



OPEN ACCESS

EDITED BY

Barbora Kozlikova,
Masaryk University, Czechia

REVIEWED BY

Nayim Gonzalez-Rodriguez,
Spanish National Cancer Research
Center, Spain
Jakob Ruickoldt,
University of Potsdam, Germany

*CORRESPONDENCE

Kyle L. Morris,
✉ kyle@ebi.ac.uk

RECEIVED 08 December 2025

REVISED 16 January 2026

ACCEPTED 28 January 2026

PUBLISHED 06 March 2026

CITATION

Fonseca N, Duraisamy AK, Wang Z,
Somasundharam S, Tayebinia M, de
Oliveira LC, Ma M, Turner J,
Patwardhan A, Kleywegt GJ, Hartley M
and Morris KL (2026) Chart builder: an
interactive tool for user driven data
visualization in the electron microscopy
data bank.
Front. Bioinform. 6:1763403.
doi: 10.3389/fbinf.2026.1763403

COPYRIGHT

© 2026 Fonseca, Duraisamy, Wang,
Somasundharam, Tayebinia, de Oliveira,
Ma, Turner, Patwardhan, Kleywegt,
Hartley and Morris. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Chart builder: an interactive tool for user driven data visualization in the electron microscopy data bank

Neli Fonseca¹, Amudha Kumari Duraisamy¹, Zhe Wang¹,
Sriram Somasundharam², Minoosadat Tayebinia¹,
Lucas C. de Oliveira¹, Miao Ma¹, Jack Turner¹,
Ardan Patwardhan³, Gerard J. Kleywegt³, Matthew Hartley² and
Kyle L. Morris^{1*}

¹Electron Microscopy Data Bank (EMDB), European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Cambridgeshire, United Kingdom, ²Electron Microscopy Public Image Archive (EMPIAR), European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Cambridgeshire, United Kingdom, ³Cellular Structure and 3D Bioimaging (CS3DB), European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Cambridgeshire, United Kingdom

The cryogenic sample-electron microscopy (cryoEM) field has generated significant amounts of 3D Electron Microscopy (3DEM) volumetric data and associated metadata, now comprehensively archived in the Electron Microscopy Data Bank (EMDB - www.emdatabank.org) and the Electron Microscopy Public Image Archive (EMPIAR - www.empiar.org). Harnessing the full potential of these resources requires robust, flexible, and publicly accessible tools for data exploration, analysis and retrieval. Here, we present Chart Builder, an interactive web-based platform that enables researchers to create customizable, publication-quality visualizations directly from archival metadata, validation assessments, and cross-reference annotations. Chart Builder integrates the same query-driven and flexible Solr search system as EMDB search, into a user interface with tools to assist users to filter, group, and compare data without programming expertise. It supports multiple chart types (including line, bar, area, scatter (2D and 3D), histogram, bubble, pie, geographic and Venn diagrams) with customizable axes, data series, and statistical operators. Users can apply global filters, define temporal, categorical, or custom query-based axes, and explore multi-dimensional relationships interactively. Chart data-points are linked to their underlying datasets, such that visualisation interaction opens entry-level or archive-level search results for inspection and datasets from charts may be exported in several ways. Findability, accessibility, interoperability and reusability of data are facilitated by these direct access and export mechanisms, including HTML embedding, persistent URL sharing and chart/data download options. By combining interactivity and ease of use with up-to-date access to the EMDB and EMPIAR archive metadata, both computational and experimental communities may explore and visualize current metadata and export to formats for further analysis or as publication-ready figures. Chart Builder promotes community-driven data analysis and empowers users to evaluate trends in the biological 3DEM field. Chart Builder is freely

accessible and fully integrated into the EMDB website at <https://www.ebi.ac.uk/emdb/statistics/builder/>.

KEYWORDS

cryoEM (cryo-electron microscopy), data repository, data visualization, FAIR (findable accessible interoperable and reusable) principles, graph

1 Introduction

The Electron Microscopy Data Bank (EMDB) is a public repository for three-dimensional electron microscopy (3DEM) volumetric reconstructions from cryogenic sample-electron microscopy (cryoEM) and cryogenic sample-electron tomography (cryoET) (wwPDB Consortium, 2024). The Electron Microscopy Public Image Archive (EMPIAR) holds cryoEM and cryoET 2D and 3D image experimental and derived data, volume electron microscopy (vEM) and soft and hard X-ray tomography (XT) data (Ludin et al., 2023). Both are entry-based deposition databases, with over one hundred schema-controlled data fields describing individual entries. Together, EMDB and EMPIAR form a comprehensive resource ecosystem that represent data and metadata from experimental data collection, that data's reduction and interpretation, through to the final 3DEM reconstruction of a biological target, predominantly via averaging techniques. Where a molecular reconstruction is targeted, these data may have model coordinates derived, which may be stored in the PDB database (wwPDB Consortium, 2019). The availability of data from these databases and connections between them creates scientific transparency, reproducibility, and innovation in cryoEM (Patwardhan et al., 2017). Importantly and more broadly this translates the same benefits to developments in experimental approaches (Kimanius et al., 2024; Jamali et al., 2024), structural knowledge (PDB-EM consortium, 2020) and predictive structural biology data (Jumper et al., 2021; Varadi et al., 2022) made possible by the reuse of FAIR data.

The cryoEM field has been experiencing rapid growth, driven by rising investments in structural biology research, technological advances in throughput and quality of cryoEM imaging (Kühlbrandt, 2014) and the success of image processing softwares such as RELION (Zivanov et al., 2018) and cryoSPARC (Punjani et al., 2017). Most often molecular structures are the targets of these studies leading to these data being archived in the core archives, EMDB, EMPIAR and PDB. This surge in data availability is highly valuable but presents mounting challenges for the community in terms of data analysis, quality assessment and underscores the essential need for accessible, standardised statistical summaries and visualization tools (Lawson et al., 2021).

Statistics plots have been available in the EMDB since its inception through the statistics page (formerly named EMstats), a collection of predefined charts developed in collaboration with the EM community. Initially containing a few dozen, it now includes over one hundred charts, summarizing the archives' overall contents and trends (Tagari et al., 2002). These charts provide high-level insights into features such as resolution distribution, database growth over time, and methodological parameters (<https://www.ebi.ac.uk/emdb/statistics>). Additionally, for any set of search results, a set of nine dynamic plots

can be viewed from the search statistics. The chart types are fixed but summarize the subset of retrieved entries. However, users wishing to generate any deeper statistical summaries or even analytical queries of the EMDB and EMPIAR are required to write bespoke scripts to access and retrieve the database content.

The motivation for this work is to provide an accessible, flexible, user-friendly, query-based visualization plotting tool designed to facilitate data exploration and enable the rapid creation of statistical and analytical enquiries without requiring coding skills or deep technical knowledge of the database. The EMDB Chart Builder extends and integrates EMDB REST API endpoints into an interactive, query-based framework that enables users to design customized visual summaries of EMDB data. It represents the next stage in the evolution of EMDB statistics, offering greater flexibility, transparency, and direct access to analytical tooling for dataset discovery and data-driven insights across a rapidly growing cryoEM archive.

2 Materials and methods

2.1 Data sources and availability

The statistical visualizations in this work are based on four main data sources: author metadata provided during database submission to deposition systems either to EMDB via OneDep (Young et al., 2017) or to EMPIAR directly, and added-value annotations from EMDB EMICSS (Duraismy et al., 2025) and EMDB Validation Analysis (Wang et al., 2022). Data from these sources are indexed in the EMDB Solr search engine and provided to the chart builder via access to public REST API endpoints. The user of the chart builder may identify metadata that they wish to explore and plot by searching all fields available in the search system (<https://www.ebi.ac.uk/emdb/documentation/search/fields>).

2.1.1 EMDB metadata

The EMDB metadata is provided by the authors (depositors) during the deposition of cryoEM maps data and its associated metadata in the OneDep system (Young et al., 2017) through the OneDep deposition interface (<https://deposit-pdbe.wwpdb.org/>). Captured data includes but is not limited to key information describing each three-dimensional EM (3DEM) reconstruction, such as authors, dates, sample information, experimental information, map details and metadata for deposited files (Westbrook et al., 2022). These metadata are stored by the OneDep system in the mmCIF dictionary format (<https://mmcif.wwpdb.org/>) and translated into the EMDB

schema controlled XML data model for dissemination (https://www.ebi.ac.uk/emdb/documentation#data_model).

2.1.2 EMPIAR metadata

EMPIAR metadata is supplied by EMPIAR depositors, from the EMPIAR deposition system (<https://www.ebi.ac.uk/empair/deposition>), describing administrative and technical data for raw imaging datasets, including contributor information, dataset descriptions, sample identifiers and cross references to EMDB.

2.1.3 Validation analysis metadata

The validation of EMDB entries provides standardized assessments of map and model quality using a broad set of metrics calculated by the Validation Analysis package (Wang et al., 2022). Key metrics that are made publicly available include resolution values from internally calculated (Chen et al., 2013) and deposited Fourier Shell Correlation (FSC) plots with multiple cutoffs, Q-score (Pintilie et al., 2020), Atom inclusion (Lagerstedt et al., 2013), CCC (Liebschner et al., 2019) and SMOC score (Joseph et al., 2017) for model-map fitting quality.

2.1.4 Cross-reference annotations

Database cross-references delivering added-value annotations are systematically gathered by the EMICSS framework (Duraismy et al., 2025). It generates comprehensive cross-references for each EMDB entry, providing direct links to a vast array of biological and structural databases. Specifically, it connects to resources related to protein sequences: UniProt (The UniProt Consortium, 2025), ontological labels: Gene Ontology (Aleksander et al., 2023); structures: PDB (wwPDB Consortium, 2019), PDBe-KB (PDBe-KB Consortium, 2022), AlphaFold DB (Varadi et al., 2022), EMPIAR (Iudin et al., 2023); complexes: Complex Portal (Meldal et al., 2015), domains: Pfam (Mistry et al., 2021), Rfam (Kalvari et al., 2021), InterPro (Paysan-Lafosse et al., 2023), CATH (Sillitoe et al., 2021), SCOP (Andreeva et al., 2020); ligands: ChEBI (Hastings et al., 2016), ChEMBL (Zdrzil et al., 2024), DrugBank (Knox et al., 2024); publications: ORCID (Haak et al., 2012), ISSN, journal names, PubMed, PubMed Central, DOI; and standardised molecular weight obtained by PDBe assemblies (Velankar et al., 2016).

2.2 Query system

The EMDB query system is powered by Apache Solr (Apache Software Foundation, 2025), utilizing the Lucene query language to enable a flexible and powerful filter mechanism for searching the database. This system allows users to combine multiple filters using Boolean operators and range queries, providing high adaptability in constructing complex searches tailored to specific criteria. For example, users can filter entries by resolution ranges (e.g., resolution [3 TO 5]) or apply time-based filters to select data deposited within a certain date range (e.g., deposition_date [2024-01-01T00:00:00Z TO 2024-12-31T00:00:00Z]). Full documentation on how to use the EMDB search system and the list of available search fields can be

found in the website (<https://www.ebi.ac.uk/emdb/documentation/search>, <https://www.ebi.ac.uk/emdb/documentation/search/fields>).

2.3 Implementation details

The implementation of the EMDB Chart Builder relies on a backend architecture composed of REST API endpoints developed using Django (Django Software Foundation, 2025) and Python. These backend services communicate directly with the Solr search engine to execute user queries, retrieve, and format the relevant metadata and validation information from EMDB and EMPIAR. On the front-end, Highcharts (Highsoft, 2025) is used as the visualization library to render interactive and customizable statistical plots. This architecture allows flexible data exploration with rapid responses, offering an intuitive user experience that does not require users to handle raw data or scripting directly.

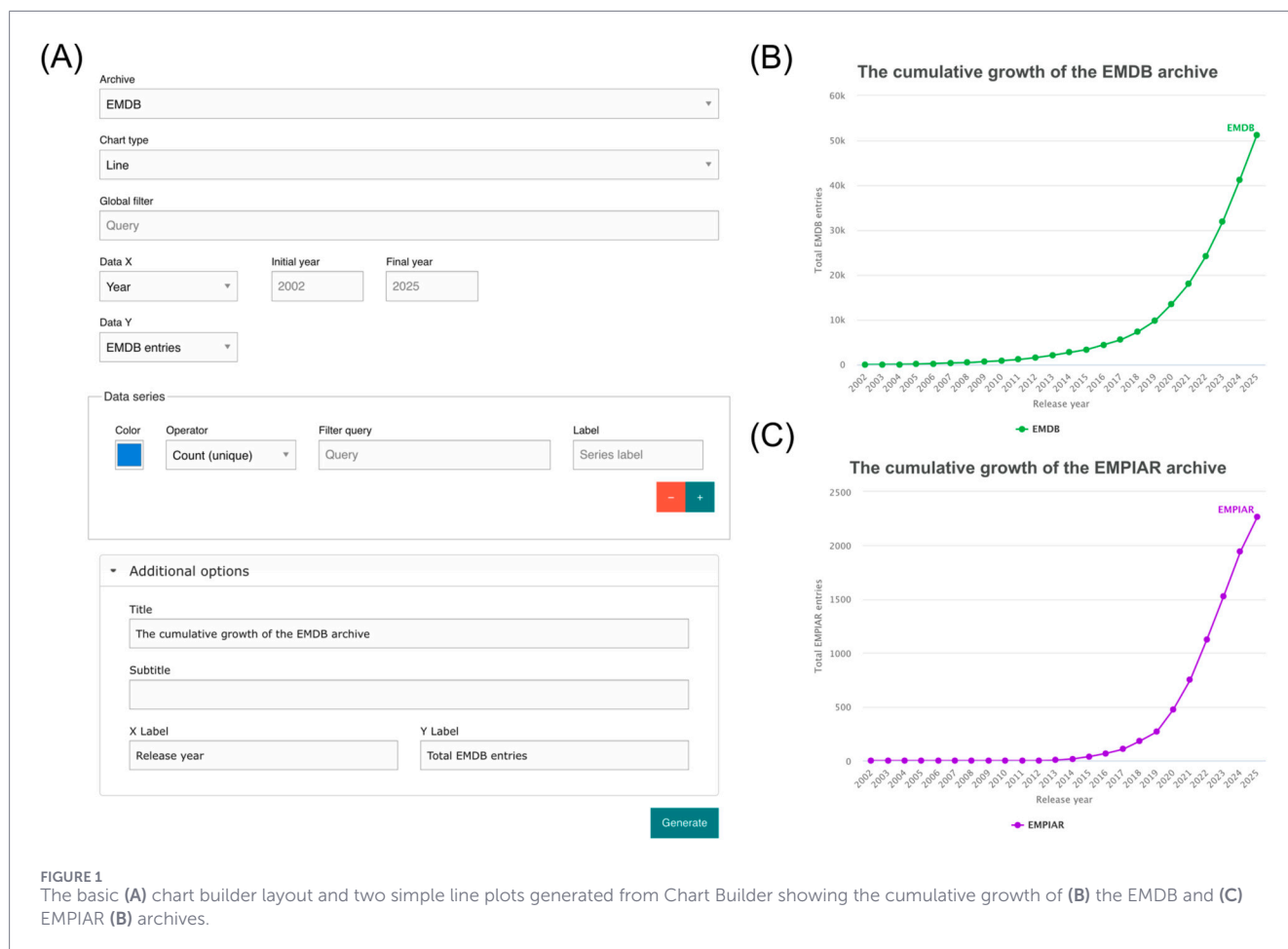
The data underlying the Search System is updated on a weekly schedule, during the EMDB weekly release cycle that occurs every Wednesday at midnight UTC (wwPDB Consortium, 2024). For every new release and updated entries scheduled to the current weekly release, EMDB indexes the metadata and performs a full run of the Validation Analysis package, indexing the results in the Search System. The EMICSS pipeline is executed for every entry of the weekly archive release, with regular updates of the archive (Duraismy et al., 2025), updating all the Solr search documents to the latest version of data from cross-referenced databases.

2.4 Reproducibility and usability

The EMDB Chart Builder promotes reproducibility and usability by providing open access through the EMDB website, ensuring that all users can freely explore, create, and share custom charts without restrictions, in line with EMDB's open data policies. A key aspect of reproducibility is the ability to trace and adapt the visualizations that are already in use across the resource. To support this, most of the charts displayed on the EMDB statistics page and within search result statistics now include an "Edit" button that opens the corresponding plot directly in the Chart Builder. This feature allows users to examine the underlying configuration of predefined charts, use them as worked examples, and then adapt or extend them to address their own research questions. In addition, the tool guarantees consistency between chart visualizations and corresponding tabular output, enabling users to seamlessly switch between graphical and numerical representations for validation and deeper analysis. The Chart Builder is fully integrated into the existing EMDB website and can be accessed at <https://www.ebi.ac.uk/emdb/statistics/builder>.

3 Results

The Chart Builder provides an interactive user interface where users can generate custom visualizations from EMDB and EMPIAR metadata without the need for programming. The tool integrates a query-driven interface (Figure 1A) with multiple chart types, flexible axis definitions, and customizable data series, allowing users to explore archive statistics in a reproducible and transparent manner. In the following subsections, we demonstrate the main



functionalities of this web based user driven visualisation and analysis tool.

3.1 Archive selection

Users start by selecting the archive from which the data will be drawn: either EMDB or EMPIAR. The selection of the archive will change what search fields are available for use when building the chart. Figure 1B contains two graphs generated by the Chart Builder showing the growth of EMDB and EMPIAR archives on an annual basis.

3.2 Chart types

Chart Builder supports a diverse range of visualization types to accommodate different data exploration and analytical needs. The tool includes standard formats such as line, area, bar and stream; statistical formats such as histograms, pie charts and scatter plots; as well as more specialized visualizations such as bubble plots, Venn diagrams and map charts.

3.2.1 Bar and area charts

Bar and area charts can be displayed in stacked, unstacked, or percentage-stacked (Figure 2A) formats, making them suitable for

analysing distributions and proportions across time or metadata categories.

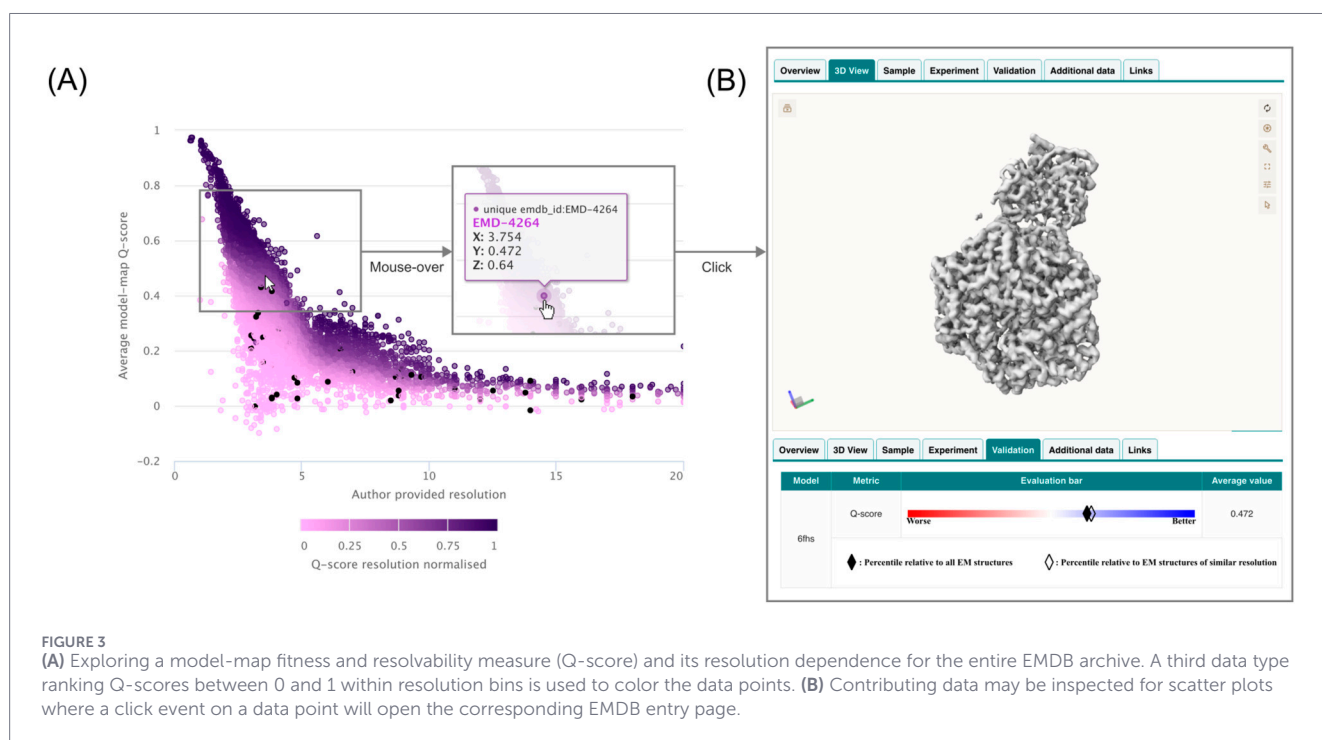
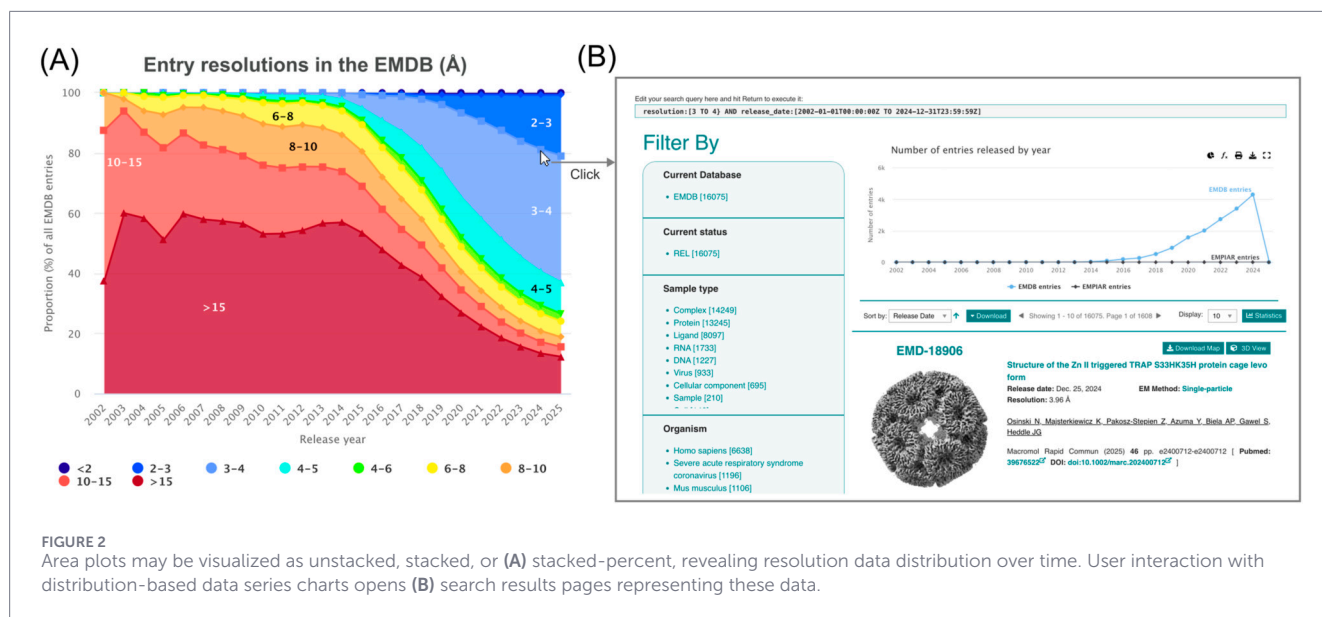
Each chart is interactive and can be customized to display data series according to user-defined filters, colours, and statistical operators, such as maximum, minimum, average, count or unique. All charts can be interacted with, where hovering with the cursor will display an inspection of the data for that distribution data, and a click event will redirect to the EMDB search results (Figure 2B) for the corresponding data of that part of the chart. Most of the chart types share the same controls, some exceptions are listed below.

3.2.2 Pie charts

Pie charts are limited to a single data series and provide a clear overview of proportional breakdowns (e.g., fraction of entries by EM method). The pie chart slices are defined by the dataset provided to the X-axis option, and the values are defined by the Y-axis option.

3.2.3 Scatter plots

Scatter (2D and 3D) and bubble plots allow the exploration of relationships between numerical variables. Instead of defining x- and y-axes through categorical or temporal variables, users directly select numerical variables (search fields) from a dropdown (e.g., molecular weight, resolution, map size). When entries are



identified with data available in each named variable, the number of entries identified and plotted is reported. Each point in these plots corresponds to an individual EMDb or EMPIAR entry and the click event will redirect to the respective entry page instead of the search results. Bubble and Scatter 3D plots further extend this by incorporating a third numerical dimension (z-axis), such as dataset size, resolution or other validation metrics, which determines the z-axis scaling (bubble size or point color) as seen in Figure 3.

3.2.4 Histograms

Histograms enable users to examine the distribution of numerical variables within EMDb or EMPIAR datasets, with examples shown in Figure 4. Similar to scatter plots, histograms require the user to select one numerical variable from the available search fields to define the X-axis (for example, resolution, map voxel size, or molecular weight). The frequency or count of entries within user defined value intervals is then represented along the Y-axis.

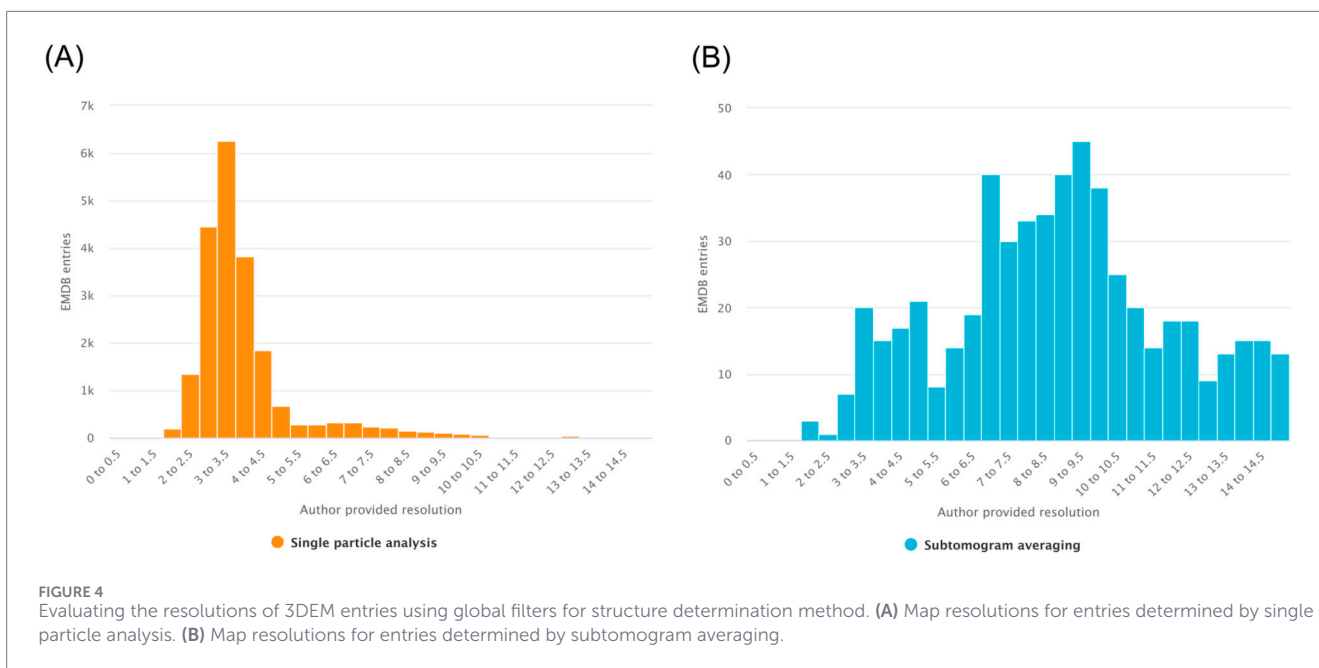


FIGURE 4 Evaluating the resolutions of 3DEM entries using global filters for structure determination method. **(A)** Map resolutions for entries determined by single particle analysis. **(B)** Map resolutions for entries determined by subtomogram averaging.

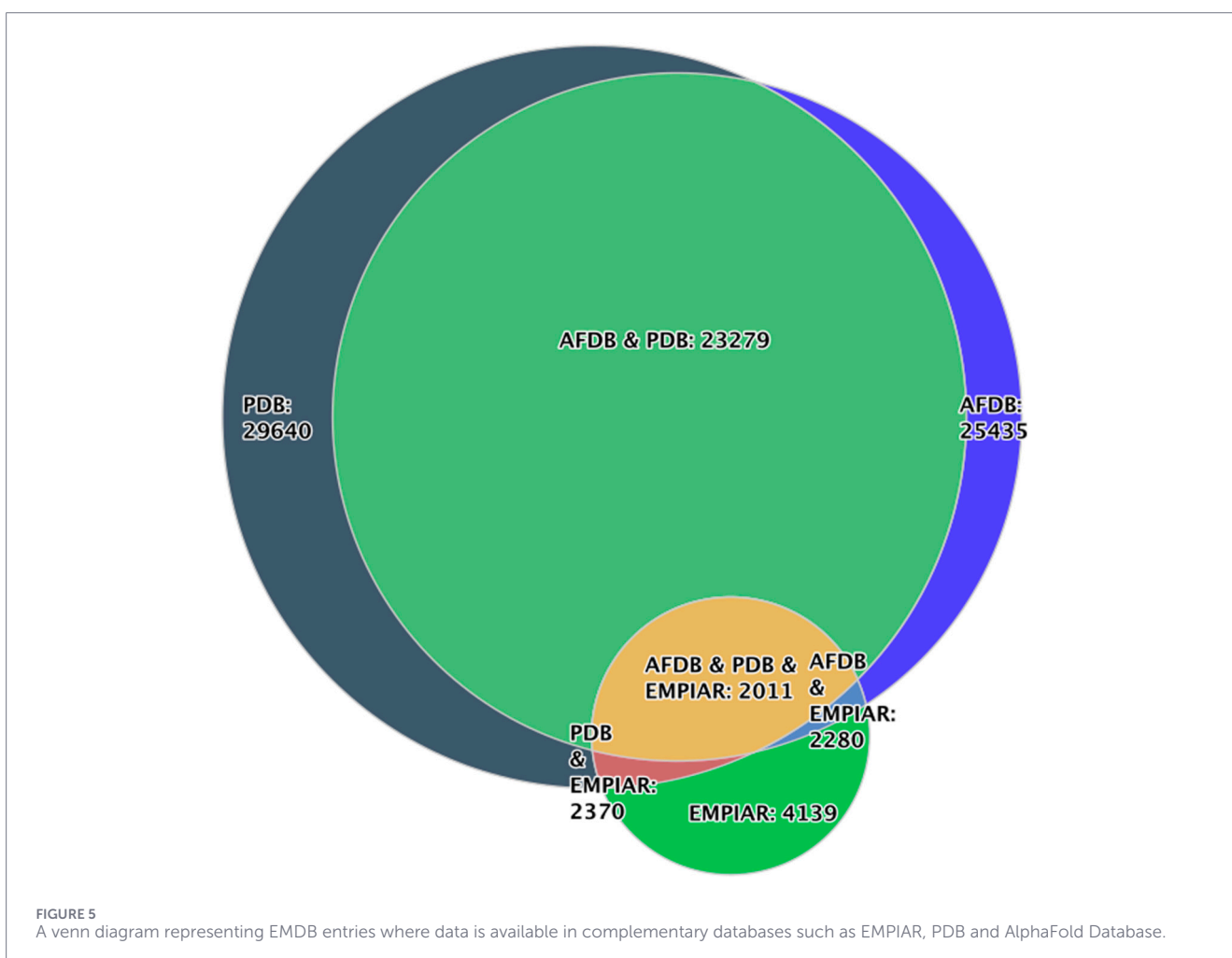
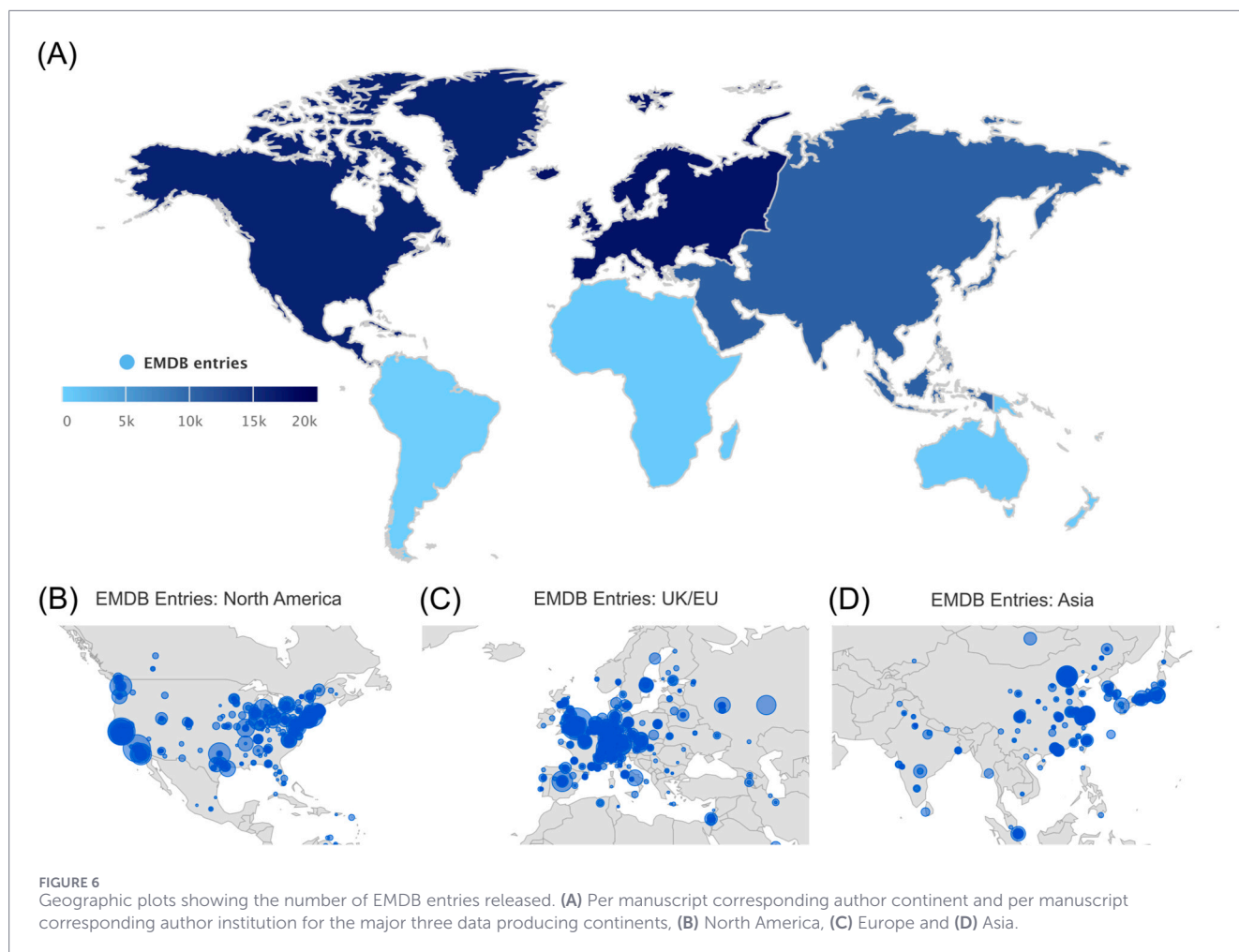


FIGURE 5 A Venn diagram representing EMDb entries where data is available in complementary databases such as EMPIAR, PDB and AlphaFold Database.



User-defined value intervals are defined for the histogram by specifying the start value, end value, and bin (gap) size, which allows control over the resolution and range of the displayed distribution. This flexibility makes histograms particularly useful for identifying trends, outliers, or biases in data distributions.

3.2.5 Venn diagrams

Venn diagrams provide a unique way to visualize overlaps between sets defined by user-specified queries. In this plot, users also do not define x- and y-axes but instead use the data series to define each set. Each data series corresponds to a set, and intersections illustrate entries that satisfy multiple conditions, as shown in Figure 5 for entries in the EMDB with fitted models in PDB, corresponding models in AlphaFold DB and raw data in EMPIAR. This feature enables users to investigate relationships between subsets of the archives in an intuitive manner.

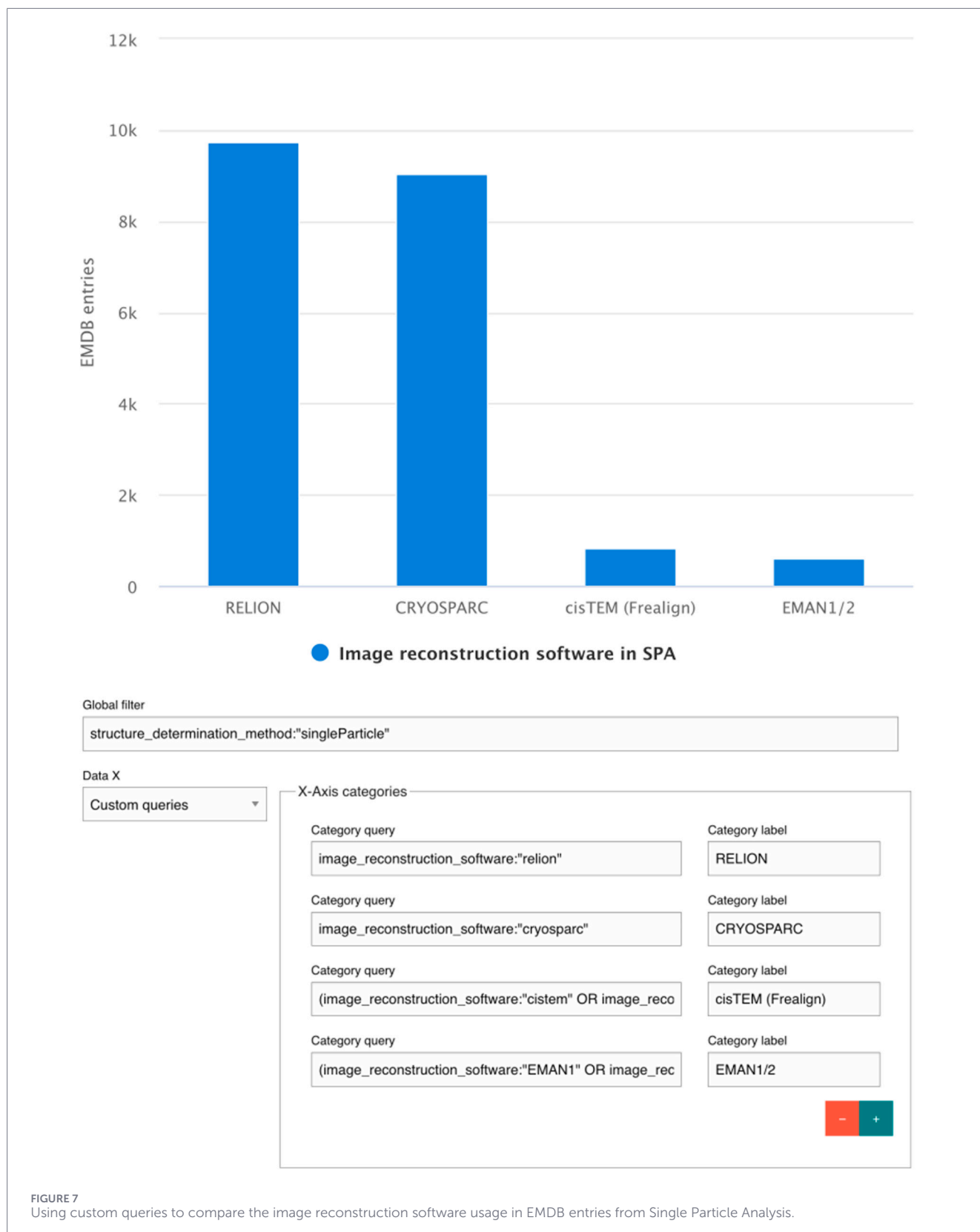
3.2.6 Geographic plots

Geolocation annotations are available for both EMDB and EMPIAR entries. For EMPIAR, we provide statistics for the principal investigator's institutional affiliation's country and continent, since

its information is available in the EMPIAR metadata. For EMDB, we provide access to wwPDB sites (deposition and annotation) and grant country, both collected from the deposition; and we also have four enriched fields collected from the Europe PMC: corresponding author country, continent, city and institution (Rosonovski et al., 2024). These will only include EMDB entries with an associated publication that is in open access format. Depending on the type of data, the geographic plots can be coloured as heatmap or bubbles. In Figure 6 we show both examples, the number of EMDB entries released per continent as a heatmap, and the number of EMDB entries released by institutions at the Research Organization Registry (Research Organization Registry, 2024) visualized as a bubble map.

3.3 Global filter

Before generating a chart, users can apply a global filter to restrict the entire dataset that will be provided to the chart. This filter is expressed as a query in the EMDB query language, i.e., `structure_determination_method:"singleParticle"`. A global filter can target any available metadata field, such as reconstruction method, release year, or sample type. By narrowing the dataset at the outset, researchers can focus their analysis on a specific subset of interest. For example, a query restricted to single-particle reconstructions



allows users to investigate resolution trends only within that method, while excluding tomography or subtomogram averaging entries. Documentation is available for the search query syntax used for

filtering at <https://www.ebi.ac.uk/emdb/documentation/search> and via the use of the 'Filter' button to the right of the Global filter field.

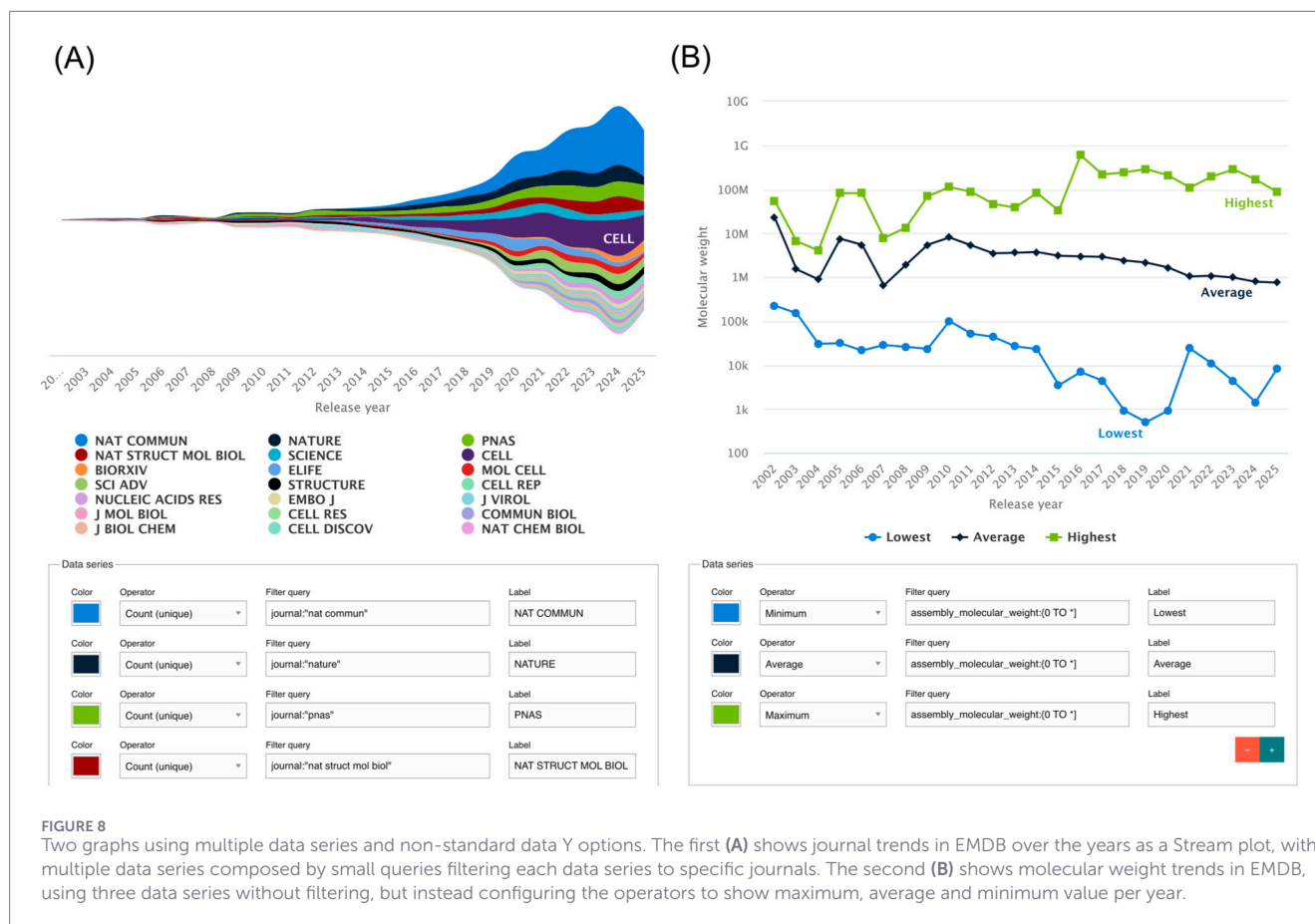


FIGURE 8

Two graphs using multiple data series and non-standard data Y options. The first (A) shows journal trends in EMDB over the years as a Stream plot, with multiple data series composed by small queries filtering each data series to specific journals. The second (B) shows molecular weight trends in EMDB, using three data series without filtering, but instead configuring the operators to show maximum, average and minimum value per year.

TABLE 1 Descriptions of the data export and chart modification.

Icon	Function
	Show all curves hide all curves
	HTML for embedding
	URL for sharing
	Print
	Download chart as image PNG/JPEG/PDF/SVG download data table CSV/XLS
	Full screen
	Y-axis scaling: Linear/Logarithmic
	Bar/Area behaviour: Unstacked stacked stacked percentage

3.4 X-axis options

The independent variables (x-axis) of a chart can be defined in three different ways, enabling both temporal and categorical analyses:

Yearly - Users may specify a start and end year to visualize trends over time. This is particularly effective for examining archive growth,

such as the steady increase in high-resolution structures deposited since 2015 as seen in Figure 2A.

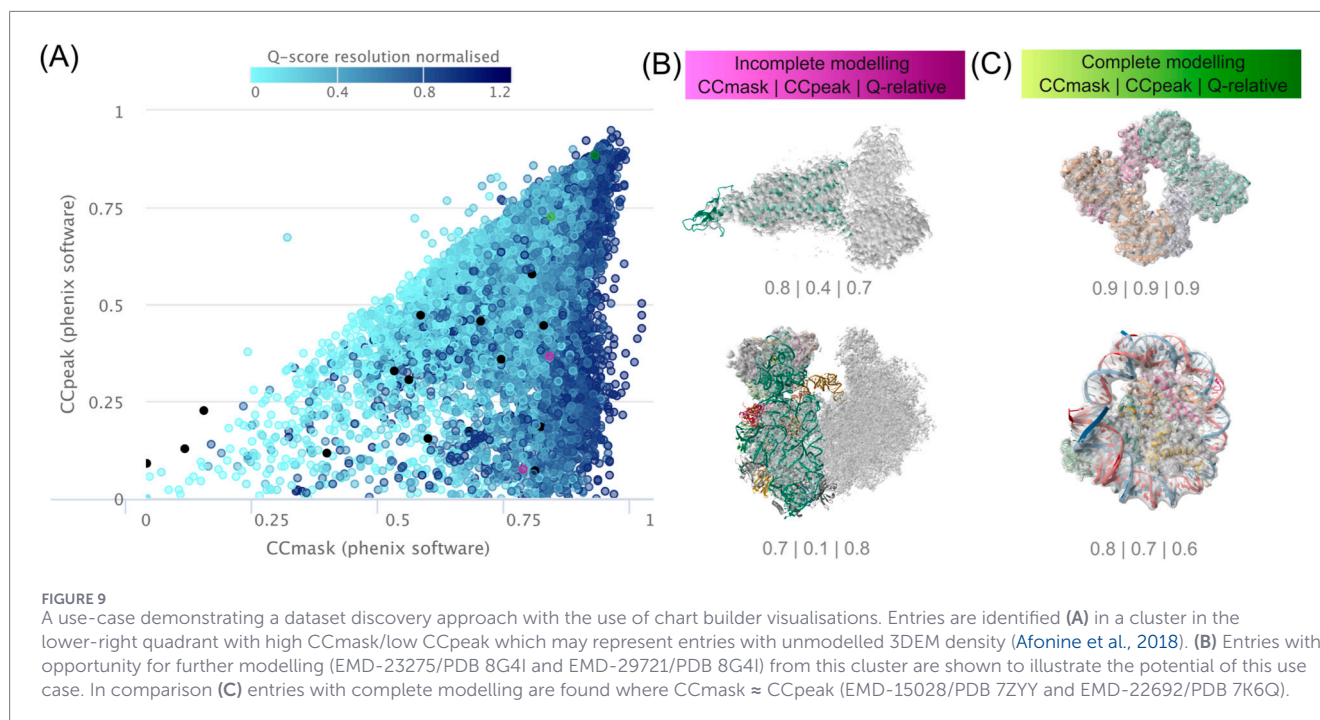
By Experimental Metadata—A categorical variable (e.g., EM method, sample type, or microscope model) can be chosen as the x-axis, with all categories automatically included in the visualization. This option enables comparisons across experimental conditions, for example, showing the relative contribution of different microscopes to deposited structures.

By Custom Queries—Users can define a set of queries manually, with each query corresponding to one label on the x-axis. This provides maximum flexibility, allowing the creation of user-defined categories such as a distinct set of software as seen in Figure 7.

Together, these options allow users to construct temporal, categorical, or bespoke groupings, making it possible to address a wide range of exploratory questions with a single interface.

3.5 Y-axis options

By default, the dependent variables (y-axis) are displayed as the number of EMDB/EMPIAR entries. However, users can select alternative variables, such as for EMDB: number of authors, number of fitted models, number of associated raw data entries in EMPIAR, number of masks, number of additional maps, resolution, number of publications, molecular weight, atom inclusion, Q-score, map size, and frames per image; and for EMPIAR: number of authors, number of associated EMDB entries, number of



Publications, dataset size, and number of images. Figure 8 shows two examples using distinct data Y options: count of publications and minimum/average/maximum molecular weight.

3.6 Series filter

The Chart Builder allows the inclusion of multiple data series within a single plot, enabling comparative and multi-dimensional analyses. Each data series can be customized with a user-defined color, statistical operator (count, sum, average, maximum, or minimum), and an optional Series Filter to refine the subset of data being represented. Users may also assign custom labels ("Category label"), which are displayed in the chart legend to ensure clarity in interpretation. An example on how data series can be used in this application is shown in Figure 7. The same search query syntax may be used for Series Filters as for Global Filters.

3.7 Chart customisation and export

The Chart Builder provides a range of interactive and customization options that enhance both the usability and adaptability of the visualizations. Once generated, charts remain fully interactive: users can dynamically switch between display modes such as stacked, unstacked, or percentage views (for bar and area plots—Figure 2) and toggle between linear and logarithmic scales (Figure 1) to effectively represent data spanning multiple orders of magnitude. Data series can also be interactively managed; clicking on a legend item allows users to toggle individual series on or off, facilitating clearer comparisons between subsets of data. These features are accessible via a menu at the top right of charts, the buttons of which are chart specific and summarised in Table 1.

Additional interactive features support exploration and validation of the dataset which has been developed for plotting in a

chart. Hovering over chart elements reveals precise numerical values relevant to that data point or series. While clicking on a data point in a scatter plot opens that specific entry for exploration. For the chart types histogram, pie, bar, when a chart section is clicked it opens a new window with search filters enabled to show results from that section of the chart. Export functionality is versatile by allowing the generation of shareable URL links to the customized chart builder page itself, generating HTML code for chart embedding, printable charts, and downloadable figures in multiple formats, including high-quality images (PNG, JPEG, PDF or SVG). Full screen views are suitable for presentations with support for granular customization for presentation and reporting including; Titles, subtitles, and axis labels directly editable within the interface. Finally, for further external data analyses, tabular files (CSV or XLS) of the data generating the current plot may be downloaded.

4 Discussion

The EMDB Chart Builder was developed to enable the statistical exploration of the expanding 3DEM landscape through intuitive, interactive configurable visualisations rather than scripts or specialist tools. As EMDB and EMPIAR continue to grow in size and complexity, tools are required to understand the archive at a glance to assess emerging patterns or contextualise deposited entries within broader community trends. The Chart Builder addresses this by leveraging the EMDB Solr search system into a visual, user-driven discovery environment where a broad user base can build, adapt, and share statistical summaries of the 3DEM structural biology archives. As a major step forward from the original fixed EMstats plots, the Chart Builder allows users to test unique hypotheses through configurable filters and metadata combinations, or refining existing EMDB curated visualisations. Lastly, these capabilities directly align

with FAIR principles: they make data easier to find, easier to access, interoperable across metadata sources, and readily reusable through persistent URLs and export options.

A central strength of the tool lies in the harmonisation of metadata from multiple sources—including depositor-supplied fields, EMICSS cross-references, and Validation Analysis metrics—into a single visual interface. This integration empowers the community to develop rich data driven insights: how 3DEM data properties inferred by validation are distributed, which software pipelines are shaping current practice, how molecular weights or resolutions vary across methods and time, or how data production is geographically distributed. User-level interactions support exploring these insights; hovering exposes metadata and clicking linking directly back to underlying entries or archive searches.

Figure 9 presents a use case where a data visualization may be used to identify a cluster of the EMDB archive with properties valuable to molecular modelling methods development. Further examples in this manuscript illustrate how visual analytics can support monitoring archive growth, exploring model-map fit, examining software usage and molecular weight trends, analysing cross-archive relationships, and mapping global 3DEM methodology trends. Collectively, these use cases demonstrate how visual tools can reveal clusters, anomalies, or methodological signals that are useful for benchmarking, training-set creation, and methodological innovation.

Altogether, the Chart Builder aims to support users by lowering barriers to data exploration and enabling researchers to interact with the archive in ways that are visual, open, and collaborative. It provides a modern platform for transparent and reproducible exploration of EMDB and EMPIAR, helping make electron microscopy data more accessible and interpretable, and supporting the need for data-driven insight, innovation, and discovery.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding author.

Author contributions

NF: Conceptualization, Data curation, Formal Analysis, Methodology, Software, Visualization, Writing – review and editing, Writing – original draft. AD: Data curation, Writing – review and editing. ZW: Data curation, Writing – review and editing. SS: Data curation, Writing – review and editing. MT: Data curation, Writing – review and editing. LO: Data curation, Writing – review and editing. MM: Data curation, Writing – review and editing. JT: Data curation, Writing – review and editing. AP: Funding acquisition, Writing – review and editing, Project administration. GK: Conceptualization, Supervision, Writing – review and editing, Project administration. MH: Supervision, Writing – review and editing. KM: Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review and editing, Project administration.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work has been supported by the European Molecular Biology Laboratory–European Bioinformatics Institute (EMBL-EBI); and the Wellcome Trust (grant number 212977/Z/18/Z and 310300/Z/24/Z).

Acknowledgements

We thank Matt Jeffryes and Melissa Harrison in EuropePMC (EMPC, EMBL-EBI) for discussion around EMPC API endpoints to provide geographical annotations to EMDB entries. We thank the wwPDB in the worldwide collaborative effort to archive structural biology data and the OneDep team for continued team work to deliver this vision.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. The author(s) verify and take full responsibility for the use of generative AI in the preparation of this manuscript. Generative AI was used. We acknowledge the use of ChatGPT (OpenAI) as a tool to assist with spell checking, grammatical corrections and editorial assistance during the preparation of this manuscript. All intellectual content, methods, codes, interpretations, and conclusions are the sole responsibility of the authors.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbinf.2026.1763403/full#supplementary-material>

References

- Afonine, P. V., Klaholz, B. P., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., et al. (2018). New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D. Struct. Biol.* 74, 814–840. doi:10.1107/s2059798318009324
- Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., et al. (2023). The gene ontology knowledgebase in 2023. *Genetics* 224 (1), iyad031. doi:10.1093/genetics/iyad031
- Andreeva, A., Kulesha, E., Gough, J., and Murzin, A. G. (2020). The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Research* 48 (D1), D376–D382. doi:10.1093/nar/gkz1064
- Apache Software Foundation (2025). Apache solr. Available online at: <https://solr.apache.org/> (Accessed October 1, 2025).
- Chen, S., McMullan, G., Faruqi, A. R., Murshudov, G. N., Short, J. M., Scheres, S. H., et al. (2013). High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* 135, 24–35. doi:10.1016/j.ultramic.2013.06.004
- Django Software Foundation (2025). Django: the web framework for perfectionists with deadlines. Available online at: <https://www.djangoproject.com/> (Accessed October 1, 2025).
- Duraisamy, A. K., Fonseca, N., Kleywegt, G. J., Patwardhan, A., and Morris, K. L. (2025). EMICSS: added-value annotations for EMDB entries. *Bioinforma. Adv.* 5 (1), vbaf203. doi:10.1093/bioadv/vbaf203
- Haak, L. L., Fenner, M., Paglione, L., Pentz, E., and Ratner, H. (2012). ORCID: a system to uniquely identify researchers. *Learn. Publishing* 25 (4), 259–264. doi:10.1087/20120404
- Hastings, J., Owen, G., Dekker, A., Ennis, M., Kale, N., Muthukrishnan, V., et al. (2016). ChEBI in 2016: improved services and an expanding collection of metabolites. *Nucleic Acids Research* 44 (D1), D1214–D1219. doi:10.1093/nar/gkv1031
- Highsoft (2025). Highcharts: interactive JavaScript charts. Available online at: <https://www.highcharts.com/> (Accessed October 1, 2025).
- Iudin, A., Korir, P. K., Somasundharam, S., Weyand, S., Cattavillo, C., Fonseca, N., et al. (2023). “EMPIAR: the electron microscopy public image archive,” 51. *Nucleic Acids Res.*, D1503–D1511. doi:10.1093/nar/gkac1062
- Jamali, K., Käll, L., Zhang, R., Brown, A., Kimanius, D., and Scheres, S. H. W. (2024). Automated model building and protein identification in cryo-EM maps. *Nature* 628, 450–457. doi:10.1038/s41586-024-07215-4
- Joseph, A. P., Lagerstedt, I., Patwardhan, A., Topf, M., and Winn, M. (2017). Improved metrics for comparing structures of macromolecular assemblies determined by 3D electron-microscopy. *J. Structural Biology* 199 (1), 12–26. doi:10.1016/j.jsb.2017.05.007
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (7873), 583–589. doi:10.1038/s41586-021-03819-2
- Kalvari, I., Nawrocki, E. P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K., Marz, M., et al. (2021). Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Research* 49 (D1), D192–D200. doi:10.1093/nar/gkaa1047
- Kimanius, D., Jamali, K., Wilkinson, M. E., Lövestam, S., Velazhahan, V., Nakane, T., et al. (2024). Data-driven regularization lowers the size barrier of cryo-EM structure determination. *Nat. Methods* 21, 1216–1221. doi:10.1038/s41592-024-02304-8
- Knox, C., Wilson, M., Klinger, C. M., Franklin, M., Oler, E., Wilson, A., et al. (2024). DrugBank 6.0: the DrugBank knowledgebase for 2024. *Nucleic Acids Research* 52 (D1), D1265–D1275. doi:10.1093/nar/gkad976
- Kühlbrandt, W. (2014). The resolution revolution. *Science* 343 (6178), 1443–1444.
- Lagerstedt, I., Moore, W. J., Patwardhan, A., Sanz-García, E., Best, C., Swedlow, J. R., et al. (2013). Web-based visualisation and analysis of 3D electron-microscopy data from EMDB and PDB. *J. Structural Biology* 184 (2), 173–181. doi:10.1016/j.jsb.2013.09.021
- Lawson, C. L., Kryshafovich, A., Adams, P. D., Afonine, P. V., Baker, M. L., Barad, B. A., et al. (2021). Cryo-EM model validation recommendations based on outcomes of the 2019 EMDataResource challenge. *Nat. Methods* 18 (2), 156–164. doi:10.1038/s41592-020-01051-w
- Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Biol. Crystallogr.* 75 (10), 861–877. doi:10.1107/S2059798319011471
- Meldal, B. H., Forner-Martinez, O., Costanzo, M. C., Dana, J., Demeter, J., Dumousseau, M., et al. (2015). The complex portal—an encyclopaedia of macromolecular complexes. *Nucleic Acids Research* 43 (D1), D479–D484. doi:10.1093/nar/gku975
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., et al. (2021). Pfam: the protein families database in 2021. *Nucleic Acids Research* 49 (D1), D412–D419. doi:10.1093/nar/gkaa913
- Patwardhan, A., Brandt, R., Butcher, S. J., Collinson, L., Gault, D., Grünwald, K., et al. (2017). Building bridges between cellular and molecular structural biology. *Elife* 6, e25835. doi:10.7554/eLife.25835
- Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., et al. (2023). InterPro in 2022. *Nucleic Acids Research* 51 (D1), D418–D427. doi:10.1093/nar/gkac993
- PDBE-KB consortium (2022). “PDBE-KB: collaboratively defining the biological context of structural data,” 50. *Nucleic Acids Res.*, D534–D542. doi:10.1093/nar/gkab988
- 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 (3), 328–334. doi:10.1038/s41592-020-0731-1
- Punjani, A., Rubinstein, J. L., Fleet, D. J., and Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* 14 (3), 290–296. doi:10.1038/nmeth.4169
- Research Organization Registry (ROR) (2024). Research Organization registry. *Res. Organ. Registry*. Available online at: <https://ror.org> (Accessed November 12, 2025).
- Rosonovski, S., Levchenko, M., Bhatnagar, R., Chandrasekaran, U., Faulk, L., Hassan, I., et al. (2024). Europe PMC in 2023. *Nucleic Acids Res.* 52 (D1), D1668–D1676. doi:10.1093/nar/gkad1085
- Sehnal, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., et al. (2021). Mol Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res.* 49 (W1), W431–W437. doi:10.1093/nar/gkab314
- Sillitoe, I., Bordin, N., Dawson, N., Waman, V. P., Ashford, P., Scholes, H. M., et al. (2021). CATH: increased structural coverage of functional space. *Nucleic Acids Research* 49 (D1), D266–D273. doi:10.1093/nar/gkaa1079
- Tagari, M., Newman, R., Chagoyan, M., Carazo, J. M., and Henrick, K. (2002). New electron microscopy database and deposition system. *Trends Biochemical Sciences* 27 (11), 589. doi:10.1016/s0968-0004(02)02176-x
- The UniProt Consortium (2025). “UniProt: the universal protein knowledgebase in 2025,” 53. *Nucleic Acids Res.*, D609–D617. doi:10.1093/nar/gkac1010
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* 50 (D1), D439–D444. doi:10.1093/nar/gkab1061
- Velankar, S., Van Ginkel, G., Alhroub, Y., Battle, G. M., Berrisford, J. M., Conroy, M. J., et al. (2016). PDBE: improved accessibility of macromolecular structure data from PDB and EMDB. *Nucleic Acids Research* 44 (D1), D385–D395. doi:10.1093/nar/gkv1047
- Wang, Z., Patwardhan, A., and Kleywegt, G. J. (2022). Validation analysis of EMDB entries. *Biol. Crystallogr.* 78 (5), 542–552. doi:10.1107/S205979832200328X
- Westbrook, J. D., Young, J. Y., Shao, C., Feng, Z., Guranovic, V., Lawson, C. L., et al. (2022). PDBx/mmCIF ecosystem: foundational semantic tools for structural biology. *J. Molecular Biology* 434 (11), 167599. doi:10.1016/j.jmb.2022.167599
- wwPDB consortium (2019). “Protein Data Bank: the single global archive for 3D macromolecular structure data, Nucleic Acids Research,” 47. *Nucleic Acids Res.*, D520–D528. doi:10.1093/nar/gky949
- wwPDB consortium (2024). *EMDB—The electron microscopy data bank, Nucleic Acids Research*, 52(D1). Oxford University Press, D456–D465.
- Young, J. Y., Westbrook, J. D., Feng, Z., Sala, R., Peisach, E., Oldfield, T. J., et al. (2017). OneDep: unified wwPDB system for deposition, biocuration, and validation of macromolecular structures in the PDB archive. *Structure* 25 (3), 536–545. doi:10.1016/j.str.2017.01.004
- Zdrzil, B., Felix, E., Hunter, F., Manners, E. J., Blackshaw, J., Corbett, S., et al. (2024). The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Research* 52 (D1), D1180–D1192. doi:10.1093/nar/gkad1004
- Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E., et al. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife*. Nov. 7, e42166. doi:10.7554/eLife.42166