News Letter

Vol. 2

June 2003

Protein Data Bank Japan

http://www.pdbj.org/

PDBj

Protein Data Bank Japan

PDBj
Protein Data Bank Japan

Protein Data Bank Japan

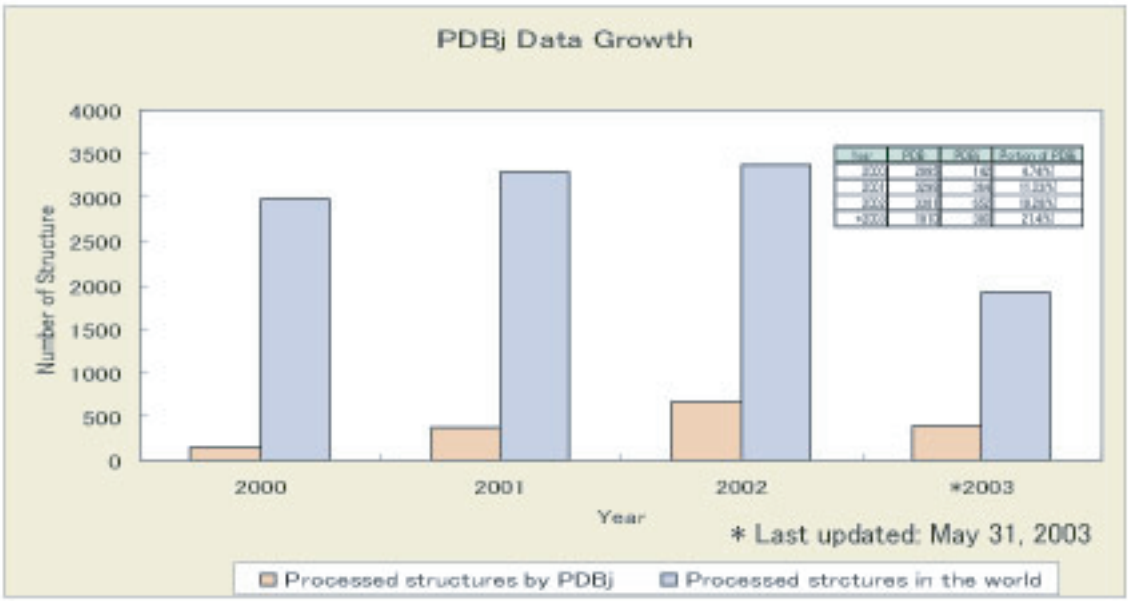
http://www.pdbj.org/

News Letter Vol. 2

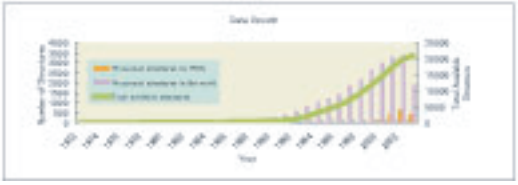
PDBj

Today gene information for many species is just at the point of being revealed. To make use of this information, it is necessary to look at the proteins for which the genes code. In particular, elucidation of an individual protein structure can reveal its specific function, and so structures are the key to understanding gene evolution and biological function on a molecular level. We, PDBj (Protein Data Bank Japan), maintain the database for these protein structures with financial assistance from the Institute for Bioinformatics Research and Development of Japan Science and Technology Corporation (BIRD-JST), collaborating with the Research Collaboratory for Structural Bioinformatics (RCSB) and the MSD in the European Bioinformatics Institute (EBI) in EU. All three organizations serve as deposition, data processing and distribution sites. For further information, see our Web site, http://www.pdbj.org/.

PDB Deposition Statistics

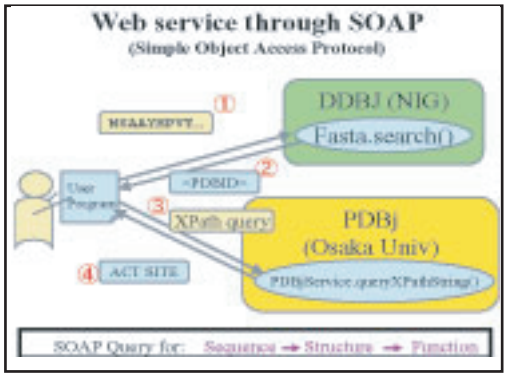


As of May 31st, 2003 the number of registered structures in the PDB (Protein Data Bank) is 21055, and the contribution from PDBj is ever increasing. On the PDBj Web page, the average number of access requests per day is 6,894, and the average number of page views delivered per day is 717. Both of these figures are higher than in the previous fiscal year.



News Letter Vol. 2

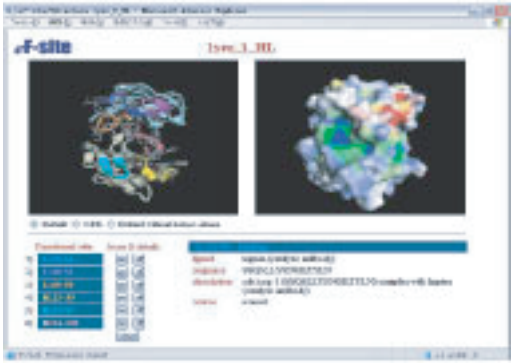
XML description of PDB data and advanced search systems



We developed our own preliminary XML (eXtensible Markup Language) description of PDB data along several fronts since the last fiscal year, keeping the data contents in coincidence with that described by the mmCIF format. Using this preliminary development of the XML database, RCSB, EBI and PDBj have made efforts to establish a new canonical XML description of PDB data by June, 2003. The new schema file with the fully tagged data is now open at http: //deposit.pdb.org/pdbML/pdbx-v0.904.xsd, and the schema for the alternative description with separated coordinate files is available at http://deposit.pdb.org/pdbML/pdbx-v0.904-alt.xsd.
Using the extensible nature of the XML format, we have also made efforts to include more information relating to individual proteins. Currently, PDB files are lacking a detailed description of function, experimental conditions, and the like. We add this information to the XML files by extracting it from journal articles that reference the PDB entries. The number of these annotated PDB entries is now over 2000. Based on the schema of the canonical XML description, we have prepared an experimental XML database (XML-DB_PDBj), including the additional information above.
For querying the XML database, we provide multiple search strategies using X-Path. The best way to make use of an XML database is with Web Services. This is implemented using a protocol called SOAP (Simple Object Access Protocol). For information on our experimental services, see our Web site, http://www.pdbj.org/XML-DB_PDBj/.

eF-site database

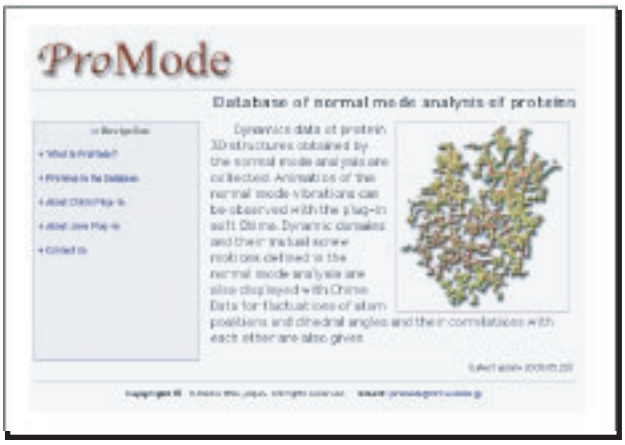
At eF-Site (electrostatic molecular surface of Functional-Site), we probe the relationship between structure and function with an emphasis on the surface properties of the functional site. This is because proteins interact primarily at the molecular surface. Since the important properties that characterize a given functional site are the electrostatic potential and the geometry at the surface, we display these properties graphically, and enable surface-based database queries. The functional site information is grouped in five categories: antibody, prosite, active site, membrane proteins, and binding site. We have also recently added two new categories: mononucleotides and DNA. Moreover, to enhance convenience for the user, we developed our own web-based viewer (PDBjViewer) that enables visualization of the molecular surfaces of functional sites and the surface properties at the same time with conventional molecular graphics facilities.



In the near future, eF-Site will display a more exhaustive binding site database, consisting of the targets to all registered small molecule-binding proteins in the PDB database.

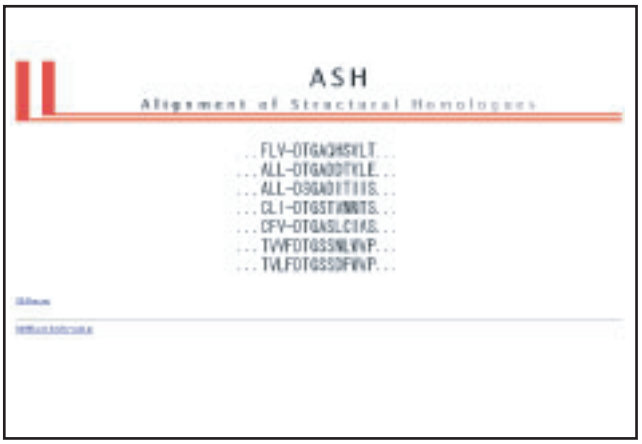
News Letter Vol. 2

*Pro*Mode database



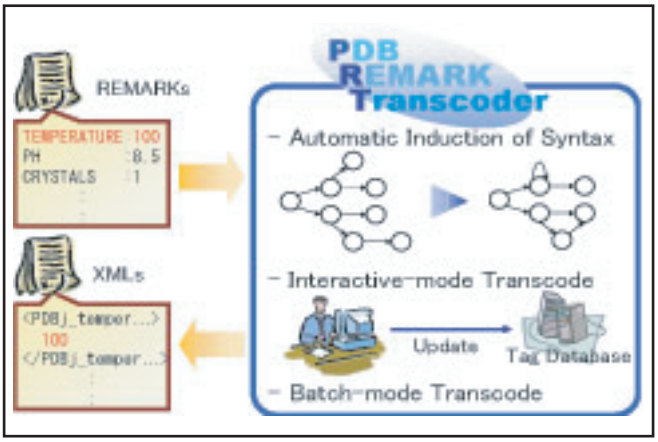
*Pro*Mode is a database of normal mode analysis (NMA) of proteins, developed by Prof. Hiroshi Wako at Waseda Univ. Dynamics simulations of protein molecules are important to understand the principles of protein structure and function. NMA is based on the simple harmonic approximation, and it has been confirmed that the results from NMA are reasonable qualitatively in most cases. NMA is much less time-consuming in computation and more systematic and adequate for the analysis carried out routinely for many proteins than Monte Carlo or Molecular Dynamics simulations. The most time-consuming part of the NMA is regularization and energy minimization of the original structures in the PDB. *Pro*Mode uses the program FEDER developed by the group of Prof. Nobuhiro Go at Japan Atomic Energy Research Institute for rapid and efficient computations. (About FEDER, see T. Noguti and N. Go (1983) *J. Phys. Soc. Jpn* **52**, 3685; H. Wako and N. Go (1987) *J. Comp. Chem.* **8**, 625) Now, 80 proteins are available, and much more numbers of proteins will soon be analyzed and registered.

Three-dimensional comparison tool: ASH



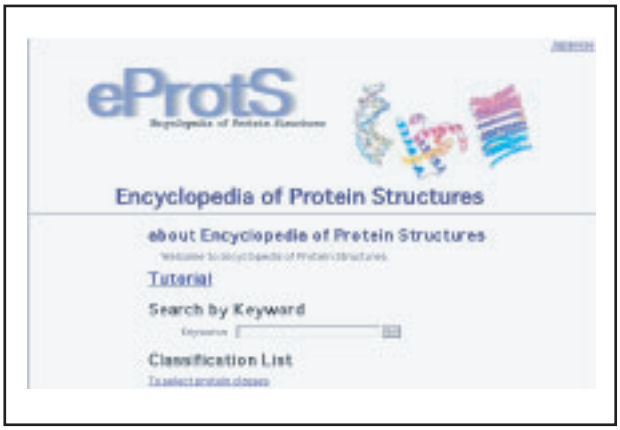
"Structural Alignment", in parallel with other residue-based properties, makes use of protein conformational information. The reliability of the extracted information is significantly enhanced by simultaneously aligning many structures (multiple structure alignment). The team of Prof. Hiroyuki Toh at Kyoto Univ. has developed the protein structural multiple alignment system -ASH- based on the double dynamic programming algorithm that was originally proposed by Taylor and Orengo in 1989. This system utilizes the "distance cut-off approximation", introduced by Prof. Toh in 1997, as well as the "two-step alignment method" and the secondary structural forward alignment technique. These improvements have not only increased the processing speed, but have also improved the accuracy of the resulting alignments. The program can be accessed on the web at: http://timpani.genome.ad.jp /~ash/ and also on the PDBj home page.

PDB REMARK transcoder



The team of Prof. Takenao Okawa at Osaka Univ. has developed "PDB REMARK transcoder", which is an effective translation tool for automatically extracting relevant information from PDB file REMARK lines and converting the information to XML format. At first, the PDB REMARK transcoder memorized description patterns automatically based on REMARK data from 100 entries in the PDB. With this data, 8,906 entries (about 5.2 million words) were converted to XML, and the accuracy of tagging for 10 entries that were selected at random was examined. The tagging is ranked fifth-level automatically based on the distance of property value from cluster to token when extracting description patterns. The highest reliable level accounts for 85% of all 8,906 entries. Thus PDB REMARK transcoder is significantly effective at the preliminary step toward the final artificial adjustment. We have now been optimizing the reliability of the transcoder by training on known data.

eProtS



The PDB is aimed at experts in structural biology and is described only in English. This format may not be very convenient for users who are not specialists of structural biology, or for Japanese students. For this reason, we have started to develop the encyclopedia of Protein Structures "eProtS". This database takes advantage of XML by summarizing important genetic information in a very comprehensive form. The resulting document serves as an educational resource for non-experts. We currently introduce about a hundred proteins in this manner, both in English and Japanese.

PDBj Staff

Director

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

Group for PDB Database Curation

Kusunoki, Masami (Associate Prof., IPR, Osaka Univ.)
Kosada, Takashi (Technical Assistant, IPR, Osaka Univ.)
Igarashi, Reiko (Research Assistant, JST)
Kengaku, Yumiko (Research Assistant, JST)
Saeki, Shinobu (Research Assistant, JST)
Morita, Yasuyo (Research Assistant, JST)

Group for Development of XML description

Ito, Nobutoshi (Prof., School of Biomedical Science, Tokyo Medical and Dental Univ.)
Standley, M. Daron (Researcher, JST)
Kobayashi, Kaori (Technical Assistant, JST)
Sakamoto, Hisashi (Technical Assistant, JST)
Shimizu, Yukiko (Research Assistant, JST)
Takahashi, Aki (Research Assistant, JST)

eF-Site database

Kinoshita, Kengo (Instructor, Grad., Sch., of Integrated Sci., Yokohama City Univ.)

*Pro*Mode database

Wako, Hiroshi (Prof., School of Social Sciences, Waseda Univ.)

Three-dimentional comparison tool: ASH

Toh, Hiroyuki (Prof., Bioinformatics Center, Institute for Chem. Research, Kyoto Univ.)

PDB REMARK transcoder

Okawa, Takenao (Associate Prof., Grad. Sch. of Informatics Sci., Osaka Univ.)
Kaneda Yoshikazu(Assistant Researcher, JST)

Group for NMR database (NMR)

Akutsu, Hideo (Prof., IPR, Osaka Univ.)
Matsuki, Yoh (Researcher, JST)
Nakatani, Eiichi (Technical Assistant, JST)

Secretary

Kamada, Chisa (JST)

At the IPR, (Left back to right):
Nakamura, H., Kamada, C., Sakamoto, H., Standley, M. D., Kosada, T., Nakatani, E.,
Saeki, S., Takahashi, A., Shimizu, Y., Kobayashi, K.,
Morita, Y., Igarashi, R., Kengaku, Y., Kusunoki, M.

Office address

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)6-6879-8636
FAX: +81-(0)6-6879-8636
E-mail: pdbj@protein.osaka-u.ac.jp