

生体分子の分子動力学シミュレーション (1) 方法

古明地 勇人^{a,c*}, 上林 正巳^b, 長嶋 雲兵^a

^a 産業技術融合領域研究所 〒 305-8562 つくば市東 1-1-4

*e-mail: komeiji@nair.go.jp

^b 生命工学工業技術研究所 〒 305-8566 つくば市東 1-1

^c 電子技術総合研究所 〒 305-8568 つくば市梅園 1-1-4

(Received: June 10, 1999 ; Accepted for publication: June 30, 1999 ; Published on Web: September 20, 1999)

分子動力学シミュレーションは、タンパク質やDNAなどの生体高分子の構造・物性・機能相関の研究に有用であることが認められている。本総説では、筆者らが開発したソフトウェア PEACH (Program for Energetic Analysis of bioChemical molecules) に導入した機能を中心に、生体高分子の分子動力学シミュレーションの方法を概観する。

キーワード : Program for Energetic Analysis of bioChemical molecules (PEACH), GRAvity piPE (GRAPE), Molecular dynamics, Proteins, DNA, RNA

目次

1. 古典分子動力学シミュレーションの概要
2. 初期構造と初期速度
3. 力とポテンシャルの計算
4. 時間積分
5. アンサンブル (温度制御など)
6. トrajジェクトリーの解析方法
7. おわりに

1 古典分子動力学シミュレーションの概要

本総説では、生体分子分子動力学用ソフトウェア PEACH (Program for Energetic Analysis of bioChemical molecules, 古明地, 1995) に導入されたアルゴリズムを中心に、タンパク質やDNAなど生体高分子の分子動力学シミュレーション (MD) の方法を解説する。

古典分子動力学法 (古典 MD) とは、一般に「 N 体多体系の状態の時間発展を、ニュートンの運動方程式の数値解を差分法により求めてシミュレーションすること」と定義できる。つまり、連立微分方程式

$$m_i \frac{d^2}{dt^2} \mathbf{r}_i = -\nabla_i U(\mathbf{r}_1 \dots \mathbf{r}_N) \quad i = 1 \sim N \quad 1(1)$$

を、時間 t に関する差分法に直して解くことである。ここで、 m_i 、 \mathbf{r}_i 、は粒子 i の質量と位置、 U は系全体のポテンシャルエネルギーである。この場合、粒子は生体高分子の構成原子である。

最近は、量子力学と組み合わせた MD も盛んであるが、本総説では省略し、古典力学に基づいた MD についてのみ解説する。古典 MD の教科書を巻末に何冊か挙げておいた [1-10]。

Figure 1 に、生体高分子の MD の大まかな手順を示す。この手続きは、どんな物質の MD でも基本的には変わらない。計算機に入力するのは、分子の初期構造と力場定数である。(1) 初期構造 (座標) を入力し、速度を発生させ、(2) その座標と力場定数を用いて力を計算し、(3) 時間刻み (time step、 Δt) 後の座標と速度を求め、(4) 必要なら温度や圧力を制御する。(2)-(4) の各段階を繰り返すと、各原子の座標と速度、さらに座標と速度から得られるエネルギーなどの物理量の時間変化が得られる。Figure 1 の (5) で、これらを統計処理したり画像表示したりして、その生体高分子の構造や物性などについて議論をするわけである。なお、各原子の三次元座標の時間変化をトラジェクトリー (trajectory、軌跡) と呼ぶ。

本総説では、(1)-(5) の方法にそれぞれについて、PEACH に導入した機能を中心に、詳しく説明する。

なお、PEACH は、MD-GRAPe (GRAvity piPE for MD、商品名 ITL-MD-ONE、画像技研社、Taiji et al. [11], Komeiji et al. [12]) という、多体問題計算用の専用プロセッサと組み合わせると有効であるが、今回紹介するアルゴリズムは、特記したものを除けば、普通の汎用計算機でも利用でき、広い応用範囲を持つ。また、PEACH には、GRAPe を使わないで MD 計算するための、汎用機用バージョンもある。

2 初期構造と初期速度

本節では、MD 計算を始めるための初期データとしての初期構造 (Figure 1(1)) および初期速度について説明する。

2.1 初期構造

生体分子の初期構造 (座標) としては、Protein Data Bank (PDB) などに登録されている、X 線結晶解析や多次元 NMR による三次元座標を用いることが多い。ただし、PDB の座標をそのまま MD に入力できるのではなく、足りない原子 (特に水素) や残基を足したり、イオンを加えたり、と言ったモデリング操作を行ってから入力する必要がある。また、大抵は溶媒分子を加える。溶媒分子の扱いについては 2.2 で詳しく説明する。

2.2 境界条件と溶媒の扱い方

Figure 2.2 に、生体分子によく使われる溶媒 (生体系では、ほとんどの場合水) と境界条件を示す。

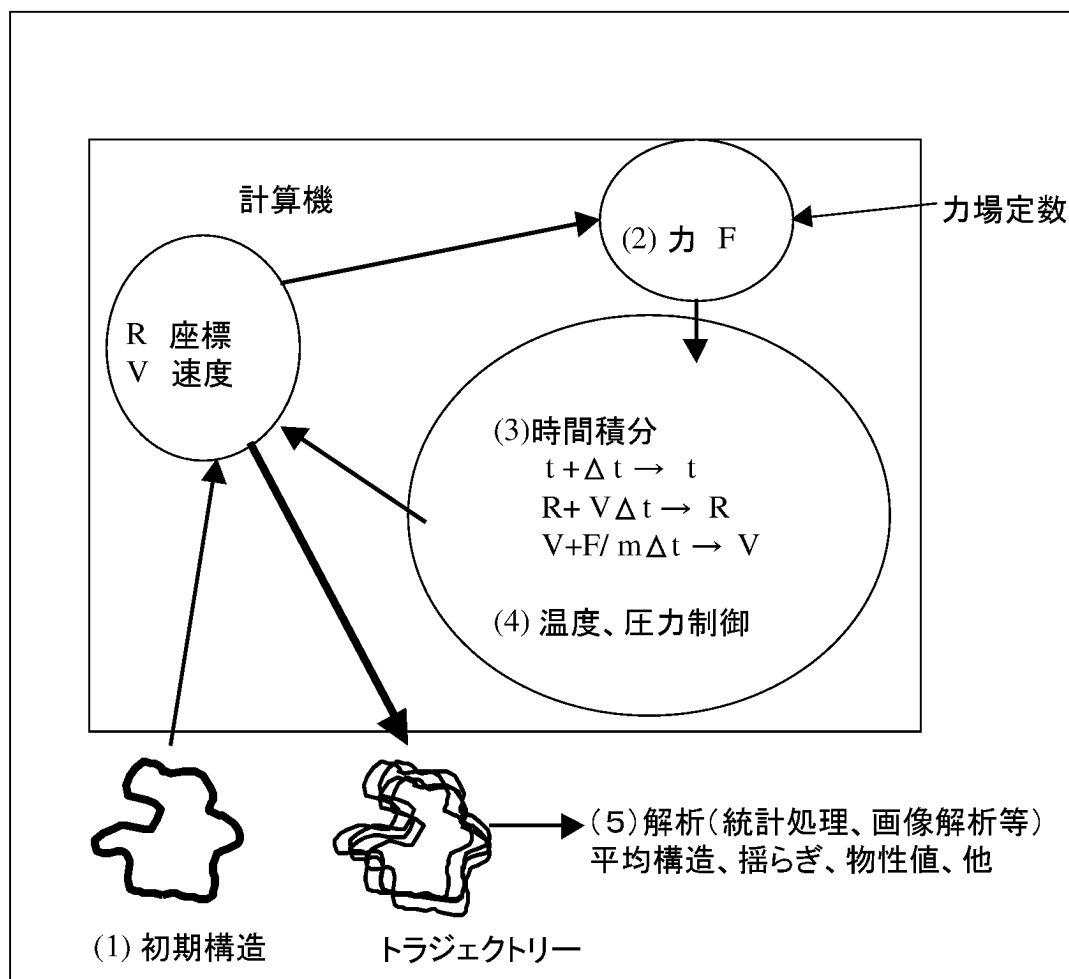


Figure 1. 分子動力学シミュレーション (Molecular Dynamics, MD)

(1)の孤立系のように溶媒を直接取り入れない場合は、溶媒を連続体近似して誘電率を工夫することが必要である (Guenot & Kollman [13] など)。いずれにせよ、溶媒分子を直接取り入れない MD は計算結果に問題点が多く (Komeiji et al. [14], Saito [15] など)、最近あまり行われない。

(2)は Shell water と呼ばれ、生体分子の周りに何層かの水を配置したもの。溶媒の分子数が少ない割には、比較的よい結果が得られる (Guenot & Kollman [13] など)。ただし、水分子に束縛を掛けないと、エントロピー効果のため水分子が蒸発してしまうことがある。

(3)は Cap water と呼ばれ、ある原子 (あるいは溶質の重心など) を中心にして球状に水を発生させる。球の中の水は自由だが、外に出て行こうとすると、軽い束縛力 (バネの力) により中に押し戻されるようにする。

従来は Figure 2.2(3)A のように、タンパク質の一部だけを覆って活性部位のダイナミクスを追ったり自由エネルギー計算を行ったりする場合に使うことが多かった。水が野球帽のよう

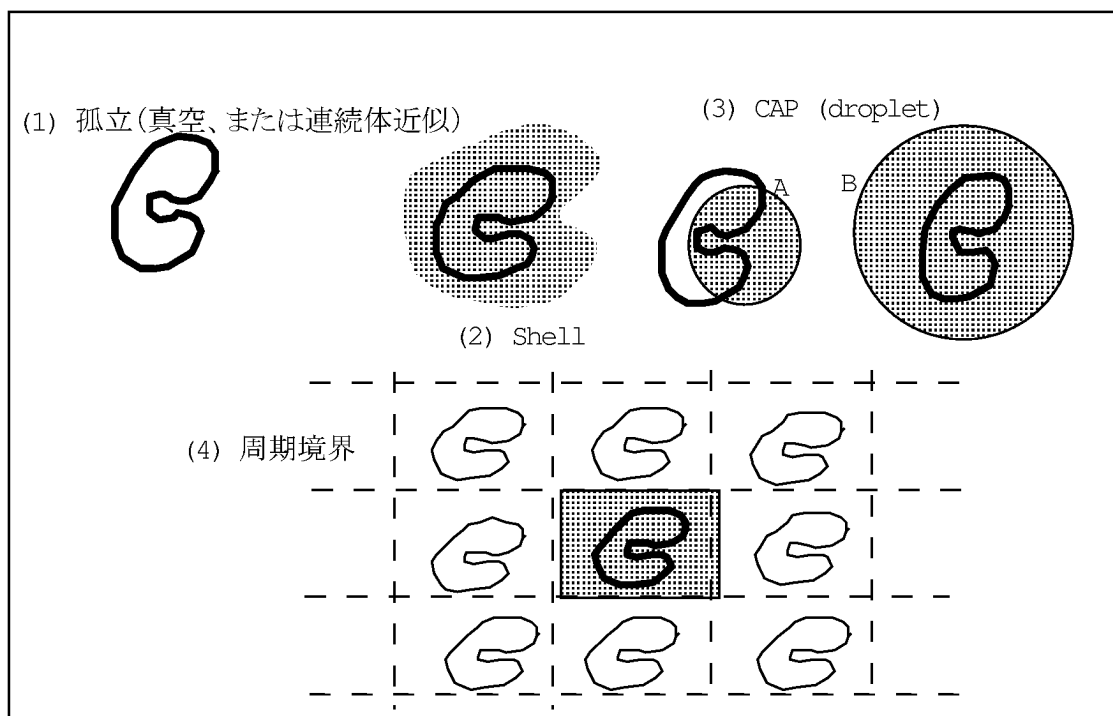


Figure 2.2. 生体分子のMDにおける溶媒の扱い

な形になることから Cap water と呼ばれる。だが、PEACH では、Figure 2.2(3)B のように生体分子をすっぽりと覆った、言わば、ゴム風船の中に水と分子を閉じ込めた形にして使うことが多く、Cap よりもむしろ Droplet と呼ぶほうがふさわしい。この Droplet でも (2) の Shell でも、水の表面が真空に露出しているという問題点がある。それが、水や溶質の挙動にどの程度影響を及ぼすかは、1999 年の時点では系統的な報告は見あたらない。

周期境界条件 (Periodic boundary condition、Figure 2.2(4)) とは、溶質と溶媒を基本セル (実像) の中に入れ、そのレプリカ (虚像) からの寄与も計算に考慮する方法である。この方法では、(2) や (3) と違って表面が真空に露出しない。溶液や固体の MD では、周期境界条件を使うのが普通であり、また、計算機の進歩や高速アルゴリズムの普及により (Darden et al. [16] など) 生体分子の MD にも多く使われるようになった。

周期的でない生体分子を周期的に扱うことによる問題点については、まだ、詳細な調査はない。しかし、Smith et al. [17, 18]、De Souza & Ornstein [19] などは、問題なさそうだとしている。溶質部分に注目するか、溶媒部分に注目するかでも、結論は異なるだろう。

なお、Figure 2.2 のどの溶媒モデルがよいかは一概に言えない。対象とする分子の性質と大きさ、シミュレーションしたい条件、求めたい物理量、利用できる計算機資源等により、選択は変わるからである。現在、生体高分子の MD でよく使われるのは、droplet((3)B) と周期境界 (4) である。

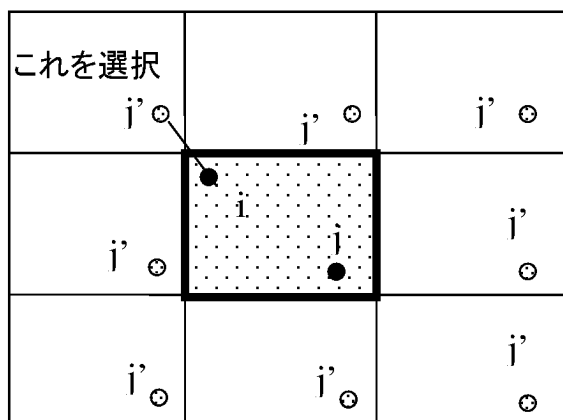


Figure 2.2.1. 最近接粒子の選択

2.2.1 周期境界条件での最近接粒子の選択

周期境界条件の場合 (Figure 2.2(4))、Figure 2.2.1 のように数多い虚像の中から、最近接粒子のみを選択して相互作用を計算することが多い。これは、粒子 i と j の相互作用を計算するとき、粒子 j の実像と 26 個の虚像 (図は 2 次元だが、実際は 3 次元) の計 27 個の中から、最も i に近いものを選んで計算する方法である。Allen & Tildesley [6] の教科書にいくつか方法が載っているが、PEACH では、以下のアルゴリズムを利用している。

$$x_{ij} = x_i - x_j$$

$$x_{ij} = x_{ij} - x_{box} * \text{anint}(x_{ij}/x_{box})$$

x_{box} は、箱の一辺の長さである。また、 anint は fortran の組み込み関数で、 $\text{anint}(a)$ は a を四捨五入した整数である。 y 、 z についても同様に計算する。

2.2.2 系の電荷の中和

エワルド法および、類似の高速化法を使って周期境界条件でクーロン力を計算する場合は、系の電荷を中和する (つまり 0 にする) 必要がある。以下に代表的な方法を挙げておく。

A. 中和できる数の電荷を溶質全体に加える

たとえば、タンパク質全体が + 2 の電荷を持っている場合、合計 - 2 の電荷を全原子に均等にばらまけば、系は中和する。筆者の経験では、タンパク質分子量 1 万につき、電荷が 1 ~ 2 までなら、この方法が有効である。しかし、電荷がこれ以上大きくなると、しばしば構造が崩れる場合があるので、あまり勧められない。

B. アミノ酸の荷電状態を変えて調節する

一部のアミノ酸の電荷を調節して、電荷を 0 にする。不自然な電荷分布になるからと嫌がる人もいるが、分子全体のダイナミクスに限れば、普通は大きな影響はないようである。系の一部の挙動の解析のために MD を行うならば、無関係な部分に余計な電荷を押し付け

るのも一つの方法であろう。

C. イオンを発生させる

系の電荷が大きい場合は、中和するだけの数のイオンを発生させる。一般的には、これが最も現実的で汎用性がある方法である。PEACHでは、発生させたイオンと溶質の間のクーロン相互作用が最小になるような位置の溶媒分子を選んで、それとイオンと置換することで発生させている。結果的に電荷をもった残基の近くに発生されることが多い。なお、電荷を中和させるためのイオンを発生させる方法には、それらを全くランダムに配置する、あるいは電荷を持った残基の近くに配置する、などの方法もある。

2.3 エネルギー極小化 (Energy Minimization)

初期構造を作成した後は、エネルギー極小化計算 (Energy Minimization, EM) を行うことが多い。エネルギー極小化計算にはさまざまな方法があり (Brooks et al. [20] に詳しい) 目的によって使いわける。MD の下準備としての EM は、初期構造が持っている不自然な構造の歪みを取って、MD 初期における時間積分の発散を避けるために行うのである。

PEACH では、現在のところ、単純な最急降下法 (Steepest descent method、[20] および引用文献参照) しかサポートしていない。ただし、低温で MD を行えば EM と同様の効果が得られるので (Quenched dynamics [20]) 利用して欲しい。

2.3.1 最急降下法 (Steepest descent method)

エネルギーの一次勾配 (つまり力) を計算して、エネルギーが低くなる方向に系を動かす方法である。以下に PEACH で利用したアルゴリズムの疑似コードを示す。

```
 $\Delta R = \Delta R_{ini}$ 
do i = 1, nstep
  compute F and  $ENE_{new}$  ! 力とエネルギーを計算
   $\Delta ENE = ENE_{new} - ENE_{old}$ 
   $dRMS = \sqrt{\sum \mathbf{F}^2} / \sqrt{Ndfree}$  !  $Ndfree$  は系の自由度
  EXIT if either  $\Delta ENE$  or  $dRMS$  is converged
  if  $\Delta ENE > 0$  then
     $\Delta R = 1/2 \Delta R$ 
  else
     $\Delta R = \min(\Delta R_{max}, 1.2\Delta R)$ 
  end if
   $\mathbf{R} = \mathbf{R} + \Delta R / dRMS * \mathbf{F}$ 
   $ENE_{old} = ENE_{new}$ 
end do
```

ΔR はステップの長さで、距離の次元を持つ。 ΔR は ΔR_{ini} の初期値を与えられるが、計算の途中でいろいろな値に調節される。

上記の do ループでは、nstep 回だけ極小化が行われる。まず、力とエネルギーを計算する。こ

のとき、 F と ENE_{new} の計算法は、第 3 章で説明する MD の場合と同じである。前ステップとのエネルギーの差 (ΔENE) と力の自乗平均 ($dRMS$) をそれぞれ計算する。ここで、収束条件 (後述) が満たされたならば計算を終える。

if 構文の中では、ステップの長さ ΔR を調節している。この部分は任意性があり、論文やソフトのマニュアルに記述されていないことが多い。上記のアルゴリズムでは、エネルギーが増加してしまった場合は、ステップの長さを半分に減らし、そうでない場合は 1.2 倍にする。ただし、上限 (ΔR_{max}) を設けておく。

最後に、座標 \mathbf{R} を力の働く方向に動かして、1 ステップが終りである。

計算の条件を左右する入力パラメータを改めて以下にまとめておく。

- $nstep$: ステップ数。
- ΔR_{ini} : ΔR の初期値。
- ΔR_{max} : ΔR の最大値。

さらに、収束条件として、以下の二つを入力する。どちらか一方が満たされれば、ステップの途中でも計算を打ち切る。

- ΔE_{conv} : $|\Delta E| < \Delta E_{conv}$ になったら収束とみなす。
- $dRMS_{conv}$: $|dRMS| < dRMS_{conv}$ になったら収束とみなす。

これら 5 つのパラメータをどう設定すればよいかは、シミュレーションする系の性質により大きく変わるから一概に言えない。なお、AMBER 等の市販プログラムで使用するパラメータが上記のものと全く同じかどうかは不明である。いずれにせよ、最急降下法は収束しにくい方法なので、収束しなくても数 10 ~ 数 100 ステップ程度で打ち切って、Quenched dynamics に移ったほうがよい。

2.4 初期速度

初期速度 (初期温度と言い換えてもよい) の与え方にはいくつか流儀がある。

簡単なのは、初期速度を与えず (ゼロのまま) MD を開始することである。初期構造のエネルギーが完全に極小化されていることはほとんどなく、普通はわずかに歪みを持っている。従って、初期速度を与えなくとも、その歪みを解消する方向に力が働き MD を開始させることができる。この方法の利点は、自然に力が働く方向に粒子が動き始めるため、初期 (ほんの数ステップの間である) に時間積分が発散してしまう恐れが少ないことである。一方、問題点としては、温度緩和がしにくい条件 (クーロン力を切断した場合など) での MD では、初期の温度格差が長く残ってしまうことが挙げられる (hot spot と呼ばれる)。

一般には、ボルツマン分布に従うように初期速度を発生させることが広く行われている。すなわち、粒子の速度の x 成分を v_x 、質量を m 、系の初期温度を T 、 k をボルツマン因子とすると、平均 0、分散 kT/m の正規乱数を発生させ v_x に与えればよい (y 、 z 成分についても同様)。この方法だと hot spot はできにくい、一方、初期構造の歪みが大きい場合、時間積分が発散し

やすいという欠点もある。ただし、最初期の時間積分の発散は、初期だけ時間幅を短くしたり、温度制御を強くかけたりすることで、大抵は避けられる。初期温度 T の与え方は人により系によりまちまちであり、いきなり 300K にしてしまう場合もある。しかし、経験的には 0.1–5K 程度の低温を与えて徐々に温度を上げていくほうが発散の危険が少ない。

なお、平衡化、サンプリングに十分時間を掛けて結果を出した場合は、初期速度の与え方によらない結果が得られるはずである。逆に、初期速度の与え方に結果が大きく依存する場合、平衡化やサンプリングの時間が足りないことになる。

2.5 その他

ここでは、初期条件とは必ずしも関係ないが、PEACH に導入した MD の特殊なテクニックを挙げておく。

2.5.1 系の一部だけを動かす Belly 法

MD や EM で系の一部だけを動かす方法を Belly 法と呼ぶ。ベリーダンス(腹踊り)に似ているから、この名前が付いたらしい。動かす部分を Belly と呼ぶ。計算機のパワーが低い時代に、たとえば酵素活性中心の回りだけ動かして解析する、というような形で使われた。また、90 年代初頭ぐらいまでの自由エネルギー計算では、ごく普通に Belly を使っていた。

Belly を使う意義は二つある。

1. 計算時間を短縮する。
2. 系の一部を止めて他の部分だけ動かして最適化等をする。

しかし、PEACH の場合 (1) の目的では使えない。力を計算する部分に BELLY を導入すると、プログラムが複雑になり計算効率が落ちるので、BELLY を使う使わないにかかわらず、全ての原子への力を計算するからである(つまり、PEACH で BELLY を使っても、計算時間が短縮されない)。PEACH で Belly を使うのは (2) の目的の場合のみで、たとえば、タンパク質や核酸部分は固定して、水やイオンのみ最適化させる場合に使われる。

PEACH では、Belly として、溶質、イオン、溶媒というようにも指定できるし、特定の分子、残基、原子を指定することもできる。また、水素原子だけを動かすこともできる。

2.5.2 重心運動の凍結

PEACH では、オプションとして、イオンや溶媒を含めた系全体の重心の速度をゼロにすることができる。初期速度を重心運動がゼロになるように設定しても、計算誤差が積もることで、重心が速度を持つようになってしまう。そこで、ステップ毎に重心運動の分は、全体の速度から引くようにしているわけである。

一般には、2.5.1 の Belly を使う場合以外は、重心運動は凍結してしまったほうがよい。特に、温度一定アルゴリズムを使う場合は、重心運動を凍結しないと、「Flying ice cube」の問題が生ずる。これは、重心運動だけがどんどん速くなって行って、系の内部運動が低下してしま

う、つまり温度が低くなってしまいう現象である (Harvey et al. [21])。

3 力とポテンシャルの計算

MDやEMにおいては、各ステップ毎に全粒子に掛かる力を計算する (Figure 1(2))。この力の関数形、およびパラメータは、計算対象や要求される精度等により様々である。タンパク質などの生体分子用には古くから何種類かパラメータのセット (力場 force field と呼ぶ) が開発、改良され続けている。現在 PEACH では、AMBER の 3 種類の力場 (Weiner et al. [22, 23], Cornell et al. [24]) と OPLS 力場 (Jorgensen & Tirado-Rives [25]) の計 4 種類を利用できる。

ここでは、最新の AMBER 力場 (Cornell et al. [24]) を例にして説明することにする。

系の全ポテンシャルエネルギー U は次の式で表わされる。

$$\begin{aligned}
 U &= \sum_{\text{bonds}} K_r (r - r_0)^2 && \text{bond} \\
 &+ \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 && \text{angle} \\
 &+ \sum_{\text{torsions}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] && \text{torsion} \\
 &+ \sum_{i>j} \left[\frac{a_{ij}}{r_{ij}^{12}} - \frac{b_{ij}}{r_{ij}^6} \right] && \text{VDW} \\
 &+ \frac{1}{SC_{14VDW}} \sum_{1-4\text{pairs}} \left[\frac{a_{ij}}{r_{ij}^{12}} - \frac{b_{ij}}{r_{ij}^6} \right] && \text{1-4 VDW} \\
 &+ \sum_{i>j} \frac{q_i q_j}{r_{ij}} && \text{Coulomb} \\
 &+ \frac{1}{SC_{14ELC}} \sum_{1-4\text{pairs}} \frac{q_i q_j}{r_{ij}} && \text{1-4 Coulomb} \quad 3(1)
 \end{aligned}$$

共有結合 (bond)、結合角 (angle)、捻れ角 (torsion) の力は、結合に関係する力であり N 体系の場合は $O(N)$ の計算時間で済む。プログラムは複雑である。一方、VDW 力 (ファン・デル・ワールス力) やクーロン力 (Coulomb 力、静電相互作用) は、非共有結合力と呼ばれ、何の工夫もせずに正確に計算しようとする $O(N^2)$ の時間が掛かる。プログラムは単純である。1-4 VDW や 1-4 クーロンは、1-4 ペアのみについて計算する (Figure 3 参照)。

式 3(1) に定義されたポテンシャルを距離微分すれば力を得ることができる。

$$\mathbf{F}_i = -\nabla_i U \quad 3(2)$$

式 3(1) のそれぞれについて、以下で詳しく説明するが、原子の結合トポロジーを Figure 3 のように定義して説明する。1 の原子を基準にすると、共有結合している原子は 2 の位置、さらにその隣は 3 の位置というように、位置により番号をつける。共有結合している原子のペアは 1-2

ペア、間に別の原子が一つ入っているペアは 1-3 ペア、二つ入っているペアは 1-4 ペア、というように呼ぶ。

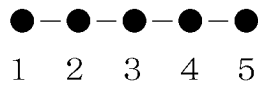


Figure 3. 原子の結合トポロジーの定義

3.1 共有結合力 (bond)

共有結合をしている原子同士 (1-2 ペア) に関しては、下のようにバネで近似したポテンシャルと力を計算する。

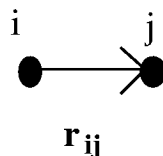


Figure 3.1. 共有結合

$$U_{bondij} = Kb(r_{ij} - r_{eq})^2 \quad 3.1(1)$$

$$\mathbf{F}_i = -\mathbf{F}_j = -2Kb(r_{ij} - r_{eq}) \frac{\mathbf{r}_{ij}}{r_{ij}} \quad 3.1(2)$$

3.2 結合角 (angle)

共有結合をしている 3 つの原子間 (1-2-3) は、次の式でポテンシャルと力を計算する。

$$U_{angleijk} = Ka(\theta_{ijk} - \theta_{eq})^2 \quad 3.2(1)$$

$$\mathbf{F}_i = -2Ka(\theta_{ijk} - \theta_{eq}) \frac{1}{r_{ij} \sin \theta_{ijk}} \left(\frac{\mathbf{r}_{kj}}{r_{kj}} - \cos \theta_{ijk} \frac{\mathbf{r}_{ij}}{r_{ij}} \right) \quad 3.2(2)$$

$$\mathbf{F}_k = -2Ka(\theta_{ijk} - \theta_{eq}) \frac{1}{r_{kj} \sin \theta_{ijk}} \left(\frac{\mathbf{r}_{ij}}{r_{ij}} - \cos \theta_{ijk} \frac{\mathbf{r}_{kj}}{r_{kj}} \right) \quad 3.2(3)$$

$$\mathbf{F}_j = -\mathbf{F}_i - \mathbf{F}_k$$

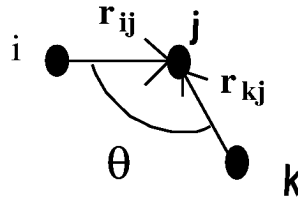


Figure 3.2. 結合角

ただし、

$$\cos \theta_{ijk} = \frac{\mathbf{r}_{ij} \cdot \mathbf{r}_{kj}}{r_{ij} r_{jk}} \quad 3.2(4)$$

である。

3.3 捻れ角 (torsion angle)

1-2-3-4の位置にある原子が作る捻れ角(二面角 dihedral angle と呼ぶこともある)によるポテンシャルと力は、以下のような手続きで計算する。フーリエ級数で展開した場合も同様に計算する(duplicated torsion と呼ばれる)

原子(i, j, k, l)によって作られるねじれ角 Φ を、Figure 3.3 のように定義する。

ただし、 $-\pi < \Phi < \pi$ とする。法線ベクトルを

$$\mathbf{n}_1 = \mathbf{r}_{ij} \times \mathbf{r}_{kj}, \quad \mathbf{n}_2 = \mathbf{r}_{kj} \times \mathbf{r}_{kl} \quad 3.3(1)$$

とすれば、

$$\cos \phi = \frac{\mathbf{n}_1 \cdot \mathbf{n}_2}{n_1 n_2} \quad 3.3(2)$$

ここで、 $-\pi < \Phi < \pi$ となるように、 Φ を求めると、

$$\phi = -\text{sign}(\mathbf{r}_{kj} \cdot (\mathbf{n}_1 \times \mathbf{n}_2)) \text{acos}(\cos \phi) \quad 3.3(3)$$

となる。

上式の捻れ角によるポテンシャルは

$$U_{\text{torsion}} = \frac{V}{2} [1 + \cos(\eta\phi - \gamma)] \quad 3.3(4)$$

とあらわされる。ここで、与えられたパラメータは V が barrier height、 η が周期、 γ が位相(0 か π)である。このとき、原子 i, j, k, l にかかる力は以下の4つの式で計算される。

$$\begin{aligned} \mathbf{F}_i &= f_a(\mathbf{f}_b \times \mathbf{r}_{kj}) \\ \mathbf{F}_j &= f_a(-\mathbf{f}_c \times \mathbf{r}_{ki} + \mathbf{f}_b \times \mathbf{r}_{ik}) \\ \mathbf{F}_k &= f_a(-\mathbf{f}_b \times \mathbf{r}_{ij} + \mathbf{f}_c \times \mathbf{r}_{jl}) \\ \mathbf{F}_l &= f_a(\mathbf{f}_c \times \mathbf{r}_{kj}) \end{aligned} \quad 3.3(5)$$

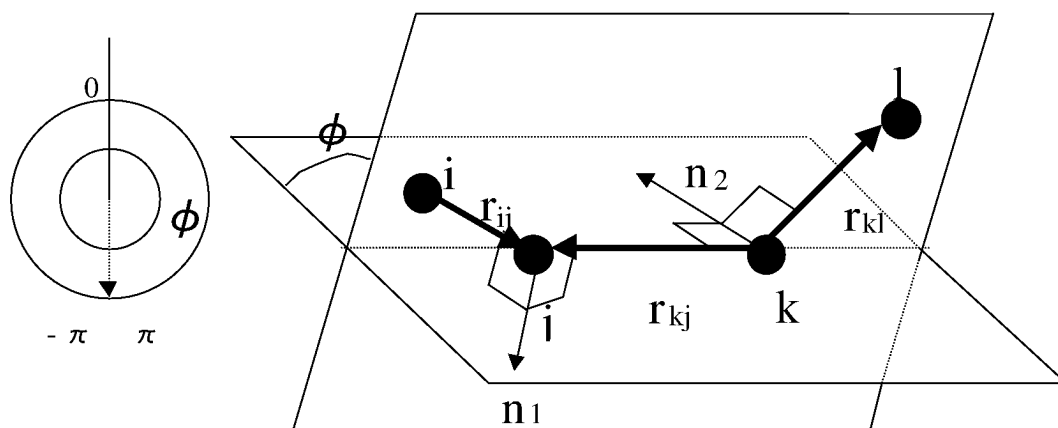


Figure 3.3. ねじれ角

ただし、 f_a 、 f_b 、 f_c は

$$\begin{aligned}
 f_a &= -\frac{V\eta \sin(\eta\phi - \gamma)}{2 \sin\phi} \\
 f_b &= \left(\frac{\mathbf{n}_2}{n_2} - \cos\phi \frac{\mathbf{n}_1}{n_1} \right) \frac{1}{n_1} \\
 f_c &= \left(\frac{\mathbf{n}_1}{n_1} - \cos\phi \frac{\mathbf{n}_2}{n_2} \right) \frac{1}{n_2}
 \end{aligned} \tag{3.3(6)}$$

である。この式では、 f_a は $\sin\phi = 0$ のとき計算できない。だが、 $\sin\gamma = 0$ ($\gamma = 0$ or π) を用いて、 $\eta = 1 \sim 4$ の場合にのみ求めると、

$$\begin{aligned}
 \frac{\sin(\eta\phi - \gamma)}{\sin\phi} &= \cos\gamma & (\eta = 1) \\
 &= 2\cos\phi \cos\gamma & (\eta = 2) \\
 &= (-4\sin^2\phi + 3)\cos\gamma & (\eta = 3) \\
 &= 4(\cos\phi(2\cos^2\phi - 1))\cos\gamma & (\eta = 4)
 \end{aligned} \tag{3.3(7)}$$

として計算できる。

なお、 $ijkl$ が、この順序に並んでいない場合は、improper torsion と呼ばれているが、同様の式で計算できる。

3.4 ファン・デル・ワールス (VDW) 力および水素結合

VDW 力は Figure 3 に示された結合トポロジーでの距離が 1-5 の位置以上に遠い原子間に適用される。共有結合していない原子ペアは、近い距離だと強く反発し合うが、遠くなると小さな力で引き合う (Figure 3.4)。これをファン・デル・ワールス (VDW) 力と呼ぶ。この力は Lennard-Jones が提唱した以下のような簡単なポテンシャルで近似されることが多い。

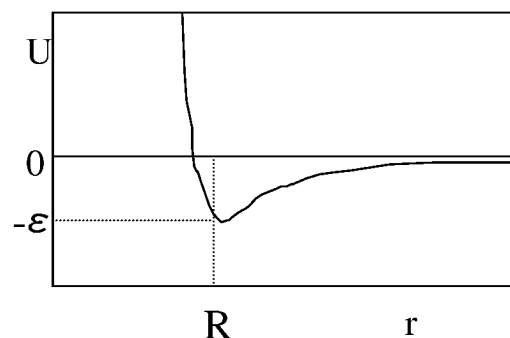


Figure 3.4. VDW ポテンシャル

$$U_{ij} = \varepsilon_{ij} \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}}{r_{ij}} \right)^6 \right] \quad 3.4(1)$$

この式で、引力項が 6 乗なのは量子力学計算に基づいているが、反発項が 12 乗であることに理論的必然性はない（本来は、引力項のみを VDW 力と呼ぶのだが、生体分子の力場では反発項も含めて VDW と呼ぶことが多い）。力は、

$$\mathbf{F}_{ij} = -\frac{12\varepsilon_{ij}}{R_{ij}} \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{13} - \left(\frac{R_{ij}}{r_{ij}} \right)^7 \right] \frac{\mathbf{r}_{ij} - \mathbf{r}_{ij}}{r_{ij}} \quad 3.4(2)$$

となる。

なお、古い AMBER 力場では (Weiner et al. [22, 23]) 1 対の原子ペアについて、VDW 力が水素結合か力のどちらか一方を計算することになっている。水素結合のポテンシャルは下の 3.4(3) 式を使っている。

$$U_{ij} = \frac{A}{r_{ij}^{12}} - \frac{C}{r_{ij}^{10}} \quad 3.4(3)$$

VDW 力や水素結合力は、一般的には原子間距離 8–15 程度で切断してしまって構わない。切断方法は 3.5.2 節参照。

PEACH では、計算機として GRAPE を使う場合は VDW 力とクーロン力は別々に計算するが、汎用機を使う場合、VDW 力とクーロン力は同じペアリストを使って (3.5.2 節) ひとつのサブルーチンで同時に計算している。

3.5 クーロン力 (Coulomb、Electrostatic、静電相互作用)

クーロン力も VDW 力と同様、Figure 3 の 1–5 結合以上に遠い原子間に適用される。クーロン力の計算は、一般に MD の計算で最も時間が掛かる部分で、さまざまな工夫がなされている。また、境界条件によっても計算法が変わる。以下に、基本的な方法を上げる。

3.5.1 非周期境界条件における直接和

Figure 2.2(1)–(3) のような非周期境界において、‘no cutoff’ 計算 (遠距離のクーロン力も切断しないで正確に計算すること) をしたければ、単純に、

$$U = \sum_{i>j} \frac{q_i q_j}{r_{ij}} \quad 3.5.1(1)$$

$$\mathbf{F}_i = \sum_j \frac{q_i q_j}{r_{ij}^3} (\mathbf{r}_i - \mathbf{r}_j) \quad 3.5.1(2)$$

を計算すればよい。だが、この方法は系の大きさが数千原子以上の時は非実用的である。ただし、GRAPE のような特殊な専用計算機を使う場合はその限りではない。

3.5.2 周期境界条件および非周期境界条件におけるカットオフ計算

生体分子の場合は、従来クーロン力を 8–15 程度の距離で切断することが多かった。これをカットオフ (cutoff) と呼ぶ。PEACH も汎用機用バージョンは、周期、非周期境界ともに、カットオフを使うことができる。ただし、原子間の距離に直接適用する atom-based cutoff ではなく、残基間の距離に適用する residue-based cutoff である。これは、Figure 3.5.2 のように、残基 i と残基 j の間の最も近い原子ペアの距離が切断半径 (cutoff radius) 内にある場合は、 i と j に属する全ての原子間のクーロン力を計算する方法である。これにより dipole split を避けることができる。

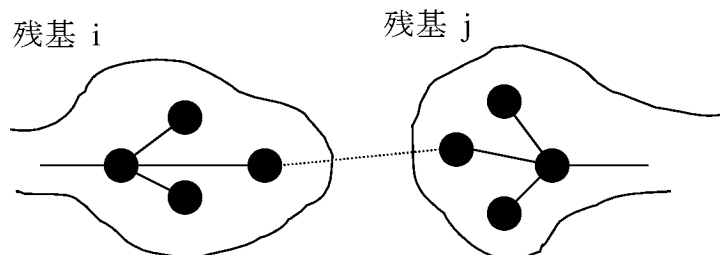


Figure 3.5.2. Residue-based cutoff

何ステップかに一回、どの原子とどの原子が相互作用するかを残基間の距離に基づいて調べてペアリストを作っておき、そのリストを使って力を計算する。PEACH では、以下のように作成する。MD を開始する前に、それぞれのアミノ酸残基や核酸塩基の原子の中で、残基の重心にもっとも近い原子を調べておく。その原子の座標を用いて、50 fs に一回ほどの頻度で残基間ペアリストを作る (最終的に作りたいリストの切断半径よりも 10 以上長い切断半径を使う)。その残基間ペアリストを利用して、最終的に使用する原子間ペアリストを 10 fs に一回ぐらいの頻度で作る。この方法により、ペアリスト作成時間を短縮できる。

なお、VDW 力や水素結合力 (3.4) も同じペアリストを使って、クーロン力と同じサブルーチンで計算する。

周期境界の場合、距離を最近接粒子間距離 (2.2.1) によって測ること以外は非周期境界と同様である。ただし、切断半径は、箱の一辺の長さの半分以下にしなければならない。

3.5.3 周期境界条件におけるエワルド和 –表式–

Figure 2.2(4)の周期境界において、基本セル内、および無限個の虚像からの基本セルへのクーロンポテンシャルの寄与の総和は、 $\mathbf{r}_{ijn} = \mathbf{r}_i - \mathbf{r}_j - \mathbf{N}$ として、

$$U_{tot} = \frac{1}{2} \sum_n \sum_i q_i \sum_j' \frac{q_j}{r_{ijn}} \quad 3.5.3(1)$$

と表わすことができる。ただし、 $\mathbf{n} = (n_x, n_y, n_z)$ を、その成分が整数のベクトル ($n_x, n_y, n_z = 0, \pm 1, \pm 2, \dots$)として、 $\mathbf{N} = (xbox \cdot n_x, ybox \cdot n_y, zbox \cdot n_z)$ 。xbox等は基本セルの辺の長さである。また、 \sum_j' のダッシュは、 $\mathbf{n} = (0, 0, 0)$ の場合に、 $i = j$ の項は計算に入れないことを示す。

式3.5.3(1)は、収束が遅い級数であり、現実のシミュレーションには使うことができない。しかし、この式を数学的にフーリエ変換することにより、以下の2つの収束の速い級数と、定数に分割することができる(エワルド法)つまり $U_{tot} = U_{real} + U_{wave} + U_{self}$ である。

$$U_{real} = \frac{1}{2} \sum_N \sum_i q_i \sum_j' \frac{q_j}{r_{ijn}} \operatorname{erfc} \left(\frac{r_{ijn}}{\eta} \right) \quad \text{実空間} \quad 3.5.3(2)$$

$$U_{wave} = \frac{1}{2} \sum_k \frac{\exp(-\pi^2 \eta^2 \mathbf{K}^2)}{\mathbf{K}^2} \left\{ \left(\sum_i q_i \cos 2\pi \mathbf{K} \cdot \mathbf{r}_i \right)^2 + \left(\sum_i q_i \sin 2\pi \mathbf{K} \cdot \mathbf{r}_i \right)^2 \right\} \quad \text{波動空間} \quad 3.5.3(3)$$

$$U_{self} = -\frac{1}{\pi^{1/2} \eta} \sum_i q_i^2 \quad \text{自己エネルギー} \quad 3.5.3(4)$$

式(3)のベクトル \mathbf{K} は、ベクトル $\mathbf{k} = (k_x, k_y, k_z)$ を \mathbf{n} 同様の整数ベクトルとすると、 $\mathbf{K} = (k_x/xbox, k_y/ybox, k_z/zbox)$ である。パラメータ η は長さの次元を持ち、上記3項間のバランスを決める。式3.5.3(2)-(4)の導出は、巻末の教科書(岡田、大沢、[2]、Allen & Tildesley, [6])を参照のこと。

実空間の和(式3.5.3(2))は、通常のクーロン力(式3.5.3(1))に補誤差関数 erfc が掛かった形である。 erfc は収束が速い関数なので、式3.5.3(2)は、素早く収束する。よって、カットオフの場合と同様にペアリストを使って、ある切断半径内の原子からの寄与のみを計算すればよい(ただし、3.5.2で説明した residue-based cutoffではなく、直接、原子間距離に切断半径を適用する)。PEACHの汎用機バージョンでは、エワルドの実空間はVDW力と同じサブルーチンで一緒に計算している。エワルド法は、物理学的にはカットオフを使わないで遠距離からのクーロン力の寄与を計算できるが、プログラム上はカットオフを使うわけである。波数空間の和(式2.5.3(3))にはガウス型関数 $\exp(-x^2)$ が掛かっているのでこちらも収束が速い。

ポテンシャル3.5.3(2)と(3)を距離で微分すると力になる(3.5.3(5), (6))。ポテンシャル3.5.3(4)は定数だから力には関係ない。

$$\mathbf{F}_{real} = \frac{q_i}{\eta^3} \sum_n \sum_j' q_j \left[\frac{\operatorname{erfc} \left(\frac{r_{ijn}}{\eta} \right)}{\left(\frac{r_{ijn}}{\eta} \right)^3} + \frac{2}{\sqrt{\pi} \left(\frac{r_{ijn}}{\eta} \right)^2} \exp \left(- \left(\frac{r_{ijn}}{\eta} \right)^2 \right) \right] \mathbf{r}_{ijn} \quad 3.5.3(5)$$

$$\mathbf{F}_{wave} = \frac{2q_i}{L^3} \sum_{\mathbf{K}} \frac{\exp(-\pi^2 \eta^2 \mathbf{K}^2)}{\mathbf{K}^2} \times \left(\sin 2\pi \mathbf{K} \cdot \mathbf{r}_i \sum_j q_j \cos 2\pi \mathbf{K} \cdot \mathbf{r}_j - \cos 2\pi \mathbf{K} \cdot \mathbf{r}_i \sum_j q_j \sin 2\pi \mathbf{K} \cdot \mathbf{r}_j \right) \mathbf{K} \quad 3.5.3(6)$$

なお、Komeiji et al. [12] では、立方体のみを扱っているため、 \mathbf{k} と \mathbf{n} で表示してあるが、ここでは、直方体も扱えるように、 \mathbf{K} と \mathbf{N} で表示した。

3.5.4 周期境界におけるエワルド和 –条件決め–

3.5.3 でエワルド和のポテンシャルと力の表式 (3.5.3(2)-(6)) を示した。これらの級数の計算では、精度を上げるためには可能な限り多くの項を取るべきである。しかし、計算時間は当然、項数が上がれば増大してしまう。よって、実空間の切断距離 R_{cut} 、波数ベクトル \mathbf{k} の絶対値の上限 k_{max} 、収束パラメータ η を調節して、実現可能な速度範囲で、要求される計算精度を達成する必要がある。

上記3つのパラメータの調節の仕方は様々であるが、現時点では、Fincham[27] の理論的な決め方を推奨する。概略は以下のとおり。なお、ここでは一辺が L の立方体を扱い、エネルギーについてのみ議論する。また、実空間の和 3.5.3(2) と (5) においては、最近接粒子 (2.2.1 節) のみを考慮することにするので、 r_{ijn} ではなく r_{ij} と表記する。

目的は、エネルギーの誤差が $\varepsilon (= \exp(-p))$ になるような、パラメータの組を求めることである。

実空間と波数空間 (3.5.3 (2),(3)) の収束性は、それぞれ、 $\text{erfc}(r_{ij}/\eta)$ と $\exp(-(\pi\eta k/L)^2)$ によって概ね決められる。補誤差関数 $\text{erfc}(x)$ は x が大きいところでは $\exp(-x^2)$ になる。よって、 $r_{ij} = R$ と $k = k_{max}$ において、以下の式が成り立てば、実空間も波数空間も の誤差以下で求められることになる。

$$p = \left(R_{cut} / \eta \right)^2 = \left(\pi \eta k_{max} / L \right)^2 \quad 3.5.4(1)$$

上式より、

$$k_{max} = LR_{cut} / \pi \eta^2 \quad 3.5.4(2)$$

$$\eta = R_{cut} / \sqrt{p} \quad 3.5.4(3)$$

の二つが導かれる。

Fincham[27] は、普通の計算機を利用することを前提に、3.5.4(1)-(3) をもとに、計算時間を求め、それを最小にするようなパラメータの組を求めている。しかし、以下では GRAPE を利用する場合に限定して議論を進める (Komeiji & Uebayasi [26])。

GRAPE を利用する場合、実空間は自動的に箱の辺の長さ L の半分になる。よって

$$R_{cut} = L / 2 \quad 3.5.4(4)$$

となる。これを式 3.5.4(2) と (3) に代入すれば、

$$\eta = L / 2\sqrt{p} \quad 3.5.4(5)$$

$$k_{max} = 2p / \pi \quad 3.5.4(6)$$

という式が得られる。つまり、 η と k_{max} は自動的に決まる。

以上の議論では、力については考慮していない。しかし、現実に適用すると、上記の方法で決めたパラメータが、力でもポテンシャルでも経験的に選んだ最適値とよく一致する (Komeiji & Uebayasi [26])

なお、エwald法でクーロン力を計算する場合、

「生体分子は本当は周期的に並んでいるわけではないので、波数空間の項は沢山取ってはいけない。」

という議論があるが、これは間違った認識である。この表現では、波数ベクトル \mathbf{K} を沢山とることが、Figure 2.2(4)の虚像を沢山とることに対応するかのようと思われるが、それは誤解である。エwald法では、 \mathbf{K} を多く取るとはフーリエ成分を多く取ることに対応するので、力やポテンシャルがより精密に計算されることを意味する。周期条件による問題点については、できるだけ高精度のエwald計算した上で改めて解析すべきものであり、低精度のエwald計算すれば周期依存性が避けられるというものではない。

3.5.5 非共有結合力の計算法について

以上、クーロン力の計算法について、いくつか例を挙げた。PEACHに導入されたこれらのアルゴリズムは、基本的なものであって、これ以外に様々な高速計算法が考案され、実用化されている。たとえば、周期境界を利用する場合の高速計算法は、上記のオリジナルなエwald法だけではない。波数空間部分を高速計算できる Particle Mesh Ewald 法 (Darden et al. [16]) が生体分子、特に核酸のMDに適用され、大きな成果を収めている。さらに、Particle-Particle Particle-Cell 法 (PPPC 法、Saito [15]) や Fast Multipole Method (FMM, Greengard, [29]) など、多重極子展開を利用した高速化法もある。これらの高速化法の詳細については原著論文を参照のこと。

以下では、エwald法も含めた、これらの高速化法の共通原理を簡単にまとめておく。

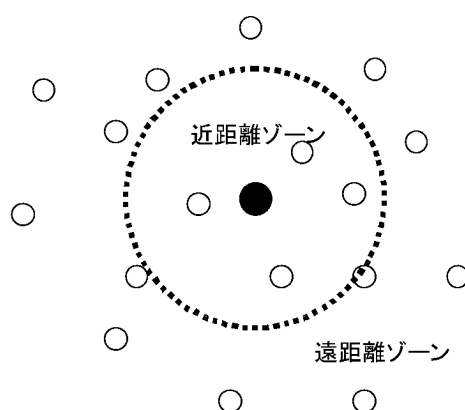


Figure 3.5.5. クーロン力の計算

クーロン力を切断せずに計算する場合、非周期境界の場合は式 3.5.1(1)、周期境界の場合は式 3.5.3(1) を計算することになる。粒子数を N とすれば、前者は $O(N^2)$ 、後者は $O(MN^2)$ (M は虚像の数で、理想的には) の計算量が必要となる。このような計算をまともにしていたら、

計算機がどんなに速くても足りない。そこで、Figure 3.5.5のように、クーロン力を近距離の力と、遠距離からの力にわけて計算することになる。

カットオフ法は、遠距離からの寄与を無視した場合である。また、エワルド法の場合、(数学的に正しい表現ではないが)実空間の計算が近距離力、波数空間が遠距離力に対応する。

近距離の寄与の計算は、決められた距離内に存在する全ての原子間について、クーロン力の直接和を計算する。普通はVDW力と一緒に計算するため、最低8、余裕を持って15以内の距離にある原子ペアについて力を計算する必要がある。時間刻みは、1-2 fs程度であり、これ以上長くできない。

一方、遠距離の粒子からの寄与の計算は、様々な工夫がされている。フーリエ変換(エワルド法)や多重極子展開(PPPC法やFMM法)を利用して空間的にまとめることで、計算量を短縮できる。また、遠距離力は近距離力に比較して長い時間刻み幅をとることができる(多時間刻み幅法、4.2節参照)。空間的にまとめる方法と時間刻みを長くする方法は併用できるので、遠距離からの力の計算時間は精度を下げずにかなり短縮することができる。もちろん近似としての限界はある(Bishop et al. [30])。

近距離粒子間の力は直接和で計算せざるを得ず、高速化は難しい(もちろん、並列化などでの高速計算は可能である)よって、ある系を8カットオフでMDをした場合と、高速化手法で遠距離力を正確に計算した場合では、原理的に、後者は前者の計算速度を越えることができない。また、高速化アルゴリズムにより、計算時間は $O(N^{3/2})$ 、 $O(N\log N)$ 、 $O(N)$ 等になるが、そのオーダーの係数は一般に大きく、粒子数が少ないと効果がない。高速化アルゴリズムを使うのならば、自分の扱いたい系の大きさと必要な精度を考えて、計算速度と精度の上昇が見込めるようなアルゴリズムを選択すべきである。

3.6 1-4非共有結合力(1-4VDW、1-4Coulomb)

式2(1)に出てくる1-4VDW力と1-4Coulomb力は1-4非共有結合力(1-4 Nonbonded interaction)と呼ばれ、Figure 3で1-4の位置にある原子間に適用される。通常のVDW力やクーロン力に較べて、 SC_{14VDW} と SC_{14ELC} の係数で割って、やや力を弱める(AMBERでは、SCNB、SCEEと呼ばれる)。これらの値は使う力場によって違うので、Table 3.6にその推奨値を挙げておく。

なお、AMBER84とAMBER86の力場で使われる水素結合について、1-4水素結合ペアについては、式3.4(3)の10-12ポテンシャルは計算せず、代わりに1-4VDW力を計算する。

Table 3.6. 1-4非共有結合のスケールファクター

力場	AMBER84 [22]	AMBER86 [23]	AMBER95 [24]	OPLS [25]
SC_{14VDW}	2.0	2.0	2.0	8.0
SC_{14ELC}	2.0	2.0	1.2	2.0

3.7 排除原子の扱い方 (Excluded atoms)

Figure 3 の、1-2 ペア、1-3 ペア、1-4 ペアは、Excluded pairs と呼ばれ、3.4 節と 3.5 節で説明した非共有結合力は計算しない。この Excluded pairs の考慮の仕方を説明する。

3.7.1 VDW 力およびクーロン力の直接和の場合

普通の汎用計算機でペアリストを使う場合は、排除原子をあらかじめペアリストから抜いてしまう。そうすれば、排除原子からの力やポテンシャルは計算しなくてすむ。

MD-GRAPE を使う場合は、排除原子の分まで力やポテンシャルを計算するしかない。そこで、排除原子による力やポテンシャルを計算して、GRAPE で計算した力やポテンシャルの値から引いて補正する。この方法は計算精度を考えると望ましくはないが、現時点の MD-GRAPE の仕様では、これ以外によい方法がない。また、今のところこの方法で特に問題点は出ていない。

3.7.2 クーロン力のエワルド和の場合

エワルド和の場合は、波数空間は、式 3.5.3(3)(6) をそのまま計算する。実空間は、排除粒子を除いたペアリストを用いて式 3.5.3 (2)(5) を計算する。その値から以下の値を引けばよい。

$$U = \frac{1}{2} \sum_{(i,j):excluded} \frac{q_i q_j}{r_{ij}} \operatorname{erf} \left(\frac{r_{ij}}{\eta} \right) \quad 3.7.2(1)$$

ただし、 $\operatorname{erf}(x)=1-\operatorname{erfc}(x)$ である。

GRAPE を使う場合は、排除原子の寄与まで含んだ形でエネルギーと力を計算し (式 3.5.3(2)-(6))、そこから、排除原子によるエネルギーと力を直接和 (式 3.5.1(1), (2)) の形で計算して引き算をする。

4 時間積分

Figure 1 ではニュートンの運動方程式の時間積分 (3) を $R+V\Delta t$ R と書いてあるが、実際には、もっと計算精度のよい方法が使われている。今回は、PEACH に導入された速度 Verlet 法と、その応用である多時間刻み幅法 RESPA を紹介する。他の様々な方法については、巻末に挙げた教科書を参照して頂きたい。

4.1 速度 Verlet 法 (Velocity-Verlet method)

速度 Verlet 法は、Swope et al. [31] によって提案され、より一般的な導出は Tuckerman et al. [32] によってなされた。以下に疑似コードを挙げる。

```

do i = 1, nstep
  v = v + Δt/2 * f/m
  r = r + Δt * v
  call calforce(f)    ! 力を計算する
  v = v + Δt /2 * f/m
end do

```

このように、単純な積分法だが、安定性が高い。また、次項の RESPA に拡張できる。

4.2 多時間刻み幅法 RESPA (REference System Propagator Algorithm)

生体分子や有機分子では、3章で説明したように、何種類もの力が働いている。このような系の MD を行う場合、最も速い運動に時間幅を合わせ必要がある。具体的には、水素原子の伸縮運動が 20 fs 程度の周期を持っているので、それより 2 桁下の 0.2 fs ぐらいの時間幅を使わなければならない。しかし、クーロン力や VDW 力のような柔らかい力はそんなに速く変化するわけではなく、1-2 fs 程度の時間幅でも十分で、0.2 fs では不経済である。また 3.5.5 節で説明した長距離のクーロン力は、もう少し長い時間幅でも構わない。そこで、力の種類に合わせて時間幅を変えるのが実用的である。このような方法を、多時間刻み幅法 (Multiple time step method) と呼び、いくつかのアルゴリズムが提案されてきた。たとえば、タンパク質などで今でも時々使われる Twin-range 法 (Komeiji et al. [14]) も、この一種である。

従来の多時間刻み幅法は、時間反転性 (time reversibility) * が保証されないことが多く、安定性に問題があった。だが、Tuckerman et al. [32] により、時間反転性を満たす積分法 (Integrator) の一般的導出法が示された (RESPA)

RESPA の導出法は複雑なので省略するが、得られたアルゴリズム自体は、単純でわかりやすい。以下に疑似コードを挙げる。力は、 f_h 、 f_m 、 f_s (hard, medium, soft) の三つに分けてある。

```

do i = 1, nstep
  v = v + Δts/2 * fs/m
  do j1 = 1, L
    v = v + Δtm/2 * fm/m
    do j2 = 1, K
      v = v + Δth/2 * fh/m
      r = r + Δth * v
      call calforce(fh)    ! 固い力の計算
      v = v + Δth/2 * fh/m
    end do
    call calforce(fm)    ! 中間の力の計算
    v = v + Δtm/2 * fm/m
  end do
  call calforce(fs)    ! 柔らかい力の計算
  v = v + Δts/2 * fs/m
end do

```

柔らかい力 f_s は 1 ステップに一回だけ、中間の力 f_m はより多い回数で、固い力はさらに頻繁に f_h 、というように、それぞれに対して、違う時間刻み (Δt_s , Δt_m , Δt_h) を与えている。この方法により、一律 Δt_h の時間刻みを与える場合に比べて、計算精度をあまり下げずに速度を数倍にすることができる。具体例は、Komeiji et al. [26] および引用文献を参照のこと。

それぞれの時間刻み (Δt_s , Δt_m , Δt_h) をどう与えるかは、MD の対象や力場により変わるので、経験的に決める必要がある。その場合、時間刻みの組み合わせをいろいろ変えてミクロカノニカル MD (5.1 節参照) を行い 4.2(1) または (2) (全エネルギー E の初期値 E_{init} からのずれの時間平均 $\langle \rangle$ の相対値) の値を調べる。

$$\frac{\sqrt{\langle (E_{init} - E)^2 \rangle}}{|E_{init}|} \quad 4.2(1)$$

$$\frac{\langle |E_{init} - E| \rangle}{|E_{init}|} \quad 4.2(2)$$

どの程度の精度が必要かは一概には言えないが、上の式で定義した値が、0.003 付近になるぐらいが上限の目安とされている。

* 時間反転性とは、ある時刻の粒子系の速度は完全に逆向きにした場合、計算機の精度が無限大ならば、最初の時刻の位置にぴったり戻ることを言う。時間反転性を満たす積分法のほうが、そうでないものより安定である。

** RESPA は、正確には、Tuckerman らの方法によって導出される任意のアルゴリズムのことを指す。その中で、上述のアルゴリズムは、drRESPA (double reversible REference System Propagator Algorithm) と呼ぶ。

5 アンサンブル (温度制御など)

MD では温度制御や圧力制御を行うことにより (Figure 1(4))、様々な統計的集合 (アンサンブル) を実現させることができる。生体分子の MD では、NVE (粒子数・体積・エネルギー一定、ミクロカノニカルアンサンブル)、NTV (粒子数・温度・体積一定、カノニカルアンサンブル)、NTP (粒子数・温度・圧力一定、圧力アンサンブル) のどれかが使われる。PEACH には NTP アンサンブルは導入していないので、ここでは NVE と NTV のみ説明する。

5.1 NVE アンサンブル (ミクロカノニカル)

よく知られているように、孤立系の全エネルギー (Hamiltonian、 H) は保存される。

$$H = \sum \frac{1}{2}mv^2 + U \quad 5.1(1)$$

温度制御や圧力制御等をしないで MD を行うと、このミクロカノニカルアンサンブルが実現する。その意味で最も基本的なアンサンブルであり、Classical MD と呼ばれたこともある。ただし、最近では、Classical MD は「Quantum MD ではなく古典力学の範囲での MD である」ということを示す場合が多いので、ミクロカノニカル MD と呼ぶことを推奨する。

新しく MD のプログラムを開発する際は、まずマイクロカノニカル MD でエネルギーが保存されるかどうかでデバッグする。また、4 章で説明した積分法の安定性などを調べるときにも使われる。全エネルギーが保存されるのだから、その揺らぎをチェックして、積分法の良否を調べることができる (式 4.2(1) または (2) を利用)。

マイクロカノニカル MD は、粒子の温度制御 (= 速度のスケーリング) をしないので、分子の自然なダイナミクスを調べるときには理想的なアンサンブルと言われている。ただし、マイクロカノニカル MD は系が本当に平衡化していないと温度が不安定になる。特にクーロン力をカットオフした場合、温度が急上昇してしまうことが多い。

5.2 NVT アンサンブル (Woodcock)

温度一定にしたいときに、最も簡単なのは、Woodcock による速度スケーリング法である (Woodcock [33])。温度 T と粒子の速度の関係は

$$\frac{3NkT}{2} = \sum \frac{mv^2}{2} \quad 5.2(1)$$

の関係にあるので、各ステップで、上式を満たすように速度 v をスケーリングすればよい。

この方法は精度が悪いので今日ではあまり用いられない。ただし、MD の開始時に、低温で数 10 ステップシミュレーションをしたりする時には便利である。

5.3 NVT アンサンブル (Nose-Hoover)

上記の Woodcock 法は計算精度が悪い。また生体分子の分野で広く使われている Berendsen et al. [34] の定温アルゴリズムは得られるアンサンブルの理論的性質がよく解っていない。そこで、PEACH には、カノニカルアンサンブルが実現することが理論的に保証されている Nose-Hoover の定温アルゴリズムを導入した。Nose [35] に Nose-Hoover をはじめ、定温アルゴリズムの詳しい解説がある。また、田中、山本の教科書 [4] にも解りやすい解説が載っている。

Nose の方法にはいくつかのバリエーションがあるが、基本的には、以下のような原理である。体系の熱浴として自由度 1 の仮想的な粒子 (座標 η 、運動量 p_η 、質量 Q) を考える。ただし、これらの物理量の次元は、距離、運動量、質量、ではない。この粒子が摩擦力を通して系に熱を供給したり奪ったりして、系の温度を一定に保つ。この仮想粒子を含めた系全体のハミルトニアン H' は、

$$H' = \sum_{i=1}^N \frac{m_i \mathbf{v}_i^2}{2} + U(\mathbf{r}_0 \dots \mathbf{r}_N) + \frac{p_\eta^2}{2Q} + 3NkT\eta \quad 5.3(1)$$

で表わされるが、この仮想的なハミルトニアンは保存される、つまりマイクロカノニカル分布に従う。一方、現実の粒子の部分 $E = \sum 1/2 m v^2 + U$ は、統計力学でいうカノニカル分布に従うことが証明されている [35]。このように、系を仮想粒子まで含めて考えている、という意味で、拡張系の方法 (Extended system method) と呼ばれることもある。

PEACH では、大筋は Tuckerman et al. [32] に従って、以下のようなアルゴリズムを使っている。

```

do i = 1, nstep
  pη = pη + Δts/2 * Fη
  η = η + Δts * pη / Q
  v = v * exp(-Δts/2 * pη / Q)

  ***** begin V-Verlet *****
  do i = 1, nstep
    v = v + Δt/2 * f/m
    r = r + Δt * v
    call calforce(f) ! calculate force
    v = v + Δt / 2 * f/m
  end do
  ***** end V-Verlet *****

  v = v * exp(-Δts/2 * pη / Q)
  call calfehta(Fη) ! calculate Fη
                    ! (=Σ v2/m - 3NkT)
  pη = pη + Δts/2 * Fη
end do

```

仮想粒子の質量 Q は $Q=3NkT\tau^2$ である。ここでは、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。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。昇温過程などではを短くして温度を素早く調節すべきだが、サンプリングの過程では、は長めにして、系の自然なダイナミクスを調べるようにすべきである。また、温度を効率よく制御するには、まずマイクロカノニカルMDを行い、その温度変化の周期とのオーダーを合わせるとよい。理想的には、平衡に達した時点でマイクロカノニカルMDに切り替えてサンプリングを行うことが望ましい。

6 トrajekトリーの解析方法

MDのトラjekトリーから何をどう引き出すかについて (Figure 1(5)), まず、基本的なことを述べる。ある物理量 A をMDで求めたいとしよう。 A はエネルギーでも $RMSD$ でも回転半径でもなんでもよい。 Figure 6(1) に示すようにMDの過程は平衡化とサンプリングの二つに分けられる。 A の時間変化をプロットしてみて、値が安定に達してからサンプリングを行い A の平均と揺らぎを求める。研究発表では、まず Figure 6 のようなグラフを出し、さらに A の平均と揺らぎを表にして示すのが基本である。

Figure 6(1) はあくまで理想であって、そうは行かない場合もある。例えば、構造転移などにより、 Figure 6(2) のようになる場合もある。また、(3) のように、大体安定に達したように見えるが、よくみると少しずつ値が変わっていく場合もある ($RMSD$ などによく見られる現象)。こういった場合は、個々のケースに応じて対処する必要がある。

理想的には、平衡化もサンプリングも長ければ長い程よいわけだが、現実にはそうもいかな

い。また、求めたい物理量の種類により、平衡化に掛かる時間もサンプリングに必要な時間も違う。一般的に、ある物理量を求めるのに必要なシミュレーションの時間は、
 平均 < ゆらぎ < ゆらぎの相関
 と考えてよい。

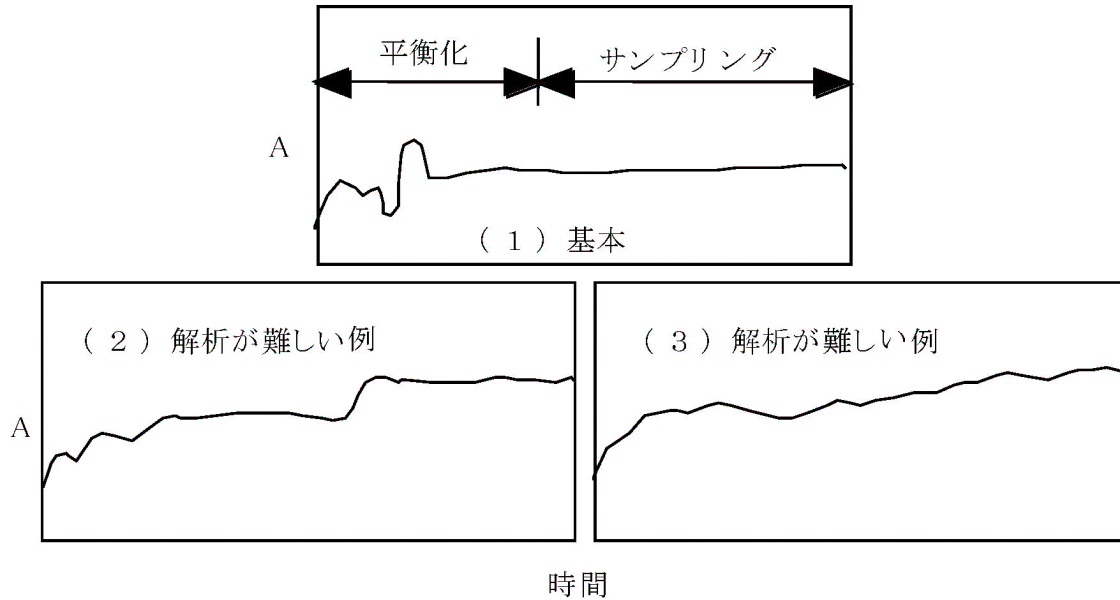


Figure 6. MDによる物理量の求め方

MDのトラジェクトリーから得られる情報は様々であり、全てを記すことはできない。巻末の教科書や文献を参照して欲しい。PEACHに導入した解析方法については Komeiji et al. [14, 26, 28, 36] に記述しておいたが、重要なものについて説明する。

以下の説明では、ある物理量 A の時間平均を $\langle A \rangle$ と書く。すなわち、

$$\langle A \rangle = \frac{1}{N_{step}} \sum_{j=1}^{N_{step}} A_j \quad (6.1)$$

である。ただし、添字 j は時間ステップを示し、 N_{step} は全時間ステップ数である。また、系の粒子数を N 、個々の粒子 i の座標、速度、質量は、 \mathbf{r}_i 、 \mathbf{v}_i 、 m_i であらわす。

6.1 エネルギーに関する物理量

6.1.1 エネルギーの平均、ゆらぎ

全エネルギー、ポテンシャルエネルギー、運動エネルギーなどのエネルギーについて（まとめて E と書く）その平均と揺らぎ（根平均自乗、RMS、Root Mean Square）を解析して、MDの精度を較べることが多い。

$$\begin{aligned} \text{平均エネルギー} &: \langle E \rangle \\ \text{エネルギーの揺らぎ} &: \sqrt{\langle (E - \langle E \rangle)^2 \rangle} \end{aligned}$$

NVE アンサンブルの場合、全エネルギーが保存量なので、揺らぎが小さい (5.1 節)。Nose の定温 MD では全エネルギーではなく Nose Hamiltonian が保存量である (5.3 節)。どちらの場合も、ポテンシャルエネルギーや運動エネルギーは保存量ではないが、十分時間を掛ければ平均値もゆらぎもほぼ一定になる。

6.1.2 温度

一般に、ある系の温度 T は、

$$T = \frac{1}{N_{df}k} \sum_{i=1}^N m_i \langle \mathbf{v}_i^2 \rangle = \frac{2}{N_{df}k} \langle E_{kin} \rangle \quad 6.1.2(1)$$

で定義される。ただし、 k はボルツマン定数。 N_{df} は系の自由度で、通常は $3N$ に等しい。系に束縛を掛けたり (たとえば、重心運動を止めたり)、Belly を使ったりする場合、凍結した自由度の数を $3N$ から引く。

PEACH では、温度は、溶質 (タンパクや核酸など)、イオン、溶媒のそれぞれについて、個別の計算して出力される。さらに、解析プログラムを使えば、残基毎の温度を計算できる。個別の温度は、

$$T_{in} = \frac{1}{(3n-3)k} \sum_{i=i_1}^{i_2} m_i \langle \mathbf{v}_i^2 \rangle \quad \text{内部} \quad 6.1.2(2)$$

$$T_G = \frac{1}{3k} M_G \langle \mathbf{V}_G^2 \rangle \quad \text{重心} \quad 6.1.2(3)$$

として、内部運動と重心運動それぞれに関する温度を分けて計算している。上式では、計算したい残基が原子 i_1 に始まり i_2 に終わり、原子数が $n (= i_2 - i_1 + 1)$ としている。 M_G は残基の質量、 \mathbf{V}_G は残基の重心の速度である。計算している単位がアミノ酸残基のように小さい場合、 T_G と T_{in} は全体の温度 T と一致する。だが、溶質全体や溶媒全体のような大きい単位について計算した場合、 T_{in} は T と一致するが、 T_G はまず一致しない。また、ナトリウムイオンのような 1 原子分子の場合は、当然 T_G は T と一致するが、 T_{in} は定義できない。

なお、上記で T と一致するというのは、あくまで、系が充分平衡に達し、サンプリング時間も長くたった場合であり、しかも系に不自然なノイズ等が入らない場合のことである。カットオフをしたりすれば不均一な温度分布が現われる場合もある (Oda et al. [37] など)。そうでなくとも、サンプリング時間が短ければ温度が不均一に見えることもある。

6.2 分子構造とダイナミクス

次に、溶質の構造やダイナミクスに関する物理量の求め方について説明する。なお、2 つの構造 (\mathbf{R} と \mathbf{r} とする) を重ね合わせるが必要な場合、

- (1) ふたつの構造の重心の位置を合わせ、
 - (2) さらに最小自乗法で重ね合わせる (Least Square fit)
- という方法をとる。現在のところ、McLachlan [38] の方法により、

$$w = \sum_i m_i (\mathbf{R}_i - \mathbf{r}_i)^2 \quad 6.2(1)$$

の値が最小になるように重ね合わせている。また、系全体でなく、一部分を選んで重ね合わせることもできる。

6.2.1 根平均二乗変位 (Root Mean Square Deviation, *RMSD*)

分子トポロジーが同じである二つの構造の差は、以下の式で *RMSD* (Root Mean Square Deviation) として計算する。一般に、アミノ酸残基や核酸のヌクレオチド単位で計算することが多い。その場合、

$$RMSD = \frac{1}{N} \sum_{i=1}^N (\mathbf{R}_i - \mathbf{r}_i)^2 \quad 6.2.1(1)$$

として計算する。残基毎、主鎖毎、側鎖毎、に分けて出すこともできる。原子の質量で重みを付ける場合もあるが、PEACHでは、今のところそのオプションはない。

RMSD は、結晶構造からの MD 構造のずれを調べる場合などに使う。

6.2.2 溶媒露出面積 (Accessible Surface Area, *ASA*)

溶質が、どの程度溶媒に露出しているかを測る手段として、(solvent) Accessible Surface Area (*ASA*) を計算する (Shrake & Ruply [39])。これは、溶媒分子をある半径 (1.4 ぐらいに設定することが多い) の球と仮定して、溶質の周囲を転がした場合に、中心が描く軌跡の曲面の面積である (Figure 6.2.2)

6.2.3 回転半径

以下の定義による量 R_g を回転半径 (radius of gyration) と呼ぶ。

$$R_g = \sqrt{\frac{\sum_i m_i r_i^2}{\sum_i m_i}} \quad 6.2.3(1)$$

ただし、 r_i は、原子 i の分子の重心からの距離、 m_i は質量である。

回転半径は、分子中の質量の拡がり具合を示すパラメータである。回転半径は、タンパク質のフォールディングやアンフォールディングのシミュレーションの際に、そのほどこけ具合を示すパラメータとしてよく使われる。

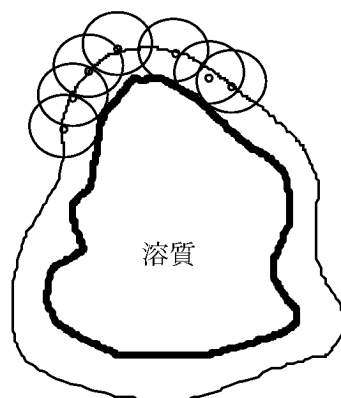


Figure 6.2.2. ASA

6.2.4 根平均二乗揺らぎ (Root Mean Square Fluctuation, *RMSF*)

平衡状態に達してからの MD のトラジェクトリーの構造の揺らぎは、次のように *RMSF* (Root Mean Square Fluctuation) として計算する。なお、アミノ酸残基やヌクレオチド単位で計算することが多いので、残基 j の *RMSF* の形で表わす。

$$RMSF(j) = \sqrt{\langle (\mathbf{R}_j - \langle \mathbf{R}_j \rangle)^2 \rangle} \quad 6.2.4(1)$$

ここで、残基 j の座標 \mathbf{R}_j は、単なる残基中の全原子の平均でもよいし、残基の重心でもよい。主鎖や側鎖も選ぶことができるし、あるいは、 C_α のような、特定の原子でもよい。また、残基の原子すべての揺らぎの自乗平均を取ることもある。

RMSF は、次式で X 線結晶解析の等方性温度因子 (B) と比較できる (Brooks et al. [7])。

$$RMSF(j) = \sqrt{\frac{3B}{8\pi^2}}, \quad B = \frac{8\pi^2}{3} (RMSF(j))^2 \quad 6.2.4(2)$$

なお、残基 j の座標については、以下の *DM*、*DFM*、*DCCM* でも *RMSF* の場合と同様に選ぶことができる。

6.2.5 Distance Map (*DM*)

DM (Distance Map) は、残基間 (あるいは原子間) の距離に基づいて、タンパク質などのドメイン構造を解析する方法である。二つの残基間の距離 R_{ij} を以下のように定義する。

$$R_{ij} = |\langle \mathbf{R}_i \rangle - \langle \mathbf{R}_j \rangle| \quad 6.2.5(1)$$

この値に基づいて描いた Figure 6.2.7 のような二次元の地図を *DM* と呼ぶ。

この図は、結晶構造や NMR 構造に基づいて描いたものと通常は大差なく、その意味ではわざわざ MD をして描く程のものではない。だが、以下の *DFM* を解析する場合のレファレンスとなる。

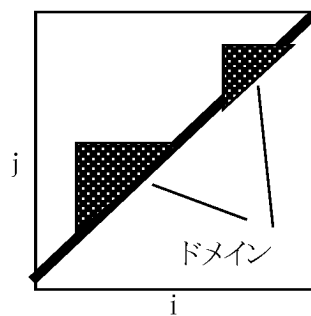


Figure 6.2.5. DM

6.2.6 Distance Fluctuation Map (DFM)

二つの残基間の距離の揺らぎ D_{ij} を以下のように定義する。

$$D_{ij} = \sqrt{\langle (r_{ij} - \langle r_{ij} \rangle)^2 \rangle} \quad 6.2.6(1)$$

この D_{ij} を Figure 6.2.5 と同様に二次元等高図にしたものを Distance Fluctuation Map (DFM, Komeiji et al. [28]) と呼ぶ (DFM は筆者の造語である)。

上記の DM で決めたドメインの構造が固いか柔らかいかを、DFM で解析することができる。

6.2.7 動的相関図 (Dynamical Cross Correlation Map, DCCM)

二つの残基の揺らぎの相関 C_{ij} を以下のように定義する。

$$C_{ij} = \frac{\langle \Delta \mathbf{r}_i \bullet \Delta \mathbf{r}_j \rangle}{\sqrt{\langle \Delta \mathbf{r}_i^2 \rangle \langle \Delta \mathbf{r}_j^2 \rangle}} \quad 6.2.7(1)$$

ただし、 $\Delta \mathbf{r}_i$ は残基 i の平均構造からの揺らぎをしめす。 C_{ij} は、統計学上の相関係数に対応し、 $-1 \leq C_{ij} \leq 1$ の値を取り得る。 $+1$ の場合は、 i と j の揺らぎが完全に相関していることを示し、 0 の場合は相関なし、 -1 の場合は負の相関があることになる (Figure 6.2.7)。

この C_{ij} を Dynamical Cross Correlation (動的相関) と呼び、これを Figure 6.2.5 と同様に等高線図にしたものを DCCM (Dynamical Cross Correlation Map) と呼ぶ。具体例は、Harte et al. [40]、Komeiji et al. [14, 28] など。

DCCM は、分子内の協同運動 (collective motion) やドメイン構造を解析するために導入されたものである (Harte et al. [40] および引用文献参照)。しかし、問題点も指摘されている。DCCM では式 6.2.9 からわかるように、2 つの残基の平行な協同運動しか観測されない。また、中心付近のコアドメイン内の協同運動は観測できない。さらに、揺らぎの相関であるため収束が遅く、再現性が低い (Hunenberger et al. [41])。私見としては、DCCM でタンパク質全体の協同運動やドメイン構造の議論はしないほうがよい。だが、タンパク質と小さな分子の相互作用の解析なら、収束が速いので充分実用になると考えられる (Komeiji et al. [28])。

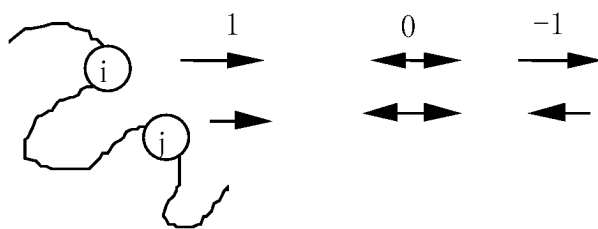


Figure 6.2.7. 動的相関

6.2.8 Essential Dynamics および Principal Component Analysis

DCCM よりも詳しく分子内協同運動を調べたい場合、Essential Dynamics (*ED*, Amadei et al. [42]) や Principal Component Analysis (*PCA*, Kitao et al. [43]) などの方法が有効である。

ED と *PCA* に共通するアイデアは、原子の座標の揺らぎから、独立な基準座標の揺らぎを求める、ということである。ここでは、まず、*PCA* について一通り説明する。

N 粒子系の MD を行った場合の座標の時間発展は、 x, y, z をそれぞれ独立に扱えば、 $r_1(t), r_2(t), \dots, r_{3N}(t)$ とあらわせる。重心運動と回転運動はあらかじめ除いておく。ここで、 $r_1(t) \dots r_{3N}(t)$ を、それぞれの平均座標からの揺らぎに変換し、さらに、質量の平方根で重み付けをする (mass-weighted coordinates)

$$R_i(t) = \sqrt{m_i}(r_i(t) - \langle r_i \rangle) \quad 6.2.8(1)$$

PCA は、この R_i ($i=1 \dots 3N$) を、独立な基準座標 q_i ($i=1 \dots 3N$) (Principal Component) に変換する手続きである。ここで、

$$\begin{aligned} \langle q_i(t)q_k(t) \rangle &= \lambda_j(j=k) \\ &= 0(j \neq k) \end{aligned} \quad 6.2.8(2)$$

を満たすものとする。係数の組 $a_{ij}(i=1, 3N, j=1, 3N)$ 、ただし

$$\sum_{j=1}^{3N} a_{ij}^2 = 1 \quad 6.2.8(3)$$

を用いて、 $R_i(t)$ を以下のように表わすことができる。

$$R_i(t) = \sum_{j=1}^{3N} a_{ij}q_j(t) \quad 6.2.8(4)$$

この q_j が、分子内協同運動を表す物理量である。6.2.8(2) からは、 q_j が調和振動子であると仮定しているように見えるが、さしあたって、そうである必要はない。

さて、6.2.8(2)-(4) を満たす、 q_j と a_{ij} の組は、以下の手続きで求めることができる。ベクトル \mathbf{R} 、 \mathbf{Q} 、行列 \mathbf{A} を定義する。 \mathbf{Q} と \mathbf{R} は時間の関数であることを強調して (t) をつけた。一方、 \mathbf{A} は時間に依存しない。

$$\mathbf{R}(t) = \begin{bmatrix} R_1 \\ \cdot \\ \cdot \\ R_{3N} \end{bmatrix} \quad 6.2.8(5)$$

$$\mathbf{Q}(t) = \begin{bmatrix} q_1 \\ \cdot \\ \cdot \\ q_{3N} \end{bmatrix} \quad 6.2.8(6)$$

$$\mathbf{A} = \begin{bmatrix} a_{11} \dots \dots \dots a_{1,3N} \\ \dots \dots a_{22} \\ \cdot \\ \cdot \\ a_{3N,1} \dots \dots \dots a_{3N,3N} \end{bmatrix} \quad 6.2.8(7)$$

MDで得られる結果は、 $\mathbf{R}(t)$ である。この $\mathbf{R}(t)$ より分散共分散行列 (variance-covariance matrix) を以下の式で求める。

$$\text{cov}\mathbf{R} = \langle \mathbf{R}\mathbf{R}^\tau \rangle = \begin{bmatrix} \langle R_1 R_1 \rangle \dots \dots \dots \langle R_1 R_{3N} \rangle \\ \langle R_2 R_1 \rangle \dots \dots \dots \cdot \\ \cdot \\ \cdot \\ \langle R_{3N} R_1 \rangle \dots \dots \dots \langle R_{3N} R_{3N} \rangle \end{bmatrix} \quad 6.2.8(8)$$

$\text{cov}\mathbf{R}$ は対称行列であるが、これは、行列 \mathbf{A} を用いて以下のように対角化され、

$$\text{cov}\mathbf{R} = \mathbf{A}\mathbf{\Lambda}\mathbf{A}^\tau \quad 6.2.8(9)$$

さらに、対角行列 $\mathbf{\Lambda}$ は

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 \dots \dots \dots O \\ \dots \lambda_2 \dots \dots \dots \\ \dots \dots \lambda_3 \dots \dots \dots \\ \cdot \\ \cdot \\ O \dots \dots \dots \lambda_{3N} \end{bmatrix} = \begin{bmatrix} \langle q_1^2 \rangle \dots \dots \dots O \\ \dots \langle q_2^2 \rangle \dots \dots \dots \\ \cdot \\ \cdot \\ O \dots \dots \dots \langle q_{3N}^2 \rangle \end{bmatrix} \quad 6.2.8(10)$$

となっている (式 6.2.8(2) と (3) を用いた)

求めた正規直行行列 \mathbf{A} を用いれば、

$$\mathbf{Q}(t) = \mathbf{A}^\tau \mathbf{R}(t) \quad 6.2.8(11)$$

として、基準座標 (Principal component) の時間発展を求めることができる。また、行列 \mathbf{A} の列ベクトル

$$\mathbf{A}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \cdot \\ \cdot \\ a_{3Nj} \end{bmatrix} \quad 6.2.8(11)$$

を、 j 番目の Principal Axis と呼ぶ。これと、Principal component q_j は、

$$q_j(t) = \mathbf{A}_j^T \mathbf{R}(t) = a_{1j}R_1(t) + a_{2j}R_2(t) + \dots + a_{3Nj}R_{3N}(t) \quad 6.2.8(12)$$

の関係にある。

以上が PCA のあらましである。ED の場合は式 6.2.8(1) で、質量を掛けない形の座標を使う以外は、まったく同様の操作を行う。ED と PCA は、よく似ているが、PCA のほうが物理的な意味付けがしやすいという利点がある。以下で説明する。

これまでは、Principal component q は調和振動子であると仮定する必要はなかった。しかし、これ以降は調和振動子であると仮定しよう。式 6.2.8(1) で質量を掛けた形にしてあるため、 $\langle q_i^2 \rangle$ は調和振動子の $m \langle r^2 \rangle$ に対応するとみなすことができる。この振動子の有効周期を ω_{eff} とおけば、平均の運動エネルギーは

$$\langle K \rangle = \frac{\omega_{eff}^2}{2} \langle q_i^2 \rangle \quad 6.2.8(13)$$

となる。一方、温度 T における等分配則により、自由度 1 の調和振動子の持つ平均運動エネルギーはボルツマン定数 k を用いて、

$$\langle K \rangle = \frac{kT}{2} \quad 6.2.8(14)$$

と表わすことができる。式 (13) と (14) より、 i 番目の Principal component の持つ有効振動数は、

$$\omega_{effi} = \sqrt{\frac{kT}{\langle q_i^2 \rangle}} = \sqrt{\frac{kT}{\lambda_i}} \quad 6.2.8(15)$$

として求めることができる。これを利用して、スペクトル密度などを調べることができる。なお、調和振動子であるとした仮定が正しいかどうかは、それぞれの Principal Component のトラジェクトリーから判断できる。一般に、値が大きい component は、調和振動から外れることが多いが、小さくなると調和振動性が強くなる。

PCA や ED を用いることにより、分子内の様々な時間領域の協同運動を抜き出して解析することが可能である。具体例は Hayward & Go [44] のレビューおよび引用文献を参照のこと。

なお、PCA も ED も、ともに揺らぎとその相関に関する量であるため、やはり、収束性は悪い。MD のトラジェクトリーの中のどの時間領域を使うかで、時間に依存しないはずの λ が大きく変わってしまうことも指摘されている (Balsera et al.[45])。だからと言って、使っていけないということにはならないが、結果の解釈には注意が必要である。

7 おわりに

以上、生体高分子のMDシミュレーションについて、基本的なアルゴリズムとその原理を紹介した。生体分子へのMDの応用範囲は広がっている。今回は紹介できなかったが、タンパク質設計に有用な自由エネルギー計算、フォールディングのシミュレーションで注目されているマルチカノニカルMD、三次元構造予測法のホモロジーモデリング、など、枚挙に暇がない。これらの方法に関しては、PEACHに導入した時点で、改めて解説したい。

プログラム PEACH の開発に関して、多くの先生方にご指導をいただきました。お名前を挙げて感謝いたします。東京理科大学・山登一郎教授、弘前大学・斎藤稔教授、大正製薬・宮川博夫博士、物質工学工業技術研究所・三上益弘博士、電子技術総合研究所・横山浩博士、長谷泉博士。また、多体問題専用計算機 MD-GRAPE は、東京大学の杉本大一郎教授（現・放送大）、泰地真弘人博士（現・統計数理研）、福重俊幸博士と、（株）画像技研の高田亮氏、清水昭浩氏、齋藤圭司氏の共同開発によるものです。

なお、PEACH は、最新バージョン公開に向けて現在整備中です。テストバージョンならば配布可能ですので、希望者は komeiji@nair.go.jp まで連絡してください。

参考文献

<教科書>

- [1] 上田あきら, コンピューターシミュレーション, 朝倉書店 (1990).
- [2] 岡田勲、大沢映二, 分子シミュレーション入門, 海文堂 (1989).
- [3] 川添良幸、三上益弘、大野かおる, コンピュータシミュレーションによる物質科学, 共立 (1996).
- [4] 田中実、山本良一, 計算物理学と計算化学, 海文堂 (1990).
- [5] フーヴァー (田中実訳), フーヴァー 分子動力学入門, 共立 (1998).
- [6] Allen, M. P., Tildesley, D. J., *Computer Simulation of Liquids*, Clarendon Press, Oxford (1987).
- [7] Brooks, C. L., Karplus, M., Pettitt, B. M., *Proteins: a theoretical perspective of dynamics, structure, and thermodynamics*, John Wiley & Sons, N. Y. (1988).
- [8] Frenkel, D., Smit, B., *Understanding Molecular Simulation*, Academic Press, S. D. (1996).

[9] Haile, J. M., *Molecular dynamics simulation: Elementary Methods*, John Wiley Sons, N. Y. (1992).

[10] Rapport, D. C., *The art of molecular dynamics simulation*, Cambridge Univ. Press (1995).

< 原著論文 >

[11] Taiji, M., Fukushige, T., Makino, J., Ebisuzaki, T., Sugimoto, D., MD-GRAPE: A parallel special-purpose computer system for classical molecular dynamics simulations., *Proceedings of the 6th joint EPS-APS international conference on Physics Computing* (Gruber, R., and Tomassini, M.), 200-203, European Physical Soc., Geneve. (1994).

[12] Komeiji, Y., Uebayasi, M., Takata, R., Shimizu, A., Itsukashi, K., Taiji, M., Fast and accurate molecular dynamics simulation of a protein using a special-purpose computer, *J. Comp. Chem.*, **18**, 1546-1563 (1997).

[13] Guenot, J., Kollman, P. A., Molecular dynamics studies of a DNA binding protein 2, *Prot. Sci.*, **1**, 1185-1205 (1992).

[14] Komeiji, Y., Uebayasi, M., Someya, J., Yamato, I., Molecular dynamics simulation of trp-aporepressor in a solvent, *Prot. Engng.*, **4**, 871-875 (1991).

[15] Saito, M., Molecular dynamics simulations of proteins in water without the truncation of long-range Coulomb interactions, *Mol. Simul.*, **8**, 321-333 (1992).

[16] Darden, T. A., York, D. M., Pedersen L. G., Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems, *J. Chem. Phys.*, **98**, 10089-92 (1993).

[17] Smith, P. E., Pettitt, M., Ewald artifacts in liquid state molecular dynamics simulations, *J. Chem. Phys.*, **105**, 4289-4293 (1996).

[18] Smith, P. E., Blatt, H. D., Pettitt, M., On the presence of rotational Ewald artifacts in the equilibrium and dynamics properties of a zwitterionic tetrapeptide in aqueous solution, *J. Phys. Chem.*, **B 101**, 3886-3890 (1997).

[19] Noberto de Souza, O., Ornstein, R. L., Effect of periodic box size on aqueous molecular dynamics simulation of a DNA dodecamer with Particle-Mesh Ewald method, *Biophys. J.*, **72**, 2395-2397 (1997).

[20] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M., CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *J. Comp. Chem.*, **4**, 187-217 (1983).

[21] Harvey, S. C., Tan, R. K.-Z., Cheatham III, T. E., The flying ice cube: velocity rescaling in molecular dynamics leads to violation of energy equipartition, *J. Comput. Chem.*, **19**, 726-740 (1998).

- [22] Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Weiner, P., A new force field for molecular mechanical simulation of nucleic acids and proteins, *J. Am. Chem. Soc.*, **106**, 765-784 (1984).
- [23] Weiner, S. J., Kollman, P. A., Nguyen, D. T., Case, D. A., An all atom force field for simulations of proteins and nucleic acids, *J. Comp. Chem.*, **7**, 230-252 (1986).
- [24] Cornell, W. D., Cieplak, P., Layly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc.*, **117**, 5179-5197 (1995).
- [25] Jorgensen, W. L., Tirado-Rives, J., The OPLS potential functions of proteins, *J. Am. Chem. Soc.*, **110**, 1657-1666 (1988).
- [26] Komeiji, Y., Uebayasi, M., Molecular dynamics simulation of the Hin-recombinase DNA complex, *Mol. Simul.*, **21**, 303-324 (1999).
- [27] Fincham, D., Optimization of the Ewald sum for large system, *Mol. Simul.*, **13**, 1-9 (1994).
- [28] Komeiji, Y., Uebayasi, M., Yamato, I., Molecular dynamics simulations of trp- apo- and holorepressors, *Proteins*, **20**, 248-258 (1994).
- [29] Greengard, L., Fast algorithms for classical physics, *Science*, **265**, 909-914 (1994).
- [30] Bishop, T. C., Skeel, R. D, Schulten, K., Difficulties with multiple time stepping and fast multipole algorithm in molecular dynamics, *J. Comput. Chem.*, **18**, 1785-1791 (1997).
- [31] Swope, W. C., Andersen, H. C., Berens, P. H., Wilson, K. R., A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules, *J. Chem. Phys.*, **76**, 637-649 (1982).
- [32] Tuckerman, M., Berne, B. J., Martyna, G. J., Reversible multiple time scale molecular dynamics, *J. Chem. Phys.*, **97**, 1990-2001 (1992).
- [33] Woodcock, L. V., Isothermal molecular dynamics calculations for liquid salts, *Chem. Phys. Lett.*, **10**, 257-261 (1971).
- [34] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W., DiNola, A., Haak, J. R., Molecular dynamics with coupling to an external bath, *J. Chem. Phys.*, **81**, 3684-3690 (1984).
- [35] Nose, S., Constant temperature molecular dynamics methods, *Prog. Theor. Phys. Suppl.*, **103**, 1-46 (1991).

- [36] Komeiji, Y., Uebayasi, M., Change in conformation by DNA-peptide association: Molecular dynamics of the Hin-recombinase:hixL complex, *Biophys. J.*, **77**, 123-138 (1999).
- [37] Oda, K., Miyagawa, H., Kitamura, K., How does the electrostatic force cutoff generate non-uniform temperature distributions in proteins?, *Mol. Simul.*, **15**, 167-177 (1996).
- [38] McLachlan, A. D., Gene duplications in the structural evolution of chymotrypsin, *J. Mol. Biol.*, **128**, 49-79 (1979).
- [39] Shrake, A., Ruply, J. A., Environment and exposure of solvent of protein atoms, *J. Mol. Biol.*, **79**, 351-371 (1973).
- [40] Harte, W. E. Jr., Swaminathan, S., Beveridge, D. L., Molecular Dynamics of HIV Protease., *Proteins*, **13**, 175-194 (1992).
- [41] Hunenberger, P. H., Mark, A. E., Van Gunsteren, W. F., Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations, *J. Mol. Biol.*, **252**, 492-503 (1995).
- [42] Amadei, A., Linssen, A. B. M., Berendsen, H. J. C., Essential Dynamics of Proteins, *Proteins*, **17**, 412-425 (1993).
- [43] Kitao, A., Hirata, F., Go, N., The effects of solvent on the conformation and the collective motions of protein: normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum, *Chem. Phys.*, **158**, 447-472 (1991).
- [44] Hayward, H., Go, N., Collective variable description of native protein dynamics, *Annu. Rev. Phys. Chem.*, **46**, 223-250 (1995).
- [45] Balsera, M. A., Wriggers, W., Oono, Y., Schulten, K., Principal component analysis and long time protein dynamics, *J. Phys. Chem.*, **100**, 2567-2572 (1996).

Molecular Dynamics Simulation of Biological Molecules (1) Methods

Yuto KOMEIJI^{a,c*}, Masami UEBAYASI^b and Umpei NAGASHIMA^a

^aNational Institute for Advanced Interdisciplinary Research
1-1-4 Higashi, Tsukuba, 305-8562 Japan
**e-mail: komeiji@nair.go.jp*

^bNational Institute of Bioscience and Human-Technology
1-1-3 Higashi, Tsukuba, 305-8566 Japan

^cElectrotechnical Laboratory
1-1-4 Umezono, Tsukuba, 305-8568 Japan

Molecular dynamics simulation has been extensively applied to analyses of structure, dynamics, and function of biological molecules. Various methods implemented in a newly developed software, PEACH (Program for Energetic Analysis of bioCHEMical molecules), are described. This review covers: (1) an outline of classical molecular dynamics, (2) generation of the initial structure and temperature, (3) computation of force and potential, (4) time integration, (5) ensemble, and (6) analyses of the generated trajectories.

Keywords: Program for Energetic Analysis of bioCHEMical molecules (PEACH), GRAvity piPE (GRAPE), Molecular dynamics, Proteins, DNA, RNA