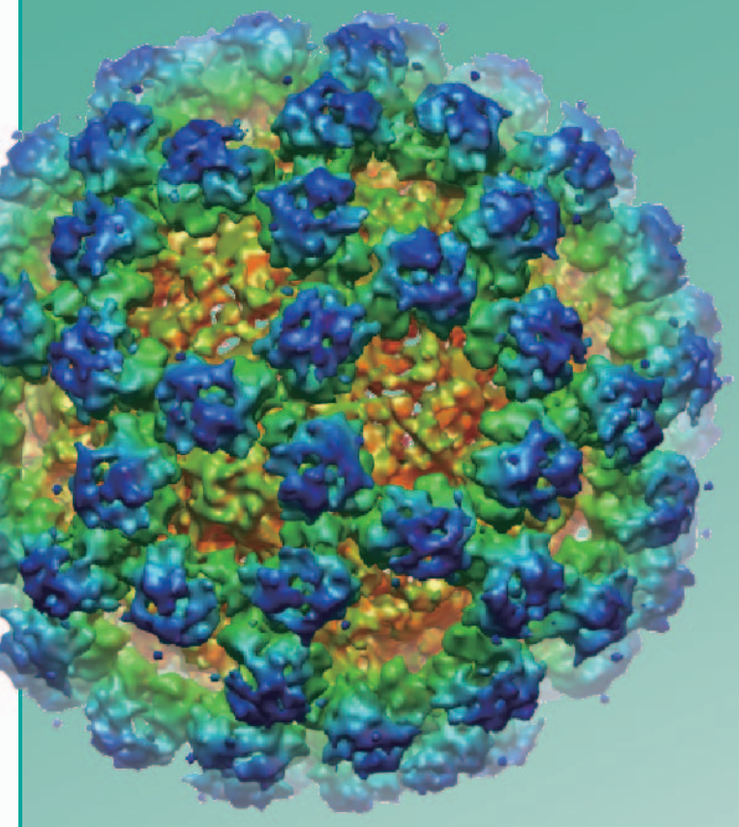
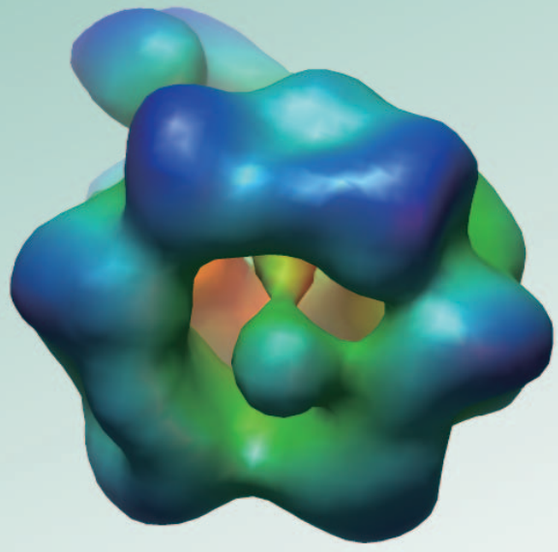
PDBj

Protein Data Bank Japan

http://pdbj.org/





Newsletter vol.15

June, 2013

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

More than 20,000 entries have been processed at PDBj

The PDBj (Protein Data Bank Japan, http://pdbj.org/) is a member of the wwPDB (worldwide PDB, http://wwpdb.org/), and has been curating the structure data of biological macromolecules, such as proteins, nucleic acids and carbohydrates, which are deposited by researchers in the structural biology field, mainly from Asian and Oceania regions. The PDBj has been processing and editing such atomic-resolution structures with high qualities, and collaborating the other members of the wwPDB to construct a single and public archive, PDB, all the data in which are released through internet.

Since the PDBj started the first data processing in 2000, 13 years have passed, and more than 20,000 entries were processed by the PDBj at the end of 2012. The number of these processed entries is much larger than those available in the world in 2000. The total available data in PDB became larger than 90,000 in May 2013. Now, the wwPDB collects varieties of structural biology data, covering the NMR chemical shifts and molecular images taken by the electron microscope (EM), collaborating with the BMRB and EMDB (http://emdatabank.org/), respectively. The PDBj has so far been editing and processing the deposited NMR chemical shift data, as the PDBj-BMRB group (http://bmrb.protein.osaka-u.ac.jp/) collaborating with the BMRB in USA.

The databases in the life science field in Japan have been integrated by the JST-NBDC (Japan Science and Technology Agency-National Bioscience Database Center), as the governmental program. The PDBj have succeeded in being granted by the NBDC for the Integration Program with the title "Global Construction and Integration of PDB" applying the Semantic web technology for three years from April 2011 to March 2014. First, the PDBj constructed the ontology in the Web Ontology Language (OWL), based on the PDBx/mmCIF dictionary. At the weekly update of the original PDBML data, which are described in the canonical XML format, the PDB data in the Resource Description Framework (RDF) format are automatic generated. The RDF data in the individual category of each entry can be identified with the unique URL, and they are provided as the wwPDB canonical data from the web site (PDB/RDF, http://rdf.wwpdb.org/). In the same way, the library data for small molecules in PDB are also available with the RDF format. The UniProt (http://uniprot.org/) has now added links from their own data to this PDB/RDF. People, who are interested in the semantic web technology, are welcome to our PDB/RDF page. In addition, the PDBj makes our own effort to construct a new portal site, applying the recent web technology, as mentioned below.

For the more secure and stable management of our database, the PDBj provides an additional portal site other than Kansai area, collaborating with the NBDC, so as to prevent any inconvenience for users. However the frequent updates of the web browsers sometimes make it difficult for us to follow all the versions of many kinds of web browsers.

The PDBj would welcome any comments from the users through our web page (http://pdbj.org/pdbj_contact.html). Any comments are also welcome at the workshops, seminars or booth presentations at the scientific meetings.

Announcing the new PDBj web interface

Fig.1: New web Top Page

The new PDBj web interface provides users with a modern interface to the PDB data. The web interface is not just to offer an improvement of the visual experience, but this release is the beginning for more features, which could allow us to add protein functions, annotations and associated data. The major strength of the new interface is its dynamic properties to integrate these various data. In combination with its customizability, it allows users to select what information they would like to see and where.

Various services are also integrated into the new web interface more tightly than our previous portal, “legacy portal”, which we offered so far. In the new web interface, “PDBj Mine” as the general RDB query system, “Sequence Navigator” to find homolog proteins, “Structure Navigator” to search similar structures, “ProMode Elastic” as the normal mode analysis by the Elastic Network model, “Molecule of the Month” for introduction of proteins every month, and the “Help page”, have all been integrated and inter-linked into a

single web interface. From the various services, in case there is an association to another service, a link to that resource is provided.

One of the most useful features in the new web interface is the multifunctional “Search bar” available in the header of the interface. By default the search bar searches through PDBj Mine, Information pages (such as News, Help pages and Molecule of the Month) and Status search. In addition, this “Search bar” can also be used to send a search query to EMDB and BMRB to search their respective databases. Using the “Search bar” Advanced searches can also be performed directly using a set of commands to specify which fields to be searched in and to omit results containing specific keywords. Similar to the legacy portal, the new web interface also offers users an Advanced search option to specify their parameters using a graphical interface.

The new interface also offers multi-language support for English, Japanese, Chinese (simplified and traditional) and Korean, as well as the legacy portal. For each language, the basic interface such as the menus and the introduction page for deposition have been translated, but services and information pages are generally only available in English and Japanese, as well as in the legacy portal. The new interface does offer increased accessibility by allowing users to switch language on each page, which can be useful enough even in the case where a user reached a page from a search engine.

From a programmer's point of view, because the new interface works completely by using a REST design (REpresentational State Transfer), all data provided by the new web interface is also accessible via any programming language, which can request data via URLs and can parse JSON and XML. Via these REST services, users can integrate PDBj data and services into their own software.

The release of the new PDBj web interface marks a new milestone for PDBj. The customizability of the interface will personalize the experience for each individual user to optimize their preferred view of the information offered by PDBj.

It is known that the new web interface does not work well on some browsers of their old versions (in particular for older versions of IE, Internet Explore, than IE8), and so the legacy portal is continuously available from the PDBj: http://legacy.pdbj.org/.

Any questions, suggestions, and comments to the new PDBj web interface are welcome from our web site: http://pdbj.org/pdbj_contact.html.

Introduction to the PDBx/mmCIF format

The “PDB format” means that fixed-column length, line-based flat file format to most users of the PDB. But the PDB format is more than 40 years old and it is unfortunately failing to meet the requirements today's structural biology imposes. For example, the number of atom that can be contained in a single entry is less than 100,000 so that some huge supramolecular structures are currently divided into several PDB entries. For example, the structural data of Vault are described in the three separated PDB files (2zuo, 2zv4, and 2zv5). Also, today's structural biology cannot be done without referencing the results of other fields than structural biology so that functional, taxonomic, and other annotations are becoming as valuable as atomic coordinates themselves. But in the traditional PDB files it is sometimes difficult to parse such information without excessive “exception handling.”

In order to address these issues, it was decided in 2011 wwPDB advisory committee that a new PDB format should be devised, and that new format turned out to be the PDBx/mmCIF format which is currently adopted as the official archiving format of the wwPDB.

So, what's good about PDBx/mmCIF? First of all, the grammar of PDBx/mmCIF is rigorously (formally) defined so that no “exception handling” is necessary as far as the syntax is concerned. Second, the categories and terms in the format are precisely defined in the PDBx dictionary. By looking up the dictionary, we can learn which category describes what kind of data, as well as the relationships between different categories. The syntax of the PDBx dictionary itself is formally defined so that we can parse the dictionary to generate a program than manipulates PDBx/mmCIF files, for example. At present, there are over 300 categories defined in the dictionary, among them are

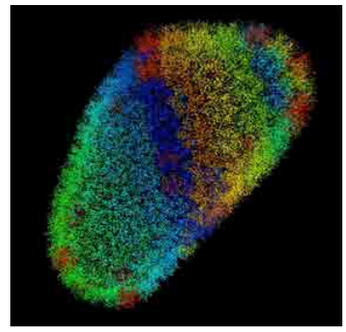


Fig.2: The structure of HIV capsid (PDB ID : 3J34, 3J3Q, 3J3Y) released on 29 May contains more than 2 million atoms and 1000 peptide chains, and is divided into 25 PDB entries (1VU4 etc.) .

references to external data resources, which facilitates integration of structural data with other biological and/or chemical annotations. Third, compared to PDBML (a “direct” translation of mmCIF into XML), PDBx/mmCIF files are (arguably) easy to read both by humans and by machines: tags describing category and category items are simple enough to not interfere with human eyes.

Given the above benefits, what are obstacles for adopting the PDBx/mmCIF format? Apparently, many programs do not support the format yet. However, some widely used applications already support the PDBx/mmCIF format (e.g., Jmol, OpenRasMol, Chimera, CCP4, Phenix). Programmers may feel reluctant to modify their codes. But there are PDBx/mmCIF parsers already available in some major programming languages such as C/C++, Python, Java, and Perl. By using such libraries, the amount of codes to be modified may be greatly reduced.

A comprehensive list of PDBx/mmCIF-related information can be found at;
http://mmcif.pdb.org.

EVENT

■ Science Agora 2012 (Nov. 10-11, Odaiba, Tokyo)

Science Agora is an annual scientific event for the public supported by Japan Science and Technology Agency. We, PDBj participated in the event for the first time and talked about what is the use of protein structural analysis, using 3D viewer, “Yorodumi Prime”, which is developed by PDBj.

Yorodumi Prime URL: http://pdbj.org/prime

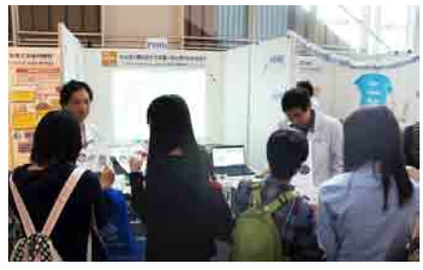


Fig.3: Science Agora 2012

■ PDBj Exhibition Booth at the 35th Annual Meeting of the Molecular Biology of Japan

(Dec. 11-14, Marinemesse Fukuoka, Fukuoka-shi)

Fig.4: PDBj Exhibition Booth at the Bio Database corner

PDBj set up an exhibition booth to introduce our web services and recent activities at the Bio Database corner organized by JST-NBDC.

All Seminar/Workshop Materials Download Available

■ PDBj Workshop at Tokyo University of Agriculture and Technology

(Mar. 7, Koganei-shi, Tokyo)

PDBj holds workshops in different areas of Japan periodically, for the purpose of education and training of database users, and took place at Koganei Campus, Tokyo University of Agriculture and Technology this year.

Fig.5: PDBj Workshop

Data Growth

The statistics data is also available at the wwPDB web page (http://wwpdb.org/stats.html).

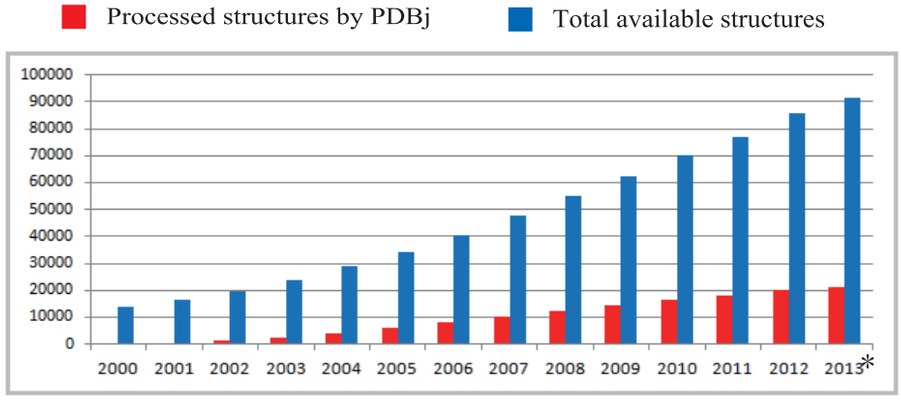


Fig.6 *As of June 5, 2013

Database of experimental information derived by NMR: BMRB

PDBj-BMRB has been collaborating with the BMRB (Wisconsin, USA) and the wwPDB to develop and implement deposition and annotation management strategies for large-scale NMR databases. Our current project aims to enhance data sharing and access among the Japanese life science databases through database integration. The BMRB mirror server, which provides services originating from the University of Wisconsin has been re-designed. This server incorporates an easy-to-understand interface with integrated data visualization tools for the submission of NMR data. The new BMRB website provides an interface that enables advanced SQL queries. A new PDBj-BMRB portal site reflects the feedback from researchers who have made database submissions, and is now online. We have produced not only updated data submission guides (in Japanese and English) but also support tools; MagRO which analyzes NMR data and BMRBxTool which converts NMR-STAR format into XML format, originally developed at PDBj-BMRB. Combined with the new visualization and support tools, the data submission procedure is greatly enhanced by making it convenient and user friendly.

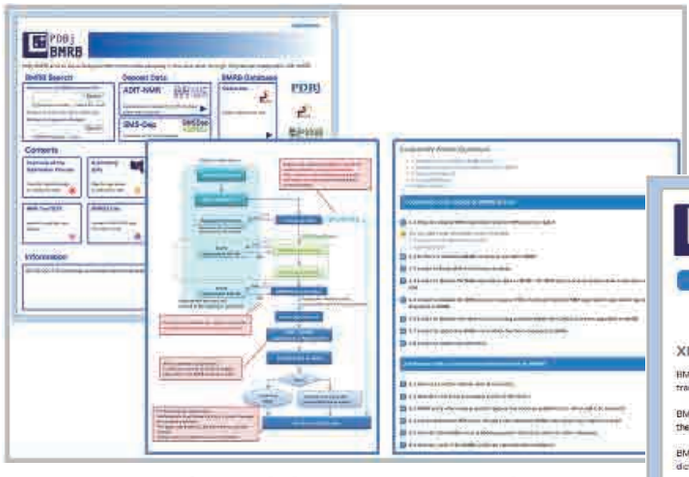


Fig.7: Contents of upgraded PDBj-BMRB portal site, Overview of the submission process and Frequently Asked Questions

Fig.8: BMRBxTool, XML converter for NMR-STAR

STAFF

Head

Nakamura, Haruki, Ph. D. (Prof., IPR, Osaka Univ.)

Group for PDB Database Curation

Nakagawa, Atsushi, Ph. D. (Group Leader, Prof., IPR, Osaka Univ.)
Matsuda, Makoto, Ph. D. (IPR, Osaka Univ.)
Igarashi, Reiko (IPR, Osaka Univ.)
Kengaku, Yumiko (IPR, Osaka Univ.)
Cho, Hasumi, Ph. D. (IPR, Osaka Univ.)
Ikegawa, Yasuyo (IPR, Osaka Univ.)
Sato, Junko (IPR, Osaka Univ.)

Group for Development of new tools and services

Kinjo, Akira R., Ph. D. (IPR, Osaka Univ.)
Iwasaki, Kenji, Ph. D. (IPR, Osaka Univ.)
Suzuki, Hirofumi, Ph. D. (IPR, Osaka Univ.)
Yamashita, Reiko (IPR, Osaka Univ.)
Kudou, Takahiro (IPR, Osaka Univ.)
Nishikawa, Ken, Ph. D. (Guest Prof., IPR, Osaka Univ.)
Bekker, Gert-Jan (IPR, Osaka Univ.)

Group for BMRB

Fujiwara, Toshimichi, Ph. D. (Group Leader, Prof. Osaka Univ.)
Akutsu, Hideo, Ph. D. (Prof., IPR, Osaka Univ.)
Kojima, Chojiro, Ph. D. (IPR, Osaka Univ.)
Kobayashi, Naohiro, Ph. D. (IPR, Osaka Univ.)
Iwata, Takeshi (IPR, Osaka Univ.)
Takahashi, Ami (IPR, Osaka Univ.)
Yokochi, Masashi (IPR, Osaka Univ.)

Collaboratory Researchers

Wako, Hiroshi, Ph. D. (Prof., Waseda Univ.) (for Pro Mode)
Ito, Nobutoshi, Ph. D. (Prof., Tokyo Medical and Dental Univ.)
Kinoshita, Kengo, Ph.D. (Prof., Tohoku Univ.) (for e F-site)
Standley, Daron, Ph. D. (IFReC, Osaka Univ.)
(for SeqNavi, StructNavi, SeSAW, and ASH)
Katoh, Kazutaka, Ph. D. (IFReC, Osaka Univ.) (for MAFFTash)

Secretary

Haruki, Nahoko (IPR, Osaka Univ.)

Contacting Protein Data Bank Japan (URL : http://pdbj.org/)

Research Center for State-of-the-Art, Functional Protein Analysis
Institute for Protein Research, Osaka University
3-2 Yamadaoka, Suita, Osaka 5650871, JAPAN
TEL(PDBj Office): +81-6-6879-4311
TEL(PDBj Deposition Office): +81-6-6879-8634