



# CEITEC



Central European Institute of Technology  
BRNO | CZECH REPUBLIC

## Structural Bioinformatics and Molecular Modeling

Jaroslav Koča

[jkoca@ceitec.cz](mailto:jkoca@ceitec.cz)



EUROPEAN UNION  
EUROPEAN REGIONAL DEVELOPMENT FUND  
INVESTING IN YOUR FUTURE



OP Research and  
Development for Innovation





# CEITEC



Central European Institute of Technology  
BRNO | CZECH REPUBLIC

## Structural bioinformatics



EUROPEAN UNION  
EUROPEAN REGIONAL DEVELOPMENT FUND  
INVESTING IN YOUR FUTURE

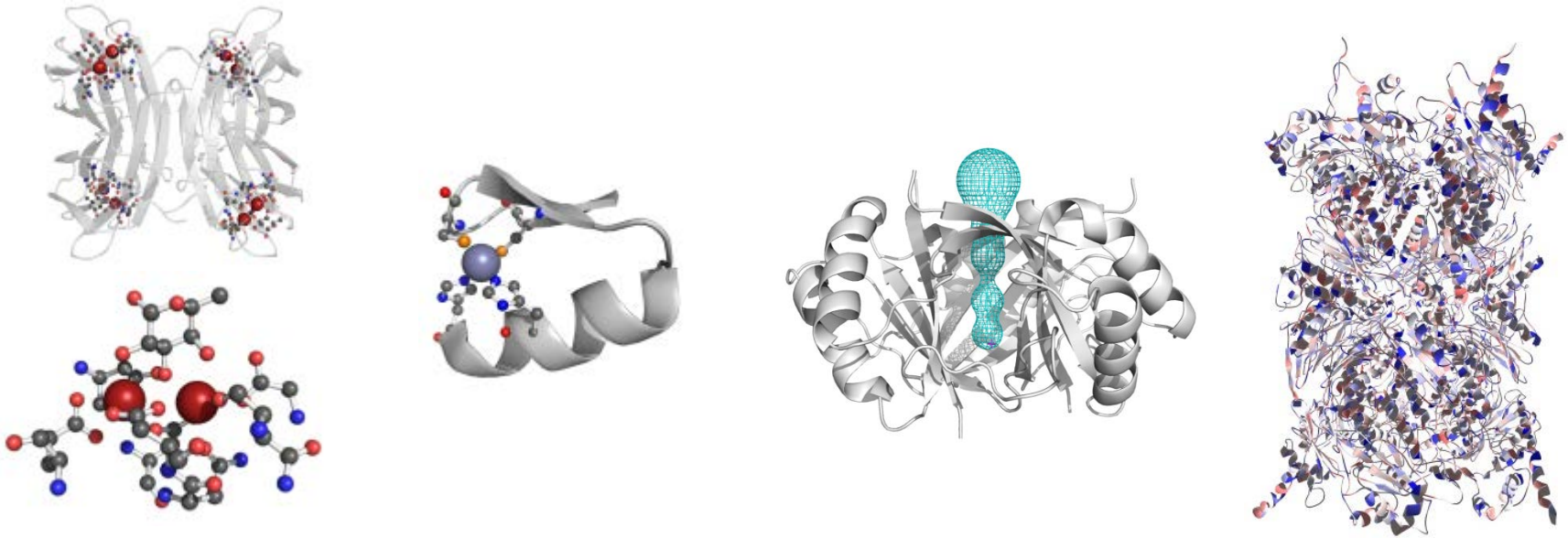


**OP Research and  
Development for Innovation**



# Mission

**Input:** Information about biomacromolecular structure



**Outputs:**

- Understanding of biomacromolecular function
- Prediction of structural change influence
- Classification of biomacromolecules
- Understanding of relations between biomacromolecules

# Model of a molecule in computer

## Atoms:

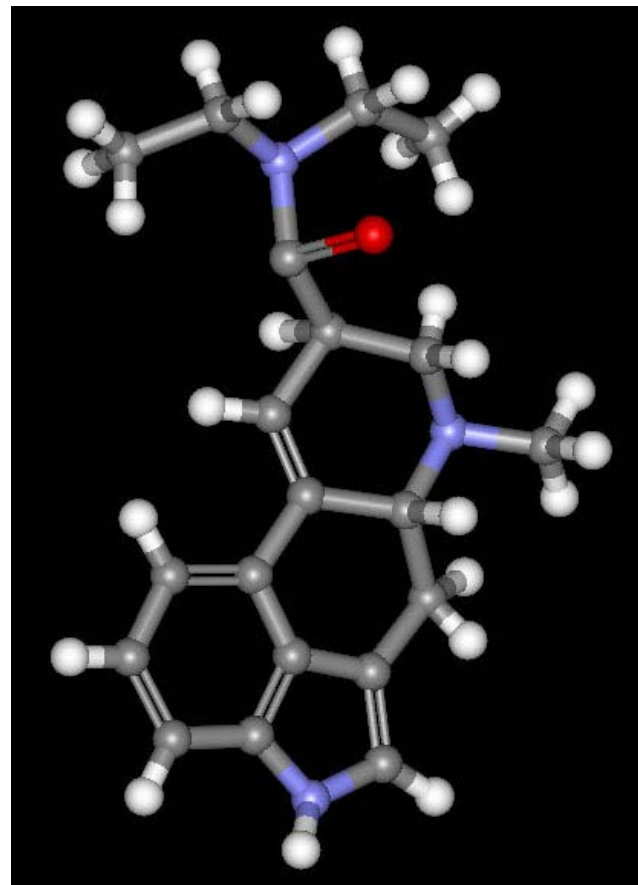
Points in a 3D space

Information about a chemical element

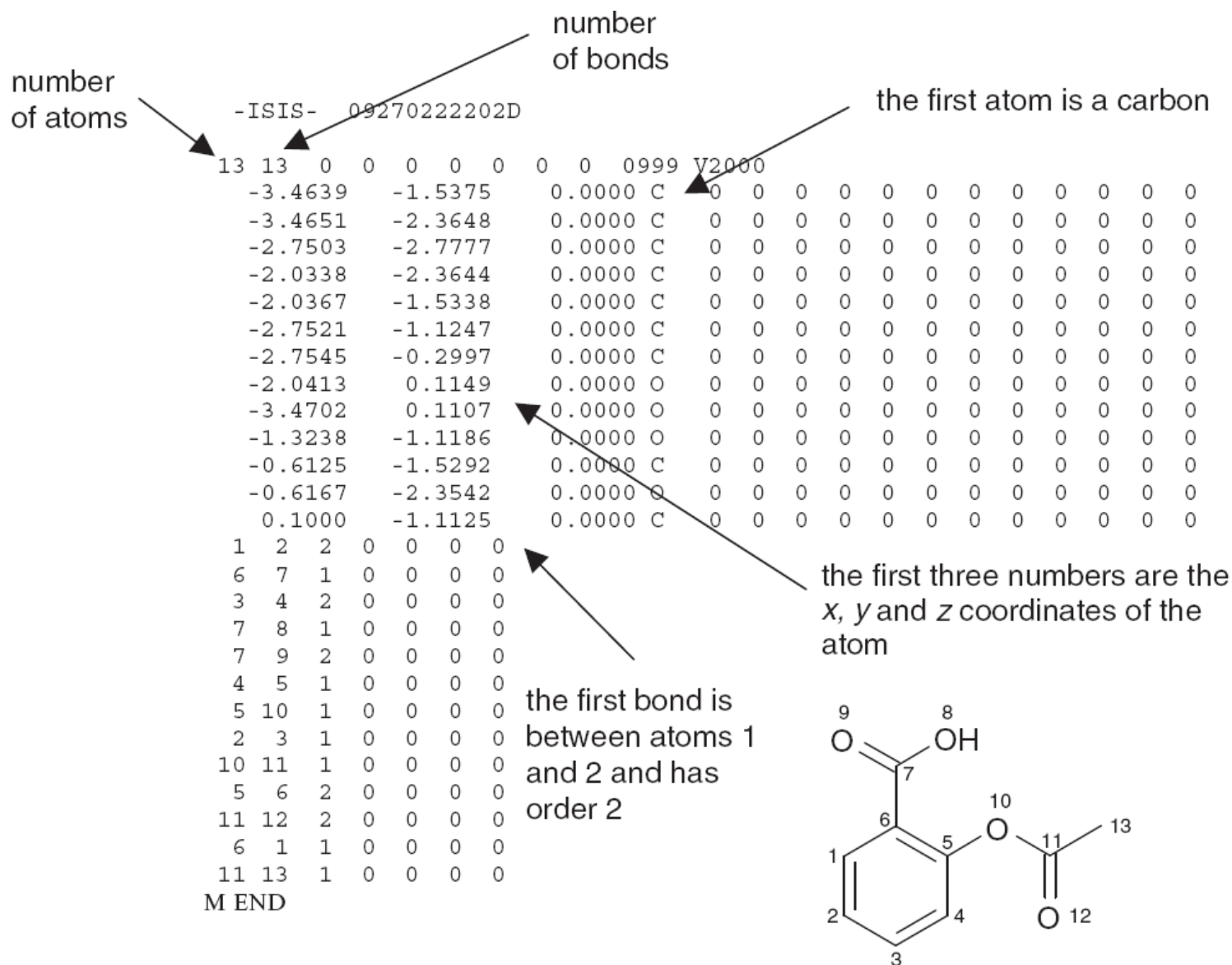
## Bonds:

Information about two bound atoms

Bond order (single, double, triple, ...)



# Molecule in a computer – MOL format



# Current databases of bio(macro)molecules

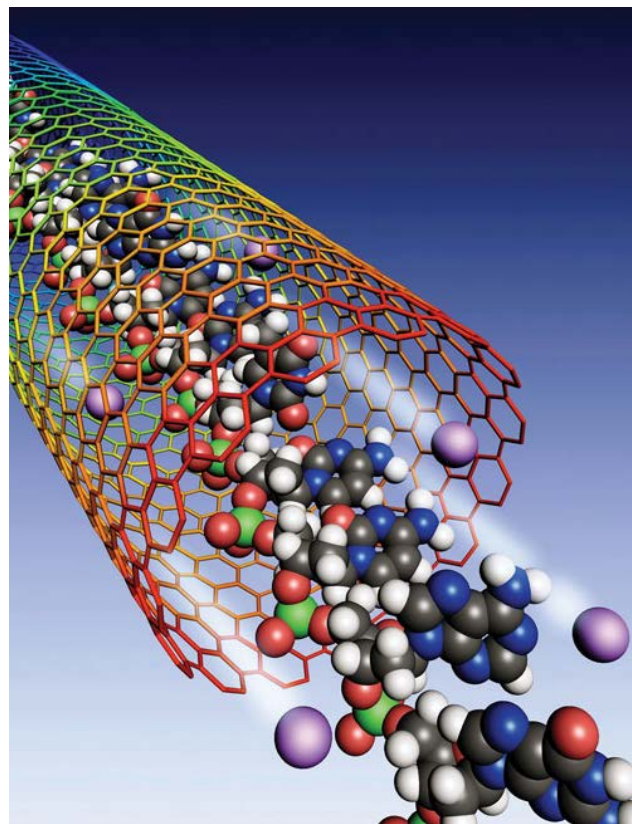
„**Information boom**“ in the field of bio(macro)molecular structures

**Why?:** High performance techniques of structural analysis were developed.

## **Results:**

- We are able to get information about a genome of a person in a week and for low price
- More than 120 000 structures of proteins and nucleic acids are available in Protein Data Bank
- Millions of structures of small chemical compounds (drug-like molecules, ligands...) are stored in Pubchem, Zinc, Drugbank and other database

**Most of this information is publicly and freely accessible:-)**





# Protein Data Bank

## – example of biomacromolecular structure DB

RCSB PDB

Deposit

Search

Visualize

Analyze

Download

Learn

More

MyPDB Login

RCSB PDB

PROTEIN DATA BANK

An information Portal to  
119480 Biological  
Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands

Go

Advanced Search | Browse by Annotations

PDB-101

EMDBank

StructuralBio

Worldwide Protein Data Bank Foundation

Facebook

Twitter

YouTube

LinkedIn

Welcome

Deposit

Search

Visualize

Analyze

Download

Learn

A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Insulin and Diabetes

Poster

All Resources

June Molecule of the Month



Beta-galactosidase

Latest Entries

As of Tuesday Jun 07



5JVD

PDB Entry

Tubulin-TUB092 complex

New Features

February 2016 Release



Redesigned Search Results Page

New Organization. Improved Layout. Clean. Usable. Faster Load.

Electron Density Map Visualization

Genetic Variation Track on Protein Feature View

Wild Type Search

News

Publications



Join our Development Team in San Diego

Looking for a Postdoctoral Fellow to develop innovative tools for analysis, integration, query, and visualization of the 3D biomolecular structure information found in the PDB. » 06/07/16

Vote Now! » 05/31/16

Join the Sequence-Structure Hackathon at ISMB Orlando 2016 » 05/04/16

AutoDoc Annotation System to

Contact Us

# Protein Data Bank

## – example of biomacromolecular structure DB II

RCSCB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

**RCSCB PDB** An Information Portal to 119480 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands **Go**

Advanced Search | Browse by Annotations | Search History (2) | Previous Results (136)

PDB-101 EMBL-EBI EMDatabank Structural Biology Knowledgebase Worldwide Protein Data Bank Foundation

Structure Summary 3D View Annotations Sequence Sequence Similarity Structure Similarity Experiment Literature

Biological Assembly 1

**1FDH**

STRUCTURE OF HUMAN FOETAL DEOXYHAEMOGLOBIN

DOI: 10.2210/pdb1fdh/pdb

Classification: **OXYGEN TRANSPORT**

Deposited: 1976-08-18 Released: 1976-08-19

Deposition author(s): [Frierjunior, J.A.](#)

Organism: [Homo sapiens](#)

Structural Biology Knowledgebase: 1FDH (2 models >17 annotations) [SBKB.org](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.5 Å

R-Value Work:

wwPDB Validation [Full Report](#)

Metric	Percentile Ranks	Value
Clashscore		35
Ramachandran outliers		2.5%
Sidechain outliers		17.9%

Legend:   
█ Percent outliers to all X-ray structures   
█ Percent outliers to X-ray structures of similar resolution

Literature [Download Primary Citation](#)

Structure of human foetal deoxyhaemoglobin.

[Frier, J.A., Perutz, M.F.](#)

(1977) J.Mol.Biol. 112: 97-112

PubMed: [881729](#) [Search on PubMed](#)

View in 3D: [JSmol](#) or [PV](#) (in Browser)

Standalone Viewers

[Simple Viewer](#) [Protein Workshop](#)

[Ligand Explorer](#) [Kiosk Viewer](#)

Protein Symmetry: Cyclic - C2 ([View in 3D](#))

Protein Stoichiometry: Hetero 4-mer - A2B2

Biological assembly 1 assigned by authors

Macromolecule Content

- Unique protein chains: 2

<http://www.rcsb.org/pdb/explore.do?structureId=1FDH#carousel-structuregallery> | es

Contact Us



# DrugBank

## – example of drug DB

**DRUGBANK** Browse ▾ Search ▾ Downloads About ▾ Help ▾ Contact Us

Search Drugs 🔍

Get DrugBank to go! The DrugBank app for iOS and Android is coming soon. [Sign up to get early access](#)

# DRUGBANK

Drug & Drug Target Database

### DrugBank Version 4.5

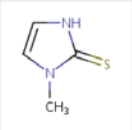
The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. The database contains 8206 drug entries including 1991 FDA-approved small molecule drugs, 207 FDA-approved biotech (protein/peptide) drugs, 93 nutraceuticals and over 6000 experimental drugs. Additionally, 4333 non-redundant protein (i.e. drug target/enzyme/transporter/carrier) sequences are linked to these drug entries. Each DrugCard entry contains more than 200 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data. [More about DrugBank](#)

DrugBank is offered to the public as a freely available resource. Use and re-distribution of the data, in whole or in part, for commercial purposes (including internal use) requires a license. We ask that users who download significant portions of the database cite the DrugBank paper in any resulting publications.

Citing DrugBank:

Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. *DrugBank: a comprehensive resource for in silico drug discovery and exploration*. Nucleic Acids Res. 2006 Jan 1;34(Database issue):D668-72. [16381955](#)

Drug of the day: **Methimazole**




A thioureylen antithyroid agent that inhibits the formation of thyroid hormones by interfering with the incorporation of iodine into tyrosyl residues of thyroglobulin. This is done by interfering with the oxidation of iodide ion and iodotyrosyl groups through inhibition of the peroxidase enzyme. [PubChem]

For the treatment of hyperthyroidism, goiter, Graves disease and psoriasis.


[Learn more about Methimazole](#)

PEOPLE RECENT POPULAR


Recent Comments



Ross Barnard  
Thanks Adam :)  
DrugBank: ado-trastuzumab emtansine (DB05773) · 1 week ago



Adam Maciejewski  
Hi Ross, we have since placed this drug into the biotech class.  
DrugBank: ado-trastuzumab emtansine (DB05773) · 1 week ago



Ross Barnard  
Thankyou Adam  
DrugBank: ado-trastuzumab emtansine (DB05773) · 2 weeks ago

# DrugBank

## – example of drug DB II

DrugBank

Browse Search Downloads About Help Contact Us

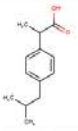
Search Drugs

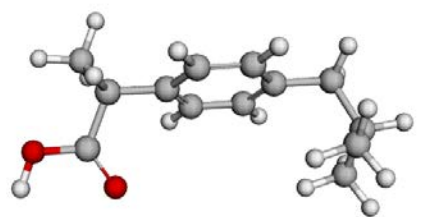
Identification Taxonomy Pharmacology ADMET Pharmacoeconomics Properties Spectra References Interactions 2 Comments

Targets (6) Enzymes (10) Carriers (1) Transporters (8) Biointeractions (24) Show Drugs with Similar Structures

Get DrugBank to go! The DrugBank app for iOS and Android is coming soon. Sign up to get early access

### Identification

Name	Ibuprofen
Accession Number	DB01050 (APRD00372)
Type	Small Molecule
Groups	Approved
Description	Ibuprofen, a propionic acid derivative, is a prototypical nonsteroidal anti-inflammatory agent (NSAIA) with analgesic and antipyretic properties.
Structure	 <a href="#">MOL</a> <a href="#">SDF</a> <a href="#">3D-SDF</a> <a href="#">PDB</a> <a href="#">SMILES</a> <a href="#">InChI</a> <a href="#">View 3D Structure</a>
Synonyms	<ul style="list-style-type: none"><li>(+)-2-(P-Isobutylphenyl)propionic acid</li><li>(+)-alpha-Methyl-4-(2-methylpropyl)benzeneacetic acid</li><li>(+)-Ibuprofen</li><li>(+)-P-Isobutylhydratropic acid</li><li>(4-Isobutylphenyl)-alpha-methylacetic acid</li><li>(RS)-ibuprofen</li><li>2-(4-Isobutylphenyl)propanoic acid</li><li>4-Isobutylhydratropic acid</li><li>Adran</li><li>Advil</li><li>alpha-(4-Isobutylphenyl)propionic acid</li><li>alpha-(P-Isobutylphenyl)propionic acid</li><li>Amibufen</li><li>Anco</li></ul>

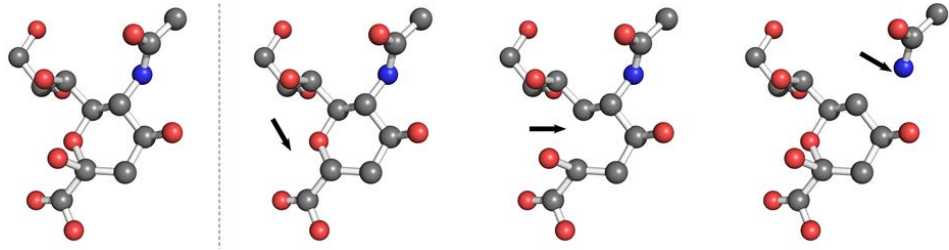


Download: [MOL](#) [SDF](#) [3D-SDF](#) [PDB](#) [SMILES](#) [InChI](#)

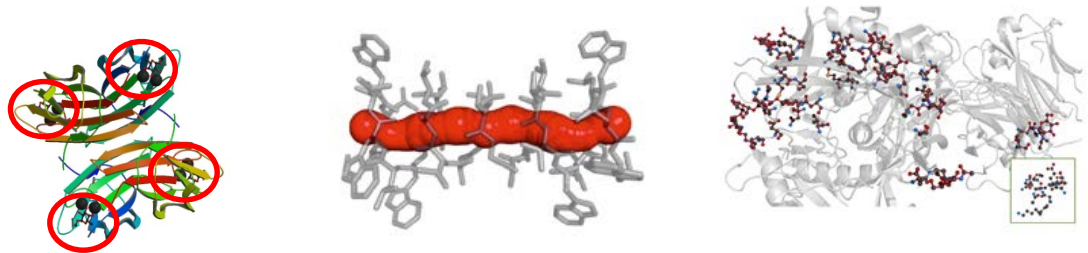
# Methodologies of structural bioinformatics

## - processing and analysis of the structures

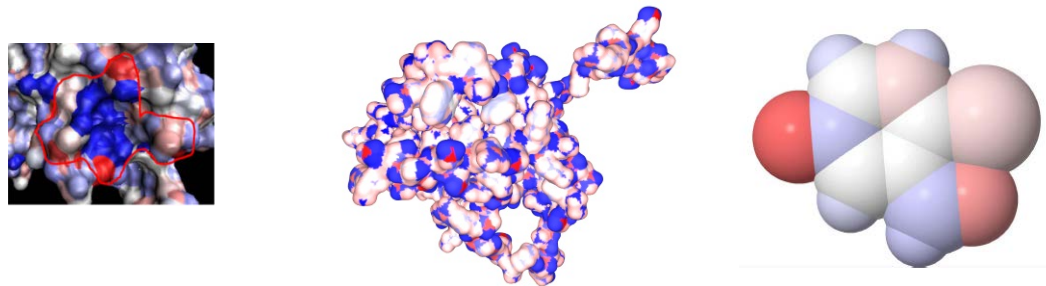
### Validation



### Detection of biologically important parts

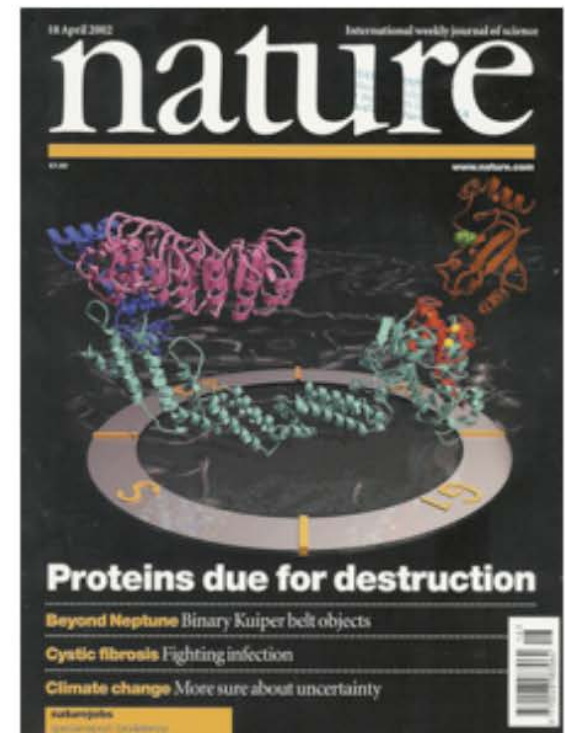
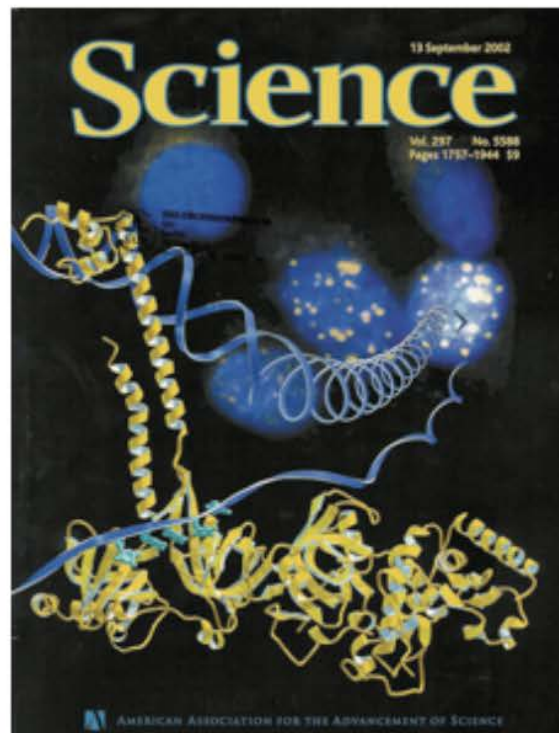
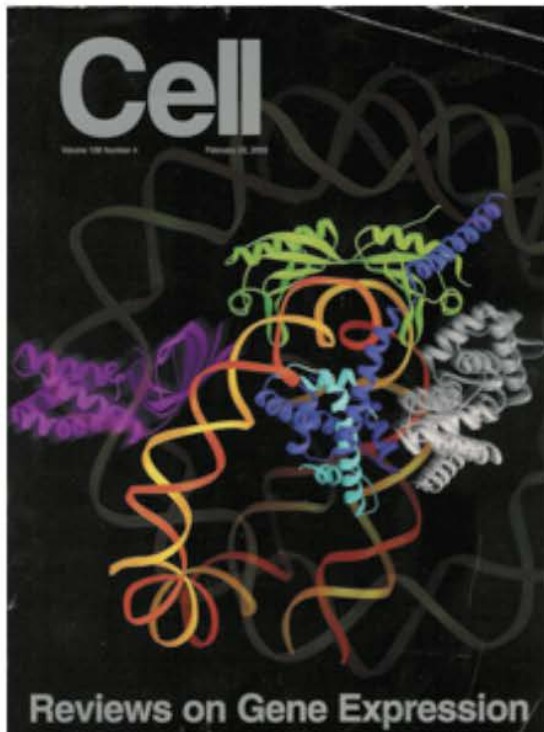


### Characterization



# Validation: Why to validate?

- High-throughput experimental techniques produce a large amount of data on the 3D structure of proteins and their complexes
- This allows us to produce impressive research results





# Validation: Why to validate?

- But are our results correct?
- **Structural biology community found that some published structures contained serious errors**
- For this reason, validation of biomolecular structures arose as a major issue

## Retraction

WE WISH TO RETRACT OUR RESEARCH ARTICLE "STRUCTURE OF MsbA from *E. coli*: A homolog of the multidrug resistance ATP binding cassette (ABC) transporters" and both of our Reports "Structure of the ABC transporter MsbA in complex with ADP•vanadate and lipopolysaccharide" and "X-ray structure of the EmrE multidrug transporter in complex with a substrate" (1–3).

The recently reported structure of Sav1866 (4) indicated that our MsbA structures (1, 2, 5) were incorrect in both the hand of the structure and the topology. Thus, our biological interpretations based on these inverted models for MsbA are invalid.

An in-house data reduction program introduced a change in sign for anomalous differences. This program, which was not part of a conventional data processing package, converted the anomalous pairs (I+ and I-) to (F- and F+), thereby introducing a sign change. As the diffraction data collected for each set of MsbA crystals and for the EmrE crystals were processed with the same program, the structures reported in (1–3, 5, 6) had the wrong hand.

The error in the topology of the original MsbA structure was a consequence of the low resolution of the data as well as breaks in the elec-

tron density for the connecting loop regions. Unfortunately, the use of the multicopy refinement procedure still allowed us to obtain reasonable refinement values for the wrong structures.

The Protein Data Bank (PDB) files 1JSQ, 1PF4, and 1Z2R for MsbA and 1S7B and 2F2M for EmrE have been moved to the archive of obsolete PDB entries. The MsbA and EmrE structures will be recalculated from the original data using the proper sign for the anomalous differences, and the new C $\alpha$  coordinates and structure factors will be deposited.

We very sincerely regret the confusion that these papers have caused and, in particular, subsequent research efforts that were unproductive as a result of our original findings.

GEOFFREY CHANG, CHRISTOPHER B. ROTH,  
CHRISTOPHER L. REYES, OWEN PORNILLOS,  
YEN-JU CHEN, ANDY P. CHEN

Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA.

### References

1. G. Chang, C. B. Roth, *Science* **293**, 1793 (2001).
2. C. L. Reyes, G. Chang, *Science* **308**, 1028 (2005).
3. O. Pornillos, Y.-J. Chen, A. P. Chen, G. Chang, *Science* **310**, 1950 (2005).
4. R. J. Dawson, K. P. Locher, *Nature* **443**, 180 (2006).
5. G. Chang, *J. Mol. Biol.* **330**, 419 (2003).
6. C. Ma, G. Chang, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2852 (2004).

# Validation: Validation of biomacromolecules

Various tools for the validation of the protein and nucleic acid 3D structures are well established:

*WHAT\_CHECK, PROCHECK, MolProbity, OOPS, Mogul, Coot, PHENIX*

They are focused on checking of structure and geometry properties:

- Electron density
- Atom clashes
- Bond length
- Bond angles
- Chirality and planarity



# Validation: Ligand validation

## Very important:

- Ligands play a key role in a function of biomacromolecules
- Ligands are the main source of errors in structures

## Challenging:

- High diversity and nontriviality of their structure
- Lack of information about correct structures

## Validation against tabular values of properties:

- Compares geometrical properties of molecules with tabular value
- Example of errors: Atom clashes, bond length errors,
- Tools: ValLigURL, Mogul, Coot, PHENIX

## Validation against a template molecule:

- Compares a validated molecule with a correct (template) molecule
- Example of errors: Missing atoms, wrong chirality, atom substitutions
- Tools: PDB care, MotiveValidator, ValidatorDB, PDB validation reports

# Validation: Example - PDB entry validation

## Validation of 3D12

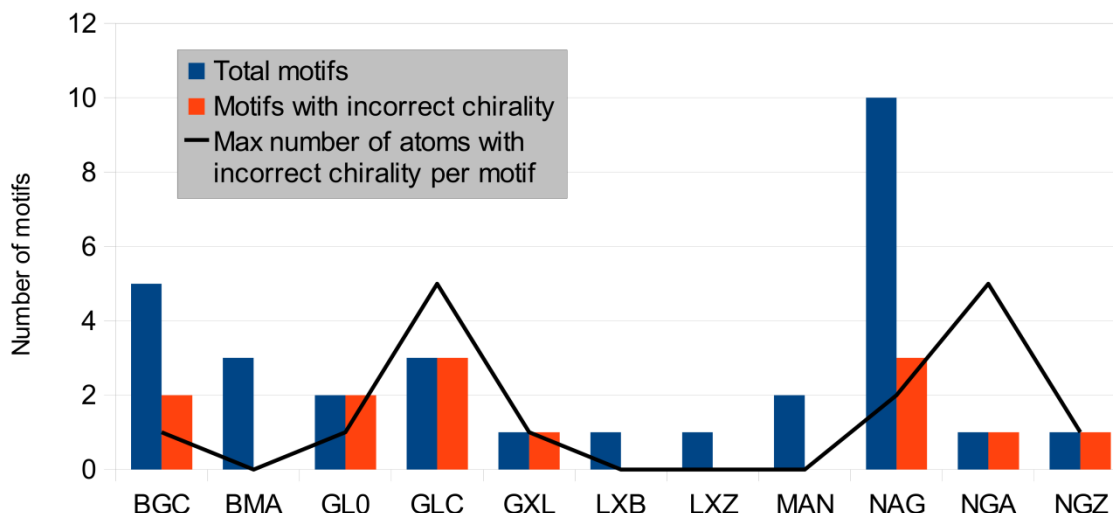
### Nipah G attachment glycoprotein

Contains 30 instances of 11 different carbohydrates, each with one ring and five chiral atoms.

### Results:

- 13 of these ligands have incorrect chirality
- In a few cases, all chiral atoms exhibit incorrect chirality

Not very good ☹️



# Validation: Exercise

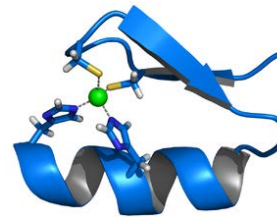
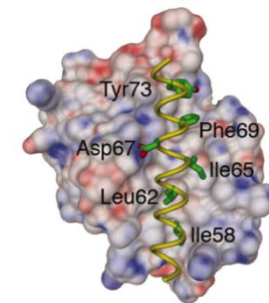
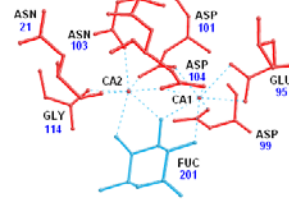
## Validation of Nipah G attachment glycoprotein (3D12):

- See validation results for 3D12 in ValidatorDB:  
Open ValidatorDB (<http://ncbr.muni.cz/ValidatorDB>), use:  
Search -> PDB Entry -> 3D12 -> Quick Search. Browse the bookmarks  
“Overview”, “Summary” and “Details” to get all the required information.
- See PDB validation report for 3D12:  
Go to Protein Data Bank Europe (<http://www.ebi.ac.uk/pdbe/node/1>), use:  
Search 3D12, Download files, Validation, Full report (PDF)

# Detection: Which biomacromolecular parts can we detect?

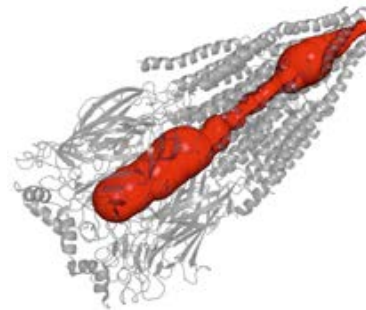
## Fragments:

- Binding sites
- Elements of secondary structure
- Supersecondary motifs



## Channels:

- Pathways from surface to a binding site
- Pores



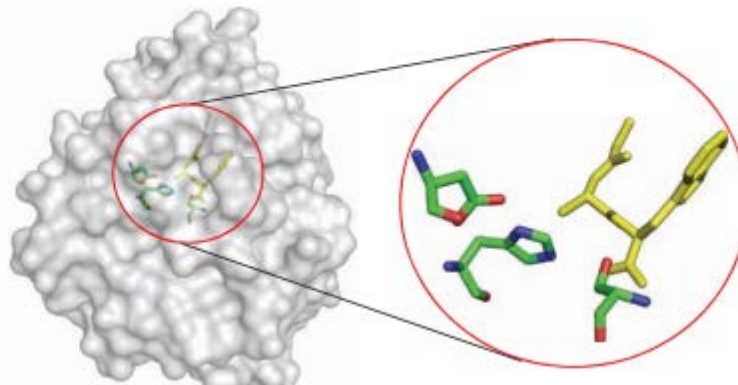
# Detection: Why to detect biomacromolecular parts?

## Fragments:

- Patterns for drug design
- Comparison of biomacromolecules
- Understanding / discovery of biomacromolecule function

## Channels:

- Key objects for biomacromolecule function
- Influence the binding site selectivity (only some substrate can path through)



# Detection: How to detect biomacromolecular fragments?

- **Methodology:**

- Describe a fragment via a defined expression (query)
- Find all suitable fragments

- **Tools:** PatternQuery, RASMOT-3D PRO, Promotif, Prosite, IMAAAGine, PDBeMotif, SPRITE& ASSAM, 3Dfit, SPASM, Protein segment finder

`Residues("TES").AmbientResidues(4)`

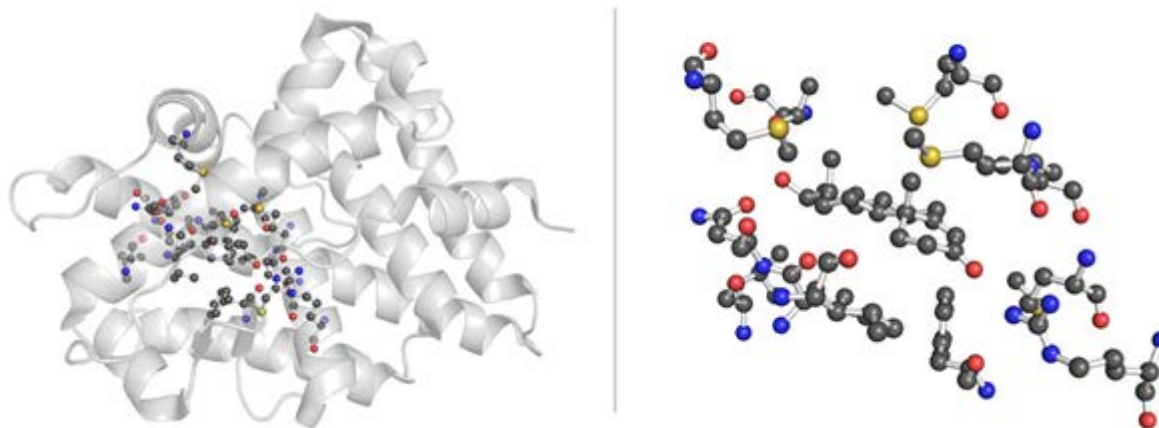


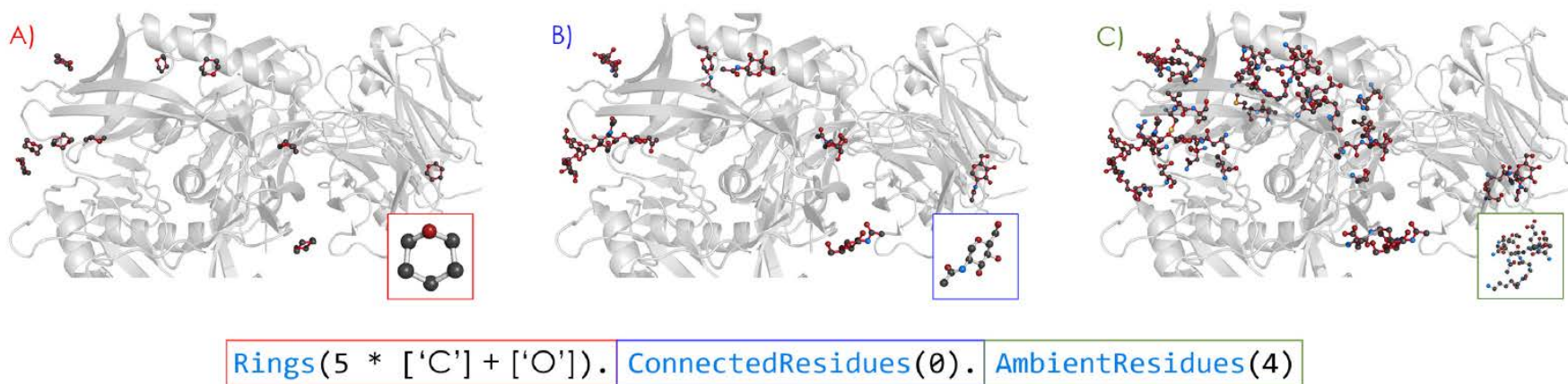
Figure: Detection of testosterone (TES) and its 4 Å large surrounding via PatternQuery: A query and a picture of the detected fragment.



# Detection: Example – binding pocket detection

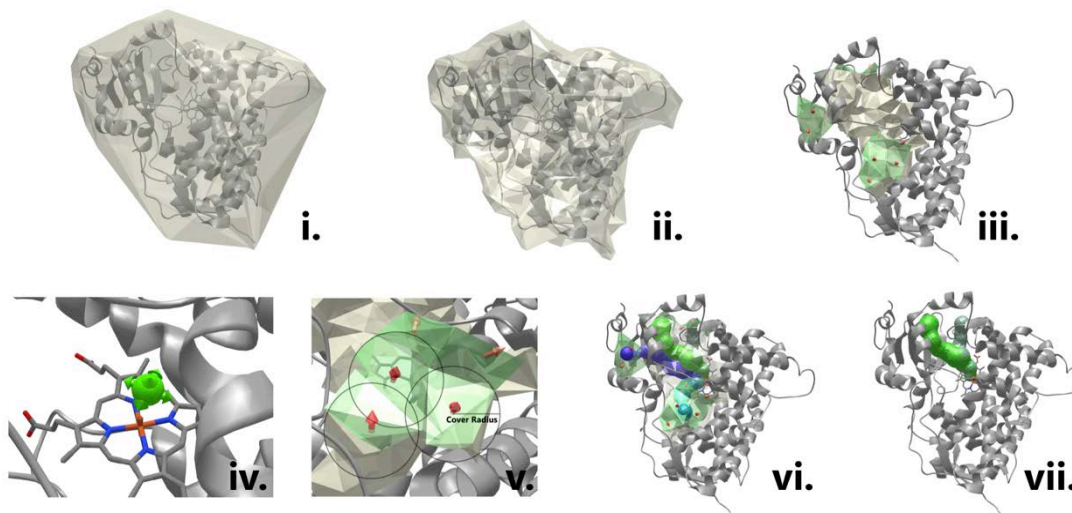
## Detection of fragments within 3U7Y

- **Glycoprotein gp160** from Human immunodeficiency virus 1 in complex with Homo sapiens immunoglobulins (PDB ID: 3U7Y).
- **Goal:** Detect a binding pocket of any residue containing a pyranose
- **Results** - via PatternQuery (<http://ncbr.muni.cz/PatternQuery>):



A) First, the query identifies a pyranose moiety (a ring composed of 5 carbons and an oxygen atom). B) Then, all residues which include this pattern in their structure are identified. C) Finally, all the residues that are at most 4Å from any of the pyranose containing residues are detected as well.

# Detection: How to detect biomacromolecular channels?



- **Methodology:**

- Delaunay triangulation/Voronoi diagram (i.)
- Approximating the molecular surface (ii.) and identifying cavities (iii.)
- Identifying possible start (iv.) and end (v.) points of channels
- Computing channels via Dijkstra's algorithm (vi.)
- Filtering of channels – removing too similar channels (vii.)

- **Tools:** MOLE, Caver, MolAxis

## Detection: Example – channels detection

### Detection of channels within 1TQN

- **Microsomal cytochrome P450 3A4** (PDB ID: 1TQN).
- **Goal:** Detect channels from a buried active site (in Glu 308 and Thr 309 residues, according to Catalytic Site Atlas) to a surface of the cytochrome
- **Results** - via MOLEonline (<http://ncbr.muni.cz/mole>) :

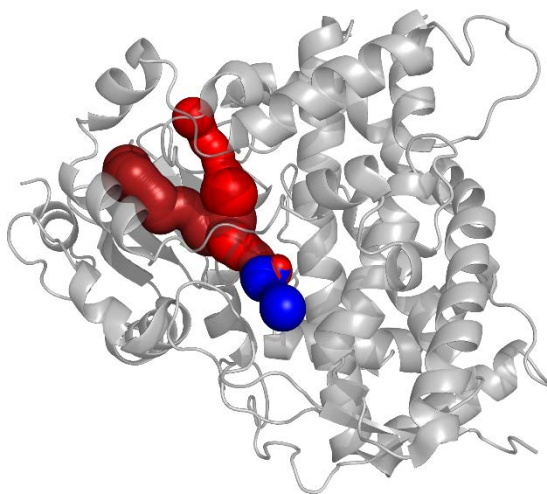


Figure: Results of channel analysis of Cytochrome P450 3A4. Three channels found from user-specified starting point (i.e., Glu 308 and Thr 309) are shown – the solvent channel (in blue), the channel 2a (in dark red) and the channel 2e (in light red).

# Detection: Exercise

## Exercise 1 - Fragment detection:

- Detect a binding pocket of any residue containing a pyranose within glycoprotein gp160 (3U7Y).

Open PatternQuery (<http://ncbr.muni.cz/PatternQuery>), do:

Query Protein Data Bank

Queries:

Unique Name ...: Test111

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

Add+

PDB ID List: 3U7Y

Submit

Details: Test111 10/1

## Exercise 2 – Channel detection:

- Detect channels from a buried active site (in Glu 308 and Thr 309 residues, according to Catalytic Site Atlas) to a surface of the cytochrome P450 3A4 (1TQN).

Open MOLEonline (<http://ncbr.muni.cz/mole>), do:

Quick start -> 1TQN -> Next -> Starting point: A Glu 308, A thr 309 -> Submit

# Characterization: Which characteristics can we calculate?

## Geometrical properties:

- Biomacromolecular surface
- Channel length, channel volume, ....
- Binding site size
- ....

## Biochemical composition:

- Channel lining residues
- Binding site residues composition

## Physico-chemical properties:

- **Partial atomics charges**
- Hydrophobicity
- Partition coefficients
- ...

We will now focus on them,  
because they are very useful

+

They present an illustrative  
example of structure  
characterization

# Characterization: Why are partial atomic charges useful?

- **Real numbers describing a distribution of electron density within a molecule**
- **Provide clues to the chemical behaviour of molecules**
- **Applications:**
  - **Computational chemistry and molecular modeling:**
    - molecular dynamics
    - docking
    - conformational searches
    - binding site predictions
  - **Chemoinformatics:**
    - descriptors for QSAR and QSPR modelling
    - virtual screening
    - similarity searches
  - **Structural bioinformatics:**
    - study of mechanisms and effects connected with certain chemical action, e.g.:
      - an activation of some biomacromolecule
      - a binding of some ligand
    - predict influences of structural changes, e.g.:
      - an influence of a certain point mutation



# Characterization: How to calculate partial atomic charges?

- **Methodology:**

- Quantum mechanical (QM) charge calculations: Accurate, but very time demanding
- Empirical charge calculations: Accuracy comparable to QM, markedly faster

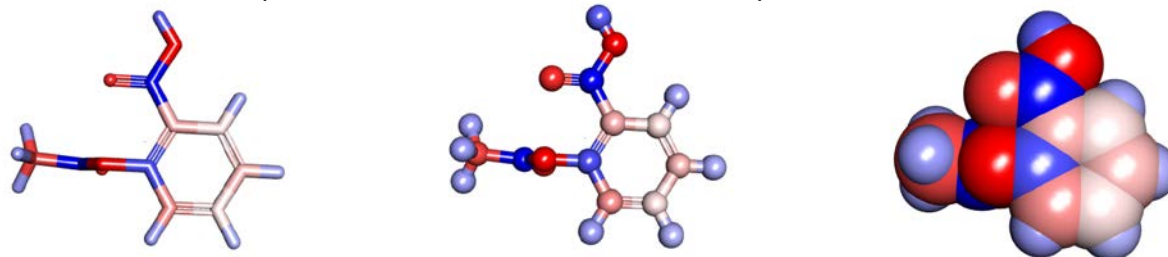
- **Tools:**

- **QM charges:** Gaussian
- **Empirical charges:** AtomicChargeCalculator, OpenBabel, NEEMP, EEM\_Solver

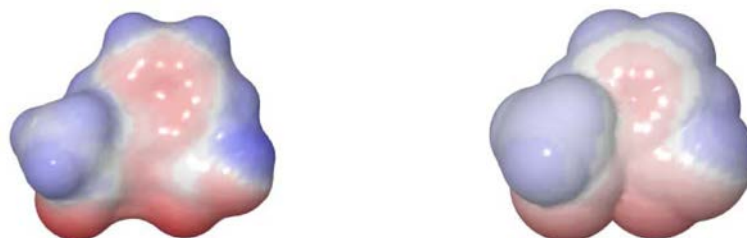
# Characterization: How to visualize partial atomic charges?

## Coloring of molecular structure models according charges:

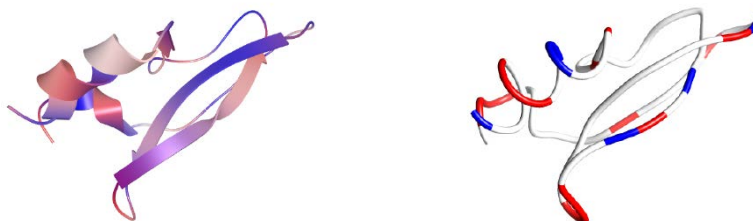
- Standard models (sticks, balls & sticks, CPK):



- Surface models (van der Waals surface, solvent accessible surface):

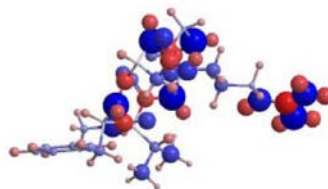


- Schematic models (cartoon, ribbon):



## Size of atoms based on charges

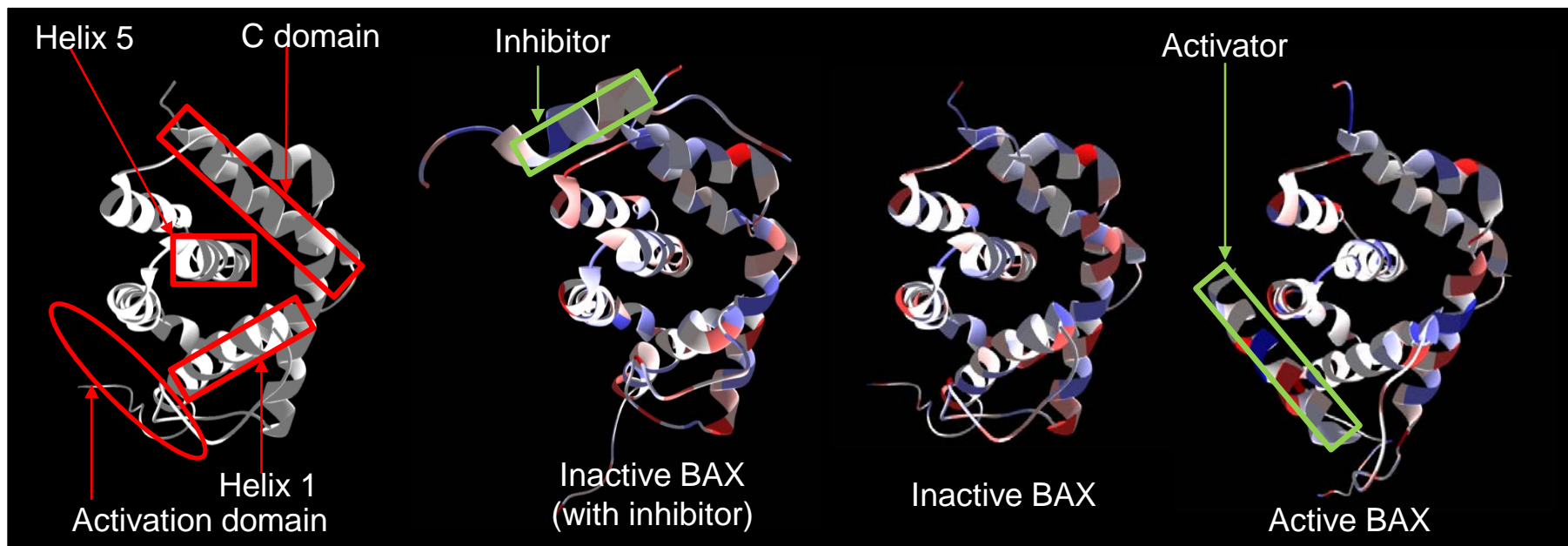
(only for balls & stick models):



# Characterization: Example – charge calculation in proteins

## Calculation of charges in inactive and active BAX protein

- **Apoptotic protein BAX:** inactive BAX (PDB ID: 1F16), inactive BAX with inhibitor (PDB ID: 2LR1), active BAX with activator (PDB ID: 2K7W)
- **Goal:** Detect a binding pocket of any residue containing a pyranose
- **Results** - via AtomicChargeCalculator (<http://ncbr.muni.cz/ACC>):



### Figure:

- Inactive BAX with and without an inhibitor show very similar charge distribution
- Activation of BAX is performed via binding of a strongly charged activator
- Binding of this activator causes vanishing of charge in Helix 1, Helix 5 and C domain (they are white in the figure of activated BAX).
- It causes a release of C domain in activated BAX

# Characterization: Exercise

## Exercise 1 – Charge calculation in small organic molecule:

- Toluene is known to have more negative charges in the positions ortho- and para- (comparable to meta- position) and therefore it directs a follow-up substitution into these positions. Calculate charge distribution in toluene (it can be obtained from Pubchem, CID 1140) and check, if the results agree with this fact.

Download toluene 3D structure (in SDF format) from Pubchem and save it to a file Structure3D\_CID\_1140.sdf.

Open AtomicChargeCalculator (<http://ncbr.muni.cz/ACC>), do:

Submit a Computation -> Select file: Structure3D\_CID\_1140.SDF -> Upload -> Compute -> Click on Structure3D\_CID\_1140 -> 3D model -> hover mouse on the atoms and you can see charge values.

## Exercise 2 – Charge calculation in protein:

- Calculate charges for inactive BAX (PDB ID: 1F16), inactive BAX with inhibitor (PDB ID: 2LR1) and active BAX with activator (PDB ID: 2K7W). Study differences in charge distributions of these structures.

Download PDB files for structures 1F16, 2LR1 and 2K7W from Protein Data Bank.

Open AtomicChargeCalculator (<http://ncbr.muni.cz/ACC>) and do the same steps as in the previous exercise.

## Further reading:

Koča J., Svobodová Vařeková R., Pravda L., Berka L., ...

### **Structural bioinformatics tools for drug design**

Extraction of biologically relevant information from structural databases

Springer, should appear in the autumn of 2016

# Thank you for your attention



Central European Institute of Technology  
Masaryk University  
Kamenice 753/5  
625 00 Brno, Czech Republic

[www.ceitec.muni.cz](http://www.ceitec.muni.cz) | [info@ceitec.muni.cz](mailto:info@ceitec.muni.cz)



EUROPEAN UNION  
EUROPEAN REGIONAL DEVELOPMENT FUND  
INVESTING IN YOUR FUTURE



**OP Research and  
Development for Innovation**

