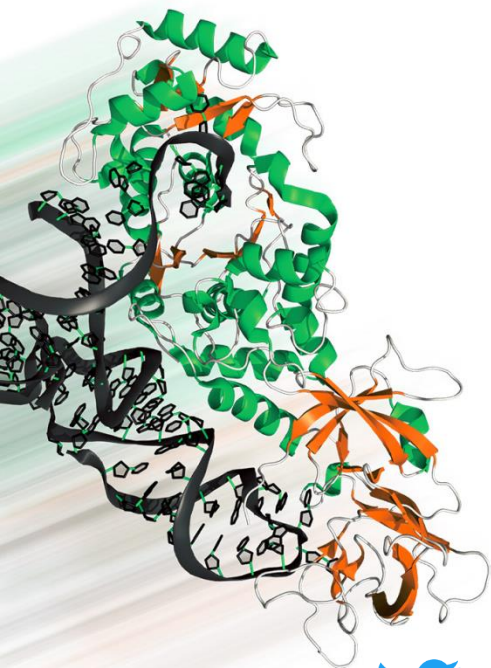




central european institute of technology

BRNO | CZECH REPUBLIC



Detection and extraction of key structural regions (patterns)

Lukáš Pravda



@WebchemTools



ncbr.muni.cz/webchemistry



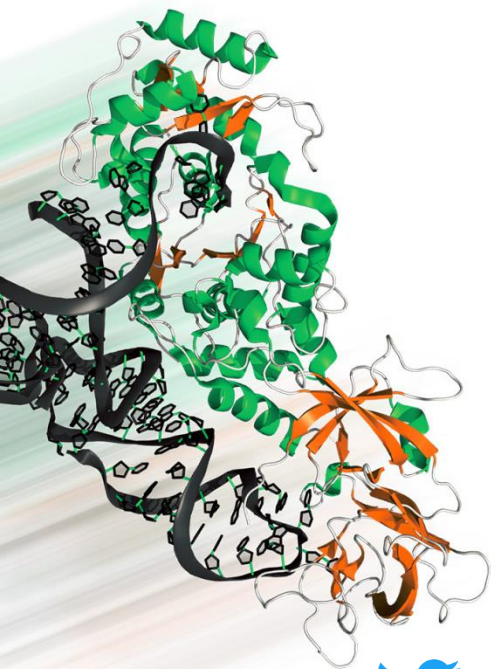
Outline

- Variety of structural patterns
 - How can we find them?
 - Software tools
 - PatternQuery
- Channels/tunnels/pores
 - Identification of channels and their properties
 - Software tools
 - MOLE
- Practical session with examples



central european institute of technology

BRNO | CZECH REPUBLIC



Variety of structural patterns

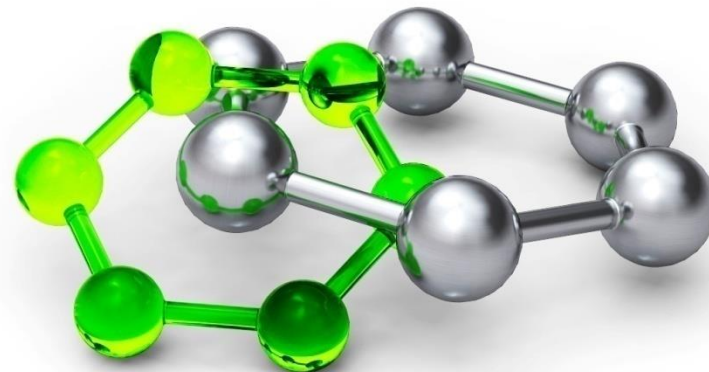
Lukáš Pravda



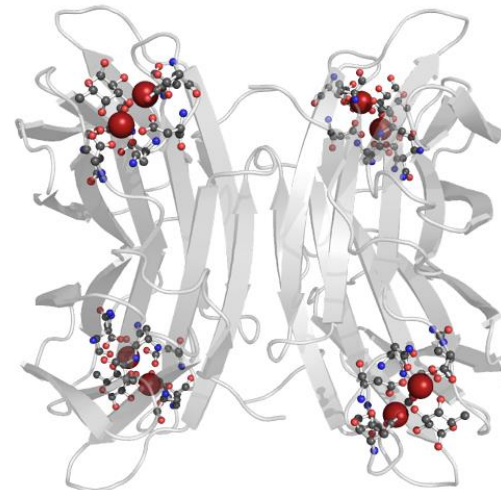
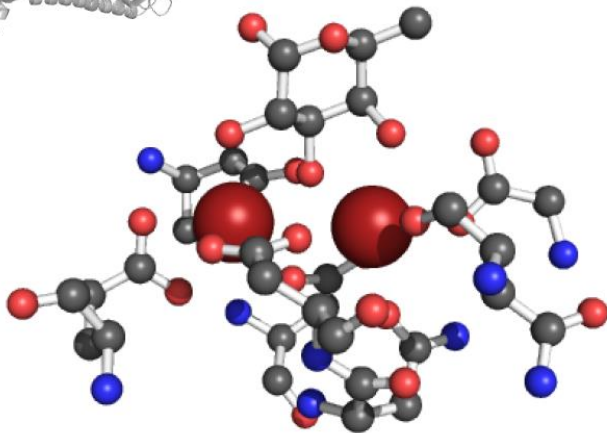
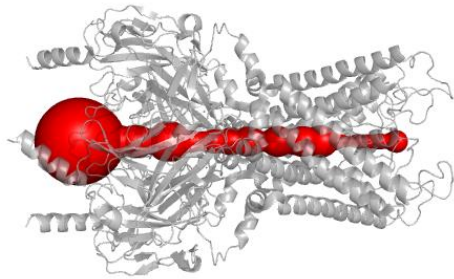
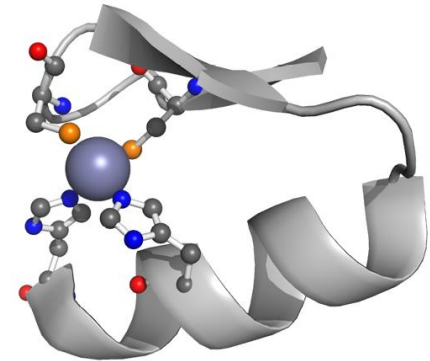
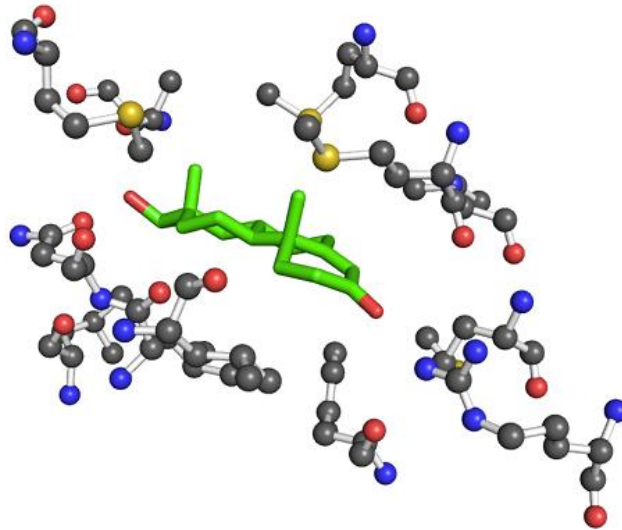
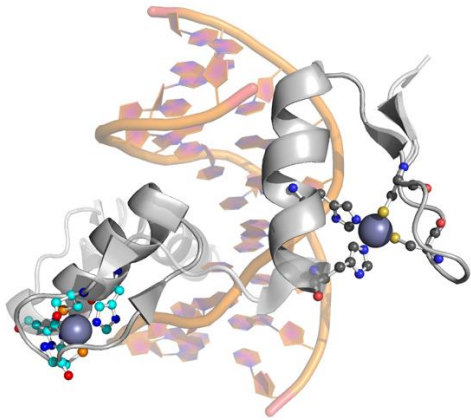
@WebchemTools



ncbr.muni.cz/webchemistry



Variety of structural patterns



Detection

- The goal is to identify and possibly extract biological regions of ones interest within biomolecular structure.
- Including but not limited to:
 - Active/binding/interaction sites
 - Sequences of amino acids or nucleic acids
 - Pockets/channels or void.
 - Super secondary motifs.

OK, but wait why do we need them?

- Database wide detection enables us to carry out experiments which not has been feasible before.
- Output of these searches are often an input for further analyses:
 - Structural and functional assignment of newly determined structures.
 - Comparative analyses
 - Design and engineering of novel functional sites
 - Study of binding modes of certain atoms/residues

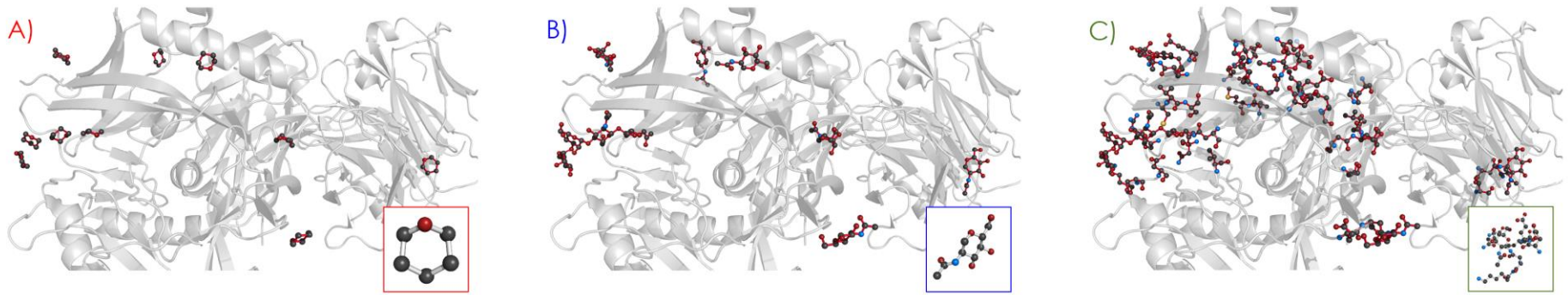
Software tools

- A plethora of different software tools – these are usually a single purpose:
 - Detection of ligands
 - Binding site identification
 - Pockets/cavities
 - Channels
 - In house scripts and tools
- The question is, can we do any better?

PatternQuery

- Web-based application designed for detection and extraction of molecular (sub)structures - patterns of user interest.
- Uses unique python like query language to define composition, topology and connectivity of these patterns.
- Allows querying single structures as well as the entire PDB or its subset based on a number of criteria (organism of origin, resolution, date of release, ...)

How does it look like?



`Rings(5 * ['C'] + ['O']). ConnectedResidues(0). AmbientResidues(4)`

<http://ncbr.muni.cz/PatternQuery>

PatternQuery – Structure of language

- Generator queries
 - `Atoms()`, `Residues()`, `RegularMotifs()`
- Modifier queries
 - `ConnectedResidues()`, `AmbientAtoms()`, `Filter()`
- Combinatory queries
 - `Or()`, `Near()`, `Cluster()`
- So far some 50 different queries, which can be readily used!

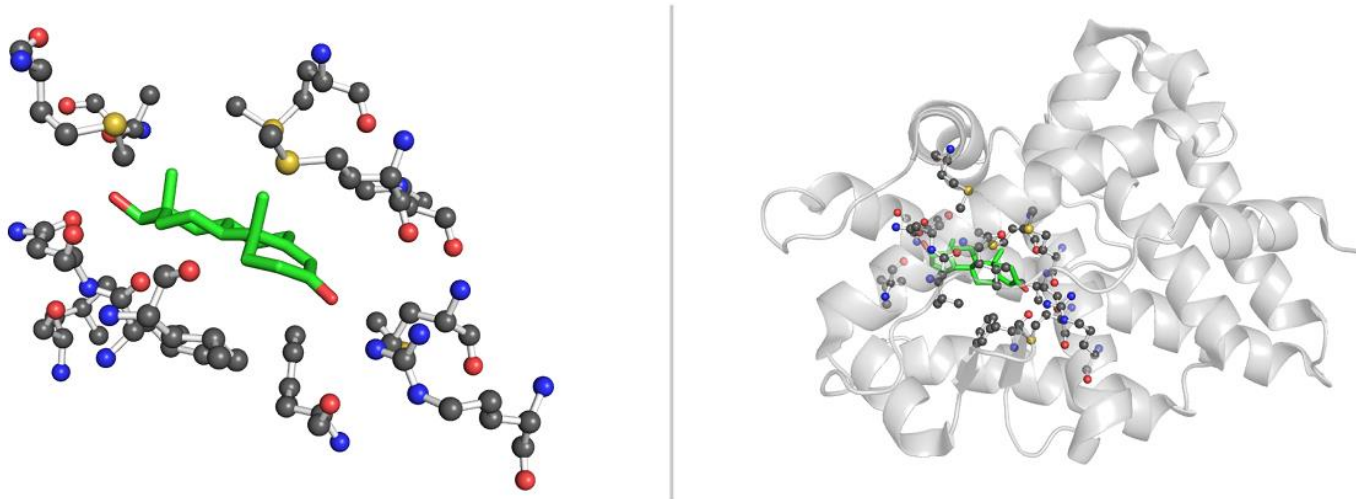
PatternQuery – Thinking in queries

- Find binding pocket of all ligands in the protein structure (distance $\leq 4\text{\AA}$)

```
1 temp = List()
2
3 for residue in molecule.Residues():
4     if residue.IsHet():
5         temp.Add(residue)
6
7 neighborhoodLookup = NeighborhoodLookup(molecule.Atoms())
8 result = List()
9
10 for residue in temp:
11     surroundings = neighborhoodLookup.Find(residue.Atoms, <= 4.0)
12     result.Add(union(residue, surroundings))
13
14 return result
```

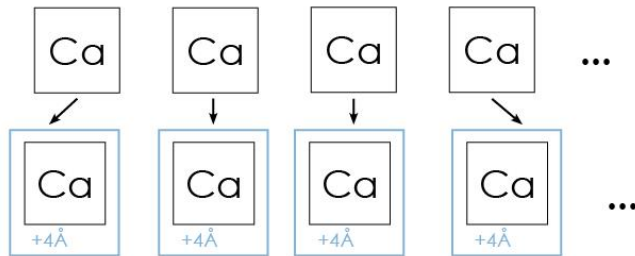
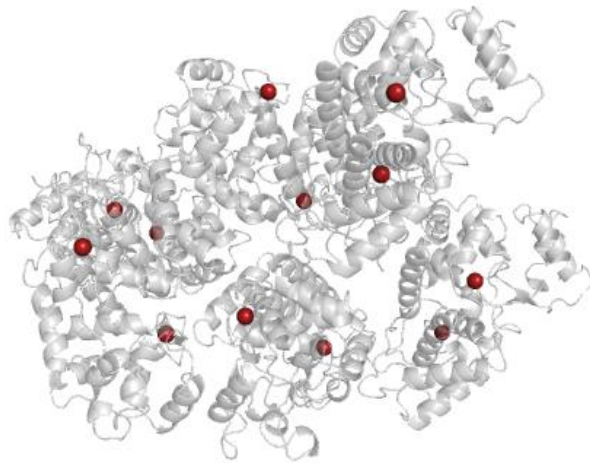
PatternQuery – Thinking in queries

- Find binding pocket of all ligands in the protein structure (distance $\leq 4\text{\AA}$)

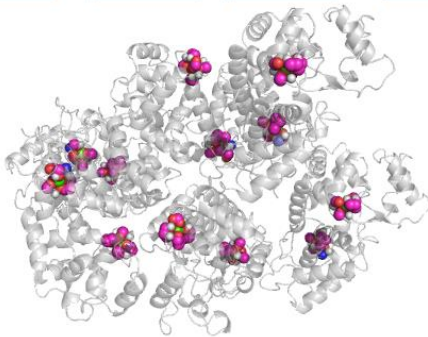


```
1 HetResidues().AmbientAtoms(4.0)
```

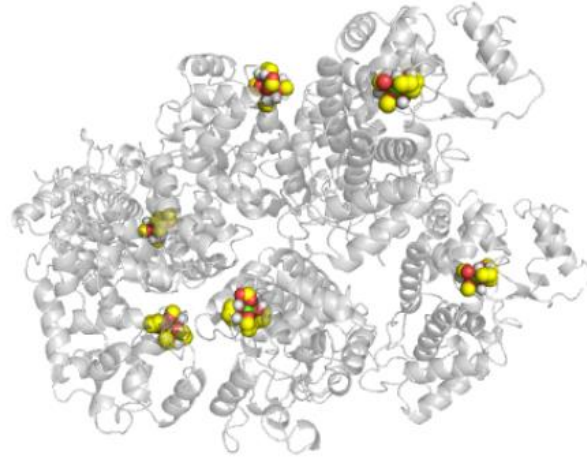
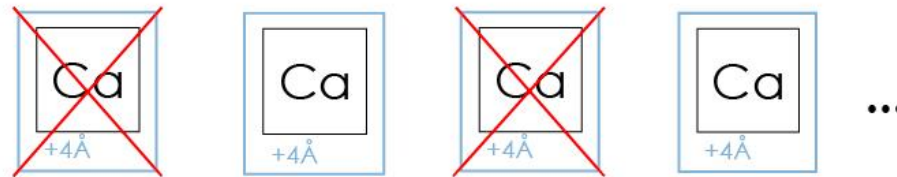
Build a query I



`Atoms("Ca"). AmbientResidues(4)`



Build a query II



```
Atoms("Ca")  
. AmbientResidues(4)  
. Filter(lambda l: l.Count(Atoms()) > 6))
```

Biologically interesting queries I.

- Post-translationally modified aminoacids

```
NotAminoAcids()                               ModifiedResidues()  
. Filter(lambda l: HetResidues() == 6))
```

- Het atoms not covalently bound to protein

```
HetResidues()  
. Filter(lambda l: l.IsNotConnectedTo(AminoAcids()))
```

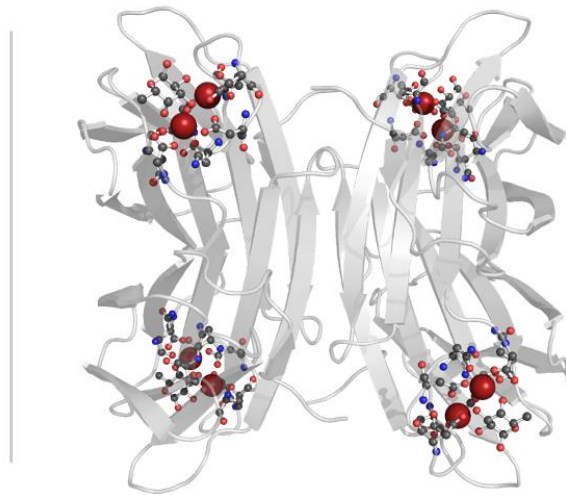
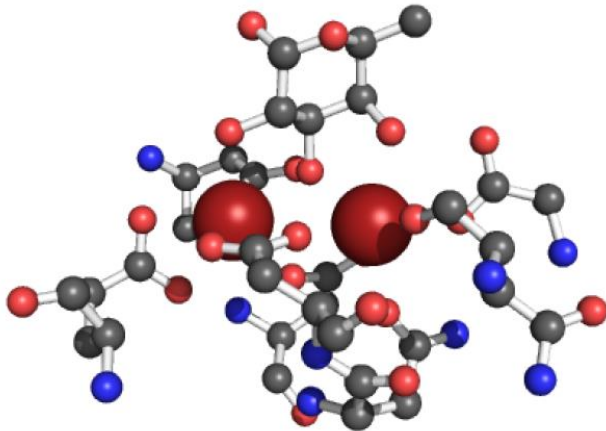
- Residues with a sugar moiety

```
Or(Rings(4 * ["C"] + ["O"]).ConnectedResidues(0),  
   Rings(4 * ["C"] + ["O"]).ConnectedResidues(0))
```

Biologically interesting queries II.

- PA Lec-B sugar binding site

```
Near(4, Atoms("Ca"), Atoms("Ca"))  
.AmbientResidues(3)  
.Filter(lambda l:  
    l.Count(Or(Rings(5 * ["C"] + ["O"]), Rings(4 * ["C"] + ["O"]))) > 0)  
.Filter(lambda l: l.Count(Atoms("P")) == 0)
```



Questions?



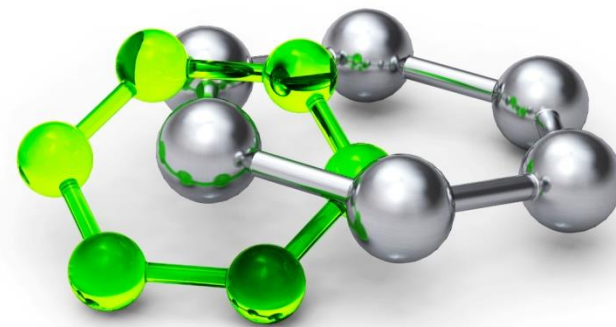
OK, let's move ON!



@WebchemTools



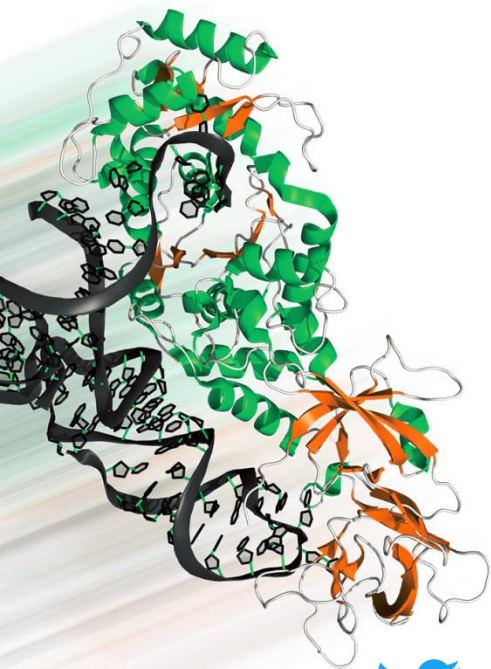
ncbr.muni.cz/webchemistry





central european institute of technology

BRNO | CZECH REPUBLIC



Channels

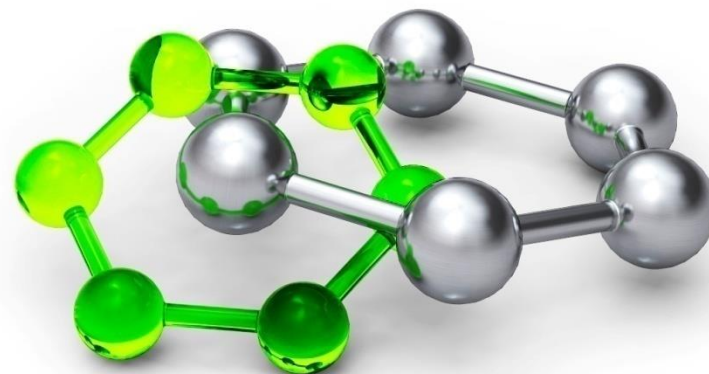
Lukáš Pravda



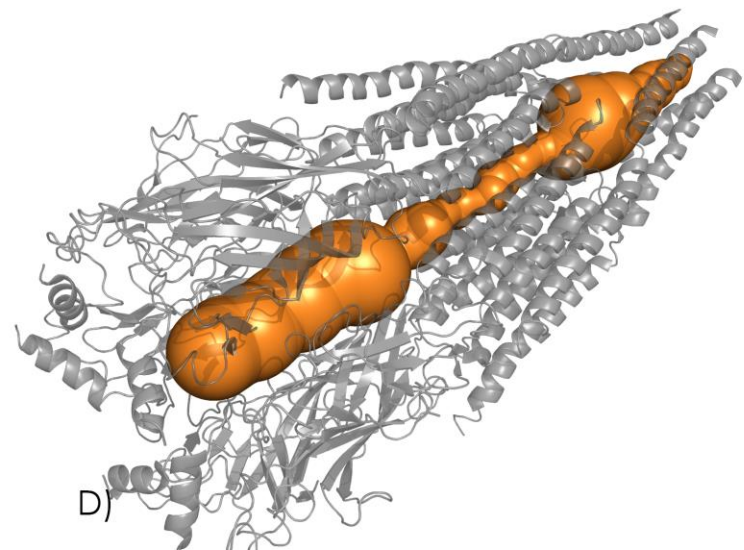
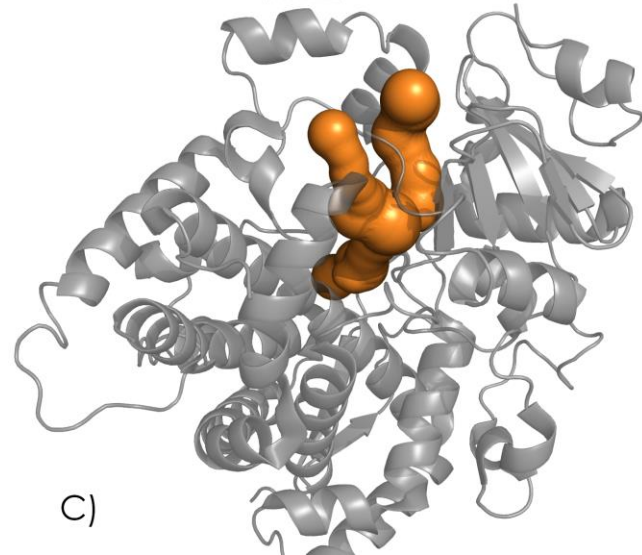
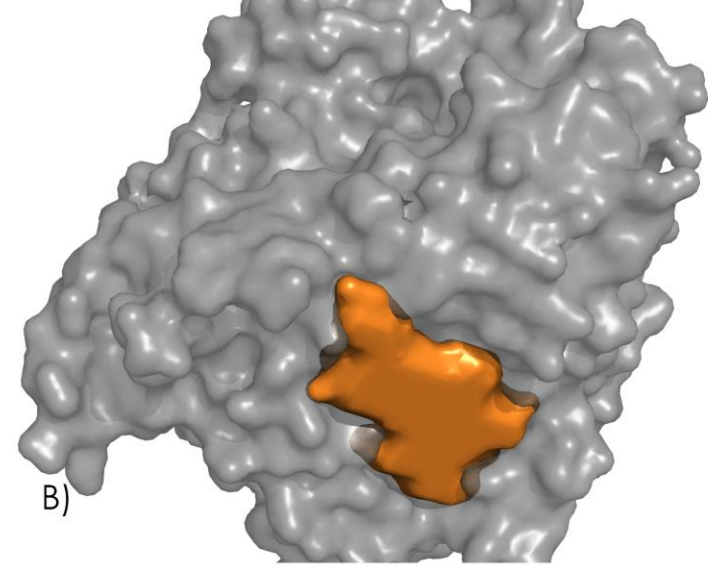
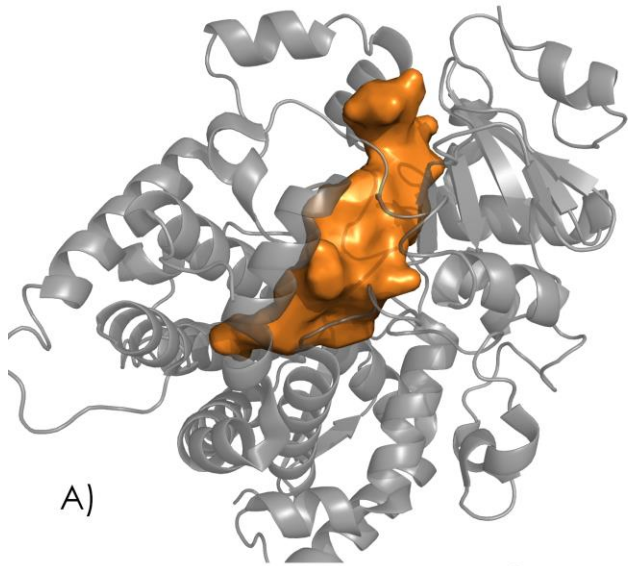
@WebchemTools



ncbr.muni.cz/webchemistry

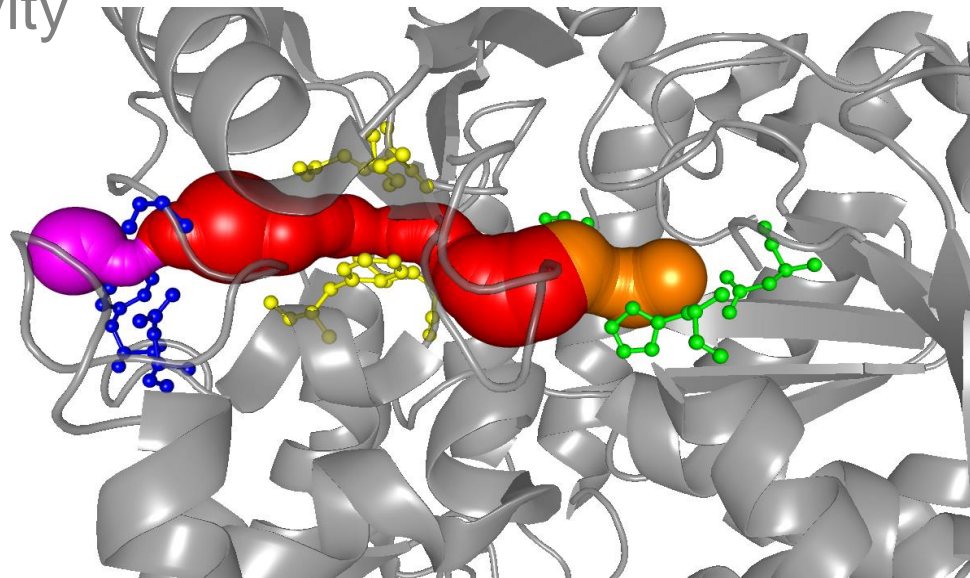


Protein empty voids



What are the channels/tunnels?

- A type of protein empty void.
- Connects active/binding site with the bulk solvent.
- Spans through membrane
- They greatly influence protein specificity, selectivity and rate of chemical processes.
- They look pretty(-ish 😊)

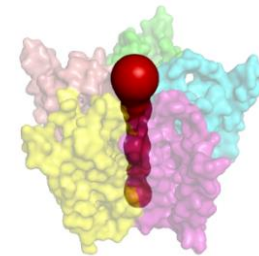


How can we find them?

- Over the time a number of approaches has been developed.
- Presently the most successful one relies on Delaunay Triangulation and Dijkstra's algorithm.
- Other approaches involves:
 - Grid search
 - Slice and optimization algorithms
 - Sphere-filling methods

Software tools

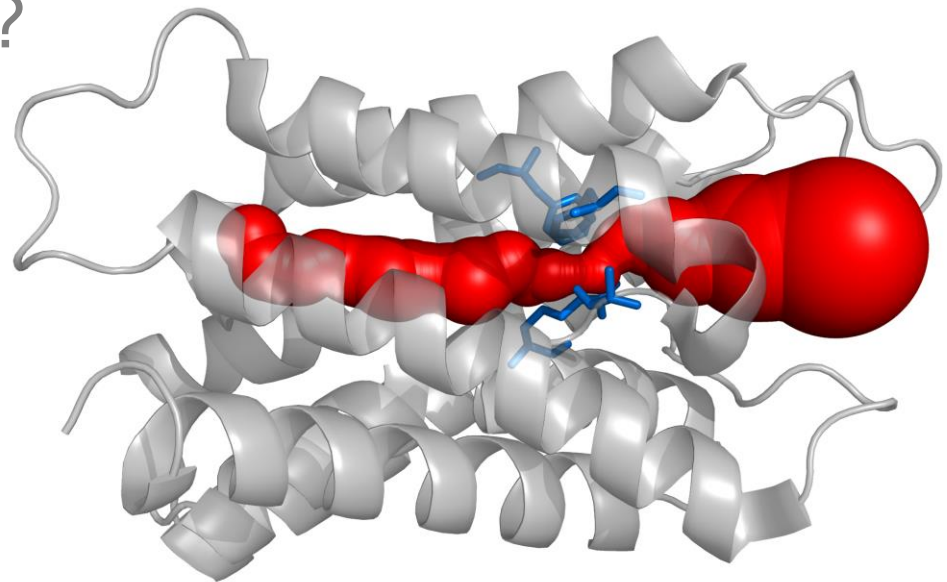
- MOLE
- CAVER
- MolAxis
- ChexVis
- BetaVoid
- HOLE
- And others...



ChExVis

Use case – aquaporin 0

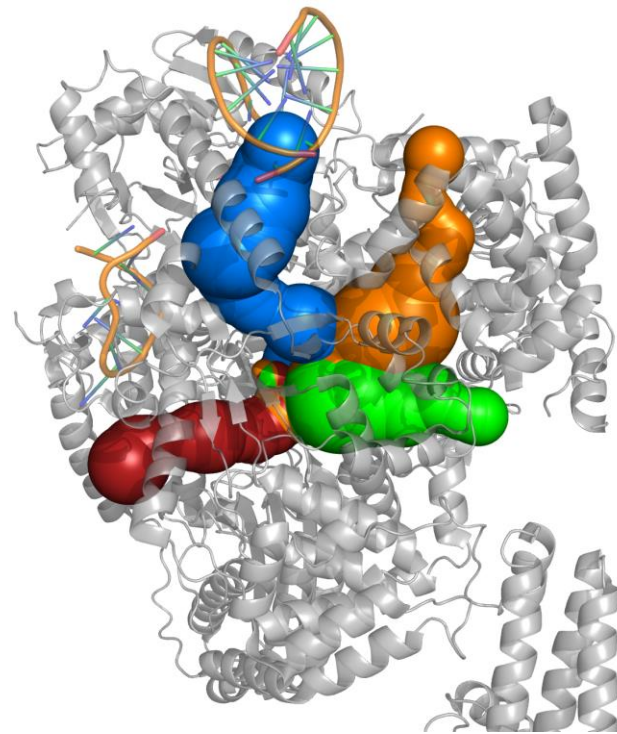
- Large family of proteins permitting permeation of various molecules – mainly water.
- Channel is a tight fit for water molecules.
- How can water permeate through the channel, while protons don't?



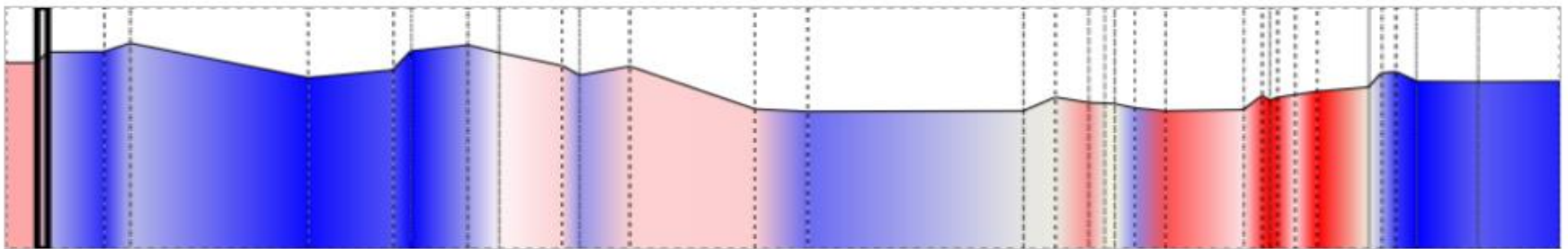
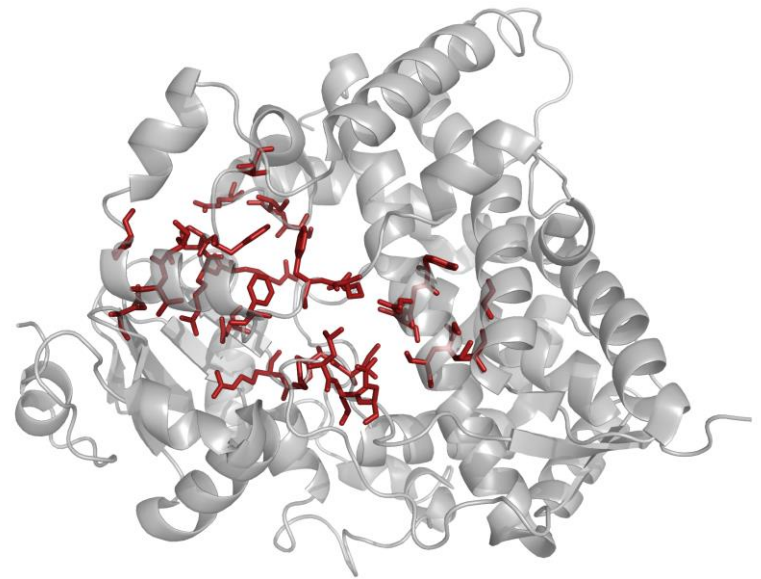
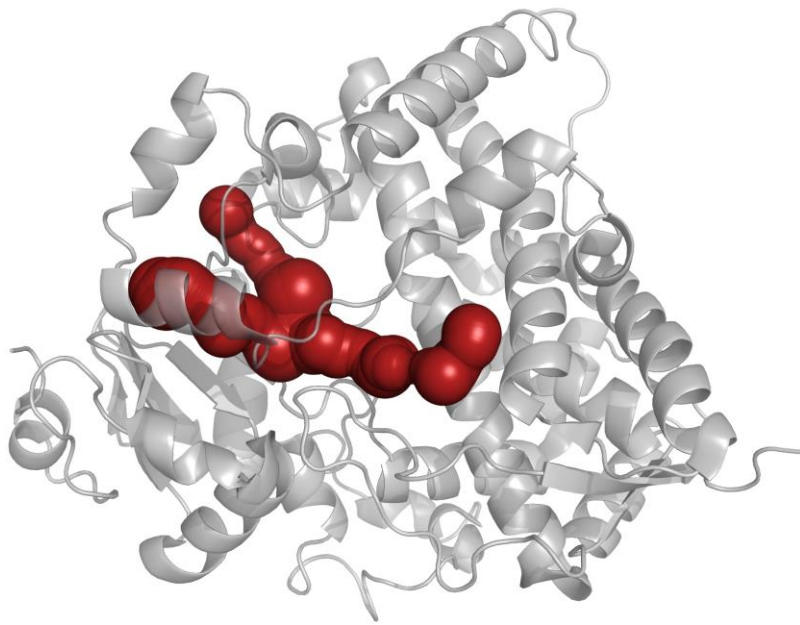
- ar/R region in blue

Use case – bunyavirus

- Negative-strand RNA viruses are serious human pathogens (Crimean-congo fever, Lassa fever, influenza).
- How one can kill a virus?
- Design a channel inhibitor!

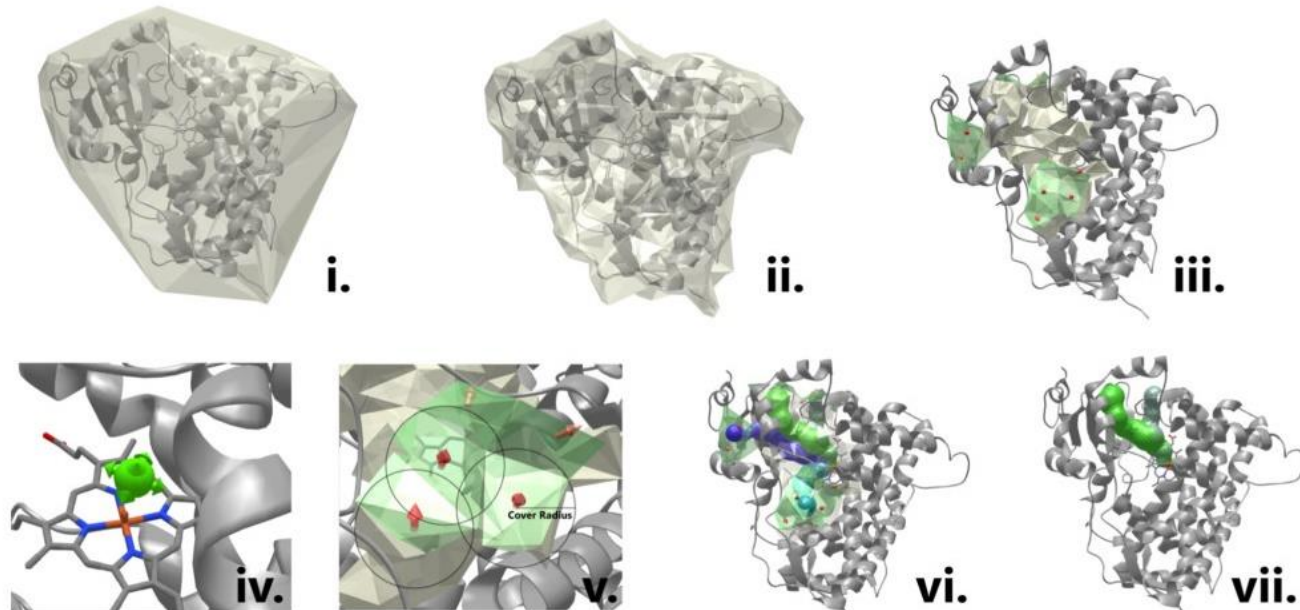


Physicochemical properties – channel duality



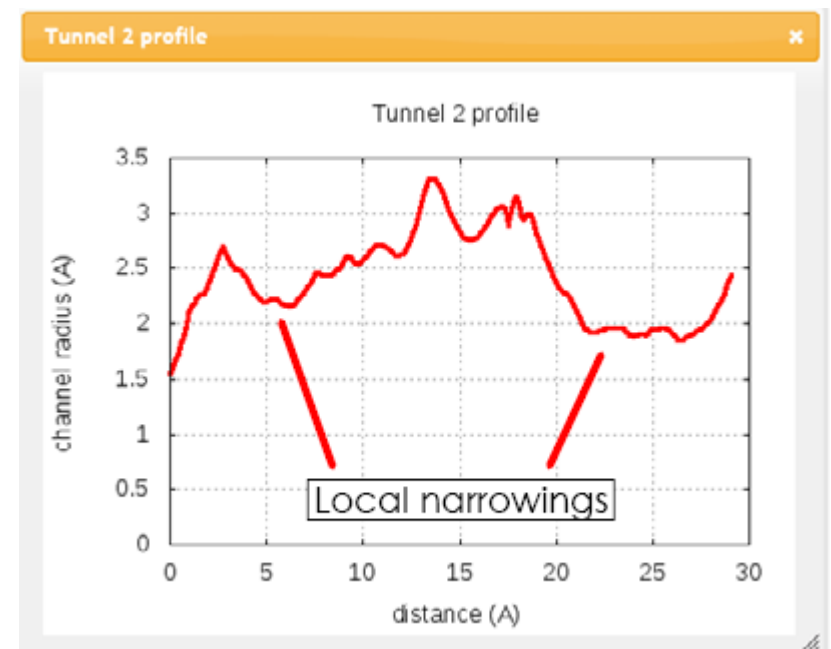
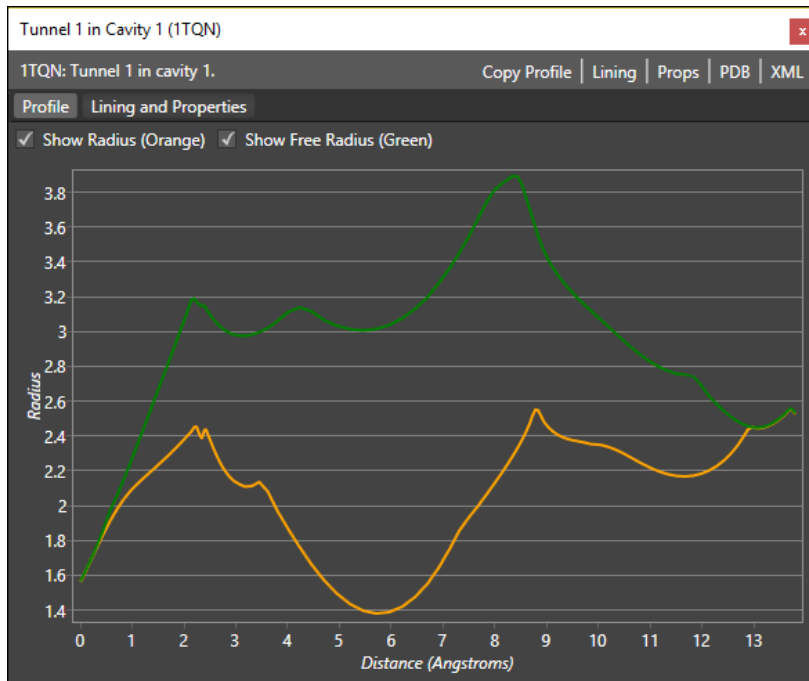
MOLE computation

- Input: Protein structure + set of parameters
- Output: Channel profile, properties and lining r.



Result analysis - properties

- Channel length vs channel radius
- Check presence of bottlenecks and local narrowings.
- Channel flexibility



Result analysis - properties

- Hydropathy, polarity, mutability, formal charge
- Evaluate independent layers as well as entire channel.

Tunnel 1 in Cavity 1 (1TQN)

ITQN: Tunnel 1 in cavity 1. Copy Profile | Lining | Props | PDB | XML

Profile Lining and Properties

Physicochemical properties of lining side-chains
 Charge: 1 (2-1) Hydropathy: -2.00 Hydrophobicity: -0.70 Polarity: 24.21 Mutability: 81.00

Layer-weighted Physicochemical properties of lining side-chains
 Hydropathy: -2.38 Hydrophobicity: -0.85 Polarity: 25.11 Mutability: 89.00

	R	E	S	L	Q	K	S	G	K	Y	Rad	FRad	Dist	Hdry	Hdph	Pol
1	R	E									1.56	1.56	0.60	-4.00	-0.78	50.95
2	R	E	S								1.92	1.98	2.00	-2.93	-0.84	34.52
3	R	E	S	L							2.37	3.06	2.44	-1.25	-0.35	25.93
4	E	S	L								2.13	2.97	3.38	-0.17	-0.32	17.23
5	E	S	L	Q							2.12	2.98	3.98	-1.00	-0.52	13.81
6	E	S		Q							1.38	3.00	8.57	-2.60	-1.07	18.37
7	E	S		Q	K						2.39	3.30	9.61	-2.93	-0.91	26.15
8	E	S			K	S					2.23	2.84	11.40	-2.25	-0.87	25.69
9	E				K	S					2.17	2.44	13.34	-2.73	-0.84	33.69
10	E				K	S	G				2.46	2.46	13.56	-2.15	-0.83	26.11
11	E				S	G	K	Y			2.51	2.51	13.79	-1.28	-0.52	11.99

Lining residues for tunnel 2

1 Unique lining residues set - all
 212 ARG A, 308 GLU A, 312 SER A, 308 GLU A, 369 ILE A, 309 THR A, 482 LEU A, 483 LEU A, 369 ILE A, 371 MET A, 213 PHE A, 215 PHE A, 370 ALA A, 372 ARG A, 370 ALA A, 105 ARG A, 108 PHE A, 119 SER A, 106 ARG A, 120 ILE A, 120 ILE A, 107 PRO A, 122 GLU A, 107 PRO A, 111 VAL A

Unique lining residues set - sidechains
 212 ARG A, 308 GLU A, 312 SER A, 369 ILE A, 482 LEU A, 371 MET A, 215 PHE A, 370 ALA A, 105 A, 108 PHE A, 120 ILE A, 122 GLU A, 107 PRO A, 111 VAL A

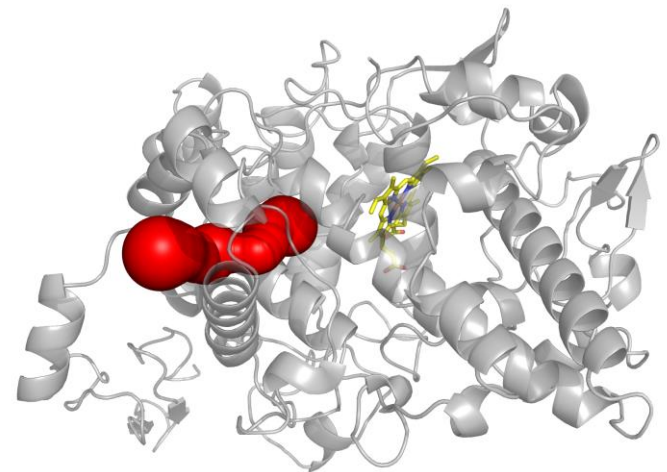
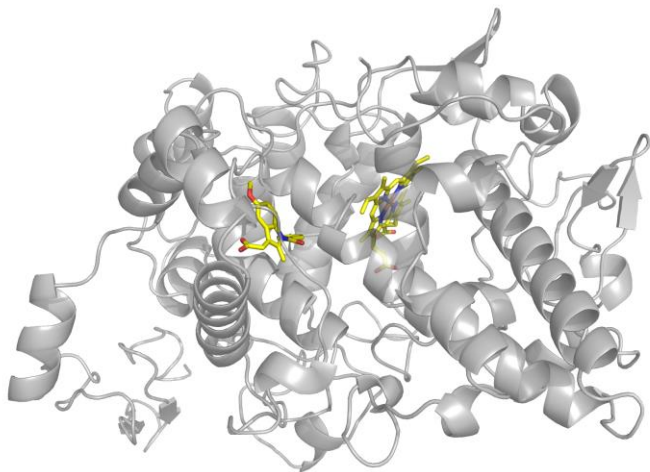
2 Physicochemical properties of lining side-chains
 Charge: 0 (2-2)
 Hydropathy: -0.1
 Hydrophobicity: 0.15
 Polarity: 16.52
 Mutability: 82

3 Lining residues

#	Res	Bltn	Dist	Hpa	Hpb	Pol	mut
1	<input type="checkbox"/> 212 ARG A, 308 GLU A	1.56	0.84	-4	-0.78	50.95	80
2	<input type="checkbox"/> 212 ARG A, 308 GLU A, 312 SER A	1.99	1.36	-2.93	-0.84	34.52	92
3	<input type="checkbox"/> 212 ARG A, 312 SER A, 308 GLU A, 369 ILE A	2.21	1.51	-0.3	-0.1	14.3	101
4	<input type="checkbox"/> 212 ARG A, 312 SER A, 308 GLU A, 369 ILE A, 309 THR A	2.26	1.97	-0.32	-0.24	12.11	101
5	<input type="checkbox"/> 212 ARG A, 312 SER A, 308 GLU A, 369 ILE A	2.34	2.26	-0.3	-0.1	14.3	101
6	<input type="checkbox"/> 212 ARG A, 312 SER A, 369 ILE A	2.49	2.49	-0.27	0.14	17.93	101
7	<input type="checkbox"/> 212 ARG A, 312 SER A, 369 ILE A, 482 LEU A	2.61	2.69	0.75	0.39	13.48	89
8	<input type="checkbox"/> 212 ARG A, 312 SER A, 369 ILE A, 482 LEU A, 483 LEU A	2.7	2.81	0.52	0.15	11.46	89
9	<input type="checkbox"/> 212 ARG A, 312 SER A, 369 ILE A, 482 LEU A	2.66	3.01	0.75	0.39	13.48	89

Where are my channels - I?

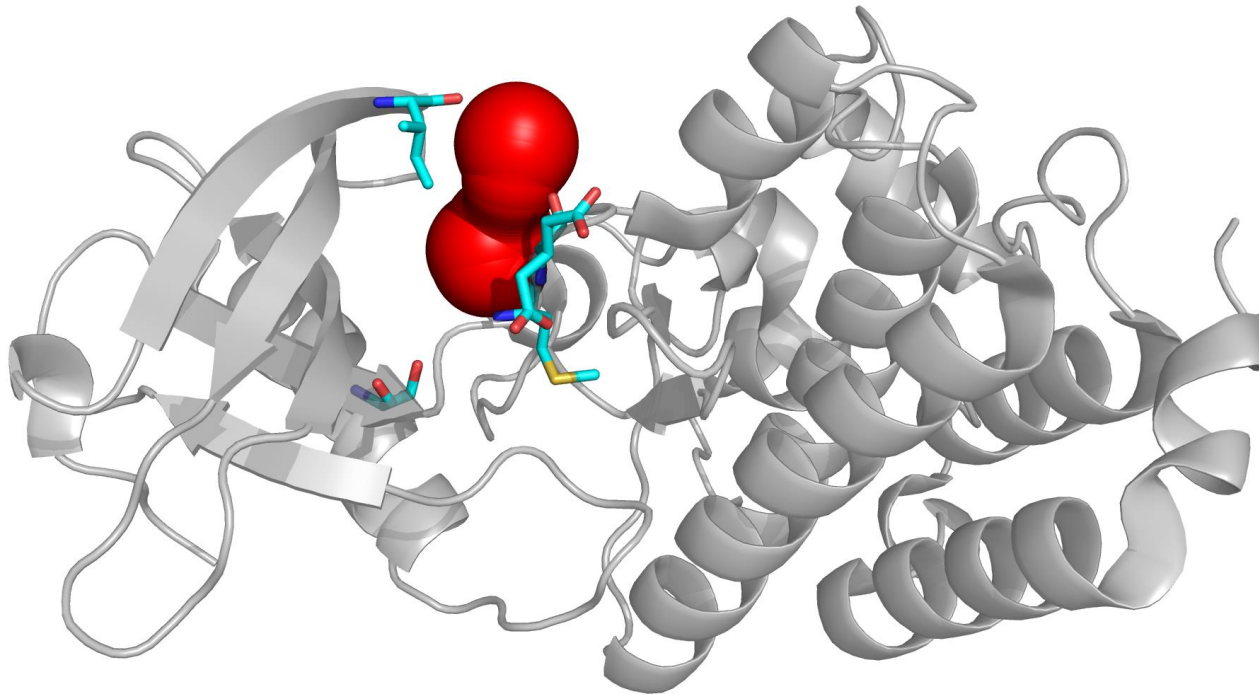
- Q: No channel has been identified ☹️ Why?
 - A: Wrong set up of **ProbeRadius**, **InteriorThreshold** or **Filtering** criteria.
 - A: Substrate is blocking channel



- Cyclooxygenase-2 (PDB ID: 4cox) complexed with non-selective inhibitor indomethacin

Where are my channels - II?

- Q: No channel has been identified ☹️ Why?
 - A: Active site is located on the surface on its vicinity



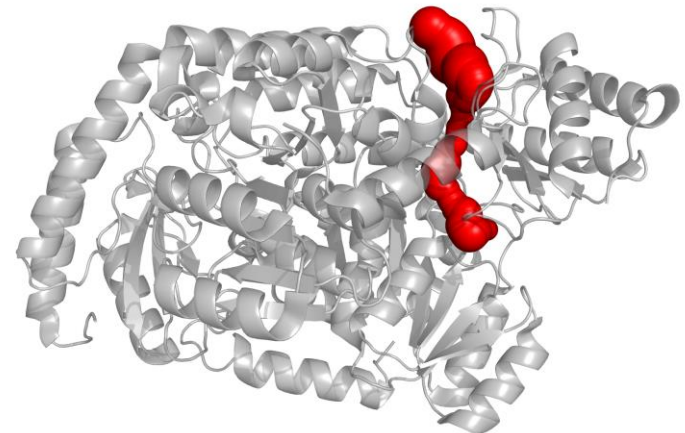
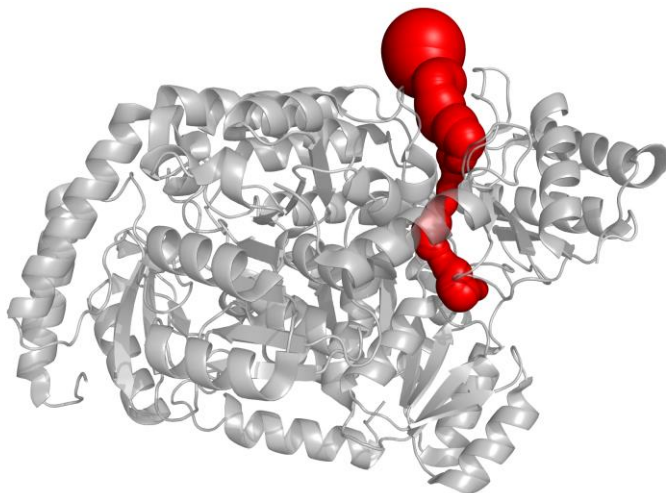
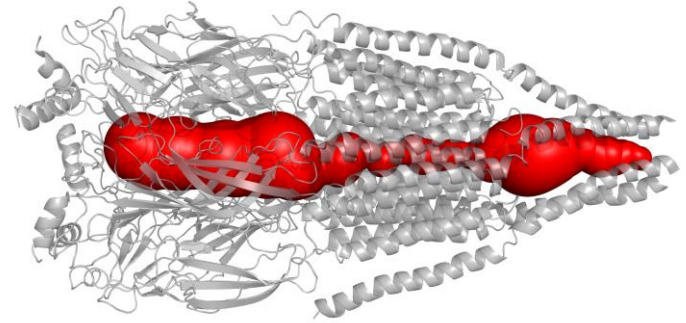
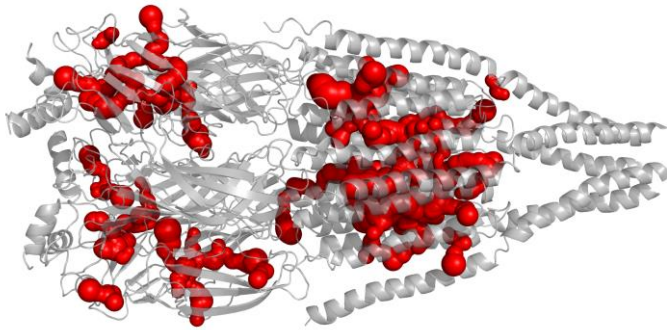
- Pocket-like channel found in tyrosine kinase EPh4 (PDB ID 2vwx)

Where are my channels - III?

- Q: No channel has been identified 😞 Why?
 - A: No channel is there whatsoever

Where are my channels - IV?

- Q: None of the found channels is relevant to me?
 - A: Multiple reasons. Usually wrong set up of **ProbeRadius** or **InteriorThreshold** parameters



Questions?



After a break we can continue with the hands-on experience!

<https://goo.gl/f5YrcE>



@WebchemTools



ncbr.muni.cz/webchemistry

