Nucleic Acids Research, 2024, **52**, D456–D465
https://doi.org/10.1093/nar/gkad1019
Advance access publication date: 22 November 2023
Database issue

EMDB—the Electron Microscopy Data Bank

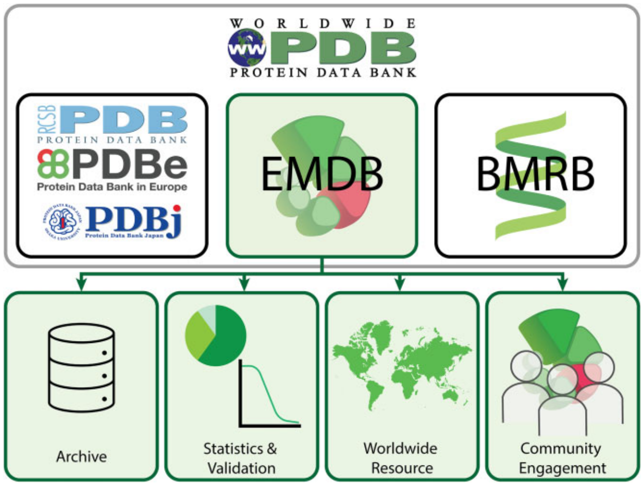
The wwPDB Consortium[*,†]

[*] To whom correspondence should be addressed. Tel: + 44 1223 494 444; Email: kyle@ebi.ac.uk
Correspondence may be also addressed to Ardan Patwardhan. Tel: + 44 1223 494 444; Email: ardan@ebi.ac.uk
[†] Authors details present in appendix.

Abstract

The Electron Microscopy Data Bank (EMDB) is the global public archive of three-dimensional electron microscopy (3DEM) maps of biological specimens derived from transmission electron microscopy experiments. As of 2021, EMDB is managed by the Worldwide Protein Data Bank consortium (wwPDB; wwpdb.org) as a wwPDB Core Archive, and the EMDB team is a core member of the consortium. Today, EMDB houses over 30 000 entries with maps containing macromolecules, complexes, viruses, organelles and cells. Herein, we provide an overview of the rapidly growing EMDB archive, including its current holdings, recent updates, and future plans.

Graphical abstract


WORLDWIDE PDB PROTEIN DATA BANK
RCSB PDB PROTEIN DATA BANK
PDBe Protein Data Bank in Europe
PDBj
EMDB
BMRB
Archive
Statistics & Validation
Worldwide Resource
Community Engagement

Introduction

In the past decade, 3DEM techniques including cryogenic-sample electron microscopy and tomography (cryoEM and cryoET, respectively) (1), have become leading structure-determination techniques in the field of structural biology (2–4). Since the advent of the 'Resolution Revolution' (5) 3DEM is routinely being used to study biological structures on scales from atoms (6,7) to molecules in cells (8,9), to whole cells (e.g. EMD-11073 (10)). In addition, 3DEM has proved valuable for the study of sample heterogeneity, as is well reviewed in (11), and new software is pushing the study of protein heterogeneity even further (12–16).

In 2002, the Macromolecular Structure Database group (from which resources such as Electron Microscopy Data Bank (EMDB) (17), Electron Microscopy Public Image Archive (EMPIAR) (18) and the Protein Data Bank in Europe (PDBe) (19) have emerged) established the EMDB for the archiving and dissemination of 3DEM volumes (20). EMDB is located at the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) in Hinxton, UK. In 2007, the Research Collaboratory for Structural Bioinformatics (RCSB) (21) and the National Center for Macromolecular Imaging (NCMI) (22) joined forces with the EMDB. The Protein Data Bank Japan (PDBj) (23) joined this collaboration in 2013. In 2021, the EMDB archive became a core wwPDB archive alongside the PDB (24) and BMRB (25), with the EMDB team at EMBL-EBI becoming a core wwPDB member and serving as the wwPDB-designated Archive Keeper for the EMDB core archive. Protein Data Bank China (PDBc) recently joined the wwPDB as an associate member (26), further expanding the international collaboration managing these archives. wwPDB core archives are made available at no charge and with no limitation on usage under the CC0 1.0 Creative Commons licence (https://creativecommons.org/share-your-work/public-domain/cc0/). All three core archives managed jointly by the wwPDB operate under the FAIR principles of Findability, Accessibility, Interoperability and Reusability (27). Following the FAIR principles facilitates

Received: September 28, 2023. **Revised:** October 18, 2023. **Editorial Decision:** October 19, 2023. **Accepted:** October 20, 2023

data discovery, collaboration, and data reproducibility—all of which are important for accelerating research and innovation.

Interoperation between archives is essential for serving the complete data associated with a 3DEM experiment. The 3DEM Coulomb potential map (henceforth referred to as 'map' or 'volume') is stored in the EMDB archive, whereas derived atomic coordinate structures are stored in the PDB archive. The raw data from which maps are derived are collected by the Electron Microscopy Public Image Archive (EMPIAR) (18). In addition to the aforementioned structural archives, entries in EMDB can also be associated with entries in the AlphaFold Protein Structure Database (28), Small Angle Scattering Biological Data Bank (SASBDB) (29), PDB-Dev (30) and EMPIAR (18).

The wwPDB members support its core mission of sustaining freely accessible, interoperating Core Archives of structure data and metadata for biological macromolecules as an enduring public good to promote basic and applied research and education across the sciences (24). Here, we provide an overview of the recent developments in the EMDB archive including its content, deposition protocols, use and future prospects.

Archive content

Growth and statistics

EMDB archives volumes from single-particle analysis (SPA), subtomogram averaging (STA), helical reconstruction (HR), tomography, and electron crystallography (EC). All these techniques produce 3D volume data with the exception of EC, which produces diffraction data from which such volumes can be calculated.

EMDB holds >30 000 entries as of 4 October 2023, approximately 55% of which have associated atomic coordinates archived in the PDB (Figure 1A). The number of 3DEM entries released per year is growing exponentially (Figure 1b). If this growth continues EMDB is predicted to hold 50 000 entries in 2025 and 100 000 entries in 2028 (at the current archive-doubling time of ∼2.5 years), with ∼13 500 and ∼31 500 releases predicted in these years, respectively. Based on the current trends it is expected that the number of PDB 3DEM entries released per year will surpass PDB Macromolecular Crystallography or MX releases in 2025. The rapid growth of 3DEM entries highlights the increased accessibility and utility of 3DEM methods.

The fraction of entries in the EMDB archive determined by each modality is shown in Figure 2A. SPA is by far the most popular method used to generate volumes that are deposited to the archive, making up 82.8% of the total archive at the end of 2022, an increase of >4% since 2016 (17). STA, Tomography, HR, and EC make up the remaining 17.2% in descending order. Using metadata harvested from EMDB, the effect of the resolution revolution (5) can be visualised (Figure 2B). The biological insight being sought by the researcher will ultimately dictate the resolution that is required from the 3DEM volume obtained from the investigation. Macromolecular structure, interactions and function may be sufficiently described by 'lower resolution' volumes. In 2022 alone, 307 SPA maps equal to or below 10 Å resolution were released, each providing novel insights to the scientific community. Where the goal of the investigation is to model a macromolecular structure with atomic accuracy, a high resolution 3DEM reconstruction is sought. More than 60% of entries in 2022 were at better than 4 Å resolution (4034 out of 6139) and over 20% of entries from the same year were sub-3 Å (1299 out of 6139). At present, the archive also contains structures with atomic resolvability (<1.2 Å, as per the Sheldrick criterion (31,32)), including 2 SPA and 28 EC entries.

Archive data

Each EMDB entry describes a macromolecular complex or subcellular structure represented in a 3DEM volume. In certain cases, multiple entries may be linked to fully represent the potential multiple volumes output from a 3DEM experiment. All entries contain a primary 3DEM volume around which an entry is based. Unfiltered, unmasked and unsharpened half-maps, generated by default in most SPA, STA and HR workflows, have been mandatory for all relevant modalities deposited since 25 February 2022 (https://www.wwpdb.org/news/news?year=2022#6218da3152988f064bf8c4a3). Each entry contains an image provided by the depositor, enabling them to showcase the primary volume according to their preferences. A number of optional additional files may also be present in an entry, including additional volumes, masks, a Fourier Shell Correlation (FSC) curve, and layer-line files. Table 1 gives an overview of all file types that may be present in an entry.

An entry's primary volume must be adequately described with additional metadata to ensure adherence to FAIR principles (27). These metadata are stored in Extensible Markup Language (XML) and mmCIF files with items and attributes defined in the EMDB data model (www.ebi.ac.uk/emdb/documentation#version30) and mmCIF dictionary (mmcif.wwpdb.org), respectively. These definitions include a hierarchy of information which allows description of cellular structures, supramolecules, and macromolecules, all of which may exist in a single volume. Various other types of experimental metadata are also present in these files, including information on specimen preparation, microscopy instrumentation, data-collection protocol, and software used during volume generation and processing. The EMDB data model includes definitions of mandatory items and support for enumerations and allowed data ranges. All metadata files in the EMDB archive can, therefore, be validated against the XML schema.

Global data deposition

Depositions to EMDB are managed by the wwPDB global OneDep deposition (33), validation (34–36), and biocuration software system (37). OneDep is a unified deposition system for 3DEM, X-ray, and NMR 3D biostructures, experimental data and related metadata. OneDep is hosted at wwPDB data centres around the world (located in the USA, Europe and Asia) providing a consistent deposition experience independent of geographic location. Biocuration of incoming entries is geographically distributed among wwPDB partner sites as follows: RCSB PDB handles all depositions from the Americas and Oceania, PDBe and the EMDB team at EMBL-EBI manage depositions from Europe and Africa, and PDBj and PDBc handle all depositions from Asia. This arrangement divides the effort of biocuration and ensures that depositors can communicate with wwPDB biocurators within, or close to, their local timezone.

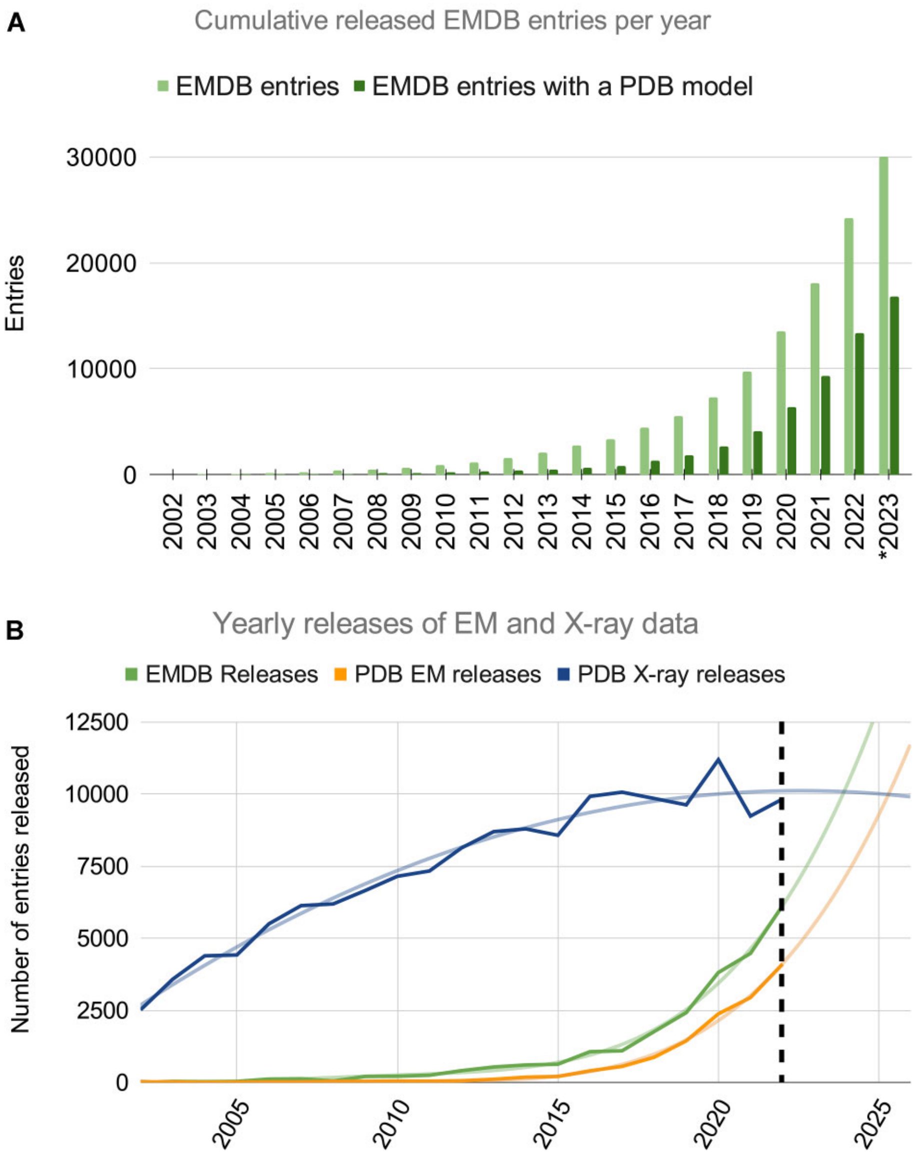


Figure 1. (**A**) Cumulative number of all EMDB entries and of those with an atomic coordinate structure in the PDB at the end of each year. *2023 data are through October 4th. (**B**) Number of annually released EMDB and PDB entries. EMDB data includes all modalities. PDB data is split according to X-ray and EM models released per year. Data is shown until the end of 2022 (dashed black line) with trendlines showing predicted future growth.

The OneDep system makes use of the PDBx/mmCIF framework during deposition, validation, and biocuration. All data and metadata related to a deposition are defined in the EMDB data model, which in turn informs the PDBx/mmCIF dictionary (mmcif.wwpdb.org) (38). Thus for 3DEM, the PDBx/mmCIF dictionary faithfully represents the EMDB model described in the EMDB XML schema. Data specific to 3DEM methods in the PDBx/mmCIF dictionary use the 'em' namespace (*e.g.* 'em_imaging' for metadata on the electron microscope setup (mmcif.wwpdb.org/dictionaries/mmcif_pdbx_v50.dic/Categories/em_imaging.html)). The dictionary definitions can also include rules such as relationships between different data items, enumerations, and allowed ranges. Finally, the PDBx/mmCIF format is extensible, allowing the dictionary to grow with the data models for all three wwPDB core archives (PDB, EMDB and BMRB).

At the end of the deposition, validation and biocuration processes, a wwPDB validation report for reviewers is provided to the depositor. This report contains a range of community-recommended validation metrics for both the volume and, if present, the atomic coordinates (34,36,39). It is strongly recommended by wwPDB that depositors provide their confidential post-biocuration wwPDB validation reports in PDF format to scientific journals when submitting related manuscripts (24). Validation metrics related to 3DEM volumes are implemented in a tiered approach (36), allowing testing and gathering of community feedback by presenting the metrics on the EMDB website (tiers 1 and 2) before potential implementation in the wwPDB validation reports (tier 3). *Q*-score (40) is the most recently added tier-3 metric, providing a new metric which complements the previously implemented atom-inclusion score (41).

Upon depositor request, manuscript publication (including on a preprint server), or when one year has elapsed since

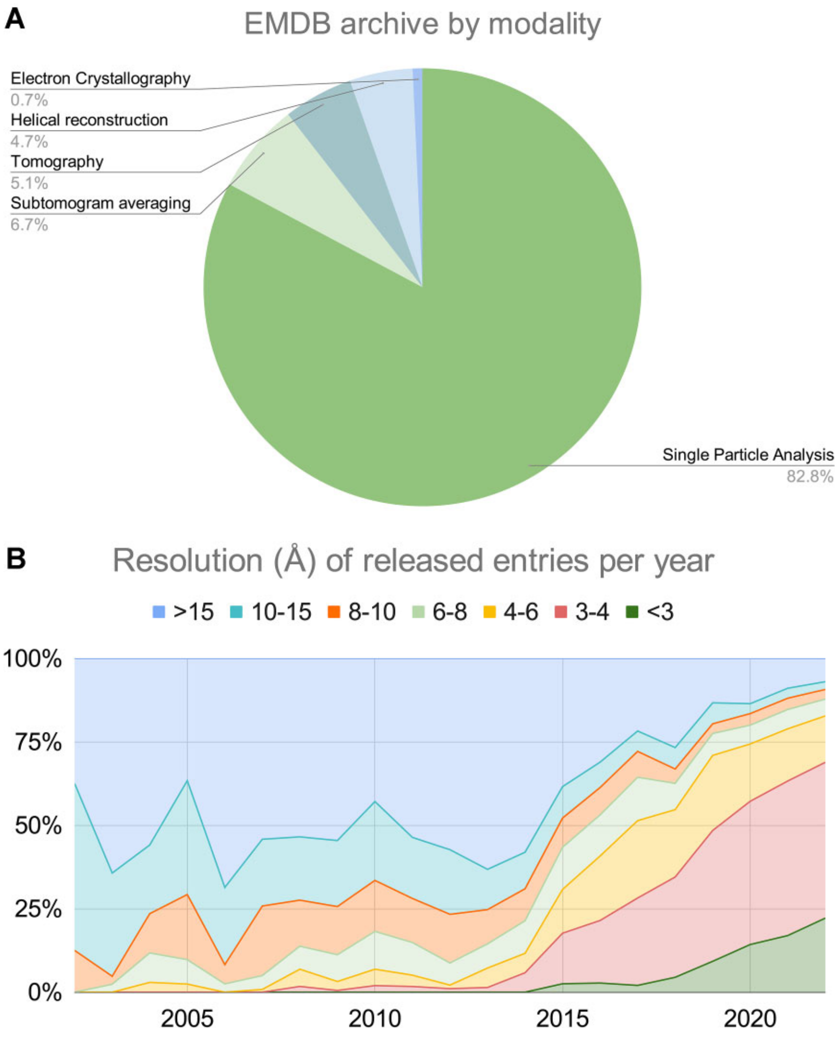


Figure 2. (**A**) Proportion of the EMDB archive contributed by each modality as of the end of 2022. (**B**) Proportion of released entries in various resolution (Å) shells per year from 2002 to 2022. Note the huge impact of the resolution revolution since 2014.

Table 1. An overview of possible entry content held in the EMDB archive

Category (number)	Description
Primary map (1)	EM map or tomogram that is described in the associated publication
Half-maps (0 or 2)[a]	Unfiltered, unsharpened, and unmasked raw half-maps for SPA, SPA based helical reconstructions, or STA
Masks (0 or more)	Primary/raw map masks, segmentation/focused-refinement masks, and half-map masks
Additional cryo-EM maps (0 or more)	Examples include difference maps, maps showing alternative conformational states and/or compositions, and maps with different processing (for example filtering, sharpening, and masking)
Auxiliary files (0 or more)	Examples include author-determined FSC curves (half-maps, map–model, …), structure factors, and layer lines
Metadata files (2)	XML and mmCIF files containing an entry's metadata
Validation Report (1)	Volume-only validation reports (3 file formats).

[a] Deposition of half-maps was made mandatory in February of 2022.

the date of deposition, an entry is prepared for public release in the next weekly EMDB update cycle. Entries can be prepared for release from Monday to Thursday in any given week. Release of atomic coordinates cannot precede release of the associated EMDB entry. Every Thursday, after all entries are ready for release, the EMDB archive begins the release process, which includes data-integrity checks, running of automated validation and data-enrichment pipelines, and manual inspection of validation outputs for all entries. Depositors are contacted in the event of any peculiarities noticed during final validation checks. New data are then made public every Wednesday at 00:00 Universal Time Coordinated or UTC.

Data dissemination

The EMDB archive is served *via* an FTP site (https://ftp.ebi.ac.uk/pub/databases/emdb/), which is mirrored by wwPDB (https://files.wwpdb.org/pub/emdb), RCSB PDB (https://files.rcsb.org/pub/emdb) and PDBj (https://files.pdbj.org/pub/emdb/). The EMDB FTP site provides access to volumes, entry images, FSC data, metadata, and validation files for released entries. The EMDB FTP site also serves files describing the EMDB data model and the status of all released and unreleased 3DEM entries in the '/doc/' and '/status/' folders of the EMDB FTP area, respectively. The number of unique IP addresses from which downloads are initiated in a given calendar month and total data downloaded are shown in Figure 3A. In 2022, total volume of downloaded data exceeded 24 TB (the total archive size at the end of 2022 was 8.9 TB), a 33% increase on the previous year. When assessing total downloads of EMDB data across all wwPDB FTP sites (Figure 3B) a seasonal variation is observed. In 2020 and 2021, most downloads occurred during Q4, whereas in 2022 download requests peaked slightly earlier, in Q3.

Data, news, and statistics relevant to EMDB can be found on the EMDB website (ebi.ac.uk/emdb/). News and statistics pertaining to the wider wwPDB, including EMDB, are available on the wwPDB website (wwpdb.org). Volume data is served in CCP4/MRC map format (ftp.ebi.ac.uk/pub/databases/emdb/doc/Map-format/). Metadata is provided in XML and mmCIF format, with the EMDB data model described in the 'docs/' section of the EMDB FTP site.

Archive updates

A major change to the EMDB archive involved the transition to a new version of the EMDB data model, which underpins the data structure of the archive. The new data model, fully implemented in 2021, contains a rich set of metadata and supports hierarchical description of the sample composition. Descriptions of both molecular and cellular samples are supported and metadata has been modularised such that different 3DEM modalities have specific data items. Taken together these features facilitate a more accurate and complete description of 3DEM entries in the EMDB archive relative to the previous version of the data model. The schema and documentation for the data model are available online (https://www.ebi.ac.uk/emdb/documentation).

When extending the EMDB data model, archive remediation is required to add new information to legacy entries, where possible. One recent addition to the data model is the option to link EMDB entries to SASBDB entries. These data are now available for newly released entries and a remediation will be carried out to add this information for relevant legacy entries. When major changes are made to the archive or policies, they are communicated via the EMDB website and wwPDB channels, including emails to major mailing lists (e.g. CCPEM and 3DEM).

Community engagement

Community engagement is at the heart of archive management, such as annotation practices, archive policies and future development plans. To this end, wwPDB is guided by an expert international advisory committee. The committee advises on the activities of the wwPDB partners and outcomes of its annual meetings are published on the wwPDB website (http://www.wwpdb.org/about/advisory).

In addition to the annual advisory committee meetings, several workshops specific to the planning of EMDB's future endeavours have been conducted. Community experts were engaged for the purposes of ensuring optimal data management (42), annotation of cellular data (43) and ensuring integration of data between multiple resources (44). The validation methods in 3DEM are constantly evolving, and no one metric can provide an overall description of volume quality. An expert Validation Task Force was set up and first met in 2010 (39), with a second meeting held in 2020. This group contributes and advises on the validation metrics EMDB includes in its Validation Analysis resource (36) and the wwPDB validation reports (33).

The EMDB team is committed to training new scientists in EMDB and OneDep, and communicating science via a number of outreach channels. EMDB staff members have given lectures at various training courses around the world. They also run an X (previously Twitter) and YouTube channel to highlight recently released entries and share EMDB news.

Current trends and future outlook

Rapid growth

The EMDB archive continues to experience exponential growth year-on-year (Figure 1), while at the same time depositors are creating more archive entries per publication (Figure 4). The wwPDB partners are addressing these developments through improvements to the OneDep deposition, validation and biocuration system. Biocuration and validation workflows are under continuous review, to inform, plan, design and implement improvements such as increased automation. Depositor experience and efficiency are also important considerations. The ORCID-based login feature allows depositors to access all their depositions within their browser, rather than having to keep a logbook of all deposition IDs. A deposition API is also in development, which will allow software packages to directly deposit data to the OneDep system. In the most favourable cases, this could entirely obviate the need for depositor intervention, bringing a level of efficiency and convenience to the deposition process that is necessary to accommodate the growth of 3DEM volume generation by the community. Programmatic collection and deposition of data also represents a method for improving metadata accuracy and completeness. wwPDB is currently working with developers from popular structural biology software packages to accelerate the adoption of the deposition API.

Rapidly evolving science

Every 3DEM modality supported by EMDB is experiencing continuous and rapid innovation and improvement. Single particle-based methods (SPA and STA) can now be used to analyse continuous conformational variability of a macromolecule, sometimes referred to as heterogeneity analysis (12–15). These approaches allow many structural states to be resolved using a single imaging dataset. Simultaneously, tomography is seeing explosive growth in the amount of data that can be recorded per unit time (45–47). Such advances require frequent updates to the EMDB data model and PDBx/mmCIF dictionary. wwPDB plans to establish a wwPDB 3DEM working group to accelerate implementation of such additions and promote adoption of data and methodological standards across the community.

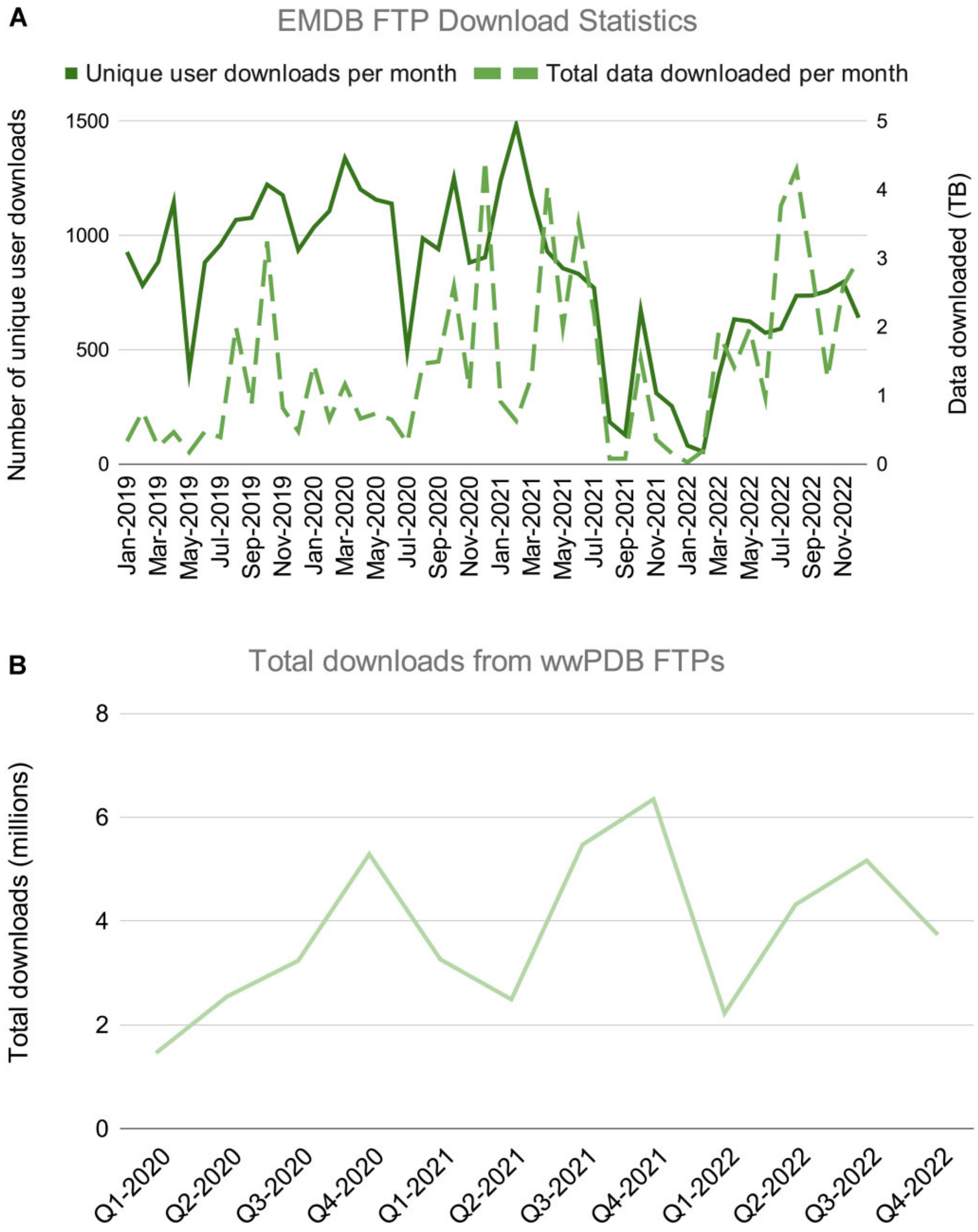


Figure 3. (**A**) Monthly counts of the number of downloads by unique users each month (solid line, left axis) and the total amount of data served (dashed line, right axis). (**B**) Quarterly counts of the total number of downloads of EMDB data from all wwPDB FTP sites.

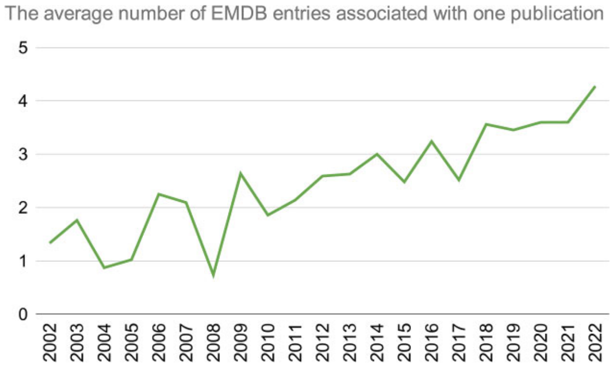


Figure 4. The average number of EMDB entries associated with a single publication per year.

Software development in the field of artificial intelligence and machine learning (AI/ML) has been nothing short of spectacular in recent years. EMDB data have already been used in the development of new AI-based software tools for various applications, including particle picking (48), particle pruning (49), volume sharpening (50), local-resolution estimation (51), secondary-structure detection (52), residue-level quality estimation (53) and model building (54,55). EMDB endeavours to continue supporting the development of new ML algorithms through expert curation, and thorough and accurate labelling of entries *via* the EMDB data model.

Looking forward

As the 3DEM field continues to expand and evolve, the importance of archiving and appropriately describing new types of data will continue to be essential. To this end, EMDB plans to continue to expand the 3DEM data model in collaboration with the community and various wwPDB working groups. Examples of enhancements include provenance description of synthesised macromolecules and labelling of composite maps. Furthermore, as 3DEM experiments are expected to be performed downstream of increasingly heterogeneous preparation methods and/or preceding multimodal imaging experiments, it will be important to be able to capture metadata describing more complex origins of a final 3DEM volume. Planned improvements of metadata describing 3DEM experiments will both support and enhance interoperability of archived data. As an example, metadata is now available in mmCIF format for newly released entries and we plan to expand this to the whole archive, in addition to the existing XML files.

The EMDB Validation Analysis package (36) uses several software tools to generate an extensive set of validation statistics, plots and images, which enable assessment of various volume and map-model features. Comprehensive metrics are important, but the sheer volume of metrics can be overwhelming to some users, particularly those less familiar with 3DEM techniques. Furthermore, calculating an ever-expanding set of metrics for every entry in the exponentially growing archive is a significant computational burden. Future thought will need to be aimed at summarising important results from the Validation Analysis pipeline, whilst limiting the computational and environmental costs of running the software. Summarising results is likely to take a similar form to the 'sliders' already presented within wwPDB validation reports for atomic models deposited to the PDB archive (56).

Validation of deposited metadata is essential for ensuring that the archive retains its fidelity and value in the long term. wwPDB has implemented a number of 3DEM-related checks within the OneDep system to improve data reliability. For example, minimum defocus values can no longer be greater than maximum defocus values, and provided pixel sizes of volumes are compared with those reported in the volume header. Additionally, deposition of atomic coordinates that include parts of structures extending beyond the deposited volume's bounding box is now stopped at the file-upload stage. Several multi-file checks are in development, including assessment of whether uploaded half-maps are identical, or whether the primary map and half-maps are offset from one another or have different grid parameters. In rare cases, depositors accidentally select the wrong modality at the start of the deposition process, this results in the instantiation of metadata items designed for a different modality. To mitigate this issue, EMDB is experimenting with deep-learning approaches to predict the experimental method from which a volume was derived. Taken together, these checks will help to ensure that data and metadata archived in EMDB are as accurate as possible and adhere to the FAIR principles.

Finally, developing the EMDB archive to accommodate 3DEM volumes with metadata captured accurately, automatically and robustly validated, with increased description of a dataset's experimental context, will enable the archive to support the field going forward. We expect to support new technologies that generate 3DEM volumes in new ways and to more deeply integrate these data with other structural and non-structural archives. Together, these improvements will support insight into macromolecules in their experimental and biological context, maintaining an archive of high-value to the 3DEM, data science and machine learning communities.

Data Availability

The data underlying Figure 1 are available in the EMDB* and PDB** archives. The data underlying Figures 2 and 4 are available in the EMDB* archive. The data underlying Figure 3 will be shared on reasonable request to the corresponding author. * accessible at 'https://ebi.ac.uk/emdb'. ** accessible at 'https://www.rcsb.org/stats'.

Acknowledgements

We wish to thank the scientific community, including depositors, users and advisors for their longstanding support of EMDB. We would like to thank the very many members of the 3DEM community who have enthusiastically deposited data, participated in workshops and engaged with EMDB at many levels, in particular providing help, advice and feedback. We finally thank the many developers who have contributed software and validation tools for the OneDep pipeline.

Funding

EMDB is supported by the European Molecular Biology Laboratory-European Bioinformatics Institute; EMBL IT Support is acknowledged for provision of computer and data storage servers; EMDB is further supported by funding from the Wellcome Trust [212977/Z/18/Z]; the Protein Data Bank in Europe is supported by European Molecular Biology Laboratory-European Bioinformatics Institute; Wellcome Trust [218303/Z/19/Z, 104948/Z/14/Z]; Biotechnology and Biological Sciences Research Council [BB/P025846/1, BB/V004247/1, 20-BBSRC/NSF-BIO]; RCSB PDB core operations are jointly funded by the National Science Foundation [DBI-1832184, PI: S.K. Burley]; US Department of Energy [DESC0019749, PI: S.K. Burley]; National Cancer Institute, the National Institute of Allergy and Infectious Diseases;National Institute of General Medical Sciences of the National Institutes of Health [R01GM133198, PI: S.K. Burley]; other funding awards to RCSB PDB by the NSF and to PDBe by the UK Biotechnology and Biological Research Council are jointly supporting development of a Next Generation PDB archive [DBI-2019297, PI: S.K. Burley; BB/V004247/1, PI: S. Velankar]; new Mol* features [DBI-2129634, PI: S.K. Burley; BB/W017970/1, PI: S. Velankar]; the Protein Data Bank Japan is supported by the the Database Integration Coordination Program from the Department of NBDC program, Japan Science and Technology Agency [JPMJND2205, PI: Genji Kurisu]; Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research; BINDS) from AMED [23ama121001, PI: Genji Kurisu]; joint usage program of Institute for Protein Research, Osaka University; the BMRB is supported by the U.S. National Institutes of Health [R01GM109046]; PDBc is supported by the Shanghai Advanced Research Institute (SARI), Chinese Academy of Sciences and the ShanghaiTech University; staff in India are supported by grants from the Ministry of Electronics and Information Technology, GOI [MeitY project no: CORP:DG:3191] under the National Supercomputing Mission program and Indian National Science Academy under its Honorary Scientist Fellowship scheme. Funding for open access charge: European Molecular Biology Laboratories Core Funding.

Conflict of interest statement

None declared.

References

1. Henderson,R. and Hasnain,S. (2023) 'Cryo-EM': electron cryomicroscopy, cryo electron microscopy or something else? *IUCrJ*, **10**, 519–520.
2. Nogales,E. (2016) The development of cryo-EM into a mainstream structural biology technique. *Nat. Methods*, **13**, 24–27.
3. Liedtke,J., Depelteau,J.S. and Briegel,A. (2022) How advances in cryo-electron tomography have contributed to our current view of bacterial cell biology. *J. Struct. Biol. X*, **6**, 100065.
4. Shi,D., Nannenga,B.L., Iadanza,M.G. and Gonen,T. (2013) Three-dimensional electron crystallography of protein microcrystals. *eLife*, **2**, e01345.
5. Kühlbrandt,W. (2014) The resolution revolution. *Science*, **343**, 1443–1444.
6. Yip,K.M., Fischer,N., Paknia,E., Chari,A. and Stark,H. (2020) Atomic-resolution protein structure determination by cryo-EM. *Nature*, **587**, 157–161.
7. Nakane,T., Kotecha,A., Sente,A., McMullan,G., Masiulis,S., Brown,P.M.G.E., Grigoras,I.T., Malinauskaite,L., Malinauskas,T., Miehling,J., *et al.* (2020) Single-particle cryo-EM at atomic resolution. *Nature*, **587**, 152–156.
8. Zhu,Y., Koo,C.W., Cassidy,C.K., Spink,M.C., Ni,T., Zanetti-Domingues,L.C., Bateman,B., Martin-Fernandez,M.L., Shen,J., Sheng,Y., *et al.* (2022) Structure and activity of particulate methane monooxygenase arrays in methanotrophs. *Nat. Commun.*, **13**, 5221.
9. O'Reilly,F.J., Xue,L., Graziadei,A., Sinn,L., Lenz,S., Tegunov,D., Blötz,C., Singh,N., Hagen,W.J.H., Cramer,P., *et al.* (2020) In-cell architecture of an actively transcribing-translating expressome. *Science*, **369**, 554–557.
10. Mahajan,M., Seeger,C., Yee,B. and Andersson,S.G.E. (2020) Evolutionary remodeling of the cell envelope in bacteria of the Planctomycetes phylum. *Genome Biol. Evol.*, **12**, 1528–1548.
11. Toader,B., Sigworth,F.J. and Lederman,R.R. (2023) Methods for cryo-EM single particle reconstruction of macromolecules having continuous heterogeneity. *J. Mol. Biol.*, **435**, 168020.
12. Zhong,E.D., Bepler,T., Berger,B. and Davis,J.H. (2021) CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods*, **18**, 176–185.
13. Punjani,A. and Fleet,D.J. (2023) 3DFlex: determining structure and motion of flexible proteins from cryo-EM. *Nat. Methods*, **20**, 860–870.
14. Harastani,M., Eltsov,M., Leforestier,A. and Jonic,S. (2021) HEMNMA-3D: cryo electron tomography method based on normal mode analysis to study continuous conformational variability of macromolecular complexes. *Front. Mol. Biosci.*, **8**, 663121.
15. Powell,B.M. and Davis,J.H. (2023) Learning structural heterogeneity from cryo-electron sub-tomograms with tomoDRGN. bioRxiv doi: https://doi.org/10.1101/2023.05.31.542975, 02 June 2023, preprint: not peer reviewed.
16. Nakane,T., Kimanius,D., Lindahl,E. and Scheres,S.H. (2018) Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife*, **7**, e36861.
17. Lawson,C.L., Patwardhan,A., Baker,M.L., Hryc,C., Garcia,E.S., Hudson,B.P., Lagerstedt,I., Ludtke,S.J., Pintilie,G., Sala,R., *et al.* (2016) EMDataBank unified data resource for 3DEM. *Nucleic Acids Res.*, **44**, D396–D403.
18. Iudin,A., Korir,P.K., Somasundharam,S., Weyand,S., Cattavitello,C., Fonseca,N., Salih,O., Kleywegt,G.J. and Patwardhan,A. (2023) EMPIAR: the electron microscopy public image archive. *Nucleic Acids Res.*, **51**, D1503–D1511.
19. Gutmanas,A., Alhroub,Y., Battle,G.M., Berrisford,J.M., Bochet,E., Conroy,M.J., Dana,J.M., Fernandez Montecelo,M.A., Van Ginkel,G., Gore,S.P., *et al.* (2014) PDBe: protein Data Bank in Europe. *Nucleic Acids Res.*, **42**, D285–D291.
20. Tagari,M., Newman,R., Chagoyen,M., Carazo,J.-M. and Henrick,K. (2002) New electron microscopy database and deposition system. *Trends Biochem. Sci.*, **27**, 589.
21. Burley,S.K., Bhikadiya,C., Bi,C., Bittrich,S., Chao,H., Chen,L., Craig,P.A., 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.
22. Lawson,C.L., Baker,M.L., Best,C., Bi,C., Dougherty,M., Feng,P., Van Ginkel,G., Devkota,B., Lagerstedt,I., Ludtke,S.J., *et al.* (2011) EMDataBank.org: unified data resource for CryoEM. *Nucleic Acids Res.*, **39**, D456–D464.
23. Bekker,G., Yokochi,M., Suzuki,H., Ikegawa,Y., Iwata,T., Kudou,T., Yura,K., Fujiwara,T., Kawabata,T. and Kurisu,G. (2022) Protein Data Bank Japan: celebrating our 20th anniversary during a global pandemic as the Asian hub of three dimensional macromolecular structural data. *Protein Sci.*, **31**, 173–186.
24. wwPDB consortium (2019) Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.*, **47**, D520–D528.
25. 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.
26. Xu,W., Velankar,S., Patwardhan,A., Hoch,J.C., Burley,S.K. and Kurisu,G. (2023) Announcing the launch of Protein Data Bank China as an Associate Member of the Worldwide Protein Data Bank Partnership. *Acta Crystallogr. Sect. Struct. Biol.*, **79**, 792–795.
27. 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**, 160018.
28. Varadi,M., Anyango,S., Deshpande,M., Nair,S., Natassia,C., Yordanova,G., Yuan,D., Stroe,O., Wood,G., Laydon,A., *et al.* (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.*, **50**, D439–D444.
29. Valentini,E., Kikhney,A.G., Previtali,G., Jeffries,C.M. and Svergun,D.I. (2015) SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Res.*, **43**, D357–D363.
30. Sali,A., Berman,H.M., Schwede,T., Trewhella,J., Kleywegt,G., Burley,S.K., Markley,J., Nakamura,H., Adams,P., Bonvin,A.M.J.J., *et al.* (2015) Outcome of the first wwPDB hybrid/integrative methods task force workshop. *Structure*, **23**, 1156–1167.
31. Sheldrick,G.M. (1990) Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallogr. A*, **46**, 467–473.
32. Wlodawer,A. and Dauter,Z. (2017) `Atomic resolution': a badly abused term in structural biology. *Acta Crystallogr. Sect. Struct. Biol.*, **73**, 379–380.
33. 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.
34. 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., *et al.* (2017) Validation of Structures in the Protein Data Bank. *Structure*, **25**, 1916–1927.
35. Feng,Z., Westbrook,J.D., Sala,R., Smart,O.S., Bricogne,G., Matsubara,M., Yamada,I., Tsuchiya,S., Aoki-Kinoshita,K.F., Hoch,J.C., *et al.* (2021) Enhanced validation of small-molecule

ligands and carbohydrates in the Protein Data Bank. *Structure*, **29**, 393–400.

36. Wang,Z., Patwardhan,A. and Kleywegt,G.J. (2022) Validation analysis of EMDB entries. *Acta Crystallogr. Sect. Struct. Biol.*, **78**, 542–552.
37. 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.
38. Westbrook,J.D., Young,J.Y., Shao,C., Feng,Z., Guranovic,V., Lawson,C.L., 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.
39. Henderson,R., Sali,A., Baker,M.L., Carragher,B., Devkota,B., Downing,K.H., Egelman,E.H., Feng,Z., Frank,J., Grigorieff,N., *et al.* (2012) Outcome of the first electron microscopy validation task force meeting. *Structure*, **20**, 205–214.
40. Pintilie,G., Zhang,K., Su,Z., Li,S., Schmid,M.F. and Chiu,W. (2020) Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat. Methods*, **17**, 328–334.
41. Lagerstedt,I., Moore,W.J., Patwardhan,A., Sanz-García,E., Best,C., Swedlow,J.R. and Kleywegt,G.J. (2013) Web-based visualisation and analysis of 3D electron-microscopy data from EMDB and PDB. *J. Struct. Biol.*, **184**, 173–181.
42. Patwardhan,A., Carazo,J.-M., Carragher,B., Henderson,R., Heymann,J.B., Hill,E., Jensen,G.J., Lagerstedt,I., Lawson,C.L., Ludtke,S.J., *et al.* (2012) Data management challenges in three-dimensional EM. *Nat. Struct. Mol. Biol.*, **19**, 1203–1207.
43. Patwardhan,A., Ashton,A., Brandt,R., Butcher,S., Carzaniga,R., Chiu,W., Collinson,L., Doux,P., Duke,E., Ellisman,M.H., *et al.* (2014) A 3D cellular context for the macromolecular world. *Nat. Struct. Mol. Biol.*, **21**, 841–845.
44. Patwardhan,A., Brandt,R., Butcher,S.J., Collinson,L., Gault,D., Grünewald,K., Hecksel,C., Huiskonen,J.T., Iudin,A., Jones,M.L., *et al.* (2017) Building bridges between cellular and molecular structural biology. *eLife*, **6**, e25835.
45. Liu,H.-F., Zhou,Y. and Bartesaghi,A. (2022) High-resolution structure determination using high-throughput electron cryo-tomography. *Acta Crystallogr. Sect. Struct. Biol.*, **78**, 817–824.
46. Bouvette,J., Liu,H.-F., Du,X., Zhou,Y., Sikkema,A.P., Da Fonseca Rezende E Mello,J., Klemm,B.P., Huang,R., Schaaper,R.M., Borgnia,M.J., *et al.* (2021) Beam image-shift accelerated data acquisition for near-atomic resolution single-particle cryo-electron tomography. *Nat. Commun.*, **12**, 1957.
47. Eisenstein,F., Yanagisawa,H., Kashihara,H., Kikkawa,M., Tsukita,S. and Danev,R. (2023) Parallel cryo electron tomography on in situ lamellae. *Nat. Methods*, **20**, 131–138.
48. Wagner,T., Merino,F., Stabrin,M., Moriya,T., Antoni,C., Apelbaum,A., Hagel,P., Sitsel,O., Raisch,T., Prumbaum,D., *et al.* (2019) SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.*, **2**, 218.
49. Sanchez-Garcia,R., Segura,J., Maluenda,D., Carazo,J.M. and Sorzano,C.O.S. (2018) *Deep Consensus*, a deep learning-based approach for particle pruning in cryo-electron microscopy. *IUCrJ*, **5**, 854–865.
50. Sanchez-Garcia,R., Gomez-Blanco,J., Cuervo,A., Carazo,J.M., Sorzano,C.O.S. and Vargas,J. (2021) DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.*, **4**, 874.
51. Dai,M., Dong,Z., Xu,K. and Zhang,Q.C. (2023) CryoRes: local resolution estimation of cryo-EM density maps by deep learning. *J. Mol. Biol.*, **435**, 168059.
52. Maddhuri Venkata Subramaniya,S.R., Terashi,G. and Kihara,D. (2019) Protein secondary structure detection in intermediate-resolution cryo-EM maps using deep learning. *Nat. Methods*, **16**, 911–917.
53. Terashi,G., Wang,X., Maddhuri Venkata Subramaniya,S.R., Tesmer,J.J.G. and Kihara,D. (2022) Residue-wise local quality estimation for protein models from cryo-EM maps. *Nat. Methods*, **19**, 1116–1125.
54. Pfab,J., Phan,N.M. and Si,D. (2021) DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2017525118.
55. Jamali,K., Käll,L., Zhang,R., Brown,A., Kimanius,D. and Scheres,S.H.W. (2023) Automated model building and protein identification in cryo-EM maps Molecular Biology.
56. Read,R.J., Adams,P.D., Arendall,W.B., Brunger,A.T., Emsley,P., Joosten,R.P., Kleywegt,G.J., Krissinel,E.B., Lütteke,T., Otwinowski,Z., *et al.* (2011) A new generation of crystallographic validation tools for the Protein Data Bank. *Structure*, **19**, 1395–1412.

Appendix

Current wwPDB Consortium Members with Affiliations

EMDB

Jack Turner[1], Sanja Abbott[1], Neli Fonseca[1], Ryan Pye[1], Lucas Carrijo[1], Amudha Kumari Duraisamy[1], Osman Salih[1], Zhe Wang[1], Gerard J. Kleywegt[1], Kyle L. Morris[1,*], Ardan Patwardhan[1,*]

RCSB PDB

Stephen K. Burley[2,3,4,5], Gregg Crichlow[2], Zukang Feng[2], Justin W. Flatt[2], Sutapa Ghosh[2], Brian P. Hudson[2], Catherine L. Lawson[2], Yuhe Liang[2], Ezra Peisach[2], Irina Persikova[2], Monica Sekharan[2], Chenghua Shao[2], Jasmine Young[2]

PDBe

Sameer Velankar[6], David Armstrong[6], Marcus Bage[6], Wesley Morellato Bueno[6], Genevieve Evans[6], Romana Gaborova[7], Sudakshina Ganguly[6], Deepti Gupta[6], Deborah Harrus[6], Ahsan Tanweer[6], Manju Bansal[8], Vetriselvi Rangannan[8]

PDBj

Genji Kurisu[9,10], Hasumi Cho[9], Yasuyo Ikegawa[9], Yumiko Kengaku[10], Ju Yaen Kim[10], Satomi Niwa[10], Junko Sato[10], Ayako Takuwa[10], Jian Yu[10]

BMRB

Jeffrey C. Hoch[11], Kumaran Baskaran[11]

PDBc

Wenqing Xu[12,13], Weizhe Zhang[12], Xiaodan Ma[12]

1. Cellular Structure and 3D Bioimaging, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, United Kingdom

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

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

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

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

6. Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-

EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, United Kingdom

7. CEITEC - Central European Institute of Technology, Masaryk University, Kamenice 5, 62500 Brno, Czech Republic

8. Molecular Biophysics Unit, Indian Institute of Science, Bangalore. India

9. Protein Data Bank Japan, Protein Research Foundation, Minoh, Osaka 562–8686, Japan

10. Protein Data Bank Japan, Institute for Protein Research, Osaka University, Suita, Osaka 565–0871, Japan

11. Biological Magnetic Resonance Data Bank, Department of Molecular Biology and Biophysics, UConn Health, 263 Farmington Ave., Farmington, CT 06030,–3305, USA

12. National Facility for Protein Science, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China

13. School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China

Received: September 28, 2023. **Revised:** October 18, 2023. **Editorial Decision:** October 19, 2023. **Accepted:** October 20, 2023