

# BIOINFORMATICĂ APLICATĂ ÎN BIOLOGIA STRUCTURALĂ

## Seminar VIII

Operații simple cu secvențe.  
Clasificarea proteinelor folosind CATH

# Din ce familie face parte proteina cu secvența

>NM\_000558.4\_aac

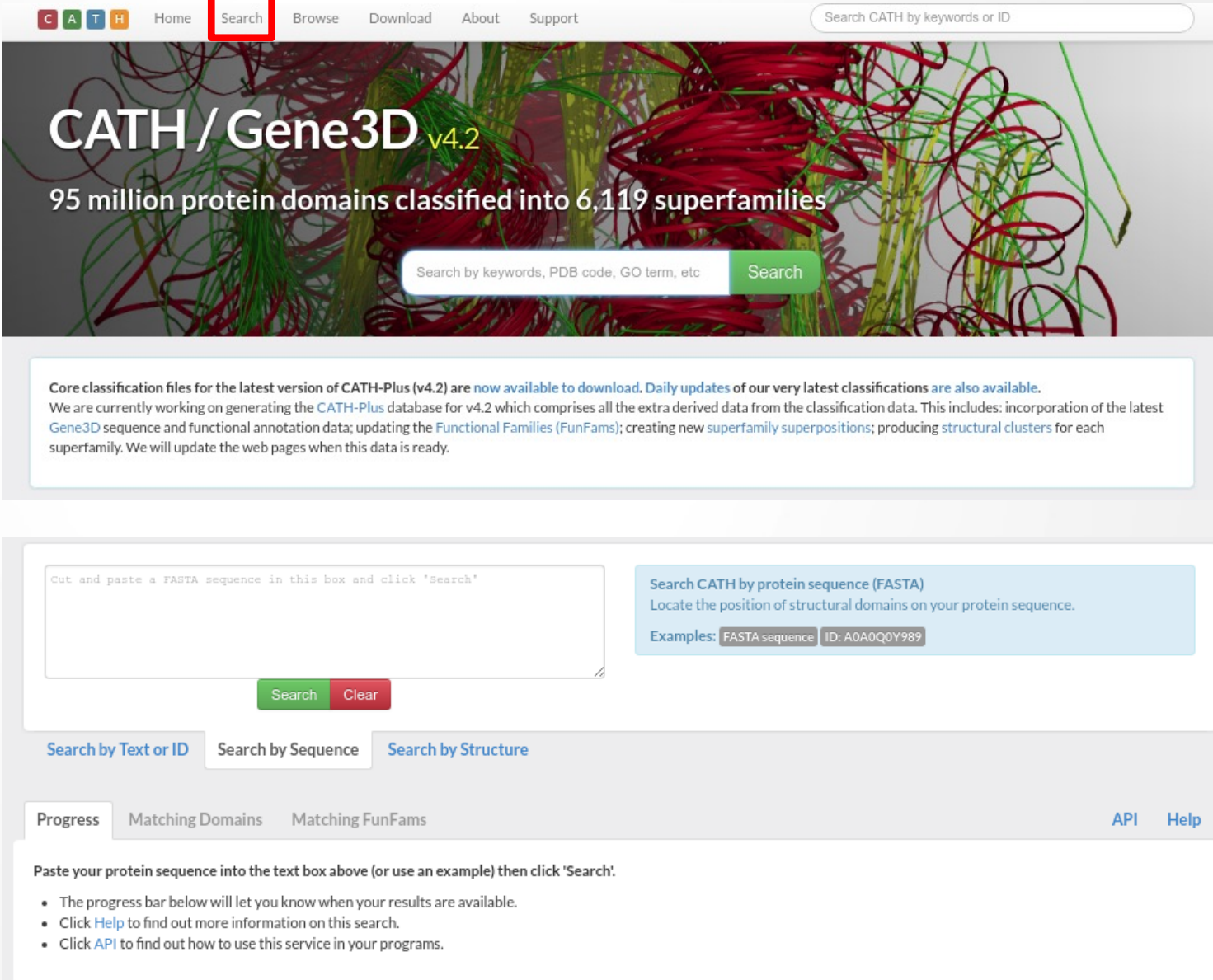
MVLSPADKTNVKAAWGKVGGAHAGEYGAEALERMFLSFPTTKTYFPHFDSLHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALS  
DLHAHKLKRVDPVNFKLLSHCLLVTLAAHLPAEFTP AVHASLTKFLASVSTVLTISKYR\*

Deschideți fișierul word salvat în seminarul anterior (NM\_000558.doc), identificați secvența de aminoacizi și copiați-o.

# Clasificarea unei proteine pornind de la secvență

1. Accesați <http://www.cathdb.info/>

2. Accesați funcția Search



The screenshot shows the CATH / Gene3D v4.2 website. The header includes navigation links: Home, Search (highlighted with a red box), Browse, Download, About, and Support. A search bar is located in the top right corner with the placeholder text "Search CATH by keywords or ID". The main banner features the text "CATH / Gene3D v4.2" and "95 million protein domains classified into 6,119 superfamilies". Below the banner is a search bar with the placeholder text "Search by keywords, PDB code, GO term, etc" and a green "Search" button. A text box below the banner contains information about the latest version of CATH-Plus (v4.2) and its updates. The search interface includes a text box for pasting a FASTA sequence, a "Search" button, and a "Clear" button. Below the text box are three tabs: "Search by Text or ID", "Search by Sequence" (selected), and "Search by Structure". A blue box on the right provides instructions for searching by protein sequence (FASTA) and includes examples: "FASTA sequence" and "ID: A0A0Q0Y989". At the bottom, there are tabs for "Progress", "Matching Domains", and "Matching FunFams", along with "API" and "Help" links. A text box at the bottom provides instructions for using the search service and lists three bullet points: "The progress bar below will let you know when your results are available.", "Click Help to find out more information on this search.", and "Click API to find out how to use this service in your programs."

3. Copiați secvența în căsuța de căutare

# Clasificarea unei proteine pornind de la secvență

C A T H Home Search Browse Download About Support

## Search CATH

```
>QUERY
MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMP
NALSALSDLHAHKLRLVDPVNFKLLSHCLLVTLAAHLPAEFTF AVHASLDFKFLASVSTVLTSKYR
```

Search Clear

Search CATH by protein sequence (FASTA)  
Locate the position of structural domains on your protein sequence.  
Examples: `FASTA sequence` ID: `A0A0Q0Y989`

Search by Text or ID Search by Sequence Search by Structure

Progress **401 Matching Domains** **3 Matching FunFams** [API](#) [Help](#)

Your search has been submitted to a queue and your results should be available shortly (< 1 minute).  
Searching for CATH structural domains in your protein...

Waiting Queued Running Done

Found 401 matches

Searching for CATH Functional Families in your protein...

Waiting Queued Running Done

Found 3 matches

**Proteine cu care proteina are domenii similare**

**Familii din care face parte proteina de interes**

# Rezultatele interogării CATH

**Pagina cu domenii** proteice – domeniile ce au membri similari la nivel de secvență cu proteina țintă sunt prezentate ierarhic, funcție de semnificația scorului de omologie.

CATH Home Search Browse Download About Support

## Search CATH

>QUERY  
MVLSPADKTNVKAAMGKVGVAHAGEYGAELERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMP  
NALSALSDDLHAHKLKRVDPVNFKLLSHCLLVTLAAHLPAEPTF AVHASLDRKFLASVSTVLTSKYR

Search Clear

Search CATH by protein sequence (FASTA)  
Locate the position of structural domains on your protein sequence.  
Examples: FASTA sequence ID: A0A0Q0Y989

Search by Text or ID Search by Sequence Search by Structure

Progress 401 Matching Domains 3 Matching FunFams API Help

Match	Matching Regions	Evalue
1irdA00		7.3e-78
1c7dA02		7.3e-78
4it1B01		5.5e-1
4wchD00		9.4e-1
1xp3A00		9.4e-1
1b0bA00		9.4e-1
3u7bA00		2.1e+0
2aeeB00		2.1e+0
1d8hA00		3.6e+0

# Rezultatele interogării CATH

## Domeniu cu scor mare de omologie

### CATH Classification

Level	CATH Code	Description
1	1	Mainly Alpha
2	1.10	Orthogonal Bundle
3	1.10.490	Globin-like
4	1.10.490.10	Globins

### CATH Clusters

Superfamily	Globins
Functional Family	Hemoglobin beta chain subunit

### Enzyme Information

### UniProtKB Entries (1)

[P68871](#) **HBB\_HUMAN**  
*Homo sapiens*  
 Hemoglobin subunit beta

### PDB Structure



PDB 1IRD

External Links  

- PDBSum
- Proteopedia

Method X-RAY DIFFRACTION

Organism

Primary Citation  
 Carbonmonoxy-Haemoglobin at 1.25 Å Resolution  
 Park, S.-Y., Tame, J.R.H.  
*To be Published*

## Domeniu cu scor mic de omologie

### CATH Classification

Level	CATH Code	Description
3	3	Alpha Beta
4	3.20	Alpha-Beta Barrel
5	3.20.100	mRNA Triphosphatase Cet1; Chain A
6	3.20.100.10	mRNA triphosphatase Cet1-like

### CATH Clusters

Superfamily	mRNA triphosphatase Cet1-like
Functional Family	mRNA-capping enzyme subunit beta

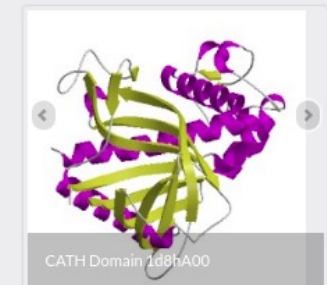
### Enzyme Information

**3.1.3.33 Polynucleotide 5'-phosphatase.**  
*based on mapping to UniProt O13297*  
 A 5'-phosphopolynucleotide + H(2)O = a polynucleotide + phosphate.  
 -|- Does not act on nucleoside monophosphates. -|- Induced in *Escherichia coli* by T-even phages.

### UniProtKB Entries (1)

[O13297](#) **CET1\_YEAST**  
*Saccharomyces cerevisiae* S288C  
 MRNA-capping enzyme subunit beta

### PDB Structure



PDB 1D8H

External Links  

- PDBSum
- Proteopedia

Method X-RAY DIFFRACTION

Organism *Escherichia*

Primary Citation  
 Structure and mechanism of yeast RNA triphosphatase: an essential component of the mRNA capping apparatus.  
 Lima, C.D., Wang, L.K., Shuman, S.  
*Cell(Cambridge,Mass.)*

# Rezultatele interogării CATH

**Pagina cu familii** proteice – familiile ce au membri similari la nivel de secvență cu proteina țintă sunt prezentate ierarhic, funcție de semnificația scorului de omologie.

The screenshot displays the CATH database search interface. At the top, there are navigation links: Home, Search, Browse, Download, About, and Support. A search bar on the right contains the text "Search CATH by keywords or ID". Below this, the "Search CATH" section features a text input field with a protein sequence: `>QUERY  
MVLSPADKTNVKAAWGKVGSAHAGEYGAELERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMP  
NALSALSIDLRAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTF AVHASLKDPLASVSTVLTSKYR`. Below the input field are "Search" and "Clear" buttons. To the right, a blue box provides instructions: "Search CATH by protein sequence (FASTA) Locate the position of structural domains on your protein sequence. Examples: FASTA sequence ID: A0A0Q0Y989". Below the search area, there are three tabs: "Search by Text or ID", "Search by Sequence" (which is active), and "Search by Structure". At the bottom, there are two progress indicators: "401 Matching Domains" and "3 Matching FunFams". On the far right, there are links for "API" and "Help". The search results are presented in a table with the following columns: Match, 3D, EC, Matching Regions, and Evalue. The first result, "Hemoglobin beta chain subunit (1.10.490.10/FF/5159)", is highlighted with a red box. The second result is "Mutant hemoglobin beta chain (1.10.490.10/FF/4588)" and the third is "Globin 1, isoform A (1.10.490.10/FF/5170)".

Match	3D	EC	Matching Regions	Evalue
<b>Hemoglobin beta chain subunit (1.10.490.10/FF/5159)</b>	3D	EC		2.5e-50
Mutant hemoglobin beta chain (1.10.490.10/FF/4588)				1.7e-17
Globin 1, isoform A (1.10.490.10/FF/5170)	3D			2.8e-12

# Rezultatele interogării CATH

Denumirea familiei din care face parte proteina țintă

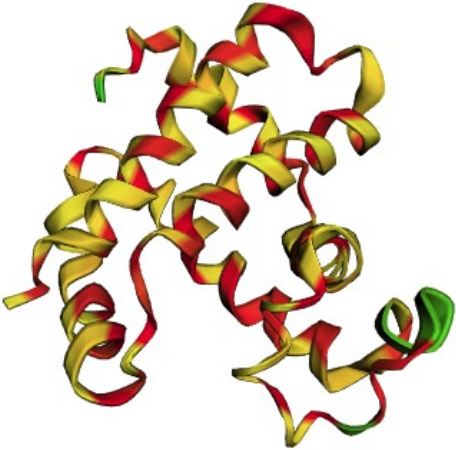
Structura 3D a domeniului comun tuturor membrilor acestei familii proteice

Alinierea secvenței tuturor membrilor acestei familii proteice în zona domeniului identificat.

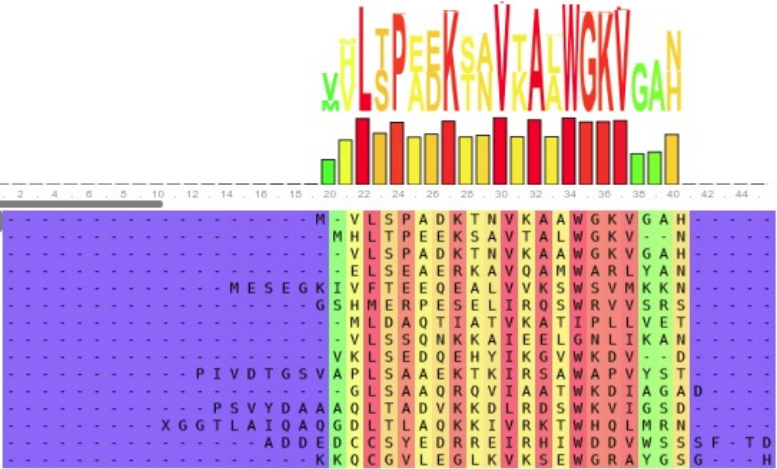
FunFam 5159: Hemoglobin beta chain subunit

Summary Alignment GO EC Taxonomy

Representative CATH Domain 1a00B00



Sequence alignment 200 sequences



More information on the open source software used on this web site can be found in the [Tools](#) page.

**Functional Families**

Overview of the Structural Clusters (SC) and Functional Families within this CATH Superfamily. Clusters with a representative structure are represented by a filled circle.

- SC:1
- Diguanylate cyclase Dc
- Group III hemoglobin
- Neuroglobin
- Hemoglobin-like, oxyge
- Body wall hemoglobin
- Group 1 hemoglobin gl
- Heme-based aerotactic
- Hemoglobin beta cha**
- Group 1 truncated hem
- Globin 1, Isoform A
- Globin-like protein 9
- Methyl-accepting chem
- Methyl-accepting chem
- Globin
- Globin-like protein 26
- Group 1 truncated hem
- Neuroglobin
- Mutant hemoglobin bet
- Neuroglobin
- GLB-10



# Exercitiu pentru prezentare

Deschideți fișierul fasta ce conține secvența moleculei proteice primite. Realizați o interogare CATH și stabiliți clasa, arhitectura, topologia și superfamilia din care face parte cel puțin un domeniu din proteina alocată. Salvați o imagine care să demonstreze concluzia dvs., introduceți-o în prezentare și scrieți o scurtă explicație.