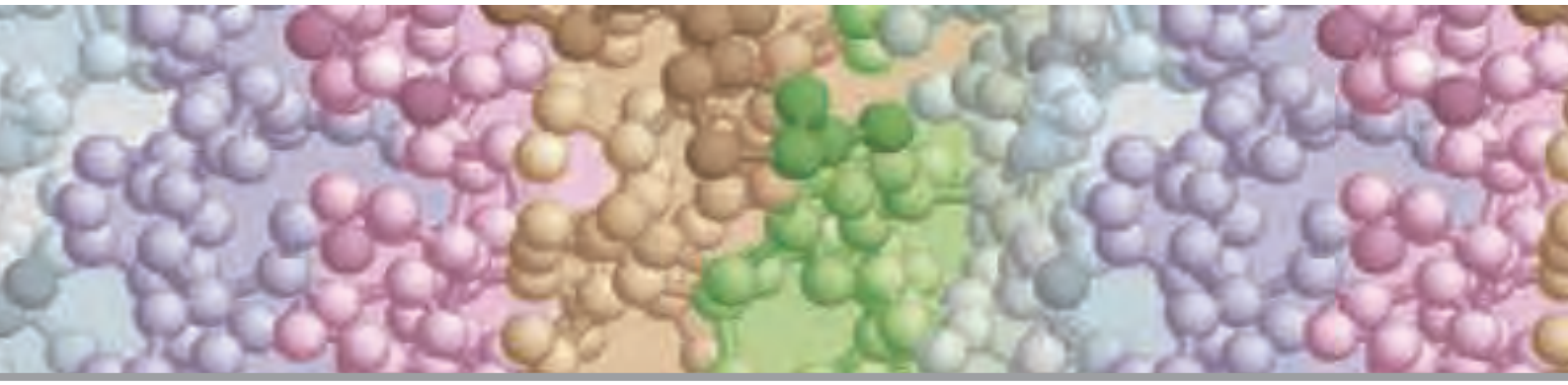
PDBj
Protein Data Bank Japan

http://www.pdbj.org/

News Letter Vol. 11, No. 1

April 2009



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

News

Change to the new formats V 3.15 and V 3.2

The wwPDB has created the new PDB formats, and all the data in the new formats were released on March 17, 2009. The files originally released before December 2, 2008 follow the PDB format V 3.15, and the files originally released after that date follow V 3.20. The format V 3.15 is essentially the same as V3.2, except some REMARK information, where the previously described data could not be converted properly. For details about the differences between V3.15 and V3.2, see: http://www.wwpdb.org/documentation/changesv3.15.pdf. The format change was not only for the flat PDB format data, but also for those of the mmCIF and the PDBML data described by XML. The full description of the new format V 3.2 is found at: http://www.wwpdb.org/documentation/format32/v3.2.html

Summary of the changes is as follows:

1) A new record, SPLIT, identifies multiple entries that are part of one big complexed structure.
2) A new record, NUMMDL, indicates the number of structural models, when multiple model structures are deposited.
3) A new record, MDLTYP, specifies the minimized average model for the NMR structure, and the entries containing only Cα atoms in case of proteins and P atoms in nucleic acids.
4) New additional records, DBREF1 and DBREF2, contain sequence references, when the accession codes or sequence numbering is too long to be fit with the DBREF format.
5) REMARKs 475 and 480 indicate Zero occupancy residues and atoms, respectively.
6) Metal coordination information is provided in REMARK 620.
7) Database references:
7-1) The source organism as listed in NCBI Taxonomy database is indicated by the Taxonomy ID.
7-2) PubMed IDs and DOIs are available for the primary citations.
8) Biological assemblies: The quaternary assembly as calculated by PISA/PQS as well as author provided biological unit are included in the files (REMARK 350).
9) Binding sites: SITE records define any residues that interact with ligands and metal ions, based on distance. Author provided information is also included. An evidence code has been added to identify whether the SITE records are software calculated or author provided. The distance restraints are limited to non H-atom contacts less than 3.7Å.
10) Electron microscopy and NMR templates have been updated and standardized.
11) Small molecule chemistry: The Chemical Component Dictionary (http://www.wwpdb.org/ccd.html) has been enhanced with consistent chemical and systematic naming, re-generation of SMILES strings, and chirality checks.
12) Other minor problems have been fixed.

Soon!

We will have a luncheon-seminar on May 21st, 2009
at the 9th Annual Meeting of the Protein Science Society of Japan

Date and Time: May 21st Thursday, 11:45AM to 12:45PM
Place: Room C (Wakakusa), ANA Hotel KUMAMOTO NEW SKY, Kumamoto, Japan

DDBJing & KEGGing & PDBjing Database Workshop in Kyoto

DDBJing & KEGGing & PDBjing Database Workshop was held at Bioinformatics Center, Institute for Chemical Research, Kyoto University on November 27th -28th. We introduced our services and had training using PCs in cooperation with DDBJ, KEGG and DBLCS (Database Center for Life Science).

A snapshot of the Workshop.

The staff members of the database workshop from DDBJ, KEGG, PDBj and DBLCS.

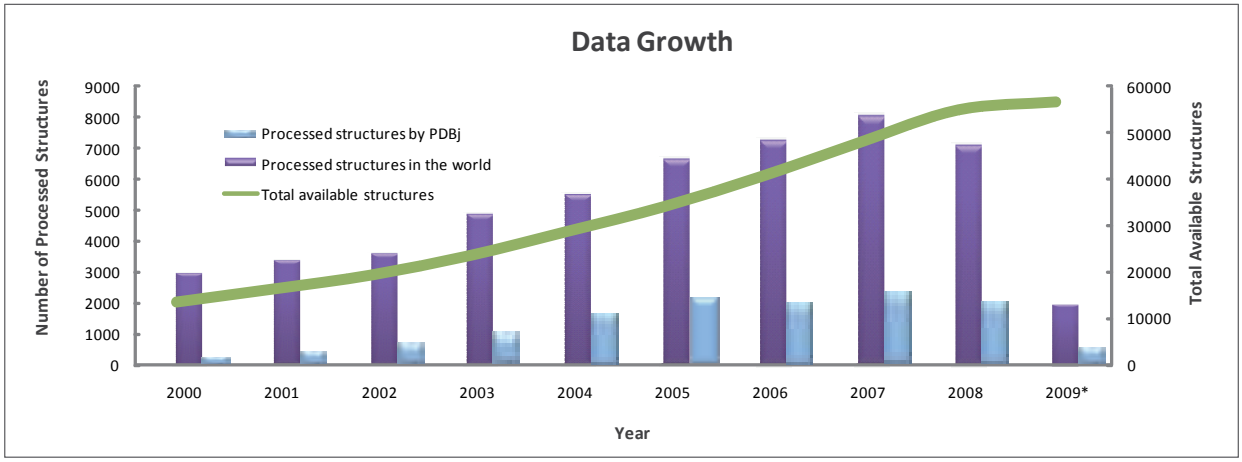
The 46th of the Annual Meeting of the Biophysical Society of Japan

The 46th of the Annual Meeting of the Biophysical Society of Japan was held on December 3rd - 5th at Fukuoka Convention Center in Fukuoka. We introduced our databases and services.

Snapshots of the Workshop.

Statistics

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



* Last updated : April 1, 2009

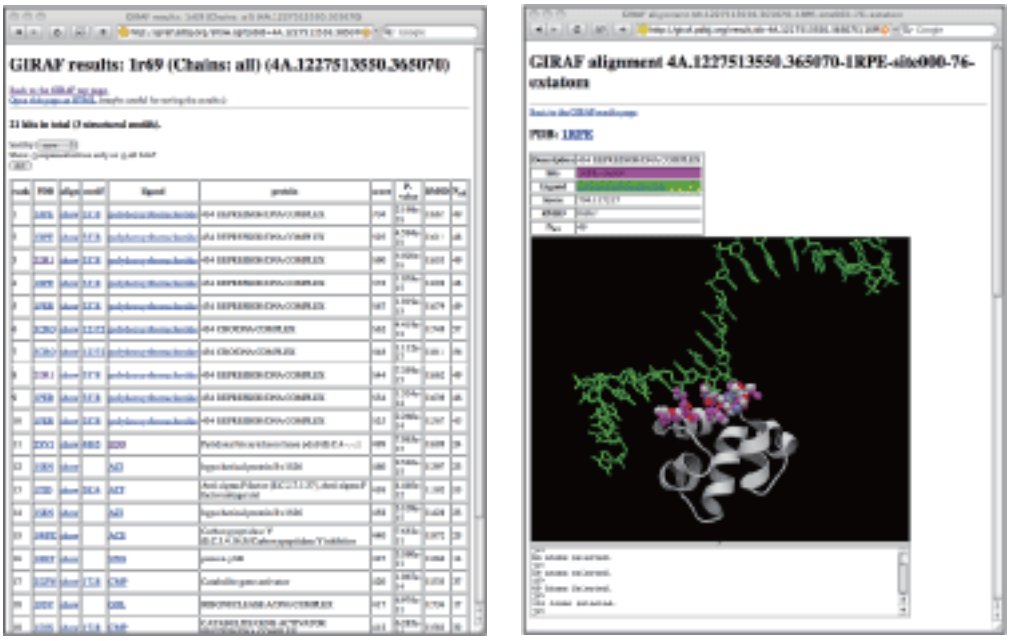
Services

GIRAF: Searching for similar ligand binding site structures

Summary: PDBj released a new service GIRAF (Geometric Indexing and Refined Alignment Finder) which, given a PDB ID or uploaded PDB file, searches and aligns substructures that are similar to known ligand binding sites in the PDB. The underlying database contains more than 186,000 ligand binding site structures including DNA, RNA and peptide binding sites. Although it depends on the size of the query structure, a typical search is completed within 3 to 15 minutes. The search results are summarized in a list of matching sites (Figure, center) which contains links to the pages that contain detailed annotations and molecular graphics of each hit (Figure, right). Molecular graphics is implemented with the jV, a graphics software developed by PDBj, so that you can look into the details of the superposition interactively.

The GIRAF frontpage.

How GIRAF works: GIRAF searches a big database of ligand binding sites in a relatively short period of time by using a technique called "Geometric Indexing" developed at PDBj. In this method, the ligand binding sites extracted from PDB are preprocessed by changing the coordinate frames with structural annotations, and then stored into a relational database. The stored data are properly indexed according to their geometric features. When a search is conducted, GIRAF first refers to the index, thus avoiding exhaustive comparisons. Consequently, refined alignments are constructed only for those binding sites found in the index, which greatly reduces the search time. For the details of the GIRAF method, please refer to Kinjo & Nakamura, BIOPHYSICS, 3:75-84 (2007); for its applications, see Kinjo & Nakamura, Structure, 17:234-246 (2009).



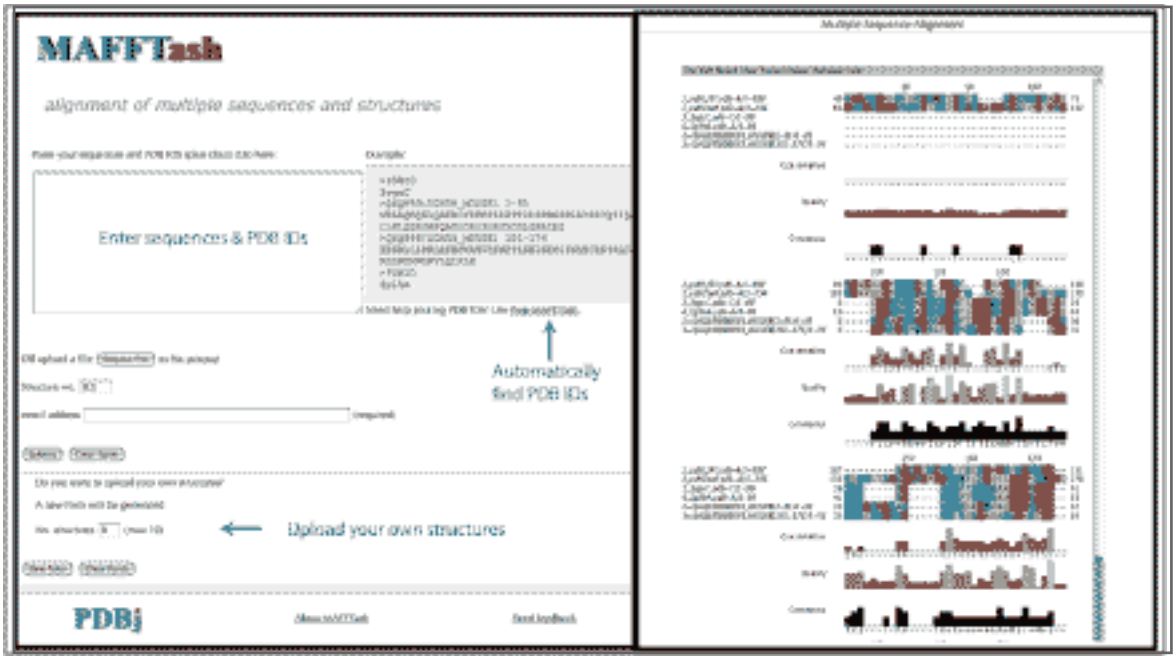
An Example of the GIRAF.

MAFFTash: Multiple alignment for protein sequences and structures

MAFFTash is a web-based tool for calculating multiple sequence alignments (MSAs) from sequences and, if available, structures. Structures can be provided by their PDB identifiers, or by uploading the PDB-formatted files directly. MAFFTash works by first aligning all structural domain pairs using the programs ASH and Protein Domain Parser. The use of domain decomposition means that the structural alignments are “rigid” within domains but 'flexible' between domains. Then MAFFTash extracts continuous stretches of residue pairs that align well and uses this information to improve the MSA. The final result is computed using the program MAFFT, a widely-used package that employs a number of different algorithms to efficiently construct an accurate MSA.

To run MAFFTash you must provide a list of sequences and either PDB identifiers or external PDB-formatted files. Note that chain identifiers are now mandatory for all PDB entries, so whitespaces, dashes, and underbars are not acceptable chain identifiers. If you are uncertain about which chain IDs to use, please use the xPSSS search engine: type in your PDB ID, then click on “sequence information (FASTA format)” . You will see the PDB sequence for each chain in FASTA format. Note also, that MAFFTash provides a tool for automatically picking up a set of PDB IDs, given a set of (FASTA-formatted) sequences. To use this feature, click “Prep-MAFFTash” under the Example on the MAFFTash top page. It is possible to adjust the weight of the structural information. Currently, this weight must be adjusted by hand, and should be increased (e.g., to 1) if hundreds of sequences or more are input.

An example is shown below. In this example, two PDB IDs, two external PDB files, and two sequences, corresponding to Caspase activating and recruitment (CARD) domains and their homologs, were input. The resulting alignment is displayed in Jalview, and can be downloaded as a FASTA-formatted text file.



An example of the MAFFTash.

Staff

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

Group for PDB Database Curation
Nakagawa, Atsushi, Prof. (IPR, Osaka Univ.)
Matsuura, Takanori, Dr. (IPR, Osaka Univ.)
Igarashi, Reiko (JST-BIRD)
Kengaku, Yumiko (JST-BIRD)
Matsuura, Kanna (JST-BIRD)
Inoue, Mayumi (IPR, Osaka Univ.)
Chen, Minyu (IPR, Osaka Univ.)

Group for Development of new tools and services
Kinjo, Akira R., Dr. (IPR, Osaka Univ.)
Iwasaki, Kenji, Dr. (IPR, Osaka Univ.)
Suzuki, Hirofumi, Dr. (IPR, Osaka Univ.)
Yamashita, Reiko (JST-BIRD)
Kamada, Chisa (JST-BIRD)
Shimizu, Yukiko (JST-BIRD)
Kudou, Takahiro (IPR. Osaka Univ.)

Group for NMR Database
Fujiwara, Toshimichi, Prof. (IPR, Osaka Univ.)
Akutsu, Hideo, Prof. (IPR, Osaka Univ.)
Kobayashi, Naohiro, Dr. (IPR, Osaka Univ.)
Nakatani, Eiichi（JST-BIRD)
Harano, Yoko (IPR, Osaka Univ.)

Group for Medical Institute of Bioregulation, Kyushu University
Toh, Hiroyuki, Prof. (MIB, Kyushu Univ.)
Katou, Kazutaka, Dr. (DMI, Kyusyu Univ.)
Ohtsu, Miki (MIB, Kyushu Univ.)

Collaboratory Reserchers
Wako, Hiroshi, Prof. (Waseda Univ.)
Ito, Nobutoshi, Prof. (Tokyo Med. Dent. Univ.)
Kinoshita, Kengo, Dr. (IMS, Univ. of Tokyo)
Standley, Daron M., Dr. (iFReC, Osaka Univ.)

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/