσοσφαιρίνες, T-cell receptors, αλληλουχίες μικρότερες από 8 αμινοξέα, συνθετικές αλληλουχίες, αλληλουχίες που καλύπτονται από ευρεσιτεχνίες, κοκ.

Άλλες βάσεις : NR

Η NR (Non-Redundant protein dataBase) είναι μια σύνθετη βάση πρωτεϊνικών αλληλουχιών η οποία προκύπτει από το συνδυασμό των PDB, Swiss-Prot, PIR, GenPept (η οποία προκύπτει από την αυτόματη μετάφραση των CDS της GenBank), SPupdate (εβδομαδιαίες προσθήκες της Swiss-Prot) και GenPeptupdate (ημερήσιες προσθήκες της GenBank). Είναι από τις πλέον πλήρεις και ενημερωμένες βάσεις, αλλά περιέχει πληθώρα επαναλαμβανόμενων καταχωρήσεων.

Άλλες βάσεις :

UniProt: Swiss-Prot + TrEMBL

Είναι μια σύνθετη βάση πρωτεϊνικών αλληλουχιών. Προκύπτει από το συνδυασμό της Swiss-Prot με την TrEMBL. Θεωρείται σαν μία από τις πλέον πλήρεις και μη επαναλαμβανόμενες (non-redundant) βάσεις.

Αναζήτηση πληροφοριών

Το βασικό πρόβλημα είναι η εύρεση ενός ενιαίου τρόπου αναζήτησης πληροφοριών μέσα στο πλήθος των διαθέσιμων βάσεων δεδομένων. Το πρόβλημα μεγεθύνεται από το ότι

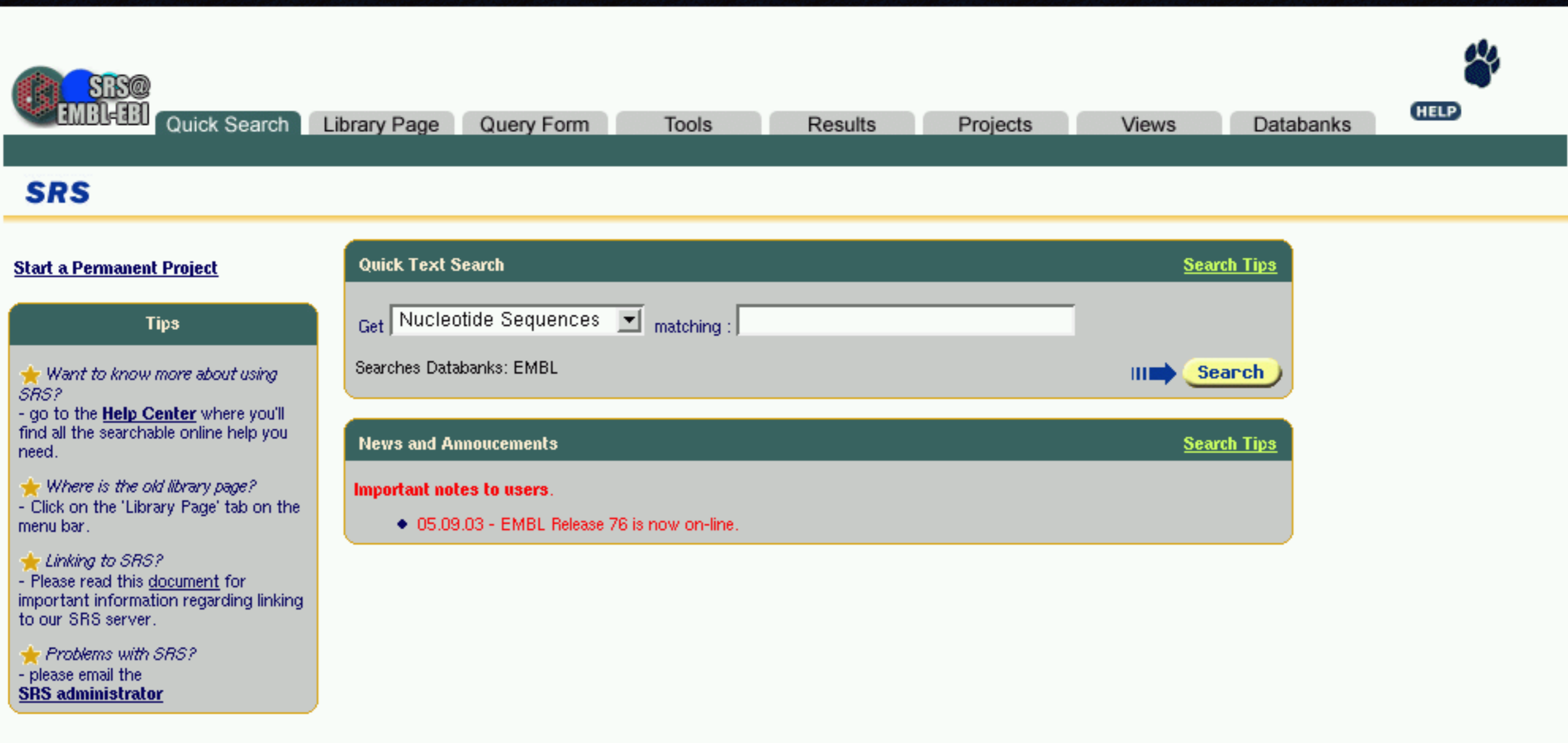
- Σχεδόν κάθε βάση έχει το δικό της τρόπο οργάνωσης των δεδομένων για κάθε καταχώρηση (entry format).
- Το συντακτικό αναζήτησης πληροφοριών (query language) επίσης διαφέρει από βάση σε βάση.

Αναζήτηση πληροφοριών : SRS

Το SRS (Sequence Retrieval System) είναι η Ευρωπαϊκή απόπειρα (μέσω του EMBnet) να δημιουργηθεί ένας ηλεκτρονικός μεσολαβητής ανάμεσα στην πληθώρα των βάσεων δεδομένων και τον τελικό χρήστη. Το SRS επιτρέπει την αρχειοθέτηση οποιασδήποτε flat-file βάσης και την διασύνδεση της με τις υπόλοιπες βάσεις. Ο τελικός χρήστης μπορεί να αναζητήσει με έναν ενιαίο τρόπο πληροφορία που μπορεί να προέρχεται από οποιαδήποτε από τις δεκάδες βάσεις που υποστηρίζει το SRS χωρίς να χρειάζεται να θυμάται λεπτομερώς τις δομές δεδομένων που χρησιμοποιούν.

Αναζήτηση πληροφοριών : SRS

User interface



The screenshot displays the SRS (Sequence Retrieval System) user interface. At the top left is the SRS@EMBL-EBI logo. A horizontal navigation bar contains the following tabs: Quick Search, Library Page, Query Form, Tools, Results, Projects, Views, and Databanks. A HELP button with a paw print icon is located at the top right. Below the navigation bar is a large 'SRS' header. On the left side, there is a 'Start a Permanent Project' link and a 'Tips' section. The 'Tips' section contains three items: 'Want to know more about using SRS?' with a link to the Help Center, 'Where is the old library page?' with a link to the Library Page, and 'Linking to SRS?' with a link to a document. The main content area features a 'Quick Text Search' box with a dropdown menu set to 'Nucleotide Sequences', a 'matching:' input field, and a 'Search' button. Below the search box is a 'News and Announcements' section with a 'Search Tips' link and a red 'Important notes to users' section containing a note about EMBL Release 76.

SRS@EMBL-EBI Quick Search Library Page Query Form Tools Results Projects Views Databanks **HELP**

SRS

[Start a Permanent Project](#)

Tips

- ★ *Want to know more about using SRS?*
- go to the [Help Center](#) where you'll find all the searchable online help you need.
- ★ *Where is the old library page?*
- Click on the 'Library Page' tab on the menu bar.
- ★ *Linking to SRS?*
- Please read this [document](#) for important information regarding linking to our SRS server.
- ★ *Problems with SRS?*
- please email the [SRS administrator](#)

Quick Text Search [Search Tips](#)

Get **Nucleotide Sequences** matching :

Searches Databanks: EMBL

News and Announcements [Search Tips](#)

Important notes to users.

- ◆ 05.09.03 - EMBL Release 76 is now on-line.

Αναζήτηση πληροφοριών : NCBI

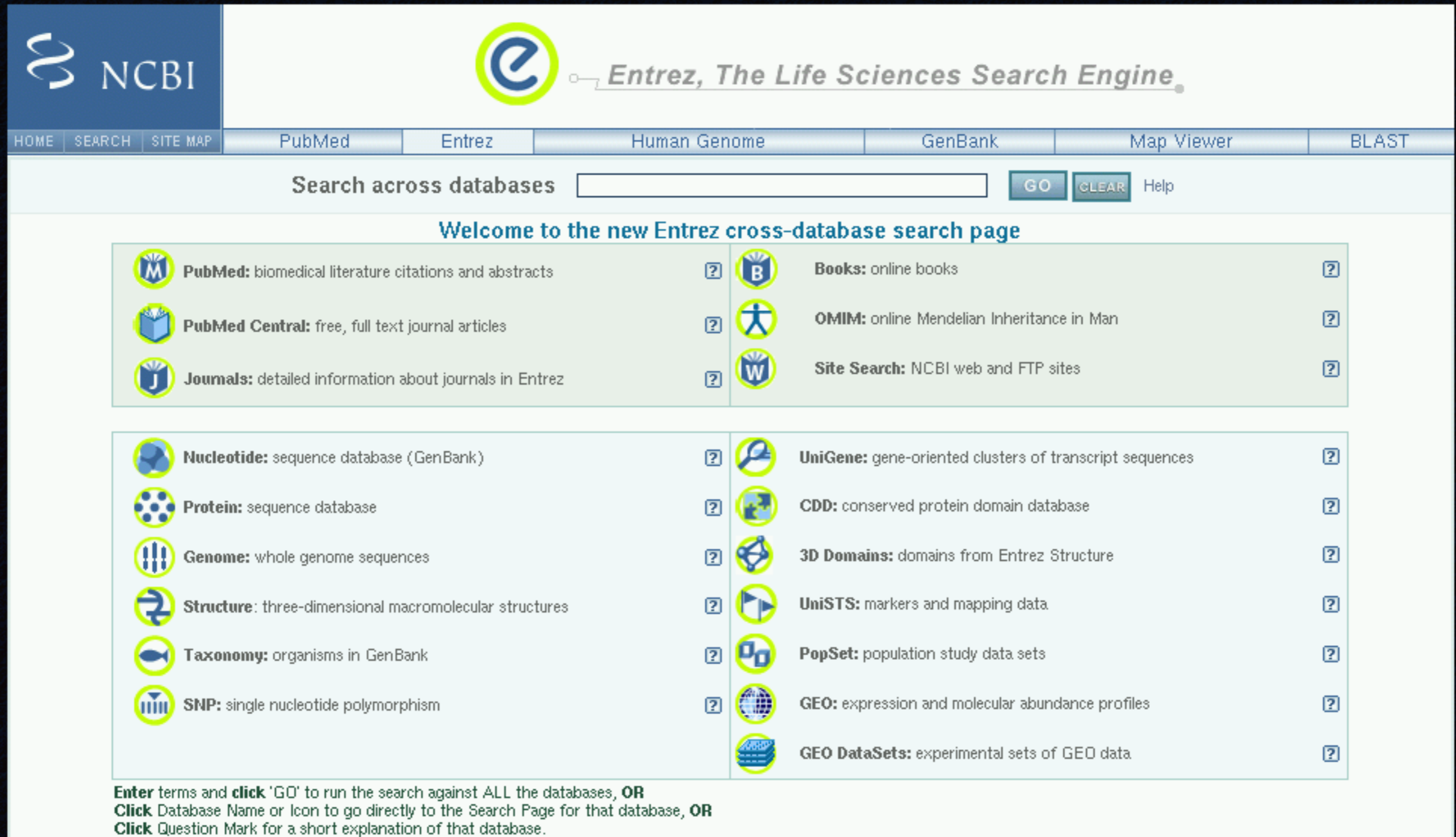
Το NCBI (National Center for Biotechnology Information) έχει δημιουργήσει το Entrez, ένα δικτυακό εργαλείο αναζήτησης και ανάκτησης πληροφοριών από βάσεις δεδομένων βιολογικού περιεχομένου. Όπως και το SRS, το Entrez επιτρέπει την με ενιαίο τρόπο αναζήτηση σε αλληλουχίες DNA (GenBank, EMBL, DDBJ), πρωτεϊνικές αλληλουχίες (Swiss-Prot, PIR, PDB, μεταφρασμένες DNA αλληλουχίες), δεδομένα χαρτογράφησης χρωμοσωμάτων και γονιδιωμάτων, δομές μακρομορίων (από την PDB), και κυριότερα, βιβλιογραφικές αναφορές από την PubMed.

Αναζήτηση πληροφοριών : NCBI

Ιδιαίτερα σε ότι αφορά τις βιβλιογραφικές αναζητήσεις, το Entrez επιτρέπει συγγενή άρθρα που ανήκουν σε διαφορετικές βάσεις να συνδεθούν μεταξύ τους ακόμα και σε περιπτώσεις που δεν υπάρχουν άμεσες αναφορές από το ένα άρθρο στο άλλο.

Αναζήτηση πληροφοριών : NCBI

User interface



The screenshot displays the NCBI Entrez search engine interface. At the top left is the NCBI logo. To its right is the Entrez logo and the text "Entrez, The Life Sciences Search Engine". Below this is a navigation menu with tabs for HOME, SEARCH, SITE MAP, PubMed, Entrez, Human Genome, GenBank, Map Viewer, and BLAST. A search bar is located below the menu, with the text "Search across databases" and buttons for GO, CLEAR, and Help. The main content area is titled "Welcome to the new Entrez cross-database search page" and contains a grid of database links. Each link includes an icon, the database name, a brief description, and a question mark icon for help. The databases listed are: PubMed, PubMed Central, Journals, Books, OMIM, Site Search, Nucleotide, Protein, Genome, Structure, Taxonomy, SNP, UniGene, CDD, 3D Domains, UniSTS, PopSet, GEO, and GEO DataSets.







































NCBI

Entrez, The Life Sciences Search Engine

HOME SEARCH SITE MAP PubMed Entrez Human Genome GenBank Map Viewer BLAST

Search across databases GO CLEAR Help

Welcome to the new Entrez cross-database search page

 PubMed: biomedical literature citations and abstracts 	 Books: online books 
 PubMed Central: free, full text journal articles 	 OMIM: online Mendelian Inheritance in Man 
 Journals: detailed information about journals in Entrez 	 Site Search: NCBI web and FTP sites 
 Nucleotide: sequence database (GenBank) 	 UniGene: gene-oriented clusters of transcript sequences 
 Protein: sequence database 	 CDD: conserved protein domain database 
 Genome: whole genome sequences 	 3D Domains: domains from Entrez Structure 
 Structure: three-dimensional macromolecular structures 	 UniSTS: markers and mapping data 
 Taxonomy: organisms in GenBank 	 PopSet: population study data sets 
 SNP: single nucleotide polymorphism 	 GEO: expression and molecular abundance profiles 
	 GEO DataSets: experimental sets of GEO data 

Enter terms and **click** 'GO' to run the search against ALL the databases, **OR**
Click Database Name or Icon to go directly to the Search Page for that database, **OR**
Click Question Mark for a short explanation of that database.