STRUCTURAL DATABASES OF BIOMOLECULES, THE PDB AND NDB

Bohdan Schneider

Center for Biomolecules and Complex Molecular Systems, Institute of Organic Chemistry and Biochemistry, AS CR, Fleming Sq. 2, CZ-16610 Prague, Czech Republic

THE PROTEIN DATA BANK, PDB [1], http://www.pdb.org/, serves for deposition, data processing and distribution of the biomolecules; it is the international archive for three-dimensional structural data of proteins, nucleic acids, virus particles, and saccharides determined experimentally by x-ray crystallography, NMR techniques, electron (cryo)microscopy, and neutron diffraction. It is jointly operated by the Worldwide Protein Data Bank (wwPDB, http://www.wwpdb.org/). The PDB has developed many tools for the deposition and validation of structures easily available from the web. The archive is available in a form of the relations database with extensive tools to query and report; the web also provides many resources for understanding the structure of biological macromolecules and access to files of both coordinates and "experimental" files (structure facotrs for x-ray, distance constraint files for NMR structures). The tools developed for internal processing of depositions are also used for depositions [2] taking advantage of the mmCIF dictionaries (//ndbserver.rutgers.edu/mmcif/, [3]). The wwPDB has collaborated on a project to remediate the PDB archive and create a new set of corrected files; the whole archive has been remediated. The most significant changes include extensive corrections in the ligand dictionary ("Chemical component dictionary") where redundant ligand definitions have been removed, ligand naming corrected, including stereochemistry at all chiral centers, atom naming was harmonized with the IUPAC nomenclature as much as possible. Of importance are also updates of many sequence and publication references; all existing sequence references point uniformly to the UniProt (UNP) database.

NUCLEIC ACID DATABASE, NDB [4], http://ndbserver.rutgers.edu, was established in 1991 as a resource for experts on nucleic acid structures. and contains both X-ray and NMR structures containing dinucleotide and longer sequences. The core of the NDB is its relational database of primary and derivative data with rich query and reporting capabilities. A popular feature of the NDB is the "Structure Atlas". The Atlas pages include not only the basic information on structures as publication details and thumbnail image but also links to several tables of nucleic acid-specific properties as base morphology, backbone geometrical features, and tables of base pairing interactions, including their 2D diagrams; these are especially useful for larger RNA molecules with complicated folds.

DATA DISTRIBUTION.
Coordinate files, database reports, software programs, and other resources are freely available from the web pages of both databases.

1. Berman H.M., Battistuz T., Bhat T.N., Bluhm W.F., Bourne P.E., Burkhardt K., Feng Z., Gilliland G.L., Iype L., Jain S., Fagan P., Marvin J., Padilla D., Ravichandran V., Schneider B., Thanki N., Weissig H., Westbrook J.D., Zardecki, C. (2002): The Protein Data Bank. **Acta Crystallogr D, 58**, 899-907.

2. Westbrook J., Feng Z., Berman H.M. (1998): ADIT—The AutoDep Input Tool. Department of Chemistry, Rutgers, the State University of New Jersey, RCSB-99.

3. Bourne P.E., Berman H.M., Watenpaugh K., Westbrook J.D., Fitzgerald P.M.D. (1997): The macromolecular crystallographic information file (mmCIF). *Methods Enzymol.* **277**, 571–590.

4. Berman H.M., Olson W.K., Beveridge D.L., Westbrook J., Gelbin A., Demeny T., Hsieh S.-H., Srinivasan A.R., Schneider B. (1992): The Nucleic Acid Database—a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**,751–759.