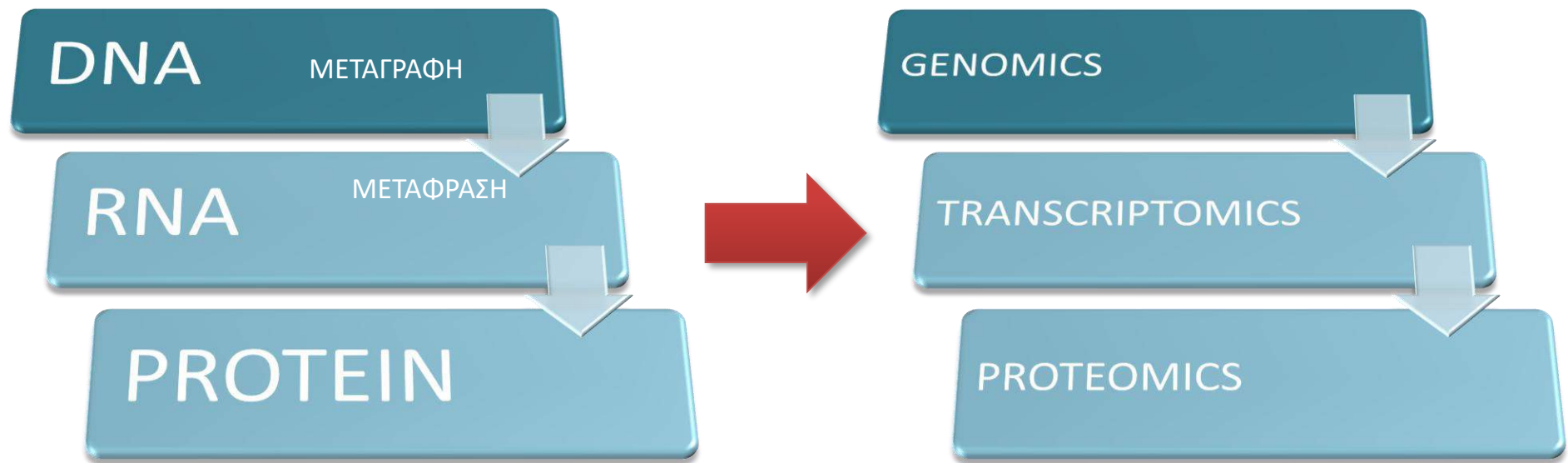


Βιοπληροφορική Φροντιστήριο

Μονάδες ζωής σε ψηφιακή πληροφορία

How life can be translated into data



Αλληλουχία

- Είναι μια σειρά από νουκλεοτίδια ή αμινοξέα όπως αυτά εμφανίζονται σε ένα μόριο DNA ή RNA και μια πρωτεΐνη αντίστοιχα

Πχ TACTAGAGC

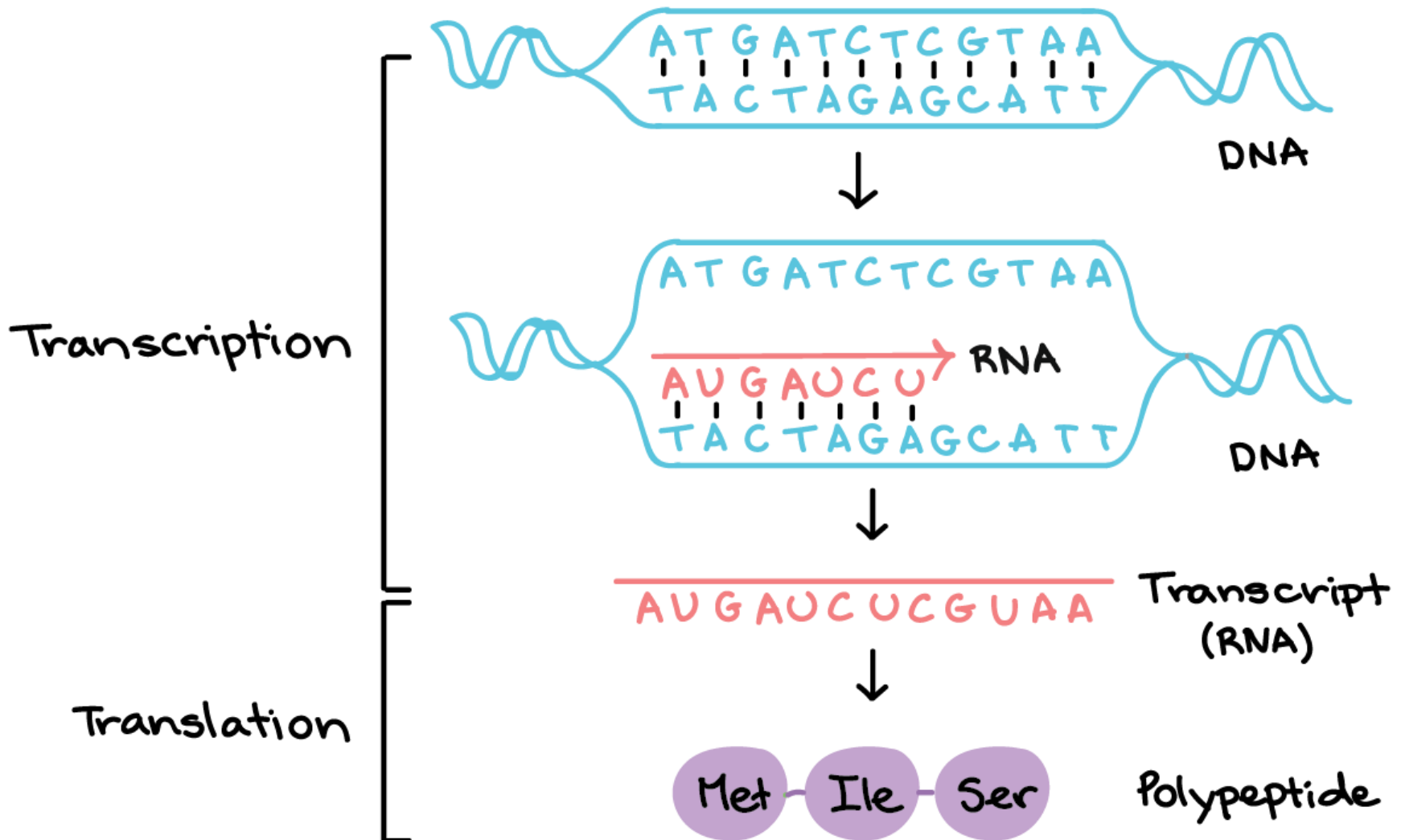
Nucleic acid 1-letter code

A = Αδενίνη | T = Θυμίνη

G = Γουανίνη | C = Κυτοσίνη

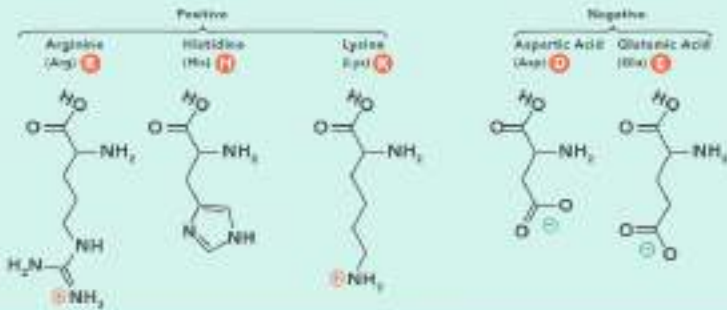
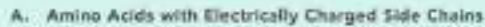
Μεταγραφή

Οι François Jacob και Jacques Monod ανακάλυψαν επίσης ότι η αλληλουχία ενός mRNA μορίου είναι συμπληρωματική με ένα τμήμα μίας από τις δύο αλυσίδες DNA.

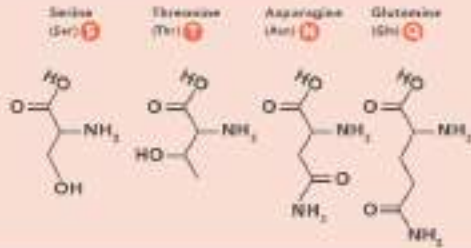


20 Αμινοξέα

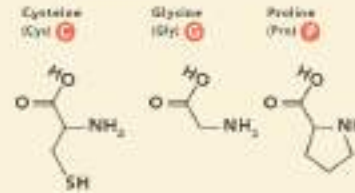
64 Κωδικόνια



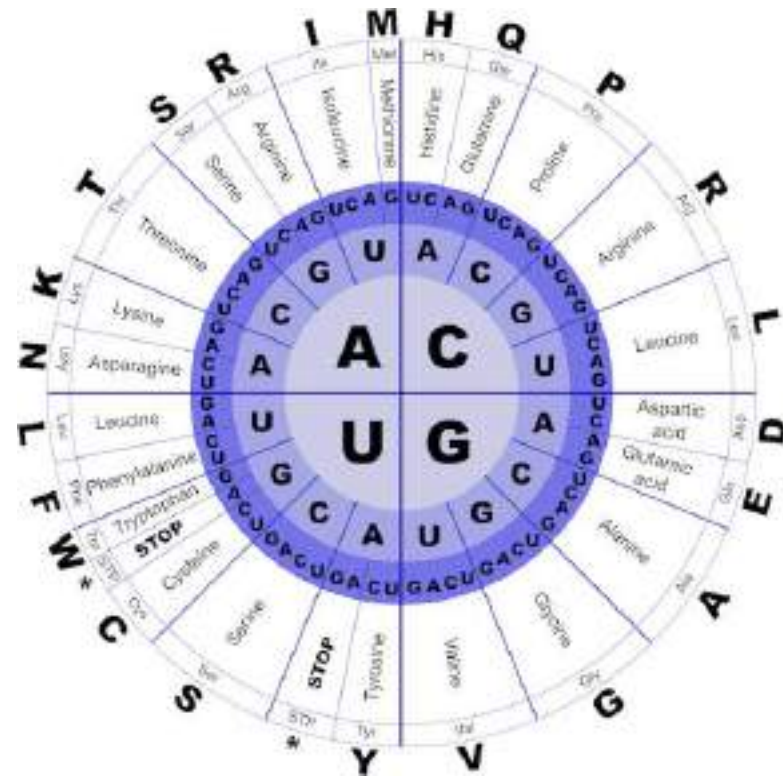
B. Amino Acids with Polar Uncharged Side Chains



C. Special Cases



D. Amino Acids with Hydrophobic Side Chains

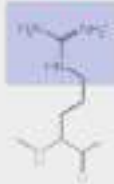


Amino acid abbreviations

Amino acid	3-letter code	1-letter code
alanine	ala	A
arginine	arg	R
asparagine	asn	N
aspartic acid	asp	D
asparagine or aspartic acid	asx	B
cysteine	cys	C
glutamic acid	glu	E
glutamine	gln	Q
glutamine or glutamic acid	glx	Z
glycine	gly	G
histidine	his	H
isoleucine	ile	I
leucine	leu	L
lysine	lys	K
methionine	met	M
phenylalanine	phe	F
proline	pro	P
serine	ser	S
threonine	thr	T
tryptophan	trp	W
tyrosine	tyr	Y
valine	val	V

Αποτέλεσμα πολυμορφισμών

- Synonymous (Silent)
- Non-Synonymous
 - Nonsense
 - A random stop codon is introduced, and the protein is incomplete
 - Missense
 - In a conservative missense mutation, the amino acid replaced is similar in function and shape to the amino acid being replaced.
 - In a non-conservative missense mutation, a completely different kind of amino acid is added to the chain.

		Point mutations			
		No mutation	Silent	Nonsense	Missense
					conservative non-conservative
DNA level		TTC	TTT	ATC	TCC TGC
mRNA level		AAG	AAA	UAG	AGG ACG
protein level		Lys	Lys	STOP	Arg Thr
					 
					basic polar

<https://biologydictionary.net/missense-mutation/>

NCBI Nucleotide

The screenshot shows the NCBI Nucleotide database homepage. At the top, there is a navigation bar with 'NCBI', 'Resources', and 'How To' links. A 'Sign in to NCBI' link is on the right. Below the navigation bar, there is a search section with a 'Nucleotide' dropdown menu, a search input field, and a 'Search' button. A blue arrow points from the text 'Database selection' to the search input field. Below the search section, there is a large banner for 'Nucleotide' with a description: 'The Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript sequence data provide the foundation for biomedical research and discovery.' Below the banner, there are three columns of links: 'Using Nucleotide' (Quick Start Guide, FAQ, Help, GenBank FTP, RefSeq FTP), 'Nucleotide Tools' (Submit to GenBank, LinkOut, E-Utilities, BLAST, Batch Entrez), and 'Other Resources' (GenBank Home, RefSeq Home, Gene Home, SRA Home, INSDC).

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search Help

Advanced

Nucleotide

The Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript sequence data provide the foundation for biomedical research and discovery.

Using Nucleotide

- [Quick Start Guide](#)
- [FAQ](#)
- [Help](#)
- [GenBank FTP](#)
- [RefSeq FTP](#)

Nucleotide Tools

- [Submit to GenBank](#)
- [LinkOut](#)
- [E-Utilities](#)
- [BLAST](#)
- [Batch Entrez](#)

Other Resources

- [GenBank Home](#)
- [RefSeq Home](#)
- [Gene Home](#)
- [SRA Home](#)
- [INSDC](#)

Search BC054498 (=accession number)

Homo sapiens collagen, type I, alpha 2, mRNA (cDNA clone MGC:57503 IMAGE:6474549), complete cds

GenBank: BC054498.1

[FASTA](#) [Graphics](#)

Go to: ☐

LOCUS BC054498 5060 bp mRNA linear PRI 15-JUL-2006
DEFINITION Homo sapiens collagen, type I, alpha 2, mRNA (cDNA clone MGC:57503 IMAGE:6474549), complete cds.
ACCESSION BC054498
VERSION BC054498.1
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5060)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
CONSRM Mammalian Gene Collection Program Team
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED [12477932](#)

FEATURES

source

gene

CDS

CDS
(coding
sequence)

Location/Qualifiers
1..5060
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:57503 IMAGE:6474549"
/tissue_type="Uterus, leiomyosarcoma"
/clone_lib="NIH_MGC_71"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

1..5060
/gene="COL1A2"
/db_xref="GeneID:1278"
/db_xref="HGNC:HGNC:2198"
/db_xref="MIM:120160"
101..4201
/gene="COL1A2"
/codon_start=1
/product="collagen, type I, alpha 2"
/protein_id="AAH54498.1"
/db_xref="GeneID:1278"
/db_xref="HGNC:HGNC:2198"
/db_xref="MIM:120160"

Protein accession number

/translation="MLSFVDTRTLLLLAVTLCLATCQSLQEETVRKGPAGDRGPRGER
GPPGPPGRDGEDGPTGPPGPPGPPGPPGLGNGFAAQYDGKGVGLGPGFMGLMGPRGPP
GAAGAPGPQGPQGPAGEPGEPTGTGPAGARGPAGPPGKAGEDGHPGKPGRPGERGVVG
PQGARGFPPTGLPGFKGIRGHNGLDGLKQPGAPGVKGEPCAPGENGTPGQTGARGL
PGERGRVGPAGPAGARGSDGSVGFVGPAGPIGSAGPPGPPGAPGPKGEIGAVGNAGPA
GPAGPRGEVGLPGLSGPVGPPGNPFGANGLTGAKGAAGLPGVAGAPGLPGPRGIPGPVG
AAGATGARGLVGEPGPAGSKGESGNKGEPSAGPQGPSPGSGEEGKRGPNGEAGSAGP
PGPPGLRGSPSRGLPGADGRAGVMGPPGSRGASGPAGVRGPNGDAGRPPGEPGLMGPR
GLPGSPGNIGPAGKEGFVGLPGIDGRFGPIGPAGARGEFGNIGFPFGPKGPTGDPGKNG
DKGHAGLAGARGAPGPDGNNGAQGPPGPQGVQGGKEQGPAGPPGQGLPGPSGPAGE
VGKPGERGLHGEFGLPGPAGPRGERGPPGESGAAGPTGPIGSRGPSGPPGPDGNKGEF
GVVGAVGTAGPSGPGSLPGERGAAGIPGGKGEKGEPLRGEIGNPGRDGARGAPGAVG
APGPAGATGDRGEAGAAGPAGPAGPRGSPGERGEVGPAGPNGFAGPAGAAGQPGAKGE
RGAKGPKGENGVVGPTGPFVGAAGPAGPNPPGAGSRGDDGGPPGHTGFFGAAGRTGPP
GPSGISGPPGPPGPAGKEGLRGPRGDQGPVGRTEVGAVGPPGFAGEKGPSGEAGTAG
PPGTGPGQGLLGA PGILGLPGSRGERGLPGVAGAVGEPGLGIAGPPGARGPPGAVGS
PGVNGAPGEAGRDGNPCNDGPPGRDQGPCHKGERGYPNIGPVGAAGAPGPHGPFVGP
GKHGNRSETGPSGPVGPAGAVGPRGPSGPQGIKGDKEPGEKGRGLPGLKGNHNLQG
LPGIAGHHGDQAPGSVGPAGPRGPAGPSGPAGKDGRTGHPGTVPAGIRGPQGHQGP
AGPPGPPGPPGPPGVSGGYPDFGYDGFYRADQPRSAFSLRPKDYEV DATLKS LN NQI
ETLLTPEGSRKNPARTCRDLRLSHPEWSSCYWIDPNQGC TMDAIKVYCD FSTGETCI
RAQPENIPAKNWYRSSKDKKHVWLGETINAGSQFEYNVEGVTSKEMATQLAFMRLLAN
YASQNITYHCKNSIAYMDEETGNLKKAVILQGSNDVELVAEGNSRPTYTVLVDGCSKK
TNEWGKTIIEYKTNKPSRLPFLDIAPLDIGGADQEFFVDIGPVCFK"

Protein
sequence

```

1  gtactggcca  egactgeatg  cccgcgccc  ccaggtgata  cctccgccc  tgacccagg
61  gctctgcgac  accaggagtc  tgcattgtta  agtgctagac  atgctcagct  ttgtggatac
121  ggggactttg  ttgctgcttg  cagtaacctt  atgctagaca  acatgccaat  ctttacaaag
181  ggaaaactgt  agaaaaggcc  cagccggaga  tagaggacca  cgtggagaaa  ggggtccacc
241  agggccccc  ggcagagatg  gtgaagatgg  tcccacaggc  cctcctggtc  caectggtec
301  tcttgcccc  cctggctctg  gtgggaactt  tgetgctcag  tatgatggaa  aaggagttgg
361  acttgccct  ggaccaatgg  gcttaatggg  acctagaggc  ccacctggtg  cagctggagc
421  cccaggccct  caaggtttcc  aaggacctgc  tggtagacct  ggtgaacctg  gtcaaatctg
481  tcttgccagt  gctcgtggtc  cagctggccc  tcttgccaa  gctggtagag  atggtccacc
541  tggaaaacc  ggcagacctg  gtgagagagg  agttgttgg  ccacagggtg  ctctgggtt
601  ccttggaact  cctggacttc  ctggcttcaa  aggcattagg  ggacacaatg  gtctggatgg
661  attgaaggga  cagcccggtg  ctctgggtgt  gaagggtgaa  cctgggtccc  ctggtgaaaa
721  tggaaactca  ggtcaaacag  gagcccggtg  gcttctgtgt  gagagaggac  gtgttggtgc
781  ccttgcccca  gctgggtccc  gtggcagtga  tggaaagtgt  ggtcccggtg  gtctgctgtg
841  tcccatggg  tctgctggcc  ctccaggctt  cccagggtcc  cctggcccca  agggtgaaat
901  tggagctgtt  ggtaacgctg  gtctgctgtg  tcccgcgggt  ccccggtgtg  aagtggtctt
961  tccaggccct  tccggccccc  ttggacctcc  tggtaactct  ggagcaaacg  gcttactctg
1021  tggcaagggt  gctgctggcc  tcccgggtgt  tctgggggt  cccggccctc  ctggaccccg
1081  cggtaattct  ggcctgtgtg  gtgctgcggg  tctgactggt  gccagaggac  ttgttggtga
1141  gcttggtcca  gctgggtcca  aaggagagag  cggtaacaag  ggtgagccc  gctctgctgg
1201  gcccgaagg  cctcctggtc  ccagtgtgtg  agaaaggaaa  agaggcccta  atggggagag
1261  tggatctgct  ggcctctcag  gacctctctg  gctgagaggt  agtctgtgt  ctctgtgtct
1321  tcttgagct  gatggcagag  ctggcgtcat  gggccctct  ggtagtctg  gtgcaagtgg
1381  cctgctgga  gtcagaggac  ctaatggaga  tctgctctc  cctggggagc  ctgctctcat

```

...

```

3501  actcagccac  ccagagtgg  gcagtgttta  ctactggatt  gacctaac  aaggatgcac
3661  tatggatgct  atcaaggtat  actgtgattt  ctctactggc  gaaacctgta  tccgggcccc
3721  acctgaaaa  atccagcca  agaactggta  taggagctcc  aaggacaaga  aacacgtctg
3781  gctagggaaa  actatcaatg  ctggcagcca  gtttgaaata  aatgtagaag  gactgacttc
3841  caaggaaatg  gctacccaac  ttgecttcat  gcgctgctg  gccaaactatg  cctctcagaa
3901  catcaacctac  caetgcaaga  acagcattgc  atacatggat  gaggagactg  gcaacctgaa
3961  aaaggctgtc  attctacagg  gctctaatga  tgttgaaact  gttgtgagg  gcaacagcag
4021  gttccattac  actgttcttg  tagatggctg  ctctaaaaag  accaatgaat  ggggaagagc
4081  aatcattgaa  taaaaaacaa  ataagccatc  acgctgccc  ttcttgata  ttgcaccttt
4141  ggcactcgt  ggtgctgacc  aggaattctt  tgtgacatt  ggcacagct  gtttcaasta
4201  aatgaactca  atctaaatta  aaaaagaaa  aaatttgaaa  aaactttctc  ttgacattt
4261  cttctctctt  ttttttaact  gaaagctgaa  tcttccatt  tcttctgac  atctacttgc
4321  ttaatttgt  ggcaaaagag  aaaaagaaag  attgacaga  gcattgtgca  atacagtctc
4381  attactctct  tcccgcgtc  ccccaaaaat  ttgaattttt  ttttaaacac  tcttaccctt
4441  gttatggaaa  atgcaacct  ttgtaagaaa  acaaaaataa  aattgaaa  ataaaaacca
4501  taaacatttg  caacacttgt  ggcttttgaa  tatcttccac  agagggaagt  ttaaaaccca
4561  aacttccaaa  gggttaact  acctcaaaa  acttcccat  gagtgtgac  cacttggtta
4621  ggtgctgacc  tagacagaga  tgaactgagg  tcttgtttt  gttttgttca  taatacaaa
4681  gtgctaatta  atagtatttc  agatacttga  agaattgtga  tgggtctaga  agaatttgag
4741  aagaataact  cctgtattga  gttgtatcgt  gtggtgtatt  ttttaaaaa  ttgatttag
4801  cattcatatt  ttcatctta  ttcccaatta  aaagtatgca  gattatttgc  ccaactctt
4861  ttcagattca  gcatttgttc  ttgcccagtc  tcattttcat  cttcttccat  ggttccacag
4921  aagctttgtt  tcttgggcaa  gcagaaaaat  taattgtac  ctattttgta  tatgtgagat
4981  gtttaataaa  attgtgaaa  aaatgaata  aagcatgtt  ggttttccaa  aaaaaaaa
5041  aaaaaaaa  aaaaaaaa

```

Nucleotide sequence

Expasy translate

Translate is a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence.

DNA or RNA sequence

Please enter a DNA or RNA sequence - numbers and blanks are ignored

Output format

- ☐ Verbose: Met, Stop, spaces between residues
- ☒ Compact: M, -, no spaces
- ☐ Includes nucleotide sequence
- ☐ Includes nucleotide sequence, no spaces

DNA strands

- ☒ forward
- ☒ reverse

Genetic codes - [See NCBI's genetic codes](#)

Standard

reset

TRANSLATE!

Insert nucleotide sequence

AUGAUGGCA

1. Υπάρχουν τρία πιθανά reading frames, ανάλογα με το πού ξεκινά η ανάγνωση:
AUG AUG GCA → Μεθειονίνη (Met) - Μεθειονίνη (Met) - Αλανίνη (Ala)
2. **U GA UGG CA...** → Δεν σχηματίζεται λειτουργική πρωτεΐνη.
3. **G AU GGC A...** → Δεν σχηματίζεται λειτουργική πρωτεΐνη.

5'3' Frame 1

TAETASAPRAQSLAFADPPFVHTRQLTMDDDIAALVWVNGSGMCKAGFAGDDAPRAVFRSIVGRPRHQGVVVGMEQKDSYVVGDEAQSKRGILTLKY
PIEBSGIVTNWDDMEKIWHHTFYNELRVAFEEHPVLLTEAPLNPKANHEKMTQINFETENTFAMYVAIQAVLSLYASGHTTGIVMDSGDGVTHTVPIY
EGYALPHAILPLDLAGHDLT DYLMKILTERTGYSEFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSYELEDDGQVITIGNEPRERCPEALFO
PSFLGMSOGIHETTENSIMKCEVDIRKDLYANTVLSGGTTMYPGIADRMQKEITALAPSTMKINIIAPPERKYSVWIGGSILASLSTFQQMWISNQ
EYDESGPSIVHRRKEF-ADYDLVALHPFLTTPNLRKQDEICMALFVEFVLFWFFFTFWLDSGFKWNWNGEGDSSRLERASPKVHNVAEDFDCTLLFF-
-SFQI-DALLQEVPCHEPKSHPTSL-GEWPSPLPSPHRGGDSIAFV-IM-CKIFLIFALILFYFVLF-MMLRAPPSEFFVPQLEMYEGHWSFWEWVE
AARAYLYTDLRFVE-KCIP-K-

5'3' Frame 2

PPRPRPREHRASPLPIRRPSTFAASSFWMMISPRSSSTTAPACARPASBATWPPGPSSPPSWGARGTRA-WWANVREIFMWATRPRAREASS-
PSSTASSPTGTTWRKSGTTPSTMSCVNLPRSTPCC-PRPP-TPRPTARR-PRSCLRPSTPQPCITLSRLCYPCTFLAVPLAS-WTPVTGSPITLCPST
RGMPFSIMFSCVWTWLAGT-LTTS-RSSPSAATASPPRPSGKSCVTLRSCATSPWTSSKRWPRLLPAPPNRRATSCILTARSSPLAMSGSAALRHSSS
LPSWAMSPVASTKLPSIFS-SVTWTSAKTCTPTQCCLAAPCTTALPTGCRRRSLPWHPAQ-RSRSLLLLSASTPOGSAAPSWFRCPFPSSRCGSASR
SMTSPAPPSSTANASPRIMT-LRYTIS-QNLTCAENKMLALWLYLEFLFCFGEFFFEGLTQDLNLTGTVKVTAVGWSEHPPKFTMWPRTLIAHCCFFN
SHSKYEMRCYRKSLLAILKATPLLSKENGFPVLSQVHTGEVIALLSCKLCHAKFE-SSP-YFFILFYFE--AFVPPLPFLSPNLRCKAKGLPGSGWE
QPSLTCTLT-DQLNKSAHLKN

5'3' Frame 3

KADRVRPASTEPRLCRSAAARPHPPPFAHHG--YRARARRQRLRHVQGRLRGRRCPPGRLEPLHRGAPQAPGRDGGHSGEGFLCGRRGPEQERHPHPEVP
HRARHRHQLGHHGENLAPHLLQ-AACGSRGAPRAADRGPPEFQGGPREDDEPDHV-DLQHPSEHVRCYPGCAIPVRLWPHYNHRDGLR-RGHFPCAHLR
GVCFFPCHPASGPGWPGFD-LPHEDPHARLQLHHHGRAGNRA-H-GEAVLRPPGLRARDGHCFCQLLPGEELRAA-RPGHHHWQ-AVFLP-GTLFA
FLPGHGVLMHFRNYLQLHMEV-RGHFQRFVRQMSAVWRHHNVFNMCRQDAEGDHCFCGTQHNEQDQNCSS-AQVLRVDRLMHPGLAVHLEADVDQAG
V-RVRLPHRPPQMLLGGI-LSCVTPFLDKT-LAQKTR-DWHGFCFCFCFVLVFFFLA-LRI-KLER-R-QQSVGASIPQSSQCGRGL-LHIVVELI
VIPNMHCVVTTGSPFPS-KPPRFSLRMAQSSPKSTQGR--HCFRVNVYMQNFFNLRLNLTFLFCFILNDEPSCFPFFLECPPT-DV-RLLVSLGVGGG
SQGLEVH-LETS-IKVHTLKM

Ανοιχτά
πλαίσια
ανάγνωσης

Το **open reading frame (ORF)** είναι ένα τμήμα του mRNA που περιλαμβάνει:

1.Ένα κωδικόνιο έναρξης (start codon):

Συνήθως το AUG.

2.Μία σειρά κωδικονίων που κωδικοποιούν αμινοξέα.

3.Ένα κωδικόνιο τερματισμού (stop codon):

UAA, UAG, ή UGA.

5'-UTR: Αμετάφραστη. Βοηθά στη ρύθμιση.

ORF: Μεταφράζεται σε πρωτεΐνη:

3'-UTR: Αμετάφραστη. Ρυθμίζει τη σταθερότητα και τη μεταφορά του mRNA

Ψάχνοντας στις βάσεις

- Basic Local Alignment Search Tool (BLAST)

Το βασικό εργαλείο αναζήτησης τοπικής στοίχισης (**BLAST**) είναι ένα πρόγραμμα που αναφέρει περιοχές τοπικής στοίχισης μεταξύ **αλληλουχίας αναζήτησης (query)** και **αλληλουχιών σε μια βάση δεδομένων (subject-target)**. Η ικανότητα ανίχνευσης της **ομολογίας** αλληλουχίας μας επιτρέπει να προσδιορίσουμε εάν ένα γονίδιο ή μια πρωτεΐνη σχετίζεται με άλλα γνωστά γονίδια ή πρωτεΐνες. Η ανίχνευση **ομόλογων** αλληλουχιών διευκολύνει επίσης την αναγνώριση των **διατηρημένων τομέων** που μοιράζονται τα πολλαπλά γονίδια και την αναγνώριση των μελών μιας οικογένειας γονιδίων



blastn **blastp** blastx tblastn tblastx

Enter Query Sequence

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

```
MLSFVDTRTLILLAVTLCATCQSLQEETVVRKGPAGDRQPRGERQPPGPPGRD  
GEDGPTGPPGPPGPPGPPGLGGNFAADYDGKGVGLGPGMGLMGPRGPPG  
AAGAPGPGGFGGPPAGEPGEPPQTGPAGARGPAGPPKAGEDGHHPGKPPRP  
GERGVVGPGGARGFPPTPLPGFGKGRGHNGLDGLKGQGPAGPVKGEPPGAR
```

Query subrange [?](#)From To

Or, upload file

Choose file

No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)☐ Align two or more sequences [?](#)

Insert protein sequence

Choose Search Set

Databases

☒ Standard databases (nr etc.) [New](#) ☐ Experimental databases

Compare

☐ Select to compare standard and experimental database [?](#)

Standard

Database

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

Organism

Optional

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

Exclude

Optional

☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Restrict search to specific taxon

Program Selection

Algorithm

- ☐ Quick BLASTP (Accelerated protein-protein BLAST)
- ☒ **blastp** (protein-protein BLAST)
- ☐ PSI-BLAST (Position-Specific Iterated BLAST)
- ☐ PHI-BLAST (Pattern Hit Initiated BLAST)
- ☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)**BLAST**

Search database nr using Blastp (protein-protein BLAST)

☐ Show results in a new window

BLASTn: Nucleotide query vs Nucleotide db

BLASTp: Protein query Protein db

BLASTx: Nucleotide query (auto-translated) vs Protein db

BLAST

- Εύρεση της μέγιστης δυνατής ομοιότητας μεταξύ αλληλουχιών query - subject

Descriptions									
Sequences producing significant alignments									
Download Select columns Show 100									
select all 100 sequences selected									
GenPept Graphics Distance tree of results Multiple alignment MSA Viewer									
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☒	Collagen type I alpha 2 [Homo sapiens]	Homo sapiens	2469	2469	100%	0.0	100.00%	1366	AAH42586.1
☒	collagen type I alpha 2 chain [Homo sapiens]	Homo sapiens	2468	2468	100%	0.0	99.93%	1366	KAJ2546688.1
☒	collagen alpha-2(I) chain precursor [Homo sapiens]	Homo sapiens	2466	2466	100%	0.0	99.93%	1366	NP_000080.2
☒	collagen alpha-2(I) chain [Pan troglodytes]	Pan troglodytes	2465	2465	100%	0.0	99.85%	1366	XP_001168894.1
☒	collagen alpha-2(I) chain [Pan paniscus]	Pan paniscus	2464	2464	100%	0.0	99.93%	1366	XP_003609763.2
☒	pro-alpha2(I) collagen [Homo sapiens]	Homo sapiens	2464	2464	100%	0.0	99.85%	1366	CAA98969.1
☒	collagen alpha-2(I) chain [Gorilla gorilla gorilla]	Gorilla gorilla gorilla	2464	2464	100%	0.0	99.78%	1366	XP_018886394.1
☒	collagen alpha-2(I) chain [Hylobates moloch]	Hylobates moloch	2461	2461	100%	0.0	99.63%	1366	XP_032618930.1
☒	collagen alpha-2(I) chain [Nomascus leucogenys]	Nomascus leucogenys	2461	2461	100%	0.0	99.71%	1366	XP_003252555.1
☒	collagen alpha-2(I) chain [Pongo abelii]	Pongo abelii	2458	2458	100%	0.0	99.63%	1366	XP_024105396.1
☒	pro-alpha 2(I) collagen [Homo sapiens]	Homo sapiens	2451	2451	100%	0.0	99.41%	1366	AAB93981.1
☒	collagen alpha-2(I) chain [Ptilocobus tephrosceles]	Ptilocobus tephrosceles	2449	2449	100%	0.0	99.27%	1366	XP_023082504.2

Multiple Sequence alignment (MSA)

- Όταν θέλουμε να ελέγξουμε πολλαπλές αλληλουχίες ίδιου μήκους για ομολογία μεταξύ τους
- Συνήθως κάνει ανά-δύο συγκρίσεις (pairwise).
- Συχνότεροι Αλγόριθμοι : Clustal (W or OMEGA), MUSCLE, MAFFT

```
A G C T T A C T A A T C C G G G C C G A A T T A G G T C
A G T T T A T T A A T T C G A G C T G A A C T A G G T C
A G T C T A T T A A T T C G A G C A G A A C T T G G T C
A G T T T A T T A A T T C G A G C T G A A C T T G G C C
A G T C T A C T A A T T C G A G C T G A A T T A G G T C
A G A T T A T T A A T T C G A G C T G A A C T T G G T C
A G A T T G C T A A T T C G A G C C G A A T T A G G T C
A G A T T A T T A A T C C G G G C T G A A T T A G G T C
A G T C T A T T A A T T C G A G C T G A A T T A G G A C
A G C T T A T T A A T T C G T G C T G A A C T C G G A C
A G C T T A T T A A T T C G A G C T G A A C T C G G A C
A G C T T A T T A A T T C G A G C C G A A C T C G G G C
A G T C T T T T A A T T C G A G C T G A A T T A G G A C
```

Clustal Omega

[Input form](#)[Web services](#)[Help & Documentation](#)[Feedback](#)[Share](#)

Tools > Multiple Sequence Alignment > Clustal Omega

Service Retirement

[Wise2DBA](#) and [Promoterwise](#) are scheduled for retirement on 15th April 2018. Alternatives can be found at [Exonerate](#), [BWA](#) or [BLAT](#). If you have any concerns, please contact us via [support](#).

Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

Important note: This tool can align up to 4000 sequences or a maximum file size of 4 MB.

STEP 1 - Enter your input sequences

[Enter or paste](#) a set of

PROTEIN

[sequences](#) in any supported format:

Στοίχιση και πολλαπλή στοίχιση

- Πρωτεΐνη

Histone H1 (residues 120-180)

HUMAN	KKASKPKKAASKAPT	KKPKATPVKKAKKKL	AATPKKAKKPKTV	KA	KPVKASKPKKAKPVK
MOUSE	KKAAKPKKAASKAP	SKKPKATPVKKAKKKP	AATPKKAKKPKV	VKV	KPVKASKPKKAKTVK
RAT	KKAAKPKKAASKAP	SKKPKATPVKKAKKKP	AATPKKAKKPKI	VKV	KPVKASKPKKAKPVK
COW	KKAAKPKKAASKAP	SKKPKATPVKKAKKKP	AATPKKTKKPKTV	KA	KPVKASKPKKTKPVK
CHIMP	KKASKPKKAASKAPT	KKPKATPVKKAKKKL	AATPKKAKKPKTV	KA	KPVKASKPKKAKPVK
	:	**:	*****:	*****	**

NON-CONSERVED AMINO ACIDS

Conservative

Conservative

Non-conservative

Conservative

Non-conservative

Semi-conservative

Conservative

Non-conservative