Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Finding the right fit: chiseling structures out of cryo-electron microscopy maps

Elizabeth Villa[1] and Keren Lasker[2]

Cryo-electron microscopy is a central tool for studying the architecture of macromolecular complexes at subnanometer resolution. Interpretation of an electron microscopy map requires its computational integration with data about the structure's components from all available sources, notably atomic models. Selecting a protocol for EM density-guided integrative structural modeling depends on the resolution and quality of the EM map as well as the available complimentary datasets. Here, we review rigid, flexible, and *de novo* integrative fitting into EM maps and provide guidelines and considerations for the design of modeling experiments. Finally, we discuss efforts towards establishing unified criteria for map and model assessment and validation.

Addresses
[1] Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093, United States
[2] Department of Developmental Biology, Stanford University, Stanford, CA 94305, United States

Corresponding authors: Villa, Elizabeth (evilla@ucsd.edu) and Lasker, Keren (klasker@stanford.edu)

Current Opinion in Structural Biology 2014, **25**:118–125

This review comes from a themed issue on **Theory and simulation**

Edited by **Rommie E Amaro** and **Manju Bansal**

0959-440X/$ – see front matter,

http://dx.doi.org/10.1016/j.sbi.2014.04.001.

Introduction

Cryo-electron microscopy (cryo-EM) is a central tool in structural biology ideally suited for studying macromolecular complexes that are challenging for other structure-determination methods. In the modality of single-particle analysis (SPA) [1], 2D micrographs are combined into three-dimensional maps where voxels of varying intensities represent the structure of the macromolecular complex under observation. Typically, EM maps generated by SPA range in resolution from 3 to 20 Å, strongly depending on the nature of the complex under study [2••]. The higher end of this range is achieved for symmetric and stable structures, such as viruses. Labile and transient complexes are difficult to purify and freeze homogeneously resulting in lower resolution EM maps. The level of interpretation, and thus the biological questions that can be addressed, depends on the resolution of the EM map and varies from rough localization of component proteins to precise assignment of amino acid side chains.

In most cases, the resolution of an EM map is insufficient to provide a full atomic description of its underlying structure. Thus, computational integration of atomic structures with EM maps is regularly performed. The resolution of the EM map is often adequate for accurate placement of atomic structures of the subunits into the density map, resulting in a model of the entire assembly. Given sufficient resolution, flexible fitting can be used to further refine the model by adjusting initial atomic models to better represent their conformation in the EM map while maintaining their physicochemical properties. In some cases, the resolution enables the atomic structure to be modeled *de novo*.

Here, we provide guidelines and considerations for the design of modeling experiments based on EM maps. We have found that concepts that are well known in the field of SPA are not appreciated by the molecular modeling community and vice versa. As the fields of SPA and molecular modeling integrate due to technical advances in cryo-EM, we hope that this review will encourage experts in one discipline to become conversational in the other. The literature covered in this review relates to the challenges encountered in EM-guided structural modeling, without emphasis on particular methods. A comprehensive list of techniques and data sources for building atomic structures into EM maps can be found in Table S1.

The unresolved resolution criterion

The resolution criterion for SPA reconstructions remains a matter of debate. The most common measure is the Fourier shell correlation (FSC), which estimates signal-to-noise as a function of spatial frequency [3]. While FSC is a good measure of resolution, the resolution is overestimated if the reconstruction protocol fails to prevent the enhancement of noise at high frequencies caused by overfitting of 2D images during alignment [4]. Overestimation of resolution affects many published EM maps and must be considered when planning modeling experiments. With the increasing applicability of SPA, the community is working to establish stricter criteria for estimating the quality and resolution of SPA reconstructions [5–8].

In practice, resolution is not uniform across the EM map and may vary considerably due to factors such as flexibility or low occupancy in certain regions of the structure. This apparent limitation is potentially one of the greatest

strengths of cryo-EM: since the cryo-frozen particles resemble macromolecules in solution, capturing heterogeneous conformations or subunit compositions within the same sample can yield information on structural dynamics [9–11] and kinetics [12]. However, varying resolution can present a challenge for structural modeling. Using a single measure of resolution leads to overinterpretation of certain areas of the map while discarding high resolution information in others. An alternative approach is to use estimates of local resolution, such as FSC calculated for masked regions of the map defined by a sliding 3D window. Recent studies have incorporated local resolution to assess the quality of the EM map [13•,30•] and to filter the final reconstruction before fitting atomic structures into it [11•,15]. Tools for determining local resolution [16••,17] furnish 3D grids with resolution independently defined at each voxel. For both rigid and flexible fitting applications, we strongly encourage local-resolution filtering of EM maps. Local-resolution filtering is also advised for atomic models when these models are used for cross-correlation calculations.

Choosing the right integrative modeling approach

The resolution of an EM map and the available atomic information often do not allow for an unambiguous determination of the underlying structure. In such cases, the interpretation of an EM map must be assisted by the incorporation of heterogeneous experimental datasets at different resolutions. This approach, broadly termed 'integrative modeling' [18,19], compensates for coverage, accuracy, and precision limitations of a single data set.

Selecting a protocol for EM density-guided integrative modeling depends on the resolution of the EM map as well as the available datasets. An EM map has the unique property of providing both global and local information for integrative modeling, and thus acts as both a scaffold and a source of fine structural information. In order to guide conformational sampling and to improve the convergence and accuracy of the resulting model, modeling should be implemented by integrating complimentary datasets from different sources. These datasets may include composition and copy numbers from proteomics, location of individual subunits through labeling or deletions, distances between subunits and stoichiometry from FRET, single-molecule and super-resolution light-microscopy, protein and residue proximities from chemical cross-linking and proteomics experiments, the overall shape of subunits from SAXS profiles, and atomic structures from X-ray crystallography and nuclear magnetic resonance. Additionally, statistical inferences from biological databases and physicochemical properties of macromolecules are typically incorporated in the analysis [14,20–22]. Integrative modeling play a critical role in determining the molecular architecture of key macromolecular complexes including the nuclear pore complex [23] and its Nup84 subcomplex [24], the anaphase-promoting complex [25,26], the human polycomb repressive complex 2 [27], the human protein phosphatase 2A [28], the telomerase holoenzyme [29], and the 26S proteasome [13•,30•].

Rigid-body and multiple fitting localize components of the complex

Rigid-body-fitting (RBF) methods perform a six-dimensional search to localize an atomic structure in an EM map by maximizing the goodness-of-fit between the EM map and a simulated density map of the fitted atomic model [31] (Table S1). Maximization of the cross correlation (CC) score is a good predictor for goodness-of-fit. Several other scores, including Laplacian-filtered CC, difference least-squares, envelope score, and mutual information can be useful predictors at different resolution regimes [32]. When applying rigid body fitting, one must give consideration to the representation of the atomic model as a simulated density map, conformational changes, and fitting of multiple atomic models into one map.

A simulated density is typically obtained by projecting an atomic structure into a cubic lattice of equivalent dimensions to the EM map and convoluting it with a kernel emulating the point-spread function (e.g. a Gaussian kernel with a half radius determined by the resolution of the EM map). This description is too simplistic when simulating a density map at subnanometer resolution as local variations in resolution affect the accuracy of the goodness-of-fit measurement. The simulated density map can be normalized to that effect by adjusting the width of the kernel per voxel as a function of the local resolution of the corresponding voxel in the EM map. An alternative method produces a more accurate description of the structure factors that accounts for local variation in resolution while performing model optimization [33•]. This iterative resolution-dependent density simulation should improve the accuracy of rigid and flexible fitting and of model verification algorithms.

The goodness-of-fit score is also affected by the conformation of the atomic structure. If the conformation of the fitted structure differs from that found in the complex, the goodness-of-fit score might be artificially low. In cases where conformational differences considerably affect the reliability of the fit, it is possible assign some flexibility to the atomic models. A common approach is to divide the atomic models into smaller rigid bodies and to independently and iteratively fit them to better reflect the conformation in the EM map [34]. Caution must be placed on how these rigid bodies are selected for fitting and how they are connected after fitting, as these choices affect the quality of the final model [35].

When rigid bodies of the complex are fitted separately, the uncertainties in RBF may lead to overlaps in their positions if they are fitted independently and sequentially. For such