Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement

Frank DiMaio[1,12], Yifan Song[1,2,12], Xueming Li[3,11], Matthias J Brunner[4–7], Chunfu Xu[8], Vincent Conticello[8], Edward Egelman[9], Thomas C Marlovits[4–7], Yifan Cheng[3] & David Baker[1,10]

We describe a general approach for refining protein structure models on the basis of cryo-electron microscopy maps with near-atomic resolution. The method integrates Monte Carlo sampling with local density-guided optimization, Rosetta all-atom refinement and real-space *B*-factor fitting. In tests on experimental maps of three different systems with 4.5-Å resolution or better, the method consistently produced models with atomic-level accuracy largely independently of starting-model quality, and it outperformed the molecular dynamics–based MDFF method. Cross-validated model quality statistics correlated with model accuracy over the three test systems.

Recent developments in direct electron detectors as well as improved image data analysis have led to vast improvement in the resolution achievable by single-particle cryo-electron microscopy (cryo-EM)[1,2]. Tools for automatic structure determination, model rebuilding, all-atom refinement and model validation are needed for single-particle reconstructions with near-atomic resolution. Currently available X-ray crystallographic tools[3–5] perform relatively poorly with density-map resolutions worse than 3 Å, failing to converge on an accurate atomic model[5]. Methods have been developed specifically for building and refining structures into cryo-EM density maps, including tools to fit crystal structures into density[6–8], to fit secondary-structure elements and assign sequence and then perform all-atom refinement[9,10], and to rebuild missing regions of protein backbone on the basis of experimental density data[11,12]. However, these methods rely on the existence of a high-quality starting model or known secondary-structure assignment.

Here we present a unified approach to model building, refinement and model validation using near-atomic-resolution cryo-EM reconstructions. Starting from homologous structures, and using density maps over a wide range of resolutions, we show that when the resolution is better than 4.5 Å, the approach converges on accurate all-atom models largely independently of starting-model accuracy. The approach could automatically correct sequence registration errors and has a substantially better radius of convergence than the widely used molecular dynamics flexible fitting (MDFF) method[13].

RESULTS

We sought to develop a cryo-EM model-refinement protocol that follows experimental density as much as possible while maintaining the physicochemical accuracy of the model. We integrated approaches from crystallographic refinement, *ab initio* structure prediction, and segment rebuilding and refinement in comparative modeling to enable progression from a poor starting model—having correct topology but large local backbone deviations—to an atomically accurate model in one seamless protocol.

We adapted to density-guided model building a sampling strategy previously developed for comparative modeling[14]. In this strategy, backbone fragments collected from the Protein Data Bank (PDB)[15] are inserted via superposition onto the current model, and Cartesian-space minimization against a low-resolution energy function 'stitches closed' the broken peptide bonds (**Supplementary Fig. 1**). To sample more effectively, before stitching we optimize each fragment to fit the density in the region. A Monte Carlo trajectory is carried out with trial moves consisting of replacement of a region that fits the density poorly with a backbone fragment from the PDB preminimized to fit the density. In the preminimization step, coordinate constraints at the fragment end points maintain proper peptide-bond geometry, and Ramachandran and rotameric constraints maintain reasonable backbone and side-chain geometry. Because the fragments are minimized in isolation, testing of a large set of fragments at each position is computationally feasible, quickly identifying backbone conformations consistent with both local sequence and

---

[1]Department of Biochemistry, University of Washington, Seattle, Washington, USA. [2]Cyrus Biotechnology, Seattle, Washington, USA. [3]Keck Advanced Microscopy Laboratory, Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California, USA. [4]Center for Structural Systems Biology (CSSB), University Medical Center Eppendorf-Hamburg (UKE), Hamburg, Germany. [5]Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany. [6]Institute of Molecular Biotechnology (IMBA), Austrian Academy of Sciences, Vienna, Austria. [7]Research Institute of Molecular Pathology (IMP), Vienna, Austria. [8]Department of Chemistry, Emory University, Atlanta, Georgia, USA. [9]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, Virginia, USA. [10]Howard Hughes Medical Institute, University of Washington, Seattle, Washington, USA. [11]Present address: School of Life Science, Tsinghua University, Beijing, China. [12]These authors contributed equally to this work. Correspondence should be addressed to D.B. (dabaker@u.washington.edu).

RECEIVED 20 FEBRUARY 2014; ACCEPTED 11 DECEMBER 2014; PUBLISHED ONLINE 23 FEBRUARY 2015; DOI:10.1038/NMETH.3286

the experimental data. All steps of modeling take into account the native symmetry of the complex, with all subunits present in multisubunit complexes.

Because cryo-EM maps are frequently better in resolution in some regions than others, atomic *B* factors are fit against cryo-EM density data to maximize the real-space correlation between model and map. The protocol alternates between *B*-factor refinement and model rebuilding until the correlation between the map and the model converges. Finally, because the fit of a model to experimental data alone provides little information on model quality (as the model was refined against that same experimental data), we developed a cross-validation metric using independently collected data that provides a map quality–independent assessment of model accuracy.

Model building and refinement

To evaluate the new refinement protocol and compare it to alternate approaches, we used three recently collected experimental data sets: the 20S proteasome at 3.3-Å resolution[2]; PrgH and PrgK (PrgH/PrgK), the periplasmic domains of the needle complex, at 4.6-Å resolution; and a peptide fiber assembly at 4.3-Å resolution[16]. To test the dependence of the method on the number of images used in the reconstruction and map resolution, we generated reconstructions of 20S at resolutions of 4.1, 4.4, 5.0 and 6.0 Å, and reconstructions of PrgH/PrgK at 5.4 and 7.1 Å, using subsets of the particle images to realistically evaluate the challenges arising from limited data. At the highest resolution, much side-chain density is visible; at 4.0- to 4.5-Å resolution, limited side-chain density is observed, but individual strands and the pitch of helices are visible; and at 5 Å and worse, individual β-strands and the pitch of helices are indistinguishable.

We focused first on the 20S proteasome, as the large amount of collected data and the wide range of structural variation in homologs allows for systematic investigation of the dependence of the method on the resolution of the map and the accuracy of the starting model. We refined the 20S crystal structure (PDB code: 1PMA) into the highest-resolution reconstruction to generate a reference model for comparison; to allow for evaluation of the refined model versus the crystal structure, we first split the particle images into two sets and generated two independent reconstructions. We set atomic *B* factors to uniform and refined coordinates and *B* factors iteratively into the reconstruction. After refinement, the *B* factors fit to the reconstruction showed very good agreement with those of the crystal structure ($R^2 = 0.74$; **Fig. 1a**,**b**). Refinement of the crystal structure led to subtle changes in the backbone in several loops (**Fig. 1c**) that increased agreement with the independent reconstruction, a result suggesting that the refined model represents the structure on the electron microscopy grid more accurately than does the crystal structure.

To explore the effect of starting-model accuracy on model refinement, we constructed comparative models from 11 homologs of 20S with sequence identity ranging from 12% to 40% (PDB codes: 1G0U, 1G3K, 1IRU, 1M4Y, 1Q5Q, 1RYP, 2X3B, 3H4P, 3NZJ, 3UNF and 4HNZ) and another crystal structure with 100% sequence identity (PDB code: 1YAR). Errors in these starting structures are diverse and cover challenges commonly seen in structure refinement, including rigid-body movement of helices and strands, missing residues from the template, changes in loop conformation and misaligned residues. Compared to the reference model, input models had 15–96% of Cα atoms with 1 Å and r.m.s. deviations of 1.0–7.1 Å.

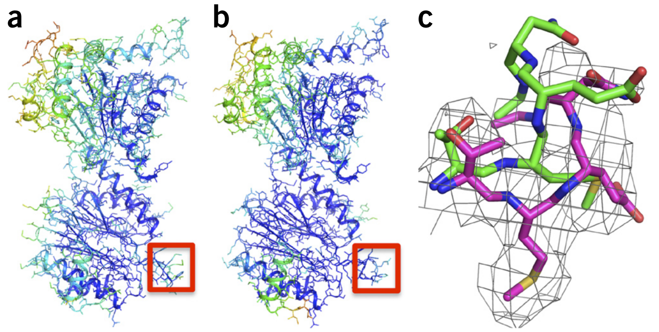

Figure 1 | Refinement of 20S proteasome crystal structure into high-resolution cryo-EM density. (**a**,**b**) The crystal structure (**a**; PDB code: 1PMA) and the cryo-EM model (**b**) refined against the 3.3-Å map. The model is colored by the *B* factor in the crystal structure (**a**) and by the Rosetta real-space *B*-factor fit to the cryo-EM map (**b**). (**c**) Example of a loop region (red box in **a**,**b**) adopting a different conformation in the cryo-EM model: green, crystal structure; magenta, Rosetta-refined model; remaining colors indicate noncarbon atoms. The independent map density (not used in refinement) is shown.

To simultaneously evaluate (i) the dependence of refined-model accuracy on starting-model quality and map resolution and (ii) the accuracy of model-validation metrics, we generated independent training and testing maps at 3.3-, 4.1-, 4.4-, 5.0- and 6.0-Å resolution and refined each of the starting models into each of the training maps (**Fig. 2**). In all calculations, peptide fragments from structures with higher sequence identity than that of the worst starting model (12%) were excluded. For comparison to a widely used current method, we built full-length comparative models from each starting model using Modeller[17] and refined them using the MDFF protocol[13]. As above, the refined models were evaluated by determining the fraction of residues with Cα atoms within 1 Å of those in the refined crystal structure.

Although the accuracy of the starting models fell dramatically with decreasing sequence identity, the accuracy of the refined Rosetta models for the 3.3-, 4.1- and 4.4-Å maps was quite good, even with distant starting models (**Fig. 2a–e** and **Supplementary Fig. 2**). For the majority of input templates, the refined models were over 75% accurate, with errors primarily in surface loops. However, at 5- and 6-Å resolution, the performance of the method was less good. The extensive backbone sampling carried out during refinement is a double-edged sword: it allows dramatic improvement of starting models at high resolution, but it can degrade starting models when the experimental data provide insufficient restraints. The dramatic decrease in performance going from 4.4 to 5.0 (and 6.0) Å may reflect the blurring of β-strands in β-sheets or the difficulty placing Cβ atoms in helices. Because resolution alone does not provide a perfect picture of map quality, this approach should not be used on maps lacking features such as resolution of individual strands or the pitch of helices.

The widely used MDFF method was much less sensitive to the resolution of the density map but more sensitive to the accuracy of the starting model, likely owing to stronger tethering to the starting model; this reduces model degradation at low resolution, but at high resolution makes it difficult to improve distant starting models. To increase the range of motion in MDFF, we tried reducing the restraints to the starting model and incrementally increasing

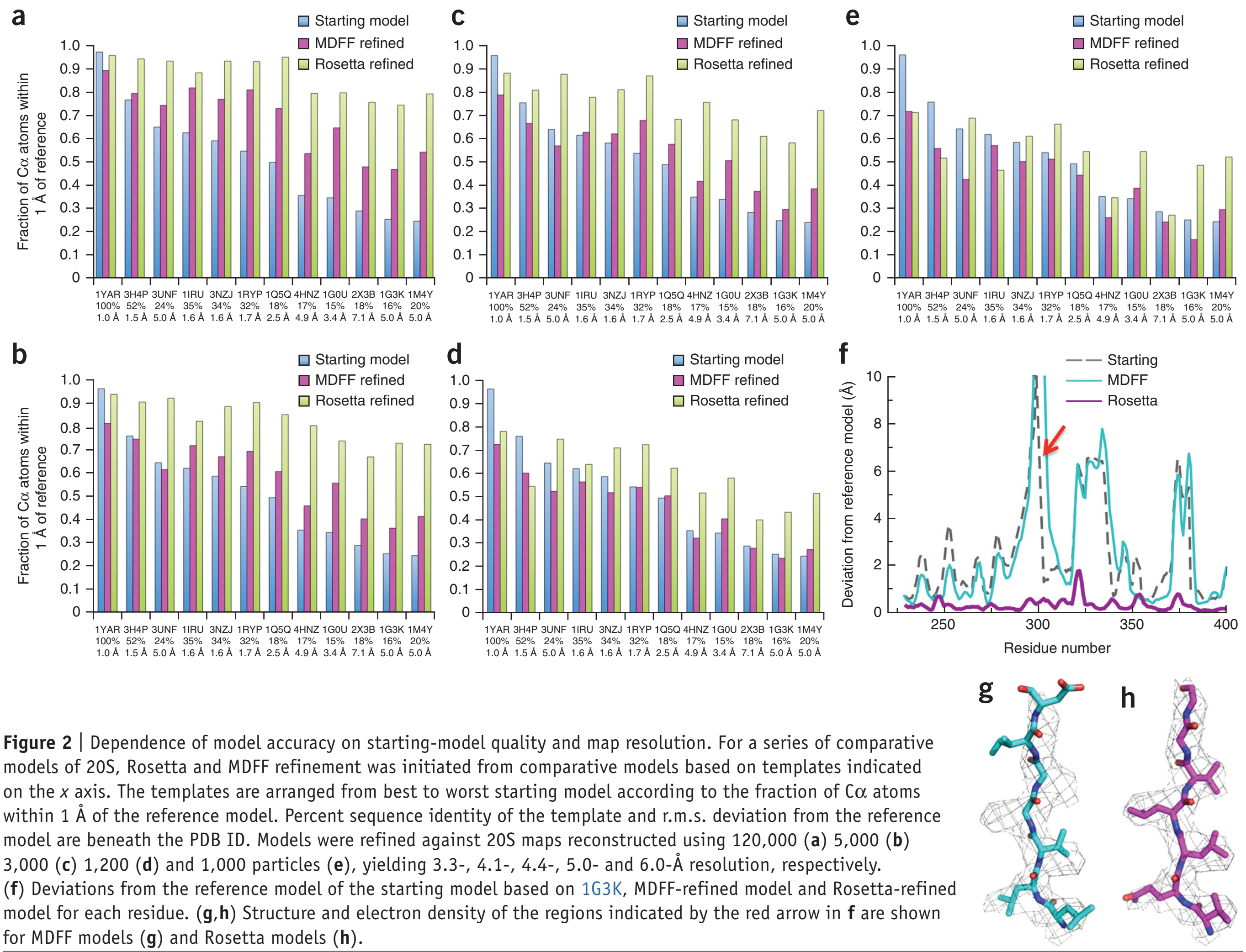

Figure 2 | Dependence of model accuracy on starting-model quality and map resolution. For a series of comparative models of 20S, Rosetta and MDFF refinement was initiated from comparative models based on templates indicated on the *x* axis. The templates are arranged from best to worst starting model according to the fraction of Cα atoms within 1 Å of the reference model. Percent sequence identity of the template and r.m.s. deviation from the reference model are beneath the PDB ID. Models were refined against 20S maps reconstructed using 120,000 (**a**) 5,000 (**b**) 3,000 (**c**) 1,200 (**d**) and 1,000 particles (**e**), yielding 3.3-, 4.1-, 4.4-, 5.0- and 6.0-Å resolution, respectively. (**f**) Deviations from the reference model of the starting model based on 1G3K, MDFF-refined model and Rosetta-refined model for each residue. (**g**,**h**) Structure and electron density of the regions indicated by the red arrow in **f** are shown for MDFF models (**g**) and Rosetta models (**h**).

the density weight. With lower restraint weight and higher density weight, the models moved further to better fit the density, but overall model geometry was compromised as indicated by increased MolProbity[18,19] scores (data not shown). In contrast, models could move substantially in Rosetta to fit the density while maintaining good MolProbity scores. Errors in the starting models are largely retained through MDFF refinement but are often reduced in Rosetta-refined models (**Fig. 2f**). When starting from distant models, Rosetta generated superior-quality models (**Fig. 2a–e**) on all but the 5.0- and 6.0-Å maps.

The Rosetta refinement protocol was able to correct the majority of errors from the input structure for 3.3- to 4.4-Å maps because the rebuilding procedure can quickly overcome local barriers. Density data are used to select and then optimize individual fragments, making backbone conformational sampling focused and efficient. In many cases, the starting models are incorrectly fit into the density with errors in sequence registration and misplaced secondary-structure elements (**Fig. 2g**). With map resolutions better than 4.5 Å, most of these errors were fixed in the Rosetta-refined models (**Fig. 2h**). These errors are not fixed using previous published protocols (**Supplementary Table 1**). On the other hand, with lower-resolution maps, density information was not enough to guide correct placement of fragments, and many incorrect sampled models fit the density equally well.

Model validation

The fit of a refined model to an independent test map provides an unbiased measurement of model quality. We found previously that the medium-resolution Fourier shell correlation (FSC) was a better predictor of model accuracy than real-space correlation[20]. Although the entire model-map FSC curve is informative, when many models are generated, it is valuable to have a single number that reflects model quality; thus, we integrate the FSC over the medium-resolution range. This integrated FSC on an independent map (or free FSC) correlated well with model accuracy, particularly at high resolution (**Fig. 3a–e**). Furthermore, the real-space correlation between models and the independent testing map over segments of the chain correlated with the local accuracies of models. In high-resolution maps, as the local correlation decreased, the fraction of incorrectly modeled residues increased (**Supplementary Fig. 3**). This allowed for the identification of local errors that were not eliminated by the automated protocol.

Testing on other systems

For the PrgH/PrgK ring and the peptide fiber, there were no crystal structures of the complexes in the same multimeric configuration to use as gold standards, so we relied on the FSC against an independent reconstruction to evaluate model accuracy.

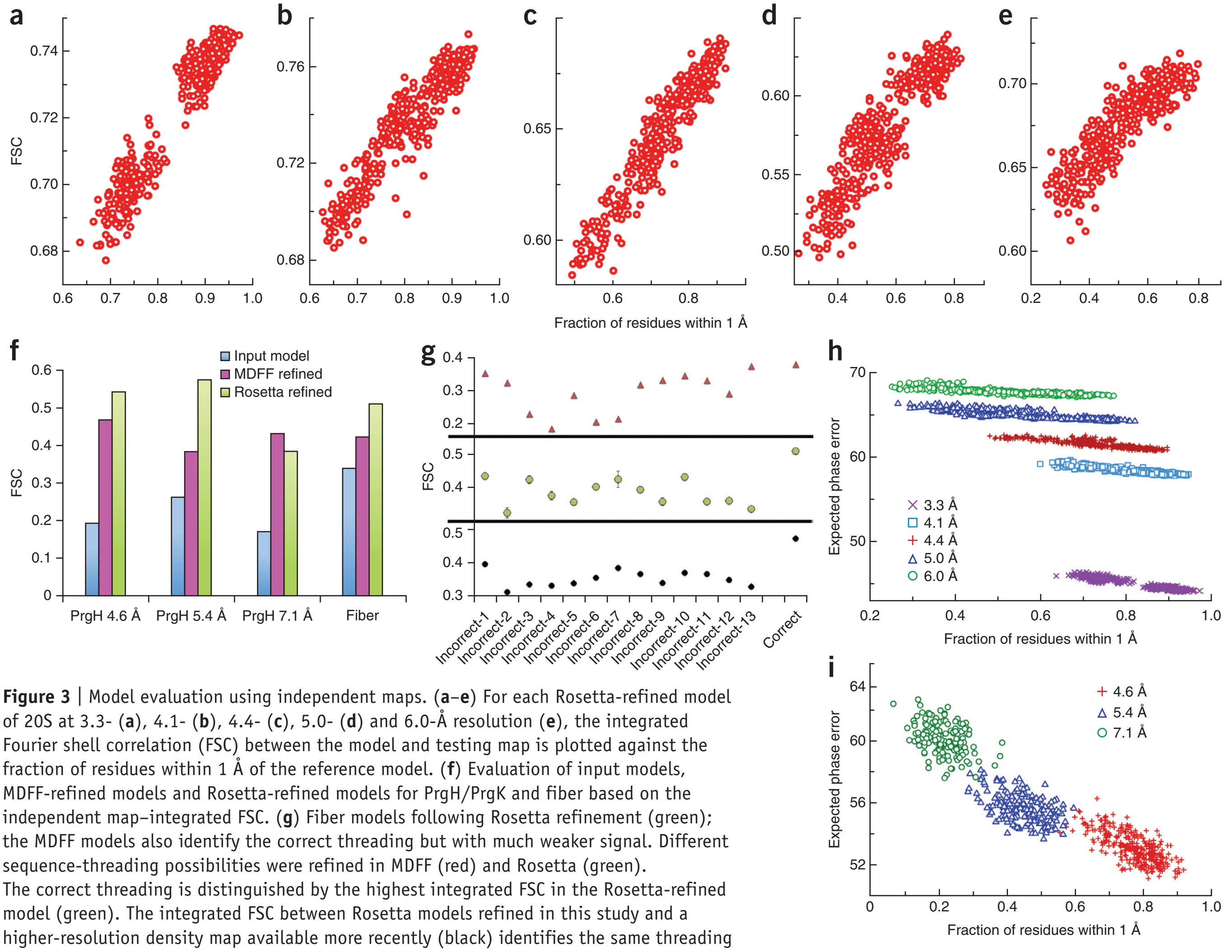

Figure 3 | Model evaluation using independent maps. (**a**–**e**) For each Rosetta-refined model of 20S at 3.3- (**a**), 4.1- (**b**), 4.4- (**c**), 5.0- (**d**) and 6.0-Å resolution (**e**), the integrated Fourier shell correlation (FSC) between the model and testing map is plotted against the fraction of residues within 1 Å of the reference model. (**f**) Evaluation of input models, MDFF-refined models and Rosetta-refined models for PrgH/PrgK and fiber based on the independent map–integrated FSC. (**g**) Fiber models following Rosetta refinement (green); the MDFF models also identify the correct threading but with much weaker signal. Different sequence-threading possibilities were refined in MDFF (red) and Rosetta (green). The correct threading is distinguished by the highest integrated FSC in the Rosetta-refined model (green). The integrated FSC between Rosetta models refined in this study and a higher-resolution density map available more recently (black) identifies the same threading as the lower-resolution map (green). (**h**,**i**) Expected phase error correlates with the accuracies of refined models. Refinement was carried out with reconstructed maps of 20S proteasome (**h**) and with maps of PrgH/PrgK (**i**) at the indicated resolutions. The expected phase error tracks absolute model quality better than does the integrated FSC (**Supplementary Fig. 4**).

The first system, PrgH/PrgK, is a C24 symmetric ring, with data to 4.6-Å resolution (M.J.B., F.D. and T.C.M., unpublished data). We also utilized lower-resolution reconstructions made with subsets of the entire data set, which had estimated resolutions of 5.4 and 7.1 Å. The starting model is a hybrid derived from two sources: one subunit comes from a crystal structure in a different multimeric conformation, and the second subunit comes from a homologous structure. At each resolution, we fit Rosetta and MDFF models against a training map reconstructed from half of the images, and we measured the FSC against a testing map. As with 20S, there were substantially better fits to the independent data with the Rosetta models than with the MDFF models at 4.6- and 5.4-Å resolution, but at 7.1 Å the MDFF-generated model had better FSC than the Rosetta-generated model (**Fig. 3f**).

In the fiber case, the map was of a repeating helical fiber structure. The challenge was not in identifying the backbone conformation but rather determining the orientation and sequence registration of the helix in the two density maps. The 4.3-Å map had a single copy of the helix in the asymmetric unit. Even at this resolution, the nearly palindromic nature of the sequence made sequence registration difficult. Instead of using fragment-based assembly, we enumerated the 14 possible sequence registrations and refined each model. There was a clear signal for one particular sequence registration (**Fig. 3g**), with an independent map–agreement improvement of over 0.02 compared to the next-best registration. With MDFF, the overall independent-map agreement was worse for this registration, and there was little signal for this registration relative to other possible registrations. A more recent higher-resolution reconstruction had strong signal for this registration (**Fig. 3g**), further suggesting that it is correct.

Estimated phase error

Even at high resolution, the integrated model-map FSC, although very effective at evaluating the relative accuracy of multiple models to a single map, does not provide a measure of absolute accuracy of models in different maps (**Fig. 3a–e** and **Supplementary Fig. 4**). FSC also has a number of weaknesses that make it somewhat undesirable as an evaluation metric: (i) different resolution ranges are summed over for different maps, so the values are not comparable, and (ii) it does not take into account the signal-to-noise ratio in each shell, which may vary even in maps at the same resolution. An absolute measure assessing the accuracy of a model in a map is thus desirable.

We sought to develop a likelihood-based measure for evaluating the agreement between model and map that gives reasonable accuracy measures independent of map. As described in the Online Methods, we developed a measure of expected phase error (EPE) in reciprocal space. Although it is not perfect, the EPE is more comparable between different-resolution maps (**Fig. 3h,i**) than is the integrated FSC (**Supplementary Fig. 4**). Obtaining an absolute scale measure of model quality that is less sensitive to noise remains an important area of research.

DISCUSSION

Starting from experimental density maps with 4.5-Å resolution or better, we have shown, for three different systems, that it is possible to consistently generate models with near-atomic-level accuracy. As there is not a standard definition of map resolution, we also provide a more qualitative description of the map quality necessary for our method to be usefully applied: inspection of the maps in our test set (**Supplementary Fig. 5**) suggests that the pitch of helices, individual strands and some large aromatic side chains should be at least in part visible.

By using ideas from crystallographic refinement, such as independent model validation and atomic *B*-factor fitting, we have improved model generation from cryo-EM maps. These methods should prove useful when used in conjunction with automated *de novo* chain tracing methods[21] for fully automated structure determination from cryo-EM density. A next step is to use modeling to reduce map error, as is done in crystallography through map rephasing and density modification. Although single-particle reconstructions contain equally accurate amplitude and phase information, we may still use modeling to reduce errors in the image reconstruction process. For example, using intermediate models rather than heuristic scaling factors (as in ref. 22) to rescale map intensities as a function of resolution should more accurately recapitulate high-resolution details. Models may also be used to reduce errors in determining particle orientation or particle conformation in heterogeneous systems. Such methodological advances could substantially improve the determination of atomic models from cryo-EM reconstructions.

All the refinement tools prevented in this manuscript are freely available for academic use through the Rosetta software suite (weekly releases after 15 February 2015), available at https://www.rosettacommons.org/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. EMDataBank: map of fiber, EMD-6123; lower-resolution maps of 20S, EMD-6245, EMD-6246, EMD-6247 and EMD-6248.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors thank K. Laidig and D. Alonso for setting up and maintaining computational resources. This work was supported by the US National Institutes of Health (NIH) grants R01GM092802 (D.B.), R01GM082893 and R01GM098672 (Y.C.), and EB001567 (E.E.).

AUTHOR CONTRIBUTIONS

F.D. and Y.S. developed the methods and ran experiments; F.D., Y.S. and D.B. wrote the manuscript. X.L. and Y.C. provided the 20S low-resolution data sets and provided feedback on the method. M.J.B. and T.C.M. analyzed the PrgH data set and provided feedback on the method. C.X., V.C. and E.E. collected the fiber data set and provided feedback on the method. All authors helped in editing the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Milazzo, A.C. *et al.* Initial evaluation of a direct detection device detector for single particle cryo-electron microscopy. *J. Struct. Biol.* **176**, 404–408 (2011).
2. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
3. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
4. Langer, G., Cohen, S.X., Lamzin, V.S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179 (2008).
5. Terwilliger, T.C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 61–69 (2008).
6. Tjioe, E., Lasker, K., Webb, B., Wolfson, H.J. & Sali, A. MultiFit: a web server for fitting multiple protein structures into their electron microscopy density map. *Nucleic Acids Res.* **39**, W167–W170 (2011).
7. Woetzel, N., Lindert, S., Stewart, P.L. & Meiler, J. BCL::EM-Fit: rigid body fitting of atomic structures into density maps using geometric hashing and real space refinement. *J. Struct. Biol.* **175**, 264–276 (2011).
8. Saha, M. & Morais, M.C. FOLD-EM: automated fold recognition in medium- and low-resolution (4–15 Å) electron density maps. *Bioinformatics* **28**, 3265–3273 (2012).
9. Lindert, S. *et al.* EM-fold: *de novo* atomic-detail protein structure determination from medium-resolution density maps. *Structure* **20**, 464–478 (2012).
10. Baker, M.L., Baker, M.R., Hryc, C.F., Ju, T. & Chiu, W. Gorgon and pathwalking: macromolecular modeling tools for subnanometer resolution density maps. *Biopolymers* **97**, 655–668 (2012).
11. Topf, M., Baker, M.L., Marti-Renom, M.A., Chiu, W. & Sali, A. Refinement of protein structures by iterative comparative modeling and cryoEM density fitting. *J. Mol. Biol.* **357**, 1655–1668 (2006).
12. DiMaio, F., Tyka, M.D., Baker, M.L., Chiu, W. & Baker, D. Refinement of protein structures into low-resolution density maps using Rosetta. *J. Mol. Biol.* **392**, 181–190 (2009).
13. Trabuco, L.G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683 (2008).
14. Song, Y. *et al.* High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735–1742 (2013).
15. Berman, H.M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
16. Egelman, E.H. *et al.* Structural plasticity of helical nanotubes based on coiled-coil assemblies. *Structure* **23**, 280–289 (2015).
17. Eswar, N. *et al.* Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **15**, 5.6 (2006).
18. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
19. Davis, I.W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
20. DiMaio, F. *et al.* Improved low-resolution crystallographic refinement with Phenix and Rosetta. *Nat. Methods* **10**, 1102–1104 (2013).
21. Wang, R.Y.-R. *et al.* *De novo* protein structure determination from near-atomic-resolution cryo-EM maps. *Nat. Methods* doi:10.1038/nmeth.3287 (23 February 2015).
22. Fernández, J.J., Luque, D., Castón, J.R. & Carrascosa, J.L. Sharpening high resolution information in single particle electron cryomicroscopy. *J. Struct. Biol.* **164**, 170–175 (2008).

ONLINE METHODS

An overview of the model-building process is illustrated in **Supplementary Figure 1**. Initial models are derived from a crystal structure of an alternate state (PrgH), a crystal structure of a homolog (20S), a manually built comparative model based on a low-resolution structure (PrgK) or an idealized helix (fiber). When starting with an alignment to a known structure, rather than using a full-length model, we used RosettaCM[14]—guided by the experimental data[12]—to rebuild gaps in the alignment. For the proteasome, 200 comparative models were generated from each starting point; for the fiber, ten models were generated. Rosetta forcefield used for optimization, including fit-to-density, was used to select the best model.

Map generation. For all data sets, 'gold-standard' independent reconstructions[23] were made using maximum-likelihood reconstruction[24]. The reported resolutions in the manuscript correspond to the FSC = 0.143 value of the two half maps. One of these reconstructions was used only for rebuilding and refinement (the 'training' map), whereas the other was used only for validation (the 'testing' map). Subsets of the complete particle set were selected and split into two halves; each half-set was used to create lower-resolution training and testing maps. In all cases, *B*-factor correction[22] was applied to the map before refinement to amplify data in high-resolution shells.

MDFF. Models were initially built with Modeller[17,25] in the cases where no crystal structure was available. For each starting homolog, five Modeller models were built, with unaligned terminal residues removed. Each of these starting points was used as input for MDFF. MDFF modeling was carried out using the protocol described by Schulten and coworkers[26]. Energy minimization was used to optimize bond geometries and remove clashes in the input model; a molecular dynamics simulation was carried out for 100 ps and followed by a final energy minimization. The MDFF electron density term was used in all three steps with a weight of 1, 0.3 and 10, respectively.

Density-guided model building. Multiple independent Monte Carlo trajectories are carried out, each consisting of several hundred of the density refined fragment moves described below; trajectories begin with 17-residue fragments and then shift to 9-residue fragments. At each step of the trajectory, a random position in the protein is chosen, with frequency weighted by local density agreement: residues with a local correlation less than 0.6 are sampled frequently (4× base), those with correlation 0.6–0.8 are sampled occasionally (base), and those with correlation above 0.8 are sampled rarely (0.04× base). A set of 25 fragments of 17 or 9 residues in length is selected on the basis of the sequence identity to the target structure. Each fragment is then superimposed on the current model so that the two N- and C-terminal residues overlap with the corresponding residues in the current model. Then, for each fragment, we: (i) rigid body–minimize the fragment into density; (ii) optimize side-chain rotamers to best fit the density; and (iii) minimize all torsions against a forcefield assessing agreement with density, agreement of the terminal residues of the fragment with the corresponding positions in the current model, and backbone and side-chain torsional probabilities. Because this optimization is done with small fragments, ignoring interactions with the remainder of the protein, it is very quick, allowing the 25 fragments to be optimized and evaluated in about 1 CPU-second. At each position, the fragment with best fit to the density that has an r.m.s. of less than 0.5 Å over the terminal residues is selected. Backbone atomic positions from the selected fragment then replace the corresponding backbone in the current model, and the entire structure is minimized in Cartesian space (as in ref. 20) to regularize backbone geometry at the stitching site. The minimization is done using a smooth version of the Rosetta centroid level energy function[14], which primarily consists of sterics and backbone hydrogen bonding supplemented with density agreement. For each input, 200 trajectories were chosen. The amount of sampling was considered converged as additional calculations do not increase the maximum FSC found in the resulting models.

Real-space *B*-factor refinement. To better model the density maps and generate more accurate models, we refined atomic *B* factors against the maps optimizing the real-space correlation between model and map. Given that atom i has a B factor B_i, we calculate the density of the model as

$$\rho_{\mathrm{c}} = \sum_{\text{atoms } i} \left(\frac{\pi}{f_i + B_i/4} \right)^{\frac{3}{2}} \exp\left(-\frac{\pi^2}{f_i + B_i/4} \left\| x - x_i \right\|^2 \right)$$

Here, f is a scattering factor fit to each element. Our implementation makes use of a single-Gaussian scattering for each atom type, but it is straightforward to extend this to a standard five-Gaussian scattering model[27].

B-factor refinement is carried out using quasi-Newton optimization, with the gradient of the B factor of atom i (located at coordinates x_i) given in real space by

$$\frac{\partial \mathrm{RSCC}}{\partial B_i} = \frac{1}{\sigma_{\mathrm{c}}^2} \left(\sigma_{\mathrm{c}} \frac{\partial \sum \rho_{\mathrm{c}} \rho_{\mathrm{o}}}{\partial B_i} - \sum \rho_{\mathrm{c}} \rho_{\mathrm{o}} \frac{\partial \sum \rho_{\mathrm{c}}^2}{\partial B_i} \right)$$

Here, ρ_{c} and ρ_{o} are the calculated and observed density, σ_{c} is the s.d. of the calculated density, the observed density has been standardized to mean = 0 and s.d. = 1 over a mask around the protein, and sums are over the density map. Then

$$\frac{\partial \sum \rho_{\mathrm{c}} \rho_{\mathrm{o}}}{\partial B_i} = \sum_x \rho_{\mathrm{o}}(x) \left(2\left(x - x_i\right)^2 - \frac{3f_i + 3B_i/4}{\pi^2} \right) \left(\frac{-\pi^2}{\left(f_i + B_i/4\right)^2} \right)$$

$$\frac{\partial \sum \rho_{\mathrm{c}}^2}{\partial B_i} = \sum_x 2\rho_{\mathrm{c}}(x) \left(2\left(x - x_i\right)^2 - \frac{3f_i + 3B_i/4}{\pi^2} \right) \left(\frac{-\pi^2}{\left(f_i + B_i/4\right)^2} \right)$$

To prevent overfitting of B values, we also use restraints so that nearby atoms have similar B values, using the same formulation as phenix.refine[28]

$$E = \sum_{\substack{\text{atoms } i,j \\ \left\| x_i - x_j \right\| < 5\,\text{Å}}} \pi \frac{1}{\left\| x_i - x_j \right\|} \times \frac{\left(B_i - B_j\right)^2}{\left(B_i + B_j\right)}$$

As atomic-coordinate errors can lead to artificially high B values in refinement, which leads to reduced forces acting on these (incorrect) atom positions in subsequent rounds of coordinate refinement, we perform several rounds of refinement with uniform B values before our first cycle of B-factor refinement.

Atomic refinement. Atomic refinement is based on the Rosetta 'relax' protocol, where cycles of discrete side-chain optimization are alternated with cycles of quasi-Newton optimization. In all cases, the relevant symmetry was included in the Rosetta refinement to model the full biological unit. An additional term assesses agreement to density. For speed considerations, we use approximate model-map correlation as our metric: an atom's density is convoluted over the entire map, with spline interpolation used to quickly compute the $\Sigma\rho_c\rho_o$ term in the correlation, with ρ_c the computed map and ρ_o the experimental map. With proper normalization of ρ_c and ρ_o, this approximation differs from a real-space correlation only by the term $\Sigma{\rho_c}^2$; assuming this is constant is equivalent to assuming a constant atom density, which is not unreasonable.

B-factor optimization is carried out using a similar approximation for computational efficiency. Our fast density formulation precomputes a three-dimensional (3D) grid where $f(x) = \Sigma_z\rho_c(z + x)$, that is, the overlap between calculated and observed density when a single atom is placed at x; this was extended to a 4D grid where $f(x, B) = \Sigma_z\rho_c(z + x)$. Grid spacing was uniform in $1/B^2$, which allows for 8–12 grid points in the B dimension to accurately approximate this space.

Using this approximation, B factors can be very quickly fit by refining atoms along the B dimension of the 4D surface. However, when refining along the B dimension, the assumption of a relatively constant $\Sigma{\rho_c}^2$ is violated. To remedy this, we compute the exact correlation at a number of fixed values of B_{mean} (corresponding to each discrete sample in the B dimension). These values are used as a scaling factor for the spline coefficients of each B slice. This allows us to use 4D interpolation to both fit B values and refine atomic coordinates taking into account atomic B factors. All of our refinement steps are followed by exact B-factor refinement at the end, which tends to further improve real-space correlation by about 0.01–0.02.

Finally, previous work has shown that relaxing bond ideality is important for both structure prediction and refinement against crystallographic data[28]. Thus, the final two cycles of refinement are carried out in Cartesian space, allowing for bond-angle and bond-length deviations to slightly improve energetics and fit to the experimental data.

Validation metrics. Following previous work[20], models were validated against an independent reconstruction using the integrated FSC of the model and independent reconstruction in high-resolution shells. Additionally, an alternate likelihood-based model validation metric was explored. This metric is formulated on the basis of the probability of the data given the model. Assuming each structure factor is independent,

$$\begin{aligned} P\left(E_{\text{obs}} \middle| E_{\text{model}}\right) &= \sum_{\substack{\text{resolution}\\ \text{range}}} P\left(e_{\text{obs}} \middle| e_{\text{model}}\right) \\ &= \sum_{\substack{\text{resolution}\\ \text{range}}} \int_{\substack{\text{complex}\\ \text{plane}}} P\left(e_{\text{obs}} \middle| e_{\text{true}}\right) P\left(e_{\text{true}} \middle| e_{\text{model}}\right) \mathrm{d}e_{\text{true}} \end{aligned}$$

Here, E_{model} and E_{obs} are the model and map structure factors (with lowercase e referring to individual structure factors), normalized in resolution bins so that $\Sigma|E(r_i)|2 = 1$. The term e_{true} represents the (unknown) ground-truth structure factors. In the integral, the first term accounts for errors in the reconstruction and the second accounts for errors in the model. Although fully exploring this formulation remains an important topic of future research, parameterization of each of these terms is not straightforward and is out of the scope of this manuscript.

Instead, in this work, we explore a more computationally tractable formulation of model error: the expected phase error (EPE). By computing errors in phase space, we no longer need to worry about integration over different resolution ranges because the EPE goes to 90° in the limit of completely random data. We can integrate over all resolutions—independent of estimated map resolution—and have a reasonable measure of model quality comparable between different maps. Our measure assumes phase errors are normally distributed in phase space with deviation σ_k

$$P\left(\measuredangle\alpha_{hkl}^{\text{train}}\alpha_{hkl}^{\text{test}}\right) \sim \mathcal{N}\left(x; 0, \sigma_k\right)$$

These deviations are estimated from the independent reconstructions and are computed separately for each resolution bin by calculating deviations in model phase error between different bins.

Under this assumption, given the phase error $\delta = \alpha_{\text{model}} - \alpha_{\text{map}}$ between model and test map, we can compute

$$\begin{aligned} \text{EPE} &= \arg\left(\int_{-\infty}^{+\infty} e^{i|x|}\mathcal{N}\left(x; \delta, \sigma_k\right)\mathrm{d}x\right) \\ &= \arg\left(\frac{1}{2}e^{-\frac{\sigma_k^2}{2}-i\delta}\left(\operatorname{erf}\left(\frac{\delta - i\sigma_k^2}{\sigma_k\sqrt{2}}\right) + e^{2i\delta}\operatorname{erf}\left(\frac{\delta - i\sigma_k^2}{\sigma_k\sqrt{2}}\right) + 1\right)\right) \end{aligned}$$

In bins where the two maps agree (for example, in low-resolution bins), the error is simply the difference between model phase and the independent-map phase. As the agreement drops in higher-resolution bins, the error is smoothed out; at the extreme—at resolutions that contain no information—the error is uniformly 90° given the model-map agreement.

Structural data sets. *PrgH/PrgK map reconstruction.* Cryo-EM image data for the PrgH/PrgK sample have been recorded previously[29,30] on a Polara transmission electron microscope equipped with a field emission gun (FEG) operated at 300 kV and a 4K camera (Gatan Ultrascan 4000 UHS, physical pixel size of 15 µm) with a total electron dose of 25–35 e⁻/Å². Defocus values were determined using CTFFIND3 (ref. 31). From a total of >70,000 particles (pixel size, 1.33 Å/pixel), random subsets containing 40,000, 30,000, 20,000, 10,000, 3,000 or 1,000 particles were extracted. 3D maps were independently generated using gold-standard reconstructions in RELION (v1.2).

20S map reconstruction. The 3D density maps of *Thermoplasma acidophilum* 20S proteasome used in this study were determined using a subset of the particle images from a full data set described previously[2]. Therefore, the sample preparation, data acquisition and image processing were identical to those of this previous

doi:10.1038/nmeth.3286

study. Briefly, *T. acidophilum* 20S proteasome was expressed and purified from *Escherichia coli* according to the established protocols[32]. A drop of 2 µl of purified 20S proteasome at a concentration of ~0.9 µM sample was applied to glow-discharged Quantifoil holey carbon grids (Quantifoil, Micro Tools) and plunge frozen by using a Vitrobot Mark III (FEI). Grids of frozen hydrated samples were imaged using a FEI TF30 Polara electron microscope (FEI) equipped with a field emission electron source and operated at an accelerating voltage of 300 kV. Images were recorded at a nominal magnification 31,000× using a Gatan K2 Summit camera operated at super-resolution counting mode with a calibrated physical pixel size of 1.22 Å at 31,000×. A 10-s exposure time at a dose rate of ~10 counts/pixel/s leads to a total dose ~30 $e^-/Å^2$. The defocus was in the range of ~0.8–1.9 µm. CTFFIND3 (ref. 31) was used to determine the defocus values. Side-view particles of 20S proteasome were picked automatically by using FindEM[33]. All picked particles were first subject to standard procedure of multiple rounds of multireference alignment and classification. Particles within bad 2D classes were removed. All remaining particles were subject to further manual inspection, and more bad particles were removed. The complete data set contains 126,729 particles.

Maps were constructed using subsets of full data set containing the first 5,000, 3,000, 1,200 and 1,000 particles from the full set. For each subset, GeFREALIGN[34] was used to refine and determine the 3D reconstructions with a D7 symmetry following a frequency-limited refinement procedure[2,24]. The atomic structure of archaeal 20S proteasome (PDB code: 2C92) filtered to 15 Å was used as the initial model. The final 3D reconstructions have resolutions of 4.1 Å, 4.4 Å, 5.0 Å and 6.0 Å using Fourier shell correlation 0.143 criteria[35]. Structure features in the amplitude-sharpened maps confirm these claimed resolutions.

23. Henderson, R. *et al.* Outcome of the first electron microscopy validation task force meeting. *Structure* **20**, 205–214 (2012).
24. Scheres, S.H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
25. Eswar, N. *et al.* Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein. Sci.* **50**, 2.9 (2007).
26. Trabuco, L.G., Villa, E., Schreiner, E., Harrison, C.B. & Schulten, K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **49**, 174–180 (2009).
27. Peng, L.-M., Ren, G., Dudarev, S.L. & Whelan, M.J. Robust parameterization of elastic and absorptive electron atomic scattering factors. *Acta Crystallogr. A* **52**, 257–276 (1996).
28. Afonine, P.V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
29. Radics, J., Königsmaier, L. & Marlovits, T.C. Structure of a pathogenic type 3 secretion system in action. *Nat. Struct. Mol. Biol.* **21**, 82–87 (2014).
30. Schraidt, O. & Marlovits, T.C. Three-dimensional model of *Salmonella*'s needle complex at subnanometer resolution. *Science* **331**, 1192–1195 (2011).
31. Mindell, J.A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
32. Yu, Y. *et al.* Interactions of PAN's C-termini with archaeal 20S proteasome and implications for the eukaryotic proteasome-ATPase interactions. *EMBO J.* **29**, 692–702 (2010).
33. Roseman, A.M. FindEM—a fast, efficient program for automatic selection of particles from electron micrographs. *J. Struct. Biol.* **145**, 91–99 (2004).
34. Li, X., Grigorieff, N. & Cheng, Y. GPU-enabled FREALIGN: accelerating single particle 3D reconstruction and refinement in Fourier space on graphics processors. *J. Struct. Biol.* **172**, 407–412 (2010).
35. Rosenthal, P.B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).

doi:10.1038/nmeth.3286