

RCSB Protein Data Bank: Delivering Integrative Structures Alongside Experimental Structures and Computed Structure Models

Brinda Vallat^{1,2,*}, Yana Rose³, Dennis W. Piehl¹, Jose M. Duarte³, Sebastian Bittrich³, Chunxiao Bi³, Joan Segura³, Arthur Zalevsky⁴, Monica R. Sekharan¹, Benjamin M. Webb⁴, Charmi Bhikadiya³, Maria Voigt¹, Jeremy Henry³, Henry Chao¹, Aditya Pingale¹, David Raymond¹, Jared Sagendorf⁴, Ezra Peisach¹, Zukang Feng¹, Vladimir Guranovic¹, James Smith¹, Alexander S. Rose⁵, Jasmine Y. Young¹, Helen M. Berman^{1,6,7}, Andrej Sali⁴, Stephen K. Burley^{1,2,3,6,8}

¹Research Collaboratory for Structural Bioinformatics Protein Data Bank and the Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

²Rutgers Cancer Institute, Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA

³Research Collaboratory for Structural Bioinformatics Protein Data Bank, San Diego Supercomputer Center, University of California, La Jolla, CA 92093, USA

⁴Research 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

⁵Mol* Consortium, San Diego, California, USA

⁶Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

⁷Department of Quantitative and Computational Biology, University of Southern California, Los Angeles CA 90089, USA

⁸Rutgers Artificial Intelligence and Data Science (RAD) Collaboratory, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

*Contact Author: brinda.vallat@rcsb.org

Abstract

The Protein Data Bank (PDB) archives three-dimensional (3D) structures of macromolecules determined experimentally using various methods. It is jointly managed by the Worldwide Protein Data Bank (wwPDB) consortium. Research Collaboratory for Structural Bioinformatics (RCSB) PDB, the US data center for the PDB, provides streamlined access to more than 240,000 structures through a variety of research-focused tools on [RCSB.org](https://www.rcsb.org). In addition, RCSB.org makes available over 1 million computed structure models predicted using deep learning methods and archived in the AlphaFold Database and ModelArchive. The PDB-IHM system was developed as a wwPDB project based on community recommendations to archive structures determined using integrative/hybrid methods (IHM). These structures are computed by combining information from multiple experimental and computational techniques to overcome the limitations of traditional single methods (e.g., macromolecular crystallography, 3D electron microscopy, nuclear magnetic resonance spectroscopy). In 2024, PDB-IHM was unified with the PDB to archive integrative structures alongside single-method experimental structures. These integrative structures have been made accessible *via* the RCSB.org website, facilitating efficient

delivery of IHM data to a broad community of PDB users. Herein, we describe the expanded capabilities of RCSB.org that support discovery, analysis, and visualization of integrative structures together with single-method experimental structures and computed structure models.

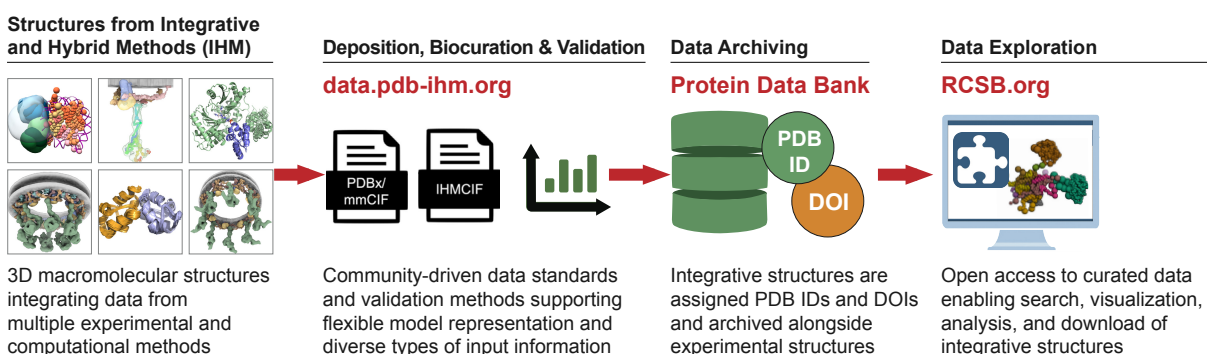
Key Points

- Integrative structures are archived alongside experimental structures in the PDB
- Integrative structures are now accessible via the RCSB.org research-focused web portal
- RCSB.org provides full support for discovery, exploration, and analysis of integrative structures

Keywords

PDBx/mmCIF; IHMCIF; PDB-IHM; wwPDB; Integrative Methods; RCSB.org; Structure Validation

Graphical Abstract



Introduction

The Protein Data Bank (PDB) is the single global repository for experimentally determined three-dimensional (3D) atomic structures of biomolecules and their complexes (1,2). Managed by the worldwide Protein Data Bank (wwPDB) organization (3), PDB promotes the FAIR (Findable, Accessible, Interoperable, and Reusable) (4) and FACT (Fairness, Accuracy, Confidentiality, and Transparency) (5) principles of scientific data management. The PDB primarily archives structures determined by macromolecular crystallography (MX), nuclear magnetic resonance (NMR) spectroscopy, and 3D electron microscopy (3DEM). Since its establishment in 1971, the archive has grown more than 30,000-fold and currently holds more than 240,000 3D biostructures that are expertly curated and rigorously validated based on community guidelines.

The Research Collaboratory for Structural Bioinformatics (RCSB) PDB (6) is the US data center for the wwPDB. It provides access to PDB data through the research-focused RCSB.org web portal. RCSB.org offers a variety of tools and features that support searching, browsing, visualizing, validating, and analyzing macromolecular structures. As a living digital resource, RCSB.org is updated weekly to incorporate new structures and enrich our understanding of

every structure in the archive with annotations from ~50 trusted external resources. In 2022, we expanded RCSB.org to provide access to approximately 1 million computed structure models (CSMs) from AlphaFold Protein Structure Database (AlphaFoldDB) (7) and the ModelArchive (8)—including complete proteomes of model organisms and select pathogens—thereby enhancing the availability of 3D biostructure data to the scientific community (9).

Advances in structure determination led to the development of integrative/hybrid methods (IHM) that combine information from multiple experimental and computational sources (10). These sources can include 3DEM, small-angle scattering (SAS), chemical crosslinking mass spectrometry (crosslinking-MS), Förster resonance energy transfer (FRET), electron paramagnetic resonance (EPR) spectroscopy, and many others. In addition, available structures of assembly components can be used as starting models. Integrative methods are complementary to traditional structure determination methods and are typically used to determine structures of larger macromolecules or complex assemblies that may be difficult or impossible to resolve with a single technique. Examples of macromolecular complexes studied using integrative approaches include the structure of the human LINE-1 ORF2p protein (11), the structure of the human SNAPc-DNA (12), and the assembly pathway of the human nuclear pore complex (13,14).

Based on community recommendations made by the wwPDB IHM Task Force (15,16), the PDB-Dev system was developed for deposition, biocuration, validation, and dissemination of integrative structures (17,18). This prototype system was built alongside the wwPDB OneDep system (19) that handles experimental structures archived in the PDB. In 2024, PDB-Dev was unified with the PDB and rebranded as PDB-IHM to denote the “IHM” branch of the PDB archive (20). Importantly, integrative structures are now issued PDB accession codes and archived alongside experimental structures in the PDB. Following unification, RCSB.org capabilities were extended to support dissemination of integrative structures. Herein, we describe the new functionalities of RCSB.org that provide expanded access to integrative structures together with single-method experimental structures and CSMs, thus making it a comprehensive resource for 3D macromolecular structure data.

Results and Discussion

With the increasing application of integrative methods, the structural biology community has encouraged the PDB to create robust mechanisms for archiving, validating, and disseminating integrative structures in a FAIR manner (15,16). Archiving integrative structures introduced additional requirements when compared to single-method experimental structures in the PDB archive to accommodate new data representations and metadata:

1. Integrative structures can be multi-scale, including both atomic and coarse-grained representations.
2. Integrative structures can be multi-state, corresponding to multiple conformational and compositional states jointly satisfying the input restraints. In addition, the multiple states can be ordered to depict a pathway, such as an enzymatic reaction or a biochemical pathway.
3. Integrative modeling can result in a collection of models (ensembles), where each model independently satisfies the input restraints within an acceptable threshold.
4. Input restraints used in integrative modeling can be of diverse types, ranging from pairwise distances between atoms, residues, or domains fit to 2D images and 3D volumes.

5. Starting structure models of system components can come from experimental, computational, or integrative methods, and can be rigid or flexible during modeling.
6. Integrative approaches make use of many different modeling algorithms and software tools.

Accordingly, archiving integrative structures required creation of new data standards, tools, and workflows that satisfy the above requirements, capture the provenance information for all input data, enable automated data processing, support rigorous assessment of these structures, and facilitate efficient data dissemination to a broad community of users. In the following sections, we describe aspects of PDB-IHM infrastructure and extensions of the RCSB.org web portal that accomplish these goals.

PDB-IHM Data Deposition, Biocuration, and Validation

The foundational data standard for archiving integrative structures is provided by IHMCIF (21), an extension of the PDBx/mmCIF data dictionary (22,23) used by the PDB. At its core, IHMCIF follows the same set of definitions for representing 3D atomic structures of macromolecules as PDBx/mmCIF. Information specific to integrative structure determination, such as definitions for multi-scale, multi-state, and ordered-state structural models, spatial restraints from various experimental data types (*e.g.*, 3DEM, crosslinking-MS, FRET, SAS, EPR), starting structure models used, and references to experimental data archived in external repositories, are defined in the IHMCIF extension.

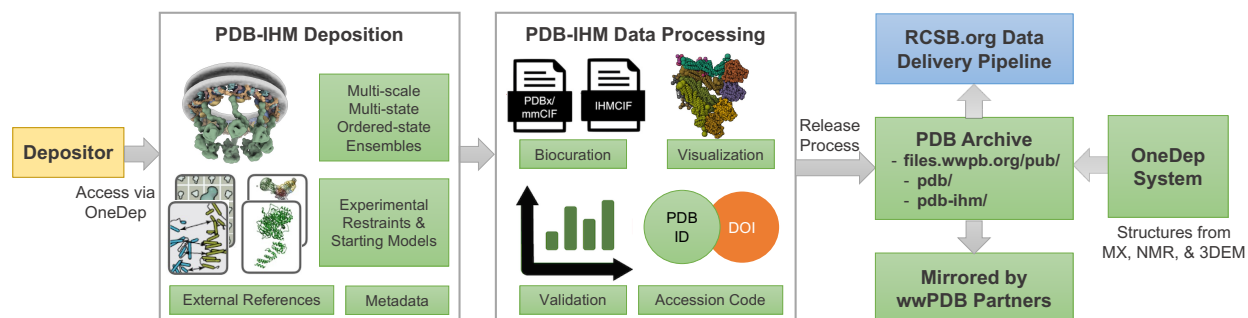


Figure 1: The PDB-IHM data pipeline supports (i) deposition of integrative structures, associated experimental restraints and starting models, references to related data in external resources, and other metadata details; (ii) data processing steps for manual and automated biocuration, structure validation, visualization, and issue of PDB accession codes and digital object identifiers (DOIs); and (iii) release process to the PDB archive. The PDB-IHM deposition interface can be accessed via the wwPDB OneDep system and works in parallel to OneDep, which handles structures from MX, NMR, and 3DEM. Bulk deposition via an API is also supported. Integrative structures archived in the PDB along with experimental structures are mirrored by the wwPDB partners and are the primary source of data for the RCSB.org data delivery pipeline.

The PDB-IHM system for deposition, biocuration, and validation (data.pdb-ihm.org; Figure 1) relies on the PDBx/mmCIF and IHMCIF data standards and related tools for processing integrative structures. It can be accessed through and operates in parallel with the wwPDB OneDep system and was built using the DERIVA scientific asset management software (24,25). The system supports harvesting and deposition of heterogeneous data and metadata associated with integrative structure modeling investigations (18,20).

PDB-IHM biocuration includes both automated and manual steps to ensure data consistency, completeness, and compliance with archive requirements, similar to OneDep biocuration (26).

Once the process is completed, PDB accession codes are issued, and validation reports are generated. Methods for validating integrative structures have been developed (unpublished results) based on recommendations from the wwPDB IHM Task Force (15,16) and in collaboration with community experts. A validation software package has been implemented within the IHM deposition and curation pipeline, so that users can obtain full and summary versions of validation reports along with accession codes and use them as part of the manuscript review process.

Unification of PDB-IHM with the PDB archive

Although PDB-IHM (pdb-ihm.org) (20) was developed in parallel with OneDep, the eventual goal was to unify it with the PDB such that integrative structures are archived alongside single-method experimental structures. In 2024, unification was completed, and integrative structures are now provided under the “pdb_ihm” directory within the PDB archive.

Following unification, integrative structures can be downloaded from the PDB archive, which is available across multiple wwPDB partner sites, including rcsb.org, pdbe.org, and pdbj.org. Repository contents include mmCIF files, validation reports, and files reporting archive holdings. Within mmCIF files, a specific data item called “struct.structure_determination_methodology” is populated as “integrative” to clearly identify structures determined using integrative methods. Both new and pre-existing integrative structures are now identified with PDB accession codes (previously issued PDB-Dev accession codes are preserved in the “database_2” mmCIF category for provenance).

The PDB-IHM data release workflow further ensures timely delivery of updated/released IHM entries via the PDB archive, synchronously with the weekly update process for single-method PDB structures. New integrative structures are issued digital object identifiers (DOIs) based on the PDB accession codes and the corresponding DOI landing pages are made available on the wwPDB.org website. DOIs allow online versions of primary publications to be cross-referenced to every structure stored in the PDB archive.

Incorporation of Integrative Structures into the RCSB.org Data Delivery Pipeline

Following the unification of PDB-IHM with PDB and the availability of integrative structures in the PDB archive, RCSB.org was extended to support parallel delivery of integrative structures. In 2022, a similar extension was carried out to incorporate ~1 million CSMs (27) from AlphaFoldDB (7) and the ModelArchive (8). We followed a similar approach to incorporate integrative structures into the RCSB.org data delivery pipeline (Figure 2).

Data management is based on a schema-first approach, wherein every step of the data delivery pipeline together with supporting services and tools is informed by metadata encoded in a standard JSON (JavaScript Object Notation) schema representation (28). These schemas are derived from the PDBx/mmCIF definitions for single-method PDB structures and the ModelCIF (29) definitions for CSMs. We extended the RCSB.org schemas to include the IHMCIF extension dictionary (21), which allows information to be extracted and consumed from IHMCIF data files. Additional definitions were introduced to the internal RCSB.org schema to support efficient search and data delivery. For example, the “rcsb_entry_info” category was extended with attributes that identify features specific for integrative structures, such as multi-scale, multi-state, ordered-state, and/or ensemble representations.

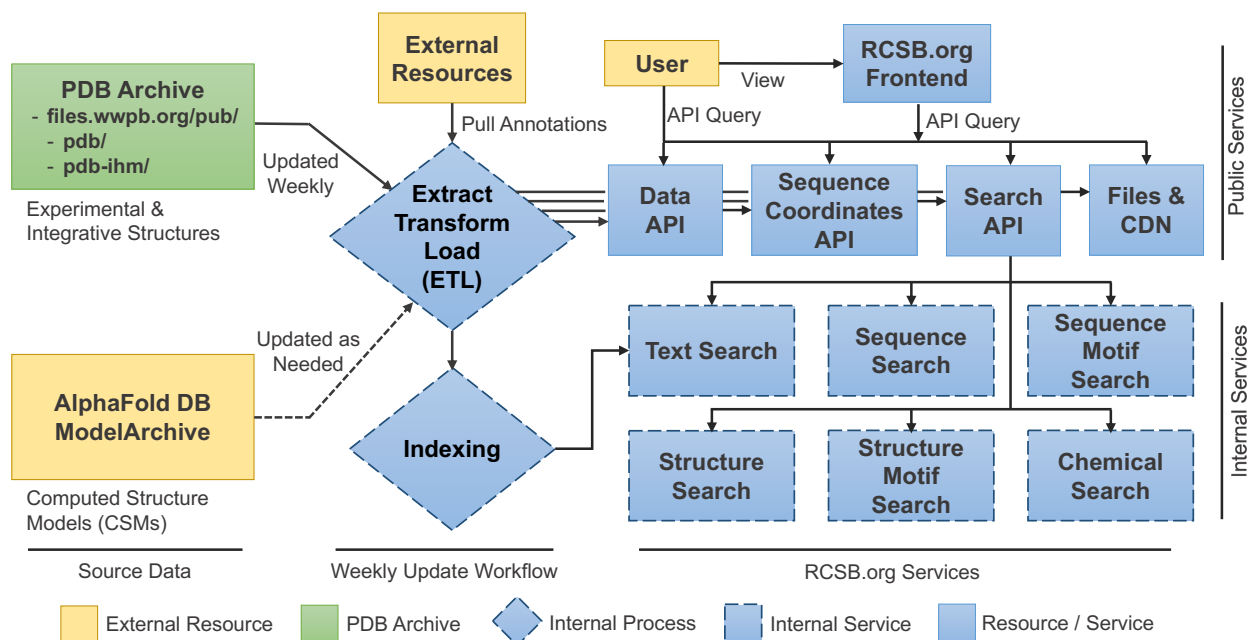


Figure 2: Overview of the RCSB.org modular architecture and data flow. Experimental and integrative structures are obtained from the PDB archive weekly, while CSMs are sourced from AlphaFoldDB and ModelArchive on an as-needed basis and stored on our servers. Annotations obtained from approximately 50 external resources are integrated with the source data through an Extract–Transform–Load (ETL) process, which automates data retrieval, standardization, and incorporation of heterogeneous datasets into the RCSB.org data warehouse. The front-end web portal integrates with multiple back-end services, including content delivery via a Content Delivery Network (CDN) that distributes static assets such as images and files; metadata access through the Data API (GraphQL and REST endpoints at data.rcsb.org); search operations through the Search API (REST endpoints at search.rcsb.org); and sequence annotation delivery through the Sequence Coordinates API (GraphQL endpoints at sequence-coordinates.rcsb.org). These services are supported by a suite of dedicated APIs and microservices.

The RCSB.org weekly update pipeline has been extended to support PDB-IHM data, including modifications of backend tools for extracting, transforming, and loading data into the RCSB.org data warehouse, plus libraries for integrating annotations from ~50 trusted external biodata resources. PDB accession codes serve as unique identifiers for integrative structures. Primary and computed data, together with external annotations, are processed and loaded for a representative model in a manner consistent with single-method experimental structures. If an entry contains multiple structure models (e.g., ensemble), then Depositor(s) can select one of these models to be the representative conformation. If provided by the Depositor(s), the representative model is identified based on annotations in the mmCIF file; otherwise, it is chosen as the first model containing the largest number of molecules. Moreover, provenance information is explicitly captured (*via* the “`rcsb_entry_info.structure_determination_methodology`” data item) to distinguish integrative structures from single-method experimental structures and CSMs. Finally, backend workflows for indexing, sequence clustering, image generation, and BinaryCIF (30) file generation were extended to fully support integrative structures, including multi-scale structures with coarse-grained representations.

The RCSB.org user interfaces, data delivery, and search services are supported by a set of modular microservices (28) (Figure 2). The front-end service, accessible *via* the RCSB.org web

portal, integrates with multiple back-end application programming interface (API) services. Coordinate files in text and binary formats are distributed through the Files Download service (files.rcsb.org) and ModelServer API (models.rcsb.org), while static content, such as images, is delivered *via* a content delivery network (CDN). Structured metadata are available programmatically through the Data API (data.rcsb.org), which offers both Representational State Transfer (REST) and GraphQL endpoints. Search operations are supported by the Search API (search.rcsb.org), which offers a unified interface to heterogeneous search modalities, including 3D structure and sequence similarity searches, 3D and sequence motif searches, and chemical structure searches—each supported by a dedicated microservice. Sequence alignments and annotations are served through the Sequence Coordinates API (sequence-coordinates.rcsb.org). All single-method experimental structures and integrative structures released in the archive are immediately accessible via the API services and RCSB.org, which are updated in synchrony with the weekly data releases every Wednesday at 00:00 Universal Time Coordinated (UTC).

The screenshot shows the RCSB.org homepage. At the top, the RCSB PDB logo is on the left. Next to it are two boxes: one with a flask icon and '241,760 Structures from the PDB archive' (labeled A), and another with a computer monitor icon and '1,068,577 Computed Structure Models (CSM)' (labeled B). To the right is a search bar with the placeholder 'Enter search term(s), Ligand ID or sequence' (labeled C), an 'Include CSM' toggle, and a magnifying glass icon. Below the search bar are links for 'Advanced Search' and 'Browse Annotations'. A navigation bar below the search bar contains logos for PDB-101, PDB, EMDDataResource, NAKB, WW-PDB Foundation, and PDB-IHM, along with social media icons. A banner below the navigation bar reads 'Redesigned PDB Statistics Support Enhanced Functionality' with an 'Explore Statistics' button. The main content area has a left sidebar with links: Welcome, Deposit, Search, Visualize, Analyze, Download, and Learn. The main content area includes a 'September Molecule of the Month' section with a green molecular model of the Absciscic Acid Receptor, a 'NEW Explore Integrative Structures' section with a puzzle piece icon, and a 'PDB-101 Training Resources' section with a lightbulb icon.

Figure 3: RCSB.org homepage (as of September 2025). (A) Display of structure counts accompanied by dedicated icons that distinguish structure types: dark-blue “puzzle piece” for integrative structures, dark-blue “Erlenmeyer flask” for experimentally determined structures, and light-blue “computer monitor” for computed structure models (CSMs). Clicking on these icons launches searches across the entire corresponding datasets. (B) Link to supporting documentation that provides guidance on data access and search functionality for integrative structures. (C) Keyword-based search bar for full-text queries.

RCSB.org Functionalities Supporting Integrative Structures

The extended RCSB.org web portal provides comprehensive support for the discovery, exploration, validation, and analysis of integrative structures, fully incorporated into the website search, navigation, and visualization functionalities. From the homepage, the total count for the PDB archive reflects both single-method experimental structures and integrative structures (Figure 3A). To ensure clear visual distinction among structure types, consistent use of icons and colors conveys data provenance across RCSB.org: a dark-blue puzzle piece icon represents the integrative structures, a dark-blue Erlenmeyer flask icon denotes the single-

method experimental structures, and a light-blue computer monitor icon denotes CSMS. Search links associated with each icon type allow direct access to the corresponding dataset. In addition, the RCSB.org homepage provides a link to supporting documentation with guidance on data access and search functionality for integrative structures (Figure 3B).

Integrative structures can be accessed via keyword-based searches (Figure 3C), which return both single-method experimental and integrative structures (Figure 4). Search result snippets for integrative structures are explicitly marked with a puzzle-piece icon (Figure 4A) and report the structure determination methodology (Figure 4B), ensuring that users are immediately informed of the provenance. Faceted filters, including *Structure Determination Methodology* (Figure 4C) and *Integrative Input Data* (Figure 4D), allow refinement of search results to exclusively display integrative structures, facilitating targeted exploration of specific modeling approaches.

The screenshot displays the RCSB.org search results interface. On the left, a sidebar titled 'Refinements' contains several filter categories: 'Structure Determination Methodology' (with sub-options for experimental and integrative), 'Scientific Name of Source Organism' (Bos taurus), 'Taxonomy' (Eukaryota), 'Experimental Method' (ELECTRON MICROSCOPY), 'Integrative Input Data' (3DEM volume, Comparative model, Crosslinking-MS data, De Novo model, Experimental model), 'Polymer Entity Type' (Protein), and 'Refinement Resolution'. The main content area shows two search results. The first result, 6VNW, is a Cryo-EM structure of apo-BBSome, determined by Yang, S., Walz, T., Nachury, M., and Chou, H. in 2020. The second result, 8ZZI, is titled 'The molecular architecture of the BBSome and its implications for facilitated transition zone crossing' and was determined by Chou, H., Apelt, L., Farrell, D.P., White, S.R., Woodsmith, J., Svetlov, V., Goldstein, J.S., Nager, A.R., Li, Z., Muller, J., Dollfus, H., Nudler, E., Stelzl, U., DiMaio, F., Nachury, M.V., and Walz, T. in 2019. This entry is marked as integrative with a puzzle-piece icon and a yellow background. The 'Structure Determination Methodology' for 8ZZI is listed as 'Integrative'. The 'Organisms' are 'Bos taurus' and the 'Macromolecule' consists of 'Unique protein chains: 8'.

Figure 4: Search results page displaying both experimental and integrative structures. (A) Distinct structure-type icons: a puzzle piece for integrative structures and a flask for experimentally determined structures. (B) Search result snippets for integrative structures explicitly report the *Structure Determination Methodology* used. Refinement option for filtering results by (C) *Structure Determination Methodology* and (D) *Integrative Input Data*.

The Advanced Search interface (rcsb.org/search/advanced) offers fine-grained control for working with integrative structures (Figure 5). Search parameters include model features (e.g., multi-scale, multi-state, ordered-state models) (Figure 5A), types of input experimental data (e.g., crosslinking-MS, 3DEM maps, NMR restraints) (Figure 5B), and associated dataset accessions from external repositories, including Biological Magnetic Resonance Data Bank (BMRB) (31), Electron Microscopy Data Bank (EMDB) (32), small-angle scattering biological data bank (SASBDB) (33), and PRoteomics IDentification Database (PRIDE) (34) (Figure 5C). In addition to these metadata-based criteria, Advanced Search supports scientific queries, including sequence similarity (Figure 5D), 3D structure similarity (Figure 5E), and motif-based sequence or structural searches (Figure 5F and 5G). These capabilities allow retrieval of integrative structures that meet precisely defined criteria.

Advanced Search Query Builder Help

Full Text ?

Structure Attributes ? Help

AND	IHM Multi-Scale Flag	x	is	N	A	+ NOT	Count	x
AND	IHM Dataset Name	x	is	Crosslinking-MS data	B	+ NOT	Count	x
AND	IHM Dataset Source DB Accession Code - AlphaFoldDB	x	is not empty		C	+ NOT	Count	x

AND / OR Add Attribute Add Subquery Remove Subquery

Add Subquery

Chemical Attributes ?

Sequence Similarity ? **D** Help

AND SNAELFNLESRVIEKSLTQMEDYKALQMKLWEASKLSFATCKS

Entry ID 1MBN Sequence Type Protein E-Value Cutoff 0.1 Identity Cutoff 0 % (Integer only) ? Count Clear

Sequence Motif ? **F**

Structure Similarity ? **E**

Structure Motif ? **G**

Chemical Similarity ?

Return Structures grouped by No Grouping Include Computed Structure Models (CSM) ? ☐ Count Clear Search

Figure 5: Example query in the Advanced Search interface for integrative structures. The query combines metadata-based criteria, including (A) atomic resolution integrative structures that (B) utilize crosslinking-MS experimental data and (C) AlphaFoldDB *de novo* models as starting structures, with (D) sequence-based searches. Additional scientific searches, such as (E) 3D structure similarity and motif-based (F) sequence or (G) structure searches, can also be applied to further refine results.

Each integrative structure is presented through its own Structure Summary Page (SSP) that provides an overview of the representative model, associated metadata, and supporting experimental data. To facilitate recognition, each page is marked with a dedicated icon identifying integrative structures (Figure 6A) and an explanatory note describing integrative structures (Figure 6B). Coordinate data are available for download in both mmCIF and BinaryCIF formats compatible with the IHMCIF dictionary (Figure 6C). Two types of validation reports are provided in the PDF format: a full report and a summary report (Figure 6D). SSPs further illuminate high-level model features, such as whether the model is multi-scale, multi-state, and/or ordered-state, along with a clear indication of the currently loaded representative model (Figure 6E). In addition, users are provided with direct links to launch similarity-based searches (Figure 6F).

SSPs are organized into specialized tabs (Figure 6G): Structure tab – interactive visualization of the integrative structures *via* the Mol* viewer (35) directly within the web browser, with support for displaying multi-scale assemblies that combine atomic and coarse-grained representations (Figure 7A and 7B). In addition to these representations, crosslink restraint data can be displayed, with the option to color-code restraints according to differences between the experimental input restraints and observed distances in the deposited structure, enabling immediate visual assessment of the agreement between model and underpinning data (Figure 7C and 7D). Hovering over a restraint visualization displays detailed information about the crosslink restraint parameters (e.g., type, threshold, uncertainties Ψ , σ_1 , & σ_2 , (36) and distance) (Figure 7E). Annotations tab – cross-references to external biological resources. Experiment tab – a summary of input datasets used in modeling (e.g., 3DEM maps, crosslinking-MS restraints, and starting models), together with links to related data in EMDB, PRIDE, and other repositories (Figure 8). Sequence tab – mappings of sequence features to structural elements, including domains and functional annotations, with interactive visualization bridging sequence and structure views.



Figure 6: Structure summary page (SSP) for the integrative structure of the Mis18 complex (PDB ID: [pdb_00009a8g](#)) (37). (A) A puzzle piece icon used to identify integrative structures. (B) An explanatory note describing integrative structures. (C) Coordinate data available for download in mmCIF and BinaryCIF formats. (D) Validation reports provided in PDF format (full and summary). (E) High-level model features (e.g., multi-scale, multi-state, ordered) and indication of the currently loaded model. (F) Link to launch a 3D similarity scientific search. (G) Specialized tabs provide focused views on different aspects of a structure.

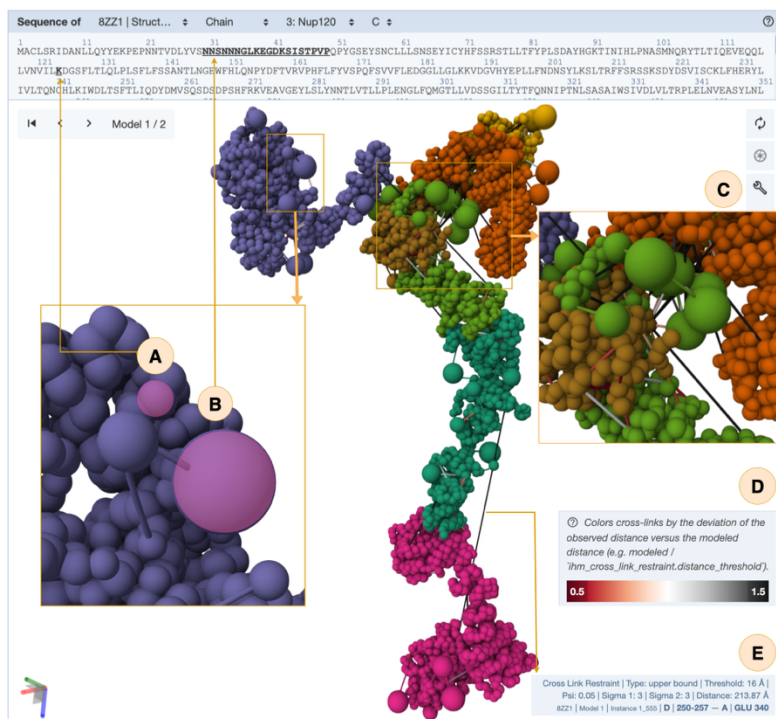


Figure 7. Visualization of the Nup84 sub-complex of the Nuclear Pore Complex (PDB ID: [pdb_00008zz1](#)) (38) in the Mol* viewer. Multi-scale representation, with coarse-grained spheres corresponding to (A) single residues and (B) residue ranges. (C) Visualization of crosslinked residues connected by lines. (D) Crosslink restraints that are bound by an upper limit are color-coded according to the deviation between the experimentally-derived input restraints used for modeling and the distances observed in the deposited

model. (E) Hovering over a restraint visualization displays detailed information about the crosslink restraint parameters (e.g., type, threshold, uncertainties Ψ , σ_1 , & σ_2 , and distance).

Structure Summary
Structure
Annotations
Experiment
Sequence
Genome
Versions

Display Files
Download Files
Data API

9A8G | **pdb_00009a8g**

Structural Basis for Mis18 Complex Assembly

INTEGRATIVE STRUCTURE

Input Data A

Name	Type	Count
3DEM volume	Experimental data	3
Crosslinking-MS data	Experimental data	1
De Novo model	Starting model	3
Experimental model	Starting model	2

Related Data Resources B

Database Name	Accession Code
PRIDE	PXD047345
EMDB	EMD-50218 EMD-50219 EMD-50220
PDB	7SFZ 7SFY
AlphaFoldDB	AF-Q9NYP9-F1 AF-O43482-F1 AF-Q6P0N0-F1

Figure 8: [Experimental details](#) for the integrative structure of the Mis18 complex (PDB ID: [pdb_00009a8g](#)). (A) Input experimental datasets used for modeling. (B) Related datasets deposited in external repositories.

Integrative structures are also included in “Groups” pages on RCSB.org (39) that cluster experimental structures, integrative structures, and CSMs based on sequence similarity or mapping to reference UniProt (40) sequences. This unified framework allows side-by-side comparison of 3D information from different sources. Moreover, the pairwise structure alignment tool (41) enables quantitative and visual comparison of atomic integrative structures, single-method experimental structures, and CSMs. Alignments can be performed for any structure or UniProt accession, supporting cross-methodology structural comparison and analysis.

Conclusions and Future Perspectives

Unifying PDB-IHM with the PDB archive and making integrative structures accessible *via* RCSB.org has broadened the scope and capabilities of the PDB. Given the growing importance of integrative modeling in elucidating 3D structures of larger biomolecules and macromolecular complexes, these efforts address important community requirements and support sustained growth of the archive.

Upcoming developments include the creation of PDB-IHM beta archive alongside the PDB beta archive as planned by wwPDB. The beta archive will support extended PDB IDs, which will become the primary IDs when 4-character accession codes are consumed (expected by 2028).

We are also working on delivering IHM validation data in mmCIF format so that software tools can consume this information. Furthermore, we are working on streamlining data exchange with EMDDB and PRIDE data resources to access pre-release experimental data for structure validation.

Results from a preliminary analysis carried out using artificial intelligence tools relying on large language models (LLMs) estimate that roughly 25% of lower-resolution 3DEM structures (< 3.5 Å) deposited into the PDB archive as single-method experimental structures are likely to be integrative structures (we acknowledge that this metric may be an overestimate). The OpenAI GPT-3.5-turbo LLM was provided with publicly available primary publications from 2023 reporting lower-resolution 3DEM structures archived in the PDB and was tasked with identifying structures that use both a 3DEM map and other types of experimental data supported by PDB-IHM, such as crosslinking-MS. PDB structures thus identified are currently not annotated as integrative structures, and some of the data used to generate them are not archived in the PDB (*i.e.*, there has been incomplete data capture). PDB-IHM provides a robust mechanism for deposition, biocuration, validation, and archiving integrative structures together with all the supporting data and metadata details regarding the input restraints, starting models, and modeling methods used, thus enabling FAIR data delivery. In addition, the PDB-IHM pipeline facilitates enhanced structure validation based on spatiotemporal restraints and starting models used in structure determination. Thus, we encourage the structural biology community to deposit integrative structures using the PDB-IHM.org deposition tool accessible from the wwPDB OneDep system, especially 3DEM-based integrative structures. Depositors wishing to correct the scientific record and ensure complete data capture are strongly encouraged to re-deposit their 3DEM-crosslinking-MS (or similar) integrative structures using the PDB-IHM deposition pipeline.

As structure determination methods advance and the diversity of data used by integrative methods increases, PDB-IHM and RCSB.org will also continue to expand in scope, including improvements in the PDB data standards, deposition and curation workflows, validation methods, and visualization and analysis tools. These improvements will be informed and encouraged by collaborations across the structural biology community, in turn maintaining the PDB in service of structural biology as it maximizes its impact on cell biology and beyond.

Funding

RCSB PDB core operations are jointly funded by NSF (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 NIH (R01GM157729, PI: S.K. Burley). B. Vallat acknowledges funding from the United States National Science Foundation (NSF) award DBI-2112966 (PI: B. Vallat). A. Sali acknowledges funding from NSF and the United States National Institutes of Health (NIH) (NSF DBI-2112967, PI: A. Sali; NIH R01GM083960, PI: A. Sali; NIH P41GM109824, PI: M.P. Rout). B.M. Webb is a Nebius Fellow.

Acknowledgments

We thank all members of the RCSB PDB team and the wwPDB consortium for their support. We acknowledge Dr. Carl Kesselman and his team for supporting the development and maintenance of the PDB-IHM infrastructure. We are grateful to members of the wwPDB IHM Task Force for their continued support and recommendations. We thank the more than 60,000

researchers worldwide who have deposited structures to PDB and PDB-IHM, and the many millions of public 3D biostructure data consumers working and learning in nearly every country and territory recognized by the United Nations.

References

1. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res*, **28**, 235-242.
2. wwPDB consortium. (2019) Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res*, **47**, D520-D528.
3. Berman, H.M., Henrick, K. and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank. *Nat Struct Biol*, **10**, 980.
4. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, **3**, 1-9.
5. van der Aalst, W.M.P., Bichler, M. and Heinzl, A. (2017) Responsible data science. *Business & Information Systems Engineering*, **59**, 311-313.
6. Burley, S.K., Bhatt, R., Bhikadiya, C., Bi, C., Biester, A., Biswas, P., Bittrich, S., Blaumann, S., Brown, R., Chao, H. *et al.* (2025) Updated Resources for Exploring Experimental PDB Structures and Computed Structure Models at the RCSB Protein Data Bank. *Nucleic Acids Research*, **53**, D564–D574.
7. Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov, M., Nair, S., Mirdita, M., Yeo, J. *et al.* (2024) AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res*, **52**, D368-D375.
8. Tauriello, G., Waterhouse, A.M., Haas, J., Behringer, D., Bienert, S., Garello, T. and Schwede, T. (2025) ModelArchive: a deposition database for computational macromolecular structural models. *J Mol Biol*, **437**, 168996.
9. Burley, S.K., Bhikadiya, C., Bi, C., Bittrich, S., Chao, H., Chen, L., Craig, A.P., Crichlow, G.V., Dalenberg, K., Duarte, J.M. *et al.* (2023) RCSB Protein Data Bank (RCSB.org): Delivery of experimentally-determined PDB structures alongside one million Computed Structure Models of proteins from Artificial Intelligence/Machine Learning. *Nucleic Acids Res*, **51**, D488–D508.
10. Rout, M.P. and Sali, A. (2019) Principles for Integrative Structural Biology Studies. *Cell*, **177**, 1384-1403.
11. Baldwin, E.T., van Eeuwen, T., Hoyos, D., Zalevsky, A., Tchesnokov, E.P., Sanchez, R., Miller, B.D., Di Stefano, L.H., Ruiz, F.X., Hancock, M. *et al.* (2024) Structures, functions and adaptations of the human LINE-1 ORF2 protein. *Nature*, **626**, 194-206.
12. Shah, S.Z., Perry, T.N., Graziadei, A., Cecatiello, V., Kaliyappan, T., Misiaszek, A.D., Muller, C.W., Ramsay, E.P. and Vannini, A. (2025) Structural insights into distinct mechanisms of RNA polymerase II and III recruitment to snRNA promoters. *Nature communications*, **16**, 141.
13. Latham, A.P., Zhang, W., Tempkin, J.O.B., Otsuka, S., Ellenberg, J. and Sali, A. (2025) Integrative spatiotemporal modeling of biomolecular processes: Application to the assembly of the nuclear pore complex. *Proc Natl Acad Sci U S A*, **122**, e2415674122.
14. Otsuka, S., Tempkin, J.O.B., Zhang, W., Politi, A.Z., Rybina, A., Hossain, M.J., Kueblbeck, M., Callegari, A., Koch, B., Morero, N.R. *et al.* (2023) A quantitative map of nuclear pore assembly reveals two distinct mechanisms. *Nature*, **613**, 575-581.
15. Sali, A., Berman, H.M., Schwede, T., Trewhella, J., Kleywegt, G., Burley, S.K., Markley, J., Nakamura, H., Adams, P., Bonvin, A.M. *et al.* (2015) Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop. *Structure*, **23**, 1156-1167.

16. 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. *et al.* (2019) Federating Structural Models and Data: Outcomes from A Workshop on Archiving Integrative Structures. *Structure*, **27**, 1745-1759.
17. Vallat, B., Webb, B., Westbrook, J.D., Sali, A. and Berman, H.M. (2018) Development of a Prototype System for Archiving Integrative/Hybrid Structure Models of Biological Macromolecules. *Structure*, **26**, 894-904 e892.
18. Vallat, B., Webb, B., Fayazi, M., Voinea, S., Tangmunarunkit, H., Ganesan, S.J., Lawson, C.L., Westbrook, J.D., Kesselman, C., Sali, A. *et al.* (2021) New system for archiving integrative structures. *Acta crystallographica. Section D, Structural biology*, **77**, 1486-1496.
19. Young, J.Y., Westbrook, J.D., Feng, Z., Sala, R., Peisach, E., Oldfield, T.J., Sen, S., Gutmanas, A., Armstrong, D.R., Berrisford, J.M. *et al.* (2017) OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the PDB Archive. *Structure*, **25**, 536-545.
20. Vallat, B., Webb, B.M., Zalevsky, A., Tangmunarunkit, H., Sekharan, M.R., Voinea, S., Shafaeibejestan, A., Sagendorf, J., Hoch, J.C., Kurisu, G. *et al.* (2025) PDB-IHM: A System for Deposition, Curation, Validation, and Dissemination of Integrative Structures. *J Mol Biol*, **437**, 168963.
21. Vallat, B., Webb, B.M., Westbrook, J.D., Goddard, T.D., Hanke, C.A., Graziadei, A., Peisach, E., Zalevsky, A., Sagendorf, J., Tangmunarunkit, H. *et al.* (2024) IHMCIF: An Extension of the PDBx/mmCIF Data Standard for Integrative Structure Determination Methods. *J Mol Biol*, **436**, 168546.
22. Fitzgerald, P.M.D., Westbrook, J.D., Bourne, P.E., McMahon, B., Watenpaugh, K.D. and Berman, H.M. (2005) In Hall, S. R. and McMahon, B. (eds.), *International Tables for Crystallography G. Definition and exchange of crystallographic data*. Springer, Dordrecht, The Netherlands, pp. 295-443.
23. Westbrook, J.D., Young, J.Y., Shao, C., Feng, Z., Guranovic, V., Lawson, C., Vallat, B., Adams, P.D., Berrisford, J.M., Bricogne, G. *et al.* (2022) PDBx/mmCIF Ecosystem: Foundational semantic tools for structural biology. *J Mol Biol*, **434**, 167599.
24. Bugacov, A., Czajkowski, K., Kesselman, C., Kumar, A., Schuler, R.E. and Tangmunarunkit, H. (2017) Experiences with Deriva: An Asset Management Platform for Accelerating eScience. *Proc IEEE Int Conf Escience*, **2017**, 79-88.
25. Schuler, R.E., Kesselman, C. and Czajkowski, K. (2016), *2016 IEEE 12th International Conference on e-Science (e-Science)*. IEEE, pp. 31-40.
26. Young, J.Y., Westbrook, J.D., Feng, Z., Peisach, E., Persikova, I., Sala, R., Sen, S., Berrisford, J.M., Swaminathan, G.J., Oldfield, T.J. *et al.* (2018) Worldwide Protein Data Bank biocuration supporting open access to high-quality 3D structural biology data. *Database*, **2018**, bay002.
27. Bittrich, S., Bi, C., Bhikadiya, C., Chao, H., Duarte, J.M., Dutta, S., Fayazi, M., Henry, J., Khokhriakov, I., Lowe, R. *et al.* (2023) RCSB Protein Data Bank: Efficient searching and simultaneous access to one million Computed Structure Models alongside the PDB structures enabled by architectural advances. *J Mol Biol*, **435**, 167994.
28. Rose, Y., Duarte, J.M., Lowe, R., Segura, J., Bi, C., Bhikadiya, C., Chen, L., Rose, A.S., Bittrich, S., Burley, S.K. *et al.* (2021) RCSB Protein Data Bank: Architectural advances towards integrated searching and efficient access to macromolecular structure data from the PDB archive. *J Mol Biol*, **433**, 166704.
29. Vallat, B., Tauriello, G., Bienert, S., Haas, J., Webb, B.M., Zidek, A., Zheng, W., Peisach, E., Piehl, D.W., Anischanka, I. *et al.* (2023) ModelCIF: An extension of PDBx/mmCIF data representation for Computed Structure Models. *J Mol Biol*, **435**, 168021.

30. Sehna, D., Bittrich, S., Velankar, S., Koca, J., Svobodova, R., Burley, S.K. and Rose, A.S. (2020) BinaryCIF and CIFTools-Lightweight, efficient and extensible macromolecular data management. *PLoS Comput Biol*, **16**, e1008247.
31. Hoch, J.C., Baskaran, K., Burr, H., Chin, J., Eghbalnia, H.R., Fujiwara, T., Gryk, M.R., Iwata, T., Kojima, C., Kurisu, G. *et al.* (2023) Biological Magnetic Resonance Data Bank. *Nucleic Acids Res*, **51**, D368-D376.
32. ww, P.D.B.C. (2024) EMDB-the Electron Microscopy Data Bank. *Nucleic Acids Res*, **52**, D456-D465.
33. Kikhney, A.G., Borges, C.R., Molodenskiy, D.S., Jeffries, C.M. and Svergun, D.I. (2020) SASBDB: Towards an automatically curated and validated repository for biological scattering data. *Protein Sci*, **29**, 66-75.
34. Perez-Riverol, Y., Bandla, C., Kundu, D.J., Kamatchinathan, S., Bai, J., Hewapathirana, S., John, N.S., Prakash, A., Walzer, M., Wang, S. *et al.* (2025) The PRIDE database at 20 years: 2025 update. *Nucleic Acids Res*, **53**, D543-D553.
35. Sehna, D., Bittrich, S., Deshpande, M., Svobodova, R., Berka, K., Bazgier, V., Velankar, S., Burley, S.K., Koca, J. and Rose, A.S. (2021) Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res*, **49**, W431–W437.
36. Erzberger, J.P., Stengel, F., Pellarin, R., Zhang, S., Schaefer, T., Aylett, C.H.S., Cimerancic, P., Boehringer, D., Sali, A., Aebersold, R. *et al.* (2014) Molecular architecture of the 40S·eIF1·eIF3 translation initiation complex. *Cell*, **158**, 1123-1135.
37. Thamkachy, R., Medina-Pritchard, B., Park, S.H., Chiodi, C.G., Zou, J., de la Torre-Barranco, M., Shimanaka, K., Abad, M.A., Gallego Paramo, C., Feederle, R. *et al.* (2024) Structural basis for Mis18 complex assembly and its implications for centromere maintenance. *EMBO Rep*, **25**, 3348-3372.
38. Shi, Y., Fernandez-Martinez, J., Tjioe, E., Pellarin, R., Kim, S.J., Williams, R., Schneidman-Duhovny, D., Sali, A., Rout, M.P. and Chait, B.T. (2014) Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol Cell Proteomics*, **13**, 2927-2943.
39. Segura, J., Rose, Y., Bi, C., Duarte, J., Burley, S.K. and Bittrich, S. (2023) RCSB Protein Data Bank: visualizing groups of experimentally determined PDB structures alongside computed structure models of proteins. *Front Bioinform*, **3**, 1311287.
40. UniProt Consortium. (2023) UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res*, **51**, D523–D531.
41. Bittrich, S., Segura, J., Duarte, J.M., Burley, S.K. and Rose, Y. (2024) RCSB Protein Data Bank: exploring protein 3D similarities via comprehensive structural alignments. *Bioinformatics*, **40**, btae370.