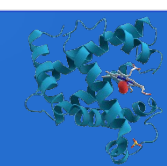


1.04.2026

Curs 7 – Stabilirea funcției unei proteine. Similaritate și omologie la nivel de secvență

Legătura secvență –funcție



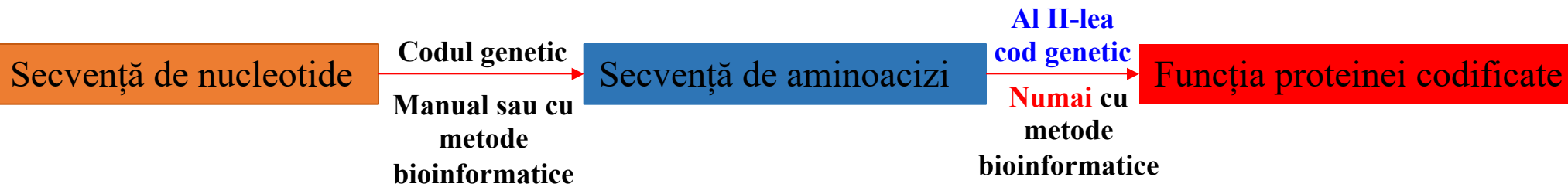
Folosind metodele experimentale de secvențiere enumerate anterior, se stabilește secvența unei gene. Aceasta codifică următoarea secvență de aminoacizi:

MAAKYRIGYFVGSLATGSINRVLSQALINLAPEDLEFSEIPIRDLPLYSYDYDADFPPEGR

Care este funcția acestei peptide și implicit a genei codificatoare?

Cunoașterea secvenței de nucleotide a unui fragment de ADN și implicii a secvenței de aminoacizi a unei proteine nu înseamnă obligatoriu și cunoașterea rolului (funcției) moleculei respective.

Și totuși, secvența de aminoacizi este cea ce coordonează structura tridimensională a peptidei și deci reacția enzimatică/funcția pe care peptidea o are/realizează.



JOURNAL OF BACTERIOLOGY, May 2006, p. 3431–3432
0021-9193/06/\$08.00+0 doi:10.1128/JB.188.10.3431–3432.2006
Copyright © 2006, American Society for Microbiology. All Rights Reserved.

Vol. 188, No. 10

Makrythanasis and Antonarakis *Genome Medicine* 2011, 3:21
<http://genomemedicine.com/content/3/4/21>



GUEST COMMENTARY

The Difficult Road from Sequence to Function

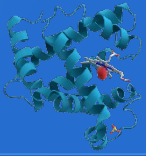
Robert H. White*

Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0308

RESEARCH HIGHLIGHT

From sequence to functional understanding: the difficult road ahead

Periklis Makrythanasis¹ and Stylianos E Antonarakis^{1,2*}



Pentru identificarea computerizată a funcției unei proteine sau gene necunoscute se pleacă de la următoarele **premize**:

1. **toate genele/proteinele au evoluat din alte gene/proteine** prin **mutația** secvenței primare;



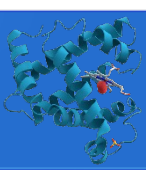
Ce înseamnă și cum funcționează evoluția?

Mutațiile reprezintă modificări spontane nederivate a mesajului genetic. Cel mai frecvent mutațiile apar în procesul de replicarea ADN-ului sau **prin acțiunea factorilor de mediu asupra ADN-ului.** Mutațiile reprezintă materialul de bază pentru variabilitatea și evoluția organismelor vii. **Cum?**

Funcție de **amplarea** lor mutațiile se clasifică în:

- A. Mutații punctiforme
- B. Mutații de amplare mică
- C. Mutații de amplare mare

Mutațiile ca sursă de diversificare a informației genetice



A. Mutațiile punctiforme - mutații ce afectează o singură bază azotată din secvența acizilor nucleici.

Au fost identificate **trei tipuri de mutații punctiforme**:

- **Substituții** – înlocuirea unei baze azotate cu alta
5'ACCGTCTA3' → 5'ACGGTCTA3'
- **Insertii** – adăugarea unei baze azotate suplimentare
5'ACCGTCTA3' → 5'ACCTGTCTA3'
- **Deleții** – pierderea uneia sau a mai multor baze azotate
5'ACCGTCTA3' → 5'AGTCTA3'

Funcție de **efectul** lor asupra produsului codificat de gena în care apar, mutațiile punctiforme se clasifică în:

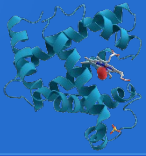
a. **mutație non-sens** – modificare unei baze duce la schimbarea mesajului unui codon în STOP, proteina codificată de genă fiind astfel mai scurtă și nefuncțională;

5' AUGGUC **UAU** CUAGGCGAUUAA 3' → 5' AUGGUC **UAA** CUAGGCGAUUAA 3'
START V T L G D Stop START V STOP

b. **mutație cu sens greșit** – modificare unei baze duce la schimbarea mesajului unui codon și duce la încorporarea unui alt aminoacid în molecula proteică care afectează funcția proteinei codificate

5' AUGGUCUAUCUAGGCGAUUAA 3' → 5' AUGGUCUAUCUAGGCGA**A**UAA 3'
START V T L G D Stop START V T L G **E** Stop

Mutațiile ca sursă de diversificare a informației genetice



c. mutație neutră – modificare unei baze ce duce la schimbarea mesajului unui codon, încorporarea unui aminoacid echivalent în molecula proteică, dar nu modifică funcția proteinei codificate; **Cum se explică acest lucru?**

d. mutație silențioasă – modificare unei baze ce duce la schimbarea mesajului unui codon dar care:

1. are loc într-o zonă ne-tradusă în proteine sau ARN – Ex: introni
2. nu modifică aminoacizii încorporați – **de ce? - mutație sinonimă**

5 ' AUGGUCUAUCUAGGCGAUUAA 3 ' → 5 ' AUGGUCUAUCUAGGAGAUUAA 3 '
START V T L G D Stop START V T L G D Stop

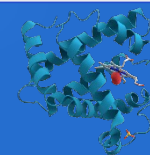
e. mutații cu schimbarea cadrului de lectură (frame-shifts) – inserția sau deleția unei baze azotate ce duce la modificarea modului în care ribosomul citește mesajul genetic de pe molecula de ARNm.

5 ' AUGGUCUAUCUAGGCGAUUAA 3 ' → 5 ' AUGGUCUUCUAGGCGAUUAA 3 '
START V T L G D Stop START V F STOP

B. Mutații de amploare mică – sunt asemănătoare mutațiilor punctiforme d.p.v. al tipurilor și efectelor, dar cuprind câteva baze azotate

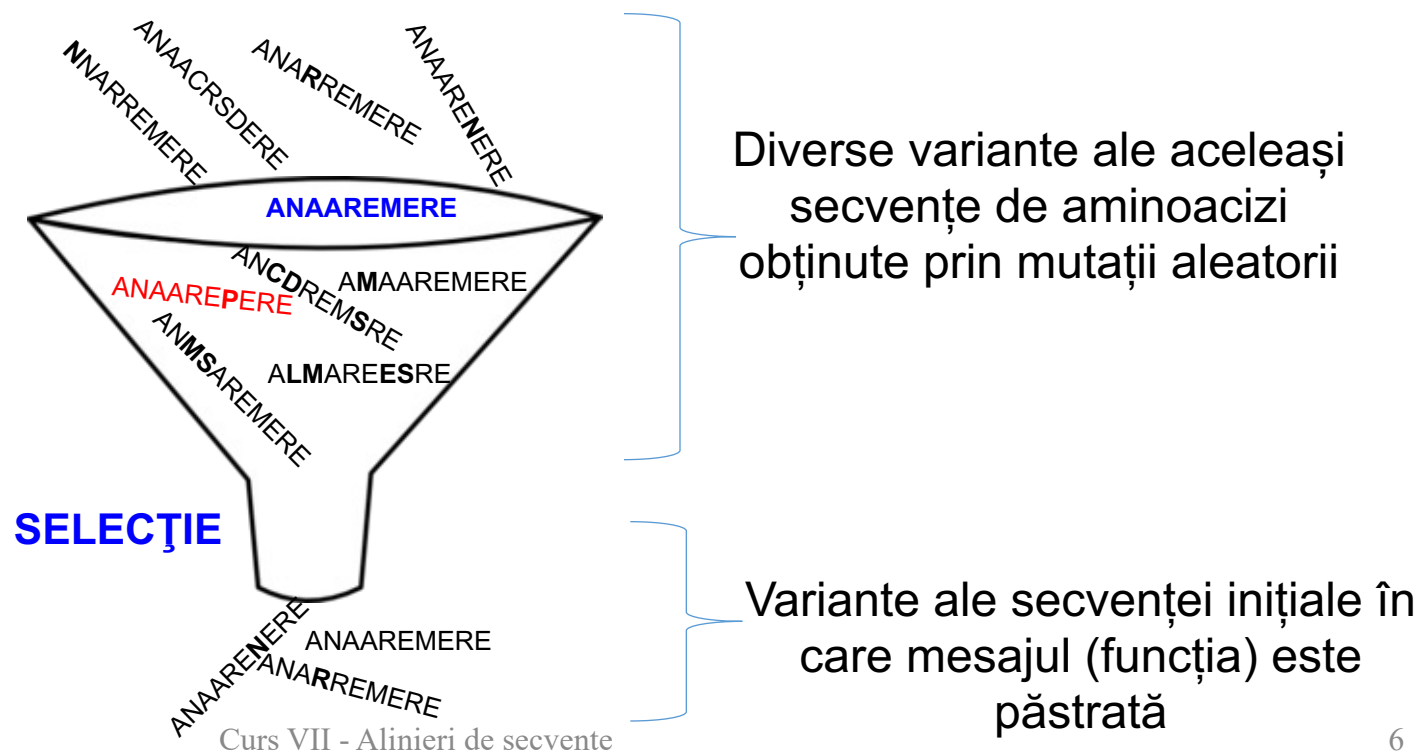
C. Mutații de amploare mare – mutații de dimensiuni mari ce afectează poziția unei gene în cadrul cromozomului și modul de organizarea a materialului genetic – duplicări, inserții de gene, rearanjări cromozomiale.

Legătura secvență –funcție

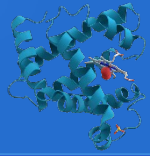


2. Înlocuirea unui aminoacid cu altul într-o proteină nu este întotdeauna aleatoare ci este corelată cu rolul aminoacidului în cadrul proteinei. Din punct de vedere al frecvenței cu care sunt înlocuiți au fost descrise **3 categorii distincte** de aminoacizi într-o secvență proteică:

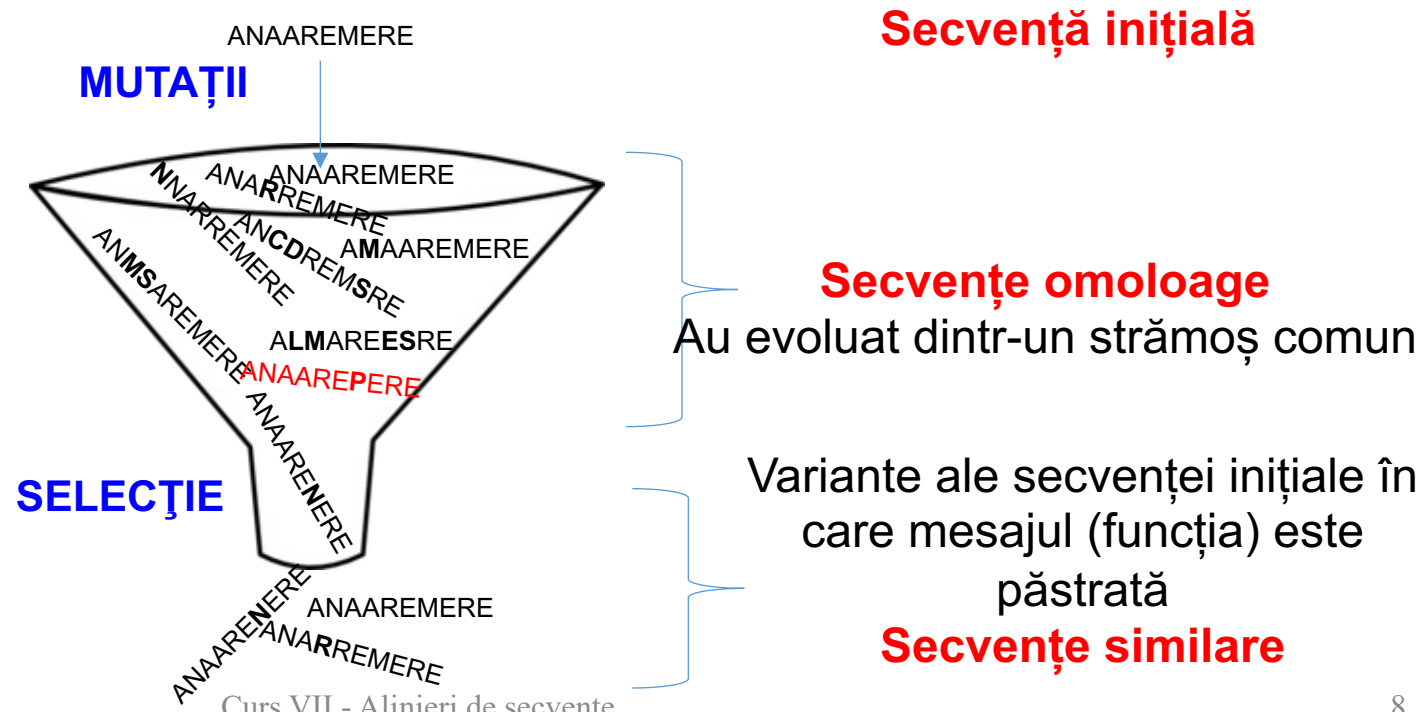
- a) **aminoacizi înalt conservați** - nu sunt înlocuiți decât extrem rar - sunt aminoacizii din **situsul catalitic** sau **funcțional**, implicați în mod direct de realizarea funcției;
- b) **aminoacizi conservați** - sunt înlocuiți destul de rar - sunt aminoacizii implicați în realizarea structurilor secundare și terțiare;
- c) **aminoacizi puțin conservați** - sunt înlocuiți frecvent - sunt în general aminoacizii de pe suprafața proteinelor, înlocuirea lor nu modifică semnificativ funcția proteinei.

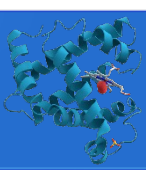


Alinieri de secvențe



- b. **identitate** - la nivel de secvență - se referă la două secvențe care prezintă asemănări una în raport cu cealaltă datorită unui număr mai mare sau mai mic de **aminoacizi identici**; similaritatea se exprimă ca **procente identitate** - % de aminoacizi ce sunt identici între 2 secvențe;
- c. **similaritate** la nivel de secvență - se referă la două secvențe care prezintă asemănări una în raport cu cealaltă datorită unui număr mai mare sau mai mic de **aminoacizi identici** dar ia în calcul și **semnificația substituțiilor dintre aminoacizi** ;
- c. **omologie** - se referă la faptul că două secvențe se aseamănă una cu cealaltă deoarece au evoluat dintr-un strămoș comun, dar nu au obligatoriu aceeași funcție;
- d. **secvența de aminoacizi identici înalt conservați** dintr-o aliniere a două sau mai multe secvențe se numește **secvență consens (consensus)**;





Gradul de similaritate a două proteine la nivel de secvență este dictat, pe de o parte, de **numărul de mutații ce le diferențiază** (distanța evolutivă) și, pe de altă parte, de **structurile lor tridimensionale și de funcțiile specifice** pe care cele două proteine le îndeplinesc.

Două secvențe de nucleotide similare vor codifica un mesaj genetic similar și deci vor avea funcții similare.

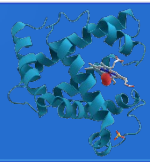
Două proteine similare vor avea structuri similare și deci funcții similare.

Întrebarea inițială: `Care este funcția acestei peptide și implicit a genei codificatoare?`

MAAKYRIGYFVGSLATGSINRVLSQALINLAPEDLEFSEIPIRDLPLYSYDYDADFPPEGR

devine: **cu ce peptidă cunoscută este similară această secvență?**

BLAST – identificarea de secvențe similare



BLAST - Basic Local Alignment Search Tool

1. **identifică**, dintr-o bază de date, **secvențele similare cu o secvență țintă (tinta analizei, experimentului)**. Aceste secvențe identificate poartă numele de **secvențe “subiect”**, iar identificarea lor se bazează pe **alinieri locale**.

Secvența „subiect este „suprapusă” peste cea țintă la nivelul alinierilor locale astfel încât secvențele comparate vor fi alcătuite din zone perfect aliniate și zone nealiniat (așa numitele **GAP’s**) care formează bucle între o aliniere locală și următoarea aliniere locală.

2. **cuantifică nivelul de similaritate** dintre secvențele “subiect” și secvența țintă prin utilizarea unor **matrici de substituție**. O matrice de substituție arată frecvența cu care un aminoacid este înlocuit cu altul și are la bază observațiile experimentale.

298

Biologie moleculară. Metode experimentale

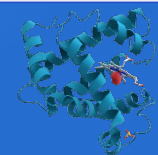
| | A | C | D | E | F | G | H |
|---|----|----|----|----|----|----|----|
| A | 4 | 0 | -2 | -1 | -2 | 0 | -2 |
| C | 0 | 9 | -3 | -4 | -2 | -3 | -3 |
| D | -2 | -3 | 6 | 2 | -3 | -1 | -1 |
| E | -1 | -4 | 2 | 5 | -3 | -2 | 0 |
| F | -2 | -2 | -3 | -3 | 6 | -3 | -1 |
| G | 0 | -3 | -1 | -2 | -3 | 3 | -1 |
| H | -2 | -3 | -1 | 0 | -1 | -1 | 3 |

BLOSUM 62

FIGURA 35. Exemplu de matrice BLOSUM.

Cu litere mari sunt reprezentați aminoacizii. Cel mai mare punctaj au identitățile, iar funcție de frecvența de substituție (observată practic în laborator) primesc un anumit punctaj și substituțiile. Se poate observa și existența unor punctaje negative, alocate pentru substituțiile cel mai puțin întâlnite.

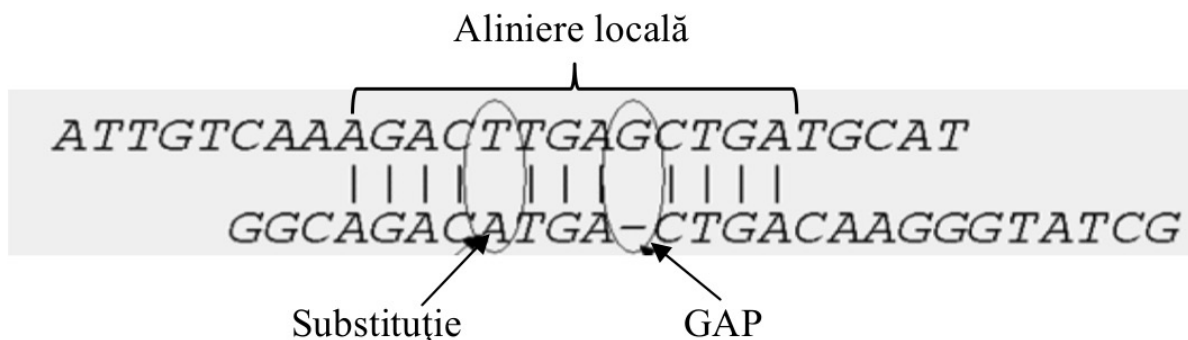
BLAST – identificarea de secvențe similare



3. Calculează un **scor de similaritate** prin însumarea punctelor pentru fiecare pereche aminoacid-aminoacid și **ierarhizează secvențele** țintă funcție de valoarea acestui scor.

Scoruri de similaritate calculate de BLAST:

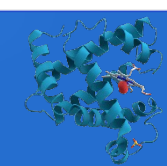
- **punctaj brut** (engl. **Raw score**) notat cu S , este calculat prin însumarea punctelor pentru fiecare pereche aminoacid-aminoacid, aminoacid-nimic și penalizărilor pentru GAP; nu permite ierarhizarea secvențelor, valoare lui depinde de lungimea secvențelor analizate;



$$S = \sum_{(\text{identități, substituții})} - \sum_{(\text{penalizări GAP})}$$

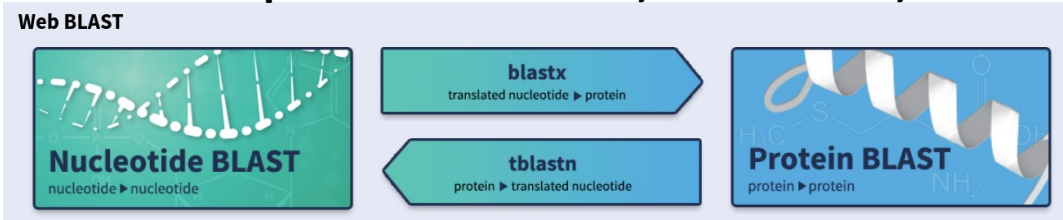
- **scorul în biți notat cu S'** - se calculează prin normalizarea lui S în funcție de diverse variabile statistice care depind, la rândul lor, de tipul de matrice utilizat. **Cu cât punctajul S' obținut este mai mare cu atât asemănarea dintre cele două secvențe este mai mare;**
- **parametru statistic E** - care se definește ca număr de potriviri care apar doar datorită șansei într-o bază de date de o anumită dimensiune. **Cu cât valorile lui E sunt mai mici, cu atât rezultatele sunt considerate ca având un înalt grad de semnificație (alinierea fiind deci statistic semnificativă).**

Cum se realizează o analiză BLAST?



1. Accesează: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

2. Selectează tipul de analiză funcție de secvența de interes:



utilizate

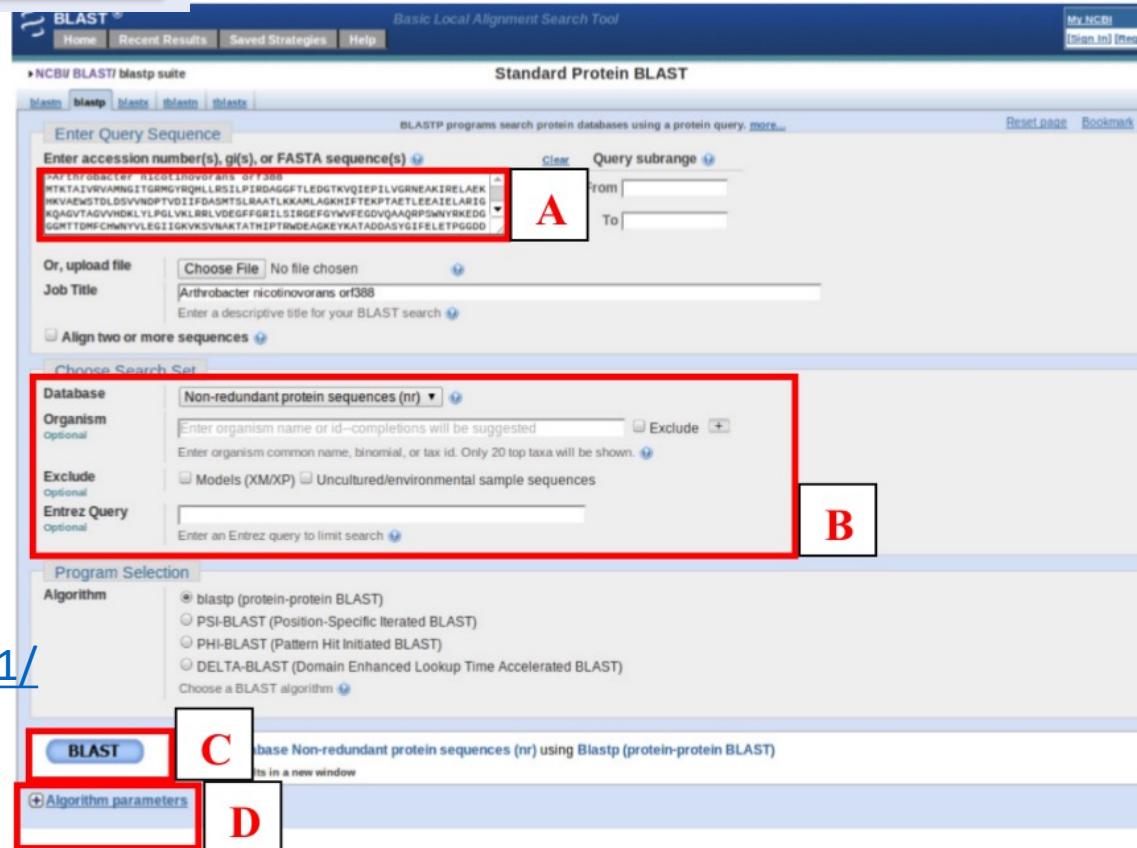
3. Copie secvența în căsuța pentru secvența țintă (**query**), setează parametrii căutării și apasă BLAST

A – căsuța text în care a fost inserată secvența țintă în format FASTA;

B – zona cu parametrii utilizați pentru restrângerea spațiului de căutare;

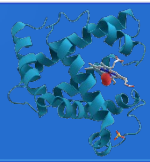
C – buton pentru lansarea investigației;

D – zona cu parametrii algoritmului de căutare



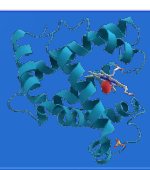
<http://www.ncbi.nlm.nih.gov/books/NBK21101/>

Parametri pentru restrângerea spațiului de căutare



- „**Query subrange**” – se utilizează pentru a reduce secvența investigată doar la un anumit fragment precizat prin poziția de început și de sfârșit;
- „**Or, upload file**” – permite încărcarea unui fișier text în format FASTA; în acest caz nu mai este necesară precizarea secvenței în căsuța text;
- „**Database**” – permite selecția bazei de date unde se va realiza investigarea; funcția este utilă pentru a restrânge investigarea în diverse direcții; cea mai mare bază de date utilizabilă este „non-redundant protein sequences (nr)”; cea mai mică este „Protein Data Bank proteins (pdb)” care are avantajul de a fi alcătuită doar din proteine cu structură determinată experimental și funcție cunoscută;
- „**Algorithm**” – permite selectarea diverșilor algoritmi de identificare și aliniere a secvențelor similare; de asemenea, tot în această pagină, apăsând semnul „+” din fața textului „Algorithm parameters”, se pot modifica o serie de parametri ai algoritmului de căutare, precum:
 - „**Max target sequences**” – numărul maxim de secvențe subiect ce va fi luat în calcul;
 - „**Expect**” – vor fi afișate doar rezultatele care au o valoare a lui E mai mică decât cea specificată;
 - „**Matrix**” – tipul de matrice care va fi utilizat pentru calcularea punctajului similarității (BLOSUM62, PAM70 etc.)
 - „**Gap Costs**” – schema de punctaj corespunzătoare prezenței unui GAP.

3. Rezultate BLAST și semnificația lor



```

LOCUS       YP_002488236             389 aa             linear   BCT 27-JAN-2012
DEFINITION   oxidoreductase [Arthrobacter chlorophenolicus A6].
ACCESSION   YP_002488236
VERSION     YP_002488236.1   GI:220912927
DBLINK     Project: 58969
DBSOURCE    REFSEQ: accession NC\_011886.1
KEYWORDS    .
SOURCE     Arthrobacter chlorophenolicus A6
  ORGANISM  Arthrobacter chlorophenolicus A6
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Micrococcineae; Micrococcaceae; Arthrobacter.
REFERENCE   1 (residues 1 to 389)
  AUTHORS   Lucas,S., Copeland,A., Lapidus,A., Glavina del Rio,T., Tice,H.,
            Bruce,D., Goodwin,L., Pitluck,S., Goltsman,E., Clum,A., Larimer,F.,
            Land,M., Hauser,L., Kyrpides,N., Mikhailova,N., Jansson,J. and
            Richardson,P.
  CONSRTM   US DOE Joint Genome Institute
  TITLE     Complete sequence of chromosome of Arthrobacter chlorophenolicus A6
  JOURNAL   Unpublished
REFERENCE   2 (residues 1 to 389)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2009) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
    
```

A

Secvență nepublicată, funcție teoretică

```

LOCUS       YP_002541450             385 aa             linear   BCT 01-DEC-2011
DEFINITION   oxidoreductase [Agrobacterium radiobacter K84].
ACCESSION   YP_002541450
VERSION     YP_002541450.1   GI:222082085
DBLINK     Project: 58269
DBSOURCE    REFSEQ: accession NC\_011983.1
KEYWORDS    .
SOURCE     Agrobacterium radiobacter K84
  ORGANISM  Agrobacterium radiobacter K84
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
REFERENCE   1 (residues 1 to 385)
  AUTHORS   Slater,S.C., Goldman,B.S., Goodner,B., Setubal,J.C., Farrand,S.K.,
            Nester,E.W., Burr,T.J., Banta,L., Dickerman,A.W., Paulsen,I.,
            Otten,L., Suen,G., Welch,R., Almeida,N.F., Arnold,F., Burton,O.T.,
            Du,Z., Ewing,A., Godsy,E., Heisel,S., Houmiel,K.L., Jhaveri,J.,
            Lu,J., Miller,N.M., Norton,S., Chen,Q., Phoolcharoen,W., Ohlin,V.,
            Ondrusek,D., Pride,N., Stricklin,S.L., Sun,J., Wheeler,C.,
            Wilson,L., Zhu,H. and Wood,D.W.
  TITLE     Genome sequences of three agrobacterium biovars help elucidate the
            evolution of multichromosome genomes in bacteria
  JOURNAL   J. Bacteriol. 191 (8), 2501-2511 (2009)
  PUBMED    19251847
    
```

B

Secvență publicată, funcție teoretică

```

LOCUS       2GLX_A                   332 aa             linear   BCT 24-SEP-2008
DEFINITION   Chain A, Crystal Structure Analysis Of Bacterial 1,5-Af Reductase.
ACCESSION   2GLX_A
VERSION     2GLX_A   GI:114794099
DBSOURCE    pdb: molecule 2GLX, chain 65, release Aug 27, 2007;
            deposition: Apr 5, 2006;
            class: Oxidoreductase;
            source: Mol_id: 1; Organism_scientific: Ensifer Adhaerens;
            Organism_common: Bacteria; Strain: S-30.7.5; Gene: Afr;
            Expression_system: Escherichia Coli; Expression_system_common:
            Bacteria; Expression_system_strain: Bl21(De3);
            Expression_system_vector_type: Plasmid; Expression_system_plasmid:
            Pet24a;
            Exp. method: X-Ray Diffraction.
KEYWORDS    .
SOURCE     Ensifer adhaerens (Sinorhizobium morelense)
  ORGANISM  Ensifer adhaerens
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Rhizobiaceae; Sinorhizobium/Ensifer group; Ensifer.
REFERENCE   1 (residues 1 to 332)
  AUTHORS   Kuhn,A., Yu,S. and Giffhorn,F.
  TITLE     Catabolism of 1,5-anhydro-D-fructose in Sinorhizobium morelense
            S-30.7.5: discovery, characterization, and overexpression of a new
            1,5-anhydro-D-fructose reductase and its application in sugar
            analysis and rare sugar synthesis
  JOURNAL   Appl. Environ. Microbiol. 72 (2), 1248-1257 (2006)
  PUBMED    16461673
    
```

C

Secvență publicată, funcție demonstrată experimental