

# パスサンプリングを使った分子動力学と ベイズ推定

藤崎 弘士<sup>†</sup>

(受付 2014 年 1 月 31 日 ; 改訂 7 月 29 日 ; 採択 7 月 29 日)

## 要 旨

近年, 分子動力学の分野ではパスサンプリング (path sampling) と総称される非平衡のパス・アンサンブルを生成する手法が盛んに使われるようになってきているが, この手法はベイズ推定を用いたパラメータ推定と密接な関係がある. 本稿ではその関連について, 主に分子動力学の例を用いて, どのように分子系の動的なパラメータや反応座標などが推定されるのかということを解説する.

キーワード: 分子動力学, パスサンプリング, ベイズ推定, 遷移状態, 遷移パス, 最小自由エネルギー経路.

## 1. パスサンプリングとは何か

パスアンサンブル (path ensemble) (藤崎, 2014) とは, ある力学系によって生成される軌道を統計的なアンサンブルとして考える上での基本的な概念である. 具体的には, 時間の関数として軌道 (簡単のために 1 次元で考えるが, 多次元への概念的な拡張は容易である)  $x(t)$  を考えたときに, その重み  $P[x(t)]$  を  $x(t)$  の汎関数として与えるということである. 軌道の重みを考えるという概念は歴史が古く, たとえば Wiener 過程を考えるときに定義される Wiener 測度がその一例である. より一般の確率過程に対しては Wentzell-Freidlin 作用 (Freidlin and Wentzell, 2012; E, 2011) と呼ばれるものが定義されており, これは物理や化学の分野では Onsager-Machlup (OM) 作用 (Onsager and Machlup, 1953; Adib, 2008) と呼ばれるものと数学的に等価である.

たとえば, 1 次元の位置  $x(t)$  に対する強摩擦のランジュバン方程式を考える.

$$(1.1) \quad \frac{dx}{dt} = \frac{1}{\zeta} F(x) + \sqrt{2D} \eta(t)$$

ここで  $F$  は力,  $\zeta$  は質量を含んだ摩擦係数,  $D$  は拡散定数であり,  $\eta(t)$  は  $\langle \eta(t)\eta(0) \rangle = \delta(t)$  を満たすガウス型の白色ノイズである. すると, OM 作用は以下のように定義される (Adib, 2008).

$$(1.2) \quad S = \frac{\Delta U}{2} + \frac{1}{2} \int_0^t ds \left[ \left( \frac{dx}{dt} \right)^2 + \left( \frac{F}{\zeta} \right)^2 + \frac{2D}{\zeta} \frac{dF}{dx} \right]$$

ただし, ここでは力がポテンシャル関数の微分として書かれる ( $F = -\partial U(x)/\partial x$ ) 場合を考えており,  $U(x)$  はそのポテンシャル関数,  $\Delta U \equiv U(x_f) - U(x_i)$  は経路の始点  $x_i$  と終点  $x_f$  の間のポテンシャル差である. この作用を用いることで, 経路の重みは

<sup>†</sup> 日本医科大学 医学部: 〒180-0023 東京都武蔵野市境南町 1-7-1

$$(1.3) \quad P[x(t)] = \int_{x(0)=x_i}^{x(t)=x_f} \mathcal{D}x(s) e^{-S[x(s)]/2D}$$

と具体的に表される。ここで測度  $\mathcal{D}x(s)$  はこれが経路積分として定義されていることを意味する。

これに類するような作用、つまり軌道の重みを用いることで、軌道を生成するためのアルゴリズムが作れる。それは通常のモンテカルロ法と同様に、トライアルの軌道をまず生成し、この重みを使って、例えばメトロポリス法でそれをアクセプト/リジェクトするかどうかを決めればよい。これは分子動力学法(molecular dynamics, MD)のような方法とは非常に異なる軌道の生成のアルゴリズムであり、このような方法の総称をパスサンプリング(path sampling)と言う。分子動力学法の場合は初期状態を与え、あとは力学系の運動方程式を差分して解くことで、時系列  $x(t)$  を得る。しかし、パスサンプリングの場合は  $x(t)$  そのものを1つの「実体」と考えて、その全体をパス空間で動かすわけである。これが既存のMDと比べたときの、パスサンプリングの概念・アルゴリズム的な違いである。

これは一見、MDと比べて、過剰なコストのかかる方法のように見える。しかし、例えば、始点と終点が決まっているときの力学系の性質を知りたい場合が往々にしてある。そのときは始点からMDを始めるとしても、それが終点に行き着くという保証はない。また、後で述べるように、作用を使った方法であれば、 $\Delta t$  を比較的大きくとれる。そこで、作用に基づく方法が優位になるような状況が出現することが期待される。

このことから容易に分かるように、配置空間  $x$  の場合と異なり、今の場合はパス空間  $x(t)$  なので、サンプルする空間は膨大なものになり(配置空間の  $t/\Delta t$  倍は必要になる)、効率的なアルゴリズムの開発は欠かすことはできない。特に系の自由度が大きくなったときにそれは本質的な問題になるので、ここ10年ほどで様々なアルゴリズムが提案されている。パスサンプリングは、複雑な系における非平衡の振る舞いを調べるためには重要な概念であるので、軌道の重みに基づく軌道アンサンブルの生成法は今後ますます重要になるだろう。(このことに関しては、藤崎, 2014 を参照.)

本稿では上記の考えを「ひっくり返した」手法について考える。例えば、実験的な測定などにより、時系列  $x(t)$  が与えられたときに、軌道の重みに基づいて系の性質を知ることはいかぬのか、という問題である。例えば生体分子などであれば、そのダイナミクスを決定する力場(force field)がある程度精度よく決められており、そこから軌道アンサンブルを作ることができる。しかし、場合によってはこの力場の精度が十分でない場合があり、力場のパラメータを実験的な時系列から refine したいという欲求が生じる。これはまさに統計科学のデータ同化(data assimilation)と同じ考え方であり、現在では物理・化学の分野の研究者たちもこの考えを用いるようになってきている。つまり、パスサンプリングで用いる軌道の重みを尤度(likelihood)関数のように考えて、それからパラメータ推定を行うということである。本稿では、著者の専門である化学物理・生物物理の分野からいくつかの具体例を取り出し、それらについて解説するが、最後に一般的な議論も行う。

## 2. ベイズ推定と分子動力学

上記のパスサンプリングは複雑な系の非平衡現象を理解する上で非常に強力な方法であるが、いくつかの弱点・難点がある。その一つは前節で述べたパス空間のサンプリングの問題であるが、それが解決できたとしてもその結果得られたパスアンサンブルから何を計算するか、それをどう解析・解釈するかという問題が残る。もちろん系の個別性が表に出る問題ではあるが、一般的な問いとしては、

- ある状態から別の状態にどのくらいの時間スケールで遷移するか？
- その遷移のメカニズムは何か？

といったものが考えられる。化学物理・生物物理の分野では、前者に関しては、遷移レート (reaction rate), もしくは拡散定数 (diffusion constant) のような動的な指標となる量をどのように推定するか, 後者に関しては反応座標 (reaction coordinate) をどのように推定するかという問題と考えることができる。

### 2.1 動的パラメータの推定

上で動的なパラメータとして遷移レートと拡散定数を挙げたが、本稿では拡散定数の話題を取り上げよう。(遷移レートを拡散定数を使って表現する問題は古くから Kramers の問題として知られている。確率過程の教科書 Risken, 1989 などを参照のこと。)

まず、単純で一様な系であれば、拡散定数を求めるのは容易であり、以下の estimator を使えばよい。

$$(2.1) \quad D = \lim_{T \rightarrow \infty} \frac{\langle [x(T) - x(0)]^2 \rangle}{2T}$$

しかし、これは非一様な系に対しては単純には使えない。たとえば、タンパク質の折りたたみの問題で、ある集団座標 (collective variable)  $Q$  に注目し、その拡散定数を考えたときに、 $D$  は  $Q$  に依存することがある。つまり、 $Q$  の関数としての拡散「係数」 $D(Q)$  を考えなければならない。これにはいくつかの方法が考えられるが(例えば、Hummer, 2005 など)、ここでは上の OM 作用を用いた重みを尤度関数として用いた Micheletti らの仕事 (Micheletti et al., 2008) を紹介しよう。

彼らは具体的な系としては、Asakura-Oosawa 相互作用を感じる粗視化されたモデルポリマー (混みあった環境中のクロマチンを想定している) を考え、その末端間距離を集団座標  $s$  としたときの、最適なランジュバン・モデルは何かという問題を調べた。(これと同じ系のパスサンプリングの問題に関しては、Fujisaki et al., 2013 を参照。) その際に上で述べたように拡散定数が  $s$  の関数となるので、これは式 1.1 を以下のように拡張しなければならない (Risen, 1989 の 3.3.2 節を見よ)。

$$(2.2) \quad ds_i = v_i(s(t))dt + \sqrt{2}D_{ij}^{1/2}(s(t))dW_j(t)$$

ただし、 $v_i$  は

$$(2.3) \quad v_i(s) = D_{ij}(s)\partial_j \mathcal{F}(s) - \partial_j D_{ji}(s)$$

と定義されるドリフト項であり、その中身の  $\mathcal{F}$  は平均力ポテンシャルに相当する項、 $D_{ij}$  は  $s$  に依存する拡散定数 (係数) である。また、 $dW_j(t)$  は Wiener 過程を表す。ここで重要なことは拡散定数が  $s$  に依存するので、これは乗法性雑音 (multiplicative noise) の場合になっているということと、その帰結としてドリフト項に余分な項が付け足される (式 (2.3) の第 2 項) ということである。

これから  $\chi_i(t) = ds_i(t) - v_i(s)dt$  を定義すると、経路の重みは

$$(2.4) \quad P[s(t)] = \prod_t \pi(s(t), ds(t))$$

$$(2.5) \quad \pi(s, ds) \propto \frac{1}{\det D(s(t))^{1/2}} \exp \left[ -\frac{1}{4dt} D_{ij}^{-1}(s(t)) \chi_i(t) \chi_j(t) \right]$$

となる。微小時間の遷移に対する重み  $\pi(s, ds)$  の指数部 (に  $\det D$  の寄与も更にも含めたものは

微小時間に対する OM 作用を拡散定数が一定でない場合に拡張したものと見なすことができる。  $\pi(s, ds)$  を尤度関数と見なすと、最適な  $v_i, D_{ij}$  を求めるためには、  $\pi(s, ds)$  の対数をとって、  $v_i, D_{ij}$  に関して微分(実際は変分)し、ゼロとすればよい。(その際に、  $D_{ij}$  が対称行列であること、  $\ln \det D$  を  $D_{ij}$  で微分すると  $D$  の逆行列の転置行列が得られるということ、  $D^{-1}(x)$  を  $x$  で微分すると  $-D^{-1}(\partial D/\partial x)D^{-1}$  となることなどを使う。)ただし、それから得られるものは Kramers-Moyal 展開して得られる通常定義

$$(2.6) \quad v_i(s) = \frac{1}{dt} \langle ds_i \rangle,$$

$$(2.7) \quad D_{ij}(s) = \frac{\langle ds_i ds_j \rangle - \langle ds_i \rangle \langle ds_j \rangle}{2dt}$$

と同じである。ここでの平均  $\langle * \rangle$  は集団座標が左辺の  $s$  にある状態でのものであり、拘束付きの平均である。ここまでは外力なしの自然のダイナミクスを考えている。

そこで、この系を外力  $\theta(t)$  で引っ張ってみよう。実験的には、カンチレバー(cantilever)や光ピンセット(optical tweezer)などで分子を引っ張ることに対応する(Hummer and Szabo, 2001)。この結果として得られるダイナミクスはもちろん天然のものと違うが、外力まで含めた OM 作用の重みを考えると、最適化された  $v_i, D_{ij}$  として外力のかかっている自然の状態の推定ができるはずである。効率の点からの期待は、バイアスをかけたダイナミクスによるサンプリングが高速に行われれば、以下の式によって元のダイナミクスの情報を効率よく見積もれるということである。集団座標が1次元のときには  $v_i, D_{ij}$  として、それぞれ1次元の量  $v, D$  を考えればよいので、それらは

$$(2.8) \quad v(s) = \frac{1}{dt} \langle ds \rangle - D(s) \langle \theta \rangle$$

$$D(s) = \frac{-1 + \sqrt{1 + (\langle \theta^2 \rangle - \langle \theta \rangle^2) (\langle ds^2 \rangle - \langle ds \rangle^2)}}{dt (\langle \theta^2 \rangle - \langle \theta \rangle^2)}$$

となる(Micheletti et al., 2008)。また、ここでのベイズ推定を使った議論は尤度関数(重み)に基づいているので、これらの推定値の誤差を見積もることもできる。

その結果得られた  $s$  の関数としての  $v, D$  の推定値と誤差は図1に示されている。それらを用いてランジュバン・ダイナミクスを走らせると、自由エネルギー地形や滞在分布が得られるが、それは外力を加えずに MD を長時間走らせたものから計算したものとほぼ一致する(Micheletti et al., 2008)。また、このランジュバン・モデルは強摩擦を仮定しているので、記憶のある一般化されたランジュバン方程式ではなく、記憶項のない(雑音の相関が急速に減衰する)マルコフ的な方程式である。Micheletti らはマルコフ性についても調べているが、ある程度以上長い  $dt$  を用いれば、マルコフ性が成立することを確かめている。しかし、  $dt$  が長すぎると今度はエネルギー地形を十分にサンプルしなくなるので、よいモデリングではなくなる(ノイズのガウス性も成り立たなくなる)。つまり、ランジュバン・モデルが精度よく成立する  $dt$  の上限と下限が存在するというのである。

OM 作用を用いたパラメータ推定に関しては、Miyazaki, Harada の仕事も面白い(Miyazaki and Harada, 2011)。そこでは、系を外部から観測することで、内部のパラメータを推測しているが、観測の条件をきちんと選ばないと「精度が失われる」という相転移のような現象が起こることが報告されている。また、パスの重みを尤度関数と考えて動的なパラメータを求める方法を実際のペプチドに応用した例としては、Marinelli et al.(2009)などを参照されたい。

## 2.2 反応座標の推定

反応座標とは、「反応」を記述するために「最適化された」座標と考えることができる。ここで

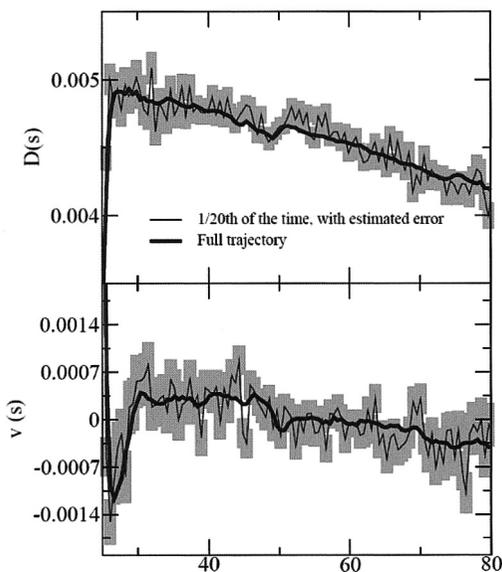


図 1. 粗視化ポリマーモデルの末端間距離に対して、ランジュバン・モデリングを行った結果。上：末端間距離の関数としての拡散定数。下：末端間距離の関数としてのドリフト。Reprinted with permission from Micheletti, C., Bussi, G., and Laio, A. (2008). *The Journal of Chemical Physics*, **129**, 074105. Copyright 2008 AIP Publishing LLC.

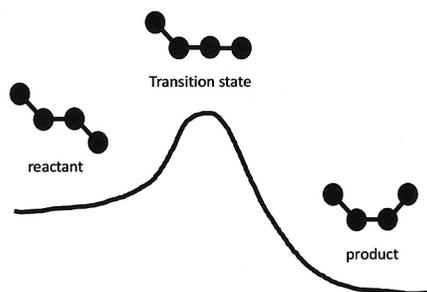


図 2. 分子系におけるレアイベントの概念図。

反応とはある状態 A から別の状態 B に遷移することであり、それは従来の化学反応や構造変化などの一般的な状態遷移も含むものとする。まず、状態をどのように定義するかという問題からして自明ではないが、ここでは化学反応のように反応物 (reactant) と生成物 (product) があり、それが何らかの障壁 (barrier) で隔てられていると考えておこう (図 2)。化学反応などを考える上では、この障壁上の状態を考えることがよくあり、それは遷移状態 (transition state) と呼ばれる。反応座標を求めるといことは大雑把に言って、この遷移状態とその周辺がどうなっているか、それがどのように反応物、生成物につながっていくか調べることに相当する。また、A から遷移状態を越えて B にいたるパスを遷移パス (transition path, TP) と呼ぶ。その統計的な性質を調べるのが以下では課題となる。

遷移パスの「道筋」のみに興味があるのであれば、それを求める方法は多数存在する。化学物理・生物物理・物質科学など分野で近年よく使われるものとしては nudged elastic band 法や string 法と呼ばれる方法がある(藤崎, 2011)。これらは経路を離散化して、ビーズの集まりとし、それらが一定の間隔を保つように制御しながら、系のエネルギー地形に合わせて、最小エネルギー経路 (minimum energy path) を探すという方法である。これは非常に有効な方法であり、大きな系に関しては、集団座標を考慮して、それに対する最小自由エネルギー経路 (minimum free energy path) を求めることも可能である(酵素の構造変化への応用例としては, Matsunaga et al., 2012 を参照)。

しかし、このような手法の問題点は、本来ダイナミックなパスの情報を必ずしも引き出すことができないということである(それを近似的に扱うのが遷移状態理論 (transition state theory) であり、またそれに基づく反応流束の考えである (Chandler, 1987))。しかし、パスサンプリングであれば、そこには動的な情報が乗っているため、パスの重みを使って反応経路の情報を引き出せないかと考えるのは自然である。そこで、Hummer は以下の条件付き確率の式(実質的にベイズ推定(伊庭, 2003)の式)を用いることを考えた(Hummer, 2004)。

$$(2.9) \quad p(x|\text{TP})p(\text{TP}) = p(\text{TP}|x)p_{\text{eq}}(x)$$

ここで  $p_{\text{eq}}(x)$  は平衡状態の分布(つまりボルツマン分布)、 $p(x|\text{TP})$  は遷移パスの配置空間  $x$  における分布、 $p(\text{TP})$  は遷移パスの占める割合(平衡における長時間の軌道を考えて、そのトータルな時間に対する遷移パスが費やす時間として定義できる)、 $p(\text{TP}|x)$  は  $x$  が遷移パス上にある確率を表す。 $x = (q, p)$  は相空間上の 1 点を表す記号とする。ここで一番知りたいことは  $p(\text{TP}|x)$  であり、もちろん

$$(2.10) \quad p(\text{TP}|x) = \frac{p(x|\text{TP})p(\text{TP})}{p_{\text{eq}}(x)}$$

によって得ることができる。これは TP 上にあれば有限の値を持つが、遷移状態で最大になることが期待される。よって、これを尤度関数と見なして最大化することで動的な遷移状態を求めるというアルゴリズムを考えることができる。

一方、自由度の大きな系に関しては、詳細な相空間の座標  $x$  ではなく、集団座標  $r(x)$  を使うことが多い。ただし、ここでは議論を簡単にするために、 $r$  は 1 次元の座標とする。それに対しては条件付き確率の式は

$$(2.11) \quad p(\text{TP}|r) = \frac{p(r|\text{TP})p(\text{TP})}{p_{\text{eq}}(r)}$$

となる。ただし、ここで

$$(2.12) \quad p(\text{TP}|r) = \frac{\int p(\text{TP}|x)\delta[r - r(x)]p_{\text{eq}}(x)dx}{\int \delta[r - r(x)]p_{\text{eq}}(x)dx}$$

である。この場合も遷移状態は  $p(\text{TP}|r)$  が最大となる  $r$  としてよいだろう。

集団座標  $r(x)$  は原理的には  $x$  の関数として任意に選べる。しかし、それが反応座標と呼べるためには、この条件付き確率  $p(\text{TP}|r)$  が  $r$  の関数として 1 つのシャープなピークをもっていたほうがよい。つまり、これは図 2 のような描像が明確に現れるということであり、1 つの遷移状態があることを仮定していることになる。(もちろん、ある現実の系に対して、1 次元の反応座標の問題も含め、この描像が成り立つかどうかは別の検証がいる。)

比較的小さなタンパク質の折りたたみに関しては、native contact  $Q$  と呼ばれる量がよい 1 次元の反応座標であることが知られている (Best and Hummer, 2006) が、Best と Hummer はこの

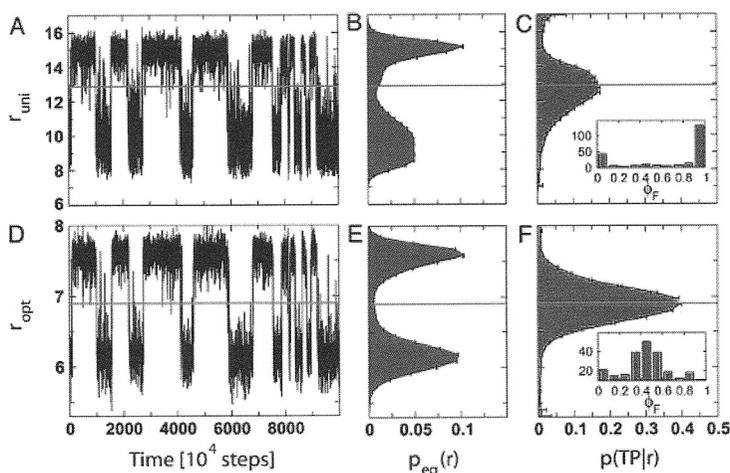


図3. 粗視化されたタンパク質モデルに対する、反応座標のダイナミクス(左)、その分布(中央)、遷移パス上にある確率(右). 上は最適化されていない反応座標, 下はベイズ推定から最適化された反応座標. Reprinted with permission from Best, R. B. and Hummer, G. (2005). *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6732. Copyright 2005 National Academy of Sciences, U.S.A.

問題を上の条件付き確率を使って再検討した (Best and Hummer, 2005). まず,  $Q$  によく似た量を以下で定義した.

$$(2.13) \quad r_W = \frac{1}{2} \sum_{i,j} w_{ij} q_{ij}$$

ここで  $q_{ij}$  は contact matrix と呼ばれる量で, 残基  $i$  と  $j$  の間の距離が  $12\text{\AA}$  以下であれば 1, それ以外は 0 となるようなものである.  $w_{ij}$  は重みで,  $\sum_{i,j} w_{ij} = 1$  を満たす. そして, これがよい反応座標になるために, どのように重み  $w_{ij}$  を選ばばいいのかという問題を考えた. そこで上で述べたように,  $p(\text{TP}|r_W)$  を尤度関数と思って,  $w_{ij}$  を最適化する解法を提案した. これはまさに動的な軌道を使ったパラメータ推定を行うということに相当する. その際には,  $w_{ij}$  をランダムに 2 つ選んで, それらを変更する Monte Carlo (MC) ムーブを考えるが, そのとき  $w_{ij}$  の確率としての規格化を満たすようにして変更する (つまり,  $w_{ij}, w_{kl}$  を  $w_{ij} + \Delta w_{ij}, w_{kl} + \Delta w_{kl}$  と変えるときに,  $\Delta w_{ij} + \Delta w_{kl} = 0$  となるようにする).

まず, MC ムーブを行う前の初期状態として, 一様な  $w_{ij}$  を考えると, 図 3 上のような結果が得られる.  $p(\text{TP}|r_W)$  のピークはあまり高くなく,  $r$  が大きいところで別の分布の山も見られる. それに対して,  $p(\text{TP}|r_W)$  の最大値の最適化を行った結果が図 3 下である. まず,  $r$  の関数としての分布は非常にガウシアンに近く, ピークの値も大きい. よって, こちらの  $r_W$  のほうがより反応座標としてふさわしいと判断できる.

反応座標が得られれば, それに基づいて, 「反応」(いまの場合はタンパク質の折りたたみ)の理解をすることができる. その結果,  $r_W = 6.5$  付近の非折りたたみの状態から,  $r_W = 6.9$  付近で 2 つのヘリックスが形成される状態を経て,  $r_W = 7.4$  付近の 3 つのヘリックスが形成される折りたたみ状態に落ち着くことが分かった.

このような計算を行うときは 2 通りのやり方がある. まず, 理想的な状況として, 長時間の

軌道が大量にある場合を考える. 上の場合,  $r_W$  の計算は簡単なので, 図3のように  $r_W$  の軌道から  $p_{\text{eq}}(r)$ ,  $p(r|\text{TP})$  などを計算すればよい. しかし, それほど大量の軌道がない場合は,

$$(2.14) \quad p(\text{TP}|x) = \phi_A(x)\phi_B(x) + \phi_A(x)\phi_B(x)$$

という関係 (Hummer, 2004) を使って, コミッター (**committor**) 関数  $\phi_{A,B}(x)$  から直接  $p(\text{TP}|x)$  を計算する. ここでコミッター関数とは, 相空間の一点  $x = (q, p)$  から出発して, A, B どちらかのベイスンを最初に訪れるかを表す確率である. (この概念に確率が割り振られるためには, ランジュバン・ダイナミクスのようなもの考えるのが普通である.) また,  $x = (q, -p)$  であり, これは逆向きに軌道を走らせるという意味である.

もっと粗視化された集団座標  $r$  を考えて,  $r$  が従うダイナミクスが強摩擦のランジュバン方程式に従うと仮定すれば,

$$(2.15) \quad p(\text{TP}|r) = 2\phi_A(r)(1 - \phi_A(r))$$

が成り立つ. 2つのベイスンを考える場合は, コミッター関数の最大値は  $1/2$  であり, そのとき状態は遷移状態付近にあると考えることができる.

また, 慣習的なパスサンプリングのアルゴリズム (例えば, Chandler らの shooting アルゴリズム) を使うと, MC ムーブをする軌道の断片 (ビーズ) が遷移状態付近にないとリジェクトされてしまうことが多い (reactant か product 側に大部分の軌道が流れていってしまうため) ので, Hummer はもっと巧みな方法を考えた. それは遷移状態付近のビーズのみを MC ムーブするという方法であり, その際にはまたパスサンプリングの時間を決め打ちしなければならない問題も回避される. また, そうして得られたパスサンプリングから反応レートを見積もることも可能である. これは原理的に非常によい方法であるが, 複雑な分子系に適用する時にはまず遷移状態を求めるのが難しいために簡単にはいかない. (Best と Hummer は粗視化されたタンパク質モデルやナノチューブ中の水の集団的な双極子のダイナミクスにこの手法を用いたが, これらの場合は遷移状態を見つけるのは比較的簡単である.) この考えをさらに一般化し, 現実的な分子系に適用しやすくすることは今後の課題である (Peters and Trout, 2006).

### 3. まとめと展望

生体分子の機能においては「揺らぎ」が非常に重要だということはよく言われている. しかし, それらは平衡状態付近での揺らぎを指していることが多く, タンパク質の折りたたみや構造変化のような, より動的で複雑な非平衡の「揺らぎ」とは区別しなければならない. この後者の部分を強調するような生体分子の教科書 (Zuckerman, 2010) も出版されており, そこでも軌道の重みという概念が強調されている (この本の 11 章には, OM 作用の分かりやすい解説がある). 軌道の重みに基づいて, 生体分子の経路を統計的に生成したり, 本稿で解説したように動的なパラメータや反応座標を推定することは, 今後ますます行われるようになってくるだろう.

また, ここでは生体分子のシミュレーション周辺の話を取り上げたが, 経路の重みや OM 作用という概念はランダムなダイナミクスに一般的なものであるので, 生体分子に限った話ではない. 例えば, Apte et al. (2007) らの研究では, 気象データのようなものに対するデータ同化を行う際に, ダイナミクスとしてはランジュバン方程式を仮定しているが, それは OM 作用を用いたパラメータ推定と同等になる. また, もちろん多変数に対する拡張も (概念的には) 容易であり, Ginzburg-Landau 方程式や反応拡散方程式に対応する OM 作用 (等価だが Wentzell-Freidlin 作用) を考えることもできる (E et al., 2004). その結果として, パターン形成のようなダイナミクスをレアイベントと捉えてサンプルする, もしくはそのダイナミクスからパラメータ推定を

することもできる。よって、このような概念・方法を分子レベルから接続していくことで、ミクロからセミマクロにつながるマルチスケールな方法論(E, 2011)が確立できると期待される。

古田忠臣氏(東京工業大学)と重田育照氏(大阪大学)には原稿を読んでいただき、貴重なコメントをいただいた。ここに感謝します。

## 参 考 文 献

- Adib, S. (2008). Stochastic actions for diffusive dynamics: Reweighting, sampling, and minimization, *The Journal of Physical Chemistry B*, **112**, 5910–5916.
- Apte, A., Hairer, M., Stuart, A. M. and Voss, J. (2007). Sampling the posterior: An approach to non-Gaussian data assimilation, *Physica D*, **230**, 50–64.
- 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.
- Best, R. B. and Hummer, G. (2006). Diffusive model of protein folding dynamics with Kramers turnover in rate, *Physical Review Letters*, **96**, 228104-1–228104-4.
- Chandler, D. (1987). *Introduction to Modern Statistical Mechanics*, Oxford University Press, New York.
- E, W. (2011). *Principles of Multiscale Modeling*, Cambridge University Press, New York.
- E, W., Ren, W. and Vanden-Eijnden, E. (2004). Minimum action method for the study of rare events, *Communications on Pure and Applied Mathematics*, **57**, 637–656.
- Freidlin, M. I. and Wentzell, A. D. (2012). *Random Perturbations of Dynamical Systems*, 3rd ed., Springer, Heidelberg.
- 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.
- 藤崎弘士(2014). 分子系に対するパスサンプリングについて, 分子シミュレーション研究会会誌アンサンブル, **16**, 8–15.
- Fujisaki, H., Shiga, M., Moritsugu, K. and Kidera, A. (2013). Multiscale enhanced path sampling based on the Onsager-Machlup action: Application to a model polymer, *The Journal of Chemical Physics*, **139**, 054117-1–054117-9.
- Hummer, G. (2004). From transition paths to transition states and rate coefficients, *The Journal of Chemical Physics*, **120**, 516–523.
- Hummer, G. (2005). Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations, *New Journal of Physics*, **7**, 34-1–34-14.
- Hummer, G. and Szabo, A. (2001). Free energy reconstruction from nonequilibrium single-molecule pulling experiments, *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 3658–3661.
- 伊庭幸人(2003). 『ベイズ統計と統計物理』, 岩波書店, 東京.
- Marinelli, F., Pietrucci, F., Laio, A. and Piana, S. (2009). A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations, *PLoS Computational Biology*, **5**, e1000452-1–e1000452-18.
- 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**, e1002555-1–e1002555-12.
- Micheletti, C., Bussi, G. and Laio, A. (2008). Optimal Langevin modelling of out-of-equilibrium molecular dynamics simulations, *The Journal of Chemical Physics*, **129**, 074105-1–074105-8.
- Miyazaki, M. and Harada, T. (2011). Bayesian estimation of the internal structure of proteins from

- single-molecule measurements, *The Journal of Chemical Physics*, **134**, 085108-1–085108-10.
- Onsager, L. and Machlup, S. (1953). Fluctuations and irreversible processes, *Physical Review*, **91**, 1505–1511.
- Peters, B. and Trout, B. L. (2006). Obtaining reaction coordinates by likelihood maximization, *The Journal of Chemical Physics*, **125**, 054108-1–054108-10.
- Risken, H. (1989). *The Fokker-Planck Equation: Methods of Solution and Applications*, Springer, Berlin.
- Zuckerman, D. M. (2010). *Statistical Physics of Biomolecules: An Introduction*, CRC Press, New York.
- (藤崎弘士, 藤崎百合 訳(2014). 『生体分子の統計力学入門—タンパク質の動きを理解するために—』, 共立出版, 東京.)

## Molecular Dynamics Using Path Sampling and Bayesian Inference

Hiroshi Fujisaki

Department of Physics, Nippon Medical School

Recently “path sampling” methods to generate non-equilibrium path ensembles have been extensively used along with molecular dynamics simulations. These techniques are closely related to the parameter estimation using the Bayesian estimation in statistics. Reviewing related works of molecular dynamics simulations, this article explains how the dynamic parameters and/or reaction coordinates in such a molecular system can be efficiently estimated using both path sampling and Bayesian inference.