

BIOL312

Practical 3

BLASTP

- What is Modification?
 - Change in DNA and protein sequence and corresponding changes in protein function.
- Mutations accumulate in sequences from an ancestral sequence
- Derived sequence diverge one from another
- Some sequences remain similar that allow identification of common ancestry.



Evolution


Sequence Identity

Same residues being present at corresponding position between 2 sequences.

For proteins; same amino acids

 Query 4 KLVVRPW#
Sbjct 2 KLVVRPW#
KLVVRPW#

For nucleic acids; same bases

 Query 46 CAGACC
Sbjct 47 |||||
CAGAC#

Sequence Similarity

Similar residues being present at corresponding position between 2 sequences.

- ❖ For nucleic acids, sequence similarity and identity are the same.
- ❖ For proteins, sequence similarity involves a.a with similar physicochemical and functional properties.

Similar substitution: conservative substitution

Not expect to disrupt the structure/function of the protein

Some examples of similar substitution:

-Substitution of lysine - arginine



Positively charged hydrophilic a.a.

-Substitution of isoleucine, leucine and valine



Aliphatic hydrophobic side chains

Sequence Identity & Sequence Similarity

- % Identity (PID) = (# of identical a.a. / # of total a.a) * 100
- % Similarity = [(# of identical a.a. + # of similar substitutions)/ # of total a.a.] *100

E-Value

Goal is to understand the likelihood that a BLAST search of an entire database produces a result by chance.

- E-value: Expectation value that indicates the number of alignments with a score that are expected to occur by chance in a database.
- This provides an estimate of the number of false positive results from a BLAST search.
- E-value depends on the **database size & query length**.
- The closer the E-value to 0, the better the alignment is.
- **As E-value approaches zero, the probability that the alignment occurred by chance approaches zero.**

E.g.: $E=1e-2$ ($= 1 \times 10^{-2} = 0.01$)

Using BLASTP to Identify an Unknown DNA Sequence

```
GGCATGAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGACACAACACTGTGTTCACTAGCAAC
CTCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGT
GAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTACAAGACAGGTTTAAGGAG
ACCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCTCTGC
CTATTGGTCTATTTTCCCACCCTTAGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCT
TTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCT
CGGTGCCTTTAGTGATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCCACTGAGTGAGCTG
CACTGTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGGTGAGTCTATGGGACCCTTGATGTTTTCTTT
CCCCTTCTTTTCTATGGTTAAGTTCATGTCATAGGAAGGGGAGAAGTAACAGGGTACAGTTTAGAATGGG
AACAGACGAATGATT
```

II. Identifying the Unknown DNA Sequence by Investigating its Protein Product

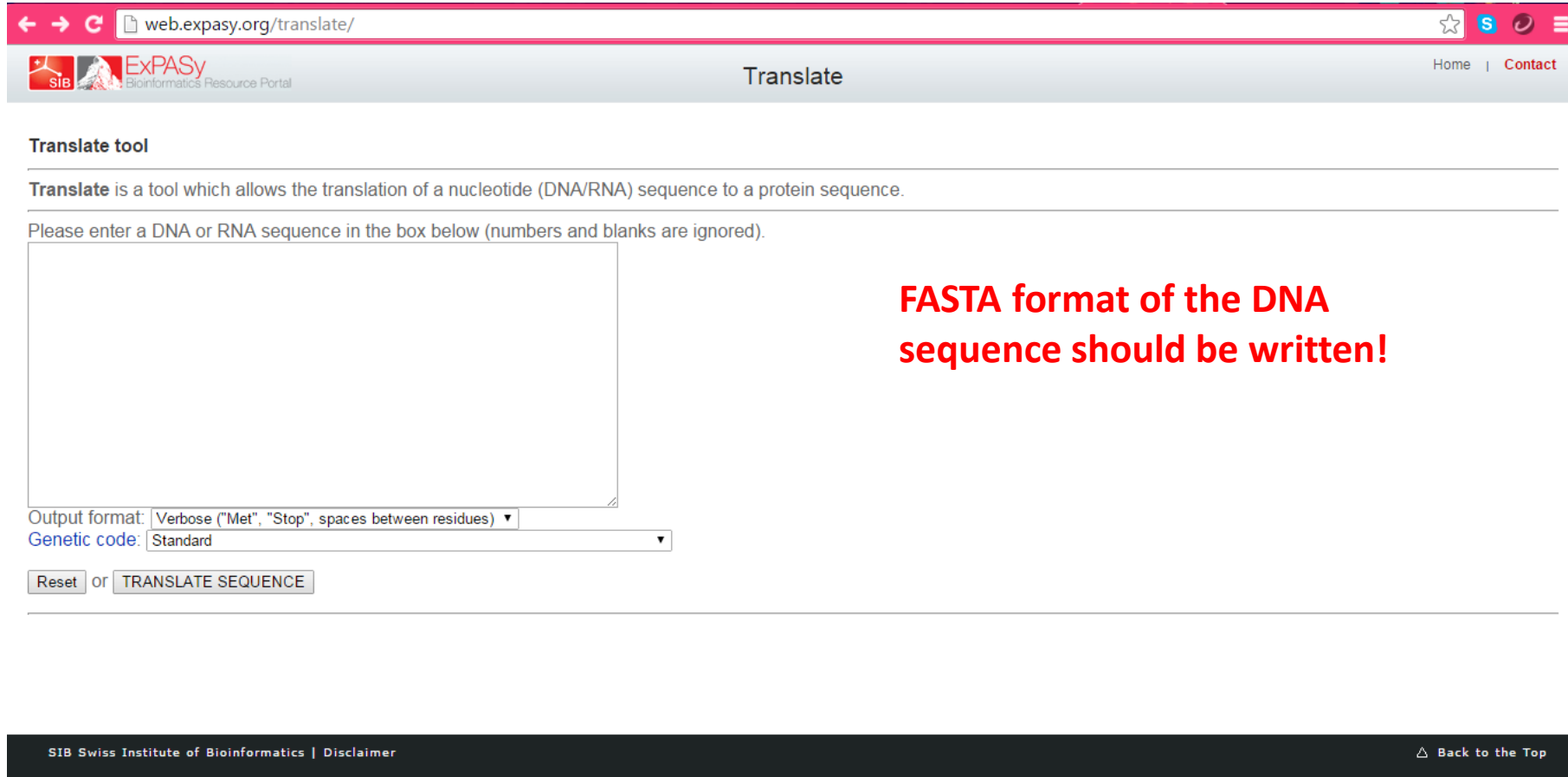
Before translating a nucleotide sequence to a.a. sequence;

- Which strand is the coding strand
- The correct frame should be determined.

- EXPASY is a bioinformatics resource portal that is used for translating a DNA sequence in all six reading frames. This solve the problem;
 - Having start translation at wrong nucleotide
 - Inhibit frameshift mutation

EXPASY

(<http://web.expasy.org/translate/>)



The screenshot shows the Expasy Translate tool interface. At the top, there is a browser address bar with the URL web.expasy.org/translate/. Below the address bar is the Expasy logo and the text "SIB Swiss Institute of Bioinformatics Resource Portal". The main heading is "Translate". In the top right corner, there are links for "Home" and "Contact".

The main content area is titled "Translate tool". It contains the following text: "Translate is a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence." Below this, it says: "Please enter a DNA or RNA sequence in the box below (numbers and blanks are ignored)." There is a large empty text input box for the sequence.

To the right of the input box, there is a red text annotation: **FASTA format of the DNA sequence should be written!**

Below the input box, there are two dropdown menus: "Output format: Verbose ('Met', 'Stop', spaces between residues)" and "Genetic code: Standard". At the bottom of the form, there are two buttons: "Reset" and "TRANSLATE SEQUENCE".

At the bottom of the page, there is a footer with the text "SIB Swiss Institute of Bioinformatics | Disclaimer" and a link "Back to the Top".

In order to translate nucleotide sequence use ExPasy Translate Tool

web.expasy.org/translate/

SIB ExPASy Bioinformatics Resource Portal

Translate

Translate tool

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

Please enter a DNA or RNA sequence in the box below (numbers and blanks are ignored).

```
GGCATGAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGACACAACCTGTGTCACTAG
CAAC
CTCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCA
AGGT
GAACGTGGATGAAGTTGGTGGTGGAGCCCTGGGCAGGTTGGTATCAAGGTTACAAGACAGGTTTAA
GGAG
ACCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCT
CTGC
CTATTGGTCTATTTCCACCCCTTAGGCTGCTGGTGGTCTACCCCTGGACCCAGAGGTTCTTTGAG
TCCT
TTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAG
TGCT
CGGTGCCTTTAGTGATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCCACTGAGTGA
GCTG
CACTGTGACAAGCTGCACGTGGATCCTGAGAACTCAGGGTGAGTCTATGGGACCCCTTGATGTTTT
```

Output format: Verbose ("Met", "Stop", spaces between residues) ▼

Genetic code: Standard ▼

Reset or TRANSLATE SEQUENCE

Insert the longest a.a. sequence to BLASTP

Translate Tool - Results of translation

Open reading frames are highlighted in red. Please select one of the following frames - in the next page, you will be able to select your initiator and retrieve your amino acid sequence:

5'3' Frame 1
XXXXXXXXXXXXXXXXXXXXXXXXXG **Met** KVRAEPSIAYICF **Stop** HNCVH **Stop** QPQTD **Met** VHLTPEEKSAVTALWGKVNVDVGG
EALGRLVSRQLQDRFKETNRNWACGDREDSWVSDRH **Stop** LSLPIGLFSHP **Stop** AAGGLPLDPEVL **Stop** VLWGSVHS **Stop** CCYG
QP **Stop** GEGSWQESARCL **Stop** **Stop** WPGSPGQPQGHLCHE **Stop** AAL **Stop** QAARGS **Stop** ELQGES **Met** GPL **Met** FSFPFFS **Met** VKF
Met S **Stop** EGEK **Stop** QGTV **Stop** NGKQTN

5'3' Frame 2
XXXXXXXXXXXXXXXXXXXXXXXXXA **Stop** KSGQSHLLTFASDITVFTSNLKQTPWCT **Stop** LLRRSLPLLPCGAR **Stop** TW **Met** KLVV
RPWAGWYQGYKTGLRRPIETGHVETKTLGFLIGTDSLCLLVYFPTLRLLVVYPWTQRFFESFGDLSTPDAV **Met** GNPVKVAH
GKKVLGAFSDGLAHLNKGTFATLSELHCDKLHVDPENFRVSLWDP **Stop** CFLSPSFLWLSSCHRKGRSNRVQFR **Met** GNRR
Met

5'3' Frame 3
XXXXXXXXXXXXXXXXXXXXXXXXXHESQGRAIYCLHLLTQLCSLATSNRHHGAPDS **Stop** GEVCRYCPVGQGERG **Stop** SWW **Stop**
GPGQVGIVTRQV **Stop** GDQ **Stop** KLG **Met** WRQRLLGF **Stop** **Stop** ALTLSAYWSIFPPLGCWWSTLGRGSLSPGLICPLL **Met** LLW
ATLR **Stop** RL **Met** ARKCSVPLV **Met** AWLTWTTSRAPLPH **Stop** VSCVTSTCTWILRTSG **Stop** VYGTLDVFFPLLYG **Stop** VHVIGRGE
VTGYSLEWETDE **Stop**

3'5' Frame 1
NHSSVSHSKLYPVTSP **Met** T **Stop** T **Stop** P **Stop** KRRGKKT SRVP **Stop** THPEVLRIHVQLVTVQLTQCGKGALEVVQVSQAITKGT
EHFLA **Met** SLHLRVAHNSIRSGQIPKGLKEPLGPRVDHQQPKGGKIDQ **Stop** AERVSAYQKPKSLLCLH **Met** PSFYWSP **Stop** TCLV
TLIPTCPGPHQLHPRSPCPTGQ **Stop** RQTSPQESGAPWCLFEVASEHSCVRSKCKQ **Stop** **Met** ALP **Stop** LSCXAXXXXXXXXXXXXX
XXXXXXXXXX

3'5' Frame 2
IIRLFPILNCTLLLLPFL **Stop** HELNHRKEGERKHQGSRLTLKFSGSTCSLSQCSSLVAKVPLRLSR **Stop** ARPSLKAPSTFLP
Stop AFTLGLPITASGVDRSPKDSKNLWVQG **Stop** TTSSLRVGK **Stop** TNRQRESVPIRNPRVFSVSTCPVSIKLLKPV **Stop** P **Stop** Y
QPAQLTTNFIVHLAPQGSNGRLLRSQVHHGVCLRLLVNTVVSEANVSNRWLCPDFHAXLXXXXXXXXXXXXXXXXXXXXX

3'5' Frame 3
SFVCFPF **Stop** TVPCYFSPSYD **Met** NLTIEKKGKENIKGPIDSP **Stop** SSQDPRAACHSAHHSVWQRCP **Stop** GCPGEPGHH **Stop** RH
RALSCHEPSP **Stop** GCP **Stop** QHQEWTDPTQRTQRTSGSKGRPPAA **Stop** GWENRPIGRESQCLSETQESSLSPHAQFLLVSLNLS
CNLDTNLPRASPPTSSTFTLPHRAVTADFSSGVRCT **Met** VSV **Stop** GC **Stop** **Stop** TQLCQKQ **Met** **Stop** AIDGSALTF **Met** XX **Stop** XXXX
XXXXXXXXXXXXXXXXXX

Translate Tool - Results of translation

Open reading frames are highlighted in red. Please select one of the following frames - in the next page, you will be able to select your initiator and retrieve your amino acid sequence:

5'3' Frame 1
XXXXXXXXXXXXXXXXXXXXXXXXXXXXRATAVHRLRRPLALGPRCWFSSLLAFSSSSAPDPSSAAWSLLLPCLPAPRRPARARRPGSPRRSAAAAASAPHTV Stop APAAAAAAGVPS Stop PRRPPPPR
PDRPPRRPPESPPRR Stop RHNHRARPPDSVQY Stop SGEPERALRGAA Met RPSGTAGAALLALLAALCPASRALEEKVKCGGTSNKLTQLGTFEDHFLSLQR Met FNNCEVVLGNLEITYVQRNYDLSFL
KTIQEVAGYVLIANTVERIPLNLIIRGN Met YYENSYALAVLSNYDANKTGLKELP Met RNLQEIILHGAVRFSNNPALCNVESIQWRDIVSSDFLSN Met S Met DFNHLGSCQKCDPSCPNGSCWAGAEEN
CQKLTIIICAQQCSGRGCRGKSPSDCCHNQAAGCTGPRESDCLVCRKFRDEATCKDTCPLP Met LYNPTTYQ Met DYNPEGKYSFGATCVKCKPRNYVVDHGSCVRACGADSYE Met EEDGVRKCKKCE
GPKRKYCNGIGIGEFKDTLSINATNIKHFKNCTISGGDLHILPVAFRGDSFHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAFENLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVIISG
KNKNLCYANTINWKKLFTSGGQTKIISNRGENSCKATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCNILEGEPREFVENSECICQCHPECLPQA Met NITCTGRGPDNCIQCAHYIDGPHCVK
TCPAGV Met GENNTLVWKYADAGHVCHLCHPNCTYGPNESLRA Met LFCFLKSSCNQSDGVSVHSGSPAAQESCLGWIPSLPSEFQLGWGGCSHLHAWPSASVIITASSCH Stop ASCLHASVPPAF
LSATPARGPPGSQEYPTHFLPSTPFASASHPN Stop Stop LTIPTRTDLFLTAR Stop LFKAQNLSSETCRDS Met LPAFVSIS Met APVHASSLSGFCPSV Stop NSRASCATIICVKSQ Met Stop KHLNNKLAALFNA
RTTPVTE Stop KLKNIC Stop VNEQ Met NKIINN Stop PLIQHPDTL Stop F Stop LFKSSLAYIAL Stop FN Stop KSP

5'3' Frame 2
XXXXXXXXXXXXXXXXXXXXXXXXXXXXGPPPLSTASGGRWPVPAAGSPSSSHSPPLLLPIPPPPPGPSSSRPASPRPGPRELDDRAAPGAARPQQPPPPTRCERPPRRRPESRASPGGRRRDPRT
TGHLVGVRPSPRLAANATTTAHGLTPSSIDRESSELFGEQRCDPPGRPGQORSWRCLRSARRVGLWRKRKFAKARVTSRRAWLLKIFASRGCISATVRWSLGIWKL Met CRGI Met FPS Stop RPSR
RWLV Met SSLPSTQWSGFLWKTCSRSEICTTKIPTI Stop QSYLT Met Met QIKPD Stop RSCP Stop EIYRKS Met APCGSATTLPCATWKASSGGT Stop SAVTFSATCQWTSRTTWAAKSVIQAVP Met GAAG
VQERRARN Stop PKSSVPSSAPGAAVASPPVTAATTSVPQAARAPGRATAWCAANSETKPRARTAPHSKSTTPPRTRW Met Stop TPRANTALVPPA Stop RSVPI Met W Stop QITARASEVGPPTA Met RWR
KTASASVRSAGLAACKVTE Stop VLVNLTHTSP Stop Met LRLINTSKTAPPSVAISTSCRWHLGVTPSHILRLWIHNRWF Stop KP Stop RKSQGF Stop FRLLGLKTGRS Met LLRT Stop KSYAAGPSN Met VSF
LRLSA Stop T Stop HPWDYAPSR Stop V Met E Met Stop Stop FLETKICAMet QIQ Stop T GKNC LGPPVRRKKL Stop ATEVTKAARPAARSA Met PCAPRAAAGARSPTASLAG Met SAEAGNATSATFWRASQS
LWRTLSAYSATQSACLR Stop TSPAQDGDQTTVSSVPTLTAPTASRPARQESWEKTPWSGSTQ Met PA Met CATCAIQTAPTQGE Met RVSEPCYSAFLNYHPVIKV Met Met AACPTRAGAQLLRSHA Stop
DGSLLFCRQSFVWVDAATS Met PGLLHL Stop SSRPPPATEPHAFTRLFPLSLPPLHVGRRQVPSILPISFLPLPLPVPLTPTSS Stop PSPGLTSSSSLPDDCSKHRICQKPAAGTCCQPSP Stop LAWP
QS Met LLALVPSAPLFEILEPAVRQLSVSKARCEINISITNWLCS Met LEQRLSQSRNSKIFAE Stop Met NKI Stop INA Stop IINH Stop SNIQTHCDFNYLRVV Stop HILLYDLIKNL

5'3' Frame 3
XXXXXXXXXXXXXXXXXXXXXXXXXXXXGHRCPPPAAAGLGSPLLVLPPRILLLLCSSRSLRRLVPPPPALPPRAPARAS Stop TTGPPAQRGRSSLRPPHGVARRGRGGRSPELAPAAAAAQTGR
QATSSASARVPASPLTPQPPRTAP Stop LRPVLIGRAGASSSGSDATLDRGRSAPGAAGCALPGESESGGKESLPRHE Stop QAHAVGHF Stop RSFSQPPEDVQ Stop L Stop GGPWEFGNYLCAEEL Stop SF
LLKDHPPGGWLCPCPCQHSADSF GKPADHQRKYVLRKFLRLSSLI Stop L Stop CK Stop NRTEGAHEKFTGNPAWRRRAVQQQPCPVQRGKHPVAGHSQQ Stop LSQQHVNLPEPPGQLPKV Stop SKLSQ
WELLGCRRGELPETDQNHLCPAVLRALPWVQVPQ Stop LLPQVCRRLHGPPGERLPGVQIPRRSHVQGHLPPTHALQPHHPVDGCEPRGQIQWLCHLREEVSP Stop LCGDRSRLVRPSLWGRQL Stop DG
GRRRPQV Stop EVRRALPQSV Stop RNRVW Stop RHTLHKCYEY Stop TLQKLHLHQRWSPHAGGI Stop G Stop LLHTYSASGSTGTGYSENKGNHRVADSGLA Stop KQDGPCCF Stop EPRNHTRQDQ
ATWSVFCGCGPEHNILGITLPGQDK Stop WRCDNFRKQKQFVLCYKNKLEKTVWDLRSENENYKQR Stop KQLGHRPGLPCLVLRGLLGPGAQGLRLLPECCPRQG Met RGQVQHS GGRAKGVCGEL
Stop VHTVPRVPASGHEHHLRTGTRQLYVPCPLH Stop RPLRQDLGRSHGRKQHPGLEVRRCRPCVPPVPSKHLRARK Stop ESQSHVILPF Stop TIIL Stop SK Stop WQVPPEREPSCSGV Met L
Met DPFSSAVRVSAGLGMet QPPCLAFCDHGGLLPLSL Met PSRVCSRFPPFCHPCTWAARFPRVSYPPFSFSLCQCLSPQLVANHHQQD Stop PLPHCQ Met IVQSTEFVRNLQGLHAASLLRN Stop
HGSPCF Stop PWFLPLCLKF Stop SQLCDNYLCQKPDVTSQ Stop QGCFVQ Stop NINACHRYETKYLLE Stop TNE Stop Met HNK Stop LTTNPTSRHIVILII Stop E Stop FSIYCF Met I Stop LKIS

3'5' Frame 1
WRFLIKS Stop SNIC Stop TTLK Stop LKSCVW Met LD Stop WLIYYAFIHLFIHSANIFEFLCDDRRCSSEIQSSQFVIE Met FSHLAFDITDNCRTAGSRISNRGAEGTKARS Met DWGHANYGEGWQHGVPAFG
Stop QILCFEQSSGSEEEVSPGGDG Stop LLVGVGRTGKGSGRKE Met GRILLGTWRPTCRGGRKESGGNRRVKA Stop GSVAGGRRDDHRCRRPG Met EVAASTPTQLKL Stop RQKRDDPS Stop A Stop LLSS
WAPALVGHAAIITLITG Stop Stop FKKA Stop HGSETLISWPVAVW Met AQVAH Met AGICVLPDQGVVFSHDS CRAGLDAV GAVNVVGTLDTVVWSPSCAGDVHGLRQALWVALYALRVLHKLPLWALQNV
LVHAFPASADIPARDAVPLRAPAALGGAQG Met ADLACGLAAVFTSVAYNFRFLTGGPKQFPVYICIAQIFVS Stop NYHISITYLLEGA Stop SQGCYVQADNRKRKL Met LLGPAAYDF Stop VLKS Met EV
RPVFRPSLNQKQPCDFLYGFQNIQFLWIQRSS Met CEGVTPKCHRQDVEIATDGGAVFEVFNIRSIYGEVFKFTNTYSVTHFAARFALLTLADAVFLHLIAGVPTGSDARAVICHIIITGTLHAGGTKAV
FALGVHILVRRGGVVEHWGAGVLRGPFVSEFAAHQAVALPGARAACGLVVAAVTGGLATAAPGALLGTDGQFLAVLLSCTPAPIGTAWITLLAAAQVLLVH Stop HVAEKVTADYVPPPLDAFHVAQ
GRVVAEPHGA Met QDFL Stop ISHGQLLSGFIICIVR Stop DC Stop GVGIFVVIHSSDDLQVQFRRNPLHCVTEGNETISHLLDGL Stop EGKIIIPHLIGNFQIPKDLTVIEHPLEAEK Met IFKSAQLRELVTRALA
NFLFLQSPTRRAERSQQRERCPGRPGGSHRCSPKSSLRSLRSILDGVRGPCAVVVLAARRGLGRTPRVPVVRSGRRRPPGLARDSGRLGRGGRSHRVGGGGCCGRAAPGAARSSSSRGPGRGEA
GREEEGPGGGGIGRSRGGGECIEEGGEPAAQTGGRRPEAVDSGGPXXXXXXXXXXXXXXXXXXXXXXXXXXXX

3'5' Frame 2
GDF Stop LNHKAIYAKLLLN Stop NHNVSGCWISG Stop LF Met HLFICFSFTQIFLSFYSTGVVLLNKAASLLRCHFHWLLTQIIVAQLALEFQTEGQKEPRLEAWTGA Met LITEKAGS Met ESLQVSDKFC
LNNHLAVRRKRSVLGV Met VSY Stop LG Stop EALAKGVRKWWGYSWEPGPRGVAERAGGTDAS Stop RHEAQWQEEAV Met IDAEGQAWRWLHPPQS Stop NSDGRREGIHPKHS Stop AAGLPLW
DTLPSLL Stop LQDDSLKRQNN Met ALRSLFPGP Stop VQFGWRHWHWPASAYFQTRVLFSP Met TPAGQVLTQWGSPS Met Stop WAHWIQLSGPRPVQV Met F Met A Stop GRHSGWHC Met HSEFSTNSLGSRS
Met HSTHSDIETEDTOSLSCQOOSGELKAWTWDVALQLESPLLLEVE Stop REVRNSEEQIIEVA Stop HKELEREITSRSLISLBERNRKQV Met ERTTAREN Stop RCLVLR Met LSRBSKAW

Standard Protein BLAST

blastp blastx tblastn tblastx

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

```
Met KLVV R P W A G W Y Q G Y K T G L R R P I E T G H V E T E K T L G F L  
I G T D S L C L L V Y F P T L R L L V V Y P W T Q R F F E S F G D L S T P D  
A V Met G N P K V K A H G K K V L G A F S D G L A H L D N L K G T F A T L S  
E L H C D K L H V D P E N F R V S L W D P
```

Clear Query subrange

From

To

Or, upload file Dosya seçilmedi

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database

Organism Exclude

Exclude Models (XM/XP) Uncultured/environmental sample sequences

Entrez Query [YouTube](#) [Create custom database](#)

Program Selection

Algorithm

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

Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)

Show results in a new window

[+ Algorithm parameters](#)

New Analyze your query with [SmartBLAST](#)

Graphic Summary

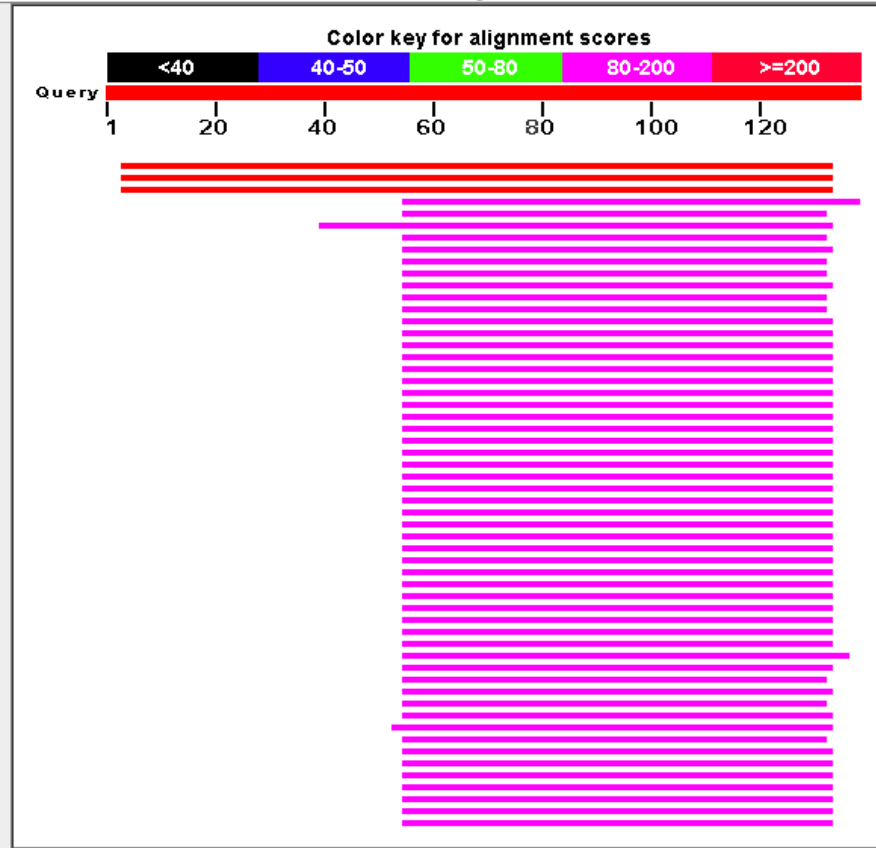
Show Conserved Domains

Putative conserved domains have been detected, click on the image below for detailed results.



Distribution of 100 Blast Hits on the Query Sequence

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



Perfect matches: red
Next best: purple
Mediocre: green
Poor matches: blue
Very poor or no match: black

Descriptions

Descriptions

Sequences producing significant alignments:

Select: All None Selected: 0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> PREDICTED: hemoqlobin subunit beta isoform X1 [Pongo abelii]	247	247	94%	9e-81	94%	XP_009244997.1
<input type="checkbox"/> PREDICTED: hemoqlobin subunit beta isoform X1 [Nomascus leucogenys]	244	244	94%	9e-80	94%	XP_012359140.1
<input type="checkbox"/> PREDICTED: hemoqlobin subunit beta isoform X1 [Mandrillus leucophaeus]	237	237	94%	8e-77	89%	XP_011830555.1
<input type="checkbox"/> mutant beta qlobin [Homo sapiens]	164	164	60%	5e-49	96%	AAK20080.1
<input type="checkbox"/> hemoqlobin beta qlobin chain [Homo sapiens]	154	154	56%	7e-45	97%	ADV59922.1
<input type="checkbox"/> beta-qlobin [Pan troglodytes]	154	154	67%	8e-45	85%	CAA26204.1
<input type="checkbox"/> hemoqlobin beta [Homo sapiens]	153	153	56%	1e-44	97%	AAT36651.1
<input type="checkbox"/> mutant beta-qlobin [Homo sapiens]	154	154	56%	1e-44	96%	AAG46184.1
<input type="checkbox"/> hemoqlobin [Homo sapiens]	153	153	56%	1e-44	97%	ABG47031.1
<input type="checkbox"/> beta qlobin [Homo sapiens]	153	153	56%	1e-44	97%	AAY15222.1
<input type="checkbox"/> hemoqlobin beta subunit variant [Homo sapiens]	154	154	56%	1e-44	96%	AAF00488.1
<input type="checkbox"/> mutant hemoqlobin beta chain [Homo sapiens]	153	153	56%	1e-44	97%	AAQ63175.1
<input type="checkbox"/> beta qlobin [Homo sapiens]	153	153	56%	1e-44	97%	ACF16762.1
<input type="checkbox"/> mutant beta-qlobin [Homo sapiens]	154	154	56%	1e-44	96%	AAL68978.1
<input type="checkbox"/> Chain B. Oxygen Affinity Modulation By The N-Termini Of The Beta Chains In Human And Bovine Hemoqlobin	154	154	56%	1e-44	96%	2HHE_B
<input type="checkbox"/> Chain B. T-To-T(High) Quaternary Transitions In Human Hemoqlobin: Betan108a Deoxy Low-Salt (1 Test Set)	154	154	56%	1e-44	96%	1Y7Z_B
<input type="checkbox"/> hemoqlobin beta chain variant Hb S-Wake [Homo sapiens]	154	154	56%	1e-44	96%	AAN11320.1
<input type="checkbox"/> Chain B. T-To-T(High) Quaternary Transitions In Human Hemoqlobin: Deshis146beta Deoxy Low-Salt	154	154	56%	1e-44	96%	1Y85_B
<input type="checkbox"/> Chain B. Structure Of The T. Brucei Haptoglobin-haemoqlobin Receptor Bound To Human Haptoglobin-haemoqlobin	154	154	56%	1e-44	96%	5HU6_B
<input type="checkbox"/> unnamed protein product [Homo sapiens]	154	154	56%	1e-44	96%	CAA23759.1
<input type="checkbox"/> Chain B. (Alpha-Oxy, Beta-(C112q)deoxy) T-State Human Hemoqlobin	154	154	56%	1e-44	96%	1GBV_B
<input type="checkbox"/> Chain B. Refined Crystal Structure Of Deoxyhemoqlobin S. I. Restrained Least-Squares Refinement At 3.0-Angstroms Resolution	154	154	56%	1e-44	96%	1HBS_B
<input type="checkbox"/> hemoqlobin beta chain variant [Homo sapiens]	154	154	56%	1e-44	96%	ACD39349.1
<input type="checkbox"/> Chain B. Cyanomet Rhb 1.1 (Recombinant Hemoqlobin)	154	154	56%	1e-44	96%	1ABY_B
<input type="checkbox"/> beta qlobin chain variant [Homo sapiens]	154	154	56%	1e-44	96%	AAK29639.1
<input type="checkbox"/> hemoqlobin beta [synthetic construct]	154	154	56%	1e-44	96%	AAX29557.1
<input type="checkbox"/> Chain B. High-resolution X-ray Study Of Deoxy Recombinant Human Hemoqlobins Synthesized From Beta-globins Having Mutated Amino Termini	154	154	56%	2e-44	96%	1DXV_B

Alignments

Download GenPept Graphics Next Previous Descriptions

PREDICTED: hemoglobin subunit beta isoform X1 [Pongo abelii]
 Sequence ID: [ref|XP_009244997.1](#) Length: 169 Number of Matches: 1

Related Information

[Gene](#) - associated gene details
[Map Viewer](#) - aligned genomic context

Range 1: 2 to 128 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
247 bits(631)	9e-81	Compositional matrix adjust.	121/129(94%)	122/129(94%)	2/129(1%)
Query 4	KLWVRPWAGWYQGYKTGLRRPIETGHVETKTLGFLIGTDSLCLLVYFPTLRLLWVYPWT				63
Sbjct 2	KLWVRPWAGWYQGYKTGLRR IE GHVETKTLGFLIGTDSLCLLVYFPTLRLLWVYPWT				61
Query 64	QRFFESFGDLSTPDAMVETGNPKVKAHGKKVLFVAFSDGLAHLNLTGTFATLSELHCDKLL				123
Sbjct 62	QRFFESFGDLSTPDAMV--GNPKVKAHGKKVLFVAFSDGLAHLNLTGTFATLSELHCDKLL				119
Query 124	HVDPENFRV 132				
Sbjct 120	HVDPENFR+ 128				

Download GenPept Graphics Next Previous Descriptions

PREDICTED: hemoglobin subunit beta isoform X1 [Nomascus leucogenys]
 Sequence ID: [ref|XP_012359140.1](#) Length: 161 Number of Matches: 1

Related Information

[Gene](#) - associated gene details

Range 1: 2 to 128 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
244 bits(624)	9e-80	Compositional matrix adjust.	121/129(94%)	122/129(94%)	2/129(1%)
Query 4	KLWVRPWAGWYQGYKTGLRRPIETGHVETKTLGFLIGTDSLCLLVYFPTLRLLWVYPWT				63
Sbjct 2	KLWVRPWAGWYQGYKTGLRR IET HVETEK LGFLIGTDSLCLLVYFPTLRLLWVYPWT				61
Query 64	QRFFESFGDLSTPDAMVETGNPKVKAHGKKVLFVAFSDGLAHLNLTGTFATLSELHCDKLL				123
Sbjct 62	QRFFESFGDLSTPDAMV--GNPKVKAHGKKVLFVAFSDGLAHLNLTGTFATLSELHCDKLL				119
Query 124	HVDPENFRV 132				
Sbjct 120	HVDPENFR+ 128				

Download GenPept Graphics Next Previous Descriptions

PREDICTED: hemoglobin subunit beta isoform X1 [Mandrillus leucophaeus]
 Sequence ID: [ref|XP_011830555.1](#) Length: 169 Number of Matches: 1

Related Information

[Gene](#) - associated gene details

Range 1: 2 to 128 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
237 bits(605)	8e-77	Compositional matrix adjust.	115/129(89%)	119/129(92%)	2/129(1%)
Query 4	KLWVRPWAGWYQGYKTGLRRPIETGHVETKTLGFLIGTDSLCLLVYFPTLRLLWVYPWT				63
Sbjct 2	KLWVRPWAGWYQGYKT LRR IE GHVE EKT LG FLIGTDSLCLLVYFPTLRLLWVYPWT				61
Query 64	QRFFESFGDLSTPDAMVETGNPKVKAHGKKVLFVAFSDGLAHLNLTGTFATLSELHCDKLL				123
Sbjct 62	QRFFESFGDLSSPDAMV--GNPKVKAHGKKVLFVAFSDGLNHLNLTGTFATLSELHCDKLL				119
Query 124	HVDPENFRV 132				
	HVDPENF++				

Download GenPept Graphics

Next Previous Descriptions

liver transporter list-1 [Mus musculus]

Sequence ID: [dbj|BAB03272.1](#)

[See 4 more title\(s\)](#)

Range 1: 1 to 689 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
1166 bits(3016)	0.0	Compositional matrix adjust.	560/691(81%)	613/691(88%)	6/691(0%)

Query	1	MDHTQQSRKAAEAQPSR\$KQTRFCDFGKFLAALSFSYICKALGGVVMKSSITQIERRFD	60
Sbjct	1	MDQTQHP\$KAA--QPLR\$EKTRHCDGFRIFLAALSFSYICKALGGVIMKSSITQIERRFD	58
Query	61	IPSSISGLIDGGFEIGNLLVIVFVSFYFGSKLHRPKLIGIGCFIMGIGSILTALPHFFMGY	120
Sbjct	59	IPSSISGLIDGGFEIGNLLVIVFVSFYFGSKLHRPKLIGTGCFFIMGIGSILTALPHFFMGY	118
Query	121	YKYAKENDIGSLGNSTLTFCINQMTSPTGPSPEIVEKGCEKGLKSHMWIYVLMGNMLRGI	180
Sbjct	119	YRYATENDISSLHNSTLTCLVNQTTSLTGT\$PEIMEKGCEKGSNSYTWIYVLMGNMLRGI	178
Query	181	GETPIVPLGISYLD\$FAKEGHTSMHLGTLHTIAMIGPILGFIMSSVFAKIYVDVGVVDLN	240
Sbjct	179	GETPIVPLGVS\$YID\$FAKEGNSSMYLGTHTIAMIGPILGFIMSSVFAKIYVDVGVVDLR	238
Query	241	SVRITPNDARWVGAWLSFIVNGLLCITSSIPFFFLPKIPKRSQERKNSVSLHAPK\$TDE	300
Sbjct	239	SVRITPQDARWVGAWLGFIVNGLLCIIC\$IPFFFLPKIPKRSQERKNSASLHV\$K\$TDE	298
Query	301	EKKHMTNLT\$KQEEQDPSNMTGFLRSLRSILTNEIYVIFLILTL\$QVSGFIGSFTYLFKFI	360
Sbjct	299	DKNPVTNPTTQEQ\$QAPANLTGLW\$SLRSILTNEQYVIFLILTL\$QISSFIGSFTYLFKFI	358
Query	361	EQQFGRTASQANFLLGIITIP\$MATAMFLGGYIVKFKLTSVGI\$AKFVFFITSSVAYAFQF	420
Sbjct	359	EQQFGRTASQANFLLG\$VITIP\$MASGMFLGGYLIKRLKLTLLGITK\$FVFFITTTMAYVYFL	418
Query	421	LYFPLLCENKPFAGLTLTYDGMNPVDSHIDVPLSYCNSDC\$CDKNQWEPICGENGVTYIS	480
Sbjct	419	SYFLLICENKAFAGLTLTYDGMNPVDSHIDVPLSYCNSDC\$CDKNQWEPVCGENGVTYIS	478
Query	481	PCLAGCK\$FRGD\$KPNNTEFYDC\$CISNS---GNNSAHLGECPRYKCKTNYFYIILQV	536
Sbjct	479	PCLAGCK\$FRGD\$KLNIEFYDC\$CVS\$G\$G\$FQKGNH\$SARLGECP\$RDKCKTKYFYITFQV	538
Query	537	TVSFF\$TAMG\$PSLILILM\$K\$VQPEL\$KSLAMGFHSLIIRALGGILAPIY\$YGAFIDRTCIKW	596
Sbjct	539	IISFF\$TALG\$TSLMLILIR\$VQPEL\$KSLMGMGFHSLVVRTLGGILAPV\$Y\$GALIDRTCMKW	598
Query	597	SVTSCGKRGACRLYNSRLFGFSYLG\$NLAL\$KTPPLF\$LYV\$LIYFTKRYK\$RNDN\$KLENG	656
Sbjct	599	SVTSCGARGACRLYNSRLFGMIYVGL\$IAL\$KTPILL\$LYVALIYV\$MKR\$K\$RNDN\$KILENG	658
Query	657	RQFTDEGNPDSV\$NKN\$GYCVPYDEQ\$NETPL 687	
Sbjct	659	RKFTDEGNPEP\$VNN\$NGY\$CVP\$DE\$E\$T\$PL 689	

Related Information

- [Gene](#) - associated gene details
- [Map Viewer](#) - aligned genomic context
- [Identical Proteins](#) - Identical proteins to NP_065241.1

% 81 Identical

% 88 Similar

+ similar substitutions