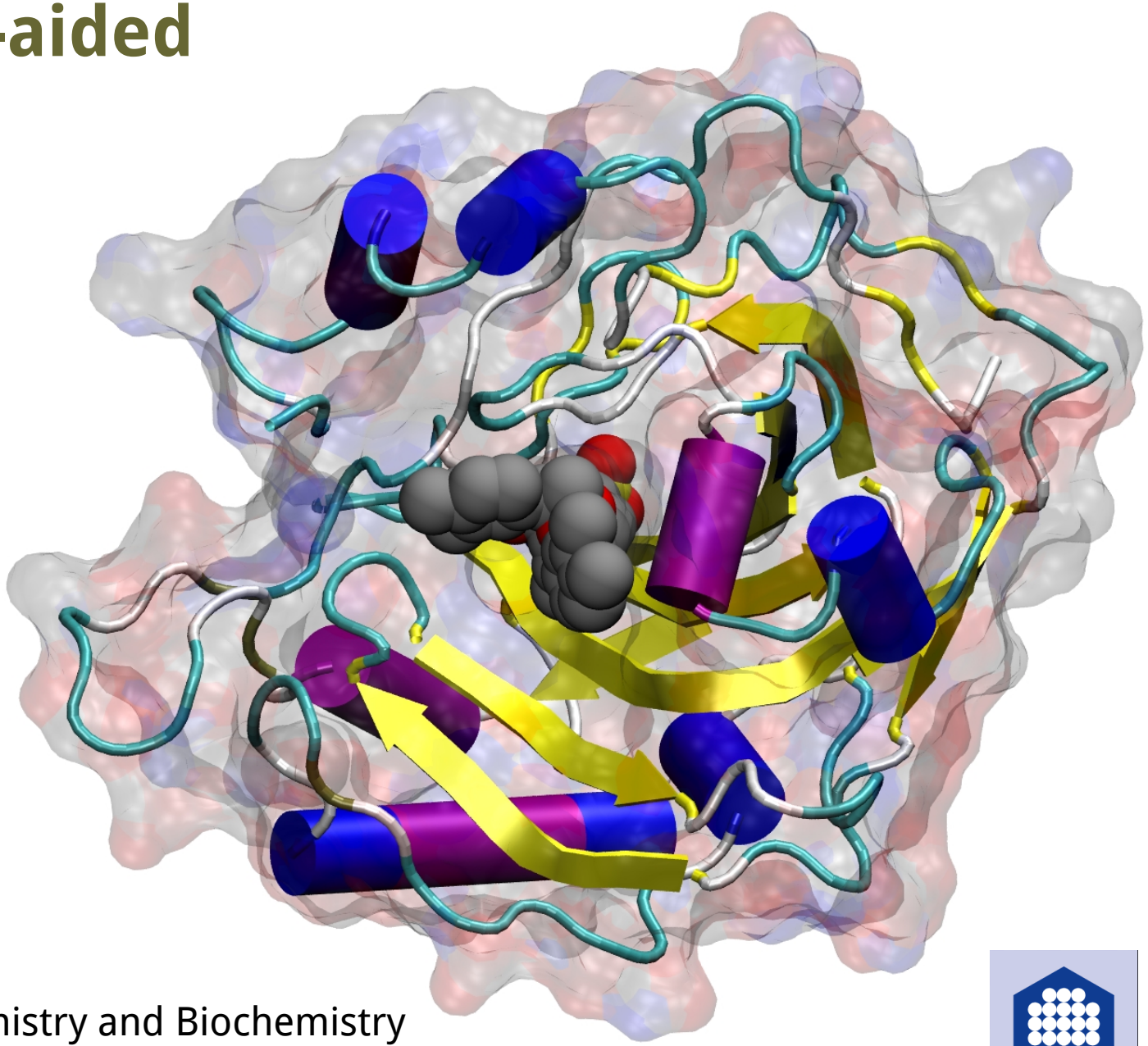
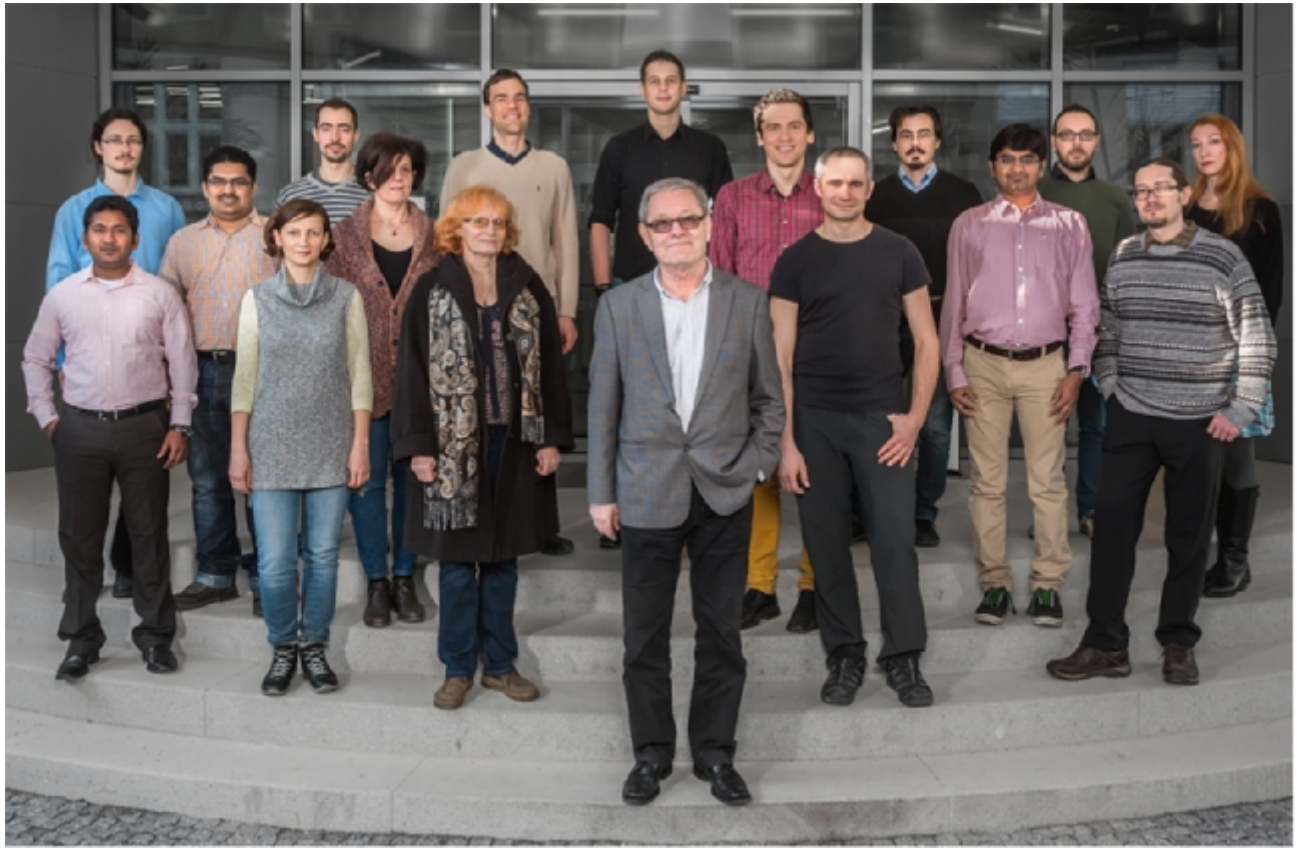
Large-scale quantum-mechanical calculations for computer-aided drug design



Jan Řezáč

Institute of Organic Chemistry and Biochemistry
Academy of Sciences of the Czech Republic

Pavel Hobza group

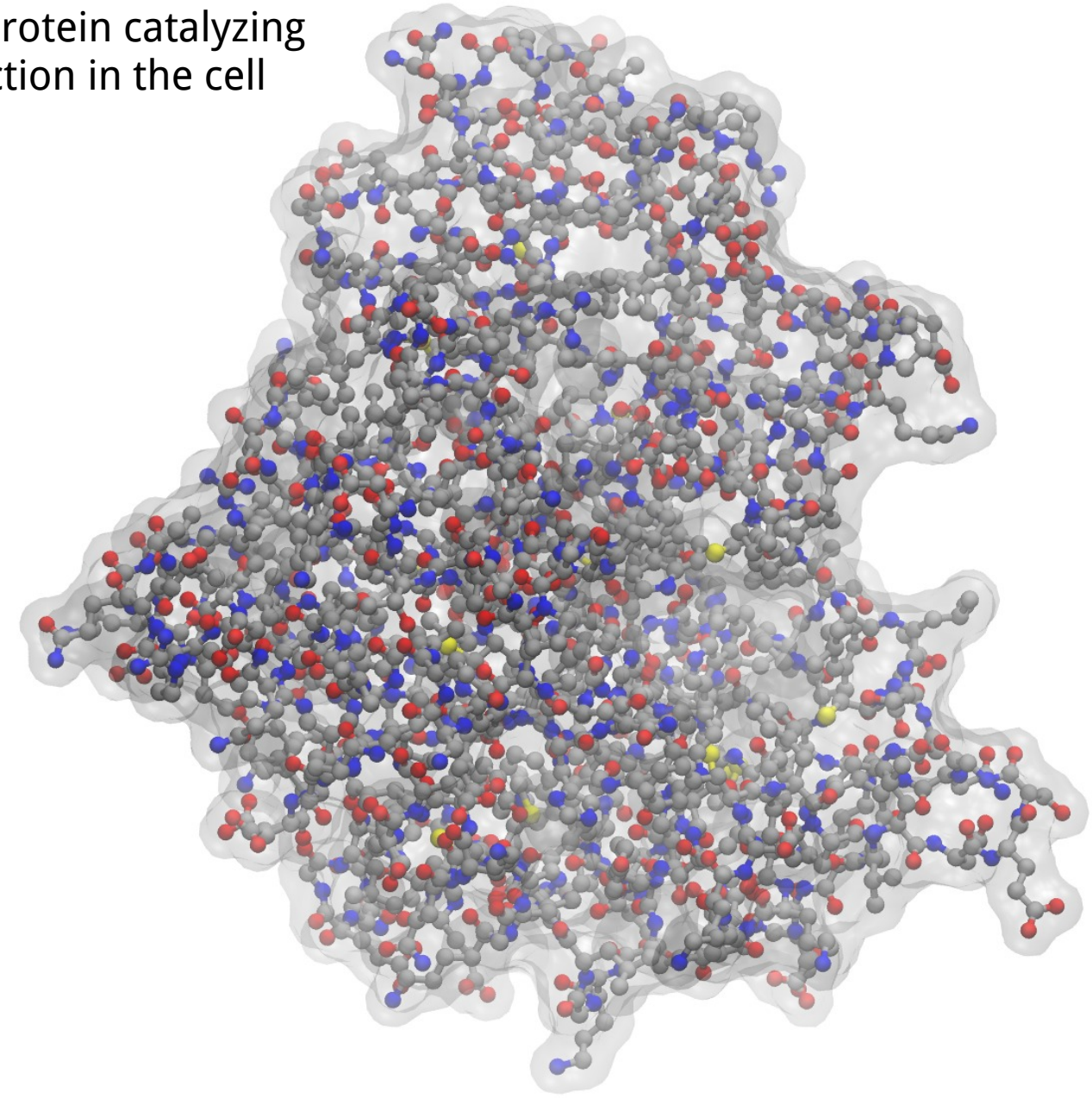


Pavel Hobza
Martin Lepšík
Jindřich Fanfrlík
Jan Řezáč
Adam Pecina
Cemal Köprülüoğlu
Saltuk Eyrilmez
Haresh Ajani

Institute of Organic Chemistry and Biochemistry
Academy of Sciences of the Czech Republic
Prague

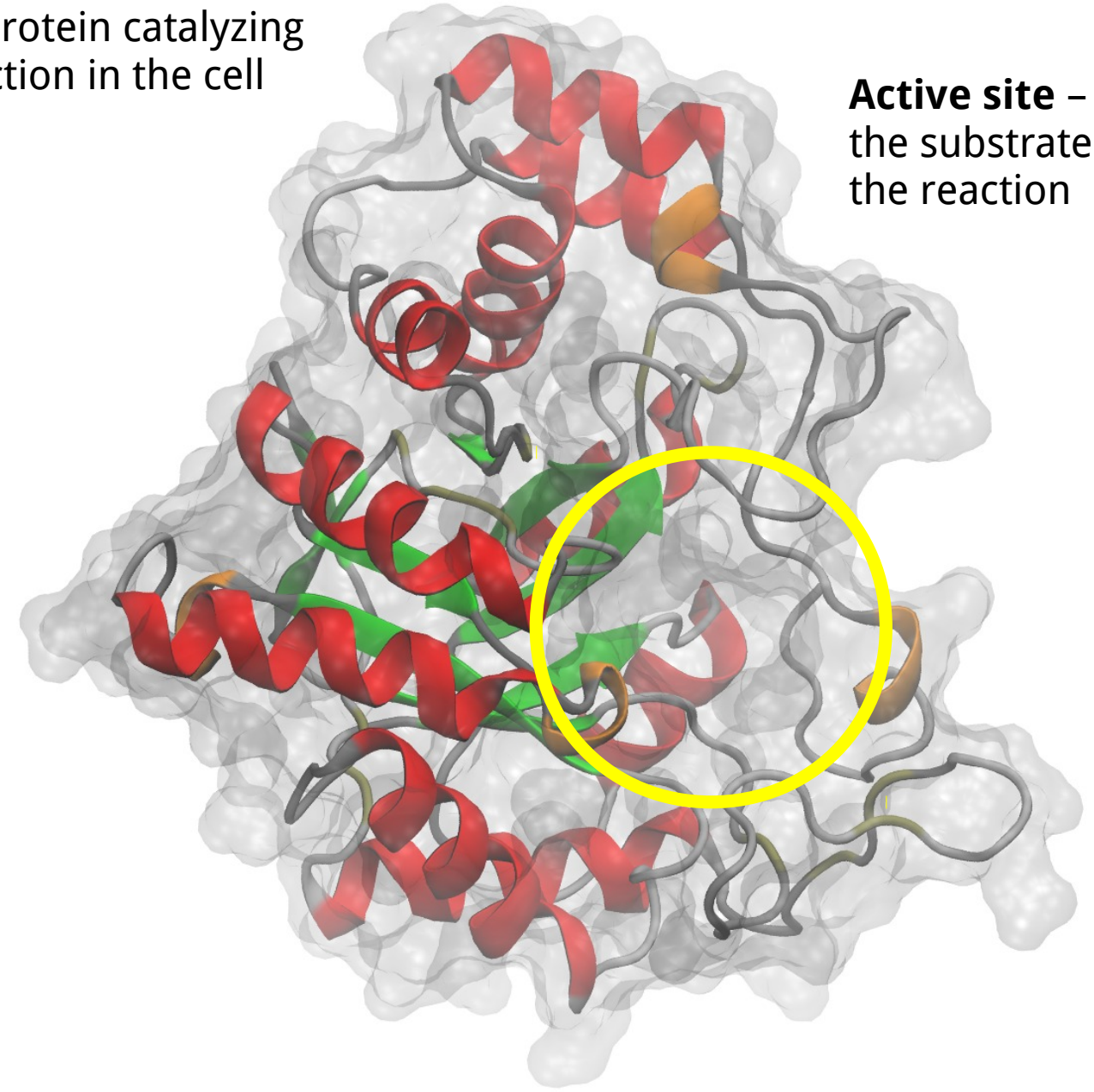
Enzyme inhibition

Enzyme – a protein catalyzing chemical reaction in the cell



Enzyme inhibition

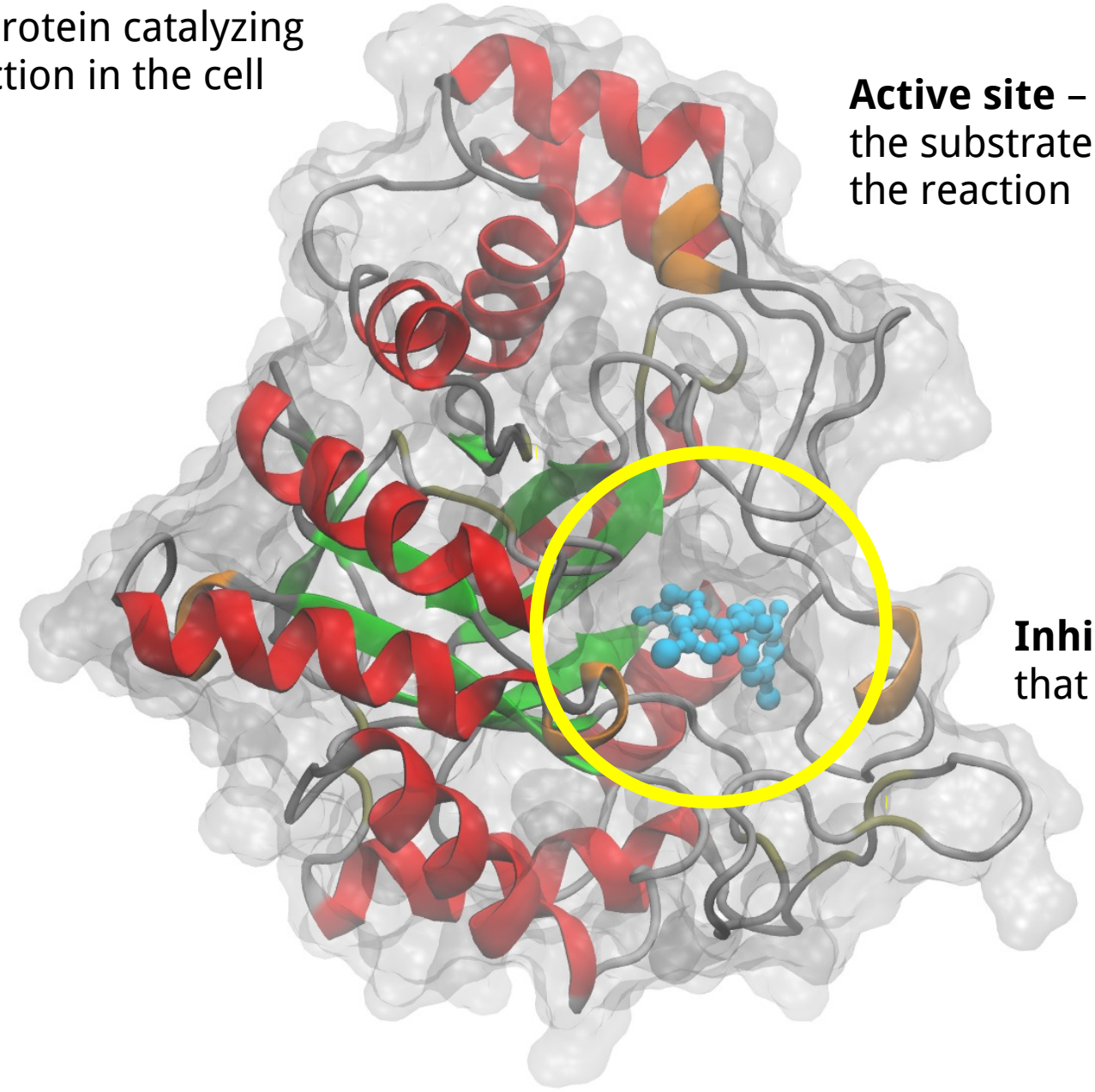
Enzyme – a protein catalyzing chemical reaction in the cell



Active site – cavity that binds the substrate and enables the reaction

Enzyme inhibition

Enzyme – a protein catalyzing chemical reaction in the cell



Active site – cavity that binds the substrate and enables the reaction

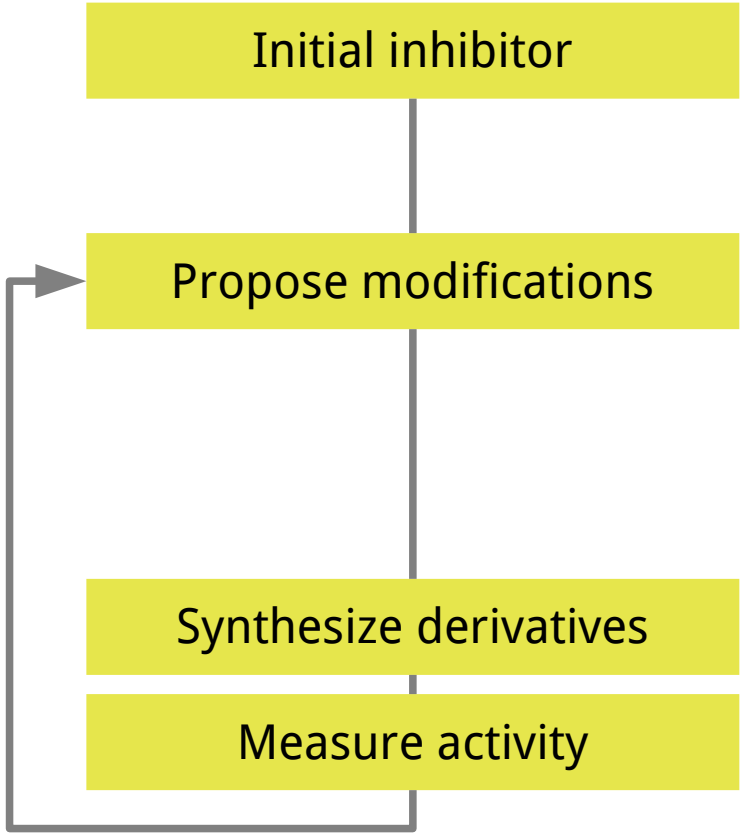
Inhibitor – molecule that blocks the active site

Rational drug design

- Identify metabolic pathways involved in the disease
- Identify structure and mechanism of a key enzyme
- Design an inhibitor of this enzyme
  - binds stronger than the substrate
  - specific to the target
  - can be delivered into the cell
  - is not toxic

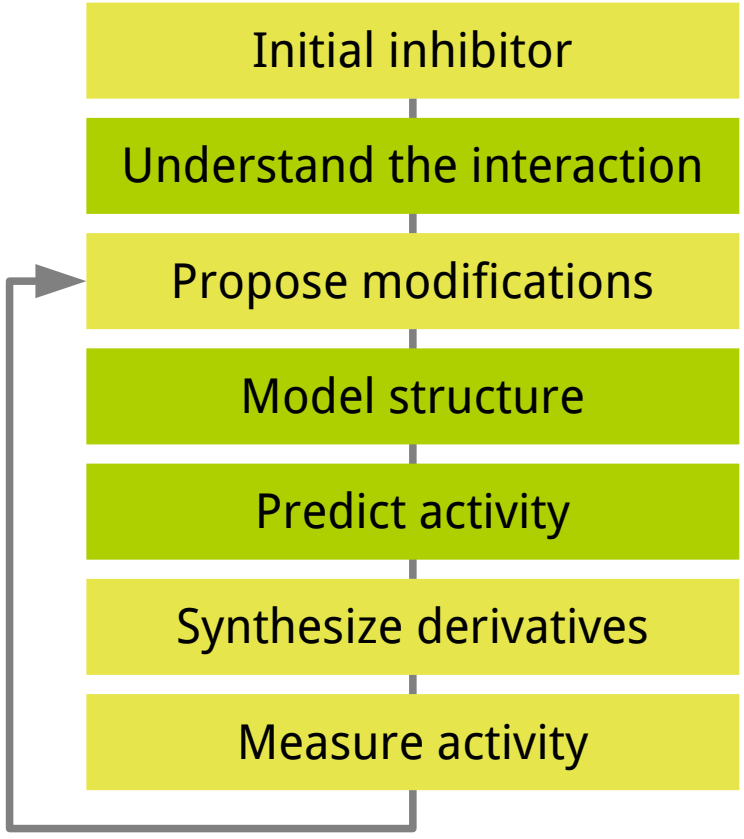
Rational drug design

- Identify metabolic pathways involved in the disease
- Identify structure and mechanism of a key enzyme
- Design an inhibitor of this enzyme



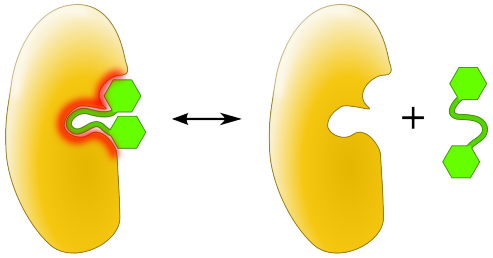
Computer-aided drug design

- Identify metabolic pathways involved in the disease
- Identify structure and mechanism of a key enzyme
- Design an inhibitor of this enzyme

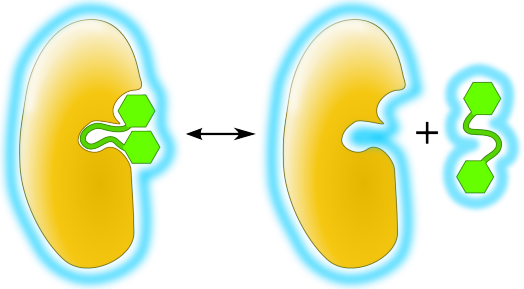


Modeling protein-ligand binding

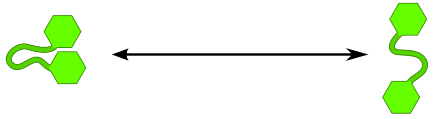
- Protein structure – X-ray data + modeling
- Geometry of the protein-ligand complex – **docking**
- Binding free energy calculation – **scoring**



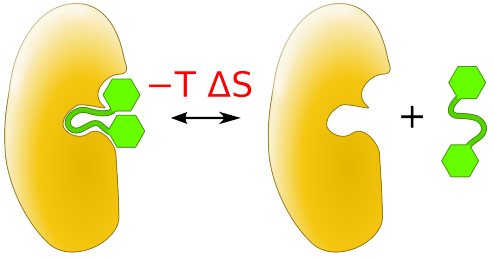
Binding energy



Solvation



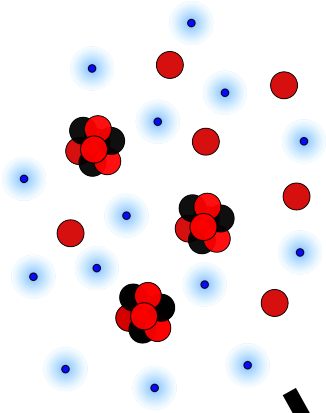
Geometry changes



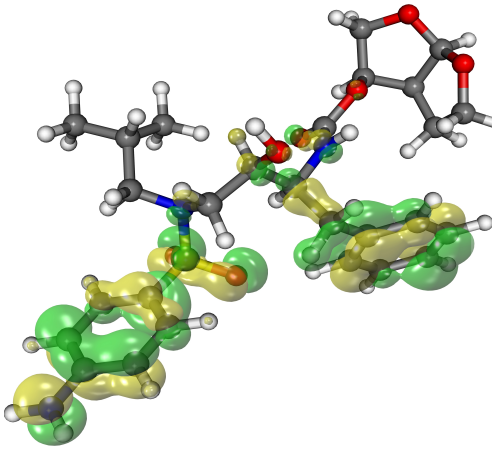
Entropy

Computational approaches

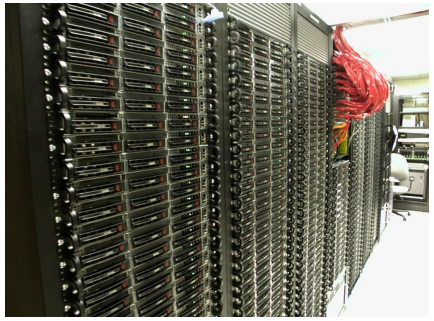
- Statistics-based models
- Molecular mechanics – approximate empirical models
- Quantum-mechanical calculations



$$\hat{H}\Psi = E\Psi$$



$$E = 2\sum_{i=1}^{N} H_{ii} + \sum_{i=1}^{N}\sum_{j=1}^{N}\left(2J_{ij} - K_{ij}\right)$$



theoretical approximations

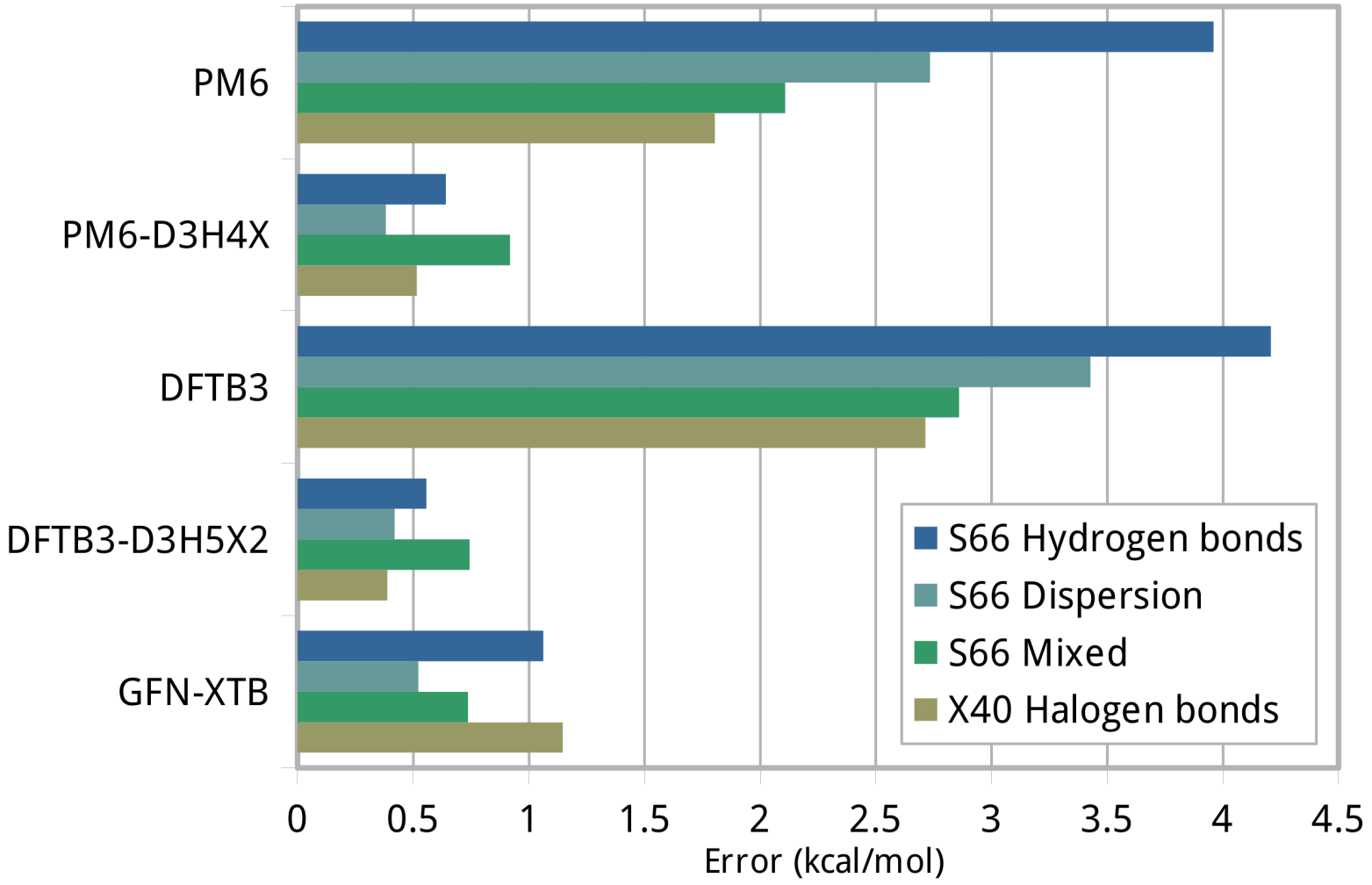
numerical calculations

Semiempirical QM methods

- Full QM to demanding for large-scale applications
- Semiempirical methods
  - Approximations compensated with empirical parameters
  - 1000s of atoms computed in minutes
  - No system-specific parameterization
  - Covers all phenomena using appropriate physics
- Limitations
  - Poor description of non-covalent (intermolecular) interactions

Method development

- Corrections for PM6 and DFTB methods
- Fixing London dispersion, hydrogen and halogen bonds



Our method development

- Corrections for description of non-covalent interactions[1-5]
- Benchmark calculations for their parametrization[6]
- Solvation model improvements

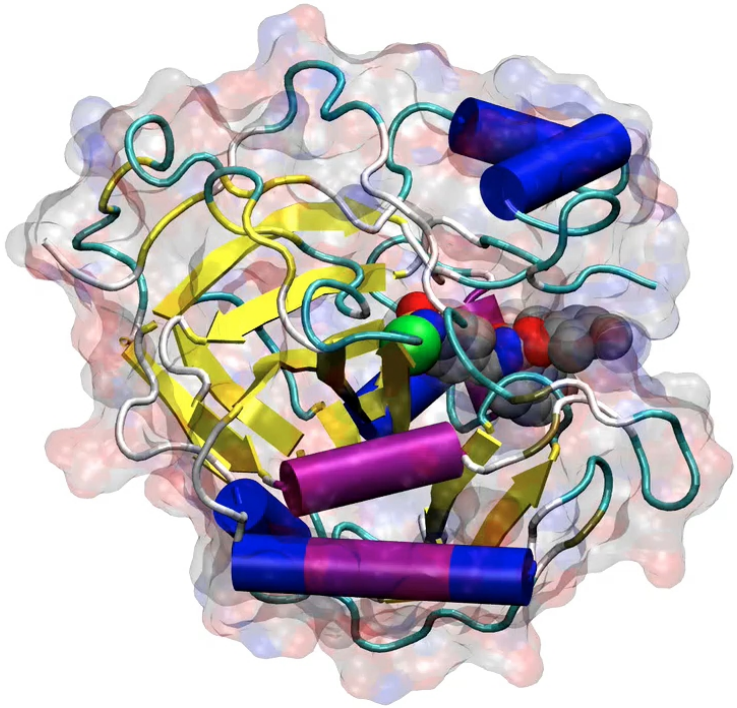
- Computational drug design protocol based on SQM[7]
- Software framework for automating the calculations[8]

1) Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. J. Chem. Theory Comput. 2009, 5 (7), 1749–1760.
2) Korth, M.; Pitoňák, M.; Řezáč, J.; Hobza, P. J. Chem. Theory Comput. 2010, 6 (1), 344–352.
3) Řezáč, J.; Hobza, P. Chem. Phys. Lett. 2011, 506 (4-6), 286–289.
4) Řezáč, J.; Hobza, P. J. Chem. Theory Comput. 2012, 8 (1), 141–151.
5) Hostaš, J.; Řezáč, J.; Hobza, P. Chem. Phys. Lett. 2013, 568–569, 161–166.
6) Řezáč, J.; Hobza, P. Chem. Rev. 2016, 116 (9), 5038–5071.
7) Lepšík, M.; Řezáč, J.; Kolář, M.; Pecina, A.; Hobza, P.; Fanfrlík, J. ChemPlusChem 2013, 78 (9), 921–931.
8) http://cuby4.molecular.cz

Virtual screening capabilities

- Docking
  - large databases of compunds
  - **10^6 structures** generated
- Fast filtering – SQM scoring on fixed geometry
  - from 5 minutes
  - **100 000 calculations** / project
- Full scoring – including SQM optimization
  - few days
  - **1000 calculations** / project

Carbonic anhydrase II



- Perfect model with accurate experimental data for 10 inhibitors[1]
- Zinc metaloprotein

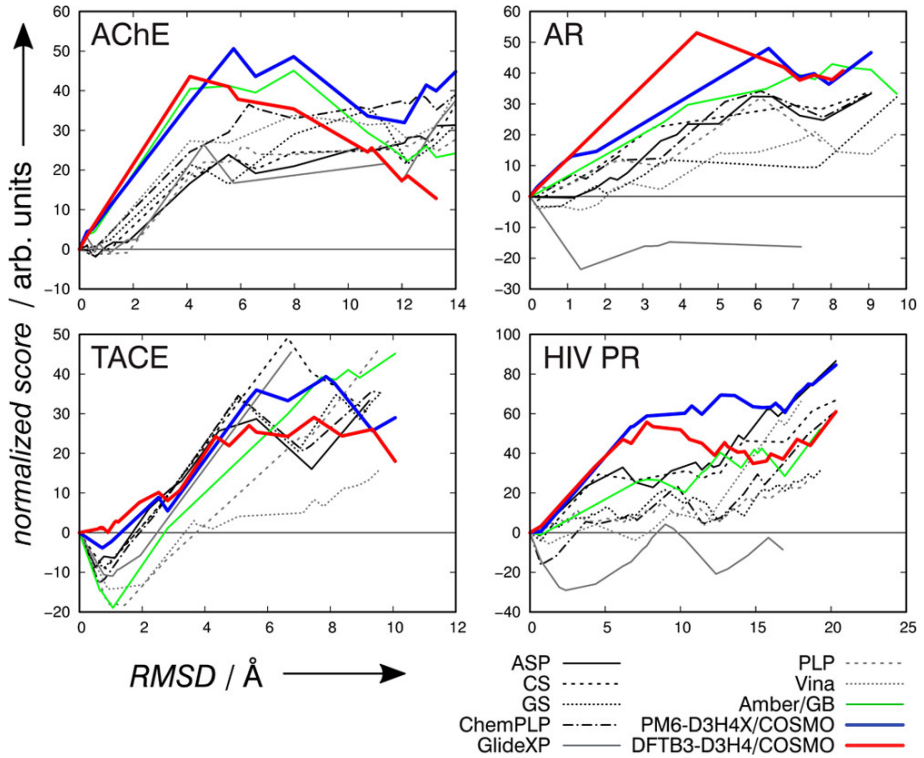
- New solvation model COSMO2[2]

Crystal waters	YES		NO	
	R^2	PI	R^2	PI
SQM/COSMO	0.58	0.87	0.45	0.73
SQM/COSMO2	0.68	0.91	0.69	0.75
AMBER/GB	0.09	0.45	0.13	0.3
Vina	0.39	0.67	0.07	0.2
DOCK 6	0.38	0.5	0.1	-0.22
Autodock 4	0.19	0.53	0.09	0.06
Gold ASP	0.15	0.47	0.12	0.35
GoldPLP	0.12	0.39	0.13	0.39
GoldScore	0.01	0.15	0.01	-0.1
Chemscore	0.01	0.08	0.01	-0.1

1) Pecina A., Řezáč, J. et al. ChemPhysChem 2018.
2) Kříž K., Řezáč J.; submitted manuscript

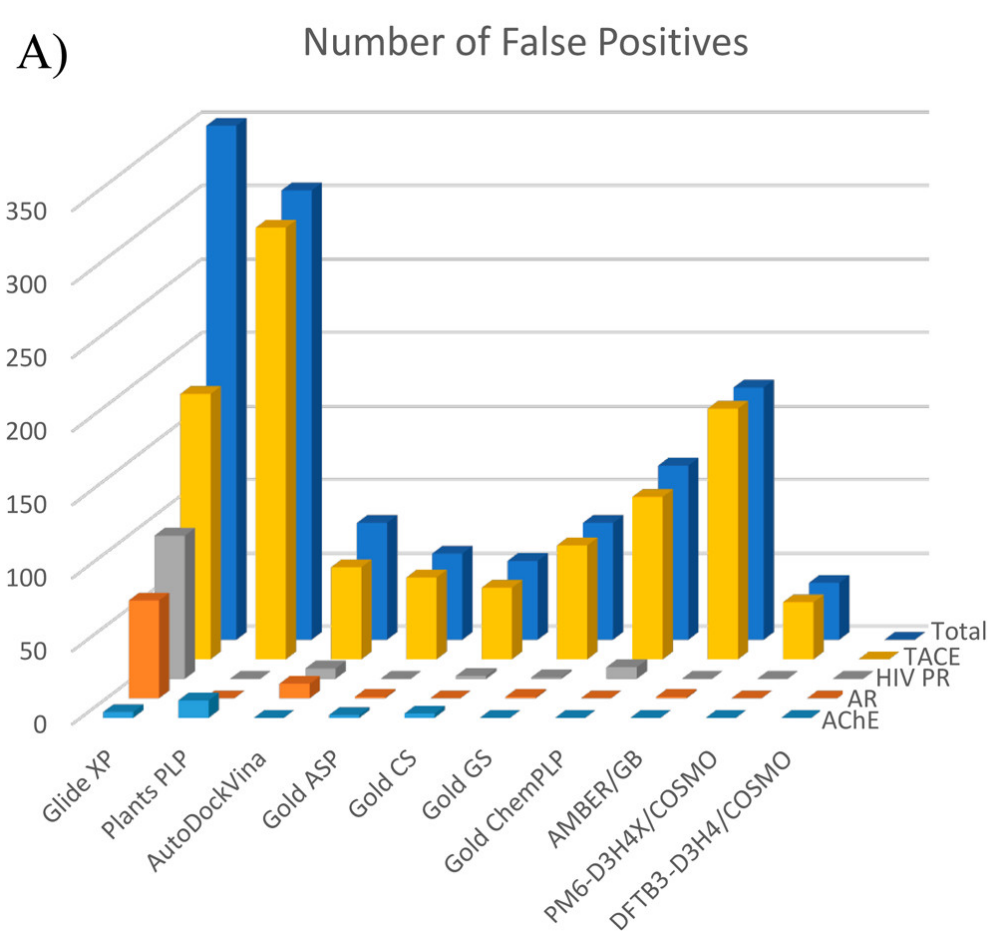
Identification of Native Protein–Ligand Poses

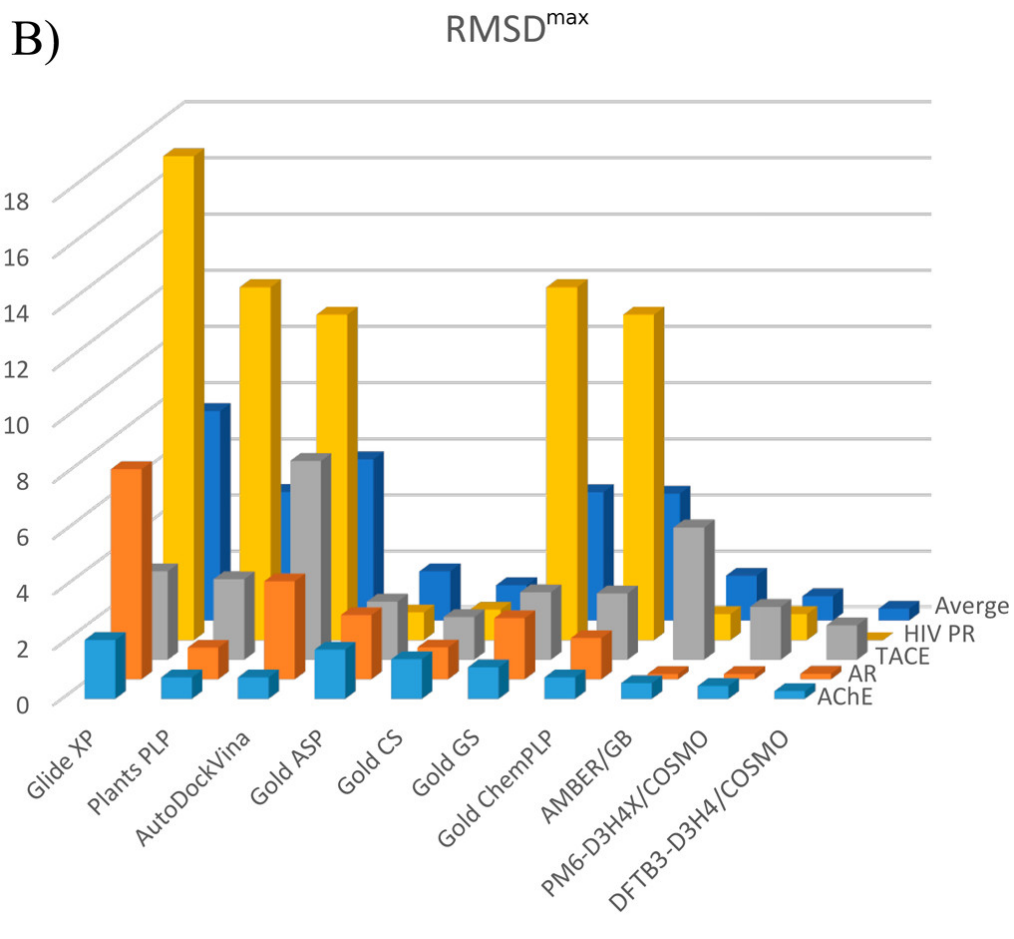
- Can SQM identify native geometry of P-L complex?
- large set of structures from docking
- Native pose should be energy minimum
- SQM tested against multiple scoring functions used in the field



1) Pecina, A.; Haldar, S.; Fanfrlík, J.; Meier, R.; Řezáč, J.; Lepšík, M.; Hobza, P. J. Chem. Inf. Model. 2017.

Identification of Native Protein–Ligand Poses





1) Pecina, A.; Haldar, S.; Fanfrlík, J.; Meier, R.; Řezáč, J.; Lepšík, M.; Hobza, P. J. Chem. Inf. Model. 2017.

Conclusions

- SQM method perform significantly better than standard scoring functions and molecular mechanics
- The computational cost is well justified
- Synergy between method development and applications

GAČR
GRANTOVÁ AGENTURA ČESKÉ REPUBLIKY

IT4Innovations
národní
superpočítačové
centrum