THE JOURNAL OF PHYSICAL CHEMISTRY B

pubs.acs.org/JPCB Article

Fitting Low-Resolution Protein Structures into Cryo-EM Density Maps by Multiobjective Optimization of Global and Local Correlations

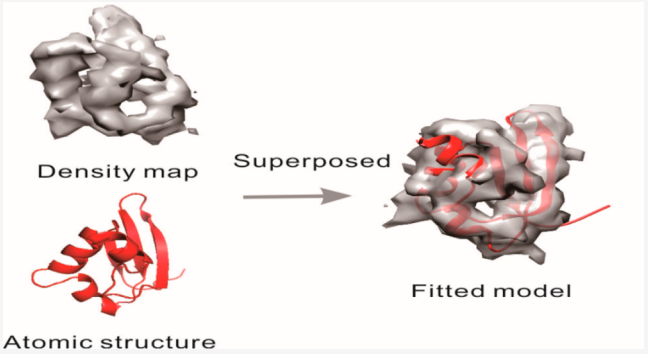
Published as part of The Journal of Physical Chemistry virtual special issue "Ruth Nussinov Festschrift".

Biao Zhang, Wenyi Zhang, Robin Pearce, Yang Zhang,* and Hong-Bin Shen*

Cite This: *J. Phys. Chem. B* 2021, 125, 528−538 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The rigid-body fitting of predicted structural models into cryo-electron microscopy (cryo-EM) density maps is a necessary procedure for density map-guided protein structure determination and prediction. We proposed a novel multiobjective optimization protocol, MOFIT, which performs a rigid-body density-map fitting based on particle swarm optimization (PSO). MOFIT was tested on a large set of 292 nonhomologous single-domain proteins. Starting from structural models predicted by I-TASSER, MOFIT achieved an average coordinate root-mean-square deviation of 2.46 Å, which was 1.57, 2.79, and 3.95 Å lower than three leading single-objective function-based methods, where the differences were statistically significant with *p*-values of 1.65×10^{-6}, 6.36×10^{-8}, and 6.44×10^{-11} calculated using two-tail Student's *t* tests. Detailed analyses showed that the major advantages of MOFIT lie in the multiobjective protocol and the extensive PSO search simulations guided by the composite objective functions, which integrates complementary correlation coefficients from the global structure, local fragments, and individual residues with the cryo-EM density maps.



INTRODUCTION

Obtaining high-resolution protein structures is vital to understanding the functions of proteins and the mechanisms of disease pathways they are involved in. Several methods, such as X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy (cryo-EM),[1−4] have been successfully used to obtain macromolecule structures.[5−9] In recent years, a revolution has taken place in cryo-EM imaging, leading to a large amount of cryo-EM density maps, which often have relatively medium-to-high resolutions (e.g., ∼3−10 Å).[10,11] Although the resolution of some electron microscopy (EM)-derived density maps is not sufficiently high, they can usually provide an accurate topological description of the molecular structure and, thus, can act as an optimization restraint for a predicted atomic structure.[12] Such a density map-constrained refinement has been a promising direction in the field of protein structure prediction.[13−16] This refinement modeling procedure usually includes three steps, which are (1) atomic structure prediction using cutting-edge methods such as I-TASSER,[17] which identifies and assembles fragments from numerous threading templates in order to produce full-length models, (2) fitting the atomic structure into an EM density map, and (3) atomic refinement iterations. Fitting the structure of a protein into a cryo-EM density map is an essential step to decrease the conformational search space during the subsequent atomic refinement step and to ensure that refinement is accurately guided by the density map. It is however a nontrivial task, especially for low- and medium-resolution density maps, and advanced computational techniques are required to fill the void of misinformation from cryo-EM density maps in order to obtain high-resolution protein structures.[18−20]

Currently, several computational methods have been developed to fit atomic structures into experimental cryo-EM density maps. These programs include EMFIT,[21] colores of Situs,[22,23] 3SOM,[24] MultiFit,[25] ADP_EM,[26] Attract-EM,[27] EMatch,[28] PowerFit,[29] and UCSF Chimera,[30] which have been developed to provide relevant structural insights into macromolecular complex structures. These programs usually perform an automated search of all the possible relative rotations and translations to maximize a cross-correlation

Received: November 3, 2020
Revised: December 17, 2020
Published: January 5, 2021

https://dx.doi.org/10.1021/acs.jpcb.0c09903
J. Phys. Chem. B 2021, 125, 528−538

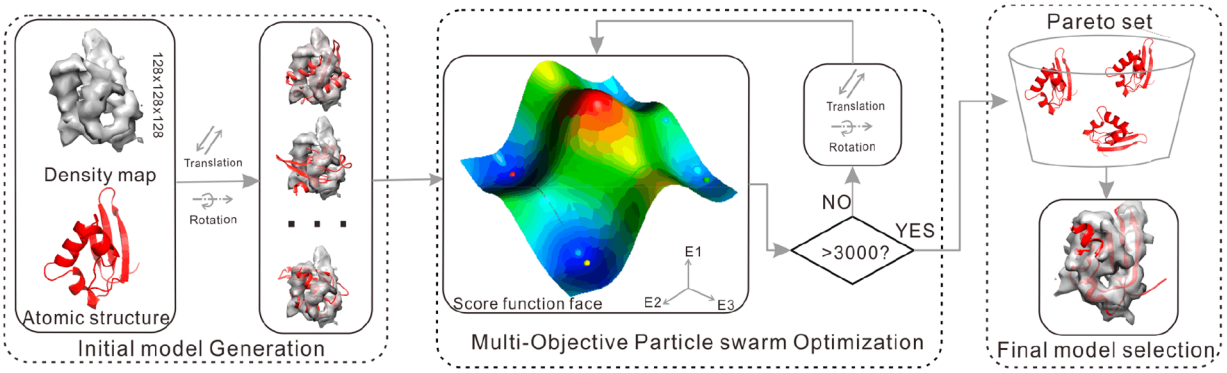


Figure 1. Flowchart of the MOFIT procedure that rigidly fits a predicted protein structure into a cryo-EM density map. MOFIT includes three objective functions: E1 (global CC), E2 (fragment-level CC), and E3 (residue-level CC).

function to improve a structure's fit into a density map. As an example of these methods, UCSF Chimera[30] is a program for the interactive visualization and analysis of molecular structures and related data that can be used to locally optimize the fitting of atomic coordinates into a density map. Additionally, ADP_EM is a multiresolution docking method that executes a search in rotational space to maximize the correlation using a fast-rotational matching method.[26] Moreover, colores of Situs is a contour-based matching method that utilizes a six-dimensional search using fast Fourier transforms to rapidly scan the rigid-body degrees of freedom for a probe molecule relative to a fixed target density map.[23] Lastly, EMatch performs a fitting using template-matching procedures, which identifies secondary structure elements in the three-dimensional (3D) EM maps.[28]

Although many successes have been achieved, several issues still limit the efficacy of existing fitting algorithms. First, the metric used to evaluate the fitting quality is an important criterion that will guide the search direction. Most existing algorithms use a linear combination of different objective functions, such as the structure-map correlation coefficient,[26] the protrusion term score between the subunit and density envelope,[31] and molecular contour information,[23] as the optimization criteria. The fitting between a predicted structure and the EM density map can be very complicated and will result in different fitted poses due to several complicating factors, such as the quality of the predicted atomic structure, the resolution of the density map, the signal-to-noise ratio (SNR) of the density map, misalignment between the predicted structure and the density map, and other factors. Linear combination can consider multiple factors, but generally it is difficult to derive proper weights in order to balance each term for different targets. Furthermore, cascading multiple factors into a single objective function would also weaken the conflict among different factors, which can be a much more positive effect to avoid the local optimum in the optimization iterations. Second, since the search space is large during the fitting process, an exhaustive search performed without processing, such as using Fourier techniques or filters, is a time-consuming component. A heuristic searching algorithm should help balance the efficiency and robustness of the procedure.

In this work, we propose a new method named MOFIT, which utilizes multiobjective particle swarm optimization[32] (PSO) to perform a rigid-body fitting of a predicted structure into a density map. As shown in Figure 1, the pipeline of MOFIT contains three consecutive steps of model generation, multiobjective particle swarm optimization, and final model selection. Given the advancement in structure prediction, we first used I-TASSER[17] to generate initial models for the target sequences. Next, three objective functions, which account for correct global topology and an accurate local fitting using fragment- and residue-level information, were constructed to guide the heuristic PSO search, where the final models are selected from the knee points of the nondominated Pareto solutions.[33] To carefully examine the strengths and weaknesses of MOFIT, we performed large-scale benchmark tests with density maps created from both noise-free and noisy simulations. The results showed significant advantages of the MOFIT pipeline for a cryo-EM density map-based protein atomic structure fitting compared to other state-of-the-art approaches. The online server and standalone package of MOFIT, together with a detailed tutorial of how to install and use the program, are available at www.csbio.sjtu.edu.cn/bioinf/MOFIT/. Similarly, the I-TASSER server and package are freely available to academic users at https://zhanglab.ccmb.med.umich.edu/I-TASSER/.

METHODS

MOFIT consists of three steps of initial structure generation, multiobjective particle swarm optimization, and final model selection. A flowchart of the program is depicted in Figure 1.

Calculated Density Map. When fitting a protein atomic structure into a cryo-EM density map, the protein structure must first be translated to a calculated density map in order to compute the correlation. For a given atomic structure of a protein that contains only Cα atoms with coordinates x_1–x_N, the calculated density map on grid y can be obtained by

$$\rho_c(y) = \sum_{x_i \in N} C \cdot \exp(-k \cdot \|x_i - y\|^2) \quad (1)$$

where $k = (\pi/(2.4 + 0.8R_0))^2$ and $C = a \cdot (k/\pi)^{1.5}$ are parameters that describe the shape of the Gaussian kernel,[34] R_0 is the map resolution, and a is the mass of the Cα atom.

Multiobjective Function. MOFIT uses three different objective functions that account for the global and local evaluation of the fitting quality.

Objective Function 1 is the correlation coefficient between the calculated density map $\rho_c(y)$ from the atomic structure and experimental density map $\rho_o(y)$, which can be calculated by

https://dx.doi.org/10.1021/acs.jpcb.0c09903
J. Phys. Chem. B 2021, 125, 528−538

$$\begin{cases} E1 = 1.0 - \mathrm{CC}(r) \\ \mathrm{CC}(r) = \dfrac{\sum\limits_{y\in \mathrm{RB}(l)} (\rho_o(y) - \bar{\rho}_o)(\rho_c(y) - \bar{\rho}_c)}{\sqrt{\sum\limits_{y\in \mathrm{RB}(l)} (\rho_o(y) - \bar{\rho}_o)^2} \sqrt{\sum\limits_{y\in \mathrm{RB}(l)} (\rho_c(y) - \bar{\rho}_c)^2}} \end{cases} \tag{2}$$

where $\bar{\rho}_c$ and $\bar{\rho}_o$ are the average values of the calculated density map and experimental density map, respectively. RB (l) represents all the grid point sets, and r represents all the residues. The CC is a global score that is sensitive to the shape of the density map.

Objective Function 2 accounts for the local correlation coefficient, which can be calculated by

$$E2 = 1.0 - \frac{L+1}{L^2} \sum_{i}^{L_1} \frac{1}{1 + \frac{1}{L * \mathrm{CC}(i)}} \tag{3}$$

where CC (i) represents the correlation coefficient between residue i and the corresponding grid point in the density map; L_1 represents all residues that have CC $(i) > 0$; L is the length of the amino acid sequence for the atomic structure. This function represents the similarity of the local residues.

Objective Function 3 evaluates the fragments for which all residues have CC (i) scores greater than 0 and can be calculated as follows

$$E3 = 1.0 - \frac{2}{L} \sum_{i}^{L_2} \frac{1}{1 + \frac{L}{\mathrm{CV}(i)}} \tag{4}$$

where $\mathrm{CV}(i) = \sum_{i_{\mathrm{down}}}^{i_{\mathrm{up}}} f(i)$; $f(i) = 1$ if CC $(i) > 0$ or $f(i) = 0$ otherwise; i_{up} represents continued residues ahead of residue i with CC $(i) > 0$; i_{down} represents continued residues after residue i with CC $(i) > 0$; L_2 represents the set for which CC (i) is greater than 0.

Multiobjective Optimization. *Formulating A Density Map Fitting as a Multiobjective Optimization Problem.* Given the three objective functions (eqs 2−4), a unique solution probably does not exist that lies at the global optimum of all the objectives, since the objective functions can have conflicts for some proteins where optimal decisions need to be balanced with trade-offs between two or more different objectives. The fittings for which none of the objective functions can be improved in value without degrading the other objectives are called the Pareto solutions. Without additional subjective preference information, all Pareto optimal solutions are considered equally good. The multiobjective optimization problem can be studied from different viewpoints, and there exist various solution philosophies and goals. A set of Pareto-optimal solutions can be obtained and quantified with respect to the trade-offs in satisfying the different objectives. The standard paradigm for multiobjective models can be written as

$$\begin{aligned} &\min F(x) = [f_1(x), f_2(x), f_3(x)]^T \\ &\quad s.t.\ g_i(x) > 0,\ i = 1, 2, \ldots, L \end{aligned} \tag{5}$$

where f_1, f_2, and f_3 are the three objective functions that are defined in eqs 2−4; $g_i(x)$ is the corresponding constraint function; x is the coordinate of the atomic model; and i is the residue index of the atomic structure.

Multiobjective Particle Swarm Optimization (MOPSO). In recent years, multiobjective optimization technology based on heuristics has been dramatically improved, which is now more practical and efficient than the classical method.[35] The PSO algorithm is an evolutionary technology based on group intelligence that simulates social behavior. It has been widely used in theoretical and engineering optimization due to its unique search mechanism, excellent convergence performance, and convenient computer implementation.[36] Multiobjective optimization is an efficient solution to the vast conformation space that must be searched in our specific problem of protein structure and density map fitting. We use the multiobjective particle swarm optimization (MOPSO) algorithm, proposed by Coello,[37] to search for the global best position and the best position by the dominance swarm relationship, where all swarms move toward the optimal direction, whose implementation has three steps.

Step 1. Initialization. The model predicted by I-TASSER will first be translated into the density map center and then randomly rotated to generate N different poses. Each pose is considered one particle in the MOPSO optimization. The ith particle is represented by a two-dimensional vector $C = [t, r]$. Here, t and r represent the random rigid body translation and rotation movement scales that range from [−1 Å, 1 Å] and [−90°, 90°], respectively.

Step 2. Optimization and Searching. Each model's position is updated by rigid-body translation and rotation from the main simulation's cycle. In each simulation, the three objective functions are calculated according to a conformation's coordinates. Then, a nondominated solution that has at least two decreasing objectives functions is put into the Pareto set. The new conformations are updated according to the particle formulas of

$$\begin{aligned} v_i^k = {} & \omega \times v_i^{k-1} + c_1 \times \gamma \times (P_i^k - C_i^{k-1}) \\ & + c_2 \times \gamma \times (G_i^k - C_i^{k-1}) \end{aligned} \tag{6}$$

where v_i^k is the change of movement in model i at the kth iteration; $C_i^k = C_i^{k-1} + v_i^k$ is the new conformation of model i at the kth iteration; ω is the inertia weight, which we linearly decrease from 1.5 to 0.5 during the simulation according to our local test and recommended empirical value;[38] γ is a random value in the range of [0,1], which is used to introduce perturbation; c_1 and c_2 are set to 2;[39] P_i^k is the best conformation of model i in the previous iteration; and G_i^k is randomly selected from the Pareto set in the current iteration representing the best conformation in the current iteration.

Step 3. Solution Ranking. We rank all the models of the Pareto set and select the top conformation as the final structure.

Final Model Selection. Usually, there exist multiple Pareto-optimal solutions for multiobjective optimization problems. For single-objective optimization, we can achieve the best performance for one solution. However, it is not easy to judge which solution is better in the Pareto set, since all are so-called nondominated solutions. Here, we rank the Pareto solution using a knee score and select the knee point as the final solution.[33] Knee points can give better answers among the set of nondominated solutions,[36] and they are widely used as a criterion for a final solution selection from the Pareto set. However, while the importance of knee points and front regions have long been recognized in multiobjective

https://dx.doi.org/10.1021/acs.jpcb.0c09903
J. Phys. Chem. B 2021, 125, 528−538

optimization, they have not yet been tested in the atomic structure and density map fitting problem. We also tried to rank the results from the Pareto set using a clustering algorithm,[40] but the performance was not as good as the knee-based ranking. Note that the knee algorithm is used to choose an output model from the larger set of a Pareto solutions, and it is independent of the multiobjective optimization main iterations.

After finding the Pareto solutions, each Pareto solution is projected into a three-dimensional score space with score axes *E1*, *E2*, and *E3* as defined in eqs 2−4. Knee points are usually those solutions for which a small deprivation in one objective will simultaneously cause considerable improvement in other objectives. The margin utility method[41] is used to search for more essential knee solutions in the Pareto front. The specifics are described as follows

$$U_{x,\lambda} = \lambda_1 f_1(x) + \lambda_2 f_2(x) + \lambda_3 f_3(x)$$
$$s.t. \quad \lambda_1 + \lambda_2 + \lambda_3 = 1 \text{ and } \lambda_1,\ \lambda_2,\ \lambda_3 > 0 \tag{7}$$

where x is the nondominated solution in the Pareto set, and λ_1, λ_2, and λ_3 are the weight variants. The knee algorithm is used to rank the Pareto solutions and select the top conformation as the final output model. It uses the value of the three objective functions used for multiobjective optimization to rank the final Pareto set solution. The expected marginal utilities can be simply approximated by calculating the marginal utility for all individuals. Several utility functions are generated by randomly setting λ_1, λ_2, and λ_3 to random values. The average utility is expected to be the marginal utility for each nondominated solution. The conformation with the largest marginal utility is taken as the final output.

RESULTS

Benchmark Data Set. To test the performance of MOFIT, a nonredundant set of 251 single-domain proteins was collected randomly from the protein data bank (PDB) library with a pairwise sequence identity less than 30%. In addition, we randomly selected 41 domains from the CASP12−13 experiments.[42] Overall, the test data set contain 292 proteins, including 48 α-, 36 β-, and 208 $\alpha\beta$-proteins, following the structural classification of proteins (SCOP) categorization[43] (see list in Table S1). Simulated noise-free density maps were generated from the target structures using EMAN2,[44] with resolutions nearly evenly distributed in the 5−10 Å range as shown in Figure S1 in the Supporting Information. One reason for choosing this resolution range is that this is the range at which atomic structure predictors are generally needed, as high-resolution structure determination is not possible from the density maps alone.[45] Table S1 in the Supporting Information lists the specific resolution of the density map for each protein, where the resolution for each protein was randomly selected from the 5−10 Å integer range. We also randomly added Gaussian noise with a standard deviation of 0.01 and an average value of 1 into the simulated density maps using Xmipp.[46,47] The cross-correlation between the noise-free and noisy maps was equal to 0.8, which was the value that was determined to be the proper setting for mimicking experimental density maps.[48]

To assess the effectiveness of our method, we calculated the coordinate root-mean-square deviations (RMSDs), cRMSD= $\sqrt{\sum_{i=1}^{L} |\vec{x}_i - \vec{y}_i|^2 / L}$, of the superposed models, where L is the protein length, $\vec{x}$ represents the coordinates of the I-TASSER model superposed onto the target structure using the TM-score program,[49] and $\vec{y}$ represents the coordinates of the model fitted into the EM density map by MOFIT. Compared to RMSD and template modeling (TM)-score, which measure the similarity between a fitted model and the target structure directly, the major advantage of using cRMSD is that cRMSD is not affected by the inherent quality of the I-TASSER models themselves, since cRMSD measures the distance between the modeled position and the best position for the same I-TASSER model obtained by TM-score superposition with the known native. The value of cRMSD falls in a range from 0 to infinity, where a cRMSD of 0 indicates a perfect match between the TM-score superposed structure and the fitted structure obtained by each atomic structure-density map superposition method. If a model that is identical with the experimental (native) structure is used for fitting, the TM-score program should perfectly superpose the model to the experimental structure. Therefore, for these cases, the cRMSD will be equal to the actual RMSD of the fitted model relative to the target structure.

Control Methods. As a control, we compared MOFIT with three state-of-the-art density map fitting programs, including PowerFit (http://www.bonvinlab.org/education/powerfit/),[29] Situs (https://situs.biomachina.org/fguide.html),[22] and ADP_EM (https://chaconlab.org/hybrid4em/adp-em/adpem-user).[26] These programs represent a set of density map fitting methods that are either widely used in the community or newly developed. Meanwhile, they provide standalone packages, which are downloadable and very helpful for us to run them on our own benchmark proteins. A rigid-body fitting experiment was performed using all control methods starting from the same set of cryo-EM density maps and structures predicted by I-TASSER. For PowerFit, a rotational sampling interval was defined to restrict a sampling space with the default value (-a 10). For ADP_EM, as recommended by the user manual, the bandwidth for harmonic transformation was set to 32, and the density threshold value for the experimental map was set to 0.06. For Situs, the angular granularity was set to 8 (-deg 8), and the density map cutoff value was set to 0.0. All other parameters for the programs were set to the default values.

Benchmark Results. A reasonable model fitting method should be capable of fitting most of the native structures into their cryo-EM density maps. Therefore, we first applied MOFIT to fit experimentally determined structures into density maps. As shown in Figure S2, MOFIT fit the experimental structures into their density maps achieving cRMSDs < 0.5 Å, in the (0.5 Å, 1 Å) range, in the (1 Å, 1.5 Å) range, and greater than 1.5 Å for 164, 109, 18, and 1 case(s), respectively. The average cRMSD of our method (0.56 Å) was higher than that of Situs (0.12 Å), while it was lower than that of both ADP_EM (0.62 Å) and PowerFit (1.64 Å). One reason for MOFIT being outperformed by Situs is that MOFIT was optimized on predicted structure models, while Situs was parametrized on a native structure. Thus, Situs is able to perform better on the native structures, but its performance is worse than MOFIT for fitting predicted structures into density maps. However, MOFIT was able to outperform ADP_EM and PowerFit on both experimental and predicted structures using its effective heuristic searching and the multiobjective strategy.

https://dx.doi.org/10.1021/acs.jpcb.0c09903
J. Phys. Chem. B 2021, 125, 528−538

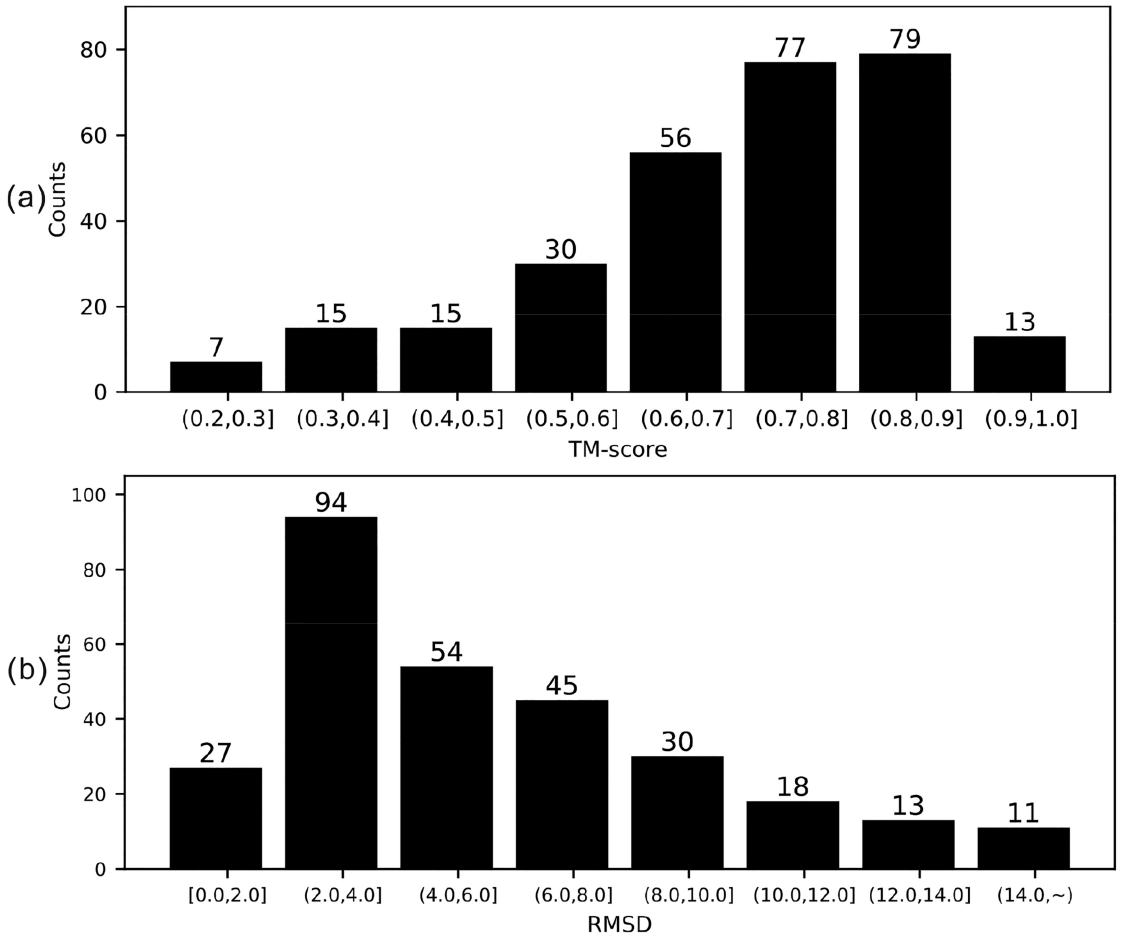


Figure 2. Histogram distribution for the TM-scores (a) and RMSDs (b) of the I-TASSER models for the 292 benchmark test proteins.

Next, we tested MOFIT on more challenging cases whose structures were predicted by I-TASSER. For each target, all homologous templates with a sequence identity greater than 30% to the query sequence were excluded from the I-TASSER template library. As a result, the average TM-score of the predicted structural models was 0.698. There were 255 of 292 targets with correct folds, that is, TM-scores greater than 0.5.[50] A more detailed TM-score and RMSD distribution of the I-TASSER models is displayed in Figure 2.

In Table 1 (columns 2 and 3), we present a summary of fitted models by MOFIT and three control methods on the noise-free density maps. On average, MOFIT achieved a cRMSD of 2.46 Å, which was 0.99 Å lower than that by ADP_EM (3.45 Å), 1.12 Å lower than that by Situs (3.58 Å), and 1.97 Å lower than that by PowerFit (4.43 Å). The differences were also statistically significant, as the p-values were 3.51×10^{-3}, 6.35×10^{-4}, and 7.92×10^{-9} for the comparison with ADP_EM, Situs, and PowerFit, respectively, as determined by a two-tailed Student's t test between MOFIT and the control methods. In Figure 3, we also list a head-to-head comparison of the cRMSD results, where MOFIT outperformed ADP_EM, Situs, and PowerFit on 64.4%, 55.5%, and 89.7% of the cases, respectively, as determined by the lower cRMSD values of the MOFIT-fitted models.

Table 1. Summary of Model Fitting Results on the 292 Test Proteins

	noise-free density map		noisy density map	
methods	cRMSD[a] (Å) (n_b[b])	p-value[c]	cRMSD[a] (Å) (n_b[b])	p-value[c]
ADP_EM	3.45 (188)	3.51×10^{-3}	5.14 (200)	1.19×10^{-4}
Situs	3.58 (162)	6.35×10^{-4}	3.75 (148)	1.11×10^{-3}
PowerFit	4.43 (262)	7.92×10^{-9}	6.54 (270)	6.27×10^{-13}
MOFIT	2.46		2.57	

[a]cRMSD: RMSD between fitted model and the best-fit position obtained by TM-score superposed. [b]n_b: Number of cases for which the cRMSD of MOFIT was lower than that of the control method. [c]p-Value: between cRMSDs obtained by MOFIT and the control methods using two-tailed Student's t tests.

The improvements made by MOFIT are mainly due to the design of the parallel score landscape of MOFIT, which is driven by global, fragment, and residue-level CC scores between the structure and the density map. The utilization of multiple objective functions balances the global, fragment, and residue-level similarity between the predicted structures and density maps, which is important for fitting models when different target and density map combinations are involved. To further evaluate the stability of our method, we list in Table 1 (columns 4 and 5) the model fitting results on density maps with simulated noise added. Although the noisy data resulted in reduced performance for all methods (which is expected), the average cRMSD of MOFIT increased only slightly by 0.11 Å (from 2.46 to 2.57 Å), which was lower than all three control methods, that is, 1.69 Å for ADP_EM, 0.17 Å for Situs, and 2.11 Å for PowerFit. The overall cRMSD of MOFIT was also

https://dx.doi.org/10.1021/acs.jpcb.0c09903
J. Phys. Chem. B 2021, 125, 528−538

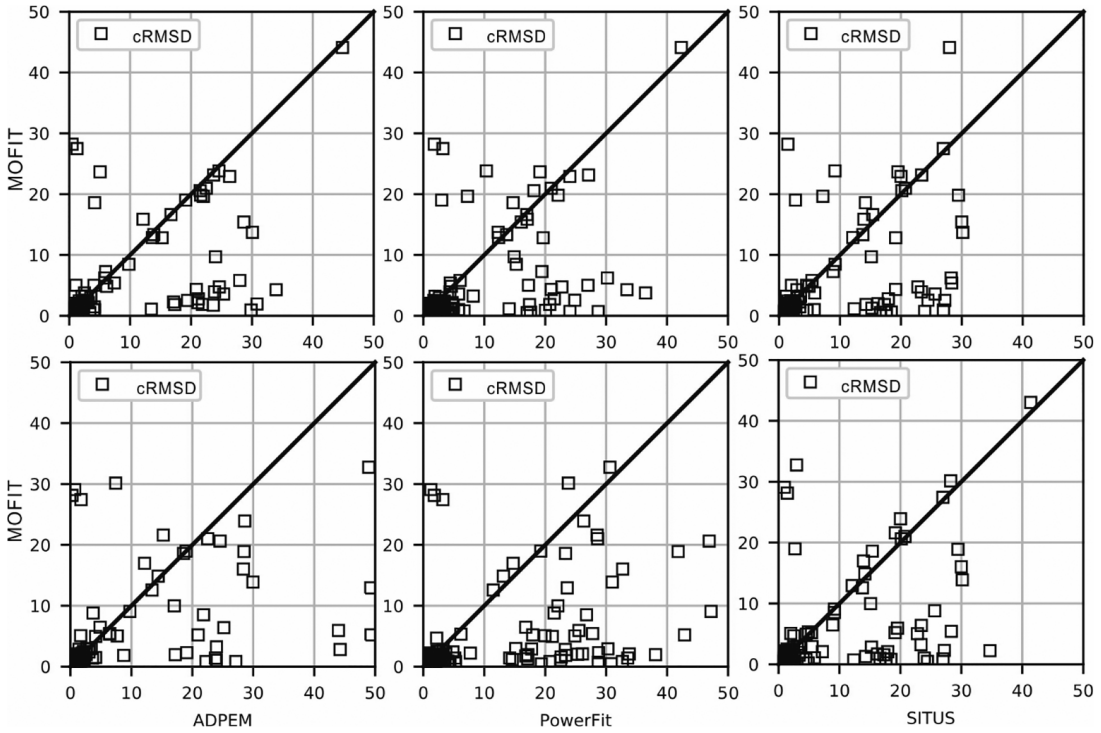


Figure 3. Head-to-head cRMSD comparison among different methods on the 292 noise-free density maps (upper row) and 292 noisy density maps (lower row). (left) MOFIT vs ADP_EM; (middle) MOFIT vs PowerFit; (right) MOFIT vs Situs.

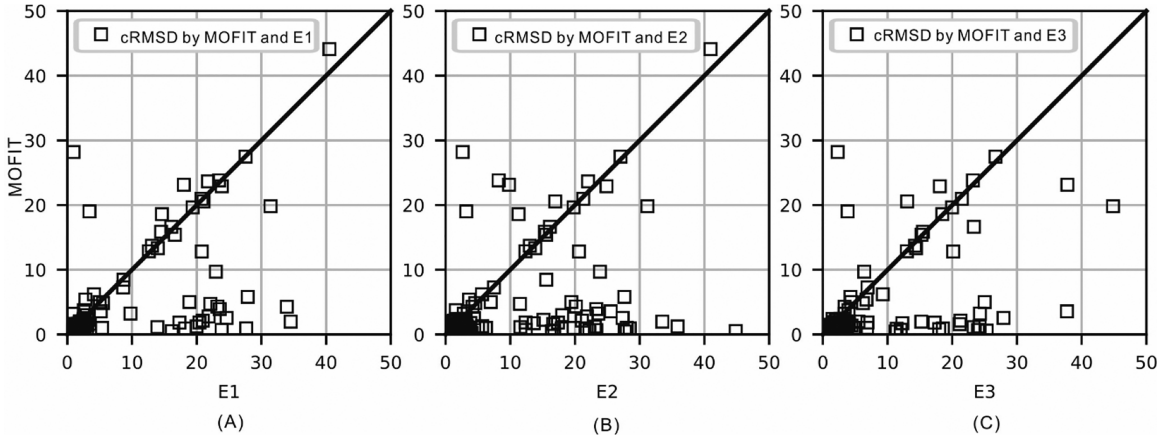


Figure 4. Head-to-head comparison between each single objective function (x-axis) and the overall multiobjective function (y-axis) for the 292 test proteins. Each block represents the cRMSD value on the corresponding axis.

significantly lower than the control methods with p-values of 1.19×10^{-4}, 1.11×10^{-3}, and 6.27×10^{-13} for the comparison with ADP_EM, Situs, and PowerFit, respectively. Of the 292 test targets, MOFIT achieved better cRMSDs than ADP_EM, Situs, and PowerFit for 68.5%, 50.7%, and 92.4% of the cases, respectively (Figure 3).

To examine the relationship between model quality and the meaningfulness of structure-density fitting, in Figure S3, we present the correlation coefficient (CC) of fitted models to the EM density map versus TM-score of the initial fitted models for the 292 test proteins. The results show that, when the initial model has an incorrect fold (i.e., TM-score < 0.5), most of the fits have a low CC and there is nearly no correlation between CC and the TM-score (Pearson correlation coefficient = 0.18). However, for models with a correct fold (TM-score > 0.5), there is a strong correlation between CC and the TM-score (Pearson correlation coefficient = 0.74), showing such fitting problems are still meaningful even when the I-TASSER models are not perfect but they have the correct fold. Fortunately, as shown in Figure 2, the majority (87%) of the I-TASSER models for the test set had a TM-score greater than 0.5, which is part of the reason why we mainly benchmarked our method on the real-world case where the experimental structure is unknown but predicted structural models may be used, such as those constructed by I-TASSER.

In Table S2, we provide a summary of the modeling results by I-TASSER and the follow-up EM density map fitting methods on an additional set of seven randomly selected proteins from the EMDR (Electron Microscopy Data Resource) that have experimentally determined density maps. For these seven proteins, the average cRMSD of MOFIT was 1.49 Å, compared to 2.85 Å for ADP_EM, 2.18 Å for Situs, and

https://dx.doi.org/10.1021/acs.jpcb.0c09903
J. Phys. Chem. B 2021, 125, 528−538

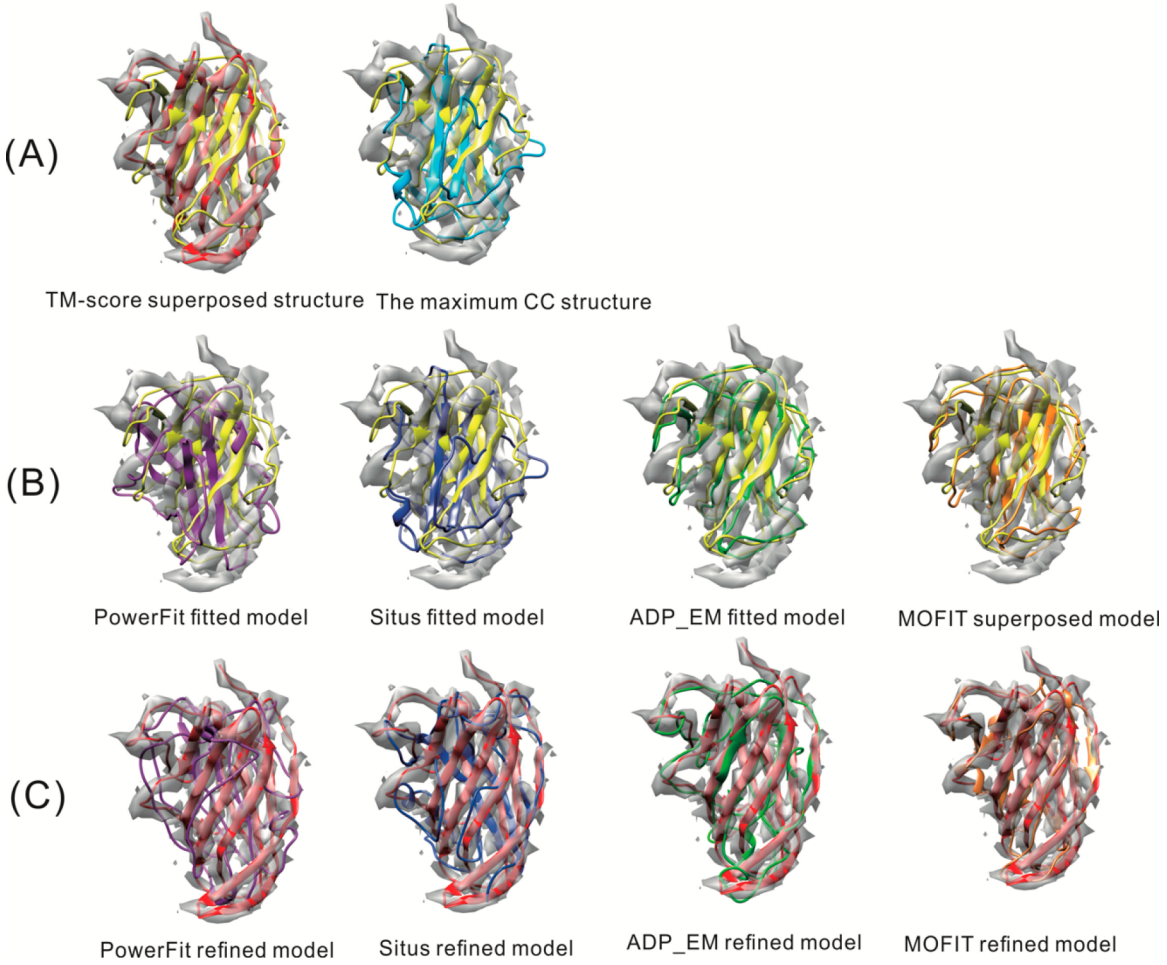


Figure 5. Illustrative examples of fitting for the C-terminal domain of the MADV2 fiber from mouse adenovirus 2 (ID: 5NC1). (A) The maximum CC structure (the maximum CC score structure, cyan) with a CC of 0.383 and a cRMSD of 25.33 Å and the TM-score superposed model (the best superposed structure, yellow) with a CC of 0.357 and a cRMSD of 0; native model (red). (B) Different colors represent different fitting results for all methods. PowerFit fitted model (purple, cRMSD = 20.89 Å), Situs fitted model (blue, cRMSD = 25.33 Å), ADP_EM fitted model (green, cRMSD = 2.89 Å), MOFIT fitted model (oranges, cRMSD = 2.20 Å), and TM-score superposed model (yellow). (C) Different colors represent different refined models using EM-Refiner. PowerFit refined model with an RMSD of 12.11 Å and a TM-score of 0.460 (purple), Situs refined model with an RMSD of 11.67 Å and a TM-score of 0.465 (blue), ADP_EM refined model with an RMSD of 11.69 Å and a TM-score of 0.503 (green), MOFIT refined model with an RMSD of 10.60 Å and a TM-score of 0.504 (orange) and native structure (red).

3.27 Å for PowerFit, respectively. These results show the robustness of the MOFIT performance on both the simulated (with/without noise) and experimental density maps.

In MOFIT, the knee algorithm is used to select the final fitted model after obtaining the Pareto set. We also compared the results obtained using a clustering algorithm[40] and the knee algorithm[33] on the 292 cases. For the clustering algorithm, the model closest to the centroid of the largest cluster was the final output model. For the knee algorithm, we selected the model with the best score as the final output model. On average, the cRMSD by the knee algorithm was 2.46 Å, which was slightly better than that obtained by the clustering algorithm (2.56 Å). In Figure S4, we present the three-objective distributions of the nondominated particles (PDBID: 5NC1) in the Pareto set, where each point indicates a nondominated solution. After the fitting iterations terminated, MOFIT obtained the non-dominated particles of the Pareto set in score space, where the final output structure was selected by the knee algorithm. For the case of 5NC1, the knee point had the largest marginal utility of all nondominated solutions as shown in Figure S4.

Comparison between Single-Objective and Multi-objective Function Optimization. To examine the impact of using multiple objective functions, we present in Figure 4 a head-to-head comparison between the fitting performance obtained by each single-objective function and the overall multiobjective function used by MOFIT. The subfigures A, B, and C of Figure 4 show the cRMSD distribution for the 292 targets using the functions E1 (global CC), E2 (fragment-level CC), and E3 (residue-level CC) versus MOFIT. Overall, there were 261, 259, and 268 of the 292 cases for which MOFIT created a fitted model with a lower cRMSD than the single-objective functions E1, E2, and E3, respectively. As shown in Table S3, the average cRMSD by MOFIT was 2.46 Å, which was 1.57 Å lower than obtained by E1 (4.03 Å), 2.80 Å lower than that by E2 (5.26 Å), and 3.95 Å lower than that by E3 (6.41 Å); this corresponds to the p-values of 1.65×10^{-6}, 6.36×10^{-8}, and 6.44×10^{-11}, respectively, as determined by two-tailed Student's t tests, thereby demonstrating that the improvement achieved by the multiobjective function is statistically significant.

https://dx.doi.org/10.1021/acs.jpcb.0c09903
J. Phys. Chem. B 2021, 125, 528−538

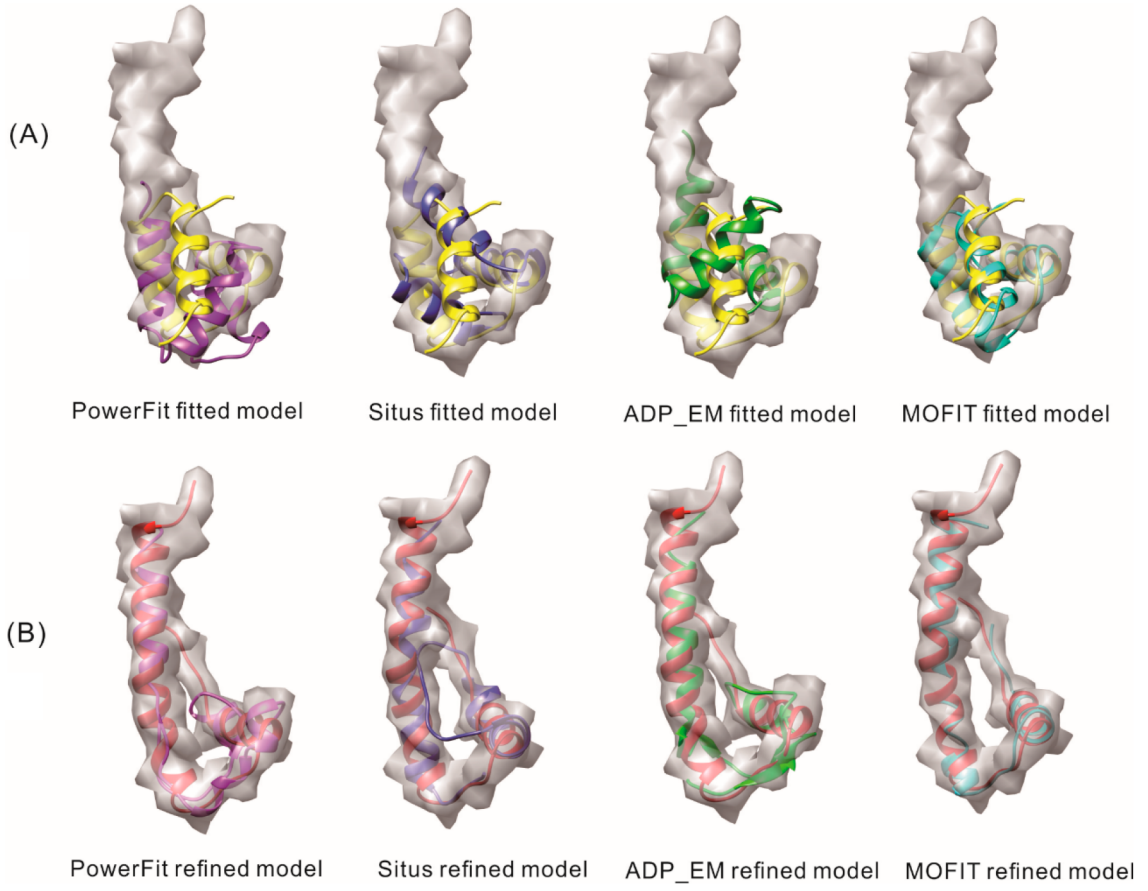


Figure 6. Illustrative examples of superposition affecting the refinement for the single domain of 3a1iA using a simulated density map with a 5 Å resolution. (A) Different colors represent different fitting results. PowerFit fitted model with a cRMSD of 17.06 Å (purple), Situs fitted model with a cRMSD of 17.04 Å (blue), ADP_EM fitted model with a cRMSD of 17.04 Å (green), MOFIT fitted model with a cRMSD of 2.75 Å (cyan), and the TM-score superposed model (yellow). (B) Different colors represent different refined models using EM-Refiner. PowerFit refined model with an RMSD of 10.19 Å (purple), Situs refined model with an RMSD of 11.72 Å (blue), ADP_EM refined model with an RMSD of 10.19 Å (green), MOFIT refined model with an RMSD of 6.34 Å (cyan), and native model (red).

In Figure S5, we present a head-to-head comparison between each individual objective function in order to determine which single objective function gives the best results. We found that no single objective function could achieve a dominantly lower cRMSD than any other function, and different single scoring functions had different preferences among all 292 test cases. Further, Figure S6 shows the value of each objective function at different PSO searching iterations for a selected protein (PDBID: 1efdN). From the figure, it can be seen that E1, E2, and E3 display asynchronous behavior, suggesting a conflict between E1, E2, and E3 during the searching iterations. Multiobjective optimization will search the diverse conformations where optimal decisions need to be balanced in the presence of the trade-offs between two or more different objectives. This example case illustrates that E2 and E3 are important to give favorable cRMSD values and that all three single objective functions converge by the end of the simulation. Since models with incorrect folds (like those with disordered and misaligned regions, among other complicating factors) interfere with the optimization of objective function 1, which accounts for the global correlation, the addition of local correlation functions such as E2 and E3 can improve the fitting performance by focusing on specific and complementary correlation regions. Multiobjective optimization enables us to utilize the complementarity of multiple objective functions to realize the distinct point-of-view evaluation of the object.

Case Studies. The overall fitting results indicate that the correct folding of a predicted structure is important to obtain the correct superposition conformations. Figure 5 presents such an example for the C-terminal domain of the MAdV2 fiber from the mouse adenovirus 2 (PDBID: 5NC1) (CASP target ID: T0880) to illustrate the results of different methods on a challenging target. The resolution of the density map for this target was 5 Å, but I-TASSER generated an initial structure with an incorrect fold (TM-score = 0.47) compared with the native structure. The CC between the superposed model and the density map was 0.357 based on the best superposition obtained by TM-score superposing the predicted structure onto the native structure.

However, the pose corresponding to the maximum CC score was not the pose with the minimal cRMSD when we used the single CC objective function as the score function for the fitting simulation. This is particularly true when the I-TASSER model is low quality with an incorrect fold. In Figure 5A, for example, the maximum CC score structure (cyan) resulted in a conformation with a cRMSD of 25.33 Å. Given that the correlation between the model and the density map using the single global correlation function did not reflect the quality of

https://dx.doi.org/10.1021/acs.jpcb.0c09903
J. Phys. Chem. B 2021, 125, 528−538

the superposition, a single-objective function is most likely not the best choice for this case and other similar ones.

Figure 5B compares the results of MOFIT (multiobjective optimization) and the three control methods that all utilize a single-objective optimization. After the superposition, MOFIT achieved a cRMSD of 2.20 Å, which was dramatically lower than that of PowerFit (20.89 Å) and Situs (25.33 Å), and slightly lower than that of ADP_EM (2.89 Å). These cases correspond to the points in Figure 3 that are far away and near the diagonal lines in Figure 3, respectively. To further illustrate the impact of density map fitting on the final protein structure model prediction, we showed in Figure 5C the full-length models constructed by EM-Refiner (https://zhanglab.ccmb.med.umich.edu/EM-Refiner),[16] starting from the initial superpositions obtained by the different density map fitting programs. Here, EM-Refiner is a recently developed cryo-EM based structure refinement algorithm, which constructs final full-length protein structure models using replica-exchange Monte Carlo (REMC) simulations. Although EM-Refiner has its own density map fitting algorithm, it is completely different than MOFIT, because EM-Refiner uses a single CC score searched by a short REMC simulation, while MOFIT uses multiple optimization functions (combining local, fragment-level, and global CC scores) and searches with multiobjective particle swarm optimization to search the conformational space.

As shown in Figure 5C, the refined model produced by EM-Refiner starting from the MOFIT superposition had a TM-score of 0.504 (RMSD = 10.60 Å), compared to a TM-score of 0.460 (RMSD = 12.11 Å) starting from the PowerFit superposition, a TM-score of 0.465 (RMSD= 11.67 Å) from the Situs superposition, and a TM-score of 0.503 (RMSD = 11.69 Å) from the ADP_EM superposition. These results indicate that the higher-quality superposition achieved by MOFIT improved the density map-constrained refinement results, where our method outperformed PowerFit and Situs at structure superposition for this example. In fact, on the one hand, the refined structures starting from the PowerFit and Situs initial superpositions had worse TM-scores than the initial model due to the poor structure-map superposition. On the other hand, both MOFIT and ADP_EM were able to produce high enough quality initial superpositions to draw the predicted structural model closer to the native structure, resulting in a correctly folded model with a TM-score greater than 0.5. This result demonstrates the importance of utilizing multiple objective functions, which can help generate higher-quality fitted models, thus improving the model quality for the cryo-EM-based protein structure prediction.

Figure 6 shows another example from PDB ID: 3a1iA, for which I-TASSER generated a correct fold with a TM-score of 0.57, where the resolution of the simulated density map was 5 Å. As shown in Figure 6A, MOFIT generated a fitted model with a cRMSD of 2.75 Å, which was significantly lower than that obtained by PowerFit, Situs, and ADP_EM with cRMSD values of 17.06 Å, 17.04 Å, and 17.04 Å, respectively. Starting from the MOFIT fitted model, EM-Refiner constructed a full-length model with a TM-score of 0.59 (RMSD = 6.34 Å), which was significantly closer to the native than those generated when starting from PowerFit (TM-score = 0.26, RMSD = 11.19 Å), Situs (TM-score = 0.39, RMSD = 11.72 Å), or ADP_EM (TM-score = 0.32, RMSD = 10.19 Å) superposition, as shown in Figure 6B. This result further illustrates the importance of density-map fitting methods. Even if the initial predicted structures have a correct fold, different methods will produce different density map superpositions, which may result in dramatically different final structure model quality.

CONCLUSION

In this work, we proposed a novel rigid-body fitting method, MOFIT, which performs multiobjective particle swarm optimization (MOPSO) to fit predicted structures into 3D cryo-EM density maps, which is an essential procedure for cryo-EM-based protein structure determination. In our protocol, three objective functions were used during the MOPSO process, which enable the method to achieve a better fitting performance via the introduction of additional scoring criteria. When fitting a predicted structure into a density map, we considered both the global and local cross-correlations, which account for full model-level, fragment-level, and residue-level correlation scores. MOPSO allows for the presence of complementary trade-offs among multiple objectives to obtain optimal fitting solutions,[36] where the best solutions are achieved from the nondominated Pareto set. The global correlation accounts for the preference of the fit for the overall protein shape, while the local correlation scores assess the fitting results in accordance with the correct local topology. Moreover, our algorithm samples a wide region of conformational space by starting from diverse initial models with different poses. The benchmark results on 292 nonhomologous proteins showed that MOFIT generated fitted models significantly closer to the optimal positions compared to those of other state-of-art programs, including ADP_EM, Situs, and PowerFit, as well as the internal programs based on each individual MOFIT objective functions. This demonstrates the effectiveness of the multiobjective functions and the MOPSO search protocol.

The density map fitting experiments showed that the final model accuracy depends on both model quality and the density map resolution. In Figure S7, we present the correlation between the cRMSD and TM-score of each I-TASSER model at different density map resolutions. We found that the Pearson correlation coefficient between the cRMSD and TM-score was −0.61, −0.52, and −0.50 for 5−6, 7−8, and 9−10 Å maps, respectively. The average cRMSDs were 0.81, 0.97, and 1.13 Å for 5−6, 7−8, and 9−10 Å resolution density maps, where the TM-score of the atomic structures used for fitting were more than 0.5. These data further confirmed that the major challenge to the fitting experiment mainly comes from the cases with low-quality structure prediction and low-resolution density maps. Given the efficiency of multiobjective optimization, in the future, we will continue to extend the current MOPSO protocol to cryo-EM-based structure fitting and refinement with a primary focus on the targets with low TM-score structure prediction and low-resolution density maps.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.0c09903.

The resolution of the simulated density maps, experimental density maps information, statistical information for the benchmark data set, additional analysis results (PDF)

https://dx.doi.org/10.1021/acs.jpcb.0c09903
J. Phys. Chem. B 2021, 125, 528−538

AUTHOR INFORMATION

Corresponding Authors

Hong-Bin Shen − *Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai, China;* orcid.org/0000-0002-4029-3325; Email: hbshen@sjtu.edu.cn

Yang Zhang − *Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 48109, United States;* orcid.org/0000-0002-2739-1916; Phone: (734) 647-1549; Email: zhng@umich.edu; Fax: (734) 615-6553

Authors

Biao Zhang − *Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai, China; Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 48109, United States*

Wenyi Zhang − *School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310024, China*

Robin Pearce − *Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 48109, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.0c09903

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (Nos. 62073219, 61725302, and 61671288), the Science and Technology Commission of Shanghai Municipality (No. 17JC1403500), the National Institute of General Medical Sciences (GM083107 and GM136422), the National Institute of Allergy and Infectious Diseases (AI134678), and the National Science Foundation (DBI1564756, IIS1901191).

REFERENCES

(1) Nguyen, N. X.; Armache, J.-P.; Lee, C.; Yang, Y.; Zeng, W.; Mootha, V. K.; Cheng, Y.; Bai, X.-c.; Jiang, Y. Cryo-Em Structure of a Fungal Mitochondrial Calcium Uniporter. *Nature* **2018**, *559*, 570−574.

(2) Dang, S.; Feng, S.; Tien, J.; Peters, C. J.; Bulkley, D.; Lolicato, M.; Zhao, J.; Zuberbühler, K.; Ye, W.; Qi, L.; et al. Cryo-Em Structures of the Tmem16a Calcium-Activated Chloride Channel. *Nature* **2017**, *552*, 426−429.

(3) Chen, Z.; Sun, L.; Zhang, Z.; Fokine, A.; Padilla-Sanchez, V.; Hanein, D.; Jiang, W.; Rossmann, M. G.; Rao, V. B. Cryo-Em Structure of the Bacteriophage T4 Isometric Head at 3.3-Å Resolution and Its Relevance to the Assembly of Icosahedral Viruses. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E8184−E8193.

(4) Yuan, S.; Wang, J.; Zhu, D.; Wang, N.; Gao, Q.; Chen, W.; Tang, H.; Wang, J.; Zhang, X.; Liu, H. Cryo-Em Structure of a Herpesvirus Capsid at 3.1 Å. *Science* **2018**, *360*, eaao7283.

(5) Voss, J. E.; Vaney, M.-C.; Duquerroy, S.; Vonrhein, C.; Girard-Blanc, C.; Crublet, E.; Thompson, A.; Bricogne, G.; Rey, F. A. Glycoprotein Organization of Chikungunya Virus Particles Revealed by X-Ray Crystallography. *Nature* **2010**, *468*, 709−712.

(6) Pilla, K. B.; Otting, G.; Huber, T. 3D Computational Modeling of Proteins Using Sparse Paramagnetic NMR Data. In *Bioinformatics;* Springer, 2017; pp 3−21.

(7) Wriggers, W. R.; Galkin, V. E.; Hunter, W. A.; Kovacs, J. A. Damped Dynamics as a Validation Platform for the Flexible Refinement of Atomic Models against Cryo-Em Maps. *Biophys. J.* **2020**, *118*, No. 292a.

(8) de Groot, B. L.; Engel, A.; Grubmüller, H. The Structure of the Aquaporin-1 Water Channel: A Comparison between Cryo-Electron Microscopy and X-Ray Crystallography. *J. Mol. Biol.* **2003**, *325*, 485−493.

(9) Oling, F.; Sopkova-de Oliveira Santos, J.; Govorukhina, N.; Mazères-Dubut, C.; Bergsma-Schutter, W.; Oostergetel, G.; Keegstra, W.; Lambert, O.; Lewit-Bentley, A.; Brisson, A. Structure of Membrane-Bound Annexin A5 Trimers: A Hybrid Cryo-Em-X-Ray Crystallography Study. *J. Mol. Biol.* **2000**, *304*, 561−573.

(10) Patwardhan, A. Trends in the Electron Microscopy Data Bank (Emdb). *Acta Crystallographica Section D: Structural Biology* **2017**, *73*, 503−508.

(11) Liang, Y.-L.; Khoshouei, M.; Deganutti, G.; Glukhova, A.; Koole, C.; Peat, T. S.; Radjainia, M.; Plitzko, J. M.; Baumeister, W.; Miller, L. J.; et al. Cryo-Em Structure of the Active, G S-Protein Complexed, Human Cgrp Receptor. *Nature* **2018**, *561*, 492−497.

(12) Perilla, J. R.; Zhao, G.; Lu, M.; Ning, J.; Hou, G.; Byeon, I.-J. L.; Gronenborn, A. M.; Polenova, T.; Zhang, P. Cryoem Structure Refinement by Integrating Nmr Chemical Shifts with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2017**, *121*, 3853−3863.

(13) 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*, 361−365.

(14) Joseph, A. P.; Malhotra, S.; Burnley, T.; Wood, C.; Clare, D. K.; Winn, M.; Topf, M. Refinement of Atomic Models in High Resolution Em Reconstructions Using Flex-Em and Local Assessment. *Methods* **2016**, *100*, 42−49.

(15) Yan, S.; Guo, C.; Hou, G.; Zhang, H.; Lu, X.; Williams, J. C.; Polenova, T. Atomic-Resolution Structure of the Cap-Gly Domain of Dynactin on Polymeric Microtubules Determined by Magic Angle Spinning Nmr Spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 14611−14616.

(16) Zhang, B.; Zhang, X.; Pearce, R.; Shen, H.-B.; Zhang, Y. A New Protocol for Atomic-Level Protein Structure Modeling and Refinement Using Low-to-Medium Resolution Cryo-Em Density Maps. *J. Mol. Biol.* **2020**, *432*, 5365−5377.

(17) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-Tasser Suite: Protein Structure and Function Prediction. *Nat. Methods* **2015**, *12*, 7.

(18) Wolf, S. G.; Rez, P.; Elbaum, M. Phosphorus Detection in Vitrified Bacteria by Cryo-Stem Annular Dark-Field Analysis. *J. Microsc.* **2015**, *260*, 227−233.

(19) Chebotareva, N.; Bomans, P. H.; De Haas, F.; Hubert, D.; Frederik, P.; Sijbesma, R. P.; Sommerdijk, N. The Three-Dimensional Constitution of Micelle Forming Surfactants as Studied by Cryo Electron Microscopy and Tomography. *Microsc. Microanal.* **2006**, *12*, 1544−1545.

(20) Baker, M. L.; Zhang, J.; Ludtke, S. J.; Chiu, W. Cryo-Em of Macromolecular Assemblies at near-Atomic Resolution. *Nat. Protoc.* **2010**, *5*, 1697.

(21) Rossmann, M. G.; Bernal, R.; Pletnev, S. V. Combining Electron Microscopic with X-Ray Crystallographic Structures. *J. Struct. Biol.* **2001**, *136*, 190−200.

(22) Wriggers, W. J. B. r. Using Situs for the Integration of Multi-Resolution Structures. *Biophys. Rev.* **2010**, *2*, 21−27.

(23) Chacón, P.; Wriggers, W. J. J. o. m. b. Multi-Resolution Contour-Based Fitting of Macromolecular Structures. *J. Mol. Biol.* **2002**, *317*, 375−384.

(24) Ceulemans, H.; Russell, R. B. Fast Fitting of Atomic Structures to Low-Resolution Electron Density Maps by Surface Overlap Maximization. *J. Mol. Biol.* **2004**, *338*, 783−793.
(25) Lasker, K.; Topf, M.; Sali, A.; Wolfson, H. J. Inferential Optimization for Simultaneous Fitting of Multiple Components into a Cryoem Map of Their Assembly. *J. Mol. Biol.* **2009**, *388*, 180−194.
(26) Garzón, J. I.; Kovacs, J.; Abagyan, R.; Chacón, P. J. B. Adp_Em: Fast Exhaustive Multi-Resolution Docking for High-Throughput Coverage. *Bioinformatics* **2007**, *23*, 427−433.
(27) de Vries, S. J.; Zacharias, M. J. P. o. Attract-Em: A New Method for the Computational Assembly of Large Molecular Machines Using Cryo-Em Maps. *PLoS One* **2012**, *7*, e49733.
(28) Lasker, K.; Dror, O.; Shatsky, M.; Nussinov, R.; Wolfson, H. J. bioinformatics, Ematch: Discovery of High Resolution Structural Homologues of Protein Domains in Intermediate Resolution Cryo-Em Maps. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2007**, *4*, 28−39.
(29) CP van Zundert, G.; MJJ Bonvin, A. Fast and Sensitive Rigid-Body Fitting into Cryo-Em Density Maps with Powerfit. *AIMS Biophysics* **2015**, *2*, 73−87.
(30) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. Ucsf Chimera—a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(31) Lasker, K.; Sali, A.; Wolfson, H. J. Determining Macromolecular Assembly Structures by Molecular Docking and Fitting into an Electron Density Map. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 3205−3211.
(32) Weir, T.; Mond, B. Pre-Invex Functions in Multiple Objective Optimization. *Journal of Mathematical Analysis and Applications* **1988**, *136*, 29−38.
(33) Zhang, X.; Tian, Y.; Jin, Y. A Knee Point-Driven Evolutionary Algorithm for Many-Objective Optimization. *IEEE Transactions on Evolutionary Computation* **2015**, *19*, 761−776.
(34) DiMaio, F.; Tyka, M. D.; Baker, M. L.; Chiu, W.; Baker, D. J. J. o. m. b. Refinement of Protein Structures into Low-Resolution Density Maps Using Rosetta. *J. Mol. Biol.* **2009**, *392*, 181−190.
(35) Leung, M.-F.; Ng, S.-C.; Cheung, C.-C.; Lui, A. K. A New Strategy for Finding Good Local Guides in Mopso. In *2014 IEEE Congress on Evolutionary Computation (CEC), IEEE*: 2014; pp 1990−1997.
(36) Wang, D.; Geng, L.; Zhao, Y.-J.; Yang, Y.; Huang, Y.; Zhang, Y.; Shen, H.-B. Artificial Intelligence-Based Multi-Objective Optimization Protocol for Protein Structure Refinement. *Bioinformatics* **2019**, *36*, 437−448.
(37) Coello, C. A. C.; Pulido, G. T.; Lechuga, M. S. Handling Multiple Objectives with Particle Swarm Optimization. *IEEE Transactions on evolutionary computation* **2004**, *8*, 256−279.
(38) Tripathi, P. K.; Bandyopadhyay, S.; Pal, S. K. Multi-Objective Particle Swarm Optimization with Time Variant Inertia and Acceleration Coefficients. *Inf. Sci.* **2007**, *177*, 5033−5049.
(39) Parsopoulos, K. E.; Vrahatis, M. N. Particle Swarm Optimization Method in Multiobjective Problems. In *Proceedings of the 2002 ACM symposium on Applied computing, Proceedings of the 2002 ACM symposium on Applied computing*, 2002; pp 603−607.
(40) Zhang, Y.; Skolnick, J. J. J. o. c. c. Spicker: A Clustering Approach to Identify near-Native Protein Folds. *J. Comput. Chem.* **2004**, *25*, 865−871.
(41) Branke, J.; Deb, K.; Dierolf, H.; Osswald, M. Finding Knees in Multi-Objective Optimization. In *International conference on parallel problem solving from nature;* Springer, 2004; pp 722−731.
(42) Xu, Y.; Xu, D.; Gabow, H. N. Protein Domain Decomposition Using a Graph-Theoretic Approach. *Bioinformatics* **2000**, *16*, 1091−1104.
(43) Lo Conte, L.; Ailey, B.; Hubbard, T. J.; Brenner, S. E.; Murzin, A. G.; Chothia, C. Scop: A Structural Classification of Proteins Database. *Nucleic acids research* **2000**, *28*, 257−259.
(44) Tang, G.; Peng, L.; Baldwin, P. R.; Mann, D. S.; Jiang, W.; Rees, I.; Ludtke, S. J. Eman2: An Extensible Image Processing Suite for Electron Microscopy. *J. Struct. Biol.* **2007**, *157*, 38−46.
(45) Cheng, Y. J. C. Single-Particle Cryo-Em at Crystallographic Resolution. *Cell* **2015**, *161*, 450−457.
(46) Sorzano, C.; Marabini, R.; Velázquez-Muriel, J.; Bilbao-Castro, J. R.; Scheres, S. H.; Carazo, J. M.; Pascual-Montano, A. Xmipp: A New Generation of an Open-Source Image Processing Package for Electron Microscopy. *J. Struct. Biol.* **2004**, *148*, 194−204.
(47) Lindert, S.; Alexander, N.; Wötzel, N.; Karakaş, M.; Stewart, P. L.; Meiler, J. Em-Fold: De Novo Atomic-Detail Protein Structure Determination from Medium Resolution Density Maps. *Structure* **2012**, *20*, 464−478.
(48) Woetzel, N.; Lindert, S.; Stewart, P. L.; Meiler, J. Bcl:: Em-Fit: Rigid Body Fitting of Atomic Structures into Density Maps Using Geometric Hashing and Real Space Refinement. *J. Struct. Biol.* **2011**, *175*, 264−276.
(49) Zhang, Y.; Skolnick, J. Scoring Function for Automated Assessment of Protein Structure Template Quality. *Proteins: Struct., Funct., Genet.* **2004**, *57*, 702−710.
(50) Xu, J.; Zhang, Y. How Significant Is a Protein Structure Similarity with Tm-Score = 0.5? *Bioinformatics* **2010**, *26*, 889−95.

https://dx.doi.org/10.1021/acs.jpcb.0c09903
J. Phys. Chem. B 2021, 125, 528−538