PDB Depositions with RCSB PDB Tools

Zardecki Christine

The RCSB PDB (Berman *et al.*, 2000) has developed tools to make depositing structure data easier. There are a few steps a depos itor can take to make the process of depositing a structure to the PDB quick, easy, and accurate! It is an iterative process – if you encounter problems at a particular step, please make the correction(s) and go through the steps again.

1.Use the pdb_extra ct Program Suite (Yang *et al.*, 2004)

This software tool extracts information about data collection, phasing, density modification, and the final structure refinement from the output files output files produced by many applications used for structure determination. The collected information is organized into an mmCIF file that is ready for deposition. Since pdb_extract fills in many data fields in this mmCIF file, fewer data items have to be manually entered -- saving time and minimizing errors.

The online version of pdb-extract (http://pdb-extract.rutgers.edu/) is used to upload output files and create mmCIF files. The files can be used as input to ADIT for a complete deposition.

pdb_extract can be downloaded in source and binary versions for Linux, SGI, SUN, OSF and Mac OSX from http://sw-tools.pdb.org/. Source and Linux binary versions of ADIT are also available.

pdb_extract is also part of the CCP4i inte rface.

2. Check your structure with the Validation Suite and Server

(Westbrook *et al.*, 2003)

to ensure that the data being deposited is accurate and reflects what you intend to submit (http://deposit.pdb.org/validate/).

This tool is used by the RCSB PDB for checking an d processing structure data. The format consistency of the coordinates can be reviewed, along with validation reports

from MOLPROBITY (Davis *et al.*, 2004), NUCheck (Feng *et al.*, 1998), PROCHECK (Laskowski *et al.*, 1992), and SFCHECK (Vaguine *et al.*, 1999). Sequence/coordinate alignment, missing and extra atoms or residues, and data inconsistencies are also reported here.

The Validation Server is also available from PDB Japan (PDBj) at http://pdbdep.protein.osaka-u.ac.jp/validate/.

3. Run BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) to compare your sequence to sequence database references (Wheeler *et al.*, 2005). Any necessary corrections can then be made to your sequence and coordinates.

4. Use Ligand Depot

at http://ligand-depot.rutgers.edu/ to find the proper codes for existing ligands, to link to other entries with a particular ligand, and to search for substructures (Feng *et al.*, 2004). If the ligand is not present in the dictionary, please email detailed information (complete chemical name, 2D figure showing connectivity, bond order and sterochemistry) along with the RCSB and PDB IDs of the entries they are associated to expedite the processing of these files.

5. Deposit your structure using ADIT (http://deposit.pdb.org/adit/).

Using its editor to add any missing information to the deposition. A desktop version of ADIT is also available.

ADIT is also available from PDBj at http://pdbdep.protein.osaka-u.ac.jp/adit/.

For First Time Depositors

For a detailed packet of information about these tools, including reprints about validation and Ligand Depot, please send your postal address to info@rcsb.org with the subject line "first time depositor packet".

Acknowledgements

The Protein Data Bank (PDB) is operated by Rutgers, The State University of New Jersey and the San Diego Supercomputer Center at the University of California, San Diego -- two members of the Research Collaboratory for Structural Bioinformatics (RCSB). The RCSB PDB is supported by funds from the National Science Foundation (NSF), the National Institute of General Medical Sciences (NIGMS), the Office of Science, Department of Energy (DOE), the National Library of Medicine (NLM), the National Cancer Institute (NCI), the National Center for Research Resources (NCRR), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), and the National Institute of Neurological Disorders and Stroke (NINDS).

The RCSB PDB is a member of the wwPDB.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235-242.

Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2004). *Nucleic Acids Res* **32**, W615-619.

Feng, Z., Chen, L., Maddula, H., Akcan, O., Oughtred, R., Berman, H. M. & Westbrook, J. (2004). *Bioinformatics* **20**, 2153-2155.

Feng, Z., Westbrook, J. & Berman, H. M. (1998). Report NDB-407. Rutgers University, New Brunswick, NJ.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1992). *PROCHECK v.2.0 - programs to check the stereochemical quality of protein structures.*

Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Crystallogr D Biol Crystallogr* **55**, 191-205.

Westbrook, J., Feng, Z., Burkhardt, K. & Berman, H. M. (2003). *Meth Enz* **374**, 370-385.

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S., Helmberg, W., Kenton, D. L., Khovayko, O., Lipman, D. J., Madden, T. L., Maglott, D. R., Ostell, J., Pontius, J. U., Pruitt, K. D., Schuler, G. D., Schriml, L. M., Sequeira, E., Sherry, S. T., Sirotkin, K., Starchenko, G., Suzek, T. O., Tatusov, R., Tatusova, T. A., Wagner, L. & Yaschenko, E. (2005). *Nucleic Acids Res* **33**, D39-45.

Yang, H., Guranovic, V., Dutta, S., Feng, Z., Berman, H. M. & Westbrook, J. (2004). *Acta Crystallogr D Biol Crystallogr* **60**, 1833-1839.