

# FOR THE RECORD

## Cryo-EM model validation using independent map reconstructions

Frank DiMaio,<sup>1</sup> Junjie Zhang,<sup>2</sup> Wah Chiu,<sup>3</sup> and David Baker<sup>1\*</sup>

<sup>1</sup>Department of Biochemistry, University of Washington, Seattle, Washington 98195

<sup>2</sup>Department of Structural Biology, Stanford University School of Medicine, Stanford, California 94305

<sup>3</sup>National Center for Macromolecular Imaging, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030

Received 26 October 2012; Accepted 3 April 2013

DOI: 10.1002/pro.2267

Published online 17 April 2013 proteinscience.org

**Abstract:** An increasing number of cryo-electron microscopy (cryo-EM) density maps are being generated with suitable resolution to trace the protein backbone and guide sidechain placement. Generating and evaluating atomic models based on such maps would be greatly facilitated by independent validation metrics for assessing the fit of the models to the data. We describe such a metric based on the fit of atomic models with independent test maps from single particle reconstructions not used in model refinement. The metric provides a means to determine the proper balance between the fit to the density and model energy and stereochemistry during refinement, and is likely to be useful in determining values of model building and refinement metaparameters quite generally.

**Keywords:** cryo-electron microscopy; protein structure; model validation; model refinement

### Introduction

There has been considerable progress over the past decade in the generation of subnanometer cryo-electron microscopy (cryo-EM) density maps of protein complexes,<sup>1</sup> with an increasing number of maps sufficiently resolved to recover a backbone trace.<sup>2–4</sup> In parallel, computational methods have been

developed to generate and refine models using subnanometer resolution cryo-EM density.<sup>5–8</sup> However, while there are a number of ways of assessing the reliability<sup>9</sup> and resolution<sup>10,11</sup> of single-particle reconstructions, there is no standard method by which models built and refined into such reconstructions may be independently validated. Recent studies (e.g.<sup>12</sup>) use some combination of density correlation and protein structure geometry<sup>13</sup> to assess model accuracy.

Validation criteria are very important both for assessing the quality of structure models and for guiding the modeling process itself as illustrated by the extensive use of the free R factor to assess and guide crystallographic model refinement.<sup>14</sup> Indeed, development of independent validation criteria is essential to progress in building detailed atomic models based on cryo-EM data using methods such

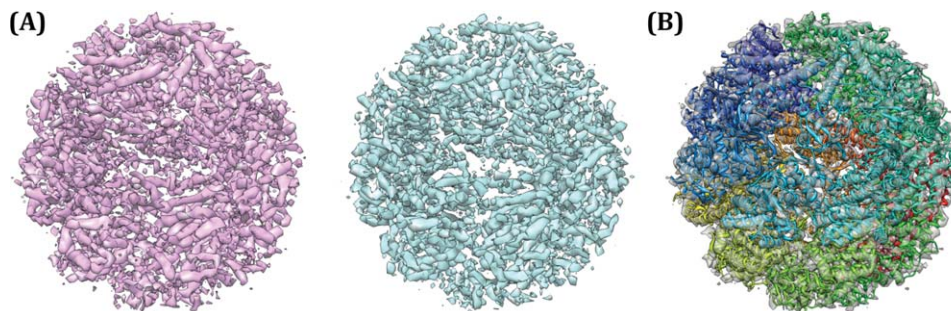
---

*Abbreviations:* cryo-EM, cryo-electron microscopy; FSC, Fourier shell correlation.

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: NIH; Grant numbers: P41GM103832, PN2EY016525, R01GM079429, and R01GM092802.

\*Correspondence to: David Baker, Department of Biochemistry, University of Washington, Seattle, WA 98195.  
E-mail: dabaker@u.washington.edu



**Figure 1.** A: Independent cryo-EM reconstructions of Mm-Cpn. The training map is in magenta and the testing map in cyan. B: A model refined into the training set density (at  $w_a = 0.1$ ).

as Rosetta.<sup>7</sup> Decisions that arise during modeling such as when to incorporate explicit side-chain representations, the extent of flexibility to allow for backbone and side-chain conformations, the extent of refinement against the density, and the balance between the density and the physical chemistry implicit in the force field used in modeling all require an independent measure of model quality. The pitfall in using the fit of a model to the density data used to guide model building as a validation criterion is that the higher the weight on the density term the better the overall fit.

In this paper, we explore the use of independent reconstructions for model validation. We split a large set of particle images into two independent sets and build density reconstructions from each sets. One of these maps—the “training map”—is used for model building and refinement, and the other—the “testing map”—is used for cross validation.

## Results

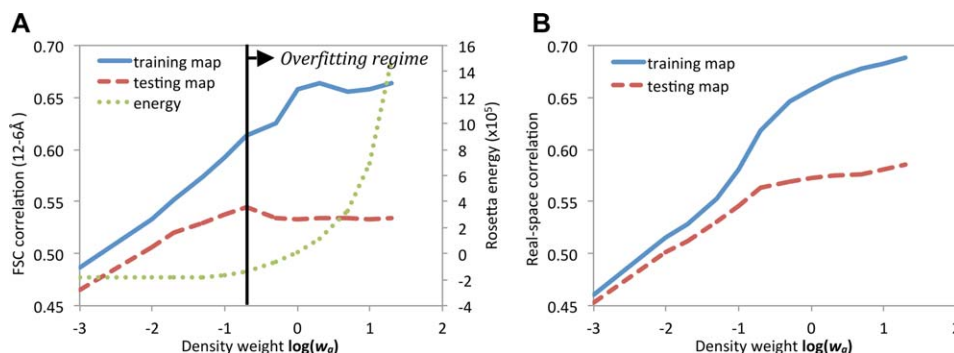
To evaluate this approach to cross validation, we focused on the single-particle reconstruction of wild-type Mm-cpn in the ATP/AlFx induced closed state. The 616 CCD frames used in the original 4.3 Å resolution reconstruction<sup>2</sup> were divided into two separate image data sets. The contrast transfer function parameters were independently determined for each frame. The training map was reconstructed using 22,571 particles boxed out from one set of 308 CCD frames while the testing map was reconstructed using 22,446 particles boxed out from another set of 308 CCD frames. Two different initial density models, generated independently from program **startc-sym** in EMAN1<sup>15</sup> were used to initiate the refinement. While the resultant independently reconstructed maps appear to have consistent backbone connectivity, some of the protruding densities which likely correspond to sidechains differ, probably due to the limited map resolution exacerbated by use of only half of the data-set (Supporting Information Movie 1). One of these reconstructions, the training map, was used for model refinement, while

the other, the testing map, was held aside and only used after refinement for model validation (analogous to the crystallographic  $R_{\text{free}}$  reflections<sup>16</sup>).

We used the Rosetta structure prediction methodology to build and refine models into the training map by adding a score term to the Rosetta energy function assessing agreement between model and map.<sup>7</sup> A weighing term  $w_a$  controls the contribution of the experimental data relative to the Rosetta all atom energy. Models were built from the homologous thermosome KS-1 structure (PDB accession code 1Q3Q), and refined into the training set density using 12 different values of  $w_a$ . The Rosetta symmetric refinement protocol<sup>17</sup> was used to model the entire D8 symmetric complex [Fig. 1(B)]. All modeling used a voxel spacing of 1.30 Å, which is different from the previous value of 1.33 Å.<sup>2</sup> This recalibration reduced the Rosetta all atom energy and improved the fit of models to both the testing and training maps.

The Rosetta all atom energy of the models is shown as a function of  $w_a$  in the green dotted line in Figure 2. The energy of the models is roughly constant as  $w_a$  increases to 0.2, and then rises steeply. The steep increase suggests that models above this point may be overfit to the training map. Molprobit validation<sup>13</sup> (not shown) similarly suggests a sharp decrease in model quality above  $w_a \approx 0.2$ .

The fit of the refined models to the training map and the independent testing map was assessed using the Fourier shell correlation (FSC) in the highest resolution shells (12–6 Å). As expected, the fit of the refined models to the training map increases with increasing  $w_a$  during model refinement [Fig. 2(A), blue solid line]. The fit of the refined models to the testing map [Fig. 2(A), red dashed line] increases in parallel with the fit to the training map for  $w_a < 0.1$ . However, above 0.1–0.2 there is an “inflection point,” where the fit of the refined models to the training map continues to increase but the fit to the testing map decreases, suggesting overfitting to the training map. The overfitting is not as pronounced when the density



**Figure 2.** A: Assessment of the extent of model overfitting with increasing  $w_a$ . Solid line: average model Fourier shell correlation (12–6 Å) with the training map; dashed line: correlation with testing map; dotted line: Rosetta all atom energy. The black vertical line indicates “inflection point” around  $w = 0.1$ – $0.2$ , above which the training map correlation increases, the testing map correlation decreases, and the Rosetta energy increases markedly, suggesting overfitting of models to noise in the data. B: The same data plotted using real-space correlation instead; above the inflection point the correlation with the testing map continues to increase, albeit much more slowly than the correlation with the training map. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

correlation rather than the FSC is used to assess the fit between the models and maps [Fig. 2(B)]; while a similar inflection point is observed around  $w_a = 0.1$ – $0.2$ , the agreement to the testing map continues to increase, albeit slowly, as  $w_a$  increases. The individual model-versus-map FSC curves (Supporting Information Fig. 2) suggest that this residual correlation arises from the lowest resolution Fourier shells—in these shells, the models agree less well with either dataset than the datasets agree with one another. This may be due to missing features in modeling, such as bulk solvent.

It is notable that the steep increase in Rosetta energy, and maximal agreement to the independent test map, both occur in the same  $w_a$  range (from 0.1 to 0.2). Taken together, the increase in fit to the testing map and the lack of increase in Rosetta energy as  $w_a$  varies from 0 to 0.1 suggests that model accuracy increases with increasing  $w_a$  over this range. Above  $w_a = 0.2$ , the sharp increase in Rosetta energy and the decreasing agreement to the testing map suggest that model quality decreases in this region due to overfitting to the training map.

The two independent maps also provide insight into the issue of resolution assessment in cryo-EM maps. The resolution of a cryo-EM map has been commonly estimated from the Fourier Shell Correlation (FSC) between maps computed from split odd and even sets of image data during the last iteration of map refinement. Using this criterion, the resolutions of each of the two reconstructed density maps [Fig. 1(A)] were assessed to be 4.6 Å and 4.7 Å, respectively, at a 0.5 threshold (Supporting Information Fig. 1). However, the FSC between the maps independently reconstructed from two halves of data split from the onset of the refinement indicates a resolution of only 6.7 Å, at 0.143 threshold<sup>10</sup> (Supporting Information Fig. 4). The actual resolution is

likely somewhere between these two values: the former estimate is likely overly optimistic<sup>18</sup> and the latter estimate is overly pessimistic. The resolvability of the Mm-cpn map using the entire data set is adequate to resolve  $\beta$ -strands in the equatorial domains of the subunit and to trace the polypeptide backbone.<sup>2</sup> The low value of the FSC correlation between the independent maps may reflect significant structural variation in the apical domains and loop regions (Supporting Information Fig. 5). Development of a more informative resolution definition for cryo-EM maps remains an open area of investigation.

## Discussion

The development of metrics for cryo-EM model validation will be a critical step towards robustly generating high accuracy all-atom models from cryo-EM density. The split dataset approach explored here provides a metric that should be immediately useful: the Fourier correlation between the refined model and an independently reconstructed density map over shells with resolution higher than 12 Å. This metric can be used to determine the balance between the fit to the map and model energy and stereochemistry during refinement, as described in this paper, as well as any other metaparameter(s) related to model building and refinement.

The overall density correlation is less sensitive to overfitting than the FSC estimate at subnanometer resolution range because at very low resolution the training and test maps are more similar to each other than to the models. The source of the discrepancy between model and map at low resolution is not clear; it could reflect inaccuracies in the physical model (failure to treat model heterogeneity, bulk solvent effects, etc.), incorrect modeling of the contrast transfer function, or biases in the reconstruction

process itself. Whatever the source, the problem can be avoided by discarding the lowest resolution shell during cross validation.

Splitting of the data-set into two halves inevitably limits the resolution and the quality of the reconstructed maps; this problem can be overcome by using the split data set approach for weight and other modeling parameter selection, after which a final refinement could be carried out using all of the data. It also might be possible to use individual particle images directly for validation provided alignment of the images to the model can be carried out sufficiently rapidly and accurately; this could require much less independent data and would have the added benefit of reducing possible bias introduced in the reconstruction process. Such an approach may be applicable for large virus particles, which may have sufficient inherent contrast for individual particle orientation and conformation assessment.

### Acknowledgments

The authors thank Drs. Steve J. Ludtke, Michael F. Schmid, and Matthew L. Baker for discussions, and Cathy Lawson for her comments on the paper. They also thank an anonymous reviewer for valuable suggestions.

The cryo-EM maps have been deposited to EMDB (accession numbers EMD-5645 and EMD-5646). The model was deposited to PDB (id: 3j3x).

### References

1. Chiu W, Baker ML, Jiang W, Dougherty M, Schmid MF (2005) Electron cryomicroscopy of biological machines at subnanometer resolution. *Structure* 13:363–372.
2. Zhang J, Baker ML, Schröder GF, Douglas NR, Reissmann S, Jakana J, Dougherty M, Fu CJ, Levitt M, Ludtke SJ, Frydman J, Chiu W (2010) Mechanism of folding chamber closure in a group II chaperonin. *Nature* 463:379–383.
3. Zhang X, Settembre E, Xu C, Dormitzer PR, Bellamy R, Harrison SC, Grigorieff N (2008) Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc Natl Acad Sci U S A* 105:1867–1872.
4. Zhang X, Jin L, Fang Q, Hui WH, and Zhou ZH (2010) 3.3 Å cryo-EM structure of a nonenveloped virus

- reveals a priming mechanism for cell entry. *Cell* 141:472–482.
5. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* 16:673–683.
6. Schröder GF, Brunger AT, Levitt M (2007) Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure* 15:1630–1641.
7. DiMaio F, Tyka MD, Baker ML, Chiu W, Baker D (2009) Refinement of protein structures into low-resolution density maps using rosetta. *J Mol Biol* 392:181–190.
8. Baker ML, Abeyasinghe SS, Schuh S, Coleman RA, Abrams A, Marsh MP, Hryc CF, Ruths T, Chiu W, Ju T (2011) Modeling protein structure at near atomic resolutions with Gorgon. *J Struct Biol* 174:360–373.
9. van Heel M, Schatz M (2005) Fourier shell correlation threshold criteria. *J Struct Biol* 151:250–262.
10. Rosenthal PB, Henderson R (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* 333:721–745.
11. Henderson R, Chen S, Chen JZ, Grigorieff N, Passmore LA, Ciccarelli L, Rubinstein JL, Crowther RA, Stewart PL, Rosenthal PB (2011) Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy. *J Mol Biol* 413:1028–1046.
12. Baker ML, Zhang J, Ludtke SJ, Chiu W (2010) Cryo-EM of macromolecular assemblies at near-atomic resolution. *Nature Protoc* 5:1697–1708.
13. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst D* 66:12–21.
14. Read RJ, Adams PD, Arendall WB, Brunger AT, Emsley P, Joosten RP, Kleywegt GJ, Krissinel EB, Lutheke T, Otwinowski Z, Perrakis A, Richardson JS, Sheffler WH, Smith JL, Tickle IJ, Vriend G, Zwart PH (2011) A new generation of crystallographic validation tools for the protein data bank. *Structure* 19:1395–1412.
15. Ludtke SJ, Baldwin PR, Chiu W (1999) EMAN: semi-automated software for high-resolution single-particle reconstructions. *J Struct Biol* 128:82–97.
16. Brünger AT (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355:472–475.
17. DiMaio F, Leaver-Fay A, Bradley P, Baker D, André I (2011) Modeling symmetric macromolecular structures in Rosetta3. *PLoS One* 6:e20450.
18. Scheres SH, Chen S (2012) Prevention of overfitting in cryo-EM structure determination. *Nature Meth* 9:853–854.