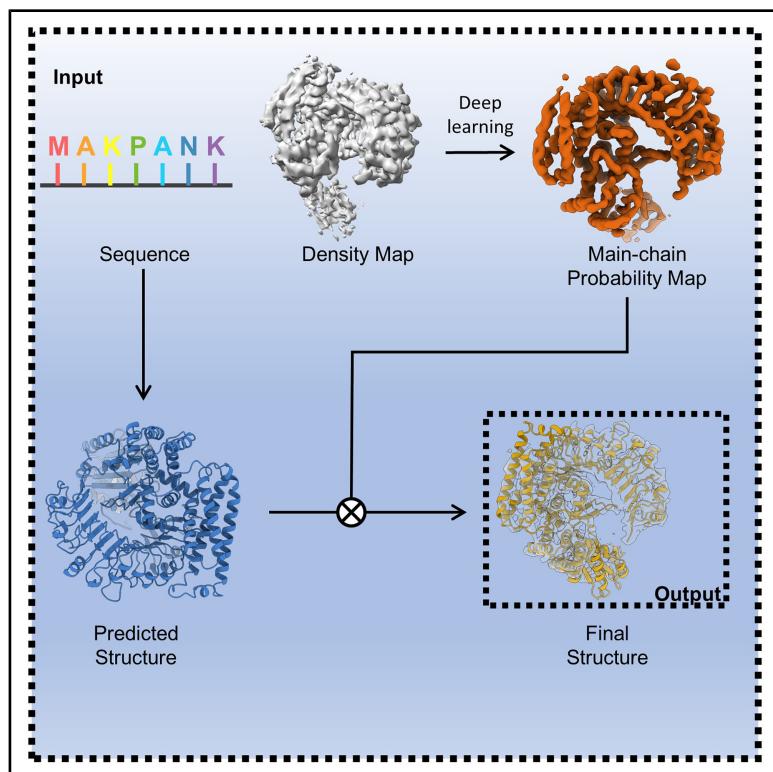


Protein model building for intermediate-resolution cryo-EM maps by integrating evolutionary and experimental information

Graphical abstract



Authors

Ji Chen, Tao Li, Jiahua He,
Sheng-You Huang

Correspondence

jiahua.he@ucsf.edu (J.H.),
huangsy@hust.edu.cn (S.-Y.H.)

In brief

Chen et al. present CryoEvoBuild, an automated method for improved model building from intermediate-resolution EM maps that integrates evolutionary and experimental information. An extensive evaluation demonstrates the superior performance of CryoEvoBuild compared with state-of-the-art methods.

Highlights

- Method for accurate model building from intermediate-resolution cryo-EM maps
- CryoEvoBuild utilizes both experimental and evolutionary information
- CryoEvoBuild uses a novel domain-based flexible model-to-map fitting method
- Improved performance of CryoEvoBuild compared to state-of-the-art methods

Resource

Protein model building for intermediate-resolution cryo-EM maps by integrating evolutionary and experimental information

Ji Chen,¹ Tao Li,¹ Jiahua He,^{1,2,*} and Sheng-You Huang^{1,3,*}

¹School of Physics, Huazhong University of Science and Technology, Wuhan, Hubei 430074, P.R. China

²Present address: Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA, USA

³Lead contact

*Correspondence: jiahua.he@ucsf.edu (J.H.), huangsy@hust.edu.cn (S.-Y.H.)

<https://doi.org/10.1016/j.str.2025.11.004>

SUMMARY

Accurate model building in intermediate-resolution cryo-EM maps normally requires flexible fitting of reliable initial structures. However, while deep learning-based methods such as AlphaFold2 can predict highly accurate structures, the predicted structures often differ from experimental EM maps on both global and local scales, which poses a great challenge to accurate model building in intermediate-resolution EM maps with such initial structures. Addressing the challenge, we propose CryoEvoBuild, an automated method for improved protein model building from intermediate-resolution EM maps through the effective integration of evolutionary and experimental information. CryoEvoBuild implements a novel domain-wise fitting, refinement, assembly, and rebuilding pipeline with a recycling framework guided by AlphaFold2. Extensive benchmarking on a diverse test set of 117 maps at 4.0–10.0 Å resolutions demonstrates that CryoEvoBuild significantly improves the accuracy of AF2-predicted structures and outperforms state-of-the-art approaches, including EMBuild and phenix.dock_and_rebuild.

INTRODUCTION

Cryogenic electron microscopy (cryo-EM), including single-particle analysis (SPA)¹ and cryo-electron tomography (cryo-ET),² has become a cornerstone of structural and cellular biology.³ Recent advancements in microscopy hardware,⁴ sample preparation,⁵ and data processing algorithms^{6,7} have driven a significant increase in the number of three-dimensional (3D) electron density maps deposited in the Electron Microscopy DataBank (EMDB).⁸ Interpreting the density maps and constructing accurate structure models offer irreplaceable insights into understanding the underlying mechanisms that govern cellular processes.

Generally, cryo-EM maps at atomic or near-atomic resolution (<4 Å) allow for highly accurate model building, either through manual building⁹ or computational *de novo* modeling methods.^{10–15} However, because intermediate-resolution maps (4–10 Å) lack sufficient atomic details, it is difficult to build structural models solely based on density data. This limitation is evident in the substantial gap between the number of density maps in the EMDB and the number of corresponding structure models in the Protein DataBank (PDB).¹⁶ As of April 17, 2024, there are 7649 density maps at 4.0–10 Å resolutions in the EMDB, but only about half of them have associated structural models in the PDB. Therefore, there is a pressing need for accurate structural interpretation methods in intermediate-resolution EM density maps.

For maps at intermediate resolutions, prior knowledge about the underlying structure is essential for model building. In most cases, the model building process starts from an initial structure, which can either be a previously determined high-resolution homology structure or a predicted model obtained through homology modeling,^{17,18} fragment threading¹⁹ or deep learning.^{20–23} This initial structure is first aligned to the density map through rigid fitting^{24–29} and then adjusted through flexible fitting^{30–36} in order to obtain an optimized model that conforms to the density map. Within this framework, the feasibility and accuracy of fitting-based modeling depend heavily on the quality of the initial model, particularly when no homologous structure is available.

Recent advances in deep learning have harnessed evolutionary information embedded in multiple sequence alignments or protein language models to predict protein structures directly from sequences. Compared with traditional comparative modeling methods, cutting-edge deep learning-based prediction algorithms such as AlphaFold,^{20,21} RoseTTAFold,²² and ESM-Fold²³ achieve unprecedented accuracy in structure prediction. This breakthrough greatly enhances the potential of fitting-based approaches for structure determination on intermediate-resolution cryo-EM density maps. As a result, model building can now be reframed as a problem of reconciling the inconsistency between the predicted structure representing evolutionary information and the density map representing true experimental information. To address this, we previously

developed EMBuild,³⁷ the first fully automated pipeline, to our knowledge, for modeling protein structures from intermediate-resolution cryo-EM maps. EMBuild iteratively fits, refines, and assembles predicted protein models using a main-chain probability map generated by deep learning, enabling the automated construction of high-quality models.

However, automated computational methods such as EMBuild and related approaches,^{38,39} cannot fully replace the labor-intensive and expert-guided model building in the majority of practical cryo-EM projects. This is because, despite the fact that deep learning-based predictions often yield accurate models for individual protein domains, they frequently fall short of capturing the structure with residue-level accuracy. Limitations arise from several factors, including the presence of intrinsically disordered regions, inadequate utilization of residue covariation information, lack of consideration for the impact from non-protein components and environments, and poorly predicted domain-domain interactions.^{40,41} These issues result in inevitable discrepancies between predicted models and experimental density maps, which can hardly be satisfactorily resolved by existing algorithms in a simple fitting-based manner. Consequently, the resulting final fitted models still require substantial manual correction guided by prior structural knowledge to address errors that are beyond the reach of existing automated methods.

With the continued exponential growth in cryo-EM data—particularly cryo-ET data that often fall within intermediate resolution range, automation is becoming the crux of the matter in alleviating bottlenecks in structure determination and reducing reliance on expert intervention. The key aspect of automation lies in effectively leveraging both evolutionary and experimental information to construct high-quality models that require minimal manual correction and refinement. To tackle this challenge, we present a major advancement beyond EMBuild, CryoEvoBuild, a fully automated method for structure determination from intermediate-resolution EM density maps through the integration of evolutionary and experimental information.

CryoEvoBuild adopts a novel domain-based fitting, refinement, assembling, and rebuilding pipeline to flexibly fit an initial model predicted by AlphaFold2 (AF2)²⁰ into the density map. Crucially, the fitted model is then used as the template to guide AlphaFold2 in predicting an improved initial model in a subsequent round of the modeling pipeline. By combining these steps, the evolutionary information utilized by AlphaFold2 and the experimental information encoded in the density map are closely integrated, yielding a consensus model that is expected to exhibit high fidelity to experimental data while maintaining structural plausibility. Compared with our previous EMBuild method, CryoEvoBuild uses a better domain assignment strategy and thus can address molecular flexibility more efficiently during structure fitting. In addition, CryoEvoBuild also includes a recycling step to refine the initial structure using the built model as a template, which is especially important for those poorly predicted regions in the initial structure. We extensively evaluate CryoEvoBuild on a benchmark set of 117 experimental EM density maps with resolutions in the range of 4–10 Å. The results demonstrate that CryoEvoBuild significantly improves the accuracy of AF2-predicted structures and outperforms state-of-the-art methods, including EMBuild and phenix.dock_and_rebuild.

RESULTS

Overview of the CryoEvoBuild framework

An overview of CryoEvoBuild is illustrated in Figure 1. The inputs of CryoEvoBuild are a cryo-EM map and its corresponding protein sequence. The entire workflow can be organized into four modules, including (a) the main-chain probability prediction module, (b) the protein structure prediction module, (c) the domain fitting and assembling module, and (d) the rebuilding and recycling module. In the main-chain probability prediction module (Figure 1A), a UNet++⁴²-based deep learning module is applied to convert the input density map to a main-chain probability map. In the protein structure prediction module (Figure 1B), AlphaFold2²⁰ is employed to predict an initial model from the input sequence. The predicted model is further segmented into multiple structural domains using SWORD.⁴³ In the domain fitting and assembling module (Figure 1C), individual domains are first fitted into the main-chain probability map, then undergo semi-flexible sub-domain refinement, and are finally assembled using the Bron-Kerbosch maximum clique problem solver. In the rebuilding and recycling module (Figure 1D), phenix.dock_and_rebuild⁴⁰ is applied to rebuild poorly modeled regions. The rebuilt model is used as a template for AlphaFold2 to predict a new initial model in the next round of the modeling pipeline. The recycling process can be repeated several times as needed, yielding the final output model.

Evaluating built models against the Protein DataBank structures

CryoEvoBuild was evaluated on a non-redundant test set of 117 cryo-EM maps at 4–10 Å resolutions. We first assessed the quality of the built model using the PDB structure as the reference. Here, we used two metrics, template modeling score (TM-score) and root-mean-square deviation (RMSD) between the built model and the reference PDB structure calculated using TM-align.⁴⁴ The evaluation results of CryoEvoBuild were compared with those of EMBuild and phenix.dock_and_rebuild. Since all methods used AlphaFold2 to predict initial models from sequences, we also included the evaluation results of initial AlphaFold2-predicted structures for comparison (Table S1).

Figure 2 shows a comparison of TM-scores and RMSDs between CryoEvoBuild and other methods. It can be seen from the figure that CryoEvoBuild substantially improves over the initial AlphaFold2 predictions and outperforms the other methods. Specifically, the models built by CryoEvoBuild achieve an average TM-score of 0.910 and an average RMSD of 1.96 Å, compared with 0.764 and 3.66 Å for AlphaFold2 predictions, 0.849 and 2.45 Å for EMBuild models, and 0.834 and 2.70 Å for phenix.dock_and_rebuild, respectively. Figure 2A illustrates the distribution of TM scores. It can be observed from the figure that while all AlphaFold2 predictions achieve an acceptable accuracy with TM-scores above 0.5, only one of them reached a TM-score above 0.9. This indicates that although AlphaFold2 generally captures the correct global topology, substantial deviations exist in local conformations. In contrast, CryoEvoBuild is able to properly fit and refine the models into the density maps, resulting in a high TM-score of >0.9 for 88 out of the total of 117 models. We also examined the impact of resolutions on CryoEvoBuild. It is shown that the performance is not decreasing with lower resolutions. Specifically,

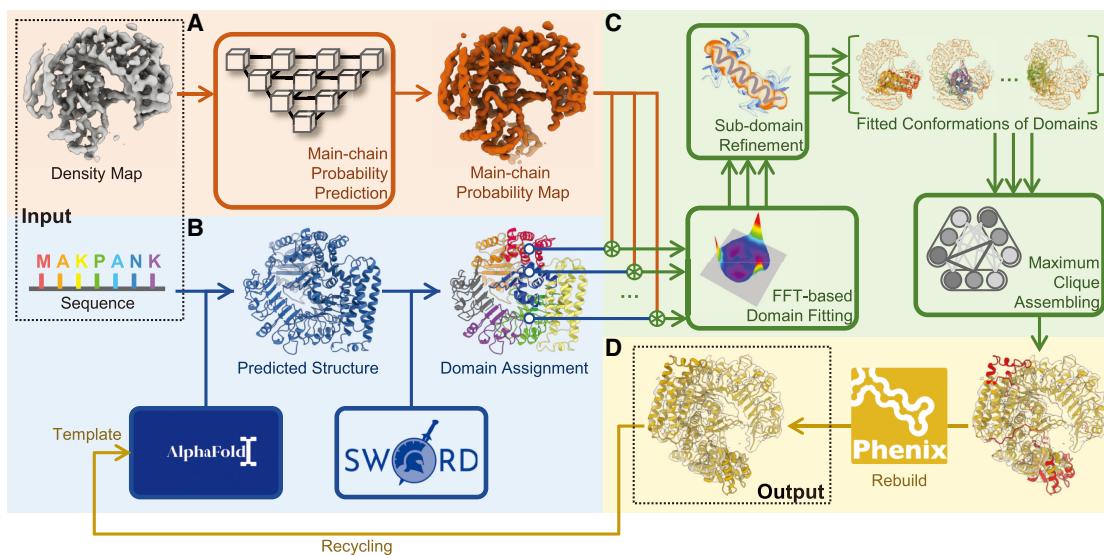


Figure 1. Overview of the CryoEvoBuild framework

Taking the EM density map and the corresponding protein sequence as input, CryoEvoBuild builds the model through four key modules.

(A) Main-chain probability prediction module. A UNet++-based deep learning model converts the density map into a main-chain probability map.

(B) Protein structure prediction module. AlphaFold2 predicts an initial protein model, which is then segmented into structural domains using SWORD.

(C) Domain fitting and assembling module. Domains are fitted into the main-chain probability map, refined through semi-flexible sub-domain refinement, and assembled using a Bron-Kerbosch maximum clique solver.

(D) Rebuilding and recycling module. Poorly modeled regions are refined using phenix.dock_and_rebuild. The resulting model then serves as a template for AlphaFold2 to generate an improved initial model for a second round of modeling.

CryoEvoBuild achieves an average TM-score of 0.912, 0.838, and 0.956 for the cases at 4–6, 6–8, and 8–10 Å resolutions, respectively. These results demonstrate the robustness of CryoEvoBuild against map resolutions.

Figure 2B presents a head-to-head comparison of individual cases for the TM-scores between CryoEvoBuild and other methods, showing that CryoEvoBuild produces better models for most of the test cases. A similar trend is observed in the RMSD values. As shown in Figure 2C, the RMSD distribution of CryoEvoBuild is centered around 2 Å. Out of the total of 117 test cases, CryoEvoBuild yields better RMSD values in 110 cases compared to AlphaFold2, 87 cases compared to EMBuild, and 84 cases compared to phenix.dock_and_rebuild. Figure 3 exhibits several examples of the test cases. These findings suggest that CryoEvoBuild can reliably build atomic models that rival the quality of the PDB models on intermediate-resolution EM density maps, in a fully automated manner.

In addition to TM-score and RMSD, we also assessed the quality of built models using two other metrics, C α coverage and C α score, which are expected to exclude the impact of outliers. Here, the C α coverage is the number of C α atoms in the built model within 3 Å from those in the PDB structure regardless of their residue types divided by that in the PDB structure after superimposing the two models, which was calculated using phenix.chain_comparison.¹⁵ The C α score is defined as the C α coverage divided by RMSD of the covered C α atoms.⁴⁵ It is shown that CryoEvoBuild still achieves the highest average C α coverage of 89.8%, compared with 82.3% for EMBuild, 80.2% for phenix.dock_and_rebuild, and 68.6% for AF2 (Figure 4A). In terms of C α score, CryoEvoBuild also exhibits its superior perfor-

mance and achieves an average C α score of 0.76, which is significantly higher than 0.63 for EMBuild, 0.52 for phenix.dock_and_rebuild, and 0.45 for AF2 (Figure 4B).

Examples of built models

Figure 3A shows an example of the built model for EMD-13020 at 4.12 Å resolution, where the target chain is the DNA polymerase alpha catalytic subunit (PDB ID: 7OPL, chain A). It can be seen from the figure that the AlphaFold2 prediction deviates significantly from the deposited model and the density map, especially in the top region. In spite of the fact that all methods are able to fit the bottom part into the density map, only CryoEvoBuild successfully fits and refines the top part as well. As such, the TM-score and RMSD of the CryoEvoBuild model reach 0.979 and 1.69 Å, respectively, which are drastically improved from 0.746 to 4.81 Å for AlphaFold2 predictions and also significantly higher than the other methods.

The second example is for EMD-18701 at 4.60 Å resolution (Figure 3B). The target chain is the vacuolar protein sorting-associated protein (PDB ID: 8QX8, chain B). Comparison between the AlphaFold2 model and the reference PDB structure reveals that the relative orientations among repeating helices in the α -solenoид are not correctly predicted by AlphaFold2, resulting in a TM-score of 0.842 and an RMSD of 3.89 Å. Although phenix.dock_and_rebuild is able to properly fit the bottom β -propeller into the density map, the top part of the α -solenoид model is not fitted into the density map and completely deviates from the reference PDB structure, probably due to poor local resolution. As a result, the overall RMSD of 4.24 Å for the phenix.dock_and_rebuild model is even worse than that for its initial

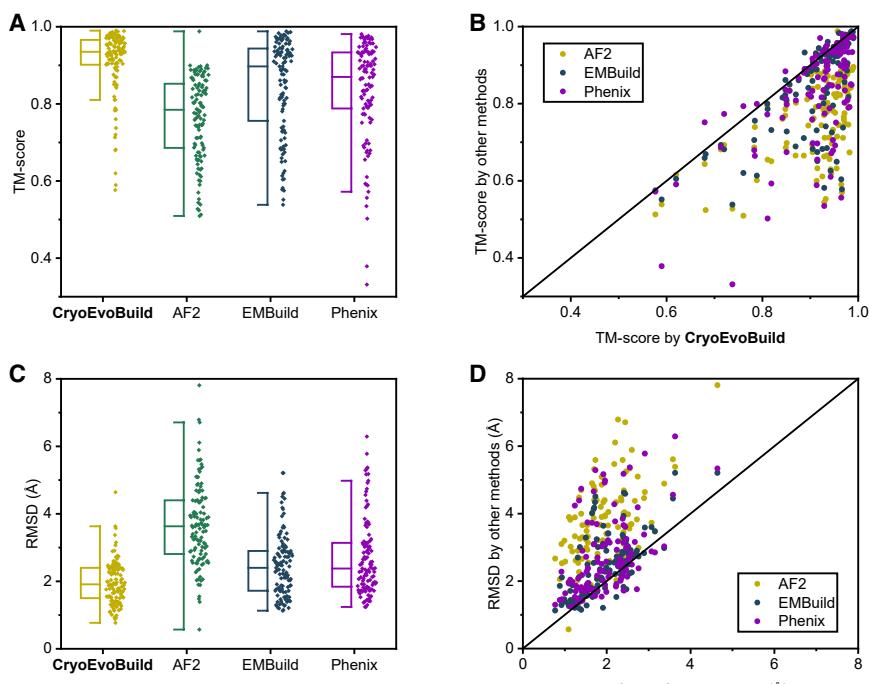


Figure 2. Comparison of TM-score and RMSD against the reference PDB structure on the test set of 117 intermediate resolution maps

(A and C) Boxplots of TM-score (A) and RMSD (C) for the models generated by CryoEvoBuild, AlphaFold2 (AF2), EMBuild, and phenix.dock_and_rebuild (Phenix) on the 117 test cases. The center line of each box indicates the median; the lower and upper hinges correspond to the first and third quartiles, respectively; whiskers extend to 1.5 times the interquartile range from the hinges. Individual data points are plotted to the right of each corresponding box.

(B and D) Comparisons of TM-score (B) and RMSD (D) between CryoEvoBuild and other methods for individual test cases.

AlphaFold2 model. In contrast, EMBuild and CryoEvoBuild are both able to properly fit the α -solenoid into the density. The comparison between CryoEvoBuild and EMBuild also highlights the effectiveness of the rebuilding and recycling module: CryoEvoBuild achieved a TM-score of 0.984 and RMSD of 1.24 Å, compared with 0.978 and 1.46 Å for EMBuild.

The third example shown in Figure 3C presents the P fimbria tip G-adhesin PapG-II (PDB ID: 7LHH, chain G), whose corresponding map is at 7.20 Å resolution (EMD-23340). This structure consists of two domains connected by a linker. Starting from the AlphaFold2 prediction with a TM-score of 0.630 and an RMSD of 4.40 Å, where the two domains are incorrectly oriented, phenix.dock_and_rebuild fails to fit one of the domains into the density map, resulting in a poor-quality model. While EMBuild successfully positions both domains within the density, one domain is completely inverted. This issue highlights a limitation of EMBuild, where the entire protein is fitted into the density as a rigid body, followed by local refinements of domain orientations. When the map resolution is relatively low and the initial pose is highly biased, as shown in this case, EMBuild is less likely to yield an accurate model. In contrast, CryoEvoBuild overcomes this limitation by globally fitting individual domains into the density map and assembling them with appropriate structural restraints, resulting in a high-quality model with a TM-score of 0.942 and an RMSD of 1.7 Å.

Impact of initial models

To examine the influence of initial AF2 predictions, Figure 5A shows a comparison of the performance in different ranges of TM-scores for initial AF2-predicted models. It can be seen from the figure that CryoEvoBuild tends to improve more for the cases with lower-quality AF2 models. Specifically, for initial AF2 models with 0.5–0.7 TM-scores, CryoEvoBuild achieves an improved average TM-score of 0.801, compared with 0.626 for AF2, which

gains an improvement of 28.0%. For the cases with initial TM-scores between 0.7 and 0.8, CryoEvoBuild yields an average TM-score of 0.910, compared with 0.758 for AF2, resulting in an improvement of 20.1%. For the high-quality AF2 cases with 0.8–1.0 TM-scores, CryoEvoBuild gives an average TM-score of 0.939, compared with 0.859 for AF2, leading to an improvement of 9.3%. Similar trends can also be observed in results of RMSDs (Figure 5B). For example, for those cases with poor initial models (i.e., RMSD >4 Å), CryoEvoBuild achieves the greatest improvement and gives an average RMSD of 2.55 Å, compared with 4.99 Å for AF2. These results demonstrate the robustness of CryoEvoBuild against the quality of initial AF2-predicted models.

Despite the overall superior performance of CryoEvoBuild, it may not improve initial AF2 models for some cases. Such cases without improvements may be grouped into two types. One is the cases whose initial AF2 models have a poor quality, where the AF2-predicted models either do not have an accurate secondary structure or have an intrinsically disordered domain. In such cases, CryoEvoBuild normally cannot improve initial AF2 models because of the wrong structure within domains and/or segments. The other is those cases with very high-accuracy AF2 models but a low resolution map, where CryoEvoBuild tends to make models a little worse because the low-quality map would drive the fitting process in the wrong direction. Here are two examples of the two types of cases. Figure S1A is an example of EMD-13186 at 4.3 Å resolution. In this case, CryoEvoBuild does not improve the AF2 model because the initial AF2 model has a wrong structure in the helix domain, which is difficult to correct through fitting. The second example is EMD-12220 at 8.9 Å resolution (Figure S1B). In this case, CryoEvoBuild slightly reduces the quality of the AF2 model because the initial AF2 model already has a very high TM-score of 0.988, but the map has a low resolution of 8.9 Å. In light of these observations, it may be advisable to retain the initial AF2 models in cases where the predicted structures exhibit very high confidence ($pIDDT > 0.9$) while the corresponding cryo-EM maps provide only limited structural detail (resolution poorer than 8 Å).

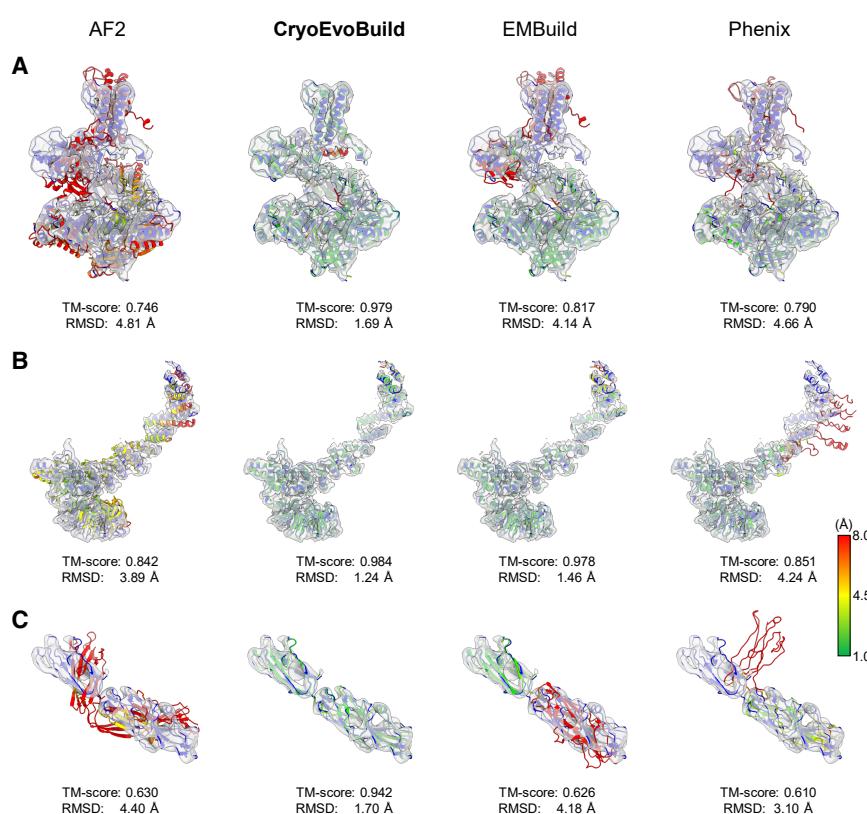


Figure 3. Examples of the models built by EMBuild and other methods from intermediate-resolution cryo-EM maps

The reference PDB structures are shown in blue, and the corresponding EM density maps are displayed as semi-transparent isosurfaces. The models generated by AlphaFold2 (AF2), CryoEvoBuild, EMBuild, and phenix.dock_and_rebuild (Phenix) are colored from green to red based on C α displacements relative to the PDB structure.

(A) PDB ID: 7OPL, chain A (EMD-13020 at 4.12 Å resolution).

(B) PDB ID: 8QX8, chain B (EMD-18701 at 4.6 Å resolution).

(C) PDB ID: 7LHH, chain G (EMD-23340 at 7.2 Å resolution).

Evaluating built models against the density maps

Taking the PDB structures as the reference, we have demonstrated the superior quality for the models built by CryoEvoBuild in terms of TM-score and RMSD. However, such quality metrics may be biased if the reference PDB model is sub-optimally built or contains errors. Moreover, in real applications, we normally do not have such a reference structure. Therefore, it is necessary to evaluate how well the built model coincides with the target density map. Here, we report two real space correlation coefficient (CC) values, CC_box and CC_mask calculated by phenix.map_model_cc⁴⁶ between the built model and the target density map on the test set of 117 intermediate-resolution maps (Table S1). The CC_box is a global metric that compares a density box region around the model, while the CC_mask uses the map values inside a mask calculated around the macromolecule and therefore focuses more on local quality. The models built by CryoEvoBuild are compared with the PDB structures, as well as those built by EMBuild, phenix.dock_and_rebuild, and AF2.

Figure 6A shows the boxplots of CC_box of different methods. It can be seen from the figure that CryoEvoBuild achieves overall higher CC_box values than the other methods. Specifically, CryoEvoBuild achieves a high average CC_box value of 0.829, which is close to 0.859 for the PDB models and is significantly improved from 0.767 for EMBuild models, 0.703 for phenix.dock_and_rebuild models, and 0.603 for AF2 predictions. The head-to-head comparison of CC_box between the PDB models and the models built by CryoEvoBuild is shown in Figure 6B, from which we can see that all the CryoEvoBuild

models have a CC_box value above 0.55. Surprisingly, 23 models that are automatically built by CryoEvoBuild achieve a CC_box value even higher than their corresponding PDB models. Similar improvement can also be observed in CC_mask. As illustrated in Figure 6C, CryoEvoBuild achieves a significantly higher average CC_mask of 0.744 on the entire test set, compared with 0.661 for EMBuild, 0.526 for phenix.dock_and_rebuild, and 0.368 for AF2. Remarkably, the average CC_mask of the models built by CryoEvoBuild even slightly surpasses the average value of 0.739 for the PDB models. Specifically, all CryoEvoBuild models had a CC_mask value of above 0.5, of which more than half (66 out of the total of 117 models) have a CC_mask value higher than their corresponding PDB models (Figure 6D). From both global and local fit-to-map perspectives, these quantitative results demonstrate that CryoEvoBuild is capable of automatically building high-quality models. Furthermore, in a significant number of cases, the models generated by CryoEvoBuild exhibit better consistency with the maps than the PDB models that are typically manually constructed by experts.

Figure 7 shows several examples where the CryoEvoBuild model fits better into the map than the PDB model in terms of fit-to-map metrics. For better visualization, the residues in the built models are colored according to their per-residue CC values reported by phenix.map_model_cc. In the first example of mouse pendrin (PDB ID: 7WLB, chain B), the target density map (EMD-32580) is at 4.1 Å resolution (Figure 7A). It can be clearly seen from the figure that while the PDB model and the CryoEvoBuild model both fit well into the density, which is reflected in their similar CC_box values (0.738 for the CryoEvoBuild model versus 0.736 for the PDB model), the CryoEvoBuild model exhibits significantly better local fitting into the density, yielding a CC_mask of 0.735, compared with only 0.708 for the PDB model. Similar improvement over the PDB model can also be found in the second example of EMD-7453, displayed in Figure 7B. It can be seen from the figure that CryoEvoBuild builds a better model of the AP-1 complex

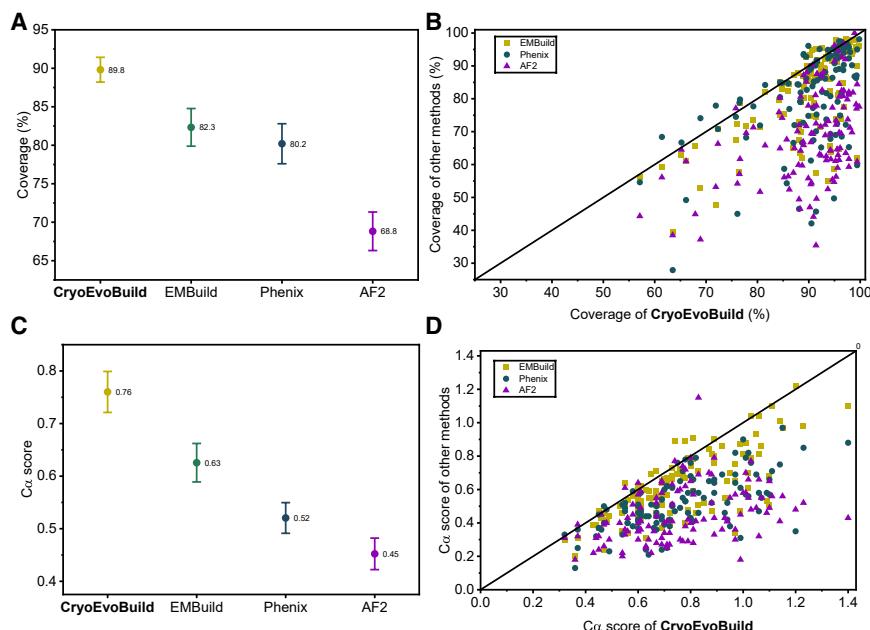


Figure 4. Comparison of residue coverage and C α score against the reference PDB structure on the test set of 117 intermediate resolution maps

(A and C) The C α coverage (A) and C α score (C) for the models generated by CryoEvoBuild, EMBuild, and phenix.dock_and_rebuild (Phenix). The center markers in each interval represent the mean value for the group of categories. The range bars represent the confidence intervals.

(B and D) Comparisons of C α coverage (A) and C α score (C) between CryoEvoBuild and other methods for individual test cases.

predictions are properly informed by experimental data, which in turn facilitates improving the model quality.

To quantitatively assess the impact of these innovations, we performed a detailed ablation study. Specifically, we compared the performance of the baseline CryoEvoBuild model with AlphaFold2, our previously developed EMBuild, and a

variant of CryoEvoBuild without the recycling mechanism on the test set of 117 intermediate-resolution maps (Table S1). Figure 8A shows the comparison of different methods. It can be seen from the figure that, compared with AlphaFold2, EMBuild gains notable improvements. Specifically, the average TM-score increases from 0.764 (AlphaFold2) to 0.849 (EMBuild), while the RMSD is reduced from 3.66 Å to 2.45 Å. On top of that, the CryoEvoBuild variant without recycling further substantially enhances the model quality, leading to an average TM-score of 0.891 and an average RMSD of 2.17 Å. Finally, with the addition of the recycling process, the baseline CryoEvoBuild pipeline reaches the highest structural accuracy, achieving the highest TM-score and RMSD of 0.910 and 1.96 Å, respectively.

Figure 8B illustrates the built model for the heat-labile enterotoxin (PDB ID: 8U5B, chain B), where the resolution of the density map is at 5.3 Å (EMD-41915). It can be seen from the figure that the initial AlphaFold2 model exhibits large local deviations compared with the PDB model. Moreover, the N-terminal tail of the AlphaFold2 model is completely misoriented (highlighted by the dashed box), resulting in a relatively low TM-score of 0.666 and an RMSD of 2.90 Å. It can also be seen from the figure that although the initial model is slightly refined by EMBuild, leading to an improved TM-score of 0.685 and an RMSD of 2.70 Å, a noticeable mismatch is still retained. Similarly, the CryoEvoBuild variant without recycling still cannot correct the errors in the initial structure, giving a TM-score of 0.687 and RMSD of 2.76 Å. After incorporating the recycling process, the CryoEvoBuild pipeline yields a significant improvement and achieves a high TM-score of 0.935 and a low RMSD of 1.31 Å, along with a visibly refined local arrangement of secondary structure units, a corrected N-terminal orientation, and an improved alignment with the density map.

Computational efficiency

We also benchmarked the running times of CryoEvoBuild and other tested methods on the test set of 117 cases. All three

subunit mu-1 (PDB ID: 6D83, chain M) into the 4.27 Å density map, achieving significantly higher CC_box and CC_mask of 0.802 and 0.776, compared with 0.780 and 0.753 for the PDB model, respectively. The third example is shown in Figure 7C, where the target density map is EMD-34896 with a resolution of 4.7 Å. Due to the relatively weaker density signal, the bottom part of the PDB model for the tetratricopeptide repeat protein (PDB ID: 8HMD, chain C) does not fit well into the density map. In contrast, the model built by CryoEvoBuild significantly improves this region. As such, the CC_box and CC_mask are drastically improved from 0.886 to 0.563 for the PDB model to 0.934 and 0.788 for the CryoEvoBuild model, respectively.

Dissecting key improvements in CryoEvoBuild

CryoEvoBuild introduces several key innovations over our previously developed EMBuild method, which can be broadly categorized into two aspects. First, CryoEvoBuild incorporates a newly developed domain-level fitting, assembly, and refinement algorithm. Unlike EMBuild, which only performs semi-flexible refinement, CryoEvoBuild fits and refines each domain independently into the main-chain probability map, followed by reassembling under appropriate structural restraints. This strategy significantly enlarges the domain search space, making the method more resilient to incorrect domain orientations in the initial model. In addition, CryoEvoBuild integrates the phenix.dock_and_rebuild module to reconstruct poorly modeled regions, further enhancing model quality. Collectively, these improvements are expected to enable CryoEvoBuild to achieve more accurate fitting to the density map and greater robustness to errors in initial AlphaFold predictions. Second, CryoEvoBuild leverages both experimental density information and evolutionary constraints from AlphaFold2 through a recycling protocol. Specifically, the model generated by CryoEvoBuild in the first round is used as a template to guide a new AlphaFold2 prediction, thereby producing a refined initial model for subsequent model building. This recycling framework ensures that the AlphaFold2

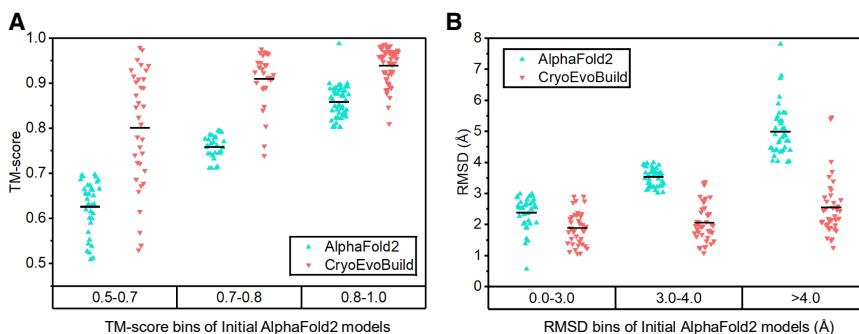


Figure 5. Impact of initial AF2 models on CryoEvoBuild on the test set of 117 intermediate resolution maps

(A) The TM-scores of the models built by CryoEvoBuild for initial AF2 models with different ranges of TM-scores.
(B) The RMSDs of the models built by CryoEvoBuild for initial AF2 models with different ranges of RMSDs. The center black line indicates the average values of each range.

algorithms were initialized from the same AF2-predicted structures. The computations were performed on an Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40GHz. Detailed results are provided in Table S2. On average, CryoEvoBuild requires 111.3 min for model building in a single case, compared with 1.9 min for EMBuild and 384.4 min for phenix.dock_and_rebuild. The longer running times of CryoEvoBuild and phenix.dock_and_rebuild, compared to EMBuild can be attributed to their use of template-based iterative refinement, which involves multiple refinement cycles and substantially increases computational cost. Figure S2 shows the running times of CryoEvoBuild as a function of protein sequence length, demonstrating an approximately linear correlation. This linear scaling can be ascribed to the highly efficient hierarchical fitting procedure implemented in CryoEvoBuild.³⁷

DISCUSSION

We have developed CryoEvoBuild, a powerful model building pipeline for protein structure determination from intermediate-

resolution cryo-EM maps through the integration of evolutionary and experimental information. By combining domain-wise fitting, refinement, assembling, and rebuilding with a recycling framework guided by AlphaFold2 predictions, CryoEvoBuild achieves improved structural accuracy and better agreement with the density map. CryoEvoBuild is extensively evaluated on a test set of 117 maps at 4–10 Å resolutions and compared with state-of-the-art methods, including EMBuild and phenix.dock_and_rebuild. On average, CryoEvoBuild produces highly accurate models with a TM-score of 0.91 and an RMSD of 1.96 Å. Importantly, CryoEvoBuild reliably builds models that fit well to the density map, while other automated approaches often fail. In a substantial number of cases, it even surpasses expert-built PDB models in terms of real-space correlation, highlighting its potential to assist or complement manual modeling. The evaluation results suggest that CryoEvoBuild can produce accurate structural models that only require further minimal correction. Overall, CryoEvoBuild is expected to provide valuable assistance for structure determination from intermediate-resolution

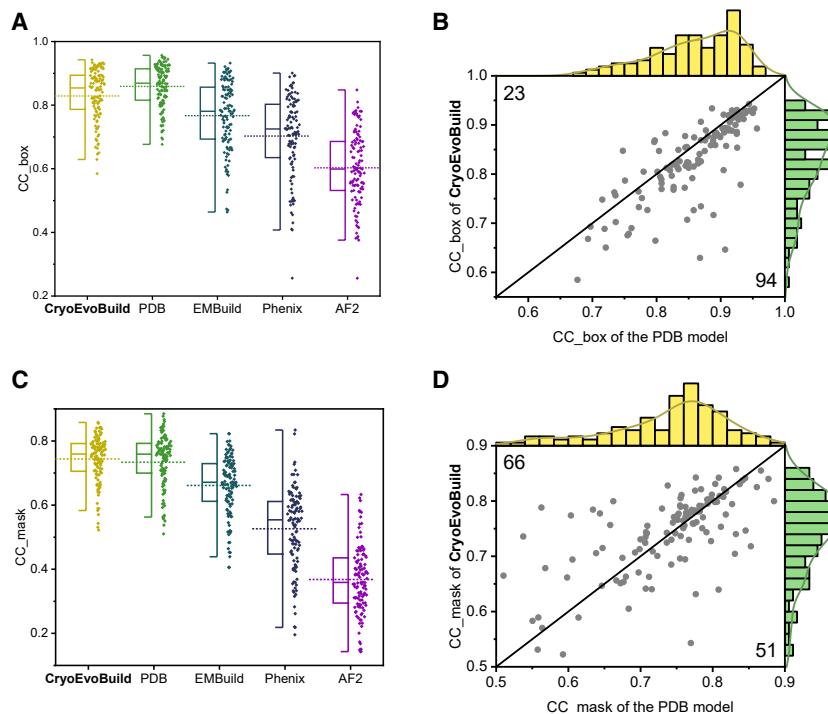


Figure 6. Comparison of CC_box and CC_mask against the EM density map among different model-building methods

(A and C) Boxplots of CC_box (A) and CC_mask (C) for the models generated by CryoEvoBuild, AlphaFold2 (AF2), EMBuild, and phenix.dock_and_rebuild (Phenix) on the 117 test cases. The dashed line indicates the mean; the center line of each box marks the median; the lower and upper hinges correspond to the first and third quartiles, respectively; whiskers extend to 1.5 times the interquartile range from the hinges. Individual data points are plotted to the right of each corresponding box.

(B and D) Head-to-head comparisons of CC_box (B) and CC_mask (D) between PDB models and the CryoEvoBuild models for individual test cases. Distributions are shown along the corresponding axes.

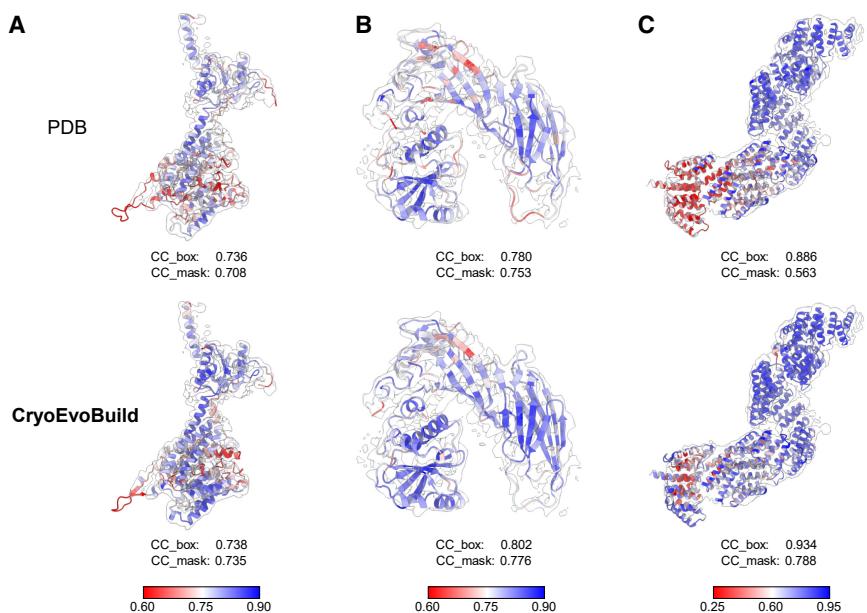


Figure 7. Examples comparing correlation coefficients against the EM density map between the CryoEvoBuild model and the PDB model

Structural models are colored from red to blue according to per-residue real-space correlation coefficients (CCs) with the density map. The corresponding density maps are shown as semi-transparent isosurfaces.

(A) PDB ID: 7WLB, chain B (EMD-32580 at 4.1 Å resolution).

(B) PDB ID: 6D83, chain M (EMD-7453 at 4.27 Å resolution).

(C) PDB ID: 8HMD, chain C (EMD-34896 at 4.7 Å resolution).

cryo-EM maps, and represents a significant step toward a fully automated, high-accuracy model building pipeline.

Although CryoEvoBuild is evaluated on single chains here for validation purposes, it can also build the structures of protein complexes from cryo-EM maps. Figure S3 shows an example of EMD-13020 at 4.12 Å resolution, which is the cryo-EM structure of DNA polymerase alpha-primase bound to SARS CoV nsp1. In this case, EMBuild achieves a TM-score of 0.880 and an RMSD of 3.74 Å in the built model, where the upper-left part does not fit well (Panel a). As a comparison, CryoEvoBuild successfully builds the complete complex structure, and

achieves a TM-score of 0.942 and an RMSD of 2.04 Å (Panel b). In addition, it should be noted that CryoEvoBuild can model the complex with any number of chains, though the running time will significantly increase when the complex being modeled becomes larger. In addition, CryoEvoBuild currently focuses on protein structure modeling, primarily relying on protein structure prediction algorithms, and thus does not support RNA modeling at this time. With the development of RNA structure prediction, RNA support could be a direction for future expansion, particularly by integrating predicted RNA structures and/or available RNA structures into the PDB.

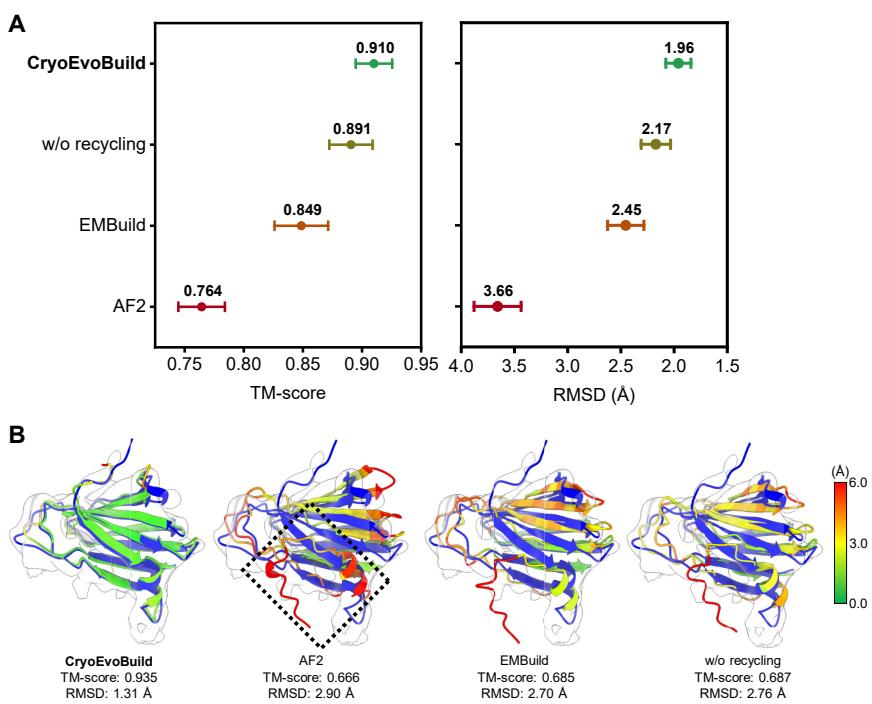


Figure 8. Ablation experiments dissecting key improvements in CryoEvoBuild

(A) TM-score and RMSD of CryoEvoBuild, a CryoEvoBuild variant without applying the recycling step (w/o recycling), EMBuild, and AlphaFold2 (AF2) on the test set of 117 intermediate resolution maps. Points represent means, and error bars indicate 95% bootstrap percentile intervals.

(B) Comparison of the models generated by AlphaFold2, EMBuild, CryoEvoBuild variant without recycling, and CryoEvoBuild for EMD-41915 at 5.3 Å resolution (PDB ID: 8U5B, chain B). Models are color-coded from green to red based on C α displacements relative to the reference PDB structure. The density map is shown as a semi-transparent isosurface. The region highlighted by the dashed box indicates the N-terminal tail of the protein, where the AlphaFold2 model is completely misoriented.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Sheng-You Huang (huangsy@hust.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The raw data of the evaluation results are provided in **Tables S1, S2, S3, and S4**. All published datasets used in this article were taken from the EMDB and PDB (accession codes specified in the figure captions and in **Tables S1, S2, S3, and S4**). The CryoEvoBuild package is freely available for academic or non-commercial users via <http://huanglab.phys.hust.edu.cn/CryoEvoBuild/>. Any additional information required to reanalyze the data reported in this article is available from the **lead contact** upon request.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (grant nos. 32161133002, 32430020, and 62072199) and the startup grant of Huazhong University of Science and Technology.

AUTHOR CONTRIBUTIONS

S.H. conceived the project. S.H. and J.H. supervised the project. J.C., J.H., and S.H. designed and performed the experiments. All authors analyzed the data. J.H., J.C., and S.H. wrote the article. All authors have read and approved the final version of the article.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **METHOD DETAILS**
 - Data collection
 - Main-chain probability prediction
 - Protein structure prediction and domain assignment
 - Domain fitting into the density map
 - Domain assembling
 - Rebuilding poorly modeled regions
 - Structure recycling
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Evaluation metrics

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.str.2025.11.004>.

Received: June 24, 2025

Revised: October 31, 2025

Accepted: November 3, 2025

REFERENCES

1. Chari, A., and Stark, H. (2023). Prospects and limitations of high-resolution single-particle cryo-electron microscopy. *Annu. Rev. Biophys.* 52, 391–411.
2. Young, L.N., and Villa, E. (2023). Bringing structure to cell biology with cryo-electron tomography. *Annu. Rev. Biophys.* 52, 573–595.
3. Nogales, E., and Mahamid, J. (2024). Bridging structural and cell biology with cryo-electron microscopy. *Nature* 628, 47–56.
4. Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* 10, 584–590.
5. Xu, Y., and Dang, S. (2022). Recent technical advances in sample preparation for single-particle cryo-EM. *Front. Mol. Biosci.* 9, 892459.
6. Punjani, A., Zhang, H., and Fleet, D.J. (2020). Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* 17, 1214–1221.
7. Zivanov, J., Otón, J., Ke, Z., von Kügelgen, A., Pyle, E., Qu, K., Morado, D., Castaño-Díez, D., Zanetti, G., Bharat, T.A.M., et al. (2022). A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0. *eLife* 11, e83724.
8. wwPDB Consortium (2024). EMDB—the Electron Microscopy Data Bank. *Nucleic Acids Res.* 52, D456–D465.
9. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501.
10. Jamali, K., Käll, L., Zhang, R., Brown, A., Kimanis, D., and Scheres, S.H.W. (2024). Automated model building and protein identification in cryo-EM maps. *Nature* 628, 450–457.
11. Li, T., He, J., Cao, H., Zhang, Y., Chen, J., Xiao, Y., and Huang, S.Y. (2025). All-atom RNA structure determination from cryo-EM maps. *Nat. Biotechnol.* 43, 97–105.
12. Li, T., Cao, H., He, J., and Huang, S.Y. (2024). Automated detection and *de novo* structure modeling of nucleic acids from cryo-EM maps. *Nat. Commun.* 15, 9367.
13. Wang, X., Terashi, G., and Kihara, D. (2023). CryoREAD: *de novo* structure modeling for nucleic acids in cryo-EM maps using deep learning. *Nat. Methods* 20, 1739–1747.
14. He, J., and Huang, S.Y. (2021). Full-length *de novo* protein structure determination from cryo-EM maps using deep learning. *Bioinformatics* 37, 3480–3490.
15. Terwilliger, T.C., Adams, P.D., Afonine, P.V., and Sobolev, O.V. (2018). A fully automatic method yielding initial models from high-resolution cryo-electron microscopy maps. *Nat. Methods* 15, 905–908.
16. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242.
17. Webb, B., and Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* 54, 5.6.1–5.6.37.
18. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Remppen, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303.
19. Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5, 725–738.
20. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589.
21. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500.
22. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876.

23. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 379, 1123–1130.
24. Petersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
25. Kawabata, T. (2008). Multiple subunit fitting into a low-resolution density map of a macromolecular complex using a gaussian mixture model. *Biophys. J.* 95, 4643–4658.
26. Lasker, K., Sali, A., and Wolfson, H.J. (2010). Determining macromolecular assembly structures by molecular docking and fitting into an electron density map. *Proteins* 78, 3205–3211.
27. C P van Zundert, G., and M J J Bonvin, A. (2015). Fast and sensitive rigid-body fitting into cryo-EM density maps with PowerFit. *AIMS Biophys.* 2, 73–87.
28. Zhang, B., Zhang, W., Pearce, R., Zhang, Y., and Shen, H.B. (2021). Fitting low-resolution protein structures into cryo-EM density maps by multiobjective optimization of global and local correlations. *J. Phys. Chem. B* 125, 528–538.
29. Cragnolini, T., Sahota, H., Joseph, A.P., Sweeney, A., Malhotra, S., Vasishyan, D., and Topf, M. (2021). TEMPy2: a Python library with improved 3D electron microscopy density-fitting and validation workflows. *Acta Crystallogr. D Struct. Biol.* 77, 41–47.
30. Orzechowski, M., and Tama, F. (2008). Flexible fitting of high-resolution x-ray structures into cryoelectron microscopy maps using biased molecular dynamics simulations. *Biophys. J.* 95, 5692–5705.
31. López-Blanco, J.R., and Chacón, P. (2013). iMODFIT: efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *J. Struct. Biol.* 184, 261–270.
32. Kirmizialtin, S., Loerke, J., Behrmann, E., Spahn, C.M.T., and Sanbonmatsu, K.Y. (2015). Using molecular simulation to model high-resolution cryo-EM reconstructions. *Methods Enzymol.* 558, 497–514.
33. Dou, H., Burrows, D.W., Baker, M.L., and Ju, T. (2017). Flexible fitting of atomic models into cryo-EM density maps guided by helix correspondences. *Biophys. J.* 112, 2479–2493.
34. Kovacs, J.A., Galkin, V.E., and Wriggers, W. (2018). Accurate flexible refinement of atomic models against medium-resolution cryo-EM maps using damped dynamics. *BMC Struct. Biol.* 18, 12.
35. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczki, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* 75, 861–877.
36. Zhou, X., Li, Y., Zhang, C., Zheng, W., Zhang, G., and Zhang, Y. (2022). Progressive assembly of multi-domain protein structures from cryo-EM density maps. *Nat. Comput. Sci.* 2, 265–275.
37. He, J., Lin, P., Chen, J., Cao, H., and Huang, S.Y. (2022). Model building of protein complexes from intermediate-resolution cryo-EM maps with deep learning-guided automatic assembly. *Nat. Commun.* 13, 4066.
38. Alnabati, E., Esquivel-Rodriguez, J., Terashi, G., and Kihara, D. (2022). MarkovFit: Structure fitting for protein complexes in electron microscopy maps using Markov random field. *Front. Mol. Biosci.* 9, 935411.
39. Zhang, Z., Cai, Y., Zhang, B., Zheng, W., Freddolino, L., Zhang, G., and Zhou, X. (2024). DEMO-EM2: assembling protein complex structures from cryo-EM maps through intertwined chain and domain fitting. *Brief. Bioinform.* 25, bbae113.
40. Terwilliger, T.C., Poon, B.K., Afonine, P.V., Schlicksup, C.J., Croll, T.I., Millán, C., Richardson, J.S., Read, R.J., and Adams, P.D. (2022). Improved AlphaFold modeling with implicit experimental information. *Nat. Methods* 19, 1376–1382.
41. Terwilliger, T.C., Liebschner, D., Croll, T.I., Williams, C.J., McCoy, A.J., Poon, B.K., Afonine, P.V., Oeffner, R.D., Richardson, J.S., Read, R.J., and Adams, P.D. (2024). AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nat. Methods* 21, 110–116.
42. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., and Liang, J. (2020). UNet++: redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans. Med. Imaging* 39, 1856–1867.
43. Postic, G., Ghouzam, Y., Chebrek, R., and Gelly, J.C. (2017). An ambiguity principle for assigning protein structural domains. *Sci. Adv.* 3, e1600552.
44. Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 33, 2302–2309.
45. Dai, X., Wu, L., Yoo, S., and Liu, Q. (2023). Integrating AlphaFold and deep learning for atomistic interpretation of cryo-EM maps. *Brief. Bioinform.* 24, bbad405.
46. Afonine, P.V., Klaholz, B.P., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C., Adams, P.D., and Urzhumtsev, A. (2018). New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D Struct. Biol.* 74, 814–840.
47. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). PyTorch: an imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* 32, 8024–8035.
48. Andersen, C.A.F., Palmer, A.G., Brunak, S., and Rost, B. (2002). Continuum secondary structure captures protein flexibility. *Structure* 10, 175–184.
49. DiMaio, F., Tyka, M.D., Baker, M.L., Chiu, W., and Baker, D. (2009). Refinement of protein structures into low-resolution density maps using Rosetta. *J. Mol. Biol.* 392, 181–190.
50. Wang, Z., Bovik, A.C., Sheikh, H.R., and Simoncelli, E.P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612.
51. Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A.A., Aflalo, C., and Vakser, I.A. (1992). Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. USA* 89, 2195–2199.
52. Wen, Z., He, J., and Huang, S.Y. (2020). Topology-independent and global protein structure alignment through an FFT-based algorithm. *Bioinformatics* 36, 478–486.
53. Dantzig, G.B. (1951). Maximization of a linear function of variables subject to linear inequalities. *Activity Analysis Product. Allocat* 13, 339–347.
54. Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., and Adams, P.D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.* 74, 531–544.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Electron Microscopy Data Bank (EMDB)	wwPDB ⁸	https://www.ebi.ac.uk/emdb/
Protein Data Bank (PDB)	Berman et al. ¹⁶	https://www.rcsb.org/
Software and algorithms		
Python, version 3.8+	Python Software Foundation	RRID:SCR_008394
PyTorch, version 1.8.1 or later	Paszke et al. ⁴⁷	RRID:SCR_018536
AlphaFold2	Jumper et al. ²⁰	https://github.com/google-deepmind/alphafold
Phenix	Terwilliger et al. ⁴⁰	https://www.phenix-online.org/
CryoEvoBuild	This paper	http://huanglab.phys.hust.edu.cn/CryoEvoBuild/
EMBuild	He et al. ³⁷	http://huanglab.phys.hust.edu.cn/EMBuild/
TM-align, version 20220412	Zhang et al. ⁴⁴	N/A
stride	Andersen et al. ⁴⁸	http://webclu.bio.wzw.tum.de/stride/
SWORD	Postic et al. ⁴³	https://github.com/DSIMB/SWORD2

METHOD DETAILS

Data collection

All EM density maps and their atomic structures in our dataset are downloaded from the EMDB⁸ and the PDB,¹⁶ respectively. To build a test set that covers the density maps of diverse proteins, all EM maps and associated PDB structures that meet the following criteria are collected: (i) resolution within 4.0–10.0 Å; (ii) reconstructed by single-particle analysis (SPA) or subtomogram averaging (STA); (iii) a unique correspondence between structure and density map; (iv) without any unknown atom or residue; (v) more than 95% atoms have valid B-factor and occupancy; (vi) more than 70% of the protein sequence is modeled in the PDB structure; (vii) the total number of residues falls within the range of 50–2000; (viii) the PDB model fits well to the corresponding density map. Each PDB structure is split into individual protein chains. The density map for each chain is segmented from its corresponding density map using a mask calculated within 4.0 Å from heavy atoms of this chain. To ensure criterion (viii), the CC_mask values between PDB models and the corresponding density maps are calculated using phenix.map_model_cc.⁴⁶ The pair of model and map with CC_mask lower than 0.6 is removed from our dataset. Further manual inspection is conducted to exclude the entries that exhibit obvious misfits between map and model. For CryoEvoBuild, we directly use the main-chain probability prediction module trained in EMBuild, where the training set consists of 209 pairs of experimental cryo-EM density maps and associated PDB structures at 4–8 Å resolutions.³⁷ To avoid overlap between the training and test sets, any protein in our test set that has a sequence identity of above 30% to any protein in the training set is excluded. To remove redundancy, we further cluster our test set using a sequence identity cutoff of 30%, which finally results in 117 entries.

Main-chain probability prediction

CryoEvoBuild uses the same main-chain probability prediction module from EMBuild.³⁷ This module adopts a three-dimensional (3D) UNet++-based deep learning architecture, which is trained on a non-redundant dataset of 209 pairs of density maps and PDB models.³⁷ The network is implemented with Pytorch (version 1.8.1).⁴⁷ During training, the network takes each density map as input, and is iteratively updated by minimizing the differences between the output and a main-chain probability map derived from the associated PDB model as its learning target. Specifically, the main-chain probability $p : \mathbb{R}^3 \rightarrow (0, 1]$ at grid point $\mathbf{x} \in \mathbb{R}^3$ is defined as

$$p(\mathbf{x}) = \max_{\mathbf{a} \in A} e^{-k\|\mathbf{x} - \mathbf{a}\|^2}, \quad (\text{Equation 1})$$

where $\mathbf{a} \in A \subset \mathbb{R}^3$ stands for the position vectors of main-chain atoms (N, C, and C α), and factor k is derived from the map resolution R , i.e., $k = (\pi/(2.4+0.8R))^2$.⁴⁹ The difference between the network output and the learning target is measured by a combined loss function consisting of two terms,³⁷ which include smooth L1 loss and structural similarity (SSIM)⁵⁰ loss. By iteratively training the network through back-propagation with the Adam optimizer, the trained network can be used to predict the main-chain probability maps from intermediate-resolution density maps. To obtain a refined representation of the main-chain probability map, mean-shift

algorithm is applied to find a set of main-chain points (MCPs) for each map towards local maxima of main-chain probabilities, by iteratively shifting the seed points $\mathbf{z}_i \in Z \subset \mathbb{R}^3$ ($i = 1, \dots, N; t = 0, 1, \dots$) as

$$\mathbf{z}_i^{t+1} = \frac{\sum_{n=1}^N K(\mathbf{z}_i^t - \mathbf{x}_n)p(\mathbf{x}_n)\mathbf{x}_n}{\sum_{n'=1}^N K(\mathbf{z}_i^t - \mathbf{x}_{n'})p(\mathbf{x}_{n'})}, \quad (\text{Equation 2})$$

where $K : \mathbb{R}^3 \rightarrow (0, 1]$ is the Gaussian kernel function presented as $K(\mathbf{z}_i^t - \mathbf{x}_n) = e^{-k\|\mathbf{z}_i^t - \mathbf{x}_n\|^2}$. The $p(\mathbf{x}_n)$ is the probability to find a main-chain atoms at the grid point \mathbf{x}_n . The probability of a shifted seed point is $q(\mathbf{z}_i) = \frac{1}{N} \sum_{n=1}^N K(\mathbf{z}_i - \mathbf{x}_n)p(\mathbf{x}_n)$. Then, the shifted seed points will be clustered and the highest probability value is selected as the representative of this cluster. These resulted points \mathbf{z}_i ($i=1, \dots, L$) are the main-chain points, where L is the number of main-chain points. Detailed numbers of main-chain points for each map are listed in [Table S3](#).

Protein structure prediction and domain assignment

Starting from the input protein sequence, CryoEvoBuild first utilizes AlphaFold2 (AF2)²⁰ to predict an initial model. It should be noted that during the development of CryoEvoBuild, AlphaFold3 (AF3)²¹ has not been released yet, so we used AF2 to predict the initial protein structures. However, the pipeline of CryoEvoBuild is readily applicable to other structure prediction algorithms such as AF3. To ensure that our evaluation is not biased by homologous structures, we do not provide any template to AF2 when predicting this initial model. Since the predicted model usually does not perfectly conform to the target density map, we introduce a two-stage fitting into the pipeline of CryoEvoBuild.

Specifically, we first use SWORD⁴³ to assign structure domains of the predicted model at two different levels: domains and sub-domains. At the level of domains, the predicted model is segmented into several large domains, which are separately fitted into the density map as rigid bodies, and then re-assembled together to yield the model. Each of the domains is further split into multiple sub-domains, where our previously developed semi-flexible domain-refinement algorithm³⁷ is applied to optimize the relative orientations of these sub-domains according to the density map. To increase diversity, CryoEvoBuild takes multiple domain assignments into account as long as the domain assignment quality metrics reported by SWORD satisfy certain criteria. Specifically, different domain assignments are distinguished by varying minimum domain length thresholds. For each threshold, the assignment with the most domains and a \mathcal{K} value of ≥ 3.0 , as reported by SWORD, is retained. In this study, the minimum domain lengths are set to [30, 60, 90, 150]. In all cases, each domain must have a SWORD-assigned quality rating of at least three stars. An exception is made for particularly short proteins (fewer than 200 residues), where the minimum quality threshold is relaxed to two stars to allow for more flexible domain assignments. For sub-domains, we simply select the assignment that contains the most sub-domains with a minimum length of 30 residues and a SWORD-assigned quality rating of at least one star.

Domain fitting into the density map

The fitting of each domain into the target density map is conducted using a fast Fourier transform (FFT)-based exhaustive search over all six degrees of freedom,^{51,52} where all domains are searched in order in one round of fitting. CryoEvoBuild simplifies the direct matching between the full-atom model and the density map to the matching between main-chain atoms and the predicted MCPs. For intermediate-resolution maps, the predicted main-chain atom locations often contain some uncertainties, so we use the MCPs instead of the whole map to reduce the error during fitting. Main-chain points are merged at the local highest probability point to enhance the accuracy of main-chain atom locations and reduces the cost of calculation. We first map the main-chain atoms of the predicted model $\mathbf{y}_i \in Y$ and the predicted MCPs $\mathbf{z}_i \in Z$ onto two 3D grids of size $N \times N \times N$ according to [Equation 1](#), where the grid spacing α is empirically set to 1.5 Å. The grid-based match score τ for a superimposition between main-chain atoms and MCPs when the main-chain atom grid \mathcal{Y} is shifted relative to the fixed MCP grid Z by respectively $i, j, k = 1-N, \dots, 0, \dots, N-1$ grid points in three translational dimensions is simply defined as

$$\tau(i, j, k) = -\lambda \sum_{l=1}^N \sum_{m=1}^N \sum_{n=1}^N \mathcal{Z}_{l,m,n} \times \mathcal{Y}_{l+i, m+j, n+k}, \quad (\text{Equation 3})$$

where λ is defined as $\lambda = (k/\pi)^{1.5}$.⁴⁹ The match scores for all possible translations of main-chain atoms, i.e., $\{\mathbf{t} = (-\alpha i, -\alpha j, -\alpha k)\}_{i,j,k=1-N}^{N-1}$ can be computed in a single round of convolution in Fourier space. The exhaustive search in rotational degrees of freedom is performed by exploring 4392 evenly distributed orientations at 15° intervals in the Euler space. That is, for each of the 4392 orientations, the main-chain atoms are first rotated by the rotation matrix $\mathbf{R} \in SO(3)$, and then the FFT-based translational search is conducted for this orientation.

After the exhaustive grid-based search, the top-scored poses $\phi = (\mathbf{R}, \mathbf{t}) \in SO(3) \times \mathbb{R}^3$ are further optimized by the SIMPLEX⁵³ method, a derivative-free optimization technique that iteratively refine the poses guided by a scoring function as $\underset{\phi}{\operatorname{argmin}} s(\phi, Y, Z)$

measuring the fitting between main-chain atoms and MCPs. Here, we term such a scoring function as the main-chain match score, which is defined as

$$s(\phi, Y, Z) = -\lambda \sum_{\mathbf{y} \in Y} \max_{\mathbf{z} \in Z} q(\mathbf{z}) e^{-k\|\mathbf{R}\mathbf{y} + \mathbf{t} - \mathbf{z}\|^2}. \quad (\text{Equation 4})$$

The top 50 best poses are input to semi-flexible refinement at the level of sub-domains. Specifically, a graph is first built based on the structural connectivity of the sub-domains. Starting from one of the sub-domains, each sub-domain is separately refined one by one in the order of breadth-first search (BFS) on the graph. Namely, starting from a selected node, the algorithm explores the neighboring nodes at the current depth before moving on to the nodes at the next depth level. The graph is built according to the connectivity of sub-domains. Starting from a seed domain, we sequentially optimize its adjacent sub-domains. After all neighboring sub-domains have been optimized using the SIMPLEX method,⁵³ the recently refined sub-domain is treated as the new seed domain. This procedure is repeated until all sub-domains are optimized. By changing the starting sub-domain, each pose will result in multiple conformations. All sub-domains are iteratively selected as seed domains. For each domain, we keep 50 top-scored conformations for subsequent assembling.

Domain assembling

After independent fitting and refinement of all domains of a protein into the density map, the next step is to assemble them back together. To maintain structural rationality during assembling, instead of selecting the highest-scored conformation of each domain and putting them together, two additional issues must be addressed. On one hand, atomic clashes should be prevented. On the other hand, structural connectivity of domains must be retained. For the atomic clash issue, we define the clash score of one main-chain atom $\mathbf{y}_{i,l_i} \in Y_{i,l_i}$ in the l_i -th conformation of the i -th domain relative to the l_j -th conformation of the j -th domain ($i \neq j$) with main-chain atoms Y_{j,l_j} as

$$c(\mathbf{y}_{i,l_i}, Y_{j,l_j}) = \max_{\mathbf{y}_{i,l_i} \in Y_{i,l_i}} e^{-k \left(\min \left\{ \|\mathbf{y}_{i,l_i} - \mathbf{y}_{j,l_j}\| - d_{\text{clash},0.0} \right\} \right)^2}, \quad (\text{Equation 5})$$

where d_{clash} is the cutoff distance below which the clash score reaches its maximum value of 1. The clash score of domain Y_{i,l_i} relative to domain Y_{j,l_j} is the average clash score of all main-chain atoms as follows,

$$\bar{c}(Y_{i,l_i}, Y_{j,l_j}) = \frac{1}{|Y_{i,l_i}|} \sum_{\mathbf{y}_{i,l_i} \in Y_{i,l_i}} c(\mathbf{y}_{i,l_i}, Y_{j,l_j}) \quad (\text{Equation 6})$$

For the structural connectivity issue, we define the restraint score between any pair of structurally connecting domains as

$$r(Y_{i,l_i}, Y_{j,l_j}) = \begin{cases} 1.0, & d(Y_{i,l_i}, Y_{j,l_j}) \geq d_{\text{upper}}, \\ \frac{d(Y_{i,l_i}, Y_{j,l_j}) - d_{\text{lower}}}{d_{\text{upper}} - d_{\text{lower}}}, & d_{\text{upper}} > d(Y_{i,l_i}, Y_{j,l_j}) \geq d_{\text{lower}}, \\ 0.0, & d_{\text{lower}} > d(Y_{i,l_i}, Y_{j,l_j}), \end{cases} \quad (\text{Equation 7})$$

where $d(Y_{i,l_i}, Y_{j,l_j})$ is the distance between the C α atoms of two connecting residues. Here, the lower bound of the distance cutoff, d_{lower} is set to 6 Å, which is approximately twice the range of average distance between two neighboring C α atoms. Within this range, we can consider two adjacent domains connected. The upper bound d_{upper} is set to 12 Å, which corresponds to about the 95% confidence interval of the overall domain-domain distance distribution (Figure S4), indicating the distance thresholds at which the restraint is fully satisfied or vanished, respectively.

According to the match scores of domains, and the clash scores and restraint scores among different domains, the problem of assembling domains into a structurally rational protein can be converted to the maximum clique problem. Assuming N ($N=50$ in this study) conformations are kept for each of the M domains in the target protein, first, an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{A}, \mathcal{E})$, $\mathcal{V} \in \mathbb{R}^{N \times M, 1}$, $\mathcal{A} \in \{0, 1\}^{N \times M, N \times M}$, $\mathcal{E} \in \mathbb{R}^{N_{\text{edge}}, 1}$ is built, where each vertex indicates one fitted conformation of a domain. The values of vertices are the corresponding match scores. Conformations from different domains are connected by edges, while conformations of the same domain are not connected, resulting in $N_{\text{edge}}=N^2M(M-1)/2$ edges. The value of edges are defined as follows,

$$e(Y_{i,l_i}, Y_{j,l_j}) = 1 - ((1 - \bar{c}(Y_{i,l_i}, Y_{j,l_j})) \times (1 - r(Y_{i,l_i}, Y_{j,l_j}))), \quad (\text{Equation 8})$$

which takes values in the range of [0,1], where a value close to 1 indicates the two domains can be assembled together without causing any structural irrationality. To prune the graph, edges with values below a given threshold $e_{\text{threshold}}$ of 0.95 are removed from \mathcal{G} . The goal of domain assembling becomes finding the best-scored complete subgraph (i.e., clique) of \mathcal{G} . Since different conformations of one domain are not connected in \mathcal{G} , the resulted clique contains at most one conformation for each domain. The scoring function is the sum of main-chain match scores of containing domains weighted by the edge values, i.e.,

$$s_{\text{total}} = \sum_{\substack{i,j \\ i \neq j}} e(Y_{i,l_i}, Y_{j,l_j}) \times (s(Y_{i,l_i}) + s(Y_{j,l_j})) + \theta \sum_{\substack{i,j \\ A_{ij}^{\text{struct}} = 1}} (d(Y_{i,l_i}, Y_{j,l_j}) - d_{C_\alpha - C_\alpha})^2, \quad (\text{Equation 9})$$

where the latter term represents a spring constraint that ensures the distance between the C_α atoms of two adjacent residues, belonging to structurally connecting domains, remains close to the standard value $d_{C\alpha-C\alpha}$ of 3.8 Å. $\mathcal{A}^{\text{struct}} \in \{0, 1\}^{MM}$ is the structural adjacency matrix, where $\mathcal{A}_{ij}^{\text{struct}} = 1$ indicates that the i -th domain and the j -th domain are structurally connecting to each other, and vice versa. The spring constraint term is weighted by $\theta=0.02$, which corresponds approximately to the maximum s_{total} of spring constant distribution (Figure S5). The Bron-Kerbosch algorithm, which is an effective recursive backtracking maximum clique problem solver, is applied to find those top-scored combinations of domain conformations, i.e.,

$$\underset{\{(l_i)\}_{i=1}^M, \{(l_j)\}_{j=1}^M}{\operatorname{argmin}} (s_{\text{total}}) \quad (\text{Equation 10})$$

In case some domains cannot be assembled into the protein with only one round of Bron-Kerbosch algorithm, an iterative strategy is adopted by CryoEvoBuild to gradually assemble all the domains into the protein. Namely, after each round of assembling, map regions with fitted models are blocked out from subsequent rounds by reducing the probability values of corresponding MCPs, based on their clash scores (Equation 5) relative to the current assembled model D by the following formula:

$$q'(\mathbf{z}_i, D) = q(\mathbf{z}_i) \times (1 - c(\mathbf{z}_i, D)). \quad (\text{Equation 11})$$

The remaining domains are iteratively assembled to the current model by fitting and refining in accordance with the updated MCPs. The iterative assembling procedure is carried out at most 4 rounds or until all the domains are assembled, as illustrated in Figure S6. Among the models generated from different domain assignments, the one with the highest s_{total} is selected for further processing in the pipeline.

Rebuilding poorly modeled regions

After assembling the domains into the target protein, CryoEvoBuild applies further refinements to poorly modeled regions within the constructed model. To determine which regions require refinement, two quantitative metrics are assessed at residue level. The first metric, the pLDDT score reported by AF2, measures the confidence level of each residue in the predicted initial model, where a higher value indicates higher expected accuracy. The second metric is the main-chain match score of each residue, which measures the structural fit to the density map. The formula to calculate main-chain match score for each main-chain atom in the model, i.e., $\mathbf{y} \in Y \subset \mathbb{R}^3$, is derived from Equation 4, as

$$s(\mathbf{y}, Z) = -\lambda \max_{\mathbf{z} \in Z} q(\mathbf{z}) e^{-k||\mathbf{y} - \mathbf{z}||^2}. \quad (\text{Equation 12})$$

The main-chain match score for each residue is computed as the average score of three main-chain atoms. We further smooth the main-chain match scores for each residue within each domain, then within each secondary structure fragment assigned by STRIDE⁴⁸ and finally at residual level using a sliding window.³⁷ Letting $\{s_i\}_{i=1}^L$ denote the set of main-chain match scores for a protein with L residues, \bar{s} denote the average main-chain match score, i.e., $\bar{s} = \frac{1}{L} \sum_{i=1}^L s_i$, we can calculate the relative main-chain match score for each residue as $s'_i = s_i / \bar{s}$. A higher relative main-chain match score indicates that the residue better conforms to the map. Combining these metrics together, residues with $\text{pLDDT} < 50$ or $(\text{pLDDT} \times s'_i) < 70$ are marked to be refined. The refinement of these poorly modeled regions is performed using phenix.dock_and_rebuild,⁴⁰ which combines the functions of processing, docking and rebuilding predicted models into a cryo-EM map. In this workflow, since CryoEvoBuild has already fitted the predicted model into the density map, only the rebuilding function of phenix.dock_and_rebuild is utilized. The assembled model is provided as the ‘morphed_model_file’, which allows low-quality/confidence regions to be rebuilt with regard to the density map while preserving the overall structural framework.

Structure recycling

By fitting and refining the predicted AF2 model into the density map, the resulting model is expected to achieve the best possible agreement with the density map. However, this optimal fit is still constrained by the quality of the initial AF2 prediction. To overcome this limitation, CryoEvoBuild employs a recycling strategy to further improve the initial prediction, and consequently, the final model. Specifically, after the first round of fitting and refinement, the resulting model is used as a template for AF2 to generate an updated prediction. The underlying idea is that the experimental information encoded in the fitted model can be leveraged by AF2 to produce a refined prediction that better conforms to the density map while maintaining structural plausibility. This updated model is then used as the starting point for a second round of fitting and refinement. Finally, the refined model is further optimized at the atomic level with respect to the density map using phenix.real_space_refine.⁵⁴

QUANTIFICATION AND STATISTICAL ANALYSIS

Evaluation metrics

Two types of metrics are used to evaluate the quality of models built by CryoEvoBuild on the test set of 117 cases. The first category assesses the structural similarity between the built model and the reference PDB structure, including the template modeling score (TM-score) and Root Mean Square Deviation (RMSD), which are calculated using the TM-align program.⁴⁴ The second category

consists of fit-to-map metrics, which measure the agreement between the built model and the density map. Specifically, `phenix.map_model_cc46` is used to compute two real-space correlation coefficient (CC) values: CC_box and CC_mask. Based on these metrics, CryoEvoBuild is quantitatively compared with EMBuild and `phenix.dock_and_rebuild` on the entire test set. For each test case, the initial model is generated by AF2 and is used by all tested methods. The CC values for AF2 are calculated through rigid fitting of the AF2-predicted structure to the map. Default parameters are applied to both EMBuild and `phenix.dock_and_rebuild` during the evaluations.

Supplemental Information

Protein model building

**for intermediate-resolution cryo-EM maps by integrating
evolutionary and experimental information**

Ji Chen, Tao Li, Jiahua He, and Sheng-You Huang

Accurate protein model building for intermediate-resolution cryo-EM maps by integrating evolutionary and experimental information

Ji Chen¹, Tao Li¹, Jiahua He^{1,2,*} and Sheng-You Huang^{1,3,*}

¹School of Physics, Huazhong University of Science and Technology, Wuhan, Hubei 430074, P. R. China

²Current address: Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, California, USA

³Lead contact

*e-mail: jiahua.he@ucsf.edu (J.H.), huangsy@hust.edu.cn (S.-Y.H.)

Supplementary Figures

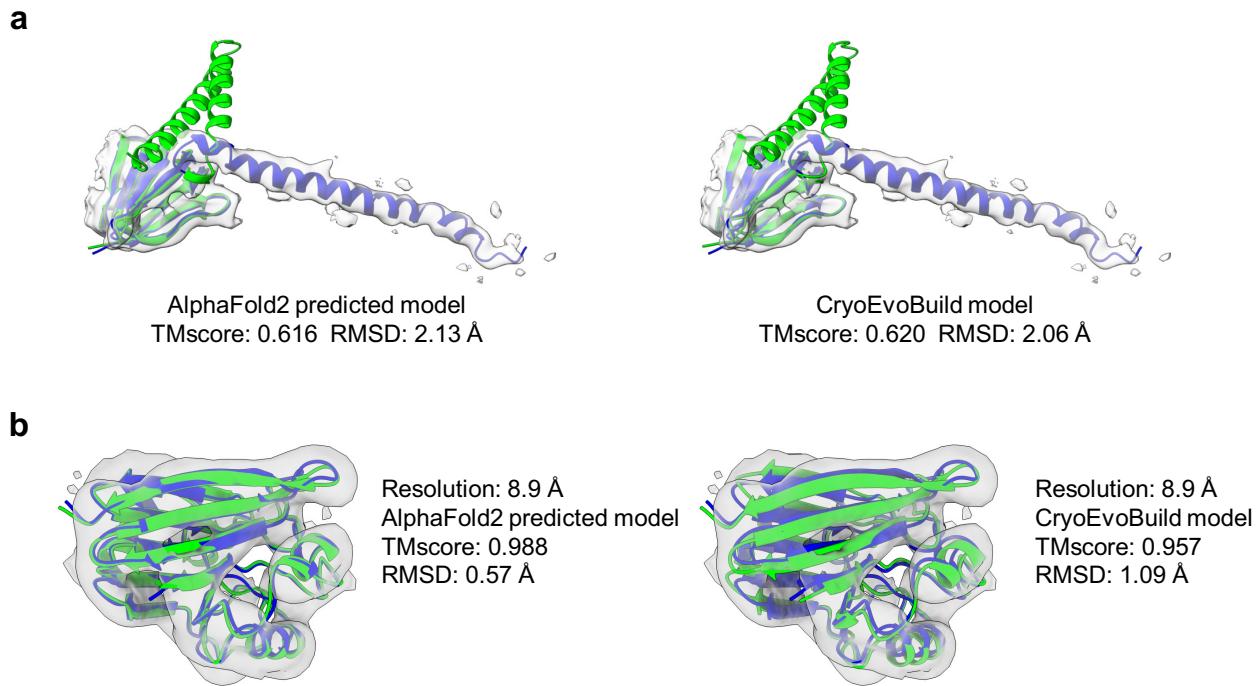


Figure S1: Examples of two edge cases built by CryoEvoBuild from intermediate-resolution cryo-EM maps, related to Figure 2. The reference PDB structures are colored in blue, and the corresponding EM density maps are displayed as semi-transparent isosurfaces. The structures of AF2 models and CryoEvoBuild models are colored in green. **a**, PDB ID: 7P3W, chain w (EMD-13186 at 4.3 Å resolution). **b**, PDB ID: 7BLN, chain D (EMD-12220 at 8.9 Å resolution)

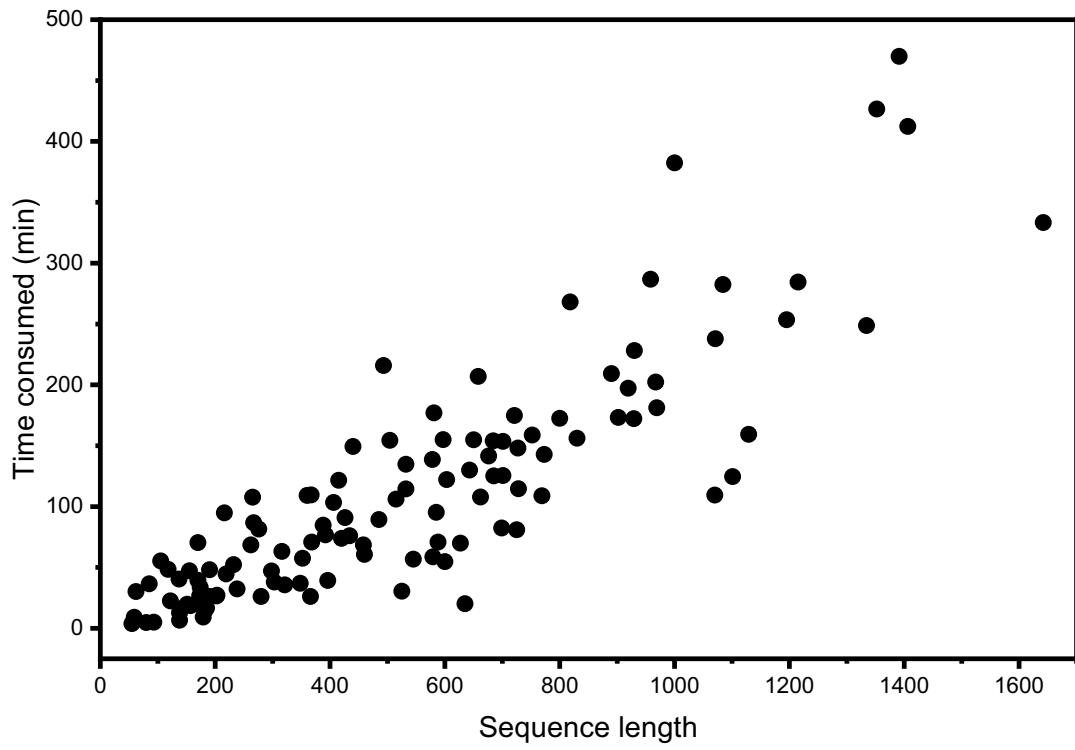


Figure S2: Running times of CryoEvoBuild on the test set of 117 intermediate-resolution maps, related to Figure 2. The running time is plotted as a function of protein sequence length.

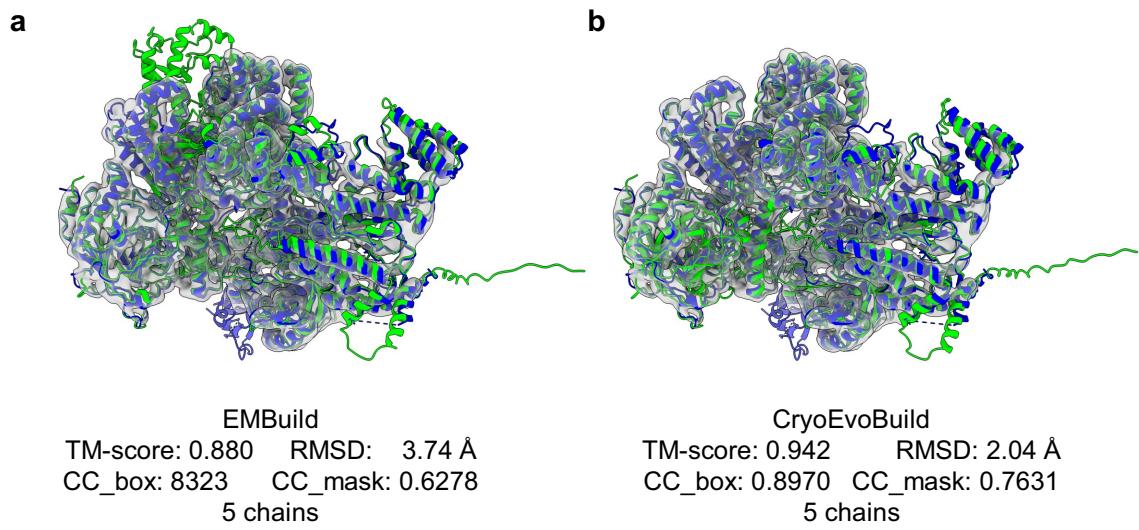


Figure S3: An example of protein complex structure built by EMBuild and CryoEvoBuild, related to STAR Methods. The reference PDB structure is colored in blue, and the built model is colored in green. The target is EMD-13020 at 4.12 Å resolution (PDB ID: 7OPL).

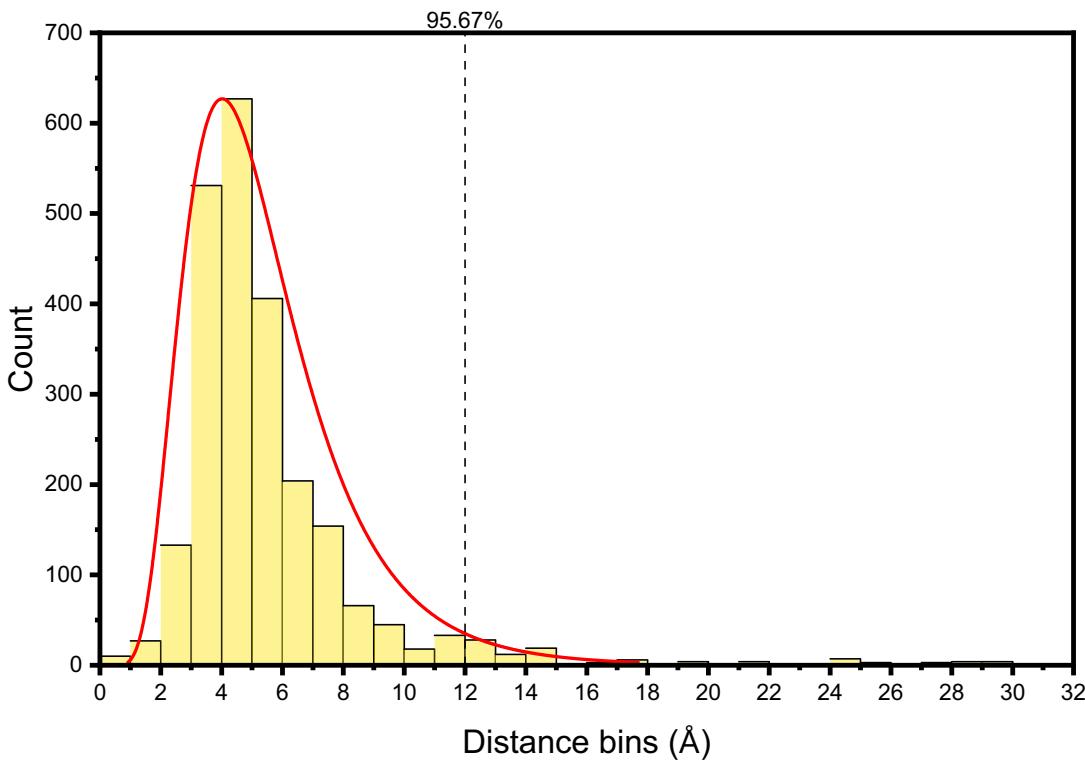


Figure S4: Domain-domain distance distribution of 117 intermediate resolution maps, related to STAR Methods. The red line represents the log-normal fit of the data. The dashed line indicates the $d_{\text{upper}} = 12 \text{ \AA}$, which corresponds to approximately 95% confidence interval of the overall distribution.

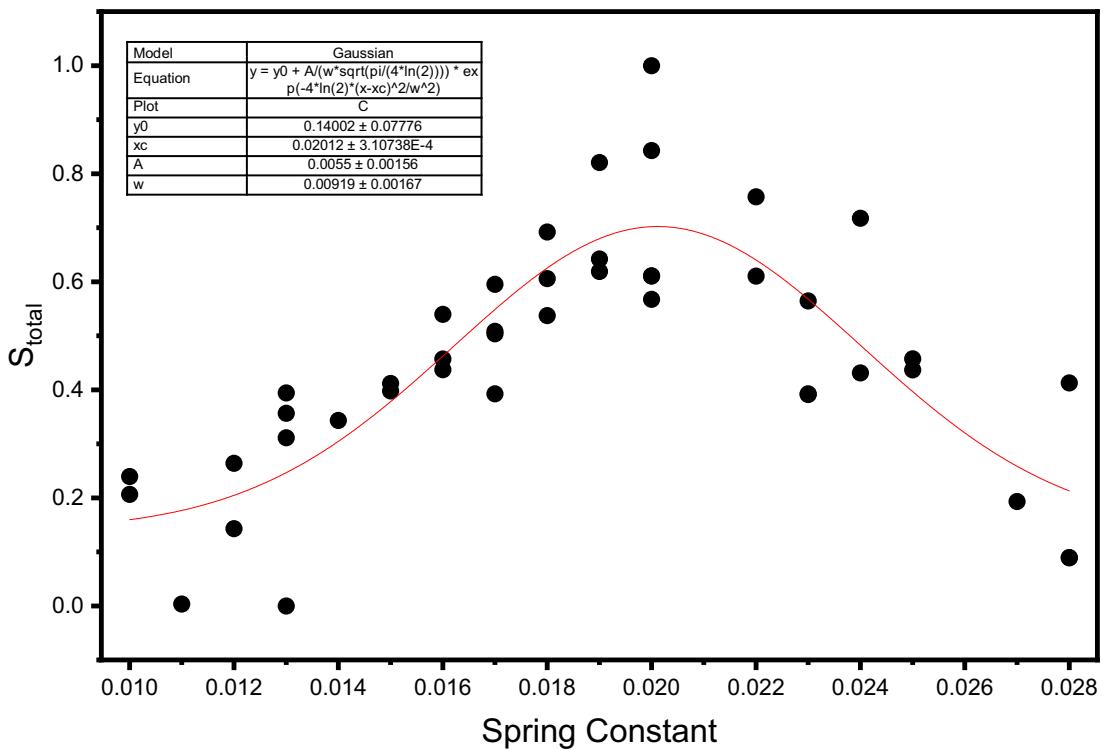


Figure S5: The s_{total} scores versus spring constants θ fitting on an additional data set of 43 intermediate resolution maps filtered with a CC_{mask} of 0.7, related to STAR Methods. The scores were normalized to the range of [0, 1]. Details of 43 intermediate-resolution maps are provided in Supplementary Table 4.

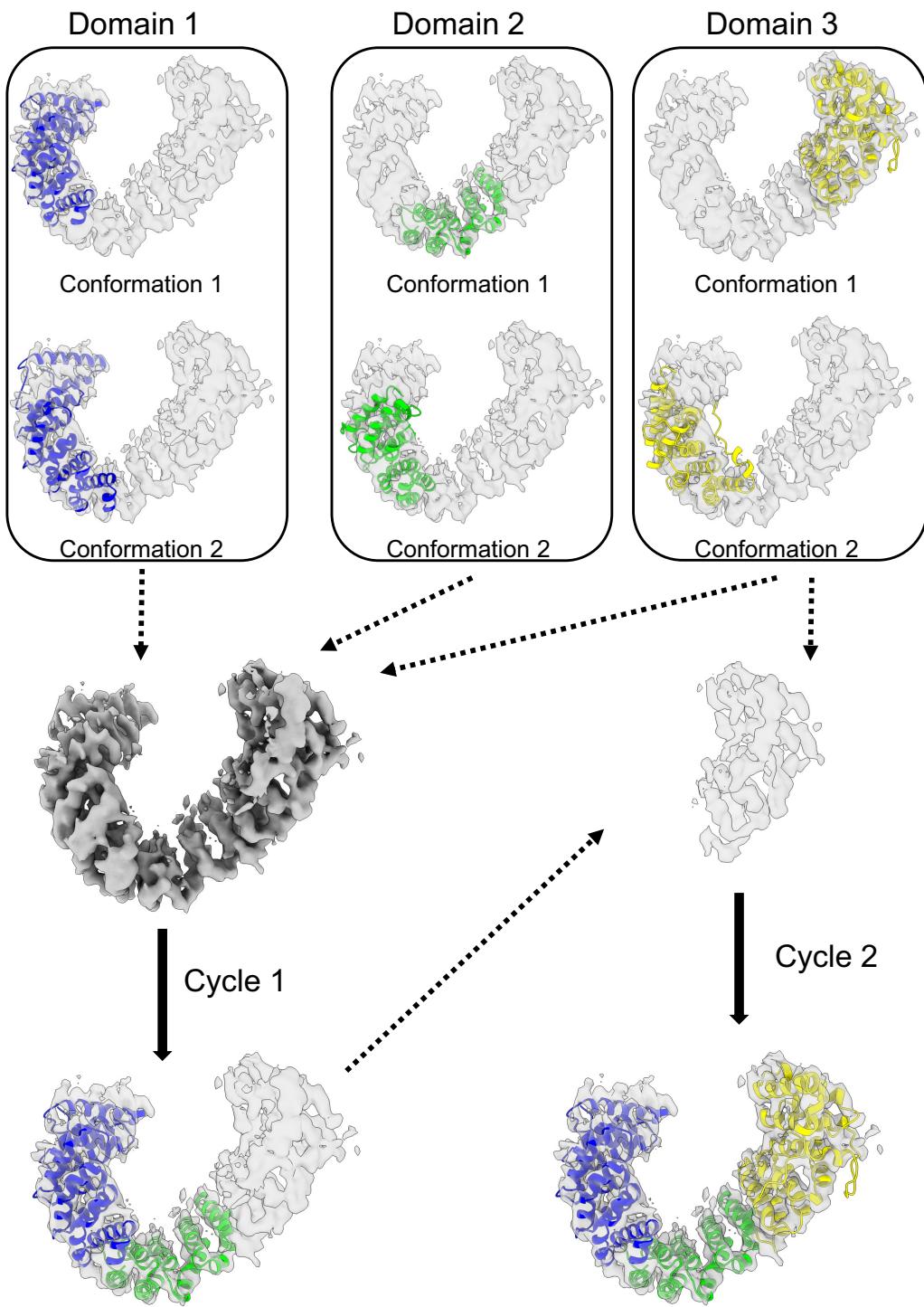


Figure S6: Schematic illustration of the domain assembling procedure, related to STAR Methods. The entire structure is divided into three domains (blue, green, and yellow), each has multiple alternative conformations fitted to the density map. After multiple rounds of fitting operations, the complete structure is assembled into the cryo-EM map.