X-tal Raw Data Archive (XRDa):
A crystallographic raw diffraction image archive in Asia

Gert-Jan Bekker, *Genji Kurisu
Institute for Protein Research, Osaka University

In 1971:

CRYSTALLOGRAPHY

Protein Data Bank

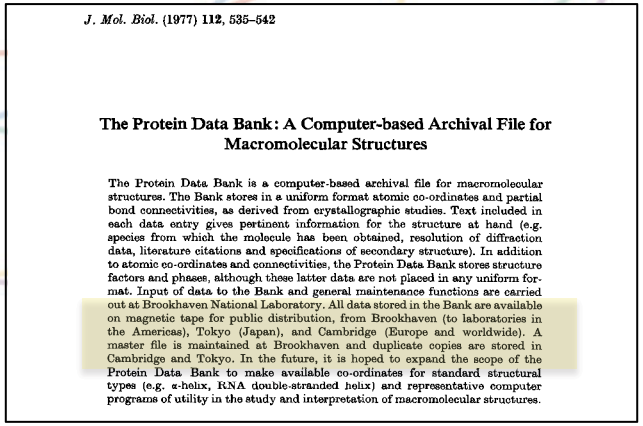
A repository system for protein crystallographic data will be operated jointly by the Crystallographic Data Centre, Cambridge, and the Brookhaven National Laboratory. The system will be responsible for storing atomic coordinates, structure factors and electron density maps and will make these data available 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)

Cytochrome c @4.5 Å in 1971 and @2.3 Å in 1975, which was the 1st Asian and 21st PDB entry in the entire archive.

J. Biochem. (Tokyo) **70**, 913-924 (1971)
J. Biochem. (Tokyo) **77**, 147-162 (1975)

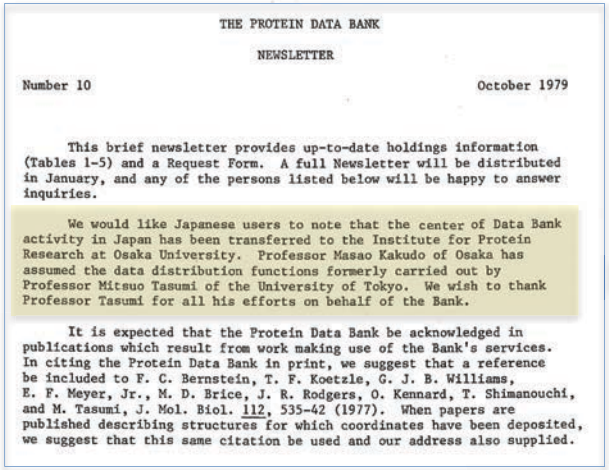
Early Data-out services of PDB in Japan



J. Mol. Biol. (1977) **112**, 535–542

The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures

The Protein Data Bank is a computer-based archival file for macromolecular structures. The Bank stores in a uniform format atomic co-ordinates and partial bond connectivities, as derived from crystallographic studies. Text included in each data entry gives pertinent information for the structure at hand (e.g. species from which the molecule has been obtained, resolution of diffraction data, literature citations and specifications of secondary structure). In addition to atomic co-ordinates and connectivities, the Protein Data Bank stores structure factors and phases, although these latter data are not placed in any uniform format. Input of data to the Bank and general maintenance functions are carried out at Brookhaven National Laboratory. All data stored in the Bank are available on magnetic tape for public distribution, from Brookhaven (to laboratories in the Americas), Tokyo (Japan), and Cambridge (Europe and worldwide). A master file is maintained at Brookhaven and duplicate copies are stored in Cambridge and Tokyo. In the future, it is hoped to expand the scope of the Protein Data Bank to make available co-ordinates for standard structural types (e.g. α-helix, RNA double-stranded helix) and representative computer programs of utility in the study and interpretation of macromolecular structures.



THE PROTEIN DATA BANK

NEWSLETTER

Number 10 October 1979

This brief newsletter provides up-to-date holdings information (Tables 1-5) and a Request Form. A full Newsletter will be distributed in January, and any of the persons listed below will be happy to answer inquiries.

We would like Japanese users to note that the center of Data Bank activity in Japan has been transferred to the Institute for Protein Research at Osaka University. Professor Masao Kakudo of Osaka has assumed the data distribution functions formerly carried out by Professor Mitsuo Tasumi of the University of Tokyo. We wish to thank Professor Tasumi for all his efforts on behalf of the Bank.

It is expected that the Protein Data Bank be acknowledged in publications which result from work making use of the Bank's services. In citing the Protein Data Bank in print, we suggest that a reference be included to F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, J. Mol. Biol. 112, 535-42 (1977). When papers are published describing structures for which coordinates have been deposited, we suggest that this same citation be used and our address also supplied.

Table 24.1.3.2. *PDB mirror sites as of May 1999*

Official PDB mirror sites
Argentina: University of San Luis
Australia: Australian National Genomic Information Service, Sydney; The Walter and Eliza Hall Institute of Medical Research, Melbourne
Brazil: ICB UFMG, Inst. de Ciencias Biologicas, Univ. Federal de Minas Gerais
China: Institute of Physical Chemistry, Peking University, Beijing
France: Institut de Génétique Humaine, Montpellier
Germany: GMD, German National Research Center for Information Technology, Sankt Augustin
India: Bioinformatics Centre, University of Pune
Israel: Weizmann Institute of Science, Rehovot
Japan: Institute of Protein Research, Osaka University
Poland: ICM Interdisciplinary Centre for Modelling, Warsaw University
Taiwan: National Tsing Hua University, HsinChu
United Kingdom: Cambridge Crystallographic Data Centre, Cambridge; EMBL Outstation, EBI, Hinxton
United States: Bio Molecular Engineering Research Center, Boston University; North Carolina Supercomputing Center, Research Triangle Park; University of Georgia, Athens, Georgia; PDB at Brookhaven National Laboratory

J. Mol. Biol., **112**, 535-342, 1977

PDB Newsletter, No.10, Oct. 1979

Inter. Table of Cryst. Vol F, 1999

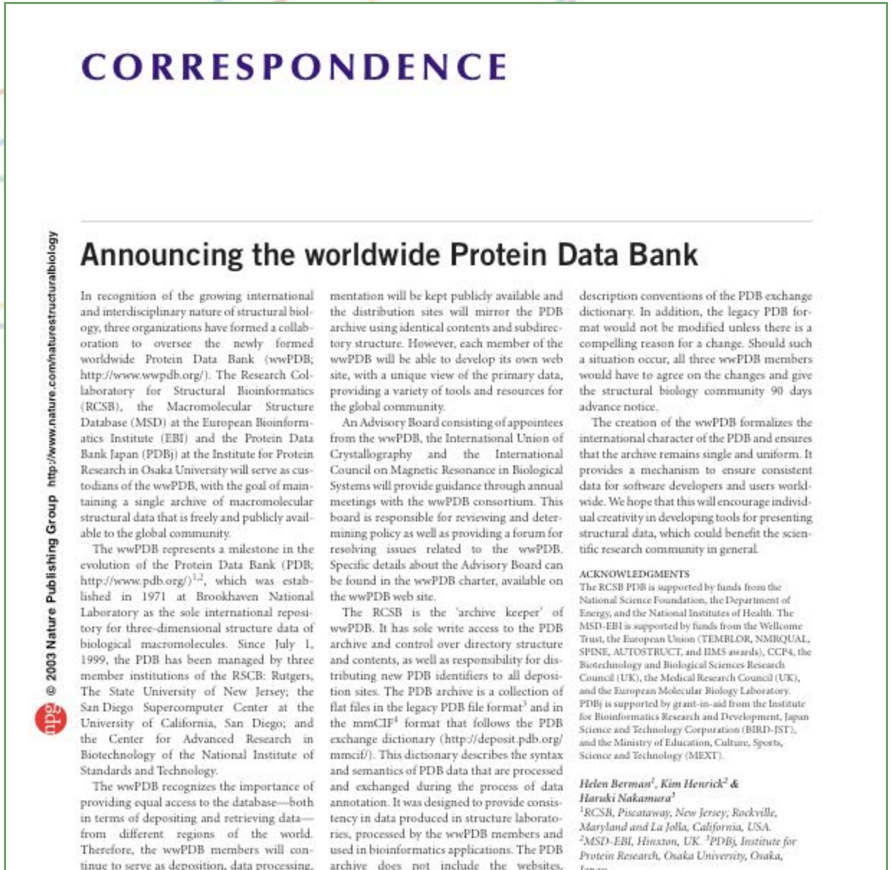
Magnetic Tape ➡ **CD-ROM** ➡ **Internet** (BNL PDB mirror site @Osaka 1998)

PDBj was founded in 2000

Data-in and Data-out services in Japan.

https://pdbj.org. in English, Japanese, Chinese and Korean

wwPDB was founded in 2003



CORRESPONDENCE

Announcing the worldwide Protein Data Bank

In recognition of the growing international and interdisciplinary nature of structural biology, three organizations have formed a collaboration to oversee the newly formed worldwide Protein Data Bank (wwPDB; http://www.wwpdb.org/). The Research Collaboratory for Structural Bioinformatics (RCSB), the Macromolecular Structure Database (MSD) at the European Bioinformatics Institute (EBI) and the Protein Data Bank Japan (PDBj) at the Institute for Protein Research in Osaka University will serve as custodians of the wwPDB, with the goal of maintaining a single archive of macromolecular structural data that is freely and publicly available to the global community.

The wwPDB represents a milestone in the evolution of the Protein Data Bank (PDB; http://www.pdb.org/)[1,2], which was established in 1971 at Brookhaven National Laboratory as the sole international repository for three-dimensional structure data of biological macromolecules. Since July 1, 1999, the PDB has been managed by three member institutions of the RSCB: Rutgers, The State University of New Jersey; the San Diego Supercomputer Center at the University of California, San Diego; and the Center for Advanced Research in Biotechnology of the National Institute of Standards and Technology.

The wwPDB recognizes the importance of providing equal access to the database—both in terms of depositing and retrieving data—from different regions of the world. Therefore, the wwPDB members will continue to serve as deposition, data processing, mentation will be kept publicly available and the distribution sites will mirror the PDB archive using identical contents and subdirectory structure. However, each member of the wwPDB will be able to develop its own web site, with a unique view of the primary data, providing a variety of tools and resources for the global community.

An Advisory Board consisting of appointees from the wwPDB, the International Union of Crystallography and the International Council on Magnetic Resonance in Biological Systems will provide guidance through annual meetings with the wwPDB consortium. This board is responsible for reviewing and determining policy as well as providing a forum for resolving issues related to the wwPDB. Specific details about the Advisory Board can be found in the wwPDB charter, available on the wwPDB web site.

The RCSB is the 'archive keeper' of wwPDB. It has sole write access to the PDB archive and control over directory structure and contents, as well as responsibility for distributing new PDB identifiers to all deposition sites. The PDB archive is a collection of flat files in the legacy PDB file format[3] and in the mmCIF[4] format that follows the PDB exchange dictionary (http://deposit.pdb.org/mmcif/). This dictionary describes the syntax and semantics of PDB data that are processed and exchanged during the process of data annotation. It was designed to provide consistency in data produced in structure laboratories, processed by the wwPDB members and used in bioinformatics applications. The PDB archive does not include the websites, description conventions of the PDB exchange dictionary. In addition, the legacy PDB format would not be modified unless there is a compelling reason for a change. Should such a situation occur, all three wwPDB members would have to agree on the changes and give the structural biology community 90 days advance notice.

The creation of the wwPDB formalizes the international character of the PDB and ensures that the archive remains single and uniform. It provides a mechanism to ensure consistent data for software developers and users worldwide. We hope that this will encourage individual creativity in developing tools for presenting structural data, which could benefit the scientific research community in general.

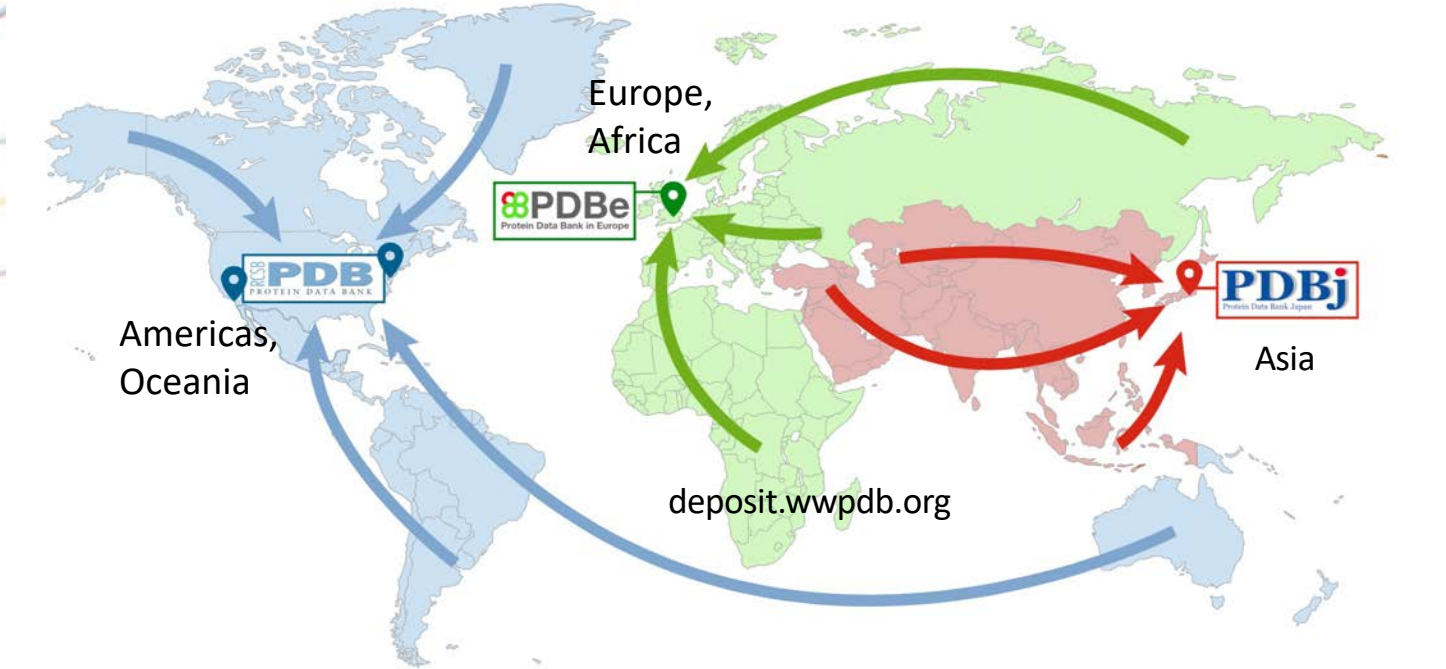
ACKNOWLEDGMENTS
The RCSB PDB is supported by funds from the National Science Foundation, the Department of Energy, and the National Institutes of Health. The MSD-EBI is supported by funds from the Wellcome Trust, the European Union (TEMBLOR, NMRQUAL, SPINE, AUTOSTRUCT, and IIMS awards), CCP4, the Biotechnology and Biological Sciences Research Council (UK), the Medical Research Council (UK), and the European Molecular Biology Laboratory. PDBj is supported by grant-in-aid from the Institute for Bioinformatics Research and Development, Japan Science and Technology Corporation (BIRD-JST), and the Ministry of Education, Culture, Sports, Science and Technology (MEXT).

Helen Berman[1], Kim Henrick[2] & Haruki Nakamura[3]
[1]RCSB, Piscataway, New Jersey; Rockville, Maryland and La Jolla, California, USA. [2]MSD-EBI, Hinxton, UK. [3]PDBj, Institute for Protein Research, Osaka University, Osaka,

Nature Structural Biology **10**, 980 (2003)
doi: 10.1038/nsb1203-980

Helen Berman
Kim Henrick Haruki Nakamura

PDBj is in charge of processing and annotating the depositions from Asia



*PDB China started contributing to Data Processing of Asian depositions from 2022.
Acta Cryst. D. available *online ahead of print* DOI:10.1107/S2059798323006381

The history of PDBj from 2000

PDBj founded

SF become mandatory

CS become mandatory

PDB40 Symposium

OneDep at PDBj

PDBx/mmCIF become mandatory

PROTEIN DATA BANK 20 Years

2000 01 02 03 04 05 06 07 08 09 2010 11 12 13 14 15 16 17 18 19 2020 21 23

Launch of *Promode service* at PDBj

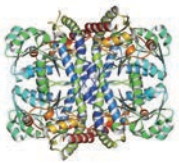
Launch of *eF-site service* at PDBj

20,000 depositions at PDBj

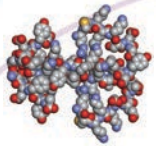
20th Anniversary of PDBj,

50,000 depositions at PDBj

The 1st MX structure processed by PDBj was 1GC0 (L-Met γ-lyase), deposited from Japan



The 1st NMR structure processed by PDBj was 1EJQ (Syndecan-4), deposited from Korea



Structural Biology Data Archiving in Asia

1. NMR Spectroscopy

WORLDWIDE PDB
PROTEIN DATA BANK

Spectral Data

Chemical shifts
Coupling constants
Time-domain data
Spectral peak lists

Relaxation Data

R1/T1, R2/T2, R1rho/T1rho, NOEs

BMRBj
Biological Magnetic Resonance Data Bank Japan

→

WORLDWIDE PDB
PROTEIN DATA BANK

Coordinates (XYZ)

PDB Core Archive with the metadata including the links to other databases & source of proteins

PDBj
Protein Data Bank Japan

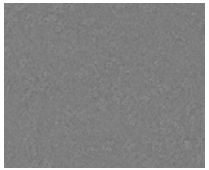
Structural Biology Data Archiving in Asia

2. Cryo Electron Microscopy

2D Images → **3D Maps** → **Coordinates (XYZ)**

EMPIAR

Raw 2D images of Cryo-EM including motion correction movies (~5 TB/entry)

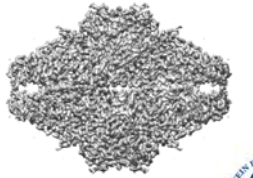


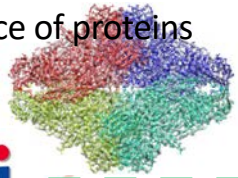
PDBj Protein Data Bank Japan

WORLDWIDE PDB PROTEIN DATA BANK

EMDB Electron Microscopy Data Bank

Core Archives with the metadata including the links to other databases, experimental condition & source of proteins



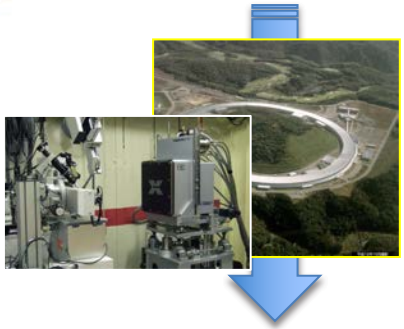


PDBj Protein Data Bank Japan

Structural Biology Data Archiving in Asia

3. X-ray Crystallography

2D raw images → Structure Factors → Coordinates (XYZ)



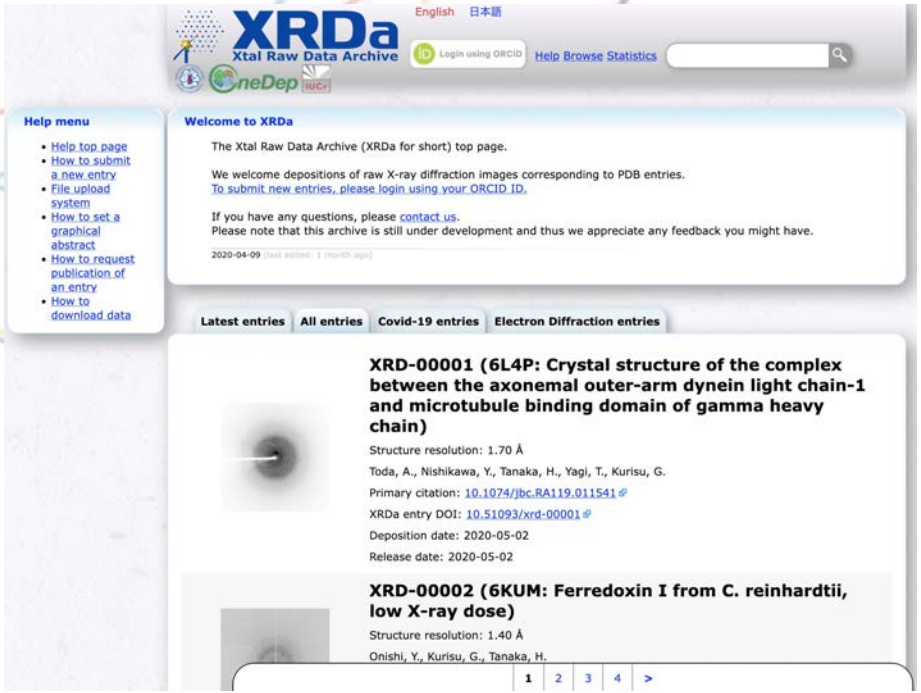
PDB Core Archive with the metadata including the links to other databases, experimental condition & source of proteins

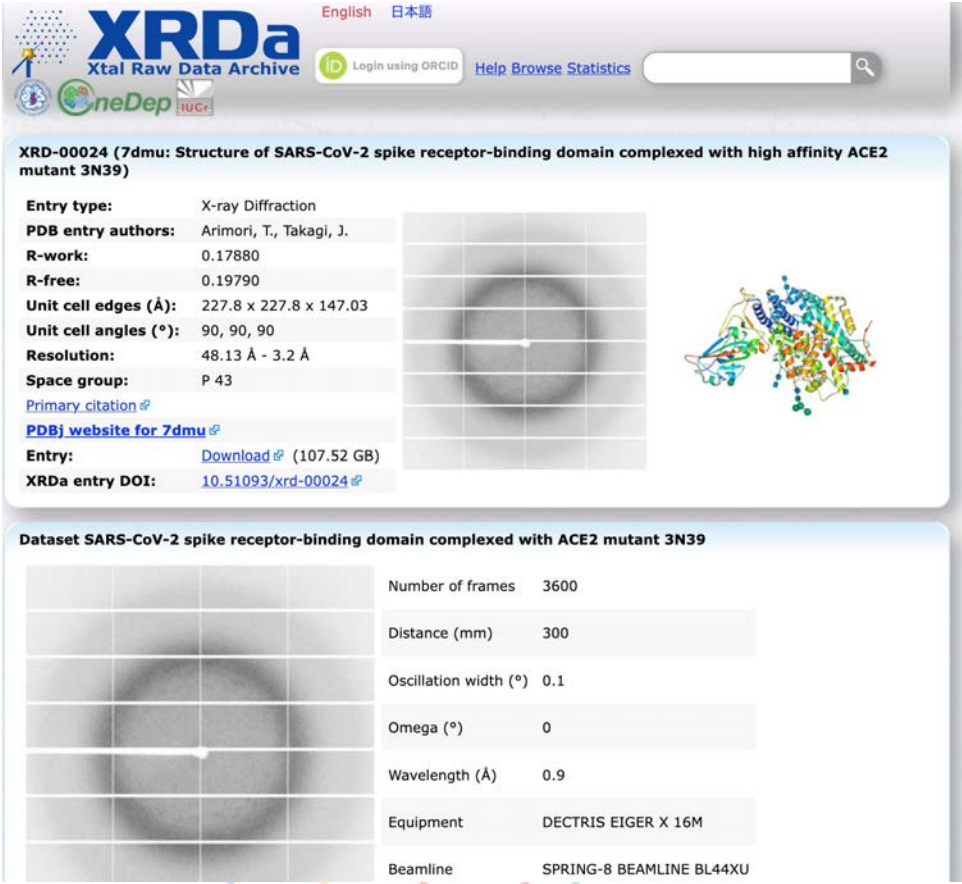
- ProteinDiffraction.org - https://proteindiffraction.org
- SBGrid - https://sbgrid.org
- CXIDB - http://www.cxidb.org

No public database for X-ray images available in Asia

Archive for X-ray Diffraction Images

ORIGINAL





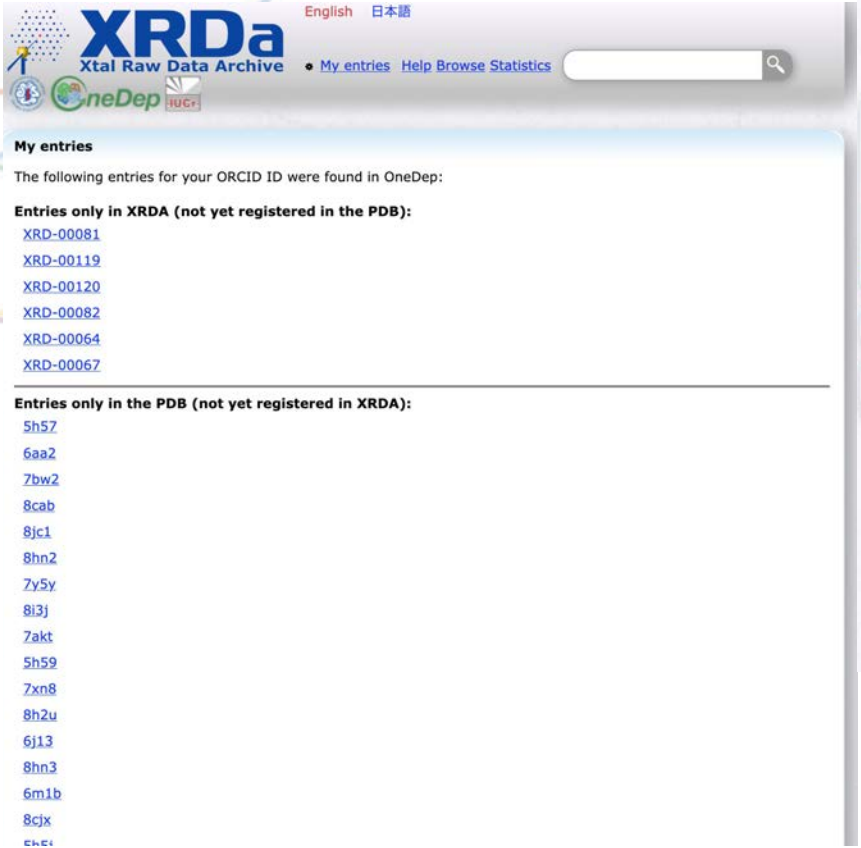
https://xrda.pdbj.org

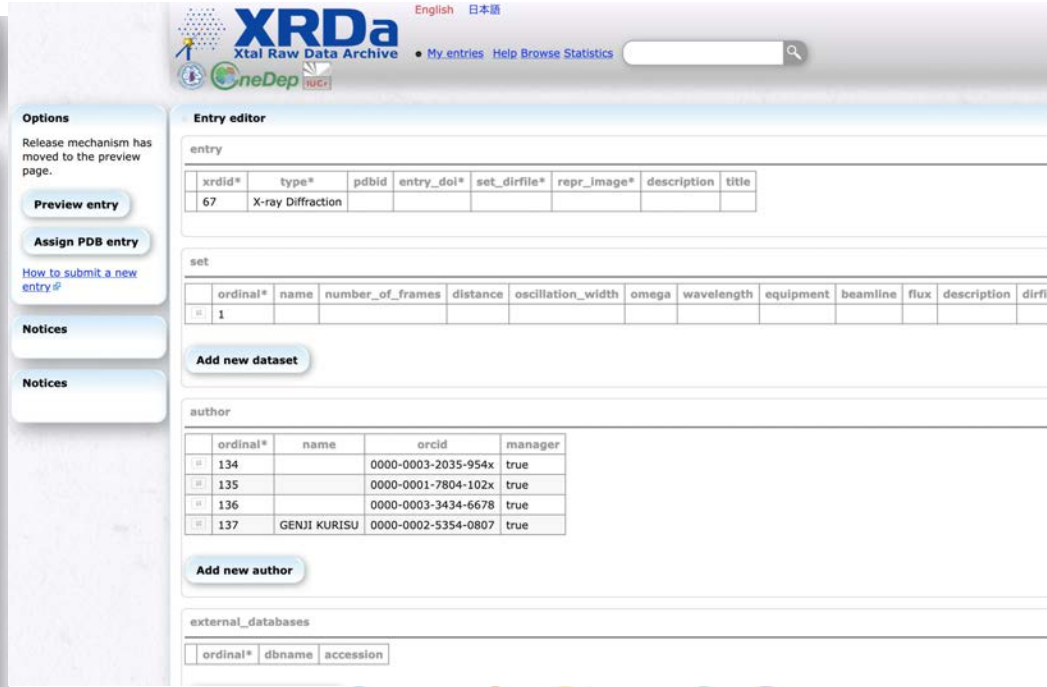
Now 81 entries are released

Archive for X-ray Diffraction Images

https://xrda.pdbj.org

ORIGINAL



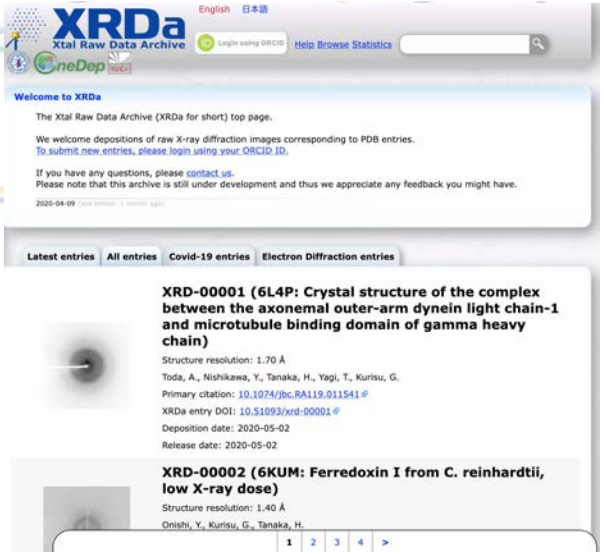


Further development

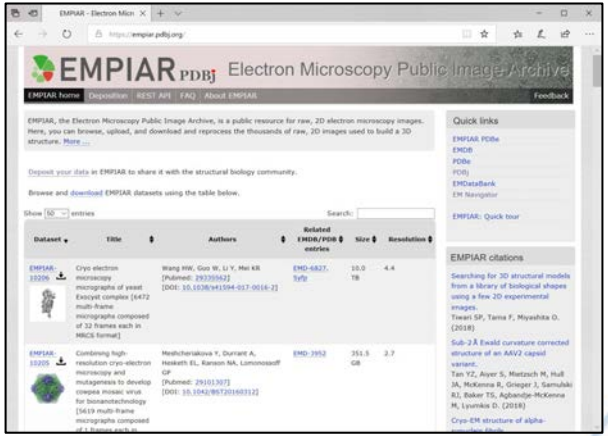
1) Semi-automatic data upload
2) Automatic data quality check

Please deposit your experimental data to PDBj

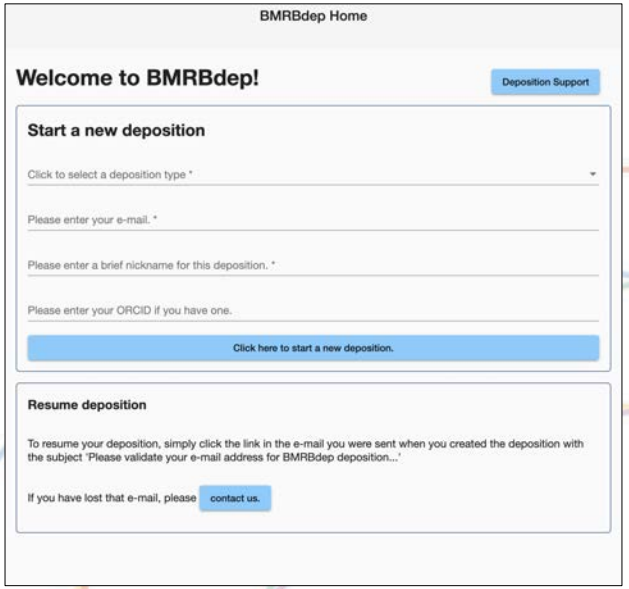
ORIGINAL
https://xrda.pdbj.org



ORIGINAL
https://empiar.pdbj.org



https://deposit-bmrb.pdbj.org

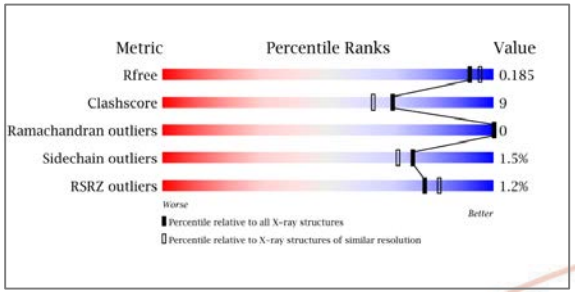


All three experimental data types are now archived in Asia!

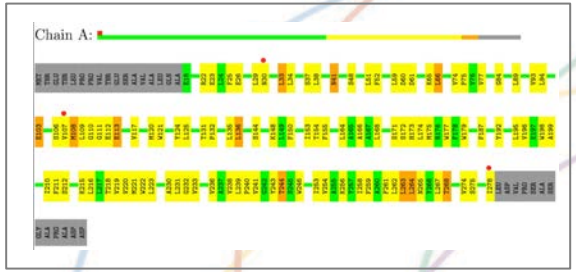
wwPDB validation report

- 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

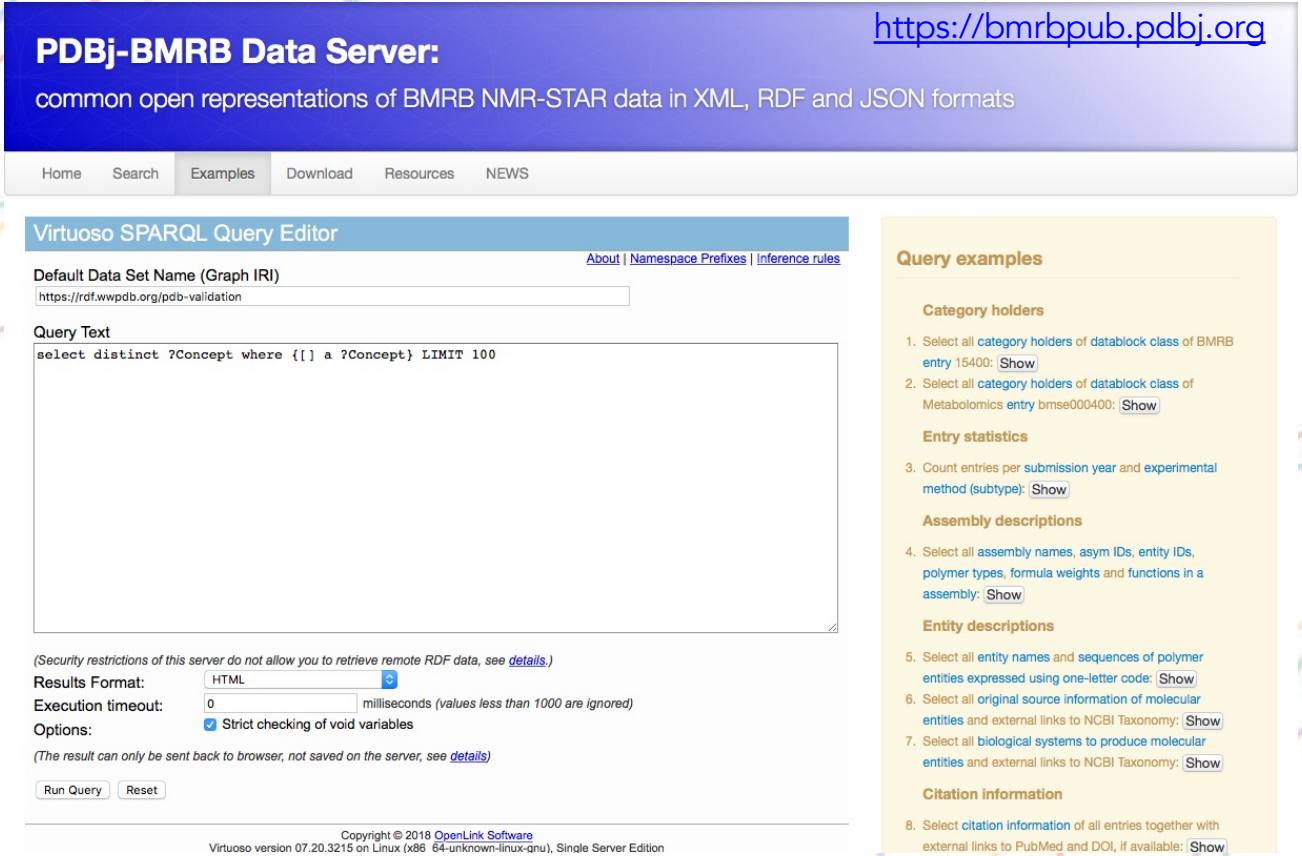
Overall Quality Summary



Residue Plots

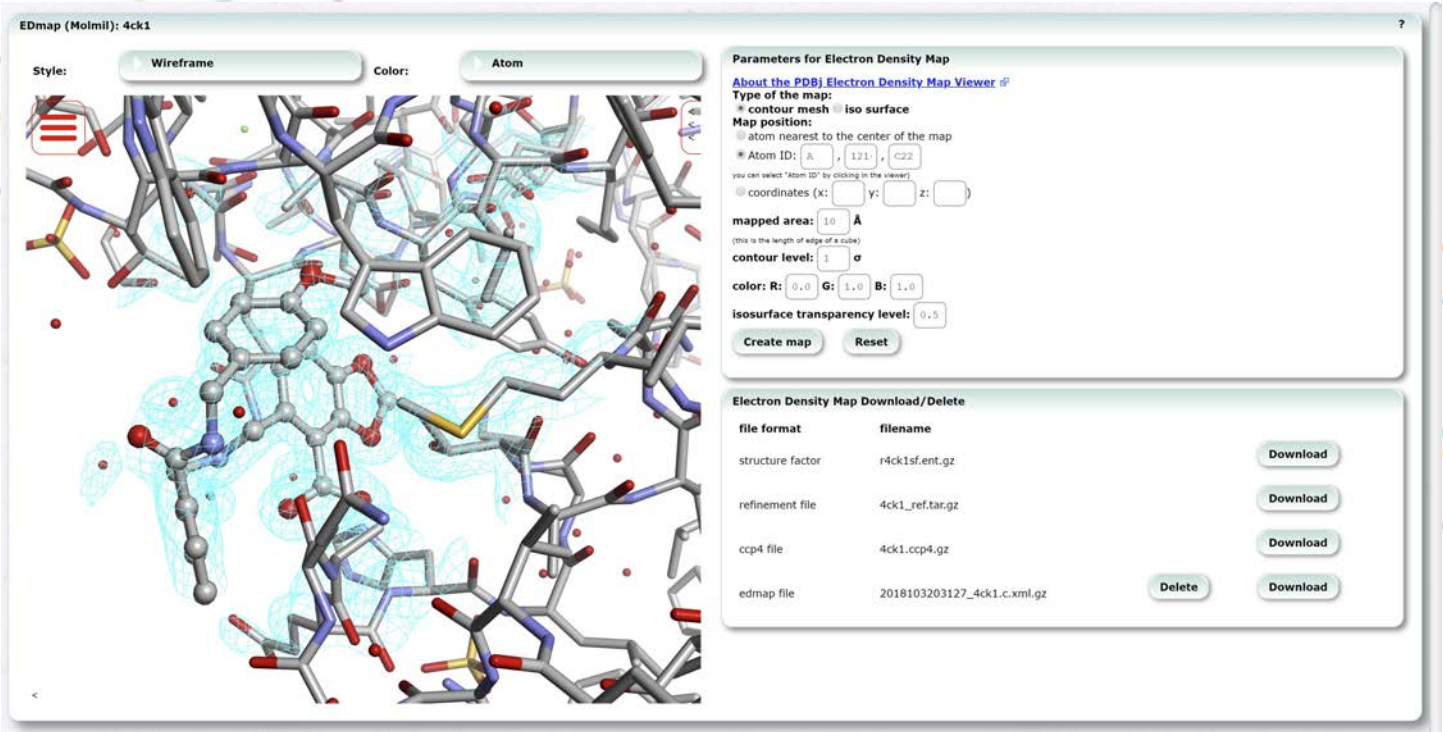


wwPDB/RDF-validation graph



Example: Entry search based on wwPDB/RDF-validation

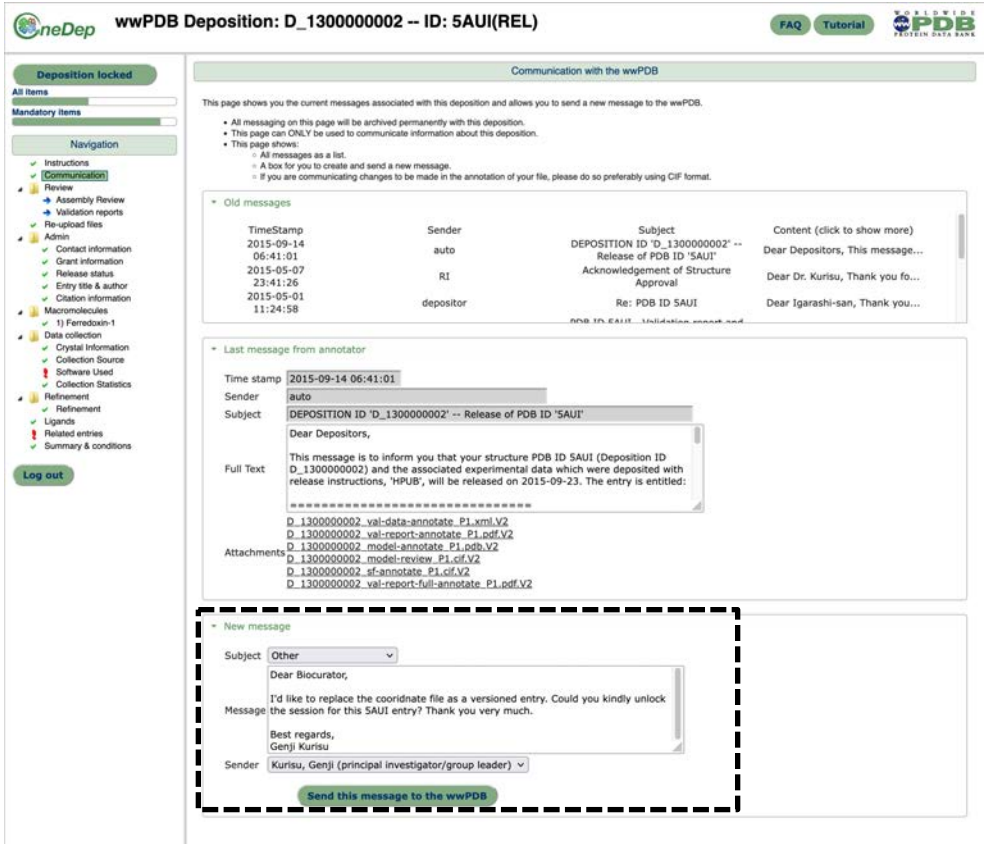
"4CK1",
"INTEGRASE",
"(4-CARBOXY-1,3-BENZODIOXOL-5-YL)METHYL-[[2-[(4-METHOXYPHENYL)METHYLCARBAMOYL]PHENYL]METHYL]AZANIUM",
"OM1",
"0.081"



Depositor-Initiated Atomic Coordinate Versioning

- Improve the quality of PDB structures
- Preserve original PDB IDs and maintain connections to the scientific literature
- Reasoning captured
- The latest versioned file can be accessed at FTP
- The latest minor version of each major version is available at versioned FTP

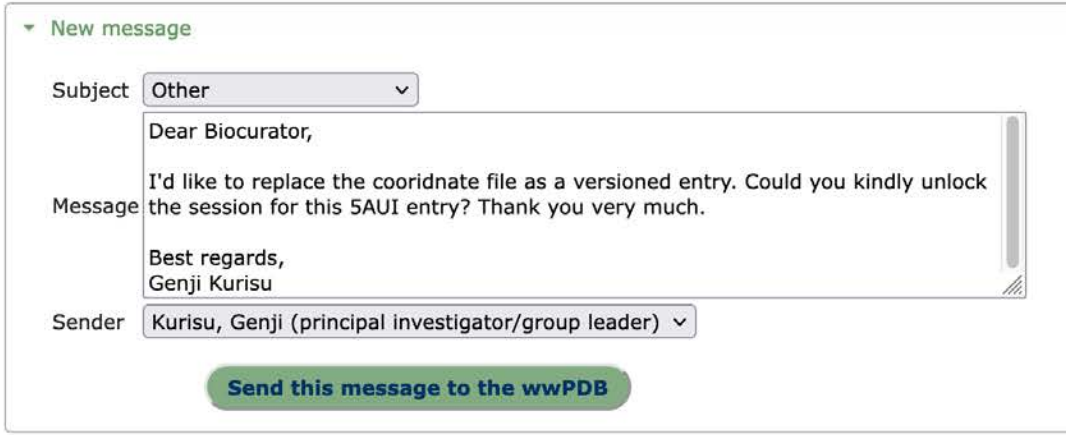
ftp-versioned.wwpdb.org



Depositor-Initiated Atomic Coordinate Versioning

- Improve the quality of PDB structures
- Preserve original PDB IDs and maintain connections to the scientific literature
- Reasoning captured
- The latest versioned file can be accessed at FTP
- The latest minor version of each major version is available at versioned FTP

ftp-versioned.wwpdb.org



Diffraction Data Case Study for GOSC

Deliverables

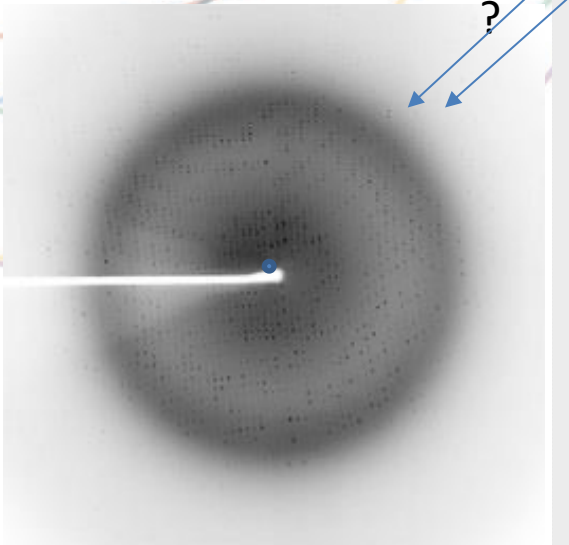
Reproducibility of data sets is paramount.

The team aim for a single point of contact for definitive molecular models, namely at the PDBj and its XRDa, the X-ray Diffraction Data Archive based at the Institute for Protein Research in Japan.

We aim to avoid dispersed multiple versions of a protein model derived from a single raw diffraction data set. Controlled versioning procedure of PDB entries should be tightly linked.

A critical deliverable is to realise metrics of ‘definitive reusability’ which would then be applicable to the individual diffraction data sets held in the XRDa. These metrics and the definitive diffraction data files are a bedrock of interoperability.

A glimpse of the variation of X-ray diffraction resolution limit choice involving the commonly used metrics in macromolecular crystallography:-

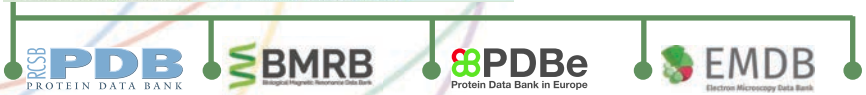


Resolution cut off estimates:-

resolution of all data : 1.913
based on CC(1/2) >= 0.33 : 1.946
based on mean(I/sigma) >=2.0 : 3.037
based on R-merge < 0.5 : 2.411
based on R-meas < 0.5 : 2.497
based on completenes >=90% : 2.335
based on completeness >=50% : 2.155

Via the Diederichs and Karplus method, using the XRDa entry the resolution limit should be 2
The depositor, Sato et al (Biochem. J. 478, 1023–1042) used 2.40 Å.

Acknowledgements



- **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.)
  - Tsunezumi, Kiyo (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))
  - 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.)

Diffraction Data Case Study for GOSC

John R Helliwell (UK)

Loes Kroon-Batenburg (The Netherlands)

CCD ID becomes 5-letters

- 3-letter ID codes of the Chemical Component Dictionary (CCD) will run out in 2022.
- Only the PDBx/mmCIF format will be provided for CCD IDs with 5-letters.