

ストリング法によるタンパク質構造変化解析

松永 康佑[†]

(受付 2014 年 1 月 31 日; 改訂 8 月 26 日; 採択 8 月 29 日)

要 旨

タンパク質の大規模構造変化は細胞内での重要な化学反応の制御に関わっている。しかしながら、それらの構造変化はおよそミリ秒で起こる稀なイベント(レアイベント)であるため、全原子モデルを用いた通常の分子動力学シミュレーションでサンプリングすることが難しい。本稿では、レアイベントであるタンパク質構造変化パスウェイを効率よく探索する手法であるストリング法を解説する。特に、ストリング法を応用する際のプラクティカルな問題が統計数理の手法で解決し得ることを説明し、最後に酵素タンパク質であるアデニル酸キナーゼの構造変化へ応用した研究を紹介する。

キーワード：レアイベント, 最小自由エネルギー経路, ストリング法, 自由エネルギー計算, アデニル酸キナーゼ。

1. はじめに

近年の計算機ハードウェア・ソフトウェアの急速な進歩によって、分子動力学(Molecular Dynamics; MD)シミュレーションは、生体分子研究において理論・実験を補完する有力なアプローチのひとつとなっている。特に、全原子モデルを用いたMDシミュレーションは、しばしば計算顕微鏡(computational microscope)とも呼ばれるように、タンパク質等の生体分子のダイナミクスを原子レベルで観察することができる強力な手法である。タンパク質は細胞内で一つの構造周りを揺らいでいるわけではなく、いくつかの安定構造間をダイナミックに構造遷移する。そしてこれらの大規模な構造変化が、酵素反応やシグナル伝達といった細胞内の重要な化学反応の制御に関わっていることが知られている(Fuchigami et al., 2011)。そこで、全原子MDシミュレーションを用いてタンパク質構造変化のメカニズムをマイクロなレベルから解明したいという欲求に駆られるが、現在の通常のMDシミュレーションで達成できるシミュレーション時間がたかだかマイクロ秒(10^{-6} 秒)なのに対し、タンパク質の構造変化はおよそミリ秒(10^{-3} 秒)で起こるといふ稀なイベント(レアイベント)であるため、通常の全原子MDシミュレーションで構造変化を研究することは未だ困難な状況である。

そこでこれまでにMDシミュレーション分野では、タンパク質の構造空間を効率よくサンプリングするための手法が開発されてきた。例えば、拡張アンサンブル法のひとつであるマルチカノニカル法(Nakajima et al., 1997)や、レプリカ交換分子動力学法(Sugita and Okamoto, 1999)は、現在のMDシミュレーション研究で頻繁に用いられる強力な手法である。一方で最近、構造変化前と後の2つの安定構造が事前に分かっているとして(実際これらの構造は、異なる条件下でのX線結晶構造解析やNMRで事前に特定されていることが多い)、その間を繋ぐ構造変化

[†]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

パスウェイのサンプリングに特化した手法が研究・開発されている(藤崎, 2011). これら一連の手法はパスサンプリングと総称され, 通常のMDシミュレーションのようにニュートン方程式に基づいて時間積分を行うのではなく, 作用(action)や他の物理量に基づいて構造変化パスウェイをまとめて動かすところに特徴がある. 構造変化パスウェイは多数のレプリカ系で離散的に近似された上で, それらレプリカ系をまとめて並列に動かすので, 並列計算機環境で効率よくパスウェイをサンプリングすることができる.

本稿では, パスサンプリング法の代表的手法であるストリング法(E et al., 2002; Maragliano et al., 2006)について解説する. ストリング法は, 構造変化における典型的なパスウェイである最小自由エネルギー経路(minimum free energy path)と呼ばれるものを探索する手法である. まず次章では, 構造変化パスウェイに関する理論を簡単に紹介し, 最小自由エネルギー経路とはどのようなものかを解説する. 次に, ストリング法のアルゴリズムについて解説した後で, 実際に応用するにあたってのプラクティカルな問題とその解決法について説明する. 最後に, ストリング法を酵素タンパク質の構造変化へ応用した著者らの研究(Matsunaga et al., 2012)について紹介する. なおストリング法にはいくつかのバリエーションが存在するが, 本稿で解説する方法は, 現在の生体分子シミュレーション研究で用いられることの多いString Method in Collective Variable Space(Maragliano et al., 2006)と呼ばれるものである.

2. 最小自由エネルギー経路

本章では, タンパク質が2つの構造間, 例えば不活性状態から活性状態へ大規模に構造変化する際の, 構造変化パスウェイに関する理論を簡単に紹介する. 一般にタンパク質の運動は, 分子内部の複雑な相互作用や溶媒からの揺動によって確率的なものになる. したがって, もし2つの構造間の遷移トラジェクトリをミクロな原子レベルで観察すると, 多様な遷移パスウェイが観測されるだろう. しかし一方で, 適切に選ばれたタンパク質の部分自由度へそれらミクロな遷移トラジェクトリを射影したならば, それらの遷移は一本もしくは数本のドミナントなパスウェイ付近を通過すると期待される. もしそれら射影された遷移トラジェクトリが形作るトラジェクトリ集団のチューブの半径が十分に小さいならば, そのドミナントなパスウェイは構造遷移を代表すると言ってよいだろう. 本章では, このドミナントなパスウェイを, Most Probable Pathway(MPP; 最も実現される確率の高いパスウェイ)として数学的に定式化し, いくつかの仮定の下でそれが自由エネルギー地形上の最小自由エネルギー経路で近似されることを示す. 本稿の主題であるストリング法はこの最小自由エネルギー経路を求める方法であり, そのアルゴリズムについては次章で紹介する.

タンパク質とその周りの溶媒も含めた N 原子系のデカルト座標 $\mathbf{x} \in \mathbb{R}^{3N}$ を考える. 系がポテンシャルエネルギー関数 $U(\mathbf{x})$ を感じて運動しているとする. 温度 T における \mathbf{x} の平衡分布はボルツマン分布となる,

$$(2.1) \quad p(\mathbf{x}) = e^{-\beta U(\mathbf{x})} / Z.$$

ここで, $\beta = 1/(k_B T)$ はボルツマン定数 k_B と温度 T の積の逆数であり, $Z = \int d\mathbf{x} e^{-\beta U(\mathbf{x})}$ は座標に関する分配関数(configurational partition function)である.

現在のタンパク質MDシミュレーション研究では, 溶媒も含めた総原子数 N として少なくとも1万個, 最大で1千万個という大規模なサイズが使用される. しかしながら, 我々が興味があるのはそれら全ての自由度ではなく, 構造変化に関わるタンパク質の部分自由度空間であることが多い. そこで, 注目する部分自由度を $\mathbf{z} = \{z_1, z_2, \dots, z_n\}$ ($n \ll N$) とし, これを collective variable と呼ぶことにする. 元々のデカルト座標 \mathbf{x} から collective variable への射影を $\mathbf{z}'(\mathbf{x})$ と

すると, collective variable z が平均的に感じる実効的なエネルギー (自由エネルギーまたは平均力ポテンシャルと呼ばれる) $W(z)$ は以下となる,

$$(2.2) \quad W(z) = -\beta^{-1} \ln Z^{-1} \int dx \delta(z - z'(x)) e^{-\beta U(x)}.$$

ここで, $\delta(\cdot)$ は Dirac のデルタ関数である.

以下では collective variable z の運動を近似して表現し, collective variable 空間における構造変化の MPP を定式化する. まず, collective variable z の運動の特徴的な時間スケールが他の自由度に比べて非常に遅いと仮定する. その時, collective variable z の運動は, 自由エネルギー $W(z)$ 上の拡散現象として近似することができる. 具体的には, z 以外の自由度の緩和の時間スケールよりも長く, $W(z)$ が一定として近似できる短い時間スケールを Δt とすると, $z(t)$ から $z(t + \Delta t)$ への時間発展方程式は以下となる (Ermak and McCammon, 1978),

$$(2.3) \quad z_i(t + \Delta t) = z_i(t) + \sum_{j=1}^n \left(-\beta D_{ij}(z(t)) \frac{\partial W(z(t))}{\partial z_j} + \frac{\partial D_{ij}(z(t))}{\partial z_j} \right) \Delta t + R_i(t).$$

ここで $D(z)$ は拡散テンソル (diffusion tensor) と呼ばれる量である. $D(z)$ の対角項 D_{ii} は z_i が感じる摩擦係数の逆数に対応し, 非対角項 D_{ij} は z_i と z_j 間のタンパク質内部自由度を介した相互作用や, 溶媒を介した流体力学的相互作用等を非等方的な拡散として繰り込んで表現している. z が変化するとともにタンパク質内部の相互作用や周りの溶媒の配置も変化するので, 一般に拡散テンソル $D(z)$ は z に依存する. $R(t)$ はガウス熱雑音 (Gaussian thermal noise) であり, $\langle R_i(t) \rangle = 0$ と $\langle R_i(t) R_j(t) \rangle = 2D_{ij} \Delta t$ を満たす.

次に collective variable 空間における構造変化パスウェイを表現するために, 媒介変数 s ($0 \leq s \leq 1$) を用いた曲線 $z(s)$ を定義しよう. 曲線 $z(s)$ は, 構造変化前の始状態に対応する $z(0)$ と, 構造変化後の終状態である $z(1)$ を繋ぐ曲線である. この時, MPP が満たすべき性質を, 「パスウェイ上の点 $z(s')$ を初期条件として時間発展させた時に, Δt 後において最も実現されやすい値が同じパスウェイに乗っている」とする. Δt 後に最も実現されやすい z は, 式 (2.3) から求めることができる. 式 (2.3) では, R_i がガウシアンであるので, $R_i = 0$ が最も実現されやすい値となる. したがって, $z(s)$ が MPP であるときに満たすべき式は以下となる (Pan et al., 2008),

$$(2.4) \quad z_i(s) = z_i(s') + \sum_{j=1}^n \left(-\beta D_{ij}(z(s')) \frac{\partial W(z(s'))}{\partial z_j} + \frac{\partial D_{ij}(z(s'))}{\partial z_j} \right) \Delta t.$$

Δt が十分に小さいならば, 式 (2.4) は, 右辺第二項が MPP の接線成分に比例していることを意味している. すなわち, 右辺第二項のパスウェイに直交した成分がゼロとなる. したがって, パスウェイの直交成分を $[\cdot]^\perp$ としてベクトル表記を用いると, 式 (2.4) は以下と等価となる,

$$(2.5) \quad [-\beta D(z) \nabla W(z) + \nabla \cdot D(z)]^\perp = 0.$$

以上の式 (2.4), (2.5) が MPP の定義であり, collective variable z の運動が他の自由度に比べて非常に遅いというシンプルな仮定のみを使って得られた. しかしながら, 式 (2.4), (2.5) の拡散テンソル $D(z)$ とその微分の計算は困難であるため, 更に大胆な近似を行って数値計算で取り扱いきやすい形にする必要がある. 以下では, 2つの仮定を使って更に近似を行い, 数値計算で取り扱える形である最小自由エネルギー経路として定式化する.

まず, 始状態と終状態の間の自由エネルギーバリアの高さに比べて熱エネルギー $k_B T$ が小さいと仮定し, 式 (2.5) の $-\beta D(z) \nabla W(z)$ と $\nabla \cdot D(z)$ のどちらが大きな値を持つかを見てみる. 拡散テンソル $D(z)$ は, Einstein の関係により $k_B T (= \beta^{-1})$ に比例するため, 変数変換を行って

$k_B T$ に依らない量 $M(z) = \beta D(z)$ を使って式(2.5)を書き直す,

$$(2.6) \quad [-M(z)\nabla W(z) + k_B T \nabla \cdot M(z)]^\perp = 0.$$

自由エネルギーバリアの高さに比べて熱エネルギー $k_B T$ が小さい場合, 式(2.6)の $\nabla W(z)$ に比べ, $k_B T \nabla \cdot M(z)$ は小さな値になることが期待される. したがって, 式(2.6)を以下のように近似する (Maragliano et al., 2014),

$$(2.7) \quad [-M(z)\nabla W(z)]^\perp = 0.$$

最後の問題は, 計算困難な量である $M(z) = \beta D(z)$ を如何に扱いやすい形で近似するかである. $D(z)$ は様々な相互作用を繰り込んだ統計量であるため, 数値計算でも正しく評価することは困難である. また仮に計算が可能だとしても, $D(z)$ の評価のために計算量が膨大になっては元も子もない. そこで最もシンプルな近似として, 元々のデカルト座標 x を collective variable 空間へ射影する際のスケール変換の影響だけを考慮して, $M(z)$ を計量テンソル (metric tensor) $M^{\text{metric}}(z)$ で置き換える (Johnson and Hummer, 2012). 元々のデカルト座標 x から collective variable への射影を $z'(x)$ とすると, 計量テンソル $M^{\text{metric}}(z)$ は, 計量の $z'(x) = z$ における期待値 (カノニカル集団平均) として以下のように定義される (Maragliano et al., 2014),

$$(2.8) \quad M_{ij}^{\text{metric}}(z) = e^{\beta W(z)} Z^{-1} \int dx \sum_{k=1}^{3N} \frac{1}{m_k} \frac{\partial z'_i}{\partial x_k} \frac{\partial z'_j}{\partial x_k} e^{-\beta U(x)} \delta(z'(x) - z) \\ = \left\langle \sum_{k=1}^{3N} \frac{1}{m_k} \frac{\partial z'_i}{\partial x_k} \frac{\partial z'_j}{\partial x_k} \right\rangle_{z'(x)=z}.$$

ここで, $\langle \cdot \rangle_{z'(x)=z}$ は, $z'(x) = z$ におけるカノニカル集団平均を意味する. Collective variable への射影 $z'(x)$ が線形の場合, 計量テンソル $M^{\text{metric}}(z)$ は定数となり z に依らない (更にアンサンブル計算も不要になる). しかし一般的には, タンパク質主鎖に沿った二面角等の曲がった座標系へ射影することが多く, 計量テンソル $M^{\text{metric}}(z)$ は z の関数となる. 計量テンソルの計算に必要な微分 $\partial z'_i / \partial x_k$ は, 多くの MD シミュレーションコードが束縛ルーチン等で行っている計算であり, コードを少し改変するだけで計量テンソル計算へ再利用できる.

以上, 式(2.7)において $M(z)$ に計量テンソル $M^{\text{metric}}(z)$ を用いたものが最小自由エネルギー経路の定義であり, 以下で書き表される (Maragliano et al., 2014),

$$(2.9) \quad [-M^{\text{metric}}(z)\nabla W(z)]^\perp = 0.$$

直感的には, もし $M^{\text{metric}}(z)$ が単位行列ならば, 式(2.9)は最小自由エネルギー経路が自由エネルギー地形の等高線に直交する最急降下パスウェイと等価であることを示している. 式(2.9)が MPP の近似であることを導出するために以下の2つの仮定を用いた: (1) 始状態と終状態間の自由エネルギーバリアの高さに比べて熱エネルギー $k_B T$ が小さい. (2) 拡散テンソル $D(z)$ がスケール変換のみを考慮した計量テンソル $M^{\text{metric}}(z)$ で近似できる. このうち仮定(1)については, 多くのタンパク質構造変化の自由エネルギーバリアの高さが $10k_B T$ のオーダーであり, 熱エネルギー $k_B T$ に比べて大きいことを考えると妥当な近似だと思われる. 一方で仮定(2)は大胆な近似であり, 近似の妥当性について現在も研究が行われている. 例えば Maragliano らは, 小分子系 (alanine dipeptide) において, 特に自由エネルギーバリア領域で最小自由エネルギー経路が MPP を良く近似することを示している (Maragliano et al., 2014). 一方で Johnson と Hummer は, 非等方的な拡散テンソル $D(z)$ を持つ2次元系 (Muller potential 系) を使って, MPP と最小自由エネルギー経路が一致しない場合があることを細かく検証している (Johnson

and Hummer, 2012). 他に, 複数本の短いトラジェクトリを発生させて $D(z)$ をダイレクトに評価する手法も提案されているが (Pan et al., 2008), 計算時間との兼ね合いから現在のところ応用研究で使われることは少ない. これらの先進的な話題は本稿の範囲を超えるので割愛する. 興味のある読者は文献を参照されたい.

3. ストリング法

3.1 ストリング法のアルゴリズム

本節では, 前章の式 (2.9) で定義される最小自由エネルギー経路を効率良く探索するストリング法のアルゴリズムを紹介する. まずおおまかにストリング法とは, collective variable z の値を全原子レベルの MD シミュレーションの情報を使って動かし, 今度は z を追って全原子 MD シミュレーションが移動するということを繰り返す, いわばマルチスケールのアルゴリズムである. したがって, collective variable z と全原子 MD シミュレーションは, 互いに情報をやりとりさせながらも, それぞれは別個の系として動かしていくことに注意されたい.

前章と同様に, collective variable 空間における構造変化パスウェイを媒介変数 s ($0 \leq s \leq 1$) を用いた曲線 $z(s)$ で表す. そして数値計算にのせるために $z(s)$ を離散化して, R 個の点 $\{z_\alpha\}$ ($\alpha = 1, \dots, R$) で近似する. 以下ではこの各 z_α をイメージと呼ぶことにする. イメージ z_1 は構造変化前の始状態, イメージ z_R は構造変化後の終状態に対応する.

式 (2.9) から, パスウェイ $z(s)$ が最小自由エネルギー経路であるとき, $-M^{\text{metric}}(z(s))\nabla W(z(s))$ のパスウェイに対する直交成分はゼロとなる. したがって最小自由エネルギー経路の探索では, パスウェイ $z(s)$ の離散近似であるイメージ $\{z_\alpha\}$ をそれぞれ $[-M^{\text{metric}}(z_\alpha)\nabla W(z_\alpha)]^\perp$ がゼロに収束するまで動かしていくことが基本的アルゴリズムとなる. ここでの $-M^{\text{metric}}(z_\alpha)\nabla W(z_\alpha)$ の推定には, 各イメージ z_α 周りで全原子 MD シミュレーションを行う. しかしながら何度もイメージを動かしていくと, 最初はパスウェイ上に等間隔に配置されていたイメージが次第にバラバラとなり, パスウェイとしての近似が破綻してしまう. そこでストリング法では, 一回イメージを動かした後で, パスウェイを連続的な曲線で補間し, その曲線上でイメージが等間隔になるよう再配置するというステップを追加する. 以上をまとめたストリング法のアルゴリズムは以下となる (Maragliano et al., 2006) :

- (1) 各イメージ z_α 周りで全原子 MD シミュレーションを行い, $M^{\text{metric}}(z_\alpha)$ と $-\nabla W(z_\alpha)$ を推定する
- (2) 各イメージ z_α を $[-M^{\text{metric}}(z_\alpha)\nabla W(z_\alpha)]^\perp$ にしたがって動かす
- (3) イメージ $\{z_\alpha\}$ を平滑化し, パスウェイの補間とイメージの再配置を行う
- (4) (1) から (3) をイメージ $\{z_\alpha\}$ が動かなくなるまで繰り返す

ステップ (1) の全原子 MD シミュレーションでは, 通常用いる分子力場のポテンシャルエネルギー $U(\mathbf{x})$ に対して z_α を中心とする調和的な束縛 (harmonic restraint) を加えた以下のポテンシャル関数 $U_\alpha(\mathbf{x})$ の下でシミュレーションを行う,

$$(3.1) \quad U_\alpha(\mathbf{x}) = U(\mathbf{x}) + \frac{\kappa}{2} |\mathbf{z}'(\mathbf{x}) - z_\alpha|^2.$$

ここで, κ はバネ定数であり, $\mathbf{z}'(\mathbf{x})$ はデカルト座標 \mathbf{x} から collective variable への射影である. 全原子 MD シミュレーションから得られるサンプルを \mathbf{x}_t ($t = 1, \dots, T$) とすると, $M^{\text{metric}}(z_\alpha)$ と $-\nabla W(z_\alpha)$ の推定量は以下となる (Maragliano et al., 2014) :

$$(3.2) \quad \hat{M}_{ij}^{\text{metric}}(z_\alpha) = \frac{1}{T} \sum_{t=1}^T \sum_{k=1}^{3N} \frac{1}{m_k} \frac{\partial z'_i(\mathbf{x}_t)}{\partial x_k} \frac{\partial z'_j(\mathbf{x}_t)}{\partial x_k},$$

$$-\frac{\widehat{\partial W}(z_\alpha)}{\partial z_i} = \frac{\kappa}{T} \sum_{t=1}^T (z'_i(\mathbf{x}_t) - z_{\alpha,i}).$$

これらは $\kappa \rightarrow \infty$ の極限で不偏推定量となる．したがってストリング法では，アンブレラサンプリングなどの通常の束縛付き MD シミュレーションに比べて大きな値の κ が用いられる．例えば著者らが行った酵素タンパク質の応用研究においては，アンブレラサンプリングで $\kappa = 0.1 \text{ kcal}/(\text{mol } \text{\AA}^2)$ を用い，ストリング法で $\kappa = 10 \text{ kcal}/(\text{mol } \text{\AA}^2)$ を用いた．

ステップ(2)では，適当に小さな Δt を選び，各イメージ z_α を以下のように動かす，

$$(3.3) \quad z_\alpha \leftarrow z_\alpha + [-\hat{M}^{\text{metric}}(z_\alpha) \widehat{\nabla W}(z_\alpha)]^\perp \Delta t.$$

直交成分 $[\cdot]^\perp$ の抽出計算では，まずパスウェイの接線方向を一次近似で $(z_\alpha - z_{\alpha\pm 1}) / |z_\alpha - z_{\alpha\pm 1}|$ として求め，その後 $[-\hat{M}^{\text{metric}}(z_\alpha) \widehat{\nabla W}(z_\alpha)]$ の接線方向への射影を取り除く操作を行う．接線成分は，次のステップ(3)の再配置でも近似的に取り除かれると期待されるため，文献によっては直交成分を抽出せずに $[-\hat{M}^{\text{metric}}(z_\alpha) \widehat{\nabla W}(z_\alpha)] \Delta t$ で動かすものもある(例えば Maragliano et al., 2014)．

ステップ(3)ではまず，イメージ移動の際の統計誤差をなめすために簡単な Low-pass フィルタ

$$(3.4) \quad z_\alpha \leftarrow (1-h)z_\alpha + \frac{h}{2}(z_{\alpha-1} + z_{\alpha+1}),$$

を用いて平滑化する． h ($0 \leq h \leq 1$) は平滑化の強さを決めるパラメータであり，経験的には $h = 0.1$ 付近の値が用いられる．その後イメージ間の距離が等間隔になるように再配置を行う．再配置では，隣り合ったイメージ間を区分線形近似してパスウェイを補間した後で，その補間パスウェイの全長を計算し，等間隔になるようイメージを置き直すという操作を行う．

以上のアルゴリズムのステップ(1)は各イメージ毎に独立に計算できる．例えば並列計算機環境上では，ステップ(1)の束縛付き全原子 MD シミュレーションを異なる計算ノードで並列に走らせることができる．ステップ(2)と(3)は，隣接するイメージの座標や距離の情報が必要なために計算ノード間の集団通信が必要となるが，通常 collective variable z の次元 n は高々 $10 \sim 1000$ のオーダーであることが多いので通信コストはほとんど無視できる．したがって，ストリング法は最近の大規模並列計算機にマッチした効率的な並列アルゴリズムといえる．

3.2 応用にあたってのプラクティカルな話題

前節ではストリング法のアルゴリズムの基本部分を解説した．しかし，実際にストリング法をタンパク質の構造変化へ応用する場合，幾つかのプラクティカルな問題が存在する．具体的には，collective variable z をどう選ぶべきだろうか？ また，最小自由エネルギー経路が MPP を近似しているか否かを如何にして検証すればよいのだろうか？ 得られたパスウェイに沿った自由エネルギー変化や統計量を如何に計算するのか？ といった問題が挙げられる．本節では，これらを克服するために著者らが提案した幾つかの方法を解説する．

まず，タンパク質の構造変化の場合，collective variables z としてどの部分自由度を選ぶべきだろうか？ これはストリング法で一番難しい問題で，MPP の導出で仮定した「他の自由度に比べて運動の時間スケールが遅い」部分自由度をタンパク質のような複雑系において事前に知ることは難しい．ストリング法では collective variable 空間の次元が大きくなっても計算量は発散しないので，とにかく構造変化に関係のありそうな自由度を片っ端から collective variables とする，というアプローチも考えられるが，collective variable 空間の次元が大きくなりすぎると，

イメージを等間隔に再配置する際に誤って不安定な箇所配置してしまったり、(一般にエネルギー地形の凸凹は次元とともに爆発的に増えるので)イメージがエネルギー地形のローカルなミニマムにトラップされてしまう危険性がある。そこで著者らは、事前に始状態と終状態まわりで全原子 MD シミュレーションを行った上で主成分分析(北尾, 2001)による次元削減を行い、得られた主成分を collective variables とすることを提案している(Matsunaga et al., 2012)。主成分は揺らぎの大きな座標なので自由エネルギー地形が滑らかであることが期待できると同時に、これまでの研究からタンパク質のドメイン運動(まとまった剛体的部位に分かれて動く運動)に関わる遅い自由度を抽出できることが知られている。また、元々のデカルト座標 x から主成分への射影は線形(第 i 主成分は $z_i = \sum_k c_{i,k}(x_k - \langle x_k \rangle)$)なので、計量テンソル $M^{\text{metric}}(z)$ が定数となり、式(3.2)第1式の計算を省略することができる。例えば、タンパク質の不斉炭素原子(C_α 原子)のデカルト座標の主成分を collective variable z とすると、計量テンソルは z に依らない対角行列

$$(3.5) \quad M^{\text{metric}} = \text{diag} \left(\frac{1}{m_{C_\alpha}}, \dots, \frac{1}{m_{C_\alpha}} \right),$$

となる。ここで、 m_{C_α} は不斉炭素原子の質量である。

選んだ collective variable が適切だったか否かはストリング法計算が終了した後で検証することができる。まず、得られた最小自由エネルギー経路が MPP を良く近似しているかどうかは、最小自由エネルギー経路に沿ったサンプリング(アンブレラサンプリング等)の結果が実験データと整合的であるかを調べることで分かる。もし MPP を良く近似しているならば、他のマイナーなパスウェイを無視してそのパスウェイ付近のサンプリングデータだけを採用しても実験データと合うことが期待できる。実験データとしては、FRET(Fluorescence/Förster Resonance Energy Transfer)計測の結果を使うことができる。FRET 計測では、タンパク質の2箇所の残基にラベルした(波長の異なる)蛍光色素分子の蛍光発光を測定し、そこから残基間距離の分布情報を得ることができる。そこで、ストリング法で得られた最小自由エネルギー経路パスウェイ付近のサンプリングデータから後に解説する多状態ベネット受容比法(Multistate Bennett Acceptance Ratio; MBAR, Shirts and Chodera, 2008)を適用することで、同じ残基間の距離の分布を計算し比較することができる。FRET 計測について詳しくは、本特集内の岡本氏の論文を参照されたい。また、collective variable が遷移状態(構造変化前と後の状態を分ける分割面)を捉えているかの検証も重要である。これは最小自由エネルギー経路上の自由エネルギーのバリア付近において束縛無しの全原子 MD シミュレーションを何本か走らせ、五分五分の確率で行き先が分かれるか否かを調べることで検証できる(コミッターテストと呼ばれる, Bolhuis et al., 2002)。

これらの検証のためには、パスウェイに沿った自由エネルギー変化の計算が必要となる。以下では、その計算法について解説する。ストリング法では、一般に 10~1000 次元の collective variable 空間中での曲がったパスウェイが手に入るが、そのパスウェイに沿った自由エネルギー変化をどのように計算できるだろうか? ナイーヴに考えると、ストリング法のアルゴリズムのステップ(1)で既に自由エネルギーの勾配 $\nabla W(z_\alpha)$ を推定しているのだから、 z_α と z_1 との間の自由エネルギー差はパスウェイに沿った線積分の形で

$$(3.6) \quad W(z_\alpha) - W(z_1) \sim \sum_{\alpha'=1}^{\alpha-1} \widehat{\nabla W}(z_{\alpha'}) \cdot (z_{\alpha'+1} - z_{\alpha'}),$$

と推定できる。しかし、タンパク質サイズの巨大系においては $\widehat{\nabla W}(z_{\alpha'})$ の推定誤差と離散化による誤差が積分とともに積み重なり無視できないオーダーになってしまう。実際、例えば著者らが行った酵素タンパク質の応用研究においては、蓄積した誤差が数 kcal/mol 以上におよん

でしまった．そこで著者らは，ストリング法によるイメージ $\{z_\alpha\}$ の時間発展を止めて，各イメージ z_α 周りのアンブレラサンプリングを行い，以下で解説する多状態ベネット受容比法を適用する計算法を提案している (Matsunaga et al., 2012)．

以下，厳密性を持たせるために collective variable z_α の揺らぎも考慮に入れて，collective variable 空間を $\{z_\alpha\}$ を母点としてポロノイ分割した上での各ポロノイセル B_α の自由エネルギー $W(B_\alpha)$ を評価することにする．通常，アンブレラサンプリングの後処理としてはヒストグラム法の一つである Weighted Histogram Analysis Method (WHAM; Kumar et al., 1992) が用いられることが多い．しかしながら WHAM は，各サンプル点を所属するセルの中心(今の場合はイメージ z_α) で置き換える操作をするため， z_α が高次元の場合には推定精度が悪くなる(パスウェイが曲がっているので接線方向のみに束縛をかけることが困難であるのと，高次元空間ではサンプル点がセルの中心に集まらずその周りのシェルに集まってしまうのが理由である)．以下で解説する多状態ベネット受容比法では，セルの中心で置き換える操作をせずに，各サンプル点毎にカノニカル集団平均を求めるための重みを計算するのでこのような問題を回避できる．

アンブレラサンプリングでは，ストリング法と同様に z_α を中心とする調和的な束縛を加えた式(3.1)のポテンシャル関数 $U_\alpha(x)$ の下で全原子 MD シミュレーションを行う(ただしストリング法の時よりは弱いバネ定数 κ を用いる)．各アンブレラ系 α ($\alpha = 1, \dots, R$) から T_α 点のサンプルが得られたとして，まとめて $\{\mathbf{x}_{\alpha t}\}_{t=1}^{T_\alpha}$ と表す．多状態ベネット受容比法では，各アンブレラ系の(無次元化された)自由エネルギー $f_\alpha = -\ln \int d\mathbf{x} \exp[-\beta U_\alpha(\mathbf{x})]$ の推定量 \hat{f}_α を，以下の式を自己無撞着に解いて求める (Shirts and Chodera, 2008)：

$$(3.7) \quad \hat{f}_\alpha = -\ln \frac{\sum_{\kappa=1}^R \sum_{t=1}^{T_\kappa} \frac{\exp[-\beta U_\alpha(\mathbf{x}_{\kappa t})]}{\sum_{\gamma=1}^R T_\gamma \exp[\hat{f}_\gamma - \beta U_\gamma(\mathbf{x}_{\kappa t})]}}{\sum_{\kappa=1}^R \sum_{t=1}^{T_\kappa} \frac{\exp[-\beta U_\alpha(\mathbf{x}_{\kappa t})]}{\sum_{\gamma=1}^R T_\gamma \exp[\hat{f}_\gamma - \beta U_\gamma(\mathbf{x}_{\kappa t})]}}$$

この推定量 \hat{f}_α は，サンプル数 $T_\alpha \rightarrow \infty$ ($\alpha = 1, \dots, R$) の極限で不偏であり，かつ有効性(分散が一番小さい推定量という意味)を持つ漸近有効推定量である．この推定法では，各 \hat{f}_α の推定に全てのアンブレラサンプリングのサンプルデータを使うので，式(3.6)の線積分のように誤差が蓄積するという問題は生じない．

各アンブレラ系の自由エネルギーの推定値 \hat{f}_α を求めたら，任意の物理量 $A(\mathbf{x})$ に対して，分配関数 $Z = \int d\mathbf{x} \exp[-\beta U(\mathbf{x})]$ と $Z_A = \int d\mathbf{x} A(\mathbf{x}) \exp[-\beta U(\mathbf{x})]$ を考え，それぞれサンプル点数ゼロとして式(3.7)を“外挿”すると $A(\mathbf{x})$ のカノニカル集団平均 $\langle A(\mathbf{x}) \rangle$ の推定値 \hat{A} が求まる：

$$(3.8) \quad \hat{A} = \frac{\hat{Z}_A}{\hat{Z}} = \sum_{\alpha=1}^R \sum_{t=1}^{T_\alpha} w_{\alpha t} A(\mathbf{x}_{\alpha t}) = \sum_{m=1}^M w_m A(\mathbf{x}_m).$$

式(3.8)の2行目では，簡便のために α と t を，ひとつの記号で m でまとめた ($M = \sum_{\alpha=1}^R T_\alpha$)．ここで， w_m は，カノニカル集団平均を求めるための各サンプル点の重みとして解釈できる．

ポロノイセル B_α の自由エネルギー $W(B_\alpha)$ を求めるためには，サンプル点がポロノイセル B_α にいるときにだけ1をとり他は0をとるような指示関数(indicator function) $\chi_\alpha(z'(\mathbf{x}))$ の期待値をとって，そのセルの確率 p_α を計算する ($z'(\mathbf{x})$ は \mathbf{x} から collective variable 空間への射影)：

$$(3.9) \quad \hat{p}_\alpha = \sum_{m=1}^M w_m \chi_\alpha(z'(\mathbf{x}_m)).$$

したがって，自由エネルギーの推定値は $\hat{W}(B_\alpha) = -\beta^{-1} \ln(\hat{p}_\alpha)$ となる．指示関数の計算ではポロノイセルの分割面を陽に計算する必要はなく，単に事前に各サンプル点の射影 $z'(\mathbf{x}_m)$ と全ポロノイセルの母点 $\{z_\alpha\}$ との距離を計算し，一番距離の近いポロノイセルのインデックス α を割り当てておけばよい．このときの計算量は，collective variable の次元が大きくても発散はし

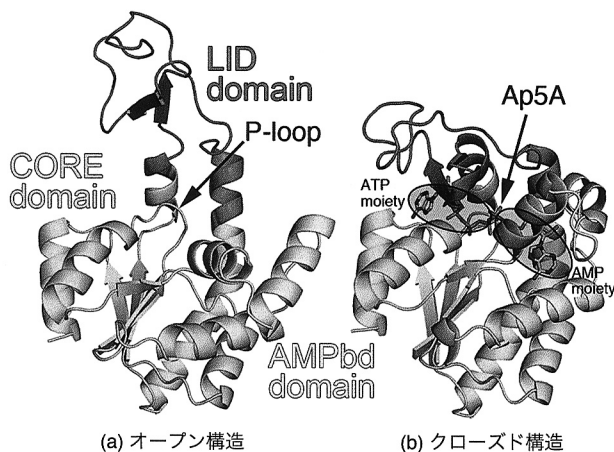


図 1. アデニル酸キナーゼの結晶構造. (a) 基質無しオープン構造(PDB ID: 4ake)と, (b) 基質が結合しているクローズド構造(PDB ID: 1ake). Matsunaga et al. (2012) から改変・転載.

ない.

指示関数として, ボロノイセルの代わりに FRET 計測でラベルされる 2 つの残基間距離をピン化する関数を用いると, その残基間距離の確率分布・自由エネルギー変化を計算することができる. これによって FRET 計測データと比較することが可能となる.

更に, 各ボロノイセル B_α 内における $A(x)$ のカノニカル集団平均の推定値 \hat{A}_α を, 条件付き確率を考慮して以下のように計算することができる (Matsunaga et al., 2012) :

$$(3.10) \quad \hat{A}_\alpha = \frac{\sum_{m=1}^M w_m \chi_\alpha(z'(x_m)) A(x_m)}{\sum_{m=1}^M w_m \chi_\alpha(z'(x_m))} = \frac{\sum_{m=1}^M w_m \chi_\alpha(z'(x_m)) A(x_m)}{\hat{p}_\alpha}.$$

式(3.10)は, 各ボロノイセル内の統計量(構造や揺らぎ, 相互作用など)がパスウェイに沿ってどのように変化するかを調べるのに便利な式である. 例えば著者らが行った酵素タンパク質の応用研究においては, collective variable 以外の自由度も含めたタンパク質全体の平均構造・揺らぎが, パスウェイに沿ってどのように変化するかを式(3.10)を使って調べた.

4. アデニル酸キナーゼへの応用

本章では, 酵素タンパク質であるアデニル酸キナーゼ (Adenylate Kinase; AK, 214 残基) の構造変化に対してストリング法を応用した著者らの研究を紹介する (Matsunaga et al., 2012). AK は, ATP と AMP から, ADP を 2 分子生成するリン酸基転移反応 $ATP + AMP \rightleftharpoons 2ADP$ を可逆的に触媒する酵素であり, 細胞内のエネルギーバランスを調節することで細胞の恒常性維持に寄与している. 図 1 に示した AK の結晶構造からは, 基質 (ATP と AMP) が結合していない場合はオープン構造, 結合している場合はクローズド構造をとり溶媒から隔離した環境でリン酸基転移反応を触媒する, と考えられている. では, どのようなメカニズムでこのような構造変化が起きるのか? より具体的には, 基質結合はどのようにして構造変化を引き起こすのか? この問いに答えるため, 著者らは基質が結合していない系 (apo-AK) と結合している系 (holo-AK) の 2 つの系に対してストリング法を応用し, 両者のオープン・クローズド構造の構造変化パスウェイを比べることで, 構造変化が引き起こされるメカニズムを調べた.

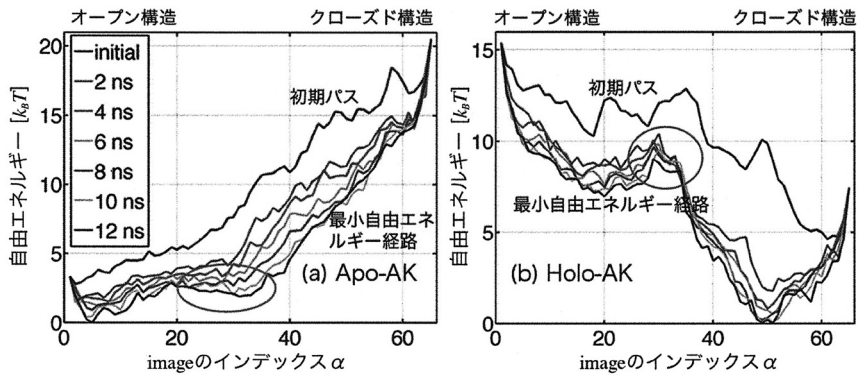


図2. スtring法で得られたパスに沿った自由エネルギー変化. (a) Apo-AK. (b) Holo-AK. Matsunaga et al.(2012)から改変・転載.

全原子 MD シミュレーションでは、タンパク質に AMBER ff03 力場、溶媒に TIP3P 水分子モデルを使用した。基質分子には ATP と AMP が繋がったアナログ分子 (Ap5A) を用いた。Apo-AK, holo-AK ともに溶媒を含めておよそ 62,000 原子のサイズとなった。まず collective variable を定義するために、apo-AK についてオープン構造・クローズド構造それぞれの周りで 3 ns の束縛無し全原子 MD シミュレーションを行い、両者を混ぜたトラジェクトリの不斉炭素原子のデカルト座標に対して主成分分析を行った。その結果、第 1 主成分は構造変化の方向、それ以外の主成分は、結晶構造で観測される 2 つのドメイン部分 (ATP が結合するドメインである LID ドメインと、AMP が結合する AMPb ドメイン) の動きに対応していた。AK は主に、この LID ドメイン・AMP ドメインと、それ以外の部分 (CORE ドメイン) が剛体的に相対運動して構造変化すると考えられる。したがって collective variable として使う主成分の数は、少なくとも $6 \times 3 - 6 = 12$ 個必要であると予想される。そこで、ここでは第 1~第 20 主成分までを collective variable として定義して String 法計算を行った。初期パスウェイは、クローズド構造からオープン構造へ、collective variable 空間内を線形に繋ぐパスウェイ上を 1 ns かけて引っ張って作成した。その後、初期パスウェイを離散化して 65 個のイメージを作成し、各イメージ周りの全原子モデル構造を平衡化後に String 法計算を行った。String 法計算では、結晶構造に相当する両端点 (イメージ z_1 をオープン構造、イメージ z_{65} をクローズド構造とした) は動かさずに固定した。また計算中、主成分分析で求めた平均構造周りで、タンパク質全体の並進・回転 (構造変化に無関係な外部運動) を止めるための拘束を課した。String 法計算が自由エネルギー地形上の局所的な凸凹にトラップされていないかを検証するために、最初の String 法計算で得られた最小自由エネルギー経路上のドメインの動きを確認した後で、敢えてそれとは異なるドメイン運動をする初期パスウェイを作成して再度 String 法計算を行い、同様のドメイン運動に収束することを確認した。以上の計算には、理研・次世代生命体統合シミュレーションソフトウェアの研究開発プロジェクトにおいて寺田透氏 (東大) を中心に開発された MD ライブラリである μ^2 lib (Terada et al., 2012) を使用した。計算資源には、理研のスーパーコンピュータである RIKEN Integrated Cluster of Clusters (RICC) の約 500 コアを用いた。

図 2 は apo-AK と holo-AK の両者について、探索された構造変化パスウェイに沿った自由エネルギー変化を示している (式 (3.9) を使用)。最もエネルギーが高いパスウェイが初期パスウェイ、エネルギーが収束しているパスウェイが最小自由エネルギー経路、その間は String 法計算

が収束するまでに得られた途中のパスウェイである。結晶構造から予想されたように、apo-AK ではオープン構造が安定なのに対し、holo-AK ではクローズド構造が安定であることがわかる。特に、最小自由エネルギー経路上においては、apo-AK で部分的にクローズした構造で準安定状態が存在し(丸で囲んだ箇所)、また holo-AK で自由エネルギーバリア(丸で囲んだ箇所)が現れている。まず著者らは、この最小自由エネルギー経路が MPP に一致しているかを検証するために、最小自由エネルギー経路周りのサンプリングデータから、Henzler-Wildman et al. (2007) が行った FRET 計測でラベルされた 2 残基間の距離を軸として自由エネルギー変化を求めた。その結果、定性的なレベルではあるが FRET 実験の結果と一致しており、見つかった最小自由エネルギー経路が MPP であることが支持された。以下では、この最小自由エネルギー経路周りのサンプリングデータを解析することで判明した構造変化の詳細を紹介する。

Apo-AK で見つかった部分的にクローズした準安定状態では何が起きているのだろうか？式 (3.10) を用いて、最小自由エネルギー経路に沿った AK の平均構造とその周りの構造揺らぎの変化を解析した結果、準安定状態においては LID ドメインと CORE ドメインを繋ぐヒンジ領域の揺らぎが増大していることが分かった。この局所部分の揺らぎの増大は cracking と呼ばれ、エントロピーゲインによりエネルギーを下げるメカニズムとして粗視化モデル研究で提案されている (Miyashita et al., 2003)。相互作用を調べた結果、揺らぎが大きくなる原因は ASP118-LYS136 間の salt bridge が壊れるためであることが分かった。一方で、holo-AK では ASP118-LYS136 のちょうど反対側(ヒンジが折れる内側)の残基と (Ap5A の) ATP 部位が相互作用しており、ヒンジが折れて LID がクローズするのを誘導していた。これらの結果から著者らは、Apo-AK における cracking した準安定状態は、基質が結合したときに相互作用を組み替えやすくし、LID ドメインをスムーズにクローズさせるための状態であると解釈した。

一方 holo-AK においては、自由エネルギーのバリア部分と、(Ap5A の) AMP 部位の AMPb ドメインへの結合が関係していることが分かった。しかしまず相互作用の詳細を調べる前に、本当にバリアが遷移状態に対応しているかを検証するために、バリア付近において東縛無しの全原子 MD シミュレーションを何本か走らせ五分五分の確率で行き先が分かれるかをチェックするテスト(コミッターテスト)を行った。バリア付近のイメージ $\{z_\alpha | \alpha = 32, 33, 34\}$ 周辺から、それぞれ 10 ns の東縛無しの MD を 40 個計算した後の分布が図 3 左である。バリアの前後によって 10 ns 後の分布が分かれており、確かに遷移状態となっていることが分かる。その後、AMP 部位と AK との相互作用を調べた結果、 $\alpha = 32$ から $\alpha = 34$ の間に少しだけ AMPb ドメインが動くだけで、AMP 部位と AMPb ドメインとの間に水素結合が 3 つ形成されることが判明した(図 3 右)。この 3 つの相互作用は全て AMP にとって特異的な相互作用である。つまり AMPb ドメインの場合、構造変化の早い段階で基質特異性が効いており、かつ構造変化を誘導していることが分かった。一方で、LID ドメインが完全にクローズするのは AMPb ドメインの後であり、ATP の結合部位である P-loop 付近に水分子が長く留まることも一因となり、ATP に対する基質特異性が現れるのは構造変化の最終段であることが分かった。これは、AMP の濃度を増やした時に、誤って AMP が LID ドメインへ結合し AK の機能が阻害されるという実験研究のシナリオ (Sinev et al., 1996) とも整合的である。以上の観察をもとに、LID ドメインに関しては 2 段階の構造変化モデル、AMPb ドメインに関しては誘導適合(induced-fit)に則った構造変化モデルを提案した。より詳細な内容については Matsunaga et al. (2012) を参照されたい。

5. まとめ

本稿では、レアイベントであるタンパク質構造変化のパスウェイに関する理論を簡単に紹介し、構造変化パスウェイがいくつかの仮定の下で最小自由エネルギー経路で代表されることを

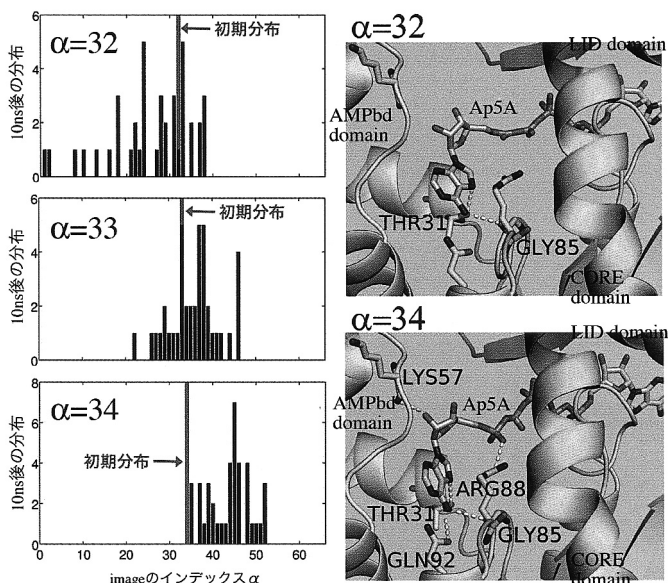


図 3. コミッターテストの結果. (左)Holo-AK においてバリア付近のイメージ $\{z_\alpha | \alpha = 32, 33, 34\}$ 周辺からそれぞれ束縛無し of 全原子 MD シミュレーションを 40 個ながし, 10 ns 後の構造をイメージを母点とするボロノイセルへ射影した分布. (右)バリアの前後における基質 (Ap5A) とタンパク質の相互作用の変化の様子. バリアの前後で相互作用が 3 つ増減していることを示している. Matsunaga et al. (2012) から改変・転載.

説明した. そして, 最小自由エネルギー経路を効率良く探索する手法であるストリング法のアルゴリズムを解説した. ストリング法のアルゴリズムはシンプルで実装も容易であるが, 実際に応用しようとすると幾つかのプラクティカルな問題があり, それを解決し得るのが主成分分析や多状態ベネット受容比法といった統計数理の手法であることを述べた.

ストリング法は, 構造変化のトラジェクトリ集団が一本ないしは数本の MPP 付近に集まっている, という前提の下で MPP を近似する最小自由エネルギー経路上の構造変化を主にエネルギーの観点から特徴付ける. 本稿では触れなかったが, 最近では構造変化をより動的な観点から特徴付けるための理論的なフレームワークが整備されてきている (Best and Hummer, 2005). 例えば Transition Path Theory (TPT) は, 平衡状態のアンサンブルの中から構造変化に対応するトラジェクトリのみを抜き出し, その分布や確率流を取り扱う統計理論を展開している (Vanden-Eijnden, 2006). 最近では TPT を理論的基盤として, ストリング法に動的要素を付加した新たなパスウェイ探索法が開発されたり (Vanden-Eijnden and Venturoli, 2009), タンパク質フォールディングシミュレーションのデータからどの構造がどの程度フォールディングしやすいかという動的な指標を与える研究が行われている (Noé et al., 2009). この方面に関して興味のある読者は文献を参照されたい.

MD シミュレーションの手法・データ解析では, 本稿で紹介した内容の他にも様々な未解決の統計数理の問題が残されている. 本稿との関連に限っても, 例えばボロノイセル間の遷移確率をアンブレラサンプリングデータ等から推定することができれば (Maragliano et al., 2009), そこから構造変化のレートを推定し, 単分子 FRET 計測で得られるレートと比較することが可能となる. また, 重回帰分析や判別分析等を上手く使えばパスに沿った自由エネルギー変化の

原因となる重要な残基間相互作用を特定できるかもしれない。本稿の内容が、ミクロな複雑系の統計数理の問題に少しでも興味を持っていただくきっかけになれば幸いである。

謝 辞

共同研究者である藤崎弘士氏(日本医科大)、寺田透氏(東大)、木寺詔紀氏(横浜市大)に感謝致します。また杉田有治氏(理研)と、原稿を読んでいただいた小林千草氏(理研)、小室靖明氏(中央大)、狩野康人氏(理研)に併せて感謝致します。理研の基礎科学特別研究員制度のサポートに感謝致します。

参 考 文 献

- Best, R. B. and Hummer, G. (2005). Reaction coordinates and rates from transition paths, *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6732–6737.
- Bolhuis, P. G., Chandler, D., Dellago, C. and Geissler, P. L. (2002). Transition path sampling: Throwing ropes over rough mountain passes, in the dark, *Annual Review of Physical Chemistry*, **53**, 291–318.
- E, W., Ren, W. and Vanden-Eijnden, E. (2002). String method for the study of rare events, *Physical Review B*, **66**, 052301.
- Ermak, D. L. and McCammon, J. A. (1978). Brownian dynamics with hydrodynamic interactions, *The Journal of Chemical Physics*, **69**, 1352–1360.
- Fuchigami, S., Fujisaki, F., Matsunaga, Y. and Kidera, A. (2011). Protein functional motions: Basic concepts and computational methodologies, *Advances in Chemical Physics*, **145**, 35–82.
- 藤崎弘士(2011). 生体分子におけるパスサーチおよびパスサンプリングについて, 日本医科大学基礎科学紀要, **40**, 83–98.
- Henzler-Wildman, K. A., Thai, V., Lei, M., Ott, M., Wolf-Watz, M., Fenn, T., Pozharski, E., Wilson, M. A., Petsko, G. A., Karplus, M., Hübner, C. G. and Kern, D. (2007). Intrinsic motions along an enzymatic reaction trajectory, *Nature*, **450**, 838–844.
- Johnson, M. E. and Hummer, G. (2012). Characterization of a dynamic string method for the construction of transition pathways in molecular reactions, *The Journal of Physical Chemistry B*, **116**, 8573–8583.
- 北尾彰朗(2001). 主成分分析を使って眺めた蛋白質のエネルギー地形, 統計数理, **49**, 43–56.
- Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. and Kollman, P. A. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, *Journal of Computational Chemistry*, **13**, 1011–1021.
- Maragliano, L., Fischer, A., Vanden-Eijnden, E. and Ciccotti, G. (2006). String method in collective variables: Minimum free energy paths and isocommittor surfaces, *The Journal of Chemical Physics*, **125**, 24106.
- Maragliano, L., Roux, B. and Vanden-Eijnden, E. (2009). Free energy and kinetics of conformational transitions from Voronoi tessellated milestoning with restraining potentials, *Journal of Chemical Theory and Computation*, **5**, 2589–2594.
- Maragliano, L., Roux, B. and Vanden-Eijnden, E. (2014). Comparison between mean forces and swarms-of-trajectories string methods, *Journal of Chemical Theory and Computation*, **10**, 524–533.
- Matsunaga, Y., Fujisaki, H., Terada, T., Furuta, T., Moritsugu, K. and Kidera, A. (2012). Minimum free energy path of ligand-induced transition in adenylate kinase, *PLOS Computational Biology*, **8**, e1002555.
- Miyashita, O., Onuchic, J. N. and Wolynes, P. G. (2003). Nonlinear elasticity, proteinquakes, and the

- energy landscapes of functional transitions in proteins, *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 12570–12575.
- Nakajima, N., Nakamura, H. and Kidera, A. (1997). Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides, *The Journal of Physical Chemistry B*, **101**, 817–824.
- Noé, F., Schütte, C., Vanden-Eijnden, E. and 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.
- Pan, A. C., Sezer, D. and Roux, B. (2008). Finding transition pathways using the string method with swarms of trajectories, *The Journal of Physical Chemistry B*, **112**, 3432–3440.
- Shirts, M. R. and Chodera, J. D. (2008). Statistically optimal analysis of samples from multiple equilibrium states, *The Journal of Chemical Physics*, **129**, 124105.
- Sinev, M. A., Sineva, E. V., Ittah, V. and Haas, E. (1996). Towards a mechanism of AMP-substrate inhibition in adenylate kinase from *Escherichia coli*, *FEBS Letters*, **397**, 273–276.
- Sugita, Y. and Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding, *Chemical Physics Letters*, **314**, 141–151.
- Terada, T., Moritsugu, K., Matsunaga, Y. and Kidera, A. (2012). μ^2 lib: A library for developing multicopy and multiscale molecular dynamics simulation programs, <http://www.mu2lib.org/>.
- Vanden-Eijnden, E. (2006). Transition path theory, *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1* (eds. M. Ferrario, G. Ciccotti and K. Binder), 439–478, Springer-Verlag, Berlin Heidelberg.
- Vanden-Eijnden, E. and Venturoli, M. (2009). Revisiting the finite temperature string method for the calculation of reaction tubes and free energies, *The Journal of Chemical Physics*, **130**, 194103.

Conformational Transition Pathways in Proteins Explored by the String Method

Yasuhiro Matsunaga

RIKEN Advanced Institute for Computational Science

Conformational transitions of proteins are essential for the regulation of important chemical reactions in cells. Although all-atom molecular dynamics simulation is a powerful technique to investigate atomic details of protein dynamics, it is hard to access the time-scale of such slow transitions with conventional brute-force simulations. This article reviews a method, called the string method, which is designed for efficiently finding transition pathways in proteins. In particular, practical issues of the method, such as free energy calculations along the pathways, are discussed in detail. The method is applied to the conformational transition of Adenylate Kinase. It is shown that the transition pathway identified by the method reveals atomic details of the transition mechanism.