

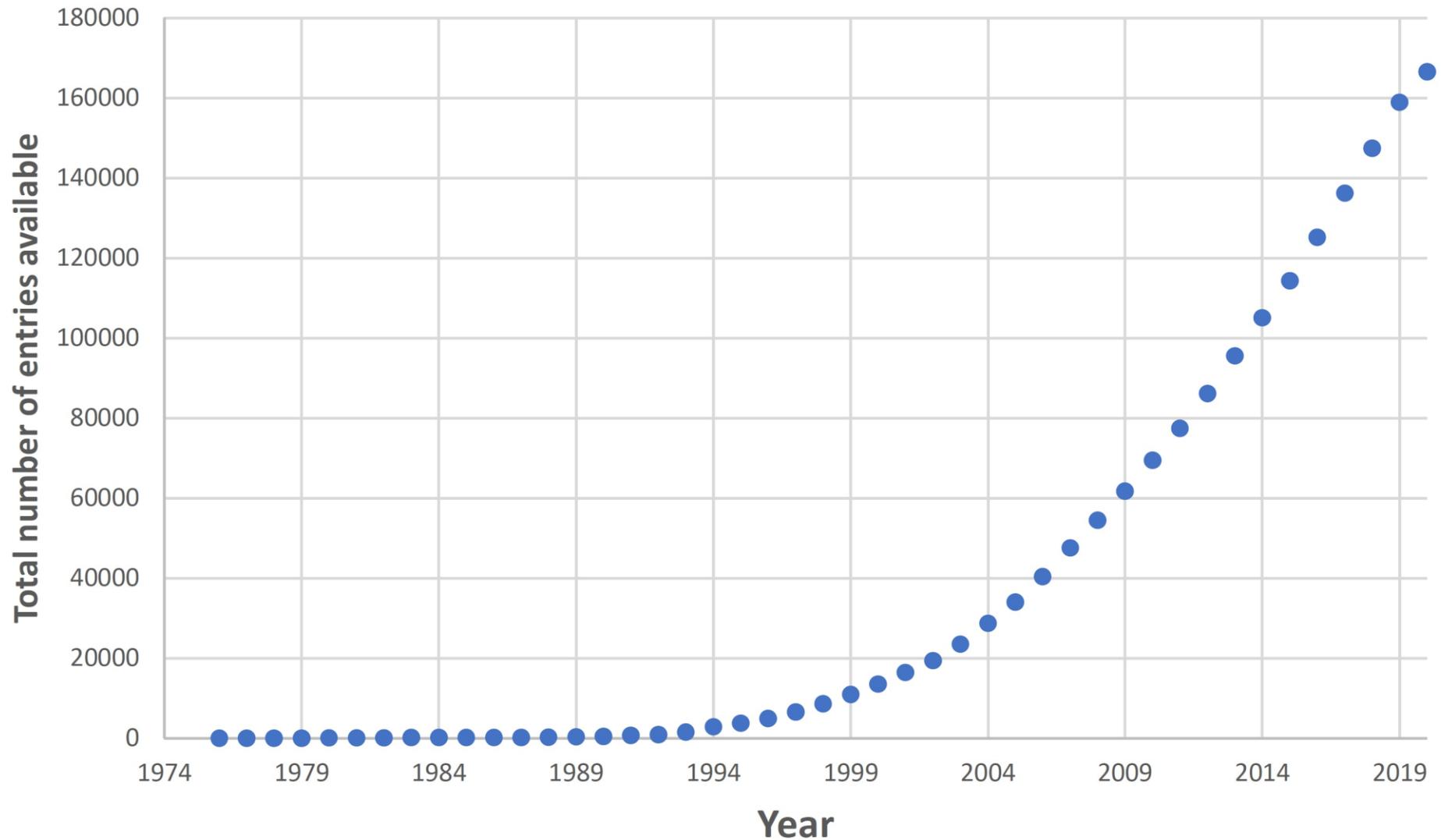
# Secondary structure diagrams of proteins, protein families and ligands

Radka Svobodová

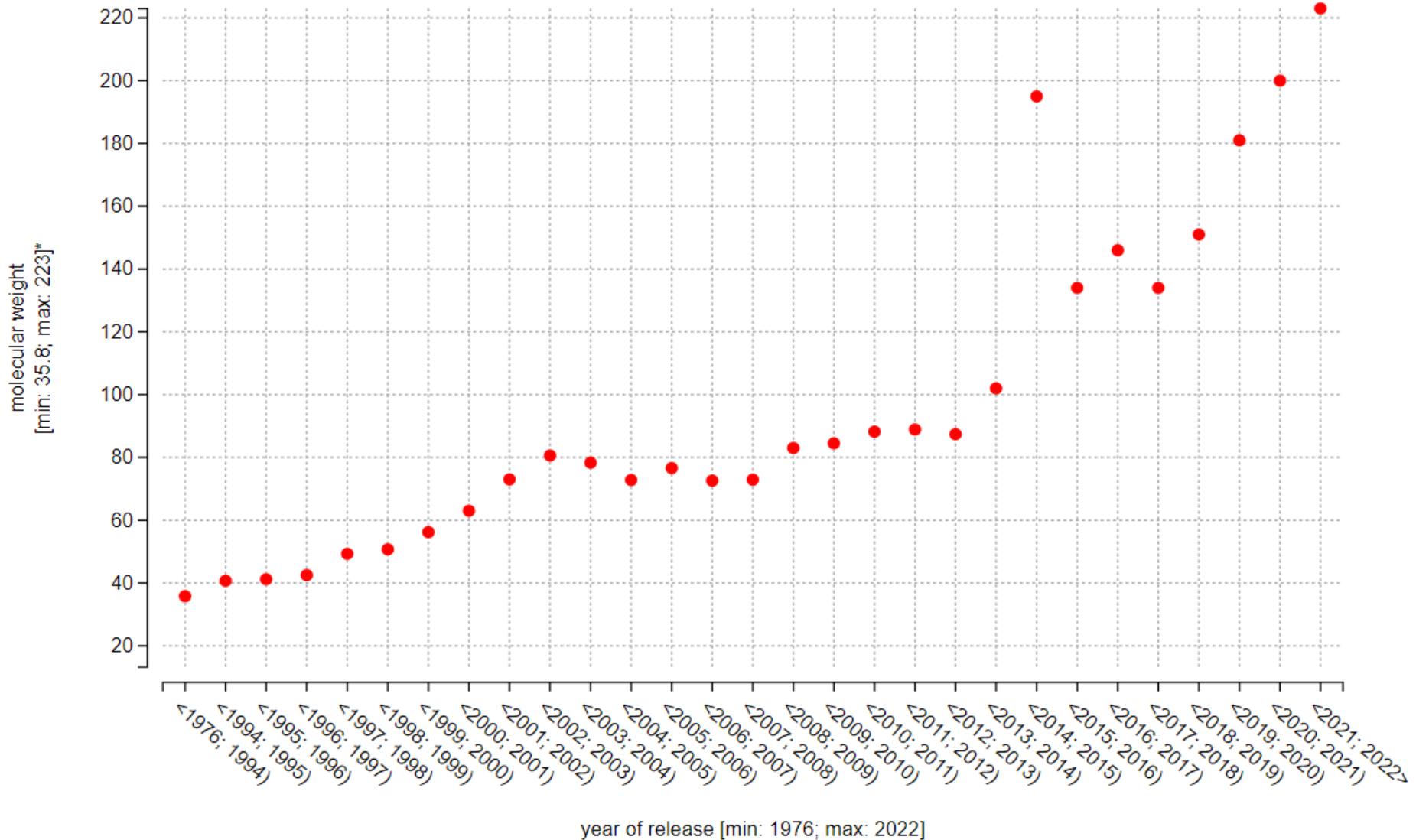
NCBR, CEITEC  
MASARYK UNIVERSITY



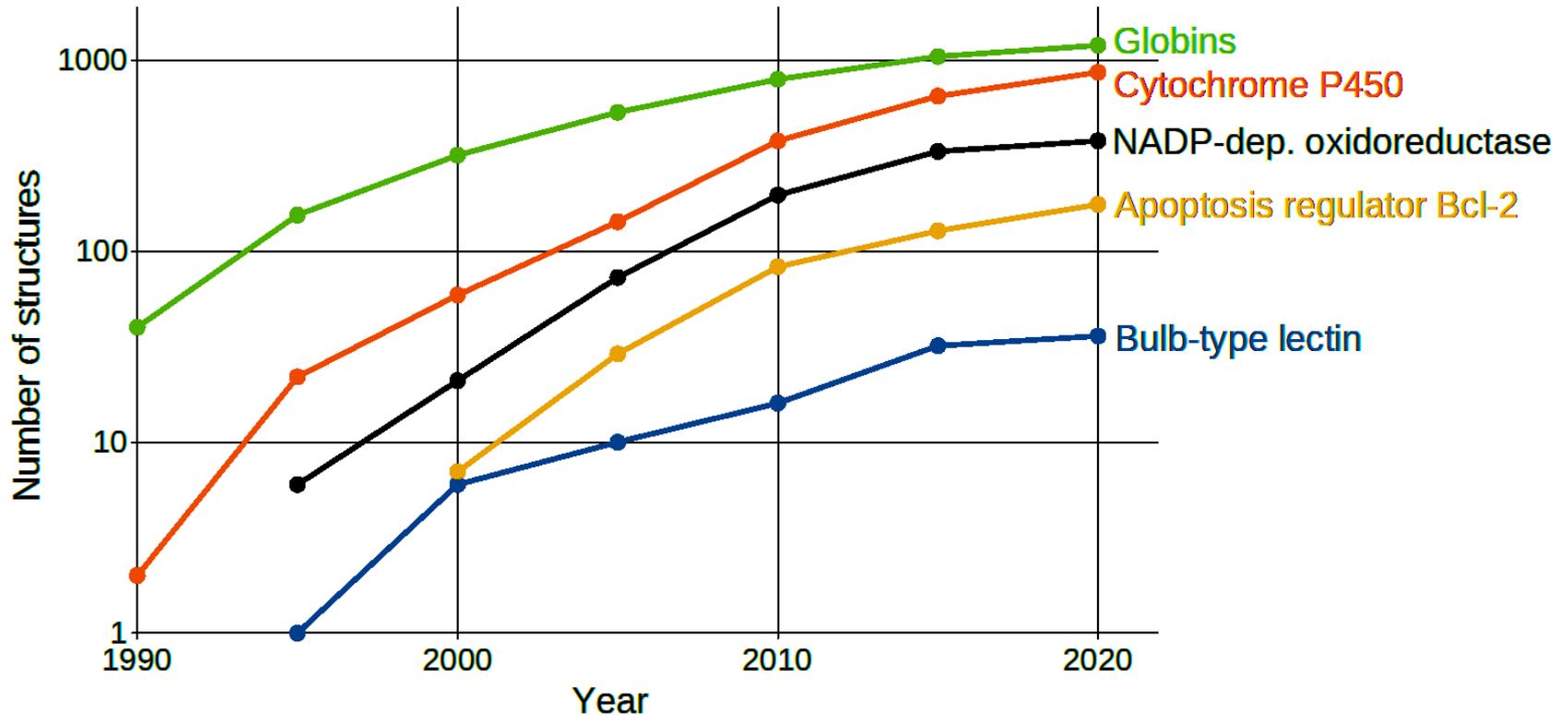
# Current trends: Number of available structures grows



# Current trends: Size of deposited structures also grows



# Current trends: Protein families are getting bigger



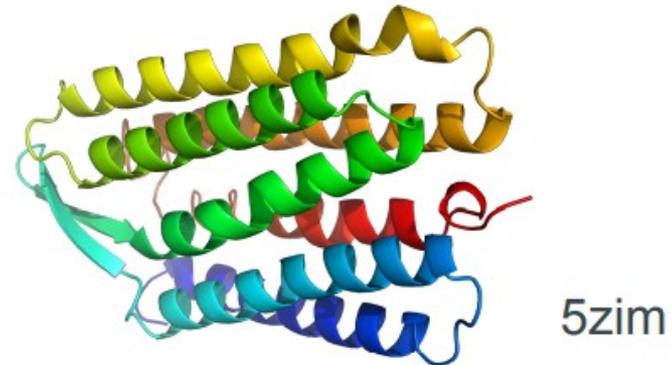
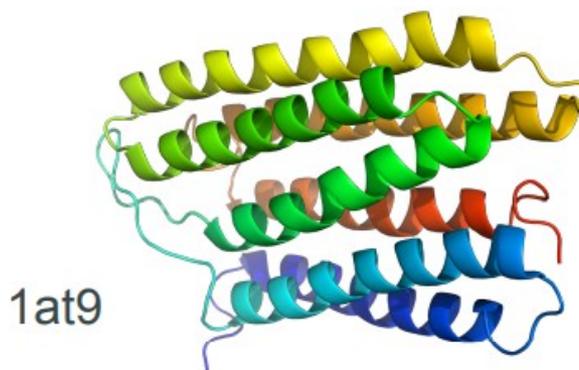
Analysis of individual structure



Analysis of a whole family

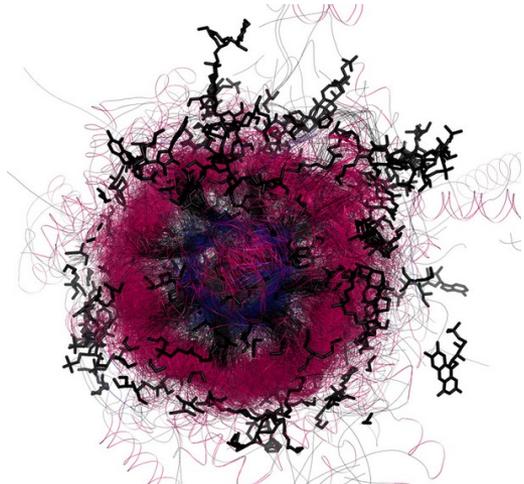
# Protein family structures and their analysis

- Comparison of protein family members
  - Different species
  - Different substituents
  - Mutations
  - Active and inactive forms
- Firm (conserved) and flexible regions
- Binding of ligands



# Protein family structures and their analysis

## How to do it?



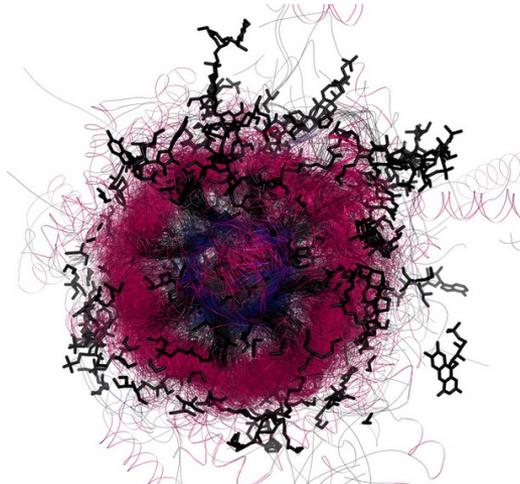
Aldolase class I  
(protein family 3.20.20.70)



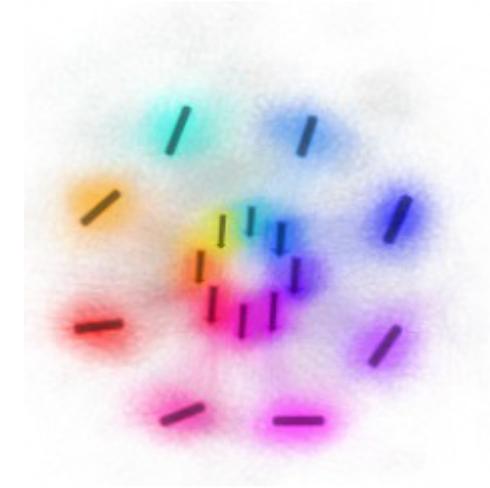
Cytochrome P450  
(protein family 1.10.630.10)

# Protein family structures and their analysis

## How to do it?



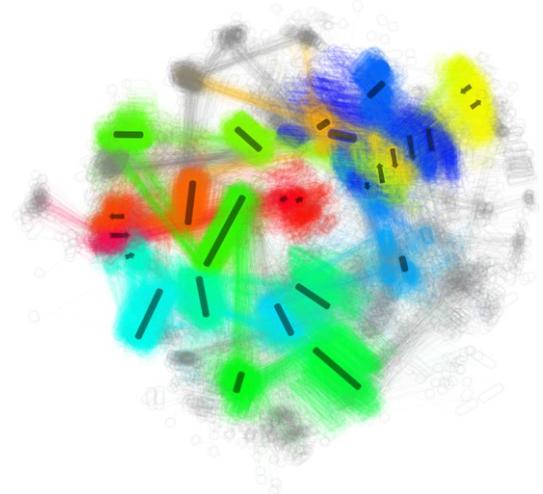
Aldolase class I  
(protein family 3.20.20.70)



**Insight into protein family:**  
Secondary structure  
2D diagrams



Cytochrome P450  
(protein family 1.10.630.10)



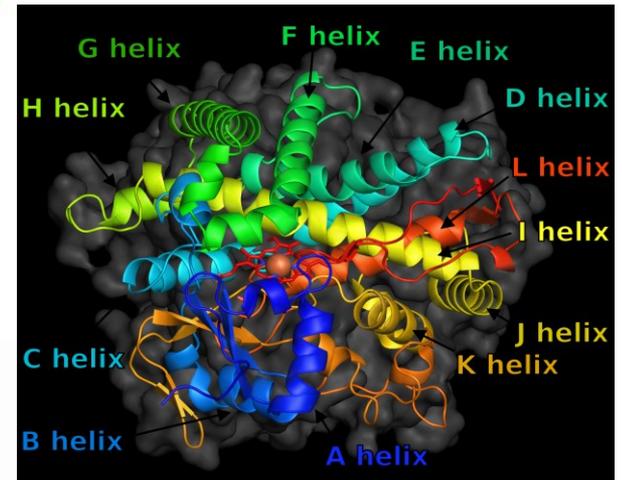
# Protein family structures and their analysis

## Secondary structure utilization – necessary steps

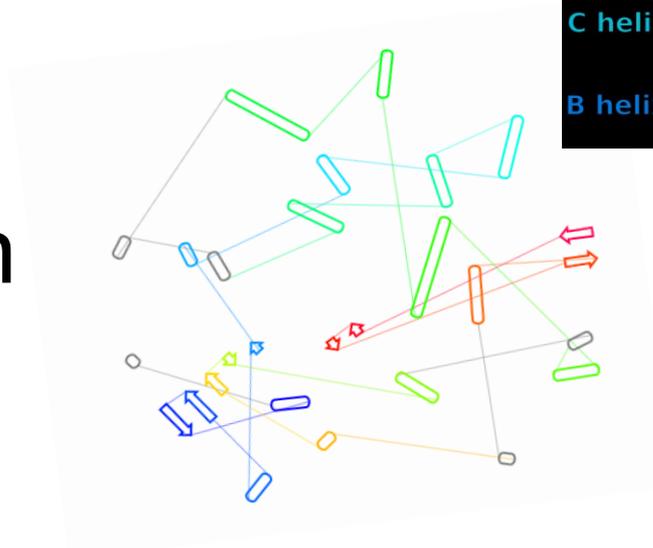
- Detection



- Annotation

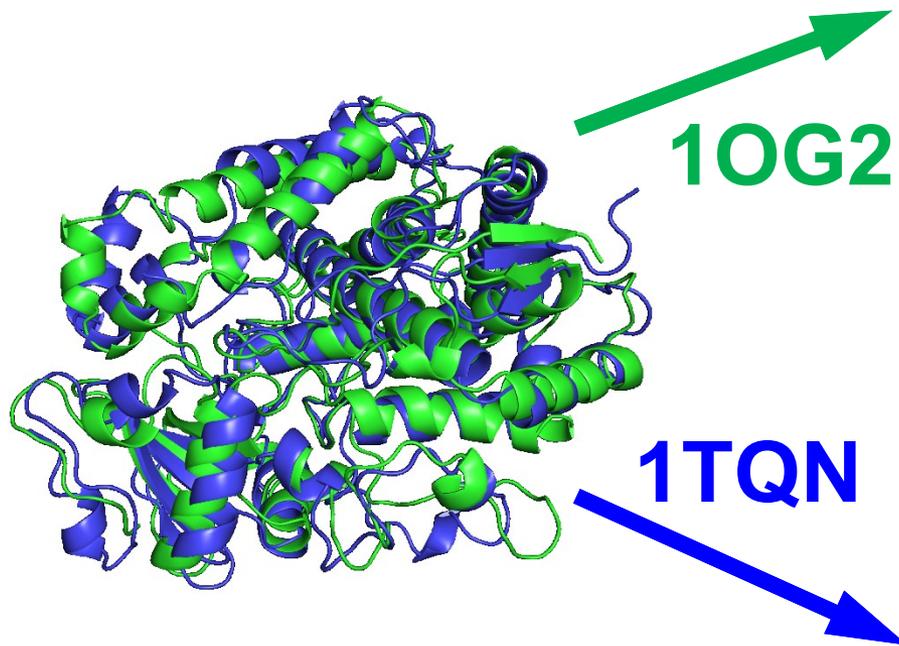


- Visualization



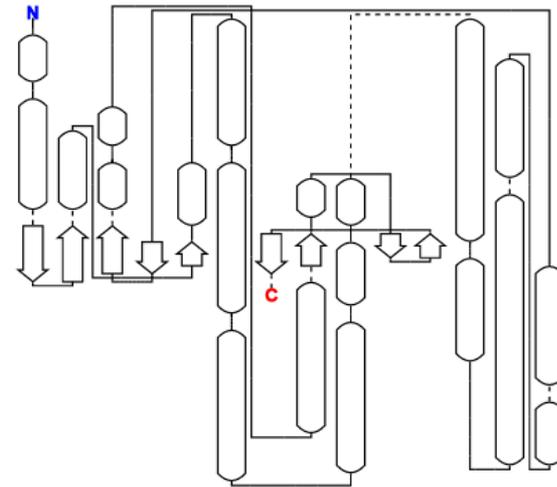
# Visualization of secondary structure in 2D: Solved in past? Not for protein families!

ISSUE 1: Similar proteins have  
different 2D diagrams

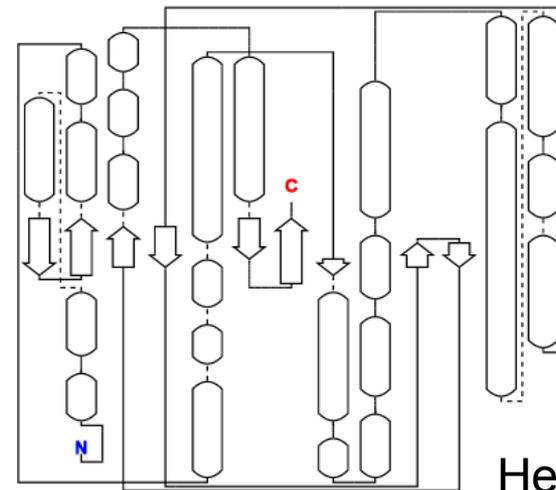


1OG2

1TQN



RMSD: 2.295 Å



Hera, PDBe

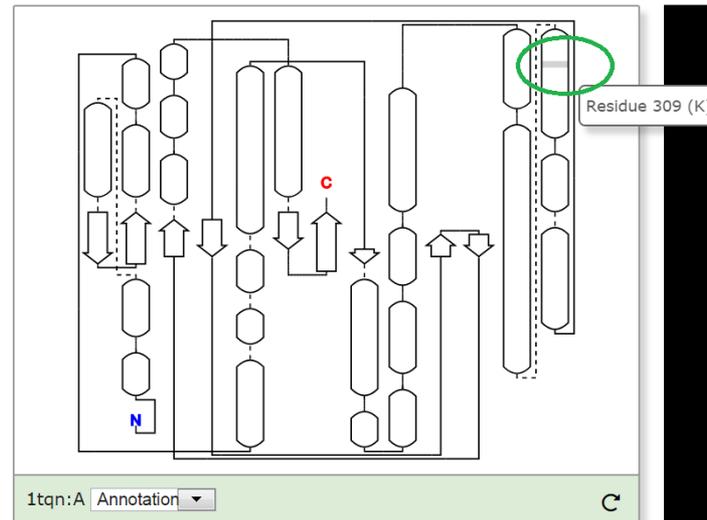
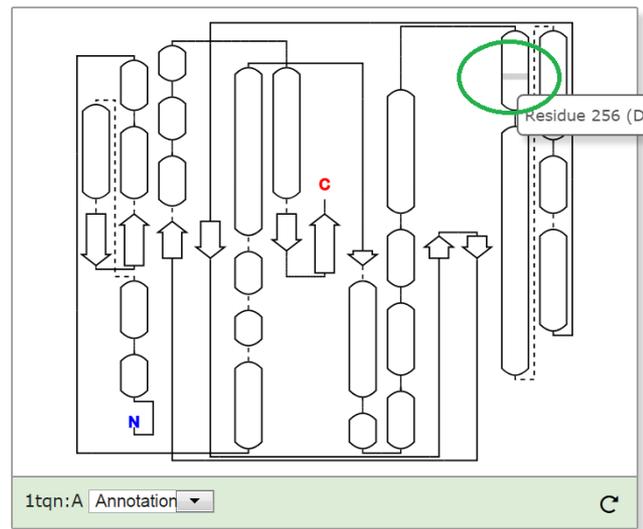
# Visualization of secondary structure in 2D: Solved in past? Not for protein families!

## ISSUE 2:

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

# 1TQN

Hera,  
PDBe

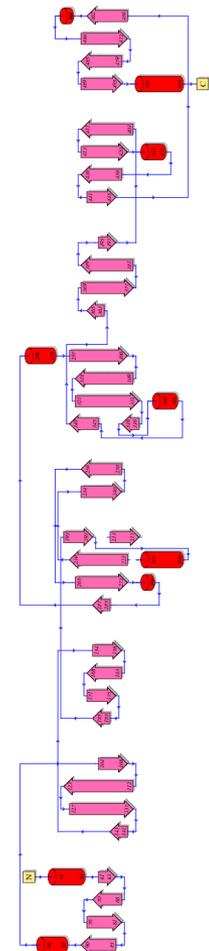


# Visualization of secondary structure in 2D: Solved in past?

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



**1ORW**



**HERA**

# Protein family based 2D diagrams

## How to get them?

Input:



### Step 1: Detection & annotation

- Find secondary structure elements (SSE)
- Annotate them

### Step 2: Statistics

- Average length of SSE
- Average occurrence of SSE

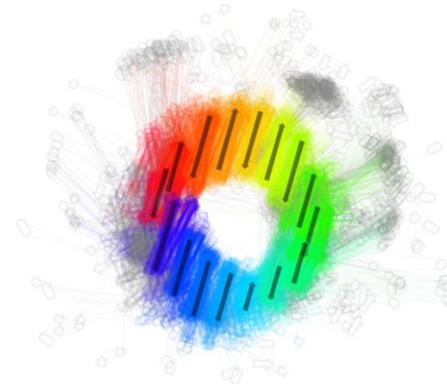
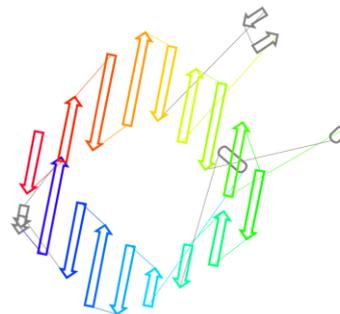
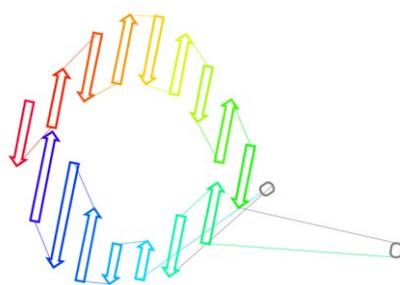
# Protein family based 2D diagrams

## How to get them?

### Step 3: Construct the 2D diagram

- Group all  $\beta$ -strands into sheets
- Divide the **helices and sheets into primary** (common for most of the domains) **and secondary** (the remaining ones).
- **Place all primary** helices and sheets into the 2D diagram.
- **Adjust the angles** of the primary helices and sheets.
- **Add all secondary** helices and sheets into the 2D diagram.
- **Adjust the angles** of the secondary helices and sheets.

### Step 4: Draw the 2D diagrams



# Protein family 2D diagrams

## 2DProts database

<https://2dprots.ncbr.muni.cz>

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

## Examples

Click headings below to expand:



Porin (protein family 2.40.160.10)

# Protein family 2D diagrams

## 2DProts database

2DProts

Custom entry

User manual

Description of methods

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

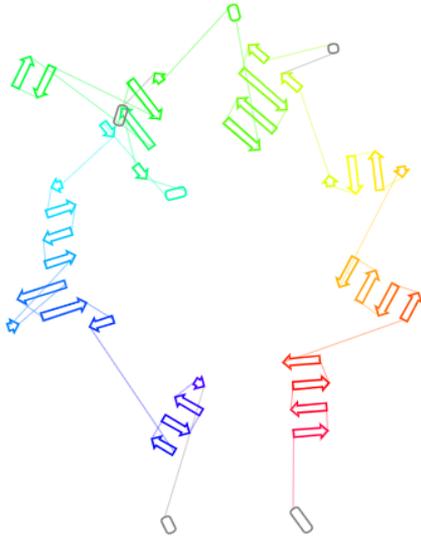
Search 2DProts

## Domain 1j2eA01

### 2D domain diagram

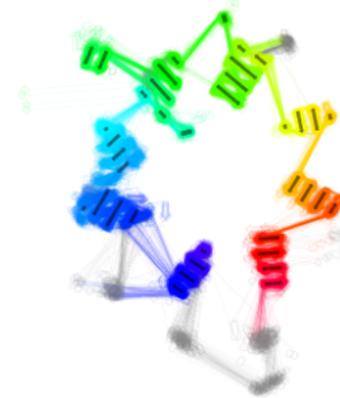
No ligands

With ligands



« no prev. | [1j2eB01](#) »

Member of  
family:  
[2.140.10.30](#)



Part of protein:  
[1j2e](#)

# Protein family 2D diagrams

## 2DProts database

### Family 2.140.10.30

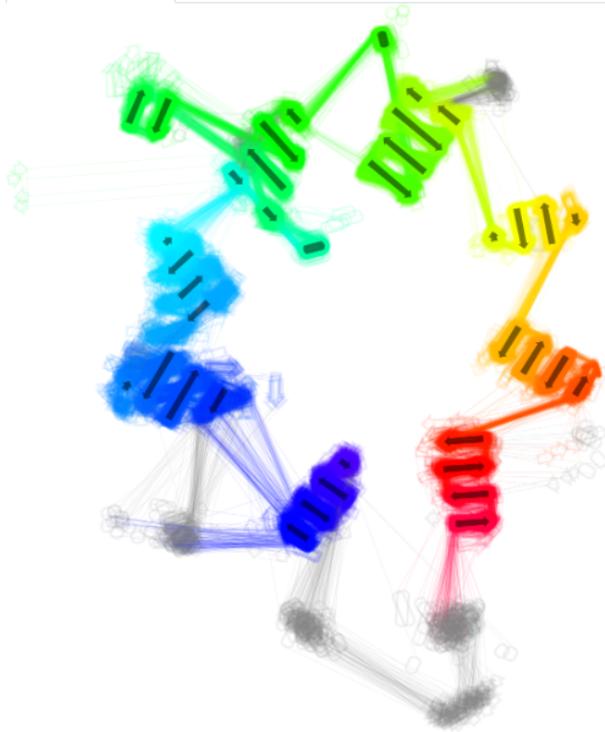
#### 2D multi diagram

No ligands

With ligands

No ligands, opaque

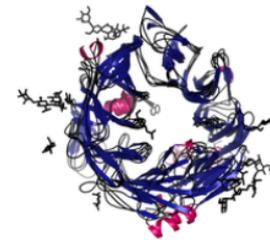
No ligands, opaque, no averages



Multi image for 2.140.10.30 ([svg source](#))

[« 2.140.10.20](#) | [2.150.10.10 »](#)

#### 3D model ([CATH](#))

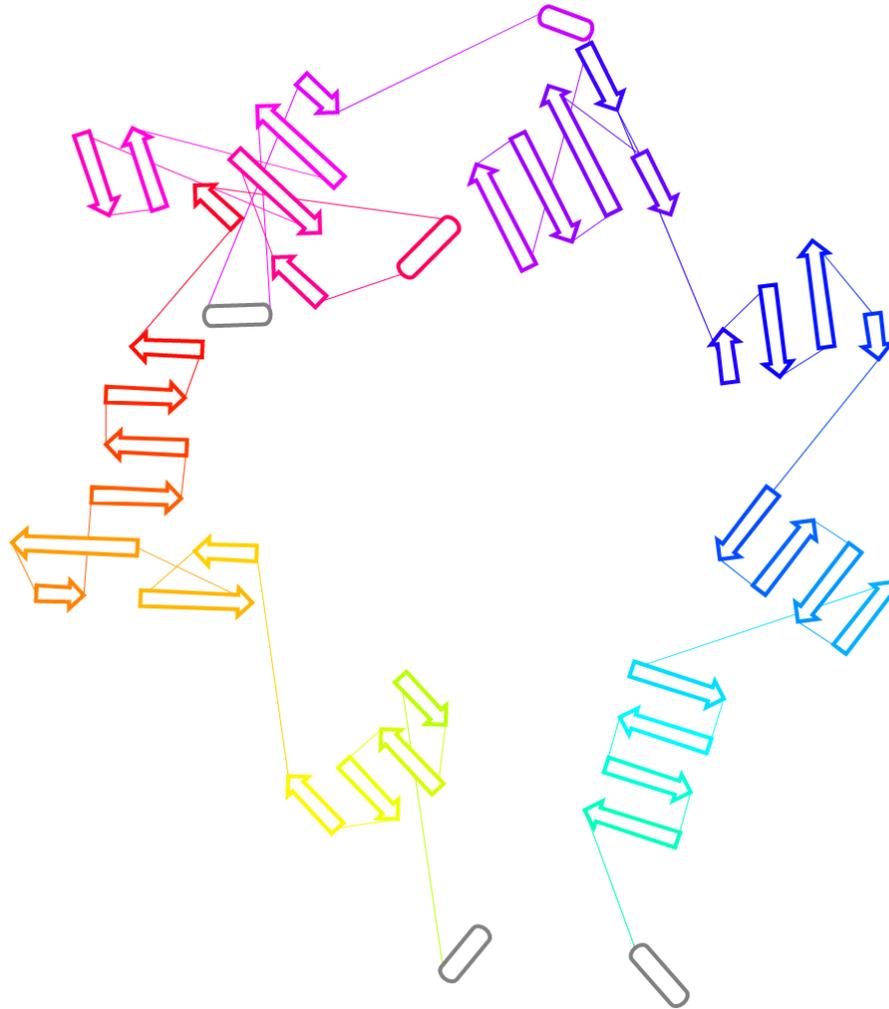


#### Domains (325)

- [1j2eA01](#)
- [1j2eB01](#)
- [1n1mA02](#)
- [1n1mB01](#)
- [1nu6A01](#)
- [1nu6B01](#)
- [1nu8A01](#)
- [1nu8B01](#)
- [1orvA01](#)

# 2DProts outputs

## 2D diagram of a protein domain



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



With opacity



No opacity

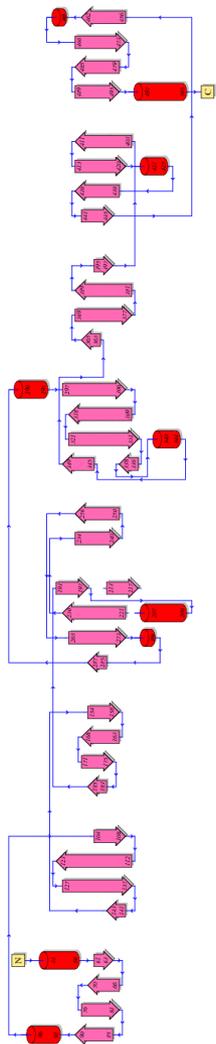
# Superfamily: Dipeptidylpeptidase IV (2.140.10.30)

**PROTEIN**

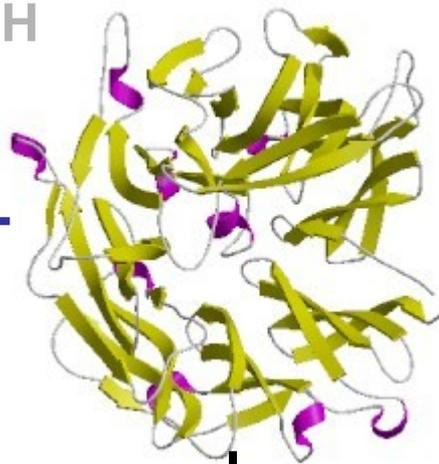
**PROTEIN FAMILY**

**Current solution**

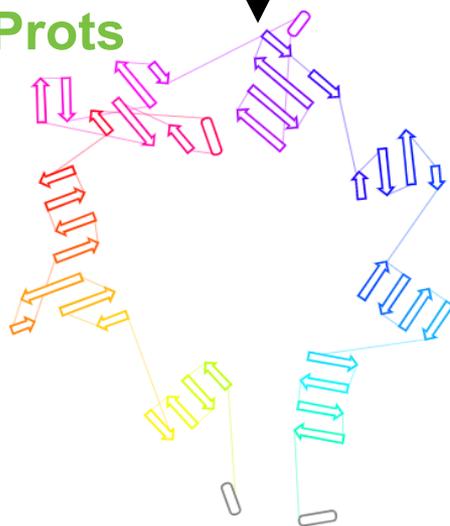
**HERA**



**CATH**



**2DProts**



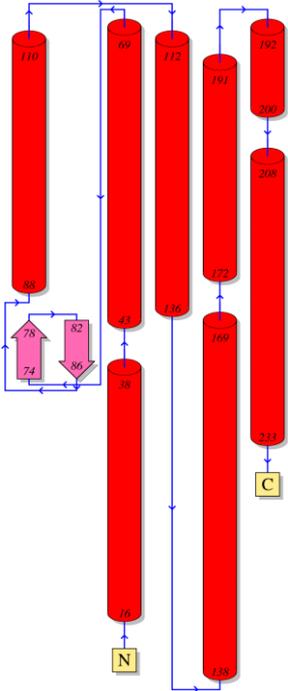
# Superfamily: Rhodopsin 7-helix transmembrane proteins

**PROTEIN** (1.20.1070.10)

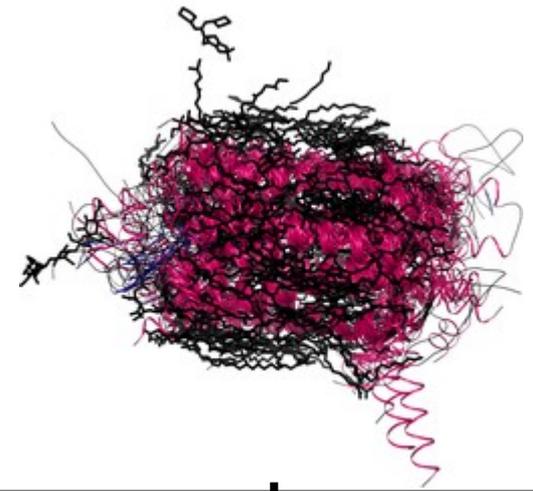
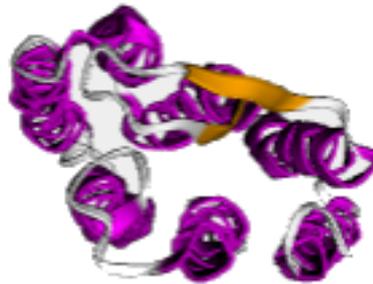
**PROTEIN FAMILY**

**Current solution**

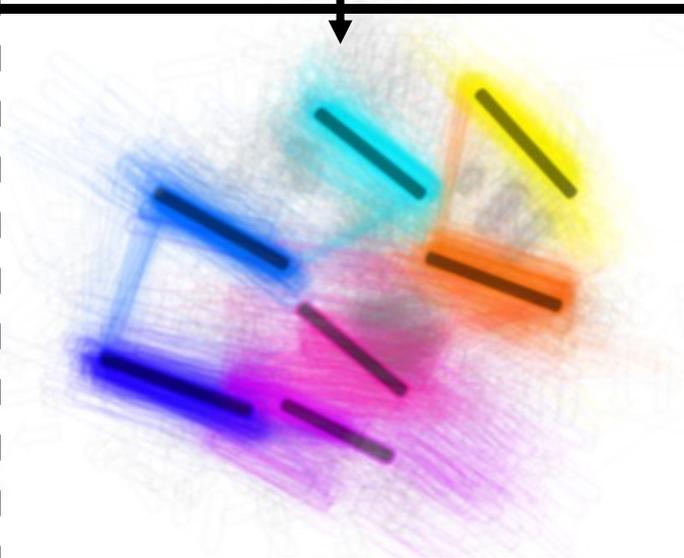
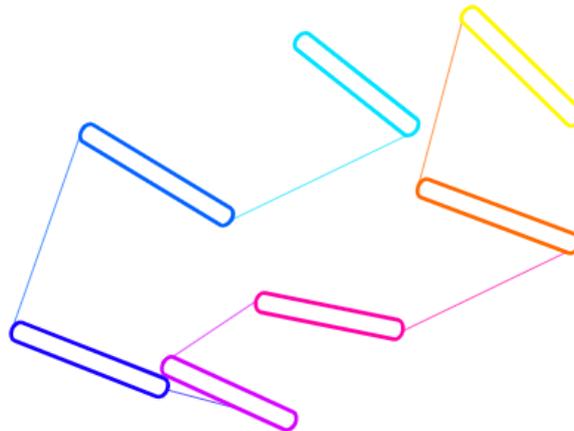
**HERA**



**CATH**



**2DProts**



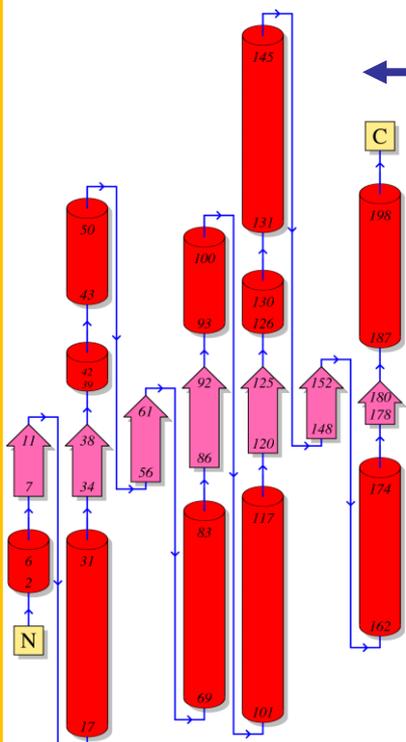
# Superfamily: Aldolase class I (3.20.20.70)

**PROTEIN**

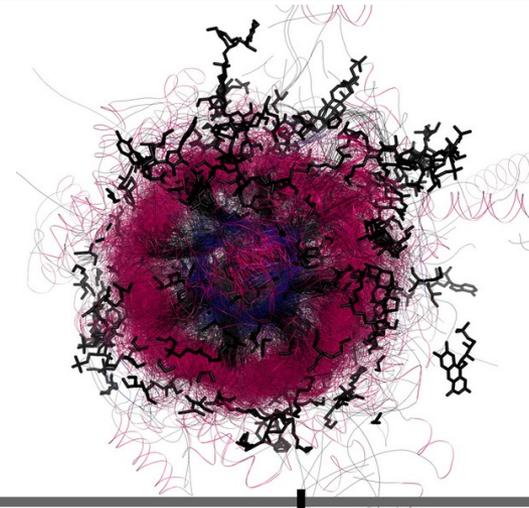
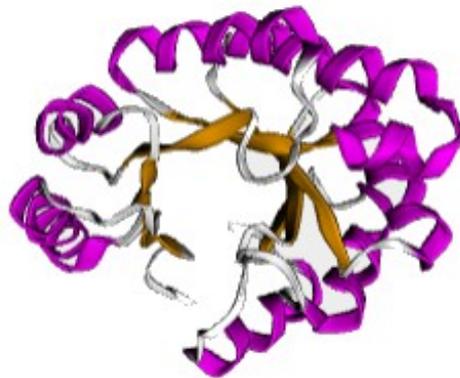
**PROTEIN FAMILY**

Current solution

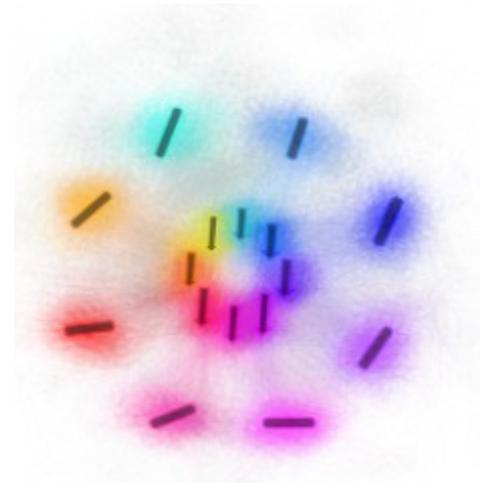
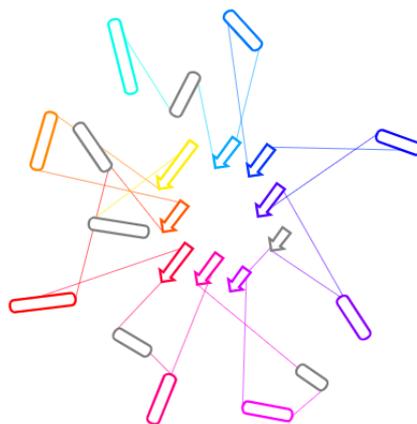
**HERA**



**CATH**



**2DProts**



# 2DProts integration to CATH

The screenshot shows the CATH database interface for Superfamily 1.10.60.10. The page title is "CATH Superfamily 1.10.60.10" with a subtitle "Iron dependent repressor, metal binding and dimerisation domain". A search bar and navigation menu are at the top. The main content area features a "SUPERFAMILY LINKS" sidebar with options like "Summary", "Superfamily Superposition", "Classification / Domains", "Functional Families", and "Structural Neighbourhood". The "Functional Families" section includes a tree diagram with a red circle at the root and branches for "Diphtheria t Transcriptio", "Transcriptio", "Iron (Metal)", and "Manganese". The main content area has tabs for "SS", "Alignbow", "2DProts", and "2DProts". A large 3D superposition figure is displayed, showing multiple protein structures in different colors (purple, yellow, red) overlaid on a grey wireframe. A green button "Visit 2DProts (1.10.60.10)" is visible. A blue text box at the bottom states: "These superposition figures provide an indication of the relative distance and position of secondary structure elements within CATH superfamilies. Image are generated by the 2DProts database".

CATH Superfamily 1.10.60.10

Iron dependent repressor, metal binding and dimerisation domain

Home / Superfamily 1.10.60.10

SUPERFAMILY LINKS

- Summary
- Superfamily Superposition**
- Classification / Domains
- Functional Families
- Structural Neighbourhood

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
  - Diphtheria t Transcriptio
  - Transcriptio
  - Iron (Metal)
  - Manganese

SS Alignbow 2DProts 2DProts

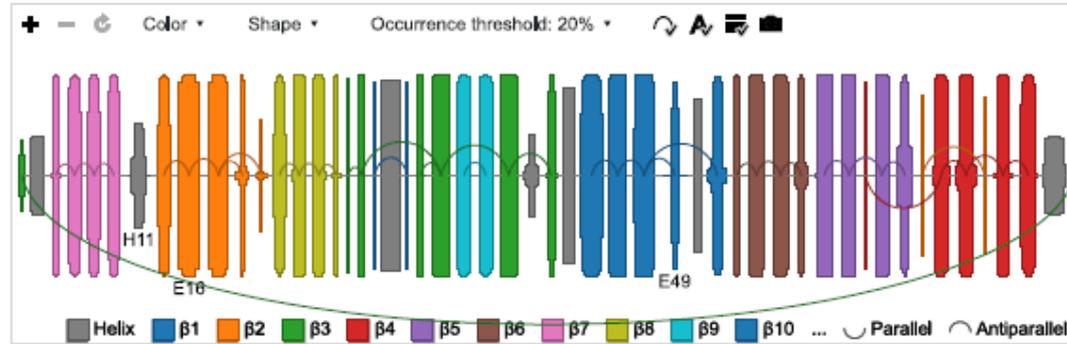
Visit 2DProts (1.10.60.10)

These superposition figures provide an indication of the relative distance and position of secondary structure elements within CATH superfamilies. Image are generated by the 2DProts database

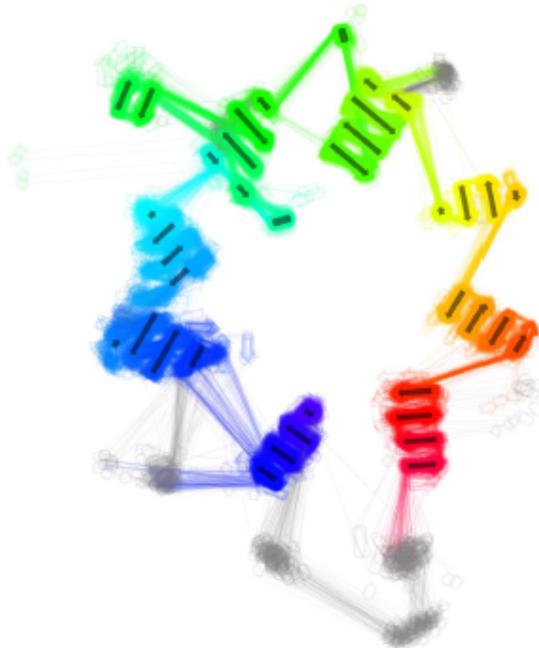
# 2DProts integration into OverProt

<https://overprot.ncbr.muni.cz>

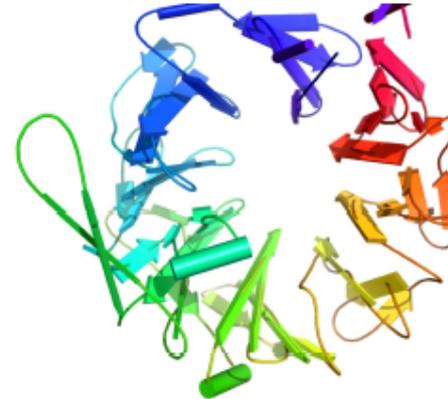
Family: 2.140.10.30 *Dipeptidylpeptidase IV, N-terminal domain*



2D view (2DProts)



3D view (MAPSCI + OverProt)



# 2DProts integration into OverProt

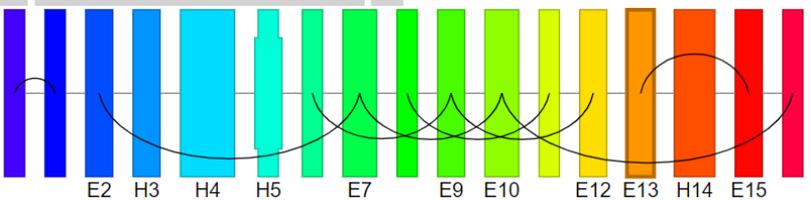
<https://overprot.ncbr.muni.cz>

Integrated Viewer

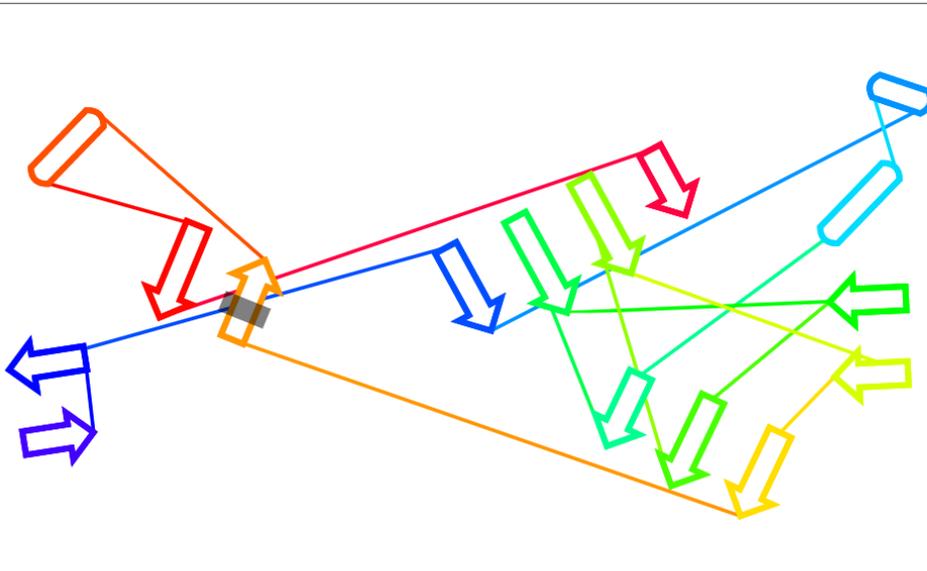
Current family (CATH): 2.160.10.30  
Current domain: 3ogzA02

Protein family (CATH): 2.160.10.30  
Protein domain: 3ogzA02 Submit

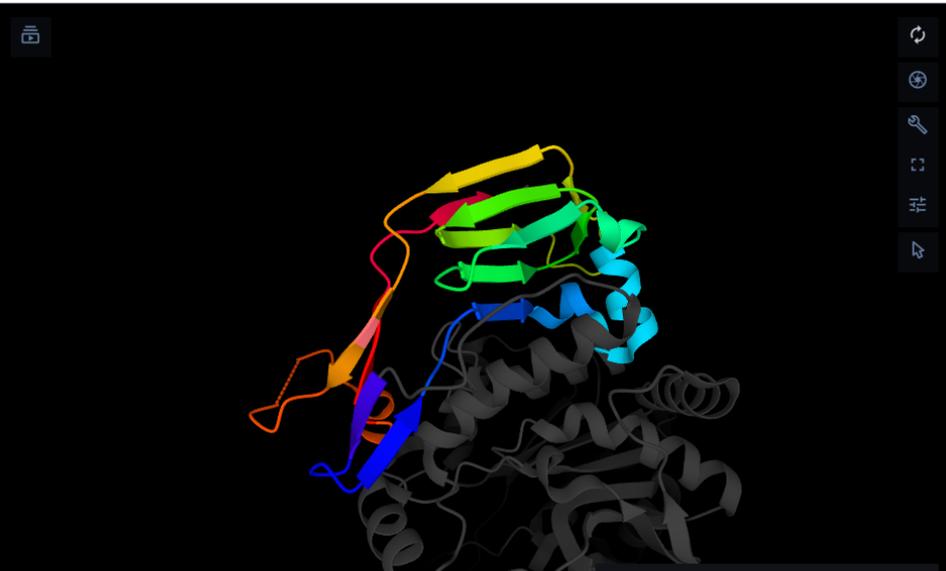
+ - ↺ Color ▾ Shape ▾ Beta-connectivity ▾ Occurrence threshold: 25% ▾ 📷



E2 H3 H4 H5 E7 E9 E10 E12 E13 H14 E15



3ogz | Entity 1 | Chain A Annotation ▾



UDP-sugar pyrophosphorylase  
3OGZ | Model 1 | Instance\_1\_555 | A | VAL 575

# 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**, 37(23), 4599-4601.

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

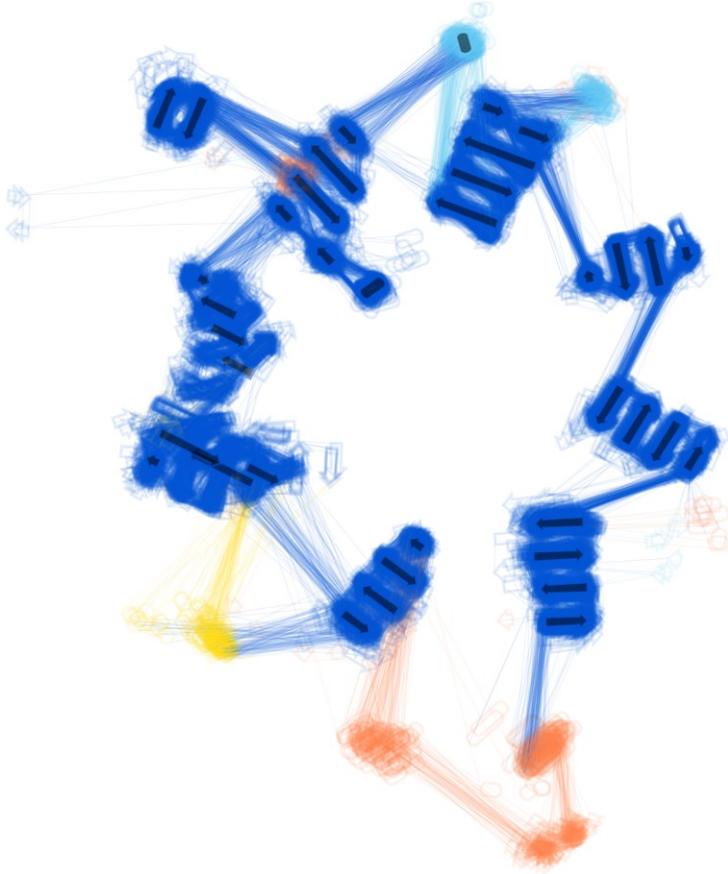
**Open Access**

No barriers to access – all articles freely available online

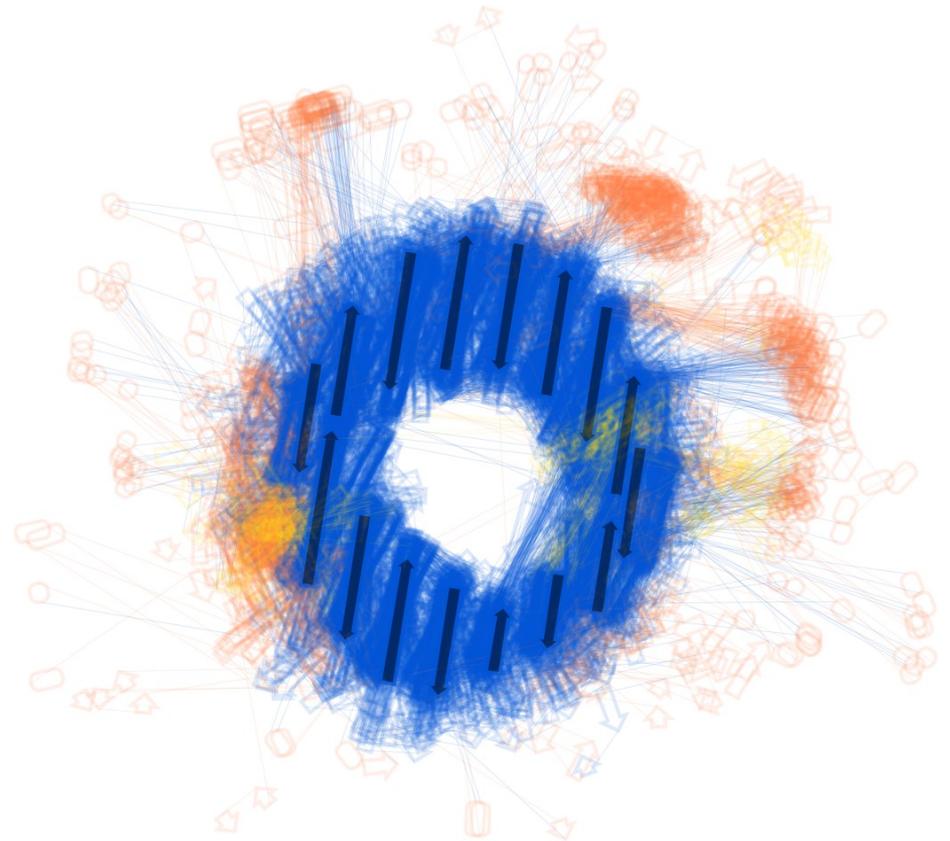


# 2DProts: Coloring by structure properties

## Example: Occurrence of secondary structures

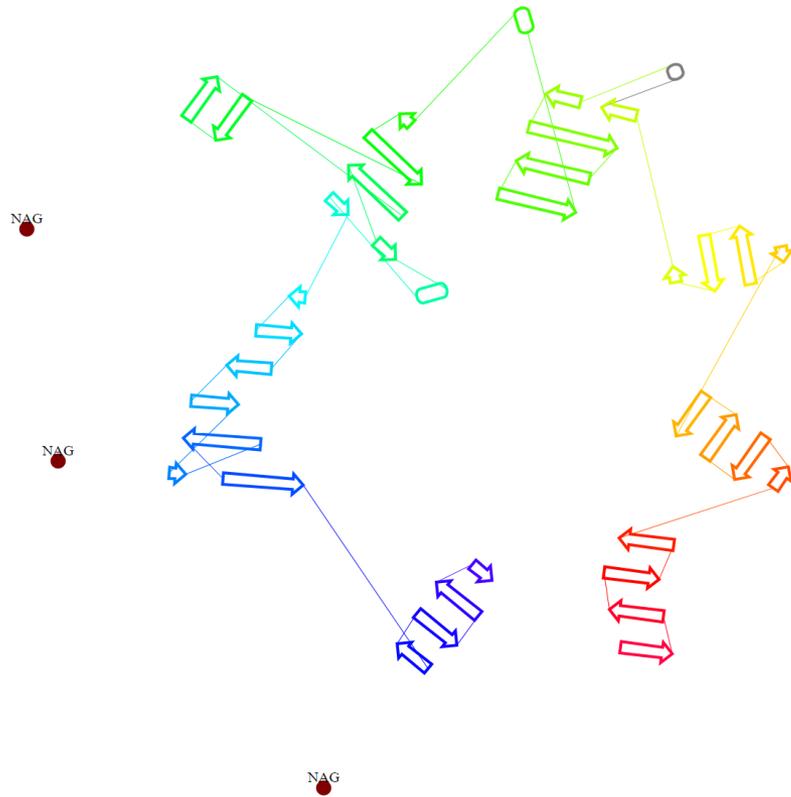


**Cytochrome reductase,  
Family 2.140.10.30**

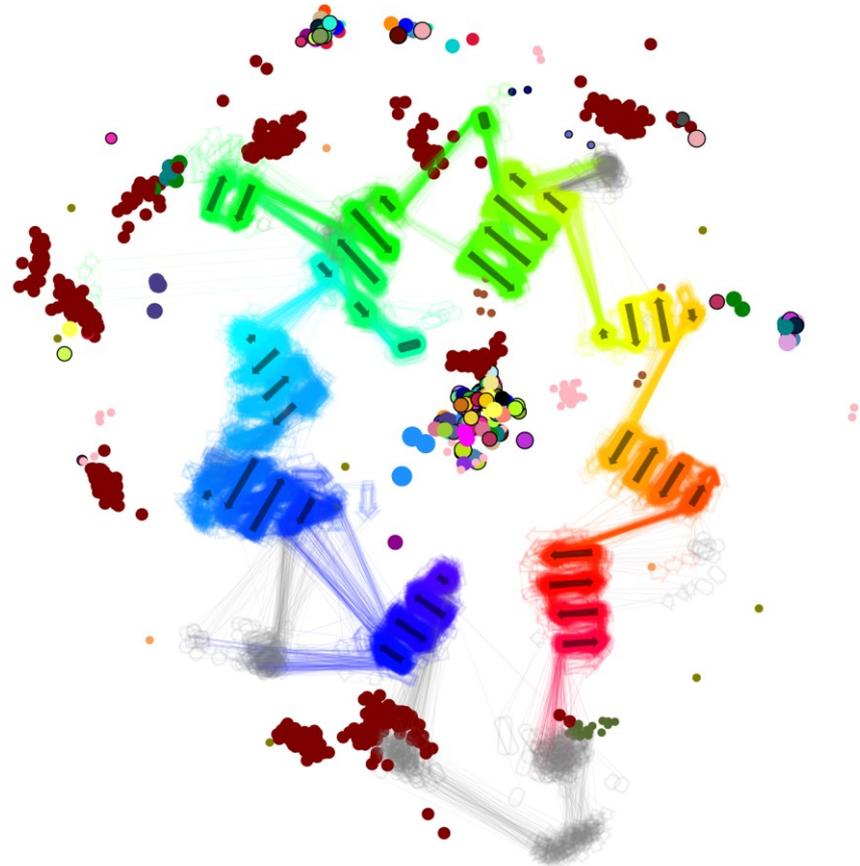


**Porin  
Family 2.40.160.10**

# 2DProts: Integration of ligands



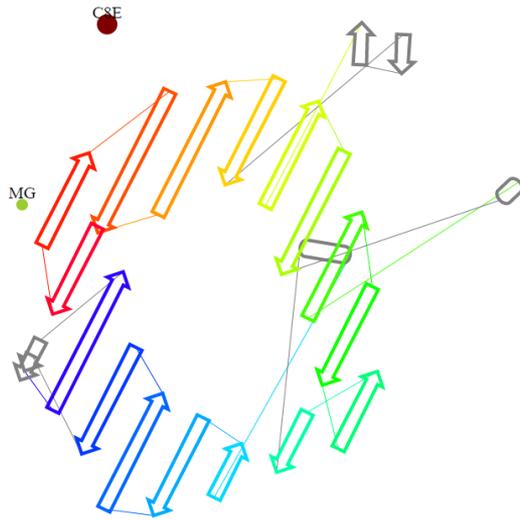
**PDB ID 2bgn,  
domain A00**



**Cytochrome reductase, family 2.140.10.30**

● ...NAG (742 pcs)	● ...SY1 (8 pcs)	● ...T22 (4 pcs)	● ...RUF (4 pcs)	● ...P2Y (4 pcs)	● ...0WG (4 pcs)	● ...10T (4 pcs)	● ...B2Q (2 pcs)
● ...S04 (56 pcs)	● ...SC3 (8 pcs)	● ...6RL (4 pcs)	● ...B2Y (4 pcs)	● ...8O3 (4 pcs)	● ...0QG (4 pcs)	● ...008 (4 pcs)	● ...P54 (2 pcs)
● ...NA (20 pcs)	● ...6Z8 (8 pcs)	● ...8OL (4 pcs)	● ...7AC (4 pcs)	● ...BPR (4 pcs)	● ...AES (4 pcs)	● ...01T (4 pcs)	● ...677 (2 pcs)
● ...HG (10 pcs)	● ...XIH (5 pcs)	● ...8VU (4 pcs)	● ...LF7 (4 pcs)	● ...007 (4 pcs)	● ...13Z (4 pcs)	● ...LUI (4 pcs)	● ...277 (2 pcs)
● ...EDO (9 pcs)	● ...715 (4 pcs)	● ...RUM (4 pcs)	● ...PHI (4 pcs)	● ...PEG (4 pcs)	● ...9K4 (4 pcs)	● ...75L (2 pcs)	● ...D3C (2 pcs)

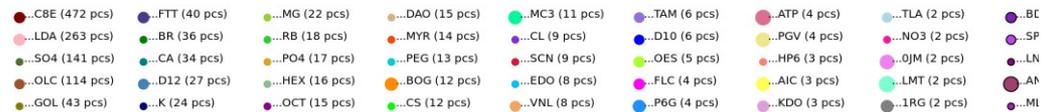
# 2DProts: Integration of ligands



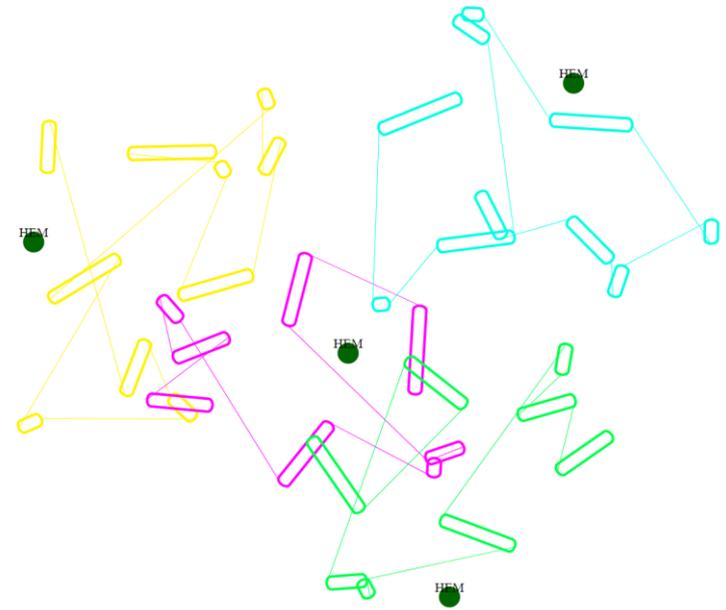
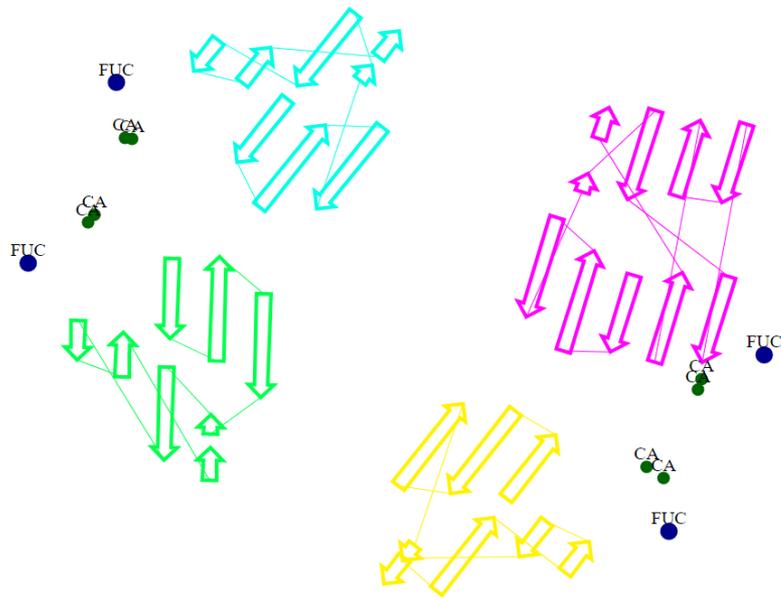
**OMPF Porin**  
**PDB ID 2zfg,**  
**domain A00**



**Porin, Family 2.40.160.10**



# 2DProts: 2D diagrams for proteins



**Pseudomonas  
aeruginosa lectin II  
PDB ID 1gzt**

**Hemoglobine  
PDB ID 1v4w**

# 2DProts: Integration of AlphaFoldDB

AlphaFold Protein Structure Database

Home About FAQs Downloads

# AlphaFold Protein Structure Database

Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism BETA Search

Examples: Free fatty acid receptor 2 At1g58602 Q5VSL9 E. coli Help: AlphaFold DB search help

Feedback on structure: Contact DeepMind

# 2DProts: Integration of AlphaFoldDB

AlphaFold Protein Structure Database

EMBL-EBI home Services Research Training About us EMBL-EBI

AlphaFold Protein Structure Database

Home About FAQs Downloads

AlphaFold DB provides open access to over 200 million protein structure predictions to accelerate scientific research.

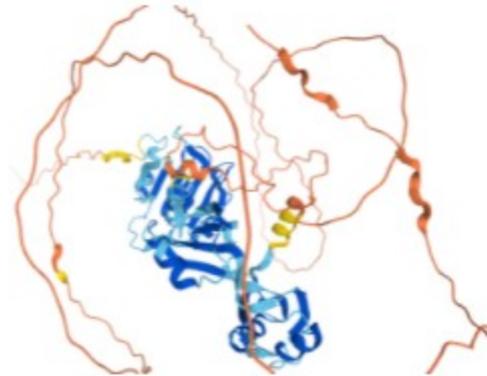
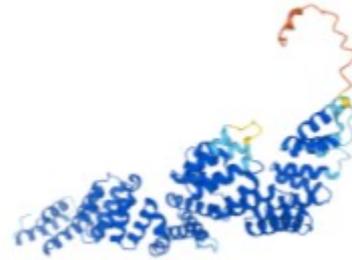
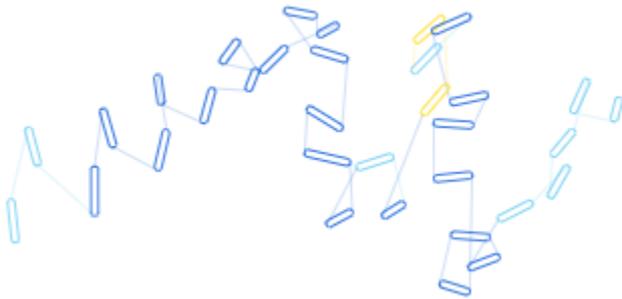
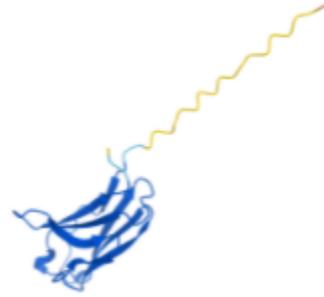
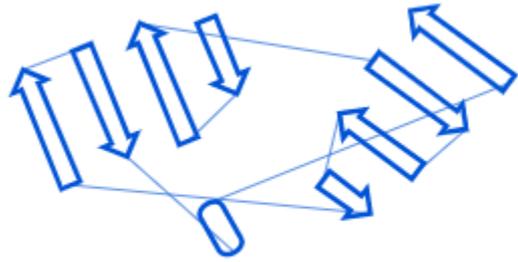
Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism BETA Search

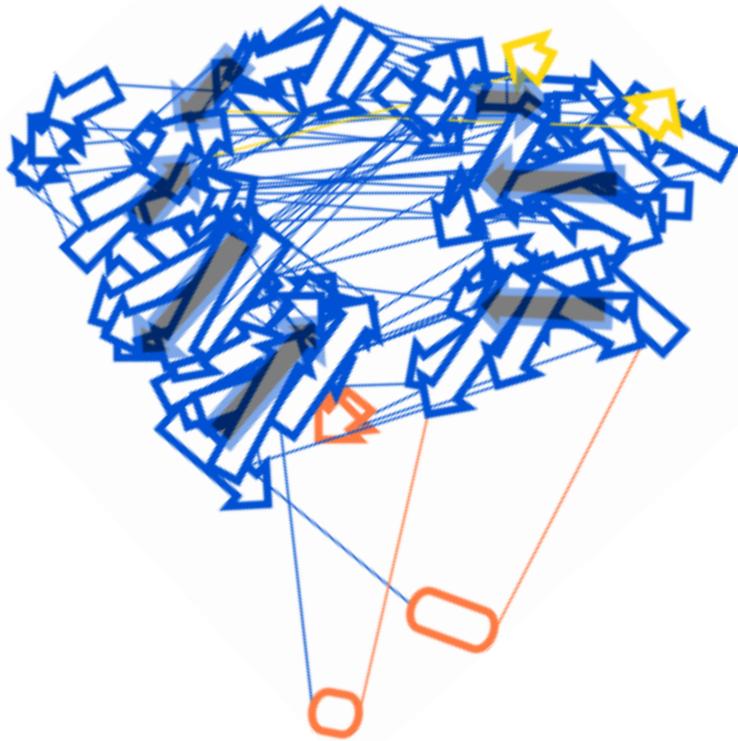
Examples: Free fatty acid receptor 2 At1g58602 Q5VSL9 E. coli Help: AlphaFold DB search help

Feedback on structure: Contact DeepMind

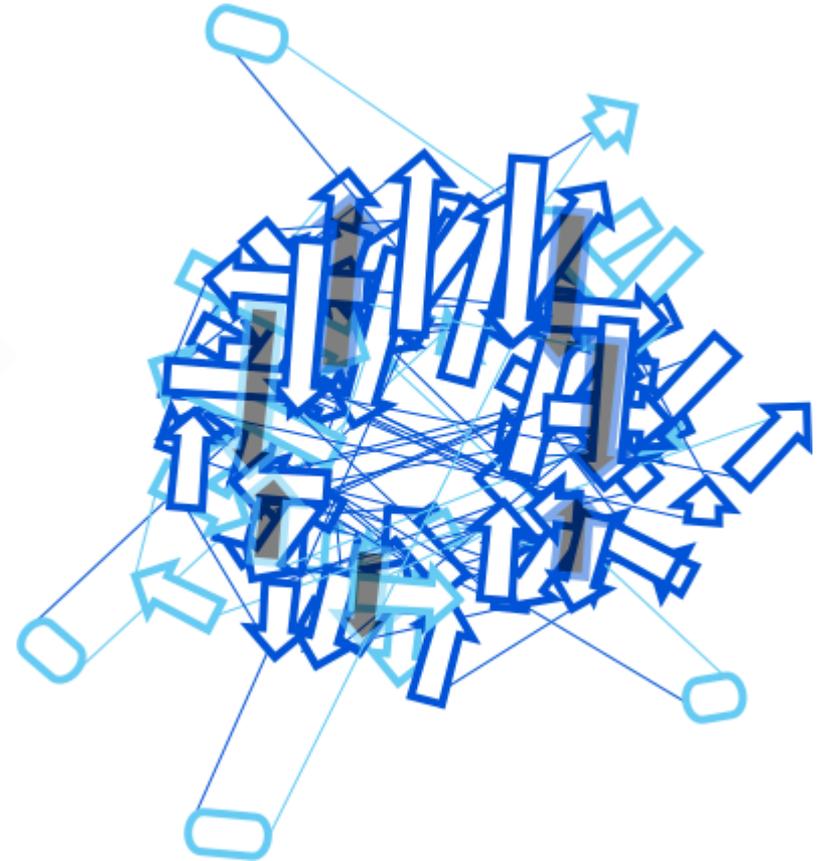
# 2DProts: Integration of AlphaFoldDB



# 2DProts: Integration of AlphaFoldDB



**Structures  
from PDB**



**Structures from  
AlphaFoldDB**

**E. coli PapC protein, C-terminal domain  
Family 2.60.40.2070**



## 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

