Published in final edited form as:

J Struct Biol. 2018 November ; 204(2): 276–282. doi:10.1016/j.jsb.2018.09.004.

Assessing the quality of single particle reconstructions by atomic model building

Joshua H. Mendez^a and Scott M. Stagg^{b,c,*}

^aDepartment of Physics, 77 Chieftan Way, Tallahassee, FL 32306, United States

^bInstitute of Molecular Biophysics, 91 Chieftain Way, Florida State University, Tallahassee, FL 32306, United States

^cDepartment of Chemistry and Biochemistry, 95 Chieftain Way, Florida State University, Tallahassee, FL 32306, United States

Abstract

The 2015/2016 Map Challenge challenged cryo-EM practitioners to process a series of publicly available cryoEM datasets. As part of the challenge, metrics needed to be developed to assess and compare the quality of the different map submissions. The most common metric for assessing maps is determining the resolution by Fourier shell correlation (FSC), but there are well known instances where the resolution can be misleading. In this manuscript, we present a new approach for assessing the quality of a map by determining the map “modelability” rather than on resolution. We used the automated map tracing and modeling algorithms in Rosetta to generate populations of models, and then compared the populations between different map entries by the Rosetta score, RMSD to a reference model provided by the map challenge, and by pair-wise RMSDs between different models in the population. These metrics were used to determine statistically significant rankings for the map challengers for each dataset. The rankings revealed inconsistencies between the resolution by FSC, emphasized the importance of the interplay between number of particles contributing to a map and map quality, and revealed the importance of software familiarity on single particle reconstruction results. However, because multiple variables changed between map entries, it was challenging to derive best practices from the map challenge results.

Keywords

cryo-EM map assessment; FSC; Map challenge; Best practices

Introduction

The 3DEM map challenge invited “challengers” to process several public cryo-EM datasets with the hopes of 1) establishing a benchmark set of datasets suitable for high resolution

*Corresponding author at: Institute of Molecular Biophysics, 91 Chieftain Way, Florida State University, Tallahassee, FL 32306, United States, sstagg@fsu.edu (S.M. Stagg).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jsb.2018.09.004>.

cryoEM, 2) encouraging developers and users of 3DEM software packages to analyze these datasets and come up with best practices, 3) evolve criteria for evaluation and validation of the results of the reconstruction and analysis, and 4) compare and contrast the various reconstruction approaches to achieve high efficiency and accuracy. Along with such a map challenge comes a need for metrics for comparing and assessing the resolution and quality of the submitted maps. The most common metric for assessing cryo-EM maps is the Fourier shell correlation method (FSC) (reviewed in (Sorzano et al., 2017)) where the particles contributing to a single particle reconstruction are split into two halves, reconstructed, and then the reconstructions compared in increasing frequency shells in Fourier space. The resolution for the overall reconstruction is determined as the frequency the Fourier shell correlation falls below some threshold, and many different cutoffs have been proposed including 0.5 (Harauz and van Heel, 1986) 0.143 (Rosenthal and Henderson, 2003), and a moving threshold based on information theory (van Heel and Schatz, 2005). Using FSC to assess map challengers is problematic because FSC doesn't necessarily measure map quality. For instance, it is possible that for any two maps, one might have a higher $FSC_{0.5}$ but a lower $FSC_{0.143}$. The map with higher $FSC_{0.5}$ may be higher quality and biological interpretability but be lower resolution by the $FSC_{0.143}$ metric. Another challenge with FSC-based assessment is that FSC values can be artificially inflated by overly aggressive masking or overfitting due to the alignment of noise during the single particle refinement (Scheres and Chen, 2012; Sousa and Grigorieff, 2007). Other metrics for determining resolution include local resolution determination (Kucukelbir et al., 2014) and Fourier neighbor correlation (Sousa and Grigorieff, 2007), but these methods can also be influenced by masking and/or correlated noise.

Here we developed an approach for comparing the quality of different reconstructions that does not rely on the FSC. The idea behind our analysis is that the cryo-EM map quality metrics that have been developed so-far are mostly directed at assessing the resolution, however the driving force for determining structures is what biological interpretations can be derived from the map. Most of the maps in the map challenge were determined to high enough resolution to build atomic models, so our metric for model quality was not resolution, but rather, how well atomic models could be built from the maps. We call this the 'model-ability' of a map. In our approach, we generated families of models automatically using the EM modeling tools developed in Rosetta (DiMaio et al., 2015). We then used two metrics for assessing map quality; how closely the atomic models derived from the map match the published atomic structure and are scored in Rosetta, and how closely the atomic models in the family of models resemble each other. This is similar to a method that was proposed for assessing models built from EM maps (Herzik et al., 2017), but that method depended on a model being already available instead of building models ab initio from the maps like we are doing here.

Our modelability metrics proved quite powerful for comparing and ranking different maps. For instance, even when maps reported similar and/or identical resolutions by FSC-based metrics, our method was capable of discerning subtle but significant differences between maps. Most significantly, our analyses showed the maps that were the most model-able, and thus presumably the most interpretable, were not necessarily the ones with the highest resolutions by FSC. The purpose of this manuscript is to 1) describe our method and validate

it, 2) report our rankings for the different map challengers, and 3) report on whether any best practices could be discerned from the results.

Approach

The Map Challenge 2015/2016 from EMDataBank was composed of seven targets, each with submissions from different researchers. The goal of the assessment stage was to devise a protocol capable of assessing and validating maps. Our protocol is driven by how much can be interpreted from the map, what we call “modelability”, rather than the traditional approach of FSC resolution. Since we are focusing on model-ability, we chose to use the sharpened maps that the challengers deposited for their submissions. Our view was that these represent the maps that would be interpreted biologically and presumably represent the best efforts of the challenger to produce a highly detailed map. We note that the sharpening could certainly influence the quality of the modeling, but we viewed our role as assessors to focus only on what the challengers submitted, so we processed the user-submitted sharpened maps.

The protocol for automated model building was divided in three major steps: initial model building, loop extension & refinement, and model comparison. The first step, initial model building, starts by aligning each map from a specific target. This was done by loading each map into Chimera (Pettersen et al., 2004) for an initial visual inspection. The map which visually appeared to be the best quality was selected as the reference map. Subsequently, each map was automatically aligned using the Chimera tool fit in map. Modeling the entire atomic structures of all maps would be computationally untenable, so only representative segments of the maps were modeled. In order to segment the same region from all maps, the reference pdb model that was provided as part of the Map Challenge was loaded and aligned to the maps using the same procedure as the maps, first manual and then using fit in map. The next step was to identify a section of the map that would be a good representation of the entire map. For the symmetrical specimens, the extracted region represented one asymmetric unit. For the ribosome, we chose the protein uL15 to be representative of the entire map quality because it had regions on the outside of the ribosome and loops that extended into the interior. Once the representative segment was identified for the individual maps, the segment was extracted and was low pass filtered until the surface was smooth with no discernable features. Then the region was expanded by five shells with Gaussian decay and converted to a mask using *e2proc3d* from the software package EMAN2 (Tang et al., 2007). This mask was then used for the extraction of the selected region for each map without introducing hard edges or artifacts in the extracted map. Next, an initial atomic model was created using from the extracted segment using the default parameters of the Rosetta protocol *denovo_density* and the amino acid sequence. If the initial round of *denovo_density* did not cover seventy percent of the extracted map more rounds were performed using the previous initial model as the input for the new round until enough coverage was achieved or no improvement in the initial model was observed; in our case we had a limit of 5 extra rounds. For the GroEL maps, none of the maps produced realistic models, so this target was dropped from further analysis.

For the second step, the amino acid sequence, the extracted map, and the initial atomic model were used as inputs for the Rosetta function `rosetta_scripts` with the default parameters for loop extension and refinement. During this process, Rosetta completed the coverage of the extracted map while optimizing the structure simultaneously. We used this strategy to generate two thousand atomic models for each map in an attempt to populate as many conformations as were allowed by the given map.

In the third and final step, models were scored by two different metrics called Combined Score and Internal root mean square (RMSD) that assessed how consistently Rosetta was able to generate models from the given EM maps. The Internal RMSD evaluation was generated by calculating the pair-wise RMSDs between each generated atomic model for a given map, while the Combined Score was generated by dividing each atomic model's Rosetta energy score by its RMSD to the reference atomic model provided by the Map Challenge. The aim with the Internal RMSD was to score entries using no other information than how well the given map constrained the automatic modeling. The Combined Score combined two external pieces of information to generate a score: the RMSD to the Map Challenge reference structure and the Rosetta energy score. Assuming the Map Challenge reference structure is the ground truth, then models with a low RMSD would be the best, but it is possible that the reference structures were not necessarily the best that could be generated from the cryo-EM data and that Rosetta might do a better job. Presumably the best structure would have the lowest Rosetta energy, but then again this depends on the parameterization of the Rosetta force field and the relative weights of the map (Fig. 1A). Thus, the Combined Score combines both pieces of data by dividing a model's Rosetta score by its RMSD to the reference model. Assuming Rosetta was able to completely and correctly model a given map, then the Internal RMSD should be the best representation of the quality of a map and should agree with the Combined Score, but in cases where Rosetta was unable to completely and accurately model a map or was systematically off in its modeling, then the Combined Score is more reliable. For this reason, we ranked maps based on both scores. The score distributions were plotted using box and whisker plots to visualize the distribution of values (Fig. 1B,C). Good quality maps have a mean Internal RMSD value near zero and a Combined Score as negative as possible while maintaining the smallest possible spread (Fig. 1).

Finally, the mean Internal RMSD and Combined Scores were used to generate challenger rankings for each Map Challenge dataset (Fig. 1D). In order to test whether the differences between the means were statistically significant, Kruskal-Wallis one-way Analysis of Variance (Kruskal-Wallis) test (Kruskal and Wallis, 1952) and Dunn's Pairwise Comparison (Dunn's test) (Dunn, 1961) were performed. In the Kruskal-Wallis test, if the null hypothesis is rejected, at least one of the map's mean is different from the rest. To identify which differences were statistically significant, a post-hoc Dunn's test was used. Dunn's test is a pairwise comparison that identifies if the mean difference of a pair is significant at a specific p-value. In our case, all targets showed to have at least one map to be significant, which led us to proceed with the Dunn's test with a p-value of 0.05.

The full results and rankings for each map challenge sample, Rosetta score vs. RMSD plots, Combined Score distributions, Internal RMSD distributions, and statistical testing results for each map challenger that we analyzed are found in the Supplemental Information.

Results and discussion

Using the methods described above, we have ranked the submissions to the 3DEM Map Challenge by their mean Combined Score and their mean Internal RMSD (Supplemental Information). It should be noted that our method only worked when the resolution of the entries was high-enough to be automatically modeled. We attempted to assess all 66 Map Challenge entries, but some entries like the GroEL ones could not be modeled at all because they were too low in resolution. Another limitation was the large amount of computation power needed, requiring extensive use of a computer cluster and over 2,000,000 CPU hours (228 CPU years) for all models combined. Nonetheless, we were able to produce high-quality rankings for the remaining 57 maps.

Metric validation

First, we set out to assess the validity of our metrics. If the model-ability rankings are correct, then they should correlate with map details as assessed by eye. A good example of this is found in the β -galactosidase entries. Comparing emcd138 (1st by our metrics), emcd106 (tied for 2nd), emcd154 (4th), and emcd113 (5th) (Fig. 1D), there was a clear trend in the observable details with the top scorer having the clearest side chain densities and the last having the least clear (Fig. 2A).

Interestingly, in the β -galactosidase entries, the top scorer did not have the highest resolution as reported by FSC. There are many other examples of this phenomenon in the map challenge. For instance, among the apoferritin entries, the top Internal RMSD scorer had a resolution of 3.5 Å, while the 2nd ranked entry had a resolution of 3.1 Å (Fig. 2C). Similarly, the top ranked TRPV1 entry had a resolution of 3.3 Å while the number two entry had a resolution of 3.1 Å (Fig. 2D). In each case, when we inspected the top scorers visually, the higher ranked entries had better-resolved details than the lower ranked ones. Another interesting observation from the β -galactosidase entries was that by FSC alone, the top scorer by model-ability, which also has more features by visual inspection, would have been tied for third with 5 other entries with 3.3 Å reported resolution, emcd139, 134, 167, 159, and 164 (note: maps 159 and 164 appeared to be duplicate entries). This suggests that FSC alone isn't always the best way to compare maps. Along those lines, there was a surprisingly broad range in quality among the 3.3 Å maps. When sorted by model-ability, there was a clear trend in the appearance of high-resolution features in the maps (Fig. 2B). Maps emcd159 (and its duplicate 164) and 167 couldn't be modeled despite their relatively high-resolution. Inspection of the maps shows that they have poorly resolved features (Fig. 2B). It is unclear why there was a disconnect between the resolution and the features for those maps.

There were a few cases where the model-ability metrics were misleading. For instance, in the Brome Mosaic Virus (BMV) submissions, emcd110 was the top scorer by Internal RMSD (Fig. 3B) but was next to the worst scorer by Combined Score (Fig. 3A).

Comparison of Rosetta models generated for that entry to the map challenge reference model revealed that Rosetta frequently mistraced the density (Fig. 3C) for that map. That likely contributed to its inconsistent Internal RMSD score, especially if Rosetta mistraced it in similar ways for different models. Nonetheless, examination of the distribution of Internal RMSD values revealed that emcd110 had a long tail trending towards large RMSDs which is a clue that Rosetta was prone to mistracing that map (Fig. 3B). Another challenge for our approach was large multi-component particles like the ribosome. It was unrealistic to model the entire ribosome, so we chose to focus only on ribosomal protein uL15. The extended nature of that protein proved to be challenging for Rosetta, and it frequently mistraced it. Nonetheless, a trend could be observed in the detail of the maps that generally matched the modelability scores. The top 4 scorers by Internal RMSD in order were emcd123, 125, 114, and 149, and that trend was generally observed in the visual inspection of extracted uL15 density (Fig. 3F) except for emcd125, which had the poorest details of the four (Fig. 3F). Moreover, the resolution for reconstructions of the ribosome are usually quite anisotropic and could be strongly influenced by the custom mask used for refinement. To fairly compare the different maps, one would need to assess multiple different areas for the different maps.

Observations from the map challenge

Examining the parameters different challengers used for processing their data, it was clear that some software packages were overrepresented compared to others. In the case of frame alignment, Motioncorr (Li et al., 2013) was used in 55% of the submissions we evaluated. This may be due to its speed due to its GPU acceleration coupled with the fact that it generates an overall alignment without having to do per-particle alignment. Similarly, the CTF parameter estimation the software packages CTFFIND 3/4 (Mindell and Grigorieff, 2003; Rohou and Grigorieff, 2015) were strongly preferred, with a usage of around 66%. Lastly, Relion (Scheres, 2012) was also strongly preferred in the last steps of map reconstruction (ie. 2D/3D classification and map refinement). It was used in 64% of the entries at the refinement step.

Influence of number of particles on modelability

Despite the overrepresentation of certain software packages, it was difficult to compare different processing strategies because multiple variables were changed from map to map in the different submissions. This made it difficult to discern any “best practices” from the assessments. A possible exception to this was found in the ribosome submissions. In the 80S Ribosome submissions emcd123, 126, 127, and 128 all particles used in each submission originated from the same dataset and were processed similarly except for the 3D classification step, where they were sorted by 3D classification and further refined independently. Thus, the only variable to change between the maps was the number of particles contributing to the final reconstructions. The reported resolution of each map increased with the number of particles, and the highest resolution map, emcd123, was also the top ranked by model-ability (Fig. 4). As previously stated, the resolution, anisotropy, and size of the ribosome made overall assessment challenging for our method. Nonetheless, trends could be observed when the Rosetta Score vs. RMSD plots were analyzed for emcd123, 126, 127, and 128 (Fig. 4B). Interestingly, the plots indicate that emcd127 is nearly equivalent to emcd126 with despite having only half the particles. Though emcd127

has fewer particles, the overall quality of that subset of particles is likely better than the particles contributing to emcd126. This effect is shown in another way in map emcd128, which has a similar number of particles to emcd127 but is inferior in terms of resolution and modelability (Fig. 4). Again, this likely indicates that the quality of particles contributing to emcd128 is lower than 127, but it is unclear what makes some particles better than others with these data. In future map challenges, we are hopeful that more variables can be controlled in the data provided to the challengers, such as providing challengers with fixed CTF, particle coordinates, and aligned compensated movies so that it may be possible to deconvolute the contributions of all those variables when assessing map challengers.

User familiarity

Another phenomenon we noticed was that familiarity with the software processing packages appeared to have a large effect on the quality of the deposited maps. An example of this can be seen in the apoferritin submissions emcd124 and emcd147. The entire data analysis was done identically for those entries with the exception of map refinement; emcd124 was processed in Relion while emcd147 was processed in Frealign. As seen in Fig. 5, emcd147 performed worse than its counterpart, and this is also reflected on their reported resolutions of 6.8 Å and 4.8 Å respectively. Emcd147 was refined in Frealign with a global search of the particle orientation (Mode 3) while neglecting the local search (Mode 1), while a more experienced user might choose a different approach. Indeed, it is clear that Frealign was capable of producing very high-quality maps for the apoferritin dataset as emcd121, which was processed in Frealign by the author of that software, was one of the highest ranked entries by our metrics (Fig. 5). This indicates that there are some best practices for processing in Frealign (and by extension, most software packages), but it is unclear from the map challenge data what those may be. It is noted, however, that the Frealign algorithms have been incorporated into a new software package called cisTEM (Grant et al., 2018) that has a new automated processing workflow that may incorporate the best practices that the Frealign authors have determined.

Influence of sharpening on reconstruction quality

In assessing the map challengers, we performed all of the modeling on the challenger-supplied sharpened maps because we were interested in assessing what were presumably the challenger's best efforts at generating high-quality maps. Nonetheless, sharpening could affect the results of the modeling, so we tested the effects of different sharpening procedures on emcd 166, which is an apoferritin challenge submission from our group. The submitted raw map for emcd 166 was sharpened using Bsoft, Relion, or cisTEM. The resulting sharpened maps were then assessed for their model-ability as we described above. This showed that indeed, the sharpening can have a distinct effect on map model-ability (Fig. 6). Compared to the submitted sharpened map (Fig. 6, left), sharpening with different software tools made the model-ability slightly worse or slightly better depending on the software package. Of note, the submitted emcd 166 map was ranked third by our metrics but would have been tied for first if it had been sharpened in a different way. This again demonstrates the importance of establishing best practices for all stages of single particle cryo-EM.

Conclusion

We have demonstrated that assessing models by their model-ability, although computationally expensive, is capable of discerning map differences that FSC isn't able to. By ranking maps according to their Combined Score and Internal RMSD assessments, "best" maps can be identified (Supplemental Information). Using the model-ability rankings as a guide, we were able to make some observations about the processing strategies from the metadata provided by map challengers. One of our most striking observations was that resolution as reported by FSC was often inconsistent with map quality. Among the β -galactosidase entries, five had the same resolution by FSC but differed widely in their model-ability and the visual details apparent in the maps (Fig. 2B). Also, in many cases, the highest resolution maps were not necessarily the highest quality maps (Fig. 2A,C,D). Another observation from our analyses was the importance of the interplay between the number of particles contributing to a map and the particle quality. In the case of the ribosome maps, one challenger submitted multiple maps that were processed identically except for the number of particles contributing to the reconstruction. The top ranked map by our metrics also had the highest resolution and included the most particles (Fig. 4). However, by splitting this into subsets with fewer particles, it was clear that some subsets were better than others (Fig. 4), but in the case of the ribosome data it was better to include more particles than to subclassify into smaller subsets. Finally, user familiarity with the processing software packages had a strong influence on the quality of the submitted maps, and this was particularly apparent in the apoferritin submissions; but with the current map challenge, it was hard to discern best practices from the submitted maps. In most cases, multiple variables changed from entry to entry, and it was unclear which variables among CTF estimation, particle picking, movie alignment, damage compensation or dozens of other variables were the ones that contributed most to the reconstruction quality. With future map challenges, it is hoped that some of these variables can be controlled so that best practices can be determined.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was partially supported by a grant from the NIH, R01GM108753.

References

- DiMaio F, Song Y, Li X, Brunner MJ, Xu C, Conticello V, Egelman E, Marlovits T, Cheng Y, Baker D, 2015 Atomic accuracy models from 4.5 Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat. Methods* 12, 361–365. 10.1038/nmeth.3286. [PubMed: 25707030]
- Dunn OJ, 1961 Multiple comparisons among means. *J. Am. Stat. Assoc* 56, 52–64. 10.2307/2282330.
- Grant T, Rohou A, Grigorieff N, 2018 cisTEM, user-friendly software for single-particle image processing. *Elife* 7, e14874 10.7554/eLife.35383.
- Harauz G, van Heel M, 1986 Exact filters for general geometry three dimensional reconstruction. In: *Proceedings of the IEEE Computer Vision and Pattern Recognition Conf.*, pp. 146–156.

- Herzik MA, Fraser JS, Lander GC, 2017 A multi-model approach to assessing local and global cryo-EM map quality. *bioRxiv*. 10.1101/128561.
- Kruskal WH, Wallis WA, 1952 Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc* 47, 583–621. 10.2307/2280779.
- Kucukelbir A, Sigworth FJ, Tagare HD, 2014 The local resolution of cryo-EM density maps. *Nat. Methods* 11, 63–65. 10.1038/nmeth.2727. [PubMed: 24213166]
- Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, Agard DA, Cheng Y, 2013 Electron counting and beam-induced motion correction enable nearatomic-resolution single-particle cryo-EM. *Nat. Methods* 10, 584–590. 10.1038/nmeth.2472. [PubMed: 23644547]
- Mindell JA, Grigorieff N, 2003 Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol* 142, 334–347. 10.1016/S1047-8477(03)00069-8. [PubMed: 12781660]
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE, 2004 UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem* 25, 1605–1612. 10.1002/jcc.20084. [PubMed: 15264254]
- Rohou A, Grigorieff N, 2015 CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol* 192, 216–221. 10.1016/j.jsb.2015.08.008. [PubMed: 26278980]
- 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. [PubMed: 14568533]
- Scheres SHW, 2012 RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol* 180, 519–530. 10.1016/j.jsb.2012.09.006. [PubMed: 23000701]
- Scheres SHW, Chen S, 2012 Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* 9, 853–854. 10.1038/nmeth.2115. [PubMed: 22842542]
- Sorzano COS, Vargas J, Otón J, Abrishami V, de la Rosa-Trevín JM, GómezBlanco J, Vilas JL, Marabini R, Carazo JM, 2017 A review of resolution measures and related aspects in 3D electron microscopy. *Prog. Biophys. Mol. Biol* 124, 1–30. 10.1016/j.pbiomolbio.2016.09.005. [PubMed: 27666962]
- Sousa D, Grigorieff N, 2007 Ab initio resolution measurement for single particle structures. *J. Struct. Biol* 157, 201–210. 10.1016/j.jsb.2006.08.003. [PubMed: 17029845]
- Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ, 2007EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol* 157, 38–46. 10.1016/j.jsb.2006.05.009. [PubMed: 16859925]
- van Heel M, Schatz M, 2005 Fourier shell correlation threshold criteria. *J. Struct. Biol* 151, 250–262. 10.1016/j.jsb.2005.05.009. [PubMed: 16125414]

Fig. 1. Assessment of Map Challenge β -galactosidase entries by modelability. A) Plot of the Rosetta energy score vs. RMSD to the map challenge reference model for entry emcd138. B) Box-and-whisker plots of Combined Scores for the β -galactosidase entries. C) Box-and-whisker plots of Internal RMSDs for the β -galactosidase entries. D) Modelability rankings for the β -galactosidase entries.

Fig. 2.

Comparison of models, visual details in EM maps, modelability ranking, and $FSC_{0.143}$ resolution. In each panel, the maps are contoured such that they enclose the same volume. In all panels the green asterisk highlights areas where there is a difference in the side chain density. A) β -galactosidase maps in order of their ranking. B) β -galactosidase asymmetric units that have the same resolution by FSC. C) The top two apoferritin entries. D) The top two TRPV1 entries. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 3.

Examples where the model-ability measurements were misleading. A) Combined Score distributions for BMV. B) Internal RMSD distributions for BMV. C) Map challenge reference model for BMV fit into the asymmetric unit density for emcd110. D) An example Rosetta model for emcd110 where the chain is traced incorrectly. In (C), (D), and (E) the models are colored by residue number. E) The location where Rosetta mistraced the density (arrow). F) The top model-ability scorers in the ribosome dataset. Ribosomal protein uL15 is shown. Emcd125 appears misranked when inspecting the map details. In each panel, the maps are contoured such that they enclose the same volume.

Fig. 4. Influence of the number of particles contributing to a map on map quality and model-ability. A) Maps emcd123, 126, 127, and 128. The maps are shown from left to right in order of the number of particles contributing to the maps. In each panel, the maps are contoured such that they enclose the same volume. B) Rosetta score vs. reference model RMSD for the maps in (A).

Fig. 5. Combined Score (A) and Internal RMSD (B) for models that illustrate the influence of user expertise on map quality. Emcd124 and 147 were submissions from the same user. The higher ranked entry, emcd124, was processed in Relion and the emcd147 was processed in Frealign. On the other hand, emcd121, which was one of the top apoferritin submissions, was processed in Frealign.

Fig. 6.

Influence of sharpening on modelability. A) Internal RMSD for emcd 166 comparing the submitted sharpened map to maps sharpened with Bsoft, Relion, or cisTEM. B) Combined score for emcd 166 sharpened as in (A). C) An extracted α -helix from the sharpened maps in (A) and (B).