

Evolving data standards for cryo-EM structures

Cite as: *Struct. Dyn.* **7**, 014701 (2020); doi: [10.1063/1.5138589](https://doi.org/10.1063/1.5138589)

Submitted: 14 November 2019 · Accepted: 7 January 2020 ·

Published Online: 24 January 2020



View Online



Export Citation



CrossMark

Catherine L. Lawson,^{1,a)}  Helen M. Berman,^{2,3}  and Wah Chiu^{4,5} 

AFFILIATIONS

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

²Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, USA

³Department of Biological Sciences and Bridge Institute, University of Southern California, Los Angeles, California 90089, USA

⁴Division of CryoEM and Biomaging, SLAC National Accelerator Laboratory, Stanford University, Stanford, California 94025, USA

⁵Department of Bioengineering, Stanford University, Stanford, California 94305, USA

Note: This article is part of the Special Issue: Transactions from the 69th Annual Meeting of the American Crystallographic Association: Data Best Practices: Current State and Future Needs.

^{a)}Author to whom correspondence should be addressed: cathy.lawson@rutgers.edu

ABSTRACT

Electron cryo-microscopy (cryo-EM) is increasingly being used to determine 3D structures of a broad spectrum of biological specimens from molecules to cells. Anticipating this progress in the early 2000s, an international collaboration of scientists with expertise in both cryo-EM and structure data archiving was established (EMDataResource, previously known as EMDataBank). The major goals of the collaboration have been twofold: to develop the necessary infrastructure for archiving cryo-EM-derived density maps and models, and to promote development of cryo-EM structure validation standards. We describe how cryo-EM data archiving and validation have been developed and jointly coordinated for the Electron Microscopy Data Bank and Protein Data Bank archives over the past two decades, as well as the impact of evolving technology on data standards. Just as for X-ray crystallography and nuclear magnetic resonance, engaging the scientific community via workshops and challenging activities has played a central role in developing recommendations and requirements for the cryo-EM structure data archives.

© 2020 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/1.5138589>

INTRODUCTION

Electron cryo-microscopy (cryo-EM) has very recently become a mainstream area of structural biology and medicine, enabling 3D visualization of a wide variety of biologically important complexes that were previously inaccessible to science. Early cryo-EM 3D density maps typically lacked atomic detail, yielding only the overall molecular shape, but could still sometimes be interpreted at a “pseudo-atomic” level via fitting of previously known coordinates or homology models [Fig. 1(a)].¹ Recent major technological advances now make it increasingly possible to directly visualize atomic details [Fig. 1(b)].^{2,3} These achievements were recognized by the award of the 2017 Chemistry Nobel Prize to cryo-EM pioneers Dubochet, Frank, and Henderson.⁴

The development of cryo-EM is directly reflected by the growth of cryo-EM structure depositions contributed worldwide to public data archives [Fig. 2(a)]. The archiving systems and underlying data standards supporting deposition, annotation, release, and validation of cryo-EM structures and the associated metadata describing cryo-EM

experiments have been developed over time to support this growth.⁵ We outline here the history of these systems and describe the process by which data standards have been developed, highlighting the role of engaging the scientific community to develop recommendations and requirements. The archiving systems and standards continue to evolve as technology drives the need for new descriptors and validation metrics.

CRYO-EM STRUCTURE DATA ARCHIVING

The Protein Data Bank (PDB), established in 1971 as a public archive for atomic coordinates of biological structures derived from X-ray crystallography,⁶ began accepting models derived from nuclear magnetic resonance spectroscopy (NMR) in 1988,⁷ and from electron microscopy (EM) and electron crystallography (EC) in 1990.⁸ In recognition of the fact that publicly available 3D density maps could accelerate discovery in structural biology and medicine, the Electron Microscopy Data Bank (EMDB) at the European Bioinformatics

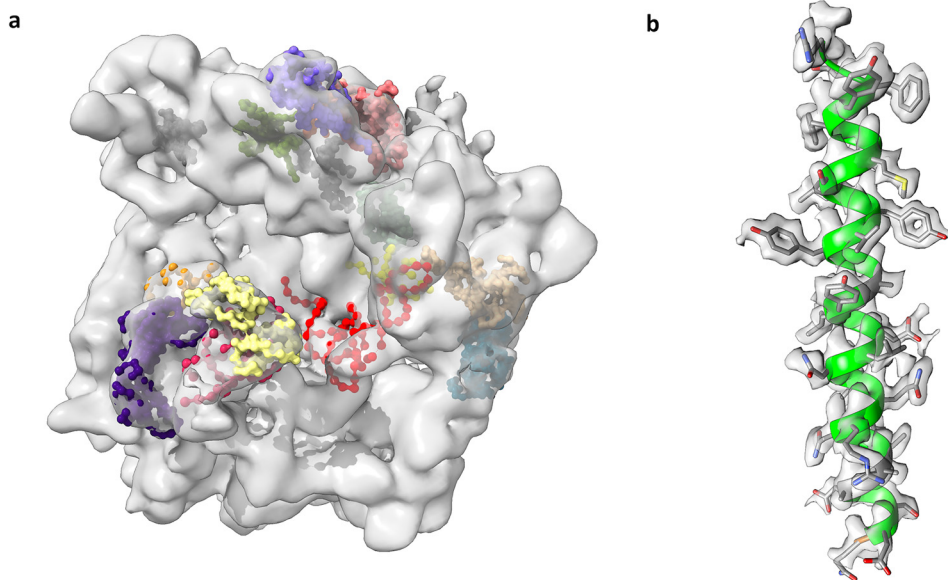


FIG. 1. Cryo-EM: contrasting early (2000) vs recent (2019). (a) Cryo-EM structure of the *E. coli* 70S ribosome determined by the Frank group at 11.5 Å, one of the first maps deposited in the EMDB archive (EMD-1003). It is shown here superimposed with manually fitted components deposited to PDB (1eg0).³⁷ (b) Helical segment of the 1.8 Å apoferritin map used as a target in the 2019 model challenge,³⁶ with a fitted model.

Institute (EBI) was launched in 2002 with support from the European Union.⁹ EMDB's launch was quickly followed by a pair of editorials in *Structure* and *Nature Structural Biology* encouraging electron microscopists to deposit their density maps.^{10,11} Similar to BioMagResBank (BMRB), which archives experimental data from NMR,¹² the EMDB accepts maps determined using any cryo-EM method, including single particle reconstruction with any symmetry, helical filament reconstruction, subtomogram averaging, tomography, and electron

crystallography, along with metadata describing the full experimental workflow (Fig. 3).

In 2006, scientists in the UK (EMDB) and USA [Research Collaboratory for Structural Bioinformatics (RCSB) and the National Center for Macromolecular Imaging (NCMI)] initiated a collaboration funded by National Institutes of Health (NIH) aimed to ensure that data archiving and validation standards for cryo-EM maps and models would be coordinated internationally.^{13,14} The project, formerly

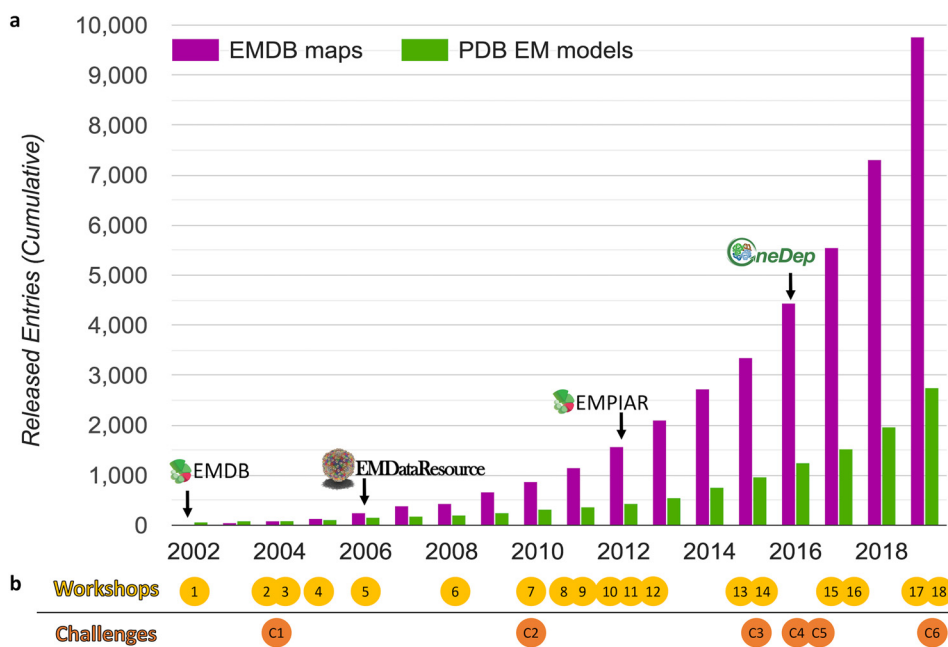


FIG. 2. Growth of data archives and community activity timeline. (a) Released map entries in EMDB and released EM model coordinate entries in PDB are shown, cumulative by year. Milestones (indicated with arrows) are described in the main text. Plot source: emdatabase.org. (b) Workshops (yellow circles) and challenges (orange circles) related to data standards and validation development are plotted according to the year they were held. Numbers within the circles correspond to Tables II and III rows.

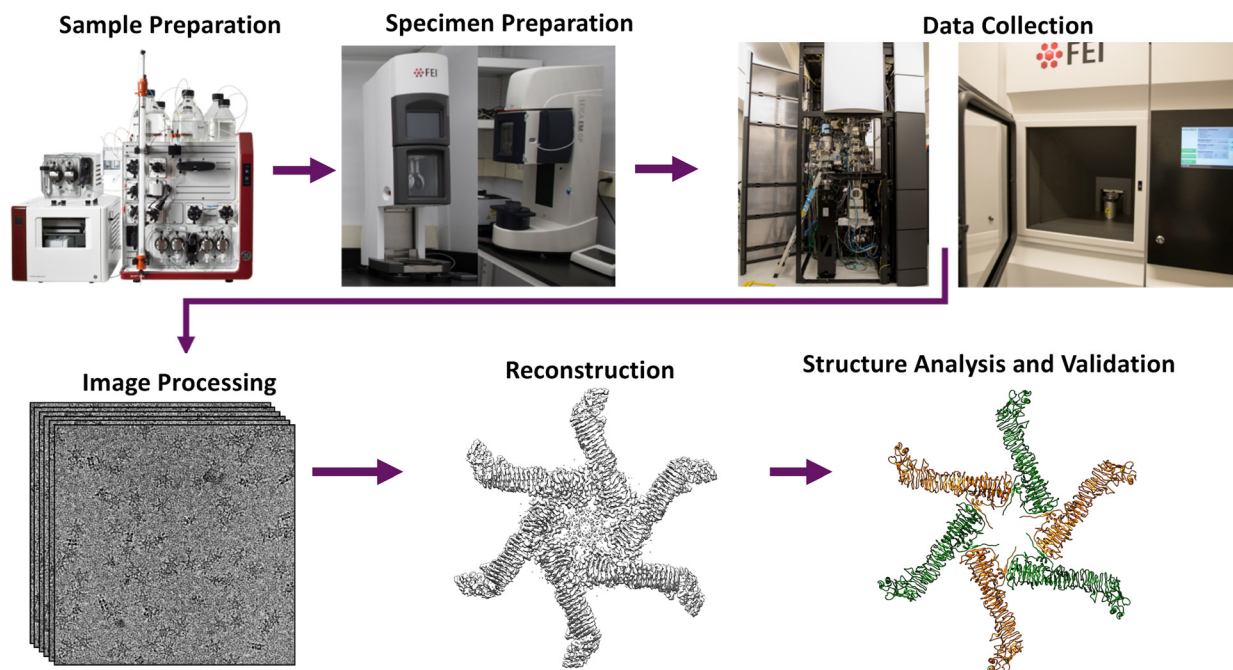


FIG. 3. Cryo-EM experimental workflow. The major steps for determining a structure using cryo-EM single particle reconstruction are shown. The specimen shown is *Helicobacter pylori* vacuolating cytotoxin A oligomer.³⁶

TABLE I. PDBx/mmCIF EM Dictionary used by wwPDB OneDep. As described in Fig. 4, all workflow metadata categories are collected by the OneDep system. Most categories are archived in both PDB and EMDB; asterisked categories are archived only in EMDB.

<i>Top level</i>	<i>Sample/specimen preparation</i>	<i>Image processing & reconstruction</i>	<i>Experimental data</i>
em_experiment em_software	em_buffer em_buffer_component em_crystal_formation em_embedding em_sample_support em_specimen em_staining em_vitrification	em_3d_reconstruction em_image_processing em_particle_selection em_volume_selection em_ctf_correction	em_map* em_structure_factors* em_layer_lines*
<i>Sample description</i> em_entity_assembly em_entity_assembly_molwt em_entity_assembly_naturalsource em_entity_assembly_recombinant em_virus_entity em_virus_natural_host em_virus_shell	em_fiducial_markers* em_focused_ion_beam* em_grid_pretreatment* em_high_pressure_freezing* em_shadowing* em_support_film* em_tomography* em_tomography_specimen* em_ultramicrotomy*	em_2d_crystal_entity em_3d_crystal_entity em_helical_entity em_single_particle_entity em_euler_angle_assignment* em_final_classification* em_start_model*	
<i>Data collection</i> em_diffraction em_diffraction_shell em_diffraction_stats em_image_recording em_image_scans em_imaging em_imaging_optics		<i>Structure analysis</i> em_3d_fitting em_3d_fitting_list em_fsc_curve*	

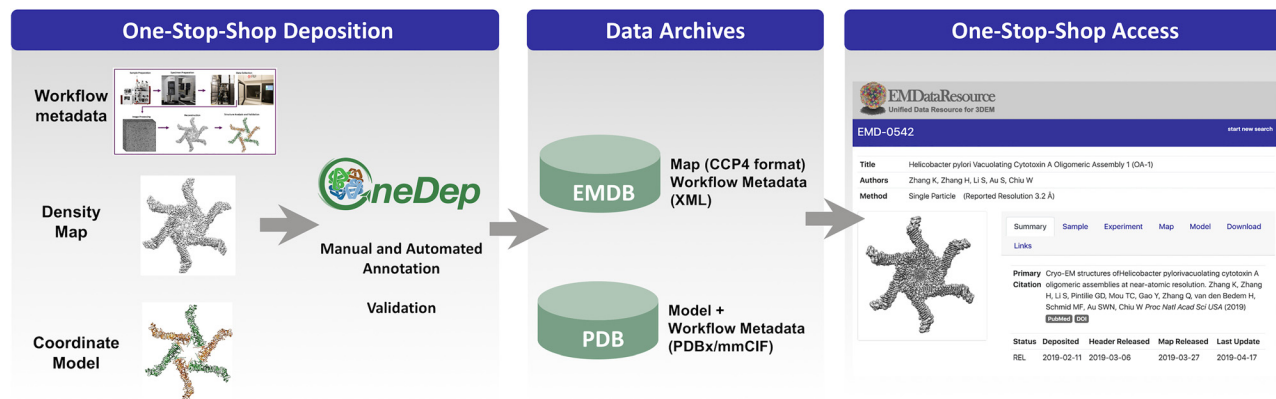


FIG. 4. Current systems for deposition, archiving, and accessing cryo-EM structures. Worldwide, every cryo-EM structure (map, experimental metadata, and optionally coordinate model) is deposited and processed through the wwPDB OneDep system (deposit.wwpdb.org), following the same annotation and validation workflow also used for X-ray crystallography and NMR structures.^{17,18} Map-only depositions yield an EMDB entry, while joint map+model depositions yield both EMDB and PDB entries. Workflow metadata collected in OneDep are passed to both EMDB and PDB. EMDB holds all workflow metadata while PDB holds a subset of the metadata; see [Table I](#). The PDB and EMDB archives are accessible by FTP and rsync at wwPDB mirror sites in the US, UK, and Japan. Released cryo-EM structure data from both archives can be accessed via EMDDataResource, EMDB, and wwPDB partner websites.

known as EMDDataBank, was recently rebranded as EMDDataResource (EMDR; emdataresource.org). The EMDR project website serves as a global resource for cryo-EM structure data archiving and retrieval, news, events, software tools, data standards, validation methods, and community challenges.

EMDep, designed and implemented at EBI, was the first system designed to collect and annotate maps and associated metadata for EMDB.⁹ In 2008, the EMDR team created a joint map+model deposition system for cryo-EM structures by connecting EMDep with AutoDep and ADIT (AutoDep Input Tool), the PDB data collection

TABLE II. Cryo-EM community data archiving and validation workshops.

#	Year	Title/location	Organizers	Description	Key outcomes
1	2002	IIMS Workshop, UK	Kim Henrick, José-María Carazo, Stephen Fuller	Promote software development in the field of 3DEM molecular structure determination	Guidelines and release policies for the new EMDB archive. ^{10,11,39} Deposition system to collect EMDB data and maps ⁹
2	2004	3DEM Developers workshop, UK	Kim Henrick	Developer review of tools and software practices used in the field of cryoEM	Priorities developed for EMDB including electron tomography, PDB-EMDB cross-referencing, lossless map conversion, review of community map standards and conventions
3	2004	Cryo-EM Structure Deposition Workshop, NJ, USA	Helen Berman, Wah Chiu, Michael Rossmann	Develop community consensus on data items needed for deposition of maps and atomic models derived from cryoEM	Need for deposition one-stop-shop articulated. Recommendations incorporated in extended EM data dictionary ¹³
4	2005	3DEM Developers Workshop, UK	Kim Henrick	Introduced EM data dictionary to software developers to facilitate its integration into 3DEM packages and electronic notebooks	Agreement to adopt a common set of conventions for maps ²²
5	2006	CryoEM Standards Task Force, TX, USA	Wah Chiu, David Belnap, José-María Carazo	Gather cryoEM map conventions and formats with associated metadata used by different developers	Key conventions (e.g., Euler angles) were evaluated for interoperability and conversion tools were created ^{40,41}
6	2008	Electron Crystallography Data	Wah Chiu, Cathy Lawson	Gather expert advice on metadata items in the EM dictionary	Recommendations incorporated into EM data dictionary ¹³

TABLE II. (Continued.)

#	Year	Title/location	Organizers	Description	Key outcomes
7	2010	Model Workshop, CA, USA EM Validation Task Force, NJ, USA	Helen Berman, Wah Chiu, Gerard Kleywegt, Cathy Lawson	pertaining to electron crystallography Expert review of potential validation standards for maps and models produced by 3DEM reconstruction	Recommendations summarized in white paper, ²⁵ and implemented in OneDep validation reports ^{14,18}
8	2011	Model Challenge Workshop, HI, USA	Steve Ludtke, Cathy Lawson, Gerard Kleywegt, Helen Berman, Wah Chiu	Computational groups described and compared tools they used to model a selected set of published cryoEM structures	Results published in <i>Biopolymers</i> journal special issue ⁴²
9	2011	Data Management Challenges in 3DEM, UK	Ardan Patwardhan, Gerard Kleywegt, Jason Swedlow	Gather expert advice on key topics in data archiving and validation for 3DEM data, including data model, validation, raw-data archiving	Recommendations summarized in white paper. ²³ Web-based visualization tools developed. ²⁸ EMPIAR raw data archive created. ¹⁵ Extended data model implemented ¹⁷
10	2012	3DEM Modeling Workshop, TX, USA	Wah Chiu	Current challenges in creating and validating coordinate models built into cryo-EM maps	Recommendations gathered for modeling and validation standards and future model challenges
11	2012	3D Cellular Context for the Macromolecular World, UK	Ardan Patwardhan, Gerard Kleywegt, Jason Swedlow	Discussions on data archiving and validation for emerging 3D cellular imaging techniques	Recommendations summarized in white paper ²⁴
12	2012	Instruct Image Processing Center Developer Workshop, Spain	José-María Carazo	Effort to standardize information exchange in single particle reconstruction and to establish algorithm benchmarking	CTF benchmarking challenge ⁴³ and EMX convention for CTF and single-particle parameters developed ⁴⁴
13	2015	CryoEM Model Validation Workshop, MA, USA	Wah Chiu, Cathy Lawson, Paul Adams	Modeling experts met to present and discuss challenges in modeling into cryoEM maps	Gathered recommendations ⁵ directly used in development of 2016 Model Challenge
14	2015	Building Bridges between Cellular and Molecular Structural Biology, UK	Ardan Patwardhan, Gerard Kleywegt	Expert discussions on how to integrate structural data from a diverse range of public archives covering cellular and molecular structural biology	Recommendations to develop tools/file formats for map segmentation, and tools to support biological structure annotation described in white paper ⁴⁵
15	2017	Model Challenge Assessment, LA, USA	Wah Chiu, Cathy Lawson, Paul Adams	First pass analyses of models and data submitted to the 2016 Model Challenge	Recommended metrics implemented on model challenge comparison website ⁴⁶
16	2017	CryoEM Structure Joint Challenges Workshop, CA, USA	Cathy Lawson, Wah Chiu	Joint review of the 2016 Map and Model Challenge activities	Results described in <i>Journal of Structural Biology</i> special issue; ³⁰ also featured in <i>Nature Methods</i> editorial ⁴⁷
17	2019	Frontiers in cryo-EM Validation, UK	Gerard Kleywegt, Garib Murshudov, Elena Orlova, Ardan Patwardhan, Alan Roseman, Peter Rosenthal, Maya Topf, Martyn Winn	Discuss current and future community needs/challenges for validation tools to support maps and models from single-particle analysis	Meeting featured in 2018 <i>Nature</i> editorial. ⁴⁸ Recommendations white paper is in preparation
18	2019	Model Metrics Workshop, CA, USA	Cathy Lawson, Wah Chiu	Review modeling processes and assessment results of the 2019 Model Metrics Challenge and plan for future events	Recommendations editorial and full event manuscript are in preparation

TABLE III. Cryo-EM community challenge activities.

#	Year	Event	Organizing group(s)	Description	Key outcomes
C1	2004	Particle Picking Challenge	National Resource for Automated Molecular Microscopy (La Jolla)	Compare diverse particle selection algorithms	Algorithms from 12 developer groups were compared and contrasted ⁴⁹
C2	2010	Model Challenge	EMDataResource, NCMI	Computational groups were asked to apply their tools to a selected set of published cryoEM structures.	Results published in <i>Biopolymers</i> journal special issue ⁴²
C3	2015	CTF Challenge	Instruct Image Processing Center (Madrid), NCMI	Evaluate community/algorithm abilities to estimate key parameters of EM Contrast Transfer Function for a wide set of experimental conditions	CTF benchmarking challenge summary and results published ⁴³
C4	2016	Map Challenge	EMDataResource	Establish benchmark datasets, develop best practices, evolve criteria for validation of 3DEM reconstructions	Results described in <i>Journal of Structural Biology</i> special issue; ³⁰ also featured in <i>Nature Methods</i> Editorial ⁴⁷
C5	2016	Model Challenge	EMDataResource	Establish benchmark datasets, develop best practices, evolve criteria for validation of 3DEM map-derived models	Results described in <i>Journal of Structural Biology</i> special issue; ³⁰ also featured in <i>Nature Methods</i> Editorial ⁴⁷
C6	2019	Model Metrics Challenge	EMDataResource	Identify metrics most suitable for evaluating/comparing fit of atomic coordinate models into cryo-EM maps in 1.8–3.0 Å resolution range	Recommendations editorial and full event manuscript are in preparation

systems at the EBI and RCSB sites.¹³ The system that was implemented enabled a one stop shop for cryo-EM model and map depositions. Joint curation ensured that maps and models were deposited at the same physical scale and in the same coordinate frame. Journals that publish cryo-EM structures began to require authors to deposit maps to EMDB and models to PDB. This system supported the processing and release of nearly 4000 maps and 1000 models over a nine-year period (2008–2015).¹⁴

In 2012, the Electron Microscopy Public Image Archive (EMPIAR) was established at EBI.¹⁵ Supported by the UK Medical Research Council and UK Biotechnology & Biological Sciences Research Council, EMPIAR enables cryo-EM scientists to archive and share raw images and intermediate data files associated with their maps deposited to EMDB. Making recently collected image data broadly available has multiple benefits, including accelerating development of reconstruction software, and enriching resources for cryo-EM scientists in training. EMPIAR has its own deposition and curation system, but accesses metadata from the related EMDB entry. Individual entry storage sizes can be up to 15 TB. Approximately 4% of EMDB entries deposited since 2012 have associated EMPIAR entries.

The Worldwide PDB (wwPDB) is the global organization that manages the PDB archive.¹⁶ In 2016, deposition, annotation, and release of cryo-EM structure maps and models were migrated to the wwPDB OneDep system (Fig. 4), using requirements that were initiated and developed by EMDR.¹⁷ At that time, it became mandatory to deposit maps to EMDB for all cryo-EM models deposited to PDB. In addition, structure validation reports, which can be provided by depositors in an official PDF format to journal editors and reviewers as part

of manuscript review, began to be produced for all cryo-EM structures.¹⁸

CREATING A DATA DICTIONARY

The foundation of any data repository is its data representation scheme. Based in part on the International Union of Crystallography dictionary for small molecule crystallography Crystallographic Information File (CIF),¹⁹ the Macromolecular Crystallographic Information File (mmCIF) was developed in the 1990s to support rich data content for the macromolecular crystallographic experiment and its results, with precise data type definitions, logical groupings for related data items, explicit parent-child relationships, enumerations for controlled vocabulary, extensibility, and many other features embedded in a computer-readable format.²⁰ This dictionary is now the Master Format for the PDB. Particularly relevant for cryo-EM, very large complexes are readily represented, since mmCIF has no limits on the number of atoms or polymer chains.

Following the lead of the crystallographic community, an mmCIF extension dictionary containing data terms for cryo-EM experiments was drafted jointly in the early 2000s based on requirements provided by the cryo-EM community. The dictionary was vetted and expanded by the scientific community via multiple workshops, and subsequently integrated by EMDR into the PDBx/mmCIF dictionary for use in the hybrid joint map+model deposition system.¹³ In 2015, based on feedback from additional workshops, the EMDR team further modified and expanded the dictionary in several ways. Hierarchical descriptions of complex specimens were enabled, and experimental descriptions

for each of the cryo-EM methods were extended.⁵ The >500 term EM dictionary (Table 1) is now the basis for cryo-EM depositions to both EMDB and PDB in the wwPDB OneDep system. The dictionary continues to be updated regularly to support the evolving needs of the scientific community.

GATHERING COMMUNITY REQUIREMENTS

Developing a trusted scientific data repository requires careful attention to the interplay among science, technology, and community interest.²¹ Workshops and Challenges are two types of community outreach activities that are effective in bringing these three elements together; both have been employed multiple times to move EM data and validation standard development forward [Fig. 2(b)]. Workshops (typically 2–3 days) enable groups of experts to review current practices and develop recommendations, while Challenges (taking place over several months to a year) provide forums for experts to exercise and demonstrate current workflows and test novel procedures. Challenges can incorporate one or more workshops for planning or results review. Tables II and III list and summarize goals and outcomes of 18 international workshops and six challenges held over the past two decades. Below we provide additional descriptions of selected activities, as well as a montage of workshop photos (Fig. 5).

EM extension dictionary development

The main goal of the 2004 *Cryo-EM Structure Deposition Workshop* [Fig. 5(a)], attended by ~30 scientists including cryo-EM,

image processing, crystallography, database, funding agency, and journal representatives, was to develop a global community consensus on data items for deposition of density maps and atomic models derived from cryo-EM studies. Terms were reviewed category-by-category in two focus groups, and recommendations for revisions and extensions were obtained (Fig. 6). Furthermore, participants unanimously requested a “one-stop shop” for deposition and retrieval of the cryo-EM map and model data. Following the workshop, the dictionary was further revised with follow-up input from attendees. The resulting dictionary was presented at the 2005 3DEM Gordon Research Conference, and EMDR’s project website became the requested one-stop-shop portal.

The EM extension dictionary was next reviewed by software developers at the 2005 *3DEM Developers Workshop* to facilitate its integration with major 3DEM packages and electronic notebook systems. There were two important outcomes: (a) the draft dictionary was unanimously accepted by the participants and (b) a set of proposed conventions for describing EM micrographs and density maps was developed.²² The conventions enable a standardized approach to image interpretation and presentation, with recommended units for common parameters, rotation and symmetry notations, and common sense principles such as “objects should have overall positive density” (early image correction procedures sometimes generated objects darker than their background depending on image processing and display software). The conventions were subsequently incorporated into the EM extension dictionary to facilitate representation of map-related data items in PDB and EMDB.



FIG. 5. Workshop Participant Photos. (a) 2004 Cryo-EM structure deposition workshop; (b) 2010 EM Validation Task Force (EM VTF) Workshop; (c) 2011 model challenge workshop; (d) 2017 joint challenges workshop. Image in (d) Reprinted with permission from C. L. Lawson and W. Chiu, *J. Struct. Biol.* **204**(3), 523–526 (2018). Copyright 2018 Elsevier.

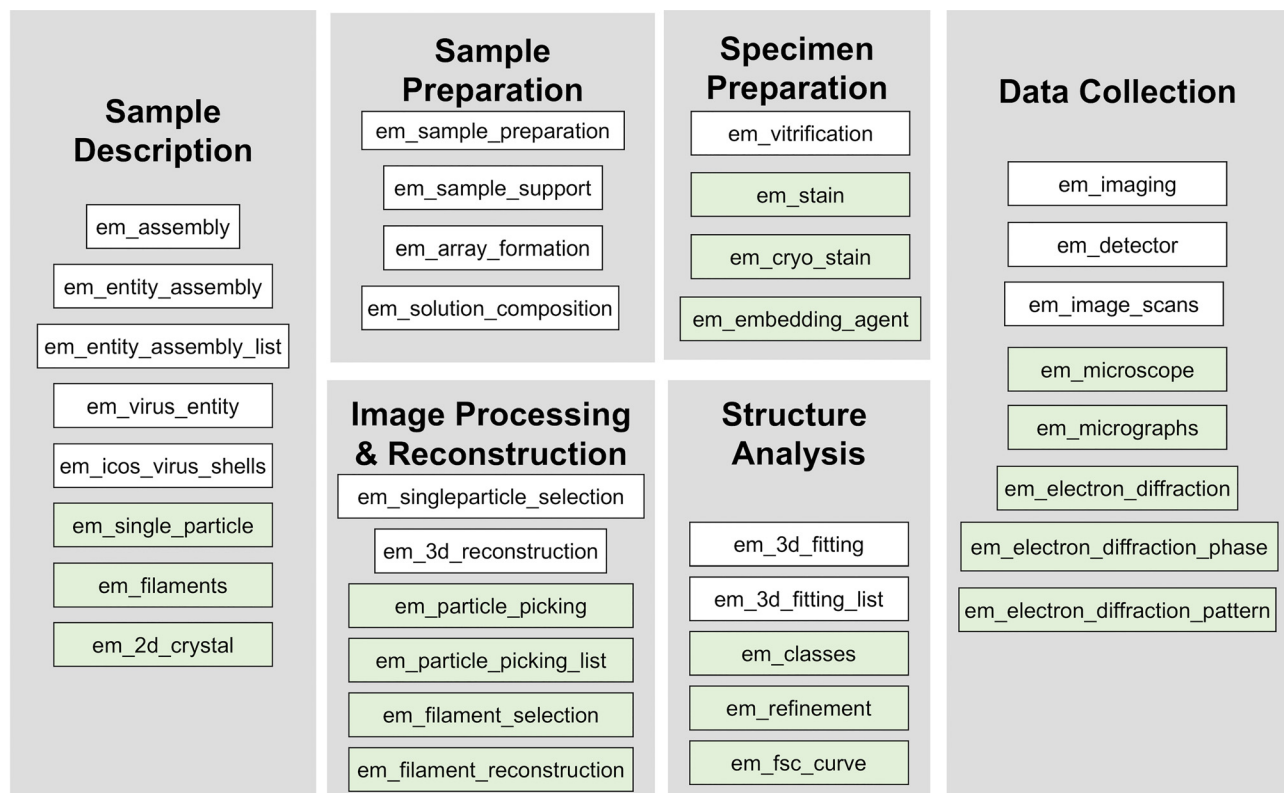


FIG. 6. Overall structure of the EM extension data dictionary circa 2005. New categories of data items recommended by participants of the 2004 Cryo-EM Structure Deposition Workshop are shown in green.

Data standards for cryo-EM structures were further addressed at the 2011 *Data Management Challenges in 3D Electron Microscopy Workshop*²³ and the 2015 *Building Bridges between Cellular and Molecular Structural Biology Workshop*.²⁴ Needs for hierarchical sample description as well as extensions to cryo-EM experimental sub-method descriptions were recognized. A future archival segmentation file format, for which requirements were gathered at the 2015 meeting, will make use of the hierarchy, enabling map regions to be connected with biological annotations.^{5,24}

Developing validation standards

At the 2010 *EM Validation Task Force (EM VTF) Workshop* [Fig. 5(b)], an international group of experts explored how to assess cryo-EM maps, models, and other data deposited into EMDB and PDB. For maps, participants recognized a critical need to develop standards for assessing map resolution and accuracy. They recommended establishing two fully independent image datasets at the outset for evaluating resolution by Fourier Shell Correlation (FSC); at the time, this was not typically done, but it is now the standard procedure. However, they also advised that maps still be carefully inspected to ensure that the resolution estimate by FSC is in accordance with the map's visible features.

The EM VTF's 2012 white paper notably called for the scientific community to develop new criteria for the evaluation of maps and for the evaluation of fit of the model to the experimental map density.²⁵

In contrast, in 2011 the VTF for X-ray crystallography published a comprehensive and detailed set of recommendations to validate structures and experimental data determined using X-ray crystallography.²⁶ The difference reflects the fact that cryo-EM is still a rapidly evolving field.

Validation standards and raw image data archiving were additional topics of discussion at the 2011 *Data Management Challenges in 3D Electron Microscopy Workshop*.²³ Several services were developed and implemented at EBI in response to workshop recommendations. The EMPIAR raw data archive was created,¹⁵ and stand-alone FSC and tilt-pair servers were developed for depositors to validate their cryo-EM maps.^{5,27} In addition, Visual Analysis web pages were designed to display an informative series of images and plots for every EMDB entry, and to help users assess data quality of released cryo-EM maps and models.^{28,29}

Two EMDR-sponsored challenges subsequently aimed to address the 2010 EM VTF's call for improved metrics to evaluate both maps and fit of models to experimental data (2016 *Map and Model Challenges*). Following the 2017 *Joint Challenges Workshop* at Stanford, which had over 90 participants [Fig. 5(d)]; key results and recommendations were collated into a virtual special issue of the *Journal of Structural Biology* published in December 2018.³⁰

The Map Challenge provided a unique forum for critically evaluating the standard method for estimating map resolution by FSC (Fig. 7, inset). A key observation was that as currently practiced, the

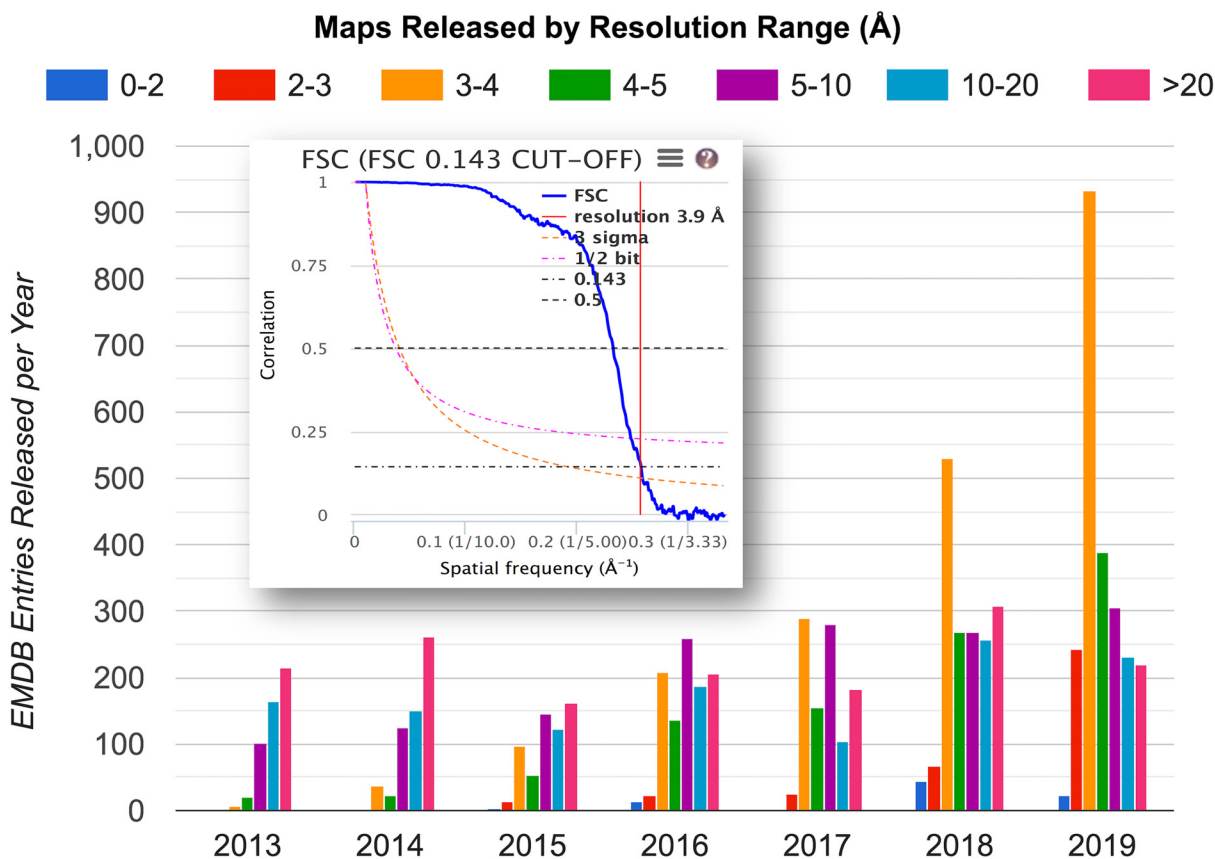


FIG. 7. Changing cryo-EM resolution landscape. Annual distribution of depositor-reported resolution for map entries released into EMDB. The sharp increase at 2–4 Å resolution is a direct consequence of the recent advances in image detection and processing.³ Inset: example Fourier Shell Correlation (FSC) plot, which is the current standard for estimating map resolution.²⁵ The correlation between two independent half-map reconstructions (blue curve) falls with decreasing spatial frequency; the resolution estimate (in this case 3.9 Å) is read at FSC = 0.143 (dash-dotted horizontal line). Plot source: emdatabase.org. Inset FSC plot source: EMDB visual analysis.²⁸

procedure is not sufficiently standardized: a number of different variables (e.g., map box size, voxel size, filtering and masking practice, and threshold value for interpretation) can substantially impact the outcome.³¹ As a result, different expert practitioners can arrive at different resolution estimates for the same level of map details. For example, two of the apoferritin maps submitted to the challenge had practitioner-estimated resolutions of 3.1 Å and 3.5 Å, respectively, though they were indistinguishable by eye. A direct conclusion is that any “reported-resolution”-based search or ranking for maps or associated models will have limited reliability. In follow-up discussions at the 2019 *Frontiers in Cryo-EM Validation Workshop*, one suggestion made was to have the archives independently estimate resolution by FSC from deposited unmasked, minimally filtered half-maps. This procedure would likely make comparisons between maps less susceptible (though not completely impervious) to variations in practitioner practice.

The 2017 *Joint Challenges Workshop* also sparked lively discussions about the potential for model-based metrics to estimate not only model quality, but also to provide one or more independent measures of map resolvability. Several procedures of this type have been proposed and tested. EMRinger evaluates whether density peaks in the

map fall within the possible rotameric configuration for the carbon-β atom in a side chain.³² Other procedures have been developed to measure map quality. For example, Z-scores capture how much larger the cross-correlations score (CCS) is for atoms in such features at their placed location compared to the CCS at displaced positions.^{33–35} Another recently devised experimental metric, Q-score, measures resolvability of the individual atom(s) in reference to the model.³⁶

Changing validation goals

Looking at the distribution of reported resolution of maps released into EMDB annually over the past few years (Fig. 7), one can readily see a striking sharp recent increase in maps in the 2–4 Å range. This development is a direct result of recent technological improvements, and it changes the “goal-posts” for developing validation methods, adding urgency to the need for metrics to validate structures at near-atomic to atomic resolution.

The 2019 *Model Metrics Challenge* and associated 2019 *Model Metrics Workshop* were designed with the goal of evaluating metrics for map-model fit of moderately high-resolution maps (3.1–1.8 Å). A full write-up will be published elsewhere, and we note two findings

here. First, the new metrics that by some means combine both model and map quality (e.g., EMRinger and Q-score) appear to be quite useful for ranking sets of structures. Second, refined Atomic Displacement Parameters (ADPs), which were included in about half of the models submitted by challenge participants, could modestly improve fit of the model to the map, particularly for the highest resolution (1.8 Å) target map. The meaning of refined ADPs/B-factors in the context of a cryo-EM density map is less clear. Best practices (e.g., to avoid overfitting) will need to be investigated.

WHERE WE ARE, WHAT'S NEXT

The initial EM validation report format released in 2016 focused on assessment of model geometry for PDB entries.¹⁸ As will be reported in more detail in a future publication, additional sections covering map analysis and visualization and map-model fit analysis and visualization will become available to EMDB and PDB depositors by early 2020. The Visual Analysis web pages hosted at EBI since 2012²⁸ have served as a test-bed for the development of the new features, which will include (a) several types of orthogonal images of the deposited map and map superimposed with model; (b) FSC curves to support depositor-reported map resolution; and (c) map-model fit statistics via “atom inclusion,” the percentage of modeled atoms falling inside a map at its recommended contour level. The new features will enable scientists (depositors, annotators, journal editors, and manuscript reviewers) to make initial assessments of map features, map quality, and map-model fit, bypassing the need to first download/view files in a graphics program.

A planned meeting in January 2020 at EBI organized by wwPDB will bring together cryo-EM and data archiving experts to discuss the current state of data archiving for cryo-EM structures derived from the single-particle reconstruction method, and to solicit recommendations on what data should be included and/or made mandatory in depositions and associated validation reports. The following points might be considered as part of the deliberations:

- Can estimation of map resolution be better standardized across the community? This would enable fairer comparisons among maps determined in different laboratories and using different software packages.
- Additional metrics (beyond atom inclusion) are available that describe map-model fit, including several novel procedures that effectively yield a joint assessment of map and model quality in a broad resolution range. How should map-model fit be reported as part of a structure determination and in a joint map+model deposition?
- What best practice recommendations can be made for refinement of ADPs in cryo-EM models at different resolutions?
- How should we evaluate multiple structures determined from a single specimen that may have variable quality and resolution?

ACKNOWLEDGMENTS

EMDataResource is funded by the U.S. National Institutes of Health/National Institute of General Medical Science, No. R01GM079429-12. We thank current and past EMDR colleagues for their contributions to data standards development, with special recognition to former and current directors of the EMDB archive

including Kim Henrick, Gerard Kleywegt, and Ardan Patwardhan. We are also tremendously grateful to the cryo-EM community for their enthusiastic participation and support for data standards development.

REFERENCES

- ¹W. Chiu, M. L. Baker, W. Jiang, M. Dougherty, and M. F. Schmid, “Electron cryomicroscopy of biological machines at subnanometer resolution,” *Structure* **13**(3), 363–372 (2005).
- ²K. R. Vinothkumar and R. Henderson, “Single particle electron cryomicroscopy: Trends, issues and future perspective,” *Q. Rev. Biophys.* **49**, e13 (2016).
- ³W. Kuhlbrandt, “Biochemistry. The resolution revolution,” *Science* **343**(6178), 1443–1444 (2014).
- ⁴Y. Cheng, R. M. Glaeser, and E. Nogales, “How cryo-EM became so hot,” *Cell* **171**(6), 1229–1231 (2017).
- ⁵A. Patwardhan and C. L. Lawson, “Databases and archiving for CryoEM,” *Methods Enzymol.* **579**, 393–412 (2016).
- ⁶“Crystallography: Protein data bank,” *Nat. New Biol.* **233**(42), 223–223 (1971).
- ⁷P. C. Driscoll, A. M. Gronenborn, L. Beress, and G. M. Clore, “Determination of the three-dimensional solution structure of the antihypertensive and antiviral protein BDS-I from the sea anemone *Anemonia sulcata*: A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing,” *Biochemistry* **28**(5), 2188–2198 (1989).
- ⁸R. Henderson, J. M. Baldwin, T. A. Ceska, F. Zemlin, E. Beckmann, and K. H. Downing, “Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy,” *J. Mol. Biol.* **213**(4), 899–929 (1990).
- ⁹K. Henrick, R. Newman, M. Tagari, and M. Chagoyen, “EMDep: A web-based system for the deposition and validation of high-resolution electron microscopy macromolecular structural information,” *J. Struct. Biol.* **144**(1–2), 228–237 (2003).
- ¹⁰S. D. Fuller, “Depositing electron microscopy maps,” *Structure* **11**(1), 11–12 (2003).
- ¹¹Editorial, A database for ‘em,’ *Nat. Struct. Biol.* **10**(5), 313 (2003).
- ¹²E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. K. Wenger, H. Yao, and J. L. Markley, “BioMagResBank,” *Nucl. Acids Res.* **36**, D402–408 (2007).
- ¹³C. L. Lawson, M. L. Baker, C. Best, C. Bi, M. Dougherty, P. Feng, G. van Ginkel, B. Devkota, I. Lagerstedt, S. J. Ludtke, R. H. Newman, T. J. Oldfield, I. Rees, G. Sahni, R. Sala, S. Velankar, J. Warren, J. D. Westbrook, K. Henrick, G. J. Kleywegt, H. M. Berman, and W. Chiu, “EMDataBank.org: Unified data resource for CryoEM,” *Nucl. Acids Res.* **39**, D456–464 (2011).
- ¹⁴C. L. Lawson, A. Patwardhan, M. L. Baker, C. Hryc, E. S. Garcia, B. P. Hudson, I. Lagerstedt, S. J. Ludtke, G. Pintilie, R. Sala, J. D. Westbrook, H. M. Berman, G. J. Kleywegt, and W. Chiu, “EMDataBank unified data resource for 3DEM,” *Nucl. Acids Res.* **44**(D1), D396–403 (2016).
- ¹⁵A. Iudin, P. K. Korir, J. Salavert-Torres, G. J. Kleywegt, and A. Patwardhan, “EMPIAR: A public archive for raw electron microscopy image data,” *Nat. Methods* **13**(5), 387–388 (2016).
- ¹⁶wwPDB Consortium, “Protein Data Bank: The single global archive for 3D macromolecular structure data,” *Nucl. Acids Res.* **47**(D1), D520–D528 (2019).
- ¹⁷J. Y. Young, J. D. Westbrook, Z. Feng, R. Sala, E. Peisach, T. J. Oldfield, S. Sen, A. Gutmanas, D. R. Armstrong, J. M. Berrisford, L. Chen, M. Chen, L. Di Costanzo, D. Dimitropoulos, G. Gao, S. Ghosh, S. Gore, V. Guranovic, P. M. S. Hendrickx, B. P. Hudson, R. Igarashi, Y. Ikegawa, N. Kobayashi, C. L. Lawson, Y. Liang, S. Mading, L. Mak, M. S. Mir, A. Mukhopadhyay, A. Patwardhan, I. Persikova, L. Rinaldi, E. Sanz-Garcia, M. R. Sekharan, C. Shao, G. J. Swaminathan, L. Tan, E. L. Ulrich, G. van Ginkel, R. Yamashita, H. Yang, M. A. Zhuravleva, M. Quesada, G. J. Kleywegt, H. M. Berman, J. L. Markley, H. Nakamura, S. Velankar, and S. K. Burley, “OneDep: Unified wwPDB system for deposition, biocuration, and validation of macromolecular structures in the PDB archive,” *Structure* **25**(3), 536–545 (2017).
- ¹⁸S. Gore, E. Sanz Garcia, P. M. S. Hendrickx, A. Gutmanas, J. D. Westbrook, H. Yang, Z. Feng, K. Baskaran, J. M. Berrisford, B. P. Hudson, Y. Ikegawa, N. Kobayashi, C. L. Lawson, S. Mading, L. Mak, A. Mukhopadhyay, T. J.

- Oldfield, A. Patwardhan, E. Peisach, G. Sahni, M. R. Sekharan, S. Sen, C. Shao, O. S. Smart, E. L. Ulrich, R. Yamashita, M. Quesada, J. Y. Young, H. Nakamura, J. L. Markley, H. M. Berman, S. K. Burley, S. Velankar, and G. J. Kleywegt, "Validation of structures in the Protein Data Bank," *Structure* **25**(12), 1916–1927 (2017).
- ¹⁹S. R. Hall, F. H. Allen, and I. D. Brown, "The crystallographic information file (Cif)—A new standard archive file for crystallography," *Acta Crystallogr., Sect. A* **47**, 655–685 (1991).
- ²⁰P. M. D. Fitzgerald, J. D. Westbrook, P. E. Bourne, B. McMahon, K. D. Watenpugh, and H. M. Berman, in *International Tables for Crystallography G. Definition and Exchange of Crystallographic Data*, edited by S. R. Hall and B. McMahon (Springer, Dordrecht, The Netherlands, 2005), pp. 295–443.
- ²¹H. M. Berman, C. L. Lawson, B. Vallat, and M. J. Gabanyi, "Anticipating innovations in structural biology," *Q. Rev. Biophys.* **51**, e8 (2018).
- ²²J. B. Heymann, M. Chagoyen, and D. M. Belpap, "Common conventions for interchange and archiving of three-dimensional electron microscopy information in structural biology," *J. Struct. Biol.* **151**(2), 196–207 (2005).
- ²³A. Patwardhan, J. M. Carazo, B. Carragher, R. Henderson, J. B. Heymann, E. Hill, G. J. Jensen, I. Lagerstedt, C. L. Lawson, S. J. Ludtke, D. Mastronarde, W. J. Moore, A. Roseman, P. Rosenthal, C. O. Sorzano, E. Sanz-Garcia, S. H. Scheres, S. Subramaniam, J. Westbrook, M. Winn, J. R. Swedlow, and G. J. Kleywegt, "Data management challenges in three-dimensional EM," *Nat. Struct. Mol. Biol.* **19**(12), 1203–1207 (2012).
- ²⁴A. Patwardhan, A. Ashton, R. Brandt, S. Butcher, R. Carzaniga, W. Chiu, L. Collinson, P. Doux, E. Duke, M. H. Ellisman, E. Franken, K. Grunewald, J. K. Heriche, A. Koster, W. Kuhlbrandt, I. Lagerstedt, C. Larabell, C. L. Lawson, H. R. Saibil, E. Sanz-Garcia, S. Subramaniam, P. Verkade, J. R. Swedlow, and G. J. Kleywegt, "A 3D cellular context for the macromolecular world," *Nat. Struct. Mol. Biol.* **21**(10), 841–845 (2014).
- ²⁵R. Henderson, A. Sali, M. L. Baker, B. Carragher, B. Devkota, K. H. Downing, E. H. Egelman, Z. Feng, J. Frank, N. Grigorieff, W. Jiang, S. J. Ludtke, O. Medalia, P. A. Penczek, P. B. Rosenthal, M. G. Rossmann, M. F. Schmid, G. F. Schroder, A. C. Steven, D. L. Stokes, J. D. Westbrook, W. Wriggers, H. Yang, J. Young, H. M. Berman, W. Chiu, G. J. Kleywegt, and C. L. Lawson, "Outcome of the first electron microscopy validation task force meeting," *Structure* **20**(2), 205–214 (2012).
- ²⁶R. J. Read, P. D. Adams, W. B. Arendall, A. T. Brunger, P. Emsley, R. P. Joosten, G. J. Kleywegt, E. B. Krissinel, T. Lutteke, Z. Otwinowski, A. Perrakis, J. S. Richardson, W. H. Sheffler, J. L. Smith, I. J. Tickle, G. Vriend, and P. H. Zwart, "A new generation of crystallographic validation tools for the Protein Data Bank," *Structure* **19**(10), 1395–1412 (2011).
- ²⁷S. Wasilewski and P. B. Rosenthal, "Web server for tilt-pair validation of single particle maps from electron cryomicroscopy," *J. Struct. Biol.* **186**(1), 122–131 (2014).
- ²⁸I. Lagerstedt, W. J. Moore, A. Patwardhan, E. Sanz-Garcia, C. Best, J. R. Swedlow, and G. J. Kleywegt, "Web-based visualisation and analysis of 3D electron-microscopy data from EMDB and PDB," *J. Struct. Biol.* **184**(2), 173–181 (2013).
- ²⁹S. Abbott, A. Iudin, P. K. Korir, S. Somasundharam, and A. Patwardhan, "EMDB web resources," *Curr. Protoc. Bioinf.* **61**(1), 5.10.11–5.10.12 (2018).
- ³⁰C. L. Lawson and W. Chiu, "Comparing cryo-EM structures," *J. Struct. Biol.* **204**(3), 523–526 (2018).
- ³¹J. B. Heymann, R. Marabini, M. Kazemi, C. O. S. Sorzano, M. Holmdahl, J. H. Mendez, S. M. Stagg, S. Jonic, E. Palovcak, J. P. Armache, J. Zhao, Y. Cheng, G. Pintilie, W. Chiu, A. Patwardhan, and J. M. Carazo, "The first single particle analysis map challenge: A summary of the assessments," *J. Struct. Biol.* **204**(2), 291–300 (2018).
- ³²B. A. Barad, N. Echols, R. Y. Wang, Y. Cheng, F. DiMaio, P. D. Adams, and J. S. Fraser, "EMRinger: Side chain-directed model and map validation for 3D cryo-electron microscopy," *Nat. Methods* **12**(10), 943–946 (2015).
- ³³G. Pintilie and W. Chiu, "Assessment of structural features in Cryo-EM density maps using SSE and side chain Z-scores," *J. Struct. Biol.* **204**(3), 564–571 (2018).
- ³⁴J. H. Mendez and S. M. Stagg, "Assessing the quality of single particle reconstructions by atomic model building," *J. Struct. Biol.* **204**(2), 276–282 (2018).
- ³⁵M. A. Herzik, Jr., J. S. Fraser, and G. C. Lander, "A multi-model approach to assessing local and global cryo-EM map quality," *Structure* **27**(2), 344–358.E3 (2019).
- ³⁶G. Pintilie, K. Zhang, Z. Su, S. Li, M. F. Schmid, and W. Chiu, "Measurement of atom resolvability in cryo-EM maps with Q-scores," *Nat. Methods* (in press) (2020).
- ³⁷I. S. Gabashvili, R. K. Agrawal, C. M. Spahn, R. A. Grassucci, D. I. Svergun, J. Frank, and P. Penczek, "Solution structure of the E. coli 70S ribosome at 11.5 Å resolution," *Cell* **100**(5), 537–549 (2000).
- ³⁸K. Zhang, H. Zhang, S. Li, G. D. Pintilie, T. C. Mou, Y. Gao, Q. Zhang, H. van den Bedem, M. F. Schmid, S. W. N. Au, and W. Chiu, "Cryo-EM structures of *Helicobacter pylori* vacuolating cytotoxin A oligomeric assemblies at near-atomic resolution," *Proc. Natl. Acad. Sci. U. S. A.* **116**(14), 6800–6805 (2019).
- ³⁹M. Tagari, R. Newman, M. Chagoyen, J. M. Carazo, and K. Henrick, "New electron microscopy database and deposition system," *Trends Biochem. Sci.* **27**(11), 589 (2002).
- ⁴⁰J. B. Heymann and D. M. Belpap, "Bsoft: Image processing and molecular modeling for electron microscopy," *J. Struct. Biol.* **157**(1), 3–18 (2007).
- ⁴¹G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, and S. J. Ludtke, "EMAN2: An extensible image processing suite for electron microscopy," *J. Struct. Biol.* **157**(1), 38–46 (2007).
- ⁴²S. J. Ludtke, C. L. Lawson, G. J. Kleywegt, H. Berman, and W. Chiu, "The 2010 cryo-EM modeling challenge," *Biopolymers* **97**(9), 651–654 (2012).
- ⁴³R. Marabini, B. Carragher, S. Chen, J. Chen, A. Cheng, K. H. Downing, J. Frank, R. A. Grassucci, J. Bernard Heymann, W. Jiang, S. Jonic, H. Y. Liao, S. J. Ludtke, S. Patwari, A. L. Piotrowski, A. Quintana, C. O. Sorzano, H. Stahlberg, J. Vargas, N. R. Voss, W. Chiu, and J. M. Carazo, "CTF challenge: Result summary," *J. Struct. Biol.* **190**(3), 348–359 (2015).
- ⁴⁴R. Marabini, S. J. Ludtke, S. C. Murray, W. Chiu, J. M. de la Rosa-Trevin, A. Patwardhan, J. B. Heymann, and J. M. Carazo, "The electron microscopy exchange (EMX) initiative," *J. Struct. Biol.* **194**(2), 156–163 (2016).
- ⁴⁵A. Patwardhan, R. Brandt, S. J. Butcher, L. Collinson, D. Gault, K. Grunewald, C. Hecksel, J. T. Huiskonen, A. Iudin, M. L. Jones, P. K. Korir, A. J. Koster, I. Lagerstedt, C. L. Lawson, D. Mastronarde, M. McCormick, H. Parkinson, P. B. Rosenthal, S. Saalfeld, H. R. Saibil, S. Sarntivijai, I. Solanes Valero, S. Subramaniam, J. R. Swedlow, I. Tudose, M. Winn, and G. J. Kleywegt, "Building bridges between cellular and molecular structural biology," *Elife* **6**, e25835 (2017).
- ⁴⁶A. Kryshafyovych, P. D. Adams, C. L. Lawson, and W. Chiu, "Evaluation system and web infrastructure for the second cryo-EM model challenge," *J. Struct. Biol.* **204**(1), 96–108 (2018).
- ⁴⁷Editorial, "Challenges for cryo-EM," *Nat. Methods* **15**, 985 (2018).
- ⁴⁸M. Baker, "Cryo-electron microscopy shapes up," *Nature* **561**, 565–567 (2018).
- ⁴⁹Y. Zhu, B. Carragher, R. M. Glaeser, D. Fellmann, C. Bajaj, M. Bern, F. Mouche, F. de Haas, R. J. Hall, D. J. Kriegman, S. J. Ludtke, S. P. Mallick, P. A. Penczek, A. M. Roseman, F. J. Sigworth, N. Volkman, and C. S. Potter, "Automatic particle selection: Results of a comparative study," *J. Struct. Biol.* **145**(1–2), 3–14 (2004).