THE DATA UNIVERSE OF STRUCTURAL BIOLOGY

Helen M Berman[1,2,*], Brinda Vallat[3], Catherine L Lawson[3]
IUCrJ, 7, 630-638

[1]Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, USA
[2]Department of Biological Sciences and Bridge Institute, University of Southern California, Los Angeles, California 90089, USA
[3]Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, USA
*Corresponding author: berman@rcsb.rutgers.edu

Abstract

The Protein Data Bank (PDB) has grown from a small data resource for crystallographers to a worldwide resource serving structural biology. The history of the growth of the PDB and the role that the community played in developing standards and policies are described. We also illustrate how other biophysics communities are collaborating with the worldwide PDB to create a network of interoperating data resources. This network will expand the capabilities of structural biology and enable the determination of increasingly complex structures.

Synopsis

The history of the growth of the Protein Data Bank and the role that the community has played in developing standards and policies are described.

Keywords

Protein Data Bank; Structural biology; X-ray crystallography; Data resources; Data standards

1.0 Introduction

Crystallographers have a long tradition of effective data management practices. It is intriguing to speculate on the origins of these practices. Perhaps the requirement for ordered crystals carries over into a need for ordered results. Or perhaps it is a consequence of the fact that crystallographic experiments generate large volumes of data, yielding definitive results that are utilized by many other scientists. From its inception, the International Union of Crystallography (IUCr) took a leadership role in promoting data standards; one of the stated objectives of the IUCr is “to facilitate standardization of methods, units, nomenclatures and symbols” (iucr.org). This high level of standardization has enabled us to efficiently turn the relatively high volume of data produced by the crystallographic experiment first into information, and then into knowledge. Another objective of the IUCr, “to promote international cooperation in

crystallography", beyond creating the necessary standards, created a framework for data sharing.

Data sharing in the crystallographic community has been achieved by the development of databases, some of which are summarized in a recent article (Bruno *et al.*, 2017). One of the first data resources to be established was The International Centre for Diffraction Data's Powder Diffraction File (Faber & Fawcett, 2002, Kabekkodu *et al.*, 2002). Established in 1941, it currently houses more than one million datasets. The Cambridge Structural Database (CSD), established in 1965 by Olga Kennard, currently contains over one million small molecule structures (Groom *et al.*, 2016). Inspired in part by these resources, the Protein Data Bank (PDB) was established in 1971 to serve as an archive for the structures of biological macromolecules (Protein Data Bank, 1971). Since that time, the PDB has evolved from a data archive for bio macromolecular crystal structures to a resource for all structural biology methods. In this review, we describe this evolution with an emphasis on how the community has worked together to develop standards and policies for data sharing.

2.0 The Protein Data Bank

2.1 Early History

The Protein Data Bank began as a grassroots movement in the 1960's when the very first protein structures began to be published (Kendrew *et al.*, 1960, Perutz *et al.*, 1960). In an era when punched cards and magnetic tapes were the media for data storage and the post office was the only way to distribute information, the task of sending data to a colleague was overwhelming. At the same time, there was an increasing interest in protein folding and it was recognized that protein structure data could be enormously useful in tackling the challenge of structure prediction (Levinthal, 1968). Starting in the 60's, a series of informal meetings were held among the producers and potential users of atomic coordinate data. At the Cold Spring Harbor Symposium on protein structure held in June 1971 (Cold Spring Harbor Laboratory, 1972), Walter Hamilton offered to set up the Protein Data Bank at Brookhaven National Laboratory. He immediately flew to England and made an agreement with Olga Kennard, the head and founder of the CSD, to collaborate on such an enterprise. The announcement of the PDB appeared in Nature New Biology in October 1971 (Protein Data Bank, 1971). Hamilton worked with Edgar Meyer and Helen Berman to set up the PDB; after Hamilton's untimely death in 1973, Tom Koetzle became the head of the PDB.

In the early days, data submission was entirely voluntary. To encourage data deposition, Tom Koetzle wrote letters to protein crystallographers making them aware of the resource. The PDB Format (Bernstein *et al.*, 1977) (**Figure 1)**, based on the 80-column punched card, contained data fields for the coordinates and meta data describing the crystallographic experiment and the chemistry of the molecules in the crystal. Data distribution was accomplished using magnetic tapes and a newsletter announced the PDB holdings (Protein Data Bank, 1974).

The earliest structures in the PDB were determined using X-ray crystallography (X-ray). In 1985, the first structure determined using Nuclear Magnetic Resonance (NMR) spectroscopy was published (Williamson *et al.*, 1985) and in 1990, the first structure determined using three dimensional electron microscopy (3DEM) was incorporated into the PDB (Henderson *et al.*, 1990).

The PDB was managed by the Brookhaven National Laboratory from 1971 until 1999. In 1999 the Research Collaboratory for Structural Bioinformatics (RCSB) (Berman *et al.*, 2000)--a consortium consisting of researchers from San Diego Supercomputer Center (SDSC), Rutgers, and National Institute of Standards and Technology (NIST)—began to manage the archive. In 2003, the Worldwide PDB (wwPDB) was created to formally recognize the global reach of the PDB (Berman *et al.*, 2003). The initial partners were RCSB PDB, the Macromolecular Structure Database (MSD (Boutselakis *et al.*, 2003); now PDBe), and PDB Japan (PDBj; (Nakamura *et al.*, 2002)). The wwPDB partners formalized an agreement that there would be a single global archive with data that are freely and publicly available. Informed by advice from a Scientific Advisory Committee, the wwPDB sets the standards and procedures for processing and distributing the data.

2.2 Deposition Guidelines

The 1980's saw a new kind of activism in the crystallographic community. Many felt very strongly that data sharing should be a condition of publication. Among them were Richard Dickerson (Barinaga, 1989) who wrote letters to colleagues and to journals promoting the idea that coordinate data should be deposited to the PDB. Fred Richards circulated a petition signed by almost 200 colleagues urging the same. In that same period, several different committees were set up to study the issue. One organized by the IUCr Commission on Biological Macromolecules discussed in detail what data should be deposited; after several years of discussion and deliberation Guidelines were published (International Union of Crystallography, 1989). It was recommended that coordinates should be submitted to the PDB; deposition of structure factors was optional. Hold periods were allowed before data release. Once these Guidelines were in place and backed by strong sentiments in the community, the campaign to require data deposition succeeded. Although it took some time, virtually all journals that publish macromolecular structures now require data deposition into the PDB.

2.3 mmCIF standard

During the 1980's a new format called the Crystallographic Information File (CIF) was developed (Hall *et al.*, 1991). It is a self-defining text format that contains the key definitions for most aspects of crystallographic experiment. Its design is suitable for small molecules and allows for easy validation of these structures. CIF was adopted by the IUCr and American Chemical Society journals. In the early 1990's the IUCr set up a new committee to create a CIF like format for macromolecular structures. It soon became apparent that because of the complexity of macromolecular structures, the syntax of CIF would not be suitable. A new variant called Macromolecular Crystallographic Information File (mmCIF) was created (Bourne *et al.*, 1997).

mmCIF is a self-defining format that specifies the standards for representing macromolecular structures (Fitzgerald *et al.*, 2005). These standards include definitions for describing the experimental procedures, the chemistry of the components, and the results of a biomacromolecular crystallographic structure determination. mmCIF also provides mechanisms to enforce data consistency, which is important for archiving. A comparison of coordinate records in PDB and mmCIF is shown in **Figure 1**.

mmCIF has been designed to be extensible. Over time, the wwPDB has extended mmCIF to build the PDBx/mmCIF metadata framework (Westbrook *et al.*, 2005, Westbrook & Fitzgerald, 2009, Westbrook, 2013) that enables archiving of structural models obtained from X-ray, NMR, and 3DEM experiments. In addition to definitions for representing macromolecular structures, the framework also includes descriptions of the supporting metadata such as information about source organisms, samples, workflows, authors, citations, software, and model quality metrics.

Because of its syntax, mmCIF allows for the creation of relational databases and it was clear that it would be useful for storing PDB data. However, the pushback on mmCIF by the community was very strong. The PDB format was simple and human readable. It was used by hundreds of software programs for structure determination and analysis. However, the 80-column format meant strict limitations on the number of atoms that could be stored within a single file and large structures had to be split into multiple files. mmCIF was first adopted by the Nucleic Acid Database (Berman *et al.*, 1992), and by the PDB when its management was taken over by the RCSB in 1999 but it was not until 2011 that crystallographic software developers agreed to adopt mmCIF. Once that occurred, PDBx/mmCIF became the Master Format for the PDB. All of the very large structures that had needed to be split into multiple files in the PDB Format were then converted to single mmCIF files; access to these complex structures is now greatly simplified. A PDBx/mmCIF Working Group was set up under the auspices of the wwPDB to enable the use of the format in major software packages. Starting in 2019, all X-ray structure depositions are required to be in mmCIF (Adams *et al.*, 2019).

2.4 Validation

In the very early days of the PDB, the primary focus of annotation was to ensure that the data were formatted correctly and that there were no obvious errors. In time, validation procedures were set up to check the geometry, nomenclature and chemistry of the coordinate files. Among the items checked are the stereochemistry including valence geometry, dihedral angles, planarity and chirality. Non bonded contacts as well as crystallographic and non-crystallographic symmetry are assessed. The primary sequence of the polymer is checked against sequence databases and the geometry of small molecules is evaluated.

Notably absent in these early assessments were checks against the primary data. In 2000, the IUCr recommended that structure factors be a requirement of deposition and enforced this requirement for its journals (Commission on Biological Macromolecules, 2000). Although this was endorsed by many in the community (Wlodawer, 2007), it was

not until 2008 that the deposition of structure factors became mandatory. This requirement, plus suspicions that there were some fraudulent structures in the PDB (Berman *et al.*, 2010), led the wwPDB to convene an X-ray Validation Task Force (VTF). The X-ray VTF, led by Randy Read and consisting of thought leaders in crystallographic methods, studied possible checks that could be done on the full complement of data. The entire corpus of data (70,000) were run against these checks to assess outliers. A final set of recommendations were made (Read *et al.*, 2011).

The X-ray VTF recommended that a small set of validation data be presented in an easily understood format, with comparisons made to both the full PDB archive and to the structure's resolution class. The suggested validation criteria included measures that evaluate fit of the structure to the experimental data (R_{free} and Real-space residual Z-scores (Brunger, 1992, Kleywegt *et al.*, 2004)), assess the quality of the coordinates (clashes, protein backbone, side-chain rotamers, and buried unsatisfied H-bonds (Laskowski *et al.*, 1993, Chen *et al.*, 2010, Dunbrack & Cohen, 1997, Hooft *et al.*, 1996)), and check the crystal lattice for underpacking (Sheffler & Baker, 2009). The VTF developed a novel "sliders" representation that compactly displays a structure's score values for each of the key criteria, as well as its percentile rank in the archive and compared it to other structures in the same resolution range. They also listed criteria that should be flagged for review in any incoming PDB structure entry: poor overall geometry or extreme local geometry distortion, inverted chirality, structure factor intensity outliers, incorrect data labels, missed symmetry, missed twinning, incomplete structure, poor ligand density or geometry, and inconsistent carbohydrate nomenclature.

In addition to the X-ray VTF, an NMR VTF (Montelione *et al.*, 2013) and 3DEM VTF (Henderson *et al.*, 2012) have been established and have produced their respective recommendations for validation.

Concurrent with the work of the VTF's, the wwPDB partners began a project to create a unified system for deposition, curation and validation of structures that have been submitted to PDB. The recommendations of the VTF's became the basis of the validation suite (Gore *et al.*, 2017) for the new system, called OneDep (Young *et al.*, 2017). The validation report includes most of the indicators recommended by the X-ray VTF including a slider (**Figure 2a**), various geometric checks, and graphical summaries of chain quality. (**Figure 2b**). Many journals now require authors to submit the PDB validation reports with their manuscripts. Thus, the structural biology community has set a very high bar for responsible reporting of research results.

2.5 Current state of the PDB

The rate of growth of PDB holdings has increased dramatically (wwPDB consortium, 2019). From seven relatively small crystal structures there are now more than 160,000. **Figure 3** shows the growth charts for structures determined by the three methods currently supported by the PDB. Of the three, 3DEM shows the greatest growth rate.

The complexity of structures archived in the PDB has increased over time starting from the single chain structures of myoglobin to more complex macromolecular assemblies

such as the ribosome and viruses. There are now more than 600 full ribosome structures and several structures of viruses including Zika, Ebola, dengue, and enteroviruses (Rossmann, 2013, Kaelber *et al.*, 2017). Most notably the RCSB PDB set up a resource page for the 2019-nCoV (coronavirus) related structures (RCSB PDB, 2020).

The usage of the PDB is remarkable with 900 million downloads in 2019 (**Figure 4**). These structures are used in many ways including as starting models for crystal structures being solved by molecular replacement and for fitting 3DEM maps. Modelers make particularly heavy use of the PDB. For example, the CASP project uses PDB data to develop methods for structure prediction (Kryshtafovych *et al.*, 2019). Biochemists and biophysicists use structures to help explain their findings and structures in the PDB have facilitated the discovery of several new drugs (Westbrook & Burley, 2019). The wwPDB partners maintain heavily accessed websites that offer many scientific and educational services.

3.0 Other Structural Biology Databases

As the field of structural biology grew, new data resources were developed to complement and supplement the data in the Protein Data Bank. These include repositories for new types of primary data, and knowledgebases that integrate data from resources in other fields of biology. A summary of some of these resources is given here.

The Nucleic Acid Database (NDB) (Berman *et al.*, 1992) was originally developed as a resource for annotated nucleic acid structures. Although the PDB did accept nucleic acid structures, the focus of annotation was on proteins. In the late 1990's, an agreement was reached with the PDB for NDB to do the primary annotation on nucleic acid structures and transfer them to the PDB. The NDB also became the proving ground for developing the mmCIF standard. The internal format for NDB was mmCIF which allowed the data to be easily loaded into a relational database. When the management of the PDB moved to RCSB, the NDB became a knowledgebase used by specialists in nucleic acids. It contains annotations for the nucleic acid base pairs, backbone conformations, and structural motifs as well as functional descriptions of proteins bound to nucleic acids.

Recognizing that publicly available 3D density maps could accelerate discovery in structural biology and medicine, the Electron Microscopy Data Bank (EMDB) at the European Bioinformatics Institute (EBI) was launched in 2002 (Henrick *et al.*, 2003). EMDB accepts maps determined using any cryo-EM method, including single particle reconstruction with any symmetry, helical filament reconstruction, subtomogram averaging, tomography, electron crystallography, and micro electron diffraction, along with metadata describing the full experimental workflow.

In 2006, scientists from EMDB, RCSB, and the National Center for Macromolecular Imaging (NCMI) initiated a collaboration to ensure that data archiving and validation standards for cryo-EM maps and models are coordinated internationally (Lawson *et al.*, 2011). The project, now known as EMDataResource (EMDR; emdataresource.org) hosted the first 3DEM VTF (Henderson *et al.*, 2012). The EMDR project website (**Figure 5**) serves as a global resource for cryo-EM structure data, and EM related news, events,

software tools, data standards, validation methods, and community challenges (Lawson *et al.*, 2016, Lawson & Chiu, 2018). The site also offers growth statistics for 3DEM structures in the PDB and maps in EMDB (**Figure 5**).

In 2012 the Electron Microscopy Public Image Archive (EMPIAR) was established at EBI (Iudin *et al.*). EMPIAR enables cryo-EM scientists to archive and share raw images and intermediate data files associated with their maps deposited to EMDB. Making raw image data broadly available has multiple benefits, including accelerating development of reconstruction software, and enriching resources for cryo-EM scientists in training. Approximately 4% of EMDB entries deposited since 2012 have associated EMPIAR entries.

Creation of a publicly available database for experimental NMR data was first proposed in 1989 (Ulrich *et al.*, 1989). The design and implementation of the NMR database called BioMagResBank (BMRB) began in 1991 (Seavey *et al.*, 1991). BMRB is a repository for data obtained from NMR spectroscopy experiments carried out on biological systems (Ulrich *et al.*, 2008, Romero *et al.*, 2020) and employs the NMR-STAR (Ulrich *et al.*, 2018) data standards to describe NMR experiments as well as many kinds of NMR spectral data and derived data (e.g., assigned chemical shifts, restraints, coupling constants, relaxation parameters, etc.). BMRB became a core member of the wwPDB in 2007 (Markley *et al.*, 2008), allowing for common practices to be established for depositions of biomolecular NMR data in BMRB and the associated structural models in the PDB. Currently, about 10% of structures deposited in the PDB have been determined using NMR spectroscopy. An extension of PDBx/mmCIF, called NMR Exchange Format (NEF) (Gutmanas *et al.*, 2015) has been created to facilitate data exchange.

Small angle scattering (SAS) of X-rays and neutrons provides information regarding 3D structures and structural changes of biological macromolecules in solution. Recent advances have led to the use of SAS in conjunction with X-ray, NMR and 3DEM as a complementary method to determine the structures of macromolecules. In 2013, the wwPDB set up a SAS Validation Task Force (SASVTF) to address the requirements for archiving SAS data (Trewhella *et al.*, 2013). Following the SASVTF recommendations (Trewhella *et al.*, 2013), the Small Angle Scattering Biological Data Bank (SASBDB) (Valentini *et al.*, 2015) was established in 2015 at the European Molecular Biology Laboratory, Hamburg Outstation. SASBDB is a curated repository for data obtained from SAS experiments. The archival standards for SASBDB are encoded in the sasCIF data dictionary (Malfois & Svergun, 2000), an extension of the PDBx/mmCIF data representation. The sasCIF dictionary describes SAS experimental data, SAS derived models, and additional metadata required for analysis and validation (Kachala *et al.*, 2016).

Structures of complex macromolecular assemblies are increasingly determined using integrative modeling (Rout & Sali, 2019), where a combination of complementary experimental and computational techniques is employed. In addition to traditional structure determination methods such as X-ray, NMR, and 3DEM, experimental techniques such as SAS, atomic force microscopy (AFM), chemical cross-linking (CX),

mass spectrometry (MS), Hydrogen/Deuterium exchange (HDX), Förster resonance energy transfer (FRET), electron paramagnetic resonance (EPR), and various proteomics and bioinformatics approaches contribute to integrative modeling. Spatial restraints derived from the different kinds of experimental and computational methods are combined to determine integrative structures of the macromolecular assembly. In 2014, the wwPDB established an Integrative/Hybrid Methods (IHM) Task Force and sponsored a Workshop that engaged a community of experts to address the challenges involved in archiving integrative structures. A white paper was published (Sali *et al.*, 2015) with recommendations for archiving integrative structures.

Based on the recommendations of the wwPDB IHM Task Force, an IHM extension of the PDBx/mmCIF dictionary (**Figure 6**) has been developed to describe integrative structures and their associated spatial restraints (Vallat *et al.*, 2018, Vallat *et al.*, 2019). The IHM dictionary extension contains definitions for multi-scale models with atomic and coarse-grained representations, ensembles in multiple conformational states, spatial restraints derived from different kinds of experimental techniques, starting structural models used in integrative modeling, and simplified definitions of the modeling workflow.

A prototype archiving system, PDB-Dev (https://pdb-dev.wwpdb.org) (**Figure 7**) has been created to archive integrative structural models (Burley *et al.*, 2017, Vallat *et al.*, 2018, Vallat *et al.*, 2019). PDB-Dev was built based on the definitions in the IHM dictionary and consists of about 40 integrative structures of macromolecular complexes as of March 2020.

In 2019, a Biophysical Society (BPS) Satellite Workshop assessed progress and discussed further requirements for archiving integrative structures. One of the recommendations that emerged was development of common data standards to enable efficient data exchange among the scientific repositories contributing to structural biology (Berman *et al.*, 2019). The recommendations provide the foundation for building a global federation of interoperating scientific resources that follow common data management practices and enable efficient data sharing and archiving.

Following the Workshop, practitioners of several different experimental methods have engaged in further community-building activities. For instance, the HDX-MS community has published a whitepaper with recommendations for performing, interpreting and reporting HDX-MS experiments (Masson *et al.*, 2019), the CX-MS community is in the process of finalizing their recommendations with regard to standards and archiving of CX-MS data, the FRET community has established a platform for joint scientific efforts in the field of FRET (www.fret.community), the 3DEM community is working on recommendations for validating 3DEM maps and models, and the integrative modeling community is focused on building a comprehensive infrastructure for PDB-Dev and creating methods for validating integrative structures.

4.0 Perspectives

In this review we show how the crystallographic community has played a leadership role in establishing data standards and creating an effective framework for responsible data management. The PDB has set an example for bottom-up, community-driven establishment of data management practices, paving the way for development of standards and for creation of several other structural biology resources. Now other biological communities that contribute to integrative structural biology are coming together to develop data standards and promote data sharing. This steady progression ensures that, in time, there will be a global network of interoperating data resources that enable scientific research. Given this trajectory, it is not overly optimistic to speculate that in the next decade, it will be possible to tackle very large structure determination challenges such as the creation of a spatio-temporal model of an entire cell (Singla *et al.*, 2018).

Acknowledgements

We gratefully acknowledge funding from NSF, DBI-1756248 and NIGMS, R01GM079429-12.

We also thank the members of the extended PDB family for the many years of collaboration and support.

Figure legends

Figure 1: Formats for representation of atomic coordinates. (a) PDB format. All data items are in fixed-sized fields and definitions are implicit. (b) mmCIF format. The names of the data items as defined in the mmCIF dictionary are listed first using a loop directive. The values of the data items then follow in a tabular form. This representation enables mmCIF to be flexible, self-consistent and software compatible.

Figure 2. Key elements of the wwPDB validation report for X-ray structures are shown for PDB entry 6pzd, a recent crystal structure of Influenza A neuraminidase, determined at 1.12 Å (Zhu *et al.*, 2019). (a) Graphical display of key metrics (“sliders”). For each metric, two percentile ranks are calculated: an absolute rank with respect to the entire PDB archive and a relative rank with respect to structures determined at similar resolution. Slider markers in the blue region on the right are indicative of a high quality structure. Lower quality structures have the markers in the red region on the left. (b) Residue Property Plot: Residues are color-coded green if no issues are detected, yellow if there are outliers for one criterion (e.g., unusual bond lengths), orange if there are outliers for two criteria with outliers (e.g., unusual bond lengths and clashes), and red for three or more criteria. A horizontal stack bar plot presents the fraction of residues with each color code. Unmodeled regions of the chain, if present, are represented by a gray segment. The upper red bar indicates the fraction of residues with poor fit to the electron density.

Figure 3. Cumulative holdings of PDB at the end of each decade for each of the three major structure determination methods, X-ray crystallography, NMR, and 3DEM, respectively. 3DEM methods include structures determined by electron microscopy

(single particle, helical, subtomogram averaging, and tomography) and electron crystallography.

Figure 4. Total Annual Downloads of PDB Archive files. Plotted values represent the sum of annual downloads from all of the wwPDB partner ftp and web sites. Data Source: https://www.wwpdb.org/stats/download.

Figure 5. (a) Recent screenshot of the EMDataResource website (https://www.emdataresource.org). The website is updated weekly to highlight all newly released EMDB maps. (b) Cumulative number of 3DEM maps available in EMDB and coordinate models available in PDB, by year. 2020 statistics are through February 2. source: https://www.emdataresource.org/statistics.html.

Figure 6: The IHM dictionary provides definitions for (a) spatial restraints from experimental methods such as X-ray, NMR, 3DEM, CX-MS, SAS and FRET, (b) multi-scale assemblies consisting of both atomic coordinates and coarse-grained representations, (c) ensembles representing multiple conformational states or ensembles related by time or other criteria such as events in a sequential pathway, and (d) starting structural models used in integrative modeling.

Figure 7: Screen shot of the current PDB-Dev website (https://pdb-dev.wwpdb.org). PDB-Dev currently consists of over 40 integrative structures including several that are on hold for publication. The structures archived in PDB-Dev vary in complexity from simple atomic structures in a single conformational state to complex coarse-grained assemblies in multiple conformational states. The data model underlying PDB-Dev supports the representation of these complex structures as well as the diverse set of spatial restraints used in building them.

References

Adams, P. D., Afonine, P. V., Baskaran, K., Berman, H. M., Berrisford, J., Bricogne, G., Brown, D. G., Burley, S. K., Chen, M., Feng, Z., Flensburg, C., Gutmanas, A., Hoch, J. C., Ikegawa, Y., Kengaku, Y., Krissinel, E., Kurisu, G., Liang, Y., Liebschner, D., Mak, L., Markley, J. L., Moriarty, N. W., Murshudov, G. N., Noble, M., Peisach, E., Persikova, I., Poon, B. K., Sobolev, O. V., Ulrich, E. L., Velankar, S., Vonrhein, C., Westbrook, J., Wojdyr, M., Yokochi, M. & Young, J. Y. (2019). *Acta crystallographica. Section D, Structural biology* **75**, 451-454.

Barinaga, M. (1989). *Science* **245**, 1179-1181.

Berman, H. M., Adams, P. D., Bonvin, A. A., Burley, S. K., Carragher, B., Chiu, W., DiMaio, F., Ferrin, T. E., Gabanyi, M. J., Goddard, T. D., Griffin, P. R., Haas, J., Hanke, C. A., Hoch, J. C., Hummer, G., Kurisu, G., Lawson, C. L., Leitner, A., Markley, J. L., Meiler, J., Montelione, G. T., Phillips, G. N., Jr., Prisner, T., Rappsilber, J., Schriemer, D. C., Schwede, T., Seidel, C. A. M., Strutzenberg, T. S., Svergun, D. I., Tajkhorshid, E., Trewhella, J., Vallat, B., Velankar, S., Vuister, G. W., Webb, B., Westbrook, J. D., White, K. L. & Sali, A. (2019). *Structure* **27**, 1745-1759.

Berman, H. M., Henrick, K. & Nakamura, H. (2003). *Nature Structure Biology* **10**, 980.

Berman, H. M., Kleywegt, G. J., Nakamura, H., Markley, J. L. & Burley, S. K. (2010). *Nature* **463**, 425.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S. H., Srinivasan, A. R. & Schneider, B. (1992). *Biophys J* **63**, 751-759.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res* **28**, 235-242.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer Jr., E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *Journal of Molecular Biology* **112**, 535-542.

Bourne, P. E., Berman, H. M., McMahon, B., Watenpaugh, K. D., Westbrook, J. D. & Fitzgerald, P. M. (1997). *Methods Enzymol* **277**, 571-590.

Boutselakis, H., Dimitropoulos, D., Fillon, J., Golovin, A., Henrick, K., Hussain, A., Ionides, J., John, M., Keller, P. A., Krissinel, E., McNeil, P., Naim, A., Newman, R., Oldfield, T., Pineda, J., Rachedi, A., Copeland, J., Sitnov, A., Sobhany, S., Suarez-Uruena, A., Swaminathan, J., Tagari, M., Tate, J., Tromm, S., Velankar, S. & Vranken, W. (2003). *Nucleic Acids Res* **31**, 458-462.

Brunger, A. T. (1992). *Nature* **355**, 472-475.

Bruno, I., Gražulis, S., John R Helliwell, Soorya N Kabekkodu, Brian McMahon & John Westbrook (2017). *Data Science Journal* **16**, 38.

Burley, S. K., Kurisu, G., Markley, J. L., Nakamura, H., Velankar, S., Berman, H. M., Sali, A., Schwede, T. & Trewhella, J. (2017). *Structure* **25**, 1317-1318.

Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Crystallographica. Series D* **66**, 12-21.

Cold Spring Harbor Laboratory (1972). *Cold spring harbor symposia on quantitative biology, vol. 36*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Commission on Biological Macromolecules (2000). *Acta Crystallographica. Series D* **56**, 2.

Dunbrack, R. L., Jr. & Cohen, F. E. (1997). *Protein Sci* **6**, 1661-1681.

Faber, J. & Fawcett, T. (2002). *Acta Crystallogr B* **58**, 325-332.

Fitzgerald, P. M. D., Westbrook, J. D., Bourne, P. E., McMahon, B., Watenpaugh, K. D. & Berman, H. M. (2005). *International tables for crystallography g. Definition and exchange of crystallographic data*, edited by S. R. Hall & B. McMahon, pp. 295-443. Dordrecht, The Netherlands: Springer.

Gore, S., Sanz Garcia, E., Hendrickx, P. M. S., Gutmanas, A., Westbrook, J. D., Yang, H., Feng, Z., Baskaran, K., Berrisford, J. M., Hudson, B. P., Ikegawa, Y., Kobayashi, N., Lawson, C. L., Mading, S., Mak, L., Mukhopadhyay, A., Oldfield, T. J., Patwardhan, A., Peisach, E., Sahni, G., Sekharan, M. R., Sen, S., Shao, C., Smart, O. S., Ulrich, E. L., Yamashita, R., Quesada, M., Young, J. Y., Nakamura, H., Markley, J. L., Berman, H. M., Burley, S. K., Velankar, S. & Kleywegt, G. J. (2017). *Structure* **25**, 1916-1927.

Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. (2016). *Acta Crystallogr B* **72**, 171-179.

Gutmanas, A., Adams, P. D., Bardiaux, B., Berman, H. M., Case, D. A., Fogh, R. H., Guntert, P., Hendrickx, P. M., Herrmann, T., Kleywegt, G. J., Kobayashi, N., Lange, O. F., Markley, J. L., Montelione, G. T., Nilges, M., Ragan, T. J., Schwieters, C. D., Tejero, R., Ulrich, E. L., Velankar, S., Vranken, W. F., Wedell, J. R., Westbrook, J., Wishart, D. S. & Vuister, G. W. (2015). *Nat Struct Mol Biol* **22**, 433-434.

Hall, S. R., Allen, F. H. & Brown, I. D. (1991). *Acta Crystallogr A* **47**, 655-685.

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E. & Downing, K. H. (1990). *J Mol Biol* **213**, 899-929.

Henderson, R., Sali, A., Baker, M. L., Carragher, B., Devkota, B., Downing, K. H., Egelman, E. H., Feng, Z., Frank, J., Grigorieff, N., Jiang, W., Ludtke, S. J., Medalia, O., Penczek, P. A., Rosenthal, P. B., Rossmann, M. G., Schmid, M. F., Schroder, G. F., Steven, A. C., Stokes, D. L., Westbrook, J. D., Wriggers, W., Yang, H., Young, J., Berman, H. M., Chiu, W., Kleywegt, G. J. & Lawson, C. L. (2012). *Structure* **20**, 205-214.

Henrick, K., Newman, R., Tagari, M. & Chagoyen, M. (2003). *J Struct Biol* **144**, 228-237.

Hooft, R. W., Sander, C. & Vriend, G. (1996). *Proteins : Structure, Function, and Genetics* **26**, 363-376.

International Union of Crystallography (1989). *Acta Cryst* **A45**, 658.

Iudin, A., Korir, P. K., Salavert-Torres, J., Kleywegt, G. J. & Patwardhan, A. (2016). *Nat Methods* **13**, 387-388.

Kabekkodu, S. N., Faber, J. & Fawcett, T. (2002). *Acta Crystallogr B* **58**, 333-337.

Kachala, M., Westbrook, J. & Svergun, D. (2016). *J Appl Crystallogr* **49**, 302-310.

Kaelber, J. T., Hryc, C. F. & Chiu, W. (2017). *Annual review of virology* **4**, 287-308.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. C. (1960). *Nature* **185**, 422-427.

Kleywegt, G. J., Harris, M. R., Zou, J. Y., Taylor, T. C., Wahlby, A. & Jones, T. A. (2004). *Acta Crystallographica. Series D* **60**, 2240-2249.

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. & Moult, J. (2019). *Proteins : Structure, Function, and Genetics* **87**, 1011-1020.

Laskowski, R. A., McArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *Journal of Applied Crystallography* **26**, 283-291.

Lawson, C. L., Baker, M. L., Best, C., Bi, C., Dougherty, M., Feng, P., van Ginkel, G., Devkota, B., Lagerstedt, I., Ludtke, S. J., Newman, R. H., Oldfield, T. J., Rees, I., Sahni, G., Sala, R., Velankar, S., Warren, J., Westbrook, J. D., Henrick, K., Kleywegt, G. J., Berman, H. M. & Chiu, W. (2011). *Nucleic Acids Research* **39**, D456-464.

Lawson, C. L. & Chiu, W. (2018). *J Struct Biol* **204**, 523-526.

Lawson, C. L., Patwardhan, A., Baker, M. L., Hryc, C., Garcia, E. S., Hudson, B. P., Lagerstedt, I., Ludtke, S. J., Pintilie, G., Sala, R., Westbrook, J. D., Berman, H. M., Kleywegt, G. J. & Chiu, W. (2016). *Nucleic Acids Res* **44**, D396-403.

Levinthal, C. (1968). *Extrait du Journal de Chimie Physique* **65**, 44-45.

Malfois, M. & Svergun, D. I. (2000). *Journal of Applied Crystallography* **33**, 812-816.

Markley, J. L., Ulrich, E. L., Berman, H. M., Henrick, K., Nakamura, H. & Akutsu, H. (2008). *Journal of Biomolecular NMR* **40**, 153-155.

Masson, G. R., Burke, J. E., Ahn, N. G., Anand, G. S., Borchers, C., Brier, S., Bou-Assaf, G. M., Engen, J. R., Englander, S. W., Faber, J., Garlish, R., Griffin, P. R., Gross, M. L., Guttman, M., Hamuro, Y., Heck, A. J. R., Houde, D., Iacob, R. E., Jorgensen, T. J. D., Kaltashov, I. A., Klinman, J. P., Konermann, L., Man, P., Mayne, L., Pascal, B. D., Reichmann, D., Skehel, M., Snijder, J., Strutzenberg, T. S., Underbakke, E. S., Wagner, C., Wales, T. E., Walters, B. T., Weis, D. D., Wilson, D. J., Wintrode, P. L., Zhang, Z., Zheng, J., Schriemer, D. C. & Rand, K. D. (2019). *Nat Methods* **16**, 595-602.

Montelione, G. T., Nilges, M., Bax, A., Guntert, P., Herrmann, T., Richardson, J. S., Schwieters, C. D., Vranken, W. F., Vuister, G. W., Wishart, D. S., Berman, H. M., Kleywegt, G. J. & Markley, J. L. (2013). *Structure* **21**, 1563-1570.

Nakamura, H., Ito, N. & Kusunoki, M. (2002). *Tanpakushitsu Kakusan Koso* **47**, 1097-1101.

Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G. & North, A. C. T. (1960). *Nature* **185**, 416-422.

Protein Data Bank (1971). *Nature (London), New Biol.* **233**, 223-223.

Protein Data Bank (1974). Brookhaven National Laboratory.

RCSB PDB (2020). *RCSB PDB Covid-19/SARS-COV-2 resources,* https://www.rcsb.org/news?year=2020&article=5e74d55d2d410731e9944f52&feature=true.

Read, R. J., Adams, P. D., Arendall, W. B., 3rd, Brunger, A. T., Emsley, P., Joosten, R. P., Kleywegt, G. J., Krissinel, E. B., Lutteke, T., Otwinowski, Z., Perrakis, A., Richardson, J. S., Sheffler, W. H., Smith, J. L., Tickle, I. J., Vriend, G. & Zwart, P. H. (2011). *Structure* **19**, 1395-1412.

Romero, P. R., Kobayashi, N., Wedell, J. R., Baskaran, K., Iwata, T., Yokochi, M., Maziuk, D., Yao, H., Fujiwara, T., Kurusu, G., Ulrich, E. L., Hoch, J. C. & Markley, J. L. (2020). *Methods Mol Biol* **2112**, 187-218.

Rossmann, M. G. (2013). *Q Rev Biophys* **46**, 133-180.

Rout, M. P. & Sali, A. (2019). *Cell* **177**, 1384-1403.

Sali, A., Berman, H. M., Schwede, T., Trewhella, J., Kleywegt, G., Burley, S. K., Markley, J., Nakamura, H., Adams, P., Bonvin, A. M., Chiu, W., Peraro, M. D., Di Maio, F., Ferrin, T. E., Grunewald, K., Gutmanas, A., Henderson, R., Hummer, G., Iwasaki, K., Johnson, G., Lawson, C. L., Meiler, J., Marti-Renom, M. A., Montelione, G. T., Nilges, M., Nussinov, R., Patwardhan, A., Rappsilber, J., Read, R. J., Saibil, H., Schroder, G. F., Schwieters, C. D., Seidel, C. A., Svergun, D., Topf, M., Ulrich, E. L., Velankar, S. & Westbrook, J. D. (2015). *Structure* **23**, 1156-1167.

Seavey, B. R., Farr, E. A., Westler, W. M. & Markley, J. L. (1991). *J Biomol NMR* **1**, 217-236.

Sheffler, W. & Baker, D. (2009). *Protein Sci* **18**, 229-239.

Singla, J., McClary, K. M., White, K. L., Alber, F., Sali, A. & Stevens, R. C. (2018). *Cell* **173**, 11-19.

Trewhella, J., Hendrickson, W. A., Kleywegt, G. J., Sali, A., Sato, M., Schwede, T., Svergun, D. I., Tainer, J. A., Westbrook, J. & Berman, H. M. (2013). *Structure* **21**, 875-881.

Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent Wenger, R., Yao, H. & Markley, J. L. (2008). *Nucleic Acids Res* **36**, D402-408.

Ulrich, E. L., Baskaran, K., Dashti, H., Ioannidis, Y. E., Livny, M., Romero, P. R., Maziuk, D., Wedell, J. R., Yao, H., Eghbalnia, H. R., Hoch, J. C. & Markley, J. L. (2018). *Journal of Biomolecular NMR* 1-5.

Ulrich, E. L., Markley, J. L. & Kyogoku, Y. (1989). *Protein Seq Data Anal* **2**, 23-37.

Valentini, E., Kikhney, A. G., Previtali, G., Jeffries, C. M. & Svergun, D. I. (2015). *Nucleic Acids Res* **43**, D357-363.

Vallat, B., Webb, B., Westbrook, J., Sali, A. & Berman, H. M. (2019). *J Biomol NMR*

Vallat, B., Webb, B., Westbrook, J. D., Sali, A. & Berman, H. M. (2018). *Structure* **26**, 894-904.

Westbrook, J. (2013). *Pdbx/mmcif dictionary resources,* http://mmcif.wwpdb.org.

Westbrook, J., Henrick, K., Ulrich, E. L. & Berman, H. M. (2005). *International tables for crystallography*, edited by S. R. Hall & B. McMahon, pp. 195-198. Dordrecht, The Netherlands: Springer.

Westbrook, J. D. & Burley, S. K. (2019). *Structure* **27**, 211-217.

Westbrook, J. D. & Fitzgerald, P. M. D. (2009). *Structural bioinformatics, second edition*, edited by P. E. Bourne & J. Gu, pp. 271-291. Hoboken, NJ: John Wiley & Sons, Inc.

Williamson, M. P., Havel, T. F. & Wuthrich, K. (1985). *J Mol Biol* **182**, 295-315.

Wlodawer, A. (2007). *Acta Crystallographica. Series D* **63**, 421-423.

wwPDB consortium (2019). *Nucleic Acids Res* **47**, D520-D528.

Young, J. Y., Westbrook, J. D., Feng, Z., Sala, R., Peisach, E., Oldfield, T. J., Sen, S., Gutmanas, A., Armstrong, D. R., Berrisford, J. M., Chen, L., Chen, M., Di Costanzo, L., Dimitropoulos, D., Gao, G., Ghosh, S., Gore, S., Guranovic, V., Hendrickx, P. M., Hudson, B. P., Igarashi, R., Ikegawa, Y., Kobayashi, N., Lawson, C. L., Liang, Y., Mading, S., Mak, L., Mir, M. S., Mukhopadhyay, A., Patwardhan, A., Persikova, I., Rinaldi, L., Sanz-Garcia, E., Sekharan, M. R., Shao, C., Swaminathan, G. J., Tan, L., Ulrich, E. L., van Ginkel, G., Yamashita, R., Yang,

H., Zhuravleva, M. A., Quesada, M., Kleywegt, G. J., Berman, H. M., Markley, J. L., Nakamura, H., Velankar, S. & Burley, S. K. (2017). *Structure* **25**, 536-545.

Zhu, X., Turner, H. L., Lang, S., McBride, R., Bangaru, S., Gilchuk, I. M., Yu, W., Paulson, J. C., Crowe, J. E., Jr., Ward, A. B. & Wilson, I. A. (2019). *Cell Host Microbe* **26**, 729-738 e724.

(a)

```
ATOM      1  N   VAL A   1       6.204  16.869   4.854  1.00 49.05           N
ATOM      2  CA  VAL A   1       6.913  17.759   4.607  1.00 43.14           C
ATOM      3  C   VAL A   1       8.504  17.378   4.797  1.00 24.80           C
ATOM      4  O   VAL A   1       8.805  17.011   5.943  1.00 37.68           O
ATOM      5  CB  VAL A   1       6.369  19.044   5.810  1.00 72.12           C
ATOM      6  CG1 VAL A   1       7.009  20.127   5.418  1.00 61.79           C
ATOM      7  CG2 VAL A   1       5.246  18.533   5.681  1.00 80.12           C
ATOM      8  N   LEU A   2       9.096  18.040   3.857  1.00 26.44           N
ATOM      9  CA  LEU A   2      10.600  17.889   4.283  1.00 26.32           C
ATOM     10  C   LEU A   2      11.265  19.184   5.297  1.00 32.96           C
ATOM     11  O   LEU A   2      10.813  20.177   4.647  1.00 31.90           O
ATOM     12  CB  LEU A   2      11.099  18.007   2.815  1.00 29.23           C
ATOM     13  CG  LEU A   2      11.322  16.956   1.934  1.00 37.71           C
ATOM     14  CD1 LEU A   2      11.468  15.596   2.337  1.00 39.10           C
ATOM     15  CD2 LEU A   2      11.423  17.268   0.300  1.00 37.47           C
```

(b)

```
loop_
_atom_site.group_PDB
_atom_site.id
_atom_site.type_symbol
_atom_site.label_atom_id
_atom_site.label_alt_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_entity_id
_atom_site.label_seq_id
_atom_site.pdbx_PDB_ins_code
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
_atom_site.pdbx_formal_charge
_atom_site.auth_seq_id
_atom_site.auth_comp_id
_atom_site.auth_asym_id
_atom_site.auth_atom_id
_atom_site.pdbx_PDB_model_num
ATOM   1 N N   . VAL A 1 1 ? 6.204  16.869 4.854  1.00 49.05 ? 1 VAL A N   1
ATOM   2 C CA  . VAL A 1 1 ? 6.913  17.759 4.607  1.00 43.14 ? 1 VAL A CA  1
ATOM   3 C C   . VAL A 1 1 ? 8.504  17.378 4.797  1.00 24.80 ? 1 VAL A C   1
ATOM   4 O O   . VAL A 1 1 ? 8.805  17.011 5.943  1.00 37.68 ? 1 VAL A O   1
ATOM   5 C CB  . VAL A 1 1 ? 6.369  19.044 5.810  1.00 72.12 ? 1 VAL A CB  1
ATOM   6 C CG1 . VAL A 1 1 ? 7.009  20.127 5.418  1.00 61.79 ? 1 VAL A CG1 1
ATOM   7 C CG2 . VAL A 1 1 ? 5.246  18.533 5.681  1.00 80.12 ? 1 VAL A CG2 1
ATOM   8 N N   . LEU A 1 2 ? 9.096  18.040 3.857  1.00 26.44 ? 2 LEU A N   1
ATOM   9 C CA  . LEU A 1 2 ? 10.600 17.889 4.283  1.00 26.32 ? 2 LEU A CA  1
ATOM  10 C C   . LEU A 1 2 ? 11.265 19.184 5.297  1.00 32.96 ? 2 LEU A C   1
ATOM  11 O O   . LEU A 1 2 ? 10.813 20.177 4.647  1.00 31.90 ? 2 LEU A O   1
ATOM  12 C CB  . LEU A 1 2 ? 11.099 18.007 2.815  1.00 29.23 ? 2 LEU A CB  1
ATOM  13 C CG  . LEU A 1 2 ? 11.322 16.956 1.934  1.00 37.71 ? 2 LEU A CG  1
ATOM  14 C CD1 . LEU A 1 2 ? 11.468 15.596 2.337  1.00 39.10 ? 2 LEU A CD1 1
ATOM  15 C CD2 . LEU A 1 2 ? 11.423 17.268 0.300  1.00 37.47 ? 2 LEU A CD2 1
```

Figure 1

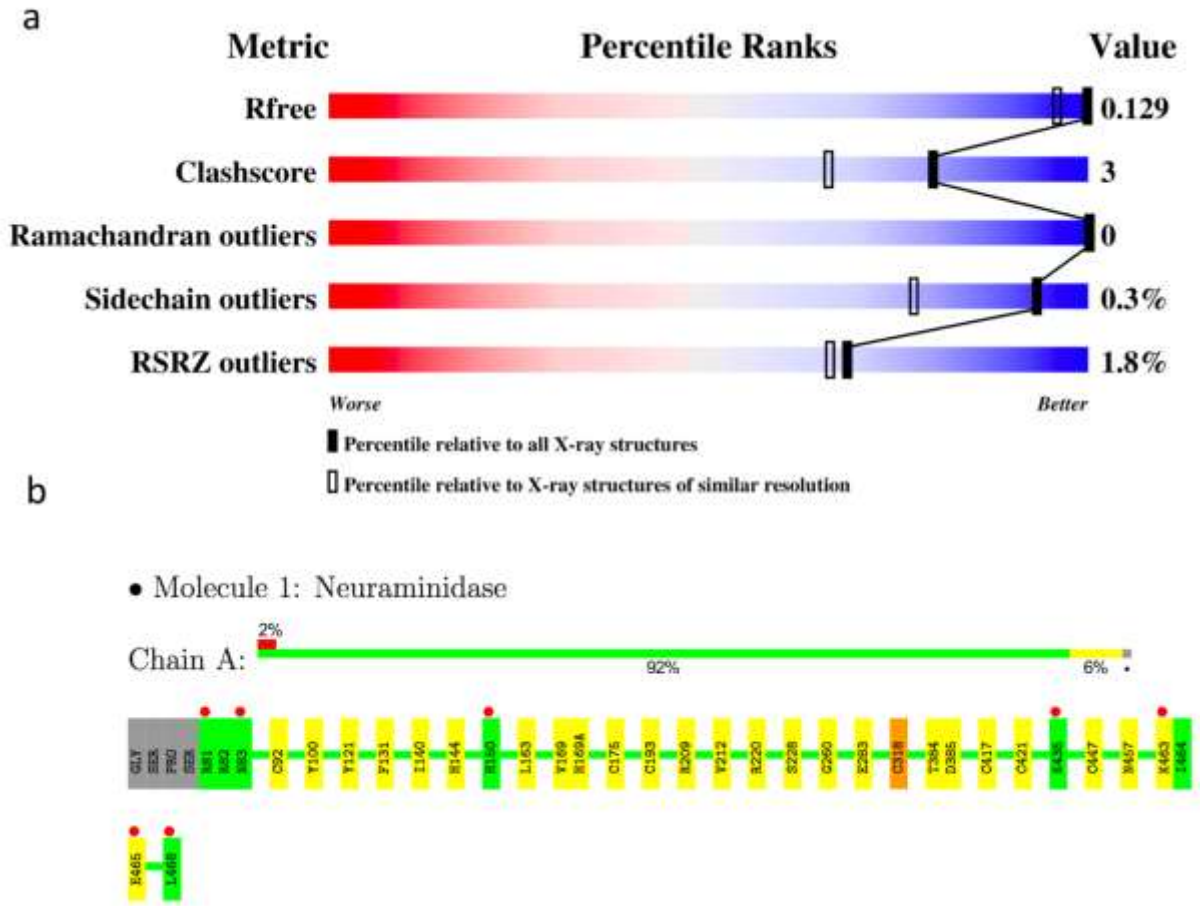
a
Metric
Percentile Ranks
Value
Rfree
0.129
Clashscore
3
Ramachandran outliers
0
Sidechain outliers
0.3%
RSRZ outliers
1.8%
Worse
Better
Percentile relative to all X-ray structures
Percentile relative to X-ray structures of similar resolution
b
• Molecule 1: Neuraminidase
Chain A:
2%
92%
6%

Figure 2

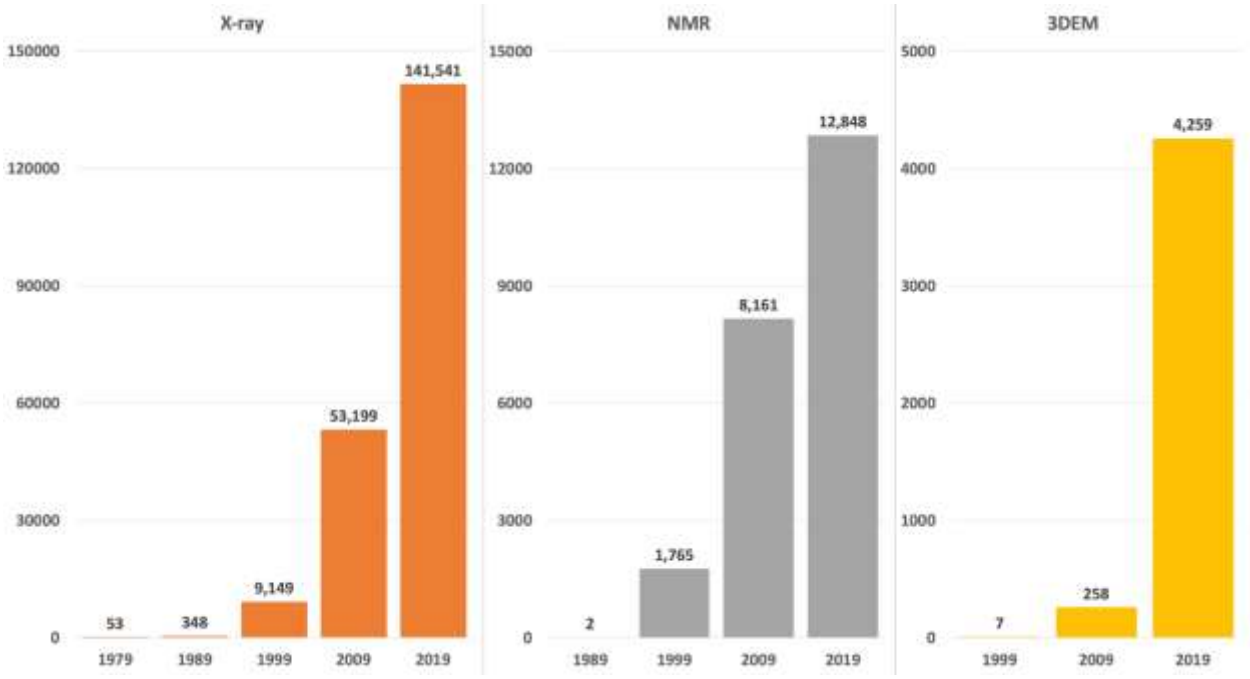
X-ray
150000
120000
90000
60000
30000
0
53
348
9,149
53,199
141,541
1979
1989
1999
2009
2019
NMR
15000
12000
9000
6000
3000
0
2
1,765
8,161
12,848
1989
1999
2009
2019
3DEM
5000
4000
3000
2000
1000
0
7
258
4,259
1999
2009
2019

Figure 3

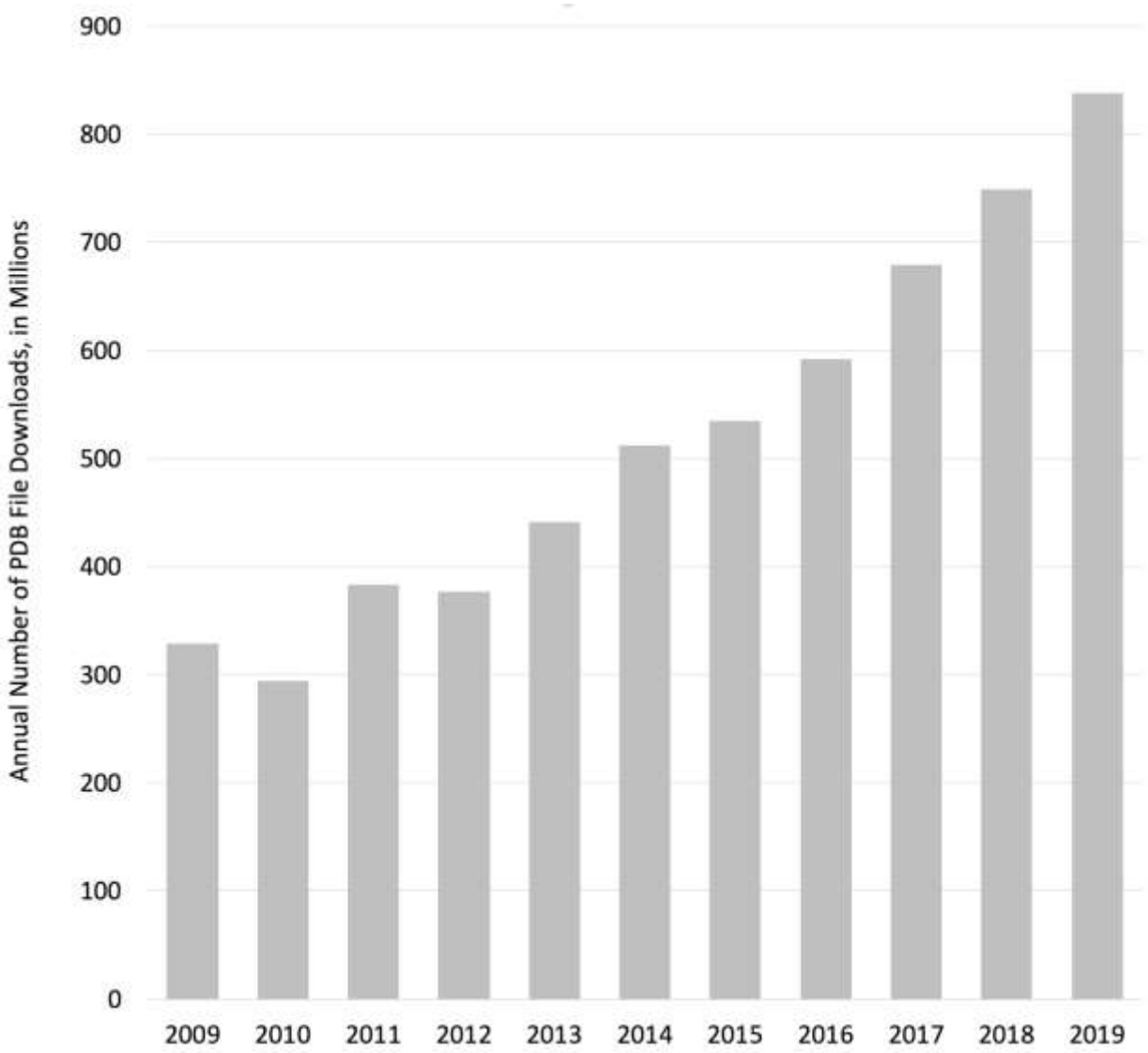
Annual Number of PDB File Downloads, in Millions
900
800
700
600
500
400
300
200
100
0
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019

Figure 4

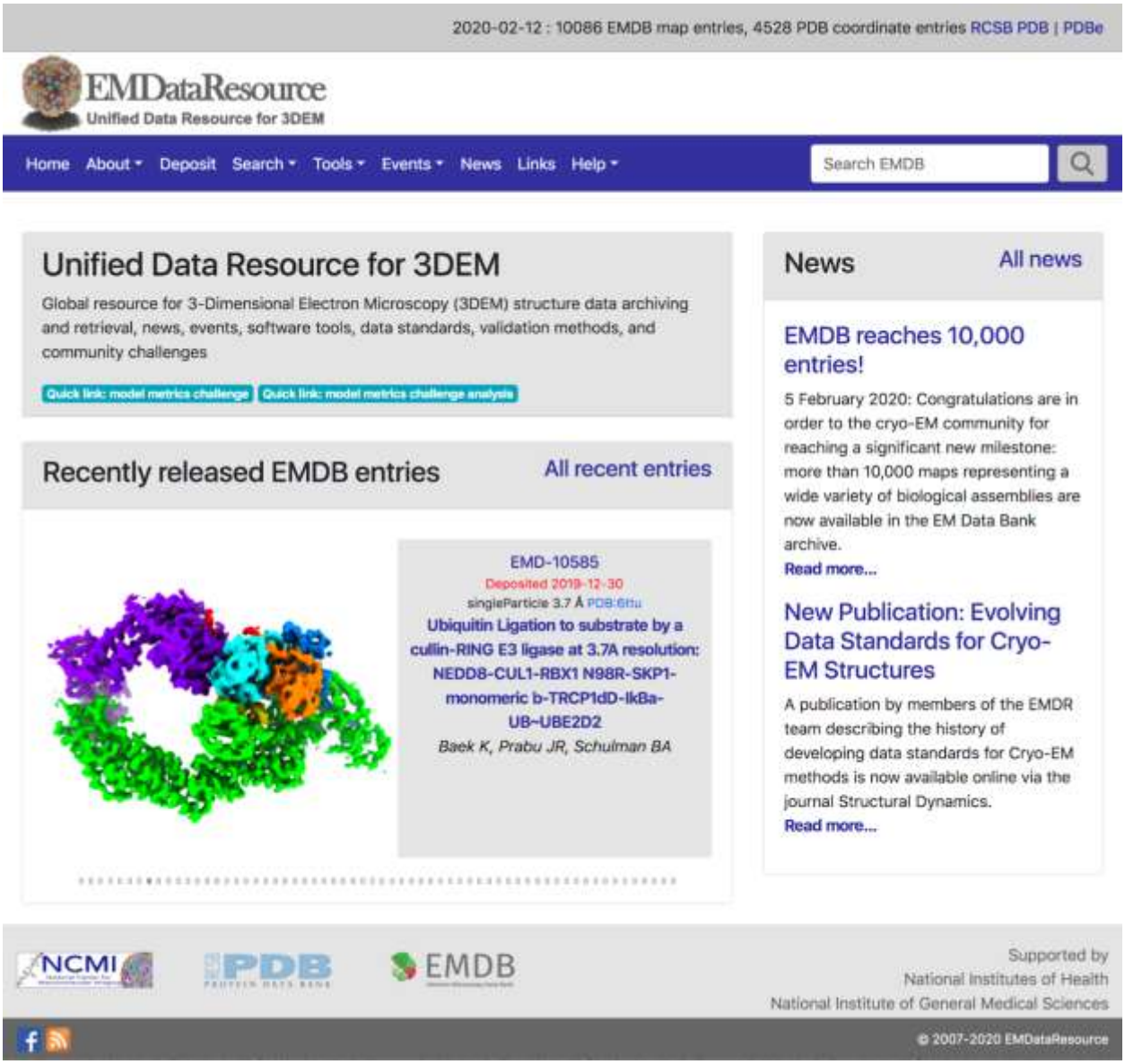
2020-02-12 : 10086 EMDB map entries, 4528 PDB coordinate entries RCSB PDB | PDBe
EMDataResource
Unified Data Resource for 3DEM
Home About Deposit Search Tools Events News Links Help
Search EMDB
Unified Data Resource for 3DEM
Global resource for 3-Dimensional Electron Microscopy (3DEM) structure data archiving and retrieval, news, events, software tools, data standards, validation methods, and community challenges
Quick link: model metrics challenge
Quick link: model metrics challenge analysis
Recently released EMDB entries
All recent entries
EMD-10585
Deposited 2019-12-30
singleParticle 3.7 Å PDB:6tlu
Ubiquitin Ligation to substrate by a cullin-RING E3 ligase at 3.7A resolution: NEDD8-CUL1-RBX1 N98R-SKP1-monomeric b-TRCP1dD-IkBa-UB~UBE2D2
Baek K, Prabu JR, Schulman BA
News
All news
EMDB reaches 10,000 entries!
5 February 2020: Congratulations are in order to the cryo-EM community for reaching a significant new milestone: more than 10,000 maps representing a wide variety of biological assemblies are now available in the EM Data Bank archive.
Read more...
New Publication: Evolving Data Standards for Cryo-EM Structures
A publication by members of the EMDR team describing the history of developing data standards for Cryo-EM methods is now available online via the journal Structural Dynamics.
Read more...
NCMI
PDB
EMDB
Supported by
National Institutes of Health
National Institute of General Medical Sciences
© 2007-2020 EMDataResource

Figure 5a

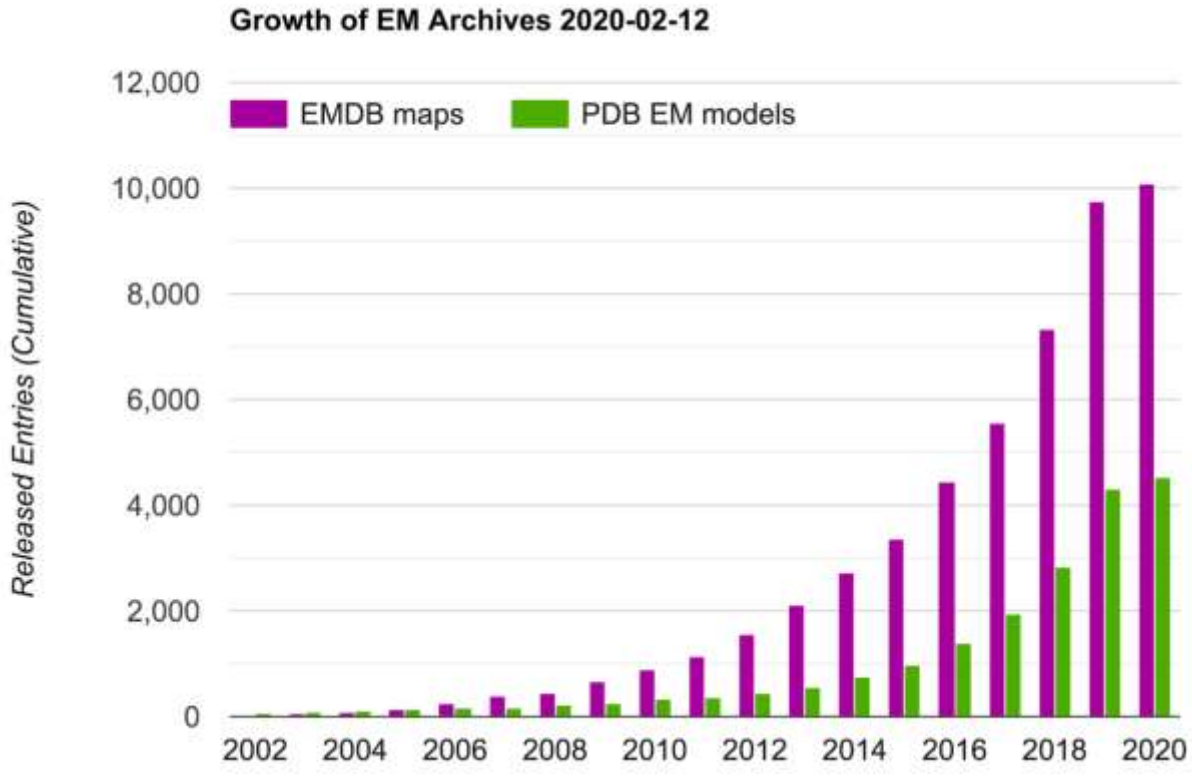
Growth of EM Archives 2020-02-12
EMDB maps
PDB EM models
Released Entries (Cumulative)
12,000
10,000
8,000
6,000
4,000
2,000
0
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020

Figure 5b

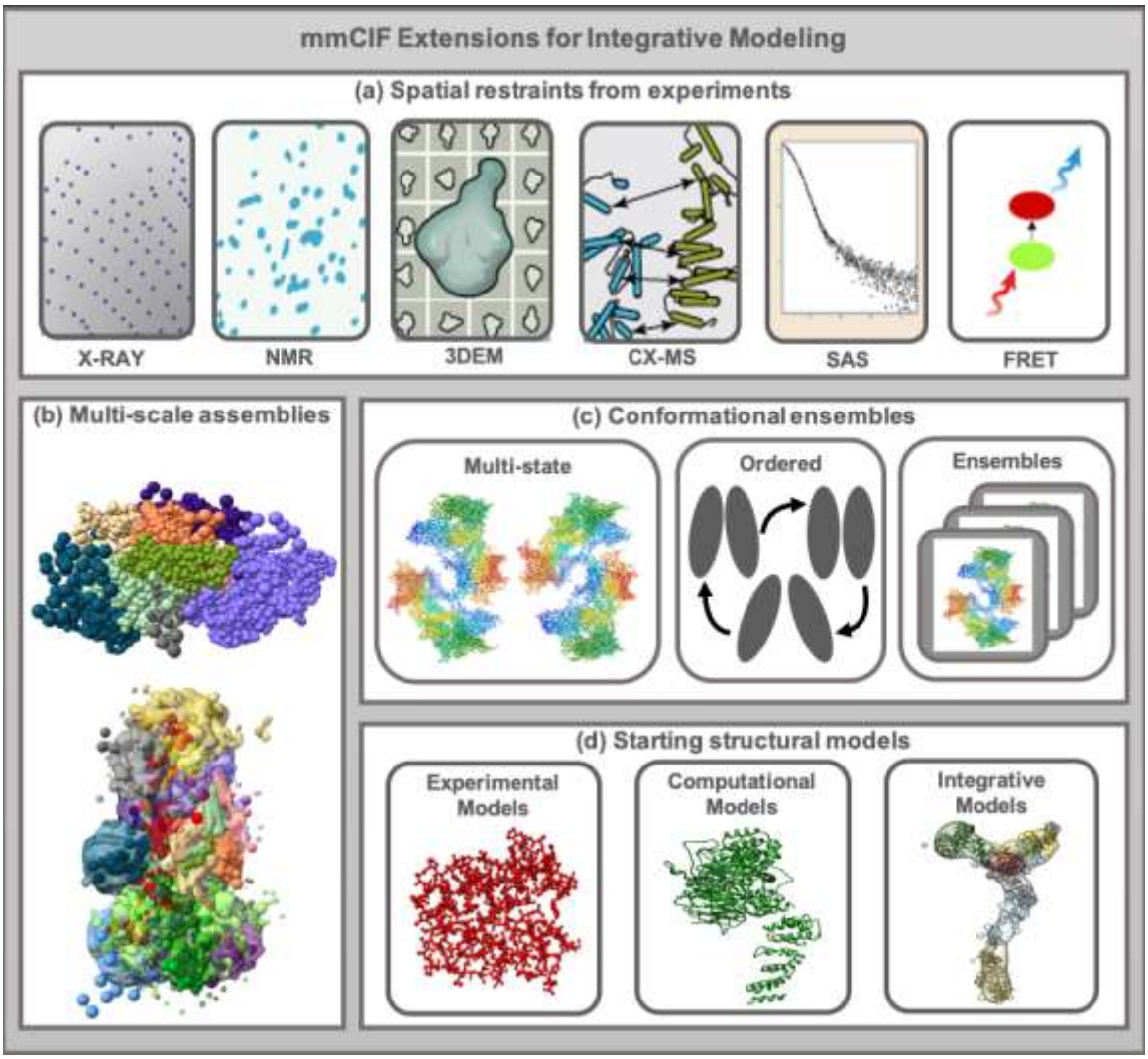
mmCIF Extensions for Integrative Modeling
(a) Spatial restraints from experiments
X-RAY
NMR
3DEM
CX-MS
SAS
FRET
(b) Multi-scale assemblies
(c) Conformational ensembles
Multi-state
Ordered
Ensembles
(d) Starting structural models
Experimental Models
Computational Models
Integrative Models

Figure 6

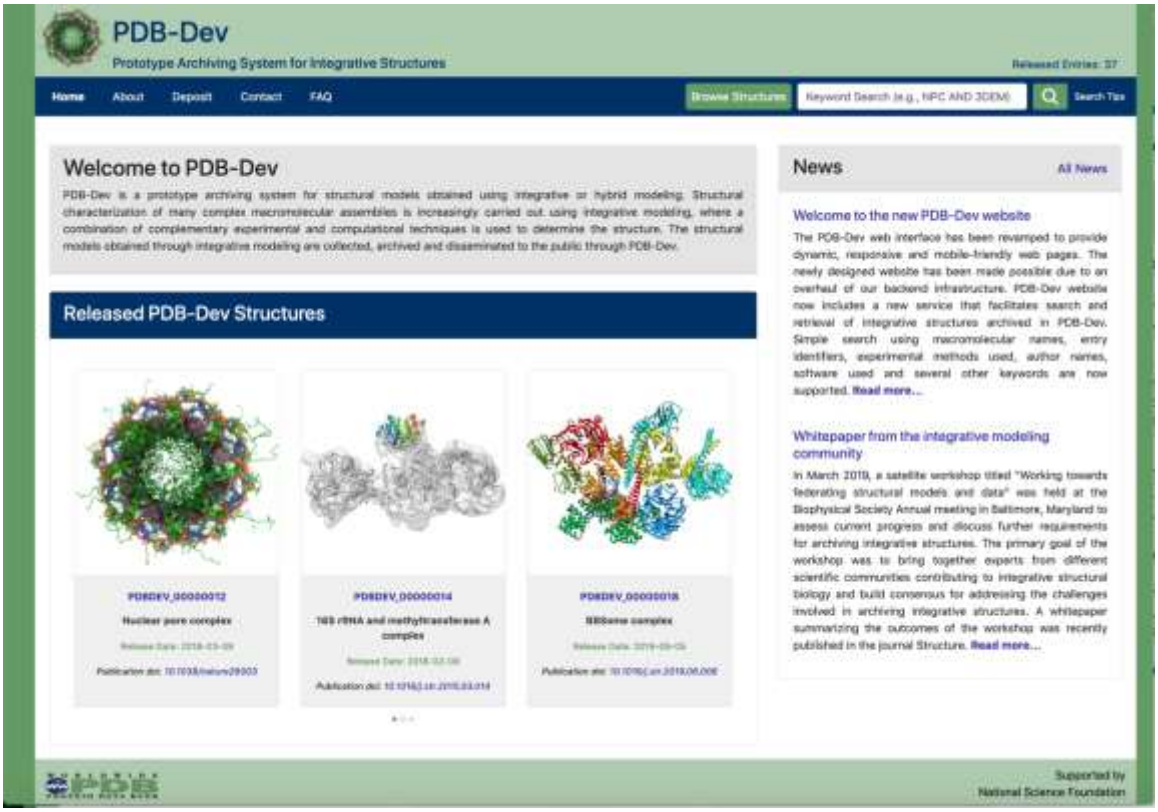
PDB-Dev
Prototype Archiving System for Integrative Structures
Home
About
Deposit
Contact
FAQ
Welcome to PDB-Dev
Released PDB-Dev Structures
News
All News
Welcome to the new PDB-Dev website
Whitepaper from the integrative modeling community
Supported by
National Science Foundation

Figure 7