Structure validation Introduction

Low quality structures: We can not use them!

Garbage in garbage out

Validation – beginning

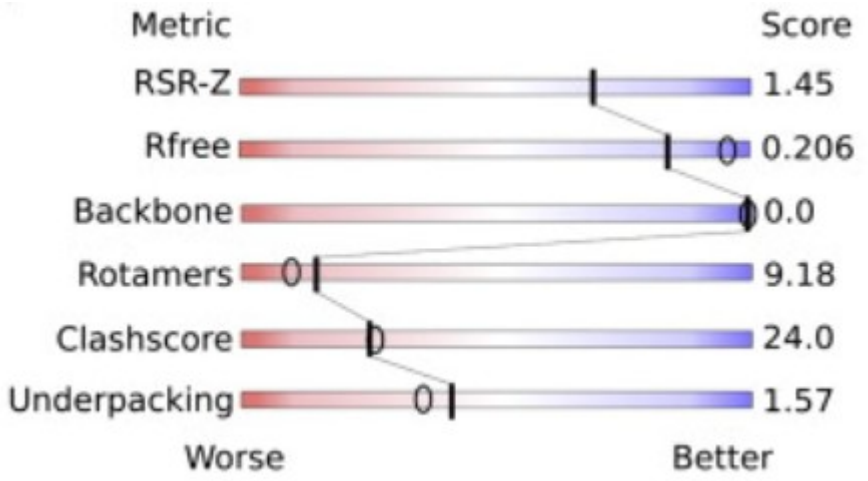
- **Great structural biology results, excellent papers**
- 2006: First problems - retraction of papers because wrong 3D structure data:
  - Matthews B.W. (2007) *Five retracted structure reports: inverted or incorrect?* **Protein science**, 16, 1013–6.
  - Johnston C.A., Kimple A.J., Giguere P.M., Siderovski D.P. (2008) *RETRACTED: Structure of the Parathyroid Hormone Receptor C Terminus Bound to the G-Protein Dimer Gb1g2*. **Structure**, 16, 1086–1094.
- 2008: X-ray Validation Task Force
  - Read R.J. et. al (2011). *A new generation of crystallographic validation tools for the protein data bank*. **Structure**, 19(10), 1395-1412.
- 2009: NMR Validation Task Force
  - Montelione G.T., et al. (2013). *Recommendations of the wwPDB NMR validation task force*. **Structure**, 21(9), 1563-1570.
- 2010: Electron Microscopy Validation Task Force
  - Henderson R., (2012). *Outcome of the first electron microscopy validation task force meeting*. **Structure**, 20(2), 205-214.

Nightmare before Christmas

Retraction

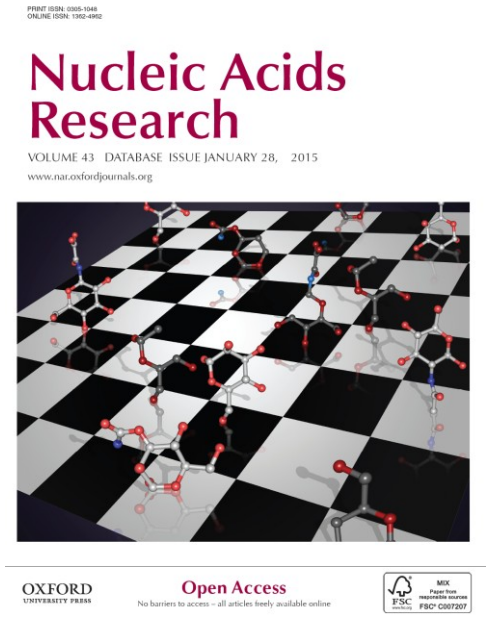
WE WISH TO RETRACT OUR RESEARCH ARTICLE "STRUCTURE OF MsbA from *E. coli*: A homolog of the multidrug resistance ATP binding cassette (ABC) transporters" and both of our Reports "Structure of the ABC transporter MsbA in complex with ADP·vanadate and lipopolysaccharide" and "X-ray structure of the EmrE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

SCIENCE VOL 314 22 DECEMBER 2006



Validation – ligands

- **2015 ValidatorDB**
  - Sehnal D., Svobodová R., et al (2015). *ValidatorDB: database of up-to-date validation results for ligands and non-standard residues from the Protein Data Bank*. **Nucleic acids research**, 43(D1), D369-D375.
- **2016 wwPDB/CCDC/D3R Ligand Validation Workshop**
  - Adams P.D. …, Svobodová R., … (2016). *Outcome of the first wwPDB/CCDC/D3R ligand validation workshop*. **Structure**, 24(4), 502-508.



Validation – reports and metr

- **2017 Validation reports**
  - Gore S., García E.S., Hendrickx P.M., Gutmanas A., Westbrook J.D., Yang H., ..., Kleywegt G.J. (2017). *Validation of structures in the Protein Data Bank*. **Structure**, *25*(12), 1916-1927.

- **2018 Combined validation metrics**
  - Smart O.S., Horský V., Gore S., Svobodová R., Bendová V., Kleywegt G.J., Velankar S. (2018). *Worldwide Protein Data Bank validation information: usage and trends*. **Acta Crystallographica Section D: Structural**

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: HEM, PIM

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#|Z| >5	RMSZ	#|Z| >5
1	A	0.28	0/3132	0.58	1/4287 (0.0%)

There are no bond length outliers.

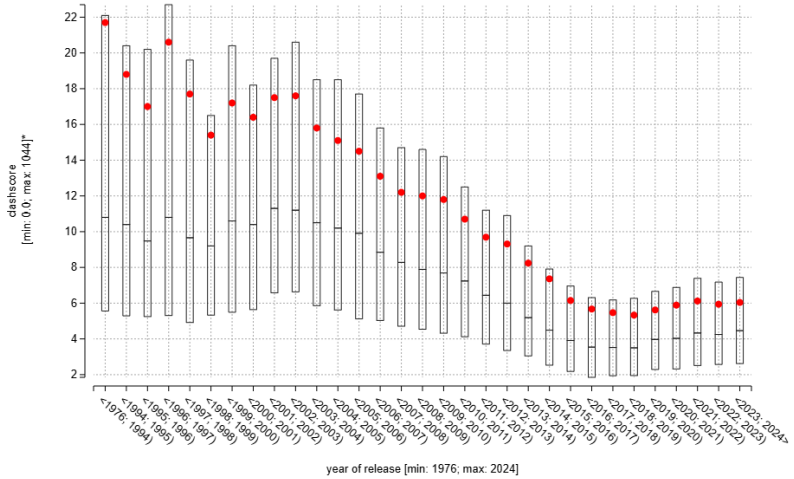
All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	309	ILE	N-CA-C	-5.24	96.85	111.00

Vladimír Horský

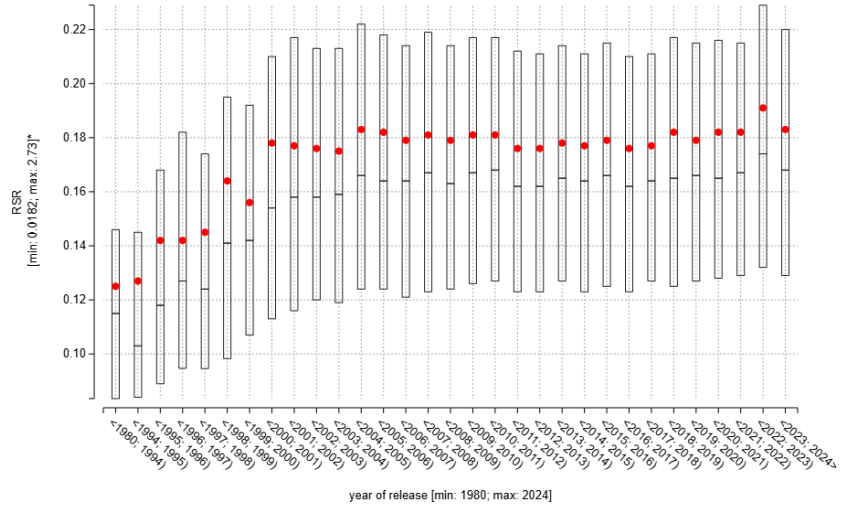
$$M_{\text{geom}} = \left(\frac{M_{\text{Rama}}^{-1} + M_{\text{side}}^{-1} + M_{\text{clash}}^{-1} + M_{\text{RNAs}}^{-1}}{4} \right)^{-1} \qquad M_{\text{xray}} = \left(\frac{M_{Rfree}^{-1} + M_{\text{RSRZ}}^{-1}}{2} \right)^{-1}$$

Validation: Does it help?



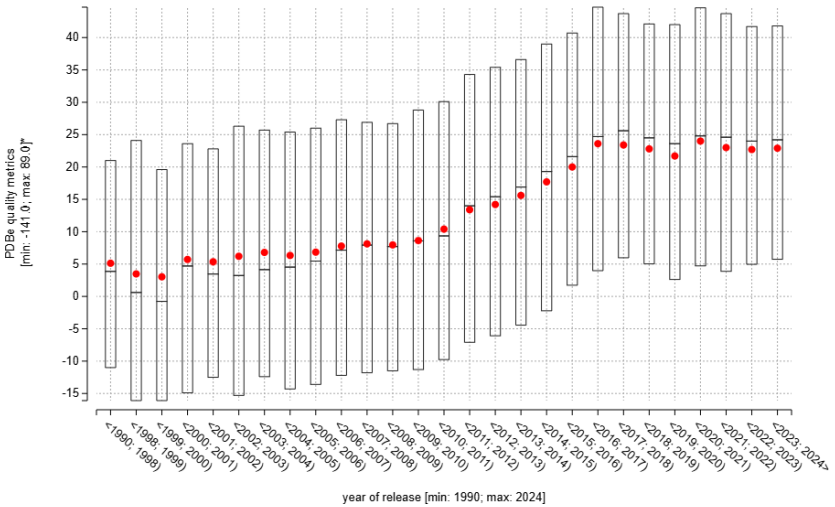
*clashscore structure quality factor

current 2024-06-13



*average RSR of ligands in structure

current 2024-06-13



*combined overall quality

current 2024-06-13

Horský V., Bendová V., Toušek D., Koča J., Svobodová R. (2019). *ValTrendsDB: bringing Protein Data Bank validation information closer to the user*. **Bioinformatics**, *35*(24), 5389-5390.

Validation – lectures today

- Protein validation
- Ligand validation
- Validation trends