

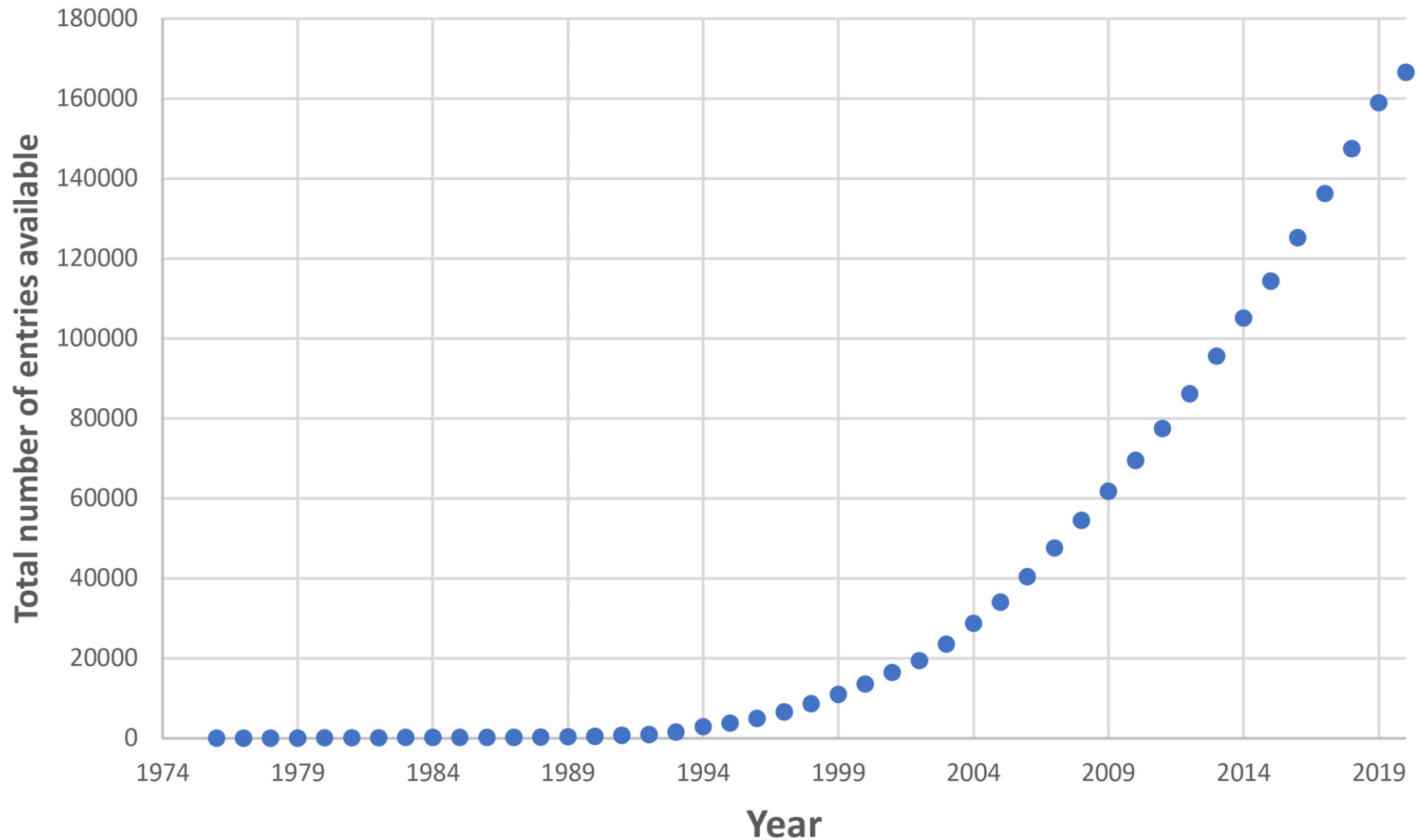
2DProts: Family-wide 2D diagrams of protein secondary structure

Radka Svobodová

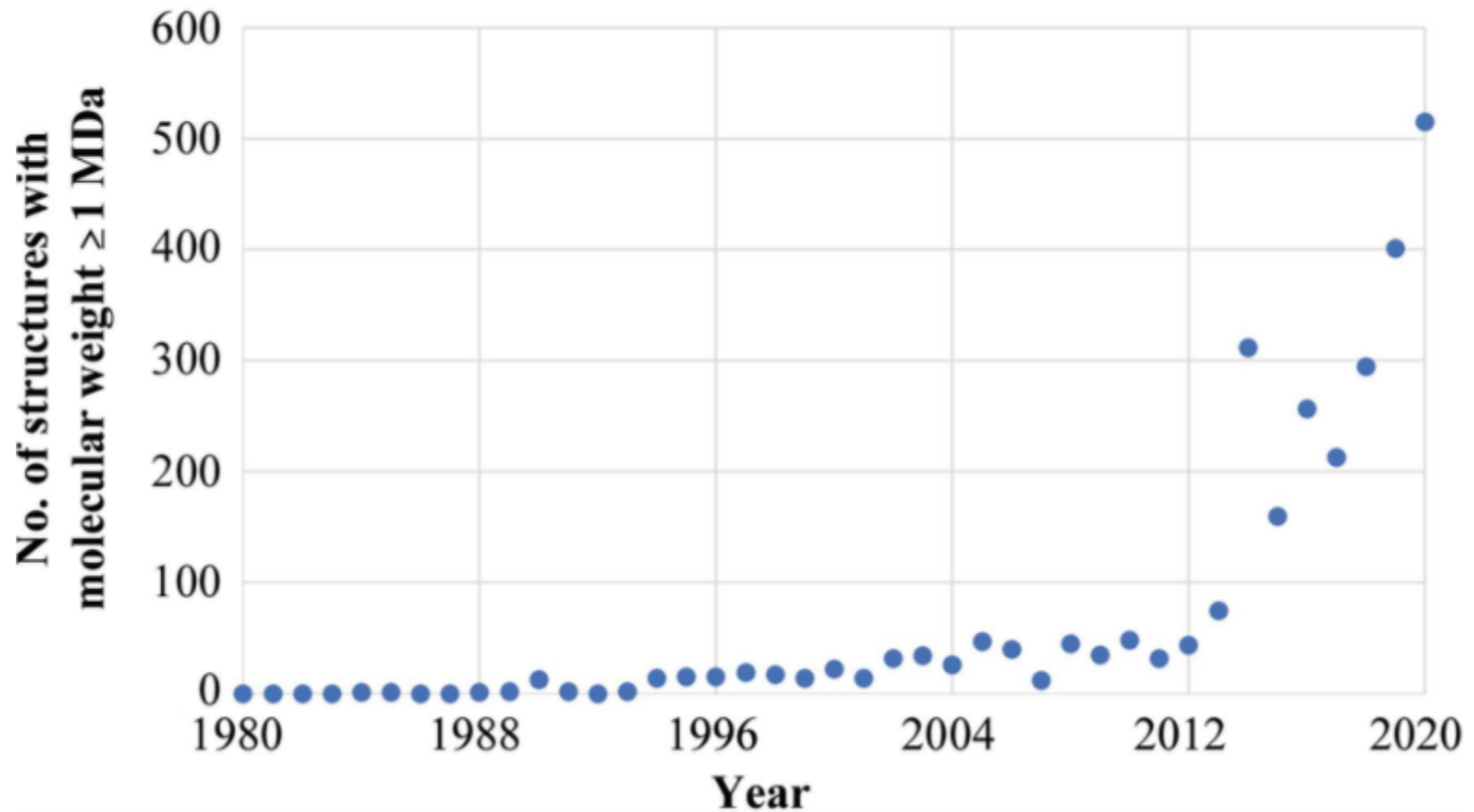
CEITEC
MASARYK UNIVERSITY



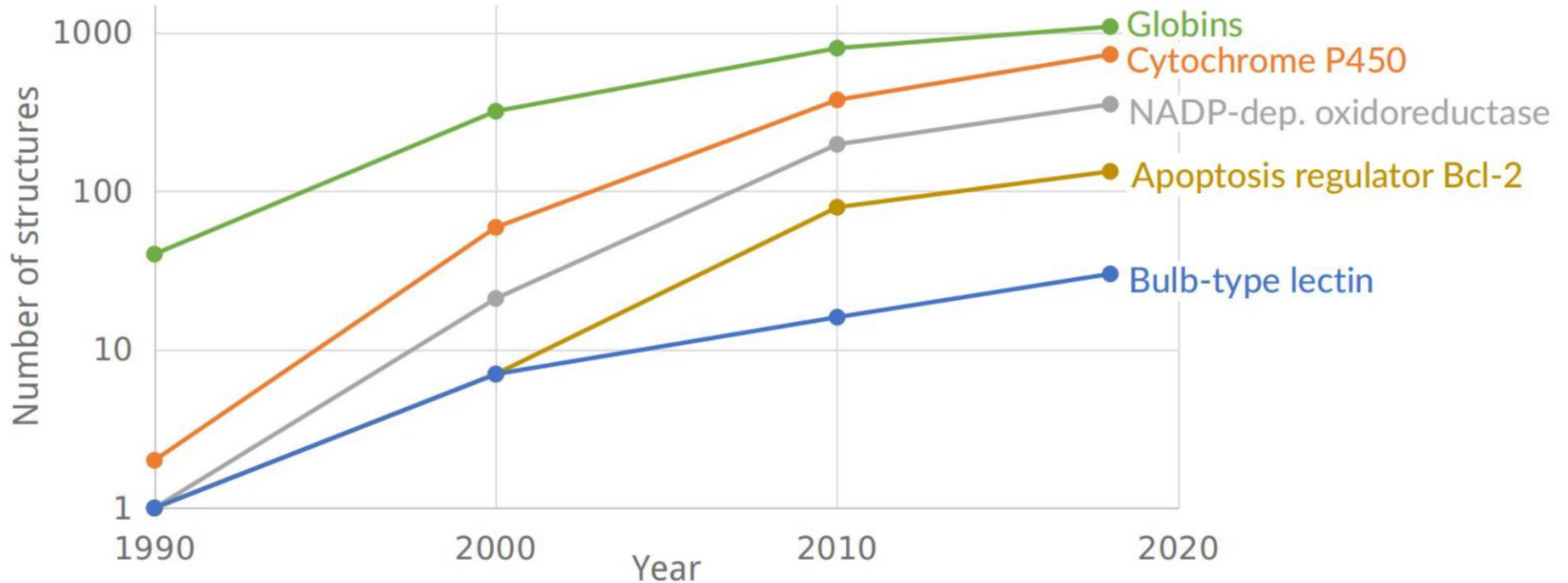
Current trends: Number of available structures grows



Current trends: Size of deposited structures also grows



Current trends: Protein families are getting bigger



Analysis of individual structure



Analysis of a whole family

Protein families – members

Search the PDB archive < PDBe < | x +

https://www.ebi.ac.uk/pdbe/entry/search/index/?searchParams=%7B"q_cath_topology":%5B%7B"value":"...

EMBL-EBI Services Research Training About us

PDBe

PROTEIN DATA BANK

P450

Search

Advanced search

Feedback

Download

/page ▾

Molecule name (187)	Sequence family (19)	Structure domain (4)
NADPH--cytochrome P450 redu... (166)	PF00067 : p450 (1109)	Cytochrome p450 (93)
Cytochrome P450 (133)	IPR001128 : Cytochrome P450 (967)	Cytochrome P450 (162)
Bifunctional cytochrome P450/N... (127)	IPR036396 : Cytochrome P450 ... (967)	Cytochrome p450-Terp; domain 2 (68)
Cytochrome P450-cam (127)	IPR002397 : Cytochrome P450,... (498)	NADPH-cytochrome p450 Redu... (40)
Cytochrome P450cam (127)	IPR002401 : Cytochrome P450,... (276)	
Cytochrome P450 102A1 (123)	IPR002403 : Cytochrome P450,... (103)	
Cytochrome P450(BM-3) (123)	IPR002402 : Cytochrome P450,... (58)	
Cytochrome P450BM-3 (123)	IPR008072 : Cytochrome P450,... (58)	
Flavocytochrome P450 BM3 (123)	IPR023173 : NADPH-cytochrom... (46)	
Cytochrome P450 3A3 (77)	IPR008068 : Cytochrome P450,... (36)	
More...	More...	

Protein families – members

The screenshot shows the PDB search interface. At the top, there is a navigation bar with links for EMBL-EBI, Services, Research, Training, and About us. Below this is the PDBe logo and a search bar containing the text "Ex. - hemoglobin, BRCA1_HUMAN". A red circle highlights the "Advanced search" button. Below the search bar, there is a filter bar showing "CATH topology : Cytochrome p450". Below the filter bar, there are buttons for "Show filters", "Advanced search" (highlighted with a red circle), and "Download". Below these buttons is a tabbed interface with "Entries", "Macromolecules", "Compounds", and "Protein families". Below the tabs is a pagination bar showing "Entries 1 to 10 of 930" and a dropdown menu for "Release date (asc)" and "10 /page".

Search the PDB archive < PDBe < | x +

https://www.ebi.ac.uk/pdbe/entry/search/index/?searchParams=%7B"q_cath_topology":%5B%7B"value":"...

EMBL-EBI Services Research Training About us

EMBL-EBI

PDBe
Protein Data Bank in Europe

Ex. - hemoglobin, BRCA1_HUMAN

Search

Examples: [hemoglobin](#), [BRCA1_HUMAN](#)

Advanced search

Feedback

PDBE / SEARCH

CATH topology : Cytochrome p450

Show filters **Advanced search** Download

Entries Macromolecules Compounds Protein families

< 1 2 3 ... 93 > Entries 1 to 10 of 930

Release date (asc) 10 /page

Protein families – members

The image shows a web browser window displaying the PDB search interface. An "Advanced search form" modal is open, showing search criteria for "CATH topology" and "Release date".

Advanced search form

CATH topology ⓘ

condition: Equal to

Example: sh3 type barrels

Cytochrome p450

Release date ⓘ

condition: AND Less than equal to

Example: 4/20/2013

2000

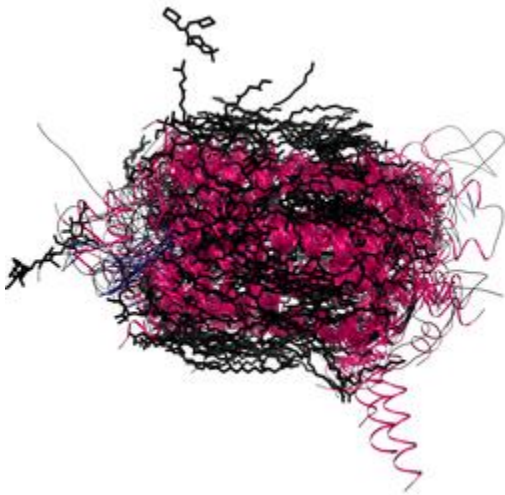
Select to add a search field :

Select / type to filter options

Buttons: Submit, Cancel

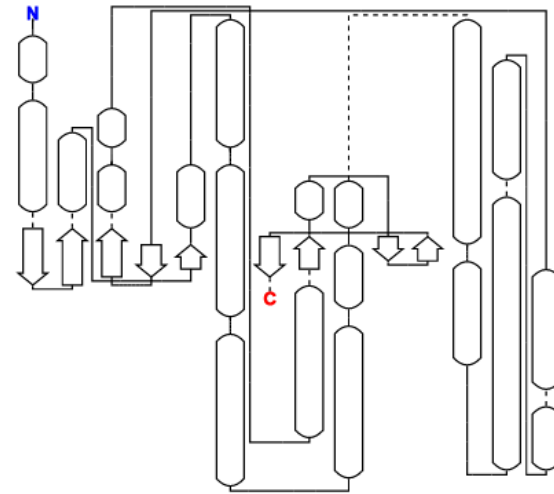
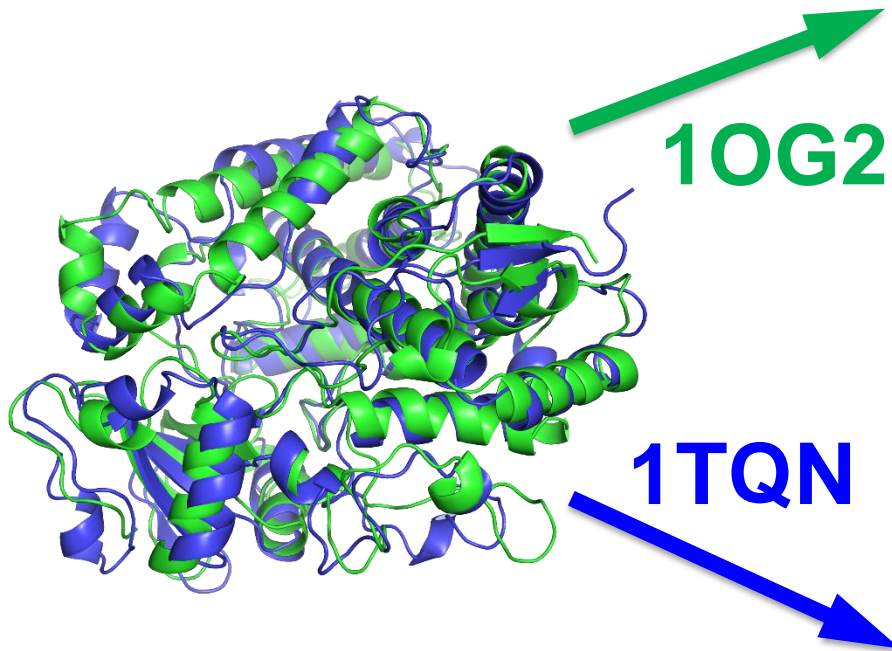
Protein secondary structure: A clue for protein family analysis

- Comparison of protein family members
 - Different species
 - Different substituents
 - Mutations
 - Active and inactive forms
- Firm and flexible secondary structure elements
- Binding of ligands

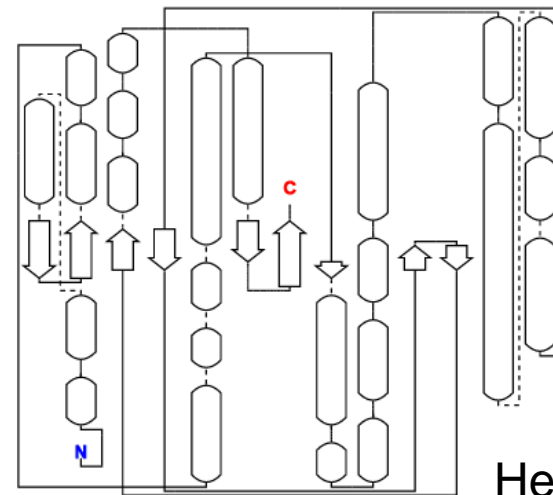


Visualization of secondary structure in 2D: Solved in past? Not for protein families!

ISSUE 1: Similar proteins have
different 2D diagrams



RMSD: 2.295 Å



Hera, PDBe

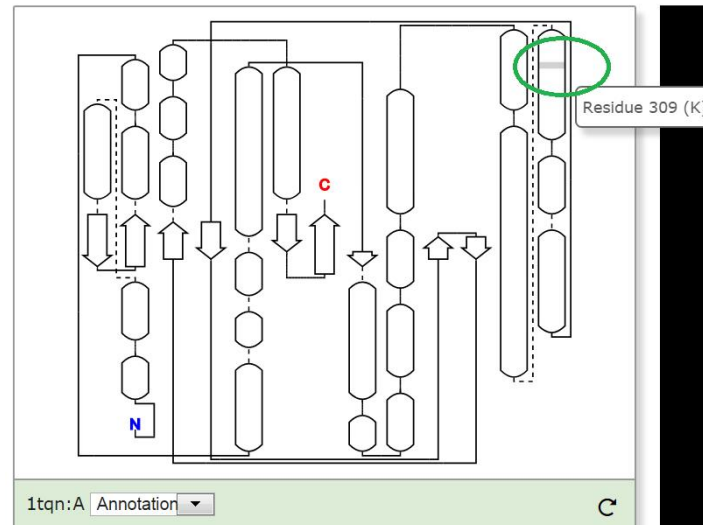
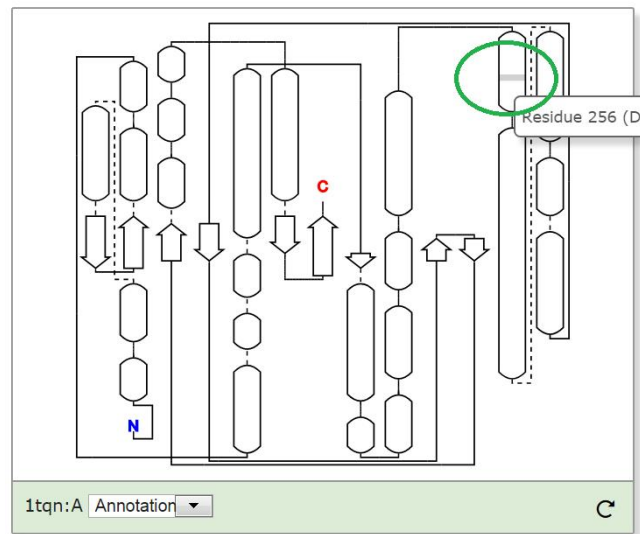
Visualization of secondary structure in 2D: Solved in past? Not for protein families!

ISSUE 2:

Secondary structure elements close in 2D diagrams are far in reality

1TQN

Hera,
PDBe

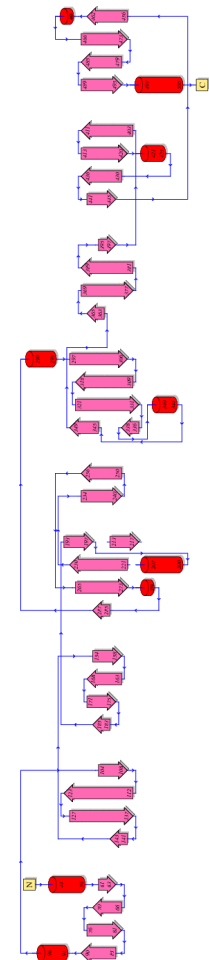


Visualization of secondary structure in 2D: Solved in past?

ISSUE 3: 2D diagrams does not reflect
a shape of a protein



1ORW



HERA

2DProts

Protein family based 2D diagrams

Input:

- A CATH superfamily (e.g. 2.60.120.400)
- the list of its domains (e.g. 1gztA00, 1ourA00, ...)
- PDB structures of these domains.

Step 1: For each domain in the given family, **find its SSEs** (via SecStrAnnotator) and annotate them:

- topologically equivalent SSEs have the same name.

Step 2: For each group of SSEs with the same name, **compute average length and frequency of SSE** occurrence.

2DProts

Protein family based 2D diagrams

Step 3: For each domain in the family:

Step 3.1: Try to select an appropriate **starting layout** among the previously computed domains.

Step 3.2: Group all **β -strands into sheets** and compute a 2D model of each individual sheet.

Step 3.3: Divide the **helices and sheets into primary** (common for most of the domains) **and secondary** (the remaining ones).

Step 3.4: **Place all primary** helices and sheets into the 2D diagram.

Step 3.5: **Adjust the angles** of the primary helices and sheets.

Step 3.6: **Add all secondary** helices and sheets into the 2D diagram.

Step 3.7: **Adjust the angles** of the secondary helices and sheets.

Step 4: Draw an individual 2D diagram for each domain and a common multiple 2D diagram for the whole family

2DProts Database

2DProts

Custom entry

User manual

Description of methods

e.g., 1r9nA01, 1r9n, 2.140.10.

Search 2DProts

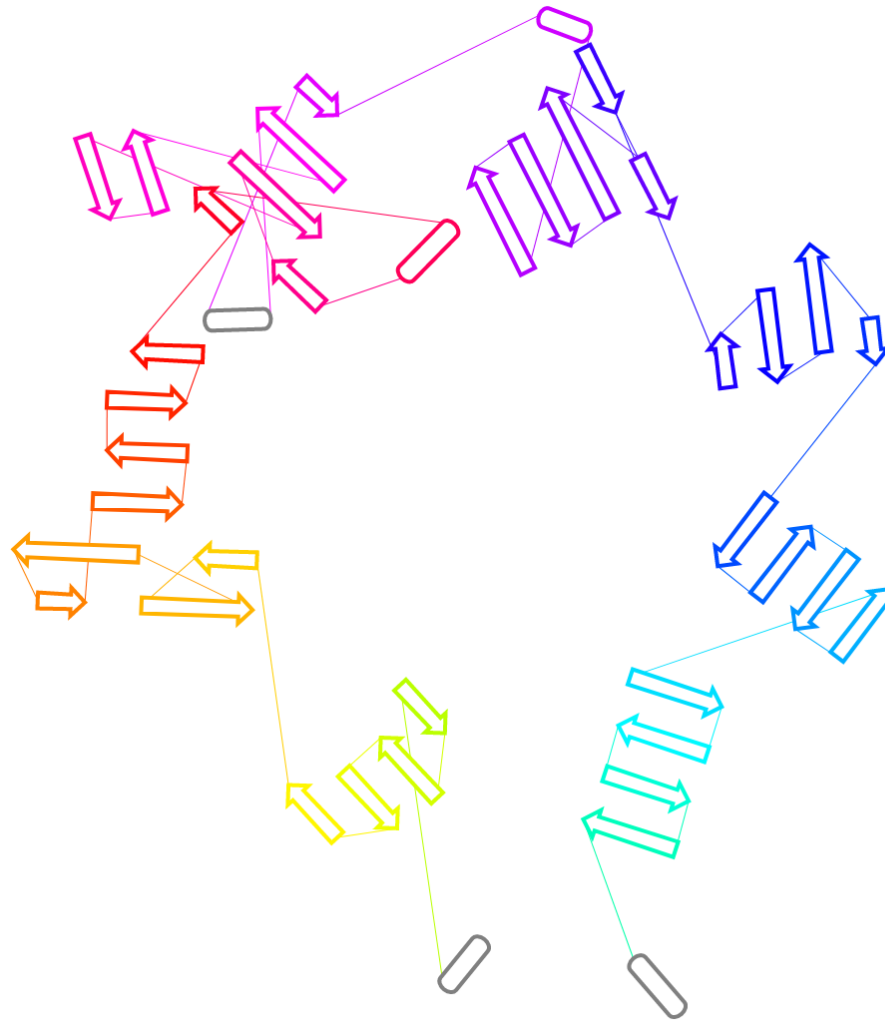
2DProts

Database of 2D diagrams of domain secondary structures

- Precalculated **2D diagrams** of domains from all CATH families
- Includes **multiple 2D diagrams** for a whole protein family
- Freely available at: <http://ncbr.muni.cz/2DProts/>
- Updated each week

2DProts outputs

2D diagram of a protein domain



2DProts outputs: Multiple 2D diagram of protein domains in a family



With opacity



No opacity

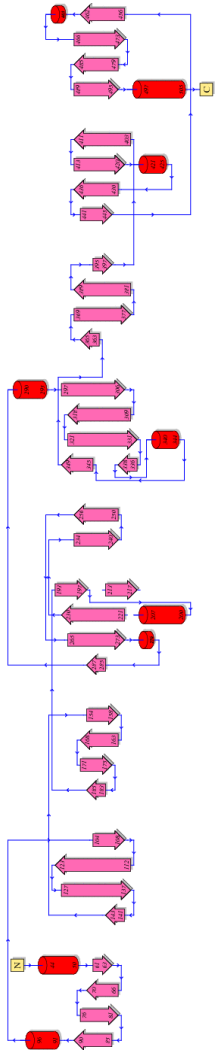
Superfamily: Dipeptidylpeptidase IV (2.140.10.30)

PROTEIN

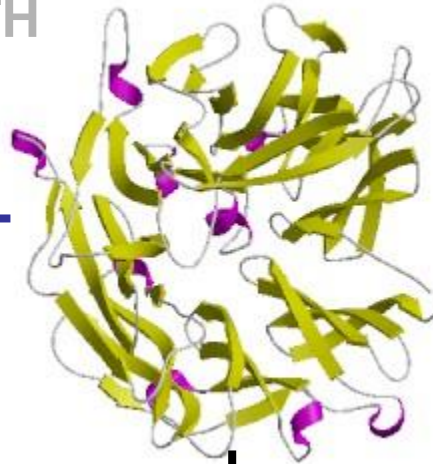
PROTEIN FAMILY

Current solution

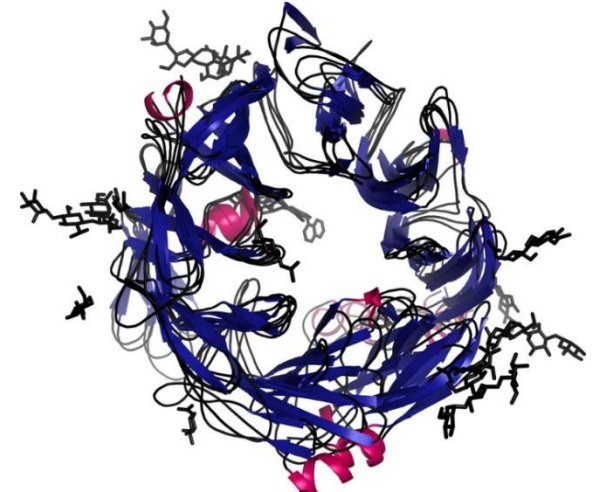
HERA



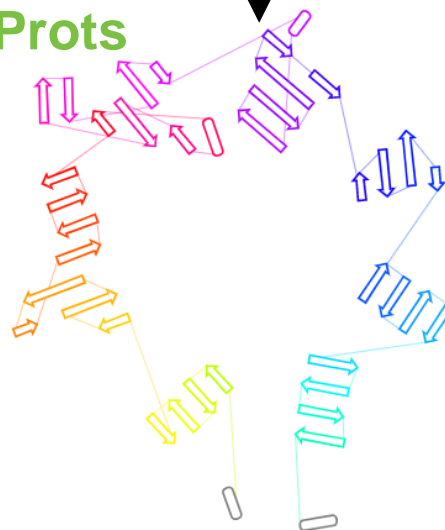
CATH



PROTEIN FAMILY



2DProts



TEC

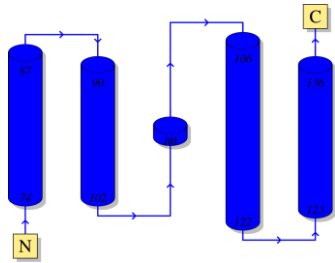
Superfamily: Iron dependent repressor (1.10.60.10)

PROTEIN

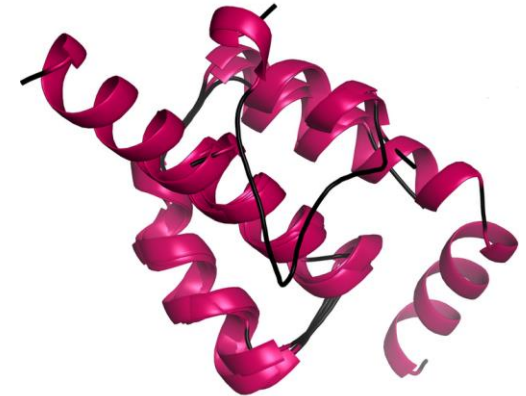
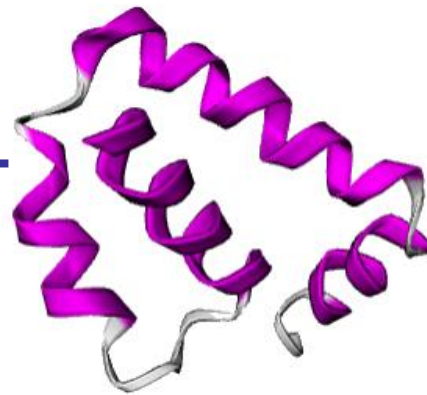
PROTEIN FAMILY

Current solution

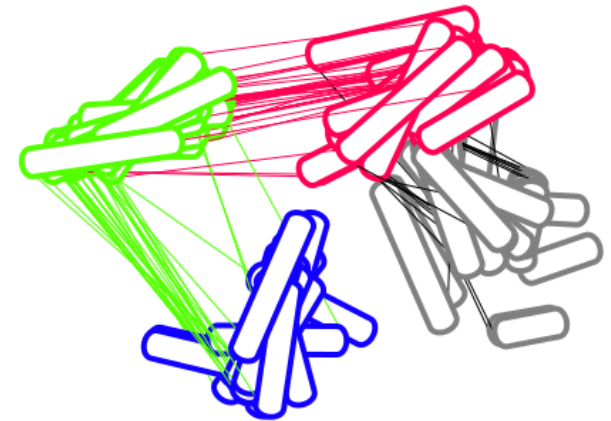
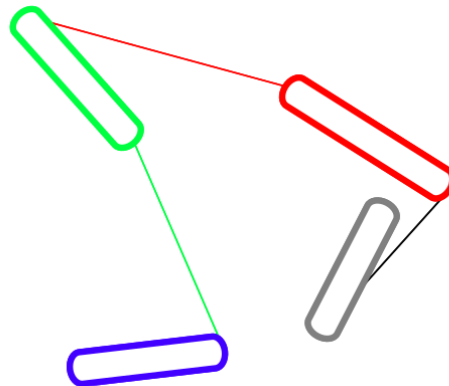
HERA



CATH



2DProts



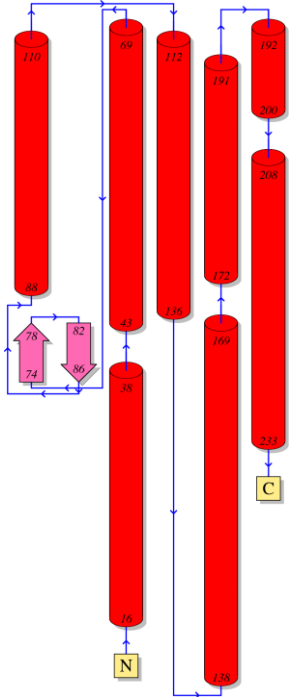
Superfamily: Rhodopsin 7-helix transmembrane proteins

PROTEIN (1.20.1070.10)

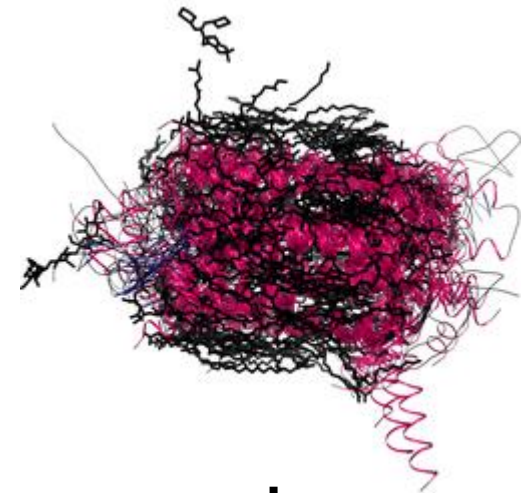
PROTEIN FAMILY

Current solution

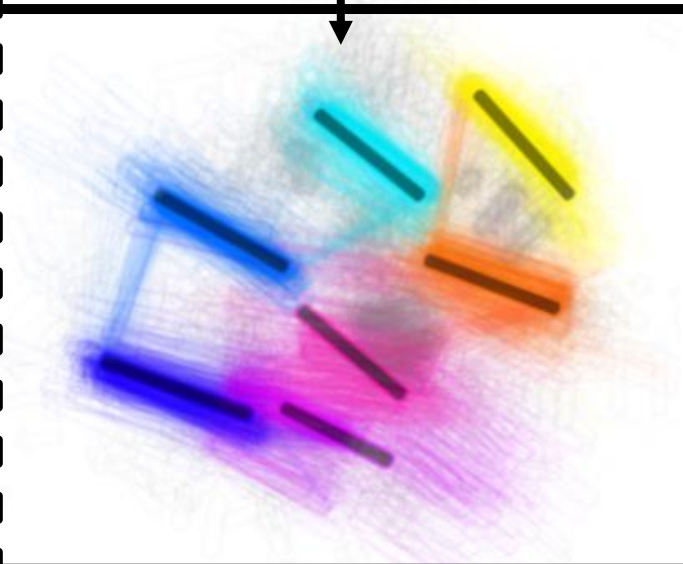
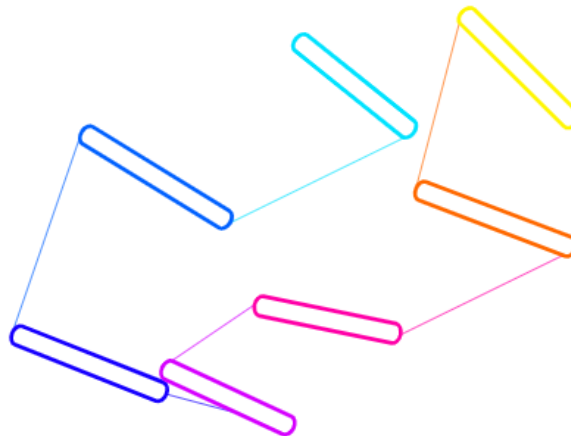
HERA



CATH



2DProts



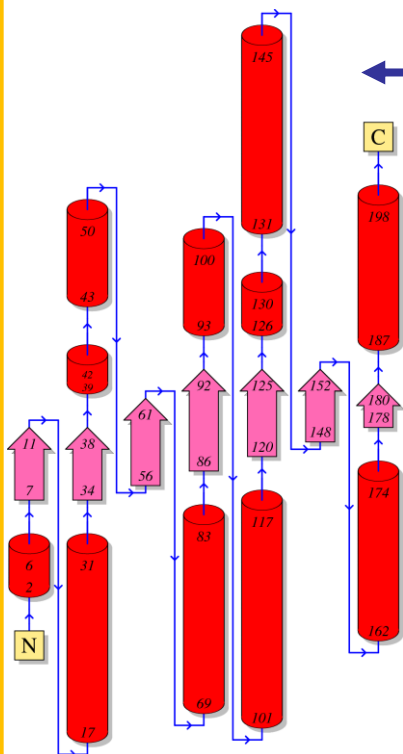
Superfamily: Aldolase class I (3.20.20.70)

PROTEIN

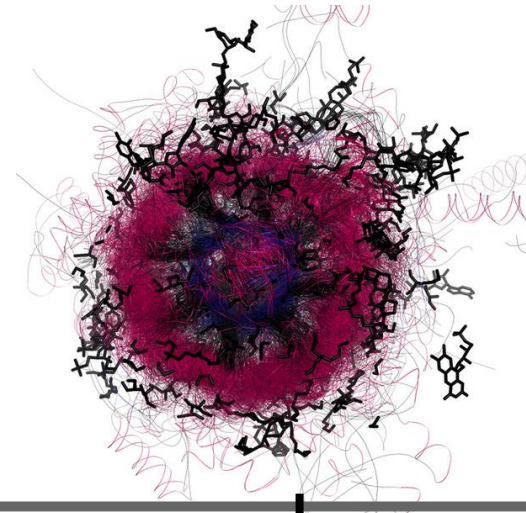
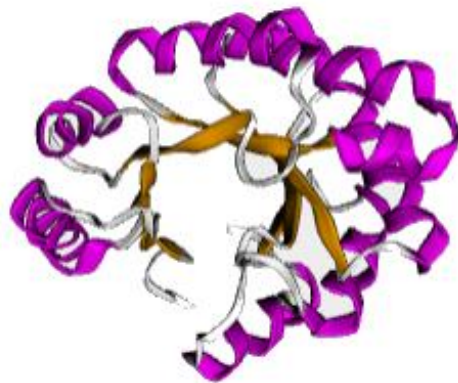
PROTEIN FAMILY

Current solution

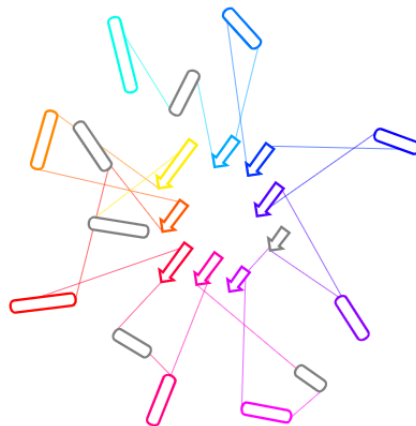
HERA



CATH



2DProts



2DProts

Other features

- Precalculated results for CATH structural clusters
- Possibility to process user defined sets of domains (e.g., select some organism, resolution, experimental method, etc.)

2DProts

User defined sets

[2DProts](#)[Custom entry](#)[User manual](#)[Description of methods](#)[Search 2DProts](#)

2DProts

Create custom image

Here you can create custom multi image from provided domains. All the domains in their latest version have to be from one family. Enter one domain (e.g. **1tqnA00**) per line.

Domains:

```
1gztA00
1gztB00
1gztC00
1gztD00
```



For example if you want to show only all domains of **1ecp** protein from family **3.40.50.1580**, just copy&paste following list:

```
1ecpA00
1ecpB00
1ecpC00
1ecpD00
1ecpE00
1ecpF00
```

2DProts

User defined sets

2DProts

Custom entry

User manual

Description of methods

e.g., 1r9nA01, 1r9n, 2.140.10.3

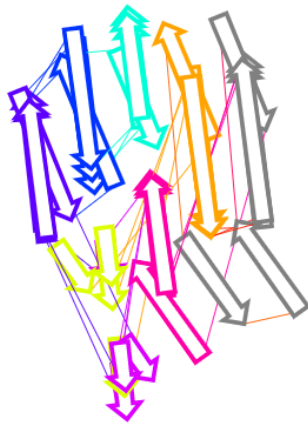
Search 2DProts

2DProts

Job f55d4171-8420-4455-add6-b5b6e950e2f2

2D multi diagram

No transparency, no averages



From family
2.60.120.400

Domains (4)

- 1gztA00
- 1gztB00
- 1gztC00
- 1gztD00

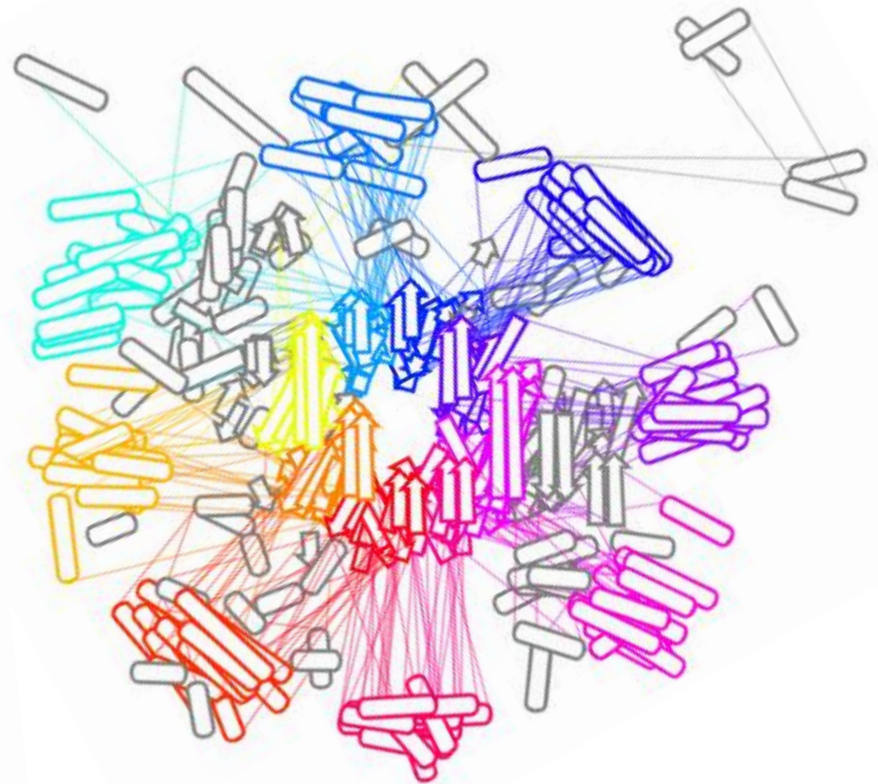
2DProts

User defined sets

Aldolase class I (3.20.20.70)

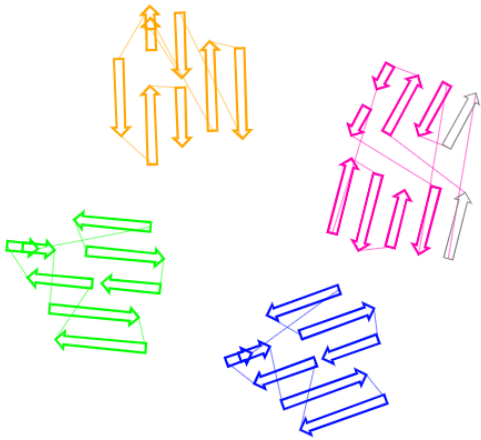


Archea

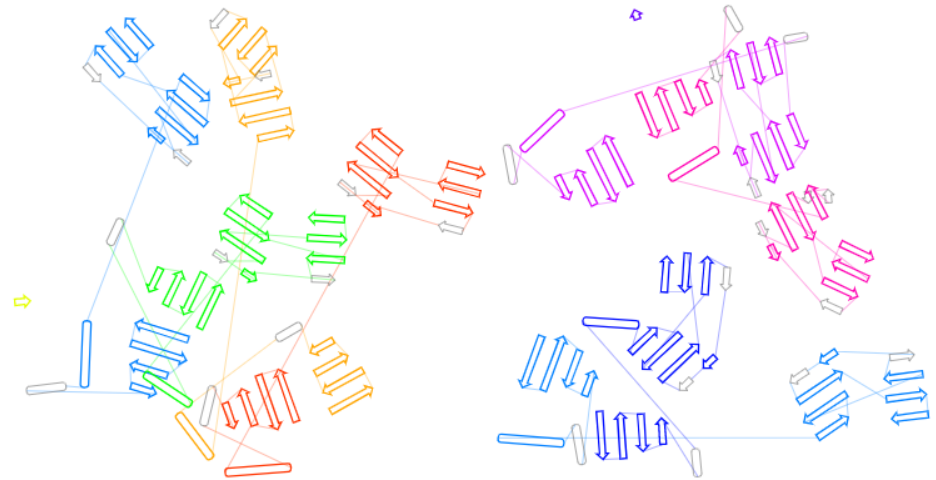
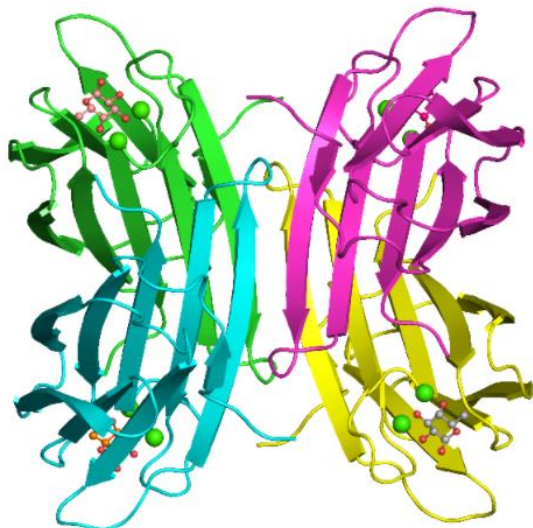


Thermotoga maritima

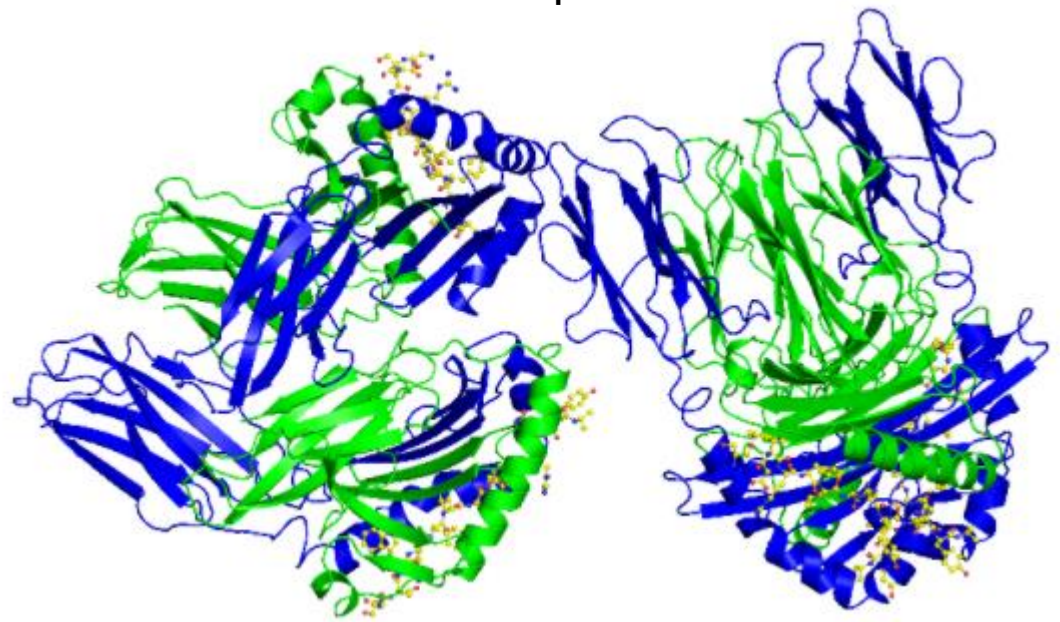
2DProts new features Proteins



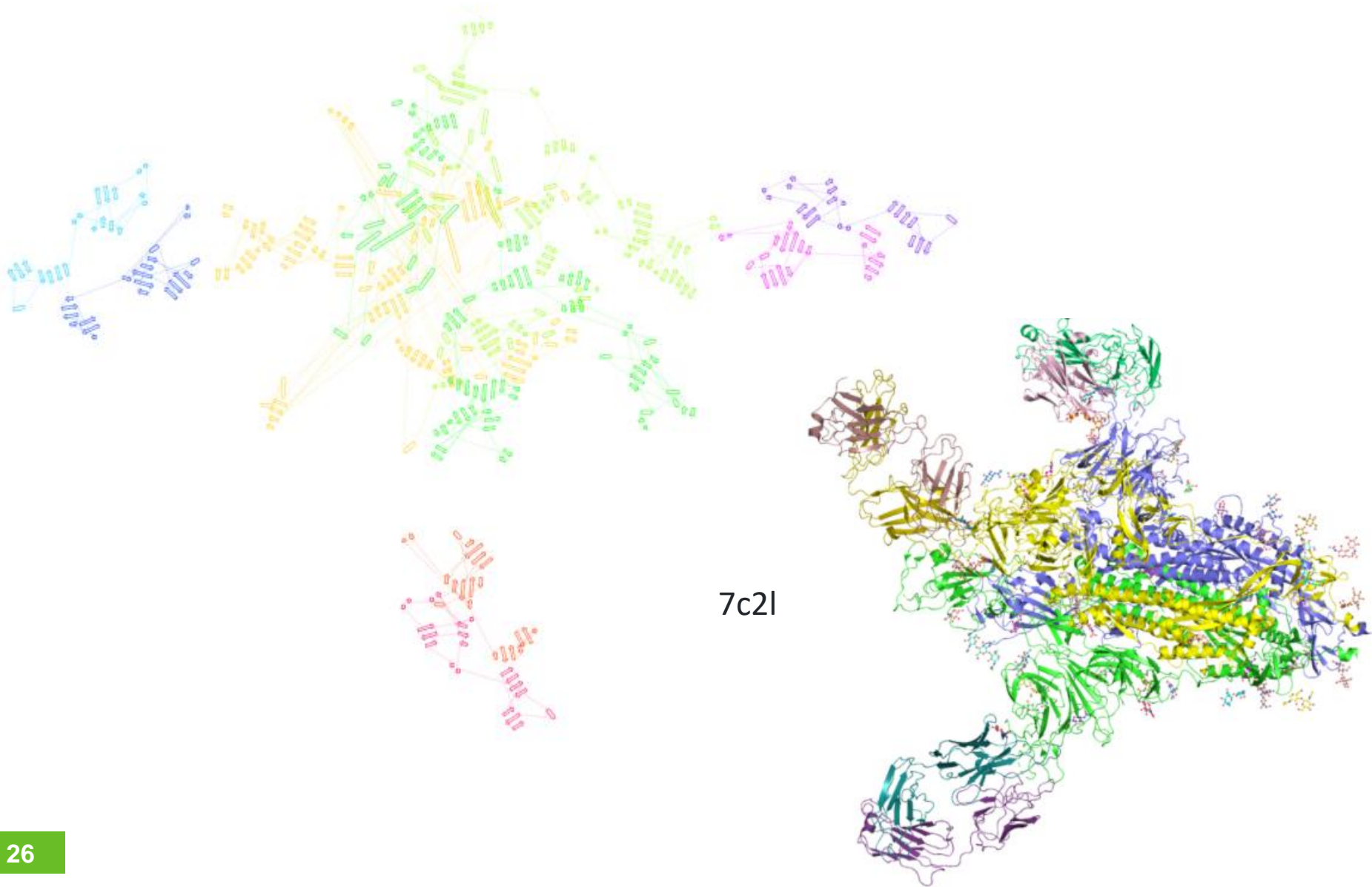
1gzt



1aqd

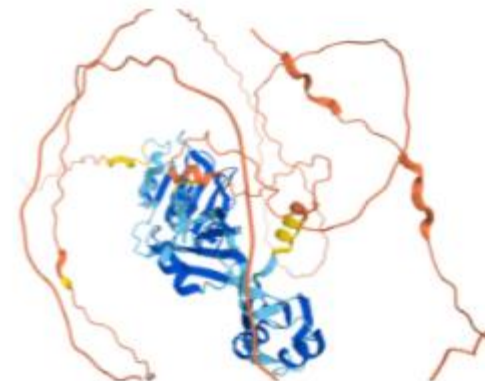
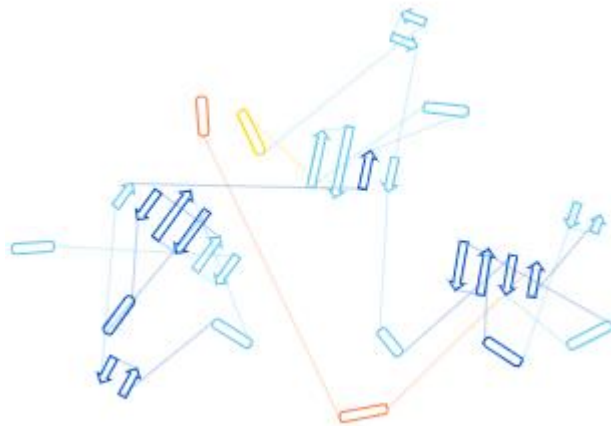
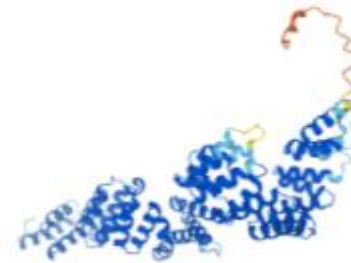
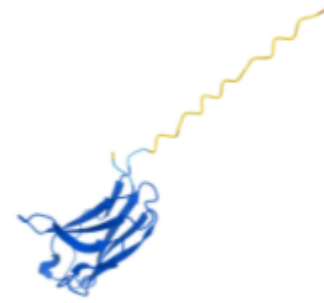
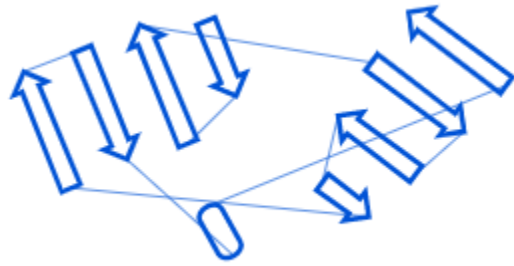


2DProts new features Proteins



2DProts new features

Alpha fold



2DProts integration to CATH

The screenshot shows a web browser window displaying the CATH Superfamily 1.10.60.10 page. The browser address bar shows the URL www.cathdb.info/version/latest/superfamily/1.10.60.10/superposition. The page title is "CATH Superfamily 1.10.60.10" with a description "Iron dependent repressor, metal binding and dimerisation domain". A "View in Gene3D" button is visible in the top right.

The page features a navigation menu with "Home", "Search", "Browse", "Download", "About", and "Support". A search bar is located in the top right corner. Below the navigation, there are several tabs for different views: "SS", "Alignbow", "2DProts", and another "2DProts" tab. The "2DProts" tabs are highlighted, indicating the current view.

On the left side, there is a "SUPERFAMILY LINKS" section with a "Summary" link and a "Superfamily Superposition" button. Below this, there are links for "Classification / Domains", "Functional Families", and "Structural Neighbourhood".

The "Functional Families" section provides an overview of structural clusters (SC) and functional families within the CATH Superfamily. It includes a diagram showing a red circle representing SC.1, which is linked to four functional families: Diphtheria t Transcriptio, Transcriptio, Iron (Metal), and Manganese.

The main content area displays a 3D superposition figure of structural domains within the CATH superfamily. The figure shows several domains in different colors (purple, yellow, red) and orientations, illustrating their relative distance and position. A green button labeled "Visit 2DProts (1.10.60.10)" is located in the top right of the main content area.

A blue text box at the bottom of the main content area states: "These superposition figures provide an indication of the relative distance and position of secondary structure elements of structural domains within CATH superfamilies. Image are generated by the 2DProts database".

Publications

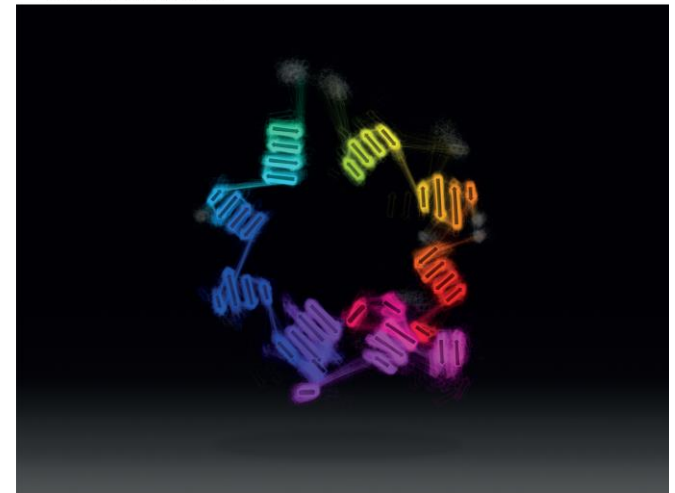
Sillitoe I, ..., Berka K, Hutařová Vařeková I, Svobodová R., et al., 2021. CATH: increased structural coverage of functional space. *Nucleic Acids Research*, 49(D1), D266-D273.

Hutařová Vařeková, I., Hutař, J., Midlik, A., Horský, V., Hladká, E., Svobodová, R., & Berka, K. (2021). 2DProts: database of family-wide protein secondary structure diagrams. *Bioinformatics*.

PRINT ISSN: 0305-1048
ONLINE ISSN: 1362-4962

Nucleic Acids Research

VOLUME 49 DATABASE ISSUE JANUARY 8, 2021
<https://academic.oup.com/nar>



OXFORD
UNIVERSITY PRESS

Open Access
No barriers to access – all articles freely available online



Acknowledgement

Ivana Hutařová Vařeková

Jan Hutař

Adam Midlik

Karel Berka

Aliaksei Charesneu

Eva Hladká

ELIXIR and ELIXIR CZ



CATH, EBI: Ian Sillitoe, Christine A Orengo



EMBL-EBI, PDBe: Dr. Sameer Velankar



Thank you for your attention