Chapter 16
Archiving of Integrative Structural Models

Helen M. Berman, Jill Trewhella, Brinda Vallat, and John D. Westbrook

Abstract Integrative or hybrid structural biology involves the determination of three-dimensional structures of macromolecular assemblies by combining information from a variety of experimental and computational methods. Archiving the results of integrative/hybrid modeling methods have complex requirements and existing archiving mechanisms are insufficient to handle these pre-requisites. Three concepts important for archiving integrative/hybrid models are presented in this chapter: (1) building a federated network of structural model and experimental data archives, (2) development of a common set of data standards, and (3) creation of mechanisms for interoperation and data exchange among the repositories in a federation. Methods proposed for achieving these objectives are also discussed.

Keywords Protein Data Bank · Integrative/hybrid modeling methods · PDBx/mmCIF · Data standards · Data exchange · Structural biology federation

H. M. Berman (✉)
RCSB Protein Data Bank, Department of Chemistry and Chemical Biology, Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ, USA
e-mail: berman@rcsb.rutgers.edu

J. Trewhella
School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

Department of Chemistry, University of Utah, Salt Lake City, UT, USA

B. Vallat
RCSB, Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ, USA

J. D. Westbrook
RCSB Protein Data Bank, Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ, USA

H. Nakamura et al. (eds.), *Integrative Structural Biology with Hybrid Methods*,
Advances in Experimental Medicine and Biology 1105,
https://doi.org/10.1007/978-981-13-2200-6_16

16.1 Introduction

The field of structural biology has undergone dramatic growth and change in the 60 plus years since Kendrew determined the structure of myoglobin (Kendrew et al. 1958) and Perutz the structure of hemoglobin (Perutz et al. 1960) – the first atomic structures of macromolecular proteins determined using X-ray crystallography. Today, while individual biomolecular structures of the highest resolution and accuracy remain central to the field, the next frontier in structural molecular biology is characterization of the large, complex and dynamic macromolecular networks and machinery that drive fundamental biological processes such as replication, transcription, concerted movement, defense against infection, etc. These targets are elusive to traditional approaches to structure determination that use a single technique, such as X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy or 3D Electron Microscopy (3DEM). To address this problem, integrative or hybrid (I/H) methods are being developed that combine data from complementary experimental techniques and computational models in innovative ways (Sali et al. 2015, Ward et al. 2013). For example, I/H methods have been used to develop detailed molecular models of the molecular machines and assemblies that control protein biosynthesis (ribosome) (Leitner et al. 2016), the movement of proteins across the nuclear membrane in a cell (nuclear pore complex) (Kim et al. 2018), sensing in pathogenic bacteria that enables infection (bacterial type III secretion system) (Loquet et al. 2012), and the regulation of the degradation of damaged, malfunctioning or toxic proteins in the cell (proteasomal lid sub-complex) (Politis et al. 2014).

The Protein Data Bank (*PDB*), founded in 1971 with only seven protein structures (Protein Data Bank 1971), is today a searchable, open global archive that holds more than 140,000 structures of biological macromolecules and their complexes, all of which are freely accessible. The vast majority of deposited structures have been determined by a single technique: X-ray crystallography, NMR spectroscopy or 3D electron microscopy. The Model Archive (*MA*) (Haas et al. 2013; Haas and Schwede 2013), managed by the Protein Model Portal (PMP), archives about 1400 *in silico* models derived using purely computational techniques. Well-developed infrastructure is in place for these structural model archives, with efficient deposition and data processing procedures along with data standards, validation and curation methods.

The increasingly diverse data types used in I/H methods has led to models that can span multiple spatiotemporal scales and conformational states. Therefore, existing archiving mechanisms that are designed for individual atomistic structures, are insufficient to capture the details of an I/H model. The necessary requirements for processing and archiving I/H models have yet to be fully established. In recognition of this problem, the worldwide PDB (wwPDB) (Berman et al. 2007) established the I/H Methods Task Force, and in October 2014, a workshop was held (Sali et al. 2015) at the European Bioinformatics Institute, Hinxton, UK. Thirty-eight leaders in experimental structural biology, *in silico* and integrative modeling,

visualization, and data archiving discussed the steps required to make the results of I/H modeling publicly available. They converged on the set of recommendations summarized below (Sali et al. 2015):

Recommendation 1: In addition to archiving the models themselves, all relevant experimental data and metadata as well as experimental and computational protocols should be archived; inclusivity is key.

Recommendation 2: A flexible model representation needs to be developed, allowing for multi-scale models (with atomistic and non-atomistic representations), multi-state models (existing in various conformations), ensembles of models, and models related by time or other order.

Recommendation 3: Procedures for estimating the uncertainty of integrative models should be developed, validated, and adopted.

Recommendation 4: A federation of model and data archives should be created.

Recommendation 5: Publication standards for integrative models should be established.

Implementation of these recommendations will take years of research and community building efforts. However, the key recommendations involving the creation of a federated system of model and data archives and the development of a flexible data representation are crucial for archiving I/H models and hence are being addressed presently.

16.2 The Structural Biology Federation

Models determined by I/H methods utilize the data from a wide range of biophysical methodologies, including but not limited to: X-ray crystallography, NMR spectroscopy, 3DEM, Small Angle Scattering (SAS), Förster Resonance Energy Transfer (FRET), Chemical Crosslinking and Mass Spectrometry (CX-MS), Electron Paramagnetic Resonance (EPR) spectroscopy, Atomic Force Microscopy (AFM), deep sequencing and coevolution methods and other proteomics and bioinformatics techniques (Ward et al. 2013; Whitehead et al. 2012; Hopf et al. 2014). Experimental data from complementary methods are combined to provide a set of spatial restraints and structural information that are used in the determination of the three-dimensional structures of macromolecular assemblies. Currently, these data are stored in a variety of places. The atomic coordinates of structural models derived by X-ray crystallography, NMR spectroscopy, and 3DEM are archived in the *PDB* (Berman et al. 2000) along with data needed for model validation such as the structure factors from X-ray crystallography and NMR chemical shifts. There are also several experimental data repositories that store information belonging to the particular domain: the Electron Microscopy Data Bank (*EMDB*) (Patwardhan and Lawson 2016) (Lawson et al. 2011) archives the 3DEM maps as well as extensive metadata; BioMagResBank (*BMRB*) (Ulrich et al. 2008) contains NMR spectra, chemical shifts and other NMR-derived information such as NOE restraints and coupling constants; Small Angle Scattering Biological Data Bank (*SASBDB*) (Valentini et al. 2015) and *BIOISIS* (Rambo et al. 2017) contain small-angle scattering data

and models; members of the ProteomeXchange consortium (Vizcaino et al. 2014) including PRIDE (Vizcaino et al. 2016) and PeptideAtlas (Desiere et al. 2006) archive proteomics data as well as results from chemical crosslinking and mass spectrometry experiments. For other experimental methods, such as FRET and EPR, there are no standard mechanisms to archive the experimental data. As a result, there may be cloud-hosted data sets on external sites such as GitHub (GitHub Inc. 2007), or perhaps most commonly, un-hosted data sets not usually accessible to the public that reside in individual research laboratories.

In addition to archiving the three-dimensional coordinates of structural models, it is necessary to archive metadata describing the chemistry and the protocols used to determine the model, as well as the subset of experimental data needed to validate the models. Furthermore, many communities want and need a broader set of experimental data and metadata archived so that they can be available for future research.

To accommodate the need for an archive of validated models, and archives for the different experimental methods used to compute these models, a federated system of model and data archives was recommended by the I/H Methods Task Force (Sali et al. 2015). A conceptual diagram of this Federation is shown in Fig. 16.1. At the center of the figure are the principal structural biology model repositories, including the existing ***PDB*** and ***MA*** archives, along with a prototype ***PDB-Dev*** system, which hosts I/H models and associated spatial restraints (Vallat et al. 2016c, 2018; Burley et al. 2017). The outside ring includes complementary experimental data repositories that would share a subset of experimental data and metadata with the structural model repositories at the center, while continuing to provide the full complement of data for their specialist communities. An important component

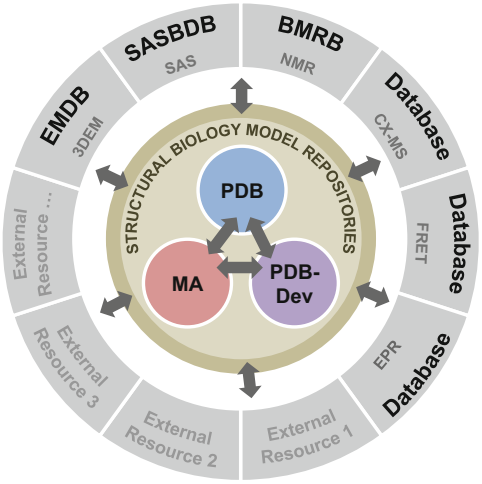

Fig. 16.1 A conceptual diagram of the proposed members of the federation. Repositories that focus on macromolecular structural models are shown in the center of the figure (structural biology model repositories), while examples of repositories that contain primary experimental data and/or derived restraints and associated metadata are shown in the outer circle. This outer circle contains only some examples of experimental data archives

of this federation is the establishment of methods for data exchange among the individual repositories. The data definitions supporting these repositories need to be well-aligned and software tools required for this purpose need to be developed. The I/H models of complex biological systems will likely evolve with time as new and different kinds of data become available. Therefore, the data exchange mechanisms should be able to support these evolutionary improvements. The creation of a Federation will provide a unified network of resources for structural biology models and data and will further enable the development of mechanisms for communication and interoperation among the different scientific communities contributing to structural biology.

16.3 Creation of Data Standards

One of the important pre-requisites for building an archive is the creation of data standards. The data standards, usually defined in a "dictionary" of data terms, provide the descriptions and specifications for the information stored in an archive. These data specifications include precise definitions for the data terms including their units and allowed ranges, software features, storage data formats, and data relationships and dependencies. To build an interoperable federated system of structural biology resources, it is necessary that each participating repository has well-defined data standards.

The scope of the contents to be archived varies among the data repositories. Ideally, the archived content contains the minimum information needed to accurately represent a complete and reproducible experiment. Experimental data repositories typically capture the sample conditions, the experimental methods and software tools used, the primary results and derived data, and associated metadata. Structural model repositories capture atomic and molecular descriptions along with metadata related to the structure determination method. The scope of the data content and formats for data standards among different repositories are not always the same.

The *PDB* archive uses the PDBx/mmCIF data standard (Fitzgerald et al. 2005) that grew out of an effort by the crystallographic community to define the many elements of the crystallographic experiments and the results derived from those experiments. The initial dictionary contained about 3000 data items, which is now expanded to about 6500. Terms specific for NMR and for 3DEM were added as structural models derived from those methods were deposited and processed by the *PDB*. In addition to the atomic coordinates of the models, the *PDB* also stores experimental data that are essential for validating these structures. These include X-ray structure factors, NMR chemical shifts and restraints, all of which are defined in the PDBx/mmCIF data dictionary (Fitzgerald et al. 2005).

The experimental data repositories that are members of the Federation, archive method-specific data and metadata. They require a compatible data representation that serves the needs of the community. BMRB (Ulrich et al. 2008) has a large array of NMR specific spectral data such as the chemical shifts, NOE restraints

and coupling constants. The underlying data representation is based on the NMR-Star format (BioMagResBank 2004), which is a close relative of the PDBx/mmCIF data representation. EMDB (Patwardhan and Lawson 2016) (Lawson et al. 2011) contains 3DEM-derived maps expressed in CCP4 format (Winn et al. 2011) and a database that follows an internally defined XML format. SASBDB (Valentini et al. 2015) archives the results of solution scattering experiments and has adopted an extension of the PDBx/mmCIF dictionary, called sasCIF (Kachala et al. 2016; Malfois and Svergun 2000). The sasCIF extension provides SASBDB the advantages of pre-aligned data definitions and seamless interoperation with the *PDB*. Other communities that generate *in silico structural models*, CX-MS data, FRET data, EPR data, and deep genome sequencing are in various stages of creating standards for their disciplines.

The creation of an I/H model archive requires the development of a flexible data representation as recommended by the wwPDB I/H Methods Task Force. The existing data pipeline of the *PDB* archive is insufficient to handle I/H models because the *PDB* currently handles mono-scale atomistic structures derived from experimental techniques such as X-ray crystallography, NMR spectroscopy, and 3DEM. The data representation for I/H models should account for ensembles of multi-scale structural models (comprising of atomistic and coarse-grained representations of macromolecular assemblies), conformations in multiple states and models related by time or other order. It is envisioned that multi-scale I/H models can span a broad range of structures including those of individual molecules, their complexes, cellular neighborhoods, and even the entire cell. Furthermore, the input spatial restraints used in I/H modeling can be obtained from a variety of experimental and computational techniques and hence, the data representation should be able to comprehensively capture this information together with details of modeling workflows and other relevant metadata.

An I/H methods data dictionary has been created (Berman et al. 2016, Vallat et al. 2016a, b, 2018) that defines the data contents from an I/H investigation to be archived. This dictionary is an extension of the PDBx/mmCIF dictionary (Fitzgerald et al. 2005) and therefore is complementary to the definitions already present in the PDBx/mmCIF dictionary such as descriptions of the molecular system, atomic coordinates, metadata related to authors, citations, and software use. New definitions have been created to represent multi-scale structural models (including coarse-grained spheres and three-dimensional Gaussian volumes), multi-state and time ordered ensembles, starting structural models used as input in the I/H modeling and restraints derived from experimental methods such as CX-MS, 2DEM, 3DEM and SAS. Preliminary information regarding the modeling workflows and validation metrics are also defined in the dictionary. The initial set of definitions have been created based on the I/H models obtained from the Integrative Modeling Platform (IMP, (Russel et al. 2012)) software package. Figure 16.2 shows a schematic representation of the contents of the I/H methods data dictionary.

The PDB-Dev system (Vallat et al. 2016c, 2018; Burley et al. 2017) has been built based on the new I/H methods extension dictionary. At present, twenty two structures covering a variety of I/H modeling software and experimental data types

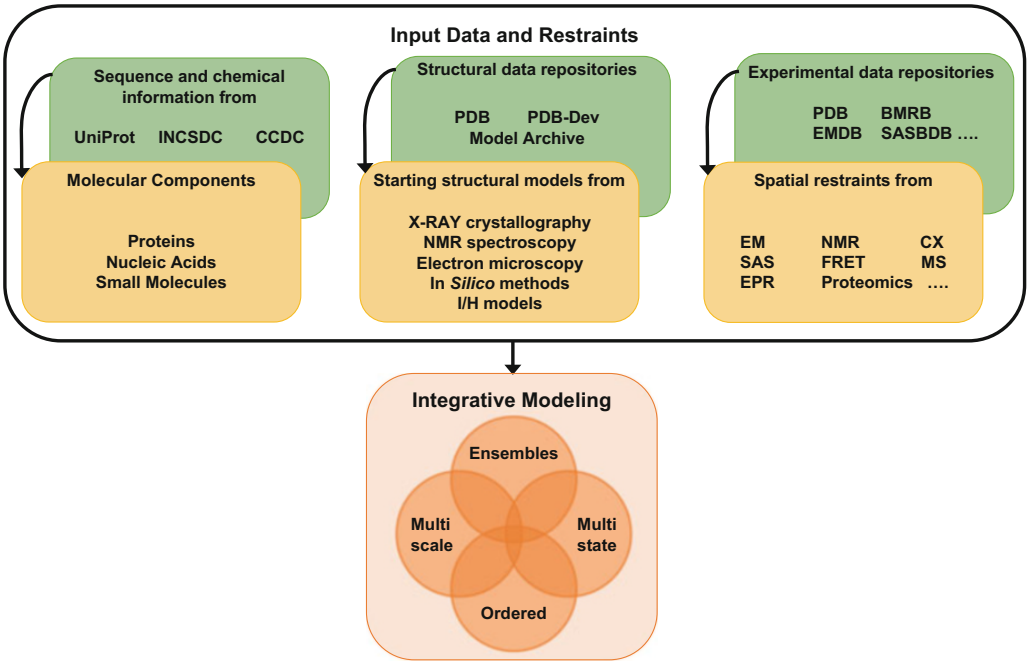

Fig. 16.2 Illustration of the data content captured in the integrative/hybrid methods dictionary (Berman et al. 2016; Vallat et al. 2016a, b, 2018). The green boxes represent existing external repositories that archive sequence, chemical, structural, and experimental data for biological macromolecules. The yellow, orange, and blue boxes represent the information captured in the recently developed I/H methods dictionary. This information includes details of the molecular components, the starting structural models of individual molecular components, and the spatial restraints derived from various experimental methods. The details of the integrative modeling algorithm are also captured in the dictionary including definitions for multi-scale, multi-state and ordered structural ensembles of macromolecular assemblies

have been deposited into PDB-Dev. These structures and associated spatial restraints are available from the *PDB-Dev* website (Vallat et al. 2016c, 2018; Burley et al. 2017) in a format compliant with the new I/H methods data dictionary (Berman et al. 2016; Vallat et al. 2016a, b, 2018). The ChimeraX visualization software (Ferrin et al. 2017) provides basic support to visualize the multi-scale I/H models obtained from *PDB-Dev*.

Following the recommendations of the wwPDB I/H Methods Task Force, we have assembled a set of data standards and a prototype deposition and archiving system that lays the foundation for building a full-fledged archive for I/H models. The development of a comprehensive data pipeline to curate and validate these I/H structural models to provide cleaner and richer data content to the users, is the focus of ongoing research projects.

16.4 Methods for Data Exchange

The proposed federation comprises a network of information resources that contribute to the field of structural biology. The creation of a federation will greatly streamline the process of data preservation and access. The basis of such a federation is the establishment of mechanisms for exchanging information among its various members. This important process requires extensive participation and consensus building among the communities involved.

Experience suggests that the organization of the structural biology federation be based on autonomous repositories networked via a set of mutually agreed communication and data exchange protocols. The diversity of archived data types and data validation protocols require the greatest local autonomy in establishing data formats and standards, and to build and maintain each individual repository. Mutually agreed mechanisms are then required to enable member repositories to interoperate with each other in an effective manner including efficient methods for communication and data exchange. The objective of seamless interoperation with the federation can be achieved in several ways, as proposed below, and these may be adopted based on community consensus.

References to data residing in other repositories will rely on high level identifiers such as Digital Object Identifiers (DOIs), stable accession codes and persistent URLs. While experimental data and structural models will reside in their respective repositories, the spatial restraints and associated information derived from the experimental data, required for validation of the structural model, will be shared among the repositories. The limited set of commonly shared information need to be identified and defined accordingly to avoid duplication and to enable semantically precise data exchange. Software tools need to be developed to facilitate seamless interoperation among the repositories in the federation. These tools include development of methods for data harvesting, format conversion, semantic mapping and alignment of data residing in different repositories as well as mechanisms for exchange of data using secure industry-standard web services. Figure 16.3 shows a schematic representation of different layers of interoperation among various structural model and experimental data repositories in the proposed federation as well as developers of I/H modeling software.

To account for refinements of the structural models arising from revisions to the underlying experimental data and/or modeling methods, data exchange mechanisms should support versioning and updates to data residing in a particular repository. Timely propagation of updated information to other repositories within the federation will also need to be supported. These objectives can be achieved through mutual agreements on maintaining explicitly versioned data files and unambiguous descriptions of accession codes and version numbers within the commonly shared data definitions. Furthermore, automated messaging and communication tools are required to enable downstream dissemination of data updates.

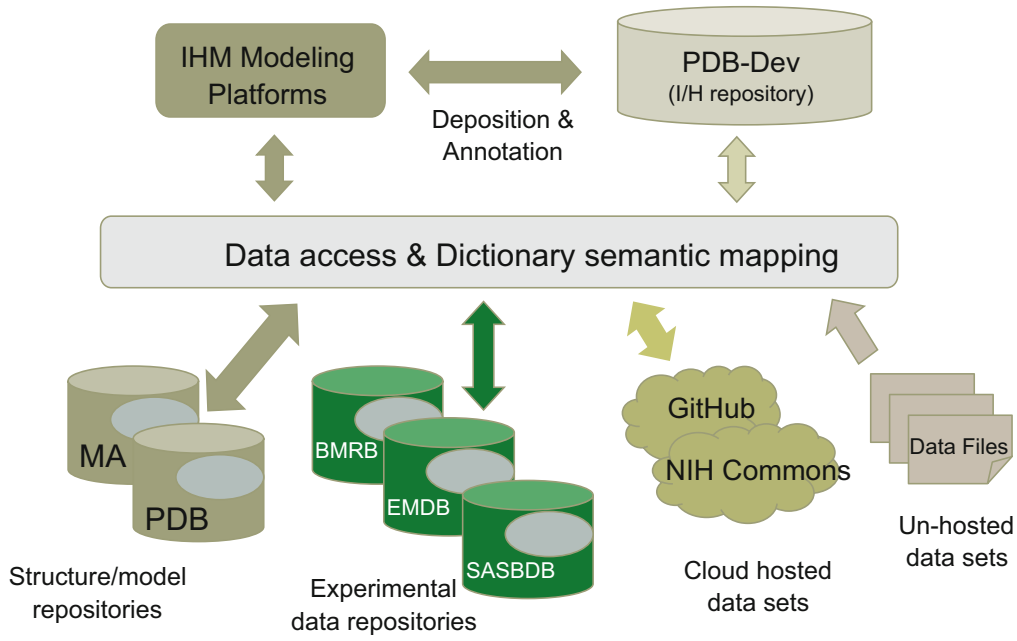

Fig. 16.3 A schematic portrayal of the data exchange among the structural model and experimental data repositories in the proposed structural biology federation

16.5 Conclusion

The future of structural biology relies heavily on the development of integrative/hybrid methods that combine information from a variety of experimental data sources with computational methods to elucidate the structures of complex macromolecular assemblies. These I/H methods are evolving into techniques that provide spatiotemporal information regarding molecular events at the cellular level. From an archival perspective, it is important to capture every structural and functional detail so that the knowledge gained from I/H models can be available for other applications in biotechnology and medicine as well as to guide future research. The structural biology community and the worldwide PDB (wwPDB, (Berman et al. 2007)) have combined their efforts to enable the archiving and dissemination of I/H models and associated experimental data and computational protocols in a concerted manner. Although the long-term vision of a comprehensive structural biology federation is yet to be fully materialized, the first steps in this direction have been productive and basic building blocks have been developed. These steps include bringing together several research communities contributing to the field of structural biology and the development of preliminary data standards and a prototype archiving system for I/H models. Further progress towards the establishment of a unified, global and interoperable network of structural biology resources that provides rich content of curated and validated structural data to the users, is the focus of ongoing and future research and community building efforts.

Acknowledgements This work was supported by NSF EAGER grant DBI-1519158. We thank our collaborators Andrej Sali and Benjamin Webb for their contributions towards the development of the I/H methods data dictionary and the PDB-Dev prototype system. We thank all our colleagues in the I/H methods Task Force and the members of the wwPDB for their help and support with this project.

References

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242

Berman H, Henrick K, Nakamura H, Markley JL (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Res 35:D301–D303

Berman HM, Westbrook J, Vallat B, Webb B, Sali A (2016) A data dictionary for archiving integrative/hybrid models. 66th annual meeting of the American Crystallographic Association, Denver, CO, USA, pp. 85–SA

BioMagResBank (2004) NMRSTAR: dictionary version 3.1.1.78 [Online]. The Board of Regents of the University of Wisconsin System. Available: http://www.bmrb.wisc.edu/dictionary/. Accessed 15 Dec 2015

Burley SK, Kurisu G, Markley J, Nakamura H, Velankar S, Berman HM, Sali A, Schwede T, Trewhella J (2017) PDB-Dev: a prototype system for depositing integrative/hybrid structural models. Structure 25:1317–1318

Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, Chen S, Eddes J, Loevenich SN, Aebersold R (2006) the peptide atlas project. Nucleic Acids Res 34:D655–D658

Ferrin T, Huang C, Peeterson E, Goddard T, Couch G, Meng E, Morris S (2017) UCSF ChimeraX [Online]. Available: https://www.rbvi.ucsf.edu/chimerax/. Accessed 5 July 2017

Fitzgerald P MD, Westbrook JD, Bourne PE, McMahon B, Watenpaugh KD, Berman HM (2005) 4.5 macromolecular dictionary (mmCIF). In: Hall SR, Mcmahon B (eds) International tables for crystallography G. Definition and exchange of crystallographic data. Springer, Dordrecht, pp. 295–443

GitHub Inc (2007) GitHub: how people build software [Online]. Available http://www.github.com/. Accessed 1 Nov 2013

Haas J, Schwede T (2013) Model archive [Online]. Available: http://www.modelarchive.org/. Accessed 12 Oct 2016

Haas J, Roth S, Arnold K, Kiefer F, Schmidt T, Bordoli L, Schwede T (2013) The protein model portal – a comprehensive resource for protein structure and model information. Database (Oxford), 2013, bat031

Hopf TA, Scharfe CP, Rodrigues JP, Green AG, Kohlbacher O, Sander C, Bonvin AM, Marks DS (2014) Sequence co-evolution gives 3D contacts and structures of protein complexes. elife 3

Kachala M, Westbrook J, Svergun D (2016) Extension of the sasCIF format and its applications for data processing and deposition. J Appl Crystallogr 49:302–310

Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature 181:662–666

Kim SJ, Fernandez-Martinez J, Nudelman I, Shi Y, Zhang W, Raveh B, Herricks T, Slaughter BD, Hogan JA, Upla P, Chemmama IE, Pellarin R, Echeverria I, Shivaraju M, Chaudhury AS, Wang J, Williams R, Unruh JR, Greenberg CH, Jacobs EY, Yu Z, de la Cruz MJ, Mironska R, Stokes DL, Aitchison JD, Jarrold MF, Gerton JL, Ludtke SJ, Akey CW, Chait BT, Sali A, Rout MP (2018) Integrative structure and functional anatomy of a nuclear pore complex. Nature 555(7697):475–482

Lawson CL, Baker ML, Best C, Bi C, Dougherty M, Feng P, van Ginkel G, Devkota B, Lagerstedt I, Ludtke SJ, Newman RH, Oldfield TJ, Rees I, Sahni G, Sala R, Velankar S, Warren J, Westbrook JD, Henrick K, Kleywegt GJ, Berman HM, Chiu W (2011) EMDataBank.org: unified data resource for CryoEM. Nucleic Acids Res 39:D456–D464

Leitner A, Faini M, Stengel F, Aebersold R (2016) Crosslinking and mass spectrometry: an integrated technology to understand the structure and function of molecular machines. Trends Biochem Sci 41:20–32

Loquet A, Sgourakis NG, Gupta R, Giller K, Riedel D, Goosmann C, Griesinger C, Kolbe M, Baker D, Becker S, Lange A (2012) Atomic model of the type III secretion system needle. Nature 486:276–279

Malfois M, Svergun DI (2000) sasCIF: an extension of core crystallographic information file for SAS. J Appl Crystallogr 33:812–816

Patwardhan A, Lawson CL (2016) Databases and archiving for CryoEM. Methods Enzymol 579:393–412

Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North ACT (1960) Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5 Å resolution, obtained by X-ray analysis. Nature 185:416–422

Politis A, Stengel F, Hall Z, Hernandez H, Leitner A, Walzthoeni T, Robinson CV, Aebersold R (2014) A mass spectrometry-based hybrid method for structural modeling of protein complexes. Nat Methods 11:403–406

Protein Data Bank (1971) Protein Data Bank. Nature New Biol 233:223

Rambo RP, Tainer JA, Hura GL (2017) BIOISIS [Online]. Available: http://www.bioisis.net/. Accessed 7 Aug 2017

Russel D, Lasker K, Webb B, Velazquez-Muriel J, Tjioe E, Schneidman-Duhovny D, Peterson B, Sali A (2012) Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. PLoS Biol 10:e1001244

Sali A, Berman HM, Schwede T, Trewhella J, Kleywegt G, Burley SK, Markley J, Nakamura H, Adams P, Bonvin AM, Chiu W, Peraro MD, Di Maio F, Ferrin TE, Grunewald K, Gutmanas A, Henderson R, Hummer G, Iwasaki K, Johnson G, Lawson CL, Meiler J, Marti-Renom MA, Montelione GT, Nilges M, Nussinov R, Patwardhan A, Rappsilber J, Read RJ, Saibil H, Schroder GF, Schwieters CD, Seidel CA, Svergun D, Topf M, Ulrich EL, Velankar S, Westbrook JD (2015) Outcome of the first wwPDB hybrid/integrative methods task force workshop. Structure 23:1156–1167

Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL (2008) BioMagResBank. Nucleic Acids Res 36:D402–D408

Valentini E, Kikhney AG, Previtali G, Jeffries CM, Svergun DI (2015) SASBDB, a repository for biological small-angle scattering data. Nucleic Acids Res 43:D357–D363

Vallat B, Webb B, Westbrook J, Sali A, Berman H (2016a) Integrative/hybrid methods PDBx/mmCIF dictionary extension [Online]. Available: https://github.com/ihmwg/IHM-dictionary. Accessed 9 June 2016

Vallat B, Webb B, Westbrook J, Sali A, Berman H (2016b) Integrative/hybrid methods PDBx/mmCIF dictionary extension documentation [Online]. Available: https://github.com/ihmwg/IHM-dictionary/tree/master/dictionary_documentation. Accessed 9 June 2016

Vallat B, Webb B, Westbrook J, Sali A, Berman HM (2016c) The PDB-Dev prototype deposition and archiving system [Online]. Available: https://pdb-dev.wwpdb.org/. Accessed 31 Aug 2016

Vallat B, Webb B, Westbrook JD, Sali A, Berman HM (2018) Development of a prototype system for archiving integrative/hybrid structure models of biological macromolecules. Structure 26(6):894–904 e2

Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, Dianes JA, Sun Z, Farrah T, Bandeira N, Binz PA, Xenarios I, Eisenacher M, Mayer G, Gatto L, Campos A, Chalkley RJ, Kraus HJ, Albar JP, Martinez-Bartolome S, Apweiler R, Omenn GS, Martens L, Jones AR, Hermjakob H (2014) ProteomeX change provides globally coordinated proteomics data submission and dissemination. Nat Biotechnol 32:223–226

Vizcaino JA, Csordas A, Del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, Xu QW, Wang R, Hermjakob H (2016) 2016 update of the PRIDE database and its related tools. Nucleic Acids Res 44:D447–D456

Ward AB, Sali A, Wilson IA (2013) Biochemistry. Integrative structural biology. Science 339: 913–915

Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, Myers CA, Kamisetty H, Blair P, Wilson IA, Baker D (2012) Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat Biotechnol 30:543–548

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 67:235–242