

# PDBj (PDB Japan) の活動と wwPDB

## Activity of PDBj and wwPDB

中村 春木

Haruki Nakamura



大阪大学蛋白質研究所

Institute for Protein Research

<http://pdbj.org/>

<http://wwpdb.org/>

# wwPDB

**WORLDWIDE PDB**  
PROTEIN DATA BANK

**The worldwide Protein Data Bank**  
[www.wwPDB.org](http://www.wwPDB.org) • [info@wwPDB.org](mailto:info@wwPDB.org)

**RCSB PDB**  
PROTEIN DATA BANK  
Research Collaboratory for Structural Bioinformatics  
[www.pdb.org](http://www.pdb.org)

**EMBL-EBI PDB<sub>e</sub>**  
PROTEIN DATA BANK EUROPE  
Protein Data Bank in Europe  
[pdbe.org](http://pdbe.org)

**PDBj**  
Protein Data Bank Japan  
Protein Data Bank Japan  
[www.pdbj.org](http://www.pdbj.org)

**BMRB**  
BioMagResBank  
[www.bmrwisc.edu](http://www.bmrwisc.edu)

**NMR実験情報DB**

Kleywegt, G   Markley, JL  
Berman, HM   Nakamura, H

**wwPDB  
Advisory Committee  
on Sept 27, 2013  
Rutgers Univ.**



# PDBj @ 大阪大学蛋白質研究所

## Protein Data Bank Japan

日本蛋白質構造データバンク

<http://pdbj.org/>

2001年度から、(独)科学技術  
振興機構(JST-NBDC:バイオ  
サイエンスデータベースセンター)の支援  
を受けて活動中

PDBj トップページ(日英中韓)



PDBj スタッフ(2013年4月)

## PDBjの活動

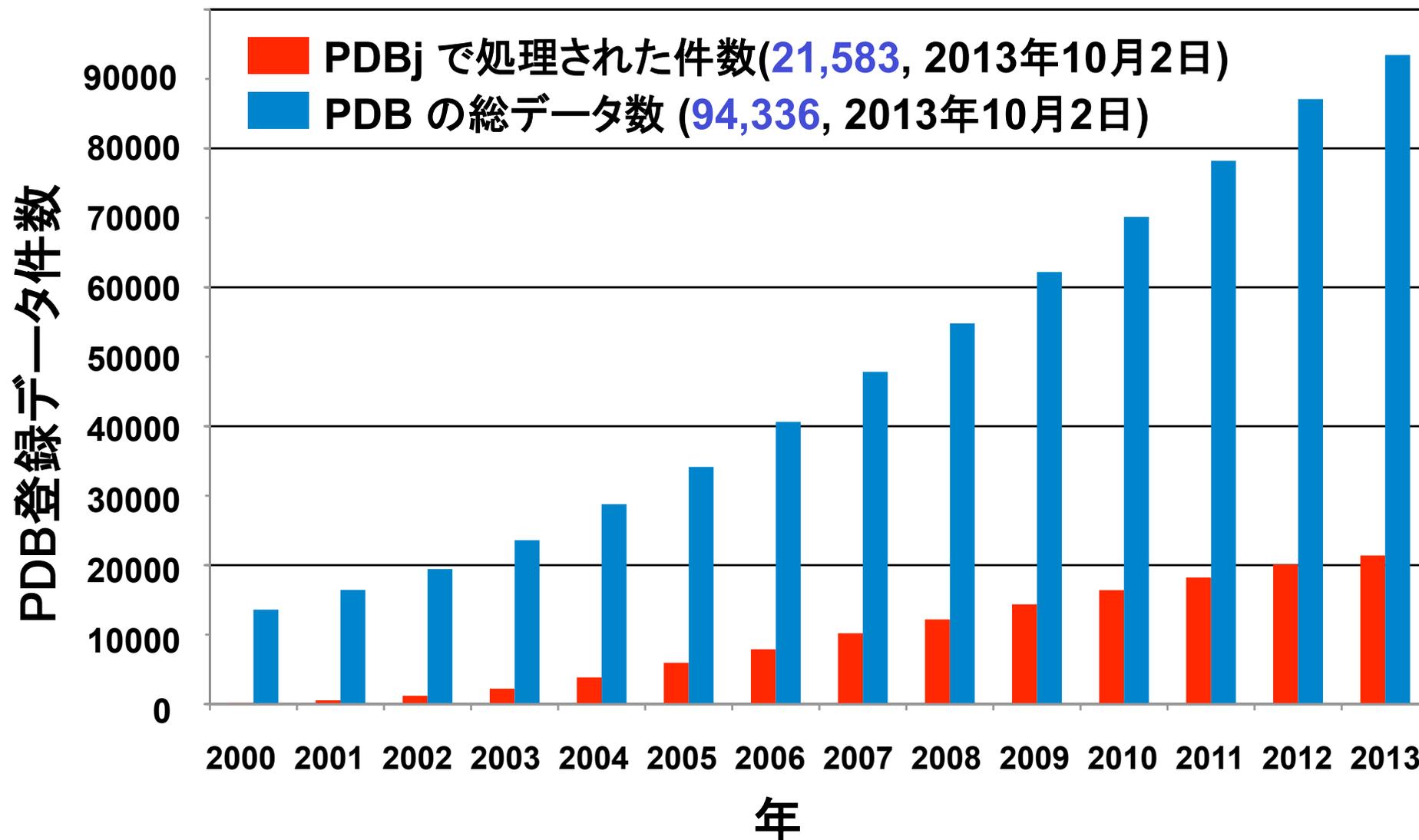
1) “Data-in” の活動: wwPDBの一員として

- ・品質管理がなされた登録作業の実施
- ・新たな統一的登録システムの開発
- ・新たな共通フォーマットの開発(XML, RDF)

2) “Data-out” の活動: 共通データの

- ・ダウンロードサイト(毎週アップデート)の運営
- ・種々のサービスや二次データベースの提供
- ・データベース統合化環境の開発・提供
- ・上記環境での配列→構造→機能の同定・推定

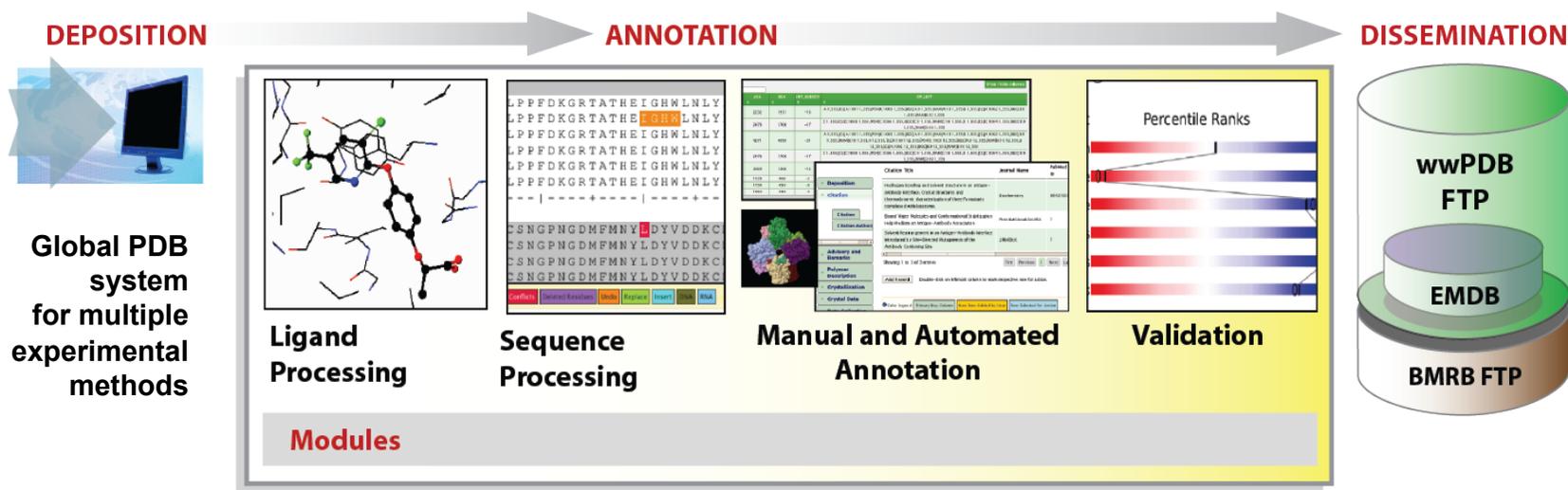
# PDBjのData-inの活動



世界中で決定された構造の約1/4の登録処理をPDBjで実施

# 新しいデータ処理システムの開発

## データ増大・高品質化に対処

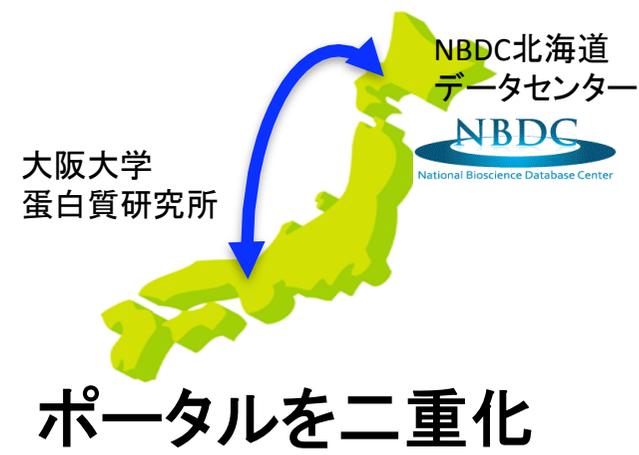


- Enables workload balancing and increased **productivity**
- Better quality assurance of **ligand** chemistry and **polymer sequences**
- **PDBx/mmCIF** is the master file format
- **Validation suites** based on recommendations from expert task forces; X-ray validation pipeline is available as a stand-alone server
- System will support **all accepted experimental methods**

# PDBjのData-out活動

## <http://pdbj.org/>

The screenshot shows the PDBj website interface. At the top left, it displays '93970' items available as of 2013-09-18. The main navigation menu includes Home, Data Registration, Search, Services & Software, Secondary Databases, Materials, PDBj Information, and Customization. The central content area features a search bar, a welcome message, and several news items under 'Latest News' and 'Maintenance Notice'. The right sidebar contains 'Molecule of the Month', 'BMRB Search', and 'Partners'.



# PDBjへのアクセス

<http://pdbj.org/>

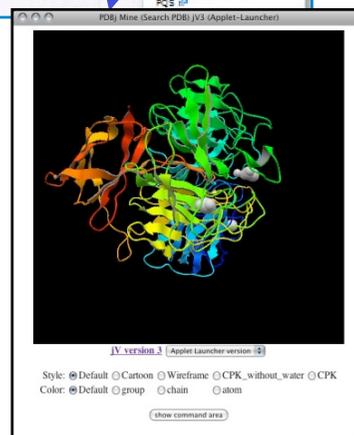
Amino acid sequence (FASTA)

```

ftp://ftp.pdbj.org/XML/pdbmlplus/fasta_seq_data/1gof_seq.txt
>1GOFA:GALACTOSE OXIDASE
ASAP1GSAISRNNWAVTCDASQSGNECNKAIDGNKDTFHHTFYGGANGDKPPHTTYIDMK
TTQNVNGLSMLPRDQNGQNGWIGRHEVYLSDDTNGSPVAGSGWADSTTKYSNFETRP
ARVYRLVAITEANGOPWTSIAEINVOASSYTAPOPLGRGPTIDLPVFAAAALEPTS
GRVLMSSYRNDAPGGSPGIIILTSMDPSTQIVSDRTVYFKDMFCGISMDGNGQIV
VTGONDAKRTSLYDSSSDSWIPGDMQVARGYQSATMSDGRVFTIGGSSGQVFKNGE
VYSPSSKTWTLFNAKVNPHLTKQGLYRSDNHAWLFGWKKGSVFOGQSPAMNNYTS
GSGDVSAGKRQENRQVAPDAMCGNAVYDAVKGKILTFGGSPDYQDSDATNAHILTLG
EPGTSPTVFASNGLIYARTFHTSVLFDGSGTFTGGQRGIFPEDSTVFPFPIYVPEO
DTPYKQPNISIVRYHSISILLPGRVPMGGGLGDCFTWRFQAIFFPFWLYNSGML
ATRFKITRTSTQSVKVGGRITISDSSISKASLRYGTATHVTNDORRPLTLTNGNG
SYSFQVPSDGGVALPGYVWMLFVHNSAGVPSVASTIRVTO
    
```

Functional site	focus & details	description
1) A:495	<input type="checkbox"/> on <input type="checkbox"/> off	sequence Y
2) A:272	<input type="checkbox"/> on <input type="checkbox"/> off	description Proton acceptor
3) A:495	<input type="checkbox"/> on <input type="checkbox"/> off	source Swiss-Prot: 1
4) A:496	<input checked="" type="checkbox"/> on <input type="checkbox"/> off	
5) A:581	<input type="checkbox"/> on <input type="checkbox"/> off	
6) A:228	<input type="checkbox"/> on <input type="checkbox"/> off	
7) A:590	<input type="checkbox"/> on <input type="checkbox"/> off	
8) A:272	<input type="checkbox"/> on <input type="checkbox"/> off	
9) A:495	<input type="checkbox"/> on <input type="checkbox"/> off	
10) A:75-87	<input type="checkbox"/> on <input type="checkbox"/> off	
11) A:194	<input type="checkbox"/> on <input type="checkbox"/> off	
12) A:227-228	<input type="checkbox"/> on <input type="checkbox"/> off	
13) A:272	<input type="checkbox"/> on <input type="checkbox"/> off	

Data viewer at PDBj



Graphic viewer: jV  
<http://pdbj.org/jV/>

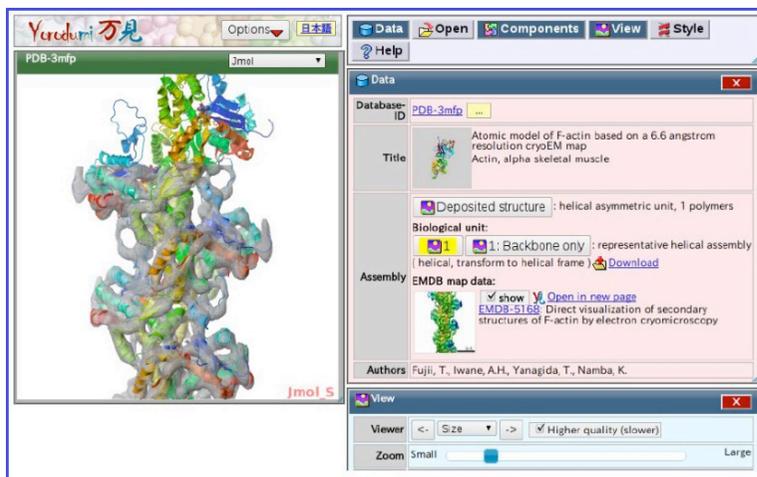
Molecular surface DB: eF-site  
<http://ef-site.hgc.jp/eF-site/>

Kinjo et al. NAR 40, D453 (2012)

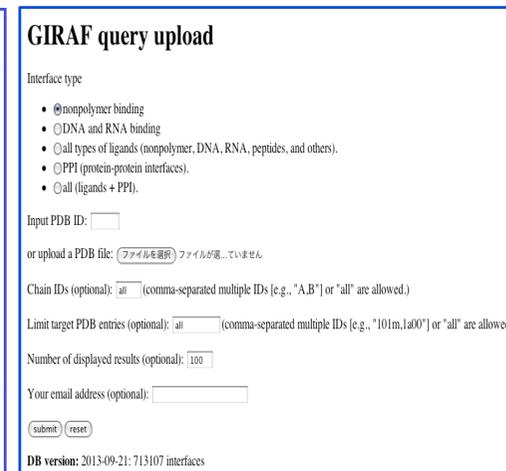
# PDBjのサービス・ツール



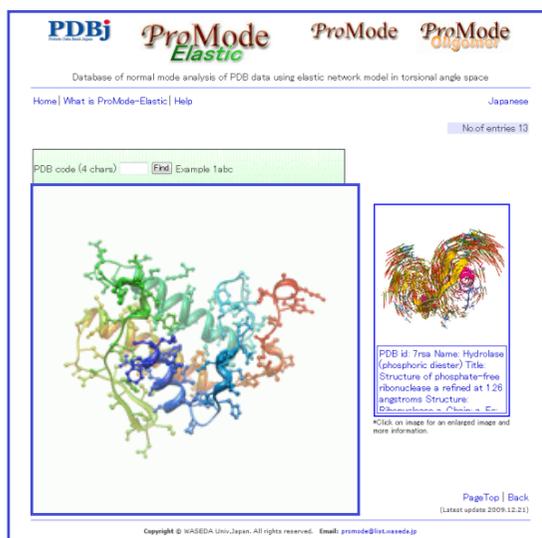
**EM Navi:** 電子顕微鏡画像EMDB



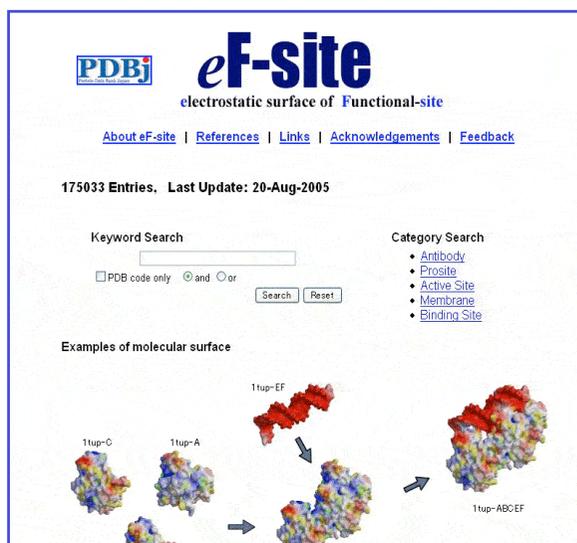
**Yorodumi:** PDBとEMDBの統合化



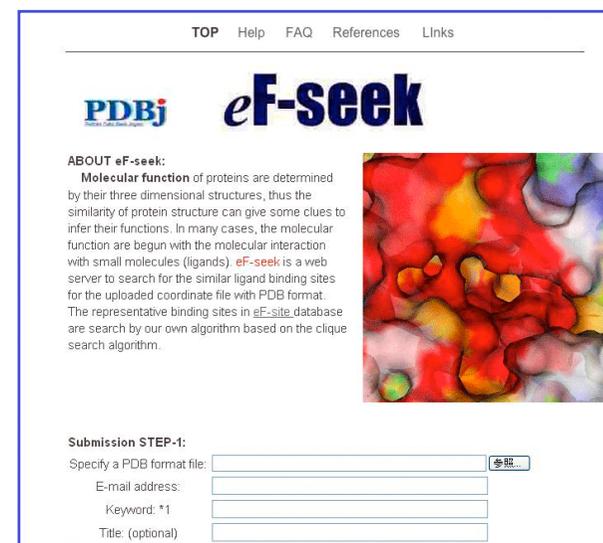
**GIRAF:** 類似リガンド結合部位



**ProMode:** 基準振動解析



**eF-site:** 分子表面DB



**eF-seek:** 類似分子表面検索

EM Navigator
3次元電子顕微鏡データナビゲーター [ [English](#) / [日本語](#) ]

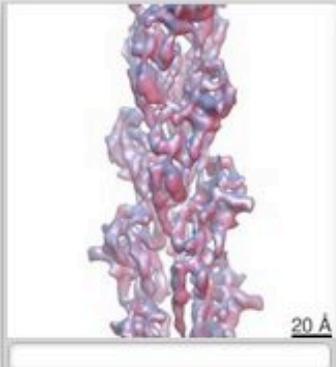
トップ
ギャラリー
リスト
分布図
統計情報
ビューア
解説

[PDBj](#) > [EM Navigator](#)

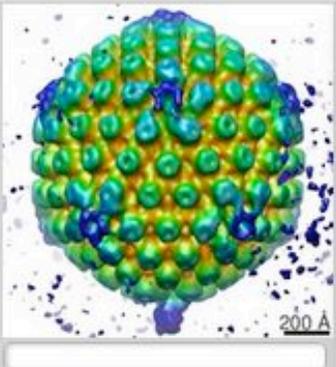
データを見る
詳しく

- さがす : (キーワード / EMDB ID / PDB ID)
- 眺める : [ギャラリー](#) [リスト](#) [表](#)

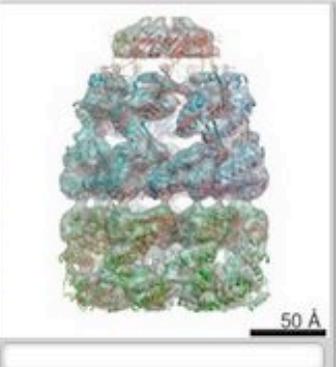
ムービースロット
▶再生
方向
✕シャッフル



EMDB-1980



EMDB-5452



EMDB-2325

情報

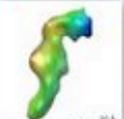
#### EM Navigatorとは？

- 生体分子や生体組織の3次元電子顕微鏡データを、**気軽にわかりやすく眺めるためのウェブサイト**です。
- [EMDB](#) と [PDB](#) のデータを利用しています ([統計情報](#))
- 分子・構造生物学の専門家にも、初心者や専門外のかたにも利用していただけるサイトを目指しています。
- PDBjが運営しています。

[詳しくはこちら](#)

#### お知らせ

- **2013-09-11**: 公開データ
 

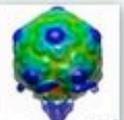





その他の最新データ: [EMDB付随情報](#), [EMDB更新](#), [PDB更新](#)
- **2013-09-04**: 公開データ
 







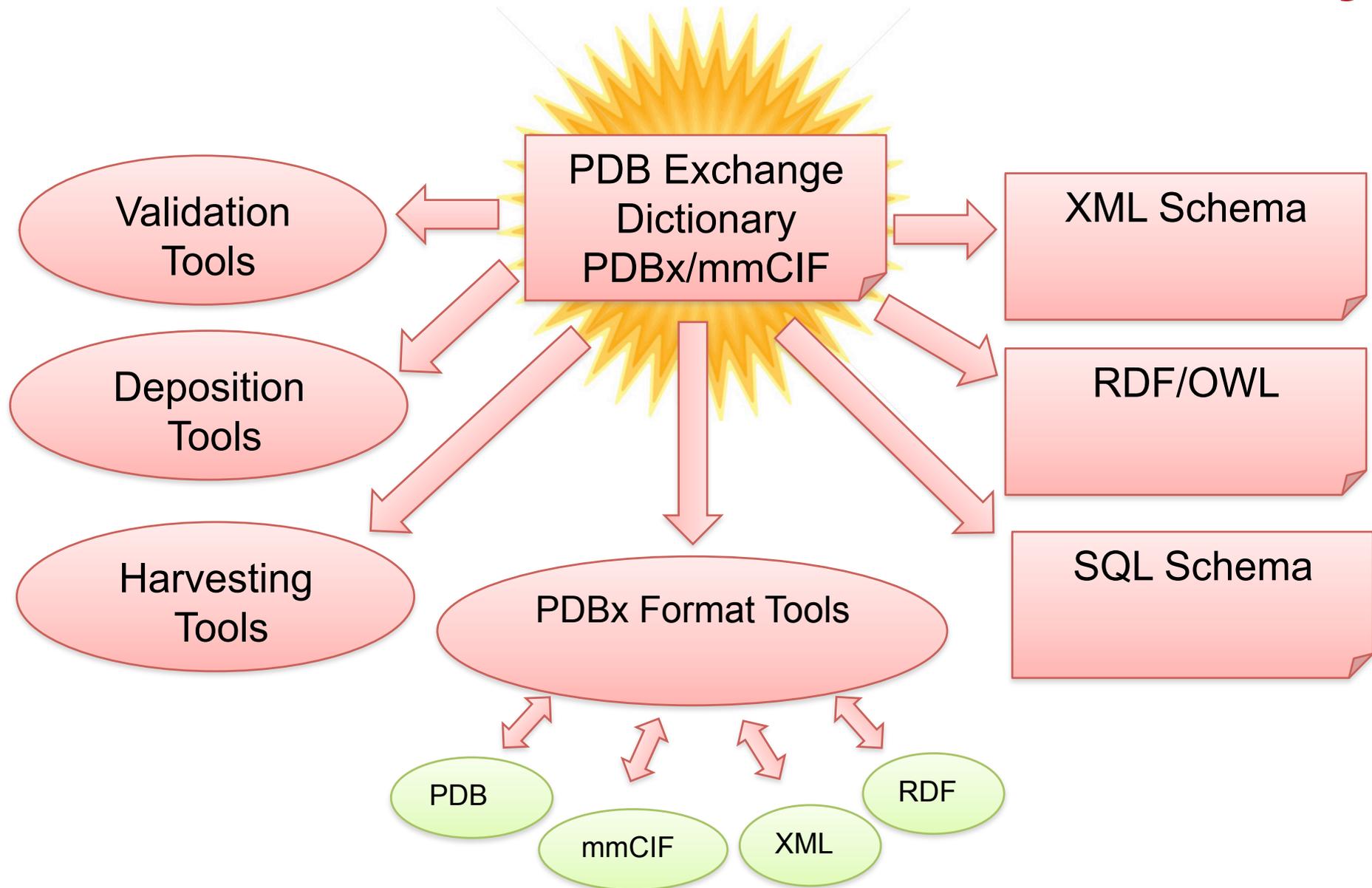
## PDBの新たなフォーマット: PDBx/mmCIF

- 問題点: PDBフォーマットは **40年前**に開発されたものであり、現代科学の要求には適合できなくなっている
- 記述における現実的な限界
  - 最大 **62** の高分子鎖
  - 最大 **99,999** の原子数
  - 低分子リガンドにおける結合次数や光学活性を特定できない
  - NMR, EM(電子顕微鏡), それらの融合法の記載が困難
  - メタデータの記載が厄介で柔軟性がない



- **backward compatibility** をできるだけ保持
- これまでの **PDB フォーマット data**を作れる**Web site**
- **2013年中**にスタートの予定

# The Central Role of the Data Dictionary



## Current Supported Archival Formats *protein structure format universe*

PDB (*ca.* 1974)

PDBx/mmCIF (*ca.* 1997)

PDBML (*ca.* 2005)

RDF (*ca.* 2011)



In managing the formats, PDBx is the master format.

# PDB Format Example

```

REMARK 3 DATA USED IN REFINEMENT.
REMARK 3 RESOLUTION RANGE HIGH (ANGSTROMS) : 1.57
REMARK 3 RESOLUTION RANGE LOW (ANGSTROMS) : 23.00
REMARK 3 DATA CUTOFF (SIGMA(F)) : 0.000
REMARK 3 COMPLETENESS FOR RANGE (%) : NULL
REMARK 3 NUMBER OF REFLECTIONS : 43316
REMARK 3
REMARK 3 FIT TO DATA USED IN REFINEMENT.
REMARK 3 CROSS-VALIDATION METHOD : NULL
REMARK 3 FREE
REMARK 3 R VAL
REMARK 3 R VAL
REMARK 3 FREE
REMARK 3 FREE
REMARK 3 FREE
REMARK 3
  
```

- Record-oriented with fixed column format
- Metadata in semi-structured remarks
- Documentation by example
- Most widely used and supported archival format

```

ATOM 1 N VAL A 363 21.741 1.557 11.725 1.00 33.52 N
ATOM 2 CA VAL A 363 21.557 -0.831 11.024 1.00 32.13 C
ATOM 3 C VAL A 363 20.954 -1.757 9.943 1.00 31.73 C
ATOM 4 O VAL A 363 19.737 -1.906 9.845 1.00 30.94 O
ATOM 5 CB VAL A 363 21.883 0.552 10.391 1.00 33.45 C
  
```

# PDBx/mmCIF Format Example

- Name – value pairs

```

_expt1.entry_id          1XBB
_expt1.method            'X-RAY DIFFRACTION'
_expt1.crystals_number   1
  
```

- Table

```

loop_
_data_
_data_
_data_
_data_
_database_PDB_rev.replaces
_database_PDB_rev.status
1 2004-11-02 2004-08-30 0 1XBB ?
2 2005-03-22 ?          1 1XBB ?
3 2009-02-24 ?          1 1XBB ?
  
```

- Simple syntax
- Named data items
- Data semantics defined in the PDBx data dictionary
- Software support in most popular languages

# PDBML Example

```
<PDBx:entity_polyCategory>
  <PDBx:entity_poly entity_id="1">
    <PDBx:type>polypeptide(L)</PDBx:type>
    <PDBx:nstd_linkage>no</PDBx:nstd_linkage>
    <PDBx:nstd_monomer>no</PDBx:nstd_monomer>
    <PDBx:pdbx_seq_one_letter_code>
      DIVLTOSPASLSASVGFETVTTITCRASCNIHNYLAWYOOKOCKSPOLLVYVYTTTLADG
    </PDBx:pdbx_seq_one_letter_code>
  </PDBx:entity_poly>
</PDBx:entity_polyCategory>
```

- Three flavors of XML files:
  - fully marked-up files
  - files without atom records
  - files with a more space efficient encoding of atom records
- Follows naming and semantics of the PDBx data dictionary



# Transitional Home for Large Structures

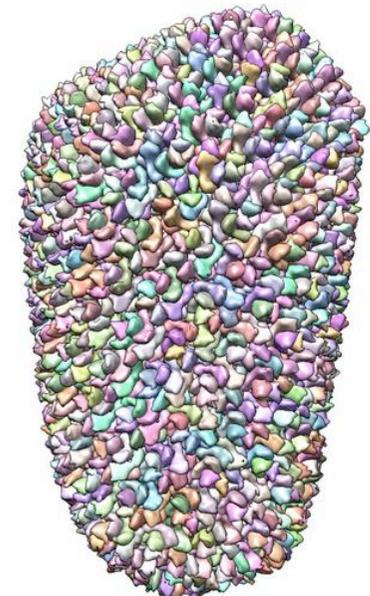
Large single entries are now stored separately on the wwPDB ftp site, and PDB internally produces *divided/split* PDB format files.

[ftp://ftp.wwpdb.org/pub/pdb/data/large\\_structures/mmCIF/](ftp://ftp.wwpdb.org/pub/pdb/data/large_structures/mmCIF/)  
[ftp://ftp.wwpdb.org/pub/pdb/data/large\\_structures/XML/](ftp://ftp.wwpdb.org/pub/pdb/data/large_structures/XML/)

- HIV-1 Capsid 3J3Q –
- 1356 chains
  - >2M atoms
  - 25 – PDB format entries



3J3Y



3J3Q

# Providing Format Compatibility

- Adopt a *PDB friendly* mmCIF/PDBx style -
  - All records on a single text line
  - Columns presented in standard column order.
  - Tabular presentation with leading record names  
(e.g. ATOM, CELL, REFIN)
  - Method independent features in left-most columns  
(e.g. identifiers & coordinates)
  - Method specific features in the right-most columns  
(e.g. ADPs, NMR order/disorder parameters)
  - Continue to support PDB nomenclature semantics  
(e.g. PDB style chains, residue numbering, and insertion codes)
- Large entries will be internally converted to divided/split PDB format files.

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

# 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** - **Chimera, Jmol, OpenRasMol**  
PDB actively working with community developers to help fill in missing functionalities. Two workshops scheduled in Fall 2013 ...

# PDBj スタッフ (2013年度)

## 統括責任者

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)

松田 真 (大阪大学蛋白質研究所・特任研究員)

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 玲 (大阪大学蛋白質研究所・准教授)

岩崎 憲治 (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)

西川 建 (大阪大学蛋白質研究所・客員教授)

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)

阿久津 秀雄 (大阪大学蛋白質研究所・招聘教授)

児嶋 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)

高橋 あみ (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

輪湖 博 (早稲田大学社会科学総合学術院・教授)

(for Pro Mode)

伊藤 暢聡 (東京医科歯科大学大学院・教授)

木下 賢吾 (東北大学大学院情報科学研究科・教授)

(for e F-site)

Standley, M.Daron (大阪大学免疫学フロンティア  
研究センター・准教授)

(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・  
特任准教授) (for MAFFTash)

## 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)