Development of Tools for the Interpretation of Cryo-EM data

Gabriella Reggiano

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading committee:

Frank DiMaio, Chair

Justin Kollman

Charles Asbury

Program Authorized to Offer Degree:

Biochemistry

University of Washington

Abstract

Development of Tools for the Interpretation of Cryo-EM Data

Gabriella Reggiano

Chair of the Supervisory Committee:

Frank DiMaio

Department of Biochemistry

In this dissertation, I describe my efforts to build tools to address two gaps in the field of cryo-electron microscopy: deriving structural details about the conformational landscape from cryo-EM data and model validation for moderate resolution cryo-EM maps. Currently, there are few model validation metrics that can precisely evaluate the local quality of atomic models built into maps solved to the resolutions common for cryo-EM. I developed MEDIC (Model Error Detection in Cryo-EM), a robust statistical model to identify local errors in protein structures built into cryo-EM maps. In the second half of this dissertation, I describe my efforts to use atomic models to guide single particle analysis of cryo-EM datasets to obtain a mechanistic understanding of the protein conformational space. Revealing the protein conformational landscape contained in a cryo-EM dataset is notoriously difficult as individual 2D images have a very low signal-to-noise ratio. State of the art methods are only capable of resolving a few very distinct states or describing the motion at low resolutions.

Table of Contents

List of Figures

List of Supplemental Figures

List of Tables

Acknowledgements

It's a gift to be able to look back at myself five years ago and so clearly see my intellectual and personal growth. There are many people to thank for that, but the largest contribution belongs to Frank. He is incredibly generous with his time: always willing to look through code with you, to critique a presentation slide by slide, to discuss the minute details of Rosetta. More importantly, Frank is always enthusiastic. I have come to him many (many) times with bad data and always left his office invigorated and full of ideas. I don't think I would be writing this now if not for his optimism about my middling results (not everything in this thesis is middling, I promise) and his kindness in allowing me to switch projects halfway through my PhD.

I also want to thank my committee: Justin Kollman, Chip Asbury, David Veesler and David Baker, whose thoughtful discussion about my research strengthened my work. I especially want to thank Justin, who allowed me to drop in on his group meetings when I needed extra help from a microscopist, and Chip, whose support (and wonderful feedback at every BPSD seminar) has given me some of the confidence I needed to finish out this PhD.

The BPSD program wouldn't run without Erin Kirschner, who has a bajillion jobs and still manages to make sure that no student falls through the cracks. She was a wonderful resource for navigating the department and to chat with as I was figuring out my career path. Thank you, Erin, for always making personal time for students despite your busy schedule.

To my friends and coworkers in the DiMaio lab: Daniel Farrell, Guangfeng Zhou, Carson Adams, Andrew Muenks, Marisa Brandys, and Ryan McHugh, I am very thankful that I got to come into the office and start arguments, and take quizzes on wax figurines, and put together tier lists. I recently read an article that said one of the greatest determinants of your happiness at work is not what you do, but who you are doing it with, and working with all of you has brought me so much joy.

I want to thank my friends, who have listened to me rant, made me laugh, and reminded me of my worth outside of my career. I'm especially thankful to the friends I made because of this program: Inez Pranoto, Robby Divine, Halli Benasutti, Ryan Kibler, Celia Bisbach, Ha Dang, Olivia Thibeault, Mimi Divine, Meg Adams, James Griffin, Anna Lauko, and Erin Yang. Grad school and working in science can be such particular and unique experiences, and you all have helped me stay sane. I'm very grateful for my cohort, who made me feel welcome in a new state, thousands of miles from my family and friends. Every year, our cohort would rent a cabin, and we would split off for activities during the day, relaxing in whatever way we wished and then come back together for meals or for games or for chit chat at the end of the day. That type of simple, effortless camaraderie is not easy to find, and I'm so glad that I found it here when I needed it.

I want to thank Marisa and Inez, especially. Being vulnerable with these two has led to fulfilling friendships that I look forward to maintaining. Marisa continues to inspire me to advocate for myself, without her encouragement it would have taken me much longer to reach out for help. I'm not sure that I would be writing this if she hadn't initiated our friendship and shared her own fears and worries about graduate school with me. Marisa has such a wonderful talent for leading conversation, something I cherish (and envy) in a friend. Inez, I think unknowingly, just in her natural state of being, facilitated many of the friendships I've made here. I admire her work ethic, her ability to maintain friendships across states (that will come in handy), and her positivity. She has so much faith in me and my abilities, and I have leaned on that many times throughout my graduate career.

For our regular video chats and their insight and support, I'm so thankful for Lianne Barney and Elizabeth McSheery. When I think of the most important things that came out of my time at UConn, you two are first in my mind. I hope that the three of us can be in one place more often than we are now (it should be much easier, considering Lianne and I have consolidated – Elizabeth, you are welcome to consolidate with us). I also want to thank so many of my friends not listed here, outside of the program and back home, who I looked to for joy and for support.

Finally, I want to thank my family: my parents, my sisters, and my brother. My sisters, Michele and Bianca, for somehow always sending me a care package on my worst days. My brother, AJ, for helping me feel at home, even when I felt caught between two different states. And my parents, for their endless support. I remember asking my family to watch a public talk I was giving online, thinking maybe they would, it's 10pm their time, they probably won't stay up. My dad stayed up to watch, with the help of several cups of coffee, and he called me up the next day to tell me how my talk was the best of the three that night (he is obviously very unbiased). And then he told me how he had called the hosts to get a link to share and sent the video to everyone. I can't wait to come home to all the noise and ruckus of our family.

Dedication

For another, smarter me.
May you continue to become less dumb.

Chapter 1 : Introduction

1.1 The rise of cryo-electron microscopy

Many of our body's critical functions: the breakdown of food into energy, the chemical signals sent from our brain to our muscles, the physical work required to contract those muscles, are mediated by proteins [1-2]. Every protein in our body, of which there are around 20,000 unique structures, is given a specific task. Take, for example, glycolysis – the breakdown of glucose to create 2 ATP – which requires 10 different enzymes to extract only some of the energy contained in this single molecule of sugar [3]. Our understanding of glycolysis is entirely dependent on our knowledge of each enzyme: how each catalyzes a specific reaction, how each is regulated by ligands or post translational modifications. Understanding how and when a protein does its task provides insight into how the most basic functions are carried out in our cells and our bodies [4].

A 3-dimensional structure aids that process, allowing scientists to create testable hypotheses about the specifics of how a protein performs its function. X-ray crystallography has been the preferred method for obtaining a 3D structure of a protein since the 1950s, with 86% of deposited structures in the RCSB solved by this method. Cryo-electron microscopy (cryo-EM) structures make up a much smaller portion of the database, but the number deposited has been growing exponentially [5]. This growth can be attributed to advances in image processing and direct electron detectors, which have made it possible to obtain resolutions that rival X-ray crystallography [6-10]. Most cryo-EM density maps, however, are obtained at resolutions in the range of 3-5Å due to conformational and structural heterogeneity [11].

Despite the limited resolution compared to X-ray crystallography, cryo-EM still has several advantages: (1) It does not require a crystal, which often necessitates a large amount of purified protein. (2) It captures proteins closer to their physiological state, as they are kept in solution until the point of freezing. (3) It is amenable to large protein complexes as well as membrane proteins. (4) Finally, it can provide insight into the conformational landscape of a protein [12-13].

1.2 Obtaining a 3D map from cryo-electron microscopy datasets

In cryo-electron microscopy, a protein sample is flash frozen in a thin layer of vitreous ice and the sample is imaged with a transmission electron microscope. Individual 2D images of the protein are collected from the micrograph and averaged to increase the signal-to-noise ratio. Each averaged image captures the protein in a different orientation, and we back project them to build a 3D map. Averaging individual images is essential to obtain higher resolutions but can be problematic if the protein is structurally or conformationally heterogenous. Regions

of flexibility or binding partners and ligands with partial occupancy will be captured at lower resolutions or omitted from the 3D reconstruction entirely.

Although the details of flexible regions are lost in the 3D reconstruction, that data is still present in the individual 2D images and can be extracted with more complex image processing techniques [14-18]. Some of these techniques can reveal how domains move in relation to each other but are unable to improve the resolution [16-18]. Others can improve the resolution of large domains but details at the interface get lost in the analysis [15]. As such, interpreting the conformational landscape at the atomic level remains an open problem.

1.3 Interpretation of cryo-electron microscopy maps

Once a microscopist has their density map, an atomic model that can explain the coulombic potential must be created. Building an accurate structure at the resolutions (3-5Å) common for cryo-EM is difficult; side-chain density at resolutions between 3-4Å is limited and carbonyls are indistinguishable from the tube of the backbone. At 5Å, the backbone can no longer be traced in the density: alpha-helices become rods and beta-strands blur together. And once the resolution drops below 5Å, cryo-EM maps become interpretable only at the domain level.

Modelers can take advantage of previously solved structures of components to aid the interpretation of their map, even at low resolutions. Crystal structures can be docked into the density and then refined with a variety of protocols that use force fields derived from high-resolution structures to improve the geometry of the model while maintaining the density fit [19-20]. Homology modelling, combined with flexible fitting, is a powerful tool to build a starting model if no crystal structure is available [21-22].

If a microscopist is studying a protein with low homology or if regions of the structure need to be rebuilt to fit the density, they often build the backbone and assign residue identities manually. Manual building has become standard in the field because tools for building structures into X-ray data are unreliable at the resolutions typical for cryo-EM [23-24]. Automatic methods for *de novo* modelling at these resolutions have been developed to reduce bias and errors and take two different approaches: fragment-based rebuilding [25-26] or chain-tracing followed by sequence assignment [27-28]. Fragment based methods tend to be more computationally expensive and struggle to build large proteins or segments greater than 100 residues in length. On the other hand, chain-tracing methods rely heavily on side chain information to properly assign sequence, so accuracy drops significantly around 3.5Å. Both methods require continuous density with a clear trace of the main chain to build the backbone. As a result, building structures *de novo* was only possible at resolutions higher than 4-4.5Å for many years.

1.4 Incorporation of machine-learning in cryo-EM model building

The advent of machine learning for structure prediction has made it possible to model structures at much lower resolutions, even below 5Å [29]. These highly accurate deep-learning methods leverage co-evolution information gleaned from multiple sequence alignments (MSA) to accurately predict protein structures, while more recent methods swap the MSA for language models [30-32]. However, these methods still struggle with proteins that have shallow MSA information, with protein complexes, and with flexible regions [33-34]. Cryo-EM data complements these weaknesses: domain orientations are clear down to 9Å resolution and flexible regions or regions with poor covariance information in the MSA may be visible at higher resolutions. In fact, incorporation of implicit density into the AlphaFold prediction pipeline was shown to improve density fit for a small number of cases [35].

Instead of editing or refining a predicted structure into cryo-EM data with traditional tools, new deep learning methods have been trained on the explicit task of building and refining models into density. DeepTracer and ModelAngelo use neural networks to predict where the backbone chain is likely to reside in the density and then assign amino acid identities using the side chain density before building an atomic model. As such, these two methods are only suitable for higher-resolution structures; performance starts to drop around 3.5Å [36-37]. EMBuild, on the other hand, was designed with the explicit task of refining models into lower-resolution maps and was tested on maps ranging from 4Å to 10Å. EMBuild similarly predicts where the backbone will reside based on the density map and uses this backbone probability to refine domains of predicted models into the density [38].

1.5 Validation metrics in cryo-EM model building

Deep learning methods have drastically reduced the labor and expertise necessary to build cryo-EM structures. Microscopists can start with an AlphaFold prediction, only needing to rebuild small portions of the model to fit the density, or they can take advantage of these new deep learning tools that will build their structure from scratch. In the former scenario, a modeler can introduce errors in the regions which require rebuilding. In the latter, it is possible for the method itself to make a mistake. Deep learning models are only as accurate as their training data, so mistakes made by microscopists could be learned and propagated by these methods. Thus, the ability to identify errors will not only aid during hand-building but can create more accurate data for deep learning methods to train with.

Previous efforts in identification of model errors rely on metrics that primarily fall into one of two categories: model quality metrics that focus on atomic geometry [39-40], and fit-to-density metrics that focus on local fit-to-density [41-45]. Model quality metrics usually identify residues or bonds with geometry that differ from statistics collected on high-resolution crystal structures. Often, these metrics are not precise enough to catch local mistakes at the

resolutions common for cryo-EM, because refinement protocols can easily push a wrong model to have good quality under these metrics. CaBLAM addresses this by defining a new geometry term for the backbone which is not used by refinement protocols; however, due to its high cutoff value, CaBLAM is unsuitable for residue-level accuracy [46]. Density-based metrics have two major weaknesses: many are noisy at the level of individual residues and are better suited to evaluate a model's global quality [42,43], while density-based metrics that robustly evaluate local fit rely heavily on side chain density, making them less reliable at resolutions below 3.5Å [45]. Furthermore, some scientists may overfit their models to low-resolution density, so density fit by itself is not always enough to evaluate whether an error has been made [47-48]. Neither model quality metrics nor density-based metrics are robust enough to identify mistakes made at these resolutions.

Chapter 2 : Residue-level error detection in cryo-EM models

2.1 Overview of MEDIC

In this chapter, I present MEDIC (Model Error Detection in Cryo-EM), a statistical model that weighs the contributions of structural information with local model-map agreement to identify local backbone errors in a cryo-EM structure. An overview of training and usage of MEDIC is shown schematically in Figure 2.1. MEDIC is trained to predict a probability of error for every residue, based on three features: energy guided metrics for Ramachandran angles and bond deviations from Rosetta's energy function [49], expected fit-to-density for a residue given the local resolution and the amino acid identity, and predicted model error from DeepAccuracyNet (Figure 2.1A). DeepAccuracyNet is a deep convolutional neural network trained to distinguish native protein structures from Rosetta-determined decoys [50]. It predicts per-residue local Distance Difference Test (p-lDDT), a measure of the number of atom pair distances that are maintained between a native structure and a decoy [51]. We compute DeepAccuracyNet's predicted error on substructures of the input model and combine substructure predictions. Without this step, DeepAccuracyNet predicts that all residues in a structure are placed correctly (p-lDDT of 1.0) for large complexes typical in cryo-EM studies. For the fit-to-density metric, I used masked real-space cross-correlation to measure density fit, and then normalize that value based on statistics for each residue identity at its local resolution, gathered from a set of deposited map-model pairs between the resolutions of 1.5 to 5Å.

Given these three features, a combined model was trained using a set of seven obsoleted protein structures which had been edited months after the initial deposition, presumably to correct structure errors. In this dataset, we identified a variety of errors in the obsoleted versions: loops drawn through helical density, sequence registry errors, squished loops, and misplaced carbonyls (Supplemental Figure 2.1). The combined logistic regression model was trained to predict the residues that changed between the original and most-recent deposition. I validated this initial model on an additional 3 obsoleted structures which had been withheld from training. I compared MEDIC's error probabilities to the residues that changed between these depositions and found that MEDIC had a precision of 76% at a recall of 60% (Supplemental Figure 2.2). Given the high performance on this initial set, I then used this model to generally evaluate deposited structures (Figure 2.1B). Throughout my analysis, I divide these probabilities into three categories: definite error, possible error, and non-error (see Methods). Data analysis is performed only with the definite errors, while each image is colored according to the three categories.

2.1.1 Description of structural features

The energy guided metrics in the model are pulled from Rosetta's realistic energy function [49]. Every pdb is refined in Rosetta, so that the energy scores are meaningful. Then,

the energies for Ramachandran angles (*rama_prepro*) and bond deviations (*cart_bonded*) are evaluated for each residue in the structure and fed directly to MEDIC.

The final structural feature, predicted lDDT, comes from DeepAccuracyNet [50]. Because DeepAccuracyNet was trained on smaller structures, <300 residues in length, I run the model on portions of the structure at a time: a sequence of 20 residues and the context within 20Å of that query sequence. DeepAccuracyNet also predicts the mask to use during the lDDT calculation and I additionally mask out nearby residues (i-4 to i+4) in the predicted mask to downweight DeepAccuracyNet's tendency to favor alpha helices. The predicted lDDT values are saved for only the query sequence and then passed to MEDIC.

2.1.2 Description of density feature

To calculate expected fit-to-density for amino acids, I collected statistics on a set of 24 deposited map-model pairs, using atomic B-factors as a substitute for local resolution. Each model is refined as described in Methods. The masked real-space density cross-correlation was calculated for every residue using an atom mask of 3.2Å and each was placed into a bin according to its amino acid identity and the average B-factor of the residues within an 8Å neighborhood. Outliers were removed from each bin using the interquartile range method. A mean of the cross-correlation scores was then computed for each amino acid/B-factor bin and a standard deviation was calculated across each B-factor bin.

Now that I have collected statistics, I can apply them during error prediction. The means and standard deviations are used to transform the cross-correlation of each residue in a protein model into a z-score. A very negative density z-score is indicative of a residue which fits the density worse than expected, given its amino acid identity and the average B-factor. The density z-score is then passed to MEDIC. This process of collecting statistics and transformation of raw scores is carried out for the cross-correlation of the residue by itself and the cross correlation of a three-residue window centered on the residue of interest.

2.1.3 Training on obsoleted structures

I probed the RCSB for pdbs which had been edited after deposition, pulling all cryo-EM structures between 2.5 and 4Å resolution that had coordinates replaced [5]. Upon manual inspection, 10 models of the 46 were chosen, eliminating cases where changes were made to ligands, nucleotides or only rotamers, or where the obsoleted model didn't resemble a globular protein. 3 of the 10 models were withheld from training and used for validation. PDB IDS for this training set are: 6K1H, 6ZA9, 7BV2, 6TT7, 6CP3, 7BW4, and 7KSM. We added two structures to the training set which we believed to contain very few errors and labeled every residue as a non-error: 6VRW and 6U8E. PDB IDS for the validation set are: 5V8F, 7JSN, and 6L42.

Now I have a set of pdbs that contain mistakes made by microscopists and need to generate labels for training, marking each residue in a model as an "error" or "non-error." I compare the obsoleted pdb with the newer version, removing any domains or regions that exist

in only one of the models. Each residue in which the backbone atoms have an RMSD greater than or equal to 1Å between the two models is marked as an error. To capture sequence registration errors, any residue that appears in the obsoleted model but not the new version is marked as an error. This process resulted in approximately 800 errors out of a total of 21000 residues. I then trained a logistic regression classifier, with balanced class weights, to predict the errors using the structural and density features.

In the earlier stages of this project, I also trained a multilayer perceptron (MLP) to predict errors. Unlike a logistic regression, which learns only a linear combination of the features, a multilayer perceptron can learn about relationships between features and apply more complex logic operations. The performance of the MLP on the small cross-validation set was equivalent to the logistic regression classifier. Moreover, when I tested it on a larger set of structures (those discussed in section 2.3.2), I found that the MLP marked many more single residues as mistakes, a hallmark of overfitting. I suspect that the training set and the associated features may be too simple for an MLP. It's possible that with a larger training set and more features, an MLP would perform better.

2.3 Testing MEDIC on deposited structures

2.3.1 Validation on low resolution structures later solved to higher resolutions

To validate my approach, I considered EMDB-deposited structures between 3.5 and 5Å resolution, which were subsequently solved to better than 3.5Å (and at least 1Å better than the original deposition). There were 68 cases, of which I manually removed 41 with domain orientation changes between the high-resolution and low-resolution structures. The results on this dataset are summarized in Figure 2.2A. On this set of 27 structures, our method has a precision of 67% at a recall of 60%. MEDIC outperforms the widely used density-only metric, Q-score [41], which has a precision of only 34% at the same 60% recall.

I next examined which features were predictive of the true positives identified by MEDIC. Approximately 81% are predicted by the p-lDDT alone, while the remaining 19% require at least 2 features to be considered an error. The reliance on p-lDDT to predict most of the errors could be because of bias in the training set, which primarily contains long segments that were corrected. It might also simply reflect the types of errors microscopists tend to make; hand-built models are much more likely to fit the density well but have poor geometry and structural features.

Some of the errors identified by MEDIC in the low-resolution structures are highlighted in Figure 2.2B, with the corresponding model in its high-resolution density map in Figure 2.2C. In a structure of a voltage-gated calcium channel (PDB 5GJW), it is difficult to trace the backbone while properly accounting for the large aromatic side chain density (Figure 2.2B, top panel). The mistake is identified by MEDIC with relatively equal contributions from the p-lDDT and bond geometry scores. Likewise, the error found in an insulin degradation enzyme (PDB

6B70) is captured by multiple features, this time the density and bond geometry scores (Figure 2.2B, middle panel). The backbone is hardly visible in the density map, which may explain why the microscopists had difficulty properly fitting the serines into the density. In contrast, the mistake found in a transmembrane channel (PDB 6M66) is dominated by the p-lDDT score (Figure 2.2B, bottom panel). It would be difficult to catch this error by visual inspection, as the model seems reasonable given the density.

To better understand any shortcomings of MEDIC, I looked at two structures for which the performance was worse than the aggregate results. MEDIC falsely marks an entire stretch of residues as a mistake (Figure 2.2D) in a partial complex of an ATP synthase (PDB 6F36). It does not see the proper structural context for this sequence as it is unmodelled in the low-resolution structure (Figure 2.2E). The other case which MEDIC performed poorly on, a dehydrogenase (PDB 7E5Z), contained many errors fewer than 3 residues in length which MEDIC failed to identify, two of which are shown in Figures 2.2F-I. MEDIC fails to mark an incorrect carbonyl as an error in the low-resolution model (Figure 2.2F) that is supported by the higher-resolution data (Figure 2.2G). However, MEDIC finds zero high-probability errors in a region of the low-resolution model (Figure 2.2H) which appears to be an error in the high-resolution model (Figure 2.2I).

Given the worse performance on the errors in the dehydrogenase (PDB 7E5Z), I manually examined 30 differences across 4 low-resolution structures that MEDIC failed to identify. Among these, 16 were mistakes in the model built against low-resolution data, while 14 were either ambiguous in the high-resolution density or seemingly incorrect in the high-resolution model. Three examples are highlighted in Supplemental Figure 2.3: one difference where the high-resolution structure has an error (Supplemental Figure 2.3A-B), and two more where the high-resolution structure is not supported by the density (Supplemental Figure 2.3C-F).

2.3.2 Using MEDIC to guide model rebuilding

With the understanding that MEDIC is relatively precise when identifying errors, I next wanted to assess the usefulness of the model to aid in a manual structure rebuilding process. To that end, I evaluated MEDIC on a selection of 12 models with diverse topologies and resolutions and attempted to fix – using Rosetta refinement tools and AlphaFold – all the segments marked as errors (see Methods). There were 237 segments predicted to be definite errors (with high error probability), 33 of which were disordered regions with little or no visible density (Supplemental Figure 2.4). Of the remaining 204 segments, 133 (65%) were 1-3 residues in length, 38 (19%) between 4-9 residues, and 33 (16%) were greater than 10 residues. I was able to rebuild and fix 120 (59%) of these segments; for an additional 26 segments, I was able to significantly reduce the number of definite errors in that region. The fixable mistakes included 2 sequence registration errors, where the sequence is shifted on the backbone relative

to the correct placement, 51 incorrect loops, 51 cases of poor secondary structure, and 16 flipped carbonyls (Table 2.1).

A representative subset of errors that MEDIC was able to address are highlighted in Figures 2.3 and 2.4. In these cases, I was able to correct 2 significant sequence registration errors (Figure 2.3). Figure 2.3A compares the deposited structure of a lipid scramblase (PDB 6E1O) with my new model. Notably, my model has better hydrophobic packing and I explain the large side chain density with a phenylalanine as opposed to a lysine residue (Figure 2.3B). This sequence registration error was propagated from a previously solved crystal structure (PDB 4WIS), in which the density for this region was poorly resolved. In both structures, this helix is preceded and followed by unresolved regions, making proper sequence placement more difficult. Conversely, the sequence registration error found in a hedgehog receptor (PDB 6DMB) occurs because the pitch of the helix is not visible in the density (Figure 2.3C). The addition of a bulge in the repaired model (Figure 2.3D), justified by the preceding proline, pushes a phenylalanine into large side chain density which was poorly explained by an alanine in the original model.

MEDIC is also capable of finding gross backbone errors, including areas with poor secondary structure and incorrect loops. In Figure 2.4A, it is clear by eye that the beta strands of this kinesin motor domain (PDB 5MM4) have poor hydrogen bonding. After rebuilding (Figure 2.4B), the method marks these regions as correct, as MEDIC balances proper structural features with density fit. In addition to identifying poor structural features, MEDIC can recognize if a stretch of residues is assigned the incorrect secondary structure, such as the region from a hedgehog receptor (PDB 6DMB) depicted in Figure 2.4C. However, the fixed model is supported by more than the p-lDDT score; it has less unexplained density, which is reflected by large improvements in the density scores (Figure 2.4D).

Furthermore, MEDIC can identify some shorter, subtler backbone errors, such as incorrectly placed carbonyls, by combining multiple features (Figure 2.4E-H). The deposited model of the bluetongue virus (PDB 3J9E) has a Ramachandran angle that falls just in the "Allowed" region (Figure 2.4E). MEDIC uses the p-lDDT and the bond geometry scores to predict this error, and after rebuilding, both Ramachandran angles and density fit improve (Figure 2.4F). Similarly, the structure for a neurotoxin (PDB 7QFQ) contains Ramachandran angles which Molprobity also classifies as "Allowed" (Figure 2.4G). I find this error with relatively equal contributions from p-lDDT, density, and geometry energies. The rebuilt model improves the density fit for the tryptophan and alanine residues while removing the problematic Ramachandran angles (Figure 2.4H). Of the over 1300 residues identified as errors across these 12 models, approximately 66.5% of them were predicted by the p-lDDT score alone, 1.4% by the density, and 0.4% by the Ramachandran energy, while 32% required at least 2 features.

To quantify MEDIC's performance on this set of structures, I used the differences between the deposited structures and the rebuilt models (see Methods) to determine that

MEDIC has a precision of 67% at recall of 60% (Supplemental Figure 2.5). The increased performance of MEDIC at high recall values compared to the low- vs. high-resolution validation set could be attributed to a few factors. In the set of validation structures, it is possible that the high-resolution models may contain errors. Moreover, there could still be conformational differences between the high- and low-resolution structures, such as flexible loops or shifts that occur at interfaces contained in only one of the depositions. Both would hurt MEDIC's perceived performance.

2.3.3 Identifying errors in all deposited structures in the EMDB

After confirming MEDIC's high accuracy and utility in model building, I ran MEDIC on all structures in the EMDB between the resolutions of 3 to 5Å to gauge the reliability of the method on over 1500 depositions. The aggregate statistics from this run are shown in Figure 2.5. Upon inspection, several models were composed of docked crystal structures with no visible density for one or more domains, so I removed residues with a model-map correlation of less than 0.4. In Figure 2.5A, I show the fraction of residues marked as errors in every EMDB deposition. There is only a slight trend with resolution, which is unsurprising given that as we move to lower resolutions, microscopists are more likely to dock crystal structures or use homology modeling than hand-build structures. Because cryo-EM maps are rarely homogenous in resolution, I also report the fraction of residues marked as errors after grouping by atomic B-factors (Figure 2.5B). At very low atomic B-factors (indicating well-resolved density), very few errors are made. As the atomic B-factors increase, more mistakes are made.

I manually inspected the outliers in the data: errors with low atomic B-factors and maps with very high error fractions. Although the fraction of errors is greater than 40% on the 20 model-map pairs I examined, the errors do seem real. In some cases, entire domains have little to no secondary structure (Supplemental Figure 2.6A-B). All these structures were built pre-Alphafold, using outdated (then state-of-the-art) structure prediction software or by hand-tracing into low-resolution data. Unsurprisingly, I find that 88% of the errors in this set are predicted by the p-lDDT alone. In the structures that contain errors with low atomic B-factors, I find that while some errors appear to be real, there also appear to be false positives. There are several causes for the perceived false positives, including residues marked as errors because they are involved with ligand or metal binding, or they correspond to very short disordered segments (Supplemental Figure 2.6C-E).

2.3.4 MEDIC can guide rebuilding of and find errors in AlphaFold predictions

Although MEDIC can identify errors in hand-built structures, many microscopists will now start model-building from an AlphaFold prediction [35]. I compare MEDIC's performance to AlphaFold models, highlighting loops which I identified as an error in the original deposition (Figure 2.6A & 2.6D) and where AlphaFold predictions do not fit the density. The loop predicted by AlphaFold for the motor protein, prestin (PDB 7S9D), would require significant rebuilding

(Figure 2.6B). MEDIC identifies the new model, built with tools in Rosetta, as correct (Figure 2.6C). The shorter loop predicted by Alphafold for the bluetongue virus (PDB 3J9E) is not only a poor fit to density (Figure 2.6E); the carbonyls are placed incorrectly when compared to the final model (Figure 2.6F). Of the 12 models I rebuilt, 23 regions (from 7 different AlphaFold models) would have required rebuilding. AlphaFold was confident (predicted lDDT > 70) in 10 of these regions, which means that modelers would need to manually identify these mistakes, not just remove low confidence regions, and then rebuild, presumably by hand. MEDIC will be useful for this editing process: my method was able to identify that the deposited structure or the rebuilt model was correct in 18 of those 23 regions. In the remaining 5 cases, I was unable to build a structure that satisfied MEDIC.

I next investigated whether MEDIC could identify mistakes made by AlphaFold predictions. I highlight two subtle mistakes in refined AlphaFold predictions, identified by comparison with my rebuilt models, which are unlikely to be captured by density metrics alone. Rosetta's refinement protocol is unable to assign the correct rotameric state for the tryptophan in the AlphaFold prediction of a malate transporter (PDB 7VOJ), suggesting that the backbone is not modeled correctly despite AlphaFold's confidence in this region (Figure 2.6G). In the rebuilt model, subtle changes in the backbone allow the valine and tryptophan to fit the side chain density better (Figure 2.6H). The AlphaFold prediction for the bluetongue virus (PDB 3J9E) places the alanine incorrectly and the glycine pops out of the density (Figure 2.6I). We slide over the isoleucine and then reorient the alanine to explain a small bump in the density (Figure 2.6J). On our set of 4 refined AlphaFold predictions, I found that MEDIC had a precision of 80% at recall 60% (Supplemental Figure 2.7). Of the true positives that MEDIC identifies, approximately 36% are found by the p-lDDT alone, 42% by the density score, and the remaining 22% required at least two features. Unsurprisingly, the density score plays a larger role here, as AlphaFold models tend to have good geometry and structural features but poor fit-to-density.

2.3.5 MEDIC performs similarly when using half-maps for refinement and evaluation

In a previous study, it was shown that half-maps – maps generated from two independent subsets of the raw cryo-EM data – can be used during refinement to reduce and evaluate overfitting [47]. Given my goal to identify errors in structures, I compared MEDIC's performance when using the full map and when using half-maps (see Methods). Our statistics used for density evaluation were collected on full maps, so it was unclear how using half-maps may affect performance. On two randomly chosen structures from the EMDB, I noticed that the density z-scores differed significantly between the full map and the half-maps (Supplemental Figure 2.8A). In one case, the half-maps had better density scores, because the atomic B-factors used for evaluation were lower in the half-maps (Supplemental Figure 2.8A, top). In the other case, the half-maps had worse density scores, because half-maps tend to be noisier than full maps (Supplemental Figure 2.8A, bottom). To combat these two effects, I used phenix's

auto_sharpen protocol to separately sharpen each half-map [67]. The error probabilities using the sharpened half-maps were much closer to those obtained with full maps (Supplemental Figure 2.8B).

Given this result, I then took 11 structures from our high- vs. low-resolution validation set which were deposited with half-maps, sharpened each of the half-maps, and evaluated the structures with MEDIC. MEDIC's performance with the half-maps was very similar to the full maps (Supplemental Figure 2.8C). It appears disordered regions are more frequently marked as errors, likely because these regions are more poorly resolved in half-maps than full maps. I did not test using the raw half-maps with these 11 structures, largely due to time constraints. I am doubtful that using the raw half-maps will perform better than the sharpened ones, but that possibility cannot be ruled out. In addition, a lower density weight may need to be used for half-maps, although that has not been tested either. Overall, it seems as though using MEDIC in conjunction with half-maps is a viable option, given that they are sharpened first.

2.4 Discussion of MEDIC

2.4.1 Possible improvements to MEDIC

As mentioned previously, I believe that MEDIC may have difficulty identifying mistakes that are less than 3 residues in length. This could be for a few reasons: the statistical model could be biased towards longer errors, or the model does not have features that capture short errors. To determine if the training data is the issue, I can train with only the short errors contained in the dataset and a stark improvement in the ability to detect short errors would suggest that bias is the issue. Randomly pulling an equal number of residues from long errors and short errors could help us balance the two.

If training only on short errors doesn't improve performance, I may need to add additional features to the model. I have a strong suspicion that many of the short errors with misplaced carbonyls that MEDIC could identify were found because of Ramachandran outliers that were introduced after refinement. In these cases, the Ramachandran scores are particularly poor; but if there's a flipped carbonyl and no outlier, I believe MEDIC currently has no feature to identify a mistake has been made. CaBLAM was created for this explicit purpose and combining it with other features might mitigate its tendency to mark so many false positives. In addition, I have noticed that modelers tend to overfit sidechains into density at the expense of backbone density fit. This type of error gets lost in a simple residue cross-correlation, as the sidechain which is overfit to density will make up for the backbone which fits poorly. Therefore, a metric that evaluates fit-to-density on only the backbone atoms rather than the full residue may help capture these mistakes.

I also noticed that MEDIC tends to mark prolines as incorrect. Prolines, both those marked as errors and non-errors, have higher contributions from the Rosetta energy scores compared to the other amino acids. Instead of feeding Rosetta energy scores directly to MEDIC,

I could normalize Rosetta scores based on amino acid identity, which may help reduce the number of prolines marked as false positives.

Currently, MEDIC cannot find rotamer errors, when the sidechain flips out of the density, but the backbone is correct. MEDIC was only trained on backbone errors, although I did attempt to develop a method for rotamer identification as well. I generated a rotamer error training set by taking the training data and flipping residues out of the density at random with programs in Rosetta. I found that p-lDDT was not a good predictor of rotamer errors, and the single residue density score separated the rotamers from the non-errors but not the backbone errors. I then generated a density score that would assess whether the residue of interest fit the density worse compared to all residues within a 15Å neighborhood (Supplemental Figure 2.9A). While this metric seemed to separate rotamers errors from non-errors and backbone errors, combining this score with the previous features and training a multilevel perceptron still resulted in poor performance (Supplemental Figure 2.9B). Moving forward, it may be better to start with these new density neighborhood scores, add the original features back one at a time and train a logistic regression model. Additional features that may prove useful: a density score where the cross-correlation uses a wide mask to capture unexplained density and rotamer energies derived from the Dunbrack library [52].

2.4.2 Future directions

As it becomes more commonplace to model large protein complexes into lower resolution density maps [29], validation metrics that can evaluate these structures and help guide rebuilding are necessary. MEDIC's performance on structures with resolutions worse than 5Å has not been tested and given that the statistics for density did not include these resolutions, it is unclear how reliable my method will be in those cases. MEDIC could be extended to lower resolutions by gathering more statistics and by measuring density fit across longer stretches of sequence, making it suitable for use with cryo-electron tomography. A training set could be curated from low resolution structures which are later solved to higher resolutions by removing regions with different domain orientations and regions of ambiguity. Incorporating AlphaFold models into the training set may also be useful, so that MEDIC more explicitly learns to find regions which have good structural features but do not fit the density well.

AlphaFold has not only made it possible to model lower-resolution structures, it also drastically changed the model building process for higher resolution structures as well. Now microscopists will edit loops or interaction sites rather than build entire structures. For large complexes, identifying and fixing errors in AlphaFold models can still be error-prone and time consuming, especially if these are flexible regions solved to lower local resolutions. Creating a program to automatically dock these models and fix any remaining errors would reduce the amount of time and expertise necessary to solve structures. MEDIC could be used to guide this

rebuilding process; the method's high precision would substantially reduce the sampling space, which makes the problem of automatically fixing local errors much more tractable. Based on the observations described here, I believe that MEDIC will be a powerful validation tool for cryo-EM microscopists.

2.5 Acknowledgements

The development of a ChimeraX plugin (highlighted in Supplemental Figure 2.10) for MEDIC was done by Wolfgang Lugmayr and Thomas C. Marlovits.

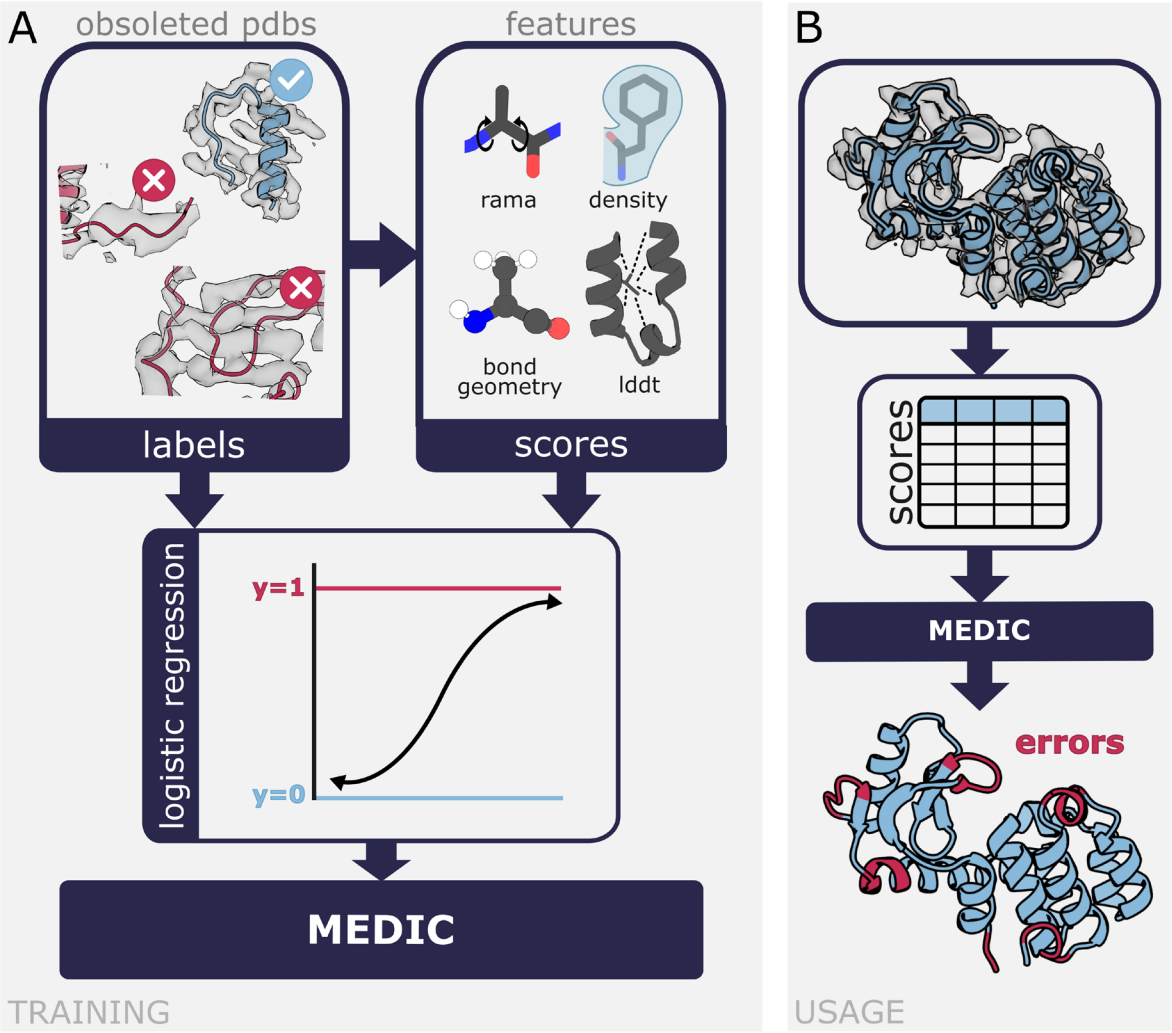


Figure 2.1 Overview of training and usage of MEDIC.

(A) Pulling pdbs that had been edited after deposition, I marked every residue for which the backbone moved between the two versions as an error (red) and collected scores from each of the features on all residues. These labels and scores were fed to logistic regression, which gives us the statistical model, MEDIC. **(B)** To use MEDIC, provide a map/model pair to the program. I calculate the scores for each of the features, which are then passed to MEDIC. MEDIC predicts a probability that each residue is an error, where higher probability is indicative of an error. (back)

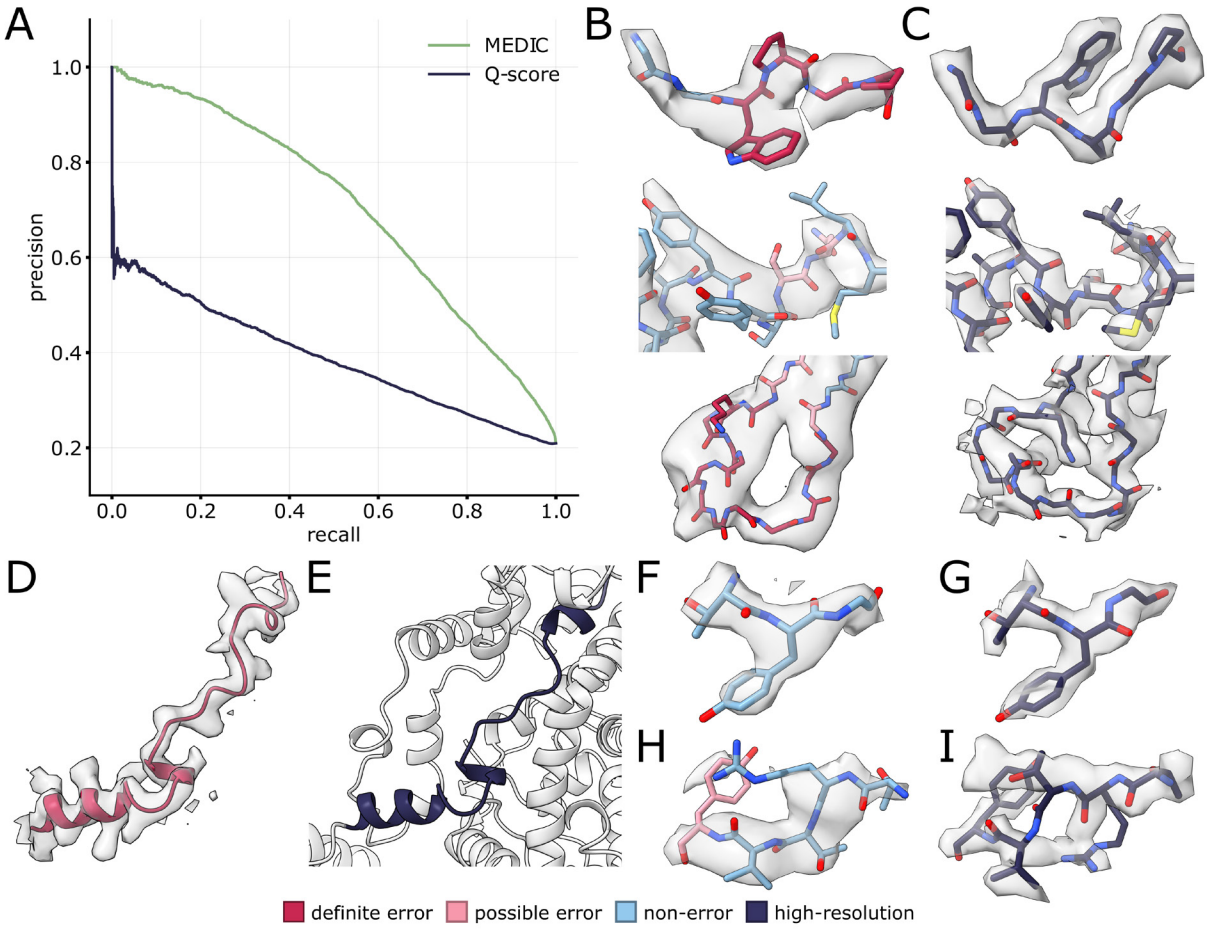


Figure 2.2 MEDIC identifies errors that are validated by high-resolution data.

For panels B, D, F, and H, residues are colored by MEDIC error prediction. **(A)** Precision-recall curve of MEDIC error prediction and Q-scores on differences between low-resolution and high-resolution structures. **(B)** Examples of successful identification of errors in low-resolution structures: voltage-gated calcium channel (PDB 5GJW, residues 263-268A) **(top)**, insulin degradation enzyme (PDB 6B70, residues 237-253C) **(middle)**, transmembrane channel (PDB 6M66, residues 246-263) **(bottom) (C)** The analogous region in the high-resolution structure: voltage-gated calcium channel (PDB 6JPA) **(top)**, insulin degradation enzyme (PDB 7K1F) **(middle)**, transmembrane channel (PDB 6WBF) **(bottom)**. **(D)** False positive predicted by MEDIC in ATP synthase (PDB 6F36, residues 122-151N). **(E)** High-resolution structure ATP synthase (PDB 6RD5) with missing context from low-resolution structure colored in gray. **(F)** MEDIC misses an incorrect carbonyl in low-resolution structure of a dehydrogenase (PDB 7E5Z, residue 63B). **(G)** Analogous region in high-resolution structure (PDB 7VW6). **(H)** MEDIC does not mark a region in the dehydrogenase (PDB 7E5Z, residues 282-286A) that matches the high-resolution data. **(I)** Mistake made in the high-resolution model (PDB 7VW6). (back)

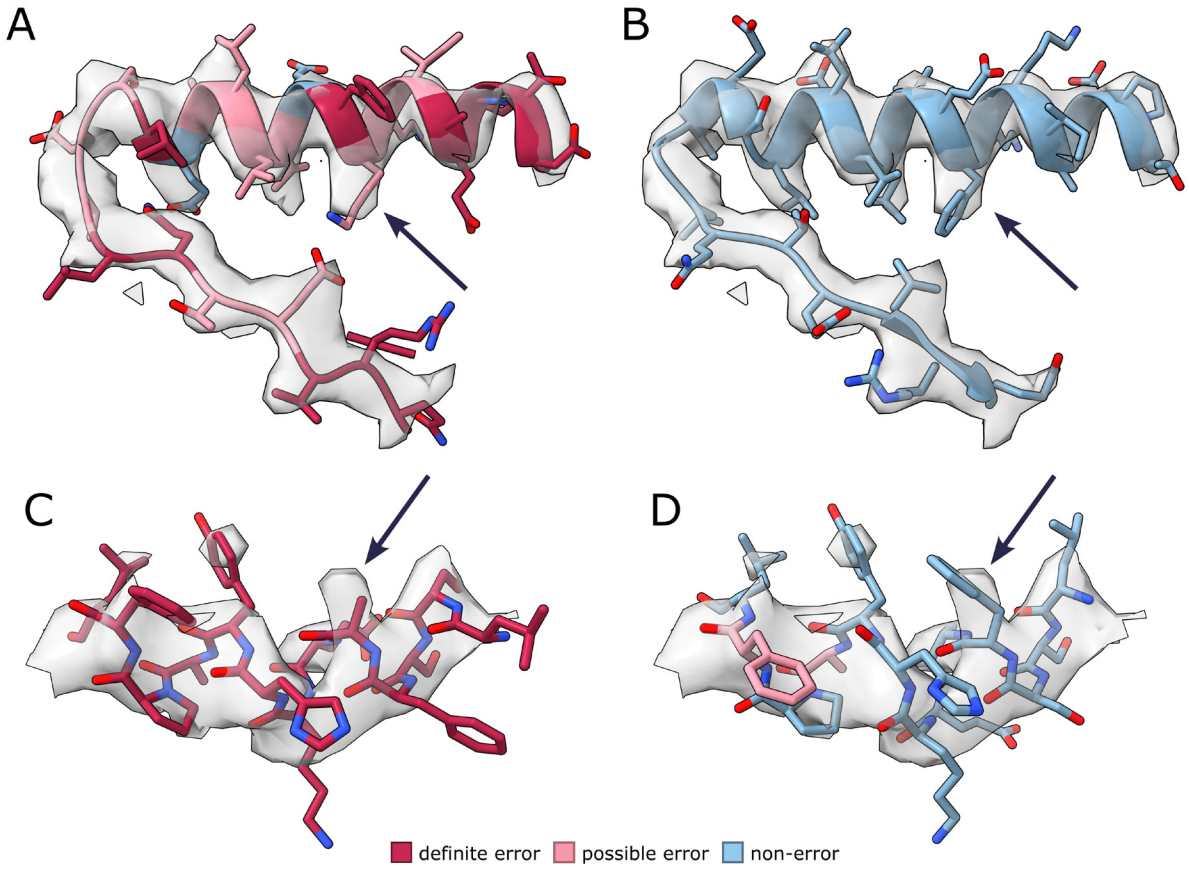


Figure 2.3 Sequence registration errors identified in deposited structures.

All residues colored by predicted error from MEDIC. **(A)** Sequence registration error in lipid scramblase (PDB 6E1O, residues 28-53A). **(B)** Rebuilt model of **A**, where phenylalanine fills large side-chain density. **(C)** Sequence registration error in hedgehog receptor (PDB 6DMB, residues 753-765A). **(D)** Rebuilt model of **C** with a bulge added, where phenylalanine fills large side-chain density. (back)

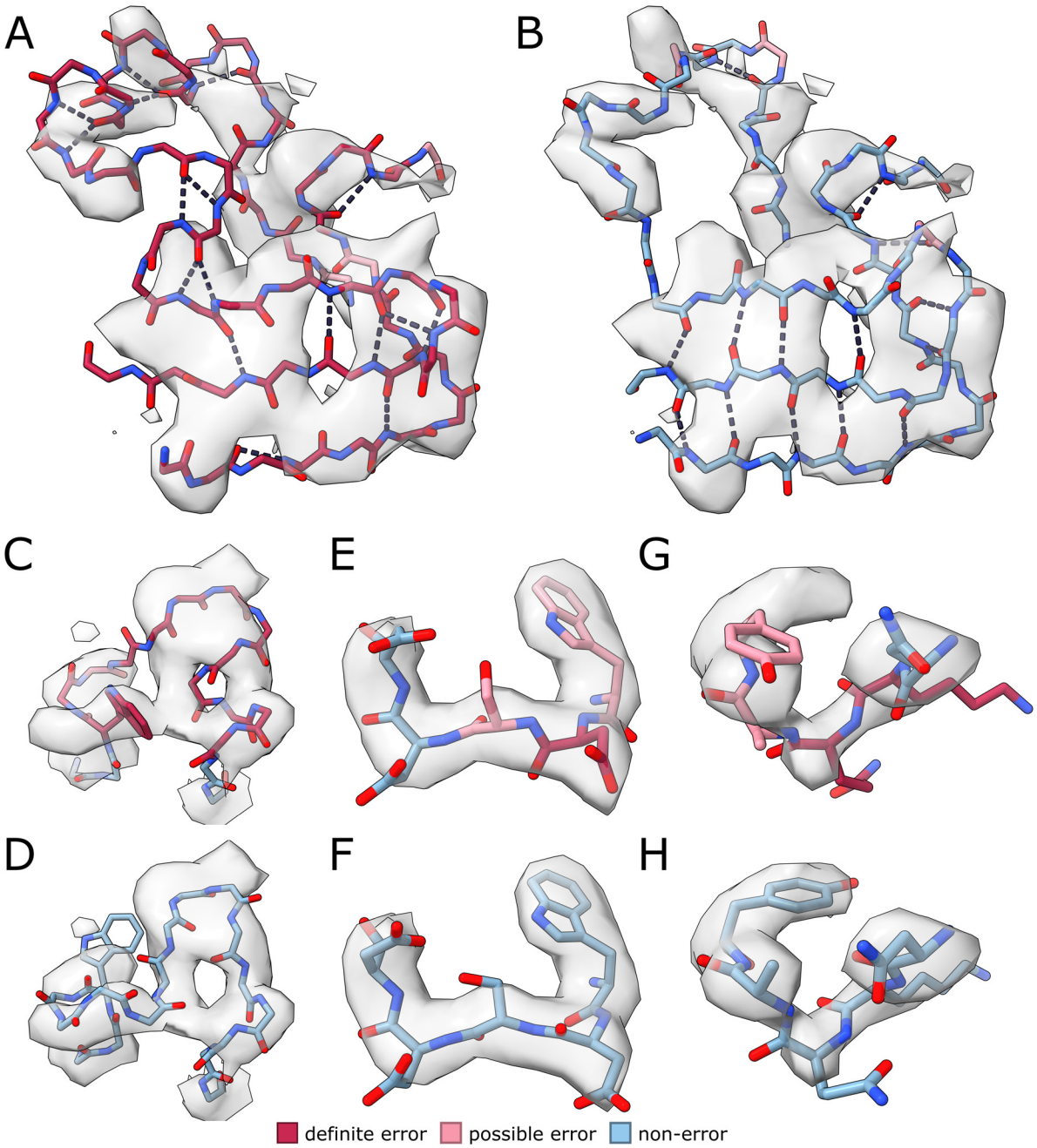


Figure 2.4 Backbone errors identified in deposited structures.

All residues colored by predicted error from MEDIC. **(A)** Predicted errors in kinesin motor domain (PDB 5MM4, residues 10-45K and 65-80K) **(B)** Rebuilt model of **A** with better hydrogen-bonding. **(C)** Small loop in hedgehog receptor (PDB 6DMB, residues 883-897A) that poorly explains density. **(D)** Rebuilt loop of **C**, which has less unexplained density. **(E)** Protein backbone with incorrect carbonyls in bluetongue virus (PDB 3J9E, residues 370-375D). **(F)** Rebuilt backbone of **E** with improved Ramachandran angles. **(G)** Deposited structure in neurotoxin (PDB 7QFQ, residues 416-421A). **(H)** Rebuilt model of **G** with better fit to density and improved Ramachandran angles. (back)

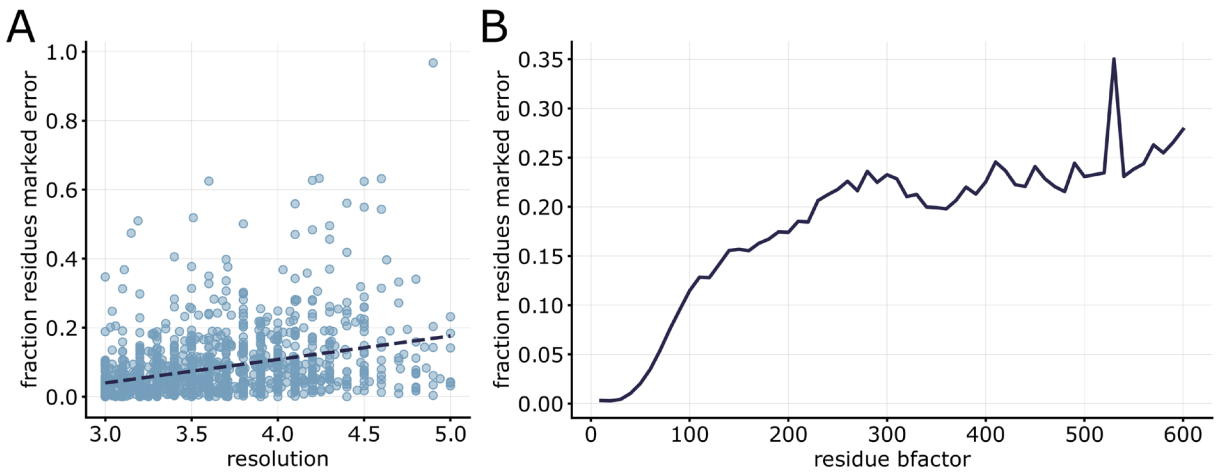


Figure 2.5 MEDIC identifies errors in over 1000 deposited structures in the EMDB.

For both plots, residues with low density cross correlation, less than 0.4, were not included. **(A)** Fraction of residues marked as an error by MEDIC in each deposited structure. **(B)** Fraction of residues marked as an error with atomic B-factors between X-10 and X. (back)

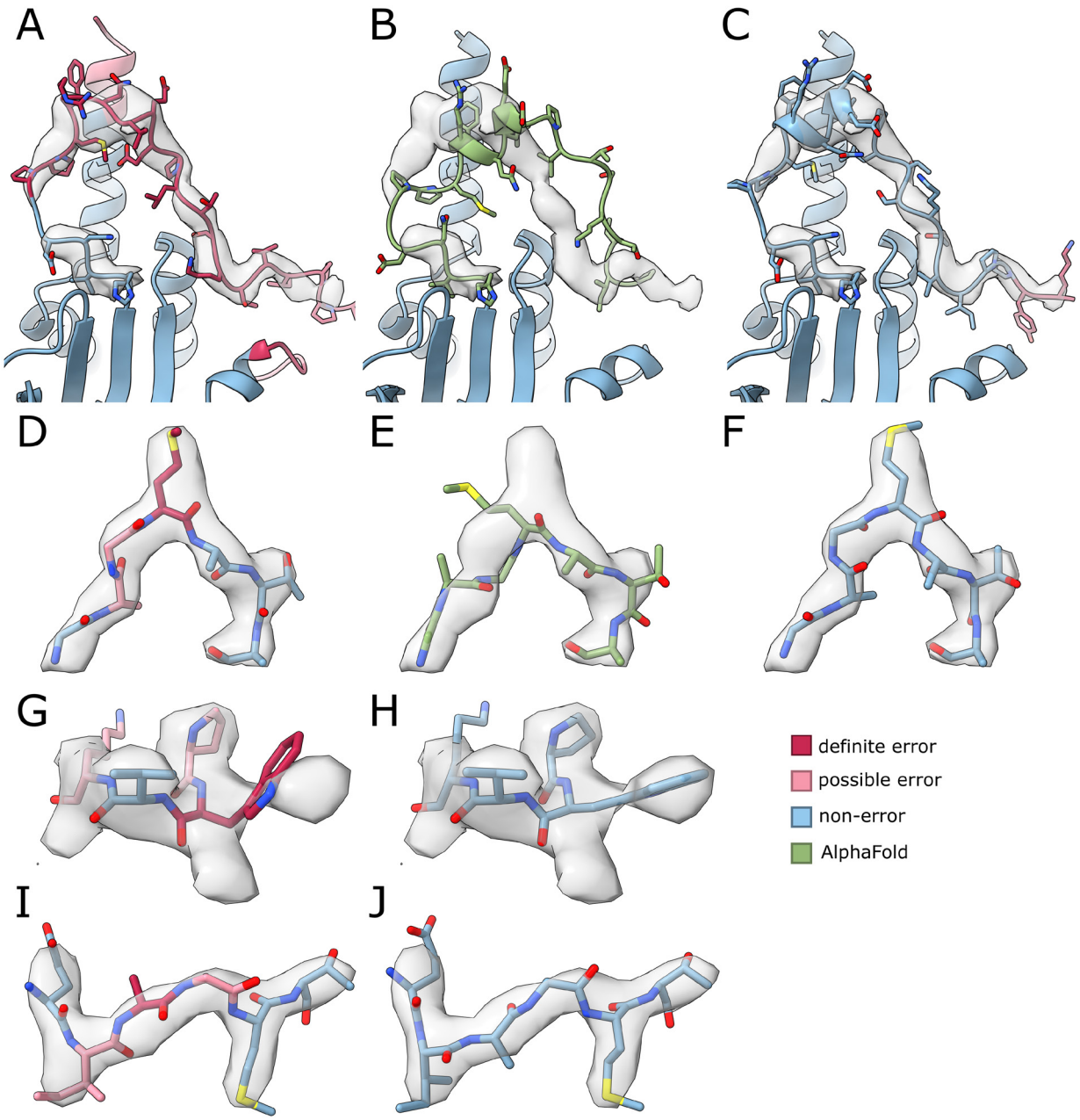


Figure 2.6 MEDIC can guide rebuilding of and identify errors in AlphaFold models.

(A) Deposited structure of prestin (PDB 7S9D, residues 615-636A), colored by error prediction. **(B)** AlphaFold model for prestin after docking the relevant domain into the density. **(C)** Rebuilt structure of loop in **C**, colored by error prediction. **(D)** Deposited structure for bluetongue virus (PDB 3J9E, 235-239D) in density map, colored by error prediction. **(E)** AlphaFold model for bluetongue virus after refining the model into the density. **(F)** Rebuilt structure of loop in **D**, colored by error prediction. **(G)** Refined AlphaFold prediction for malate transporter (7VOJ, residues 128-131) colored by error prediction. **(H)** Rebuilt structure of region in **G**, colored by

error prediction. **(I)** Refined AlphaFold prediction for bluetongue virus (PDB 3J9E, residues 219-224), colored by error prediction. **(J)** Rebuilt structure of region in **I**, colored by error prediction. (back)

PDB ID	Total Segments Marked by MEDIC	Fixed	Improved
6JT1	7	2	1
incorrect loop	3	2	1
7R9U	8	6	1
poor secondary structure	5	4	1
incorrect loop	3	2	0
7QFQ	30	26	1
poor secondary structure	16	16	0
incorrect loop	9	8	1
incorrect carbonyls	2	2	0
3J9E	4	3	1
incorrect loop	2	1	1
incorrect carbonyls	2	2	0
7S9D	24	16	1
poor secondary structure	11	10	1
incorrect loop	7	6	0
5MM4	56	25	9
poor secondary structure	17	14	3
incorrect loop	17	9	5
incorrect carbonyls	3	2	1
6C14	10	5	0
poor secondary structure	4	3	0
incorrect loop	2	2	0
6XOW	4	3	0
incorrect loop	3	3	0

PDB ID	Total Segments Marked by MEDIC	Fixed	Improved
7VOJ	1	0	1
incorrect loop	1	0	1
6DMB	36	22	10
poor secondary structure	3	1	2
incorrect loop	18	10	8
sequence registry	1	1	0
incorrect carbonyls	9	9	0
6E1O	11	9	0
poor secondary structure	3	3	0
incorrect loop	5	5	0
sequence registry	1	1	0
6C0V	13	4	1
poor secondary structure	1	0	1
incorrect loop	4	3	0
incorrect carbonyls	1	1	0
Totals	204	120	26
poor secondary structure	60	51	8
incorrect loop	74	51	17
sequence registry	2	2	0
incorrect carbonyls	17	16	1

Table 2.1 Summary of identified and corrected high-probability errors in 12 deposited models, excluding disordered regions.

A segment is fixed if the new model contains no high-probability errors in that region. A segment is improved if the number of high-probability errors in that region is significantly reduced, indicating that an error was made in the original model, but we lack the tools to build the correct model. Any remaining segments could not be fixed nor improved. The categories of

error are defined by the following: (1) poor secondary structure – secondary structure that is missing important features, like backbone hydrogen bonds, or secondary structure that is assigned improperly, a loop is drawn instead of a helix. (2) incorrect loop – a loop which is drawn incorrectly through the density, but secondary structure has been assigned correctly. (3) sequence registry – the backbone is drawn correct and the sequence is placed incorrectly. (4) incorrect carbonyls – carbonyls are misplaced. (back)

Chapter 3 : Revealing protein conformational space through model-guided image processing

3.1 Overview of model-guided image processing

Although methods have been developed to reveal the conformational space contained in a cryo-EM dataset, the field still lacks robust procedures to obtain a structural understanding of these conformational changes. cryoDRGN and cryoSPARC's 3D variability analysis produce trajectories of how the density changes across their defined latent space, but it is unclear how these trajectories should be translated to protein structures [16-17]. A more detailed understanding of the conformational space is usually obtained via 3D classification, where the images are sorted into discrete classes and refined to high resolutions. However, classification of cryo-EM data is challenging because we have many images, a low signal to noise ratio, and no knowledge of what or how many conformations are present in solution.

I can reduce the complexity of this sorting problem by leveraging the power of atomic models and Rosetta's energy function to approximate what conformations the protein may adopt and use these hypothetical states to guide image processing. In crystallography, atomic models are used to supplement missing information as a solution to the phase problem during molecular replacement; therefore, using models in EM is a natural extension [53]. 3D classification is also made difficult by the noise caused by averaging different conformations. Density maps generated from atomic models reduce that noise so particles may be sorted more accurately and more precisely, which may produce classes with more subtle changes in conformation.

An overview of model-guided image processing is shown schematically in Figure 3.1. Model-guided image processing starts with the same steps of a single particle analysis: particle picking followed by 2D and 3D classification to remove images without protein present and then refinement to obtain a high-resolution consensus map. At this stage, I build an initial model into the map with existing methods in Rosetta [23,25,26,54,55]. Based off this atomic model, an ensemble of hypothetical states that the protein may adopt in solution is generated. Density maps are created from these hypothetical states and fed to 3D classification to guide particle sorting. Each class is then refined separately using a FSC gold standard refinement. FSC gold standard refinements spilt the data into two independent halves during refinement and combine them at the end, which calculates the resolution of the map more accurately. The data can then support or reject any of these hypothetical conformations, where a low number of images and low resolution after refinement denote rejection. Supported states are then passed to subsequent rounds by building a new atomic model into the refined density map, allowing the program to adjust the hypothetical state to one that better represents the data.

3.1.1 Passing hypothetical density maps to 3D classification

Using existing code in Rosetta, I generate a density map at the desired resolution based on the atomic model, where each atom is represented as a Gaussian sphere and the width of the sphere is dependent on the desired resolution and the B-factor for the given atom. Each density map is altered by a histogram matching protocol so that the density histogram agrees with that of the consensus map and is then fed as input to 3D classification in RELION, a cryo-EM image processing suite [9]. Histogram matching is an essential step; without it, RELION placed all particles into a single class.

3.1.2 Generation of hypothetical states

Given an atomic model(s), I have developed methods to generate an ensemble of low energy conformations representative of the lowest energy state(s). In cases where there are two or more conformational states, I use energy-guided state interpolation methods to propose putative intermediate states. Using Rosetta's forcefield, I find energy minima near the interpolated coordinates of the two states that may correspond to stable intermediates I can pull out from the data [54]. To describe non-linear motion, a different approach uses the normal modes to find a state nearby the interpolated coordinates. The normal modes have been shown to describe large scale conformational changes for many proteins; therefore, this method may better approximate the collective motion of the protein [56].

If only a single atomic model is available, I calculate the normal modes for the system and then perturb every alpha-carbon by the low-frequency normal modes. As described above, I can use Rosetta's forcefield to find energy minima near these new coordinates. Generating an ensemble of states can be done in a few different ways: multiple normal modes can be combined and used at once, each of the normal modes can be used separately, or the normal modes can be added or subtracted from the coordinates.

3.1.3 Measures to prevent overfitting

Model bias – generating maps which look more like the model than reality – is a well-identified problem in cryo-EM and multiple measures were used to prevent overfitting at different stages in the pipeline [57-58]. The resolution of the input maps to 3D classification are low-pass filtered to resolutions of at least 5Å and models with deleted regions can be used to generate input maps. If I obtain a higher resolution than was given to the program and if omitted regions improve with the rest of the structure, I can be sure that I am finding information that is not present in the inputs and is driven by the data. At the map refinement stage, the reference map is known to pull the refinement protocol into local minima [9]. To combat this bias, each reference map is low-pass filtered to a resolution of at least 10Å and when appropriate, the consensus map is provided as the reference. The density map obtained

at the end of refinement should reflect the images assigned to the class and not the starting point of the refinement protocol.

Model bias can also occur at the modelling stage: the atomic model can be fit to noise in the density map and the starting model can bias the refinement. The atomic model obtained from the consensus map is used as the starting point for refinement, instead of the model used for 3D classification. To prevent refinement of the atomic model to noise in the density map, half-maps are built from the data, which are reconstructed from only half of the given images, and the atomic model is built into one of the half-maps [47].

3.2 Results on preliminary datasets

3.2.1 Model-guided 3D classification recapitulates traditional methods

First, I tested whether I could pass multiple states to 3D classification and obtain similar resolutions and particle sorting to traditional methods. Using a cryo-EM dataset of a mechanotransduction channel, TRP-NOMPC, I show that feeding in conformational states can perform as well as state of the art methods (Figure 3.2). I recapitulated the image processing of the original paper, using focused classification to sort particles based only on the cytosolic portion of the channel, and found two different conformations [59]. I fit models into the two states, generated maps based on those models and fed them back to 3D classification. The particle sorting (Figure 3.2A) and the maps produced from a single iteration of model-guided image processing (Figure 3.2B) are comparable to what was obtained after 25 iterations of focused classification.

3.2.2 Recapitulating real motion based on a single model

To test our structure generation methods, a small benchmark set of six proteins that exhibited hinge and shear motions was created from a protein motion database [60]. The benchmark set contained two conformations for every protein, one of which would be used as a starting point for state generation. In conjunction with a rotation student, Audrey, we successfully recreated three of the conformational states from this set, finding a structure at least 0.5Å RMSD closer to the final conformation (Figure 3.3A). For three of the six proteins, we were unable to recapitulate the movement between the two states, likely because these proteins exhibited very local motions which the normal modes are unable to capture. An example of this type of localized movement is highlighted in Figure 3.3B, where a small loop of an outer membrane transporter moves approximately 13Å. The normal mode with the lowest RMSD and GDT is shown in Figure 3.3C, and this loop has barely moved. In contrast, larger motions seem to be easily replicated with the state generation method (Figure 3.3D). Applying a single normal mode to a ribose binding protein created a conformational state within 1Å RMSD of the goal structure (Figure 3.3E), and this model was the lowest scoring structure of all the states created when sorting by Rosetta energies (Figure 3.3F). For the other two successes,

applying normal modes in succession – creating a new conformational state with one normal mode and then perturbing that generated state by another normal mode – improved the RMSD to the goal conformation. The models that best recapitulated the motion (lowest RMSDs) were in the top four when sorting by Rosetta energies. Being able to eliminate hypothetical states before 3D classification is essential: this method of applying normal modes in succession would create a minimum of 100 states.

Given the successes on this small benchmark set, I tried to apply these methods to a cryo-EM dataset of a sodium channel that contained two major conformational states [61]. I tried both methods to recreate one of the states from the other, perturbing the model by each of the top 20 normal modes and applying multiple rounds of normal mode perturbation. The latter was the only one that increased the GDT to the native. It seemed like it would be possible to narrow the states down by Rosetta energy as well, as the structures with the lowest GDT values had much lower energies.

3.2.3 Using 3D classification to eliminate states which are not present in the 2D images

With methods to generate hypothetical states in place, I next investigated if 3D classification can accept states that exist within the data and reject those that do not. A preliminary test with the TRP-NOMPC dataset suggested 3D classification could identify both likely and unlikely intermediate states. After fitting models into the two states obtained from focused classification, I generated three intermediate states using each of the three interpolation methods described (see Methods 4.2.2). One of the hypothetical states is highlighted in Figure 3.4A-B. Each of the states from the different methods had proper structural features, scored well in Rosetta, and looked to be a reasonable intermediate given the two starting structures. 3D classification was used to sort particles into the two conformations from focused classification and one of the hypothetical intermediates. The particle sorting and resulting resolution estimates suggest that RELION can see a difference between the hypothetical intermediates and that it can choose a state that is most representative of the data (Figure 3.4C-E). One of the methods for generating intermediates, the torsion method (see Methods 4.2.2 for differences between methods), resulted in three classes with improved local resolutions (Figure 3.4E) compared to the other two methods (Figure 3.4C-D). As an additional test, I generated a state which I did not expect to find in the dataset by extrapolating the motion beyond either of the two input structures. This state received less than 1% of the particles and was resolved to a resolution of 25Å, which suggests that RELION can reject unreasonable states by assigning it very few particles.

Next, I needed to understand how to eliminate states for a system where I start with a single protein structure. I pulled a cryo-EM dataset of a GPCR for which cryoSPARC 3DVA had revealed several types of motion [62]. I used the top 10 normal modes to generate 10 different structures based off the consensus model and passed them all to 3D classification. But it was

not clear which of these models were better, as each of them received a relatively similar number of particles and were reconstructed to similar resolutions. I then tried testing conformational states with fewer maps present. For each normal mode, I ran a round of 3D classification with the consensus map and two hypothetical states, adding and subtracting the applied normal mode. These results were also inconclusive, with each run distributing particles similarly.

Because particle sorting and reconstruction resolutions were uninformative, I investigated other criteria to use for elimination of states. By deleting part of the model and measuring how well the density of the deleted region returned after refinement, I thought I could obtain a more accurate understanding of whether this hypothetical state was contained in the underlying 2D projections. A "successful" state would be one for which the density of the deleted region matched its corresponding input model. With this new measure, I was able to find one state where the density for the deleted region matched the input model better than the consensus model (Supplemental Figure 3.1A). In comparison, there were many normal modes which failed this criterion (Supplemental Figure 3.1B). However, when I attempted to use the successful hypothetical state in subsequent rounds of model-guided image processing and state generation, I was unable to improve the density for the deleted region compared to the first round. This inability to improve with multiple rounds suggests either this metric may not be the proper criteria for elimination or that the 3D classification itself is struggling to properly sort images.

3.2.4 Running iterations of model-guided processing

Finally, I needed to assess whether iterations of model guided processing would lead us closer to the true conformational states in the data. An important underlying assumption of our method, I believed that my hypothetical state could differ from the ground truth state and iterations would allow us to edit and adjust our prediction based on the density information. The sodium channel dataset, which I had tested state generation methods on, was an excellent benchmark dataset: (1) I knew the ground truth consisted of two conformational states, obtained from rounds of traditional 3D classification. (2) I had used the normal modes to try and recreate one of these states from the other. (3) That hypothetical state was not a perfect match for the conformational state which it was meant to replicate. I gave the hypothetical state, in place of the conformational state it was supposed to recreate, and the remaining ground truth state to 3D classification. Unfortunately, the refined density map from the hypothetical state seemed to be an average of the two ground truth states (Figure 3.5A). Running multiple iterations did not bring us closer to the conformations obtained from 3D classification.

I thought to emphasize the differences in the refined map of the hypothetical state by creating a difference map with the consensus map, a common practice for refinement in

crystallography data [63]. A difference map is the map which results from subtracting a simulated density map of the atomic model from the cryo-EM density derived from the data. For my purposes, this map was made manually in Chimera, only subtracting the pieces that needed to move. While this did create a map for which the differences were more obvious (Figure 3.5B), I then needed to refine the model into the new map. This would require a refinement protocol that could automatically fix local mistakes, as most of the structure fits the density well and only a small portion needs to be rebuilt. Current methods in Rosetta are not able to refine this structure into the density (Figure 3.5C).

3.3 Major roadblocks and future directions

3.3.1 Rebuilding local errors in a model

Fixing local errors in a structure is difficult because of a sampling problem: I must find a small subset of residues that poorly fits the density while the rest of the structure matches well. MEDIC, or even a portion of the MEDIC procedure, could be used to solve this problem. One of the strengths of MEDIC is its ability to evaluate local fit-to-density, and it would be possible to only run the density portion of MEDIC's protocol and look for stretches of residues which are fitting the density poorly. Those regions could be deleted and rebuilt with traditional tools in Rosetta or possibly with newer machine learning methods, assuming the map in question is at least 4-4.5Å in resolution.

3.3.2 Finding the ideal number of states in a dataset

For any cryo-EM dataset, depending on the conformational heterogeneity and the number of images, there should be an ideal number of classes to sort the particles into. In a typical image processing pipeline, finding that number is difficult because the 3D classification algorithm is trying to sort images into their respective classes at the same time it is learning what those classes may look like. I may have had difficulty finding the ideal number of states for these datasets because I never had the correct conformational states together at once. I was always supplying the hypothetical state with the consensus map or with other hypothetical states, hoping that RELION would still be able to sort particles accurately. While RELION may be capable of rejecting states with that setup, it could be harder for it to accept states, especially if the motion is continuous or subtle. In support of this hypothesis, the TRP-NOMPC dataset is the only one for which the elimination strategy seemed to work; possibly because I had two states which were already representative of the data fed to RELION. The use of programs like cryoDRGN, which can produce maps for a user specified number of states across a trajectory, would simplify the problem. The maps produced by cryoDRGN would be representative of the underlying data and could be used as seeds for my method.

3.3.3 Future directions

Although model-guided image processing yielded promising results for the TRP-NOMPC dataset, the method struggled to produce high-resolution conformational states for two others: the GPCR and sodium channel. This problem may be too complex for 3D classification, even with the addition of models into the processing pipeline. Deep learning protocols, on the other hand, have been more successful. Current machine learning methods, however, do not incorporate any protein structural information to their analysis and work entirely in the density space. A deep learning protocol may be able to learn the conformational changes present in the data and apply them directly to a protein structure, using a forcefield to maintain physically realistic models. The information gleaned from the structure could be fed back to the other half of the protocol which learns from the raw 2D images – a model-guided deep learning protocol to reveal the protein conformational landscape.

3.4 Acknowledgements

The work described in section 3.2.2 on the larger benchmark set of proteins was conducted by Audrey O'Neil.

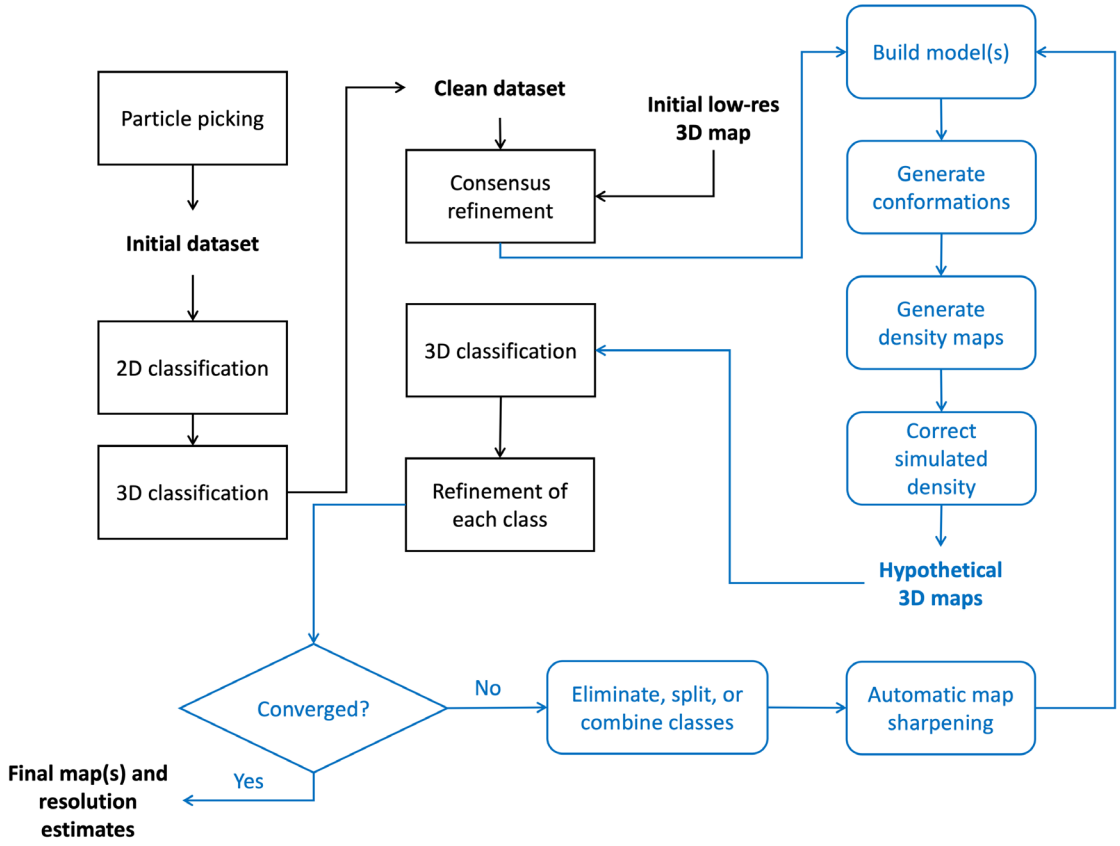


Figure 3.1 Overview of method to process cryo-EM data with atomic models.

Steps for a typical image processing pipeline for a cryo-EM dataset in black with additional steps/inputs for model-guided processing in blue. (back)

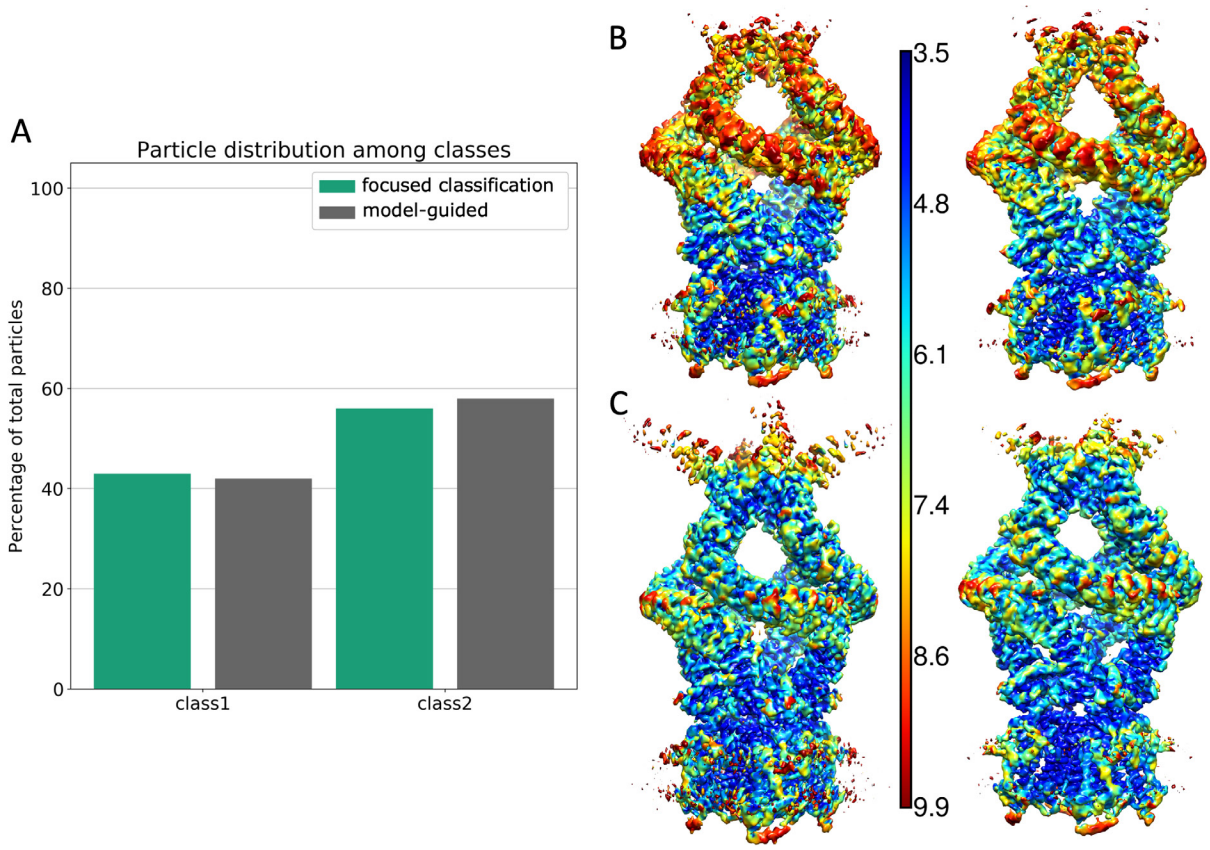


Figure 3.2 One iteration of model-guided image processing recapitulates results from focused classification.

(A) Comparison of particle sorting for class 1 and class 2 for the two methods. **(B)** Density map of class 1 colored by resolution for focused classification **(left)** and model-guided particle sorting **(right)**. **(C)** Density map of class 2 colored by resolution for focused classification **(left)** and model-guided particle sorting **(right)**. (back)

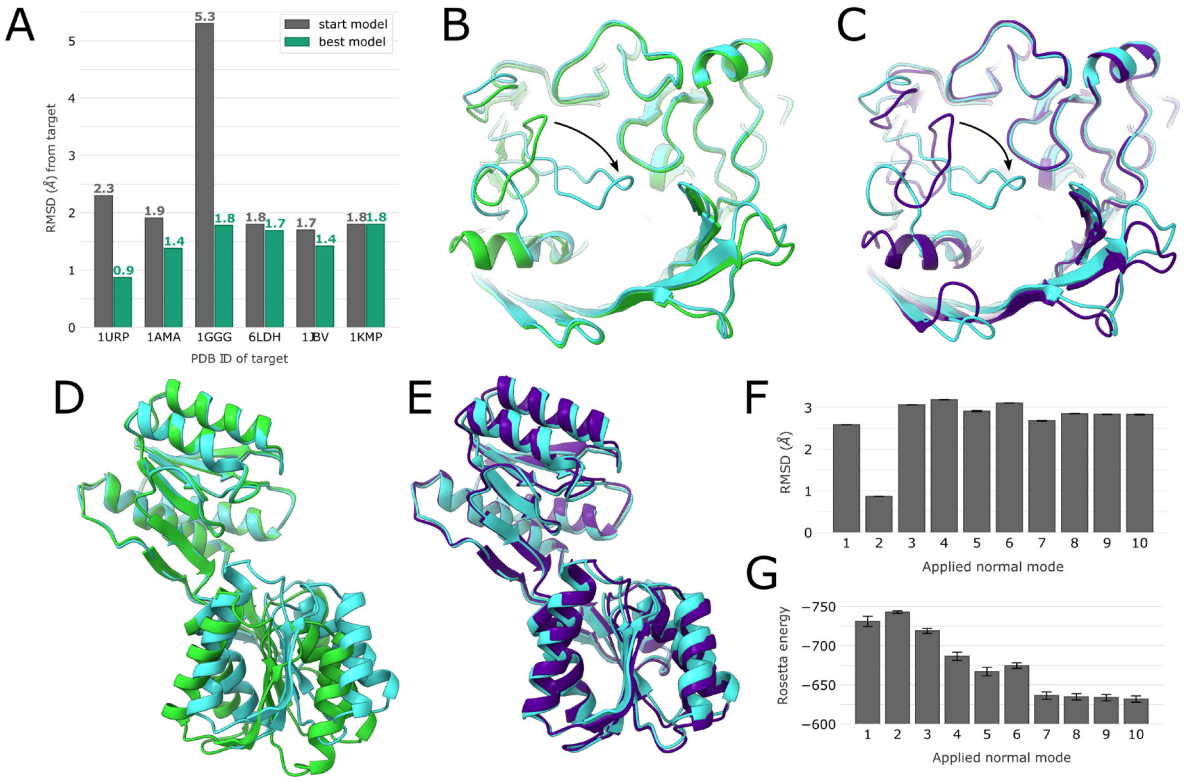


Figure 3.3 Normal modes can recapitulate large protein motions, but not small ones.

(A) RMSD from target structure for starting model and for the best model after the state generation method. **(B)** Starting model (PDB ID 1KMO) in green and target structure (PDB ID 1KMP) in blue. Localized movement of loop highlighted with arrow. **(C)** Best model from normal mode application shown in purple, superimposed on target structure (PDB ID 1KMP) **(D)** Starting model (PDB ID 1BA2) in green and target structure (PDB ID 1URP) in blue. **(E)** Best model from normal mode application shown in purple, superimposed on target structure (PDB ID 1URP) **(F)** RMSD from target structure (PDB ID 1URP) for each model created from the application of a single normal mode. **(E)** Rosetta energies of each generated model, where the model with the lowest RMSD (normal mode 2) has the lowest Rosetta energy. (back)

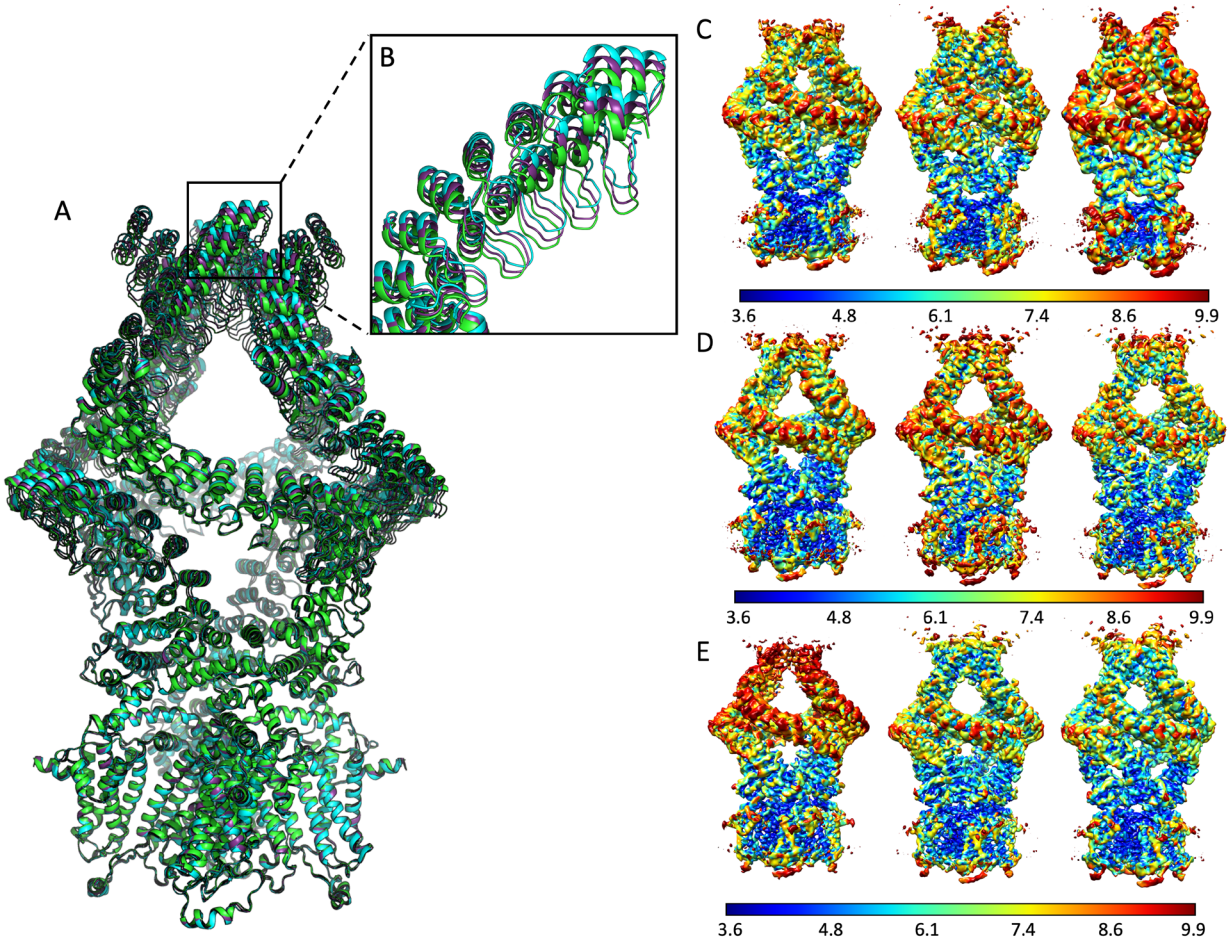


Figure 3.4 3D classification can identify the hypothetical state that best represents the data.

(A) Starting model (green), ending model (blue) and an intermediate (purple) generated by one of the interpolation methods. **(B)** Side view of the first helices in the ankyrin repeat domain. **(C-E)** Density maps of start **(left)**, intermediate **(middle)**, and end **(right)** states for the **(C)** cartesian method, **(D)** normal method and **(E)** torsion method colored by resolution (see Methods 4.2.2 for differences between these methods). (back)

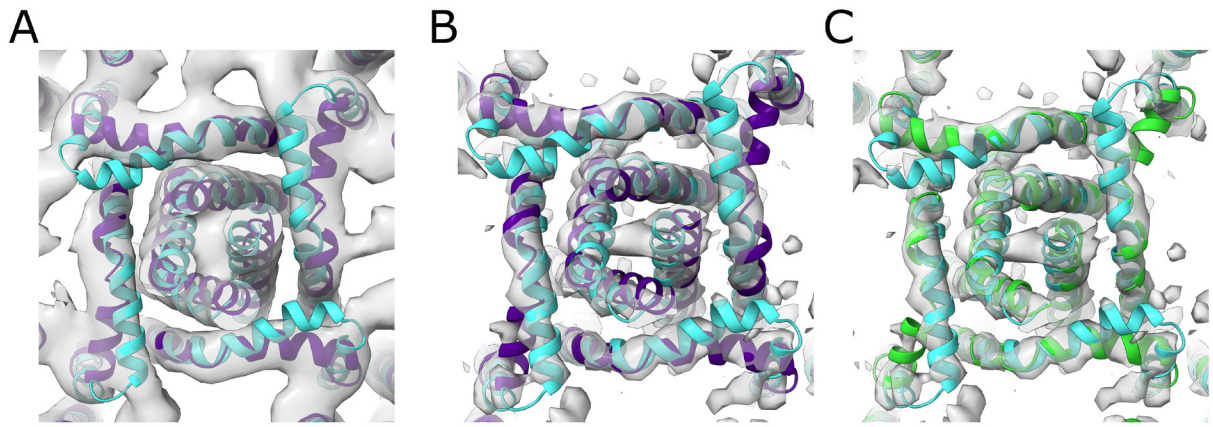


Figure 3.5 Difference maps did not improve refinement of atomic models during model guided image processing.

(A) Density map from one iteration of model guided processing. The ground truth states shown in ribbon (blue, purple). Purple model was used as starting point for hypothetical state generation. Blue model was the final state that needed to be recapitulated. **(B)** Difference map created manually from map shown in **A**. Models shown same as **A**. **(C)** Hypothetical state after refinement, shown in green, into the difference map. Final state to recapitulate shown in blue. (back)

Chapter 4 : Methods

4.1 MEDIC

4.1.1 Preparation of input pdbs

Preparation of pdbs for training or for error detection is a three-step process. First, I remove all ligands, nucleotides, or noncanonical amino acids. Then I refine the structure into the density map, first with cartesian minimization and then with Rosetta's LocalRelax protocol [55]. Finally, I perform B-factor fitting on the refined model. After this, all the scores for the model features can be calculated.

4.1.2 Evaluation of error vs non-error

To determine the threshold at which a residue is an error, I chose a threshold value from the precision-recall curve which balances the two statistical measures. I use both the precision-recall from the 12 rebuilt models and the high-resolution low-resolution validation set to choose thresholds. The precision-recall curve for the 12 rebuilt models was calculated using leave-one out validation to avoid bias. I consider every residue with a probability above 0.78 to be a definite error. At threshold 0.78, MEDIC has a precision of 70% and recall of 80% on the set of 12 rebuilt structures and a precision of 78% and recall of 49% on the validation set. All statistics and data analysis are done only with this more stringent threshold value. I consider every residue with a probability between 0.78 and 0.6 to be a possible error. At a threshold of 0.60, MEDIC has a precision of 52% and recall of 95% on the 12 rebuilt structures and a precision of 68% and recall of 61% on the validation set. Every residue with a probability less than 0.6 is a non-error.

4.1.3 Calculation of error contributions

To determine whether a single feature is predictive of an error, I take the probability equation that I have learned from the final training dataset (Eq. 1), where l is the p-lDDT score, sd the single residue density score, ld the 3-residue density score, r the Ramachandran energy, and b is bond energy:

$$f(x) = 5.15 - 10.41l - 0.38sd - 0.17ld + 0.59r - 0.41b \quad (1)$$

I replace all features, except the ones of interest, with the mean score, derived from the scores of the EMDB depositions (over 1500 cases). For example, I replace p-lDDT, Ramachandran and bond energies with the corresponding mean values to calculate how predictive the density scores are. I then take the result from Eq. 1 and plug it into Eq. 2 to get the final probability. If the final probability is above the threshold for definite errors, then that residue is predicted by a single feature.

$$P = \frac{1}{1 - e^{-f(x)}} \quad (2)$$

4.1.4 Error identification on deposited structures and retraining

I identified cryo-EM structures with less than 2000 residues and a resolution between 3 and 5Å. I then chose 12 structures with diverse topologies and resolutions to run through MEDIC, using the statistical model obtained from training on the obsoleted pdbs. I used a probability threshold of 0.62, derived from the precision-recall curve for the small set of 3 withheld obsoleted structures. I chose a slightly lower threshold, sacrificing precision (60%) for recall (85%) to ensure that I would find most of the errors.

After error identification, I attempted to rebuild every region that was predicted to be an error, following the protocol described below. I then added these 12 models to the training data. I generated error labels by looking first for residues with RMSDs greater than 1.5 after rebuilding, for which the probability was greater than 0.5 and had dropped by 0.2 after fixing. I also labeled residues with RMSDs between 0.5 and 1.5 with probabilities greater than 0.6 and that dropped by 0.2. Any 1-residue errors from this set were removed if they were not within 2 residues of other errors. These labels and scores were passed into the logistic regression with the obsoleted structures, adding an additional 1200 errors to the dataset.

4.1.5 Model rebuilding

For each rebuild, I ran AlphaFold on the sequence [30], docking the model or separately docking its domains into the density using UCSF Chimera [64]. Then, I removed all regions in the deposited model that were identified as errors plus/minus 2-3 residues on either side of the segment. I passed the AlphaFold models and the trimmed deposited model as templates to RosettaCM [23]. I ran at least 2 rounds of iterative RosettaCM, passing the top 5 models out of the total 50 to the next round. Additional rounds were run if model convergence for the top 5 was poor or if additional errors were detected by MEDIC and Molprobity. Any remaining regions which AlphaFold or RosettaCM were not able to fix were built with RosettaES [26]. Success in rebuilding was determined by how well regions matched the density by eye, Molprobity scores, and MEDIC predictions. All images of these structures were made in ChimeraX [64].

4.1.6 High- and low-resolution structure validation

I pulled all cryo-EM structures between 3.5Å and 10Å for which there was another deposition with the same UniProt ID and at least 1Å higher resolution, with a maximum of 3.5Å. If the query structure had a model-map FSC greater than 10Å, the pair was thrown out. From this initial pool of 68 structures, 41 pairs were tossed because there were significant conformational changes caused by image processing, ligand binding, or physiological conditions.

For the remaining 28 pairs of structures, the high-resolution structure was docked and refined into the low-resolution map, and the low-resolution structure was refined into its own density [55]. The backbone RMSD between the two structures was calculated for every residue and all residues with at least 1Å RMSD were labeled as errors. Residues that only existed in one model of the pair were tossed and not used in validation. Error detection was then run on the low-resolution structure using the statistical model from the larger dataset and precision-recall curves were calculated with the described labels.

4.1.7 Error identification in AlphaFold models

We performed error identification on 4 AlphaFold models for the following PDB IDS: 3J9E, 6C0V, 6C14, and 7VOJ. First, we split the AlphaFold predictions into domains, docked them into the density and refined each one separately in Rosetta [23]. Then, we generated labels for the AlphaFold models where every residue for which the backbone RMSD was at least 1Å from our rebuilt model was marked as an error. Disordered regions with little density support were manually removed from the AlphaFold models. A precision-recall curve was generated by passing the refined models to MEDIC and comparing predictions to the labels.

4.1.8 Comparison to Q-scores

To obtain a precision-recall curve for Q-scores, I first generated Q-scores for each residue in the structure. I then subtracted the Q-score for the residue from the expected Q-score based on the global resolution for that map. This procedure mimics the usage of Q-score, where modelers are advised to examine residues which drop below the expected value. The difference between expected and actual Q-score is then used to calculate the precision-recall curve.

4.1.9 Identifying errors in all deposited structures in the EMDB

I pulled every deposited cryo-EM structure with resolutions between 3 and 5Å (1713 structures), removing approximately 300 structures for which the model-map FSC at 0.5 was less than 1/10 $Å^{-1}$. Then I prepared each pdb as described above and ran the statistical model from the combined dataset to detect errors. Of the 1389 structures that met that criteria, MEDIC successfully ran on 1214 (87.4%). To remove regions of disorder, I toss out all residues for which the density cross correlation is less than 0.4 in all subsequent analyses.

4.1.10 Generating a training set for detection of rotamer errors

I added the rotamer errors to the existing training set and validation set of obsoleted pdbs. First, I deleted all sidechains in each structure, so that Rosetta would need to build them from scratch. Then I refine the structure in Rosetta with the LocalRelax protocol, using a high density weight for minimization and a low density weight to pack rotamers. By lowering the density weight for the packing step, Rosetta is allowed to pack rotamers that make energetic

sense, but that may not fit the density. To curate the labels, I called a residue a rotamer error if: (1) the density cross correlation dropped by 0.06, (2) the full atom RMSD was at least 1.5Å, and (3) the backbone RMSD was less than 1Å.

4.1.11 Evaluation of MEDIC using half-maps

To evaluate MEDIC on half-maps, a new branch of MEDIC called "halfmaps" was created on the github and is still available for use. Each of the half-maps was sharpened separately, using the default parameters for *phenix.auto_sharpen* [67].The first sharpened half-map is used for the refinement step of MEDIC. All parameters remained the same for refinement, including the weight on the density. The second sharpened half-map is then used to evaluate the density fit.

4.2 Model-guided image processing

4.2.1 Normal mode state generation

First, I calculate the normal modes of the system, using strategies described in the literature [65-66]. All alpha carbons within a 9Å radius of each other are attached by a spring, creating coupled oscillators. To find the normal modes, I must find the motions where the alpha carbons move at the same frequency. This is accomplished by using Hooke's law to describe the displacement of each mass in relation to each other and creating a matrix that can be used to derive the eigenvectors.

After calculating the normal modes, I must apply them to the structure. Any normal mode or combination of normal modes can be used during this process. For every alpha carbon, I add or subtract the vector from the coordinates, creating a "perturbed" set of coordinates. After the vector algebra, I then calculate the RMSD of the perturbed coordinates from the original. Every normal mode is scaled by a user provided RMSD value divided by the calculated RMSD. If the calculated RMSD was 2.0 and the user provided a value of 4, each vector would be multiplied by a factor of 2. Finally, these coordinates are passed to Rosetta as constraints, using a harmonic function that creates an energy penalty based on the distance of the atom from the desired coordinates. With a Rosetta refinement protocol [54], I essentially pull the starting structure into or near the desired coordinates.

4.2.2 Generation of interpolated states

Given two states, I can describe the movement between them by finding, for every atom, the vector from one state to the other. Creating an intermediate requires only to move along that vector some percentage provided by the user. Three different methods were utilized to get the final structure after obtaining the movement vectors. In the first (Cartesian method), the structure is forced into the intermediate coordinates followed by an energy minimization in Cartesian space. In the second (torsion method), the structure is pulled to the intermediate coordinates by using these coordinates as constraints in a refinement protocol, as described

above [54]. The third method (normal method) uses the normal modes to perturb the vector calculated by the start and end states. After calculating the normal modes, I sum the dot products of the original vector and each of the top 10 normal modes. This summed vector is used to calculate the final coordinates, which are then used as constraints in a refinement.

4.2.3 Density map generation and histogram matching

To generate a density map from an atomic model, I used Rosetta's *pdb_to_map* program, which simulates density using scattering factors for each element and a gaussian sphere whose radius is determined by the desired resolution. This simulated density map is then resampled on the grid of the reconstructed density map obtained from image processing. Then, for the simulated map and the reconstructed density map, I sort the density values at each voxel from smallest to largest in a 1D array. For every voxel, I replace the density value of the generated map with the corresponding value in the sorted array of the reconstructed density map. This will replace all the density values in the generated map with those from the reconstructed map while maintaining the shape of the generated map, matching the histograms.

4.2.3 Passing density maps to 3D classification and refinement

3D classification was done in RELION, as it can be run from the command line and allowed for more fine-tuned control than programs like cryoSPARC or cisTEM. Only a single round of classification was run, typically with similar parameters to those used during 3D classification for the dataset. For example, when doing traditional 3D classification on the Nav dataset, a mask was applied, no angular alignment was allowed, and the regularization parameter was set to a value of 10. I would use all those same parameters in the round of 3D classification.

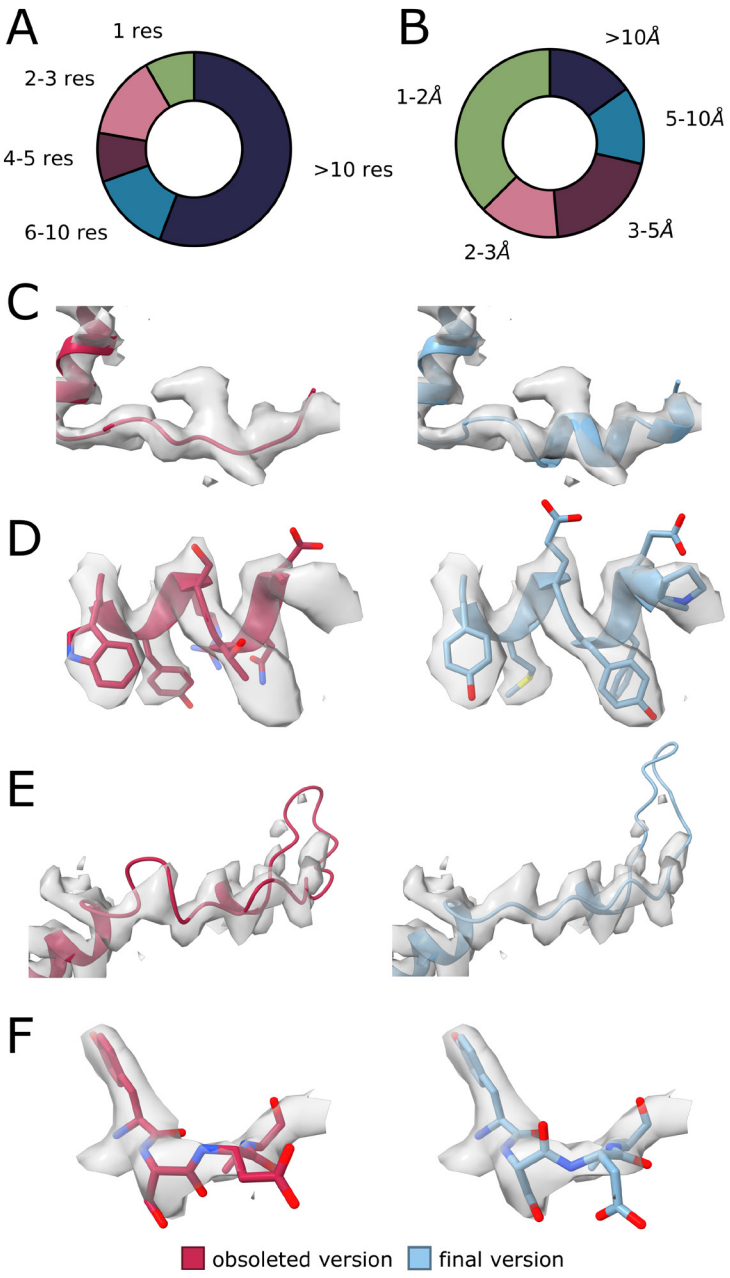
After classification, the particles for each class are split according to their assignment in the StarFile. Each class is refined locally using the consensus map low-pass filtered to 10A as the reference. If a mask was used during refinement for the consensus, that same mask is also used for refinement.

4.2.4 Running iterations of model-guided image processing

A single python script was written to run every step of model-guided image processing and to send the output to the next iteration. The user could specify parameters at all stages if they believed the defaults to be insufficient. The script began by taking atomic models specified by the users, generating density maps, and matching histograms to a user-provided map. These histogram matched maps are then passed to one round of 3D classification in RELION. The StarFile provided by RELION is used to split particles amongst their respective assignments and perform a full FSC Gold Standard refinement on each class. One of the half-maps after refinement is sharpened by *phenix.auto_sharpen,* as it's been shown that Rosetta performs

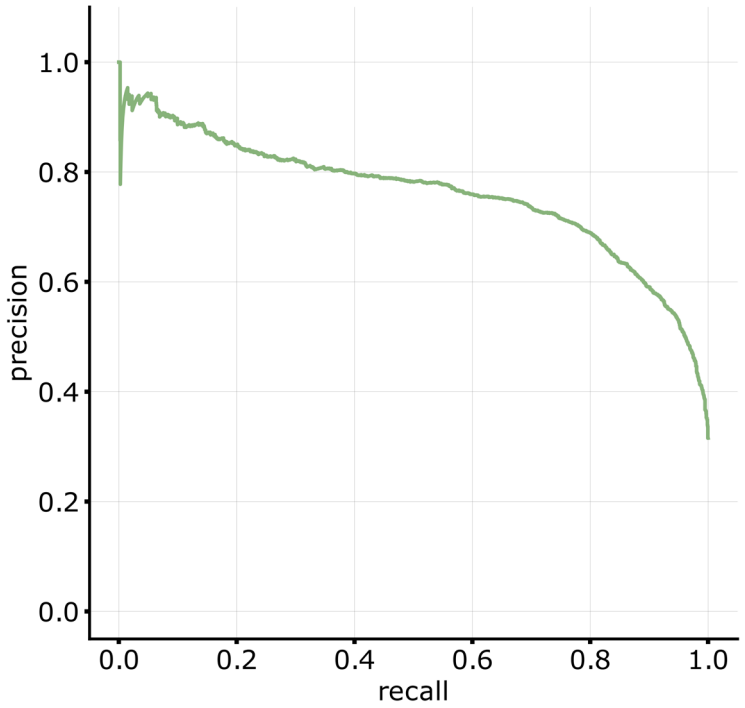
better on sharpened maps [55,67]. Finally, refinement of the atomic model into the sharpened half-map is carried out in Rosetta with a FastRelax protocol augmented by a fit-to-density score term. The top 10% of refined models by score are passed to the next iteration, and density maps will be generated based on these new models. I also collect the orientations assigned by RELION during the refinement stage and pass these to the next iteration. Updating orientations during each iteration stabilized classes and prevented classes from swapping conformations (Supplemental Figure 3.2).

Supplemental



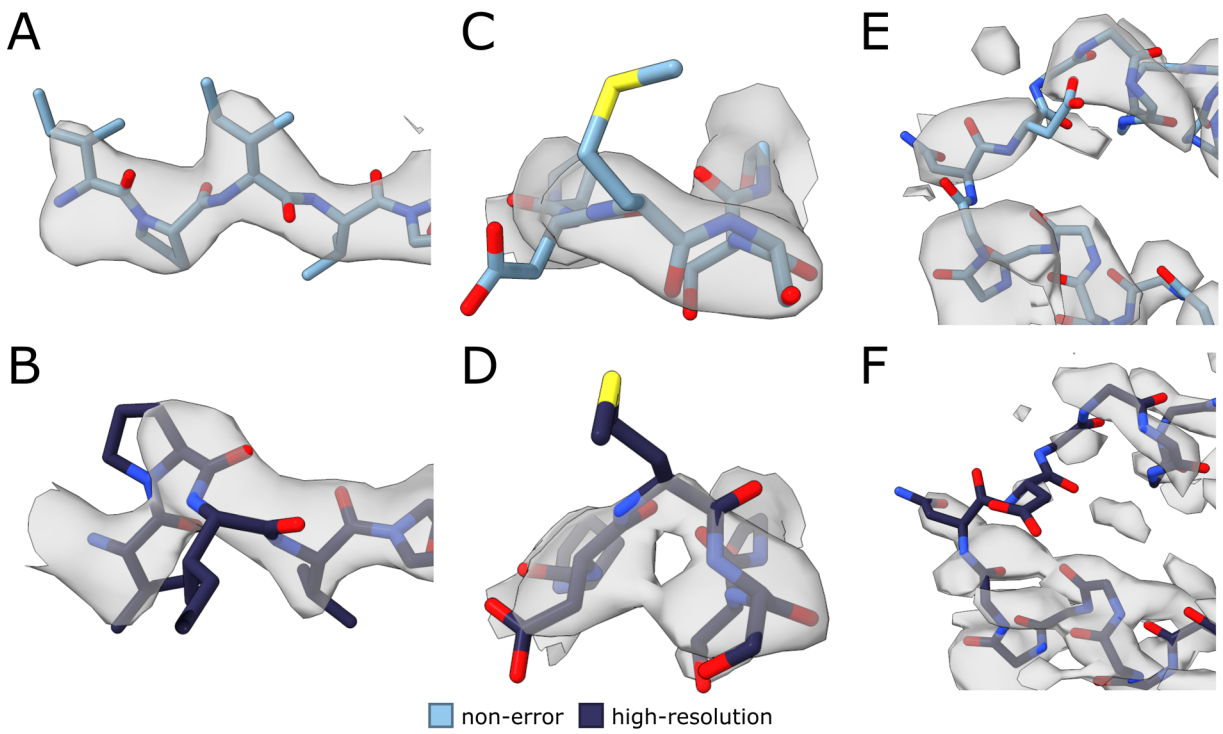
Supplemental Figure 2.1 Initial training set contains a variety of errors made by modelers.

(A) Breakdown of lengths of labeled errors contained in the dataset. **(B)** Breakdown of average RMSDs between the obsoleted and final versions of each segment. **(C)** Improper secondary structure assignment error (residues 11-28R) in 6ZA9 (left) with average RMSD of 7.2Å from final version (right). **(D)** Sequence registry error (residues 907-920A) in 7BV2 (left) with average RMSD of 12.5Å from final version (right). **(E)** Improperly drawn loop (residues 209-230C) in 6K1H (left) with average RMSD of 9.5Å from final version (right). **(F)** Misplaced carbonyl (residue 608A) in 7BV2 (left) with average RMSD of 1.6Å from final version (right). (back)



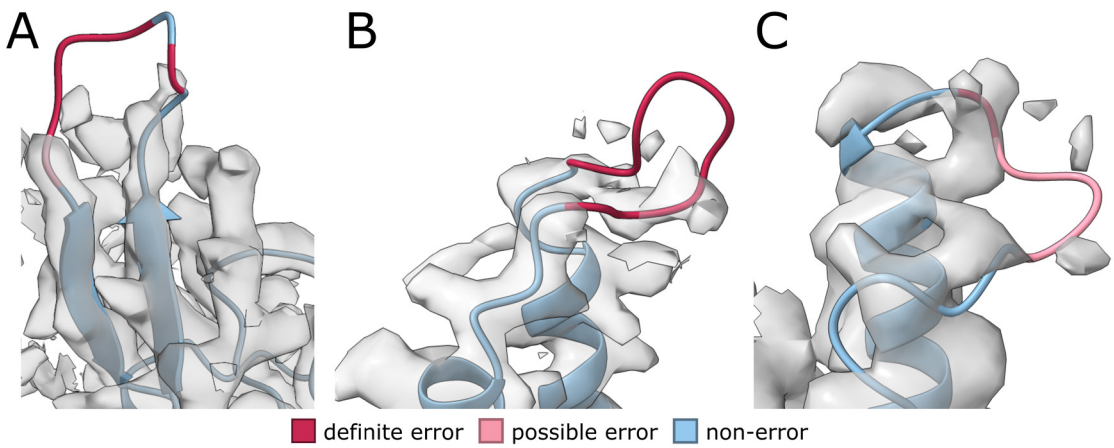
Supplemental Figure 2.2. Precision-recall of MEDIC predictions for the set of 3 withheld obsoleted structures using an early version of MEDIC.

(back)



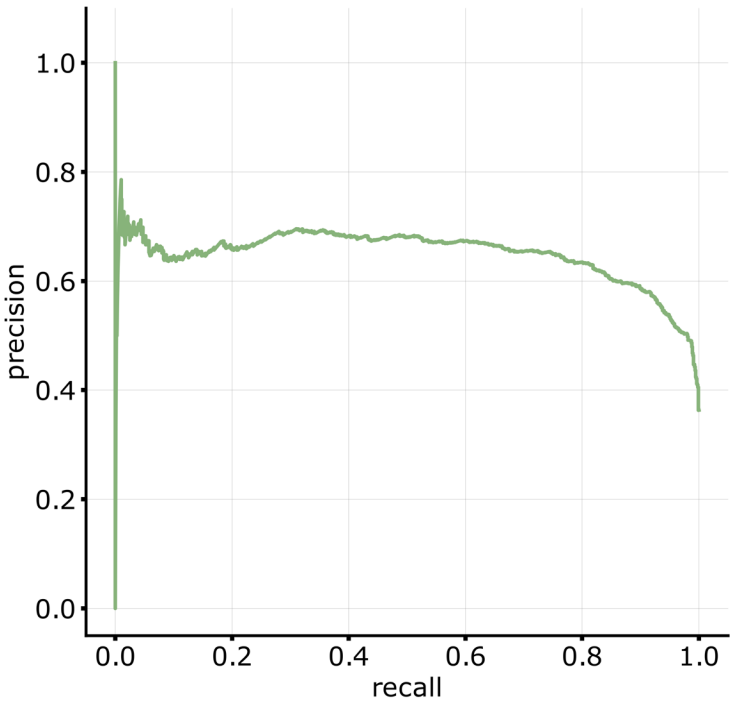
Supplemental Figure 2.3 High resolution structures contain differences which are not supported by the data.

Low-resolution structures **(A, C, E)** colored by MEDIC prediction. **(A)** Low-resolution structure of glutamate hydrogenase (PDB 3JD3, residues 427-431A) that differs from high-resolution. **(B)** High-resolution structure of protein from **A** (PDB 5K12) appears to have an error in this region. **(C)** Loop in glutamate hydrogenase (PDB 3JD3, residues 166-173C). **(D)** High-resolution structure (PDB 5K12) is poorly resolved in the same region from **C**. **(E)** Region from low-resolution structure of TRPV5 (PDB 6PBE, 48-66A) is poorly resolved. **(F)** Corresponding high-resolution structure (PDB 7T6O) is poorly resolved in the same region from **E**. (back)



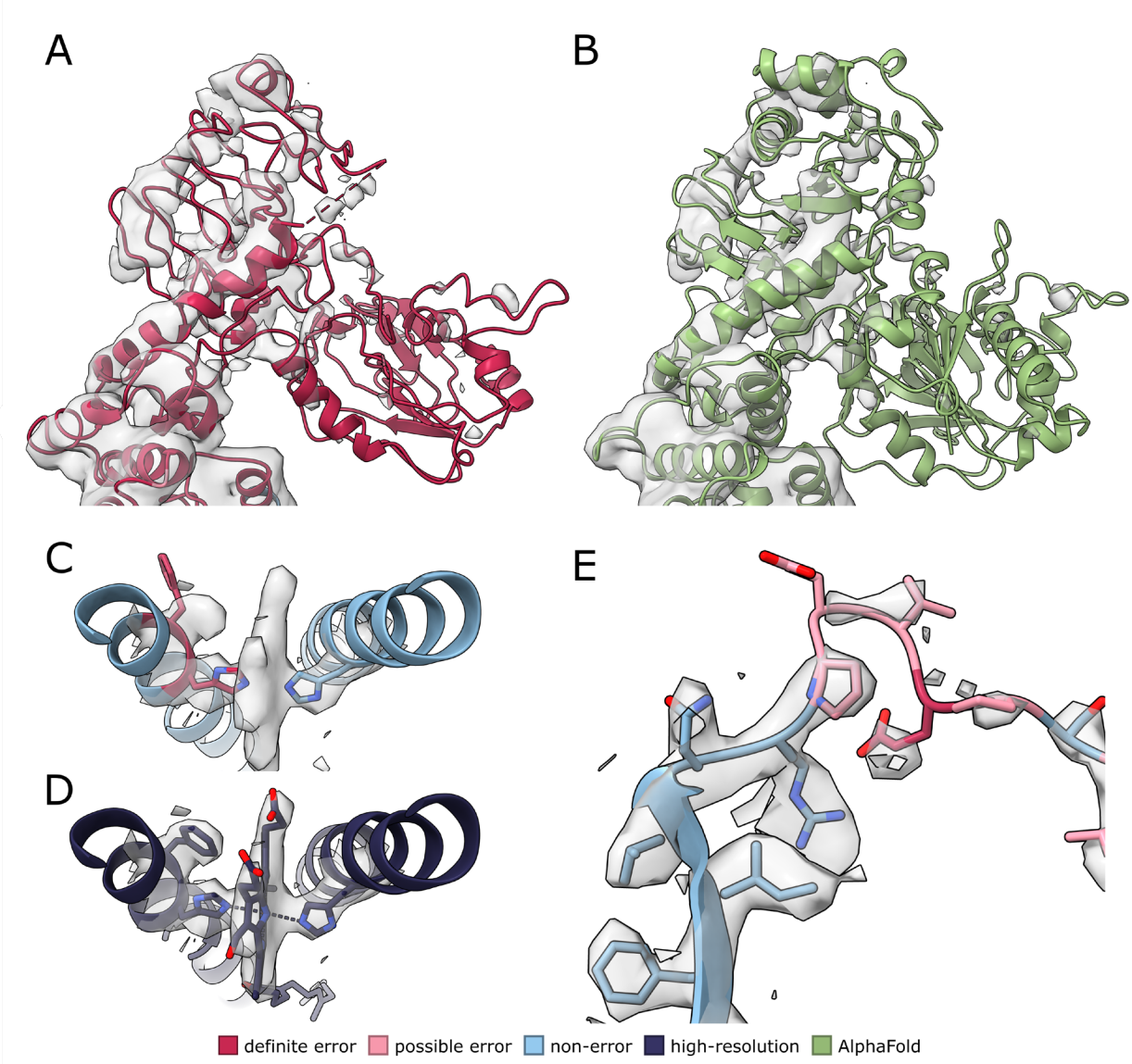
Supplemental Figure 2.4 MEDIC tends to mark disordered regions as errors.

Regions shown are from our new rebuilt models for the following structures: **(A)** neurotoxin (PDB 7QFQ), **(B)** lipid scramblase (PDB 6E1O), **(C)** prestin (PDB 7S9D). (back)



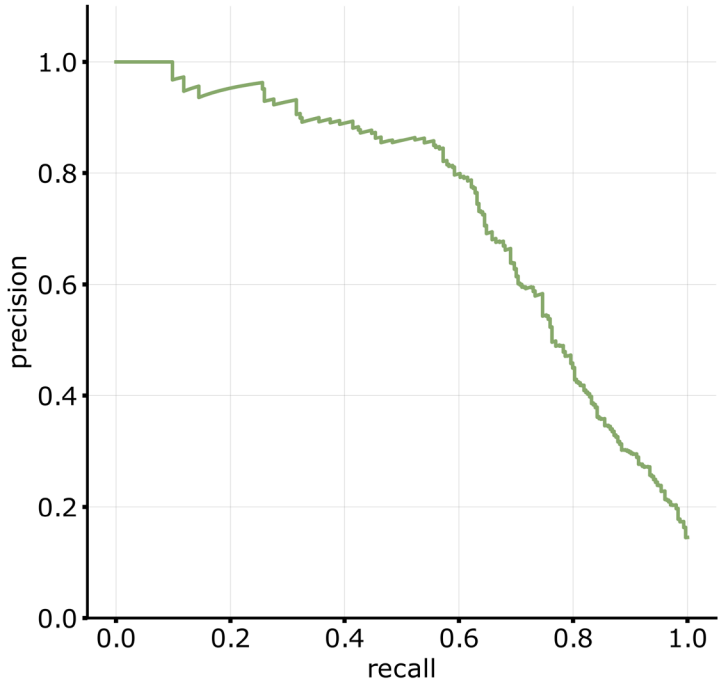
Supplemental Figure 2.5 Precision-recall of MEDIC predictions for the set of 12 rebuilt structures using the full training dataset.

I used leave-one-out validation on each of the models from the set of 12 rebuilt structures to avoid bias in training. (back)



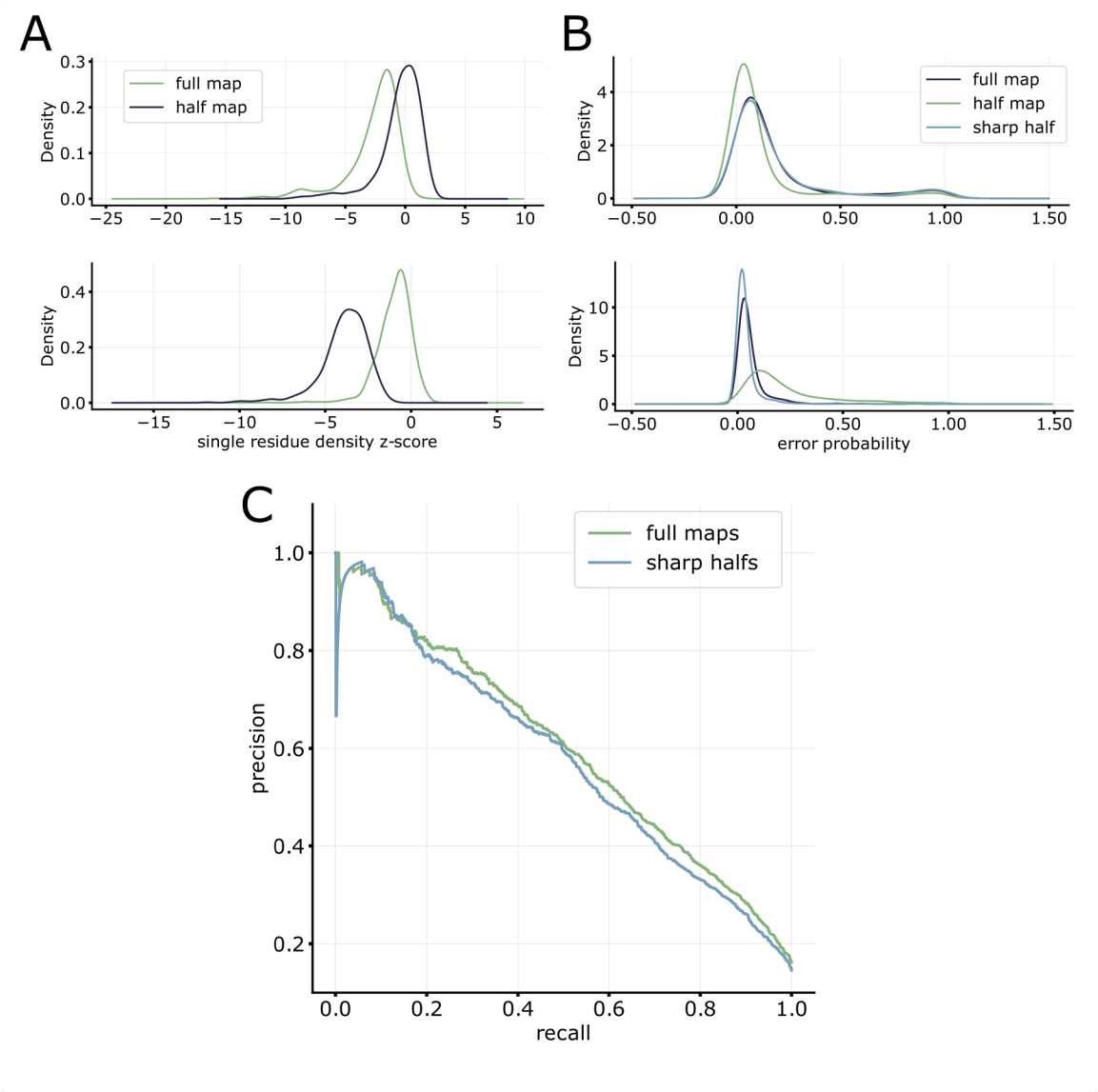
Supplemental Figure 2.6 Examples of error prediction on deposited models in the EMDB.

(A) Domain from L-fucose-1-P guanylyltransferase (PDB 5YYS) colored by MEDIC prediction. **(B)** AlphaFold prediction for protein from **A** docked into the density map. **(C)** Binding residues of cytochrome C oxidase (PDB 5Z62) after refinement and colored by MEDIC prediction. **(D)** Deposited structure of cytochrome C oxidase with ligand bound. **(E)** Rubisco activase complex (PDB 5NV3) colored by MEDIC prediction. (back)



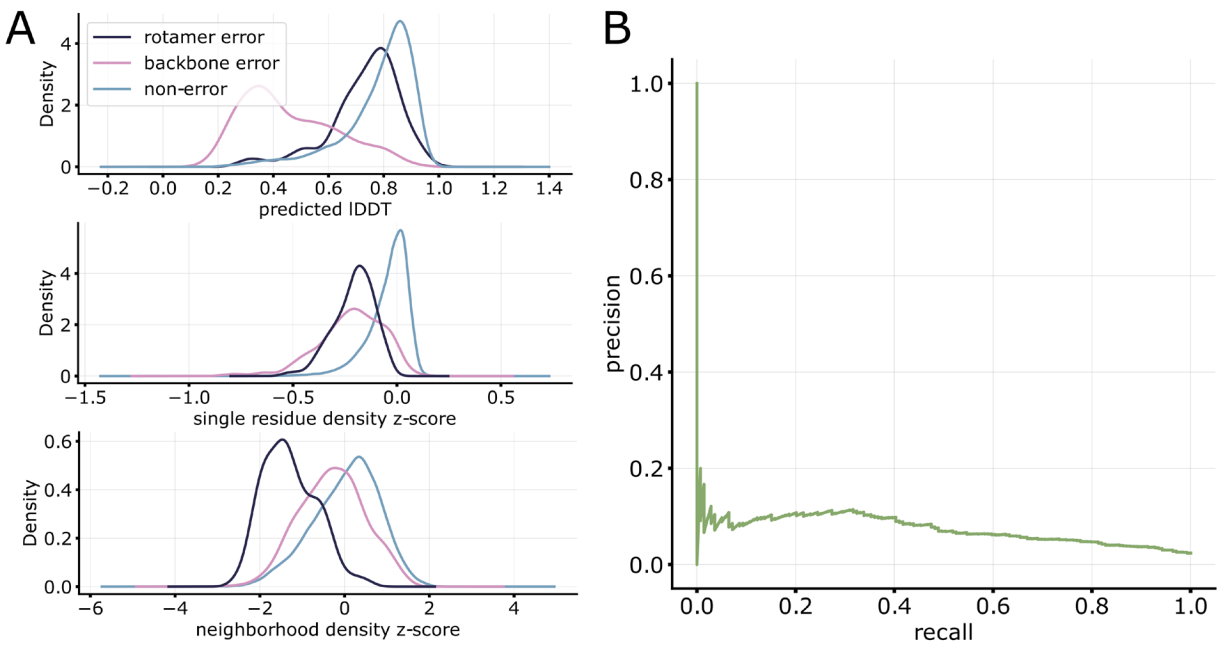
Supplemental Figure 2.7 Precision-recall of MEDIC predictions for 4 refined AlphaFold models.

(back)



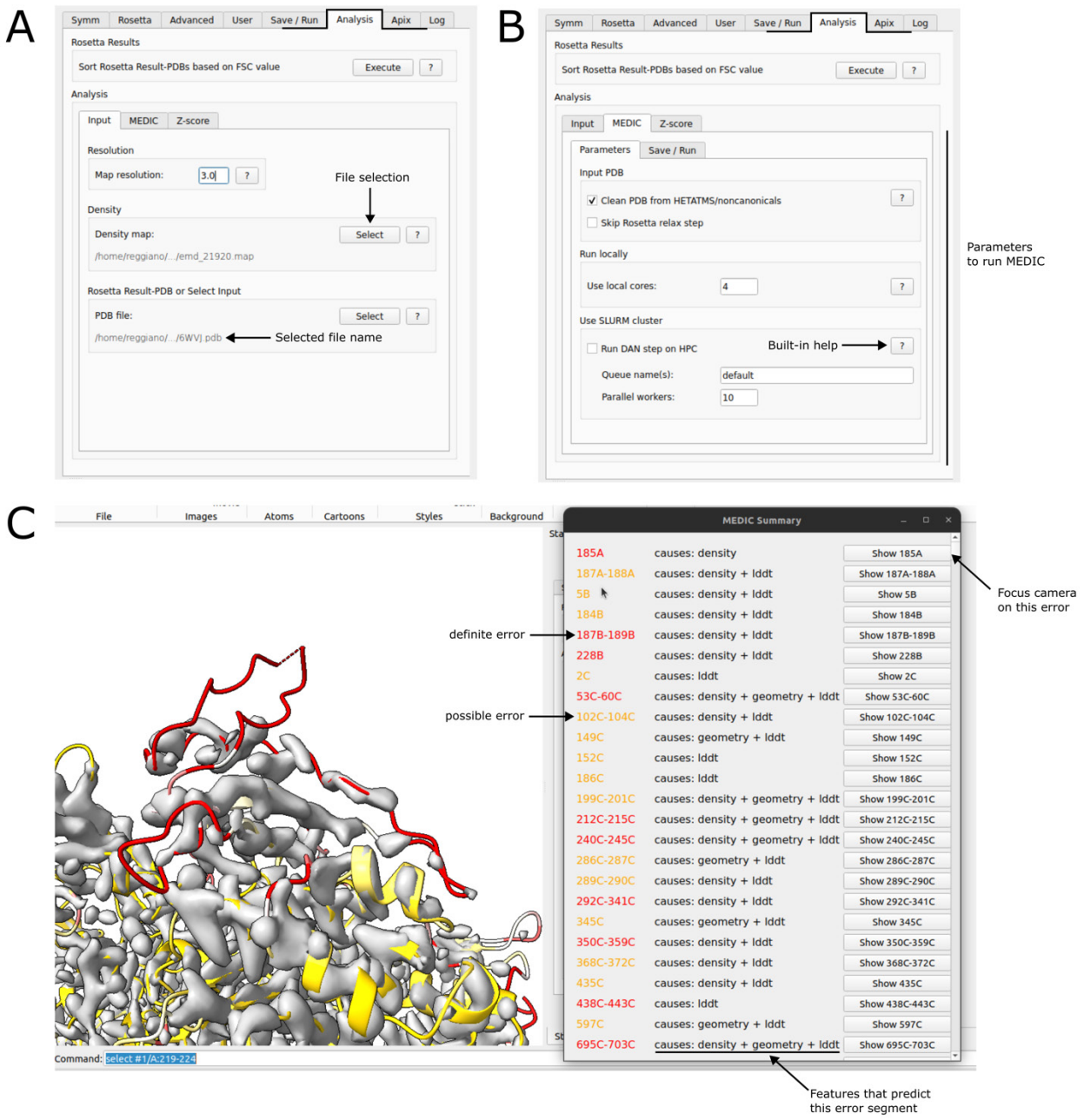
Supplemental Figure 2.8 MEDIC performs similarly on half-maps.

(A) Comparison of single residue density z-score on full-map and half-map evaluation for a ribonucleotide reductase (PDBID 7URG, **top**) and receptor-bound ferrichrome transporter (PDBID 8A8C, **bottom**). **(B)** Error probabilities using full-maps, half-maps, and half-maps sharpened with phenix for same structures as in **A**. **(C)** Precision-recall curve of MEDIC on 11 structures from our high- vs. low-resolution validation set that were deposited with half-maps.
(back)



Supplemental Figure 2.9 Feature breakdown and performance for prediction of rotamer errors.

(A) Density plots of feature scores for rotamer errors, backbone errors, and non-errors in the training data. **(B)** Precision-recall curve of MEDIC predictions for rotamer errors on a small cross-validation set of obsoleted structures. (back)



Supplemental Figure 2.10 MEDIC will be available as a plug-in to ChimeraX.

(A) Under the analysis tab of StarMap, users can provide their density map and atomic model. **(B)** Users can set the parameters required to run MEDIC. **(C)** Users can load in results and view their structure colored by MEDIC predictions, as well as a summary within ChimeraX that lists definite errors and possible errors with the predictive features listed. (back)

EMDB ID	PDB ID	Resolution
EMD-0225	6hin	4.1
EMD-0258	6hrb	4.0
EMD-0605	6o1u	2.8
EMD-0825	6lst	2.3
EMD-11957	7b05	3.8
EMD-12042	7b5o	2.5
EMD-20584	6tys	3.5
EMD-21023	6v20	2.1
EMD-21037	6v3h	3.5
EMD-21199	6vgq	3.5
EMD-22032	6x3t	2.6
EMD-22358	7jjp	1.9
EMD-22754	7k9l	4.9
EMD-2788	4v1w	4.7
EMD-3061	5a63	3.4
EMD-30614	7d8x	2.8
EMD-4553	6qi9	4.6
EMD-6618	5hi9	4.7
EMD-6630	3jcz	3.3
EMD-7088	6be1	4.3
EMD-7793	6d3r	4.3
EMD-8015	5gaq	3.1
EMD-8194	5k12	1.8
EMD-9062	6mar	4.5

Supplementary Table 2.1 Density maps and models used to fit density z-scores

	6CP3		6K1H		6TT7	
	Version 1	Version 2	Version 1	Version 2	Version 1	Version 2
R.M.S. deviations						
Bond lengths (Å)	0.0093	0.0093	0.0074	0.0070	0.0094	0.0086
Bond angles (°)	1.38	1.37	1.44	1.23	1.50	1.48
Validation						
Molprobity Score	1.92	1.98	1.89	2.43	2.05	2.02
Clash Score	8.12	9.09	5.99	20.08	8.56	8.68
Poor Rotamers (%)	0.36	1.07	0.75	0.90	1.29	1.19
Ramachandran Plot						
Favored (%)	92.18	92.41	89.41	86.50	91.41	91.89
Allowed (%)	7.62	7.29	10.59	13.31	8.03	7.58
Outliers (%)	0.20	0.30	0	0.19	0.56	0.53
	6ZA9		7BV2		7BW4	
	Version 1	Version 2	Version 2	Version 3	Version 1	Version 2
R.M.S. deviations						
Bond lengths (Å)	0.0072	0.0079	0.0045	0.050	0.0046	0.0046
Bond angles (°)	1.38	1.37	0.98	1.28	0.99	0.98
Validation						
Molprobity Score	1.55	1.55	1.48	1.28	2.21	2.20
Clash Score	4.72	5.20	5.32	5.27	15.96	16.36
Poor Rotamers (%)	0.84	0.74	0	0	1.68	1.68
Ramachandran Plot						
Favored (%)	95.57	96.04	96.78	98.70	95.23	95.47
Allowed (%)	4.43	3.96	3.22	1.3	4.63	4.34
Outliers (%)	0	0	0	0	0.09	0.19

	7KSM	
	Version 1	Version 2
R.M.S. deviations		
Bond lengths (Å)	0.0045	0.0070
Bond angles (°)	0.97	1.05
Validation		
Molprobity Score	1.80	1.74
Clash Score	9.49	9.09
Poor Rotamers (%)	0.04	0.08
Ramachandran Plot		
Favored (%)	95.73	95.56
Allowed (%)	4.27	4.44
Outliers (%)	0	0

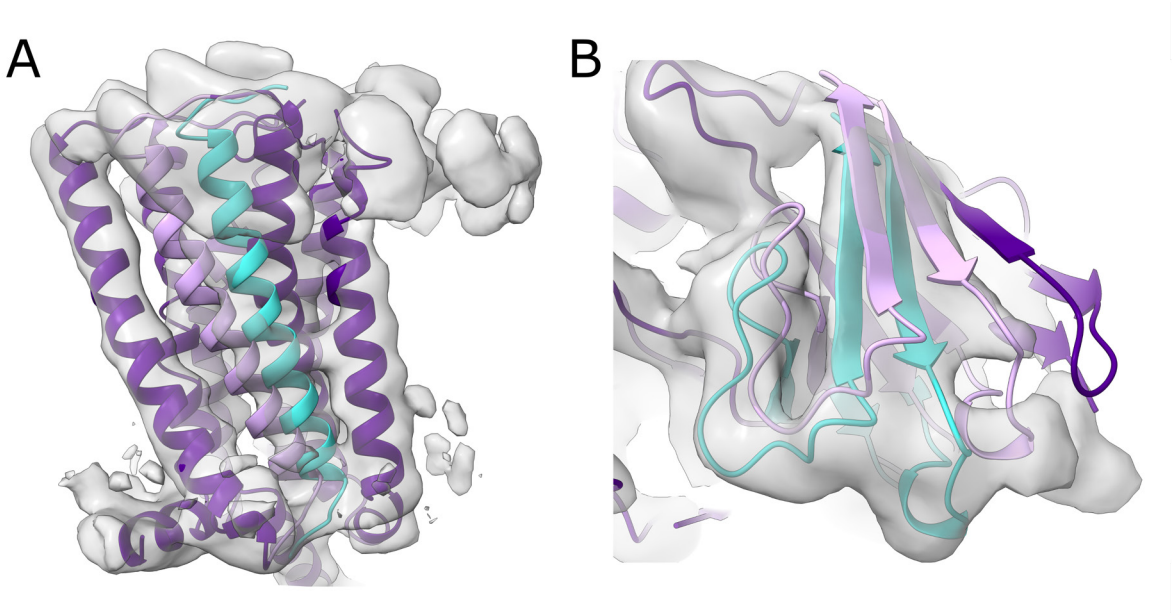
Supplementary Table 2.2 Validation statistics on seven obsoleted structures

	3J9E		5MM4		6C0V	
	before	after	before	after	before	after
R.M.S. deviations						
Bond lengths (Å)	0.0314	0.0319	0.0086	0.0227	0.0073	0.0236
Bond angles (°)	3.27	1.83	1.36	1.73	1.02	1.69
Validation						
Molprobity Score	3.48	0.93	2.18	0.99	0.92	0.86
Clash Score	20.69	1.58	12.45	2.17	0.33	1.32
Poor Rotamers (%)	24.26	0	0.1	0	0	0.1
Ramachandran Plot						
Favored (%)	87.02	97.88	89.17	98.56	95.73	98.87
Allowed (%)	9.56	1.93	10.73	1.44	4.27	1.04
Outliers (%)	3.42	0.19	0.1	0	0	0.09

	6C14		6DMB		6E1O	
	before	after	before	after	before	after
R.M.S. deviations						
Bond lengths (Å)	0.0058	0.0273	0.0091	0.0261	0.0085	0.0296
Bond angles (°)	1.25	1.9	1.56	1.74	1.44	1.73
Validation						
Molprobity Score	1.72	1.17	2.21	0.88	1.68	0.95
Clash Score	5.29	3.23	6.47	1.46	4.62	1.86
Poor Rotamers (%)	0.27	0	1.94	0.24	0.78	0
Ramachandran Plot						
Favored (%)	93.25	97.76	86.21	99.19	93.11	99.23
Allowed (%)	6.75	2.02	13.18	0.61	6.89	0.77
Outliers (%)	0	0.22	0.61	0.2	0	0

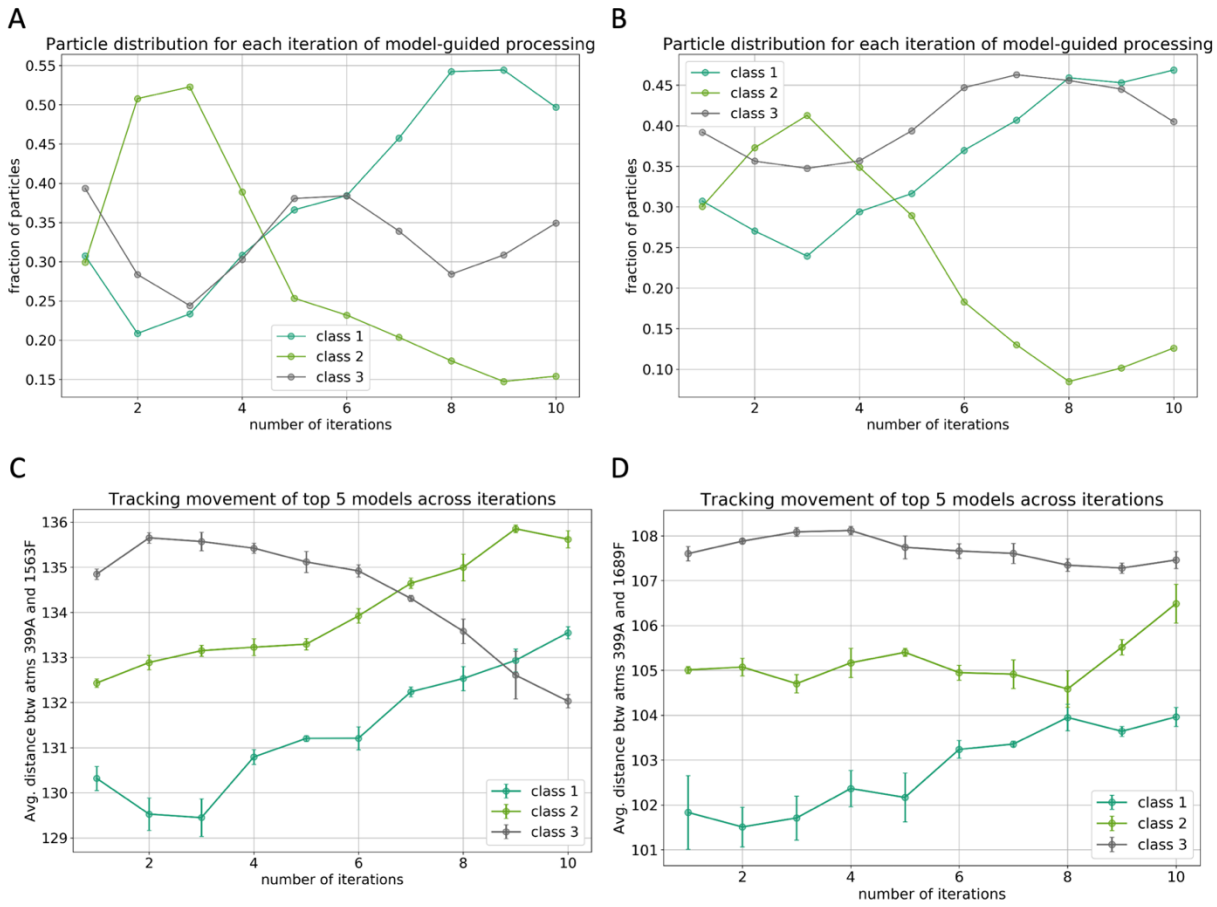
	6JT1		6XOW		7QFQ	
	before	after	before	after	before	after
R.M.S. deviations						
Bond lengths (Å)	0.0049	0.0225	0.0068	0.0257	0.0036	0.0209
Bond angles (°)	1.1	1.63	1.26	1.86	0.93	1.65
Validation						
Molprobity Score	2.45	0.83	2.16	1.15	1.82	0.76
Clash Score	10.39	1.18	24.2	3.6	8.31	0.85
Poor Rotamers (%)	4.62	0	0.23	0.12	0	0
Ramachandran Plot						
Favored (%)	93.82	98.64	95.84	98.65	94.57	98.84
Allowed (%)	5.82	1.18	4.16	1.35	5.43	1.08
Outliers (%)	0.36	0.18	0	0	0	0.08
	7R9U		7S9D		7VOJ	
	before	after	before	after	before	after
R.M.S. deviations						
Bond lengths (Å)	0.0034	0.0264	0.0068	0.0258	0.0087	0.0315
Bond angles (°)	0.86	1.94	1.11	1.8	1.25	1.87
Validation						
Molprobity Score	1.68	1.06	2.44	1.03	1.83	0.85
Clash Score	3.5	2.72	23.34	2.45	7.62	1.26
Poor Rotamers (%)	0	0.49	0.35	0.17	1.73	0
Ramachandran Plot						
Favored (%)	90.26	99.4	89	98.81	96.46	97.98
Allowed (%)	9.14	0.6	10.7	1.19	3.29	2.02
Outliers (%)	0.6	0	0.3	0	0.25	0

Supplementary Table 2.3 Validation statistics on set of 12 rebuilt models before and after.



Supplemental Figure 3.1 Identification of hypothetical states that may or may not exist in the data.

(A) Example of hypothetical state for GPCR (purple) which passes the model deletion and density recovery criterion. Deleted region of model shown in light purple and corresponding region of consensus model shown in blue. **(B)** Example of hypothetical state (purple) which fails the model deletion and density recovery criteria. Deleted region of model shown in light purple and corresponding region of consensus model shown in blue. (back)



Supplemental Figure 3.2 Orientations from map refinements of previous iterations improve convergence of model-guided image processing

(A) Particle distributions for every class across iterations of model-guided image processing where each iteration starts from orientations found during consensus refinement. **(B)** Particle distributions for every class where orientations after refinement of each class are passed to the following iteration. **(C)** One-dimensional representation of protein conformations for every class where each iteration starts from orientations found during consensus refinement. **(D)** One-dimensional representation of protein conformations for every class where orientations after refinement of each class are passed to the following iteration. (back)

Supplemental Figure 5.1 Some of the items crocheted by this author over the course of her graduate career

(A) Totoro, the first freehanded amigurumi. **(B)** A very cute sweater, the only wearable made so far (of many) that is actually wearable. **(C)** Two model organisms, drosophila and zebrafish (plus one Pesto). **(D)** Blueberry pie, probably the only pie this author will ever make. **(E)** Ice cream cone, triple scoop, because less is not always more. **(F)** A Junimo.

Bibliography

1. Li, X., Gu, J., & Zhou, Q. (2015). Review of aerobic glycolysis and its key enzymes - new targets for lung cancer therapy. *Thoracic Cancer*, *1*, 17–24. https://doi.org/10.1111/1759-7714.12148
2. Kuo, I. Y., & Ehrlich, B. E. (2015). Signaling in Muscle Contraction. *Cold Spring Harbor Perspectives in Biology*, *2*, a006023. https://doi.org/10.1101/cshperspect.a006023
3. Akram, M. (2013). Mini-review on glycolysis and cancer. *Journal of Cancer Education*, *28*(3), 454–457. https://doi.org/10.1007/s13187-013-0486-9
4. Shi, Y. (2014). A Glimpse of Structural Biology through X-Ray Crystallography. *Cell*, *5*, 995–1014. https://doi.org/10.1016/j.cell.2014.10.051
5. Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G. V., Christie, C. H., Dalenberg, K., Di Costanzo, L., Duarte, J. M., Dutta, S., Feng, Z., Ganesan, S., Goodsell, D. S., Ghosh, S., Green, R. K., Guranović, V., Guzenko, D., Hudson, B. P., … Zhuravleva, M. (2020). RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Research*, *D1*, D437–D451. https://doi.org/10.1093/nar/gkaa1038
6. Bammes, B. E., Rochat, R. H., Jakana, J., Chen, D.-H., & Chiu, W. (2012). Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3/4 Nyquist frequency. *Journal of Structural Biology*, *3*, 589–601. https://doi.org/10.1016/j.jsb.2012.01.008
7. Zivanov, J., Nakane, T., & Scheres, S. H. W. (2019). A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ*, *1*, 5–17. https://doi.org/10.1107/s205225251801463x
8. Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, *3*, 290–296. https://doi.org/10.1038/nmeth.4169
9. Scheres, S. H. W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, *3*, 519–530. https://doi.org/10.1016/j.jsb.2012.09.006
10. Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P. M. G. E., Grigoras, I. T., Malinauskaite, L., Malinauskas, T., Miehling, J., Uchański, T., Yu, L., Karia, D., Pechnikova, E. V., de Jong, E., Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., … Scheres, S. H. W. (2020). Single-particle cryo-EM at atomic resolution. *Nature*, *7832*, 152–156. https://doi.org/10.1038/s41586-020-2829-0
11. Lawson, C., Patwardhan, A., Pintilie, G. D., Sanz Garcia, E., Lagerstedt, I., Baker, M. L., Sala, R., Ludtke, S. J., Berman, H. M., Kleywegt, G., & Chiu, W. (2013). Emdatabank: Unified Data Resource for 3DEM. *Biophysical Journal*, *2*, 351a. https://doi.org/10.1016/j.bpj.2012.11.1950
12. Nguyen, T. H. D., Galej, W. P., Bai, X., Oubridge, C., Newman, A. J., Scheres, S. H. W., & Nagai, K. (2016). Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature*, *7590*, 298–302. https://doi.org/10.1038/nature16940

13. Davis, J. H., Tan, Y. Z., Carragher, B., Potter, C. S., Lyumkis, D., & Williamson, J. R. (2016). Modular Assembly of the Bacterial Large Ribosomal Subunit. *Cell*, *6*, 1610-1622.e15. https://doi.org/10.1016/j.cell.2016.11.020
14. Punjani, A., & Fleet, D. J. (2021a). 3D Flexible Refinement: Structure and Motion of Flexible Proteins from Cryo-EM. *BioArchiv*. https://doi.org/10.1101/2021.04.22.440893
15. Nakane, T., Kimanius, D., Lindahl, E., & Scheres, S. H. (2018). Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *ELife*. https://doi.org/10.7554/elife.36861
16. Punjani, A., & Fleet, D. J. (2021b). 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of Structural Biology*, *2*, 107702. https://doi.org/10.1016/j.jsb.2021.107702
17. Zhong, E. D., Bepler, T., Berger, B., & Davis, J. H. (2021). CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods*, *2*, 176–185. https://doi.org/10.1038/s41592-020-01049-4
18. Frank, J., & Ourmazd, A. (2016). Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods*, 61–67. https://doi.org/10.1016/j.ymeth.2016.02.007
19. Croll, T. I. (2018). ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallographica Section D Structural Biology*, *6*, 519–530. https://doi.org/10.1107/s2059798318002425
20. DiMaio, F., Tyka, M. D., Baker, M. L., Chiu, W., & Baker, D. (2009). Refinement of Protein Structures into Low-Resolution Density Maps Using Rosetta. *Journal of Molecular Biology*, *1*, 181–190. https://doi.org/10.1016/j.jmb.2009.07.008
21. Song, Y., DiMaio, F., Wang, R. Y.-R., Kim, D., Miles, C., Brunette, T., Thompson, J., & Baker, D. (2013). High-Resolution Comparative Modeling with RosettaCM. *Structure*, *10*, 1735–1742. https://doi.org/10.1016/j.str.2013.08.005
22. Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., & Schwede, T. (2008). Protein structure homology modeling using SWISS-MODEL workspace. *Nature Protocols*, *1*, 1–13. https://doi.org/10.1038/nprot.2008.197
23. Cowtan, K. (2006). The "Buccaneer" software for automated protein chain tracing. *Acta Crystallographica Section A Foundations of Crystallography*, *a1*, s87–s87. https://doi.org/10.1107/s0108767306098266
24. Afonine, P. V., Poon, B. K., Read, R. J., Sobolev, O. V., Terwilliger, T. C., Urzhumtsev, A., & Adams, P. D. (2018). Real-space refinement in *PHENIX* for cryo-EM and crystallography. *Acta Crystallographica Section D Structural Biology*, *6*, 531–544. https://doi.org/10.1107/s2059798318006551
25. Wang, R. Y.-R., Kudryashev, M., Li, X., Egelman, E. H., Basler, M., Cheng, Y., Baker, D., & DiMaio, F. (2015). De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nature Methods*, *12*(4), 335–338. https://doi.org/10.1038/nmeth.3287
26. Frenz, B., Walls, A. C., Egelman, E. H., Veesler, D., & DiMaio, F. (2017). RosettaES: a sampling strategy enabling automated interpretation of difficult cryo-EM maps. *Nature Methods*, *8*, 797–800. https://doi.org/10.1038/nmeth.4340

27. Terwilliger, T. C., Adams, P. D., Afonine, P. V., & Sobolev, O. V. (2018). A fully automatic method yielding initial models from high-resolution cryo-electron microscopy maps. *Nature Methods*, *11*, 905–908. https://doi.org/10.1038/s41592-018-0173-1
28. Terashi, G., & Kihara, D. (2018). De novo main-chain modeling for EM maps using MAINMAST. *Nature Communications*, *1*. https://doi.org/10.1038/s41467-018-04053-7
29. Fontana, P., Dong, Y., Pi, X., Tong, A. B., Hecksel, C. W., Wang, L., Fu, T.-M., Bustamante, C., & Wu, H. (2022). Structure of cytoplasmic ring of nuclear pore complex by integrative cryo-EM and AlphaFold. *Science*, *6598*. https://doi.org/10.1126/science.abm9326
30. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *7873*, 583–589. https://doi.org/10.1038/s41586-021-03819-2
31. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., … Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, *6557*, 871–876. https://doi.org/10.1126/science.abj8754
32. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Santos Costa, A. dos, Fazel-Zarandi, M., Sercu, T., Candido, S., & Rives, A. (2022). *Language models of protein sequences at the scale of evolution enable accurate structure prediction*. https://doi.org/10.1101/2022.07.20.500902
33. Perrakis, A., & Sixma, T. K. (2021). AI revolutions in biology. *EMBO Reports*, *11*. https://doi.org/10.15252/embr.202154046
34. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., … Hassabis, D. (2021). *Protein complex prediction with AlphaFold-Multimer*. https://doi.org/10.1101/2021.10.04.463034
35. Terwilliger, T. C., Poon, B. K., Afonine, P. V., Schlicksup, C. J., Croll, T. I., Millán, C., Richardson, Jane. S., Read, R. J., & Adams, P. D. (2022). Improved AlphaFold modeling with implicit experimental information. *Nature Methods*. https://doi.org/10.1038/s41592-022-01645-6
36. Pfab, J., Phan, N. M., & Si, D. (2020). DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes. *Proceedings of the National Academy of Sciences*, *2*. https://doi.org/10.1073/pnas.2017525118
37. Jamali, K., Kimanius, D., & Scheres, S. (2022). ModelAngelo: Automated Model Building in Cryo-EM Maps. *ArXhiv*. https://doi.org/ https://doi.org/10.48550/arXiv.2210.00006
38. He, J., Lin, P., Chen, J., Cao, H., & Huang, S.-Y. (2022). Model building of protein complexes from intermediate-resolution cryo-EM maps with deep learning-guided automatic assembly. *Nature Communications*, *1*. https://doi.org/10.1038/s41467-022-31748-9
39. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2009). Molprobity: All-atom structure

validation for Macromolecular Crystallography. *Acta Crystallographica Section D Biological Crystallography*, *66*(1), 12–21. https://doi.org/10.1107/s0907444909042073
40. Prisant, M. G., Williams, C. J., Chen, V. B., Richardson, J. S., & Richardson, D. C. (2019). New tools in molprobity validation: Cablam for CryoEM Backbone, UnDowser to rethink "Waters," and NGL viewer to recapture online 3D graphics. *Protein Science*, *29*(1), 315–329. https://doi.org/10.1002/pro.3786
41. Pintilie, G., Zhang, K., Su, Z., Li, S., Schmid, M. F., & Chiu, W. (2020). Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nature Methods*, *17*(3), 328–334. https://doi.org/10.1038/s41592-020-0731-1
42. Pintilie, G., & Chiu, W. (2021). Validation, analysis and annotation of Cryo-EM Structures. *Acta Crystallographica Section D Structural Biology*, *77*(9), 1142–1152. https://doi.org/10.1107/s2059798321006069
43. Barad, B. A., Echols, N., Wang, R. Y.-R., Cheng, Y., DiMaio, F., Adams, P. D., & Fraser, J. S. (2015). Emringer: Side chain–directed model and map validation for 3D cryo-electron microscopy. *Nature Methods*, *12*(10), 943–946. https://doi.org/10.1038/nmeth.3541
44. Ramírez-Aportela, E., Maluenda, D., Fonseca, Y. C., Conesa, P., Marabini, R., Heymann, J. B., Carazo, J. M., & Sorzano, C. O. (2021). FSC-Q: A Cryoem Map-to-atomic model quality validation based on the local Fourier shell correlation. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-020-20295-w
45. Istrate, A., Wang, Z., Murshudov, G. N., Patwardhan, A., & Kleywegt, G. J. (2021). 3D-strudel - a novel model-dependent map-feature validation method for high-resolution cryo-EM structures. https://doi.org/10.1101/2021.12.16.472999
46. Williams, Christopher Joseph (2015). *Using C-Alpha Geometry to Describe Protein Secondary Structure and Motifs.* Dissertation, Duke University. Retrieved from https://hdl.handle.net/10161/9968.
47. DiMaio, F., Zhang, J., Chiu, W., & Baker, D. (2013). Cryo-EM model validation using independent map reconstructions. *Protein Science*, *22*(6), 865–868. https://doi.org/10.1002/pro.2267
48. Pintilie, G., Chen, D.-H., Haase-Pettingell, C. A., King, J. A., & Chiu, W. (2016). Resolution and probabilistic models of components in cryoem maps of mature p22 bacteriophage. *Biophysical Journal*, *110*(4), 827–839. https://doi.org/10.1016/j.bpj.2015.11.3522
49. Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., Shapovalov, M. V., Renfrew, P. D., Mulligan, V. K., Kappel, K., Labonte, J. W., Pacella, M. S., Bonneau, R., Bradley, P., Dunbrack, R. L., Das, R., Baker, D., Kuhlman, B., Kortemme, T., & Gray, J. J. (2017). The Rosetta all-atom energy function for macromolecular modeling and Design. *Journal of Chemical Theory and Computation*, *13*(6), 3031–3048. https://doi.org/10.1021/acs.jctc.7b00125
50. Hiranuma, N., Park, H., Baek, M., Anishchenko, I., Dauparas, J., & Baker, D. (2021). Improved protein structure refinement guided by deep learning based accuracy estimation. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-21511-x
51. Mariani, V., Biasini, M., Barbato, A., & Schwede, T. (2013). LDDT: A local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, *29*(21), 2722–2728. https://doi.org/10.1093/bioinformatics/btt473

52. Dunbrack, R. L., & Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Science*, *8*, 1661–1681. https://doi.org/10.1002/pro.5560060807
53. Evans, P., & McCoy, A. (2007). An introduction to molecular replacement. *Acta Crystallographica Section D Biological Crystallography*, *1*, 1–10. https://doi.org/10.1107/s0907444907051554
54. Conway, P., Tyka, M. D., DiMaio, F., Konerding, D. E., & Baker, D. (2013). Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Science*, *1*, 47–55. https://doi.org/10.1002/pro.2389
55. Krebs, W. G., Alexandrov, V., Wilson, C. A., Echols, N., Yu, H., & Gerstein, M. (2002). Normal mode analysis of macromolecular motions in a database framework: Developing mode concentration as a useful classifying statistic. *Proteins: Structure, Function, and Genetics*, *48*(4), 682–695. https://doi.org/10.1002/prot.10168
56. van Heel, M. (2013). Finding trimeric HIV-1 envelope glycoproteins in random noise. *Proceedings of the National Academy of Sciences*, *45*. https://doi.org/10.1073/pnas.1314353110
57. Scheres, S. H., & Chen, S. (2012). Prevention of overfitting in Cryo-EM structure determination. *Nature Methods*, *9*(9), 853–854. https://doi.org/10.1038/nmeth.2115
58. Jin, P., Bulkley, D., Guo, Y., Zhang, W., Guo, Z., Huynh, W., Wu, S., Meltzer, S., Cheng, T., Jan, L. Y., Jan, Y.-N., & Cheng, Y. (2017). Electron cryo-microscopy structure of the mechanotransduction channel NOMPC. *Nature*, *547*(7661), 118–122. https://doi.org/10.1038/nature22981
59. Gerstein, M. (1998). A database of macromolecular motions. *Nucleic Acids Research*, *26*(18), 4280–4290. https://doi.org/10.1093/nar/26.18.4280
60. Xu, H., Li, T., Rohou, A., Arthur, C. P., Tzakoniati, F., Wong, E., Estevez, A., Kugel, C., Franke, Y., Chen, J., Ciferri, C., Hackos, D. H., Koth, C. M., & Payandeh, J. (2019). Structural basis of nav1.7 inhibition by a gating-modifier spider toxin. *Cell*, *176*(4). https://doi.org/10.1016/j.cell.2018.12.018
61. Krishna Kumar, K., Shalev-Benami, M., Robertson, M. J., Hu, H., Banister, S. D., Hollingsworth, S. A., Latorraca, N. R., Kato, H. E., Hilger, D., Maeda, S., Weis, W. I., Farrens, D. L., Dror, R. O., Malhotra, S. V., Kobilka, B. K., & Skiniotis, G. (2019). Structure of a signaling cannabinoid receptor 1-G protein complex. *Cell*, *176*(3). https://doi.org/10.1016/j.cell.2018.11.040
62. Henderson, R., & Moffat, J. K. (1971). The difference fourier technique in protein crystallography: Errors and their treatment. *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry*, *27*(7), 1414–1420. https://doi.org/10.1107/s0567740871004060
63. Wang, R. Y.-R., Song, Y., Barad, B. A., Cheng, Y., Fraser, J. S., & DiMaio, F. (2016). Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *ELife*. https://doi.org/10.7554/elife.17219
64. Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2020). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, *30*(1), 70–82. https://doi.org/10.1002/pro.3943

65. Yang, L., Song, G., & Jernigan, R. L. (2007). How well can I understand large-scale protein motions using normal modes of elastic network models? *Biophysical Journal*, *93*(3), 920–929. https://doi.org/10.1529/biophysj.106.095927
66. Kim, M. K., Jernigan, R. L., & Chirikjian, G. S. (2002). Efficient generation of feasible pathways for protein conformational transitions. *Biophysical Journal*, *83*(3), 1620–1630. https://doi.org/10.1016/s0006-3495(02)73931-3
67. Terwilliger, T. C., Sobolev, O. V., Afonine, P. V., & Adams, P. D. (2018). Automated map sharpening by maximization of detail and connectivity. *Acta Crystallographica Section D Structural Biology*, *74*(6), 545–559. https://doi.org/10.1107/s2059798318004655