PDBj

Protein Data Bank Japan

News Letter Vol. 7

March 2006

http://www.pdbj.org/

News

The Second wwPDBAC meeting

On August 30, 2005, the Second wwPDB Advisory Committee(wwPDBAC) meeting was held at Firenze, Italy, 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), Dr. Kei Yura (PDBj, JAEA), Prof. Soichi Wakatsuki (PDBj, Photon Factory), (2) Community Stakeholder Representatives: Prof. Edward N. Baker (IUCr, University of Auckland), Dr. R. Andrew Byrd (ICMRBS, National Cancer Institute), Dr. Jose-Maria Carazo (Macromolecular EM, CNB), (3) PDB Site Leaders: Prof. Helen M. Berman (RCSB, Rutgers University), Dr. Kim Henrick (MSD, EBI), Prof. Haruki Nakamura (PDBj, Osaka University), and (4) Funding Agency Representatives: Mr. Toshiyuki Koike (JST).

The meeting started with an overview of PDB and wwPDB by Prof. Helen Berman, and the three members of wwPDB (RCSB, MSD-EBI, and PDBj) made presentations about their current activities and grant situations. Answering the previous comments from the wwPDBAC members, made in the First wwPDBAC meeting held in November 21, 2004, the individual wwPDB members reported the results of remediation of PDB data: Ligands for RCSB, Sequences, taxonomy and entities for MSD-EBI, and Citations for PDBj. These remediation tasks will be finished by the end of 2006. In addition, several other issues were discussed: Tight relationship with BMRB (Biomolecular NMR database) and the Macromolecular EM, Database for theoretical models, Impact of increasing depositions, and wwPDBAC Advocacy with Funding Agencies.

The third wwPDBAC meeting will be held in October, 2006, near Tokyo, Japan, accompanying the satellite workshops of the ICSG 2006 (International Conference on Structural Genomics).

The 4th planning exhibition, the space-time mystery - on the occasion of the 100th Anniversary of Einstein’ s three major papers

The 4th planning exhibition, the space-time mystery - on the occasion of the 100th Anniversary of Einstein’ s three major papers- was held September 18th – 25th 2005, at Osaka University, Nakanoshima Center. This exhibition was for non- specialists including high-school students. We exhibited at the booth, where we introduced our activities. On 25th, Prof. Nakamura had a mini-lecture, which introduced eProtS database by using PCs. We got a lot of requests for our databases from the participants.

Prof. Nakamura at the booth.

The 2nd DDBJing & PDBjing – the 14th DDBJing in Osaka-

The 2nd DDBJing & PDBjing Workshop – the 14th DDBJing in Osaka- was held February 2nd and 3rd 2006, at Osaka University, Nakanoshima Center. We introduced our activities and had lectures and training using PCs by a program different from last year, in corporation with DDBJ. The total participants were 54 people.

Snapshot of the Workshop.

Reports

Report on the Remediation of Primary Citations of PDB data

The original PDB entries have very many incomplete primary citations. Accurate primary citations are key archival data, and they are essential for linking to other databases. Historically, BNL had an archive of the reprints of the primary citations, but they were not complete, although 1000 of those reprints are still valuable as some pre-date PubMed. So far, the three wwPDB members have made efforts to remediate the primary citation information, but they were not complete and were done independently. The purpose of this remediation of the primary citations in PDB data is to examine the current situation of the citation information, to plan a systematic way to correct the information, and to execute the remediation following it. Finally, a new procedure to add the corrected information to the current PDB data was considered.

We first examined the primary citation information in the PDB data on May 10, 2005, which included a total of 31,663 entries, in as systematic a manner as we possibly could. First, we assume that 16,897 entries, which had the consensus citation information among the three wwPDB members with the same PubMed numbers, had reliable citation information. In the rest of the 14,766 entries, 544 entries had no information on primary citations. 2,798 entries had the citation information only as “to be published” , and 958 entries had no PubMed ID with some citation information that is not identical among the three wwPDB members. These 4,300 (544+2,798+958) entries seemed to have the most unreliable citation information. In addition, in the 10,466 (14,766-4,300) entries with the PubMed ID, we found 4,208 entries, whose citation information was not clearly described. Thus, the primary citations in the 8,508 (4,300 + 4,208) entries needed to be remediated, by collecting the actual articles and examining them manually.

The three annotators in the PDBj group have been constructing a new literature archive by collecting primary citations, producing electronic copies, and storing them in a TByte hard disk, by using the Osaka University Library with 12,000 journals. So far, more than 14,000 such files for the PDB entries have been collected. Using this literature archive with the previous BNL archive, the PDBj group has manually examined the citation information of the above 8,508 entries. The following is the result of the examination.

For citations without the PubMed IDs (4,300 entries):

Established the correct primary citations with PubMed IDs:	1,226
Established the correct primary citations without PubMed IDs:	387
Structural genomics primary citations may not be published:	697
Confirmed that the citation is “Unpublished” by the authors:	84
Obsolete or replaced ID after May 10, 2005:	65
Stopped remediation for Theoretical models:	383
Sub-total:	2,842

(The remaining 1,458 are still being annotated at PDBj. Among them, 1,365 are "to be published".)

For citations with the PubMed IDs (4,208 entries)

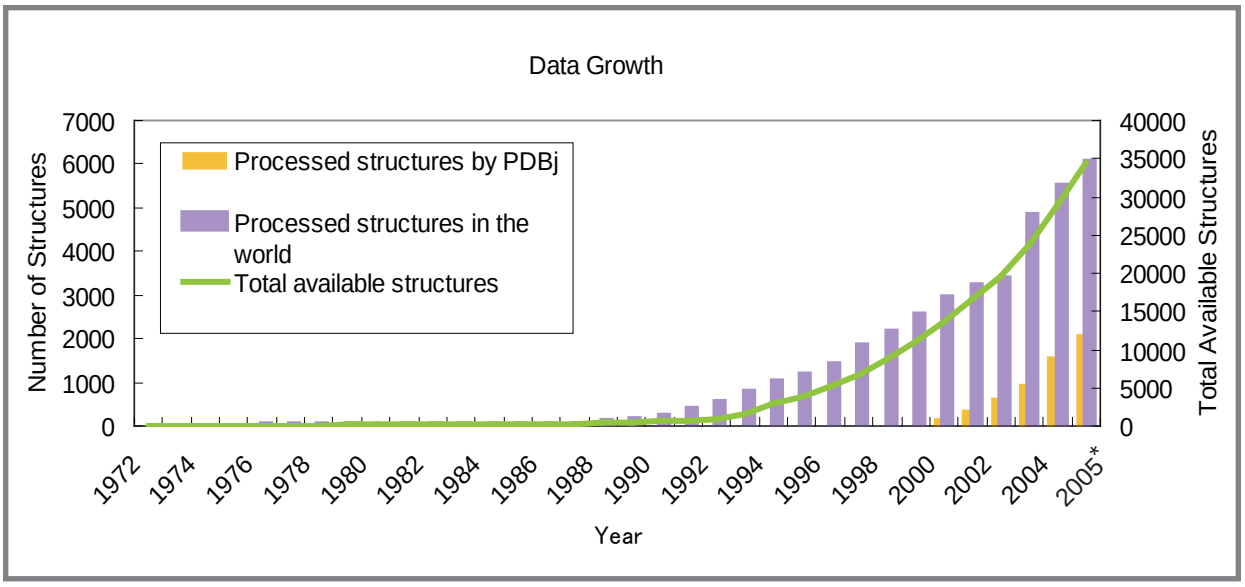
PDBj annotated and confirmed:	4,056

(The remaining 13 have not been published yet, and 139 are theoretical models)

In conclusion, the PDBj group has succeeded in confirming and annotating most of the primary citation information for the unreliable data. Among the initial 8,508 entries, only 1,471 entries were not confirmed, but most of them (about 93%) are "to be published" newly registered data. Thus, only about 100 entries were left to have ambiguous or unreliable citation information among 31,663 PDB entries dated on May 10, 2005.

The remediated data by the PDBj group should be added to the current PDB database, while keeping the original primary citation information by the authors. In addition, it is clear that "to be published" data should be systematically updated by inventing and developing a new system, where the depositors can easily revise the information by themselves.

Statistics



* Last updated : Dec. 31th, 2005

Services

Extension of ASH functions

Pazos and Valencia (2001) and Goh et al. (2000) have developed a method to predict protein-protein interactions by evaluating co-evolution for a pair of proteins. However, the method is known to include many false positives in the prediction. We assumed that the cause of the false positives is the background information about the evolutionary relationships of the source organisms. We tried to exclude such information from the prediction by using a projection operator, and found that the method can drastically reduce the number of false positives from the prediction (Sato, Yamanishi, Kanehisa, Toh (2005) Bioinformatics 21, 3482-3489). The program for the method, which is written in R, can be downloaded from the ASH HP. We are now proceeding further extension of ASH functions, which include hybrid evolutionary trace and correlated mutation analysis. The new version of ASH will be released next year.

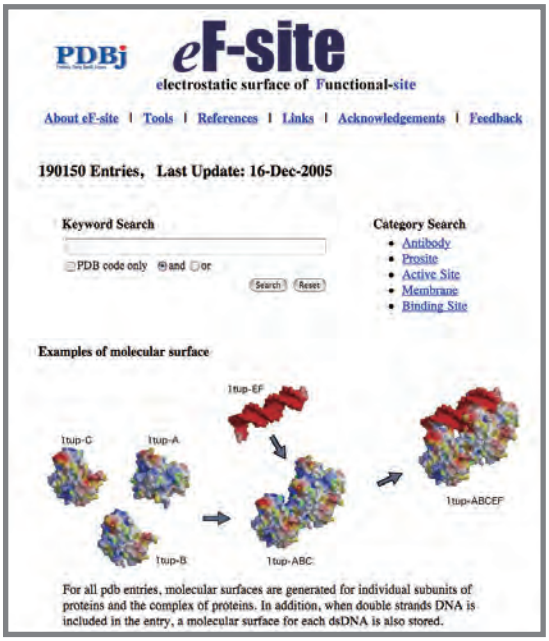
The ASH frontpage.

jV3

jV3, an interactive 3D viewer program to visualize protein and nucleic acid molecules, has been upgraded. The main change of the upgraded version is additional options: - script, which allows drawing and other functions to be run as a batch job. For instance it is possible to draw a series of graphs and then create an animation by using OTS software. The other supplements are a response to the latest JOGL: Java OpenGL and an elevation of stability by the revision of GUI implementation and its surroundings. The display of functional information on xPSSS is also revised.

eF-site

eF-site has relaunched the website (http://ef-site.hgc.jp/eF-site/). It stores almost all protein molecular surfaces in the PDB, including updates every week. For protein complexes, it contains not only the molecular surface to the entire stricture but every subunit helping study of the interaction among proteins. As of Dec 16, 2005 there are about 190,000 entries.



The eF-site frontpage.

eF-surf

We released eF-surf, which is a database that stores the protein molecular surface and the electrostatic potential using the same method of eF-site (http://ef-site.hgc.jp/eF-surf/). The user can download the calculated data of the molecular surface in the form efvet, XML and molscript by uploading the files in PDB format.

Soon !!

We will have a luncheon-seminar on April 26th at the 6th Annual Meeting of the Protein Science Society of Japan.

Date and Time: April 26th. Wednesday, 11:45AM to 12:45PM
Place: Room B-2, Kyoto International Conference Hall, Kyoto, Japan

Staff

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-BIRD)
Matsura, Kanna (JST-BIRD)

Group for Development of PDBj database
Ito, Nobutoshi, Prof. (Tokyo Med. Dent. Univ.)
Paehler, Arno, Dr. (JSTBIRD)
Standley, Daron M., Dr. (JST-BIRD)
Yamashita, Reiko (JST-BIRD)
Yoshihara, Atsuro (JST-BIRD)
Shimizu, Yukiko (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 NMR Database
Akutsu, Hideo, Prof. (IPR, Osaka Univ.)
Matsuki, Yoh, Dr. (JST-BIRD)
Nakatani, Eiichi（JST-BIRD)

Secretary
Kamada, Chisa (JST-BIRD)

Contacting

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
URL: http://www. pdbj.org/