

# Protein Data Bank Japan (PDBj): updated user interfaces, resource description framework, analysis tools for large structures

Akira R. Kinjo\*, Gert-Jan Bekker, Hirofumi Suzuki, Yuko Tsuchiya, Takeshi Kawabata, Yasuyo Ikegawa and Haruki Nakamura

Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

Received September 15, 2016; Revised October 06, 2016; Editorial Decision October 07, 2016; Accepted October 11, 2016

## ABSTRACT

The Protein Data Bank Japan (PDBj, <http://pdbj.org>), a member of the worldwide Protein Data Bank (wwPDB), accepts and processes the deposited data of experimentally determined macromolecular structures. While maintaining the archive in collaboration with other wwPDB partners, PDBj also provides a wide range of services and tools for analyzing structures and functions of proteins. We herein outline the updated web user interfaces together with RESTful web services and the backend relational database that support the former. To enhance the interoperability of the PDB data, we have previously developed PDB/RDF, PDB data in the Resource Description Framework (RDF) format, which is now a wwPDB standard called wwPDB/RDF. We have enhanced the connectivity of the wwPDB/RDF data by incorporating various external data resources. Services for searching, comparing and analyzing the ever-increasing large structures determined by hybrid methods are also described.

## INTRODUCTION

The Protein Data Bank Japan (PDBj) (1) accepts and annotates macromolecular structure data in collaboration with other worldwide Protein Data Bank (wwPDB) (2,3) partners, PDBe (4), RCSB PDB (5) and BMRB (6). The wwPDB partners recently started a new, unified deposition and annotation system called OneDep. Accordingly, depositions from Asia and the Middle East are directed to the PDBj deposition site (<http://deposit-pdbj.wwpdb.org>) and are annotated by the annotators at PDBj.

In addition to processing depositions and maintaining the archive, PDBj provides a wide range of analysis tools and derived databases to facilitate structural biology and bioinformatics research. Since the last publication of our NAR DB article (1), PDBj has undergone a number of

major changes regarding user interfaces and analysis tools as well as additional data provided. The previously described Resource Description Framework (RDF) format, PDB/RDF, is now one of the wwPDB standard formats called wwPDB/RDF and is enhanced with supplementary information in order to connect PDB data with other biological data resources.

## USER INTERFACES

User interfaces include interactive (and graphical) web interfaces for humans and RESTful web services for computer programs. We also expose our backend database in the forms of web services or dump files for enabling very complex queries. These are described in turn.

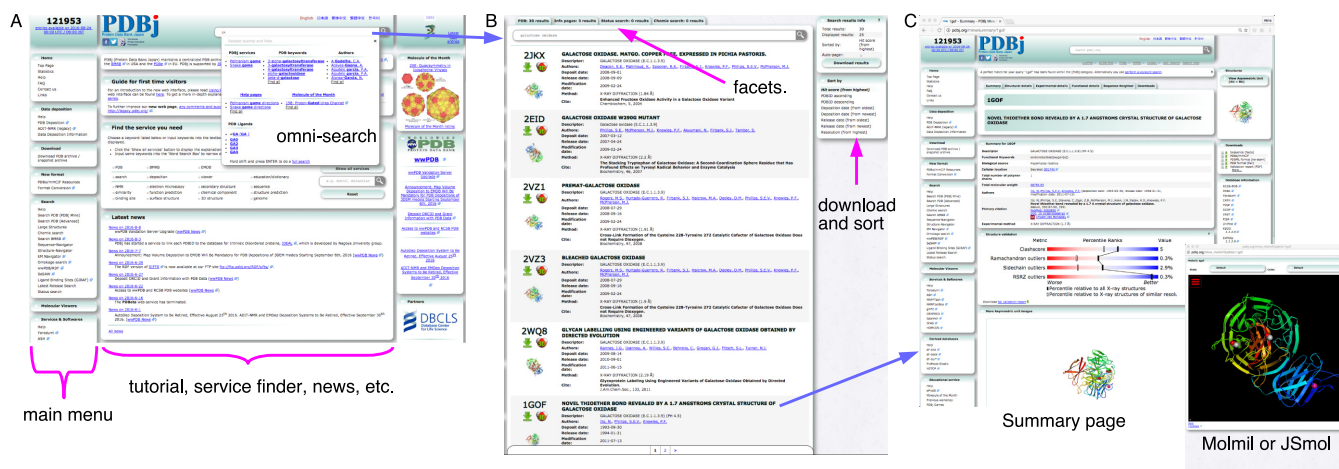
### Web interface

The web interface of PDBj was updated to provide a uniform integrated interface for the available services as well as to provide a scalable interface for devices ranging from smartphones to workstations. This update incorporates several innovative/renovative features as described below.

We have implemented various functionalities to ease novice as well as expert users to interact with inherently complex macromolecular structural information. For the novice user, we have created a series of interactive tutorials for PDBj's search services (<http://pdbj.org/help/tutorials>). In addition, the service finder on the PDBj top page helps the user to find particular services provided by PDBj. The omni-search service, which is present on the top of nearly every page at the PDBj web site, provides faceted navigation of search results categorized into PDB entries, general information pages, status search and chemical components (Chemie, see below). In fact, the omni-search provides candidate search results in these categories while the user is typing (Figure 1).

One of the most basic utilities for structural analysis is the molecular viewer. We previously developed the Java applet-based molecular viewer jV (7). However, modern devices

\*To whom correspondence should be addressed. Tel: +81 6 6879 4311; Fax: +81 6 6879 6836; Email: [akinjo@protein.osaka-u.ac.jp](mailto:akinjo@protein.osaka-u.ac.jp)



**Figure 1.** Basic web user interface. (A) On the PDBj home page, the user finds the main menu for all the services (on the left side of the page), interactive tutorials (top center) and service finder (in the middle). The omnisearch bar (top left) guides the user to input keywords that are categorized into service types. (B) A search result is shown in several facets (PDB entries, 'Info pages', release status and chemical components) where the given keywords are found. (C) In the summary page of each PDB entry are shown basic information of the entry such as the title, authors, citation and validation report. Interactive molecular graphics are available for asymmetric units as well as biological units using Molmil (8) or JSmol (asymmetric units only) (<http://wiki.jmol.org/index.php/JSmol>) (right bottom).

such as smartphones and tablets don't support Java applets. In order to offer a high performance, high quality molecular viewer on these platforms as well as on regular workstation platforms where the usage of Java applets has also diminished in recent years, we have developed a new molecular viewer named Molmil and have deployed this for various services on our web site (8), including the database of electrostatic molecular surfaces (eF-site) (7) even for very large structures. During the development of Molmil, we also developed a new JSON based format derived from PDBx/mmCIF called PDBx/mmJSON (8). Molmil, as well as PDBj's Mine PDB Explorer service use this format. The user can download the mmJSON file of PDB entries via a REST service (see <http://pdbj.org/help/file-formats>). The summary page of each entry is semantically annotated with schema.org vocabulary (<http://schema.org/>) in a microformat. The Sagace search engine (<http://sagace.nibiohn.go.jp/en/>) (9), for example, exploits these annotations to display PDB-specific search results with thumbnails of macromolecular structures.

Along with the graphical interfaces on Web browsers, we have also renewed our RESTful web services. In fact, all services provided by the main PDBj web site are also available as a REST service as described on <http://pdbj.org/help/rest-interface>.

Chemie (<http://pdbj.org/chemie-search>) is a new service offered by PDBj. Much like its counterparts, RCSB PDB's Ligand Expo (<http://ligand-expo.rcsb.org/>) and PDBe's PDBeChem (<http://www.ebi.ac.uk/pdbe-srv/pdbechem/>), PDBj's Chemie offers users an interface to search and explore the Chemical Component Dictionary of the wwPDB via text search of various fields, as well as sub-structure search via SMILES with Tanimoto scoring using fingerprints (10) generated by the Open Babel software (11).

## PDBj Mine RDB

The most basic infrastructure underlying PDBj's web services is the PDBj Mine relational database (RDB) which has been completely redesigned in recent years. When the old version of the PDBj Mine RDB was developed (12), PDBj's backend system depended on the 'PDBMLplus' files that were based on PDBj's own extension of the standard PDBML files (13). Since then, the PDBx/mmCIF format (14) has been adopted as the canonical file format of the wwPDB. Accordingly, we have separated the additional components of the PDBMLplus files from the core PDBML parts. This separation made it possible to design an easier-to-use RDB of the entire PDB data, hence PDBj Mine RDB version 2 offering a leaner database structure as well as better scalability for the recently released large structures. In a nutshell, each table in the PDBj Mine RDB schema corresponds to a category as defined in the PDBx/mmCIF dictionary and different tables are cross-referenced via foreign keys when they are defined in the dictionary and make sense in the RDB. A complete documentation of the schema is provided at <http://pdbj.org/mine-rdb-docs>. The RDB is implemented in PostgreSQL (version 9.3) (<http://www.postgresql.org/>) and the complete database dump and weekly updates are available at our ftp site (<ftp://ftp.pdbj.org/mine2/>) and see also <http://pdbj.org/help/mine2-rdb-local-install> for installation instruction) along with an extensive list of example SQL queries (<http://pdbj.org/help/mine2-sql>). SQL queries can be executed on the web browser (<http://pdbj.org/mine/sql/>) or via the REST interfaces with several options for the output file formats.

One weakness of the PDB data is the scarcity of up-to-date cross-references to other biological databases, which makes it difficult to interpret the biological relevance of macromolecular structures in a wider context. In order to solve this problem at least partially, we have also integrated the SIFTS resource developed by PDBe and UniProt (15) into the PDBj Mine RDB. Currently, the SIFTS sum-

many files provided for ‘quick access’ (<https://www.ebi.ac.uk/pdbe/docs/sifts/quick.html>) are incorporated. Example queries that combine the PDB and SIFTS data are also provided at <http://pdjb.org/help/mine2-sql>. Using this resource, the user can analyze various trends in structural biology as reflected in the PDB. As an example, we show in Figure 2 the yearly change of newly identified Pfam (16) structures, which was obtained by a single SQL query using the web interface.

## WWPDB/RDF: PDB ON THE SEMANTIC WEB

### Basic structure of wwPDB/RDF

Partly as a product of the BioHackathon (17), we started distributing PDB/RDF, the PDB data in the Resource Description Framework (RDF) format, in 2011 (1). The PDB/RDF data has become one of the standard formats of the wwPDB and is now called wwPDB/RDF provided at <http://rdf.wwpdb.org/pdb/> on the web as well as on our FTP site at <ftp://ftp.pdbj.org/RDF>. We also provide the Chemical Component Dictionary as RDF (<http://rdf.wwpdb.org/cc/>). These RDF resources are referenced by some of the major data providers such as UniProt (18) and PubChem (19). More recently, we, in collaboration with the BMRB team, have also developed BMRB/RDF which translates nuclear magnetic resonance data in the BMRB into the RDF format accompanied with many links to external biological data resources (20).

The structure of the wwPDB/RDF is mostly unchanged from that which was previously described (1) except for the base URI (now <http://rdf.wwpdb.org/> rather than <http://pdjb.org/rdf/>). That is, the ‘ontology’ of the wwPDB/RDF is based on the PDBx/mmCIF dictionary (<http://mmcif.wwpdb.org/>) (14) so that each PDBx/mmCIF category corresponds to an OWL class and each PDBx/mmCIF item is a data property in the RDF version. However, there are a few major enhancements in the current version for increased interoperability with other data resources. First, some descriptive attributes were added in the datablock layer of each entry. Second, many external resources are referred to by the URI’s based on the Identifiers.org (<http://identifiers.org/>) hosted at European Bioinformatics Institute (21). These changes were recommended in the guideline proposed by the Database Center for Life Science (DBCLS) (<http://wiki.lifesciencedb.jp/mw/RDFizingDatabaseGuideline> [in Japanese]) that coordinates a database integration project in Japan. In the previous PDB/RDF, we excluded some categories such as `entity_poly_seq` in order to reduce the data volume. In the current version, however, we have added `entity_poly_seq`, `pdbx_poly_seq_scheme` and `pdbx_nonpoly_scheme` categories to wwPDB/RDF so that we can make residue-by-residue references to external resources based on, for example, the SIFTS PDB-UniProt sequence mapping. On the other hand, the `atom_site` category is still excluded from the distribution due to its large volume.

### SIFTS enhancement

Similarly to the PDBj Mine RDB, wwPDB/RDF is also supplemented with the SIFTS resource (15) based on

the SIFTS ‘Quick access’ files (<https://www.ebi.ac.uk/pdbe/docs/sifts/quick.html>). Our SIFTS RDF data are distributed as a separate release from the wwPDB/RDF at our FTP site (<ftp://ftp.pdbj.org/RDF/sifts/>). Nevertheless, thanks to the very design of the Semantic Web technologies of which RDF constitutes the core, integration of the SIFTS resource and wwPDB/RDF as well as other data resources should be trivial. To ensure the connectivity of various data, the links to external resources described in the SIFTS data are also based on the URI’s provided by the Identifiers.org (<http://identifiers.org/>). In addition, we also provide sequence alignments between PDB and UniProt sequences using the FALDO ontology (22).

### SPARQL endpoint at National Bioscience Database Center (NBDC RDF Portal)

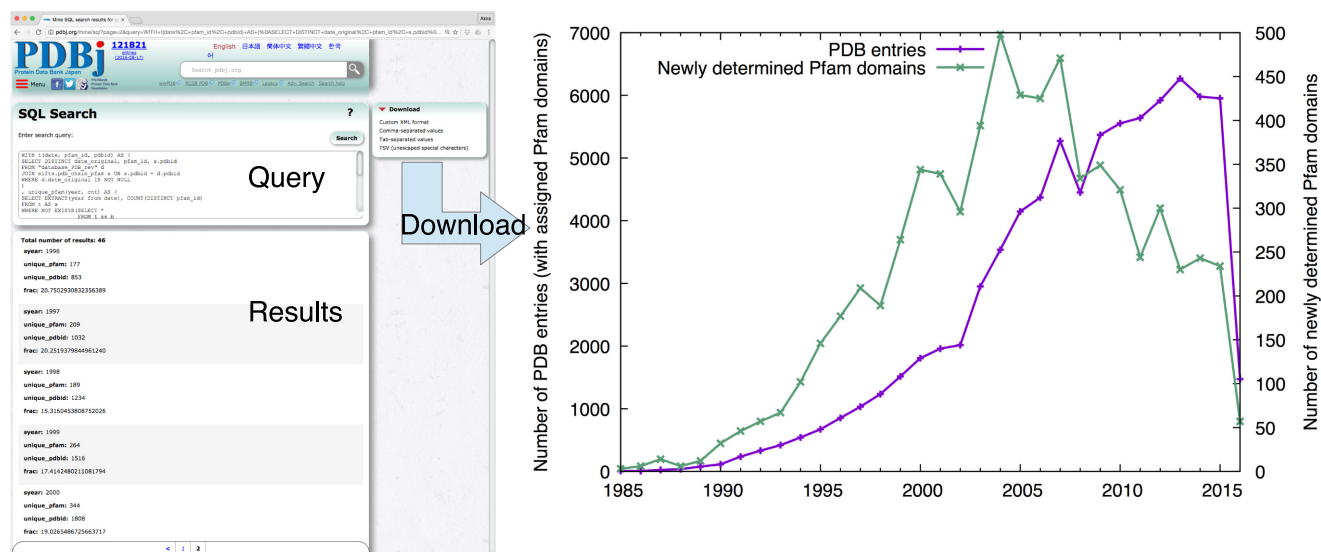
The National Bioscience Database Center of Japan (NBDC) provides the RDF Portal (<http://integbio.jp/rdf/>) where they publish various resources (mostly developed in Japan) in the RDF format and provide a SPARQL endpoint to query these resources. The RDF Portal includes the entire wwPDB/RDF data with the SIFTS data together with other data resources. The aforementioned BMRB/RDF (20) is also a part of this portal. Currently, the wwPDB/RDF data are updated quarterly in January, April, July and October. Although the RDF Portal is not necessarily up-to-date, it provides a convenient access to the wwPDB/RDF data. Some example SPARQL queries are also provided at <http://integbio.jp/rdf/?view=detail&id=pdjb>.

### ANALYSIS TOOLS FOR LARGE AND/OR COMPLEX STRUCTURES

Most of the services offered by the PDBj are summarized in Table 1 (35–49). Here, we focus on those related to analysis of large structures and complexes.

One of the exciting trends in structural biology in the recent years is the elucidation of very large complex structures by hybrid methods that combine electron microscopy (EM) and other methods. Since the information of EM structures is of very different characters than those of traditional PDB structures, it requires special treatment in visualization and analysis. We at the PDBj have been providing the EM Navigator (<http://pdjb.org/emnavi/>) for aiding the animated visualization of EM structures. The EM Navigator is now integrated with the Yorodumi browser (<http://pdjb.org/emnavi/viewtop.php>) so that entries in the PDB and EMDB can be browsed in a single integrated interface (Figure 3A). As the number of EM structures increases, there will be a demand for comparing (and classifying) those structures. The Omokage search (23) provides structural similarity search over macromolecular assemblies in both the PDB and EMDB based on electron density, independent of sequence order and number of subunits. For a given query structure (in the form of either atomic coordinates or an electron density map), Omokage first searches globally similar structures in the PDB and EMDB that have a similar one-dimensional profile (a vector of global shape features), and then generates the ‘align-





**Figure 2.** An example of PDBj Mine SQL queries. Queries can be executed on the web browser as well as via the RESTful web service (the database dump for local installation is also available). The result of a query can be saved in such formats as a custom XML, comma-separated values or tab-separated values. The plot on the right shows the yearly change of newly determined Pfam domains in the PDB, as obtained from a single SQL query against the PDBj Mine RDB. (Note that the data in the years 2015 and 2016 are incomplete because not all the PDB entries deposited in these years have been released yet.)

**Table 1.** Analysis tools and derived databases

Service <sup>a</sup>	Description	URL <sup>b</sup>	Reference
<b>PDBj Mine<sup>c</sup></b>	Simple, advanced and SQL search for PDB data	<a href="http://pdbj.org/mine">pdbj.org/mine</a>	
<b>Chemie search<sup>d</sup></b>	Searching the Chemical Component Dictionary	<a href="http://pdbj.org/chemie-search">pdbj.org/chemie-search</a>	
<b>Sequence Navigator<sup>e</sup></b>	BLAST (27) search against PDB	<a href="http://pdbj.org/seq-navi">pdbj.org/seq-navi</a>	
<b>Structure Navigator<sup>e</sup></b>	ASH structure similarity search	<a href="http://pdbj.org/struc-navi">pdbj.org/struc-navi</a>	(35)
<b>SeSAW</b>	Functional site prediction	<a href="http://pdbj.org/sesaw">pdbj.org/sesaw</a>	(36)
<b>GIRAF</b>	Interaction interface similarity search	<a href="http://pdbj.org/giraf">pdbj.org/giraf</a>	(37,47–49)
<b>Molmil<sup>d</sup></b>	WebGL-based Molecular graphics viewer	<a href="http://pdbj.org/molmil">pdbj.org/molmil</a>	(8)
<b>Yorodumi<sup>e</sup></b>	Integrated PDB/EMDB data browser	<a href="http://pdbj.org/yorodumi">pdbj.org/yorodumi</a>	
<b>ASH</b>	Pairwise structure alignment	<a href="http://pdbj.org/ash">pdbj.org/ash</a>	(35,38–39)
<b>MAFFTash</b>	Integrated multiple sequence/structure alignment	<a href="http://pdbj.org/mafftash">pdbj.org/mafftash</a>	(39)
<b>CRNPRED</b>	One-dimensional structure prediction	<a href="http://pdbj.org/crnpred">pdbj.org/crnpred</a>	(40,41)
<b>Spanner</b>	Homology modeling	<a href="http://pdbj.org/spanner">pdbj.org/spanner</a>	(42)
<b>SFAS</b>	Automated function annotation	<a href="http://pdbj.org/sfas">pdbj.org/sfas</a>	
<b>eF-site<sup>f</sup></b>	Electrostatic surface database	<a href="http://pdbj.org/ef-site">pdbj.org/ef-site</a>	(7,43)
<b>eF-seek</b>	Search against eF-site	<a href="http://pdbj.org/ef-seeek">pdbj.org/ef-seeek</a>	(44)
<b>HOMCOS<sup>d</sup></b>	Searching and modeling of protein complex	<a href="http://homcos.pdbj.org">homcos.pdbj.org</a>	(26)
<b>ProMode Elastic<sup>g</sup></b>	ProMode with elastic network model	<a href="http://pdbj.org/promode-elastic">pdbj.org/promode-elastic</a>	(31,32)
<b>Molecule of the Month<sup>e</sup></b>	Japanese translation of RCSB PDB MoM	<a href="http://pdbj.org/mom">pdbj.org/mom</a>	(45)
<b>PDBj Games<sup>e</sup></b>	Simple and fun games	<a href="http://pdbj.org/pdbj-games">pdbj.org/pdbj-games</a>	

<sup>a</sup>New or largely updated services are in **boldface**.

<sup>b</sup>The 'http://' prefix is omitted.


<sup>c</sup>Simplified RDB schema for the SQL search integrated with SIFTS resources(15), more powerful search engine and updated web interface.

<sup>d</sup>New service.








<sup>e</sup>Updated web interface.

<sup>f</sup>Extended database with large structures.

<sup>g</sup>Updated web interface and extended database.

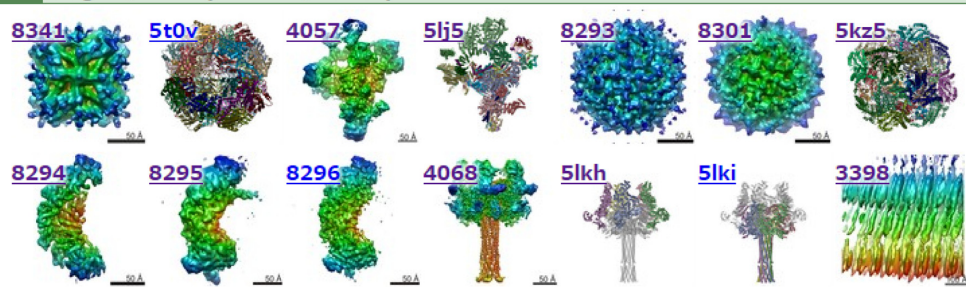
**A**  **EM Navigator** - 3D electron microscopy data navigator -


- **Browse 3DEM data**

- Keywords / EMD-ID / PDB-ID  **Submit**
-  [Advanced search, table view, etc.](#)
-  [Omokage search - shape similarity search](#)
-  [Yorodumi - integration of EMD/PDB metadata & structure viewer](#)
-  [Gallery](#)  [Statistics](#)  [3DEM papers](#)  [Species](#)

- **Recently released data**

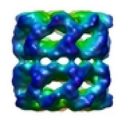
- **Aug 31, 2016 (EMDB: 9, PDB:5)**



**B**  **Omokage search** - Shape similarity search of macromolecules -

+ **Search query**

- **Subject structure**

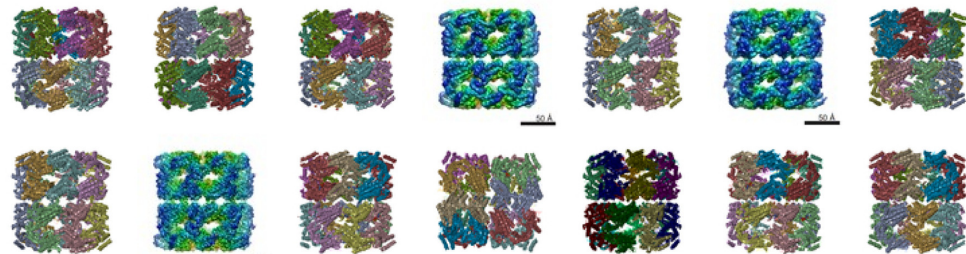
 [Database: EMD / ID: 1080](#)  
A 11.5 A single particle reconstruction of GroEL using EMAN.

- **Search result**

Showing 1 - 100 of 2,000 items found from all (231,674 structures)

Display:   [download CSV file \(for MS-EXCEL, etc.\)](#)

Re-rank top  data by  **Submit** [Detail](#)



**Figure 3.** Analysis tools for the PDB and EMD. (A) The EM Navigator/Yorodumi (<http://pdj.org/emnavi>) provides an integrated interface to the PDB and EMD combined. (B) The Omokage search (<http://pdj.org/emnavi/omo-search.php>) performs shape similarity searches through both the PDB and EMD.

ment' of structures by fitting Gaussian mixture models (Figure 3B) (24).

Another recent trend in structural biology focuses on complex structures between proteins or between proteins and chemical compounds (25). To support analyses and modeling of complex structures, we have added HOMCOS (26), a tool for searching, analyzing and modeling of complex structures based on structural similarities. The HOMCOS server supports several functionalities. First, possible dimeric complexes of proteins are searched for a given pair of protein sequences (based on sequence similarities using BLAST (27)), which serve as templates for modeling the complex structures. Second, possible protein–compound complexes are searched for a pair of a protein sequence and the structure of a chemical compound (in either 3D or 2D). While the protein similarities are detected by BLAST, the similarities between chemical compounds are detected by the KCOMBU program (28,29) and superposed by the FKCOMBU program (30). Third, for a given protein sequence, residue-wise annotations of possible ligand binding sites and effects of mutations are listed. These annotations are imported from the UniProt (18) database in addition to complex structures in the PDB.

Understanding protein functions often requires analysis of protein dynamics. However, running molecular dynamics simulations and analyzing the trajectories are not always easy for the casual user of the PDB. We therefore provide ProMode (31–33) which is a database of protein dynamics computed from normal mode analysis (NMA). The original ProMode database was derived from the NMA based on a physicochemical potential function (31). It was, however, computationally very demanding and could not scale to large structures. We have therefore developed ProMode Elastic (32) which is based on the (all-atom) elastic network model (34) of native protein structures and thus alleviates energy minimization. We are currently in the process of extending the ProMode Elastic to cover many of the (moderately) large complexes in the PDB.

## CONCLUSION

PDBj accepts and processes the deposited macromolecular structural data through the OneDep system as one of the members of the wwPDB. In addition to keeping and providing the common structural archive, PDBj develops our own services. In particular, semantic web tools based on RDF could be useful to link PDB data with other external biological data resources. Other tools to analyze large and complex macromolecular structures following the recent trends in structural biology should also support researchers in the wide range of the life science field.

## ACKNOWLEDGEMENTS

The authors thank the PDBj annotators, Reiko Igarashi, Yumiko Kengaku, Hasumi Cho, Junko Sato for their constant effort that makes the wwPDB possible at all, and Reiko Yamashita and Takahiro Kudou for technical support. A. R. K. thanks Shuichi Kawashima (DBCLS) and Katsuhiko Okubo (NBDC) for their help in improving the design of the wwPDB/RDF, and Michelle Ragsac for as-

sisting with SQL queries in Figure 2. Y. T. thanks Shigeru Endo for helping with the ProMode Elastic service.

## FUNDING

Database Integration Coordination Program from the National Bioscience Database Center (NBDC); Platform Project for Supporting in Drug Discovery and Life Science Research (Platform for Drug Discovery, Informatics, and Structural Life Science) from Japan Agency for Medical Research and Development (AMED) (in part); JSPS KAKENHI [26440078] (in part). Funding for open access charge: Database Integration Coordination Program from the National Bioscience Database Center (NBDC), JST.

*Conflict of interest statement.* None declared.

## REFERENCES

- Kinjo,A.R., Suzuki,H., Yamashita,R., Ikegawa,Y., Kudo,T., Igarashi,R., Kengaku,Y., Cho,H., Standley,D.M., Nakagawa,A. *et al.* (2012) Protein Data Bank Japan (PDBj): Maintaining a structural data archive and Resource Description Framework format. *Nucleic Acids Res.*, **40**, D453–D460.
- Berman,H., Henrick,K., Nakamura,H. and Markley,J.L. (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.*, **35**, D301–D303.
- Berman,H.M., Burley,S.K., Kleywegt,G.J., Markley,J.L., Nakamura,H. and Velankar,S. (2016) The archiving and dissemination of biological structure data. *Curr. Opin. Struct. Biol.*, **40**, 17–22.
- Velankar,S., van Ginkel,G., Alhroub,Y., Battle,G.M., Berrisford,J.M., Conroy,M.J., Dana,J.M., Gore,S.P., Gutmanas,A., Haslam,P. *et al.* (2016) PDBe: improved accessibility of macromolecular structure data from PDB and EMDB. *Nucleic Acids Res.*, **44**, D385–D395.
- Rose,P.W., Prlić,A., Bi,C., Bluhm,W.F., Christie,C.H., Dutta,S., Green,R.K., Goodsell,D.S., Westbrook,J.D., Woo,J. *et al.* (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.*, **43**, D345–D356.
- Ulrich,E.L., Akutsu,H., Doreleijers,J.F., Harano,Y., Ioannidis,Y.E., Lin,J., Livny,M., Mading,S., Maziuk,D., Miller,Z. *et al.* (2008) BioMagResBank. *Nucleic Acids Res.*, **36**, D402–D408.
- Kinoshita,K. and Nakamura,H. (2004) eF-site and PDBjViewer: database and viewer for protein functional sites. *Bioinformatics*, **20**, 1329–1330.
- Bekker,G.-J., Nakamura,H. and Kinjo,A.R. (2016) Molmil: a molecular viewer for the PDB and beyond. *J. Cheminform.*, **8**, 42.
- Morita,M., Igarashi,Y., Ito,M., Chen,Y.-A., Nagao,C., Sakaguchi,Y., Sakate,R., Masui,T. and Mizuguchi,K. (2012) Sagace: a web-based search engine for biomedical databases in Japan. *BMC Res. Notes*, **5**, 1–5.
- Cereto-Massagué,A., Ojeda,M.J., Valls,C., Mulero,M., Garcia-Vallvé,S. and Pujadas,G. (2015) Molecular fingerprint similarity search in virtual screening. *Methods*, **71**, 58–63.
- O'Boyle,N.M., Banck,M., James,C.A., Morley,C., Vandermeersch,T. and Hutchison,G.R. (2011) Open babel: an open chemical toolbox. *J. Cheminform.*, **3**, 1–14.
- Kinjo,A.R., Yamashita,R. and Nakamura,H. (2010) PDBj mine: design and implementation of relational database interface for Protein Data Bank Japan. *Database*, **2010**, baq021.
- Westbrook,J., Ito,N., Nakamura,H., Henrick,K. and Berman,H.M. (2005) PDBML: the representation of archival macromolecular structure data in XML. *Bioinformatics*, **21**, 988–992.
- Westbrook,J.D. and Bourne,P.E. (2000) STAR/mmCIF: an ontology for macromolecular structure. *Bioinformatics*, **16**, 159–168.
- Velankar,S., Dana,J.M., Jacobsen,J., van Ginkel,G., Gane,P.J., Luo,J., Oldfield,T.J., O'Donovan,C., Martin,M.J. and Kleywegt,G.J. (2013) SIFTS: structure integration with function, taxonomy and sequences resource. *Nucleic Acids Res.*, **41**, D483–D489.
- Finn,R.D., Bateman,A., Clements,J., Coghill,P., Eberhardt,R.Y., Eddy,S.R., Heger,A., Hetherington,K., Holm,L., Mistry,J. *et al.*



- (2014) The Pfam protein families database. *Nucleic Acids Res.*, **42**, D222–D230.
17. Katayama, T., Wilkinson, M.D., Micklem, G., Kawashima, S., Yamaguchi, A., Nakao, M., Yamamoto, Y., Okamoto, S., Oouchida, K., Chun, H.-W. *et al.* (2013) The 3rd DBCLS BioHackathon: improving life science data integration with Semantic Web technologies. *J. Biomed. Semantics*, **4**, 1–17.
18. Bairoch, A., Apweiler, R., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M. *et al.* (2005) The universal protein resource (UniProt). *Nucleic Acids Res.*, **33**, D154–D159.
19. Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A. *et al.* (2016) PubChem Substance and Compound databases. *Nucleic Acids Res.*, **44**, D1202–D1213.
20. Yokochi, M., Kobayashi, N., Ulrich, E.L., Kinjo, A.R., Iwata, T., Ioannidis, Y.E., Livny, M., Markley, J.L., Nakamura, H., Kojima, C. *et al.* (2016) Publication of nuclear magnetic resonance experimental data with semantic web technology and the application thereof to biomedical research of proteins. *J. Biomed. Semantics*, **7**, 16.
21. Wimalaratne, S.M., Bolleman, J., Juty, N., Katayama, T., Dumontier, M., Redaschi, N., Le Novère, N., Hermjakob, H. and Laibe, C. (2015) SPARQL-enabled identifier conversion with Identifiers.org. *Bioinformatics*, **31**, 1875–1877.
22. Bolleman, J.T., Mungall, C.J., Strozzi, F., Baran, J., Dumontier, M., Bonnal, R. P.J., Buels, R., Hoehndorf, R., Fujisawa, T., Katayama, T. *et al.* (2016) FALDO: a semantic standard for describing the location of nucleotide and protein feature annotation. *J. Biomed. Semantics*, **7**, 1–12.
23. Suzuki, H., Kawabata, T. and Nakamura, H. (2016) Omokage search: shape similarity search service for biomolecular structures in both the PDB and EMD. *Bioinformatics*, **32**, 619–620.
24. Kawabata, T. (2008) Multiple subunit fitting into a low-resolution density map of a macromolecular complex using a gaussian mixture model. *Biophys. J.*, **95**, 4643–4658.
25. Schiebel, J., Krimmer, S.G., Röwer, K., Knörlein, A., Wang, X., Park, A.Y., Stieler, M., Ehrmann, F.R., Fu, K., Radeva, N. *et al.* (2016) High-throughput crystallography: reliable and efficient identification of fragment hits. *Structure*, **24**, 1398–1409.
26. Kawabata, T. (2016) HOMCOS: an updated server to search and model complex 3D structures. *J. Struct. Funct. Genomics*, doi:10.1007/s10969-016-9208-y.
27. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.L. (1997) Gapped Blast and PSI-Blast: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
28. Kawabata, T. (2011) Build-up algorithm for atomic correspondence between chemical structures. *J. Chem. Info. Model.*, **51**, 1775–1787.
29. Kawabata, T., Sugihara, Y., Fukunishi, Y. and Nakamura, H. (2013) LigandBox: a database for 3D structures of chemical compounds. *Biophysics*, **9**, 113–121.
30. Kawabata, T. and Nakamura, H. (2014) 3D flexible alignment using 2D maximum common substructure: dependence of prediction accuracy on target-reference chemical similarity. *J. Chem. Info. Model.*, **54**, 1850–1863.
31. Wako, H., Kato, M. and Endo, S. (2004) ProMode: a database of normal mode analyses on protein molecules with a full-atom model. *Bioinformatics*, **20**, 2035–2043.
32. Wako, H. and Endo, S. (2013) Normal mode analysis based on an elastic network model for biomolecules in the Protein Data Bank, which uses dihedral angles as independent variables. *Comput. Biol. Chem.*, **44**, 22–30.
33. Tsuchiya, Y., Kinoshita, K., Endo, S. and Wako, H. (2012) Dynamic features of homodimer interfaces calculated by normal-mode analysis. *Protein Sci.*, **21**, 1503–1513.
34. Tirion, M.M. (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.*, **77**, 1905–1908.
35. Standley, D.M., Toh, H. and Nakamura, H. (2007) ASH structure alignment package: sensitivity and selectivity in domain classification. *BMC Bioinformatics*, **8**, 116.
36. Standley, D.M., Yamashita, R., Kinjo, A.R., Toh, H. and Nakamura, H. (2010) SeSAW: balancing sequence and structural information in protein functional mapping. *Bioinformatics*, **26**, 1258–1259.
37. Kinjo, A.R. and Nakamura, H. (2007) Similarity search for local protein structures at atomic resolution by exploiting a database management system. *Biophysics*, **3**, 75–84.
38. Standley, D.M., Toh, H. and Nakamura, H. (2005) GASH: an improved algorithm for maximizing the number of equivalent residues between two protein structures. *BMC Bioinformatics*, **6**, 221.
39. Katoh, K. and Toh, H. (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.*, **9**, 286–298.
40. Kinjo, A.R. and Nishikawa, K. (2005) Predicting secondary structures, contact numbers, and residue-wise contact orders of native protein structure from amino acid sequence using critical random networks. *Biophysics*, **1**, 67–74.
41. Kinjo, A.R. and Nishikawa, K. (2006) CRNPRED: highly accurate prediction of one-dimensional protein structures by large-scale critical random networks. *BMC Bioinformatics*, **7**, 401.
42. Lis, M., Kim, T., Sarmiento, J., Kuroda, D., Dinh, H., Kinjo, A.R., Amada, K., Devadas, S., Nakamura, H. and Standley, D.M. (2011) Bridging the gap between single-template and fragment based protein structure modeling using Spanner. *Immunome Res.*, **7**, 1–8.
43. Kinoshita, K. and Nakamura, H. (2003) Identification of protein biochemical functions by similarity search using the molecular surface database eF-site. *Protein Sci.*, **12**, 1589–1595.
44. Murakami, Y., Kinoshita, K., Kinjo, A.R. and Nakamura, H. (2013) Exhaustive comparison and classification of ligand-binding surfaces in proteins. *Protein Sci.*, **22**, 1379–1391.
45. Goodsell, D.S., Dutta, S., Zardecki, C., Voigt, M., Berman, H.M. and Burley, S.K. (2015) The RCSB PDB 'Molecule of the Month': Inspiring a Molecular View of Biology. *PLoS Biol.*, **13**, e1002140.
46. Standley, D.M., Toh, H. and Nakamura, H. (2004) Detecting local structural similarity in proteins by maximizing number of equivalent residues. *Proteins*, **57**, 381–391.
47. Kinjo, A.R. and Nakamura, H. (2009) Comprehensive structural classification of ligand binding motifs in proteins. *Structure*, **17**, 234–246.
48. Kinjo, A.R. and Nakamura, H. (2010) Geometric similarities of protein-protein interfaces at atomic resolution are only observed within homologous families: an exhaustive structural classification study. *J. Mol. Biol.*, **399**, 526–540.
49. Kinjo, A.R. and Nakamura, H. (2012) GIRAF: a method for fast search and flexible alignment of ligand binding interfaces in proteins at atomic resolution. *Biophysics*, **8**, 79–94.