

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. CHARMMm 力場での相互作用計算	4
2.1 CHARMMm ポテンシャルエネルギー関数	4
2.2 スイッチング関数による長距離相互作用のカットオフ計算	5
3. GB / ASA	6
3.1 Accessible Surface Area (ASA)	6
3.2 Generalized Born accessible surface area	9
4. Multicanonical 法の拡張	12
4.1 Force-Biased Multicanonical Sampling 法	12
4.2 Simulated Tempering 法	13
5. RATTLE 法	14
6. Hybrid MD/HD 法	17
7. 拘束ポテンシャル	22
7.1 CAP 拘束の対象の拡張	22
7.2 楕円体 CAP	23
7.3 範囲付き位置拘束	25
参考文献	26

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。

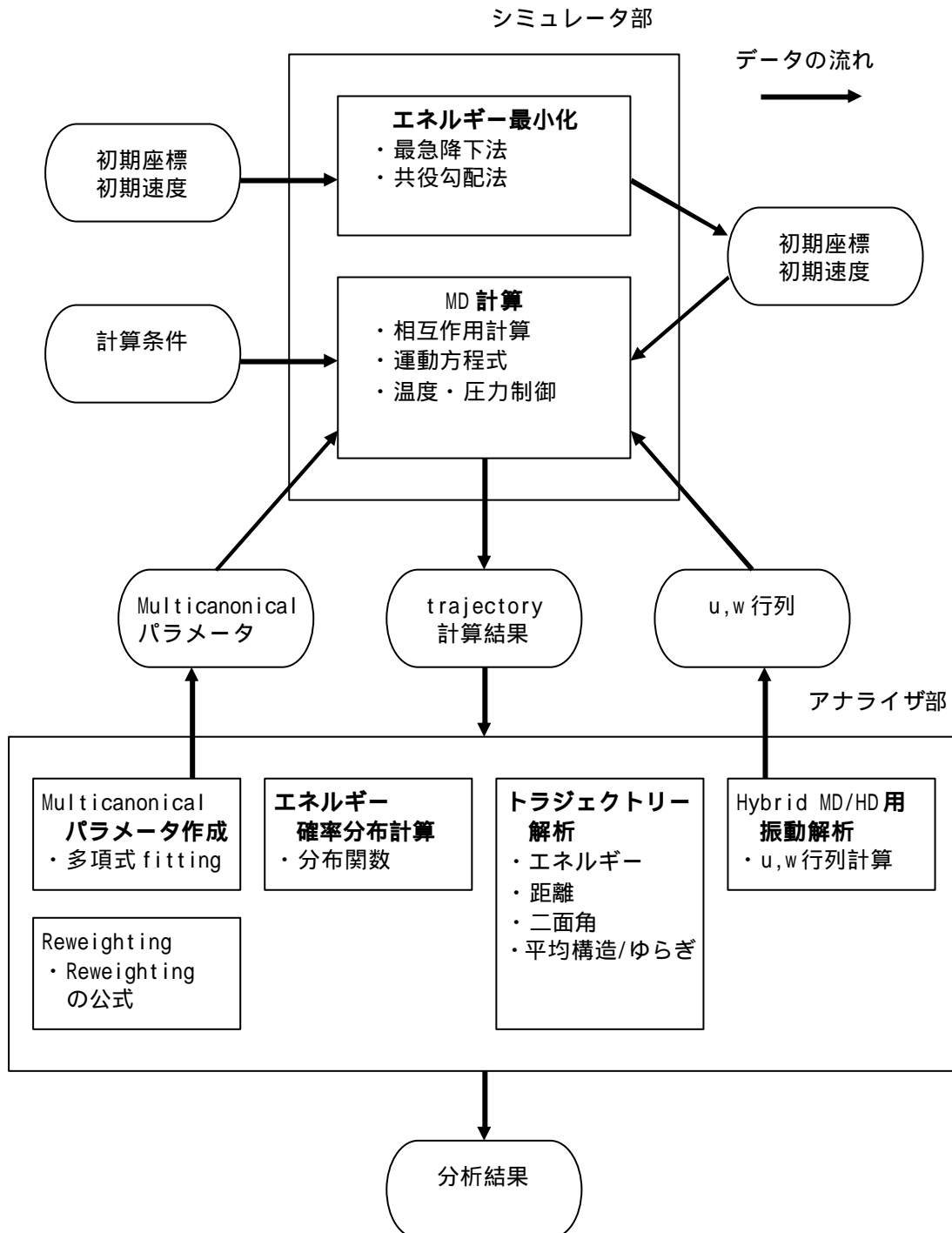


図1 構造探索エンジンの構成

2. CHARMMm 力場での相互作用計算

2.1 CHARMMm ポテンシャルエネルギー関数

CHARMM力場は、生体分子の力場パラメータセットとして、AMBER力場と並んで広く利用されている。CHARMMは米国ハーバード大学のKarplusらが中心に作ったプログラムで、以下のポテンシャル関数を実装している。

$$E_{CHARMM} = E_{bond} + E_{angle} + E_{Urey-Bradley} + E_{torsion} + E_{improper} + E_{nonbond}$$

$$E_{nonbond} = E_{van-der-Waals} + E_{elecstatic}$$

CHARMM、AMBERなど、異なる力場パラメータセットを使用するには、それぞれの力場パラメータに対応するポテンシャルエネルギー関数を使用する。CHARMMとAMBERの違いは、Urey-Bradley、Improper-Torsion、van der Waals、のエネルギー関数である。

表1 CHARMMとAMBERで使用されるポテンシャルエネルギー関数の相違

エネルギー項	CHARMM	AMBER
Urey-Bradley	$E_{ub} = k_{ub} (s - s_0)^2$ (k_{ub} : 力の係数、 s, s_0 : 距離)	なし
Improper-Torsion	$E_{imp} = k_{imp} (\mathbf{q} - ini)^2$ (k_{imp} : 力の係数、 \mathbf{q} : 下図で定義される平面 ABC と平面 BCD との間の二面角、 ini : 理想値) <div style="text-align: center;"> </div>	$E_{imp} = k_{imp} (1 + \cos(n \cdot \mathbf{j} - ini))$ (n : シンメトリ、 \mathbf{j} : 下図で定義される平面 ABC と平面 BCD との間の二面角、 ini : 理想値) <div style="text-align: center;"> </div>
van der Waals	$E_{vdw} = eps_{ij} \left(\left(\frac{R \min_{ij}}{r_{ij}} \right)^{12} - 2 \cdot \left(\frac{R \min_{ij}}{r_{ij}} \right)^6 \right)$ (eps_{ij} : 力の係数、 r_{ij} : 原子 $i - j$ 間の距離、 $R \min_{ij}$: 原子 $i - j$ 間のエネルギー最安定距離)	$E_{vdw} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6}$ (A_{ij}, B_{ij} : 力の係数、 r_{ij} : 原子 $i - j$ 間の距離)

2.2 スイッチング関数による長距離相互作用のカットオフ計算

長距離相互作用の計算は全ての原子ペアに対して行うため、計算量は N^2 (N は原子数) に比例し、多くの時間を消費することになる。多くの場合「カットオフ距離」を導入して計算時間を節約する。これは「二つの原子間の距離がカットオフ距離より大きければ、その原子間の長距離相互作用は計算しない」というカットオフ計算を行うものである。

カットオフ計算を行う場合、長距離相互作用はカットオフ距離のところで非連続となる。MDの繰り返しの中で二つの原子間距離がカットオフ距離付近で変動するとき、この瞬間的な変動分が全体（温度制御など）に影響を与えることがある。そこで、長距離相互作用の項に変動を低減するためのスイッチング関数を導入する。

$$E_{CHARMm} = E_{bond} + E_{angle} + E_{Urey-Bradley} + E_{torsion} + E_{improper} + E_{nonbond}$$

$$E_{nonbond} = (E_{van-der-Waals} + E_{elecstatic}) \cdot SW(r_{ij}, r_{on}, r_{off})$$

以下のスイッチング関数を導入する。

$$SW(r_{ij}, r_{on}, r_{off}) = \begin{cases} 1 & \dots \text{if } r_{ij} < r_{on} \\ \left[\frac{(r_{off}^2 - r_{ij}^2)^2 (r_{off}^2 + 2r_{ij}^2 - 3r_{on}^2)}{(r_{off}^2 - r_{on}^2)^3} \right] & \dots \text{if } r_{on} < r_{ij} < r_{off} \\ 0 & \dots \text{if } r_{ij} > r_{off} \end{cases}$$

3. GB / ASA

3.1 Accessible Surface Area (ASA)

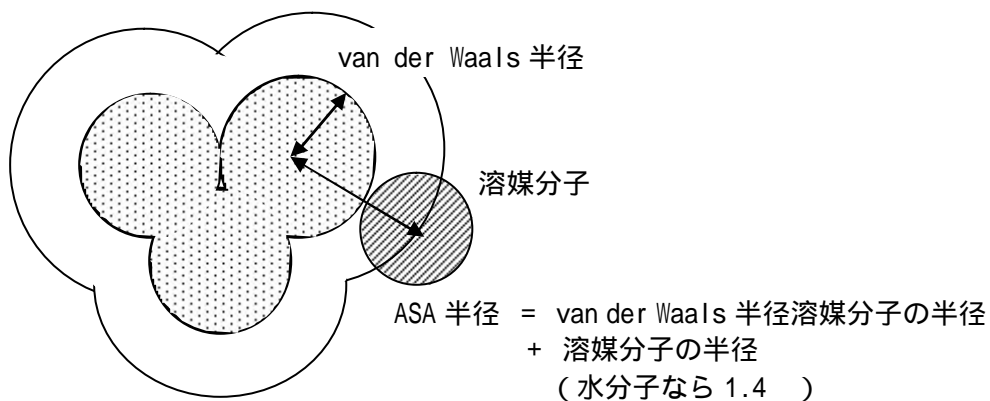
溶質と溶媒の相互作用（溶媒和エネルギー）は、タンパク質など溶質の構造を仮に固定したとき、溶媒のあるなしに伴う全系の自由エネルギー差として定義される。

$$G = G_{\text{溶媒}} - G_{\text{真空}}$$

$G_{\text{溶媒}}$ 、 $G_{\text{真空}}$ の分子内ポテンシャル項は差分である G で打ち消される。さらに、溶質の影響を受けるのは最近傍の溶媒のみであると考え、溶媒自由エネルギーは最近傍の溶媒分子の数に、ひいてはその溶媒分子の中心によって描かれた露出表面積（Accessible Surface Area : ASA）に比例すると考える。溶質分子の熱力学量と露出表面積との間の経験的な比例性を使って、溶媒和自由エネルギー G を以下のように導入する。

$$\Delta G = \sum_i g_i A_i$$

ここでは A_i は i 種の原子団と水の接触する露出表面積で、分子構造をもとに計算される。 g_i は対応する比例定数で真空中から溶媒へ原子団を移相するときのエネルギー利得である。 g_i は低分子の熱力学実験から求められ、パラメータとして与えられる。



NOTE

大井ら[11]は、原子団や基ごとの水和エネルギーパラメータ g を求めている。彼らはタンパク質分子の配座エネルギー計算と同じ枠組みのもと水和自由エネルギー G を求めるため、小さな有機分子ガスがガス状態から水溶液へ移行する際の自由エネルギーのデータを用いた。 G の値としてエタン、プロパンなどの実験値を用いて、 g を最小二乗法で求めている。

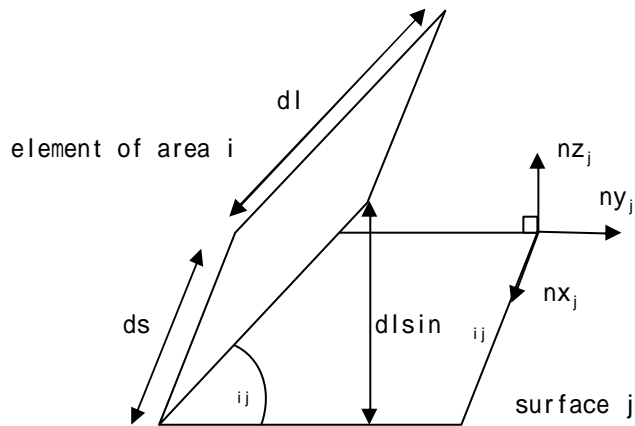
Accessible Surface Area における、エネルギーと力の計算：

微小平面 i ($dl \times ds$) において、平面 j との接触パラメータを θ_{ij} とする。平面 j のベクトル nz_j 、平面 j の tangent ベクトル nx_j, ny_j 、微小平面 i と平面 j とのなす角を θ_{ij} とすると、エネルギーの変化は、 nz_j で与えられる。よって、 ds における力は、

$$df_{ij} = \frac{\Delta S_i dl ds}{dl \sin \theta_{ij}} nz_j$$

原子 i と原子 j についての力は、境界 c_{ij} についての積分で表される。

$$F_{SAsi} = \sum_j \oint_{c_{ij}} \frac{1}{|nz_i \times nz_j|} (nz_j \Delta S_i - nz_i \Delta S_j) ds$$



T.J.Richmond[12]は、Accessible Surface Area が重なる系において、弧の長さとお交差点における角度を求め、エネルギーと力の計算を行う方法を体系化している。

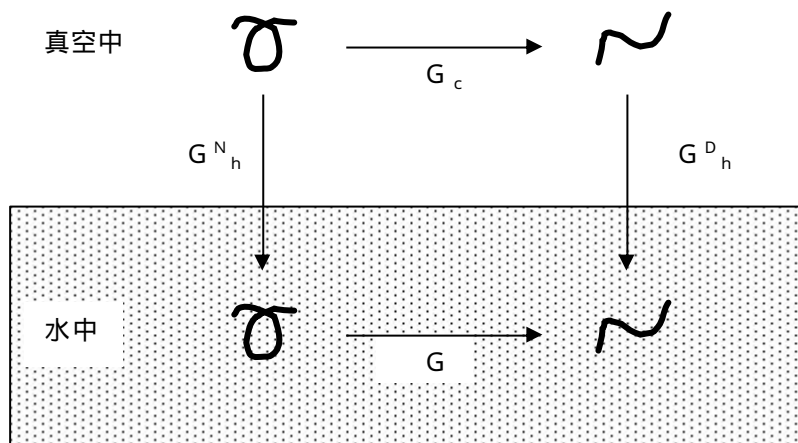
- (1) 重なりのある球 (Accessible Surface Area の球) のリストを準備する。
- (2) 重なりのパターンを解析する。
- (3) 弧の半径、交差点の角度を求め、エネルギーと力の計算を行う。

NOTE

タンパク質の天然状態 (N 状態) と変性状態 (D 状態) のギブスの自由エネルギーを G^N 、 G^D とすると、水溶液中の変性のギブスの自由エネルギー G は次のように表される。

$$G = G^D - G^N$$

ここで、タンパク質が水溶液中で変性する場合を熱力学サイクル (下図) で考える。



このとき G は、真空中での変性のギブスの自由エネルギー G_c と、変性のギブスの水和自由エネルギー G_h に分けられる。ここで、 G^D_h と G^N_h は D 状態と N 状態のギブスの水和自由エネルギーである。

$$G = G_c + G_h$$

$$G_h = G^D_h - G^N_h$$

N 状態と D 状態の各原子団の ASA はそれぞれの構造から計算できるので、上式から G を求めることができる。

3.2 Generalized Born accessible surface area

溶媒をあらわに取り扱わないモデルでは、溶媒は、実際の溶媒の平均的誘電率を持ち、van der Waals 半径で描かれる表面を持つ溶質を取り囲む連続体として扱われる。このような連続体近似を行うモデルの中で、Generalized Born (GB) Surface Area (SA) モデルは良く利用されている。

Still [14]らは GB/SA モデルで、溶媒和自由エネルギー G_{sol} を、溶媒 - 溶媒間の空洞部分のエネルギーの項 (G_{cav})、溶質 - 溶媒間の van der Waals エネルギーの項 (G_{vdW}) および溶質 - 溶媒間の静電極性エネルギーの項 (G_{pol}) の合計として与えた。

$$G_{sol} = G_{cav} + G_{vdW} + G_{pol} = \sum s_k A_k + G_{pol} \quad --(1)$$

(s_k : Atomic solvation parameter)

GB/SA モデルでは、 $G_{cav} + G_{vdW}$ の部分を Accessible Surface Area (ASA) の一次関数で計算する。ここで A_k は溶媒の露出表面積、 s_k は経験的に決めた溶媒和パラメータ (Atomic solvation parameter) である。

(1) Generalized Born の方程式

GB の方程式は、次のように定義される。

$$G_{pol} = -\frac{1}{2} \left(1 - \frac{1}{\epsilon_w} \right) \sum_i \sum_j \frac{q_i q_j}{f_{GB}} \quad --(2)$$

$$f_{GB} = \sqrt{r_{ij}^2 + R_i R_j \exp \left(-\frac{r_{ij}^2}{4 R_i R_j} \right)} \quad --(3)$$

- q_i 、 q_j : 原子 i と原子 j の電荷
- ϵ_w : 溶媒の誘電率
- r_{ij} : 原子 i 、原子 j 間の原子間距離
- R_i 、 R_j : 原子 i および原子 j の Born 半径

電荷の 2 つのはすべての原子 (i と j) のペアに対して行う。

(2) Born 半径

単純なイオンの場合、電荷 q で半径 r のときの静電自由エネルギーは解析的に求めることができる。いわゆる Born モデルである。

$$G = -\frac{1}{2} \left(1 - \frac{1}{\epsilon_w} \right) \frac{q^2}{r} \quad \text{--(4)}$$

ϵ_w : 溶媒の誘電率

半径 r は、イオンの Born 半径とよばれる。このときのエネルギーは、真空中では自己エネルギー、溶媒中ではイオンの Born エネルギーとよばれる。

分子中の原子の場合、それぞれの溶媒との相互作用を考慮する必要があるが、それらの溶媒部分は近似により、分子内の他の原子に置き換えて取り扱う。GB モデルの基本的な考えは、Born エネルギーを求めるためには、原子 i に対する有効半径 R_i を定義してやれば良いということである。

ただし、一般的には R_i は r_i (原子 i の固有半径) だけでなく、他のすべての原子の半径および相対位置にも依存する。そこで、Born 半径を少ない計算コストで求める方法として、エネルギー密度に基づき、クーロン場近似を使用した近似計算方法が開発されている。

$$R_i^{-1} = r_i^{-1} - \frac{1}{4\pi} \int_{\text{solute}} \frac{dV}{r^4} \quad \text{--(5)}$$

($r > r_i$)

ここでは、分子内部の領域を計算する (原子 i の中心から半径 r_i の領域を除く)、原子の Born 半径 R の計算は、式(5)中の積分を評価することと等価である。

NOTE

理想的には、原子 i の位置に置かれた点電荷 q_i および分子の全原子の固有半径によって決定された境界領域から Poisson 方程式を解き、その電荷 i の自己エネルギーがそれらの Born エネルギーと等しくなるように Born 半径を設定されるべきである。しかしこれでは、直接 Poisson 方程式の数値解を求めることに比べ、近似する効果が得られない。

$$\nabla e(r) \nabla f(r) = -4\pi r(r)$$

(3) スケールファクタを用いる方法 (Hawkins et al.)

この方法の基本的なアイデアは、各原子の寄与の合計によって式(5)の中の積分を近似することである。ここで、位置 r_{ij} の原子 i に球体があるとし、分子が半径 r_j のオーバーラップしない球体の集合から成るとした場合、式(5)は球体の体積の積分の合計として書くことができる。

$$R_i^{-1} = r_i^{-1} - \frac{1}{4\pi} \sum_j \int_{\text{sphere-}j} \frac{dV}{r^4} \quad \text{--(6)}$$

球体に関する積分は、解析的に計算することができる。

$$R_i^{-1} = r_i^{-1} - \sum_j \left[\frac{r_j}{2(r_{ij}^2 - r_j^2)} - \frac{1}{4r_{ij}} \log \left(\frac{r_{ij} - r_j}{r_{ij} + r_j} \right) \right] \quad \text{--(7)}$$

ここで、式(7)の二体間の合計部分では、近隣の原子が互いにオーバーラップするので、溶質の領域として重複してカウントされる。Hawkins は、このオーバーラップ部分を相殺するために、近隣の R_i の値をスケールリングによって経験的に修正する方法を提案している。

$$R_i^{-1} = r_i^{-1} - \sum_j H(r_{ij}, S_j r_j) \quad \text{--(8)}$$

ここで、 S_j はスケールリングファクタであり、実験値などで適合させる。

NOTE

Still が Born 半径を三重積分で求めていたのに対し、二体間計算で一方の van der Waals 半径にスケールファクタをかけ、二体の球体がオーバーラップするのを避けた。これにより、Born 半径を単積分で計算するようになった。

4. Multicanonical 法の拡張

4.1 Force-Biased Multicanonical Sampling 法

Multicanonical アンサンブルは、サンプリングにバイアスをかける umbrella sampling 法の一種といえるが、反応座標がエネルギーそのものであるため、バイアスのかけ方の定義が明確で汎用性に富む利点がある。ここでは温度は一定で、ポテンシャルの方をスケールする、ポテンシャルスケールリングの方法が用いられている。

Force-Biased Multicanonical Sampling 法では、エネルギー確率分布関数 P_{mc} のヒスト

グラムから直接 $\frac{\partial \ln P_{mc}^i(E)}{\partial E}$ を求めるようにしており、ここから重み関数 のエネルギー

微分 (力のバイアス) の更新値である $\tilde{b}^{i+1}(E)$ が求められる。これにより、重み関数の精密化を自動化することができ、より容易に Multicanonical Sampling を実現することができる。

Multicanonical MD での Force-Biased Multicanonical Sampling 法のスキーム：

(1) Multicanonical MD の重み関数 のエネルギー微分 を、反復解法で収束させる。

$$\mathbf{a}^{i+1}(E) = \mathbf{a}^i(E) + \frac{1}{\mathbf{b}_0} \ln P_{mc}^i(E)$$

$$\tilde{\mathbf{b}}^{i+1}(E) = \tilde{\mathbf{b}}^i(E) + \Gamma^i(E)$$

$$\tilde{\mathbf{b}}^i(E) = \frac{\partial \mathbf{a}^i}{\partial E}, \quad \Gamma^i(E) = k_B T_0 \frac{\partial \ln P_{mc}^i(E)}{\partial E}$$

MD シミュレーションでの $P(E)$ の閾値 (P_{min}) により、 $\tilde{\mathbf{b}}^{i+1}(E)$ の更新を制御する。

- ・ 閾値を超えた場合 ($P(E) > P_{min}$) : $\tilde{\mathbf{b}}^{i+1}(E) = \tilde{\mathbf{b}}^i(E) + \Gamma^i(E)$
- ・ 閾値を超えない場合 : $\tilde{\mathbf{b}}^{i+1}(E) = \tilde{\mathbf{b}}^i(E)$ (更新しない)

(2) $\tilde{\mathbf{b}}^{i+1}(E)$ を上式で更新し、(短時間の) Multicanonical MD を繰り返していく。

(3) 異なる T によるシミュレーション結果を multiple-histogram method で統合する。

(4) Reweighting により、任意の温度 T での Canonical Sampling を再構成することができる。

4.2 Simulated Tempering 法

Simulated Tempering 法は、温度に対して一様な分布を得るための方法である。ここで、温度に対して一様な分布を実現するような重みを $\exp\left(-\frac{E}{k_B T} - g(T)\right)$ とすると、 $g(T)$ は Multicanonical 法のとおり同様に反復解法によって精密化することができる。最終的に得られた温度分布関数のヒストグラムから、任意の温度のデータを集計することによって Canonical Sampling を再構成することができる。

Simulated Tempering 法のスキーム：

(1) Simulated Tempering の重み関数 $g(T)$ を、反復解法で収束させる。

$$g^{i+1}(T) = g^i(T) + \ln P_{ST}(T)$$

上式を、温度 T での Canonical アンサンブルの平均エネルギー U を用いて表す。

$$U_F^{i+1}(T) = U_F^i(T) + \Pi(T)$$

$$U'(T) = \frac{U(T)}{k_B T^2}, \quad \Pi(T) = \frac{\partial \ln P_{ST}^i(T)}{\partial T}$$

$g(T)$ が与えられたとして、重み $\exp\left(-\frac{E}{k_B T} - g(T)\right)$ でのサンプリングを実際に MD シ

ミュレーションで行うには、以下の手続きをふめばよいことになる。

ある温度で通常の MD (力 = $-E/r$) によって様々な構造 (状態) をサンプリングする。

ある構造 (状態) のときに、系の温度を変える。

その際の温度 T から T' への遷移確率は、次のように与えられる。

$$a(T \rightarrow T') = \min \left[1, \exp \left\{ \left(\frac{1}{k_B T} - \frac{1}{k_B T'} \right) E + g(T) - g(T') \right\} \right]$$

(2) 重み g を上式で更新することにより、温度の確率分布関数 $P_{ST}(T)$ を平らにしていく。

(3) 最終的に、得られた温度のヒストグラムから、任意の温度のデータを集計することによって Canonical Sampling を再構成することができる。

5. RATTLE 法

共有結合の結合伸縮の周期は、特に水素原子が関連しているとき短くなる。この場合、正確に運動方程式を解くためには、タイムステップ Δt を小さくとる必要がある。しかし、このような結合伸縮は、全体の運動から見た場合はさほど重要ではない。そこで、共有結合の原子間距離を一定の長さに拘束して運動方程式を解く方法が利用される。これにより、周期の小さい結合伸縮がなくなるため Δt を大きくすることができる。SHAKE 法は verlet 法と結びついた数値解法として用いられるが、位置に関する拘束だけを問題にしていた。

原子間の距離 d に対する拘束条件と対応する運動方程式は以下で与えられる。

$$g = (\mathbf{r}_i - \mathbf{r}_j)^2 - d_{ij}^2 = 0 \quad --(1)$$

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\mathbf{r}}_i} \right) - \frac{\partial L}{\partial \mathbf{r}_i} = \mathbf{1} \frac{\partial g}{\partial \mathbf{r}_i} \quad --(2)$$

RATTLE 法は位置だけでなく、SHAKE 法であらわに扱っていなかった速度に関する拘束条件も満足させる。RATTLE 法は velocity verlet 法と結びついた数値解法として用いられる。velocity verlet 法では以下の差分式で運動方程式を解く。ここで、 \mathbf{r}_i 、 \mathbf{v}_i 、 \mathbf{f}_i 、 m_i は、原子 i の座標、速度、力、質量である。

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

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \cdot \mathbf{v}_i \left(t + \frac{\Delta t}{2} \right) \quad --(4)$$

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i \left(t + \frac{\Delta t}{2} \right) + \frac{\Delta t}{2} \cdot \frac{\mathbf{f}_i(t + \Delta t)}{m_i} \quad --(5)$$

velocity verlet 法では位置と速度の両方が系の時間発展を記述するため、式(1)の位置に関する拘束条件に加え、式(1)の微分で表される式(6)の拘束条件が、式(3)での \mathbf{v} に関して満たすようにする。

$$\begin{aligned} \frac{dg}{dt} &= 2(\dot{\mathbf{r}}_i - \dot{\mathbf{r}}_j) \cdot (\mathbf{r}_i - \mathbf{r}_j) = 0 \\ (\mathbf{v}_i - \mathbf{v}_j) \cdot (\mathbf{r}_i - \mathbf{r}_j) &= 0 \end{aligned} \quad --(6)$$

RATTLE 法を組み合わせた時間積分 (velocity verlet 法) 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 rattle_coordinate : 位置の拘束条件による補正処理

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

  call rattle_velocity : 速度の拘束条件による補正処理
enddo

```

(1) 位置の拘束条件による補正

式(1), (2), (3), (4)より、新しい位置は、

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \cdot \mathbf{v}_i(t) + \frac{\Delta t^2}{2} \cdot \frac{\mathbf{f}_i(t)}{m_i} + \frac{\Delta t^2}{2} \cdot \frac{2\mathbf{I} \{ \mathbf{r}_i(t) - \mathbf{r}_j(t) \}}{m_i} \quad --(7)$$

系には拘束による力だけが単独に働いていると考え、その変位は式(7)の第4項から次のようにすることができる。jについては変位の方が逆になる。

$$\Delta \mathbf{r}_i(t + \Delta t) = C_{ij} \frac{\mathbf{r}_i(t) - \mathbf{r}_j(t)}{m_i} \quad --(8)$$

式(1)を満たすようにCを決めることによって、補正後の位置を求めることができる。

位置の補正に対する速度の補正は、式(8)から次のようになる。

$$\mathbf{v}'_i \left(t + \frac{\Delta t}{2} \right) = \mathbf{v}_i \left(t + \frac{\Delta t}{2} \right) + C_{ij} \frac{\mathbf{r}_i(t) - \mathbf{r}_j(t)}{m_i} \cdot \frac{1}{\Delta t} \quad --(9)$$

(2) 速度の拘束条件による補正

式(1),(2),(5)より、新しい速度は、

$$v_i(t + \Delta t) = v_i\left(t + \frac{\Delta t}{2}\right) + \frac{\Delta t}{2} \cdot \frac{f_i(t + \Delta t)}{m_i} + \frac{\Delta t}{2} \cdot \frac{2\mathbf{I} \{r_i(t + \Delta t) - r_j(t + \Delta t)\}}{m_i} \quad --(10)$$

位置の場合と同様に、その変位は式(10)の第3項から次のようにすることができる。

$$\Delta v_i(t + \Delta t) = V_{ij} \frac{r_i(t + \Delta t) - r_j(t + \Delta t)}{m_i}$$

式(6)を満たすようにVを決めることによって、補正後の速度を求めることができる。

補正処理の具体的な方法は SHAKE 法と同様の反復法を用いて、全ての拘束条件が満たされる位置と速度を求める。

6. Hybrid MD/HD 法

- Hybrid method of Molecular Dynamics and Harmonic Dynamics -

リガンドがタンパク質にドッキングする際、タンパク質の大きな構造変化を伴うことがしばしばある。しかし、このようなドッキング過程を記述する際にタンパク質全体をそのまま用いるのは計算時間が膨大にかかるため、あまり得策ではない。そこで開発されたのがこの手法である。この手法は、タンパク質全体がフレキシブルであるときに結合ポケットを含む局所領域がとっていた運動を、ドッキングの際に切り出したその局所領域に反映させるというものである。具体的には、タンパク質全体のシミュレーションを行い、そのときの C 原子(あるいは主鎖原子)の運動をトラジェクトリの主成分分析により取り出す。それを基準振動の重ね合わせで近似して、ドッキングの際に切り出した局所領域の C 原子(あるいは主鎖原子)の運動に取り込む。

Hybrid MD/HD 法は、調和振動子から成る部分 A (タンパク質全体の C 原子あるいは主鎖原子) あらわに原子を考慮する部分 B (切り出したタンパク質局所領域に属する原子のうち部分 A 以外の原子、溶媒分子とリガンド) の 2 つからなり、全系のハミルトニアンは次のように書ける。

$$H = E_{AA} + E_{A'B} + E_{BB} + T_A + T_B \quad --(1)$$

それぞれの意味は以下の通り。

- ・ A' : 部分 A のうち、切り出したタンパク質局所領域に含まれる部分 (A' A)
- ・ E_{AA} : 部分 A の調和振動子ポテンシャル【振動子】
- ・ E_{A'B} : 部分 A' に属する原子と部分 B に属する原子との 1-2, 1-3, 1-4, 1-5 相互作用ポテンシャル【通常】
- ・ E_{BB} : 部分 B に属する原子どうしの相互作用【通常】
- ・ T_A : 部分 A の調和振動子の運動エネルギー
- ・ T_B : 部分 B の原子の運動エネルギー

表 1 各部分との関係

	A (A')	B
A	振動子	通常 (A' - B)
B	通常 (A' - B)	通常

運動方程式は、調和振動子の部分 A に対する基準振動空間での運動方程式と、部分 B に対するデカルト空間での運動方程式の 2 つからなる。部分 B に属する原子の運動方程式は、部分 B の原子座標を X として、

$$\dot{p} = -\frac{\partial H}{\partial X}, \quad \dot{X} = \frac{\partial H}{\partial p} \quad --(2)$$

から、単に部分 A'、部分 B 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。

$$\begin{aligned} \dot{p} &= -\frac{\partial E_{BB}}{\partial X} - \frac{\partial E_{BA'}}{\partial X} \\ &= f_{B-B} + f_{B-A'} \end{aligned} \quad --(3)$$

問題は、部分 B に属する原子との相互作用を受けながら、特定の限られた基準振動モードのみで振動する部分 A の運動方程式である。デカルト空間での座標 x 、速度 v の代わりに、基準振動空間での座標（振動の変位） c と速度 \dot{c} を使ってハミルトニアンを記述し、 c についての運動方程式を解くことにより、部分 A の座標と速度を求める。

デカルト空間での部分 A の原子、すなわち、 C 原子（あるいは主鎖原子）の座標 x 、速度 v は、基準振動空間での座標 c 、速度 \dot{c} から、以下のような変換によって得られるものとする。

$$x_i = x_i^0 + \sum_I^b u_{i,I} c(t)_I \quad --(4)$$

$$v_i = \frac{dx_i}{dt} = \sum_I^b u_{i,I} \dot{c}(t)_I \quad --(5)$$

$i: 1, \dots, 3N$
 $N: \text{部分 A に属する原子の数}$

x_i^0 は振動中心座標（デカルト空間）である。また、基準振動の順番を I とし、各基準振動 I の原子座標の変化を $c(t)_I$ とする。 \sum_I^b の和は、取ってくる基準振動モードの数 b であり、ドッキングの際に切り出す結合ポケットを含む領域の運動を記述するのに十分なモードを取ってくるものとする。 $(3N \times b)$ の変換行列 $U (= u_{i,I})$ は、タンパク質全体のシミュレーションでのトラジェクトリから得られる。分散行列の対角化によって求められる部分 A' と部分 B の原子間相互作用 $E_{A'B}$ を $E(c, X)$ とし、 X は部分 B の原子座標とする。 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、

$$E_{pot} = E(c, X) + \sum_I^b \frac{k_I}{2} c(t)_I^2 \quad --(6)$$

c に関するハミルトニアン¹の運動エネルギー部分は、

$$p_i = m_i \frac{dx_i}{dt} = \sum_I^b m_i u_{i,I} \dot{c}(t)_I \quad --(7)$$

$$E_k = \sum_i^{3N} \frac{p_i^2}{2m_i} = \sum_i^{3N} \frac{1}{2m_i} \left(\sum_I^b m_i u_{i,I} \dot{c}(t)_I \right)^2 \quad --(8)$$

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。

$$\begin{aligned} p_I &= \frac{\partial L}{\partial \dot{c}_I} = \sum_i^{3N} \frac{2}{2m_i} \left(\sum_h^b m_i u_{i,h} \dot{c}(t)_h \right) m_i u_{i,I} \\ &= \sum_i^{3N} \left(\sum_h^b m_i u_{i,h} \dot{c}(t)_h \right) u_{i,I} \\ &= \sum_h^b \sum_i^{3N} m_i u_{i,I} u_{i,h} \dot{c}(t)_h \end{aligned} \quad --(9)$$

この運動量 p をもとに、ハミルトンの方程式、

$$\dot{p} = -\frac{\partial H}{\partial q}, \quad \dot{q} = \frac{\partial H}{\partial p} \quad --(10)$$

を解く。

$$\dot{p}_I = -\frac{\partial H}{\partial c_I} \quad --(11)$$

左辺は、

$$\dot{p}_I = \sum_h^b \sum_i^{3N} m_i u_{i,I} u_{i,h} \ddot{c}(t)_h \quad --(12)$$

右辺は、

$$-\frac{\partial E_{pot}}{\partial c_I} = -k_I c(t)_I - \frac{\partial E(\mathbf{c}, \mathbf{X})}{\partial c_I} \quad --(13)$$

右辺第二項を通常のカから計算できるようにする。式(4)を用いて、

$$\frac{\partial E(\mathbf{c}, \mathbf{X})}{\partial c_l} = \sum_i^{3N'} \frac{\partial E(\mathbf{x}, \mathbf{X})}{\partial x_i} \frac{\partial x_i}{\partial c_l} = - \sum_i^{3N'} f_i \frac{\partial \left(x_i^0 + \sum_x^b u_{i,x} c(t)_x \right)}{\partial c_l} = - \sum_i^{3N'} f_i u_{i,l} \quad --(14)$$

N' : 部分 A' に属する原子の数

結局、 \mathbf{c} の運動方程式は、両辺を合せて、

$$\sum_h^b \sum_i^{3N} m_i u_{i,l} u_{i,h} \ddot{c}(t)_h = -k_l c(t)_l + \sum_i^{3N'} f_i u_{i,l} \quad --(15)$$

このままでは \mathbf{c} について解きにくいので、 $\sum_i^{3N} m_i u_{i,l} u_{i,h}$ を $(b \times b)$ 行列 \mathbf{D} ($= d_{l,h}$)

の (l, h) 成分とみなし、その逆行列 \mathbf{W} ($= w_{x,l}$) を定義する。

$$\sum_l^b w_{x,l} d_{l,h} = \mathbf{d}_{x,h} \quad --(16)$$

式(15)の左辺に \mathbf{W} をかけ、 \mathbf{c} について和をとると、

$$\begin{aligned} \sum_l^b w_{x,l} \dot{p}_l &= \sum_l^b w_{x,l} \sum_h^b \sum_i^{3N} m_i u_{i,l} u_{i,h} \ddot{c}(t)_h \\ &= \sum_l^b \sum_h^b w_{x,l} d_{l,h} \ddot{c}(t)_h \\ &= \sum_h^b \ddot{c}(t)_h \mathbf{d}_{x,h} \\ &= \ddot{c}(t)_x \end{aligned} \quad --(17)$$

式(15)の右辺に行列 \mathbf{W} をかけると、 \mathbf{c} に関する時間発展の微分方程式が得られる。

$$\ddot{c}(t)_x = \sum_l^b w_{x,l} \left(-k_l c(t)_l + \sum_i^{3N'} f_i u_{i,l} \right) \quad --(18)$$

マルチカノニカルの場合、 $E = E_{AA} + E_{A'B} + E_{BB}$ とすると、ポテンシャルが $E_{mc}(E)$ であるときの温度一定の計算をすることに対応する。

部分Bに属する原子の運動方程式は、

$$\begin{aligned}
 \dot{\boldsymbol{p}} &= -\frac{\partial E_{mc}(E)}{\partial \boldsymbol{X}} \\
 &= -\frac{dE_{mc}(E)}{dE} \frac{\partial E}{\partial \boldsymbol{X}} \\
 &= -\frac{dE_{mc}(E)}{dE} \left(\frac{\partial E_{BB}}{\partial \boldsymbol{X}} + \frac{\partial E_{BA'}}{\partial \boldsymbol{X}} \right) \\
 &= \frac{dE_{mc}(E)}{dE} (\boldsymbol{f}_{B-B} + \boldsymbol{f}_{B-A'}) \quad \text{--(19)}
 \end{aligned}$$

部分Aに属する原子の運動方程式は、

$$\begin{aligned}
 \dot{p}_1 &= -\frac{\partial E_{mc}(E)}{\partial c_1} \\
 &= -\frac{dE_{mc}(E)}{dE} \frac{\partial E}{\partial c_1} \\
 &= -\frac{dE_{mc}(E)}{dE} \frac{\partial E_{pot}}{\partial c_1} \quad \text{--(20)}
 \end{aligned}$$

上式を式(11) ~ (18)と同様に解くと、最終的に、

$$\ddot{c}(t)_x = \frac{dE_{mc}(E)}{dE} \sum_I^b w_{x,1} \left(-k_I c(t)_I + \sum_i^{3N} f_i u_{i,1} \right) \quad \text{--(21)}$$

が得られる。

7. 拘束ポテンシャル

7.1 CAP 拘束の対象の拡張

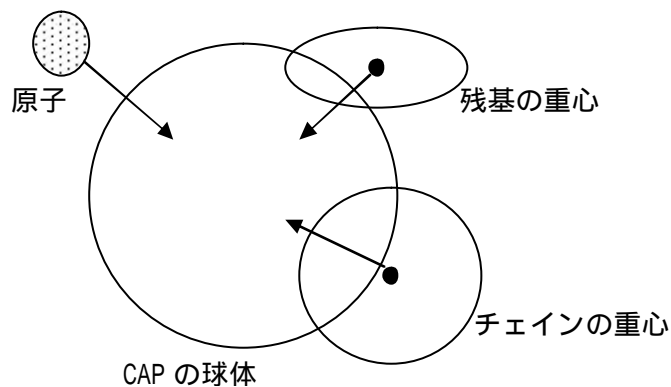
ある原子をある範囲にとどめておくために、CAP 拘束を使用することができる。CAP 拘束は、任意の位置に球体を定義し、その球体の外にある原子に対して球体の内側に帰るように軽いバネの力 (CAP 拘束力) がかかるようにする。

ここでは、「チェーン」、「残基」、「原子」の単位で CAP 拘束力の計算対象を指定できるようにする。また、CAP の球体の内外判定をそれぞれ次のように定義する。

チェーン： チェーンの重心

残基： 残基の重心

原子： 原子の座標



残基、チェーン (分子) は、重心位置で CAP の内外を判定する。

CAP の外になる場合は、指定単位内 (チェーン、残基、原子) の原子に力がかかる。

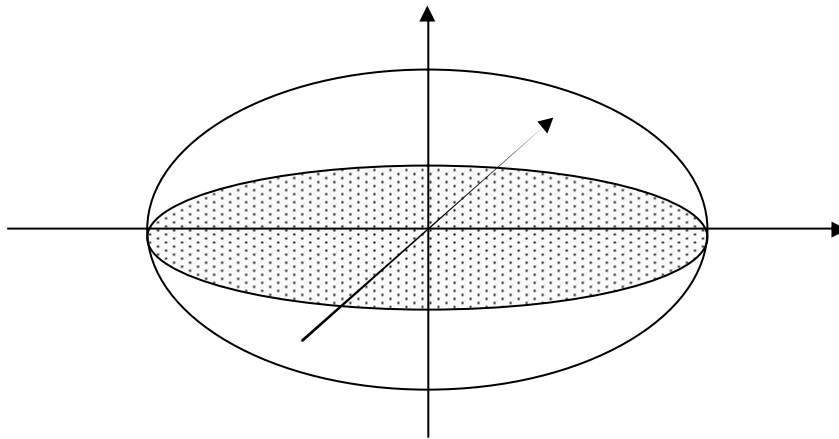
7.2 楕円体 CAP

CAP 拘束の形状を球体だけではなく、楕円体が使用できるように拡張する。

(1) 「中心座標 + 各軸方向への距離」指定

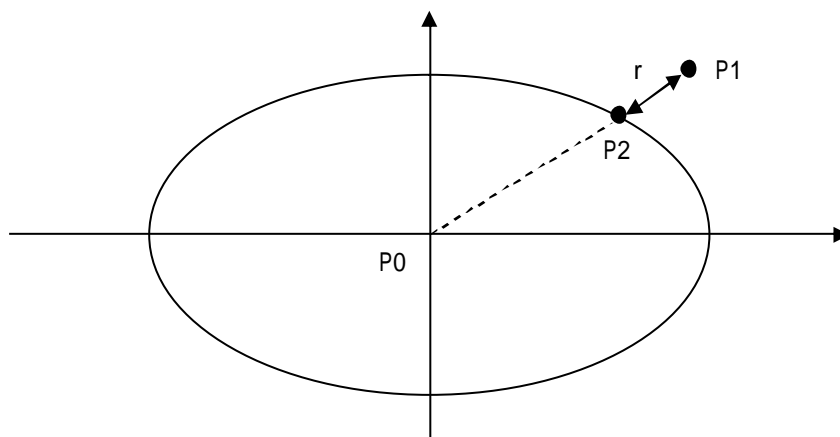
楕円の方程式で与える楕円体を指定する。各軸方向で大きさを変えることができる。

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$



楕円面から出た対象の位置 P_1 と、楕円面中心 P_0 とを結ぶ直線が、楕円面と交わる点を P_2 とする。拘束力は、点 P_1 、 P_2 間距離 r にかかる力が楕円中心方向 (r 方向) にはらたく。

$$E = \frac{1}{2} h r^2 \quad (h : \text{力の係数})$$



残基、チェーン (分子) は、重心位置で CAP の内外を判定する。

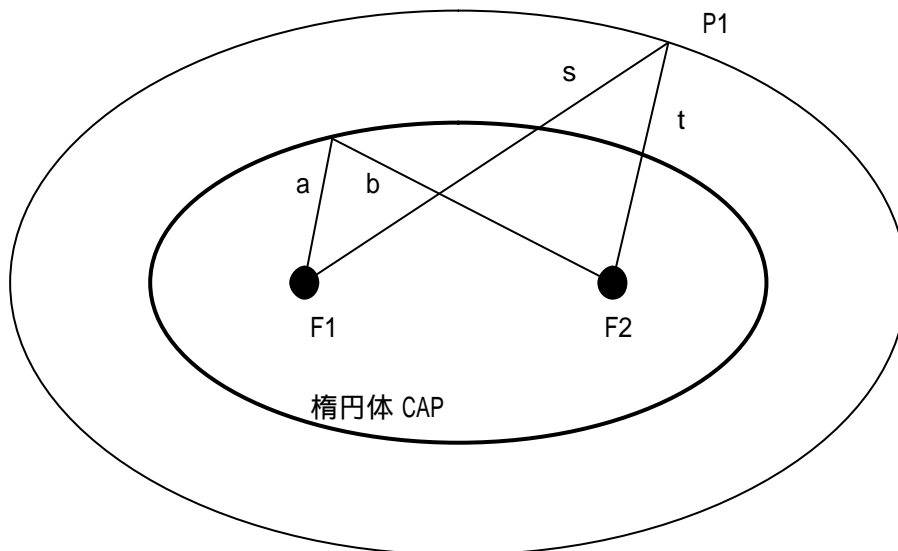
(2) 「焦点座標 + 距離」指定

「2つの焦点と、そこからの距離の和が一定」という条件を満たす楕円体を指定する。
焦点 F_1 との距離を a 、 F_2 との距離を b とすると、常に以下を満たす。

$$a + b = 2l$$

楕円面から出た対象の位置 P_1 と、焦点 F_1 との距離を s 、 F_2 との距離を t とする。等ポテンシャル面が楕円面になるとして、次の拘束力がはらたく。

$$E = \frac{1}{2} h \left(\frac{s+t-2l}{2} \right)^2 \quad (h : \text{力の係数})$$



残基、チェーン（分子）は、重心位置で CAP の内外を判定する。

7.3 範囲付き位置拘束

任意の点Pを基点に、距離 D_L から距離 D_U の範囲に指定した原子を拘束するような拘束ポテンシャルを定義できるようにする。拘束ポテンシャルは、点Pと指定した原子との距離 d により、次のように定義する。

(1) $d < D_L$ の場合 (下図、点A)

$$E = \frac{h}{k_L} k_B T_C (d^2 - D_L^2)^2$$

(2) $D_L \leq d \leq D_U$ の場合 (下図、点B)

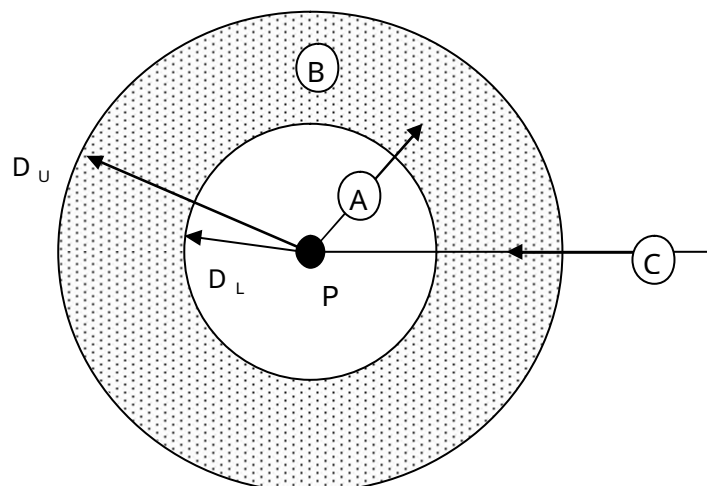
$$E = 0$$

(3) $d > D_U$ の場合 (下図、点C)

$$E = \frac{h}{k_U} k_B T_C (d^2 - D_U^2)^2$$

ここで、 k_B はボルツマン定数、 T_C は拘束温度、 h は力の係数。 k_L および k_U はそれぞれの範囲にかける係数 (デフォルト 1.0) である。

また、拘束の対象は原子指定だけではなく、「チェーン」、「残基」、「原子」の単位で指定できるようにする。範囲の内外判定は、拘束対象の重心位置で行う。



残基、チェーン (分子) は、重心位置で範囲の内外を判定する。

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.
- [9] J. G. Kim, Y. Fukunishi, H. Nakamura, Dynamical origin of uniform sampling in multicanonical ensemble, *Physical Review E* 67, 011105 (2003).

Multicanonical WHAM 法 :

- [10] 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.

GB / ASA :

- [11] Ooi, T., Oobatake, M., Nemethy, G. & Scheraga, H. A. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proc Natl Acad Sci U S A 84, 3086-90. (1987).
- [12] Richmond, T. J. Solvent accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect. J Mol Biol 178, 63-89. (1984).
- [13] Kinjo, A. R., Kidera, A., Nakamura, H. & Nishikawa, K. Physicochemical evaluation of protein folds predicted by threading. Eur Biophys J 30, 1-10. (2001).
- [14] Still, W. C., Tempczyk, A., Hawley, R., C. & Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. J. Am. Chem. Soc. 112, 6127 (1990).
- [15] Hawkins, D., Gregory., Cramer, J., Christopher. & Truhlar, G., Donald. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. J Phys Chem 100, 19824-39 (1996).
- [16] Hawkins, D., Gregory., Cramer, J., Christopher. & Truhlar, G., Donald. Pairwise solute descreening of solute charges from a dielectric medium. Chemical Physics Letters 246, 122-129 (1995).
- [17] Onufriev, A., Bashford, D. & Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. J Phys Chem. B 104, 3712-3720 (2000).
- [18] Bashford, D. & Case, D. A. Generalized born models of macromolecular solvation effects. Annu Rev Phys Chem 51, 129-52. (2000).
- [19] Schaefer, M. & Karplus, M. A Comprehensive Analytical Treatment of Continuum Electrostatics. J Phys Chem 100, 1578-1599 (1996).