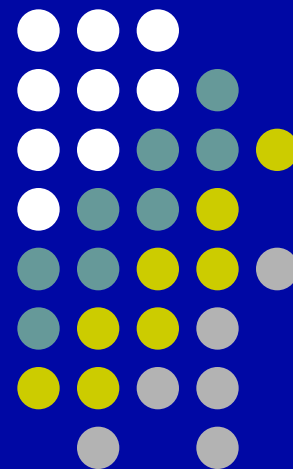


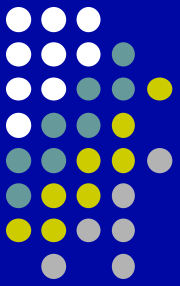
Στοιχεία Βιοπληροφορικής



Ιωάννης Μακρής, Γιώργος Ματσόπουλος



Η ζωή βασίζεται σε 3 σημαντικά μόρια



- **DNA**

Περιέχει όλες τις γενετικές πληροφορίες που καθορίζουν τη βιολογική ανάπτυξη όλων των κυτταρικών μορφών ζωής

- **RNA**

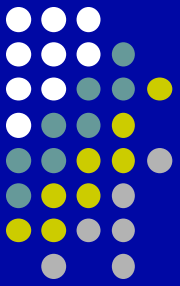
παίζει βασικό ρόλο στη διαδικασία της μετάφρασης του DNA σε πρωτεϊνικά προϊόντα

- **Πρωτεΐνες**

παίζουν σημαντικό ρόλο σε όλες τις βιολογικές διεργασίες



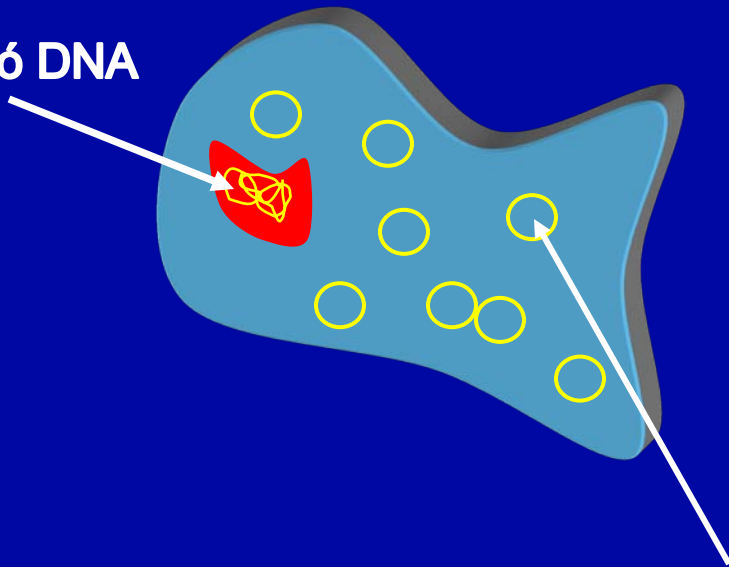
Που βρίσκεται το DNA;



Κύτταρο

Κυτταρικοί τύποι

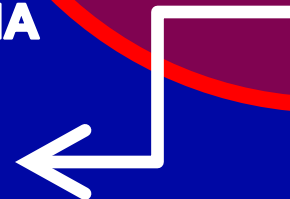
Πυρηνικό DNA



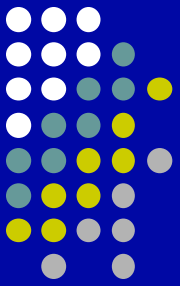
Μιτοχονδριακό DNA

Αίμα
Ρίζες μαλλιών
Σάλιο
Ιδρώτας
Σπέρμα
Διάφοροι ιστοί

Το ίδιο!!!



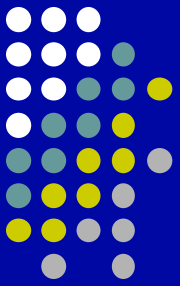
DNA (δεοξυριβονουκλεϊκό οξύ)



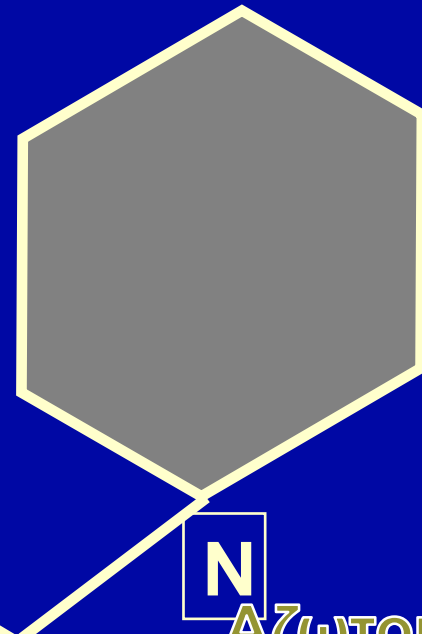
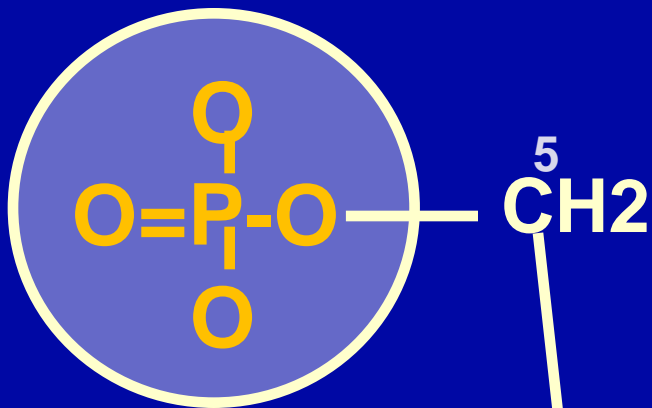
- Νουκλεϊκό οξύ (πολυμερές) που αποτελείται από μεγάλο αριθμό δεοξυριβονουκλεοτιδίων
- Το καθένα περιέχει μια βάση, ένα σάκχαρο και μια φωσφορική ομάδα
- Σχηματίζει διπλή ελικοειδή αλυσίδα
- Οι βάσεις του DNA μεταφέρουν τη γενετική πληροφορία (γονιδίωμα) ενώ το σάκχαρο και η φωσφορική ομάδα έχουν δομικό ρόλο



DNA νουκλεοτίδιο



Φωσφορική ομάδα

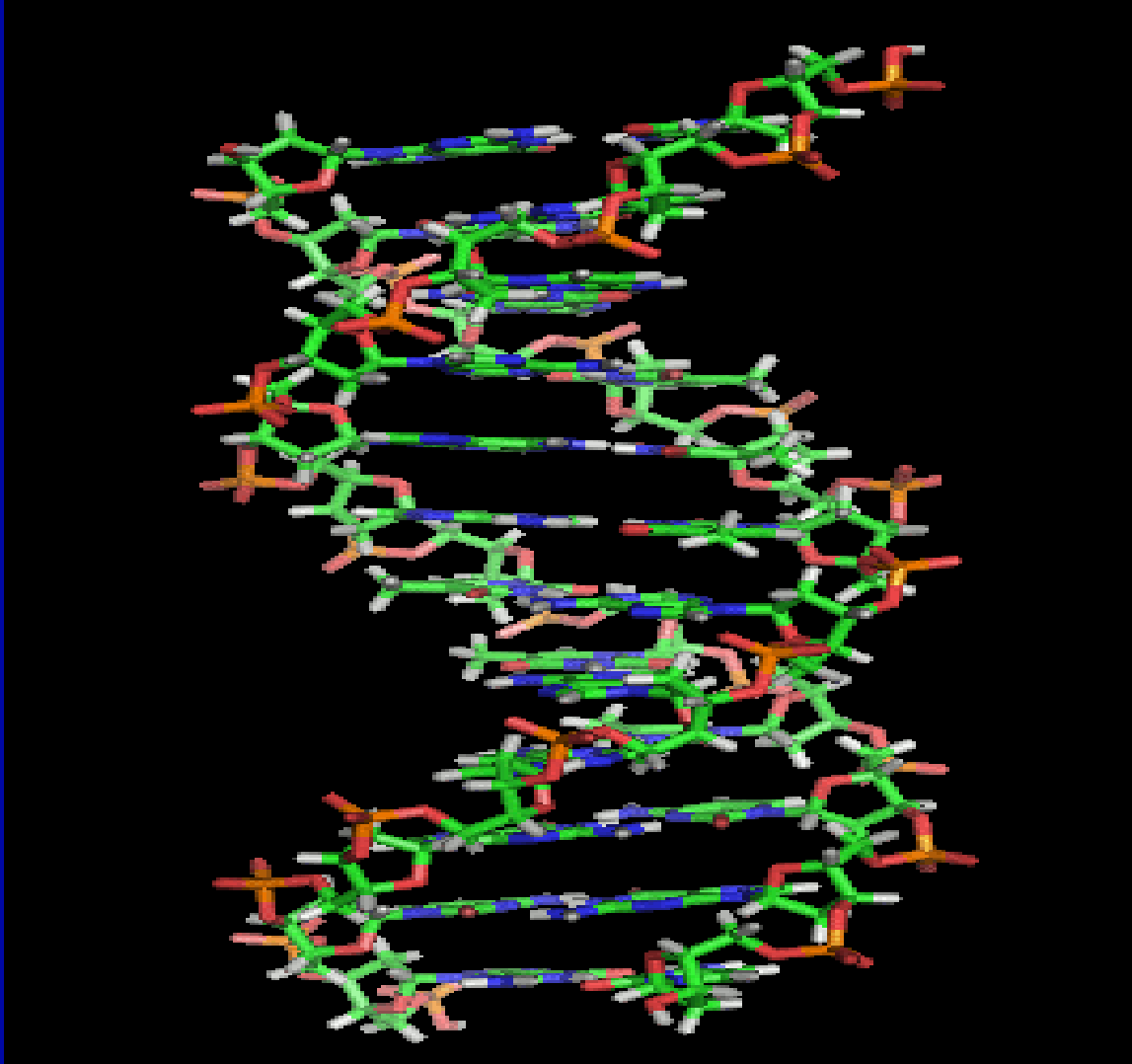
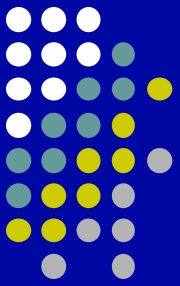


Αζωτούχος βάση
(A, G, C, T)

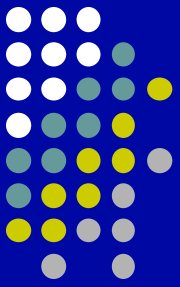
Σάκχαρο
(Δεοξυριβόζη)



DNA – διπλή έλικα

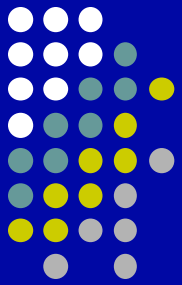


Η δομή του DNA



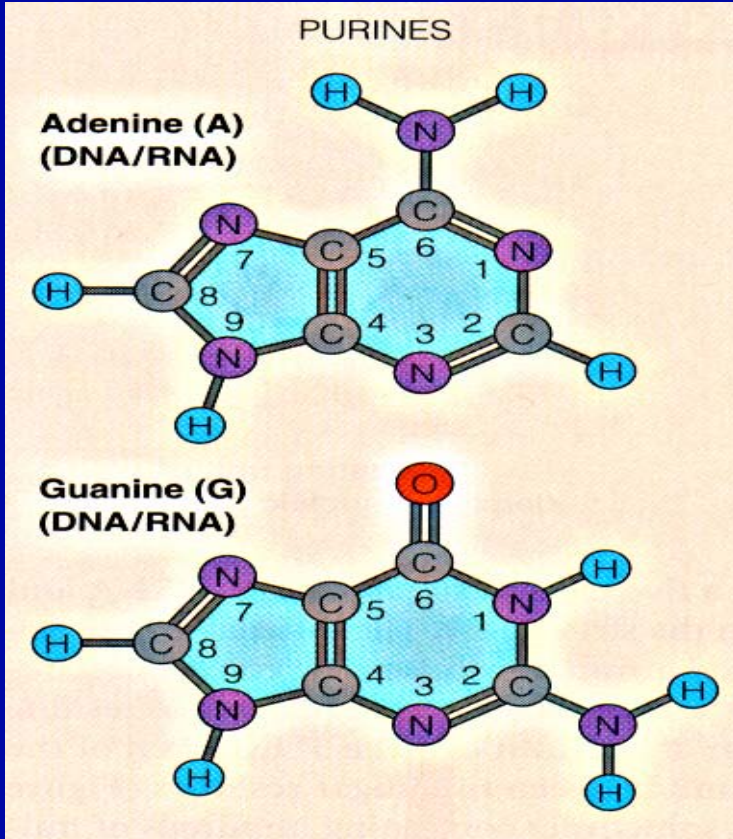
- Αποτελείται από 2 πολυνουκλεοτιδικές αλυσίδες σε διπλή έλικα που διατάσσονται αντιπαράλληλα
- Κάθε αλυσίδα αποτελείται από 4 είδη νουκλεοτιδίων (A=αδενίνη, T=θυμίνη, C=κυτοσίνη, G=γουανίνη)
- Τα νουκλεοτίδια αποτελούνται από μια πεντόζη με την οποία συνδέονται μία ή περισσότερες φωσφορικές ομάδες και μια αζωτούχο βάση
- Στην περίπτωση του DNA η πεντόζη είναι η **δεοξυριβόζη** (δεοξυριβονουκλεϊκό οξύ)
- Δεσμοί υδρογόνου σχηματίζονται μόνο μεταξύ των βάσεων αδενίνης-θυμίνης (A-T) και γουανίνης-κυτοσίνης (G-C)
- Κάθε κλώνος DNA περιέχει μια αλληλουχία νουκλεοτιδίων που είναι ακριβώς **συμπληρωματική** (complementary) προς την αλληλουχία των νουκλεοτιδίων του άλλου κλώνου



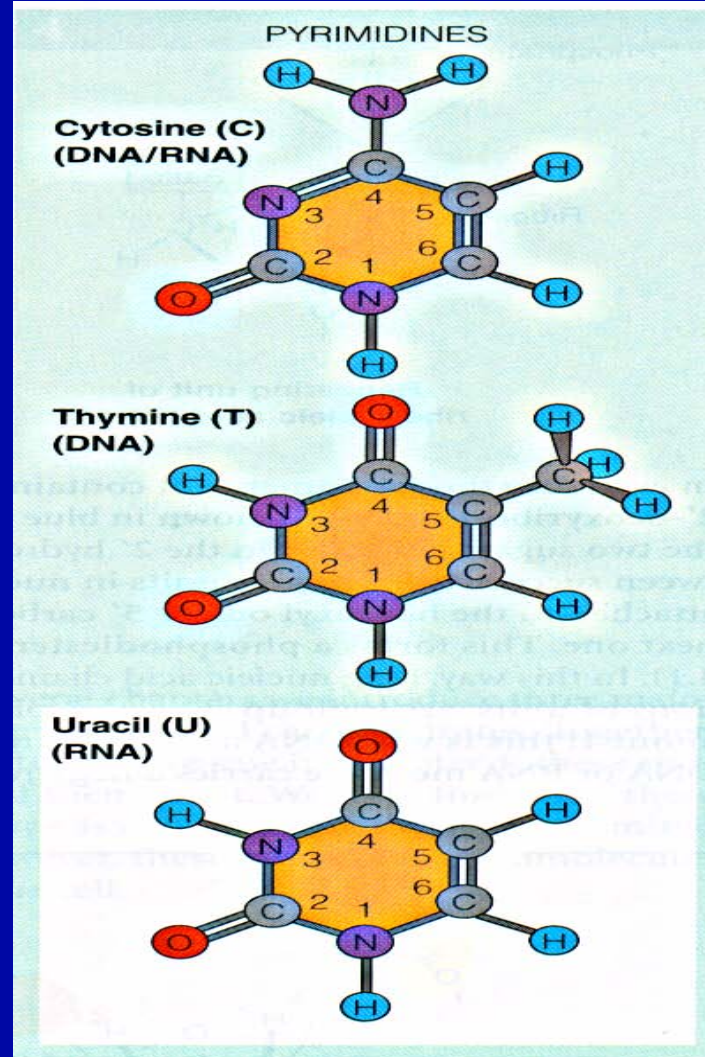


ΟΙ ΒΑΣΕΙΣ ΠΟΥ ΣΥΝΑΝΤΩΝΤΑΙ ΣΤΑ ΝΟΥΚΛΕΪΚΑ ΟΞΕΑ

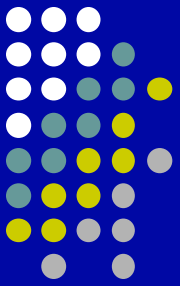
ΠΟΥΡΙΝΕΣ



ΠΥΡΙΜΙΔΙΝΕΣ

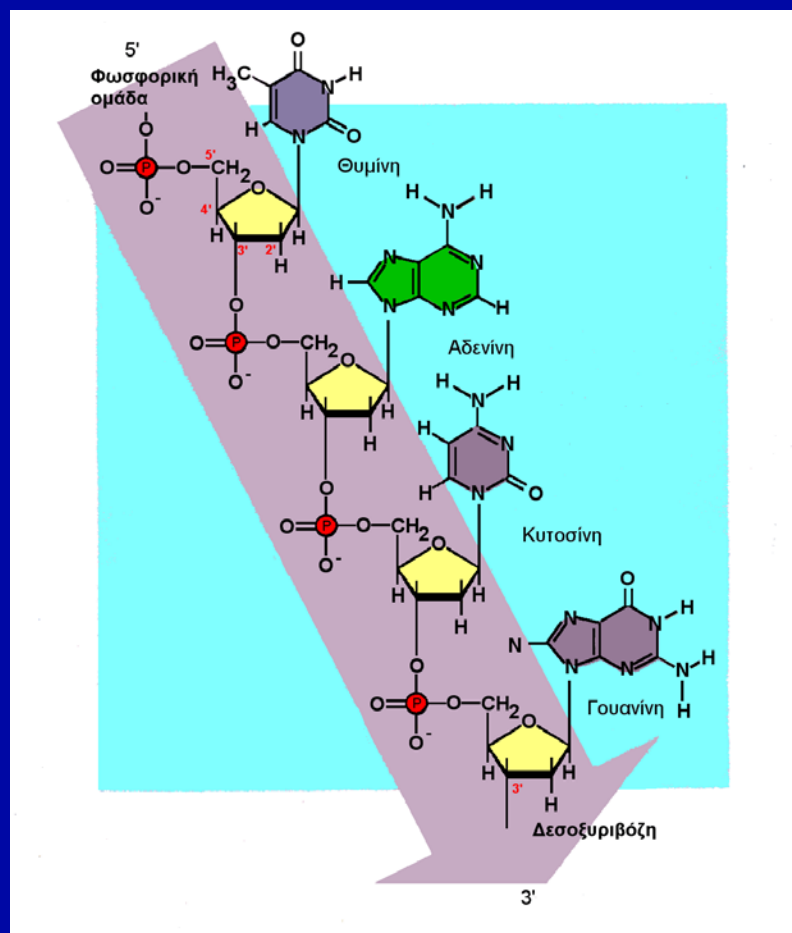
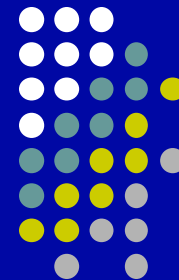


Προσανατολισμός DNA

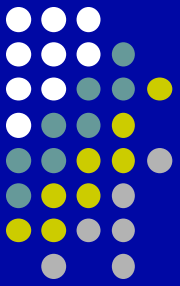


- Ανεξάρτητα από τον αριθμό των νουκλεοτιδίων από τα οποία αποτελείται η πολυνουκλεοτιδική αλυσίδα, το πρώτο της νουκλεοτίδιο έχει πάντα μια ελεύθερη φωσφορική ομάδα συνδεδεμένη στον 5' άνθρακα της πεντόζης του και το τελευταίο νουκλεοτίδιό της έχει ελεύθερο το υδροξύλιο του 3' άνθρακα της πεντόζης του. Όταν επιμηκύνεται μια αλυσίδα DNA ή RNA αυξάνεται πάντα με κατεύθυνση 5' προς 3' δηλαδή δημιουργείται πάντα ένα καινούργιο 3' άκρο. Για το λόγο αυτό αναφέρεται ότι ο προσανατολισμός της πολυνουκλεοτιδικής αλυσίδας είναι **5' προς 3'**





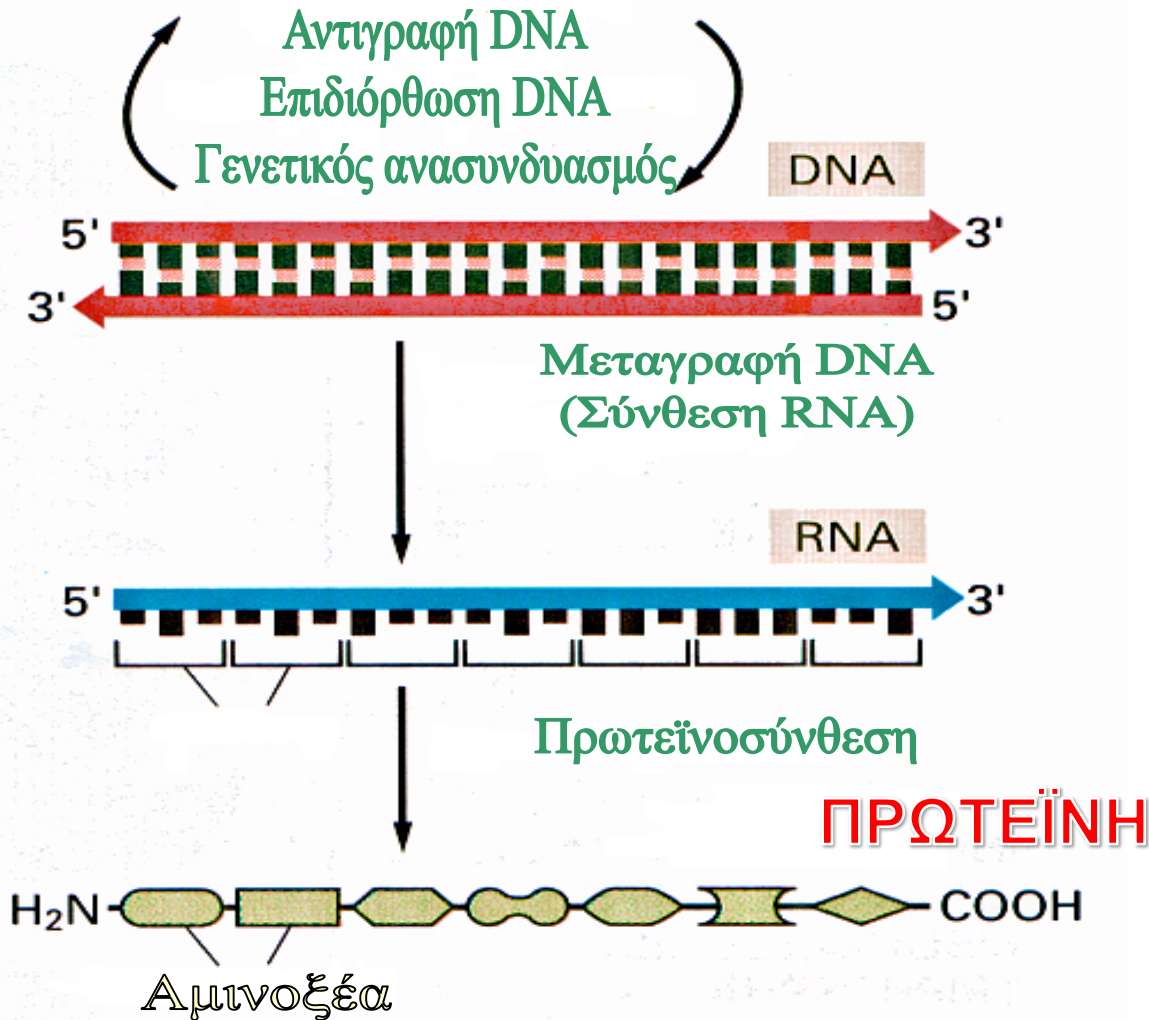
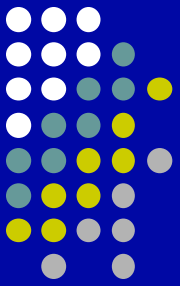
Διαφορές του RNA από το DNA



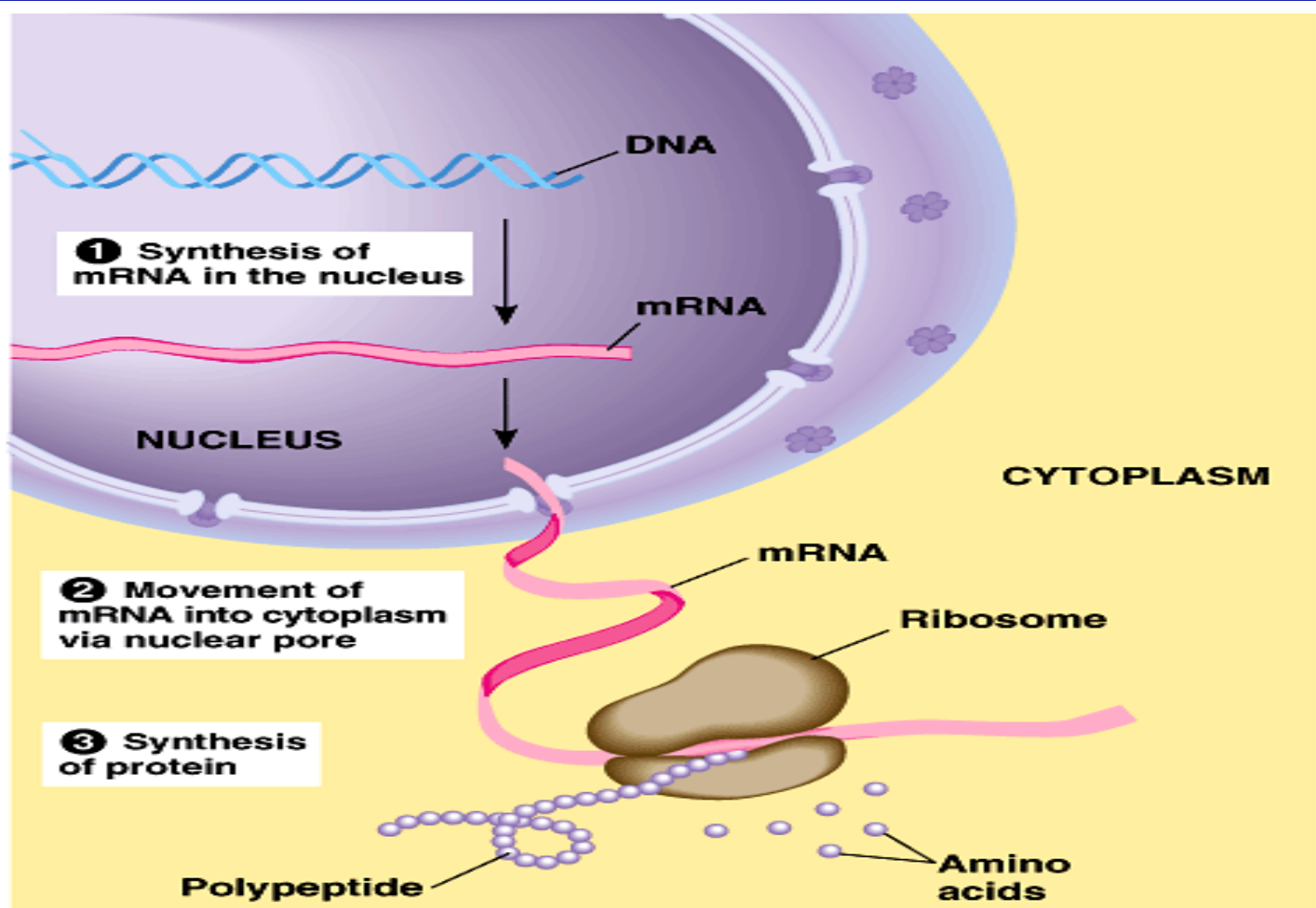
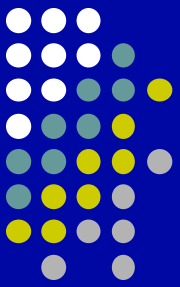
- τα νουκλεοτίδια του είναι ριβονουκλεοτίδια, δηλαδή περιέχουν το σάκχαρο **ριβόζη** (απ' όπου και η ονομασία ριβονουκλεϊνικό οξύ) αντί του σακχάρου δεοξυριβόζη
- όπως και το DNA, περιέχει τις βάσεις αδενίνη (A), γουανίνη (G) και κυτοσίνη (C), αλλά αντί για θυμίνη (T) περιέχει **ουρακίλη** (U). Η U, όπως και η T, ζευγαρώνει (με δεσμούς υδρογόνου) με την A
- Το DNA πάντοτε βρίσκεται στα κύτταρα υπό μορφή δίκλωνης έλικας, ενώ το RNA είναι **μονόκλωνο**



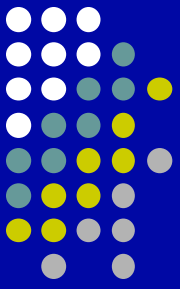
Το Κεντρικό Δόγμα της Μοριακής Βιολογίας



DNA-RNA-πρωτεΐνη – Διάγραμμα της ροής πληροφορίας στο κύτταρο

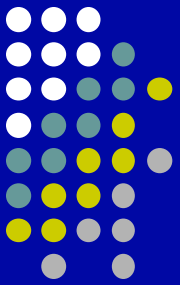


Ο γενετικός κώδικας



- Με τη μεταγραφή, οι πληροφορίες που βρίσκονται στα γονίδια μεταφέρονται στο mRNA με βάση τη συμπληρωματικότητα των νουκλεοτιδικών βάσεων. Η αλληλουχία των βάσεων του mRNA καθορίζει την αλληλουχία των αμινοξέων στις πρωτεΐνες με βάση ένα κώδικα αντιστοίχισης νουκλεοτιδίων RNA με αμινοξέα πρωτεϊνών ο οποίος ονομάζεται **γενετικός κώδικας**. Γι' αυτό η πρωτεϊνοσύνθεση είναι πραγματικά μια διαδικασία «μετάφρασης» από τη γλώσσα των βάσεων στη γλώσσα των αμινοξέων.
- Επειδή ο αριθμός των διαφορετικών αμινοξέων που συγκροτούν τις πρωτεΐνες είναι είκοσι και αντίστοιχα ο αριθμός των διαφορετικών νουκλεοτιδίων που σχηματίζουν το RNA είναι τέσσερα, θεωρήθηκε πιθανόν ότι τρία νουκλεοτίδια αντιστοιχούν σε ένα αμινοξύ και γι' αυτό ο γενετικός κώδικας ονομάστηκε **κώδικας τριπλέτας**. Ο κώδικας τριπλέτας είναι φυσική συνέπεια του γεγονότος ότι τέσσερα νουκλεοτίδια, αν συνδυαστούν ανά ένα ($4^1=4$) ή ανά δύο ($4^2=16$), δε δίνουν αρκετούς συνδυασμούς για να κωδικοποιηθούν τα είκοσι αμινοξέα. Αν όμως συνδυαστούν ανά **τρία** ($4^3=64$), οι συνδυασμοί είναι παραπάνω από αρκετοί.





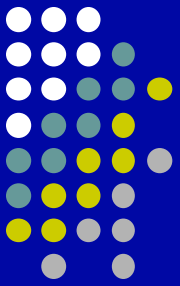
Συμπερασματικά

- Κάθε ομάδα τριών διαδοχικών νουκλεοτιδίων του RNA αποκαλείται κωδικόνιο (codon) και **κάθε κωδικόνιο καθορίζει ένα αμινοξύ**
- Οι κανόνες με τους οποίους η αλληλουχία των νουκλεοτιδίων ενός γονιδίου, μέσω του mRNA, μεταφράζεται στην αλληλουχία των αμινοξέων μιας πρωτεΐνης συλλογικά αναφέρονται ως γενετικός κώδικας (genetic code)

	AGA									UUA										AGC
	AGG									UUG										AGU
GCA	CGA						GGA			CUA				CCA	UCA	ACA			GUA	
GCC	CGC						GGC		AUA	CUC				CCC	UCC	ACC			GUC	UAA
GCG	CGG	GAC	AAC	UGC	GAA	CAA	GGG	CAC	AUC	CUG	AAA		UUC	CCG	UCG	ACG		UAC	GUG	UAG
GCU	CGU	GAU	AAU	UGU	GAG	CAG	GGU	CAU	AUU	CUU	AAG	AUG	UUU	CCU	UCU	ACU	UGG	UAU	GUU	UGA
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	τερματισμός
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	



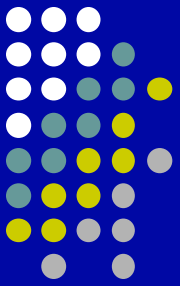
Τα βασικά χαρακτηριστικά του γενετικού κώδικα είναι:



- Ο γενετικός κώδικας είναι **κώδικας τριπλέτας**, δηλαδή μια τριάδα νουκλεοτιδίων του mRNA, το **κωδικόνιο**, κωδικοποιεί ένα αμινοξύ.
- Ο γενετικός κώδικας είναι **συνεχής**, δηλαδή το mRNA διαβάζεται συνεχώς ανά τρία νουκλεοτίδια χωρίς να παραλείπεται κάποιο νουκλεοτίδιο.
- Ο γενετικός κώδικας είναι **μη επικαλυπτόμενος**, δηλαδή κάθε νουκλεοτίδιο ανήκει σε ένα μόνο κωδικόνιο.
- Ο γενετικός κώδικας είναι **σχεδόν καθολικός (παγκόσμιος)**. Όλοι οι οργανισμοί έχουν τον ίδιο γενετικό κώδικα. Αυτό πρακτικά σημαίνει ότι το mRNA από οποιονδήποτε οργανισμό μπορεί να μεταφραστεί σε εκχυλίσματα φυτικών, ζωικών ή βακτηριακών κυττάρων in vitro και να παράγει την ίδια πρωτεΐνη.
- Ο γενετικός κώδικας είναι **εκφυλισμένος**. Με εξαίρεση δύο αμινοξέα (μεθειονίνη και τρυπτοφάνη) τα υπόλοιπα 18 κωδικοποιούνται από δύο μέχρι και έξι διαφορετικά κωδικόνια. Τα κωδικόνια που κωδικοποιούν το ίδιο αμινοξύ ονομάζονται **συνώνυμα**.
- Ο γενετικός κώδικας έχει **κωδικόνια έναρξης** και **κωδικόνια λήξης**. Το κωδικόνιο έναρξης σε όλους τους οργανισμούς είναι το AUG και κωδικοποιεί το αμινοξύ μεθειονίνη. Υπάρχουν τρία κωδικόνια λήξης, τα UAG, UGA και UAA. Η παρουσία των κωδικονίων αυτών στο μόριο του mRNA οδηγεί στον τερματισμό της σύνθεσης της πολυπεπτιδικής αλυσίδας.



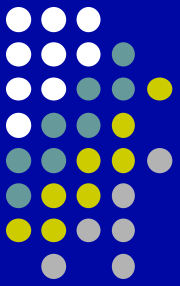
Πρωτεΐνες



- Είναι μεγάλα σύνθετα βιομόρια
- Αποτελούνται από αμινοξέα τα οποία ενώνονται μεταξύ τους με πεπτιδικούς δεσμούς
- Όλες οι πρωτεΐνες περιέχουν άνθρακα, οξυγόνο και άζωτο και οι περισσότερες εξ'αυτών και θείο
- Η ακολουθία αμινοξέων σε μια πρωτεΐνη καθορίζεται από ένα γονίδιο και κωδικοποιείται από το γενετικό κώδικα DNA
- Παράγονται στα ριβοσώματα



Επίπεδα πρωτεϊνικής δομής

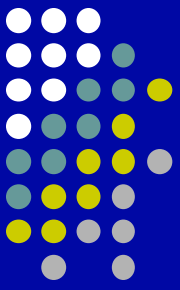


Η λειτουργικότητα μιας πρωτεΐνης είναι αποτέλεσμα της στερεοδιάταξής της

5 επίπεδα δομών

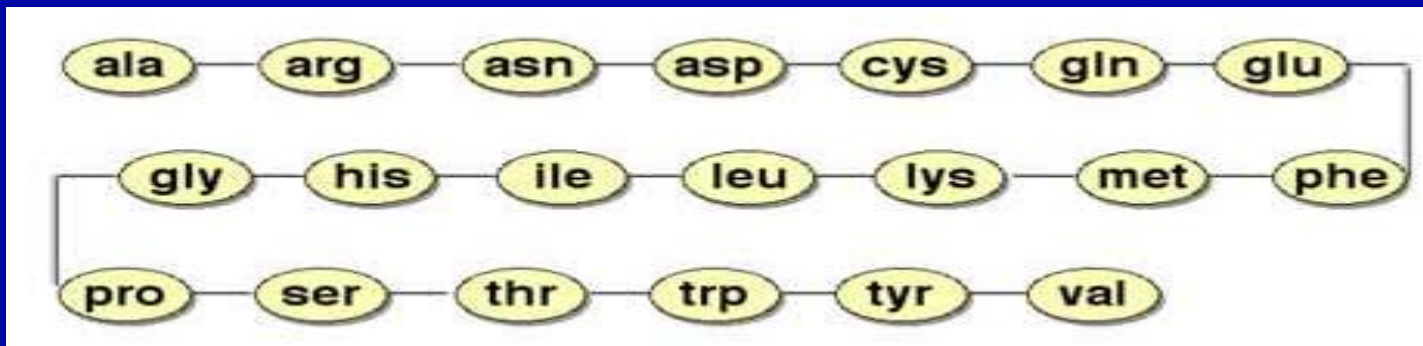
- Πρωτοταγής δομή
- Δευτεροταγής δομή
- Τριτοταγής δομή
- Τεταρτοταγής δομή
- Πενμπτοταγής δομή

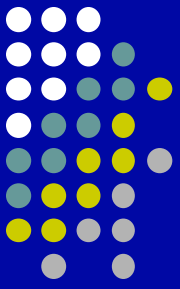




Πρωτοταγής δομή

- Είναι η γραμμική αλληλουχία των αμινοξέων πάνω στην αλυσίδα
- Η αλληλουχία των αμινοξέων καθορίζει μία και καλά καθορισμένη τρισδιάστατη διάταξη των ατόμων
- Η πρωτοταγής δομή είναι μοναδική για κάθε πρωτεΐνη





Δευτεροταγής δομή

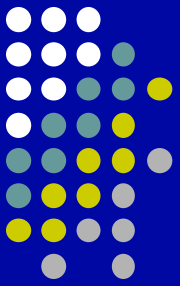
Οι πολυπεπτιδικές αλυσίδες μπορούν να αναδιπλώνονται σε κανονικές επαναλαμβανόμενες δομές. Αυτές είναι τρεις και ορίζουν την *δευτεροταγή δομή των πρωτεϊνών*:

- α) α-Έλικα
- β) β-Πτυχωτή Επιφάνεια
- γ) Έλικα Κολλαγόνου

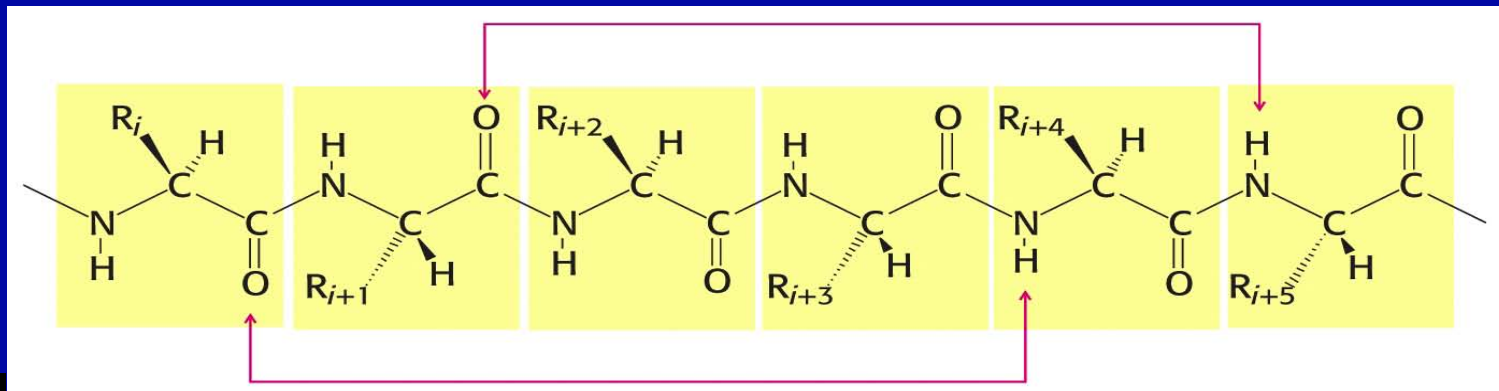
Και οι τρεις βασίζονται στους δεσμούς υδρογόνου



α-έλικα

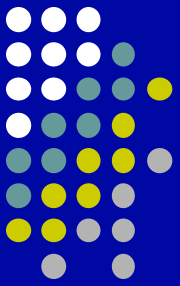
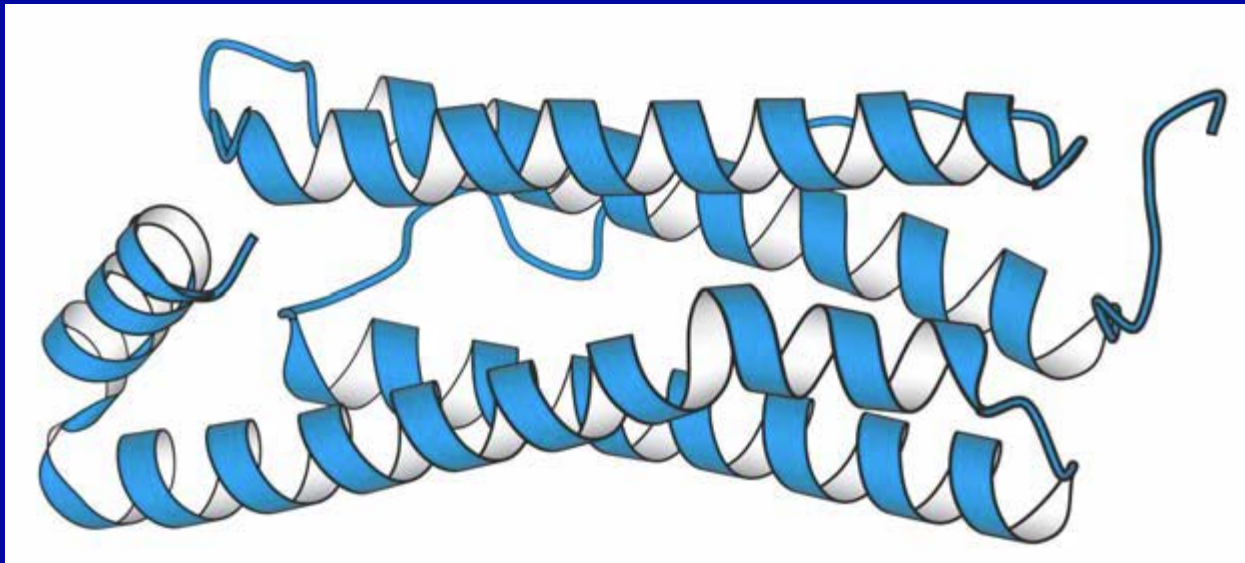


- Έχει ραβδωτή μορφή με την κύρια πεπτιδική αλυσίδα στο εσωτερικό και τις πλευρικές αλυσίδες να εκτείνονται προς τα έξω σε ελικοειδή διαμόρφωση
- Σταθεροποιείται με δεσμούς υδρογόνου μεταξύ των ομάδων NH και CO της κύρια αλυσίδας που απέχουν κατά 4 μονομερή.

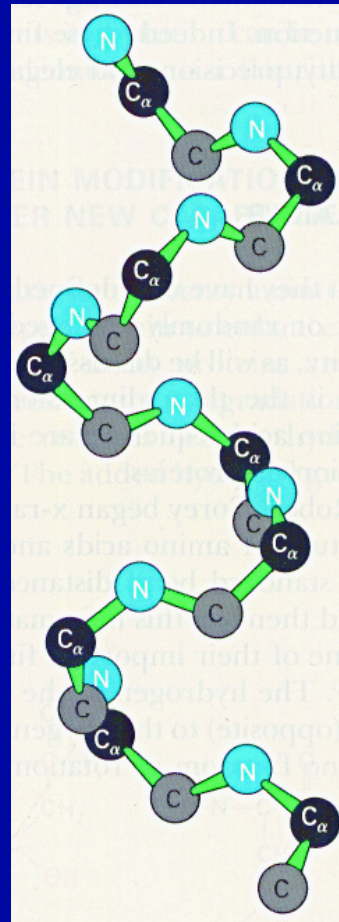
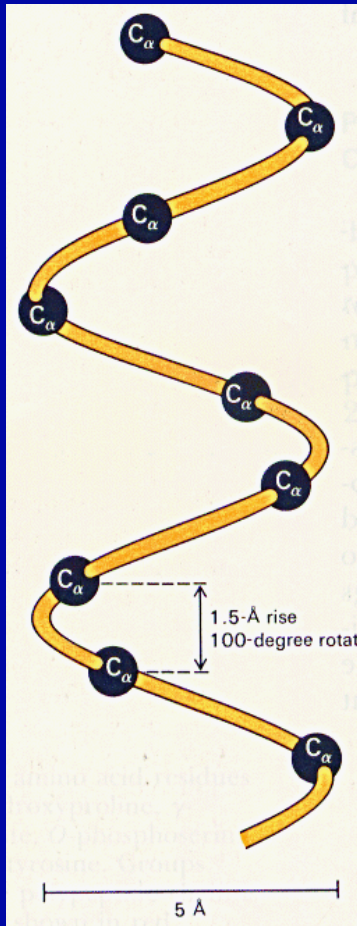
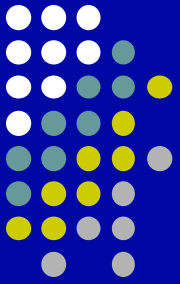


α-έλικα

- Η συνήθισμένη απεικόνιση της α-Έλικας σαν ελικοειδείς κορδέλες:

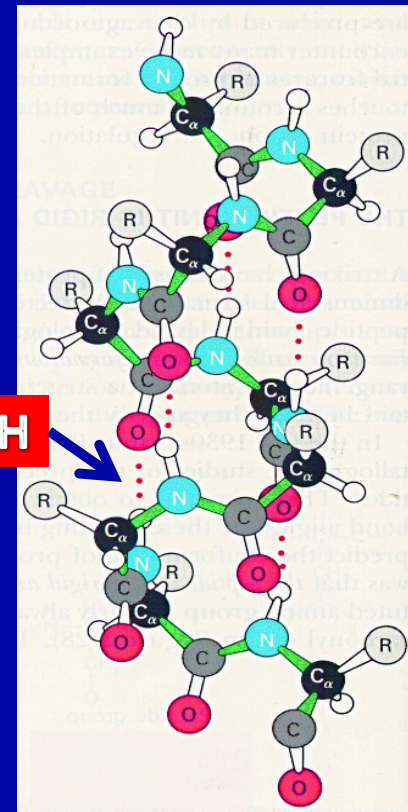


α-έλικα



Υδρόφιλα R: προς τα έξω

Δεσμοί Η

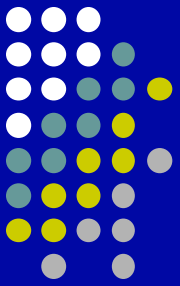


3,6 αμινοξέα ανά στροφή

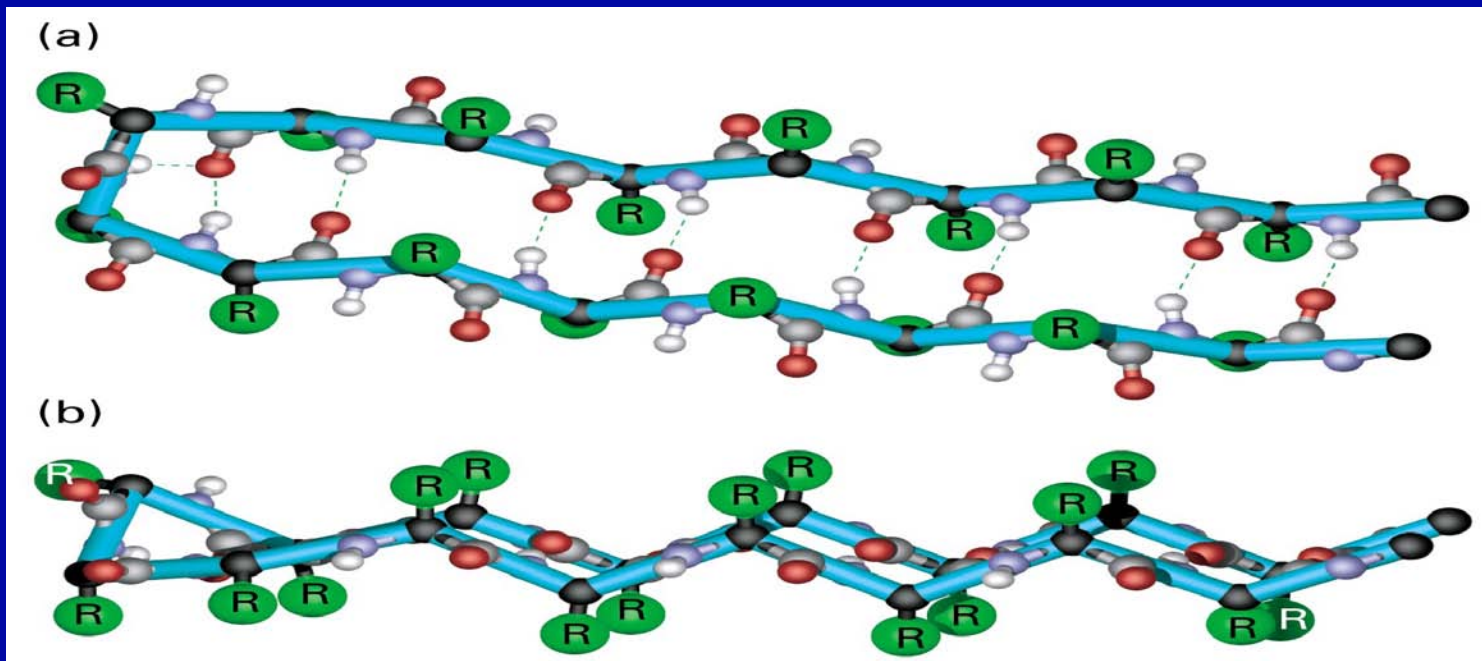
Υδρόφοβα R: προς τα μέσα



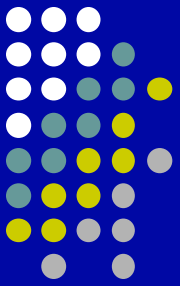
β-πτυχωτή επιφάνεια



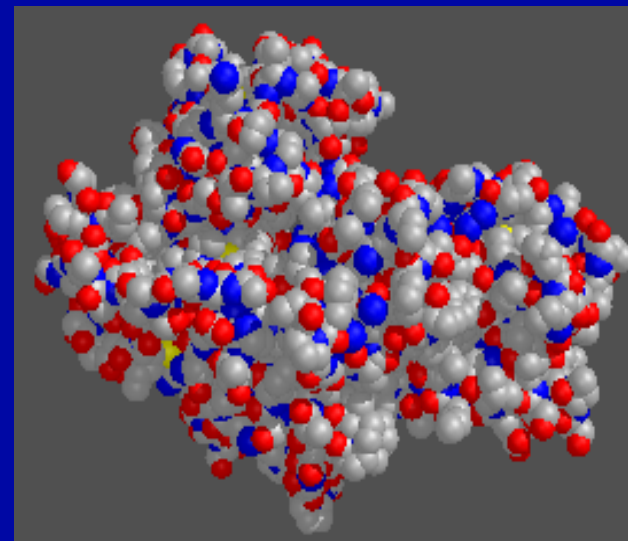
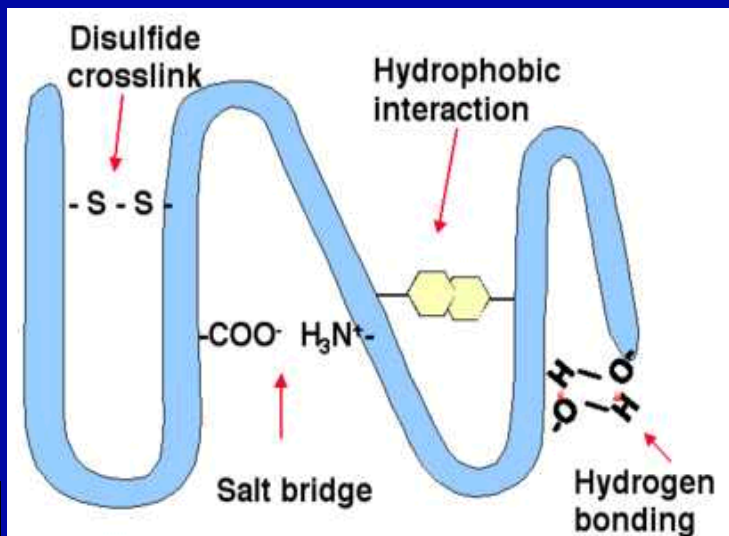
- Είναι επίπεδη επιφάνεια όπου η πολυπεπτιδική αλυσίδα είναι σχεδόν τελείως ανοιχτή. Σταθεροποιείται με δεσμούς υδρογόνου ανάμεσα σε διαφορετικές αλυσίδες
- Οι αλυσίδες μπορεί να έχουν αντίθετη (α) ή ίδια (β) κατεύθυνση



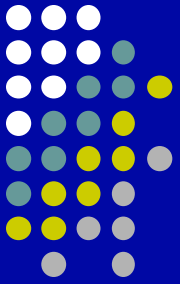
Τριτοταγής δομή



- Η τριτοταγής δομή των πρωτεϊνών αναφέρεται στην πλήρη στερεοδιάταξη όλων των δομικών μονάδων και είναι αυτή που τελικά καθορίζει την βιολογική λειτουργία τους
- Η φύση όμως της επιφάνειας διαφέρει από το εσωτερικό της (υδρόφοβες/ υδρόφιλες, μη πολικές/ πολικές)



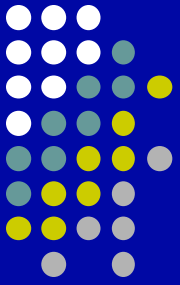
Τεταρτοταγής δομή



- Ένα υψηλότερο επίπεδο οργάνωσης της δομής των πρωτεϊνικών μορίων.
- Είναι ένα σύμπλοκο δύο ή περισσότερων ομοίων ή διαφορετικών πολυπεπτιδικών αλυσίδων, που λέγονται υπομονάδες ή πρωτομερή.
- Τα πρωτομερή συνδέονται μεταξύ τους όχι με χημικούς δεσμούς, αλλά με δευτερεύοντες ηλεκτροστατικούς δεσμούς π.χ.
 - ✓ ιονικούς,
 - ✓ δεσμούς υδρογόνου,
 - ✓ δυνάμεις Van der Waals κ.λπ.



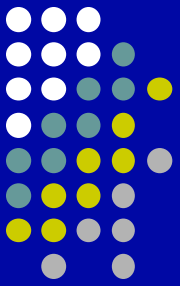
Πεμπτοταγής δομή



- Το υψηλότερο επίπεδο οργάνωσης των πρωτεϊνών (πολυπρωτεϊνικά ή πολυ-ενζυμ(ατ)ικά συστήματα).
- Κάθε συνδεδεμένη μονάδα διαθέτει μια εξειδικευμένη δραστηριότητα την οποία και διατηρούν αναλλοίωτη και μετά τη διάσπαση του συστήματος.
- Με τη βοήθεια ενός τέτοιου πολυενζυμικού συστήματος επιτυγχάνεται να έλθει σε πέρας μια ολόκληρη σειρά δράσεων που αντιπροσωπεύει είτε μια διάκριτη πορεία βιοχημικών αντιδράσεων είτε ένα βιολογικό μηχανισμό (π.χ. μυική συστολή



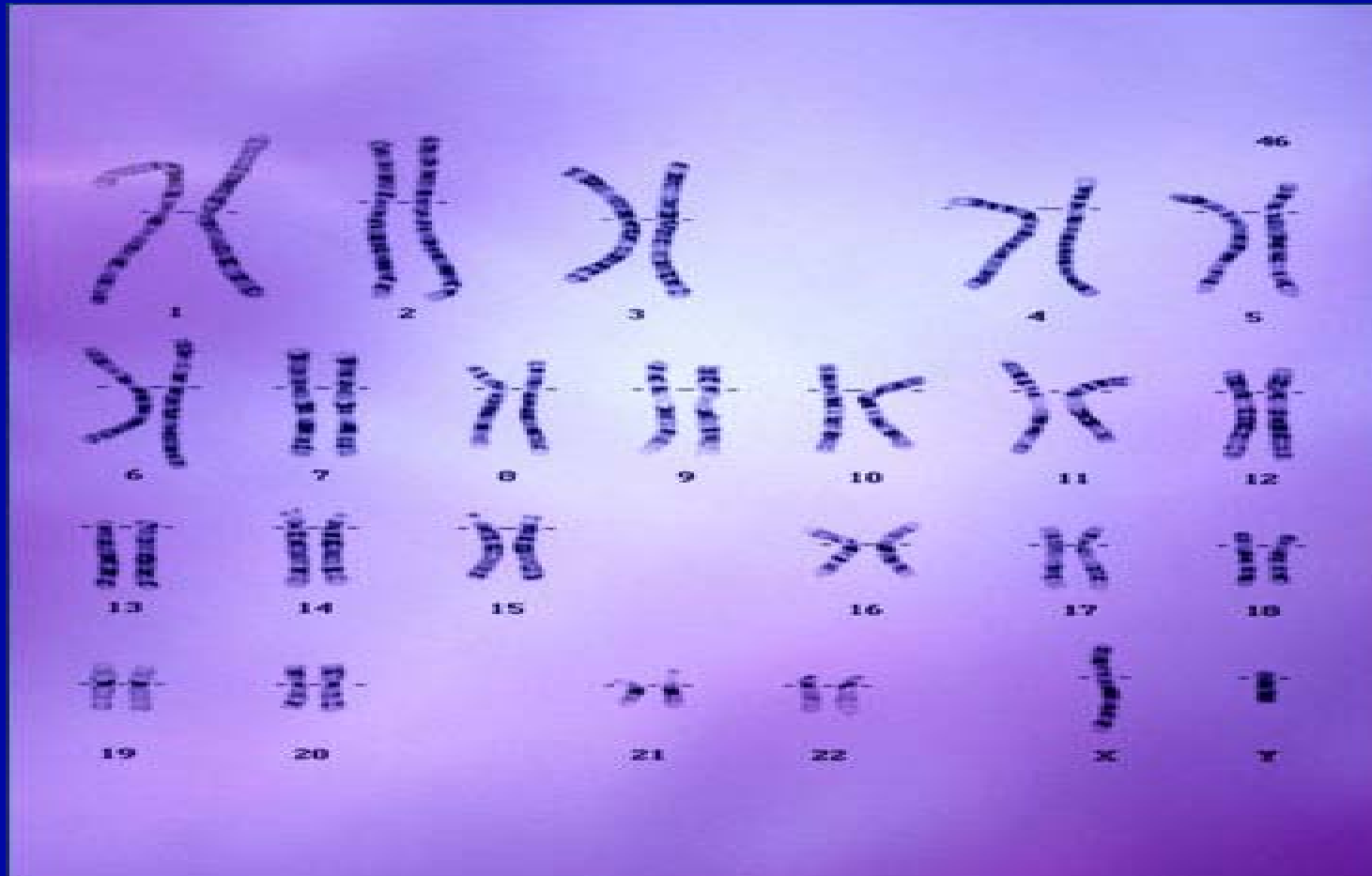
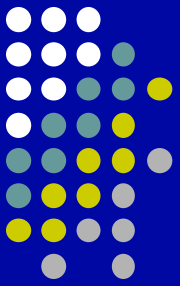
Χρωμοσώματα



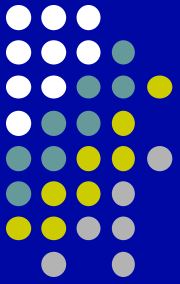
- Τα χρωμοσώματα αποτελούνται από DNA που περιέχει πολλά γονίδια και βρίσκονται στον πυρήνα του κυττάρου
- Τα χρωμοσώματα στον άνθρωπο βρίσκονται ανά ζεύγη σχηματίζοντας **23 ζευγάρια**
 - **22 ζευγάρια αυτοσωμικά** (ίδια μορφολογικά στα αρσενικά και στα θηλυκά)
 - **1 ζευγάρι φυλετικών χρωμοσωμάτων (X και Y)**
- Τα κύτταρα των θηλυκών ατόμων περιέχουν δύο χρωμοσώματα X, ενώ τα κύτταρα των αρσενικών ατόμων ένα X και ένα μικρότερο Y χρωμόσωμα
- Η ταξινόμηση των χρωμοσωμάτων ενός ατόμου σύμφωνα με το μέγεθος και το σχήμα τους ονομάζεται **καρυότυπος**



Καρυότυπος ανθρώπου



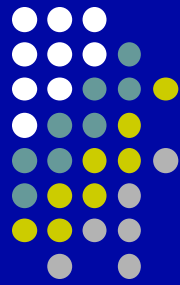
Χρωμοσώματα



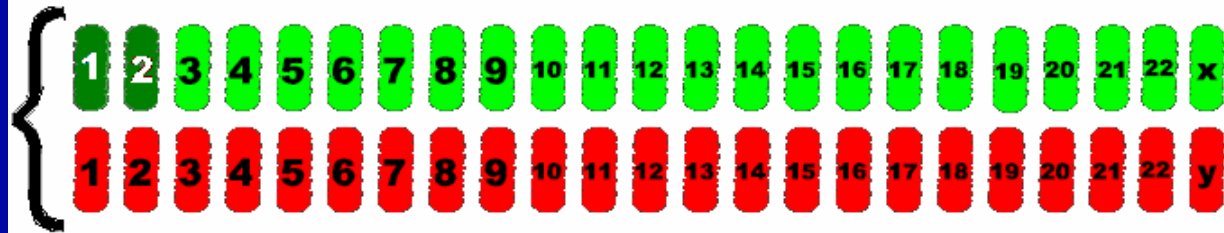
- Κάθε γονέας συνεισφέρει μια πλήρη σειρά χρωμοσωμάτων στο παιδί (από 23 χρωμοσώματα)
- Με αυτόν τον τρόπο οι γονείς περνούν γονίδια στο παιδί
- Κάθε παιδί αποκτά τα μισά χρωμοσώματά του από τον πατέρα του και τα άλλα μισά από τη μητέρα του
- Αυτή η μεταβίβαση συμβαίνει όταν τα σπερματοζωάρια ενώνονται με τα ωάρια
- Τα σπερματοζωάρια και τα ωάρια είναι απλοειδή, δηλαδή περιέχουν μια σειρά χρωμοσωμάτων (23 χρωμοσώματα)
- Όταν αυτά ενώνονται δημιουργούν ένα απλό κύτταρο, που ονομάζεται ζυγωτό, που έχει 2 πλήρεις σειρές από 23 χρωμοσώματα ($2 \times 23 = 46$)



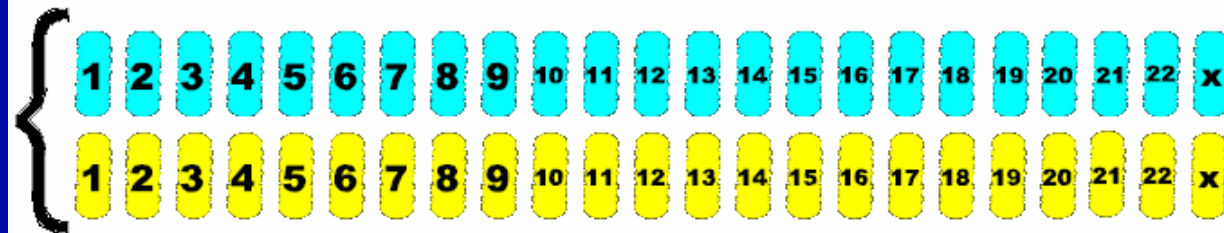
Κάθε γονέας συνεισφέρει μια πλήρη σειρά χρωμοσωμάτων στο παιδί



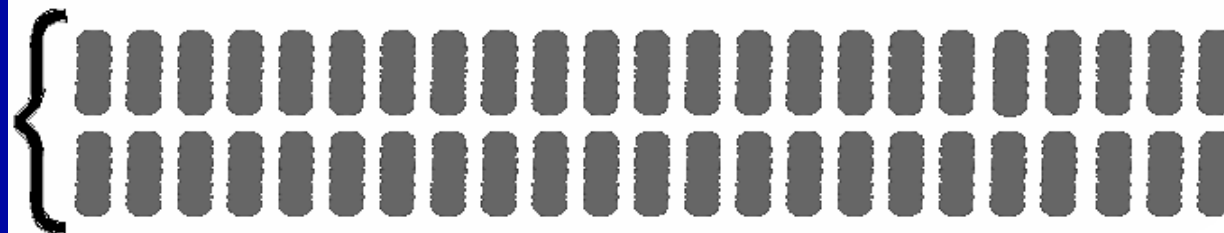
Χρωμοσώματα του πατέρα



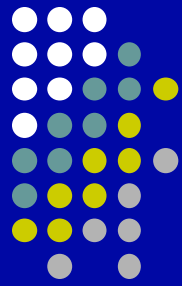
Χρωμοσώματα της μητέρας



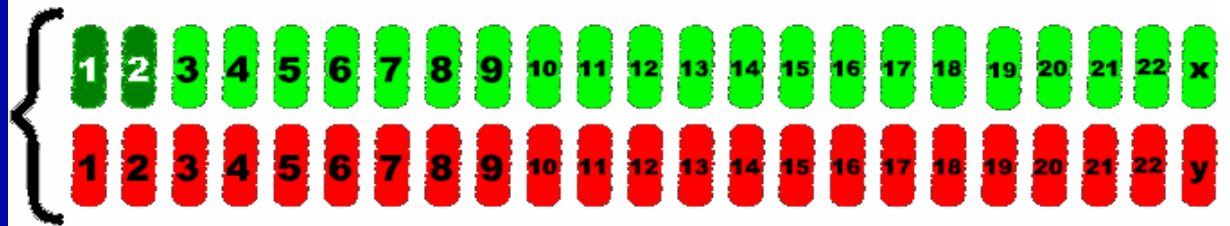
Πιθανά χρωμοσώματα 1^{ου} παιδιού



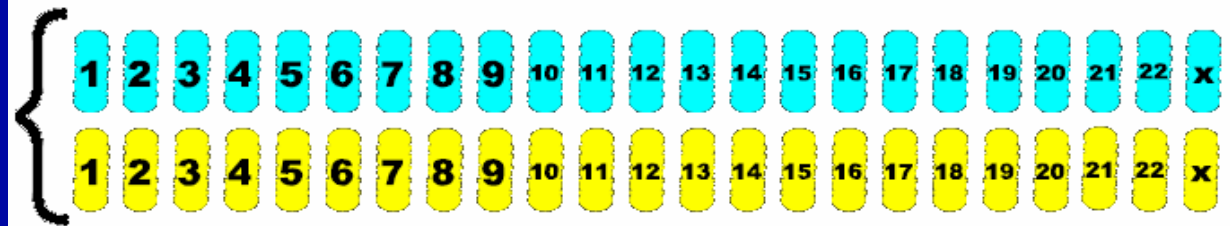
Ο μόνος κανόνας είναι ότι το παιδί πρέπει να αποκτήσει 1 από κάθε χρωμόσωμα



Χρωμοσώματα του πατέρα



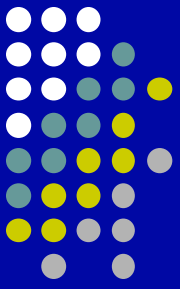
Χρωμοσώματα της μητέρας



Πιθανά χρωμοσώματα 1^{ου} παιδιού



Χρωμοσωμικές μεταλλάξεις

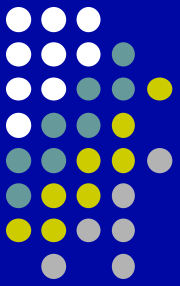


Έχουν ως αποτέλεσμα:

- Την αλλαγή της **δομής** των χρωμοσωμάτων (χρωματοσωματική αναδιάταξη)
- Τον **αριθμό** των χρωματοσωμάτων (γενοτυπική μετάλλαξη)



Χρωματοσωμικές αναδιατάξεις

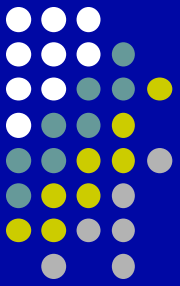


Έλλειμμα

- Συμβαίνει όταν ένα τμήμα ενός χρωμοσώματος λείπει
- Τα ελλείμματα έχουν συνήθως σημαντικές γενετικές επιπτώσεις
- Όταν βρίσκονται σε ομοζυγωτική κατάσταση είναι θνησιγόνα
- Η μετάλλαξη που δημιουργεί έλλειμμα, μπορεί να γίνει επικίνδυνη, αν το έλλειμμα πάρει την θέση ενός υγιούς και επικρατούς αλληλόμορφου, που βρίσκεται απέναντι σε ένα υπολειπόμενο θνησιγόνο
- Το σύνδρομο Cri-du-chat προέρχεται από έλλειμμα του ενός άκρου του χρωμοσώματος 5, με αποτέλεσμα το άτομο να έχει μόνο ένα αντίγραφο όλων των γονιδίων του στο άκρο αυτού του χρωμοσώματος. Οι άνθρωποι με το σύνδρομο αυτό έχουν φυσική και νοητική καθυστέρηση



Γενοτυπικές μεταλλάξεις



- **Ανευπλοειδία**
η αλλαγή του αριθμού των χρωμοσωμάτων
- **Ευπλοειδία**
η αλλαγή των χρωμοσωματικών σειρών

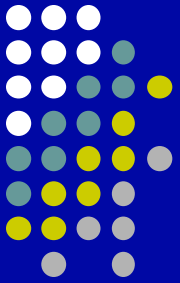
χωρίζεται σε:

Απλοειδία – απουσία μιας χρωμοσωμικής σειράς

Ευπλοειδία – αύξηση των χρωμοσωμικών σειρών



Ανευπλοειδία

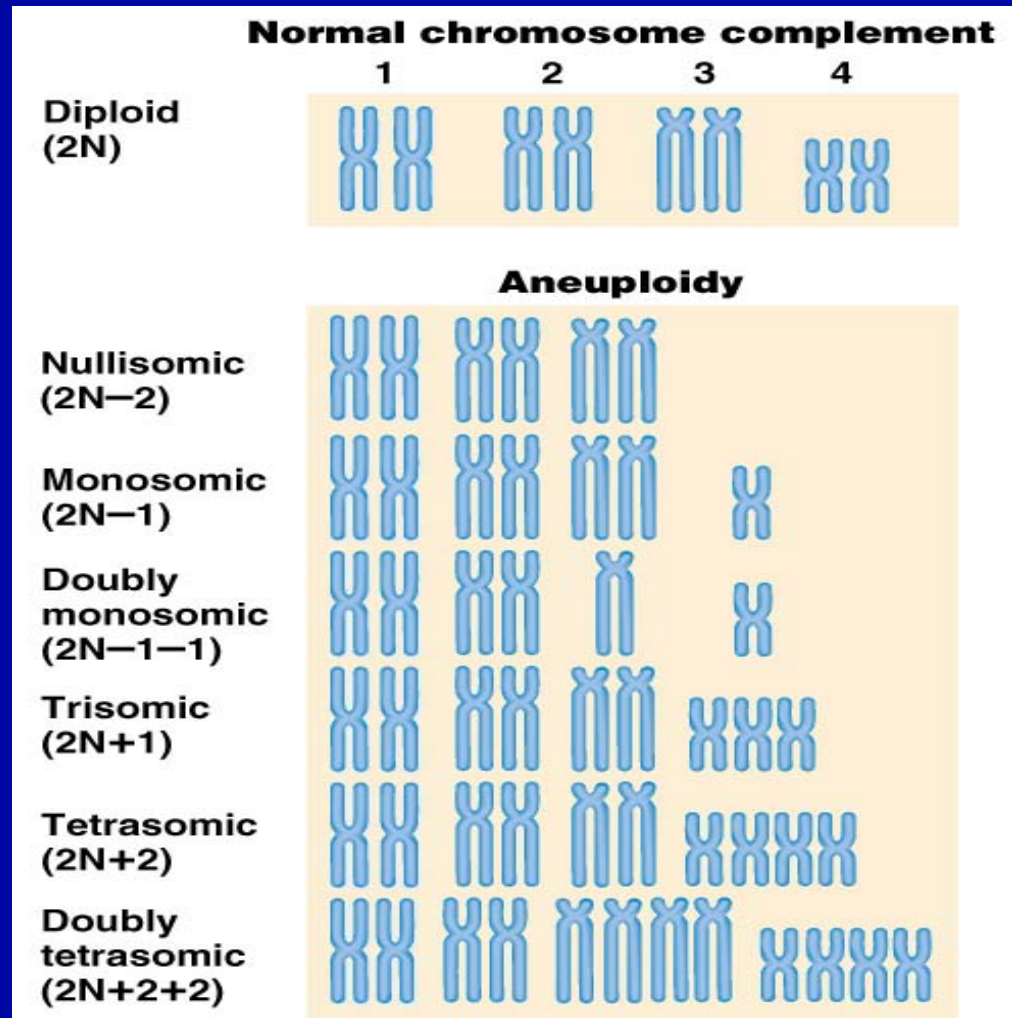


Ασωμία = έλλειψη ενός ομόλογου χρωμοσωμικού Ζευγαριού

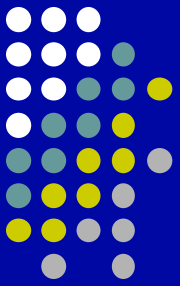
Μονοσωμία = έλλειψη ενός Χρωμοσώματος

Τρισωμία = αύξηση των χρωμοσωμάτων κατά ένα

Τετρασωμία = αύξηση των χρωμοσωμάτων κατά δύο



Γονίδιο



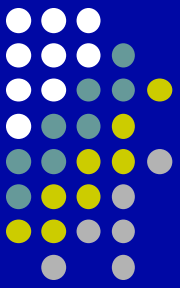
- Η βασική μονάδα της κληρονομικότητας
- Μεταβιβάζει μια πληροφορία από ένα κύτταρο και κατ'επέκταση από τη μια γενιά στην άλλη
- Αποτελείται από DNA, βρίσκεται στα χρωμοσώματα και κωδικοποιεί πρωτεΐνες

Αποτελείται από διάφορα μέρη και περιλαμβάνει

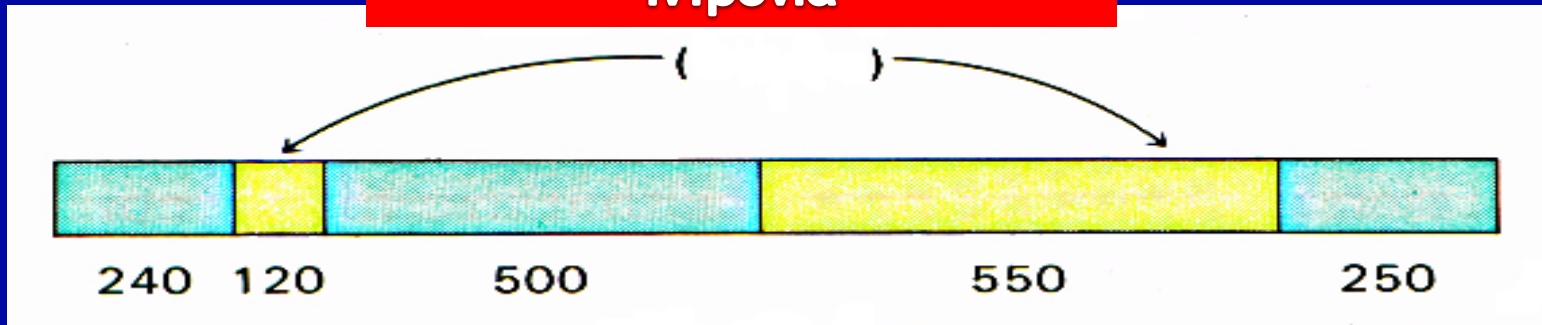
- **Εξώνια** (**exon** - the **ex**pressed parts) – περιοχές που κωδικοποιούν γονιδιακά προϊόντα
- **Ιντρόνια** (**intron** – the **int**ervening parts) – περιοχές που δεν κωδικοποιούν γονιδιακά προϊόντα



Μερικά γονίδια διακόπτονται από ιντρόνια. Αυτό συμβαίνει κυρίως στα ευκαρυωτικά γονίδια



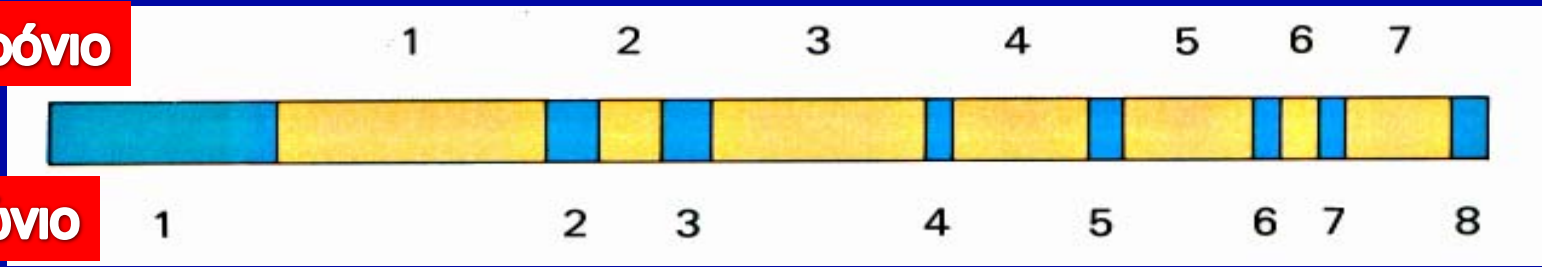
Παρεμβαλλόμενες ακολουθίες ιντρόνια



Γονίδιο β-globin

Ζεύγη βάσεων

ΙΝΤΡΟΝΙΟ

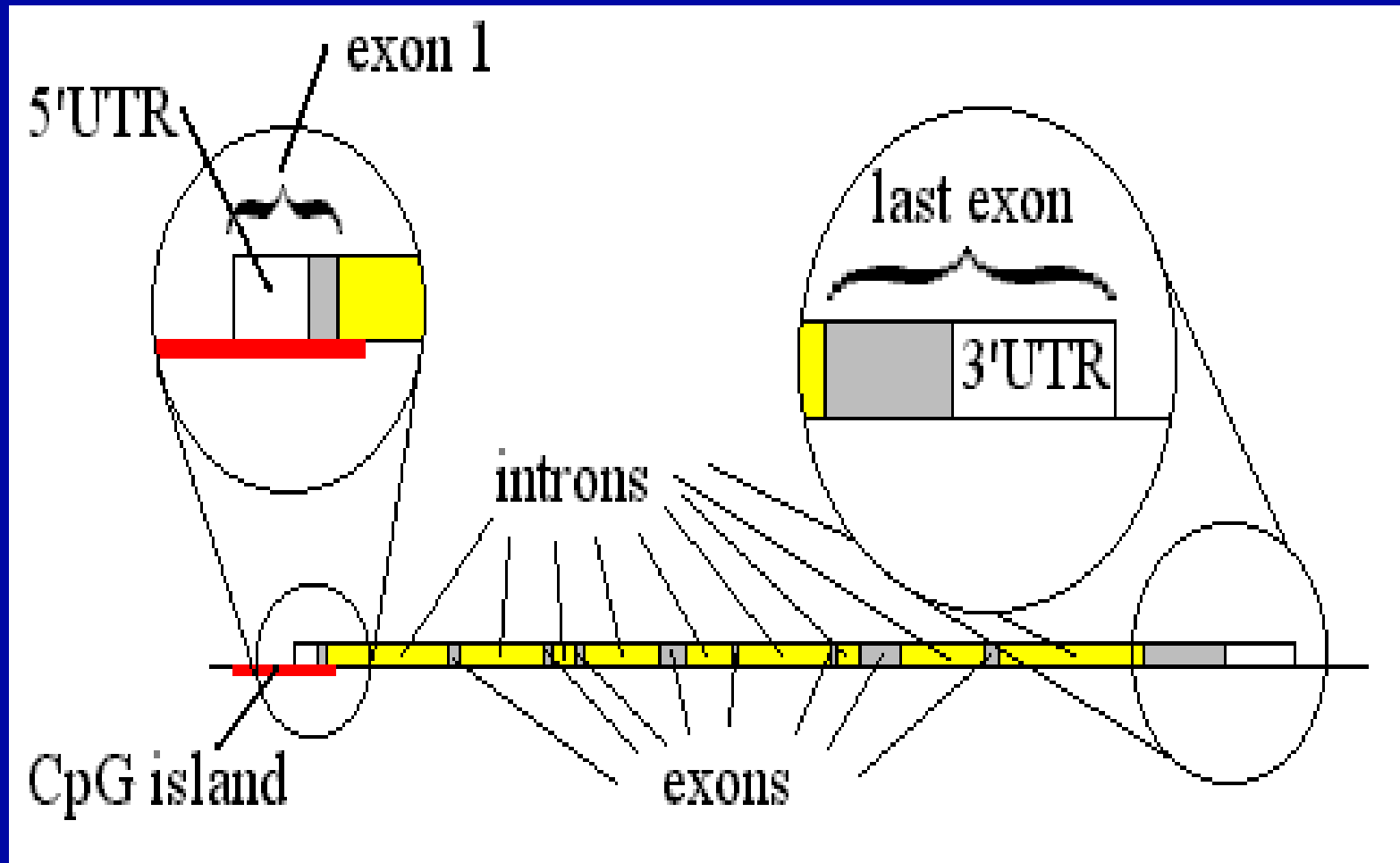
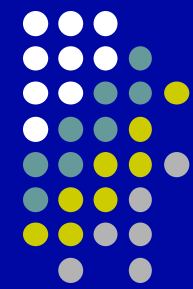


ΕΞΩΝΙΟ

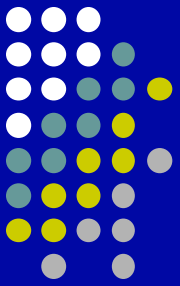
Γονίδιο Ovalbumin



“Τυπικό” γονίδιο

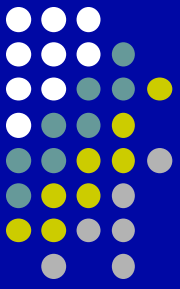


Συμπεράσματα προγράμματος ανθρώπινου γονιδιώματος



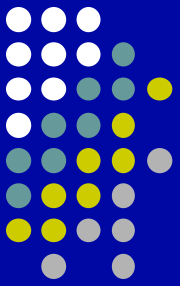
- Το ανθρώπινο γονιδίωμα περιλαμβάνει 3.164.700.000 νουκλεοτιδικές βάσεις
- Το μέσο μέγεθος για ένα γονίδιο είναι 3.000 βάσεις αλλά τα μεγέθη διαφέρουν σημαντικά με το πιο μεγάλο γονίδιο να είναι αυτό της μυϊκής δυστροφίας που έχει 2.4 εκατομμύρια βάσεις
- Ο συνολικός αριθμός των γονιδίων του ανθρώπινου γονιδιώματος είναι περίπου 25.000, πολύ μικρότερος από τον αριθμό των 80.000-100.000 που είχε αρχικά υπολογιστεί
- Περίπου 99.9% των νουκλεοτιδικών βάσεων είναι ίδιες σε όλους τους ανθρώπους





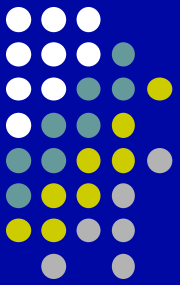
- Οι περιοχές του γονιδιώματος που είναι πλούσιες σε γονίδια έχουν μεγάλη περιεκτικότητα σε γουανίνη (G) και κυτοσίνη (C)
- Αντίθετα, οι “έρημες” από γονίδια περιοχές του γονιδιώματος είναι πλούσιες στα νουκλεοτίδια αδερίνη (A) και θυμίνη (T)
- Τα γονίδια βρίσκονται διάσπαρτα σε διάφορες περιοχές του γονιδιώματος με μεγάλα μη κωδικοποιήσιμα τμήματα ανάμεσά τους
- Τμήματα επαναλαμβανόμενων βάσεων γουανίνης (G) και κυτοσίνης (C) μέχρι και 30.000 βρίσκονται κοντά στις περιοχές που είναι πλούσιες σε γονίδια. Αυτές οι λεγόμενες νησίδες CpG πιστεύεται ότι βοηθούν στη ρύθμιση της γονιδιακής δραστηριότητας





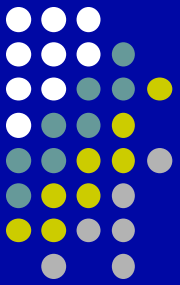
- **Λιγότερο από 1.5% του γονιδιώματος κωδικοποιεί για πρωτεΐνες**
- **Οι επαναληπτικές ακολουθίες (που δεν κωδικοποιούν για πρωτεΐνες) πιστεύεται ότι βοηθούν στη διαμόρφωση της αρχιτεκτονικής του γονιδιώματος με την ανακατανομή του, καθώς και στη δημιουργία νέων γονιδίων και τροποποίηση των ήδη υπαρχόντων**
- **Έχουν ανακαλυφθεί περισσότερα από 3 εκατομμύρια απλοί νουκλεοτιδικοί πολυμορφισμοί (SNPs). Αυτή η πληροφορία είναι πολύ σημαντική για την εύρεση ακολουθιών που σχετίζονται με ασθένειες και για τη μελέτη της ανθρώπινης συγγένειας και ιστορίας**





- Το χρωμόσωμα 1 έχει τα περισσότερα γονίδια (2968) και το χρωμόσωμα Y τα λιγότερα (231)
- Οι λειτουργίες για περισσότερο από 50% των γονιδίων είναι άγνωστες
- Τα χρωμοσώματα ποικίλλουν ευρέως στον αριθμό γονιδίων που περιέχουν
- υπάρχουν μόνο 300 μοναδικά γονίδια στο ανθρώπινο γονιδίωμα, τα οποία δεν είναι στο ποντίκι
- 113 γονίδια έχουν μεταφερθεί στο ανθρώπινο γονιδίωμα από τα βακτηρίδια

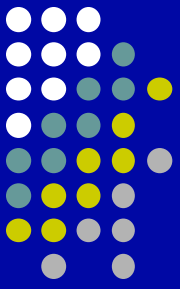




- το ψάρι *fuugu* έχει το πιο συνοπτικό γονιδίωμα. Δεν έχει «άχρηστο» (μη κωδικοποιήσιμο) DNA
- οι περισσότερες μεταλλάξεις εμφανίζονται στα αρσενικά
- υπάρχουν 250.000 πρωτεΐνες που γίνονται από τα περίπου 25.000 γονίδια
- το σκυλί είναι 85% ίδιο με έναν άνθρωπο από την άποψη της γενετικής ακολουθίας και πολλές από τις 380 κληρονομημένες ασθένειες στα σκυλιά είναι παρόμοιες με εκείνους στους ανθρώπους
- 1.778 γονίδια έχουν προσδιοριστεί με ασθένειες μέχρι τώρα, από το άσθμα μέχρι το Alzheimer



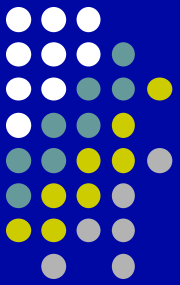
Μεταλλάξεις



- Μετάλλαξη χαρακτηρίζεται οποιαδήποτε αλλαγή στην ποσότητα ή την δομή του γενετικού υλικού ενός οργανισμού
- Οι αλλαγές αυτές είναι μόνιμες, κληρονομήσιμες και μπορεί να οδηγούν και σε αλλαγή του φαινοτύπου
- Οι μεταλλάξεις αφορούν μετατροπές της νουκλεοτιδικής αλληλουχίας ενός γονιδίου, ομάδας γονιδίων ή ολόκληρων χρωματωσώματων



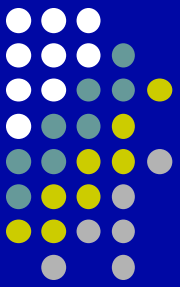
Τύποι μετάλλαξης



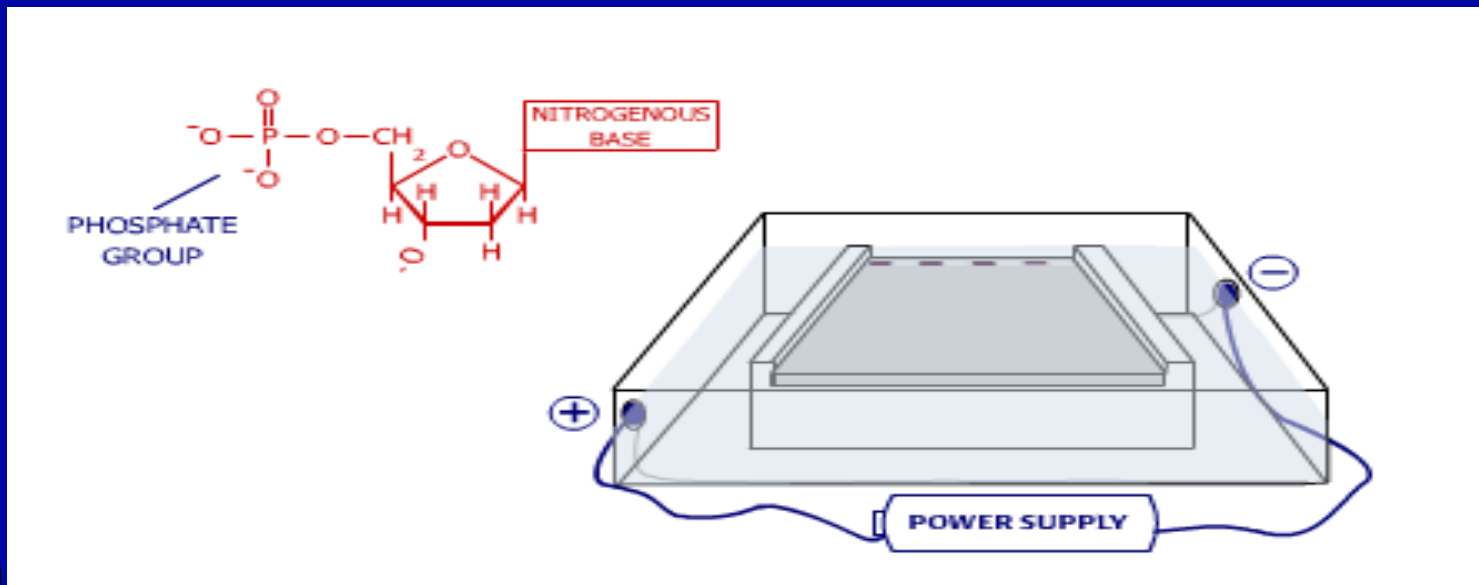
Τύπος μετάλλαξης	Περιγραφή
Σημειακή μετάλλαξη	Αλλαγή βάσης
Μετάλλαξη αφαίρεσης	Αφαίρεση μιας ή περισσότερων βάσεων
Μετάλλαξη παρεμβολής	Παρεμβολή μιας ή περισσότερων βάσεων
Μετάλλαξη αλλαγής πλαισίου ανάγνωσης (frameshift)	Αφαίρεση ή παρεμβολή ενός αριθμού βάσεων που δε μπορεί να διαιρεθεί με το 3
Μετάλλαξη αναστροφής	Αναστροφή μιας ακολουθίας βάσεων (μπορεί να προκαλέσει αλλαγή του πλαισίου ανάγνωσης)

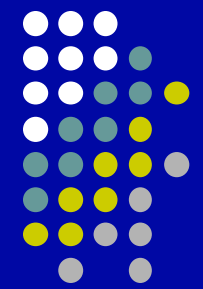


Η ηλεκτροφόρηση πηγμάτων αγαρόζης χρησιμοποιείται για την ανάλυση τμημάτων DNA

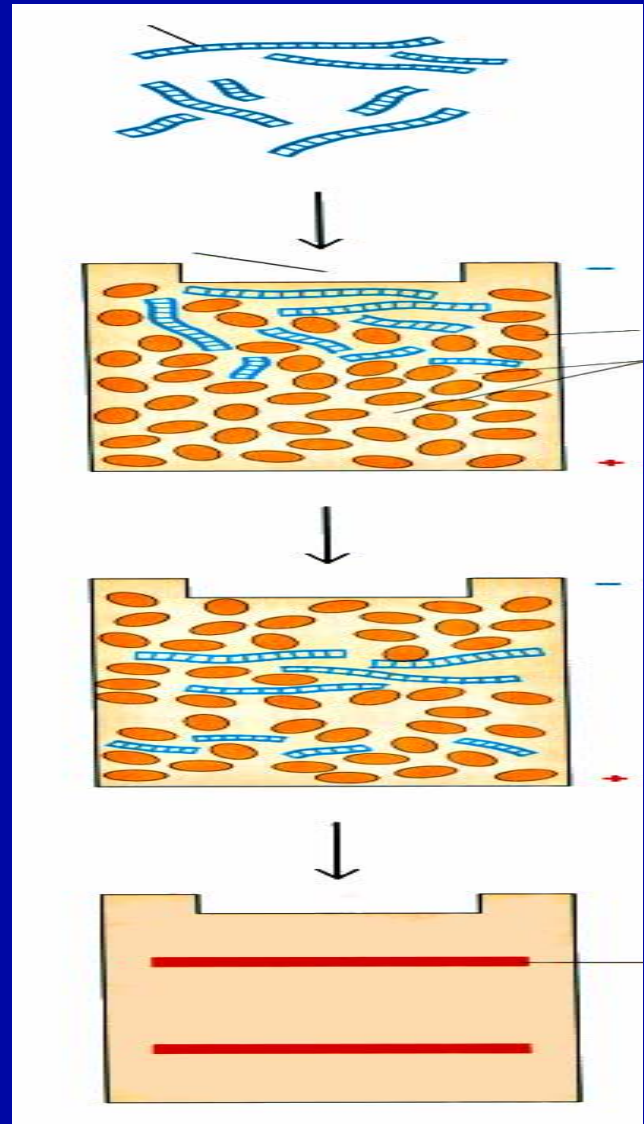
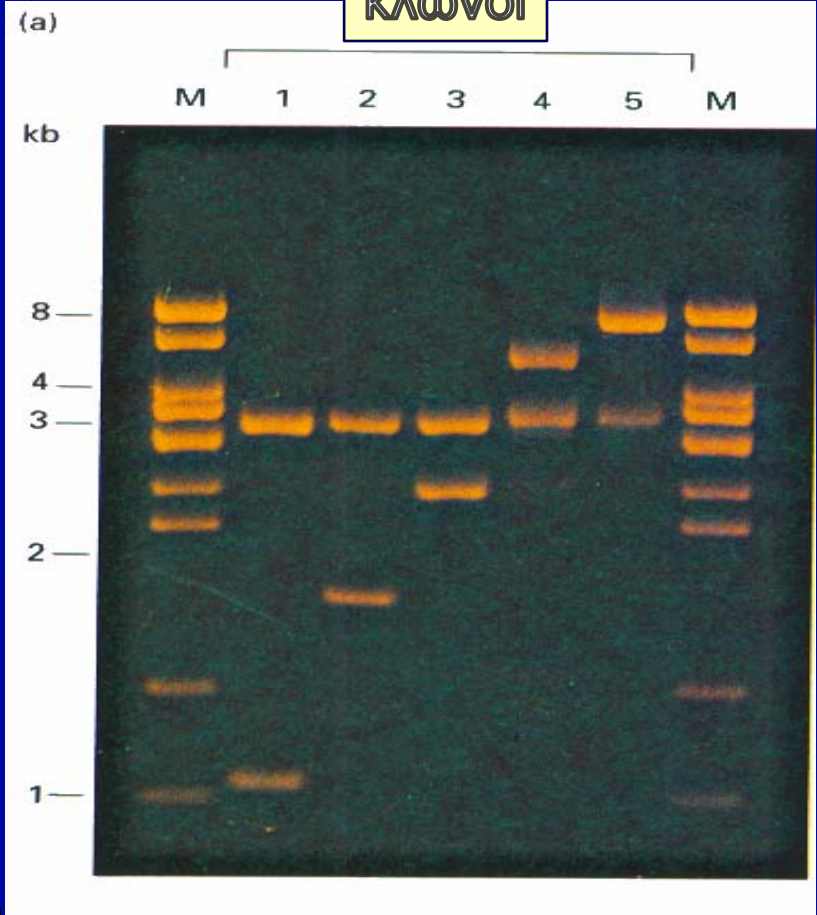


- Οι φωσφορικές ομάδες στο DNA περιλαμβάνουν αρνητικά φορτισμένα οξυγόνα δίνοντας στο μόριο του DNA ένα ολικό αρνητικό φορτίο. Σε ηλεκτρικό ρεύμα, το αρνητικά φορτισμένο DNA μετακινείται προς το θετικό πόλο σε μια συσκευή ηλεκτροφόρησης

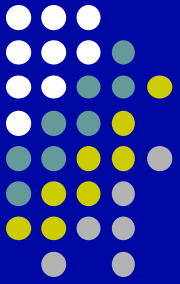




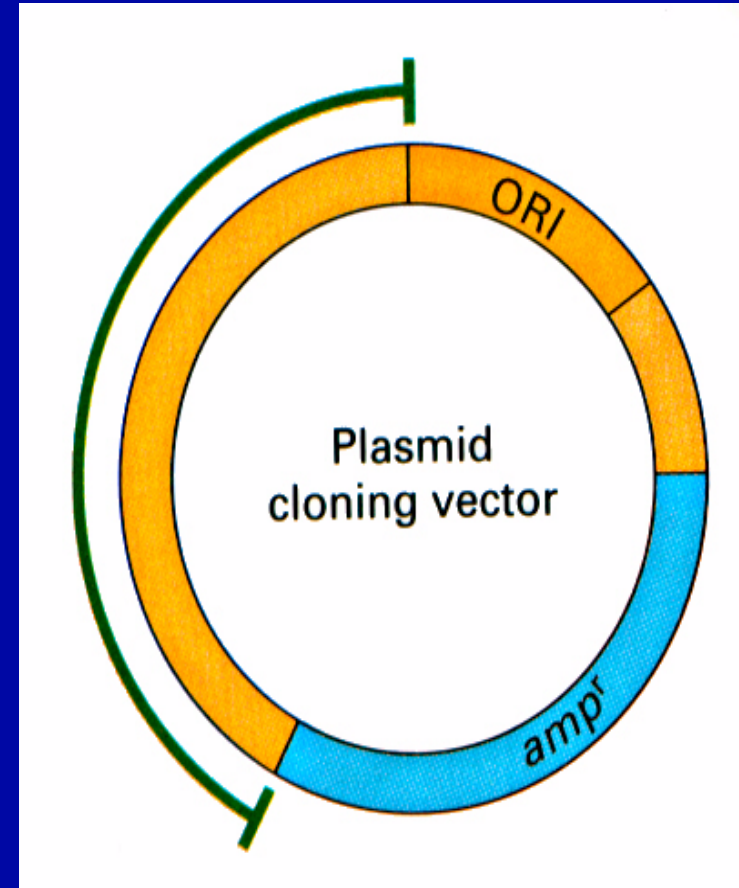
κλώνοι



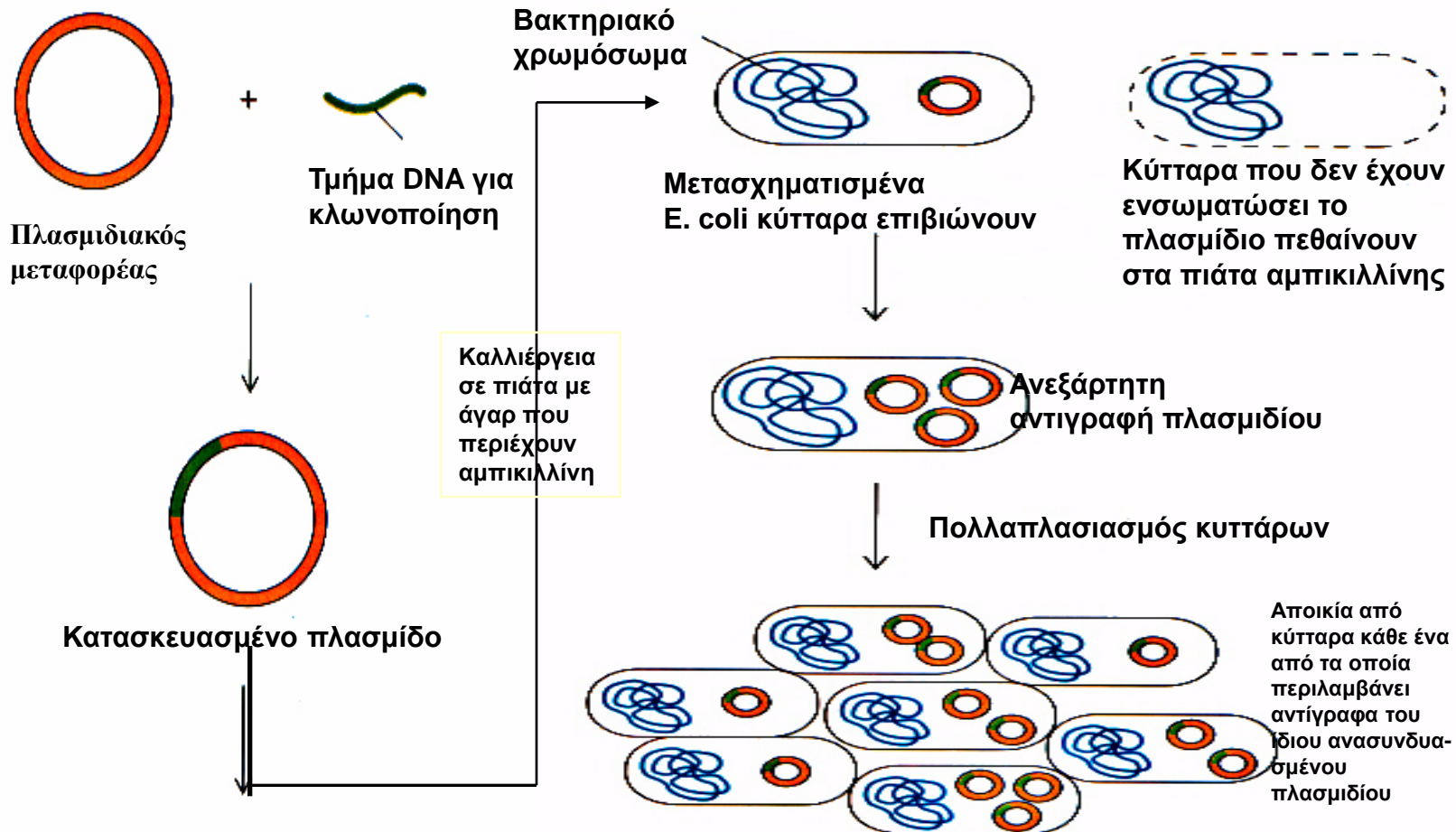
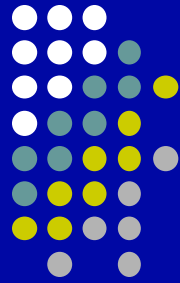
Κλωνοποίηση



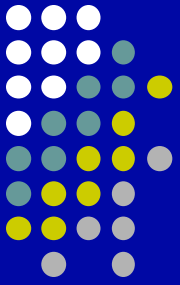
- Η ουσία της γενετικής μηχανικής έγκειται στην παρασκευή μεγάλων αριθμών ίδιων μορίων DNA προερχόμενων από ένα γονίδιο (κλωνοποίηση).
- Αυτό επιτυγχάνεται με την εισαγωγή της ακολουθίας ενός γονιδίου σε ένα μεταφορέα (vector) που μπορεί να είναι ένα πλασμίδιο ή ένας βακτηριοφάγος. Το εισερχόμενο DNA θα αντιγραφεί μαζί με το DNA του μεταφορέα.
- Το κατασκευασμένο πλασμίδιο (vector) DNA εισάγεται σε ένα βακτήριο με μια διαδικασία που λέγεται **μετασχηματισμός**.
- Το επόμενο βήμα είναι η επιλογή των βακτηρίων που έχουν ενσωματώσει ένα αντίγραφο του πλασμιδίου. Αυτό επιτυγχάνεται μέσα από ένα γονίδιο του πλασμιδίου που κωδικοποιεί για αντοχή στα αντιβιοτικά, με αποτέλεσμα κάθε βακτήριο που έχει ενσωματώσει το πλασμίδιο να είναι ανθεκτικό στο συγκεκριμένο αντιβιοτικό



Διαδικασία για την κλωνοποίηση ενός τμήματος DNA σε ένα πλασμιδιακό μεταφορέα



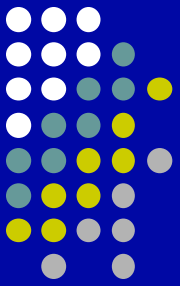
Η αλυσιδωτή αντίδραση της πολυμεράσης (PCR)



Η Αλυσιδωτή αντίδραση πολυμεράσης (ή **PCR – polymerase chain reaction**) είναι μία τεχνική για την απομόνωση και τον πολλαπλασιασμό αλληλουχίας **DNA**, μέσω της **ενζυμικής** αναπαραγωγής του **DNA** χωρίς τη χρήση ζωντανών μικροοργανισμών (όπως *E.coli* ή οι **ζύμες**). Η PCR είναι μία **in vitro** μέθοδος και μπορεί να πραγματοποιηθεί χωρίς περιορισμούς στη μορφή του χρησιμοποιούμενου DNA, και μπορεί να διαφοροποιηθεί εκτενώς για την πραγματοποίηση πολλών μεθόδων γενετικής επέμβασης. Συγκεκριμένα DNA θραύσματα μπορούν να κλωνοποιηθούν σε ένα δοκιμαστικό σωλήνα (απουσία ζωντανών κυττάρων) με τη χρήση της αλυσιδωτής αντίδρασης πολυμεράσης (PCR).



Η αλυσιδωτή αντίδραση της πολυμεράσης (PCR)



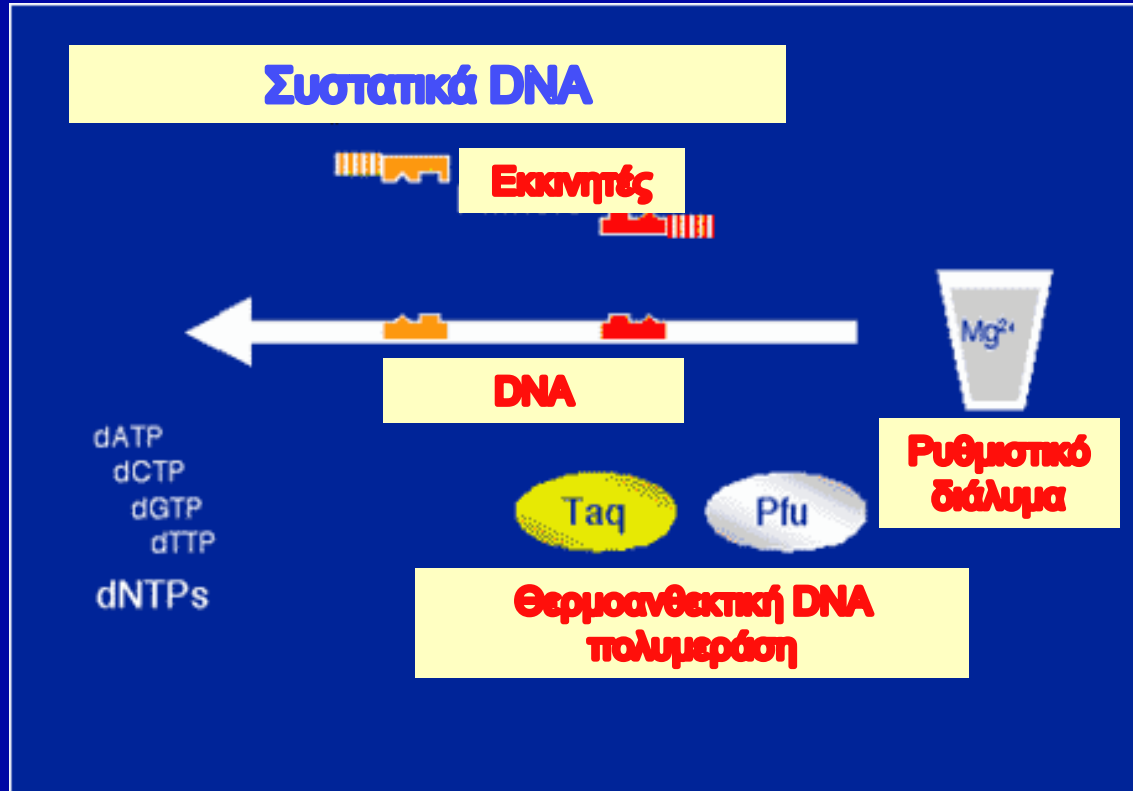
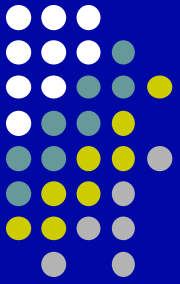
Με την PCR μια συγκεκριμένη περιοχή του γονιδιώματος μπορεί να πολλαπλασιαστεί μέχρι και δισεκατομμύρια φορές, δεδομένου ότι είναι γνωστή η **νουκλεοτιδική του αλληλουχία**.

Ένας πλήρης κύκλος μιας PCR αντίδρασης περιλαμβάνει τρία στάδια:

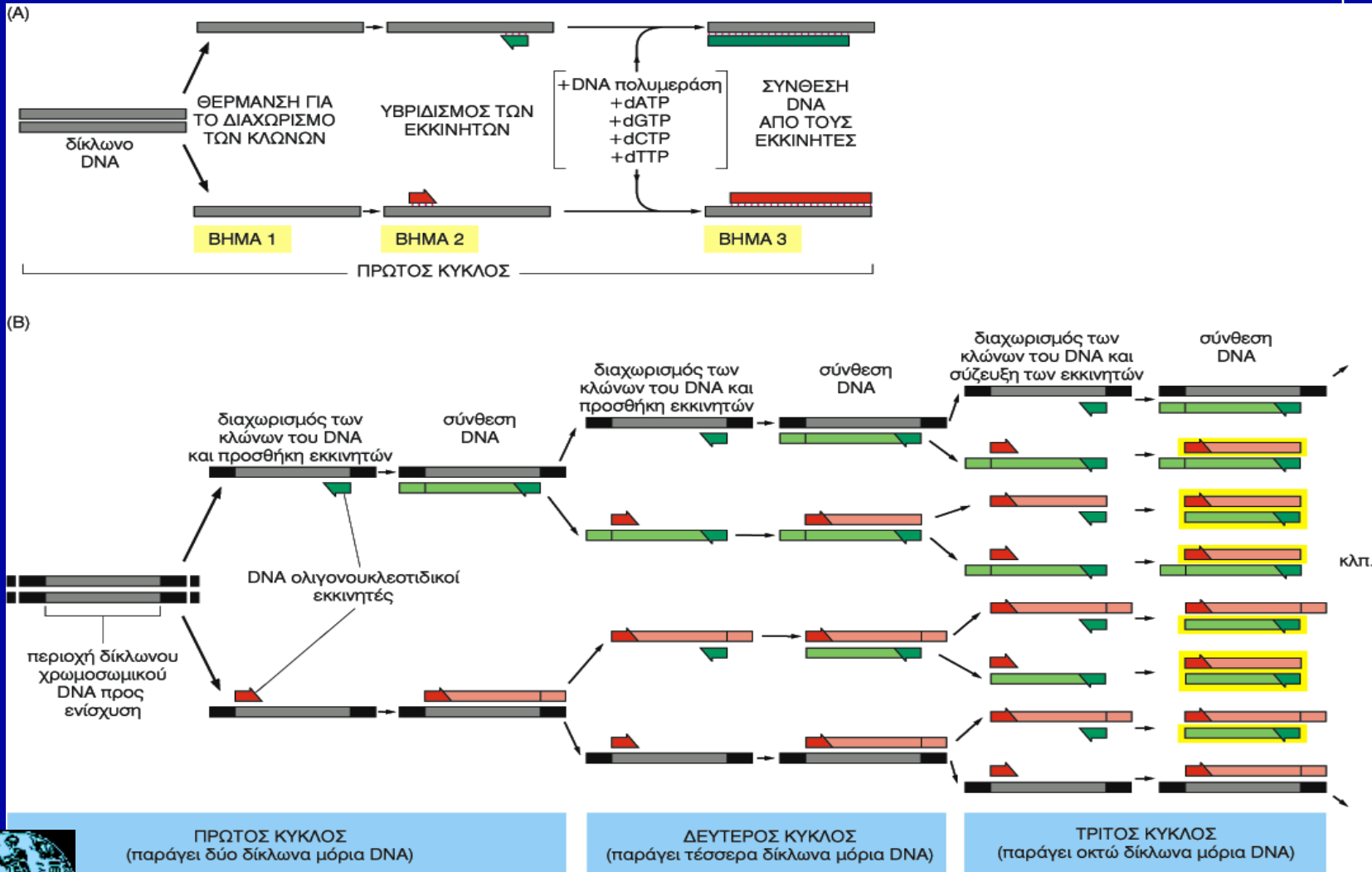
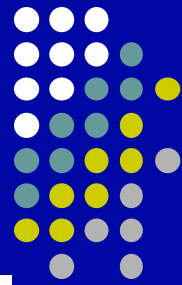
- Αποδιάταξη του DNA (denaturation)
- Προσαρμογή των εκκινητών (primers) στο DNA εκμαγείο (annealing)
- Επιμήκυνση των εκκινητών (extension).



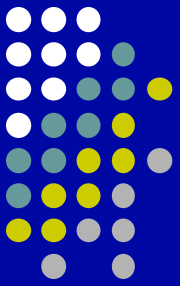
Συστατικά PCR



Η αλυσιδωτή αντίδραση πολυμεράσης (PCR)



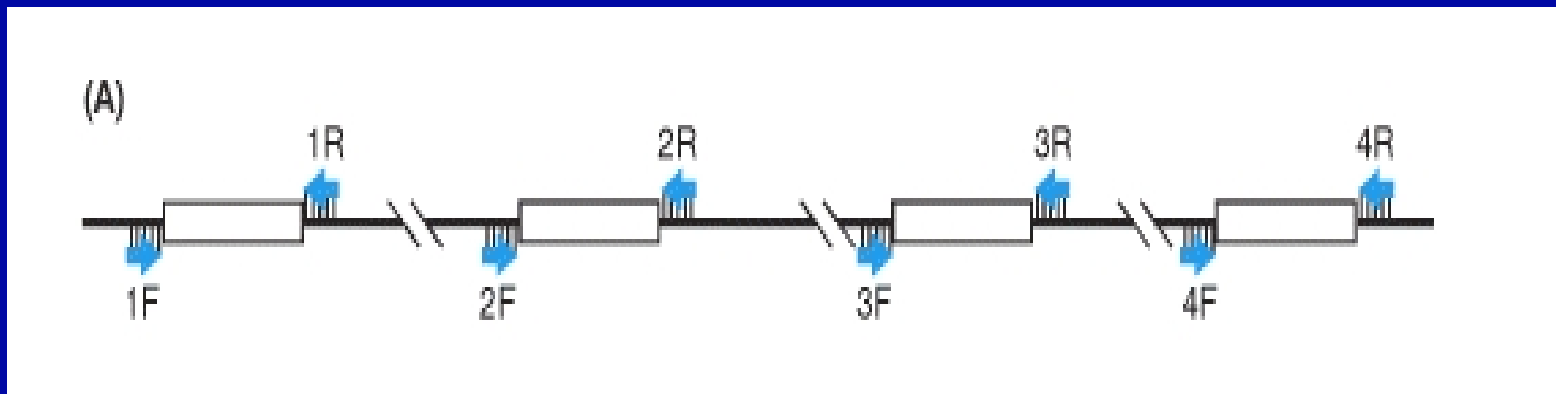
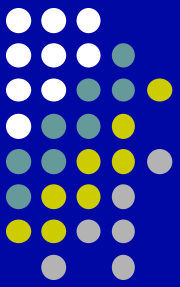
Εφαρμογές της PCR



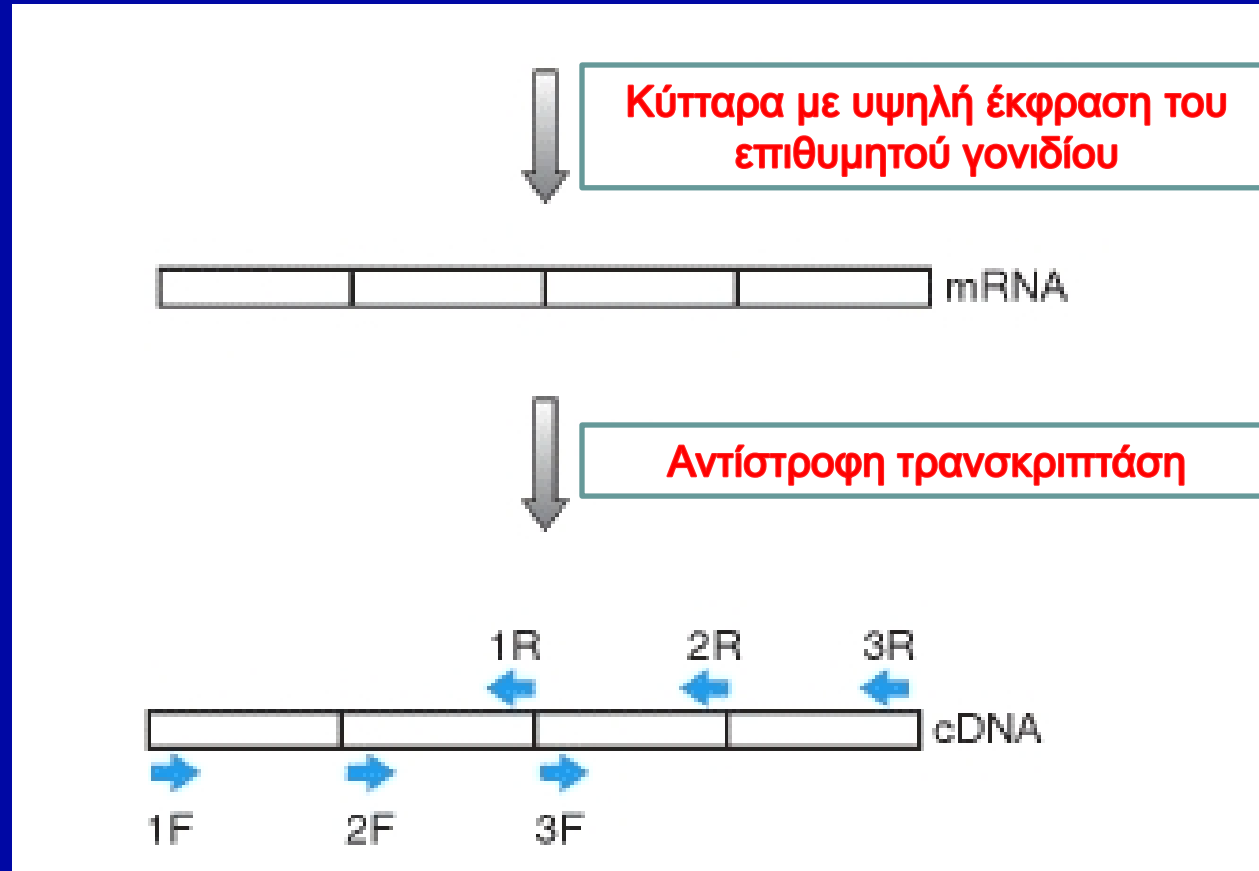
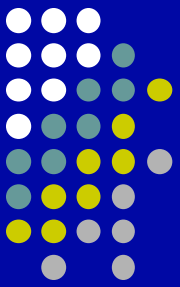
1. Ιατρική διάγνωση γενετικών ασθενειών
 - Προσδιορισμός μεταλλάξεων σε αλληλόμορφα γονίδια
 - Διάγνωση της παρουσίας μολυσματικών συστατικών
 - Επιδημιολογικές μελέτες
2. Ιατροδικαστικές εφαρμογές
 - Τεστ πατρότητας
 - Καταχώρηση του DNA για προσδιορισμό του ατόμου
 - Εγκληματολογικές έρευνες
3. Χειρισμός του DNA για πειράματα γενετικής μηχανικής
4. Προσδιορισμός της αλληλουχίας του DNA
5. Απομόνωση καινούριων γονιδίων
6. Ανθρωπολογικές μελέτες
 - Γενετική πληθυσμών
 - Μελέτες μετανάστευσης
- 7. Εξελικτικές σχέσεις



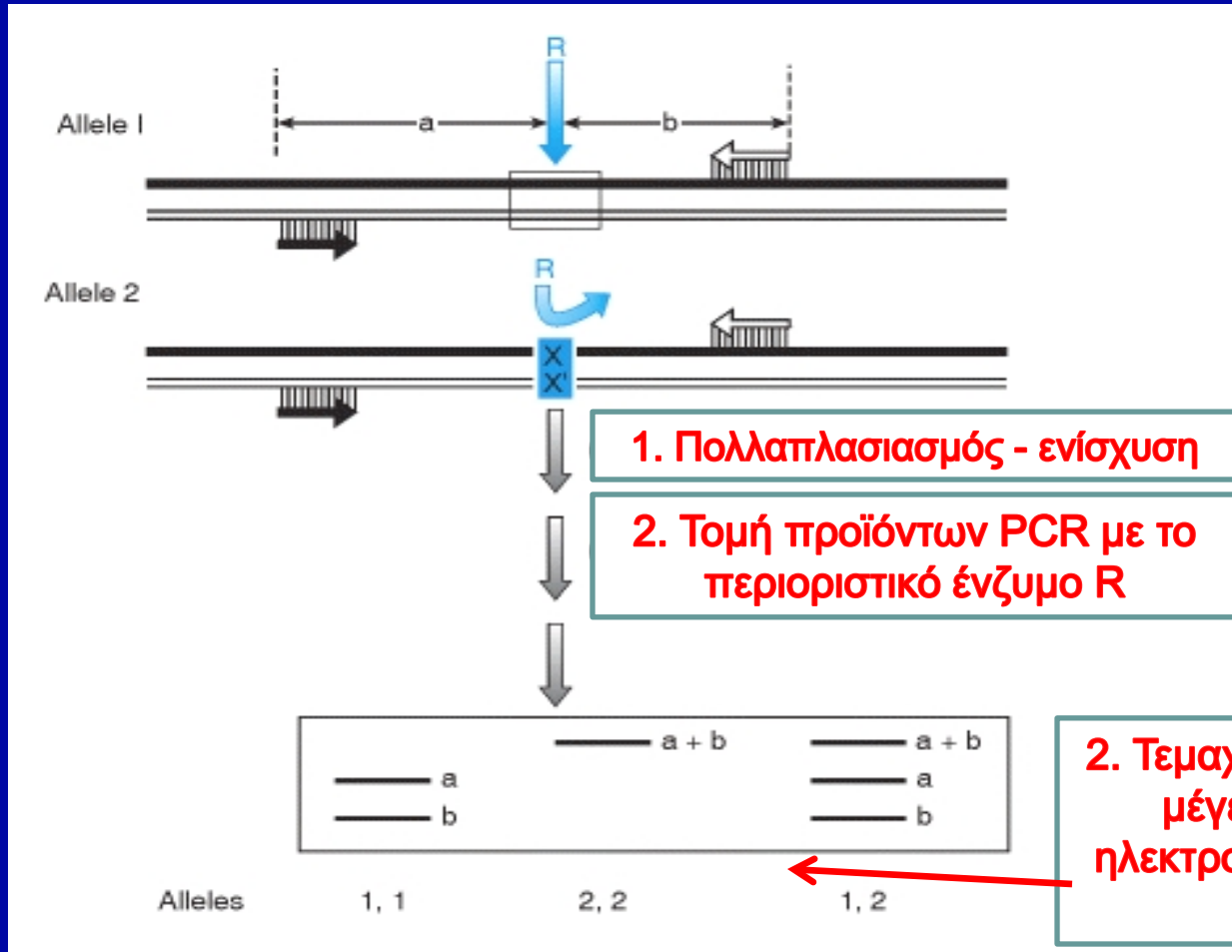
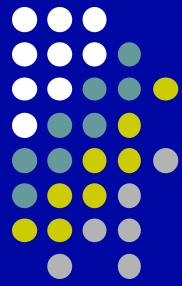
A. Ανίχνευση μεταλλάξεων



Β. Ανίχνευση μεταλλάξεων, RT-PCR



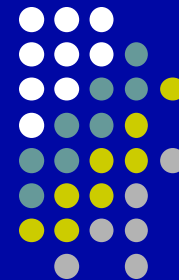
Γ. Προσδιορισμός πολυμορφισμών τμημάτων που λαμβάνονται με περιοριστικά ένζυμα(RFLPs)



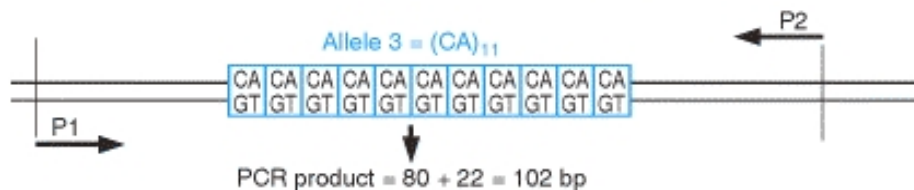
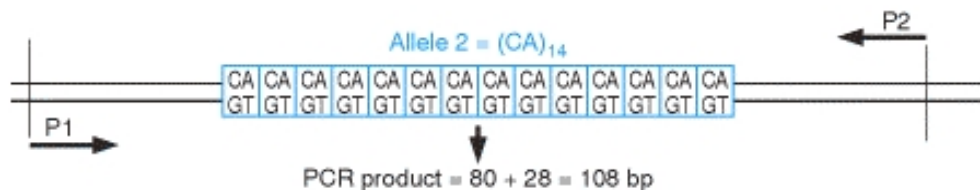
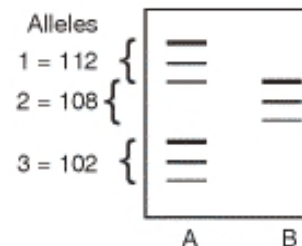
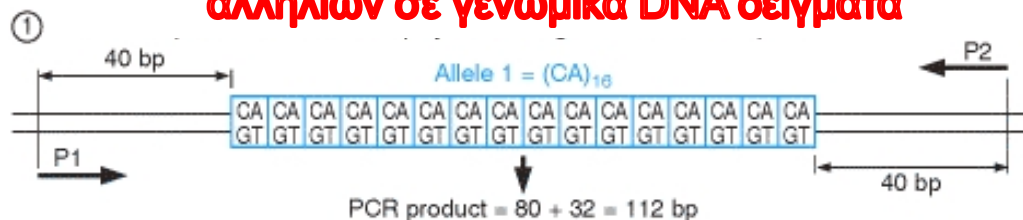
2. Τεμαχισμός ανάλογα με το μέγεθος με τη χρήση ηλεκτροφόρησης πηγμάτων αгарόζης



Δ. Προσδιορισμός πολυμορφισμών μικρών νουκλεοτιδικών επαναλήψεων (STRs)



Χρήση PCR εκκινήτων για την ενίσχυση αλληλίων σε γενωμικά DNA δείγματα

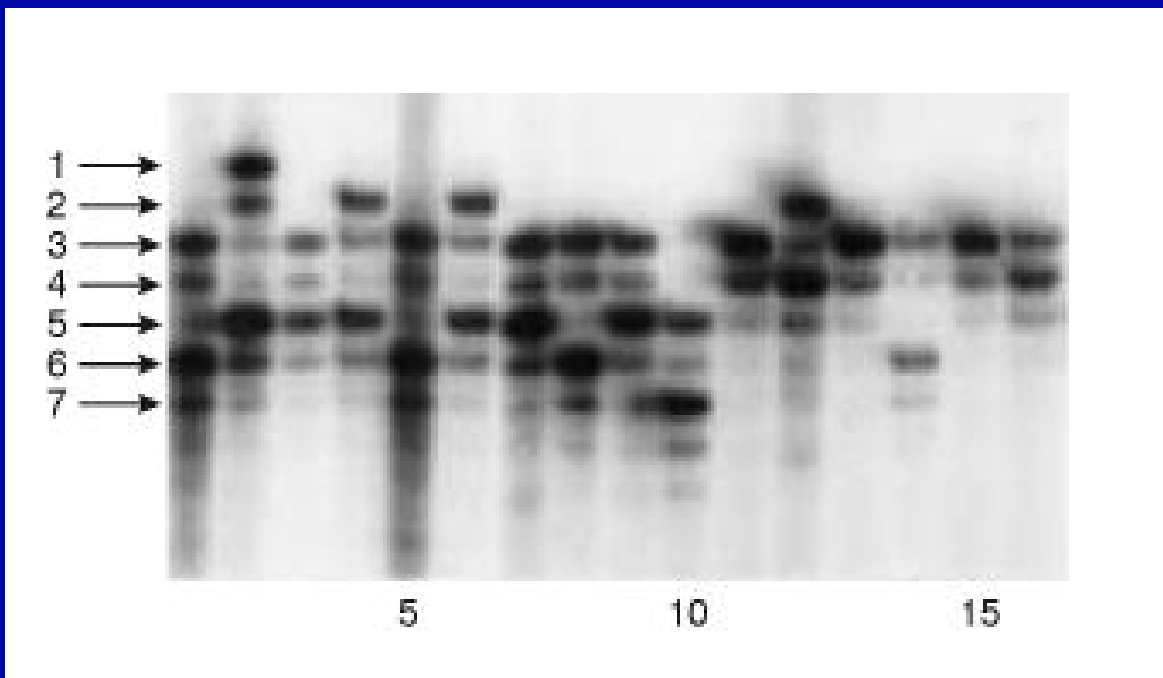
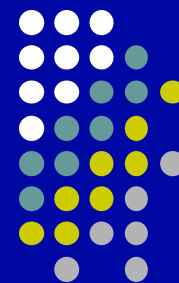


② Denature PCR products and size-fractionate by polyacrylamide gel electrophoresis

③ Autoradiography



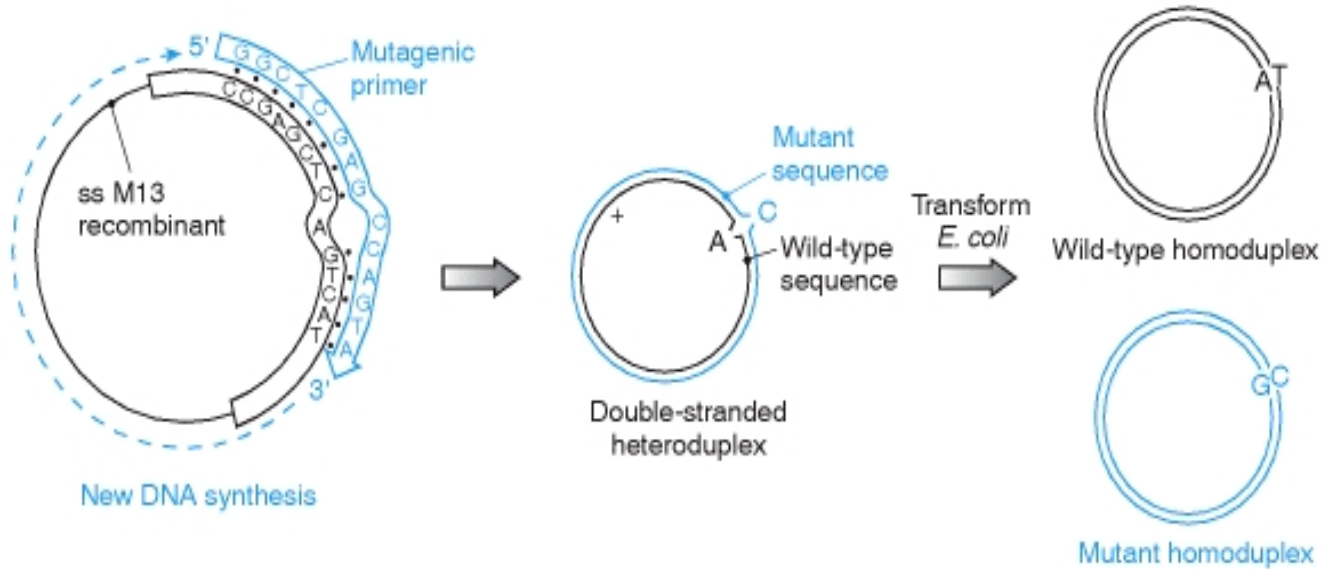
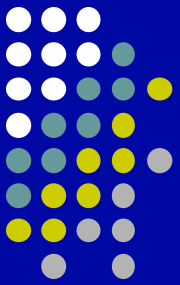
Δ. Προσδιορισμός πολυμορφισμών μικρών νουκλεοτιδικών επαναλήψεων (STRs)



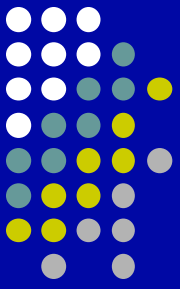
1 (3,6); 2 (1,5); 3 (3,5); 4 (2,5); 5 (3,6); 6 (2,5);
7 (3,5); 8 (3,6); 9 (3,5); 10 (5,7); 11 (3,3); 12
(2,4); 13 (3,3); 14 (3,6); 15 (3,3); 16 (3,4)



Δημιουργία απλών μεταλλάξεων



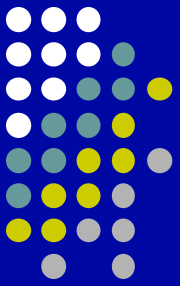
Βιοπληροφορική



Η επιστήμη που γεννάται και αναπτύσσεται από τη σύγκλιση της Βιοτεχνολογίας και της Πληροφορικής και στοχεύει στο σχεδιασμό εργαλείων λογισμικού (software tools), με σκοπό τη διευκόλυνση και την επιτάχυνση της βιολογικής της ιατρικής και της φαρμακευτικής έρευνας



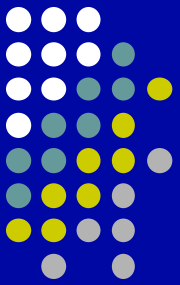
Εφαρμογές της βιοπληροφορικής



- **Ανεύρεση λειτουργίας πρωτεϊνών, ανεύρεση αλληλεπιδράσεων πρωτεϊνών μεταξύ τους, κατανόηση της πολυπλοκότητας των βιολογικών συστημάτων**
- **Η σύγκριση του γονιδιώματος διαφόρων ειδών επιτρέπει την εξαγωγή πολύτιμων συμπερασμάτων σχετικά με τις εξελικτικές σχέσεις των οργανισμών μεταξύ τους**
- **Προσπάθεια αντιμετώπισης διαφόρων ασθενειών με την ανάπτυξη νέων διαγνωστικών μέτρων και θεραπευτικών μεθόδων**
- **Γονιδιακή θεραπεία**
- **Γενετική τροποποίηση φυτών και ζώων**

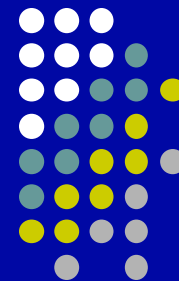


ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ



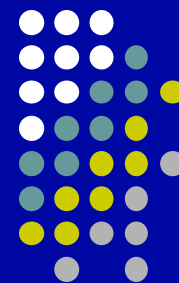
Νουκλεοτιδική ακολουθία	GenBank	Η μεγαλύτερη δημόσια βάση δεδομένων ακολουθίας
	DDBJ	DNA τράπεζα δεδομένων της Ιαπωνίας
	EMBL	Ευρωπαϊκό εργαστήριο Μοριακής Βιολογίας
	MGDB	Βάση δεδομένων γονιδιώματος ποντικών
	GSX	Βάση δεδομένων γονιδιακής έκφρασης ποντικών
	NDB	Βάση δεδομένων νουκλεϊνικών οξέων





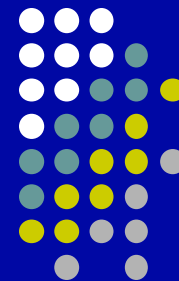
Πρωτεϊνική ακολουθία	SWISS-PROT	Ελβετικό ίδρυμα για τη βιοπληροφορική και ευρωπαϊκό ίδρυμα βιοπληροφορικής
	TrEMBL	Συμπλήρωμα της SWISS-PROT με σχόλια
	TrEMBLnew	Εβδομαδιαία, προεπεξεργασμένη αναπροσαρμογή σε TrEMBL
	PIR	Πρωτεϊνική πηγή πληροφοριών
Τρισδιάστατες δομές	PDB	Πρωτεϊνική τράπεζα δεδομένων
	MMDB	Βάση δεδομένων μοριακής διαμόρφωσης





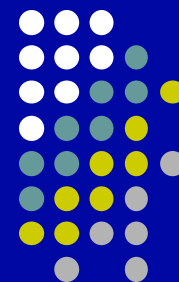
Ένζυμα και ενώσεις	LIGAND	Χημικές ενώσεις και αντιδράσεις
Μοτίβα ακολουθίας (Ευθυγράμμιση)	PROSITE	Μοτίβα ακολουθίας
	BLOCKS	Προερχόμενη από την PROSITE
	PRINTS	Προερχόμενη από την BLOCKS
	Pfam	Πρωτεϊνική βάση δεδομένων των ευθυγραμμίσεων και των κρυμμένων μοντέλων Markov
	ProDOM	Πρωτεϊνικές περιοχές





Οδοί και συμπλέγματα	Pathway	Μεταβολικοί και ρυθμιστικοί χάρτες διαφόρων οδών
Μοριακή ασθένεια	OMIM	Μεντελική κληρονομιά στον άνθρωπο
Βιοϊατρική βιβλιογραφία	PubMed	Περιλαμβάνει τη MEDLINE
	Medline	Ιατρική βιβλιογραφία
Άνυσμα μεταφοράς (vector)	UniVec	Χρησιμοποιείται για να προσδιορίσει τη μόλυνση των ανυσμάτων μεταφοράς
Πρωτεϊνικές μεταλλάξεις	PMD	Βάση δεδομένων των πρωτεϊνικών μεταλλάξεων

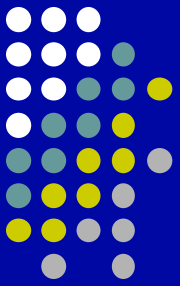




Εκφράσεις γονιδίων	GEO	Καθολική έκφραση γονιδίων
Δείκτες αμινοξέος	Aaindex	Βάση δεδομένων δεικτών αμινοξέος
Πρωτεϊνική/πεπτιδική βιβλιογραφία	LITDB	Βιβλιογραφική βάση δεδομένων για τις πρωτεΐνες και τα πεπτίδια
Κατάλογος γονιδίων	GENES	Γονιδιακή βάση δεδομένων της KEGG

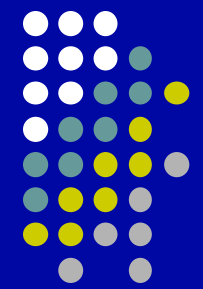


Ευκαρυωτική γενωμική εισαγωγή στη GenBank



- Εισάγουμε τη διεύθυνση www.ncbi.nlm.nih.gov/entrez/
- Διαλέγουμε nucleotide από το μενού του search
- Πληκτρογραφούμε το GenBank ID AF018430 στο πεδίο for και μετά πατάμε το Go
Μια σελίδα αποτελεσμάτων προκύπτει η οποία παρουσιάζει ένα σύντομο ορισμό για αυτή τη νουκλεοτιδική ακολουθία
- Πατάμε το X01714 υπερσύνδεσμο





Entrez Nucleotide

EGCTCAGGATATGACTTCGTCGCTAGGATCGGATC CCGGGATTATATAGCTCGATCGATCT
 TTCTCTATATCCGCGGATGGGATATACACACACATCCCGCGGATAGCATGACTGATCTA
 CCCCACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTA
 CACAGAGACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTA

[All Databases](#) | [PubMed](#) | [Nucleotide](#) | [Protein](#) | [Genome](#) | [Structure](#) | [OMIM](#)

Search for

Limits

Format:

- Nucleotide
- GSS
- EST
- Structure
- Genome
- BioSystems
- Books
- CancerChromosomes
- Conserved Domains
- dbGaP
- 3D Domains
- Gene
- Genome Project
- GENSAT
- GEO Profiles
- GEO DataSets
- HomoloGene
- Journals
- MeSH
- NCBI Web Site
- NLM Catalog
- OMIA
- OMIM
- Peptidome
- PMC
- PopSet
- Probe
- Protein Clusters
- PubChem BioAssay
- PubChem Compound

[History](#) | [Clipboard](#) | [Details](#)

[Graphics](#) | [More Formats](#) ▼

Pase (DUT) gene, exon 3

1177 bp
DNA
linear
PRI 28-SEP-

ns dUTPase (DUT) gene, exon 3.

GI:2443576

ns (human)

[ns](#)
Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Hominidae; Homo.

1 to 1177)
.E.

mic nuclear and mitochondria dUTPase gene

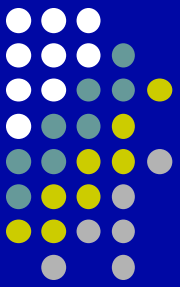
d
1 to 1177)
.E.

mission
(11-AUG-1997) Biology, York University, 4700 Keele St
, ONT M3J 1P3, Canada

[Location/Qualifiers](#) ▼



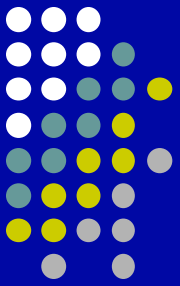
Επεξήγηση των λέξεων-κλειδιά σε μια **ευκαρυωτική** GenBank εισαγωγή



- **LOCUS** – το όνομα της περιοχής είναι HSDUT2. Η υπόλοιπη πληροφορία μας λέει ότι πρόκειται για γραμμικό DNA με μήκος 1177 νουκλεοτιδικές βάσεις
- **DEFINITION** – αυτή η γραμμή δείχνει ότι το όνομα του γονιδίου είναι DUT και ότι η εισαγωγή περιλαμβάνει το εξώνιο 3 του γονιδίου
- Οι **ACCESSION**, **VERSION** και **KEYWORDS** γραμμές παρέχουν τις συνηθισμένες πληροφορίες. Για το AF018430 το **KEYWORDS** είναι κενό που σημαίνει ότι δε μπορεί κάποιος με λέξεις-κλειδιά να αναζητήσει την εισαγωγή σε άλλες βάσεις δεδομένων

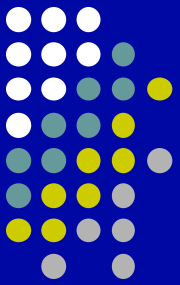


Επεξήγηση των λέξεων-κλειδιά σε μια **ευκαρυωτική** GenBank εισαγωγή



- **SEGMENT** – αυτό το τμήμα σχετίζεται με τη μωσαϊκή δομή των ευκαρυωτικών γονιδίων. Δείχνει ότι η τωρινή GenBank εισαγωγή είναι το δεύτερο τμήμα μιας εισαγωγής που αποτελείται από τέσσερα τμήματα. Χρειάζονται και οι τέσσερις εισαγωγές για να μπορέσει κάποιος να ξαναχτίσει όλη την mRNA ακολουθία
- Τα **ORGANISM**, **SOURCE** και **REFERENCE** παρέχουν τις ίδιες πληροφορίες όπως και στο προκαρυωτικό γονίδιο





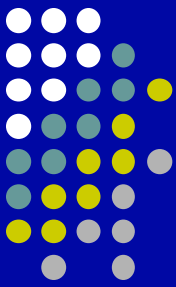
```
FEATURES             Location/Qualifiers
    source             1..1177
                       /organism="Homo sapiens"
                       /mol_type="genomic DNA"
                       /db_xref="taxon:9606"
                       /map="15q15-q21.1"
    gene                 order(AF018429.1:<1..1735,1..1177,AF018431.1:1..45,
                       AF018432.1:658..732,AF018432.1:884..954,
                       AF018432.1:1391..>1447)
                       /gene="DUT"
    mRNA                join(AF018429.1:<282..561,AF018429.1:1034..1172,
                       560..651,AF018431.1:1..45,AF018432.1:658..732,
                       AF018432.1:884..954,AF018432.1:1391..>1447)
                       /gene="DUT"
                       /product="dUTPase"
                       /note="alternatively spliced; encodes mitochondrial form
of the protein"
    CDS                 join(AF018429.1:282..561,AF018429.1:1034..1172,560..651,
                       AF018431.1:1..45,AF018432.1:658..732,AF018432.1:884..954,
                       AF018432.1:1391..1447)
                       /gene="DUT"
                       /note="DUT-M; alternatively spliced; mitochondrial form of
the protein; similar to H. sapiens dUTPase encoded by
GenBank Accession Number U90224"
                       /codon_start=1
                       /product="dUTPase"
                       /protein_id="AAB71393.1"
                       /db_xref="GI:2443580"
                       /translation="MTPLCPRPALCYHFLTSLLRSAMQNARGTAEGRSRGTLRARPAP
RPPAAQHGIIPRPLSSAGRLSQGCRGASTVGAAGWKGELPKAGGSPAPGPETPAISPSK
RARPAEVGGMQLRFARLSEHATAPTRGSARAAGYDLYSAYDYTIIPMEKAVVKTDIQI
ALPSGCYGRVAPRSGLAAKHFIDVGAGVIDEDYRGNVGVVLFNFGKEKFEVKKGDRIA
QLICERIFYPEIEEVQALDDTERGSGGFGSTGKN"
    mRNA                join(AF018429.1:<1018..1172,560..651,AF018431.1:1..45,
                       AF018432.1:658..732,AF018432.1:884..954,
                       AF018432.1:1391..>1447)
                       /gene="DUT"
                       /product="dUTPase"
                       /note="alternatively spliced; encodes nuclear form of the
protein"
    CDS                 join(AF018429.1:1018..1172,560..651,AF018431.1:1..45,
                       AF018432.1:658..732,AF018432.1:884..954,
                       AF018432.1:1391..1447)
                       /gene="DUT"
```



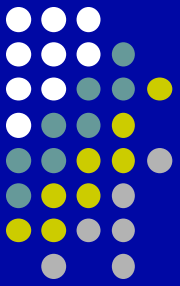
Επεξήγηση των λέξεων-κλειδιά σε μια **ευκαρυωτική** GenBank εισαγωγή

ΠΙΝΑΚΑΣ “FEATURES”

- Το **source** τμήμα δείχνει για το AF018430 ότι βρίσκεται στο χρωμόσωμα 15 και πιο συγκεκριμένα στην q21.1 κυτταρογενετική περιοχή
- Η λέξη **gene** έχει σκοπό να περιγράψει λεπτομερώς την ανακατασκευή διαφόρων mRNAs που βρίσκονται σε διαφορετικές εισαγωγές
- Η λέξη **mRNA** σχετίζεται με τον τρόπο με τον οποίο η GenBank αναπαριστάνει το εναλλακτικό μάτισμα
- **Exon** – παρέχει πληροφορίες για την ακριβή θέση του (μοναδικού στο συγκεκριμένο παράδειγμα) εξωνίου



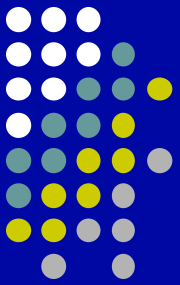
Χρησιμοποιώντας τη βάση δεδομένων για εξαγωγή γονιδιακών πληροφοριών



- Εισάγουμε τη διεύθυνση www.ncbi.nlm.nih.gov/entrez/
- Από το μενού του search επιλέγουμε Gene
- Πληκτρογραφούμε DUT [gene] human [organism] στο πεδίο for και μετά πατάμε Go



Η εισαγωγική σελίδα του Entrez gene για το ανθρώπινο γονίδιο DUT



NCBI Entrez Gene

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search Gene for DUT[gene] human[organism] Go Clear [Save Search](#)

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Sort by Relevance Send to

All: 1 Current Only: 1 Genes Genomes: 1 SNP GeneView: 1

1: [DUT](#) Order cDNA

Official Symbol DUT and Name: deoxyuridine triphosphatase [*Homo sapiens*]
Other Aliases: FLJ20622, dUTPase
Other Designations: dUTP nucleotidohydrolase; dUTP pyrophosphatase; deoxyuridine 5'-triphosphate nucleotidohydrolase
Chromosome: 15; **Location:** 15q15-q21.1
Annotation: Chromosome 15, NC_000015.9 (48623621..48635570)
MIM: 601266
GeneID: 1854



Πληροφορίες της εισαγωγής για το ανθρώπινο γονίδιο DUT



Entrez Gene



All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search Gene for DUT[gene] human[organism] Go Clear Save Search

Limits Preview/Index History Clipboard Details

Display Full Report Show 20 Sort by Relevance Send to

All: 1 Current Only: 1 Genes Genomes: 1 SNP GeneView: 1

1: DUT deoxyuridine triphosphatase [*Homo sapiens*]

GeneID: 1854

updated 24-Jun-2009

Summary

Official Symbol DUT

provided by [HGNC](#)

Official Full Name deoxyuridine triphosphatase

provided by [HGNC](#)

Primary source [HGNC:3078](#)

See related [Ensembl:ENSG00000128951](#); [HPRD:03165](#); [MIM:601266](#)

Gene type protein coding

RefSeq status REVIEWED

Organism [Homo sapiens](#)

Lineage *Eukaryota*; *Metazoa*; *Chordata*; *Craniata*; *Vertebrata*; *Euteleostomi*; *Mammalia*; *Eutheria*; *Euarchontoglires*; *Primates*; *Haplorrhini*; *Catarrhini*; *Hominidae*; *Homo*

Also known as dUTPase; FLJ20622; DUT

Summary This gene encodes an essential enzyme of nucleotide metabolism. The encoded protein forms a ubiquitous, homotetrameric enzyme that hydrolyzes dUTP to dUMP and pyrophosphate. This reaction serves two cellular purposes: providing a precursor (dUMP) for the synthesis of thymine nucleotides needed for DNA replication, and limiting intracellular pools of dUTP. Elevated levels of dUTP lead to increased incorporation of uracil into DNA, which induces extensive excision repair mediated by uracil glycosylase. This repair process, resulting in the removal and reincorporation of dUTP, is self-defeating and leads to DNA fragmentation and cell death. Alternative splicing of this gene leads to different isoforms that localize to either the mitochondrion or nucleus. A related pseudogene is located on chromosome 19. [provided by RefSeq]

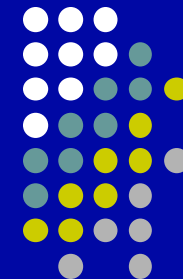
[Entrez Gene Home](#)

Table Of Contents

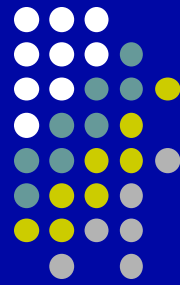
- Summary
- Genomic regions, transcripts...
- Genomic context
- Bibliography
- Interactions
- General gene information
- General protein information
- Reference Sequences
- Related Sequences
- Additional Links

Links

- Order cDNA clone
- BioSystems
- CCDS
- Conserved Domains
- Genome
- GEO Profiles
- HomoloGene
- Map Viewer
- Nucleotide
- EST
- OMIM
- PubChem Compound



Πατήστε εδώ για λεπτομερή χάρτη και πειραματικές αποδείξεις

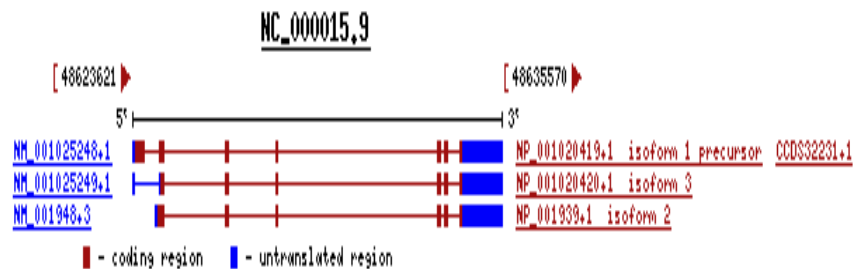


Genomic regions, transcripts, and products



Go to [reference sequence details](#)

[Try our new Sequence Viewer](#)

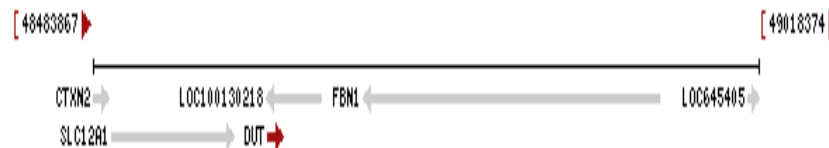


Genomic context

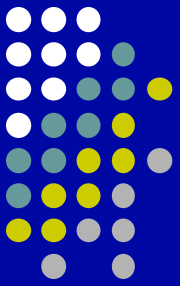


chromosome: 15; Location: 15q15-q21.1

[See DUT in MapViewer](#)



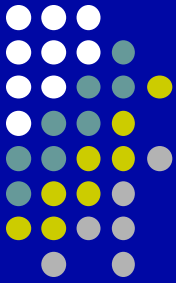
Το πρόγραμμα Ensembl



- Εισάγουμε τη διεύθυνση www.ensembl.org/
- Πατάμε το σύνδεσμο human όπου μας παρουσιάζεται μια σχηματική περιγραφή όλων των χρωμοσωμάτων αριθμημένα από το μεγαλύτερο προς το μικρότερο
- Μπορούμε να επιλέξουμε οποιονδήποτε οργανισμό καθώς η σελίδα είναι έτσι διαμορφωμένη ώστε πατώντας οποιοδήποτε σύνδεσμο ανοίγει ένα καινούριο παράθυρο που σχετίζεται με αυτό το σύνδεσμο
- Για το συγκεκριμένο παράδειγμα επιλέγουμε το χρωμόσωμα 15
- Επιλέγουμε Gene από το pull down menu
- Εισάγουμε DUT στο πεδίο search και μετά πατάμε το κόκκινο Go πεδίο



Το πρόγραμμα Ensembl






Search Ensembl

Search:
e.g. human gene

Browse a Genome

The Ensembl project produces genome data for other eukaryotic species, and makes this information freely available online.
Click on a link below to go to the species' home page.

Popular genomes ([Log in to customize](#))

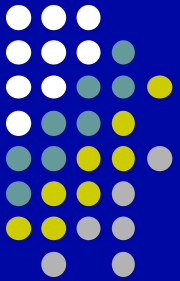
-  **Human**
NCBI36
-  **Mouse**
NCBIM37
-  **Zebrafish**
Zv8




















































Species List:

- All species
-
- Aedes
- Alpaca
- Anole Lizard
- Anopheles
- Armadillo
- Bushbaby
- C.elegans
- C.intestinalis
- C.savignyi
- Cat
- Chicken
- Chimpanzee
- Cow
- Dog
- Dolphin
- Elephant
- Fruitfly
- Fugu
- Gorilla
- Guinea Pig
- Hedgehog
- Horse
- Human
- Hyrax
- Kangaroo rat
- Lesser hedgehog tenrec
- Macaque
- Medaka



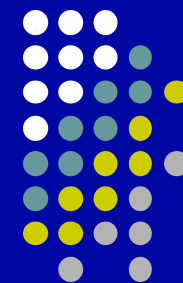
Όλοι οι οργανισμοί που είναι διαθέσιμοι Από το πρόγραμμα Ensembl

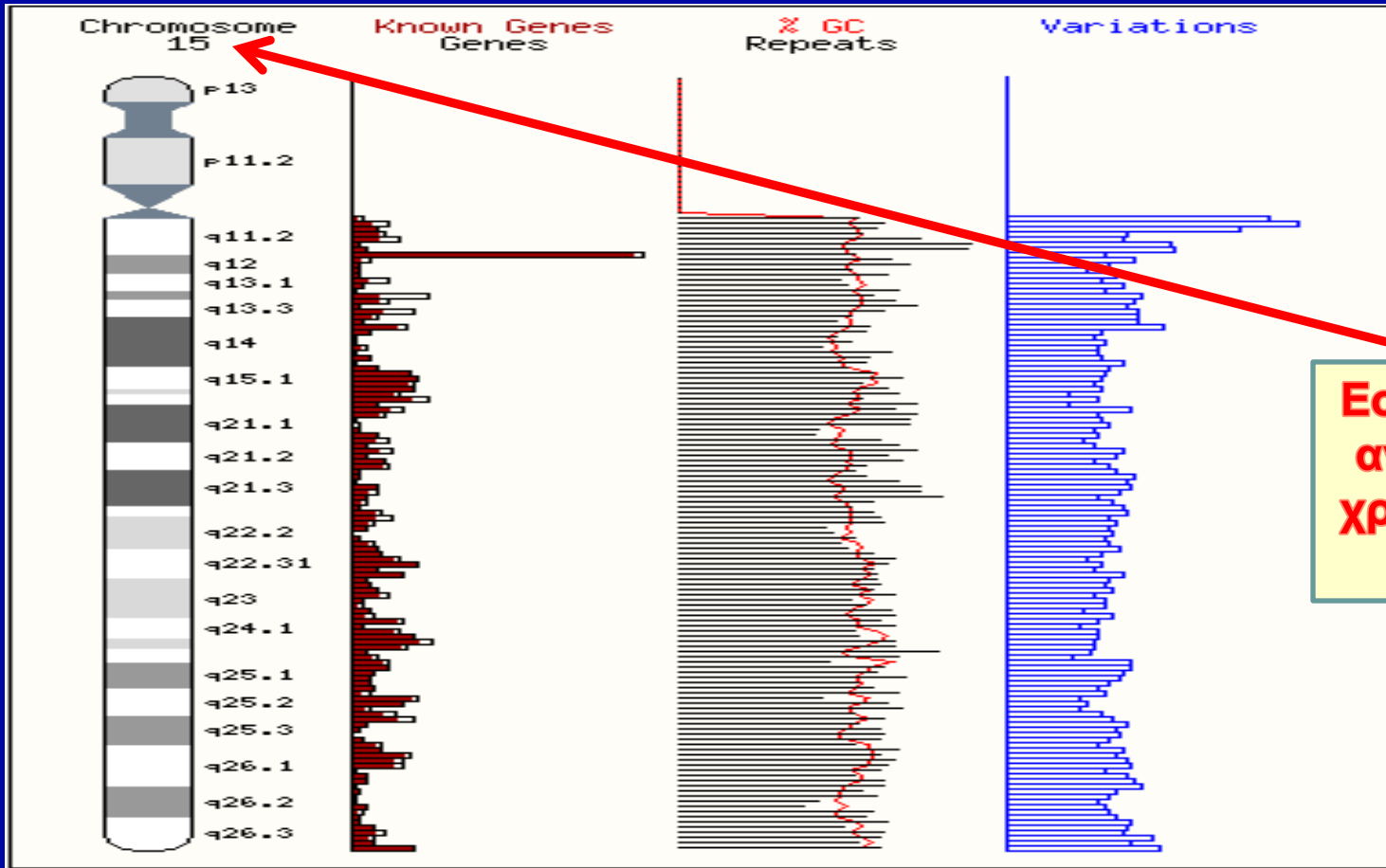
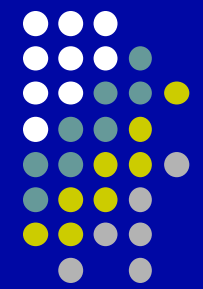


 Alpaca <i>Vicugna pacos</i>	 Fugu <i>Takifugu rubripes</i>	 Orangutan <i>Pongo pygmaeus</i>
 Anole Lizard <i>Anolis carolinensis</i>	 Gorilla <i>Gorilla gorilla</i>	 Pig (preview - assembly only) <i>Sus scrofa</i>
 Anopheles <i>Anopheles gambiae</i>	 Guinea Pig <i>Cavia porcellus</i>	 Pika <i>Ochotona princeps</i>
 Armadillo <i>Dasypus novemcinctus</i>	 Hedgehog <i>Erinaceus europaeus</i>	 Platypus <i>Ornithorhynchus anatinus</i>
 Bushbaby <i>Otolemur gamettii</i>	 Horse <i>Equus caballus</i>	 Rabbit <i>Oryctolagus cuniculus</i>
 Caenorhabditis elegans	 Human <i>Homo sapiens</i>	 Rat <i>Rattus norvegicus</i>
 Ciona intestinalis	 Hyrax <i>Procavia capensis</i>	 Saccharomyces cerevisiae
 Ciona savignyi	 Kangaroo rat <i>Dipodomys ordii</i>	 Shrew <i>Sorex araneus</i>
 Cat <i>Felis catus</i>	 Lamprey (preview - assembly only) <i>Petromyzon marinus</i>	 Sloth <i>Choloepus hoffmanni</i>
 Chicken <i>Gallus gallus</i>	 Lesser hedgehog tenrec <i>Echinops telfairi</i>	 Squirrel <i>Sciurus harrisi</i>
 Chimpanzee <i>Pan troglodytes</i>	 Macaque <i>Macaca mulatta</i>	 Stickleback <i>Gasterosteus aculeatus</i>
 Cow <i>Bos taurus</i>	 Medaka <i>Oryzias latipes</i>	 Tarsier <i>Tarsius syrichta</i>
 Dog <i>Canis familiaris</i>	 Megabat <i>Pteropus vampyrus</i>	 Tetraodon <i>Tetraodon nigroviridis</i>
 Dolphin <i>Tursiops truncatus</i>	 Microbat <i>Myotis lucifugus</i>	 Tree Shrew <i>Tupaia belangeri</i>
 Elephant <i>Loxodonta africana</i>	 Mouse <i>Mus musculus</i>	 Xenopus tropicalis
 Fruitfly <i>Drosophila melanogaster</i>	 Mouse Lemur <i>Microcebus murinus</i>	 Zebra Finch <i>Taeniopygia guttata</i>
 Fugu <i>Takifugu rubripes</i>	 Opossum <i>Monodelphis domestica</i>	 Zebrafish <i>Danio rerio</i>



Ο καρυότυπος του ανθρώπινου γονιδιώματος του προγράμματος Ensembl





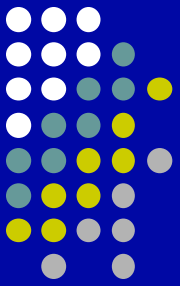
**Εστίαση στο
ανθρώπινο
χρωμόσωμα
15**

Chromosome Statistics

Length (bps):	100,338,915
Known Protein-coding Genes:	652
Novel Protein-coding Genes:	3
Pseudogene Genes:	71
miRNA Genes:	73
rRNA Genes:	8
snRNA Genes:	44
snoRNA Genes:	97
Misc RNA Genes:	70
SNPs:	439,456



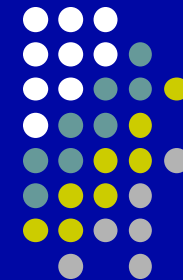
Το πρόγραμμα Ensembl



- Η επόμενη πολύπλοκη παράσταση μας δείχνει τη δομή και το περιεχόμενο του γονιδίου DUT σε διαφορετικές ευκρίνειες:
 - σχηματική επισκόπηση (εύρος 1Mb)
 - λεπτομερή εικόνα (εύρος 10-Kbp)
 - εικόνα σε εύρος βάσης (εύρος 100-bp)
- Πατώντας με το ποντίκι όλες τις περιοχές της εικόνας μπορούμε να εξετάσουμε οποιαδήποτε χρωμοσωμική περιοχή (μέχρι το επίπεδο του γονιδίου) και να πάρουμε πλήθος πληροφοριών για την εσωτερική δομή, τις διαφορετικές μορφές έκφρασης του συγκεκριμένου γονιδίου και να εκτιμήσουμε τις επιπτώσεις των απλών νουκλεοτιδικών πολυμορφισμών μέσα σε αυτό το γονίδιο



Εστιάζοντας στο ανθρώπινο χρωμόσωμα 15 με το πρόγραμμα Ensembl



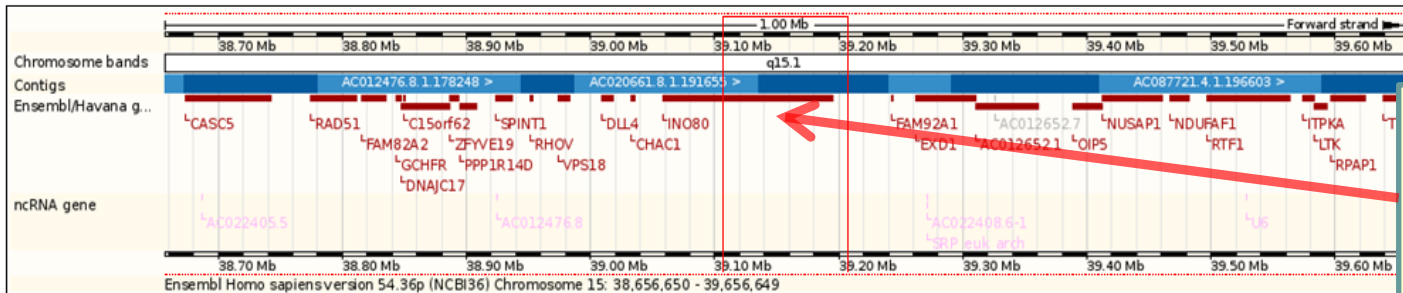
Chromosome 15: 39,106,650-39,206,650



«Region overview

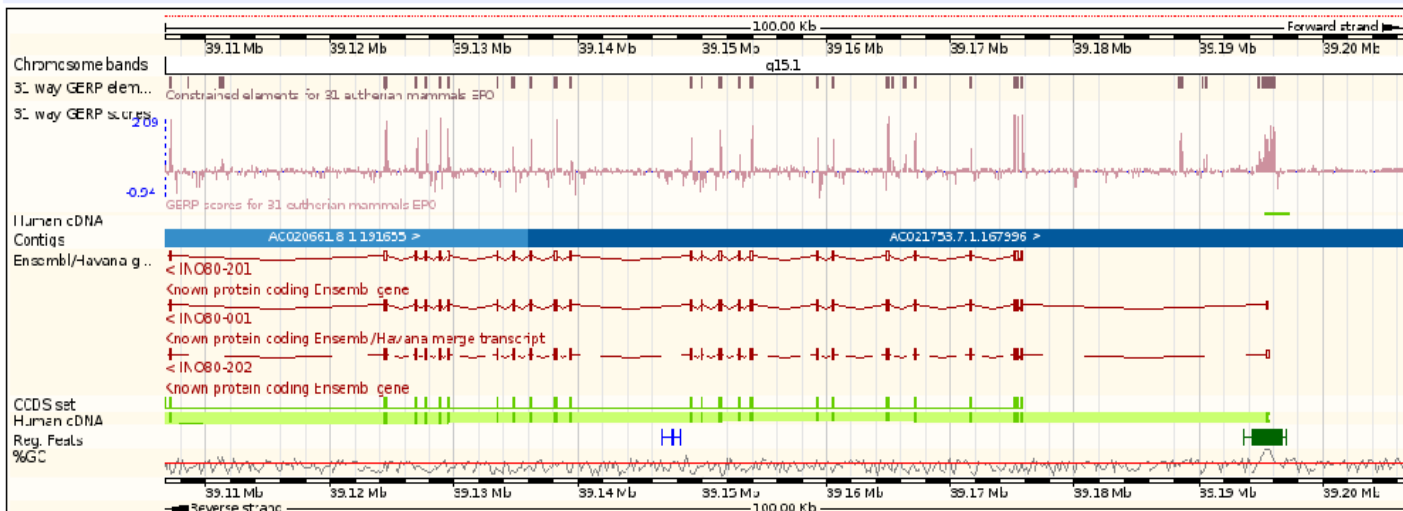
Region in detail [help](#)

Genomic alignments »



Το ανθρώπινο γονίδιο TP53B51

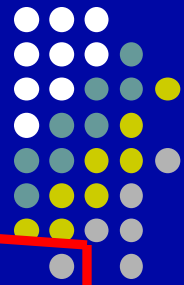
Location: 15 39106650 39206650 [Go>](#)



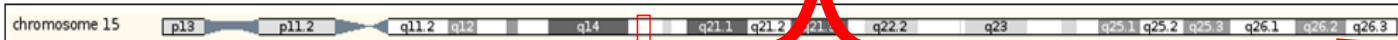
Gene Legend: Known protein coding (red), Gene associated (blue), Promoter associated (green). There are currently 178 trackstuned off. Ensembl Homo sapiens version 54.36p (NCBI36) Chromosome 15: 39,106,650 - 39,206,650



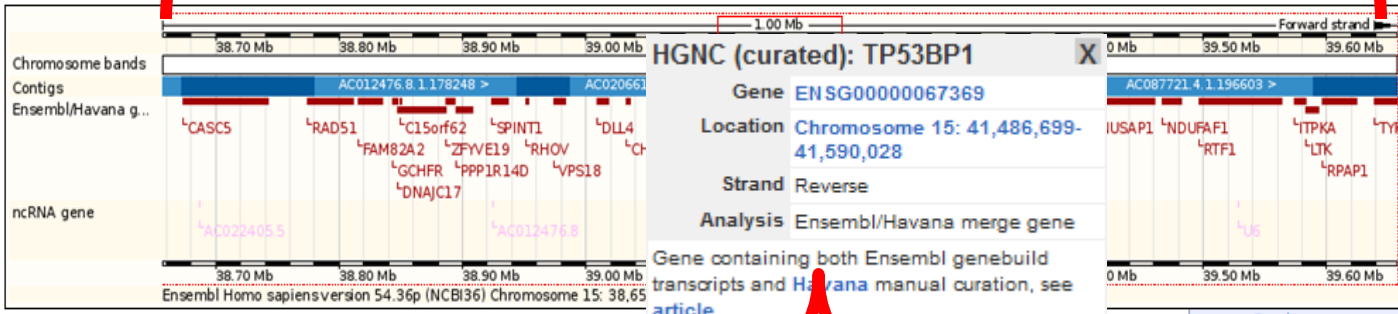
Το ανθρώπινο γονίδιο TP53B51



Chromosome 15: 39,106,650-39,206,650

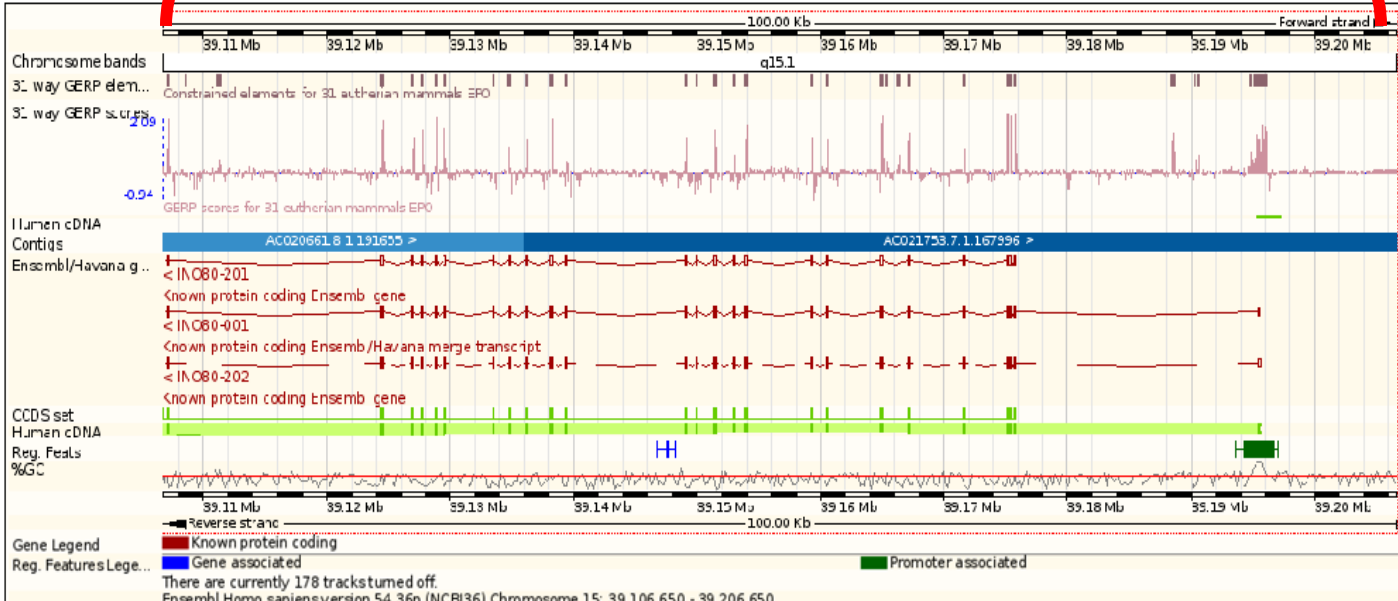


« Region overview Region in detail [help](#) Genomic alignments »



Εύρος 1Mb

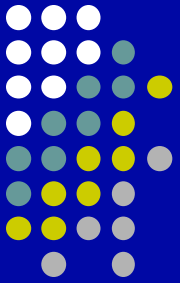
Location: 15 39106650 39206650 [Go>](#)



Εύρος 100 Kb

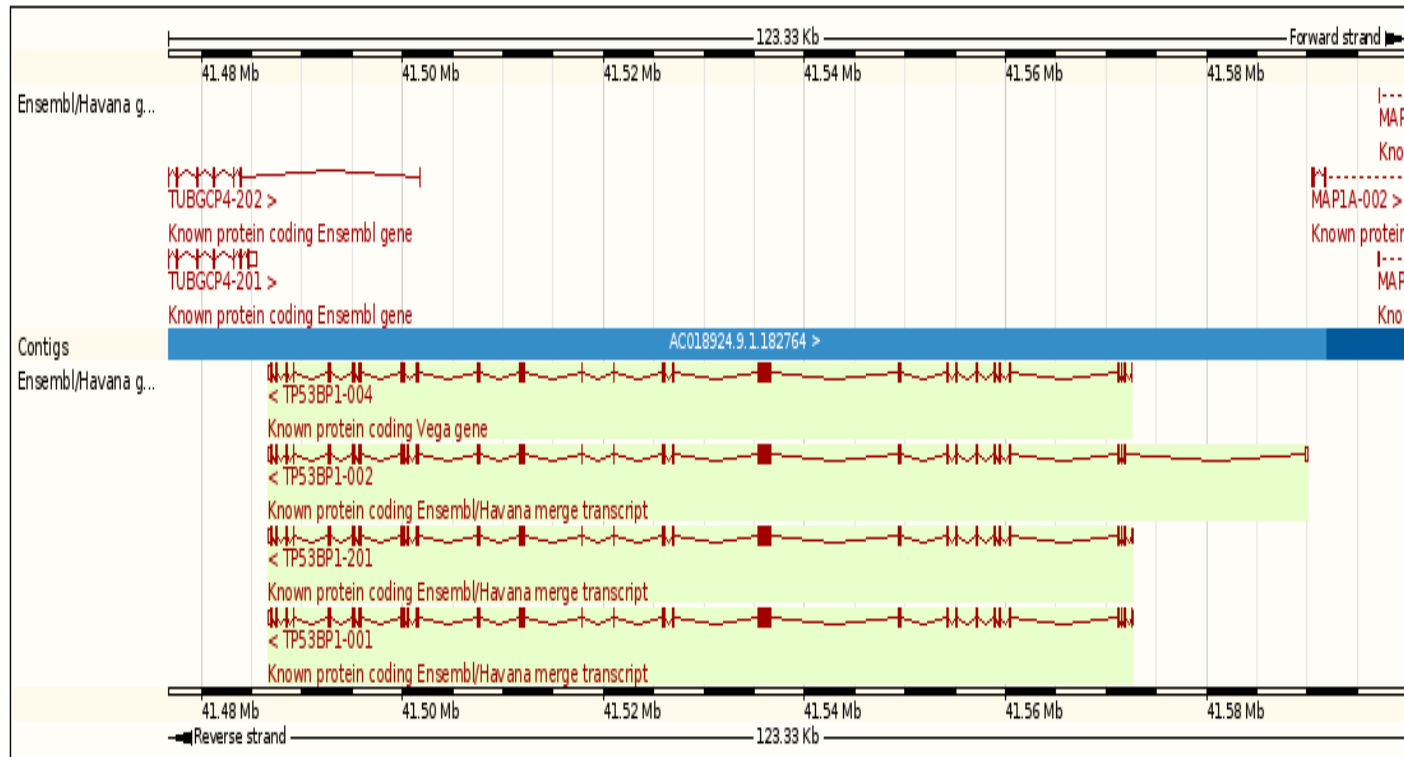


Το ανθρώπινο γονίδιο TP53B51

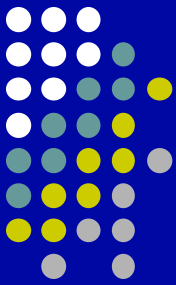


Name [TP53BP1](#) (HGNC (curated))
Synonyms 53BP1, p202 [To view all Ensembl genes linked to the name [click here](#).]
CCDS This gene is a member of the Human CCDS set: [CCDS10096](#)
Gene type Known protein coding
Prediction Method Gene containing both Ensembl genebuild transcripts and [Havana](#) manual curation, see [article](#).

Transcripts



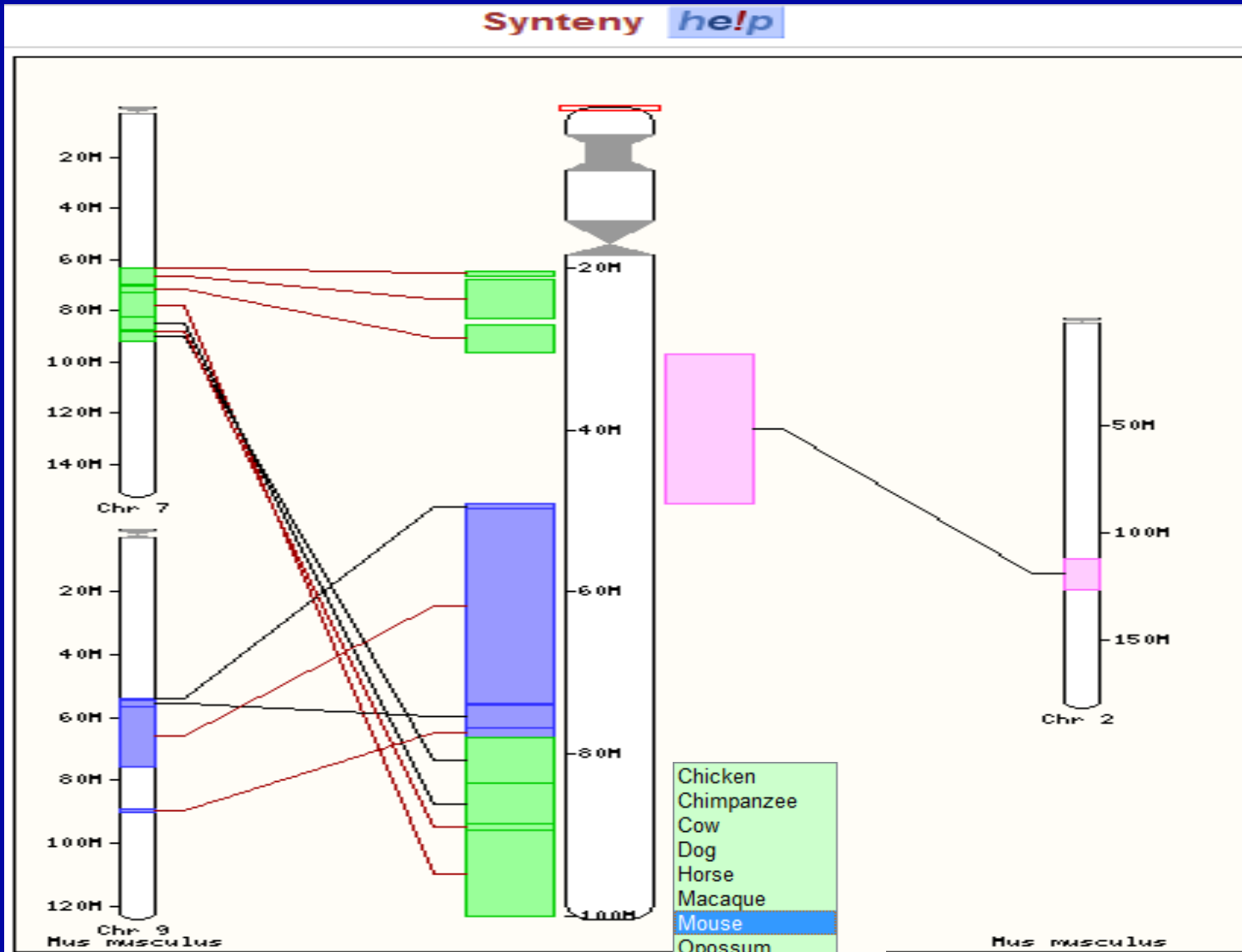
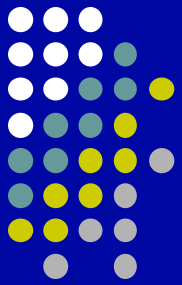
Περισσότερες πληροφορίες για το TP53BP1 με τη βοήθεια του Ensembl

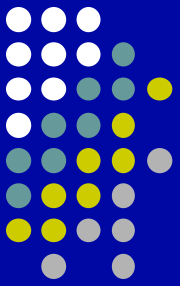


Core Data		Database Links			
Approved Symbol +	TP53BP1	Accession Numbers +			
Approved Name +	tumor protein p53 binding protein 1	U09477	GenBank	EMBL	DDBJ UCSC
HGNC ID +	HGNC:11999	Rat Genome Database ID (mapped data supplied by RGD) +			
Status +	Approved	RGD:1308039	RGDID		
Chromosome +	15q15-q21	Entrez Gene ID +			
Previous Symbols +		7158	Gene	Map Viewer	
Previous Names +	"tumor protein p53-binding protein, 1"	CCDS IDs +			
Aliases +	53BP1, p202	CCDS10096.1	CCDSID		
Name Aliases +		Pubmed IDs +			
Locus Type +	gene with protein product	8016121, 9748285	PMID	CiteXplore	
		VEGA IDs +			
Gene Symbol Links		OTTHUMG00000059757	VEGA GeneView		
GENATLAS GeneCards GeneClinics / GeneTests GoPubmed		Ensembl ID (mapped data supplied by Ensembl) +			
HCOP HInvDB Treefam wikigenes		ENSG00000067369	Ensembl GeneView	UCSC	
		RefSeq (mapped data supplied by NCBI) +			
		NM_001141980	GenBank	EMBL	DDBJ UCSC
Specialist Database Links		OMIM ID (mapped data supplied by NCBI) +			
COSMIC		605230	OMIM		
		UCSC ID (mapped data supplied by UCSC) +			
		uc001zrs.1	UCSC Index		
		UniProt ID (mapped data supplied by UniProt) +			
		Q12888	UniProt	UCSC	



Συνταινία του γονιδίου TP53B51

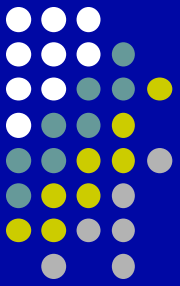




Πρωτεϊνικές βάσεις δεδομένων

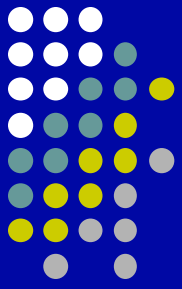


Από τα ανοικτά πλαίσια ανάγνωσης στις ώριμες πρωτεΐνες

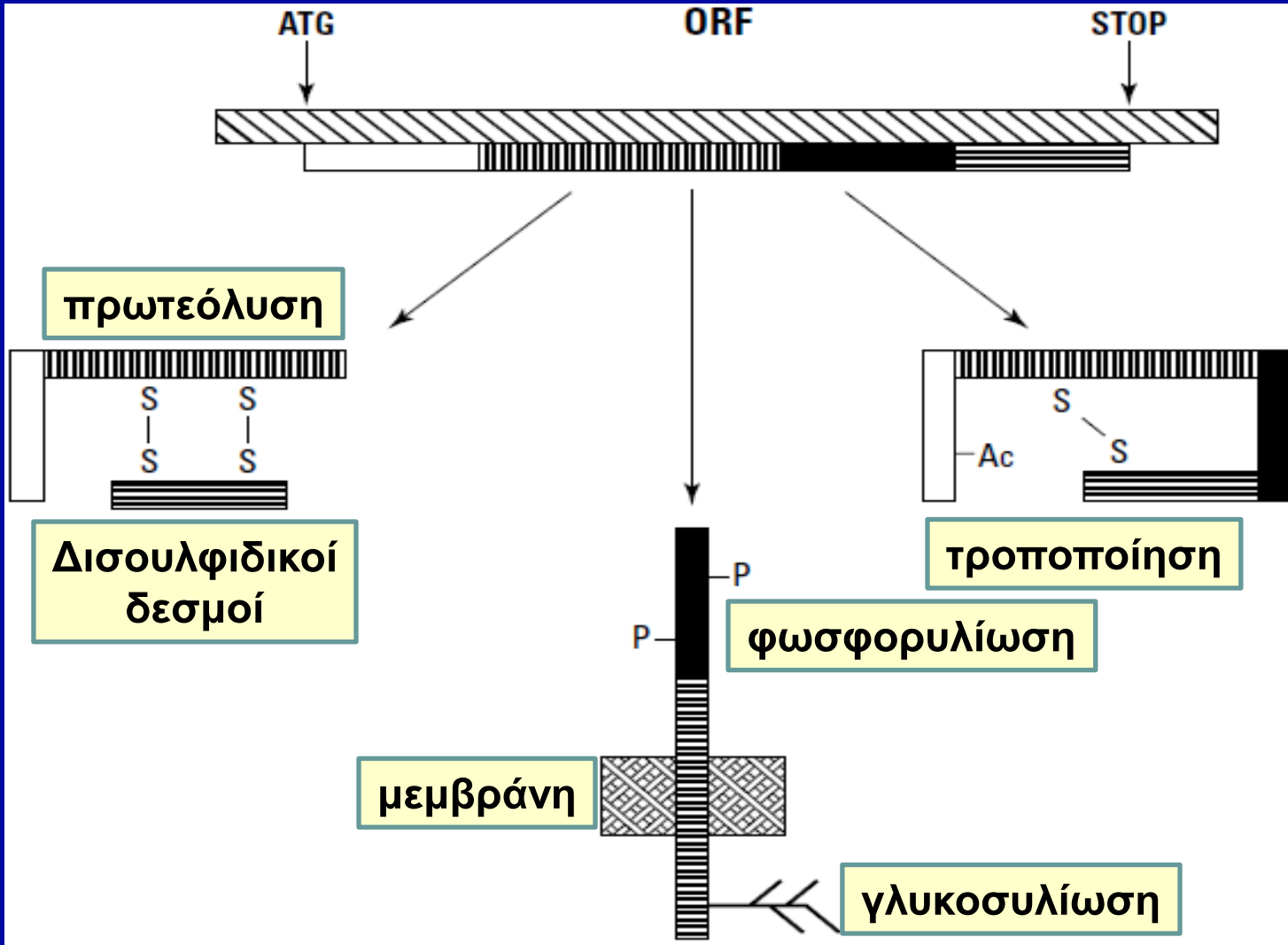


- Οι πρωτεΐνες σπανίως συμπεριφέρονται όπως θα περίμενε κανείς με βάση τη μετάφραση που δίνεται από τους υπολογιστές
- Ο λόγος είναι ότι η αλυσίδα των αμινοξέων τροποποιείται κατά την πορεία της προς μια ώριμη πρωτεΐνη
- Το αποτέλεσμα είναι ότι ακόμα και κάποιες απλές φυσικοχημικές ιδιότητες της ώριμης πρωτεΐνης (μέγεθος, μοριακό βάρος, ισοηλεκτρικό σημείο) είναι δύσκολο να προβλεφθούν
- Αυτή η περίπλοκη διαδικασία της πρωτεϊνικής ωρίμανσης ονομάζεται μετα-μεταφραστική τροποποίηση (post translational modification)

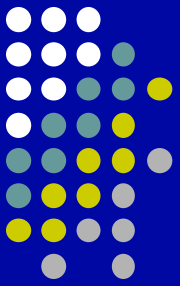




Πιθανές τροποποιήσεις που συναντώνται κατά τη διάρκεια της πρωτεϊνικής ωρίμανσης



Η Swissprot εισαγωγή



Διαιρείται στα εξής τμήματα:

- γενικών πληροφοριών
 - βιβλιογραφικών πληροφοριών
 - λειτουργικών πληροφοριών
 - τον πίνακα με τα χαρακτηριστικά
 - το τμήμα με την ακολουθία
-
- ✓ Εισάγουμε τη διεύθυνση www.expasy.ch/sprot/
 - ✓ Πληκτρολογούμε για το συγκεκριμένο παράδειγμα το Swiss-Prot ID P00533 στο πεδίο search
 - ✓ Πατάμε το κουμπί Go



Η Swissprot εισαγωγή

Search for



The UniProt Knowledgebase consists of:

- **UniProtKB/Swiss-Prot**, a curated protein sequence database which strives to provide a high level of annotation (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases [[More details](#) / [References](#) / [Linking to Swiss-Prot](#) / [User manual](#) / [Recent changes](#) / [Disclaimer](#)].
- **UniProtKB/TrEMBL**; a computer-annotated supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Swiss-Prot.

These databases are developed by the Swiss-Prot groups at [SIB](#) and at [EBI](#).

UniProt Knowledgebase Release 15.4 consists of:

UniProtKB/Swiss-Prot Release 57.4 of 16-Jun-2009: 470369 entries ([More statistics](#))

UniProtKB/TrEMBL Release 40.4 of 16-Jun-2009: 8594382 entries ([More statistics](#))

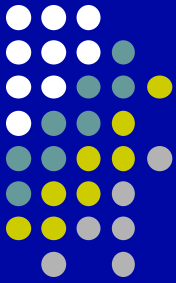
> **Swiss-Prot headlines**

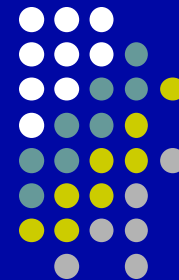
Rotavirus: a serial killer in UniProtKB/Swiss-Prot ([Read more...](#))

Access to the UniProt Knowledgebase


- **New UniProt web site** new
- UniProtKB Taxonomy browser
- **ViralZone** - Portal to viral UniProtKB/Swiss-Prot entries new
- **BLAST** similarity search

- Retrieve a list of UniProtKB entries
- Randomly retrieve a UniProtKB entry
- UniProtKB Sequence/Annotation Version Database
- Swiss-Prot ID tracker





★ Reviewed, UniProtKB/Swiss-Prot P00533 (EGFR_HUMAN)

Last modified November 3, 2009. Version 154.  [History...](#)

Contribute

 [Send feedback](#)

 [Read comments \(0\) or add your own](#)

 [Clusters with 100%, 90%, 50% identity](#) |  [Documents \(7\)](#) |  [Third-party data](#) |  [Customize display](#)


[TEXT](#) | [XML](#) | [RDF/XML](#) | [GFF](#) | [FASTA](#)





[Names and origin](#) · [Protein attributes](#) · [General annotation \(Comments\)](#) · [Ontologies](#) · [Binary interactions](#) · [Alternative products](#) · [Sequence annotation \(Features\)](#) · [Sequences](#) · [References](#) · [Web resources](#) · [Cross-references](#) · [Entry information](#) · [Relevant documents](#)



Όνόματα και προέλευση

★ Reviewed, UniProtKB/Swiss-Prot **P00533** (EGFR_HUMAN)

Last modified June 16, 2009. Version 148.  [History...](#)

 Clusters with 100%, 90%, 50% identity |  Documents (7) |  Third-party data |  Customize display

[Names and origin](#) · [Protein attributes](#) · [General annotation \(Comments\)](#) · [Ontologies](#) · [Binary interactions](#) · [Alternative information](#) · [Relevant documents](#)

Names and origin

Protein names

Recommended name:

Epidermal growth factor receptor

EC=2.7.10.1 

Alternative name(s):

Receptor tyrosine-protein kinase ErbB-1

Gene names

Name: **EGFR**

Synonyms: ERBB1

Organism

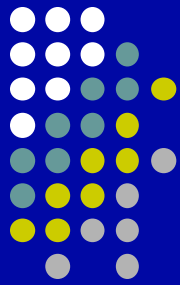
Homo sapiens (Human)

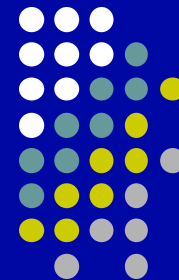
Taxonomic identifier

9606 [NCBI]

Taxonomic lineage

[Eukaryota](#) › [Metazoa](#) › [Chordata](#) › [Craniata](#) › [Vertebrata](#) › [Homo](#)





NiceZyme View of ENZYME: EC 2.7.10.1

Accepted Name

Receptor protein-tyrosine kinase.

Alternative Name(s)

Receptor protein tyrosine kinase.

Reaction catalysed

ATP + a [protein]-L-tyrosine <=> ADP + a [protein]-L-tyrosine phosphate

Comment(s)

- The receptor protein-tyrosine kinases, which can be defined as having a transmembrane domain, are a large and diverse multigene family found only in metazoans.
- In the human genome, 58 receptor-type protein-tyrosine kinases have been identified and these are distributed into 20 subfamilies.
- Formerly EC 2.7.1.112.

Cross-references

PROSITE [PDOC00100](#)

BRENDA [2.7.10.1](#)

EC2PDB [2.7.10.1](#)

PRIAM enzyme-specific profiles [2.7.10.1](#)

KEGG Ligand Database for Enzyme Nomenclature [2.7.10.1](#)

IUBMB Enzyme Nomenclature [2.7.10.1](#)

IntEnz [2.7.10.1](#)

MEDLINE [Find literature relating to 2.7.10.1](#)

MetaCyc [2.7.10.1](#)

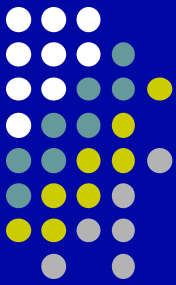
[P13368](#), 7LESS_DROME;
[Q9UM73](#), ALK_HUMAN;
[P79750](#), CSF11_FUGRU;

[P20806](#), 7LESS_DROVI;
[P97793](#), ALK_MOUSE;
[Q8UVR8](#), CSF12_FUGRU;

[Q54N73](#), 7TMK1_DICDI;
[Q9VBW3](#), CAD96_DROME;
[Q9I8N6](#), CSF1R_DANRE;



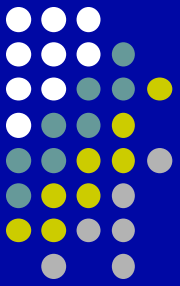
Πρωτεϊνικά χαρακτηριστικά - Σχόλια



Protein attributes		Hide Top
Sequence length	1210 AA.	
Sequence status	Complete.	
Sequence processing	The displayed sequence is further processed into a mature form.	
Protein existence	Evidence at protein level.	
General annotation (Comments)		Hide Top
Function	<p>Receptor for EGF, but also for other members of the EGF family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding EGF-like growth factor, GP30 and vaccinia virus growth factor. Is involved in the control of cell growth and differentiation. Phosphorylates MUC1 in breast cancer cells and increases the interaction of MUC1 with C-SRC and CTNNB1/beta-catenin. Ref.32</p> <p>Isoform 2/truncated isoform may act as an antagonist. Ref.32</p>	
Catalytic activity	ATP + a [protein]-L-tyrosine = ADP + a [protein]-L-tyrosine phosphate.	
Subunit structure	Binds RIPK1. CBL interacts with the autophosphorylated C-terminal tail of the EGF receptor. Part of a complex with ERBB2 and either PIK3C2A or PIK3C2B. The autophosphorylated form interacts with PIK3C2B, maybe indirectly. Interacts with PELP1. Binds MUC1. Ref.32 Ref.26 Ref.30 Ref.31 Ref.36	
Subcellular location	<p>Cell membrane; Single-pass type I membrane protein.</p> <p>Isoform 2: Secreted.</p>	
Tissue specificity	Ubiquitously expressed. Isoform 2 is also expressed in ovarian cancers. Ref.39	
Post-translational modification	<p>Phosphorylation of Ser-695 is partial and occurs only if Thr-693 is phosphorylated. Ref.17 Ref.25 Ref.29 Ref.33 Ref.34 Ref.35 Ref.37 Ref.38 Ref.40 Ref.41 Ref.42</p> <p>Ref.43</p> <p>Monoubiquitinated and polyubiquitinated upon EGF stimulation; which does not affect tyrosine kinase activity or signaling capacity but may play a role in lysosomal targeting. Polyubiquitin linkage is mainly through 'Lys-63', but linkage through 'Lys-48', 'Lys-11' and 'Lys-29' also occur.</p>	
Involvement in disease	Defects in EGFR are associated with lung cancer [MIM:211980].	
Miscellaneous	Binding of EGF to the receptor leads to dimerization, internalization of the EGF-receptor complex, induction of the tyrosine kinase activity, stimulation of cell DNA synthesis, and cell proliferation.	
Sequence similarities	<p>Belongs to the protein kinase superfamily. Tyr protein kinase family. EGF receptor subfamily.</p> <p>Contains 1 protein kinase domain.</p>	



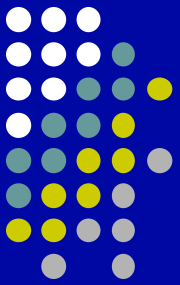
Πρωτεϊνικά χαρακτηριστικά - Σχόλια



- **FUNCTION** – υποδοχέας του EGF, εμπλεκόμενος στον έλεγχο της κυτταρικής ανάπτυξης
- **CATALYTIC ACTIVITY** – πρωτεϊνική κινάση (μεταφέρει φωσφορικά γκρουπ από υψηλά ενεργειακά μόρια σε συγκεκριμένα υποστρώματα) τυροσίνης
- **SUBUNIT** – άλλες πρωτεΐνες με τις οποίες σχηματίζει σύμπλοκα
- **SUBCELLULAR LOCATION** – πρωτεΐνη κυτταρικής μεμβράνης
- **PTM** – λίστα από τροποποιήσεις
- **DISEASE** – ατέλειες στην EGFR εμπλέκονται στη δημιουργία καρκίνου του πνεύμονα
- **MISCELLANEOUS** – περιγραφή του μοριακού μηχανισμού που σχετίζεται με τη λειτουργία της πρωτεΐνης
- **SIMILARITY** – με άλλες ακολουθίες της πρωτεΐνης EGF



ΟΝΤΟΛΟΓΙΕΣ

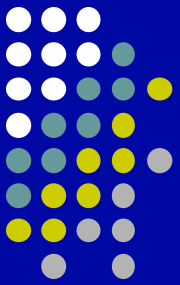


Ontologies	
Keywords	
Biological process	Cell cycle
Cellular component	Cell membrane Membrane Secreted
Coding sequence diversity	Alternative splicing Polymorphism
Disease	Disease mutation
Domain	Repeat Signal Transmembrane
Ligand	ATP-binding Nucleotide-binding
Molecular function	Anti-oncogene Kinase Receptor Transferase Tyrosine-protein kinase
PTM	Disulfide bond Glycoprotein Isopeptide bond Phosphoprotein Ubl conjugation
Technical term	3D-structure Direct protein sequencing
Gene Ontology (GO)	
Biological process	<p>activation of phospholipase A2 activity by calcium-mediated signaling Traceable author statement. Source: UniProtKB</p> <p>activation of phospholipase C activity Traceable author statement. Source: UniProtKB</p> <p>cell proliferation Traceable author statement. Source: Protinc</p> <p>cell-cell adhesion Inferred from mutant phenotype. Source: UniProtKB</p>

Cellular component	<p>Shc-EGFR complex Inferred from sequence or structural similarity. Source: UniProtKB</p> <p>basolateral plasma membrane Inferred from direct assay. Source: UniProtKB</p> <p>endoplasmicreticulum Inferred from direct assay. Source: HPA</p> <p>endosome Inferred from direct assay. Source: UniProtKB</p> <p>extracellular space ^{Ref. 4} Non-traceable author statement. Source: UniProtKB</p> <p>integral to membrane Inferred from electronic annotation. Source: UniProtKB-KW</p> <p>nucleus Inferred from direct assay. Source: UniProtKB</p>
Molecular function	<p>ATP binding Inferred from electronic annotation. Source: UniProtKB-KW</p> <p>MAP/ERK kinase kinase activity Non-traceable author statement. Source: UniProtKB</p> <p>actin filament binding Inferred from direct assay. Source: UniProtKB</p> <p>double-stranded DNA binding ^{Ref. 22} Non-traceable author statement. Source: UniProtKB</p> <p>epidermal growth factor receptor activity ^{Ref. 22} ^{Ref. 24} Inferred from direct assay. Source: UniProtKB</p> <p>identical protein binding Inferred from physical interaction. Source: IntAct</p> <p>protein heterodimerization activity Inferred from direct assay. Source: UniProtKB</p> <p>protein phosphatase binding Inferred from physical interaction. Source: UniProtKB</p>



Εναλλακτικά προϊόντα - Ισομορφές



Alternative products

This entry describes **4** isoforms produced by **alternative splicing**. [[Align](#)] [[Select](#)]

Isoform 1 (identifier: **P00533-1**)

Also known as: p170;

This isoform has been chosen as the 'canonical' sequence. All positional information in this entry refers to it.

Isoform 2 (identifier: **P00533-2**)

Also known as: p60; Truncated; TEGFR;

The sequence of this isoform differs from the canonical sequence as follows:

404-405: FL → LS

406-1210: Missing.

Isoform 3 (identifier: **P00533-3**)

Also known as: p110;

The sequence of this isoform differs from the canonical sequence as follows:

628-705: CTGPGLEGCP...GEAPNQALLR → PGNESLKAML...SVIITASSCH

706-1210: Missing.

Isoform 4 (identifier: **P00533-4**)

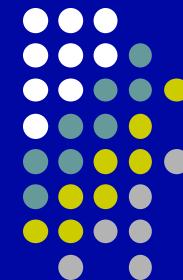
The sequence of this isoform differs from the canonical sequence as follows:

628-628: C → S

629-1210: Missing.



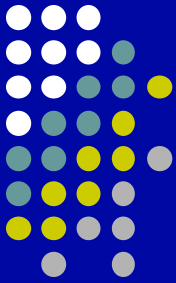
Πληροφοριακά στοιχεία ακολουθίας



Sequence annotation (Features)			
Feature key	Position(s)	Length	Description
Molecule processing			
<input type="checkbox"/>	Signal peptide	1–24	24 Ref.18
<input type="checkbox"/>	Chain	25–1210	1186 Epidermal growth factor receptor
Regions			
<input type="checkbox"/>	Topological domain	25–645	621 Extracellular Potential
<input type="checkbox"/>	Transmembrane	646–668	23 Potential
<input type="checkbox"/>	Topological domain	669–1210	542 Cytoplasmic Potential
<input type="checkbox"/>	Repeat	75–300	226 Approximate
<input type="checkbox"/>	Repeat	390–600	211 Approximate
<input type="checkbox"/>	Domain	712–979	268 Protein kinase
<input type="checkbox"/>	Nucleotide binding	718–726	9 ATP By similarity
<input type="checkbox"/>	Compositional bias	1025–1071	47 Ser-rich
Sites			
<input type="checkbox"/>	Active site	837	1 Proton acceptor By similarity
<input type="checkbox"/>	Binding site	745	1 ATP By similarity
<input type="checkbox"/>	Site	1016	1 Important for interaction with PIK3C2B
Amino acid modifications			
<input type="checkbox"/>	Modified residue	678	1 Phosphothreonine; by PKC
<input type="checkbox"/>	Modified residue	693	1 Phosphothreonine Ref.17 Ref.29 Ref.33 Ref.37 Ref.40 Ref.41 Ref.42 Ref.43
<input type="checkbox"/>	Modified residue	695	1 Phosphoserine Ref.17 Ref.42 Ref.43
<input type="checkbox"/>	Modified residue	725	1 Phosphothreonine Ref.42
<input type="checkbox"/>	Modified residue	869	1 Phosphotyrosine Ref.38
<input type="checkbox"/>	Modified residue	978	1 Phosphotyrosine Ref.35
<input type="checkbox"/>	Modified residue	991	1 Phosphoserine Ref.33 Ref.37 Ref.41 Ref.42 Ref.43
<input type="checkbox"/>	Modified residue	993	1 Phosphothreonine Ref.43
<input type="checkbox"/>	Modified residue	995	1 Phosphoserine Ref.37 Ref.43



Πληροφοριακά στοιχεία ακολουθίας



<input type="checkbox"/>	Glycosylation	56	1	N-linked (GlcNAc...) (complex); atypical; partial Ref.33 Ref.28
<input type="checkbox"/>	Glycosylation	128	1	N-linked (GlcNAc...) Ref.33 Ref.27
<input type="checkbox"/>	Glycosylation	175	1	N-linked (GlcNAc...) Ref.33 Ref.27
<input type="checkbox"/>	Disulfide bond	31 ↔ 58		Ref.21
<input type="checkbox"/>	Disulfide bond	157 ↔ 187		Ref.21
<input type="checkbox"/>	Disulfide bond	190 ↔ 199		Ref.21
<input type="checkbox"/>	Cross-link	716		Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin) Ref.20
<input type="checkbox"/>	Cross-link	737		Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin) Ref.20
<input type="checkbox"/>	Cross-link	754		Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in ubiquitin) Ref.20

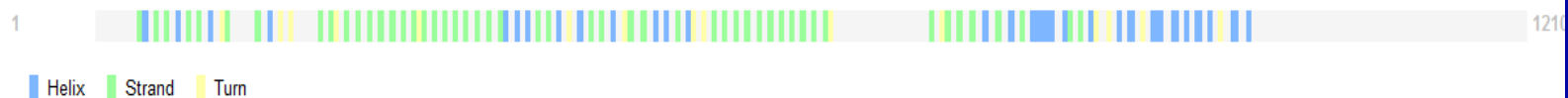
Natural variations

<input type="checkbox"/>	Alternative sequence	404 – 405	2	FL → LS in isoform 2.
<input checked="" type="checkbox"/>	Alternative sequence	406 – 1210	805	Missing in isoform 2.
<input checked="" type="checkbox"/>	Alternative sequence	628 – 705	78	CTGPG...QALLR → PGNESLKAMLFCLFKLSSCN QSNDSVSHQSGSPAAQESC LGWIPSLLPSEFQLGWGGCS HLHAWPSASVITASSCH in isoform 3.
<input type="checkbox"/>	Alternative sequence	628	1	C → S in isoform 4.
<input checked="" type="checkbox"/>	Alternative sequence	629 – 1210	582	Missing in isoform 4.
<input checked="" type="checkbox"/>	Alternative sequence	706 – 1210	505	Missing in isoform 3.
<input type="checkbox"/>	Natural variant	98	1	R → Q: dbSNP rs17289589 . Ref.7
<input type="checkbox"/>	Natural variant	266	1	P → R: dbSNP rs17336639 . Ref.7
<input type="checkbox"/>	Natural variant	521	1	R → K: dbSNP rs2227983 . Ref.7 Ref.48

Experimental info

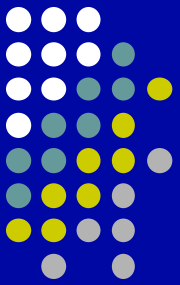
<input type="checkbox"/>	Mutagenesis	1016	1	Y → F: 50% decrease in interaction with PIK3C2B. 65% decrease in interaction with PIK3C2B; when associated with F-1197. Abolishes interaction with PIK3C2B; when associated with F-1197 and F-1092. Ref.30
<input type="checkbox"/>	Mutagenesis	1092	1	Y → F: No change in interaction with PIK3C2B. Abolishes interaction with PIK3C2B; when associated with F-1197 and F-1016. Ref.30

Secondary structure



[Details...](#)

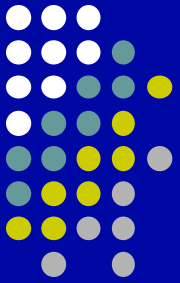
Πληροφοριακά στοιχεία ακολουθίας



- **SIGNAL** – τα αμινοξέα 1-24 αποτελούν ένα πεπτίδιο σήμανσης
- **CHAIN** – η ώριμη πεπτιδική αλυσίδα (από 25-1210)
- **TOPO_DOM** – τοπολογικό πεδίο που δείχνει ποια περιοχή της ακολουθίας σχετίζεται με ότι συμβαίνει εξωκυτταρική ή ενδοκυτταρικά
- **TRANSMEM** – τα αμινοξέα 646-668 αποτελούν την τρανσμεμβρανική περιοχή της πρωτεΐνης
- **DOMAIN** – τα αμινοξέα 669-1210 δείχνουν ένα άλλο πεδίο. Είναι όλα ενδοκυτταρικά***
- **BINDING** – το ακριβές σημείο πρόσδεσης με το ATP***
- **ACT_SITE** – τα νούμερα αναφέρουν τα αμινοξέα που εμπλέκονται στη δραστηριότητα του ενζύμου
- **COMPBIAS** – περιοχή πλούσια σε κάποιο συγκεκριμένο/α αμινοξύ***



Πληροφοριακά στοιχεία ακολουθίας



P00533[669-1210], Epidermal growth factor receptor, Homo sapiens (Human)

```

10      20      30      40      50      60
MRPSGTAGAA LLALLAALCP ASRALEEKV CQQTSNKLTQ LGTFEDHFLS LQRMFNNCEV

70      80      90      100     110     120
VLGNLEITYV QRNYDLSFLK TIQEVAGYVL IALNIVERIP LENLQIIRGN MYYENSYALA

130     140     150     160     170     180
VLSNYDANKI GLKELPMMNL QEILHGAVER SNNPALCNVE SIQWRDIVSS DFLSNMSMDF

190     200     210     220     230     240
QNHLGSCQKQ DFSCFNGSCW GAGEENCQKI TKIICAQQCS GRCRGKSPSD CCHNQCAAGC

250     260     270     280     290     300
TGPRESDECLV CRKFRDEATC KDTCPPLMLV NPTTYQMDVN PEGKYSPGAT CVKKCPRNYV

310     320     330     340     350     360
VTDHGSCVRA CGADSYEMEE DGVRKCKKCE GPCRKVCNGI GIGEFKDSLS INATNIKHPK

370     380     390     400     410     420
NCTSISSGDLH ILPVAFRGDS FTHTPPLDFQ ELDILKTVKE ITGFLLIQAW PENRTDLHAF

430     440     450     460     470     480
ENLEIIRGRT KQHQQPFLAV VSLNITSLG RSLKEISDGD VIISGNKILC YANTINWKKL

490     500     510     520     530     540
FGTSGQKTKI ISNRGENSCK ATGQVCHALC SPEGCGWPEP RDCVSCRNVV RGRECVDKCN

550     560     570     580     590     600
LLEGEPEFVY ENSECIQCHP ECLPQAMNIT CTGRGPDNCT QCAHYIDGPH CVKTCFAGVM

610     620     630     640     650     660
GENNVLVWKY ADAGHVCHLC HPNCTYGCCTG PGLEGCPYNG PKIPSIATGM VGALLLLLVV

670     680     690     700     710     720
ALGIGLFMRK RHIVRKRTLR RLLQERELVE PLTPSGEAPN QALLRILKET EFKKIKVLGS

730     740     750     760     770     780
GAFGTVYKGL WIPEGEKVKI FVAIKELREA TSPKANKEIL DEAYVMASVD NPHVCALLGI

790     800     810     820     830     840
CLTSTVQLIT QLMPPGCLLD YVREHKDNIG SQVLLNWCVO IAKGQNYLED RRLVHRDLAA

850     860     870     880     890     900
RNVLVKTPQH VKITDFGLAK LLGAEKHYE AEGCKVPIKW MALESILHRI YTHQSDVWSY

910     920     930     940     950     960
GVTVWELMTF GSKPYDGIPA SEISSILEKG ERLPQPPICT IDVVMIMVKC WMIDADSRPK

970     980     990     1000    1010    1020
FRELIIEFSK MARDPORYLV IQGDERMHLF SPTDSNFYRA IMDEEDMDDV VDADEYLIPQ

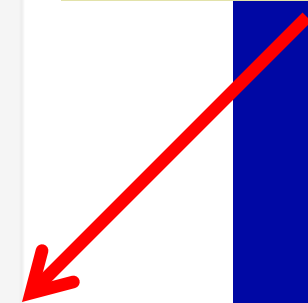
1030    1040    1050    1060    1070    1080
QGFFSSPSTS RTPLLSSLSA TSNMSTVACI DRNGLQSCPI KEDSFLQRYV SDPTGALTED

1090    1100    1110    1120    1130    1140
SIDDTFLPVE EYINQSVPKR FAGSVQNFVY HNQPLNAPS RDPHYQDPHS TAVGNPEYLN

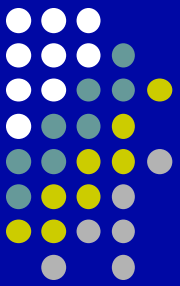
1150    1160    1170    1180    1190    1200
TVQPTCVNST FDSPAHWAQR CSHQISLDNF DYQQDFFPKE AKPNGIFPKG TAENAAYLRV

1210
RPOSSEPIGR
    
```

domain



Πληροφοριακά στοιχεία ακολουθίας



430 440 450 460 470 480
ENLEIIRGRT KQHGGQFSLAV VSLNITSLGL RSLKEISDGD VIISGNKNLC YANTINWKKL

490 500 510 520 530 540
FGTSGQKTKI ISNRGENSCK ATGQVCHALC SPEGCWGPEP RDCVSCRNVS RGRECVDKCN

550 560 570 580 590 600
LLEGEPREFV ENSECIQCHP ECLPQAMNIT CTGRGPDNCI QCAHYIDGPH CVKTCPAGVM

610 620 630 640 650 660
GENNTLVWKY ADAGHVCHLC HPNCTYGCTG PGLEGCPYNG PKIPSIATGM VGALLLLLVV

670 680 690 700 710 720
ALGIGLFMRR RHIVRKRTL RLLQERELVE PLTPSGEAPN QPFRILKET EFKKIKVLGS

730 740 750 760 770 780
GAFGTVYKGL WIPEGEKVKI PVAIKELREA TSPKANKEIL DEAYVMASVD NPHVCRLGLI

790 800 810 820 830 840
CLTSTVQLIT QLMFPGCLLD YVREHKDNIG SQYLLNWCVQ IAKGMNYLED RRLVHRDLAA

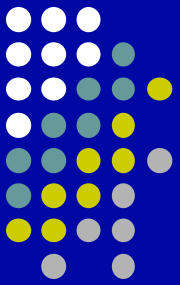
850 860 870 880 890 900
RNVLVKTPQH VKITDFGLAK LLGAEKEYH AEGGKVPIKW MALESILHRI YTHQSDVWSY

910 920 930 940 950 960
GVTVWELMTF GSKPYDGIPA SEISSILEKG ERLPQPPICT IDVYMIMVKC WMIDADSRPK

binding
site



Πληροφοριακά στοιχεία ακολουθίας



GAFGTVYKGL WIPEGEKVKI PVAIKELREA TSPKANKEIL DEAYVMASVD NPHVCRLLEGI

790 800 810 820 830 840
CLTSTVQLIT QLMPFGCLLD YVREHKDNIG SQYLLNWCVQ IAKGMNYLED RRLVHRDLAA

850 860 870 880 890 900
RNVLVKTPQH VKITDFGLAK LLGAEKEYH AEGGKVPIKW MALESILHRI YTHQSDVWSY

910 920 930 940 950 960
GVTVWELMTF GSKPYDGIPA SEISSILEKG ERLPQPPICT IDVYMIMVKC WMIDADSRPK

970 980 990 1000 1010 1020
FRELIIIEFSK MARDPQRYLV IQGDERMHLP SPTDSNFYRA LVELEDMDV VDADEYLIPQ

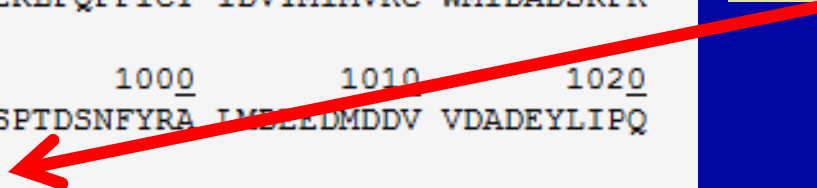
1030 1040 1050 1060 1070 1080
QGFFSSPSTS RTPLLSSLSA TSNNSTVACI DRNGLQSCPI KEDSFLQRYYS SDPTGALTED

1090 1100 1110 1120 1130 1140
SIDDTFLPVP EYINQSVPKR PAGSVQNPVY HNQPLNPAPS RDPHYQDPHS TAVGNPEYLN

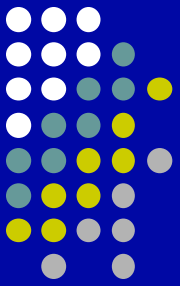
1150 1160 1170 1180 1190 1200
TVQPTCVNST FDSPAHWQK GSHQISLDNP DYQQDFFPKE AKPNGIFKGS TAENAEYLRV

1210
APQSSEFIGA

compbias

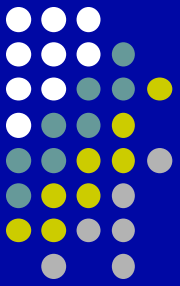


Πληροφοριακά στοιχεία ακολουθίας



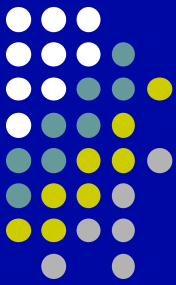
- MOD_RES – τα νούμερα δείχνουν τα αμινοξέα που υπόκεινται φωσφορυλίωση
- DISULPHID – τα νούμερα δείχνουν ζευγάρια κυστεΐνης που σχηματίζουν δεσμό στην ώριμη πρωτεΐνη
- VARIANT – το πεδίο αυτό δείχνει παραλλαγή της ακολουθίας της EGFR πρωτεΐνης (οι περισσότερες έχουν συσχετιστεί με την εμφάνιση του καρκίνου του πνεύμονα)
- MUTAGEN – το πεδίο αυτό χρησιμοποιείται για να καταγράψει τις αλλαγές στην ακολουθία της πρωτεΐνης οι οποίες έχουν εισαχθεί πειραματικά
- CONFLICT – δείχνει τις αποκλίσεις που υπάρχουν σε διάφορες πηγές της πρωτεϊνικής ακολουθίας, δηλαδή λάθη ή πολυμορφισμούς





ΑΝΑΛΥΣΗ ΝΟΥΚΛΕΟΤΙΔΙΚΗΣ ΚΑΙ ΠΡΩΤΕΪΝΙΚΗΣ ΑΚΟΛΟΥΘΙΑΣ

Σχεδιασμός PCR εκκινήτων



εισάγουμε τη διεύθυνση www.biotoools.umassmed.edu

Primer3Plus

pick primers from a DNA sequence

[Primer3Manager](#) [Help](#)
[About](#) [Source Code](#)

Task: Detection *Select primer pairs to detect the given template sequence. Optionally targets and included/excluded regions can be specified.*

Main | General Settings | Advanced Settings | Internal Oligo | Penalty Weights | Sequence Quality

Sequence Id:

Paste source sequence below Or upload sequence file:

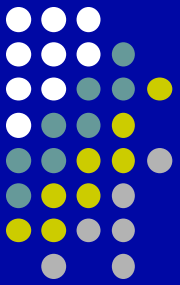
Mark selected region:

Excluded Regions: < >
Targets: []
Included Region: { }

<input checked="" type="checkbox"/> Pick left primer or use left primer below.	<input type="checkbox"/> Pick hybridization probe (internal oligo) or use oligo below.	<input checked="" type="checkbox"/> Pick right primer or use right primer below (5'→3' on opposite strand).
<input type="text"/>	<input type="text"/>	<input type="text"/>



Ανάλυση σύνθεσης DNA



www.genomatix.de/cgi-bin/tools/tools.pl

Inspecting sequence AA03518 [[U03518](#)] (1 - 237):

[DNA] [Aspergillus awamori internal transcribed spacer 1 (ITS1) and 18S rRNA and 5.8S rRNA genes, partial sequence.]

Base-Content:

The AT-content of this sequence is 39.24%, the GC-content 60.76%, others: 0.00%

Mono-Nucleotides:

	in bps	in %
A	41	17.30%
C	77	32.49%
G	67	28.27%
T	52	21.94%

Di-Nucleotides:

		second nucleotide			
		A	C	G	T
first nucl.	A	14	11	7	9
	C	11	31	19	15
	G	11	20	22	14
	T	4	15	19	14

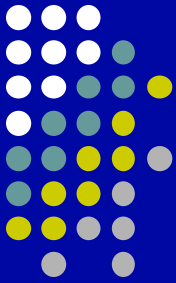
Tri-Nucleotides:

		second/third nucleotide															
		AA	AC	AG	AT	CA	CC	CG	CT	GA	GC	GG	GT	TA	TC	TG	TT
first nucl.	A	3	6	2	3	3	5	1	2	1	1	1	4	0	4	2	3
	C	5	2	2	2	3	12	11	5	2	5	9	3	1	3	6	5
	G	4	1	3	3	2	10	5	2	4	7	9	2	1	5	4	4
	T	1	2	0	1	3	4	2	6	4	7	3	5	2	3	7	2



Genome scan

genes.mit.edu/genomescan/



Run GenomeScan:

Organism:

Sequence name (optional)

Print option:

Upload your DNA sequence file (one-letter code, upper or lower case, spaces/numbers ignored):

Or paste your DNA sequence here (one-letter code, upper or lower case, spaces/numbers ignored):

```
GGTAGGAAAGGCGGGGAGGGGCTCCGGCCGCTCTGGAAAGGAATCCAACGCGGCTTGAGGCTGTGGGGAAG
TAGGGTGGCGAGCGSTCCTTCTGCGCGCGGGGGGCGCGGGGGTGGGGTGGTCCATTAGGGTCCCTGGC
GAGGGGGCGGCTTCTAGTGTGTGAAGGCGACGCCCTAGAAGCTCCCTTCAAAGTTGGCCCCACGCGCT
GAATGTGGAAAGTTGACTGGGACCCAGTAGTTTCCCATCCCAAACCTGCTTTCGAGAGAGGGCTTCAAAC
CCAAAATGTGAATCCCGCTCCCTCTCAGCCAGAACTGTGACTCGTCCCGGGGAGGGGGCGGTGGGTGG
GGCGGGGCTGGCGGGAATTTGSGTITGGCGCGCTCCCTGCGGCGACGCTCATCGTGGCTCTCCTCTT
CCCCCGGTGGTCTCTCGCTCGCTTCTGGCTCTGCCATGCCCCCTGCTCTGAAGAGACAACCCGCCATTTCA
CCCASTAAAGCGGGCCCGGCTGCGGAGGTGGCGGGCATGCAGCTCCGCTTTGCCCGGCTCTCCGAGCAG
CCACGGCCCCCACCCGGGGCTCCGCGCGCGCGCGGGCTACGACCTGTACAGGTGAGCGGGGACCTGCGG
GCGAGGAGGCTGGGAAGGCGCGGCCGCTGCGCCACAGCTAGAAACAGTCAACGGAGAGATCACAGGAA
CACACTAGCTATAAATAGGATTTCTGCCITTTTGGTITAAAATTTAGCTTTCATCTTTGGCATAAAT
TAAATAGAGATTTGGGCAAGACTGCAGAAATAAGTAAAATAGCTATACGGTGTCTAGCAAGGCGTTACTT
```

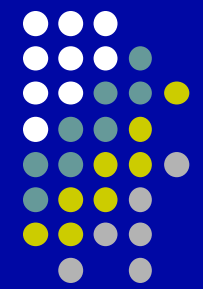
Upload protein(s) with similarity to genomic sequence (FastA format):

Or paste your protein(s) here (FastA format):

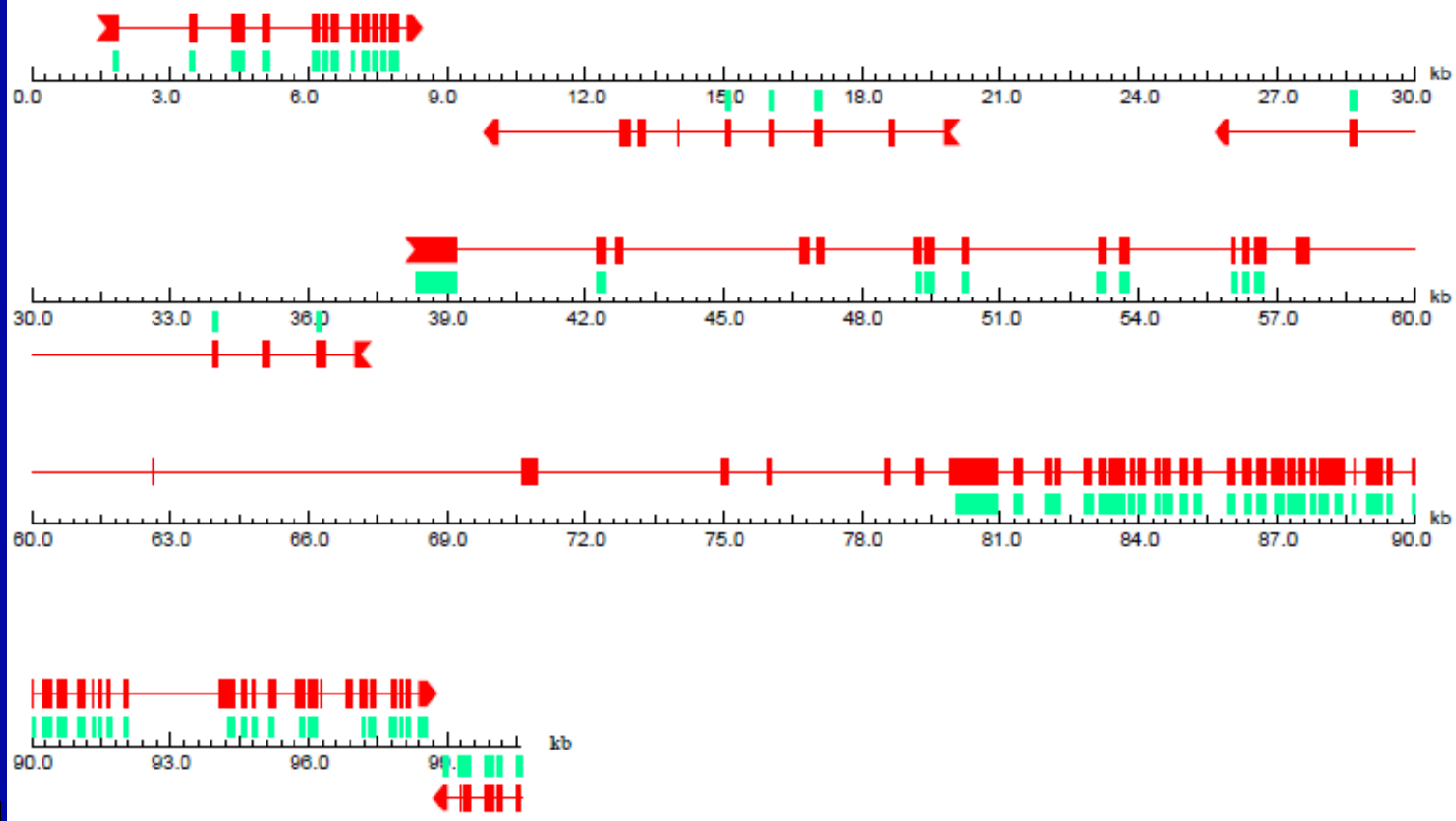
```
>sp|P33316|DUT_HUMAN Deoxyuridine 5'-triphosphate nucleotidohydrolase,
mitochondrial OS=Homo sapiens GN=DUT PE=1 SV=3
MTPLCPRPALCYHFLTSLLRSAMQNARGTAEGRSRGTLRARPAFRPFAAQHGI PRPLSSA
GRLSQGCRGASTVGAAGWKGELPKAGGSPAPGPETPAISPSHRARP AEVGGMQLR FARLS
EHATAPTRGSARAAGYDLYSAYDYTIIPMEKAVVKTDIQIALPSGCYGRVA PRSGLAAKH
FIDVAGVIDE DYRGNVGVVLENFPGKEKFEVHKGDRIAQLICERIFYPEIEEVQALDDTE
RSGGGFSGTGRN
```

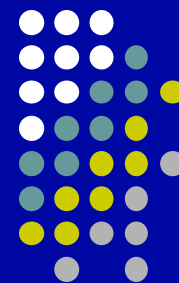
To have the results mailed to you, enter your email address here (optional)





GenomeScan predicted genes in sequence gb

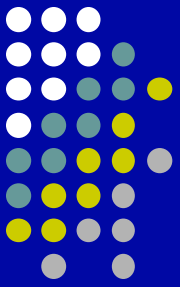




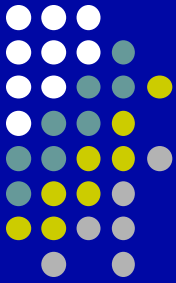
Δωρεάν διαδικτυακοί τόποι για αναζήτηση σημάτων σε ακολουθίες DNA

Εταιρία/Πρόγραμμα	URL	Περιγραφή
Zhang Lab, Cold Spring Harbor, USA	rulai.cshl.edu/	Αναζήτηση για εκκινητές και διάφορα άλλα συντηρημένα στοιχεία
Center for Information Technology, NIH, USA; WWW Signal Scan	bimas.dcrct.nih.gov/molbio/signal/	Αναζήτηση για μεταγραφικά στοιχεία
Center for Information Technology, NIH, USA; WWW Promoter Scan	bimas.dcrct.nih.gov/molbio/proscan/	Πρόβλεψη ευκαρυωτικών περιοχών εκκινητών
Munich Bioinformatics Center, GER, CREDO	mips.gsf.de/services/analysis	Εντοπισμός <i>cis</i> -regulatory στοιχείων
San Diego Supercomputer Center, USA; MEME & MAST	meme.sdsc.edu/meme/intro.html	Ανακάλυψη μοτίβων σε ομάδες σχετιζόμενες με πρωτεϊνικές ή DNA ακολουθίες
Université Libre de Bruxelles, Belgium	rsat.ulb.ac.be/rsat/	Ανάλυση ρυθμιστικών ακολουθιών





ΑΝΑΖΗΤΗΣΕΙΣ ΟΜΟΙΟΤΗΤΑΣ ΑΚΟΛΟΥΘΙΩΝ ΣΕ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ



BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

- [Human](#)
- [Mouse](#)
- [Rat](#)
- [Arabidopsis thaliana](#)
- [Oryza sativa](#)
- [Bos taurus](#)
- [Danio rerio](#)
- [Drosophila melanogaster](#)
- [Gallus gallus](#)
- [Pan troglodytes](#)
- [Microbes](#)
- [Apis mellifera](#)

Basic BLAST

Choose a BLAST program to run.

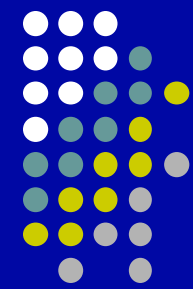
- | | |
|----------------------------------|--|
| nucleotide blast | Search a nucleotide database using a nucleotide query
<i>Algorithms: blastn, megablast, discontinuous megablast</i> |
| protein blast | Search protein database using a protein query
<i>Algorithms: blastp, psi-blast, phi-blast</i> |
| blastx | Search protein database using a translated nucleotide query |
| tblastn | Search translated nucleotide database using a protein query |
| tblastx | Search translated nucleotide database using a translated nucleotide query |

Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- Make specific primers with [Primer-BLAST](#)
- Search [trace archives](#)
- Find [conserved domains](#) in your sequence (cds)
- Find sequences with similar [conserved domain architecture](#) (cdart)
- Search sequences that have [gene expression profiles](#) (GEO)
- Search [immunoglobulins](#) (IgBLAST)
- Search for [SNPs](#) (snp)
- Screen sequence for [vector contamination](#) (vecscreen)
- [Align two \(or more\) sequences using BLAST \(b2seq\)](#)
- Search [protein](#) or [nucleotide](#) targets in PubChem BioAssay
- Search SRA [transcript libraries](#)
- Constraint Based Protein [Multiple Alignment Tool](#)





Enter Query Sequence

Enter accession number, gi, or FASTA sequence [?](#) [Clear](#)

Query subrange [?](#)

From

To

Or, upload file

 [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Database

 [?](#)

Organism

Optional

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Entrez Query

Optional

Enter an Entrez query to limit search [?](#)

Program Selection

Algorithm

- blastp (protein-protein BLAST)
- PSI-BLAST (Position-Specific Iterated BLAST)
- PHI-BLAST (Pattern Hit Initiated BLAST)

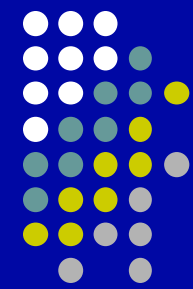
Choose a BLAST algorithm [?](#)

BLAST

Search database nusing Blastp (protein-protein BLAST)

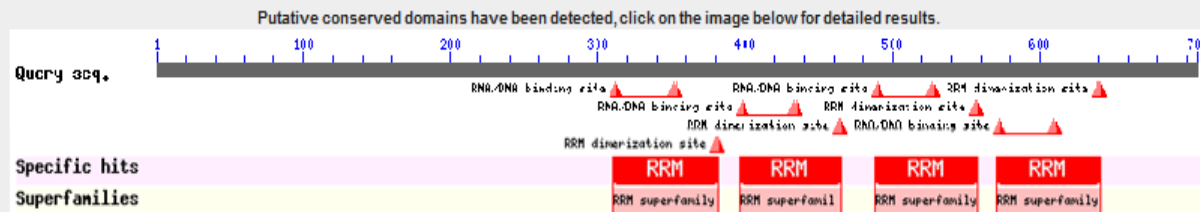
Show results in a new window





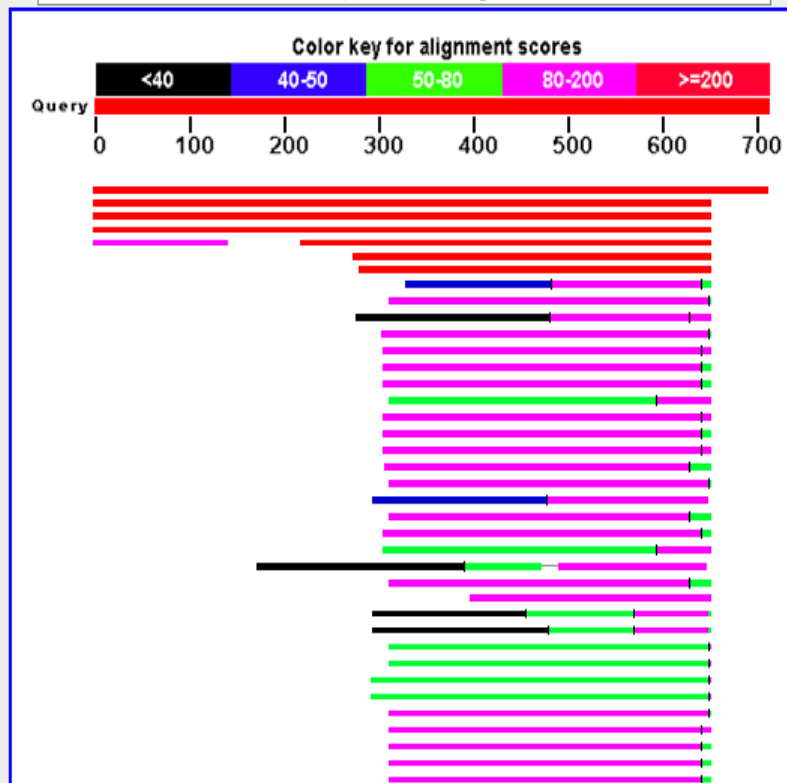
Graphic Summary

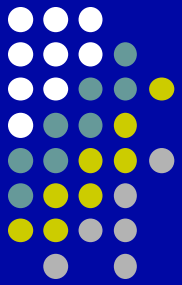
Show Conserved Domains



Distribution of 228 Blast Hits on the Query Sequence

Mouse-over to show define and scores, click to show alignments

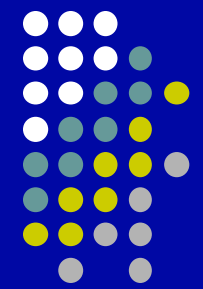




Descriptions

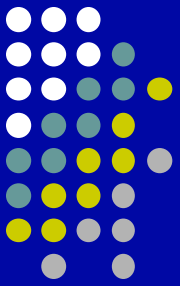
Sequences producing significant alignments:			Score (Bits)	E Value
sp P09405.2 NUCL_MOUSE	nucleolin [Mus musculus] >gi 128843 sp...		1379	0.0
sp P13383.3 NUCL_RAT	RecName: Full=Nucleolin; AltName: Full=P...		941	0.0
sp P08199.2 NUCL_MESAU	RecName: Full=Nucleolin; AltName: Full...		919	0.0
sp Q4R4J7.3 NUCL_MACFA	RecName: Full=Nucleolin		768	0.0
sp P19338.3 NUCL_HUMAN	RecName: Full=Nucleolin; AltName: Full...		657	0.0
sp P15771.1 NUCL_CHICK	RecName: Full=Nucleolin; AltName: Full...		448	5e-125
sp P20397.3 NUCL_XENLA	RecName: Full=Nucleolin; AltName: Full...		391	9e-108
sp Q5RF26.3 NUCL_PONAB	RecName: Full=Nucleolin		136	4e-31
sp Q13310.1 PABP4_HUMAN	RecName: Full=Polyadenylate-binding p...		104	2e-21
sp A5DW14.1 PABP_LODEL	RecName: Full=Polyadenylate-binding pr...		104	2e-21
sp P04147.4 PABP_YEAST	RecName: Full=Polyadenylate-binding pr...		101	2e-20
sp P31209.2 PABP_SCHPO	RecName: Full=Polyadenylate-binding pr...		100	3e-20
sp Q9EPH8.1 PABP1_RAT	RecName: Full=Polyadenylate-binding pro...		100	4e-20
sp Q6IP09.1 PABPB_XENLA	RecName: Full=Polyadenylate-binding p...		99.4	8e-20
sp Q5R8F7.1 PABP1_PONAB	RecName: Full=Polyadenylate-binding p...		99.0	8e-20
sp Q1DXH0.1 PABP_COCIM	RecName: Full=Polyadenylate-binding pr...		99.0	1e-19
sp P11940.2 PABP1_HUMAN	RecName: Full=Polyadenylate-binding p...		98.6	1e-19
sp Q9H361.2 PABP3_HUMAN	RecName: Full=Polyadenylate-binding p...		98.6	1e-19
sp P29341.1 PABP1_MOUSE	RecName: Full=Polyadenylate-binding p...		97.8	2e-19
sp Q6FKG4.1 PABP_CANGA	RecName: Full=Polyadenylate-binding pr...		96.3	6e-19
sp Q6BI95.2 PABP_DEBHA	RecName: Full=Polyadenylate-binding pr...		96.3	6e-19
sp P42731.1 PABP2_ARATH	RecName: Full=Polyadenylate-binding p...		95.9	8e-19
sp Q6CSV3.1 PABP_KLULA	RecName: Full=Polyadenylate-binding pr...		94.0	3e-18
sp P20965.3 PABPA_XENLA	RecName: Full=Polyadenylate-binding p...		93.2	5e-18
sp A4QUF0.1 PABP_MAGGR	RecName: Full=Polyadenylate-binding pr...		91.7	1e-17
sp P32588.3 PUB1_YEAST	RecName: Full=Nuclear and cytoplasmic ...		91.7	1e-17
sp Q74ZS6.1 PABP_ASHGO	RecName: Full=Polyadenylate-binding pr...		91.3	2e-17
sp Q2UK72.1 PABP_ASPOR	RecName: Full=Polyadenylate-binding pr...		91.3	2e-17
sp Q96DU9.1 PABP5_HUMAN	RecName: Full=Polyadenylate-binding p...		91.3	2e-17
sp Q7JGR2.1 PABP5_MACMU	RecName: Full=Polyadenylate-binding p...		90.5	3e-17
sp Q0CR95.1 PABP_ASPTN	RecName: Full=Polyadenylate-binding pr...		89.7	6e-17
sp A1D4K4.1 PABP_NEOF1	RecName: Full=Polyadenylate-binding pr...		89.7	6e-17
sp Q4WK03.1 PABP_ASPFU	RecName: Full=Polyadenylate-binding pr...		89.4	7e-17
sp A1CRM1.1 PABP_ASPCL	RecName: Full=Polyadenylate-binding pr...		89.0	1e-16
sp Q5AI15.1 PABP_CANAL	RecName: Full=Polyadenylate-binding pr...		88.2	2e-16
sp Q6DEY7.1 EPAB_XENTR	RecName: Full=Embryonic polyadenylate-...		87.8	2e-16
sp Q6GR16.1 EPABB_XENLA	RecName: Full=Embryonic polyadenylate...		87.4	3e-16
sp Q4VXU2.1 PAP1L_HUMAN	RecName: Full=Polyadenylate-binding p...		86.7	4e-16



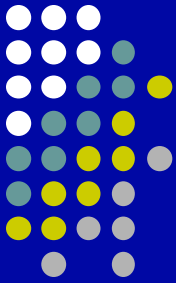


sp A2Q848.1 PABP ASPNC	RecName: Full=Polyadenylate-binding pr...	86.3	7e-16	G
sp Q5KBW2.1 PABP CRYNE	RecName: Full=Polyadenylate-binding pr...	85.9	8e-16	
sp Q05196.2 PABP5 ARATH	RecName: Full=Polyadenylate-binding p...	85.1	1e-15	G
sp A5DM21.2 PABP PICGU	RecName: Full=Polyadenylate-binding pr...	84.3	2e-15	
sp Q7S6N6.1 PABP NEUCR	RecName: Full=Polyadenylate-binding pr...	84.3	3e-15	
sp A3LXL0.1 PABP PICST	RecName: Full=Polyadenylate-binding pr...	83.6	4e-15	
sp P21187.3 PABP DROME	RecName: Full=Polyadenylate-binding pr...	82.4	8e-15	
sp Q0U1G2.3 PABP PHANO	RecName: Full=Polyadenylate-binding pr...	82.4	1e-14	
sp Q2GSX8.1 PABP CHAGE	RecName: Full=Polyadenylate-binding pr...	82.4	1e-14	
sp Q4P8R9.1 PABP USTMA	RecName: Full=Polyadenylate-binding pr...	81.3	2e-14	
sp Q5B630.2 PABP EMENI	RecName: Full=Polyadenylate-binding pr...	80.5	3e-14	
sp Q54BM2.1 PAP1A DICDI	RecName: Full=Polyadenylate-binding p...	79.7	5e-14	
sp P27476.1 NSR1 YEAST	RecName: Full=Nuclear localization seq...	77.4	3e-13	G
sp P19684.1 ROC5 NICSY	RecName: Full=33 kDa ribonucleoprotein...	77.0	3e-13	
sp P28644.1 ROC1 SPIOL	RecName: Full=28 kDa ribonucleoprotein...	77.0	4e-13	
sp O64380.1 PABP3 ARATH	RecName: Full=Polyadenylate-binding p...	75.9	9e-13	G
sp Q6CDH3.1 PABP YARLI	RecName: Full=Polyadenylate-binding pr...	74.3	3e-12	G
sp Q9ZQA8.1 PABPX ARATH	RecName: Full=Probable polyadenylate-...	73.2	5e-12	G
sp Q8R3C6.1 RBM19 MOUSE	RecName: Full=Probable RNA-binding pr...	70.9	2e-11	G
sp P41891.2 GAR2 SCHPO	RecName: Full=Protein gar2	70.1	4e-11	G
sp Q1ZXC2.1 PAP1B DICDI	RecName: Full=Polyadenylate-binding p...	68.9	1e-10	G
sp P53883.1 NOP13 YEAST	RecName: Full=Nucleolar protein 13	68.6	1e-10	G
sp O74400.1 YOCE SCHPO	RecName: Full=Uncharacterized RNA-bind...	65.1	1e-09	G
sp Q2HJ60.1 ROA2 BOVIN	RecName: Full=Heterogeneous nuclear ri...	63.2	6e-09	G
sp Q5RBU8.1 ROA2 PONAB	RecName: Full=Heterogeneous nuclear ri...	63.2	6e-09	G
sp O88569.2 ROA2 MOUSE	RecName: Full=Heterogeneous nuclear ri...	63.2	6e-09	G
sp P22626.2 ROA2 HUMAN	RecName: Full=Heterogeneous nuclear ri...	63.2	6e-09	G
sp A7VJC2.1 ROA2 RAT	RecName: Full=Heterogeneous nuclear ribo...	62.8	8e-09	G
sp P51990.1 RO22 XENLA	RecName: Full=Heterogeneous nuclear ri...	62.4	1e-08	G
sp Q9TTV2.1 ROA2 SAGOE	RecName: Full=Heterogeneous nuclear ri...	61.6	2e-08	G
sp P04256.3 ROA1 RAT	RecName: Full=Heterogeneous nuclear ribo...	60.8	3e-08	G
sp P49312.2 ROA1 MOUSE	RecName: Full=Heterogeneous nuclear ri...	60.5	3e-08	G
sp Q32P51.2 RA1L2 HUMAN	RecName: Full=Heterogeneous nuclear r...	60.5	4e-08	G
sp P0C7M2.1 RA1L3 HUMAN	RecName: Full=Putative heterogeneous ...	60.5	4e-08	G
sp P09651.4 ROA1 HUMAN	RecName: Full=Heterogeneous nuclear ri...	60.5	4e-08	G
sp Q14011.1 CIRBP HUMAN	RecName: Full=Cold-inducible RNA-bind...	60.1	5e-08	G
sp P60826.1 CIRBP CRIGR	RecName: Full=Cold-inducible RNA-bind...	59.3	8e-08	G
sp Q28521.3 ROA1 MACMU	RecName: Full=Heterogeneous nuclear ri...	59.3	9e-08	G
sp Q9SVM8.1 GRP2 ARATH	RecName: Full=Glycine-rich RNA-binding...	58.9	1e-07	G
sp P48810.2 RB87F DROME	RecName: Full=Heterogeneous nuclear r...	58.9	1e-07	G
sp Q99070.1 GRP2 SORBI	RecName: Full=Glycine-rich RNA-binding...	58.2	2e-07	G
sp P17130.1 ROA1 XENLA	RecName: Full=Heterogeneous nuclear ri...	57.8	3e-07	G
sp Q03878.1 GRP1 DAUCA	RecName: Full=Glycine-rich RNA-binding...	57.4	3e-07	G
sp P51992.1 RO32 XENLA	RecName: Full=Heterogeneous nuclear ri...	57.4	3e-07	G





ΑΝΑΖΗΤΗΣΕΙΣ ΟΜΟΙΟΤΗΤΑΣ ΑΚΟΛΟΥΘΙΩΝ ΣΕ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ



BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

- [Human](#)
- [Mouse](#)
- [Rat](#)
- [Arabidopsis thaliana](#)
- [Oryza sativa](#)
- [Bos taurus](#)
- [Danio rerio](#)
- [Drosophila melanogaster](#)
- [Gallus gallus](#)
- [Pan troglodytes](#)
- [Microbes](#)
- [Apis mellifera](#)

Basic BLAST

Choose a BLAST program to run.

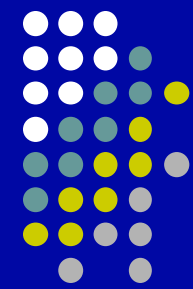
- | | |
|----------------------------------|--|
| nucleotide blast | Search a nucleotide database using a nucleotide query
<i>Algorithms: blastn, megablast, discontinuous megablast</i> |
| protein blast | Search protein database using a protein query
<i>Algorithms: blastp, psi-blast, phi-blast</i> |
| blastx | Search protein database using a translated nucleotide query |
| tblastn | Search translated nucleotide database using a protein query |
| tblastx | Search translated nucleotide database using a translated nucleotide query |

Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- Make specific primers with [Primer-BLAST](#)
- Search [trace archives](#)
- Find [conserved domains](#) in your sequence (cds)
- Find sequences with similar [conserved domain architecture](#) (cdart)
- Search sequences that have [gene expression profiles](#) (GEO)
- Search [immunoglobulins](#) (IgBLAST)
- Search for [SNPs](#) (snp)
- Screen sequence for [vector contamination](#) (vecscreen)
- [Align two \(or more\) sequences using BLAST \(b2seq\)](#)
- Search [protein](#) or [nucleotide](#) targets in PubChem BioAssay
- Search SRA [transcript libraries](#)
- Constraint Based Protein [Multiple Alignment Tool](#)





Enter Query Sequence

Enter accession number, gi, or FASTA sequence [?](#) [Clear](#)

Query subrange [?](#)

From

To

Or, upload file

 [Browse...](#) [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

[Align two or more sequences](#) [?](#)

Choose Search Set

Database

Non-redundant protein sequences (nr) [?](#)

Organism

Optional

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Entrez Query

Optional

Enter an Entrez query to limit search [?](#)

Program Selection

Algorithm

- blastp (protein-protein BLAST)
- PSI-BLAST (Position-Specific Iterated BLAST)
- PHI-BLAST (Pattern Hit Initiated BLAST)

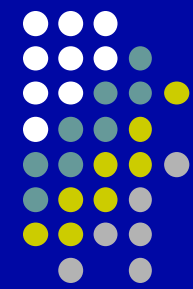
Choose a BLAST algorithm [?](#)

BLAST

Search database nusing Blastp (protein-protein BLAST)

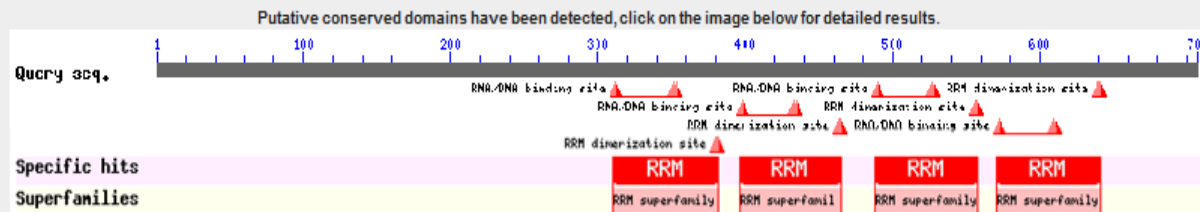
Show results in a new window





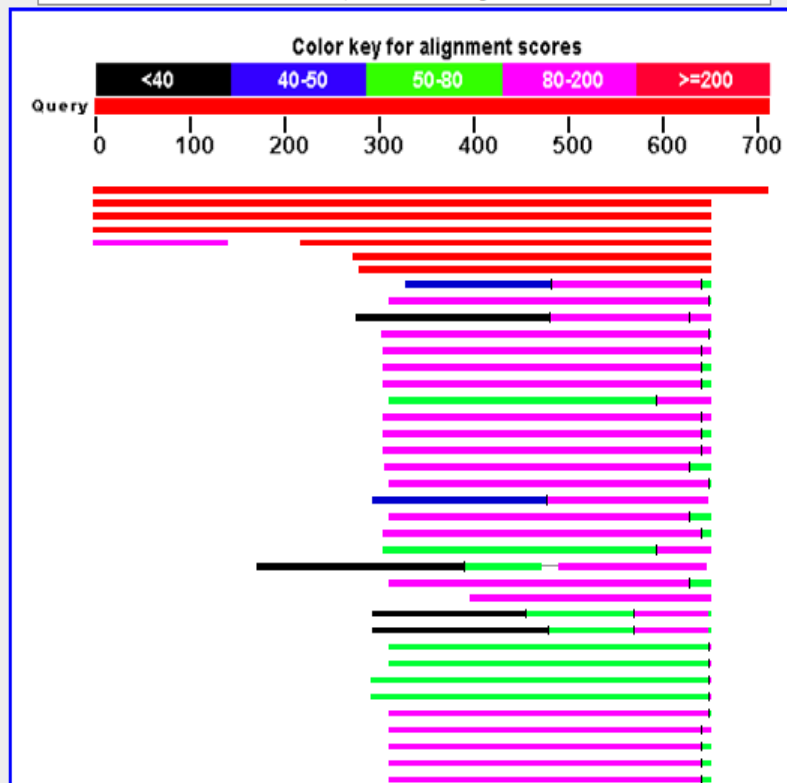
Graphic Summary

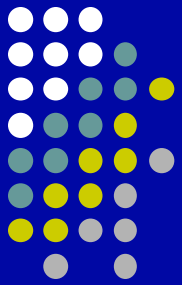
Show Conserved Domains



Distribution of 228 Blast Hits on the Query Sequence

Mouse-over to show define and scores, click to show alignments

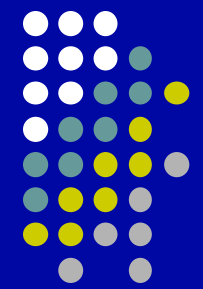




Descriptions

Sequences producing significant alignments:			Score (Bits)	E Value
sp P09405.2 NUCL_MOUSE	nucleolin [Mus musculus] >gi 128843 sp...		1379	0.0
sp P13383.3 NUCL_RAT	RecName: Full=Nucleolin; AltName: Full=P...		941	0.0
sp P08199.2 NUCL_MESAU	RecName: Full=Nucleolin; AltName: Full...		919	0.0
sp Q4R4J7.3 NUCL_MACFA	RecName: Full=Nucleolin		768	0.0
sp P19338.3 NUCL_HUMAN	RecName: Full=Nucleolin; AltName: Full...		657	0.0
sp P15771.1 NUCL_CHICK	RecName: Full=Nucleolin; AltName: Full...		448	5e-125
sp P20397.3 NUCL_XENLA	RecName: Full=Nucleolin; AltName: Full...		391	9e-108
sp Q5RF26.3 NUCL_PONAB	RecName: Full=Nucleolin		136	4e-31
sp Q13310.1 PABP4_HUMAN	RecName: Full=Polyadenylate-binding p...		104	2e-21
sp A5DW14.1 PABP_LODEL	RecName: Full=Polyadenylate-binding pr...		104	2e-21
sp P04147.4 PABP_YEAST	RecName: Full=Polyadenylate-binding pr...		101	2e-20
sp P31209.2 PABP_SCHPO	RecName: Full=Polyadenylate-binding pr...		100	3e-20
sp Q9EPH8.1 PABP1_RAT	RecName: Full=Polyadenylate-binding pro...		100	4e-20
sp Q6IP09.1 PABPB_XENLA	RecName: Full=Polyadenylate-binding p...		99.4	8e-20
sp Q5R8F7.1 PABP1_PONAB	RecName: Full=Polyadenylate-binding p...		99.0	8e-20
sp Q1DXH0.1 PABP_COCIM	RecName: Full=Polyadenylate-binding pr...		99.0	1e-19
sp P11940.2 PABP1_HUMAN	RecName: Full=Polyadenylate-binding p...		98.6	1e-19
sp Q9H361.2 PABP3_HUMAN	RecName: Full=Polyadenylate-binding p...		98.6	1e-19
sp P29341.1 PABP1_MOUSE	RecName: Full=Polyadenylate-binding p...		97.8	2e-19
sp Q6FKG4.1 PABP_CANGA	RecName: Full=Polyadenylate-binding pr...		96.3	6e-19
sp Q6BI95.2 PABP_DEBHA	RecName: Full=Polyadenylate-binding pr...		96.3	6e-19
sp P42731.1 PABP2_ARATH	RecName: Full=Polyadenylate-binding p...		95.9	8e-19
sp Q6CSV3.1 PABP_KLULA	RecName: Full=Polyadenylate-binding pr...		94.0	3e-18
sp P20965.3 PABPA_XENLA	RecName: Full=Polyadenylate-binding p...		93.2	5e-18
sp A4QUF0.1 PABP_MAGGR	RecName: Full=Polyadenylate-binding pr...		91.7	1e-17
sp P32588.3 PUB1_YEAST	RecName: Full=Nuclear and cytoplasmic ...		91.7	1e-17
sp Q74ZS6.1 PABP_ASHGO	RecName: Full=Polyadenylate-binding pr...		91.3	2e-17
sp Q2UK72.1 PABP_ASPOR	RecName: Full=Polyadenylate-binding pr...		91.3	2e-17
sp Q96DU9.1 PABP5_HUMAN	RecName: Full=Polyadenylate-binding p...		91.3	2e-17
sp Q7JGR2.1 PABP5_MACMU	RecName: Full=Polyadenylate-binding p...		90.5	3e-17
sp Q0CR95.1 PABP_ASPTN	RecName: Full=Polyadenylate-binding pr...		89.7	6e-17
sp A1D4K4.1 PABP_NEOF1	RecName: Full=Polyadenylate-binding pr...		89.7	6e-17
sp Q4WK03.1 PABP_ASPFU	RecName: Full=Polyadenylate-binding pr...		89.4	7e-17
sp A1CRM1.1 PABP_ASPCL	RecName: Full=Polyadenylate-binding pr...		89.0	1e-16
sp Q5AI15.1 PABP_CANAL	RecName: Full=Polyadenylate-binding pr...		88.2	2e-16
sp Q6DEY7.1 EPAB_XENTR	RecName: Full=Embryonic polyadenylate-...		87.8	2e-16
sp Q6GR16.1 EPABB_XENLA	RecName: Full=Embryonic polyadenylate...		87.4	3e-16
sp Q4VXU2.1 PAP1L_HUMAN	RecName: Full=Polyadenylate-binding p...		86.7	4e-16





sp A2Q848.1 PABP ASPNC	RecName: Full=Polyadenylate-binding pr...	86.3	7e-16	G
sp Q5KBW2.1 PABP CRYNE	RecName: Full=Polyadenylate-binding pr...	85.9	8e-16	
sp Q05196.2 PABP5 ARATH	RecName: Full=Polyadenylate-binding p...	85.1	1e-15	G
sp A5DM21.2 PABP PICGU	RecName: Full=Polyadenylate-binding pr...	84.3	2e-15	
sp Q7S6N6.1 PABP NEUCR	RecName: Full=Polyadenylate-binding pr...	84.3	3e-15	
sp A3LXL0.1 PABP PICST	RecName: Full=Polyadenylate-binding pr...	83.6	4e-15	
sp P21187.3 PABP DROME	RecName: Full=Polyadenylate-binding pr...	82.4	8e-15	
sp Q0U1G2.3 PABP PHANO	RecName: Full=Polyadenylate-binding pr...	82.4	1e-14	
sp Q2GSX8.1 PABP CHAGE	RecName: Full=Polyadenylate-binding pr...	82.4	1e-14	
sp Q4P8R9.1 PABP USTMA	RecName: Full=Polyadenylate-binding pr...	81.3	2e-14	
sp Q5B630.2 PABP EMENI	RecName: Full=Polyadenylate-binding pr...	80.5	3e-14	
sp Q54BM2.1 PAP1A DICDI	RecName: Full=Polyadenylate-binding p...	79.7	5e-14	
sp P27476.1 NSR1 YEAST	RecName: Full=Nuclear localization seq...	77.4	3e-13	G
sp P19684.1 ROC5 NICSY	RecName: Full=33 kDa ribonucleoprotein...	77.0	3e-13	
sp P28644.1 ROC1 SPIOL	RecName: Full=28 kDa ribonucleoprotein...	77.0	4e-13	
sp O64380.1 PABP3 ARATH	RecName: Full=Polyadenylate-binding p...	75.9	9e-13	G
sp Q6CDH3.1 PABP YARLI	RecName: Full=Polyadenylate-binding pr...	74.3	3e-12	G
sp Q9ZQA8.1 PABPX ARATH	RecName: Full=Probable polyadenylate-...	73.2	5e-12	G
sp Q8R3C6.1 RBM19 MOUSE	RecName: Full=Probable RNA-binding pr...	70.9	2e-11	G
sp P41891.2 GAR2 SCHPO	RecName: Full=Protein gar2	70.1	4e-11	G
sp Q1ZX2.1 PAP1B DICDI	RecName: Full=Polyadenylate-binding p...	68.9	1e-10	G
sp P53883.1 NOP13 YEAST	RecName: Full=Nucleolar protein 13	68.6	1e-10	G
sp O74400.1 YOCE SCHPO	RecName: Full=Uncharacterized RNA-bind...	65.1	1e-09	G
sp Q2HJ60.1 ROA2 BOVIN	RecName: Full=Heterogeneous nuclear ri...	63.2	6e-09	G
sp Q5RBU8.1 ROA2 PONAB	RecName: Full=Heterogeneous nuclear ri...	63.2	6e-09	G
sp O88569.2 ROA2 MOUSE	RecName: Full=Heterogeneous nuclear ri...	63.2	6e-09	G
sp P22626.2 ROA2 HUMAN	RecName: Full=Heterogeneous nuclear ri...	63.2	6e-09	G
sp A7VJC2.1 ROA2 RAT	RecName: Full=Heterogeneous nuclear ribo...	62.8	8e-09	G
sp P51990.1 RO22 XENLA	RecName: Full=Heterogeneous nuclear ri...	62.4	1e-08	G
sp Q9TTV2.1 ROA2 SAGOE	RecName: Full=Heterogeneous nuclear ri...	61.6	2e-08	G
sp P04256.3 ROA1 RAT	RecName: Full=Heterogeneous nuclear ribo...	60.8	3e-08	G
sp P49312.2 ROA1 MOUSE	RecName: Full=Heterogeneous nuclear ri...	60.5	3e-08	G
sp Q32P51.2 RA1L2 HUMAN	RecName: Full=Heterogeneous nuclear r...	60.5	4e-08	G
sp P0C7M2.1 RA1L3 HUMAN	RecName: Full=Putative heterogeneous ...	60.5	4e-08	G
sp P09651.4 ROA1 HUMAN	RecName: Full=Heterogeneous nuclear ri...	60.5	4e-08	G
sp Q14011.1 CIRBP HUMAN	RecName: Full=Cold-inducible RNA-bind...	60.1	5e-08	G
sp P60826.1 CIRBP CRIGR	RecName: Full=Cold-inducible RNA-bind...	59.3	8e-08	G
sp Q28521.3 ROA1 MACMU	RecName: Full=Heterogeneous nuclear ri...	59.3	9e-08	G
sp Q9SVM8.1 GRP2 ARATH	RecName: Full=Glycine-rich RNA-binding...	58.9	1e-07	G
sp P48810.2 RB87F DROME	RecName: Full=Heterogeneous nuclear r...	58.9	1e-07	G
sp Q99070.1 GRP2 SORBI	RecName: Full=Glycine-rich RNA-binding...	58.2	2e-07	G
sp P17130.1 ROA1 XENLA	RecName: Full=Heterogeneous nuclear ri...	57.8	3e-07	G
sp Q03878.1 GRP1 DAUCA	RecName: Full=Glycine-rich RNA-binding...	57.4	3e-07	G
sp P51992.1 RO32 XENLA	RecName: Full=Heterogeneous nuclear ri...	57.4	3e-07	G



Basic BLAST

Advanced BLAST

Basic BLAST

Usage: Choose the the suitable BLAST program and database for your query sequence. Paste your sequence in one of the supported [formats](#) into the sequence field below and press the "Run BLAST" button. Don't forget your e-mail address, so that we can send you the results in case of traffic jam...

Make sure that the format button (next to the sequence field) shows the correct format .

See also our [BLAST database description](#).

Please select the program:

blastp

Please select the database:

DNA databases

Please select

Protein databases

Please select

Gapped alignment on/off

blosum62

BLAST filter on/off

SwissProt ID or AC

Graphic output on/off

Query title (option)

Paste your sequence here:
(or ID or accession number)

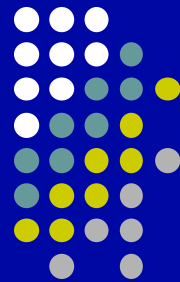
P09405|

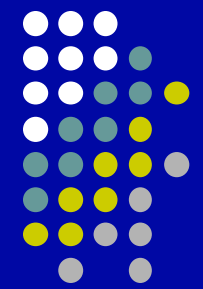
E-mail address

HTML

Run BLAST

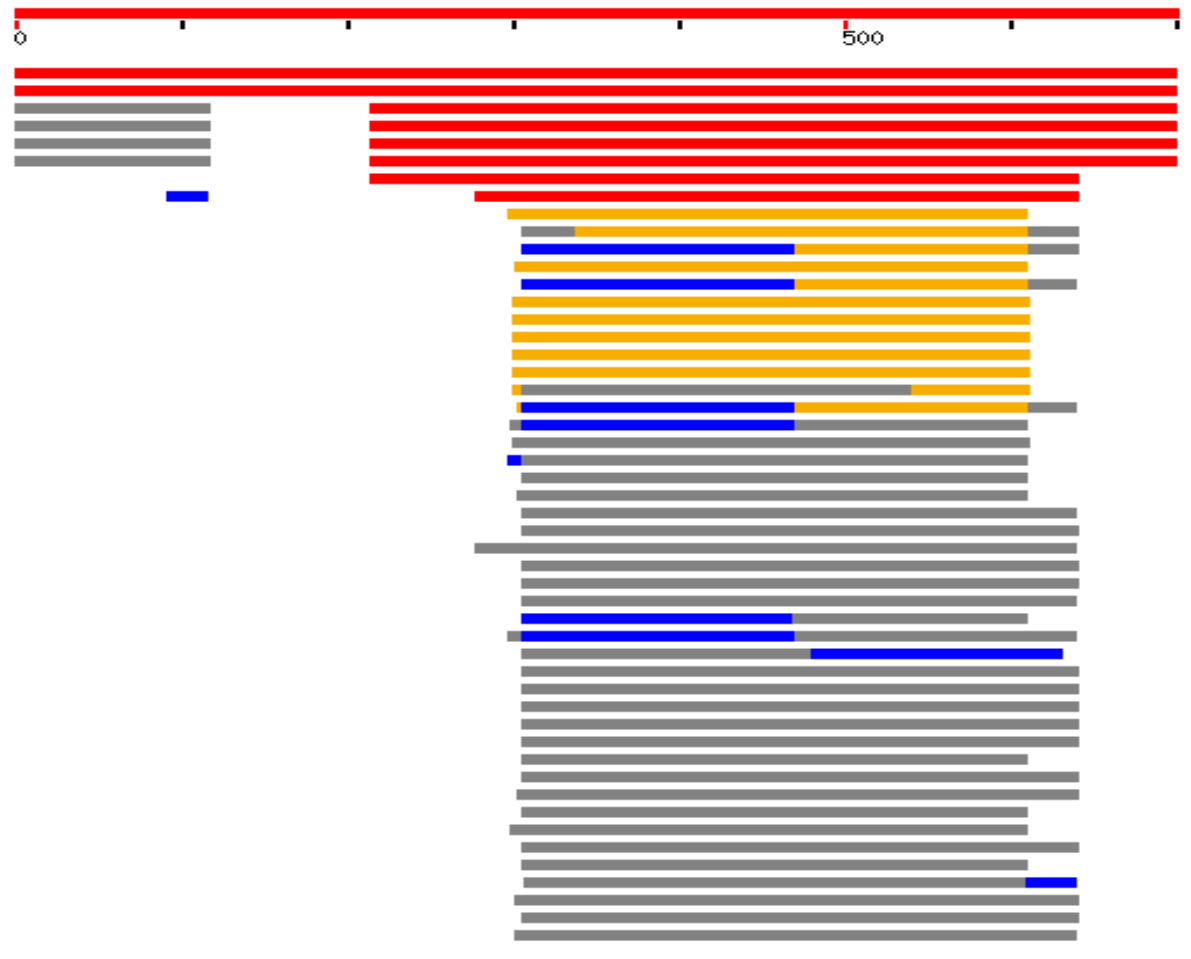
Clear form



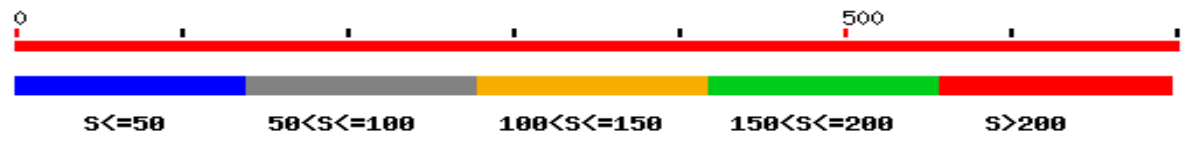


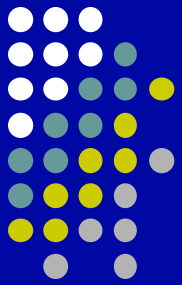
sp

NUCL_MOUSE
NUCL_RAT
NUCL_MESAU
NUCL_HUMAN
NUCL_MACFA
NUCL_POMAB
NUCL_CHICK
NUCL_XENLA
PABP_CANGA
PABP4_HUMAN
PABP_YEAST
PABP3_HUMAN
PABP_LODEL
PABP1_RAT
PABP1_MOUSE
PABP1_POMAB
PABP1_BOVIN
PABP1_HUMAN
PABPB_XENLA
PABP_DEBHA
PABP_SCHPO
PABPA_XENLA
PABP2_ARATH
PABP_KLULA
PABP_ASHGO
PABP_CANAL
EPABB_XENLA
NSR1_YEAST
EPAB_XENTR
EPABA_XENLA
PABP_PICST
PABP_DROME
PABP_PICGU
PAP1A_DICDI
PABP5_MACHU
PABP5_HYLLA
PABP5_GORGO
PABP5_HUMAN
PABP5_PONPY
PAP1L_HUMAN
PABP5_PANTR
PABP_CRYNE
PABP5_ARATH
PABPX_ARATH
PABP_COICM
PAP1B_DICDI
RBM28_MOUSE
PABP_MAGGR
PABP_USTMA
PABP_NEUCR



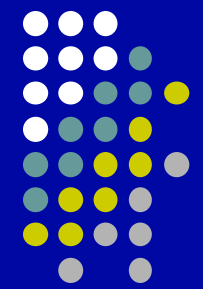
sp





Sequences producing significant alignments:			(bits)	Value
sp	P09405	NUCL_MOUSE (Ncl)RecName: Full=Nucleolin; AltName: Full...	784	0.0
sp	P13383	NUCL_RAT (Ncl)RecName: Full=Nucleolin; AltName: Full=P...	734	0.0
sp	P08199	NUCL_MESAU (NCL)RecName: Full=Nucleolin; AltName: Full...	664	0.0
sp	Q4R4J7	NUCL_MACFA (NCL)RecName: Full=Nucleolin; [Macaca fascic...	596	e-169
sp	P19338	NUCL_HUMAN (NCL)RecName: Full=Nucleolin; AltName: Full...	596	e-169
sp	Q5RF26	NUCL_PONAB (NCL)RecName: Full=Nucleolin; [Pongo abelii]	595	e-169
sp	P15771	NUCL_CHICK (NCL)RecName: Full=Nucleolin; AltName: Full...	423	e-117
sp	P20397	NUCL_XENLA (ncl)RecName: Full=Nucleolin; AltName: Full...	376	e-103
sp	Q6FKG4	PABP_CANGA (PAB1)RecName: Full=Polyadenylate-binding p...	110	3e-23
sp	Q13310	PABP4_HUMAN (PABPC4)RecName: Full=Polyadenylate-bindin...	110	4e-23
sp	P04147	PABP_YEAST (PAB1)RecName: Full=Polyadenylate-binding p...	108	2e-22
sp	Q9H361	PABP3_HUMAN (PABPC3)RecName: Full=Polyadenylate-bindin...	107	4e-22
sp	Q9EPH8	PABP1_RAT (Pabpc1)RecName: Full=Polyadenylate-binding ...	106	5e-22
sp	A5DW14	PABP_LODEL (PAB1)RecName: Full=Polyadenylate-binding p...	106	6e-22
sp	Q5R8F7	PABP1_PONAB (PABPC1)RecName: Full=Polyadenylate-bindin...	105	8e-22
sp	P11940	PABP1_HUMAN (PABPC1)RecName: Full=Polyadenylate-bindin...	105	1e-21
sp	P61286	PABP1_BOVIN (PABPC1)RecName: Full=Polyadenylate-bindin...	105	1e-21
sp	P29341	PABP1_MOUSE (Pabpc1)RecName: Full=Polyadenylate-bindin...	105	1e-21
sp	Q6IP09	PABPB_XENLA (pabpc1-B)RecName: Full=Polyadenylate-bind...	104	2e-21
sp	Q6BI95	PABP_DEBHA (PAB1)RecName: Full=Polyadenylate-binding p...	103	4e-21
sp	Q6CSV3	PABP_KLULA (PAB1)RecName: Full=Polyadenylate-binding p...	100	3e-20
sp	Q74ZS6	PABP_ASHGO (PAB1)RecName: Full=Polyadenylate-binding p...	100	5e-20
sp	P20965	PABPA_XENLA (pabpc1-A)RecName: Full=Polyadenylate-bind...	100	5e-20
sp	P42731	PABP2_ARATH (PAB2)RecName: Full=Polyadenylate-binding ...	100	5e-20
sp	P31209	PABP_SCHPO (pab1)RecName: Full=Polyadenylate-binding p...	100	6e-20
sp	Q5AI15	PABP_CANAL (PAB1)RecName: Full=Polyadenylate-binding p...	96	1e-18
sp	Q6GR16	EPABB_XENLA (epabp-B)RecName: Full=Embryonic polyadeny...	96	1e-18
sp	Q6DEY7	EPAB_XENTR (epabp)RecName: Full=Embryonic polyadenylat...	95	1e-18
sp	P27476	NSR1_YEAST (NSR1)RecName: Full=Nuclear localization se...	95	2e-18
sp	A3LXL0	PABP_PICST (PAB1)RecName: Full=Polyadenylate-binding p...	93	6e-18
sp	Q98SP8	EPABA_XENLA (epabp-A)RecName: Full=Embryonic polyadeny...	93	7e-18
sp	P21187	PABP_DROME (pAbp)RecName: Full=Polyadenylate-binding p...	91	2e-17
sp	A5DM21	PABP_PICGU (PAB1)RecName: Full=Polyadenylate-binding p...	90	6e-17
sp	Q54BM2	PAP1A_DICDI (pabpc1A)RecName: Full=Polyadenylate-bindi...	88	2e-16
sp	Q4VXU2	PAP1L_HUMAN (PABPC1L)RecName: Full=Polyadenylate-bindi...	88	2e-16
sp	P60050	PABP5_PONPY (PABPC5)RecName: Full=Polyadenylate-bindin...	88	2e-16
sp	P60049	PABP5_PANTR (PABPC5)RecName: Full=Polyadenylate-bindin...	88	2e-16
sp	Q7JGR2	PABP5_MACMU (PABPC5)RecName: Full=Polyadenylate-bindin...	88	2e-16
sp	P60048	PABP5_HYLLA (PABPC5)RecName: Full=Polyadenylate-bindin...	88	2e-16
sp	Q96DU9	PABP5_HUMAN (PABPC5)RecName: Full=Polyadenylate-bindin...	88	2e-16
sp	P60047	PABP5_GORGO (PABPC5)RecName: Full=Polyadenylate-bindin...	88	2e-16
sp	Q5KBW2	PABP_CRYNE (PAB1)RecName: Full=Polyadenylate-binding p...	87	3e-16
sp	Q05196	PABP5_ARATH (PAB5)RecName: Full=Polyadenylate-binding ...	86	9e-16





```

> F|NP\_035010.3 | nu UGlin [Mus musculus]
  sp|P09405.2|NUCL\_MOUSE GacName: Full=Nucleolin; AltName: Full=Protein C23
  gb|AAK07920.1|AF318184.1 Gucleolin [Mus musculus]
  @ more sequence titles
  Length=707

  GENE ID: 17975 Ncl | nucleolin [Mus musculus] (Over 10 PubMed links)

  Score = 1379 bits (3569), Expect = 0.0, Method: Compositional matrix adjust.
  Identities = 707/707 (100%), Positives = 707/707 (100%), Gaps = 0/707 (0%)

  Query 1      MVKLAAGKTHGEAKKMAPPPKEVEEDSSEDEEMSEDEDDSSGEEEVVVI P Q K K G K K A T T T P 60
  Sbjct 1      MVKLAAGKTHGEAKKMAPPPKEVEEDSSEDEEMSEDEDDSSGEEEVVVI P Q K K G K K A T T T P 60

  Query 61     AKKVVVSQTKKA AVPT PAKKAAVT PGKKAVAT PAKKNIT PAKVI PT PGKKGAAQAKALVP 120
  Sbjct 61     AKKVVVSQTKKA AVPT PAKKAAVT PGKKAVAT PAKKNIT PAKVI PT PGKKGAAQAKALVP 120

  Query 121    TPGKKGAATPAKGAKNGKNAKKEDSDEDEDEDEDEDDSDDEDEDEDEE DEFEPPIVKGVKPA 180
  Sbjct 121    TPGKKGAATPAKGAKNGKNAKKEDSDEDEDEDEDEDDSDDEDEDEDEE DEFEPPIVKGVKPA 180

  Query 181    KAAPAAPASEDEEDEDDEEDEDDEEEDDDSEEEVMEITTAKGKKT PAKVVPMKAKSVA 240
  Sbjct 181    KAAPAAPASEDEEDEDDEEDEDDEEEDDDSEEEVMEITTAKGKKT PAKVVPMKAKSVA 240

  Query 241    EEEDDEEDEDDEDEDEDDEEEDDDDEEEEEEEEPVKAAPGKRKKEMTKQKEAPEAKKQKV 300
  Sbjct 241    EEEDDEEDEDDEDEDEDDEEEDDDDEEEEEEEEPVKAAPGKRKKEMTKQKEAPEAKKQKV 300

  Query 301    EGSEPTTTFNLFIGNLNPKNKSVNELKFAISELFAKNDLAVVDVRTGTNRKFGYVDFESAE 360
  Sbjct 301    EGSEPTTTFNLFIGNLNPKNKSVNELKFAISELFAKNDLAVVDVRTGTNRKFGYVDFESAE 360

  Query 361    DLEKALELTGLKVFNGNEIKLEKPKGRDSKKVRAARTLLAKNLSFNITEDELKEVFEDAME 420
  Sbjct 361    DLEKALELTGLKVFNGNEIKLEKPKGRDSKKVRAARTLLAKNLSFNITEDELKEVFEDAME 420

  Query 421    IRLVSQDGKSKGIAYIEFKSEADA EKNLEEKQGA EIDGRSVSLYYTGEKGRQERTGKTS 480
  Sbjct 421    IRLVSQDGKSKGIAYIEFKSEADA EKNLEEKQGA EIDGRSVSLYYTGEKGRQERTGKTS 480

  Query 481    TWSGESKTLVLSNLSYSATKETLEEVFEKATFIKVPQNP HGKPKGYAFIEFASFEDA KEA 540
  Sbjct 481    TWSGESKTLVLSNLSYSATKETLEEVFEKATFIKVPQNP HGKPKGYAFIEFASFEDA KEA 540

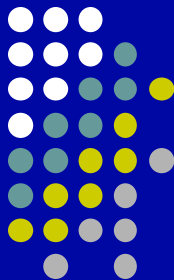
  Query 541    LNSCNKMEIEGRTIRLELQGSNSRSQPSKTLFVKGLSEDTTEETLKE SFEGSVRARIVTD 600
  Sbjct 541    LNSCNKMEIEGRTIRLELQGSNSRSQPSKTLFVKGLSEDTTEETLKE SFEGSVRARIVTD 600



  Query 601    RETGSSKGFVDFNSEEDAKAAKEAMEDGEIDGNKVITLDWAKPKGEGGF GGRGGGRGGF 660
  Sbjct 601    RETGSSKGFVDFNSEEDAKAAKEAMEDGEIDGNKVITLDWAKPKGEGGF GGRGGGRGGF 660

  Query 661    GGRGGGRGGGRGGFGGRGRGGFGGRGGFRGGRGGGGDFKPKQ GKKT KFE 707
  Sbjct 661    GGRGGGRGGGRGGFGGRGRGGFGGRGGFRGGRGGGGDFKPKQ GKKT KFE 707

```





>  P13383.3|NUCL_RAT Re  Game: Full=Nucleolin; AltName: Full=Protein C23
Length=713

GENE ID: 25135 Ncl | nucleolin [Rattus norvegicus] (10 or fewer PubMed links)

Score = 941 bits (2433), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 595/655 (90%), Positives = 633/655 (96%), Gaps = 10/655 (1%)

```

Query 1  MVKLAKAGKTHGEAKKMAPPPKEVEEDSEDEEMSEDEDDSSGEEVVIPQKKGKKATTP 60
Sbjct 1  MVKLAKAGKTHGE+KKMAPPPKEVEEDSEDEEMSEDEDDSSGEEVVIPQKKGKKATTP 60

Query 61  AKKVVVSQTKKAAVPTPAKKAAPTGGKAVATPAKKNITPAKVIPTGGKKGAAQAKALVP 120
Sbjct 61  AKKVVVSQTKKAAVPTPAKKAAPTGGKAAATPAKAVTPAKVPTGGKKGAAQAKALVP 120

Query 121  TPGKKGAATPAKGAKNGKNAKKEDSDEDEDEDEDDSDSEDEDEDEEDEFEPPIVKGVKPA 180
Sbjct 121  TPGKKGA TPAKGAKNGKNAKKEDSDEDEDEDEDDSDSEDEDEDEDEFEPPIVKGVKPA 178

Query 181  KAAPAAPASEDEDEDEDEDEDEDDDEEEDDSEEE---VMEITTAKGKKTPAKVVPKAK 237
Sbjct 179  KAAPAAPASEDEDEDEDEDEDDDDDEEEDDSEEEVMEITPAKGKKTPAKVVPKAK 238

Query 238  SVAEEEDDEEDEDDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 296
Sbjct 239  SVAEEEDDEEDEDDEEDEDDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 298

Query 297  KQKVEGSEPTTPEFNLFIGNLNPNKSVNELKFAISELFAKNDLAVVDVRTGTNRKFGYVDF 356
Sbjct 299  KQKIEGSEPTTPEFNLFIGNLNPNKSV AELKVAISELFAKNDLAAVDVRTGTNRKFGYVDF 358

Query 357  ESAEDLEKALELTGLKVFNEIKLEKPKGRDSKKVRAARTLLAKNLSFNITEDELKEVFE 416
Sbjct 359  ESAEDLEKALELTGLKVFNEIKLEKPKGRDSKKVRAARTLLAKNLSFNITEDELKEVFE 418

Query 417  DAMEIRLVSQDGKSGIAYIEFKSEADAENLEEKQGAEIDGRSVSLYYTGEKGQRQERT 476
Sbjct 419  DAVEIRLVSQDGRSGIAYIEFKSEADAENLEEKQGAEIDGRSVSLYYTGEKGQRQERT 478

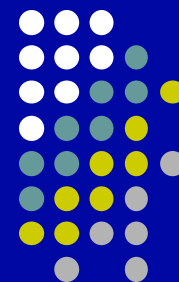
Query 477  GKTSTWSGESKTLVLSNLSYSATKETLEEVFEKATFIKVPQNPHGKPKGYAFIEFASFED 536
Sbjct 479  GKNSTWSGESKTLVLSNLSYSATEETLQEVFEKATFIKVPQNPHGKSKGYAFIEFASFED 538

Query 537  AKEALNSCNKMEIEGRTIRLELQG----SNSRSQPSKTLFVKGLSEDTEETLKESFEFS 592
Sbjct 539  AKEALNSCNKMEIEGRTIRLELQG N+RSQPSKTLFVKGLSEDTEETLKESFEFS 598

Query 593  VRARIVTDRETGSSKGFVDFNSEEDAKAAKEAMEDGEIDGNKVTLTDWAKPKGE 647
Sbjct 599  VRARIVTDRETGSSKGFVDFNSEEDAKAAKEAMEDGEIDGNKVTLTDWAKPKGE 653

```

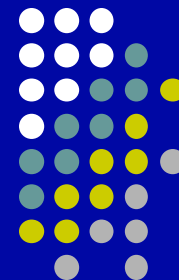




ΕΠΙΛΟΓΗ ΤΩΝ ΚΑΤΑΛΛΗΛΩΝ ΠΡΟΓΡΑΜΜΑΤΩΝ BLAST ΓΙΑ ΤΟ DNA

Ερώτηση	Απάντηση
Ενδιαφέρομαι για μη κωδικοποιήσιμο DNA;	Ναι. Χρησιμοποιείστε blastn. Το blastn είναι μόνο για συγγενικές DNA ακολουθίες. (Περισσότερο από 70% ίδιες)
Θέλω να ανακαλύψω καινούριες πρωτεΐνες;	Ναι. Χρησιμοποιείστε tblastx
Θέλω να ανακαλύψω πρωτεΐνες που κωδικοποιούνται στη δική μου αλληλουχία DNA;	Ναι. Χρησιμοποιείστε blastx
Δεν είμαι σίγουρος για την ποιότητα του DNA που έχω;	Ναι. Χρησιμοποιείστε blastx αν υποψιάζεστε ότι η DNA ακολουθία σας κωδικοποιεί για πρωτεΐνη αλλά μπορεί να περιέχει λάθη (κατά τον προσδιορισμό της αλληλουχίας)





▼ Algorithm parameters

General Parameters

Max target sequences

100 ▼

Select the maximum number of aligned sequences to display ?

Short queries

Automatically adjust parameters for short input sequences ?

Expect threshold

10 ?

Word size

28 ▼ ?

Scoring Parameters

Match/Mismatch Scores

1,-2 ?

Gap Costs

Linear ▼ ?

Filters and Masking

Filter

Low complexity regions ?

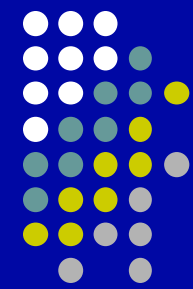
Species-specific repeats for: Human ▼ ?

Mask

Mask for lookup table only ?

Mask lower case letters ?





blastn **blastp** blastx tblastn tblastx

BLASTP programs search protein databases using a protein query. [more...](#)

Enter Query Sequence

Enter accession number, gi, or FASTA sequence [?](#) Clear Query subrange [?](#)

P01922 From
To

Or, upload file Browse... [?](#)

Job Title
Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Database [?](#)

Organism Optional ... suggested

Entrez Query Optional

- Non-redundant protein sequences (nr)
- Reference proteins (refseq_protein)
- Swissprot protein sequences (swissprot) ... id. Only 20 top taxa will be shown. [?](#)
- Patented protein sequences (pat)
- Protein Data Bank proteins (pdb)**
- Environmental samples (env_nr)

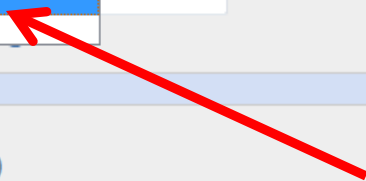
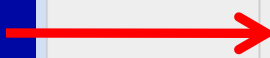
Program Selection

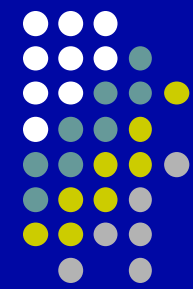
Algorithm

- blastp (protein-protein BLAST)
- PSI-BLAST (Position-Specific Iterated BLAST)**
- PHI-BLAST (Pattern Hit Initiated BLAST)

Choose a BLAST algorithm [?](#)

BLAST Search database **pdb** using **PSI-BLAST (Position-Specific Iterated BLAST)**





PSI blast Iteration 1

P69907:RecName: Full=Hemoglobin subunit alpha;...

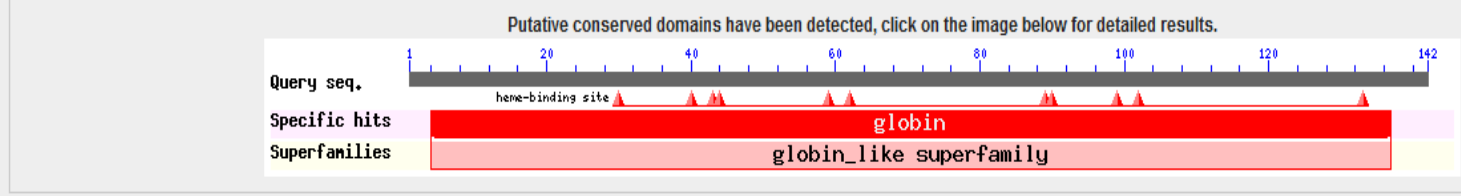
Query ID gi|122412|sp|P01922|HBA_HUMAN
Description Hemoglobin alpha chain.
Molecule type amino acid
Query Length 142

Database Name pdb
Description PDB protein database
Program BLASTP 2.2.21+ [Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [Related Structures](#) [Multiple alignment](#)

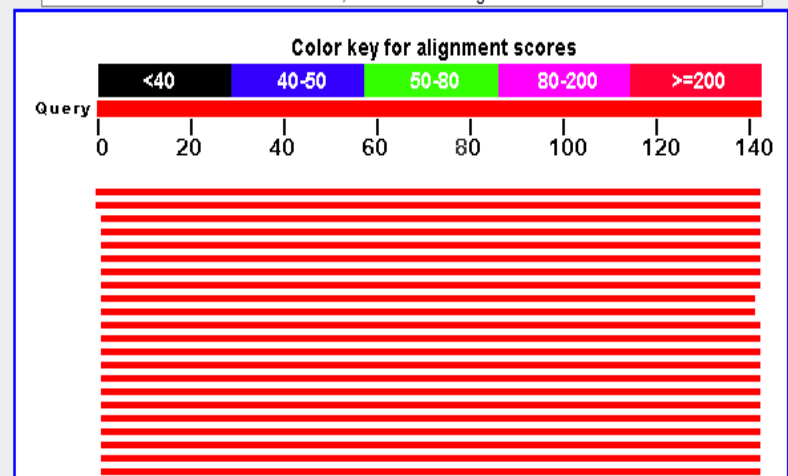
Graphic Summary

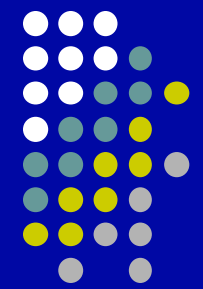
[Show Conserved Domains](#)



Distribution of 106 Blast Hits on the Query Sequence

Mouse-over to show define and scores, click to show alignments

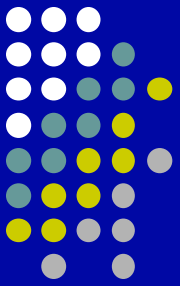




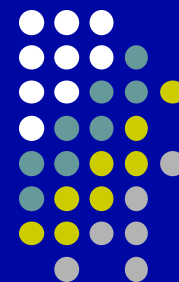
Sequences with E-value BETTER than threshold

Sequences producing significant alignments:				Score	E	
				(Bits)	Value	
NEW	<input checked="" type="checkbox"/>	pdb 1BZ1 A	Chain A, Hemoglobin (Alpha + Met) Variant >pdb 1BZ...	286	1e-78	
NEW	<input checked="" type="checkbox"/>	pdb 1BAB A	Chain A, Hemoglobin Thionville: An Alpha-Chain Var...	284	7e-78	
NEW	<input checked="" type="checkbox"/>	pdb 1ABY A	Chain A, Cyanomet Rhb1.1 (Recombinant Hemoglobin) ...	284	8e-78	
NEW	<input checked="" type="checkbox"/>	pdb 1C7D A	Chain A, Deoxy Rhb1.2 (Recombinant Hemoglobin)	284	9e-78	
NEW	<input checked="" type="checkbox"/>	pdb 1COH A	Chain A, Structure Of Haemoglobin In The Deoxy Qua...	284	1e-77	
NEW	<input checked="" type="checkbox"/>	pdb 1BZZ A	Chain A, Hemoglobin (Alpha V1m) Mutant >pdb 1BZZ C...	283	2e-77	
NEW	<input checked="" type="checkbox"/>	pdb 1AJ9 A	Chain A, R-State Human Carbonmonoxyhemoglobin Alph...	283	2e-77	
NEW	<input checked="" type="checkbox"/>	pdb 1O1O A	Chain A, Deoxy Hemoglobin (A,C:v1m,V621; B,D:v1m,V...	281	5e-77	
NEW	<input checked="" type="checkbox"/>	pdb 2H35 A	Chain A, Solution Structure Of Human Normal Adult ...	281	5e-77	
NEW	<input checked="" type="checkbox"/>	pdb 1Y0D A	Chain A, T-To-Thigh Quaternary Transitions In Huma...	281	5e-77	
NEW	<input checked="" type="checkbox"/>	pdb 1A3O A	Chain A, Artificial Mutant (Alpha Y42h) Of Deoxy H...	281	5e-77	
NEW	<input checked="" type="checkbox"/>	pdb 1O1N A	Chain A, Deoxy Hemoglobin (A-Glyglygly-C:v1m,L29w;...	281	6e-77	
NEW	<input checked="" type="checkbox"/>	pdb 1YDZ A	Chain A, T-To-T(High) Quaternary Transitions In Hu...	281	6e-77	
NEW	<input checked="" type="checkbox"/>	pdb 1RVW A	Chain A, R State Human Hemoglobin [alpha V96w], Ca...	281	6e-77	
NEW	<input checked="" type="checkbox"/>	pdb 1R1X A	Chain A, Crystal Structure Of Oxy-Human Hemoglobin...	281	6e-77	
NEW	<input checked="" type="checkbox"/>	pdb 1O1M A	Chain A, Deoxy Hemoglobin (A-Glyglygly-C:v1m,L29f,...	281	9e-77	
NEW	<input checked="" type="checkbox"/>	pdb 1XZ5 A	Chain A, T-To-Thigh Quaternary Transitions In Huma...	280	9e-77	
NEW	<input checked="" type="checkbox"/>	pdb 1O1J A	Chain A, Deoxy Hemoglobin (A-Gly-C:v1m,L29f,H58q; ...	280	9e-77	
NEW	<input checked="" type="checkbox"/>	pdb 1XY0 A	Chain A, T-To-Thigh Transitions In Human Hemoglobi...	280	1e-76	
NEW	<input checked="" type="checkbox"/>	pdb 2W72 A	Chain A, Deoxygenated Structure Of A Distal Site H...	280	1e-76	
NEW	<input checked="" type="checkbox"/>	pdb 1XZV A	Chain A, T-To-Thigh Quaternary Transitions In Huma...	280	1e-76	
NEW	<input checked="" type="checkbox"/>	pdb 1XZU A	Chain A, T-To-Thigh Quaternary Transitions In Huma...	280	1e-76	
NEW	<input checked="" type="checkbox"/>	pdb 1Y09 A	Chain A, T-To-T(High) Quaternary Transitions In Hu...	280	1e-76	
NEW	<input checked="" type="checkbox"/>	pdb 1GLI A	Chain A, Deoxyhemoglobin T38w (Alpha Chains), V1g ...	280	2e-76	
NEW	<input checked="" type="checkbox"/>	pdb 1XZ7 A	Chain A, T-To-Thigh Quaternary Transitions In Huma...	280	2e-76	
NEW	<input checked="" type="checkbox"/>	pdb 2DXM A	Chain A, Neutron Structure Analysis Of Deoxy Human...	280	2e-76	
NEW	<input checked="" type="checkbox"/>	pdb 1O1L A	Chain A, Deoxy Hemoglobin (A-Gly-C:v1m,L29w,H58q; ...	280	2e-76	
NEW	<input checked="" type="checkbox"/>	pdb 1O1I A	Chain A, Cyanomet Hemoglobin (A-Gly-C:v1m,L29f,H58...	279	2e-76	





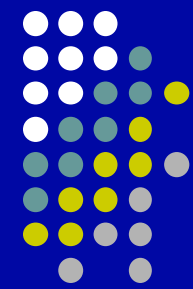
ΕΥΘΥΓΡΑΜΜΙΣΕΙΣ ΠΟΛΛΑΠΛΩΝ ΑΚΟΛΟΥΘΙΩΝ



ΚΥΡΙΑ ΚΡΙΤΗΡΙΑ ΓΙΑ ΤΗ ΔΗΜΙΟΥΡΓΙΑ ΜΙΑΣ ΕΥΘΥΓΡΑΜΜΙΣΗΣ ΠΟΛΛΑΠΛΩΝ ΑΚΟΛΟΥΘΙΩΝ

Κριτήριο	Σημασία
Δομική ομοιότητα	Τα αμινοξέα που παίζουν τον ίδιο ρόλο σε κάθε δομή είναι στην ίδια κολώνα
Εξελικτική ομοιότητα	Τα αμινοξέα ή νουκλεοτίδια που σχετίζονται με το ίδιο αμινοξύ (ή νουκλεοτίδιο) με τον κοινό πρόγονο όλων των ακολουθιών τοποθετούνται στην ίδια κολώνα
Λειτουργική ομοιότητα	Τα αμινοξέα ή νουκλεοτίδια με την ίδια λειτουργία βρίσκονται στην ίδια κολώνα
Ομοιότητα ακολουθίας	Τα αμινοξέα που βρίσκονται στην ίδια κολώνα είναι αυτά που παρέχουν ευθυγράμμιση με τη μεγαλύτερη ομοιότητα.





SIB BLAST Network Service

This NCBI BLAST2 service is maintained by the [Swiss Institute of Bioinformatics](#).

Click on the icons to access the [online BLAST help](#).

◆ Accession number or sequence

Enter a Swiss-Prot/TrEMBL accession number or a **PROTEIN** sequence in **RAW** format.

P20472

Output format: HTML

Run BLAST or Reset Form

Choose the appropriate BLAST program and database:

⊙ blastp - query against the UniProt Knowledgebase (Swiss-Prot+TrEMBL)

◆ Taxonomic groups (not available for PDB and translated EST):

select a database subsection: To restrict the search to a particular taxon, it is much faster to select a database subsection from the drop-down list on the left than to specify your own taxonomic group in the box below. This also gives more accurate statistics.

or specify a taxonomic group: Enter a species name, a TaxID or the latin name of a taxonomic group (elements of the OG, OS and OX lines) to restrict your search to a particular taxon. You may enter a list separated by semicolons (;). Example: Fungi; Homo sapiens.

or select a microbial proteome: Non-redundant Swiss-Prot+TrEMBL complete proteome sets, see the HAMAP pages.

Search only Swiss-Prot (curated sequences) Exclude fragment sequences

⊙ blastp - query against another protein database address if you select the translated ESTs.

⊙ tblastn - query against the six-frame translation of a nucleotide database

Taxonomic groups:

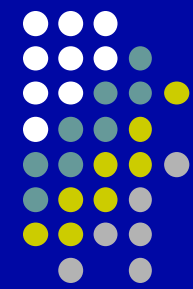
or select a microbial genome: EMBL genome records, including plasmids.

◆ Your email address:

If an e-mail address is provided, results will be automatically mailed back (recommended for tblastn searches)

Run BLAST or Reset Form





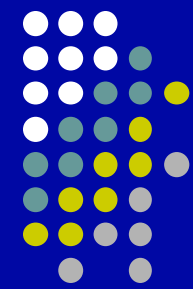
List of potentially matching sequences

Send selected sequences to

Include query sequence

	Db	AC	Description	Score	E-value
<input checked="" type="checkbox"/>	sp	P20472	PRVA_HUMAN Parvalbumin alpha [PVALB] [Homo sapiens (Hu...	187	2e-46
<input type="checkbox"/>	tr	B2R4H7	_HUMAN Parvalbumin, isoform CRA_a (cDNA, FLJ92094, Homo...	187	2e-46
<input type="checkbox"/>	sp	P80050	PRVA_MACFU Parvalbumin alpha (Parvalbumin, muscle) [PV...	181	1e-44
<input type="checkbox"/>	tr	B8ZZ19	_HUMAN Parvalbumin (Fragment) [PVALB] [Homo sapiens (Hu...	176	4e-43
<input type="checkbox"/>	tr	B1AH73	_HUMAN Parvalbumin (Fragment) [PVALB] [Homo sapiens (Hu...	173	4e-42
<input type="checkbox"/>	sp	P02625	PRVA_RAT Parvalbumin alpha [Pvalb] [Rattus norvegicus ...	172	7e-42
<input type="checkbox"/>	sp	Q0VCG3	PRVA_BOVIN Parvalbumin alpha [PVALB] [Bos taurus (Bovi...	171	2e-41
<input type="checkbox"/>	sp	P80080	PRVA_GERSP Parvalbumin alpha [PVALB] [Gerbillus sp. (G...	170	4e-41
<input type="checkbox"/>	sp	P02624	PRVA_RABIT Parvalbumin alpha [PVALB] [Oryctolagus cuni...	169	8e-41
<input type="checkbox"/>	sp	P32848	PRVA_MOUSE Parvalbumin alpha [Pvalb] [Mus musculus (Mo...	165	9e-40
<input type="checkbox"/>	tr	Q545M7	_MOUSE Parvalbumin (Putative uncharacterized protein) (...	165	9e-40
<input checked="" type="checkbox"/>	sp	P80079	PRVA_FELCA Parvalbumin alpha [PVALB] [Felis silvestris...	164	3e-39
<input type="checkbox"/>	tr	Q80WI0	_9MURI Parvalbumin (Fragment) [Pva] [Mus sp]	154	2e-36
<input type="checkbox"/>	tr	B5G1U4	_TAEGU Putative parvalbumin variant 3 [Taeniopygia gutt...	152	8e-36
<input type="checkbox"/>	sp	P80026	PRVM_CHICK Parvalbumin, muscle [Gallus gallus (Chicken)]	151	2e-35
<input type="checkbox"/>	tr	Q66KT9	_XENLA MGC85441 protein [MGC85441] [Xenopus laevis (Afr...	147	2e-34
<input type="checkbox"/>	tr	A1A637	_XENLA LOC100036820 protein [LOC100036820] [Xenopus lae...	145	1e-33
<input type="checkbox"/>	sp	P51434	PRVA_CAVPO Parvalbumin alpha (Fragment) [PVALB] [Cavia...	130	2e-29
<input type="checkbox"/>	tr	A5I875	_9NEOB Parvalbumin [ran m 1.01] [Limnonectes macrodon (...	127	4e-28
<input type="checkbox"/>	tr	Q8JIU0	_9NEOB Parvalbumin alpha [Rana sp. CH-2001]	126	6e-28
<input type="checkbox"/>	tr	B1AH72	_HUMAN Parvalbumin [PVALB] [Homo sapiens (Human)]	125	1e-27
<input type="checkbox"/>	tr	COHAT9	_SALSA Parvalbumin alpha [PRVA] [Salmo salar (Atlantic ...	124	3e-27
<input type="checkbox"/>	sp	P18087	PRVA_RANCA Parvalbumin alpha (PA 4.97) [Rana catesbeia...	122	7e-27
<input type="checkbox"/>	tr	Q8JIU2	_RANES Parvalbumin alpha [Rana esculenta (Edible frog)]	121	1e-26
<input type="checkbox"/>	tr	Q804W1	_DANRE Parvalbumin isoform 4b (Pvalb6 protein) [pvalb6]...	121	2e-26
<input checked="" type="checkbox"/>	sp	P02627	PRVA_RANES Parvalbumin alpha [Rana esculenta (Edible f...	120	4e-26





Taxonomic view NiceBlast view Printable view

List of potentially matching sequences

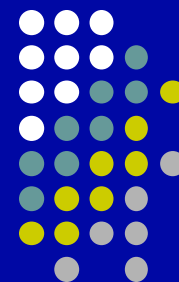
Send selected sequences to Clustal W (multiple alignment) ▾ Submit Query

Include query sequence

Db	AC	Description	Score
<input checked="" type="checkbox"/>	sp P20472	PRVA_HU	(Hu... 187
<input type="checkbox"/>	tr B2R4H7	_HUMAN	Homo... 187
<input type="checkbox"/>	sp P80050	PRVA_MA	[PV... 181

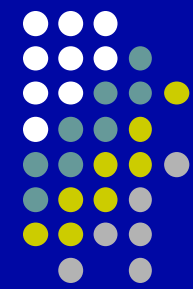
- Clustal W (multiple alignment)
- T-COFFEE (multiple alignment)
- MAFFT (multiple alignment)
- Reduce redundancy
- JACOP (build clusters)
- PRATT (find conserved patterns)
- Retrieve entries (Swiss-Prot format)
- SHOPS (show operon structures, for complete prokaryotic genomes)
- Retrieve sequences (FASTA format)
- Retrieve accession numbers





```
>sp|P20472|PRVA_HUMAN Parvalbumin alpha OS=Homo sapiens GN=PVALB PE=1 SV=2
MSMTDLLNAEDIKKAVGAFSATDSFDHKKFFQMVGLKKKSADDVKKVFHMLDKDKSGFIE
EDELGFILKGFSPDARDLSAKETKMLMAAGDKDGDGKIGVDEFSTLVAES
>sp|P80079|PRVA_FELCA Parvalbumin alpha OS=Felis silvestris catus GN=PVALB PE=1 SV=2
MSMTDLLGAEDIKKAVEAFTAVDSFDYKFFQMVGLKKKSPDDIKKVFHILDKDKSGFIE
EDELGFILKGFYPDARDLSVKETKMLMAAGDKDGDGKIDVDEFFSLVAKS
>sp|P02627|PRVA_RANES Parvalbumin alpha OS=Rana esculenta PE=1 SV=1
PMTDLLAAGDISKAVSAFAAPESFNHKKFFELCGLKSKSKEIMQKVFHVLDQDQSGFIEK
EELCLILKGFTPEGRSLSDKETTALLAAGDKDGDGKIGVDEFVTLVSES
>sp|P02626|PRVA_AMPME Parvalbumin alpha OS=Amphiuma means PE=1 SV=1
SMTDVIPEADINKAIHAFKAGEAFDFKFFVHLLGLNKRSPADVTKAFHILDKDRSGYIEE
EELQLILKGFSKEGRELTDKETKDLLIKGDKDGDGKIGVDEFTSLVAES
>sp|P02619|PRVB_ESOLU Parvalbumin beta OS=Esox lucius PE=1 SV=1
SFAGLKDADVAAALAACSAADSFKHKEFFAKVGLASKSLDDVKKAFYVIDQDKSGFIEED
ELKLFLQNFSPSARALTDAETKAFLADGDKDGDGMIGVDEFAAMIKA
>sp|P43305|PRVU_CHICK Parvalbumin, thymic CPV3 OS=Gallus gallus PE=2 SV=2
MSLTDILSPSDIAAALRDCQAPDSFSPKFFQISGMSKSSSQLKEIFRILDNDQSGFIE
EDELKYFLQRFECGARVLTASETKTFLAAADHDGDGKIGAEEFQEMVQS
>sp|P32930|ONCO_HUMAN Oncomodulin OS=Homo sapiens GN=OCM PE=1 SV=3
MSITDVLSADDIAAALQECQDPDTFEPQFFQTSGLSKMSASQVKDVFRFIDNDQSGYLD
EEELKFFLQKFESGARELTESETKSLMAAADNDGDGKIGAEEFQEMVHS
>sp|Q91482|PRVB1_SALSA Parvalbumin beta 1 OS=Salmo salar PE=1 SV=1
MACAHLCKEADIKTALEACKAADTFSFKTFFHTIGFASKSADDVKKAFKVIDQDASGFIE
VEELKLFLQNFCPKARELTDAETKAFLKAGDADGDGMIGIDEFAVLVKQ
>sp|P02620|PRVB_MERME Parvalbumin beta OS=Merluccius merluccius PE=1 SV=1
AFAGILADADITAALAACKAEGSFKHGEFFTKIGLKGKSAADIKKVFGIIDQDKSDFVEE
DELKLFLQNFSAGARALTDAETATFLKAGDSGDGKIGVEEFAAMVKG
>sp|P02622|PRVB_GADCA Parvalbumin beta OS=Gadus callarias PE=1 SV=1
AFKGILSNADIKAAEAACFKEGSFDEDGFYAKVGLDAFSADELKKLFKIADEDKEGFIEE
DELKLFLIAFAADLRALTDAETKAFLKAGDSGDGKIGVDEFGALVDKWGAKG
```





File name:

Format Swiss-Prot or FASTA

Note: If you do not enter a filename, the sequences will be sent directly to your browser.

- Upload a file from your computer which contains a list of Swiss-Prot/TrEMBL accession numbers (AC) or entry names (ID) or protein_ids, one per line:

or

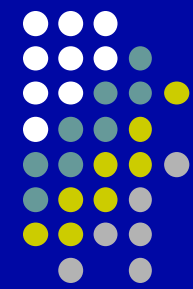
- Or enter a list of Swiss-Prot/TrEMBL accession numbers (AC), entry names (ID) or protein_ids, one per line:

```
P20472
P80079
P02626
P02619
P43305
P32930
Q91482
P02620
P02622
```

or

Last modified 23/Jun/2004



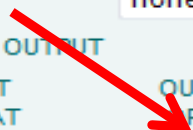


YOUR EMAIL <input type="text"/>	ALIGNMENT TITLE <input type="text" value="Sequence"/>	RESULTS <input type="text" value="interactive"/>	ALIGNMENT <input type="text" value="fast"/>		
KTUP (WORD SIZE) <input type="text" value="def"/>	WINDOW LENGTH <input type="text" value="def"/>	SCORE TYPE <input type="text" value="percent"/>	TOPDIAG <input type="text" value="def"/>	PAIRGAP <input type="text" value="def"/>	
MATRIX <input type="text" value="def"/>	GAP OPEN <input type="text" value="def"/>	NO END GAPS <input type="text" value="yes"/>	GAP EXTENSION <input type="text" value="def"/>	GAP DISTANCES <input type="text" value="def"/>	
	ITERATION <input type="text" value="none"/>		NUMITER <input type="text" value="1"/>		
OUTPUT OUTPUT FORMAT <input type="text" value="aln w/numbers"/>	OUTPUT ORDER <input type="text" value="input"/>	TREE TYPE <input type="text" value="none"/>	CORRECT DIST. <input type="text" value="off"/>	IGNORE GAPS <input type="text" value="off"/>	CLUSTERING <input type="text" value="NJ"/>

Enter or paste a set of sequences in any supported format: [Help](#)

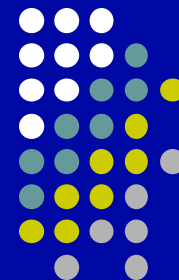
```
>sp|P20472|PRVA_HUMAN Parvalbumin alpha OS=Homo sapiens GN=PVALB  
PE=1 SV=2  
MSMTDLLNAEDIKKAVGAFSATDSFDHKKKFQMVGLKKKSADDVKKVFHMLDKDKSGFIE  
EDELGFILKGFSPDARDLSAKETKMLMAAGDKDGDGKIGVDEFSTLVAES  
>sp|P80079|PRVA_FELCA Parvalbumin alpha OS=Felis silvestris catus  
GN=PVALB PE=1 SV=2  
MSMTDLLGAEDIKKAVEAFTAVIDSFYDKKFQMVGLKKKS PDDIKKVFHILDKDKSGFIE  
EDELGFILKGFYPDARDLSVKETKMLMAAGDKDGDGKIDVDEFPSLVAKS  
>sp|P02626|PRVA_AMPME Parvalbumin alpha OS=Amphiuma means PE=1 SV=1  
SMTDVIPEADINKAIHAFKAGEAFDFKKFVHLLGLNKRSPADVTKAFHILDKDRSGYIEE
```

Upload a file:



- input
- aligned
- input





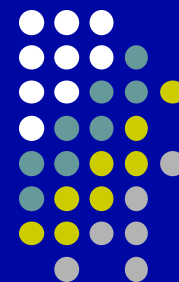
ClustalW2 Results

Results of search	
Number of sequences	9
Alignment score	11717
Sequence format	Pearson
Sequence type	aa
JalView	<input type="button" value="Start Jalview"/>
Output file	clustalw2-20090709-1446361394.output
Alignment file	clustalw2-20090709-1446361394.aln
Guide tree file	clustalw2-20090709-1446361394.dnd
Your input file	clustalw2-20090709-1446361394.input

To save a result file right-click the file link in the above table and choose "Save Target As".

If you cannot see the JalView button, reload the page and check your browser settings to enable Java Applets.





CLUSTAL 2.0.10 multiple sequence alignment

```
sp|P20472|PRVA_HUMAN      MSMTDLLNAEDIKKAVGAFSATDSFDHKKFFQMVGLKKKSADDVKKVFHM 50
sp|P80079|PRVA_FELCA     MSMTDLLGAEDIKKAVEAFTAVDSFDYKKFFQMVGLKKKSPDDIKKVFHI 50
sp|P02626|PRVA_AMPME     -SMTDVIPEADINKAIHAFKAGEAFDFKKFVHLLGLNKRSPADVTKAFHI 49
sp|P02619|PRVB_ESOLU    -SFAGL-KDADVAAALAACSAADSFKHKEFFAKVGLASKSLDDVKKAFYV 48
sp|P43305|PRVU_CHICK     MSLTDILSPSDIAAALRDCQAPDSFSPKKFFQISGMSKKSSSQKKEIFRI 50
sp|P32930|ONCO_HUMAN     MSITDVL SADDIAAALQECQDPDTFEPQKFFQTSGLSKMSASQVKDVFRF 50
sp|Q91482|PRVB1_SALSA    MACAHLCKEADIKTALEACKAADTFSFKTFFHTIGFASKSADDVKKAFKV 50
sp|P02620|PRVB_MERME    -AFAGILADADITAALAACKAEGSFKHGEFFTKIGLKGKSAADIKKVFGE 49
sp|P02622|PRVB_GADCA    -AFKGILSNADIKAAEAACFKEGSFDELDGFIYAKVGLDAFSADELKLFKI 49
      :   :   * : *       :*. *   * : *   :... * .

sp|P20472|PRVA_HUMAN      LDKDKSGFIEEDELGFILKGFSPDARDLSAKETKMLMAAGDKDGDGKIGV 100
sp|P80079|PRVA_FELCA     LDKDKSGFIEEDELGFILKGFYPDARDLSVKETKMLMAAGDKDGDGKIDV 100
sp|P02626|PRVA_AMPME     LDKDRSGYIEEEELQLILKGF SKEGRELTDKETKDLLIKGDKDGDGKIGV 99
sp|P02619|PRVB_ESOLU    IDQDKSGFIEEDELKFLQNFSPSARAL TDAETKAFLADGDKDGDGMIGV 98
sp|P43305|PRVU_CHICK     LDNDQSGFIEEDELKYFLQRFECGARVLTASETKTFLAAADHDGDGKIGA 100
sp|P32930|ONCO_HUMAN     IDNDQSGYLDEEELKFFLQKFESGARELTESETKSLMAAADNDGDGKIGA 100
sp|Q91482|PRVB1_SALSA    IDQDASGFIEVEELKFLQNF SAGARAL TDAETATFLKAGDSGDGKIGV 100
sp|P02620|PRVB_MERME    IDQDKSDFVEEDELKFLQNF SAGARAL TDAETATFLKAGDSGDGKIGV 99
sp|P02622|PRVB_GADCA    ADEDKEGFIEEDELKFLIAFAADLRAL TDAETKAFLKAGDSGDGKIGV 99
      * : * ..::: : **  : * *   * * : **  : : . * **** * .

sp|P20472|PRVA_HUMAN      DEFSTLVAES---- 110
sp|P80079|PRVA_FELCA     DEFFSLVAKS---- 110
sp|P02626|PRVA_AMPME     DEFTSLVAES---- 109
sp|P02619|PRVB_ESOLU    DEFAAMIKA----- 107
sp|P43305|PRVU_CHICK     EEFQEMVQS----- 109
sp|P32930|ONCO_HUMAN     EEFQEMVHS----- 109
sp|Q91482|PRVB1_SALSA    DEFVAVLVKQ----- 109
sp|P02620|PRVB_MERME    EEFAAMV-----KG 108
sp|P02622|PRVB_GADCA    DEFGALVDKKGAKG 113
```





Swiss Institute of
Bioinformatics

TCoffee

[HOME](#) | [references](#) | [help](#) |

TCOFFEE::Regular

Computes a **multiple sequence alignment** and the associated **phylogenetic tree**.
Use T-Coffee to align **Protein**, **RNA** and **DNA** sequences.

Limitations: Maximum number of sequences is **50**
Maximum length of sequences is **2000**

Sequence Input

switch to [TCOFFEE :: Advanced](#)

Paste or upload your set of sequences in FASTA format and Press the SUBMIT button [[Sample File](#)].

[Upload File](#)

Browse...

or paste data

```
>gi|131100|sp|P20472.2|PRVA_HUMAN RecName:
Full=Parvalbumin alpha
MSMTDLLNAEDIKKA VGAFSATDSFDHKKFFQMVGLKKKSADDVKKVFHMLDKDKSGFI
EEDELGFILKGFSPDARDLSAKETKMLMAAGDKDGDGKIGVDEFSTLVAES
>gi|118600941|sp|P80079.2|PRVA_FELCA RecName:
```

Computation Mode

switch to [TCOFFEE :: Advanced](#)

regular: regular T-Coffee.
expresso: structural extension (E-mail recommended).
rcoffee: RNA secondary structure extension (E-mail recommended).

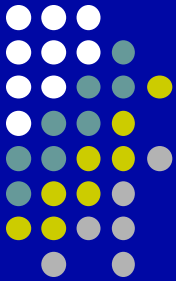
[Computation mode](#)

regular ▼

You may paste your e-mail address:

Submit

Reset





TCOFFEE :: Regular

Your data will remain available on this server over the next 9 days. It will then be deleted.
Do not forget to bookmark this [URL](#) or save it for further reference.

RESULTS

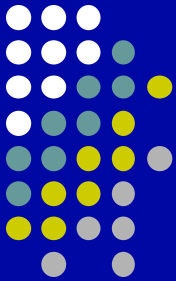
Multiple Alignment	clustalw aln	fasta aln	phylip	score html	score pdf
Tree	dnd file				
System files	LOG	Command line			
Input(s)	tcfTCOR99129_28500.in0				

SEND RESULTS

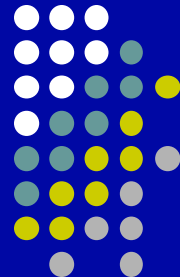
PROTOGENE: turning amino acid alignments into bona fide CDS nucleotide alignments

MyHits: a new interactive resource for protein annotation and domain identification

Home Server: [TCOFFEE :: Regular](#) | [TCOFFEE :: Advanced](#)



clustalw aln



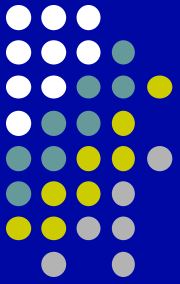
CLUSTAL FORMAT for T-COFFEE Version_7.71 [http://www.tcoffee.org] [MODE: regular], CPU=0.26 sec, SCORE=98, Nseq=9, Len=114

```
gi|131100|sp|P20472.2|PRVA_HUMAN MSMTDLLNAEDIKKAVGAFSATDSFDHKKFFQMVGLKKKSADDVKKVFHMLDKDKSGFIE
gi|118600941|sp|P80079.2|PRVA_FELCA MSMTDLIGAEDIKKAWEAFTAVDSFDYKKFFQMVGLKKKSPDDIKKVFHILDKDKSGFIE
gi|131095|sp|P02626.1|PRVA_AMPME -SMTDVIPEADINKAIHAFKAGEAFDFKKFVHLLGLNKRSPADVTKAFHILDKDRSGYIE
gi|131111|sp|P02619.1|PRVB_ESOLU -SFAG-LKDADVAALAAACSAADSFKHKEFFAKVGLASKSLDDVKKAFYVIDQDKSGFIE
gi|1346814|sp|P43305.2|PRVU_CHICK MSLTDILSPSDIAAALRDCQAPDSFSPKFFQISGMSKSSSQLEIFRILDNDQSGFIE
gi|251757454|sp|P32930.3|ONCO_HUMAN MSITDVLSSADIAAALQECQDPDTFEPQKFFQTSGLSKMSASQVKDVFRFIDNDQSGYLD
gi|2493445|sp|Q91482.1|PRVB1_SALSA MACAHLCKEADIKTALEACKAADTFSFKTFFHTIGFASKSADDVKKAFKVIDQDASGFIE
gi|131116|sp|P02620.1|PRVB_MERME -AFAGILADADITAAALAAACKAEGSFKHGEFFTKIGLKGSAADIKKVFGIIDQDKSDFVE
gi|131112|sp|P02622.1|PRVB_GADCA -AFKGILSNADIKAAEAACFKEGSFDEDGFYAKVGLDAFSADELKKLFKIADEDKEGFIE
: * : * :* * * :.. * . ** :..:
```

```
gi|131100|sp|P20472.2|PRVA_HUMAN EDELGFILKGFSPDARDLSAKETKMLMAAGDKDGDGKIGVDEFSTLV----AES
gi|118600941|sp|P80079.2|PRVA_FELCA EDELGFILKGFYPDARDLSVKETKMLMAAGDKDGDGKIDVDEFFSLV----AKS
gi|131095|sp|P02626.1|PRVA_AMPME EEELQLILKGFSGEGLRELTDKETKDLLIKGDKDGDGKIGVDEFSTLV----AES
gi|131111|sp|P02619.1|PRVB_ESOLU EDELKFLQNFSPSARALTAETKAFLADGDKDGMIGVDEFAMMI-----KA
gi|1346814|sp|P43305.2|PRVU_CHICK EDELKYFLQRFECGARVLTASETKTFLAAADHDGDKIGAEFFQEMV-----QS
gi|251757454|sp|P32930.3|ONCO_HUMAN EEELKFFLQKFESGARELSETKSLMAAADNDGDGKIGAEFFQEMV-----HS
gi|2493445|sp|Q91482.1|PRVB1_SALSA VEELKFLQNFPCKARELTAETKAFLKAGDADGDMIGIDEFAVLV-----KQ
gi|131116|sp|P02620.1|PRVB_MERME EDELKFLQNFSAARALTAETATFLKAGDSGDGKIGVEEFAAMV-----KG
gi|131112|sp|P02622.1|PRVB_GADCA EDELKFLIAFAADLRALTAETKAFLKAGDSGDGKIGVDEFGALVDKWKAGK
:** :* * * : ** :: .* **** * . : ** :: .
```



fasta aln



```
>gi|131100|sp|P20472.2|PRVA_HUMAN RecName: Full=Parvalbumin alpha
MSMTDLLNAEDIKKAVGAFSATDSFDHKKFFQMVGLKKKSADDVKKVHMLDKDKSGFIE
EDELGFILKGFSPDARDLSAKETKMLMAAGDKDGDGKIGVDEFSTLV----AES
>gi|118600941|sp|P80079.2|PRVA_FELCA RecName: Full=Parvalbumin alpha
MSMTDLLGAEDIKKAVEAFTAVDSFDYKFFQMVGLKKKSPDDIKKVFHILDKDKSGFIE
EDELGFILKGFYPDARDLSVKETKMLMAAGDKDGDGKIDVDEFFSLV----AKS
>gi|131095|sp|P02626.1|PRVA_AMPME RecName: Full=Parvalbumin alpha
-SMTDVIPEADINKAIHAFKAGEAFDFKKFVHLLGLNKRSPADVTKAFHILDKDRSGYIE
EEELQLILKGFSGRELTDKETKDLLIKGDKDGDGKIGVDEFTSLV----AES
>gi|131111|sp|P02619.1|PRVB_ESOLU RecName: Full=Parvalbumin beta; AltName: Full=Parvalbumin-2; AltName: Full=Parvalbumin II; AltName: Full=Parvalbumin pI 4.10
-SFAG-LKDADVAAALAACSAAVSFKHKEFFAKVGLASKSLDDVKKAFYVIDQDKSGFIE
EDELKLFQNFSPSARALDAETKAFLADGDKDGDGMIGVDEFAAMI-----KA
>gi|1346814|sp|P43305.2|PRVU_CHICK RecName: Full=Parvalbumin, thymic CPV3; AltName: Full=Parvalbumin-3
MSLTDILSPSDIAAALRDCQAPDSFSPKFFQISGMSKSSQLKEIFRILDNDQSGFIE
EDELKYFLQRFECGARVLIASETKTFLAAADHDGDGKIGAEFFQEMV-----QS
>gi|251757454|sp|P32930.3|ONCO_HUMAN RecName: Full=Oncomodulin; Short=OM; AltName: Full=Parvalbumin beta
MSITDVLASADDIAAALQECQDPDTFEPQKFFQTSGLSKMSASQVKDVFRFIDNDQSGYLD
EEELKFFLQKFESGARELSESETKSLMAAADNDGDGKIGAEFFQEMV-----HS
>gi|2493445|sp|Q91482.1|PRVB1_SALSA RecName: Full=Parvalbumin beta 1; AltName: Full=Major allergen Sal s 1; AltName: Allergen=Sal s 1
MACAHLCKEADIKTALEACKAADTFSEKTFHTIGFASKSADDVKKAFKVIDQDASGFIE
VEELKLFQNFPCKARELDAETKAFLKAGDAGDGMIGIDEFAVLV-----KQ
>gi|131116|sp|P02620.1|PRVB_MERME RecName: Full=Parvalbumin beta
-AFAGILADADITAALAACKAEGSFKHGEFFTKIGLKGKSAADIKKVFGIIDQDKSDFVE
EDELKLFQNFSAARALDAETATFLKAGSDGDGKIGVEEFAAMV-----KG
>gi|131112|sp|P02622.1|PRVB_GADCA RecName: Full=Parvalbumin beta; AltName: Full=Allergen Gad c I; AltName: Full=Allergen M; AltName: Allergen=Gad c 1
-AFKGILSNADIKAAEAACFKEGSFDEDFYAKVGLDFADELKLFKIADEDKEGFIE
EDELKFLIAFAADLRALDAETKAFLKAGSDGDGKIGVDEFGALVDKKGAKG
```

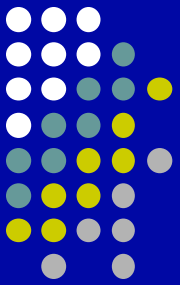


phylip

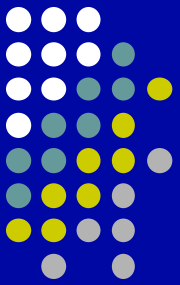
9 114

```
gi|131100| MSMTDLLNAE DIKKAVGAFS ATDSFDHKKF FQMVGLKKKS ADDVKKVFHM LDKDKSGFIE
gi|1186009 MSMTDLLGAE DIKKAVEAFT AVDSFDYKFF FQMVGLKKKS PDDIKKVFHI LDKDKSGFIE
gi|131095| -SMTDVIPEA DINKAIHAFK AGEAFDFKFF VHLLGLNKRS PADVTKAFHI LDKDRSGYIE
gi|131111| -SFAG-LKDA DVAAALAACS AADSFKHKEF FAKVGLASKS LDDVKKAFYV IDQDKSGFIE
gi|1346814 MSLTDILSPS DIAAALRDCQ APDSFSPKFF FQISGMSKKS SSQKKEIFRI LDNDQSGFIE
gi|2517574 MSITDVLSAD DIAAALQECQ DPDTFEPQFF FQTSGLSKMS ASQVKDVFRF IDNDQSGYLD
gi|2493445 MACAHLCKEA DIKTALEACK AADTFSEKTF FHTIGFASKS ADDVKKAFKV IDQDASGFIE
gi|131116| -AFAGILADA DITAALAACK AEGSFKHGEF FTKIGLKGKS AADIKKVFGI IDQDKSDFVE
gi|131112| -AFKGILSNA DIKAAEAACF KEGSFDDEGF YAKVGLDAFS ADELKCLFKI ADEDKEGFIE
```

```
EDELGFILKG FSPDARDLSA KETKMLMAAG DKDGDGKIGV DEFSTLV--- -AES
EDELGFILKG FYPDARDLSV KETKMLMAAG DKDGDGKIDV DEFFSLV--- -AKS
EEELQLILKG FSKEGRELTD KETKDLLIKG DKDGDGKIGV DEFTSLV--- -AES
EDELKFLQON FSPSARALTD AETKAFLADG DKDGDGMIGV DEFAAMI--- --KA
EDELKYFLQR FECGARVLTA SETKTFLAAA DHDGDGKIGA EEFAQEMV--- --QS
EEELKFFLQK FESGARELTE SETKSLMAAA DNDGDGKIGA EEFAQEMV--- --HS
VEELKFLQON FCPKARELTD AETKAFLKAG DADGDGMIGI DEFAVLV--- --KQ
EDELKFLQON FSAGARALTD AETATFLKAG DSDGDGKIGV EEFAAMV--- --KG
EDELKFLIA FAADLRALTD AETKAFLKAG DSDGDGKIGV DEFGALVDKW GAKG
```



score html



★
BAD AVG GOOD
★

```
gi|131100|sp|P2 : 98
gi|118600941|sp : 98
gi|131095|sp|P0 : 98
gi|131111|sp|P0 : 97
gi|1346814|sp|P : 99
gi|251757454|sp : 99
gi|2493445|sp|Q : 98
gi|131116|sp|P0 : 98
gi|131112|sp|P0 : 98
cons : 98
```

```
gi|131100|sp|P2 MSMTDLLNAED I KKAVGAFSATDSFDHKKFFQMVGLKKKSADDVK
gi|118600941|sp MSMTDLLGAED I KKAVEAFTAVDSFDYKKFFQMVGLKKKSPDD I K
gi|131095|sp|P0 -SMTDVIPEAD INKAIHAFKAGEAFDFKKFVHLLGLNKRSPADVT
gi|131111|sp|P0 -SFAG-LKDADVAAALAACSAADSFKHKEFFAKVGLASKSLDDVK
gi|1346814|sp|P MSLTDILSPSDIAAALRDCQAPDSFSPKKFFQISGMSKKSSSQLR
gi|251757454|sp MSITDVLSADDIAAALQECQDPDTFEPQKFFQTSGLSKMSASQVK
gi|2493445|sp|Q MACAHLCKEADIKTALEACKAADTF SFKTF FHTIGFASKSADDVK
gi|131116|sp|P0 -AFAGILADADITAAALAACKAEGSFKHGEFFTKIGLKGKSAADIK
gi|131112|sp|P0 -AFKGI LSNADIKAAEAACFKEGSFDEDGFYAKVGLDAFSADELK
```

```
cons : * : * : * . * * : * : .
```

```
gi|131100|sp|P2 KVFHMLDKDKSGFIEEDELGFILKGFSPDARDLSAKETKMLMAAG
gi|118600941|sp KVFHILDKDKSGFIEEDELGFILKGFYPDARDLSVKETKMLMAAG
gi|131095|sp|P0 KAFHILDKDRSGYIEEEELQLILKGF SKEGRELTDKETKDLLIKG
gi|131111|sp|P0 KAFYVIDQDKSGFIEEDELKFLQNFSP SARAL TDAETKAF LADG
gi|1346814|sp|P EIFRILDNDQSGFIEEDELKYFLQRFECGARVLTASETKTFLAAA
gi|251757454|sp DVFRFIDNDQSGYLDEEELKFFLQKPFESGARELTESETKSIMAAA
gi|2493445|sp|Q KAFKVIDQDASGFIEVEELKFLQNF C PKAREL TDAETKAF LKAG
gi|131116|sp|P0 KVFGIIDQDKSDFVEEDELKFLQNF SAGARAL TDAETATFLKAG
gi|131112|sp|P0 KLFKIADEDKEGFIEEDELKFL I AFAADLRAL TDAETKAF LKAG
```

```
cons . * . * : * . : : : : * * : * * * : * * : * * : : .
```



