

BIOINFORMATICĂ APLICATĂ ÎN BIOLOGIA STRUCTURALĂ

Seminar VII

Operații simple cu secvențe. BLAST

SMS2

Se dă secvența (fișierul fasta NM_000558.fasta)

```
>NM_000558  
ATGGTGCTGTCTCCTGCCGACAAGACCAACGTCAAGGCCGCCTGGGGTAAGGTCGGCGCGCACGCTGGCG  
AGTATGGTGCGGAGGCCCTGGAGAGGATGTTCCCTGTCCTTCCCCACCACCAAGACCTACTTCCCGCACTT  
CGACCTGAGCCACGGCTCTGCCCAGGTTAAGGGCCACGGCAAGAAGGTGGCCGACGCGCTGACCAACGCC  
GTGGCGCACGTGGACGACATGCCCAACGCGCTGTCCGCCCTGAGCGACCTGCACGCGCACAAAGCTTCGGG  
TGGACCCGGTCAACTTCAAGCTCCTAAGCCACTGCCTGCTGGTGACCCTGGCCGCCACCTCCCCGCCGA  
GTTACCCCTGCGGTGCACGCCTCCCTGGACAAGTTCCTGGCTTCTGTGAGCACCGTGCTGACCTCCAAA  
TACCGTTAA
```

Ce codifică această secvență?

SMS2

1. Ce reprezintă secvența dată – ADN, ARN sau o secvență de aminoacizi?

>NM_000558.4

```
ATGGTGCTGTCTCCTGCCGACAAGACCAACGTCAAGGCCGCCTGGGGTAAGGTCGGCGCGCACGCTGGCG  
AGTATGGTGCGGAGGCCCTGGAGAGGATGTTCCCTGTCCTTCCCCACCACCAAGACCTACTTCCC  
CGACCTGAGCCACGGCTCTGCCCAGGTTAAGGGCCACGGCAAGAAGGTGGCCGACGCGCTGACCAACGCC  
GTGGCGCACGTGGACGACATGCCCAACGCGCTGTCCGCCCTGAGCGACCTGCACGCGCACAAGCTTCGGG  
TGGACCCGGTCAACTTCAAGCTCCTAAGCCACTGCCTGCTGGTGACCCTGGCCGCCACCTCCCCGCCGA  
GTTACCCCTGCGGTGCACGCCTCCCTGGACAAGTTCCTGGCTTCTGTGAGCACCGTGCTGACCTCCAAA  
TACCGTTAA
```

SMS2



Format Conversion

- Combine FASTA
- EMBL to FASTA
- EMBL Feature Extractor
- EMBL Trans Extractor
- Filter DNA
- Filter Protein
- GenBank to FASTA
- GenBank Feature Extractor
- GenBank Trans Extractor
- One to Three
- Range Extractor DNA
- Range Extractor Protein
- Reverse Complement
- Split Codons
- Split FASTA
- Three to One
- Window Extractor DNA
- Window Extractor Protein

Sequence Analysis

- Codon Plot
- Codon Usage
- CpG Islands
- DNA Molecular Weight
- DNA Pattern Find
- DNA Stats
- Fuzzy Search DNA
- Fuzzy Search Protein
- Ident and Sim
- Multi Rev Trans
- Mutate for Digest
- ORF Finder
- Pairwise Align Codons
- Pairwise Align DNA
- Pairwise Align Protein
- PCR Primer Stats
- PCR Products
- Protein GRAVY
- Protein Isoelectric Point
- Protein Molecular Weight
- Protein Pattern Find
- Protein Stats
- Restriction Digest
- Restriction Summary
- Reverse Translate
- Translate

Sequence Figures

- Color Align Conservation
- Color Align Properties
- Group DNA
- Group Protein
- Primer Map
- Restriction Map
- Translation Map

Random Sequences

- Mutate DNA
- Mutate Protein
- Random Coding DNA
- Random DNA Sequence
- Random DNA Regions
- Random Protein Sequence

2. Deschideți documentul Microsoft Word numit NM_000558 și copiați secvența fasta pe care o conține. **Folosiți SMS2 pentru a transforma secvența dată într-o secvență de aminoacizi**

https://mail.uaic.ro/~marius.mihasan/research/mirrored_sites_tools/sms2/index.html

sau

<http://www.bioinformatics.org/sms2/>

sau

Google.ro căutați după cuvântul cheie SMS2

Care dintre funcțiile listate de SMS2 le veți alege?

SMS2

3. Ștergeți textul predefinit din căsuța text (butonul Clear) și copiați secvența FASTA din documentul Word. Apăsați Submit.

Sequence Manipulation Suite:

Translate

Translate accepts a DNA sequence and converts it into a protein in the reading frame you specify. Translate supports the entire IUPAC alphabet and several genetic codes.

Paste a raw sequence or one or more FASTA sequences into the text area below. Input limit is 200,000,000 characters.

```
>sample sequence
gckugcgaygartty

>sample sequence 2
ggwgggggaggtggcgaggaagatgacgtggtagttgtcgcggcagctgcc
agcaagaaaaataacatgataattatcacgacaactacctggtgatgttc
```

- Translate in on the strand.
- Use the genetic code.

*This page requires JavaScript. See [browser compatibility](#).

*You can [mirror this page](#) or [use it off-line](#).

Sequence Manipulation Suite - Google Chrome

Translate results

```
>rf 1 Untitled
MLLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG
KKVADALTNVAHVDDMPNALSALSDLHAHKLRVDPVNFLLSHCLLVTLAAHLPAEFTP
AVHASLDKFLASVSTVLTSKYR*
```

Secvența obținută este secvența de aminoacizi codificată de secvența ADN inițială. Copiați-o în fișierul word sub forma unei secvențe în format FASTA cu titlul

NM_000558_aac.

SMS2

Realizați procesul de traducere in-silico utilizând diverse valori ale listei Translate in:

Sequence Manipulation Suite:

Translate

Translate accepts a DNA sequence and converts it into a protein in the reading frame you specify. Translate supports the entire IUPAC alphabet and several genetic codes.

Paste a raw sequence or one or more FASTA sequences into the text area below. Input limit is 200,000,000 characters.

```
>sample sequence
gckugcgaygartty

>sample sequence 2
ggwgggggaggtggcgcgaggaagatgacgtggtagttgtcgcggcagctgccaggagaagt
agcaagaaaaataacatgataattatcacgacaactacctggtgatgttgctagtaatat
```

Submit Clear Reset

- Translate in on the strand.
- Use the genetic code.

*This page requires JavaScript. See [browser compatibility](#).

*You can [mirror this page](#) or [use it off-line](#).

Ce obțineți?

De ce?

Care este secvența corectă?

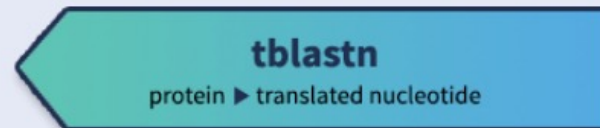
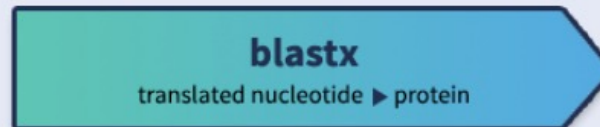
BLAST

Cunoașterea secvenței de aminoacizi nu este suficientă pentru a cunoaște rolul proteinei codificate de secvența inițială.

Cu ce peptidă cunoscută este similară secvența NM_000558?

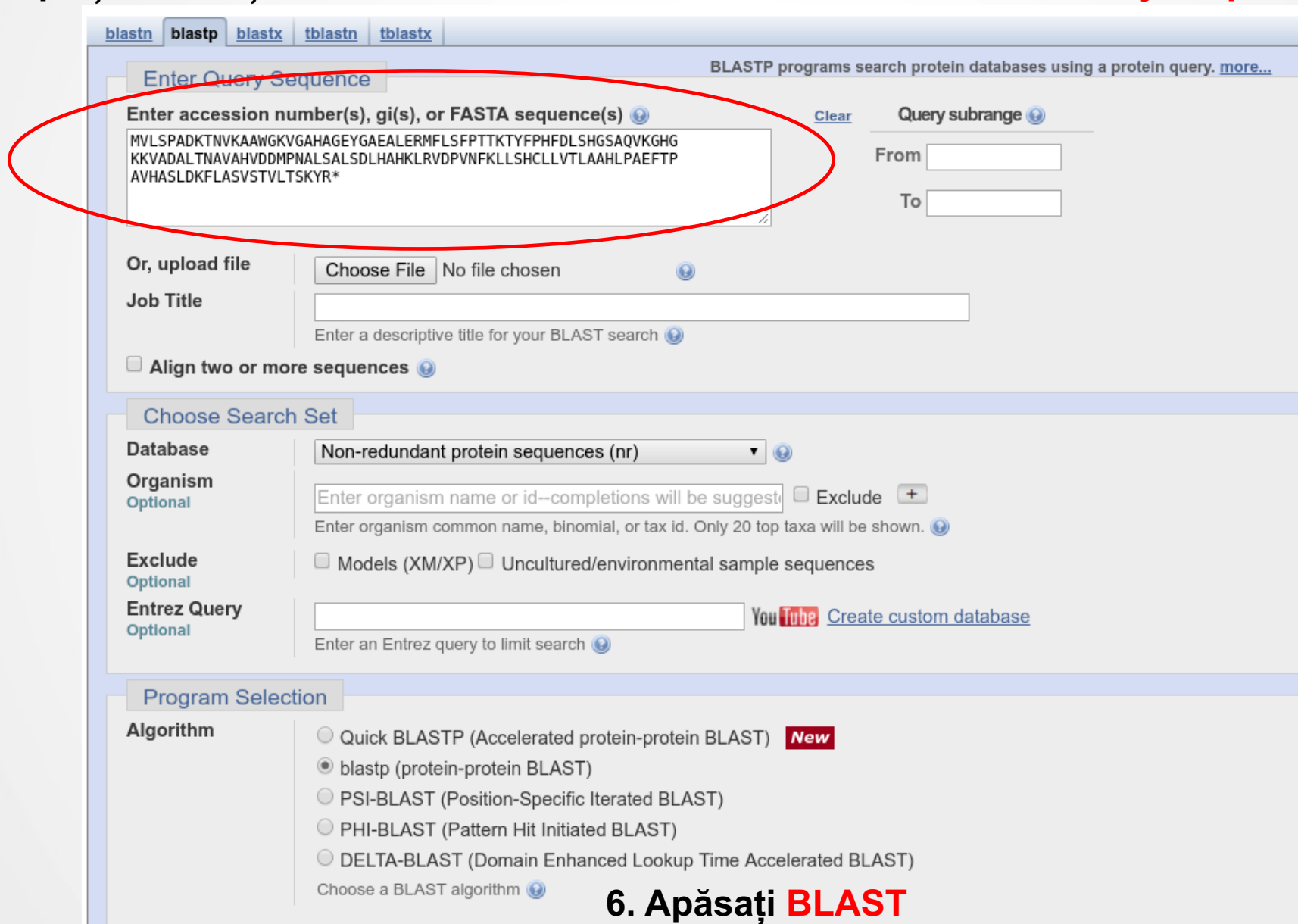
4. Accesați: <https://blast.ncbi.nlm.nih.gov> și selectați Protein BLAST

Web BLAST



BLAST

5. Copiați secvența de aminoacizi în caseta text intitulată: **Enter Query Sequence**



The screenshot displays the BLAST web interface. The 'Enter Query Sequence' section is highlighted with a red oval. It contains a text input field with the following amino acid sequence: MVLSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSSDLHAHKLRVDPVFNKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR*. Below the text field are options to 'Or, upload file' (with a 'Choose File' button and 'No file chosen' text) and a 'Job Title' field. There are also checkboxes for 'Align two or more sequences' and 'Exclude' options. The 'Choose Search Set' section includes a 'Database' dropdown set to 'Non-redundant protein sequences (nr)', an 'Organism' field, and 'Exclude' checkboxes for 'Models (XM/XP)' and 'Uncultured/environmental sample sequences'. The 'Entrez Query' field is empty. The 'Program Selection' section shows radio buttons for different algorithms: 'Quick BLASTP (Accelerated protein-protein BLAST)' (marked 'New'), 'blastp (protein-protein BLAST)' (selected), 'PSI-BLAST (Position-Specific Iterated BLAST)', 'PHI-BLAST (Pattern Hit Initiated BLAST)', and 'DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)'. A '6. Apăsați BLAST' instruction is at the bottom.

blastn blastp blastx tblastn tblastx

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

Enter Query Sequence

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

MVLSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG
KKVADALTNAVAHVDDMPNALSALSSDLHAHKLRVDPVFNKLLSHCLLVTLAAHLPAEFTP
AVHASLDKFLASVSTVLTISKYR*

From

To

Or, upload file No file chosen [?](#)

Job Title

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

Align two or more sequences [?](#)

Choose Search Set

Database [?](#)

Organism [Optional](#) Exclude

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

Exclude [Optional](#) Models (XM/XP) Uncultured/environmental sample sequences

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

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

Program Selection

Algorithm

Quick BLASTP (Accelerated protein-protein BLAST) **New**

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 [?](#)

6. Apăsați BLAST

Rezultate BLAST

Sumarul căutării

BLAST® » blastp suite » RID-8DZ4U15801R [Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

BLAST Results

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#) [YouTube](#) [How to read this page](#) [Blast report description](#)

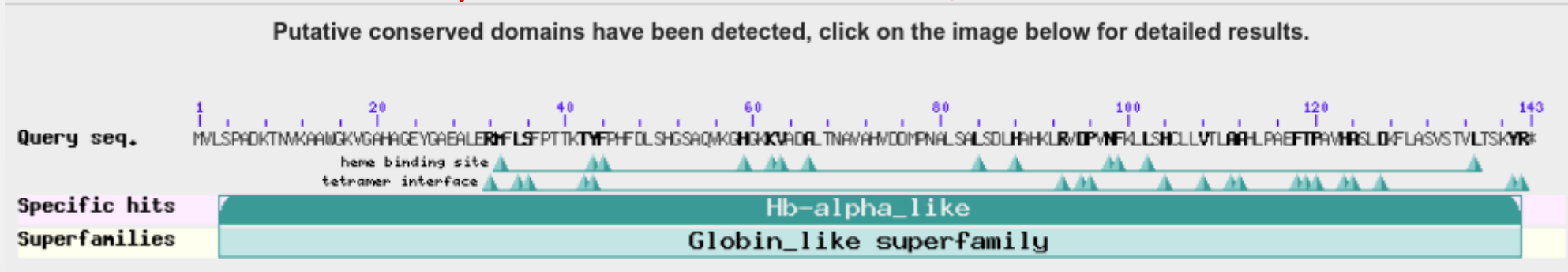
Job title: Protein Sequence (143 letters)

RID 8DZ4U15801R (Expires on 02-17 22:45 pm)	Database Name nr
Query ID Icl Query_129995	Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Description None	Program BLASTP 2.8.0+ Citation
Molecule type amino acid	
Query Length 143	

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

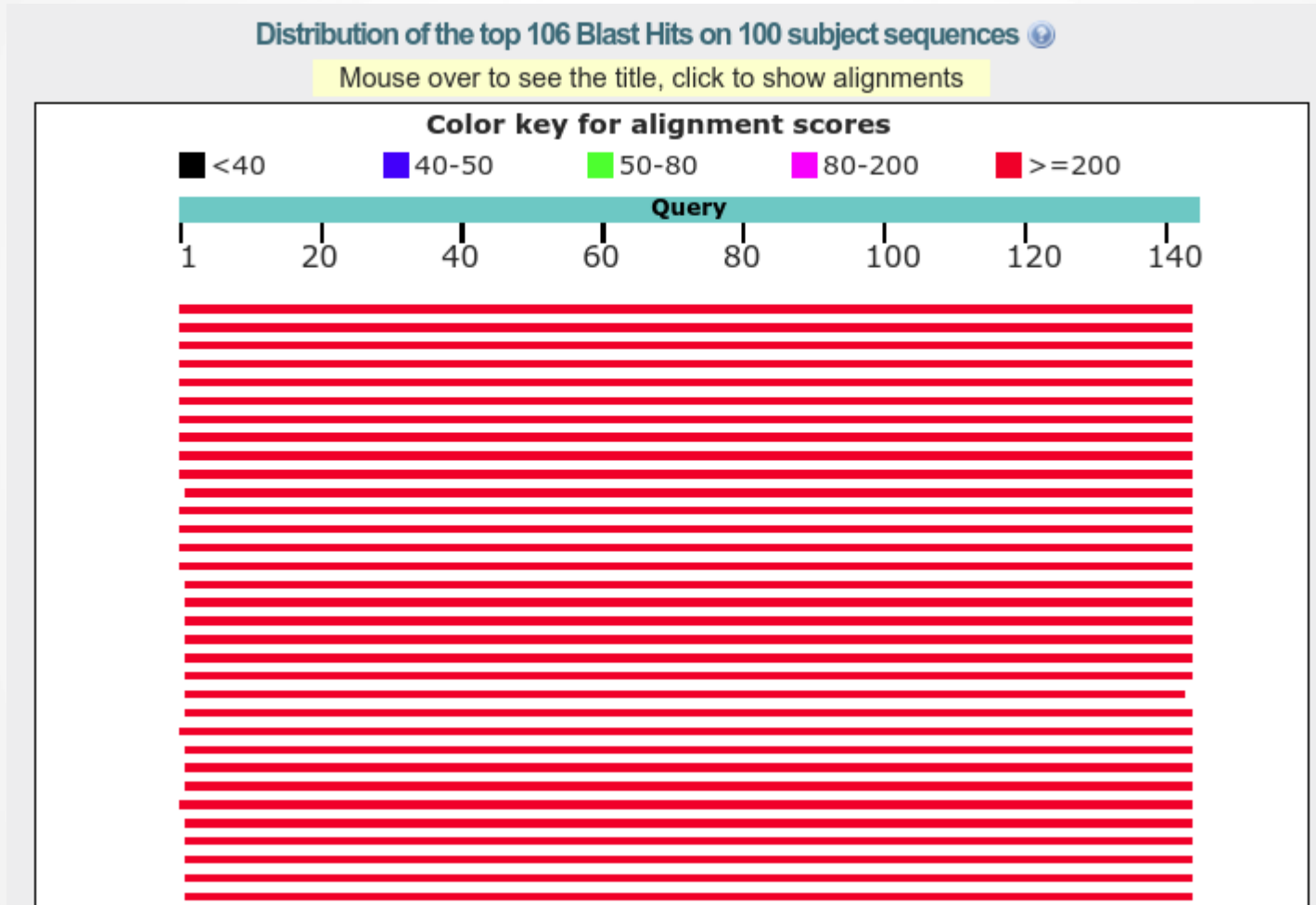
New Analyze your query with [SmartBLAST](#)

Domenii înalt conservate și situs-uri cunoscute de legare



Rezultate BLAST

Prezentarea grafică de ansamblu a rezultatelor;



Rezultate BLAST

Tabel cu secvențele identificate și scoruri BLAST

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	hemoglobin alpha 2 [synthetic construct]	287	287	99%	4e-98	100%	AAX29522.1
<input type="checkbox"/>	hemoglobin subunit alpha [Homo sapiens]	286	286	99%	6e-98	100%	NP_000508.1
<input type="checkbox"/>	Chain B, A Cis-Proline In Alpha-Hemoglobin Stabilizing Protein Directs The Structural Reorganization Of Alpha-Hemoglobin	286	286	99%	7e-98	100%	3IA3_B
<input type="checkbox"/>	mutant hemoglobin alpha 2 globin chain [Homo sapiens]	286	286	99%	1e-97	99%	AKZ66543.1
<input type="checkbox"/>	hemoglobin alpha-2 [Homo sapiens]	285	285	99%	2e-97	99%	AAN04486.1
<input type="checkbox"/>	alpha-2-globin [Homo sapiens]	285	285	99%	2e-97	99%	AAF72612.1
<input type="checkbox"/>	TPA: globin C1 [Homo sapiens]	286	286	99%	3e-97	100%	SAI82135.1
<input type="checkbox"/>	hemoglobin alpha 1-2 hybrid [Homo sapiens]	284	284	99%	5e-97	99%	ABF56145.1
<input type="checkbox"/>	Chain A, Hemoglobin Thionville: An Alpha-Chain Variant With A Substitution Of A Glutamate For Valine At Na-1 And Having An Acetylated Methionine Nh2 Term	284	284	99%	5e-97	99%	1BAB_A
<input type="checkbox"/>	hemoglobin alpha-1 globin chain [Homo sapiens]	284	284	99%	6e-97	99%	AAK37554.1
<input type="checkbox"/>	Chain A, Structure Of Haemoglobin In The Deoxy Quaternary State With Ligand Bound At The Alpha Haems	284	284	98%	7e-97	100%	1COH_A
<input type="checkbox"/>	alpha 2 globin variant [Homo sapiens]	284	284	99%	8e-97	99%	BAD97112.1
<input type="checkbox"/>	PREDICTED: hemoglobin subunit alpha [Rhinopithecus roxellana]	283	283	99%	8e-97	99%	XP_010380159.1
<input type="checkbox"/>	alpha-2 globin [Pongo pygmaeus]	283	283	99%	1e-96	99%	AAR29173.1
<input type="checkbox"/>	hemoglobin alpha 1 [Homo sapiens]	283	283	99%	1e-96	99%	AFI57164.1
<input type="checkbox"/>	Chain A, Hemoglobin (Alpha V1m) Mutant	283	283	98%	2e-96	99%	1BZZ_A
<input type="checkbox"/>	Chain A, Carbonmonoxy Structure Of Hemoglobin Evans Alphav62mbetawt	283	283	98%	2e-96	99%	4MQC_A
<input type="checkbox"/>	Chain A, Structure Of The Human Hemoglobin Mutant Hb Providence (a-gly-c:v1m; B,d:v1m,k82d; Ferrous, Carbonmonoxy Bound)	282	282	98%	2e-96	99%	1AJ9_A
<input type="checkbox"/>	Chain A, Structure Of The Human Hemoglobin Mutant Hb Providence (a-gly-c:v1m; B,d:v1m,k82d; Ferrous, Carbonmonoxy Bound)	282	282	98%	2e-96	99%	5SW7_A

Secvența de ADN conține gena ce codifică subunitate α a hemoglobinei

Exercitiu pentru prezentare

Deschideți fișierul fasta ce conține secvența moleculei proteice primite. Realizați o analiză BLAST și stabiliți numele/funcția proteinei respective. Salvați o imagine care să demonstreze concluzia dmv., introduceți-o în prezentare și scrieți o scurtă explicație.