

BIOINFORMATICĂ APLICATĂ ÎN BIOLOGIA STRUCTURALĂ

Seminar X

Predicția proprietăților proteinelor pe
baza secvenței de aminoacizi

Care este Mw, pI, coeficientul de extincție și stabilitatea în *E.coli* a proteinei codificate de gena de mai jos:

????

????

????

>AJ507836.1:c60415-59915 *Arthrobacter nicotinovorans* pAO1 megaplasmid
sequence, strain ATCC 49919

```
ATGAGTAGCTATGCACGACGAACTCCAGTCAGGACGGTCTTGAACTTTTGCACAGCGATTCGGCAGATCA
TCACAGGAGAGGCAGGAAGCGTGGCGGCTGACAAAGGTAACAGGGGGCAGAAACGTGCGCCGTTACTTCA
TAGACATCGCCTTCATGCTTGGTTGTATCTTGGAAAGTGCCATAACCACCGAGGTGACTGGCACAGTAATT
CTCGACTTTTTCCGAAGGCTTTCAACTCCCGGCACAGACAACCGCGGCCATGGCACTATAACGCCTTTTCCT
TCTTTCTGCTAACCCGCGCGCTTAGGGCAGTACCACTAAGCGTCGCTTATGCTACGTGGTCTGGTCTAGG
CACAGTAGCCGTTGCCTTTGCTGGCGCGATTATCCACGGAGAGGCAGTCACTTTGGGAAGGATAACAGCC
ATTACTGCAGTGATCGGGGGCATCGTTATATTGAACCTTGCCACCACGCGCCAACACTCGGCACGACGCA
AAGACGTCTAA
```

Deschideți fișierul fasta corespunzător secvenței afișateBABS/nepB_dna.fasta

1. Traducerea mesajului în secvență de aminoacizi

Folosiți SMS2 pentru a traduce secvența de ADN a genei în secvența de aminoacizi a catenei codificate:

```
>rf 1 Untitled
MSSYARRTPVRTVLNFC TAI RQ I I TGEAGSVAADKGNRGQKRAPLLHRHRLHAWLYLGSA
ITTEVTGTVILDFSEGFQLPAQTTAAMALYAFSFFLLTRALRAVPLSVAYATWSGLGTVA
VAFAGAI IHGEAVTLGRITAITAVIGGIVILNLATTRQHSARRKDV*
```

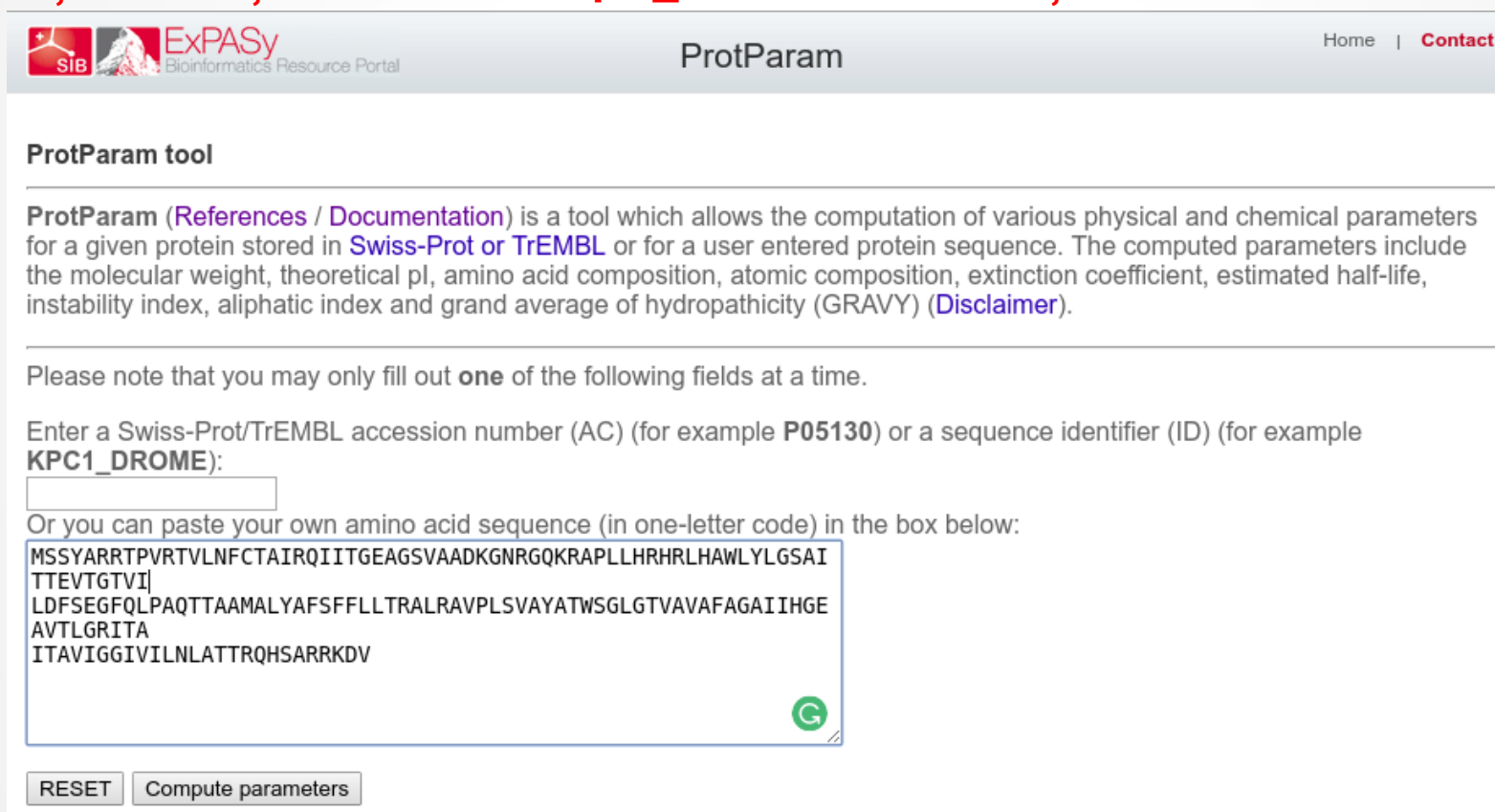
Copiați secvența obținută și salvati-o într-un fișier fasta intitulat nepB_aac.fasta

2. Calcularea Mw și pl

2a. Accesați <https://web.expasy.org> sau (sau Google și căutați expasy server)

2.b. Dați click pe secțiunea Proteomics (stânga sus) și identificați programul ProtParam; accesați link-ul corespunzător;

2.c. copiați secvența de aac din nepB_aac.fasta în căsuța text din fereastra deschisă:



The screenshot shows the ProtParam tool interface. At the top, there is a header with the Expasy logo (SIB Bioinformatics Resource Portal) and the title "ProtParam". Navigation links for "Home" and "Contact" are visible. Below the header, the tool is titled "ProtParam tool". A descriptive paragraph explains that ProtParam is a tool for computing various physical and chemical parameters for a protein, such as molecular weight, theoretical pI, amino acid composition, and GRAVY. Below this, a note states that only one of the input fields should be filled at a time. There are two input options: a text box for a Swiss-Prot/TrEMBL accession number (AC) or a sequence identifier (ID), and a larger text area for pasting a custom amino acid sequence. The sequence provided in the text area is: MSSYARRTPVRTVLNFCFAIRQIITGEAGSVAADKGNRGQKRAPLLHRHRLHAWLYLGSATTEVTGTVI|LDFSEGFQLPAQTAAAMLYAFSFFLLTRALRAVPLSVAYATWSGLGTVAVAFAGAIIHGEAVTLGRITAITAVIGGIVILNLATTRQHSARRKDV. At the bottom, there are two buttons: "RESET" and "Compute parameters".

ProtParam tool

ProtParam ([References](#) / [Documentation](#)) is a tool which allows the computation of various physical and chemical parameters for a given protein stored in [Swiss-Prot](#) or [TrEMBL](#) or for a user entered protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) ([Disclaimer](#)).

Please note that you may only fill out **one** of the following fields at a time.

Enter a Swiss-Prot/TrEMBL accession number (AC) (for example **P05130**) or a sequence identifier (ID) (for example **KPC1_DROME**):

Or you can paste your own amino acid sequence (in one-letter code) in the box below:

```
MSSYARRTPVRTVLNFCFAIRQIITGEAGSVAADKGNRGQKRAPLLHRHRLHAWLYLGSATTEVTGTVI|LDFSEGFQLPAQTAAAMLYAFSFFLLTRALRAVPLSVAYATWSGLGTVAVAFAGAIIHGEAVTLGRITAITAVIGGIVILNLATTRQHSARRKDV
```

2.d. apăsați butonul **Compute parameters**.

2. Calcularea Mw și pI

2.e. identificați în fereastra deschisă informațiile de interes:

Number of amino acids: 166

Molecular weight: 17749.64

Theoretical pI: 11.09

Extinction coefficients:

Extinction coefficients are in units of $M^{-1} \text{ cm}^{-1}$, at 280 nm measured in water.

Ext. coefficient 16960
Abs 0.1% (=1 g/l) 0.956, assuming all pairs of Cys residues form cystines

Ext. coefficient 16960
Abs 0.1% (=1 g/l) 0.956, assuming all Cys residues are reduced

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index:

The instability index (II) is computed to be 35.95
This classifies the protein as stable.

Aliphatic index: 106.51

Grand average of hydropathicity (GRAVY): 0.385

Ce alte informații sunt calculate de ProtParam?

Codifică gena indicată anterior o proteină transmembranară?

- 3a. Accesați <http://www.cbs.dtu.dk/services/TMHMM/> (sau Google și căutați TMHMM server)
3b. copiați secvența de aac din nepB_aac.fasta în căsuța text din fereastra deschisă:

DTU Bioinformatics
Department of Bio and Health Informatics
[Home](#)

TMHMM Server v. 2.0

Prediction of transmembrane helices in proteins



[Instructions](#)

SUBMISSION

Submission of a local file in **FASTA** format (HTML 3.0 or higher)

No file chosen

OR by pasting sequence(s) in **FASTA** format:

```
MSSYARRTPVRTVLNFCTAIRQIITGEAGSVAADKGNRGQKRAPLLHRHRLHAWLYLGSAITTEVTGTVI
LDFSEGFQLPAQTAAALYAFSFFLLTRALRAVPLSVAYATWSGLGTVAVAFAGAIHGEAVTLGRITA
ITAVIGGIVILNLATTRQHSARRKDV
```

Output format:

Extensive, with graphics
 Extensive, no graphics
 One line per protein

Other options:

Use old model (version 1)

3.c. apăsați butonul **Submit**.

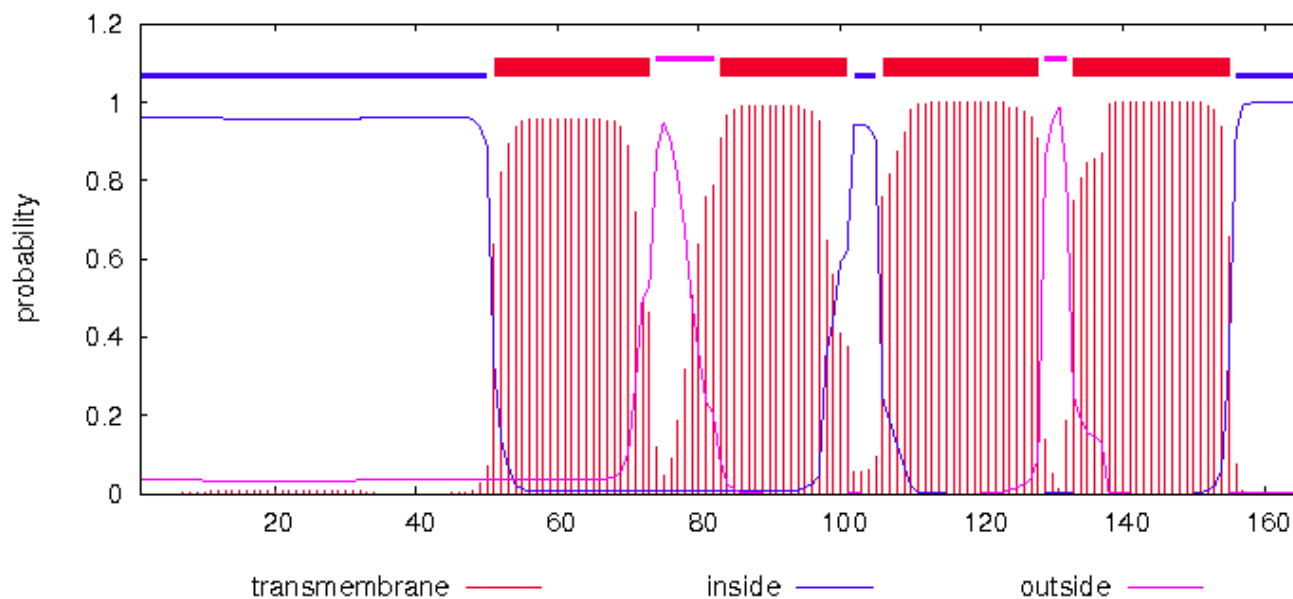
Codifică gena indicată anterior o proteină transmembranară?

TMHMM result

[HELP](#) with output formats

```
# WEBSEQUENCE Length: 166
# WEBSEQUENCE Number of predicted TMHs: 4
# WEBSEQUENCE Exp number of AAs in TMHs: 85.19899
# WEBSEQUENCE Exp number, first 60 AAs: 9.33735
# WEBSEQUENCE Total prob of N-in: 0.96225
WEBSEQUENCE TMHMM2.0 inside 1 50
WEBSEQUENCE TMHMM2.0 TMhelix 51 73
WEBSEQUENCE TMHMM2.0 outside 74 82
WEBSEQUENCE TMHMM2.0 TMhelix 83 101
WEBSEQUENCE TMHMM2.0 inside 102 105
WEBSEQUENCE TMHMM2.0 TMhelix 106 128
WEBSEQUENCE TMHMM2.0 outside 129 132
WEBSEQUENCE TMHMM2.0 TMhelix 133 155
WEBSEQUENCE TMHMM2.0 inside 156 166
```

TMHMM posterior probabilities for WEBSEQUENCE



Da, proteina conține 4 helix-uri transmembranare conectate prin 2 bucle în afara celulei și 1 buclă în interiorul celulei.

Exercițiu de introdus în prezentare

1. Deschideți fișierul fasta alocat cu un editor de text pentru a vizualiza conținutul;
2. Folosiți secvența pentru a răspunde la următoarele întrebări:
 - a. Care este masa moleculară a proteinei dmV?
 - b. La ce pH proteina dmV nu este încărcată electric?
 - c. Care este cea mai bună gazdă pentru a exprima proteina alocată?
 - d. Este o proteină globulară sau trans-membranară?
 - e. Dacă este o proteină transmembranară, câte helix-uri transmembranare conține?

Atașați figuri care să susțină răspunsurile dmV. și scrieți câte o scurtă legendă pentru fiecare figură introdusă.