Assessing Macromolecular Crystal Structures for Agreement with Experimental Data at the Individual Amino Acid Residue Level

Stephen Burley[1], Chenghua Shao[2], Sijian Wang[3]

[1]RCSB Protein Data Bank, Rutgers University [2]RCSB Protein Data Bank, [3]Department of Statistics,

Stephen.Burley@rcsb.org

More than 85% of the atomic-level, three-dimensional (3D) structures of biological macromolecules currently comprising the Protein Data Bank (PDB) archive were determined using macromolecular crystallography (MX). Agreement between atomic coordinates and experimental data for >100 million individual amino acid residues occurring within ~148,000 MX structures in PDB has been analyzed in detail. The Real-Space-Correlation-Coefficient (RSCC) calculated using the 3D atomic structure of each residue and experimental electron density enables rigorous outlier detection of unreliable atomic coordinates (particularly important for sidechains) and ready evaluation of structure quality by any PDB user, even those who are not experts in structural biology. For human protein structures represented in the PDB, detailed comparisons of the individual residue RSCC experimental-agreement metric with AlphaFold2 computed structure model confidence (pLDDT-predicted local distance difference) document (i) that RSCC and pLDDT are well correlated, and (ii) that experimentally-determined MX structures (3.5Å resolution or better) are more reliable than predicted structure models and should be used preferentially whenever available.

RCSB Protein Data Bank is funded by the National Science Foundation (DBI-1832184), the US Department of Energy (DE-SC0019749), and the National Cancer Institute, National Institute of Allergy and Infectious Diseases, and National Institute of General Medical Sciences of the National Institutes of Health under grant R01GM3198.

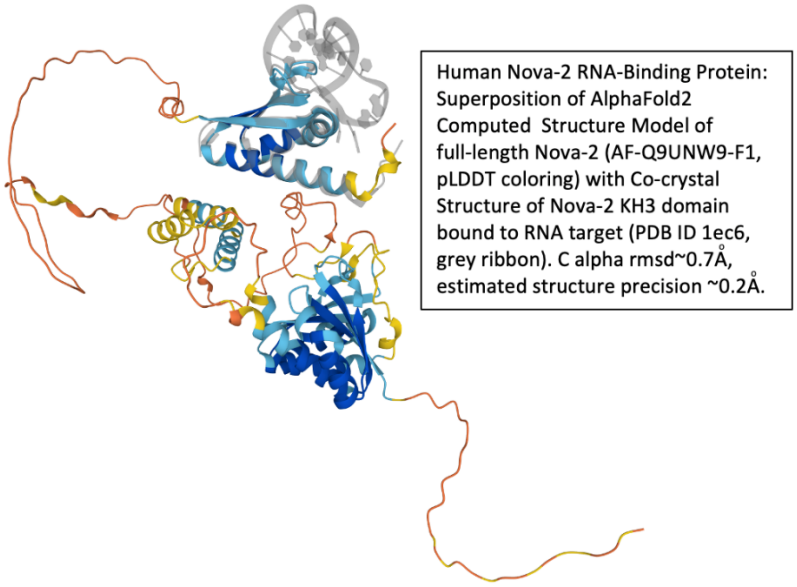

Figure 1