

## Research Data Management: Structure Factors and Atomic Coordinates

Stephen K. Burley, M.D., D.Phil.  
Director, RCSB Protein Data Bank  
Institute for Quantitative Biomedicine  
Rutgers, The State University of New Jersey  
174 Frelinghuysen Road  
Piscataway, NJ 08854

### Abstract

The Protein Data Bank (PDB; [pdb.org](http://pdb.org)) was established in 1971 as the first open-access digital data resource in biology. Today, the PDB archive serves as the single global repository for more than 125,000 experimentally determined atomic-level structures of biological macromolecules (protein, DNA, RNA) and their complexes. The worldwide PDB (wwPDB; [wwpdb.org](http://wwpdb.org)) partnership, the international collaboration that manages the PDB archive, supports Deposition, Biocuration, Validation, and Distribution of PDB data. The mission of the wwPDB organization is to ensure that the PDB archive will continue in perpetuity as a high-quality, open-access digital data resource with no limitations on usage.

Through its global collaboration, the wwPDB has developed OneDep, a unified platform for Deposition, Biocuration, and Validation of 3D biological macromolecules experimentally determined by X-ray crystallography, NMR spectroscopy, and 3D Electron Microscopy. Data are submitted to the PDB archive *via* this OneDep system. OneDep is designed to help the wwPDB and the global structural biology research community meet the challenges of rapidly changing technologies and keep pace with evolving data archiving needs over the coming decades. The PDB archive and the OneDep system are underpinned by an extensible data architecture based on the PDBx/mmCIF dictionary ([mmcif.wwpdb.org](http://mmcif.wwpdb.org)). Community involvement in the development of this data dictionary is coordinated by the wwPDB PDBx/mmCIF Working Group (Chaired by Paul Adams, LBL/UC Berkeley).

At present, ~90% of PDB holdings were derived from diffraction methods. The earliest PDB entry in the PDB archive for which structure factors are available was deposited in 1976. Deposition of structure factors became mandatory in 2008, and ~90% of all crystallographic entries now include

these data.

Management of structure factors and atomic coordinates within the PDB archive will be discussed, with emphasis on current efforts to extend the range and the complexity of the diffraction data and metadata items that can be deposited.

Acknowledgements: The RCSB Protein Data Bank (RCSB PDB; [rcsb.org](http://rcsb.org)) is a founding member of the Worldwide Protein Data Bank organization (wwPDB; [wwpdb.org](http://wwpdb.org)). Additional members of the wwPDB include the Protein Data Bank in Europe (PDBe; [pdbe.org](http://pdbe.org)), Protein Data Bank Japan (PDBj; [pdbj.org](http://pdbj.org)), and BioMagResBank (BMRB; [bmr.org](http://bmr.org)). Core RCSB PDB operations are funded by a grant to SKB (NSF DBI-1338415) from the National Science Foundation, the National Institutes of Health, and the US Department of Energy.