

第 II 部 分子動力学計算の方法と解析方法

はじめに

第 II 部では、タンパク質を例にとり、MD 計算の方法と得られたトラジェクトリの解析方法を述べる。第 1 章では、MD 計算で用いる分子のエネルギー関数やアンサンブルの発生方法について詳述する。第 2 章では、系のサンプリングを効率よく行

うために工夫された MD 計算手法を説明し、得られた結果から自由エネルギーを算出する方法を示す。第 3 章では、MD 計算から得られたトラジェクトリの典型的な解析方法を紹介する。

1. 分子動力学 (MD) 法

1.1 MD 計算の概要

MD 計算は、多数の原子や分子の集団系である液体、固体、ガラス、タンパク質などの構造やダイナミクスの研究に用いられる。MD 計算はこの集団系を構成している全ての原子 1 つ 1 つの運動について Newton の運動方程式を逐一解くことにより、その軌跡を追跡するものである。このためには、系全体の自由度の数だけ運動方程式を解く必要が

あるが、解析的に解くことは不可能であるため時間に沿って数値的に解く。運動方程式を解くには原子に作用する力を求めることが必要である。もしこの力の評価が十分な精度で正しいものであり、かつ運動方程式に関しても十分に精度の高い数値解が得られるとすれば、系の全自由度に対して得られた分子の軌跡は自然界で観測されるものと同

等であると考えられる。機器による測定では、さまざまな方法で、物質の構造やダイナミクスを反映した物理量を得ることができるが、実際に運動している原子の様子を直接観察することはできない。これに対して MD 計算では、系の全自由度に

対する原子の軌跡を得るため、系の構造や分子運動の直接観察が可能である。さらに現象を特徴付けるような熱力学量を軌跡から直接計算して求めることができる。

1.2 系のエネルギー関数¹⁻²

MD 計算を実行する際には、系の性質を表すポテンシャルエネルギーの決定が重要である。第一原理に従い量子力学計算から求めることもできるが、計算時間が莫大になり実用上不可能である。その代わり一般的に用いられている方法は、古典力学に基づく経験的なポテンシャルエネルギーの関数形を予め与えて、それを基に MD 計算を行う方法である。この中では系の分子の化学結合の情報も含められるため、計算を実行する前にどの粒子が化学結合をしているかを指定する必要がある。このことから、通常の MD 計算では化学結合の生成・解離を扱うことはできない。また相互作用を短い

時間で計算するために、複雑な分子を単純な系へとモデル化してポテンシャルを簡単な関数で表現する必要がある。このため分子に対して相互作用点を導入する。これは斥力や分散力、静電相互作用、水素結合や共有結合まで含めたすべての相互作用をこの点の間の距離だけに依存した関数 $U_{\alpha\beta}(r_{\alpha\beta})$ に渡る和として表す。 $r_{\alpha\beta}$ は相互作用点とした α と β の距離とする。通常、相互作用点は原子核の位置に置かれる。ポテンシャルエネルギー関数には経験的なパラメータが含まれる。分子内ポテンシャル関数と分子間ポテンシャル関数について以下に記す。

1.2.1 分子内ポテンシャル関数

分子内ポテンシャル関数は以下のように与えられる。結合長に関しては原子間を調和振動子でつなぐ。

$$U_{bond}(r_{\alpha\beta}) = \frac{1}{2} k_{bond} (r_{\alpha\beta} - r_0)^2 \quad (2.1)$$

ここで、 $r_{\alpha\beta}$ は 2 つの原子間の結合距離、 k_{bond} は結合の force constant、 r_0 は平衡の結合長である。

結合角については、2 本の結合間を調和振動子でつなぐ。

$$U_{angle}(\theta_{\alpha\beta\gamma}) = \frac{1}{2} k_{angle} (\theta_{\alpha\beta\gamma} - \theta_0)^2 \quad (2.2)$$

ここで、 $\theta_{\alpha\beta\gamma}$ は結合角、 k_{angle} は変角のばね定数、 θ_0 は平衡の結合角である。

二面角については、以下の(a), (b)に場合分けする。

(a) 分子内に分岐なく 1 次元的に並んだ原子の場合

これは図 2-1-2-1a に示すように、分岐なく 1 次元的に並んだ 4 つの原子の場合である。この場合は二面角 ϕ_{ijkl} に依存する項と原子 i と l の間の相互作用項との和で表す。

$$U_{\text{torsion}}(\phi_{ijkl}) = \frac{1}{2} k_{\text{torsion}} \left\{ 1 - \cos(n\phi_{ijkl} - \phi_0) \right\} + CV_{ij}(r_{ij}) \quad (2.3)$$

ここで C はポテンシャル調整のためのパラメータである。 n はポテンシャルミニマムの数を示し、飽和炭化水素鎖などでは両端の炭素原子に結合する水素原子などの影響により $n=3$ であり、1 個のトランス構造と 2 個のゴーシュ構造を表す。また二重結合があるような場合には、二重結合は自由には回転しないため $n=2$ となる。このポテンシャル関数はタンパク質の高次構造や脂質分子のコンフォメーションに敏感である。

(b) 分子内に化学結合の分岐が存在する場合

これは図 2-1-2-1b に示すように、分子内に化学結合の分岐が存在し 1 次元的に並んでいない場合である。この場合は平衡位置では l 原子は i - j - k の 3 つの原子が作る平面上にあるが、 l 原子はこの平面上からずれる運動も可能である。このとき i - j - k 原子で構成する平面と i - j - l 原子で構成する平面の作る角度を improper torsion と呼び ψ_{ijkl} で表す。

$$U_{\text{improper torsion}}(\psi_{ijkl}) = \frac{1}{2} k_{\text{improper torsion}} (\psi_{ijkl} - \psi_0)^2 \quad (2.4)$$

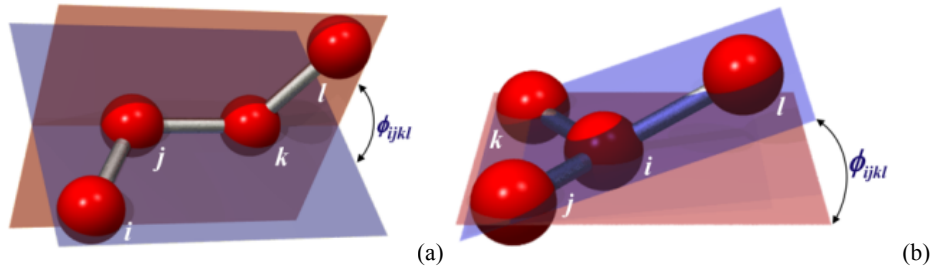


図 2-1-2-1 二面角と improper torsion を構成する分子構造³

(a) 原子団 i - j - k - l が分岐なく 1 次元的に結合した分子構造中の二面角 ϕ_{ijkl} 。 (b) 分子内に分岐が存在し、原子団 i - j - k - l が 1 次元的に並んでいない場合の improper torsion ψ_{ijkl} 。³ 改変。

1.2.2 分子間ポテンシャル関数

(a) van der Waals 相互作用

1 つの分子を構成している多数の電子のある瞬間の配置を考えると、ほとんどの場合これは原子核に対して対称ではなく分極を生じている。この分極によって生じる電場に、他の分子の電子が応答すると分極が発生する。これらの瞬間的に分極した分子同士の相互作用は、電子の存在確率に従って平均を取った場合でも 0 とはならず引力相互作用として残る。これが van der Waals ポテンシャルあるいは分散力によるポテンシャルと呼ばれているものであり、以下の式で表される。

$$U_{\text{vdW}}(r_{\alpha\beta}) = 4\epsilon_{\alpha\beta} \left\{ \left(\frac{\sigma_{\alpha\beta}}{r_{\alpha\beta}} \right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r_{\alpha\beta}} \right)^6 \right\} \quad (2.5)$$

式中の $\varepsilon_{\alpha\beta}$, $\sigma_{\alpha\beta}$ は原子の種類に依存した定数である。分散力は距離に対しての減衰は速いが、原子に働く長距離相互作用をすべて計算に含めると、莫大な計算量になってしまう。そこで、ある一定の距離までの相互作用は計算し、それ以上の遠い位置にある原子からの相互作用は無視するカットオフ近似を用いる。

(b) 静電相互作用

水分子を例にとると、酸素原子や孤立電子対の周辺に負の電荷を持った電荷が分布し、また水素原子に正の電荷が分布している。この場合、水分子には永久双極子モーメントが存在し、さらにもう 1 つの水分子を離れた場所に置くと、分子間相互作用としてこれら電荷間に静電相互作用が働く。つまり、双極子-双極子相互作用がある。これは水分子に限らず、イオンの場合には電荷があり、二酸化炭素 (CO_2) やベンゼン (C_6H_6) など構成する原子数が増えるほど多重極の双極子モーメント間での静電相互作用が起きる。この相互作用は 2 つの電荷の間に働くクーロンポテンシャルによって表現される。

$$U_{\text{electrostatic}}(r_{\alpha\beta}) = \frac{Q_\alpha Q_\beta}{4\pi\varepsilon_0 r_{\alpha\beta}} \quad (2.6)$$

Q_α , Q_β はそれぞれ原子 α , β の電荷の大きさ、 ε_0 は真空の誘電率である。このクーロンポテンシャルは長距離力の代表的なものであり、距離に対して極めてゆっくりとしか減衰しない。そのため途中でカットオフしてしまうと大きな誤差を招いてしまう。そのため効率よく見積もる方法として MD 法では Ewald の方法⁴が広く用いられている。

この他にも重要な相互作用として水素結合がある。これは水分子を例にとると、2 つの水分子が接近して、1 つの水分子の水素原子が相手分子の酸素原子に存在する孤立電子対の方を向くとき安定化要素として働く相互作用である。水分子に限らず、タンパク質の二次構造の安定化にも大きく寄与する。この水素結合に対しては特別なポテンシャル関数は与えない。van der Waals 相互作用と静電相互作用のパラメータの中に組み込んで表現する。以上を合わせると系のポテンシャルエネルギー関数は

$$\begin{aligned} U = & \sum_{l=1}^{N_{\text{bond}}} \frac{1}{2} k_{\text{bond}} (r_l - r_0)^2 \\ & + \sum_{l=1}^{N_{\text{angle}}} \frac{1}{2} k_{\text{angle}} (\theta_l - \theta_0)^2 \\ & + \sum_{l=1}^{N_{\text{torsion}}} \frac{1}{2} k_{\text{torsion}} \{1 - \cos(n\phi_l - \phi_0)\} + CV_l(r_l) \\ & + \sum_{l=1}^{N_{\text{improper torsion}}} \frac{1}{2} k_{\text{improper torsion}} (\psi_l - \psi_0)^2 \\ & + \sum_{l=1}^{N_{\text{vdW}}} 4\varepsilon_l \left\{ \left(\frac{\sigma_l}{r_l} \right)^{12} - \left(\frac{\sigma_l}{r_l} \right)^6 \right\} \\ & + \sum_{l=1}^{N_{\text{electrostatic}}} \frac{q_l Q_l}{4\pi\varepsilon_0 r_l} \end{aligned} \quad (2.7)$$

と与えられる。ここで各項における l は各相互作用に付けた番号であり、 N は全相互作用数である。

1.3 Newton の運動方程式の数値計算法^{1,2}

分子集団系における全自由度に対する運動方程式は連立微分方程式であり、これは一般に解析的に解くことはできない。そこである時間刻みごとに時間に沿って逐次数値的に解を求めていく。MD 計算で良く用いられる運動方程式の数値計算法には、大きく分けて差分近似法と予測子-修正子法、そしてリウビル演算子を用いた時間発展演算子法がある²。これらの方法は微分方程式に対する数値計算法を運動方程式に適用したものである。以下では差分近似法の 1 つである速度ベルレ (velocity verlet) 法¹を説明する。

まず速度ベルレ法の基礎となるベルレ法¹から述べる。現在の時刻 t から Δt だけ未来 ($t+\Delta t$) と過去 ($t-\Delta t$) の粒子 i の位置 $r_i(t\pm\Delta t)$ を、時刻 $t=t$ のまわりでテイラー展開する。

$$r_i(t+\Delta t) = r_i(t) + \Delta t \dot{r}_i(t) + \frac{\Delta t^2}{2!} \ddot{r}_i(t) + \frac{\Delta t^3}{3!} \dddot{r}_i(t) + O(\Delta t^4) \quad (2.8)^i$$

$$r_i(t-\Delta t) = r_i(t) - \Delta t \dot{r}_i(t) + \frac{\Delta t^2}{2!} \ddot{r}_i(t) - \frac{\Delta t^3}{3!} \dddot{r}_i(t) + O(\Delta t^4) \quad (2.9)$$

式(2.8), (2.9)の両辺を足し合わせると、

$$r_i(t+\Delta t) + r_i(t-\Delta t) = 2r_i(t) + \Delta t^2 \ddot{r}_i(t) + O(\Delta t^4) \quad (2.10)$$

となり、Newton の運動方程式

$$F_i = m_i \frac{d^2 r_i}{dt^2}$$

より、

$$r_i(t+\Delta t) = 2r_i(t) - r_i(t-\Delta t) + \frac{\Delta t^2}{m} F_i(t) + O(\Delta t^4) \quad (2.11)$$

が得られる。これは、粒子 i の現在の位置と力、そして Δt 秒前の位置から Δt 後の位置を求めるものである。二階微分方程式は 2 つの境界条件を必要とするが (通常は初期位置と初速度、あるいは最初と最後の位置)、この場合は現在と Δt 秒前の位置である。力は分子同士に働くポテンシャルエネルギーの位置微分として予め解析的に求めておいて、時刻 t での分子集団の配置から直接計算することができる。すべての粒子に対して Δt 秒だけ未来の位置を求めることができるから、この計算を時間軸に沿って逐次実行することにより系の時間発展、つまり全粒子の軌跡を求めることができる。このとき粒子の速度は式(2.8)から式(2.9)を差し引いて得られる式

$$\dot{r}_i(t) = \frac{r_i(t+\Delta t) - r_i(t-\Delta t)}{2\Delta t} + O(\Delta t^2) \quad (2.12)$$

ⁱ右辺最後の項の $O(\Delta t^4)$ は、 Δt に対して 4 次以上の高次の誤差が含まれており、右辺の残りの項は Δt の 3 次まで正しいことを意味している。

から求める．しかし式(2.11)，式(2.12)をそのまま用いるのではなく，数値計算上の桁落ちⁱⁱを防ぐために式

(2.11)を以下のように少し変形した形式を用いる．これが速度ベルレ法である．式(2.11)に $\frac{1}{2}r_i(t-2\Delta t)$ を足

して引いて，

$$\begin{aligned}
 r_i(t+\Delta t) &= r_i(t) + \frac{1}{2}r_i(t) + \frac{1}{2}r_i(t) - r_i(t-\Delta t) \\
 &+ \frac{1}{2}r_i(t-2\Delta t) - \frac{1}{2}r_i(t-2\Delta t) + \frac{\Delta t^2}{m_i}F_i(t) \\
 &= r_i(t) + \Delta t \frac{r_i(t) - r_i(t-2\Delta t)}{2\Delta t} \\
 &+ \frac{1}{2} \left\{ 2r_i(t-\Delta t) - r_i(t-2\Delta t) + \frac{\Delta t^2}{m_i}F_i(t-\Delta t) \right\} \\
 &- r_i(t-\Delta t) + \frac{1}{2}r_i(t-2\Delta t) + \frac{\Delta t^2}{2m_i}F_i(t) + \frac{\Delta t^2}{2m_i}F_i(t) \\
 &= r_i(t) + \Delta t \left\{ \dot{r}_i(t-\Delta t) + \frac{\Delta t}{m_i} \frac{F_i(t) + F_i(t-\Delta t)}{2} \right\} + \frac{\Delta t^2}{2m_i}F_i(t)
 \end{aligned} \tag{2.13}$$

となる．第 2 式の中括弧内は $r_i(t)$ を式(2.11)に対応する展開で表した．ここで MD 計算において時刻 $t=t$ における速度を

$$v_i(t) = v_i(t-\Delta t) + \frac{\Delta t}{m_i} \frac{F_i(t) - F_i(t-\Delta t)}{2} \tag{2.14}$$

により計算することになると位置は，

$$r_i(t+\Delta t) = r_i(t) + \Delta t v_i(t) + \frac{\Delta t^2}{2m_i}F_i(t) + O(\Delta t^4) \tag{2.15}$$

から計算される．速度と位置に関する一組の時間発展の式(2.14)と(2.15)は式(2.11)と数学的に等価であり同じトラジェクトリを与える．しかし数値的には異なり，式(2.14)と式(2.15)に対しては n ステップ後においても，

$$v_i(t+n\Delta t) = v_i(t) + \frac{\Delta t}{m_i} \sum_{k=1}^n \frac{F_i[t+(k-1)\Delta t] - F_i(t+k\Delta t)}{2} \tag{2.16}$$

ⁱⁱ桁落ち

例えば，6 桁の有効数字を持った計算機で，1.23456 に 0.00123456 を 1000 回足し合わせることを考える．正確には， $(1.23456+0.00123456*1000)=2.46912$ となる．しかし，計算機では 1 つずつ加えて行くと $1.23456+0.00123+0.00123+0.00123 \cdots = 2.46456$ となり，精度は 3 桁でしかない．これを桁落ちという．式(2.11)において右辺第 3 項の $(\Delta t^2/m_i)F_i(t)$ は第 1 項，第 2 項に比べて極めて小さく，桁落ちが生じる．

$$r_i(t + n\Delta t) = r_i(t) + \Delta t \sum_{k=1}^n v_i \left[t + (k-1)\Delta t \right] + \frac{\Delta t^2}{2m_i} \sum_{k=1}^n F_i \left[t + (k-1)\Delta t \right] \quad (2.17)$$

のように大きな数と小さな数を分離し、小さい数ばかりの和を取った後でこれを大きな数に加えることができ桁落ちを防ぐことができる。

1.4 アンサンブルの生成方法^{1,2}

系のハミルトニアン H が次のように与えられているとする。

$$H(q, p) = \sum_{i=1}^N \frac{p_i^2}{2m_i} + U(q) \quad (2.18)$$

このハミルトニアンに対する正準方程式は

$$\begin{aligned} \frac{dq_i}{dt} &= \frac{\partial H}{\partial p_i} = \frac{p_i}{m_i} \\ \frac{dp_i}{dt} &= -\frac{\partial H}{\partial q_i} = -\frac{\partial U}{\partial q_i} \end{aligned} \quad (2.19)$$

となり、Newton の運動方程式に帰着する。また式(2.19)から

$$\frac{dH}{dt} = 0 \quad (2.20)$$

となることから、この系は等エネルギー面上を運動する。与えられた全エネルギーを E とすると分布関数 $P(q, p)$ は、

$$P(q, p) \propto \delta[H(q, p) - E] \quad (2.21)$$

と得られる。しかし我々が必要とするのは式(2.20)を満たすアンサンブルではない。生体環境の再現や実験で得られた物理量との比較のため、等温・等圧条件を満たすアンサンブルが望ましい。そこで MD 法における等温等圧アンサンブルを得る手続きを示す。

温度・圧力制御法：等温等圧 (NPT) アンサンブルの生成

NPT アンサンブルの生成において、温度制御には Langevin dynamics 法を用い、これとカップルさせて圧力制御を行う Langevin piston Nosé-Hoover 法⁵⁻⁶を示す。温度制御である Langevin dynamics 法は Langevin 方程式を解くことによりカノニカル分布を与える方法である¹。Langevin 方程式は以下のように与えられる

$$\begin{aligned} \frac{dq_i}{dt} &= \frac{p_i}{m_i} \\ \frac{dp_i}{dt} &= -\frac{\partial U}{\partial q_i} - \gamma_i p_i + R_i \end{aligned} \quad (2.22)$$

ここで、 γ_i は摩擦係数であり、また R_i は

$$\langle R_{i\alpha}(t)R_{i\beta}(t') \rangle = 2m_i\gamma_i k_B T \delta_{ij} \delta_{\alpha\beta} \delta(t-t') \quad (2.23)$$

となるようなランダム力である。この方程式を数値的に解く際には、乱数を用いてランダム力を取り入れる。この方程式を解くことにより得られる Langevin ダイナミクスの分布関数 $P(q, p)$ はカノニカル分布であり、

$$P(q, p) \propto \exp \left[-\beta \left(\sum_{i=1}^N \frac{p_i^2}{2m_i} + U(q) \right) \right] \quad (2.24)$$

で表される。圧力制御の Langevin piston Nosé-Hoover 法は、Nosé-Hoover 法による圧力制御⁵と Langevin dynamics 法によるピストンの熱揺動を組み合わせた方法である。まず Langevin piston 法⁶のピストンの Langevin 方程式は

$$\dot{r}_i = \frac{p_i}{m_i} + \frac{1}{3} \frac{\dot{V}}{V} r_i \quad (2.25)$$

$$\dot{p}_i = f_i - \frac{1}{3} \frac{\dot{V}}{V} p_i \quad (2.26)$$

$$\ddot{V} = \frac{1}{W} [P(t) - P_{ext}] - \gamma \dot{V} + R(t) \quad (2.27)$$

ここで V はピストンの体積、 $P(t)$ はピストンの瞬間の圧力、 P_{ext} はピストンに指定した圧力、 r_i 、 p_i 、 m_i はそれぞれ粒子 i の位置、運動量、質量、力である。 W はピストンの質量としている（次元は kgm^{-4} ）。 γ はピストンに対する摩擦係数（次元は s^{-1} ）であり、 $R(t)$ はランダムな力による効果（ピストンに対する雑音）であり以下の式を満たす。

$$\langle R(0)R(t) \rangle = \frac{2\gamma k_B T \delta(t)}{W} \quad (2.28)$$

ここで k_B はボルツマン定数である。ピストンの自由度を Langevin dynamics による熱浴と共役させることで NPT アンサンブルを満たすトラジェクトリが得られる。この場合熱浴との共役はピストンの 1 自由度のみを介して行われるため、エネルギーの交換が MD 計算のタイムスケールと比較して遅くなる。そこでより効率的に NPT アンサンブルを実現するため、粒子を熱浴と共役させる Nosé-Hoover 法⁷を追加して導入する場合が多い。

2. MD 計算の各手法と自由エネルギーの算出法

2.1 自由エネルギー計算の意義とその困難²

生体系でのさまざまな現象や化学現象を理解する上で最も重要な物理量は、自由エネルギーである。自由エネルギーは状態の出現のしやすさの尺度を与え、さまざまな状態量の分布や平衡の位置を決定する。

これは状態の安定性を示すものであり、大きく 2 つに分けられる。1 つは分子内自由度に依存した安定性であり、もう 1 つは分子間の相互作用に依存した安定性である。前者はタンパク質の折り畳み（フォールディング）などがあり、後者はタンパク質と基質の結合や脂質二重膜を貫通する物質の透過現象などが挙げられる。安定性の議論に限らず、ダイナミクスにおいても自由エネルギー解析は有効である。一般にある状態から別の状態に遷移する系を調べるとき、反応座標における自由エネルギー地形に従って議論することが多い。遷移経路に障壁がある場合、自由エネルギーの大きさから遷移の速度論を展開することができる。

しかし精度よく系の自由エネルギーを見積もることは実際上困難が伴う。最大の困難は系の緩和時間よりも MD 計算で実行できる時間が短いということである。タンパク質のような多自由度系はエネルギー極小状態（準安定状態）が無数に存在し、このエネルギー極小状態を経巡りながらエネルギー最小状態（最安定状態）に緩和するので、非常に長い時間を要する（図 2-2-1-1）。一般に存在するタンパク質は、自然な環境下でミリ秒から秒のオーダーでランダムコイルから特異的な立体構造に折り畳まれる。しかし現在の計算機能力で MD 計算が可能な時間スケールは、アミノ酸が数百個程度の水溶性タンパク質の場合でも 1 マイクロ秒が限界であり、この範囲ではタンパク質のフォールディング現象を正確に記述することはできない。加えて一度エネルギー極小状態に陥ってしまうと、準安定状態であるため抜け出すのにさらに時間を費やすことになる。ゆえに通常の MD 計算時間ではタンパク質のすべての状態が出現しないことから、正確な自由エネルギーを算出することは困難である。

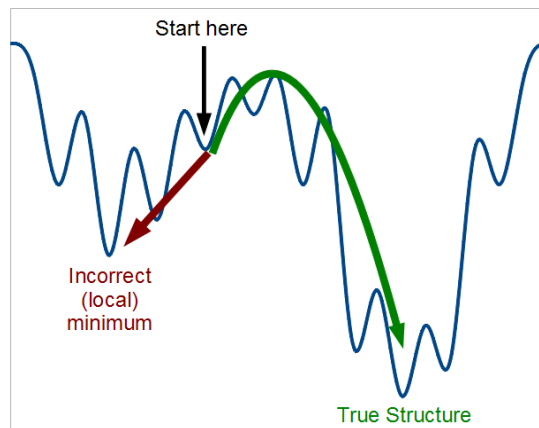


図 2-2-1-1 タンパク質の典型的なエネルギー面⁸

横方向は任意の反応座標、縦方向はエネルギー。一般的にタンパク質にエネルギー面は無数のエネルギー極小状態 (local minimum) が存在する。初期構造から各状態を遷移する途中で、このエネルギー極小状態に陥ってしまうことがある（赤い矢印）。エネルギー最小 (global minimum) の最安定構造に到達するには（緑の矢印）かなりの長時間を要する。⁸ 改変。

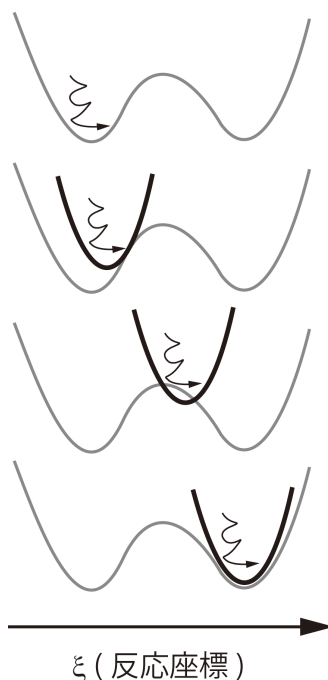
このような困難を克服するためには、工夫された MD 計算手法が必要である。系の緩和を促進しさまざまな状態を出現させ、得られた系の軌跡から統計力学・熱力学量を算出することが可能な方法でなければならない。以下では、これまでに提唱されている多くの MD 計算手法の中から本研究で実際に行った手法を取り上げ、それらの詳細と自由エネルギーを始めとする物理量の期待値を算出する方法を示す。

2.2 アンブレラサンプリング法と WHAM (Weighted Histogram Analysis Method)

例としてタンパク質への基質の結合を考える。現象を記述する反応座標が明確に存在する場合（タンパク質の結合サイトと基質の距離など）これを変数として自由エネルギーを計算することができる。まず反応座標を ξ とし、それに沿った自由エネルギー $F(\xi)$ は、以下のように

$$F(\xi) = -k_B T \ln P(\xi) \quad (2.29)$$

反応座標に沿った系の状態の存在確率 $P(\xi) = N(\xi)/n$ を用いて記述できる。但し、 $N(\xi)$ は MD 計算中に反応座標が ξ となるステップ数であり、 n は全ステップ数である。従って MD 計算でこの分布を計算することにより自由エネルギーを得ることができるが、自由エネルギーの高い領域の分布は、MD 計算の時間内でトラジェクトリが自発的に巡らないか、または非常にまれであるため統計平均が十分でなくなる。そこで図 2-2-2-1 に示すように、記述したい反応座標の領域にトラジェクトリが巡るよう系のポテンシャルエネルギーに人為的なポテンシャル関数（バイアスポテンシャル）を加えた状態で MD 計算を行う。このポテンシャル関数は反応座標上を十分覆うように複数配置する。ここで 1 つのバイアスポテンシャルに対応する MD 計算をウィンドウと呼ぶ。このときウィンドウは周囲のウィンドウと互いの存在確率分布が重なるように配置しなければならない。アンブレラサンプリング法とは、ウィンドウ計算後に得られた確率分布からバイアスポテンシャルによる寄与を差し引いて系の自由エネルギー面を構成する方法である。

図 2-2-2-1 アンブレラサンプリング法の概念図⁹

グレーが系のポテンシャルエネルギー曲線。黒がバイアスポテンシャル。横軸は任意の反応座標 ξ 。矢印は系のトラジェクトリを表す。1 段目はバイアスポテンシャルがない場合。2 段目から 4 段目に掛けてバイアスポテンシャルを系のポテンシャルエネルギー障壁を乗り越えるように配置した場合の様子。バイアスポテンシャルによって、エネルギー障壁を乗り越えるトラジェクトリを実現することが可能である。系の自由エネルギー曲線はこのトラジェクトリから、バイアスポテンシャルの寄与を取り除いて再構成する。⁹ 改変。

アンブレラサンプリングから系の実質的な自由エネルギーを構成する WHAM¹⁰⁻¹³ について述べる。WHAM の目的は、トラジェクトリからアンブレラポテンシャルの影響を取り除き、バイアスのない状態での確率分布 $P_0(\xi)$ (ξ はある反応座標) を求めることである。ここで、ポテンシャルエネルギー $U_0(\mathbf{R})$ をもった系を考える。 \mathbf{R} は原子座標の集合である。反応座標 ξ は、 $\xi(\mathbf{R})$ のようにこの原子座標の関数で表現できる。この系にバイアスポテンシャル $W_i(\xi(\mathbf{R}))$ が課された MD 計算のポテンシャルエネルギーは式(2.30)で表される。

$$U_0(\mathbf{R}) + W_i(\xi(\mathbf{R})) \quad (2.30)$$

この MD 計算からバイアスのある確率分布 $\{P_i^{(b)}(\xi)\}$ が得られる。 $\{P_i^{(b)}(\xi)\}$ は i 番目の MD 計算から反応座標 ξ の規格化されたヒストグラムとして得られる。これに対応する i 番目の MD 計算のバイアスのない確率分布 $P_i^{(u)}(\xi)$ は式(2.31)で表される¹⁰。

$$P_i^{(u)}(\xi) = e^{\beta[W_i(\xi) - f_i]} P_i^{(b)}(\xi) \quad (2.31)$$

f_i は系にバイアスポテンシャル $W_i(\xi(\mathbf{R}))$ を課したことによって得られる自由エネルギーである。この f_i はのちに経験的に決定する。WHAM では N ($i=1, 2, \dots, N$) 個のバイアスポテンシャルの掛かった MD 計算のトラジェクトリを用いて、バイアスのない確率分布 $P_0(\xi)$ を得る。

$$P_0(\xi) = C \sum_{i=1}^N P_i(\xi) P_i^{(u)}(\xi) \quad (2.32)$$

ここで、 C は規格化定数である。重み $P_i(\xi)$ は $P_0(\xi)$ の規格化のために必要である。

$$\sum_{i=1}^N P_i(\xi) = 1 \quad (2.33)$$

バイアスのない確率分布 $P_0(\xi)$ の統計誤差を最小にするため、任意の反応座標 ξ において式(2.34)を課す。

$$\frac{\partial(\sigma^2[P_0(\xi)])}{\partial P_i} = 1 \quad (2.34)$$

このとき、 $P_0(\xi)$ は以下の式を満たす。

$$P_0(\xi) = C \sum_{i=1}^N \frac{n_i e^{-\beta[W_i(\xi) - f_i]}}{\sum_{j=1}^N n_j e^{-\beta[W_j(\xi) - f_j]}} P_i^{(u)}(\xi) \quad (2.35)$$

$$= C \sum_{i=1}^N \frac{n_i}{\sum_{j=1}^N n_j e^{-\beta[W_j(\xi) - f_j]}} P_i^{(b)}(\xi) \quad (2.36)$$

ここで、バイアスポテンシャルによる自由エネルギー $-f_i$ は自己無撞着に求めることができる。自由エネルギー $-f_i$ の定義式は以下である。

$$e^{-\beta f_i} = \int d\xi P_0(\xi) e^{-\beta W_i(\xi)} \quad (2.37)$$

$$= C \int d\xi \sum_{i=1}^N \frac{n_i e^{-\beta W_i(\xi)}}{\sum_{j=1}^N n_j e^{-\beta [W_j(\xi) - f_j]}} P_i^{(b)}(\xi) \quad (2.38)$$

ここで、自由エネルギーの集合 $\{f_i\}$ は式(2.37)の左辺と式(2.38)の右辺の両方に現れており、反復計算によって解くことができる。初期値 $\{f_i^0\}$ を式(2.38)の右辺に代入することで、式(2.37)の左辺から $\{f_i^1\}$ が得られる。これを同様に式(2.38)の右辺に代入することで、式(2.37)の左辺から $\{f_i^2\}$ が得られる。これを $\{f_i\}$ が収束するまで繰り返す。規格化定数 C を消去するため、全ての $\{f_i\}$ からオフセット f_0 を差し引く。これは、確率分布 $P_0(\xi)$ を規格化したことと等価である。収束したバイアスポテンシャルによる自由エネルギー $\{f_i\}$ を式(2.36)に代入することで、バイアスのない確率分布 $P_0(\xi)$ が得られる。さらにこの確率分布 $P_0(\xi)$ を式(2.39)に代入することで、アンブレラサンプリングからバイアスポテンシャルの寄与を取り除いた系の自由エネルギー $F_0(\xi)$ を得ることができる。

$$F_0(\xi) = -k_B T \ln P_0(\xi) \quad (2.39)$$

2.3 拡張アンサンブル法と REMD 法

一定温度の MD 計算では、ボルツマン因子 $W_B(E; T)$ に基づいて状態が発生するため、エネルギー E の分布 $P_B(E; T)$ は状態密度 $n(E)$ を用いて

$$P_B(E; T) \propto n(E) W_B(E; T) \equiv n(E) e^{-\beta E} \quad (2.40)$$

と与えられる。ここで β は逆温度である。状態密度はエネルギーの増大に伴って急激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ので、カノニカル分布は一般にベル型をしている(図 2-2-3-1 点線)。

これに対し拡張アンサンブル法は、非ボルツマン因子による人為的なアンサンブルに基づきポテンシャルエネルギー空間上の一次元ランダムウォークを実現する手法である¹⁴⁻¹⁵。代表的な例として、マルチカノニカル法¹⁶ や焼き戻し法¹⁷ がある。マルチカノニカル法では、各状態は非ボルツマン的な重み因子

$W_{\text{multicanonical}}(E)$ に比例して発生し、ポテンシャルエネルギーの一様分布 $P_{\text{multicanonical}}(E)$ が得られる。

$$P_{\text{multicanonical}}(E) \propto n(E) W_{\text{multicanonical}}(E) \equiv \text{const.} \quad (2.41)$$

すなわち、系は人為的なアンサンブルの下でポテンシャルエネルギー空間の一次元ランダムウォークが実現される(図 2-2-3-1 実線)。

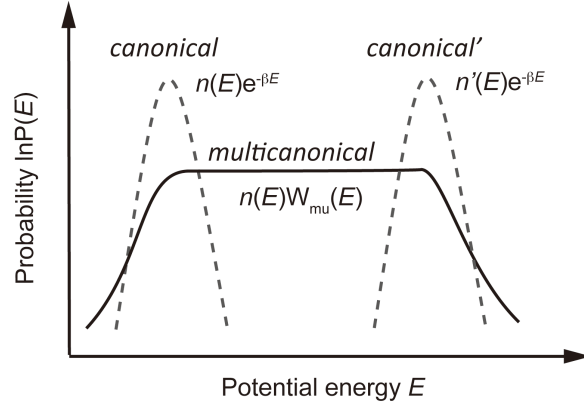


図 2-2-3-1 マルチカノニカルおよびカノニカルアンサンブルのポテンシャルエネルギー確率分布の模式図¹⁸

横軸はポテンシャルエネルギー，縦軸はポテンシャルエネルギーの確率分布の対数．実線はマルチカノニカル法における確率分布，点線はカノニカルアンサンブルにおける確率分布を示す．¹⁸ 参考．

このときポテンシャルエネルギーはさまざまな値を取ることができるので，系がポテンシャルエネルギー極小状態に陥ってもいずれそこから抜け出すことが期待される．よって従来の方法と比べより広範囲な構造空間のサンプリングが実現できる．この手法では，はじめに $W_{\text{multicanonical}}(E)$ を決定するため短い試行 MD 計算を行う必要がある．通常は非 Boltzmann 的な重み因子として，状態密度 $n(E)$ の逆数に比例する値を用いる．しかしながら， $W_{\text{multicanonical}}(E)$ を決定する操作は系が複雑になるにつれて難しくなり，十分な経験がないと適切な値の決定ができない．

このような中，1999 年に杉田と岡本は REMD 計算¹⁹を開発した．以下ではこの方法の詳細を述べる．質量が $m_k (k=1, \dots, N)$ の N 個の粒子で構成され，座標が $q \equiv \{q_1, \dots, q_N\}$ ，運動量が $p \equiv \{p_1, \dots, p_N\}$ である系を考える．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 $E(q)$ を用いて

$$H(q, p) = K(p) + E(q) \quad (2.42)$$

と与えられる．ここで系の温度 T での運動エネルギー $K(p)$ は

$$\sum_{k=1}^N \frac{p_k^2}{2m_k} \quad (2.43)$$

であり，その期待値は以下の通りである．

$$\langle K(p) \rangle_T = \left\langle \sum_{k=1}^N \frac{p_k^2}{2m_k} \right\rangle_T = \frac{3}{2} N k_B T \quad (2.44)$$

この手法では，相互作用しない互いに独立な M 個の系のコピー（レプリカ）を用意し，それらが取る M 個の温度 $T_m (m=1, \dots, M)$ を用意する．これらをひとつにまとめた系を REMD での系と定義する．レプリカを表すラベルを i, j とし，温度を表すラベルを m, n とすると，これらは 1 対 1 に対応する．そこで，以下のような置換関数 σ およびその逆関数 σ^{-1} を導入する．

$$\begin{aligned} i &= \sigma(m), m = \sigma^{-1}(i) \\ j &= \sigma(n), n = \sigma^{-1}(j) \end{aligned} \quad (2.45)$$

温度 T_m でのレプリカ i の状態 $x_m^{[i]} \equiv (q^{[i]}, p^{[i]})_m$ (添字は上付きがレプリカ, 下付きが温度) によって系の状態 X は, 以下のように指定される.

$$X = (x_1^{[\sigma(1)]}, \dots, x_M^{[\sigma(M)]}) = (x_{\sigma^{-1}(1)}^{[1]}, \dots, x_{\sigma^{-1}(M)}^{[M]}) \quad (2.46)$$

このときレプリカ同士は相互作用しないので, 状態 X の重み因子 $W_{REM}(X)$ は各レプリカ (またはそれに対応する温度) のボルツマン因子の積で与えられる.

$$W_{REM}(X) = \exp\left[-\sum_{i=1}^M \beta_{\sigma^{-1}(i)} H(q^{[i]}, p^{[i]})\right] = \exp\left[-\sum_{m=1}^M \beta_m H(q^{[\sigma(m)]}, p^{[\sigma(m)]})\right] \quad (2.47)$$

ここで, $q^{[i]} \equiv \{q_1^{[i]}, \dots, q_N^{[i]}\}$ と $p^{[i]} \equiv \{p_1^{[i]}, \dots, p_N^{[i]}\}$ は, レプリカ i における N 個の粒子の位置と運動量である. REMD 法は, 異なる温度をもつ系のコピー (レプリカ) の並列計算を実行し, 計算途中で詳細釣り合いを満たすように系の温度を交換することで高温では構造変化を実現し, 低温ではエネルギーの低い最安定状態を探索する方法である. REMD 計算は具体的には以下のように行われる (図 2-2-3-2). まず系の M 個のレプリカは指定したステップの間, 通常温度一定の MD 計算を並列に実行する. 次にレプリカの交換操作をある手続きに従って行う. そしてこの交換操作が終了した後は, また通常温度一定の MD 計算を並列して行う. このサイクルを繰り返すことにより, ある 1 つのレプリカを見ていけば, 温度軸上を動き回り, 長時間の振る舞いはランダムウォークと見なすことができる. それによって温度と強い相関のあるポテンシャルエネルギーも, その振る舞いはランダムウォークと見なせる.

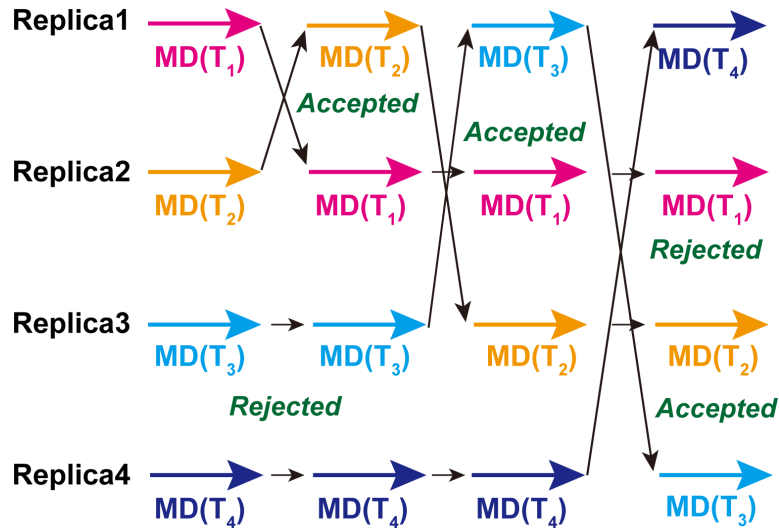


図 2-2-3-2 REMD 計算の模式図

横方向は MD 計算の時間発展, 縦方向はレプリカの並列を表す. 例として 4 つのレプリカを用いて 3 回のレプリカ交換を行った場合を示した. レプリカの色が MD 計算での温度を表している. レプリカ交換の判定は隣り合う温度ペアに対してのみ行われる. 2 種類の隣り合うペア (上限と下限の温度を除いて, 1 つの温度に対して隣り合う温度は高温側と低温側の 2 つ存在する) に対して交互にレプリカの交換を試行する.

レプリカの交換操作についてくわしく見る。ここで温度がそれぞれ T_m と T_n に対応するレプリカ i と j の交換を考える。すなわち、系の状態は以下のように変更される。

$$X = (..., x_m^{[i]}, ..., x_n^{[j]}, ...) \rightarrow X' = (..., x_m^{[j]}, ..., x_n^{[i]}, ...) \quad (2.48)$$

ここで、 $x_m^{[j]}, x_n^{[i]}$ はそれぞれ以下のように定義される。

$$\begin{aligned} x_m^{[j]} &= \{q^{[j]}, p^{[j]}\}_m \\ x_n^{[i]} &= \{q^{[i]}, p^{[i]}\}_n \end{aligned} \quad (2.49)$$

交換されるとレプリカと温度の対応関係が変更されるので、新たな置換関数 σ' を導入する。

$$\begin{aligned} j &= \sigma'(m), \quad m = \sigma'^{-1}(j) \\ i &= \sigma'(n), \quad n = \sigma'^{-1}(i) \end{aligned} \quad (2.50)$$

式()の $p^{[i]}, p^{[j]}$ は交換された後のレプリカ中の運動量であり、以下のように割り当てられる。

$$\begin{aligned} p^{[i]} &\equiv \sqrt{\frac{T_n}{T_m}} p^{[i]} \\ p^{[j]} &\equiv \sqrt{\frac{T_m}{T_n}} p^{[j]} \end{aligned} \quad (2.51)$$

これはレプリカ中の粒子の速度が、交換する 2 つの温度の比の平方根によってスケールされることを意味しており、式(2.44)を満たしている。このとき、交換操作によって系が平衡状態へ収束しなければならない。一般に平衡状態に緩和していると言える必要条件是以下の 2 つである。

(a) 詳細釣り合いの条件

系が状態 X から X' へ遷移するには、遷移確率 $w(X \rightarrow X')$ についての関係式

$$W_{REM}(X) w(X \rightarrow X') = W_{REM}(X') w(X' \rightarrow X) \quad (2.52)$$

を満たす。

(b) エルゴード条件

任意の 2 つの状態 X から X' の間の遷移確率がゼロではないか、あるいは有限個のゼロでない遷移確率の式で表される。

式(2.43), (2.44), (2.45), (2.47), (2.48)を用いると、詳細釣り合いの条件式は以下のように記述される。

$$\begin{aligned}
 \frac{w(X \rightarrow X')}{w(X' \rightarrow X)} &= \frac{W_{REM}(X')}{W_{REM}(X)} \\
 &= \frac{\exp[-\beta_m (K(p^{[j]}) + E(q^{[i]}))] \cdot \exp[-\beta_n (K(p^{[i]}) + E(q^{[j]}))] }{\exp[-\beta_m (K(p^{[i]}) + E(q^{[i]}))] \cdot \exp[-\beta_n (K(p^{[j]}) + E(q^{[j]}))] } \\
 &= \exp \left[\begin{aligned} &-\beta_m \frac{T_m}{T_n} K(p^{[j]}) - \beta_n \frac{T_n}{T_m} K(p^{[i]}) + \beta_m K(p^{[i]}) + \beta_n K(p^{[j]}) \\ &-\beta_m \{E(q^{[j]}) - E(q^{[i]})\} - \beta_n \{E(q^{[i]}) - E(q^{[j]})\} \end{aligned} \right] \\
 &= \exp[-(\beta_n - \beta_m)(E(q^{[i]}) - E(q^{[j]}))] \\
 &= \exp(-\Delta)
 \end{aligned} \tag{2.53}$$

このとき第 3 式から第 4 式で運動量の項がキャンセルされている。しかしながら、詳細釣り合いの条件を課してもなお、遷移確率 $\omega(X \rightarrow X')$ の取り方は一意的には決まらず任意性が残る。そこで、もっとも一般的な系に適用できて実用上も良く使われている遷移確率の取り方にメトロポリス法 (Metropolis method) がある。メトロポリス法では、現在の系の状態 X から新たな状態 X' を作り、その状態 X' を実際に次のステップとして採用するか否かを以下の遷移確率に従って決める。

$$\begin{aligned}
 \omega(X \rightarrow X') &= w(x_m^{[i]} | x_n^{[j]}) \\
 &= \min \left(1, \frac{W_{REM}(X')}{W_{REM}(X)} \right) \\
 &= \begin{cases} 1 & \Delta \leq 0 \\ \exp(-\Delta) & \Delta > 0 \end{cases}
 \end{aligned} \tag{2.54}$$

途中で式(2.53)を用いた。すなわち、 $\Delta \leq 0$ のとき新たな状態 X' を確率 1 で受け入れ、 $\Delta > 0$ のとき新たな状態 X' を確率 $\exp(-\Delta)$ で受け入れる。実際の MD 計算では、 $[0,1]$ の乱数を発生させ、その大小を比較して受け入れの決定を判断する。メトロポリス法の遷移確率は、(a)の詳細釣り合いの条件を満たしており、 $\Delta > 0$ のとき遷移確率をゼロとしていないため、(b)のエルゴード性を保証している。

交換するレプリカペアは、隣り合った温度対に対応するレプリカ同士に限定する。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。マルチカノニカル法や焼き戻し法に比べて REMD 法は、重み因子が始めから分かっているため、その決定を行う必要がない。また各レプリカを計算機の各 CPU に対応させれば、座標と運動量の代わりに温度を交換させるだけであるので CPU 間の情報通信量を最小にすることができる。そのため並列化計算に適した方法である。

カノニカルアンサンブルでのレプリカに割り振った温度 T_m ($m = 1, \dots, M$) の物理量の期待値 A は、

$$\langle A \rangle_{T_m} = \frac{1}{N_{sim}} \sum_{t=1}^{N_{sim}} \sum_{i=1}^M A[x_{\sigma^{-1}(i;t)}^{[i]}(t)] \delta_{\sigma^{-1}(i;t), m} \tag{2.55}$$

N_{sim} はレプリカの物理量 A を算出した総数、 $\sigma^{-1}(i;t)$ は t 回目の算出の置換関数、 δ_{ij} はクロネッカーのデルタ関数である。すなわち、温度 T_m のレプリカにおける物理量 A の算術平均である。レプリカに割り振っ

た温度 T_m ($m = 1, \dots, M$) の任意の中間温度での物理量 A の期待値は, WHAM の拡張版である多ヒストグラム再重法¹⁹⁻²¹を用いて見積もることができる. R 個の温度 m ($1, \dots, R$) に対応した, R 個の互いに独立なレプリカがあり, 任意の温度での物理量 A の期待値は,

$$\langle A \rangle_T = \frac{\sum_E A(E) P(E; \beta)}{\sum_E P(E; \beta)} \quad (2.56)$$

ここで分布 $P(E; \beta)$ は,

$$P(E; \beta) = \frac{\sum_{m=1}^R g_m^{-1} N_m(E) e^{-\beta E}}{\sum_{m=1}^R n_m g_m^{-1} e^{f_m - \beta_m E}} \quad (2.57)$$

$$e^{-f_m} = \sum_E P(E; \beta_m)$$

となる. 但し, $g_m^{-1} = 1 + 2\tau_m$ であり τ_m は温度 T_m での積分された自己相関時間²²である. ある温度 T において物理量 A を色々な時刻で観測するとき, その規格化された自己相関関数は $\phi_A(t) = C_A(t)/C_A(0)$ で与えられる. ここで $C_A(t) = \langle A(0)A(t) \rangle_T - \langle A(0) \rangle_T^2$ である. 積分された自己相関時間は $\tau_A = \int_0^\infty \phi_A(t) dt$ で定義される. ここでは, 物理量としてポテンシャルエネルギー E を考える. しかし, タンパク質の系ではこの量は幅広い温度において温度に依らず, 定数においても多次元 WHAM の結果に影響がないことが示されており¹³, 実際に式(2.57)でも g_m^{-1} は消去された形で計算される. $N_m(E)$ は温度 T_m でのエネルギー E のヒストグラム, n_m は得られたデータの総数, f_m は無次元化されたヘルムホルツの自由エネルギー (すなわち, β_m とヘルムホルツ自由エネルギーの積) である. これを繰り返すことによって自己無撞着に解く.

2.4 非平衡統計力学からのアプローチ

2.4.1 Targeted MD 法

タンパク質や基質の構造変化を外部からの力を加えることによって誘起し, それによるタンパク質の構造変化を調べる MD 計算方法が提案されている²³⁻²⁵. 特に, 対象としたタンパク質の始状態 (初期構造) と終状態 (Targeted 構造) を定義し, 初期構造に外力を加えることで Targeted 構造までの構造変化を実現する手法が Targeted MD 法²⁶である. このとき系には, 外部からのバースポテンシャル U_{TMD} が掛かる.

$$U_{\text{TMD}} = \frac{1}{2} \frac{k}{N} [RMSD(t) - RMSD'(t)]^2 \quad (2.58)$$

k はばね定数, N は選択した原子数, $RMSD(t)$ は t ステップ目の構造と Targeted 構造との RMSD 値, $RMSD'(t)$ は初期構造とターゲット構造を, 指定した MD 計算時間で遷移させるときの t ステップ目に割り振られた RMSD 値. RMSD は 3.1 節で説明をしている. 通常は始状態と終状態を線形に繋ぐ. 各原子には, x, y, z 軸方向にそれぞれの位置で微分した力が加えられる.

2.4.2 Jarzynski 等式と Crooks の揺らぎの定理を用いる方法

Targeted MD 法では外力によって構造変化を引き起こした。さらに一步進んで、構造変化に伴う自由エネルギー変化を得ることを考える。ここで問題となるのは、Targeted MD では外部からの力が加わっており、非平衡不可逆過程ということである。従って平衡状態での物理量である自由エネルギー変化を Targeted MD から直接算出するのは一見困難なように思われる。しかしながらこの問題を解決する理論が 1997 年に Jarzynski によって提唱された²⁷。この理論を用いると、構造変化に伴う平衡過程の自由エネルギー変化 ΔF が、その構造変化を引き起こす非平衡不可逆過程における仕事量 W を指数関数の肩に乗せたもののアンサンブル平均によって記述される。

$$\Delta F = -k_B T \ln \left\langle \exp \left[-\frac{W}{k_B T} \right] \right\rangle_{\text{nonequilibrium}} \quad (2.59)$$

ここで、右辺のアンサンブル平均は Targeted MD による非平衡過程に対して取られる。すなわち、Targeted MD を行う際に系に対して行った仕事量 $W(\xi)$ を求め、多数の構造変化過程のトラジェクトリによるアンサンブル平均を取ることによって、平衡過程での自由エネルギー変化を得ることができる²⁶。

それでは実際の手続きを見てみる。バイアスポテンシャルとして式(2.58)を用いたときの非平衡過程での仕事量 W は

$$W = \int_0^t \frac{\partial H_\lambda(\mathbf{x}(t))}{\partial \lambda} \dot{\lambda} dt \quad (2.60)$$

$\mathbf{x}(t)$ は全原子に対する位相空間の時刻 t での点、 λ は始状態と終状態を繋ぐコントロールパラメータ、 H は系のハミルトニアン。Targeted MD では、 λ が $RMSD$ に相当し、 H が U_{TMD} に置き換わる（これ以外のエネルギー項は λ に依らない）。Jarzynski 等式をよく見ると、 β を逆温度として

(i) 「よく起こる、小さい W のトラジェクトリ」は大きな $\exp[-\beta W]$ を与える。

(ii) 「まれにしか起こらない、大きな W のトラジェクトリ」は小さな $\exp[-\beta W]$ を与える。

これらが釣り合って ΔF が決まるが、(ii)の「まれにしか起こらないトラジェクトリ」をいかにサンプリングできるかが Jarzynski 等式を用いるかどうかの判断基準となる。その意味では Jarzynski 等式を用いる場合、かなり多くの試行回数を行う必要がある。試行回数が少ないと上で述べたように $\exp[-\beta W]$ を過大評価してしまう²⁸。

そこで Crooks の揺らぎの定理 (Crooks fluctuation theorem, CFT)²⁹⁻³⁰を使うことも考えられる。この方法では forward (始状態から終状態) と backward (終状態から始状態) の両方向の操作を行い、仕事量 W のアンサンブル平均ではなく、その分布の形を用いる。

$$\frac{P_f[W]}{P_b[-W]} = \exp[\beta(W - \Delta G)] \quad (2.61)$$

$P_f[W]$ は forward へ遷移させたときの仕事量 W の分布、backward へ遷移させたときの仕事量 $-W$ の分布であり、2つの分布の交点での仕事量 W が始状態と終状態の自由エネルギー差 ΔG に等しいことを主張している。

CFT は Jarzynski 等式と比較して MD 計算の試行回数は少なく済み、精度よく見積もることが期待できる³¹。しかし両方向への仕事量の分布を得るには、多数の MD 計算が要求されることから決して計算コストは小さくない。また Jarzynski 等式と CFT のいずれも自由エネルギー差を見積もることは可能であるが、外力による構造変化であるためその構造変化の遷移経路に関する議論ができない。タンパク質では構造変化の遷移経路が、その機能を理解する上で重要な要素の一つであり、その観点からもこれらの方法は本研究には適していないと判断した。

3. MD 計算結果の解析方法

3.1 RMSD (root-mean-square deviation)

2 つのタンパク質立体構造の類似性の比較には、RMSD (root-mean-square deviation, 平均 2 乗偏差) が広く用いられる。\$N\$ 個の原子からなるタンパク質 \$a, b\$ の構造を直交座標で以下のように表す。

$$\mathbf{r}_a(t) = (x_1^a(t), y_1^a(t), z_1^a(t), \dots, x_N^a(t), y_N^a(t), z_N^a(t)) \quad (2.62.a)$$

$$\mathbf{r}_b(t) = (x_1^b(t), y_1^b(t), z_1^b(t), \dots, x_N^b(t), y_N^b(t), z_N^b(t)) \quad (2.62.b)$$

ここで \$\mathbf{x}_i^j(t)\$ は時刻 \$t\$ でのタンパク質 \$j\$ の \$i\$ 番目の原子の直交座標を表す。通常 2 つのタンパク質の立体構造の比較は、構造の並進・回転に起因する原子の位置のずれを消去するために、始めに 2 つの構造の重ね合わせを行ったあとに行う。そのため、式(2.62.a), (2.62.b)で定義された座標も、一方の立体構造に他方を重ね合わせてあると考える。すると RMSD は次式で定義される。

$$RMSD(t) = \sqrt{\frac{\sum_{i=1}^N \{(x_i^a(t) - x_i^b(t))^2 + (y_i^a(t) - y_i^b(t))^2 + (z_i^a(t) - z_i^b(t))^2\}}{N}} \quad (2.63)$$

RMSD は直交座標に限らず、系の構造を指定する任意の変数で計算できる。ただし、立体構造の重ね合わせは直交座標で表現された立体構造間の比較を行うときに必要となる特別な操作である。原子間距離や二面角は、座標空間に置かれたタンパク質の位置や配向とは無関係に、タンパク質構造を特徴付ける座標である。このような座標を内部座標と呼ぶ。内部座標を用いて RMSD を計算するときには、立体構造の重ね合わせを行う必要はない。

3.2 平均力ポテンシャル (potential of mean force, PMF)

ある反応座標 \$\xi\$ で規定された座標空間 \$\Gamma(\xi)\$ の確率分布 \$P(\Gamma(\xi))\$ は、ボルツマン因子に基づいて座標空間 \$\Gamma\$ での積分として次のように記述できる。

$$P(\Gamma(\xi)) = \frac{\int \delta(\Gamma - \Gamma(\xi)) \exp[-\beta U(\Gamma)] d\Gamma}{\int \exp(-\beta U(\Gamma)) d\Gamma} \quad (2.64)$$

$$= \langle \delta(\Gamma - \Gamma(\xi)) \rangle_{U(\Gamma)} \quad (2.65)$$

但し、運動量での積分はすでに行っているとし、ポテンシャルエネルギー U は座標空間 Γ のみに依存するとする。 δ はDiracのデルタ関数であり、座標空間 Γ が $\Gamma(\xi)$ と一致するものだけを抜き出す関数である。式(2.65)はポテンシャルエネルギー $U(\Gamma)$ の下で座標空間 $\Gamma(\xi)$ のアンサンブル平均をとることを意味する。このときPMFである $F(\xi)$ は

$$F(\Gamma(\xi)) = -\frac{1}{\beta} \ln P(\Gamma(\xi)) \quad (2.66)$$

と定義できる。式(2.64), (2.65)から明らかなように、 $P(\Gamma(\xi))$ は ξ 以外のすべての自由度を積分して得られているので、 $F(\xi)$ は状態 ξ における自由エネルギーとも見なせる。式(2.66)によって ξ で表されるさまざまな状態における自由エネルギー、すなわち ξ が張る空間における自由エネルギー地形を算出することができる。特に式(2.65)で示した座標空間 $\Gamma(\xi)$ のアンサンブル平均で確率分布 $P(\Gamma(\xi))$ が得られることは重要である。すなわち、位相空間での積分が困難であっても、MD計算によってさまざまな状態がサンプリングされ十分なアンサンブルが得られていれば、自由エネルギー地形が得られることを示している。

3.3 主成分分析 (principle component analysis, PCA)

タンパク質の立体構造の揺らぎは、機能を発現する上で重要であると考えられる。主成分分析はタンパク質の熱揺らぎの特徴を得るのに有用な方法である³²⁻³³。揺らぎの特徴を決めているのはタンパク質の自由エネルギー地形であり、反対に揺らぎを詳細に調べることで自由エネルギー地形が分かる。以下では、主成分分析がなぜタンパク質の揺らぎと自由エネルギー地形を解析するのに有用であるのか述べ、その方法の詳細を説明する。

3.3.1 揺らぎの異方性⁹

タンパク質はアミノ酸がペプチド結合で繋がった鎖状の高分子であり、天然の状態ではアミノ酸配列に特異的な立体構造を形成している。系の平衡状態では、天然構造の近傍で揺らいでいる。タンパク質の平衡状態での揺らぎは、ある構造のまわりで揺らいでいると見なしてよく、その揺らぎは異方性が高い。この異方的な揺らぎを表すため、タンパク質の揺らぎを記述するには各原子の直交座標ではなく、直交座標の線形結合で表される集団座標を用いるのが適している。

3.3.2 主成分分析の詳細

主成分分析は、多くの変数をできるだけ情報を失わずに少数個のパラメータ（主成分）で代表させる方法である³⁴⁻³⁶。そのため、多数のパラメータを含む複雑な現象を解析する際によく用いられる。温度 T での構造集合の中の構造 i の座標を

$$\mathbf{q}_i = (q_1^i, q_2^i, \dots, q_M^i) \quad (2.67)$$

で表す。 M は構造を表すための自由度の数である。全原子の直交座標を使うと $M=3N$ である。ここで分散共分散行列として

$$C_{jk} = \langle q_j q_k \rangle - \langle q_j \rangle \langle q_k \rangle \quad (2.68)$$

を定義する．ここで， C_{jk} は行列の (j,k) 成分である．平均操作 $\langle \dots \rangle$ は，構造集団での単純平均を表す．

$$\langle q_j q_k \rangle = \sum_{i=1}^{N_{all}} \frac{q_j^i q_k^i}{N_{all}} \quad (2.69)$$

$$\langle q_j \rangle = \sum_{i=1}^{N_{all}} \frac{q_j^i}{N_{all}} \quad (2.70)$$

N_{all} は構造集団中に含まれるすべての構造数である．次にこの分散共分散行列を対角化して M 個の固有ベクトル $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M)$ と固有値 $(\lambda_1, \lambda_2, \dots, \lambda_M)$ が得られる．ここで得られた固有ベクトルの組は，集団座標と見なすことができる．ある構造 i を固有ベクトル \mathbf{v}_j の上に射影すると，その値は

$$c_j^i = \mathbf{v}_j \cdot (\mathbf{q}_i - \langle \mathbf{q} \rangle) \quad (2.71)$$

となる．上の式で \mathbf{q}_i から平均構造 $\langle \mathbf{q} \rangle$ を引くのは，座標軸の原点を分布の平均値に設定するためである．2つの固有ベクトル \mathbf{v}_j と \mathbf{v}_k で構成された2次元平面上に構造 i を射影するには，この2つの固有ベクトルに対し式(2.71)を適用する．結果として2個の数 (c_j^i, c_k^i) が得られる．平面の中で座標 (c_j^i, c_k^i) が構造 i の点となる．この操作をすべての構造に対して行くと，平面上には構造集団の揺らぎを表す分布が得られる．また分布の標準偏差はその固有値 λ_i に等しい．そのため最も大きい固有値を第一主成分，次に大きい固有値を第二主成分とすると，固有値が大きい順に構造変化の揺らぎが大きくなっている．従って，大きな固有値を持つ固有ベクトルで構成された分布はその系の構造変化の特徴をよく反映している．そのため，得られた分布から自由エネルギー地形を求めることで，構造の安定性を調べることができる．

3.4 構造のクラスター分析法³⁷

クラスター分析とは，データがある方針のもとで，類似しているいくつかのかたまり（クラスター）にまとめる方法である．対象間の距離を定義して，距離の近さによって対象を分類する手続きを行う．クラスター分析には，階層的クラスタリングと非階層的クラスタリング（分割最適化クラスタリング）の2種類がある．

3.4.1 階層的クラスタリング

この手法は， N 個の対象からなるデータが与えられたとき，1 個の対象だけを含む N 個のクラスター $C_i (i=1, 2, \dots, N)$ がある状態をまず作成する．この状態から始めて，対象 x_1, x_2 間の距離 $d(x_1, x_2)$ からクラスター間の距離 $d(C_1, C_2)$ を計算し，最もこの距離に近い2つのクラスターを統合する．この操作をすべての対象が1つのクラスターに統合されるまで繰り返すことで，対象としたデータの階層的構造を得る（図

2-3-4-1). クラスター間の距離 $d(C_1, C_2)$ には複数の定義がある. 以下に代表的な 2 つの例(a), (b)を示す.

(a) Nearest neighbor method

$$d(C_1, C_2) = \min_{x_1 \in C_1, x_2 \in C_2} d(x_1, x_2) \quad (2.72)$$

(b) Further neighbor method

$$d(C_1, C_2) = \max_{x_1 \in C_1, x_2 \in C_2} d(x_1, x_2) \quad (2.73)$$

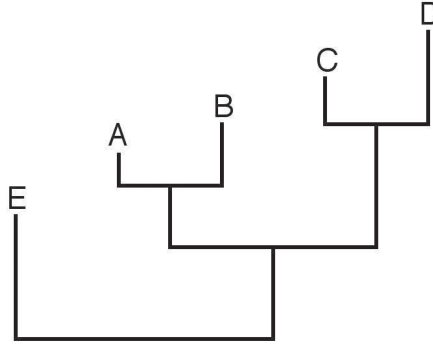


図 2-3-4-1 階層的リングによって得られる樹形図

始めの対象データ A, B, C, D, E からデータの統合を行って, 最終的に 1 つのクラスターに統合させる. この段階で対象としたデータの階層的構造が得られる.

3.4.2 非階層的 (分割最適化) クラスタリング

この手法では, 分割の良さの評価関数を定義してその評価関数を最適にする分割を探索する (図 2-3-4-2). 可能な分割の総数はデータの総数 N に対して指数関数的に増加するので, 実際は準最適解を求めることになる. 本研究でも用いた代表的な k -means 法³⁸は, セントロイド (クラスターの重心点) c_i をクラスターの代表点として評価関数

$$\sum_{i=1}^k \sum_{x \in c_i} [d(x, c_i)]^2 \quad (2.74)$$

を最小にするように最適解を探索する. ここで, k はあらかじめ決定したクラスター数 (セントロイド数) を表す. しかし, この手法では局所最適解しか得られない (初期値に依存する) ため, ランダムに初期値を変更する必要がある. 以下の手順で行う.

1. k 個のセントロイド $c_i (i=1, 2, \dots, k)$ をランダムに選択.
2. 各対象を最も近いセントロイドに割り当てクラスターを形成する.
3. すべての対象の割り当てが 1 つ前のステップと等しい場合終了する. 割り当てが異なる場合, 各クラスターの重心を求め, これを新たなセントロイドとする.

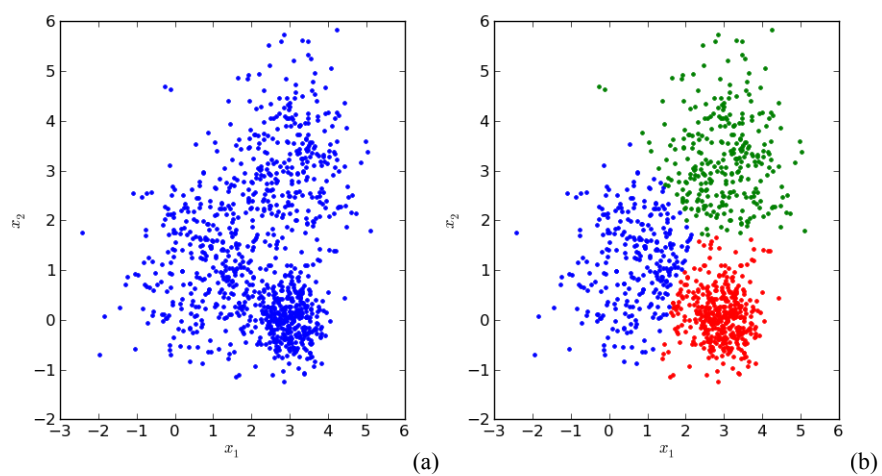


図 2-3-4-2 k -means クラスタリングによる各クラスターへの分割³⁹

(a) クラスタリング前の元となるデータの 2 次元 (x_1, x_2) 分布. (b) (a) のデータに対して行った k -means クラスタリングの結果. 3 つのクラスター (赤, 緑, 青の集団) に分割されている.³⁹ 改変.

引用文献

1. M. P. Allen; Tildesley, D. J., *Computer Simulation of Liquids*. Oxford university press: Clarendon Press, 1989; p 408.
2. 岡崎進, 吉井範行, コンピュータ・シミュレーションの基礎 (第2版). 化学同人: 2000.
3. Bart Markvoort, P. S., Peter Hilbers, Koen Pieterse
PumMa.
<http://cbio.bmt.tue.nl/pumma/index.php/Theory/Potentials> (accessed 3, Dec., 2014).
4. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.
5. Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177.
6. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R., Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613-4621.
7. Hoover, W., Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695-1697.
8. Clarkson, M. W. Conformational Flux.
<http://conflux.mwclarkson.com/2011/09/foldit-pmv-pr/> (accessed 3, Dec., 2014).
9. 後藤祐児, 桑島邦博, 谷澤克行【編】, タンパク質科学 構造・物性・機能. 化学同人: 2005.
10. Souaille, M.; Roux, B. t., Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* **2001**, *135*, 40-57.
11. Ferrenberg, A.; Swendsen, R., New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.* **1988**, *61*, 2635-2638.
12. Ferrenberg, A.; Swendsen, R., Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* **1989**, *63*, 1195-1198.
13. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011-1021.
14. 岡本祐幸, 多自由度複雑系で有効な拡張アンサンブル法の開発. *分子研レターズ* **49**.
15. 小久保裕功, 岡本祐幸, 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. *統計数理* **2006**, *54* (2), 211-222.
16. Berg, B.; Neuhaus, T., Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Phys. Rev. Lett.* **1992**, *68*, 9-12.
17. Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N., New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *J. Chem. Phys.* **1992**, *96*, 1776.
18. 森義治, 拡張アンサンブル法とその生体分子系の定圧定温シミュレーションへの適用. 名古屋大学博士学位論文 **2010 年度 (平成 23 年 3 月 25 日)**, 甲 9189.
19. Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141-151.
20. Sugita, Y.; Okamoto, Y., Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Chem. Phys. Lett.* **2000**, *329*, 261-270.
21. Sugita, Y.; Kitao, A.; Okamoto, Y., Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000**, *113*, 6042.
22. 杉田有治, 光武亜代理, 岡本祐幸, 拡張アンサンブル法によるタンパク質の折り畳みシミュレーション. *日本物理学会誌* **2001**, *56*, 591-599.
23. Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K., Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J. Chem. Phys.* **2003**, *119*, 3559.
24. Cuendet, M. a.; Michielin, O., Protein-protein interaction investigated by steered molecular dynamics: the TCR-pMHC complex. *Biophys. J.* **2008**, *95*, 3575-3590.
25. Kosztin, I.; Barz, B.; Janosi, L., Calculating potentials of mean force and diffusion coefficients from nonequilibrium processes without Jarzynski's equality. *J. Chem. Phys.* **2006**, *124*, 64106.
26. Fornili, A.; Giabbai, B.; Garau, G.; Degano, M., Energy Landscapes Associated with Macromolecular Conformational Changes from Endpoint Structures. *J. Am. Chem. Soc.* **2010**, *132*, 17570-17577.
27. Jarzynski, C., Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **1997**, *78*, 0-3.
28. Liphardt, J.; Dumont, S.; Smith, S. B.; Tinoco, I.; Bustamante, C., Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science (New York, N.Y.)* **2002**, *296*, 1832-1835.
29. Crooks, G., Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E* **2000**, *61*, 2361-2366.
30. Crooks, G. E., Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics* **1999**, *60*, 2721-2726.
31. Collin, D.; Ritort, F.; Jarzynski, C.; Smith, S. B.; Tinoco, I.; Bustamante, C., Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature* **2005**, *437*, 231-234.
32. Kokubo, H.; Okamoto, Y., Classification and Prediction of Low-Energy Membrane Protein Helix

Configurations by Replica-Exchange Monte Carlo Method. *J. Phys. Soc. Jpn.* **2004**, 73, 2571-2585.

33. 北尾彰朗, 主成分分析を使って眺めた蛋白質のエネルギー地形. 特集「地図を描く・風景を眺める-主成分分析・多次元尺度法とその周辺」. *統計数理* **2001**, 49, 43-56.

34. Smith, L. I., A tutorial on Principal Components Analysis Introduction. *Statistics* **2002**, 51.

35. Kitao, A.; Hirata, F.; Gō, N., The effects of solvent on the conformation and the collective motions of protein: Normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *Chem. Phys.* **1991**, 158, 447-472.

36. Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C., Essential dynamics of proteins. *Proteins: Struct., Funct., Bioinf.* **1993**, 17, 412-425.

37. 神嶋敏弘, データマイニング分野のクラスタリング手法(1). *人工知能学会誌* **2003**, 18, 59-65.

38. MacQueen, J. B., Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press* **1967**, 1, 281-297.

39. Petersen, J. P. PyPR v0.1rc3 documentation.

<http://pypr.sourceforge.net/kmeans.html> (accessed 4, Dec., 2014).

