

# 拡張アンサンブル法による膜タンパク質の 立体構造予測

小久保 裕功<sup>1</sup>・岡本 祐幸<sup>2</sup>

(受付 2006年3月23日;改訂 2006年7月10日)

## 要 旨

統計力学的アンサンブルを再現する計算機シミュレーション法にマルコフ連鎖モンテカルロ法があるが、この手法は普通ボルツマン因子に基づいて状態を発生させ、温度一定のカノニカルアンサンブルを実現する。しかし、系の自由度の数が大きいとき、エネルギー極小状態が無数に存在するとともに、それらの間に高いエネルギー障壁が存在するため、シミュレーションがそれらエネルギー極小状態に留まってしまっており、低温における正確なカノニカルアンサンブルを再現するのは困難である。この困難を克服するために開発されたのが、拡張アンサンブル法と総称されるシミュレーション手法である。拡張アンサンブル法は非ボルツマン因子に基づいて、エネルギー空間上の1次元酔歩を実現することによって、エネルギー極小状態に留まるのを避ける。ここでは、拡張アンサンブル法の一つである、レプリカ交換法とその一般化版について解説する。そして、具体例として、レプリカ交換モンテカルロ法を膜タンパク質の立体構造予測問題に適用した結果を紹介する。

キーワード: 拡張アンサンブル法, レプリカ交換法, 膜タンパク質, 膜貫通ヘリックス, タンパク質の立体構造予測。

## 1. はじめに

ある多変数関数の最小値を求めることは多くの分野で必要とされることである。その解を予測するために広く使われる手法としては、いくつかの変数値における関数値を求める(観測すること)による、または、解が既知の類似関数を求めることによる、帰納的なアプローチがある。一方、計算機シミュレーションなどによって、演繹的に解を求めようとするのが、本稿のアプローチである。

多自由度の複雑系では、エネルギー極小状態が無数に存在し、それらが高いエネルギー障壁で隔てられているため、従来の温度一定のカノニカルアンサンブル上のマルコフ連鎖モンテカルロ法(MC)などによる計算機シミュレーションでは、それらのエネルギー極小状態に留まってしまっており、定量的で正確な予測を行うのが絶望的に難しくなってしまうという難点がある。

このような計算機シミュレーションの困難を克服する有効な方法として、拡張アンサンブル法(generalized-ensemble algorithm)と総称される手法が存在する(筆者による解説としては、例えば、Mitsutake et al., 2001; 岡本, 2002などを参照)。この方法は、非ボルツマン因子による

<sup>1</sup> Department of Chemistry, University of Houston, Houston, TX 77204, U.S.A.

<sup>2</sup> 名古屋大学 大学院理学研究科: 〒464-8602 愛知県名古屋市千種区不老町

人工の統計集団に基づいており、ポテンシャルエネルギー空間上の一次元酔歩を実現することで、エネルギー極小状態に留まるのを避ける。そして、従来の手法と比べて格段に幅広い状態空間を探索することを可能とするのである。

本稿では、特に広く使われている拡張アンサンブル法の一つである、レプリカ交換法(replica-exchange method) [モンテカルロ法版については、Hukushima and Nemoto, 1996; 分子動力学法版については、Sugita and Okamoto, 1999; この手法の歴史及び文献の詳細については、Iba, 2001; 伊庭 他, 2005 を参照]とその改良版について解説する。そして、具体例として、レプリカ交換モンテカルロ法の膜タンパク質の立体構造予測シミュレーションへの適用結果を紹介する。

## 2. シミュレーション手法

ここでは、レプリカ交換法の詳しい説明を述べる。

### 2.1 レプリカ交換法

生体系を表す系として、質量が  $m_k$  ( $k=1, \dots, N$ ) の  $N$  個の原子の系を考え、それらの座標と運動量がそれぞれ  $q \equiv \{q_1, \dots, q_N\}$  と  $p \equiv \{p_1, \dots, p_N\}$  で表されるとする。系のハミルトニアン  $H(q, p)$  は運動エネルギー  $K(p)$  とポテンシャルエネルギー  $E(q)$  の和で次のように与えられる。

$$(2.1) \quad H(q, p) = K(p) + E(q).$$

ここで、運動エネルギーは次式のとおりである。

$$(2.2) \quad K(p) = \sum_{k=1}^N \frac{p_k^2}{2m_k}.$$

温度一定のカノニカルアンサンブルでは、各状態  $x \equiv (q, p)$  は次のボルツマン因子の確率重みで実現される。

$$(2.3) \quad W_B(x; T) = e^{-\beta H(q, p)}.$$

ここで、 $\beta$  はボルツマン定数  $k_B$  と温度  $T$  の積の逆数である ( $= 1/k_B T$ )。このとき、温度  $T$  における平均運動エネルギーは次で与えられる。

$$(2.4) \quad \langle K(p) \rangle_T = \left\langle \sum_{k=1}^N \frac{p_k^2}{2m_k} \right\rangle_T = \frac{3}{2} N k_B T.$$

式(2.1)では座標  $q$  と運動量  $p$  が分離されているので、運動エネルギーについては分配関数中でガウス積分を実行できて、考慮からはずすことができる。よって、ボルツマン因子は次のように書くことができる。

$$(2.5) \quad W_B(x; T) = W_B(E; T) = e^{-\beta E}.$$

すると、温度一定のカノニカルアンサンブルでは、ポテンシャルエネルギー  $E$  の分布  $P_B(E; T)$  は状態密度  $n(E)$  とボルツマン因子  $W_B(E; T)$  の積で以下のように与えられる。

$$(2.6) \quad P_B(E; T) \propto n(E) W_B(E; T).$$

状態密度はエネルギーと共に急速に増加する関数であり、ボルツマン因子は指数関数的に減少する関数であるから、カノニカル分布は一般にベル型をしている。

この温度  $T$  におけるカノニカル分布を再現する計算機シミュレーション法としては、メトロポリスのモンテカルロ法が広く使われている。この手法では、マルコフ過程に基づきながら、

以下のメトロポリス判定に従って状態を発生させて行く。すなわち、ポテンシャルエネルギー  $E$  をもつ状態  $x$  から  $E'$  をもつ状態  $x'$  への遷移確率は次で与えられる。

$$(2.7) \quad w(x \rightarrow x') = \min\left(1, \frac{W_B(x')}{W_B(x)}\right) = \min(1, \exp(-\beta\Delta E)).$$

ここで、 $\Delta E$  は次で定義される。

$$(2.8) \quad \Delta E \equiv E(x') - E(x) = E' - E.$$

実際の計算では、状態  $x$  から次の状態の候補  $x'$  を発生させ、 $\Delta E$  を計算する。もし、 $\Delta E \leq 0$  (ポテンシャルエネルギーが下がる) なら、この候補  $x'$  を採択する。逆に、 $\Delta E > 0$  (ポテンシャルエネルギーが上がる) なら、0 から 1 の間の値をとる一様乱数  $r$  を発生させ、 $r \leq \exp(-\beta\Delta E)$  なら、候補  $x'$  を採択するという手順を踏めばよいことになる。

このメトロポリス法は温度一定のカノニカル分布を再現する手法であるが、生体系のように自由度の数が大きく、エネルギー極小状態が無数に存在する系の場合、熱的揺らぎが小さい低温ではシミュレーションがそれらに留まってしまうため、平衡に達するまで膨大な計算時間を必要とする。

この困難を克服する方法として拡張アンサンブル法が考えられたが、まず、レプリカ交換法を説明する。

レプリカ交換法における系は、 $M$  個の違う温度  $T_m$  ( $m=1, \dots, M$ ) をもつ、互いに相互作用しない  $M$  個の独立なコピー(レプリカ)から成る。それぞれのレプリカには常に一つの温度が対応するとする。よって、レプリカと温度には 1 対 1 対応が成り立ち、レプリカを表すラベルを  $i$  ( $i=1, \dots, M$ )、温度を表すラベルを  $m$  ( $m=1, \dots, M$ ) とするとき、 $i$  と  $m$  は置換関係にある。

$$(2.9) \quad \begin{cases} i = i(m) \equiv f(m), \\ m = m(i) \equiv f^{-1}(i). \end{cases}$$

ここで、 $f(m)$  は  $m$  の置換であり、 $f^{-1}(i)$  はその逆置換である。

この系における「状態」を  $X = \{x_1^{[i(1)]}, \dots, x_M^{[i(M)]}\} = \{x_{m(1)}^{[1]}, \dots, x_{m(M)}^{[M]}\}$  で表すことにしよう。すなわち、系の状態  $X$  は各レプリカの状態

$$(2.10) \quad x_m^{[i]} \equiv (q^{[i]})_m$$

により指定される。そして、温度  $T_m$  にあるレプリカ  $i$  の状態  $x_m^{[i]}$  は  $N$  個の原子の座標  $q^{[i]}$  により決まる。

レプリカ同士は相互作用しないので、状態  $X$  の重み因子はボルツマン因子の積で以下のように与えられる。

$$(2.11) \quad W_{\text{REM}}(X) = \exp\left\{-\sum_{i=1}^M \beta_{m(i)} E(q^{[i]})\right\} = \exp\left\{-\sum_{m=1}^M \beta_m E(q^{[i(m)]})\right\}.$$

ここで、 $i(m)$  と  $m(i)$  は式(2.9)の置換関数である。

温度がそれぞれ  $T_m$  と  $T_n$  に対応するレプリカ  $i$  と  $j$  を交換することを考える。

$$(2.12) \quad X = \{\dots, x_m^{[i]}, \dots, x_n^{[j]}, \dots\} \longrightarrow X' = \{\dots, x_m^{[j]}, \dots, x_n^{[i]}, \dots\}.$$

ここで、 $i, j, m, n$  は式(2.9)の置換関数による関係を持つ。よって、レプリカの交換は以下のような新しい置換関数  $f'$  を導入することになる。

$$(2.13) \quad \begin{cases} i = f(m) \longrightarrow j = f'(m), \\ j = f(n) \longrightarrow i = f'(n). \end{cases}$$

このレプリカの交換は次のように書くことができる。

$$(2.14) \quad \begin{cases} x_m^{[i]} \equiv (q^{[i]})_m \longrightarrow x_m^{[j]} \equiv (q^{[j]})_m, \\ x_n^{[j]} \equiv (q^{[j]})_n \longrightarrow x_n^{[i]} \equiv (q^{[i]})_n. \end{cases}$$

ちなみに、この操作はレプリカがそれぞれ  $i$  と  $j$  に対応する温度  $T_m$  と  $T_n$  を交換することと同等である。よって、式(2.14)は次のように書いても良い。

$$(2.15) \quad \begin{cases} x_m^{[i]} \equiv (q^{[i]})_m \longrightarrow x_n^{[i]} \equiv (q^{[i]})_n, \\ x_n^{[j]} \equiv (q^{[j]})_n \longrightarrow x_m^{[j]} \equiv (q^{[j]})_m. \end{cases}$$

さて、このレプリカ対の交換の操作は、遷移確率  $w(X \rightarrow X')$  に次の詳細釣り合いの条件を課すことによって、平衡状態への収束が保証される。

$$(2.16) \quad \frac{W_{\text{REM}}(X)}{Z} w(X \rightarrow X') = \frac{W_{\text{REM}}(X')}{Z} w(X' \rightarrow X),$$

ここで、 $Z$  は系全体の分配関数である。よって、式(2.11)と式(2.16)から以下が得られる。

$$(2.17) \quad \begin{aligned} \frac{W_{\text{REM}}(X')}{W_{\text{REM}}(X)} &= \exp\{-\beta_m E(q^{[j]}) - \beta_n E(q^{[i]}) + \beta_m E(q^{[i]}) + \beta_n E(q^{[j]})\} \\ &= \exp\{-\beta_m [E(q^{[j]}) - E(q^{[i]})] - \beta_n [E(q^{[i]}) - E(q^{[j]})]\} \\ &= \exp(-\Delta). \end{aligned}$$

ここで、 $\Delta$  は次で与えられる。

$$(2.18) \quad \Delta = \beta_m (E(q^{[j]}) - E(q^{[i]})) - \beta_n (E(q^{[i]}) - E(q^{[j]}))$$

$$(2.19) \quad = (\beta_m - \beta_n) (E(q^{[j]}) - E(q^{[i]})).$$

ここで、ラベル  $i, j, m, n$  は交換前の置換関数(式(2.9))で次のように関係している。

$$(2.20) \quad \begin{cases} i = f(m), \\ j = f(n). \end{cases}$$

よって、レプリカ対の交換確率は、次のメトロポリスの判定条件で与えられる。

$$(2.21) \quad w(X \rightarrow X') \equiv w(x_m^{[i]} | x_n^{[j]}) = \min\left(1, \frac{W_{\text{REM}}(X')}{W_{\text{REM}}(X)}\right) = \min(1, \exp(-\Delta)).$$

ここで、2つ目の表現  $w(x_m^{[i]} | x_n^{[j]})$  では、交換されるレプリカ対(と温度対)をあらわに書いた。

レプリカ交換シミュレーションは、次の2つのステップを交互に繰り返すことで実行される( $T_1 < T_2 < \dots < T_M$  とする)。

- (1) 各レプリカ  $i$  ( $i = 1, \dots, M$ )において、一定温度  $T_m$  ( $m = 1, \dots, M$ )の独立なカノニカルアンサンブル上の従来の MC シミュレーションを同時にあるステップ数実行する。
- (2) 隣接した温度( $T_m$  と  $T_{m+1}$ )に対応するレプリカ対( $i$  と  $j$ )を式(2.21)の遷移確率  $w(x_m^{[i]} | x_{m+1}^{[j]})$  に従って交換する。

ステップ2では隣り合う温度値に対応するレプリカ対を交換することに注意されたい。これは、レプリカ交換の受け入れ確率が温度差に従って指数関数的に減少するからである(式(2.19)と式(2.21)を参照)。また、ステップ2において、レプリカ交換が採択されるたびに、式(2.9)の置換関数が更新されることになる。

レプリカ交換シミュレーションは並列計算機に特に適している。各レプリカを各計算機(CPU)に対応させることによって、計算機間の情報通信量を最小にできる訳である(座標の代わりに温度を交換するだけなので)。これは、シミュレーションの間に、MCステップ $t$ の関数として置換関数 $m(i;t) = f^{-1}(i;t)$ を記録していくことになる。

長いレプリカ交換シミュレーションの後、物理量 $A$ の任意の温度 $T = 1/k_B\beta$ におけるカノニカルアンサンブル平均は、次式で与えられる。

$$(2.22) \quad \langle A \rangle_T = \frac{\sum_E A(E) P_B(E;T)}{\sum_E P_B(E;T)} = \frac{\sum_E A(E) n(E) e^{-\beta E}}{\sum_E n(E) e^{-\beta E}}.$$

ここで、状態密度 $n(E)$ の最適解は次の2つの方程式を逐次的に解くことによって得られる(多ヒストグラム再重法(multiple-histogram reweighting techniques) Ferrenberg and Swendsen, 1989)。

$$(2.23) \quad n(E) = \frac{\sum_{m=1}^M N_m(E)}{\sum_{m=1}^M n_m e^{f_m - \beta_m E}},$$

$$(2.24) \quad e^{-f_m} = \sum_E n(E) e^{-\beta_m E}.$$

ここで、式中のヒストグラム $N_m(E)$ は、レプリカごとではなく、温度 $T_m = 1/k_B\beta_m$  ( $m = 1, \dots, M$ )ごとに得られたポテンシャルエネルギー分布のヒストグラムであり、 $n_m$ はその全サンプル数である。また、 $f_m$ は無次元化されたヘルムホルツ自由エネルギーである。上の方程式は例えば次のように逐次的に解かれる。まず、全ての $f_m$  ( $m = 1, \dots, M$ )を0にセットする。次に式(2.23)を使って、 $n(E)$ を求める。得られた $n(E)$ を式(2.24)に代入して、新しい $f_m$ の値を得る。この操作を $n(E)$ と $f_m$ の値が収束するまで繰り返す。

更に、ポテンシャルエネルギーの関数でない場合を含む一般の物理量 $A$ のカノニカルアンサンブル平均は、次式で与えられる。すなわち、まず、上の式(2.23)と(2.24)を解いて、 $f_m$  ( $m = 1, \dots, M$ )を求める。そして、それを次の式に代入することによって、平均値が得られる(Mitsutake et al., 2003)。

$$(2.25) \quad \langle A \rangle_T = \frac{\sum_{m=1}^M \sum_{x_m} A(x_m) \frac{\exp(-\beta E(x_m))}{\sum_{\ell=1}^M n_\ell \exp(f_\ell - \beta_\ell E(x_m))}}{\sum_{m=1}^M \sum_{x_m} \frac{\exp(-\beta E(x_m))}{\sum_{\ell=1}^M n_\ell \exp(f_\ell - \beta_\ell E(x_m))}}.$$

ここで、 $x_m$ についての和は温度 $T_m$ に対応するシミュレーションの途中の(トラジェクトリーの)状態についての和を意味する。

## 2.2 多次元レプリカ交換法

ここで、レプリカ交換法の多次元(多変数)への拡張版である、多次元レプリカ交換法(multi-dimensional replica-exchange method; Sugita et al., 2000)を紹介しよう。この手法の開発に至った重要なアイデアは、従来のレプリカ交換法において、 $M$ 個のレプリカが相互作用しない限り、各レプリカのハミルトニアン $H(q,p)$ は同一である必要がないということである。例えば、ハミルトニアンはあるパラメータに依存し、それぞれのレプリカが違うパラメータ値をとって

いても良いことになる。このとき、温度  $T_m$  に対応する  $i$  番目のレプリカのハミルトニアンは次のように書くことができる。

$$(2.26) \quad H_m(q^{[i]}, p^{[i]}) = K(p^{[i]}) + E_{\lambda_m}(q^{[i]}).$$

すなわち、ポテンシャルエネルギー  $E_{\lambda_m}$  はパラメーター  $\lambda_m$  に依存しているとする。

従来のレプリカ交換法では、レプリカ  $i$  と温度  $T_m$  が 1 対 1 対応しているのに対し、ここでは、レプリカ  $i$  と「パラメーターセット」 $\Lambda_m \equiv (T_m, \lambda_m)$  が 1 対 1 対応している。従来のレプリカ交換法のパラメーター空間が 1 次元 ( $\Lambda_m = T_m$ ) であるので、この手法はその多次元への拡張版ということが出来る。これによって、温度空間上ばかりでなく、様々なパラメーター空間上の酔歩が実行されて、より効率的な状態空間のサンプリングが実現されると期待される。さて、レプリカ同士は相互作用しないので、状態  $X$  の重み因子は各レプリカに対応するボルツマン因子の積で以下のように与えられる(式(2.11)を参照)。

$$(2.27) \quad W_{\text{MREM}}(X) = \exp \left\{ - \sum_{i=1}^M \beta_{m(i)} E_{\lambda_{m(i)}}(q^{[i]}) \right\} = \exp \left\{ - \sum_{m=1}^M \beta_m E_{\lambda_m}(q^{[i(m)]}) \right\}.$$

ここで、 $i(m)$  と  $m(i)$  は式(2.9)の置換関数である。よって、従来のレプリカ交換の採択判定式を導いたのと同じことをすることによって、レプリカ交換法の遷移確率は式(2.21)で与えられることが分かる。ただし、ここでは  $\Delta$  は以下ようになる(式(2.18)参照)。

$$(2.28) \quad \Delta = \beta_m (E_{\lambda_m}(q^{[j]}) - E_{\lambda_m}(q^{[i]})) - \beta_n (E_{\lambda_n}(q^{[j]}) - E_{\lambda_n}(q^{[i]})).$$

一般に、全ポテンシャルエネルギー  $E_{\lambda_m}$  と  $E_{\lambda_n}$  は違う関数なので、新たに、交換後の  $E_{\lambda_m}(q^{[j]})$  と  $E_{\lambda_n}(q^{[i]})$  を評価する必要があることに注意されたい。

### 3. レプリカ交換法による膜タンパク質の立体構造予測

ここでは、レプリカ交換モンテカルロ法の適用例として、膜タンパク質の立体構造予測の結果を紹介する(Kokubo and Okamoto, 2004a, b, c, d)。

具体的には(SOSUI, Hirokawa et al., 1998 などの WWW サーバーを使って)膜貫通ヘリックスの部分だけを取り出し、その膜貫通ヘリックスだけのレプリカ交換モンテカルロシミュレーションを実行した。すなわち、膜貫通ヘリックスの主鎖の部分は固定して、ヘリックスの平行移動と回転及び側鎖の自由度についてのモンテカルロシミュレーションを行った。ポテンシャルエネルギー関数(力場)としては、CHARMM param19 のものを使った。膜及びその外側の水の存在やタンパク質のヘリックス以外の部分の存在を考慮するために、以下のような拘束エネルギーも課した(人工的に取り入れたバネの力により、膜から外へ出たり、ヘリックス同士が不自然に遠くまで行ってしまわないようにした)。

$$(3.1) \quad E_{\text{constr}} = \sum_{i=1}^{N_H-1} k_1 \theta(r_{i,i+1} - d_{i,i+1}) [r_{i,i+1} - d_{i,i+1}]^2 \\ + \sum_{i=1}^{N_H} \{ k_2 \theta(|z_i^L - z_0^L| - d_i^L) [|z_i^L - z_0^L| - d_i^L]^2 \\ + k_2 \theta(|z_i^U - z_0^U| - d_i^U) [|z_i^U - z_0^U| - d_i^U]^2 \} \\ + \sum_{C_\alpha} k_3 \theta(r_{C_\alpha} - d_{C_\alpha}) [r_{C_\alpha} - d_{C_\alpha}]^2.$$

ここで,  $N_H$  は膜貫通ヘリックスの総数,  $\theta(x)$  はステップ関数,  $k_1, k_2, k_3$  はバネ定数,  $r_{i,i+1}$  は  $i$  番目と  $i+1$  番目のヘリックスの隣り合う末端間の距離,  $d_{i,i+1}$  はその許容値,  $z_i^L (z_i^U)$  は  $i$  番目のヘリックスの下端(上端)の  $z$  座標値,  $z_0^L (z_0^U)$  はその標準値,  $d_i^L (d_i^U)$  は  $z_i^L (z_i^U)$  の  $z_0^L (z_0^U)$  からのずれの許容値,  $r_{C_\alpha}$  は  $C_\alpha$  原子の原点からの距離,  $d_{C_\alpha}$  は  $r_{C_\alpha}$  の許容値である. ここで, 許容値とは, 対応するずれの値がそれを越えると, バネの力が働いて元に戻そうとするようになる「臨界値」のことである.

最初の例は, 2本の膜貫通ヘリックスからなる, glycophorin A である. ヘリックスは18個のアミノ酸からなり, 両方とも(一文字表記で)以下のアミノ酸配列を持つ. TLHFGVMAGVIGTILLI. レプリカ数は13をとり, 以下の13個の温度を使った. 200, 239, 286, 342, 404, 489, 585, 700, 853, 1041, 1270, 1548, 1888 K. この場合, 式(3.1)中のパラメータは以下のようにとった.  $N_H = 2, k_1 = k_2 = 0.5 \text{ kcal}/(\text{mol } \text{Å}^2), k_3 = 0.05 \text{ kcal}/(\text{mol } \text{Å}^2), d_{i,i+1} = 20 \text{ Å}, z_0^L = -13.35 \text{ Å}, z_0^U = +13.35 \text{ Å}, d_i^L = d_i^U = 1.0 \text{ Å}, d_{C_\alpha} = 50 \text{ Å}$ .

まず, レプリカ交換が効率的に行われたことを確かめるために, 4つの時系列を図1に示す. 図1(a)は一つの温度(200 K)が MC sweep とともにどのレプリカに対応したかを示し, 図1(b)~図1(d)は一つのレプリカ(レプリカ6)における, それぞれ, 温度, ポテンシャルエネルギー, 自然の構造からの根二乗平均距離(RMSD; 2つの構造がどれくらい似ているかの指標でゼロのとき, 2つの構造が完全に一致する)の MC sweep の関数としての変化である. これらの値は全て最小値と最大値の間を酔歩して, レプリカ交換が効率よくなされたことが分かる.

図2には, レプリカ交換シミュレーション中に得られた構造のスナップショットを示す. いろいろな構造がサンプルされていることが分かる.

図3にはレプリカ交換シミュレーションで得られた立体構造の主成分分析を行った結果を示す. 第一主成分軸と第二主成分軸への射影である. 高温では, 構造の分布がばらばらであるが, 低温では, 3つの構造に分類できることが分かった.

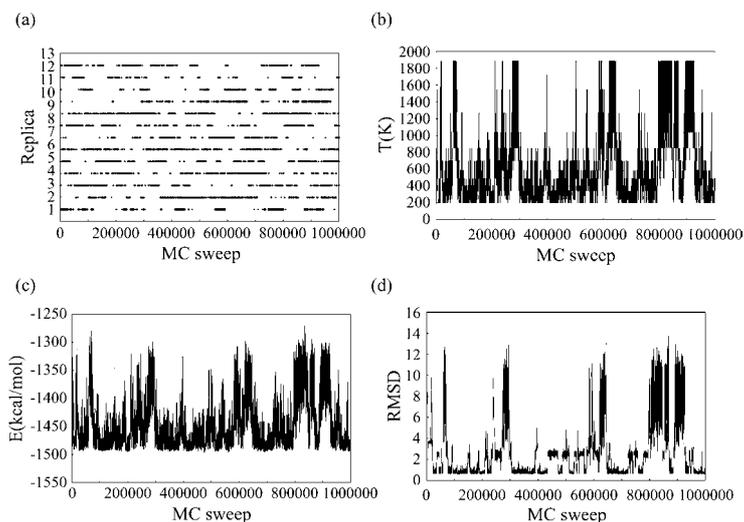


図1. glycophorin A のレプリカ交換モンテカルロシミュレーションにおける経時変化. (a)一つの温度(200 K)に対応するレプリカの経時変化, 一つのレプリカ(レプリカ6)における, (b)温度, (c)ポテンシャルエネルギー, (d)自然の構造からの根二乗平均距離, の経時変化.

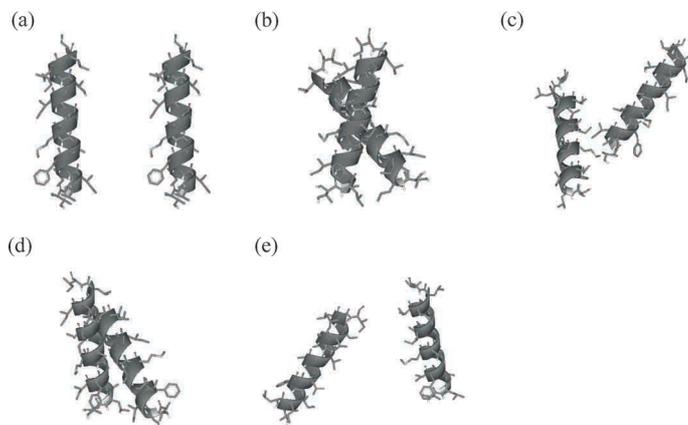


図 2. glycoprotein A のレプリカ交換モンテカルロシミュレーションにおけるスナップショット (a)は初期配置 .

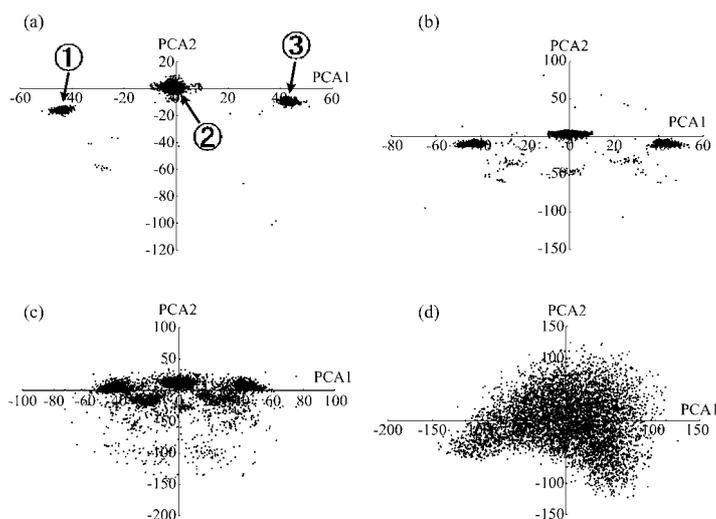


図 3. glycoprotein A のレプリカ交換シミュレーションで得られた立体構造の主成分分析 . PCA1, PCA2 はそれぞれ第一主成分軸と第二主成分軸に対応 (a)200 K (b)342 K, (c)585 K (d)1888 K.

図 4 では、実験で求められた自然の構造とレプリカ交換シミュレーションで得られた 3 つの構造 (図 3) を比較した . クラスタ 2 の構造が自然の構造と一致していることが分かる . また、クラスタ 1 とクラスタ 3 は同一の構造であることも分かる . 後者の違いは、2 本のヘリックスの番号の付け方の違いに起因するものであり、低温では 2 つの構造しかとらなことが分かる . そして、自然の構造に対応するクラスタ 2 が自由エネルギー (及びポテンシャルエネルギー) の最小状態であることが判明した . つまり、この場合、立体構造予測に完全に成功したと言える .

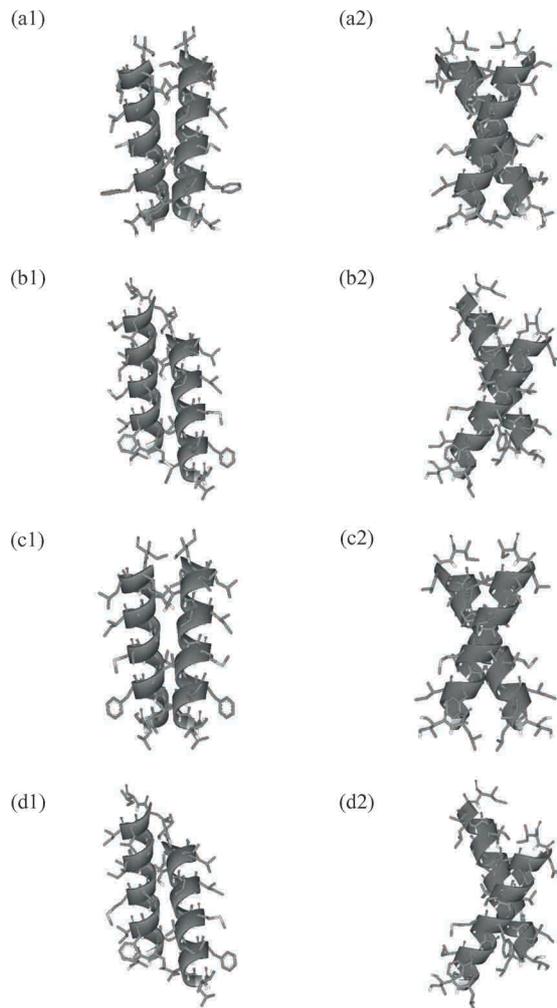


図 4. glycophorin A の自然の構造(a)とレプリカ交換シミュレーションで得られた 3 つの構造(b)(c)(d)の比較(図 3 のクラスター 1, 2, 3 に対応)。(a1)(a2)などの記号の後の番号は一つの構造を違う方向から見た図であることを示す。

最後に、より複雑な系である、bacteriorhodopsin における立体構造予測の結果を紹介する。すなわち、7 本の膜貫通ヘリックスのレプリカ交換シミュレーションの結果である。この場合、式(3.1)中のパラメータは以下のようにとった。 $N_H = 7$ ,  $k_1 = k_2 = 1.0 \text{ kcal}/(\text{mol } \text{Å}^2)$ ,  $k_3 = 0.05 \text{ kcal}/(\text{mol } \text{Å}^2)$ ,  $d_{i,i+1} = 20 \text{ Å}$ ,  $z_0^L = 0.0 \text{ Å}$ ,  $z_0^U = 31.5 \text{ Å}$ ,  $d_i^L = d_i^U = 2.0 \text{ Å}$ ,  $d_{C_\alpha} = 100 \text{ Å}$ 。

図 5 にエネルギー極小状態として得られた構造と自然の構造を比較する。7 本のヘリックスの配置が一致していることが分かる。

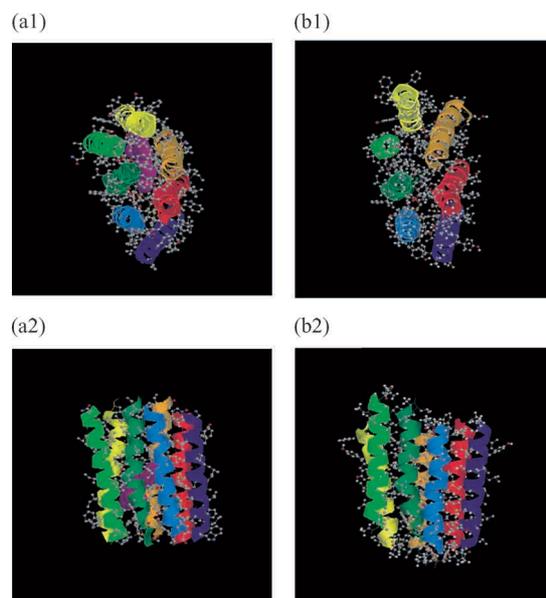


図 5. bacteriorhodopsin の自然の構造(a)とレプリカ交換シミュレーションで得られたエネルギー極小構造(b)の比較。(a1)(a2)などの記号の後の番号は一つの構造を違う方向から見た図であることを示す。自然の構造(a)では、シミュレーションでは取り入れなかったレチナル分子を紫で表した。

#### 4. おわりに

本稿では、タンパク質の立体構造予測問題において有効な演繹的手法として、拡張アンサンブル法について解説した。特にレプリカ交換法は従来の温度一定のカノニカルシミュレーションを並行して実行するだけなので、適用がきわめて容易である。実際、レプリカ交換分子動力学法はタンパク質の折り畳み問題などにおいて広く使われている。

#### 参 考 文 献

- Ferrenberg, A. M. and Swendsen, R. H. (1989). Optimized Monte Carlo data analysis, *Physical Review Letters*, **63**, 1195–1198.
- Hirokawa, T., Boon-Chieng, S. and Mitaku, S. (1998). SOSUI: Classification and secondary structure prediction system for membrane proteins, *Bioinformatics*, **14**, 378–379.
- Hukushima, K. and Nemoto, K. (1996). Exchange Monte Carlo method and application to spin glass simulations, *Journal of Physical Society of Japan*, **65**, 1604–1608.
- Iba, Y. (2001). Extended ensemble Monte Carlo, *International Journal of Modern Physics C*, **12**, 623–656.
- 伊庭幸人, 種村正美, 大森裕浩, 和合 肇, 佐藤整尚, 高橋明彦 (2005). 『計算統計 II マルコフ連鎖モンテカルロ法とその周辺』, 統計科学のフロンティア 12, 岩波書店, 東京.
- Kokubo, H. and Okamoto, Y. (2004a). Prediction of transmembrane helix configurations by replica-exchange simulations, *Chemical Physics Letters*, **383**, 397–402.
- Kokubo, H. and Okamoto, Y. (2004b). Prediction of membrane protein structures by replica-exchange

- Monte Carlo simulations: Case of two helices, *Journal of Chemical Physics*, **120**, 10837–10847.
- Kokubo, H. and Okamoto, Y. (2004c). Classification and prediction of low-energy membrane protein helix configurations by replica-exchange Monte Carlo method, *Journal of Physical Society of Japan*, **73**, 2571–2585.
- Kokubo, H. and Okamoto, Y. (2004d). Self-assembly of transmembrane helices of bacteriorhodopsin by a replica-exchange Monte Carlo simulation, *Chemical Physics Letters*, **392**, 168–175.
- Mitsutake, A., Sugita, Y. and Okamoto, Y. (2001). Generalized-ensemble algorithms for molecular simulations of biopolymers, *Biopolymers (Peptide Science)*, **60**, 96–123.
- Mitsutake, A., Sugita, Y. and Okamoto, Y. (2003). Replica-exchange multicanonical and multicanonical replica-exchange Monte Carlo simulations of peptides. I. Formulation and benchmark test, *Journal of Chemical Physics*, **118**, 6664–6675.
- 岡本祐幸 (2002). 拡張アンサンブルの方法, 『生体系のコンピュータ・シミュレーション』(岡崎 進, 岡本祐幸 編), 化学フロンティア No. 8, 25–42, 化学同人, 京都.
- Sugita, Y. and Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding, *Chemical Physics Letters*, **314**, 141–151.
- Sugita, Y., Kitao, A. and Okamoto, Y. (2000). Multidimensional replica-exchange method for free-energy calculations, *Journal of Chemical Physics*, **113**, 6042–6051.

## Prediction of Membrane Protein Structures by Generalized-ensemble Algorithms

Hironori Kokubo<sup>1</sup> and Yuko Okamoto<sup>2</sup>

<sup>1</sup>Department of Chemistry, University of Houston

<sup>2</sup>Department of Physics, Nagoya University

The Markov-chain Monte Carlo method is a computer simulation algorithm that reproduces statistical ensembles. It is usually based on the Boltzmann weight factor and realizes a fixed-temperature canonical ensemble. However, when the number of degrees of freedom of the system is large, there exist a huge number of local-minimum-energy states that are separated by high-energy barriers. This forces the simulation to get trapped in energy-local-minimum states and makes it very difficult to reproduce an accurate low-temperature canonical ensemble. Generalized-ensemble algorithms are generic terms for those methods that are based on non-Boltzmann weight factors and overcome the above-mentioned difficulty by realizing a one-dimensional random walk in energy space. We review one of the generalized-ensemble algorithms, namely, the replica-exchange method, and its extensions. As an example of its application, we present the results of replica-exchange Monte Carlo simulation applied to the prediction of membrane protein structures.