

日本と世界の蛋白質構造データバンク: PDBjとwwPDB

**Protein Data Bank Japan (PDBj) and
the Worldwide PDB (wwPDB)**

Haruki Nakamura

Inst. Protein Research, Osaka Univ.





Protein Structure Databank: PDB

PDB History

1960s: Protein crystallography started

- Myoglobin, Hemoglobin,
Lysozyme, etc

1970s □ PDB started with 7 data

(October, 1971) at Brookhaven

- National Laboratory, USA
- PR-Osaka Univ. determined
cytochrome c and deposited it to PDB
- Magnetic tapes were distributed



Max Perutz & □
□ John Kendrew (Nobel
Prize in Chemistry, 1962) □

CRYSTALLOGRAPHY

Protein Data Bank

A repository system for protein crystallographic data will be operated jointly by the Crystallographic Data Centre, Cambridge, and the Brookhaven National Laboratory. The system will be responsible for storing atomic coordinates, structure factors and electron density maps and will make these data available on request. Distribution will be on magnetic tape in machine-readable form whenever possible. There will be no charge for the service other than handling costs. Files will be updated as new material is received. The total holding will be

Protein Structure Databank: PDB

1980s: Rapid data increase (IUCr recommendation

“Data deposition to PDB is mandatory for paper submission to journals”)

1990s: RCSB-PDB started in USA

2000s: Foundation of wWPDB (worldwide PDB)

Further many data by structural genomics

2010s: New methodologies are applied

More than 100,000 data

Towards “Big Data”

The logo for PDBj (Protein Data Bank Japan) features the letters "PDBj" in a large, bold, blue serif font. The letter "j" has a red dot at its top right corner. Below the letters, the text "Protein Data Bank Japan" is written in a smaller, dark blue sans-serif font.

Protein Data Bank Japan

http://pdbj.org/

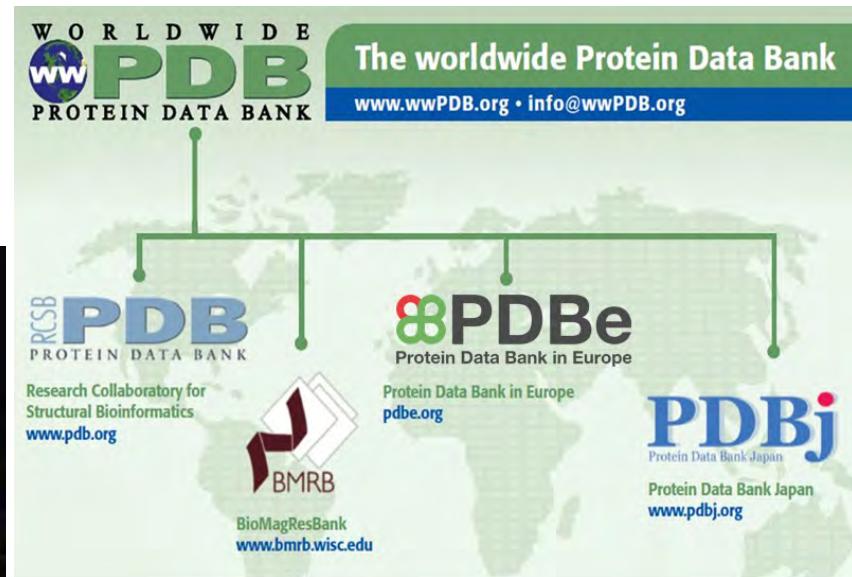
Since 2001, PDBj has been managed at [Institute for Protein Research, Osaka University](#) as a member of the [wwPDB](#), to curate and process the deposited data for an open and single archive.



wwPDB.org



PDBj staffs (April 2015)

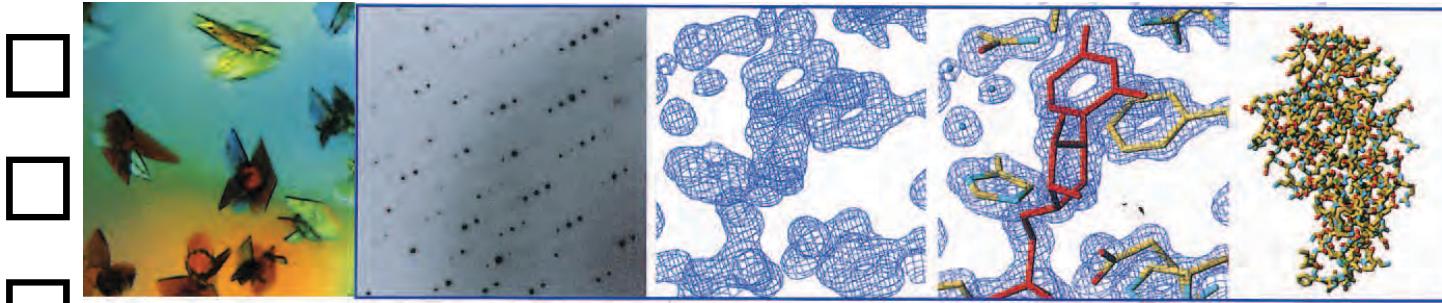


wwPDB members and their heads



Protein Structure Databank: PDB

- X-ray Crystallography

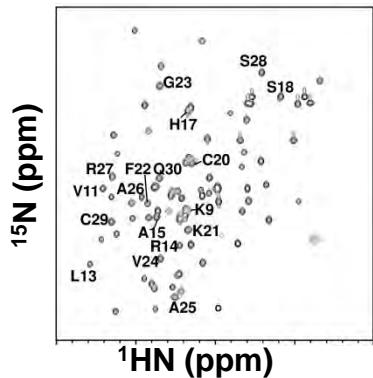


□ Crystal □ X-ray diffraction □ Electron density map □ Atomic model □

SPring-8
synchrotron □



- Nuclear Magnetic Resonance (NMR)



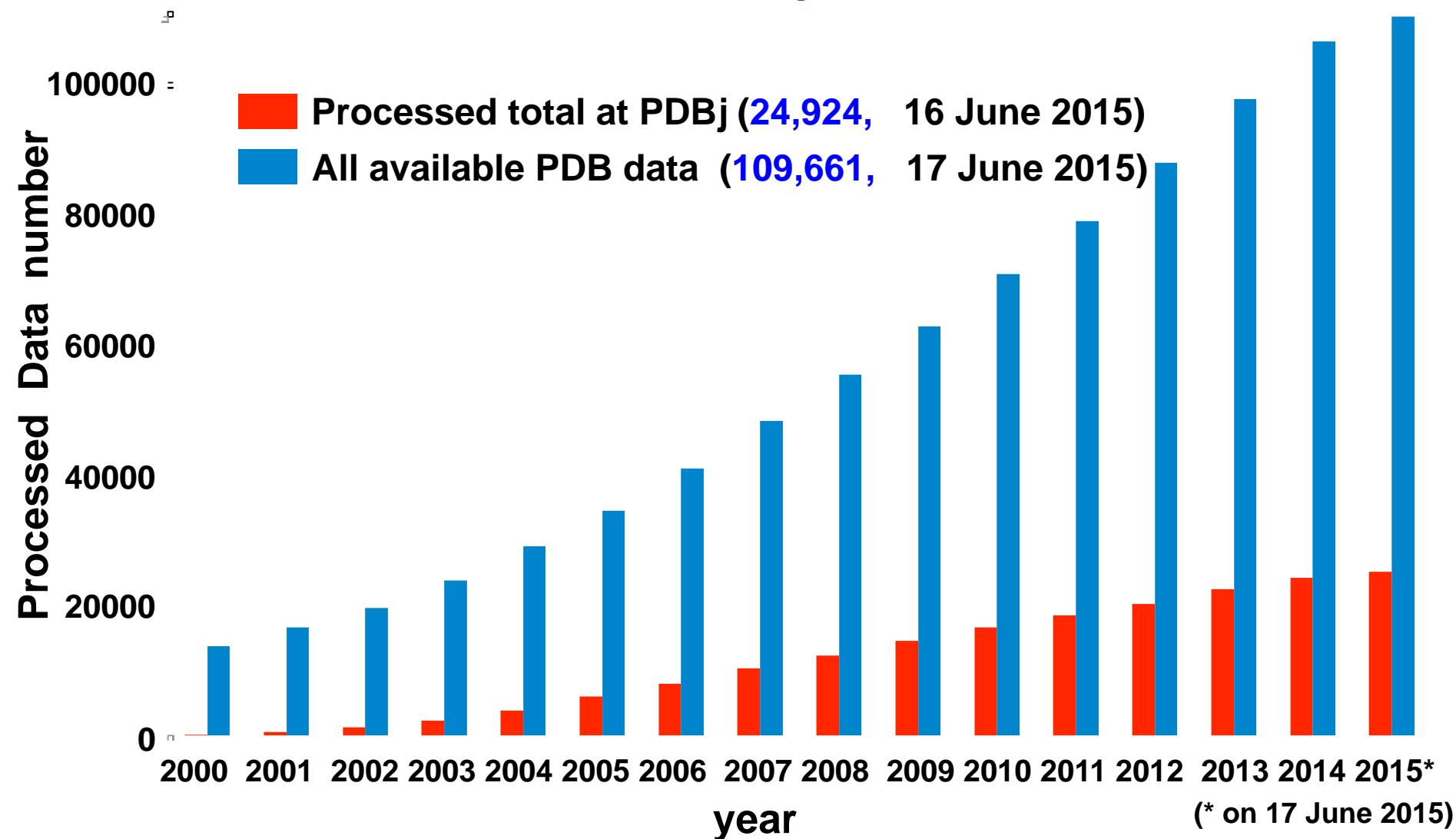
950 MHz NMR
spectrometer
with super-
conducting
magnet □

- Cryo-Electron Microscopy □

Activities/Services of each member of the wwPDB

- “**Data-in**” activity, common in all the wwPDB members with high quality control. For that purpose, new format, data deposition, and validation system are developed
- “**Data-out**” services, common archive as the ftp site and the characteristic services by each wwPDB member

Data-in at PDBj and wwPDB



PDBj curates and processes about a Quarter of the deposited data, mainly from Asian and Oceania regions

Activities/Services of each member of the wwPDB

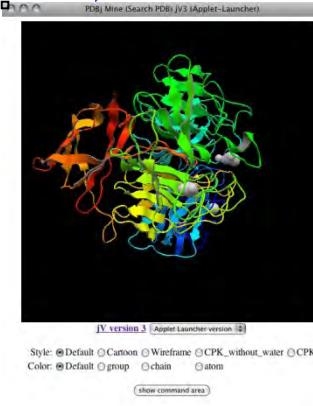
- “Data-in” activity, common in all the wwPDB members with high quality control. For that purpose, new format, data deposition, and validation system are developed
- “**Data-out**” services, common archive as the ftp site and the characteristic services by each wwPDB member

Data-out from PDB

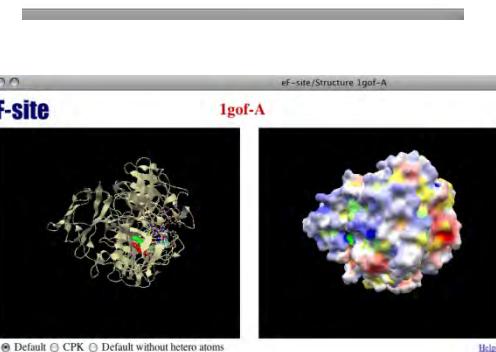
<http://pdbj.org/>

Data viewer at PDBj

Graphic viewer: *jV* and *Molmil* <http://pdbj.org/jV/>



Amino acid sequence (FASTA)



Functional site	locus & details	AA-495	AA-T	AA-TL
1) A-495	(on) (off)	Y		
2) A-272	(on) (off)		Proton acceptor	
3) A-495	(on) (off)			source
4) A-496	(on) (off)			Swiss-Prot : 1
5) A-581	(on) (off)			
6) A-238	(on) (off)			
7) A-390	(on) (off)			
8) A-272	(on) (off)			
9) A-495	(on) (off)			
10) A-75-87	(on) (off)			
11) A-194	(on) (off)			
12) A-227-228	(on) (off)			
13) A-177	(on) (off)			

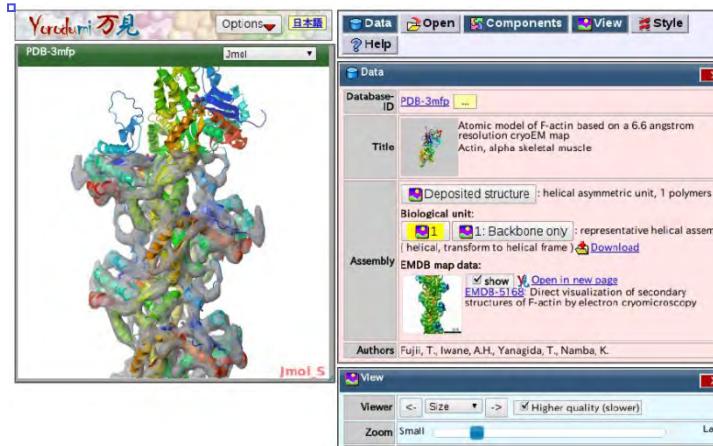
Molecular surface DB: eF-site

Kinjo et al. NAR 40, D453 (2012)

PDBj services



EM Navi: EMDB browser



Yorodumi: PDB & EMDB

GIRAF query upload

Interface type

- Nonpolymer binding
- DNA and RNA binding
- All types of ligands (nonpolymer, DNA, RNA, peptides, and others).
- PPI (protein-protein interfaces).
- All (ligands + PPI).

Input PDB ID:

or upload a PDB file:

Chain IDs (optional): (comma-separated multiple IDs [e.g., "A,B"] or "all" are allowed.)

Limit target PDB entries (optional): (comma-separated multiple IDs [e.g., "101m,1a00"] or "all" are allowed.)

Number of displayed results (optional): [100]

Your email address (optional):

DB version: 2013-09-21: 713107 interfaces

GIRAF: Similar binding site



ProMode: Normal Mode



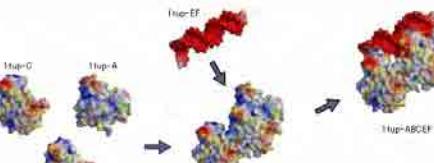
[About eF-site](#) | [References](#) | [Links](#) | [Acknowledgements](#) | [Feedback](#)

175033 Entries. Last Update: 20-Aug-2005

Keyword Search
 PDB code only and or
 Search Reset

- Category Search
- Antibody
 - Protein
 - Active Site
 - Membrane
 - Binding Site

Examples of molecular surface



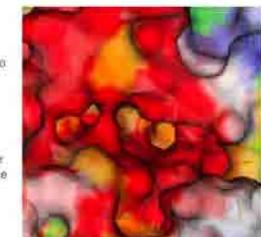
eF-site: Mol. Surface DB

[TOP](#) [Help](#) [FAQ](#) [References](#) [Links](#)



ABOUT eF-seek:

Molecular function of proteins are determined by their three dimensional structures, thus the similarity of protein structure can give some clues to infer their functions. In many cases, the molecular functions are begun with the molecular interaction with small molecules (ligands). eF-seek is a web server to search for the similar ligand binding sites for the uploaded coordinate file with PDB format. The representative binding sites in eF-site database are search by our own algorithm based on the clique search algorithm.



Submission STEP-1:

Specify a PDB format file:

E-mail address:

Keyword: #

Title: (optional)

eF-seek: Similar surface

Integration with EMDB: Search for similar SHAPE

Query: human RNA polymerase II with RNA (EMDB: 2190)

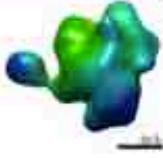
Similar shapes from 195,658 images

 Omokage search - Shape similarity search of macromolecules - (English) 日本語

Search query

Subject structure

Database: EMDB / ID: 2190
human RNA polymerase II in complex with AluRA RNA
[Quick](#), [Yorodumi](#), [EM Navigator](#)



Search result

Showing 1 - 100 of 2000 structures found from all (195658 structures)

Pages: [1](#) [2](#) [3](#) [4](#) [10](#) [20](#) [Previous](#) [Next](#)

Display: [images only](#) [as list](#)

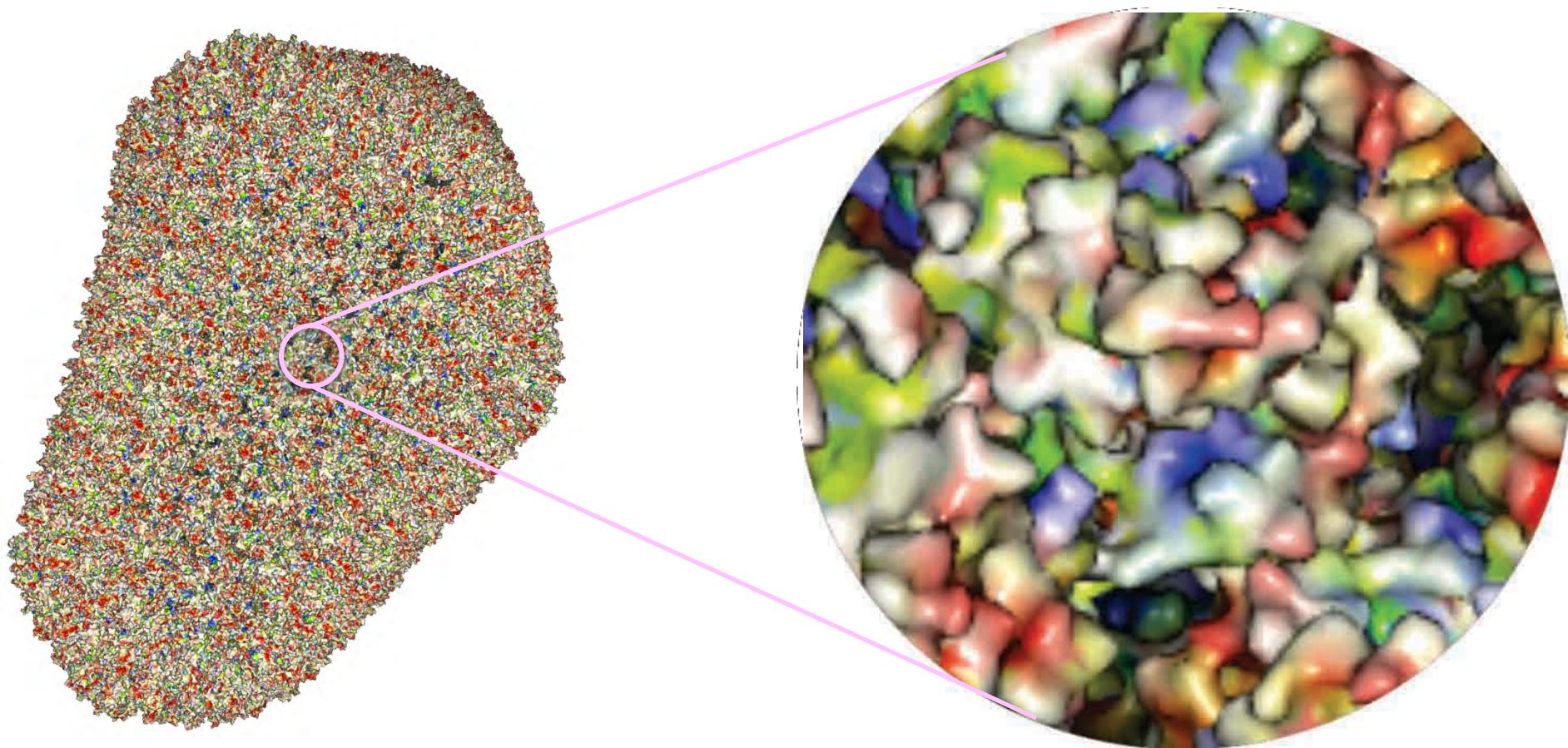
Molmil: Molecular Viewer based on WebGL



- Based on JavaScript/WebGL.
- Supports PDB, mmCIF, PDBML formats.
- Links to PDB, chem_comp (Compound), and ProMode Elastic.
- Outputs screenshots.
- Available for iOS8 (iPAD, iPhone etc)

Molmil Viewer for eF-site

even for large structures (coming soon)



HIV-1 Capsid 3J3Q, 1356 chains, >2M atoms

Task force of wwPDB for the hybrid approach was held in EBI on 6-7 October, where Iwasaki and Nakamura attended.

Data bank struggles as protein imaging ups its game

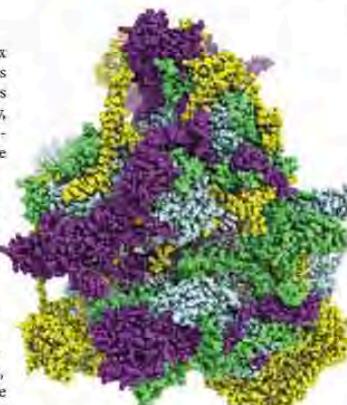
Hybrid methods to solve structures of molecular machines create a storage headache.

BY EWEN CALLAWAY

Structural biology, the mapping of complex biological molecules such as proteins, is in the grip of a revolution. The field has long been dominated by X-ray crystallography, a technique made iconic by its role in decoding the DNA double helix in the 1950s. But the need to tackle more complex structures and to watch 'molecular machines' function in real time is fuelling a shift towards hybrid imaging methods that can create moving models.

That is posing a challenge for the world's official repository for protein structures: the Protein Data Bank (PDB), which relies almost exclusively on crystallography data and lacks the standards and software infrastructure to archive structures described by hybrid methods. This month, leaders of the four organizations around the world that host the data bank held a workshop in Hinxton, UK, to hatch a plan to ensure that hybrid models and their insights into fundamental biology and disease do not get lost.

Historically, structural biology has focused on generating three-dimensional (3D) descriptions of individual proteins. In many cases, this is a task perfect for crystallography, in which



A subunit of a ribosome, a molecular machine.

crystallography. "These days, being a crystallographer is not good enough," says Gerard Kleywegt, a structural biologist at the European Bioinformatics Institute in Hinxton, who heads the European annex of the PDB.

that alter the proteasome's activity. This year, another team published a hybrid model of the key HIV proteins that sneak the virus into a cell, which may help in vaccine design (M. Pancera *et al. Nature* <http://doi.org/wfz>; 2014).

The hybrid approach has also tackled the ribosome, which produces proteins; the nuclear pore complex, which provides a gateway between the genome in the nucleus and the rest of the cell; and the molecular syringes made by bacteria that inject proteins into cells. Models of many more molecular machines are expected. "We're going to enter a period of exponential growth in the generation of these hybrid structures," says Stephen Burley, a structural biologist at Rutgers University in Piscataway, New Jersey, who heads one of the two US annexes of the PDB.

At the PDB workshop, on 6–7 October, Kleywegt, Burley and three dozen others hashed out the challenges that these triumphs are creating for the PDB. Crystallography yields a standardized set of data files in which a structure and its level of precision are self-evident; by contrast, the underlying data for the hybrid models exist in a mishmash of formats such as X-ray diffraction patterns or electron-micrograph pictures. And going from raw data to a



Task force of wwPDB at EBI

Nature (2014) 514, pp.416

“Ultimately, we want to have the molecular structure of an entire cell. That’s still science fiction at the moment – but it’s somewhere we can get to in 10, 20 years” (by Jan Ellenberg, EMBL)

The wwPDB Symposium: Integrative Structural Biology with Hybrid Methods

Date and Time: 9:00 AM to 18:00 PM on 3 October (Saturday) 2015

Place: Osaka University Hall, Auditorium, Toyonaka, Osaka, Japan

Organizer: The wwPDB Foundation

- 9:00 Opening Remark **Haruki Nakamura** (PDBj, Osaka Univ)
- 9:10 **Helen M. Berman** (wwPDB, Rutgers Univ)
- 9:40 **Andrej Sali** (UCSF)
- 10:10 **Wah Chiu** (Baylor College of Medicine)
- 10:40 - 11:00 Coffee Break
- 11:00 **Takashi Fujii** (Osaka Univ)
- 11:30 **Helen Saibil** (Birkbeck College)
- 12:00 **Kenji Iwasaki** (Osaka Univ)
- 12:30 - 14:00 Lunch
- 14:00 **Angela Gronenborn** (Univ Pittsburgh)
- 14:30 **Florence Tama** (RIKEN)
- 15:00 **Takeshi Kawabata** (Osaka Univ)
- 15:30 - 15:50 Coffee Break
- 15:50 **Mitsunori Ikeguchi** (Yokohama City Univ)
- 16:20 **Paul Adams** (Lawrence Berkeley Laboratory)
- 16:50 **Gerard Kleywegt** (PDBe, EBI)
- 17:20 **R. Andrew Byrd** (NCI at Frederick)
- 17:50 Closing Remark **Stephen K Burley** (RCSB-PDB, Rutgers Univ)

Data Formats of PDB data

- **PDB (conventional and flat)**
- **PDB Exchange (mmCIF)**
 - Mechanism for extension based on new demands
- **PDBML**
 - Derived from mmCIF
 - All entries converted to XML
 - Automatic translation from mmCIF data files and dictionaries
 - 3-styles of translation released

*(Westbrook, Ito, Nakamura, Henrick, Berman (2005)
Bioinformatics, 21, 988-992)*

New standard PDB format: PDBx/mmCIF

- Current PDB format is almost **40 years old** and does not support today's science.
- PDB Record format limitations
 - **Max. 62 chains**
 - **Max. 99,999 atoms**
 - No bond orders or chirality specified for ligands
 - No support for NMR, EM, hybrid methods, ...
 - Meta-data specification cumbersome and inflexible



- **Preserve backward compatibility where possible**
- **PDBML (XML) and RDF format files are available.**
- **Start in 2014 and the current PDB format will be phased out in 2016.**

```

ATOM      1   N    GLN A  39       24.690  -27.754  24.275  1.00  60.76          N
ATOM      2   CA   GLN A  39       23.581  -26.768  24.416  1.00  60.98          C
ATOM      3   C    GLN A  39       23.990  -25.379  23.905  1.00  59.98          C
ATOM      4   O    GLN A  39       25.070  -25.209  23.330  1.00  60.25          O
ATOM      5   CB   GLN A  39       23.136  -26.685  25.878  1.00  60.69          C
ATOM      6   N    VAL A  40       23.115  -24.395  24.122  1.00  59.58          N
ATOM      7   CA   VAL A  40       23.342  -23.010  23.690  1.00  57.26          C
ATOM      8   C    VAL A  40       24.000  -22.152  24.778  1.00  56.00          C
ATOM      9   O    VAL A  40       23.992  -20.920  24.692  1.00  55.53          O
ATOM     10   CB   VAL A  40       22.015  -22.337  23.275  1.00  57.32          C

```

PDB

```

loop_
_atom_site.group_PDB
_atom_site.id
_atom_site.auth_atom_id
_atom_site.type_symbol
_atom_site.auth_comp_id
_atom_site.auth_asym_id
_atom_site.auth_seq_id
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.pdbx_PDB_model_num
_atom_site.occupancy
_atom_site.pdbx_auth_alt_id
_atom_site.B_iso_or_equiv

```

PDBx/mmCIF

ATOM	1	N	N	GLN	A	39	24.690	-27.754	24.275	1	1.000	.	60.760
ATOM	2	CA	C	GLN	A	39	23.581	-26.768	24.416	1	1.000	.	60.980
ATOM	3	C	C	GLN	A	39	23.990	-25.379	23.905	1	1.000	.	59.980
ATOM	4	O	O	GLN	A	39	25.070	-25.209	23.330	1	1.000	.	60.250
ATOM	5	CB	C	GLN	A	39	23.136	-26.685	25.878	1	1.000	.	60.690
ATOM	6	N	N	VAL	A	40	23.115	-24.395	24.122	1	1.000	.	59.580
ATOM	7	CA	C	VAL	A	40	23.342	-23.010	23.690	1	1.000	.	57.260
ATOM	8	C	C	VAL	A	40	24.000	-22.152	24.778	1	1.000	.	56.000
ATOM	9	O	O	VAL	A	40	23.992	-20.920	24.692	1	1.000	.	55.530
ATOM	10	CB	C	VAL	A	40	22.015	-22.337	23.275	1	1.000	.	57.320
ATOM	11	N	N	ALA	A	41	24.560	-22.804	25.797	1	1.000	.	54.570

wwPDB Service site for a new format

<http://mmcif.wwpdb.org/> or <http://mmcif.pdbj.org/>



The screenshot shows the homepage of the PDBx/mmCIF Dictionary Resources. At the top, there is a navigation bar with links for "Home", "Dictionaries", "Documentation", "Downloads", and "Contact Us". To the right of the navigation bar is the "wwPDB" logo. Below the navigation bar, the main title "PDBx/mmCIF Dictionary Resources" is displayed in large, bold, blue text. Underneath the title, a descriptive text states: "This site provides information about the format, dictionaries and related software tools used by the Worldwide Protein Data Bank (wwPDB) to define data content for deposition, annotation and archiving of PDB entries." A green button labeled "Browse the current dictionary »" is located below the descriptive text. The background of the page has a light gray gradient.

Dictionaries

- [Browse the current dictionary»](#)
- [Download/view all dictionaries »](#)
- [Search dictionaries»](#)

Documentation

- [PDB > PDBx/mmCIF correspondences »](#)
- [PDBx/mmCIF for large structures »](#)
- [Software resources »](#)
- [C++ » and Python » programming examples](#)
- [File syntax » and dictionary organization »](#)
- [Atomic » and molecular » descriptions](#)
- [References »](#)
- [Glossary »](#)

FAQs

Questions about PDBx/mmCIF format, and data content, or software tools?
Check out the [FAQ»](#)

PDBx/mmCIF Software Support

- **Phenix and Refmac** – produce native PDBx files for deposition
- **MMDB** - macromolecular object library in CCP4
- **iotbx.cif/ucif** - CCTBx C++/Python IO library with dictionary validation
- **CCIF** – CCP4 C++ library with FORTRAN support and dictionary validation
- **CBFLib** - ANSI-C library for CIF & imgCIF files
- **mmLIB** - Python toolkit supporting CIF & mmCIF
- **BioPython** - Python toolkit for computational biology
- **PyCifRW** - Python CIF/mmCIF parsing tools
- **BioJava** - Java mmCIF IO package
- **STAR::Parser** – Perl mmCIF parser and molecular object library
- **RCSBTools** - C++/Python parsing and dictionary validation tools plus many other supporting format conversion and data management applications
- **Visualization** - **UCSF Chimera, Jmol, OpenRasMol, Coot, CCP4mg, jV, Molmil**

wwPDB Service site for a new format

<http://mmcif.pdbj.org/converter/index.php?l=en>

PDBx/mmCIF

[Home](#)[Dictionaries](#)[Documentation](#)[Downloads](#)[Format Conversion](#)[Contact Us](#)[English](#) 日本語

PDB format - PDBx/mmCIF conversion service

You can convert a molecular structural data into another format. The type of uploaded file is determined automatically. When the type is mmCIF and PDB format, it is converted into PDB format and mmCIF, respectively. The gzip compressed files that end ".gz" of the name, are also available. When they are gzipped, the converted files are also gzipped.

1. Specify a source file to convert

Specify a source file to convert. The maximum size of the file is 1GB.

選択... ファイルが選択されていません。

2. Confirm the contents of operations

3. Execute conversion & Download the converted file

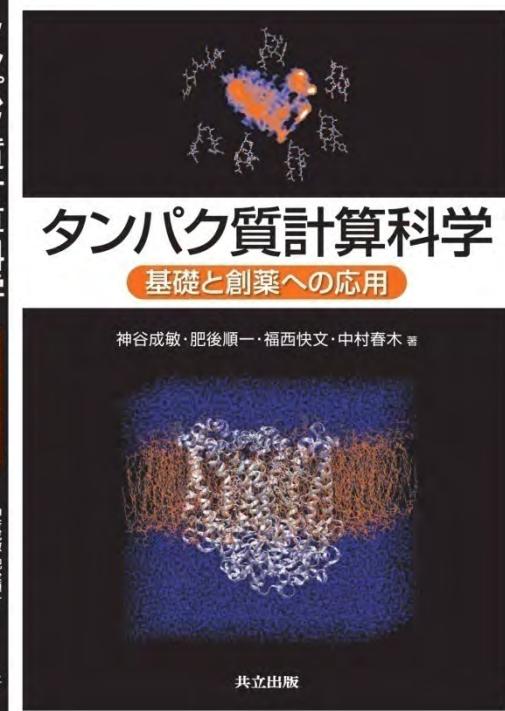
When you convert a large structure mmCIF file into the PDB format, It will be treated as following:

- When it includes more than 99999 atoms, all the atomids larger than 99999 are rewrited to 99999.
- When the chain id (auth_asym_id) has two letters, it will be described as it is by using unused 21th column and defined 22th column.

構造生命科学： 生体高分子の構造を基に進める生命科学

神谷・肥後・福西・中村 著

タンパク質計算科学



(共立出版2009年8月初版)

中村 編

見てわかる 構造生命科学

見てわかる 構造生命科学

▶生命科学研究へのタンパク質構造の利用◀

中村春木 編



化学同人



化学同人

(化学同人2014年4月初版)

工藤・西川・中村 訳

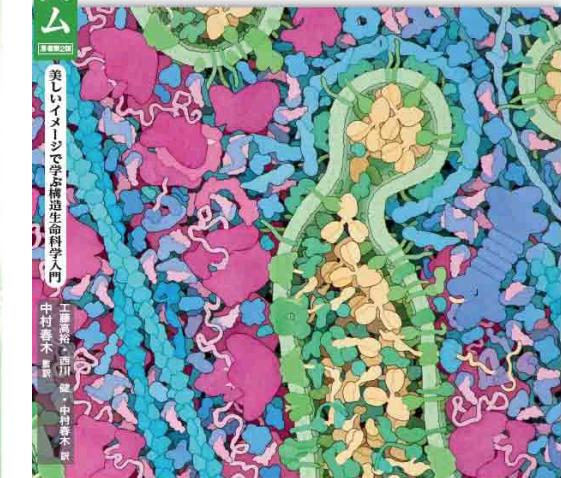
生命のメカニズム
THE MACHINERY OF LIFE

原著第2版

生命のメカニズム

美しいイメージで学ぶ構造生命科学入門

David S. Goodsell ■
中村春木 訳
工藤高裕・西川健・中村春木 訳



(シナジー2015年2月初版)

PDBj Staffs

- 総括責任者

- 中村 春木  (大阪大学蛋白質研究所・教授)

- PDBjデータベース構築グループ

- 中川 敦史  (大阪大学蛋白質研究所・教授)
- 五十嵐 令子 (大阪大学蛋白質研究所・特任研究員)
- 見学 有美子 (大阪大学蛋白質研究所・特任研究員)
- 張 羽澄 (大阪大学蛋白質研究所・特任研究員)
- 池川 恭代 (大阪大学蛋白質研究所・特任研究員)
- 佐藤 純子 (大阪大学蛋白質研究所・特任研究員)

- PDBj国際的な運営高度化グループ

- 金城 琳  (大阪大学蛋白質研究所・准教授)
- 岩崎 煉治 (大阪大学蛋白質研究所・准教授)
- 鈴木 博文 (大阪大学蛋白質研究所・特任研究員)
- 山下 鈴子 (大阪大学蛋白質研究所・特任技術専門職員)
- 工藤 高裕 (大阪大学蛋白質研究所・特任研究員)
- Bekker, Gert-Jan (大阪大学蛋白質研究所・特任研究員)

- BMRBデータベース管理運営グループ

- 藤原 敏道  (大阪大学蛋白質研究所・教授)
- 久保田 秀雄 (大阪大学蛋白質研究所・客員教授)
- 芊嶋 長次郎 (大阪大学蛋白質研究所・准教授)
- 小林 直宏 (大阪大学蛋白質研究所・特任研究員)
- 岩田 武史 (大阪大学蛋白質研究所・特任研究員)
- 横地 政志 (大阪大学蛋白質研究所・特任研究員)

- 研究協力者

- 鮎湖 博  (早稲田大学社会科学総合学術院・教授) for ProMode
- 矢溝 広  (北里大学理学部・准教授) for ProMode
- 伊藤 幹鵬 (東京医科歯科大学大学院・教授)
- 木下 駿吾  (東北大学大学院情報科学研究科・教授) for eF-site
- Standley, M.Daron  (京都大学ウイルス研究所・教授) for SeqNavi, StructNavi, SeSAW and ASH
- 加藤 和貴 (大阪大学免疫学フロンティア研究センター・准教授) for MAFFTash

H. Nakamura

A. Nakagawa
R. Igarashi
Y. Kengaku
H. Cho
Y. Ikegawa
J. Sato

A. R. Kinjo
K. Iwasaki
H. Suzuki
R. Yamashita
T. Kudou
B. Gert-Jan

T. Fujiwara
H. Akutsu
C. Kojima
N. Kobayashi
T. Iwata
M. Yokochi

H. Wako
S. Endo
N. Ito
K. Kinoshita
K. Katoh

- 事務職員

- 晴氣 菜穂子 (大阪大学蛋白質研究所・特任事務職員)

N. Haruki