

Improving the quality of 3D cryo-EM structure data at the Worldwide Protein Data Bank

Justin Flatt¹

¹N/A

justin@rcsb.rutgers.edu

The Protein Data Bank (PDB) was established in 1971 as the first open-access digital data resource in biology with just seven X-ray crystallographic structures of proteins. Today, the single global PDB archive houses more than 190,000 experimentally-determined three-dimensional (3D) structures of biological macromolecules that are made freely available to millions of users worldwide with no limitations on usage. This information facilitates basic and applied research and education across the sciences, impacting fundamental biology, biomedicine, biotechnology, bioengineering, and energy sciences. Recent technical advances have catalyzed explosive growth in both the quantity and complexity of 3D electron microscopy (3DEM) data in both the PDB and Electron Microscopy Data Bank (EMDB), a member of the Worldwide Protein Data Bank (wwPDB, wwpdb.org). The wwPDB jointly manages the PDB, EMDB, and BMRB Core Archives and is committed to making data Findable, Accessible, Interoperable, and Reusable (FAIR).

wwPDB works together with expert community guidance to develop data standards to ensure biocuration keeps pace with advances in this rapidly evolving field and ensures highest possible data quality. The wwPDB global OneDep tool for PDB structure deposition, validation, and biocuration is constantly enhanced with extended metadata, more enumeration lists, and improved data checking for incoming structures derived from macromolecular crystallography, nuclear magnetic resonance spectroscopy, and 3DEM data. To improve 3DEM data validation, OneDep now requires the inclusion of half maps in EMDB depositions determined using single particle, helical, or subtomogram averaging methods. The 3DEM-related portion of the PDBx/mmCIF data dictionary used by OneDep currently includes more than 500 terms. It is regularly updated to provide controlled vocabulary and boundary ranges that reflect the current state of 3DEM and ensure accuracy and completeness of deposited metadata. wwPDB validation reports generated within OneDep provide extensive 3DEM-specific validation, including map analysis and visualization and overall map-model fit analysis, in addition to evaluating the geometric quality of deposited atomic coordinates.

In addition, efforts are underway to incorporate community recommendations for validating near-atomic resolution structures at per-residue level (e.g., adoption of local fit-to-map metrics). Finally, the wwPDB recently implemented PDB entry versioning, which allows authors to update the atomic coordinates of their existing PDB structures to correct errors detected after release, whilst retaining original PDB IDs. We thank the many authors who have used the new versioning feature to proactively correct the atomic coordinates of more than 350 previously-deposited PDB structures.