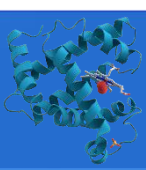


**8.04.2026**

**Curs 8 – Clasificarea structurală proteinelor – Bazele de date SCOP și CATH**

# Evoluția proteinelor



În general, se consideră că **evoluția** acționează la nivel molecular prin **3 mecanisme principale**:

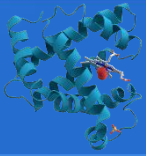
1. **Modificări aleatorii** ale secvenței ADN-ului prin încorporarea greșită de baze azotate ca urmare a efectelor factorilor mutageni sau a erorilor în replicarea ADN-ului; **rata de eroare a ADN-polimerazei: 1 nucleotidă greșită la  $10^6$ - $10^8$  nucleotide încorporate, bacteriile au  $5 \cdot 10^6$  pb**
2. **Procese reparatorii** ale ADN-ului ce au ca scop eliminarea defectelor la nivelul ADN-ului și pot duce la procese recombinatoriale (deleții, duplicații de gene, inversiuni);
3. **Presiunea selectivă** ce decide care dintre mutații/modificări vor fi păstrate în descendență.

Deși cu cei 20-22 de aminoacizi se pot constitui virtual un număr nelimitat de secvențe, conformații proteice și deci funcții, **două forțe importante limitează diversificarea extremă a structurilor proteice**:  
**A. Divergența funcției** - plecând de la modelul de mai sus al evoluției, se poate conclua logic că **toate structurile proteice tridimensionale provin din diversificarea unui număr finit de secvențe de aminoacizi numite secvențe ancestrale comune**. Aceste secvențe au fost diversificate pentru a crea noi conformații, deci funcții;

**B. Convergența funcției** – **două secvențe** ce provin din secvențe ancestrale **comune diferite pot evolua independent** dar pot fi selectate pentru aceeași funcție și deci **adopta conformații identice**.

**Similaritatea la nivel de secvență NU este suficientă pentru a identifica funcția unei proteine.**

# Evoluția proteinelor



Secvențe ancestrale comune

ANAAREMERE

PEREAREANA

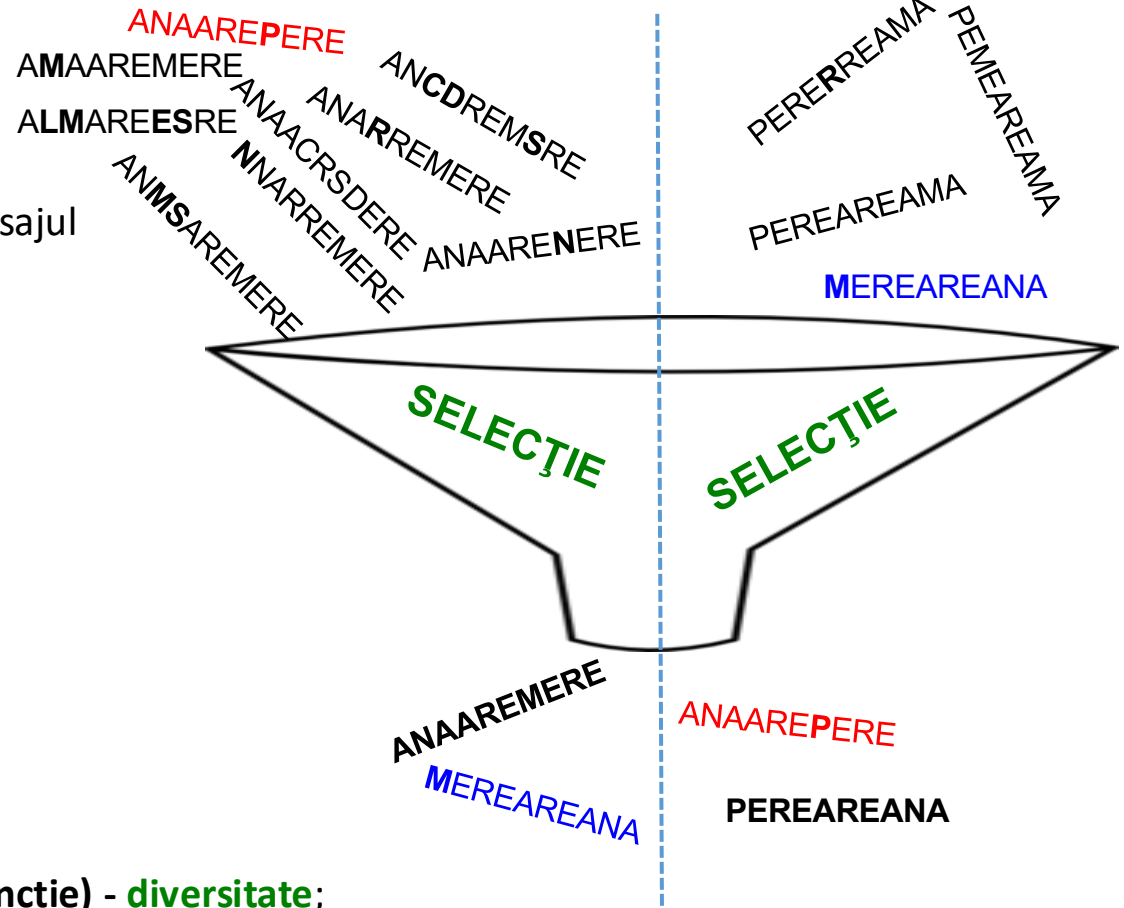
## Diversificarea mesajului prin mutații

rezultând numeroase variante de secvențe cu mesajul (funcția) similar(ă) sau nu cu cel (cea) inițial(ă);

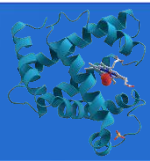
Presiunea selectivă face ca doar mesajul/funcția semnificativ(ă) să fie păstrat(ă);

Secvențele păstrate pot fi;

- similare ca secvență și codică același mesaj (funcție) - **diversitate**;
- lipsite de similaritate la nivel de secvență dar codificând același mesaj (funcție) - **evoluție convergentă**;
- similare ca secvență dar codificând mesaje (funcții) diferite – **evoluție divergentă**.



# Domeniile proteice ca unități de clasificare

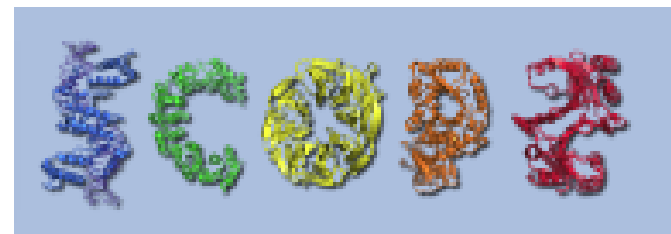


**Evoluția nu selectează secvența, ci selectează funcția. În cazul proteinelor funcția este dependentă de structura tridimensională.**

Cea mai mică unitate la nivelul căreia acționează evoluția este **domeniul proteic (curs 3)** – o **secțiune compactă dintr-o proteină independentă structural și frecvent și funcțional de restul proteinei**. Un domeniu proteic **va avea aceeași structură tridimensională și frecvent aceeași funcție** chiar dacă este separat de proteina din care provine.

În principiu, 1) **numărul de domenii proteice este finit** și 2) **între domenii apar legături evolutive**. **Domeniile proteice** reprezintă astfel singura **unitate de clasificare** ce poate fi folosită pentru a împărți proteinele într-o manieră ierarhică ce ține cont atât de **gradul de similaritate la nivel de secvență cât și de asemănările structurale**. Principalele sisteme de clasificare structural a proteinelor sunt:

**SCOP2 – Structural Classification Of Proteins**

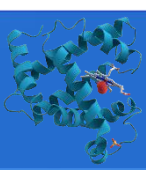


<http://scop2.mrc-lmb.cam.ac.uk/>

**CATH – Class Architecture Topology Homologous**



<http://cathdb.info/>



O bază de date ce realizează clasificarea **MANUALĂ** a domeniilor proteice prin analiza lor **VIZUALĂ**. Baza de date este organizată pe **4 nivele ierarhice**. Proteinele mici cu un singur domeniu vor apare într-un singur nivel ierarhic, proteinele de dimensiuni mari ce au mai multe domenii vor apare în mai multe nivele ierarhice.

**Nivele ierarhice de clasificare** a proteinelor in SCOP2:

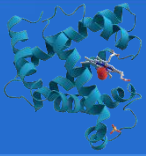
**1. Familia proteică** – cuprinde domenii **foarte apropiate ca secvență, structură și funcție ce provin din aceeași secvență ancestrală comună**. Două proteine fac parte din aceeași familie dacă:

- Au o **similaritate** la nivel de secvență de **minim 30%**; sau
- Au **aceeași structură tridimensională și funcție identică**.

Ex: Family: Voltage-gated potassium channels (4000034) – conține 5 proteine din 5 specii diferite

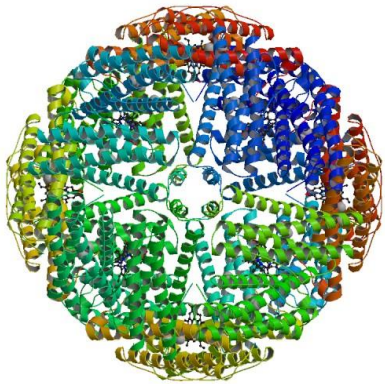
**2. Superfamilia proteică** – grupează mai multe familii proteice ce au o **similaritate de secvență redusă, dar au conformații și funcții similare**;

**3. Fold (conformație) proteic(ă)** – grupează domeniile din mai multe familii ce **au aceleași elemente ale structurii secundare dispuse în aceeași ordine și conectate în aceeași manieră**. Proteinele din același fold **diferă** una de cealaltă prin **lungimea și structura tridimensională a zonelor ce conectează elementele de structură secundară**.

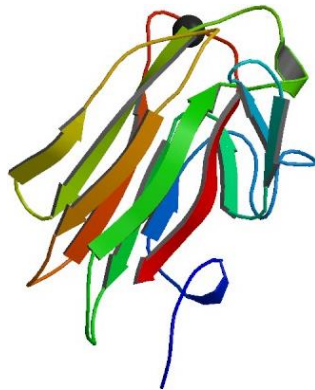


**4. Clasa proteică** – cuprinde proteine cu diverse conformații dar care au în comun același tip de organizare a structurii secundare. În SCOP2 au fost descrise 5 clase principale:

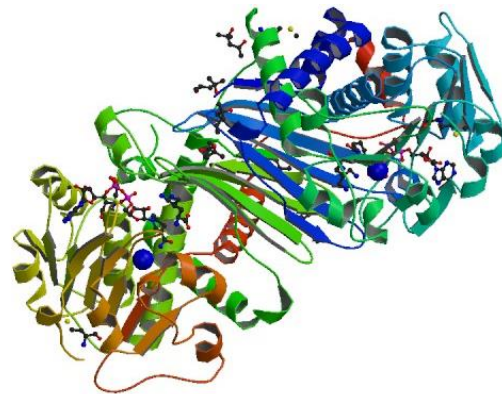
1. **All- $\alpha$**  – proteine/domenii proteice ce conțin numai structuri  $\alpha$ -helicale;
2. **All- $\beta$**  – proteine/domenii proteice ce conțin numai structuri  $\beta$ -pliate;
3.  **$\alpha/\beta$**  - proteine/domenii proteice ce conțin structuri  $\alpha$ -helicale alternând cu structuri  $\beta$ -pliate;
4.  **$\alpha+\beta$**  - proteine/domenii proteice ce conțin structuri  $\alpha$ -helicale și structuri  $\beta$ -pliate segregate;
5. **Small proteins** – proteine de dimensiuni mici ce nu au elemente de structură secundară sau acestea sunt foarte mici.



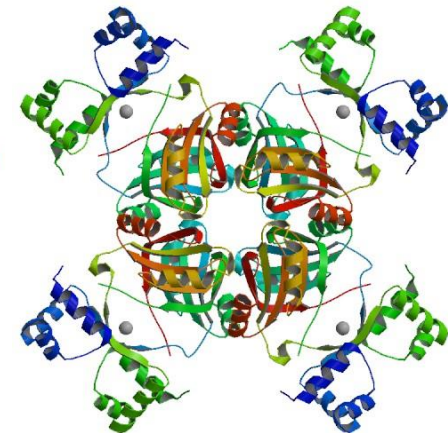
1F5N



2VMH

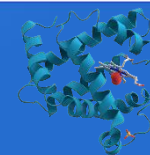


1X7D



2CG4

# Clasa Small Proteins in SCOP2

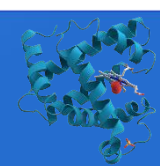


## Class

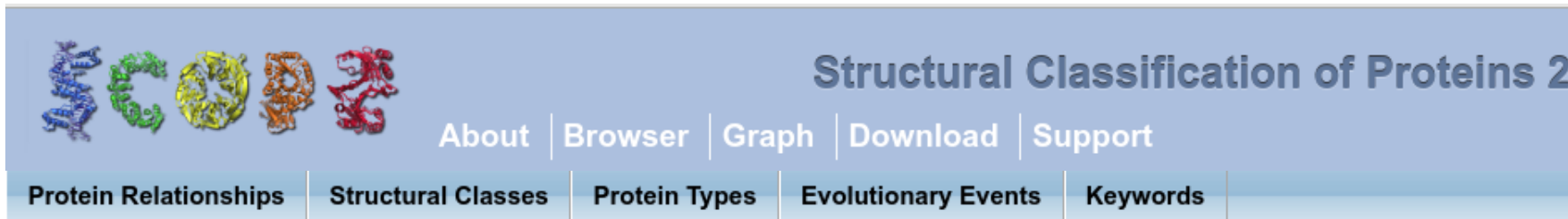
class **Small proteins** (1000004)  
proteins with little or no secondary structures.  
Derived from [SCOP 56992](#)

## Children

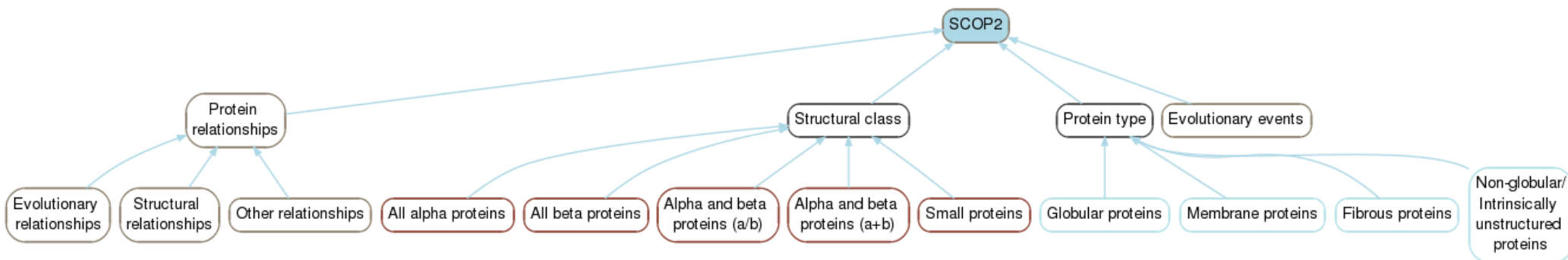
- fold **GATA zinc finger-like** (2000049)  
contains a beta-hairpin between the two pairs of zinc ion ligands and one or more turns of helix at the C-terminus  
Derived from [SCOP 57715](#)
- fold **Rubredoxin-like** (2000112)  
metal ion-binding fold comprising two beta-hairpins; each hairpin contains at its tip two metal ion-coordinating residues, usually, cysteines.  
Derived from [SCOP 57769](#)
- fold **HypF zinc finger-like** (2000160)  
coordinates zinc ion with two CxxC motifs, each motif is located at the N-end of a helical turn  
Derived from [SCOP 55820](#)
- fold **DnaJ/Hsp40 cysteine-rich domain** (2000953)  
metal(zinc)-bound extended beta-hairpin fold  
Derived from [SCOP 57937](#)
- iupr **Calmodulin binding domain (CaMBD)** (2001258)

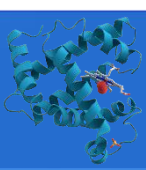


Scopul principal al SCOP2 este oferii o clasificare a tuturor structurilor tridimensionale proteice cunoscute. Clasificarea poate fi accesată prin prisma a 5 categorii distincte:



1. **Legătura** structurală sau evolutivă dintre proteine - aici se găsesc nivelele ierarhice de clasificare până la superfamilie;
2. **Clasa structurală** – echivalentul nivelului ierarhic clasă proteică;
3. **Tipul de proteină** – clasifică proteinele în **a. proteine globulare**, **b. proteine membranare**, **c. proteine fibrilare** și **d. proteine fără structură secundară**;
4. Evenimente evolutive specifice și de amploare ce au dus la apariția unei proteine date;





O bază de date ce realizează clasificarea **SEMI-AUTOMATĂ** a domeniilor proteice prin analiza **SIMILARITĂȚII** la nivel de secvență și evaluarea **VIZUALĂ** a topologiei. Cele **4 nivele ierarhice** de organizare sunt similare, dar nu identice cu cele din SCOP2.

**Nivele ierarhice de clasificare** a proteinelor in CATH:

**1. Clasa proteică** – este dată de **tipul predominant de structuri secundare**. În CATH au fost descrise 3 clase:

-**Predominat alfa** – domenii ce conțin majoritar structuri  $\alpha$ -helicale;

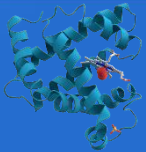
-**Predominat beta** – domenii ce conțin majoritar structuri  $\beta$ -pliate;

-**Alfa-beta** – domenii ce conțin ambele tipuri de structuri majoritare  $\alpha$ -helicale și  $\beta$ -pliate;

**2. Arhitectura proteică** – este dată de **modul în care structurile secundare sunt poziționate în spațiu** fără a se ține cont de conformația secțiunilor de legătură dintre structurile secundare;

**3. Topologia proteică** – este dată de **modul în care elementele structurii secundare sunt interconectate** între ele. Două domenii pot avea aceeași arhitectură (acesași orientare a structurilor secundare în spațiu), însă dacă ordinea de conectare a structurilor secundare este diferită vor face parte din topologii diferite.

**4. Superfamilia omoloagă** (**H**omologous superfamily) -



**4. Superfamilia omoloagă (Homologous superfamily)** – sunt grupate domenii între care există relații evolutive ce sunt dovedite prin similitudini la nivel de secvență, structură sau funcție. În cadrul unei superfamilii, domeniile sunt grupate în familii funcție de identitatea la nivel de secvență (35, 60, 95, 100%)

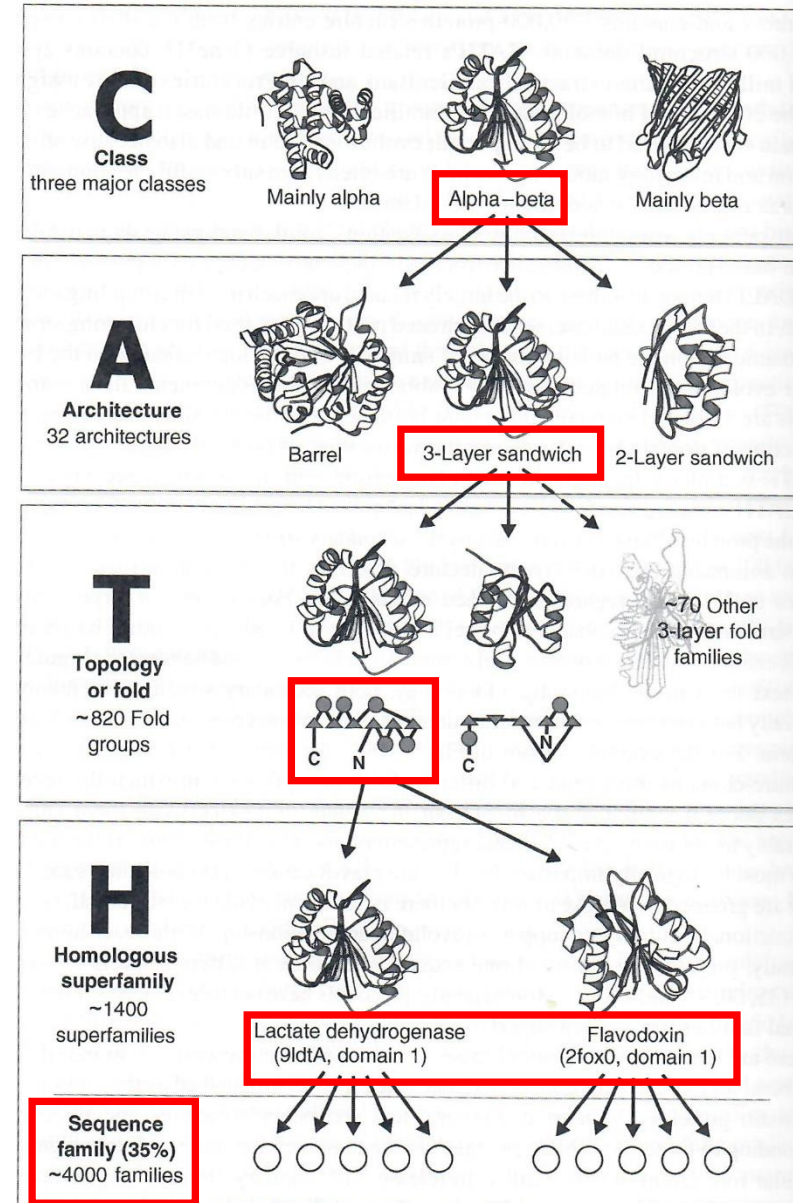
Ex. de clasificare in CATH:

**Clasa Alfa-beta din CATH conține 3 arhitecturi distincte printre care și arhitectura sandwich-ului tristratificat.**

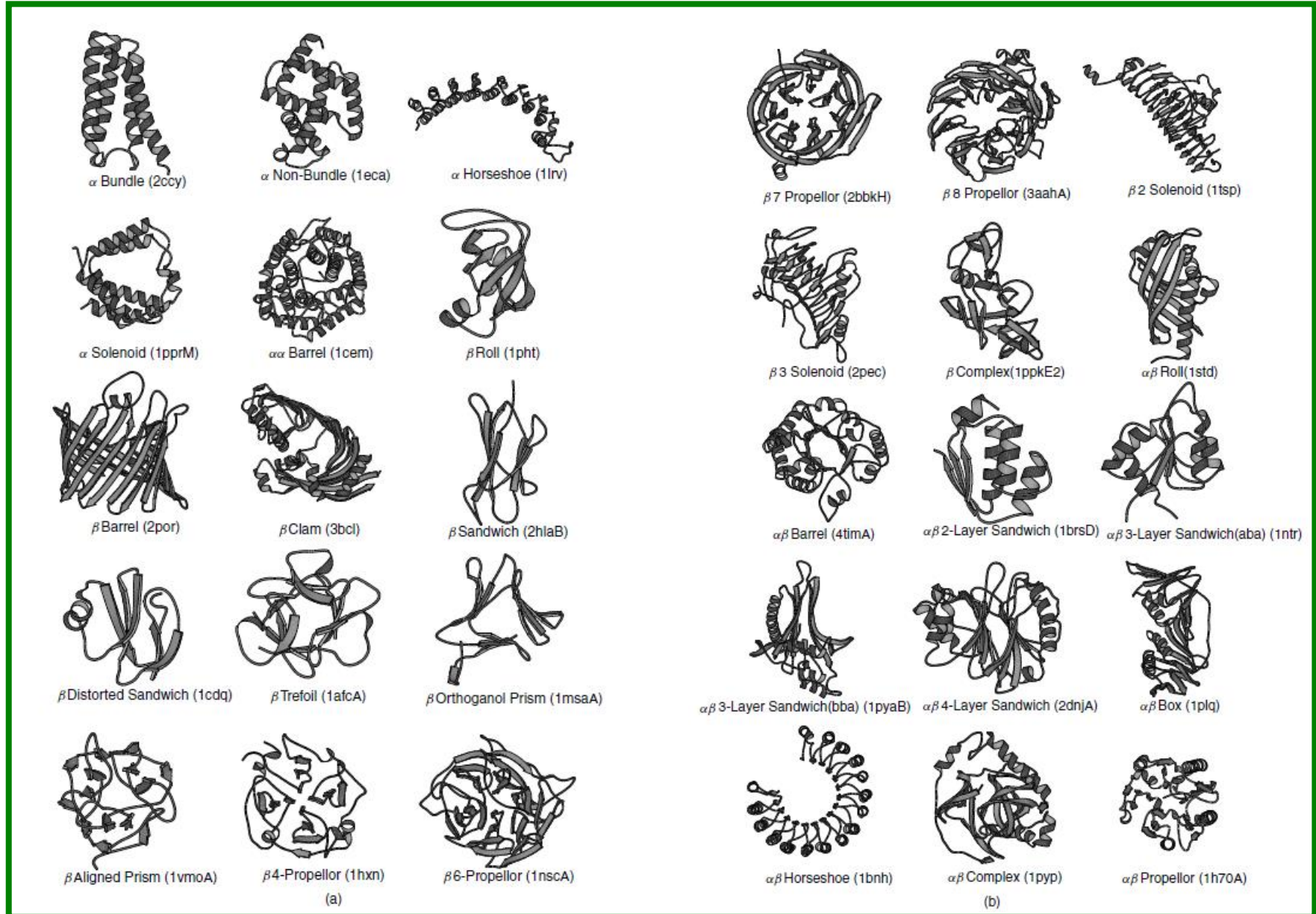
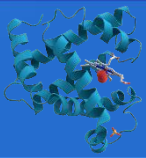
**Această arhitectură** poate fi realizată prin combinarea și aranjarea în 72 de moduri a structurilor secundare – **72 de topologii** diferite pentru aceeași arhitectură. Două dintre topologii sunt reprezentate.

**Aceași topologie** poate fi realizată de **domenii ce diferă unul de celălalt la nivel de secvență sau funcție**. În exemplul dat două proteine cu aceeași topologie dar cu funcție diferită – două superfamilii diferite.

**Aceași funcție** poate fi realizată de secvențe diferite, între care există un nivel de similaritate – lactat dehidrogenaza (aceeași enzimă) din diverse specii (alte secvențe)

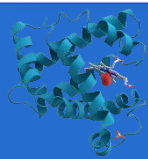


# Arhitecturile majore din baza de date CATH



# Accesarea bazei de date CATH

<http://www.cathdb.info/>



CATH

Home Search Browse Download About Support

Search CATH by keywords or ID

## CATH / Gene3D v4.2

95 million protein domains classified into 6,119 superfamilies

Search by keywords, PDB code, GO term, etc

Search

Core classification files for the latest version of CATH-Plus (v4.2) are now available to download. Daily updates of our very latest classifications are also available. We are currently working on generating the CATH-Plus database for v4.2 which comprises all the extra derived data from the classification data. This includes: incorporation of the latest Gene3D sequence and functional annotation data; updating the Functional Families (FunFams); creating new superfamily superpositions; producing structural clusters for each superfamily. We will update the web pages when this data is ready.



### 3D Structure

Find out what 3D structure your protein adopts

Find out more

Go



### Protein Evolution

Learn about a particular protein family and how it evolved

Find out more



### Protein Function

Investigate the function of your protein

Find out more

Go



### Conserved Sites

Look at protein sites that are highly conserved and implicated in function

Find out more

Go



### Download Data

Download data files and query CATH via webservice

Go



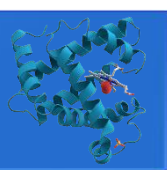
### Learn more

Find out how CATH is created and maintained, how to link to CATH and more

Go

Spre deosebire de SCOP2, CATH oferă posibilitatea de a clasifica o proteină de interes plecând de la secvență (Aplicație la seminar)

# La ce este utilă clasificarea proteinelor?



Similaritatea la nivel de secvență este utilă și poate fi folosită pentru a stabili funcția unei proteine dacă:

- 1. Există un nivel de identitate la nivel de secvență între secvența țintă și cea subiect suficient de mare (cele 2 secvențe sunt apropiate d.p.v. evolutiv).**
- 2. Există date experimentale privind funcția secvenței subiect.**

**În cazul în care una din condițiile de mai sus nu este îndeplinită, BLAST este lipsit de semnificație.**

Suplimentar, BLAST analizează secvența proteică per ansamblu și nu face distincție între diversele domenii ale unei proteine.

Prin stabilirea structurii terțiare a unei proteine necunoscute, alocarea domeniilor și stabilirea nivelelor ierarhice cărora aparține, se pot afla informații legate de funcția sa prin analogie cu proteine mult mai îndepărtate evolutiv.

**Rolul unei proteine este dată de:**

**1. Amplasarea celulară** – în citosol, integrată în membrană, atasată de membrană – **clasa și fold-ul în SCOP2, clasa și arhitectura în CATH;**

**2. Funcția** – enzimă, proteină structurală - **fold-ul în SCOP2, arhitectura în CATH;**

În cazul enzimelor:

**3. Reacția generală catalizată** – **Superfamilia** în SCOP2, **topologia și superfamilia** omoloagă în CATH;

**4. Substratul asupra căruia acționează** – poate fi indicat de **familie** în SCOP2 și CATH, dar frecvent nu poate fi dedus prin simpla clasificare a domeniilor.