Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta

Ray Yu-Ruei Wang[1,2,¶], Yifan Song[2,$], Benjamin A Barad[3,4], Yifan Cheng[5], James S Fraser[3] & Frank DiMaio[2,6*]

1 Graduate program in Biological Physics, Structure and Design, University of Washington, Seattle, WA, USA.
2 Department of Biochemistry, University of Washington, Seattle, WA, USA.
3 Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, California, USA.
4 Graduate Group in Biophysics, University of California, San Francisco, San Francisco, California, USA.
5 Keck Advanced Microscopy Laboratory, Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California, USA.
6 Institute for Protein Design, University of Washington, Seattle, WA, USA

¶ Present address: Department of Biochemistry & Biophysics, University of California, San Francisco, San Francisco, CA, USA.

$ Present address: Cyrus Biotechnology, Seattle, Washington, USA.

* Corresponding author: dimaio@uw.edu

Abstract

Cryo-EM has revealed the structures of many challenging yet exciting macromolecular assemblies at near-atomic resolution (3-4.5Å), providing biological phenomena with molecular descriptions. However, at these resolutions accurately positioning individual atoms remains challenging and error-prone. Manually refining thousands of amino acids – typical in a macromolecular assembly – is tedious and time-consuming. We present an automated method that can improve the atomic details in models manually built in near-atomic-resolution cryo-EM maps. Applying the method to three systems recently solved by cryo-EM, we are able to improve model geometry while maintaining the fit-to-density. Backbone placement errors are automatically detected and corrected, and the refinement shows a large radius of convergence. The results demonstrate the method is amenable to structures with symmetry, of very large size, and containing RNA as well as covalently bound ligands. The method should streamline the cryo-EM structure determination process, providing accurate and unbiased atomic structure interpretation of such maps.

Introduction

Advances in direct electron detectors as well as better image analysis algorithms have led cryo-electron microscopy (cryo-EM) to achieve near-atomic resolution (3-4.5 Å) using single-particle analysis [1-3]. Cryo-EM reconstructions at these resolutions, where individual β-strands are resolvable, and bulky sidechains are somewhat visible, make it possible to build an all-atom model directly from such maps [4,5]. Although sequence can be registered, density maps at this range of resolution do not grant enough information to precisely assign coordinates for each atom in the structure, from which molecular interactions for a biochemical process is captured. Furthermore, such model building and refinement is challenging and error prone [6,7]. Determination of detailed atomic interactions from these sparse sources of data is desirable, however, the inherent ambiguity in the data makes identifying these interactions extremely difficult, even for experts.

Model-building into a cryo-EM map at near-atomic resolution generally involves manually building a model into the map using a graphical user interface tool [8] followed by refinement with software repurposed from X-ray crystallography [9,10]. This process requires identification of key amino acid sidechains to register stretches of sequence within the map (possibly aided by the topology from a homologous structure), followed by extension of these short fragments of sequence to form one or more fully connected protein chains. At near-atomic resolution, this manual model-building and refinement can be error prone owing to: a) the density may not be of sufficient resolution to uniquely identify sidechain rotamers, even for bulky aromatic residues, making it difficult to accurately determine sidechain-sidechain or sidechain-backbone interactions; b) for regions of non-regular secondary structure (turns or loops) or with poor local resolution, it may be difficult to accurately position backbone atoms; and c) in these same regions, precise sequence registration may also be error prone. Getting these atomic interactions correct is crucial for understanding detailed atomic mechanisms of proteins, designing drugs with a very specific shape complementarity, and for understanding subtle conformational changes of a protein. A structure refinement procedure that can automatically improve the atomic details of a model from such density data is thus very much desired.

In this manuscript, we develop a three-stage approach for automatically refining manually traced cryo-EM models (Figure 1). While previously we have developed an iterative local rebuilding tool capable of refining homology models into near-atomic-resolution cryo-EM maps [11], several advances were required for extending this tool to successfully refine hand-built models. Our new approach includes a method for automatically detecting and correcting problematic residues in hand-built models without overfitting, a model-selection method for identifying models with good agreement to the density data and with physically realistic geometry, a voxel size refinement method for correcting errors in calibrating the magnification scaling factor of a microscope, an improved

sidechain-optimization method to correct sidechain placement errors in very large systems, and a way to estimate uncertainty in a refined model. These methods, combined, allow to correct backbone errors that significantly deviate from the starting model, but may still assign a high degree of confidence to these regions in the refined model.

Finally, we apply this approach to three recently solved cryo-EM single particle reconstructions at near-atomic resolution: the TRPV1 channel at 3.4-Å resolution (TRPV1) [12], the F_{420}-reducing [NiFe] hydrogenase (Frh) at 3.4-Å resolution [13], and the large subunit of mitochondrial ribosome at 3.4-Å resolution (mitoribosome) [14]. We show that in all three cases of diverse and large systems, we are able to automatically refine models to high-quality (as assessed by MolProbity), while maintaining or improving agreement to the density data. Significantly, in the case of TRPV1, we newly identify a biological relevant atomic interaction – a disulfide bond – not built in the originally deposited model, but supported in the literature. In the case of Frh, we show our refinement procedure led to a significant improvement of model geometry. Finally, in the case of mitoribosome, we show significant improvement in model geometry: the number of "Ramachandran favored" residues increases by 5%, and Molprobity [15] score improvement is observed in all 48 protein chains.

Results

An overview of our refinement approach is shown schematically in Figure 1 (and is fully described in *Methods*). Broadly, the approach proceeds in three stages. In the first stage, we identify problematic residues by assessing local model-strain and local agreement to density data. These regions are rebuilt against a "training" half-map using fragment-based Monte Carlo sampling with many independent trajectories followed by all-atom refinement. Secondly, the best subset of these independent trajectories are selected by identifying a subset of stereochemically correct models with best agreement to an independent "validation" half-map, to prevent overfitting. Finally, models are further optimized in the full-reconstruction with a weight optimally scaled between experimental data and the forcefield using the "validation" half map. Our approach adopts and improves upon our previous work on refining cryo-EM structures from distant homology structures [11], in which a similar fragment-based backbone rebuilding strategy is employed. However, several critical improvements were necessary in extending our previous work to successfully refine hand-traced models, larger complexes, and a more diverse set of systems.

Identification of backbone errors using local strain. In previous work [11], local fit to density is used to identify residues in a distant homology model to rebuilt. However, unlike remote homology models, hand-traced models typically fit the data very well, but are incorrect geometrically (strain). Consequently, a key improvement is to make use of model strain as a criterion in selecting regions to

refine. Moreover, when the previous approach was applied to the *de novo* hand-traced models from cryo-EM maps, we observed that – following all-atom refinement – in incorrect regions, the models still fit the density well, but did so by introducing strain in the nearby bond angles and torsions. This often occurred in near Cβ atom of aromatic residues, where strain was introduced to fit the sidechain into density (Figure 1-figure supplement 1). We reasoned that in these strained residues, the backbone was incorrect; by correcting the backbone we would be able to fit a non-strained sidechain into density. Thus, local strain can serve as an indicator to identify regions to refine to improve both the fit-to-density and model geometry. We developed an error predictor by constructing a function (see *Methods*) that assesses both local model-map agreement as well as local model-strain. Using a training dataset composed of error-containing models of a cryo-EM map in which the structure has been determined by X-ray crystallography (Figure 1-figure supplement 2), we show that the new error predictor offers better discrimination of incorrectly versus correctly placed backbone, with an AUPRC (area under precision-recall curve) of 0.80 versus 0.76 using density alone (Figure 1-figure supplement 2). In cases where models are hand-built into density, we expect this strain term to play an even larger role, as fit-to-data is expected to have larger influence on the initially constructed model.

***Better treatment of sidechain density*.** Recent works have shown that certain sidechains – particularly negatively charged amino acids (Glu/Asp) – tend to suffer from radiation damage and thus appear weaker in single-particle reconstructions [16,17]. Moreover, density from certain bulky sidechains, for example, Lys and Arg, tends to be less well-defined than their backbone density. This missing density dramatically affects the convergence of conformational sampling during structure refinement, where sidechains tend to be fit into density corresponding to backbone atoms. To compensate for this, we downweigh the contributions of sidechains which are less resolved in cryo-EM density. Down-weighing factors for each amino acid were determined by comparing the average per-amino-acid real-space B-factor on two cryo-EM reconstructions with known high-resolution crystal structures (20S proteasome [1] and β-galactosidase [16]), where the ratio of backbone and sidechain average B-factors was used to derive the scaling factors. Table 2 shows the computed scalefactors used in our refinement method.

Local sidechain refinement for large complexes. When our previous all-atom refinement approach was applied to very large complexes (800+ residues), we observed many instances where sidechains were not properly optimized into density (Figure 6-figure supplement 1). It was hypothesized that this was due to the convergence of sidechain optimization, as the number of possible sidechain states expands exponentially with the number of residues present in a protein. Here, we opted to treat this global optimization problem as a series of smaller local optimization problems, repeatedly optimizing overlapping regions of ~20-100 residues until all residues in a protein are visited at least once. This

approach resolved this sidechain fitting issue, as shown in Figure 6-figure supplement 1 (right panel).

Voxel size refinement. The voxel size of a cryo-EM reconstruction is determined by the physical pixel size on the detector scaled by a magnification factor. However, the magnification factor may be determined with some inaccuracy, leading to errors in deciding the voxel size of the resulting single-particle reconstruction. It has shown that voxel size may be off by as much as several percent from previous studies when using EM maps as molecular replacement targets [18]. Here, we develop a voxel size refinement strategy, which scales the voxel size of the map to maximize model-map real-space correlation coefficient. During refinement we alternate structure refinement and map voxel size refinement with several cycles iteratively until the voxel size converges (Figure 1). The approach is fully described in the Methods section.

Moreover, we investigate the robustness of our voxel size refinement method in the presence of model errors, and to demonstrate that our iterative approach captures the general agreement between forcefield and voxel size. We initially made use of an arbitrary target structure (PDB id: 4AKE). We calculated density to 3Å resolution on a 1Å grid. We ran several MD trajectories in Rosetta, followed by all-atom minimization, yielding 50 models that are 2.9-3.1Å RMSd fro the native structure. We initially refined voxel size against each of these models, yielding voxel sizes from 0.95 to 1.02 Å (stdev=0.011). Following our iterative procedure, the deviation was much smaller, ranging from 0.99 to 1.02 (stdev=0.004). Figure 2A illustrates the distribution of voxel sizes derived from the models before refinement (red curve) and after refinement (blue curve), showing a sharp peak at the true voxel size (1.0Å) after refinement.

The effect of B-factor sharpening in Rosetta structure refinement. To investigate to what extent of B-factor sharpening would influence the accuracy of Rosetta structure refinement, we benchmarked structure refinement using various B-factors from 0 to -200 (Figure 2B). Using the 20S proteasome 3.3Å resolution map, we refined a model starting from a template (PDB id: 3H4P, 52% sequence identity) into the training map, and reported the iFSC evaluated by the validation map. The method does not show a particularly strong dependency to B-factor sharpening values; with B-factors ranging from –40 to –130 the refinement all performed equally well as assessed by free-iFSC.

The role of independent reconstruction in Rosetta structure refinement. In our previous approaches, we have used independent reconstructions (“validation” half-map) for both model selection [11] and for determination of the balance between model geometry and fit-to-data during refinement [19]. In this manuscript, we use independent reconstructions in the same manner during the first two stages of refinement (Figure 1). However, at the very last stage we perform several steps in the context of the full reconstruction, due to the additional sidechain details that may be only present in the full reconstruction. As

shown in Figure 1, for the best 10 sampled models selected from the stage 2, we perform a final all-atom and atomic B-factor refinement against the complete reconstruction. Similar to the approach adapted by the REFMAC group [9,20], we use two independent halves of the data (training/validation half-maps) to optimize the weight used with full-reconstruction data (see *Methods* and (Figure 1-figure supplement 3), and that weight is used in refinement against the full reconstruction, as well as voxel size refinement. Following refinement against the full reconstruction, model geometry is verified (using MolProbity [15]) to ensure it is not worsening during refinement against the full reconstruction. This confers additional sensitivity during model selection.

Evaluation of refined models with Molprobity and EMRinger. Models are evaluated for geometric quality using Molprobity [15], which compares summary statistics of an all-atom model to those from high-resolution crystal structures. In addition to using MolProbity to assess model quality, we further validate the Rosetta-refined models with EMRinger [21], as an independent source to validate both model geometry and density-fit at sidechain level. EMRinger samples density around Cγ atoms as they are rotated about the χ_1 dihedral angle, and identifies the angle which presents peak density for the Cγ; based on prior statistical and chemical information, this position should generally fall into the rotamer distribution of χ_1, with angles of 60, 180, and 300 degrees. The distribution of measured peak angles at various signal-to-noise cutoffs is integrated into the EMRinger score, which reports on backbone model-to-map agreement using side chain geometry.

Application to TRPV1. We first applied our new refinement approach to the recently determined 3.4-Å cryo-EM reconstruction of the TRPV1 channel in the apo form [12]. Half-maps were reconstructed by subdividing particles into two sets randomly, with one used for initial model rebuilding and refinement, and the other used for validation. The deposited model (PDB id: 3J5P) was used as input to the protocol described previously. All refinement was carried out using the native C4 symmetry. Fragment-based rebuilding was only carried out on the transmembrane region, while a final all-atom refinement was performed on the full structure. All input files are included as Supplemental Data File 1.

The results of refinement are indicated in Figures 3 and 4, and Table 1. The refined model improves both model quality and model-data agreement compared to the deposited model: the MolProbity score improves from 3.81 to 1.45, the fit-to-data (integrated Fourier shell correlation from 10 to 3.4Å) slightly drops from 0.612 to 0.607, but the EMRinger score improves from 0.65 to 2.34, indicating the better fitting shown in the deposited model might be due to overfitting. Figure 3A–B compares the refined model and the deposited model, colored with model violations reported by MolProbity. Figure 4A illustrates the convergence of our refined ensemble, showing the 10 selected refined structures, the top model colored by per residue structural variation, and the refined B-factors. Both of these measures provide unique insights on assessing the local confidence of the

refined models, in which structural variance shows the allowed local conformations that satisfy the density data, whereas B-factors assess the local resolution of the density data at different regions of a model.

Closer inspection of the refined models identified a disulfide linkage (C386-C390) that was not built in the deposited model (Figure 4B). This disulfide has previously been identified and characterized in the literature as playing an important role in response to oxidative stress for the TRPV1 channel [22]; this, combined with our models better explaining a tube of density unaccounted for in the deposited model, let us speculate that this disulfide bond is present in the cryo-EM reconstruction. This motion also illustrates the magnitude of conformational change that may be captured by our protocol; our Monte Carlo backbone sampling strategy allows refinement to overcome energy barriers that other methods using density minimization alone cannot. Despite the magnitude of these changes, the conformational ensemble is well converged in this region (Figure 4B, right panel) providing further confidence in our refined model.

Refinement of highly-liganded complexes: application to the F_{420}-reducing [NiFe] hydrogenase complex. As our next test of the approach, we wanted to illustrate model refinement of a complex with large numbers of ligands, some of which are covalently bound, all in a system with high-order point symmetry. For this, we chose the 3.4-Å reconstruction of F_{420}-reducing [NiFe] hydrogenase complex, where the asymmetric unit contains 3 protein chains which feature with a [NiFe] cluster, two metal ions, and four [4Fe4S] clusters covalently bound to cysteine sidechains, and an FAD (Figure 5A). The complex is a dodecamer with tetrahedral symmetry, with 12 copies of a 902-residue molecule of three protein chains. We used the *-auto_setup_metals* option of Rosetta to maintain covalent linkages between protein and ligand during refinement (full input files are included as Supplemental Data File 2). The results of refinement are indicated in Figure 5 and Table 1, where the MolProbity score improves from 3.98 to 1.59, the EMRinger score improves from 1.06 to 2.17, however, the iFSC drops from 0.743 to 0.708. We reason that the decrease of fit-to-density, at high-resolution (10-3.4Å) shells, may be a result of overfitting the model to the density map, where the deposited model was forced to fit the density by deviating the geometry observed in high-resolution crystal structures. This overfitting hypothesis is well supported by a high number of bad clashes and 39% rotamer outliers found in the deposited model (Table 1 and Figure 5C).

Refinement of large complexes: application to the mitochondrial ribosome large subunit. Finally, we wanted to test the ability of our refinement to scale to large asymmetric macromolecular assemblies, more typical of cryo-EM single particle reconstruction. To do so, we considered refining models against the previously published 3.4-Å cryo-EM reconstruction of the large subunit of the human mitochondrial ribosome [14]. The deposited model had been previously refined with *REFMAC* [9], and consists of 48 chains of proteins with 7469 amino acids assigned and two chains of RNA with 1529 nucleic acid bases.

In order to make conformational sampling tractable, we used a slightly modified strategy from that shown in Figure 1 (full input files are included as Supplemental Data File 3). The first two steps of the protocol (error identification and backbone rebuilding) were carried out on each protein chain individually, while the third step was carried out on the fully assembled complex. Model selection was carried out on each individual chain; each selected model was refined as a complete assembly, with the top model of each chain refined together, the second selected model for each chain refined together, and so on. Nucleic acids were not refined but were included as rigid bodies to accurately recapitulate protein/RNA interactions.

The results of refinement are indicated in Figure 6 and Table 1. Several large-scale conformational changes again appear in converged models; these models show better geometry, fit to density and fewer unexplained regions of density. The backbone geometry improvements are in particular noticeable in proteins with β-sheet containing domains. Unlike other refinement procedures (*phenix.real_space_refine* [10] and *REFMAC* [9]), which require manual input of secondary structure restraints determined either from an initial model or homologous protein structure to maintain backbone geometry during refinement, in our approach the Rosetta forcefield is able to optimize hydrogen bond geometry in secondary structures without requiring *a priori* knowledge of secondary structures. This is particularly powerful in refining *de novo* structures where secondary structure is ambiguous due to poor local resolution. Figure 5C illustrates an example (chain k) of this from the case of mitoribosome, where a β-sheet not present in the original model is identified, the backbone geometry is improved, and the model fits the density much better than the deposited model (Figure B, left panel, red arrow); the refinement also shows a large radius of convergence.

The refined ribosome model has 1.50 MolProbity score, 0.676 iFSC, and 2.40 EMRinger score. The largest improvements tend to occur in regions of low local resolution (~5Å assessed by ResMap[23] from the original paper) on the periphery of the complex. Looking at the results on individual chains, as indicated in Figure 5A, the MolProbity score improves on all 48 protein chains, which in part is from the much improved backbone geometry assessed by the *Ramachandran favored* term in MolProbity (Figure 6A, right panel). Our Monte Carlo backbone sampling can correct these incorrect backbone placements, which often require significant compensating conformational changes. EMRinger score is also consistently improved ((Figure 6-figure supplement 2), particularly in regions where the deposited model scores poorly.

Comparison to phenix.real_space refinement. Finally, using the same set of target proteins, we compare the Rosetta refinement results with another state-of-the-art real-space refinement method from the *phenix* package (*phenix.real_space_refine*) [10]. In order to prevent refinement from fitting to

noise, starting from the deposited models we carried out phenix real-space refinement in the training maps, containing only the half of the data used by Rosetta in the first step of the refinement procedure (Figure 1). We then used the validation map to evaluate and compare the *phenix* refinement results to the Rosetta-refined models before full-map refinement. On the case of TRPV1, *phenix* used 0.24 CPU hours, generating a single model. For Rosetta, the 1000 independent trajectories take about 5 hours each, for 5000 CPU-hours. As shown in Table 3, with much shorter run time (at most 1 hour) *phenix* can yield models with geometry almost as good as Rosetta, albeit with slightly worse density fit evaluated by both real-space correlation coefficient and iFSC. However, without the Monte Carlo backbone conformational sampling of Rosetta, models generated from *phenix* tend to minimally perturb the structure, and can not provide large backbone corrections shown in this manuscript.

Discussion

In this manuscript, we develop a method for improving atomic details of manually traced models from 3-4.5Å resolution cryo-EM density. We show the applicability of the approach, by applying it to three systems: a membrane protein, an asymmetric macromolecular assembly containing large numbers of protein chains and RNAs, and a highly symmetric system with a large number of ligands. In all cases, we show that we are able to significantly improve model geometry while maintaining good agreement to the density data. We show that model convergence can be used to suggest local model uncertainty in addition to B-factors. Finally, we also show that our models also recover structure features that are supported in the literature, or in much better local agreement with the density data.

Unlike other approaches [9,24], our approach can automatically perform large-scale backbone reorganization, correcting backbone placement errors common in these 3-4.5Å resolution datasets. Two features of our refinement approach regarding the use of prior information are critical in the success of this large-scale refinement. First, the use a physically realistic forcefield throughout refinement handles the under-constrained nature of refinement at these resolutions, by using chemical “domain knowledge” learned from high-resolution crystal structures to implicitly fill in the missing information in the data. Second, our fragment-based rebuilding which explicitly samples the most likely backbone conformations given a short stretch of sequence also uses prior information gather from high-resolution protein structures, further restricting conformation space, and filling in additional information not present in the data.

We found in all cases the high-resolution density-fit (evaluated by iFSC integrated from 10-3.4Å resolution shells) all drops slightly after Rosetta structure refinement. We reason that it is likely due to the deposited models were overfit to the density maps, where models (especially sidechains) were forced to fit into the

density by violating the geometry observed from known high-resolution crystal structures. The observation of the slightly decrease of fit-to-density but significant improvement of model geometry bolsters the importance of using prior information (eg. sidechain rotatmers), as well as having a refinement scheme to monitor model overfitting.

Finally, an open question is on what way structure refinement can be further improved, particularly as refinement extends to even lower resolutions (worse than 5Å). Enhancing the predicting power of the Rosetta modeling methods is the key to to push the resolution limit of the current refinement method further. This can be achieved through: 1) improving the energy function (forcefield) used in refinement, and 2) improvements in conformational sampling methodology, particularly for systems where secondary structure prediction is poor. Further improvements in the role of B-factor sharpening and the effect on refinement are necessary, as well as better predictors of local model error. Finally, structure refinement in maps with highly heterogeneous local resolution remains challenging, where a single set of refinement parameters cannot readily be applied at all regions. Methodological improvements that allow adjustment of parameters based on local map quality will be essential to accurately refine structures from such maps. In our effort to enable automated structure refinement on large macromolecular assemblies, we hope this method can be a valuable tool for determining atomic accuracy structures from near-atomic-resolution cryo-EM data.

Methods

***Preparing maps for refinement*.** Split maps were provided by the original authors. One map was randomly chosen for refinement, and the other was used for validation. In all cases, a B-factor of -100 was applied to the map used for refinement using the "*image_handler"* tool in RELION [3]. The maps were subsequently filtered to the user-refined resolution. In the case of the mitochondrial ribosome, segmented maps were prepared using a custom Rosetta application and the deposited structure to guide segmentation:

density_tools.default.linuxgccrelease -s 3j7y0.pdb -mapfile EMD-2762.mrc -mask_radius 2 -maskonly

Some steps of the protocol also made use of the full reconstruction. As with the training map, these were sharpened using a B-factor of -90 with a low-pass filter to 3Å.

***Preparing structures for refinement*.** In the case of TRPV1, residues 111-202 in the Ankryin repeat domain from the deposited model did not have visible density, and so were deleted prior to refinement. Furthermore, automatic refinement as applied in two stages due to the highly heterogeneity between the

trans-membrane domain and the Ankryin repeat domain. The trans-membrane domain (residue 234-586) was first refined in the density masked using the deposited model. In the case of the mitoribosome, residues from chain t and chain f, in which atoms are assigned to residues “UNK,” were removed from all the refinement process, as well as data analyses or results comparisons. In the case of Frh, we found the deposited symmetry operators (the “BIOMT” lines) are not able to generate a symmetric model that can properly fit into the density map (Figure5-figure supplement 1). To assemble the symmetric complex, we manually docked each subunit of the symmetric complex into the deposited density map (EMD-2513) using Chimera, and used this model as the “deposited model” used in the paper. Refinement of ligands received special treatment: refinement started using protein only, with constraints maintaining ligand site geometry. Later, ligands were added back on and rerefined.

***Algorithm for model rebuilding*.** Model rebuilding generally follows the procedure from our previous work [11], with a few key changes highlighted below. Rebuilding starts from the deposited structure, which is first conservatively refined using one macrocycle of the Rosetta *relax* protocol to trigger local strain on sidechains, which iterates four cycles Monte Carlo rotamer optimization with all-atom minimization, ramping the weight on van der Waals repulsion in each cycle. Minimization is carried out in Cartesian space, with a term enforcing ideal bond angles, bond lengths, and planarity [25].

Following Cartesian minimization, the worst residues are selected using the following equation to evaluate the quality of the model at residue *i*:

$$Z_{error}^{(i)} = w_{dens} \cdot Z_{dens}^{(i)} + w_{lcldens} \cdot Z_{lcldens}^{(i)} + w_{bonded} \cdot Z_{bonded}^{(i)} + w_{rama} \cdot Z_{rama}^{(i)}$$

Four different terms appear in this equation, two of which assess a model’s agreement to data, two of which assess a model’s local strain. The first two, $Z_{dens}^{(i)}$ and $Z_{lcldens}^{(i)}$, assess the model-map agreement of the backbone and sidechain atoms of each residue, computing the real-space correlation coefficient in a region around a residue, and converting that to a *Z-score* compared to the entire model. For the former term, an absolute correlation coefficient is computed; for the latter term, the correlation is normalized with respect to residues nearby (those within 10 Å of residue *i*). The latter term is specifically added to deal with maps that have significant diversity in local resolution.

The second two terms, $Z_{bonded}^{(i)}$ and $Z_{rama}^{(i)}$, assess a model’s strain following model refinement. The motivation for these terms is that in cases where the model was built incorrectly into density, it will be energetically unfavorable. Following an initial refinement, these incorrect portions will either be move away from the data, or will introduce model strain to maintain the favorable agreement to the data, depending upon the balance of forces between the two. These terms compare the per-residue bond geometry term, and the per-residue

Ramachandran energy, respectively, to that over the entire structure, and compute a *Z-score* for each residue.

For each of the four terms, a Z-score is computed and is summed together, with a particular weight for each term. The weights were tuned using a 3.3-Å cryo-EM map dataset with known high-resolution structure (the 20S proteasome [1]), where a set of ~500 error-containing models was used as the training data. The results of this tuning process are shown in Figure 1-figure supplement 2. The final weights selected were w_{dens}=0.45, $w_{lcldens}$=0.05, w_{bonded}=0.15, w_{rama}=0.35.

After computing this weighted Z-score for each residue, all residues with a score below some target value (see the next section on iteration for specific values) are selected for local rebuilding. Local rebuilding uses the iterative fragment-based approach previously published [11]. In our new approach, a residue is randomly chosen from the pool tagged for rebuilding from the previous step. Given the local sequence around this selected residue, a set of 25 protein backbone conformations from high-resolution structures with similar local sequence and predicted secondary structure is sampled. Each sampled backbone is refined – as an isolated fragment – into density using the following three step procedure: a) the backbone only is minimized in torsion space using a simplified energy function, b) sidechain rotamers are optimized into density, and c) both backbone and sidechain are minimized in torsion space using a simplified energy function. Constraints on the ends of each fragment ensure the local region is reasonable in the context of the entire backbone. Of the 25 sampled fragments, the best is selected by fit to density. Finally, the replaced fragment is minimized in the context of the complete structure. This process is run as a Monte Carlo trajectory.

***Iterative rebuilding and all-atom refinement*.** Model rebuilding and all atom refinement are run iteratively, as shown in Figure 1. Four separate 200-step Monte Carlo trajectories are run with increasing coverage of predicting errors but sacrificing the accuracy of the predictions. This is done with the Z-score cutoff increased in each step, following the schedule shown in Figure 1-figure supplement 2: first residues with Z<-0.5 are selected for fragment-based rebuilding, followed by -0.3, -0.1, and finally Z<0. Between each cycle, a single iteration of *Relax* is run, in the same manner as the pre-refinement step. At the start of each stage, $Z_{error}^{(i)}$ of a model is re-evaluated as above to avoid refining fixed errors from the previous stage, and residues predicted to be in error are selected. Finally, an additional 200 step Monte Carlo trajectory is run with the $Z_{error}^{(i)}$ computing solely from $Z_{rama}^{(i)}$ to ensure the favorable Ramachandran geometry in models.

***Pre-proline Ramachandran potential*.** Following early experiments, a new term was added to Rosetta that enforces a distinct pre-proline Ramachandran potential, replacing the original 20 different potentials:

$$E_{rama} = P(\varphi_i, \psi_i | AA_i)$$

With 40 different potentials conditioned on the sequence identity of the C-terminal adjacent residue:

$$E_{rama} = P(\varphi_i, \psi_i | AA_i, is\_pro_{i+1})$$

This potential was trained using the Richardson 8000 set of high-resolution crystal structures [15], and smoothed using adaptive kernel density estimates, as with the original Ramachandran potential [26]. They are included in the released Rosetta with the energy term *rama_prepro* (using the same weight as the Rosetta term *rama*). Figure 1-figure supplement 4 illustrates the resulting potentials. For all experiments in this manuscript, this term replaced the default Ramachandran score term in Rosetta.

***Local relax*.** Following our four cycles of refinement, we run a modified version of *Relax*, which we call *LocalRelax*. Modifications were made following the observation that – when applied to very large complexes (800+ residues) – we observed many instances where sidechains were not properly optimized into density, even though the density was very clear. Figure 5-figure supplement 1 shows several such cases.

In *LocalRelax*, small overlapping regions of ~20-100 residues (discontinuous in sequence space) are selected for optimization repeatedly, until the entire protein has been optimized at least once. The approach is based upon the idea of neighbor residues, where residue neighbors are defined as all residues with a Cβ-Cβ distance less than 8Å. We first find the residue r_i with the most residue neighbors. Then we optimize the neighbors of r_i, and the neighbors-of-neighbors of r_i: the neighbors are allowed to optimize both sidechain and backbone conformation, while the neighbors-of-neighbors may only optimize sidechain conformation. This optimization is performed via Monte Carlo sampling of sidechain rotamers, followed by Cartesian minimization of all movable atoms. Following this, all neighbors of r_i (as well as r_i) are marked as visited, and the process repeats, selected a new r_i as the unmarked residue with the most neighbors. This process continues until all residues are marked. In total, 4 cycles of this procedure are carried out, increasing the weight on van der Waals repulsion in each cycle. Finally, following coordinate refinement with *LocalRelax*, we fit atomic B-factors following the scheme of our previous paper [11].

***Sidechain rescaling*.** We compute a scalefactor associated with each sidechain, that describes how much contribution to the density score each sidechain contributes. The values were computing using the 3.3-Å reconstruction of the 20S proteasome [1] and the 3.2-Å reconstruction of β-galactosidase [16]. Models were refined into the density and real-space atomic B-factors were fit for each atom. We then converted the atomic B-factors to scale factors using the following transformation:

$$scale_{AA} \approx \frac{1}{B^{3/2}}$$

Scales were normalized such that the scale for all backbone atoms was equal to 1. To prevent overfitting, each sidechain was grouped into one of three classes, and all sidechains within a given group were given the average scalefactor of the group. Finally, while maintaining the ratio of these three groups with respect to one another, we scaled the relative contribution of backbone versus sidechain density, and selected the best values based on free FSC following refinement. The final values range from 0.66 to 0.78, and are tabulated in Table 2.

Voxel size refinement. To optimize the voxel size of a map used to refine the model, we fix the model coordinates, and compute the model density. We then refine the voxel size $v=[v_x,v_y,v_z]$ and the origin $o=[o_x,o_y,o_z]$ of the map density – fixing these parameters in the model density – to maximize the real-space correlation coefficient between the two:

$$CC(v,o) = \frac{\sum \rho_o(\vec{x}) \cdot \rho_c(\hat{I}_{v,o}(\vec{x})) - \sum \rho_o(\vec{x}) \cdot \sum \rho_c(\hat{I}_{v,o}(\vec{x}))}{\left(\sigma_o^2(\vec{x}) + \sigma_c^2\left(\hat{I}_{v,o}(\vec{x})\right)\right)^{1/2}}$$

$$\hat{I}_{v,o}(x,y,z) = \left(o_x + x/a_x\,, o_y + y/a_y\,, o_z + z/a_z\right)$$

Here, ρ_o refers to the experimental map and ρ_c to the map derived from the model, while σ_o and σ_c refer to the standard deviations over the corresponding density maps. Sums are taken over the entire map. Off-grid density values are computed using cubic splines to interpolate the calculated density map. This function is optimized with respect to the voxel size paramters using l-BFGS minimization; analytic derivatives are computed for CC with respect to *v* and *o*, and the same cubic splines are used to calculate derivatives with respect to the calculated map. Voxel size may be refined isotropically or anisotropically (either 4 or 6 total parameters); all experiments in this manuscript treated this refinement isotropically (that is, all three axes are scaled together).

In this report, although we carried out voxel size refinement for all the three targets, we found only minimal changes in the cases of TRPV1 and Mitoribosome, but the Frh case. For fair comparison, for the cases of TRPV1 and Mitoribosome we report all the model-to-map metrics using maps with the deposited voxel sizes.

Refinement against the full reconstruction and model selection. The previously described protocol was run to generate 5000 independent trajectories. From these 5000 models, a set of 10 representative models is chosen, following the protocol outlined in Figure 1. We want our optimized models to simultaneously be optimal in terms of: a) independent map agreement, b) physically realistic geometry, and c) agreement to the full reconstruction. The

latter is necessary, as the full reconstruction often features details not present in the independent half maps.

Independent-map FSCs were computed against the validation map – subject to the same sharpening scheme as the training map – using the *ComputeFSC* mover in Rosetta. The integrated FSC between 10Å and the reported resolution (3.4Å in all cases) of the map was used to assess agreement with the independent map. The script computes FSC after masking the map with a mask computed from the model and filtered to 12Å with the command line:

density_tools.exe -in:file:s model.pdb -mapfile validation_map.mrc -mask_radius 12 -nresbins 50 -lowres 10 -hires 3.4 -verbose

In the case of the mitochondrial ribsosome, each segmented domain map was evaluated separately. Of the 1000 generated models, the top 50 by independent map agreement are selected.

Next, we want to identify the models from this subset that are the most physically realistic. To do this, all 50 models are rescored with MolProbity [15], and the top 10 are selected. While computing similar features to the Rosetta energy, its slightly different implementation makes it a somewhat orthogonal measure for structure evaluation.

Finally, we want to use features from the full reconstruction to further improve the model, particularly bulky sidechains that may not be visible in the half-map reconstructions. However, when refining against the full reconstruction we need to be careful not to overfit to the full reconstruction, as we no longer have an independent map with which to evaluate overfitting. We use two ideas to avoid overfitting in this case. First, we do not perform any fragment based rebuilding with the full map, and instead only perform two cycles of *LocalRelax* and B-factor refinement with the full map. Second, we use halfmaps to determine the optimal fit-to-density weight when refining against the full map. The weight is selected using the following relation where the weight is chosen to maximize the following:

$$U = FSC_{free} - 0.004 \cdot E_i$$

Here, E_i is the per-residue energy, and is included as additional regularization to avoid overfitting. The value of 0.004 was chosen to normalize the two based on the relative dynamic ranges of both terms.

The top 10 models from the previous selection are subject to refinement against the full map. The final model is then taken as the model with best integrated-FSC against the full reconstruction. Local deviation over all ten models is used to estimate model uncertainty. The per-residue structural variance of ensemble models is calculated using Theseus with the default command line [27].

Assembly of the mitochondrial ribosome. In the case of the mitochondrial ribosome, we refine separate models for each protein subunit. A final assembly step combines the full model. In this final assembly step, all subunits, plus the deposited nucleic acid chains are combined in a single model, and are subject to 2 cycles of *LocalRelax* against the full reconstruction.

EMRinger score calculation. For each of the five models following model selection, EMringer was run using the command:

phenix.emringer MODEL.pdb MAP.ccp4

To calculate per-chain EMRinger scores, pdb files were first segmented by chain ID and then emringer scores were calculated against the segmented pdb files. A script is included to automate the PDB segmentation and calculation of EMRinger scores.

EMRinger scores can be compared absolutely between structures, although model size and local resolution variation are sources of noise for the EMRinger score. Scores below one are indicators of suboptimal model to map agreement for structures better than 4-Å resolution, while a score around zero indicates no improvement beyond randomness.

Phenix real-space refinement. Starting from the deposited model for each of the three targets, real-space refinement was carried out using the Phenix package (v. 2450) with a default setting using the command:

phenix.real_space_refine MODEL.pdb MAP.mrc resolution=3.4

For the case of Frh, ligand files were appended to the above command with cif files generated using the command:

phenix.elbow ligands.pdb

Availability

All methods described are available as part of Rosetta, using weekly releases after week 35, 2016. The Rosetta XML files and flags for running all the refinements discussed in this manuscript are included as Supplemental Data Files 1-3. The scripts and the tutorial used for running the method described here is available now at the website of the corresponding author (https://faculty.washington.edu/dimaio/files/density_tutorial_sept15_2.pdf).

Acknowledgements

The authors thank Drs. Alan Brown, Alexy Amunts and Venki Ramakrishnan for sharing the half maps of mitoribosomal large subunit (EMD-2762) with us; the author especially thank Dr. Alan Brown on providing helpful comments on the Rosetta refined mitoribosome, in which the suggestions led to the new development on better optimizing sidechains in very large protein complexes. The author thanks Drs. Metteo Allegretti and Janet Vonck for sharing the half maps of Frh (EMD-2513) with us. The authors thank Dr. Erhu Cao for commenting on the refined TRPV1 model initially. The authors thank Dr. Vikram Mulligan for helping on using the "-*auto_setup_metals*" module he developed on facilitating ligand setup in modeling Frh using Rosetta. R.Y.-R.W. thanks Dr. David Baker for funding support and suggestions of the research.

Contributions

R.Y.-R.W. performed the research. R.Y.-R.W. and F.D. developed the method and prepared the manuscript. R.Y.-R.W., Y.S. and F.D. conceived and designed the research. F.D. supervised the research. B.A.B and J.S.F. performed the EMringer analysis. Y.C. provided the TRPV1 half-map data set with various b-factor sharpening, and analyzed and interpreted the results on the TRPV1 refinement. All authors edited the manuscript.

References

1. Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, Agard DA, Cheng Y: **Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-em.** *Nature methods* (2013) **10**(6):584-590.

2. Bai XC, Fernandez IS, McMullan G, Scheres SH: **Ribosome structures to near-atomic resolution from thirty thousand cryo-em particles.** *eLife* (2013) **2**(e00461.

3. Scheres SH: **A bayesian view on cryo-em structure determination.** *Journal of molecular biology* (2012) **415**(2):406-418.

4. Kudryashev M, Wang RY, Brackmann M, Scherer S, Maier T, Baker D, DiMaio F, Stahlberg H, Egelman EH, Basler M: **Structure of the type vi secretion system contractile sheath.** *Cell* (2015) **160**(5):952-962.

5. Wang RY, Kudryashev M, Li X, Egelman EH, Basler M, Cheng Y, Baker D, DiMaio F: **De novo protein structure determination from near-atomic-resolution cryo-em maps.** *Nature methods* (2015) **12**(4):335-338.

6. DeLaBarre B, Brunger AT: **Considerations for the refinement of low-resolution crystal structures.** *Acta crystallographica Section D, Biological crystallography* (2006) **62**(Pt 8):923-932.

7. Brunger AT, DeLaBarre B, Davies JM, Weis WI: **X-ray structure determination at low resolution.** *Acta crystallographica Section D, Biological crystallography* (2009) **65**(Pt 2):128-133.

8. Emsley P, Lohkamp B, Scott WG, Cowtan K: **Features and development of coot.** *Acta crystallographica Section D, Biological crystallography* (2010) **66**(Pt 4):486-501.

9. Brown A, Long F, Nicholls RA, Toots J, Emsley P, Murshudov G: **Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions.** *Acta crystallographica Section D, Biological crystallography* (2015) **71**(Pt 1):136-153.

10. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD: **Towards automated crystallographic structure refinement with phenix.Refine.** *Acta crystallographica Section D, Biological crystallography* (2012) **68**(Pt 4):352-367.

11. DiMaio F, Song Y, Li X, Brunner MJ, Xu C, Conticello V, Egelman E, Marlovits TC, Cheng Y, Baker D: **Atomic-accuracy models from 4.5-a**

cryo-electron microscopy data with density-guided iterative local refinement. *Nature methods* (2015) **12**(4):361-365.

12. Liao M, Cao E, Julius D, Cheng Y: **Structure of the trpv1 ion channel determined by electron cryo-microscopy.** *Nature* (2013) **504**(7478):107-112.

13. Allegretti M, Mills DJ, McMullan G, Kuhlbrandt W, Vonck J: **Atomic model of the f420-reducing [nife] hydrogenase by electron cryo-microscopy using a direct electron detector.** *eLife* (2014) **3**(e01963.

14. Brown A, Amunts A, Bai XC, Sugimoto Y, Edwards PC, Murshudov G, Scheres SH, Ramakrishnan V: **Structure of the large ribosomal subunit from human mitochondria.** *Science* (2014) **346**(6210):718-722.

15. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC: **Molprobity: All-atom structure validation for macromolecular crystallography.** *Acta crystallographica Section D, Biological crystallography* (2010) **66**(Pt 1):12-21.

16. Bartesaghi A, Matthies D, Banerjee S, Merk A, Subramaniam S: **Structure of beta-galactosidase at 3.2-a resolution obtained by cryo-electron microscopy.** *Proceedings of the National Academy of Sciences of the United States of America* (2014) **111**(32):11709-11714.

17. Campbell MG, Veesler D, Cheng A, Potter CS, Carragher B: **2.8 a resolution reconstruction of the thermoplasma acidophilum 20s proteasome using cryo-electron microscopy.** *eLife* (2015) **4**(

18. Jackson RN, McCoy AJ, Terwilliger TC, Read RJ, Wiedenheft B: **X-ray structure determination using low-resolution electron microscopy maps for molecular replacement.** *Nature protocols* (2015) **10**(9):1275-1284.

19. DiMaio F, Zhang J, Chiu W, Baker D: **Cryo-em model validation using independent map reconstructions.** *Protein science : a publication of the Protein Society* (2013) **22**(6):865-868.

20. Fernandez IS, Bai XC, Murshudov G, Scheres SH, Ramakrishnan V: **Initiation of translation by cricket paralysis virus ires requires its translocation in the ribosome.** *Cell* (2014) **157**(4):823-831.

21. Barad BA, Echols N, Wang RY, Cheng Y, DiMaio F, Adams PD, Fraser JS: **Emringer: Side chain-directed model and map validation for 3d cryo-electron microscopy.** *Nature methods* (2015) **12**(10):943-946.

22. Chuang HH, Lin S: **Oxidative challenges sensitize the capsaicin receptor by covalent cysteine modification.** *Proceedings of the National Academy of Sciences of the United States of America* (2009) **106**(47):20097-20102.

23. Kucukelbir A, Sigworth FJ, Tagare HD: **Quantifying the local resolution of cryo-em density maps.** *Nature methods* (2014) **11**(1):63-65.

24. Headd JJ, Echols N, Afonine PV, Grosse-Kunstleve RW, Chen VB, Moriarty NW, Richardson DC, Richardson JS, Adams PD: **Use of knowledge-based restraints in phenix.Refine to improve macromolecular refinement at low resolution.** *Acta crystallographica Section D, Biological crystallography* (2012) **68**(Pt 4):381-390.

25. Conway P, Tyka MD, DiMaio F, Konerding DE, Baker D: **Relaxation of backbone bond geometry improves protein energy landscape modeling.** *Protein science : a publication of the Protein Society* (2014) **23**(1):47-55.

26. Ting D, Wang G, Shapovalov M, Mitra R, Jordan MI, Dunbrack RL, Jr.: **Neighbor-dependent ramachandran probability distributions of amino acids developed from a hierarchical dirichlet process model.** *PLoS computational biology* (2010) **6**(4):e1000763.

27. Theobald DL, Wuttke DS: **Accurate structural correlations from maximum likelihood superpositions.** *PLoS computational biology* (2005) **preprint**(2008):e43.

Figure Legends

Figure 1. An overview of the three stages of automated refinement. (Left) In stage 1, problematic regions are predicted using a newly developed error predictor judging on local strain in the model and poor local density fit. These selected regions are subject to iterative fragment-based rebuilding with in a Monte Carlo sampling trajectory. Refinement in this stage is restricted to using one-half of the data, referred to as the training map. (Middle) In stage 2, the best models from the ~5000 independent Monte Carlo trajectories are selected. Models are selected based on: agreement to the validation map (independently constructed from the other half of the data), then by model geometry as assessed by MolProbity, and finally, based on agreement to the full reconstruction. At this point, the selected models should in general have good fit-to-density and good geometry without overfitting to the data. (Right) In stage 3, using the 10 best models selected, we then optimize against the full reconstruction. Two half maps are used for choosing the optimal density weight to refine structures using full-reconstruction. Finally, these top 10 models are optimized (without large scale backbone rebuilding) into the full-reconstruction, which alternates with voxel size refinement iteratively. Finally, these models are subject to B-factor refinement.

Figure 1-figure supplement 1. A closeup view of model strain indicating errors in density-optimized TRPV1 models using the previous Rosetta approach. Both insets show two regions of models refined by the previous approach, where strain can indicate errors in models. In both cases, phenylalanine sidechains fit the density well, but both show geometric strain around the C**β** atom. The type of strain (as evaluated by MolProbity) is indicated by model color, using the key on the right.

Figure 1-figure supplement 2. Incorporating model strain improves error detection. Guided by the 3.3-Å 20S proteasome reconstruction, we evaluated 500 models against the high-resolution crystal structure. We plot here the precision (y-axis) and recall of predicting which residues were incorrectly placed (RMS > 1Å). Using density alone (pink line) is outperformed by using a combination of density and model strain (blue line). Our refinement approach considers four points on this curve when picking density + model strain cutoffs, indicated on the plot with "Stage1-4".

Figure 1-figure supplement 3. Density weight optimization against halfmaps for Mitoribosome. Before refinement against the full reconstruction, we optimize the weight on the "fit-to-density" energy using half maps, to avoid overfitting. We plot several key metrics here as a function of weight on the fit-to-density score term (X-axis), including the FSC "overfitting" (FSCwork - FSCfree, top), the

Rosetta energy (row 2), and several Molprobity model geometry terms (rows 3-6). In all cases, we see a sharp inflection point where overfitting increases and geometry gets notably worse. As a general rule-of-thumb, we use the weight maximizing FSCfree-0.04*per-residue-energy to capture this inflection point.

Figure 1-figure supplement 4. Model geometry is improved with a separate pre-proline potential. It was found that refined models initially had poor pre-proline geometry. Thus a new backbone torsional potential was created which separately treats pre-proline and pre-non-proline residues. In the plot above we show the old potential (left), the new pre-non-proline potential (middle), and the pre-proline potential (right), for three different residue identities. The color indicates the unweighted energy values, using the key on the right.

Figure 2. The accuracy of voxel size refinement and the effect of B-factor sharpening in Rosetta refinement. (A) Voxel size refinement on perturbed models. Perturbed structures were generated by running short MD trajectories in Rosetta, followed by all-atom minimization. Voxel size is refined against the perturbed models, yielding the density distribution in red. Following cycles of iterated voxel refinement and all-atom refinement, the voxel size shows significantly better convergence (blue line). (B) Rosetta structure refinement with a range values of B-factor sharpening. We plot here free-iFSC as a function of B-factor sharpening of the training map. The results indicate that our refinement method is not particularly sensitive to the extent of B-factor sharpening, behaving similarly over a range of sharpening values between -40 and -90. The error bars show standard deviation of the free-iFSC among the top10 ensemble models (see *Methods* for the ensemble selection method).

Figure 3. Refinement of the apo TRPV1 channel (EMD-5778) shows improved model quality. (A) A comparison of the deposited and Rosetta-refined models, as assessed by MolProbity. Residues reported as violations are colored using the key shown in the far right. Blue open arrows indicate that hydrogen-bond geometry of a β-hairpin was automatically detected and improved in the Rosetta refined model. (B) An overlay the asymmetric unit of the deposited (pink) and Rosetta-refined (green) model indicates the magnitude of conformational changes that are explored by our refinement approach. (C) The agreement of models to map assessed by Fourier space correlation (Y-axis) at each resolution shell (X-axis), where the reported resolution (3.4Å) is depicted in a dashed line colored in orange. The deposited model is shown in the curve with pink color, while the Rosetta refined model is shown in the curve colored in green.

Figure 4. Refinement of the TRPV1 channel identifies a previously unmodelled disulfide bond. (A) An overview of the entire structure, estimating local model uncertainty in two ways: local structural diversity and refined B-factors. Local structure diversity is indicated by showing an overlay of the top 10 Rosetta models (left), the top model colored by per residue deviation (middle), and the refined per-atom B-factors (right). Using the model selection method illustrated in Figure 1 (middle panel), the Cα RMSDs among the selected ensemble range from 0.44 to 0.63 Å. The orange square shows the location of a newly identified disulfide bond (C386-C390) revealed by our refinement protocol. (B) A zoomed-in view of the disulfide linkage (C386-C390) identified by the automated method. Note that the sidechain coordinates of C390 were unassigned in the deposited model; for presentation, the sidechain atoms of C390 were optimally added by Rosetta based on the deposited backbone torsion angles of C390.

Figure 5. Refinement of the F_{420}-reducing [NiFe] hydrogenase (EMD-2513) improves the model geometry. (A) An illustration comparing the model geometry of the deposited (upper panel)and Rosetta refined (lower panel) models. Three chains (A/B/C) of the asymmetric unit of the complex are shown as cartoon with geometry violations reported by MolProbity colored with the key shown in the far right. Four iron-sulfur clusters [4Fe4S] and an FAD are shown in a stick representation. Metal ions are depicted as spheres, with Zn grey, Fe orange, and Ni green . (B) Model-map agreement – as assessed by Fourier shell correlation (Y-axis) as a function of resolution (X-axis) – quantifies this improvement following voxel size refinement. The pink curve corresponds to the deposited model; the green curve corresponds to a model refined by Rosetta. (C) Model quality as assessed by EMRinger and MolProbity. The X-axis shows methods used to evaluate the models, while the Y-axis shows the scores under each criterion.

Figure 5-figure supplement 1. The symmetry operators denoted in the deposited PDB (PDB 4ci0) produce a complex that could not fit into the deposited density map properly. (Left panel) The symmetric complex downloaded from protein data bank as biounit shifts the entire complex out of the deposited density map. The middle and right panels shows a zoomed-in view of two regions in the deposited models corresponding to the helix and the sheet indicated by the orange and cyan squares on the left panel.

Figure 6. Refinement of the large subunit of the human mitochondrial ribosome (EMD-2762) shows improvements to all subunits. (A) Scatterplots of model quality of each of the 48 protein chains compare the deposited (X-axis) and Rosetta (Y-axis) models using MolProbity. On the left, the MolProbity score of all 48 protein chains are compared, where lower values indicate better model geometry. On the right, the percentage of “Ramachandran favored” residues are

compared on each chain, with higher values preferable. (B) An evaluation of the fit-to-density of each protein chain. On the left, we compare the Fourier shell correlation (FSC) of each chain before and after refinement; we integrate the FSC from 10Å to 3.4Å. Higher values indicate better agreement with the data. The largest improvement, chain k, is indicated by the red arrow. On the right, the full FSC curve is shown, with the deposited model shown in pink, and the Rosetta refined model shown in green; the reported map resolution (3.4Å) is indicated in the dashed orange line. (C) A zoomed-in view indicating the large radius of convergence of the refinement of chain k. The left panel shows the density for chain k is in a region of relatively low local resolution.

Figure 6-figure supplement 1. Local relax shows better placement of sidechains for large systems. In the case of mitoribosome, refinement of a particularly well-resolved region in the map (left) led to sidechains clearly misaligned with the density (middle). This was due to the poor convergence of our Monte Carlo sidechain placing approach when applied to systems with >1000 residues. Our alternative approach, LocalRelax, which instead performs many local sidechain optimizations, correctly places sidechains consistent with density (right).

Figure 6-figure supplement 2. EMRinger analysis on refinement of the large subunit of the human mitochondrial ribosome. A scatterplot of model quality assessed by EMringer of each of the 48 protein chains compares the deposited (X-axis) and Rosetta (Y-axis) models.

Supplemental Data File Legends

Supplemental Data File 1. Input files to carry out the TRPV1 structure refinement described in the manuscript. Structure refinement of TRPV1 using Rosetta involves two steps: 1) refinement of only the transmembrane regions, and 2) local refinement of the full system, including the Ankyrin repeat domains. The package includes two folders, one for each of the two steps with the command lines and input files necessary to run TRPV1 structure refinement.

Supplemental Data File 2. Input files to carry out the Frh structure refinement described in the manuscript. Structure refinement of Frh using Rosetta involves three steps: 1) refinement of the asymmetric unit without ligands present, 2) local refinement of the asymmetric unit with the ligands present, and 3) local refinement the full symmetric complex with ligands present. The package includes three folders, one for each of the three steps with the command lines and input files necessary to run Frh structure refinement.

Supplemental Data File 3. Input files to carry out the Mitoribosome structure refinement described in the manuscript. Structure refinement of the case of Mitoribosome using Rosetta involves in two steps: 1) refinement of individual chains, and 2) local refinement the whole assembly. The package includes two folders, one for each of the two steps with the command lines and input files necessary to run Mitoribosome structure refinement.

Table 1. Structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta

	EMD ID	PDB ID	Reported resolution [Å]	Symmetry	Number of amino acid[a]	MolProbity[b]				EMRinger Score[b]	iFSC[c]
						Score	Clashscore	Rotamer outliers [%]	Ramachandran favored [%]		
TRPV1	5778	3j5p	3.4	C4	489 (1956)	3.81 / 1.45	86.35 / 1.96	28.78 / 0.00	95.65 / 91.93	0.65 / 2.34	0.612 / 0.607
Frh	2513	4ci0	3.4	T	893 (10716)[d]	3.98 / 1.59	120.42 / 3.22	39.11 / 0.27	96.51 / 92.18	1.06 / 2.17	0.743 / 0.708
Mitoribosome	2762	3j7y	3.4	N/A	7469[e]	2.71 / 1.50	8.38 / 3.51	8.49 / 0.08	89.86 / 94.86	2.09 / 2.40	0.692 / 0.676

a. Number of protein residues in the asymmetric unit and (the total residues) modelled.
b. Scores from deposited (left) versus (/) Rosetta refined (right) model.
c. Integrated Fourier shell correlation (iFSC) from 10–3.4Å resolution shells.
d. In addition to protein residues, 9 residues of ligand per asymmetric unit–including a [NiFe] cluster, two metal ions (Fe and Zn), and four [4Fe4S] clusters, and an FAD–were included in the refinement.
e. In addition to protein residues, 1529 base pairs of RNA molecule were included in the refinement.

Table 2
Sidechain scaling factors used in automated Rosetta structure refinement

Sidechain	Raw data	Factor used
ARG	0.84	0.66
LYS	0.84	0.66
GLU	0.85	0.66
MET	0.87	0.66
ASP	0.88	0.66
CYS	0.87	0.71
GLN	0.89	0.71
HIS	0.91	0.71
ASN	0.91	0.71
THR	0.94	0.71
SER	0.95	0.71
TYR	0.95	0.78
TRP	0.96	0.78
ALA	0.97	0.78
PHE	0.98	0.78
PRO	0.98	0.78
ILE	0.99	0.78
LEU	0.99	0.78
VAL	1.00	0.78

Table 3. Comparison of structure refinement results between Rosetta and phenix.real_space_refine[a]

	RSCC[a,b,c] validation map	iFSC[a,b,d] validation map	EMRinger Score[a,b] validation map	MolProbity[b]				Number of residues with better RSCC[b,e]
				Score	Clashscore	Rotamer outliers [%]	Ramachandran favored [%]	
TRPV1	0.785 / 0.790	0.546 / 0.566	1.84 / 1.90	1.59 / 1.48	4.30 / 2.14	0.00 / 0.00	94.41 / 91.72	86 / 250
Frh	0.835 / 0.835	0.504 / 0.517	1.36 / 1.27	1.68 / 1.62	7.99 / 3.66	0.68 / 0.13	96.31 / 92.67	677 / 1328
Mitoribosome	0.832 / 0.832	0.476 / 0.478	2.05 / 1.98	1.88 / 1.62	6.17 / 4.08	0.38 / 0.00	90.19 / 93.49	415 / 564

a. To avoid over-fitting, refinement using both methods was carried out using the half-map approach, in which the models were subject to refinement using the training maps. The results showing here were evaluated using the validation-maps. The input model information is the same as reported at Table 1.
b. Numbers (scores) from *phenix.real_space_refine* (left) versus (/) Rosetta refined (right) model.
c. Real-space correlation coefficients were evaluated using UCSF Chimera.
d. Integrated Fourier shell correlation (iFSC) from 10–3.4Å resolution shells.
e. We calculate per-residue real-space correlation coefficient and report the number of residues which show the value of ΔRSCC greater than 0.05.

Figure 1

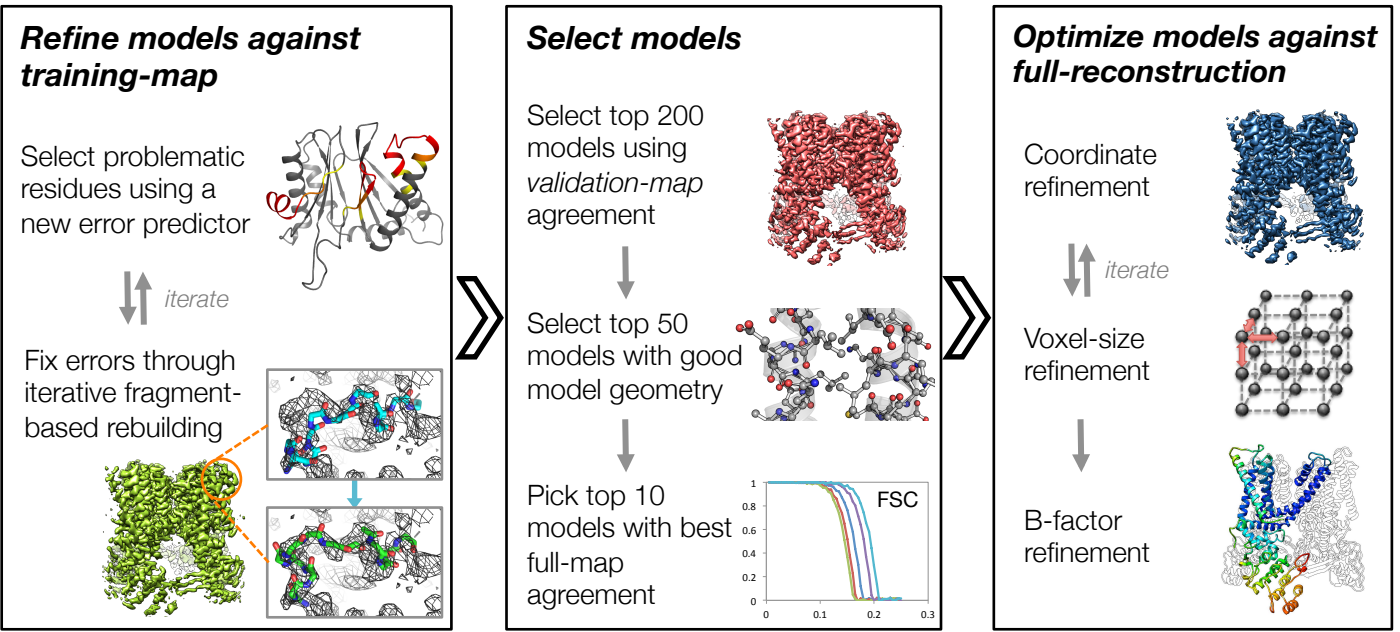
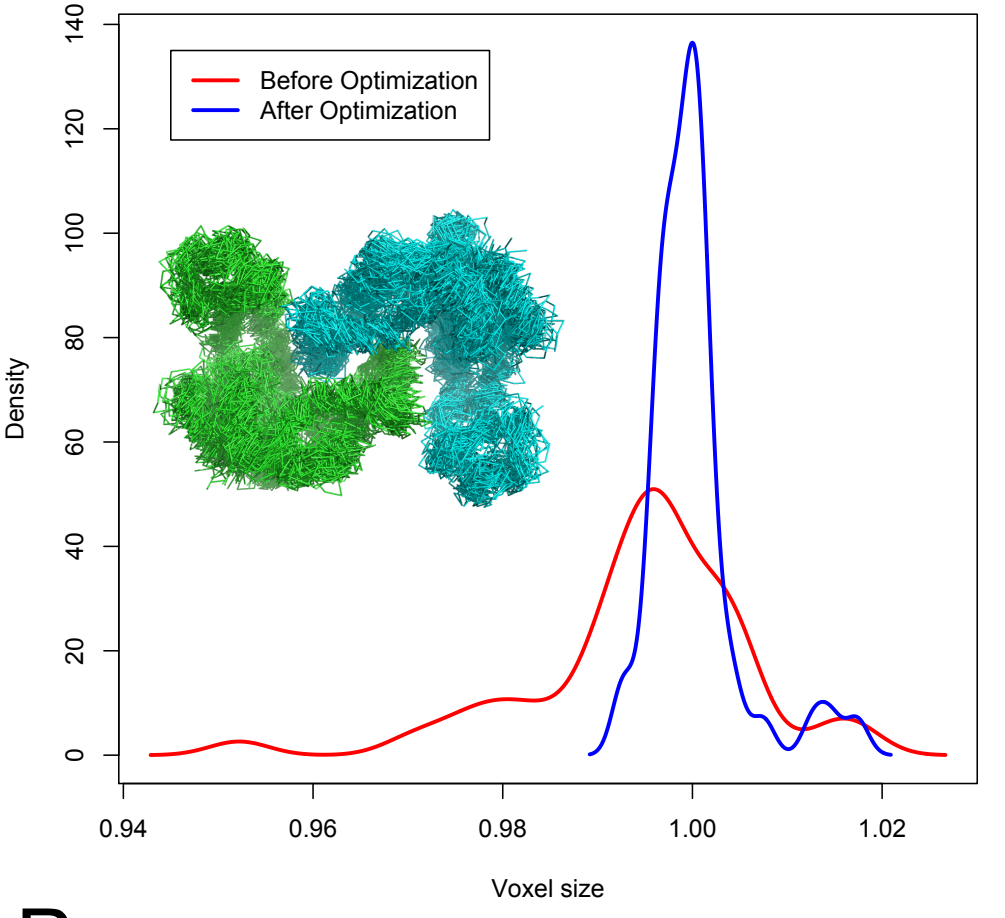

Refine models against training-map
Select problematic residues using a new error predictor
iterate
Fix errors through iterative fragment-based rebuilding
Select models
Select top 200 models using validation-map agreement
Select top 50 models with good model geometry
Pick top 10 models with best full-map agreement
FSC
1
0.8
0.6
0.4
0.2
0
0
0.1
0.2
0.3
Optimize models against full-reconstruction
Coordinate refinement
iterate
Voxel-size refinement
B-factor refinement

Figure 2

A


Before Optimization
After Optimization
Density
0
20
40
60
80
100
120
140
0.94
0.96
0.98
1.00
1.02
Voxel size

B

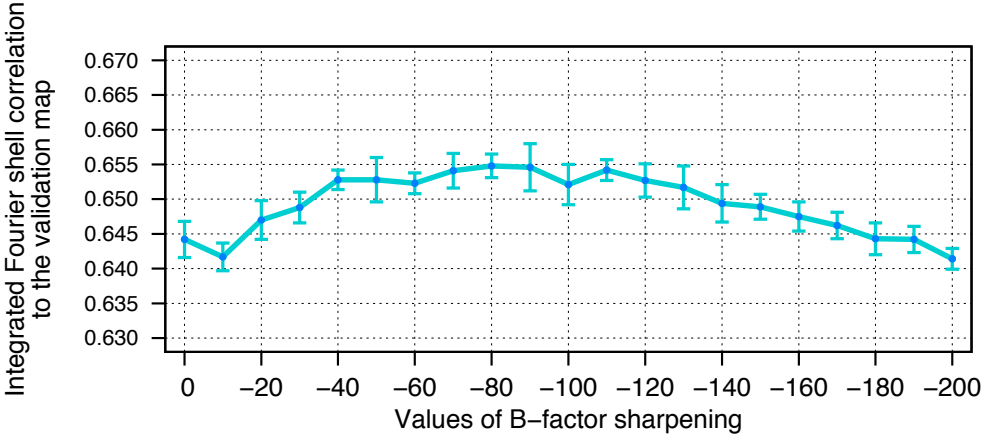
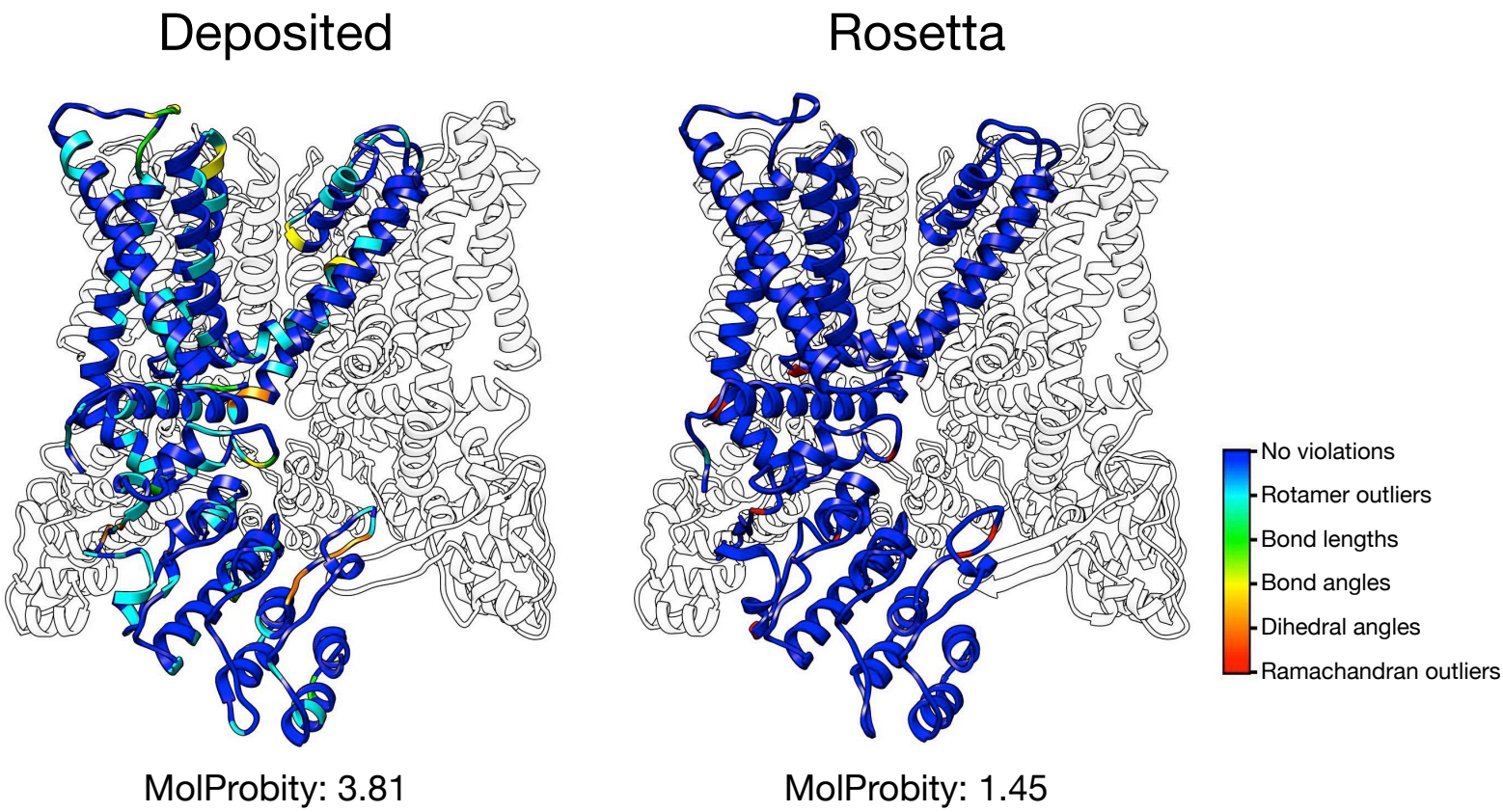

Integrated Fourier shell correlation to the validation map
0.670
0.665
0.660
0.655
0.650
0.645
0.640
0.635
0.630
0
−20
−40
−60
−80
−100
−120
−140
−160
−180
−200
Values of B−factor sharpening

Figure 3

A


Deposited
Rosetta
No violations
Rotamer outliers
Bond lengths
Bond angles
Dihedral angles
Ramachandran outliers
MolProbity: 3.81
MolProbity: 1.45

B


Overlay

C

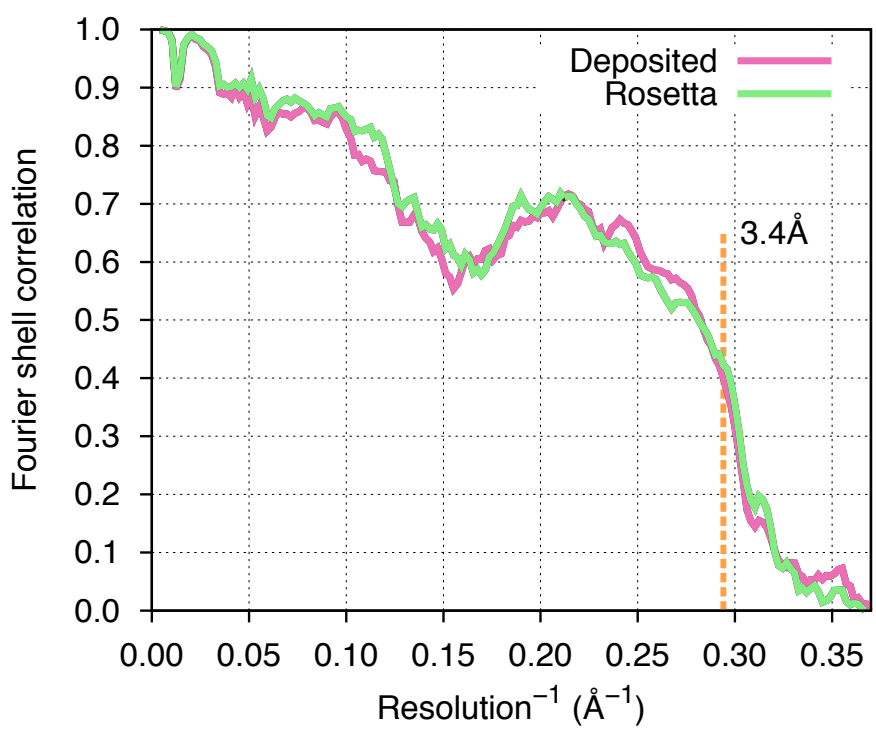

Deposited
Rosetta
3.4Å
Fourier shell correlation
Resolution^{-1} (Å^{-1})
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35

Figure 4

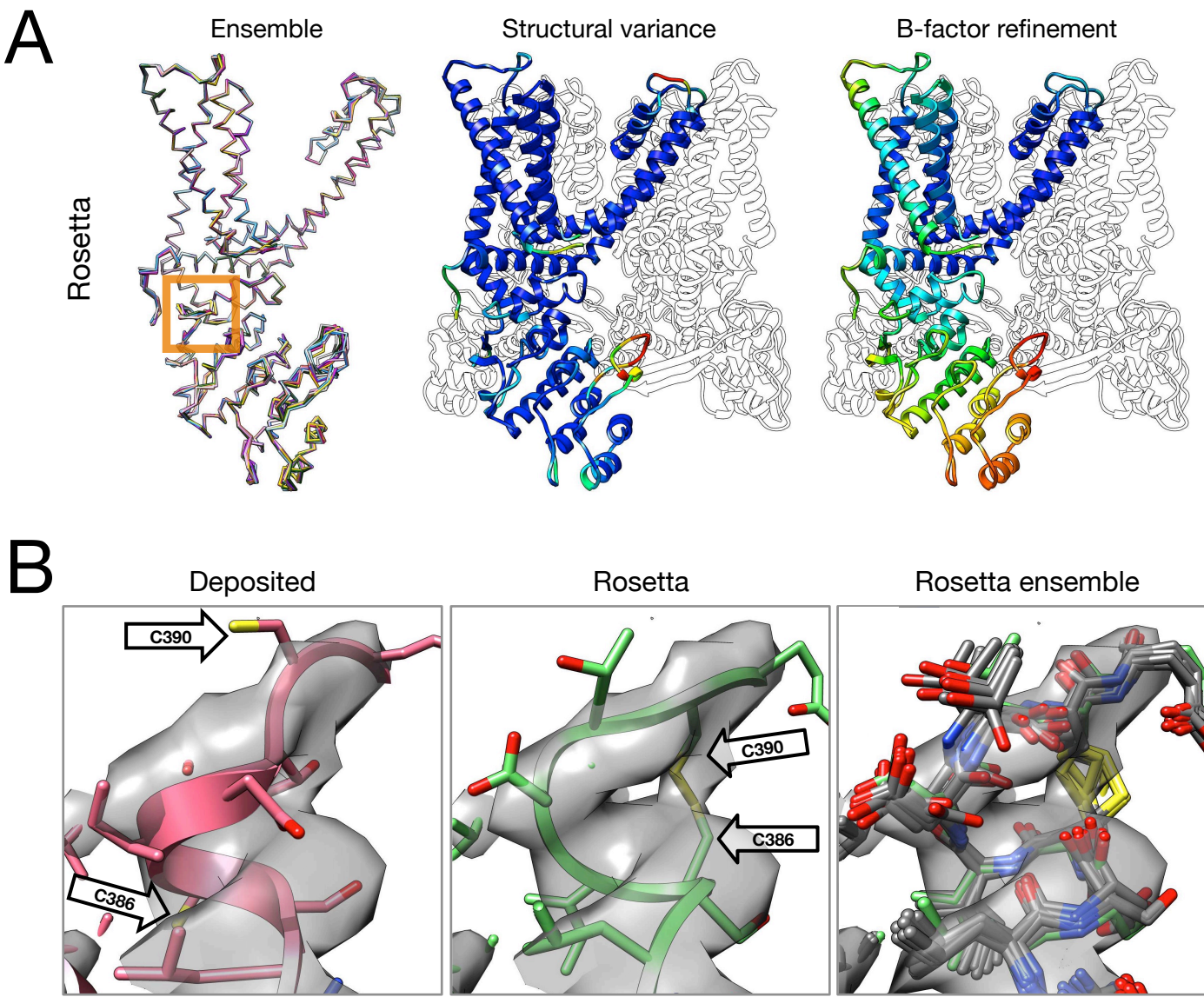

A
Ensemble
Structural variance
B-factor refinement
Rosetta
B
Deposited
Rosetta
Rosetta ensemble
C390
C386
C390
C386

Figure 5

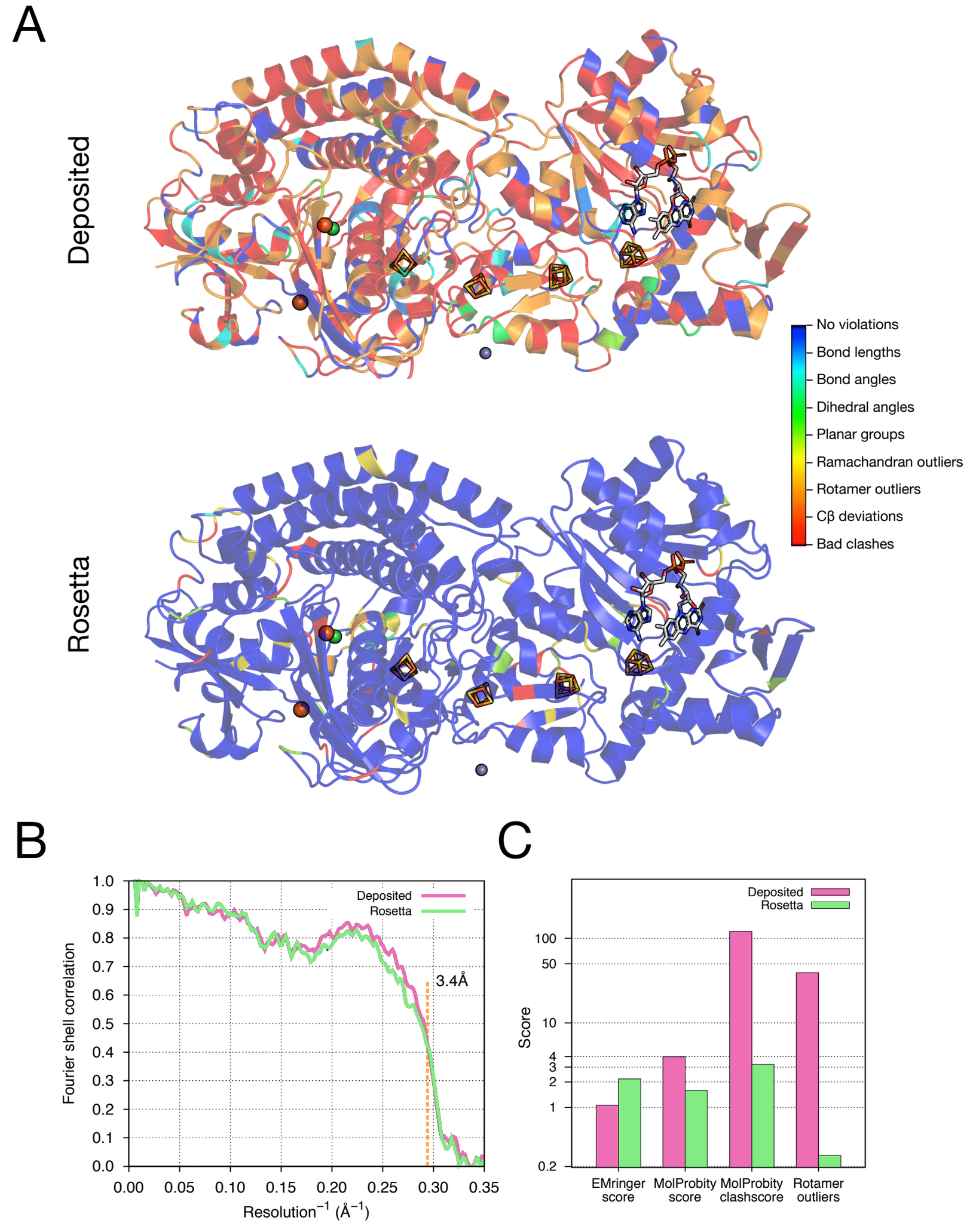

A
Deposited
Rosetta
No violations
Bond lengths
Bond angles
Dihedral angles
Planar groups
Ramachandran outliers
Rotamer outliers
Cβ deviations
Bad clashes
B
Fourier shell correlation
Deposited
Rosetta
3.4Å
Resolution^{-1} (Å^{-1})
C
Score
Deposited
Rosetta
EMringer score
MolProbity score
MolProbity clashscore
Rotamer outliers

Figure 6

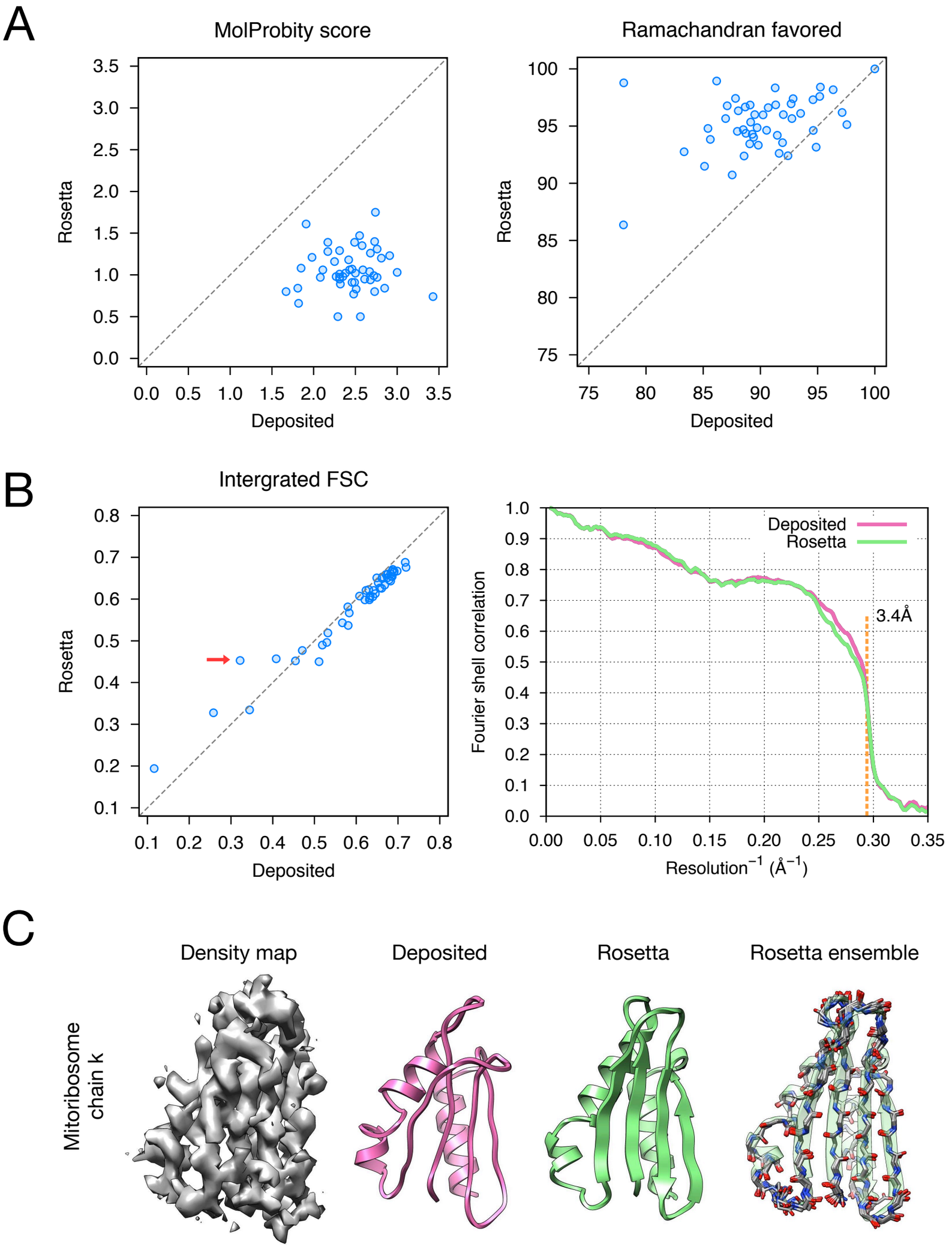
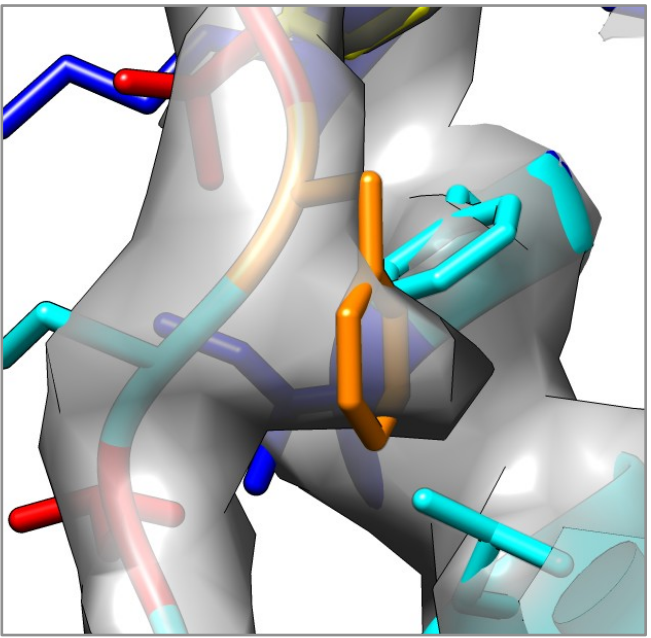
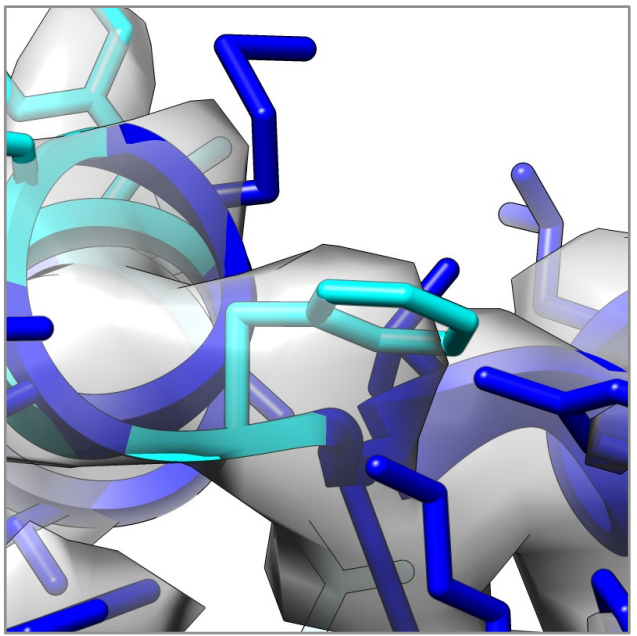

A
MolProbity score
Rosetta
Deposited
Ramachandran favored
Rosetta
Deposited
B
Intergrated FSC
Rosetta
Deposited
Deposited
Rosetta
3.4Å
Fourier shell correlation
Resolution^{-1} (Å^{-1})
C
Density map
Deposited
Rosetta
Rosetta ensemble
Mitoribosome chain k

Figure1-figure supplement 1





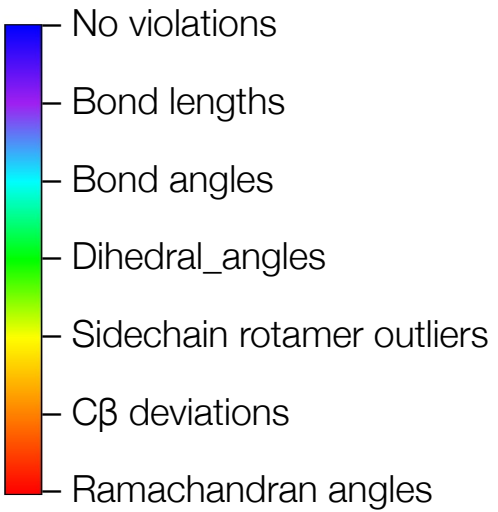


Figure 1-figure supplement 2

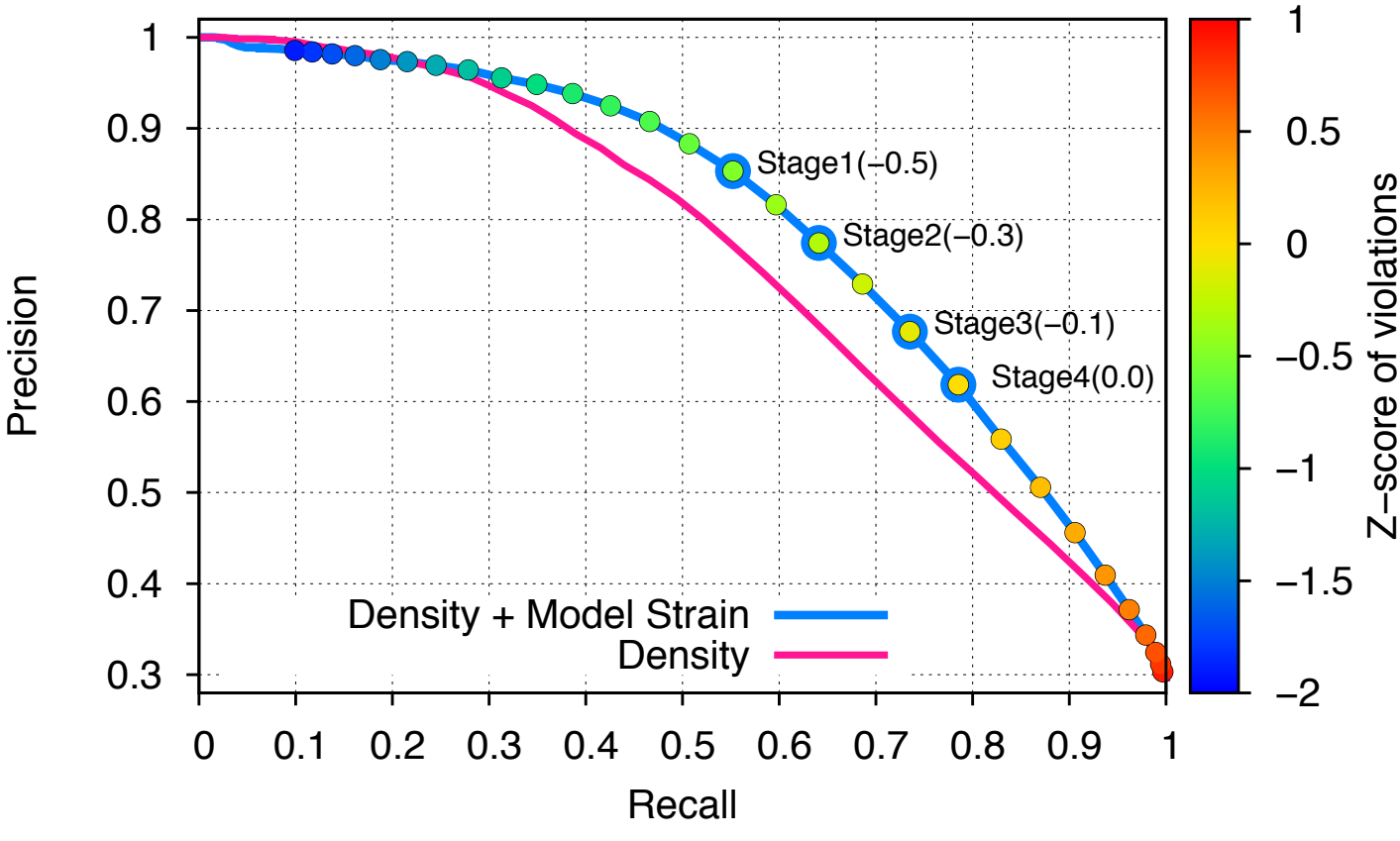


Figure 1-figure supplement 3

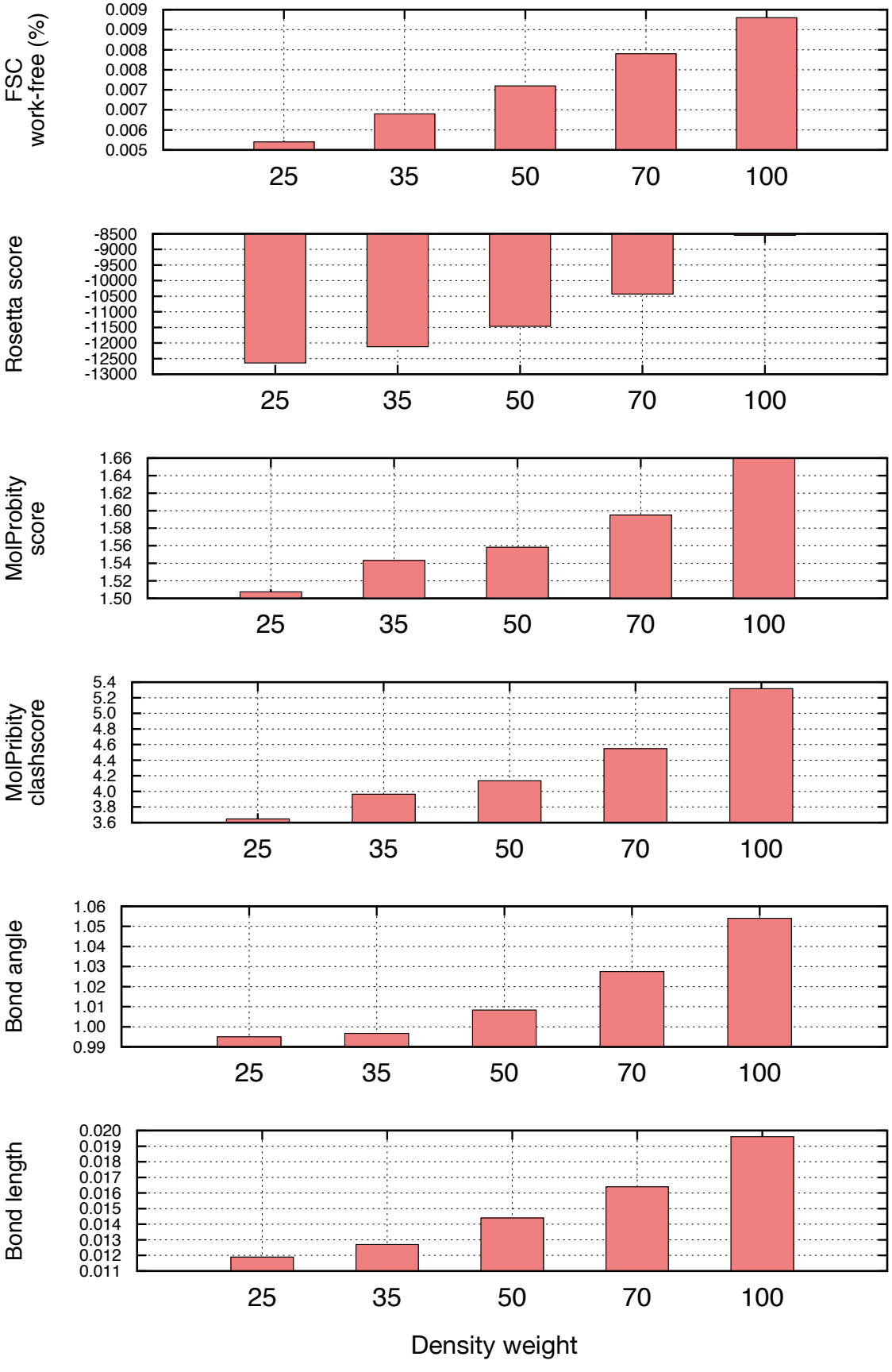


Figure 1-figure supplement 4

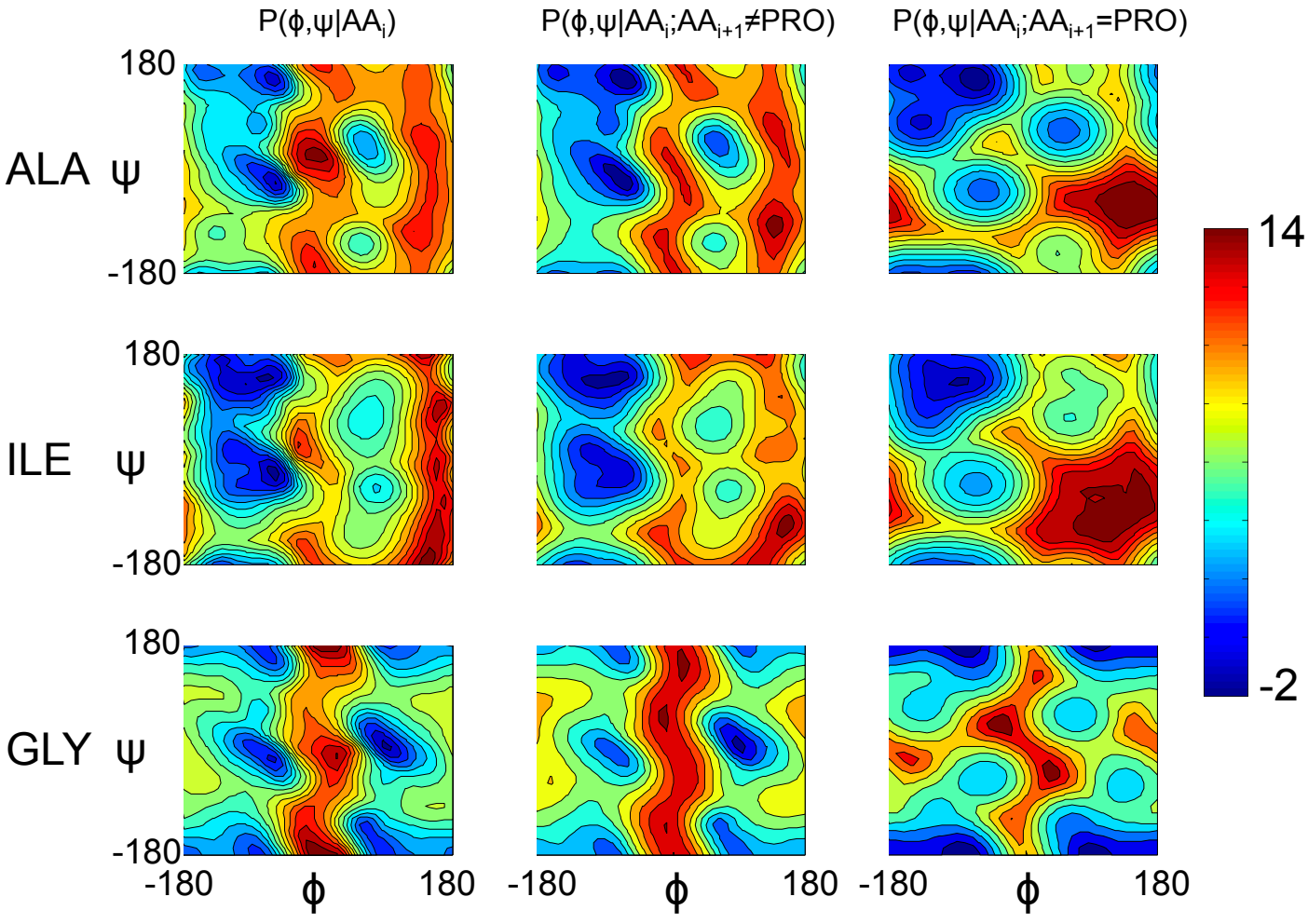


Figure 5-figure supplement 1

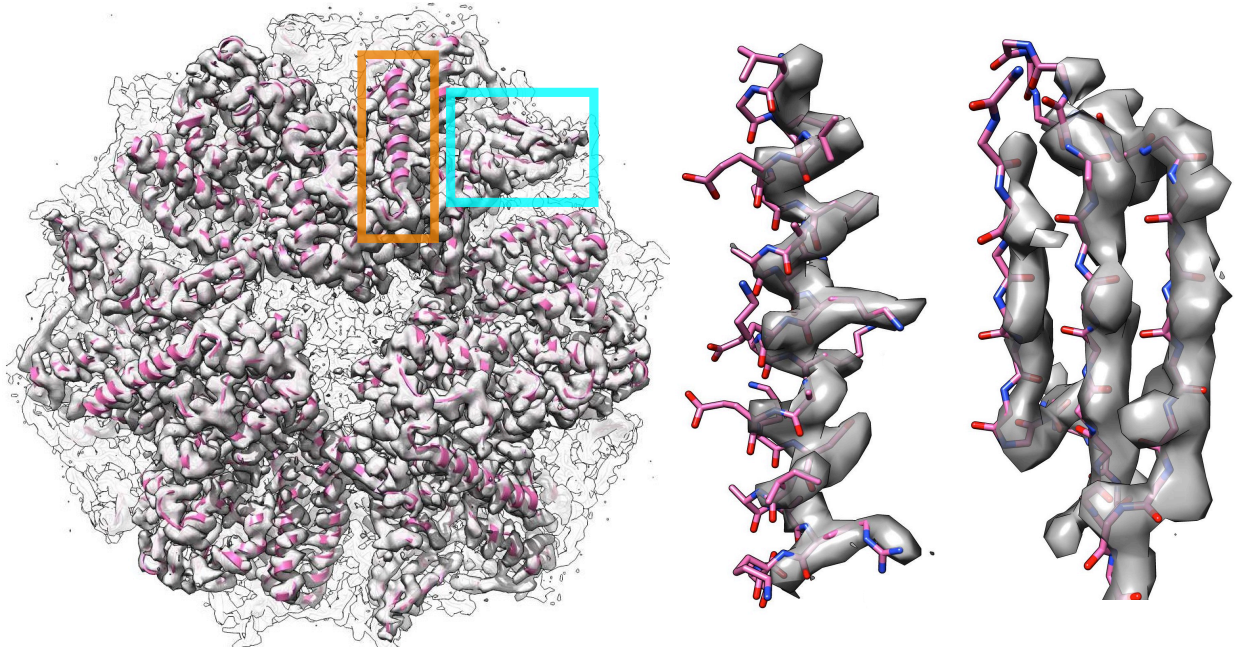


Figure 6-figure supplement 1

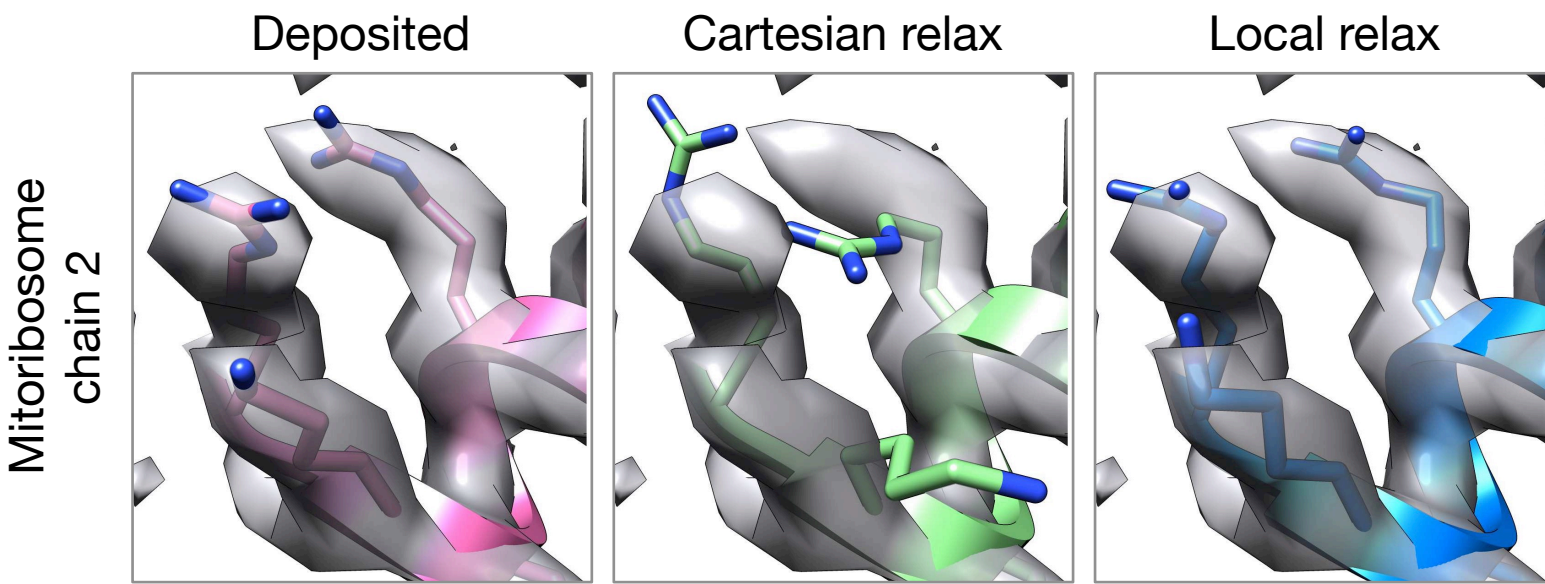


Figure 6-figure supplement 2