Celebrating the 50th Anniversary of the Protein Data Bank – Prepublication Peer Review and Validation of Small-Molecule Ligands Bound to Proteins and Nucleic Acids

Stephen K. Burley

RCSB Protein Data Bank, Institute for Quantitative Biomedicine, Department of Chemistry and Chemical Biology, Rutgers Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, United States; sburley@proteomics.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 structures of proteins. During its first 50 years of continuous operations, PDB holdings have grown to more than 175,000 structures becoming the single global archive of 3D-structures of proteins, nucleic acid, and their complexes with one another and small-molecule ligands. Open access to expertly biocurated PDB structures enables the efforts of many millions of basic and applied researchers, educators, and students around the world. Their work impacts fundamental biology, biomedicine, bioengineering, biotechnology, and energy sciences.

The Worldwide Protein Data Bank (wwPDB, wwpdb.org) manages the PDB archive according to the FACT principles of Fairness-Accuracy-Confidentiality-Transparency and the FAIR principles of Findable-Accessible-Interoperable-Reusable. Current wwPDB members include the US RCSB Protein Data Bank (RCSB PDB), Protein Data Bank in Europe (PDBe), Protein Data Bank Japan (PDBj), Electron Microscopy Data Bank (EMDB), and Biological Magnetic Resonance Bank (BMRB).

All data in the PDB archive conform to the wwPDB PDBx/mmCIF data dictionary, which is fully extensible both human- and machine-readable. PDB structures are composed of amino acids or nucleotide building blocks that comprise biopolymers, and associated small molecules such as water molecules, solute molecules, ions, co-factors, metabolites, enzyme inhibitors, drugs, etc. Every new structure coming into the PDB is processed using the wwPDB OneDep global system for deposition, validation, and biocuration. All PDB structures are accompanied by an official wwPDB Validation Report, exemplifying standards developed collaboratively with wwPDB Task Forces composed of community experts.

Small-molecule constituents of PDB structures are defined in the wwPDB Chemical Component Dictionary (CCD). This dictionary contains detailed chemical descriptions for standard and modified amino acids/nucleotides, small molecule ligands, solvent molecules, and others. Precise knowledge of interactions between macromolecules and small-molecule ligands is central to our understanding of biological and biochemical function, drug action, mechanisms of drug resistance, and drug-drug interactions.

Recent enhancements to the CCD and the wwPDB Validation Report will be described, together with value-added information concerning ligand quality now available on the US Research Collaboratory for Structural Bioinformatics Protein Data Bank PDB website (RCSB PDB, RCSB.org).

wwPDB members are US RCSB PDB (supported by NSF, NIH, DOE, and Rutgers Cancer Institute of New Jersey), PDBe (EMBL-EBI, Wellcome Trust, BBSRC, MRC, and EU), and PDBj (NBDC-JST), and BMRB (NIGMS).

Keywords: Protein Data Bank, Structure Validation, Small-Molecule