# Ligand Validation for the Protein Data Bank

**Stephen K. Burley**

RCSB PDB, Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, United States, and RCSB PDB, Skaggs School of Pharmacy and Pharmaceutical Sciences and San Diego Supercomputer Center, University of California San Diego, La Jolla, CA 92093, United States.

The Protein Data Bank (PDB) is the global repository for experimentally determined three-dimensional (3D) structures of biological macromolecules (proteins and nucleic acids). It is managed by the Worldwide Protein Data Bank partnership (wwPDB, wwpdb.org), consisting of the United States RCSB Protein Data Bank (rcsb.org), Protein Data Bank in Europe (pdbe.org), Protein Data Bank Japan (pdbj.org), and BioMagResBank (bmrb.org). In addition to biopolymer structure data (e.g., polypeptide chain), the PDB archive contains chemical and structural information for bound ligands. These ligands, also known as chemical components, are very diverse in nature, including ions, solvent molecules, crystallization precipitants, natural and non-standard amino acids and nucleic acids, and myriad ligands, such as drugs, cofactors, metal clusters, surfactants, etc. Precise knowledge of interactions between macromolecules and small ligands is central to our understanding of biological function, drug action, mechanisms of drug resistance, and drug-drug interactions.

The PDB archive now holds more than 139,000 experimentally determined 3D structures of biological macromolecules, with >25,700 unique ligands, all of which are publicly accessible without restrictions on usage. These structures provide essential information to a large, diverse user community worldwide. Structure data file downloads exceeded 670 million in 2017. Unique users of the RCSB PDB website (rcsb.org) are conservatively estimated at more than one million/year.

All small molecule ligands present in PDB structures are catalogued in the wwPDB Chemical Component Dictionary (CCD; http://www.wwpdb.org/data/ccd). The wwPDB CCD encompasses IUPAC atom nomenclature for standard amino acids and nucleotides, stereo- chemical assignments, bond order assignments, experimental model and computed ideal coordinates, systematic names, and chemical descriptors.

The wwPDB recently developed a global unified system (OneDep, deposit.wwpdb.org) to support deposition, validation, and biocuration of macromolecular structures and their ligand complexes and related metadata coming into the PDB archive. Within the Ligand Module of this expert software system, ligands are compared to existing CCD entries to ensure stereo-chemical accuracy and assignment of appropriate ligand IDs. New ligand IDs are being issued at the rate of ~2,000 per year. The OneDep Validation Module produces validation reports for structures determined by X-ray crystallography, nuclear magnetic resonance spectroscopy, and 3D electron microscopy. Validation standards were developed with the benefit of recommendations from expert task forces representing each experimental community. For X-ray structures, the fit of the ligand to electron density difference maps is assessed quantitatively using real- space R-factors (RSR and RSR Z-scores). Within the OneDep biocuration pipeline, 3D electron density difference maps are produced for review by expert wwPDB biocurators. Precomputed electron density difference maps for bound ligands can be viewed and downloaded on the Structure Summary page provided for each PDB structure at rcsb.org. wwPDB Validation Reports can be downloaded in XML and PDF formats from these Structure Summary pages.

In 2015, the wwPDB co-sponsored a Ligand Validation Workshop, which generated a comprehensive set of community driven recommendations aimed at further improving validation of ligand structures in the PDB (Adams *et al*. 2016; *Structure 24*, 502-508). Progress towards implementation of these recommendations will be reported together with ongoing enhancements to the contents of the CCD and the format of the wwPDB Validation Report.

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