PDBj

Protein Data Bank Japan

News Letter Vol. 6

August 2005

http://www.pdbj.org/

PDBj is maintained at the Protein Research Institute, Osaka University, and supported by Japan Science and Technology Agency.

1. News

Workshop about "Enhancement and standardization of biological information databases" was held at THE 42nd ANNUAL MEETING OF THE BIOPHYSICAL SOCIETY OF JAPAN, Kyoto on December 15th, 2004.

Primary databases, where experimentally observed biological data are deposited, are on the increase. Distributed and integrated databases with Data Grid technology and standardized descriptions have been developed. These recent technologies enable various applications of biological information in a multitude of directions. The DNA Data Bank of Japan (DDBj) and PDBj are the primary databases for DNA and Proteins, respectively, in Japan, and they organized this workshop at the annual meeting of the Biophysical Society of Japan. In this Workshop, DDBj and PDBj were introduced through actual examples of such new technologies, and future goals were discussed.

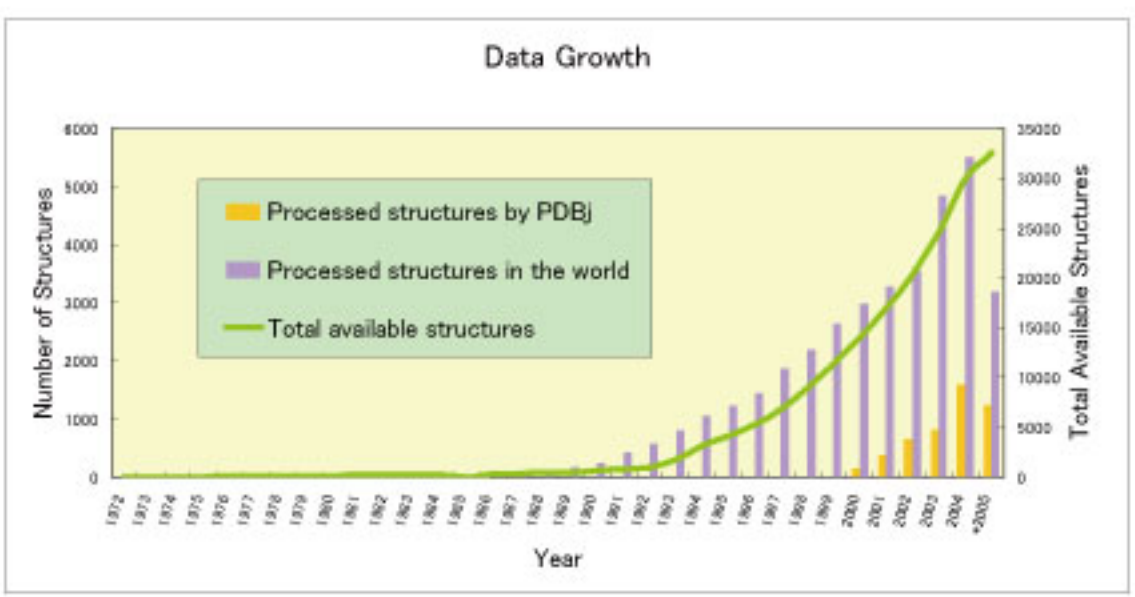
DDBJing & PDBjing - Workshop in Osaka –

The 11th DDBJing and PDBjing - Workshop in Osaka- was held March 2nd, 2005, at Osaka University, Nakanoshima Center. We introduced our activities to non-specialists using PCs and literature in cooperation with DDBj.

Snapshot of the Workshop.

2. PDB Deposition Statistics

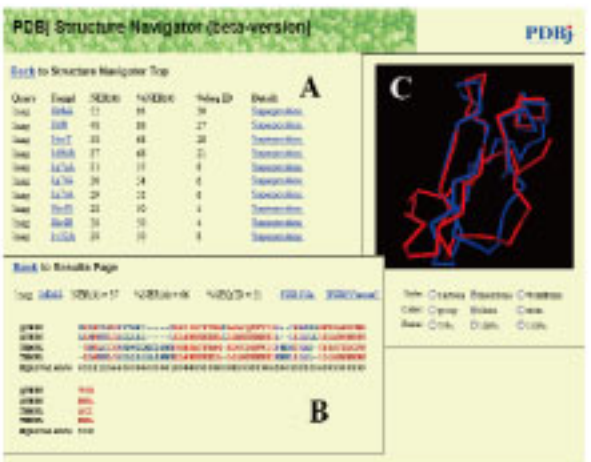
* Last updated: Jul. 20th, 2005



3. An Improved Structure Navigator and GASH Structure

In the previous newsletter we introduced "Sequence Navigator", and "Structure Navigator", web-based tools for finding sequence or structure neighbors, respectively, to a given PDB entry. We also introduced a new structure alignment program: GASH. In this article, major improvements to Structure Navigator and GASH are described.

Structure Navigator has been re-written to take advantage of our distributed computing platform: a cluster of 15 PCs. Previously, Structure Navigator had to perform a considerable number of operations in order to produce a list of structures similar to the query structure. These operations have now been carried out and the results stored in advance for nearly every PDB id so that structure neighbors can be easily retrieved.



Structure Navigator displays a list of structure neighbors (A) individual alignments (B) and structural superpositions (C) to the query structure.

When the structure neighbor lists were being computed in advance, sequence homologous structures were included. The user can now choose to filter out sequence homologs or not, meaning most of the functionality of Sequence Navigator has been incorporated into Structure Navigator. The one exception is when the query consists of an actual sequence, rather than a PDB id. For such queries, it is currently still necessary to use Sequence Navigator.

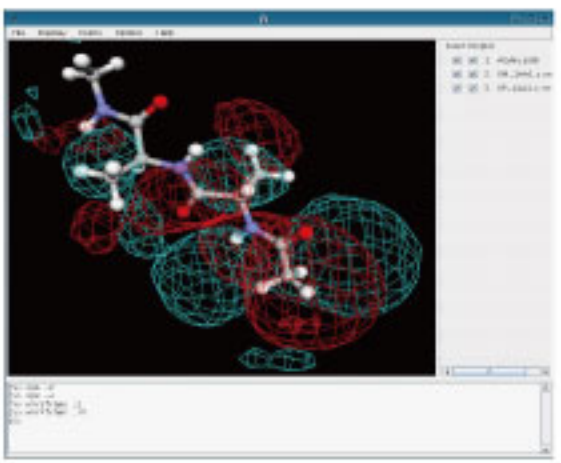
One of the challenges in constructing the new Structure Navigator and associated data bank was to store nearly a terabyte of data in such a way that it could be searched easily. This problem was solved by partitioning the data across 14 local disks on the PC cluster. These changes, in addition to a re-writing of the code has reduced the query time from up to a minute to just a few seconds.

In parallel with efforts to make the underlying code and data bank more efficient, a new XML version of the input and output data has been developed. The XML data can be accessed by a java-based application or a simple object access protocol (SOAP) server. This implementation is much more secure than our previous CGI interface.

The GASH structure alignment program has also been improved. The code has been tuned so that the average CPU time for a query under 500 residues is approximately 15 seconds, and reduces linearly with the size to about 5 seconds for a query under 100 residues. This is much better than the previous version which required up to a minute to complete an alignment. GASH has been benchmarked against DaliLite and CE on a dataset of over 3,000 structure pairs. In terms of either the number of aligned residues under a fixed RMSD, or the number of equivalent residues (NER), GASH consistently performed as well as or better than DaliLite or CE.

As with Structure Navigator, the GASH input and output data has been described in XML format. Both a web-based java interface and a command-line SOAP service that take advantage of the XML version of GASH data are available through PDBj.

4. New Developments in Electron Density Map with xPSSS



Molecular orbitals of Ala-Ala

Volume 5 of the PDBj newsletter in December 2004 introduced the concepts of the Electron Density Map (EDM) service of the PDBj. Here we report on the further development of EDM.

The initial EDM server had been limited to the PDB data available in mid-2004. Since then the number of available structural models and deposited structure factor data has grown substantially, from 12,310 entries with matching structure factor data to 16,857 entries as of July 1, 2005.

Refinement of deposited structures against deposited data and calculation of electron density maps for successfully refined structures, as described previously, is performed via Python scripts interfacing with the appropriate CCP4 software. To keep the EDM server up-to-date, we have extended these scripts further to automate the process completely. Reference is the PDB mirror maintained by the PDBj in Japan at ftp://pdb.protein.osaka-u.ac.jp/pub/pdb/. This mirror is updated weekly. We keep a local copy of PDB and structure factor data files on our EDM server, download new data every Wednesday from the mirror and carry out refinement and map calculation for new entries, while deleting obsolete entries. As a result our EDM server now contains 16,207 refined structures and 15,250 electron density maps.

Both the stand-alone *jV* viewer and the *jV* Java applet use OpenGL and require an additional jogl library as described in http://www.pdbj.org/PDBjViewer/. The Java applet has been tested successfully with Opera 8.01 and Mozilla Firefox 1.0.4 under Linux and with Internet Explorer under Windows. The recommend Java run-time environment is 1.4.2_05, but newer versions work as well and version 1.5.0_03 has been used under Linux.

The software used by the EDM server to generate electron density map contours and isosurfaces operates on three-dimensional grid data. It can naturally be extended to any other such data and we have done so for data produced by molecular modeling software.

The software that we have tested successfully produces charge or molecular orbital densities and stores these results in XML files, in a format referred to as BMSML (biomolecular structural markup language). A software API to parse these files is provided both for use with C and Fortran. We have used the Fortran API to create Python bindings.

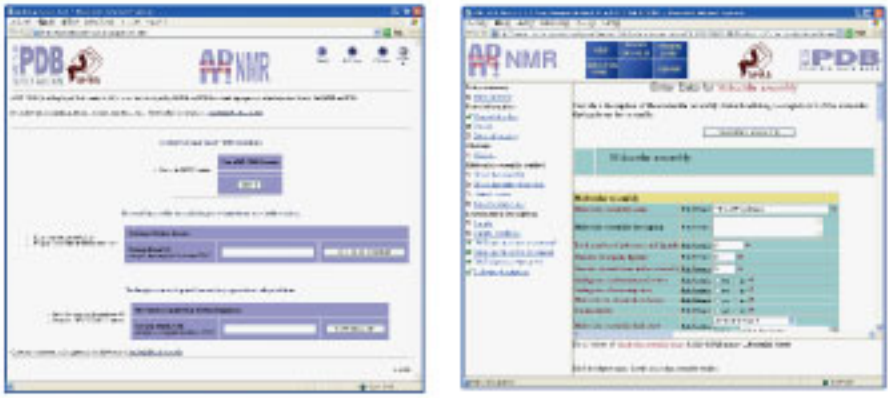
Existing contouring software used for EDM could be modified with ease to operate on such BMSML files and to produce contour or isosurface files in the XML polygon file format used previously for electron density map displays. The illustration above shows one such example, displaying molecular orbital contours, blue for positive and red for negative for an Ala-Ala dipeptide model, The positive and negative contours have been generated as separate files so that their display can be selectively turned on or off.

The table shows the substantial growth in entries. The number of refined entries is roughly 50% of the total number of entries deposited. Considering the age of the PDB, the growth is remarkable and can be attributed to the activities of structural genomics initiatives.

	June 2004	June 2005	increase[%]
refined entries	11,971	16,207	35.4
density maps	11,284	15,250	35.1

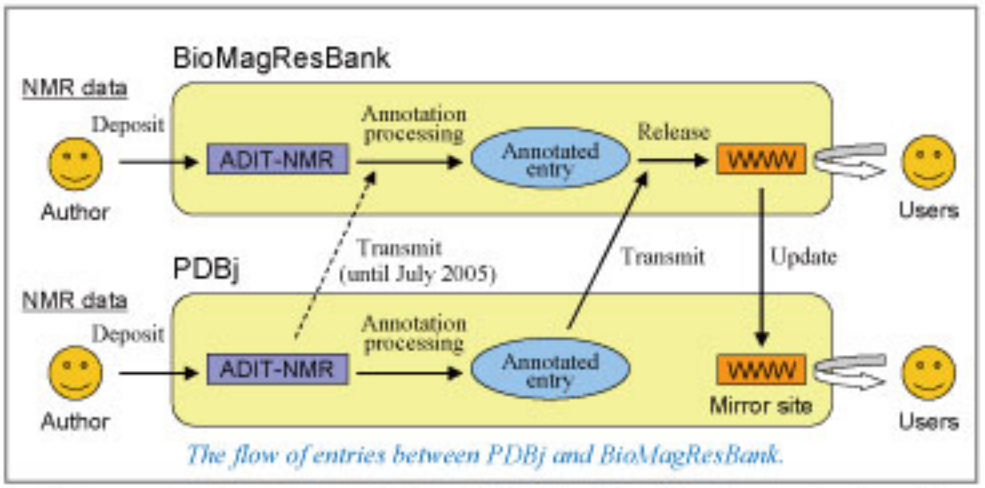
5. BMRB Mirror and Annotations

The NMR database group (Group leader: Prof. Hideo Akutsu) is maintaining a mirror site of BMRB (BioMagResBank UW-Madison, WI, USA; PI, John L. Markley) at PDBj from March 2002. Since ADIT-NMR, a web deposition interface of BMRB like ADIT for PDB, was established at PDBj, we began to accept NMR data for deposition from December 2004.



The top page (left) and a session screen (right) for data input of ADIT-NMR.

URL: http://bmrbadit.protein.osaka-u.ac.jp/bmrb-adit/



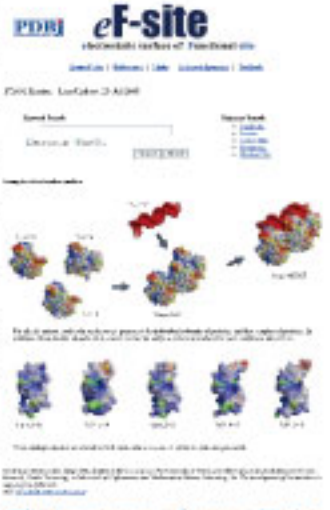
The flow of entries between PDBj and BioMagResBank.

We have started annotation processing of NMR data for deposited entries at PDBj from August 2005. Entries are annotated by each site in charge. All files for an entry generated at PDBj are transmitted to BioMagResBank after annotation processing. These entries are added to the BMRB database WWW servers and are released to the public.

6. New eF-site Database Appears

The eF-site database describing the surface of biomolecules has completely changed in internal composition. Almost all PDB entries have been registered and are open to the public. Currently only entries that are too large and the calculation too difficult have not been processed. Moreover, the new eF-site will be synchronized with the PDB, with updates nearly every day. Our aim is to contribute to research on the interaction between proteins which is becoming more and more important in the future.

Although eF-site was developed as a primary database of molecular surfaces, it will also be useful for the development of protein-protein interaction prediction methods (i.e. docking). To this end, the surface of the each molecule and its electrostatic potential are calculated for both subunits and whole structures each time a new PDB entry is registered. Moreover, all models are treated and registered in this way when two or more structures are associated with a single PDB, (e.g. NMR structures). As a result, 170,733 entries are registered as of July 6, 2005.



The top page of new eF-site.

7. The members in Osaka University

From Left to right:
Back: M. Kusunoki, Y. Ikegawa, R. Yamashita, Y. Matsuki, A. Paehler, A. Yoshihara,
Middle: Y. Kengaku, R. Igarashi, Y. Shimizu, E. Nakatani, A. Takahashi, D. Standley,
Front: C. Kamada, S. Saeki, H. Nakamura, T. Kosada

Director
Nakamura, Haruki, Prof. (IPR, Osaka Univ.)

Group for PDB Database Curation
Kusunoki, Masami, Dr. (IPR, Osaka Univ.)
Kosada, Takashi (IPR, Osaka Univ.)
Igarashi, Reiko (JST-BIRD)
Kengaku, Yumiko (JST-BIRD)
Ikegawa, Yasuyo (JST-NIRD)
Saeki, Shinobu (JST-BIRD)

Group for Development of PDBj database
Ito, Nobutoshi, Prof. (Tokyo Med. Dent. Univ.)
Paehler, Arno, Dr. (JST-BIRD)
Standley, Daron M., Dr. (JST-BIRD)
Yamashita, Reiko (JST-BIRD)
Yoshihara, Atsuro (JST-BIRD)
Shimizu, Yukiko (JST-BIRD)
Takahashi, Aki (JST-BIRD)

Group for Development of Secondary Database and Analysis Systems
Ohkawa, Takenao, Prof. (Kobe Univ.)
Kinoshita, Kengo, Dr. (IMSUT, Tokyo Univ.)
Toh, Hiroyuki, Prof. (MIB, Kyushu Univ.)
Wako, Hiroshi, Prof.(Waseda Univ.)

Group for BMRB Database
Akutsu, Hideo, Prof. (IPR, Osaka Univ.)
Matsuki, Yoh, Dr. (JST-BIRD)
Nakatani, Eiichi (JST-BIRD)

Secretary
Kamada, Chisa (JST-BIRD)

PDBj
Research Center for Structural and Functional Proteomics,
Institute for Protein Research (IPR), Osaka University
3-2 Yamadaoka, Suita, Osaka 565-0871, Japan
TEL (PDBj office): +81-(0)6-6879-4311
TEL (PDBj deposition office): +81-(0)6-6879-8638
FAX: +81-(0)6-6879-8636
E-mail: pdbjADMIN@protein.osaka-u.ac.jp