Published in final edited form as:

Arch Biochem Biophys. 2015 September 1; 581: 49–53. doi:10.1016/j.abb.2015.06.017.

Validation methods for low-resolution fitting of atomic structures to electron microscopy data

Xiao-Ping Xu and Niels Volkmann*

Bioinformatics and Structural Biology Program, Sanford-Burnham Medical Research Institute,
10901 N Torrey Pines Rd, La Jolla, CA 92037, USA

Abstract

Fitting of atomic-resolution structures into reconstructions from electron cryo-microscopy is routinely used to understand the structure and function of macromolecular machines. Despite the fact that a plethora of fitting methods has been developed over recent years, standard protocols for quality assessment and validation of these fits have not been established. Here, we present the general concepts underlying current validation ideas as they relate to fitting of atomic-resolution models into electron cryo-microscopy reconstructions, with an emphasis on reconstructions with resolutions below the sub-nanometer range.

Keywords

fitting; electron microscopy; validation

Introduction

Due to dramatic improvements in experimental methods and computational techniques, electron cryo-microscopy (cryo-EM) has matured into a powerful collection of methods that allow the high-resolution visualization of the structure and the dynamics of an extraordinary range of biological assemblies in their native aqueous environment. Recent hardware and software developments have revolutionized the field [1]. The increased signal-to-noise ratio of a new generation of cameras that detect electrons directly [2] in combination with their ability to correct for beam-induced movements, have allowed the field to obtain structural information even for particles with low or no symmetry at resolutions around 3 Å [3–8], sufficient to build de novo structural models [9].

However, the majority of reconstructions obtained by cryo-EM are of insufficient resolution for such direct structure determination. In fact, currently over 70% of the reconstructions deposited in the electron microscopy data bank [10] do not reach a resolution of better than 10 Å. While at resolutions between 5 and 10 Å secondary structural elements are often

*Corresponding author: niels@sbmri.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

visible as rods (α -helices) and sheets (β -sheets), at resolutions below the 10-Å mark, internal features of the reconstructions are not straightforward to interpret (Figure 1).

As cryo-EM methodology continues to improve, atomic-resolution reconstructions are likely to become more common. These reconstructions will likely be of highly rigid molecules. At the same time, the advances in cryo-EM technology will also open the door to structure determination of complexes that were previously too small, too heterogeneous, too flexible, or otherwise challenging, albeit at lower resolution. In addition, electron cryo-tomography has become a powerful alternative for structure determination of samples that are not amenable to single-particle approaches. New software developments and careful experimental design [11] enable the determination of structures from cryo tomograms at around 8 Å, but resolutions below 20 Å are more common. As a net-effect, the majority of cryo-EM reconstructions are likely to remain at the resolution range worse than 10 Å for the foreseeable future.

Fitting of atomic components into cryo-EM density maps insufficient for direct de-novo model building is routinely used to understand the structure and function of these macromolecular machines. Many fitting methods have been developed, but standard protocols for successful fitting remain to be established. Broadly, fitting methods can be divided into two major groups, rigid-body and flexible fitting methods. In rigid-body fitting approaches the atomic structures of components are fitted as single units. These units can be composed of entire proteins, domains, or even smaller groups of structural element. In flexible fitting approaches the entire atomic structures are allowed to distort in some way to improve the fit with the reconstruction, subject to constraints such as molecular dynamics force-fields or normal modes to counter-balance fitting of spurious noise. Comprehensive reviews of the various fitting approaches that are available were provided in several recent articles [12–14].

Despite the plethora of available fitting techniques, generally accepted criteria for assessing the accuracy and quality of the fitted models have not been established yet [15]. In this review, our aim is to present the general concepts underlying current validation ideas as they relate to fitting of atomic models into cryo-EM reconstructions, with an emphasis on reconstructions with resolutions that do not reach the sub-nanometer range.

Accuracy versus precision

In the context of fitting, it is important to emphasize the difference between accuracy and precision (Figure 2). In short, a fit is precise if similar fits are obtained with repeated runs. In contrast, a fit is accurate if it is close to the true structure (or ensemble of structures) underlying the data. Consensus approaches that compare results from different fitting methods [16,17] or from multiple scoring functions [18] using a single data set, for example, can only inform on the precision of the fit. While it has been claimed that precision gives a lower bound for the accuracy, this is not necessarily true (Figure 2). In fact, in the context of fitting atomic structures into cryo-EM reconstructions, the fitting that appears less precise can actually be more accurate than a fit with the same center position and a narrower spread. The reason is that cryo-EM reconstructions, unlike crystal structures, often represent an

ensemble of conformers that coexisted in the sample at the time of freezing, reflecting the structural dynamics of the complex in solution. This can also lead to anisotropy of the resolution in the reconstructions, further complicating the issue.

It is not immediately clear how precision can be quantified within the context of fitting atomic models. This issue is of major importance especially at low resolution, where ambiguities may arise from the geometry of the reconstruction alone [19] and an objective criterion to allow favoring one solution over another is needed. One approach towards this goal is the use of statistical methods to define confidence intervals (see below), which will then allow to define an objective precision estimate. However, the real quantity that is of interest is the accuracy or how close the obtained fit is to the true structure. Without knowledge of the true structure accuracy cannot be directly assessed.

Sources of errors

Every cryo-EM reconstruction has some uncertainty due to the presence of noise and its generally limited resolution. If only such random errors exist, a quantified precision measure may actually be a reasonable estimate for the accuracy of a fit. More troublesome in this context are potential systematic errors such as those originating in the electron microscopy data-collection and reconstruction procedures. These include misestimations of the magnification, incomplete corrections of the microscope's contrast transfer function, uneven distribution of projection images in Euler space, and misestimation of the resolution. All of these can introduce bias and throw off the fits from the true solution in a non-random fashion. Other factors that can bias fitting results in significant ways include incompleteness of the structure to be fitted in relation to the target reconstruction and the possible presence of conformational mixtures. Generally, rigid-body techniques are less susceptible to generating artifacts in the presence of errors simply because the underlying structure is kept intact and only six parameters per rigid body, three for the center-of-mass position and three for the orientation, need to be determined.

Another source of systematic errors is the fact that, at low resolution, the amount of structural information that can be obtained is limited. Care must be taken that the number of refinable parameters used during fitting does not exceed the number of independent observations in the reconstruction. Otherwise, overfitting will inevitably ensue. Even fitting of a single rigid body using only the six rotational and translational degrees of freedom can lead to ambiguities in the resulting models at intermediate resolution [19]. Generally, all available experimental restraints that are independent of the fitting process should be employed for validation purposes. However, inclusion of such restraints into the fitting process can sometimes be helpful to resolve ambiguities.

Recently developed methods for incorporating data from other sources such as proteomics [20], Förster resonance energy transfer [21], or sparse distance restraints [22] into the fitting process have shown some very encouraging potential [23,24], but they also have to be handled with care. An example is a discrepancy in fitting-based interpretation of the ryanodine receptor [25,26]. In the earlier study, a homology model was fitted into the region of a reconstruction that was indicated by antibody labeling. In the other study, fitting with an

atomic structure of a fragment was performed followed by an evaluation of ambiguities using confidence intervals. The differences in fitting protocols completely altered the final fit. More recent fits based on high-resolution reconstructions [27,28] confirmed the statistics based fit [26] to be correct indicating that the labeling constraint [25] threw off the fitting rather than improved the fit (Figure 3).

Crossvalidation

One technique that can help overcoming overfitting caused by an inadequate ratio of refinable parameters to observables is crossvalidation [29]. In crossvalidation, the data is split into a “training set” and a “test set”. Calculations are then performed on the training set while evaluation is done using the test set, so that an adequate set of parameters can be determined (Figure 4). While the fit to the training set will continue to improve due to overfitting, the test set fit will at some point worsen, thus giving optimal values for the fitting parameters involved. This scheme corresponds to a 2-fold crossvalidation also known as holdout method. This is the simplest form of k -fold crossvalidation and has the disadvantage that it usually needs to set apart a sizable portion of the data into the test set that is not accessible for use in the fitting and that an “unfortunate” split can lead to severe misestimations of the parameters. k -fold crossvalidation overcomes this problem by repeating the calculations with k different splits that cover the whole data set and thus are less amenable to artifacts caused by a specific split and also allows a larger fraction of the data to participate in the fitting calculations.

In structural biology, the most common form of crossvalidation is a 2-fold crossvalidation scheme implemented in Fourier space for X-ray crystallography in the form of the “free R-factor” [30]. A crucial prerequisite for crossvalidation to work is that the information in the test set is independent from that in the work set. For the Fourier terms in X-ray crystallography this assumption is usually justified. In electron microscopy the Fourier terms are strongly correlated so that the free R-factor it is not applicable in a direct analogy to crystallography. Several factors introduce correlations between Fourier terms in cryo-EM reconstructions. If the particle is of limited size in real space, placed in an empty box, then there are automatically correlations between neighboring Fourier components. Equally, low-pass filtering, a common operation in cryo-EM, has the same effect. The alignment of images during the reconstruction process also tends to introduce correlation in the noise of these images [31].

Crossvalidation in cryo-EM fitting

In the current context of fitting atomic structures into cryo-EM reconstructions, two conceptually different approaches have been proposed [32,33]. The first approach [32] relies on splitting the data according to frequency ranges where low-resolution frequency data with significant signal-to-noise ratio is used as a training set for model building and a shell of high-frequency data is used as a test set. As expected, significant correlations between the training and test sets can be detected. The authors propose a work-around where they compare the signal in the test set with a signal from a “perfectly overfitted” bead model, which is used as a baseline. Results using test cases with resolutions between 5 and 10 Å are

promising but whether the method works at lower resolutions remains to be seen, especially because correlations between shells will be a more serious problem for lower resolution (the Fourier components with a spatial frequency corresponding to 8 and 9 Å are closer in Fourier space than the 3 and 4 Å components).

In the second crossvalidation approach [33] the data is split randomly into two independent sets and reconstructions are built independently from each set. One of these reconstructions is used as a training set, the second is used as test set. This type of methodology can be used to select sensible weighting terms between the density and all-atom energy contributions using the program Rosetta [34]. However, the authors find that at resolution below 12 Å significant correlation between the two independent maps occur, outweighing the usefulness of the approach at resolution ranges below 12 Å [33].

Confidence intervals

A confidence interval is a range of values that is believed, with some stated level of confidence, to contain the true value of interest. In general, high levels of confidence can be achieved with wider intervals, while narrower, more precise intervals carry less confidence. Thus, there is a trade off between precision and confidence and, in fact, any statement of precision without a corresponding confidence level is incomplete. The advantage of providing a range of values for the estimate is that it will be more likely to include the correct one. Generally, the width of a confidence interval can be interpreted as a measure of precision while the confidence level of an interval is a measure of accuracy. However, this statement is only true if all error sources are accounted for correctly. If undetected systematic errors are present, this relationship is broken.

The use of confidence intervals was shown to be a powerful tool in rigid-body fitting approaches for interpreting cryo-EM reconstructions [35]. In this approach, a global search is followed by a global statistical analysis of the score distribution resulting in the definition of confidence intervals. All fits that have scores within that confidence interval satisfy the data within the error margin defined by the errors in the data and the chosen confidence level. If all errors are accounted for, this will give excellent estimates for precision and accuracy. To account for systematic errors, it is an advantage to use as many independent data sets under varying conditions as possible. Crossvalidation by splitting data into random halves is also an option [35].

Structural parameters of interest can be evaluated as properties of the sets satisfying the confidence interval criterion. For example, the uncertainty of each atom position of the fitted structure can be approximated by calculating the root-mean-square deviation for each atom using all members of the set. The statistical nature of the approach allows the use of standard statistical tests, such as Student's t-test, to evaluate the significance of differences between models in different orientations [19] or functional states [23] and to help model the corresponding conformational changes in a robust and reliable way.

Conclusions

Fitting of atomic structures into low-resolution reconstructions from cryo-EM need de facto standards and tools for assessing the quality and estimating the accuracy of the resulting fits. It is clear that rigorous and objective evaluation criteria are still needed to validate conclusions drawn from fitting of high-resolution structures into lower-resolution reconstructions from electron microscopy. Currently, several promising approaches based on cross-validation and statistical tools to obtain confidence intervals are in development but generally accepted standards are lacking. Thus, better validation criteria are likely to continue to be the subject of intense development in the near future.

Acknowledgments

This work was supported by National Institutes of Health grant R01 CA179087.

References

1. Kühlbrandt W. Cryo-EM enters a new era. *Elife*. 2014; 3:e03678. [PubMed: 25122623]
2. Grigorieff N. Direct detection pays off for electron cryo-microscopy. *Elife*. 2013; 2:e00573. [PubMed: 23426864]
3. Liao M, Cao E, Julius D, Cheng Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature*. 2013; 504:107–112. [PubMed: 24305160]
4. Allegretti M, Mills DJ, McMullan G, Kühlbrandt W, Vonck J. Atomic model of the F420-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector. *Elife*. 2014; 3:e01963. [PubMed: 24569482]
5. Bartesaghi A, Matthies D, Banerjee S, Merk A, Subramaniam S. Structure of β -galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy. *Proc Natl Acad Sci U S A*. 2014; 111:11709–11714. [PubMed: 25071206]
6. Voorhees RM, Fernández IS, Scheres SH, Hegde RS. Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution. *Cell*. 2014; 157:1632–1643. [PubMed: 24930395]
7. Wong W, Bai XC, Brown A, Fernandez IS, Hanssen E, Condrón M, Tan YH, Baum J, Scheres SH. Cryo-EM structure of the *Plasmodium falciparum* 80S ribosome bound to the anti-protozoan drug emetine. *Elife*. 2014; 3:e03080.
8. Amunts A, Brown A, Bai XC, Llácer JL, Hussain T, Emsley P, Long F, Murshudov G, Scheres SH, Ramakrishnan V. Structure of the yeast mitochondrial large ribosomal subunit. *Science*. 2014; 343:1485–1489. [PubMed: 24675956]
9. 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; 71:136–153. [PubMed: 25615868]
10. Lawson CL, Baker ML, Best C, Bi C, Dougherty M, Feng P, van Ginkel G, Devkota B, Lagerstedt I, Ludtke SJ, Newman RH, Oldfield TJ, Rees I, Sahni G, Sala R, Velankar S, Warren J, Westbrook JD, Henrick K, Kleywegt GJ, Berman HM, Chiu W. EMDDataBank.org: unified data resource for CryoEM. *Nucleic Acids Res*. 2011; 39:D456–D464. [PubMed: 20935055]
11. Schur FK, Hagen WJ, de Marco A, Briggs JA. Determination of protein structure at 8.5Å resolution using cryo-electron tomography and sub-tomogram averaging. *J Struct Biol*. 2013; 184:394–400. [PubMed: 24184468]
12. López-Blanco JR, Chacón P. Structural modeling from electron microscopy data. *WIREs Comput Mol Sci*. 2014; 5:62–81.
13. Villa E, Lasker K. Finding the right fit: chiseling structures out of cryo-electron microscopy maps. *Curr Opin Struct Biol*. 2014; 25:118–125. [PubMed: 24814094]
14. Volkmann N. The joys and perils of flexible fitting. *Adv Exp Med Biol*. 2014; 805:137–155. [PubMed: 24446360]

15. Henderson R, Sali A, Baker ML, Carragher B, Devkota B, Downing KH, Egelman EH, Feng Z, Frank J, Grigorieff N, Jiang W, Ludtke SJ, Medalia O, Penczek PA, Rosenthal PB, Rossmann MG, Schmid MF, Schröder GF, Steven AC, Stokes DL, Westbrook JD, Wriggers W, Yang H, Young J, Berman HM, Chiu W, Kleywegt GJ, Lawson CL. Outcome of the first electron microscopy validation task force meeting. *Structure*. 2012; 20:205–214. [PubMed: 22325770]
16. Ahmed A, Whitford PC, Sanbonmatsu KY, Tama F. Consensus among flexible fitting approaches improves the interpretation of cryo-EM data. *J Struct Biol*. 2012; 177:561–570. [PubMed: 22019767]
17. Ahmed A, Tama F. Consensus among multiple approaches as a reliability measure for flexible fitting into cryo-EM data. *J Struct Biol*. 2013; 182:67–77. [PubMed: 23416197]
18. Vasishtan D, Topf M. Scoring functions for cryoEM density fitting. *J Struct Biol*. 2011; 174:333–343. [PubMed: 21296161]
19. Volkmann N, Hanein D. Docking of atomic models into reconstructions from electron microscopy. *Methods Enzymol*. 2003; 374:204–225. [PubMed: 14696375]
20. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprpto A, Karni-Schmidt O, Williams R, Chait BT, Rout MP, Sali A. Determining the architectures of macromolecular assemblies. *Nature*. 2007; 450:683–694. [PubMed: 18046405]
21. Xu XP, Slaughter BD, Volkmann N. Probabilistic determination of probe locations from distance data. *J Struct Biol*. 2013; 184:78–82.
22. Campos M, Francetic O, Nilges M. Modeling pilus structures from sparse data. *J Struct Biol*. 2011; 173:436–444. [PubMed: 21115127]
23. Xu XP, Rouiller I, Slaughter BD, Egile C, Kim E, Unruh JR, Fan X, Pollard TD, Li R, Hanein D, Volkmann N. Three-dimensional reconstructions of Arp2/3 complex with bound nucleation promoting factors. *EMBO J*. 2011; 31:236–247. [PubMed: 21934650]
24. Lasker K, Förster F, Bohn S, Walzthoeni T, Villa E, Unverdorben P, Beck F, Aebersold R, Sali A, Baumeister W. Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc Natl Acad Sci U S A*. 2012; 109:1380–1387. [PubMed: 22307589]
25. Serysheva II, Ludtke SJ, Baker ML, Cong Y, Topf M, Eramian D, Sali A, Hamilton SL, Chiu W. Subnanometer-resolution electron cryomicroscopy-based domain models for the cytoplasmic region of skeletal muscle RyR channel. *Proc Natl Acad Sci U S A*. 2008; 105:9610–9615. [PubMed: 18621707]
26. Tung CC, Lobo PA, Kimlicka L, Van Petegem F. The amino-terminal disease hotspot of ryanodine receptors forms a cytoplasmic vestibule. *Nature*. 2010; 468:585–588. [PubMed: 21048710]
27. Zalk R, Clarke OB, des Georges A, Grassucci RA, Reiken S, Mancina F, Hendrickson WA, Frank J, Marks AR. Structure of a mammalian ryanodine receptor. *Nature*. 2015; 517:44–49. [PubMed: 25470061]
28. Yan Z, Bai XC, Yan C, Wu J, Li Z, Xie T, Peng W, Yin CC, Li X, Scheres SH, Shi Y, Yan N. Structure of the rabbit ryanodine receptor RyR1 at near-atomic resolution. *Nature*. 2015; 517:50–55. [PubMed: 25517095]
29. Efron B, Gong G. A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *Am Stat*. 1983; 37:36–48.
30. Brünger AT. Free R-Value - a Novel Statistical Quantity for Assessing the Accuracy of Crystal-Structures. *Nature*. 1992; 355:472–475. [PubMed: 18481394]
31. Stewart A, Grigorieff N. Noise bias in the refinement of structures derived from single particles. *Ultramicroscopy*. 2004; 102:67–84. [PubMed: 15556702]
32. Falkner B, Schröder GF. Cross-validation in cryo-EM-based structural modeling. *Proc Natl Acad Sci U S A*. 2013; 110:8930–8935. [PubMed: 23674685]
33. DiMaio F, Zhang J, Chiu W, Baker D. Cryo-EM model validation using independent map reconstructions. *Protein Sci*. 2013; 22:865–868. [PubMed: 23592445]
34. 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; 392:181–190. [PubMed: 19596339]
35. Volkmann N. Confidence intervals for fitting of atomic models into low-resolution densities. *Acta Crystallogr D Biol Crystallogr*. 2009; 65:679–689. [PubMed: 19564688]

Highlights

- Low-resolution electron microscopy data does not allow atomic modeling
- Fitting of high-resolution structures allows structural interpretation
- Validation criteria are needed

Figure 1. Spring 2015 snapshot of EM reconstructions deposited in the EM database [10]. Over 70% of the deposited structure do not reach subnanometer resolution and less than 10% reach the 5-Å mark. Some recently deposited example structures are shown to demonstrate the effect of resolution on interpretability in light of atomic resolution structures.

Figure 2.
Difference between accuracy and precision.

Figure 3. Fitting of N-terminal domain structures (residues 1-205 blue; 206-394 green; 395-557 red) to EM densities of the ryanodine receptor

A. Fitting to 9.6-Å EM reconstruction based on labeling constraints [25].

B. Fitting to 9.6-Å EM reconstruction based on confidence intervals [26].

C. Fitting to 3.8-Å EM reconstruction [28].

Figure 4. The principle of crossvalidation

A. Split data into training set and test set.

B. Parameters are modified to actively improve score in training set. Score is monitored in test set to determine optimal parameter value.

C. k -fold ($k=4$) crossvalidation. Data set is split independently k times before calculations are run independently.