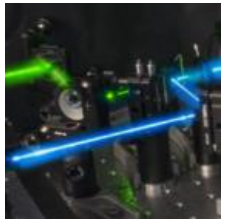
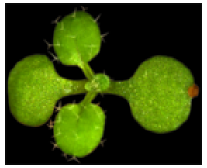
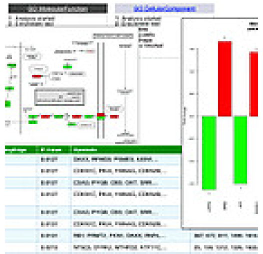
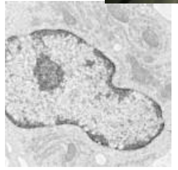
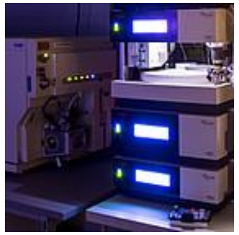


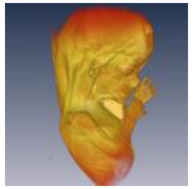
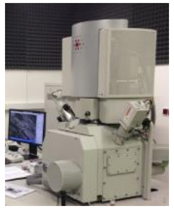


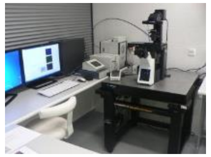









Data management solutions - making experimental data available through web interfaces

Radka Svobodová (CEITEC-MU)

24. 09. 2018

EUROPEAN UNION

RI AT CZ

Interreg
EUROPEAN UNION
Austria-Czech Republic
European Regional Development Fund

Pilot project introduction

Project partners:

- Bioinformatics CF (CEITEC-MU):
  Radka Svobodová
  Vladimír Horský
  David Sehnal
  & colleagues from Laboratory of Computational Chemistry
- Bioinformatics & Scientific Computing (VBCF):
  Attila Gyenesei and his team

Equipment used:

- MetaCenter computing cluster, Masaryk University

Pilot project introduction

Project goal:

- Web application for visualization, sharing and analyses of experimental data.

Particular goals:

- Quick data delivery and visualization of biomacromolecular structures (**LiteMol suite**)
- Visualization of quality trends in biomacromolecular data (**ValTrends^{DB}**)

Potential end-users:

= users of core biomolecular databases (Protein Data Bank, UniProt, …)

- Researchers – structural biology, biology, bioinformatics, chemoinformatics, biochemistry, etc.
- Students
- Companies – pharmaceutical, life science software development, ...
- Infrastructures – ELIXIR, Instruct, ...

Project implementation

Methodology:

- Programming languages:
  JavaScript, TypeScript
  C++
- Data inputs:
  Protein Data Bank – structures, validation reports

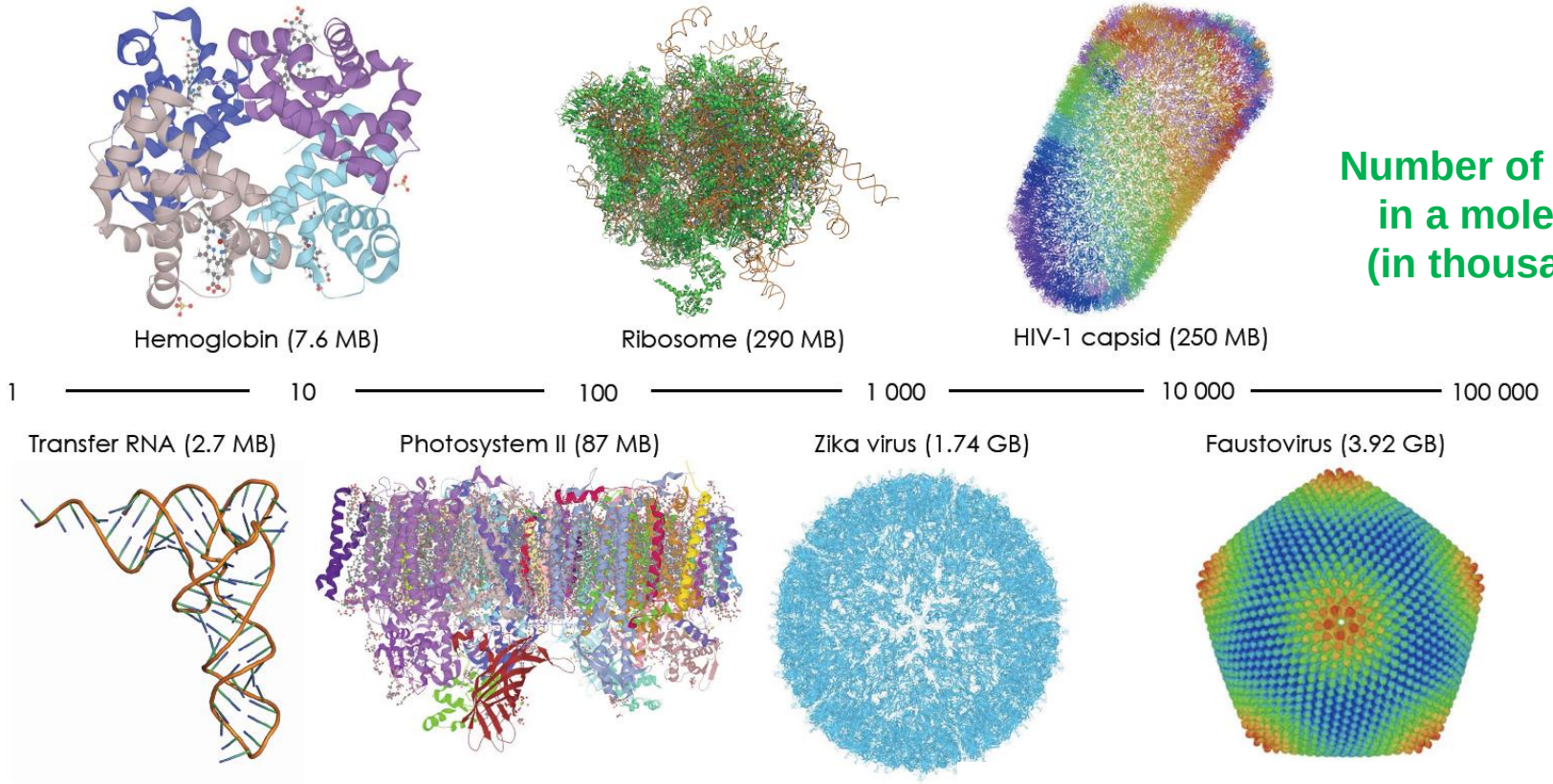
The way of cooperation:

- Development: CEITEC
- Feedback and testing: VBCF

Project results: LiteMol suite

Protein Data Bank data volume

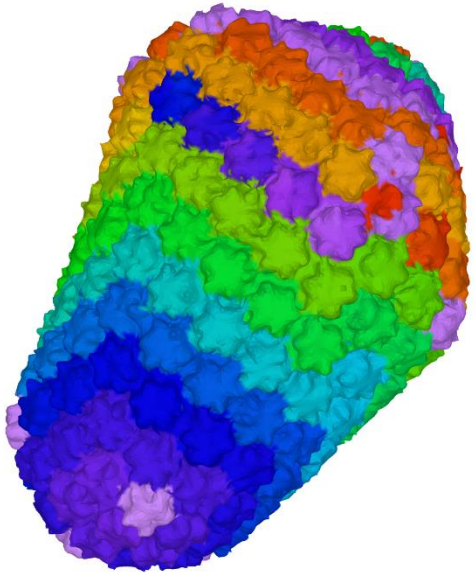
- Size of biomacromolecules markedy grows
- Increasing number of atoms

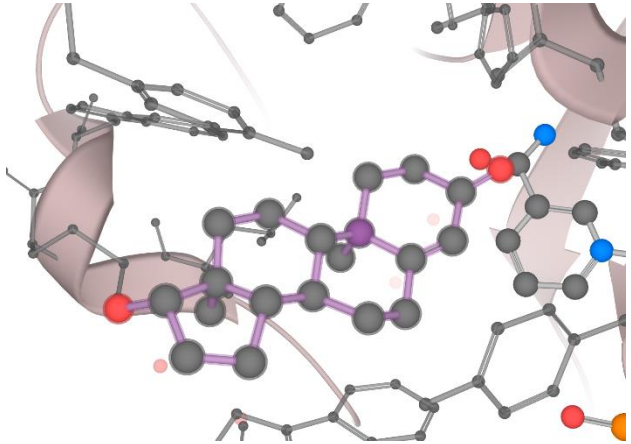


Project results: LiteMol suite

Challenges

- We need to deliver and visualize **all the layers**
- We need to do it also for **large structures**
- We need to show **entire structure**, and also its **details**
- We have to do it **quickly** and **interactively**





Project results: LiteMol suite

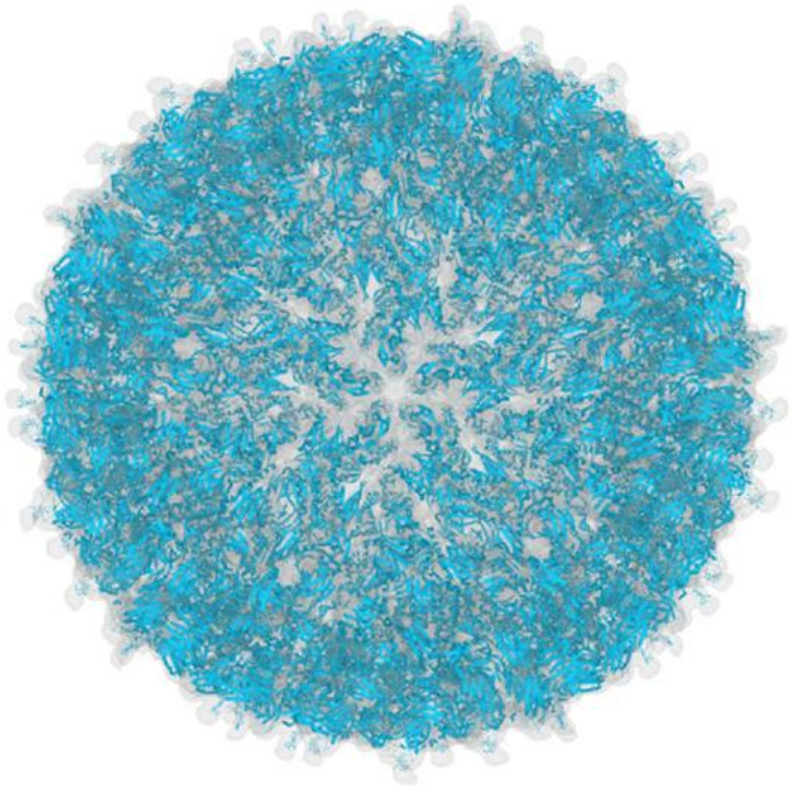
Solution – LiteMol suite

Principle:

- Works like an interactive microscope
- Deals only with data, which user needs

Project results: LiteMol suite

LiteMol suite use case: ZIKA virus



1572× less

1.03 MB

Cryo-EM structure of Zika Virus
Cartoon model + EM Surface

mmCIF+CCP4
All data

BinaryCIF
CS+DS

Project results: ValTrends^{DB}

Validation: Why to validate?

Nightmare before Christmas

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 electron 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 Cα 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.

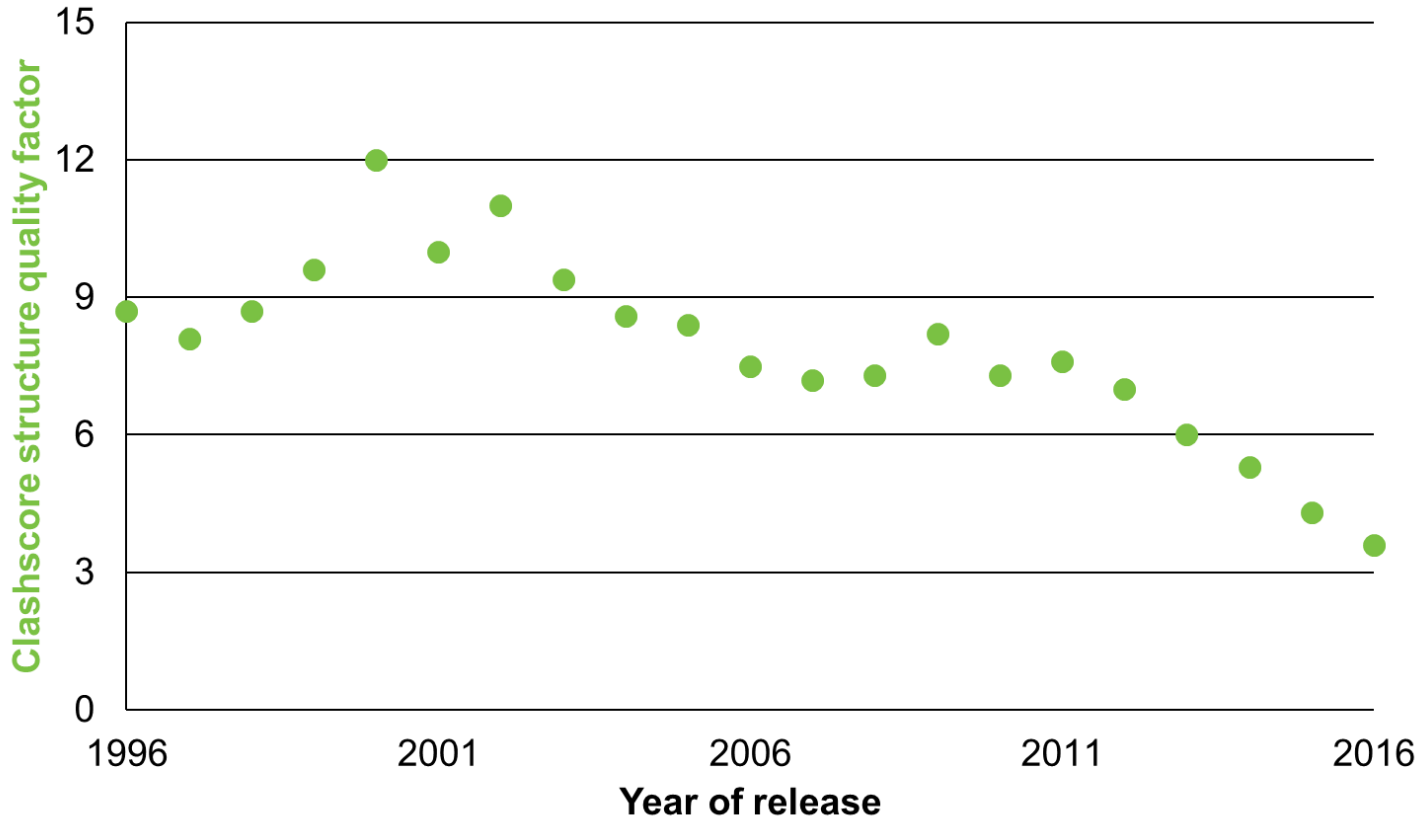
References

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).

Structural biology community found that some published structures contained serious errors

Project results: ValTrends^{DB}

A fairy tail of structure quality

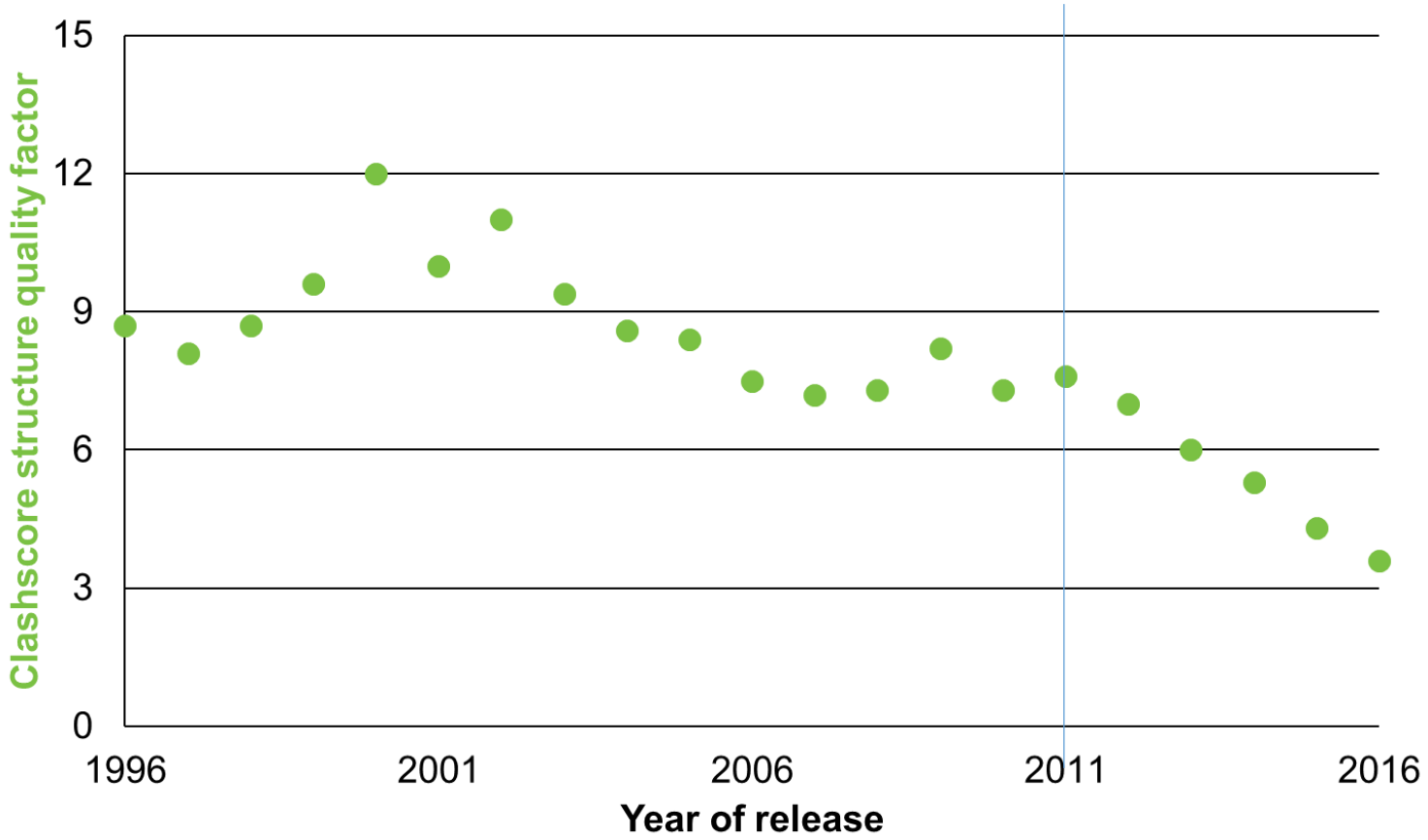


Clashscore structure quality factor represents the number of atom clashes per thousand atoms of a PDB structure

Based on data from ncbr.muni.cz/ValTrendsDB

Project results: ValTrends^{DB}

Validation report paper[1] published

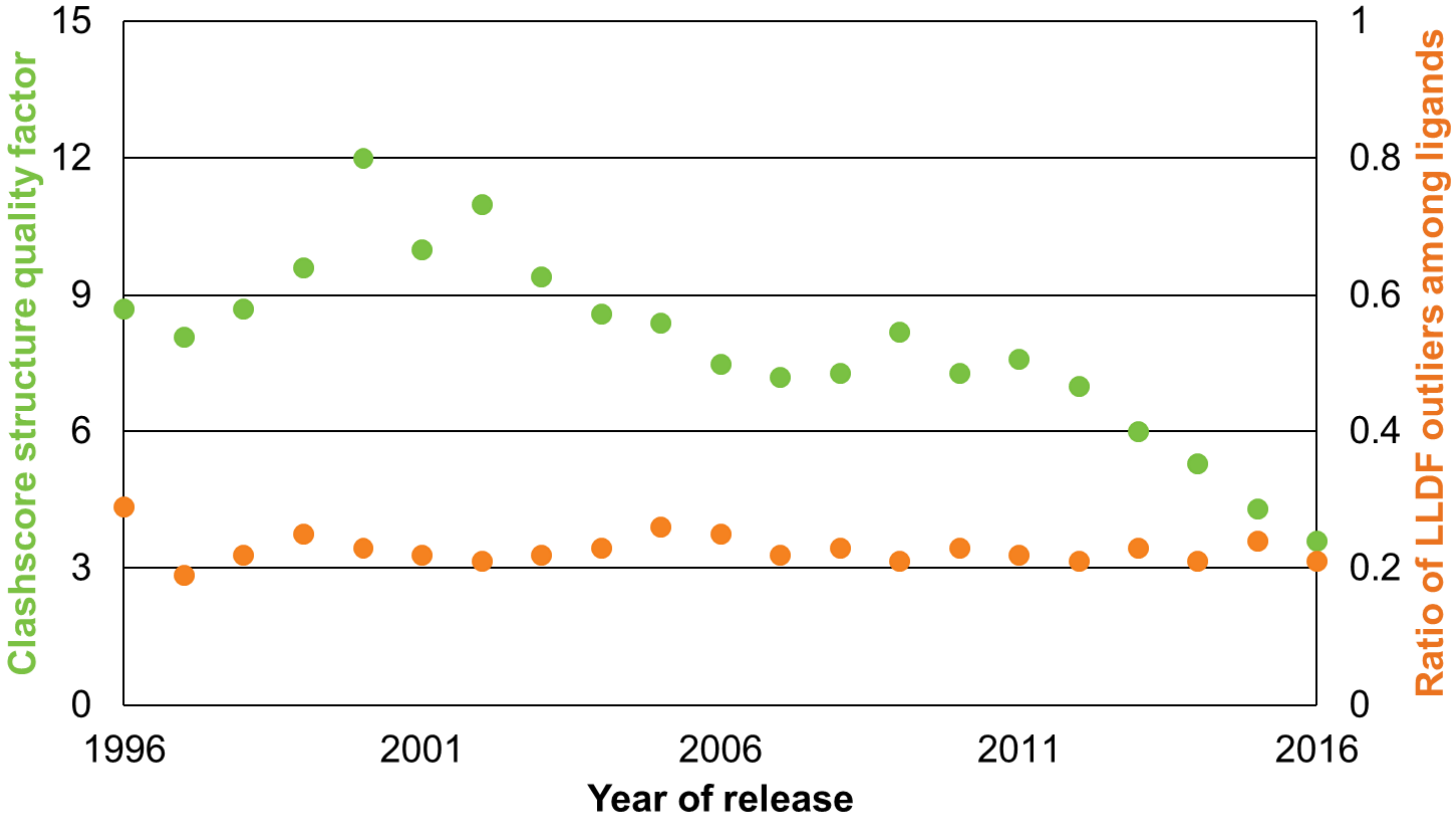


[1]Read, R. J. et al. (2011). A new generation of crystallographic validation tools for the protein data bank. *Structure (London, England: 1993), 19*(10), 1395–412.

Based on data from ncbr.muni.cz/ValTrendsDB

Project results: ValTrends^{DB}

A fairy tail of structure quality?

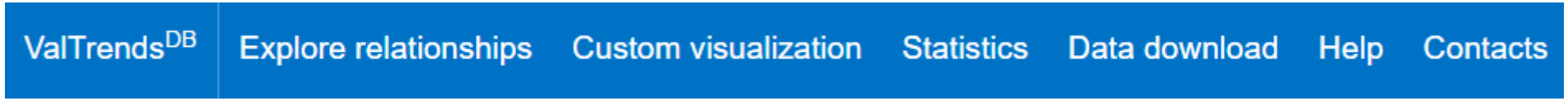


Ratio of LLDF outliers among ligands represents statistical comparison of ligand model quality to adjacent (5 Å) residue model quality

Based on data from ncbr.muni.cz/ValTrendsDB

Project results: ValTrendsDB

The ValTrends^{DB} database



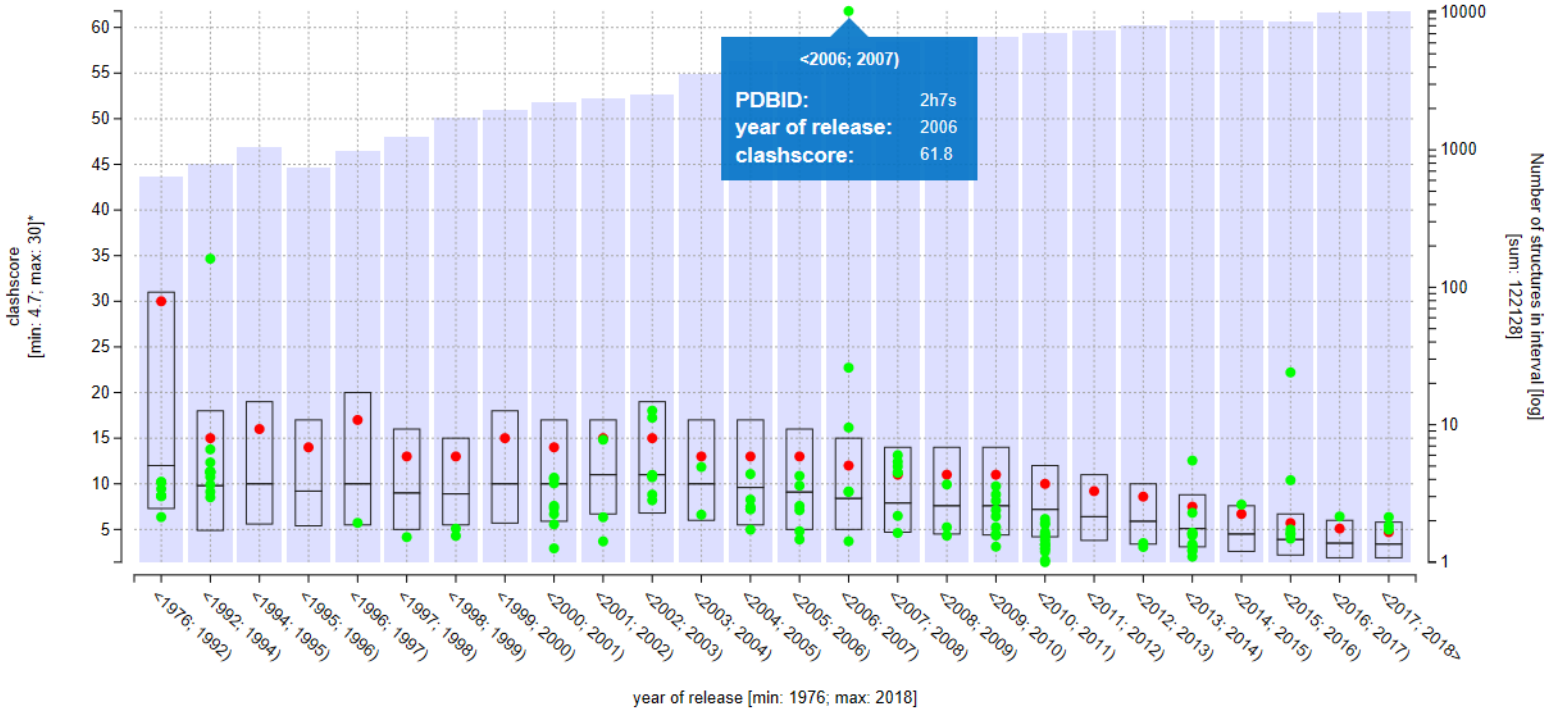
- 88 factors considered,1852 meaningful pairs of factors evaluated
- Updated weekly
- Results available in **ValTrends^{DB}** database at ncbr.muni.cz/ValTrendsDB
  - Predefined plots
  - Custom plots
  - Value distribution
  - Detailed description of the analysis

Project results: ValTrends^{DB}

Quality of protein class – example

Dependency of **clashscore** on **year of release**

Green points are structures of cytochromes P450-cam



*clashscore structure quality factor

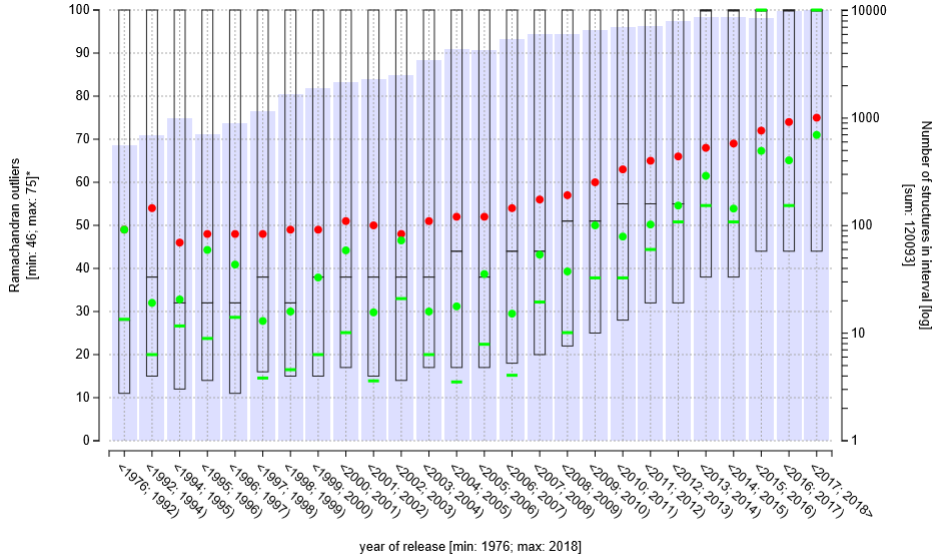
Note: The graph shows, that the structure 2h7s has markedly lower quality than other structures in the dataset.

Project results: ValTrends^{DB}

Quality of proteins published in selected journal – example
Dependency of **Ramachandran outliers** on **year of release**

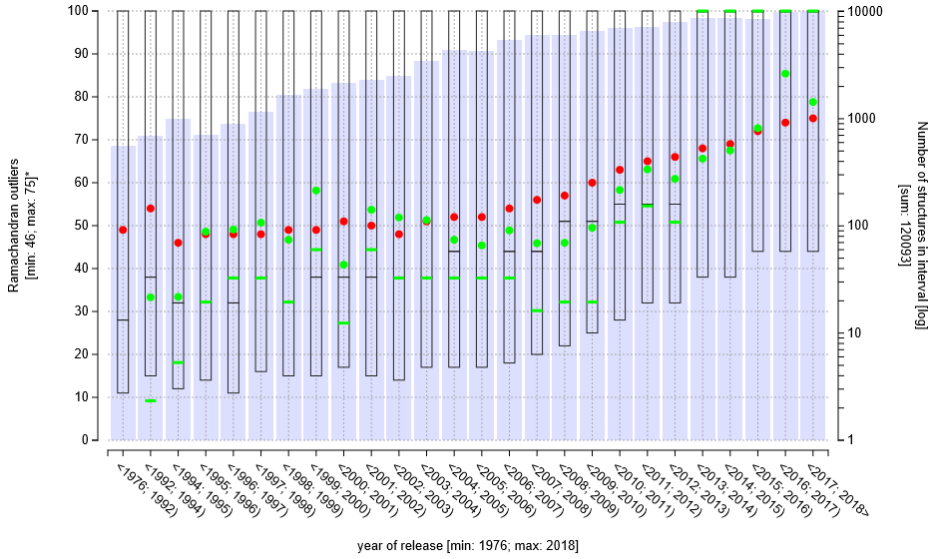
Green points and lines are average values and medians of quality criteria for structures published in:

Nature



*Ramachandran outliers structure quality factor - percentile version
(the higher percentage, the better quality)

Structure



*Ramachandran outliers structure quality factor - percentile version
(the higher percentage, the better quality)

Note: The graph shows, structures published in Nature have lower quality than the structures published in Structure.

Conclusion

- **Web applications LiteMol suite and ValTrendsDB were developed**
- **The tools showed as very perspective:**
  - Used by many users
  - Parts of LiteMol suite integrated in Protein Data Bank and Uniprot
  - LiteMol suite was published in Nature Methods
  - ValTrends^{DB} integration is in progress
  - Two articles were published based on ValTrends^{DB} analyses

- **Future plans:**
  - Extending the tools and their functionality
  - Preparation for Open access