# Protein structure quality trends

Svobodová Radka1,2, Horský Vladimír1,2, Bučeková Gabriela1,2, Porubská Jana1,2, Doshchenko Viktoriia2

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

2 National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 753/5, 625 00 Brno, Czech Republic

Information about protein 3D structure, collected in the Protein Data Bank (PDB), is a valuable input for structural bioinformatics research and other life science fields. An important topic is the quality of the data because the reliability of research results based on these protein structure data depends on their quality. For this reason, there is a strong focus on protein structure validation. Various quality criteria were published, and information about the quality of individual protein structures is summarised in the PDB validation reports [1].

Here, we utilise our tools ValTrendsDB [2] and ValidatorDB [3] and examine selected interesting trends in protein structure quality. First, we focus on the evolution of protein structure quality in time. Afterwards, we analyse the relationship between the quality of a structure and the journal in which it has been published. Last but not least, we uncover several ligand quality issues, which are not considered by common validation procedures yet.

[1] Smart, O. S., Horský, V., Gore, S., Svobodová Vařeková, R., Bendová, V.,Kleywegt, G. J., Velankar, S. (2018). Worldwide Protein Data Bank validation information: usage and trends. Acta Crystallographica Section D, 74, 237-244.

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

[3] Sehnal, D., Svobodová Vařeková, R., Pravda, L., Ionescu, C.-M., Geidl, S., Horský, V., Jaiswal, D., Wimmerová, M., Koča, J. (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.