

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)

第 4 版

目次 -

1. はじめに	2
2. 新 MD Simulation System の概要	3
2.1 目的	3
2.2 システム概要	4
3. エネルギー最小化機能	6
3.1 概要	6
3.2 機能構成	7
3.3 最急降下法	8
3.4 拘束条件つき最急降下法	9
3.5 共役勾配法	10
4. MD 計算結果分析機能	13
4.1 概要	13
4.2 機能構成	14
4.3 構造情報分析	15
4.4 計算値分析	16
参考文献	19

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「エネルギー最小化機能」および「MD 計算結果分析機能」の方式について説明する。

また、本ソフトウェアの背景と目的、研究開発の内容を以下に示す。

背景と目的

ゲノム情報は直接の産物である蛋白質として具現することにより初めて生体内で目的とする機能を発現することができる。現在、急速に進行しつつあるゲノム解析は、結果として膨大な量の蛋白質の一次構造データを生産し続けており、ゲノム情報を人間社会の発展、健康・福祉に役立てるためには、蛋白質がいかなるものであるかを明らかにすることが 21 世紀の重大な研究課題である。そのためには、まず蛋白質がどのような機構で機能を発揮しているかを明らかにすることが必要であり、蛋白質及びそれらと相互作用する生体高分子(蛋白質、核酸、脂質、多糖等)の立体構造を原子レベルで解析し、それに基づいて分子間相互作用を詳細に解明することが有効かつ重要である。また、今後得られる構造や相互作用に関する膨大な情報を効率よく解析し、蛋白質の構造解析を加速化するためにはコンピューターによる構造情報解析技術の開発が不可欠である。本研究開発は、膜蛋白質等及びその複合体について、その立体構造や相互作用部位を原子レベルで測定・解析する構造解析装置(X線、電子線、核磁気共鳴(NMR))を開発し、膜蛋白質等の機能や分子間相互作用を効率的かつ高精度に解析する技術を開発する。また、構造情報解析技術の確立のため、高精度モデリング技術やシミュレーション技術の開発を併せて行う。

研究開発の内容

- ・電子線及びX線による蛋白質の構造と分子機構解析技術の開発
- ・核磁気共鳴法(NMR)による分子間相互作用解析技術の開発
- ・データベースとシミュレーション計算を活用した構造情報解析技術の開発

2. 新 MD Simulation System の概要

2.1 目的

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持ち、大きな蛋白質複合体や膜蛋白質の取り扱いにも適した相互作用の効率的な計算や境界条件の設定が可能である、汎用的でかつ高効率なシミュレーション計算システムを実現する。

Multicanonical method などの高効率構造探索手法をシミュレーション計算システムとして実現し、蛋白質複合体や膜蛋白質を対象として、境界条件を工夫したシミュレーション計算が行えるようにする。

特に、膜蛋白質の系を立方体セルではなく斜方体セルで扱い、静電相互作用を正しく取り込んだ脂質二重膜のシミュレーションを精度良く行える機能を実現する。

さらに、自由エネルギー地形は力場パラメータに強く依存するが、その問題を回避するために、最近中村らによって提案された *ab initio* 分子軌道計算による力場補正の手法を援用し、より精度の高い結果が得られるようにする。

上記の方法を、独立性の高い多くのシミュレーション計算の集合体として、大規模なクラスター型並列計算機上でも実現できるようにする。

Multicanonical アンサンブルを得るためには、iterative な計算かレプリカ交換法を用いるかしてエネルギー確率分布を平坦にする必要があるが、現状ではこの過程に膨大な時間がかかり経験も必要である。この計算過程を効率化し自動化することができれば、より迅速にシミュレーション計算を進めることが可能となる。現状では研究者の経験が必要であるため、未だに大きな自由度の系には適応できておらず、数残基から最大 10 残基程度の水溶液中のペプチドにしか応用されていない。新たな技術的革新によって、この過程を迅速化・自動化する必要がある。

2.2 システム概要

新 MD Simulation System のシステム概要を以下に示す。システム全体としては、大きく分けて次の機能からなる。

初期データ準備 (初期座標、初期速度の決定)

MD 計算

MD 計算結果分析

Multicanonical-MD 用パラメータ作成 (Multicanonical-MD の場合)

各機能の関係を以下に示す。

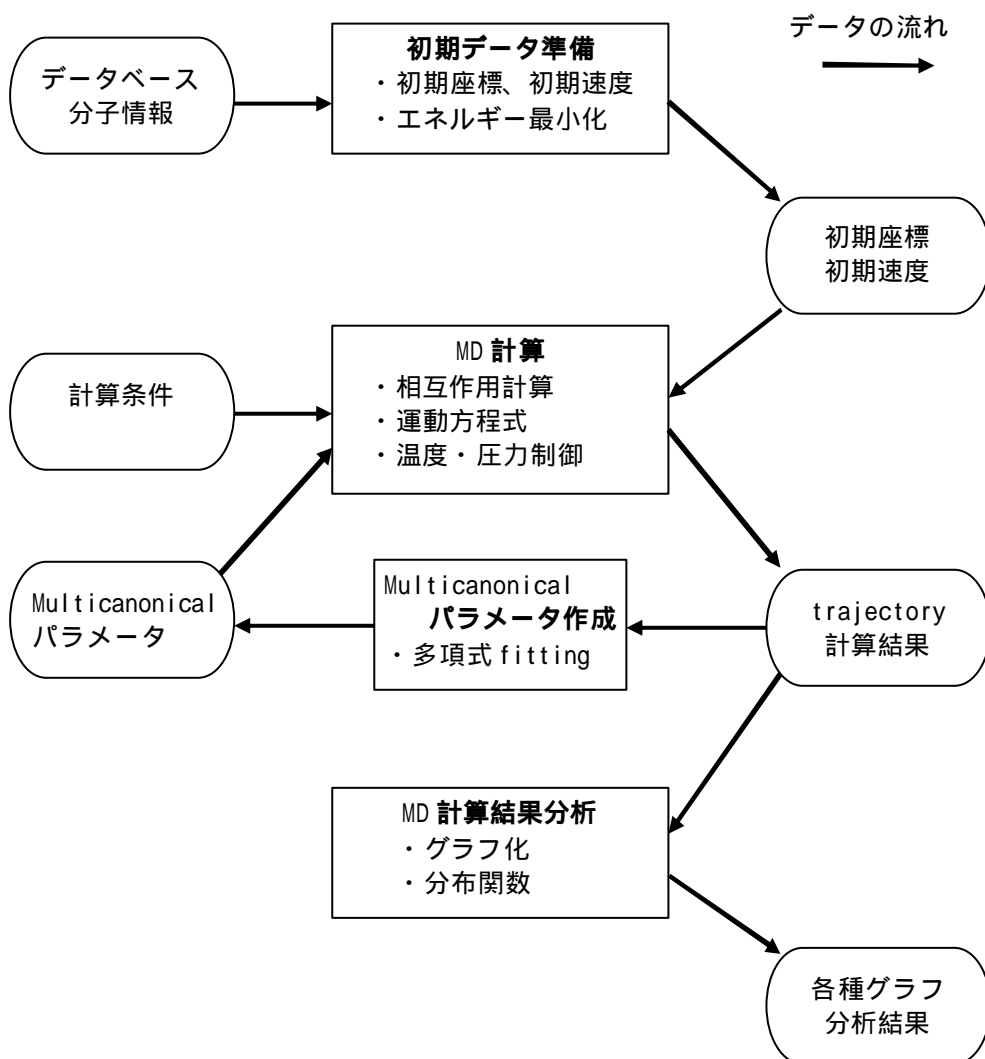


図1 各機能の関係

(1) 初期データ準備

MD 計算を行うためにはまず、対象とする分子の初期座標と初期速度を決定する必要がある。分子の情報としては実験から求められた座標情報を使用するが、この情報には水素原子の情報が入っていない場合が多い。このため、水素原子を付加する作業が必要になる。また、全体の系のエネルギー的に安定な状態を作っておく必要もある。これらの作業を人手ですべて行うのは非常に多くの時間と労力を必要とする。

新 MD Simulation System では初期データ準備支援機能として、トポロジー情報作成機能、座標情報作成機能、エネルギー最小化機能を実装する。

(2) MD 計算

初期状態での分子の初期座標（トポロジー情報と原子座標）と初期速度を使用して、分子の conformational エネルギーについて計算する。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方程式により全ての原子を一斉に運動させる。そして、時間に対する原子の位置と速度などの情報から種々の物理量を求める。

新 MD Simulation System では、各種 trajectory 情報、自由エネルギー計算結果などの情報、さらに、観測したい構造部分の trajectory 情報（モニター情報）を出力する。

(3) MD 計算結果分析

MD 計算によって出力される各種データに対して分析を行う。分析結果に対してはグラフ化も行う。新 MD Simulation System では、各種 trajectory 等の情報を出力する。これらの情報から、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。

(4) Multicanonical パラメータ作成

canonical 分布（または Multicanonical 分布）に対して polynomial fitting を行い、Multicanonical-MD 用のパラメータを作成する。このパラメータは Multicanonical-MD を実施する際の入力データのの一つである。

3. エネルギー最小化機能

3.1 概要

トポロジージェネレータなどで対象とする分子の初期構造を作成した後は、エネルギー最小化計算 (Energy Minimization) を行う。ほとんどの場合、そのままの初期構造ではエネルギー的に理想的な状態ではない。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ておく必要がある。エネルギー最小化は、与えられた初期構造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方法である。

エネルギー最小化によって求められた分子構造は、MD 計算の出発座標となる。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。

(1) エネルギー最小化手順

以下にエネルギー最小化の手順を示す。

与えられた原子の配置、原子間に働く力の様式にもとづき、原子の相互配置 (原子間距離、角度) から、個々の原子のエネルギーと働く力を求める。

エネルギーが小さくなる方向に、原子を移動させる。

収束条件との比較を行い、条件を満たせば次へ、そうでなければ最初の処理へ戻る。

(ここまでの処理を繰り返し行う)

(2) 最小化アルゴリズム

新 MD Simulation System では、以下のアルゴリズムでの最小化計算を実現する。

最急降下法 (Steepest descent method)

拘束つき最急降下法 (Steepest descent method with constraints)

共役勾配法 (Conjugate gradient method)

NOTE

最小化のアルゴリズムは、ポテンシャルエネルギー面上の局所的な極小点 (local minimum) を見つけ出すだけであり、必ずしも大域的な極小点 (global minimum) が発見できるわけではない。

3.2 機能構成

新 MD Simulation System のエネルギー最小化機能の機能構成を図 2 に示す。

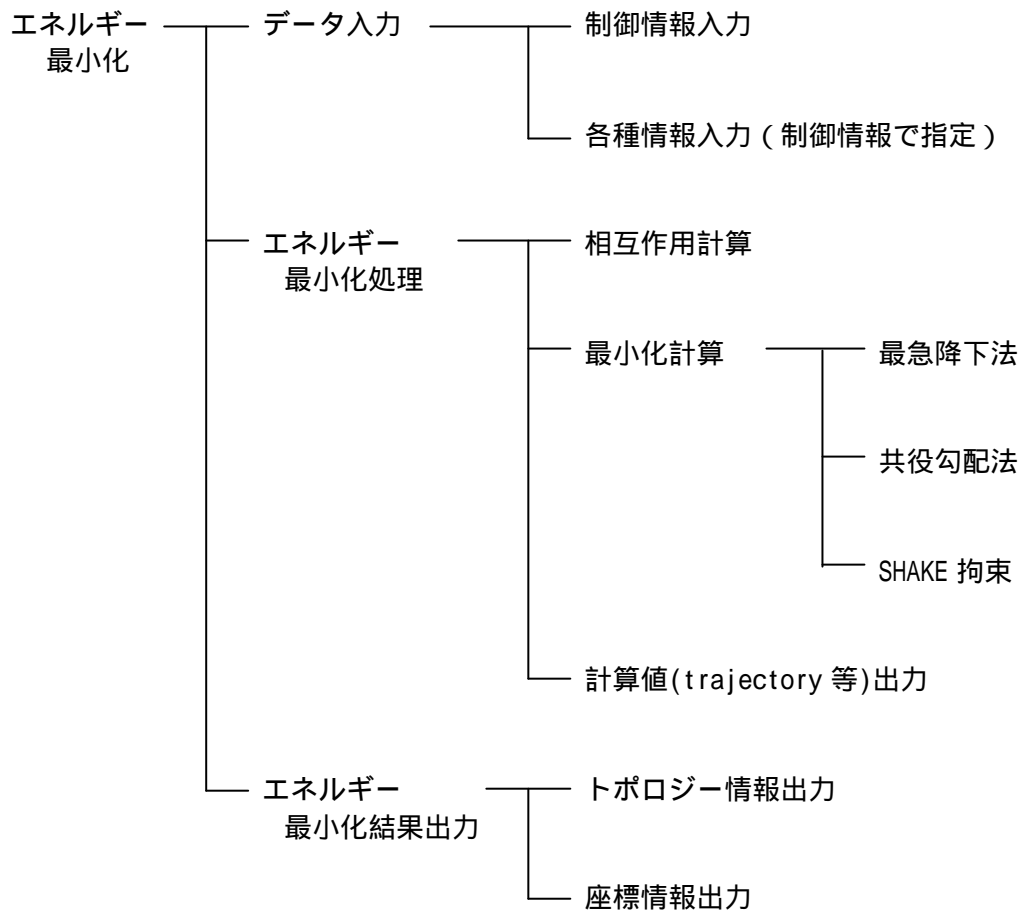


図 2 機能構成

以下では、新 MD Simulation System のうち、「エネルギー最小化機能」の実現方式について説明する。なお、「相互作用計算」および「SHAKE 拘束」については、MD 計算機能と同様の計算を行う。これらの実現方式については「新 MD Simulation System 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)」を参照のこと。

3.3 最急降下法

最急降下法では、数値的に計算されたポテンシャルエネルギー関数の一次微分（つまり力）を利用してエネルギー極小点へ近づけていく。この方法による計算手順を以下に示す。

まず、初期構造に対しポテンシャルエネルギーと力を計算する。

構成原子を座標系の軸方向に沿って順番に少しずつ動かし、その都度エネルギーと力を再計算する。

上記のプロセスを全原子に対して繰り返し、全原子をポテンシャルエネルギー面の下り坂方向の新しい位置まで動かす。

操作は、あらかじめ定めておいた判定条件を満たした時点で打ち切る。

(1) 探索方向の計算

$$\mathbf{S} = -\mathbf{G}$$

\mathbf{S} : Search direction.

\mathbf{G} : Gradient.

(2) 探索（位置の計算）

$$\mathbf{R}_t = \mathbf{R}_c + (L \times \mathbf{S})$$

\mathbf{R}_t : Trial molecule structure.

\mathbf{R}_c : Current molecule structure.

L : Step length.

\mathbf{S} : Search direction.

If $E_t \geq E_c$, L is lowered by constant rate (1.0) multiplication.

If $E_t < E_c$, L is raised by constant rate (1.0) multiplication.

E_t : Potential energy of trial molecule structure.

E_c : Potential energy of current molecule structure.

3.4 拘束条件つき最急降下法

最急降下法の小さい変更によって、拘束条件つきの最小化計算を実行することが可能になる。これには、SHAKE 拘束（共有結合の結合距離を抑制する）を使う。SHAKE 拘束によるこの最小化は、エネルギーが最小となる分子構造、理想的な長さに固定される結合距離、を見つけることができる。拘束のない通常の最小化では、多くのエネルギー項に影響を受けるので、結合距離と結合角が理想的な値になるように維持することはできない。

拘束条件つき最急降下法は、通常の最急降下法とは次の点で異なる。

探索方向の計算では、拘束した結合に沿ってだけ振舞うエネルギー項を除く。

新しい分子構造 (R_t) におけるエネルギーについて計算する前に、新しい分子構造に対して SHAKE 拘束を適用する。

長さ係数 (L) の変更方法は、通常の最急降下法の場合と同じである。

NOTE

最急降下法は、極小点付近では収束が遅くなるため、極小点から遠く離れた構造に対して適用されることが多い。この方法は、十分精密化されていない結晶データをもとに低エネルギー構造を生成させたり、二次元構造式から組み立てられた三次元分子構造を緩和させるのに特に適している。最急降下法はこのような特徴であるため、共役勾配法のような、より高等なアルゴリズムを用いる前の予備的な最小化計算として使用される。

3.5 共役勾配法

共役勾配法は、ある繰返しサイクルから次のサイクルへ至る関数の情報をすべて蓄積していく。そのためこの方法では、前のサイクルで進めた位置が次のサイクルで逆戻りすることはない。最小化計算の各ステップで勾配を計算し、新しい方向ベクトルを計算するための追加情報として利用する。ベクトルの方向は、ステップを進めるごとに極小点へ向かい絶えず改善されていく。

(1) 探索方向の計算

新 MD Simulation System では、共役勾配法での探索方向の計算に、次の3タイプの方法を使用する。

最急降下法と同じ。 : 初期の探索方向としてのマイナス勾配

、 を導入したもの。 : 最小化手続きで使用

$$\mathbf{S}^{<i+1>} = -\mathbf{G}^{<i>} + (\alpha^{<i>} \times \mathbf{S}^{<i>}) + (\beta^{<i>} \times \mathbf{S}^{<k>})$$

を導入したもの。 : restart の探索方向として使用 (未収束時)

$$\mathbf{S}^{<i+1>} = -\mathbf{G}^{<i>} + (\alpha^{<i>} \times \mathbf{S}^{<i>})$$

$\mathbf{S}^{<i>}$: i-th iteration's search direction.

$\mathbf{G}^{<i>}$: Gradient at the final molecule structure in i-th line search.

$\alpha^{<i>}$, $\beta^{<i>}$: Constant value defined by the gradient and the search direction.

k : Iteration number of restart.

NOTE

共役勾配法は最急降下法に比べ、計算に多大な時間と記憶容量 (原子の数Nに比例したメモリサイズ) が必要である。しかし、共役勾配法による最小化計算の効率は良く、ポテンシャルエネルギー関数がM個の変数の二次方程式ならば、エネルギーの極小点は高々M回の繰返しによって見つけることができる (見つからない場合は、restart)。共役勾配法により達成できる極小点への効率的な収束は、繰返しサイクル当りの計算時間の増加を十分にかくす。

(2) 探索 (位置の計算)

$$R_t = R_c + (L \times S)$$

R_t : Trial molecule structure.

R_c : Current molecule structure.

L : Step length.

S : Search direction.

ここで、エネルギーの最小値を与える L をどのように選ぶか、ということが問題となる。必要な条件は、 R_t (R_t が探索方向でエネルギー最小の構造になる) を求める際に計算される「Directional derivative ($G \times S$) がゼロ」であるということである。

新 MD Simulation System では、共役勾配法における探索手順は以下の通りである。

Directional derivative が L で「負」になる場合、新しい L を次のように定義する。

$$L = L \times 2.0$$

Directional derivative が L で「正」になる場合、新しい L を次のように定義する。

$$L = \frac{(L_n \times D_p) - (L_p \times D_n)}{D_p - D_n}$$

L_n : Step length in which the directional derivative is negative.

L_p : Step length in which the directional derivative is positive.

D_n : Directional derivative at L_n .

D_p : Directional derivative at L_p .

絶対値 $|(D_c / D_i)|$ が、評価基準値 V_{conv} 以下になるまで探索を繰り返す。条件を満たしたなら、次の探索方向の計算処理を行う。

D_c : Current directional derivative.

D_i : Initial directional derivative in the line search.

V_{conv} : Criterion of the convergence in the line search.

NOTE

エネルギー最小化のアルゴリズムの選択は、対象とする系の大きさと最適化の現在段階の2つの因子に依存する。対象が極小点から遠く離れた構造である場合、繰返しの最初の10~100サイクル程度において最良の結果を与える方法は、一般に最急降下法である。この最小化のプロセスは、そのあと共役勾配法やNewton-Raphson法などを用いて収束させることができる。対象とする系が大きすぎて二次微分行列を保存したり計算したりできない場合には、共役勾配法が使用できる唯一の方法になる。

エネルギー最小化の際の収束判定基準は、最急降下法では、分子系の現在の構造の適否は、エネルギーや座標の増分を用いて判定するしかないが、共役勾配法などでは、原子の勾配を利用することができる。この場合において最良の結果を与える方法は、分子の各構成原子上における力の二乗平均勾配を計算する方法である。望ましくない構造領域は、最大微分を吟味することにより検出可能である。すべての微分が与えられた値よりも小さければ、エネルギー極小構造の質について疑う余地はない。収束判定値は、たとえば最大微分の場合、極小化の目的に応じて変化する。もし、歪んだ分子を単に緩和させるだけで良いのであれば、最大微分に対して緩い(大きな値の)収束条件でも十分であろう。しかし、最終的な最小点を見つけ出すためには、より厳密な(小さな値の)判定基準が最大微分に対して要求される。

収束条件は、次の間の釣合いを考えて選択しなければならない。

- ・合理的な精度でエネルギー最小構造が決定できること。
- ・進展がそれ以上望めなくなったときに、不必要な計算が極力避けられること。

4. MD 計算結果分析機能

4.1 概要

MD 計算によって出力される各種データに対して分析を行う。分析結果に対してはグラフ化も行う。新 MD Simulation System では、各種 trajectory 等の情報を出力する。これらの情報から、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。また、トポロジージェネレータなどで生成した初期構造や MD 計算結果の最終構造から、その構造におけるエネルギーや力の状態を求める。

(1) 構造情報分析 (静的情報)

トポロジージェネレータなどで生成した初期構造や MD 計算結果の最終構造から、以下の計算を行う。

構造情報のチェック

エネルギーおよび力の計算

静電ポテンシャルエネルギーの計算

(2) 計算値分析 (trajectory 情報)

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

4.2 機能構成

新 MD Simulation System の MD 計算結果分析機能の機能構成を図 3 に示す。

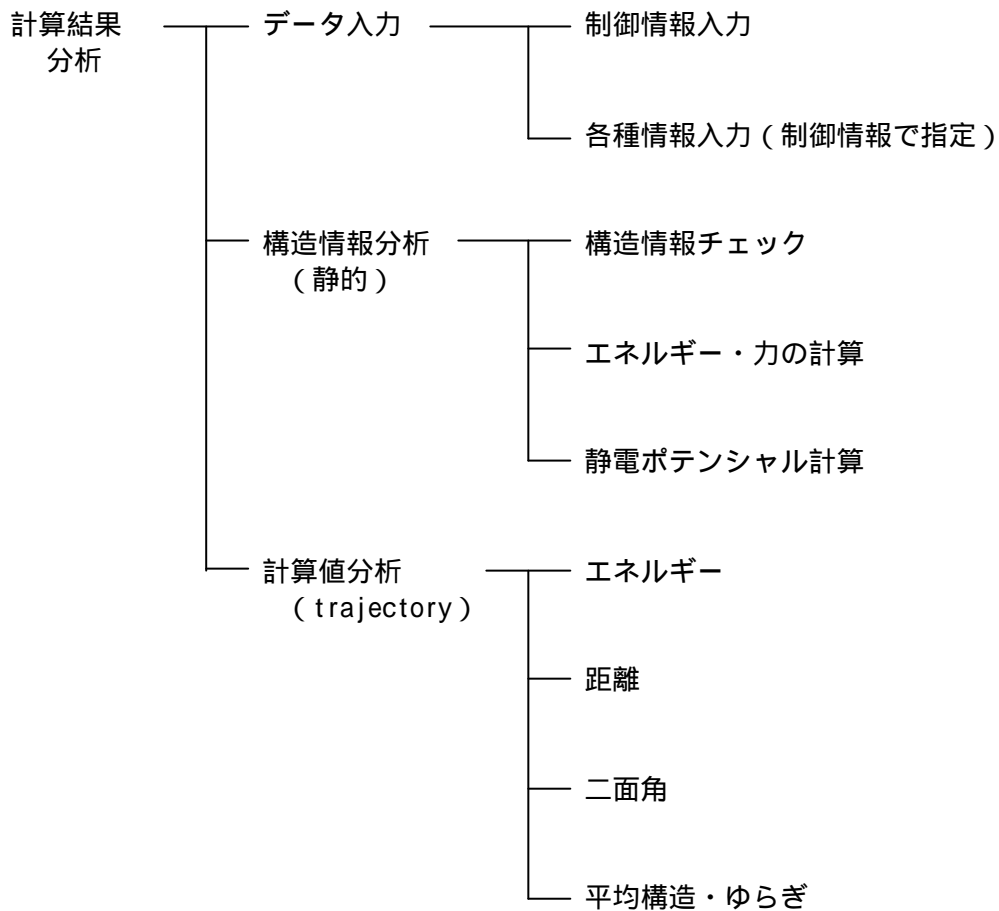


図 3 機能構成

以下では、新 MD Simulation System の「MD 計算結果分析機能」の実現方式について説明する。なお、「相互作用計算」および「SHAKE 拘束」については、MD 計算機能と同様の計算を行う。これらの実現方式については「新 MD Simulation System 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）」を参照のこと。

4.3 構造情報分析

トポロジージェネレータなどで生成した初期構造や MD 計算結果の最終構造から、その構造におけるエネルギーや力の状態を求める。

(1) 構造情報チェック

対象分子の構造情報について、結合距離、結合角、van der Waals 接触の状態を調査し、報告する。それぞれの情報に対して基準値（エネルギーが最小になる構造での結合距離、結合角。van der Waals 接触の場合は接触がない状態）との比較を行う。基準値を指定した割合以上に越えたものについて、その構造情報を出力する。

主に以下の情報を出力する。

構造情報（結合 ID、結合角 ID、原子 ID）

計算値（結合距離、結合角、van der Waals 接触）

基準値、誤差率、誤差率の根平均二乗（Root Mean Square : RMS）

(2) エネルギー・力の計算

対象分子の構造情報から、各原子におけるポテンシャルエネルギーと力の計算を行う。各原子に作用する力は、ポテンシャルエネルギーのマイナス勾配で求められる。

主に以下の情報を出力する。

構造情報（原子 ID、残基 ID）

計算値（各原子における力）

全ポテンシャルエネルギー、力の根平均二乗

(3) 静電ポテンシャル計算

対象分子の構造情報から、各原子における静電ポテンシャルエネルギーの計算を行う。

主に以下の情報を出力する。

構造情報（原子 ID、残基 ID）

計算値（各原子における静電ポテンシャルエネルギー）

4.4 計算値分析

MD 計算によって出力される各種データに対して分析を行う。分析結果に対してはグラフ化も行う。新 MD Simulation System では、各種 trajectory 等の情報を出力する。これらの情報から、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) 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

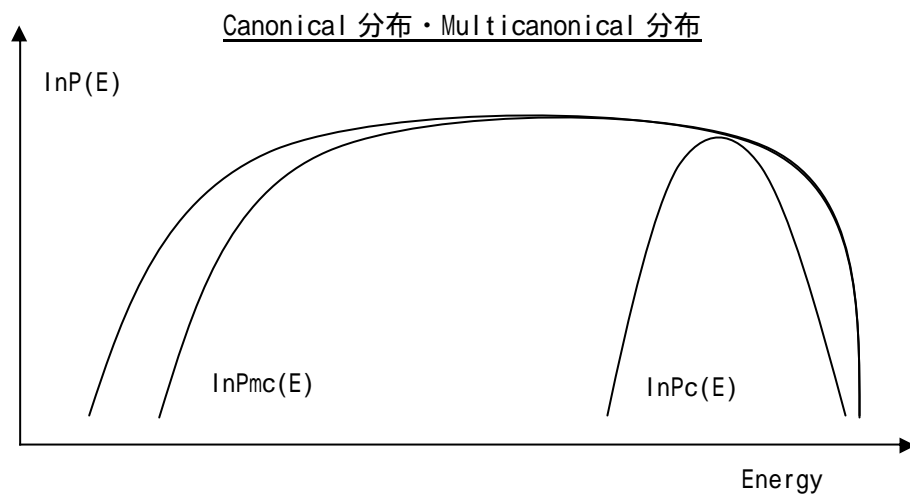
新 MD Simulation System では、主に以下の項目について情報を出力する。

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。

新 MD Simulation System では、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。

新 MD Simulation System では、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 >

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

R : 座標
 $\langle R \rangle$: 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)

新 MD Simulation System では、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)

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.