

Obiectivele lucrării practice sunt:

- căutări în bazele de date (BD) după secvențe de nucleotide
- determinarea secvențelor de nucleotide responsabile de apariția unor diagnostice
- căutare în BD după secvența de aminoacizi
- găsirea gradului de potrivire al secvenței de aminoacizi introduse.

Colaborarea dintre informaticieni, medici cercetători, biologi, matematicieni și biochimisti le-a permis acestora studiul bazei moleculare a unei boli cu ajutorul instrumentelor matematice și a tehnicii de calcul prin:

- analiza secvenței unei gene sau a produsului genei de interes;
- înțelegerea mai bună a organizării genelor analizate;
- predicția structurii moleculelor analizate (proteine).

Disponibilitatea informației genomice oferă bioinformaticianului un nou set de provocări. În prezent, ariile predominante ale analizei datelor bioinformaticice includ:

- aliniamentul secvențelor
- predicția structurii proteinelor.

Studiile de aliniament a secvențelor sunt (în general) de două tipuri:

- de *aliniament al secvențelor în pereche* realizat cu ajutorul programului **BLAST**
- de *aliniament multiplu al secvențelor* realizat cu programe de tipul **CLUSTAL**.

În ambele cazuri ideea este de a găsi similaritățile sau diferențele dintre seturile de secvențe. Analiza secvenței reprezintă un instrument foarte important în studiul relațiilor de evoluție în genoame, duplicarea genelor, îmbinarea genelor etc.

Asocierea dezvoltărilor tehnologiei de calcul și moleculare deschide noi oportunități cercetărilor genetice. Folosirea combinată a informației oferită de secvențe, a instrumentelor de calcul, a bazelor de date și a biologiei tradiționale crește speranța înțelegerii funcției și reglajelor tuturor genelor și proteinelor, precum și a descifrării funcțiilor celulei.

Programul BLAST

Un program popular de comparare a secvențelor de ADN este **BLAST** (*Basic Local Alignment Search Tool*). BLAST face parte dintr-un pachet de programe destinat căutării de secvențe proteice, accesibil în diverse forme la diferiți furnizori, sau prin intermediul NCBI, care mai oferă și **Entrez**, un instrument de meta-căutare care acoperă mare parte a bazelor de date de la NCBI, inclusiv cele

care găzduiesc structuri tridimensionale a proteinelor, genoamele complete ale organismelor și trimiteri la jurnale științifice care însoțesc intrările din bazele de date.

BLAST reprezintă instrumentul de căutare a aliniamentului local de bază, fiind un set de programe de căutare a similarităților, creat pentru identificarea clasificării și a omologilor potențiali pentru o secvență dată.

Pentru a înțelege mai bine programele BLAST, trebuie cunoscute aspectele de bază ale aliniamentelor secvențelor. Acestea sunt folosite în special pentru găsirea potențialilor omologi ce vor fi folosiți ulterior pentru prezicerea posibilelor funcții ale secvenței necunoscute sau pentru modelarea structurii sale tridimensionale.

Aliniamentul global este cel mai bun aliniament, pe întreaga lungime a secvențelor specificate. Introducerea spațiilor (gaps) în secvențele respective permite alinierea lor pe întreaga lungime. Principalul avantaj al aliniamentului global este optimizarea sa pentru secvențele care au un grad înalt de similaritate, fiind astfel folositor în etapa de aliniere a secvențelor din procesul de modelare a structurii tridimensionale (bazat pe secvențele omologe cu structură tridimensională cunoscută).

Metodele de căutare ale aliniamentului local găsesc aliniamentul optim între subregiuni sau regiuni locale ale secvențelor specificate. Aliniamentul local este cel mai potrivit pentru secvențe care au regiuni localizate de similarități. Un program de căutare a aliniamentului local este folosit de exemplu pentru găsirea motivelor, domeniilor și altor unități repetitive din secvențele respective, precum și pentru găsirea secvențelor similare pentru secvența necunoscută într-o bază de date. Pe scurt, un program de căutare al aliniamentului local este cel mai bine folosit pentru identificarea unor regiuni secvențiale mai scurte, cu un grad foarte mare de similaritate.

Toți algoritmi de comparare a secvențelor se bazează pe anumite scheme de calcul a scorului aliniamentului. Scorul aliniamentului este suma scorurilor mai mici, atribuite pentru fiecare din perechile sale de aminoacizi sau nucleotide. Majoritatea acestor algoritmi folosesc o matrice de scor pentru calcularea unui scor total fiecărui aliniament. Teoria statistică folosită în programele BLAST a fost creată de Samuel Karlin și Steven Altschul. Toate programele BLAST folosesc o matrice de substituție, atât în etapa de scanare a bazelor de date cât și în procesul de aliniere a secvențelor.

Schemele de substituție sunt considerate a fi cele mai bune metode de calcul al scorului aliniamentelor și se bazează pe analiza frecvenței cu care un aminoacid observat este înlocuit de un alt aminoacid în proteinele ale căror secvențe sunt alinate.

Codurile acceptate de programele **BLAST** sunt:

A alanină	P prolină
B aspartat sau asparagină	Q glutamină
C cisteină	R arginină
D aspartat	S serină
E glutamat	T threonină
F fenilalanină	U selenocisteină
G glicină	V valină
H histidină	W triptofan
I izoleucină	Y tirozină
K lizină	Z glutamat sau glutamină
L leucină	X orice aminoacid
M methionină	* oprirea translației
N asparagină	- spațiu de lungime nedeterminată

Exemple

La adresa <https://blast.ncbi.nlm.nih.gov/Blast.cgi> se găsesc opțiunile programului BLAST :

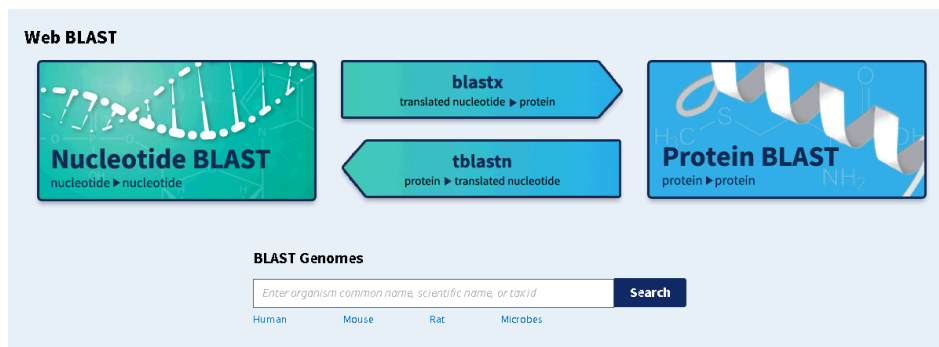


Figura 1. Interfața BLAST

- Dacă dorim să căutăm secvența de gene care codează receptorul pentru endotelină la om (figura 2.) introducem în căsuța search următoarele: **Endothelin receptor**.

Baza de Date integrată NCBI. Programul BLAST

sl.dr. Anca Tudor

The screenshot shows the NCBI Nucleotide search interface. The search query is "(Endothelin receptor) AND 'Homo sapiens'[porgn]_bid9806". The results show 86 nucleotide sequences. The first few results are listed with their accession numbers and descriptions:

- 1. 1,470 bp linear mRNA. Accession: M74921.1. GI: 182275. [GenBank](#) [FASTA](#) [Graphics](#)
- 2. [Homo sapiens ETR mRNA for endothelin receptor, complete cds](#). Accession: D90402.1. GI: 219651. [GenBank](#) [FASTA](#) [Graphics](#)
- 3. [Homo sapiens endothelin receptor type A \(EDNRA\), transcript variant 4, non-coding RNA](#). Accession: NR_045958.1. GI: 372266195. [GenBank](#) [FASTA](#) [Graphics](#)
- 4. [Homo sapiens endothelin receptor type B \(EDNRB\) mRNA, complete cds](#). Accession: U06623.1. GI: 181958. [GenBank](#) [FASTA](#) [Graphics](#)
- 5. [Homo sapiens endothelin receptor type A \(EDNRA\) mRNA, complete cds](#). Accession: U06622.1. GI: 181956. [GenBank](#) [FASTA](#) [Graphics](#)
- 6. [Homo sapiens endothelin receptor type A \(EDNRA\), transcript variant 3, mRNA](#). Accession: NM_001256283.1. GI: 372266193. [GenBank](#) [FASTA](#) [Graphics](#)
- 7. [Homo sapiens endothelin receptor type A \(EDNRA\), transcript variant 2, mRNA](#). Accession: NM_001166055.1. GI: 260593649. [GenBank](#) [FASTA](#) [Graphics](#)

Figura 2. Exemplu de căutare a unei secvențe de nucleotide

Din mulțimea de răspunsuri vom selecta varianta receptorului pentru specia umană (homo sapiens) si alegem urmatorul articol (Figura 2a):

The screenshot shows the NCBI Gene page for the gene **Homo sapiens endothelin receptor type A (EDNRA), transcript variant 2, mRNA**. The page displays various details about the gene, including its location, definition, accession numbers, keywords, and references.

LOCUS: NM_001166055 3841 bp mRNA linear PRI 15-OCT-2016

DEFINITION: Homo sapiens endothelin receptor type A (EDNRA), transcript variant 2, mRNA.

ACCESSION: NM_001166055

VERSION: NM_001166055.1

KEYWORDS: RefSeq

SOURCE: Homo sapiens (human)

ORGANISM: [Homo sapiens](#)

REFERENCE: 1 (bases 1 to 3841)

AUTHORS: Sek AC, Xie Z, Terai K, Long LM, Nelson C, Dudek AZ and Druey KM.

TITLE: Endothelial Expression of Endothelin Receptor A in the Systemic Capillary Leak Syndrome

JOURNAL: PLoS ONE 10 (7), E0133266 (2015)

PUBMED: [26176954](#)

REMARK: GeneRIF: Ednra, encoding Endothelin receptor A (ETA)-the target of Endothelin 1 (ET-1)-was significantly increased in Systemic Capillary Leak Syndrome blood-outgrowth endothelial cells compared to healthy controls. Erratum:[PloS One. 2015;10(9):e0137373. PMID: 26325591] Publication Status: Online-Only

REFERENCE: 2 (bases 1 to 3841)

AUTHORS: Salem SA, Gamai Aly D, Salah Yousef N and Moneim El-Sheer MA.

TITLE: Immunohistochemical assessment of endothelin-1 axis in psoriasis, basal cell carcinoma and squamous cell carcinoma

JOURNAL: G Ital Dermatol Venereol 150 (3), 283-291 (2015)

PUBMED: [25946671](#)

REMARK: GeneRIF: Endothelin receptor A was detected in all patients with squamous cell carcinoma and psoriasis, with a higher frequency and grade of expression than controls and basal cell carcinoma.

REFERENCE: 3 (bases 1 to 3841)

AUTHORS: Cipriani P, Di Benedetto P, Ruscitti P, Verzella D, Fischietti M, Zaccaroni F, Liakouli V, Carubbi F, Berardicurti O, Alesse E and Giacomelli P.

TITLE: Macitentan inhibits the transforming growth factor-beta profibrotic action, blocking the signaling mediated by the ETR/BetaRI complex in systemic sclerosis dermal fibroblasts

JOURNAL: Arthritis Res. Ther. 17, 247 (2015)

PUBMED: [26337964](#)

REMARK: GeneRIF: Data show that macitentan interferes with the profibrotic action of transforming growth factor-beta (TGF-beta), blocking the endothelin receptor type A (ET-1 receptor) portion of the ET-1/TGF-beta receptor complex. Publication Status: Online-Only

REFERENCE: 4 (bases 1 to 3841)

AUTHORS: Miyamoto Y, Yoshimasa T, Arai H, Takaya K, Ogawa Y, Itoh H and Nakao K.

Customize view: [Run BLAST](#) [Pick Primers](#) [Highlight Sequence Features](#) [Find in this Sequence](#)

Articles about the EDNRA gene

- Macitentan inhibits the transforming growth factor-β profibrotic actio [Arthritis Res Ther. 2015]
- The BioPlex Network: A Systematic Exploration of the Human Interactome. [Cell. 2015]
- Endothelial Expression of Endothelin Receptor A in the Systemic Capillary Leak [PloS One. 2015]

Pathways for the EDNRA gene

- Gastrin-CREB signalling pathway via PKC and MAPK
- G alpha (q) signalling events
- GPCR downstream signaling

Reference sequence information

RefSeq alternative splicing
See 3 reference mRNA sequence splice variants for the EDNRA gene.

RefSeq protein product
See the reference protein sequence for endothelin-1 receptor isoform b precursor (NP_001159527.1).

More about the EDNRA gene
This gene encodes the receptor for endothelin-1, a peptide that plays a role in potent and long-lasting vasoconstriction. This receptor asso... Also Known As: ET-A, ETA, ETAR, ETAR...

Figura 2a. Informatii despre articol

Va fi afișat cromozomul pe care se află secvența de nucleotide (cromozomul 40) – figura 2b, localizarea genei pe cromozom, secvența de nucleotide și secvența codantă.

FEATURES	Location/Qualifiers
source	1..3841 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /chromosome="4" /map="4q31.22-q31.23"
gene	1..3841 /gene="EDNRA" /gene_synonym="ET-A; ETA; ETA-R; ETAR; ETRA; hET-AR; MFDA" /note="endothelin receptor type A" /db_xref="GeneID:1909" /db_xref="HGNC:HGNC:3179" /db_xref="MIM:131243"
exon	1..460 /gene="EDNRA" /gene_synonym="ET-A; ETA; ETA-R; ETAR; ETRA; hET-AR; MFDA" /inference="alignment:Splice:1.39.8"
misc feature	30 /gene="EDNRA" /gene_synonym="ET-A; ETA; ETA-R; ETAR; ETRA; hET-AR; MFDA" /note="alternative transcription initiation site"
STS	46..236 /gene="EDNRA" /gene_synonym="ET-A; ETA; ETA-R; ETAR; ETRA; hET-AR; MFDA" /standard_name="SHGC-67921" /db_xref="UniSTS:3684"
exon	461..950 /gene="EDNRA" /gene_synonym="ET-A; ETA; ETA-R; ETAR; ETRA; hET-AR; MFDA" /inference="alignment:Splice:1.39.8"
misc feature	501..503 /gene="EDNRA" /gene_synonym="ET-A; ETA; ETA-R; ETAR; ETRA; hET-AR; MFDA" /note="upstream in-frame stop codon"
CDS	531..1487 /gene="EDNRA" /gene_synonym="ET-A; ETA; ETA-R; ETAR; ETRA; hET-AR; MFDA" /note="isoform b precursor is encoded by transcript variant 2; endothelin receptor subtype A;

Datele despre secventa de nucleotide
cromozomul, bratul cromozomului, gena

Figura 2b.

Ca secventa de nucleotide si aminoacizi, gasim urmatoarea formulare (fig.2b).

```

424 agtcacccg cgtgtctgac gattgtggag aggcgggtgga gaggcttcat ccatccacc
181 agtcacccg cgtgtctgac gattgtggag aggcgggtgga gaggcttcat ccatccacc
241 cgtgtctgac cgtgtctgac gattgtggag aggcgggtgga gaggcttcat ccatccacc
301 gggttttctga agccggggaa gctgtgcagc ggaagccgcc gccgcgccgg agcccgggag
361 accgggaccc ctcgagccca cccaccctcg ccggtccggg cttctcttgg cccagggccc
421 gcggggaccc ggcagctgtc tgcgcacgcc ggcctccagc gtgaaaaaaa agtgaagggt
481 taaaagcagc acaagtgcaa taagagatat ttctcaaat ttgcctcaag atggaacccc
541 ttgcctcag ggcctcctt ttgcctggcag tgggtggatg tgaatcagt gataatcctg
601 agagatagac cacaaatcta agcaatcatg tggatgatt caccactttt cgtggcacag
661 agctcagctt cctgtttacc actcatcaac ccactaattt ggtcctaccc agcaatggct
721 caatgcacaa ctattgcccc cagcagacta aaattacttc agctttcaaa tacattaaca
781 ctgtgatctc ttgtactatt ttcatctgga gaatgggtgg gaatgcaact ctgtcagga
841 tcatttacca gaacaaatgt atgaggaatg gcccaacgc cgtgatagcc agtcttgccc
901 ttggagacct tatctatgtg gtcattgac ttctatcaa tgaatttaag ttctaccaag
961 atgtaaggaa ctgttgcttc ttccggttct atttctgat gccctgggtg tgcactgcca
1021 tcttctacac cctcatpact tctgagatgt tgaacagaa gaatggcagc ttgagaattg
1081 ccttcagtg aacatcttaag cagcgtcag aatgggcaa aacagttttc tgcctgggtg
1141 taatttttgc tcttctgctg ttccctcttc atttaagcgc tatattgaag aaaaactgtgt
1201 ataacagatg ggaacagaa cgaatggaat tacttatttt ctactgtctc atggaattaca
1261 tctgtattaa ctgtgcaacc atgaattcat gtataaacc catagctctg tatittgtga
1321 gcaagaaatt taaaatttgt ttccagtcac gcctctgctg ctgctgttac cagtccaaaa
1381 gtctgtgac ctgcgtcccc atgaacggaa caagcatcca gtggaagaac cagatcaaaa
1441 acaaccacaa cacagaccgg agcagccata aggaagcat gaactgcca cccttagaag
1501 cactctctcg tactcccata atcctctcgg agaaaaaaat cacaaggcaa ctgtgagtc
1561 gggaatctct tctctgaccc ttcttctcta attcactccc acaccaaga agaatgctt
1621 tccaaaaccg caaggtaga ctggtttatc caccacaac atctacgaat cgtactcttt
1681 taattgatct aatttacata ttctgcgtgt tgtattcagc actaaaaaat ggtggggact
1741 gggggagaaat gaagactgtt aaatgaaac agaagatat ttactacttt tgcataaaaa
1801 tagagctttc aagcatatgg ctagctttta tggcagttct ggtgaatgtt caatgggaac
1861 ttgtcaccat gaactcttag agattaaaga caagattttc tacttttttt aagtattttt
1921 ttgtcctctc agccaaacac aatatgggct caagtcactt tactttttaa tgcattttg
1981 ttgcagctatt tttaactgc ataatagcct aacatgatta ttgaactta ttacacata
2041 gtttgaaaaa aaaaagacaa aaatagattt cagtgagca attagattag tattttccac
2101 gtcactgttt ttttttttaa aacaaaatt ctaaagctac aacaaatact acaggccctt

```

secventa de nucleotide

secventa de aminoacizi

```

531..1487
/gene="EDNRA"
/gene_synonym="ET-A; ETA; ETA-R; ETAR; ETRA; hET-AR; MFDA"
/note="isoform b precursor is encoded by transcript variant 2; endothelin receptor subtype A; endothelin-1-specific receptor; G protein-coupled receptor"
/codon_start=1
/product="endothelin-1 receptor isoform b precursor"
/protein_id=" NP_001159527.1 "
/db_xref="CCDS: CCDS54810.1 "
/db_xref="GeneID: 1909 "
/db_xref="HGNC: HGNC:3179 "
/db_xref="MIM: 131243 "
/translation="METLCRLASFNLVGCVISDNPERYSTNLSNHVDDFTFRGTE
LSFLVTTHTPTNLVLPNSGSIHIVCPQQTITSAFYINTVISTCIFIIVMGVGNATLL
RIIVQNKCHRNPHIALSLALGDLIYVVDLPINVKFYQDVVDMLFGFYFCHPLV
CTAIFYLTCEMLNRRNGSLRLSEHLKQREKATVFCLVVFALCFPLHLSRI
LKKTVYNEMDKNRCELLSFLLDYIDIGINLATMNSCINPIALYFVSKKFKNCFQSLC
CCCVQSKSLMSTVPPNWTISIQKINHQNHNHNTDRSSHKDSMN"

```

Figura 2c. Formatul GenBank de vizualizare

Daca alegem optiunea „Fasta” vom avea ca rezultat secventa de nucleotide din figura 2c.

Daca optam pentru varianta „Graphics” obtinem reprezentarea din figura 2d.

Homo sapiens endothelin receptor type A (EDNRA), transcript variant 2, mRNA

NCBI Reference Sequence: NM_001166055.1

[GenBank](#) [FASTA](#)

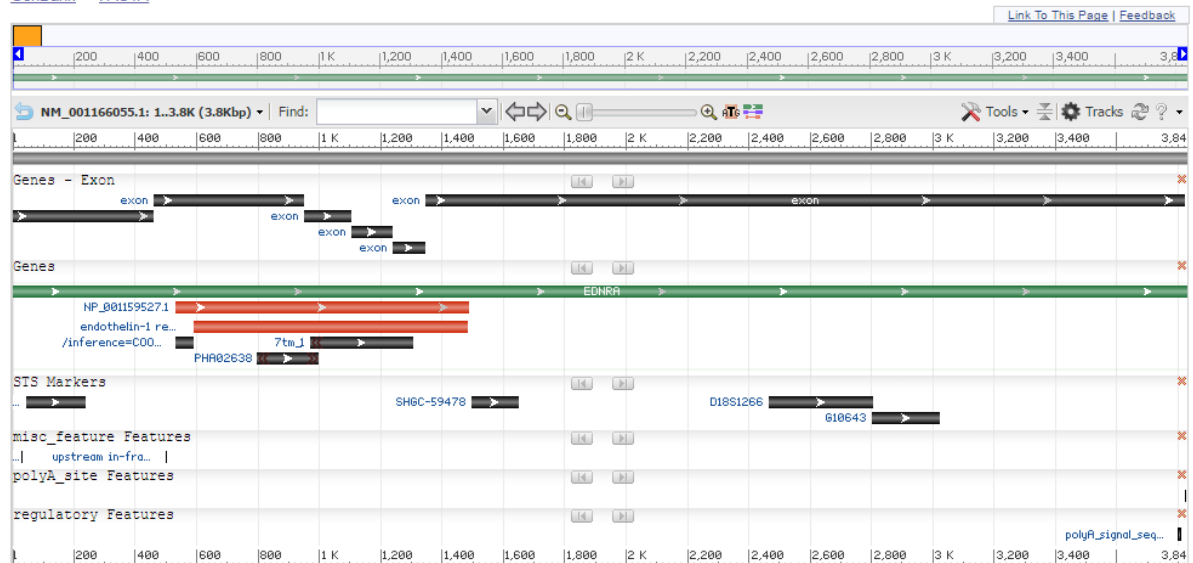


Figura 2d. Vizualizare prin „Graphics”

BIOINFORMATICA 1. Exercitii practice

Anca TUDOR, Diana LUNGEANU

NOTA: pentru secvențele-text accesați fișierul *Bioinfo-2017-practic1.txt*

Exercitiu 1

National Center for Biotechnology Information: <https://www.ncbi.nlm.nih.gov/>

BLAST ← Basic Local Alignment Search Tool

Nucleotide BLAST → Enter Query Sequence:

```
1 ctagaaaccg tatgctatat aattatatac tataaagtaa taatgtatac agtghtaatgg
61 atcatggggcc atgtgctttt caaactaatt gtacataaaa caagcatcta ttgaaaatat
121 ctgacaaact catcttttat ttttgatgtg tgtgtgtgtg tgtgtgtgtg tttttttaac
181 aggggatttgg gg
```

(căutarea este comandată cu butonul BLAST)

Observați scorul de potrivire → investigați/căutați informațiile referitoare la gena *Homo sapiens Human cystic fibrosis transmembrane conductance regulator* (CFTR), Locus AH002646.

Observați adnotările/referințele bibliografice – coloana din dreapta

Exercitiu 2

National Center for Biotechnology Information: <https://www.ncbi.nlm.nih.gov/>

Protein → Search for "insulin" → alegeți "insulin[Homo sapiens]"

Observați adnotările/referințele bibliografice – coloana din dreapta

Observați secvența de aminoacizi ai proteinei – **ORIGIN:**

```
1 malwmrllpl lallalwgp paaafvnhl cgshlvealy lvcgergffy tpktrreaed
61 lqvqqvelgg gpgagslqpl alegslqkrg iveqcctsic slyqlenycn
```

LOCUS AAA59172 110 aa linear PRI 10-JUN-2016

DEFINITION insulin [Homo sapiens].

ACCESSION AAA59172

VERSION AAA59172.1

DBSOURCE accession AH002644.2

KEYWORDS .

SOURCE Homo sapiens (human)

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

REFERENCE 1 (residues 1 to 110)

AUTHORS Bell, G.I., Pictet, R. and Rutter, W.J.

TITLE Analysis of the regions flanking the human insulin gene and sequence of an Alu family member

JOURNAL Nucleic Acids Res. 8 (10), 4091-4109 (1980)

PUBMED 6253989

COMMENT On Aug 28, 1993 this sequence version replaced gi:186432. Method: conceptual translation.

FEATURES

source 1..110

/organism="Homo sapiens"

/db_xref="taxon:9606"

/map="tip15.5"

/tissue_type="liver"

/dev_stage="foetus"

1..110

/product="Insulin"

1..24

/gene="INS"

/note="G00-119-349"

25..110

/gene="INS"

/product="C peptide; G00-119-349"

26..110

/region_name="T1Gf_insulin_like"

/note="T1Gf_like family, insulin_like subgroup, specific to vertebrates. Members include a number of peptides including insulin and insulin-like growth factors I and II, which play a variety of roles in controlling processes such as metabolism, growth and...; cd04367"

db_xref="CDD:239533"

order(32,35,37,108)

/site_type="other"

/note="putative receptor binding surface"

/db_xref="CDD:239533"

1..110

/gene="INS"

/coded_by="join(AH002644.2:2424..2610, AH002644.2:3397..3542)"

/note="precursor"

/db_xref="G00-119-349"

ORIGIN

1 malwmrllpl lallalwgp paaafvnhl cgshlvealy lvcgergffy tpktrreaed

61 lqvqqvelgg gpgagslqpl alegslqkrg iveqcctsic slyqlenycn

Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Protein 3D Structure

Insulin With Proline Analog Hsp A1 Position B28 In The R6 State

PDB: 5HRQ

Source: Homo sapiens

Method: X-Ray Diffraction

Resolution: 1.28 Å

See all 226 structures...

Articles about the INS gene

Associations of Dietary Glucose, Fructose, and Sucrose with β-Cell Function. In [Nutrients. 2017]

Fibrin scaffold enhances function of insulin producing cells differentiated in Tissue Cell. 2017

Polymorphic distribution of proteins in solution by mass spectrometry. The analy [Biologicals. 2017]

See all...

Pathways for the INS gene

Longevity regulating pathway - multiple species

Longevity regulating pathway

Phospholipase D signaling pathway

See all...

Reference sequence information

RefSeq genomic sequence

See the genomic reference sequence for the INS gene (NG_050578.1).

RefSeq protein sequence

See the genomic reference sequence for the INS gene (NG_007114.1).

RefSeq protein isoforms

See 4 reference sequence protein isoforms for the INS gene.

Exercitiu 3

3.1. Protein BLAST → Enter Query Sequence:

malwmrllpl

BLAST® → blastp suite

Standard Protein BLAST

Enter Query Sequence

malwmrllpl

1

2

BLAST

Observați scorul de potrivire....

3.2. Protein BLAST → Enter Query Sequence:

malwmrllpl lallalwgpdpaaafvnqhlcgshlvealy

Observați scorul de potrivire → investigați/căutați informațiile referitoare la proteina *insulin* [*Homo sapiens*]

Explorați reprezentarea grafică și adnotările bibliografice.

Exercitiu 4

National Center for Biotechnology Information: <https://www.ncbi.nlm.nih.gov/>

Protein BLAST → Enter Query Sequence:

```
1 mqnqagasrt stflngnre rplnvfcdme tdgggwlvfq rrmqgqtdfw rdwedyahgf
61 gnisgefhwlg nealhsltqa gdysirvdlr agdeavfaqy dsfhvdsaae yyrhllegyh
121 gtagdmsyh sgsvfsardr dpnslisla vsyrgawwyr nchyanlngl ygstvdhggv
181 swyhwkgfef svpftemklr prnfrspagg g
```

Pentru *fibrinogen* [*Homo sapiens*] → accesați structura 3D.

GenPept

fibrinogen [Homo sapiens]

GenBank: CAA50740.1

Identical Proteins FASTA Graphics

LOCUS CAA50740 211 aa Linear PRI 14-NOV-2006

DEFINITION fibrinogen [Homo sapiens].

ACCESSION CAA50740

VERSION CAA50740.1

RESOURCE ebl accession X7A337.1

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorhini; Catarrhini; Hominidae; Homo.

REFERENCE

1. (residues 1 to 211)

AUTHORS Bristow, J., Ten, J.K., Gitzelmann, S.E., Hellen, S.H. and Hiller, H.L.

TITLE Tenascin-X: a novel extracellular matrix protein encoded by the human X8 gene overlapping P450C218

JOURNAL Cell Biol. 122 (3): 369-378 (1993)

PMID 7581653

REFERENCE

2. (residues 1 to 211)

AUTHORS Bristow, J.D.

TITLE Direct Subcloning

JOURNAL Submitted (12-NOV-1993) J.D. Bristow, University of California, Dept of Pediatrics, Box 0544, 3rd and Parnassus Ave., San Francisco, CA 94143, USA

LOCATION/Qualifiers

source

1..211

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/clone="cDNA:2-1"

/clone_id="cDNA:2-1"

/clone_id="cDNA:2-1"

/dev_stage="adult"

/note="Class III region of HNC"

1..211

/product="fibrinogen"

1..202

/region_name="RBD"

/note="Fibrinogen-related domains (FRDs); C-terminal globular domain of fibrinogen. Fibrinogen is involved in blood clotting, being activated by thrombin to assemble into fibrin clots. The N-terminal of 2 times 3 chains come together to form a globular...; c100085"

/db_xref="CDD:226054"

113

/site_type="other"

/note="gamma-gamma dimer interface [polypeptide binding]"

/db_xref="CDD:226055"

order(141,145,145)

/site_type="other"

Protein

Region

Site

Site

Protein 3D Structure

The Solution Structure Of The 3rd Fibronectin Type III Domain Of Human Tenascin-PDB: 2CUM

Source: Homo sapiens

Method: Solution Nmr

See all 3 structures...

Articles about the TNXB gene

Recognizing the tenascin-X deficient type of Ehlers-Danlos syndrome: a ctc [Clin Genet. 2017]

Rare variants in tenascin genes in a cohort of children with primary ves [Pediatr Nephrol. 2015]

Serotonin identifies Tenascin-X as a Potent Marker of Ovarian Cancer [Biomol Res Int. 2015]

See all...

Pathways for the TNXB gene

MicroRNAs in cancer

PI3K-Akt signaling pathway

ECM-receptor interaction

See all...

Reference sequence information

RefSeq genomic sequence

See the genomic reference sequence for the

Exercitiu 5

National Center for Biotechnology Information: <https://www.ncbi.nlm.nih.gov/>

Protein → Advanced Search for "human fibrinogen"

Protein Advanced Search Builder

(fibrinogen[Protein Name]) AND human[Organism]

Edit Clear

Builder

Protein Name fibrinogen Show index list

AND Organism human Show index list

AND All Fields Show index list

Search or Add to history

History Download history Clear history

Search	Add to builder	Query	Items found	Time
#9	Add	Search fibrinogen	112309	14:11:24
#6	Add	Search insulin	83725	13:47:56

You are here: NCBI > Proteins > Protein Database Support Center

Căutați informații pe Wikipedia: <https://en.wikipedia.org/wiki/Fibrinogen>

Exercitiu 6

National Center for Biotechnology Information: <https://www.ncbi.nlm.nih.gov/>

Protein BLAST → Enter Query Sequence:

```
1 mkwvtfisll flfssaysrg vfrrdahkse vahrfdklge enfkalvlia faqylqqcpf
61 edhvkvlnev tefaktcvad esaencdksl htlfgdklct vatlretyge madccakqep
121 ernecflqhk ddnplnprlv rpevdvmcta fhdneetflk klyeiarrh pyfyapellf
181 fakrykaaft eccqaadkaa clpkldelr degkassakq rlkcaslqkf gerafkawav
241 arlsqrfpka efaevsklvt dltkvhtecc hgdlllecadd radlakyice nqdsissklk
301 eccekpllek shciaevend empadlpsla adfveskdvc knyaeakdvf lgmflyeyar
361 rhpdydvll lrlaktyett lekccaaadp hecyakvfde fkplveepqn likqncelfe
421 qlgeykfqna llvrytkkvp qvstptlvev srnlgkvgs cckhpeakrm pcaedylsvv
481 lnqlcvlhek tpsdrvtkc cteslvnrrp cfsalevdet yvpkefnaet ftfhadictl
541 sekerqikkq talvelvkhk pkatkeqlka vmddfaafve kcckaddket cfaeegkklv
601 aasqaalgl
```

Observați și analizați rezultatele.

Accesați opțiunea "Analyze your query with SmartBLAST".

Exercitiu 1 (solutie):

În căsuța de interogare (query) sa introducem următoarea secvență:

```
"1 ctgaaaccg tatgctatat aattatgtac tataaagtaa taatgtatac agtgaatgg
61 atcatgggcc atgtgctttt caaactaatt gtacataaaa caagcatcta ttgaaaatat
121 ctgacaaact catcttttat tttgatgtg tgtgtgtgtg tgtgtgtgtg ttttttaac
181 agggatttgg gg"
```

Apoi vom obtine aliniamentul, ca în figura 3.

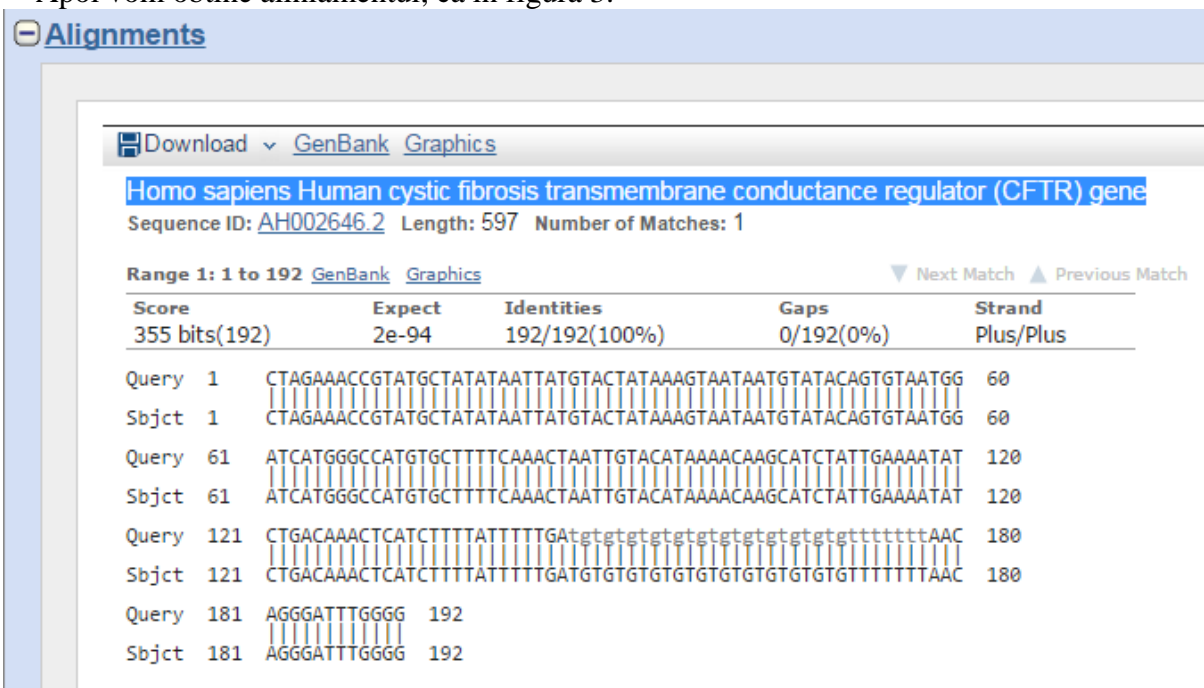


Figura 3. Rezultatul interogării

Adica, secvența de nucleotide responsabilă de apariția fibrozei chistice.

Exercitiu 2 (solutie):

Daca introducem in „BLAST Protein” secventa de aminoacizi:

```
„metlclrasfwlalvgcvisdnperrystnlshvddfttfrgtelsflvtthqptnlv
lpsngsmhnycpqtkitsafkyintvisctifivgmvgnatllriiyqnkcmrngpnalia
slalgdliyvvidlpinvfkfyqdvkdwlfgfyfcmplvctaifytlmtcemlnrrngslr
ialsehlkqrrevaktvfcclvifalcwfpplhlsrikktvynemdknrcellsflllmdyig
inlatmnsncinpialyfvskkfkncfqsclccccyqskslmtsvpmngtsiqwknhdqnnh
tdrsshkdsnm"
```

Vom obtine urmatorul rezultat de potrivire al secventei – figura 4.

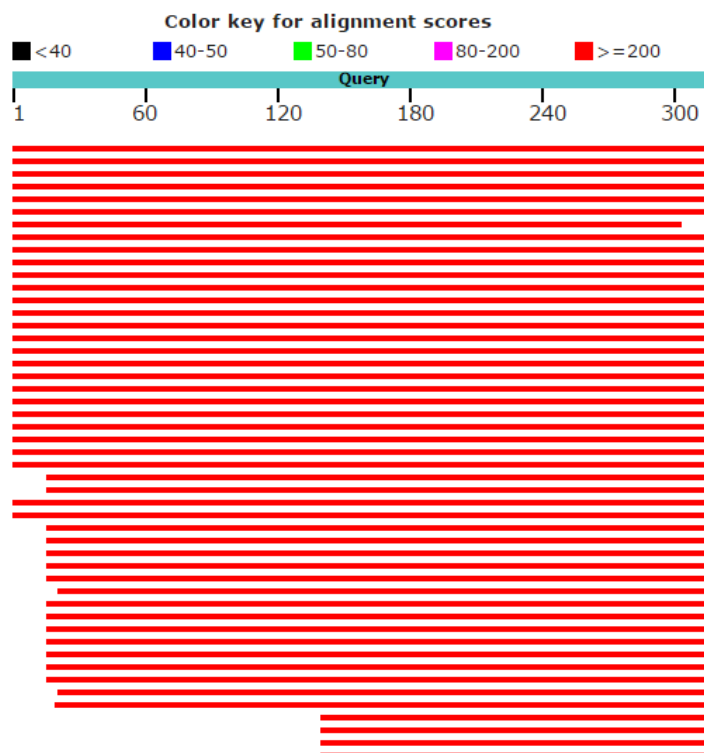


Figura 4. Paleta de culori pentru scorurile alinierii secventelor proteice

Culoarea rosie pentru aliniament ne arata ca avem o potrivire perfecta cu o secventa cunoscuta, iar aceasta potrivire, semnificativa, cu scorul cel mai mare, este pentru „*endothelin-1 receptor isoform b precursor [Homo sapiens]*” – scorul 654 din scorul total de 654, iar aliniamentul se poate observa in figura 5.

Download [GenPept](#) [Graphics](#)

endothelin-1 receptor isoform b precursor [Homo sapiens]
 Sequence ID: [NP_001159527.1](#) Length: 318 Number of Matches: 1
[See 2 more title\(s\)](#)

Range 1: 1 to 318 [GenPept](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
654 bits(1687)	0.0	Compositional matrix adjust.	317/318(99%)	317/318(99%)	1/318(0%)
Query 1	METLCLRASFWLALVGCVISDNPERYSTNLSNHVDDFTTFRGTLSFLVTTHOPTNLVLP				60
Sbjct 1	METLCLRASFWLALVGCVISDNPERYSTNLSNHVDDFTTFRGTLSFLVTTHOPTNLVLP				60
Query 61	SNGSMHNYCPOOTKITSFAFKYINTVISCTIFIVGMVGNATLLRIIYONKCMRNGPNALIA				120
Sbjct 61	SNGSMHNYCPOOTKITSFAFKYINTVISCTIFIVGMVGNATLLRIIYONKCMRNGPNALIA				120
Query 121	SLALGDLIYVVIDLPINVFKFYQDVKDWWLFGFYFCMLPVCTAIFYTLMTCEMLNRRNGS				180
Sbjct 121	SLALGDLIYVVIDLPINVFKFYQDVKDWWLFGFYFCMLPVCTAIFYTLMTCEMLNRRNGS				180
Query 181	LRIALSEHLKQRRVAKTVFCLVIFALCWFLHLSRI - KKTVYNEMDKNRCELLSFLLL				239
Sbjct 181	LRIALSEHLKQRRVAKTVFCLVIFALCWFLHLSRI KKTVYNEMDKNRCELLSFLLL				240
Query 240	MDYIGINLATMNSCINPIALYFVSKKFKNCFQSCLCCLCCYQSKSLMTSVPMMNGTSIQWKN				299
Sbjct 241	MDYIGINLATMNSCINPIALYFVSKKFKNCFQSCLCCLCCYQSKSLMTSVPMMNGTSIQWKN				300
Query 300	HDQNNHNTDRSSHKDSMN 317				
Sbjct 301	HDQNNHNTDRSSHKDSMN 318				

Figura 5.

Dacă ne propunem să căutăm proteina insulină introducem în căsuța de search **insulin** (figura 6).

Protein [Create alert](#) [Advanced](#) [Help](#)

[GI numbers no longer appear by default on sequence records. Read more.](#)

Species: [Animals \(32,037\)](#) [Plants \(149\)](#) [Fungi \(2,061\)](#) [Protists \(91\)](#) [Bacteria \(24,473\)](#) [Archaea \(32\)](#) [Viruses \(118\)](#) [Customize ...](#)

Source databases: [PDB \(2,723\)](#) [RefSeq \(18,837\)](#) [UniProtKB / Swiss-Prot \(2,477\)](#) [Customize ...](#)

Genetic compartments: [Chloroplast \(1\)](#) [Mitochondrion \(1\)](#) [Plasmid \(5\)](#) [Plastid \(1\)](#)

Sequence length: [Customize ...](#)

Summary [20 per page](#) [Sort by Default order](#) [Send to:](#) [Filters: Manage Filters](#)

See [PIN \(INSULIN\) insulin precursor](#) in the Gene database
 insulin reference sequences [Transcript \(1\)](#) [Protein \(1\)](#)

Items: 1 to 20 of 77218

1. [insulin precursor \[Aplysia californica\]](#)
 156 aa protein
 Accession: NP_001191615.1 GI: 325296757
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

2. [insulin \[Octodon degus\]](#)
 109 aa protein
 Accession: AAA40590.1 GI: 202472
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

3. [insulin preproprotein \[Danio rerio\]](#)
 108 aa protein

Results by taxon

Top Organisms [\[Tree\]](#)

[Bacillus anthracis \(5666\)](#)
[Bacillus megaterium \(5616\)](#)
[Bacillus circulans \(4950\)](#)
[Homo sapiens \(4721\)](#)
[Bacillus altitudinis \(3985\)](#)
[All other taxa \(52280\)](#)
[More...](#)

Find related data

Database: [Select](#) [Find items](#)

Search details

[insulin\[All Fields\]](#)

Figura 6. Exemplu de căutare a insulinei

Vom alege insulina corespunzătoare speciei umane (homo sapiens) și vom găsi informația potrivit căreia insulina umană este o proteină formată din 110 aminoacizi a căror reprezentare o puteți vedea în figura 6a.

```

Site      order(32,35,37,108)
          /site_type="other"
          /note="putative receptor binding surface"
          /db_xref="CDD:239833"
CDS       1..110
          /gene="INS"
          /coded_by="join(AH002844.2:2424..2610,
          AH002844.2:3397..3542)"
          /note="precursor"
          /db_xref="GDB:G00-119-349"
ORIGIN
1 malwmrllpl lallalwgpd paaafvnqhl cgshlvealy lvcgergffy tpktrreaed
61 lqvgqvelgg gpgagslqpl alegslqkrg iveqcctsic slyqlenycn
//
    
```




Figura 6a. Secvența de aminoacizi ai proteinei insulina umană

În josul paginii va fi afișată secvența de aminoacizi din care e formată insulina, începând de la origine. După cum știți de la curs, fiecare aminoacid e codat cu o singură literă.

Exercitiu 3 (soluție):

Introduceți secvența de aminoacizi **malwmrllpl** (figura 7).

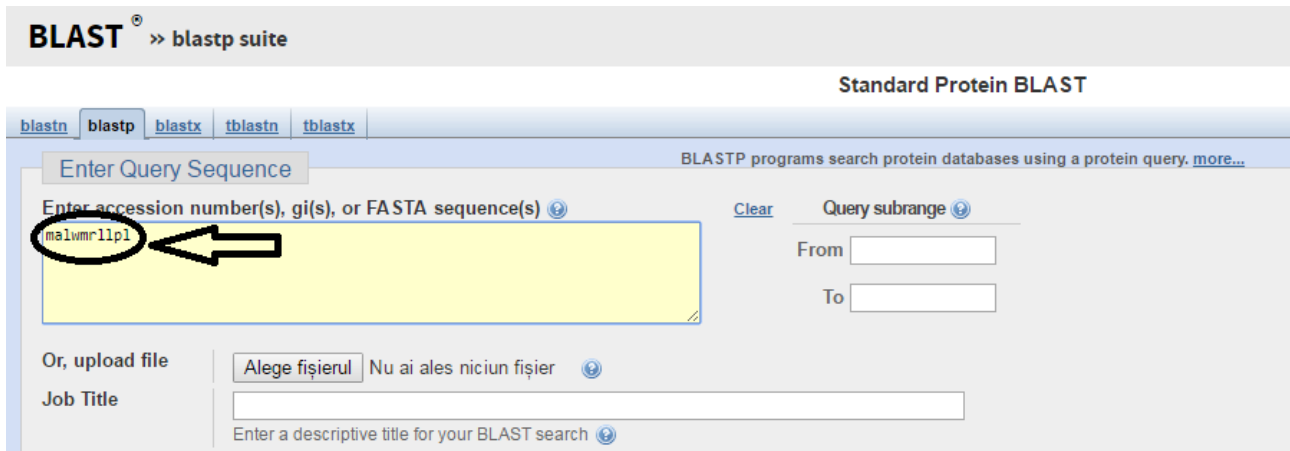


Figura 7. Interogarea BD după o secvență de aminoacizi

Apoi alegeți opțiunea cu butonul BLAST din josul paginii. Ca rezultat veți avea un grafic care va afișa gradul de potrivire al secvenței de aminoacizi introdusă (figura 7a). Sub grafic obțineți și rezultate text în ordinea potrivirii lor cu secvența introdusă inițial.

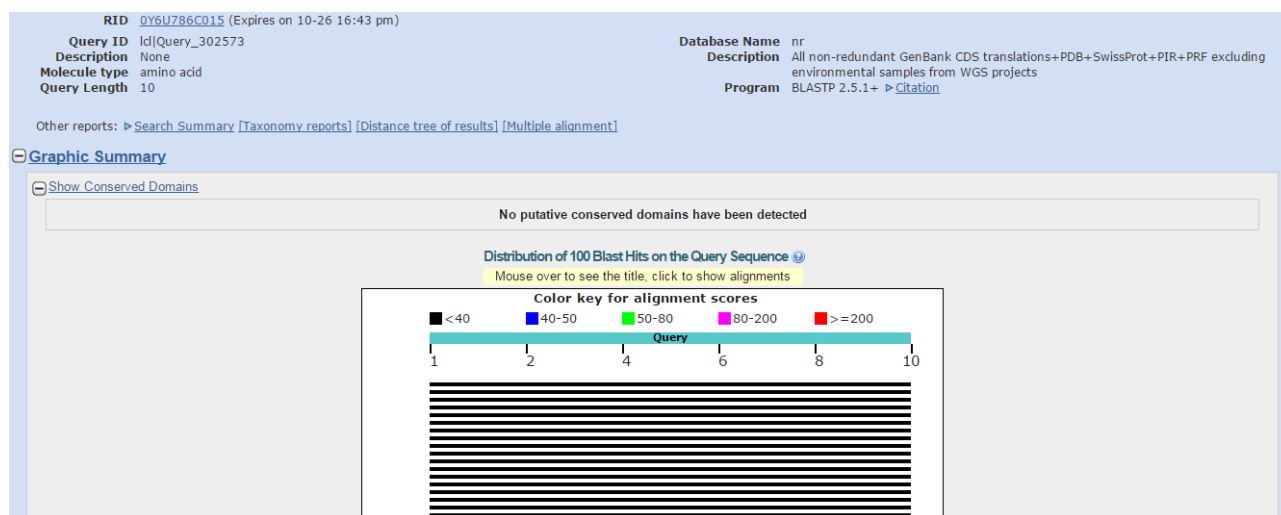


Figura 7a. Rezultatul căutării gradului de potrivire al secvenței de aminoacizi