Condensed Matter Physics 2007, Vol. 10, No 4(52), pp. 495–508

CONDENSED MATTER PHYSICS

Combination of generalized-ensemble algorithms and one-dimensional reference interaction site model theory

A.Mitsutake[1], M.Kinoshita[2], F.Hirata[3,4], Y.Okamoto[5]

[1] Department of Physics, Keio University, Yokohama, Kanagawa 223–8522, Japan

[2] Institute of Advanced Energy, Kyoto University, Uji, Kyoto 611–0011, Japan

[3] Department of Theoretical Studies, Institute for Molecular Science, Okazaki, Aichi 444–8585, Japan

[4] Department of Functional Molecular Science, The Graduate University for Advanced Studies, Okazaki, Aichi 444–8585, Japan

[5] Department of Physics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464–8602, Japan

Received October 15, 2007, in final form November 26, 2007

The article reports an attempt to study the protein folding problem by generalized-ensemble Monte Carlo simulations with reference interaction site model theory. Generalized-ensemble algorithms greatly enhance the configurational space sampling in computer simulations. The reference interaction site model theory treats solvent effects with solvent molecular shape and estimate solvation free energy around proteins. We have developed simulation algorithms which combine generalized-ensemble algorithms and one-dimensional reference interaction site model theory. This treatment can also use a simulation with three-dimensional reference interaction site model theory. In this review, we describe the methods and present the results of these simulations for a peptide.

Key words: *generalized-ensemble algorithm, multicanonical algorithm, replica-exchange method, RISM, Monte Carlo, protein folding*

PACS: *31.15.Qg, 61.25.Em, 87.15.Aa*

1. Introduction

In usual simulations, a peptide molecule and quite numerous surrounding solvent molecules are treated simultaneously. However, the number of solvent molecules required becomes huge for a large peptide or a protein. Moreover, in the usual simulations with explicit molecules, the treatment of a mixture of more than one solvent species, such as a saline solution, is nontrivial. The reference interaction site model (RISM) theory [1–4] was developed from the statistical mechanics of molecular liquid and can treat not only an infinite number of solvent molecules but also the salt effects. The theory estimates the solvation free energy of a solute molecule with fixed structures in infinite dilution. Therefore, the RISM theory is particularly suitable for these treatments.

Besides the solvation theory, effective simulation algorithms are also important for studying protein folding and stability. The generalized-ensemble algorithm is one of the most powerful configurational space sampling methods (for a review, see [5]). This method is based on non-Boltzmann probability weight factors so that a random walk in potential energy space may be realized. The random walk allows the simulation to go over any energy barrier and to sample a much wider configurational space than by conventional methods. From the simulation run, one can not only locate the energy global minimum but also calculate the canonical-ensemble average of any physical quantity at any temperature. Two of the most well-known generalized-ensemble algorithms for simulating protein systems are perhaps the multicanonical algorithm (MUCA) [6,7] and the replica-exchange method (REM) [8–11].

In the previous works, we developed an algorithm for solving the full one-dimensional (1D)

RISM equations for the system of a solute molecule with many atoms in water and salt solution that is more robust and orders of magnitude faster than the conventional one [4,12]. We analyzed the solvation structure and conformational stability of a pentapeptide Met-enkephalin with 1D RISM theory [13]. We also reported the results of the attempt to combine Monte Carlo (MC) simulated annealing and 1D RISM theory [14].

Moreover, we tried to combine the generalized-ensemble algorithms and 1D RISM theory to perform simulations of protein system [15,16]. First, we combined 1D RISM theory and multicanonical MC algorithm and performed a simulation of Met-enkephalin in aqueous solution. The effectiveness of the combinational simulation was shown in [15]. In the simulation, however, we only used the solvation free energy at "a room temperature" as the solvent effects because the multicanonical algorithm cannot treat temperature-dependent energy even though the solvent free energy depends on temperature. In order to include the effects of the temperature dependence of the solvation free energy, we then used the replica-exchange MC simulation. In this case, we had to modify the formula for the replica-exchange criterion to treat the temperature-dependent energy and performed the replica-exchange MC simulation with 1 D RISM theory [16]. In the article, we review the above attempt to combine the generalized-ensemble MC algorithms and 1D RISM theory.

2. Method

2.1. Hybrid RISM/MC method

In usual MC simulations, the energy function is given by an instant value of the potential energy for the entire system configuration. In the simulations given in this article, the total energy function E_{T} at a temperature T is given by the sum of the conformational energy E_{C} and the solvation free energy E_{SOL} of the biomolecule in a fixed conformation $\mathbf{r}_{\mathrm{C}}$:

$$E_{\mathrm{T}}(\mathbf{r}_{\mathrm{C}},T) = E_{\mathrm{C}}(\mathbf{r}_{\mathrm{C}}) + E_{\mathrm{SOL}}(\mathbf{r}_{\mathrm{C}},T). \tag{1}$$

Here, the conformational energy E_{C} consists, for example, of electrostatic term, Lennard-Jones term, and torsion-energy term. In the article, we use the RISM theory to calculate E_{SOL}. The total energy function in equation (1) is different from the usual potential energy, because we use the "*solvation free energy*" E_{SOL}. We refer to the total energy function as "*the solvent-induced potential*". The solvent-induced potential depends on temperature because the solvation free energy is a function of temperature. The validity of using the solvent-induced potential as total energy in Monte Carlo simulations was discussed in book [4] and [16]. From the argument, the ensemble average of a physical quantity $A(\mathbf{r}_{\mathrm{C}})$, which depends solely on coordinates $\mathbf{r}_{\mathrm{C}}$, is given by

$$\langle A \rangle_{\mathrm{T}} = \frac{\displaystyle\int \mathrm{d}\mathbf{r}_{\mathrm{C}} A(\mathbf{r}_{\mathrm{C}}) \exp\{-\beta(E_{\mathrm{C}} + E_{\mathrm{SOL}})\}}{\displaystyle\int \mathrm{d}\mathbf{r}_{\mathrm{C}} \exp\{-\beta(E_{\mathrm{C}} + E_{\mathrm{SOL}})\}} \,. \tag{2}$$

The "exact" equation, equation (2), implies that the ensemble average of $A(\mathbf{r}_{\mathrm{C}})$ can be calculated using the integration over the conformation of the biomolecule alone when the energy function in the Boltzmann factor is taken to be the solvent-induced potential $E_{\mathrm{T}} = E_{\mathrm{C}} + E_{\mathrm{SOL}}$ in equation (1).

In the case of the MC simulation at a temperature T with the weight factor $W(\mathbf{r}_{\mathrm{C}},T) = \exp[-\beta\{E_{\mathrm{T}}(\mathbf{r}_{\mathrm{C}},T)\}]$, the canonical transition probability from the state $X = \{\mathbf{r}_{\mathrm{C}}^{[i]}\}$ to $X' = \{\mathbf{r}_{\mathrm{C}}^{[j]}\}$ is given by the Metropolis criterion [17]:

$$w(X \to X') = \begin{cases} 1 \,, & \text{for } \Delta \leqslant 0 \,, \\ \exp(-\Delta) \,, & \text{for } \Delta > 0 \,, \end{cases} \tag{3}$$

where $\Delta = \beta(E_{\mathrm{T}}\left(\mathbf{r}_{\mathrm{C}}^{[j]},T\right) - E_{\mathrm{T}}\left(\mathbf{r}_{\mathrm{C}}^{[i]},T\right))$ and $\beta = 1/k_bT$. After performing the simulation, we can exactly obtain the average of physical quantity $A(\mathbf{r}_\mathrm{C})$ by

$$\langle A\rangle_\mathrm{T} = \frac{\displaystyle\int \mathrm{d}\mathbf{r}_\mathrm{C} A(\mathbf{r}_\mathrm{C}) W(\mathbf{r}_\mathrm{C},T)}{\displaystyle\int \mathrm{d}\mathbf{r}_\mathrm{C} W(\mathbf{r}_\mathrm{C},T)} = \frac{1}{N_\mathrm{SIM}} \sum_{i=1}^{N_\mathrm{SIM}} A_i(\mathbf{r}_\mathrm{C}), \tag{4}$$

where N_SIM is the number of samples taken.

2.2. One-dimensional reference interaction site model

E_SOL is calculated using 1D RISM theory [4,12–14]. It is assumed that the solute is present in the solvent at an infinite-dilution limit. The calculation process is split into two steps where the bulk solvent (step 1) and the solvent near a solute molecule (step 2) are successively treated. The basic equations for step 2 comprise the site-site Ornstein-Zernike (SSOZ) relation and the hypernetted-chain (HNC) closure equation. Let the subscript S and the subscript V denote the solute molecule and the solvent molecule, respectively. The model of the water molecule that we adopted is the SPC/E model [18]. Suppose that the solute molecule has N_S atomic sites and the solvent molecule has N_V atomic sites ($N_\mathrm{V} = 3$ for water). The SSOZ relation in the Fourier space is given by

$$\boldsymbol{\eta}_\mathbf{sv} = \boldsymbol{w}_\mathbf{sv} \boldsymbol{c}_\mathbf{sv} \boldsymbol{H}_\mathbf{vv} - \boldsymbol{c}_\mathbf{sv} , \tag{5}$$

$$\boldsymbol{\eta}_\mathbf{sv} = \boldsymbol{h}_\mathbf{sv} - \boldsymbol{c}_\mathbf{sv} , \tag{6}$$

$$\boldsymbol{H}_\mathbf{vv} = \boldsymbol{w}_\mathbf{vv} + \boldsymbol{\rho}_\mathbf{v} \boldsymbol{h}_\mathbf{vv} , \tag{7}$$

where $\boldsymbol{H}_\mathbf{vv}$, $\boldsymbol{\eta}_\mathbf{sv}$, and $\boldsymbol{w}_{\boldsymbol{ss}}$ are $\boldsymbol{N}_\mathbf{V} \times \boldsymbol{N}_\mathbf{V}$, $\boldsymbol{N}_\mathbf{S} \times \boldsymbol{N}_\mathbf{V}$, and $\boldsymbol{N}_\mathbf{S} \times \boldsymbol{N}_\mathbf{S}$ matrices, respectively, $\boldsymbol{\rho}_\mathbf{v}$ is the matrix of the number density of the solvent, $\boldsymbol{h}$ is the matrix of the site-site intermolecular total correlation functions, $\boldsymbol{c}$ is the matrix of the site-site intermolecular direct correlation functions, and $\boldsymbol{w}$ is the intramolecular correlation matrix. $\boldsymbol{H}_\mathbf{vv}$ is calculated in step 1 and is part of the input data for step 2. The HNC closure equation is given by

$$\begin{aligned} c_{AB}(r) &= \exp\{-u_{AB}(r)/k_BT + \eta_{AB}(r)\} - \eta_{AB}(r) - 1, \\ &\quad A = 1,\ldots,N_\mathrm{S}\,, \quad B = 1,\ldots,N_\mathrm{V}, \end{aligned} \tag{8}$$

$$\eta_{AB}(r) = h_{AB}(r) - c_{AB}(r), \tag{9}$$

where $u_{AB}(r)$ is the site-site interaction potential energy. A is an atomic site in the solute molecule, and B is an atomic site in the solvent molecule. After the site-site intermolecular correlation functions h_{AB}(r) and c$_{AB}$(r) are obtained by solving the RISM-HNC equations, the solvation free energy E_SOL for the solute molecule is calculated from [19]

$$E_\mathrm{SOL}(\mathbf{r}_\mathrm{C},T) = 4\pi k_BT \int_0^\infty F(r) r^2 \mathrm{d}r, \tag{10}$$

$$F(r) = \sum_{A=1}^{N_\mathrm{S}} \sum_{B=\ \mathrm{H,O}} \rho_B \left[\frac{h_{AB}(r)^2}{2} - c_{AB}(r) - \frac{h_{AB}(r) c_{AB}(r)}{2} \right], \tag{11}$$

$$\rho_{\ \mathrm{H}} = 2\rho_\mathrm{V}\,, \quad \rho_\mathrm{O} = \rho_\mathrm{V}\,, \tag{12}$$

where ρ are number densities. The dimensionless number density of water $\rho_\mathrm{V} d^3$ (d=2.8 Å) is 0.7317.

3. Generalized-ensemble algorithms

The generalized-ensemble algorithm is based on non-Boltzmann probability weight factors so that a random walk in potential energy space may be realized. We introduce two generalized-ensemble algorithms, multicanonical algorithm and replica-exchange method, which are commonly used in computer simulations of biomolecules.

3.1. Multicanonical algorithm

We briefly summarize the idea and implementations of the multicanonical algorithm [6,7]. In the multicanonical ensemble the probability distribution of energy $P_{\rm mu}(E)$ is defined in such a way that a configuration with any energy has equal probability: The multicanonical ensemble is based on the following non-Boltzmann weight factor $W_{\rm mu}(E)$ so that the probability distribution of potential energy $P_{\rm mu}(E)$ is flat and a one-dimensional random walk in potential energy space is realized:

$$P_{\rm mu}(E) \propto n(E) W_{\rm mu}(E) = {\rm const}~, \tag{13}$$

where $n(E)$ is the density of states. We then have

$$W_{\rm mu}(E) \propto n^{-1}(E)~. \tag{14}$$

Unlike in the case for the canonical ensemble where the Boltzmann weight factor is used, the multicanonical weight factor $W_{\rm mu}$ is not *a priori* known, and one needs its estimator for a numerical simulation. The estimator of the multicanonical weight factor $W_{\rm mu}$ is usually calculated from iterations of short multicanonical simulations (see, for instance, [21–23]). Once the weight factor $W_{\rm mu}$ has been determined, one performs with this weight factor a multicanonical simulation with high statistics. Monitoring the energy in this simulation, one would see that a random walk between high-energy states and the ground-state configurations is realized. Finally, from this simulation one can not only locate the energy global minimum but also calculate the ensemble average of any physical quantity X for a wide range of temperatures ($T = 1/k_b\beta$) by the reweighing techniques [24]:

$$\langle A \rangle_{\rm T} ~=~ \frac{\displaystyle\int {\rm d}E~ A(E) P_{\rm mu}(E) w_{\rm mu}^{-1}(E) {\rm e}^{-\beta E}}{\displaystyle\int {\rm d}E~ P_{\rm mu}(E) w_{\rm mu}^{-1}(E) {\rm e}^{-\beta E}}~, \tag{15}$$

where $P_{\rm mu}(E)$ is the distribution of energy obtained by the final simulation. We remark that we use $E_{\rm T}(= E_{\rm C} + E_{\rm SOL})$ for E, where $E_{\rm SOL}$ depends on temperature. However, E in the multicanonical simulation should be independent of temperature. Therefore, we fixed the temperature in $E_{\rm SOL}$ at 298.15 K in the present simulation. This means that the expectation value is accurate only in the vicinity of this temperature in the present simulation.

3.2. Modified replica-exchange MC method

In our simulations, we use solvent-induced potential as total potential energy of the system. The solvent-induced potential is a function of temperature because the solvation free energy $E_{\rm SOL}$ depends on temperature. When the potential energy depends on temperature, we need to modify the formula for the regular replica-exchange method. In this subsection, we describe the modified replica-exchange method which can treat the temperature-dependent energy function (see [16,20]).

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

$$\left\{ \begin{array}{rclcl} i &=& i(m) &\equiv& f(m)~, \\ m &=& m(i) &\equiv& f^{-1}(i)~, \end{array} \right. \tag{16}$$

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

Let $X = \{x_1^{[i(1)]}, \ldots, x_M^{[i(M)]}\} = \{x_{m(1)}^{[1]}, \ldots, x_{m(M)}^{[M]}\}$ stand for a "state" in the replica-exchange ensemble. The state X is specified by the M sets of coordinates $x_m^{[i]}$ in replica i at temperature

T_m. In the replica-exchange simulation with the solvent-induced potential, the weight factor for the state $X = \{\cdots, x_m^{[i(m)]}, \cdots\}$ is given by the product of $\exp[-\beta\{E_{\mathrm{T}}(\mathbf{r}_{\mathrm{C}}^{[i]}, T_m)\}]$ for each replica:

$$W_{\mathrm{REM}}(X) = \prod_{i=1}^{M} \exp\left(-\beta_{m(i)} \left\{E_{\mathrm{T}}\left(\mathbf{r}_{\mathrm{C}}^{[i]}; T_{m(i)}\right)\right\}\right), \tag{17}$$

$$= \prod_{m=1}^{M} \exp\left(-\beta_{m} \left\{E_{\mathrm{T}}\left(\mathbf{r}_{\mathrm{C}}^{[i(m)]}; T_{m}\right)\right\}\right), \tag{18}$$

where $i(m)$ and $m(i)$ are the permutation functions in equation (16) and $\mathbf{r}_{\mathrm{C}}^{[i]}$ is a coordinate of a protein in replica i at temperature T_m. Note that the temperature dependence of the total energy E_{T} comes from the solvation free energy E_{SOL} (see equations (8), (10), and (11)).

We now consider exchanging a pair of replicas, i and j, which are at temperatures T_m and T_n, respectively: $X = \{\cdots, x_m^{[i]}, \ldots, x_n^{[j]}, \cdots\} \to X' = \{\cdots, x_m^{[j]}, \ldots, x_n^{[i]}, \cdots\}$. From the condition of detailed balance of replica-exchange, we obtain the following relation:

$$W_{\mathrm{REM}}(X) w(X \to X') = W_{\mathrm{REM}}(X') w(X' \to X), \tag{19}$$

where $w(X \to X')$ is the transition probability of going from state X to state X' and $w(X' \to X)$ is the transition probability of going from state X' to state X. From equations (18) and (19), we have (see also [16,20])

$$\frac{w(X \to X')}{w(X' \to X)} = \exp(-\Delta), \tag{20}$$

where

$$\Delta = \beta_m \left\{E_{\mathrm{T}}\left(\mathbf{r}_{\mathrm{C}}^{[j]}, T_m\right) - E_{\mathrm{T}}\left(\mathbf{r}_{\mathrm{C}}^{[i]}, T_m\right)\right\} - \beta_n \left\{E_{\mathrm{T}}\left(\mathbf{r}_{\mathrm{C}}^{[j]}, T_n\right) - E_{\mathrm{T}}\left(\mathbf{r}_{\mathrm{C}}^{[i]}, T_n\right)\right\}. \tag{21}$$

Here, $\mathbf{r}_{\mathrm{C}}^{[i]}$ and $\mathbf{r}_{\mathrm{C}}^{[j]}$ are the set of the coordinates of atoms in replica i and replica j, respectively. We remark that if E_{T} is independent of temperature, we have

$$\Delta = (\beta_m - \beta_n)\left(E_{\mathrm{T}}\left(\mathbf{r}_{\mathrm{C}}^{[j]}\right) - E_{\mathrm{T}}\left(\mathbf{r}_{\mathrm{C}}^{[i]}\right)\right), \tag{22}$$

which we usually use in the regular replica-exchange simulation.

As a result, the transition probability of the replica-exchange process is given by the Metropolis criterion [17]:

$$w(X \to X') = \begin{cases} 1, & \text{for } \Delta \leqslant 0, \\ \exp(-\Delta), & \text{for } \Delta > 0, \end{cases} \tag{23}$$

where Δ is given by equation (21). Without loss of generality we can assume $T_1 < T_2 < \cdots < T_M$. A REM MC 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 certain MC steps (see equation (3)).

2. A pair of replicas at neighboring temperatures, say $x_m^{[i]}$ and $x_{m+1}^{[j]}$, are exchanged with the probability $w(X \to X')$ in equation (23).

Note that in step 2 we exchange only the pairs of replicas corresponding to the neighboring temperatures.

4. Computational details

The peptide we considered in the present article is Met-enkephalin, and it has the amino-acid sequence Tyr-Gly-Gly-Phe-Met. The peptide has 75 atomic sites, i.e., $N_{\mathrm{S}} = 75$. The 1D RISM-HNC equations and the replica-exchange method have been incorporated in the program KONF90

[25,26]. A hybrid of the Newton-Raphson method and the Picard iteration strategy for solving the RISM equations have been employed as described in detail in [4]. The conformational energy is based on ECEPP/2 [27]–[29]. In ECEPP/2, the conformational energy is given by electrostatic energy $E_{\rm ele}$, hydrogen-bond energy $E_{\rm h}$, Lennard-Jones energy $E_{\rm v}$, and torsion energy $E_{\rm tor}$. The peptide-bond dihedral angles ω were set equal to 180° for simplicity. The remaining dihedral angles ϕ and ψ in the main chain and χ in the side chains constitute the variables to be updated in the simulations (for Met-enkephalin the number of degrees of freedom is 19). One MC sweep consists of updating all these angles once with Metropolis evaluation [17] for each update.

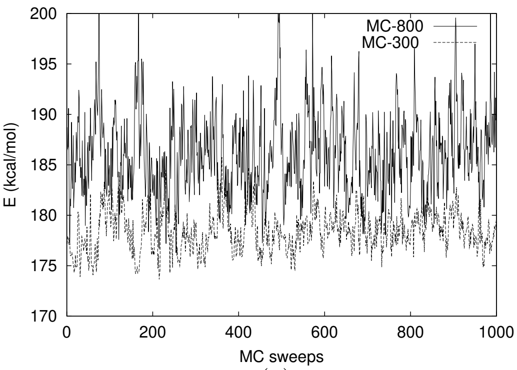
For the calculation of multicanonical weight factors, five iterations of 100 MC sweeps (500 MC sweeps in total) were required. Initial conformations were randomly generated. All thermodynamic quantities were then calculated from one production run of 10000 MC sweeps, which followed additional 100 MC sweeps for equilibration. $E_{\rm SOL}$ are calculated at 298.15.

For the replica-exchange MC simulation, we used four replicas with four temperature values, 200 K, 298 K, 368 K, and 500 K. $E_{\rm SOL}$ was calculated at the corresponding temperatures. Before taking the data, we made equilibration simulations of 5,000 MC sweeps per replica. The total number of MC sweeps for the subsequent REM production run was 20,000 MC sweeps per replica. Replica exchange was tried every 20 MC sweeps.

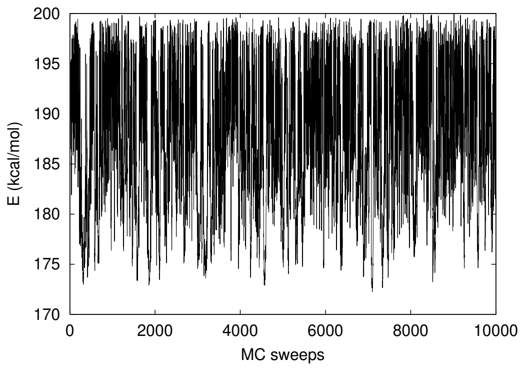
5. Results and discussion

5.1. Results of the multicanonical MC simulation

In figure 1 we show the time series of the solvent-induced potential ($E_{\rm T}$) for Met-enkephalin from regular canonical MC simulations at 300 K and 800 K (we refer to those simulations as MC–300 and MC–800, respectively) and from the multicanonical MC simulation. The time series of MC–300 and MC–800 consist of 1000 MC sweeps, while those of the multicanonical MC simulation consist of 10000 MC sweeps. $E_{\rm SOL}$ were calculated at 298.15 K for MC–300 and MC–800. The solvent-induced potential ranges from about 173 to 185 kcal/mol in MC–300 and from 178 to 220 kcal/mol in MC–800, respectively (see figure 1(a)). On the other hand, the solvent-induced potential range between 172 and 200 kcal/mol is observed (already within 1000 MC sweeps) by the multicanonical MC simulation (see figure 1(b)). It is found that the multicanonical MC simulation covers a much wider configurational space than the canonical simulations. A random walk in energy space covering a range of about 170-200 kcal/mol was successfully realized during 10000 MC sweeps.



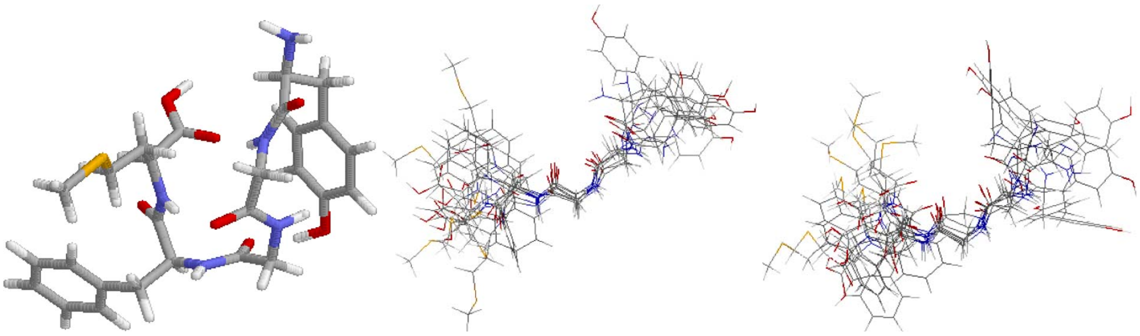
(a)



(b)

Figure 1. Time series of the solvent-induced potential $E_{\rm T}$ (kcal/mol) for: (a) the regular canonical Monte Carlo simulations at 300 K (MC–300) and 800 K (MC–800) (1000 MC sweeps); (b) the multicanonical MC simulation (10000 MC sweeps) [15]. In all simulations, $E_{\rm SOL}$ was calculated at 298.15 K.

We now consider the low-energy conformations of Met-enkephalin obtained during the multicanonical MC simulation in aqueous solution (which we refer to as MURISM). Multicanonical MC simulations in the gas phase (which we refer to as MUGAS) [31,32] and MC simulated annealing runs in aqueous solution based on 1D RISM theory (which we refer to as SARISM) [14] were also performed in previous works. We compare the low-energy conformations obtained by MURISM with those from MUGAS and SARISM. In figure 2 we show the lowest-energy conformations obtained from MUGAS, SARISM, and MURISM. For the latter two cases (figures 2(b) and 2(c)), eight low-energy conformations are superimposed. In figure 2(b) the eight conformations are the lowest-energy conformations from eight different simulated annealing runs [14]. In figure 2(c) the eight representative conformations were chosen from those in which the energy was less than 174.0 kcal/mol and the root-mean-square distances of the backbone atoms from Gly2 to Phe4 were larger than 0.4 Å (this implies that similar structures are excluded as much as possible).



(a) (b) (c)

Figure 2. The lowest-energy conformations of Met-enkephalin obtained by the previous multicanonical simulation in gas phase [32] (a), superposition of the eight lowest-energy conformations from the previous simulated annealing runs in aqueous solution [14] (b), and superposition of the eight representative low-energy conformations of Met-enkephalin obtained by the multicanonical MC simulation in aqueous solution [15] (c).

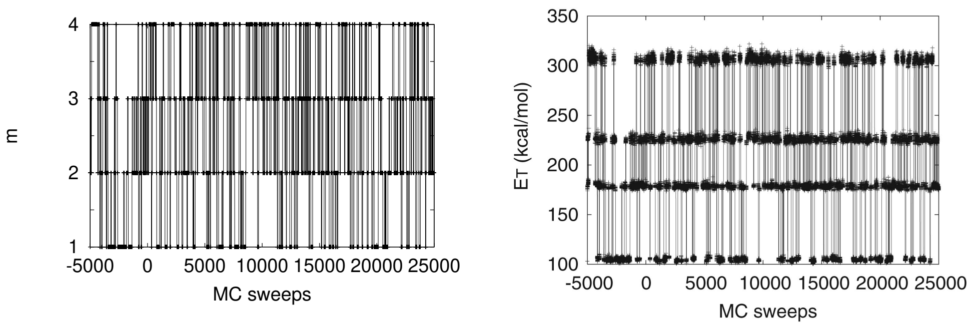
As is shown in figure 2(a), conformations in gas phase form intrachain hydrogen bonds between the amide nitrogen and carbonyl oxygen in the backbone, which make the structure compact. Especially, the lowest-energy conformation has two intrachain hydrogen bonds between Gly2 and Met5 and forms a type II' β turn involving Gly-Gly-Phe-Met (see figure 2(a)) [32]. The end-to-end distance of this conformation is 4.8 Å and the conformation is indeed compact. Here, the end-to-end distance is defined as the distance between the nitrogen atom at the N terminus and the oxygen atom at the C terminus.

On the other hand, for the multicanonical MC simulation in aqueous solution, the eight conformations in figure 2(c) are fully extended. The results are in good agreement with those of the simulated annealing runs (see figure 2(b)). The end-to-end distance of the lowest-energy conformation of MURISM is 11.7 Å, close to that of SARISM (13.6 Å), and the average end-to-end distance at 298.15 K is also about 12 Å. It is clear that the conformations tend to be extended in aqueous solution. Apparently, the amide nitrogen and carbonyl oxygen in the backbone of low-energy conformation in aqueous solution form hydrogen bonds with oxygen and hydrogen of water molecules, respectively. The solvent effect based on 1D RISM theory tends to make hydrogen bonds with backbone atoms and water molecules. The results are in qualitatively good agreement with the observations in the NMR experiments [30].

The solvent-induced potential of the lowest-energy conformation in the multicanonical MC simulation is 172.1 kcal/mol and lower than that obtained by the simulated annealing runs (SARISM) (175.4 kcal/mol) [14]. Moreover, the areas of dihedral angles of backbones are widely sampled in the multicanonical MC simulation (see [16]). The multicanonical MC simulation can thus sample much wider configurational space and provide lower-energy conformations than the simulated annealing simulation.

5.2. Results of the replica-exchange MC simulation

The replica-exchange MC simulation was performed with four replicas at the four temperatures T_m: 200 K, 298 K, 368 K, and 500 K. In figure 3 the time series of the temperature label m and of the solvent-induced potential $E_{\rm T}$ for one of the replicas are shown. We see that the replica takes every temperature value and randomly walks in temperature space as expected (figure 3(a)). Correspondingly, this replica undergoes a random walk in energy space (figure 3(b)). Other replicas also exhibit similar behavior (data not shown). In table 1 we list the acceptance ratios of replica exchange. It is clear that they are of the same order (the values vary between 16 % and 29 %). For an optimal performance of REM simulations, the acceptance ratios of replica exchange should be sufficiently uniform and large (say, > 10 %). We can see that the REM simulation has indeed achieved an optimal performance.



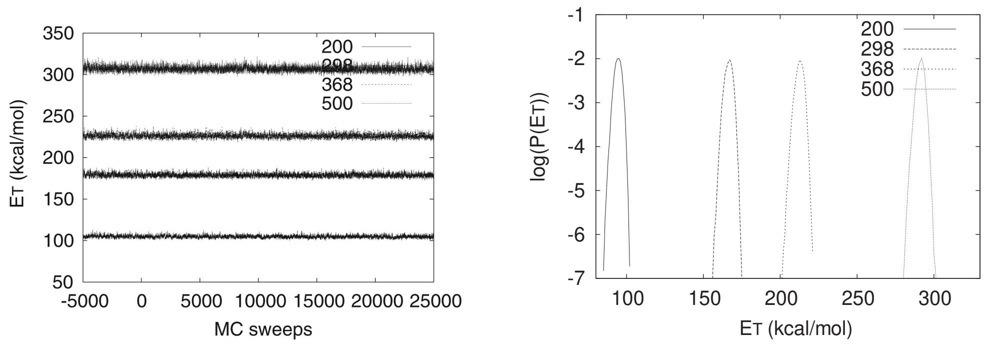
(a) (b)

Figure 3. Time series of the temperature label m (a) and of the solvent-induced potential $E_{\rm T}$ (b) for one of the replicas [16]. The negative values of MC sweeps correspond to the part of equilibration.

Table 1. Acceptance ratios of replica exchange.

Pair of temperature [K]	Acceptance ratio
200 $\longleftrightarrow$ 298	0.16
298 $\longleftrightarrow$ 368	0.29
368 $\longleftrightarrow$ 500	0.25

In figure 4 the time series of the solvent-induced potential and the histogram of the solvent-induced potential distribution at the four temperatures are shown. The difference of solvent-induced potential $E_{\rm T}$ between the neighboring temperatures increases rapidly as the temperature difference increases, because one of its component energies, solvation free energy $E_{\rm SOL}$, depends on temperature (see equation (10)) and is almost proportional to temperature (see [16]). From figure 4(a) and 4(b) we find that there are no overlaps between pairs of neighboring distributions. In the regular replica-exchange method, the histograms of the total potential energy distributions at neighboring temperatures need to overlap for the replica exchange to occur (see equation (22)). Therefore, at first glance, it seems that the number of temperatures chosen above (four) is too small. However, it turned out that the four temperatures were sufficient for ideal replica-exchange performances (see table 1). The point is that we exchange replicas by using Δ in equation (21) instead of Δ of the regular replica-exchange method in equation (22). This can be understood as follows. $E_{\rm T}(\mathbf{r}_{\rm C}^{[i]}, T)$ increases rapidly as temperature T increases and is almost proportional to temperature (see equation (10)). The term $\beta\{E_{\rm T}(\mathbf{r}_{\rm C}^{[j]}, T) - E_{\rm T}(\mathbf{r}_{\rm C}^{[i]}, T)\}$ is therefore almost temperature-independent. The magnitude of Δ in equation (21) is then of the order of one. The acceptance probability $\exp(-\Delta)$ in equation (23) is thus sufficiently large for the replica exchange to occur. Hence, replica exchange is



(a) (b)

Figure 4. Time series of the solvent-induced potential at the four temperature values (200, 298, 368, and 500 K) (a) and probability distributions of the solvent-induced potential at the four temperatures (b) [16].

indeed possible between all pairs of neighboring temperatures. This system requires much less number of replicas than the case for a regular replica-exchange simulation where the energy function is temperature independent.

(a) (b)

Figure 5. The lowest-energy conformations of Met-enkephalin obtained by the replica-exchange simulation in aqueous solution at 200 K (a) and 298 K (b) [16].

We now consider the lowest-energy conformations of Met-enkephalin obtained by the replica-exchange MC simulation in aqueous solution. The lowest-energy conformations obtained at 200 K and 298 K are shown in figure 5(a) and figure 5(b), respectively. These conformations are fully extended as in the results of the multicanonical MC simulation. The results are also in qualitatively good agreement with the observations in the NMR experiments [30].

In table 2, we show the energy components and end-to-end distance for the lowest-energy conformations at 200 K (REP1) and 298 K (REP2) obtained by the replica-exchange simulation. We also list these values obtained by the multicanonical MC simulation in gas phase (MUGAS), simulated annealing runs (SARISM), the multicanonical MC simulation in aqueous solution (MURISM). The solvent-induced potential, conformational energy, and solvation free energy of MUGAS in parentheses were obtained from [14] in which these values of conformation almost identical to the lowest-energy conformation in gas phase (see figure 2(a)) were calculated.

The values of the solvent-induced potential at 200 K (REP1) and 298 K (REP2) are 99.8 kcal/mol and 172.1 kcal/mol, respectively. The values of the conformational energy components of REP1 and REP2 are similar to each other, while the solvation free energy values are different due to different temperatures (see equation (10)). Each energy component of REP2 is almost identical

Table 2. The energy components (kcal/mol), end-to-end distance $d_{\rm e-e}$ (Å) for the lowest-energy conformations obtained by a previous multicanonical MC simulation in gas phase (MUGAS), previous simulated annealing runs (SARISM), the multicanonical MC simulation in aqueous solution (MURISM), and at 200 K (REP1) and 298 K (REP2) from the replica-exchange MC simulation. The values in parentheses were calculated for the similar conformation of the lowest-energy conformation (see figure 2(a)) in gas phase in [14].

	$E_{\rm T}$	$E_{\rm conf}$	$E_{\rm ele}$	$E_{\rm h}$	$E_{\rm v}$	$E_{\rm t}$	$E_{\rm sol}$	$d_{\rm e-e}$
MUGAS	(184.8)	–12.2(–12.0)	20.3	–6.2	–27.5	1.2	(196.8)	4.8
SARISM	175.4	9.6	26.5	–1.0	–16.3	0.5	165.8	13.6
MURISM	172.2	11.6	27.2	–0.9	–15.5	0.9	160.6	11.7
REP1	99.8	12.3	29.2	–1.3	–16.0	0.4	87.5	10.9
REP2	172.1	11.2	28.4	–1.6	–16.2	0.6	160.9	9.2

with that of MURISM, reflecting the fact that the two cases correspond to the same temperature (298.15 K). The lower-energy conformations of SARISM, MURISM, and REP are similar to one another (i.e., extended) and have similar energy components. The conformational energy $E_{\rm C}$ is about 10 kcal/mol and the solvent free energy $E_{\rm SOL}$ is about 160 kcal/mol. On the other hand, $E_{\rm C}$ of the lowest-energy conformation in gas phase is about –12 kcal/mol [32]. $E_{\rm SOL}$ is about 196.8 kcal/mol [14]. $E_{\rm T}$ of the lowest-energy conformation for gas phase is about 184.6 kcal/mol. $E_{\rm C}$ of the lowest-energy conformation in aqueous solution is about 10 kcal/mol and it is much higher than that in the gas phase because energy components of the conformational energy, $E_{\rm ele}$, $E_{\rm h}$, and $E_{\rm v}$, of the simulations with 1D RISM are higher than those in the gas phase. On the contrary, $E_{\rm SOL}$ of SARISM, MURISM, and REP are much lower than that in MUGAS. As a result of the summation of $E_{\rm SOL}$ and $E_{\rm C}$, in aqueous solution, the extended structures obtained by SARISM, MURISM, and REP are more stable than the compact ones by MUGAS.

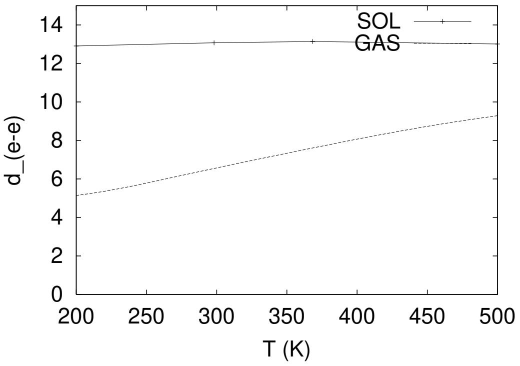


Figure 6. Average end-to-end distance obtained by the replica-exchange MC simulation in aqueous solution (SOL) and the previous multicanonical simulation in gas phase (GAS) [16]. Here, the end-to-end distance is defined as the distance between the nitrogen atom at the N terminus and the oxygen atom at the C terminus.

We calculated some thermodynamic quantities at the given temperatures by equation (4). In figure 6, the average end-to-end distance in aqueous solution as a function of temperature is shown. These average values vary only between 12 Å and 13 Å; the conformations are extended in the entire temperature range. The end-to-end distance of the peptide in gas phase exhibits a coil-globule-like transition with increasing temperature. In aqueous solution, on the other hand, no such transition is observed, since the conformation is always extended due to hydrogen bonds between the peptide and water molecules. The results are similar to the replica-exchange molecular dynamics simulation with explicit water molecules [5]. We also calculated the pair distribution function as a function of temperature to study how the peptide is hydrated. In figure 7 we show the distribution functions

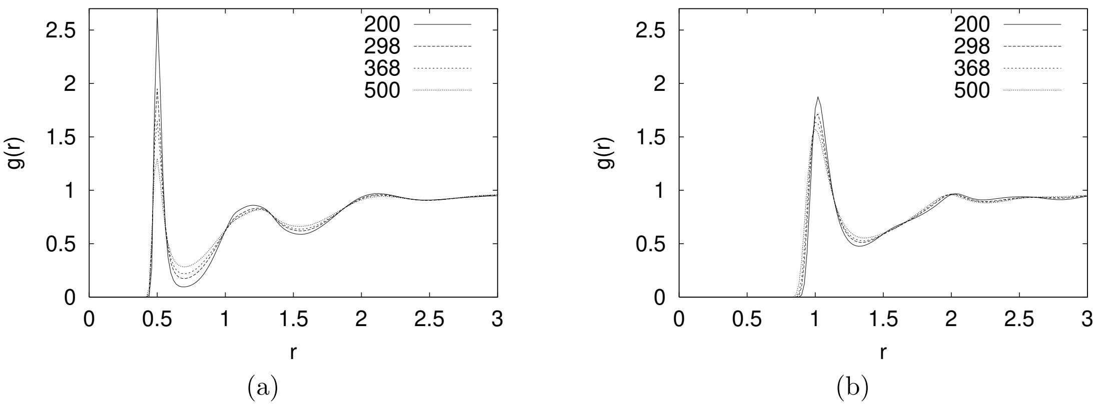


Figure 7. The pair distribution functions between water atoms and backbone atoms averaged over the different conformations of the peptide at each temperature [16]. The pair distribution function is with respect to the oxygen atom in the backbone of the 3rd residue and the hydrogen atom (a) or the oxygen atom (b) in water.

averaged over the different conformations of the peptide. The peaks in figure 7(a) and 7(b) imply that the oxygen in the backbone of Gly 3 makes hydrogen bonds with water molecules. With the increasing temperature, the peaks become low as the hydrogen bonds become weak. The data clearly show that in aqueous solution the backbone atoms of conformations make hydrogen bonds with water molecules. As a result, the conformations are extended.

We now study the relation between the conformational energy (E_{C} and its energy components: electrostatic energy E_{ele}, hydrogen bond energy E_{h}, Lennard-Jones energy E_{v}, or torsion energy E_{tor}) and the solvation free energy E_{SOL}. The results at 298.15 K are shown in figure 8. We also show the relation between E_{SOL} and the sum of E_{ele} and E_{h}. The energy ranges of both abscissa and ordinate are set to the same (25 kcal/mol) to understand the contributions of their energies to the total energy clearly. From figure 8(a), we find a clear inverse correlation between E_{C} and E_{SOL}. The tendency is the same for the results obtained at the given four different temperatures and becomes more pronounced as the temperature is lowered (see [14]). The dotted and solid line in figure 8(a) is the area in which the solvent-induced potential E_{T} is 172.1 kcal/mol (corresponding to the energy of the lowest-energy conformation in aqueous solution) and 184.5 kcal/mol corresponding to that in gas phase, respectively (see table 2). The conformations corresponding to the points between the solid line and the dotted line have E_{T} between 172.1 kcal/mol and 184.5 kcal/mol. Most conformations are in the energy region. In an aqueous solution these conformations are more preferable than the lowest-energy conformation in the gas phase. There are some conformations corresponding to the point near the dotted line. Considering this fact and the inverse correlation between E_{C} and E_{SOL}, it is established that the competition between the conformational energy and the solvation free energy occurs in aqueous solution. From figure 8(a) to figure 8(e), we also see that E_{ele}, E_{h}, and E_{v} which are components of the conformational energy E_{C} have clear inverse correlations between E_{SOL}. It implies that the solvent effect based on 1D RISM theory tends to break the interactions between intra-chain atoms and to make more interactions between peptide atoms and water molecules.

6. Conclusions

We have developed an algorithm combined with the generalized-ensemble algorithms and 1D RISM theory. This is the first attempt of the kind. For this purpose, we reasonably introduced the solvent-induced potential and developed a modified replica-exchange method.

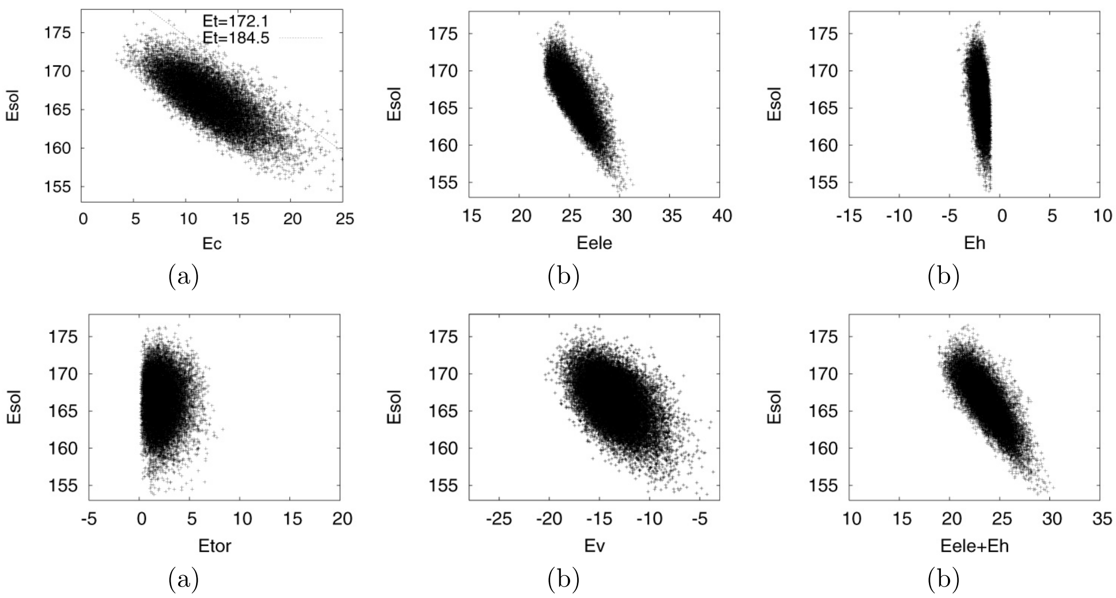


Figure 8. The relations between the solvation free energy E_{SOL} and the total conformational energy E_{C} at 298.15 K (a), and its components, electrostatic energy E_{ele} (b), hydrogen bond energy E_{h} (c), Lennard-Jones energy E_{v} (d), or torsion energy E_{tor} (e), or the sum of electrostatic energy and hydrogen bond energy (f) and solvation free energy.

Firstly, we performed a multicanonical MC simulation of Met-enkephalin in aqueous solution. Secondly, we performed the modified replica-exchange MC simulation of Met-enkephalin in aqueous solution to study temperature dependence. At a room temperature, both simulations have similar results. We obtained a lot of information concerning the solvent effects from these simulations. The conformations in aqueous solution tend to be extended because the backbone atoms make hydrogen bonds to water molecules although the lower-energy conformations in a gas phase tend to make hydrogen bonds between intra-chain backbone atoms. Thermodynamic quantities were also obtained as a function of temperature. The present approach is beneficial for performing protein folding simulations. We shall also apply these modifications to other systems which use a solvent-induced potential such as three-dimensional reference interaction theory.

7. Acknowledgements

The simulations were performed on the computers at the Research Center for Computational Science, Institute for Molecular Science. This work was supported, in part, by the Grants-in-Aid for Scientific Research on Priority Areas "Water and Biomolecules", for Young Scientists (B), 14740170, and for the Next Generation Super Computing Project, Nanoscience Program from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

1. Chandler D., Andersen H.C., J. Chem. Phys., 1972, **57**, 1930–1937.
2. Hirata F., Rossky P.J., Chem. Phys. Lett., 1981, **83**, 329–334.
3. Perkyns J., Pettitt B.M., Chem. Phys. Lett., 1992, **190**, 626–630.
4. Kinoshita M. – In Molecular Theory of Solvation, edited by Hirata, F. (Kluwer Academic, Dordrecht, 2003), pp.101–168.
5. Mitsutake A., Sugita Y., Okamoto Y., Biopolymers (Peptide Science), 2001, **60**, 96–123.
6. Berg B.A., Neuhaus T., Phys. Lett. B, 1991, **267**, 249–253.

7. Berg B.A., Neuhaus T., Phys. Rev. Lett., 1992, **68**, 9–12.
8. Hukushima K., Nemoto K., J. Phys. Soc. Jpn., 1996, **65**, 1604–1608.
9. Hukushima K., Takayama H., Nemoto K., Int. J. Mod. Phys., 1996, **C7**, 337–344.
10. Geyer C.J. – In Computing Science and Statistics: Proc. 23rd Symp. on the Interface, edited by Keramidas, E.M. (Interface Foundation, Fairfax Station, 1991), pp. 156–163.
11. Sugita Y., Okamoto Y., Chem. Phys. Lett., 1999, **314**, 141–151.
12. Kinoshita M., Okamoto Y., Hirata F., J. Comput. Chem., 1997, **18**, 1320–1326; Kinoshita M., Okamoto Y., Hirata F., J. Comput. Chem., 1998, **19**, 1724–1735.
13. Kinoshita M., Okamoto Y., Hirata F., J. Chem. Phys., 1997, **107**, 1586–1599.
14. Kinoshita M., Okamoto Y., Hirata F., J. Am. Chem. Soc.,1998, **120**, 1855–1863.
15. Mitsutake A., Kinoshita M., Okamoto Y., Hirata F., Chem. Phys. Lett., 2000, **329**, 295–303.
16. Mitsutake A., Kinoshita M., Okamoto Y., Hirata F., J. Phys. Chem. B, 2004, **108**, 19002–19012.
17. Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E., J. Chem. Phys., 1953, **21**, 1087–1092.
18. Berendsen H.J.C., Grigera J.R., Straatsma T.P.J., J. Phys. Chem., 1987, **91**, 6269–6271.
19. Singer S.J., Chandler D., Mol. Phys., 1985, **55**, 621–625.
20. La Penna G., Mitsutake A., Masuya M., Okamoto Y., Chem. Phys. Lett., 2003, **380**, 609–619.
21. Okamoto Y., Hansmann U.H.E., J. Phys. Chem., 1995, **99**, 11276–11287.
22. Berg B.A., Int. J. Mod. Phys. C, 1992, **3**, 1083.
23. Berg B.A., Nucl. Phys. B (Proc. Suppl.), 1998, **63A–C**, 982–984.
24. Ferrenberg A.M., Swendsen R.H., Phys. Rev. Lett., 1988, **61**, 2635–2638.
25. Kawai H., Okamoto Y., Fukugita M., Nakazawa T., Kikuchi T., Chem. Lett., 1991, **1991**, 213–216.
26. Okamoto Y., Fukugita M., Nakazawa T., Kawai H., Protein Eng., 1991, **4**, 639–647.
27. Momany F.A., McGuire R.F., Burgess A.W., Scheraga H.A., J. Phys. Chem., 1975, **79**, 2361–2381.
28. Némethy G., Pottle M.S., Scheraga H.A., J. Phys. Chem., 1983, **87**, 1883–1887.
29. Sippl M.J., Némethy G., Scheraga H.A., J. Phys. Chem., 1984, **88**, 6231–6233.
30. Graham W.H., Carter II S.E., Hickes P.R., Biopolymers, 1992, **32**, 1755–1764.
31. Hansmann U.H.E., Okamoto Y., J. Comput. Chem., 1993, **14**, 1333–1338.
32. Mitsutake A., Hansmann U.H.E., Okamoto Y., J. Mol. Graph. Mod., 1998, **16**, 226–238.

Комбінація алгоритмів узагальненого ансамблю і одновимірної базисної моделі взаємодіючих силових центрів

А.Міцутаке[1], М.Кіношіта[2], Ф.Хірата[3,4], Ю.Окамото[5]

[1] Фізичний факультет, університет м. Кейо, Йокагама, Канагава, Японія
[2] Інститут прикладної енергетики, університет м. Кіото, Японія
[3] Факультет теоретичних досліджень, Інститут молекулярних наук м. Оказакі, Аічі, Японія
[4] Факультет функціональних молекулярних наук, Університет м. Оказакі, Аічі, Японія
[5] Фізичний факультет, університет м. Нагоя, Аічі, Японія

Отримано 15 жовтня 2007 р., в остаточному вигляді – 26 листопада 2007 р.

Стаття присвячена вивченню проблеми згортання протеїну в рамках комбінації узагальненого ансамблю Монте-Карло та теорії базисної моделі силових центрів. Алгоритми узагальненого ансамблю можуть ефективно вибирати конфігураційний простір в комп'ютерних моделюваннях. Теорія базисних взаємодіючих силових центрів може враховувати ефекти розчинника в залежності від форми молекул, визначаючи енергію сольватації протеїнів. Ми розробили комп'ютерні алгоритми, що комбінують алгоритми узагальненого ансамблю і одновимірну базисну модель взаємодіючих силових центрів. Цей підхід може бути також поєднаний з тривимірною моделлю взаємодіючих силових центрів. В даному огляді описуються методи і представлено результати для пептиду.

Ключові слова: *алгоритм узагальненого ансамблю, багатоканонічний алгоритм, метод реплічного обміну, RISM, Монте Карло, згортання протеїну*

PACS: *31.15.Qg, 61.25.Em, 87.15.Aa*