

myPresto/cosgene 方式設計書
(構造探索エンジン編)

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. エネルギー最小化処理	4
2.1 エネルギー最小化処理部の構成	4
3. MD 計算処理	5
3.1 MD 計算処理部の構成	5
3.2 Velocity Verlet 法	6
3.3 Nose-Hoover 法	7
3.4 Ewald 法	9
3.5 Particle Mesh Ewald 法	11
3.6 蛋白はりばてモデル	13
4. 入出力ファイル	18
4.1 入力ファイル	18
4.2 出力ファイル	21
4.3 エラーチェック機能	23
5. ユーティリティ	25
5.1 エネルギー確率分布計算	25
5.2 Multicanonical パラメータ作成	25
5.3 Reweighting	25
5.4 トラジェクトリー解析	26
5.5 はりばて用振動解析	28
参考文献	29

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」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。

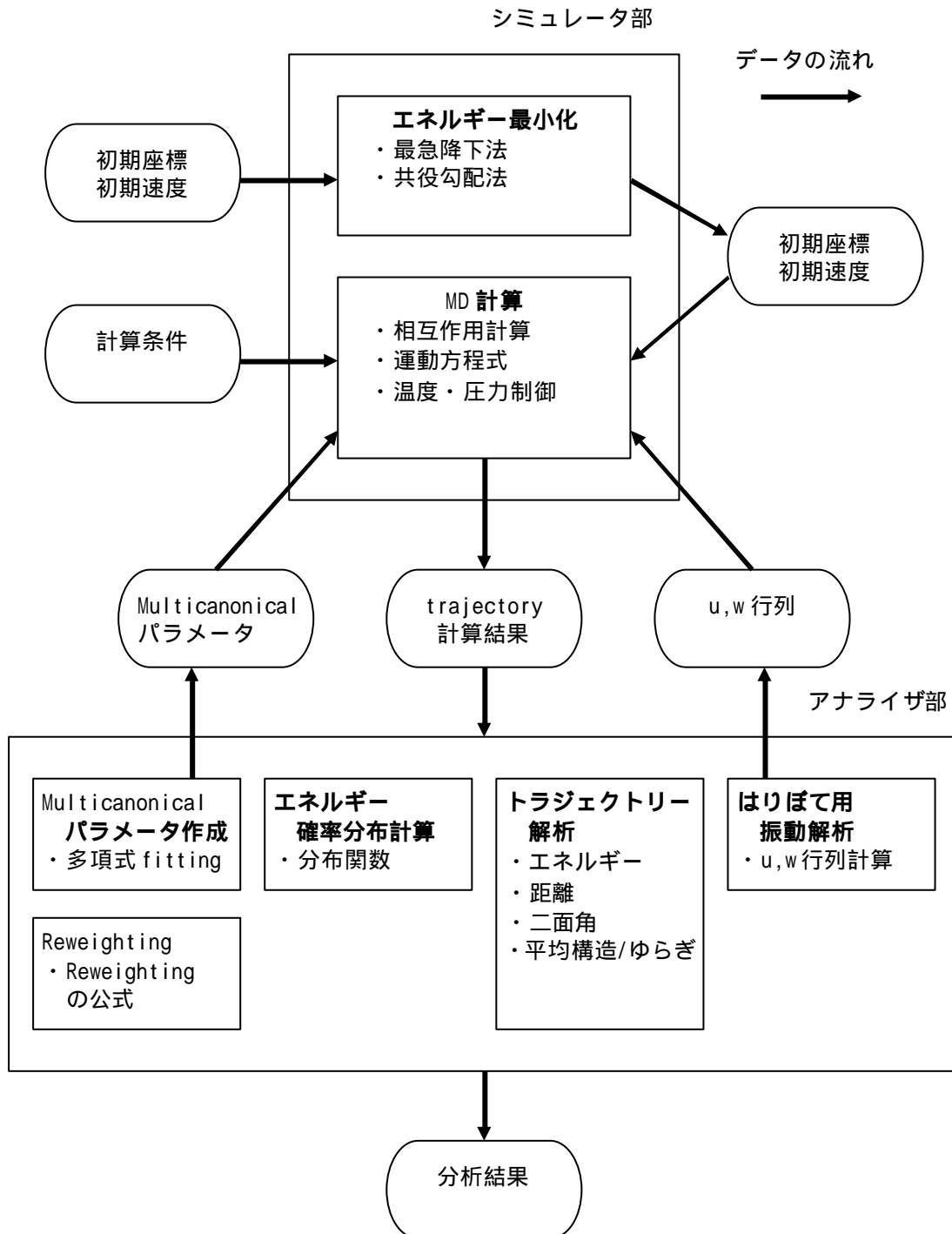


図1 構造探索エンジンの構成

2. エネルギー最小化処理

2.1 エネルギー最小化処理部の構成

エネルギー最小化処理の入出力情報およびそれらの関連図を以下に示す。エネルギー最小化機能の実現方式については、「myPresto/cosgene 方式設計書（エネルギー最小化 / MD 計算結果分析機能編）」を参照のこと。

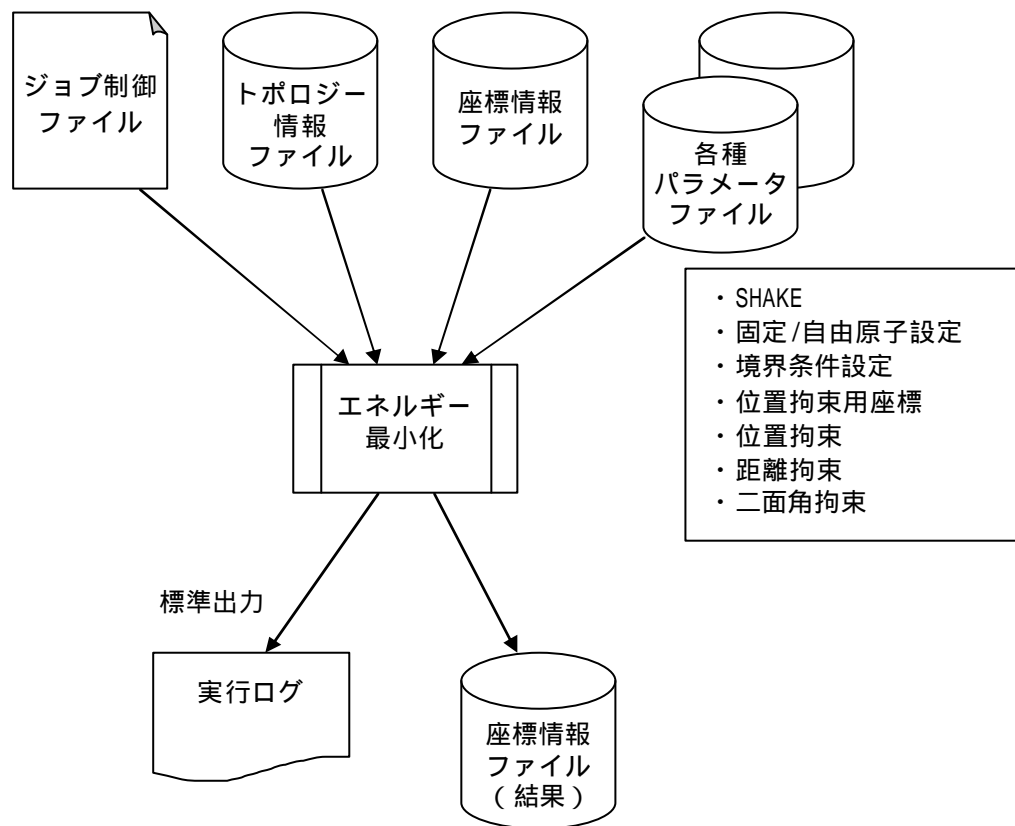


図2 エネルギー最小化処理部の構成

3. MD 計算処理

3.1 MD 計算処理部の構成

MD 計算処理の入出力情報およびそれらの関連図を以下に示す。MD 計算機能の実現方式については、「myPresto/cosgene 方式設計書 (NVE および NVT アンサンブルにおける MD 計算機能編)」および「myPresto/cosgene 方式設計書 (Multicanonical アンサンブルにおける MD 計算機能編)」を参照のこと。

次節以降に主な計算手法の概要を示す。

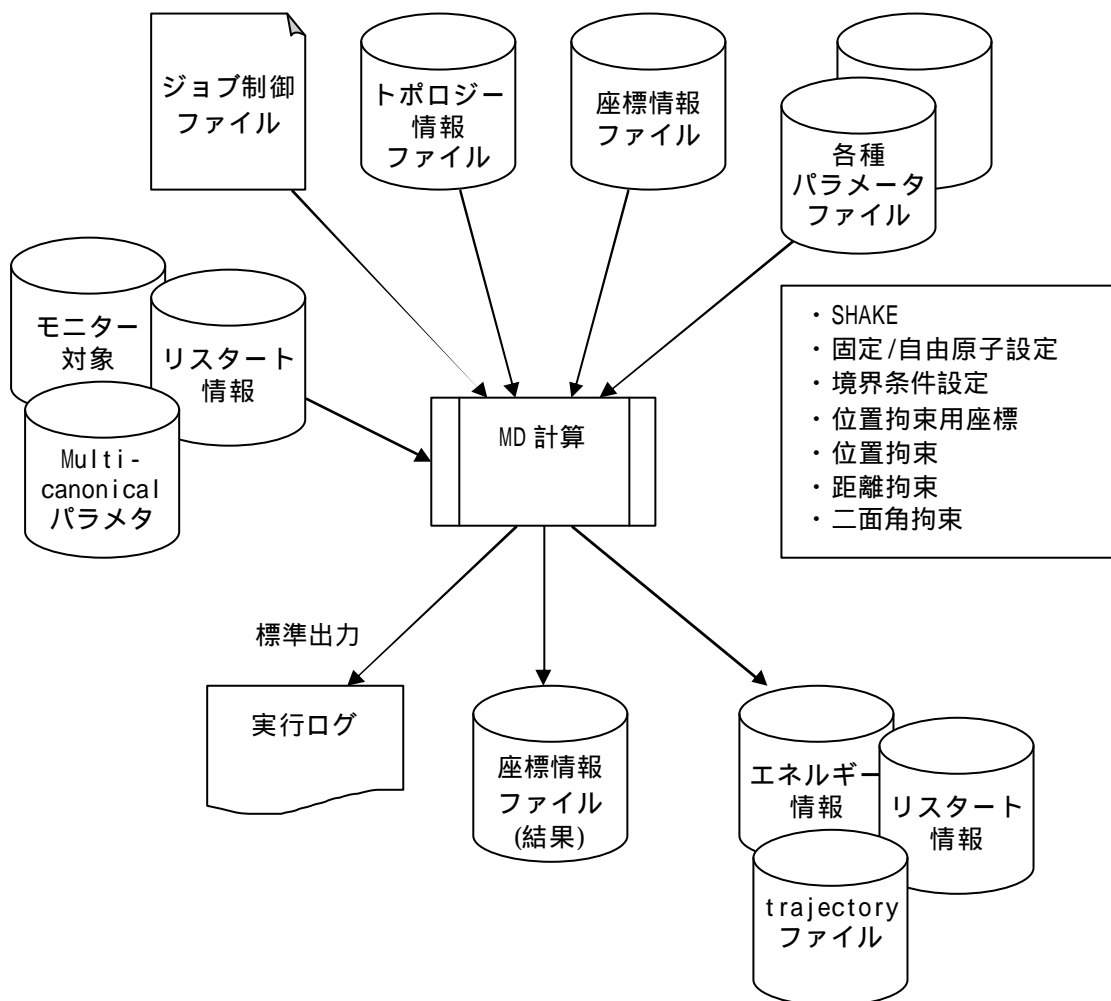


図3 MD 計算処理部の構成

3.2 Velocity Verlet 法

Velocity Verlet 法では以下の差分式で原子の運動方程式を解く。

$$\mathbf{v}_i(t + \frac{\Delta t}{2}) = \mathbf{v}_i(t) + \frac{\Delta t}{2} \cdot \frac{\mathbf{f}_i(t)}{m_i} \quad (1)$$

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \cdot \mathbf{v}_i(t + \frac{\Delta t}{2}) \quad (2)$$

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t + \frac{\Delta t}{2}) + \frac{\Delta t}{2} \cdot \frac{\mathbf{f}_i(t + \Delta t)}{m_i} \quad (3)$$

ここで、 $\mathbf{r}_i, \mathbf{v}_i, \mathbf{f}_i, m_i$ は、それぞれ原子 i の座標、速度、力、質量、 t はタイムステップである。この場合の擬似プログラムを以下に示す。

```

given [r(t),v(t)][m,dt]
call force(f)
do loop=1,nstep
  v = v + (dt/2) * (f/m)
  r = r + dt*v
  call force(f)
  v = v + (dt/2)*(f/m)
enddo

```

一方、leap-frog 法では、

$$\mathbf{v}_i(t + \frac{\Delta t}{2}) = \mathbf{v}_i(t - \frac{\Delta t}{2}) + \Delta t \cdot \frac{\mathbf{f}_i(t)}{m_i} \quad (4)$$

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \cdot \mathbf{v}_i(t + \frac{\Delta t}{2}) \quad (5)$$

この場合の擬似プログラムは、

```

given [r(t),v(t)][m,dt]
call force(f)
v(t-dt/2) = v(t) - (dt/2)*(f/m)
do loop=1,nstep
  v = v + dt * (f/m)
  r = r + dt*v
  call force(f)
enddo

```

となる。

3.3 Nose-Hoover 法

Nose-Hoover 法での仮想粒子も含めた運動方程式は、

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i} \quad (6)$$

$$\dot{\mathbf{p}}_i = \frac{\mathbf{f}_i}{m_i} - \frac{p_x}{Q} \mathbf{p}_i \quad (7)$$

$$\dot{\mathbf{x}} = \frac{p_x}{Q} \quad (8)$$

$$\dot{p}_x = \sum_i \frac{\mathbf{p}_i^2}{m_i} - N_f kT = 2K(\mathbf{p}) - N_f kT \quad (9)$$

で与えられる。ここで、 p_i は原子 i の運動量、 x, p, Q は仮想粒子の座標、運動量、質量である。また、 N_f は系の自由度で $3n$ 、 k, T はそれぞれボルツマン定数、温度である。

擬似プログラムは、

```

given [r(t),v(t)][m,dt][Nf,k,T][pxi=0,Q]
call force(f)
do loop=1,nstep
  2K = sum[mv**2]
  pxi = pxi + (dt/4)*(2K-NfkT)
  xi = xi + (dt/2)*(pxi/Q)
  v = v*exp[-(dt/2)*(pxi/Q)]
  2K = sum[mv**2]
  pxi = pxi + (dt/4)*(2K-NfkT)

  v = v + (dt/2) * (f/m)
  r = r + dt*v
  call force(f)
  v = v + (dt/2)*(f/m)

  2K = sum[mv**2]
  pxi = pxi + (dt/4)*(2K-NfkT)
  xi = xi + (dt/2)*(pxi/Q)
  v = v*exp[-(dt/2)*(pxi/Q)]
  2K = sum[mv**2]
  pxi = pxi + (dt/4)*(2K-NfkT)
enddo

```

普通 velocity verlet

となる。

仮想粒子の質量は、

$$Q = N_f k T t^2 \tag{10}$$

で与えられ、 t は仮想粒子と現実系のカップリングを調節するパラメータで時間の次元を持つ。 t を短くすればきつい温度調節になり、逆に長くするとゆるやかになる。

3.4 Ewald 法

クーロン相互作用は MD 計算において最も計算時間を要する部分であり、シミュレーション全体のパフォーマンスに大きな影響を及ぼす。そのため、計算対象となる系のサイズは、クーロン相互作用の計算方法に強く依存することになる。

一般に大きな系を対象にする場合、周期境界条件のもとで計算を行うが、演算量を軽減するためにカットオフ法が広く用いられてきた。しかしながら、実際は無限遠まで続くポテンシャルを特定の距離で切断してしまうカットオフ法では、信頼性の高い結果を得ることが難しく、現在ではノーカットオフ法である Ewald 法が広く用いられている。[11,12]

今、3つのベクトル $\mathbf{a}_a, a=1,2,3$ で定義されるユニットセル U 内にある N 個の原子を考えると、この系のクーロンエネルギーは、周期境界条件のもとで、

$$E = \frac{1}{2} \sum_{\mathbf{n}} \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}|} \quad (\text{a1})$$

で表される。ここで、 \mathbf{r}_i, q_i は原子 i の座標、電荷であり、 $q_1 + q_2 + \dots + q_N = 0$ である。

また $\sum_{\mathbf{n}}$ は格子ベクトル $\mathbf{n} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3$ (n_1, n_2, n_3 は整数) についての和をとるが、 $i = j, \mathbf{n} = 0$ の場合を除くことを意味する。

式(a1)で表されるポテンシャルエネルギーにおいて、 $1/r$ は0への収束が遅い関数である。Ewald 法では、これを収束の速い2つの項、direct space sum, E_{dir} および reciprocal space sum, E_{rec} と、定数項(または self term), E_{self} の和として以下のように表す。

$$E = E_{dir} + E_{rec} + E_{self} \quad (\text{a2})$$

$$E_{dir} = \frac{1}{2} \sum_{\mathbf{n}} \sum_{i,j=1}^N \frac{q_i q_j \text{erfc}(\mathbf{b} |\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}|)}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}|} \quad (\text{a3})$$

$$E_{rec} = \frac{1}{2pV} \sum_{\mathbf{m} \neq 0} \frac{\exp(-\mathbf{p}^2 \mathbf{m}^2 / \mathbf{b}^2)}{\mathbf{m}^2} S(\mathbf{m}) S(-\mathbf{m}) \quad (\text{a4})$$

$$E_{self} = -\frac{\mathbf{b}}{\sqrt{\mathbf{p}}} \sum_{i=1}^N q_i^2 \quad (\text{a5})$$

ここで、逆格子ベクトル \mathbf{m} は、 $\mathbf{a}_a^* \mathbf{a}_b = \mathbf{d}_{ab}, a, b = 1,2,3$ で定義される \mathbf{a}_a の逆ベクトル \mathbf{a}_a^* を用いて $\mathbf{m} = m_1 \mathbf{a}_1^* + m_2 \mathbf{a}_2^* + m_3 \mathbf{a}_3^*$ (m_1, m_2, m_3 は整数) と表される。

式(a4)において $V = \mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3$ はユニットセルの体積、 $S(\mathbf{m})$ は構造因子

$$S(\mathbf{m}) = \sum_{j=1}^N q_j \exp(2\pi i \mathbf{m} \cdot \mathbf{r}_j) \quad (\text{a6})$$

である。

Ewald 法では、direct space におけるカットオフ r_{cut} 、reciprocal space におけるカットオフ m_{cut} 、およびガウス分布で表される電荷の広がりを規定するパラメータ α がフリーパラメータになる。これらは、 E_{dir} と E_{rec} の演算バランスと、必要な計算精度から最適な値が決定される。

3.5 Particle Mesh Ewald 法

Ewald 法が開発されて以来、現在まで様々な高速化手法が提案されている。中でも Particle Mesh Ewald 法 (PME) は計算精度と高速性の点から最も優れた計算方法の一つであると言える。基本的な考え方は Ewald 法と同じであり、direct space sum、式(a3)は通常の Ewald 法と同様にカットオフを用いて計算する。特徴は reciprocal space sum にあり、式(a4)を、電荷分布を内挿したグリッド上での畳み込みを伴う FFT (Fast Fourier Transform) を用いて評価することにより高速化を実現している。[13,14]

PME 法では、式(a6)の構造因子を

$$\begin{aligned} S(\mathbf{m}) \simeq \tilde{S}(\mathbf{m}) &= \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} \sum_{k_3=0}^{K_3-1} Q(k_1, k_2, k_3) \exp \left[2\mathbf{p} \cdot i \left(\frac{m_1 k_1}{K_1} + \frac{m_2 k_2}{K_2} + \frac{m_3 k_3}{K_3} \right) \right] \\ &= F(Q)(m_1, m_2, m_3) \end{aligned} \quad (\text{a7})$$

のように近似する。ここで、 Q は計算セルを $K_1 \times K_2 \times K_3$ 個に分割したグリッド上に各原子の電荷を内挿して得られる 3 次元の行列であり、 $F(Q)$ はその 3 次元 FFT である。

式(a7)を用いると、式(a4)は

$$\begin{aligned} E_{rec} \simeq \tilde{E}_{rec} &= \frac{1}{2pV} \sum_{\mathbf{m} \neq 0} \frac{\exp(-\mathbf{p}^2 \mathbf{m}^2 / b^2)}{\mathbf{m}^2} F(Q)(m_1, m_2, m_3) F(Q)(-m_1, -m_2, -m_3) \\ &= \frac{1}{2} \sum_{m_1=0}^{K_1-1} \sum_{m_2=0}^{K_2-1} \sum_{m_3=0}^{K_3-1} Q(m_1, m_2, m_3) \cdot (\mathbf{y}_{rec} * Q) Q(m_1, m_2, m_3) \end{aligned} \quad (\text{a8})$$

と近似的に評価できる。

ここで、 \mathbf{y}_{rec} は reciprocal pair potential,

$$\begin{aligned} \mathbf{y}_{rec}(l_1, l_2, l_3) &= \frac{1}{pV} \sum_{\mathbf{m} \neq 0} \frac{\exp(-\mathbf{p}^2 \mathbf{m}^2 / b^2)}{\mathbf{m}^2} \exp \left[2\mathbf{p} \cdot i \left(\frac{m_1 l_1}{K_1} + \frac{m_2 l_2}{K_2} + \frac{m_3 l_3}{K_3} \right) \right] \\ &= F(C)(l_1, l_2, l_3) \end{aligned} \quad (\text{a9})$$

$$C(m_1, m_2, m_3) = \begin{cases} \frac{1}{pV} \sum_{\mathbf{m} \neq 0} \frac{\exp(-\mathbf{p}^2 \mathbf{m}^2 / b^2)}{\mathbf{m}^2} & \text{for } \mathbf{m} \neq 0 \\ 0 & \text{for } \mathbf{m} = 0 \end{cases} \quad (\text{a10})$$

であり、'*' は畳み込みを表す。点電荷のグリッド上への内挿には、当初ラグランジュ補間が使用されていたが、その後より計算精度の高い B スプライン補間が提案されおり、本件でも B スプライン補間を使用する。

PME 法では、Ewald 法における r_{cut} 、 に加え、補間式の次数およびグリッドサイズがフリーパラメータとなり、全体の演算バランスと計算精度から最適な値を選ぶ必要がある。

3.6 蛋白はりばてモデル

- HARmonic potential Interacting BOdy TEchnique (HARIBOTE) -

蛋白全体のシミュレーションを1度行い、リガンドドッキングに必要な部分(または蛋白全体)の運動状態のトラジェクトリーの基準振動解析から、蛋白を弾性体などで近似する。特に蛋白の主鎖の運動を基準振動の重ね合わせで近似し、その振動をばねで近似したときのフックのばね定数を逆算、このはりばて状の弾性体のうち、リガンドとの相互作用に必要な部分にのみ側鎖を生やし、原子をあらわに運動方程式にとりこめるようにする。こうして、リガンドと結合する部分のみ正確にしたモデルを用いれば、主鎖部分のゆらぎを残したまま、フレキシブルドッキングを少ない計算量で行うことができる。

はりばてモデルでは、振動子でできた土台部分 A(特に蛋白の主鎖部分)、あらわに原子を考慮する部分 B(蛋白側鎖、溶媒分子とリガンド)、A,B をとりまく境界条件を与える周辺環境部分 C の3つからなり、全系のハミルトニアンは、次のように書ける。

$$H = H(A,A) + H(A,B) + H(A,C) + H(B,B) + H(B,C) + H(C,C)$$

それぞれの意味は以下の通り。

- ・ $H(A,A)$: 部分 A の調和振動子ポテンシャルと調和振動子の運動エネルギー。【振動子】
- ・ $H(A,B)$: A に含まれる原子と B に含まれる原子との 1-2, 1-3, 1-4, 1-5 相互作用ポテンシャル。【通常】
- ・ $H(B,B)$: B の領域の原子の相互作用と運動エネルギー。【通常】
- ・ $H(B,C)$: C の境界条件 (C に並べた vdW や電荷をあらわす点など) と B の原子との 1-5 相互作用を表している。【vdW + Coulomb】
- ・ $H(A,C), H(C,C)$: 相互作用なし。

表1 各部分との関係

	A	B	C
A	振動子	通常	0
B	通常	通常	vdW+Coulomb
C	0	vdW+Coulomb	0

運動方程式は、振動子の部分 A に対する振動強度の係数 c の運動方程式と、原子部分 B に対する Newton 方程式の 2 つからなる。B に属する原子の運動方程式は、B の原子座標 X として、

$$\dot{\mathbf{p}} = -\frac{\partial H}{\partial X}, \quad \dot{\mathbf{X}} = \frac{\partial H}{\partial \mathbf{p}} \quad (\text{b7'})$$

から、単に A, B, C からの力を受けて運動する通常の MD の運動方程式に書ける。

$$\begin{aligned} \dot{\mathbf{p}} &= -\frac{\partial H(B, B)}{\partial X} - \frac{\partial H(B, A)}{\partial X} - \frac{\partial H(B, C)}{\partial X} \\ &= \mathbf{f}_{B-B} + \mathbf{f}_{B-A} + \mathbf{f}_{B-C} \end{aligned}$$

問題は A の部分の、外力を受けながら、特定の限られた運動モードのみで振動するはりぼて部分の運動方程式である。座標 x 、速度 v の代わりに、振動の強度 c とその時間微分 \dot{c} を使ってハミルトニアンを記述し、これからはりぼて部分の c についての運動方程式と、リアルな原子の部分の座標 x についての運動方程式との連立微分方程式を得る。まず、はりぼて部分のハミルトニアンを作る。固有振動の順番を I とし、各固有振動 I の原子座標の変化を v_I とする。 v の和は、振動のゆっくりしたものに限ってとる。式(b2)左辺は、原子 i の速度ベクトルである。

$$\mathbf{x}_i = x_i^0 + \sum_I c(t)_I v_{I,i} \quad (\text{b1})$$

$$\mathbf{v}_i = \frac{dx_i}{dt} = \sum_I \dot{c}(t)_I v_{I,i} \quad (\text{b2})$$

領域 A (はりぼて) と領域 B (あらわに原子を扱う領域) の間の相互作用を $U(c, X)$ とし、 c は領域 A の振動の係数、 X は領域 B の原子の座標とする。 c に関するハミルトニアンのうち、ポテンシャルエネルギー部分 E_{pot} は、

$$E_{\text{pot}} = U(c, X_i) + \sum_I \frac{k_I}{2} c(t)_I^2 \quad (\text{b3})$$

c に関するハミルトニアンの運動エネルギー部分は、原子 i の運動量 p_i と質量 m_i から、

$$p_i = m_i \frac{dx_i}{dt} = \sum_I \dot{c}(t)_I v_{I,i} m_i \quad (\text{b4})$$

$$E_k = \sum_i \frac{p_i^2}{2m_i} = \sum_i \frac{1}{2m_i} \left(\sum_I \dot{c}(t)_I v_{I,i} m_i \right)^2 \quad (\text{b5})$$

ここで座標 c に共役な運動量 p を、ラグランジュアンから次のように求める、

$$\begin{aligned} p_I &= \frac{\partial L}{\partial \dot{c}_I} = \sum_i \frac{2}{2m_i} \left(\sum_I \dot{c}(t)_I v_{I,i} m_i \right) v_{I,i} m_i \\ &= \sum_i \left(\sum_I \dot{c}(t)_I v_{I,i} m_i \right) v_{I,i} \end{aligned} \quad (\text{b6})$$

この運動量 p をもとに、ハミルトンの方程式

$$\dot{\mathbf{p}} = -\frac{\partial H}{\partial q}, \quad \dot{\mathbf{q}} = \frac{\partial H}{\partial p} \quad (\text{b7})$$

を解く。

$$\dot{p}_I = -\frac{\partial H}{\partial c_I} \quad (\text{b8})$$

左辺は、

$$\dot{p}_I = \sum_i \left(\sum_I \ddot{c}(t)_I v_{I,i} m_i \right) v_{I,i} \quad (\text{b9})$$

右辺は、

$$-\frac{\partial E_{pot}}{\partial c_I} = -k_I c(t)_I - \frac{\partial U(c, \mathbf{X})}{\partial c_I} \quad (\text{b10})$$

右辺第二項を通常の方法から計算できるようにする。式(b1)を用いて、

$$\frac{\partial U(c, \mathbf{X})}{\partial c_I} = \sum_i \frac{\partial U(x, \mathbf{X})}{\partial x_i} \frac{\partial x_i}{\partial c_I} = -\sum_i \bar{f}_i \frac{\partial \left(x_i^0 + \sum_I c(t)_I v_{I,i} \right)}{\partial c_I} = -\sum_i \bar{f}_i v_{I,i} \quad (\text{b11})$$

結局、 c の運動方程式は、両辺をあわせて

$$\sum_i \left(\sum_h \ddot{c}(t)_h v_{h,i} m_i \right) v_{I,i} = -k_I c(t)_I + \sum_i \bar{f}_i v_{I,i} \quad (\text{b12})$$

温度コントロールの場合は、Evans の Gaussian 拘束法を用いることとし、力に原子の速度に比例する摩擦力を加える。 $f = f + \mathbf{xv}$

$$\sum_i \left(\sum_h \ddot{c}(t)_h v_{h,i} m_i \right) v_{I,i} = -k_I c(t)_I + \sum_i \bar{f}_i v_{I,i} + \mathbf{x} \sum_i \left(\sum_h \dot{c}(t)_h v_{h,i} m_i \right) v_{I,i} \quad (\text{c12})$$

このままでは c について解きにくいので、 $v_{,i}$ を行列 v の i 要素とみなし、その逆行列 u を定義する。ここで、逆行列を定義するため、振動モードの数と領域 A の原子数は同じにする必要があることに注意されたい。

$$\sum_I u_{i,I} v_{I,j} = \mathbf{d}_{i,j} \quad (\text{b13})$$

式(b12)の左辺に u をかけ、 i について和をとると

$$\begin{aligned} \sum_I u_{i,I} \dot{p}_I &= \sum_I u_{j,I} \sum_i \left(\sum_h \ddot{c}(t)_h v_{h,i} m_i \right) v_{I,i} \\ &= \sum_I \sum_i \left(\sum_h \ddot{c}(t)_h v_{h,i} m_i \right) u_{j,I} v_{I,i} \\ &= \sum_i \left(\sum_h \ddot{c}(t)_h v_{h,i} m_i \right) \mathbf{d}_{i,j} \\ &= \sum_h \ddot{c}(t)_h v_{h,j} m_j \end{aligned} \quad (\text{b14})$$

式(b12)の右辺に u をかけ、 i について和をとると

$$\begin{aligned} \sum_I u_{j,I} \left(k_I c(t)_I - \sum_i \vec{f}_i v_{I,i} \right) &= \sum_I u_{j,I} (k_I c(t)_I) - \sum_I u_{j,I} \sum_i \vec{f}_i v_{I,i} \\ &= \sum_I u_{j,I} (k_I c(t)_I) - \vec{f}_j \end{aligned} \quad (\text{b15})$$

ここでさらに、逆行列 w を定義する。ここでも式(b13)同様、逆行列を定義するため、振動モードの数と領域 A の原子数は同じにする必要があることに注意されたい。

$$\sum_j w_{j,h} v_{I,j} m_j = \mathbf{d}_{h,I} \quad (\text{b16})$$

式(b14)に行列 w をかけると、

$$\begin{aligned} \sum_j w_{j,I} \sum_h \ddot{c}(t)_h v_{h,j} m_j &= \sum_j \sum_h \ddot{c}(t)_h w_{j,I} v_{h,j} m_j \\ &= \sum_h \ddot{c}(t)_h \mathbf{d}_{h,I} \\ &= \ddot{c}(t)_I \end{aligned} \quad (\text{b17})$$

式(b15)に行列 w をかけると、 c に関する時間発展の微分方程式が得られる。

$$\ddot{c}(t)_I = \sum_j w_{j,I} \left(- \sum_h u_{j,h} (k_h c(t)_h) + \vec{f}_j \right) \quad (\text{b18})$$

温度一定（運動エネルギー一定）となるように、項を付け加える必要がある。温度コントロールの場合は、式を展開すると以下ようになる。

$$\ddot{c}(t)_I = \sum_j w_{j,I} \left(- \sum_h u_{j,h} (k_h c(t)_h) + \vec{f}_j \right) - \mathbf{x} \cdot \dot{c}(t)_I \quad (\text{c18})$$

運動を許す振動モードの数を N としたとき、 \mathbf{x} は、

$$\mathbf{x} = - \frac{\sum_i \frac{\partial U}{\partial x_i} \dot{x}_i}{\sum_i \frac{p_i^2}{2m_i}} = - \frac{dU}{dt} / NkT$$

式の第2項の添え字 i と x_i は原子の番号と座標であるが、第3項ではあらわには含まない形になっている。またこの \mathbf{x} は、はりばて領域 A の振動子の運動方程式でも、領域 B のあらわに扱う原子の運動方程式

$$\frac{dv(t)}{dt} = \frac{f(t)}{m} - \mathbf{x}v(t)$$

でも共通の値となることに注意されたい。

4. 入出力ファイル

4.1 入力ファイル

4.1.1 ジョブ制御ファイル

ジョブ制御ファイルには、エネルギー最小化处理、MD 計算処理およびトラジェクトリー解析処理で使用するパラメータとデータファイル名などの情報を記述する。

4.1.2 トポロジー情報ファイル

トポロジー情報ファイルには、分子のトポロジー情報を記述する。

- ・分子（原子）の構成情報
- ・ポテンシャルエネルギー関数の計算で使用する force constant、など
(Bond, Angle, Torsion, Improper torsion, van der Waals, Electrostatic)

4.1.3 座標情報ファイル

座標情報ファイルには、計算対象の分子（原子）の座標情報を記述する。

データのフォーマットは、PDB 形式である。

(また、PDB 形式のデータをバイナリー表現した形式 (Binary 形式) も使用可能である)

4.1.4 SHAKE ファイル

SHAKE ファイルには、SHAKE 拘束をかける原子のグループと、それらにかかる SHAKE の係数を記述する。

4.1.5 固定 / 自由原子設定ファイル

固定 / 自由原子設定ファイルには、固定原子および自由原子を設定するための情報を記述する。

4.1.6 境界条件設定ファイル

境界条件設定ファイルには、境界条件および基本セルを設定するための情報を記述する。

- ・境界条件に入れる残基
- ・周期境界条件 (periodic boundary condition) の基本セルの長さ
- ・球体イメージ境界 (sphere-image boundary) の半径

4.1.7 位置拘束用座標情報ファイル

位置拘束用座標情報ファイルには、位置拘束をかける分子（原子）の座標情報を記述する。データのフォーマットは、PDB 形式である。

座標情報ファイル（4.1.3）と同形式のファイルである。

4.1.8 位置拘束ファイル

位置拘束ファイルには、位置拘束をかける原子の情報を記述する。

4.1.9 距離拘束ファイル

距離拘束ファイルには、距離拘束をかける原子の組の情報、拘束範囲を記述する。

4.1.10 二面角拘束ファイル

二面角拘束ファイルには、二面角拘束をかける原子の組の情報、拘束範囲を記述する。

4.1.11 モニター対象設定ファイル

モニター対象設定ファイルには、トラジェクトリー出力対象とする構造情報を記述する。

- ・分子（原子）の情報（1 原子）
- ・結合距離の情報（2 原子）
- ・結合角の情報（3 原子）
- ・二面角の情報（4 原子）

4.1.12 リスタート情報ファイル

リスタート情報ファイルには、MD 計算のリスタート用の情報を記述する。

- ・リスタート用パラメータ（それまでの実行回数(時間)、エネルギー情報など）
- ・全原子の座標情報
- ・全原子の速度情報

4.1.13 Multicanonical パラメータファイル

Multicanonical パラメータファイルには、canonical 分布 $\ln P_c(E)$ の情報を記述する。多項式フィッティングした結果の係数などをパラメータ化したもので、重み関数 $W(E)$ を求める際に使用する。Multicanonical MD 実行時に使用する。

Multicanonical パラメータファイルは、Multicanonical パラメータ作成 (5.2) の結果のファイルである。

Multicanonical パラメータ作成の入出力は、以下の通り。

入力 : canonical 分布 $\ln P_c(E)$

出力 : Multicanonical パラメータ

4.1.14 振動解析ファイル

振動解析ファイルには、はりばて用振動解析で計算した以下の情報を記述する。

- ・ u, w 行列 (式(b13)、式(b16))
- ・ 基準振動ベクトル (式(b1)、式(b2)の $v_{,i}$)
- ・ 振動ベクトルの固有値 (式(b3)の k)

振動解析ファイルは、はりばて用振動解析 (5.5) の結果のファイルである。

はりばて用振動解析の入出力は、以下の通り。

入力 : はりばて領域指定ファイル

振動解析を行い、振動子で置換すべき領域を指定するための情報を記述する。

- ・ 対象とする原子番号

出力 : 振動解析ファイル

4.2 出力ファイル

4.2.1 実行ログファイル

実行ログファイルには、エネルギー最小化および MD 計算の各処理で出力するログ情報を出力する。

4.2.2 トポロジー情報ファイル

トポロジー情報ファイルには、分子のトポロジー情報を出力する。

- ・分子（原子）の構成情報
- ・ポテンシャルエネルギー関数の計算で使用する force constant、など
(Bond, Angle, Torsion, Improper torsion, van der Waals, Electrostatic)

4.2.3 座標情報ファイル

計算結果の分子（原子）の座標情報を出力する。

データのフォーマットは、PDB 形式である。

(また、PDB 形式のデータをバイナリー表現した形式 (Binary 形式) も使用可能である)

座標情報ファイル (4.1.3) と同形式のファイルである。

4.2.4 リスタート情報ファイル

リスタート情報ファイルには、MD 計算のリスタート用の情報を出力する。

- ・リスタート用パラメータ (それまでの実行回数(時間)、エネルギー情報など)
- ・全原子の座標情報
- ・全原子の速度情報

MD 計算終了時に出力する。リスタート実行時に、入力ファイルとして使用する。

4.2.5 トラジェクトリーファイル

トラジェクトリーファイルには、一定の時間間隔ごとの構造情報を出力する。

(時間間隔は、ジョブ制御情報ファイルに指定する)

- ・全自由原子の座標
- ・全自由原子の速度
- ・全自由原子のエネルギー
- ・総エネルギー
- ・モニター対象の構造情報

(モニター対象設定ファイルに記述した構造に対して出力する)

NOTE

Binary 形式のファイルを取り扱う際は、エンディアンに注意する必要がある。

出力した計算機と入力する計算機のエンディアンの仕様が異なる場合は、データを正しく読み込むことができない。

4.3 エラーチェック機能

4.3.1 入力データのエラーチェック

入力ファイルに対して、入力データのエラーチェックを行う。入力したファイルや計算のためのパラメータの指定誤りを可能な限り検出し、エラーメッセージなどで通知する。

特に MD 計算は非常に多くの計算時間を消費するものである。もし、入力に誤りがあり、その誤りに気が付かないでいると誤った入力のまま計算し続けることになり、計算機資源と時間の浪費になってしまう。さらに、この計算の結果を正しい計算結果として取り扱ってしまう可能性すらある。このため、MD 計算の前に入力値の誤りの検出を（厳密に）行っておく必要がある。

(1) 処理方式

エラー検出は計算処理のデータ読み込み時に行うのではなく、一旦、一通りのファイルを読み込んで全てのエラーを検出する方式とする。計算処理のデータ読み込み時にエラー検出をする方式だと、計算処理の後半で読み込まれるデータのエラー検出がその時点まで遅れることになる。（これは、冒頭に述べたエラーチェック機能の目的と合致しない）

<対象データ>

- ・ジョブ制御ファイル
- ・各種入力ファイル（4.2 節の入力ファイルを参照）

<出力エラー情報>

以下の情報を出力する。

- ・エラーメッセージ（エラー種別、内容）
- ・エラーを検出した、行、カラム（ジョブ制御ファイルが対象）

<出力例>

ERROR> INPUT	: エラー種別
NOT INTEGER VALUE. LINE=15	: エラーメッセージ、行
LOOPLI=1000*10	: エラー行の内容
^	: エラーの位置（カラム）

4.3.2 ジョブ制御ファイルの作成支援

MD 計算処理では多くのパラメータを使用する。これらのパラメータは、ジョブ制御ファイルに記述する。ジョブ制御ファイルの記述ミスを防ぐため、ジョブ制御ファイルの生成を支援する機能を設ける。

ジョブ制御ファイルには指定しなければならないパラメータが多いため、ジョブ制御ファイルの作成時にパラメータの指定ミスをすることが多い。これらのエラーはエラーチェック（前節）である程度検出することができるが、最初から無いことが望ましい。

（１）処理方式

ジョブ制御ファイルの作成者は、対話的に作成していく。

作成者は、画面に表示された質問に答える形で入力していく。

作成者の回答を受けて、二重定義や矛盾がある組合せとならない質問を表示する。

また、入力エラーは対話的に解決する。

すべての入力が終了した後、ジョブ制御ファイルを作成する。

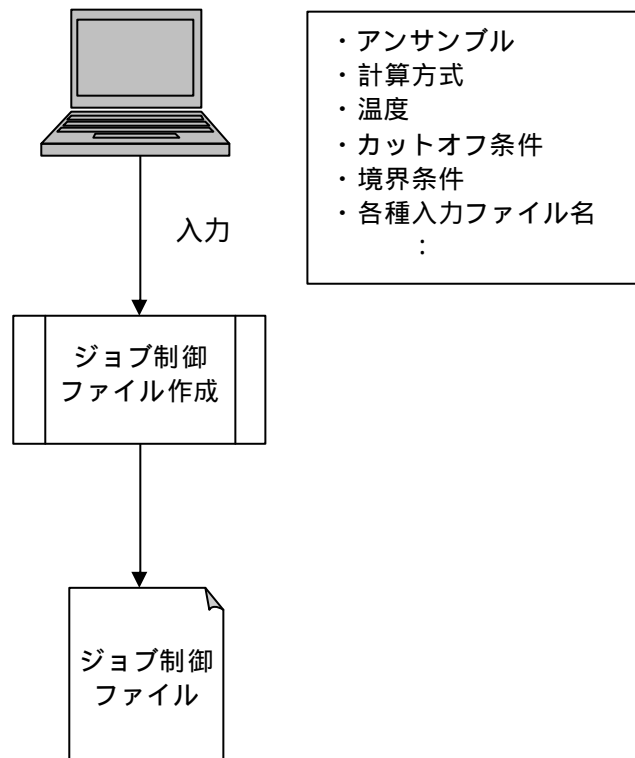


図4 ジョブ制御ファイル作成処理

5. ユーティリティ

5.1 エネルギー確率分布計算

MD 計算によって求めたポテンシャルエネルギーのトラジェクトリー（各タイムステップ毎のポテンシャルエネルギー）から、エネルギー確率分布 $\ln P(E)$ を計算する。

5.2 Multicanonical パラメータ作成

Canonical 分布 $\ln P_c(E)$ に対して多項式フィッティングを行い、 $\ln P_c(E)$ を多項式として求める。Multicanonical アンサンブルにおける MD 計算では、1 回前の MD 計算によって求めた Canonical 分布 $\ln P_c(E)$ を必要とする。求めた多項式（次数、各次の係数の情報）は、次の回の Multicanonical MD 計算および Reweighting 処理のパラメータとして与える。

5.3 Reweighting

Multicanonical アンサンブルにおける MD 計算によって求めた Multicanonical 分布 $\ln P_{mc}(E)$ および Multicanonical パラメータ（多項式の $\ln P_c(E)$ ）から、任意の温度における Canonical 分布 $\ln P_c(E)$ を計算する。（Reweighting の公式を使う）

5.4 トrajジェクトリー解析

MD 計算によって出力される各種 trajectory 等の情報から、以下の各項目に関する物理量を求める。

エネルギーに関する物理量の計算

距離に関する物理量の計算

二面角に関する物理量の計算

平均構造・ゆらぎに関する物理量の計算

(1) エネルギー

MD 計算によって出力される各種 trajectory 等の情報から、エネルギーに関する物理量を求める。エネルギーの平均とゆらぎ (Root Mean Square : RMS) を分析することによって、MD 計算の精度を比較することができる。

< エネルギーのゆらぎの計算 >

$$\langle E \rangle = \frac{1}{N} \sum (E_i)$$

E_i : エネルギー
 N : 全時間ステップ数

$$\text{RMS}(E) = \sqrt{\langle (E - \langle E \rangle)^2 \rangle}$$

E : エネルギー
 $\langle E \rangle$: 平均エネルギー

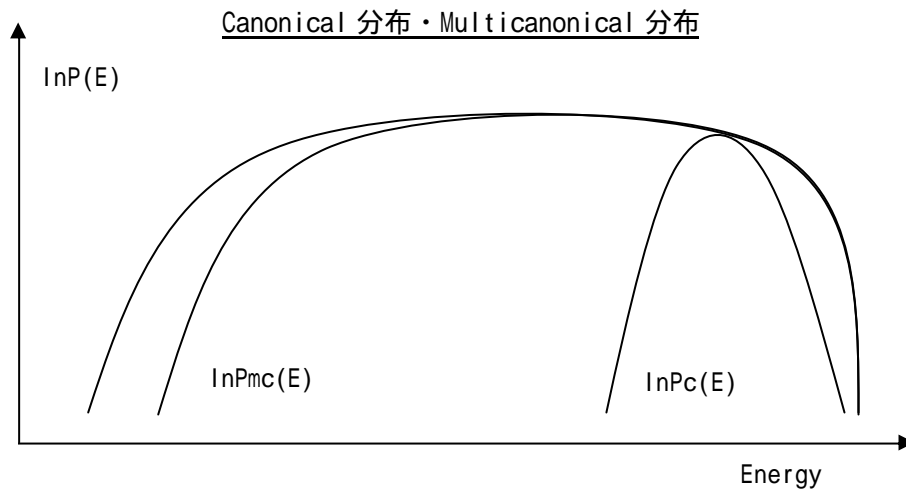
主に以下の項目について情報を出力する。

各種エネルギー (全エネルギー、運動エネルギー、ポテンシャルエネルギー) の平均および標準偏差

温度の平均および標準偏差

Canonical 分布 ($\ln P_c(E)$)

Multicanonical 分布 ($\ln P_{mc}(E)$)



(2) 距離

MD 計算によって出力される各種 trajectory 等の情報から、原子間の距離に関する物理量を求める。

主に以下の項目について情報を出力する。

- 原子の情報 (原子名、残基名)
- 原子間の距離の平均と標準偏差
- 距離に関する分布

(3) 二面角

MD 計算によって出力される各種 trajectory 等の情報から、二面角に関する物理量を求める。

主に以下の項目について情報を出力する。

- 原子の情報 (原子名、残基名)
- 原子間の二面角の平均と標準偏差
- 角度に関する分布

(4) 平均構造・ゆらぎ

MD 計算によって出力される各種 trajectory 等の情報から、平均構造とゆらぎに関する物理量を求める。トポロジーが同じである二つの構造の差は、根平均二乗変位 (Root Mean Square Deviation : RMSD) として計算する。平衡状態になってからの構造のゆらぎは、根平均二乗ゆらぎ (Root Mean Square Fluctuation : RMSF) で計算する。また、基準構造との重ね合わせを行い、その結果の構造についての物理量も求める。

< RMSD の計算 >

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum (r_i - R_i)^2}$$

r_i : 座標
 R_i : 基準座標

< RMSF の計算 >

$$\text{RMSF} = \sqrt{\langle (\mathbf{R} - \langle \mathbf{R} \rangle)^2 \rangle}$$

\mathbf{R} : 座標
 $\langle \mathbf{R} \rangle$: 平均座標

< 重ね合わせ手順 >

- ・重ね合わせる二つの構造の重心を合わせる。
- ・最小二乗法によって、重ね合わせる。(Least Square Fitting)

主に以下の項目について情報を出力する。

根平均二乗変位 (Root Mean Square Deviation : RMSD)

平均構造および標準偏差

根平均二乗ゆらぎ (Root Mean Square Fluctuation : RMSF)

5.5 はりばて用振動解析

MD 計算によって出力されるトラジェクトリー情報から、各原子の RMSF^2 の行列を求めて、対角化する。結果として、振動解析ファイルを出力する。

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.

Ewald Method :

- [11] M.P. Allen and D.J. Tildesley, "Computer Simulation of Liquids", Oxford Science, Oxford (1989).
[12] 岡崎 進, "コンピュータシミュレーションの基礎", 化学同人 (2000).

Particle Mesh Ewald Method :

- [13] T. Darden, D. York, and L. Pedersen, J. Chem. Phys., 98, 1008 (1993).
[14] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L.G. Pedersen, J. Chem. Phys., 103, 8577 (1995).