The IMB Jena Image Library of Biological Macromolecules: 2002 update

Jan Reichert and Jürgen Sühnel*

Biocomputing Group, Institut für Molekulare Biotechnologie, Postfach 100813, D-07708 Jena, Germany

ABSTRACT

The IMB Jena Image Library of Biological Macromolecules (http://www.imb-jena.de/IMAGE.html) is aimed at a better dissemination of information on three-dimensional biopolymer structures with an emphasis on visualization and analysis. It provides access to all structure entries deposited at the Protein Data Bank (PDB) and Nucleic Acid Database (NDB). In addition, basic information on the architecture of biological macromolecules is offered. Recent developments include a site database and an analysis tool that identifies all residues surrounding hetero components or sites according to geometrical criteria. This enables one to search for all structures with a certain pattern of amino acids/nucleotides/water adjacent to hetero components or sites. A new PDB/SWISS-PROT cross-reference database combines information from both PDB and SWISS-PROT, thus providing significantly more cross-references than either PDB or SWISS-PROT. The existing brief descriptions of X-ray, NMR and FTIR methods for structure determination are supplemented by information on circular dichroism.

INTRODUCTION

The IMB Jena Image Library of Biological Macromolecules (http://www.imb-jena.de/IMAGE.html) is aimed at a better dissemination of information on three-dimensional biopolymer structures with an emphasis on visualization and analysis (1,2). It provides access to all structure entries deposited at the Protein Data Bank (PDB) and Nucleic Acid Database (NDB) and also offers basic information on the architecture of biological macromolecules (3–5). One of the aims of the database is to provide as much information as possible in one place about a particular entry. In addition, with the rapidly increasing number of new structures, there is a need to generate classification schemes that allow an easy navigation through the complete structure set from different points of view.

SITE AND HETERO COMPONENT DATABASES

The IMB Jena Image Library includes a hetero components database that provides a complete listing of all hetero components occurring in the PDB. Compilations of proteins, protein–nucleic acid complexes, nucleic acids and carbohydrates listed by hetero components are offered for browsing. In addition, searches for hetero identifiers and names, chemical elements in the hetero components, and for any character string in the PDB title record are possible. In analogy to the hetero component database we have developed a new site database. It allows, for example, generation of a listing of all components involved in active sites as defined in the corresponding PDB record. Search options include the site ID and description as well as the structure title.

Both the hetero components and the site databases have been supplemented by an environment option. All PDB entries with sites or hetero components have been analyzed to identify the surrounding residues according to a geometrical criterion. All residues are taken into account that have at least one atom within a distance of 4.2 Å from any of the atoms of a specific site or hetero component. The results of this analysis are stored in the database and can be searched. A typical query is: find all entries where a site or a hetero component is surrounded by four methionines and two waters. Individual sites or hetero components can also be visualized together with the surrounding residues.

PDB/SWISS-PROT CROSS-REFERENCES

We have generated a new PDB/SWISS-PROT cross-reference database combining the information from PDB and SWISS-PROT (6). As of September 7 2001 the Image Library provides 14 970 PDB/SWISS-PROT cross-references between 4131 SWISS-PROT and 13 086 PDB entries. These figures have to be compared to the corresponding data from SWISS-PROT (10 305 cross-references for 3082 entries) and PDB (10 496 cross-references for 9161 structures). Thus, our database currently includes about 4000 more cross-references than either SWISS-PROT or PDB. It should be mentioned, however, that the PDB data given only take into account cross-references included directly in the PDB files. Our experience is that the Sequence Data category of the Other Sources section at the PDB site provides more cross-references to SWISS-PROT than indicated in the structure files. The cross-reference database can also be used to generate entry tables listed by PDB code, SWISS-PROT entry name, protein name or genus/species. Finally, the cross-reference scheme is used to list on each atlas page related entries sharing the same protein chains and also for links to a few other databases.

*To whom correspondence should be addressed. Tel: +49 3641 65 6200; Fax: +49 3641 65 6210; Email: jsuehnel@imb-jena.de

BENDING CLASSIFICATION OF NUCLEIC ACID DOUBLE HELIX STRUCTURES

The Image Library offers a tool for the uniform analysis of the helix and bending geometry of all nucleic acid structures with at least six consecutive base pairs. The results of this analysis have been summarized in a new and comprehensive classification scheme currently for about 900 nucleic acid structures. The classification is carried out according to shape of the helical axis (straight, circular, single-kink, double-kink lines) and according to the extent of bending. In this manner the most strongly bent structures currently known can be identified at a glance. The classification is updated on a regular basis.

MISCELLANEOUS

The already existing brief descriptions of experimental methods for biopolymer structure determination by diffraction methods, NMR and Fourier transform infrared spectroscopy have been supplemented by a contribution on circular dichroism.

The number of databases linked to the atlas pages is further increased and also includes relatively new data resources, such as InterPro (7) and SMART (8), that are not yet taken into account by other structure databases. Finally, a new tool allows for the visualization of user-supplied files.

ACKNOWLEDGEMENTS

REFERENCES

1. Sühnel,J. (1996) Image library of biological macromolecules. *Comput. Appl. Biosci.*, **12**, 227–229.
2. Reichert,J., Jabs,A., Slickers,P. and Sühnel,J. (2000) The IMB Jena Image Library of biological macromolecules. *Nucleic Acids Res.*, **28**, 246–249.
3. Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,E.E.,Jr, Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) The PDB: a computer-based archival file for macromolecule structures. *J. Mol. Biol.*, **112**, 535–542.
4. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242. Updated article in this issue: *Nucleic Acids Res.* (2002), **30**, 245–248.
5. Berman,H.M., Olson,W.K., Beveridge,D.L., Westbrook,J., Gelbin,A., Demeny,T., Hsieh,S.H., Srinivasan,A.R. and Schneider,B. (1992) The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.*, **63**, 751–759.
6. Bairoch,A. and Apweiler,R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, **28**, 45–48.
7. Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Birney,E., Biswas,M., Bucher,P., Cerutti,L., Corpet,F., Croning,M.D. *et al.* (2001) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.*, **29**, 37–40.
8. Schultz,J., Copley,R.R., Doerks,T., Ponting,C.P. and Bork,P. (2000) SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.*, **28**, 231–234. Updated article in this issue: *Nucleic Acids Res.* (2002), **30**, 242–244.