

ADVANCED REVIEW



WILEY

Integrating model simulation tools and cryo-electron microscopy

Joseph George Beton¹ | Tristan Cragolini² | Manaz Kaleel¹ |
Thomas Mulvaney¹ | Aaron Sweeney¹ | Maya Topf¹

¹Centre for Structural Systems Biology (CSSB), Leibniz-Institut für Virologie (LIV), Hamburg, Germany

²Institute of Structural and Molecular Biology, Birkbeck and University College London, London, UK

Correspondence

Maya Topf, Centre for Structural Systems Biology (CSSB), Leibniz-Institut für Virologie (LIV), Hamburg, Germany.
Email: maya.topf@cssb-hamburg.de

Funding information

Leibniz ScienceCampus InterACT; Wellcome Trust, Grant/Award Number: 209250/Z/17/Z

Edited by: Modesto Orozco, Associate Editor

Abstract

The power of computer simulations, including machine-learning, has become an inseparable part of scientific analysis of biological data. This has significantly impacted the field of cryogenic electron microscopy (cryo-EM), which has grown dramatically since the “resolution-revolution.” Many maps are now solved at 3–4 Å or better resolution, although a significant proportion of maps deposited in the Electron Microscopy Data Bank are still at lower resolution, where the positions of atoms cannot be determined unambiguously. Additionally, cryo-EM maps are often characterized by a varying local resolution, partly due to conformational heterogeneity of the imaged molecule. To address such problems, many computational methods have been developed for cryo-EM map reconstruction and atomistic model building. Here, we review the development in algorithms and tools for building models in cryo-EM maps at different resolutions. We describe methods for model building, including rigid and flexible fitting of known models, model validation, small-molecule fitting, and model visualization. We provide examples of how these methods have been used to elucidate the structure and function of dynamic macromolecular machines.

This article is categorized under:

Structure and Mechanism > Molecular Structures

Structure and Mechanism > Computational Biochemistry and Biophysics

Software > Molecular Modeling

KEYWORDS

cryo-EM, model validation, molecular modeling, molecular simulations

1 | INTRODUCTION

Over the past two decades, cryogenic electron microscopy (cryo-EM) has had a vast impact on the field of structural biology by solving the structures of many large and dynamic macromolecular assemblies in different functional forms

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *WIREs Computational Molecular Science* published by Wiley Periodicals LLC.

in near-native state. Overcoming the limitations of traditional techniques, such as X-ray crystallography and NMR spectroscopy, in addition to recent advances in direct electron detectors and improvements in image processing methods, have made it possible to routinely obtain resolutions better than 3.5 Å even for low purity samples.^{1,2} In fact, in the last 8 years there has been an exponential increase in the number of deposited structures in the Electron Microscopy Data Bank (EMDB).³ From 4 maps having a resolution better than 4 Å in 2011, 7556 maps have reached this level in 2021 (~40% of the database; Figure 1a).

The most widely-used cryo-EM method is single-particle cryo-EM; where thousands of noisy projection images containing up to 100 s of randomly oriented “particles” (imaged molecules) suspended in vitreous ice are collected in a transmission electron microscope at cryogenic temperatures. These imaged particles are computationally extracted from images, and their relative orientations are determined, using sophisticated software packages, before the particle images are back projected to generate a 3D reconstruction (cryo-EM map) of the original imaged molecule. Another important method is cryo electron tomography (cryo-ET); here a series of projection images are collected while tilting the sample stage. Again, the relative orientations of each image are determined and the images are back projected to generate a 3D reconstruction: a tomogram of the imaged sample. Cryo-ET can elucidate the structure of complex samples, entire cell components, tissues, or full viral particles in situ.^{4–6} This means that in combination with subtomogram averaging, cryo-ET can be used to generate reconstructions of biological structures in their native, cellular environment.⁷ However, the average resolution of reconstruction by cryo-ET from cells is currently 20 Å, although resolution as good as ~3.5 Å has recently been achieved for the 70S ribosome bound to an antibiotic inside intact bacterial cells.⁸

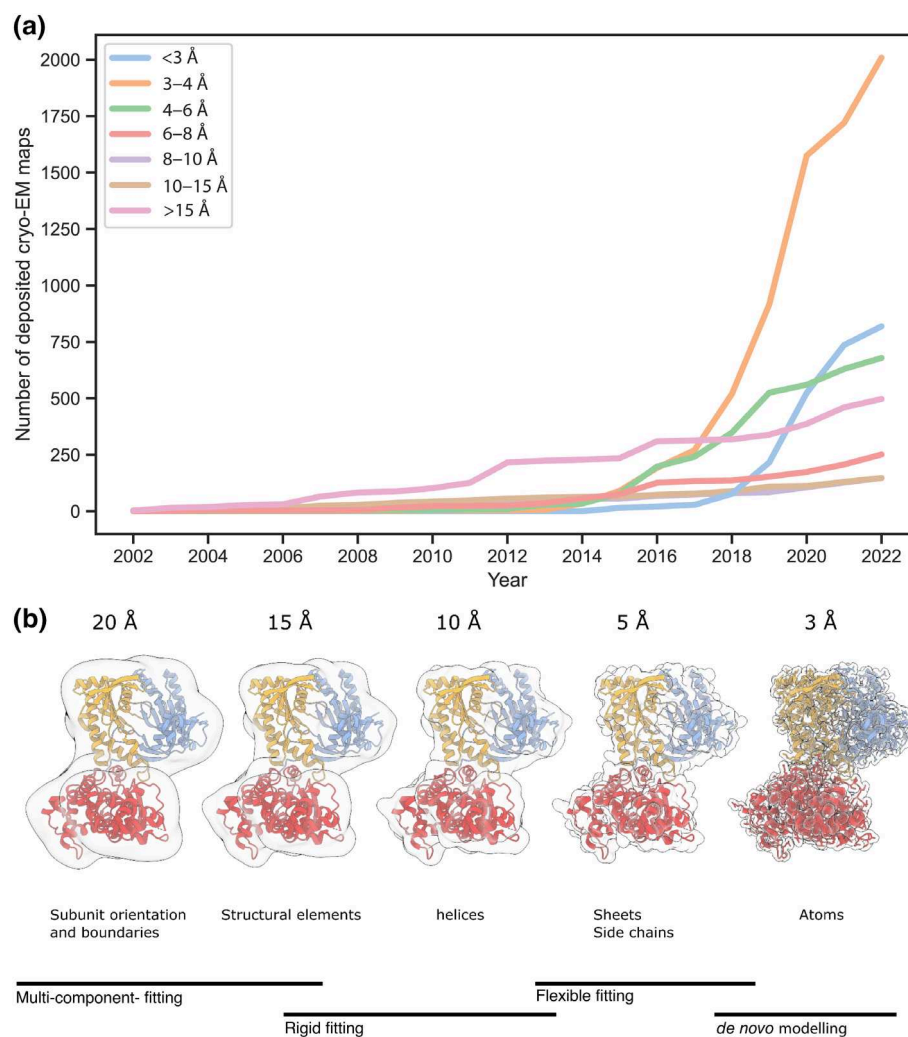


FIGURE 1 Cryo-EM map resolutions and the relevant model fitting methods. (a) Plot showing the number of cryo-EM maps deposited in the EMDB, per resolution shell, in each year since 2002. (b) Illustration of a cryo-EM map with a well-fit structure at different resolutions, with the structural features and class of methods that can be used to fit the structure as a function of resolution.

In almost all cases, after solving a cryo-EM map of a macromolecule, researchers will use it to build an atomistic model, to assist the interpretation of its structure and function. Depending on the resolution of the maps, different methods can be adopted to derive these models (Figure 1b). When the resolution is reaching atomic detail, methods derived from X-ray crystallography can be applied to refine the positions of the atoms within a structure. However, at lower resolutions, where such information is not present, different density segmentation and fitting algorithms have to be used to rigidly fit large portions of structure (e.g., secondary structure or whole chains). In those cases, a combination of multiple methods is often needed to achieve an interpretable model that agrees with the density map.

In cryo-EM/ET experiments, one obtains a snapshot of the conformational heterogeneity associated with a molecule's structural fluctuations and, therefore, the dynamic information of a molecule is intrinsically manifested in the data. However, extracting information from this ensemble of images can be complex, in large part due to the very low signal to noise ratio of single cryo-EM images.⁹ Information is typically extracted from this ensemble in one of two ways: first, classification methods can be integrated into the reconstruction process, allowing the 3D reconstruction of multiple conformations from the same set of single-particle images.^{10,11} Second, computational methods are designed to extract the dynamic information from the cryo-EM/ET density map by flexibly refining an atomic model in the map while fitting it.

Whether it is the determination of maps at multiple conformations or the building of atomistic models, to gain insights into the dynamics of the assembly structure from cryo-EM data, simulation tools are required. Many different computational methods and tools have been developed in the last two decades, and are often bundled together in software suites to perform structure classification, fitting and validation steps at the same time (e.g., CCP-EM,¹² PHENIX,¹³ Rosetta,¹⁴ and SCIPION).¹⁵ Below, we describe computational methods that are designed to extract information directly from the map, then continue with methods that combine information from the map and the corresponding atomic model.

2 | RESOLUTION HETEROGENEITY AND DYNAMICS IN CRYO-EM/-ET

In addition to the flexibility of the imaged macromolecules, the quality and information content of cryo-EM maps can be degraded by artifacts arising due to microscope defects, particle alignment errors, preferred particle orientations and the high noise in collected images.¹⁶ Therefore, quantifying the resolution of cryo-EM reconstructions (i.e., which features in a map arise due to signal generated by the imaged molecule, rather than noise) is necessary. The Fourier Shell Correlation (FSC) is the most commonly used method to do this. The FSC evaluates the similarity between two independently refined halves of a data set (half-maps) in Fourier space, by calculating the average correlation of Fourier components in each resolution (radial) shell.¹⁷ A global resolution estimate can be derived from such FSC curves by finding the frequency shell that drops below a defined threshold,¹⁸ typically 0.143 is used,¹⁹ although alternative cutoff values or regimes have been proposed.^{20–22} Once the resolution of a map is determined, one can boost the signal amplitude at higher resolutions, which are attenuated by microscope optics and image processing errors.²³ This increases the visibility of the more detailed features in the map, which, depending on the resolution, can include secondary structure elements and amino acid side chains.²⁴

Regardless of the criteria for the global resolution, there are a number of issues that cannot be fully represented by such measures, including conformational heterogeneity within the sample, as well as preferred orientations and other sources of unequal distribution of angles between particles that represent different views. In fact, the EMDB contains many maps within which the resolution distribution can be broad. Estimates of this heterogeneous “local” resolution can be used to aid the validation of the atomic models associated with the map, identify regions of uncertainty, identify dynamic versus stable regions and give insights into biological processes. A number of methods to determine local resolution estimation have been developed,²⁵ the first of which is *blocres*,²⁶ implemented in the Bsoft software.²⁷ The method uses a local form of the FSC by moving a window of a predetermined size along the two half-maps. ResMap,²⁷ which does not rely on FSC calculation to estimate the local resolution, uses local sinusoidal features to assign a resolution to every voxel. MonoRes estimates the local resolution by establishing a comparison (hypothesis test) at different frequencies between the energy of local signals and the energy of noise.^{28,29} DeepRes is based on deep learning from filtered atomic models at different frequencies.³⁰ The map sharpening should reflect the local resolution, that is, which details should be boosted, and which suppressed.³¹ There are multiple methods to achieve this, including using model based sharpening,³² sharpening based on the local resolution,³³ and using deep learning models.³⁴

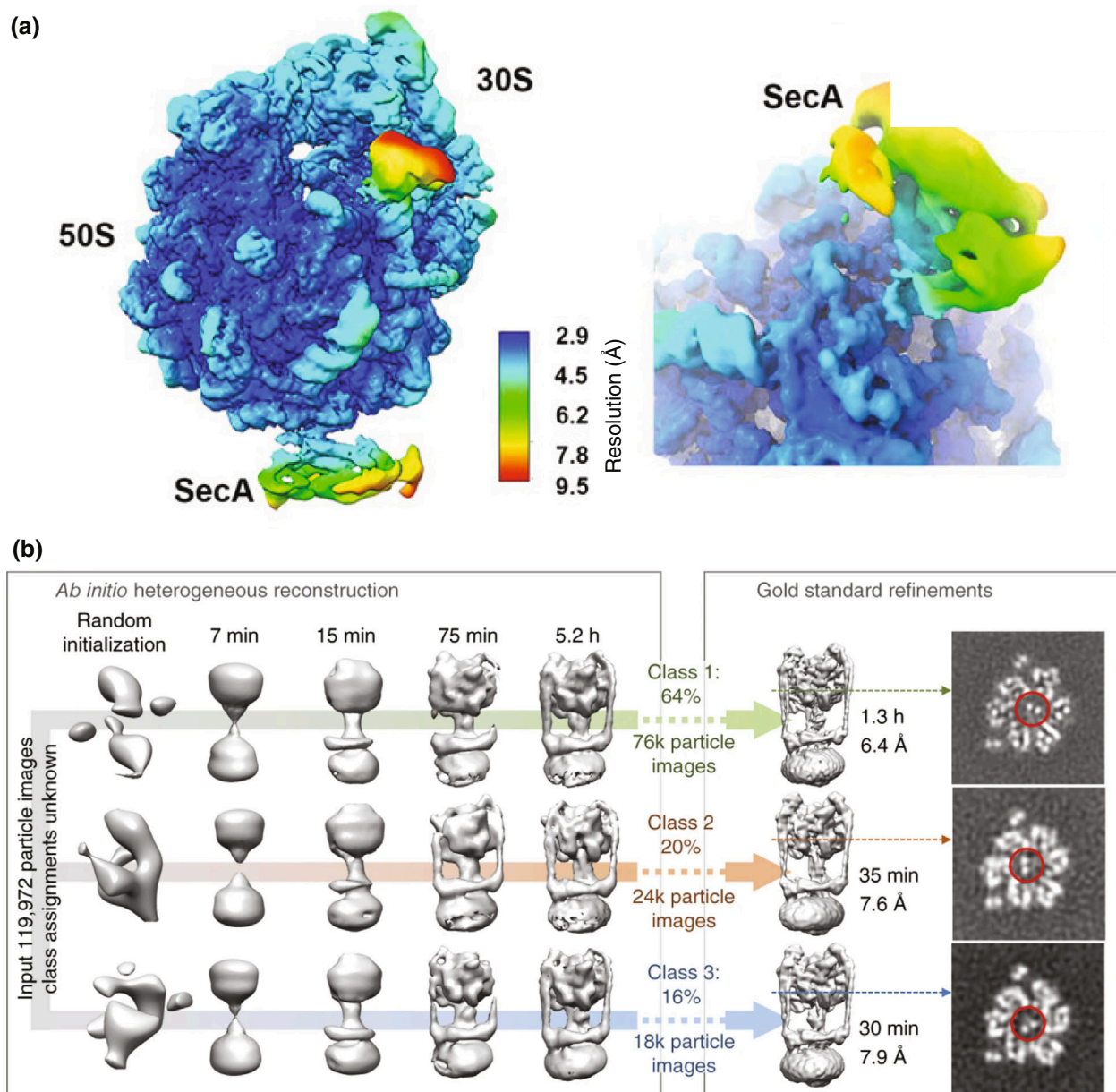


FIGURE 2 Local resolution heterogeneity and identifying structural sub-populations from a mixed population using cryo-EM. (a) Cryo-EM reconstruction of the *E. coli* ribosome in complex with SecA colored based on the local resolution. The global resolution of the map is 3.1 Å, but the SecA component is at considerably lower resolution. Figure adapted from Reference 35. (b) An example classification, and subsequent refinement, for ATP synthase using the cryoSPARC software package.³⁶ Here, one can see that the ab initio classification isolates three sub-populations of the ATP-synthase (left hand panel), each with a unique structure that is clearly visible after refinement (right hand panel). This figure is modified from Reference 36.

Heterogeneous resolution in cryo-EM maps presents a problem in interpretation and further processing, and can even lead to subunits in a complex being unresolved (Figure 2a). One cause of this can be the underlying conformational heterogeneity of the imaged molecule. Sophisticated computational methods have been developed to disentangle this mixed population of imaged molecules (particles) from a cryo-EM experiment into a series of reconstructions that each represent an individual state.^{37,38} This can be achieved by grouping sets of particles into discrete individual classes, where, ideally, each class should represent a unique conformation of the imaged molecule (Figure 2b). As both 2D and 3D classification methods have developed alongside improvements in cryo-EM sample preparation^{39,40} and imaging methods,^{41,42} subpopulations with increasingly fine structural deviations can be separated.⁴³ Classification into discrete groups is often done using maximum likelihood, as implemented in RELION and other software packages, but other

statistical methods, such as multivariate statistical analysis (MSA) have been used.^{36,44–52} The output 3D classes are normally subjected to further refinement, to boost their resolution. This has been a widely applied method for identifying structural states, with some notable applications being the rotary movement of the ATP synthase,⁵³ large conformational changes associated with RNA splicing,⁵⁴ and the movements of the ribosome associated with translation.^{38,55}

With these discrete states in hand, one can use additional methods to infer the underlying dynamics that links them. For example, principal component analysis of reconstructed volumes has been used, in conjunction with massive (>300 classes) classifications, to determine the energy landscapes for small molecule binding to the proteasome⁵⁶ and for the conformational dynamics of the human spliceosome.⁵⁷ A similar approach was used to study glutamate dehydrogenase (GDH): here, the principle components of 28 representative structures derived from molecular dynamics (MD) simulations were analyzed to identify the main motions of the NAD domain. These principal components were used to generate a free-energy landscape for the NAD domain motion, and the positions of each reconstructed class was mapped onto this landscape by determining the weighted similarity between each model-map combination.⁵⁸ A specialized software package, Fluctuating Finite Element Analysis (FFEA), has been developed to analyze large structural movements in potentially very large molecules,^{59,60} using low resolution (>5 Å) cryo-EM/-ET maps. In FFEA, each component of a large complex is treated as a single “viscoelastic” unit, which is deformed by the thermodynamics of the solvent.

However, grouping particles into discrete classes requires, to some extent, that the imaged molecule inhabits discrete states itself. This is frequently not the case as many molecules exhibit a more continuous trajectory of movement, which often corresponds to lower local resolution density in a reconstruction. However, specialized methods can extract and correct such molecular motions.^{43,61–66} An early example of such an approach was used to visualize the Brownian motions of the yeast cytosolic ribosome fluctuating in solution (Figure 3).³⁸ Here, structural subclasses of the ribosome were identified by mapping groups of particles with similar alignment angles (relative to a global reference) to an N-dimensional manifold. This manifold was analyzed in a multistep procedure, to generate a novel mapping of particles, which can be used to identify sub-classes of particles, and to derive the energy landscape sampled by the imaged system. Analyzing this mapping for the ribosome revealed it undergoes motions similar to those associated with elongation, despite the absence of any mRNA and most tRNAs.³⁸ The popular software RELION uses an alternate “Multi-Body Refinement” implementation, in which a multi-component complex can be manually segmented into multiple “bodies” (masked regions), which are then independently refined using a maximum likelihood approach.⁶³ The most significant molecular movements can be deduced and visualized after this refinement, by subjecting the relative alignment angles of each body to principal component analysis. CryoSPARC uses a 3D variability analysis algorithm to identify classes from a continuously variable population of particles⁶²: here, probabilistic principal component analysis (where principle components are estimated using an expectation maximization algorithm)^{67,68} is used to identify the main structural fluctuations within the imaged molecules. This approach was used to identify and visualize large structural movements

BOX 1 Machine learning methods commonly used in cryo-EM

Autoencoders (AEs): An autoencoder is composed of two parts: the first is an encoder, which transforms input data into a form that is typically compressed and lower dimensional. This is sometimes referred to as the code, or “latent” form. Typically, this latent form is expected to only contain information related to the most prominent identifying aspects of the input data. The second part is the decoder, which converts this latent representation into an output that is as similar to the input data as possible. A more powerful variation of Autoencoders is Variational Autoencoders.

CNNs: Convolutional neural networks (CNNs) are one of the most widely used artificial neural network architectures. These networks effectively scan the input data to extract features. CNNs excel in pattern recognition on image data.⁷⁰ These exceptional performances of CNNs are partially due to their ability to model complicated images into much smaller feature maps while keeping the complexity of the model relatively low. Furthermore, CNN kernels read adjacent pixels of an image at a given time, enabling them to learn the locality of images. This advantage is further reinforced by the hierarchical nature of CNN where each layer extracts progressively more abstract features.

U-nets: For pixel level annotation (semantic segmentation), U-nets are the most widely used architecture. CNNs are used as building blocks to build U-nets.

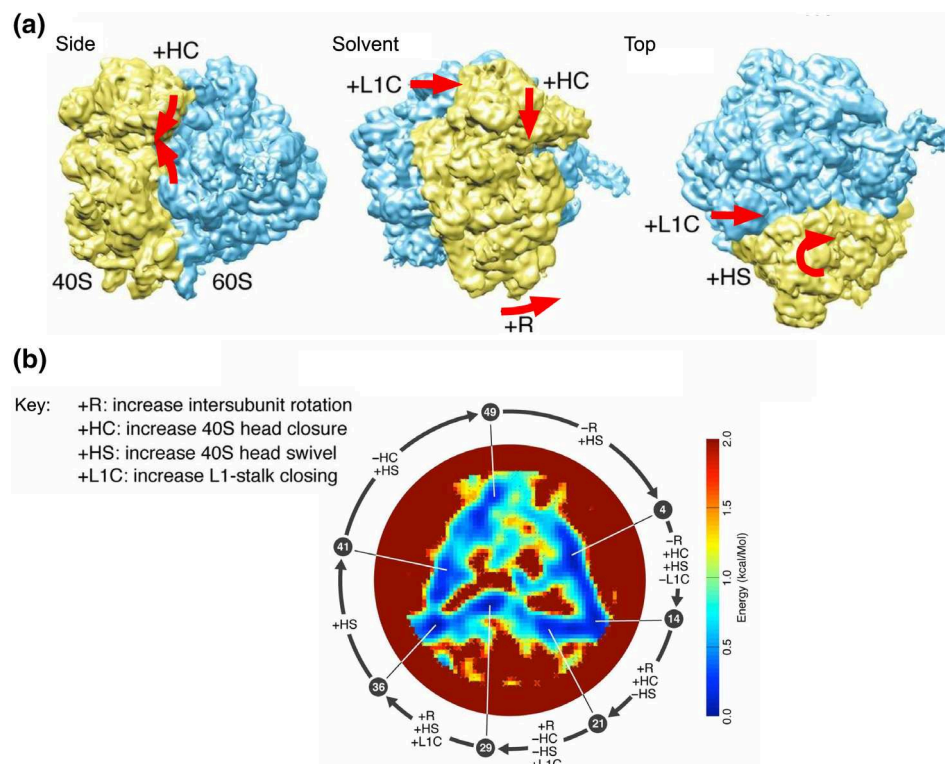


FIGURE 3 Energy landscape for the ribosome derived from cryo-EM. (a) Three views of a ribosome reconstruction with the main movements labeled with red arrows and letter codes, which are further clarified in the key. (b) The energy landscape that the ribosome traverses. The triangular shape represents the main structures occupied during its fluctuations in solvent, which is divided into 50 states. Seven states are highlighted (numbers) and the arrows and labels show the structural changes undergone as the ribosome transitions between these states. Figure modified from Reference 38.

in the human motor protein Dynein, when it was bound to a regulator protein Lis1. These movements took place at the flexible dynein linker domain, and were speculated to include the beginning of the dynein power stroke.⁶⁹

Other methods utilize machine learning algorithms to determine the conformational landscape for heterogeneous sets of particles. CryoDRGN uses a variational autoencoder, image encoder-decoder architecture⁷¹ (Box 1) to initially transform particle images and their alignment angles into a low-dimensional latent space, which can then be used to identify clusters of particles that belong to distinct conformations.⁴³ A similar encoder-decoder architecture is used in *e2gmm*: a decoder is used to transform the images and their alignment angles into latent space, and an encoder to transform this latent space information into a Gaussian mixture model representation, which can be used to identify structural sub-populations.⁶⁴

Another example of a machine learning based approach has been developed that can accurately infer protein dynamics from a cryo-EM map.⁷² Here, a 3D convolutional neural network (CNN)(Box 1) was trained with cryo-EM maps (obtained from the EMDB⁷³) and all-atom MD simulations, derived from the fitted models (obtained from the PDB⁷⁴). The fully trained neural network was able to derive accurate root mean squared fluctuations for all atoms in a protein model, based only on the EM map. These values showed good agreement with all-atom MD simulations and agreed with hydrogen-deuterium exchange experimental data.⁷²

3 | DETECTION OF ELEMENTS FROM DENSITY

Detecting important biomolecular elements from cryo-EM maps such as secondary structure elements (SSEs) still remains an open problem although several software tools for the element detection problem have been introduced. This was a particularly relevant problem prior to the “resolution revolution” in cryo-EM, where maps were generally

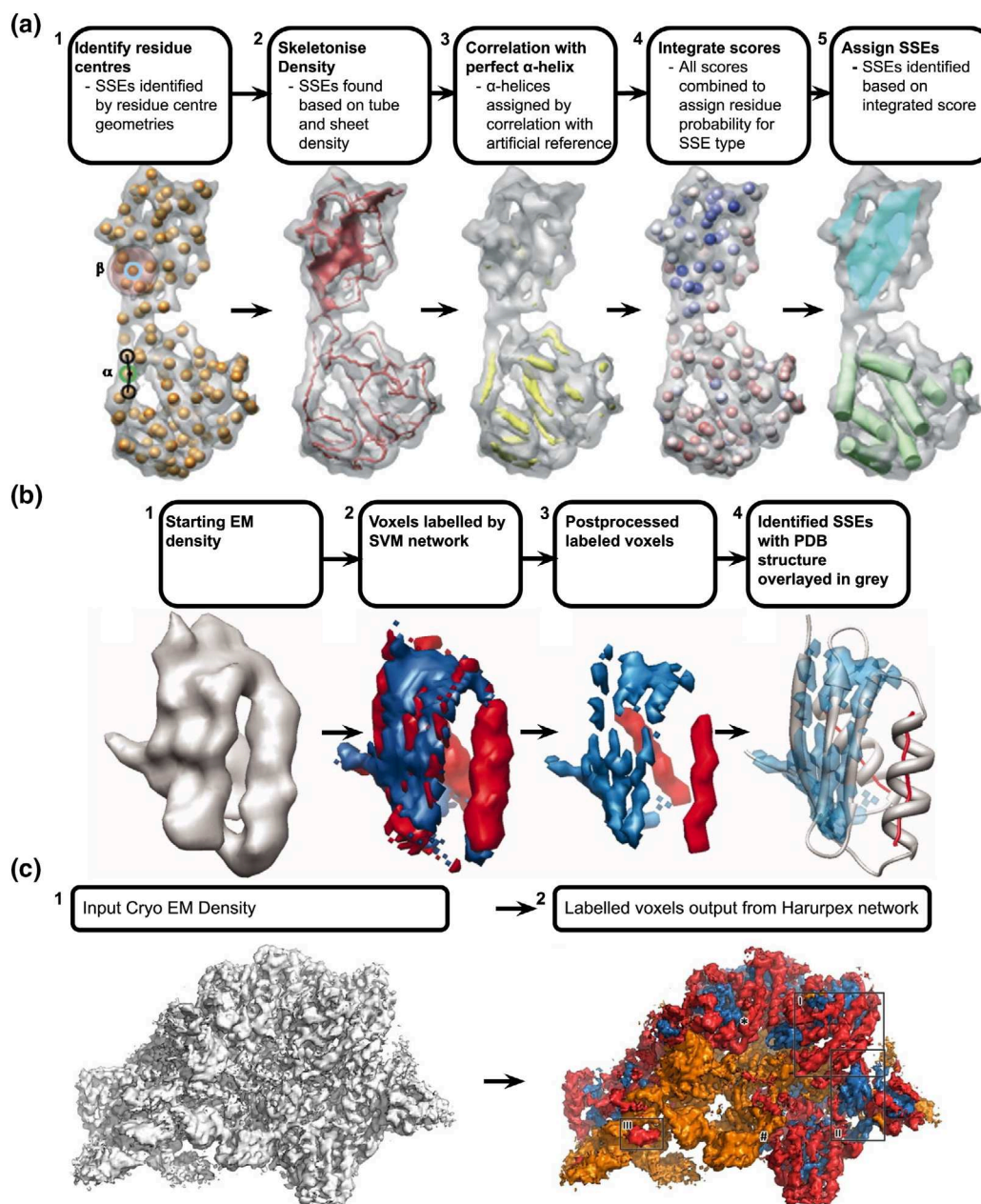


FIGURE 4 SSEHunter, SSElearner and haruspex methods for detecting elements in cryo-EM maps. (a) The multistep workflow of SSEHunter, where independent algorithms are integrated to identify SSEs. Figure adapted from Reference 77. (b) The workflow for SSElearner shown as an example for identifying SSEs from a 8 Å simulated map. Note the workflow is effectively reduced to two steps (2–3) by integration of machine learning (SVM classifier) architecture for SSE identification. Figure adapted from Reference 78. (c) Combined SSE and DNA identification using the haruspex CNN architecture from a 3.7 Å experimental cryo-EM map (EMD-9627). Here, a single pass of the cryo-EM map (1) through the network is sufficient to identify all SSEs in the map and DNA (2, α -helices in red, β -sheets in blue, DNA in orange). Example adapted from Reference 79.

restricted to medium resolution (intermediate, ~ 4 – 10 Å). However, there are still many maps being deposited in the medium resolution range, resulting from heterogeneous samples as well as sub-tomogram averaging.

One of the earliest software systems to detect biomolecular elements is Helixhunter.⁷⁵ Helixhunter detects α -helix position, orientation and length in density maps of resolutions 6–10 Å. This is using a five-dimensional cross-correlation search of the 3D cryo-EM map, followed by feature extraction. Sheetminer⁷⁶ is a tool that detects β -sheets in 6–10 Å resolution cryo-EM maps using multi-step ad hoc morphological analysis (Figure 4a). Another early work is SSEhunter.⁷⁷ SSEhunter detects both α -helices and β -sheets in density maps of intermediate-resolution with the aid of

template-based search, density skeletonization, and local geometry calculations. Detecting structural homologues of proteins in density maps is another important task. EMatch⁸⁰ is detecting helices in 6–10 Å resolution cryo-EM maps. After detecting helices, EMatch utilizes various computational methods to recognize structural homologues of protein domains. Detection of fully folded domains can be done by fitting domains representatives, as done in SPIEM.⁸¹

EM-fold⁷⁰ is a tool to build topological models for large proteins using computational structure prediction methods and experimental cryo-EM density maps. First, α -helical regions are predicted in the density map and in the protein sequence using a consensus of jufo,^{82,83} psipred,⁸⁴ and sam.^{85,86} Then, an in-house Monte Carlo assembly algorithm is used to place the predicted α helices into α -helical density rods of the density map. Finally, Rosetta⁸⁷ is used to add loop regions and side-chain coordinates. EM-fold was tested on cryo-EM density maps with ~ 6 Å resolution. Pathwalking⁸⁸ is a semi-automated template-free protocol to enumerate C α models from 3.5 to 6 Å resolution density maps. Pathwalking is capable of rapidly generating initial models for individual proteins using an approach derived from the Traveling Salesman Problem.

VolTrac is a tool to detect helices in 4–10 Å resolution cryo-EM density maps by utilizing a combination of genetic algorithms with stochastic search and a bidirectional expansion with a tabu search strategy to detect helices.⁸⁹ SSEtracer⁹⁰ is another SSE finding tool that detects helices and β -sheets in ~ 5 –10 Å resolution maps using local feature analysis based on the characteristics of the local shape of helices and β -sheets. There are several other tools available to detect SSE and related elements from cryo-EM density maps.^{91–93}

Beta barrel (β -barrel) is a special β -sheet closed toroidal structure. *BarrelMiner*⁹⁴ is a tool developed to automatically detect β -barrels from cryo-EM density maps with resolution between 5.5 and 10 Å. *BarrelMiner* uses Random Sample Consensus (RANSAC)⁹⁵ and a variation of this technique⁹⁶ to search for the location and orientation of a β -barrel in 3D density volumes. Various other tools that combine existing tools to harness higher performance to detect SSE have also been developed.^{97,98}

Given that machine learning is a proven area of study for various pattern recognition problems, such as object detection in images,^{99,100} it has been successfully applied to identifying SSEs from cryo-EM maps, and recently applied also to higher resolution maps. One of the earliest tools is *SSELearner*⁷⁸ (Figure 4b). *SSELearner* detects helices and β -sheets in density maps with the aid of a secondary structure annotator, *SSID*, that predicts helices and β -strands from the backbone C α trace. *SSELearner* uses Support Vector Machine (SVM) to detect SSEs. The predictor was trained on volumetric maps extracted from EMDB and tested on 10 simulated density maps at 8 Å resolution and 13 experimentally derived cryo-EM density maps with resolution between 3.8 and 9 Å. A widely used deep learning algorithm for SSE detection on cryo-EM maps is Convolutional Neural Networks (CNN). A CNN based predictor¹⁰¹ assigns a label for each voxel in the tomogram and predicts a probability of label for each voxel with respect to SSEs. Another CNN based predictor¹⁰² that predicts backbone structure and C α atoms along with SSEs. The model was trained on simulated maps produced from 7024 PDB files with various resolutions and tested on 50 experimental maps with the resolution ranging between 2.6 and 4.4 Å.

*Haruspex*⁷⁹ detects Oligonucleotides along with SSEs in cryo-EM density maps (Figure 4c). The prediction model is trained on 293 experimentally derived EMDB maps with an average resolution of 4 Å or better and tested on a test set of 122 EMDB maps with a resolution of 4.4 Å or better. *Emap2sec+*¹⁰³ is a new residual network (ResNet)¹⁰⁴ based tool to detect protein secondary structure and nucleic acids in cryo-EM maps. *Emap2sec+* classifies each voxel in the density map into four classes, alpha-helix, β -strand, coil, and DNA/RNA. *Emap2sec+* is designed for intermediate resolution maps while *Haruspex* is designed for 4 Å or higher resolution maps.

It is also important to note that the performance of machine learning methods can vary between experimental maps and simulated maps. For instance, *SSELearner* detects helices on 10 simulated density maps with the averaged specificity and sensitivity of 94.9% and 95.8%, respectively. This performance drops to 91.8% and 74.5%, respectively for 13 experimentally derived cryo-EM density maps obtained from EMDB. For the β -sheet detection problem, *SSELearner* detects β -sheets on above mentioned 10 simulated density maps with specificity and sensitivity of 86.7% and 96.4%, respectively while this performance drops to 85.2% and 86.5%, respectively, for above mentioned 13 EMDB cryo-EM density maps.

DeepTracer¹⁰⁵ takes the concept of element detection a step further, by determining the position of all atoms in the model, using a machine learning based approach. It is powered by four modified U-Net architectures (Box 1).¹⁰⁶ U-Net is a popular CNN based deep learning architecture for semantic segmentation where each pixel is classified. Each of these U-Nets reads the preprocessed cryo-EM density map and classifies each voxel into a class. The first U-Net (Atoms U-Net) predicts each voxel in the preprocessed cryo-EM density map into the following four classes, alpha-carbon atom, nitrogen atom, carbon atom, and no atom. The second U-Net (Backbone U-Net) classifies each voxel into three classes,

backbone, side chain, and not a part of the protein. The third U-Net (Secondary Structure U-Net) predicts the SSE of each voxel. This network classifies each voxel into four classes, loops, sheets, helices, and no structure. Finally, the fourth U-Net (Amino Acid Type U-Net) determines the amino acid type of each voxel. This network classifies each voxel into 21 classes where the first 20 classes represent the standard amino acid type and the last class represents non-amino acid. The final structure modeling is determined by the prediction of all four U-Nets. Deeptimizer was used to build an initial model of the yeast Ubr1 ubiquitinase, which was then further refined with phenix (see Section 5).¹⁰⁷

4 | RIGID-BODY ASSEMBLY FITTING

Solving the structure of very large complexes at high resolution (<4 Å) can still be challenging, often due to flexibility of the imaged molecule, or difficulties with sample preparation.¹⁰⁸ As a result, reconstructions are often restricted to low resolution (>4 Å). In such cases, one can still construct a model of the imaged sample using assembly fitting, whereby the structures of each component are placed into the cryo-EM density as rigid bodies. Given the low resolution, and thus low information, in the map, the overall architecture of the associated complexes can be probed by assembly fitting complemented by other experimental techniques, such as cross-linking Mass Spectrometry (integrative modeling).¹⁰⁹ Modeling into these lower resolution maps remains highly relevant: more than half of the deposited maps in EMDB have a resolution worse than 4 Å, including more than a third (36%) of maps deposited in 2021. At the lower end of resolution, the position of entire protein chains may be ambiguous. For large multimeric assemblies, this presents a significant challenge to assembly fitting procedures.

When attempting to solve the structure of a large protein complex, if the structure of components in the assembly are known, it is possible to place them within the map. Examples of such assemblies can be found on the PDB-Dev repository,¹¹⁰ which includes large assemblies such as polymerases, ribosomes, nuclear pore complexes, or proteasomes among others. Notably, the assembly of the proteasome and nuclear pore complex were both solved using integrative modeling algorithms and data from cryo-EM, crystallography, and mass spectrometry.¹⁰

When trying to fit multiple components rigidly within a map, each component can be described as having 6 degrees of freedom, 3 to describe its spatial position, and 3 for its orientation. The combination of those degrees of freedom results in a very large number of possible conformations, making it difficult to ascertain the best possible arrangement.

This problem has been tackled with numerous approaches. The search can be accelerated using FFT-based searches, spherical harmonics or Zernike descriptors of the volumes; and the use of maximization methods, gaussian mixtures, or density vectors can iteratively improve the fit until it reaches a local maximum (Table 1). A summary of proposed methods is shown in Table 1, and a description of how assembly fitting can be performed using these tools is provided below.

The original methods that were developed to solve this problem, such as Situs, DOCKEM, or Mod-EM tackled this problem by maximizing the correlation between the map and a blurred model,^{111,118,126} but multiple components were typically fitted sequentially.

Alternative representations of the density offer powerful ways to describe and improve the fit of a model to it. *gmfit* uses a simplified gaussian mixture representation, and optimizes their position and rotation to maximize an energy term maximizing the fit between the sum of the Gaussians and the map while minimizing subunits overlap.¹²² Later methods such as Multifit¹²⁴ used simultaneous rather than sequential fitting, as initial errors in the fitting procedure

TABLE 1 Table of assembly fitting methods for cryo-EM, ordered chronologically

Method name	Core concept	Method name	Core concept
Situs ^{111,112}	Grid-based full search	IMP ¹¹³	Restrains; Bayesian inference
DOCKEM ¹¹⁴	Correlation maximization	EMLZerD ¹¹⁵	Zernike descriptors
Chimera ¹¹⁶	Correlation maximization	γ -Tempty ¹¹⁷	Genetic algorithm
Mod-EM ¹¹⁸	Correlation maximization	HADDOCK ¹¹⁹	Restrains
ADP-EM ¹²⁰	Spherical harmonics	PRISM-EM ¹²¹	Template matching
<i>gmfit</i> ¹²²	Gaussian mixture	Vesper ¹²³	Density vector
MultiFit ¹²⁴	Belief propagation	Assemblin ¹²⁵	Combined restraint scores based on IMP

will impede finding an overall best fit. In the EMLZerD method, Zernike descriptors are used to represent the surface of a complex, and to find best matching components.

In γ -Tempy, a genetic algorithm is used to iteratively improve the match between candidate assemblies and the experimental map, as estimated by a global score, usually cross-correlation coefficient (CCC) or mutual information (MI),^{117,127} combined with a penalty clash score. The positions and rotations of components within each assembly are initialised using a vector quantisation procedure,¹²⁸ producing the initial set of assembly fits. The next set of assemblies is generated by mutating (i.e., randomly changing) the position and rotation values for the components and by combining values from two “parent” assemblies. The set of “parents” and “children” are scored and ranked, and the best candidates are kept for the next set. This method was shown to be capable of producing reasonable fits for assemblies up to eight components.¹¹⁷

PRISM-EM attempts to generate candidate complexes by using protein–protein docking, minimizing an energy score, and then fitting those complexes in the density map.¹²¹ The two main parts of the scoring in PRISM-EM are a common theme in many assembly fitting methods, with a part related to the density quality of fit, and the other dedicated to scoring the interface. However, the density map is used for scoring and not for optimizing the configuration of the components directly.

Generally, the quality of the fit is limited by the quality of the scoring, and better scoring methods are enabling improvements in fitting methods, as can be seen in VESPER. The latter is a recent method to align cryo-EM maps by converting the density into multiple local vectors that are directed toward high local density.¹²³ These vector maps are then aligned using an exhaustive search, where the degree of overlap between two maps is determined by the dot-product between the vector representations of each map. This approach was found to be very accurate for aligning two experimental cryo-EM maps, and for locally aligning one map containing, for example, a single subunit of a complex into another map of the full complex. Using this local map alignment, VESPER can be used for assembly fitting once the model for each component in the assembly has been converted into a simulated map. Using this approach, VESPER was shown to produce more accurate assemblies (i.e., with a smaller RMSD to a known ground truth solution) compared to other popular methods for the TFIIF, γ -secretase, yeast RNA polymerase I, and other complexes.¹²³

In addition to fitting structures using cryo-EM data, methods have been developed that take advantage of multiple types of data, coming from native mass spectrometry, crosslinking mass spectrometry or small-angle x-ray scattering (SAXS), and other experiments that may provide spatial restraints for the internal organization of a biomolecular complex. Methods that combine multiple experimental information include Haddock,^{119,129} IMP,¹¹³ ROSETTA,¹³⁰ and power.¹³¹ This type of integrative approach can be used to resolve assemblies that would remain too ambiguous or difficult to solve by direct fitting in the density, the extra information from other methods providing the necessary information to either find the correct solution, or separate it from others fitting similarly well in the density. Assemblin combines an IMP assembly pipeline with an efficient sampling procedure to fit very large models into lower resolution (>4 Å) maps, potentially integrating restraints from other experiments (e.g., cross linking mass spectrometry).¹²⁵ Efficient computations are achieved by generating “fit-libraries,” in which a global search is performed on each subunit independently, using the fit-to-map tool in UCSF Chimera.¹¹⁶ p -values are derived from these fit libraries, which determines the likelihood that a better fitting position can be identified, relative to each placed subunit. These p -values are integrated into the flexible Assemblin scoring function, which can include symmetry restraints, restraints from cross-linking mass spectrometry and pulldown experiments and structural restraints, for example, to favor interactions between trans-membrane domains and lipid bilayers. Structures are refined using Monte Carlo simulated annealing to optimize the placement and rotation of subunits for candidate models, which is greatly accelerated using the precalculated scores from the fit-libraries. This approach was used to generate models for the pathogenic type VII secretion system from *Mycobacterium smegmatis*,¹³² as well as multiple models of the human 10 MDa + Nuclear Pore Complex (NPC).^{133,134}

Several recent cryo-EM studies of the NPC also highlighted the utility of recent deep-learning structure prediction methods such as AlphaFold2¹³⁵ and RoseTTAfold¹³⁶ for model building.^{133,134,137,138} Two studies used cryo-ET with subtomogram averaging to generate low (20–30 Å) resolution maps of NPCs from yeast.^{133,137} In both studies homology models were used to generate models for the components of the NPC, one with Assemblin,¹³³ and the other with IMP.¹³⁷ In both cases, these approaches generated large models (~ 30 MDa), although both maps did contain unmodeled density (Figure 5). However, a significantly larger scale model was built by using deep-learning methods to predict the structures of all modeled NPC protein components, in addition to acquiring a higher resolution cryo-EM map. This facilitated building a much larger, more complete, 70 MDa, model composed of ~ 1000 subunits, into a 12 Å

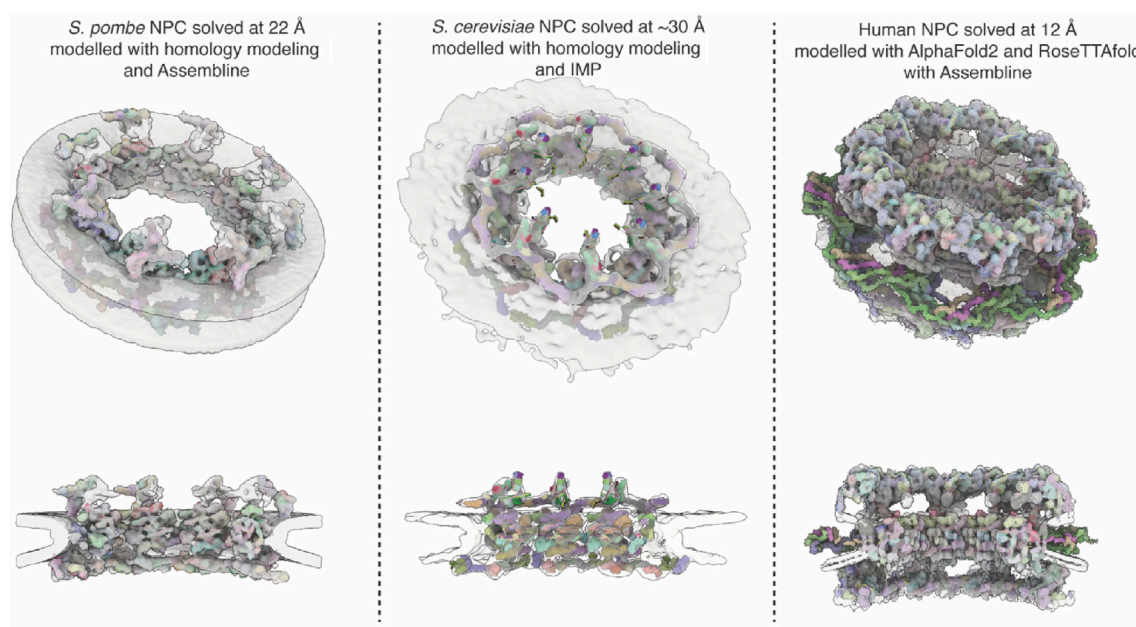


FIGURE 5 Modeling of the NPC facilitated by accurate protein structure prediction. Three cryo-EM reconstructions of the NPC are shown with the fitted models; from *S. pombe* (right, PDB-DEV ID: PDBDEV_00000094, EMD-11373),¹³³ *S. cerevisiae* (center, PDB ID: 7n9f, EMD-24258),¹³⁷ and human (right, PDB ID: 7r5k, EMD-14322).¹³⁴

resolution map of the human NPC (Figure 5).¹³⁴ A similar increase in model “completeness” was also achieved for the cytoplasmic ring of the *Xenopus laevis* NPC by integrating AlphaFold2 predictions.¹³⁸

5 | FLEXIBLE FITTING AND REFINEMENT

With atomic resolution being reachable,^{40–42} the range of map quality that is attainable in cryo-EM is very wide, necessitating different computational techniques for interpretation in terms of biomolecular structures. The range of features that can be resolved given a resolution, and in turn the techniques that are applicable to resolve these features, are presented in Figure 1b.

At the highest resolutions (better than 3–3.5 Å), tracing methods can unambiguously place atoms at local density peaks, using known geometries for common moieties, such as amino acids, nucleobases, water, ions, and common ligands. Many of those methods were first used and implemented for the processing of X-ray crystallographic data (e.g., REFMAC¹³⁹ and PHENIX¹³), and were adapted to be used in cryo-EM. At intermediate resolutions, tracing methods such as Pathwalking¹⁴⁰ or MAINMAST¹⁴¹ place pseudo-atoms within the density, which are then connected together to form the backbone chain. This reconstruction can be improved with manual inspection, as done in Gorgon¹⁴² or Coot¹⁴³ (see Section 8). After the initial reconstruction of the backbone chain, specialized tools can be used to identify the nature of amino acid side-chains such as Buccaneer,¹⁴⁴ ARP/wARP,¹⁴⁵ and Coot.^{146,147} These procedures can be used or repeated for specific parts of the model, after a well-fitted model could be generated, to further improve the quality of fit.^{130,148}

But at present, a large fraction of deposited EMDB structures fall into lower resolution ranges, with ~57% of deposited entries having a resolution worse than 4 Å, making this type of modeling directly from the density difficult (Figure 1b). As we move toward slightly lower resolutions, it is required to use an initial model, and iteratively refine its fit to the data. The initial model is often based on either a known atomistic structure from other experiments deposited in the PDB (e.g., from X-ray crystallography, NMR or cryo-EM) or homologous structures. With improvements in predictive models trained on the PDB, and in particular the rise of efficient machine learning methods that take advantage of this same data to create highly accurate models, using a prediction pipeline is becoming more common (see Section 8). We summarize many of the proposed refinement methods in the table below (Table 2). As the resolution of the map improves, methods that refine smaller structural features can be used.¹⁰

TABLE 2 Refinement methods used in cryo-EM, ordered chronologically

Method name	Core concept	Method name	Core concept
NMFF ¹⁴⁹	Normal mode-based refinement	ReMDFF ¹⁵⁰	Replica-exchange with fitting
RSRef ¹⁵¹	Real space refinement	Scipion ¹⁵	Full suite
Flex-EM/MODELER ¹⁵²	Correlation based flexible fitting	MAINMAST ¹⁴¹	Chain tracing
MDFF ¹⁵³	MD flexible fitting	Metainference with GMM ¹⁵⁴	Integrative refinement
Coot ¹⁴⁷	Interactive local refinement	PHENIX.real_space_refine ¹⁵⁵	Real-space refinement
REFMAC ^{139,156}	Maximum-likelihood	CDMD ¹⁵⁷	Correlation-based refinement
DireX ¹⁵⁸	DEN-guided fitting	ARP/WARP ¹⁴⁵	Real-space refinement
iMODFIT ¹⁵⁹	NMA-guided fitting	MDeNM-EMfit ¹⁶⁰	Normal modes and MD
Rosetta ¹⁶¹	Local density fitting	TEMPy-REF ¹⁶²	Ensemble refinement

An important technique for the simulation of biomolecules is MD, which allows for the sampling of conformational ensembles, as well as obtaining kinetic information. By integrating the information contained within a cryo-EM density map as an effective potential, it is possible to obtain conformations where the atoms are more likely to be found in regions of high density. MDFF¹⁵³ is the archetype of this class of method.

However, using the cryo-EM density as a potential has its own limitations: at lower resolution, the peaks in intensity from multiple atoms will effectively merge into one large peak (e.g., two close-by atoms may produce a single density peak unless the resolution is high enough). Attempting to drive those atoms within this density may result in unnatural structures. Given an estimate of the resolution, the forces can be computed to maximize the correlation between a simulated map at that resolution, and the experimental map. This approach was developed in Flex-EM/MODELER,¹⁵² which optimizes atomic positions using a scoring function that combines the CCC between the structure and the map with stereochemical and nonbonded interaction terms. Optimization by Monte Carlo search, a conjugate-gradients minimization, and simulated annealing MD is applied to a series of subdivisions of the structure. As the optimization progresses, the structure is divided into progressively smaller rigid bodies (e.g., using the RIBFIND clustering approach).¹⁶³ The method has been widely used for model fitting into cryo-EM maps at medium resolutions, for example to refine GroEL in multiple conformations, at 7–9 Å resolution^{37,164,165} or kinesin 8 bound to microtubules inhibited by the small molecule BTB-1 at ~5 Å resolution.¹⁶⁶ More recently, it has been adapted to the high-to-medium resolution range (3–5 Å),¹⁶⁷ often prior to refinement in Coot,¹⁴³ and Phenix,¹⁵⁵ as done for the human separase in complex with securin and CDK1–cyclin B1–CKS1.¹⁶⁸ Flex-EM can now be run via the CCP-EM suite.¹⁶⁹ Newer developments in MD-based methods such as CDMD¹⁵⁷ use an effective potential that maximizes this correlation rather than bias directly toward the density itself.

Normal mode calculation is another method that can be used to sample new structures, following low frequency modes, as in NMFF.¹⁴⁹ Normal mode analysis (NMA) is done by computing the second-order derivatives of an energy model (Hessian matrix) given a structure conformation. A commonly used model is an elastic model, with harmonic restraints put on close by atoms; this model can be used to easily compute normal modes,¹⁷⁰ or directly to move atoms.¹⁵⁸ Each mode is one eigenvector, with an associated eigenvalue. The modes with lowest eigenvalue (often referred to as “slow” modes) have been shown to correspond to global motions in protein structures. Protein structures can be refined by following those modes in the direction that maximize a correlation with cryo-EM density data, with limited local deformation of the structure, or used to bias or guide an MD procedure, as in MDeNM-EMfit¹⁶⁰ and NMMD.¹⁷¹

On the other hand, approaches that attempt to quantify the uncertainty in the data and iteratively improve the quality of the fit with respect to the model provide an alternative approach to obtaining high-quality structures. Maximum-likelihood method is used in REFMAC to improve the fit of a model to the data.¹⁵⁶ This software is also available via CCP-EM. Gaussian mixture models can be used to provide a practical representation of both experimental data and model.^{154,172}

Some statistical methods use existing structural data present in databases such as the PDB⁷⁴ or EMDB^{3,73} to provide high-likelihood structures, given a sequence and a map. Recent improvements in deep-learning methods have now found their way into refinement methods, by integrating it into custom pipelines.^{135,136}

The quality of predictions in the recent CASP14 competition showed striking patterns with respect to the map: regions of lower prediction quality were consistent across predictions from different groups, for a given map and these often correlated with ambiguous density, despite the fact that the models were generated without the knowledge of the cryo-EM map. Some protein domains are likely to have more structural heterogeneity and/or exhibit faster dynamics, which in turn causes the map to exhibit less well defined density for those regions. As larger complexes are being resolved at ever higher resolutions, it has become apparent that reconstruction (see Section 2) and fitting methods need to capture the inherent flexibility of those complexes. This can be done by using consensus approaches,^{173–175} In turn, it is not sufficient to improve the fit of a single structure to the data, since the density data itself reflects the existence of multiple structures, therefore one would expect that an ensemble of structures will provide a better representation of the data.^{176–178} Indeed, this has been demonstrated with TEMPy-REFF, a statistical method for real-space refinement based on mixture modeling. The ensemble of models generated via TEMPy-REFF can be expressed as an ensemble map that has a higher quality of fit to the data than any single model-derived map¹⁶² (see Section 7).

Even for single maps, considering an ensemble of structures may prove advantageous: early work on homology and ab initio modeling showed that considering ensemble of structures when fitting to an experimental map is informative both in terms of ranking candidate fits, and to identify best fitting structures.^{118,179} This problem has also long been appreciated and addressed in the processing of cryo-EM data, where 2D and 3D classification methods are used to reconstruct multiple densities, corresponding to multiple states (see Section 2). Refinement methods such as MDFF and Flex-EM in medium resolution maps or PHENIX and REFMAC in near-atomic resolution, can be used to generate

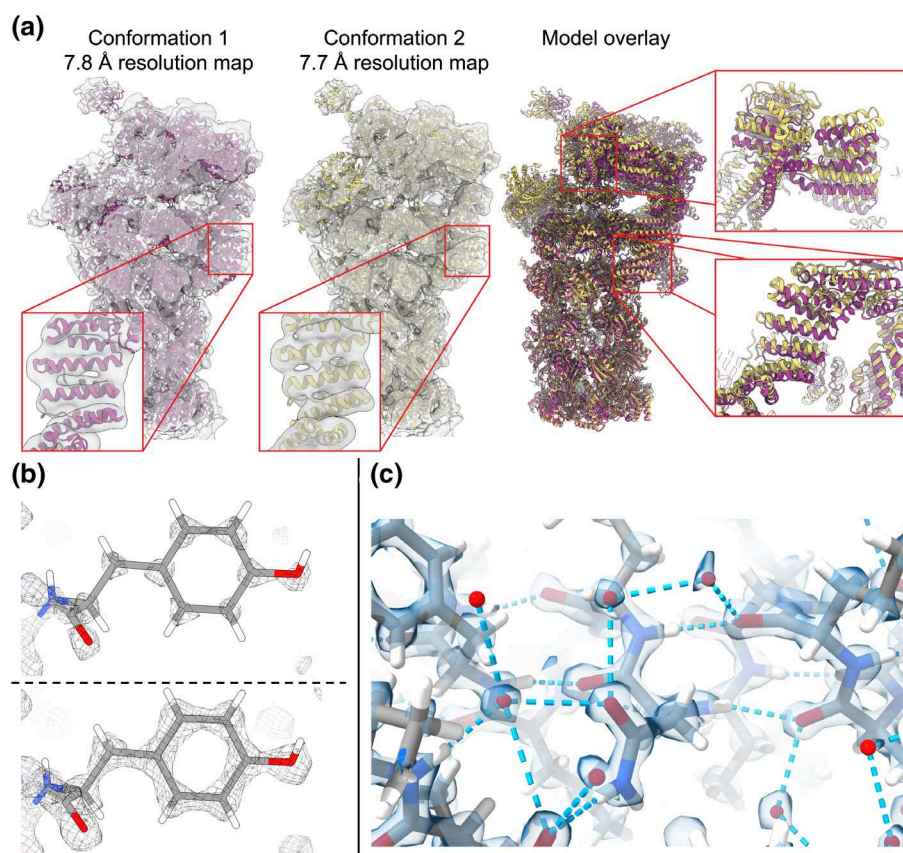


FIGURE 6 Cryo-EM model building across varying resolutions. (a) Medium resolution structures and corresponding cryo-EM maps (gray) of the proteasome in two conformations, conformation 1 (purple), and conformation 2 (yellow), bound to a slowly hydrolysing ATP analog.¹⁸⁰ The difference between the two refined models is highlighted (right, corresponding purple and yellow structures overlaid). (b) Example of a very high resolution map of apoferritin.⁴¹ At high threshold (top) only the cryo-EM density for heavy-atoms is visible, whereas at a lower threshold (bottom) hydrogen atoms start to be visible. The difference in cryo-EM density reflects expected electronegativities between atoms. (c) Another example of a very high resolution cryo-EM map of apoferritin where water molecules forming hydrogen-bonding networks are visible in the structure,⁴² with observed density in line with the expected partial charge differences between atoms.

well-fitted models for each of these maps, and allow the interpretation of their differences in terms of motions within structural elements.³⁷ An example of model refinement at multiple conformations (using MDFF) can be seen in Figure 6, where the ATPase subunit of the proteasome has been solved at medium resolution (~ 7.7 Å).

With improvements in instrumentation, notably optical hardware, and image processing methods,^{41,42,181} cryo-EM reconstructions can now reach true atomic resolution. In some maps, the densities of atoms are separate and clearly visible, including hydrogen atoms (Figure 6b,c).^{41,42} Fitting models into these very high resolution maps is often performed with methods from X-ray crystallography. For example, a model was refined into a 1.22 Å structure of aptoferritin⁴¹ (Figure 6b) using a combined approach with Coot,¹⁴³ phenix.real_space_refine,¹⁵⁵ and REFMAC.¹³⁹ Similarly, model refinement into a 1.25 Å map used Coot and REFMAC⁴² (Figure 6b). In the example of apoferritin from Reference 41 (Figure 6b), the densities are not centered on nuclei position, and show deformations consistent with charge effects deforming electron clouds as would be present in molecular orbitals. This also paves the way to direct observation of minute changes in different conditions, for example, pH changes,¹⁸² and the observation of ion binding sites and structural water molecules, which can also be used to improve subsequent computational studies of those systems, for example, provide better starting points for MD simulations.

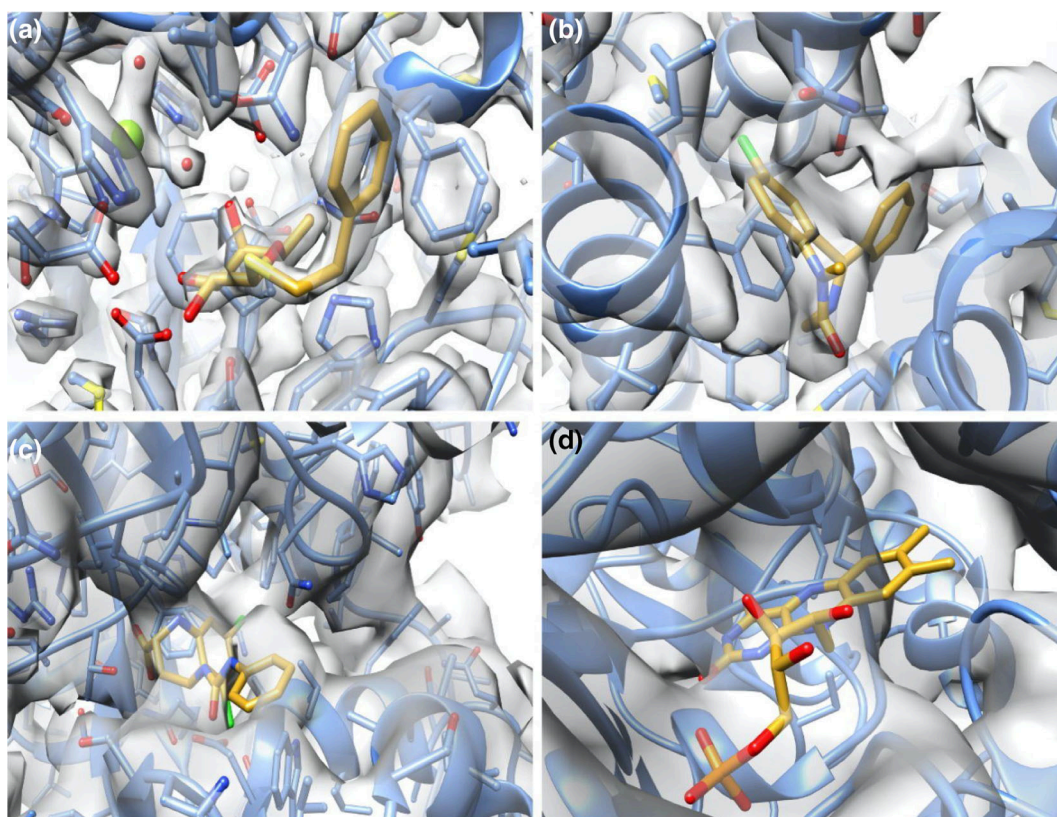


FIGURE 7 Small molecule binding sites at various resolutions. (a) An atomic model of a molecule of 2-phenylethyl 1-thio-beta-D-galactopyranoside (yellow) within a binding site on the beta-galactosidase protein (blue) within a cryo-EM map (gray) at 2.2 Å resolution. At this resolution the positions of protein side-chain atoms, solvent and ions in the map can be seen. (b) An atomic model of diazepam (yellow) bound in a subunit interface of the GABA_A receptor protein (blue) in a cryo-EM map (gray) of 2.9 Å. At this resolution the position of the ligand molecule can be seen clearly in the density along with the approximate positions of protein side chains. (c) An atomic model of the GABA_B receptor (blue) antagonist CGP54626 (yellow) in a cryo-EM map solved at 3.52 Å. At this resolution it becomes a lot harder to differentiate between protein and ligand density, although the approximate ligand position can be seen. Furthermore, only the positions of some large side chains are visible in the map. (d) A molecule of flavin mononucleotide (yellow) bound to the respiratory complex I from *T. thermophilus* (blue) within a cryo-EM map of 4.3 Å resolution. At this resolution it becomes almost impossible to distinguish the small-molecule within the protein density. Furthermore, no clear density corresponding to protein side chains is visible.

6 | FITTING SMALL MOLECULES INTO CRYO-EM MAPS

The increase in resolution of deposited cryo-EM structures has increased its relevance in drug discovery. The ability to resolve structures that classically have not been amenable to X-ray crystallography, has brought a wealth of new structures bound to small molecules, giving researchers a better understanding of the mechanism of action of such systems. Accurate small molecule refinement and fitting has been demonstrated in cryo-EM maps with resolutions >3 Å.^{183,184} This is particularly notable as it has been thought that higher resolutions would be needed for drug discovery. Although these resolutions are now achievable, the majority ($\sim 46\%$) of maps deposited in 2021 were between 3 and 4 Å resolution (Figure 1).

Methods for fitting small molecules into cryo-EM density maps have two general parts; an initial global fit of the molecule into the cryoEM map, followed by a further refinement in the context of the selected binding site. The chance of success of a fitting protocol is dependent on both the resolution of the map and the ability to generate a good initial approximation of the orientation and conformation of the small molecule within the binding site.

Issues concerning the resolution of the map can be further compounded by the fact that the resolution varies across the cryo-EM map. The local resolution is often worse at small molecule binding sites compared to the surrounding regions. This can be caused by a number of factors including: small movements of the molecule in the binding site, partial occupancy, and capture of heterogeneous binding modes of the molecule.

Should the resolution of the binding site be sufficiently high (generally, $<\sim 3$ Å) (Figure 7), the molecule can be directly refined into the cryo-EM density using structure refinement methods described above (see Section 5). One common strategy taken is to obtain an initial fit in the software Coot¹⁴⁷ manually place the ligand into the density, followed by using Coot's built-in tools such as the "jiggle-fit" protocol.¹⁸⁵ This protocol locally optimizes the placement of small molecules in density. Given a starting conformation, the translation and rotation values are subjected to small random perturbations. The generated conformations are rigidly fit to the map and the conformation with the best fit is selected and refined further in the context of the binding site, for example using the "phenix.real_space_refine" protocol¹⁵⁵ from the PHENIX software suite¹³ (see Section 5).

At high resolutions, generally, all atoms are optimized independently, while ensuring that geometric constraints such as bond lengths, bond angles and dihedral angles, remain within a reasonable range.¹⁵⁵ Geometric restraints for protein atoms are well known and generally have preferred values, this is not true for "non-standard" atoms, such as those in small molecules. Therefore, these geometry constraint values must be calculated before refinement and supplied along with the input. One method of achieving this is with the software eLBOW¹⁸⁶ supplied in the PHENIX package. This protocol has successfully been used in model-fitting pipelines for numerous protein ligand complexes obtained by cryo-EM at high resolutions including the GABAA Receptors in complex with various anesthetics and benzodiazepines (2.55–3.5 Å),¹⁸⁷ the TRPC6 receptor bound to the antagonist AM-1475 (3.08 Å),¹⁸⁸ the GLP-1 receptor complexed with agonist PF-06882961 (2.82–3.24 Å),¹⁸⁹ and the *Drosophila* solpoke potassium channel bound to various small molecule antagonists (2.59–2.73 Å).¹⁹⁰

More recently, a method for automatic determination of ligand restraints was developed with the addition of the OPLS3e forcefield¹⁹¹ into the "phenix.real_space_refine" protocol.¹⁹² The OPLS3e forcefield calculates fixed partial charges for small molecule atoms, and generates torsion angle restraints by comparing chemical moieties with values of a built-in database. This updated protocol was benchmarked against the original with a real-space refinement on 15 cryo-EM maps of resolution between 1.9 and 4.3 Å and model quality quantified using MolProbity^{193,194} and PHENIX-ramachandran Z-scores.¹⁹⁵ Little difference in model quality between the two protocols was observed, however, the inclusion of the OPLS3e forcefield led to atomic models with a significantly lower ligand strain energy.¹⁹²

Another program, GemSpot,¹⁸³ leverages molecular docking software to generate an initial fit of the ligand into the map, before applying real-space refinement with the "phenix.real_space_refine" OPLS3e forcefield protocol.¹⁹² Molecular docking aims to identify the approximate positions of atomic coordinates of a small molecule within a binding site, with a scoring function that approximates the free energy of binding and a search function that explores a large portion of the ligand binding site. An exhaustive search of the binding site would be computationally expensive, therefore many molecular docking softwares adjust the dihedral angles, translation and rotational values of small molecules in a random way. Many algorithms have been utilized for this including the genetic algorithm,¹⁹⁶ ant colony optimization,¹⁹⁷ or stochastic hill climbing,¹⁹⁸ to name a few. To speed up convergence to the optimum solution, a local search algorithm is often employed to further refine solutions, examples include the nelder-mead¹⁹⁷ and BFGS algorithms.¹⁹⁸ The accuracy of molecular docking is nowhere near that achieved by MD protocols, however the computational cost is a fraction of that for classical MD. GemSpot¹⁸³ utilizes the GLIDE molecular docking software search algorithm¹⁹⁹ with a modified scoring algorithm that is the sum of the GLIDE scoring function and the CCC of the simulated molecule with

the experimental density map. This has the advantage of being able to rapidly eliminate conformations of molecules that either have a low CCC to the experimental density map or have a worse docking score. In this way, the software automatically generates initial placements of ligands within the density that both have a good fit to the map and are chemically reasonable. A benchmark of the above mentioned 15 cryo-EM maps (to test “phenix.real_space_refine”/OPLS3e forcefield protocol) showed that solutions had a comparable CCC with the map as the original deposited structures.¹⁸³ GemSpot was also tested at lower resolution, showing greater deviations from the correct solution at 5.5 Å.

More recently, an extension to the PathWalking methodology²⁰⁰ (see Sections 3 and 5) for the ab initio generation of protein models saw the software identify and fit ligands into a protein model. Given a density map, protein model and map threshold cutoff at which ligand density is visible, the program first masks out density assigned to protein atoms. The remaining density is filtered, and density is assigned as either ligand, ion, or solvent based on the shape and distribution of pseudo-atoms fit to the density. Once density is assigned to a ligand, the ligand model is built and its fit assessed with standard correlation metrics. This method was shown to reliably fit ligands at resolutions better than 3.0 Å, with only partial success at resolutions close to 4.0 Å.

A protocol for fitting small molecules at lower resolutions (3–5 Å) based on the idea of using a molecular docking protocol for initial ligand placement followed by a real space refinement using neural-network potentials to score protein ligand interactions¹⁸⁴ has recently been developed. Once again, the GLIDE docking software¹⁹⁹ is used to achieve an initial fit of the protein/small molecule complex. This is then further refined using MDFF with neural-network potentials for the small molecule (i.e., classical force field parameters were used for protein atoms, while neural-network potentials were used for deriving small molecule parameters).

Neural-network potentials are an alternative approximation to using quantum mechanical calculations of molecular interactions, which, while accurate, are computationally expensive. Neural-network potentials are trained on large ensembles of quantum mechanical data for multiple protein ligand complexes, as a way to overcome the large computational costs of quantum mechanical calculations while retaining the accuracy of such calculations. It was reported that refinement using neural-network potentials improved the radius of convergence (i.e., the distance of the initial fit from the ground truth solution) where a correct solution can be achieved when compared to quantum mechanics/molecular mechanics-MD flexible fitting, when synthetic maps up to 4 Å were used.¹⁸⁴ Furthermore, by giving more weight to interactions between ligand atoms and protein atoms that were shown to be important from biochemical data, correct complexes were recreated with maps of 5 Å resolution.

A further strategy for refining small molecules into cryo-EM maps that has shown success with intermediate resolution maps included the use of correlation derived MD.¹⁵⁷ This protocol was shown to be able to accurately refine an atomic model of an *N*-ethylmaleimide sensitive factor bound to ATP at 3.9 Å. The model quality of the final refinement was seen to be higher than the deposited model, with a better fit to the map. When the refined model was compared to two high resolution x-ray crystallography structures, the angles of the ATP molecules were seen to be more consistent with what was reported for the x-ray structures than the deposited models.

A strategy for refining small molecules into low resolution (>4.5 Å) cryo-EM maps was touched upon by the paper that reported the use of Neural-network potentials with MD flexible fitting. This report was seen to recreate a correct fit at 5 Å when the strategy used biochemical data, while a refinement without such data failed to replicate a correct fit. The reason for this was that at this resolution density from the protein was bleeding into ligand density in the map.¹⁸⁴ This highlights one of the main challenges of fitting small molecules at this resolution range, where there is simply not enough information for the accurate placement of ligands. Nevertheless, strategies for achieving this have been reported. Two studies report a strategy involving creating a difference map²⁰¹ to mask out protein density leaving only density corresponding to the ligand for fitting.^{166,202} First, the protein model is refined into the experimental map and a synthetic map is simulated from the protein model. The amplitudes of the two maps are scaled to match and the density in the simulated map is subtracted from the experimental map leaving density corresponding to the ligand as a density difference map. In both reports, molecular docking was used along with the CCC of generated solutions to the difference map in order to find an initial placement of the ligand in the map. The top solutions were then flexibly refined into the experimental map. When both cases the approximate positions of the ligand corresponded well with experimental mutagenesis data, with respect to the distance from residues involved in drug action. However, neither was able to confidently identify a single solution, and in both cases an ensemble of high scoring solutions was reported.

This highlights the difficulty of fitting small molecules at this resolution as there is not enough information present in the experimental density maps for an accurate calculation of protein ligand complexes. However, it does seem that combining fitting with biochemical data can improve calculations.

7 | VALIDATING MODEL-MAP FIT

Once a model has been generated by rigid or flexible fitting (Sections 5 and 6), it is vital to assess whether it is an accurate and realistic (i.e., stereochemically plausible) representation of the molecule imaged in the context of the cryo-EM map. This is the goal of model validation, where a candidate model can be investigated in greater detail compared to during model refinement, where rapid evaluation of multiple models at each iteration is required. In an ideal scenario, model validation metrics should be both sufficiently informative to the researcher to identify an ideal candidate model/s, and sufficiently simple to allow non-experts to trust the deposited models that they will use as a basis for further research.

7.1 | Model-map fit scoring functions

An essential component of model validation is ensuring that a proposed model is an accurate fit to the experimental cryo-EM density. Ideally, the metric used to assess model fit should not be the same metric, which is optimized during model refinement.

Cross-correlation (CC) is the most common scoring function for assessing the similarity between a map and model but there are other scores, including those which rely on comparing surfaces of maps (e.g., the Normal Vector score or the Chamfer Distance) rather than the entire density.^{127,174,203} CC evaluates model fit by assessing the similarity between two pixelated volumes. In the case of model evaluation these would be the experimental cryo-EM map and a simulated map generated from the model (by calculating the normalized product between the densities in these two maps).^{203,204} Generally, the two maps are zero-mean normalized during calculation of the CC, such that the CC reports the covariance between them and is not biased by an offset between density values in each map. Variations of the CC score include, for example calculating the score, only on a subset of the map density values (e.g., on voxels within a mask or voxels above a set threshold). Alternatively, cryo-EM maps can be filtered prior to CC calculation, for example Laplacian filtering,²⁰⁵ that calculates the second derivative of the cryo-EM density to enhance edge features in the map. These variations are implemented in different software such as PHENIX, as CC_{mask} and CC_{vol}, respectively,²⁰⁴ and TEMPy.²⁰⁶

As mentioned above, the Fourier Shell Correlation (FSC) is another commonly used measure of model quality. For model-to-map fit assessment, first, a simulated map must be generated from the model, and the FSC computed between this and an experimental map. A single resolution value can be extracted from the FSC curve at a specified cutoff, generally 0.5, which reports on both model and map quality.²⁰⁴ Alternatively, it has been shown that integrating the FSC curves across the mid-low resolution range (6–10 Å) produces a score (FSCi) that is slightly more discriminative than real space correlation (CC).²⁰⁷ Additionally, the FSC_{average} is an implementation of map-model FSC that normalizes the correlation between each shell, based on their number of Fourier components.¹⁵⁶ Finally, by refining the model into one half-map, the FSC curve has been used to detect model overfitting (in which the model is refined against noise in the maps, rather than signal), by using the second half map as an independent test case.^{156,157,207–209} Multiple software packages implement the map-model FSC, and its variants, including Phenix,^{13,204} Rosetta,¹⁴⁸ REFMAC,¹⁵⁶ and TEMPy.²⁰⁶

A global score is necessary to evaluate overall model agreement with data, but localized deviations between the model and the experimental map can be obscured within this global average. Such deviations can be identified and investigated using “local” scoring functions (Figure 8), which can assess model fit for example, at the level of secondary structure, individual residues or even single atoms, depending on what the resolution allows. This kind of local assessment can guide the refinement process at different stages of flexible fitting.¹⁶⁷ In the following section, we introduce some popular local scoring functions, and then evaluate studies that have compared their effectiveness for scoring models in cryo-EM maps.

The CC score has been extended for use as a local scoring function in multiple packages.^{203,206,212} In Chimera and ChimeraX the CC can be calculated on a selected segment.^{116,213} Another example is the TEMPy python library,^{174,206} which contains implementations of CC (or scores similar to CC such as Mander's Overlap) as local scores. PHENIX²¹⁴ and EMDA²¹² also have their own local CC implementations; for example, in Phenix, the CC_{box} score can be applied to local regions around specific molecules or residues.²⁰⁴ For example, the segment based cross-correlation (SCCC) score is calculated for subsets of the model and corresponding portions of the map, depending on the protein's segmentation into “rigid bodies” (Table 3). These rigid bodies are essentially segments of the protein structure or sequence, for

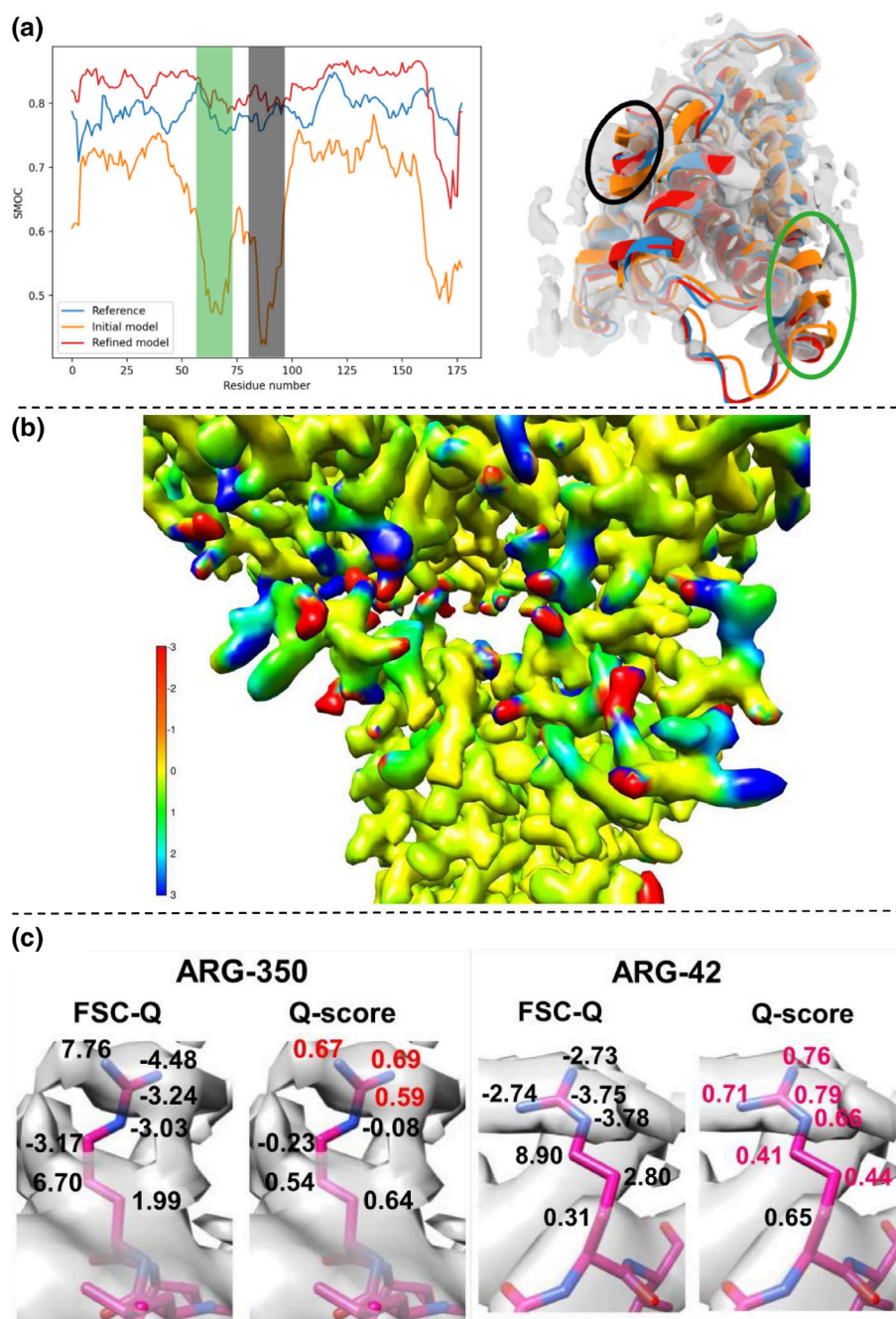


FIGURE 8 Local fit assessment. (a) SMOC scoring to evaluate target T1099 (EMD-10800) from the CASP14 competition. The residue-by-residue SMOC plot for three models is shown: the reference structure (blue), the best predicted structure before refinement (orange), and after refinement with TEMPY-REFF^{162,210} (red), with the respective models shown in the experimental cryo-EM density (right). Figure adapted from Reference 210. (b) FSC-Q values calculated for the PAC1 GPCR receptor complex (PDB ID 6p9y) superimposed onto the cryo-EM density from EMD-20278. The range of FSC-Q scores is shown in the color bar; high scoring regions (colored in red) correspond to portions of the model that lie within the detergent micelle (overfitted atoms) whereas blue regions correspond to poorly-fitted portions of the model. (c) FSC-Q and Q-scores shown side by side for specific residues (Arg350 and Arg42) shown to emphasize poorly-fitted residues that are identified by FSC-Q, but scored highly by Q-score. Incorrectly high Q-scores are highlighted in red. Figure adapted from Reference 211.

example individual α -helices or β -sheets, and can be identified using the software RIBFIND.^{163,218} This approach was used to identify, and flexibly fit, poorly fitting segments of Cocksackievirus A7 capsid protein into two low resolution maps (8.2 and 6.1 Å) that represented specific conformations associated with capsid maturation.¹⁷⁵

TABLE 3 Local scoring functions used in cryo-EM model building

Method name	Core concept	Structural level	Software package
SCCC ¹⁷⁴	Cross correlation	SSEs	TEMPy, ²⁰⁶ CCPEM ¹²
CCbox ²⁰⁴	Cross correlation	Residues	PHENIX ^{13,214}
EMRinger ²¹⁵	C _γ atom χ^1 dihedral angle	Residues	PHENIX
SMOC ¹⁶⁷	Manders' overlap coefficient	Residues	TEMPy, CCP-EM
Q-score ²¹⁶	Measuring signal fall-off	Atoms	Chimera ¹¹⁶
FSC-Q ²¹¹	Deviation between local map resolution and local map-model FSC	Atoms	Scipion ¹⁵
EMDA local correlation ²¹²	Cross correlation	Residues	EMDA ²¹²
False discovery backbone score ²¹⁷	False discovery rate	Residues (C _α)	CCP-EM
LoQFit ¹⁶²	Map-model FSC	Residues	TEMPy

For higher resolution cryo-EM maps (<4 Å), it is possible to interrogate the model fit in more detail.

EMRinger is a somewhat unique score that focuses on a very specific feature of a protein model: the position of each residue's C_γ atom χ^1 dihedral angle.²¹⁵ The score is calculated by extracting the map densities around each dihedral angle, and searching for peaks in density values at 60°, 180°, and 300°, which are known to be favored positions for the C_γ atom, based on high-resolution X-ray structures.^{215,219} This makes the EMRinger score informative for both the local map and model quality, as a clear peak at an expected position indicates a residue is within a well-resolved portion of the map, and that it is positioned correctly.

Another score interrogating the model at the residue level is the SMOC score, which is computed using the Manders' Overlap coefficient (MOC), another early-developed score that captures per-residue local fit.¹⁶⁷ The SMOC is similar to the CCC, but not normalized around the mean values of each map (cryo-EM map and simulated model map). There are two methods available to calculate the SMOC score for a model: SMOCf, which is calculated as the central residue in overlapping windows along the protein sequence, or by selecting map voxels within a cutoff distance to each residue (SMOCd). Scoring each residue in the protein using either of these implementations gives a plot of the local model to map agreement. SMOC was also applied to validate predicted structures submitted to the CASP13²²⁰ and CASP14²¹⁰ competitions (Figure 8).

As cryo-EM maps were reaching atomic, or near atomic (<3 Å), resolution, it became feasible to score the fit of individual atoms in a model. The Q-score measures each atom's "resolvability," by computing a density fall-off around each atom in the model, which is then compared to a reference Gaussian.²¹⁶ This reference Gaussian is generated in real space using values based on a high quality, high resolution (1.54 Å) apoferritin map (EMD-9865). This means that the Q-score is closely correlated with the local resolution of the experimental map, as the signal fall-off measurement is essentially probing the local resolution.

FSC-Q is another scoring function that evaluates the fit of individual atoms, using an approach similar to the assessment of local resolution by FSC.²⁶ The score essentially evaluates the difference between the local resolution of the experimental cryo-EM map, and the local FSC of the model fit into this map. Using this approach, FSC-Q can identify overfitted portions of the model. FSC-Q has been shown to be complementary to real-space correlation-based scores, such as Q-score, which can erroneously give good scores to atoms lying in noise or artifacts in cryo-EM maps (Figure 8b,c).

The relative performance (capacity to identify poorly modeled regions) of many global and local scoring functions was recently assessed in the evaluation of CASP13 cryo-EM model targets²²⁰ and the 2019 Cryo-EM modeling challenge.²²¹ In both studies the relative performance of multiple scoring functions were evaluated (Q-score was not included in the CASP13 cryo-EM model evaluation), with one important difference between the two assessments: in the CASP13 model evaluation, the submitted models were predicted from the sequence alone (i.e., they were not modeled against the cryo-EM density), and thus the quality of these models varies quite widely. This is in contrast to the 2019 cryo-EM modeling challenge, where research teams were invited to submit models for 4 maps; 3 maps of Apoferritin at reconstructed from the same dataset at varying resolutions (1.8, 2.3, and 3.1 Å) and a horse liver alcohol

dehydrogenase (ADH) map at 2.9 Å resolution. This meant that all models were a close fit to density, compared to those submitted to CASP13.

The results of the 2019 cryo-EM modeling challenge showed that there was overall poor correlation between scoring functions, and instead three clusters of similarly performing scoring functions were identified (Figure 9a). The first two clusters were notable by their relationship to the global and local resolution of the map: CC related scores (including SMOC, Figure 9a) scaled inversely with map resolution, that is, scores were lower for higher resolution maps. The inverse was true for the cluster of scoring functions that included EMRinger, Q-score, and FSC05, all of which give higher scores on average as the map resolution improves. Finally, a final cluster was identified that contained multiple

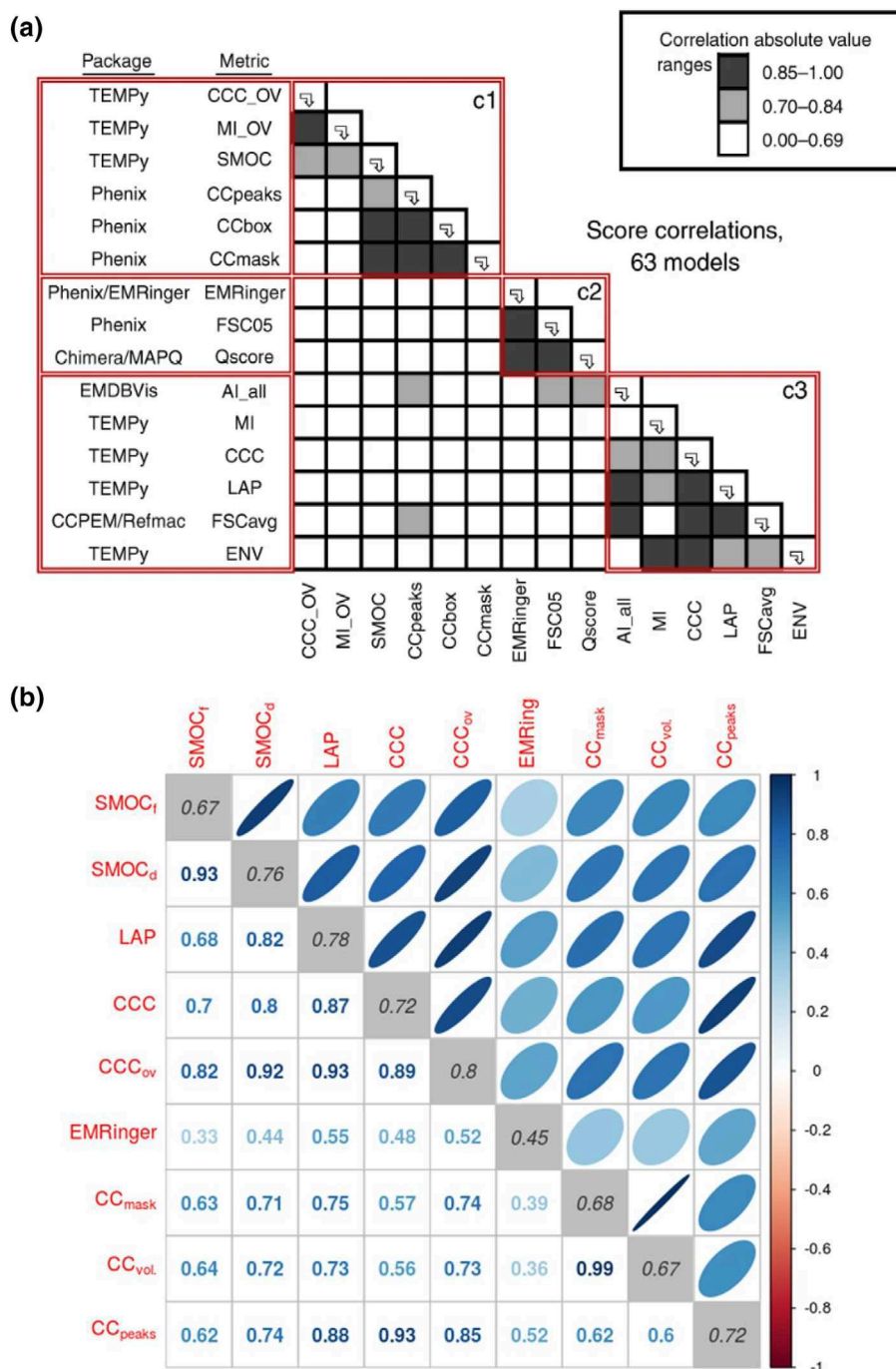


FIGURE 9 Comparing the performance of scoring functions for models in cryo-EM maps. (a) Pairwise correlation of scores for all models against all maps in the 2019 Cryo-EM modeling challenge. Figure adapted from Reference 221. (b) Pairwise correlation of scoring functions used to evaluate models submitted to the CASP13 competition.²²⁰

TEMPy scoring functions and FSC_{avg} score from REFMAC.¹⁵⁶ The scores in this cluster were all sensitive to noise in the solvent region of the cryo-EM map, compared to the other scores, which typically masked the maps around the protein signal. A somewhat similar grouping of scoring functions was observed in the CASP13 evaluation of cryo-EM targets,²²⁰ where most scoring functions showed high pairwise correlation (Figure 9b) except EMRinger, which did not correlate well with any other evaluated scoring function. However, there is the caveat that the models evaluated in CASP13 were predicted based on sequence alone, that is, without access to the cryo-EM density, so “high-resolution” scores like EMRinger, are likely to give poor scores for all but the very best predicted structures, given that correctly placing amino acid side chains in predicted models is very challenging.

7.2 | B-factors in model fitting and assessment

In addition to local deviations between the model and map, there is often substantial resolution heterogeneity across cryo-EM maps.^{26–28} Generally, the resolution decreases toward the edge of the imaged molecule, where residues are increasingly exposed to solvent, and therefore increasingly mobile. This mobility causes blurred density (i.e. worse local resolution) in the cryo-EM map, due to unaligned signal between the averaged particles in reconstruction. Similar effects are observed at mobile portions of molecules, such as loops. This local mobility can be represented within the model using the B-factor parameter (also referred to as atomic displacement factor [ADP] or temperature factor) for each atom in the model. Given B-factors are calculated per-atom, it is only practically possible for maps with resolution better than 5 Å. B-factor refinement is integrated into multiple model refinement methods.^{148,161} However, only ~50% of submitted models for the 2019 cryo-EM Model Challenge included refined B-factors,²²¹ and refined B-factors are largely not present in older models from cryo-EM maps.²²²

Atomic B-factors can be calculated from cryo-EM maps by multiple methods; cryo-EM maps can be converted into structure factors and B-factors can be calculated as they would be in an x-ray crystallography experiment.²²³ Alternatively, many packages calculate B-factors in real space, by optimizing the real space correlation between the map and model, while adjusting the atomic B-factors.^{148,161} Finally, B-factors can be approximated using each atom's Q-score,^{216,224} as the Q-score interrogates the local signal fall-off and is therefore related to the B-factor. A recent study assessed the accuracy of each of these methods for B-factor determination,²²⁵ by assessing the similarity between B-factor weighted simulated maps and experimental maps. They found the Q-score derived B-factor was the most accurate. However, a downside of the Q-score derived B-factors is they require a global scaling factor, that must be individually optimized for each assessed map. In a more recent study, TEMPy-REF¹⁶² used B-factors for refinement based on the sigma of the Gaussian function describing each atom, from a Gaussian mixture model representation of the protein.

Taking into account these refined B-factors when simulating cryo-EM maps, by using them as weighting factors, improves the accuracy for local and global scoring functions, relative to maps simulated with a constant blurring factor (Figure 10a).^{162,226} Second, the B-factors themselves can act as a tool for validation, since they reflect the local flexibility of the individual atoms. Indeed, such methods were used in part to identify fraudulent x-ray crystallographic structures.²²⁷

Singharoy et al. used the root mean square fluctuation (RMSF) as a similar value to B-factor.¹⁵⁰ The RMSF is calculated based on the MD flexible fitting (MDFF) method, which provides a distribution of fitted models. The local energy of these simulations scale based on the potentials within the cryo-EM map, meaning the RMSF score scales with the local quality of the map and reports directly on the local “reliability” of the model. Similar approaches have been used for both model and map validation.^{148,228–230} In another study,²²⁸ 10 individual models were calculated for each chain in the refined structure, and the standard deviation of each atom between the 10 structures was used to define an approximate B-factor. This idea of using an ensemble to model the uncertainty in the cryo-EM map (Figure 10b) has been used in multiple studies, including modeling of Tobacco Mosaic Virus structures at 4.3 Å¹⁷⁷ and in fitting into medium-to-low resolution maps.¹⁷⁴ This ensemble representation is similar to that used in NMR, where multiple models that are consistent with the experimental data are often deposited.²³¹ This approach was extended in a recent study¹⁶² where B-factors derived from a GMM were used to generate model ensembles. These ensembles were then used to generate an ensemble map, which was shown to be a more accurate representation of the experimental cryo-EM density than the one obtained with maps simulated from a single model (Figure 10c). Indeed the authors showed that the correlation increases as more models are included in the ensemble, up to a saturation of the CCC (Figure 10d).

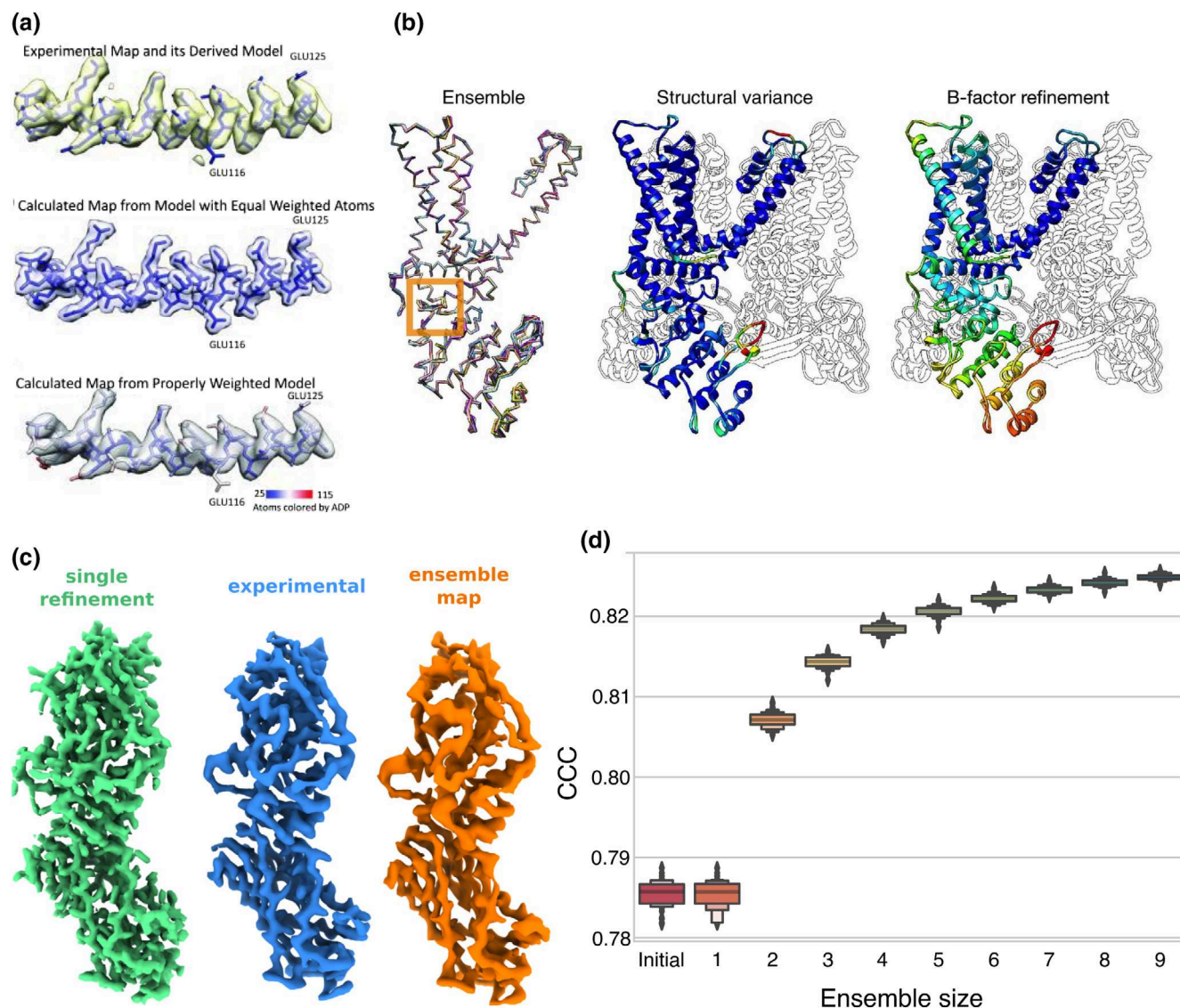


FIGURE 10 B-factor weighting and model ensemble representation to reflect variability in the data. (a) A subsection of an experimental map is shown (top) with a simulated map generated with a constant weighting (middle) and a simulated map using atomic B-factor weights (bottom). Modified from Reference 226. (b) The relationship between structural variance in an ensemble of the top-10 refined models from Rosetta automated refinement (ensemble shown left, structural variance center) and the B-factors from a representative model from the ensemble (right). Figure modified from Reference 148. (c) A simulated map from a single model (left, green color) is a poor match to the experimental density (center, blue) compared to the map generated from the model ensemble (orange) which better follows the experimental density. (d) Plot showing the CCC between EMD-9631 cryo-EM map and ensemble simulated maps with increasing number of models in the ensemble. (c) and (d) Adapted from Reference 162.

7.3 | Protein stereo-chemistry methods

Validating a model must not only take into account the fit to cryo-EM density, but also evaluate the agreement with known stereochemical properties of the imaged molecules. Typically, this assessment is done by comparing the geometry and stereochemistry of a candidate model to known baselines generated from reference datasets of high quality models. However, one must exercise caution when interpreting these values; it is commonplace to apply geometric restraints when refining models into maps with resolutions worse than 3 Å. These restraints include tightly restricting Ramachandran angles and backbone angles.²³² Model refinement methods that integrate MD simulations also intrinsically restrain some stereochemical properties through the force fields used in refinement, leading to models with less outliers, compared to those produced by other methods.^{150,162} This can complicate model validation; applying excessive

structural restraints during refinement can generate models that appear perfect, due to an absence of atom clashes or unfavorable stereochemical arrangements, but which are a poor or incorrect fit to the experimental data.^{162,233,234}

MolProbity is the most widely-used validation package for model assessment. Originally developed to validate models from X-ray crystallography, MolProbity scores models by comparing a candidate model to a carefully curated selection of 8000 high-quality and reliable models (termed the Top8000).²³³ Using this approach, MolProbity identifies problematic features, including Ramachandran outliers, deviations in bond lengths, isomeric outliers, and many other properties, including some specific to RNA and DNA.^{194,233,235} This toolbox was more recently extended to specifically deal with common problems present in models derived from medium-resolution cryo-EM maps (>3 Å) with the CaBLAM (C-Alpha Based Low-Resolution Annotation Method) tool.^{233,236} CaBLAM uses C α angular distributions to assess main-chain geometry and identify secondary structure. Although the C α trace is a prominent feature in cryo-EM maps at ~ 3 Å resolution range, the carbonyl bonds of the backbone are not visible, and peptide bond angles are commonly incorrectly orientated. CaBLAM uses a sliding 5-residue window to detect such C α outliers directly, flagging backbone and conformation mis-assignments directly. If uncorrected, these deviations can have an additive effect on downstream residues, leading to large Ramachandran outliers that are often moved to local minima when corrected, sometimes masking the underlying problem rather than fixing it.²³³ This issue was well demonstrated in the 2019 Cryo-EM modeling challenge, where more than half of submitted models contained no Ramachandran outliers, while only four had no CaBLAM outliers.²²¹

Other validation functions include statistical potentials of mean force by comparing a candidate protein to a reference dataset, including DOPE²³⁷ and QMEAN.²³⁸ More recently, AI-based approaches have been used to integrate multiple structural features into a combined score, these include ProQ3D,²³⁹ the MODFOLD7 server,²⁴⁰ and multiple scores calculated using the DeepRank network.²⁴¹ These methods have been used successfully to score predicted models in the CASP competitions,²⁴² but are also applicable to models refined into cryo-EM density.

Recently, machine-learning methods were also used to develop a more specific scoring function, Pi-score, that focuses on evaluating the quality of protein-protein interfaces between subunits in a model.²⁴³ Here, 12 interface features were used derived from high-resolution crystallographic datasets to train a support vector machine (SVM) classifier. The SVM was then used to score the interfaces in candidate models, based on their classification with the positive or negative datasets, and their distance from the hyperplane. The Pi-score was shown to be an effective validation score for CASP13 models, correlating well with other multimer scoring functions²⁴³ and has recently been used to aid in model building for the massive NPC structure assembled partly from AlphaFold2 predicted models.¹³⁴

8 | VISUALIZATION TOOLS FOR CRYO-EM STRUCTURE AND MAPS

A core, if somewhat obvious, objective in structural biology is to learn something new about the imaged molecule by visually inspecting its generated model. The first atomic structures were solved by x-ray crystallographic experiments with the aid of drawings and the assembly of physical ball-and-stick models. With the advent of computers with graphical displays, the visualization process shifted to software-based rendering. Today, the importance of visualization in the field of structural biology, is reflected by the sheer number of visualization tools available. Various ways of illustrating chemical and molecular features have been developed to aid various tasks. For example, visualizing atomic models is often enhanced with surfaces rendered at the van der Waals radius of the atoms. These surfaces can be colored according to local electrostatic and hydrophobic potentials, which could be useful for tasks such as computational drug design. In contrast, Richardson “ribbon” diagrams eschew atomic details instead focusing on the path of the backbone and the secondary structures it forms. Today, ribbon diagrams are now perhaps the most ubiquitous form of protein depiction in use.

While software has made the depiction of models a mostly automated task, building the models still requires significant user intervention. In crystallography, where high resolutions are frequently achieved, visual inspection of model and map in a residue-by-residue fashion is recommended.²⁴⁴ This advice is also valid for high-to-medium resolution cryo-EM maps. However, when only lower resolution is achieved, integrative approaches are often needed, and these, in addition to the cryo-EM information, can include existing solved structures, machine-learning models, chemical and physical restraint libraries, as well as expert insight. Regardless of the source of information or the computational methods used, it is often ultimately the responsibility of the biologists to assess these theoretical structures and their various scores and criteria. Thus, visualization in the domain of model building has an important additional aspect: to aid the model building process by conveying information from these different sources.

Before exploring the various tools, which are now available, it is worth briefly exploring how we have arrived with such a diverse offering of software. Many of these tools were developed for specific tasks. VMD for example,²⁴⁵ was developed to visualize MD simulations from the NAMD software.²⁴⁶ Due to the complexity and expertise required to build robust visualization software, many of the most popular tools have convergently evolved to support plugins and extension mechanisms, allowing their powerful graphical capabilities to be adapted to new tasks. VMD now supports other MD engines such as GROMACS,²⁴⁷ Chimera¹¹⁶ and its successor ChimeraX²¹³ were also originally developed with extensibility as a central goal, largely due to the limited capacity for extension in their predecessor software Midas.^{116,248}

While the visualization software space may seem overcrowded, competition between packages has been an important source of innovation. One of the challenges of presenting 3D data on a 2D display is conveying depth. A number of techniques to improve “depth cueing” such as ambient occlusion were pioneered in QuteMol²⁴⁹ and were swiftly adopted by other visualization tools. In the Richardson group, independent implementations of high-dimensional visualization features in two kinemage programs KiNG²⁵⁰ and Mage²⁵¹ ultimately lead to a “more general-purpose and more powerful” implementation.²⁵⁰

Visualization tools typically provide more than just the ability to inspect atomic models and volumetric data, but also the ability to either build or modify atomic models. Many map related tasks such as sharpening, cropping, and contouring are also greatly aided by the immediate feedback of interactive graphical tools. Tools such as Coot¹⁴³ aid this manual validation and fitting process and offer tools for fixing incorrectly modeled regions such as jiggle-fit¹⁸⁵ and backrub.²⁵² ISOLDE,²⁵³ which is developed as a plugin for ChimeraX and builds on the density guided force-field of MDFF,¹⁵³ provides an interactive MD simulation for fitting atomic models to density. ISOLDE and Coot are both aimed at high-resolution volumetric data where density often provides enough information to identify and fit models by visual inspection. While it is still recommended to inspect the fit to density of the resultant models using visualization software,²⁴⁴ as the resolution worsens, visually assessing fit becomes increasingly difficult as atoms and residues become less visible. However, secondary structures such as alpha-helices and to a lesser extent beta-sheets, remain recognizable both by eye and using automated techniques, between ~4.5 and 10 Å resolution. Gorgon¹⁴² is an interactive tool, which combines secondary structure detection using SSEHunter⁷⁷ and sequence data with visualization. Sculptor¹¹² combines the pattern matching and feature vector based fitting methods of Situs¹¹¹ along with interactive peak picking for exploring candidate solutions. Gorgon and Sculptor are both standalone tools which may explain why they never gained significant traction. Plugins and integrated tools get to exploit the features and inertia of the host visualization software and its ecosystem. Below are some popular tools used to visualize cryo-EM maps and fitted models:

8.1 | ChimeraX

As the successor to Chimera, ChimeraX builds on the core and extensions architecture of its predecessor but offers improved graphics, performance, and new features. ChimeraX's new rendering engine supports an implementation of ambient occlusion that runs on commodity hardware.²⁵⁴ To handle the ever larger molecules and assemblies which are being generated, representation of biological information is now largely achieved through performant array-based data structures rather than the earlier MMTK²⁵⁵ structures used in Chimera. Besides excellent tools for working with density maps, ChimeraX offers support for the mmCIF format, which is becoming an important exchange format in the integrative/hybrid modeling community.^{110,113} With new widgets for visualizing information such as from cross-linking experiments, ChimeraX is an excellent choice for exploring such data. The most recent 1.3.0 release offers integration with the new EBI AlphaFold2 database^{135,256} as well as tools to run AlphaFold2 on amino acid sequences.

8.2 | Isolde

ISOLDE is a popular plugin for ChimeraX that extends its visualization and presentation of data, by offering a mechanism for fitting structures to density interactively (Figure 11). ISOLDE offers an MDFF style density-based potential along with adaptive distance restraints.²⁵⁷ The user plays a pivotal role in fitting the structure by adding restraints and dragging atoms into density. Like Coot, ISOLDE provides a number of useful visual cues such as an interactive Ramachandran plot and a traffic light coloring scheme to indicate C-alpha atoms with problematic torsion angles. ISOLDE also offers adaptive distance restraints for maintaining local geometry.

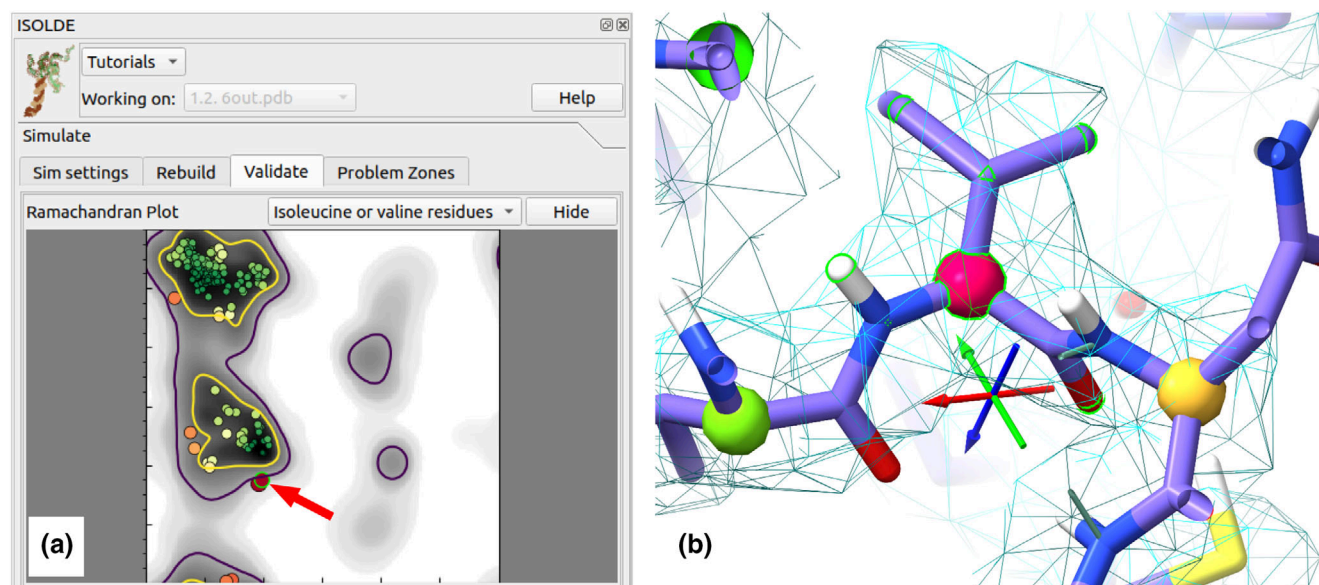


FIGURE 11 Snapshots from ISOLDE²⁵³ implementation in ChimeraX.²¹³ (a) An interactive Ramachandran plot, which shows in realtime the Ramachandran angles of the atomic model as the simulation progresses. Residues with problematic geometry can be clicked on (indicated by red arrow) which centres the 3D view on the residue (b, outlier residue highlighted with red ball). ISOLDE offers a number of facilities to correct these issues automatically or by manually tugging on atoms.

8.3 | iTEMPy

iTEMPy (Interactive TEMPy) is a TEMPy2²⁰⁶ plugin for ChimeraX which aims to provide an accessible user interface for scoring and visualizing models at worse than 5 Å. Like ISOLDE, scoring and fitting occur simultaneously and can be directed by the user interactively. However, unlike ISOLDE, which focuses its validation on model geometry, Interactive TEMPy offers feedback on local model-map fit using SMOC scores.¹⁶⁷ This is necessary as at medium and worse resolutions it is difficult to judge the fit by eye. The tool aims to provide a flexible environment for working with models and maps allowing users to use ChimeraX's facilities for fitting models and shaping maps. Due to the low resolutions targeted by the tool, a combination of MDFF style density-driven fitting and a rigid-body fitting approach similar to Flex-EM are provided. Here, users are able to break the model into small rigid-bodies by either manually grouping secondary structure elements or by using the integrated RIBFIND²¹⁸ widget. Large proteins and cryo-EM maps are particularly challenging for visual inspection. A SMOC plot which allows interactive “zoning” of the density map aids this process.

8.4 | Coot

A popular tool for model building and refinement, Coot¹⁴³ provides a set of features for working with high resolution (2.5–4 Å) density maps. In addition to backrub and jiggle-fit, Coot offers routines for loop modeling and “fragment” (short polypeptides sequences) oriented tools. Coot is extendable through Guile Scheme (<https://www.gnu.org/software/guile/>) and Python scripts. Besides visualizing structures and density, a number of methods are provided to visualize various validation metrics and make their inspection more ergonomic. Such features include the interactive Ramachandran plot as well as cues for clashes and a per-residue fit-to-density scores. A recent addition which greatly aids flexible fitting of models are the German-McClure adaptive distance restraints.²⁵⁷ These are able to maintain local-geometry but become weaker over longer distances allowing long range flexibility and conformational changes to take place.

8.5 | VMD

Originally designed to render MD trajectories from the NAMD package,²⁴⁶ VMD²⁴⁵ has a long history of visualizing dynamic molecular information. MDFF¹⁵³ and xMDFF²⁵⁸ are plugins for VMD which pioneered the density based

fitting approaches now used in ISOLDE. NMWiz²⁵⁹ is another plugin which allows the normal modes of molecules to be calculated and visualized with the aid of arrows indicating vibrational modes. VMD allows extensions to be written in C, C++, Tcl/Tk, and Python. VMD supports an optional ray-tracing rendering engine for the production of high-quality images, however a CUDA compatible NVIDIA graphics card is required to use this feature in real-time.

8.6 | PyMOL

PyMOL is a popular tool for visualization of models, particularly for drug design and rendering high-quality 3D images. As the name suggests, PyMOL is scripted via the Python programming language. Although the programming language was then relatively young, the choice paid off, with Python now being a popular programming language in the scientific community. Originally developed as an open-source project under the more permissible Python License it is now distributed as proprietary software. PyMOL has an inbuilt ray tracer for producing photo realistic renderings and can optionally export scenes for the PovRay renderer. This has made PyMOL a popular tool for producing images for publication. Its extensibility has led to the development of a wide range of plugins for tools such as AMBER²⁶⁰ and AutoDock Vina.¹⁹⁸ PyMOL has the capacity to render density from x-ray and cryo-EM experiments, however it offers less functionality for modifying this type of data compared to ChimeraX.

8.7 | Online visualization tools

Much work has been done in making data more accessible and transparent.²⁶¹ Since the establishment of data repositories such as the wwPDB⁷⁴ and the EMDb^{73,262} (for atomic models and cryo-EM maps, respectively), attention has been drawn to the complexity in interpreting this data. In the cryo-EM community it is now commonplace for authors to publish the recommended contouring level to make visual assessment of models reproducible. While this can help with interpreting map/model fit, X-ray models, and cryo-EM maps are themselves an interpretation of the experimental data they are derived from. Despite the large size of the raw data, scientists are now encouraged to archive this information as well in the IRRMC²⁶³ and EMPIAR²⁶⁴ (X-ray diffraction and cryoEM image stacks, respectively). In order to make the assessment of newly derived models and maps easier, Molstack²⁶⁵ offers a web based tool with a split screen Coot-like interface, allowing two models and maps to be easily visually compared. Indeed, web based interactive visualization is becoming more mainstream. Since the development of key technologies such as WebGL,²⁶⁶ numerous libraries and platforms for creating 3D graphics have appeared. Structural data banks have been swift to adopt these new tools such as Mol*²⁶⁷ and NGL^{268,269} for the presentation of models, replacing the static images which struggled to convey the 3D nature of molecules. These general purpose visualization tools tend to focus on the display of the structure in an unbiased manner, however, scientific communication in journals often requires more illustrative techniques to convey information. Scientists use visualization software, color, various atomic representations and potentially arrows and labels to create illustrations, which draw attention to features of interest while at the same time maintaining relevant context. However, these carefully crafted images are often still difficult to interpret as they poorly convey depth. Supplementary videos may help with interpretation, but frequently, exploring molecules from different angles is required to grasp the structure. 3dRS²⁷⁰ is an example of a web based tool which addresses this requirement. Besides being able to share and preserve a particular representation of a molecular scene, it also allows users to create their own graphical representations using a forking mechanism similar to that found in version control software like Git.²⁷¹ With protein dynamics intrinsic to function as well the interpretation of density, 3dRS has the capacity to animate trajectories. Perhaps, one of the key advantages of the web platform is its accessibility, with no requirement for the user to install additional software. EzMol²⁷² capitalizes on this simplicity, offering a tool which aims to be beginner friendly through the use of a “wizard” for step-by-step configuration of a visual scene. Adoption of tools such as 3dRS by e-journals will likely save scientists time in developing visualizations as well as improving their audience's understanding of their findings.

While the web offers significant promise for communicating scientific findings, discussing structures in a meeting or with distant collaborators can be difficult even when in the same room. Participants may struggle to communicate what feature they are looking at, particularly when faced with an unfamiliar protein. ChimeraX offers functionality to conduct meetings in a virtual reality setting. Pointing devices allow attendees to address features of a structure and for other participants to see what they are looking at, improving discourse. Virtual reality glasses also offer stereoscopic vision, which improves depth perception. Although virtual reality assists in 3D perception of objects and offers an exciting new way of exploring molecules, 3D printing²⁷³ is allowing rapid construction of physical models.

8.8 | Machine learning and structure prediction

One of the compelling aspects of machine learning methods is their ability to “learn” tasks which are difficult to describe in an algorithmic fashion. Feature detection, which is described earlier, is an important part of fitting structures both by eye and with the aid of software. Recently, a combination of Haruspex feature detection and SMOG scores were used to identify problematic structures in a large collection of published models.²⁷⁴ This is an excellent example of how scoring and machine learning can be used to lighten the cognitive load on users by drawing attention to regions of high or low confidence.

A new notion of confidence in the form of pLDDT scores from AlphaFold2^{135,275} and related machine learning methods¹³⁶ are proving beneficial to interactive fitting applications. In addition to the groundbreaking accuracy of the predicted structures themselves, the confidence scores offer another technique for gaining insight into regions of rigidity such as domains and the flexibility between them. Recent versions of ISOLDE for example can treat this confidence as a form of restraint, which can be applied to fitting structures.

9 | OUTLOOK AND CONCLUSIONS

Cryo-EM has become the structural biology technique of choice for molecules and complexes larger than ~150 kDa. Both in vitro purified samples, and molecules in their native cellular environment can now be visualized at resolutions better than 3.5 Å.^{2,8} Indeed, for a subset of sufficiently stable samples, true atomistic resolution is now achievable by cryo-EM.^{41,42} However, for a variety of reasons, including difficulties such as low number of particles, preferred orientation, and inherent flexibility of structures, many samples are still confined to the medium resolution range or even lower (Figure 1). Indeed, the resolution is often heterogeneous within single maps, including those solved at a nominal high resolution. As a result of this broad resolution range, the landscape of model building, refinement and validation tools in cryo-EM is similarly wide-ranging, and is continuously evolving (Figure 12).

Perhaps the most notable recent development in the structural biology field is the highly accurate machine learning based protein structure prediction methods.^{135,136,276–278} These methods are already having an impact in cryo-EM model building: multiple model building softwares have specific functionalities to integrate predicted models into their model refinement protocols.^{213,214,257} Such models have already been used in assembly fitting and/or refinement into cryo-EM maps, when the structures of some of the individual subunits is unknown.^{134,138,279} In addition to these recent advances, we expect further progress in structure prediction of protein-RNA/-DNA complexes, protein-ligand complexes, as well as in integrating post-translational modifications.

The development of machine-learning methods is also accelerating in other applications relevant to cryo-EM, both in the image processing as well as model building and validation.^{34,43,64,105,243} For example, machine learning-based image classification methods can help to identify more distinct states within cryo-EM datasets. Mapping this conformational heterogeneity of an imaged molecule is a unique and powerful capability of cryo-EM and could serve as a basis to better understand cellular mechanisms, particularly by being complemented with MD simulations. MD methods are important in “connecting” between states and simulating short-lived conformations and/or interactions with lipid membranes.²⁸⁰ However, such simulations are generally restricted by large computational requirements, particularly when working with large complexes such as the ones typically solved by cryo-EM. There has been recent work to address this problem, including with coarse-grained simulations, which are useful for modeling of sub-cellular systems with large complexity,^{280–282} or by using novel computational representations of protein complexes.^{59,60} Further work will also likely also focus on how local resolution and model validation could be used to explore the inherent flexibility of protein assemblies observed by cryo-EM.

Another direction is the integration of cryo-EM modeling approaches with small-molecule structure based drug design (SBDD). A recent paper has demonstrated successful fragment screening using high resolution (<3 Å) cryo-EM,²⁸³ a technique that until now has almost exclusively been applied to X-ray crystallography. In the study it was shown that not only was cryo-EM able to unambiguously resolve the positions of small fragment molecules and side-chain conformations changes upon binding, but also ligand density corresponding to fragments could be unambiguously assigned when protein targets were incubated with multiple fragment molecules. Crucially, using cryo-EM for SBDD increases the number of potential targets, as many disease related proteins, including many membrane proteins, are not amenable to X-ray crystallography. However, as mentioned above, most maps are still at a resolution worse than 3 Å, including many potential new drug targets. We expect that in the future more methods integrating underutilized

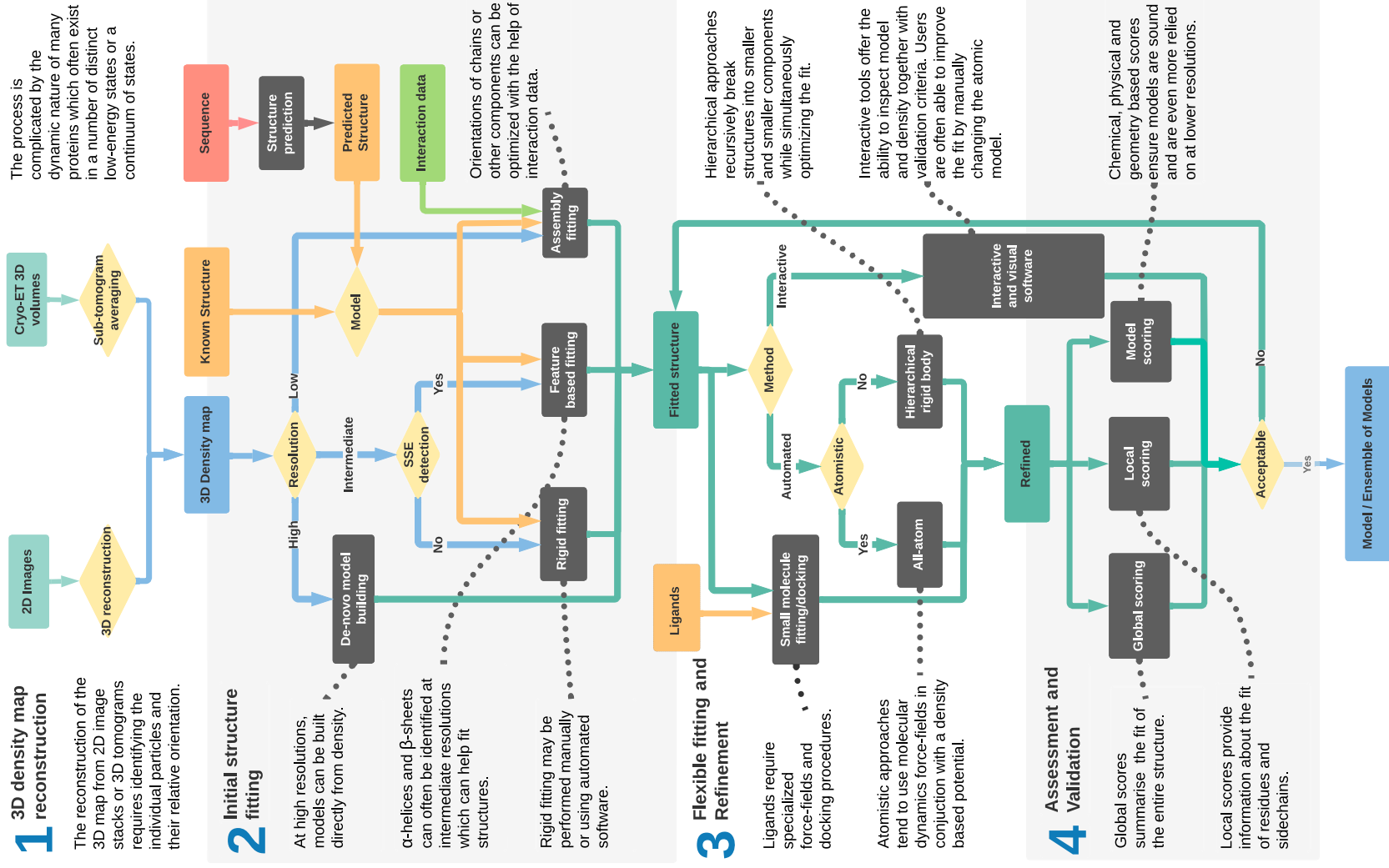


FIGURE 12 Flowchart depicting the cryo-EM model building process.

cryo-EM maps (>3.5 – 4.5 Å resolution) with small molecule docking (and machine learning approaches) will be developed to take advantage of these potential new targets.

Finally, we expect further consolidation of new software and computational developments into large software suites.^{12–15} These are gaining popularity as accessibility of scientific software can be a problem: new methods are often developed as command-line (CLI) applications, which can be useful for experts but intimidating to novices, who may wish to use these techniques on their experimental data. Graphical User Interfaces (GUIs) offer a more user-friendly approach but typically take more time to develop and maintain. For example, the CCP-EM software suite offers an easily installable graphical frontend to useful methods for reconstruction, fitting, refinement and validation and also includes visualization tools.¹² One of the advantages of the CLI approach is the possibility to build new tools or protocols by composing individual programs into pipelines. CCP-EM and Phenix¹³ along with tools such as Cow-EM and Scipion¹⁵ attempt to give some of this power to less technical users, by providing mechanisms for designing pipelines using either a “visual programming” model or a GUI for defining computational workflows.

AUTHOR CONTRIBUTIONS

Joseph George Beton: Writing – original draft (lead); writing – review and editing (lead). **Tristan Cragnolini:** Writing – original draft (equal); writing – review and editing (equal). **Manaz Kaleel:** Writing – original draft (equal); writing – review and editing (equal). **Thomas Mulvaney:** Writing – original draft (equal); writing – review and editing (equal). **Aaron Sweeney:** Writing – original draft (equal); writing – review and editing (equal). **Maya Topf:** Writing – original draft (lead); writing – review and editing (lead).

ACKNOWLEDGMENT

Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

This work was supported by Wellcome Trust grant 209250/Z/17/Z and cooperation of Leibniz Institute of Virology (LIV) and Universitätsklinikum Hamburg Eppendorf (UKE) (as part of Leibniz ScienceCampus InterACT, funded by the BWFG Hamburg and the Leibniz Association).

CONFLICT OF INTEREST


The authors have declared no conflicts of interest for this article.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Joseph George Beton  <https://orcid.org/0000-0001-7499-3867>

Tristan Cragnolini  <https://orcid.org/0000-0001-6917-7056>

Manaz Kaleel  <https://orcid.org/0000-0001-9043-0802>

Thomas Mulvaney  <https://orcid.org/0000-0002-4373-6160>

Maya Topf  <https://orcid.org/0000-0002-8185-1215>

RELATED WIREs ARTICLES

[Artificial intelligence advances for de novo molecular structure modeling in cryo-electron microscopy](#)

REFERENCES

1. Kühlbrandt W. The resolution revolution. *Science*. 2014;343(6178):1443–4.
2. Bai XC, Fernandez IS, McMullan G, Scheres SHW. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *Elife*. 2013 Feb 19;2:e00461.
3. Patwardhan A. Trends in the Electron Microscopy Data Bank (EMDB). *Acta Crystallogr D Struct Biol*. 2017 Jun 1;73(Pt 6):503–8.
4. Wan W, Briggs JAG. Cryo-electron tomography and subtomogram averaging. *Methods Enzymol*. 2016 Feb 19;579:329–67.
5. Kim SJ, Fernandez-Martinez J, Nudelman I, Shi Y, Zhang W, Raveh B, et al. Integrative structure and functional anatomy of a nuclear pore complex. *Nature*. 2018 Mar 22;555(7697):475–82.

6. Bykov YS, Schaffer M, Dodonova SO, Albert S, Plitzko JM, Baumeister W, et al. The structure of the COPI coat determined within the cell. *Elife* [Internet]. 2017 Nov 17;6:e32493. <https://doi.org/10.7554/eLife.32493>
7. Förster F. Subtomogram analysis: the sum of a tomogram's particles reveals molecular structure in situ. *J Struct Biol*. X. 2022;1(6):100063.
8. Tegunov D, Xue L, Dienemann C, Cramer P, Mahamid J. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nat Methods*. 2021 Feb;18(2):186–93.
9. Baxter WT, Grassucci RA, Gao H, Frank J. Determination of signal-to-noise ratios and spectral SNRs in cryo-EM low-dose imaging of molecules. *J Struct Biol*. 2009 Oct;166(2):126–32.
10. Malhotra S, Träger S, Dal Peraro M, Topf M. Modelling structures in cryo-EM maps. *Curr Opin Struct Biol*. 2019 Oct;58:105–14.
11. Scheres SHW. Chapter 6—Processing of structurally heterogeneous Cryo-EM data in RELION. In: Crowther RA, editor. *Methods in enzymology*. Academic Press; 2016. p. 125–57.
12. Joseph AP, Olek M, Malhotra S, Zhang P, Cowtan K, Burnley T, et al. Atomic model validation using the CCP-EM software suite. *Acta Crystallogr D Struct Biol*. 2022;78(Pt 2):152–61.
13. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010 Feb;66(Pt 2):213–21.
14. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol*. 2011;487:545–74.
15. de la Rosa-Trevín JM, Quintana A, Del Cano L, Zaldívar A, Foche I, Gutiérrez J, et al. Scipion: a software framework toward integration, reproducibility and validation in 3D electron microscopy. *J Struct Biol*. 2016 Jul;195(1):93–9.
16. Sigworth FJ. Principles of cryo-EM single-particle image processing. *Microscopy*. 2016;65(1):57–67.
17. Harauz G, van Heel M. Exact filters for general geometry three dimensional reconstruction. *Optik*. 1986;73(4):146–56.
18. Grigorieff N. Resolution measurement in structures derived from single particles. *Acta Crystallogr D Biol Crystallogr*. 2000 Oct;56(Pt 10):1270–7.
19. Rosenthal PB, Henderson R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol*. 2003 Oct 31;333(4):721–45.
20. van Heel M, Schatz M. Fourier shell correlation threshold criteria. *J Struct Biol*. 2005 Sep 1;151(3):250–62.
21. van Heel M, Schatz M. Reassessing the revolution's resolutions. *bioRxiv* [Internet]. 2017 [cited 2022 Mar 8]. p. 224402. <https://doi.org/10.1101/224402v1>.
22. Beckers M, Sachse C. Permutation testing of Fourier shell correlation for resolution estimation of cryo-EM maps. *J Struct Biol*. 2020 Oct 1;212(1):107579.
23. Rohou A. Chapter 4.7: Bfactors and map sharpening. In: Glaeser RM, Nogales E, Chiu W, editors. *Single-particle cryo-EM of biological macromolecules*. Biophysical Society; 2021. p. 4–59–67.
24. Hyrc CF, Baker ML. Chapter 6.3: building atomistic models in cryo EM density maps. In: Glaeser RM, Nogales E, Chiu W, editors. *Single-particle cryo-EM of biological macromolecules*. Biophysical Society; 2021. p. 6–6–6–25.
25. Vilas JL, Heymann JB, Tagare HD, Ramirez-Aportela E, Carazo JM, Sorzano C. Local resolution estimates of cryoEM reconstructions. *Curr Opin Struct Biol*. 2020 Oct;64:74–8.
26. Cardone G, Heymann JB, Steven AC. One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *J Struct Biol*. 2013 Nov 1;184(2):226–36.
27. Kucukelbir A, Sigworth FJ, Tagare HD. Quantifying the local resolution of cryo-EM density maps. *Nat Methods*. 2014 Jan;11(1):63–5.
28. Vilas JL, Gómez-Blanco J, Conesa P, Melero R, Miguel de la Rosa-Trevín J, Otón J, et al. MonoRes: automatic and accurate estimation of local resolution for electron microscopy maps. *Structure*. 2018 Feb 6;26(2):337–44.e4.
29. Felsberg M, Sommer G. The monogenic signal. *IEEE Trans Signal Process*. 2001 Dec;49(12):3136–44.
30. Ramirez-Aportela E, Mota J, Conesa P, Carazo JM, Sorzano COS. DeepRes: a new deep-learning- and aspect-based local resolution method for electron-microscopy maps. *IUCrJ*. 2019 Nov 1;6(Pt 6):1054–63.
31. Bharadwaj A, Jakobi A. Electron scattering properties of biological macromolecules and their use for cryo-EM map sharpening. *Faraday Discuss* [Internet]. 2022; May 12 [cited 2022 Jun 29]; Available from: <https://pubs.rsc.org/en/Content/ArticleLanding/2022/FD/D2FD00078D>
32. Jakobi AJ, Wilmanns M, Sachse C. Model-based local density sharpening of cryo-EM maps. *Elife* [Internet]. 2017 Oct 23;6:e27131. <https://doi.org/10.7554/eLife.27131>
33. Ramirez-Aportela E, Vilas JL, Glukhova A, Melero R, Conesa P, Martínez M, et al. Automatic local resolution-based sharpening of cryo-EM maps. *Bioinformatics*. 2020 Feb 1;36(3):765–72.
34. Sanchez-Garcia R, Gomez-Blanco J, Cuervo A, Carazo JM, Sorzano COS, Vargas J. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun Biol*. 2021 Jul 15;4(1):874.
35. Wang S, Jomaa A, Jaskolowski M, Yang CI, Ban N, Shan SO. The molecular mechanism of cotranslational membrane protein recognition and targeting by SecA. *Nat Struct Mol Biol*. 2019 Oct;26(10):919–29.
36. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods*. 2017 Feb 6;14(3):290–6.
37. Clare DK, Vasishtan D, Stagg S, Quispe J, Farr GW, Topf M, et al. ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. *Cell*. 2012 Mar 30;149(1):113–23.

38. Dashti A, Schwander P, Langlois R, Fung R, Li W, Hosseinizadeh A, et al. Trajectories of the ribosome as a Brownian nanomachine. *Proc Natl Acad Sci U S A*. 2014 Dec 9;111(49):17492–7.
39. Russo CJ, Passmore LA. Electron microscopy: Ultrastable gold substrates for electron cryomicroscopy. *Science*. 2014 Dec 12;346(6215):1377–80.
40. Naydenova K, Jia P, Russo CJ. Cryo-EM with sub-1 Å specimen movement. *Science*. 2020;370(6513):223–6.
41. Nakane T, Kotecha A, Sente A, McMullan G, Masiulis S, Brown PMGE, et al. Single-particle cryo-EM at atomic resolution. *Nature*. 2020 Oct 21;587(7832):152–6.
42. Yip KM, Fischer N, Paknia E, Chari A, Stark H. Atomic-resolution protein structure determination by cryo-EM. *Nature*. 2020 Oct 21;587(7832):157–61.
43. Zhong ED, Beppler T, Berger B, Davis JH. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat Methods*. 2021 Feb 4;18(2):176–85.
44. Lyumkis D, Brilot AF, Theobald DL, Grigorieff N. Likelihood-based classification of cryo-EM images using FREALIGN. *J Struct Biol*. 2013 Sep 1;183(3):377–88.
45. Grant T, Rohou A, Grigorieff N. cisTEM, user-friendly software for single-particle image processing. *Elife*. 2018 Mar 7;7:e35383.
46. Moriya T, Saur M, Stabrin M, Merino F, Voicu H, Huang Z, et al. High-resolution single particle analysis from electron cryo-microscopy images using SPHIRE. *J Vis Exp [Internet]*. 2017 May 16;(123):55448. <https://doi.org/10.3791/55448>
47. Kimanius D, Dong L, Sharov G, Nakane T, Scheres SHW. New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem J*. 2021 Dec 22;478(24):4169–85.
48. Scheres SHW. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol*. 2012 Dec;180(3):519–30.
49. Orlova EV, Saibil HR. Chapter twelve - methods for three-dimensional reconstruction of heterogeneous assemblies. In: Jensen GJ, editor. *Methods in enzymology*. Academic Press; 2010. p. 321–41.
50. van Heel M, Portugal R, Rohou A, Linnemayr C, Bebeacua C, Schmidt R, et al. Four-dimensional cryo electron microscopy at quasi atomic resolution: “IMAGIC 4D”. *Int Tables Crystallogr*. 2012;F:624–8.
51. Afanasyev P, Seer-Linnemayr C, Ravelli RBG, Matadeen R, De Carlo S, Alewijnse B, et al. Single-particle cryo-EM using alignment by classification (ABC): the structure of *Lumbricus terrestris* haemoglobin. *IUCrJ*. 2017 Sep 1;4(Pt 5):678–94.
52. Heymann JB. Guidelines for using Bsoft for high resolution reconstruction and validation of biomolecular structures from electron micrographs. *Protein Sci*. 2018 Jan;27(1):159–71.
53. Zhao J, Benlekhir S, Rubinstein JL. Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. *Nature*. 2015 May 14;521(7551):241–5.
54. Wilkinson ME, Fica SM, Galej WP, Nagai K. Structural basis for conformational equilibrium of the catalytic spliceosome. *Mol Cell*. 2021 Apr 1;81(7):1439–1452.e9.
55. Fischer N, Konevega AL, Wintermeyer W, Rodnina MV, Stark H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature*. 2010 Jul 15;466(7304):329–33.
56. Haselbach D, Schrader J, Lambrecht F, Henneberg F, Chari A, Stark H. Long-range allosteric regulation of the human 26 S proteasome by 20 S proteasome-targeting cancer drugs. *Nat Commun*. 2017 May 25;8(1):1–8.
57. Haselbach D, Komarov I, Agafonov DE, Hartmuth K, Graf B, Dybkov O, et al. Structure and conformational dynamics of the human spliceosomal bact complex. *Cell*. 2018 Jan 25;172(3):454–64.e11.
58. Oide M, Kato T, Oroguchi T, Nakasako M. Energy landscape of domain motion in glutamate dehydrogenase deduced from cryo-electron microscopy. *FEBS J*. 2020 Aug;287(16):3472–93.
59. Solernou A, Hanson BS, Richardson RA, Welch R, Read DJ, Harlen OG, et al. Fluctuating finite element analysis (FFEA): a continuum mechanics software tool for mesoscale simulation of biomolecules. *PLoS Comput Biol*. 2018 Mar 14;3:e1005897.
60. Richardson RA, Hanson BS, Read DJ, Harlen OG, Harris SA. Exploring the dynamics of flagellar dynein within the axoneme with fluctuating finite element analysis. *Q Rev Biophys*. 2020 Aug 10;10:e9.
61. Jin Q, Sorzano COS, de la Rosa-Trevín JM, Bilbao-Castro JR, Núñez-Ramírez R, Llorca O, et al. Iterative elastic 3D-to-2D alignment method using normal modes for studying structural dynamics of large macromolecular complexes. *Structure*. 2014 Mar 4;22(3):496–506.
62. Punjani A, Fleet DJ. 3D variability analysis: resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J Struct Biol*. 2021 Jun 1;213(2):107702.
63. Nakane T, Kimanius D, Lindahl E, Scheres SH. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife [Internet]*. 2018 Jun 1;7:e36861. <https://doi.org/10.7554/eLife.36861>
64. Chen M, Ludtke SJ. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. *Nat Methods*. 2021 Aug;18(8):930–6.
65. Harastani M, Eltsov M, Leforestier A, Jonic S. HEMNMA-3D: cryo electron tomography method based on normal mode analysis to study continuous conformational variability of macromolecular complexes. *Front Mol Biosci*. 2021 May 19;8:663121.
66. Harastani M, Vuillemot R, Hamitouche I, Barati Moghadam N, Jonic S. ContinuousFlex: software package for analyzing continuous conformational variability of macromolecules in cryo electron microscopy and tomography data. *J Struct Biol*. 2022 Oct;13:107906.
67. Roweis S. EM algorithms for PCA and SPCA. *Proc NIPS*. 1998;10:626–32.
68. Tipping ME, Bishop CM. Probabilistic principal component analysis. *J R Stat Soc Series B Stat Methodol*. 1999 Aug;61(3):611–22.

69. Gillies JP, Reimer JM, Karasmanis EP, Lahiri I, Htet ZM, Leschziner AE, et al. Structural basis for cytoplasmic dynein-1 regulation by Lis1. *Elife* [Internet]. 2022 Jan 7;11:e71229. <https://doi.org/10.7554/eLife.71229>
70. Lindert S, Staritzbichler R, Wötzel N, Karakaş M, Stewart PL, Meiler J. EM-fold: De novo folding of alpha-helical proteins guided by intermediate-resolution electron microscopy density maps. *Structure*. 2009 Jul 15;17(7):990–1003.
71. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017 Dec;39(12):2481–95.
72. Matsumoto S, Ishida S, Araki M, Kato T, Terayama K, Okuno Y. Extraction of protein dynamics information from cryo-EM maps using deep learning. *Nat Mach Intell*. 2021 Feb 4;3(2):153–60.
73. Patwardhan A, Lawson CL. Databases and archiving for cryoEM. *Methods Enzymol*. 2016;579:393–412.
74. Berman H, Henrick K, Nakamura H. Announcing the worldwide protein data Bank. *Nat Struct Biol*. 2003 Dec;10(12):980.
75. Jiang W, Baker ML, Ludtke SJ, Chiu W. Bridging the information gap: computational tools for intermediate resolution structure interpretation. *J Mol Biol*. 2001 May 18;308(5):1033–44.
76. Kong Y, Ma J. A structural-informatics approach for mining beta-sheets: locating sheets in intermediate-resolution density maps. *J Mol Biol*. 2003 Sep 12;332(2):399–413.
77. Baker ML, Ju T, Chiu W. Identification of secondary structure elements in intermediate-resolution density maps. *Structure*. 2007 Jan; 15(1):7–19.
78. Si D, Ji S, Nasr KA, He J. A machine learning approach for the identification of protein secondary structure elements from electron cryo-microscopy density maps. *Biopolymers*. 2012 Sep;97(9):698–708.
79. Mostosi P, Schindelin H, Kollmannsberger P, Thorn A. Haruspex: a neural network for the automatic identification of oligonucleotides and protein secondary structure in cryo-electron microscopy maps. *Angew Chem Weinheim Bergstr Ger*. 2020 Aug 24;132(35): 14898–905.
80. Lasker K, Dror O, Shatsky M, Nussinov R, Wolfson HJ. EMatch: discovery of high resolution structural homologues of protein domains in intermediate resolution cryo-EM maps. *IEEE/ACM Trans Comput Biol Bioinform*. 2007 Jan;4(1):28–39.
81. Velázquez-Muriel JA, Sorzano COS, Scheres SHW, Carazo JM. SPI-EM: towards a tool for predicting CATH superfamilies in 3D-EM maps. *J Mol Biol*. 2005 Jan 28;345(4):759–71.
82. Meiler J, Baker D. Coupled prediction of protein secondary and tertiary structure. *Proc Natl Acad Sci U S A*. 2003 Oct 14;100(21): 12105–10.
83. Meiler J, Müller M, Zeidler A, Schmäschke F. Generation and evaluation of dimension-reduced amino acid parameter representations by artificial neural networks. *Mol Model Ann*. 2001 Sep 22;7(9):360–9.
84. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*. 1999 Sep 17;292(2):195–202.
85. Chandonia JM, Karplus M. New methods for accurate prediction of protein secondary structure. *Proteins*. 1999 May 15;35(3):293–306.
86. Karplus K, Sjölander K, Barrett C, Cline M, Haussler D, Hughey R, et al. Predicting protein structure using hidden Markov models. *Proteins*. 1997;Suppl 1:134–9.
87. Rohl CA, Strauss CEM, Misura KMS, Baker D. Protein structure prediction using Rosetta. *Methods Enzymol*. 2004;383:66–93.
88. Baker MR, Rees I, Ludtke SJ, Chiu W, Baker ML. Constructing and validating initial C α models from subnanometer resolution density maps with pathwalking. *Structure*. 2012 Mar 7;20(3):450–63.
89. Rusu M, Wriggers W. Evolutionary bidirectional expansion for the tracing of alpha helices in cryo-electron microscopy reconstructions. *J Struct Biol*. 2012 Feb;177(2):410–9.
90. Si D, He J. Beta-sheet detection and representation from medium resolution Cryo-EM density maps. In: *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*; New York, NY: Association for Computing Machinery; 2013. p. 764–70. (BCB'13).
91. Kong Y, Zhang X, Baker TS, Ma J. A structural-informatics approach for tracing beta-sheets: building pseudo-C(alpha) traces for beta-strands in intermediate-resolution density maps. *J Mol Biol*. 2004 May 21;339(1):117–30.
92. Yu Z, Bajaj C. Computational approaches for automatic structural analysis of large biomolecular complexes. *IEEE/ACM Trans Comput Biol Bioinform*. 2008 Oct;5(4):568–82.
93. Si D, He J. Tracing beta strands using StrandTwister from cryo-EM density maps at medium resolutions. *Structure*. 2014 Nov 4;22(11): 1665–76.
94. Si D. Automatic detection of beta-barrel from medium resolution cryo-EM density maps. *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. New York, NY, USA: Association for Computing Machinery; 2016. p. 156–64. (BCB'16).
95. Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM*. 1981 Jun 1;24(6):381–95.
96. Torr PHS, Zisserman A. MLESAC: a new robust estimator with application to estimating image geometry. *Comput Vis Image Underst*. 2000 Apr 1;78(1):138–56.
97. Spillers C, Wriggers W, He J. Detection of protein secondary structure patterns from 3D Cryo-TEM maps at medium resolution—combining the best of SSETracer and VolTrac. *Microsc Microanal*. 2017 Jul;23(S1):242–3.
98. Haslam D, Sazzed S, Wriggers W, Kovacs J, Song J, Auer M, et al. A pattern recognition tool for medium-resolution Cryo-EM density maps and low-resolution Cryo-ET density maps. *Bioinformatics research and applications*. Cham: Springer International Publishing; 2018. p. 233–8.

99. Wieland M, Pittore M. Performance evaluation of machine learning algorithms for urban pattern recognition from multi-spectral satellite images. *Remote Sens (Basel)*. 2014 Mar 31;6(4):2912–39.
100. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. *Radiographics*. 2017 Mar;37(2):505–15.
101. Li R, Si D, Zeng T, Ji S, He J. Deep convolutional neural networks for detecting secondary structures in protein density maps from cryo-electron microscopy. In: 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2016. p. 41–6.
102. Si D, Moritz SA, Pfab J, Hou J, Cao R, Wang L, et al. Deep learning to predict protein backbone structure from high-resolution cryo-EM density maps. *Sci Rep*. 2020 Mar 9;10(1):4282.
103. Wang X, Alnabati E, Aderinwale TW, Maddhuri Venkata Subramaniya SR, Terashi G, Kihara D. Detecting protein and DNA/RNA structures in cryo-EM maps of intermediate resolution using deep learning. *Nat Commun*. 2021 Apr 16;12(1):2302.
104. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. openaccess.thecvf.com; p. 770–8.
105. Pfab J, Phan NM, Si D. DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes. *Proc Natl Acad Sci U S A* [Internet]. 2021 Jan 12;118(2):e2017525118. <https://doi.org/10.1073/pnas.2017525118>
106. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention – MICCAI 2015*. Cham: Springer International Publishing; 2015. p. 234–41.
107. Pan M, Zheng Q, Wang T, Liang L, Mao J, Zuo C, et al. Structural insights into Ubr1-mediated N-degron polyubiquitination. *Nature*. 2021 Dec;600(7888):334–8.
108. Lyumkis D. Challenges and opportunities in cryo-EM single-particle analysis. *J Biol Chem*. 2019 Mar 29;294(13):5181–97.
109. Joseph AP, Polles G, Alber F, Topf M. Integrative modelling of cellular assemblies. *Curr Opin Struct Biol*. 2017 Oct;46:102–9.
110. Burley SK, Kurisu G, Markley JL, Nakamura H, Velankar S, Berman HM, et al. PDB-dev: a prototype system for depositing integrative/hybrid structural models. *Structure*. 2017 Sep 5;25(9):1317–8.
111. Wriggers W, Milligan RA, McCammon JA. Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. *J Struct Biol*. 1999 Apr;125(2–3):185–95.
112. Birmanns S, Rusu M, Wriggers W. Using sculptor and situs for simultaneous assembly of atomic components into low-resolution shapes. *J Struct Biol*. 2011 Mar;173(3):428–35.
113. Russel D, Lasker K, Webb B, Velázquez-Muriel J, Tjioe E, Schneidman-Duhovny D, et al. Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol*. 2012 Jan;10(1):e1001244.
114. Roseman AM. Docking structures of domains into maps from cryo-electron microscopy using local correlation. *Acta Crystallogr D Biol Crystallogr*. 2000 Oct;56(Pt 10):1332–40.
115. Esquivel-Rodríguez J, Kihara D. Fitting multimeric protein complexes into electron microscopy maps using 3D Zernike descriptors. *J Phys Chem B*. 2012 Jun 14;116(23):6854–61.
116. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004 Oct;25(13):1605–12.
117. Pandurangan AP, Vasishthan D, Alber F, Topf M. γ -TEMPy: simultaneous fitting of components in 3D-EM maps of their assembly using a genetic algorithm. *Structure*. 2015 Dec 1;23(12):2365–76.
118. Topf M, Baker ML, John B, Chiu W, Sali A. Structural characterization of components of protein assemblies by comparative modeling and electron cryo-microscopy. *J Struct Biol*. 2005 Feb;149(2):191–203.
119. Neijenhuis T, Van Keulen SC, AMJJ B. Interface refinement of low- to medium-resolution Cryo-EM complexes using HADDOCK2.4. *Structure*. 2022 Apr 7;30(4):476–84.e3.
120. Garzón JI, Kovacs J, Abagyan R, Chacón P. ADP_EM: fast exhaustive multi-resolution docking for high-throughput coverage. *Bioinformatics*. 2007 Feb 15;23(4):427–33.
121. Kuzu G, Keskin O, Nussinov R, Gursoy A. PRISM-EM: template interface-based modelling of multi-protein complexes guided by cryo-electron microscopy density maps. *Acta Crystallogr D Struct Biol*. 2016 Oct 1;72(Pt 10):1137–48.
122. Kawabata T. Multiple subunit fitting into a low-resolution density map of a macromolecular complex using a gaussian mixture model. *Biophys J*. 2008 Nov 15;95(10):4643–58.
123. Han X, Terashi G, Christoffer C, Chen S, Kihara D. VESPER: global and local cryo-EM map alignment using local density vectors. *Nat Commun*. 2021 Apr 7;12(1):2090.
124. Lasker K, Topf M, Sali A, Wolfson HJ. Inferential optimization for simultaneous fitting of multiple components into a cryoEM map of their assembly. *J Mol Biol*. 2009 Apr 24;388(1):180–94.
125. Rantos V, Karius K, Kosinski J. Integrative structural modeling of macromolecular complexes using Assemblin. *Nat Protoc*. 2022 Jan; 17(1):152–76.
126. Volkman N, Hanein D. Quantitative fitting of atomic models into observed densities derived by electron microscopy. *J Struct Biol*. 1999 Apr;125(2–3):176–84.
127. Vasishthan D, Topf M. Scoring functions for cryoEM density fitting. *J Struct Biol*. 2011;174(2):333–43.
128. Zhang S, Vasishthan D, Xu M, Topf M, Alber F. A fast mathematical programming procedure for simultaneous fitting of assembly components into cryoEM density maps. *Bioinformatics*. 2010 Jun 15;26(12):i261–8.
129. van Zundert GCP, Melquiond ASJ, Bonvin AMJJ. Integrative modeling of biomolecular complexes: HADDOCKing with Cryo-electron microscopy data. *Structure*. 2015 May 5;23(5):949–60.

130. DiMaio F, Tyka MD, Baker ML, Chiu W, Baker D. Refinement of protein structures into low-resolution density maps using rosetta. *J Mol Biol*. 2009 Sep 11;392(1):181–90.
131. Degiacomi MT, Dal Peraro M. Macromolecular symmetric assembly prediction using swarm intelligence dynamic modeling. *Structure*. 2013 Jul 2;21(7):1097–106.
132. Beckham KSH, Ritter C, Chojnowski G, Ziemianowicz DS, Mullapudi E, Rettel M, et al. Structure of the mycobacterial ESX-5 type VII secretion system pore complex. *Sci Adv* [Internet]. 2021 Jun;7(26):eabg9923. <https://doi.org/10.1126/sciadv.abg9923>
133. Zimmerli CE, Allegretti M, Rantos V, Goetz SK, Obarska-Kosinska A, Zagoriy I, et al. Nuclear pores dilate and constrict in cellulose. *Science*. 2021 Dec 10;374(6573):eabd9776.
134. Mosalaganti S, Obarska-Kosinska A, Siggel M, Taniguchi R, Turoňová B, Zimmerli CE, et al. AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science*. 2022 Jun 10;376(6598):eabm9506.
135. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021 Aug;596(7873):583–9.
136. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. 2021 Aug 20;373(6557):871–6.
137. Akey CW, Singh D, Ouch C, Echeverria I, Nudelman I, Varberg JM, et al. Comprehensive structure and functional adaptations of the yeast nuclear pore complex. *Cell*. 2022 Jan 20;185(2):361–78.e25.
138. Fontana P, Dong Y, Pi X, Tong AB, Hecksel CW, Wang L, et al. Structure of cytoplasmic ring of nuclear pore complex by integrative cryo-EM and AlphaFold. *Science*. 2022 Jun 10;376(6598):eabm9326.
139. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr*. 2011 Apr;67(Pt 4):355–67.
140. Chen M, Baldwin PR, Ludtke SJ, Baker ML. De novo modeling in cryo-EM density maps with pathwalking. *J Struct Biol*. 2016 Dec;196(3):289–98.
141. Terashi G, Kihara D. De novo main-chain modeling for EM maps using MAINMAST. *Nat Commun*. 2018 Apr 24;9(1):1618.
142. Baker ML, Abeysinghe SS, Schuh S, Coleman RA, Abrams A, Marsh MP, et al. Modeling protein structure at near atomic resolutions with Gorgon. *J Struct Biol*. 2011 May;174(2):360–73.
143. Casañal A, Lohkamp B, Emsley P. Current developments in coot for macromolecular model building of electron Cryo-microscopy and crystallographic data. *Protein Sci*. 2020 Apr;29(4):1069–78.
144. Cowtan K. The buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr*. 2006 Sep;62(Pt 9):1002–11.
145. Chojnowski G, Pereira J, Lamzin VS. Sequence assignment for low-resolution modelling of protein crystal structures. *Acta Crystallogr D Struct Biol*. 2019 Aug 1;75(Pt 8):753–63.
146. Cowtan K. Completion of autobuilt protein models using a database of protein fragments. *Acta Crystallogr D Biol Crystallogr*. 2012 Apr;68(Pt 4):328–35.
147. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. 2010 Apr;66(Pt 4):486–501.
148. Wang RYR, Song Y, Barad BA, Cheng Y, Fraser JS, DiMaio F. Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *Elife*. 2016 Sep 26;5:e17219.
149. Tama F, Miyashita O, Brooks CL 3rd. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J Struct Biol*. 2004 Sep;147(3):315–26.
150. Singharoy A, Teo I, McGreevy R, Stone JE, Zhao J, Schulten K. Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron microscopy maps. *Elife*. 2016 Jul 7;5:e16105.
151. Fabiola F, Chapman MS. Fitting of high-resolution structures into electron microscopy reconstruction images. *Structure*. 2005 Mar;13(3):389–400.
152. Topf M, Lasker K, Webb B, Wolfson H, Chiu W, Sali A. Protein structure fitting and refinement guided by cryo-EM density. *Structure*. 2008 Feb;16(2):295–307.
153. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure*. 2008 May;16(5):673–83.
154. Bonomi M, Pellarin R, Vendruscolo M. Simultaneous determination of protein structure and dynamics using cryo-electron microscopy. *Biophys J*. 2018 Apr 10;114(7):1604–13.
155. Afonine PV, Poon BK, Read RJ, Sobolev OV, Terwilliger TC, Urzhumtsev A, et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol*. 2018 Jun 1;74(Pt 6):531–44.
156. Brown A, Long F, Nicholls RA, Toots J, Emsley P, Murshudov G. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr D Biol Crystallogr*. 2015 Jan 1;71(Pt 1):136–53.
157. Igaev M, Kutzner C, Bock LV, Vaiana AC, Grubmüller H. Automated cryo-EM structure refinement using correlation-driven molecular dynamics. *Elife*. 2019 Mar 4;8:e43542.
158. Wang Z, Schröder GF. Real-space refinement with DireX: from global fitting to side-chain improvements. *Biopolymers*. 2012 Sep;97(9):687–97.
159. López-Blanco JR, Chacón P. iMODFIT: efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *J Struct Biol*. 2013 Nov;184(2):261–70.
160. Costa MGS, Fagnen C, Vénien-Bryan C, Perahia D. A new strategy for atomic flexible fitting in cryo-EM maps by molecular dynamics with excited normal modes (MDenM-EMfit). *J Chem Inf Model*. 2020 May 26;60(5):2419–23.

161. DiMaio F, Song Y, Li X, Brunner MJ, Xu C, Conticello V, et al. Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat Methods*. 2015 Apr;12(4):361–5.
162. Cragnolini T, Beton J, Topf M. Cryo-EM structure and B-factor refinement with ensemble representation. *bioRxiv* [Internet]. 2022 [cited 2022 Jun 14]. p. 2022.06.08.495259. <https://doi.org/10.1101/2022.06.08.495259v1>.
163. Pandurangan AP, Topf M. Finding rigid bodies in protein structures: application to flexible fitting into cryoEM maps. *J Struct Biol*. 2012 Feb;177(2):520–31.
164. Cook AD, Roberts AJ, Atherton J, Tewari R, Topf M, Moores CA. Cryo-EM structure of a microtubule-bound parasite kinesin motor and implications for its mechanism and inhibition. *J Biol Chem*. 2021 Nov;297(5):101063.
165. Atherton J, Farabella I, Yu IM, Rosenfeld SS, Houdusse A, Topf M, et al. Conserved mechanisms of microtubule-stimulated ADP release, ATP binding, and force generation in transport kinesins. *Elife*. 2014 Sep 10;3:e03680.
166. Locke J, Joseph AP, Peña A, Möckel MM, Mayer TU, Topf M, et al. Structural basis of human kinesin-8 function and inhibition. *Proc Natl Acad Sci U S A*. 2017 Nov 7;114(45):E9539–48.
167. Joseph AP, Malhotra S, Burnley T, Wood C, Clare DK, Winn M, et al. Refinement of atomic models in high resolution EM reconstructions using flex-EM and local assessment. *Methods*. 2016 May 1;100:42–9.
168. Yu J, Raia P, Ghent CM, Raisch T, Sadian Y, Cavadini S, et al. Structural basis of human separase regulation by securin and CDK1-cyclin B1. *Nature*. 2021 Aug;596(7870):138–42.
169. Burnley T, Palmer CM, Winn M. Recent developments in the CCP-EM software suite. *Acta Crystallogr D Struct Biol*. 2017 Jun 1;73(Pt 6):469–77.
170. Chacón P, Tama F, Wriggers W. Mega-Dalton biomolecular motion captured from electron microscopy reconstructions. *J Mol Biol*. 2003 Feb 14;326(2):485–92.
171. Vuilleumot R, Miyashita O, Tama F, Rouiller I, Jonic S. NMMD: efficient Cryo-EM flexible fitting based on simultaneous Normal mode and molecular dynamics atomic displacements. *J Mol Biol*. 2022 Feb 9;434(7):167483.
172. Kawabata T. Gaussian-input Gaussian mixture model for representing density maps and atomic models. *J Struct Biol*. 2018 Jul;203(1):1–16.
173. Ahmed A, Whitford PC, Sanbonmatsu KY, Tama F. Consensus among flexible fitting approaches improves the interpretation of cryo-EM data. *J Struct Biol*. 2012 Feb;177(2):561–70.
174. Farabella I, Vasishthan D, Joseph AP, Pandurangan AP, Sahota H, Topf M. TEMPy: a python library for assessment of three-dimensional electron microscopy density fits. *J Appl Cryst*. 2015 Aug 1;48(4):1314–23.
175. Pandurangan AP, Shakeel S, Butcher SJ, Topf M. Combined approaches to flexible fitting and assessment in virus capsids undergoing conformational change. *J Struct Biol*. 2014 Mar;185(3):427–39.
176. Bonomi M, Vendruscolo M. Determination of protein structural ensembles using cryo-electron microscopy. *Curr Opin Struct Biol*. 2019 Jun;56:37–45.
177. Sachse C, Chen JZ, Coureux PD, Stroupe ME, Fändrich M, Grigorieff N. High-resolution electron microscopy of helical specimens: a fresh look at tobacco mosaic virus. *J Mol Biol*. 2007 Aug 17;371(3):812–35.
178. Lukoyanova N, Kondos SC, Farabella I, Law RHP, Reboul CF, Caradoc-Davies TT, et al. Conformational changes during pore formation by the perforin-related protein pleurotolysin. *PLoS Biol*. 2015 Feb;13(2):e1002049.
179. Baker ML, Jiang W, Wedemeyer WJ, Rixon FJ, Baker D, Chiu W. Ab initio modeling of the herpesvirus VP26 core domain assessed by CryoEM density. *PLoS Comput Biol*. 2006 Oct 27;2(10):e146.
180. Wehmer M, Rudack T, Beck F, Aufderheide A, Pfeifer G, Plitzko JM, et al. Structural insights into the functional cycle of the ATPase module of the 26 S proteasome. *Proc Natl Acad Sci U S A*. 2017 Feb 7;114(6):1305–10.
181. Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* [Internet]. 2018 Nov 9;7:e42166. <https://doi.org/10.7554/eLife.42166>
182. Radon C, Mittelstädt G, Duffus BR, Bürger J, Hartmann T, Mielke T, et al. Cryo-EM structures reveal intricate Fe-S cluster arrangement and charging in *Rhodobacter capsulatus* formate dehydrogenase. *Nat Commun*. 2020 Apr 20;11(1):1912.
183. Robertson MJ, van Zundert GCP, Borrelli K, Skiniotis G. GemSpot: a pipeline for robust modeling of ligands into cryo-EM maps. *Structure*. 2020 Jun 2;28(6):707–716.e3.
184. Vant JW, Lahey SLJ, Jana K, Shekhar M, Sarkar D, Munk BH, et al. Flexible fitting of small molecules into electron microscopy maps using molecular dynamics simulations with neural network potentials. *J Chem Inf Model*. 2020 May 26;60(5):2591–604.
185. Debreczeni JÉ, Emsley P. Handling ligands with Coot. *Acta Crystallogr D Biol Crystallogr*. 2012 Apr;68(Pt 4):425–30.
186. Moriarty NW, Grosse-Kunstleve RW, Adams PD. Electronic ligand builder and optimization workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr*. 2009 Sep 16;65(10):1074–80.
187. Kim JJ, Gharpure A, Teng J, Zhuang Y, Howard RJ, Zhu S, et al. Shared structural mechanisms of general anaesthetics and benzodiazepines. *Nature*. 2020 Sep;585(7824):303–8.
188. Bai Y, Yu X, Chen H, Horne D, White R, Wu X, et al. Structural basis for pharmacological modulation of the TRPC6 channel. *Elife*. 2020 Mar 9;9:e53311.
189. Zhang X, Belousoff MJ, Zhao P, Kooistra AJ, Truong TT, Ang SY, et al. Differential GLP-1R binding and activation by peptide and non-peptide agonists. *Mol Cell*. 2020 Nov 5;80(3):485–500.e7.
190. Raisch T, Brockmann A, Ebbinghaus-Kintscher U, Freigang J, Gutbrod O, Kubicek J, et al. Small molecule modulation of the *Drosophila* Slo channel elucidated by cryo-EM. *Nat Commun*. 2021 Dec 9;12(1):7164.
191. Roos K, Wu C, Damm W, Reboul M, Stevenson JM, Lu C, et al. OPLS3e: extending force field coverage for drug-like small molecules. *J Chem Theory Comput*. 2019 Mar 12;15(3):1863–74.

192. van Zundert GCP, Moriarty NW, Sobolev OV, Adams PD, Borrelli KW. Macromolecular refinement of x-ray and cryoelectron microscopy structures with Phenix/OPLS3e for improved structure and ligand quality. *Structure*. 2021 Aug 5;29(8):913–21.e4.
193. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*. 2007 Jul;35(Web Server):W375–83.
194. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010 Jan;66(Pt 1):12–21.
195. Sobolev OV, Afonine PV, Moriarty NW, Hekkelman ML, Joosten RP, Perrakis A, et al. A global Ramachandran score identifies protein structures with unlikely stereochemistry. *Structure*. 2020 Nov 3;28(11):1249–58.e2.
196. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol*. 1997 Apr 4;267(3):727–48.
197. Korb O, Stüttgen T, Exner TE. PLANTS: application of ant colony optimization to structure-based drug design. *Ant colony optimization and swarm intelligence*. Berlin Heidelberg: Springer; 2006. p. 247–58.
198. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010 Jan 30;31(2):455–61.
199. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*. 2004 Mar 1;47(7):1739–49.
200. Hryc CF, Baker ML. Beyond the backbone: the next generation of pathwalking utilities for model building in cryoEM density maps. *Biomolecules* [Internet]. 2022 Jun 2;12(6):773. <https://doi.org/10.3390/biom12060773>
201. Joseph AP, Lagerstedt I, Jakobi A, Burnley T, Patwardhan A, Topf M, et al. Comparing cryo-EM reconstructions and validating atomic model fit using difference maps. *J Chem Inf Model*. 2020 May 26;60(5):2552–60.
202. Peña A, Sweeney A, Cook AD, Locke J, Topf M, Moores CA. Structure of microtubule-trapped human kinesin-5 and its mechanism of inhibition revealed using cryoelectron microscopy. *Structure*. 2020 Apr 7;28(4):450–7.e5.
203. Joseph AP, Lagerstedt I, Patwardhan A, Topf M, Winn M. Improved metrics for comparing structures of macromolecular assemblies determined by 3D electron-microscopy. *J Struct Biol*. 2017 Jul 1;199(1):12–26.
204. Afonine PV, Klaholz BP, Moriarty NW, Poon BK, Sobolev OV, Terwilliger TC, et al. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr D Struct Biol*. 2018 Sep 1;74(Pt 9):814–40.
205. Chacón P, Wriggers W. Multi-resolution contour-based fitting of macromolecular structures. *J Mol Biol*. 2002 Mar 29;317(3):375–84.
206. Cragolini T, Sahota H, Joseph AP, Sweeney A, Malhotra S, Vasishtan D, et al. TEMPy2: a Python library with improved 3D electron microscopy density-fitting and validation workflows. *Acta Crystallogr D Struct Biol*. 2021 Jan 1;77(Pt 1):41–7.
207. DiMaio F, Zhang J, Chiu W, Baker D. Cryo-EM model validation using independent map reconstructions. *Protein Sci*. 2013 Jun;22(6):865–8.
208. Amunts A, Brown A, Bai XC, Llaser JL, Hussain T, Emsley P, et al. Structure of the yeast mitochondrial large ribosomal subunit. *Science*. 2014 Mar 28;343(6178):1485–9.
209. Zhao D, Liu W, Chen K, Wu Z, Yang H, Xu Y. Structure of the human RNA polymerase I elongation complex. *Cell Discov*. 2021 Oct 20;7(1):97.
210. Cragolini T, Kryshtafovych A, Topf M. Cryo-EM targets in CASP14. *Proteins* [Internet]. 2021 Aug 16; (prot.26216); Available from:89: 1949–58. <https://doi.org/10.1002/prot.26216>
211. Ramírez-Aportela E, Maluenda D, Fonseca YC, Conesa P, Marabini R, Heymann JB, et al. FSC-Q: a CryoEM map-to-atomic model quality validation based on the local Fourier shell correlation. *Nat Commun*. 2021 Jan 4;12(1):1–7.
212. Warshamanage R, Yamashita K, Murshudov GN. EMDA: a python package for electron microscopy data analysis. *J Struct Biol*. 2022 Mar;214(1):107826.
213. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci*. 2021 Jan;30(1):70–82.
214. Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol*. 2019 Oct 1;75(Pt 10):861–77.
215. Barad BA, Echols N, Wang RYR, Cheng Y, DiMaio F, Adams PD, et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat Methods*. 2015 Aug 17;12(10):943–6.
216. Pintilie G, Zhang K, Su Z, Li S, Schmid MF, Chiu W. Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat Methods*. 2020 Feb 10;17(3):328–34.
217. Olek M, Joseph AP. Cryo-EM map-based model validation using the false discovery rate approach. *Front Mol Biosci*. 2021 May 18;8: 652530.
218. Pandurangan AP, Topf M. RIBFIND: a web server for identifying rigid bodies in protein structures and to aid flexible fitting into cryo EM maps. *Bioinformatics*. 2012 Jul 12;28(18):2391–3.
219. Shapovalov MV, Dunbrack RL Jr. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure*. 2011 Jun 8;19(6):844–58.
220. Kryshtafovych A, Malhotra S, Monastyrskyy B, Cragolini T, Joseph APP, Chiu W, et al. Cryo-electron microscopy targets in CASP13: overview and evaluation of results. *Proteins: Struct Funct Bioinf*. 2019 Dec 1;87(12):1128–40.
221. Lawson CL, Kryshtafovych A, Adams PD, Afonine PV, Baker ML, Barad BA, et al. Cryo-EM model validation recommendations based on outcomes of the 2019 EMDDataResource challenge. *Nat Methods*. 2021 Feb 4;18(2):156–64.
222. Wlodawer A, Li M, Dauter Z. High-resolution Cryo-EM maps and models: a crystallographer's perspective. *Structure*. 2017 Oct 3; 25(10):1589, e1–97.

223. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr D Biol Crystallogr*. 1999 Jan;55(Pt 1):247–55.
224. Zhang K, Pintilie GD, Li S, Schmid MF, Chiu W. Resolving individual atoms of protein complex by cryo-electron microscopy. *Cell Res*. 2020 Nov 2;30(12):1136–9.
225. Pintilie G, Chiu W. Validation, analysis and annotation of cryo-EM structures. *Acta Crystallogr D Struct Biol*. 2021 Sep 1;77(Pt 9):1142–52.
226. Hryc CF, Chen DH, Afonine PV, Jakana J, Wang Z, Haase-Pettingell C, et al. Accurate model annotation of a near-atomic resolution cryo-EM map. *Proc Natl Acad Sci U S A*. 2017 Mar 21;114(12):3103–8.
227. Janssen BJC, Read RJ, Brünger AT, Gros P. Crystallography: crystallographic evidence for deviating C3b structure. *Nature*. 2007 Aug 9;448(7154):E1–3.
228. Pintilie G, Chen DH, Haase-Pettingell CA, King JA, Chiu W. Resolution and probabilistic models of components in cryoEM maps of mature P22 bacteriophage. *Biophys J*. 2016 Feb 23;110(4):827–39.
229. Mendez JH, Stagg SM. Assessing the quality of single particle reconstructions by atomic model building. *J Struct Biol*. 2018 Nov;204(2):276–82.
230. Volkman N. Confidence intervals for fitting of atomic models into low-resolution densities. *Acta Crystallogr D Biol Crystallogr*. 2009 Jul;65(Pt 7):679–89.
231. Markwick PRL, Malliavin T, Nilges M. Structural biology by NMR: structure, dynamics, and interactions. *PLoS Comput Biol*. 2008 Sep 26;4(9):e1000168.
232. Headd JJ, Echols N, Afonine PV, Grosse-Kunstleve RW, Chen VB, Moriarty NW, et al. Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr D Biol Crystallogr*. 2012 Mar 16;68(4):381–90.
233. Prisant MG, Williams CJ, Chen VB, Richardson JS, Richardson DC. New tools in MolProbity validation: CaBLAM for CryoEM backbone, UnDowser to rethink “waters,” and NGL viewer to recapture online 3D graphics. *Protein Sci*. 2020 Jan;29(1):315–29.
234. Richardson JS, Williams CJ, Hintze BJ, Chen VB, Prisant MG, Videau LL, et al. Model validation: local diagnosis, correction and when to quit. *Acta Crystallogr D Struct Biol*. 2018 Feb 1;74(Pt 2):132–42.
235. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci*. 2018 Jan;27(1):293–315.
236. Williams CJ. Using C-alpha geometry to describe protein secondary structure and motifs [PhD]. David C. Richardson JSR, editor. Duke University; 2015.
237. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci*. 2006 Nov;15(11):2507–24.
238. Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. QMEANDisCo-distance constraints applied on model quality estimation. *Bioinformatics*. 2020 Mar 1;36(6):1765–71.
239. Uziela K, Menéndez Hurtado D, Shu N, Wallner B, Elofsson A. ProQ3D: improved model quality assessments using deep learning. *Bioinformatics*. 2017 May 15;33(10):1578–80.
240. Maghrabi AHA, McGuffin LJ. Estimating the quality of 3D protein models using the ModFOLD7 server. *Methods Mol Biol*. 2020;2165:69–81.
241. Hou J, Wu T, Cao R, Cheng J. Protein tertiary structure modeling driven by deep learning and contact distance prediction in CASP13. *Proteins*. 2019 Dec;87(12):1165–78.
242. Cheng J, Choe MH, Elofsson A, Han KS, Hou J, Maghrabi AHA, et al. Estimation of model accuracy in CASP13. *Proteins*. 2019 Dec;87(12):1361–77.
243. Malhotra S, Joseph AP, Thiyaalingam J, Topf M. Assessment of protein–protein interfaces in cryo-EM derived assemblies. *Nat Commun*. 2021 Dec 7;12(1):3399.
244. Wlodawer A, Minor W, Dauter Z, Jaskolski M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J*. 2008 Jan;275(1):1–21.
245. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996 Feb;14(1):33–8. 27–8.
246. Phillips JC, Hardy DJ, Maia JDC, Stone JE, Ribeiro JV, Bernardi RC, et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys*. 2020 Jul 28;153(4):044130.
247. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015 Sep;1(1–2):19–25.
248. Ferrin TE, Huang CC, Jarvis LE, Langridge R. The MIDAS display system. *J Mol Graph*. 1988 Mar 1;6(1):13–27.
249. Tarini M, Cignoni P. QuteMol. 2006 [cited 2022 Jan 24]; Available from: <https://air.unimi.it/handle/2434/625498>
250. Chen VB, Davis IW, Richardson DC. KING (Kinemage, next generation): a versatile interactive molecular and scientific visualization program. *Protein Sci*. 2009 Nov;18(11):2403–9.
251. Richardson DC, Richardson JS. The kinemage: a tool for scientific communication. *Protein Sci*. 1992 Jan;1(1):3–9.
252. Davis IW, Arendall WB 3rd, Richardson DC, Richardson JS. The backrub motion: how protein backbone shrugs when a sidechain dances. *Structure*. 2006 Feb;14(2):265–74.
253. Croll TI. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol*. 2018 Jun 1;74(Pt 6):519–30.
254. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci*. 2018 Jan;27(1):14–25.
255. Hinsen K. The molecular modeling toolkit: a new approach to molecular simulations. *J Comput Chem*. 2000 Jan 30;21(2):79–85.

256. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 2022 Jan 7;50(D1):D439–44.
257. Croll TI, Read RJ. Adaptive Cartesian and torsional restraints for interactive model rebuilding. *Acta Crystallogr D Struct Biol.* 2021 Apr 1;77(Pt 4):438–46.
258. McGreevy R, Singharoy A, Li Q, Zhang J, Xu D, Perozo E, et al. xMDFF: molecular dynamics flexible fitting of low-resolution x-ray structures. *Acta Crystallogr D Biol Crystallogr.* 2014 Sep;70(Pt 9):2344–55.
259. Bakan A, Meireles LM, Bahar I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics.* 2011 Jun 1;27(11):1575–7.
260. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, DeBolt S, et al. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput Phys Commun.* 1995 Sep 2;91(1):1–41.
261. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data.* 2016 Mar 15;3:160018.
262. Tagari M, Newman R, Chagoyen M, Carazo JM, Henrick K. New electron microscopy database and deposition system. *Trends Biochem Sci.* 2002 Nov;27(11):589.
263. Grabowski M, Langner KM, Cymborowski M, Porebski PJ, Sroka P, Zheng H, et al. A public database of macromolecular diffraction experiments. *Acta Crystallogr D Struct Biol.* 2016 Nov 1;72(Pt 11):1181–93.
264. Iudin A, Korir PK, Salavert-Torres J, Kleywegt GJ, Patwardhan A. EMPIAR: a public archive for raw electron microscopy image data. *Nat Methods.* 2016 May;13(5):387–8.
265. Porebski PJ, Sroka P, Zheng H, Cooper DR, Minor W. Molstack-interactive visualization tool for presentation, interpretation, and validation of macromolecules and electron density maps. *Protein Sci.* 2018 Jan;27(1):86–94.
266. WebGL [Internet]. The Khronos Group. 2011 [cited 2022 Jun 21]. Available from: <https://www.khronos.org/webgl/>
267. Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, et al. Mol* viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res.* 2021 Jul 2;49(W1):W431–7.
268. Rose AS, Hildebrand PW. NGL viewer: a web application for molecular visualization. *Nucleic Acids Res.* 2015 Jul 1;43(W1):W576–9.
269. Rose AS, Bradley AR, Valasatava Y, Duarte JM, Prlic A, Rose PW. NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics.* 2018 Nov 1;34(21):3755–8.
270. Bayarri G, Hospital A, Orozco M. 3dRS, a web-based tool to share interactive representations of 3D biomolecular structures and molecular dynamics trajectories. *Front Mol Biosci.* 2021 Aug 13;8:726232.
271. Git [Internet] [cited 2022 Jun 21]. Available from: <https://git-scm.com/>
272. Reynolds CR, Islam SA, Sternberg MJE. EzMol: a web server wizard for the rapid visualization and image production of protein and nucleic acid structures. *J Mol Biol.* 2018 Jul 20;430(15):2244–8.
273. Chen TH, Lee S, Flood AH, Miljanić OŠ. How to print a crystal structure model in 3D. *CrstEngComm.* 2014;16(25):5488–93.
274. Croll TI, Diederichs K, Fischer F, Fyfe CD, Gao Y, Horrell S, et al. Making the invisible enemy visible. *Nat Struct Mol Biol.* 2021 May; 28(5):404–8.
275. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Židek A, et al. Highly accurate protein structure prediction for the human proteome. *Nature.* 2021 Aug;596(7873):590–6.
276. Hiranuma N, Park H, Baek M, Anishchenko I, Dauparas J, Baker D. Improved protein structure refinement guided by deep learning based accuracy estimation. *Nat Commun.* 2021 Feb 26;12(1):1340.
277. Terwilliger TC, Poon BK, Afonine PV, Schlicksup CJ, Croll TI, Millán C, et al. Improving AlphaFold modeling using implicit information from experimental density maps. *bioRxiv [Internet].* 2022 [cited 2022 Sep 15]. p. 2022.01.07.475350. <https://doi.org/10.1101/2022.01.07.475350v1>
278. Kappel K, Zhang K, Su Z, Watkins AM, Kladwang W, Li S, et al. Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nat Methods.* 2020 Jul;17(7):699–707.
279. Macé K, Vadakkepat AK, Redzej A, Lukoyanova N, Oomen C, Braun N, et al. Cryo-EM structure of a type IV secretion system. *Nature [Internet].* 2022 Jun 22;607:191–6. <https://doi.org/10.1038/s41586-022-04859-y>
280. Khalid S, Rouse SL. Simulation of subcellular structures. *Curr Opin Struct Biol.* 2020 Apr;61:167–72.
281. Yu A, Pak AJ, He P, Monje-Galvan V, Casalino L, Gaieb Z, et al. A multiscale coarse-grained model of the SARS-CoV-2 virion. *Biophys J.* 2021 Mar 16;120(6):1097–104.
282. Souza PCT, Alessandri R, Barnoud J, Thallmair S, Faustino I, Grünewald F, et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat Methods.* 2021 Apr;18(4):382–8.
283. Saur M, Hartshorn MJ, Dong J, Reeks J, Bunkoczi G, Jhoti H, et al. Fragment-based drug discovery using cryo-EM. *Drug Discov Today.* 2020 Mar 1;25(3):485–90.

How to cite this article: Beton JG, Cragolini T, Kaleel M, Mulvaney T, Sweeney A, Topf M. Integrating model simulation tools and cryo-electron microscopy. *WIREs Comput Mol Sci.* 2023;13(3):e1642. <https://doi.org/10.1002/wcms.1642>