JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

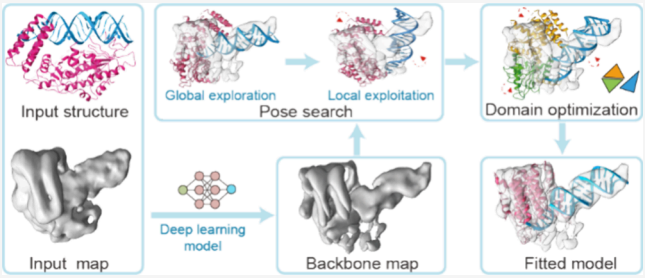
Fitting Atomic Structures into Cryo-EM Maps by Coupling Deep Learning-Enhanced Map Processing with Global-Local Optimization

Yaxian Cai,[†] Ziying Zhang,[†] Xiangyu Xu, Liang Xu, Yu Chen, Guijun Zhang,[*] and Xiaogen Zhou[*,†]

Cite This: *J. Chem. Inf. Model.* 2025, 65, 3800−3811 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: With the breakthroughs in protein structure prediction technology, constructing atomic structures from cryo-electron microscopy (cryo-EM) density maps through structural fitting has become increasingly critical. However, the accuracy of the constructed models heavily relies on the precision of the structure-to-map fitting. In this study, we introduce DEMO-EMfit, a progressive method that integrates deep learning-based backbone map extraction with a global-local structural pose search to fit atomic structures into density maps. DEMO-EMfit was extensively evaluated on a benchmark data set comprising both cryo-electron tomography (cryo-ET) and cryo-EM maps of protein and nucleic acid complexes. The results demonstrate that DEMO-EMfit outperforms state-of-the-art approaches, offering an efficient and accurate tool for fitting atomic structures into density maps.



INTRODUCTION

Proteins are essential biological macromolecules in cells responsible for a wide array of functions. Understanding their three-dimensional structure is crucial for revealing their functions and mechanisms of interaction.[1−3] Cryo-EM is a pivotal tool for elucidating structures of biological macromolecules based on the principles of electron microscopy.[4,5] In recent years, cryo-EM technology has made significant progress in instrumentation, data collection, and image reconstruction, resulting in continuous improvements in the quantity and quality of cryo-EM density maps.[6−10] However, constructing atomic structures of biological macromolecules from density maps is even more critical for further functional studies.[11−13]

Common atomic structure modeling methods often involve fitting homologous structures into density maps, particularly when dealing with maps at resolutions >3Å.[14−16] With recent breakthroughs in structure prediction technologies, numerous approaches have emerged to constructing structures through fitting models predicted by computational methods, such as AlphaFold2[17] and RoseTTAFold,[2] into density maps. Representative fitting-based modeling methods, including DiffModeler,[18] EMbuild,[19] and DEMO-EM2,[20] typically start from predicted structures and employ optimization algorithms to search for the position of each structural unit within the density map. These methods overcome the limitations of deep learning-based de novo approaches, such as DeepTracer,[21] ModelAngelo,[22] and Cryo2Struct,[23] which often suffer from low sequence recovery rates and are only applicable to density maps with resolutions <4Å.[7,24] However, the quality of the final models produced by these methods is highly dependent on the accuracy of the structure-to-map fitting algorithms.

A variety of tools have already been developed to fit atomic structures into density maps, including PowerFit,[25] gmfit,[26] MultiFit,[27] Situs,[28] VESPER,[29] and Phenix.[30] These methods typically use optimization techniques to determine the best alignment between the structural model and density maps while keeping the model rigid. Although these approaches have been employed to construct numerous structures, flexible fitting remains crucial for improving the accuracy of the final model, especially in cases where computationally predicted models have correct domain structures but inaccurate domain orientations. In addition, the successful development of an automated and precise structure fitting method for density maps at both high and low resolutions would significantly advance model construction.

We previously developed a method based on the limited-memory Broyden−Fletcher−Goldfarb−Shanno (L-BFGS) algorithm for fitting protein structures into density maps in DEMO-EM2.[20,31] Since L-BFGS is a local optimization method, the quality of its results depends heavily on the initial solution.[32] To address this, we performed multiple L-BFGS simulations concurrently, using diverse initial poses

Received: January 2, 2025
Revised: February 20, 2025
Accepted: March 13, 2025
Published: March 28, 2025

https://doi.org/10.1021/acs.jcim.5c00004
J. Chem. Inf. Model. 2025, 65, 3800−3811

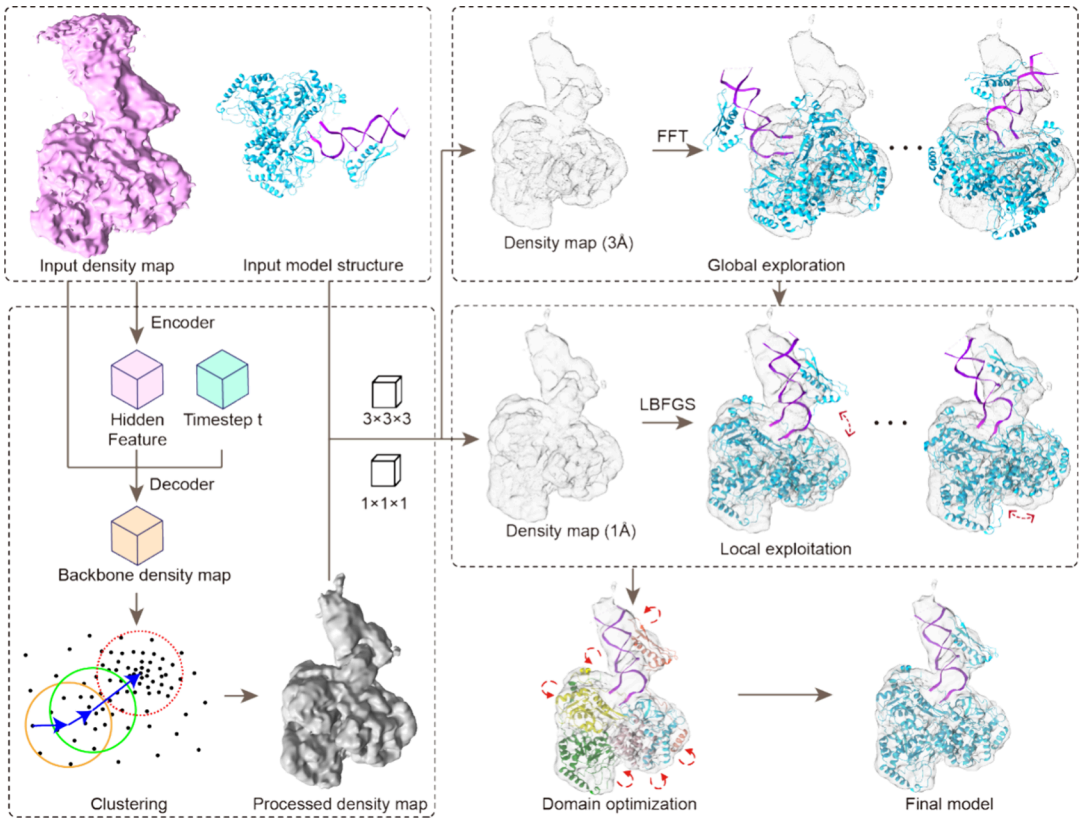


Figure 1. Flowchart depicting the fitting process of a complex structure (PDB ID 8PJJ). Initially, a backbone density map is generated using a diffusion deep learning model combined with the mean-shift algorithm. Next, raw poses of the input structure are produced through an FFT-based global search. Subsequently, the L-BFGS algorithm is applied for a local search on the raw poses to achieve improved fitting results. Finally, protein domain-level optimization is performed to refine the fitted model.

(rotation-translation parameters) determined by the radius of gyration of the structure and the density map. Experimental results demonstrated that this method achieves high-quality fittings and produces accurate protein complex structures through assembly. However, for extremely large density maps, the method can be computationally intensive, requiring significant resources and extended runtimes due to the need for numerous L-BFGS simulations.

This paper introduces DEMO-EMfit, a method for fitting atomic structures into cryo-EM density maps by combining deep learning-enhanced map preprocessing with global-local optimization. First, the density maps are preprocessed using a deep learning approach along with clustering to extract key structural features. Fast Fourier transform (FFT) and L-BFGS are then utilized for global and local searches to determine the optimal pose of the structure. Finally, domain-level optimization is applied to refine the fitted protein model. We evaluate the performance of DEMO-EMfit on a test set consisting of cryo-EM and cryo-ET maps of protein complexes and protein-nucleic acid (NA) complexes, demonstrating that our fitted models outperform those produced by state-of-the-art methods.

MATERIALS AND METHODS

Benchmark Set. To evaluate the performance of DEMO-EMfit, we use the density map data set of DEMO-EM2. This data set includes 43 density maps from the EMDB (http://www.ebi.ac.uk/emdb/), where 27 were generated through subtomogram averaging and 16 were obtained via single-particle reconstruction techniques. All density maps are sharpened and exhibit nonredundant properties, with resolutions ranging from 3 to 10 Å and evenly distributed across different resolution intervals. We test DEMO-EMfit on the data set using experimental structures and AlphaFold3-predicted structures. Since AlphaFold3 cannot predict protein structures with sequences exceeding 5000 residues, the analysis of fitting with precited structures utilizes 39 density maps (Tables S1 and S2), comprising 21 α-proteins, 9 β-proteins, and 9 mixed-type proteins.

Pipeline of DEMO-EMfit. DEMO-EMfit is a method for fitting atomic structures into density maps, as illustrated in Figure 1. Starting with an atomic structure and the corresponding density map, a deep learning diffusion approach is first applied to extract backbone atom information, followed by a clustering step to generate a backbone atom density map. Next, raw poses are determined through an FFT-based global search using an interpolated map with enlarged voxel sizes. If a promising fit is not found through the global search, then an L-BFGS-based local search is initiated, using the poses from the global search as initial positions. During this stage, the density correlation between the structure and map is used to evaluate the pose. Finally, protein domain-level optimization is conducted to refine the pose with the highest correlation score, resulting in a finely fitted structure.

Deep Learning-Enhanced Map Processing. DEMO-EMfit primarily focuses on the correlation between the density map and backbone atoms of the structure during the fitting process. To minimize the impact of nonbackbone atom information and noise, we employ the state-of-the-art deep

https://doi.org/10.1021/acs.jcim.5c00004
J. Chem. Inf. Model. 2025, 65, 3800−3811

learning method, DiffModeler, to generate a new density map that exclusively contains backbone atom information. DiffModeler transforms the task of extracting backbone atom information on density maps from a semantic segmentation problem into a categorical discrete data generation problem by combining the denoising diffusion implicit model with a U-net network.[18] Compared with other deep learning methods, DiffModeler achieves higher accuracy in extracting backbone atom information.

In DiffModeler, a mean-shift algorithm is used to refine the backbone density map generated by extracting maps featuring local representative density points from the predicted backbone density map. First, grid points with a density value >0 in the backbone density map are selected. Then, the coordinates of each grid point x_i are iteratively updated based on its neighboring grid points within 2Å. The update coordinate of x_i is

$$y_i = \frac{\sum_{n=1}^{N} H(x_i - x_n)\varphi(x_n)x_n}{\sum_{n'=1}^{N} H(x_i - x_{n'})\varphi(x_{n'})} \tag{1}$$

where N represents the number of neighboring grid points in the density map, $\varphi(x_n)$ represents the density value at grid point x_n, and H is a Gaussian kernel function.

In DiffModeler, H is a traditional Gaussian kernel function. In order to improve the accuracy of the backbone density map, we modified the Gaussian kernel function with

$$H(x) = \frac{1}{(2\pi)^{1.5}\sigma^3}\exp\left(-\frac{\left\|y_i - x_i\right\|^2}{2\sigma^2}\right) \tag{2}$$

where σ represents the bandwidth. This new kernel function, by incorporation of σ^3, can more accurately capture the characteristics of the data. Moreover, it provides better adaptability to local variations in data distribution by adjusting the normalization through a combination of the bandwidth and a constant.

Global Exploration for Initial Raw Poses. To obtain the initial fitting structure, we performed a global search using the FFT algorithm. A density map with a resolution of 3 Å, interpolated from the original backbone density map, is utilized to enhance computational efficiency and expedite the initial search process. We first convert the protein structure into a Fourier space representation and transform the density map into the frequency domain using FFT. Then, we perform an exhaustive search of all possible poses of the input structure in Fourier space based on the density correlation coefficient (CC) between the protein structure and the density map. Here, the CC score for each pose is defined as follows:

$$\mathrm{CC} = \frac{\sum_{i=1}^{N}(\rho_E(v_i) - \overline{\rho_E})(\rho_M(v_i) - \overline{\rho_M})}{\sqrt{\sum_{i=1}^{N}(\rho_E(v_i) - \overline{\rho_E})^2 \sum_{i=1}^{N}(\rho_M(v_i) - \overline{\rho_M})^2}} \tag{3}$$

where N is the number of voxels with density value larger than the set threshold, $\rho_E(v_i)$is the density value of the ith voxel in the density map, $\overline{\rho_E}$ represents the average density value of the density map, $\rho_M(v_i)$ is the density value of the ith voxel of the density map generated according to the protein structure, $\overline{\rho_M}$ is the average density value of the density map generated according to the protein structure, and $\rho_M(v_i)$ is calculated as follows:

$$\rho_M(v_i) = \sum_{j=1}^{L} m\sqrt[3]{\left(\frac{\pi}{(2.4 + 0.8R)^2}\right)^2}\exp\left(-\left(\frac{\pi}{2.4 + 0.8R}\right)^2 |v_i - x_j|^2\right) \tag{4}$$

where x_j is the Ca coordinate of the jth residue in the structure, m represents the mass of the atom, R represents the resolution of the density map, and v_i represents the position of the ith voxel. The parameter of $\frac{\pi}{(2.4 + 0.8R)^2}$ is chosen based on recommendations from previously published work to optimize the correlation between the single-Gaussian approximation and the all-atom Gaussian density of alanine.[33] For each pose, the CC is calculated and sorted in ascending order. If the maximum CC score is less than 0.6, then the top 20 poses are selected for the next stage. This method takes full advantage of the efficient computational properties of Fourier space, significantly improving the speed and accuracy of the global search.

Local Exploitation for Accurate Poses. In order to obtain better poses for the fitting, L-BFGS-based local exploitation is further performed for the poses generated in the global search. This step uses a density map with a voxel size of 1 Å interpolated from the original backbone density map. Multiple L-BFGS simulations simultaneously start from 20 nonredundant initial poses obtained in the global search phase.

Specifically, the starting poses are chosen by excluding similar ones, identified based on the distance criterion R_{min} = max(0.85R_m, 5Å), where R_m represents the model's radius of gyration. Two poses are considered similar if their distance is less than R_{min}. Each selected pose is independently subjected to an L-BFGS simulation. The CC score is used to evaluate the pose in the simulation, with a scale factor of 5000 applied to enhance the sensitivity. The L-BFGS simulation terminates upon convergence or when the number of steps reaches the maximum limit of 1000. The pose with the highest density correlation score is selected and compared to the top-scoring pose from the global search phase, and the pose with a higher score is used to generate the fitting model for the next step.

Domain-Based Structure Optimization. Given that protein structures may exhibit domain-level biases (e.g., Figure S1), we introduce a domain-level refinement strategy to enhance the accuracy of protein structure fitting. First, the FUpred[34] tool is used to identify domains within the structure. The differential evolution (DE)[35−37] algorithm is then applied to simultaneously adjust the positions and orientations of all domains while treating each domain as a rigid body. Specifically, the DE algorithm is guided by a comprehensive interdomain energy function comprising three components: the CC score between the entire structure and the density map, steric clashes between domains, and the connectivity between adjacent domains. The population size for DE is set to 50, with the crossover rate and scaling factor configured as recommended in the reference.[37] The algorithm terminates after 200 generations, and the model with the highest CC score in the last generation is selected as the final fitted structure.

RESULTS AND DISCUSSION

Evaluation on Experimental Structures. To validate the performance of DEMO-EMfit and rule out the negative impact from incorrect structures, we first test it on the benchmark set

https://doi.org/10.1021/acs.jcim.5c00004
J. Chem. Inf. Model. 2025, 65, 3800−3811

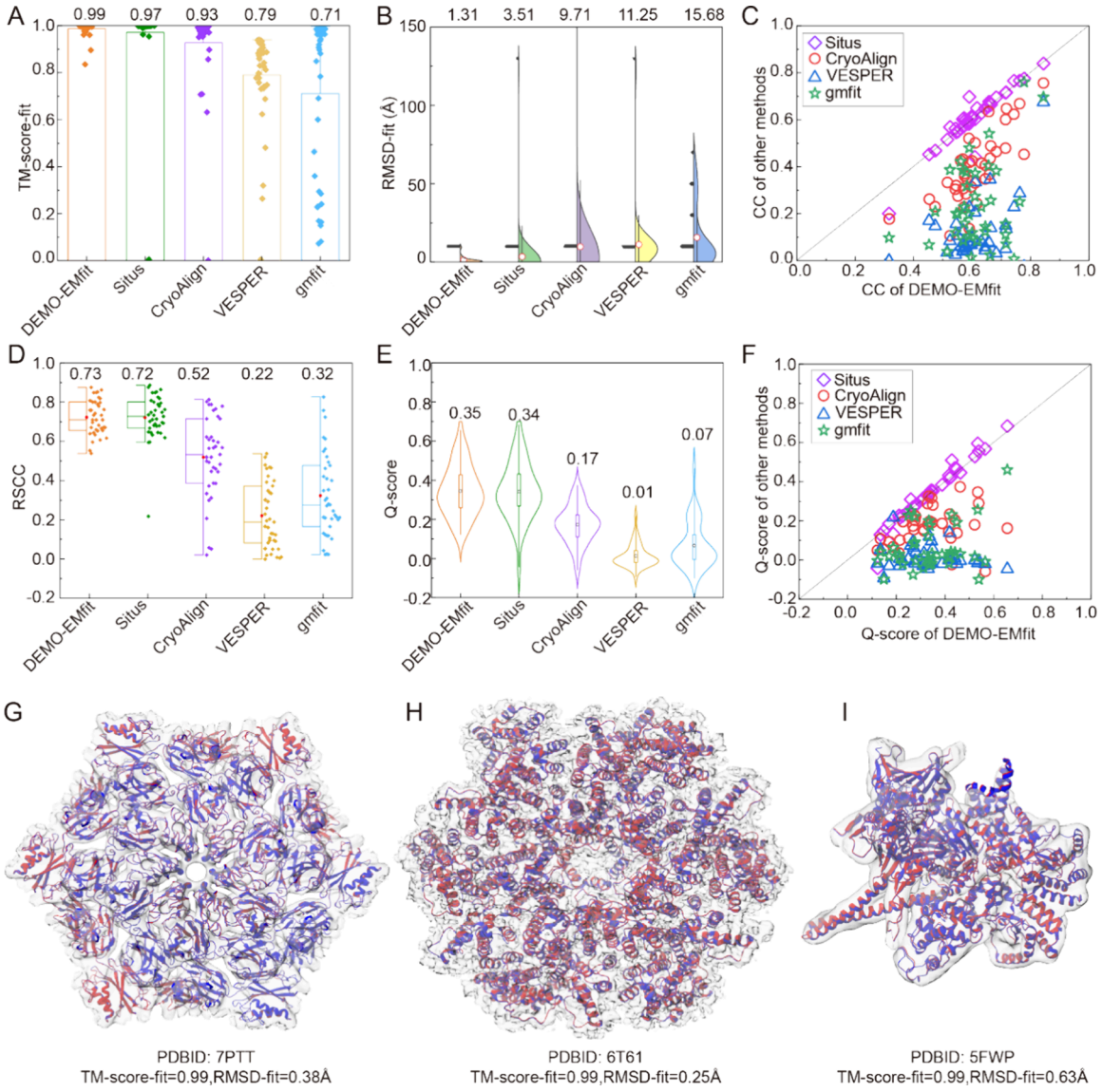


Figure 2. Results of fitting using experimental structures. (A) Box plots of the TM-score-fit for the structures fitted by different methods. The top bar of each box represents the mean value, whiskers indicate the range of outliers (1.5 times the interquartile range), and data points represent targets. (B) Half-violin plots of RMSD-fit of the fitted structures by different methods. The shape of the violin represents the distribution, whiskers show the range within one standard deviation (SD), black dots represent individual data points, and red circles indicate mean values. (C) Comparison of CC values for protein structures fitted using DEMO-EMfit and other methods. (D) Box plots of RSCC values of the models fitted by different methods. The box represents the interquartile range (25−75% of the data), the central line indicates the median, the whiskers represent the range of 1.5 times the interquartile range for outliers, differently colored dots denote data points, and red circles indicate mean values. (E) Violin plots of *Q*-scores for models fitted by different methods. The shape of the violin indicates the distribution, the box represents the interquartile range (25−75% of the data), the whiskers extend to the 1st and 99th percentiles, and black circles denote the mean values. (F) Comparison of *Q*-scores between DEMO-EMfit and other methods for each protein structure. (G−I) Comparisons of the fitted experimental structures (red) and deposited structures (blue) against the semitransparent gray density maps for the protein with PDB IDs of 7PTT, 6T61, and 5FWP.

using experimental structures. All structures are subjected to random rotations and translations before fitting. We compare DEMO-EMfit with four state-of-the-art methods: Situs, CryoAlign, VESPER, and gmfit. The key algorithmic differences between DEMO-EMfit and these methods are described in the Supporting Information, Table S3. All settings for these control methods are kept consistent with their original implementations. To assess the accuracy of the fitting, we employ multiple metrics, including TM-score-fit, RMSD-fit, CC, and *Q*-score.[30] Here, TM-score-fit and RMSD-fit represent the TM-score[38,39] and RMSD calculated by US-align[40] without structural superposition. Both TM-score/RMSD and TM-score-fit/RMSD-fit measure the similarity between the target and deposited model. However, TM-score and RMSD aligns the target model to the deposited (native) model using the Kabsch algorithm,[41] whereas TM-score-fit and RMSD-fit replace the superposition step with a fitting between the target structure and the density map. Therefore, the TM-score-fit and RMSD-fit can be used to assess both the accuracy of structure-to-map fitting and the quality of the resulting model. A TM-score-fit/RMSD-fit that is close to the TM-score/RMSD indicates a correct fitting, while a higher TM-

https://doi.org/10.1021/acs.jcim.5c00004
J. Chem. Inf. Model. 2025, 65, 3800−3811

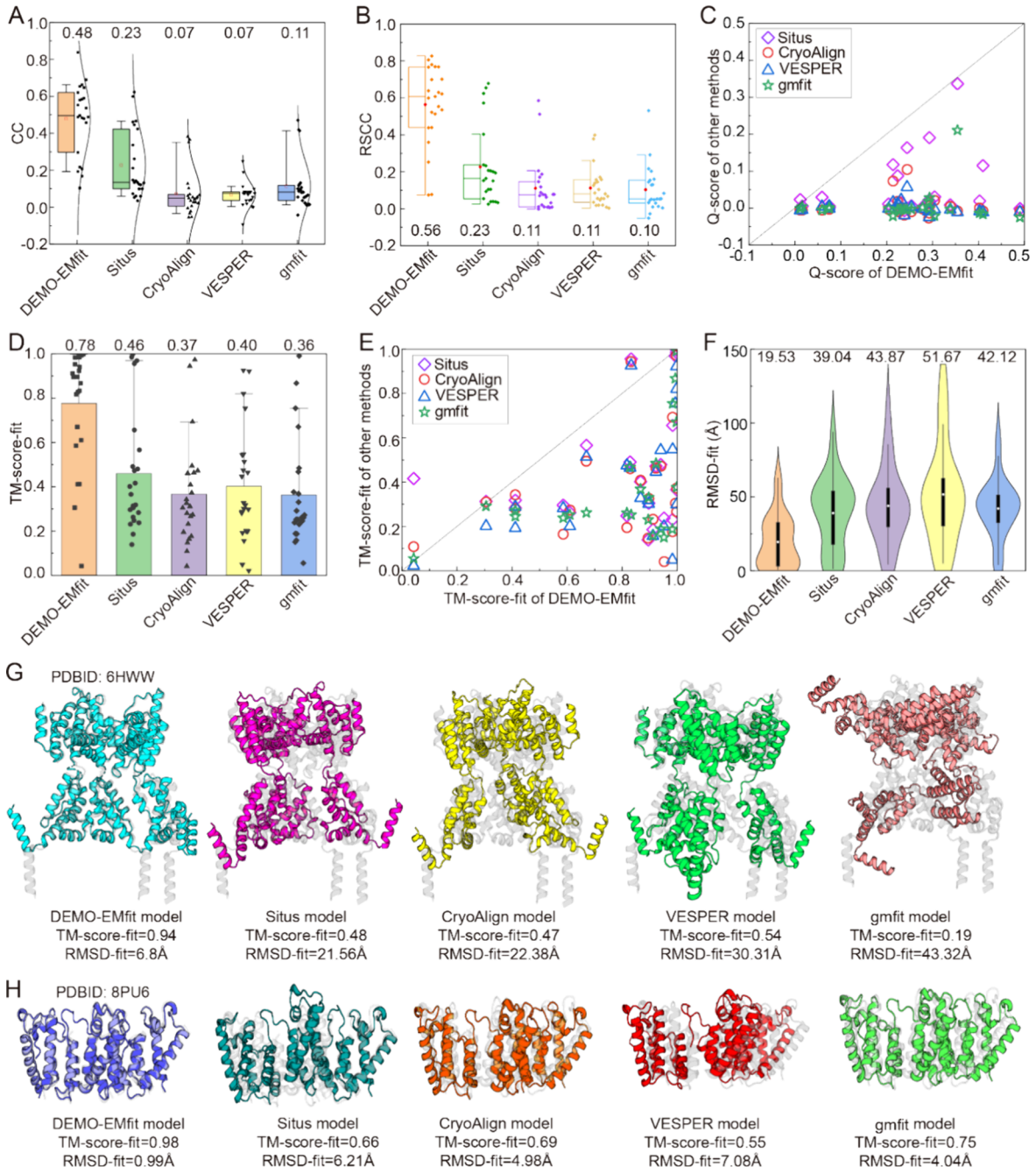


Figure 3. Results of fitting using predicted structures for 23 cryo-ET maps. (A) Box plots of CC values for models fitted by different methods. The red square represents the mean; the box covers the 25−75% range, the horizontal line within the box indicates the median, black dots depict the distribution of CC values, and whiskers extend to the 10−90% range. (B) Box plots of RSCC values for models fitted by different methods. The box represents the interquartile range (25−75%), the horizontal line and red circle mark the median and mean, respectively, and whiskers capture values up to 1.5 times the interquartile range. (C) Comparison of *Q*-scores for the structure fitted by DEMO-EMfit and other methods. (D) Box plots of TM-score-fit for models fitted by different methods, with bars indicating the mean, black dots representing the data distribution, and whiskers showing the 10−90% range. (E) Comparison of TM-score-fit for the structure fitted by DEMO-EMfit and other methods. (F) Violin-box plot of RMSD-fit. The violin shape illustrates the data distribution, the black bar denotes the interquartile range, the white circle represents the mean, and whiskers capture outliers up to 1.5 times the interquartile range. (G, H) Visual comparisons between deposited structures (translucent gray) and models fitted to PDB IDs 6HWW and 8PU6.

score-fit or a lower RMSD-fit suggests an improvement achieved through the fitting process.

Figure 2 reports the results generated by the different methods. The figure clearly demonstrates that DEMO-EMfit consistently restores structures close to the experimental, indicating its reliability and outstanding performance. As shown in Figure 2A,B, DEMO-EMfit achieves an average TM-score-fit of 0.99 and an RMSD-fit of 1.31 Å, significantly outperforming other methods, including Situs (TM-score-fit = 0.97, RMSD-fit = 3.51 Å), CryoAlign (TM-score-fit = 0.93, RMSD-fit = 9.71 Å), VESPER (TM-score-fit = 0.79, RMSD-fit = 11.25 Å), and gmfit (TM-score-fit = 0.71, RMSD-fit = 15.68 Å). Figure S2A,B illustrates the head-to-head comparison of TM-score-fit and RMSD-fit of fitting models by different

https://doi.org/10.1021/acs.jcim.5c00004
J. Chem. Inf. Model. 2025, 65, 3800−3811

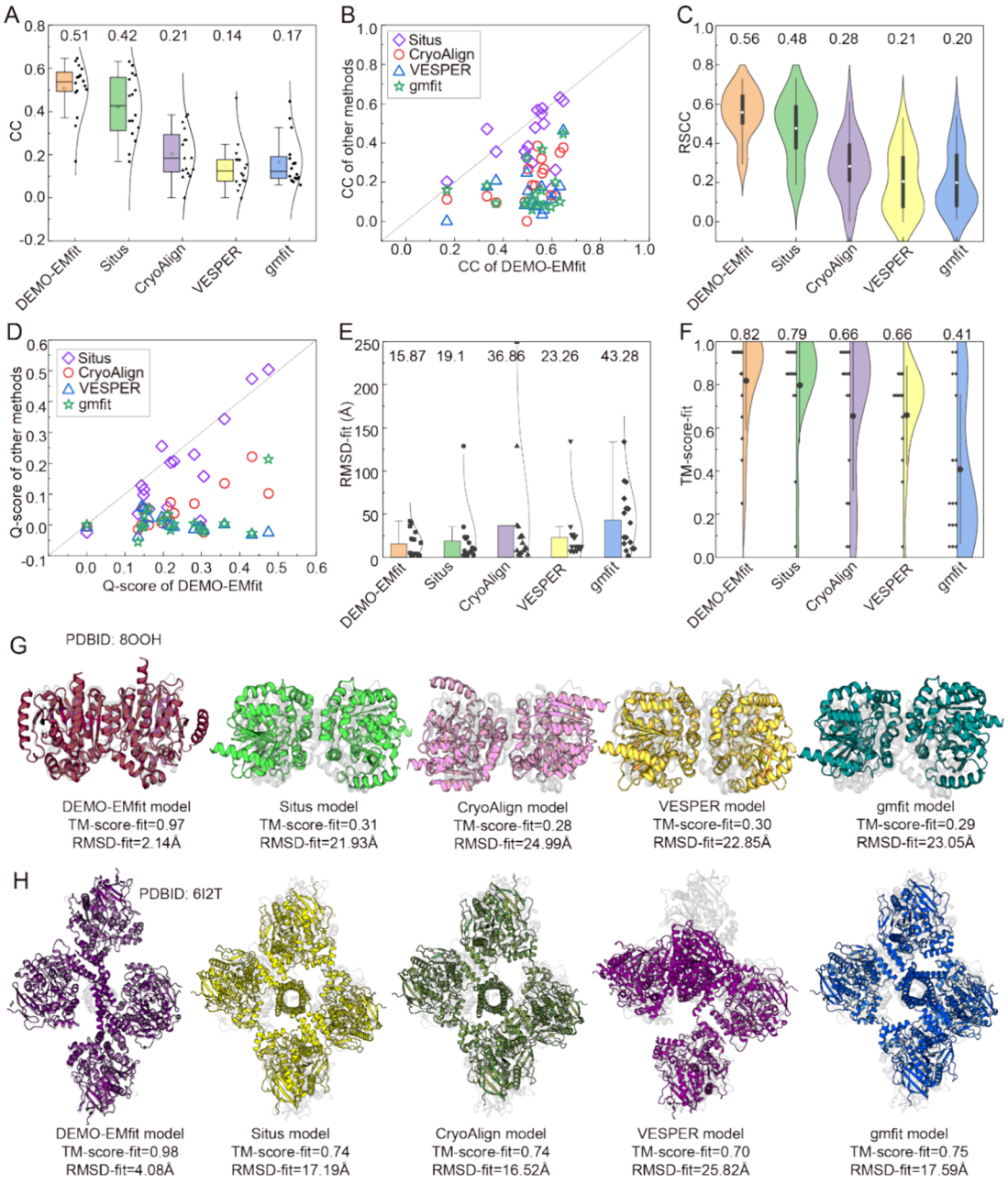


Figure 4. Results of fitting using predicted structures for 16 cryo-EM maps. (A) Box plot of CC values of models fitted by using different methods. Transparent squares indicate the mean, the horizontal line in each box represents the median, and black dots indicate the CC value for each model. Whiskers represent the distribution within a 1.5 interquartile range of outliers. (B) Comparison of CC values between DEMO-EMfit and other methods. (C) Violin plot of RSCC values obtained using different methods. Black boxes represent the interquartile range (25−75%), whiskers represent values within a 1.5 interquartile range of outliers, and white circles indicate the mean. The shape of each violin plot illustrates the data distribution. (D) Comparison of RSCC values between DEMO-EMfit and other methods. (E) Box plot of Q-scores of models fitted using different methods. Whiskers represent values within a 1.5 interquartile range of outliers. (F) Half-violin plots of TM-score-fit of fitted structures by different methods. Vertical lines represent the range of ±1 standard deviation from the mean, with half-violin plots showing distribution shapes. Diamonds indicate the TM-score-fit corresponding to domain distribution, and black circles indicate the average score for each method. (G, H) Comparison between the deposited structure (semitransparent gray) and the fitted structures for PDB IDs 8OOH and 6I2T.

methods. Additionally, Figure 2C presents the head-to-head comparison on CC of structures by different methods, where DEMO-EMfit's results significantly surpass those of other methods, highlighting its superior fitting performance (Figure S2C).

Since CC has certain biases in density correlation calculations, particularly in its inability to reflect local fitting accuracy, we present RSCC scores in Figure 2B. RSCC evaluates not only global fitting but also performance at the atomic, residue, or local region level. The details are shown in Figure 2D, and the average RSCC score for DEMO-EMfit was 0.73, which is higher than those for Situs (0.72), CryoAlign (0.52), VESPER (0.22), and gmfit (0.32).

To further evaluate DEMO-EMfit in greater detail, we utilize a more persuasive metric Q-score. Q-score is a local quantitative indicator designed to assess agreement between

https://doi.org/10.1021/acs.jcim.5c00004
J. Chem. Inf. Model. 2025, 65, 3800−3811

cryo-EM density maps and atomic models, aiming to evaluate the model quality based on the local resolution of the density map. A Q-score closer to 1 indicates a better match between each atom in the model and its corresponding density region. As shown in Figure 2E,F, DEMO-EMfit has a Q-score of 0.35, which is 2.9, 105.9, 3400, and 400% higher than Situs (0.34), CryoAlign (0.17), VESPER (0.01), and gmfit (0.07), respectively. The head-to-head comparison of Q-scores is reported in Figure S2D.

Figure 2G,H illustrates two representative examples with structures fitted into cryo-ET density maps using DEMO-EMfit. Specifically, the structure in Figure 2G, corresponding to PDB ID 7PTT, contains six chains and is fitted to a low-resolution density map at 8 Å. Despite the relatively low resolution of these density maps, our method demonstrates exceptional fitting accuracy, enabling the fitted models to align closely with the deposited structures (TM-score-fit = 0.99). Additionally, Figure 2H highlights the results for PDB ID 6T61, which is based on a high-resolution density map at 3.7 Å and represents a complex assembly with as many as 18 chains. The fitted model by DEMO-EMfit obtains a high overlap with the experiment structure. Figure 2I further shows case protein models fitted to cryo-EM density maps, with resolutions of 7.2 Å. These results also emphasize the outstanding fitting performance of our method across various resolutions. Notably, regardless of whether the density maps are of high or low resolution, our method consistently generates models that are highly congruent with the original deposited structures, achieving a TM-score-fit of 0.99.

Fitting Predicted Structures into Cryo-ET Maps. To validate the performance of DEMO-EMfit in practical application scenarios, we use AlphaFold3-predicted structures for fitting and evaluate it on cryo-ET and cryo-EM maps, respectively. In this part, we first evaluate DEMO-EMfit on 23 cryo-ET maps. The results obtained by DEMO-EMfit and the control methods are summarized in the Supporting Information, Tables S4−S8. Figure 3A illustrates the CC values of models fitted by using various methods. As shown, DEMO-EMfit significantly outperforms the other methods. The average CC value of models fitted by DEMO-EMfit is 0.48, which is 108.7, 585.7, 585.7, and 336.4% higher than those of Situs (0.23), CryoAlign (0.07), VESPER (0.07), and gmfit (0.11), respectively. Figure S3A provides the head-to-head comparison of CC values between DEMO-EMfit and the other methods.

Figure 3B reports the RSCC achieved by different methods. DEMO-EMfit again demonstrates superior performance, with an average RSCC of 0.56, substantially higher than those of Situs (0.23), CryoAlign (0.11), VESPER (0.11), and gmfit (0.10). Figure S3B further illustrates the head-to-head RSCC comparison. As shown in Figure 3C, a head-to-head comparison of the Q-score highlights the significant advantage of DEMO-EMfit. Figure S3C indicates that the average Q-score of models fitted by DEMO-EMfit is 0.24, far exceeding those of Situs (0.05), CryoAlign (0.01), VESPER (0.00), and gmfit (0.01). Additionally, Figure 3D−F shows that DEMO-EMfit achieves significantly higher TM-score-fit (0.78) and lower RMSD-fit (19.53 Å), as further detailed in Figure S3D. However, the fitting results for two protein structures are suboptimal, likely due to issues such as mean-shift or clustering during density map processing. These issues may have led to the omission of edge information, thereby reducing the accuracy of the fitting.

Figure 3G,H showcases two representative examples of structures fitted using DEMO-EMfit. Figure 3G highlights a cryo-EM density map with a resolution of 6.6 Å, corresponding to a PDB structure with two paired homologous chains (PDB ID 6HWW). Despite the low resolution, the model fitted by DEMO-EMfit achieves a remarkable TM-score-fit of 0.94 and an RMSD-fit of 6.8 Å, significantly outperforming models fitted by other methods. Figure 3H presents another case (PDB ID 8PU6, containing three homologous chains). The model fitted by DEMO-EMfit achieves an outstanding TM-score-fit of 0.98 and an RMSD-fit of 0.99 Å, which also outperforms the models fitted by the other methods. In addition, the AlphaFold2 predicted structures for these two cases achieve TM-scores of 0.47 (PDB ID 6HWW) and 0.35 (PDB ID 8PU6) relative to their respective deposited PDB structures. These values are markedly lower than those of the DEMO-EMfit models. This improvement can be attributed to the global optimization process employed by DEMO-EMfit, which effectively adjusts the positions and orientations of all domains, thereby significantly enhancing the overall structural fitting accuracy.

Fitting Predicted Structures into Cryo-EM Maps. DEMO-EMfit is further evaluated on 16 cryo-EM maps using AlphaFold3-predicted structures. Detailed results for each case are provided in the Supporting Information, Tables S9−S13. As shown in Figure 4, DEMO-EMfit continues to outperform other methods. Figure 4A,B demonstrates that the average CC value of DEMO-EMfit models is 0.51, which is 21.4, 142.9, 264.3, and 200% higher than those of Situs (0.42), CryoAlign (0.21), VESPER (0.14), and gmfit (0.17), respectively. Figure 4B further illustrates that the CC values of models fitted by DEMO-EMfit are consistently higher across cases compared to those of other methods. In terms of RSCC, DEMO-EMfit also outperforms other methods, as shown in Figure 4C and Figure S4A. The average RSCC for DEMO-EMfit is 0.56, surpassing Situs (0.48), CryoAlign (0.28), VESPER (0.21), and gmfit (0.20). Figure 4D highlights that the average Q-score of DEMO-EMfit models is 0.24, representing improvements of 33.3, 380, 2400, and 2300% compared to Situs (0.18), CryoAlign (0.05), VESPER (0.00), and gmfit (0.01), respectively (Figure S4B).

Figure 4E,F presents the TM-score-fit and RMSD-fit results, demonstrating that DEMO-EMfit achieves an average TM-score-fit of 0.82 and an average RMSD-fit of 15.87 Å. These results significantly outperform those of Situs (TM-score-fit = 0.79, RMSD-fit = 19.1 Å), CryoAlign (TM-score-fit = 0.66, RMSD-fit = 36.86 Å), VESPER (TM-score-fit = 0.66, RMSD-fit = 23.26 Å), and gmfit (TM-score-fit = 0.41, RMSD-fit = 43.28 Å). Figure S4C,D provides direct comparisons of the RMSD-fit and TM-score values between DEMO-EMfit and the other methods.

Figure 4G,H showcases comparisons between the models generated by DEMO-EMfit and those from other methods for two representative cases (PDB ID 8OOH and PDB ID 6I2T). In Figure 4G, DEMO-EMfit achieves a TM-score-fit of 0.97 and an RMSD-fit of 2.14 Å, outperforming Situs (TM-score-fit = 0.31, RMSD-fit = 21.93 Å), CryoAlign (TM-score-fit = 0.28, RMSD-fit = 24.99 Å), VESPER (TM-score-fit = 0.30, RMSD-fit = 22.85 Å), and gmfit (TM-score-fit = 0.29, RMSD-fit = 23.05 Å). Similarly, Figure 4H demonstrates that DEMO-EMfit achieves a TM-score-fit of 0.98 and an RMSD-fit of 4.08 Å, far exceeding Situs (TM-score-fit = 0.74, RMSD-fit = 17.19 Å), CryoAlign (TM-score-fit = 0.74, RMSD-fit = 16.52 Å),

https://doi.org/10.1021/acs.jcim.5c00004
J. Chem. Inf. Model. 2025, 65, 3800−3811

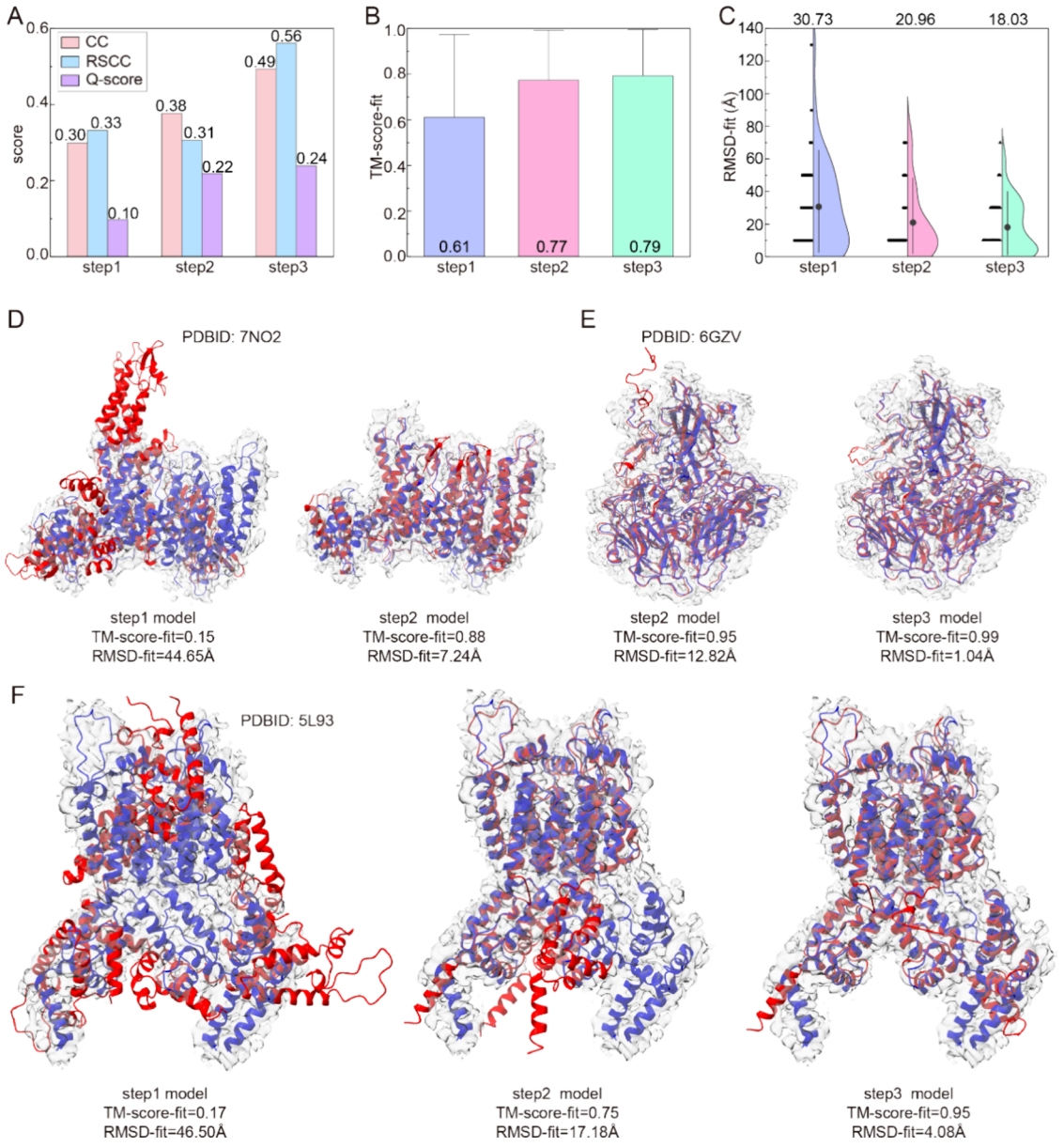


Figure 5. Results of ablation experiments for the 39 density maps, where step 1 involves FFT search; step 2 combines FFT search with LBFGS search; step 3 incorporates FFT search, LBFGS search, and domain optimization. (A) Bar charts of the average values of different scores for each scheme. (B) Box plots of TM-score-fit for each scheme, with whiskers indicating the 10−90% range. (C) Half-violin plots of RMSD-fit, where the shape indicates value distribution, black dots show data points, black circles denote mean values, and whiskers represent the 10−90% value range. (D−F) Comparison between the fitted structures (red) and the deposited structures (blue) for PDB IDs 7NO2, 6GZV, and 5L93, with the corresponding density maps shown as translucent overlays.

VESPER (TM-score-fit = 0.70, RMSD-fit = 25.82 Å), and gmfit (TM-score-fit = 0.75, RMSD-fit = 17.59 Å).

The marked improvement can be attributed to the enhanced domain orientation flexibility achieved through DEMO-EMfit's fitting process, highlighting its effectiveness in improving the overall fitting quality. Figure S5 further emphasizes the robustness of DEMO-EMfit: even when AlphaFold3-predicted structures contain inaccuracies, the final fitted models achieve significantly higher accuracy. This underscores the capability of DEMO-EMfit to overcome the limitations of initial predictions, reinforcing its effectiveness in refining the structural models. Additionally, Figure S6 presents the results of DEMO-EMfit across different protein types. The figure indicates that DEMO-EMfit performs consistently across all three protein types, with no noticeable bias toward any specific type.

Effects of Functional Modules in DEMO-EMfit. We designed ablation experiments to analyze the contribution of each functional module of DEMO-EMfit across all 39 maps. As shown in Figure 5A, we first ran DEMO-EMfit using only FFT search (step 1), achieving an average CC value of 0.30. Next, by incorporating the LBFGS search alongside the FFT search (step 2), the average CC value increased by 26.7%. Finally, when domain optimization is added to step 2 (step 3), the average CC value of DEMO-EMfit rose to 0.49. For RSCC, sequentially adding the LBFGS search and domain optimization to the FFT search increases the value from 0.33 to 0.56. Similarly, the *Q*-score improves from 0.10 to 0.22 and

https://doi.org/10.1021/acs.jcim.5c00004
J. Chem. Inf. Model. 2025, 65, 3800−3811

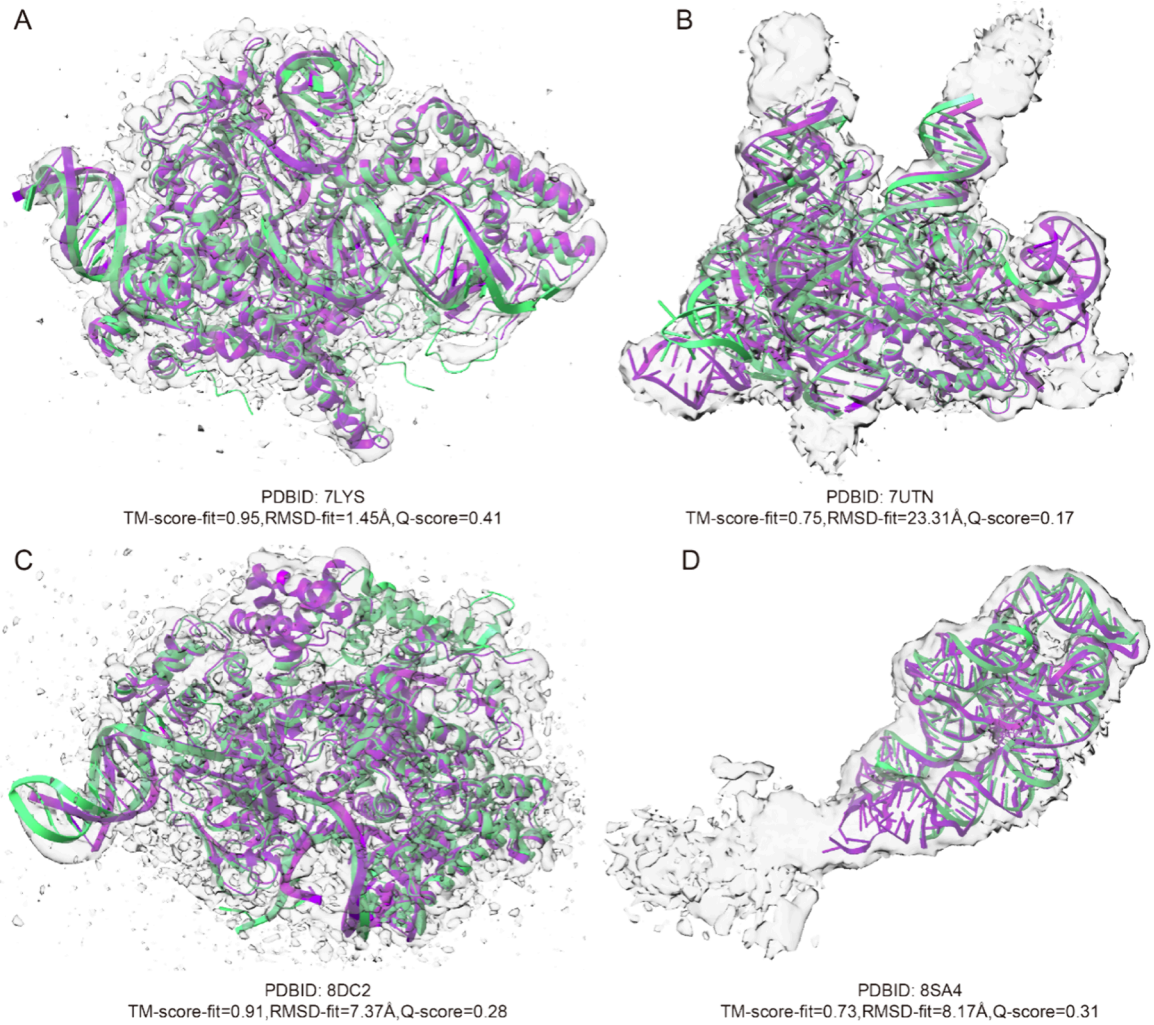


Figure 6. Representative examples of protein-NA complex models fitted by DEMO-EMfit. (A) 7LYS. (B) 7UTN. (C) 8DC2. (D) 8SA4. The deposited structure is colored purple, the DEMO-EMfit fitted model is colored green, and the density map is colored transparent gray.

eventually reaches 0.24. Figure 5B,C further demonstrates the effectiveness of local LBFGS search and domain optimization. The final fitted model achieves an average TM-score-fit of 0.79 and an RMSD-fit of 18.03 Å. Compared to step 1 (TM-score-fit = 0.61, RMSD-fit = 30.73 Å) and step 2 (TM-score-fit = 0.77, RMSD-fit = 20.96 Å), the TM-score-fit increases by 29.5 and 2.6%, respectively, while the RMSD-fit decreases by 9.77 and 2.93 Å, respectively. As shown in Figure S5, after the DEMO-EMFit fitting, most protein structures show improved or unchanged TM-score-fit and RMSD-fit.

Figure 5D illustrates a representative example (PDB ID 7NO2) where the fitting quality is significantly improved after adding a local LBFGS search. When using only the global FFT search in DEMO-EMfit, the fitted model (Figure 5D, left) achieves a TM-score-fit of merely 0.15, with an RMSD-fit as high as 44.65 Å. However, after incorporation of a local LBFGS search, the model (Figure 5D, right) achieves a TM-score-fit of 0.88, and the RMSD-fit decreases to 7.24 Å. Figure 5E provides a representative example demonstrating the effectiveness of domain optimization (PDB ID 6GZV, containing four chains). The model constructed by DEMO-EMfit (Figure 5E, left) achieves a TM-score fit of 0.95 without domain optimization. After performing domain optimization, the TM-score-fit improves to 0.99 (Figure 5E, right), and the RMSD-fit decreases significantly from 12.82 to 1.04 Å. These examples further validate the effectiveness of local search and domain optimizations.

Figure 5F illustrates a case (PDB ID 5L93, containing three chains) with stepwise improvement in fitting quality by sequentially adding local LBFGS searches and domain optimizations to the global FFT search. The left structure in Figure 5F compares the model fitted using only a global search in DEMO-EMfit with the deposited structure. The middle and right structures in Figure 5F show the comparisons of the fitted models, obtained by gradually adding local searches and domain optimizations, to the deposited structure. From the figure, we can observe that as the optimization progresses, the TM-score-fit increases steadily from 0.17 to 0.75 and ultimately to 0.95. Simultaneously, the RMSD-fit decreases progressively from 46.5 to 17.18 Å and eventually to 4.08 Å.

Figure S7 illustrates the significant advantages of deep-learning-based map processing (DLMP) in fitting. Specifically, Figure S7A shows that with the incorporation of DLMP, the average TM-score-fit of fitting results reaches 0.79, a 2.6% improvement compared to 0.77 without DLMP, indicating a positive impact of DLMP on overall fitting accuracy. Further analysis of Figure S7B reveals a detailed comparison of TM-score-fits for each model with and without DLMP. For most

https://doi.org/10.1021/acs.jcim.5c00004
J. Chem. Inf. Model. 2025, 65, 3800−3811

models, TM-score-fits significantly improve after deep learning predicts the backbone atomic density map. This result highlights the capability of DLMP to enhance the quality of the individual model fitting. Figure S7C,D shows the average RMSD-fit of fitting results. With DLMP, the average RMSD-fit is 18.03 Å, compared to 20.96 Å without it, a reduction of 2.93 Å. This finding further confirms that DLMP not only enhances the fitting precision (as reflected in the improved TM-score) but also significantly reduces deviations between the fitted and deposited structures (as indicated by the reduced RMSD). These results demonstrate that the introduction of DLMP optimizes the accuracy of the model fitting. In summary, the data and trends presented in these figures clearly showcase the remarkable improvements that DLMP brings to the fitting accuracy.

NA Complex Structure Fitting. In this section, we evaluate the performance of DEMO-EMfit on eight test cases to validate its applicability in fitting NA complex structures into density maps. These cases span a range of resolutions, sequence lengths, and molecular components, including both pure RNA/DNA structures and protein-RNA-DNA complexes. The initial models for these cases were generated using AlphaFold3, with TM-scores of 0.95, 0.71, 0.88, 0.75, 0.88, 0.84, 0.83, and 0.99 against the deposited structures for EMD-23600 (PDB ID 7LYS), EMD-26782 (PDB ID 7UTN), EMD-27320 (PDB ID 8DC2), EMD-40264 (PDB ID 8SA4), EMD-24283 (PDB ID 7R6N), EMD-33425 (PDB ID 7XSK), EMD-26816 (PDB ID 7UVT), and EMD-11022 (PDB ID 6Z0S), respectively. Here, we report only the results of DEMO-EMfit, as the control methods are applicable exclusively to protein structure fitting.

Figure 6 illustrates a comparison between the deposited structures (magenta) with the DEMO-EMfit-fitted structures (green) for the four representative cases. The fitted structures in Figure 6A,C achieve TM-score fits exceeding 0.9, with the latter case obtaining a TM-score improvement from 0.88 to 0.91. In Figure 6B, the initial AlphaFold3-predicted structure for this case has a TM-score of 0.71. DEMO-EMfit successfully fits the structure into the map, resulting in an improved model with a TM-score-fit of 0.75, compared to the started AlphaFold3 model. Figure 6D presents the fitting result for EMD-40264, a pure RNA structure. The DEMO-EMfit fitted structure achieves a TM-score-fit of 0.73 and an RMSD-fit of 8.17 Å, which are comparable to the AlphaFold3-predicted structure's TM-score of 0.75 and RMSD of 7.65 Å. The results for the four pure RNA/DNA structures are shown in Figure S8, showing performance consistent with that of the protein-NA complex cases. These results demonstrate the effectiveness of DEMO-EMfit in fitting NA complex structures to density maps.

Computational Efficiency. Based on experiments with 39 test targets, the average runtime for the entire DEMO-EMfit pipeline when starting from complex models predicted by AlphaFold3 is approximately 12.5 min. Each test was conducted on a single core of an AMD EPYC 7742 64-core CPU at 3.25 GHz. Figure S9A illustrates the relationship between the runtime and the number of residues, showing that 74.4% of the proteins are processed within 15 min. Additionally, Figure S9B presents the relationship between the runtime and the number of protein chains, indicating that 76.5% of proteins with fewer than 8 chains are completed within 15 min.

CONCLUSIONS

In this study, we introduce DEMO-EMfit, a method designed to achieve precise alignment between atomic structures and density maps through density map preprocessing and the global-local pose search strategy. DEMO-EMfit is evaluated on 16 single-particle maps and 23 subtomogram averaging maps, and its performance is benchmarked against state-of-the-art methods including Situs, CryoAlign, VESPER, and gmfit. Results on the benchmark set demonstrate that DEMO-EMfit generates more accurate fitting structures than the control methods. The average TM-score-fit achieved by DEMO-EMfit is 31.7% higher than that of the best-performing control method, with a statistically significant p-value of 5.5×10^{-8} in Student's t test for the fitting of predicted structures. Additionally, the results indicate the effectiveness of DEMO-EMfit for protein-NA complex structure fitting.

The detailed ablation analyses showed that the superior performance of DEMO-EMfit can be attributed to the following key features: (1) DEMO-EMfit leverages a deep learning approach to isolate backbone regions from density maps, effectively reducing noise and irrelevant information. (2) DEMO-EMfit involves a progressive pose search strategy to detect the global position of the structure within the map, followed by a local enhanced search to improve pose accuracy. (3) DEMO-EMfit accommodates input structures with incorrect domain orientations by employing domain-level adjustments for precise fitting.

While DEMO-EMfit has demonstrated promising results, several aspects remain for potential improvement. First, the mean-shift processing of the density map may inadvertently filter out edge information, which could impact the accuracy of the structure fitting. Applying multiple contour levels of the density map or increasing the bandwidth of the Gaussian kernel function in eq 2 could help preserve edge details. Additionally, integrating deep learning-based edge extraction with the mean-shift processing could alleviate the overfiltering. Second, the global-local search remains time-consuming for extremely large maps as it is guided by traditional CC. Deep learning-based end-to-end fitting approaches, which directly learn structural poses using point clouds extracted from maps, could significantly improve both accuracy and efficiency. Furthermore, leveraging multicore processing or distributed computing to parallelize the global-local search process, combined with deep learning-driven key voxel extraction, could further accelerate computation and improve performance. Third, traditional CC metrics may not always provide reliable evaluations of the fitting quality for certain maps. Physics-based Gaussian mixture models,[42] combined with DOT scores from VESPER, could be advantageous for enhancing fitting accuracy. Fourth, due to the weak signal of small molecules, accurately fitting protein-small-molecule complexes remains a challenge, particularly in flexible fitting scenarios. Developing more precise computational tools that enhance small-molecule density identification using deep learning could help improve the fitting accuracy. Efforts along these aspects will further strengthen DEMO-EMfit's capability in structure fitting for cryo-EM density maps.

ASSOCIATED CONTENT

Data Availability Statement

All data needed to evaluate the conclusions are present in the paper and the Supporting Information. All original code has

https://doi.org/10.1021/acs.jcim.5c00004
J. Chem. Inf. Model. 2025, 65, 3800−3811

been deposited at https://github.com/xiaogenz/DEMO-EMfit. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.5c00004.

(Figures S1−S9) Comparison of the AF2-predicted structure and native structure; results of fitting using experimental structures for all cases, using predicted structures for 23 cryo-ET maps, and using predicted structures for 16 cryo-EM maps; comparison of the final fitting results with AF2 models; bar chart of different types of protein scores; results of 39 density maps; representative examples of pure RNA/DNA models fitted using DEMO-EMfit; runtime of DEMO-EMfit; (Tables S1−S13) 23 proteins with experimental cryo-ET density maps; 16 proteins with experimental cryo-EM density maps; comparison of the core algorithms; CC, Q-score, RSCC, TM-score-fit, and RMSD-fit values of the models (PDF)

AUTHOR INFORMATION

Corresponding Authors

Guijun Zhang − *College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China;* orcid.org/0000-0002-7815-5884; Email: zgj@zjut.edu.cn

Xiaogen Zhou − *College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China;* orcid.org/0000-0001-6839-1923; Email: zxg@zjut.edu.cn

Authors

Yaxian Cai − *College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China;* orcid.org/0009-0002-3993-6490

Ziying Zhang − *College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China*

Xiangyu Xu − *College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China*

Liang Xu − *College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China*

Yu Chen − *School of Computer Science and Technology, Soochow University, Suzhou 215006, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.5c00004

Author Contributions

†Y. Cai, Z. Zhang, and X. Zhou contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China [2022ZD0115103], the National Nature Science Foundation of China [62203389 and 62201506], Fundamental Research Funds for the Provincial Universities of Zhejiang [RF-C2024006], Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang [2023R01006], Zhejiang Provincial Special Support Program for High-Level Talents (2023R5248), the "Pioneer" and "Leading Goose" R&D Program of Zhejiang [2025C01121], and General Scientific Research Projects of Zhejiang Education Department [Y202455578]. This work used the Extreme Science and Engineering Discovery Environment (XSEDE),[43] which is supported by the National Science Foundation (ACI1548562).

REFERENCES

(1) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C. C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Zemgulyte, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Zidek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493−500.

(2) 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.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871−876.

(3) Zhou, X.; Zheng, W.; Li, Y.; Pearce, R.; Zhang, C.; Bell, E. W.; Zhang, G.; Zhang, Y. I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction. *Nat. Protoc* **2022**, *17*, 2326−2353.

(4) Zhou, X. G.; Li, Y.; Zhang, C. X.; Zheng, W.; Zhang, G. j.; Zhang, Y. Progressive assembly of multi-domain protein structures from cryo-EM density maps. *Nature Computational Science* **2022**, *2*, 265−275.

(5) Li, T.; He, J.; Cao, H.; Zhang, Y.; Chen, J.; Xiao, Y.; Huang, S. Y. All-atom RNA structure determination from cryo-EM maps. *Nat. Biotechnol.* **2024**, 97.

(6) Punjani, A.; Zhang, H.; Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **2020**, *17*, 1214−1221.

(7) Yip, K. M.; Fischer, N.; Paknia, E.; Chari, A.; Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **2020**, *587*, 157−161.

(8) Liu, M.; Huang, Y.; Chengguang, Z. CryoNeFEN: High-resolution reconstruction of cryo-EM structures using neural field network. *ResearchSquare* **2024**.

(9) Lawson, C. L.; Patwardhan, A.; Baker, M. L.; Hryc, C.; Garcia, E. S.; Hudson, B. P.; Lagerstedt, I.; Ludtke, S. J.; Pintilie, G.; Sala, R.; Westbrook, J. D.; Berman, H. M.; Kleywegt, G. J.; Chiu, W. EMDataBank unified data resource for 3DEM. *Nucleic Acids Res.* **2016**, *44*, D396−403.

(10) Chen, Y.-X.; Xie, R.; Yang, Y.; He, L.; Feng, D.; Shen, H.-B. Fast Cryo-EM Image Alignment Algorithm Using Power Spectrum Features. *J. Chem. Inf. Model.* **2021**, *61*, 4795−4806.

(11) Zhang, J.; Zhang, S.; Zhou, W.; Zhang, X.; Li, G.; Li, R.; Lin, X.; Chen, Z.; Liu, F.; Shen, P.; Zhou, X.; Gao, Y.; Chen, Z.; Chao, Y.; Wang, C. A widely conserved protein Rof inhibits transcription termination factor Rho and promotes Salmonella virulence program. *Nat. Commun.* **2024**, *15*, 3187.

(12) Molodtsov, V.; Wang, C.; Firlar, E.; Kaelber, J. T.; Ebright, R. H. Structural basis of Rho-dependent transcription termination. *Nature* **2023**, *614*, 367−374.

(13) Hryc, C. F.; Baker, M. L. AlphaFold2 and CryoEM: Revisiting CryoEM modeling in near-atomic resolution density maps. *Iscience* **2022**, 25.

(14) Wang, R. Y.-R.; Kudryashev, M.; Li, X.; Egelman, E. H.; Basler, M.; Cheng, Y.; Baker, D.; DiMaio, F. De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nat. Methods* **2015**, *12*, 335−338.
(15) DiMaio, F.; Song, Y.; Li, X.; Brunner, M. J.; Xu, C.; Conticello, V.; Egelman, E.; Marlovits, T. C.; Cheng, Y.; Baker, D. Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat. Methods* **2015**, *12*, 361.
(16) Dahmani, Z. L.; Scott, A. L.; Vénien-Bryan, C.; Perahia, D.; Costa, M. G. MDFF_NM: Improved Molecular Dynamics Flexible Fitting into Cryo-EM Density Maps with a Multireplica Normal Mode-Based Search. *J. Chem. Inf. Model.* **2024**, *64*, 5151−5160.
(17) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, 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.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583−589.
(18) Wang, X.; Zhu, H.; Terashi, G.; Taluja, M.; Kihara, D. DiffModeler: large macromolecular structure modeling for cryo-EM maps using a diffusion model. *Nat. Methods* **2024**, 2307.
(19) He, J.; Lin, P.; Chen, J.; Cao, H.; Huang, S. Y. Model building of protein complexes from intermediate-resolution cryo-EM maps with deep learning-guided automatic assembly. *Nat. Commun.* **2022**, *13*, 4066.
(20) Zhang, Z.; Cai, Y.; Zhang, B.; Zheng, W.; Freddolino, L.; Zhang, G.; Zhou, X. DEMO-EM2: assembling protein complex structures from cryo-EM maps through intertwined chain and domain fitting. *Brief Bioinform* **2024**, *25*, bbae113.
(21) Pfab, J.; Phan, N. M.; Si, D. DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, No. e2017525118.
(22) Jamali, K.; Kall, L.; Zhang, R.; Brown, A.; Kimanius, D.; Scheres, S. H. W. Automated model building and protein identification in cryo-EM maps. *Nature* **2024**, *628*, 450−457.
(23) Giri, N.; Cheng, J. De novo atomic protein structure modeling for cryoEM density maps using 3D transformer and HMM. *Nat. Commun.* **2024**, *15*, 5511.
(24) Su, B.; Huang, K.; Peng, Z.; Amunts, A.; Yang, J. Improved automated model building for cryo-EM maps using CryFold. *bioRxiv* **2024**, *2024* (11), 13.623164.
(25) C.P.van Zundert, G.; M.J.J. Bonvin, A. Fast and sensitive rigid-body fitting into cryo-EM density maps with PowerFit. *AIMS Biophysics* **2015**, *2*, 73−87.
(26) Kawabata, T. Gaussian-input Gaussian mixture model for representing density maps and atomic models. *J. Struct Biol.* **2018**, *203*, 1−16.
(27) Tjioe, E.; Lasker, K.; Webb, B.; Wolfson, H. J.; Sali, A. MultiFit: a web server for fitting multiple protein structures into their electron microscopy density map. *Nucleic Acids Res.* **2011**, *39*, W167−70.
(28) Wriggers, W. Using Situs for the integration of multi-resolution structures. *Biophys Rev.* **2010**, *2*, 21−27.
(29) Han, X.; Terashi, G.; Christoffer, C.; Chen, S.; Kihara, D. VESPER: global and local cryo-EM map alignment using local density vectors. *Nat. Commun.* **2021**, *12*, 2090.
(30) Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkoczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L. W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct Biol.* **2019**, *75*, 861−877.
(31) Zhou, X. G.; Hu, J.; Zhang, C. X.; Zhang, G. J.; Zhang, Y. Assembling multidomain protein structures through analogous global structural alignments. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 15930−15938.
(32) Zhou, X.; Peng, C.; Zheng, W.; Li, Y.; Zhang, G.; Zhang, Y. DEMO2: Assemble multi-domain protein structures by coupling analogous template alignments with deep-learning inter-domain restraint prediction. *Nucleic Acids Res.* **2022**, *50*, W235−W245.
(33) DiMaio, F.; Tyka, M. D.; Baker, M. L.; Chiu, W.; Baker, D. Refinement of protein structures into low-resolution density maps using rosetta. *J. Mol. Biol.* **2009**, *392*, 181−90.
(34) Zheng, W.; Zhou, X.; Wuyun, Q.; Pearce, R.; Li, Y.; Zhang, Y. FUpred: detecting protein domains through deep-learning-based contact map prediction. *Bioinformatics* **2020**, *36*, 3749−3757.
(35) Zhou, X. G.; Zhang, G. J. Differential evolution with underestimation-based multimutation strategy. *IEEE transactions on cybernetics* **2019**, *49*, 1353−1364.
(36) Storn, R.; Price, K. Differential evolution−a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization* **1997**, *11*, 341−359.
(37) Zhou, X. G.; Peng, C. X.; Liu, J.; Zhang, Y.; Zhang, G. J. Underestimation-assisted global-local cooperative differential evolution and the application to protein structure prediction. *IEEE Transactions on Evolutionary Computation* **2020**, *24*, 536−550.
(38) Zhang, Y.; Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **2005**, *33*, 2302−9.
(39) Xu, J.; Zhang, Y. How significant is a protein structure similarity with TM-score= 0.5? *Bioinformatics* **2010**, *26*, 889−895.
(40) Zhang, C.; Shine, M.; Pyle, A. M.; Zhang, Y. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat. Methods* **2022**, *19*, 1109−1115.
(41) Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr., Sect. A* **1976**, *32*, 922−923.
(42) Chen, M.; Schmid, M. F.; Chiu, W. Improving resolution and resolvability of single-particle cryoEM structures using Gaussian mixture models. *Nat. Methods* **2024**, *21*, 37−40.
(43) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; Roskies, R.; Scott, J. R.; Wilkins-Diehr, N. XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering* **2014**, *16*, 62−74.

https://doi.org/10.1021/acs.jcim.5c00004
J. Chem. Inf. Model. 2025, 65, 3800−3811