PDBj

Protein Data Bank Japan

# News Letter Vol. 5

## December 2004

http://www.pdbj.org/

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

# 1. News

## The wwPDB Advisory Committee Meeting

The first world wide Protein Data Bank Advisory Committee (wwPDBAC) meeting was held on November 21st, 2004, at Washington, D.C., USA, chaired by Dr. Stephen K. Burley (Structural GenomiX, Inc.). The other participants were, (1) PDB Site Representatives: Prof. Wayne A. Hendrickson (RCSB, Columbia University), Dr. Gerard Kleywegt (MSD, Uppsala University), Prof. James H. Naismith (MSD, The University, St. Andrews), Prof. Yoshifumi Nishimura (PDBj, Yokohama City University), and Prof. Soichi Wakatsuki (PDBj, Photon Factory at Tsukuba), (2) Community Stakeholder Representatives: Prof. Edward N. Baker (IUCr, University of Auckland), and Dr. R. Andrew Byrd (ICMRBS, National Cancer Institute), (3) PDB Site Leaders: Prof. Helen M. Berman (RCSB, Rutgers University), Dr. Kim Henrick (MSD, EBI), and Prof. Haruki Nakamura (PDBj, Osaka University), and (4) Funding Agency Representatives: Dr. Christopher L. Greer (NSF), Dr. Michael Dunn (Wellcome Trust), Dr. Deborah Colson (Wellcome Trust), Dr. Josefina Enfedaque (European Commission), Mr. Toshiyuki Koike (JST), and Mr. Hideki Takahashi (JST).

As described in the Charter Agreement for wwPDB, the mission of the wwPDB is to maintain a single archive of macromolecular structural data that is freely and openly available to the global community. The wwPDBAC members firs agree with the wwPDBAC Mission Statement, which is to help ensure that the Protein Data Bank is maintained for the public good as a secure, single, global archive for experimental structural biology data that is freely accessible in perpetuity.

The meeting started from an overview of PDB and wwPDB by Prof. Helen Berman (RCSB), and the three members of wwPDB (RCSB, MSD-EBI, and PDBj) made brief presentations about their activities. Then, the roles and responsibilities of the wwPDBAC were confirmed among all the participants to realize the mission of the wwPDBAC. Finally, the long term issues were discussed about PDB file format, PDB Entry Remediation, inclusion of experimental structural biology data from X-ray crystallography, NMR spectroscopy, and electron microscopy, models purely built *in silico*, quality of the experimental data, and future communications among the wwPDB members and wwPDBAC.

The wwPDBAC has made a report of the meeting, describing the unanimous commentary and recommendations for the above issues. The wwPDB members appreciate their kind recommendations, and will follow them to every extent possible. We are sure that this First wwPDBAC meeting has made a great step toward the future.

## The Company from RCSB

### PDBj hosted Suzanne Richman for shareing the new annotation system. Following is her comment for the stay.

*Suzanne Richman*

To start, I would like to thank PDBj for hosting me. I appreciate how friendly, gracious and very helpful you were from the moment that I arrived.

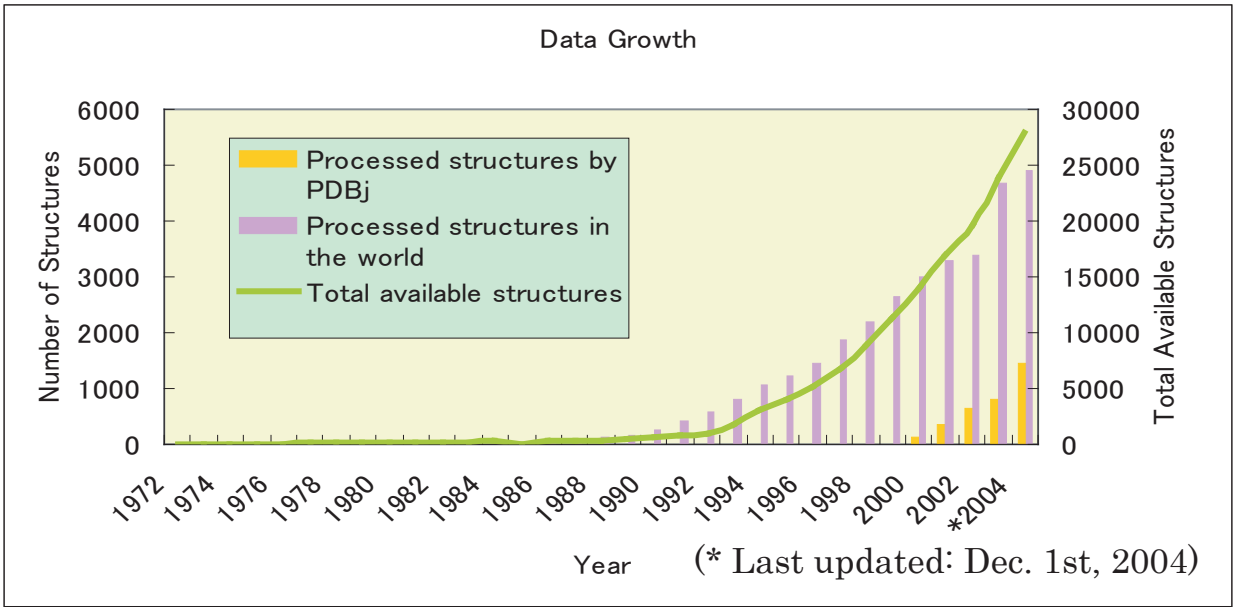
At first, I thought that I was going to stay in the dormitory for two months without leaving since the subway system seemed incomprehensible. I quickly got brave, though, and had a wonderful time exploring Japan, mainly the Kansai region. In our world of globalization, where it is possible even to buy marshmallows in Japan, it was a relief to discover that there are still many aspects of Japanese culture that are truly unique: foods that were all new to me (takoyaki, onigiri, natto, okonomiyaki, tomago sushi, otabe, umeboshi); more parking spaces for bicycles than cars; countless skyscrapers and tall apartment buildings juxtaposed with old Japanese farmhouses; brightly lit pachinko parlors on every corner; the thousands of beautiful temples in Kyoto, to name just a few.

I enjoyed learning vocabulary words and phrases and a little bit of hiragana. Sometimes it was difficult to communicate because of the language barrier; but heart-felt smiles, laughs, bows, drawing of pictures and making of gestures were often as effective as speaking. Despite our differing cultures and languages, working at PDBj felt almost

like home. We are all working on the same project, half a world apart, but with the same thoughts and feelings about it, and in an annotation room that can be just as eerily quiet, as we all work and concentrate hard.

I truly loved every moment of my stay in Osaka, and I am immensely grateful to have had such a great experience.

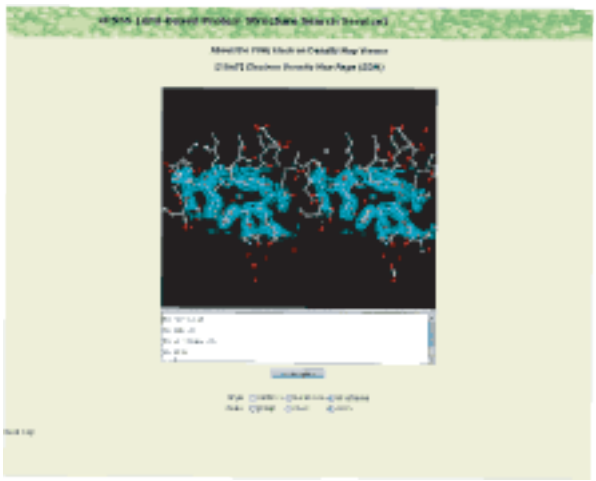
## 2. PDB Deposition Statistics




## 3. Electron Density Maps with xPSSS

The PDBj provides the PDB browser facility xPSSS. Part of this facility is an interactive viewer to visualize structural models. Within a web browser it can be used as a Java applet, but a stand-alone version is also available. We have recently extended the viewer capabilities to allow the display of electron density maps.

The PDB holds at present more than 28,000 entries. 85% of these entries have been determined using X-ray diffraction and for 65% of those entries, experimental structure factor amplitude data have also been supplied by the depositors. This makes it possible to calculate electron density maps. While they may not be familiar to many non-crystallographer users of the PDB, they are important as the basis from which the structures are derived.



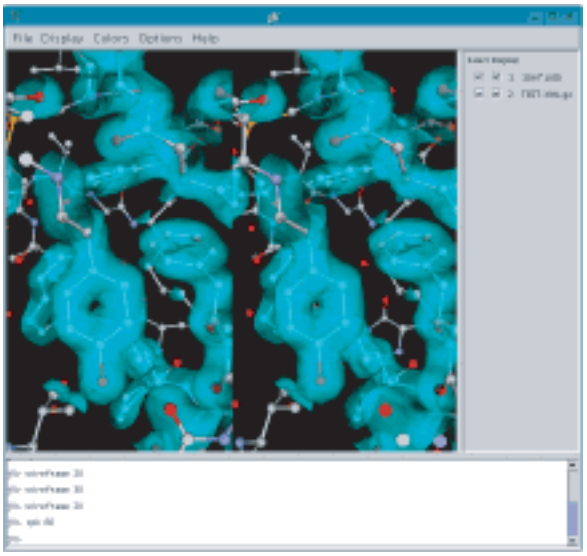
*EDM contours with xPSSS.*

Visualization of electron densities has been available for quite some time at the Uppsala Electron-Density Server (EDS) [http://eds.bmc.uu.se/eds/]. The server has been described by Kleywegt et al. (2004)[1]. Our server, which has been developed by Dr. Arno Paehler in PDBj, uses a similar overall strategy, but differs from the EDS approach in at least two important aspects.

First, while EDS uses the structure factor and model data just to calculate phases with the CCP4 program refmac, we use the same program to refine the deposited structure against the deposited data for 10 cycles.

In some cases the agreement between model and data is different from the one quoted in the deposition. If this difference is too large, EDS does not produce a map. We find that in many such cases refinement reproduces the agreement factors quoted or even leads to improved results. The mean and median of the R-factor distribution is about 1.5% lower for our refined structures than it is for the originally deposited structures.

Secondly, EDS displays only electron density contours, also colloquially known as chicken-wire representation. We offer the same electron density contours, but in addition we also offer the possibility to visualize the electron density as isosurfaces. All necessary calculations are done with Python scripts and with Fortran90 extensions to Python for computationally more expensive tasks.



*EDM isosurfaces with stand-alone jv3.*

Users can select several parameters interactively to influence the display of the electron density. Electron densities are presently available for 11,284 PDB entries, based on the status of the PDB in May 2004. Since then about 2,000 additional structure factor data have become available and we will update our map data base accordingly in the near future. Whenever a map is available for a particular entry, this is indicated on the pages displayed by xPSSS. Electron density contours or isosurfaces are generated on the fly from precalculated electron density maps upon user request and displayed by the interactive PDBj viewer.

Both the interactive viewer applet and the stand-alone viewer are written in Java, using OpenGL. It is therefore necessary to have both Java and OpenGL support for Java installed. The structure factor data used in the refinement and map calculation, the maps themselves and the refined structures can be downloaded. The map format that we use is very similar to the one used by the historic program Frodo and by its successor O, storing the map as byte values. Details of the format of the various files will be available on the PDBj website.

[1]Kleywegt,G.J., Harris,M.R., Zou,J., Taylor,T.C., Wählby,A. and Jones,T.A.
The Uppsala Electron-Density Server
Acta Crystallographica (2004) D60, 2240-2249

# 4. *jV*: PDBj Viewer

## *jV* version 3

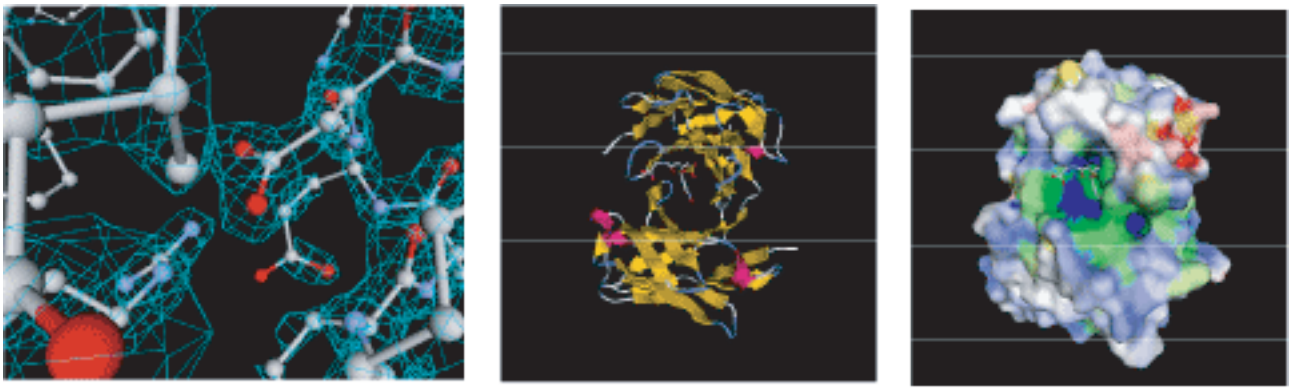
The latest version of PDBj viewer, *jV3*, was released. It is available freely in binary and source code. The latest Java with open GL technology is used to build the viewer, thus it can be used as an applet or stand-alone program on various platforms such as windows 2000/XP, Macintosh OS-X, and Linux. In this version, almost all functions found in Rasmol and Chime have been implemented along with many extensions such as

- multiple molecule handling
- polygon viewing
- animation
- communication capability with network servers via http.

Some example images are shown below and you can find some web sites using jV3 at

- eF-site: http://ef-site.protein.osaka-u.ac.jp/eF-site/

- PreDs: http://pre-s.protein.osaka-u.ac.jp/~preds/
- xPSSS: http://pdbjs3.protein.osaka-u.ac.jp/xPSSS/
- P-cats: http://bioinfo.tsurumi.yokohama-cu.ac.jp/p-cats/
- eProtS: http://eprots.protein.osaka-u.ac.jp/eProtS/

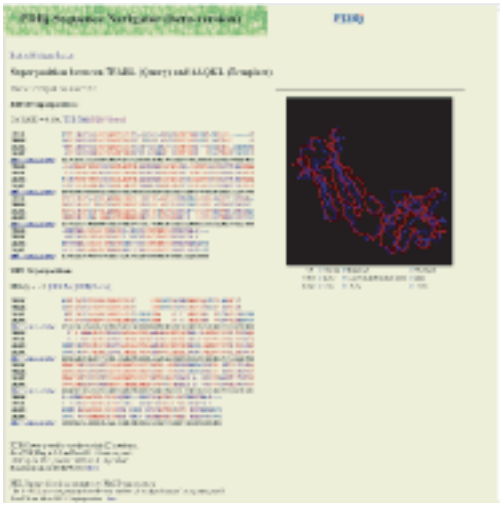


*Various display examples with jv3.*

# 5. Sequence Navigator and Structure Navigator

## Beyond Text-based Queries at PDBj

In December, 2004 PDBj added three new services to its website: Sequence Navigator (http:// www.pdbj.org/ cgi-bin/run_seq_hom.cgi) , Structure Navigator (http:// www.pdbj.org/cgi-bin/run_algn_struc.cgi) z, and GASH (http://www.pdbj.org/cgi-bin/run_gash.cgi). Sequence and Structure Navigator, developed by Daron Standley, are tools for querying the Protein Data Bank (PDB) using sequence homology and structure similarity, respectively. GASH, developed by Daron Standley and Hiroyuki Toh, is a structure alignment program that extends the ASH program developed by Hiroyuki Toh.



***Example of Sequence Navigator.***

With the advent of the World Wide PDB (see PDBj Newsletter vol. 3), PDBj has evolved beyond merely an "archive" for protein structures, to a public resource for querying and analyzing protein structure and function. The PDB currently contains over 28,000 entries, and the number of structures grows by thousands each year.The traditional method of querying this data is to search fields derived from keywords stored in the raw data files. A major effort has been undertaken to expand the set of searchable fields by integrating data from the literature and other databases directly into the portions of the PDB files, with the XML technology. To this end, PDBj utilizes a native-XML database and has added a XPath query engine to the PDBj query browser (see PDBj Newsletter vol. 4), from which very complex queries may be constructed.

In contrast to our sophisticated text-based query tools, we have also created Navigators—tools that don' t require complicated input, and that are based on the experimentally-derived portions of the PDB entry, rather than header information. By using sequence and structure to define similarity relationships between PDB entries, we are not limited to the hard-coded classifications or keywords. Moreover, we open the PDB to exploration by users without expert knowledge.

Starting from any PDB entry (the query), Sequence Navigator generates a list of sequence homologs (templates); clicking on any of the templates reveals a structural superposition of the template structure onto the query, based on a BLAST sequence alignment. The structural similarity is

quantified in terms of Root-Mean Square Deviation (RMSD), as well as a new measure, the Number of Equivalent Residues (NER)[2] . Each template on the list can be used as a new query with a single click.

Structure Navigator works in a similar manner; however, structural similarity is more ambiguous than sequence similarity, and the corresponding structural alignments more CPU intensive to carry out. For this reason, only a representative set of structure-structure alignments are computed using the ASH program, and stored on our server. When a particular query is selected, Structure Navigator first finds the best match on the representative list, then uses sequence alignment to map the stored results onto the actual query. Due to the number of entries to be considered, this process can take up to a minute.

For some cases, the alignment is so complicated, that accurate results can not be obtained using ASH, or even using popular programs, such as Dali or CE. For this reason, we have developed a new program, GASH, that extends the ASH algorithm by considering many possible alignments, instead of just one. The GASH program can take up to a minute to compute a single structural alignment of two structures, but the alignments are very accurate, even when complicated structures are compared.

[2] Standley DM, Toh H, Nakamura H.
Detecting local structural similarity in proteins by maximizing number of equivalent residues.
Proteins 2004;57(2):381-391

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

**Group for PDB Database Curation**
Kusunoki, Masami, Dr. (IPR, Osaka Univ.)
Kosada, Takashi (IPR, Osaka Univ.)

**Group for Development of PDBj database**
Ito, Nobutoshi, Prof. (Tokyo Med. Dent. Univ.)
Paehler, Arno, Dr. (JST-BIRD)
Standley, Daron M., Dr. (JST-BIRD)

**Group for Development of Secondary Database and Analysis Systems**
Ohkawa, Takenao, Dr. (ISE, Osaka Univ.)
Kinoshita, Kengo, Dr. (IMSUT, Tokyo Univ. )
Toh, Hiroyuki, Prof. (ICR, Kyoto Univ.)
Wako, Hiroshi, Prof.(Waseda Univ.)

**Group for NMR Database**
Akutsu, Hideo, Prof. (IPR, Osaka Univ.)
Matsuki, Yoh, Dr. (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