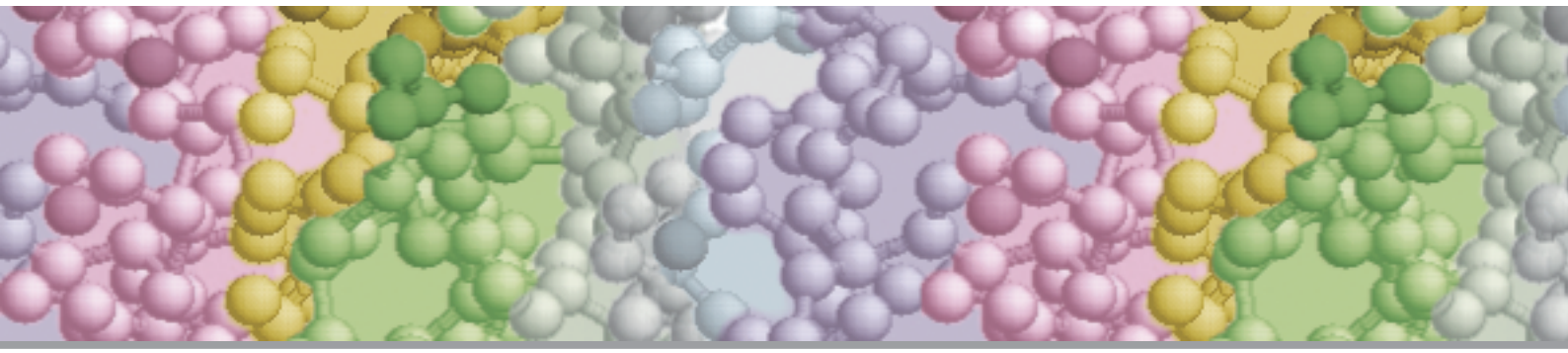
PDBj
Protein Data Bank Japan

http://pdbj.org/

# News Letter Vol.13

November 2011



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

News

## Launching New PDBj Activities

We, PDBj, have engaged in construction and enhancement of PDB database with the international collaboration since 2001, supported by JST-BIRD (Japan Science and Technology Agency, Institute for Bioinformatics Research and Development). In April, 2011, NBDC (National Bioscience Database Center http://biosciencedbc.jp) was founded in JST, so as to integrate almost all the life science databases in Japan. Many different kinds of databases, which have so far been developed by various Japanese Ministries, are now open from the common portal at NBDC. In addition, very many experimental results given by researches with the governmental supports are now provided as the common style databases, and NBDC itself works as the funding agency to support groups who develop and maintain their databases in various fields of life science. We, PDBj, have succeeded in being granted by NBDC for the Integration Program with the title "Global Construction and Integration of PDB", and we are relaunching the new activities for three years until March 2014.

In this program, PDBj will continuously construct the PDB database as the Data-in activity, and will proceed the database integration using RDF (Research Description Framework), which NBDC is going to promote. In fact, we are going to arrange the new semantic web environment for the PDB data, which have never been performed by any other groups in the world, in order to integrate PDB with other biological and medical databases. When this semantic web system is matured, a machine automatically understands the definitions and the contents of any data. Then, the necessary and essential data are expected to be readily queried and selected from various databases distributed in the cyberspace, leading to analyze and infer the biological and medical functions. The PDB/RDF is described in details in the following section “Linked Open Data". In the near future, the semantic web system with the PDB/RDF, is expected to be adopted broadly from the wwPDB (worldwide PDB).

## wwPDB Advisory Committee meeting

The wwPDB, of which the PDBj is one of the members, organizes the annual advisory Committee (wwPDBAC) meeting. This year, EBI in Hinxton, UK was a host, and the meeting was held on September 30, 2011. The participants were Dr. Stephen Burley as a chair (Eli Lilly and Company), Dr. Janet Thornton (Director of EBI), Prof. Helen M. Berman and Dr. Martha Quesada (RCSB-PDB, Rutgers Univ.), Dr. Gerard Kleywegt (PDBe, EBI), Prof. John L. Markley (BMRB, Univ. Wisconsin-Madison), Prof. Haruki Nakamura (PDBj, Osaka Univ.), Prof. Michael Rossmann (Purdue Univ.), Prof. Wayne Hendrickson (Columbia Univ.), Prof. Randy J. Read (Cambridge Univ.), Prof. Helen Saibil (Birkbeck College London), Dr. David Neuhaus (MRC), Prof. Soichi Wakatsuki (KEK, Photon Factory), Prof. Genji Kurisu (IPR, Osaka Univ.), Prof. Wah Chiu (Baylor College of Medicine), Prof. Edward Baker (Auckland Univ.) as a representative of IUCr, and Dr. R. Andrew Byrd (Center for Cancer Research, NCI) as a representative of ICMRBS. In addition, Prof. Manju Bansal (Indian Institute of Science, India) and Prof. Jianping Ding (Shanghai Institutes for Biological Sciences Chinese Academy of Sciences, China) also attended the meeting as the associated member.

First, the development of Common Deposition and Annotation (D & A) program was introduced by Martha Quesda, and the beta-version will be available in the third quarter, 2012. Next, every wwPDBAC member agreed that the conventional PDB formal, which has been used for the last 40 years, should now be updated, because of the limitation of the atom numbers, chain numbers

Fig.1: The participants of wwPDB AC Meeting

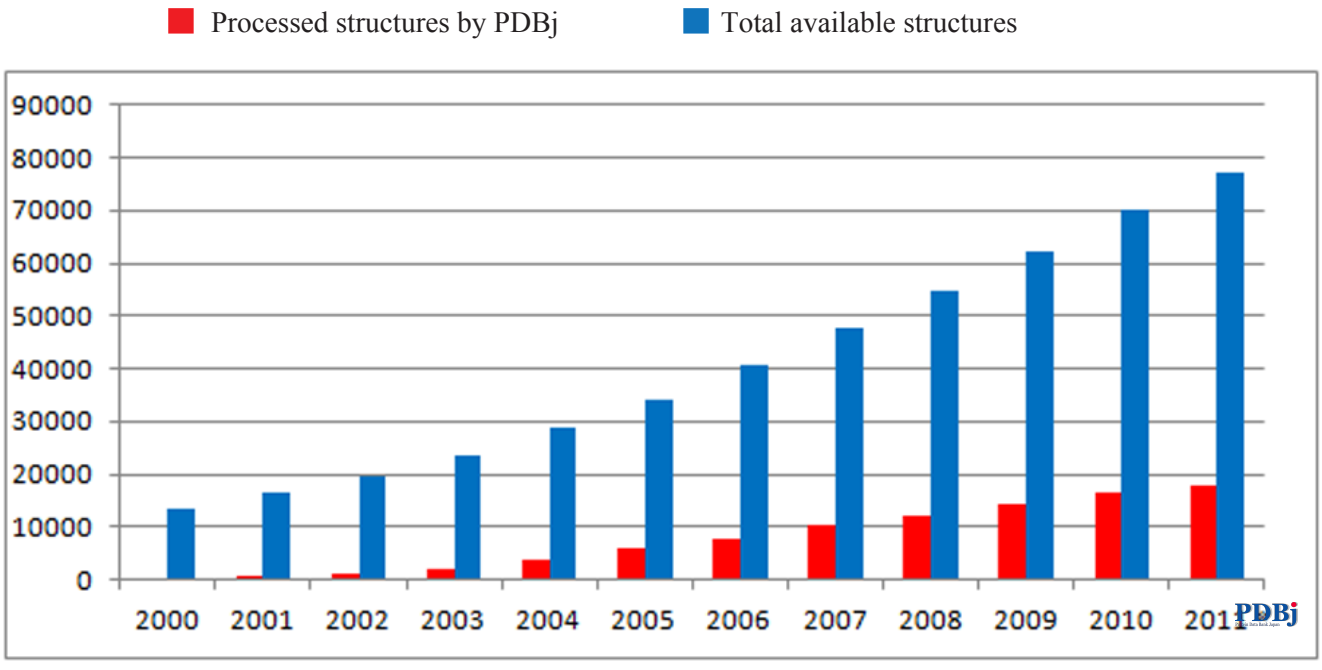
and more and more complex description of the meta-data. The recent task force members, who have been organized to invent a new format, concluded a proposal that the current PDBx (mmCIF) format is used as the basic format, and the wider description than 80 characters should be used for each ATOM record. This proposal will be further discussed. Then, the reports were shown by Validation Task Forces (VTFs) for X-ray crystallography, NMR, EM, and small-angle X-ray and Neutron scattering observations (SAXS/SANS), in order to keep the data qualities from those experiments. PDBj introduced their own relaunching activities for three years since April 2011, with the financial support from the new funding agency NBDC. Development of the RDF/PDB and a new tool to validate the data description with the schema and the dictionary were mentioned. Finally, the next archive remediation and improvement of the wwPDB corporate image are planned as the future wwPDB activities.

By the way, on September 29, one day before the wwPDBAC meeting, a contract was made between the wwPDB members and Dr. Colin R. Groom, who is a director of CCDC (Cambridge Crystallographic Data Center), about new collaboration between wwPDB and CCDC, which has long experiences for the crystal structures of small chemical compounds and has many software tools to edit them. There are many such small chemical compounds as the bound ligands to biological macromolecules, and for annotation of those chemical compounds, the wwPDB primary annotators are able to browse and query the entire CSD (Cambridge Structural Database) data given by CCDC, thus the quality of ligand data in wwPDB is expected to become much higher than before. In addition, the atom coordinates of the common chemical compounds in the both CSD and wwPDB are allowed to be made open from wwPDB.

Fig. 2: G. Kleywegt, J. L. Marklye, C. R. Groom, Helen Berman, and H. Nakamura from Left to right, at EBI on September 29, 2011.

## Data Growth

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

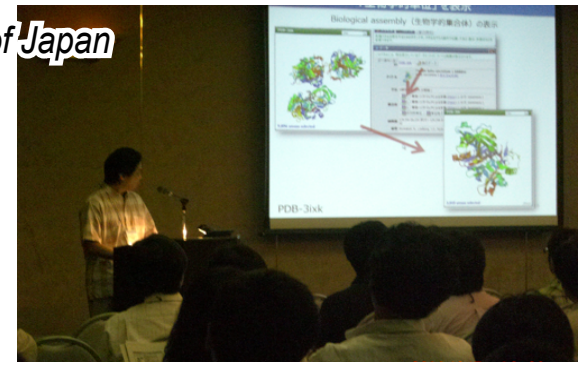



*As of Novemver 9

# Event

## *The Annual Meeting of the Protein Science Society of Japan*

The Annual Meeting of the Protein Science Society of Japan was held from June 7th to 9th, 2011 at Hotel Hankyu Expo Park, Osaka. PDBj held a luncheon seminar and workshop, which introduced our web-services and gave practical training to the users.

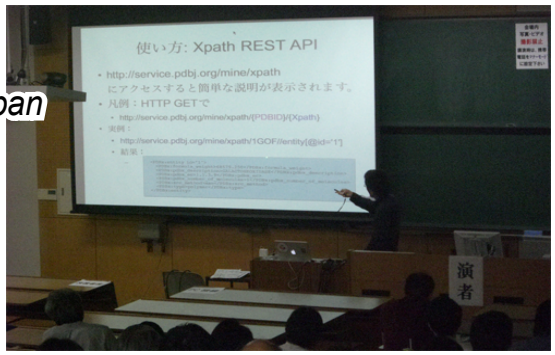


## *IUCr2011: XXII Congress and General Assembly, International Union of Crystallography*

IUCr2011 was held from August 22 to 30, at Palacio Municipal de Congresos de Madrid, Spain. PDBj set up a joint booth as a member of wwPDB, and introduced our activities and services.

## *The 49th Annual Meeting of the Biophysical Society of Japan*

The 49th Annual Meeting of the Biophysical Society of Japan was held from September 16 to 18 at Himeji Shosha Campus, University of Hyogo. We introduced our collaborative projects with NBDC and wwPDB at the luncheon seminar.

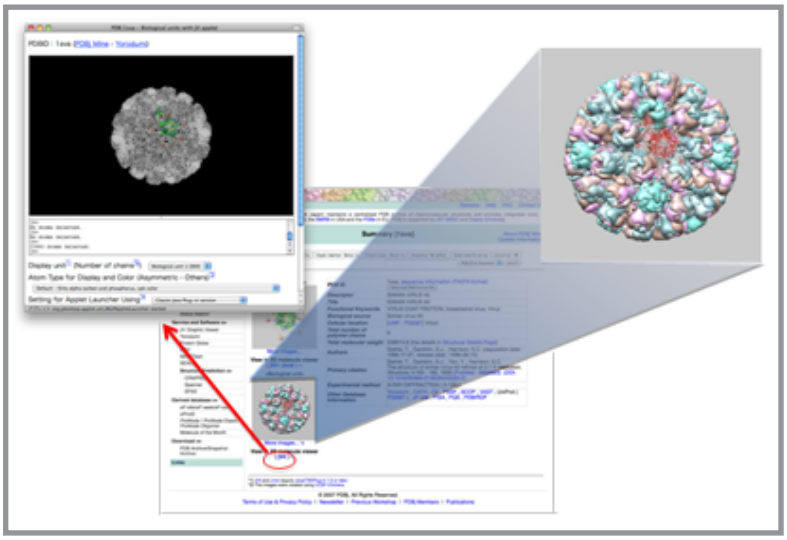



# Service

## Display of biological unit

Biologically active forms of proteins are often composed of oligomeric homo peptide chains such as homo-dimers or homo-teratrimers, which have some symmetric conformations. In addition, as in the case of Virus capsid shown in Fig. 4, units of hetero oligomers form further symmetric complex structures. Those symmetric structures are called as "Biological units", discriminating from the symmetric units, which are only formed in crystals. The Biological units are identified by PISA program and/or the authors in the corresponding PDB meta-data. However, their atomic coordinates are not directly described in the PDB origianl data, but the parameters are described in order to generate the coordinates by symmetry operations.

PDBj offers an interface to display the molecular structures of those Biological units, and to download the corresponding atomic coordinates. As shown in Fig. 4, the static image of the Virus coat proteins is constructed by a program UCSF Chimera, not showing each atom but showing the low-resolution image.



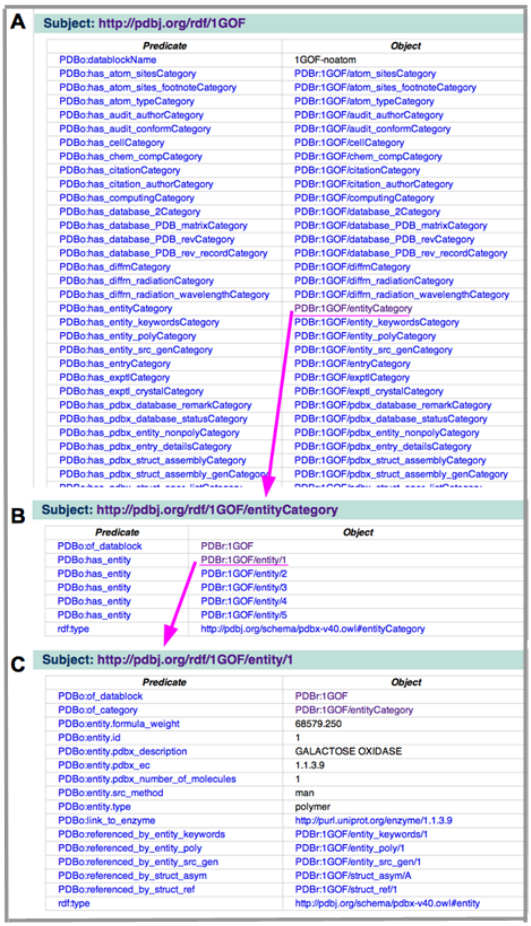
In addition, interactive display of the Biological units is also available with the molecular viewer, jV, which has been developed by PDBj. Here, the molecules in an asymmetric unit, the atomic coordinates in which are described in PDB data, are displayed as the usual atomic images, and the other parts are displayed with the solid lines linking only the Cα atoms of residues. Thus, even a huge molecular complex as a Biological unit can be smoothly manipulated. Those pictures are accessed from the Summary page of each entry in our basic query service, PDBj Mine.

Fig. 4: Example of display of the biological unit of the Virus coat protein in PDBj Mine. Top right is a static image by UCSF Chimera, and the top left is an interactive display with jV graphic applet on the web.

## PDB is ready for Linked Open Data

PDBj has started providing PDB data in the Resource Description Framework (RDF) format [1], which are accessible via URLs such as http://pdbj.org/rdf/1GOF for the PDB entry 1GOF. The RDF format is a W3C standard for the Semantic Web (see http://www.w3.org/TR/rdf-primer/ for a tutorial) which aims at connecting diverse data resources on the Web. The PDBx ontology in the Web Ontology Language (OWL) format is also available. The PDBx OWL ontology is based on the PDBML schema, which in turn is a direct translation of the mmCIF PDB exchange (PDBx) dictionary. Thus, most data types and relations between data are inherited from the PDBx dictionary. There are additional properties that are used for linking PDB data items to other databases such as UniProt. Based on the Web standard, PDB/RDF can be readily integrated with other data resources available in the RDF format.

One notable thing about PDB/RDF is that URLs are provided not only for each PDB entry, but also for each category element of PDB entries. That is, the URL http://pdbj.org/rdf/1GOF/entity/1 refers to the "entity" category element with its primary key _entity. id '1'. As such, PDB/RDF enables more fine-tuned linking to PDB data.



A Subject: http://pdbj.org/rdf/1GOF

| Predicate | Object |
|---|---|
| PDBo:datablockName | 1GOF-noatom |
| PDBo:has_atom_sitesCategory | PDBr:1GOF/atom_sitesCategory |
| PDBo:has_atom_sites_footnoteCategory | PDBr:1GOF/atom_sites_footnoteCategory |
| PDBo:has_atom_typeCategory | PDBr:1GOF/atom_typeCategory |
| PDBo:has_audit_authorCategory | PDBr:1GOF/audit_authorCategory |
| PDBo:has_audit_conformCategory | PDBr:1GOF/audit_conformCategory |
| PDBo:has_cellCategory | PDBr:1GOF/cellCategory |
| PDBo:has_chem_compCategory | PDBr:1GOF/chem_compCategory |
| PDBo:has_citationCategory | PDBr:1GOF/citationCategory |
| PDBo:has_citation_authorCategory | PDBr:1GOF/citation_authorCategory |
| PDBo:has_computingCategory | PDBr:1GOF/computingCategory |
| PDBo:has_database_2Category | PDBr:1GOF/database_2Category |
| PDBo:has_database_PDB_matrixCategory | PDBr:1GOF/database_PDB_matrixCategory |
| PDBo:has_database_PDB_revCategory | PDBr:1GOF/database_PDB_revCategory |
| PDBo:has_database_PDB_rev_recordCategory | PDBr:1GOF/database_PDB_rev_recordCategory |
| PDBo:has_diffrnCategory | PDBr:1GOF/diffrnCategory |
| PDBo:has_diffrn_radiationCategory | PDBr:1GOF/diffrn_radiationCategory |
| PDBo:has_diffrn_radiation_wavelengthCategory | PDBr:1GOF/diffrn_radiation_wavelengthCategory |
| PDBo:has_entityCategory | PDBr:1GOF/entityCategory |
| PDBo:has_entity_keywordsCategory | PDBr:1GOF/entity_keywordsCategory |
| PDBo:has_entity_polyCategory | PDBr:1GOF/entity_polyCategory |
| PDBo:has_entity_src_genCategory | PDBr:1GOF/entity_src_genCategory |
| PDBo:has_entryCategory | PDBr:1GOF/entryCategory |
| PDBo:has_exptlCategory | PDBr:1GOF/exptlCategory |
| PDBo:has_exptl_crystalCategory | PDBr:1GOF/exptl_crystalCategory |
| PDBo:has_pdbx_database_remarkCategory | PDBr:1GOF/pdbx_database_remarkCategory |
| PDBo:has_pdbx_database_statusCategory | PDBr:1GOF/pdbx_database_statusCategory |
| PDBo:has_pdbx_entity_nonpolyCategory | PDBr:1GOF/pdbx_entity_nonpolyCategory |
| PDBo:has_pdbx_entry_detailsCategory | PDBr:1GOF/pdbx_entry_detailsCategory |
| PDBo:has_pdbx_struct_assemblyCategory | PDBr:1GOF/pdbx_struct_assemblyCategory |
| PDBo:has_pdbx_struct_assembly_genCategory | PDBr:1GOF/pdbx_struct_assembly_genCategory |

B Subject: http://pdbj.org/rdf/1GOF/entityCategory

| Predicate | Object |
|---|---|
| PDBo:of_datablock | PDBr:1GOF |
| PDBo:has_entity | PDBr:1GOF/entity/1 |
| PDBo:has_entity | PDBr:1GOF/entity/2 |
| PDBo:has_entity | PDBr:1GOF/entity/3 |
| PDBo:has_entity | PDBr:1GOF/entity/4 |
| PDBo:has_entity | PDBr:1GOF/entity/5 |
| rdf:type | http://pdbj.org/schema/pdbx-v40.owl#entityCategory |

C Subject: http://pdbj.org/rdf/1GOF/entity/1

| Predicate | Object |
|---|---|
| PDBo:of_datablock | PDBr:1GOF |
| PDBo:of_category | PDBr:1GOF/entityCategory |
| PDBo:entity.formula_weight | 68579.250 |
| PDBo:entity.id | 1 |
| PDBo:entity.pdbx_description | GALACTOSE OXIDASE |
| PDBo:entity.pdbx_ec | 1.1.3.9 |
| PDBo:entity.pdbx_number_of_molecules | 1 |
| PDBo:entity.src_method | man |
| PDBo:entity.type | polymer |
| PDBo:link_to_enzyme | http://purl.uniprot.org/enzyme/1.1.3.9 |
| PDBo:referenced_by_entity_keywords | PDBr:1GOF/entity_keywords/1 |
| PDBo:referenced_by_entity_poly | PDBr:1GOF/entity_poly/1 |
| PDBo:referenced_by_entity_src_gen | PDBr:1GOF/entity_src_gen/1 |
| PDBo:referenced_by_struct_asym | PDBr:1GOF/struct_asym/A |
| PDBo:referenced_by_struct_ref | PDBr:1GOF/struct_ref/1 |
| rdf:type | http://pdbj.org/schema/pdbx-v40.owl#entity |

Fig.5: Example of PDB/RDF

[1] Kinjo, A. R. et al. (2012) Nucleic Acids Research (in press) http://dx.doi.org/10.1093/nar/gkr811

# 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 (IPR, Osaka Univ.)
Igarashi, Reiko (IPR, Osaka Univ.)
Kengaku, Yumiko (IPR, Osaka Univ.)
Cho, Hasumi (IPR, Osaka Univ.)
Ikegawa, Yasuyo (IPR, Osaka Univ.)
Takano, Tomoko (IPR, Osaka Univ.)

**Group for Development of new tools and services**
Kinjo, Akira, 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.)

**Group for BMRB**
Fujiwara, Toshimichi, Ph. D. (Group Leader, Prof. Osaka Univ.)
Akutsu, Hideo, Ph. D. (Prof., IPR, Osaka Univ.)
Kojima, Chojiro (IPR, Osaka Univ.)
Kobayashi, Naohiro (IPR, Osaka Univ.)
Iwata, Takeshi (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 ASH)**

**Secretary** Haruki, Nahoko (IPR, Osaka Univ.)

# Contacting

**Protein Data Bank japan**

Research Center for Structural and Functional Proteomics,
Institute for Protein Research (IPR), Osaka University
3-2 Yamadaoka, Suita, Osaka 5650871, JAPAN

TEL (PDBj office) : +81-6-6879-4311
TEL (PDBj Deposition office) : +81-6-6879-8634
FAX (PDBj Deposition office) : +81-6-6879-8636

URL: http://pdbj.org/