

カスケード型超並列シミュレーションによる タンパク質構造遷移のパスウェイ探索

西原 泰孝¹・原田 隆平²・北尾 彰朗¹

(受付 2013 年 12 月 30 日；改訂 2014 年 8 月 12 日；採択 8 月 21 日)

要 旨

タンパク質の構造と機能の関係を解明するうえで、構造変化に対する自由エネルギーを曲面として表す自由エネルギー地形の計算とその解析は重要である。しかし、分子シミュレーションを用いて自由エネルギー地形の計算を行う場合、長い計算時間を必要とすることが多い。それは、自由エネルギー地形上に多数存在する局所安定状態にタンパク質が留まってしまうためである。本稿では、Parallel Cascade Selection Molecular Dynamics (PaCS-MD) と Markov State Model (MSM) を組み合わせ、自由エネルギー地形をより短時間で効率よく計算する方法を紹介する。PaCS-MD は、短いシミュレーションから出現確率の低い構造を選ぶことによって、目的の構造へ高効率で遷移させるカスケード型超並列シミュレーション法である。そして、MSM は、対象としている系がマルコフ的であるとき、長いシミュレーションから状態間の遷移行列を推定し、平衡状態の確率分布を計算する方法である。これら PaCS-MD と MSM を組み合わせることで、多数の局所的な状態間遷移情報から、大域的な遷移行列と平衡状態の確率分布を得ることができる。PaCS-MD と MSM を組み合わせた手法を小タンパク質のフォールディング過程に適用し、得られた自由エネルギー地形について議論する。

キーワード：自由エネルギー地形、局所安定構造、分子動力学、フォールディング、マルコフ過程。

1. はじめに

近年、タンパク質の立体構造と機能との関係を明らかにするために、タンパク質の構造とその変化の研究が盛んに行われている。タンパク質の構造変化は多種多様であり、他分子と相互作用する部位のみが構造変化する場合もあれば、タンパク質全体が大規模に構造変化する場合もある。また、シトクロム *c* のフォールディングのようにタンパク質 1 分子の構造変化もあれば、筋肉の収縮といった複数のタンパク質が集合して構造変化を起こすものもある。

タンパク質の立体構造とその変化の研究において解析によく用いられているのが、自由エネルギー地形である。自由エネルギー地形では、構造変化に対する自由エネルギーが曲面として表されている。安定な構造は高い自由エネルギー障壁に囲まれていることから、大きな構造変化は自由エネルギー障壁を乗り越えることで実現する。したがって、自由エネルギー地形を解析することにより、構造変化の起こりやすさをエネルギーの観点から読み取ることができるの

¹ 東京大学 分子細胞生物学研究所：〒 113-0032 東京都文京区弥生 1-1-1

² 理化学研究所 計算科学研究機構：〒 650-0047 兵庫県神戸市中央区港島南町 7-1-26

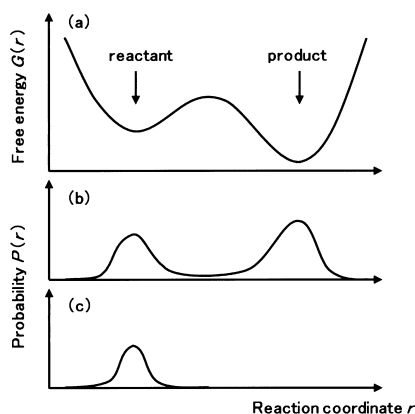


図 1. 自由エネルギー地形と確率分布の例. (a) 2 状態のモデル自由エネルギー地形. (b) (a) の自由エネルギー地形と一致する確率分布. (c) 短い時間のサンプリングによる確率分布.

で、構造遷移に沿ったパスウェイを見つけることができる。また、自由エネルギー地形を比較することによって、例えば、分子モーターを構成しているタンパク質のどの部分が分子モーターの機能発現に影響を与えているのか、酵素反応においてアミノ酸残基の変異がどのような影響を与えるのか、という点なども明らかにすることができる。

自由エネルギー地形の作成には、分子動力学 (Molecular Dynamics: MD) シミュレーションやそのバリエーションが広く用いられている。MD シミュレーションでは原子解像度で分子の描像が得られるため、構造が変化している部位を捉えやすいからである。しかし、通常の MD シミュレーションを用いて自由エネルギー地形の計算を行う場合、タンパク質の構造変化に長時間の計算を必要とすることが大きな問題である。この問題を解決するために、様々な計算法が既に提案されているものの、依然として計算に大変な労力を必要とする場合が多い。

そこで本稿では、この問題に対して 2 つの手法、Parallel Cascade Selection Molecular Dynamics (PaCS-MD) (Harada and Kitao, 2013) と Markov State Model (MSM) (Prinz et al., 2011; Pan and Roux, 2008; Chodera et al., 2007) を組み合わせて効率良く自由エネルギー計算を行う手法を提案し、解説する。PaCS-MD は、最近開発された手法で、構造既知の 2 つの状態間の遷移をバイアスなしで引き起こすことができる。具体的には、短い MD シミュレーションをカスケード的に多数行うことで大量の構造を発生させ、タンパク質の構造空間を広くサンプリングする手法である。目的の構造へ変化するパスウェイのトラジェクトリが発生するので、構造遷移の情報も同時に得ることができる。一方、MSM は、タンパク質のフォールディング過程等の研究で広く使われており、既に確立した手法である。MSM では、系がマルコフ過程に従うとしてデータの解析を行い、多数の局所的な遷移情報から平衡状態の確率分布を得ることができる。

2. PaCS-MD と MSM を用いた自由エネルギー計算

2.1 サンプリング問題

自由エネルギー地形を計算する場合、タンパク質の構造空間を広範囲にサンプリングする必要がある。簡単な例として、2 つの安定構造をもつ自由エネルギー地形を考えてみよう (図 1(a))。この自由エネルギー地形上を MD シミュレーションで十分に長い時間サンプリングできたとすると、対応する確率密度分布は図 1(b) のようになる。しかし、計算時間には限りがあるうえ、

2つの安定構造を隔てている自由エネルギー障壁が高いと、片方の安定構造付近にトラップされてしまう。MD シミュレーションを reactant の構造から始めた場合、図 1(c) のように狭い領域しかサンプリングできない。図 1(c) のような確率密度分布では、図 1(a) とは異なる不適切な自由エネルギー地形が得られてしまう。このように、サンプリング領域や初期状態の依存性などの問題は「サンプリング問題」として知られている。

このようなサンプリング問題を避けるために、これまでに様々な手法(例えば、レプリカ交換法(Hukushima and Nemoto, 1996; Sugita and Okamoto, 1999)やアンブレラサンプリング(Torrie and Valleau, 1977), Metadynamics(Laio and Parrinello, 2002; Nishihara et al., 2008), Steered MD 法(Isralewitz et al., 2001a, 2001b; Jensen et al., 2002)が提案されている。しかし、レプリカ交換法では交換の際に詳細釣り合いなどを満たす必要がある。また、これらの方法の多くは、構造変化を引き起こしたり不安定な領域の構造をサンプリングするために、系の温度を上げたり、系の特定の部位に力やポテンシャルといったバイアスを加えた計算を行っている。このようなバイアスは、かける位置や大きさ、向き、かける時間などのパラメータを最適化したうえで使用しなければならないので、サンプリングを行う前に最適化のための前計算が必要となる。しかし、最適なパラメータとは、求める自由エネルギーに応じたパラメータとなるため、パラメータの最適化自体も困難な計算となる。場合によっては、サンプリングの計算時間よりもパラメータの最適化にかかる計算時間の方が長いこともある。

2.2 PaCS-MD

PaCS-MD はバイアスなしで構造既知の2つの状態間の遷移を促す方法として提案された。また、出発状態からある方向への遷移を誘起するために用いることもできる。

図 2 に PaCS-MD の手順を示す。はじめに、始状態にある構造を用いて短時間の MD シミュレーションを行う(前計算)。得られた構造のスナップショットの中から、複数の構造を選んで初期構造とし、新たに初期速度を与えた短時間 MD シミュレーションを複数独立に行う(1st cycle)。次に、より目的の構造に近い構造を複数選んで同じ操作を繰り返し、これを目的の状態に十分近くなるまで行う(2nd to Nth cycle)。構造選択には、始状態と終状態を区別する任意の量が使用できる。典型的な例では、自由エネルギー地形を描く際の反応座標がある。また、反応座標以外の自由度の緩和が遅いために配置空間をうまくサンプリングできない場合、Root Mean Square Deviation (RMSD) を用いたり、主成分分析における第一主成分や複数の主成分の線形結合などを選べばよい。

PaCS-MD では目的の構造への近さを基準に選択を行っており、構造遷移のためのバイアス等は一切使用していない。そのため、バイアスを最適化するための前計算を必要としない。また、この計算法ではより短時間に効率よく構造遷移を観測することができる。その理由は、終状態に近い構造を抽出することが、稀にしか発生しない構造の揺らぎの出現確率を選択的に上昇させることに対応するからである(図 2)。

さらに、PaCS-MD は超並列計算に適した手法でもある。並列計算では1つのノードのダウンが全体の計算を止めてしまう。大規模並列計算の場合だと、ノードの数が膨大なため、計算の止まる頻度が上がってしまう。また、レプリカ交換法のように複数のシミュレーションのレプリカを確実に交換して計算を進めていく場合、1つの計算の停止のためにレプリカ交換の停止が起こるため計算が進まない。それに対して、PaCS-MD での1つのサイクル内の MD シミュレーションはそれぞれ独立であり、相互に依存していない。したがって、あるノードがダウンして1つの MD シミュレーションが停止しても、その部分の結果を無視して計算を進めてもよく、1つのシミュレーションの停止は、他のシミュレーションに影響を与えないのである。このことは PaCS-MD が超並列計算でより高い並列効率を持つ可能性を示している。

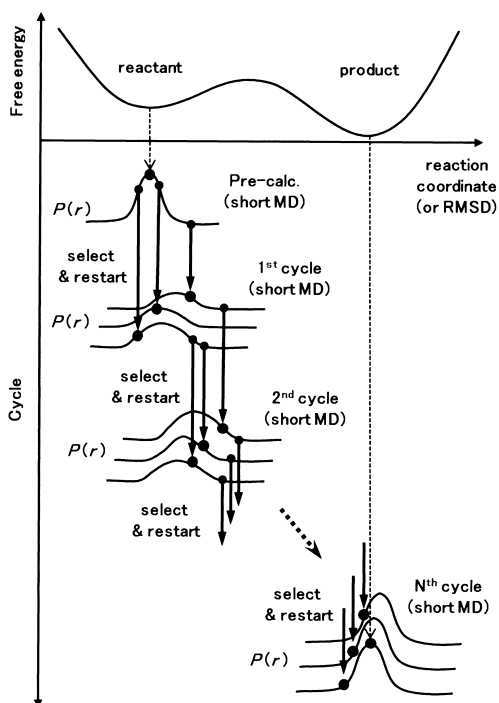


図 2. PaCS-MD の手順.

多レプリカを用いて自由エネルギー地形を計算する方法として、Weighted Ensemble (WE) (Huber and Kim, 1996; Zhang et al., 2010) や Forward Flux Sampling (FFS) (Allen et al., 2006, 2009) がある。WE や FFS も 構造既知の 2 つの状態間の経路をバイアスなしで探索する手法であるが、これらの手法や PaCS-MD では、レプリカの作成方法とレプリカ間の統計的な位置付けに違いがある。WE では、前のレプリカのトラジェクトリの最後の構造を使用して新しいレプリカを作成しており、FFS では、前のレプリカのトラジェクトリから、始状態と終状態の間を分割する特定の面上にある構造がランダムに選ばれ、新しいレプリカとして使用される。一方、PaCS-MD ではより終状態に近い構造を複数選んで新しいレプリカを作成している。これらの違いにより、終状態への収束は、PaCS-MD の方がより速いと期待される。また、レプリカ間の統計的位置付けは、WE では各レプリカ間の統計的重みが明確であるのに対し、FFS や PaCS-MD ではレプリカ間の関係が明確ではない。そのため、PaCS-MD では後述の MSM を用いてトラジェクトリの解析を行っている。

2.3 MSM

本節では MSM の理論と統計的なモデルについて概説し、速度論との関係を述べる。MSM の具体的な手順の解説は 2.4 節で行う。

ある特定の温度で平衡状態にある系のダイナミクスを考える。この系では全ての構造空間が、重ならない領域、または、離散的な状態に分割できる、と仮定する。もし、時刻 $t = 0, \tau, 2\tau, \dots$ の離散時間でこの系の発展を観察したなら、一連の系の発展は、系が訪れた状態間の遷移で表すことができる。観測された状態は、時間を離散化した確率過程の 1 つの状態と見なすことが

できる。この過程がマルコフ連鎖で記述される場合、ある時刻に各状態を観察する確率は直前の状態にのみ依存する。\$L\$ 個の状態を持つ定常状態過程の場合、この過程は観察時間間隔、いわゆる lagtime \$\tau\$ にのみ依存する \$L \times L\$ の遷移行列 \$\mathbf{T}(\tau)\$ で特徴づけられる。遷移行列の要素 \$T_{ji}(\tau)\$ は時刻 \$(t-\tau)\$ で状態 \$i\$ におり、時刻 \$t\$ で状態 \$j\$ にいる遷移確率を表している。もしこの過程が詳細釣り合いを満たしているなら、\$T_{ji}p_i^{\text{eq}} = T_{ij}p_j^{\text{eq}}\$ を満たす。ただし、\$p_i^{\text{eq}}\$ は系が平衡状態で状態 \$i\$ にいる確率を表す。

時刻 \$t\$ で \$L\$ 個の状態を占める確率を表すベクトルを \$\mathbf{p}(t)\$ で表す。もし、初期状態を表す確率ベクトルが \$\mathbf{p}(0)\$ で与えられたなら、時間 \$n\tau\$ 後の確率ベクトルは

$$(2.1) \quad \mathbf{p}(n\tau) = \mathbf{T}(n\tau)\mathbf{p}(0) = [\mathbf{T}(\tau)]^n \mathbf{p}(0)$$

と書ける。これはチャップマン-コルモゴロフ方程式である。

この過程は、連続時間で現象論的速度定数 \$\mathbf{K}\$ によっても特徴づけられる。ただし、\$\mathbf{K}\$ の要素 \$K_{ji} (i \neq j)\$ は状態 \$i\$ から状態 \$j\$ への非負の現象論的速度を表す。行の和がゼロになるように、対角項は \$K_{jj} = -\sum_{i \neq j} K_{ji}\$ から決められる。時間発展は方程式

$$(2.2) \quad \dot{\mathbf{p}}(t) = \mathbf{K}\mathbf{p}(t)$$

で支配される。ここで、\$\dot{}\$ は時間微分を表す。この時間発展方程式は形式解

$$(2.3) \quad \mathbf{p}(t) = \exp(\mathbf{K}t) \mathbf{p}(0)$$

をもつ。ただし、指数項は形式的な行列指数関数である。式(2.2)はマスター方程式とよばれ、連続時間内の離散化された状態間の発展を表す。\$\mathbf{p}(t)\$ は形式的には全ての \$t\$ で定義されるが、我々が考えている系では、式(2.3)が全ての \$t\$ で満たされるとは限らない。実際、有限の構造空間ではダイナミクスがマルコフ的である時間解像度には限界があり、この時間解像度より短い時間スケールで起こる過程はマスター方程式に従わないこともある。

遷移行列 \$\mathbf{T}(\tau)\$ と速度行列 \$\mathbf{K}\$ との間には式(2.1), (2.3)の比較から明確な関係、

$$(2.4) \quad \mathbf{T}(\tau) = \exp(\mathbf{K}\tau).$$

がある。もしこの過程が全ての時間で連続時間マルコフ過程によって特徴づけられるなら、この過程は一致する遷移行列によって同等に記述される。逆は、遷移行列 \$\mathbf{T}(\tau)\$ のサンプリング誤差によって常に成り立つとは限らないが、観測データや詳細釣り合い条件、非負の速度といった条件を用いて速度行列 \$\mathbf{K}\$ を復元する方法も考案されている (Grubmüller and Tavan, 1994; Sriraman et al., 2005)。

遷移行列 \$\mathbf{T}(\tau)\$ と速度行列 \$\mathbf{K}\$ は、それぞれ、\$\mu_k(\tau)\$ と \$\lambda_k\$ の固有値を持ち、共通の右固有ベクトル \$\mathbf{u}_k\$ をもつ。詳細釣り合い条件から全ての固有値は実数であることが保証される。ここで、互いの固有値は降順に並んでいるとすると、\$\mu_k(\tau)\$ と \$\lambda_k\$ は次の関係にある、

$$(2.5) \quad \mu_k(\tau) = \exp(\lambda_k \tau).$$

それぞれの固有値は、implied time scale とよばれる時間スケール

$$(2.6) \quad \tau_k = -\lambda_k^{-1} = -\tau / \ln \mu_k(\tau)$$

をもち、関連する固有ベクトルはこの時間スケールに応じた構造遷移の情報を与える。

2.4 マルコフモデルの構築

MSM では、マルコフ過程に従う分子の状態空間はミクロな状態に離散化され、これらの状態

間の遷移確率, または, 遷移速度が MD シミュレーションのデータから計算される. 以下に, MSM でのマルコフモデルの構築方法を簡単に述べる. 詳細は文献 (Prinz et al., 2011; Pan and Roux, 2008; Chodera et al., 2007) を参照されたい.

分子の状態空間の離散化には, クラスタリングのアプローチを用いる. MD トラジェクトリのクラスタリングを行い, ミクロな状態を定義する. クラスタリングの方法には, K-Means 法や K-Centers 法などがよく使われるが, 結果の解析がしやすいように, クラスタ中心間の距離の最小値を与えることで, クラスタ中心間の距離がほぼ等しくなるようにクラスタリングする方法も提案されている. 次に, ミクロな状態に MD シミュレーションから得られる構造を割り当てる. MD シミュレーションから系の時間発展が得られているので, ミクロな状態間の遷移を数えることができる. 得られた遷移数から遷移行列 $T(\tau)$ を, 例えば, 最尤法を用いて推定する (Prinz et al., 2011; Chodera et al., 2007).

系がマルコフ過程として記述できるなら, 遷移行列 T が得られたとき, 時刻 t での状態 $p(t)$ と時刻 $(t + \tau)$ での状態 $p(t + \tau)$ との関係は,

$$(2.7) \quad p(t + \tau) = T p(t)$$

で表される. 十分に時間が経過して, 系が平衡状態に達したとき, 上式は,

$$(2.8) \quad \pi = T \pi$$

で表される. ただし, π は平衡状態での確率分布を表し, $\sum_i \pi_i = 1$ である. 式 (2.8) は, 平衡状態の確率分布が固有値 1 の固有ベクトルに相当することを表している. クラスタリングで得られたミクロな状態が孤立することなくつながっている場合, 最も大きい固有値は 1 であり, 縮退していない. このことは, 定常状態がただ 1 つ存在することと同値である.

また, 平衡状態の自由エネルギーは平衡状態の確率分布から得られる,

$$(2.9) \quad F_i = -k_B T \ln \frac{\pi_i}{\max_j \pi_j}.$$

ただし, この式は, 最も確率分布が大きい値の自由エネルギーを 0 にしている. 得られた自由エネルギー地形から構造変化の時間スケールといったキネティックスの計算は, Mean First Passage Time (Singhal et al., 2004), Transition Path Theory (Noé et al., 2009), String Method (Pan and Roux, 2008), Milestoning (Schütte et al., 2011) 等の手法を組み合わせた先行研究がある. また, 得られる遷移行列を固有値と固有ベクトルに分解し, 解析することで, 関係する準安定状態やゆっくりとしたダイナミクスの過程とその時間スケール, フォールディング過程にあらわれる状態間の遷移経路に関する知見などが得られる. 詳しくは, 文献 (Noé et al., 2009, 2011) を参照されたい.

2.5 計算手順

既に述べた PaCS-MD と MSM を組み合わせた自由エネルギー計算全体の手順は, 図 3 のようになる. まず, PaCS-MD による構造のサンプリングを行う. 前計算によって初期構造を作成し, 短い MD シミュレーションを実行する. 得られた構造から構造のランキングを行い, 構造の選択を行う. 例えば, 目的とする構造に近い RMSD の値をもつ構造を M 個選ぶ. これら M 個の構造を初期構造にして新たにマクスウェル分布に従うように初期速度を与え, M 個の短い MD シミュレーションを行う. さらに, 構造ランキングを行い, 再び目的とする構造に近い RMSD の値をもつ構造を M 個選ぶ. これを, 目的の構造に十分近くなるまで繰り返す. MSM 解析に十分なサンプルを得るために, 我々は PaCS-MD を複数回行っている. PaCS-MD の各トラジェクトリは τ より十分長い必要がある.

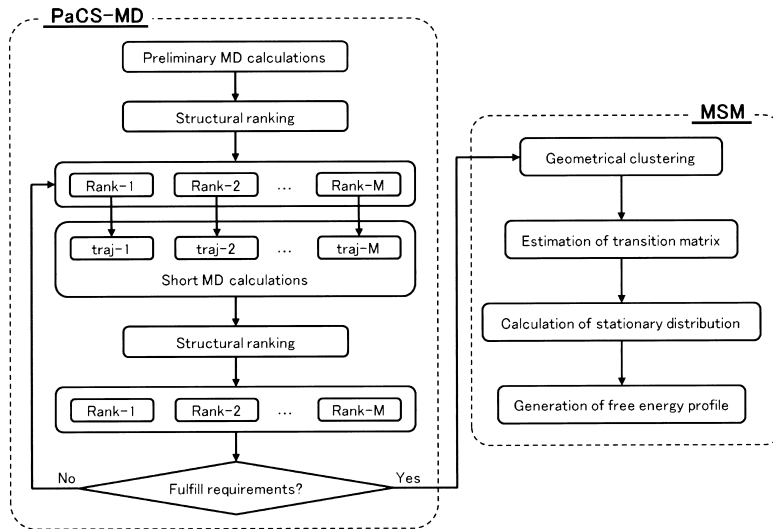


図 3. PaCS-MD + MSM を用いた自由エネルギー計算の手順.

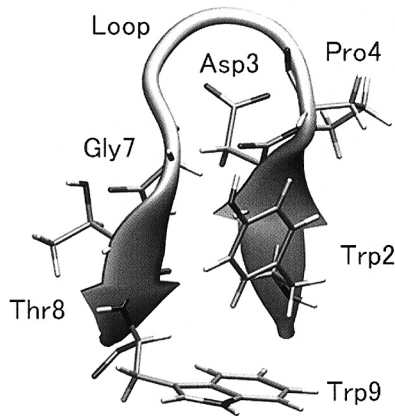


図 4. Chignolin の天然構造 (PDBid : 1UAO).

目的の構造に十分近い構造が得られたら、MSM を用いた解析に入る。PaCS-MD で作成した全ての MD トラジェクトリ (もしくは、反応座標空間に射影した値) を用いてクラスタリングを行う。遷移行列の推定後、定常状態の確率分布を求めて、自由エネルギー計算を行う。Lagtime τ は、式 (2.6) を用いて、implied time scale が収束した値を用いる。

3. 自由エネルギー計算の例

前述の PaCS-MD と MSM を用いて自由エネルギーを計算する手法を、小タンパク質の水中でのフォールディング過程に適用した例を紹介する。ここで扱ったタンパク質は 10 個のアミノ酸残基からなる chignolin である (図 4)。アミノ酸配列は、Gly-Tyr-Asp-Pro-Glu-Thr-Gly-Thr-Trp-Gly

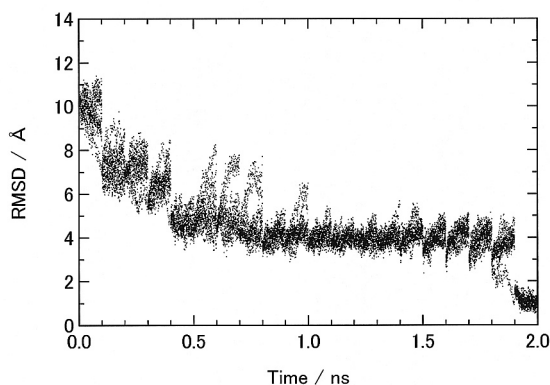


図 5. PaCS-MD での終状態への RMSD の収束性. 図は典型的な 1 セットの PaCS-MD の結果. 20 サイクルの計算を行い, 各サイクルには 10 本の 0.1-ns MD シミュレーションが含まれている.

で与えられる.

PaCS-MD では, はじめに, 伸長した chignolin の構造を用いて前計算を行い, 得られた構造の中からランダムに 20 個を選び PaCS-MD の初期構造とした. 次に, 20 セットの PaCS-MD を行い, 各 PaCS-MD で 20-cycle \times 10-run \times 100-ps MD を行った. 構造のランキングには, フォールドした chignolin を参照構造とした C_{α} 原子の RMSD を使用した.

MSM では, EMMA software package (Senne et al., 2012) を用い, クラスター中心間の最小距離が 0.25 Å となるようにクラスタリングを行った. ただし, 100 ps のトラジェクトリのうち, 最初の 20 ps は非平衡な部分を含むため, MSM を用いた解析には使用していない. Chignolin のフォールディング過程を調べるために, 3 つの水素結合距離, HB1 (Asp3:N-Thr8:O) と HB2 (Asp3:N-Gly7:O), HB3 (Asp3:O-Gly7:N) を反応座標とする空間に射影した MD トラジェクトリを用いた. これらの水素結合距離はフォールド・ミスフォールドの構造を表す指標として先行研究で用いられている (Sato et al., 2006; Harada and Kitao, 2011, 2013). HB3 の形成はフォールディングのトリガーとなっており, HB1 の形成は天然構造へのフォールディングに, HB2 の形成はミスフォールディングに関連している. MSM での解析には, $HB3 \leq 4$ Å を満たすトラジェクトリのみを使用した.

まず, PaCS-MD の終状態への収束性を見てみる. 図 5 に 1 回の PaCS-MD における C_{α} RMSD の「時間変化」を示す. サイクルが進むにつれて, RMSD の値が天然構造の値 (= 0) に近づいていることがわかる. また, RMSD の値が広く分布していることから, PaCS-MD では構造空間を広くサンプルしていることがわかる.

図 6 で chignoling の 2 つの水素結合距離の分布と自由エネルギー地形を示す. 図 6(a) からは, PaCS-MD で得られる分布は大きく広がっており, PaCS-MD が構造空間を広くサンプリングしていることがわかる. 図 6(b) では, 平衡状態の自由エネルギー地形を示してある. $(HB1, HB2) = (6.0, 2.2)$ 付近に自由エネルギー最小の構造があり, これが天然構造である. また, $(HB1, HB2) = (2.4, 5.8)$ に自由エネルギー極小の構造があり, ミスフォールドした構造を示している.

これらの構造や自由エネルギー地形上での位置は, これまでの研究とよく一致している. しかし, ミスフォールド付近の自由エネルギー地形は先行研究と若干異なっている. これは, PaCS-

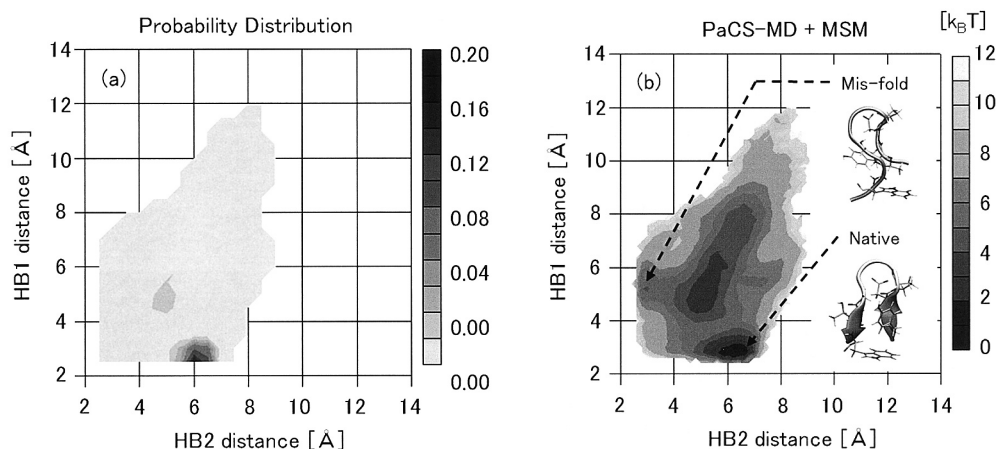


図 6. PaCS-MD と MSM を用いた計算結果. (a) PaCS-MD で得られたトラジェクトリから計算した確率密度分布. (b) PaCS-MD + MSM により作成した自由エネルギー地形とエネルギー極小状態に対応する構造.

MD の目的構造を天然構造としたため、自由エネルギー地形上で天然構造から離れたミスフォールド構造付近をうまくサンプリングできていなかったからだと考えられる。より精度の高い自由エネルギー地形を描くためには、ミスフォールド状態の構造を目的の構造として本手法を繰り返せばよい。

MSM で求めた定常状態の確率密度分布と transition path theory (E and Vanden-Eijnden, 2006) を用いて、天然構造とミスフォールド構造間の遷移速度の計算をおこなった。その結果、ミスフォールドから天然構造への遷移速度は 7.6 ns、逆は、30 ns と見積もられた。2つの構造間遷移の時間スケールは実験では明らかにされていないが、MD 計算を用いた見積もりから数百 ns のオーダーだと考えられている (Suenaga et al., 2007)。今回求めた遷移速度は、それより 1-2 オーダー小さい。ミスフォールド構造付近のサンプリングの少なさに加え、解析に使用したデータは $HB3 \leq 4 \text{ \AA}$ のみのため、構造は常にコンパクトな形をしており、状態間遷移の速度が速くなる傾向にあると考えられる。

ここで示した例では、1つの MD トラジェクトリの長さは 100 ps であり、全計算時間は 400 ns である。このように、非常に短い計算時間で自由エネルギー地形を描くことのできる今回紹介した手法は、自由エネルギー計算の強力な手法である。

4. おわりに

タンパク質が機能を発揮する時間スケールは広範囲にわたるため、MD シミュレーションを用いた構造空間内のサンプリングには限界がある。以前に比べて計算機の能力が向上しているとはいえ、自由エネルギー地形を描くことには、いまだ計算上の困難が伴う。今後も、計算機の能力向上はもちろんのこと、計算手法のさらなる発展が不可欠である。

本稿では、タンパク質の構造遷移と自由エネルギー計算という観点から、PaCS-MD と MSM を組み合わせて自由エネルギーを効率的に計算するという新たな手法を紹介した。バイアス等を使用しないことでパラメータの最適化を省き、少ない前計算でサンプリングを行える PaCS-MD を用いて大量に構造を発生させることで、自由エネルギーの計算に必要な構造空間の広いサンプリングを可能とし、PaCS-MD で得られたデータを MSM で解析することで、平衡状態の自

由エネルギー地形が得られることを解説した。さらに本手法を chignolin フォールディング過程に適用して自由エネルギー地形を作成することで、結果の妥当性および本手法の高効率性を示した。

最後に、本手法の応用可能性にも言及しておきたい。

例えば、タンパク質と小分子であるリガンドとの複合体構造探索(ドッキング)において、結合自由エネルギーの計算をする場合、タンパク質とリガンドの配向がどのように変わるかで結合自由エネルギーが大きく変わってしまう。現在、タンパク質とリガンドのドッキングでは、大量の複合体候補構造を作成し、スコア関数で自由エネルギーの値を評価する方法が広く使われているが、多くのドッキングソフトでは、計算時間の制約からタンパク質の構造を固定して結合自由エネルギーを計算するため、その精度は必ずしも高くはない。多くの場合、実験で得られる最安定構造は複合体候補構造の中に含まれるが、スコア関数で計算した最小結合自由エネルギー値をもつ構造が実験で得られる構造と一致するとは限らない。そこで、複合体候補構造を目的構造として PaCS-MD を行い、MSM で解析することで、タンパク質の動きを取りいれた、より精度の高い結合自由エネルギーを計算できると考えられる。

このように、本手法は、本稿で紹介したフォールディング過程以外にも適用できる可能性を有している。本手法を様々な事例に応用させることによって、これまで未解明の点が多いタンパク質のフォールディング過程の解明や、創薬につながるタンパク質と結合しやすいリガンドの発見・作成などが進むと期待される。今後、応用事例の開拓を早期に実現していきたい。

参 考 文 献

- Allen, R. J., Frenkel, D. and ten Wolde, P. R. (2006). Simulating rare events in equilibrium or nonequilibrium stochastic systems, *The Journal of Chemical Physics*, **124**, 024102.
- Allen, R. J., Valeriani, C. and ten Wolde, P. R. (2009). Forward flux sampling for rare event simulations, *Journal of Physics: Condensed Matter*, **21**, 463102.
- Chodera, J. D., Singhal, N., Pande, V. S., Dill, K. A. and Swope, W. C. (2007). Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics, *The Journal of Chemical Physics*, **126**, 155101.
- E, W. and Vanden-Eijnden, E. (2006). Toward a theory of transition paths, *Journal of Statistical Physics*, **123**, 503–523.
- Grubmüller, H. and Tavan, P. (1994). Molecular dynamics of conformational substates for a simplified protein model, *The Journal of Chemical Physics*, **101**, 5047.
- Harada, R. and Kitao, A. (2011). Exploring the folding free energy landscape of a β -Hairpin miniprotein, chignolin, using multiscale free energy landscape calculation method, *The Journal of Physical Chemistry B*, **115**, 8806–8812.
- Harada, R. and Kitao, A. (2013). Parallel Cascade Selection Molecular Dynamics (PaCS-MD) to generate conformational transition pathway, *The Journal of Chemical Physics*, **139**, 035103.
- Huber, G. A. and Kim, S. (1996). Weighted-ensemble Brownian dynamics simulations for protein association reactions, *Biophysical Journal*, **70**, 97–110.
- Hukushima, K. and Nemoto, K. (1996). Exchange Monte Carlo method and application to spin glass simulations, *Journal of the Physical Society of Japan*, **65**, 1604–1608.
- Israelewitz, B., Baudry, J., Gullingsrud, J., Kosztin, D. and Schulten, K. (2001a). Steered molecular dynamics investigations of protein function, *Journal of Molecular Graphics and Modeling*, **19**, 13–25.
- Israelewitz, B., Gao, M. and Schulten, K. (2001b). Steered molecular dynamics and mechanical functions

- of proteins, *Current Opinion in Structural Biology*, **11**, 224–230.
- Jensen, M. O., Prk, S., Tajkhorshid, E. and Schulten, K. (2002). Energetics of glycerol conduction through aquaglyceroporin GlpF, *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 6731–6736.
- Laio, A. and Parrinello, M. (2002). Escaping free-energy minima, *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 12562–12566.
- Nishihara, Y., Hayashi, S. and Kato, S. (2008). A search for ligand diffusion pathway in myoglobin using a metadynamics simulation, *Chemical Physics Letters*, **464**, 220–225.
- Noé, F., Schutte, C., Vanden-Eijnden, E., Reich, L. and Weikl, T. R. (2009). Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations, *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 19011–19016.
- Noé, F., Doose, S., Daidone, I., Lollmann, M., Sauer, M., Chodera, J. D. and Smith, J. C. (2011). Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments, *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 4822–4827.
- Pan, A. C. and Roux, B. (2008). Building Markov state models along pathways to determine free energies and rates of transitions, *The Journal of Chemical Physics*, **129**, 064107.
- Prinz, J., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., Chodera, J. D., Schutte, C. and Noé, F. (2011). Markov models of molecular kinetics: Generation and validation, *The Journal of Chemical Physics*, **134**, 174105.
- Satoh, D., Shimizu, K., Nakamura, S. and Terada, T. (2006). Folding free-energy landscape of a 10-residue mini-protein, chignolin, *FEBS Letters*, **580**, 3422–3426.
- Schütte, C., Noé, F., Lu, J., Sarich, M. and Vanden-Eijnden, E. (2011). Markov state models based on milestoning, *The Journal of Chemical Physics*, **134**, 204105.
- Senne, M., Trendelkamp-Schroer, B., Mey, A. S. J. S., Schutte, C. and Noé, F. (2012). EMMA: A software package for Markov model building and analysis, *Journal of Chemical Theory and Computation*, **8**, 2223–2238.
- Singhal, N., Snow, C. D. and Pande, V. S. (2004). Using path sampling to build better Markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper beta hairpin, *The Journal of Chemical Physics*, **121**, 415.
- Sriraman, S., Kevrekidis, I. G. and Hummer, G. (2005). Coarse master equation from Bayesian analysis of replica molecular dynamics simulations, *The Journal of Physical Chemistry B*, **109**, 6479–6484.
- Suenaga, A., Narumi, T., Futatsugi, N., Yanai, R., Ohno, Y., Okimoto, N. and Taiji, M. (2007). Folding dynamics of 10-residue β -hairpin peptide Chignolin, *Chemistry-An Asian Journal*, **2**, 591–598.
- Sugita, Y. and Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding, *Chemical Physics Letters*, **314**, 141–151.
- Torrie, G. M. and Valleau, J. P. (1977). Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling, *Journal of Computational Physics*, **23**, 187–199.
- Zhang, B. W., Jasnow, D. and Zuckerman, D. M. (2010). The “weighted ensemble” path sampling method is statistically exact for a broad class of stochastic processes and binning procedures, *The Journal Chemical Physics*, **132**, 054107.

Cascade-type Massive Parallel Simulation for Protein Conformational Transition Pathway Search

Yasutaka Nishihara¹, Ryuhei Harada² and Akio Kitao¹

¹Institute of Molecular and Cellular Bioscience, The University of Tokyo

²Advanced Institute for Computational Science, RIKEN

Free energy landscape calculation and analysis play important roles in elucidating the relationship between protein conformation and function. Calculating the free energy landscape of proteins by using molecular dynamics simulations often requires long calculation time. This is because many degrees of freedom in proteins generate a large number of conformational substates and proteins are frequently confined in few substates. To explore protein conformational space widely and efficiently, a novel free energy calculation method is proposed as the combination of Parallel Cascade Selection Molecular Dynamics (PaCS-MD) and Markov State Model (MSM). This method is applied to folding free energy landscape analysis of a small protein as a test case.