



Multi-Scale Flexible Fitting of Proteins to Cryo-EM Density Maps at Medium Resolution

Marta Kulik^{1†‡}, Takaharu Mori¹ and Yuji Sugita^{1,2,3*†}

¹Theoretical Molecular Science Laboratory, RIKEN Cluster for Pioneering Research, Wako-shi, Japan, ²RIKEN Center for Computational Science, Kobe, Japan, ³RIKEN Center for Biosystems Dynamics Research, Kobe, Japan

OPEN ACCESS

Edited by:

Peter J Bond,
Bioinformatics Institute (A*STAR),
Singapore

Reviewed by:

Yong Wang,
University of Copenhagen, Denmark
Simon Olsson,
Chalmers University of Technology,
Sweden

*Correspondence:

Yuji Sugita
sugita@riken.jp

†Present address:

Biological and Chemical Research
Center, Department of Chemistry,
University of Warsaw, ul. Żwirki i
Wigury 101, 02-089 Warsaw, Poland

†ORCID:

Marta Kulik
0000-0003-2381-7665
Yuji Sugita
0000-0001-9738-9216
Takaharu Mori
0000-0002-8717-2926

Specialty section:

This article was submitted to
Biological Modeling and Simulation,
a section of the journal
Frontiers in Molecular Biosciences

Received: 21 November 2020

Accepted: 26 January 2021

Published: 19 March 2021

Citation:

Kulik M, Mori T and Sugita Y (2021)
Multi-Scale Flexible Fitting of
Proteins to Cryo-EM Density Maps at
Medium Resolution.
Front. Mol. Biosci. 8:631854.
doi: 10.3389/fmolb.2021.631854

Structure determination using cryo-electron microscopy (cryo-EM) medium-resolution density maps is often facilitated by flexible fitting. Avoiding overfitting, adjusting force constants driving the structure to the density map, and emulating complex conformational transitions are major concerns in the fitting. To address them, we develop a new method based on a three-step multi-scale protocol. First, flexible fitting molecular dynamics (MD) simulations with coarse-grained structure-based force field and replica-exchange scheme between different force constants replicas are performed. Second, fitted C α atom positions guide the all-atom structure in targeted MD. Finally, the all-atom flexible fitting refinement in implicit solvent adjusts the positions of the side chains in the density map. Final models obtained via the multi-scale protocol are significantly better resolved and more reliable in comparison with long all-atom flexible fitting simulations. The protocol is useful for multi-domain systems with intricate structural transitions as it preserves the secondary structure of single domains.

Keywords: flexible fitting, multi-scale methods, replica exchange, molecular dynamics simulation, cryo-EM density map, all-atom force field, coarse-grained force field, targeted molecular dynamics

INTRODUCTION

The technical advances in single-particle cryo-EM result in a rapid escalation of the number of available density maps at increasingly higher resolutions recently crossing the 1.3 Å barrier (Yip et al., 2020) (Nakane et al., 2020). Also, the lower molecular weight limit of this method already reached much below 50 kDa (Wu and Lander, 2020). In order to extract structural information from the experimental densities, computational techniques such as rigid-docking (Wriggers et al., 1999) (Roseman, 2000), flexible fitting (Tama et al., 2004) (Topf et al., 2008) (Trabuco et al., 2008) (Whitford et al., 2011) (Dou et al., 2017) (Mori et al., 2019), and *de novo* modeling (Lindert et al., 2012) (Wang et al., 2015) (Terashi and Kihara, 2018) are essential to assign or refine the atomic positions into those maps. The choice of the computational method depends on the map resolution (Villa and Lasker, 2014). At near-atomic resolution (typically 3–5 Å), the structural details visible in the density give enough confidence for building the models *de novo*, even though the completeness and accuracy of such models would be lower than those based on atomic resolution density maps (Wang et al., 2015) (Zivanov et al., 2018). However, the density maps do not always reach so high resolutions due to intrinsic flexibility or fluctuation of biomolecules, and the methods other than *de novo* building are often required in such cases.

Flexible fitting has been widely used to guide the initial structure toward the target density map, in which normal mode analysis (Tama et al., 2004) or MD algorithms are utilized (Trabuco et al., 2008)

(Orzechowski and Tama, 2008). Here, a biasing potential is added to the molecular mechanics force field, and its force constant regulates the magnitude of the fitting. The flexible fitting usually requires a starting structure in a different state or closely related to the one prepared from other methods such as homology modeling. The results of the flexible fitting are strongly system-dependent, often fail for complex conformational changes or for globular-shaped molecules when low- or medium-resolution maps ($>5 \text{ \AA}$) are treated. This is mainly due to degeneracy of the experimental data, which prevents us from determining a unique conformation (Goh et al., 2016). Another common problem is overfitting, which leads to excessive deformations of the fitted structure and usually occurs when the applied force is too strong. On the other hand, a too weak force constant brings poor results of the flexible fitting. This issue underlines the importance of the optimization of the force constant.

REUSfit is one of the useful methods to overcome this problem (Miyashita et al., 2017), which is an extension of the replica-exchange umbrella-sampling method (REUS) (Sugita et al., 2000). Here, copies of the original system are prepared, and different force constants in the biasing potential are assigned to each replica. The flexible fitting is carried out in parallel, where the force constants are exchanged during the simulation. REUSfit showed better results than simple MD-based flexible fitting, as it enables the automatic adjustment of the force constant (Miyashita et al., 2017). However, REUSfit is computationally expensive if it is employed with the all-atom (AA) model with explicit solvent.

To date, various molecular models at different resolutions have been developed in efficient and accurate MD simulations of proteins, ranging from Coarse-grained (CG) Ca models (Kolinski, 2004) (Kmiciek et al., 2016), through AA models in implicit solvent (Schaefer et al., 1998) (Grochowski and Trylska, 2008) (Onufriev and Case, 2019) and in explicit solvent (Zhou, 2003). CG models are computationally efficient, among which the Go model has been widely used for tackling large conformational changes of proteins (Taketomi et al., 1975) (Whitford et al., 2010) (Kobayashi et al., 2015). Implicit solvent model like the GB/SA model treats solvent molecules as the continuum to reduce the computational cost for the solute-solvent or solvent-solvent interaction calculations (Anandakrishnan et al., 2015). CG model is the most suitable for REUSfit, while high-resolution molecular model is eventually desired to refine the protein structure, as it reveals the atomic interactions in biomolecules at a greater level of detail and accuracy.

In this study, we propose a multi-scale protocol in the cryo-EM flexible fitting that can connect CG and AA models in order to reach the final structural models with a quality not accessible by using any of those simulation methods separately. We focus on medium resolution density maps, i.e., around $5\text{-}8 \text{ \AA}$, since the flexible fitting is generally useful for such maps. We tested our protocol in various systems and found that it is particularly useful for the protein complexes with large conformational transitions, such as Ca^{2+} -ATPase and DNA polymerase.

METHODS

Flexible Fitting Algorithm

In MD simulations with flexible fitting, biomolecular structure is guided toward the EM density map by a biasing potential V_{EM} . This potential is simply added to the force field function and depends on a given force constant k and a cross-correlation coefficient CC between the target density map and the one calculated in the course of the simulation (Orzechowski and Tama, 2008):

$$V_{EM} = k(1 - CC).$$

CC is computed for each (i, j, k) voxel containing the target ρ^{targ} and calculated ρ^{calc} densities (Tama et al., 2004):

$$CC = \frac{\sum_{ijk} \rho^{targ}(i, j, k) \rho^{calc}(i, j, k)}{\sqrt{\sum_{ijk} \rho^{targ}(i, j, k)^2 \sum_{ijk} \rho^{calc}(i, j, k)^2}}$$

Introduction of REUS scheme (Sugita et al., 2000) to the flexible fitting simulations enables to dynamically adjust the force constant during the simulation. This approach attributes higher force constants to the replicas with higher CC , i.e., better fitted to the target density maps. The exchange probability between two replicas i and j , with force constants k_m and k_n and CC to the target map equal to CC^i and CC^j , respectively, is estimated in the following way (Miyashita et al., 2017):

$$w(X \rightarrow X') \equiv w(x_m^{[i]} | x_n^{[j]}) = \begin{cases} 1, & \text{for } \Delta \leq 0 \\ \exp(-\Delta), & \text{for } \Delta > 0 \end{cases}$$

$$\Delta = \beta(k_n - k_m)(CC^j - CC^i)$$

The parameter β stands for the inverse temperature and here is the same for all replicas. This method was implemented in GENESIS as REUSfit (Miyashita et al., 2017). For details about the generation of density maps with a Gaussian mixture model and about the domain decomposition and atomic decomposition parallelization strategies developed to speed up the flexible fitting simulations, see (Mori et al., 2019).

Targeted MD Algorithm

In targeted MD, the aim is to simulate the conformational transition of a molecule at initial configuration vector \mathbf{r}_i to the known target state given by configuration vector \mathbf{r}_F . For each configuration, a distance between the current and target configuration can be calculated as (Schlitter et al., 1994):

$$\rho = |\mathbf{r} - \mathbf{r}_F|$$

In order to force the system to undergo the desired transition, we introduce a force constraint F^c :

$$\phi(\mathbf{r}) = |\mathbf{r} - \mathbf{r}_F|^2 - \rho^2 = 0$$

$$F^c \equiv \lambda \frac{d\phi}{dr} = 2\lambda(\mathbf{r} - \mathbf{r}_F)$$

The Lagrangian multiplier λ is calculated to satisfy the above equations. At every time step, ρ is diminished by $\Delta\rho$ until the final value ρ_f is reached:

$$\Delta\rho = \frac{\Delta t(\rho_0 - \rho_f)}{T},$$

where T stands for the total simulation time. In GENESIS, the RMSD is used instead of distance. Therefore, the implemented method is based on internal coordinates rather than on Cartesian coordinates.

Chosen Systems

Seven different systems, containing PDB structures in at least two different conformations, were selected for flexible fitting simulations using simulated density maps: Ca²⁺-ATPase (Toyoshima and Nomura, 2002) (Toyoshima et al., 2000) (Toyoshima and Mizutani, 2004) (Toyoshima et al., 2007), Na⁺ K⁺-ATPase (Morth et al., 2007) (Kanai et al., 2013), Adenylate Kinase (Müller et al., 1996) (Müller and Schulz, 1992), Ribose Binding Protein (Björkman et al., 1994) (Björkman and Mowbray, 1998), Maltodextrin Binding Protein (Sharff et al., 1992) (Quioco et al., 1997), Diphtheria Toxin (Bennett et al., 1994) (Bennett and Eisenberg, 1994) and CO Dehydrogenase (Darnault et al., 2003). In the case of simulations with the experimental density maps, two systems were chosen, for which PDB structures in two states, as well as an experimental Cryo-EM density map for one of those states, are available: Magnesium Transporter CorA (Matthies et al., 2016) and DNA Polymerase (Fernandez-Leiro et al., 2015).

CG Simulations

The MMTSB Go model Web Server was used to prepare the Ca model parameters based on the potential energy function developed by Karanicolas and Brooks (Karanicolas and Brooks, 2003). For the structures containing more than one protein chain, in-home scripts were prepared to combine parameters for different chains, taking into account the inter- and intra-chain interactions. Additionally, the native contact values between $\tau 500$ and the tail of PolIII α were increased threefold to prevent $\tau 500$ from detaching from the polymerase complex, as this interaction was previously reported stable upon the studied conformational transition (Fernandez-Leiro et al., 2015). The initial and target PDB structures were superimposed using the Ca atoms. The simulated density maps were generated at 5 Å resolution with 2 Å voxel size using the target PDBs and `emmap_generator` tool from the GENESIS package (Mori et al., 2019). The resolution and voxel size of the experimental density maps were left unchanged and the negative values were removed with the SITUS `voledit` tool (Wriggers et al., 1999).

The flexible fitting molecular dynamics simulations with replica exchange (REUSfit) were performed using 32 replicas at force constants 500; 524; 555; 593; 636; 684; 737; 794; 856; 921; 991; 1,065; 1,144; 1,228; 1,318; 1,415; 1,519; 1,632; 1,755; 1,889; 2,035; 2,195; 2,371; 2,564; 2,777; 3,011; 3,269; 3,553; 3,865; 4,209; 4,586; 5,000 kcal/mol. The distribution of the force constants was based on a preliminary survey to ensure the sufficient overlap of the biasing potential between replicas. The cutoff distance for the

native and non-native contact interaction terms was set to 20 Å. All bond lengths were fixed with the SHAKE algorithm (Ryckaert et al., 1977). The temperature was kept constant at 200 K using the Berendsen thermostat (Berendsen et al., 1984). The exchanges were attempted every 2 ps and each replica was simulated for 250 ns with a time step of 20 fs, except for DNA Polymerase, where 10 fs was used as a time step. The density map was recalculated at every time step with the truncation of the Gaussian function to zero for the densities below 1% of the maximum value (Tama et al., 2004). The REUSfit of each system were repeated 20 times with different initial velocities. The highest CC structures in each simulation were chosen as the best models in each run.

AA Simulations

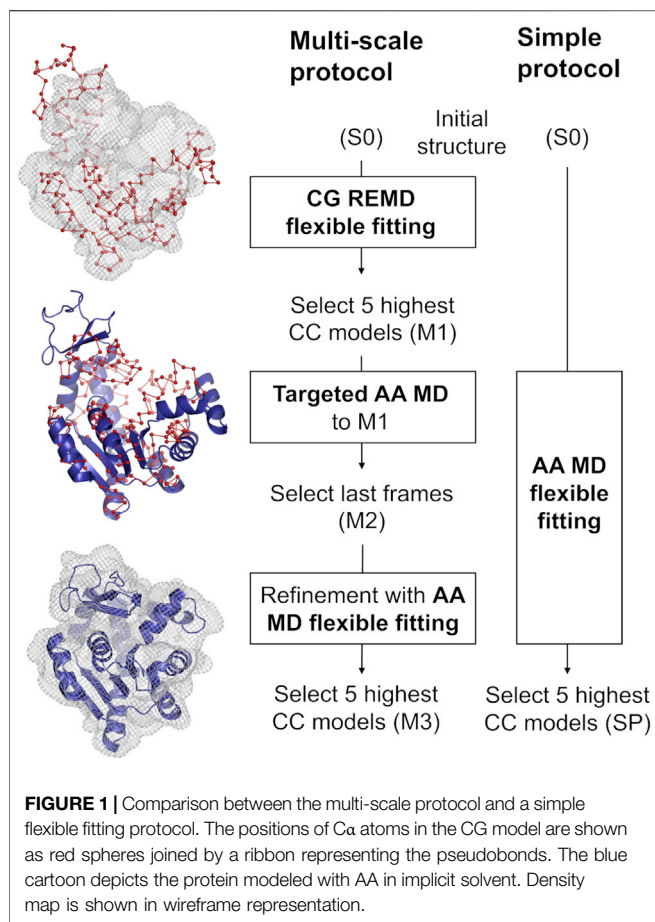
CHARMM C36 force field parameters for proteins (Best et al., 2012) were prepared using `psfgen` plugin in VMD (Humphrey et al., 1996) with the patches for disulfide bonds where necessary. GBSA implicit solvent model with OBC2 parameters (Onufriev et al., 2004) was used with salt concentration 0.15 M and the surface tension coefficient 0.005 kcal·mol⁻¹·Å⁻². The dielectric constant was set to 78.5. For the AA simulations, the 1 Å voxel size of the simulated density maps was used. The experimental density maps were the same as at CG level. In the second step of the multi-scale protocol, the Ca positions from 5 best models from CG simulations were used as a target in targeted MD simulations of AA structures in implicit solvent. The simulations in NVT ensemble with Langevin thermostat with the friction coefficient γ set at 5 ps⁻¹ were performed for 1 ns with time step 2 fs. The final structure after targeted MD will have low RMSD value with respect to the target but we perform additional alignment using the Ca atoms to shift the center of mass of the system.

The last step of the multi-scale simulations is the refinement of the last frame of the targeted MD trajectories with the flexible fitting simulations to the density maps. The simulation is performed for 1 ns using a similar set of parameters as the Targeted MD simulation. For comparison, the AA flexible fitting simulations starting from original PDB structure are carried out in 5 replicates and referred to as the simple protocol. The best models from each of those simulations are chosen according to the highest CC. All simulations were done in GENESIS version 1.4 (<https://www.r-ccs.riken.jp/labs/cbrt/>).

Analysis

Various tools were used for the analysis of the best final models M1, M2, M3, SP and also S0 and RF: CC without rigid body docking was calculated in the collage program from Situs package (Wriggers et al., 1999); RMSD in Gromacs 2016.4 (Abraham et al., 2015) with `gmx rms` tool; CaBLAM and Ramachandran outliers and MolProbity score in Phenix (Williams et al., 2018) (Adams et al., 2010).

CaBLAM and Ramachandran outliers and MolProbity score depend on the relation between the geometry of the final models and a high-quality protein dataset from PDB database, called Top8000. CaBLAM is a method to evaluate the geometry of the protein backbone using the virtual angles and dihedrals created



by the Ca atoms and carbonyl groups [details in (Williams et al., 2018)]. Ramachandran outliers use the ϕ and ψ angles of the protein backbone for analysis.

Secondary structure score is determined by comparing the secondary structure of the fitted model and the target, namely by counting the Hamming distance between two strings, representing the coded secondary structure features for each residue, as calculated by mkdssp version 2.2.1 (Touw et al., 2015) (Kabsch and Sander, 1983) and in-house scripts. The Hamming distance is normalized by dividing it by the length of the string and representing it as a percentage.

RESULTS

Multi-Scale Protocol for Flexible Fitting

Our multi-scale protocol involves three steps (Figure 1, left scheme). In step 1, REUSfit with the Ca Go model (Karanicolas and Brooks, 2003) is employed. In this model, one amino acid is represented by a single bead at their Ca atom position, and the beads interact via the structure-based potential that is constructed from the native structure. The interactions between side chains are scaled for respective amino acid types using Miyazawa-Jernigan contact energies (Miyazawa and Jernigan, 1996). In REUSfit, the

cross-correlation-coefficient (CC) between the calculated and experimental density maps is included in the biasing potential. We carry out five individual runs, and the best models, called M1, are chosen from each run according to the highest CC. In step 2, we carry out targeted MD (Schlitter et al., 1994) starting from the initial atomic structure toward M1, where the implicit solvent model (GB/SA) (Onufriev et al., 2004) is used to reduce the computational cost. Note that the experimental density map is not used in this step. Since targeted MD imposes the restraint on the root-mean-square deviation (RMSD) with respect to the target structure, the Ca coordinates in the AA model do not exactly match the target CG model. Therefore, additional alignment to M1 is necessary to shift the center of mass of the obtained structure after the targeted MD. The final atomic structure is called M2. In step 3, M2 is further refined with the CC-based flexible fitting using the experimental density map. Here, the GB/SA implicit solvent model is again used as in step 2. The final models are chosen according to their CC and named M3.

We compare the best final models obtained via two protocols, starting from the same initial structure S0: the multi-scale protocol and one, long, all-atom flexible fitting simulation called a simple protocol (SP) (Figure 1, right scheme). The computational condition is same as in step 3 of the multi-scale protocol.

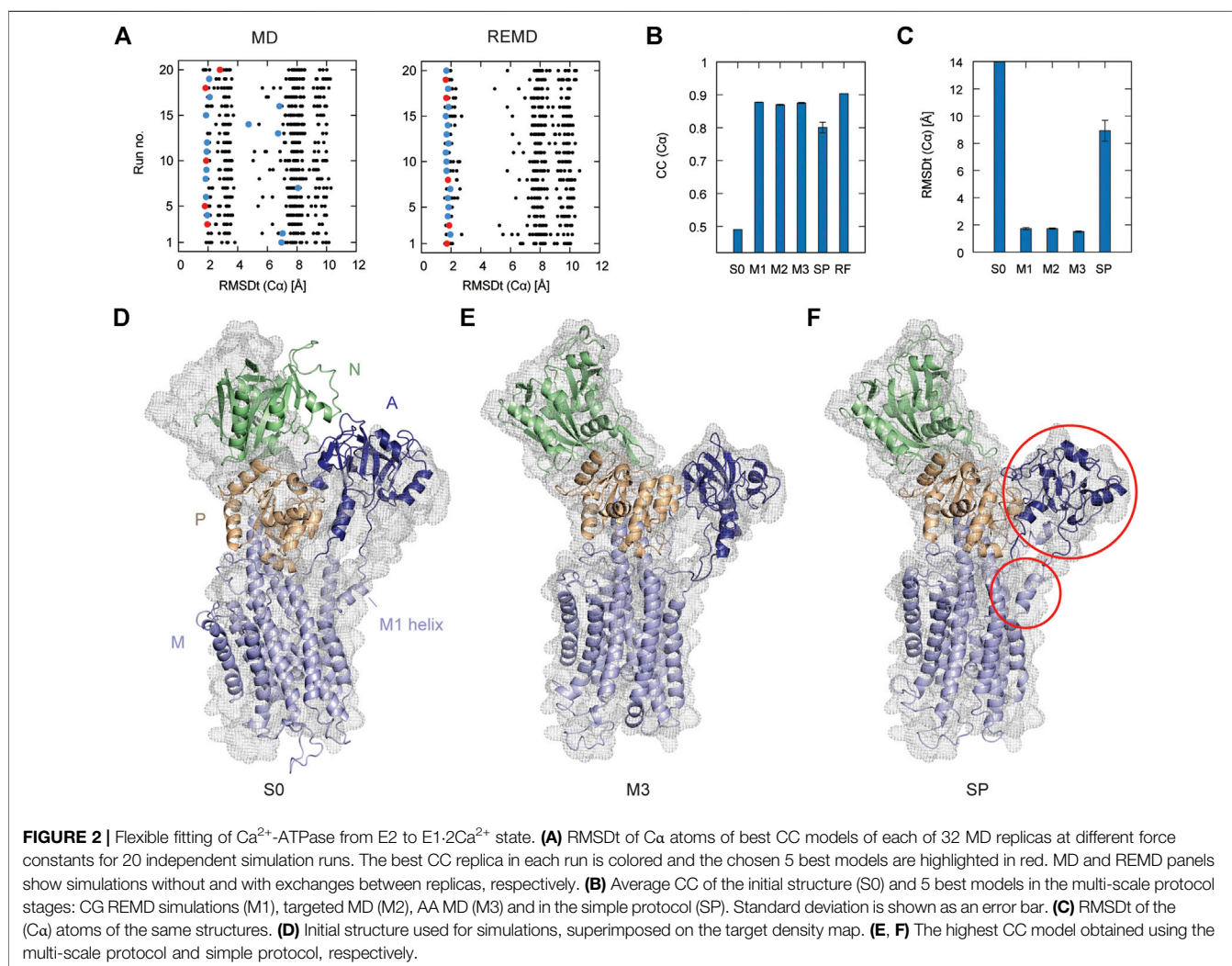
The two protocols are applied to a range of systems with simulated and experimental density maps, shown in Table 1. Simulated density maps serve as a test of accuracy and precision of fitting. The density maps were generated from X-ray crystal structures of the target biomolecules at 5 Å resolution. Tested systems include a small protein such as Adenylate Kinase as well as relatively large ones, for instance, DNA polymerase. All tested systems are classified as having two significantly different states in initial and target structures. The transitions between the two states represent varying levels of difficulty in each case. Particularly, Ca²⁺-ATPase and Diphtheria Toxin are difficult to fit correctly (initial RMSD >10 Å), while Ribose Binding Protein and Maltodextrin Binding Protein are relatively easy (initial RMSD <4 Å). For the experimental density maps, we have chosen two systems, for which both the cryo-EM density maps and fitted models were available, together with a starting structure in a different physiological state, involving complex conformational transition between the two states. Two systems fulfilling all the criteria were found, with cryo-EM density maps at resolution in the range 5-8 Å: magnesium transporter CorA (Matthies et al., 2016) and DNA polymerase (Fernandez-Leiro et al., 2015).

Large Conformational Transitions can be Fitted Well With Multi-Scale Protocol

First, we focus on Ca²⁺-ATPase, which requires large conformational transitions during the flexible fitting. Ca²⁺-ATPase undergoes a series of structural changes between E2, E1-2Ca²⁺, E1P, and E2P states (Supplementary Figure S1), linked with Ca²⁺ transport against the concentration gradient (Toyoshima, 2008). If the flexible fitting is performed in these

TABLE 1 | Systems used in the flexible fitting simulations using either simulated or experimental density maps. The RMSD was calculated between the C α atoms of the initial structure and the target structure, for the number of residues (and chains), specified in the last column.

System	State (PDB ID) of the initial \rightarrow target structures	Density map, resolution [\AA]	RMSDt (C α) [\AA]	# of residues (chains)
Ca $^{2+}$ -ATPase	E2 (1iwo) \rightarrow E1.2Ca $^{2+}$ (1su4)	simulated, 5.0	14.0	994 (1)
	E1.2Ca $^{2+}$ (1su4) \rightarrow E1-ATP (1vfp)	simulated, 5.0	13.7	994 (1)
	E1-ATP (1vfp) \rightarrow E2P (2zbf)	simulated, 5.0	11.2	994 (1)
	E2P (2zbf) \rightarrow E2 (1iwo)	simulated, 5.0	5.5	994 (1)
Na $^{+}$ K $^{+}$ -ATPase	K $^{+}$.E2 (3kdp) \rightarrow Na $^{+}$.E1P-ADP (3wgu)	simulated, 5.0	11.6	1,280 (2)
	Na $^{+}$.E1P-ADP (3wgu) \rightarrow K $^{+}$.E2 (3kdp)	simulated, 5.0	11.6	1,280 (2)
Adenylate kinase	open (4ake) \rightarrow closed (1ake)	simulated, 5.0	7.1	214 (1)
Ribose binding protein	closed (2dri) \rightarrow open (1urp)	simulated, 5.0	4.1	271 (1)
Maltodextrin binding protein	open (1omp) \rightarrow closed (1anf)	simulated, 5.0	3.8	370 (1)
Diphtheria toxin	open (1ddt) \rightarrow closed (1mdt)	simulated, 5.0	15.6	523 (2)
CO dehydrogenase	closed (1oac) \rightarrow open (1oacD)	simulated, 5.0	7.9	729 (1)
Magnesium transporter CorA	closed (3jcf) \rightarrow open (3jcg)	Experimental (EMD-6552), 7.1	9.7	1,653 (5)
DNA polymerase	Bound (5fkv) \rightarrow free (5fku)	Experimental (EMD-3201), 8.3	14.2	2,211 (7)



states, the movements of cytoplasmic domains (A, P, and N) and transmembrane helices in domain M are challenging for us, since the A domain rotates by 90° from the E1P to E2P state and

the inclination of the N domain with respect to the P domain in the E1.2Ca $^{2+}$ and E1-AMPPCP crystal structures changes by almost 90° as well.

TABLE 2 | Quality of the best models obtained using the simple and multi-scale protocol for the simulated density maps for Ca²⁺-ATPase and Diphtheria Toxin.

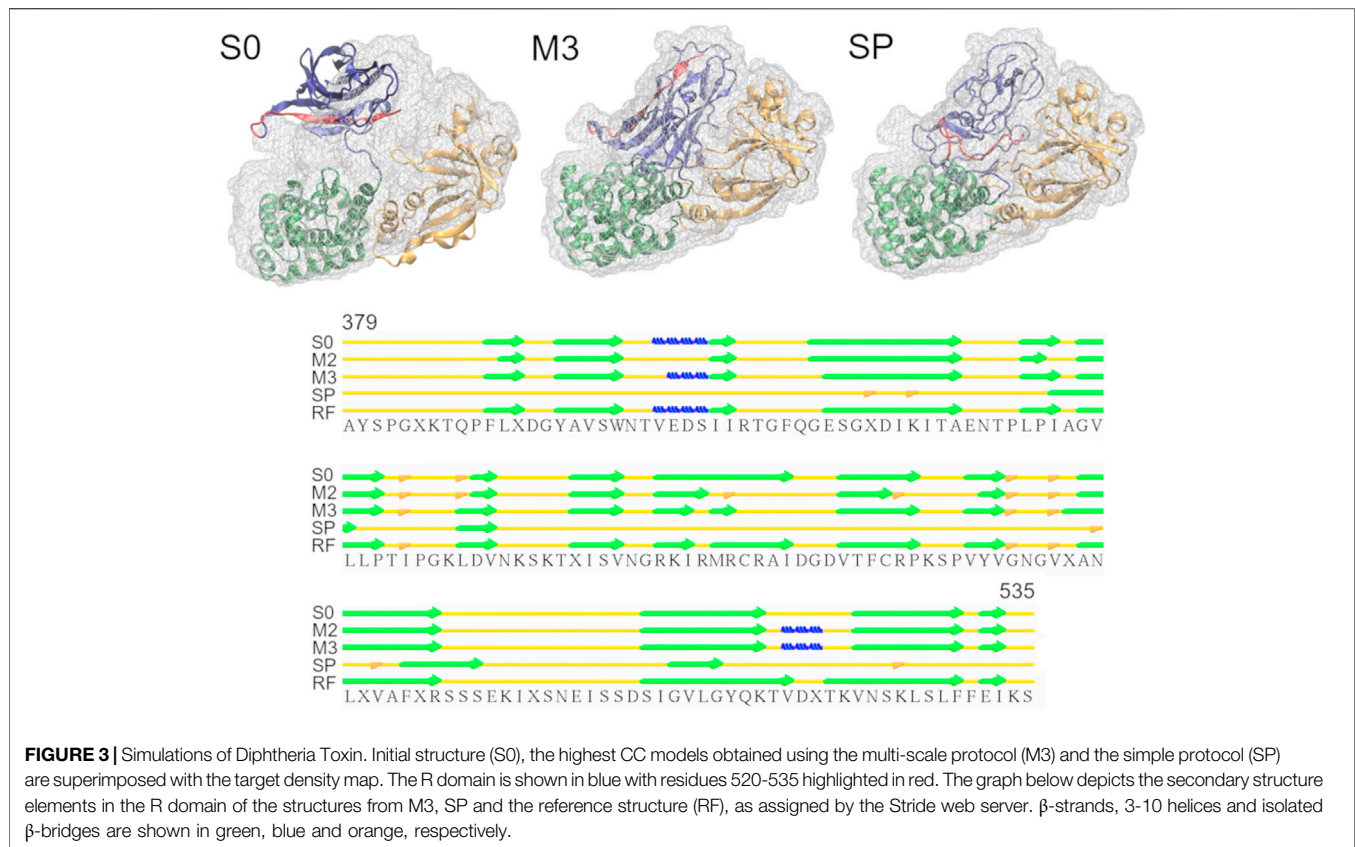
System	Str.	c.c.	RMSDt (Ca) [Å]	RMSDt (AA) [Å]	MolProbioty score	Ramachandran outlier [%]	Rotamer [%]	CaBLAM [%]	Sec. struct. index [%]
Ca ²⁺ -ATPase	MSS	0.96 (0.95 ± 0.00)	1.4 (1.5 ± 0.0)	2.3 (2.4 ± 0.1)	1.67 (1.74 ± 0.05)	2.7 (3.3 ± 0.3)	6.3 (7.3 ± 0.7)	4.0 (5.3 ± 1.0)	22 (23 ± 1)
	SP	0.91 (0.90 ± 0.02)	8.8 (8.9 ± 0.8)	9.2 (9.4 ± 0.8)	1.85 (1.80 ± 0.04)	4.2 (4.8 ± 0.4)	8.5 (8.7 ± 0.7)	7.2 (7.5 ± 1.3)	31 (32 ± 1)
E1-2Ca ₂₊	MSS	0.96 (0.95 ± 0.00)	1.5 (1.5 ± 0.1)	2.3 (2.3 ± 0.1)	1.79 (1.82 ± 0.03)	2.3 (2.4 ± 0.4)	9.1 (9.7 ± 1.1)	4.7 (4.0 ± 0.6)	21 (23 ± 1)
	SP	0.94 (0.94 ± 0.00)	3.0 (2.8 ± 0.3)	3.6 (3.5 ± 0.2)	1.83 (1.82 ± 0.05)	3.7 (3.6 ± 0.4)	9.0 (8.8 ± 1.0)	5.1 (5.8 ± 0.5)	22 (24 ± 1)
E1-ATP → E2P	MSS	0.96 (0.96 ± 0.00)	1.4 (1.4 ± 0.1)	2.3 (2.3 ± 0.1)	1.72 (1.75 ± 0.06)	2.5 (2.7 ± 0.4)	8.5 (7.9 ± 0.7)	2.8 (3.7 ± 0.5)	21 (21 ± 1)
	SP	0.93 (0.92 ± 0.02)	4.4 (5.1 ± 1.5)	4.9 (5.6 ± 1.5)	1.85 (1.84 ± 0.04)	4.3 (4.0 ± 0.7)	9.4 (8.4 ± 0.8)	6.3 (6.5 ± 0.8)	26 (28 ± 1)
E2P → E2	MSS	0.96 (0.96 ± 0.00)	1.3 (1.3 ± 0.0)	2.1 (2.1 ± 0.0)	1.56 (1.69 ± 0.02)	2.0 (2.0 ± 0.3)	6.5 (7.4 ± 0.6)	2.9 (3.2 ± 0.6)	21 (23 ± 1)
	SP	0.96 (0.95 ± 0.00)	1.4 (1.5 ± 0.1)	2.2 (2.3 ± 0.1)	1.63 (1.70 ± 0.05)	1.7 (2.1 ± 0.4)	6.5 (7.7 ± 1.2)	3.6 (3.7 ± 0.6)	24 (25 ± 2)
Diphtheria toxin	MSS	0.97 (0.97 ± 0.00)	0.8 (0.9 ± 0.1)	1.6 (1.7 ± 0.1)	1.71 (1.74 ± 0.03)	2.7 (3.0 ± 0.3)	7.8 (7.7 ± 1.0)	5.4 (5.2 ± 0.5)	18 (19 ± 2)
	SP	0.95 (0.95 ± 0.00)	9.9 (9.6 ± 0.3)	10.3 (9.9 ± 0.3)	1.93 (1.86 ± 0.05)	6.0 (5.4 ± 0.6)	8.3 (7.7 ± 0.5)	10.5 (8.9 ± 1.5)	38 (33 ± 3)

In **Figure 2A**, the left panel shows 20 individual runs of the flexible fitting from E2 to E1-2Ca²⁺ states using the CG model at different force constants (500–5,000 kcal/mol), which corresponds to REUSfit without replica exchange. The final fitted models with the highest CC are colored and chosen as the best models. Their RMSD with respect to the target structure (RMSDt) around 2 Å represent the successful simulations with well-fitted domain A. However, there are multiple failed attempts. Introducing the exchanges between the replicas brings substantial improvement, and all the attempts are successful (right panel in **Figure 2A**). Each fitting run of the multi-scale protocol is highly repeatable due to exchanges between replicas. Therefore, a single fitting simulation following the multi-scale procedure is sufficient to obtain a reliable result, which is visible in the low values of standard deviation of multiple replicates in **Table 2**. Best models from the simple protocol reveals significantly higher standard deviations. Similar results were obtained in the previous work using 10 Å resolution maps (Miyashita et al., 2017). Thus, the structures with the correct Ca atoms positions (M1) can be further used in the multi-scale protocol in targeted MD and refined to obtain M2 and M3 models, respectively.

In **Figure 2B**, we summarize the results of the multi-scaling protocol, comparing the averaged CC of the obtained M1, M2, and M3 models. For comparison, CC of the initial structure S0, reference structure RF, and the results of the simple protocol (SP) are also shown. The best models obtained using our multi-scale protocol (CC = ~0.88) are significantly better than that in the simple protocol (CC = ~0.80), indicating a successful fitting to the density map. RMSDt of the M3 model is also smaller (1.5 Å) than that of the SP model (8.9 Å) (**Figure 2C**). In addition, the MolProbioty score (Keedy et al., 2009), CaBLAM (Williams et al., 2018), and Ramachandran outliers of the M3 model are better than the SP model (**Supplementary Figure S2**). The RMSDt of the SP models are closer to the S0 structure, implying that during the single-step fitting, smaller conformational changes occurred than during the three-step multi-scale protocol. In fact, RMSDi (RMSD with respect to the initial structure) of the M2 and M3 models is larger than that of the SP model (**Supplementary Figure S2**).

In the multi-scale protocol, all three steps show similar averaged values of the CC (Ca) and RMSDt (Ca), around 0.88 and 1.5–1.7 Å, suggesting that the first CG simulation with REUSfit is the most important to determine the positions of the Ca atoms. Targeted MD is also working well to superimpose the all-atom model to the CC model by means of a gradual decrease of the RMSD of Ca atoms to their target positions. The protein side chains follow the conformational changes induced by the shift of Ca atoms. The detailed conformation of the protein side chains can be refined in step3, as the density map is not present during the targeted MD. We found that RMSDt of all non-hydrogen atoms is improved from 2.7 to 2.4 Å (**Supplementary Figure S2**). Although each domain undergoes almost rigid-body rotation and thus most side chains are already well positioned in the targeted MD, further refinement seems to be possible in step3.

One of the main differences between the M3 and SP models is the orientation of the A domain. In the initial structure, the N and A domains are contacting each other (**Figure 2D**), and they are separated during the fitting. The A domain acts as an actuator of

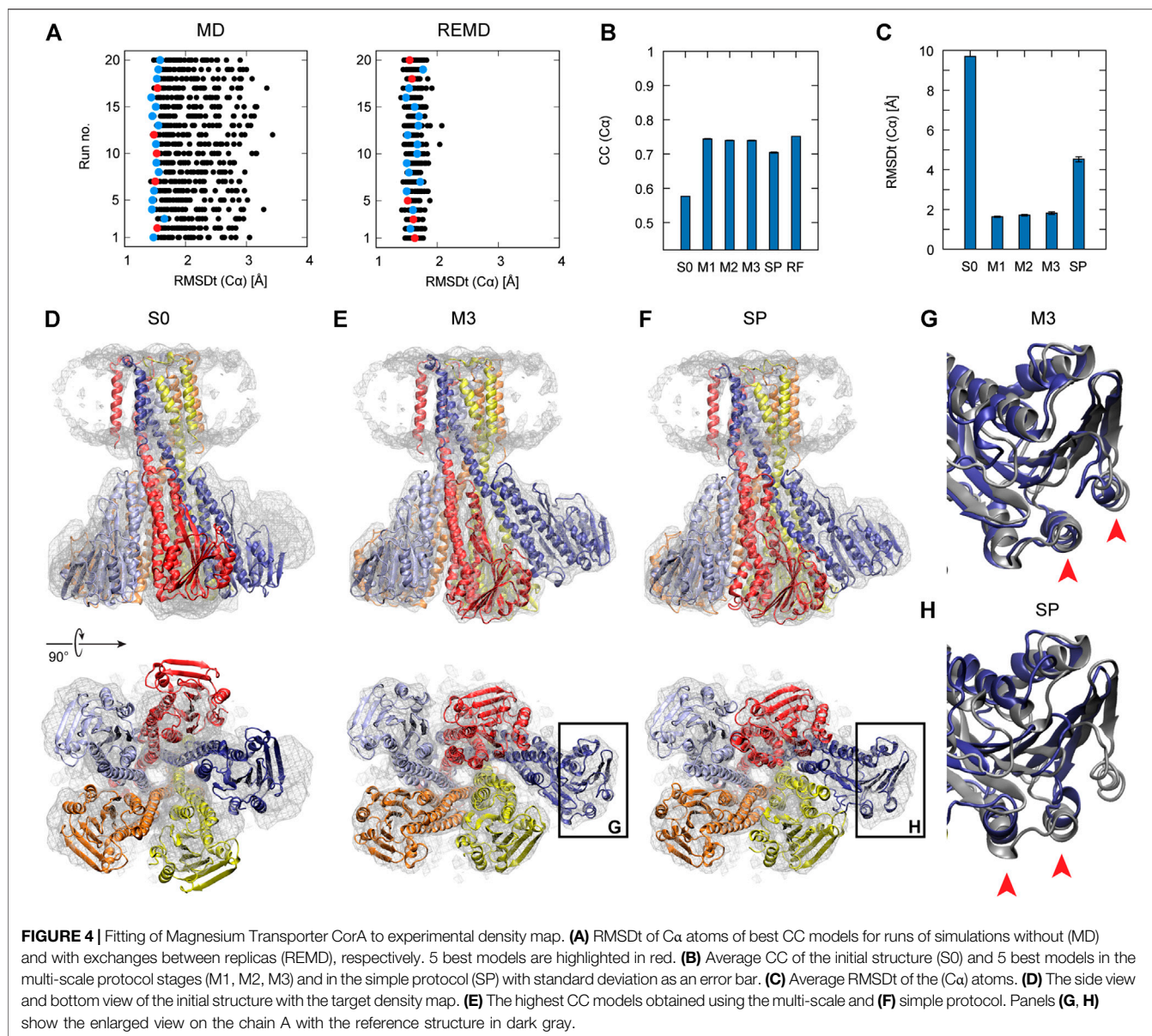


the gating of the ion pathway and it has to rotate in order to regulate the Ca^{2+} binding and release. In the CG simulation, this shift occurs with a domain rotation. This success extends to the all-atom level in the next stage of the multi-scale protocol, resulting in a correct final model (Figure 2E and Supplementary Movie S1). On the other hand, in the simple protocol the A domain was simply fitted to the density without correct rotation, and none of the simulations was successful (Figure 2F and Supplementary Movie S2). Since the orientation of the A domain can affect the transmembrane domain through the loop connecting them, the conformation of the N-terminal side of the M1 helix is also reproduced in the M3 model.

Interestingly, troublesome regions in the flexible fitting in the other physiological states of Ca^{2+} -ATPase and $\text{Na}^+ \text{K}^+$ -ATPase are similar, including the transmembrane M1 and M2 helices and A domain. (Supplementary Figure S3). In these cases, the multi-scale protocol still performs significantly better than the simple protocol (see Table 2 below). The main reason behind this is the rotation of the A domain between E1 and E2-type states, performed correctly in the multi-scale protocol and incorrectly in the simple protocol. Without the proper rotation, the secondary structure of the A domain in SP models becomes deformed to accommodate to the density map. In the transition from $\text{E1} \cdot 2\text{Ca}^{2+}$ to $\text{E1} \cdot \text{ATP}$ state, the transmembrane helices M1 and M2 undergo shifting and rearrangement with bending of the N-terminal of the M1 helix. This is reasonably well reflected in M3 models but not in SP models, where M1 helix is bent in opposite way than expected.

In order to represent the degree of disruption of the secondary structure of a protein, we are introducing in this work a new measure called a “secondary structure score.” This score is based on a Hamming distance between two strings, containing the coded secondary structure elements of each residue of two structures that are compared (see the details in the Methods section). The smaller is the value of the secondary structure score, normalized and indicated as a percentage, the more similar are two structures. The last panel of the Supplementary Figure S2 shows the secondary structure scores calculated between the target structure and each model M2, M3 and SP. Here, it is well visible that the secondary structure of the best SP model (31%) deviates much more from the target structure than the best M3 model (22%).

Diphtheria Toxin also requires correct domain rotation during the fitting. The initial structure is shown in the S0 panel in Figure 3, where RMSDt is 15.6 Å. To emphasize the change in orientation of the relevant domain (R domain), its C-terminal fragment was highlighted in red. The M3 and SP panels are the highest-CC models obtained from the multi-scale and simple protocols, respectively. The M3 model showed correct orientation of the domain R [average RMSDt (Ca) = 0.9, RMSDi (Ca) = 15.6, and CC (Ca) = 0.89], while it was wrong in the SP model [average RMSDt (Ca) = 9.6, RMSDi (Ca) = 11.7, and CC (Ca) = 0.84]. In addition, there was a disruption of the secondary structure of several β -strands in the R domain of the SP model (bottom panel in Figure 3 and Supplementary Figure S4, average secondary structure score = 33%). Such disruption is not observed in the

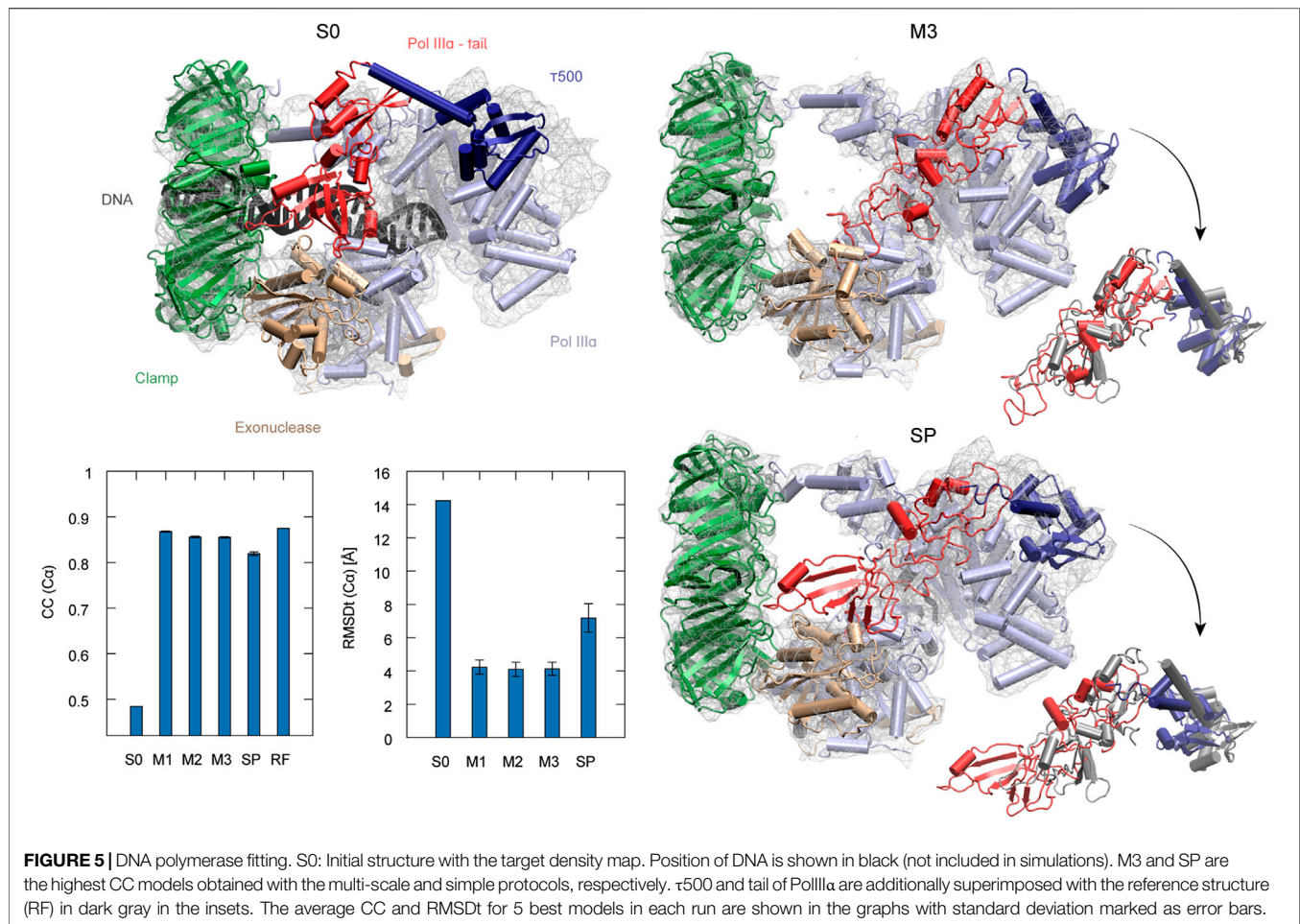


multi-scale protocol (average sec. struct. score = 19%), presumably because the CG simulations with the Go model enable keeping the native contacts within the domain. A slight improvement of the secondary structure during the final refinement in the multi-scale protocol may be caused by the adjustments of the side chains into the density map, which may influence the backbone geometry and hydrogen bonds between strands.

Simple Conformational Changes in Small Proteins do not Require Multi-Scale Procedure

The flexible fitting simulations were also performed for several small systems: Adenylate Kinase, Maltodextrin Binding Protein, Ribose Binding Protein, and CO Dehydrogenase. Even though the two conformational states of those systems are classified as

significantly different and referred to as “open” and “closed” states in the literatures, the overall transition between them is very simple and does not involve large rotation or shift in the component domains. For those systems, the best models obtained from the multi-scale protocol are similar to those from the simple protocol (**Supplementary Figure S5**). As a representative, let us take Ribose Binding Protein. The CC (Ca) in the M1, M2, M3, SP, and RF models is 0.89–0.90, and the RMSDt (Ca) is ~ 0.6 Å. We found that the multi-scale protocol gives structures of a comparable quality in terms of the MolProbity scores. The average score is 1.34 ± 0.11 in M3, while 1.23 ± 0.08 in SP (see **Supplementary Table S1**). The average values of Ramachandran, rotamer and CaBLAM outliers are similar within the standard deviation. However, the best M3 models selected basing on the CC are slightly better than the best SP models. Although the multi-scale protocol is applicable to



small proteins, the simple protocol seems to be enough in most cases. The multi-scale protocol is useful to search large conformational space. Involving the CG models in the protocol is much more beneficial for large proteins and their complexes.

The Multi-Scale Protocol Performs Well for Experimental Cryo-EM Density Maps

CorA is responsible for magnesium uptake in prokaryotes. In the closed Mg^{2+} -bound state, it forms a symmetrical pentamer. While moving to the open Mg^{2+} -free state, the 5-fold symmetry is lost. The density map of the latter state is used here as the target of flexible fitting simulations. In order to assess the performance of our protocol by calculating RMSDt, we have used the model fitted by rigid body methods and deposited by the density map authors in the RCSB PDB database. We realize that this model may not be highly accurate due to its low resolution but that is the only “answer” model that could be used for comparison and we may possibly contribute to its refinement.

To examine the usefulness of REUSfit in the multi-scale protocol, we first compare CG MD with and without replica exchange. In **Figure 4A**, we see that REUSfit produced all the models in the RMSDt range 1–2 Å, suggesting that we can select a reliable model even from

one simulation without considering the effect of the force constant. Since the targeted MD worked well in step 2, the all-atom model was superimposed to the CG model effectively. As a result, the multi-scale protocol gives significantly better models than simple protocol in terms of both CC (Cα): 0.74 vs 0.70 and RMSDt (Cα): 1.8 vs 4.5 Å (**Figures 4B,C**). Similar tendency was obtained in the averaged RMSDt for non-hydrogen atoms, CaBLAM outliers, MolProbity score, and secondary structure score (Figure S7). The reason behind it is the difficulty of fitting is the skew of chain A (dark blue in **Figures 4D–F**). The comparison of two chosen helices of chain A with the target structures shows that they were fitted correctly by the multi-scale protocol in the CG step of the protocol using REUSfit and trapped in a wrong density patch in the best model obtained with the simple protocol (red arrows in **Figures 4G,H**). The REUSfit method is able to rescue the system from being trapped in local density minima by the efficient conformational sampling.

Step 3 did not improve the structure with respect to the target model, as both CC and RMSDt values of the M2 and M3 models are similar. As already mentioned, the target model is not a high-resolution structure of the system. Also, the resolution of the density map is 7.1 Å. Therefore, it is difficult to conclude whether the final refinement step brings any improvement to the structure. It is expected that for the target density maps of

TABLE 3 | Quality of the best models obtained using the simple and multi-scale protocol for the experimental density maps.

System	Str.	c.c.	RMSDt (Ca) [Å]	RMSDt (AA) [Å]	MolProbity score	Ramachandran outlier [%]	Rotamer outlier [%]	CABLAM outlier [%]	Sec. struct. index [%]
Magnesium transporter CorA	MS3	0.77 (0.77 ± 0.00)	1.7 (1.8 ± 0.1)	2.9 (3.0 ± 0.0)	1.70 (1.70 ± 0.02)	2.8 (3.0 ± 0.6)	7.3 (7.3 ± 0.4)	5.3 (4.5 ± 0.5)	18 (19 ± 1)
	SP	0.74 (0.74 ± 0.00)	4.4 (4.5 ± 0.1)	5.1 (5.1 ± 0.1)	1.75 (1.76 ± 0.01)	2.9 (3.3 ± 0.5)	7.2 (7.7 ± 0.3)	4.8 (5.9 ± 0.7)	22 (22 ± 1)
DNA polymerase	MS3	0.85 (0.85 ± 0.00)	4.1 (4.1 ± 0.4)	4.5 (4.5 ± 0.3)	1.84 (1.80 ± 0.03)	3.6 (3.6 ± 0.4)	9.3 (8.2 ± 0.7)	6.7 (6.3 ± 0.4)	21 (21 ± 1)
	SP	0.82 (0.82 ± 0.00)	6.3 (7.2 ± 0.8)	6.6 (7.5 ± 0.8)	1.84 (1.81 ± 0.02)	4.4 (3.9 ± 0.4)	8.5 (8.1 ± 0.4)	6.1 (6.5 ± 0.3)	23 (24 ± 1)

higher resolution, step 3 would contribute more to the quality of the structure.

The second target using the experimental density map is the DNA polymerase III from *E. coli*. It serves as an enzyme for DNA replication. The complex consists of several factors of different functionality, such as the clamp for sliding along the DNA and exonuclease working as a proofreader. To build the DNA lagging strand, the polymerase needs to quickly and repeatedly release and reposition DNA. In order to understand this process, several structures of the complex have been solved, both in DNA-bound and DNA-free states (Fernandez-Leiro et al., 2015). Here, we performed flexible fitting to a density map of a DNA-free state starting from the DNA-bound state structure. DNA was removed from the original PDB structure before the simulations (Dark gray in S0 panel of Figure 5).

The best model obtained with the simple protocol reveals many inaccuracies with respect to the density map in the regions of the τ 500 domain and neighboring tail of PolIII α , when compared with the modeled reference structure (SP panel in Figure 5). Particularly, the tail of PolIII α that undergoes a rearrangement upon DNA binding is not in a correct position, significantly stretching out of the contours of the density map. Applying the multi-scale protocol allowed avoiding such problems. The secondary structure of those domains is retained with their simultaneous shift toward the correct position within the target density map (M3 panel in Figure 5). Both CC and RMSDt of the Ca atoms of the best final models are significantly improved by using the multi-scale protocol, from 0.82 to 0.85 and from 7.2 to 4.1, respectively (bottom left panel in Figure 5). Similarly, as in the case of the CorA, the CG simulations with REUSfit largely contribute to the correct fitting.

Overall Performance of Multi-Scale Protocol

In Tables 2, 3 and Supplementary Table S1, we summarize the CC, RMSD, and other parameters of the best models obtained from the simple and multi-scale protocols for all test sets. In the cases of Adenylate Kinase, Maltodextrin Binding Protein, Ribose Binding Protein, and CO Dehydrogenase, which include a simple rotation or shift of the component domains, the two protocols are almost comparable. However, in the case of Ca²⁺-ATPase, Na⁺ K⁺-ATPase, Diphtheria Toxin, Magnesium Transporter, and DNA polymerase, which includes complicated conformational changes, multi-scale protocol showed better scores than the simple protocol. In the simple protocol, RMSDt of all the heavy atoms is high,

reaching even 9.2 Å for Ca²⁺-ATPase and 10.3 Å for Diphtheria Toxin. In some cases, we also observe disruptions of the secondary structures in the rotated/shifted domain, resulting in higher secondary structure scores (>30%), and thus, poor MolProbity score and high Ramachandran and CaBLAM outliers. In the multi-scale protocol, RMSDt came down below 2.5 Å in all systems with simulated density maps and to 2.9 and 4.5 Å in the experimental density maps systems. The secondary structure scores are mostly less than 30%, and the main chain geometry is consistently at least 0.5% better than for the SP best models. We suggest that multi-scale protocol is useful for large proteins or their complexes in which a complex conformational transition, such as a rotation or domain rearrangement, occurs between two states.

DISCUSSION

To date, several methods relying on MD simulations have been developed in order to flexibly fit protein structures to the target density maps. Here, we use the CC-based approach (Tama et al., 2004). This can be combined with structure-based force fields, as in MDfit (Whitford et al., 2011), which enables us to prevent the structure from breaking secondary or tertiary structure contacts during the fitting due to native contact energy term. However, new stable contacts cannot be created during the fitting, even if there is a contact in the target structure. MDFF uses the physics-based force field such as CHARMM, which allows for creating new contacts, but requires secondary structure restraint (Villa and Lasker, 2014) (Trabuco et al., 2008). In our approach, we aim to design a fitting method that does not only keep secondary and tertiary structure relations, but also allows for the refinement of side chain positions, including breaking or creating new contacts. Therefore, we use a structure-based model for the first stage of the protocol, and CHARMM force field with the CC-based biasing potential for the third stage, joined together by targeted MD. One advantage of our method is that in the second and third stages we do not introduce any secondary structure restraints, as in CC-based methods, but we take advantage of those restraints at an earlier step by using the structure-based model. Although the protocol requires the correct secondary structure in the initial model, it allows the changes of the secondary structure in the target structures via the non-restraint MD simulations.

Another advantage of our protocol is that we are searching a large conformational space in the first stage with REUSfit (Miyashita et al., 2017), where the computational cost is reduced with the CG model.

REUSfit increases the reliability of the fitting, since exchanging the force constants between replicas can promote a feedback loop to output the structures with the highest CC in different fitting trials. Extending the normal MD simulation or using other methods increasing the sampling such as accelerated MD would also lead to a wide conformational search (Hamelberg et al., 2004). However, it may not choose and promote the same fitted pose in many runs when dealing with a complex conformational transition.

For systems with large and complex conformational changes, it is difficult to obtain reliable results just using a simple method, as the results may not be converged (Ahmed et al., 2012). Frequently, wrong fitting poses are obtained, similarly as in the case of Diphtheria Toxin presented here. Thus, for such cases, our multi-scale approach combining CG and AA model is powerful to reach a final model with reliable backbone positions by REUSfit and also containing hydrogen atoms. The consideration of the hydrogen atoms and hydrogen bonds is of importance for proteins and it allows for refinement of the side chains. Introducing a protocol consisting of three steps may be seen as a disadvantage of our method due to the relative complexity of setting up the files for both CG and AA simulations and running 3 different types of simulations consecutively in comparison with a single flexible fitting run. All necessary functionalities are implemented within the GENESIS suite of programs (Jung et al., 2015) (Kobayashi et al., 2017) (Mori et al., 2019).

The protocol uses targeted MD to convey the information about the correct C α atoms positions after structure-based model fitting to the AA simulations. The most straightforward way one could imagine for going from Ca only model to AA model would be to rebuild the protein backbone and side chains basing on conventional geometrical principles. Various tools were examined in step 2, including for example PULCHRA (Rotkiewicz and Skolnick, 2008) and SCWRL4 (Krivov et al., 2009). In our experiences, none of them was free of severe errors in rebuilt structures, especially in the case of the slightly distorted main chain after flexible fitting simulation. Common problems included strange dihedral angles of the backbone, overlapping side chains and ring penetration (i.e., insertion of a covalent bond into an aromatic ring). We consider that targeted MD is generally applicable to back mapping in multi-scale simulations that combine CGMD and AAMD when at least one reliable experimental structure is available. Another method that could be taken into consideration for retrieving atomic detail from CG model is a recently developed back mapping method using Bayesian inference and restrained MD (Peng et al., 2019).

In this study, we tried to obtain the one best model from one experimental density map. Understanding the experimental data as an ensemble average of the multiple conformations is also important, since many conformations can reproduce the same experimental data (degeneracy problem) (Rieping et al., 2005) (Cossio and Hummer, 2013) (Olsson et al., 2017). In the cryo-EM, multiple states of proteins can be classified with the 3D reconstruction protocol for a large amount of 2D images (Bai et al., 2015). However, the obtained 3D density map is still an ensemble average of molecules with sufficient conformational

fluctuations in solution. In addition, noise or random errors might be contained in the map to some extent, even if high-resolution density maps are obtained. To address these issues in the structure modeling, various inferential algorithms have been proposed (Bonomi et al., 2017). Particularly, metainference calculates the ensemble average using multiple-replica simulations, and also considers effects of the noise and errors in the energy function based on the Bayesian inference (Bonomi et al., 2016), which was recently applied to the cryo-EM structure modeling (Bonomi et al., 2018). Introduction of the ensemble average, noise, or errors will be one of the future extensions of our protocol.

CONCLUSION

We have designed a multi-scale protocol by integrating the CGMD with the AAMD in a single flexible fitting workflow. Performed simulations for several systems containing intricate domain rotations, domain shifts and transitions between open and closed states show that we can improve the accuracy and reliability of the flexible fitting procedure in comparison with a single, long AA flexible fitting simulation. The protocol performs well both for simulated and experimental density maps from cryo-EM and it is recommended to use with large proteins or their complexes.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author. The simulation protocol will be available as a tutorial “Multi-scale Cryo-EM flexible fitting” at the GENESIS webpage (<https://www.r-ccs.riken.jp/labs/cbrt/>).

AUTHOR CONTRIBUTIONS

Conceptualization, YS; Methodology and Visualization, MK, TM, and YS; Investigation, MK; Resources, YS; Writing—Original Draft, MK; Writing—Review & Editing, MK, TM, and YS; Funding Acquisition, YS; Supervision, YS.

FUNDING

This work was supported by the RIKEN Pioneering Project Dynamic Structural Biology and MEXT/Kakenhi (Grant number 19H05645 to YS). MD simulations were carried out at RIKEN on HOKUSAI BigWaterFall.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.631854/full#supplementary-material>.

REFERENCES

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1-2, 19–25. doi:10.1016/j.softx.2015.06.001
- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., et al. (2010). PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 266 (Pt. 2), 213–221. doi:10.1107/S0907444909052925
- Ahmed, A., Whitford, P. C., Sanbonmatsu, K. Y., and Tama, F. (2012). Consensus among flexible fitting approaches improves the interpretation of cryo-EM data. *J. Struct. Biol.* 177, 561–570. doi:10.1016/j.jsb.2011.10.002
- Anandakrishnan, R., Drozdetski, A., Walker, R. C., and Onufriev, A. V. (2015). Speed of conformational change: comparing explicit and implicit solvent molecular dynamics simulations. *Biophys. J.* 108, 1153–1164. doi:10.1016/j.bpj.2014.12.047
- Bai, X. C., Rajendra, E., Yang, G., Shi, Y., and Scheres, S. H. (2015). Sampling the conformational space of the catalytic subunit of human γ -secretase. *eLife* 4, e11182. doi:10.7554/eLife.11182
- Bennett, M. J., Choe, S., and Eisenberg, D. (1994). Refined structure of dimeric diphtheria toxin at 2.0 Å resolution. *Protein Sci.* 3 (9), 1444–1463. doi:10.1002/pro.5560030911
- Bennett, M. J., and Eisenberg, D. (1994). Refined structure of monomeric diphtheria toxin at 2.3 Å resolution. *Protein Sci.* 3 (9), 1464–1475. doi:10.1002/pro.5560030912
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi:10.1063/1.448118
- Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., et al. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* 8 (9), 3257–3273. doi:10.1021/ct300400x
- Björkman, A. J., Binnie, R. A., Zhang, H., Cole, L. B., Hermodson, M. A., and Mowbray, S. L. (1994). Probing protein-protein interactions. the ribose-binding protein in bacterial transport and chemotaxis. *J. Biol. Chem.* 269 (48), 30206–30211. doi:10.1016/s0021-9258(18)43798-2
- Björkman, A. J., and Mowbray, S. L. (1998). Multiple open forms of ribose-binding protein trace the path of its conformational change. *J. Mol. Biol.* 279 (3), 651–664. doi:10.1006/jmbi.1998.1785
- Bonomi, M., Camilloni, C., Cavalli, A., and Vendruscolo, M. (2016). MetaInference: a Bayesian inference method for heterogeneous systems. *Sci. Adv.* 2 (1), e1501177. doi:10.1126/sciadv.1501177
- Bonomi, M., Heller, G. T., Camilloni, C., and Vendruscolo, M. (2017). Principles of protein structural ensemble determination. *Curr. Opin. Struct. Biol.* 42, 106–116. doi:10.1016/j.sbi.2016.12.004
- Bonomi, M., Pellarin, R., and Vendruscolo, M. (2018). Simultaneous determination of protein structure and dynamics using cryo-electron microscopy. *Biophys. J.* 114 (7), 1604–1613. doi:10.1016/j.bpj.2018.02.028
- Cossio, P., and Hummer, G. (2013). Bayesian analysis of individual electron microscopy images: towards structures of dynamic and heterogeneous biomolecular assemblies. *J. Struct. Biol.* 184 (3), 427–437. doi:10.1016/j.jsb.2013.10.006
- Darnault, C., Volbeda, A., Kim, E. J., Legrand, P., Vernède, X., Lindahl, P. A., et al. (2003). Ni-Zn-[Fe4-S4] and Ni-Ni-[Fe4-S4] clusters in closed and open α subunits of acetyl-CoA synthase/carbon monoxide dehydrogenase. *Nat. Struct. Biol.* 10 (4), 271–279. doi:10.1038/nsb912
- Dou, H., Burrows, D. W., Baker, M. L., and Ju, T. (2017). Flexible fitting of atomic models into cryo-EM density maps guided by helix correspondences. *Biophys. J.* 112 (12), 2479–2493. doi:10.1016/j.bpj.2017.04.054
- Fernandez-Leiro, R., Conrad, J., Scheres, S. H., and Lamers, M. H. (2015). cryo-EM structures of the *E. coli* replicative DNA polymerase reveal its dynamic interactions with the DNA sliding clamp, exonuclease and τ . *eLife* 4, 1–16. doi:10.7554/eLife.11134
- Goh, B. C., Hadden, J. A., Bernardi, R. C., Singharoy, A., McGreevy, R., Rudack, T., et al. (2016). Computational methodologies for real-space structural refinement of large macromolecular complexes. *Annu. Rev. Biophys.* 45, 253–278. doi:10.1146/annurev-biophys-062215-011113
- Grochowski, P., and Trylska, J. (2008). Continuum molecular electrostatics, salt effects, and counterion binding—A review of the Poisson-Boltzmann theory and its modifications. *Biopolymers* 89, 93–113. doi:10.1002/bip.20877
- Hamelberg, D., Mongan, J., and McCammon, J. A. (2004). Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* 120, 11919–11929. doi:10.1063/1.1755656
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14 (1), 33–38. doi:10.1016/0263-7855(96)00018-5
- Jung, J., Mori, T., Kobayashi, C., Matsunaga, Y., Yoda, T., Feig, M., et al. (2015). GENESIS: a hybrid-parallel and multi-scale molecular dynamics simulator with enhanced sampling algorithms for biomolecular and cellular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 5, 310–323. doi:10.1002/wcms.1220
- Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22 (12), 2577–2637. doi:10.1002/bip.360221211
- Kanaï, R., Ogawa, H., Vilsen, B., Cornelius, F., and Toyoshima, C. (2013). Crystal structure of a Na⁺-bound Na⁺,K⁺-ATPase preceding the E1P state. *Nature* 502 (7470), 201–206. doi:10.1038/nature12578
- Karanicolas, J., and Brooks, C. L. (2003). Improved ϕ -like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *J. Mol. Biol.* 334, 309–325. doi:10.1016/j.jmb.2003.09.047
- Keedy, D. A., Williams, C. J., Headd, J. J., Arendall, W. B., Chen, V. B., Kapral, G. J., et al. (2009). The other 90% of the protein: assessment beyond the Calphas for CASP8 template-based and high-accuracy models. *Proteins* 77 (Suppl. 9), 29–49. doi:10.1002/prot.22551
- Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E., and Kolinski, A. (2016). Coarse-grained protein models and their applications. *Chem. Rev.* 116 (14), 7898–7936. doi:10.1021/acs.chemrev.6b00163
- Kobayashi, C., Jung, J., Matsunaga, Y., Mori, T., Ando, T., Tamura, K., et al. (2017). GENESIS 1.1: a hybrid-parallel molecular dynamics simulator with enhanced sampling algorithms on multiple computational platforms. *J. Comput. Chem.* 38, 2193–2206. doi:10.1002/jcc.24874
- Kobayashi, C., Matsunaga, Y., Koike, R., Ota, M., and Sugita, Y. (2015). Domain motion enhanced (DoME) model for efficient conformational sampling of multidomain proteins. *J. Phys. Chem. B* 119, 14584–14593. doi:10.1021/acs.jpcc.5b07668
- Kolinski, A. (2004). Protein modeling and structure prediction with a reduced representation. *Acta Biochim. Pol.* 51 (2), 349–371. doi:10.18388/abp.2004_3575
- Krivov, G. G., Shapovalov, M. V., and Dunbrack, R. L. (2009). Improved prediction of protein side-chain conformations with SCWRLA. *Proteins* 77, 778–795. doi:10.1002/prot.22488
- Lindert, S., Alexander, N., Wötzel, N., Karakaş, M., Stewart, P. L., and Meiler, J. (2012). EM-fold: de novo atomic-detail protein structure determination from medium-resolution density maps. *Structure* 20, 464–478. doi:10.1016/j.str.2012.01.023
- Matthies, D., Dalmas, O., Borgnia, M. J., Dominik, P. K., Merk, A., Rao, P., et al. (2016). Cryo-EM structures of the magnesium channel CorA reveal symmetry break upon gating. *Cell* 164, 747–756. doi:10.1016/j.cell.2015.12.055
- Miyashita, O., Kobayashi, C., Mori, T., Sugita, Y., and Tama, F. (2017). Flexible fitting to cryo-EM density map using ensemble molecular dynamics simulations. *J. Comput. Chem.* 38 (16), 1447–1461. doi:10.1002/jcc.24785
- Miyazawa, S., and Jernigan, R. L. (1996). Residue—residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* 256, 623–644. doi:10.1006/jmbi.1996.0114
- Mori, T., Kulik, M., Miyashita, O., Jung, J., Tama, F., and Sugita, Y. (2019). Acceleration of cryo-EM flexible fitting for large biomolecular systems by efficient space partitioning. *Structure* 27, 161–174. doi:10.1016/j.str.2018.09.004 e3
- Morth, J. P., Pedersen, B. P., Toustrup-Jensen, M. S., Sørensen, T. L., Petersen, J., Andersen, J. P., et al. (2007). Crystal structure of the sodium-potassium pump. *Nature* 450, 1043–1049. doi:10.1038/nature06419
- Müller, C. W., Schlauderer, G. J., Reinstein, J., and Schulz, G. E. (1996). Adenylate kinase motions during catalysis: an energetic counterweight

- balancing substrate binding. *Structure* 4 (2), 147–156. doi:10.1016/s0969-2126(96)00018-4
- Müller, C. W., and Schulz, G. E. (1992). Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor Ap5A refined at 1.9 Å resolution. a model for a catalytic transition state. *J. Mol. Biol.* 224 (1), 159–177. doi:10.1016/0022-2836(92)90582-5
- Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P. M. G. E., et al. (2020). Single-particle cryo-EM at atomic resolution. *Nature* 587, 152–156. doi:10.1038/s41586-020-2829-0
- Olsson, S., Wu, H., Paul, F., Clementi, C., and Noé, F. (2017). Combining experimental and simulation data of molecular processes via augmented Markov models. *Proc. Natl. Acad. Sci. U.S.A.* 114 (31), 8265–8270. doi:10.1073/pnas.170480311410.1073/pnas.1704803114
- Onufriev, A., Bashford, D., and Case, D. A. (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* 55, 383–394. doi:10.1002/prot.20033
- Onufriev, A. V., and Case, D. A. (2019). Generalized born implicit solvent models for biomolecules. *Annu. Rev. Biophys.* 48, 275–296. doi:10.1146/annurev-biophys-052118-115325
- Orzechowski, M., and Tama, F. (2008). Flexible fitting of high-resolution x-ray structures into cryoelectron microscopy maps using biased molecular dynamics simulations. *Biophys. J.* 95 12, 5692–705. doi:10.1529/biophysj.108.139451
- Peng, J., Yuan, C., Ma, R., and Zhang, Z. (2019). Backmapping from multiresolution coarse-grained models to atomic structures of large biomolecules by restrained molecular dynamics simulations using Bayesian inference. *J. Chem. Theory Comput.* 15, 3344–3353. doi:10.1021/acs.jctc.9b00062
- Quioco, F. A., Spurlino, J. C., and Rodseth, L. E. (1997). Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure* 5 (8), 997–1015. doi:10.1016/s0969-2126(97)00253-0
- Rieping, W., Habeck, M., and Nilges, M. (2005). Inferential structure determination. *Science* 309, 303–306. doi:10.1126/science.1110428
- Roseman, A. M. (2000). Docking structures of domains into maps from cryoelectron microscopy using local correlation. *Acta Crystallogr. D Biol. Crystallogr.* 56, 1332–1340. doi:10.1107/s0907444900010908
- Rotkiewicz, P., and Skolnick, J. (2008). Fast procedure for reconstruction of full-atom protein models from reduced representations. *J. Comput. Chem.* 29, 1460–1465. doi:10.1002/jcc.20906
- Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23, 327–341. doi:10.1016/0021-9991(77)90098-5
- Schaefer, M., Bartels, C., and Karplus, M. (1998). Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.* 284, 835–848. doi:10.1006/jmbi.1998.2172
- Schlitter, J., Engels, M., and Krüger, P. (1994). Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J. Mol. Graph* 12, 84–89. doi:10.1016/0263-7855(94)80072-3
- Sharff, A. J., Rodseth, L. E., Spurlino, J. C., and Quioco, F. A. (1992). Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry* 31 (44), 10657–10663. doi:10.1021/bi00159a003
- Sugita, Y., Kitao, A., and Okamoto, Y. (2000). Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* 113, 6042–6051. doi:10.1063/1.1308516
- Taketomi, H., Ueda, Y., and Gö, N. (1975). Studies on protein folding, unfolding and fluctuations by computer simulation. I. the effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* 7 (6), 445–459. doi:10.1111/j.1399-3011.1975.tb02465.x
- Tama, F., Miyashita, O., and Brooks, C. L. (2004). Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J. Mol. Biol.* 337, 985–999. doi:10.1016/j.jmb.2004.01.048
- Terashi, G., and Kihara, D. (2018). De novo main-chain modeling for em maps using MAINMAST. *Nat. Commun.* 9 (1), 1618. doi:10.1038/s41467-018-04053-7
- Topf, M., Lasker, K., Webb, B., Wolfson, H., Chiu, W., and Sali, A. (2008). Protein structure fitting and refinement guided by cryo-EM density. *Structure* 16, 295–307. doi:10.1016/j.str.2007.11.016
- Touw, W. G., Baakman, C., Black, J., te Beek, T. A., Krieger, E., Joosten, R. P., et al. (2015). A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* 43, D364–D368. doi:10.1093/nar/gku1028 Database issue
- Toyoshima, C. (2008). Structural aspects of ion pumping by Ca²⁺-ATPase of sarcoplasmic reticulum. *Arch. Biochem. Biophys.* 476, 3–11. doi:10.1016/j.abb.2008.04.017
- Toyoshima, C., and Mizutani, T. (2004). Crystal structure of the calcium pump with a bound ATP analogue. *Nature* 430, 529–535. doi:10.1038/nature02680
- Toyoshima, C., and Nomura, H. (2002). Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature* 418, 605–611. doi:10.1038/nature00944
- Toyoshima, C., Nakasako, M., Nomura, H., and Ogawa, H. (2000). Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* 405, 647–655. doi:10.1038/35015017
- Toyoshima, C., Norimatsu, Y., Iwasawa, S., Tsuda, T., and Ogawa, H. (2007). How processing of aspartylphosphate is coupled to lumenal gating of the ion pathway in the calcium pump. *Proc. Natl. Acad. Sci.* 104, 19831–19836. doi:10.1073/pnas.0709978104
- Trabuco, L. G., Villa, E., Mitra, K., Frank, J., and Schulten, K. (2008). Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* 16 (5), 673–683. doi:10.1016/j.str.2008.03.005
- Villa, E., and Lasker, K. (2014). Finding the right fit: chiseling structures out of cryo-electron microscopy maps. *Curr. Opin. Struct. Biol.* 25, 118–125. doi:10.1016/j.sbi.2014.04.001
- Wang, R. Y.-R., Kudryashev, M., Li, X., Egelman, E. H., Basler, M., Cheng, Y., et al. (2015). De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nat. Methods* 12, 335–338. doi:10.1038/nmeth.3287
- Whitford, P. C., Ahmed, A., Yu, Y., Hennelly, S. P., Tama, F., Spahn, C. M. T., et al. (2011). Excited states of ribosome translocation revealed through integrative molecular modeling. *Proc. Natl. Acad. Sci. U.S.A.* 108, 18943–18948. doi:10.1073/pnas.1108363108
- Whitford, P. C., Geggier, P., Altman, R. B., Blanchard, S. C., Onuchic, J. N., and Sanbonmatsu, K. Y. (2010). Accommodation of aminoacyl-tRNA into the ribosome involves reversible excursions along multiple pathways. *RNA* 16, 1196–1204. doi:10.1261/rna.2035410
- Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., et al. (2018). MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* 27, 293–315. doi:10.1002/pro.3330
- Wriggers, W., Milligan, R. A., and McCammon, J. A. (1999). Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* 125, 185–195. doi:10.1006/j.sbi.1998.4080
- Wu, M., and Lander, G. C. (2020). How low can we go? structure determination of small biological complexes using single-particle cryo-EM. *Curr. Opin. Struct. Biol.* 64, 9–16. doi:10.1016/j.sbi.2020.05.007
- Yip, K. M., Fischer, N., Paknia, E., Chari, A., and Stark, H. (2020). Atomic-resolution protein structure determination by cryo-EM. *Nature* 587, 157–161. doi:10.1038/s41586-020-2833-4
- Zhou, R. (2003). Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins* 53, 148–161. doi:10.1002/prot.10483
- Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E., et al. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *elife* 7, e42166. doi:10.7554/eLife.42166

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Kulik, Mori and Sugita. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.