



# IHMValidation: Assessment of Integrative Structure Models Deposited to the Protein Data Bank <sup>☆</sup>

Arthur O. Zalevsky <sup>1\*</sup>, Brinda Vallat <sup>2,3\*</sup>, Benjamin M. Webb <sup>1</sup>,  
 Hongsuda Tangmunarunkit <sup>4</sup>, Monica R. Sekharan <sup>2</sup>, Aref Shafeebejestan <sup>4</sup>,  
 Sai Ganesan <sup>1</sup>, Jared Sagendorf <sup>1</sup>, Cy M. Jeffries <sup>5,6</sup>, Jill Trehella <sup>7,8</sup>,  
 Andrea Graziadei <sup>9</sup>, Juan Antonio Vizcaíno <sup>10</sup>, Alexander Leitner <sup>11</sup>,  
 Juri Rappaport <sup>12,13</sup>, Ezra Peisach <sup>2</sup>, Justin W. Flatt <sup>2</sup>, Jasmine Y. Young <sup>2</sup>,  
 Kartik Majila <sup>14</sup>, Shruthi Viswanath <sup>14</sup>, Carl Kesselman <sup>4</sup>, Jeffrey C. Hoch <sup>15</sup>,  
 Genji Kurisu <sup>16</sup>, Kyle L. Morris <sup>17</sup>, Sameer Velankar <sup>18</sup>, Helen M. Berman <sup>2,19,20</sup>,  
 Stephen K. Burley <sup>2,3,19,21,22</sup>, and Andrej Sali <sup>1\*</sup>

**1 - Research Collaboratory for Structural Bioinformatics Protein Data Bank**, Department of Bioengineering and Therapeutic Sciences, the Quantitative Biosciences Institute (QBI), and the Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94157, USA

**2 - Research Collaboratory for Structural Bioinformatics Protein Data Bank and the Institute for Quantitative Biomedicine**, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

**3 - Rutgers Cancer Institute**, Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA

**4 - Information Sciences Institute**, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA

**5 - European Molecular Biology Laboratory**, Hamburg Unit, c/o Deutsches Elektronen-Synchrotron, Notkestraße 85, 22607 Hamburg, Germany

**6 - Ilse Katz Institute for Nanoscale Science and Technology**, Ben-Gurion University of the Negev, Be'er Sheva 8410501, Israel

**7 - School of Life and Environmental Sciences**, University of Sydney, NSW 2006, Australia

**8 - Department of Chemistry**, University of Utah, Salt Lake City, UT 84112, USA

**9 - Human Technopole**, Via Rita Levi Montalcini 1, 20157 Milano, Italy

**10 - European Molecular Biology Laboratory**, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust, Genome Campus, Hinxton, Cambridge CB10 1SD, UK

**11 - Institute of Molecular Systems Biology**, Department of Biology, ETH Zürich, 8093 Zurich, Switzerland

**12 - Technische Universität Berlin**, Chair of Bioanalytics, 10623 Berlin, Germany

**13 - Si-M<sup>3</sup>Der Simulierte Mensch<sup>3</sup>**, a Science Framework of Technische Universität Berlin and Charité – Universitätsmedizin Berlin, Berlin, Germany

**14 - National Center for Biological Sciences**, Tata Institute of Fundamental Research, Bangalore, Karnataka 560065, India

**15 - Biological Magnetic Resonance Data Bank**, Department of Molecular Biology and Biophysics, University of Connecticut, Farmington, CT 06030-3305, USA

**16 - Protein Data Bank Japan**, Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan

**17 - Electron Microscopy Data Bank**, European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK

**18 - Protein Data Bank in Europe**, European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK

**19 - Department of Chemistry and Chemical Biology**, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

**20 - Department of Quantitative and Computational Biology**, University of Southern California, Los Angeles, CA 90089, USA

**21 - Research Collaboratory for Structural Bioinformatics Protein Data Bank**, San Diego Supercomputer Center, University of California, La Jolla, CA 92093, USA

**22 - Rutgers Artificial Intelligence and Data Science (RAD) Collaboratory**, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

<sup>☆</sup> This article is part of a special issue entitled: 'Computation Resources 2026' published in Journal of Molecular Biology.

**Correspondence to Arthur O. Zalevsky, Brinda Vallat and Andrej Sali:** Research Collaboratory for Structural Bioinformatics Protein Data Bank, Department of Bioengineering and Therapeutic Sciences, the Quantitative Biosciences Institute (QBI), and the Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94157, USA (A.O. Zalevsky, A. Sali). Research Collaboratory for Structural Bioinformatics Protein Data Bank and the Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA (B. Vallat). [arthur.zalevsky@rcsb.org](mailto:arthur.zalevsky@rcsb.org) (A.O. Zalevsky), [brinda.vallat@rcsb.org](mailto:brinda.vallat@rcsb.org) (B. Vallat), [arthur.zalevsky@rcsb.org](mailto:arthur.zalevsky@rcsb.org) (A.O. Zalevsky), [brinda.vallat@rcsb.org](mailto:brinda.vallat@rcsb.org) (B. Vallat), [sali@salilab.org](mailto:sali@salilab.org) (A. Sali) <https://doi.org/10.1016/j.jmb.2025.169598>

**Edited by Michael Sternberg**

## Abstract

PDB-IHM is a branch of the Protein Data Bank (PDB), a Worldwide Protein Data Bank (wwPDB) Core Archive, that expands its scope by allowing for additional biomolecular structure representations and types of experimental information (*i.e.*, integrative/hybrid structure models). As of October 2025, PDB-IHM contained 374 entries, benefitting from multi-scale and multi-state representations and 17 types of experimental data. These structure models are assigned PDB accession codes and are archived alongside other experimental structures in the PDB. Rigorous interpretation of a structure model requires assessment of underlying data quality, consistency with the input data, and estimates of positional uncertainty of its components. Herein, we present the IHMValidation pipeline (<https://validate.pdb-ihm.org>; <https://github.com/salilab/IHMValidation>) based on recommendations from the wwPDB Integrative Methods Task Force plus the small-angle scattering (SAS), chemical crosslinking mass spectrometry (crosslinking-MS), and cryo-electron microscopy and tomography (3DEM) communities. The IHMValidation report (available in both PDF and HTML formats) comprises six sections: (i) overview; (ii) model details; (iii) data quality assessments; (iv) local geometry assessments (*i.e.*, model quality); (v) fit of the model to the data used to generate it; and (vi) fit of the model to the data used for validation. Future expansions of the IHMValidation pipeline will: (i) reflect recommendations coming from additional experimental communities, including Förster resonance energy transfer (FRET) and hydrogen/deuterium exchange MS (HDX-MS); (ii) include other validation criteria, such as Bayesian likelihoods for the data; and (iii) represent estimates of structure model uncertainty based on the variation among alternative models satisfying input data.

© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Introduction

Protein Data Bank (PDB) [1–3] is the single global archive of three-dimensional (3D) structure models of biological macromolecules and their complexes determined by X-ray crystallography (MX) (82% of holdings as of October 2025), three-dimensional cryo-electron microscopy (3DEM, 12%), nuclear magnetic resonance (NMR) spectroscopy (6%), and other techniques (<1%), totalling more than 243,000 entries. The PDB is one of three jointly managed Worldwide Protein Data Bank (wwPDB) [4] Core Archives, the other two being the Electron Microscopy Data Bank (EMDB) [5] and the Biological Magnetic Resonance Data Bank (BMRB) [6]. Recently, the scope of the PDB was expanded by PDB-IHM (PDB Integrative and Hybrid Methods; previously PDB-Dev [7,8]) [9]. PDB-IHM is a branch of the PDB that allows for additional biomolecular structure representations and types of experimental information, supporting deposition, validation, archival, and dissemination of integrative/hybrid struc-

ture models (hereafter abbreviated as “integrative structure models”).

Integrative modeling aims to maximize accuracy, precision, completeness, and explanatory power of a 3D structural model by combining multiple types of input information: experimental data (*e.g.*, NMR data, chemical crosslinks), physical theories (*e.g.*, stereochemistry quantified by a molecular mechanics force field), statistical preferences (*e.g.*, atomic statistical potentials extracted from known structures), and previous structural models (*e.g.*, structural models of complex subunits). For example, an integrative structure model is often computed using a 3DEM map, crosslinks determined by chemical crosslinking mass spectrometry (crosslinking-MS), and stereochemistry restraints. Depending on the resolution of the data and the scale of the modeled system, a modeler may choose to compute a model of any one of several types, including coarse-grained, multi-scaled, multi-state, and ordered-state [10]. To annotate new experimental

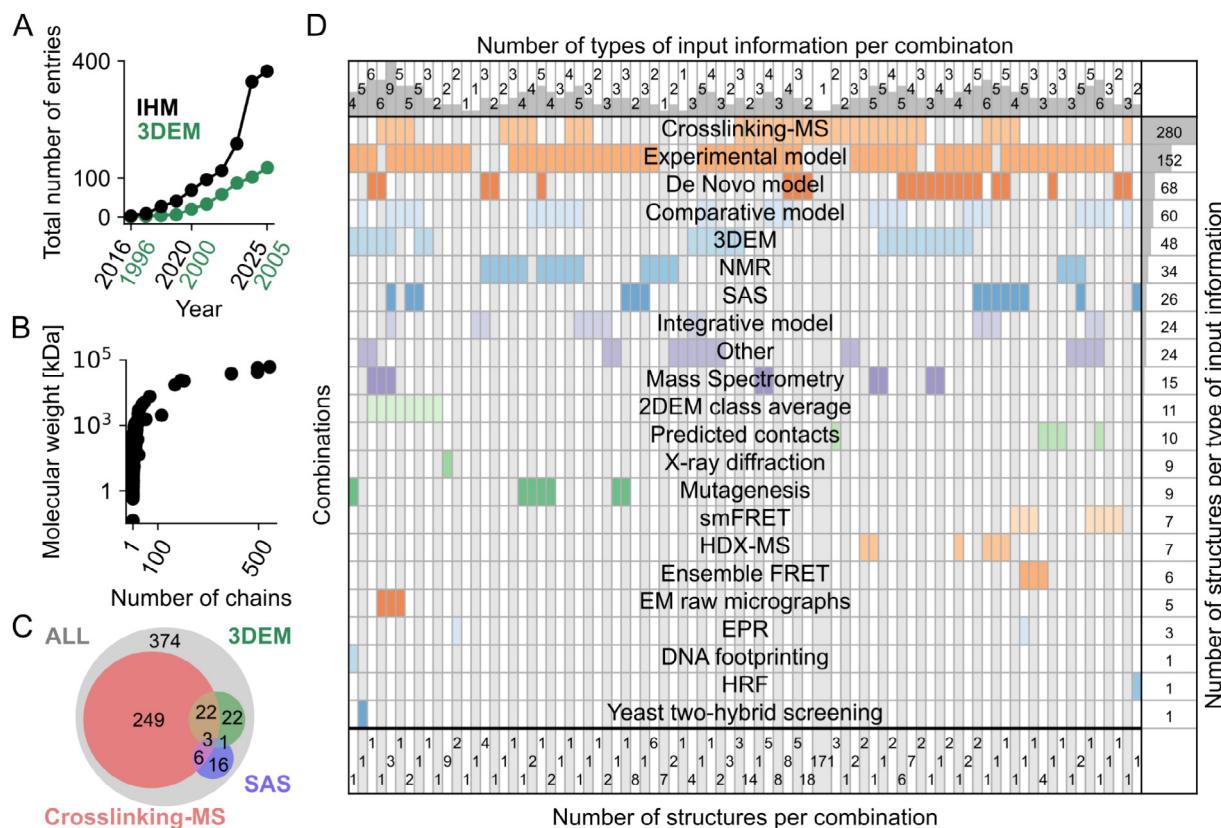
data types and model representations, the wwPDB PDB-IHM team developed the IHM extension of the PDBx/mmCIF dictionary, called IHMCF [11].

As of October 2025, PDB-IHM contained 374 publicly released entries (Figure 1A). Most (~74%) of these entries are multi-chain (homo- and hetero-oligomeric) structures; 56% of the entries have a molecular weight greater than 80 kDa, with an overall range from 5.8 kDa to 3.8 MDa (Figure 1B). This distribution is similar to that for the main PDB archive.

PDB-IHM entries vary in the types of experimental data on which they are based. Unlike the main PDB archive, which primarily consists of structures determined based on the

MX, 3DEM, and NMR data, the current PDB-IHM entries rely on 17 different types of experiments, including various forms of electron microscopy (EM), crosslinking-MS, small-angle scattering (SAS), NMR spectroscopy, mutagenesis, single-molecule Förster resonance energy transfer (smFRET), hydrogen–deuterium exchange MS (HDX-MS), X-ray diffraction, electron paramagnetic resonance (EPR), DNA footprinting, and hydroxyl radical footprinting (HRF). Overall, 83 unique combinations of these input data types were used to compute the current PDB-IHM structure models (Figure 1D).

Integrative modeling generally benefits from previously determined starting structures for



**Figure 1. PDB-IHM contents.** (A) Growth in the number of integrative structure models in PDB-IHM (black filled circles). For comparison, the initial growth in the number of 3DEM structures between 1996 and 2005 is also shown (filled green circles). We excluded the period from 1991 to 1995, because only one 3DEM structure was deposited in that timeframe. (B) Molecular weight of the PDB-IHM structure versus number of polymer chains. (C) Venn diagram of PDB-IHM entries coming from 3DEM, SAS, and crosslinking-MS data; collectively, they represent 85% of all released entries. (D) Experimental methods and structural data sources used to generate PDB-IHM structures are presented in a matrix. Each row corresponds to a source of input information, including crosslinking mass spectrometry (crosslinking-MS), various forms of electron microscopy (EM), nuclear magnetic resonance (NMR) spectroscopy, small-angle scattering (SAS), single-molecule and ensemble Förster resonance energy transfer (smFRET, Ensemble FRET), hydrogen–deuterium exchange MS (HDX-MS), electron paramagnetic resonance (EPR), and hydroxyl radical footprinting (HRF). The number of entries with a given type of input information is plotted on the right. Each column represents a particular combination of input information. Counts in the top section indicate the number of sources of input information in the column; at the bottom, the number of entries with a given combination of sources of input information.

individual components of the modeled system [10,12,13]. Half of the PDB-IHM structure models were generated using at least one starting structure; conformations of these structures can be fixed or flexible during integrative modeling. Starting structures can be determined by any structural biology method, including experimental methods (e.g., X-ray crystallography, NMR spectroscopy, 3DEM) and computational methods (e.g., comparative modeling, *de novo* structure prediction [14]). Conveniently, the atomic coordinates of these starting structures can often be downloaded from the PDB [2], PDB-IHM [9], Model Archive [15], or other public domain repositories. A substantial portion of PDB-IHM entries (43%) contain at least one polymer component extracted from an experimental structure archived in PDB (~41%) or a PDB-IHM integrative structure model (~6%). Similarly, ~29% of the PDB-IHM entries also contain at least one component polymer structure computed by comparative modeling (~16%), using software tools such as MODELER [16], SwissModel [17], and Rosetta [18], or *de novo* structure prediction (~18%), using programs such as RoseTTAFold [19] and AlphaFold2/3 [20,21]. We anticipate a rapid increase in the use of *de novo* models due to their availability, coverage, and accuracy. This trend has already been observed with recent PDB-IHM depositions (74% of entries released in 2025).

Rigorous interpretation of a structure requires assessment of underlying data quality, consistency with the input data, and estimates of positional uncertainty of the generated Cartesian coordinates. Herein, we present the IHMValidation pipeline based on recommendations from the wwPDB Integrative/Hybrid Methods Task Force (IHMtf) [22,23] plus the SAS [24,25], crosslinking-MS [26,27], and 3DEM [28] communities. We prioritized implementing support for SAS, crosslinking-MS, and 3DEM data because they accounted for ~85% of the PDB-IHM entries (Figure 1C).

## Results and Discussion

### Validation report

Inspired by previous efforts on structural validation [24,28–30], we organized information about each structure model, including experimental and structural data used for structure model generation, model representation, details of the modeling protocol, and validation metrics, into two key documents: the full validation report and a summary table (Supplementary Files 1 and 2). The full validation report includes extended statistics and detailed plots. The summary table is a concise description of the full validation report, without graphical displays. The IHMValidation pipeline generates human-readable reports in both PDF and HTML formats.

The validation report is organized into six sections: (i) overview; (ii) model details; (iii) data

quality assessments; (iv) local geometry assessments (*i.e.*, model quality); (v) fit of the model to the data used to generate it; and (vi) fit of the model to the data used for validation. Each section contains validation metrics presented as tables with numerical values and plots. Selection of metrics depends on the type of structural (atomic, coarse-grained, multi-scale) and the type of experimental data used to generate the integrative structure model, following recommendations from the IHMtf [22,23] and data-generating communities [24–28]. Next, we describe each section.

### Overview

This section provides an “executive” summary of the structure content and key quality indicators. The summary indicates serious problems with the structure, if any (Figure 2). The section contains plots of key metrics, divided into four categories, one for each validation criterion: (i) data quality assessments (Figure 2A, B); (ii) local geometry assessments (*i.e.*, model quality); (iii) fit of the model to the data used to generate it (Figure 2B, C) and (iv) fit of the model to the data used for validation.

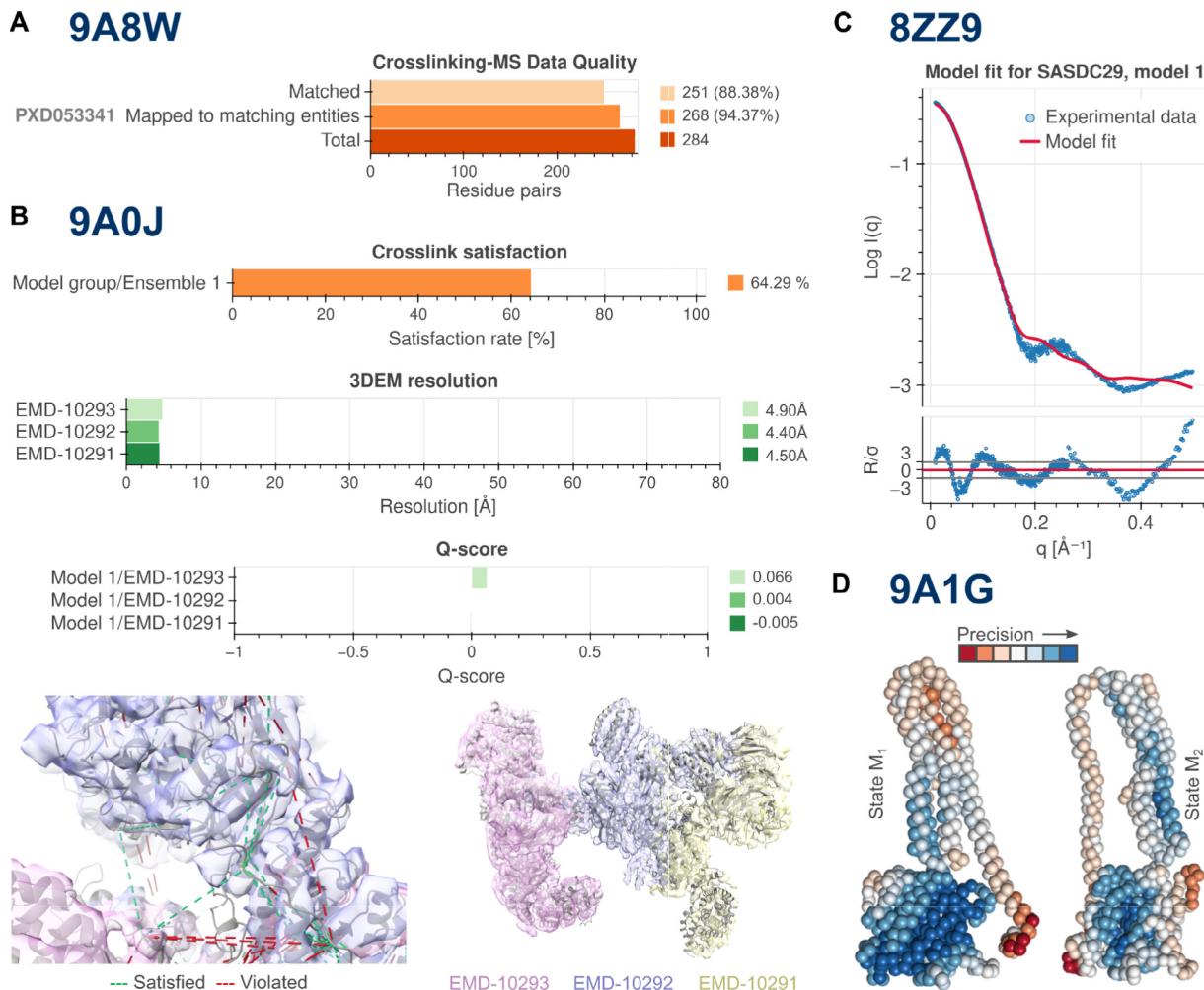
### Model details

This section contains tables with detailed information about (i) the entities present in the structure (e.g., biopolymers and ligands) and their representation (*i.e.*, atomic, coarse-grained, multi-scale, and multi-state); (ii) a description and links to datasets for computing the structure; and (iii) a summary of the modeling methodology (Supplementary File 1).

### Quality of data and fit of integrative structure models to data

Annotations can help depositors, biocurators, and data consumers identify problems with the underlying data, including artifacts resulting from experiments and data conversions into standardized formats.

With the data quality assessment in hand, we proceed to assess the integrative model against the data. IHMValidation distinguishes between two types of assessments. For both types, we use the same criteria to quantify the model-data match. First, the model is assessed against the data used to generate it. Ideally, the model should satisfy all of the input data within the specified data precision. Second, a model can be assessed with respect to holdout data (*i.e.*, data not used to compute the model), whether reported in the current study or previously published. Large deviation(s) of a model from holdout data can reveal structure overfitting [31,32], particularly when



**Figure 2. Typical issues highlighted by the IHMValidation reports.** **(A)** The number of crosslinks used for modeling versus those present in the original dataset. (N.B.: Crosslinking-MS experiments can identify thousands of proteins, of which only a fraction are used for generating a particular PDB-IHM entry). For 9A8W, 268 of 284 crosslinks reported in the PRIDE dataset (PXD053341) were identified between proteins of interest, and only 251 of them were used for integrative modeling due to truncations of flexible regions [67]. **(B)** Fit of structural model to crosslinking-MS and 3DEM data. Integrative structure models are often computed using low-to-medium resolution data (e.g., crosslinking-MS and mid-resolution cryo-EM maps). (N.B.: Lower resolution 3DEM maps ( $\sim 5$  Å) typically exhibit average per residue Q-scores of  $\sim 0.3$  [68]). Integrative modeling of 9A0J [69] utilized three different 3DEM maps (EMDB IDs: EMD-10291, EMD-10292, EMD-10293) to generate a more complete model that covers as many residues as possible. Spatial restraints derived from crosslinking-MS data provided complementary information about the relative positions of subunits across the maps. **(C)** Fit of structure to SAS data. At first glance, the agreement between the structure 8ZZ9 and the SAS data (SASBDB ID SASDC29) is unacceptably low ( $\chi^2 \sim 25$ ). However, the authors rationalized the differences in the high-q region by the intrinsic dynamics of the protein, which was confirmed using NMR spectroscopy [70]. **(D)** Regions of low-(red) and high-(blue) local precision in the two collections of multi-state models of human guanylate-binding protein 1 (hGBP-1, 9A1G), annotated and visualized using PrISM. Two deposited collections of models (100 and 106 individual models in states  $M_1$  and  $M_2$ , respectively) with substantially different representative structures for each state and noticeable differences in regions of high and low precision are shown [71].

it is generated from a sparse dataset. Currently, we report assessments based on data used to generate the integrative structure model, while holdout assessment is planned for the next IHMValidation release.

#### Quality of SAS data and fit of integrative structure models to data

Small-angle scattering (SAS) of X-rays (SAXS) or neutrons (SANS) by macromolecules in solution enables the characterization of their size and

shape. In addition, as the SAS measurement yields the time and ensemble average of the molecules present in the solution, it provides information on structural heterogeneity relating to conformational or oligomeric state(s) present in the sample [33–36]. One advantage of SAS experiments is that the characterized system is in solution. While *in situ* measurements are sometimes feasible, most studies involve samples purified to high purity. SAS data can be useful for integrative modeling, especially for large and/or flexible macromolecules that may be difficult to study using other structural techniques alone. For example, the atom-pair distance distribution function (PDDF) computed from SAS data can be used as a spatial restraint for integrative structure modeling and/or to validate the resulting models [37–39].

The SAS portion of an IHMValidation report adheres to the wwPDB SAS Validation Task Force (SASvtf) recommendations [24]. PDB-IHM depositors are required to submit their SAS data to SASBDB [40], which serves as the source of SAS data and quality metrics in sasCIF format [41] as recommended by the SASvtf [24].

For data quality assessment the scattering profile is presented with data quality metrics, including Guinier analysis for the evaluation of the radius of gyration,  $R_g$ , PDDF analysis for  $R_g$  and maximum particle dimension estimates ( $D_{max}$ ) as well as associated plots, e.g., Porod-Debye and Kratky plots that provide additional qualitative information relating to particle compactness, flatness, flexibility or extensibility (Supplementary Table 1A). PDB-IHM model metrics based on SAS data (Supplementary Table 1B) include plots of SAS profiles calculated for structures and their goodness-of-fit to experimental data in the form of the reduced  $\chi^2$  with its associated error-weighted residual difference plot between model and experiment (Figure 2C), and  $p$ -values from the Correlation Map (CorMap) test.

### Quality of crosslinking-MS data and fit of integrative structure models to data

Crosslinking-MS can identify pairs of spatially proximal residues in a protein or protein complex [42]. After the sample is exposed to a chemical crosslinker, it is proteolytically digested (usually with trypsin), followed by detecting crosslinked pairs of residues by MS. These pairs can be converted into spatial restraints for integrative modeling and/or for validation of the resulting models [43]. The development of crosslinking reagents of varying length and chemical reactivity enables the study of compositionally and structurally homogeneous and heterogeneous protein complexes *in vitro* and *in situ* [44,45]. Moreover, the approach can also be applied to nucleic acids and their complexes with proteins [46].

The crosslinking-MS community is in the process of establishing its data standards [26,27], including protocols for validating crosslinking-MS data and structures. We propose here a set of metrics and visualization approaches to facilitate further discussion about community standards for validating this type of data, in coordination with members of the crosslinking-MS [27] community and the PRIDE database for proteomics data, including crosslinking-MS [47].

In the IHMValidation report, we first note the numbers of reported and utilized crosslinks. Crosslinking-MS data are retrieved from the PRIDE database via the Crosslinking-API or directly from a data file in the open standard mzIdentML format (version 1.2 or 1.3) [48]. We then compare proteins reported in the <DBSequence> element in mzIdentML with those reported in the IHMCIF file and match the corresponding crosslinks. Percentages of crosslinks present in the IHMCIF file with respect to the crosslinking-MS data in the mzIdentML file are reported. Low percentage values indicate a potential problem with the omission of data, although sometimes only a fraction of crosslinks from an *in situ* or *in vivo* dataset is relevant for the current molecular system.

There is currently no generally accepted approach for performing a goodness-of-fit assessment between a 3D structural model and crosslinking-MS data. Widely used metrics/strategies include: (i) Euclidean or surface-based distances [49]; (ii) mapping of crosslinks onto cross-linked functional groups or  $C\alpha$  atoms [50]; (iii) cross-link satisfaction by a single conformation or a collection of conformations [51]; and (iv) continuous probabilistic scores as opposed to sharp cutoffs [52].

IHMValidation compares the structural model against the crosslink restraints used for integrative modeling as submitted by depositors. We provide a summary table listing restraints and corresponding user-provided thresholds. Next, we assess restraint satisfaction against the thresholds using Euclidean distances. The IHMValidation report attempts to account for modeling complexity during the assessment. For example, a single crosslink can be represented as a combination of lower and upper-bound distance restraints, or as a set of restraints between multiple instances of polymer entities, reflecting ambiguities in the crosslinking-MS dataset (e.g., an intra-molecular crosslink vs. an inter-molecular homodimer crosslink).

Data quality metrics based on crosslinking-MS data include the total number of crosslinks in the dataset, the number of crosslinks matched to the structure, and the number of crosslinks matched to restraints (Supplementary Table 2A, Figure 2A). Integrative structure model metrics based on crosslinking-MS data include restraint types,

distograms (i.e., histogram plots of distances), and satisfaction rates ([Supplementary Table 2B](#), [Figure 2B](#)).

### Quality of 3DEM data and fit of integrative structure models to data

3DEM, typically cryo-electron microscopy (cryoEM) and cryo-electron tomography (cryoET), is a powerful technique for studying 3D structures of biomolecules at low-to-medium-to-high resolution. CryoEM involves rapidly freezing the sample in a thin layer of vitrified ice and then imaging it using a transmission electron microscope [53]. CryoET allows studies of macromolecular structure and organization within cells or tissues [54]. Both techniques can provide unprecedented insights into the structure and function of biological systems, particularly for megadalton-scale biomolecular machines [55].

Following recommendations developed by the wwPDB 3DEM Validation Task Force [56] and community guidelines [28,57] for validation of structures generated using cryoEM and cryoET, we provide a set of numerical metrics and plots obtained from either EMDB [5] or calculated with the Validation Analysis (va) software [58] ([Figure 2B](#)). Assessment of the fit of a 3D structural model to 3DEM map data is currently available only for atomic structures.

The data quality criteria for 3DEM data include map visualization, map analysis, and Fourier-Shell Correlation (FSC) validation metrics ([Supplementary Table 3A](#)). Structural model criteria based on 3DEM data include map-model fit, Q-score, and atom inclusion ([Supplementary Table 3B](#), [Figure 2B](#)).

## Local Geometry Assessment (i.e., Integrative Structure Model Quality)

### Stereochemistry

Analysis of structural model stereochemistry is well established for 3D biostructures coming from X-ray, NMR, or 3DEM [59]. This analysis highlights potential inaccuracies by identifying statistically significant deviations in stereochemical features, such as bond lengths and angles, from reference values derived from theory or from representative high-quality residues in high-resolution protein and nucleic acid structures (e.g., X-ray protein structures with resolution limits better than 2.0 Å) [60]. Care is needed, however, in interpreting outliers as either actual errors or accurate depictions that deviate from the norm [59]. While the stereochemistry of atomic integrative structure models can be assessed using existing tools, such as MolProbity [60], many integrative structure models are coarse-grained or multi-scale and cannot therefore be evaluated with existing tools.

We currently assess the stereochemistry of integrative structure models as follows. For components represented at atomic resolution, we follow the procedure adopted long ago by the wwPDB [61], relying on MolProbity [60]. In particular, we identify deviations of a model from accepted values for bond lengths, bond angles, dihedral angles, improper dihedral angles, and atomic overlaps. For coarse-grained and multi-scale components, we highlight overlaps between pairs of coarse-grained beads or between a coarse-grained bead and an atom. This test is analogous to estimating interatomic clashes within atomic models. A total excluded volume violation of a model is defined as the percentage of overlapping pairs of beads; a pair of beads is overlapping when the center-to-center distance between the beads is smaller than the sum of their radii, as specified by the depositor.

### Local model precision

If a structure is represented as a collection of conformations, we can define local model precision (uncertainty) as the corresponding local variability among the deposited models. This variability is quantified by the Precision for Integrative Structural Models (PrISM) program [62], which efficiently identifies and visualizes regions of high- and low-precision in an integrative structural model. While the depositor can provide already superposed models, superposition can also be performed by the validation pipeline. PrISM first computes root-mean-square fluctuations (RMSFs; spreads) of individual particles in the model collection, then partitions the model into volumes at several relative levels of precision using spatial clustering of particles with similar precisions. The validation report includes a visualization of high- and low-precision regions provided by PrISM ([Figure 2D](#)).

## Summary Table

To ensure open access to all information required to reproduce and evaluate integrative structure models, we have designed a validation summary table for integrative modeling, with input from the community. This table summarizes the input information, model representation, sampling protocol, software tools, and validation output in the human-readable PDF format ([Supplementary File 2](#)). The table is inspired by X-ray [29], NMR [30], EM [56], and SAS [24] validation summary tables. A validation summary table for every entry in the PDB-IHM is generated during the deposition. Depositors are encouraged to include these tables in their respective publications to [supplement information](#) available to editors, reviewers, and readers. A recently published integrative structure model (9A05) supplemented with such a table provides a proof-of-concept [63].

## Implementation

The IHMValidation pipeline is implemented as a Python package. To simplify dependency management for the third-party validation software (e.g., MolProbity [60], ATSAS [64], mzidentml-reader, va [58], and PrISM [62]) and deployment, the IHMValidation pipeline is executed in an App-tainer (previously Singularity [65]) virtual container. The IHMValidation standalone validation server has been built using the open source DERIVA [66] scientific asset management platform.

## Future Directions

Future expansions of the IHMValidation pipeline will (i) reflect additional recommendations, including those from the Förster resonance energy transfer (FRET) and hydrogen/deuterium exchange MS (HDX MS) communities; (ii) include additional validation criteria, such as Bayesian likelihoods for the data given the structure; and (iii) represent estimates of structure uncertainty based on the variability among alternative models satisfying input information. We also welcome community contributions in the form of new methods and tools for the assessment of integrative structure models. We are committed to incorporating standards and tools for assessing integrative structure models and data as soon as corresponding communities develop them.

## Funding information

RCSB PDB core operations are jointly funded by the National Science Foundation (DBI-2321666, PI: S.K. Burley), the US Department of Energy (DE-SC0019749, PI: S.K. Burley), and the National Cancer Institute, the National Institute of Allergy and Infectious Diseases, and the National Institute of General Medical Sciences of the National Institutes of Health (R01GM157729, PI: S.K. Burley). B. Vallat acknowledges funding from the United States National Science Foundation award DBI-2112966 (PI: B. Vallat). A. Sali acknowledges funding from the National Science Foundation and the National Institutes of Health (NSF DBI-2112967, PI: A. Sali; NIH R01GM083960, PI: A. Sali; NIH P41GM109824, PI: M.P. Rout). B.M. Webb is a Nebius Fellow. J.A.V. acknowledges funding from BBSRC (BB/V018779/1, PI: J.A. Vizcaino), EPSRC (EP/Y035984/1, PI: J.A. Vizcaino), and EMBL core funding. S.V. acknowledges funding from DAE, Govt of India (RTI 4006, PI: S. Viswanath).

## CRediT authorship contribution statement

**Arthur O. Zalevsky:** Writing – original draft, Visualization, Software, Methodology,

Conceptualization. **Brinda Vallat:** Writing – review & editing, Supervision, Software, Project administration, Methodology, Funding acquisition, Conceptualization. **Benjamin M. Webb:** Writing – review & editing, Software, Methodology. **Hongsuda Tangmunarunkit:** Software, Methodology. **Monica R. Sekharan:** Writing – review & editing, Data curation. **Aref ShafaeiBejestan:** Software. **Sai Ganesan:** Software, Methodology. **Jared Sagendorf:** Writing – review & editing, Methodology. **Cy M. Jeffries:** Writing – review & editing, Methodology. **Jill Trewella:** Writing – review & editing, Methodology. **Andrea Graziadei:** Writing – review & editing, Methodology. **Juan Antonio Vizcaíno:** Writing – review & editing, Methodology. **Alexander Leitner:** Writing – review & editing, Methodology. **Juri Rappaport:** Writing – review & editing, Methodology. **Ezra Peisach:** Writing – review & editing, Methodology. **Justin W. Flatt:** Writing – review & editing, Methodology. **Jasmine Y. Young:** Writing – review & editing, Methodology. **Kartik Majila:** Software. **Shruthi Viswanath:** Writing – review & editing, Methodology. **Carl Kesselman:** Writing – review & editing, Methodology. **Jeffrey C. Hoch:** Writing – review & editing, Methodology. **Genji Kurisu:** Writing – review & editing, Methodology. **Kyle L. Morris:** Writing – review & editing, Methodology. **Sameer Velankar:** Writing – review & editing, Methodology. **Helen M. Berman:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Stephen K. Burley:** Writing – review & editing, Supervision, Funding acquisition. **Andrej Sali:** Writing – original draft, Supervision, Funding acquisition, Conceptualization.

## DATA AVAILABILITY

Summary tables and validation reports are available in PDF and HTML formats on the PDB IHM website (<https://pdb-ihm.org>) and in the PDB archive (PDF only). The validation pipeline is available for standalone use at <https://validate.pdb-ihm.org>. The source code can be downloaded from the <https://github.com/salilab/IHMValidation> repository. The software has also been deployed within the PDB-IHM deposition system, enabling depositors to access validation reports after curation is complete.

## DECLARATION OF COMPETING INTEREST

We have no conflicts of interest to declare.

## Acknowledgments

We thank the members of the wwPDB Integrative/Hybrid Methods Task Force (Alexandre Bonvin, Frank DiMaio, Gerhard Hummer, Jens Meiler,

Torsten Schwede, and Emad Tajkhorshid) for their valuable recommendations and feedback on the development of PDB-IHM. We also thank the many PDB-IHM depositors for sharing their results. Finally, we are grateful to the developers of ChimeraX (Thomas Ferrin and Thomas Goddard) and Mol\* (Alexander Rose and David Sehnal) for developing the software for visualizing integrative structure models.

## Appendix A. Supplementary material

Supplementary material to this article can be found online at <https://doi.org/10.1016/j.jmb.2025.169598>.

Received 16 October 2025;  
Accepted 13 December 2025;  
Available online xxxx

### Keywords:

integrative modeling;  
Protein Data Bank (PDB);  
Small-angle X-ray scattering (SAXS);  
crosslinking mass spectrometry (crosslinking-MS);  
three-dimensional cryo-electron microscopy (3DEM)

## References

- [1]. Crystallography: Protein Data Bank, *Nature New Biol.* **233** (1971) 223. <https://doi.org/10.1038/newbio233223b0>.
- [2]. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242. <https://doi.org/10.1093/nar/28.1.235>.
- [3]. wwPDB consortium, (2019). Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **47**, D520–D528. <https://doi.org/10.1093/nar/gky949>.
- [4]. Berman, H., Henrick, K., Nakamura, H., (2003). Announcing the worldwide Protein Data Bank. *Nature Struct. Biol.* **10**, 980. <https://doi.org/10.1038/nsb1203-980>.
- [5]. wwPDB consortium, (2024). EMDB-the Electron Microscopy Data Bank. *Nucleic Acids Res.* **52**, D456–D465. <https://doi.org/10.1093/nar/gkad1019>.
- [6]. Hoch, J.C., Baskaran, K., Burr, H., Chin, J., Eghbalnia, H. R., Fujiwara, T., Gryk, M.R., Iwata, T., Kojima, C., Kurisu, G., Maziuk, D., Miyanoiri, Y., Wedell, J.R., Wilburn, C., Yao, H., Yokochi, M., (2023). Biological Magnetic Resonance Data Bank. *Nucleic Acids Res.* **51**, D368–D376. <https://doi.org/10.1093/nar/gkac1050>.
- [7]. Burley, S.K., Kurisu, G., Markley, J.L., Nakamura, H., Velankar, S., Berman, H.M., Sali, A., Schwede, T., Trewella, J., (2017). PDB-Dev: a prototype system for depositing integrative/hybrid structural models. *Structure* **25**, 1317–1318. <https://doi.org/10.1016/j.str.2017.08.001>.
- [8]. Vallat, B., Webb, B., Westbrook, J.D., Sali, A., Berman, H. M., (2018). Development of a prototype system for archiving integrative/hybrid structure models of biological macromolecules. *Structure* **26**, 894–904.e2. <https://doi.org/10.1016/j.str.2018.03.011>.
- [9]. Vallat, B., Webb, B.M., Zalevsky, A., Tangmunarunkit, H., Sekharan, M.R., Voinea, S., ShafaeiBejestan, A., Sagendorf, J., Hoch, J.C., Kurisu, G., Morris, K.L., Velankar, S., Kesselman, C., Burley, S.K., Berman, H. M., Sali, A., (2025). PDB-IHM: a system for deposition, curation, validation, and dissemination of integrative structures. *J. Mol. Biol.* **437**, 168963. <https://doi.org/10.1016/j.jmb.2025.168963>.
- [10]. Sali, A., (2021). From integrative structural biology to cell biology. *J. Biol. Chem.* **296**, 100743. <https://doi.org/10.1016/j.jbc.2021.100743>.
- [11]. Vallat, B., Webb, B.M., Westbrook, J.D., Goddard, T.D., Hanke, C.A., Graziadei, A., Peisach, E., Zalevsky, A., Sagendorf, J., Tangmunarunkit, H., Voinea, S., Sekharan, M., Yu, J., Bonvin, A.A.M.J.J., DiMaio, F., Hummer, G., Meiler, J., Tajkhorshid, E., Ferrin, T.E., Lawson, C.L., Leitner, A., Rappaport, J., Seidel, C.A.M., Jeffries, C.M., Burley, S.K., Hoch, J.C., Kurisu, G., Morris, K., Patwardhan, A., Velankar, S., Schwede, T., Trewella, J., Kesselman, C., Berman, H.M., Sali, A., (2024). IHMCF: an extension of the PDBx/mmCIF data standard for integrative structure determination methods. *J. Mol. Biol.* **436**, 168546. <https://doi.org/10.1016/j.jmb.2024.168546>.
- [12]. Rodrigues, J.P.G.L.M., Bonvin, A.M.J.J., (2014). Integrative computational modeling of protein interactions. *FEBS J.* **281**, 1988–2003. <https://doi.org/10.1111/febs.12771>.
- [13]. Mosalaganti, S., Obarska-Kosinska, A., Siggel, M., Taniguchi, R., Turoňová, B., Zimmerli, C.E., Buczak, K., Schmidt, F.H., Margiotta, E., Mackmull, M.-T., Hagen, W. J.H., Hummer, G., Kosinski, J., Beck, M., (2022). AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science* **376**, eabm9506. <https://doi.org/10.1126/science.abm9506>.
- [14]. Baker, D., Sali, A., (2001). Protein structure prediction and structural genomics. *Science* **294**, 93–96. <https://doi.org/10.1126/science.1065659>.
- [15]. Tauriello, G., Waterhouse, A.M., Haas, J., Behringer, D., Bienert, S., Garello, T., Schwede, T., (2025). ModelArchive: a deposition database for computational macromolecular structural models. *J. Mol. Biol.* **437**, 168996. <https://doi.org/10.1016/j.jmb.2025.168996>.
- [16]. Webb, B., Sali, A., (2016). Comparative protein structure modeling using MODELLER 5.6.1–5.6.37. *Curr. Protoc. Bioinform.* **54** <https://doi.org/10.1002/cpbi.3>.
- [17]. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303. <https://doi.org/10.1093/nar/gky427>.
- [18]. Leman, J.K., Weitzner, B.D., Lewis, S.M., Adolf-Bryfogle, J., Alam, N., Alford, R.F., Aprahamian, M., Baker, D., Barlow, K.A., Barth, P., Basanta, B., Bender, B.J., Blacklock, K., Bonet, J., Boyken, S.E., Bradley, P., Bystroff, C., Conway, P., Cooper, S., Correia, B.E., Coventry, B., Das, R., De Jong, R.M., DiMaio, F., Dsilva, L., Dunbrack, R., Ford, A.S., Frenz, B., Fu, D.Y., Geniesse, C., Goldschmidt, L., Gowthaman, R., Gray, J. J., Gront, D., Guffy, S., Horowitz, S., Huang, P.-S., Huber, T., Jacobs, T.M., Jeliazkov, J.R., Johnson, D.K., Kappel, K., Karanicolas, J., Khakzad, H., Khar, K.R., Khare, S.D.,

Khatib, F., Khramushin, A., King, I.C., Kleffner, R., Koepnick, B., Kortemme, T., Kuenze, G., Kuhlman, B., Kuroda, D., Labonte, J.W., Lai, J.K., Lapidoth, G., Leaver-Fay, A., Lindert, S., Linsky, T., London, N., Lubin, J.H., Lyskov, S., Maguire, J., Malmström, L., Marcos, E., Marcu, O., Marze, N.A., Meiler, J., Moretti, R., Mulligan, V.K., Nerli, S., Norn, C., Ó'Conchúir, S., Ollikainen, N., Ovchinnikov, S., Pacella, M.S., Pan, X., Park, H., Pavlovic, R.E., Pethe, M., Pierce, B.G., Pilla, K.B., Raveh, B., Renfrew, P.D., Burman, S.S.R., Rubenstein, A., Sauer, M.F., Scheck, A., Schief, W., Schueler-Furman, O., Sedan, Y., Sevy, A.M., Sgourakis, N.G., Shi, L., Siegel, J.B., Silva, D.-A., Smith, S., Song, Y., Stein, A., Szegedy, M., Teets, F.D., Thyme, S.B., Wang, R.-Y.-R., Watkins, A., Zimmerman, L., Bonneau, R., (2020). Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nature Methods* **17**, 665–680. <https://doi.org/10.1038/s41592-020-0848-2>.

[19]. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R.D., Millán, C., Park, H., Adams, C., Glassman, C.R., DeGiovanni, A., Pereira, J.H., Rodrigues, A.V., van Dijk, A.A., Ebrecht, A.C., Opperman, D.J., Sagmeister, T., Buhlmann, C., Pavkov-Keller, T., Rathinaswamy, M.K., Dalwadi, U., Yip, C.K., Burke, J.E., Garcia, K.C., Grishin, N.V., Adams, P.D., Read, R.J., Baker, D., (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876. <https://doi.org/10.1126/science.abb8754>.

[20]. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.

[21]. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., Bodenstein, S.W., Evans, D.A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A.I., Cowie, A., Figurnov, M., Fuchs, F.B., Gladman, H., Jain, R., Khan, Y.A., Low, C.M.R., Perlin, K., Potapenko, A., Savv, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E.D., Zielinski, M., Žídek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., Jumper, J.M., (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500. <https://doi.org/10.1038/s41586-024-07487-w>.

[22]. Sali, A., Berman, H.M., Schwede, T., Trehella, J., Kleywegt, G., Burley, S.K., Markley, J., Nakamura, H., Adams, P., Bonvin, A.M.J.J., 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., Schröder, G.F., Schwiers, C.D., Seidel, C.A.M., Svergun, D., Topf, M., Ulrich, E.L., Velankar, S., Westbrook, J.D., (2015). Outcome of the first wwPDB hybrid/integrative methods task force workshop. *Structure* **23**, 1156–1167. <https://doi.org/10.1016/j.str.2015.05.013>.

[23]. 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 Jr, G.N., Prisner, T., Rappsilber, J., Schriemer, D. C., Schwede, T., Seidel, C.A.M., Strutzenberg, T.S., Svergun, D.I., Tajkhorshid, E., Trehella, J., Vallat, B., Velankar, S., Vuister, G.W., Webb, B., Westbrook, J.D., White, K.L., Sali, A., (2019). Federating structural models and data: outcomes from a workshop on archiving integrative structures. *Structure* **27**, 1745–1759. <https://doi.org/10.1016/j.str.2019.11.002>.

[24]. Trehella, J., Duff, A.P., Durand, D., Gabel, F., Guss, J. M., Hendrickson, W.A., Hura, G.L., Jacques, D.A., Kirby, N.M., Kwan, A.H., Pérez, J., Pollack, L., Ryan, T.M., Sali, A., Schneidman-Duhovny, D., Schwede, T., Svergun, D. I., Sugiyama, M., Tainer, J.A., Vachette, P., Westbrook, J., Whitten, A.E., (2017). 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. *Acta Crystallogr. D Struct. Biol.* **73**, 710–728. <https://doi.org/10.1107/S2059798317011597>.

[25]. Trehella, J., Jeffries, C.M., Whitten, A.E., (2023). 2023 update of template tables for reporting biomolecular structural modelling of small-angle scattering data. *Acta Crystallogr. D Struct. Biol.* **79**, 122–132. <https://doi.org/10.1107/S2059798322012141>.

[26]. Leitner, A., Bonvin, A.M.J.J., Borchers, C.H., Chalkley, R. J., Chamot-Rooke, J., Combe, C.W., Cox, J., Dong, M.-Q., Fischer, L., Götze, M., Gozzo, F.C., Heck, A.J.R., Hoopmann, M.R., Huang, L., Ishihama, Y., Jones, A.R., Kalisman, N., Kohlbacher, O., Mechtr, K., Moritz, R.L., Netz, E., Novak, P., Petrochenko, E., Sali, A., Scheltema, R.A., Schmidt, C., Schriemer, D., Sinz, A., Sobott, F., Stengel, F., Thalassinos, K., Urlaub, H., Viner, R., Vizcaíno, J.A., Wilkins, M.R., Rappsilber, J., (2020). Toward increased reliability, transparency, and accessibility in cross-linking mass spectrometry. *Structure* **28**, 1259–1268. <https://doi.org/10.1016/j.str.2020.09.011>.

[27]. Rappsilber, J., Bruce, J., Combe, C., Fried, S.D., Graziadei, A., Heck, A.J.R., Iacobucci, C., Leitner, A., Mechtr, K., Novak, P., O'Reilly, F., Schriemer, D.C., Sinz, A., Stengel, F., Thalassinos, K., (2025). A roadmap for improving reliability and data sharing in crosslinking mass spectrometry. *Mol. Cell. Proteomics* **24**, 101024. <https://doi.org/10.1016/j.mcpro.2025.101024>.

[28]. Kleywegt, G.J., Adams, P.D., Butcher, S.J., Lawson, C.L., Rohou, A., Rosenthal, P.B., Subramaniam, S., Topf, M., Abbott, S., Baldwin, P.R., Berrisford, J.M., Bricogne, G., Choudhary, P., Croll, T.I., Danev, R., Ganesan, S.J., Grant, T., Gutmanas, A., Henderson, R., Heymann, J.B., Huiskonen, J.T., Istrate, A., Kato, T., Lander, G.C., Lok, S.M., Ludtke, S.J., Murshudov, G.N., Pye, R., Pintilie, G. D., Richardson, J.S., Sachse, C., Salih, O., Scheres, S.H. W., Schroeder, G.F., Sorzano, C.O.S., Stagg, S.M., Wang, Z., Warshamanager, R., Westbrook, J.D., Winn,

M.D., Young, J.Y., Burley, S.K., Hoch, J.C., Kurisu, G., Morris, K., Patwardhan, A., Velankar, S., (2024). Community recommendations on cryoEM data archiving and validation. *IUCrJ* **11**, 140–151. <https://doi.org/10.1107/S205225254001246>.

[29]. Read, R.J., Adams, P.D., Arendall 3rd, W.B., Brunger, A. T., Emsley, P., Joosten, R.P., Kleywegt, G.J., Krissinel, E. B., Lütteke, T., Otwinowski, Z., Perrakis, A., Richardson, J.S., Sheffler, W.H., Smith, J.L., Tickle, I.J., Vriend, G., Zwart, P.H., (2011). A new generation of crystallographic validation tools for the protein data bank. *Structure* **19**, 1395–1412. <https://doi.org/10.1016/j.str.2011.08.006>.

[30]. Montelione, G.T., Nilges, M., Bax, A., Güntert, 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). Recommendations of the wwPDB NMR validation task force. *Structure* **21**, 1563–1570. <https://doi.org/10.1016/j.str.2013.07.021>.

[31]. Brünger, A.T., (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475. <https://doi.org/10.1038/355472a0>.

[32]. Falkner, B., Schröder, G.F., (2013). Cross-validation in cryo-EM-based structural modeling. *PNAS* **110**, 8930–8935. <https://doi.org/10.1073/pnas.1119041110>.

[33]. Chaudhuri, B., Muñoz, I.G., Qian, S., Urban, V.S., (2017). Biological Small Angle Scattering: Techniques, Strategies and Tips. Springer <https://doi.org/10.1007/978-981-10-6038-0>.

[34]. Lattman, E.E., Grant, T.D., Snell, E.H., (2018). Biological Small Angle Scattering: Theory and Practice. Oxford University Press <https://doi.org/10.1093/oso/9780199670871.001.0001>.

[35]. Svergun, D.I., Koch, M.H.J., Timmins, P.A., May, R.P., (2013). Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules. Oxford University Press <https://doi.org/10.1093/acprof:oso/9780199639533.001.0001>.

[36]. Trehewella, J., (2022). Recent advances in small-angle scattering and its expanding impact in structural biology. *Structure* **30**, 15–23. <https://doi.org/10.1016/j.str.2021.09.008>.

[37]. Schneidman-Duhovny, D., Kim, S.J., Sali, A., (2012). Integrative structural modeling with small angle X-ray scattering profiles. *BMC Struct. Biol.* **12**, 17. <https://doi.org/10.1186/1472-6807-12-17>.

[38]. Bernetti, M., Hall, K.B., Bussi, G., (2021). Reweighting of molecular simulations with explicit-solvent SAXS restraints elucidates ion-dependent RNA ensembles. *Nucleic Acids Res.* **49**, e84.

[39]. Grishaev, A., Wu, J., Trehewella, J., Bax, A., (2005). Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. *J. Am. Chem. Soc.* **127**, 16621–16628. <https://doi.org/10.1021/ja054342m>.

[40]. Kikhney, A.G., Borges, C.R., Molodenskiy, D.S., Jeffries, C.M., Svergun, D.I., (2020). SASBDB: towards an automatically curated and validated repository for biological scattering data. *Protein Sci.* **29**, 66–75. <https://doi.org/10.1002/pro.3731>.

[41]. 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. <https://doi.org/10.1107/S1600576715024942>.

[42]. Rappaport, J., Siniossoglou, S., Hurt, E.C., Mann, M., (2000). A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Anal. Chem.* **72**, 267–275. <https://doi.org/10.1021/ac991081o>.

[43]. Graziadei, A., Rappaport, J., (2022). Leveraging crosslinking mass spectrometry in structural and cell biology. *Structure* **30**, 37–54. <https://doi.org/10.1016/j.str.2021.11.007>.

[44]. 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. <https://doi.org/10.1016/j.tibs.2015.10.008>.

[45]. Fürsch, J., Kammer, K.-M., Kreft, S.G., Beck, M., Stengel, F., (2020). Proteome-wide structural probing of low-abundant protein interactions by cross-linking mass spectrometry. *Anal. Chem.* **92**, 4016–4022. <https://doi.org/10.1021/acs.analchem.9b05559>.

[46]. Sarnowski, C.P., Knörlein, A., de Vries, T., Götze, M., Beusch, I., Aebersold, R., Allain, F.-H.-T., Hall, J., Leitner, A., (2025). A highly sensitive protein-RNA cross-linking mass spectrometry workflow with enhanced structural modeling potential. *Nucleic Acids Res.* **53** <https://doi.org/10.1093/nar/gkaf523>.

[47]. Perez-Riverol, Y., Bandla, C., Kundu, D.J., Kamatchinathan, S., Bai, J., Hewapathirana, S., John, N.S., Prakash, A., Walzer, M., Wang, S., Vizcaíno, J.A., (2025). The PRIDE database at 20 years: 2025 update. *Nucleic Acids Res.* **53**, D543–D553. <https://doi.org/10.1093/nar/gkae1011>.

[48]. Combe, C.W., Kolbowski, L., Fischer, L., Koskinen, V., Klein, J., Leitner, A., Jones, A.R., Vizcaíno, J.A., Rappaport, J., (2024). mzIdentML 1.3.0 – essential progress on the support of crosslinking and other identifications based on multiple spectra. *Proteomics* **24**, e2300385. <https://doi.org/10.1002/pmic.202300385>.

[49]. Matthew Allen Bullock, J., Schwab, J., Thalassinos, K., Topf, M., (2016). The importance of non-accessible crosslinks and solvent accessible surface distance in modeling proteins with restraints from crosslinking mass spectrometry. *Mol. Cell. Proteomics* **15**, 2491–2500. <https://doi.org/10.1074/mcp.M116.058560>.

[50]. Merkley, E.D., Rysavy, S., Kahraman, A., Hafen, R.P., Daggett, V., Adkins, J.N., (2014). Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. *Protein Sci.* **23**, 747–759. <https://doi.org/10.1002/pro.2458>.

[51]. Molnar, K.S., Bonomi, M., Pellarin, R., Clinthorne, G.D., Gonzalez, G., Goldberg, S.D., Goulian, M., Sali, A., DeGrado, W.F., (2014). Cys-scanning disulfide crosslinking and Bayesian modeling probe the transmembrane signaling mechanism of the histidine kinase, PhoQ. *Structure* **22**, 1239–1251. <https://doi.org/10.1016/j.str.2014.04.019>.

[52]. Erzberger, J.P., Stengel, F., Pellarin, R., Zhang, S., Schaefer, T., Aylett, C.H.S., Cimermančič, P., Boehringer, D., Sali, A., Aebersold, R., Ban, N., (2014). Molecular architecture of the 40S<sub>e</sub>elF1<sub>s</sub>elF3 translation initiation complex. *Cell* **159**, 1227–1228. <https://doi.org/10.1016/j.cell.2014.11.001>.

[53]. Frank, J., (2023). Novel Developments in Cryo-EM of Biological Molecules: Resolution in Time and State

Space. Jenny Stanford Publishing, New York <https://doi.org/10.1201/9781003456100>.

[54]. Lucić, V., Förster, F., Baumeister, W., (2005). Structural studies by electron tomography: from cells to molecules. *Annu. Rev. Biochem.* **74**, 833–865. <https://doi.org/10.1146/annrev.biochem.73.011303.074112>.

[55]. von Appen, A., Beck, M., (2016). Structure determination of the nuclear pore complex with three-dimensional cryo electron microscopy. *J. Mol. Biol.* **428**, 2001–2010. <https://doi.org/10.1016/j.jmb.2016.01.004>.

[56]. 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., Schröder, G.F., Steven, A.C., Stokes, D.L., Westbrook, J.D., Wriggers, W., Yang, H., Young, J., Berman, H.M., Chiu, W., Kleywelt, G.J., Lawson, C.L., (2012). Outcome of the first electron microscopy validation task force meeting. *Structure* **20**, 205–214. <https://doi.org/10.1016/j.str.2011.12.014>.

[57]. Lawson, C.L., Kryshtafovych, A., Adams, P.D., Afonine, P.V., Baker, M.L., Barad, B.A., Bond, P., Burnley, T., Cao, R., Cheng, J., Chojnowski, G., Cowtan, K., Dill, K.A., DiMaio, F., Farrell, D.P., Fraser, J.S., Herzik Jr, M.A., Hoh, S.W., Hou, J., Hung, L.-W., Igae, M., Joseph, A.P., Kihara, D., Kumar, D., Mittal, S., Monastyrskyy, B., Olek, M., Palmer, C.M., Patwardhan, A., Perez, A., Pfab, J., Pintilie, G.D., Richardson, J.S., Rosenthal, P.B., Sarkar, D., Schäfer, L.U., Schmid, M.F., Schröder, G.F., Shekhar, M., Si, D., Singharoy, A., Terashi, G., Terwilliger, T.C., Vaiana, A., Wang, L., Wang, Z., Wankowicz, S.A., Williams, C.J., Winn, M., Wu, T., Yu, X., Zhang, K., Berman, H.M., Chiu, W., (2021). Cryo-EM model validation recommendations based on outcomes of the 2019 EMDataResource challenge. *Nature Methods* **18**, 156–164. <https://doi.org/10.1038/s41592-020-01051-w>.

[58]. Wang, Z., Patwardhan, A., Kleywelt, G.J., (2022). Validation analysis of EMDB entries. *Acta Crystallogr. D Struct. Biol.* **78**, 542–552. <https://doi.org/10.1107/S205979832200328X>.

[59]. Gore, S., Sanz García, 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., Kleywelt, G.J., (2017). Validation of structures in the Protein Data Bank. *Structure* **25**, 1916–1927. <https://doi.org/10.1016/j.str.2017.10.009>.

[60]. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., Jain, S., Lewis, S.M., Arendall 3rd, W.B., Snoeyink, J., Adams, P.D., Lovell, S.C., Richardson, J.S., Richardson, D.C., (2018). MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315. <https://doi.org/10.1002/pro.3330>.

[61]. Gore, S., Velankar, S., Kleywelt, G.J., (2012). Implementing an X-ray validation pipeline for the Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 478–483. <https://doi.org/10.1107/S0907444911050359>.

[62]. Ullanat, V., Kasukurthi, N., Viswanath, S., (2022). PrISM: precision for integrative structural models. *Bioinformatics* **38**, 3837–3839. <https://doi.org/10.1093/bioinformatics/btac400>.

[63]. Ganesan, S.J., Feyder, M.J., Chemmama, I.E., Fang, F., Rout, M.P., Chait, B.T., Shi, Y., Munson, M., Sali, A., (2020). Integrative structure and function of the yeast exocyst complex. *Protein Sci.* **29**, 1486–1501. <https://doi.org/10.1002/pro.3863>.

[64]. Manalastas-Cantos, K., Konarev, P.V., Hajizadeh, N.R., Kikhney, A.G., Petoukhov, M.V., Molodenskiy, D.S., Panjkovich, A., Mertens, H.D.T., Gruzinov, A., Borges, C., Jeffries, C.M., Svergun, D.I., Franke, D., (2021). Expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr.* **54**, 343–355. <https://doi.org/10.1107/S1600576720013412>.

[65]. Kurtzer, G.M., Sochat, V., Bauer, M.W., (2017). Singularity: Scientific containers for mobility of compute. *PLoS One* **12**, e0177459. <https://doi.org/10.1371/journal.pone.0177459>.

[66]. Schuler, R.E., Kesselman, C., Czajkowski, K., (2016). Accelerating data-driven discovery with scientific asset management. In: *2016 IEEE 12th International Conference on E-Science*. e-Science. <https://doi.org/10.1109/escience.2016.7870883>.

[67]. Shah, S.Z., Perry, T.N., Graziadei, A., Cecatiello, V., Kaliyappan, T., Misiaszek, A.D., Müller, C.W., Ramsay, E. P., Vannini, A., (2025). Structural insights into distinct mechanisms of RNA polymerase II and III recruitment to snRNA promoters. *Nature Commun.* **16**, 141. <https://doi.org/10.1038/s41467-024-55553-8>.

[68]. Pintilie, G., Zhang, K., Su, Z., Li, S., Schmid, M.F., Chiu, W., (2020). Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nature Methods* **17**, 328–334. <https://doi.org/10.1038/s41592-020-0731-1>.

[69]. Farrell, D.P., Anishchenko, I., Shakeel, S., Lauko, A., Passmore, L.A., Baker, D., DiMaio, F., (2020). Deep learning enables the atomic structure determination of the Fanconi Anemia core complex from cryoEM. *IUCrJ* **7**, 881–892. <https://doi.org/10.1107/S20522520009306>.

[70]. Rizzo, A.A., Vassel, F.-M., Chatterjee, N., D'Souza, S., Li, Y., Hao, B., Hemann, M.T., Walker, G.C., Korzhnev, D. M., (2018). Rev7 dimerization is important for assembly and function of the Rev1/Pol $\zeta$  translesion synthesis complex. *PNAS* **115**, E8191–E8200. <https://doi.org/10.1073/pnas.1801149115>.

[71]. Peulen, T.-O., Hengstenberg, C.S., Biehl, R., Dimura, M., Lorenz, C., Valeri, A., Folz, J., Hanke, C.A., Ince, S., Vöpel, T., Farago, B., Gohlke, H., Klare, J.P., Stadler, A. M., Seidel, C.A.M., Herrmann, C., (2023). Integrative dynamic structural biology unveils conformers essential for the oligomerization of a large GTPase. *eLife* **12** <https://doi.org/10.7554/eLife.79565>.