

## **RCSB PROTEIN DATA BANK: Enabling Breakthroughs in Biomedical Research and Structure-Guided Drug Discovery**

Christine Zardecki   Chenghua Shao   John Westbrook   Zukang Feng Jasmine Young

Protein Data Bank (PDB) was established as the 1<sup>st</sup> open access digital data resource in biology and medicine. Today, the PDB archive houses ~140,000 atomic-level biomolecular structures determined by crystallography, NMR spectroscopy, and 3D electron microscopy. It is managed by the Worldwide Protein Data Bank partnership (wwPDB; [wwpdb.org](http://wwpdb.org)).

For Data Depositors, RCSB PDB (RCSB.org) has developed the Group Deposition system (GroupDep) to support automated depositions of large numbers of X-ray structures in parallel. It allows PDB depositors to take advantage of local templates and pre-deposition tools for batch processing, data packaging, upload, review, validation, and one-click submission of many closely related structures at the same time. Data content features that can be submitted include ligand binding assay information and identification of ligand bound vs unbound structures. Structures submitted via GroupDep can be received, biocurated, validated, and publicly released in a relatively short period of time.

GroupDep was developed to support data coming from the NIH-NIGMS funded Drug Design Data Resource (D3R), which aims to advance the technology of computer-aided drug discovery through the interchange of high quality protein-ligand datasets and workflows, and by holding community-wide, blinded docking/scoring prediction challenges.

Depositors interested in testing the GroupDep system with large datasets should contact [info@rcsb.org](mailto:info@rcsb.org). GroupDep features will be incorporated into the wwPDB OneDep system in the future to support all methods.

For Data Consumers, the PDB archive provides many services that support structure-guided drug discovery. The archive provides no-cost access to structural biology data from >30,000 public and private sector researchers, with no limitations on usage. The high data quality and integrity in PDB data are ensured by 15 experienced PhD-level wwPDB data scientists. Each PDB structure is provided in context with ~140,000 structures across the archive to provide an up-to-date look at the state of structural biology. Interoperation with DrugBank and tools for visualizing ligand binding support exploration of co-crystal structures of proteins with >4,000 US FDA-approved drugs and experimental agents.

Access to PDB data contribute to patent applications, drug discovery and development, publication of scientific studies, innovations that can lead to new product development and company formation, and STEM education.

RCSB PDB is funded by the NSF (DBI-1338415), NIH, and DOE.