Special Issue
"Progress of Theoretical and Computational Biophysics"

Vol. 16, pp. 344–366 (2019)
doi: 10.2142/biophysico.16.0_344

Biophysics and Physicobiology

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

Review Article

Protein structure predictions by enhanced conformational sampling methods

Yuko Okamoto[1,2,3,4,5]

[1]*Department of Physics, Graduate School of Science, Nagoya University, Nagoya, Aichi 464-8602, Japan*
[2]*Structural Biology Research Center, Graduate School of Science, Nagoya University, Nagoya, Aichi 464-8602, Japan*
[3]*Center for Computational Science, Graduate School of Engineering, Nagoya University, Nagoya, Aichi 464-8603, Japan*
[4]*Information Technology Center, Nagoya University, Nagoya, Aichi 464-8601, Japan*
[5]*JST-CREST, Nagoya, Aichi 464-8602, Japan*

Received July 5, 2019; accepted August 7, 2019

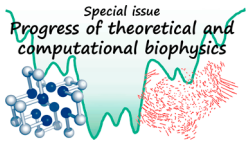
In this Special Festschrift Issue for the celebration of Professor Nobuhiro Gō's 80th birthday, we review enhanced conformational sampling methods for protein structure predictions. We present several generalized-ensemble algorithms such as multicanonical algorithm, replica-exchange method, etc. and parallel Monte Carlo or molecular dynamics method with genetic crossover. Examples of the results of these methods applied to the predictions of protein tertiary structures are also presented.

Key words: molecular simulation, protein structure prediction, multicanonical algorithm, replica-exchange method, genetic algorithm

Professor Nobuhiro Gō is probably most well-known for his *consistency principle* in protein folding [1], which states that various types of interactions that contribute to the stabilization of the native conformation of a protein are consistent with each other. This principle was deduced from the results of Monte Carlo simulations of a lattice protein model. This principle was reinterpreted as the *principle of minimal frustration* [2,3]. They showed that the phase space of the protein system is characterized by two parameters, folding temperature T_F and glass transition temperature T_G and that a protein can fold into a unique native structure if $T_G/T_F \ll 1$. Another two-parameter argument is that a protein folds into the native structure if $(T_\theta - T_F)/T_\theta \ll 1$, where T_θ is the coil-globule transition temperature [4,5]. These parameters characterize the free energy landscape of protein systems [1–6]. While these arguments were mainly given from simulations of lattice models and coarse-grained models, we confirmed these properties by calculating T_G, T_F, and T_θ of a small peptide and its free energy landscape by simulations of an all-atom model [7,8].

Besides these fundamental works on protein folding, Professor Gō has also shown that protein structures can be determined in atomistic details by utilizing the results of NMR experiments [9], which drastically improved the distance geometry methods (see, e.g., [10]). Before that time, protein structures were determined mainly by X-ray diffrac-

Corresponding author: Yuko Okamoto, Department of Physics, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan.
e-mail: okamoto@tb.phys.nagoya-u.ac.jp



◀ **Significance** ▶

This article reviews various enhanced conformational sampling methods for protein structure predictions. Several generalized-ensemble algorithms and a method based on the genetic algorithm are presented. These methods can sample much wider conformational space of biomolecules than conventional methods. Mathematical details are given so that those who are not familiar with these methods can understand them. Some examples of simulation results of protein structure predictions by these methods are also presented in order to elucidate the effectiveness of the methods.

tion experiments, and structures determined by NMR experiments were emerging for the first time. Professor Gō's method was one of very effective ones that were used to determine the three-dimensional structures of proteins from NMR experiments. After Professor Gō's works, many groups worked on this problem. One example is the uses of simulated annealing (SA) [11] to determine protein structures from the results of X-ray and NMR experiments [12–15]. SA was also applied to the protein folding simulations (see, e.g., Refs. [16–20] for earlier applications).

In this article, we discuss our methods for prediction of protein structures by Monte Carlo (MC) and molecular dynamics (MD) simulations. Conventional simulations of biomolecules suffer from the multiple-minima problem: The canonical fixed-temperature simulations at low temperatures tend to get trapped in a huge number of local-minimum-energy states, which will give wrong results. We have been advocating the uses of the *generalized-ensemble algorithms*, which overcomes the multipole-minima problem (for reviews see, e.g., Refs. [21–25]). In the generalized-ensemble algorithm, each state is weighted by an artificial, non-Boltzmann probability weight factor so that a random walk in potential energy space and/or other physical quantities (such as volume, etc.) may be realized. The random walk allows the simulation to escape from any energy-local-minimum state and to sample much wider conformational space than by conventional methods. From a single simulation run, one can obtain accurate ensemble averages as functions of temperature and/or other parameters (such as pressure, etc.) of the system by the single-histogram [26] and multiple-histogram [27,28] reweighting techniques (an extension of the multiple-histogram method is also referred to as the *weighted histogram analysis method* (WHAM) [28]).

Sucessful predictions of the three-dimensional structures of proteins are possible when both enhanced sampling techniques and accurate potential energy (or, force fields) for the protein systems are employed. If one has sufficient computational power, the former may not be necessary (see, e.g., Ref. [29] for examples of folding proteins into their native structures by conventional canonical-ensemble simulations). However, generalized-ensemble algorithms will save computation time for this purpose. Generalized-ensemble algorithms can give accurate thermodynamic averages and can also be used to judge which existing force fields are good. Several existing force fields were compared by generalized-ensemble simulations [30,31]. Those works showed that AMBER99 forms 3_{10}-helices and CHARMM22 forms π-helix in small peptides, which are not observed in experiments [30,31]. The author visited the Laboratory of Professor Charles L. Brooks, III in August, 2001, and gave a seminar talk, which included unpublished results of Ref. [30,31] (especially, Fig. 4(e) of Ref. [31], which clearly showed the formation of π-helix with CHARMM22). Although they have also independently observed tendency of π-helix formations with CHARMM22, our convincing results above encouraged them to develop a modified force field (namely, CHARMM22 with CMAP [32]) [33]. We have also proposed several methods for improving the force fields [34–37] (for a review, see Ref. [38]).

This article is organized as follows. We first describe the generalized-ensemble algorithms. We then present the results by some of the generalized-ensemble simulations. Finally, we draw conclusions.

Simulation Methods

Multicanonical algorithm

Let us consider a system of N atoms of mass m_k (k=1, ..., N) with their coordinate vectors and momentum vectors denoted by $q \equiv \{\boldsymbol{q}_1, ..., \boldsymbol{q}_N\}$ and $p \equiv \{\boldsymbol{p}_1, ..., \boldsymbol{p}_N\}$, respectively. The Hamiltonian $H(q, p)$ of the system is the sum of the kinetic energy $K(p)$ and the potential energy $E(q)$:

$$H(q, p) = K(p) + E(q), \tag{1}$$

where

$$K(p) = \sum_{k=1}^{N} \frac{\boldsymbol{p}_k^2}{2m_k}. \tag{2}$$

In the canonical ensemble at temperature T each state $x \equiv (q, p)$ with the Hamiltonian $H(q, p)$ is weighted by the Boltzmann factor:

$$W_{\mathrm{B}}(x; T) = \exp(-\beta H(q, p)), \tag{3}$$

where the inverse temperature β is defined by $\beta = 1/k_{\mathrm{B}}T$ (k_{B} is the Boltzmann constant). The average kinetic energy at temperature T is then given by

$$\langle K(p) \rangle_T = \left\langle \sum_{k=1}^{N} \frac{\boldsymbol{p}_k^2}{2m_k} \right\rangle_T = \frac{3}{2} N k_{\mathrm{B}} T. \tag{4}$$

Because the coordinates q and momenta p are decoupled in Eq. (1), we can suppress the kinetic energy part and can write the Boltzmann factor as

$$W_{\mathrm{B}}(x; T) = W_{\mathrm{B}}(E; T) = \exp(-\beta E). \tag{5}$$

The canonical probability distribution of potential energy $P_{\mathrm{B}}(E; T)$ is then given by the product of the density of states $n(E)$ and the Boltzmann weight factor $W_{\mathrm{B}}(E; T)$:

$$P_{\mathrm{B}}(E; T) \propto n(E) W_{\mathrm{B}}(E; T). \tag{6}$$

Because $n(E)$ is a rapidly increasing function and the Boltzmann factor decreases exponentially, the canonical ensemble yields a bell-shaped distribution which has a maximum around the average energy at temperature T. The conventional MC or MD simulations at constant temperature are expected to yield $P_{\mathrm{B}}(E; T)$. A MC simulation based on the Metropolis method [39] is performed with the follow-

ing transition probability from a state x of potential energy E to a state x' of potential energy E':

$$w(x \to x') = \min\left(1, \frac{W_B(E'; T)}{W_B(E; T)}\right) = \min(1, \exp(-\beta \Delta E)). \quad (7)$$

where

$$\Delta E = E' - E. \quad (8)$$

A MD simulation, on the other hand, is based on the following Newton equations of motion:

$$\dot{\boldsymbol{q}}_k = \frac{\boldsymbol{p}_k}{m_k}, \quad (9)$$

$$\dot{\boldsymbol{p}}_k = -\frac{\partial E}{\partial \boldsymbol{q}_k} = \boldsymbol{f}_k, \quad (10)$$

where $\boldsymbol{f}_k$ is the force acting on the k-th atom (k=1, ..., N). This set of equations actually yield the microcanonical ensemble, and we have to add a thermostat in order to obtain the canonical ensemble at temperature T. Here, we just follow Nosé's prescription [40,41], and we have

$$\dot{\boldsymbol{q}}_k = \frac{\boldsymbol{p}_k}{m_k}, \quad (11)$$

$$\dot{\boldsymbol{p}}_k = -\frac{\partial E}{\partial \boldsymbol{q}_k} - \frac{\dot{s}}{s}\boldsymbol{p}_k = \boldsymbol{f}_k - \frac{\dot{s}}{s}\boldsymbol{p}_k, \quad (12)$$

$$\dot{s} = s\frac{P_s}{Q}, \quad (13)$$

$$\dot{P}_s = \sum_{k=1}^{N} \frac{\boldsymbol{p}_k^2}{m_k} - 3Nk_B T = 3Nk_B(T(t) - T), \quad (14)$$

where s is Nosé's scaling parameter, Q is its mass, P_s is its conjugate momentum, and the "instantaneous temperature" $T(t)$ is defined by

$$T(t) = \frac{1}{3Nk_B}\sum_{k=1}^{N} \frac{\boldsymbol{p}_k(t)^2}{m_k}. \quad (15)$$

However, in practice, it is very difficult to obtain accurate canonical distributions of complex systems at low temperatures by conventional MC or MD simulation methods. This is because simulations at low temperatures tend to get trapped in one or a few of local-minimum-energy states. The generalized-ensemble algorithms overcome this difficulty by performaing random walks in physical quantities such as potential energy and volume, etc. (or their conjugate parameters such as temperature and pressure, etc., respectively).

One of the most well-known generalized-ensemble algorithms is perhaps the *multicanonical algorithm* (MUCA) [42,43] (for reviews see, e.g., Refs. [44,45]). The method is also referred to as *entropic sampling* [46,47] and *adaptive umbrella sampling* [48] *of the potential energy* [49]. MUCA can also be considered as a sophisticated, ideal realization of a class of algorithms called *umbrella sampling* [50]. Also closely related methods are *Wang-Landau method* [51,52], which is also referred to as *density of states Monte Carlo* [53], and *metadynamics* [54], which can be considered as an extension of Wang-Landau method where a random walk in reaction coordinate, or collective variable, instead of potential energy is performed [55]. While MUCA and its generalizations have been first applied to spin systems, MUCA was also introduced to the molecular simulation field [56]. Since then MUCA and its generalizations have been extensively used in many applications in protein and other biomolecular systems [56–86]. Molecular dynamics version of MUCA has also been developed [49,62,63]. MUCA has been extended so that flat distributions in other variables instead of potential energy may be obtained (see, e.g., Refs. [57,61,64,67,80,84]).

In the multicanonical ensemble [42,43], each state is weighted by a non-Boltzmann weight factor $W_{mu}(E)$ (which we refer to as the *multicanonical weight factor*) so that a uniform potential energy distribution $P_{mu}(E)$ is obtained:

$$P_{mu}(E) \propto n(E) W_{mu}(E) \equiv \text{const.} \quad (16)$$

The flat distribution implies that a free random walk in the potential energy space is realized in this ensemble. This allows the simulation to escape from any local minimum-energy states and to sample the configurational space much more widely than the conventional canonical MC or MD methods.

The definition in Eq. (16) implies that the multicanonical weight factor is inversely proportional to the density of states, and we can write it as follows:

$$W_{mu}(E) \equiv \exp[-\beta_0 E_{mu}(E; T_0)] = \frac{1}{n(E)}, \quad (17)$$

where we have chosen an arbitrary reference temperature, $T_0 = 1/k_B\beta_0$, and the "*multicanonical potential energy*" is defined by

$$E_{mu}(E; T_0) \equiv k_B T_0 \ln n(E) = T_0 S(E). \quad (18)$$

Here, $S(E)$ is the entropy in the microcanonical ensemble. Since the density of states of the system is usually unknown, the multicanonical weight factor has to be determined numerically by iterations of short preliminary runs [42,43].

A multicanonical MC simulation is performed, for instance, with the usual Metropolis criterion [39]: The transition probability of state x with potential energy E to state x' with potential energy E' is given by

$$w(x \to x') = \min\left(1, \frac{W_{mu}(E')}{W_{mu}(E)}\right) = \min\left(1, \frac{n(E)}{n(E')}\right) = \min(1, \exp(-\beta_0 \Delta E_{mu})), \quad (19)$$

where

$$\Delta E_{\rm mu} = E_{\rm mu}(E';\, T_0) - E_{\rm mu}(E;\, T_0). \tag{20}$$

The MD algorithm in the multicanonical ensemble also naturally follows from Eq. (17), in which the regular constant temperature MD simulation (with $T=T_0$) is performed by replacing E by $E_{\rm mu}$ in Eq. (12) [62,63]:

$$\dot{\boldsymbol{p}}_k = -\frac{\partial E_{\rm mu}(E;\, T_0)}{\partial \boldsymbol{q}_k} - \frac{\dot{s}}{s}\boldsymbol{p}_k = \frac{\partial E_{\rm mu}(E;\, T_0)}{\partial E}\boldsymbol{f}_k - \frac{\dot{s}}{s}\boldsymbol{p}_k . \tag{21}$$

If the exact multicanonical weight factor $W_{\rm mu}(E)$ is known, one can calculate the ensemble averages of any physical quantity A at any temperature T ($=1/k_{\rm B}\beta$) as follows:

$$\langle A\rangle_T = \frac{\sum_E A(E)P_{\rm B}(E;\, T)}{\sum_E P_{\rm B}(E;\, T)} = \frac{\sum_E A(E)n(E)\exp(-\beta E)}{\sum_E n(E)\exp(-\beta E)}, \tag{22}$$

where the density of states is given by (see Eq. (17))

$$n(E) = \frac{1}{W_{\rm mu}(E)}. \tag{23}$$

The summation instead of integration is used in Eq. (22), because we often discretize the potential energy E with step size ϵ ($E=E_i$; $i=1, 2, ...$). Here, the explicit form of the physical quantity A should be known as a function of potential energy E. For instance, $A(E)=E$ gives the average potential energy $\langle E\rangle_T$ as a function of temperature, and $A(E)=\beta^2(E-\langle E\rangle_T)^2$ gives specific heat.

In general, the multicanonical weight factor $W_{\rm mu}(E)$, or the density of states $n(E)$, is not *a priori* known, and one needs its estimator for a numerical simulation. This estimator is usually obtained from iterations of short trial multicanonical simulations. (The details of this process are described, for instance, in Refs. [44,45]). However, the iterative process can be non-trivial and very tedius for complex systems.

Once a satsifactory multicanonical weight factor is obtained, we perform with this weight factor a multicanonical simulation with high statistics (production run) in order to get even better estimate of the density of states. Let $N_{\rm mu}(E)$ be the histogram of potential energy distribution $P_{\rm mu}(E)$ obtained by this production run. The best estimate of the density of states can then be given by the single-histogram reweighting techniques [26] as follows (see the proportionality relation in Eq. (16)):

$$n(E) = \frac{N_{\rm mu}(E)}{W_{\rm mu}(E)}. \tag{24}$$

By substituting this quantity into Eq. (22), one can calculate ensemble averages of physical quantity $A(E)$ as a function of temperature. Moreover, ensemble averages of any physical quantity A (including those that cannot be expressed as functions of potential energy) at any temperature T ($=1/k_{\rm B}\beta$) can now be obtained as long as one stores the "trajectory" of configurations (and A) from the production run. Namely, we have [74]

$$\langle A\rangle_T = \frac{\sum_{k=1}^{n_0} A(x(k))W_{\rm mu}^{-1}(E(x(k)))\exp[-\beta E(x(k))]}{\sum_{k=1}^{n_0} W_{\rm mu}^{-1}(E(x(k)))\exp[-\beta E(x(k))]}, \tag{25}$$

where $x(k)$ is the configuration at the k-th MC (or MD) step and n_0 is the total number of configurations stored. Note that when A is a function of E, Eq. (25) reduces to Eq. (22) where the density of states is given by Eq. (24).

Replica-exchange method

The *replica-exchange method* (REM) [87,88] is one of the most widely used method in biomolecular simulations. (REM is also referred to as *parallel tempering* [89] and *multiple Markov chain method* [90]) In this method, a number of non-interacting copies (or, replicas) of the original system at different temperatures are simulated independently and simultaneously by the conventional MC or MD method. Every few steps, pairs of replicas are exchanged with a specified transition probability. The weight factor is just the product of Boltzmann factors, and so it is essentially known.

REM has already been used in many applications in protein systems [91–104]. Other molecular simulation fields have also been studied by this method in various ensembles [105–108]. Moreover, REM and other generalized-ensemble algorithms were introduced to the quantum chemistry field [109–113]. The details of molecular dynamics algorithm for REM, which is referred to as the *Replica-Exchange Molecular Dynamics* (REMD) have been worked out in Ref. [92], and this led to a wide application of REM in the protein folding and related problems (see, e.g., Refs. [114–128]).

The system for REM consists of M *non-interacting* copies (or, replicas) of the original system in the canonical ensemble at M different temperatures T_m ($m=1, ..., M$). We arrange the replicas so that there is always exactly one replica at each temperature. Then there exists a one-to-one correspondence between replicas and temperatures; the label i ($i=1, ..., M$) for replicas is a permutation of the label m ($m=1, ..., M$) for temperatures, and vice versa:

$$\begin{cases} i = i(m) \equiv f(m), \\ m = m(i) \equiv f^{-1}(i), \end{cases} \tag{26}$$

where $f(m)$ is a permutation function of m and $f^{-1}(i)$ is its inverse.

Let $X=\{x_1^{[i(1)]}, ..., x_M^{[i(M)]}\} = \{x_{m(1)}^{[1]}, ..., x_{m(M)}^{[M]}\}$ stand for a "state" in this generalized ensemble. Each "substate" $x_m^{[i]}$ is specified by the coordinates $q^{[i]}$ and momenta $p^{[i]}$ of N atoms in replica i at temperature T_m:

$$x_m^{[i]} \equiv (q^{[i]}, p^{[i]})_m. \tag{27}$$

Because the replicas are non-interacting, the weight factor for the state X in this generalized ensemble is given by the product of Boltzmann factors for each replica (or at each temperature):

$$\begin{aligned} W_{\mathrm{REM}}(X) &= \prod_{i=1}^{M} \exp\{-\beta_{m(i)} H(q^{[i]}, p^{[i]})\} \\ &= \prod_{m=1}^{M} \exp\{-\beta_m H(q^{[i(m)]}, p^{[i(m)]})\} \\ &= \exp\Big\{-\sum_{i=1}^{M} \beta_{m(i)} H(q^{[i]}, p^{[i]})\Big\} \\ &= \exp\Big\{-\sum_{m=1}^{M} \beta_m H(q^{[i(m)]}, p^{[i(m)]})\Big\}, \end{aligned} \tag{28}$$

where $i(m)$ and $m(i)$ are the permutation functions in Eq. (26).

We now consider exchanging a pair of replicas in the generalized ensemble. Suppose we exchange replicas i and j which are at temperatures T_m and T_n, respectively:

$$X = \{..., x_m^{[i]}, ..., x_n^{[j]}, ...\} \rightarrow X' = \{..., x_m^{[j]\prime}, ..., x_n^{[i]\prime}, ...\}. \tag{29}$$

Here, i, j, m, and n are related by the permutation functions in Eq. (26), and the exchange of replicas introduces a new permutation function f':

$$\begin{cases} i = f(m) \rightarrow j = f'(m), \\ j = f(n) \rightarrow i = f'(n). \end{cases} \tag{30}$$

The exchange of replicas can be written in more detail as

$$\begin{cases} x_m^{[i]} \equiv (q^{[i]}, p^{[i]})_m \rightarrow x_m^{[j]\prime} \equiv (q^{[j]}, p^{[j]\prime})_m, \\ x_n^{[j]} \equiv (q^{[j]}, p^{[j]})_n \rightarrow x_n^{[i]\prime} \equiv (q^{[i]}, p^{[i]\prime})_n, \end{cases} \tag{31}$$

where the definitions for $p^{[i]\prime}$ and $p^{[j]\prime}$ will be given below. In the original implementation of the *replica-exchange method* (REM) [87,88], Monte Carlo algorithm was used, and only the coordinates q (and the potential energy function $E(q)$) had to be taken into account. In molecular dynamics algorithm, on the other hand, we also have to deal with the momenta p. We proposed the following momentum assignment in Eq. (31) [92]:

$$\begin{cases} p^{[i]\prime} \equiv \sqrt{\dfrac{T_n}{T_m}}\, p^{[i]}, \\ p^{[j]\prime} \equiv \sqrt{\dfrac{T_m}{T_n}}\, p^{[j]}, \end{cases} \tag{32}$$

which we believe is the simplest and the most natural. This assignment means that we just rescale uniformly the velocities of all the atoms in the replicas by the square root of the ratio of the two temperatures so that the temperature condition in Eq. (4) may be satisfied.

Replica exchange is accepted according to the usual Metropolis criterion [39]:

$$w(X \rightarrow X') \equiv w(x_m^{[i]} | x_n^{[j]}) = \min(1, \exp(-\Delta)), \tag{33}$$

where in the second expression (i.e., $w(x_m^{[i]}|x_n^{[j]})$ we explicitly wrote the pair of replicas (and temperatures) to be exchanged, and

$$\Delta = \beta_m (E(q^{[j]}) - E(q^{[i]})) - \beta_n (E(q^{[j]}) - E(q^{[i]})) \tag{34}$$

$$= (\beta_m - \beta_n)(E(q^{[j]}) - E(q^{[i]})). \tag{35}$$

Without loss of generality we can assume $T_1 < T_2 < ... < T_M$. A REM simulation is then realized by alternately performing the following two steps:

1. Each replica in canonical ensemble of the fixed temperature is simulated *simultaneously* and *independently* for a certain MC or MD steps.
2. A pair of replicas at neighboring temperatures, say $x_m^{[i]}$ and $x_{m+1}^{[j]}$, are exchanged with the probability $w(x_m^{[i]}|x_{m+1}^{[j]})$ in Eq. (33).

Note that in Step 2 we exchange only pairs of replicas corresponding to neighboring temperatures, because the acceptance ratio of the exchange process decreases exponentially with the difference of the two β's (see Eqs. (35) and (33)). Note also that whenever a replica exchange is accepted in Step 2, the permutation functions in Eq. (26) are updated.

After a long REM production run, the canonical expectation value of a physical quantity A at temperature T_m (m=1, ..., M) can be calculated by the usual arithmetic mean as follows:

$$\langle A \rangle_{T_m} = \frac{1}{n_m} \sum_{k=1}^{n_m} A(x_m(k)), \tag{36}$$

where $x_m(k)$ (k=1, ..., n_m) are the configurations obtained at temperature T_m and n_m is the total number of measurements made at T=T_m. The expectation value at any intermediate temperature can also be obtained from Eq. (22), where the density of states is given by the multiple-histogram reweighting techniques [27,28] as follows. Let $N_m(E)$ and n_m be respectively the potential-energy histogram and the total number of samples obtained at temperature $T_m = 1/k_{\mathrm{B}}\beta_m$ (m=1, ..., M). The best estimate of the density of states is then given by [27,28]

$$n(E) = \frac{\sum_{m=1}^{M} N_m(E)}{\sum_{m=1}^{M} n_m \exp(f_m - \beta_m E)}, \tag{37}$$

where we have for each m(=1, ..., M)

$$\exp(-f_m) = \sum_E n(E) \exp(-\beta_m E). \tag{38}$$

Note that Eqs. (37) and (38) are solved self-consistently by iteration [27,28] to obtain the density of states $n(E)$ and the dimensionless Helmholtz free energy f_m. Namely, we can set all the f_m (m=1, ..., M) to, e.g., zero initially. We then use Eq. (37) to obtain $n(E)$, which is substituted into Eq. (38) to obtain next values of f_m, and so on.

Moreover, ensemble averages of any physical quantity A (including those that cannot be expressed as functions of potential energy) at any temperature T ($=1/k_B\beta$) can now be obtained from the "trajectory" of configurations of the production run. Namely, we first obtain f_m (m=1, ..., M) by solving Eqs. (37) and (38) self-consistently, and then we have [74]

$$\langle A\rangle_T = \frac{\displaystyle\sum_{m=1}^{M}\sum_{k=1}^{n_m} A(x_m(k))\frac{1}{\displaystyle\sum_{\ell=1}^{M} n_\ell \exp[f_\ell-\beta_\ell E(x_m(k))]}\exp[-\beta E(x_m(k))]}{\displaystyle\sum_{m=1}^{M}\sum_{k=1}^{n_m}\frac{1}{\displaystyle\sum_{\ell=1}^{M} n_\ell \exp[f_\ell-\beta_\ell E(x_m(k))]}\exp[-\beta E(x_m(k))]}, \tag{39}$$

where $x_m(k)$ (k=1, ..., n_m) are the configurations obtained at temperature T_m.

Replica-exchange multicanonical algorithm

MUCA and REM can be combined. In the *replica-exchange multicanonical algorithm* (REMUCA) [73–75] we first perform a short REM simulation (with M replicas) to determine the multicanonical weight factor and then perform with this weight factor a regular multicanonical simulation with high statistics. The first step is accomplished by the multiple-histogram reweighting techniques [27,28]. Let $N_m(E)$ and n_m be respectively the potential-energy histogram and the total number of samples obtained at temperature T_m ($=1/k_B\beta_m$) of the REM run. The density of states $n(E)$ is then given by solving Eqs. (37) and (38) self-consistently by iteration.

Once the estimate of the density of states is obtained, the multicanonical weight factor can be directly determined from Eq. (17) (see also Eq. (18)). Actually, the density of states $n(E)$ and the multicanonical potential energy, $E_{mu}(E;\ T_0)$, thus determined are only reliable in the following range:

$$E_1 \le E \le E_M, \tag{40}$$

where

$$\begin{cases} E_1 = \langle E\rangle_{T_1}, \\ E_M = \langle E\rangle_{T_M}, \end{cases} \tag{41}$$

and T_1 and T_M are respectively the lowest and the highest temperatures used in the REM run. Outside this range we extrapolate the multicanonical potential energy linearly: [73]

$$\mathcal{E}_{mu}^{\{0\}}(E) \equiv \begin{cases} \left.\dfrac{\partial E_{mu}(E;T_0)}{\partial E}\right|_{E=E_1}(E-E_1)+E_{mu}(E_1;T_0), & \text{for } E<E_1, \\ E_{mu}(E;T_0), & \text{for } E_1\le E\le E_M, \\ \left.\dfrac{\partial E_{mu}(E;T_0)}{\partial E}\right|_{E=E_M}(E-E_M)+E_{mu}(E_M;T_0), & \text{for } E>E_M. \end{cases} \tag{42}$$

The multicanonical MC and MD runs are then performed respectively with the Metropolis criterion of Eq. (19) and with the modified Newton equation in Eq. (21), in which $\mathcal{E}_{mu}^{\{0\}}(E)$ in Eq. (42) is substituted into $E_{mu}(E;\ T_0)$. We expect to obtain a flat potential energy distribution in the range of Eq. (40). Finally, the results are analyzed by the single-histogram reweighting techniques as described in Eq. (24) (and Eq. (22)).

Hence, our choice of $\mathcal{E}_{mu}^{\{0\}}(E)$ in Eq. (42) results in a canonical simulation at $T=T_1$ for $E<E_1$, a multicanonical simulation for $E_1\le E\le E_M$, and a canonical simulation at $T=T_M$ for $E>E_M$. Note also that the above arguments are independent of the value of T_0, and we will get the same results, regardless of its value.

For Monte Carlo method, the above statement follows directly from the following equation. Namely, our choice of the multicanonical potential energy in Eq. (42) gives from Eq. (17)

$$W_{mu}(E) = \exp[-\beta_0\mathcal{E}_{mn}^{\{0\}}(E)] = \begin{cases} \exp(-\beta_1 E), & \text{for } E<E_1, \\ \dfrac{1}{n(E)}, & \text{for } E_1\le E\le E_M, \\ \exp(-\beta_M E), & \text{for } E>E_M. \end{cases} \tag{43}$$

Multicanonical replica-exchange method

In the previous subsection we presented REMUCA, which uses a short REM run for the determination of the multicanonical weight factor. Here, we present the multicanonical replica-exchange method (MUCAREM) [73–75]. In MUCAREM the production run is a REM simulation with a few replicas not in the canonical ensemble but in the multicanonical ensemble, i.e., different replicas perform MUCA simulations with different energy ranges. While MUCA simulations are usually based on local updates, a replica-exchange process can be considered to be a global update, and global updates enhance the sampling further.

We now describe MUCAREM. Let M be the number of replicas. Here, each replica is in one-to-one correspondence not with temperature but with multicanonical weight factors of different energy range. Note that because multicanonical simulations cover much wider energy ranges than regular canonical simulations, the number of required replicas for the production run of MUCAREM is much less than that for the regular REM ($M\ll M$). The weight factor for this gener-

alized ensemble is now given by (see Eq. (28))

$$W_{\mathrm{MUCAREM}}(X) = \prod_{i=1}^{M} W_{\mathrm{mu}}^{\{m(i)\}}(E(x_{m(i)}^{[i]})) = \prod_{m=1}^{M} W_{\mathrm{mu}}^{\{m\}}(E(x_m^{[i(m)]})), \tag{44}$$

where we prepare the multicanonical weight factor (and the density of states) separately for m regions (see Eq. (17)):

$$W_{\mathrm{mu}}^{\{m\}}(E(x_m^{[i]})) = \exp[-\beta_m \mathcal{E}_{\mathrm{mu}}^{\{m\}}(E(x_m^{[i]}))] \equiv \frac{1}{n^{\{m\}}(E(x_m^{[i]}))}. \tag{45}$$

Here, we have introduced M arbitrary reference temperatures $T_m = 1/k_{\mathrm{B}}\beta_m$ (m=1, ..., M), but the final results will be independent of the values of T_m, as one can see from the second equality in Eq. (45) (these arbitrary temperatures are necessary only for MD simulations).

Each multicanonical weight factor $W_{\mathrm{mu}}^{\{m\}}(E)$, or the density of states $n^{\{m\}}(E)$, is defined as follows. For each m (m=1, ..., M), we assign a pair of temperatures ($T_{\mathrm{L}}^{\{m\}}$, $T_{\mathrm{H}}^{\{m\}}$). Here, we assume that $T_{\mathrm{L}}^{\{m\}} < T_{\mathrm{H}}^{\{m\}}$ and arrange the temperatures so that the neighboring regions covered by the pairs have sufficient overlaps. Without loss of generality we can assume $T_{\mathrm{L}}^{\{1\}} < ... < T_{\mathrm{L}}^{\{M\}}$ and $T_{\mathrm{H}}^{\{1\}} < ... < T_{\mathrm{H}}^{\{M\}}$. We define the following quantities:

$$\begin{cases} E_{\mathrm{L}}^{\{m\}} = \langle E \rangle_{T_{\mathrm{L}}^{\{m\}}}, \\ E_{\mathrm{H}}^{\{m\}} = \langle E \rangle_{T_{\mathrm{H}}^{\{m\}}}, \quad (m=1, ..., M). \end{cases} \tag{46}$$

Suppose that the multicanonical weight factor $W_{\mathrm{mu}}(E)$ (or equivalently, the multicanonical potential energy $E_{\mathrm{mu}}(E; T_0)$ in Eq. (18)) has been obtained as in REMUCA or by any other methods in the entire energy range of interest ($E_{\mathrm{L}}^{\{1\}} < E < E_{\mathrm{H}}^{\{M\}}$). We then have for each m (m=1, ..., M) the following multicanonical potential energies (see Eq. (42)): [73]

$$\mathcal{E}_{\mathrm{mu}}^{\{m\}}(E) = \begin{cases} \dfrac{\partial E_{\mathrm{mu}}(E_{\mathrm{L}}^{\{m\}}; T_m)}{\partial E}(E - E_{\mathrm{L}}^{\{m\}}) + E_{\mathrm{mu}}(E_{\mathrm{L}}^{\{m\}}; T_m), & \text{for } E < E_{\mathrm{L}}^{\{m\}}, \\ E_{\mathrm{mu}}(E; T_m), & \text{for } E_{\mathrm{L}}^{\{m\}} \le E \le E_{\mathrm{H}}^{\{m\}}, \\ \dfrac{\partial E_{\mathrm{mu}}(E_{\mathrm{H}}^{\{m\}}; T_m)}{\partial E}(E - E_{\mathrm{H}}^{\{m\}}) + E_{\mathrm{mu}}(E_{\mathrm{H}}^{\{m\}}; T_m), & \text{for } E > E_{\mathrm{H}}^{\{m\}}. \end{cases} \tag{47}$$

Finally, a MUCAREM simulation is realized by alternately performing the following two steps.

1. Each replica of the fixed multicanonical ensemble is simulated *simultaneously* and *independently* for a certain MC or MD steps.
2. A pair of replicas, say i and j, which are in neighboring multicanonical ensembles, say m-th and (m+1)-th, respectively, are exchanged: $X = \{..., x_m^{[i]}, ..., x_{m+1}^{[j]}, ...\} \to X' = \{..., x_m^{[j]}, ..., x_{m+1}^{[i]}, ...\}$. The transition probability of this replica exchange is given by the Metropolis criterion:

$$w(X \to X') = \min(1, \exp(-\Delta)), \tag{48}$$

where we now have (see Eq. (34)) [73]

$$\Delta = \beta_m \{\mathcal{E}_{\mathrm{mu}}^{\{m\}}(E(q^{[j]})) - \mathcal{E}_{\mathrm{mu}}^{\{m\}}(E(q^{[i]}))\} - \beta_{m+1} \{\mathcal{E}_{\mathrm{mu}}^{\{m+1\}}(E(q^{[j]})) - \mathcal{E}_{\mathrm{mu}}^{\{m+1\}}(E(q^{[i]}))\}. \tag{49}$$

Here, $E(q^{[i]})$ and $E(q^{[j]})$ are the potential energy of the i-th replica and the j-th replica, respectively.

Note that in Eq. (49) we need to newly evaluate the multicanonical potential energy, $\mathcal{E}_{\mathrm{mu}}^{\{m\}}(E(q^{[j]}))$ and $\mathcal{E}_{\mathrm{mu}}^{\{m+1\}}(E(q^{[i]}))$, because $\mathcal{E}_{\mathrm{mu}}^{\{m\}}(E)$ and $\mathcal{E}_{\mathrm{mu}}^{\{n\}}(E)$ are, in general, different functions for $m \neq n$.

In this algorithm, the m-th multicanonical ensemble actually results in a canonical simulation at $T = T_{\mathrm{L}}^{\{m\}}$ for $E < E_{\mathrm{L}}^{\{m\}}$, a multicanonical simulation for $E_{\mathrm{L}}^{\{m\}} \le E \le E_{\mathrm{H}}^{\{m\}}$, and a canonical simulation at $T = T_{\mathrm{H}}^{\{m\}}$ for $E > E_{\mathrm{H}}^{\{m\}}$, while the replica-exchange process samples states of the whole energy range ($E_{\mathrm{L}}^{\{1\}} \le E \le E_{\mathrm{H}}^{\{M\}}$).

For obtaining the canonical distributions at any intermediate temperature T, the multiple-histogram reweighting techniques [27,28] are again used. Let $N_m(E)$ and n_m be respectively the potential-energy histogram and the total number of samples obtained with the multicanonical weight factor $W_{\mathrm{mu}}^{\{m\}}(E)$ (m=1, ..., M). The expectation value of a physical quantity A at any temperature T ($=1/k_{\mathrm{B}}\beta$) is then obtained from Eq. (22), where the best estimate of the density of states is obtained by solving the WHAM equations, which now read [73]

$$n(E) = \frac{\sum_{m=1}^{M} N_m(E)}{\sum_{m=1}^{M} n_m \exp(f_m) W_{\mathrm{mu}}^{\{m\}}(E)} = \frac{\sum_{m=1}^{M} N_m(E)}{\sum_{m=1}^{M} n_m \exp(f_m - \beta_m \mathcal{E}_{\mathrm{mu}}^{\{m\}}(E))}, \tag{50}$$

and for each m ($=1, ..., M$)

$$\exp(-f_m) = \sum_E n(E) W_{\mathrm{mu}}^{\{m\}}(E) = \sum_E n(E) \exp(-\beta_m \mathcal{E}_{\mathrm{mu}}^{\{m\}}(E)). \tag{51}$$

Note that $W_{\mathrm{mu}}^{\{m\}}(E)$ is used instead of the Boltzmann factor $\exp(-\beta_m E)$ in Eqs. (37) and (38).

Moreover, ensemble averages of any physical quantity A (including those that cannot be expressed as functions of

potential energy) at any temperature T ($=1/k_B\beta$) can now be obtained from the "trajectory" of configurations of the production run. Namely, we first obtain f_m ($m=1, ..., M$) by solving Eqs. (50) and (51) self-consistently, and then we have [74]

$$\langle A\rangle_T = \frac{\sum_{m=1}^{M}\sum_{k=1}^{n_m} A(x_m(k)) \dfrac{1}{\sum_{\ell=1}^{M} n_\ell \exp(f_\ell) W_{\rm mu}^{\{\ell\}}(E(x_m(k)))} \exp[-\beta E(x_m(k))]}{\sum_{m=1}^{M}\sum_{k=1}^{n_m} \dfrac{1}{\sum_{\ell=1}^{M} n_\ell \exp(f_\ell) W_{\rm mu}^{\{\ell\}}(E(x_m(k)))} \exp[-\beta E(x_m(k))]}, \tag{52}$$

where the trajectories $x_m(k)$ ($k=1, ..., n_m$) are taken from each multicanonical simulation with the multicanonical weight factor $W_{\rm mu}^{\{m\}}(E)$ ($m=1, ..., M$) separately.

As seen above, both REMUCA and MUCAREM can be used to obtain the multicanonical weight factor, or the density of states, for the entire potential energy range of interest. For complex systems, however, a single REMUCA or MUCAREM simulation is often insufficient. In such cases we can iterate MUCA (in REMUCA) and/or MUCAREM simulations in which the estimate of the multicanonical weight factor is updated by the single- and/or multiple-histogram reweighting techniques, respectively [75].

Multidimensional replica-exchange method

We now present our multidimensional extension of REM, which we refer to as *multidimensional replica-exchange method* (MREM) [94]. (The method is also referred to as *Hamiltonian replica-exchange method* [95], *generalized parallel sampling* [129], and *Model Hopping* [130].) Some other examples of multidimensional generalized-ensemble algorithms can be found in, e.g., Refs. [120,131–137]. Another special realization of MREM is *replica-exchange umbrella sampling* (REUS) [94] and it is particularly useful in free energy calculations (see also Ref. [96] for a similar idea). REUS can be applied to ligand docking simulations and free energy calculations of ligand binding affinity [97–99].

The crucial observation that led to the new algorithm is: As long as we have M *noninteracting* replicas of the original system, the Hamiltonian $H(q, p)$ of the system does not have to be identical among the replicas and it can depend on a parameter with different parameter values for different replicas. Namely, we can write the Hamiltonian for the i-th replica at temperature T_m as

$$H_m(q^{[i]}, p^{[i]}) = K(p^{[i]}) + E_{\lambda_m}(q^{[i]}). \tag{53}$$

While replica i and temperature T_m are in one-to-one correspondence in the original REM, replica i and "parameter set" $\Lambda_m \equiv (T_m, \lambda_m)$ are in one-to-one correspondence in the new algorithm. Hence, the present algorithm can be considered as a multidimensional extension of the original replica-exchange method where the "parameter space" is one-dimensional (i.e., $\Lambda_m = T_m$). Because the replicas are non-interacting, the weight factor for the state X in this new generalized ensemble is again given by the product of Boltzmann factors for each replica (see Eq. (28)):

$$\begin{aligned} W_{\rm MREM}(X) &= \exp\Big\{-\sum_{i=1}^{M} \beta_{m(i)} H_{m(i)}(q^{[i]}, p^{[i]})\Big\} \\ &= \exp\Big\{-\sum_{m=1}^{M} \beta_m H_m(q^{[i(m)]}, p^{[i(m)]})\Big\}, \end{aligned} \tag{54}$$

where $i(m)$ and $m(i)$ are the permutation functions in Eq. (26). Then the same derivation that led to the original replica-exchange criterion follows, and the transition probability of replica exchange is given by Eq. (33), where we now have (see Eq. (34)) [94]

$$\Delta = \beta_m(E_{\lambda_m}(q^{[j]}) - E_{\lambda_m}(q^{[i]})) - \beta_n(E_{\lambda_n}(q^{[j]}) - E_{\lambda_n}(q^{[i]})). \tag{55}$$

Here, E_{λ_m} and E_{λ_n} are the total potential energies (see Eq. (53)). Note that we need to newly evaluate the potential energy for exchanged coordinates, $E_{\lambda_m}(q^{[j]})$ and $E_{\lambda_n}(q^{[i]})$, because E_{λ_m} and E_{λ_n} are in general different functions.

We remark that MUCAREM in the previous subsection is a special case of MREM. We also remark that a general formalism for multidimensional generalized-ensemble algorithms was presented in Refs. [138,139].

Parallel Monte Carlo or molecular dynamics using genetic crossover

Besides generalized-ensemble algorithms explained above, we have also proposed conformational search method using genetic crossover (PSA/GAc) [140–142]. While genetic algorithm is usually used for finding the global-minimum energy state, we have proposed to use genetic crossover to just introduce global update of conformations to enhance conformational sampling in conventional MC or MD simulations [143–145]. We refer to these methods as parallel Monte Carlo using genetic crossover (PMC/GAc) and parallel molecular dynamics using genetic crossover (PMD/GAc).

In Figure 1, we show the flow charts of the PMC/GAc or PMD/GAc method [143–145] and the original GA for comparison.

We first prepare M initial conformations of the system in study, where M is the total number of 'individuals' in GA and is usually taken to be an even integer. We then alternately perform the following two steps:

1. For the M individuals, regular canonical MC or MD simulations at temperature T are carried out simultaneously and independently for certain MC or MD steps.
2. $M/2$ pairs of conformations are selected from "parental" group randomly, and the crossover operation is performed to create "child" conformations. The obtained "child"

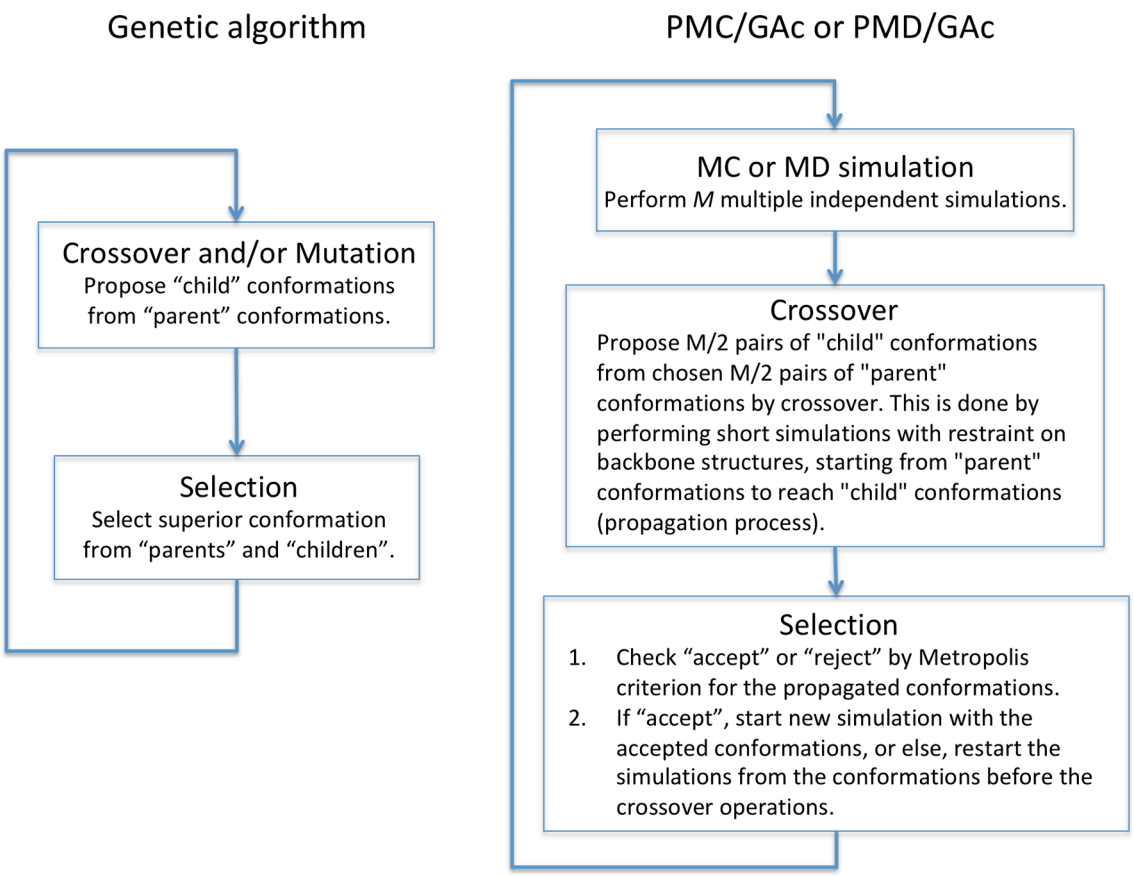


Figure 1 Flow charts of general genetic algorithm (left side) and PMC/GAc or PMD/GAc (right side).

conformations are "selected" or accepted from the parents with the following Metropolis criterion:

$$w(\mathrm{p} \to \mathrm{c}) = \min(1, \exp\{-\beta[E_\mathrm{c} - E_\mathrm{p}]\}), \tag{56}$$

where E_p and E_c stand for the potential energy of the parental conformation and the final child conformation of the parent-child pair, respectively.

In Step 2, we can employ various kinds of GAc operations. Here, we just present a case of the two-point crossover [142]. The following procedure is carried out (see Fig. 2):

1. Consecutive amino acids of length n residues in the amino-acid sequence of the conformation are selected randomly for each pair of selected conformations.
2. Dihedral angles (in only backbone or all dihedral angles) in the selected n amino acids are exchanged between the selected pair of conformations.

Note that the length n of consecutive amino-acid residues can, in general, be different for each pair of selected conformations. Motivated by the fragment assembly method [146], we take n to be an integer ranging from 2 to 10.

We need to deal with the produced "child" conformations with care. Because the produced preliminary conformations often have unnatural structures by the crossover operation, they have high potential energy and are unstable. This is particularly a serious problem for simulations with explicit solvent and lipid molecules. Therefore, a propagation process is introduced before the selection operation. As the propagation process, we perform a short MC or MD simulation with restraint potentials $E_\mathrm{rst}(\theta)$ of the (backbone) dihedral angle θ in the selected n amino acids as follows:

$$E_\mathrm{rst}(\theta) = k_\theta(\theta - \theta_\mathrm{child})^2 \tag{57}$$

where k_θ is the force constant, and θ_child is a dihedral angle proposed by exchanging dihedral angles between "parent" conformations by the crossover operation. The initial conformations for these propagation simulations are the ones before the crossover. Namely, by these propagation simulations, the corresponding backbone conformations of the n amino acids gradually transform from the ones before the crossover to the ones after the crossover.

We remark that PMC/GAc and PMD/GAc can be combined with REM for further enhancement of conformatonal space [144].

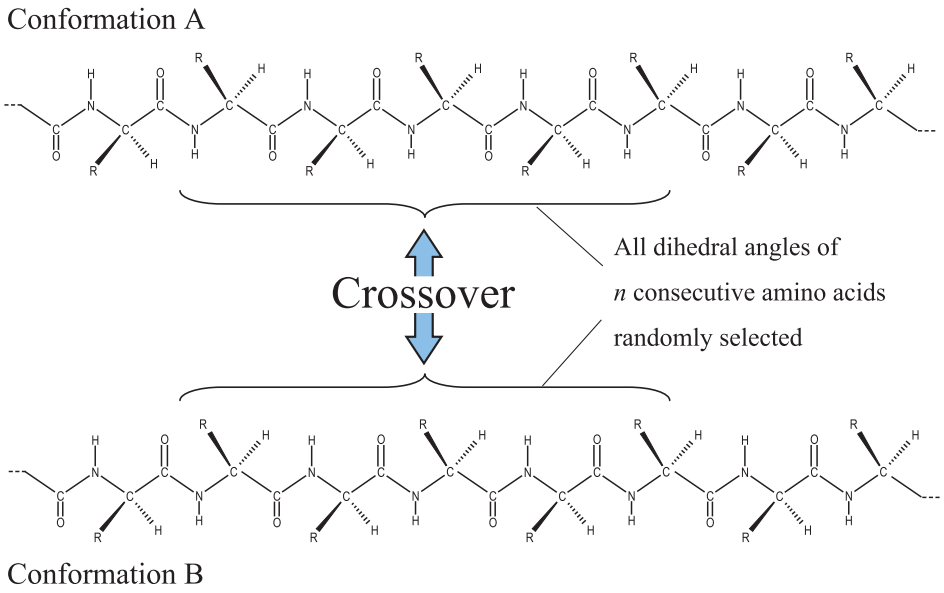


Figure 2 Schematic process of the two-point crossover operation. In this process, all dihedral angles (in backbone and side chains) within the randomly selected *n* consecutive amino acids are exchanged between a pair of conformations.

Examples of Simulation Results

The first example is the C-peptide of ribonuclease A in explicit water [85]. In the model of simulations, the N-terminus and the C-terminus of the C-peptide analogue were blocked with the acetyl group and the N-methyl group, respectively. The number of amino acids is 13 and the amino-acid sequence is: Ace-Ala-Glu$^-$-Thr-Ala-Ala-Ala-Lys$^+$-Phe-Leu-Arg$^+$-Ala-His$^+$-Ala-Nme [147,148]. The initial configuration of our simulation was first generated by a high temperature molecular dynamics simulation (at T=1000 K) in gas phase, starting from a fully extended conformation. We randomly selected one of the structures that do not have any secondary structures such as α-helix and β-sheet. The peptide was then solvated in a sphere of radius 22 Å, in which 1387 water molecules were included (see Fig. 3). Harmonic restraint was applied to prevent the water molecules from going out of the sphere. The total number of atoms was 4365. The dielectric constant was set equal to 1.0. The force-field parameters for protein were taken from the all-atom version of AMBER parm99 [151], which was found to be suitable for studying helical peptides [30], and TIP3P model [155] was used for water molecules. The unit time step, Δt, was set to 0.5 fsec.

In Table 1 the essential parameters in the simulations performed in this example are summarized.

We first performed a REMD simulation with 32 replicas for 100 psec per replica (REMD1 in Table 1). During this REMD simulation, replica exchange was tried every 200 MD steps. Using the obtained potential-energy histogram of each replica as input data to the multiple-histogram analysis in Eqs. (4) and (5), we obtained the first estimate of the multicanonical weight factor, or the density of states. We divided this multicanonical weight factor into four multicanonical weight factors that cover different energy regions [73–75] and assigned these multicanonical weight factors into four replicas (the weight factors cover the potential energy ranges from −13791.5 to −11900.5 kcal/mol, from −12962.5 to

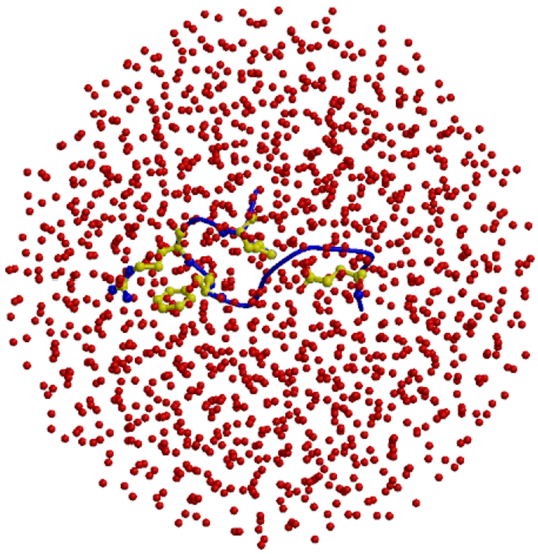


Figure 3 The initial configuration of C-peptide in explicit water, which was used in all of the 32 replicas of the first REMD simulation (REMD1 in Table 1). The red filled circles stand for the oxygen atoms of water molecules. The number of water molecules is 1387, and they are placed in a sphere of radius 22 Å. As for the peptide, besides the backbone structure (in blue), side chains of only Glu$^-$-2, Phe-8, Arg$^+$-10, and His$^+$-12 are shown (in yellow). The figure was created with Molscript [149] and Raster3D [150].

Table 1 Summary of parameters in REMD, MUCAREM, and REMUCA simulations

	Number of replicas, M	Temperature, T_m (K) (m=1, ..., M)	MD steps per replica
REMD1*	32	250, 258, 267, 276, 286, 295, 305, 315, 326, 337, 348, 360, 372, 385, 398, 411, 425, 440, 455, 470, 486, 502, 519, 537, 555, 574, 593, 613, 634, 655, 677, 700	2.0×10^5
MUCAREM1	4	360, 440, 555, 700	2.0×10^6
REMUCA1	1	700	3.0×10^7

* REMD1 stands for the replica-exchange molecular dynamics simulation, MUCAREM1 stands for the multicanonical replica-exchange molecular dynamics simulation, and REMUCA1 stands for the final multicanonical molecular dynamics simulation (the production run) of REMUCA. The results of REMD1 were used to determine the multicanonical weight factors for MUCAREM1, and those of MUCAREM1 were used to determine the multicanonical weight factor for REMUCA1.

−10796.5 kcal/mol, from −11900.5 to −9524.5 kcal/mol, and from −10796.5 to −8293.5 kcal/mol). We then carried out a MUCAREM simulation with four replicas for 1 nsec per replica (MUCAREM1 in Table 1), in which replica exchange was tried every 1000 MD steps. We again used the potential-energy histogram of each replica as the input data to the multiple-histogram analysis and finally obtained the multicanonical weight factor with high precision. As a production run, we carried out a 15 nsec multicanonical MD simulation with one replica (REMUCA1 in Table 1) and the results of this production run were analyzed in detail.

In Figure 4 we show the probability distributions of potential energy that were obtained from the above three generalized-ensemble simulations, namely, REMD1, MUCAREM1, and REMUCA1. We see in Figure 4(a) that there are enough overlaps between all pairs of neighboring canonical distributions, suggesting that there were sufficient numbers of replica exchange in REMD1. We see in Figure 4(b) that there are good overlaps between all pairs of neighboring multicanonical distributions, implying that MUCAREM1 also performed properly. Finally, the multicanonical distribution in Figure 4(c) is completely flat between around −13000 kcal/mol and around −8000 kcal/mol. The results suggest that a free random walk was realized in this energy range.

In Figure 5a we show the time series of potential energy from REMUCA1. We indeed observe a random walk covering as much as 5000 kcal/mol of energy range (note that 23 kcal/mol≈1 eV). We show in Figure 5(b) the average potential energy as a function of temperature, which was obtained from the trajectory of REMUCA1 by the reweighting techniques. The average potential energy monotonically increases as the temperature increases.

We have analyzed the data by the principal component

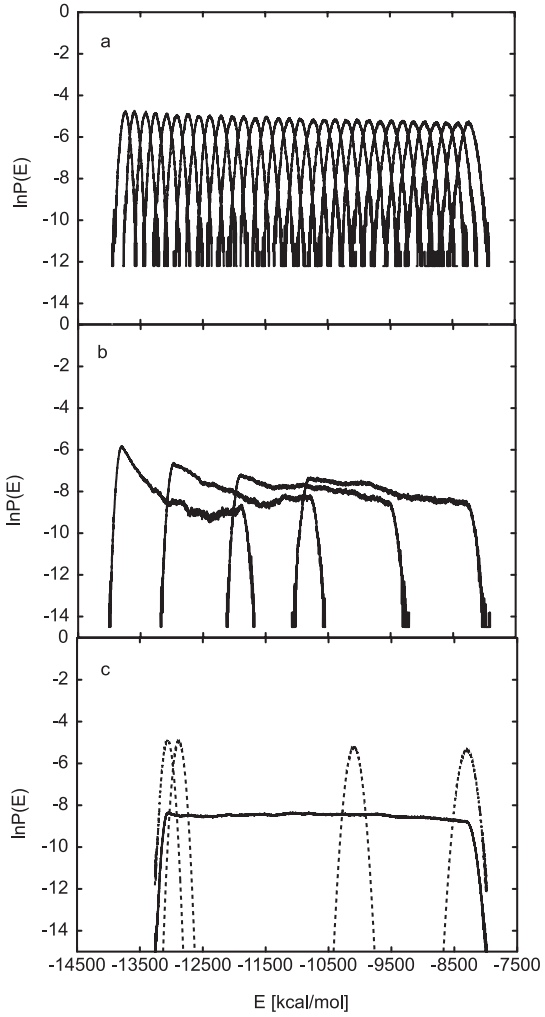


Figure 4 Probability distributions of potential energy of the C-peptide system obtained from (a) REMD1, (b) MUCAREM1, and (c) REMUCA1. See Table 1 for the parameters of the simulations. Dashed curves in (c) are the reweighted canonical distributions at 290, 300, 500, and 700 K (from left to right).

analysis (PCA) [156–158] (for a review, see Ref. [159]). In Figure 6 the potential of mean force (PMF), or free energy, along the first two principal component axes at 300 K is shown. There exist three distinct minima in the free-energy landscape, which correspond to three local-minimum-energy states. We show representative conformations at these minima in Figure 7. The structure of the global-minimum free-energy state (GM) has a partially distorted α-helix with the salt bridge between Glu^--2 and Arg^+-10. The structure is in good agreement with the experimental structure obtained by both NMR and X-ray experiments. In this structure there also exists a contact between Phe-8 and His^+-12. This contact is again observed in the corresponding residues of the X-ray structure. At LM1 the structure has a contact between Phe-8 and His^+-12, but the salt bridge between Glu^--2 and Arg^+-10 is not formed. On the other hand, the structure at LM2 has this salt bridge, but it does not have a contact

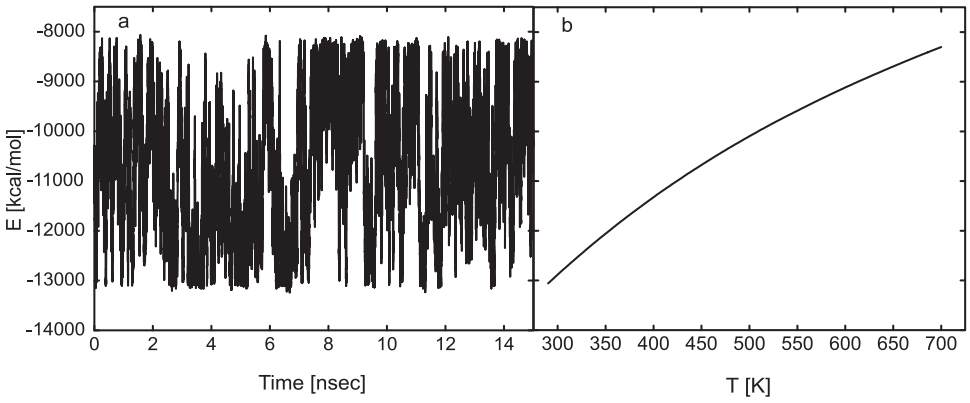


Figure 5 Time series of potential energy of the C-peptide system from the REMUCA production run (REMUCA1 in Table 1) (a) and the average potential energy as a function of temperature (b). The latter was obtained from the trajectory of REMUCA1 by the single-histogram reweighting techniques.

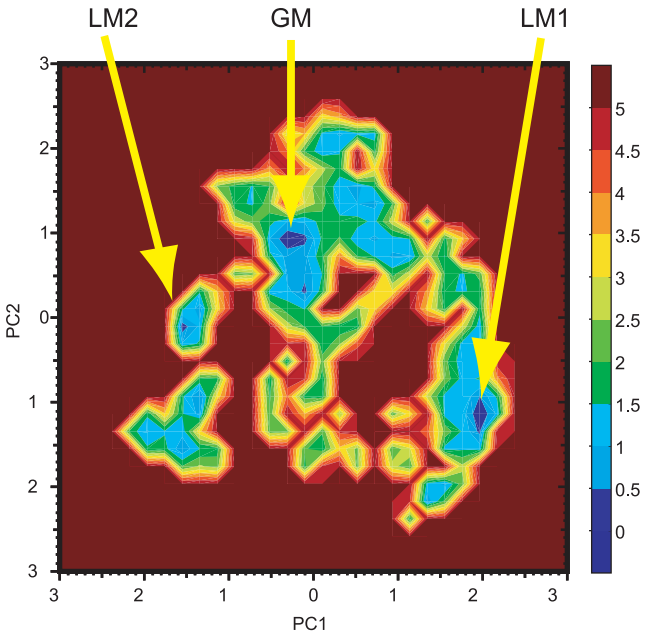


Figure 6 Potential of mean force (kcal/mol) of the C-peptide system along the first two principal components at 300 K. The free energy was calculated from the results of REMUCA production run (REMUCA1 in Table 1) by the single-histogram reweighting techniques and normalized so that the global-minimum state (GM) has the value zero. GM, LM1, and LM2 represent three distinct minimum free-energy states.

between Phe-8 and His^{+}-12. Thus, only the structures at GM satisfy all of the interactions that have been observed by the X-ray and other experimental studies.

The second example is a MUCAREM simulation of folding of a small protein [127]. The system that we simulated is chicken villin headpiece subdomain in explicit water. The number of amino acids is 36. The force field CHARMM22 [152] with CMAP [153,154] and TIP3P water model [152,155] were used. The number of water molecules was 3513. The computer code developed in Refs. [73,92,160,161], which is based on the version 2 of PRESTO [162], was used after modification for calculatoin with the CHARMM force field. The MD time step was 1.0 fsec. We made two production runs of about 1 μsec, each of which was a MUCAREM simulation with eight replicas. They are referred to as MUCAREM1 and MUCAREM2. The former consisted of 1.127 μsec covering the temperature range between 269 K and 699 K, and the latter 1.157 μsec covering the temperature range between 289 K and 699 K.

In Figure 8 we show the time series of the mainchain root-mean-square deviation (RMSD) during MUCAREM1 and MUCAREM2.

We consider that the backbone folded into the native structure from unfolded ones if the mainchain RMSD becomes $\leq$3.0 Å. The folding event is counted separately if it goes through an unfolded structure (with the backbone RMSD $\geq$6.5 Å). With this criterion, we observed 11 folding events in seven different replicas (namely, Replicas 5, 7, and 8 in MUCAREM1 and Replicas 1, 2, 4, and 5 in MUCAREM2).

In Figure 9 we show the snapshots of the replicas folding into native-like conformations for the two MUCAREM production runs.

In Figure 10 we compare the obtained low-RMSD conformations and the native structure. They are indeed very close to the native structure.

The third example is a REM MC simulation for the prediction of membrane protein structures [163]. Here, we used another implicit membrane model [164–169], where four elementary harmonic restraints are added to the original CHARMM potential energy function in order to mimic restrained membrane environment. Only the transmembrane helices are used in our simulations, and loop regions of the membrane proteins as well as lipid and water molecules were neglected. While membrane environment enhances the stability of helix structures, it also restricts sampling in

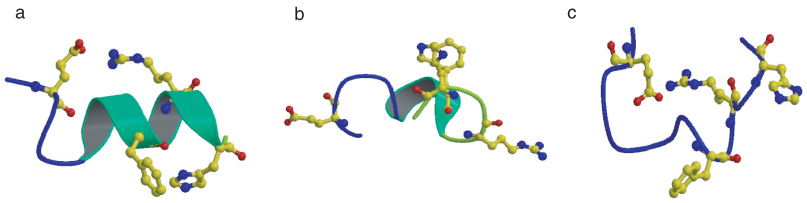


Figure 7 The representative structures at the global-minimum free-energy state ((a) GM) and the two local-minimum states ((b) LM1 and (c) LM2). As for the peptide structures, besides the backbone structure, side chains of only Glu^{-}-2, Phe-8, Arg^{+}-10, and His^{+}-12 are shown in ball-and-stick model.

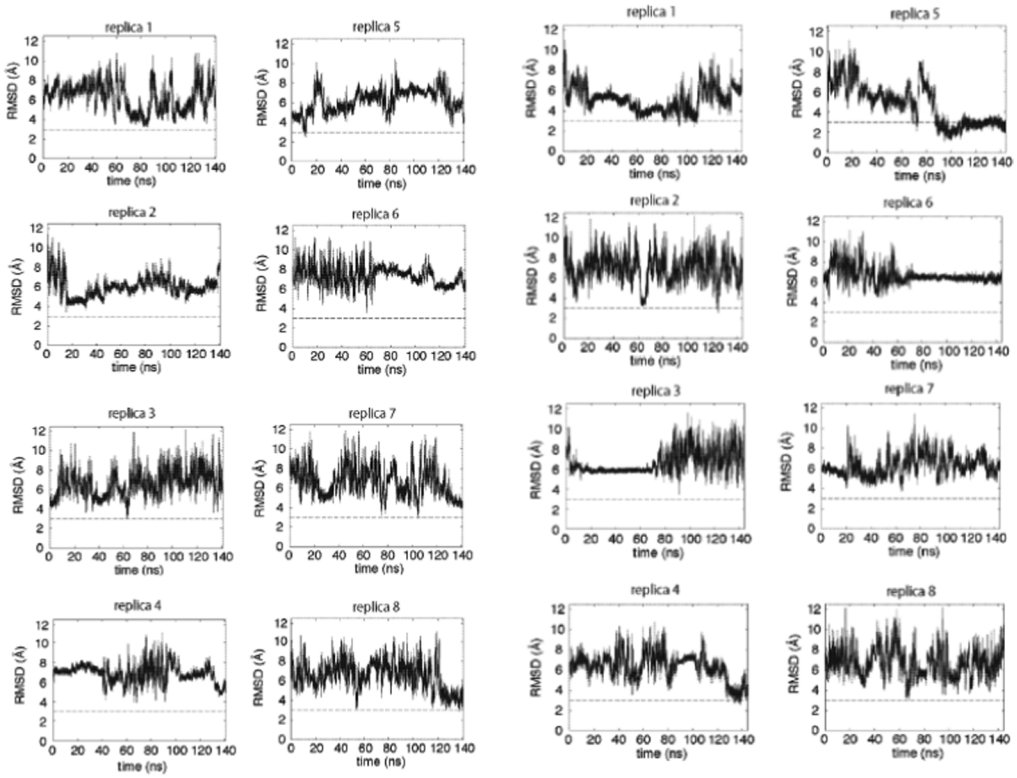


Figure 8 Time series of the backbone RMSD from the native structure of villin headpiece during MUCAREM1 (left) and MUCAREM2 (right).

conformational space in the membrane region. Thus, this model greatly reduces the search area in the conformational space during folding processes. This model is supported by many experimental data such as two-stage model (for a review, see Ref. [170]). As for a simulation method, we employed the replica-exchange Monte Carlo (MC).

Although the previous method treated helix structures as rigid bodies and required the known helix structures from the native structure, this approach reproduced nativelike structures of bacteriorhodopsin [167,168], which has seven transmembrane helices with 249 amino acids and a retinal from Halobacterium salinarum (PDB ID: 1IW6 [171], 1PY6 [172], 1BRR [173]). This protein has a function of proton pump in bio-membrane with excitations of the retinal molecule by light. However, 60% of all transmembrane helix structures are distorted in PDB, which seems to be related to their functions. Thus, we also focus our interests in this work on the reproduction of their distortions in larger proteins such as bacteriorhodopsin which has seven distorted helices.

To solve this problem we extended the above method to include the flexible treatment of backbone structures and reproduced small membrane protein structures: glycophorin A of a transmembrane helix dimer and phospholamban of a distorted transmembrane helix [169].

Our method for membrane structure predictions is as follows. We first obtain the amino-acid sequences of transmembrane helices of the target protein by bioinformatics servers such as SOSUI [174], TMHMM [175], MEMSAT [176], and HMMTOP [177]. In the present work, however, the amino-acid sequences of transmembrane helices were taken from the experimental structure as in the previous work [167,168]. Different servers may predict different

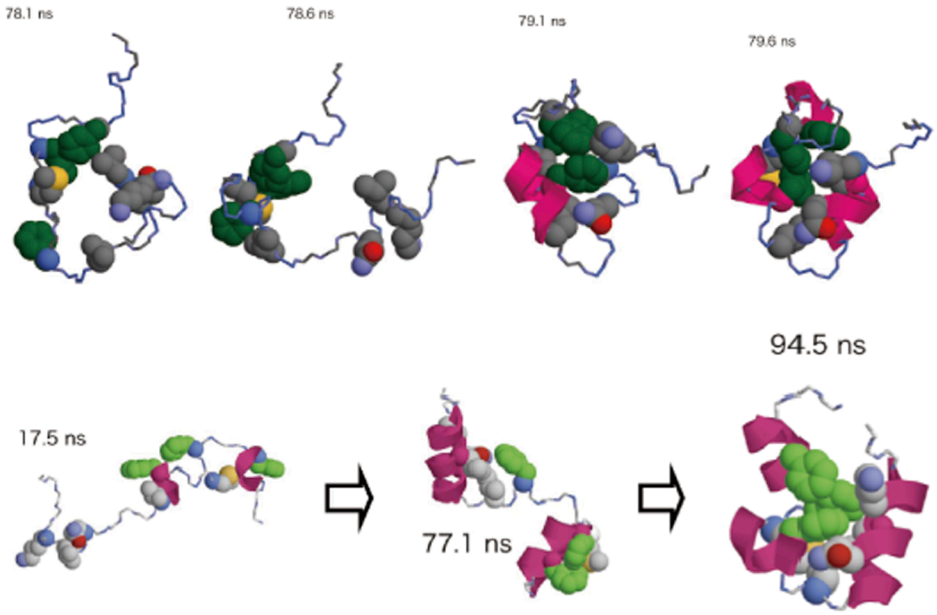


Figure 9 Snapshots of villin headpiece during the MUCAREM production runs that folded into native-like conformations: MUCAREM1 (above) and MUCAREM2 (below).

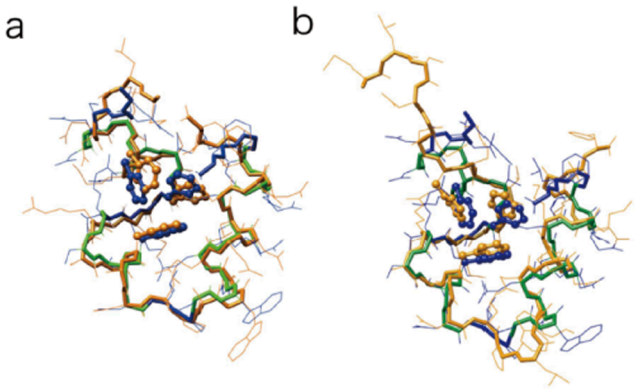


Figure 10 Low-RMSD conformations of villin headpiece obtained in MUCAREM1 and MUCAREM2 (colored in orange). The X-ray structure (PDB ID: 1YRF) is also superimposed (colored in blue and green). Here, the α-helices in the X-ray structure are colored in green and the rest in blue. Three phenylalanine side chains (Phe7, Phe11, and Phe18), which form a hydrophobic core, are shown in ball-and-stick representation. (a) The lowest-backbone-RMSD conformation observed in the two MUCAREM production runs (Replica 5 of MUCAREM2). The backbone RMSD value is 1.1 Å (for non-terminal 34 residues). (b) A low-RMSD conformation observed in MUCAREM1 (Replica 8). The RMSD value is 1.0 Å for residues 9 to 32 and 3.3 Å for non-terminal 34 residues.

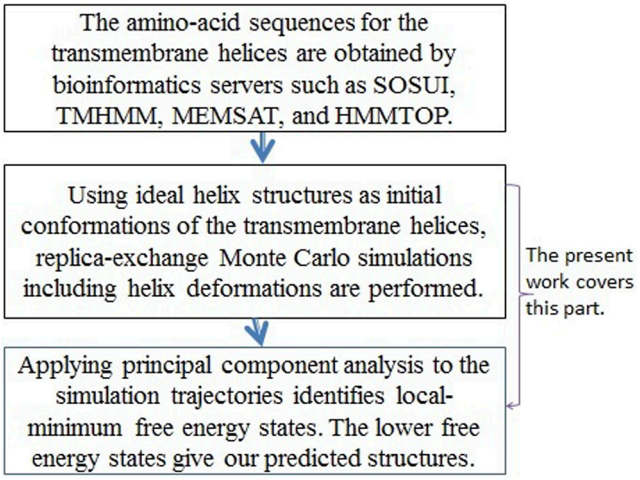


Figure 11 Our prediction procedure for membrane protein structures.

helix ends. In order to study end dependence we also made shorter simulations with a little shorter helices and confirmed that we obtained similar results. We then perform REM simulations of these transmembrane helices. Our procedure for membrane protein structure predictions is summarized in Figure 11 [169]. The MC program is based on CHARMM macromolecular mechanics program [178,179], and replica-exchange Monte Carlo method was implemented in it. Only the transmembrane helices were used in our simulations, and loop regions of the membrane proteins as well as lipid and water molecules were neglected. The membrane environment for this protein for the membrane thickness and the region of transmembrane region of the helices was taken from Orientation of Proteins in Membrane (OPM) [180]. The amino-acid sequences of the transmembrane helices are EWIWLALGTALMGLGTLYFLVKG (9–31), KFYAITTLVPAIAFTMYLSMLL (41–62), IYWARYADW LFTTPLLLLDLALL (78–100), QGTILALVGADGIMIGT GLVGAL (105–127), RFVWWAISTAAMLYILYVLFFGF (134–156), TFKVLRNVTVVLWSAYPVVWLIGSE (170–

194), and LNIETLLFMVLDVSAKVGFGLILL (201–224), which are the same as in Refs. [167,168], where two numbers in parentheses after each sequence correspond to the first and last amino-acid numbers of each helix. The N-terminus and the C-terminus of each helix were blocked with the acetyl group and the N-methyl group (first residue and last residue), respectively. The initial structure for each helix was an ideal helix structure and they were placed in the membrane region randomly. We added the following four elementary harmonic restraints as a simple implicit membrane model to the original CHARMM potential energy function of bond length, bond angle, torsion angle, van der Waals, electrostatic interaction, etc. in order to mimic the restrained membrane environment. The restraint energy function is given by

$$E_{\mathrm{restr}} = E_{\mathrm{restr1}} + E_{\mathrm{restr2}} + E_{\mathrm{restr3}} + E_{\mathrm{restr4}}, \tag{58}$$

where each term is defined as follows:

$$E_{\mathrm{restr1}} = \sum_{i=1}^{N_{\mathrm{H}}-1} k_1\,\theta(r_{i,i+1} - d_{i,i+1})[r_{i,i+1} - d_{i,i+1}]^2, \tag{59}$$

$$\begin{aligned} E_{\mathrm{restr2}} = \sum_{i=1}^{N_{\mathrm{H}}} \{ & k_2\,\theta(|z_i^{\mathrm{L}} - z_{0,i}^{\mathrm{L}}| - d^{\mathrm{L}})[|z_i^{\mathrm{L}} - z_{0,i}^{\mathrm{L}}| - d^{\mathrm{L}}]^2 \\ & + k_2\,\theta(|z_i^{\mathrm{U}} - z_{0,i}^{\mathrm{U}}| - d^{\mathrm{U}})[|z_i^{\mathrm{U}} - z_{0,i}^{\mathrm{U}}| - d^{\mathrm{U}}]^2 \}, \end{aligned} \tag{60}$$

$$E_{\mathrm{restr3}} = \sum_{C_\alpha} k_3\,\theta(r_{C_\alpha} - d_{C_\alpha})[r_{C_\alpha} - d_{C_\alpha}]^2, \tag{61}$$

$$\begin{aligned} E_{\mathrm{restr4}} = & \sum_{j=1}^{N_{BD}} k_4\,\theta(|\phi_j - \phi_0| - \alpha_j^{\phi})[|\phi_j - \phi_0| - \alpha_j^{\phi}]^2 \\ & + \sum_{j=1}^{N_{BD}} k_5\,\theta(|\psi_j - \psi_0| - \alpha_j^{\psi})[|\psi_j - \psi_0| - \alpha_j^{\psi}]^2. \end{aligned} \tag{62}$$

E_{restr1} is the energy that restrains pairs of adjacent helices along the amino-acid chain not to be apart from each other too much (loop restraints), where $r_{i,i+1}$ is the distance between the C atom of the C-terminus of the i-th helix and the C^α atom of the N-terminus of the (i+1)-th helix, and k_1 and $d_{i,i+1}$ are the force constant and the central value constant of the harmonic restraints, respectively. Each $d_{i,i+1}$ is proportional to the loop length connected between helices. $\theta(x)$ is the step function, which has 1 when x is larger than or equal to 0, otherwise zero. N_{H} is the total number of transmembrane helices in the protein.

E_{restr2} is the energy that restrains helix N-terminus and C-terminus to be located near membrane boundary planes. Here, the z-axis is defined to be the direction perpendicular to the membrane boundary planes. k_2 is the force constant of the harmonic restraints. $z_{0,i}^{\mathrm{L}}$ and $z_{0,i}^{\mathrm{U}}$ are the z-coordinate values of the C^α atom of the N-terminus or C-terminus of the i-th helix near the fixed lower membrane boundary and the upper membrane boundary, respectively. $z_{0,i}^{\mathrm{L}}$ and $z_{0,i}^{\mathrm{U}}$ are the fixed lower boundary z-coordinate value and the upper boundary z-coordinate value of the membrane planes, respectively, and here they depend on each helix atoms due to the known data from OPM [180] although constant membrane plane region is also possible like a previous research condition. d^{L} and d^{U} are the corresponding central value constants of the harmonic restraints. This term has a non-zero value only when the C^α atoms of the N-terminus or C-terminus of the i-th helix are apart more than d_i^{L} (or d_i^{U}). This restraint energy was introduced so that the helix ends are not too much apart from the membrane boundary planes.

E_{restr3} is the energy that restrains all C^α atoms within the sphere (centered at the origin) of radius d_{C_α}. r_{C_α} is the distance of C^α atoms from the origin, and k_3 and d_{C_α} are the force constant and the central value constant of the harmonic restraints, respectively.

E_{restr4} is the energy that restraints the dihedral angles of the main chain so that helix structures may not deviate too much from ideal helix structures, preventing them from forming random-coil structures. N_{BD} is the total number of (ϕ, ψ) angles in the helix backbones. Here, all the backbone dihedral angles ϕ_j and ψ_j (j=1, ..., N_{BD}) are restrained. ϕ_0 and ψ_0 are the reference value of the harmonic restraint to keep the helix structures without forming random coil structure, and α_j^{ϕ}, α_j^{ψ} are the anges of the harmonic restraints.

We set k_1=5.0, $d_{i,i+1}$=(46, 53, 34, 19, 95, 30) where i=1, 2, ..., 6, k_2=5.0, $z_{0,i}^{\mathrm{L}}$=(−14, −16, −20, −15, −19, −24, −18) where i=1, 2, ..., 7, $z_{0,i}^{\mathrm{U}}$=(12, 14, 15, 15, 14, 11, 12) where i=1, 2, ..., 7, d^{U}=d^{L}=2.0, k_3=0.5, d_{C_α}=80, k_4=30.0, k_5=30.0, ϕ_0=−62, ψ_0=−40, α_j^{ϕ}=16, and α_j^{ψ}=13.

We used 40 replicas and the following temperatures: 400, 415, 435, 455, 485, 518, 552, 589, 629, 671, 716, 764, 815, 870, 928, 990, 1056, 1127, 1202, 1283, 1369, 1460, 1558, 1662, 1774, 1892, 2019, 2154, 2298, 2452, 2616, 2791, 2978, 3177, 3390, 3616, 3808, 4050, 4250, and 4500 K. We remark that because short simulations suggested that the choice of the minimum temperature of 300 K did not change overall helix orientations compared to 400 K, we set the minimum temperature to 400 K instead of 300 K to reduce the number of replicas. We used rather high temperature values compared to experimental conditions. This is because our implicit membrane model guarantees the helix stability and enhances conformational sampling. Replica exchange was attempted at every 50 MC steps. We performed four independent simulations in total of 1,363,925,000 MC steps (the number of MC steps in each simulation was 308,000,000, 250,000,000, 343,200,000, and 462,775,000 MC steps).

We used the CHARMM19 parameter set (polar hydrogen model) for the potential energy of the system [178,179]. No cutoff was introduced to the non-bonded terms. Each helix structure was first minimized subjected to harmonic restraint on all the heavy atoms. In order to prepare random initial conformations, we first performed regular constant temperature MC simulations of all the replicas for 3,000,000 MC steps at 4500 K. We then performed equilibrium MC simulation for 3,000,000 MC steps at the above 40 temperatures,

and the last conformation for each replica was the initial structure for the REM simulations. We repeated this process four times for four independent REM simulations with different seeds. In those simulations, the dielectric constant was set to ϵ=1.0 as in the previous works [164–169], because these works confirmed that the results with ϵ=1.0 value were in better agreement with the experimental structures than ϵ=4.0. The interpretation was that in the native structure there are few lipid molecules between helices. In MC move, we updated conformations with a rigid translation and rotation of each α-helix, a rotation of torsion angles of backbones by directional manipulation and concerted rotation [181–183], and torsion rotations of side-chains. There are $2N_H+N_{SD}+N_{BD}+N_{CR}$ kinds of MC moves, where N_{SD} is the total number of dihedral angles in the side-chains of N_H helices and N_{CR} is the total number of the combination of seven successive backbone torsion angles by the concerted rotation in the helix backbone. One MC step in this article is defined to be an update of one of these degrees of freedom, which is accepted or rejected according to the Metropolis criterion.

We investigated the free energy landscape obtained by the principal component analysis. We classified the sampled structures at the minimum temperature of 400 K into clusters of similar structures by the k-means clustering method [184]. Although we can express the system more accurately as we use more principal axes in k-means clustering, we here classify and analyze the sampled structures at the lowest temperature by the first three principal components. In Figure 12, the projection of sampled structures from the REM simulations at 400 K on the first, second, and third principal component axes. We obtained five distinct clusters of similar structures. If we perform constant temperature simulations at the lowest temperature, the simulations will get trapped in any of the clusters in Figure 12, depending on the initial conformations of the simulations. However, each replica during the REM simulations did not get trapped in one of the local-minimum free energy states, by going through high temperature regions. Every replica could overcome energy barriers at higher temperatures during the simulations. This is the advantage of the replica-exchange method. Table 2 lists average quantities of five clusters of similar structures and at the temperature of 400 K. The rows of Cluster 1, Cluster 2, Cluster 3, Cluster 4 and Cluster 5 represent various average values for the structures that belong to each cluster. The number of structures in each cluster (the total number was 42238 structures) was 9123, 13146, 7457, 5121, and 4418 for Cluster 1, Cluster 2, Cluster 3, Cluster 4, and Cluster 5, respectively. Thus, the global-minimum free energy state is Cluster 2, and the second-lowest minimum state is Cluster 1.

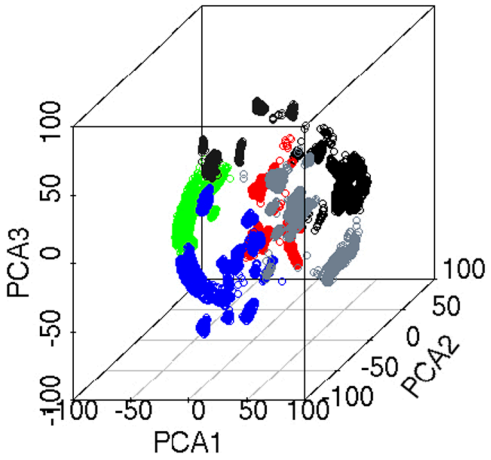


Figure 12 Projection of sampled structures at temperature 400 K on the first, second, and third principal axes from the REM simulations. Structures are classified into clusters of similar structures by k-means method and analyzed in detail. Clusters are highlighted by different colors: red, blue, yellow, green, and black. PCA1, PCA2, and PCA3 represent the principal component axes 1, 2, and 3, respectively.

Figure 13 shows the representative structure in each cluster from the highest density region. The root-mean-square-deviation (RMSD) value of each representative structure with respect to the C^α atoms was 3.6 Å, 8.8 Å, 15.8 Å, 15.9 Å, and 16.6 Å for Cluster 1, Cluster 2, Cluster 3, Cluster 4, and Cluster 5, respectively. From these RMSD values, we see that the native-like structure is the second-lowest free energy state (Cluster 1) and that the global-minimum free energy state (Cluster 2) is the second closest to the native structure. In the structure of Cluster 2, the space where the retinal molecule occupies in the native structure is filled with a helix,

Table 2 Various average quantities for each cluster and at the temperature of 400 K

	Str	E_{tot}	E_{elec}	E_{vdw}	E_{dih}	E_{geo}	RMSD
Cluster 1	9123	−7587±355	−7006±51	−1197±83	202±14	106±215	6.2±5.6
Cluster 2	13146	−7422±182	−6976±45	−1166±44	213±11	147±123	12.2±2.4
Cluster 3	7457	−7287±210	−6943±31	−1143±71	211±10	229±142	14.9±2.5
Cluster 4	5121	−7421±109	−7010±36	−1145±43	213±10	160±71	16.3±1.1
Cluster 5	4418	−7300±309	−6954±41	−1096±67	207±11	184±258	16.7±1.0
400 K	42240	−7412±263	−6979±48	−1158±68	210±12	166±170	12.5±4.8

The following abbreviations are used: Str: number of structures, E_{tot}: average total potential energy, E_{elec}: average electrostatic energy, E_{vdw}: average Lennerd-Jones energy, E_{dih}: average dihedral energy, E_{geo}: average constraint energy (all in kcal/mol), RMSD: average root-mean-square deviation of all C^α atoms (in Å).

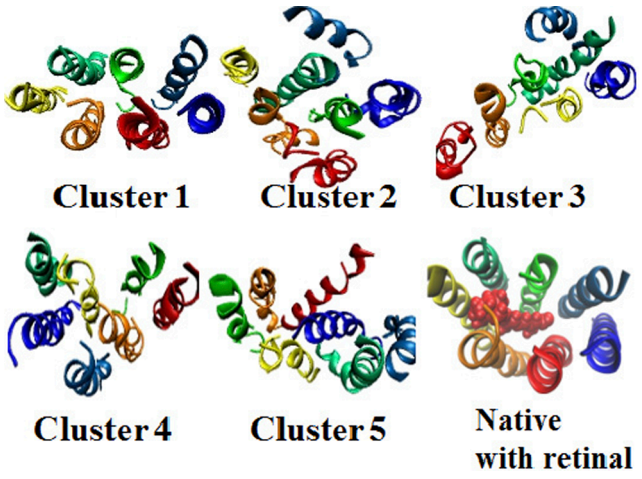


Figure 13 Typical structures in each cluster selected in the highest density region. The RMSD from the native conformation with respect to all C^{α} atoms is 3.6 Å, 8.8 Å, 15.8 Å, 15.9 Å, and 16.6 Å for Cluster 1, Cluster 2, Cluster 3, Cluster 4, and Cluster 5, respectively. Helices are colored from the N-terminus to the C-terminus: blue (Helix A), light-blue (Helix B), green (Helix C), deepgreen (Helix D), yellow (Helix E), orange (Helix F), and red (Helix G).

and this increases the contact between helices and seems to stabilize this structure more than the native-like structure of Cluster 1 with the empty space for the retinal molecule. Moreover, this result that a helix occupies the retinal space is consistent with previous works [167,168] which did not include the flexibility of helix structures. However, the previous works were not able to obtain the native-like structure such as Cluster 1. Hence, the extension of including the freedom of helix structure distortion has improved the accuracy of prediction for membrane protein structure determination by simulation. Our results suggest that in the simulations without a retinal molecule the structures can interchange between the structures of Cluster 1 and Cluster 2. After an insertion of a retinal, it then stabilizes the native-like structure. It is important that the association of helices enabled them to make a room for an insertion of a retinal molecule. This is consistent with the experimental results of bacteriorhodopsin, which observed the spontaneous insertion of a retinal molecule by a helix association [185].

We now examine the distortions of each helix of native-like structure in Cluster 1. They are compared with the native structures in Figure 14. Note that the positions of the kinks of helices are all correctly reproduced. The RMSD values with respect to backbone atoms are 0.7 Å, 0.9 Å, 1.6 Å, 2.7 Å, 1.1 Å, 2.2 Å, and 0.4 Å for Helix A, Helix B, Helix C, Helix D, Helix E, Helix F, and Helix G, respectively. These helix structures reproduced the bending of helices with smaller RMSD values for the helices apart from the retinal molecule in the native state (Helices A, B, and G), whereas the helices with larger RMSD values of about 2.0 Å are for the helices in close contact with the retinal in the native structure (Helices C, D, E, and F). They suggest that the

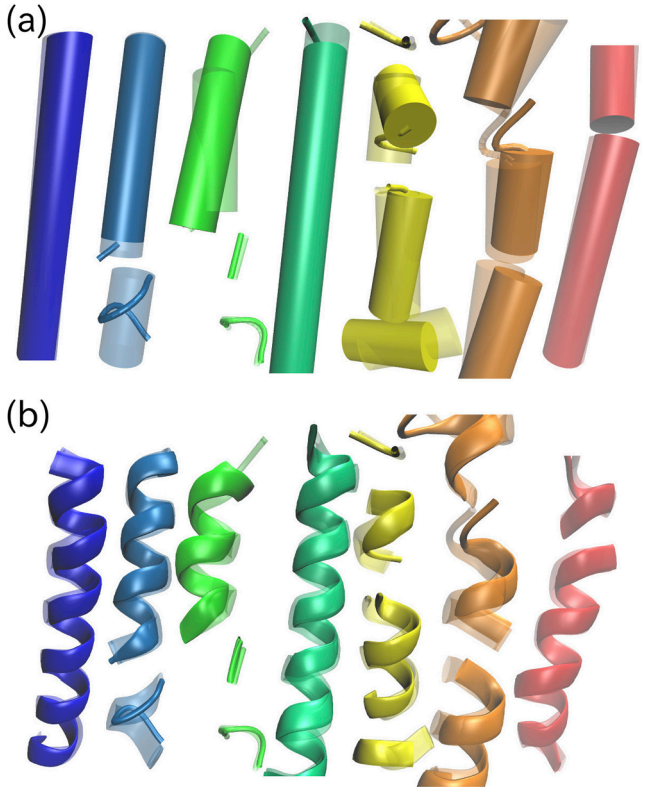


Figure 14 Comparison of distortions of each helix structure in Cluster 1 with the native helix structure by two different representation methods, (a) cartoon and (b) ribbon. The transparent structures are the native structures. The RMSD from the native conformation with respect to backbone atoms is 0.7 Å, 0.9 Å, 1.6 Å, 2.7 Å, 1.1 Å, 2.2 Å, and 0.4 Å for Helix A, Helix B, Helix C, Helix D, Helix E, Helix F, and Helix G, respectively. The coloring of helices is the same as in Figure 13.

larger RMSD values came from the neglect of the retinal molecule in our simulations.

In order to study bendings of helices more quantitatively, we analyzed the data by the software HELANAL-Plus. They are in good agreement between PDB and Cluster 1 structures. Most properties of helices are similar between them, and because of the similarity in maximum bending and average bending with standard errors, this suggests that the tendency of bending was highly reproduced, although overall helix geometry is assigned differently (L vs. K) in Helix D, Helix E, and Helix G. However, there is disagreement in data about the maximum bending location in Helix D (85T for PDB and 77G for Cluster 1). This seems to result from the lack of interactions with the retinal molecule because the retinal binding positions are 79D, 82M, and 86G near the C-terminus side.

These results imply that the position of the kinks of helices are determined by their amino-acid sequences and that the amount of bends of helices are determined by the interactions with surrounding molecules such as other helices, retinal, and lipid molecules.

The fourth example is a PMD/GAc simulation of a small

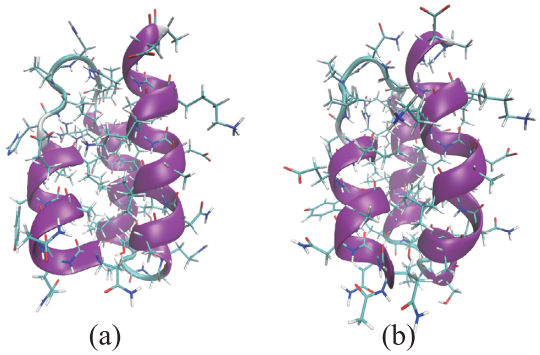


Figure 15 Structures of protein A. (a) PDB structure (PDB ID: 1BDD). (b) A conformation obtained from the present simulation, which has the lowest RMSD value from the PDB structure (RMSD=1.7 Å).

protein [143,144]. We applied the present method to protein A. Although the whole protein A has 60 amino acids, we used the truncated 46 amino-acid sequence from Gln10 to Ala55. For this simulation, we used the AMBER12 program package and incorporated the two-point genetic crossover procedure. The unit time step was set to 2.0 fs and the bonds involving hydrogen atoms were constrained by the SHAKE algorithm [186]. Each simulation for sampling was carried out for 90.0 nsec (which consisted of 45,000,000 MD steps) with 32 individuals (M=32) and performed the crossover operations 90 times during the simulation. The temperature during the simulations was kept at 300 K by using Langevin dynamics. The nonbonded cutoff of 20 Å was used. As for solvent effects, we used the GB/SA model [187] included in the AMBER12 program package (igb=5). In the crossover operations, we set the length n of consecutive amino-acid residues to be an even integer ranging from 10 to 20. This number was chosen randomly for each pair of parental conformations. As for the equilibration simulations just after each genetic crossover operation, the first simulations with the harmonic restraints on the backbone diheral angles of n amino-acid residues (the force constants were 600 kcal/mol·Å^2) lasted for 20 psec, and the following simulations without restraints also lasted for 20 psec.

We obtained a similar conformation to the experimental native structure, and its root-mean-square distance (RMSD) (for only the backbone atoms) from the native structure was 1.7 Å (see Fig. 15).

Conclusions

In this Festschrift Issue for the celebration of Professor Nobuhiro Gō's 80th birthday, we reviewed enhanced conformational sampling methods for protein structure predictions. We first introduced two well-known generalized-ensemble algorithms, namely, MUCA and REM, which can greatly enhance conformational sampling of biomolecular systems. We then presented various extensions of these algorithms, namely, REMUCA, MUCAREM, and MREM. We also presented PMC/GAc and PMD/GAc, which are based on genetic crossover.

These methods can be used not only to predict the tertiary structures of proteins but also to calculate free energy landscapes of protein folding and protein-ligand docking, etc.

Acknowledgements

The author thanks his co-workers for useful discussions. In particular, he is grateful to Drs. Ulrich H. E. Hansmann, Tomoyuki Hiroyasu, Katsuya Ishii, Akio Kitao, Hironori Kokubo, Mitsunori Miki, Ayori Mitsutake, Yoshitake Sakae, John E. Straub, Yuji Sugita, Ryo Urano, and Takao Yoda for collaborations that led to the results presented in the present article. Some of the results were obtained by the computations on the super computers at Nagoya University Supercomputer Center, the Institute for Molecular Science, Okazaki, and the Institute for Solid State Physics, University of Tokyo, Japan.

Conflict of Interest

The author declares no conflict of interest.

Author Contributions

Y. O. reviewed enhanced conformational sampling methods and wrote the manuscript.

References

[1] Gō, N. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* **12**, 183–210 (1983).
[2] Bryngelson, J. D. & Wolynes, P, G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA* **84**, 7524–7528 (1987).
[3] 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).
[4] Camacho, C. J. & Thirumalai, D. Kinetics and thermodynamics of folding in model proteins. *Proc. Natl. Acad. Sci. USA* **90**, 6369–6372 (1993).
[5] Klimov, D. K. & Thirumalai, D. Criterion that determines the foldability of proteins. *Phys. Rev. Lett.* **76**, 4070–4073 (1996).
[6] Dill, K. & Chan, H. S. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19 (1997).
[7] Hansmann, U. H. E., Masuya, M. & Okamoto, Y. Characteristic temperatures of folding of a smallpeptide. *Proc. Natl. Acad. Sci. USA* **94**, 10652–10656 (1997).
[8] Hansmann, U. H. E., Okamoto, Y. & Onuchic, J. N. The folding funnel landscape for the peptide Met-enkephalin. *Proteins* **34**, 472–483 (1999).
[9] Braun, W. & Gō, N. Calculation of protein conformations by proton-proton distance constraints: A new efficient algorithm. *J. Mol. Biol.* **186**, 611–626 (1985).
[10] Crippen, G. M. *Distance Geometry and Conformational Calculations* (Research Studies Press, New York, 1981).
[11] Kirkpatrick, S., Gelatt, C. D. Jr. & Vecchi, M. P. Optimization

by simulated annealing. *Science* **220**, 671–680 (1983).
[12] Nilges, M., Clore, G. M. & Gronenborn, A. M. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry- dynamical simulated annealing calculations. *FEBS Lett.* **229**, 317–324 (1988).
[13] Brünger, A. T. Crystallographic refinement by simulated annealing. Application to a 2.8 Å resolution structure of aspartate aminotransferase. *J. Mol. Biol.* **203**, 803–816 (1988).
[14] Harvel, T. F. An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonance. *Prog. Biophys. Mol. Biol.* **56**, 43–78 (1991).
[15] Nakai, T., Kidera, A. & Nakamura, H. Intrinsic nature of the three-dimensional structure of proteins as determined by distance geometry with good sampling properties. *J. Biomol. NMR* **3**, 19–40 (1993).
[16] Wilson, S. R., Cui, W., Moskowitz, J. W. & Schmidt, K. E. Conformational analysis of flexible molecules: location of the global minimum energy conformation by the simulated annealing method. *Tetrahedron Lett.* **29**, 4373–4376 (1988).
[17] Kawai, H., Kikuchi, T. & Okamoto, Y. A prediction of tertiary structures of peptide by the Monte Carlo simulated annealing method. *Protein Eng.* **3**, 85–94 (1989).
[18] Wilson, C. & Doniach, S. A computer model to dynamically simulate protein folding: studies with crambin. *Proteins* **6**, 193–209 (1989).
[19] Kawai, H., Okamoto, Y., Fukugita, M., Nakazawa, T. & Kikuchi, T. Prediction of α-helix folding of isolated C-peptide of ribonuclease A by Monte Calro simulated annealing. *Chem. Lett.* **20**, 213–216 (1991).
[20] Okamoto, Y., Fukugita, M., Nakazawa, T. & Kawai, H. α-Helix folding by Monte Carlo simulated annealing in isolated C-peptide of ribonuclease A. *Protein Eng.* **4**, 639–647 (1991).
[21] Hansmann, U. H. E. & Okamoto, Y. New Monte Carlo algorithms for protein folding. *Curr. Opin. Struct. Biol.* **9**, 177–183 (1999).
[22] Mitsutake, A., Sugita, Y. & Okamoto, Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* **60**, 96–123 (2001).
[23] Okumura, H., Itoh, S. G. & Okamoto, Y. Generalized-ensemble algorithms for simulations of complex molecular systems. in *Practical Aspects of Computational Chemistry II: An Overview of the Last Two Decades and Current Trends* (Leszczynski, J. & Shukla, M. K. eds.) pp. 69–101 (Springer, Dordrecht, 2012).
[24] Mitsutake, A., Mori, Y. & Okamoto, Y. Enhanced sampling algorithms. in *Biomolecular Simulations: Methods and Protocols* (Monticelli, L. & Salonen, E. eds.) pp. 153–195 (Humana Press, New York, 2013).
[25] Okamoto, Y. Structural fluctuations of proteins in folding and ligand docking studied by generalized-ensemble simulations. in *Molecular Science of Fluctuations toward Biological Functions* (Terazima, M., Kataoka, M., Ueoka, R. & Okamoto, Y. eds.) pp. 183–204 (Springer, Tokyo, 2016).
[26] Ferrenberg, A. M. & Swendsen, R. H. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.* **61**, 2635–2638 (1988).
[27] Ferrenberg, A. M. & Swendsen, R. H. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* **63**, 1195–1198 (1989).
[28] Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).
[29] Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. How fast-folding proteins fold. *Science* **334**, 517–520 (2011).
[30] Yoda, T., Sugita, Y. & Okamoto, Y. Comparisons of force fields for proteins by generalized-ensemble simulations. *Chem. Phys. Lett.* **386**, 460–467 (2004).
[31] Yoda, T., Sugita, Y. & Okamoto, Y. Secondary-structure preferences of force fields for proteins evaluated by generalized-ensemble simulations. *Chem. Phys.* **307**, 269–283 (2004).
[32] MacKerell, A. Jr., Feig, M. & Brooks, C. L. III. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415 (2004).
[33] Feig, M. private communication.
[34] Sakae, Y. & Okamoto, Y. Optimization of protein force-field parameters with the Protein Data Bank. *Chem. Phys. Lett.* **382**, 626–636 (2003).
[35] Sakae, Y. & Okamoto, Y. Secondary-structure design of proteins by a backbone torsion energy. *J. Phys. Soc. Japan* **75**, 054802 (2006).
[36] Sakae, Y. & Okamoto, Y. Improvement of the backbone-torsion-energy term in the force field for protein systems by the double Fourier series expansion. *Mol. Sim.* **39**, 85–93 (2013).
[37] Sakae, Y. & Okamoto, Y. Amino-acid-dependent main-chain torsion-energy terms for protein systems. *J. Chem. Phys.* **138**, 064103 (2013).
[38] Sakae, Y. & Okamoto, Y. Optimizations of protein force fields. in *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes—from Bioinformatics to Molecular Quantum Mechanics*, 2nd ed. (Liwo, A. ed.) pp. 203–256 (Springer-Verlag, Berlin Heidelberg, 2019).
[39] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092 (1953).
[40] Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).
[41] Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
[42] Berg, B. A. & Neuhaus, T. Multicanonical algorithms for 1st order phase transitions. *Phys. Lett. B* **267**, 249–253 (1991).
[43] Berg, B. A. & Neuhaus, T. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Phys. Rev. Lett.* **68**, 9–12 (1992).
[44] Berg, B. A. *Introduction to Monte Carlo Simulations and Their Statistical Analysis* (World Scientific, Singapore, 2004).
[45] Janke, W. Multicanonical Monte Carlo simulations. *Physica A* **254**, 164–178 (1998).
[46] Lee, J. New Monte Carlo algorithm: Entropic sampling. *Phys. Rev. Lett.* **71**, 211–214 (1993).
[47] Hao, M. H. & Scheraga, H. A. Monte Carlo simulation of a first-order transition for protein folding. *J. Phys. Chem.* **98**, 4940–4948 (1994).
[48] Mezei, M. Adaptive umbrella sampling: self-consistent determination of the non-Boltzmann bias. *J. Comput. Phys.* **68**, 237–248 (1987).
[49] Bartels, C. & Karplus, M. Probability distributions for complex systems: Adaptive umbrella sampling of the potential energy. *J. Phys. Chem. B* **102**, 865–880 (1998).
[50] Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23**, 187–199 (1977).
[51] Wang, F. & Landau, D. P. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **86**, 2050–2053 (2001).
[52] Wang, F. & Landau, D. P. Determining the density of states for classical statistical models: a random walk algorithm to produce a flat histogram. *Phys. Rev. E Stat. Nonlin Soft Matter*

Phys. **64**, 056101 (2001).
[53] Yan, Q., Faller, R. & de Pablo, J. J. Density-of-states Monte Carlo method for simulation of fluids. *J. Chem. Phys.* **116**, 8745–8749 (2002).
[54] Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA* **99**, 12562–12566 (2002).
[55] Okamoto, Y. Molecular simulations in generalised ensemble. *Mol. Sim.* **38**, 1282–1296 (2012).
[56] Hansmann, U. H. E. & Okamoto, Y. Prediction of peptide conformation by multicanonical algorithm: new approach to the multiple-minima problem. *J. Comput. Chem.* **14**, 1333–1338 (1993).
[57] Berg, B. A., Hansmann, U. H. E. & Neuhaus, T. Simulation of an ensemble with varying magnetic field: A numerical determination of the order-order interface tension in the D=2 Ising model. *Phys. Rev. B Condens. Matter* **47**, 497–500 (1993).
[58] Okamoto, Y. & Hansmann, U. H. E. Thermodynamics of helix-coil transitions studied by multicanonical algorithms. *J. Phys. Chem.* **99**, 11276–11287 (1995).
[59] Kolinski, A., Galazka, W. & Skolnick, J. On the origin of the cooperativity of protein folding: implications from model simulations. *Proteins* **26**, 271–287 (1996).
[60] Urakami, N. & Takasu, M. Multicanonical Monte Carlo simulation of a polymer with stickers. *J. Phys. Soc. Japan* **65**, 2694–2699 (1996).
[61] Kumar, S., Payne, P. & Va´squez, M. Method for free-energy calculations using iterative techniques. *J. Comput. Chem.* **17**, 1269–1275 (1996).
[62] Hansmann, U. H. E., Okamoto, Y. & Eisenmenger, F. Molecular dynamics, Langevin and hybrid Monte Carlo simulations in a multicanonical ensemble. *Chem. Phys. Lett.* **259**, 321–330 (1996).
[63] Nakajima, N., Nakamura, H. & Kidera, A. Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides. *J. Phys. Chem. B* **101**, 817–824 (1997).
[64] Bartels, C. & Karplus, M. Multidimensional adaptive umbrella sampling: Applications to main chain and side chain peptide conformations. *J. Comput. Chem.* **18**, 1450–1462 (1997).
[65] Noguchi, H. & Yoshikawa, K. First-order phase transition in a stiff polymer chain. *Chem. Phys. Lett.* **278**, 184–188 (1997).
[66] Higo, J., Nakajima, N., Shirai, H., Kidera, A. & Nakamura, H. Two-component multicanonical Monte Carlo method for effective conformation sampling. *J. Comput. Chem.* **18**, 2086–2092 (1997).
[67] Iba, Y., Chikenji, G. & Kikuchi, M. Simulation of lattice polymers with multi-self- overlap ensemble. *J. Phys. Soc. Japan* **67**, 3327–3330 (1998).
[68] Shimizu, H., Uehara, K., Yamamoto, K. & Hiwatari, Y. Structural phase transition of di-block polyampholyte. *Mol. Sim.* **22**, 285–301 (1999).
[69] Ono, S., Nakajima, N., Higo, J. & Nakamura, H. The multicanonical weighted histogram analysis method for the free-energy landscape along structural transition paths. *Chem. Phys. Lett.* **312**, 247–254 (1999).
[70] Sayano, K., Kono, H., Gromiha, M. M. & Sarai, A. Multicanonical Monte Carlo calculation of the free-energy map of the base-amino acid interaction. *J. Comput. Chem.* **21**, 954–962 (2000).
[71] Yasar, F., Celik, T., Berg, B. A. & Meirovitch, H. Multicanonical procedure for continuum peptide models. *J. Comput. Chem.* **21**, 1251–1261 (2000).
[72] Mitsutake, A., Kinoshita, M., Okamoto, Y. & Hirata, F. Multicanonical algorithm combined with the RISM theory for simulating peptides in aqueous solution. *Chem. Phys. Lett.* **329**, 295–303 (2000).
[73] Sugita, Y. & Okamoto, Y. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chem. Phys. Lett.* **329**, 261–270 (2000).
[74] Mitsutake, A., Sugita, Y. & Okamoto, Y. Replica-exchange multicanonical and multicanonical replica-exchange Monte Carlo simulations of peptides. I. Formulation and benchmark test. *J. Chem. Phys.* **118**, 6664–6675 (2003).
[75] Mitsutake, A., Sugita, Y. & Okamoto, Y. Replica-exchange multicanonical and multicanonical replica-exchange Monte Carlo simulations of peptides. II. Application to a more complex system. *J. Chem. Phys.* **118**, 6676–6688 (2003).
[76] Cheung, M. S., Garcia, A. E. & Onuchic, J. N. Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. USA* **99**, 685–690 (2002).
[77] Kamiya, N., Higo, J. & Nakamura, H. Conformational transition states of a β-hairpin peptide between the ordered and disordered conformations in explicit water. *Protein Sci.* **11**, 2297–2307 (2002).
[78] Jang, S. M., Pak, Y. & Shin, S. M. Multicanonical ensemble with Nosé-Hoover molecular dynamics simulation. *J. Chem. Phys.* **116**, 4782–4786 (2002).
[79] Terada, T., Matsuo, Y. & Kidera, A. A method for evaluating multicanonical potential function without iterative refinement: Application to conformational sampling of a globular protein in water. *J. Chem. Phys.* **118**, 4306–4311 (2003).
[80] Berg, B. A., Noguchi, H. & Okamoto, Y. Multioverlap simulations for transitions between reference configurations. *Phys. Rev. E* **68**, 036126 (2003).
[81] Bachmann, M. & Janke, W. Multicanonical chain-growth algorithm. *Phys. Rev. Lett.* **91**, 208105 (2003).
[82] Okumura, H. & Okamoto, Y. Monte Carlo simulations in multibaric-multithermal ensemble. *Chem. Phys. Lett.* **383**, 391–396 (2004).
[83] Okumura, H. & Okamoto, Y. Monte Carlo simulations in generalized isobaric-isothermal ensembles. *Phys. Rev. E* **70**, 026702 (2004).
[84] Itoh, S. G. & Okamoto, Y. Multi-overlap molecular dynamics methods for biomolecular systems. *Chem. Phys. Lett.* **400**, 308–313 (2004).
[85] Sugita, Y. & Okamoto, Y. Molecular mechanism for stabilizing a short helical peptide studied by generalized-ensemble simulations with explicit solvent. *Biophys. J.* **88**, 3180–3190 (2005).
[86] Itoh, S. G. & Okamoto, Y. Effective sampling in the configurational space of a small peptide by the multicanonical-multioverlap algorithm. *Phys. Rev. E* **76**, 026705 (2007).
[87] Hukushima, K. & Nemoto, K. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Japan* **65**, 1604–1608 (1996).
[88] Geyer, C. J. Markov chain Monte Carlo maximum likelihood. in *Computing Science and Statistics: Proc. 23rd Symp. on the Interface* (Keramidas, E. M. ed.) pp. 156–163 (Interface Foundation, Fairfax Station, 1991).
[89] Marinari, E., Parisi, G. & Ruiz-Lorenzo, J. J. Numerical simulations of spin glass systems. in *Spin Glasses and Random Fields* (Young, A. P. ed.) pp. 59–98 (World Scientific, Singapore, 1997).
[90] Tesi, M. C., van Rensburg, E. J. J., Orlandini, E. & Whittington, S. G. Monte Carlo study of the interacting self-avoiding walk model in three dimensions. *J. Stat. Phys.* **82**, 155–181 (1996).
[91] Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **281**, 140–150 (1997).
[92] Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–

151 (1999).
[93] Wu, M. G. & Deem, M. W. Efficient Monte Carlo methods for cyclic peptides. *Mol. Phys.* **97**, 559–580 (1999).
[94] Sugita, Y., Kitao, A. & Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **113**, 6042–6051 (2000).
[95] Fukunishi, F., Watanabe, O. & Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **116**, 9058–9067 (2002).
[96] Woods, C. J., Essex, J. W. & King, M. A. The development of replica-exchange-based free-energy methods. *J. Phys. Chem. B* **107**, 13703–13710 (2003).
[97] Kokubo, H., Tanaka, T. & Okamoto, Y. Ab initio prediction of protein-ligand binding structures by replica-exchange umbrella sampling simulations. *J. Comput. Chem.* **32**, 2810–2821 (2011).
[98] Kokubo, H., Tanaka, T. & Okamoto, Y. Two-dimensional replica-exchange method for predicting protein-ligand binding structures. *J. Comput. Chem.* **34**, 2601–2614 (2013).
[99] Okamoto, Y., Kokubo, H. & Tanaka, T. Prediction of ligand binding affinity by the combination of replica-exchange method and double-decoupling method. *J. Chem. Theory Comput.* **10**, 3563–3569 (2014).
[100] Gront, D., Kolinski, A. & Skolnick, J. Comparison of three Monte Carlo conformational search strategies for a proteinlike homopolymer model: Folding thermodynamics and identification of low-energy structures. *J. Chem. Phys.* **113**, 5065–5071 (2000).
[101] Verkhivker, G. M., Rejto, P. A., Bouzida, D., Arthurs, S., Colson, A. B., Freer, S. T., *et al.* Parallel simulated tempering dynamics of ligand-protein binding with ensembles of protein conformations. *Chem. Phys. Lett.* **337**, 181–189 (2001).
[102] Sikorski, A. & Romiszowski, P. Thermodynamical properties of simple models of protein-like heteropolymers. *Biopolymers* **69**, 391–398 (2003).
[103] Lin, C. Y., Hu, C. K. & Hansmann, U. H. E. Parallel tempering simulations of HP-36. *Proteins* **52**, 436–445 (2003).
[104] La Penna, G., Mitsutake, A., Masuya, M. & Okamoto, Y. Molecular dynamics of C-peptide of ribonuclease A studied by replica-exchange Monte Carlo method and diffusion theory. *Chem. Phys. Lett.* **380**, 609–619 (2003).
[105] Falcioni, M. & Deem, D. W. A biased Monte Carlo scheme for zeolite structure solution. *J. Chem. Phys.* **110**, 1754–1766 (1999).
[106] Yan, Q. & de Pablo, J. J. Hyper-parallel tempering Monte Carlo: Application to the Lennard-Jones fluid and the restricted primitive model. *J. Chem. Phys.* **111**, 9509–9516 (1999).
[107] Nishikawa, T., Ohtsuka, H., Sugita, Y., Mikami, M. & Okamoto, Y. Replica-exchange Monte Carlo method for Ar fluid. *Prog. Theor. Phys. (Suppl.)* **138**, 270–271 (2000).
[108] Okabe, T., Kawata, M., Okamoto, Y. & Mikami, M. Replica-exchange Monte Carlo method for the isobaric-isothermal ensemble. *Chem. Phys. Lett.* **335**, 435–439 (2001).
[109] Ishikawa, Y., Sugita, Y., Nishikawa, T. & Okamoto, Y. Ab initio replica-exchange Monte Carlo method for cluster studies. *Chem. Phys. Lett.* **333**, 199–206 (2001).
[110] Mori, Y. & Okamoto, Y. Free-energy analyses of a proton transfer reaction by simulated-tempering umbrella sampling first-principles molecular dynamics simulations. *Phys. Rev. E* **87**, 023301 (2013).
[111] Fedorov, D. G., Sugita, Y. & Choi, C. H. Efficient parallel implementations of QM/MM-REMD (quantum mechanical/molecular mechanics-replica-exchange MD) and umbrella sampling: isomerization of H_2O_2 in aqueous solution. *J. Phys. Chem. B* **117**, 7996–8002 (2013).
[112] Ito, S., Irle, S. & Okamoto, Y. Implementation of replica-exchange umbrella sampling in the DFTB+ semiempirical quantum chemistry package. *Comput. Phys. Commun.* **204**, 1–10 (2016).
[113] Ito, S., Fedorov, D. G., Okamoto, Y. & Irle, S. Implementation of replica-exchange umbrella sampling in GAMESS. *Comput. Phys. Commun.* **228**, 152–162 (2018).
[114] Garcia, A. E. & Sanbonmatsu, K. Y. Exploring the energy landscape of a beta hairpin in explicit solvent. *Proteins* **42**, 345–354 (2001).
[115] Zhou, R. H., Berne, B. J. & Germain, R. The free energy landscape for beta hairpin folding in explicit water. *Proc. Natl. Acad. Sci. USA* **98**, 14931–14936 (2001).
[116] Feig, M., MacKerell, A. D. & Brooks, C. L. III. Force field influence on the observation of π-helical protein structures in molecular dynamics simulations. *J. Phys. Chem. B* **107**, 2831–2836 (2003).
[117] Rhee, Y. M. & Pande, V. S. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys. J.* **84**, 775–786 (2003).
[118] Pitera, J. W. & Swope, W. Understanding folding and design: Replica-exchange simulations of "Trp-cage" miniproteins. *Proc. Natl. Acad. Sci. USA* **100**, 7587–7592 (2003).
[119] Ohkubo, Y. Z. & Brooks, C. L. III Exploring Flory's isolated-pair hypothesis: Statistical mechanics of helix-coil transitions in polyalanine and the C-peptide from RNase A. *Proc. Natl. Acad. Sci. USA* **100**, 13916–13921 (2003).
[120] Fenwick, M. K. & Escobedo, F. A. Hybrid Monte Carlo with multidimensional replica exchanges: Conformational equilibria of the hypervariable regions of a llama VHH antibody domain. *Biopolymers* **68**, 160–177 (2003).
[121] Faller, R., Yan, Q. & de Pablo, J. J. Multicanonical parallel tempering. *J. Chem. Phys.* **116**, 5419–5423 (2002).
[122] Felts, A. K., Harano, Y., Gallicchio, E. & Levy, R. M. Free energy surfaces of *β*-hairpin and *α*-helical peptides generated by replica exchange molecular dynamics with the AGBNP implicit solvent model. *Proteins* **56**, 310–321 (2004).
[123] Mitsutake, A., Kinoshita, M., Okamoto, Y. & Hirata, F. Combination of the replica-exchange Monte Carlo method and the reference interaction site model theory for simulating a peptide molecule in aqueous solution. *J. Phys. Chem. B* **108**, 19002–19012 (2004).
[124] Baumketner, A. & Shea, J. E. Free energy landscapes for amyloidogenic tetrapeptides dimerization. *Biophys. J.* **89**, 1493–1503 (2005).
[125] Roitberg, A. E., Okur, A. & Simmerling, C. Coupling of replica exchange simulations to a non-Boltzmann structure reservoir. *J. Phys. Chem. B* **111**, 2415–2418 (2007).
[126] Rosta, E., Buchete, N.-Y. & Hummer, G. Thermostat artifacts in replica exchange molecular dynamics simulations. *J. Chem. Theory Comput.* **5**, 1393–1399 (2009).
[127] Yoda, T., Sugita, Y. & Okamoto, Y. Hydrophobic core formation and dehydration in protein folding studied by generalized-ensemble simulaitons. *Biophys. J.* **99**, 1637–1644 (2010).
[128] De Simone, A. & Derreumaux, P. Low molecular weight oligomers of amyloid peptides display *β*-barrel conformations: A replica exchange molecular dynamics study in explicit solvent. *J. Chem. Phys.* **132**, 165103 (2010).
[129] Whitfield, T. W., Bu, L. & Straub, J. E. Generalized parallel sampling. *Physica A* **305**, 157–171 (2002).
[130] Kwak, W. & Hansmann, U. H. E. Efficient sampling of protein structures by model hopping. *Phys. Rev. Lett.* **95**, 138102 (2005).
[131] Bunker, A. & Dünweg, B. Parallel excluded volume tempering for polymer melts. *Phys. Rev. E Stat. Nonlin Soft Matter Phys.* **63**, 016701 (2000).
[132] Liu, P., Kim, B., Friesner, R. A. & Bern, B. J. Replica

exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. USA* **102**, 13749–13754 (2005).
[133] Affentranger, R., Tavernelli, I. & Di Iorio, E. E. A novel Hamiltonian replica exchange MD protocol to enhance protein conformational space sampling. *J. Chem. Theory Comput.* **2**, 217–228 (2006).
[134] Lou, H. & Cukier, R. I. Molecular dynamics of apoadenylate kinase: A distance replica exchange method for the free energy of conformational fluctuations. *J. Phys. Chem. B* **110**, 24121–24137 (2006).
[135] Mu, Y. Dissociation aided and side chain sampling enhanced Hamiltonian replica exchange. *J. Chem. Phys.* **130**, 164107 (2009).
[136] Itoh, S. G., Okumura, H. & Okamoto, Y. Replica-exchange method in van der Waals radius space: Overcoming steric restrictions for biomolecules. *J. Chem. Phys.* **132**, 134105 (2010).
[137] Mori, Y. & Okamoto, Y. Generalized-ensemble algorithms for the isobaric- isothermal ensemble. *J. Phys. Soc. Japan* **79**, 074003 (2010).
[138] Mitsutake, A. & Okamoto, Y. Multidimensional generalized-ensemble algorithms for complex systems. *J. Chem. Phys.* **130**, 214105 (2009).
[139] Mitsutake, A. Simulated-tempering replica-exchange method for the multidimensional version. *J. Chem. Phys.* **131**, 094105 (2009).
[140] Hiroyasu, T., Miki, M., Ogura, M. & Okamoto, Y. Examination of parallel simulated annealing using genetic crossover. *J. IPS Japan* **43**, 70–79 (2002).
[141] Sakae, Y., Hiroyasu, T., Miki, M. & Okamoto, Y. New conformational search method using genetic algorithm and knot theory for proteins. *Pac. Symp. Biocomput.* **16**, 217–228 (2011).
[142] Sakae, Y., Hiroyasu, T., Miki, M. & Okamoto, Y. Protein structure predictions by parallel simulated annealing molecular dynamics using genetic crossover. *J. Comput. Chem.* **32**, 1353–1360 (2011).
[143] Sakae, Y., Hiroyasu, T., Miki, M., Ishii, K. & Okamoto, Y. A conformational search method for protein systems using genetic crossover and Metropolis criterion. *J. Phys. Conf. Ser.* **487**, 012003 (2014).
[144] Sakae, Y., Hiroyasu, T., Miki, M., Ishii, K. & Okamoto, Y. Conformational search simulations of Trp-cage using genetic crossover. *Mol. Sim.* **41**, 1045–1049 (2015).
[145] Sakae, Y., Straub, J. E. & Okamoto, Y. Enhanced sampling method in molecular simulations using genetic algorithm for biomolecular systems. *J. Comput. Chem.* **40**, 475–481 (2019).
[146] Simons, K. T., Kooperberg, C., Huang, E. & Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. *J. Mol. Biol.* **268**, 209–225 (1997).
[147] Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M. & Baldwin, R. L. Tests of the helix dipole model for stabilization of α-helices. *Nature* **326**, 563–567 (1987).
[148] Shoemaker, K. R., Fairman, R., Schultz, D. A., Robertson, A. D., York, E. J., Stewart, J. M., *et al.* Side-chain interactions in the C-peptide helix: Phe 8 ... His 12^{+}. *Biopolymers* **29**, 1–11 (1990).
[149] Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
[150] Merritt, E. A. & Bacon, D. J. Raster3D: Photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
[151] Wang, J., Cieplak, P. & Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049–1074 (2000).
[152] MacKerell, A. D. Jr., Bashford, D., Bellott, M., Dunbrack, R. L. Jr., 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).
[153] MacKerell, A. D. Jr., Feig, M. & Brooks, C. L. III Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **126**, 698–699 (2004).
[154] MacKerell, A. D. Jr., Feig, M. & Brooks, C. L. III. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum me- chanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415 (2004).
[155] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
[156] Kitao, A., Hirata, F. & Gō, N. The effects of solvent on the conformation and the collective motions of protein: normal mode analysis and molecular-dynamics simulations of melittin in water and in vacuum. *Chem. Phys.* **158**, 447–472 (1991).
[157] Garcia, A. E. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* **68**, 2696–2699 (1992).
[158] Amadei, A., Linssen, A. B. M. & Berendsen, H. J. C. Essential dynamics of proteins. *Proteins* **17**, 412–425 (1993).
[159] Kitao, A. & Gō, N. Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol.* **9**, 164–169 (1999).
[160] Sugita, Y. & Kitao, A. Improved protein free energy calculation by more accurate treatment of nonbonded energy: Application to chymotrypsin inhibitor 2, V57A. *Proteins* **30**, 388–400 (1998).
[161] Kitao, A., Hayward, S. & Gō, N. Energy landscape of a native protein: Jumping-among-minima model. *Proteins* **33**, 496–517 (1998).
[162] Morikami, K., Nakai, T., Kidera, A., Saito, M. & Nakamura, H. Presto (protein engineering simulator): A vectorized molecular mechanics program for biopolymers. *Comput. Chem.* **16**, 243–248 (1992).
[163] Urano, R. & Okamoto, Y. Observation of helix associations for insertion of a retinal molecule and distortions of helix structures in bacteriorhodopsin. *J. Chem. Phys.* **143**, 235101 (2015).
[164] Kokubo, H. & Okamoto, Y. Prediction of transmembrane helix configurations by replica-exchange simulations. *Chem. Phys. Lett.* **383**, 397–402 (2004).
[165] Kokubo, H. & Okamoto, Y. Prediction of membrane protein structures by replica-exchange Monte Carlo simulations: case of two helices. *J. Chem. Phys.* **120**, 10837–10847 (2004).
[166] Kokubo, H. & Okamoto, Y. Classification and prediction of low-energy membrane protein helix configurations by replica-exchange Monte Carlo method. *J. Phys. Soc. Japan.* **73**, 2571–2585 (2004).
[167] Kokubo, H. & Okamoto, Y. Self-assembly of transmembrane helices of bacteriorhodopsin by a replica-exchange Monte Carlo simulation. *Chem. Phys. Lett.* **392**, 168–175 (2004).
[168] Kokubo, H. & Okamoto, Y. Analysis of helix-helix interactions of bacteriorhodopsin by replica-exchange simulations. *Biophys. J.* **96**, 765–776 (2009).
[169] Urano, R., Kokubo, H. & Okamoto, Y. Predictions of tertiary structures of α-helical membrane proteins by replica-exchange method with consideration of helix deformations. *J. Phys. Soc. Japan* **84**, 084802 (2015).
[170] Popot, J. L. & Engelman, D. M. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* **29**,

4031–4037 (1990).
[171] Matsui, Y., Sakai, K., Murakami, M., Shiro, Y., Adachi, S., Okumura, H., *et al.* Specific damage induced by X-ray radiation and structural changes in the primary photoreaction of bacteriorhodopsin. *J. Mol. Biol.* **324**, 469–481 (2002).
[172] Faham, S., Yang, D., Bare, E., Yohannan, S., Whitelegge, J. P. & Bowie, J. U. Side-chain contributions to membrane protein structure and stability. *J. Mol. Biol.* **335**, 297–305 (2004).
[173] Essen, L.-O., Siegert, R., Lehmann, W. D. & Oesterhelt, D. Lipid patches in membrane protein oligomers: crystal structure of the bacteriorhodopsin-lipid complex. *Proc. Natl. Acad. Sci. USA* **95**, 11673–11678 (1998).
[174] Hirokawa, T., Boon-Chieng, S. & Mitaku, S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**, 378–379 (1998).
[175] Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
[176] Jones, D. T., Taylor, W. R. & Thornton, J. M. A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry* **33**, 3038–3049 (1994).
[177] Tusnady, G. & Simon, I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849–850 (2001).
[178] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).
[179] Hu, J., Ma, A. & Dinner, A. R. Monte Carlo simulations of biomolecules: The MC module in CHARMM. *J. Comput. Chem.* **27**, 203–216 (2006).
[180] Lomize, M., Lomize, I. Pogozheva, I. & Mosberg, H. OPM: Orientations of proteins in membranes database. *Bioinformatics* **22**, 623–625 (2006).
[181] Gō, N. & Scheraga, H. A. Ring closure and local conformational deformations of chain molecules. *Macromolecules* **3**, 178–187 (1970).
[182] Dodd, L., Boone, T. & Theodorou, D. A concerted rotation algorithm for atomistic Monte Carlo simulation of polymer melts and glasses. *Mol. Phys.* **78**, 961–996 (1993).
[183] Dinner, A. R. Local deformations of polymers with nonplanar rigid main-chain internal coordinates. *J. Comput. Chem.* **21**, 1132–1144 (2000).
[184] MacQueen, J. Some methods for classification and analysis of multivariate observations. in *Fifth Beerkeley Symposium on Mathematics, Statistics, and Probabilities* (Neyman, U. & LeCam, L. M., eds.) vol. 1, pp. 281–297 (University of California Press, Berkeley and Los Angeles, 1967).
[185] Popot, J. L., Gerchman, S. E. & Engelman, D. M. Refolding of bacteriorhodopsin in lipid bilayers: A thermodynamically controlled two-stage process. *J. Mol. Biol.* **198**, 655–676 (1987).
[186] Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of nalkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
[187] Onufriev, A., Bashford, D. & Case, D. A. Exploring protein native states and large- scale conformational changes with a modified generalized Born model. *Proteins* **55**, 383–394 (2004).