

wwPDBとPDBjが担う PDB活動の現状

栗栖源嗣 大阪大学蛋白質研究所





http://pdbj.org/
http://wwpdb.org/





Protein Data Bank Japan Protein Data Bank Japan Protein Data Bank Japan

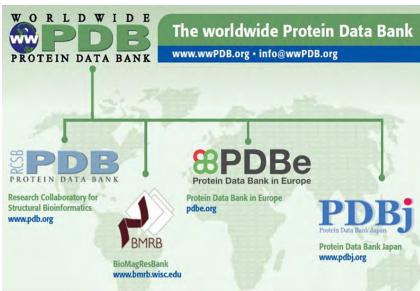
http://pdbj.org/

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.



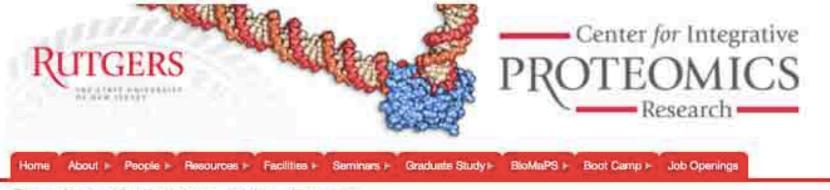


PDBj staffs (April 2014)



wwPDB members and their heads





Stephen K. Burley, M.D., D.Phil.

Director, Center for Integrative Proteomics Research Associate Director, RCSB-PDB Distinguished Professor, Department of Chemistry and Chemical Biology Member, Cancer Institute of New Jersey



A new director of the RCSB-PDB

(Helen M. Berman is going to be a chair of the wwPDB)



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

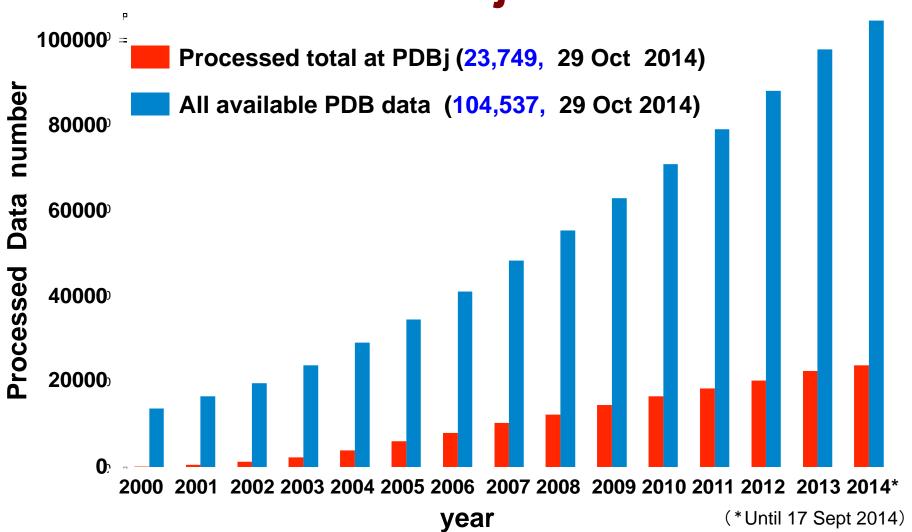


Activities/Services of each member of the wwPDB

- "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



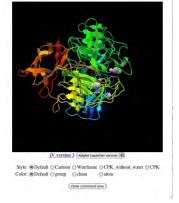
Data-out from PDBj

http://pdbj.org/



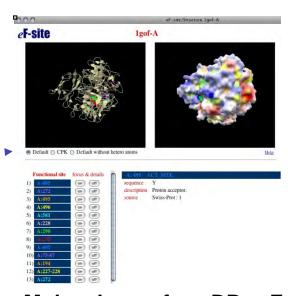
Data viewer at PDBj

Graphic viewer: JV http://pdbj.org/jV/



Amino acid sequence (FASTA)





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

Kinjo et al. NAR 40, D453 (2012)



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.

```
1
                  GLN A
                         39
                                  24.690 -27.754
                                                   24.275
                                                            1.00 60.76
ATOM
             N
                                                                                  N
                                                                                  C
              CA
                                  23.581 -26.768
                                                   24.416
                                                            1.00 60.98
ATOM
                  GLN A
                         39
ATOM
          3
             C
                  GLN A
                         39
                                  23.990 -25.379
                                                   23.905
                                                            1.00 59.98
                                                                                  C
           4
              0
                  GLN A
                         39
                                  25,070 -25,209
                                                   23,330
                                                            1.00 60.25
                                                                                  0
ATOM
                                  23.136 -26.685
                                                   25.878
                                                            1.00 60.69
                                                                                  C
ATOM
              CB
                  GLN A
                         39
ATOM
          6
             N
                  VAL A
                         40
                                  23.115 -24.395
                                                   24.122
                                                            1.00 59.58
                                                                                  N
          7
              CA
                 VAL A
                         40
                                  23.342 -23.010
                                                   23,690
                                                            1.00 57.26
                                                                                  C
ATOM
                                                                                  C
ATOM
                  VAL A
                         40
                                  24.000 -22.152
                                                   24.778
                                                           1.00 56.00
ATOM
          9
              0
                  VAL A
                         40
                                  23.992 -20.920
                                                   24.692
                                                            1.00 55.53
                                                                                  0
                                                                                                        PDB
                                  22.015 -22.337
ATOM
         10
              CB
                 VAL A
                         40
                                                   23.275
                                                           1.00 57.32
```

```
PDBx/mmCIF
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 seg 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
ATOM
           1 N
                      GLN
                                39
                                     24,690
                                              -27,754
                                                        24.275
                                                                1 1.000
                                                                              60.760
                   N
                           Α
ATOM
           2
              CA
                    C
                       GLN
                                39
                                     23.581
                                              -26.768
                                                        24.416
                                                                  1.000
                                                                              60.980
                           Α
           3
              C
                                     23.990
                                              -25.379
                                                        23.905
                                                                   1.000
                                                                              59.980
ATOM
                    C
                       GLN
                           Α
                                39
                                                                1
ATOM
              0
                    0
                       GLN
                           Α
                                39
                                     25.070
                                              -25.209
                                                        23.330
                                                                1
                                                                  1.000
                                                                              60.250
ATOM
           5
              CB
                    C
                       GLN
                                39
                                     23.136
                                              -26.685
                                                        25.878
                                                                1
                                                                  1.000
                                                                              60,690
                           Α
ATOM
           6
                                40
                                     23.115
                                              -24.395
                                                        24.122
                                                                1
                                                                   1.000
                                                                              59.580
              N
                   N
                      VAL
                           Α
ATOM
           7
              CA
                    C
                      VAL
                           Α
                                40
                                     23.342
                                              -23.010
                                                        23.690
                                                                1
                                                                  1.000
                                                                              57,260
           8
              C
                      VAL
                                40
                                     24.000
                                              -22.152
                                                        24.778
                                                                1
                                                                  1.000
                                                                              56.000
ATOM
                    C
                           Α
ATOM
              0
                      VAL
                           Α
                                40
                                     23.992
                                              -20.920
                                                        24.692
                                                                1
                                                                   1.000
                                                                              55.530
ATOM
          10
              CB
                      VAL
                                40
                                     22.015
                                              -22.337
                                                        23.275
                                                                1
                                                                   1.000
                                                                              57.320
                            Α
                                             -22.804
          11
              N
                                41
                                     24.560
                                                        25.797
                                                                1
                                                                  1.000
                                                                              54.570
ATOM
                   N ALA
                           Α
```



wwPDB Service site for a new format http://mmcif.wwpdb.org/



Dictionaries

- · Browse the current dictionary»
- · Download/view all dictionaries »
- · Search dictionaries»

Documentation

- PDB -> PDBx/mmCIF correspondences »
- PDBx/mmCIF for large structures
- 50
- 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»



wwPDB Service site for a new format http://mmcif.pdbj.org/

PDBx/mmCIF 辞書関連情報

このサイトでは、PDBエントリーの登録、アノテーション、データ保管を行う際に国際蛋白質構造データバンク(wwPDB)で利用されているファイルの書式、辞書、ソフトウェア・ツールに関する情報を提供しています。

現行の語書を見る。

辞書

- 現行の辞書を見る。
- 全ての辞書をダウンロードする×
- 辞書を検索する。

文書

- PDB→PDBx/mmClFの対応
- 巨大構造のためのPDBx/mmCIF »
 - ソフトウェアッ
 - C++ » と Python » のプログラム例
 - ファイルの書式»と辞書の構成»
 - 原子。と分子。の記述方法
 - 参考文献 "
 - 用語集。

よくある質問

PDBx/mmCIFフォーマット、データ内容、ソフトウェアツールについての質問は、よくある質問。を参照下さい。

このページの内容は wwPDB から提供されている原文を PDBj で翻訳し、適宜加筆修正したものです。



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, jV, molmil

PDB actively working with community developers to help fill in missing functionalities.



wwPDB Service site for a new format http://mmcif.pdbj.org/



PDBx/mmCIF 辞書関連情報

このサイトでは、PDBエントリーの登録、アノテーション、データ保管を行う際に 国際蛋白質構造データバンク (wwPDB) で利用されているファイルの書式、辞書、ソフ

トウェア・ツールに

PDBx/mmCIF

β-version ≡

PDBフォーマット - PDBx/mmCIF 変換サービス

善

- 現行の辞書を見る。
- 全ての辞書をダウンロード
- 辞書を検索する。

アップロードした分子構造データファイルを別の書式に変換します。 アップロードしたファイルの タイプは自動的に判定されます。 mmCIFファイルの場合はPDBファイルに、PDBファイルの場合 はmmCIFファイルに変換されます。

- 1. 変換元ファイルの指定 4m0s.cif
- 2.ファイルタイプの確認 mmCIF → PDB format
- 3. 変換実行・結果ダウンロード

変換ファイルをダウンロード ファイル選択に戻る

このページの内容は wwPDB から提供さ



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/XML/

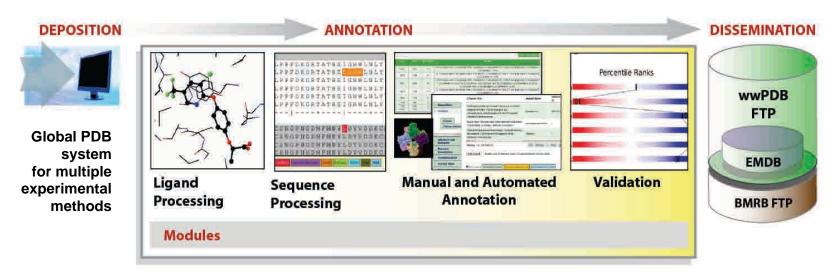


HIV-1 Capsid 3J3Q

- •1356 chains
- •>2M atoms
- •25 PDB format entries



New Deposition and Annotation System For data increase and high quality data management



- 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





A member of the



Languages: en | ja | ko | zh-cn | zh-tw

<u>登録サーバ トップ | 新登録システムチュートリアル (en, ja) | ADIT登録チュートリアル (en, ja, ko, zh-cn, zh-tw) | ADITに関するFAQ | 登録に関するFAQ | ファイルフォーマット | お問い合わせ</u>



巨大構造を登録される方は、wwPDBの新登録システムをお使いください。 PDBフォーマットの制限を超えるような巨大構造(鎖数が62、又は原子数が99,999個を超える)を登録する場合、 最新バージョンのCCP4 (REFMAC > 5.8) 又は Phenix (> 1.8.2)を使って一つのPDBx/mmCIFファイルを<u>準備した後</u>、wwPDBの登録システムからご登録ください。



X線結晶構造を登録される方は、wwPDBの新登録システムをお使いください。 http://deposit.wwpdb.org/deposition (NMRと電子顕微鏡 (3DEM) 構造については、近々公開される予定です。)

wwPDBの新登録・編集システムについての情報(チュートリアル、FAQ等)は、こちらからご覧ください。





A member of the



詳細は、wwPDB: Deposition System: Getting Started (英語) をご覧ください。

wwPDBの新登録・編集システムについてのその他の情報は、こちらの一覧ページもご覧ください。

新登録システム:はじめに

wwPDBの新しい登録システムについてご紹介します。 新システムは、PDBへの登録手続きにおける効率や、完全性、データ収集の品質を改善するために、wwPDBパートナーが開発を行ってきました。 今後の改良や開発のためにも、新システムについてのご感想・ご意見を頂ければ幸いです。

新登録システムの概要

新システム(http://deposit.wwpdb.org/deposition)は、 現在はX線結晶構造の登録にご利用頂けます。 近い将来に、NMRや電子顕微鏡によって解析された構造も登録できるようになります。

登録の前に、wwPDBの構造検証ツール (http://wwpdb-validation.wwpdb.org/) を使って、データを検証することをお勧めします。

今後増々、複雑で大きい構造がPDBに登録されることに備え、新システムは、PDBx/mmCIFの辞書とファイル形式に基づいています。 登録データの処理や配布もPDBx/mmCIFデータファイルで行われます。 PDBフォーマットのデータも、登録の際にアップロードでき、アノテータによって処理された後は、ダウンロードすることもできます。



登録システムについては、大きな変更点が4つあります。

- 1. リガンドの検証 アップロードされる座標に含まれる全てのリガンドに対して、PDBで公開済みのリガンドに一致するものがあるかどうかを検索します。 完全一致ではないが近い構造のリガンドが存在する場合、登録者は注意して確認する必要があります。 PDBに近い構造も存在しない新しいリガンドの場合、登録者はSMILESやInChI文字列、または二次元画像を求められます。
- 2. 配列の検証 座標ファイル中の全ての高分子について、登録者が入力した一次配列と座標ファイル中の配列との間で、配列を比較します。 不一致があれば、登録者に注意が促されます。 登録者は、登録のインターフェース上で一時配列を修正することもできますし、新しい座標をアップロードすることによって座標を修正することもできます。 さらに、BLAST検索でUniProtデータベースを検索したり、UniProtのアクセション番号を提供することもできます。 登録者の配列とUniProtの配列の間に不一致がある場合は、登録者はその理由についての注釈(例えば、置換変異やパリアントなど)を求められます。
- 3. X線結晶構造に対する専門的な検証 登録者が座標ファイルをアップロードすると、検証処理が始まり、処理が終了すると、登録者にメールで通知が届きます。 検証レポートは、ナビゲーションパネルのオプション「ダウンロードと表示 (Download & View) > レポートの表示 (View Reports)」から、 閲覧できます。
- 4. コミュニケーションパネル コミュニケーションパネルでは、登録者とアノテータ間でメッセージを送ることができます。 文章を送るだけでなく、このインターフェースを通して、アップロードしたデータや、編集されたもしくは議論したデータに簡単にアクセスできます。 このページで行われたコミュニケーションは、このエントリーの登録とともに永久に保管され、このエントリー以外では使用されません。 注)いかなるデータファイルのアップロードも、「ファイルのアップロード (File Upload)」オプションを使って行い、コミュニケーションパネルや電子メールでは行わないことに注意してください。



新登録システムの紹介

X線結晶構造の登録: 座標・実験データファイルのアップロード リガンドの登録: 登録構造のリガンドに一致するリガンドを確認す 構造の検証、必須項目の入力 る







チュートリアル:新登録システムを使って

- 1. 登録トップページ(http://deposit.wwpdb.org/deposition) には、2つのオプションがあります。左のパネルは登録IDとパスワードを使用して登録システムにログインするために使用します。 右側のメインパネルは、新しい登録を始めるために使用します。 新しい登録を開始するときは、登録者の国名を選択しますが、これは最善のサポートを提供するために必要となります。 フォームを入力し、登録開始ボタンを選択すると、登録IDとパスワードが電子メールで送信されます。 (パスワードは登録者自身が決めることもシステムに自動生成させることも可能です)
- 2. この情報を使用して、左側パネルからシステムへログインしてください。 ログイン後は、登録IDがページ上部に表示されます。 座標と構造因子ファイルをアップロードする際には、左側のナビゲーションパネルから、「ファイルのアップロード (File Upload)」オブションを選択してください。 ファイルのアップロード後、ドロップダウンメニューから、対応するファイルタイプを選択してください。
- 3. ファイルのアップロード後、検証処理が始まります。 検証処理が完了すると、登録者は電子メールで通知を受け取ります。 また、検証処理が完了すると、ナビゲーションパネルの最下部にある「**ダウンロードと表示 (Downloads & View) > レポートの表示 (View Reports)**」から、検証レポートを参照できるようになります。



Validation Report

- Version 1.0 in production use since August 2013
 - http://www.wwpdb.org/validation.html
 - Fixing occasional bugs
 - Collecting feedback to inform possible changes
 validation@mail.wwpdb.org
- January 2014 validation data for all X-ray structures will be made publicly available through the wwPDB ftp sites



wwPDB X-ray Structure Validation Report

Aug 13, 2013 - 09:45 AM BST

PDB ID : 1CBS

Title: CRYSTAL STRUCTURE OF CELLULAR RETINOIC-ACID-BINDI

NGPROTEINS I AND II IN COMPLEX WITH ALL-TRANS-RETINOIC

ACID AND A SYNTHETIC RETINOID

Authors : Kleywegt, G.J.; Bergfors, T.; Jones, T.A.

Deposited on : 1994-09-28 Resolution : 1.80 Å(reported)

DISCLAIMER

This is a preliminary version of a new style of wwPDB validation report.

We welcome your comments at validation@mail.wwpdb.org

The following versions of software and data (see references) were used in the production of this report:

MolProbity : 4.02b-467 Mogul : 1.15 2013 Xtriage (Phenix) : dev-1323

EDS: trunk21216
Percentile statistics: 20591

Refmac : 5.8.0043 CCP4 : 6.3.0 (Settle) Ideal geometry (proteins) : Engh & Huber (2001)

Ideal geometry (DNA, RNA) : Parkinson et. al. (1996) Validation Pipeline (wwPDB-VP) : trunk21216



Validation Report

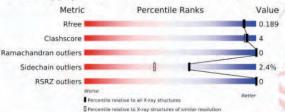
Summary

- Quality vs. all PDB X-ray
- Quality vs. entries at similar resolution
- Overview of residuebased quality for every polymer
- Table of ligands that may need attention

1 Overall quality at a glance (

The reported resolution of this entry is 1.80 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Rfree	65580	5522 (1.84-1.76)
Clashscore	76988	5040 (1.82-1.78)
Ramachandran outliers	75395	6528 (1.84-1.76)
Sidechain outliers	75377	6529 (1.84-1.76)
RSRZ outliers	65576	5522 (1.84-1.76)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	137	

Mol	Chain	Length	Quality of chain
1	A	371	
1	C	371	

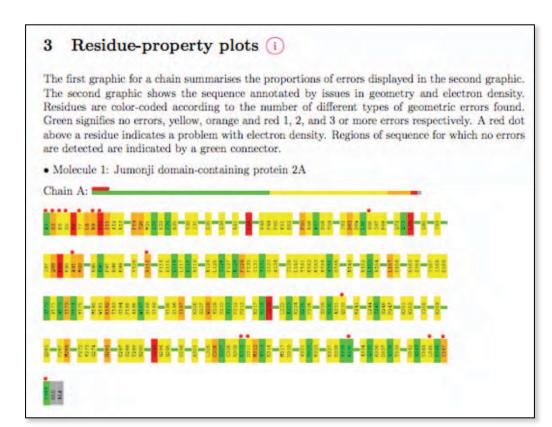
The following table lists non-polymeric compounds that contain outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
2	NAG	A	401	-	X
2	NAG	C	401	-	X



Validation Report

- Residue quality
 Red dots: poor density (RSR-Z > 2, as in EDS)
- Model/data fit ligands etc.
 - "LLDF" –Local
 Ligand Density Fit
 = Z-score of ligand
 RSR relative to
 nearby polymeric
 residues





PDBj members in 2014

Head

Nakamura, Haruki, Ph. D. M (Prof., IPR, Osaka Univ.)

Group for PDB Database Curation

- Nakagawa, Atsushi, Ph. D. (Group Leader, Prof., IPR, Osaka Univ.)
- Matsuda, Makoto, Ph. D. (IPR, Osaka Univ.)
- o Igarashi, Reiko (IPR, Osaka Univ.)
- Kengaku, Yumiko (IPR, Osaka Univ.)
- Cho, Hasumi, Ph. D. (IPR, Osaka Univ.)
- Ikegawa, Yasuyo (IPR, Osaka Univ.)
- Sato, Junko (IPR, Osaka Univ.)

Group for Development of new tools and services

- o Kinjo, Akira R., Ph. D. W (IPR, Osaka Univ.)
- Iwasaki, Kenji, Ph. D. (IPR, Osaka Univ.)
- Suzuki, Hirofumi, Ph. D. (IPR, Osaka Univ.)
- o Yamashita, Reiko (IPR, Osaka Univ.)
- Kudou, Takahiro (IPR, Osaka Univ.)
- o Bekker, Gert-Jan (IPR, Osaka Univ.)

Group for BMRB

- o Fujiwara, Toshimichi, Ph. D. (Group Leader, Prof. Osaka Univ.)
- Akutsu, Hideo, Ph. D. (Guest Prof., IPR, Osaka Univ.)
- Kojima, Chojiro, Ph. D. (IPR, Osaka Univ.)
- Kobayashi, Naohiro, Ph. D. (IPR, Osaka Univ.)
- Iwata, Takeshi (IPR, Osaka Univ.)
- Yokochi, Masashi (IPR, Osaka Univ.)

Collaboratory Researchers

- Wako, Hiroshi, Ph. D. W (Prof., Waseda Univ.) (for Pro Mode)
- o Ito, Nobutoshi, Ph. D. (Prof., Tokyo Medical and Dental Univ.)
- Kinoshita, Kengo, Ph.D.
 [™] (Prof., Tohoku Univ.) (for e F-site)
- Standley, Daron, Ph. D.
 (IFReC, Osaka Univ.) (for SeqNavi, StructNavi, SeSAW, and ASH)
- Katoh, Kazutaka, Ph. D. (IFReC, Osaka Univ.) (for MAFFTash)

Secretary

Haruki, Nahoko (IPR, Osaka Univ.)