

Evolution of macromolecular model quality

D. Brzezinski^{1,2,3}, Z. Dauter⁴, W. Minor³, M. Jaskolski^{1,5}

¹Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, ²Institute of Computing Science, Poznan University of Technology, Poznan, Poland, ³Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, USA, ⁴Synchrotron Radiation Research Section, Macromolecular Crystallography Laboratory, National Cancer Institute, Argonne National Laboratory, Argonne, USA, ⁵Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznan, Poland

dariusz.brzezinski@cs.put.poznan.pl

Over the last decades, the PDB has been developing tools and standards for the assessment of the quality of the structural models deposited in its archives. Similarly, more and more journals are now requiring validation reports generated by the PDB as a prerequisite for manuscript submission. Such quality metrics have been used previously to gauge the relationship between structural model quality and publication venues [1,2]. More recently, these indicators have been applied to assess the evolution of the quality of the PDB deposits with time [3] using the concept of a percentile (P_{QI}) metric, which combines such measures as R_{free} , RSRZ (normalized Real Space R-factor) outliers, Ramachandran outliers, Rotamer outliers, and Clashscore.

In this paper we will show how the quality of macromolecular models deposited in the PDB has changed over the years (Fig. 1) and how the P_{QI} parameter can be converted to a new measure, $P_{QI}(t,d)$, that takes into account time (t) and data resolution (d). The proposed new measure can be used to reveal how structure quality in a given moment of time was related to such issues as [4]:

- differences between proteins and nucleic acids;
- comparison with structural genomics projects;
- assessment of deposits without journal publications;
- journal impact factor (Fig. 2).

The paper will also discuss how the quality of crystallographic macromolecular structures in the PDB has improved over the last years and highlight some crucial periods in this history.

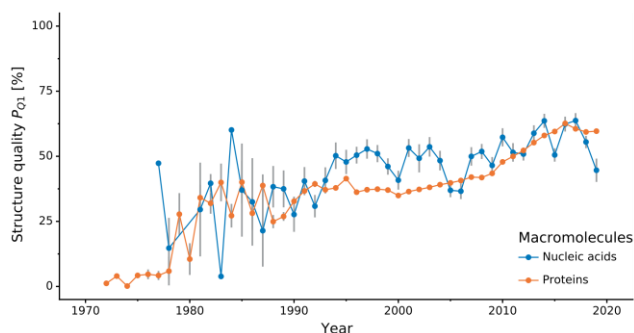


Fig. 1. Variation of mean structure quality percentile P_{QI} (y-axis), comparing nucleic acid and protein structures (color) over time (x-axis). Error bars indicate estimated unbiased standard errors of the mean.

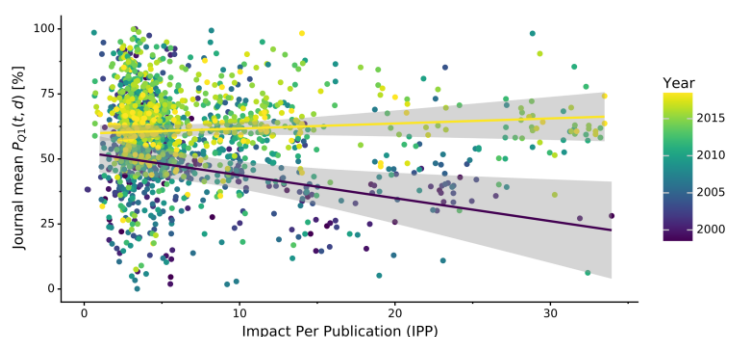


Fig. 2. Variation of mean journal structure quality $P_{QI}(t,d)$ (y-axis) in a given year (color) plotted against journal impact (x-axis). The two regression lines show linear trends for 1999 (indigo) and 2018 (yellow) together with 95% confidence intervals (gray areas).

[1] Brown, E. N. & Ramaswamy, S. (2007). *Acta Cryst.* **D63**, 941–950.

[2] Read, R. J. & Kleywegt, G. J. (2009). *Acta Cryst.* **D65**, 140–147.

[3] Shao, C., Yang H., Westbrook, J. D., Young, J. Y., Zardecki, C. & Burley, S. K. (2017). *Structure* **25**, 458–468.

[4] Brzezinski, D., Dauter, Z., Minor, W. & Jaskolski, M. (2020). *FEBS J.*, **287**, 2685–2698.

Keywords: PDB; structure quality; X-ray crystallography; proteins; nucleic acids; journal impact;

DB acknowledges the support of the Polish National Agency for Academic Exchange, grant no. PPN/BEK/2018/1/00058/U/00001. Funding for this research was also provided by NIH grant GM132595.