

myPresto/cosgene 方式設計書
(構造探索エンジン機能追加編)

第 4 版

目次 -

| | |
|---------------------------------------|----|
| 1. はじめに | 2 |
| 1.1 構造探索エンジンの概要 | 2 |
| 1.2 構造探索エンジンの構成 | 3 |
| 2. 圧力制御 | 4 |
| 2.1 アンダーソン (Andersen) の方法 | 4 |
| 2.2 パリネロ・ラーマン (Parrinello-Rahman) の方法 | 5 |
| 3. 長距離相互作用の高速計算法 | 6 |
| 3.1 Tree algorithm | 6 |
| 3.2 Fast Multipole Method | 8 |
| 4. 剛体モデル | 13 |
| 5. Tsallis Dynamics | 17 |
| 6. 有効温度マルチカノニカル法 | 19 |
| 7. レプリカ交換法 (Replica-exchange method) | 22 |
| 8. ポテンシャル平滑化法 | 25 |
| ■ 参考文献 | 31 |

1. はじめに

本設計書では、分子動力学シミュレーションソフト(cosgene)の「構造探索エンジン機能追加」の方式について説明する。

1.1 構造探索エンジンの概要

Multicanonical method などの統計物理学的に厳密な高効率の構造探索を行うアルゴリズムに基づく構造探索エンジンを実現する。構造探索エンジンは主に、エネルギー最小化処理、MD 計算処理、および、トラジェクトリー解析処理から成る。

(1) エネルギー最小化処理

対象とする分子の初期構造(トポロジー情報と原子座標)からエネルギーが最小になるように構造を変化させていき、エネルギー最小点での分子構造を探索する。

エネルギー最小化処理は、結合距離、結合角など、初期構造が持っている不自然な構造の歪みを取り除き、より安定な構造へと緩和させる。求められた分子構造は、MD 計算の初期座標となる。これにより、MD 計算の初期における時間積分の発散を避けることができる。

(2) MD 計算処理

初期状態での分子の構造(トポロジー情報と原子座標:エネルギー最小化処理の結果を使う)と速度を入力として、分子の conformational エネルギーを計算する。

多数の原子から構成される仮想的な系(原子の集合体のバルクと同じ性質を持った系)を考え、この系で与えられた原子間の相互作用を用いて各原子に働く力を求め、運動方程式の時間に関する近似式により全ての原子の運動を追跡する。そして、時間に対する原子の位置と速度などの情報からトラジェクトリー(ポテンシャルエネルギー、座標、速度、自由エネルギーなどの物理量)を結果として得る。

(3) トラジェクトリー解析処理

MD 計算処理によって出力されるトラジェクトリーを解析する。

生体高分子の安定性に関連する原子の変動、自由エネルギーの変化、生体高分子の周りの水分子の動きなどを明らかにすることができる。

1.2 構造探索エンジンの構成

構造探索エンジンは主に、エネルギー最小化処理、MD 計算処理、および、トラジェクトリー解析処理を中心としたユーティリティ群で構成する。

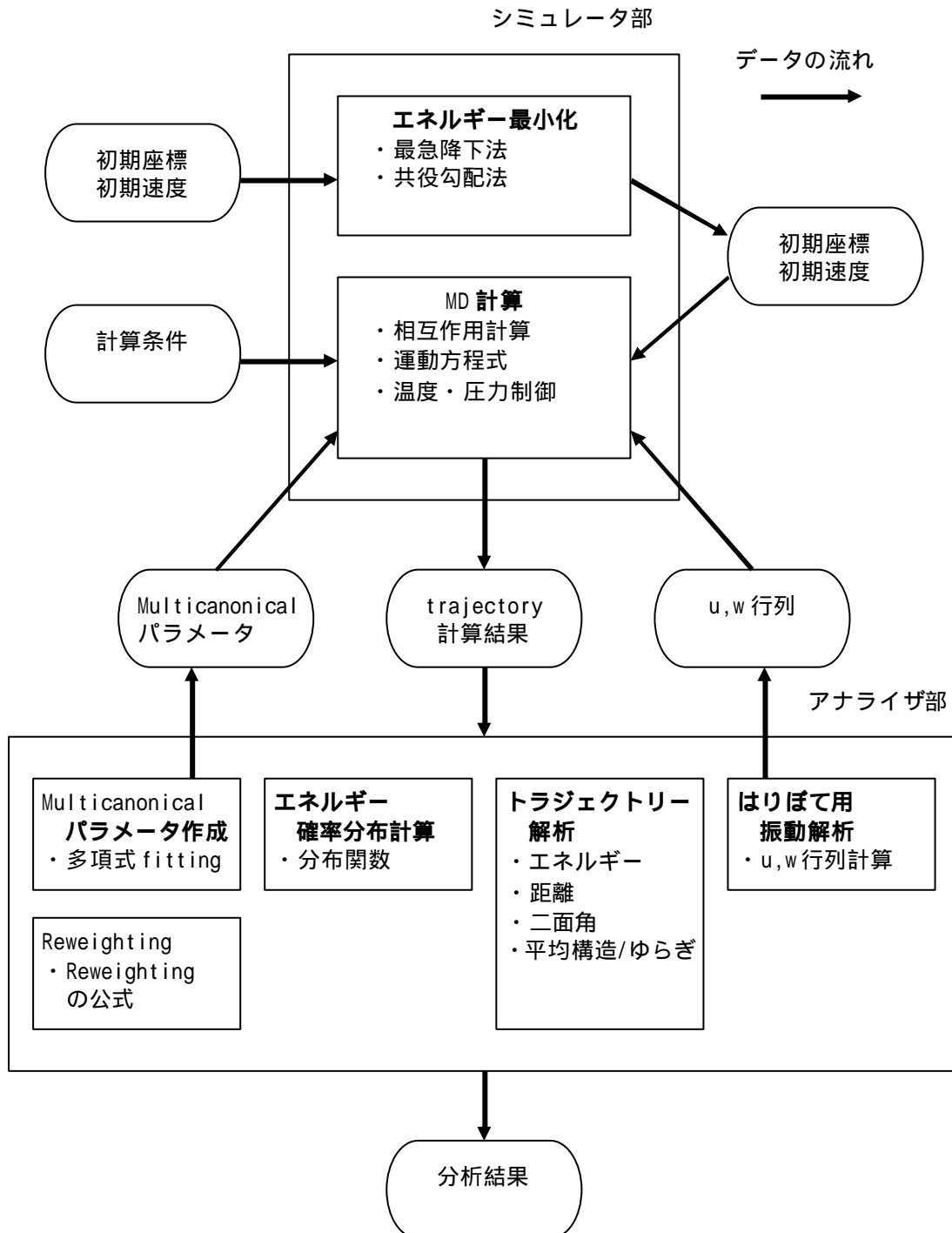


図1 構造探索エンジンの構成

2. 圧力制御

2.1 アンダーソン (Andersen) の方法

壁が動くことにより体積 V の変化する立方体セルに入った N 粒子系の運動方程式に拡張する。壁の位置を一般化座標として自由度を与え、分子の運動方程式と連立させて壁の運動そのものを解く。壁は一方向に移動するものではなく、立方体セルの三方向から等方的に圧縮、膨張させる。

内部圧力 P と外部圧力 P_{ext} が釣り合うように壁が動くこととし、壁の動きの速度 v_c とその壁の仮想的な質量 m_c を考える。この拡張系のハミルトニアン H は次式で表される。

$$H = \sum_{i=1}^N \frac{1}{2} m_i v_i^2 + \sum_{i=1}^N \sum_{j>i}^N V_{ij} + \frac{1}{2} m_c v_c^2 + P_{ext} V$$

一般化座標 q (r_i と V) に正準共役な運動量 p は、 $m_i v_i$ と $m_c v_c$ である。

ハミルトンの正準方程式 $\dot{q} = \frac{\partial H}{\partial p}$, $\dot{p} = -\frac{\partial H}{\partial q}$ より、粒子と壁の運動方程式として、

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = - \sum_{j=1, j \neq i}^N \frac{dV_{ij}}{d\mathbf{r}_{ij}} - \frac{1}{3} \left(m_i \frac{d\mathbf{r}_i}{dt} \right) \left(\frac{1}{V} \frac{dV}{dt} \right)$$

$$m_c \frac{d^2 v_c}{dt^2} = \frac{1}{3V} \left(\sum_{i=1}^N \frac{1}{2} m_i v_i^2 - \sum_{i=1}^N \sum_{j>i}^N \mathbf{r}_{ij} \cdot \frac{dV_{ij}}{d\mathbf{r}_{ij}} \right) - P_{ext}$$

が得られる。

NOTE

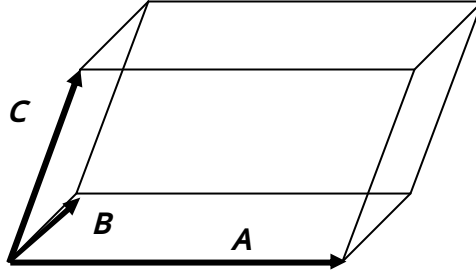
- ・ 仮想質量は経験的に決める (基本セルに含まれる粒子の全質量程度の値)。
- ・ 仮想質量が軽いと揺らぎが大きく、重いと揺らぎの周期が長くなるが、それには統計力学的意味は無い。

2.2 パリネロ・ラーマン (Parrinello-Rahman) の方法

アンダーソンの方法の一般化である。基本セルの各辺の長さが自由に変わると同時に、セルの傾きも変化し、圧力の平均値を P_0 とする。

シミュレーションセルの形状は、最も一般的な平行六面体にする。

$$\begin{aligned} \mathbf{H} &= (\mathbf{A} \ \mathbf{B} \ \mathbf{C}) \\ &= \begin{pmatrix} A_x & B_x & C_x \\ A_y & B_y & C_y \\ A_z & B_z & C_z \end{pmatrix} \end{aligned}$$



粒子座標 $\mathbf{r}_i (x_i, y_i, z_i)$ は、セル内部座標 $\mathbf{s}_i (s_{iA}, s_{iB}, s_{iC})$ ($0 < s_{iA}, s_{iB}, s_{iC} < 1$) を用いて

$$\mathbf{r}_i = \mathbf{H} \mathbf{s}_i$$

と表わせる。この拡張系のラグランジアンはセルの仮想質量 W を用いて、

$$L = \frac{1}{2} \sum_{i=1}^N m_i \dot{\mathbf{r}}_i^t \dot{\mathbf{r}}_i - \sum_{i=1}^N \sum_{j>i}^N V_{ij}(r_{ij}) + \frac{1}{2} W (\dot{\mathbf{H}}^t \dot{\mathbf{H}}) - P_{ext} \det \mathbf{H}$$

$\mathbf{r}_i = \dot{\mathbf{H}} \mathbf{s}_i + \mathbf{H} \dot{\mathbf{s}}_i$ であるが、 \mathbf{H} の時間変化が小さいため、 $\dot{\mathbf{r}}_i \cong \mathbf{H} \dot{\mathbf{s}}_i$ を用いて

$$L = \frac{1}{2} \sum_{i=1}^N m_i (\dot{\mathbf{s}}_i^t \mathbf{H}^t \mathbf{H} \dot{\mathbf{s}}_i) - \sum_{i=1}^N \sum_{j>i}^N V_{ij}(|\mathbf{H}(\mathbf{s}_i - \mathbf{s}_j)|) + \frac{1}{2} W (\dot{\mathbf{H}}^t \dot{\mathbf{H}}) - P_{ext} \det \mathbf{H}$$

粒子とセルの運動方程式は次のようになる。

$$\begin{aligned} \ddot{H}_{ab} &= \frac{1}{W} \left[\sum_i m_i (\mathbf{H} \dot{\mathbf{s}}_i)_b (\mathbf{H} \dot{\mathbf{s}}_i)_a - \sum_i \sum_{j>i} \frac{\partial V_{ij}}{\partial r_{ij}} \frac{1}{r_{ij}} [\mathbf{H}(\dot{\mathbf{s}}_i - \dot{\mathbf{s}}_j)]_a [\mathbf{H}(\dot{\mathbf{s}}_i - \dot{\mathbf{s}}_j)]_b \right] \\ &\quad - P_{ext} \frac{\partial}{\partial H_{ab}} \det \mathbf{H} \end{aligned}$$

$$\ddot{\mathbf{s}}_i = -\frac{1}{m_i} \sum_{i=1, j \neq i}^N \frac{\partial V_{ij}}{\partial r_{ij}} \frac{1}{r_{ij}} (\mathbf{s}_i - \mathbf{s}_j) - (\mathbf{H}^t \mathbf{H})^{-1} (\dot{\mathbf{H}}^t \mathbf{H} + \mathbf{H}^t \dot{\mathbf{H}}) \dot{\mathbf{s}}_i$$

3. 長距離相互作用の高速計算法

3.1 Tree algorithm

ツリー法や Fast Multipole Method (FMM) の基本的な考え方は、遠くの粒子を適当にまとめて扱うことにある。ツリー法の場合、階層的なツリー構造を用いて空間を分割し、粒子をまとめる (図 3-1)。

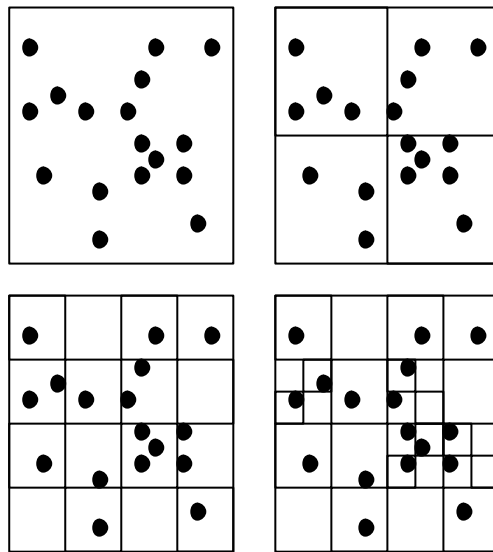
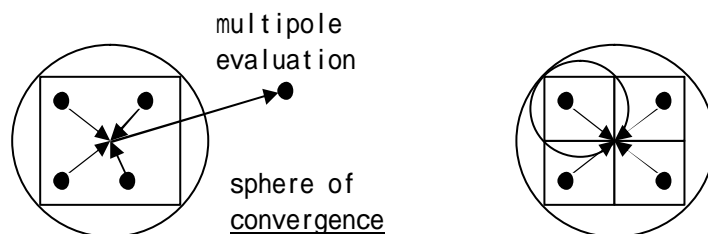


図 3-1 Tree structure

ツリーの構造を構築した後、各セルのなかの粒子がつくるポテンシャルの多重極展開 (Multipole Expansion) を以下のように計算する (図 3-2)。

- ・最下層セル：その中にある粒子が作るポテンシャルを多重極展開として求める (P2M)
- ・それ以外のセル：子セルの多重展開の展開中心をシフトして加算する (M2M)



(a) Multipole expansion

(b) Multipole translation

図 3-2 Multipole expansion and multipole translation

すべてのセルに対して多重極展開が得られた段階で、ツリー法では力の計算に入る。その計算方法は

$$\text{ある粒子への力} = \begin{cases} \text{セル多重極展開からの力 (十分離れている場合)} \\ \text{子セルからの力の合計 (それ以外)} \end{cases}$$

と再帰的に定義する (図 3-3)。ここで「十分離れている」とは、粒子とセル中心との距離を d 、セルの長さを l としたとき、あるパラメータ q に対し $\frac{l}{d} < q$ を満たすセルをさす。

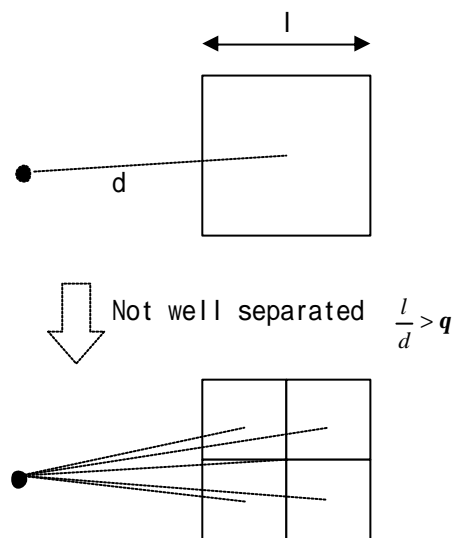


図 3-3 Well-separatedness criteria for tree algorithm

3.2 Fast Multipole Method

ツリー法では、ある粒子への力を計算するのに、遠くの粒子からの力をまとめて計算することにした。Fast Multipole Method (FMM) ではこの考え方をさらに発展させ、力を受ける側もまとめてやることで原理的に $O(N \log N)$ のツリー法に対して $O(N)$ へ計算量のオーダーを下げるができる (図 3-4)。以下にその原理を説明する。

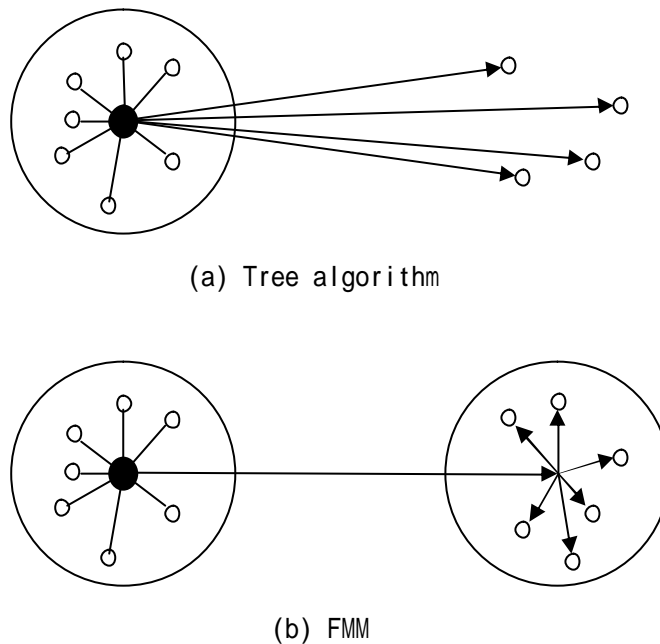


図 3-4 Approximations in tree and FMM

FMM においてもツリー法と同様の方法で階層的なツリー構造を用いて空間を分割し、それぞれのセルにおける多重極展開を求めておく (upward pass)。ただし一般には、空間全体を一様に分割する。

FMM では、1つの粒子への力を以下のように分解する。

$$\mathbf{F} = \mathbf{F}_{pp} + \mathbf{F}_{pc}$$

ここで、 \mathbf{F}_{pp} は自分のいるセルとその第 1, 2 隣接セル内にある近傍粒子からの力であり、通常の直接法によって計算する。一方 \mathbf{F}_{pc} は多重極展開から計算する。

このとき FMM では、ツリー法と異なり、各セルが (十分離れている) 他のセルから受けるポテンシャル (局所展開: Local Expansion) をあらかじめ計算しておき、それを粒子の位置で評価することで求める (downward pass)。

具体的にはまず以下の2つの操作を行う。

- ・ Interaction cell の多重極展開からセル中心での局所展開を計算する (M2L)
- ・ 自分の局所展開を子セルの中心にシフトする (L2L)

ここで、interaction cell とは、あるセルに対して、

- () そのセルと同じレベルにあって、
- () その親のセルの第1, 2隣接セルの子であって、
- () 問題のセルの第1, 2隣接セルでない、

セルである。

この操作を一番上のセルから順に、最下層のセルに達するまで繰り返すと、各セルの中心での、自分と第1, 2隣接セルに入っている粒子を除いた他のすべての粒子による局所展開が求まる (図 3-5)。したがってある粒子への力は、この最下層での局所展開を粒子の位置で評価したものと、第1, 2隣接セル内の粒子からの合計となる。

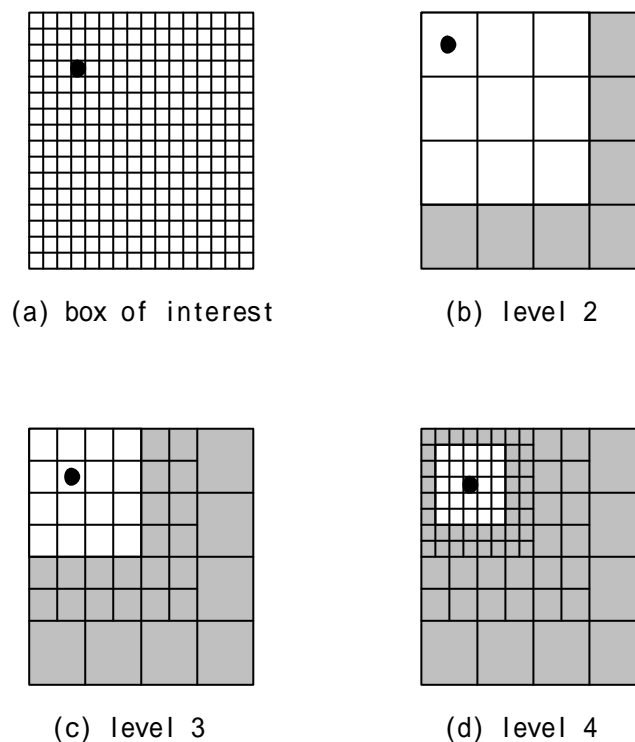


図 3-5 Creation of the local expansion from interaction cells

各セルでの多重極展開を Φ_i 、局所展開を Ψ_i 、セルに含まれる粒子に働く全ポテンシャルを f_i 、また、直接計算から得られるポテンシャルを Γ_i とすると、FMM の計算ステップは以下のようなになる (図 3-6)。

(1) Construct tree and load particles into cells at level L

(2) Upward pass

(a) DO i level L

compute Φ_i (最下層セルで多重極展開を計算する (P2M))

(b) DO $j=L-1, 2, -1$

DO i level j

$$\Phi_i = \sum_{k \in \text{children}(i)} \Phi_k \quad (\text{子セルの多重極展開をシフトして加算する (M2M)})$$

(3) Downward pass

(a) DO i level 1

$$\Psi_i = 0$$

(b) DO $j=2, L$

DO i level j

$$\Psi_i = \Psi_{\text{paren}(i)} + \sum_{k \in \text{int.cell}(i)} \Phi_k \quad (\text{多重極展開から局所展開を計算する (M2L)})$$

if ($j > 1$) $\Psi_i \rightarrow \Psi_{\text{children}(i)}$ (局所展開を子セルの中心にシフトする (L2L))

(4) DO i level L

$$f_i = \Psi_i + \sum_{k \in \text{neigh}(i) \text{ and } i} \Gamma_k \quad (\text{直接計算によるポテンシャルを加算する})$$

FMM では、空間を分割するツリー構造のレベル L と多重極展開の次数 p が計算効率と精度を決定する。また interaction cell の定義方法によってもこれらは変化する。

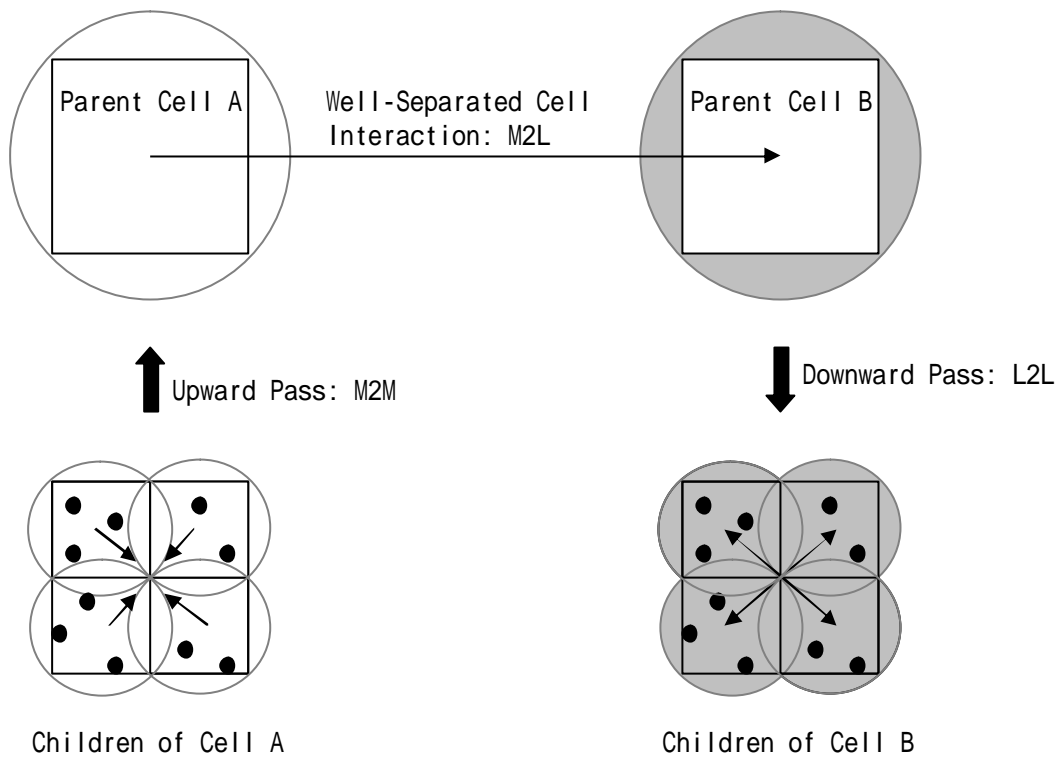


図 3-6 FMM Multipole Expansion Translations

【多重極展開とその変換式】

• Multipole Expansion (P2M)

$$\Phi(P) = \sum_{n=0}^p \sum_{m=-n}^n \frac{M_n^m}{r^{n+1}} Y_n^m(\mathbf{q}, \mathbf{f})$$

$$M_n^m = \sum_{i=1}^k q_i \mathbf{r}_i^n Y_n^{-m}(\mathbf{a}_i, \mathbf{b}_i)$$

• Multipole Translation (M2M)

$$\Phi(P) = \sum_{j=0}^p \sum_{k=-j}^j \frac{M_j^k}{r^{j+1}} Y_j^k(\mathbf{q}, \mathbf{f})$$

$$M_j^k = \sum_{n=0}^j \sum_{m=-n}^n \frac{O_{j-n}^{k-m} J_m^{k-m} A_n^m A_{j-n}^{k-m} Y_n^{-m}(\mathbf{a}, \mathbf{b}) \mathbf{r}^n}{A_j^k}$$

$$J_k^m = \begin{cases} (-1)^{\min(|m|, |k|)}, & \text{if } mk < 0; \\ 1, & \text{otherwise;} \end{cases}$$

$$A_n^m = \frac{(-1)^n}{\sqrt{(n-m)!(n+m)!}}$$

• Local Expansion (M2L)

$$\Phi(P) = \sum_{j=0}^p \sum_{k=-j}^j L_j^k Y_j^k(\mathbf{q}, \mathbf{f}) r^j$$

$$L_j^k = \sum_{n=0}^p \sum_{m=-n}^n \frac{O_n^m J_m^k A_n^m A_j^k Y_{n+j}^{m-k}(\mathbf{a}, \mathbf{b})}{A_{n+j}^{m-k} \mathbf{r}^{n+j+1}}$$

$$J_k^m = \begin{cases} (-1)^m (-1)^{\min(|m|, |k|)}, & \text{if } mk < 0; \\ (-1)^m, & \text{otherwise;} \end{cases}$$

$$A_n^m = \frac{(-1)^n}{\sqrt{(n-m)!(n+m)!}}$$

• Local Translation (L2L)

$$\Phi(P) = \sum_{j=0}^p \sum_{k=-j}^j L_j^k Y_j^k(\mathbf{q}, \mathbf{f}) r^j$$

$$L_j^k = \sum_{n=j}^p \sum_{m=-n}^n \frac{O_n^m J_{n-j}^{m,k} A_{n-j}^{m-k} A_j^k Y_{n-j}^{m-k}(\mathbf{a}, \mathbf{b}) \mathbf{r}^{n-j}}{A_n^m}$$

$$J_k^{m,k} = \begin{cases} (-1)^n (-1)^{m-k}, & \text{if } m(m-k) < 0; \\ (-1)^n (-1)^k, & \text{if } m = 0 \text{ or } mk < 0; \\ (-1)^n, & \text{otherwise;} \end{cases}$$

$$A_n^m = \frac{(-1)^n}{\sqrt{(n-m)!(n+m)!}}$$

4. 剛体モデル

ねじれ運動をもたない分子（水、二酸化炭素など）は剛体近似で取り扱うことができる。これには、剛体系の運動方程式（ニュートン・オイラーの運動方程式）を用いる。この場合、分子の自由度は6（並進3、回転3）に減らすことができ、運動方程式は複雑になるが時間積分のタイムステップを長くとれるため、計算時間を減らすことができる。

剛体*i* と *j* の間の分子間ポテンシャル U_{ij} は、分子内の原子間ポテンシャルの和として次のように表せる。

$$U_{ij} = \sum_{a=1}^m \sum_{b=1}^n V_{a(i)b(j)}$$

ここで、 $V_{a(i)b(j)}$ は分子*i* の原子*a* と分子*j* の原子*b* との間の原子間ポテンシャルであり、*m* は分子*i* の原子数、*n* は分子*j* の原子数を表す。

(a) 並進運動部分 : ニュートンの運動方程式を用いる。

分子*i* に働く力は

$$F_i = \sum_{j=1}^N \frac{\partial U_{ij}}{\partial \mathbf{r}_{ij}}$$

であるが、その成分は

$$F_{ix} = - \sum_{j=1}^N \sum_{a(i)=1}^m \sum_{b(j)=1}^n \frac{\partial V_{a(i)b(j)}}{\partial r_{a(i)b(j)}} \frac{x_{a(i)b(j)}}{r_{a(i)b(j)}}$$

$$F_{iy} = - \sum_{j=1}^N \sum_{a(i)=1}^m \sum_{b(j)=1}^n \frac{\partial V_{a(i)b(j)}}{\partial r_{a(i)b(j)}} \frac{y_{a(i)b(j)}}{r_{a(i)b(j)}}$$

$$F_{iz} = - \sum_{j=1}^N \sum_{a(i)=1}^m \sum_{b(j)=1}^n \frac{\partial V_{a(i)b(j)}}{\partial r_{a(i)b(j)}} \frac{z_{a(i)b(j)}}{r_{a(i)b(j)}}$$

となる。また、分子質量を

$$M_i = \sum_{a=1}^m m_a, \quad M_j = \sum_{b=1}^n m_b$$

としてニュートンの運動方程式を解く。

(b) 回転運動部分 : オイラーの運動方程式を用いる。

回転運動を求める手順を以下に示す。

- (1) 空間座標系でのトルクを求める。
- (2) 変換行列 R を用いて、トルクを分子座標系に変換する。
- (3) 分子座標系でのトルクを用いて、分子主軸回りの角加速度を求める。
- (4) 角加速度を用いて、角速度を更新する(数値積分)。
- (5) 角速度を用いて、分子配向(四元数)の変化速度を求める。
- (6) 分子配向(四元数)の変化速度を用いて、分子配向を更新する(数値積分)。

剛体分子 i と j との間の力により分子 i に働くトルク T_{ij} は

$$T_{ij} = \sum_{a=1}^m \sum_{b=1}^n s_{a(i)} \times F_{a(i)b(j)}$$

ここで、 $s_{a(i)}$ は分子 i の重心から分子 i の原子 a までの距離、 $F_{a(i)b(j)}$ は分子 i の原子 a が分子 j の原子 b がから受ける力を示す。その各軸方向の成分は

$$T_{ijx} = \sum_{a=1}^m \sum_{b=1}^n (s_{a(i)y} F_{a(i)b(j)z} - s_{a(i)z} F_{a(i)b(j)y})$$

$$T_{ijy} = \sum_{a=1}^m \sum_{b=1}^n (s_{a(i)z} F_{a(i)b(j)x} - s_{a(i)x} F_{a(i)b(j)z})$$

$$T_{ijz} = \sum_{a=1}^m \sum_{b=1}^n (s_{a(i)x} F_{a(i)b(j)y} - s_{a(i)y} F_{a(i)b(j)x})$$

となる。なお、 $F_{a(i)b(j)} = -F_{b(j)a(i)}$ 、 $T_{ij} = -T_{ji}$ である。

他の全ての分子 j から受ける力による分子 i のトルクは $T_i = \sum_{j=1}^N T_{ij}$ である。

トルクによる分子 i の回転運動は、 L を重心回りの角運動量とすると、

$$\frac{dL_i}{dt} = T_i$$

となる。慣性モーメント I と角速度 w を用いると、 $L_i = I_i w_i$ であり、行列表現は、

$$\frac{d}{dt} \begin{bmatrix} I_{ixx} & I_{ixy} & I_{ixz} \\ I_{iyx} & I_{iyy} & I_{iyz} \\ I_{izx} & I_{izy} & I_{izz} \end{bmatrix} \begin{bmatrix} w_{ix} \\ w_{iy} \\ w_{iz} \end{bmatrix} = \begin{bmatrix} T_{ix} \\ T_{iy} \\ T_{iz} \end{bmatrix}$$

となるが、分子は空間座標系に固定されていないので、慣性テンソルは時間とともに変化

する。座標軸を分子の慣性主軸に固定した分子座標系（主軸座標系）を用いると慣性テンソルは対角化され固定される。そこで、オイラーの運動方程式をたてるためにトルクを空間座標系から分子座標系へ変換を行う。つまり、剛体分子の主軸周りのトルクに変換するのであるが、ここで分子配向を表現するためのオイラー角（ α 、 β 、 γ ）を定義する。分子座標系への変換行列 R は次式となる。

$$R = \begin{pmatrix} \cos \alpha \cos \beta - \sin \alpha \sin \beta \sin \gamma & \sin \alpha \cos \beta + \cos \alpha \sin \beta \sin \gamma & \sin \beta \sin \gamma \\ -\cos \alpha \sin \beta - \sin \alpha \cos \beta \cos \gamma & -\sin \alpha \sin \beta + \cos \alpha \cos \beta \cos \gamma & \sin \beta \cos \gamma \\ \sin \alpha \sin \beta & -\cos \alpha \sin \beta & \cos \beta \end{pmatrix}$$

この変換は座標だけでなく、力、速度、トルク、角速度などにも共通である。

なお、分子座標系から空間座標系への逆変換を行う行列は $R^{-1} (= R^t)$ である。

この変換行列を用いて $T_{LMN} = R T_{xyz}$ で求めた分子の慣性主軸(L,M,N)周りのトルク成分を T_L 、 T_M 、 T_N とするとオイラーの運動方程式は次式となる。

$$\begin{aligned} I_L \dot{w}_L - w_M w_N (I_M - I_N) &= T_L \\ I_M \dot{w}_M - w_N w_L (I_N - I_L) &= T_M \\ I_N \dot{w}_N - w_L w_M (I_L - I_M) &= T_N \end{aligned}$$

ここで、 I_L 、 I_M 、 I_N は主軸回りの慣性モーメント、 w_L 、 w_M 、 w_N は主軸回りの角速度である。この式を \dot{w}_L 、 \dot{w}_M 、 \dot{w}_N について解くと

$$\begin{aligned} \dot{w}_L &= \frac{T_L + w_M w_N (I_M - I_N)}{I_L} \\ \dot{w}_M &= \frac{T_M + w_N w_L (I_N - I_L)}{I_M} \\ \dot{w}_N &= \frac{T_N + w_L w_M (I_L - I_M)}{I_N} \end{aligned}$$

となり、これを数値積分して主軸回りの角速度が求まる。

角速度から分子配向を求めるには、次のオイラー角との関係を用いる。

$$\begin{aligned} w_L &= \dot{\beta} \cos \gamma + \dot{\alpha} \sin \beta \sin \gamma \\ w_M &= -\dot{\beta} \sin \gamma + \dot{\alpha} \sin \beta \cos \gamma \\ w_N &= \dot{\alpha} \cos \beta + \dot{\gamma} \end{aligned}$$

これを $\dot{\alpha}$ 、 $\dot{\beta}$ 、 $\dot{\gamma}$ について解くと次のようになる。

$$\begin{aligned}\dot{\mathbf{a}} &= (\mathbf{w}_L \sin \mathbf{g} + \mathbf{w}_M \cos \mathbf{g}) \frac{1}{\sin \mathbf{b}} \\ \dot{\mathbf{b}} &= \mathbf{w}_L \cos \mathbf{g} - \mathbf{w}_M \sin \mathbf{g} \\ \dot{\mathbf{c}} &= \mathbf{w}_N - (\mathbf{w}_L \sin \mathbf{g} + \mathbf{w}_M \cos \mathbf{g}) \frac{\cos \mathbf{b}}{\sin \mathbf{b}}\end{aligned}$$

しかし、この式には分母に \sin があり、特異点 $= 0$ 、 の値をとらないとしても、その付近で大きな数値的誤差を生じるため、四元数 () 表示を用いてそれを回避する。(変数が1つ増える)

オイラー角から四元数への変換は

$$\begin{aligned}\mathbf{x} &= \sin \frac{\mathbf{b}}{2} \sin \frac{(\mathbf{g} - \mathbf{a})}{2} \\ \mathbf{h} &= \sin \frac{\mathbf{b}}{2} \cos \frac{(\mathbf{g} - \mathbf{a})}{2} \\ \mathbf{V} &= \cos \frac{\mathbf{b}}{2} \sin \frac{(\mathbf{g} + \mathbf{a})}{2} \\ \mathbf{c} &= \cos \frac{\mathbf{b}}{2} \cos \frac{(\mathbf{g} + \mathbf{a})}{2}\end{aligned}$$

であり、 $\dot{\mathbf{a}}, \dot{\mathbf{b}}, \dot{\mathbf{c}}$ の代わりに $\dot{\mathbf{x}}, \dot{\mathbf{h}}, \dot{\mathbf{V}}, \dot{\mathbf{c}}$ を用いて表すと

$$\begin{pmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{h}} \\ \dot{\mathbf{V}} \\ \dot{\mathbf{c}} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -\mathbf{V} & -\mathbf{c} & \mathbf{h} & \mathbf{x} \\ \mathbf{c} & -\mathbf{V} & -\mathbf{x} & \mathbf{h} \\ \mathbf{x} & \mathbf{h} & \mathbf{c} & \mathbf{V} \\ -\mathbf{h} & \mathbf{x} & -\mathbf{V} & \mathbf{c} \end{pmatrix} \begin{pmatrix} \mathbf{w}_L \\ \mathbf{w}_M \\ \mathbf{w}_N \\ 0 \end{pmatrix}$$

これを数値積分して分子配向 () を求める。

なお、空間座標系から分子座標系へ変換行列 \mathbf{R} を四元数で表わすと

$$\mathbf{R} = \begin{pmatrix} -\mathbf{x}^2 + \mathbf{h}^2 - \mathbf{V}^2 + \mathbf{c}^2 & 2(\mathbf{Vc} - \mathbf{xh}) & 2(\mathbf{hV} + \mathbf{xc}) \\ -2(\mathbf{Vc} + \mathbf{xh}) & \mathbf{x}^2 - \mathbf{h}^2 - \mathbf{V}^2 + \mathbf{c}^2 & 2(\mathbf{hc} - \mathbf{xV}) \\ 2(\mathbf{hV} - \mathbf{xc}) & -2(\mathbf{hc} + \mathbf{xV}) & -\mathbf{x}^2 - \mathbf{h}^2 + \mathbf{V}^2 + \mathbf{c}^2 \end{pmatrix}$$

、 、 、 の間には $^2 + ^2 + ^2 + ^2 = 1$ の関係がある。

シミュレーション中に発生する計算誤差 (その値が1からずれる) は定期的に補正する。

5. Tsallis Dynamics

Tsallis 統計は、Boltzmann-Shannon エントロピーの加法性を非加法的なものに置き換えた Tsallis entropy により基礎付けられる。Boltzmann-Gibbs (BG) 分布が Boltzmann-Shannon エントロピーを最大にする、ということで特徴づけられることに対し、Tsallis 分布は、Tsallis entropy を適当な拘束条件下で極値化する分布である。即ち、Tsallis 分布は非加法的に拡張した一般化されたカノニカル分布である。本手法で対象にする連続型の Tsallis 分布は、次の確率密度により定義される：

$$\mathbf{r}_{\text{Tsallis}}(x, p) = [1 - (1 - q)\beta E(x, p)]^{\frac{q}{1-q}} \times c. \quad (1)$$

ここで、 $E(x, p)$ は系の全エネルギーであり、Tsallis 指数 $q \rightarrow 1$ の極限が、BG 分布の密度、 $r_{\text{BG}}(x, p) = \exp[-\beta E(x, p)] \times c$ を与える。

Nose と Hoover は canonical 分布を実現する常微分方程式(ODE)を提唱し、単に一自由度の追加により canonical 分布が実現可能であることを示し、温度一定の実際の物理系の simulation を現実的なものにした。本方式では、熱浴内の物理系の候補と考えられる Tsallis 分布を実現する決定論的方程式を、Nose-Hoover の手法を基本にして導き、それを用いて物理系の simulation をできるようにする。 (x, p) 両変数を用いた Tsallis 指数 $q \geq 1$ の Tsallis 分布を対象とする(即ち、energy として、potential energy だけでなく、Kinetic energy も含める)。

本方式では、適当な数学的条件の下、ODE の解の、任意の微小領域への滞在時間率の長時間極限が、上記(1)式の $r_{\text{Tsallis}}(x, p)$ に比例するように、ODE を定義する。

次のような ODE を構築した：

$$\left. \begin{aligned} \bullet dx_i / dt &= g(x, p) p_i, \quad i=1, \dots, n, \\ \bullet dp_i / dt &= g(x, p) D_i U(x) - (?) p_i, \quad i=1, \dots, n, \\ \bullet d? / dt &= g(x, p) \|p\|^2 - n \end{aligned} \right\} \quad (2)$$

$$\text{where } g(x, p) \equiv \frac{q\beta}{[1 - (1 - q)\beta E(x, p)]},$$

$$(?) \equiv -D \log ?_c(?)$$

ここで、 x_i, p_i は各々、座標、運動量 ($i=1, \dots, n$)、 $U(x)$ は potential energy、 $\beta = 1/kT$ である。但し、 T は renormalized temperature と呼ばれる温度である。簡単のため質量は全て 1 としている。数学的付加的条件の下、この系の ergodic 性を仮定すると、この ODE を解くことにより、解は密度(1)式に比例する出現頻度にて時間変化することになり、Tsallis

分布が実現されることが示される。

即ち、任意の微小領域 $\mathcal{V}(x, p)$ に対し、次が成立する：

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \mathcal{V}_{(x,p)}(x(t), p(t)) dt &= \int_{\mathcal{V}(x,p)} \rho_{\text{Tsallis}}(x, p) dx dp / \int_{\mathbf{R}^{2n}} \rho_{\text{Tsallis}}(x, p) dx dp \quad (a.e.) \\ &\approx \rho_{\text{Tsallis}}(x, p) \mathcal{V}(x,p) / \int_{\mathbf{R}^{2n}} \rho_{\text{Tsallis}}(x, p) dx dp \\ &= \rho_{\text{Tsallis}}(x, p) \times \text{const} \end{aligned} \quad (3)$$

$$\text{where } \mathcal{V}_{(x,p)} : \mathbf{R}^{2n} \rightarrow \mathbf{R}, (x, p) \mapsto \begin{cases} 1 \cdots (x, p) \in \mathcal{V}(x, p) \\ 0 \cdots \text{otherwise} \end{cases}$$

左辺は $\mathcal{V}(x, p)$ への軌道訪問頻度の長時間極限である。従って、点 (x, p) の実現確率密度は ρ_{Tsallis} に比例する。

Tsallis 分布における物理量の期待値はサンプリングによる長時間平均にて計算できる。 x, p の関数としての物理量 A の長時間平均値は、重み ρ_{Tsallis} の空間平均に等しくなる：

$$\begin{aligned} \bar{A} &\equiv \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t A(x(t), p(t)) dt \\ &= \int_0 A(x, p) \rho_{\text{Tsallis}}(x, p) dx dp / \int_0 \rho_{\text{Tsallis}}(x, p) dx dp \\ &= \int_{\mathbf{R}^{2n}} A(x, p) \rho_{\text{Tsallis}}(x, p) dx dp / \int_{\mathbf{R}^{2n}} \rho_{\text{Tsallis}}(x, p) dx dp \quad (a.e.). \end{aligned} \quad (4)$$

即ち、式(2)によって A の長時間平均を計算することにより、Tsallis 分布における期待値が得られる。なお、これらは、通常初期値に依存しないことが示される。

ところで、Tsallis 分布を定義するものとして(1)式以外の密度が用いられることがあるが、本手法でも少しの変形で対応可能であり、(1)式の形に限定されない。なお、通常は(1)式でなく、

$$\rho_{\text{Tsallis}}(x, p) = [1 - (1 - q)\beta E(x, p)]^{\frac{1}{1-q}} \times c,$$

を用いて定義されることが多い。しかし、物理量の期待値を求める際には、上記を q 乗した(1)式の形で用いることも多い。ここでは物理量の期待値を求めることを重点としているので、便宜的に(1)式をその定義として採用したものである。また $E(x, p)$ も、さらに一般的な形の関数として取り扱うことが可能である。

6. 有効温度マルチカノニカル法

マルチカノニカル MD 法では最終的に、フラットなエネルギーの分布 ($P(E) = \text{一定}$) を実現したいのだが、 $P(E)$ はなかなかフラットにならない。そこでカノニカルアンサンブルの系のエネルギー分布 $P(E)$ では系の温度とエネルギーの関係が 1 : 1 であることを仮定することで、マルチカノニカルを連続的に分布する様々な温度でのカノニカル分布の重ね合わせで表現した。次に、ある時刻 t でエネルギー E であった場合、次の時刻 $t + \Delta t$ でエネルギー E' に遷移する確率を求めた。これによりマルチカノニカル MD での重み関数と $P(E)$ の振る舞いを解析した。またエネルギーの分布がフラットになるということは、エネルギー E から出発して高エネルギー側にも低エネルギー側にも偏った遷移をしないということであることから、マルチカノニカルを簡便に実現することができる。

ポテンシャルスケール法でのマルチカノニカル MD 法は、温度 T_0 での温度一定の条件下で以下の運動方程式を解くことで、フラットなエネルギーの分布 ($P(E) = \text{一定}$) を得る方法である。

$$\dot{p}_i = \frac{\partial E_{mc}(E)}{\partial E} f_i = g(E) f_i \quad (1)$$

ここで $g(E)$ の意味は、例えば $g(E)$ が定数 g_0 の場合、 $\dot{p}_i = g_0 f_i = \frac{\partial}{\partial q_i} g_0 E(q_i)$ からボルツマン因子を考えると、 $-g(E)E(q_i)/kT = -E(q_i)/k(T_0/g_0)$ なので、 $T = T_0/g_0$ ないし $T(E) = T_0/g(E)$ という関係が得られる。つまり、温度 T_0 でのポテンシャルスケールされた ($E_{mc}(E)$) エネルギーでのマルチカノニカル MD は、エネルギー E に依存する有効温度 $T(E) = T_0/g(E)$ での通常のポテンシャルエネルギー面 (E) での MD と同じであることがわかり、 $g(E)$ は有効温度を与える重みである。この式によりポテンシャルスケールは温度スケールに置き換えられた。

ヒストグラム法によって異なる温度での別々の計算結果を重ね合わせ、任意の温度でのカノニカルアンサンブルを構成することができる。したがって重み $g(E)$ によって定義された有効温度 $T(E)$ でのカノニカルアンサンブルを重ね合わせ、温度 T_0 でのマルチカノニカルの結果を式(2), (3)のように表現することができる。

$$p(E) = r(E) \int dT \Gamma(T) e^{-b_0(T_0/T)E} / Z(T) \quad (2)$$

$$Z(T) = \int dE \frac{p(E)}{\int dT' \Gamma(T') e^{b_0(T_0/T - T_0/T')E} / Z(T')} \quad (3)$$

ここで $r(E)$ は、温度の分布関数である。系の温度と系のエネルギーとに 1 : 1 の関係があれば

ば、式(2), (3)は(4), (5)のようにエネルギーの関数に書き直すことができる。

$$p(E) = r(E) \int dE' p(E') e^{-E/kT(E')} / Z(T(E')) \quad (4)$$

$$Z(T(E)) = \int dE' \frac{p(E')}{\int dE'' p(E'') e^{(1/kT(E)-1/kT(E''))E'} / Z(T(E''))} \quad (5)$$

式(4)の確率分布を満たすマルコフ連鎖を考えると、以下の式であって、

$$p(E) = \sum_{E'} P(E' \rightarrow E) p(E') \quad (6)$$

$$P(E' \rightarrow E) = r(E) e^{-E/kT(E')} / Z(T(E')) \quad (7)$$

理想的な場合を考えると、求めたかったエネルギーの遷移確率は、解析的に

$$P(E' \rightarrow E) = e^{-(E-U(T(E')))^2/2A} / Z(T(E')) \quad (8)$$

ここで A は、比熱 c_v により $A(T(E')) = kT(E')^2 c_v(T(E'))$ と書かれる。 $U(T(E'))$ は、温度 $T(E')$ でのカノニカルアンサンブルでの分布のピークの温度または平均の温度である。

式(8)により、重み関数に依存するマルチカノニカル MD の振る舞いを解析することができる。図 6-1 は横軸が系のエネルギー E 、縦軸が系の温度 T をとり、エネルギー E に対応する有効温度 $T = T(E)$ と温度 T でのカノニカル分布のピークのエネルギー(ないし平均エネルギー) $U = U(T)$ をプロットしたものである。

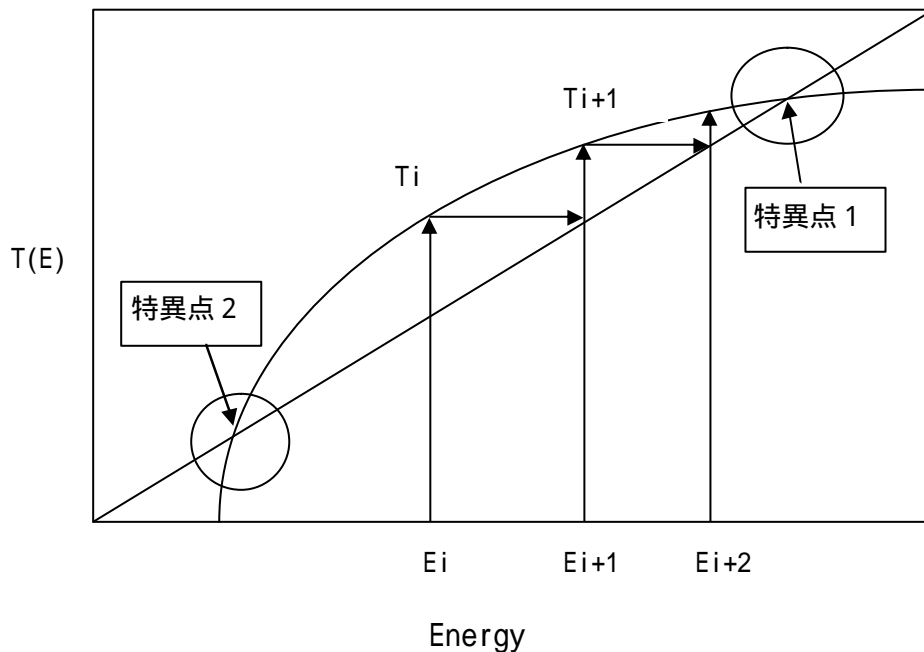


図 6-1 有効温度とエネルギー

時刻 t に系がエネルギー E_i をとっていたとき、対応する有効温度は $T_i = T(E_i)$ である。式 (8) に従えば、時刻 $t + \Delta t$ にエネルギー E_{i+1} へ遷移する確率 $P(E_i \rightarrow E_{i+1})$ は、 $E_{i+1} - U(T(E_i)) = 0$ となるとき最大となる。したがって、時刻 $t + \Delta t$ では $E_{i+1} = U(T(E_i))$ へ遷移しているのがもっとも確からしい。同様に時刻 $t + 2\Delta t$ では、 $E_{i+2} = U(T(E_{i+1}))$ へ遷移する確率が高い。

もし、有効温度 $T = T(E)$ と、系のピーク（ないし平均）エネルギーの曲線が交差しているなら、時間が進むにしたがい系の存在確率が集中する点（特異点 1）と、系がほとんど存在しなくなる点（特異点 2）とが生じる。すなわちマルチカノニカル MD においてエネルギー分布関数 $P(E)$ が平らにならず特に低温部分にピークが現れる現象は、有効温度と系のエネルギーとの関係が特異点 1 の状態になることから生じる。

上記の議論よりマルチカノニカル MD が実現するとは、全てのエネルギー領域でエネルギーが高エネルギーないし低エネルギーに遷移しなくなる、つまり全領域が特異点となり、 $E = U(T(E))$ が満たされることと同義である。逆に、マルチカノニカル MD の定義を $P(E) = \text{一定}$ から $E = U(T(E))$ に読み替えることによりマルチカノニカル MD の重み関数 $g(E)$ を決定することができる。

異なる温度 T_i での通常のカノニカル MD を行い、各シミュレーションでのエネルギー分布関数 $P(E)$ から系のピークのエネルギー（ないし平均エネルギー） $U(T_i)$ を計算することで、 $T_i = T(E_i)$ 、 $E_i = U(T_i)$ をプロットする。これらの点をエネルギーの多項式でフィットすることにより $T = T(E)$ を得て、 $T(E) = T_0/g(E)$ の関係式から重み関数 $g(E)$ を得ることができる。

7. レプリカ交換法 (Replica-exchange method)

レプリカ交換法は焼き戻し法の改良版として開発された。焼き戻し法では、温度自体が動的変数となり、シミュレーションの間に原子の座標（と運動量）及び温度が変更されていく。各状態は次の重み因子 $W_{ST}(E; T)$ に比例して発生し、温度の一様分布が得られる。

$$\begin{aligned} P_{ST}(T) &= \int dE n(E) W_{ST}(E; T) \\ &\equiv \int dE n(E) e^{-bE+a(T)} = \text{constant}. \end{aligned}$$

すなわち、この方法では「温度空間」上の自由酔歩が実現される。そして、それがポテンシャルエネルギー空間上の酔歩を誘起する。

実際の数値計算においては、温度を M 個の値 $T_m (m=1, \dots, M)$ に離散化する ($T_1 < T_2 < \dots < T_M$ とする)。最低温度 T_1 は最低エネルギー領域を探索できるほど十分低くとる。また、最高温度 T_M はエネルギー極小状態に留まらないよう、十分高くとる必要がある。すると、 $a_m = a(T_m)$ として、上式から次のことが言える。

$$e^{-a_m} \propto \int dE n(E) e^{-b_m E}$$

すなわち、焼き戻し法のパラメータ a_m は温度 T_m における「無次元化されたヘルムホルツ自由エネルギー」(a_m とヘルムホルツ自由エネルギー積) である。

上のパラメータ $a_m (m=1, \dots, M)$ は、普通短いシミュレーションによって決められる。重み因子が求めれば、焼き戻しシミュレーションは、次の2ステップを交互に繰り返すことで実行される。

- (1) 一定温度 T_m のカノニカルアンサンブル上の従来のMCまたはMDシミュレーションを、あるステップ数実行する。
- (2) 状態(座標と運動量)を固定したまま、温度 T_m を隣接温度 $T_{m \pm 1}$ に更新する。この更新の遷移確率は、次のようにメトロポリス判定で与えられる。

$$w(T_m \rightarrow T_{m \pm 1}) = \begin{cases} 1, & \text{for } \Delta \leq 0, \\ \exp(-\Delta), & \text{for } \Delta > 0. \end{cases}$$

ここで、 Δ は次のように与えられる。

$$\Delta \equiv (b_{m \pm 1} - b_m) E - (a_{m \pm 1} - a_m).$$

長い焼き戻しシミュレーションの後、物理量 A の任意の温度におけるカノニカルアンサ

サンプル平均は、

$$\langle A \rangle_T = \frac{\sum_E A(E) n(E) e^{-bE}}{\sum_E n(E) e^{-bE}}$$

で与えられるが、その中の状態密度 $n(E)$ の最適解は次の多ヒストグラム再重法で与えられる。焼き戻しシミュレーションにおいて、温度 $T_m = \frac{1}{k_B b_m}$ ($m = 1, \dots, M$) で得られたポ

テンシャルエネルギー分布のヒストグラムと全サンプル数を、それぞれ $N_m(E)$ と n_m とする。このとき、状態密度 $n(E)$ の最適解は次の2つの方程式を逐次解くことによって得られる。

$$n(E) = \frac{\sum_{m=1}^M g_m^{-1} N_m(E)}{\sum_{m=1}^M g_m^{-1} n_m e^{f_m - b_m E}},$$

$$e^{-f_m} = \sum_E n(E) e^{-b_m E}$$

ここで、 $g_m = 1 + 2 \tau_m$ であり、 τ_m は温度 T_m における積分された自己相関時間である。また、 f_m は無次元化されたヘルムホルツ自由エネルギーである。

焼き戻し法に対して、レプリカ交換法の系は M 個の違う温度 T_m ($m = 1, \dots, M$) をもつ、相互作用しない M 個の独立なコピー（レプリカ）からなる。レプリカを表すラベルを i 、温度を表すラベルを m とするとき、 i と m は置換関係にある ($i = i(m)$, $m = m(i)$)。系の状態 X は各レプリカの状態 $X_m^{[i]}$ ($q^{[i]}, p^{[i]}_m$) により決まっているが、レプリカ同士は相互作用しないので、状態 X の重み因子はボルツマン因子の積で与えられる。

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 を交換することを考える。すると、詳細釣り合いの条件を課すことによって、平衡状態への収束が保証されるが、レプリカ対の交換確率は次のメトロポリスの判定条件で与えられる。

$$w(X \rightarrow X') \equiv w(x_m^{[i]} | x_n^{[j]}) = \begin{cases} 1, & \text{for } \Delta \leq 0, \\ \exp(-\Delta), & \text{for } \Delta > 0. \end{cases}$$

ここで、 Δ は次のように与えられる。

$$\Delta \equiv (b_n - b_m)(E(q^{[i]}) - E(q^{[j]})).$$

レプリカ交換シミュレーションは、次のステップを交互に繰り返すことで実行される。

- (1) 各レプリカにおいて、一定温度 T_m ($m=1, \dots, M$) の独立なカノニカルアンサンブル上の従来の MC または MD シミュレーションを同時にあるステップ数実行する。
- (2) 隣接した温度 (T_m と T_{m+1}) に対応するレプリカ対 (i と j) を上式の遷移確率 $w(x_m^{[i]} | x_{m+1}^{[j]})$ に従って交換する。

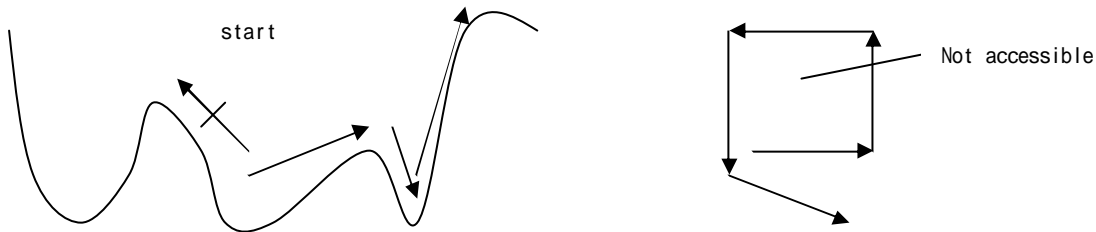
ステップ(2)では隣り合う温度値に対応するレプリカ対を交換することに注意。これは、レプリカ交換の受け入れ確率が温度差に従って指数関数的に減少するからである。(上式を参照)。レプリカ交換法は特に並列計算機に適している。各レプリカを各計算機(CPU)に対応させることによって、計算機間の情報通信量を最小にできる(座標と運動量に代わりに温度を交換するだけである)。

NOTE

レプリカ交換法の利点は、重み因子が最初から分かっていることである(すなわち、ボルツマン因子の積である)。マルチカノニカル法や焼き戻し法では、重み因子の決定のために大変な労力を必要とする場合がある。しかし、レプリカ交換法では、系の自由度が大きくなるにつれて多くのレプリカ(よって多くの計算時間)を必要とする、という困難が存在する。

8. ポテンシャル平滑化法

タブー探索(Science)は、最も効率の良い探索手法として知られている meta heuristics の手法の一つであり、通常のローカルな極小探索（最急降下法など）により極小を探索、極小に達すると今まで来たのとは別の方向であって最も勾配の緩い坂を登って鞍点に達し、再度別の極小に向かって勾配を降りていく方法である。タブー探索の、元の経路に戻らない性質はメモリーを消費するものの良い性質である。酔歩では時間 T に対し、分布の平均半径は \sqrt{T} に比例する。半径 L を探索するのに L^2 の経路が必要になる(フラクタル次元が 2)。自己回避酔歩では、フラクタル次元が 1.33、経路は $L^{1.33}$ で済むことが知られている。低次元の空間では自己回避にまつわる問題が発生し、探索が不十分になるが高次元の空間では大きな問題にならない。



タブー探索を MD に実装するには、自己回避と過去の経路をいかに実現するかが問題となる。既に Filling potential の名前で試みられているらしい。過去に経験した座標にトラジェクトリーが戻ってこないように、過去にとった構造周辺に反発ポテンシャルを発生させること、また過去の記憶（タブー長）が長くなるとメモリーや計算に支障をきたす為、代表的な特徴（ランドマーク）のみを記憶する。

以下に関数系を示す。記憶する経路・過去の構造はタイムステップ 1000step に 1 回構造を記録する。また構造の全原子を記録するのではなく、ランドマークになる特定の原子をいくつか記録する。k はその構造の番号、 x_i^k , y_i^k , z_i^k は、構造 k でのランドマークの原子の座標である。 $a_x(k, i)$, $a_y(k, i)$, $a_z(k, i)$ は、 x_i^k , y_i^k , z_i^k の揺らぎ rmsd の逆数に相当する値、 $c(k)$ (>0) はポテンシャルの高さである。ガウス型関数にしたのは、極小点でのポテンシャルに $V_{fillpot}$ を加算することによりポテンシャルをかなり平滑にできる（1 次微分と 2 次微分が 0）にすることが原理的に可能だからである。

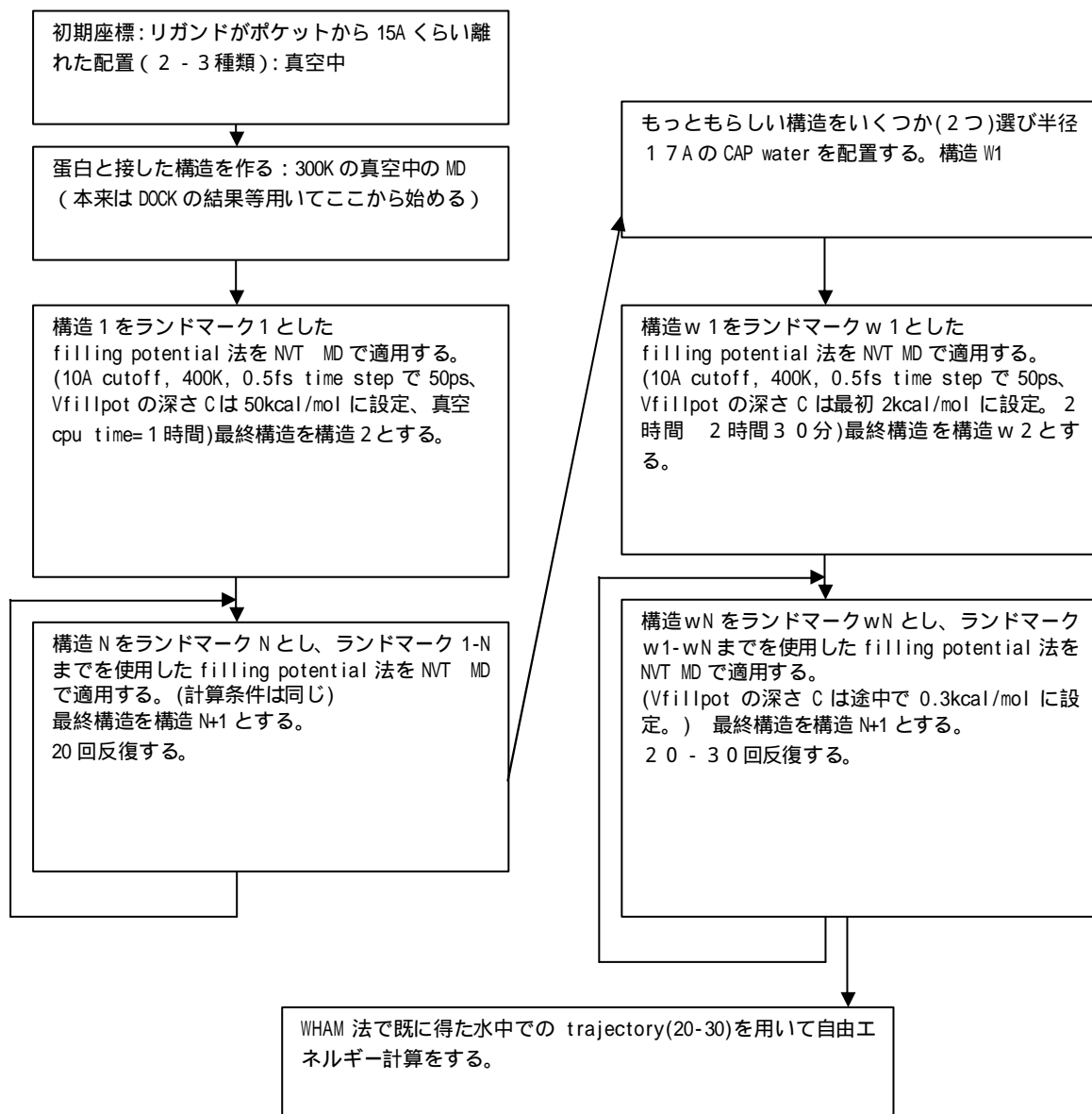
$$V_{fillpot} = \sum_{k=1}^{no_of_pot} c(k) \exp\left(-\sum_i^{landmark} a_x(k, i)(x_i - x_i^k)^2 + a_y(k, i)(y_i - y_i^k)^2 + a_z(k, i)(z_i - z_i^k)^2\right)$$

この関数系は、全てのランドマーク原子が過去の座標のランドマーク原子と重なるときの

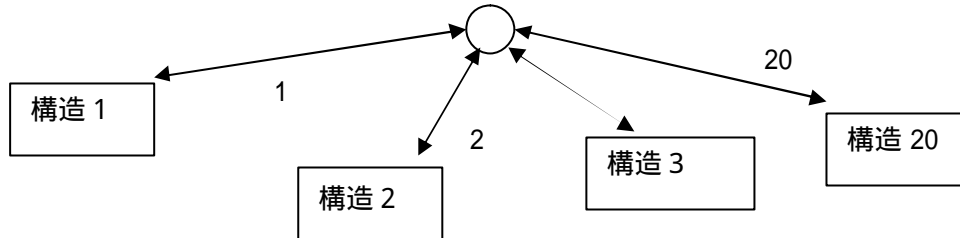
み、大きな値を持ち、1つでも原子が重ならなければ急速にポテンシャル障壁は減少する性質をもっている。リガンドの母核が蛋白に吸着した場合、母核の位置が動かなくても側鎖が回転すれば反発ポテンシャルは消失し、部分構造の最適化が生じうるように作られている。 $V_{fillpot}$ の代わりに $V_{harmonic}$ のような関数を比較してみる。

$$V_{harmonic} = \sum_{k=1}^{no_of_pot} c(k) \sum_i^{landmark} \exp(a_x(k,i)(x_i - x_i^k)^2 + a_y(k,i)(y_i - y_i^k)^2 + a_z(k,i)(z_i - z_i^k)^2)$$

これは、1つでもランドマーク原子が過去の座標の原子と重なれば、その1原子分の反発ポテンシャルが加算されるため、母核が吸着しても側鎖を動かしているあいだに反発ポテンシャルが高まり、十分な探索をする前に解離せねばならない場合がある。



filling potential 法での計算は、結果的に、水中・真空中とも、吸着 構造探索 解離の経路をたどっているため、反応経路として filling potential 法の結果をそのまま利用できる。



反応座標は $V_{fillpot}$ に従い：

$$\mathbf{z}_k = \sum_i^{landmarkatom} (a_x(k,i)(x_i - x_i^k)^2 + a_y(k,i)(y_i - y_i^k)^2 + a_z(k,i)(z_i - z_i^k)^2)$$

自由エネルギー計算

(a) 単一ヒストグラム法 : single histogram method

存在確率の分布 $P(\mathbf{x})$: \mathbf{x} は座標、 E と自由エネルギー $F(\mathbf{x})$ は、 $F(\mathbf{x}) = -kT \ln P(\mathbf{x})$ の関係にある。ヒストグラムは、次のように与えられる。

$$P(\mathbf{x}) = \frac{\int d(x-\mathbf{x}) \exp(-\mathbf{b}E(x)) dx}{\int \exp(-\mathbf{b}E(x)) dx}$$

この定義式に基づき、ポテンシャル $E+V$ のもとでのヒストグラム $P_{E+V}(\mathbf{x})$ とポテンシャル E のもとでのヒストグラム $P_E(\mathbf{x})$ との関係は；

$$\begin{aligned} P_{E+V}(\mathbf{x}) &= \frac{\int d(x-\mathbf{x}) \exp(-\mathbf{b}(E(x)+V(x))) dx}{\int \exp(-\mathbf{b}(E(x)+V(x))) dx} \\ &= \frac{\int d(x-\mathbf{x}) \exp(-\mathbf{b}V(x)) \exp(-\mathbf{b}E(x)) dx}{\int \exp(-\mathbf{b}E(x)) dx} \frac{\int \exp(-\mathbf{b}E(x)) dx}{\int \exp(-\mathbf{b}V(x)) \exp(-\mathbf{b}E(x)) dx} \\ &= \exp(-\mathbf{b}V(\mathbf{x})) P_E(\mathbf{x}) < \exp(-\mathbf{b}E(x)) >_{E+V} \end{aligned}$$

従って、凹凸のあるポテンシャル面で、サンプルされにくいところに束縛型のポテンシャル V を置けば、ポテンシャル面全体の P ないし自由エネルギー面 (pmf) を揃った精度で描ける。

(2) 重複法 : overlapping method

単一ヒストグラム法では、trajectory 1本の解析方法である。計算を分割し、2本の trajectory の結果を使いたいときは、重複法が使える。図のようにポテンシャル面で A/B の極小があり、通常 A ばかりサンプルされるが、状態 B に興味があり、A-B の差を求めたいときを考える。状態 A に拘束する potential を加え全 potential を V_0 , と状態 B に拘束する potential を加え全 potential を V_1 とし、それぞれの potential の元で別々に2つの MD を行う。

V_0 での trajectory T_0 を用い、 $V = V_1(x) - V_0(x)$ のヒストグラム $P_0(V)$ を作成する。
 V_1 での trajectory T_1 を用い、同様にヒストグラム $P_1(V)$ を作成する。単一法で座標 と置いた物がここでは反応座標 はエネルギー差 V に置いた。

$$P_0(\Delta U) = \frac{\int d((V_1(x) - V_0(x)) - \Delta V) \exp(-bV_0(x)) dx}{\int \exp(-bV_0(x)) dx}$$

$$P_1(\Delta U) = \frac{\int d((V_1(x) - V_0(x)) - \Delta V) \exp(-bV_1(x)) dx}{\int \exp(-bV_1(x)) dx}$$

関数の中身から P_0 に関する式では、 $V_0(x) = V_1(x) - \Delta V$ 、 P_1 の式では

$V_1(x) = V_0(x) - \Delta V$ だから

$$P_0(V) = \exp(-bV_1) \exp(b\Delta V) / Z_0 : \text{分母は partition function}$$

$$P_1(V) = \exp(-bV_0) \exp(b\Delta V) / Z_1$$

ヒストグラムの比をとると、各系の分配関数の比は系の間の自由エネルギー差 F なので、

$$\frac{P_1(\Delta V)}{P_0(\Delta V)} = \frac{Z_0}{Z_1} \exp(-b\Delta V) = \exp(\Delta F) \exp(-b\Delta V)$$

従って、trajectory T_0 , T_1 で、系 0 と 1 のトポロジーファイルを入れ替えてエネルギー計算をしておし、エネルギー差のヒストグラムを作成して比をとれば、状態 A-B 間の自由エネルギー差がわかる。これは、trajectory T_0 , T_1 のゲタが F ということなので、単一ヒストグラム法で拘束ポテンシャルを除き、 F の補正を加えれば全体のポテンシャル面も作れる。

上式はすべての V について定数 F を与えるが、実際にはサンプルがない場合、分母や分子がゼロになりエネルギーは発散するので、 P_1 , P_0 は十分な重なりをもっていないといけない。

(3) WHAM 法:

R 回の異なる温度での MD の trajectory の結果を 1 つにまとめる方法。摂動ポテンシャルは L 種類で m 番目のシミュレーションでは、 $\sum_{j=0}^L \mathbf{I}_{j,m} V_j$ のポテンシャルを用いているとする。m 番目の計算の Boltzman 因子は β_m 、サンプル数は n_m 、bin $\{V\}$ 、に含まれるサンプル数 (ヒストグラム) N_m とする。 f_m は、m 番目の状態の自由エネルギー (kbT 単位) である。WHAM 法は、R 個の trajectory のデータ数を考慮し、推定誤差が最小になるように各 trajectory に重みをつけてまとめ、全体のヒストグラムを以下の式

$$P_{\{\mathbf{I}\}, \mathbf{b}}(\{V\}, \mathbf{z}) = \frac{\sum_{k=1}^R N_k(\{V\}, \mathbf{z}) \exp(-\mathbf{b} \sum_{j=0}^L \mathbf{I}_j V_j)}{\sum_{m=1}^R n_m \exp(f_m - \mathbf{b}_m \sum_{j=0}^L \mathbf{I}_{j,m} V_j)}$$

$$\exp(-f_j) = \sum_{\{V\}, \mathbf{z}} P_{\{\mathbf{I}\}_j, \mathbf{b}_j}(\{V\}, \mathbf{z})$$

を同時に解くことで求める方法である。

ここで分布関数 P は各摂動ポテンシャルエネルギーの関数であり、座標 \mathbf{z} 以外にエネルギーの座標も bin にとらねばならない。分布関数 P が余りに高次の空間で定義されると、ヒストグラムのデータが疎になり、まともな計算ができない。 \mathbf{z} をタブー探索のランダムマークをもとにした値とし、摂動ポテンシャル V が \mathbf{z} のみの関数であれば、 \mathbf{z} の次元は、Vfillpot を構成した座標の数だけあることになり、上の式は以下ようになる。各シミュレーションの温度・サンプル数を揃えるなどすると更に簡単になる。

$$P_{\{\mathbf{I}\}, \mathbf{b}}(\mathbf{z}) = \frac{\sum_{k=1}^R N_k(\mathbf{z}) \exp(-\mathbf{b} \sum_{j=0}^L \mathbf{I}_j V_j)}{\sum_{m=1}^R n_m \exp(f_m - \mathbf{b}_m \sum_{j=0}^L \mathbf{I}_{j,m} V_j)}$$

$$\exp(-f_j) = \sum_{\{V\}, \mathbf{z}} P_{\{\mathbf{I}\}_j, \mathbf{b}_j}(\mathbf{z})$$

filling potential 法の計算に 20 構造使ったとすると、 \mathbf{z} は 20 次元のベクトルとなり、これでも多い。そこで精度を上げるため、ヒストグラムを構成する前のデータを用いる。

$$\exp(-f_i) = \sum_{k=1}^R \sum_{t=1}^{nk} \frac{\exp(-\mathbf{b}_i \sum_{j=0}^L \mathbf{I}_{jt} V_{jt}^{(k)})}{\sum_{m=1}^R n_m \exp(f_m - \mathbf{b}_m \sum_{j=0}^L \mathbf{I}_{jm} V_{jt}^{(k)})}$$

プログラムは、ヒストグラムにデータを溜め込まないため、トラジェクトリーを1回の計算に R^2 回読み込まねばならない。

単純な $V_{fillpot}$ では、極小を埋めたとき、もともと吸着に活性化エネルギーが必要な場合や、一度に埋める potential の高さが高すぎる場合、最後の potential を埋めたときに結果的にポテンシャル面が盛り上がる場合がある。このとき、系は遠い場所まで転がり落ち、トラジェクトリーの分布に重なりがなくなることでヒストグラム法が破綻する。そこで、反応座標が求まった後で、構造 m の V として構造 1 ~ m までの V に構造 m を中心に求心的な調和ポテンシャルを加算することでトラジェクトリーの重なりを失わないようにする。

また、 N kcal/mol 以下の binding energy のものに興味がないなら $V_{fillpot}$ をあらかじめ N kcal/mol 程度に設定し、直ちに解離するものを省くことも可能である。

$$V_{fillpot} = \sum_{k=1}^{no_of_pot} c(k) \exp\left(\sum_i^{landmark} a_x(k,i)(x_i - x_i^k)^2 + a_y(k,i)(y_i - y_i^k)^2 + a_z(k,i)(z_i - z_i^k)^2\right) + 2c(last) \sum_i^{landmark} (a_x(last,i)(x_i - x_i^{last})^2 + a_y(last,i)(y_i - y_i^{last})^2 + a_z(last,i)(z_i - z_i^{last})^2)$$

NOTE

リガンドドッキングの系では、 $P(\quad)$ を求めることはあまり重要でないように思われる。それは、吸着した構造があまりにタイトにくっついていてあえて P を求める必要がない、要するに、リガンドが解離している状態と吸着状態の自由エネルギーの差が重要なので、 f を求めることが最も大切で、 f さえあれば事足りることが多いと思えるからである。

Multicanonical 法も Umbrella sampling の1手法なので、WHAM 法と組み合わせて用いることができる。WHAM 法の結果に対し、全体の結果に Single histogram 法で Multicanonical の Weighting をかけてあげると様々な摂動ポテンシャルを Multicanonical MD で用いた場合の Multicanonical WHAM の式が得られる。

■ 参考文献

Multicanonical 法 :

- [1] N. Nakajima, H. Nakamura, A. Kidera, Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides, *J. Phys. Chem. B* (1997) 101, 817.
- [2] J. Higo, N. Nakajima, H. Shirai, A. Kidera, H. Nakamura, Two-component multicanonical Monte Carlo method for effective conformational sampling, *J. Comp. Chem.* (1997) 18, 2086.
- [3] N. Nakajima, A selectively enhanced multicanonical molecular dynamics method for conformational sampling of peptides in realistic water molecules, *Chem. Phys. Letters*, (1998) 288, 319.
- [4] H. Shirai, N. Nakajima, J. Higo, A. Kidera, H. Nakamura, Conformational sampling of CDR-H3 in antibodies by multicanonical molecular dynamics simulation, *J. Mol. Biol.* (1998) 278, 481.
- [5] S. T. Kim, H. Shirai, N. Nakajima, J. Higo, H. Nakamura, Enhanced conformational diversity search of CDR-H3 in antibodies: role of the first CDR-H3 residue, *PROTEINS: Structure, Function, and Genetics* (1999) 37, 683.
- [6] H. Nakamura, S. Ono, J. Higo, *Proc. Japan Acad.* A general ab initio approach for free energy landscapes of biological molecules around the transition states: Fusion of the classical molecular mechanics simulation and the quantum chemical calculation. (1999) *Proc. Japan Acad.* 75 Ser.B, 291.
- [7] N. Nakajima, J. Higo, A. Kidera, H. Nakamura, Free energy landscape of peptides by enhanced conformational sampling, *J. Mol. Biol.* (2000) 296, 197.
- [8] S. Ono, J. Higo, H. Nakamura, *J. Comp. Chem.* (2000) Peptide free energy profile is strongly dependent on the force field: Comparison of C96 and AMBER95. *J. Comp. Chem.* 9, 748.

Multicanonical WHAM 法 :

- [9] S. Ono, N. Nakajima, J. Higo, H. Nakamura, The multicanonical weighted histogram analysis method for the free-energy landscape along structural transition paths, *Chem. Phys. Letters*, (1999) 312, 247.

レプリカ交換法 :

- [10] Y. Sugita and Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, Chem. Phys. Letter (2000) in press.

Tsallis Dynamics :

- [11] S. Abe and Y. Okamoto eds., Nonextensive Statistical Mechanics and Its Applications, (Springer-Verlag, Berlin, 2001).
- [12] S. Abe and A. K. Rajagopal, J. Phys. A 33, 8733 (2000).
- [13] I. Andricioaei and J. E. Straub, J. Chem. Phys. 107, 9117 (1997).
- [14] I. Fukuda and H. Nakamura, Phys. Rev. E 65, 026105 (2002).
- [15] W. G. Hoover, Phys. Rev. A 31, 1695 (1985).
- [16] S. Nose, J. Chem. Phys. 81, 511 (1984).
- [17] H. A. Posch, W. G. Hoover, and F. J. Vesely, Phys. Rev. A 33, 4253 (1986).
- [18] Y. Pak and S. Wang, J. Chem. Phys. 111, 4359 (1999); J. Phys. Chem. B104, 354 (2000).
- [19] C. Tsallis, J. Stat. Phys. 52, 479 (1988).
- [20] an updated bibliography is given in <http://Tsallis.cat.cbpf.br/biblio.htm>.
- [21] C. Tsallis, R. S. Mendes, and A. R. Plastino, Physica A 261, 534 (1998).