Special Issue
"Progress of Theoretical and Computational Biophysics"

Vol. 16, pp. 310–321 (2019)
doi: 10.2142/biophysico.16.0_310

Biophysics and Physicobiology

https://www.jstage.jst.go.jp/browse/biophysico/

Regular Article

Building a macro-mixing dual-basin Gō model using the Multistate Bennett Acceptance Ratio

Ai Shinobu[1], Chigusa Kobayashi[1], Yasuhiro Matsunaga[2] and Yuji Sugita[1,3,4]

[1]*Computational Biophysics Research Team, RIKEN Center for Computational Science, Kobe, Hyogo 650-0047, Japan*
[2]*Graduate School of Science and Engineering, Saitama University, Saitama 338-8570, Japan*
[3]*Theoretical Molecular Science Laboratory, RIKEN Cluster for Pioneering Research, Wako, Saitama 351-0198, Japan*
[4]*Laboratory for Biomolecular Function Simulation, RIKEN Center for Biosystems Dynamics Research, Kobe, Hyogo 650-0047, Japan*

Received July 1, 2019; accepted August 2, 2019

The dual-basin Gō-model is a structural-based coarse-grained model for simulating a conformational transition between two known structures of a protein. Two parameters are required to produce a dual-basin potential mixed using two single-basin potentials, although the determination of mixing parameters is usually not straightforward. Here, we have developed an efficient scheme to determine the mixing parameters using the Multistate Bennett Acceptance Ratio (MBAR) method after short simulations with a set of parameters. In the scheme, MBAR allows us to predict observables at various unsimulated conditions, which are useful to improve the mixing parameters in the next round of iterative simulations. The number of iterations that are necessary for obtaining the converged mixing parameters are significantly reduced in the scheme. We applied the scheme to two proteins, the glutamine binding protein and the ribose binding protein, for showing the effectiveness in the parameter determination. After obtaining the converged parameters, both proteins show frequent conformational transitions between open and closed states, providing the theoretical basis to investigate structure-dynamics-function relationships of the proteins.

Key words: molecular dynamics simulations, coarse-grained model, structure-based potential, protein conformational transitions, multiple basin potentials

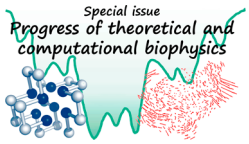
Corresponding author: Yuji Sugita, Theoretical Molecular Science Laboratory, RIKEN Cluster for Pioneering Research, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.
e-mail: sugita@riken.jp

Proteins often undergo conformational transitions during their function [1–5]. In many cases, the binding or unbinding of a ligand induces a large amplitude domain-level transition between active and inactive states [6–11]. Elucidating the mechanism for such conformational transitions connects protein structural information with functional understanding and opens the way for interference with existing processes and design of new compounds.

Molecular dynamics (MD) simulations are widely used to investigate conformational dynamics of proteins [12–18]. Recent advancement in the development of high-performance computers has sparked the interest in simulating the dynamics of systems of enormous size scales such as large biomolecular complexes and biomolecules in cellular environment

◀ ***Significance*** ▶

Structure-based coarse-grained (CG) Gō models are used with great success to sample protein dynamics on long time scales. Here, we combine two Gō type single-basin potentials to sample conformational transitions between two states of a protein. CG models are highly parameter-dependent and detecting appropriate parameters is a time-consuming task. In this study, we propose a scheme for parameter refinement of CG models using the Multistate Bennett Acceptance Ratio (MBAR), a statistical weighing method. Using MBAR, free energy profiles can be calibrated without simulations. We demonstrate conformational transitions between two reaction states of soluble proteins by applying the scheme.



[19–23]. However, all-atom MD simulations are able to simulate conformational dynamics that happen for several tens of microseconds using conventional PC-clusters or supercomputers. For reaching the millisecond time scale, in which large-scale conformational motions of proteins occur, MD simulations based on coarse-grained (CG) models [24–29] are more suitable [24,30–33].

Structure-based off-lattice Gō models [34–36] are often used in CG MD simulations with remarkable success [36–44]. Based on the original Gō model [45–47], the potential function used in such simulations consists of an attractive potential between non-bonded contacts that exist in the native structure (native contacts) and a repulsive potential between all other non-bonded pairs. The simulations realize funnel-shaped free energy landscapes for protein folding and dynamics, successfully explaining many experimental results [48–50]. A revised model was developed by Karanicolas and Brooks [51,52], in which residue-type-dependent knowledge-based energy terms (the Miyazawa-Jernigan contact energies [53]) are applied to the native contacts (hereafter, we refer to it as KBGo model). Folding pathways for proteins predicted using this model were in excellent agreement with experimental results. Moreover, the robustness of the folding pathways against competing non-native interactions was demonstrated. The Domain Motion Enhanced (DoME) model [54], developed more recently, emphasizes inter-domain motions while keeping intra-domain regions relatively rigid. This model is, in particular, useful for simulating large-scale domain motions of proteins, by making the simulation robust against temperature changes.

The off-lattice Gō models presented above predominantly stabilize a single conformation and are referred to as single-basin models. For simulating transitions between two stable states, two single-basin potentials are mixed to form a dual-basin potential. The mixing of potentials can be obtained using two main approaches, the microscopic (micro-mixing) and macroscopic (macro-mixing). In micro-mixing [39,55–57], energetic terms are added as individual terms to form the mixed potential. The mixed potential is in fact a single-basin potential of one state, to which contact energies of the other state are added as perturbations. In macro-mixing [37,38,43,58,59], the two single-basin potentials are predefined and all energetic terms are coupled to each other in the mixed potential. The micro-mixing approach is suitable for systems with a high degree of similarity, in which transitions between the two states can be described by addition of contacts. The macro-mixing approach is often successful in describing large-scale transitions with little overlap between the contact sets of the two states. For example, the sheet to helix transition in Arc repressor was successfully described using the macro-mixing potential [37].

The exponential Boltzmann weighing [37,43,60] (exponential mixing) is a macro-mixing approach in which a dual-basin potential is constructed by summing the partition functions of the two single-basin potentials. In the superposition approach [38,59], the mixed potential is resolved by solving an eigenvalue equation. Both schemes ensure a smooth transition between the two basins. Structure-based CG models in general heavily rely on system-dependent parameters, which are determined properly for successful MD simulations. Conventionally, the parameter determination in exponential mixing is done by trial and error, by performing multiple rounds of short simulations with different parameters while assessing the frequency of conformational transitions. This laborious process can be very time consuming, especially for large systems or for potentials in which more than a single parameter needs to be determined.

Here, we apply the Multistate Bennett Acceptance Ratio (MBAR) analysis method [61] for an efficient determination of parameters in the exponential-mixing dual-basin potential. The scheme that we propose consists of short simulations with multiple candidates of mixing parameters and an estimation of improved parameters based on MBAR. MBAR statistically calculates the optimal estimators for computing free energies of unsampled data using sampled data at different conditions. Without performing an enormous number of simulations, this scheme can sample the parameter space efficiently and allows a severalfold faster parameter determination. We apply this scheme to the dual-basin Gō simulations of the Glutamine Binding Protein [6] (GBP, Fig. 1) and the Ribose Binding Protein [62,63] (RBP, Supplementary Fig. S1), and discuss the quality of the predicted parameters and the stability of the extended sampling simulations.

Theory and Methods

The potential energy function of the dual-basin structure-based Gō model

We use the single-basin KBGo potential [51,52] as described in the original works [37,43]. The DoME model [54] potential carries the same functional form as the KBGo potential. It differs only in the non-bonded native contact parameters, which are set inversely proportional to the magnitude of inter-domain motions (and a constant value for intra-domain contacts). Domain classification is obtained from the Motion Tree calculation [64,65].

The macro-mixing approach of exponential Boltzmann weighing [37] (exponential mixing) for mixing two single-basin potentials, which we use in the study, is shown in Eq. (1).

$$\exp\left(-\frac{1}{k_B T_{\mathrm{mix}}} E(\boldsymbol{R})\right) = \exp\left[-\frac{1}{k_B T_{\mathrm{mix}}}\left(V_1(\boldsymbol{R})+C\right)\right] + \exp\left[-\frac{1}{k_B T_{\mathrm{mix}}} V_2(\boldsymbol{R})\right], \tag{1}$$

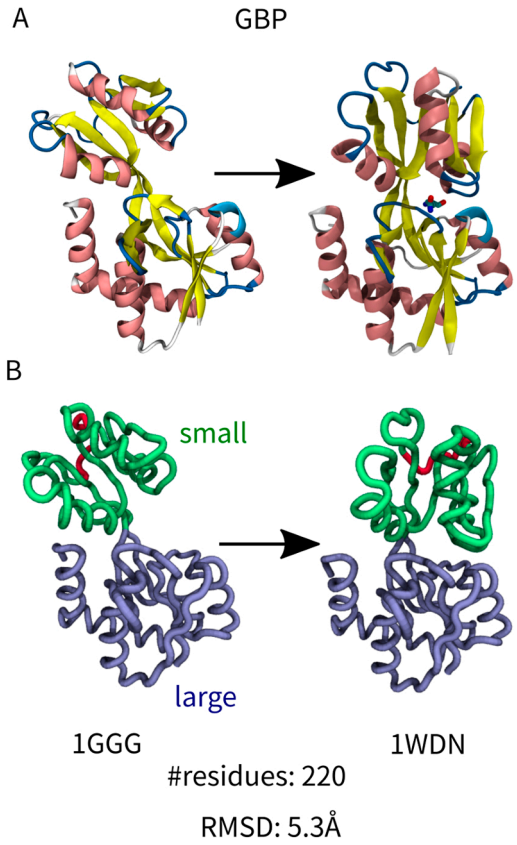


Figure 1 Structures of GBP in the unbound (left side at each panel) and bound (right side) forms, colored according to secondary elements (A) or domains (B). PDB codes written below the protein structures. RMSD values were calculated between the unbound and bound forms, fitted by their Cα atoms. The bound form in panel A shows the bound ligand (glutamine) in a stick representation. Domain classification in panel B was performed according to the Motion Tree [64,65]. Domain name notations were taken from Ref. [6].

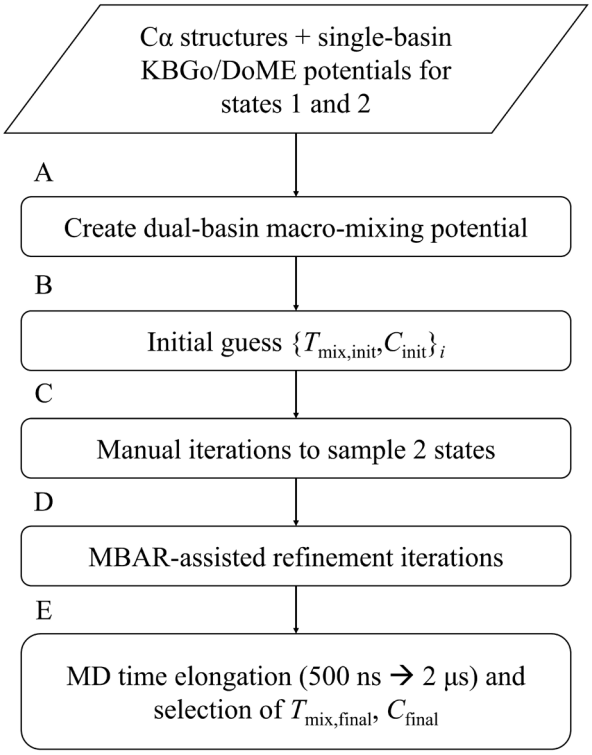


Figure 2 Overall flowchart describing the procedure presented in this work. Detailed schemes are shown in Supplementary Figures S2 and S3. Letters assigned to each stage are compatible to naming of stages in Supplementary Figures S2 and S3.

where k_B is the Boltzmann constant, V_1 and V_2 are the single-basin potentials to be mixed, and T_{mix} and C are the mixing parameters. T_{mix} is related to the barrier height between the two basins (a lower value corresponds to a higher barrier). C determines the energetic offset between the basins. The mixed potential function was implemented while adhering to all the details described in the original works [37,43]. We define the open (OP) basin as basin 1 and the closed (CL) basin as basin 2.

Parameter searching

Given the structures in the OP and CL states, we need to determine a pair of parameters (T_{mix}, C), which allow proper transitions between the two states under the dual-basin exponential mixing potential. The overall scheme for determining the mixing parameters is presented in Figure 2. Our workflow consists of three major stages that are: the manual detection of an approximate range of proper parameters (stage C and Supplementary Fig. S2), the MBAR-assisted parameter refinement (stage D and D1–D11 in Supplementary Fig. S3), and the final selection of parameters (stage E and E1–E4 in Supplementary Fig. S3). In order to estimate the potential of mean force (PMF) from MBAR, both OP and CL states as well as transitions between them needs to be sampled. Thus, manually detecting an approximate parameter range is required (stage C) prior to using the MBAR analysis for the parameter refinement (stage D). In both stages C and D, the approach is to iteratively perform simulations with a set of parameters, asses their quality, set modified parameters, and use them for the next round.

We use the distance root-mean-square displacement [66] (dRMS) as a reaction coordinate for quantifying the similarity of frames during the simulations to the OP ($dRMS_{\mathrm{OP}}$) and CL ($dRMS_{\mathrm{CL}}$) native structures. dRMS is calculated between atom pairs, which in the reference state were separated by at least four atoms in sequence and between 6–50 Å in space, and for which the difference in distance between the two reference states was larger or equal to 5 Å. Along the manuscript, time-series of $dRMS_{\mathrm{CL}}$ is used for visualizing state populations during each simulation.

The equilibrium constant K_{eq} between the CL and OP states is defined as:

$$K_{\mathrm{eq}} = \frac{p_{\mathrm{CL}}}{p_{\mathrm{OP}}}, \tag{2}$$

where p_{CL} and p_{OP} are the populations in the CL and OP states, respectively. The populations at each $dRMS_{CL}$ bin are obtained from the PMF along $dRMS_{CL}$, using the relation $dG=-k_BT\ln p$. Trajectory frames are assigned to either the OP or CL states according to their $dRMS_{CL}$ values, where the cutoff value is determined at the location of the barrier in between the two basins of the PMF plot. It is possible to determine proper mixing parameters for any target K_{eq}. In this study, we set the target K_{eq} value to 1.

Manual parameter search until two states are sampled

At this stage (C in Fig. 2 and Supplementary Fig. S2), we detect T_{mix} and C for an approximate range, which produces transitions between the two states within a single simulation. We start by using large increments of extreme values and gradually narrow the search range by reducing the increments until an acceptable range is detected. At each round, we perform a series of simulations (five in this case) with different sets of parameters, and initiate simulations from both the OP and CL states (ten simulations in total). T_{mix} or C are fixed alternatively at each round, while the other parameter is varied.

Parameters for the next round are determined according to the behavior of the current round by varying their values according to the direction which leads to proper state transitions. In case one of the states is sampled in excess, C is varied in the direction which stabilizes the other state, or in case transition frequency is too low, T_{mix} is increased.

MBAR-assisted parameter refinement

Once two states are sampled, the MBAR analysis method is used for speeding up the parameter search process (D in Fig. 2 and D1–D11 in Supplementary Fig. S3).

The MBAR method [61] allows to estimate the expectation of any physical quantity $\langle A(\boldsymbol{x})\rangle$ (in this case, the PMF) for an unsimulated condition by reweighing simulated data performed under various conditions (in this study, conditions refer to combinations of T_{mix} and C).

The first step of MBAR method estimates the free energies $\hat{f}_i$ of simulated conditions by iteratively solving the following nonlinearly coupled equations:

$$\hat{f}_i = -\ln \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{\exp[-u_i(\boldsymbol{x}_{jn})]}{\sum_{k=1}^{K} N_k \exp[\hat{f}_k - u_k(\boldsymbol{x}_{jn})]} \tag{3}$$

where i and j indicate conditions, and K is the total number of conditions. N_j is the total number of configurations (simulation frames) in condition j. u_i is the reduced potential energy defined as:

$$u_i(\boldsymbol{x}) = \beta E_i(\boldsymbol{x}) = \frac{1}{k_B T} E_i(\boldsymbol{x}), \tag{4}$$

k_B is the Boltzmann constant, and T is the simulation temperature. $E_i(\boldsymbol{x})$ is the potential energy for condition i (here, combinations of T_{mix} and C). We mark this step as MBAR1.

In the next step, we can estimate the expectation of a physical quantity $\langle A(\boldsymbol{x})\rangle$ for an unsimulated target condition (marked by 'estim' and a hat symbol) by calculating the ratio of partition functions using the results of eq. (3) [61],

$$\langle \widehat{A(\boldsymbol{x})} \rangle_{estim} = \frac{\int A(\boldsymbol{x}) \exp[-u_{estim}(\boldsymbol{x})] d\boldsymbol{x}}{\int \exp[-u_{estim}(\boldsymbol{x})] d\boldsymbol{x}} = \sum_j \sum_n w_{jn} A(\boldsymbol{x}_{jn}) \tag{5}$$

Here, for the estimation of the PMF, indicator functions are employed for $A(\boldsymbol{x})$ along the axis of the dRMS. $u_{estim}(\boldsymbol{x})$ is the reduced potential energy in the unsimulated target condition in which we attempt to estimate the PMF. w_{jn} is the weight of configuration n in condition j with respect to the target condition. $\langle \widehat{A(\boldsymbol{x})} \rangle_{estim}$ is the estimated probability density defined as the expectation of an indicator function. We mark this step as MBAR2. MBAR1 and MBAR2 were performed using the MATLAB MDToolbox (https://github.com/ymatsunaga/mdtoolbox) [67].

The procedure in the MBAR-assisted refinement stage follows a similar progression as the preliminary manual stage. However, instead of manually guessing the parameters for the next round and confirming by MD simulation, we perform several iterations of MBAR2. Once the weight factors are obtained by solving Eq. (3) using simulation trajectories, iterations for estimation of parameters can be performed without simulations (D2–D7 in Supplementary Fig. S3). The inclusion of MBAR-assisted parameter estimation reduces the number of MD simulation rounds with unsuccessful parameter choice.

The K_{eq} value for an estimated parameter set (K_{MBAR}) is obtained from the estimated PMF (PMF_{MBAR}). This process of guessing T_{mix} and C and estimating K_{MBAR} is performed iteratively until the target K_{eq} (K_{target}) is included in the range of K_{MBAR}.

The results of each simulation round are assessed and a decision is made on whether to end the parameter search (step D10 in Supplementary Fig. S3). For each of the simulations in the round, in addition to K_{sim} values, $dRMS_{CL}$ and $dRMS_{OP}$ are calculated at the free energy minimum of each of the two basins. The conditions for ending the refinement stage are as follows: 1) K_{sim} for all simulations are in an acceptable range around K_{target} (in this case we used the range 0.2–5), 2) There is at least one pair of T_{mix} and C for which K_{sim} is close enough to K_{target} (in this case we allowed a deviation of 25% from K_{target}), and 3) both basins' minima are located outside the intermediate basin region in the ($dRMS_{CL}$, $dRMS_{OP}$) space, defined as the rectangular area confined by 2 Å at the bottom limit and the dRMS value between the native states in the top limit. This ensures that the simulated ensembles do not sway too far from the native states.

Final selection of mixing parameters

Once a satisfactory parameter pair is detected in stage D, all simulations of the last round are elongated to 2 μs, K_{sim}

values are calculated, and the parameter pair used for the simulations for which K_{sim} values have the smallest average (between the OP and CL simulations) deviation from K_{target} is selected as the final parameter set.

Simulation details

For both target systems GBP and RBP, crystal (native) structures of the OP (unbound) and CL (bound) states are available (PDB codes: 1GGG/1WDN [6] and 1URP [63]/2DRI [62] for OP/CL of GBP and RBP, respectively, Fig. 1A and Supplementary Fig. S1A). Residues 5-224 were used for GBP (one residue from the N-terminus and two from the C-terminus were removed from the CL structure) and 1–271 for RBP. dRMS values between the native structures were 8.964 Å and 7.125 Å for GBP and RBP, respectively. The Cα models were created and single-basin KBGo potentials were built using the MMTSB server (https://mmtsb.org/) [51,52].

Initial values for T_{mix} and C were set in ranges of 15000 to 30000 and −50 to +50, respectively. Values were chosen based on a previous study [43], in which macro-mixing parameters for Adenylate Kinase (AdK) were determined.

Simulations were performed using the MD program GENESIS [68,69], in which both single-basin and dual-basin potential schemes were implemented. Simulations were performed in the *NVT* ensemble using the Langevin thermostat. The timestep of integration was 0.02 ps. All bonds were constrained using the SHAKE [70] algorithm. Native interactions were calculated without truncation, whereas repulsive non-native interactions were truncated at a distance of 20 Å.

The simulation temperature in the dual-basin potential was set to ~0.9 T_f. T_f is the folding temperature of the protein, and was determined in a 500 ns single-basin potential simulation in the KBGo model as the temperature for which the protein shares its time equally between the folded and unfolded states. Unfolding was identified using the RMSD of the protein to its own initial native structure. If the OP and CL simulations exhibite different T_f values, the lower value of the two is used. Dual-basin simulations were performed at 240 K and 260 K for GBP and RBP, respectively. After identifying the parameters which satisfy the mixing criteria, we performed a 10 μs sampling simulation with the final parameters. The final parameters and K_{eq} calculated for the sampling simulations are summarized in Table 1 for all four cases.

Table 1 Final parameters for macro-mixing simulations

System	Property	KBGo	DoME
GBP	C_{final}	−22.4	−29.0
	$T_{mix,final}$	21000	25000
	$K_{sim,OP/CL}$	1.25/1.25	1.00/1.12
RBP	C_{final}	−1.12	−11.25
	$T_{mix,final}$	4500	7000
	$K_{sim,OP/CL}$	0.85/0.86	1.15/0.89

Results and Discussion

Acceleration of parameter searching using MBAR

Figure 3 shows the $dRMS_{CL}$ time-series for the parameter search rounds (C–E in Fig. 2 and Supplementary Figs. S2 and S3) for GBP in the KBGo model. For this system, two manual iterations (stage C) were required for sampling two states, and five rounds of MBAR-assisted iterations and an additional 2 μs elongation for identifying $T_{mix,final}$ and C_{final}.

In round 1 (stage C1 in Supplementary Fig. S2), observing $dRMS_{CL}$ values, a few conformational transitions between the OP and CL states occur in the simulations with C=−25, while for C=−50 the protein resides solely in the OP state, and for C=0, +25, +50 solely in the CL state. We selected C=−25 and executed simulations with different T_{mix} values in round 2 (stage C5). Transitions are observed at simulations with T_{mix}=15000 (only OP simulation), 20000, and 25000 (both OP and CL). At this stage, the procedure advances to the MBAR-assisted simulations rounds. In rounds 3–7 (stages D2-D11 in Supplementary Fig. S3), we performed simulations with parameters selected by estimating PMF_{MBAR} and K_{MBAR} using $\hat{f}_i$ given from the previous simulations (stage D8). During these rounds, conformational transitions are observed 50–100 times per simulation. In round 7, criteria for ending the search are fulfilled where a pair of simulations produced K_{sim} values within the desired range (encircled). In round 8, the set of simulations with identical conditions to round 7 were performed for 2 μs, after which $T_{mix,final}$ and C_{final} were selected according to the simulation pair that produced K_{sim} values closest to K_{target} (encircled). A single simulation of 500 ns required roughly 100 minutes to complete on 8 cores of Intel® Xeon® E5-2680 v3 CPUs. The available computing resources enable running all 10 simulations of a round in parallel. Thus, detecting $T_{mix,final}$ and C_{final} (not including the 10 μs-long sampling) for the current system required approximately 18 hours.

The MBAR2 iterations (D2–D7 in Supplementary Fig. S3) serve as substitutes for MD simulations and therefore constitute the core of the speedup achieved in the current method. The number of MBAR2 rounds for the current system was between 2-4 per round (data not shown). Assuming that each MBAR2 iteration is equivalent to one round of MD simulation, we would need to perform from 10 (2 MBAR2 rounds ×5 rounds of MBAR-assisted rounds) to 20 (4×5) additional simulation rounds per system. Therefore, without using the MBAR analysis, we estimate that an extra simulation time of 50–100 μs will be required. Using the MBAR analysis results in substantial acceleration of the parameter searching process. This effect is expected to be even more prominent for larger systems in which physical simulation times are severalfold longer. The MBAR analysis calculation time is independent of system size and is negligible comparing to MD simulation times, thus the reduction in total calculation time is expected to be even greater.

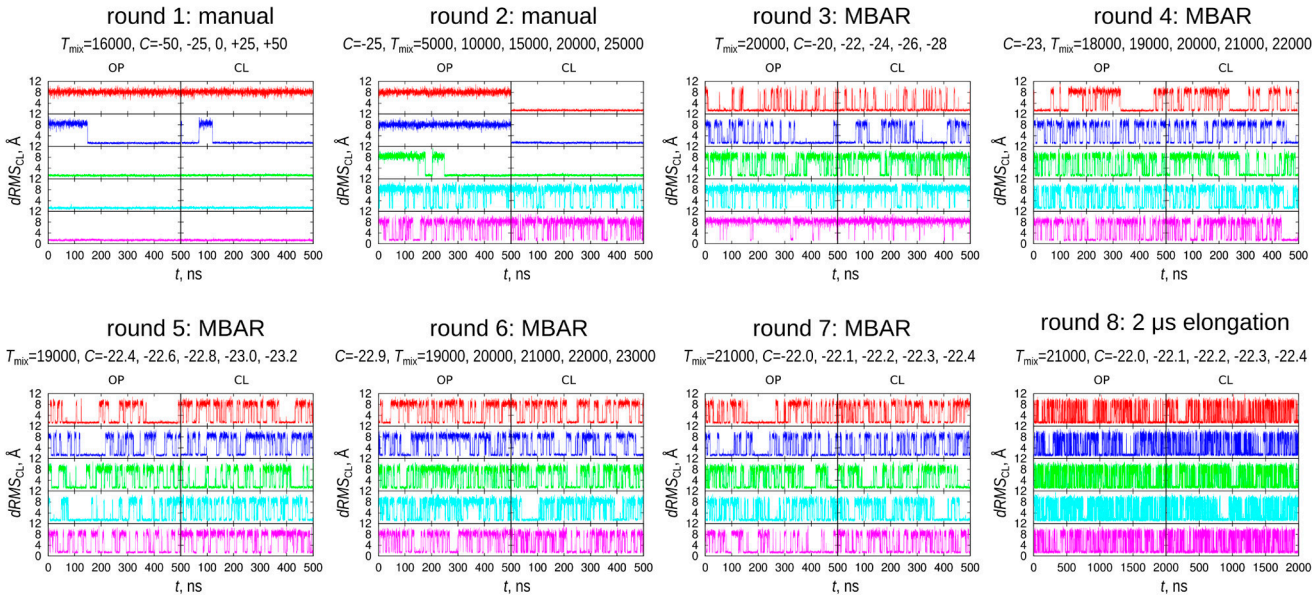


Figure 3 Time-series of $dRMS_{CL}$ for simulation rounds performed for determining the macro-mixing parameters for GBP in the KBGo model. Five simulations per initial structure (OP and CL, left and right panels at each round, respectively) were performed at each round. At each round one parameter was fixed for all simulations (a single value which is written above the plot) and the other was varied (five values written in order of top to bottom in the dRMS plots). The length of the simulation for all rounds except round 8 was 500 ns. "manual", "MBAR" and "2 μs elongation" stand for the three stages along the parameter tuning process. "manual" for the manually adjusting the parameters until two states are sampled (stage C in Fig. 2). "MBAR" for stages in which parameters were estimated using MBAR analysis using simulation data of the previous stage (stage D), and "2 μs elongation" for elongating the last "MBAR" round (round 7) to confidently determine a single set (T_{mix}, C) of optimal parameters (stage E).

Characterizing the quality of the MBAR-prediction

PMFs along $dRMS_{CL}$ and the values of K_{eq} calculated from the PMFs (K_{sim}) are shown for each round of the parameter searching process in Figure 4. In rounds 1–2, PMF_{sim} does not cover both the OP and CL states and K_{sim} values (in cases where they could be calculated) are far from the targeted K_{eq}. MBAR-predicted PMF (PMF_{MBAR}) plots are shown in black for rounds 3-8. In round 3, there is poor overlap between PMF_{sim} and PMF_{MBAR}. The overlap becomes progressively better with the rounds. Figures 5A and 5B present values for K_{sim} and the difference between K_{sim} and K_{MBAR}, respectively. From the figures we learn that the MBAR-assisted search drives K_{sim} closer to K_{target} and K_{sim} closer to K_{MBAR}.

We characterize the quality of the MBAR prediction by the similarity between PMF_{MBAR} and PMF_{sim}. There are two aspects which influence the quality of prediction. The first is the similarity between the parameters with which the weight factor $\hat{f}_i$ is calculated (data from the previous simulation round) and the parameters for which MBAR2 predicts the PMF for (to be used in the current simulation round). Higher similarity between the two parameter sets results in a more accurate prediction. Indeed, values of T_{mix} ranging from 5000 to 25000 were used for estimating PMF_{MBAR} in round 3, whereas a much smaller T_{mix} range (18000–23000) was used for all later rounds. This illustrates the importance of narrowing the parameter range for achieving accuracy in the prediction.

Another aspect regarding the quality of prediction lies in the frequency of transitions between the basins. When one basin is exceedingly stabilized with respect to the other, transitions to the less stable basin are difficult and the chances of the simulation residing in the stable basin for prolonged times increase. This brings to uneven sampling, which results in poor convergence between the two simulations as well as with predicted properties (PMF_{MBAR}). This is observed during the MBAR-assisted parameter searching where parameters which produce large or small K_{MBAR} values are used. For example, the top simulation in round 5 in Figure 4 for which K_{MBAR}=1.52 and the two K_{sim} values deviate both from each other and from K_{MBAR} (2.66 and 0.74 for the OP and CL simulations, respectively). This issue can be resolved by sufficient simulation time. Thus, if the desired K_{target} is much larger or smaller than 1, longer simulation times should be used for an accurate parameter determination.

We start the search with a wide range of parameters (5000 to 25000 and −50 to +50 for T_{mix} and C, respectively), and gradually narrow the search until the desired behavior (K_{eq}) is obtained with good enough precision, finally converging to parameter values of 21000 and −22.4 for T_{mix} and C, respectively. This ensures the robustness of the method in a sense that the final answer does not depend on the choice of initial guess. Structure-based models such as the one used here are constructed from structural information of the protein. The number of contacts, as well as the boundaries of the rigid domains are different from system to system. Therefore, the parameter values differ according to the system.

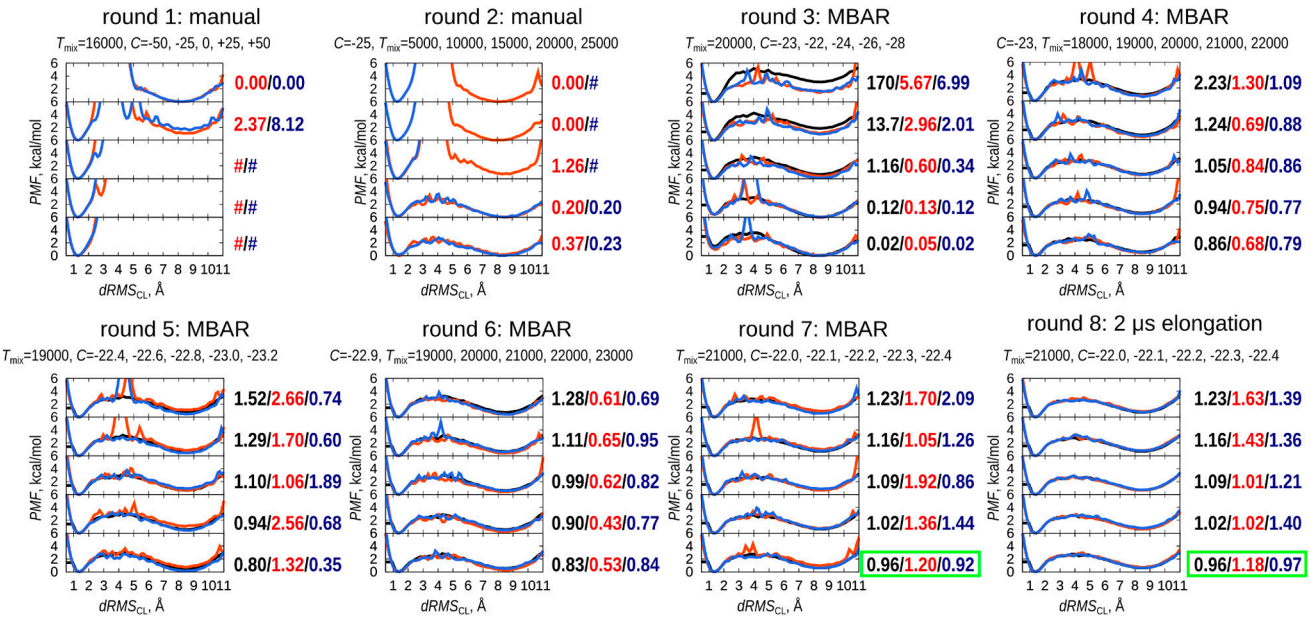


Figure 4 PMF vs. $dRMS_{CL}$, presented for simulation rounds performed for determining the macro-mixing parameters for GBP in the KBGo model. Panels representing each round correspond to those in Figure 3. Black curves ("MBAR", for rounds 3 to 8) represent PMF for each parameter set (T_{mix}, C), estimated by the MBAR analysis performed using data from the previous round (PMF_{MBAR}). Red (OP) and blue (CL) curves represent PMF calculated from simulation data of the current round (PMF_{sim}). K_{eq} values per each individual simulation are written on the right side of their respective PMF plots. For the three bottom simulations of round 1 the OP state was not sampled, thus K_{eq} is not defined (marked as "#"). Encircled text represent K_{sim} values which satisfied the criterion for ending the MBAR-assisted search stage (round 7) and the K_{sim} values of the simulations with parameters selected as $T_{mix,final}$ and C_{final}.

A robust method, such as the one presented here can be advantageous for constructing a suitable structure-based model.

Figure 6 presents free energy minima near each basin in the $dRMS_{CL}/dRMS_{OP}$ space. The most populated configuration near each basin (the configuration with the lowest free energy) does not penetrate the "intermediate region" (the grey area) for all but one simulation (in round 7 using C=−22.3). This indicates that the original native states are sampled accurately along all simulation rounds. Near the OP basin, most simulations sample configurations closer to the CL state than the native OP structure, whereas near the CL basin, simulations are spread evenly around the native configuration.

Characterizing the behavior of the mixed potential in the extended sampling simulation

Figure 7 and Table 2 display various properties calculated for the 10 μs-long sampling simulations with $T_{mix,final}$ and C_{final}. The time-series of $dRMS_{CL}$ in Figure 7A displays frequent transitions between the two states. The transition rate was calculated as 157 and 156 μs^{-1}, showing an excellent convergence between the OP and CL simulations. The total number of transitions is above 1500 for both simulations, which is sufficient for accurately calculating properties which characterize the systems' behavior. Maximal residence times were calculated as 115 and 160 ns for the OP and CL simulations, respectively, constituting under 2% of the whole simulation time. Transition rates obtained in the current study are higher than those of actual conformational transitions (which occur at time ranges of milliseconds). Indeed, simulating realistic transition rates is still challenging with the currently available computational resources. In the current work, we focus on simulating enough transition events that will allow us to characterize transition paths, or to treat very large systems.

Figure 7B shows two-dimensional free energy surfaces in the $dRMS_{CL}/dRMS_{OP}$ space. The dual-basin simulations near the CL basin sample configurations very similar to those sampled in the single-basin simulation, with the most populated configuration approximately overlapping with that in the single-basin simulation (shown as black circles). For the OP state, the most populated configuration is located closer to both the CL and the OP native states than the single basin simulations. This is likely due to the fact that under the single-basin potential, fluctuations towards the opposite direction (excessive opening in the OP state) are larger than in the dual-basin where the two domains are held more tightly against each other.

An estimated path for the states transition can be detected in Figure 7B as the line connecting the two states. From the dRMS time-series (Figs. 3 and 7A) it is apparent that coordinates shift instantaneously without sampling intermediate states. Thus, the intermediate states visible in panel B are not states which were sampled during transitions, but rather fluctuations within each basin and the most probable path

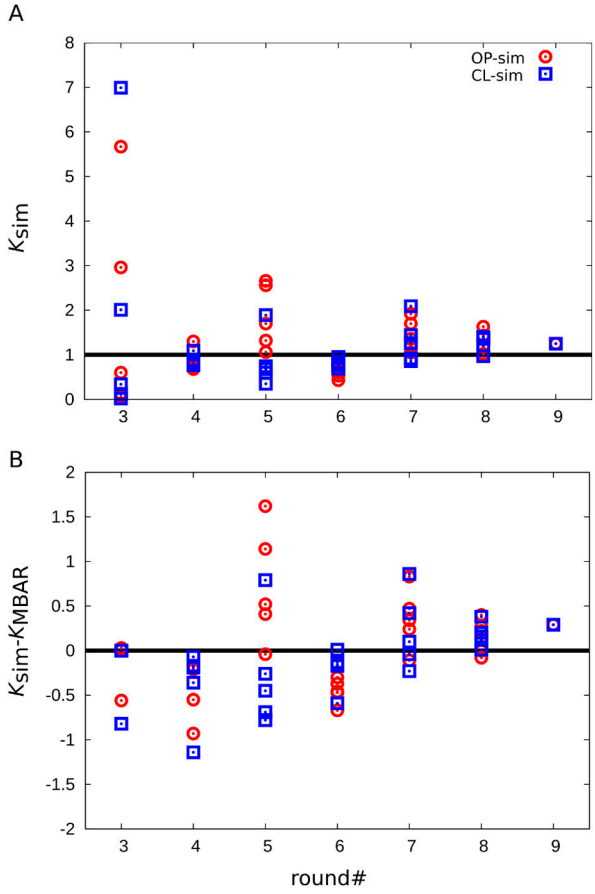


Figure 5 K_{sim} (A) and K_{sim}-K_{MBAR} (B) for rounds 3-9 of the GBP KBGo parameter search (round numbers are according to the notation in Figure 3, round 9 is the 10 μs sampling simulations with $T_{mix,final}$, C_{final}). Each point represents calculated value from one simulation. Data from simulations starting from the OP and CL states are shown in red and blue, respectively. Horizontal lines represent the value of K_{target} (A), and the location of '0' (B). For round 3 in B: four of the data points have a very large difference between the predicted K_{MBAR} and simulated K_{sim} (−165, −163, −10.8, and −11.7) and are not shown in the figure.

is estimated as the line connecting the fluctuations. Below panel B, representative structures are shown for the OP and CL most populated configurations and for an intermediate state, represented by a configuration with *dRMS* values residing midway between the two native states.

Free energy landscapes of GBP [71] and RBP [72] obtained from simulations using all-atom force fields show two basins near the crystal structures with no intermediate states in between the basins. In this aspect, our model resembles the results of the all-atom simulations.

Final K_{sim} values were 1.25 for both the OP and CL simulations (Table 2), slightly above the MBAR-predicted value of 0.96, but still within the allowed deviation range. Figure 7C shows PMF plots for the sampling simulation after 0.5, 1, 2, 4, and finally 10 μs. We observe that the OP and CL PMFs converge as the simulation time increases. Also, the transition state becomes progressively more defined. It takes approximately 4 μs until full convergence of the two simulations is observed. MBAR-assisted parameter determination was performed with 0.5 μs-long simulations, where convergence is not full. For obtaining a more accurate prediction of K_{eq}, one should either increase the length of each individual MBAR-assisted round simulation, or increase the number of MBAR-assisted rounds while modifying the guessed parameters in smaller increments.

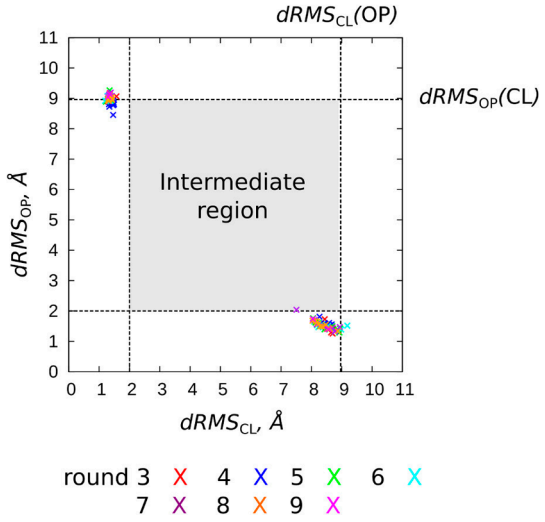


Figure 6 Location of the free energy minima near the OP (bottom right) and CL (top left) basins. Data for different simulation rounds are shown in different colors. The gray area is the rectangle confined within 2 Å from the bottom and the dRMS between the native structures from top (the "intermediate region").

The effect of system and model on the mixing behavior

We also applied the procedure to GBP in the DoME model and to RBP in the KBGo and the DoME models. The number of required rounds and simulation times are shown for all four systems in Supplementary Tables S1 and S2.

Supplementary Figure S4A shows PMF plots for the 10 μs-long simulations with $T_{mix,final}$ and C_{final} for GBP and RBP in the KBGo and DoME models. The dual-basin potential is asymmetric for all cases, with the OP basin higher in energy and broader than the CL basin. K_{sim} values close to 1 imply equal populations between the two basins. We note the relation between the basins' relative heights and the breadths where the narrower basin will be lower in energy in order to compensate for a smaller number of available configurations. This effect is demonstrated in Supplementary Figure S4A in which PMFs of GBP (top) and RBP (bottom) are displayed, and the OP basin in RBP appears broader and higher in energy than the GBP OP basin.

The position of the two single-basin potentials relative to each other is controlled by the mixing parameters, which are optimized to produce the desired mixing behavior. However, the shapes of the individual basins are innate to the system and carried on from the single-basin potentials. As noted earlier, in RBP, the OP basin is broader than in GBP (Sup-

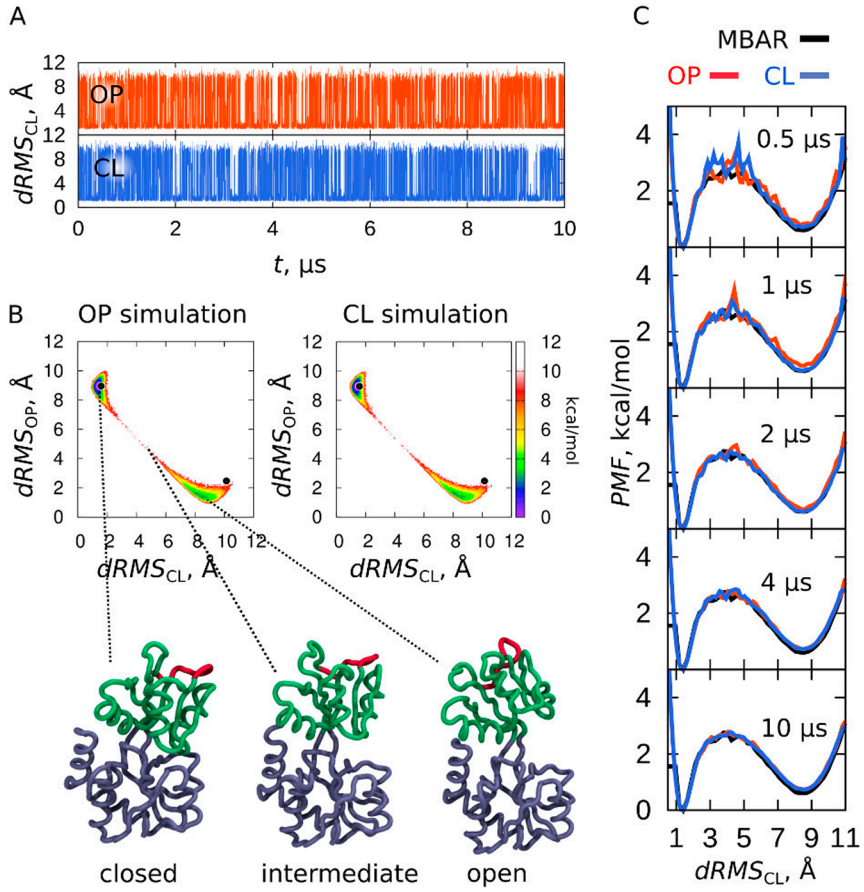


Figure 7 Results for the 10 μs sampling simulations using the final macro-mixing parameters for GBP with the KBGo model. A) Time-series of $dRMS_{CL}$ for simulations starting from the OP (top) and the CL (bottom) states. B) Two-dimensional PMF surfaces along $dRMS_{CL}$ (x-axis) and $dRMS_{OP}$ (y-axis). Black circles represent dRMS values of the most probable structures from the single-basin simulations of the OP and CL states (structure closest to the free energy minimum near each basin). Representative structures are shown for the unbound (OP), bound (CL) for the frame closest to the free energy minimum of each basin, and for an intermediate state residing along the path connecting the end states. Domain coloring corresponds to Figure 1B. C) PMF vs. $dRMS_{CL}$, calculated for 0.5, 1, 2, 4, and 10 μs along the sampling simulation.

Table 2 Properties calculated for 10 μs simulations of GBP in KBGo model using $T_{mix,final}$ and C_{final}

Property	OP simulation[a]	CL simulation[a]
K_{sim}	1.25	1.25
Transition rate[b,c], μs^{-1}	157	156
$dRMS_{near,OP}$[b,d], Å	1.30	1.35
$dRMS_{near,CL}$[b,d], Å	1.40	1.74
Average residence time[b,c], ns	6.38±0.21	6.39±0.22
Maximal residence time[b,c], ns	115	160

[a] Initial structure from which the simulation started.
[b] OP and CL states during the simulations assigned according to the native state to which the dRMS of the frame was smaller.
[c] Not including short transitions (<0.4 ns).
[d] dRMS calculated at the minimum of free energy near each basin, with respect to the native structure of each basin.

plementary Fig. S4A). Observing the two-dimensional free energy surfaces (Supplementary Figs. S4B and C) it is apparent that in RBP the OP state highly fluctuates away from its native configuration (compared to GBP in Fig. 7B). This behavior occurs in the single-basin simulations as well (data not shown).

The separation between the basins is related to the magnitude of the transition (characterized by the RMSD between the two native states). Distant basins will have a higher transition state. In order to facilitate transitions, the barrier is lowered by increasing T_{mix}. This is the case for GBP, in which the OP and CL basins are farther apart (RMSD of 5.3 Å compared to 4.1 Å in RBP), therefore its T_{mix} is higher (see Table 1). Lowering the barrier by increasing T_{mix} increases the population of intermediate states. Thus, the upper limit for T_{mix} will be as high as the native states can still be sampled accurately. This means that the transition rate can only be controlled to some extent, and is largely dictated by the magnitude of the transition. Systems with larger transition magnitudes will naturally display slower transition rates and longer simulation times are required for parameter searching as well as for the actual sampling.

Finally, comparing between the KBGo and the DoME models, we observe that C is more negative for the DoME model (the CL state is naturally more stable so the OP state needs to be stabilized with respect to it). In the DoME model, fluctuations near each basin are smaller, therefore the basins are narrower (Supplementary Fig. S4A). This mostly affects the OP basin which has less contacts, which effectively makes the OP basin narrower than in the KBGo model, reducing its population. Agreeing with the relation between basin population and relative height discussed previously, reduction of population should be compensated by lowering the basin energy.

Practical issues in the current scheme

For the systems used in the current study, which had transition RMSDs of 4-5 Å, short iterations of 500 ns MD simulations and an additional final round of 2 μs are sufficient for accurately determining the parameters. We demonstrated that using the detected parameters, the two basins are sampled accurately, and transition rate is fast enough for free energies to converge. For larger systems, barriers are naturally higher and longer simulation times will be required in order not to compensate for accuracy of basin sampling.

Other issue to be considered in the scheme is the target population ratio between the states. In the current study, the ratio was set to 1, but can, in principle be set to any desired value. This aspect is important because this way, the model can be calibrated against experimental values to construct a system which exhibits a realistic behavior. The rate is not explicitly controlled in the current application but can be controlled to some extent. The structural based CG MD simulations heavily rely on the system-dependent parameters. The target population ratio might be one of such parameters in the simulations for other proteins or protein complexes.

Conclusions

In this study, we have developed a scheme to determine parameters for exponential mixing dual-basin Gō model potential using the MBAR analysis method. Using MBAR, we can perform fast iterations for predicting the behavior of a hypothetical simulation, reducing the total simulation time. We estimate that parameter searching can be accelerated by at least three-fold using the procedure to the two protein systems. The systematic scheme for determining mixing parameters in dual-basin Gō model potential makes the CG MD simulation more powerful in future applications. The systematic scheme developed here is generally applicable to the parameter determinations that are necessary in other types of simulations.

Acknowledgment

The computer resources of K computer were provided by the RIKEN Center for Computational Science through HPCI System Research project (Project ID: ra000009). We thank the RIKEN Integrated Cluster of Clusters (RICC) and HOKUSAI GreatWave for additional computer resources. This research was supported in part by RIKEN pioneering researches of “Glycolipidologue” and “Dynamic Structural Biology” (to YS), a Grant-in-Aid for Scientific Research (C) (JSPS KAKENHI Grant no. 16K05526) (to CK).

Informed Consent

The authors declare no conflict of interest.

Author Contribution

Y. S. directed the entire study. A. S., C. K., Y. M., and Y. S. co-wrote the manuscript. A. S., C. K., and Y. M. wrote the simulation program and implemented potentials, mixing schemes and MBAR analysis code. A. S. performed the simulations and the analysis.

References

[1] Thirumalai, D. & Lorimer, G. H. Chaperonin-mediated protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 245–269 (2001).
[2] Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109**, 275–282 (2002).
[3] Margittai, M., Widengren, J., Schweinberger, E., Schröder, G., Felekyan, S., Haustein, E., *et al.* Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. *Proc. Natl. Acad. Sci. USA* **100**, 15516–15521 (2003).
[4] Munro, J. B., Sanbonmatsu, K. Y., Spahn, C. M. & Blanchard, S. C. Navigating the ribosome’s metastable energy landscape. *Trends Biochem. Sci.* **34**, 390–400 (2009).

[5] Ma, B. & Nussinov, R. Enzyme dynamics point to stepwise conformational selection in catalysis. *Cur. Opin. Chem. Biol.* **14**, 652–659 (2010).
[6] Sun, Y.-J., Rose, J., Wang, B.-C. & Hsiao, C.-D. The structure of glutamine-binding protein complexed with glutamine at 1.94 Å resolution: comparisons with other amino acid binding proteins. *J. Mol. Biol.* **278**, 219–229 (1998).
[7] Mowbray, S. L. & Björkman, A. J. Conformational changes of ribose-binding protein and two related repressors are tailored to fit the functional need. *J. Mol. Biol.* **294**, 487–499 (1999).
[8] Kern, D. & Zuiderweg, E. R. The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* **13**, 748–757 (2003).
[9] Wong, L., Lieser, S. A., Miyashita, O., Miller, M., Tasken, K., Onuchic, J. N., *et al.* Coupled motions in the SH2 and kinase domains of Csk control Src phosphorylation. *J. Mol. Biol.* **351**, 131–143 (2005).
[10] Hanson, J. A., Duderstadt, K., Watkins, L. P., Bhattacharyya, S., Brokaw, J., Chu, J.-W., *et al.* Illuminating the mechanistic roles of enzyme conformational dynamics. *Proc. Natl. Acad. Sci. USA* **104**, 18055–18060 (2007).
[11] Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789–796 (2009).
[12] MacKerell Jr, A. D., Bashford, D., Bellott, M., Dunbrack Jr, R. L., Evanseck, J. D., Field, M. J., *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
[13] Ponder, J. W. & Case, D. A. Force fields for protein simulations. in *Adv. Protein Chem.* vol. 66, pp. 27–85 (Elsevier, 2003).
[14] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
[15] Adcock, S. A. & McCammon, J. A. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* **106**, 1589–1615 (2006).
[16] Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O. & Shaw, D. E. Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* **19**, 120–127 (2009).
[17] Hospital, A., Goñi, J. R., Orozco, M. & Gelpí, J. L. Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinform. Chem.* **8**, 37–47 (2015).
[18] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015).
[19] Shaw, D. E., Deneroff, M. M., Dror, R. O., Kuskin, J. S., Larson, R. H., Salmon, J. K., *et al.* Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM* **51**, 91–97 (2008).
[20] Yu, I., Mori, T., Ando, T., Harada, R., Jung, J., Sugita, Y., *et al.* Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *Elife* **5**, e19274 (2016).
[21] Durrant, J. D., Bush, R. M. & Amaro, R. E. Microsecond molecular dynamics simulations of influenza neuraminidase suggest a mechanism for the increased virulence of stalk-deletion mutants. *J. Phys. Chem. B* **120**, 8590–8599 (2016).
[22] Trovato, F. & Fumagalli, G. Molecular simulations of cellular processes. *Biophys. Rev.* **9**, 941–958 (2017).
[23] Acun, B., Hardy, D., Kale, L. V., Li, K., Phillips, J. & Stone, J. Scalable molecular dynamics with NAMD on the Summit system. *IBM J. Res. Dev.* **62**, 4:1–4:9 (2018).
[24] Hyeon, C. & Thirumalai, D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat. Commun.* **2**, 487 (2011).
[25] Takada, S. Coarse-grained molecular simulations of large biomolecules. *Curr. Opin. Struct. Biol.* **22**, 130–137 (2012).
[26] Noid, W. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **139**, 090901 (2013).
[27] Ingólfsson, H. I., Lopez, C. A., Uusitalo, J. J., de Jong, D. H., Gopal, S. M., Periole, X., *et al.* The power of coarse graining in biomolecular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4**, 225–248 (2014).
[28] Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E. & Kolinski, A. Coarse-grained protein models and their applications. *Chem. Rev.* **116**, 7898–7936 (2016).
[29] Pak, A. J. & Voth, G. A. Advances in coarse-grained modeling of macromolecular complexes. *Curr. Opin. Struct. Biol.* **52**, 119–126 (2018).
[30] Sanbonmatsu, K. Y., Joseph, S. & Tung, C.-S. Simulating movement of tRNA into the ribosome during decoding. *Proc. Natl. Acad. Sci. USA* **102**, 15854–15859 (2005).
[31] Koga, N. & Takada, S. Folding-based molecular simulations reveal mechanisms of the rotary motor F1–ATPase. *Proc. Natl. Acad. Sci. USA* **103**, 5367–5372 (2006).
[32] Pellarin, R., Guarnera, E. & Caflisch, A. Pathways and intermediates of amyloid fibril formation. *J. Mol. Biol.* **374**, 917–924 (2007).
[33] Yao, X.-Q., Kenzaki, H., Murakami, S. & Takada, S. Drug export and allosteric coupling in a multidrug transporter revealed by molecular simulations. *Nat. Commun.* **1**, 117 (2010).
[34] Clementi, C., Nymeyer, H. & Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **298**, 937–953 (2000).
[35] Clementi, C., Jennings, P. A. & Onuchic, J. N. How native-state topology affects the folding of dihydrofolate reductase and interleukin-1β. *Proc. Natl. Acad. Sci. USA* **97**, 5871–5876 (2000).
[36] Koga, N. & Takada, S. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Gō-like model. *J. Mol. Biol.* **313**, 171–180 (2001).
[37] Best, R. B., Chen, Y.-G. & Hummer, G. Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor. *Structure* **13**, 1755–1763 (2005).
[38] Okazaki, K., Koga, N., Takada, S., Onuchic, J. N. & Wolynes, P. G. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* **103**, 11844–11849 (2006).
[39] Whitford, P. C., Miyashita, O., Levy, Y. & Onuchic, J. N. Conformational transitions of adenylate kinase: switching by cracking. *J. Mol. Biol.* **366**, 1661–1671 (2007).
[40] Takagi, F. & Kikuchi, M. Structural change and nucleotide dissociation of myosin motor domain: dual Gō model simulation. *Biophys. J.* **93**, 3820–3827 (2007).
[41] Clementi, C. Coarse-grained models of protein folding: toy models or predictive tools? *Curr. Opin. Struct. Biol.* **18**, 10–15 (2008).
[42] Hills, R. & Brooks, C. Insights from coarse-grained Gō models for protein folding and dynamics. *Int. J. Mol. Sci.* **10**, 889–905 (2009).
[43] Daily, M. D., Phillips Jr, G. N. & Cui, Q. Many local motions cooperate to produce the adenylate kinase conformational transition. *J. Mol. Biol.* **400**, 618–631 (2010).
[44] Takada, S., Kanada, R., Tan, C., Terakawa, T., Li, W. & Kenzaki, H. Modeling structural dynamics of biomolecular

complexes by coarse-grained molecular simulations. *Acc. Chem. Res.* **48**, 3026–3035 (2015).
[45] Taketomi, H., Ueda, Y. & Gō, N. 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**, 445–459 (1975).
[46] Ueda, Y., Taketomi, H. & Gō, N. Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A. Three-dimensional lattice model of lysozyme. *Biopolymers* **17**, 1531–1548 (1978).
[47] Go, N. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* **12**, 183–210 (1983).
[48] Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195 (1995).
[49] Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z. & Wolynes, P. G. Protein folding funnels: the nature of the transition state ensemble. *Fold. Des.* **1**, 441–450 (1996).
[50] Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600 (1997).
[51] Karanicolas, J. & Brooks III, C. L. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* **11**, 2351–2361 (2002).
[52] Karanicolas, J. & Brooks III, C. L. Improved Gō-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *J. Mol. Biol.* **334**, 309–325 (2003).
[53] Miyazawa, S. & Jernigan, R. L. 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 (1996).
[54] Kobayashi, C., Matsunaga, Y., Koike, R., Ota, M. & Sugita, Y. Domain Motion Enhanced (DoME) Model for Efficient Conformational Sampling of Multidomain Proteins. *J. Phys. Chem. B* **119**, 14584–14593 (2015).
[55] Whitford, P. C., Gosavi, S. & Onuchic, J. N. Conformational transitions in adenylate kinase allosteric communication reduces misligation. *J. Biol. Chem.* **283**, 2042–2048 (2008).
[56] Hyeon, C., Jennings, P. A., Adams, J. A. & Onuchic, J. N. Ligand-induced global transitions in the catalytic domain of protein kinase A. *Proc. Natl. Acad. Sci. USA* **106**, 3023–3028 (2009).
[57] Lu, Q. & Wang, J. Kinetics and statistical distributions of single-molecule conformational dynamics. *J. Phys. Chem. B* **113**, 1517–1521 (2009).
[58] Maragakis, P. & Karplus, M. Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. *J. Mol. Biol.* **352**, 807–822 (2005).
[59] Okazaki, K. & Takada, S. Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc. Natl. Acad. Sci. USA* **105**, 11182–11187 (2008).
[60] Hummer, G., Pratt, L. R. & García, A. E. Multistate Gaussian model for electrostatic solvation free energies. *J. Am. Chem. Soc.* **119**, 8523–8527 (1997).
[61] Shirts, M. R. & Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129**, 124105 (2008).
[62] Björkman, A., Binnie, R. A., Zhang, H., Cole, L. B., Hermodson, M. A. & Mowbray, S. L. Probing protein-protein interactions. The ribose-binding protein in bacterial transport and chemotaxis. *J. Biol. Chem.* **269**, 30206–30211 (1994).
[63] Björkman, A. J. & Mowbray, S. L. Multiple open forms of ribose-binding protein trace the path of its conformational change. *J. Mol. Biol.* **279**, 651–664 (1998).
[64] Koike, R., Ota, M. & Kidera, A. Hierarchical description and extensive classification of protein structural changes by Motion Tree. *J. Mol. Biol.* **426**, 752–762 (2014).
[65] Moritsugu, K., Koike, R., Yamada, K., Kato, H. & Kidera, A. Motion tree delineates hierarchical structure of protein dynamics observed in molecular dynamics simulation. *PLoS One* **10**, e0131583 (2015).
[66] Domański, J., Hedger, G., Best, R. B., Stansfeld, P. J. & Sansom, M. S. Convergence and sampling in determining free energy landscapes for membrane protein association. *J. Phys. Chem. B* **121**, 3364–3375 (2017).
[67] Matsunaga, Y. & Sugita, Y. Refining Markov state models for conformational dynamics using ensemble-averaged data and time-series trajectories. *J. Chem. Phys.* **148**, 241731 (2018).
[68] Jung, J., Mori, T., Kobayashi, C., Matsunaga, Y., Yoda, T., Feig, M., *et al.* 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 (2015).
[69] Kobayashi, C., Jung, J., Matsunaga, Y., Mori, T., Ando, T., Tamura, K., *et al.* GENESIS 1.1: A hybrid-parallel molecular dynamics simulator with enhanced sampling algorithms on multiple computational platforms. *J. Comput. Chem.* **38**, 2193–2206 (2017).
[70] Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
[71] Loeffler, H. H. & Kitao, A. Collective dynamics of periplasmic glutamine binding protein upon domain closure. *Biophys. J.* **97**, 2541–2549 (2009).
[72] Ravindranathan, K. P., Gallicchio, E. & Levy, R. M. Conformational equilibria and free energy profiles for the allosteric transition of the ribose-binding protein. *J. Mol. Biol.* **353**, 196–210 (2005).