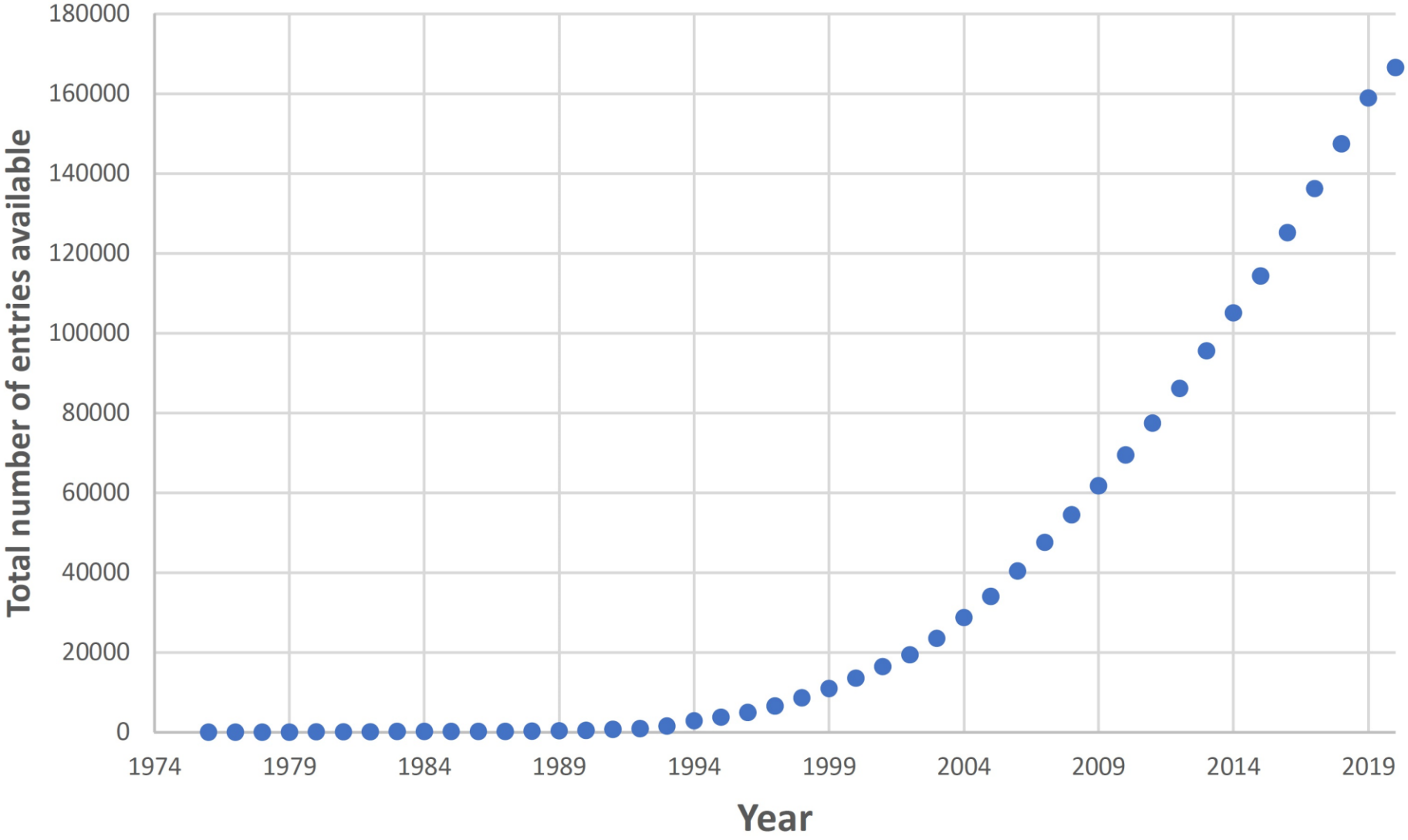
Secondary structure diagrams of proteins, protein families and ligands

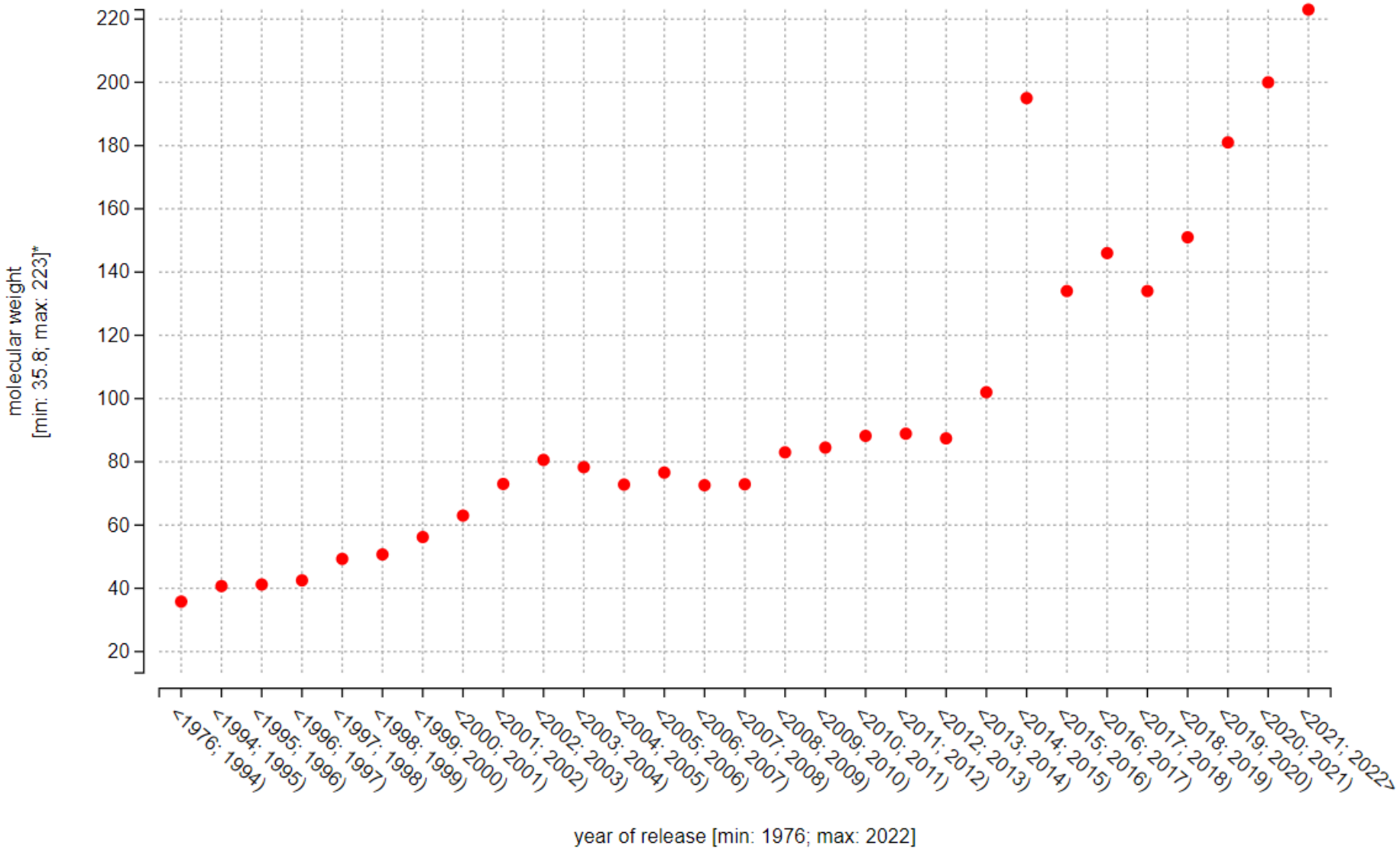
Radka Svobodová

NCBR, CEITEC
MASARYK UNIVERSITY

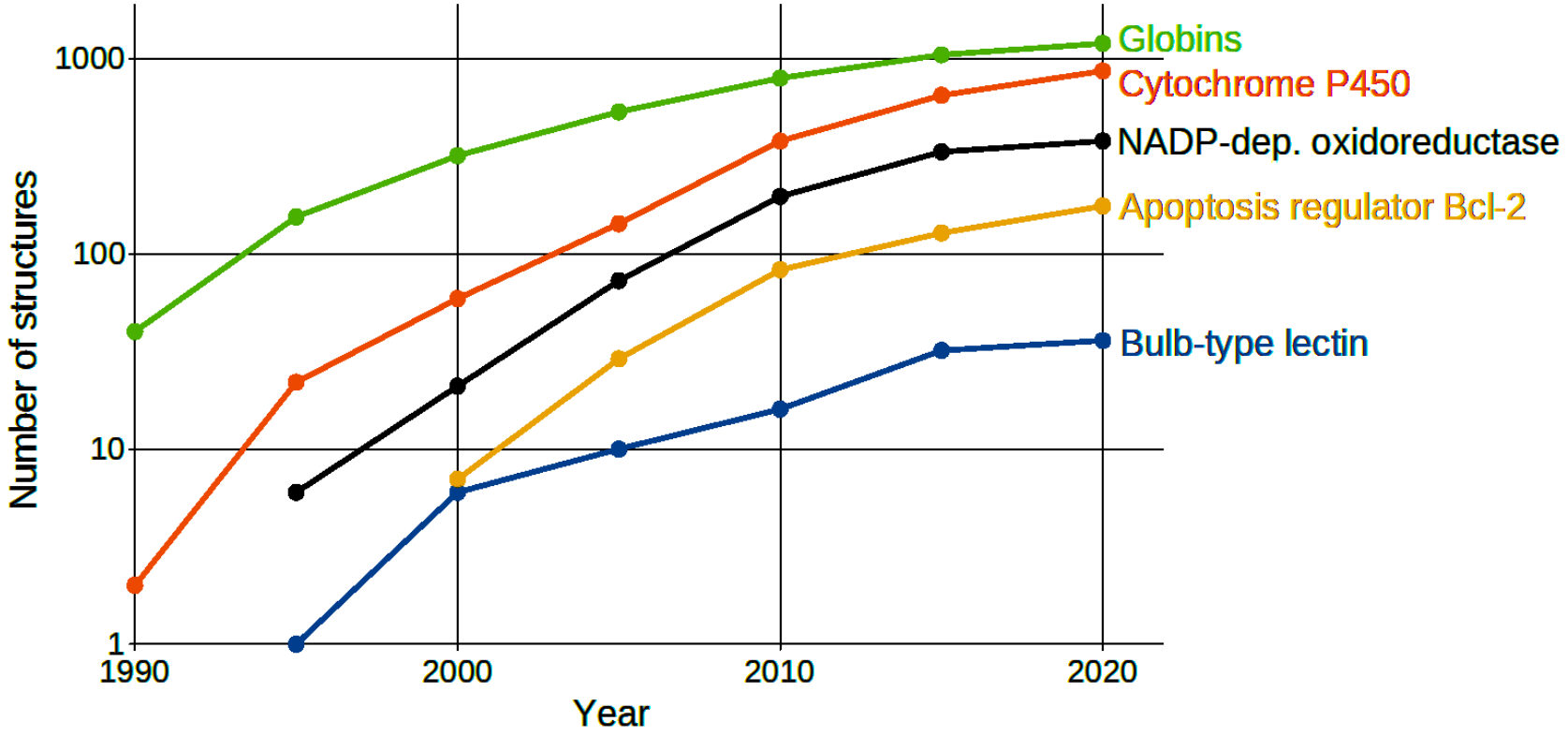
Current trends: Number of available structures grows



Current trends: Size of deposited structures also grows



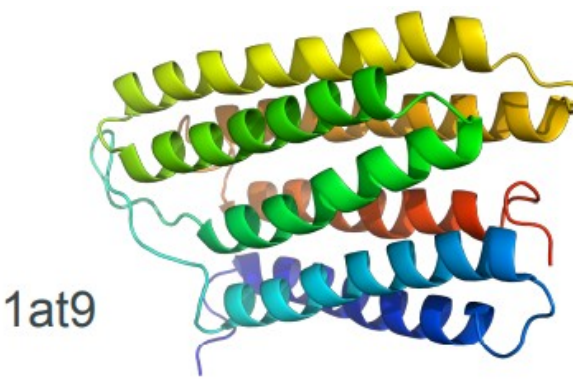
Current trends: Protein families are getting bigger

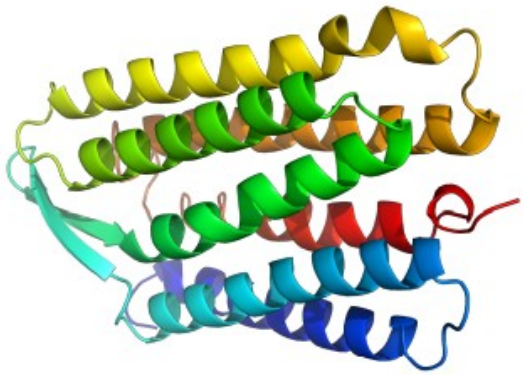


Analysis of individual structure → Analysis of a whole family

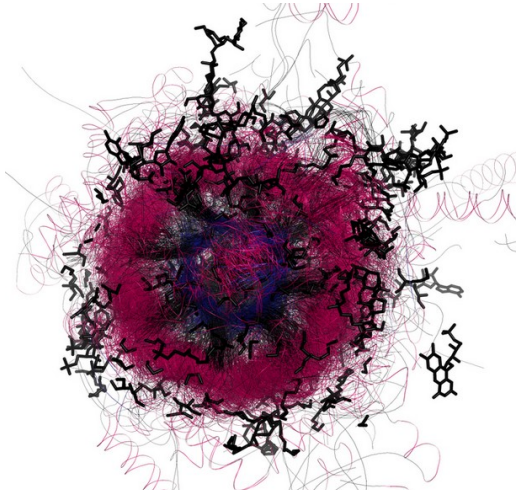
Protein family structures and their analysis

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





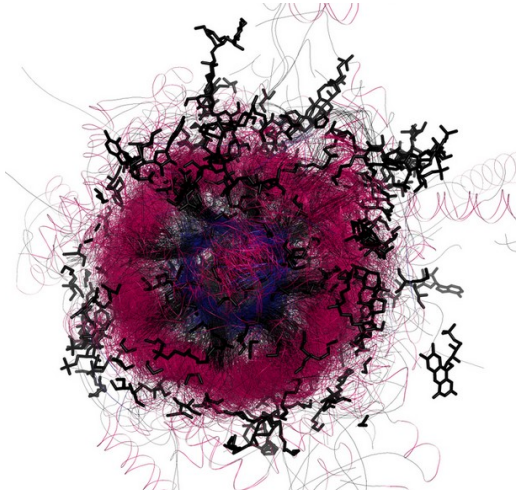
Protein family structures and their analysis
How to do it?



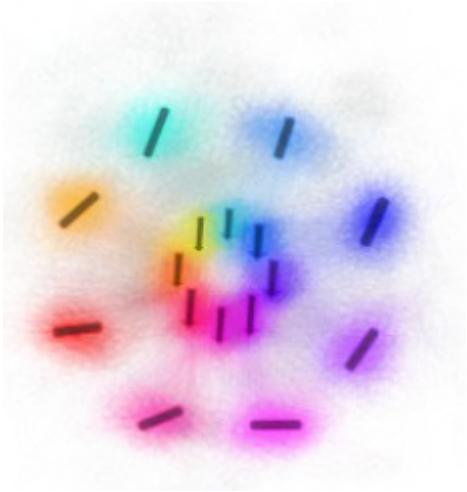
Aldolase class I
(protein family 3.20.20.70)

Cytochrome P450
(protein family 1.10.630.10)

Protein family structures and their analysis
How to do it?

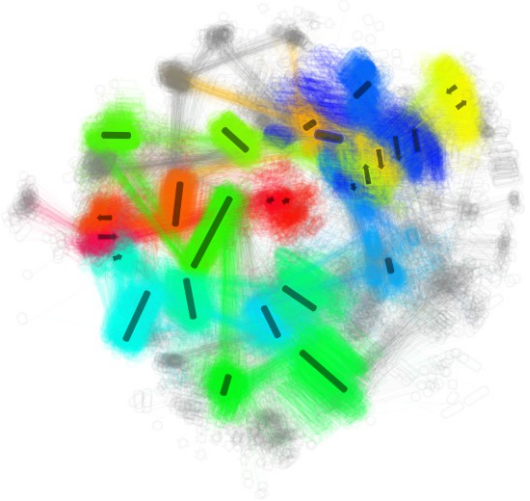


Aldolase class I
(protein family 3.20.20.70)



Insight into protein family:
Secondary structure
2D diagrams

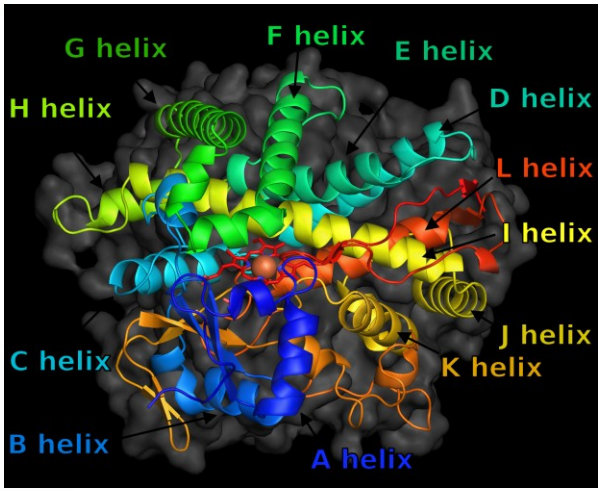
Cytochrome P450
(protein family 1.10.630.10)



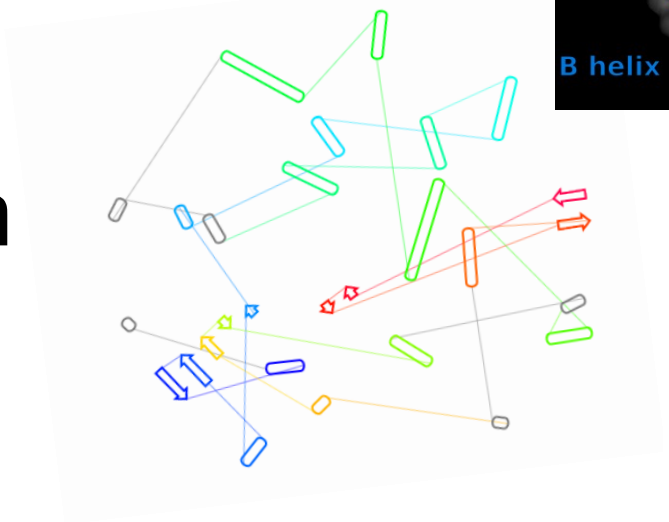
Protein family structures and their analysis
Secondary structure utilization – necessary steps

- Detection

- Annotation

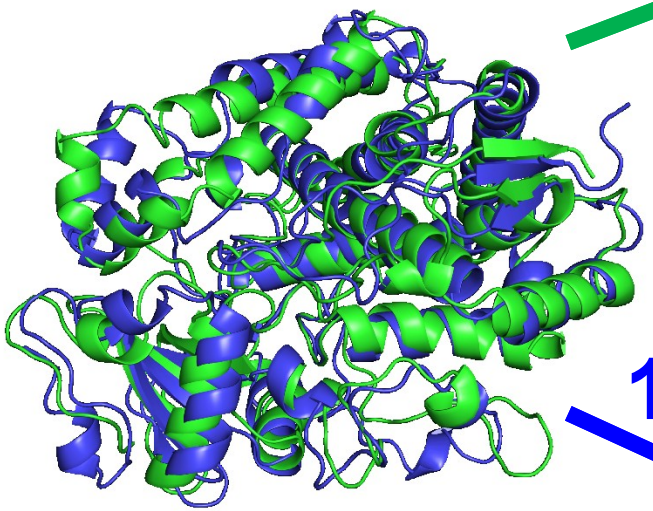


- Visualization

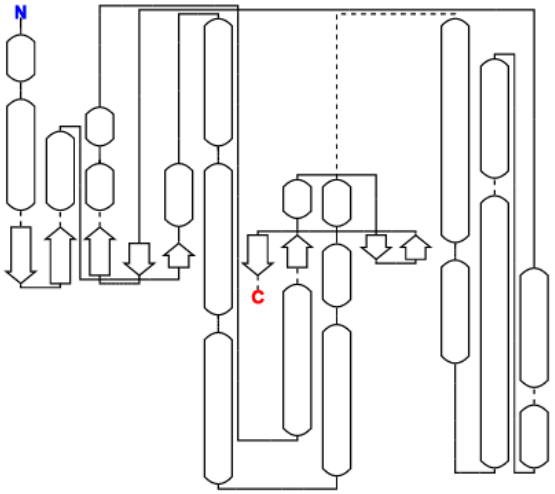


Visualization of secondary structure in 2D: Solved in past? Not for protein families!

ISSUE 1: Similar proteins have different 2D diagrams

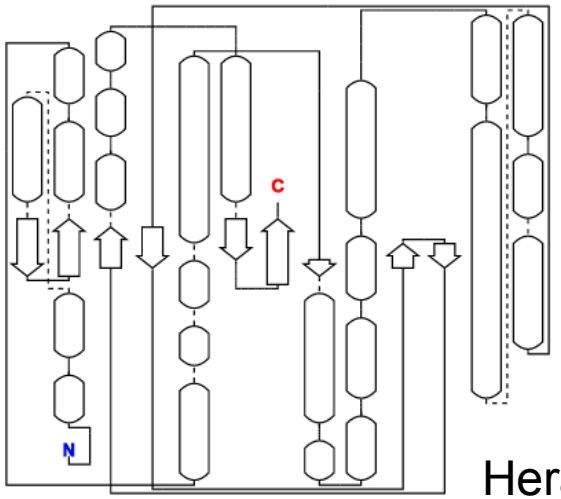


1OG2



RMSD: 2.295 Å

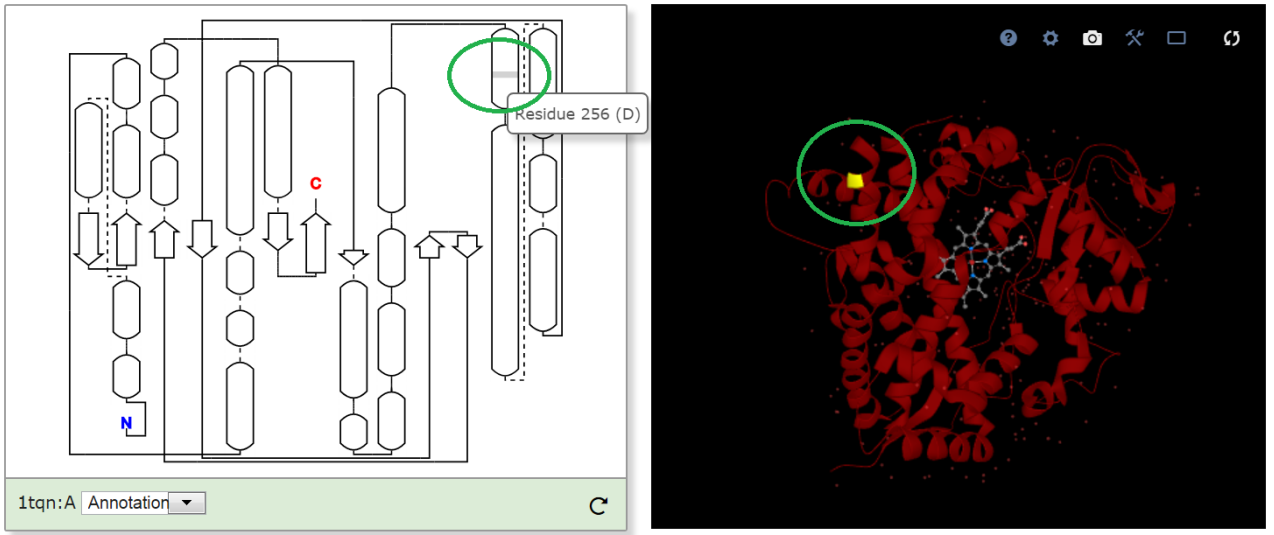
1TQN



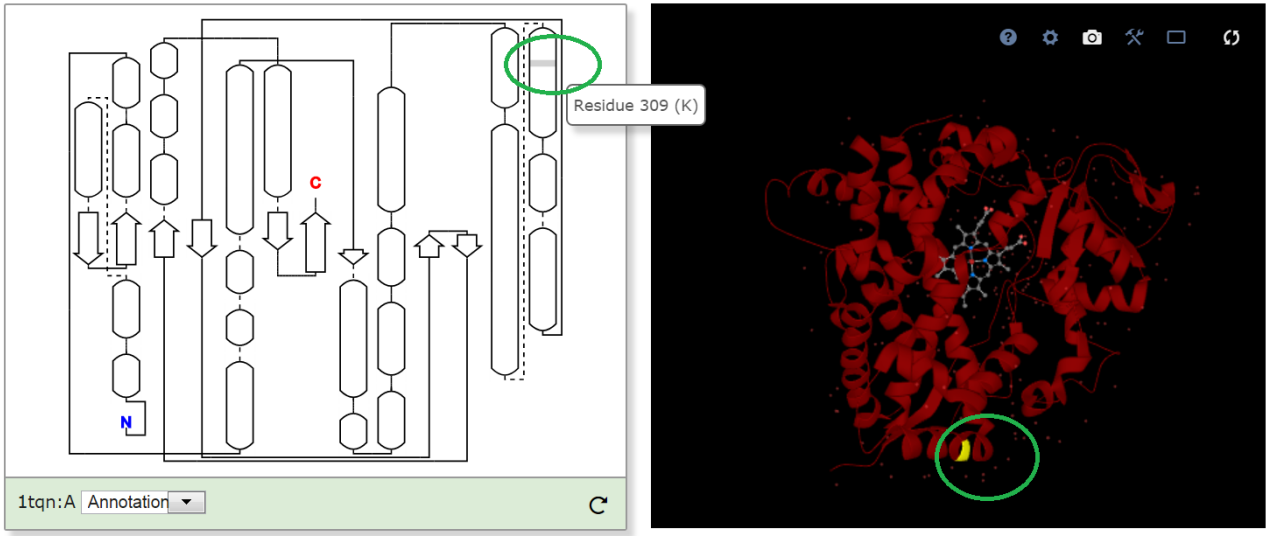
Hera, PDBe

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



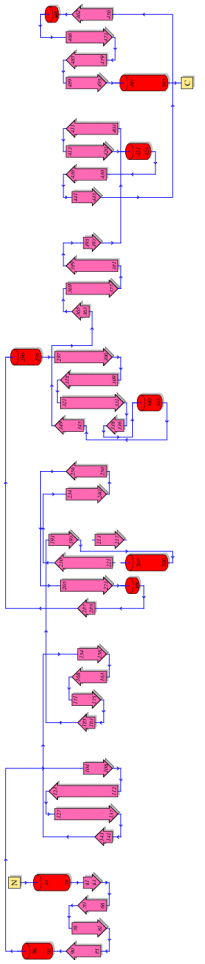
1TQN



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

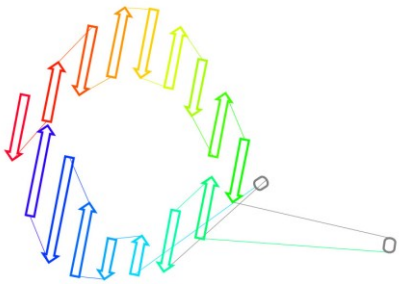
- Average length of SSE
- Average occurence of SSE

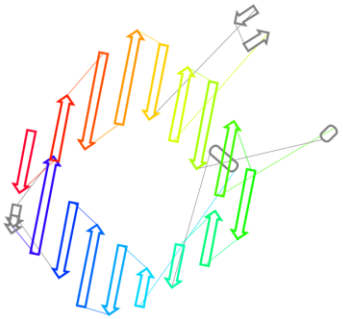
Protein family based 2D diagrams
How to get them?

Step 3: Construct the 2D diagram

- Group all β**-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.

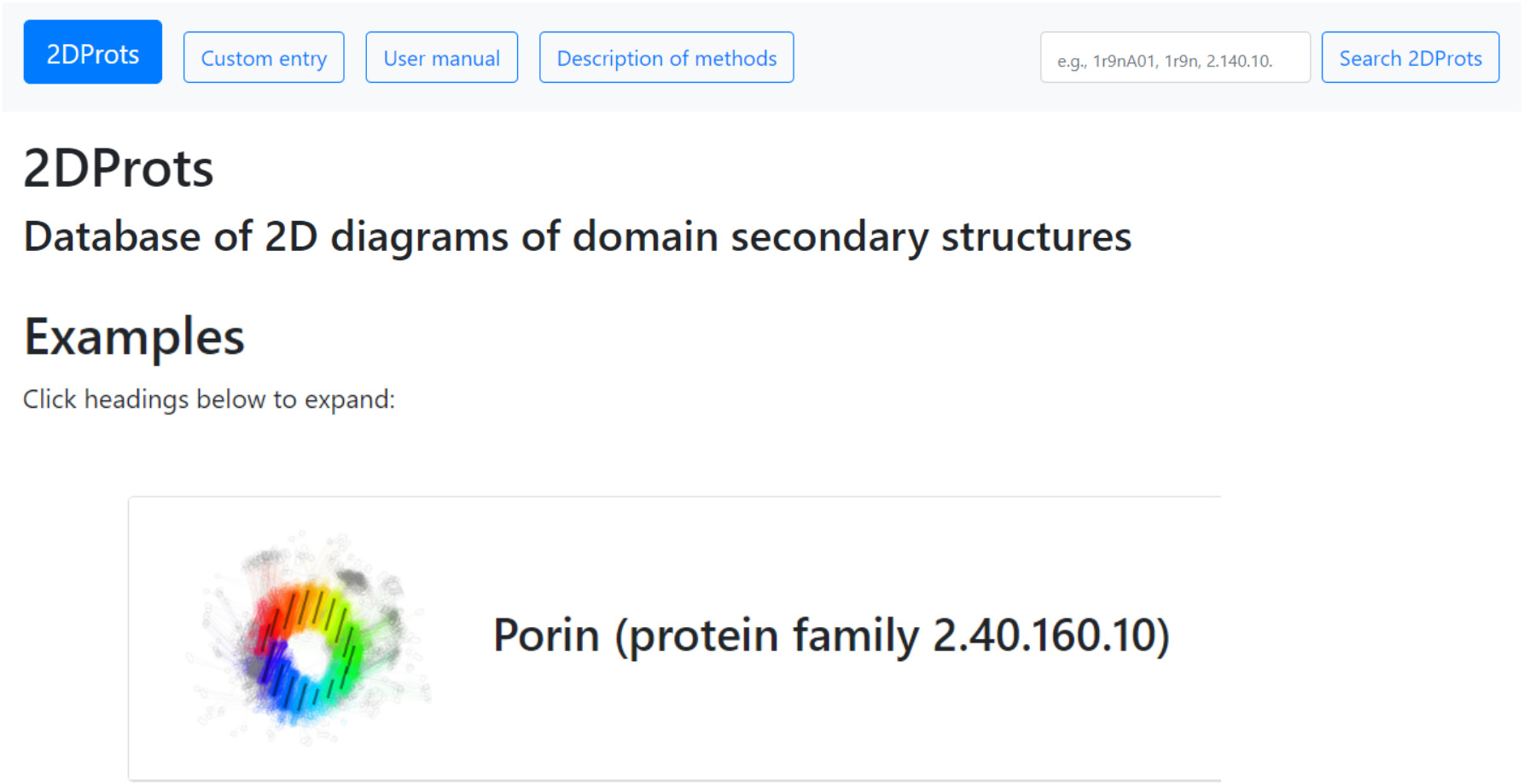
Step 4: Draw the 2D diagrams





Protein family 2D diagrams
2DProts database

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



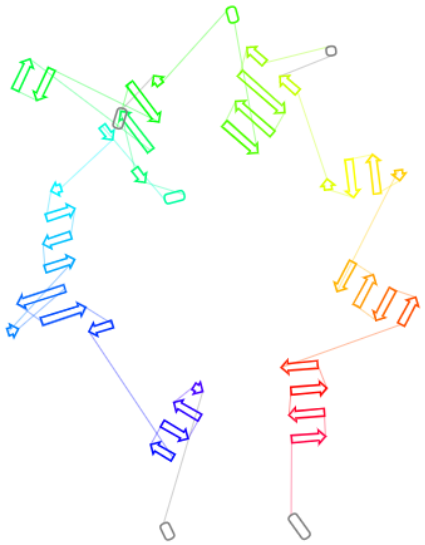
Protein family 2D diagrams
2DProts database

2DProts | Custom entry | User manual | Description of methods | e.g., 1r9nA01, 1r9n, 2.140.10.3 | Search 2DProts

Domain 1j2eA01

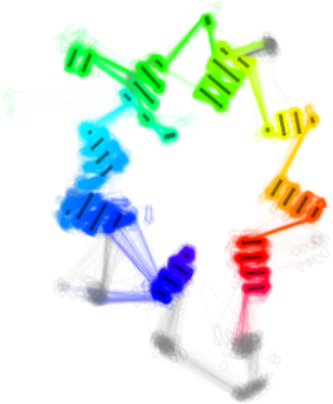
2D domain diagram

No ligands | With ligands



« no prev. | 1j2eB01 »

Member of family:
2.140.10.30



Part of protein:
1j2e

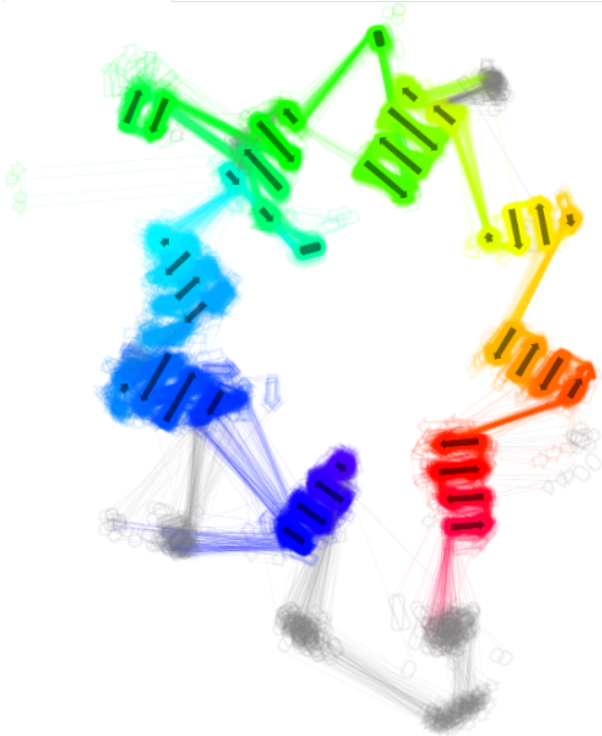
Protein family 2D diagrams 2DProts database

Family 2.140.10.30

2D multi diagram

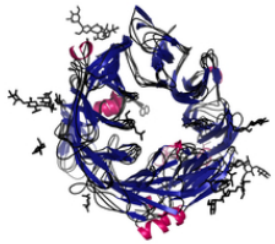
« 2.140.10.20 | 2.150.10.10 »

No ligands | With ligands | No ligands, opaque | No ligands, opaque, no averages



Multi image for 2.140.10.30 (svg source)

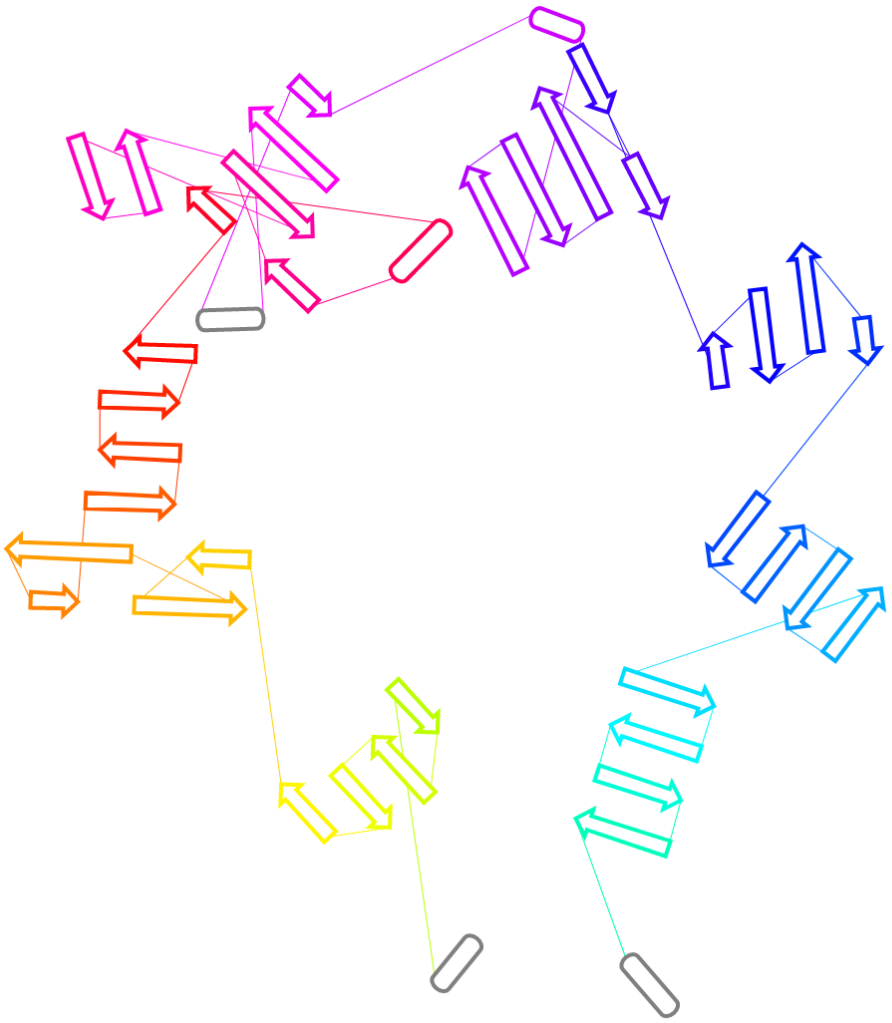
3D model (CATH)



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

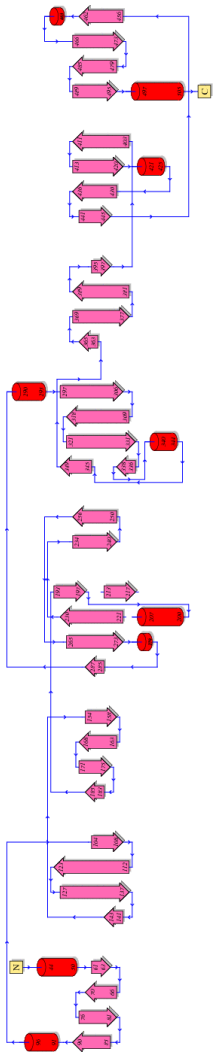
No opacity

Superfamily: Dipeptidylpeptidase IV (2.140.10.30)

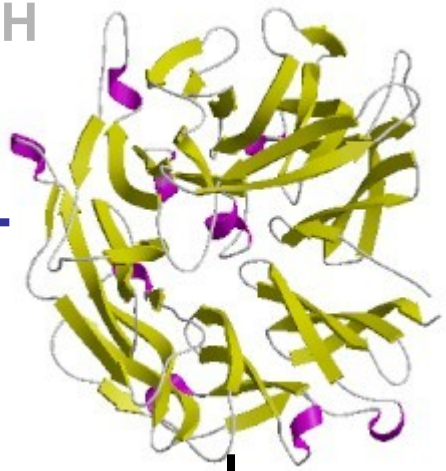
PROTEIN | **PROTEIN FAMILY**

Current solution

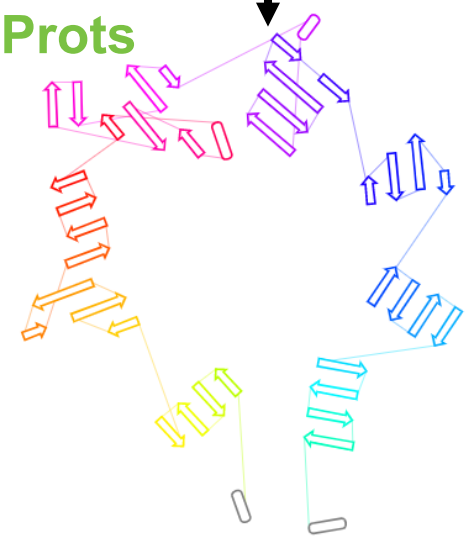
HERA



CATH



2DProts

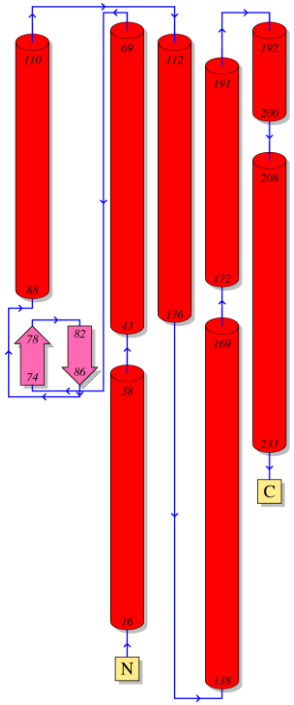


Superfamily: Rhodopsin 7-helix transmembrane proteins (1.20.1070.10)

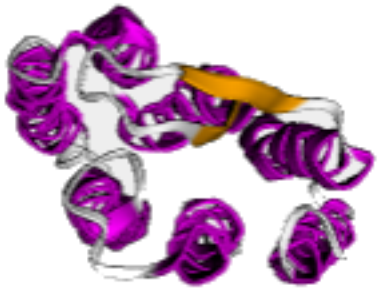
PROTEIN | **PROTEIN FAMILY**

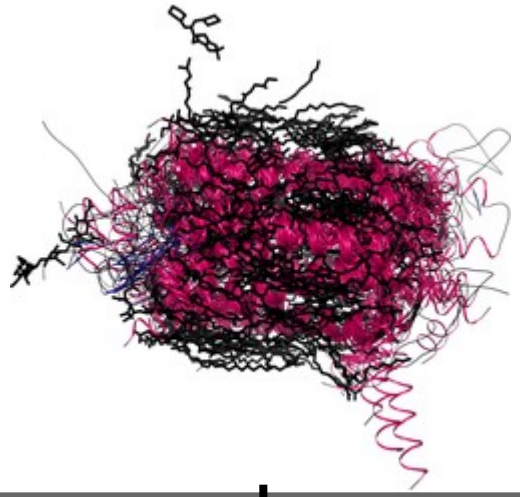
Current solution

HERA

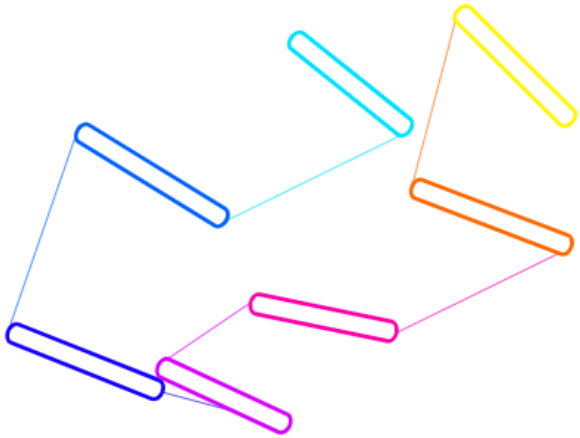


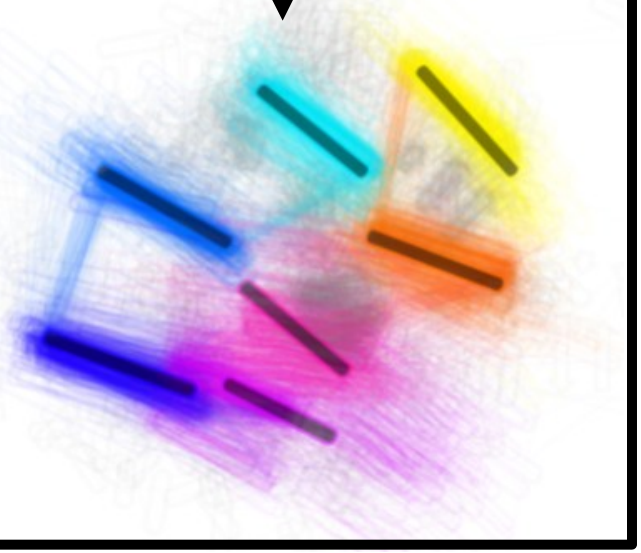
CATH





2DProts



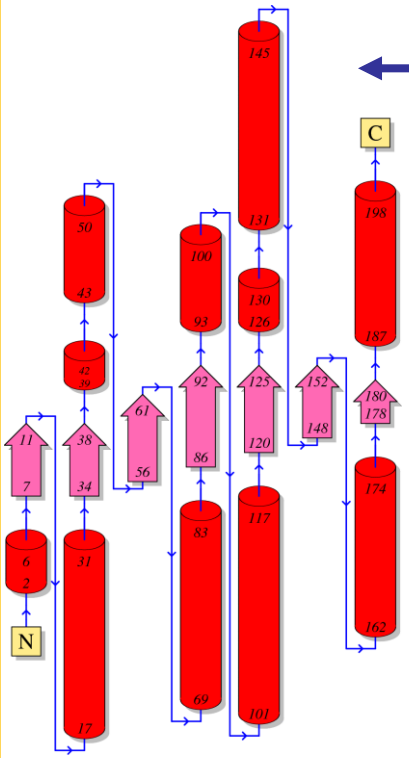


Superfamily: Aldolase class I (3.20.20.70)

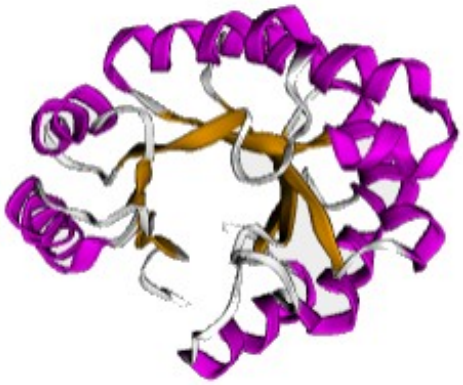
PROTEIN | **PROTEIN FAMILY**

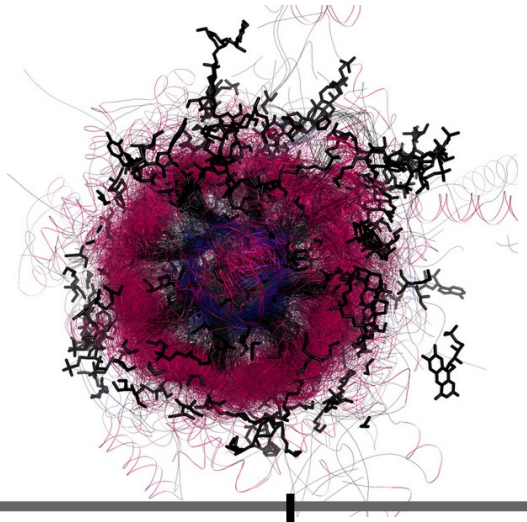
Current solution

HERA

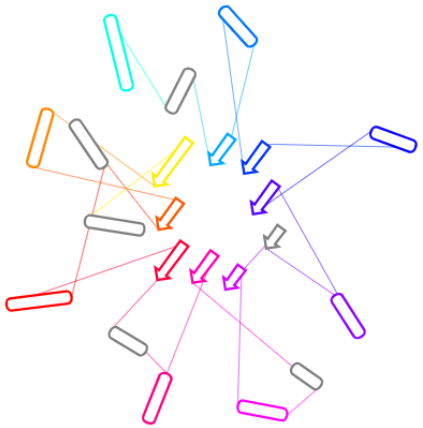


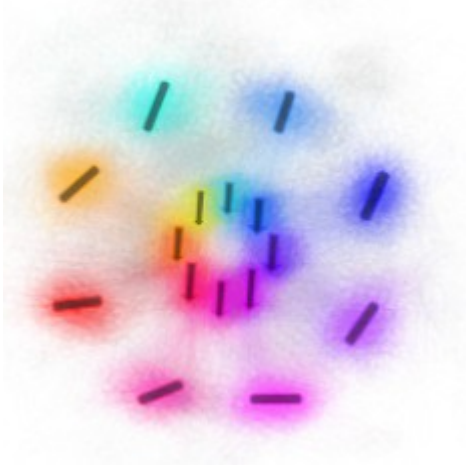
CATH



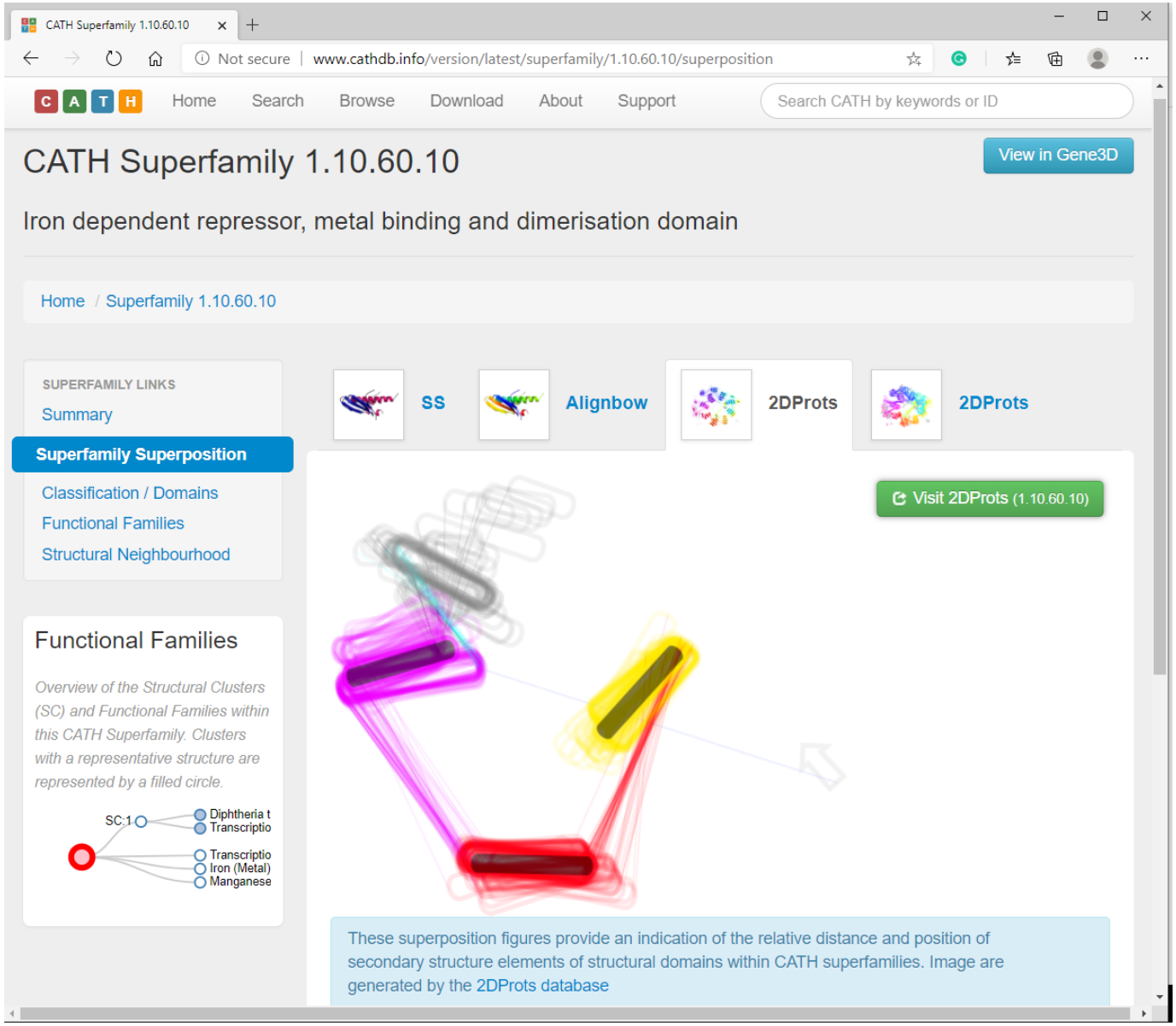


2DProts



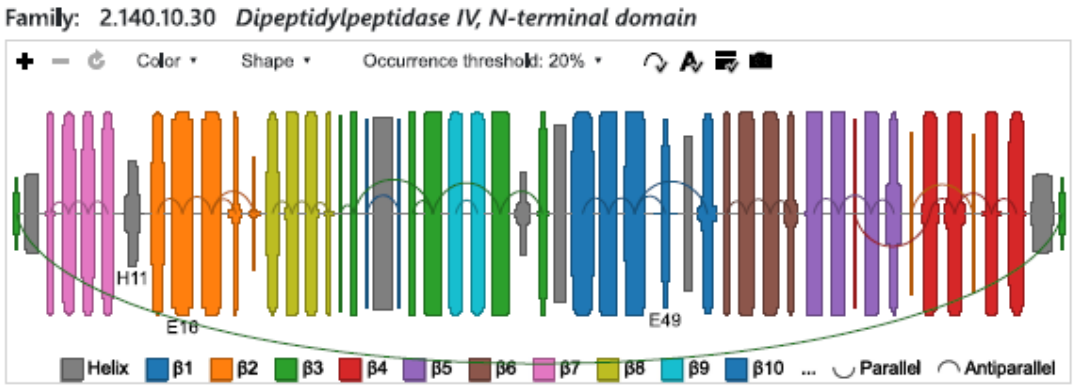


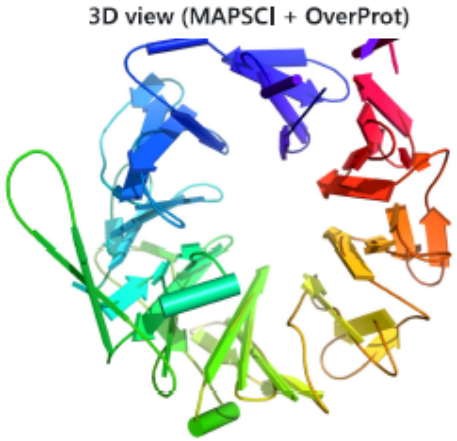
2DProts integration to CATH



2DProts integration into OverProt

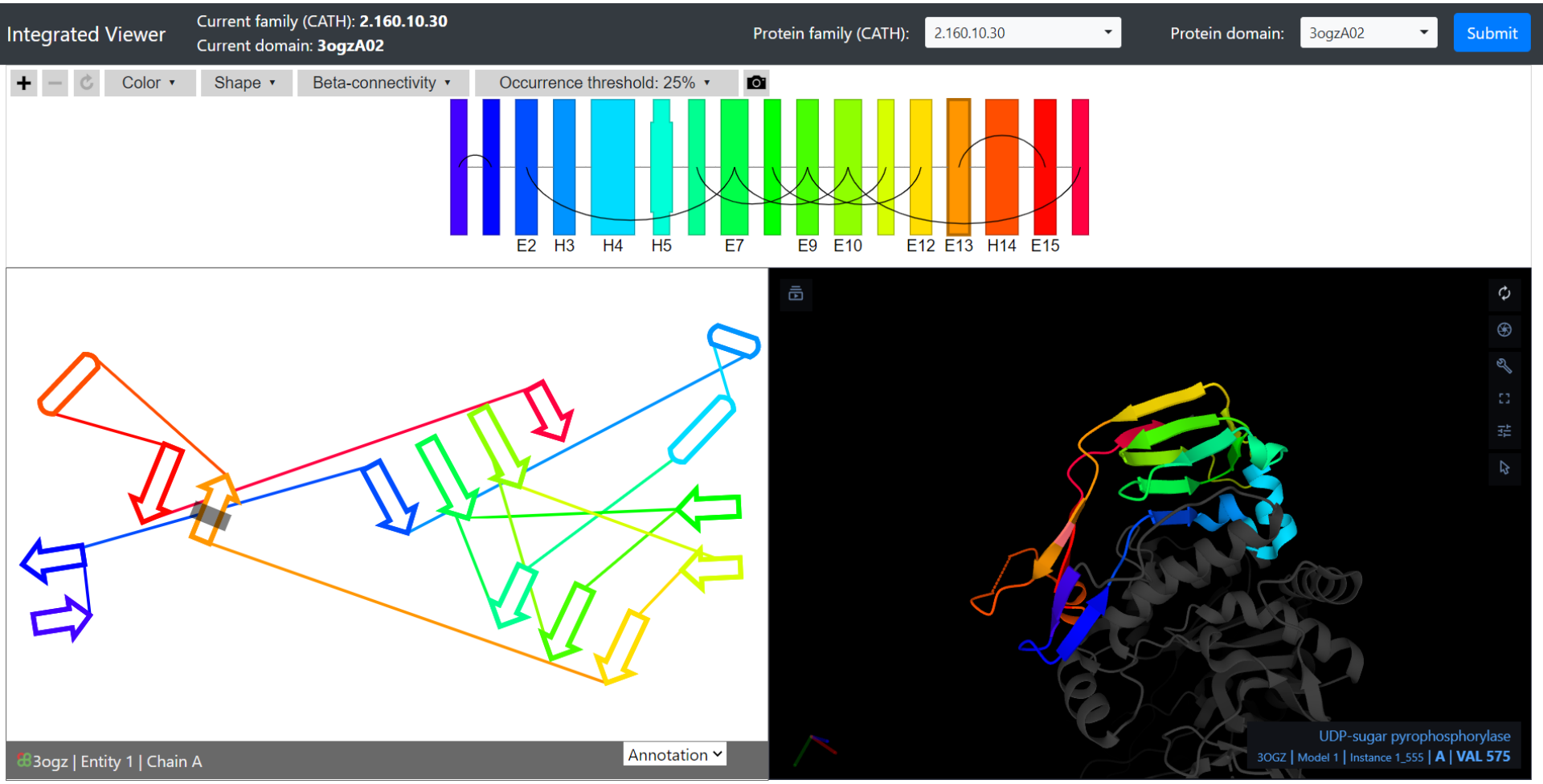
https://overprot.ncbr.muni.cz





2DProts integration into OverProt

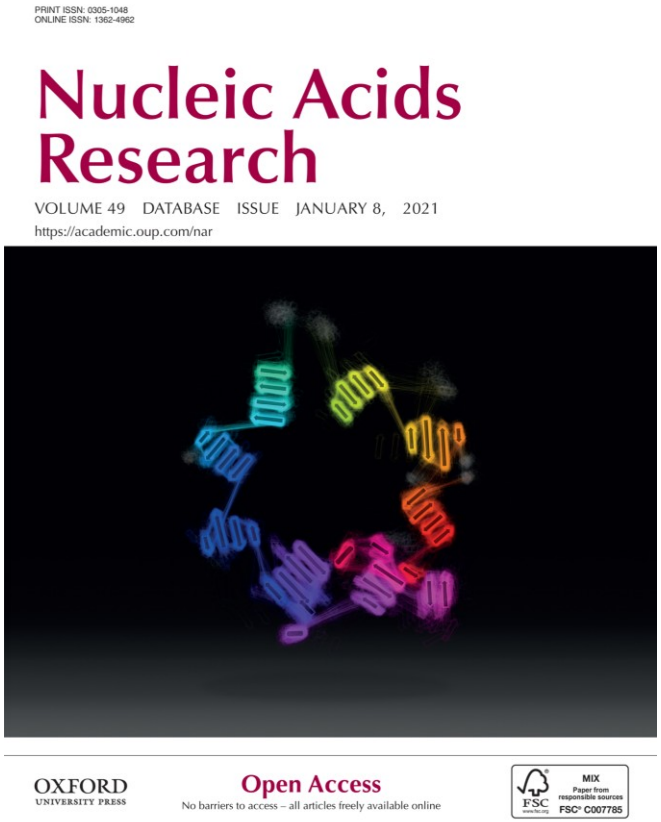
https://overprot.ncbr.muni.cz



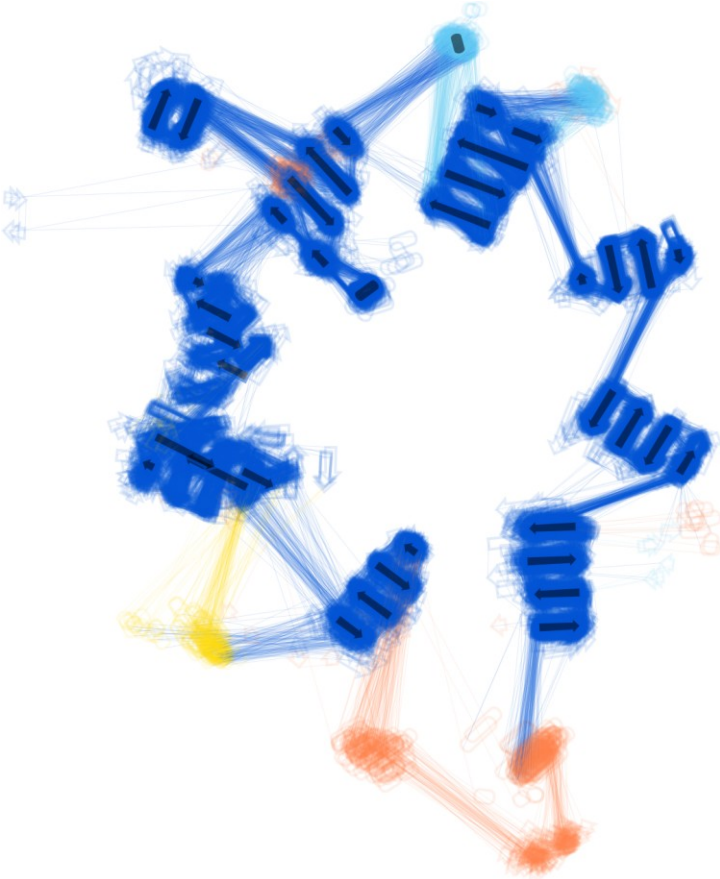
Publications

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.

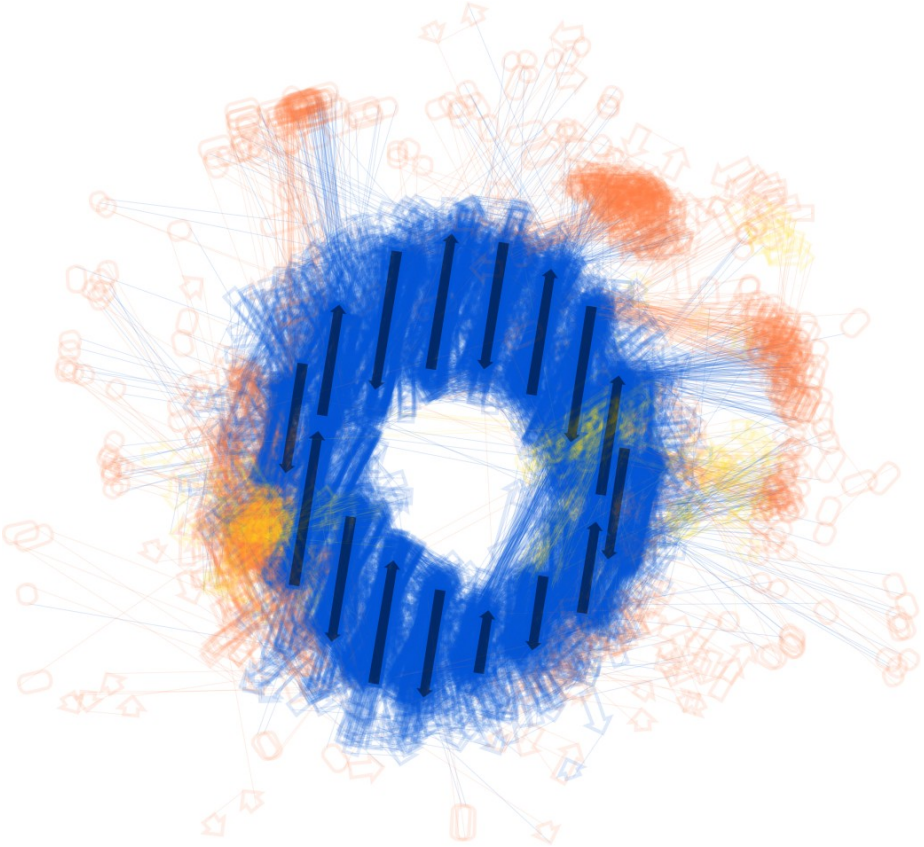
Hutařová Vařeková, I., Hutař, J., Midlik, A., Horský, V., Hladká, E., Svobodová, R., & Berka, K. (2021). *2DProts: database of family-wide protein secondary structure diagrams*. **Bioinformatics**, 37(23), 4599-4601.



2DProts: Coloring by structure properties
Example: Occurence of secondary structures

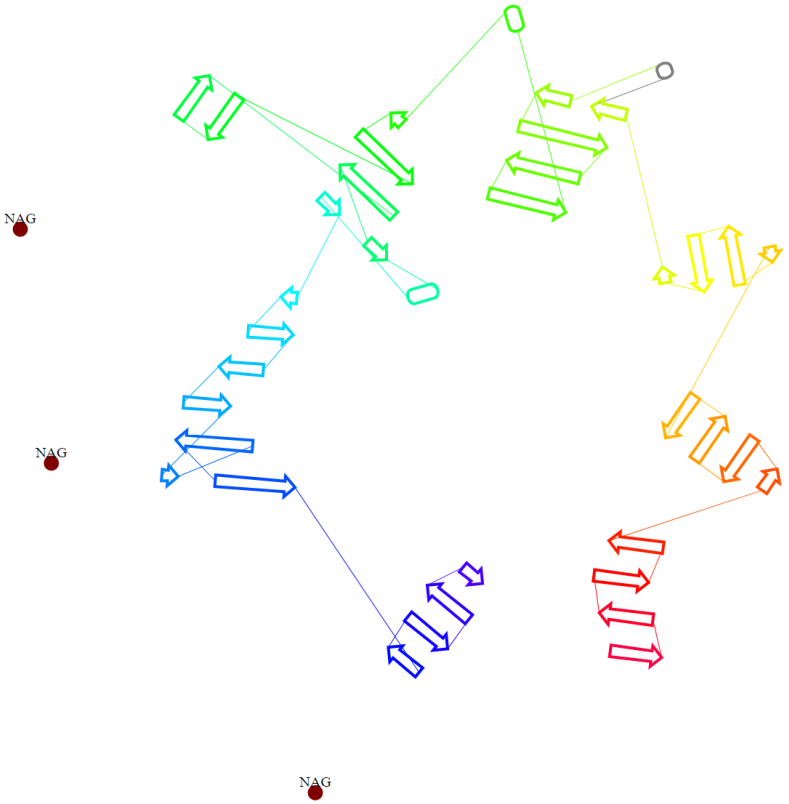


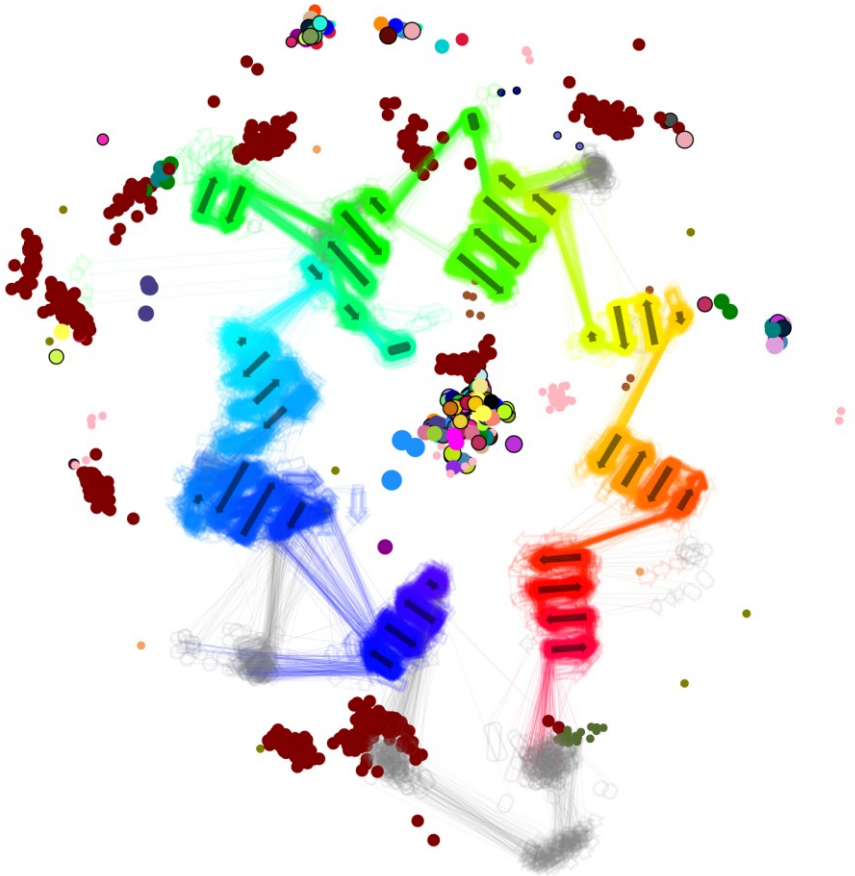
Cytochrome reductase,
Family 2.140.10.30



Porin
Family 2.40.160.10

2DProts: Integration of ligands



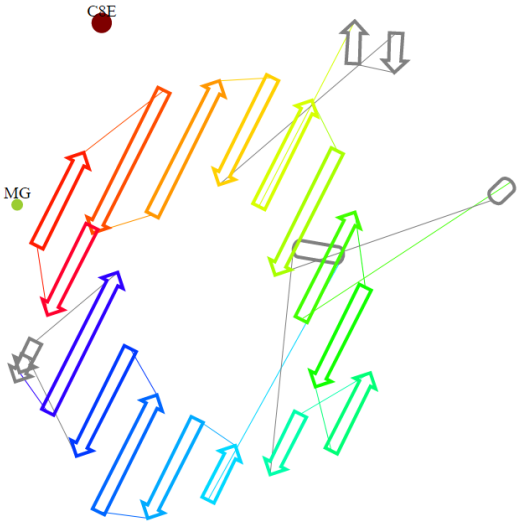


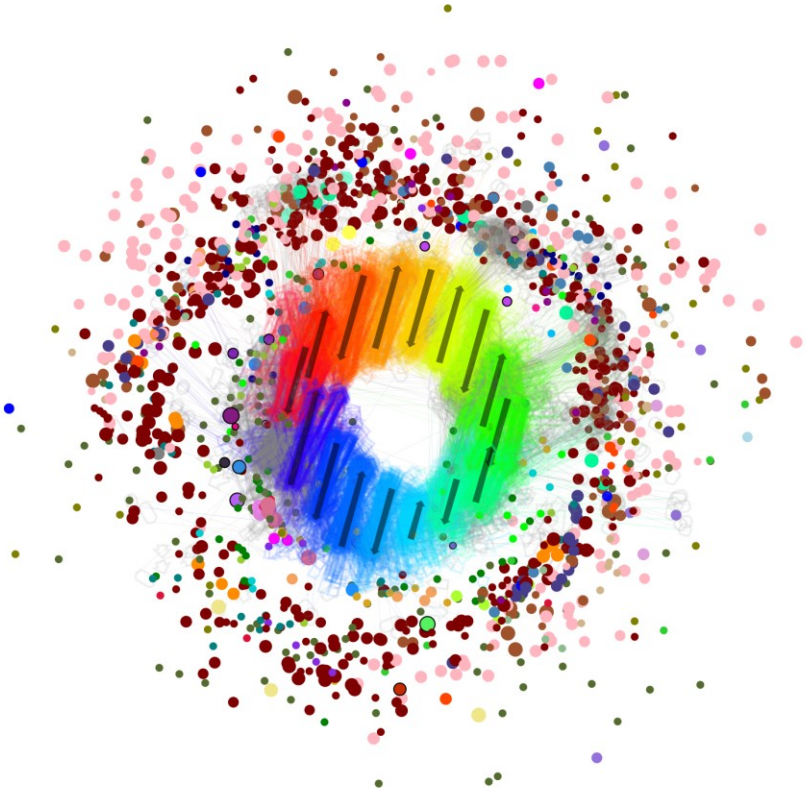
PDB ID 2bgn, domain A00

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)
...SO4 (56 pcs) ...SC3 (8 pcs) ...6RL (4 pcs) ...B2Y (4 pcs) ...8O3 (4 pcs) ...0QG (4 pcs) ...008 (4 pcs) ...PS4 (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)

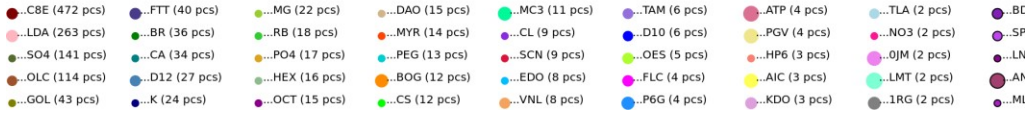
2DProts: Integration of ligands



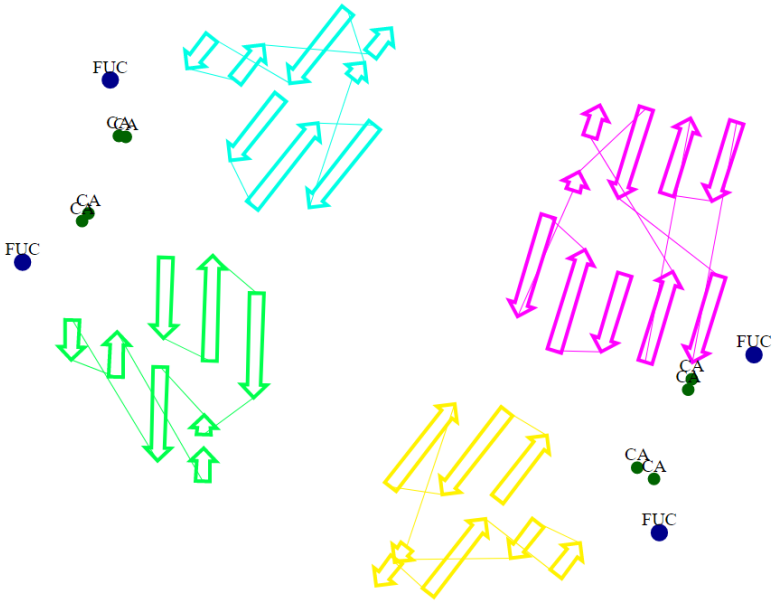


OMPF Porin PDB ID 2zfg, domain A00

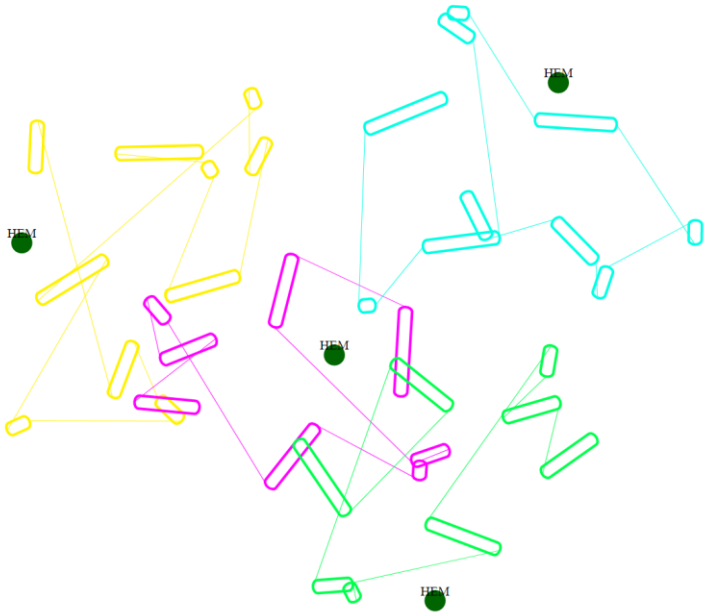
Porin, Family 2.40.160.10



2DProts: 2D diagrams for proteins

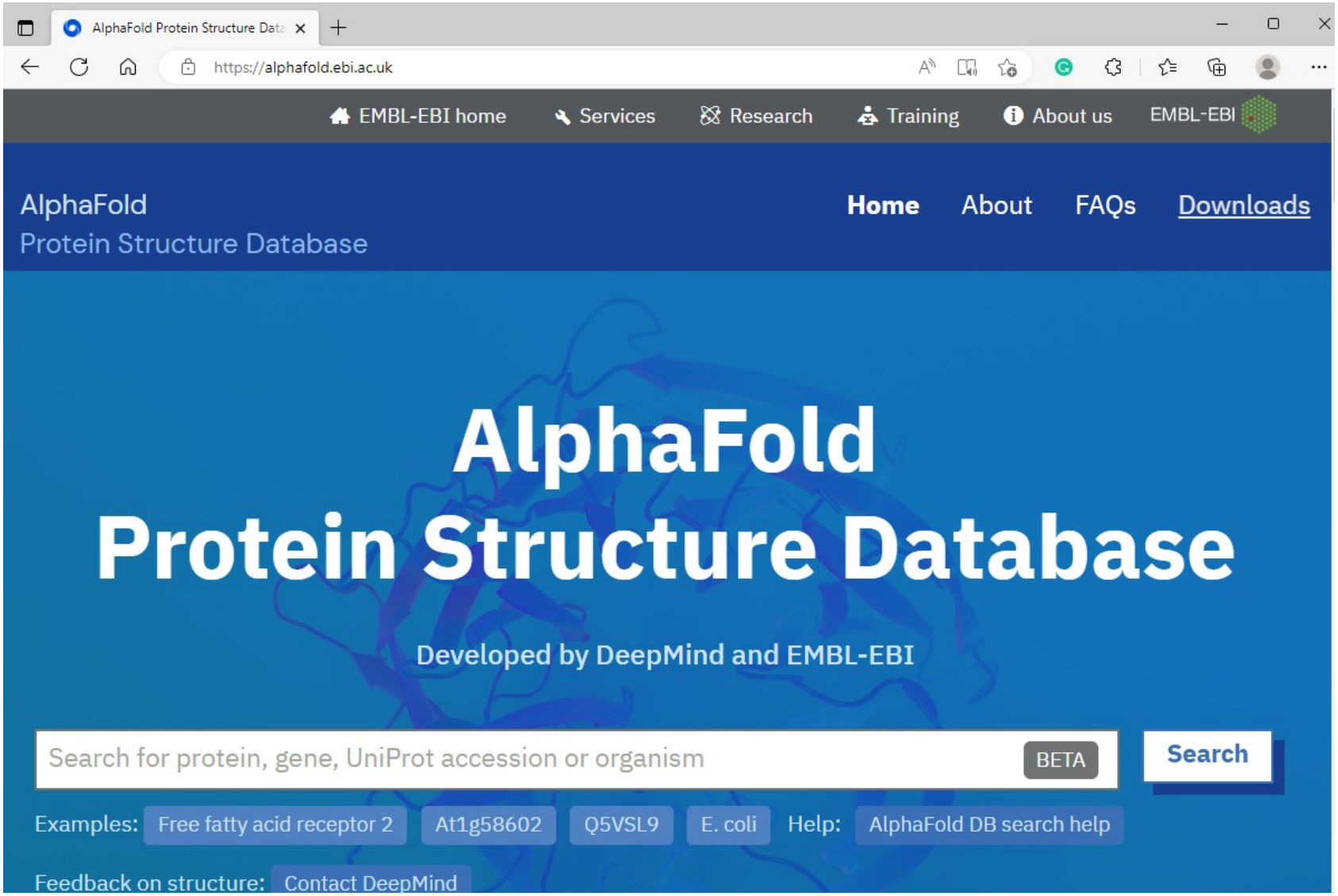


Pseudomonas aeruginosa lectin II PDB ID 1gzt

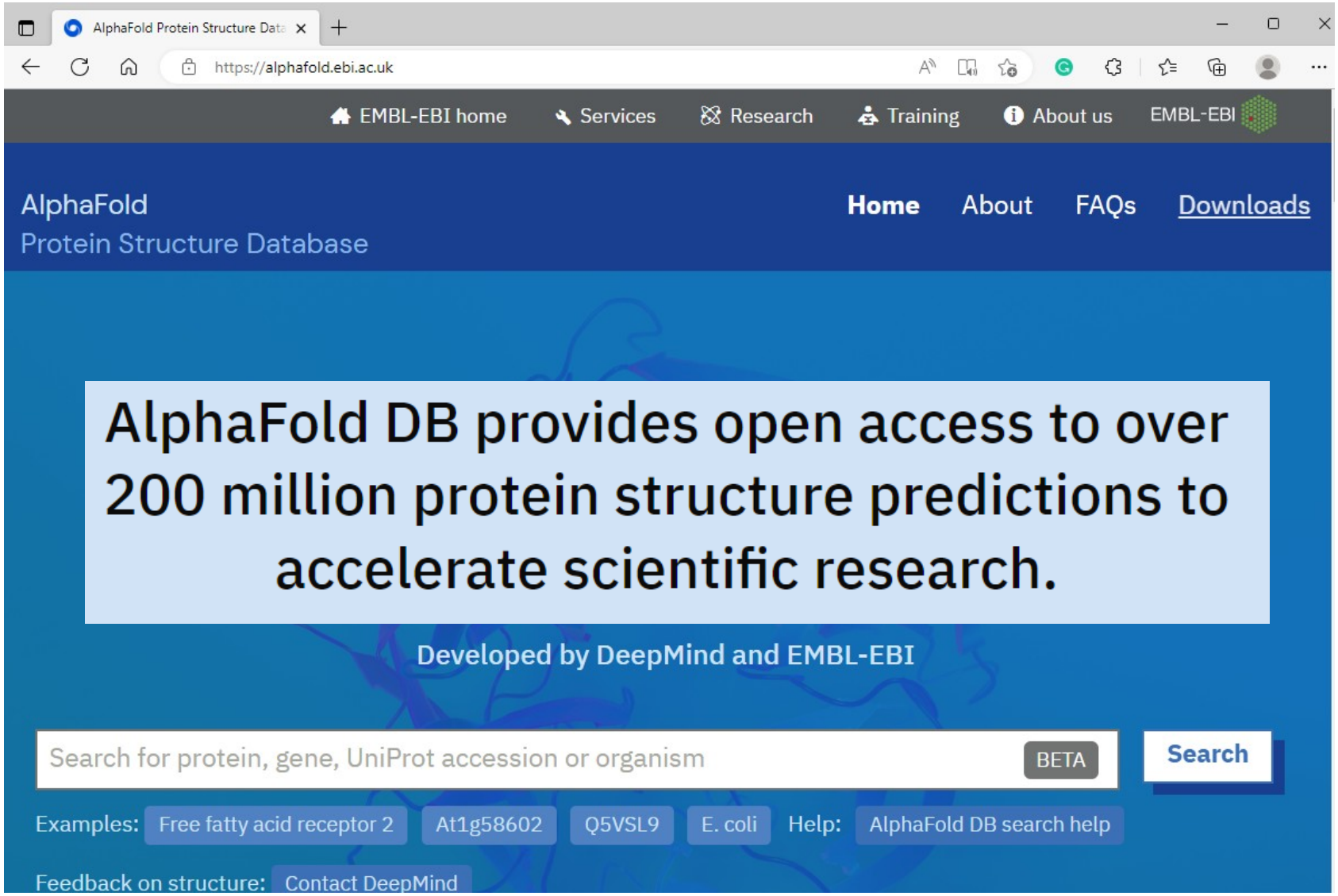


Hemoglobine PDB ID 1v4w

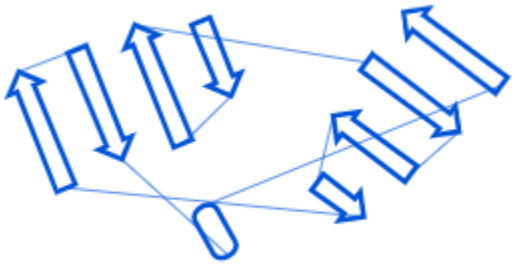
2DProts: Integration of AlphaFoldDB

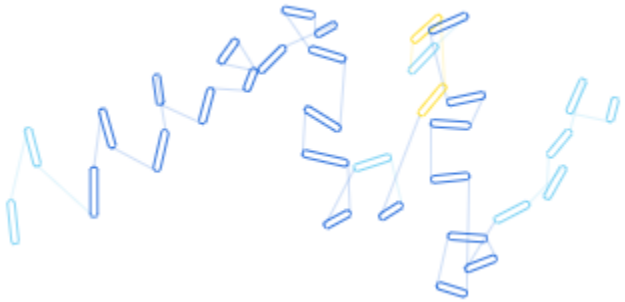


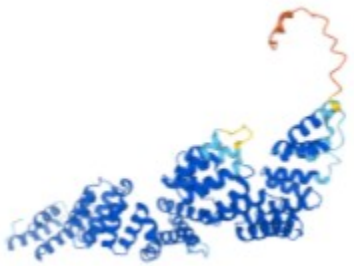
2DProts: Integration of AlphaFoldDB

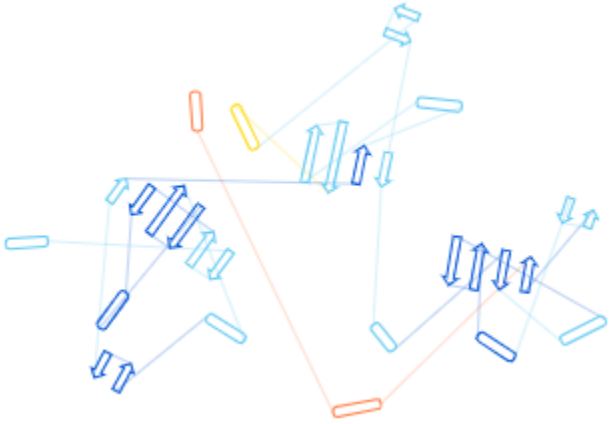


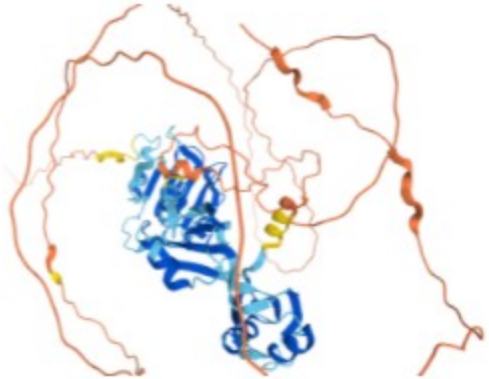
2DProts: Integration of AlphaFoldDB





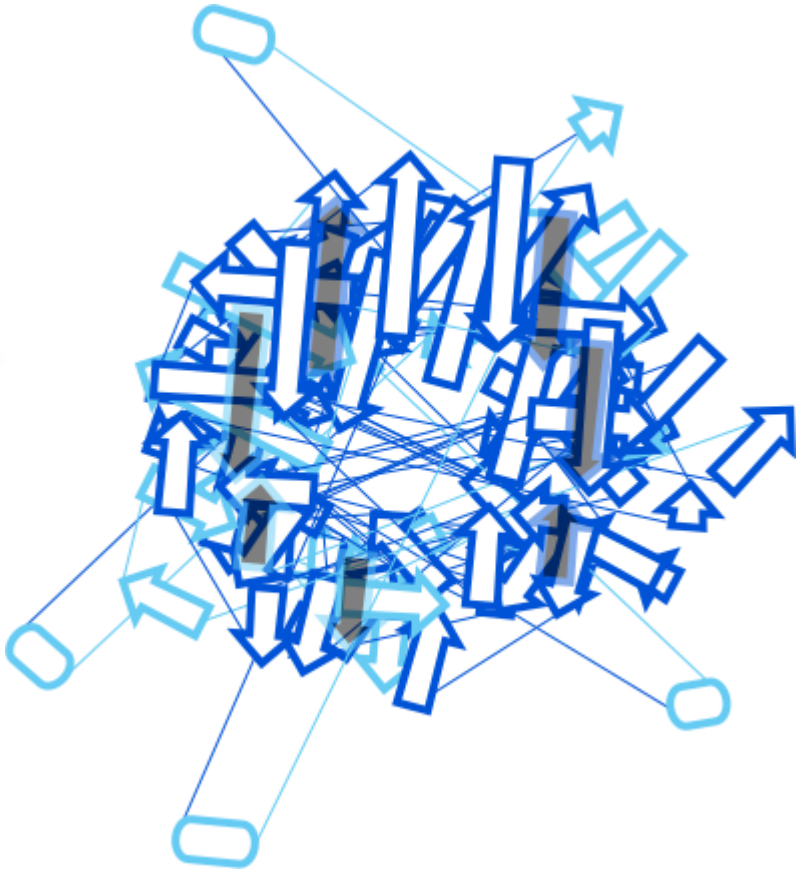






2DProts: Integration of AlphaFoldDB

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