

GPGPU を駆使したタンパク質の安定性とダイナミクスの研究

光武 亜代理

慶應義塾大学理工学部物理学科

丸山 豊

理化学研究所計算科学研究機構

1. はじめに

人体を構成するのは水と骨、そしてタンパク質である。その中でも生命活動における様々な機能を受け持つタンパク質の重要性は言うまでもない。タンパク質は 20 種類のアミノ酸が共有結合であるペプチド結合でつながった 1 本鎖の高分子化合物である。アミノ酸配列の違いによりタンパク質の立体構造や機能が特徴づけられる。複雑な分子であるのに、生理的条件下で約 100 残基程度のものは、アミノ酸配列に依存して自発的にある決まった構造に折りたたまれる。タンパク質は、折りたたまれた構造になることで機能を生じることができる。つまりタンパク質の構造を決定することが、その機能を理解していく上で重要である。しかし、例えば人体を構成するタンパク質全ての構造が明らかになっているわけではない。タンパク質の折りたたみが複雑であることや生体膜などに埋め込まれて機能する膜蛋白等は結晶化しにくいと、X 線構造解析が難しいなど様々な要因が存在する。この折りたたまれる現象一つとってもまだ未解決で興味深い現象である。本研究は、物理化学の理論を駆使した分子シミュレーションを行うことにより、計算機上でタンパク質の折れたたみを再現し、その機構を分子レベルで調べることが目的である。

タンパク質が自発的に折りたたむのにミリ秒から秒の時間スケールがかかる。実験系と同程度の時間スケールのアンサンブルを得たい場合、直接実験と対応する長時間シミュレーションを行うか、効率の良いサンプリング手法を導入するかが必要である。長時間の分子動力学法(MD)を用いたシミュレーションを実行するのはスパコンでも難しい。本研究では、後者の方法を用いて構造空間を広くサンプリングすることを試みる。効率の良いサンプリング手法として拡張アンサンブル法が知られている。これは元々スピン系で開発された手法で、現在では生体分子系などでも広く利用され、その有効性が示されている。従来のカノニカルモンテカルロ法がボルツマン因子に基づいて状態をサンプルするのに対し、この拡張アンサンブル法は非ボルツマン因子による人工の重みを用いて状態をサンプルする方法である。エネルギー空間上のランダムウォークを実現することで、エネルギー極小状態にとどまるのを避けることができる。拡張アンサンブル法の一つとして、レプリカ交換分子動力学法がある。レプリカ交換法の重み因子は、いろいろな温度のボルツマン因子の積で与えられる。同時にいろいろな温度でシミュレーションを行い、ある頻度で、この重み因子に基づいて、それぞれのレプリカの温度を交換する。この温度交換により、系が極小状態にトラップされるのを回避し構造空間を広くサンプリングできる。レプリカ交換法は、はじめにモンテカルロ法を用いて開発された[1](この方法は、Multiple Markov chain method [2]や parallel tempering [3]とも呼ばれている。また、総説は文献[4, 5]を参照)。1997 年に初めて蛋白質の系に適用され[6]、その後、分子動力学法版

が開発された[7]。さらに多次元のレプリカ交換法[8]やエネルギー関数が温度に依存する場合のレプリカ交換法[9]なども開発された。現在、蛋白質系の分子動力学シミュレーションにおいて、既存のソフト(Amber など)に導入されるなど、広く使われ始めている。広く使われている要因は、他の拡張アンサンブル法に比べて重み因子を容易に決めることができることである。また、並列計算機に適しているのも要因の1つである。本研究では、レプリカ交換分子動力学法を用いて、広く構造空間をサンプルすることを試みた。

興味があるのはタンパク質であるが、水中のタンパク質の計算を行なう場合、実際には膨大な数のタンパク質の周りの水分子の計算が付随する。そのために水を含んだ大きな位相空間でサンプリングを行なうことになる。そこで、本研究では、水を陽に取り込むのではなく、分子性液体の積分方程式理論である RISM/3D-RISM を用いて水の効果を統計力学的に取り扱うことを試みた。統計力学に基づく液体論を用いると、単分子液体のオルンシュタイン・ゼルニケ方程式に対応する分子性液体に関する RISM 方程式が得られる。分子間の相互作用をインプットとして、RISM 方程式とクロージャー方程式を自己無頓着に解くことにより、タンパク質の周りの溶媒分子の分布関数を得ることができる。通常の MD では、あるタンパク質の構造に対して無数の水の配置が存在するが、積分方程式理論では一意の分布関数として表現される。そして、分布関数から溶媒和自由エネルギーを見積もることができる。イオンなども、容易に取り入れることが可能であり、溶媒分子がタンパク質の折れたたみにどう寄与するかを調べるのに、有効な方法である。

以前、申請者は、5 アミノ酸からなるエンケファリンの系に対してレプリカ交換モンテカルロ/1D-RISM 法の理論を構築し、シミュレーションを行い、その効果を実証済みである[9]。陽に水分子を入れた場合に比べて、レプリカの数減らせることができることがわかった。以前用いた 1D-RISM 理論は、溶質分子と溶媒分子間の距離 r のみ関数として溶媒の動径分布関数を計算する理論である。しかし、タンパク質のような複雑で大きな溶質分子の周りの溶媒和構造を記述するためには、距離のみの関数である動径分布関数では不十分である。そこで今回は 3 次元分布関数を計算する 3D-RISM 理論を組み合わせ、レプリカ交換 MD/3D-RISM のプログラムを作成した。3D-RISM 理論は、タンパク質を計算セルの中心に置き、タンパク質分子の周りの等間隔なグリッド点における溶媒分子の確率分布を計算することができる。

この計算を実行するにあたり最大の問題点は 3D-RISM プログラムの計算負荷が大きいことである。共同研究者である丸山豊博士(現：理化学研究所)が GPU 用の 3D-RISM プログラムを作成済みであり、そのプログラムを利用した。GPU 用の 3D-RISM のプログラムとレプリカ交換のアルゴリズムを組み合わせることにより、実用的な計算時間で計算を実行することが可能となった。GPU 対応の 3D-RISM 理論と拡張アンサンブル法を組み合わせたアルゴリズムは、世界になく、新規なものである。

本研究では、10 アミノ酸からなる chignolin の系で、本プログラムのテストを行った。chignolin は産業技術総合研究所の本田らが作成した 10 アミノ酸からなる世界最小の人工タンパク質である(図 1) [10]。このタンパク質は室温水中で図 1 左のような折りたたまれた構造(天然構造)を取る事が知られている。つまり図 1 右のような伸びた変性構造よりも天然構造が安定である。また、他の分子シミュレーションから、このタンパク質は、天然構造以外にも折りたたまれた構造であるミスフォールド構造(図 1 中央)も形成することが知られている[11]。chignolin は、その性質を調べられているだけでなく、様々な手法のテストにも用いら



図1 chignolin の主要な構造（左から天然、ミスフォールド、変性構造）

天然とミスフォールドは折りたたみ構造を取るが、2 番目の Tyr（濃灰色）と 9 番目の Trp（黒色）の疎水性側鎖の向きが違う。天然構造では重なりあっているが、ミスフォールド構造では共に溶媒に露出している。この 2 つの構造は常温水中で同程度の存在確率を示す事が知られている。また、タンパク質内部の水素結合の組み合わせも違っている。変性構造は長く伸びてタンパク質内部の水素結合を作らない。

れている。そこで、レプリカ交換 MD/3D-RISM 法のテスト系として選択した。

2. 方法

2.1 全エネルギー

タンパク質の構造が安定かどうかを判断するための指標はエネルギーである。通常的全エネルギーは、タンパク質分子内、タンパク質-溶媒分子間、溶媒分子間のポテンシャルエネルギーであるが、ここでは、溶媒分子の効果は、長時間平均したものと考え、溶媒和自由エネルギー項として取り入れる。つまり、全エネルギー E_T をタンパク質の構造エネルギー E と溶媒和自由エネルギー $\Delta\mu$ の和で与える。

$$E_T = E + \Delta\mu$$

ここで、構造エネルギー E は、共有結合項であるボンド、結合角、捻れ角エネルギー項と非共有結合項である静電相互作用と Lennard-Jones 相互作用から成り、通常のタンパク質分子内のエネルギーを表している。もう一つの溶媒和自由エネルギー $\Delta\mu$ は、バルクの水にタンパク質が挿入された時の溶媒和自由エネルギー変化を表す。水とタンパク質の相互作用エネルギーや水の再配向エネルギー、水の並進回転エントロピーからなる。タンパク質分子は固定して、溶媒である水分子の長時間平均をとったものに対応している。分子動力学シミュレーションを行う場合は、力（エネルギーの微分）が必要であるが、溶媒和自由エネルギーに関しても、微分の方法が提案されており、溶媒からの力を平均力として導入する。

構造エネルギーの計算には Amber99SB パラメータを用いた。水は TIP3P モデルを使用した。溶媒和自由エネルギーは 3D-RISM 理論によって計算した。

2.2 3D-RISM 法

3D-RISM 理論によって溶媒和自由エネルギーを計算する方法を述べる。ここでタンパク質は溶媒で無限希釈されていると仮定した。計算プロセスは 1) バルクの溶媒の計算、2) タンパク質の周りの溶媒の計算、の 2 段階から成る。バルクの溶媒の計算は、同じパラメータ（例えば水、

295.15 K、数密度 0.03334 個/Å³、など)を使うのであれば計算を一度行なえば良い。ここではプロセス 2 のタンパク質の周りの溶媒の計算について説明する。詳細は参考文献[12]を参照されたい。

タンパク質よりも充分大きい立方体もしくは直方体の計算セルを考える。タンパク質を計算セルの中心に置き、等間隔なグリッド点における溶媒分子の確率分布を計算するのが 3D-RISM 理論である。

3D-RISM 積分方程式は、全相関関数 $h_\gamma^{uv}(\mathbf{r})$ と直接相関関数 $c_\gamma^{uv}(\mathbf{r})$ を未知関数として

$$h_\gamma^{uv}(\mathbf{r}) = \sum_{\gamma'} c_{\gamma'\gamma}^{uv}(\mathbf{r}) * (\omega_{\gamma'\gamma}^{vv}(r) + \rho^v h_{\gamma'\gamma}^{vv}(r))$$

と表される。ここで $\omega_{\gamma'\gamma}^{vv}(r) = \delta(r - l_{\gamma'\gamma}^{vv})$ は溶媒分子の分子内相関関数 ($l_{\gamma'\gamma}^{vv}$ は原子間距離)、 γ 及び γ' は溶媒分子の原子種、 ρ^v は溶媒の数密度、 $*$ は畳み込み積分を表す。 $h_\gamma^{vv}(r)$ は溶質を含まない時のバルクの溶媒の動径全相関関数であり、3D-RISM の計算に先立って計算プロセス 1 で計算される。

3D-RISM 方程式には未知関数が 2 つあるため、他の方程式が必要となる。今回は Kovalenko と Hirata によって提案された PLHNC クロージャー方程式 (KH クロージャー方程式) を用いた。

$$h_\gamma^{uv}(\mathbf{r}) = \begin{cases} \exp(\chi_\gamma^{uv}) - 1 & \text{for } \chi_\gamma^{uv} \leq 0 \\ \chi_\gamma^{uv} & \text{for } \chi_\gamma^{uv} > 0 \end{cases}$$

$$\chi_\gamma^{uv} = -\beta u_\gamma^{uv}(\mathbf{r}) + h_\gamma^{uv}(\mathbf{r}) - c_\gamma^{uv}(\mathbf{r})$$

ここで $\beta = 1/(k_B T)$ は温度の逆数 (k_B : ボルツマン定数、 T : 絶対温度) で、 $u_\gamma^{uv}(\mathbf{r})$ は、溶質と溶媒 γ の間のポテンシャルを表す。このポテンシャルはレナード・ジョーンズ項と静電相互作用項からなり

$$u_\gamma^{uv}(\mathbf{r}) = 4 \sum_\alpha \epsilon_{\alpha\gamma} \left[\left(\frac{\sigma_{\alpha\gamma}}{|\mathbf{r}_\alpha - \mathbf{r}|} \right)^{12} - \left(\frac{\sigma_{\alpha\gamma}}{|\mathbf{r}_\alpha - \mathbf{r}|} \right)^6 \right] + \frac{1}{4\pi\epsilon_0} \sum_\alpha \frac{q_\alpha q_\gamma}{|\mathbf{r}_\alpha - \mathbf{r}|}$$

と表す。ここで $\epsilon_{\alpha\gamma}$ と $\sigma_{\alpha\gamma}$ は、それぞれレナード・ジョーンズのエネルギー及びサイズパラメーターであり、 ϵ_0 は真空の誘電率、 q_α は電荷である。

3D-RISM 方程式は畳み込み積分の式なのでフーリエ空間で解き、PLHNC 方程式は実空間で解く。 $h_\gamma^{uv}(\mathbf{r})$ と $c_\gamma^{uv}(\mathbf{r})$ をフーリエ変換・逆フーリエ変換しながら 2 つの方程式を交互に解いて収束させていき、解を求める。最終的に求められた $h_\gamma^{uv}(\mathbf{r})$ と $c_\gamma^{uv}(\mathbf{r})$ から溶媒和自由エネルギー $\Delta\mu$ を次の式を用いて計算する事ができる。

$$\Delta\mu = \rho_v \sum_\alpha \int_{V_{cell}} d\mathbf{r} \left[\frac{1}{2} (h_\gamma^{uv}(\mathbf{r}))^2 \Theta(-h_\gamma^{uv}(\mathbf{r})) - c_\gamma^{uv}(\mathbf{r}) - \frac{1}{2} h_\gamma^{uv}(\mathbf{r}) c_\gamma^{uv}(\mathbf{r}) \right]$$

ここで Θ はヘビサイドステップ関数を表す。

3D-RISM のプログラムは MD に比べると計算負荷が高いため、どう高速化するかがポイントとなる。このプログラムは GPU 化されており Xeon 8 コアと比較して NVIDIA の Fermi で 9 倍程度の高速化を達成しており、今回の計算では GPU 版を用いた。3D-RISM は 3D-FFT を多用しメモリ律速なプログラムであるため、GPU に向いている[13]。

2.3 レプリカ交換法

レプリカ交換法の扱う系は、 M 個の違う温度 $T_m (m=1, \dots, M)$ をもつ相互作用しない M 個の独立な系(レプリカ)からなる。レプリカを表すラベルを i 、温度を表すラベルを m とする時、 i と m は置換関係にある ($i = i(m), m = m(i)$)。系の状態 X は各レプリカの状態 $x_m^{(i)} = (q^{(i)})_m$ により決まり、 $X = \{x_1^{(1)}, \dots, x_M^{(M)}\} = \{x_{m(1)}^{(1)}, \dots, x_{m(M)}^{(M)}\}$ で表される。ここで、 $q^{(i)}$ はレプリカ i の配位を表す。

レプリカ同士は相互作用せず、レプリカ i は温度 $T_{m(i)}$ の系であり、レプリカ交換法での重み因子は、ボルツマン因子の積で与えられる：

$$\begin{aligned} W_{REM}(X) &= \prod_{i=1}^M \exp(-\beta_{m(i)} E(x_m^{(i)})), \\ &= \prod_{m=1}^M \exp(-\beta_m E(x_m^{[i(m)]})). \end{aligned} \quad (1)$$

ここで、 $T_1 < T_2 < \dots < T_M$ とした。よって、 M 個の温度分布 $T_m (m=1, \dots, M)$ を与えてやるだけで、重み因子が決まる。

レプリカ交換法では、初めに M 個の温度 $T_m (m=1, \dots, M)$ を用意し、それぞれの温度でカノニカルシミュレーションを行う。ここまでは従来のカノニカルシミュレーションを独立に行うのと同じであるが、この方法では、ある頻度で隣接した温度に対応するレプリカ対を詳細釣り合いの条件に基づいて交換する(図2)。シミュレーション中で温度を交換しなければ、従来の M 個の温度で M 個の独立なカノニカルシミュレーションを行うことと同じである。しかし、レプリカ交換シミュレーションでは、温度の交換を行う(レプリカ交換 MD の場合は、温度交換の際に、速度のスケーリングも行う)。この温度交換により、系が極小状態にトラップされるのを回避できる。従来の低温のシミュレーション中は、構造揺らぎが小さいので、状態が極小エネルギー

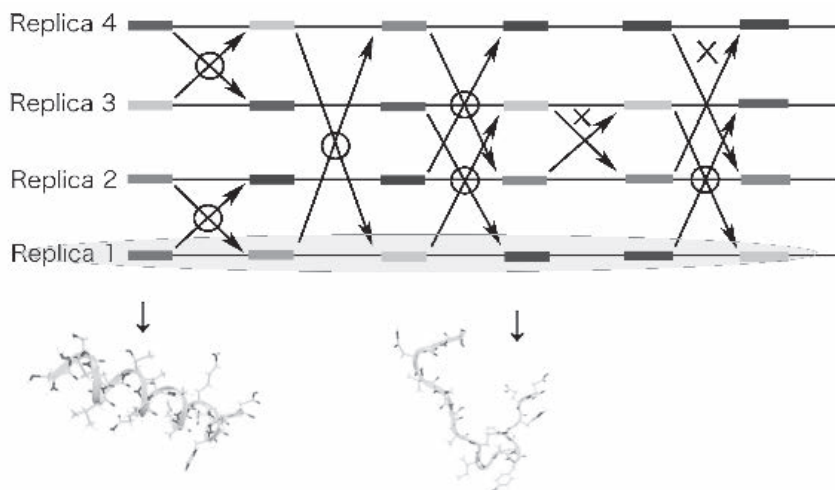


図2 レプリカ交換法の模式図

横軸は時間の進行方向を表す。青が低温、赤が高温を表す(はじめのレプリカの色(濃さ)に関して下から温度の低い色(濃さ)に対応)。矢印はレプリカ交換の組み合わせとタイミングを表す。組み合わせを交互に変えてレプリカ交換を行なう。一つのレプリカがいろいろな温度を渡り歩くことにより、温度空間(エネルギー空間)のランダムウォークが生じ、極小構造にトラップされることを避け、構造空間を広くサンプリングする。

状態にトラップされてしまう。しかし、レプリカ交換シミュレーションだと、1つのレプリカはいろいろな温度を渡り歩く（たとえば、図2のReplica1の温度の変化）。レプリカの温度が低温で、状態が極小状態にトラップされても、しだいに温度が上がることで構造の揺らぎが大きくなり、極小状態から出ることができる。そして、再度低温になると、違う極小状態を探ることができる。このように極小状態のサンプリングが良くなる。また、重み因子に従ってシミュレーションを行うので、最終的に M 個の与えた温度 $T_m(m=1,\dots,M)$ に対応する M 個のカノニカル分布を得ることができる。

どのように温度の分布を与えるか、レプリカ数をどうするか、レプリカ交換を行う頻度の設定は系に依存するため、新しい系でシミュレーションを行う場合は、最初に最適なパラメータを決める必要がある。

3. 結果

3.1 パラメータ決定のための予備計算

レプリカ交換法を実際に行なうにあたり、シミュレーションのパラメータを決めなければならない。レプリカ交換法に必要なパラメータは以下の3つである。

1. レプリカの数
2. 各レプリカの温度
3. レプリカ交換を行なう時間間隔

これらのパラメータは相互に関係し、サンプリングの精度や効率に影響する。このうちレプリカ数とレプリカの温度に関しては、各温度における全エネルギーの分布が重なるように設定する必要がある。この分布が重ならないとレプリカ交換が行なわれない。また、レプリカの温度を十分に高温にしないとローカルミニマムから脱出できずに正確なサンプリングが行なえない。そこでレプリカ数と各レプリカの温度を決めるために事前に MD/3D-RISM シミュレーションを行った。具体的には天然状態と変性状態を初期構造とし、298.15K から 450K までのいくつかの温度で短時間の一定温度での MD/3D-RISM シミュレーションを行い、全エネルギー分布を調べた。低温の分布に関しては、短いシミュレーションの場合、極小エネルギー状態にトラップされて正確な全エネルギー分布は得られないが、結果から、最終的に 298.15 と 313.11, 328.82, 345.31, 362.64, 380.83, 399.94, 420.00K の 8 つの温度を用いて計算をすることとした。残るパラメータである「レプリカ交換を行なう時間間隔」は、実際にレプリカ交換の計算を行なって決めていった。

3.2 3D-RISM 計算

レプリカ交換 MD/3D-RISM の計算を行なう前に天然構造やミスフォールド構造、変性構造についてのエネルギープロファイル調べた。具体的には、通常の分子動力学法シミュレーションによって構造のサンプリングを行い、Root Mean Square Distance (RMSD) や二面角などによって各構造を分類した。さらに、それらの構造に対して 3D-RISM 計算を行い、溶媒和自由エネルギーを求めた。図3に各構造における溶媒和構造を示す。この溶媒和構造は統計力学的な平均量であり、例えば、タンパク質の構造を固定して長時間水の計算をして分布関数を計算した場合と同じである。構造毎に溶媒和構造が変化していることがわかる。この分布関数から溶媒和自由エネルギーを求めることができる。表1に各構造の全エネルギーの平均値とその成分を示す。

天然構造とミスフォールドの全エネルギーの値はほぼ同じでミスフォールドの方が 0.1 kcal/mol 低い値となっている。この結果は使用するポテンシャルパラメータに依存し、Amber99SB ではミスフォールドの存在確率が高まるという MD の結果と一致する[11]。また、変性構造は他の構造よりも高い全エネルギーで不安定であることが示されている。つまり全エネルギーが構造の安定性の指標として正しく機能していることが示されている。

構造エネルギーをみると天然構造が一番低い値となっている。しかし変性構造は約 130 kcal/mol も高い値となっている。一方で溶媒和自由エネルギーは変性構造が最も低く、天然構造が最も高くなっている。天然構造はタンパク質内部で水素結合を形成し構造エネルギーの利得を得るが、周りの溶媒との水素結合は最小化し溶媒和自由エネルギーの利得が減少する。一方、変性構造は内部で水素結合を作らず、溶媒との水素結合を最大化する。このように構造エネルギーと溶媒和自由エネルギーは競合する関係となっている。そのため、 E や $\Delta\mu$ での各構造のエネルギー差は大きくても全エネルギーの差は小さくなる。

3.3 レプリカ交換 MD/3D-RISM

事前の計算で全エネルギーが構造の安定性の指標として、正しく機能していることがわかったので、レプリカ交換 MD/3D-RISM シミュレーションを行うことを試みた。298.15 と 313.11、328.82、345.31、362.64、380.83、399.94、420.00K の 8 個の温度を用意し、伸びた構造からシミュレーションを始めた。ただし、溶媒和自由エネルギーに関しては、298.15K の温度に固定した。レプリカ交換 MD/3D-RISM を行なうにあたり、3つのパラメータを決める必要があった。そのうちレプリカ数と各レプリカの温度は予備計算を行なって決定した。残る 1 つは、レプリ

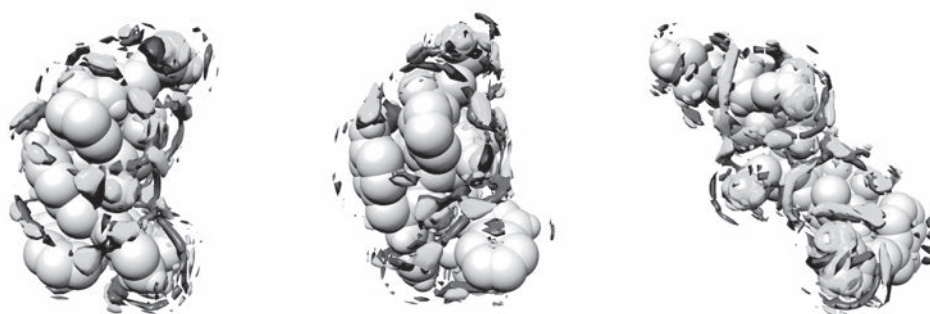


図3 天然（左）、ミスフォールド（中央）、変性構造（右）の周りの溶媒和構造
赤い(濃い色の分布)領域は水の酸素、水色（薄い色の分布）の領域は水素を表す。バルクの水よりも3倍分布確率が高い領域を示している。タンパク質の原子は球状に表示している。

表1 各構造の全エネルギーとその成分（構造エネルギーと溶媒和自由エネルギー）

Type	E_T	E	$\Delta\mu$
天然構造	-171.1	-26.4	-144.7
ミスフォールド	-171.2	-12.2	-159.0
変性構造	-148.1	103.2	-251.3

エネルギー単位は kcal/mol

カ交換を行なう時間間隔である。通常の MD を使ったレプリカ交換法では、2ps 程度の値が用いられているが、MD/3D-RISM は全く新しい手法であるため、実際にプログラムを実行させて決める必要があった。MD/3D-RISM の特性から 2ps よりも短い時間を想定して計算を行なっていたが、最終的には 5ps となった。

まず図 4 に 2ps の時のレプリカ交換のランダムウォークの様子を示す。1 から 8 へ到達し、また 1 まで戻ってくるのを 1 往復として 10 往復している。これは通常の MD を使ったレプリカ交換 MD において十分にサンプリングを行なえる条件である。この時の各温度の全エネルギーの分布を図 5 に示す。各分布は隣接する温度の分布とよく重なり充分交換可能であることがわかる。このようにレプリカ交換間隔 2fs でのサンプリングは一見機能しているようにみえる。しかし、実際 298.15K の全エネルギーの分布を見てみると、分布のピークの全エネルギーの値が -140 kcal/mol であり、事前に計算した天然構造やミスフォールド構造のアンサンブルの平均値とは違うことがわかる。レプリカ交換シミュレーションは、短いステップ数の場合、最初は初期状態に依存するため、低温のシミュレーションの全エネルギー分布は構造がトラップされたものに対応するものが得られる場合がある。今回も、温度は 10 往復しているが、構造的には、まだ初期状態にトラップされている可能性が考えられる。しかし、298.15K の構造のなかに、天然構造やミスフォールド構造よりも全エネルギーが低いものもある。表 2 に、298.15K で得られた構造のうち最も全エネルギーの低い順に 5 つ列記した。全エネルギーは予備計算における天然構造の平均値よりも低い値に達している。しかし、構造エネルギー E は表 1 の天然構造と比較して 100 kcal/mol ほど高い値となっている。これは、伸びた構造に対応する。全エネルギーが天然構造やミスフォールド構造の平均の全エネルギーより低いものは得られたが、シミュレーションとしては、まだ、低温の状態がトラップされていると考えられる。同じ温度で、構造探索を促すために、さらに交換間隔を 5 及び 10ps にした計算を実行した。

図 6 に交換間隔 5ps の時のランダムウォークの様子を示す。残念ながら計算期間の関係で 10 往復には届かず、3 往復半までの計算となった。しかし、298.15K でのサンプリングの結果は改善された。各温度の全エネルギーの分布を図 7 に示す。298.15K の分布のピーク値に対応する全エネルギーは -145kcal/mol 程度であり、2ps の時よりも、低くなっていることがわかる。また、表 3 に、298.15K で得られた構造のうち最も全エネルギーの低い順に 5 つ列記した。298.15K の全エネルギーは 2ps の時よりも 10 kcal/mol 低い構造に到達し、構造エネルギーも -30 kcal/mol を下回った。計算された構造も折りたたまれた状態になっていた。298.15K のサンプリングが改善されていることがわかった。しかし平均値は、天然構造やミスフォールド構造の平均値よりもまだ大きな値を持つため、すべての温度のカノニカル分布を得るためには、さらにシミュレーションステップが必要であると考えられる。

レプリカ交換の頻度を頻繁に起こすために、交換間隔を短くしたいと考えるが、一定温度でのサンプリングも必要なため、その兼ね合いで、交換間隔を設定する。温度の交換だけを見ても、構造が広くサンプルされていない場合があるため、構造を反映するようなパラメータの振る舞いをみる必要がある。通常は、オーダーの違う交換間隔で試して、アクセプタンス比や、構造を見て適当な交換間隔を決めるが、計算時間の兼ね合いで、2ps、5ps、10ps で試してみても、どの交換間隔が有効であるか調べた。ここでは、交換間隔 2 ps では不十分で 5 ps が、適当であることがわかった。また、本エネルギー関数を用いて、折りたたまれた構造も得ることができるということが示唆された。

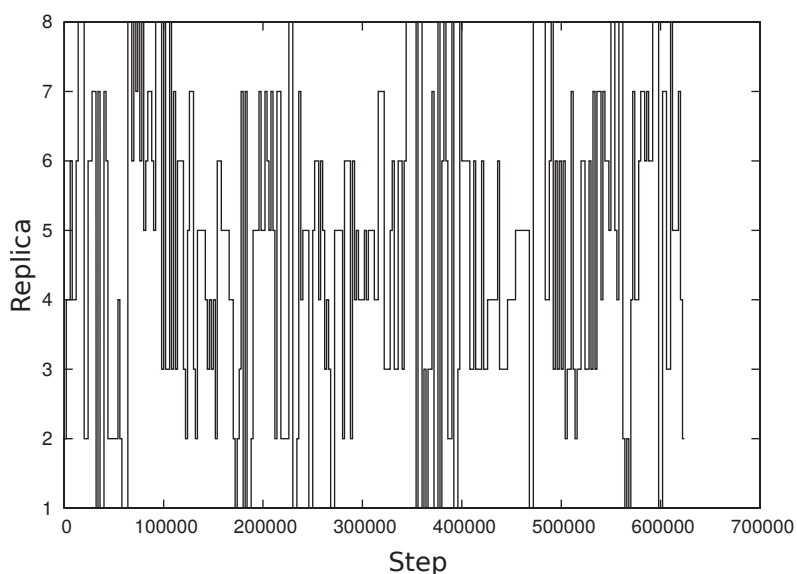


図4 交換間隔 2ps(1000 ステップ)でのレプリカ 1 のランダムウォーク

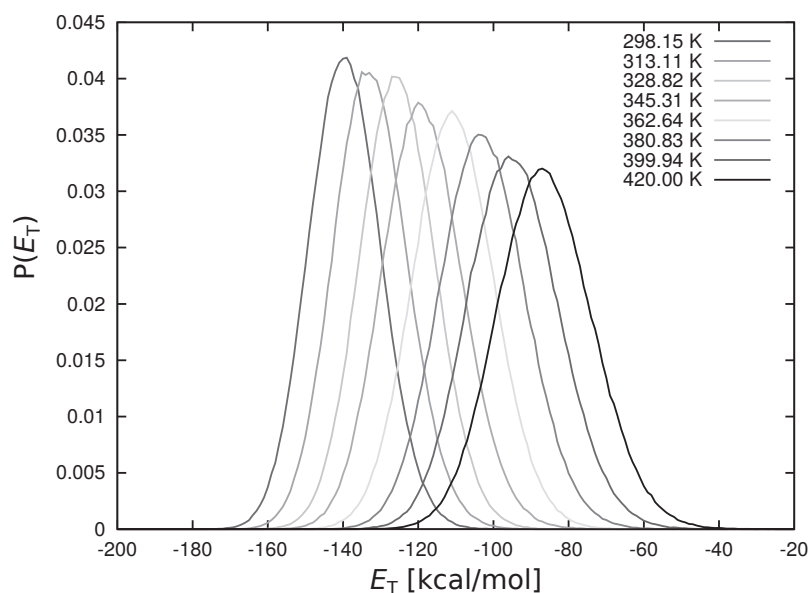


図5 交換間隔 2ps(1000 ステップ)での各温度のエネルギー分布

表2 レプリカ交換時間間隔 2ps の場合の到達最低エネルギー

	E_T	E	$\Delta\mu$
1	-178.1	73.5	-251.6
2	-177.0	102.2	-279.2
3	-177.0	107.5	-284.5
4	-176.2	67.4	-243.6
5	-175.7	85.5	-261.1

エネルギー単位は kcal/mol

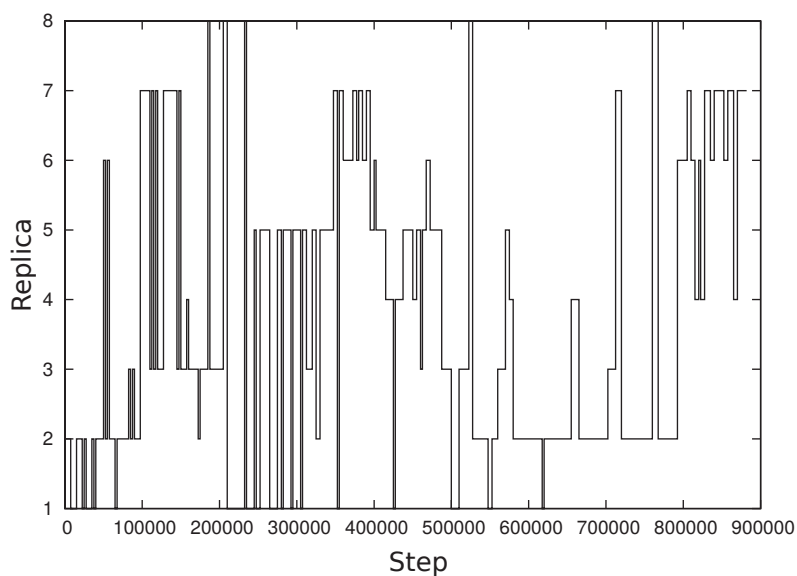


図 6 交換間隔 5ps(2500 ステップ)でのレプリカ 1 のランダムウォーク

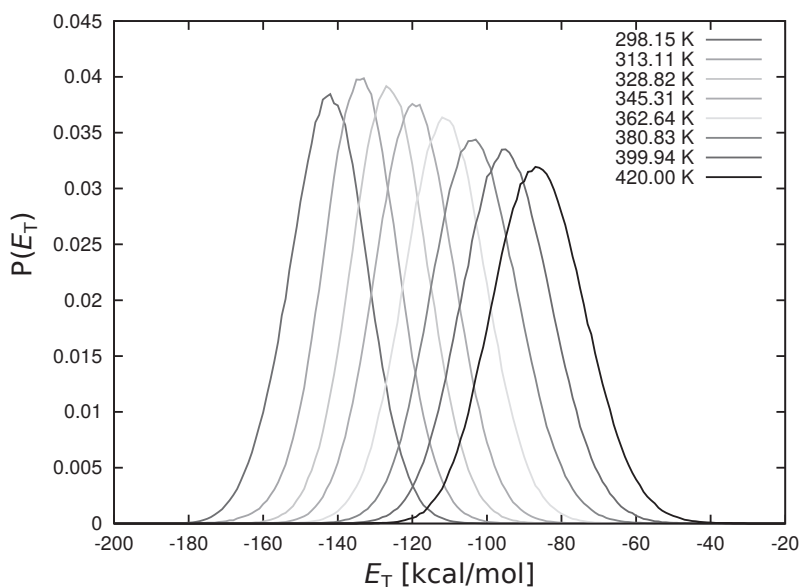


図 7 交換間隔 2ps(1000 ステップ)での各温度のエネルギー分布

表 3 レプリカ交換時間間隔 5ps の場合の到達最低エネルギー

	E_T	E	$\Delta\mu$
1	-188.8	-33.2	-155.6
2	-188.2	-24.4	-163.7
3	-187.8	-11.4	-176.3
4	-187.1	-24.8	-162.3
5	-187.0	-7.4	-179.6

エネルギー単位は kcal/mol

4. 結論

タンパク質の構造を効果的に探索するためにレプリカ交換 MD/3D-RISM 法のプログラムを開発した。負荷が大きい 3D-RISM 部分の計算には、GPU 用のプログラムを使用して計算の高速化を行い、実用的な計算時間のプログラムを作成することができた。ミニタンパク質である chignolin を対象にテスト計算を行ない、レプリカ交換 MD/3D-RISM の計算に必要なパラメータを決定した。レプリカ交換の時間間隔 5ps で折りたたまれた構造のサンプリングが行われることを確認した。chignolin を対象とした系ではレプリカ数 8 で 5ps とするのが妥当であると考えられる。これらのパラメータ決定方法は、今後の課題である。

時間間隔 5ps と 10ps の計算は計算期間内に終了しなかったが、研究室に導入した GPU クラスタで計算を継続中である。特にレプリカ交換の時間間隔がより大きい場合の振る舞いを調べておくことは今後のパラメータ決定において重要であると考ええる。そして、ステップ数を増やすことにより折れたたみを実現したいと考えている。現時点では、まだ、コストの高い計算システムであるが、3D-RISM 法を用いるとレプリカ数があまり必要ないということが経験上わかっており、大きなタンパク質にも適用できると考えられる。また、溶媒としてイオンなどの導入も可能であるため、様々な応用が可能であり有望な方法であると考ええる。

謝辞

本研究は、新学術研究「コンピューティクスによる物質デザイン：複合相関と非平衡ダイナミクス」、さががけ「ライフサイエンスの革新を目指した構造生命科学と先端基盤技術」の支援を受けて行われました。また、東京大学情報基盤センター「若手・女性利用」の GPU クラスタを用いて計算を行いました。

参 考 文 献

1. K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. 65 (1996) 1604; K. Hukushima, H. Takayama, and K. Nemoto, Int. J. Mod. Phys. C7 (1996) 337.
2. M.C. Tesi, E.J.J. van Rensburg, E. Orlandini, and S.G. Whittington, J. Stat. Phys. 82 (1996) 155.
3. E. Marinari, G. Parisi, and J.J. Ruiz-Lorenzo, Spin Glasses and Random Fields edited by A.P. Young (World Scientific, Singapore, 1998), p. 59.
4. Y. Iba, Int. J. Mod. Phys. C 12 (2001) 623.
5. A. Mitsutake, Y. Sugita, and Y. Okamoto, Biopolymers (Peptide Science) 60 (2001) 96.
6. U.H.E. Hansmann, Chem. Phys. Lett. 281 (1997) 140.
7. Y. Sugita and Y. Okamoto, Chem. Phys. Lett. 314 (1999) 141.
8. Y. Sugita, A. Kitao, and Y. Okamoto, J. Chem. Phys. 113 (2000) 6042.
9. G.L. Penna, A. Mitsutake, M. Masuya, and Y. Okamoto, Chem. Phys. Lett. 380 (2003) 609; A. Mitsutake, M. Kinoshita, Y. Okamoto, and F. Hirata, J. Phys. Chem. B 108 (2004) 19002.
10. S. Honda, K. Yamasaki, Y. Sawada, and H. Mori, Structure 12, 1507-1518, 2004.
11. P. Kuhrova, A. De Simone, M. Otyepka, and R. B. Best, Biophys. J. 102 1897-1906 2012.
12. F. Hirata (Ed.) "Molecular Theory of Solvation" Kluwer Academic Publishers 2003.

13. Y. Maruyama and F. Hirata “Modified Anderson Method for Accelerating 3D-RISM Calculations Using Graphics Processing Unit” J. Chem. Theory Comput. 8, 3015-3021, 2012.