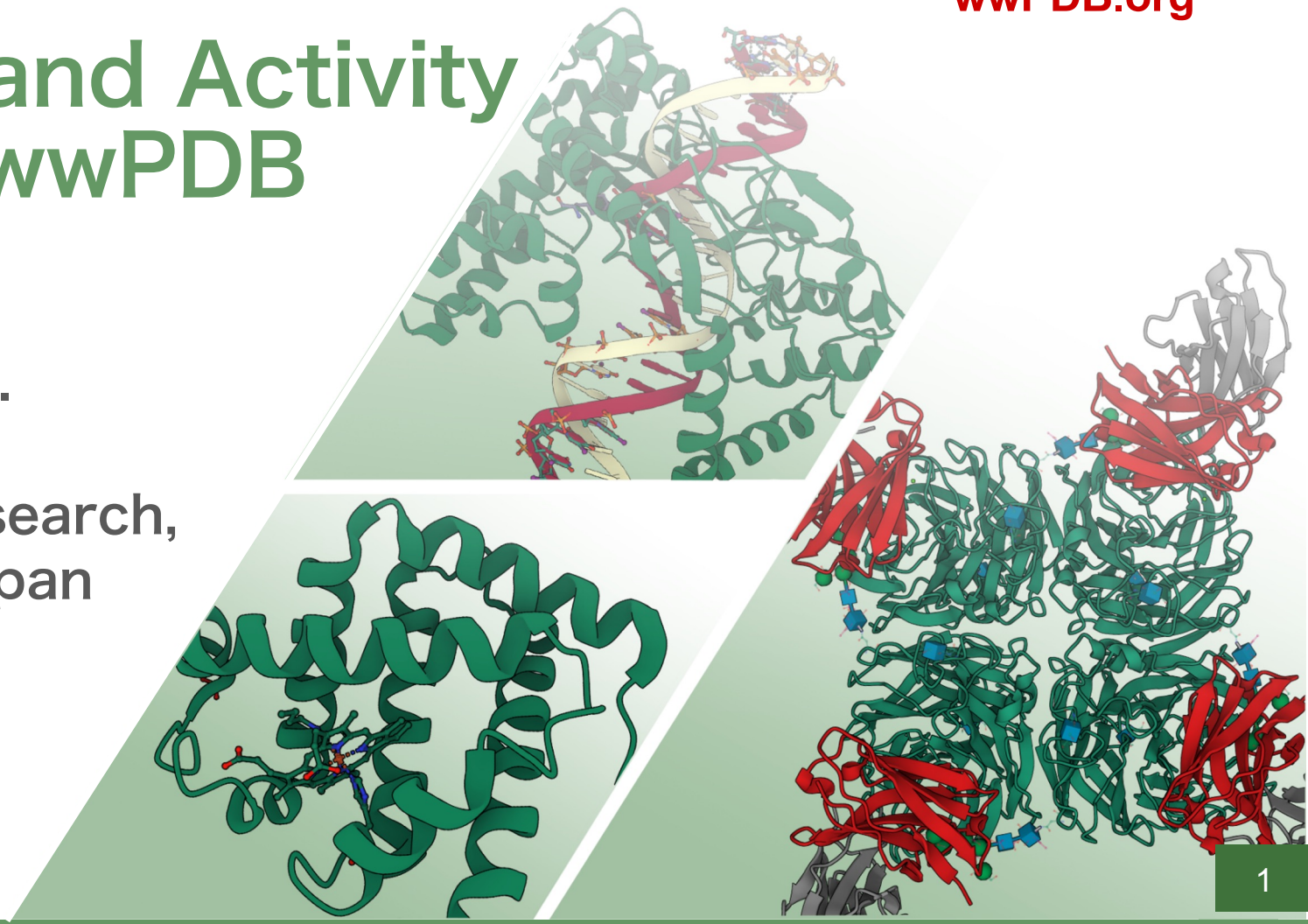
PDBj
Protein Data Bank Japan

WORLDWIDE PDB
PROTEIN DATA BANK

wwPDB.org

Recent Activities and Activity Plan of PDBj and wwPDB

Genji Kurisu, Ph.D.
栗栖源嗣　工学博士
Institute for Protein Research,
Osaka University, Japan



PDB is an Oldest Digital Resource of Life Science

PDBj Protein Data Bank Japan
as a member of wwPDB WORLDWIDE PROTEIN DATA BANK

CRYSTALLOGRAPHY

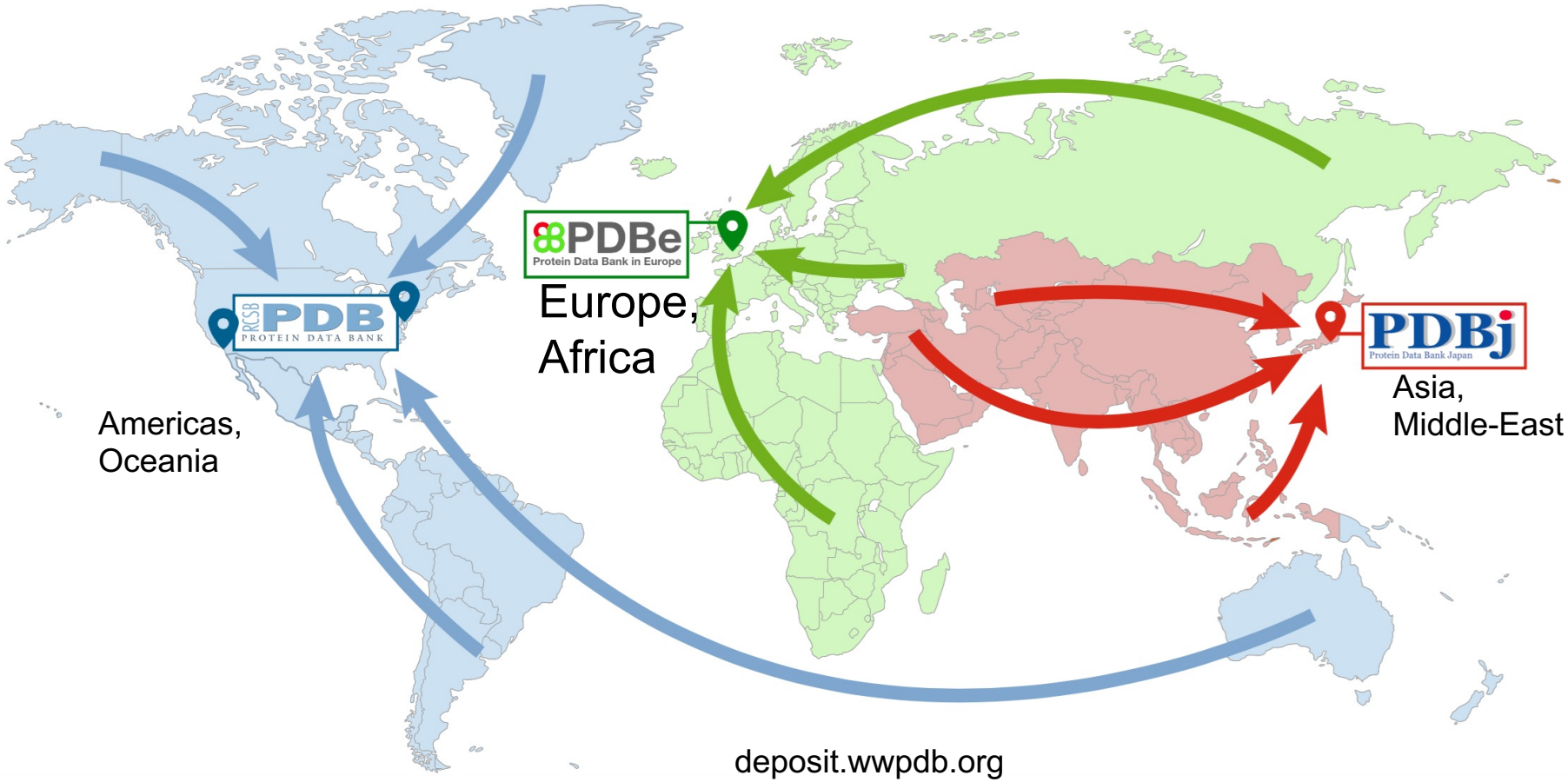
Protein Data Bank

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

Nature New Biology **233**, page 223 (1971)

Thank you Prof. Yuh-Ju Sun!

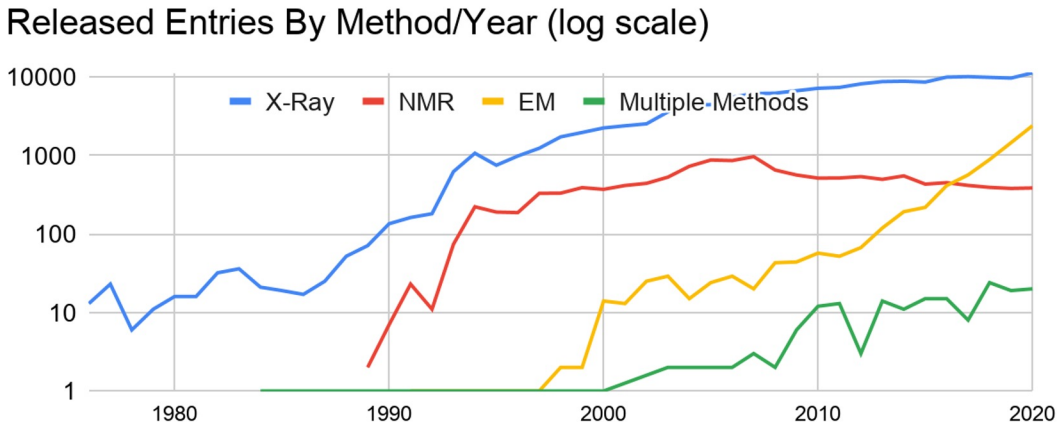
PDBj is in charge of the PDB entries from Asia

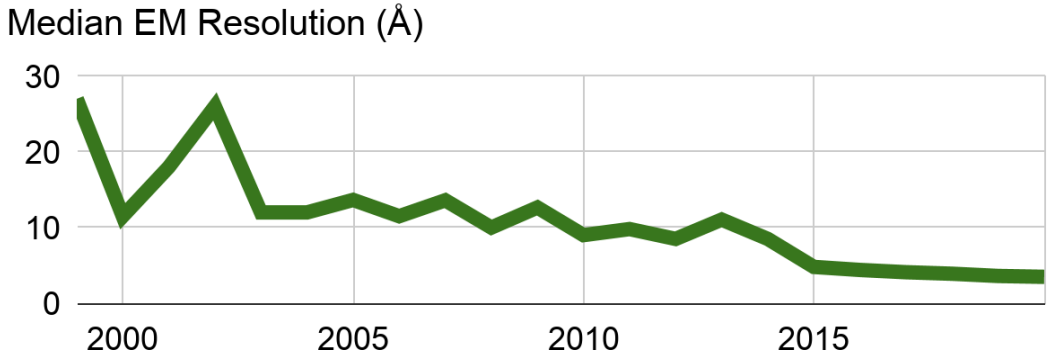


All data available at RCSB PDB, PDBe and PDBj are exactly SAME!

PDB Archive Update (1)

- Professionally curated entries, more than 200,000, are freely available under CC0 1.0 license.
  - ❖ Annually ~8% increas
- More than 400 external data resources use PDB
- Number of entries by Cryo-EM increase
  - ❖ more than 60% in 2020
  - ❖ Resolutions by 3DEM are improving, (some are ~1Å).





CoreTrustSeal certification renewed through April 2024 (CoreTrustSeal.org)

PDB Archive Update (2)

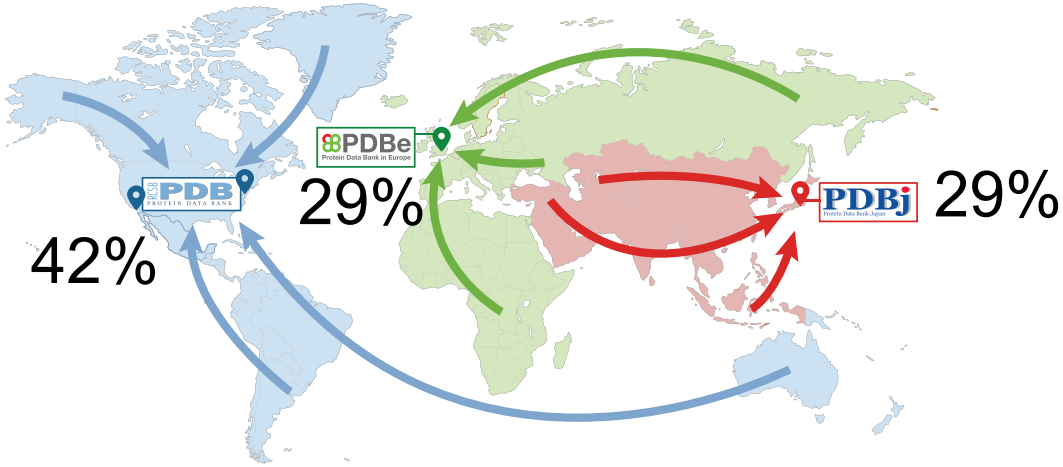
- Depositions from Asia are almost equal to those from Europe/Africa.
  Asia : 4,749/16,344 = 0.2905
  Europe/Africa : 4758/16,344 = 0.2911

  ➡ Rapid increase in depositions from mainland China

- AlphaFold initiated starting models become familiar.

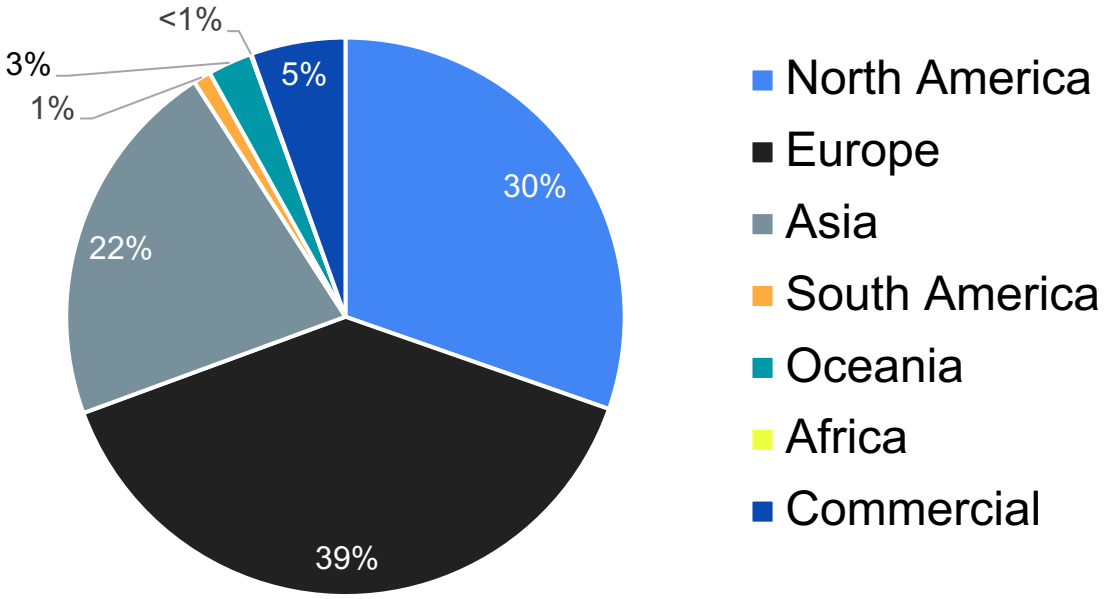
  ➡ AlphaFoldDB information started to be curated as a template for modeling/ molecular replacement.

Geographical distributions of PDB depositions in 2022

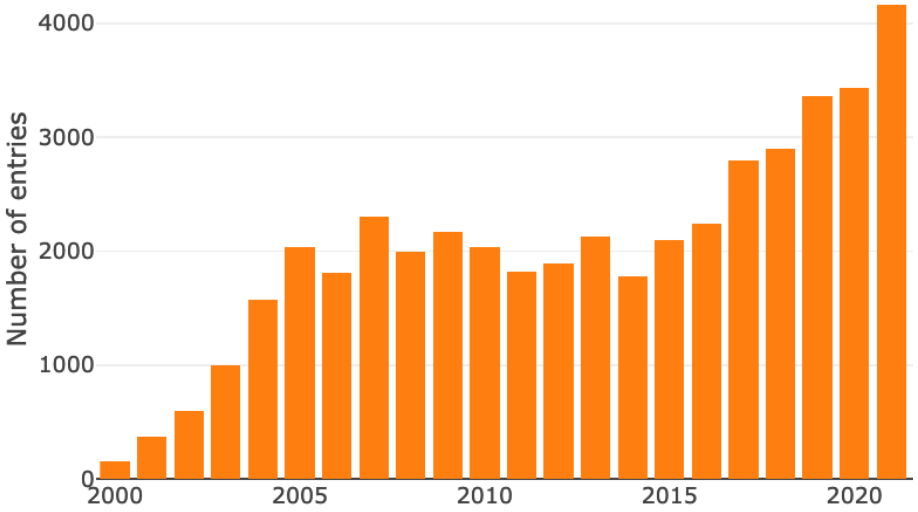


Asian Entries are increasing than the others

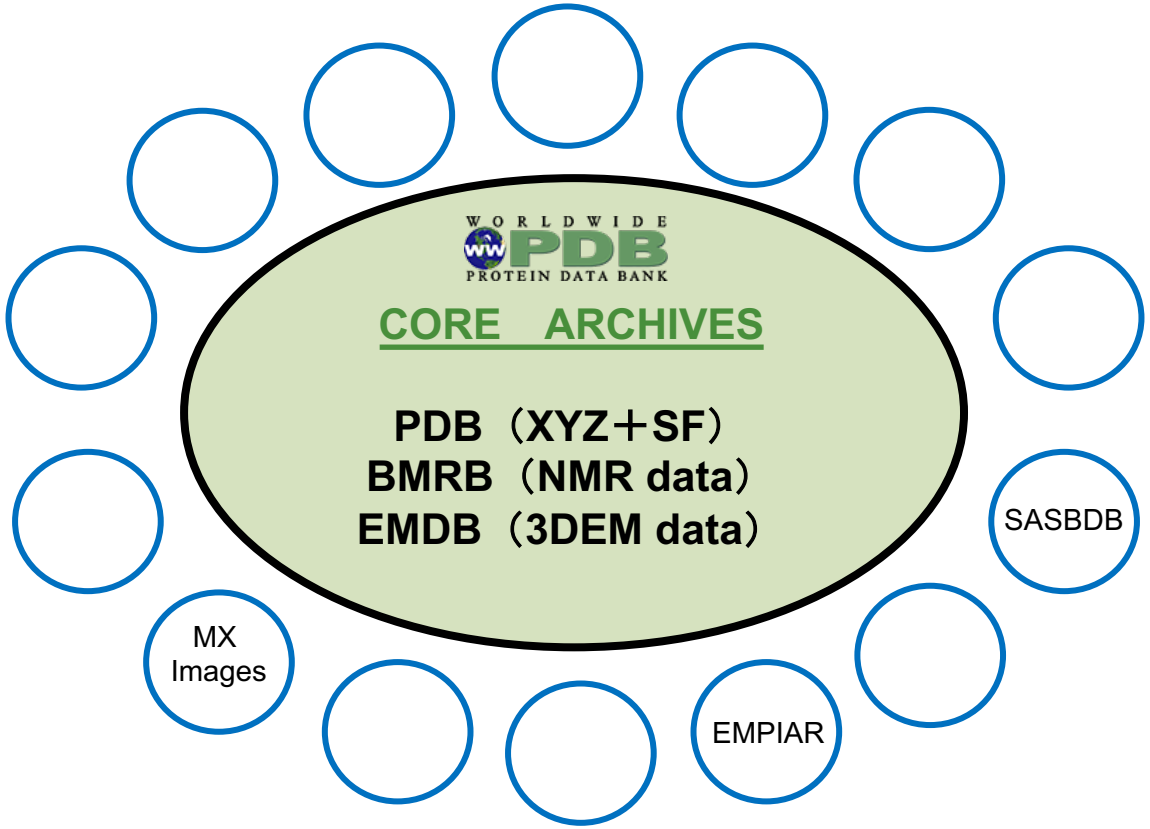
Depositor Locations

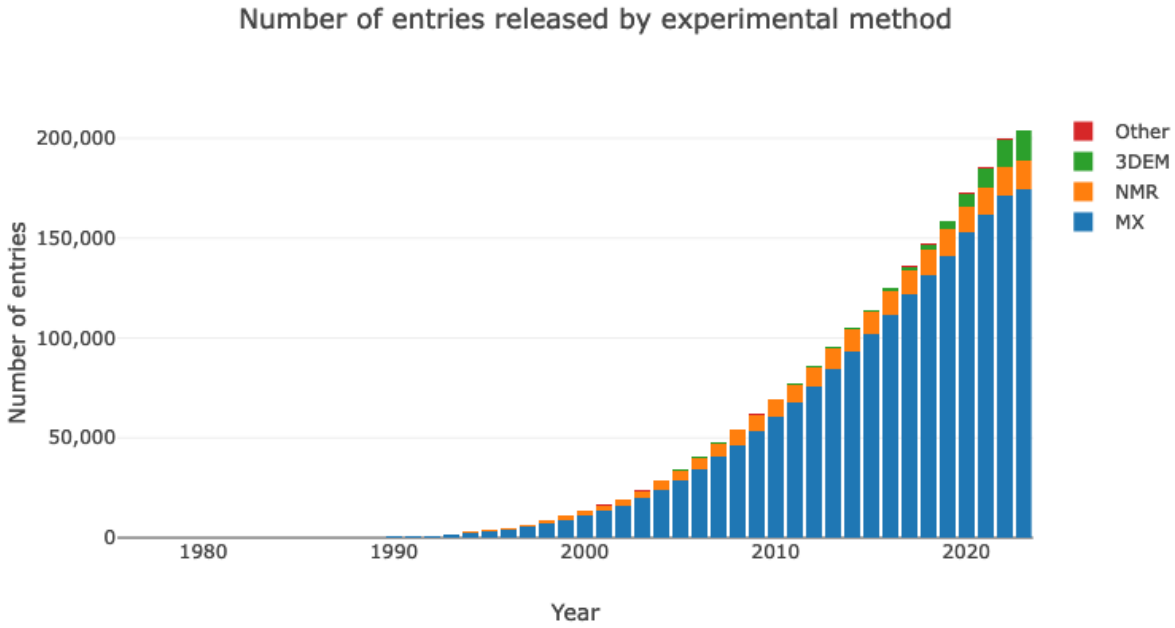


Number of entries processed by PDBj



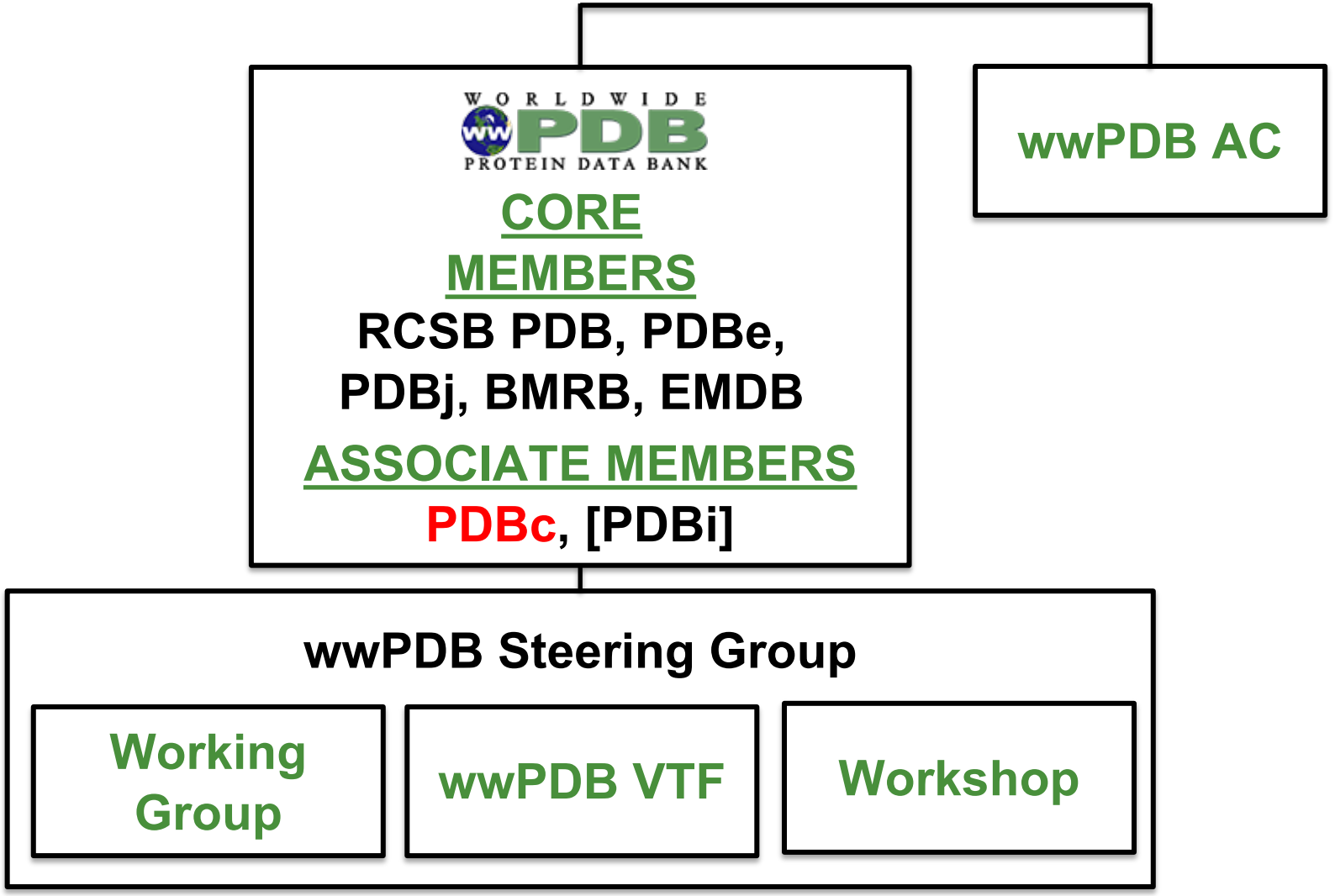
wwPDB maintain three Core Archives



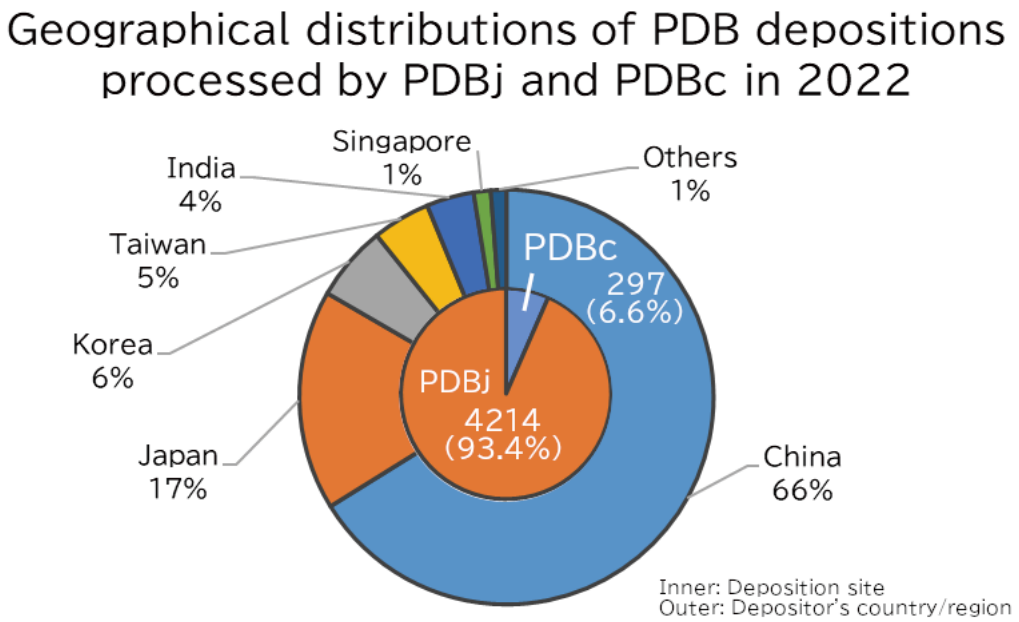


7

PDB China at Shanghai became an Associate Member in 2022



Two Annotators started to curate PDB data from Mainland China



- Deposition site in Asia is still PDBj until the performance of PDB China is fully functional.
- When deposition site in Shanghai starts, depositors in Hong Kong or Taiwan can choose the deposition site at PDBj or PDBc, as approved by wwPDB AC.

Validation report started from 2016

EDITORIAL

Nature Struct. Mol. Biology, 23 (10), 871, 2016

nature structural & molecular biology

Where are the data?

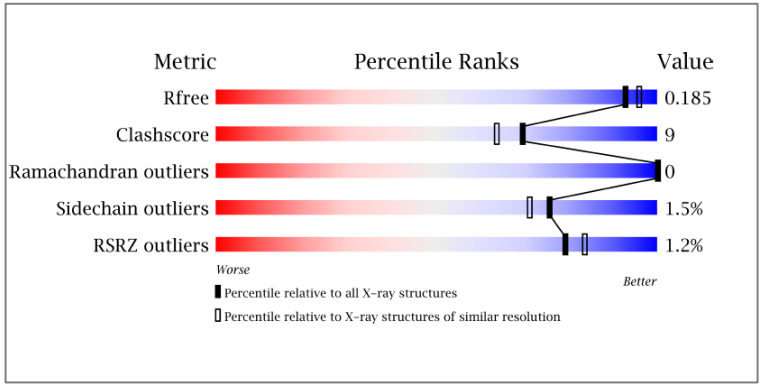
Here, we announce two policy changes across Nature journals: data-availability statements in all published papers and official Worldwide Protein Data Bank (wwPDB) validation reports for peer review.

We are now taking a further step and are requesting official wwPDB validation reports for peer review. These reports are made available by the wwPDB after data deposition (http://www.wwpdb.org/validation/validation-reports). Other Nature journals will soon follow suit.

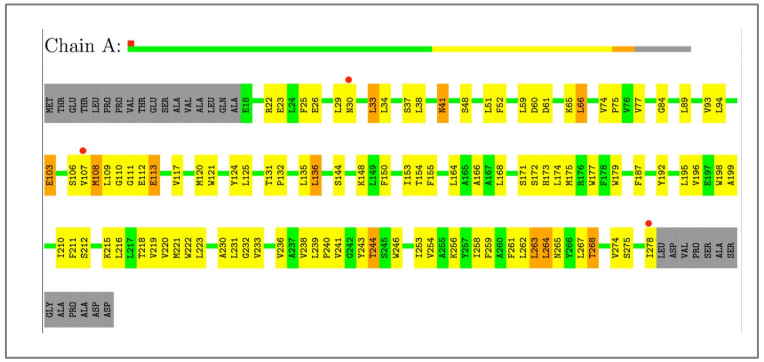
1st Generation was simple

- Model Quality
  - Bond lengths and angles (outlier info, RMS-Z)
  - Chirality, planarity
  - Close contacts (including worst clashes, MolProbity clash score)
  - Torsion angles (Ramachandran statistics, protein rotamers)
  - Ligand geometry (Mogul analysis)
- Residue Plots
  - Residues with model-quality outliers (0, 1, 2, >2)
  - Residues with RSR-Z > 5 are highlighted
  - Residues not observed

Overall Quality Summary

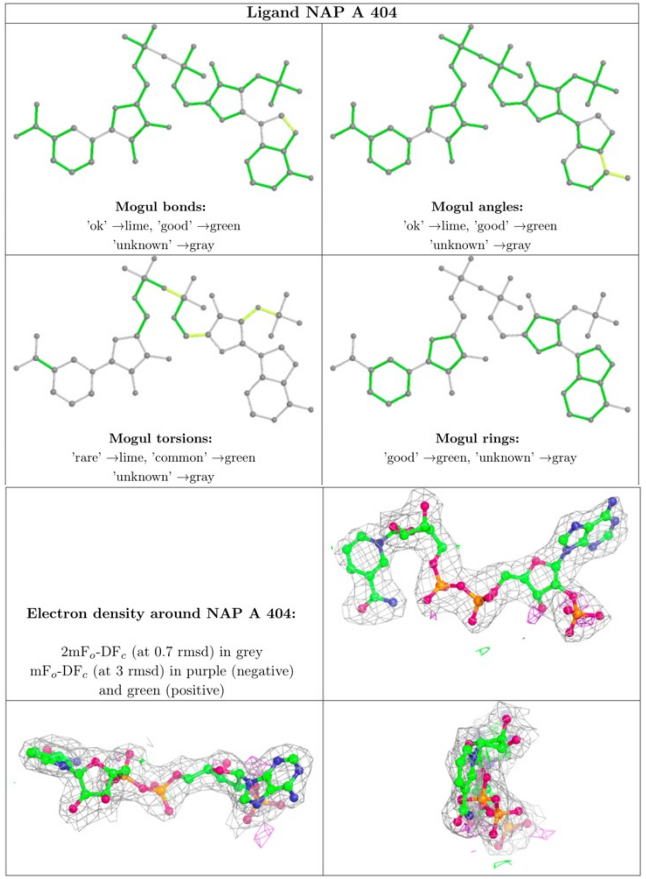


Residue Plots



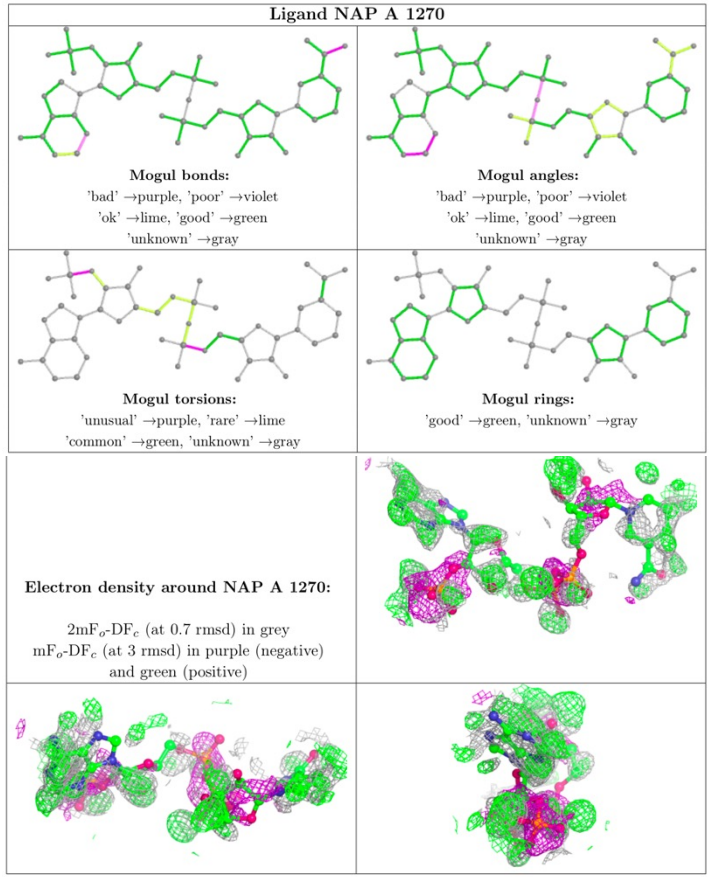
2nd Generation includes 2D and 3D information

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	NAP	A	404	48/48	0.96	0.14	31,43,66,70	0



PDB entry 5zix (Better data quality)

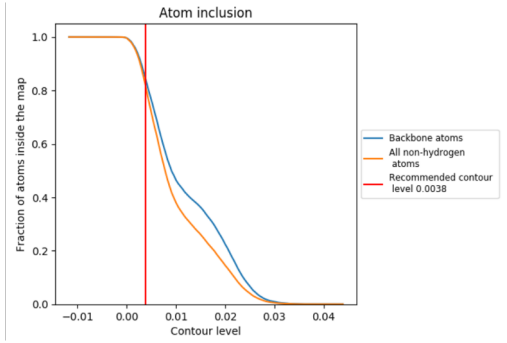
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	NAP	A	1270	48/48	-0.06	0.67	87,96,100,100	0

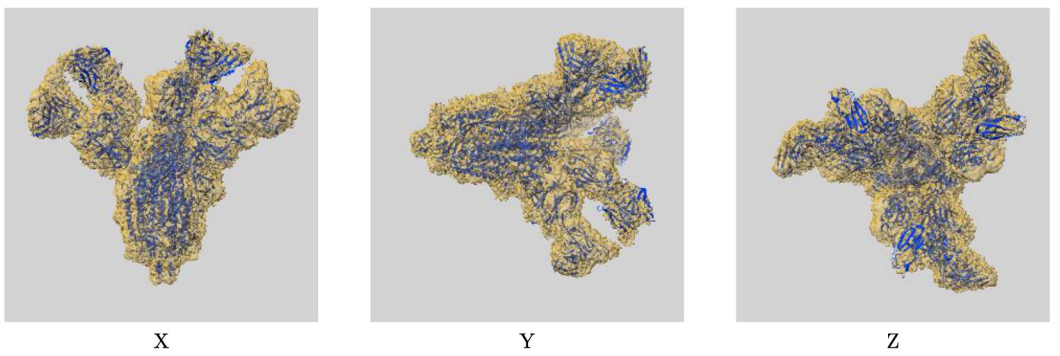


PDB entry 1zk4 (Worse data quality)

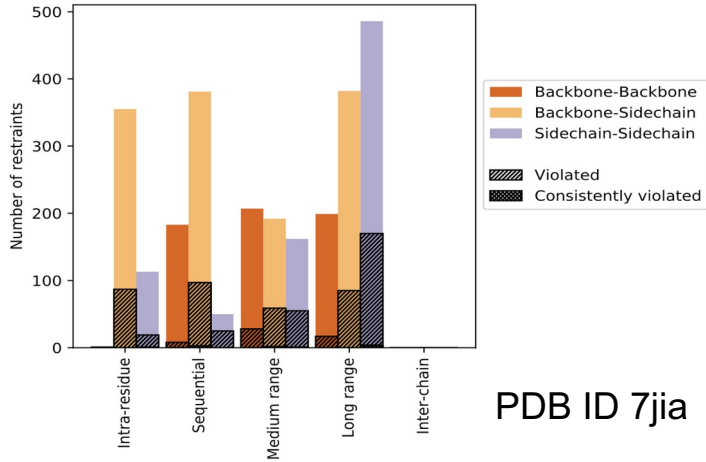
3rd Generation includes 3DEM or NMR

- EM map volume analysis
  - The fit of model to the map at residue level and global visual overlay
  - Map analysis and visualization
- NMR restraints assessments
  - Distance and dihedral angle restraints with graphical and tabular statistics
  - Available for restraints deposited in single NEF/NMR-STAR formats





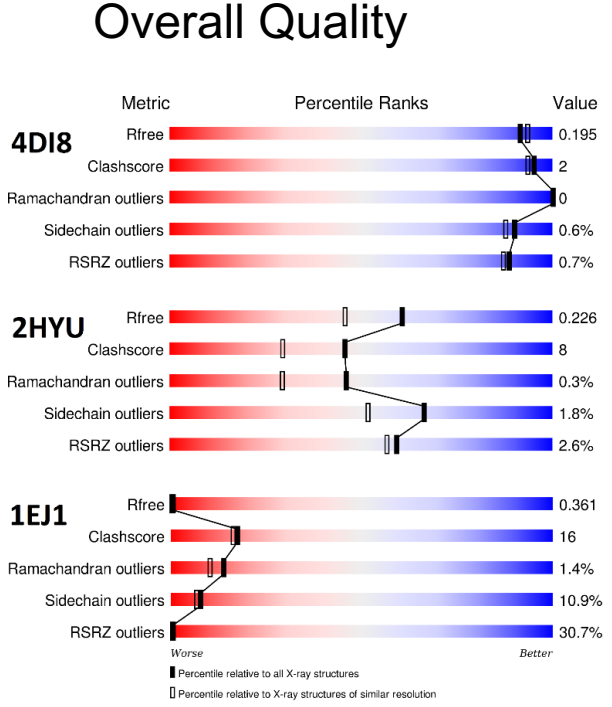
PDB ID 7cwu

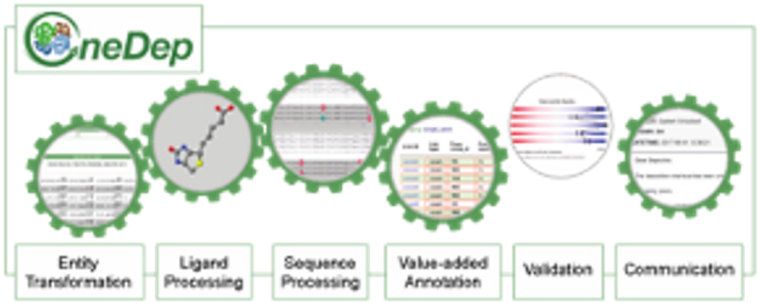


PDB ID 7jia

Validation report keeps updated!

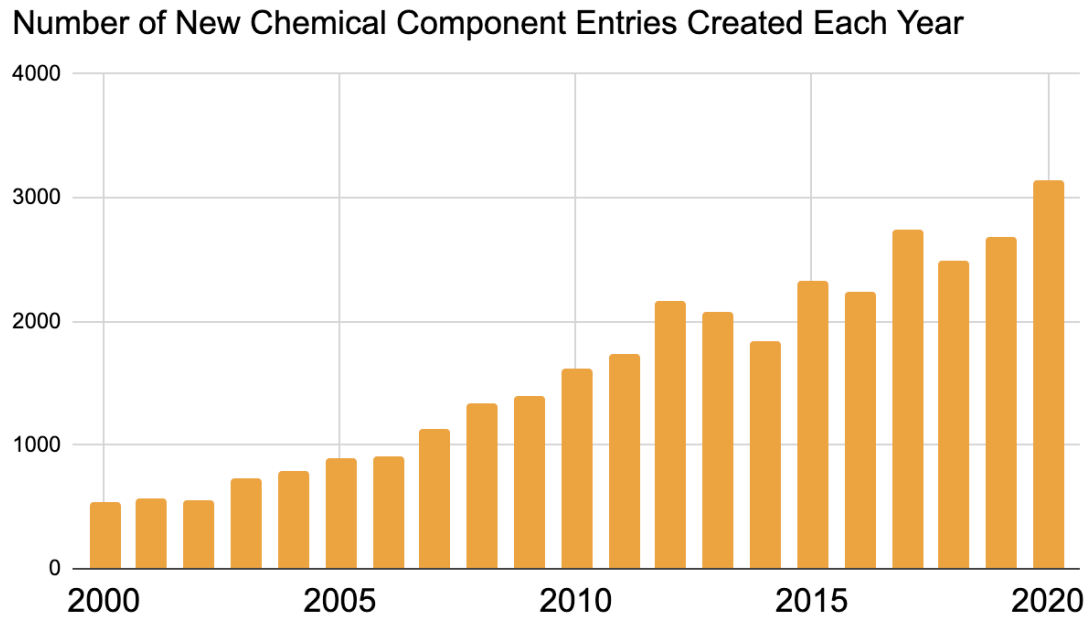
- Improved EM validation with Q-score
- Deposition of Half-maps for EM entries now Mandatory
- Enhanced Validation of Small-Molecule Ligands and Carbohydrates
- NMR Restraints Validation Available through OneDep





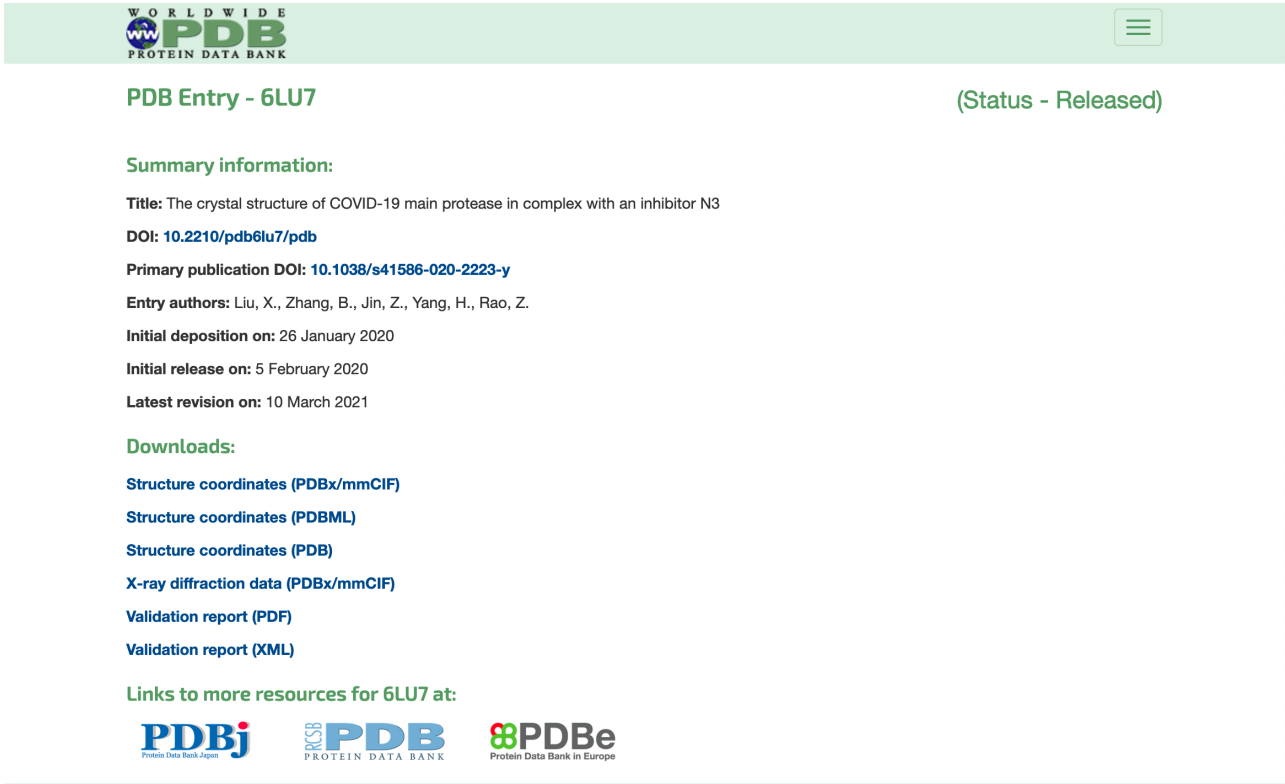
Announcement 1 ：
CCD ID becomes 5-letters

- 3-letter ID codes of the Chemical Component Dictionary (CCD) will run out in 2023 or 24
- Only the PDBx/mmCIF format will be provided for CCD IDs with 5-letters.



Announcement 2 :
PDB ID becomes 8-letters

- 4-letter PDB IDs will also run out soon.
- The wwPDB already started to serve 8-letter PDB IDs. e.g. wwpdb.org/pdb?id=**pdb_00006lu7**



Please deposit your raw data!

PDBj Protein Data Bank Japan
as a member of WORLDWIDE PDB PROTEIN DATA BANK

IUCrJ
ISSN 2052-2525
BIOLOGY | MEDICINE

Findable Accessible Interoperable Re-usable (FAIR) diffraction data are coming to protein crystallography

John R. Helliwell,[a]‡ Wladek Minor,[b]§ Manfred S. Weiss,[c]¶ Elspeth F. Garman,[d]‡‡ Randy J. Read,[e]‡‡ Janet Newman,[f]§§ Mark J. van Raaij,[g]§§ Janos Hajdu[h,i]¶¶ and Edward N. Baker[j]‡‡‡

[a]School of Chemistry, The University of Manchester, Brunswick Street, Manchester M13 9PL, United Kingdom, [b]Department of Molecular Physiology and Biological Physics, University of Virginia, 1340 Jefferson Park Avenue Pinn Hall, Charlottesville, VA 22908-0736, USA, [c]Macromolecular Crystallography (HZB-MX), Helmholtz-Zentrum Berlin, Albert-Einstein-Str. 15, D-12489 Berlin, Germany, [d]Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, United Kingdom, [e]Cambridge Institute for Medical Research, Department of Haematology, University of Cambridge, The Keith Peters Building, Hills Road, Cambridge CB2 0XY, United Kingdom, [f]Collaborative Crystallisation Centre (C3), CSIRO, 343 Royal Parade, Parkville, VIC 3052, Australia, [g]CSIC, Centro Nacional de Biotecnologia, c/Darwin 3, Madrid, 28049, Spain, [h]Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, Box 596, Uppsala, 75124, Sweden, [i]The European Extreme Light Infrastructure, Institute of Physics, AS CR, Na Slovance 2, Prague 18221 8, Czech Republic, and [j]School of Biological Sciences, University of Auckland, School of Biological Sciences, Private Bag 92-019, Auckland, New Zealand

‡ Chairman of the IUCr Committee on Data.
§ Chairman of the IUCr Commission on Biological Macromolecules.
¶ Member of the IUCr Commission on Biological Macromolecules.
‡‡ Section Editor *Acta Cryst. D*.
§§ Section Editor *Acta Cryst. F*.
¶¶ Main Editor *Journal of Applied Crystallography*.
‡‡‡ Main Editor *IUCrJ*.

Keywords: FAIR; diffraction data; IUCr policy.

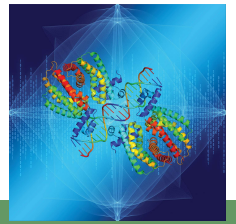
The unprecedented progress of modern science is driven, to a large extent, by the fast propagation of information. Descriptions of experiments and results, and their interpretation, are no longer disseminated solely in peer-reviewed scientific publications, but are frequently distributed through non-reviewed publication platforms as preprints, entries to data repositories, databases *etc.* As a result of ever faster computers and internet connections, many experimental results are now available instantaneously at the click of a mouse, irrespective of the location of the source or consumer.

In many instances, experiments performed and interpreted by one scientific group stimulate the interest of other scientists enough to spur research in further laboratories. Not infrequently, the results of these follow-up experiments are in disagreement with the previously obtained results and/or interpretations (Baker, 2016), notably in psychology and the clinical sciences. In some cases, the original results cannot even be reproduced well enough to allow follow-up experiments to commence (Prinz *et al.*, 2011).

Repeating an entire experiment performed by others is usually not feasible because of the significant time, effort and funds it would require (Baker, 2015). So the question is, what should be done in this new era? How can new technical developments be best exploited for furthering science and the scientific output?

The structural biology community has always been at the forefront of sharing processed, *i.e.* analysed, results. Since its creation in 1971, the Protein Data Bank (PDB; Berman *et al.*, 2000) has become an indispensable daily resource for hundreds of thousands of scientists. Initially, the PDB curated only the molecular structure coordinate files, but since 2008 the deposition of the processed diffraction data, *i.e.* intensities or structure-factor amplitudes, has been mandatory for each derived coordinate set. At present, all serious scientific journals require the deposition of the coordinates of the structures and the associated diffraction data as well as the submission of a PDB validation report with the manuscript for review. Notable also is a recent initiative by *Science* of the introduction of a Statistical Board of Reviewing Editors (McNutt, 2014*a*,*b*). This is an initiative similar to the practice of some referees insisting on access to the underpinning crystallographic data (Helliwell, 2018). Certainly, the PDB is an indispensable resource not only for structural biology but for all modern biological, biomedical and biochemical science (Burley *et al.*, 2019).

However, even with diffraction data being a part of every macromolecular crystallographic deposition in the PDB, and even assuming 'perfect' data reduction and processing of the original diffraction images, some experimental information, *e.g.* diffuse



Acta Cryst D STRUCTURAL BIOLOGY
ISSN 2059-7983

A public database of macromolecular diffraction experiments

Marek Grabowski,[a]‡ Karol M. Langner,[a]‡§ Marcin Cymborowski,[a] Przemyslaw J. Porebski,[a,b] Piotr Sroka,[a] Heping Zheng,[a] David R. Cooper,[a] Matthew D. Zimmerman,[a] Marc-André Elsliger,[c] Stephen K. Burley[d,e] and Wladek Minor[a]*

[a]Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA 22904, USA, [b]Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, Niezapominajek 8, 30-239 Cracow, Poland, [c]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 90237, USA, [d]RCSB Protein Data Bank; Center for Integrative Proteomics Research; Institute for Quantitative Biomedicine; Rutgers Cancer Institute of New Jersey; Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA, and [e]San Diego Supercomputer Center and Skaggs School of Pharmacological Sciences, University of California, San Diego, La Jolla, CA 92093, USA. *Correspondence e-mail: wladek@iwonka.med.virginia.edu

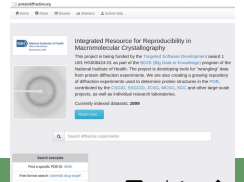
Received 23 June 2016
Accepted 17 September 2016

Edited by T. O. Yeates, University of California, USA

‡ The first two authors contributed equally.
§ Present address: Google Inc., Mountain View, CA 94043, USA.

Keywords: diffraction experiment; protein crystallography; repository; data; metadata; IRRMC.

The low reproducibility of published experimental results in many scientific disciplines has recently garnered negative attention in scientific journals and the general media. Public transparency, including the availability of 'raw' experimental data, will help to address growing concerns regarding scientific integrity. Macromolecular X-ray crystallography has led the way in requiring the public dissemination of atomic coordinates and a wealth of experimental data, making the field one of the most reproducible in the biological sciences. However, there remains no mandate for public disclosure of the original diffraction data. The Integrated Resource for Reproducibility in Macromolecular Crystallography (IRRMC) has been developed to archive raw data from diffraction experiments and, equally importantly, to provide related metadata. Currently, the database of our resource contains data from 2920 macromolecular diffraction experiments (5767 data sets), accounting for around 3% of all depositions in the Protein Data Bank (PDB), with their corresponding partially curated metadata. IRRMC utilizes distributed storage implemented using a federated architecture of many independent storage servers, which provides both scalability and sustainability. The resource, which is accessible *via* the web portal at http://www.proteindiffraction.org, can be searched using various criteria. All data are available for unrestricted access and download. The resource serves as a proof of concept and demonstrates the feasibility of archiving raw diffraction data and associated metadata from X-ray crystallographic studies of biological macromolecules. The goal is to expand this resource and include data sets that failed to yield X-ray structures in order to facilitate collaborative efforts that will improve protein structure-determination methods and to ensure the availability of 'orphan' data left behind for various reasons by individual investigators and/or extinct structural genomics projects.



1. Introduction

Issues with the reproducibility of published experimental results have recently attracted attention in many different scientific fields (Collins & Tabak, 2014). The lack of availability of original, primary scientific data represents a major factor contributing to reproducibility problems (Iqbal *et al.*,

New Archive for Xtal Diffraction Images

PDBj Protein Data Bank Japan — as a member of wwPDB PROTEIN DATA BANK

XRDa X-Ray Diffraction Archive — English 日本語 — Login using ORCID — ヘルプ エントリー探検 統計情報

Help menu

- Help top page
- How to submit a new entry
- File upload system
- How to set a graphical abstract
- How to request publication of an entry
- How to download data

XRDaへようこそ

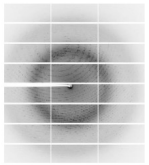
The X-Ray Diffraction Archive (XRDa for short) top page.

We welcome depositions of raw X-ray diffraction images corresponding to PDB entries.
To submit new entries, please login using your ORCID ID.

If you have any questions, please contact us.
Please note that this archive is still under development and thus we appreciate any feedback you might have.

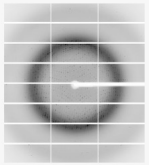
2020-04-09 (last edited: 3 months ago)

Latest entries | All entries | Covid-19 entries



7V1X: Difructose dianhydride I synthase/hydrolase (alphaFFase1) from Bifidobacterium dentium in complex with beta-D-fructofuranose

Structure resolution: 1.76 Å
Kashima, T., Arakawa, T., Yamada, C., Fujita, K., Fushinobu, S.
DOI: 10.1016/j.jbc.2021.101324
Deposition date: 2021-11-16
Release date: 2021-11-16



7DNN: Crystal structure of the AgCarB2-C2 complex with homoorientin

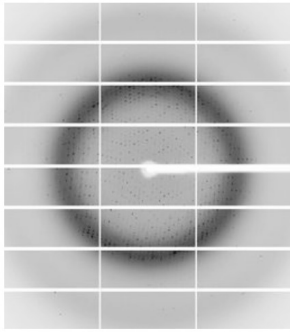
Structure resolution: 2.25 Å
Senda, M., Kumano, T., Watanabe, S., Kobayashi, M., Senda, T.
Deposition date: 2021-08-19
Release date: 2021-10-20

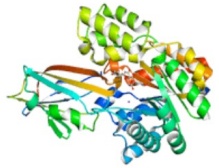
5AUI: Crystal structure of Ferredoxin

XRDa X-Ray Diffraction Archive — English 日本語 — Login using ORCID — ヘルプ エントリー探検 統計情報

7dnn: Crystal structure of the AgCarB2-C2 complex with homoorientin

Authors:	Senda, M., Kumano, T., Watanabe, S., Kobayashi, M., Senda, T.
R-work:	0.19630
R-free:	0.24710
Unit cell edges (Å):	73.905 x 102.511 x 136.09
Unit cell angles (°):	90, 90, 90
Resolution:	47.97 Å - 2.25 Å
Space group:	P 21 21 2
Primary citation	
PDBj website for 7dnn	
Entry:	Download (1.03 GB)





Dataset CarB homologue3-homoorientin complex

Number of frames	440
Distance (mm)	257.1
Oscillation width (°)	0.5
Wavelength (Å)	1
Equipment	DECTRIS PILATUS 2M-F
Beamline	PHOTON FACTORY BEAMLINE AR-NE3A

ファイルマネージャー

Path: /

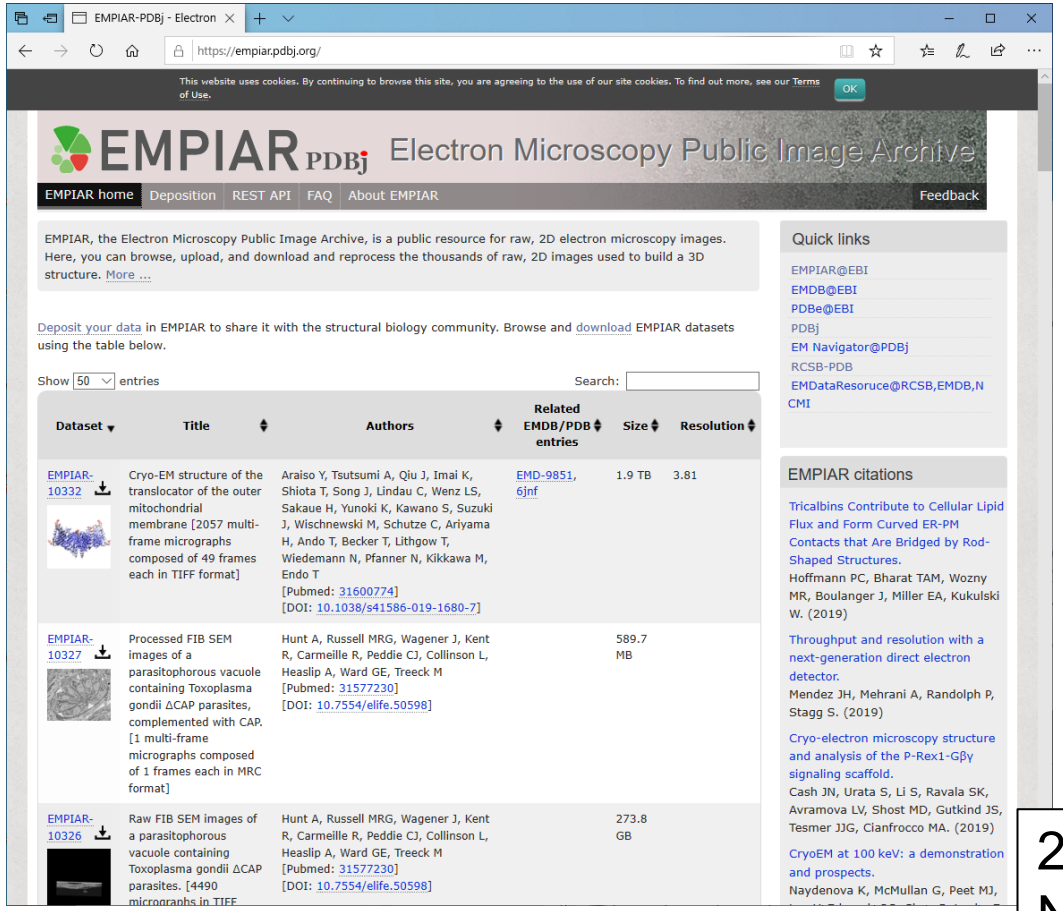
	名	説明	サイズ	変更日
	data			

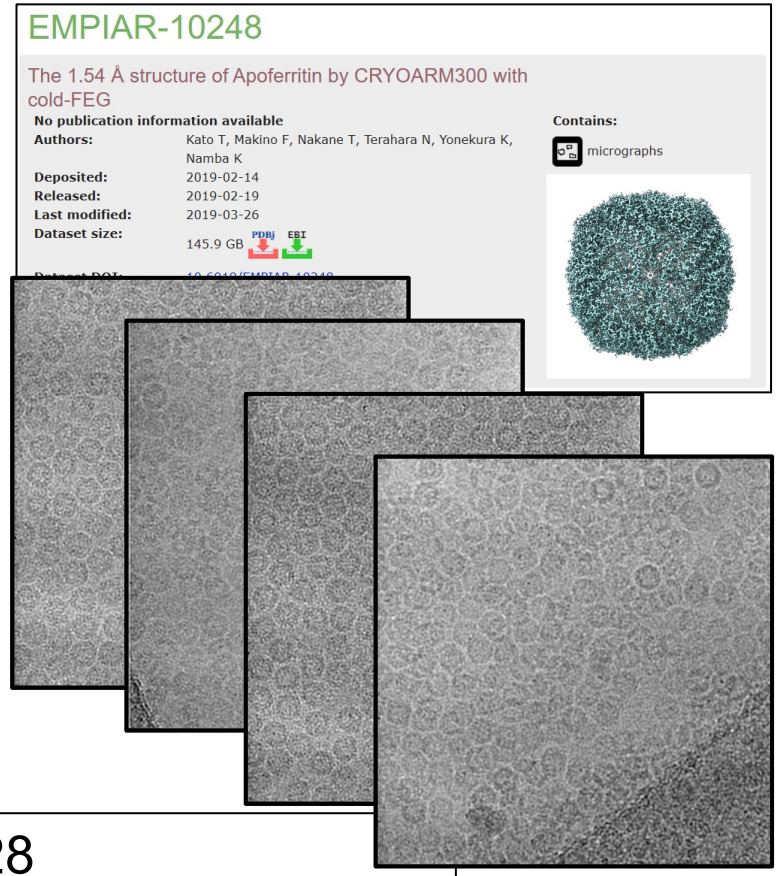
https://xrda.pdbj.org

NEW

EMPIAR-PDBj started from Dec. 2018

PDBj Protein Data Bank Japan
as a member of WORLDWIDE PDB PROTEIN DATA BANK





2023/4/28
Number of entries: 1280
Total data size : 2.6 PB

https://empiar.pdbj.org

Acknowledgements

Staff

- **Head**
  - Kurisu, Genji, Ph.D. (Prof., IPR, Osaka Univ.)
- **Group for Data-in of PDB/EMDB**
  - Nakagawa, Atsushi, Ph.D. (Prof., IPR, Osaka Univ.)
  - Yu, Jian, Ph.D. (Assoc. Prof., IPR, Osaka Univ.)
  - Kengaku, Yumiko (IPR, Osaka Univ.)
  - Sato, Junko (IPR, Osaka Univ.)
  - Kim, Ju Yaen, Ph.D. (IPR, Osaka Univ.)
  - Niwa, Satomi, Ph.D. (IPR, Osaka Univ.)
  - Takuwa, Ayako, Ph.D. (IPR, Osaka Univ.)
- **Group for Data-out of PDB/EMDB**
  - Mizuguchi, Kenji, Ph.D. (Prof., IPR, Osaka Univ.)
  - Bekker, Gert-Jan, Ph.D. (Lecturer, IPR, Osaka Univ.)
  - Nagao, Chioko, Ph.D. (Assist. Prof., IPR, Osaka Univ.)
  - Yamashita, Reiko (IPR, Osaka Univ.)
  - Kudou, Takahiro (IPR, Osaka Univ.)
- **Group for BMRB**
  - Kojima, Chojiro, Ph.D. (Prof., Yokohama National Univ.)
  - Miyanoiri, Youhei, Ph.D. (Assoc. Prof., IPR, Osaka Univ.)
  - Iwata, Takeshi (IPR, Osaka Univ.)
- **Group for EMPIAR**
  - Nakane, Takanori, MD, Ph.D. (Assoc. Prof., IPR, Osaka Univ.)
- **PRF branch office**
  - Kurisu, Genji, Ph.D. (Guest Researcher, Protein Research Foundation (PRF))
  - Isoyama, Masaharu (Informatics division, Protein Research Foundation (PRF))
  - Cho, Hasumi, Ph.D. (Protein Research Foundation (PRF))
  - Tsunezumi, Kiyo (Protein Research Foundation (PRF))
  - Yokochi, Masashi (Protein Research Foundation (PRF))
  - Ikegawa, Yasuyo (Protein Research Foundation (PRF))
- **Collaboratory Researchers**
  - Nakamura, Haruki, Ph.D. (Prof. Emer., IPR, Osaka Univ.)
  - Fujiwara, Toshimichi, Ph.D. (Prof. Emer., IPR, Osaka Univ.)
  - Yura, Kei, Ph.D. (Humanities/Natural Science Integrated AI Data Science Center, Ochanomizu University) for for EMPIAR-PDBj
  - Toh, Hiroyuki, Ph.D. (Prof., Kwansei Gakuin Univ.)
  - Kawabata, Takeshi, Ph.D. (Assoc. Prof., Tohoku Univ.) for HOMCOS, gmfit and EMPIAR-PDBj
  - Suzuki, Hirofumi, Ph.D. (Ex-Assist. Prof., IPR, Osaka Univ.) for EM Navigator
  - Kobayashi, Naohiro, Ph.D. (RIKEN SPring-8)
  - Wako, Hiroshi, Ph.D. (Prof. Emer., Waseda Univ.) for ProMode
  - Endo, Shigeru, Ph.D. (Assoc. Prof., Osaka Univ.) for ProMode
  - Ito, Nobutoshi, Ph.D. (Prof., Tokyo Medical and Dental Univ.)
  - Kinoshita, Kengo, Ph.D. (Prof., Tohoku Univ.) for eF-site
  - Standley, Daron, Ph.D. (Prof., RIMD, Osaka Univ.) for SeqNavi, DASH, SeSAW, and ASH
  - Katoh, Kazutaka, Ph.D. (Assoc. Prof., RIMD, Osaka Univ.) for MAFFTash
- **Secretary**
  - Sakuma, Ryoko (IPR, Osaka Univ.)

Created: 2020-12-02 (last edited: more than 1 year ago)

WORLDWIDE PDB PROTEIN DATA BANK

The worldwide Protein Data Bank

www.wwPDB.org • info@wwPDB.org

PDBj Protein Data Bank Japan

as a member of WORLDWIDE PDB PROTEIN DATA BANK

RCSB PDB PROTEIN DATA BANK | BMRB Biological Magnetic Resonance Data Bank | PDBe Protein Data Bank in Europe | EMDB Electron Microscopy Data Bank | PDBj Protein Data Bank Japan