

*Small-molecule ligand/drug representation and validation in the Protein Data Bank*Genji Kurisu<sup>1</sup>, Stephen K. Burley<sup>2</sup>, John L. Markley<sup>3</sup>, Haruki Nakamura<sup>1</sup>, Sameer Velankar<sup>4</sup><sup>1</sup>PDBj, Institute For Protein Research, Osaka University, Suita, Osaka, Japan, <sup>2</sup>RCSB PDB, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, United States, <sup>3</sup>BMRB, BioMagResBank, University of Wisconsin-Madison, Madison, WI 53706, United States, <sup>4</sup>PDBe, European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, United Kingdom

E-mail: gkurisu@protein.osaka-u.ac.jp

The Protein Data Bank (PDB) was established in 1971 with just seven X-ray crystallographic structures of proteins as the first open access digital resource in the biological sciences [1]. Currently, the global PDB Archive holds ~134,000 deposited and processed coordinates of biological macromolecules. Every structure deposited to the PDB undergoes an extensive validation process, which was developed in consultation with community stakeholders by the Worldwide Protein Data Bank organization (wwpdb.org) that is responsible for managing the PDB Archive. During structure determination and refinement, experimentalists are encouraged to access the anonymous wwPDB Validation Server (<http://wwpdb-validation.wwpdb.org/>) to evaluate structure quality prior to deposition. For ligands, the watermarked, official wwPDB Validation Report includes both geometrical and model fit diagnostic information.

More than three quarters of the experimentally determined PDB entries include at least one non-polymeric small molecule ligand. Small-molecule ligands/drugs found in PDB entries are described in detail in the PDB Chemical Component Dictionary (CCD), which utilizes the data representation for chemical components developed for the Macromolecular Crystallographic Information Framework data dictionary [2]. Each chemical definition in the CCD includes descriptions of chemical properties, such as stereochemical assignments, chemical descriptors, and systematic chemical names. A set of atomic model coordinates from a selected experimental entry in PDB, and a computed set of ideal atomic coordinates or the Cambridge Structural Database (CSD) are provided for each entry in the CCD. The outcome of a recent wwPDB/CCDC/D3R Ligand Validation Workshop was published in 2016 [3]. Plans of wwPDB to implement recommendations therefrom will be described.

wwPDB Members: RCSB PDB (supported by NSF DBI-1338415, NIH, and DOE); PDBe (EMBL-EBI, Wellcome Trust, BBSRC, EU, and MRC); PDBj (JST-NBDC); and BMRB (NIGMS).

[1] Berman, H. M., Henrick, K. & Nakamura, H. (2003). *Nat Struct Biol* 10, 980.

[2] Westbrook, J. D., Shao, C., Feng, Z., Zhuravleva, M., Velankar, S. & Young, J. (2015). *Bioinformatics* 31, 1274-1278.

[3] Adams, P. D., Aertgeerts, K., Bauer, C., Bell, J. A., Berman, H. M., Bhat, T. N., Blaney, J. M., Bolton, E., Bricogne, G., Brown, D., Burley, S. K., Case, D. A., Clark, K. L., Darden, T., Emsley, P., Feher, V. A., Feng, Z., Groom, C. R., Harris, S. F., Hendle, J., Holder, T., Joachimiak, A., Kleywegt, G. J., Krojer, T., Marcotrigiano, J., Mark, A. E., Markley, J. L., Miller, M., Minor, W., Montelione, G. T., Murshudov, G., Nakagawa, A., Nakamura, H., Nicholls, A., Nicklaus, M., Nolte, R. T., Padyana, A. K., Peishoff, C. E., Pieniazek, S., Read, R. J., Shao, C., Sheriff, S., Smart, O., Soisson, S., Spurlino, J., Stouch, T., Svobodova, R., Tempel, W., Terwilliger, T. C., Tronrud, D., Velankar, S., Ward, S. C., Warren, G. L., Westbrook, J. D., Williams, P., Yang, H. & Young, J. (2016). *Structure* 24, 502-508.

**Keywords:** [Protein Data Bank](#), [validation](#), [small-molecule](#)