

wwPDB Biocuration: A Global Enterprise Supporting High-Quality Open Biological Data

Jasmine Y. Young^{1*}, Sanchayita Sen², John M. Berrisford², Reiko Igarashi³, wwPDB Biocuration Team^{1,2,3,4}, wwPDB OneDep Team^{1,2,3,4}, John L. Markley⁴, Genji Kurisu³, Sameer Velankar², and Stephen K. Burley^{1,5}

¹RCSB Protein Data Bank, Center for Integrative Proteomics Research, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA.

²Protein Data Bank in Europe (PDBe), European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK.

³PDBj, Institute for Protein Research, Osaka University, Osaka, 565-0871, Japan.

⁴BMRB, BioMagResBank, University of Wisconsin-Madison, Madison, WI 53706, USA.

⁵RCSB Protein Data Bank, San Diego Supercomputer Center, University of California San Diego, La Jolla, CA 92093, USA.

The Protein Data Bank (PDB) is the single global repository for experimentally determined three-dimensional structures of biological macromolecules and their complexes with ligands. The Worldwide Protein Data Bank (wwPDB) is the international collaboration that manages the PDB archive according to the *FAIR* Principles: Findability, Accessibility, Interoperability, and Reusability. The wwPDB recently developed OneDep, a unified tool for deposition, validation, and biocuration of structures of biological macromolecules. All data deposited to the PDB undergo critical review by wwPDB biocurators. This presentation outlines the importance of biocuration for structural biology data deposited to the PDB, and describes wwPDB biocuration processes and the role of expert biocurators in sustaining a high-quality archive. Structural data submitted to the PDB are examined for self-consistency, standardized using controlled vocabularies, cross-referenced with other biological data resources, and validated for scientific/technical accuracy. We illustrate how biocuration is integral to PDB data archiving, as it facilitates accurate, consistent, and comprehensive representation of biological structure data, allowing efficient and effective usage by research scientists, educators, students, and the curious public worldwide.

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