

生体分子の分子動力学シミュレーション (2) 応用

古明地 勇人^{a, b, c*}, 田島 澄恵^{d, e}, 原口 誠^d, 高橋 伸幸^f, 上林 正巳^c, 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

^b 電子技術総合研究所, 〒 305-8568 つくば市梅園 1-1-4

^c 生命工学工業技術研究所, 〒 305-8566 つくば市東 1-1

^d (株) ベストシステムズ, 〒 305-0047 つくば市千現 2-1-6

^e お茶の水女子大学理学部情報科学科, 〒 112-8610 文京区大塚 2-1-1

^f 北海道教育大学函館校物理学教室, 〒 040-8567 函館市八幡町 1-2

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

(Received: February 8, 2000; Accepted for publication: March 14, 2000; Published on Web: July 13, 2000)

分子動力学シミュレーションは、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一般に認められている。本総説では、筆者らが開発した生体分子解析用ソフトウェア PEACH (Program for Energetic Analysis of bioChemical molecules) を使って、タンパク質と DNA の分子動力学シミュレーションを行う場合の手続きを、実践的に解説する。

キーワード: Program for Energetic Analysis of bioChemical molecules (PEACH), Molecular Dynamics (MD), Protein, DNA

目次

1. 序
2. 計算条件の決定
3. PDB ファイルの加工
4. マグネシウムイオンのデータ作成
5. 溶質分子の真空中の分子トポロジー作成
6. 溶質、イオン、溶媒を含んだ分子トポロジーの作成
7. エワルド法の条件決め
8. エネルギー極小化と昇温
9. 本計算
10. 結果の解析
11. おわりに

参考文献

付録: PEACH の MPI による並列化

1 序

本総説では、プログラム PEACH (Program for Energetic Analysis of bioCHemical molecules) ver. 3 を使って生体高分子の分子動力学 (Molecular Dynamics, MD) シミュレーションを行う方法を、実践的に解説する。筆者らは「生体高分子の分子動力学シミュレーション (1) 方法」(古明地ら [1]) において、PEACH の中に導入された MD の方法について詳述した。今回は、前回紹介した方法を使って MD を行う手順を、具体的に紹介したい。

プログラム PEACH Ver. 3.0 は、

<http://www.etl.go.jp/~komeiji/>

から入手することができる。以下では、UNIX 環境下で PEACH ソースプログラムの解凍とコンパイルが完了したのものとして、話を進める。

PEACH の使用法は、ソースプログラムと一緒に配布される英文ドキュメントに詳述してあるが、初心者にとって、ドキュメントを読んですぐに PEACH プログラムを実行するのは、なかなか困難である。そこで、一通りプログラムを動かすための実行例を作成し、ソースプログラムと併せて配布している。

peach_3.0/demo/BPTI (タンパク質の計算例)

peach_3.0/demo/DNA (核酸の計算例)

の二つのディレクトリに入っている。本総説では、これらの実行例に沿って MD 計算の手続きを解説する。

ディレクトリ BPTI/の下にはタンパク質の例として BPTI (Bovine Pancreatic Trypsin Inhibitor) の入力データと 30 ps の MD の実行結果、DNA/の下には核酸の例として 7 base pair の二本鎖 DNA の入力データと 30 ps の MD の実行結果が、それぞれ入っている。両者の下はほとんど同じディレクトリ構造なので、使用した順に各ディレクトリーの内容を説明する。なお、以下で「分子トポロジー」という場合は、分子の構成原子の種類、電荷、結合情報など、座標と速度を除いた、分子の全情報のことを示す。

data/	入力用三次元データの作成。
MG_dat/	マグネシウムイオンのトポロジー作成 (DNA のみ)
mktop_v/	溶質部分 (タンパク質か核酸) のトポロジー作成 (真空中)
mktop_bx/	箱形の水中の状態での分子トポロジー作成。
heat/	エネルギー極小化と温度上昇。
mdprod/	MD の本計算の実行。
analmd/	結果の解析。

これらのディレクトリ内のファイルの概要は、名前だけで、内容の見当が付くようになっている。以下、* は任意の文字列を表す。

*.in は、ジョブ制御用のデータやフラグパラメータが書かれた入力ファイル

*.csh は、実行用の C-shell Script ファイル

*.out は、診断用の出力ファイル

これら以外は、人間ではなく計算機が読むためのファイル
 (ただし、すべて ASCII 文字列であるので、人間も読む気になれば読める)
 従って、基本的に、この実行例を編集して新しい分子の計算を行うためには、
 *.in ファイルを計算したい分子に合わせて編集し、
 *.csh を実行し、
 *.out ファイルが出力されたら、それを読んで、結果を確認する
 という作業を行えばよい。

Figure 1 に、PEACH の各プログラムモジュールの機能と相関関係を模式的に示した。以下では、個々のモジュールの機能を簡単に説明する。

- MKDBAS プログラム アミノ酸とヌクレオチドのトポロジーデータベースを作成する。
 MKMOL プログラム 分子のトポロジー (結合情報) を作成する。
 MKCOR プログラム PDB ファイルから三次元座標を読み込み、
 イオンや溶媒の付加などを行い、
 RUNMD プログラムで使用するための
 初期座標ファイルを作成する。
 MKPARA プログラム 力場パラメータを割り当てて、RUNMD プログラムで使用する
 ための最終的なトポロジーファイルを作成する。
 RUNMD プログラム MKPARA で作成したトポロジーファイルと
 MKCOR で作成した座標ファイルを読み込み、
 MD またはエネルギー極小化計算を実行する。
 ANALMD プログラム 作成した MD トラジェクトリーファイルを加工したり
 解析するための様々なプログラムの集合。
 以下には、この総説で説明するプログラムのみを示す。

個別プログラム名	入力ファイル	プログラムの機能
ANEN	trajene_*	エネルギーの解析
GYRCRD	partop trajcrd	回転半径の解析
RMSCRD	partop restrt trajcrd	初期構造からの根平均二乗 変位の解析
TMPVEL	partop trajvel	温度分布の解析

MOSBY プログラム (作成、電総研・上野豊 [2]): RUNMD のトラジェクトリーの
 アニメーション表示を行う。このプログラムは、PEACH
 のモジュールではなく、別途に配布されている。

以上のプログラムモジュールを用いて、MD シミュレーションを行うわけである。以下の章では、MD を始めるまでの準備 (2 - 7 章、分子トポロジーファイルと座標ファイルの作成、条件決め、等)、実際の MD 計算 (8, 9 章)、結果解析 (10 章) のそれぞれについて、実行例を説明する。

この総説で説明する peach_3.0/demo/ 下のジョブは、
「Scientific な MD 計算を一通り行うための雛形」
として使ってもらうことを想定して作成したため、最終結果を得るまでかなり計算時間がかかる（1, 2 日）。単にプログラムが正常に動くかどうかだけを確かめたい場合は、

peach_3.0/test/

の下のファイルを利用したほうがよい。

なお、MD の計算手順と解析方法は、研究目的と対象によってそれぞれ異なる。「水素結合パターン」、「分子内の協同運動」、「フォールディング過程」、「溶媒と溶質の相互作用」、等々、研究したい現象により、シミュレーションの手順は変わり得る。だが、本総説では特定の目的を定めず、最大公約数的な MD の手順を示した。ユーザーは、この実行例を参考にして、各自の目的に合った方法を工夫して欲しい。

2 計算条件の決定

作業に先立って、MD の計算条件を決める（Table 2）。一番重要なのは力場であるが、ここでは最新の AMBER96 を使うことにする。溶媒は、周期境界条件を持った箱型モデルを選んだ（文献 [1] 2.2 節）。クーロン力の計算は、エワルド法で行う（文献 [1] 3.5.3 節）。

条件を決定した後、Protein Data Bank(PDB) から目的分子の三次元座標ファイルを取ってくる。

URL <http://www.rcsb.org/pdb/>

から入手可能である。今回は BPTI は 6PTI.pdb を、DNA は 312D.pdb を使うことにした。最近では NMR による溶液構造も多いが、この二つは X 線解析による結晶構造である。

以下、BPTI と DNA の 2 つの実行例を並行して解説していく。ユーザーは、自分の手で計算を行う前に、結果の検証のために、ディレクトリーごとオリジナルを複製して保存しておくことを薦める。

3 PDB ファイルの加工 (data/)

以下の章では、それぞれのディレクトリ内のファイル名とその内容を説明する。「入力ファイル」、「出力ファイル」等は、単に「入力」「出力」と略す。また、「ある分子を構成する各原子の三次元座標のデータの集まり」のことを、単に「座標」と呼ぶことにする。

<u>入力</u>	<u>ファイルの内容</u>
(BPTI)	
6PTI.pdb	PDB から入手したままの、加工前の座標。
(DNA)	
312D.pdb	"
<u>実行シェルスクリプト</u>	
(BPTI, DNA)	
pdbrenum.csh	pdbrenum プログラムを実行し、生の PDB ファイルを、mkcor 用に整形する。

出力

(BPTI, DNA)

pdbrenum.log 実行ログ。
in.pdb mkcor で利用するための、加工後 (手作業を含む) の PDB 形式の座標ファイル。

その他

(BPTI)

README ファイルの簡単な説明。

シェルスクリプトを実行させるには、

```
./pdbrenum.csh
```

とコマンドを入力すればよい。そうすると、pdbrenum.log と in.pdb のファイルが作成される。この pdbrenum.csh の中では、pdbrenum というプログラムが実行され、PDB ファイルの残基の番号を付け替えたり、原子の名前を入れ替えたりして、PEACH で読みやすくなるようになるようにする。だが、このプログラムは万能ではない。現時点では、作成された in.pdb に対し、以下のような手作業が必要である。

1. 高精度な結晶構造の場合、ひとつの原子に対して A, B ふたつの座標があることがある。ここではとりあえず、B は消して A のみ残した。
2. X 線結晶解析で得られた構造の場合、普通は PDB に水素原子の座標が入っていない。PEACH の mkcor プログラムは足りない残基や原子を自動発生して補う機能を持っているが、先頭付近の水素 (タンパクなら H1, H2, H3, HA ; DNA なら H5T, H5'1, H5'2) は、残念ながら自動発生できない。そこで、人の手で座標を作って付加する必要がある。
3. AMBER 94 と 96 の力場の場合、DNA 5' 末端と 3' 末端の塩基名は、それぞれ DT5, DT3 等に変更しなければならない。
4. CYS のうち SS ボンドを作るものを CYX に変更する。
5. HIS (ヒスチジン) がある場合は、HIP (荷電状態によっては、HID, HIE) に変更するなど。

Q: どのように水素の座標を作ればいいのか ?

A: AMBER94 の場合、H1, H2, H3 は、N に、HA は CA に、H5T は O5'T に、H5'1 と H5'2 は C5' に、それぞれ、結合している。そこで、その結合している重い原子 (N, CA, O5'T, C5') の座標をコピーして、そこから、だいたい 1 離れた場所に、水素がくっつくように、手作業で座標を作る。つまり、x, y, z のどれかの値を 1 だけ増せばよい。

「こんないい加減な水素の座標でいいのか ? 」

と思われるだろうが、あとで水素だけエネルギー極小化して最適化するので (5.4 節) 実用上は問題ない。

PEACH

Program for Energetic Analysis of bioCHemical molecules, Yuto Komeiji (1993-)

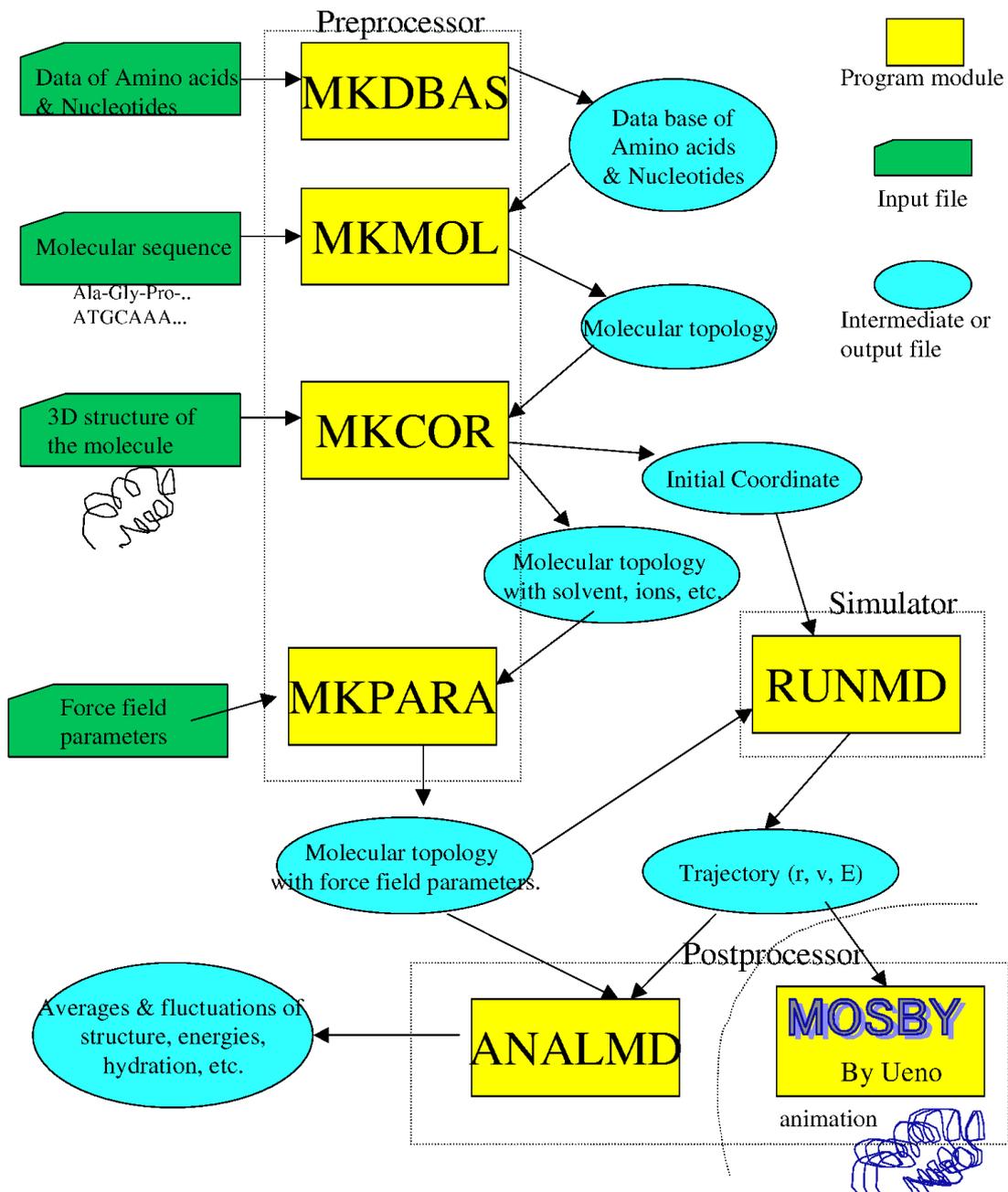


Figure 1. PEACH のプログラムモジュールと入出力ファイル

Table 2. demo ディレクトリ中の分子シミュレーションの条件

	BPTI	DNA
初期構造	6PTI.pdb	312D.pdb
溶媒の箱の一辺の長さ (Å)	立方体 46×46×46	直方体 39×42×48
原子数	892	439
溶質		
イオン	6 (Cl ⁻)	6 (Mg ²⁺)
溶媒	7941	6981
合計	8839	7426
	(以下の条件は、BPTIとDNAに共通)	
計算機	COMPAQ Alphastation 600 au (0.6 Gflops)	
力場	AMBER96 (溶質とイオン) Flexible SPC Water (溶媒)	
アンサンブル	能勢のカノニカルアンサンブル (温度一定)	
時間積分と刻み幅	Tuckermanの多時間刻み幅法 (RESPA)	
	堅い力 (共有結合、結合角、ねじれ角)	0.25 fs
	中間の力 (VDW、エワルド実空間)	2.00 fs
	柔らかい力 (エワルド波数空間)	4.00 fs
非共有結合力の計算	ファンデルワールス (VDW) 力 12 で切断	
	クーロン力 (エワルド) 実空間	12 で切断
	波数空間	Kmax = 8
	(2109 ベクトル)	
	(VDW 力とエワルド実空間は、40 fs 毎に原子ペアリストを作り、それを利用して、一つのサブルーチンで計算した)	



Figure 2. BPTI (左、リボン表示)とDNA(右、スティック表示)の結晶構造。グラフィックプログラム RASMOL[3] を使用。

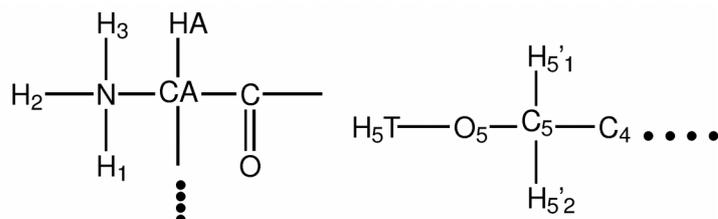


Figure 3. タンパク質 (左) の N 末端と核酸の 5 ' 末端 (右)

4 マグネシウムイオンのデータ作成 (MG_dat/) * DNA のみ

DNA の MD では、今回のデモでは、系の電荷を中和するためのカウンターイオンとしてマグネシウムイオン (Mg^{2+}) を利用することにした。しかし、AMBER94(96) の力場では、このイオンは標準データベースには含まれていない。そこで、mkdbas プログラムを使って、 Mg^{2+} のトポロジーデータベースを作成する。

<u>入力</u>	<u>ファイルの内容</u>
mg.in	分子トポロジー作成用入力データ。
<u>実行シェルスクリプト</u>	
mkdbas.csh	mkdbas プログラムを実行する。
<u>出力</u>	
mg.out	診断出力。
mg.dat	mkmol で使用するための、 Mg^{2+} イオンの分子トポロジー。

5 溶質分子の真空中の分子トポロジー作成 (mktop_v/)

はじめに、真空中の溶質 (BPTI または DNA) の分子トポロジーファイルを作る (v は vacuum の略)。一連のジョブは、

```
./MKTOP_V.csh
```

と打てば実行されるようになっているが (数分)、このシェルスクリプトの中では、

```
mkmol.csh
```

```
mkcor.csh
```

```
mkpara.csh
```

```
min.csh
```

```
mkpdb.csh
```

の 5 つのシェルスクリプトを、順番に実行させるようになっている。それぞれの行っているジョブの内容を以下で説明する。このディレクトリ下で作成されるファイルの中で、最終的に後で利用するものは、

vac.pdb

で、水素を最適化した座標ファイルを、PDB 形式に直したものである。これを 6 章の mktop_bx で利用する。

5.1 mkmol

溶質分子 (タンパク質または核酸) の一次構造 (アミノ酸または塩基配列) を入力して、分子トポロジーを作成するモジュールである。

<u>入力</u>	<u>ファイルの内容</u>
molin	ジョブ制御用入力ファイル。 この中で、シミュレーションを行う対象の タンパク質のアミノ配列や核酸の塩基配列を定義する。
../../data/dbase/amber94/amb94tot.dbas	Amber94 力場のトポロジーデータ ベース (末端以外のアミノ酸残基と核酸塩基)
../../data/dbase/amber94/amb94nt.dbas	" (N 末端のアミノ酸残基)
../../data/dbase/amber94/amb94ct.dbas	" (C 末端のアミノ酸残基)

実行シェルスクリプト

mkmol.csh mkmol プログラムを実行する。

出力

molout 診断出力。
moltop 作成された分子トポロジー。
5.2 節で mkcor の入力となる。

molin のポイント

1. SS 結合の Cross link を正しく定義しているかどうか。(今回のデモでは BPTI のみ)
(例 : CROSS RES 5 ATOM SG - RES 55 ATOM SG これは、残基 5 の SG という名の原子と、
残基 55 の SG という名の原子の間に Cross link があることを示している。)

molout のポイント

1. 系の全電荷が、整数になっているかどうか (キーワード : TOTAL CHARGE OF THE SYSTEM)。整数でない場合は、何かがおかしい。
2. プログラムが正常終了したなら、最後に
MKMOL: PROGRAM SUCCESSFULLY ENDED
というメッセージが書かれる。

5.2 mkcor (BPTI, DNA)

ここでは、moltop と、PDB ファイルを読み込み、mkcor プログラムを実行して、座標ファイルを作る

<u>入力</u>	<u>ファイルの内容</u>
corin	ジョブ制御用入力ファイル。
moltop	mkmol で作った分子トポロジーファイル
../data/in.pdb	入力用の PDB 形式の座標ファイル。
<u>実行シェルスクリプト</u>	
mkcor.csh	mkcor プログラムを実行する。
<u>出力</u>	
corout	診断出力。
cortop	作成された分子トポロジー。mkpara の入力となる。
corcor	最終的な (runmd の初期入力となる) 座標ファイル。

corout のポイント

- 途中で止まってしまうときは、molin で定義した残基名と、in.pdb での残基名が一致していないことが多い。
- どこかの原子に*INT というフラグが出ていないか？ このフラグは、「きちんとした座標を読み込めなかったし、それを作成することもできなかった」ということを意味する。もし、出ている場合は、in.pdb をチェック。
- プログラムが、正常終了したなら、最後に
MKCOR: PROGRAM SUCCESSFULLY ENDED
というメッセージが書かれる。

5.3 mkpara

以上で作られた分子トポロジーに対し、力場パラメータを割り当てる。

<u>入力</u>	<u>ファイルの内容</u>
parin	ジョブ制御用入力
cortop	mkcor で作成した分子トポロジー。
../../data/dbase/amber96/parm.list	Amber96 力場のパラメータ
<u>実行シェルスクリプト</u>	
mkpara.csh	mkpara プログラムを実行する。
<u>出力</u>	
parout	診断出力。
partop	最終的な分子トポロジー。

mkpara.csh のポイント

- mkmol ので利用したトポロジーデータベースと対応した力場パラメータファイルを使っているか？ ここで使った Amber96 の力場パラメータファイルは、Amber94 のトポロジーデータベースに対応している。

parout のポイント

1. パラメータが足りない、という表示が出たら、parm.list を調べる。
2. プログラムが、正常終了したなら、最後に
MKPARA: PROGRAM SUCCESSFULLY ENDED
というメッセージが書かれる。

5.4 min

mkcor で作成した座標 (corcor) は、水素は内部座標を用いて (あるいは人間が手で) 適当に発生したものである。よって、水素だけを、エネルギー極小化計算 (Energy Minimization, EM、文献 [1] 2.3 節) により、最適化しておく。使うモジュールは、runmd である (MD も EM も、同じプログラムモジュールを利用する)。

<u>入力</u>	<u>ファイルの内容</u>
minin	ジョブ制御用入力。
partop	mkpara で作成した分子トポロジー。
corcor	mkcor で作成した分子座標。
<u>実行シェルスクリプト</u>	
min.csh	runmd プログラムを実行する。
<u>出力</u>	
minout	診断出力。
minrep	計算の途中経過を示す、報告ファイル。 その時点での、ステップとエネルギー等が出力される。
atomout	各原子に関する情報。
min.rst	極小化計算された座標のファイル。5.5 節で使われる。

minin のポイント

1. "HYDROGEN" というフラッグを立てること。このフラッグを立てることで、水素だけが最適化される。

minout のポイント

1. EM 計算の途中でエネルギーが発散していないか ? (エネルギーがあるステップで極端に変化していたら要注意)
2. プログラムが正常終了したなら、最後に
ALL FINISHED, BYE-BYE
というメッセージが書かれる。

5.5 mkpdb

ここでは、5.4 で作成した、水素を最適化した座標を PDB 形式に変換する。

<u>入力</u>	<u>ファイルの内容</u>
partop	mkpara で作られた分子トポロジーファイル
min.rst	水素を最適化した後の構造
<u>実行シェルスクリプト</u>	
mkpdb.csh	mkpdb プログラムを実行する。
<u>出力</u>	
vac.pdb	水素を最適化した構造を、PDB 形式に変換したもの。 これを、6章の mktop_bx で利用する。

6 溶質、イオン、溶媒を含んだ分子トポロジーの作成 (mktop_bx/)

このディレクトリでは、5章で作った真空中の溶質の分子トポロジーに対して、箱形の水を発生させる。

MKTOP_BX.csh

を実行すれば、数分で全作業が終わる。このスクリプトの中では

mkmol.csh

mkcor.csh

mkpara.csh

mkpdb.csh

の4つのスクリプトが実行される。ここで作成されたファイルの中で、以後で使うものは

partop 分子トポロジーファイル

corcor 座標ファイル

の二つである。

基本的な作業は、5章の場合と同様であるが、溶媒やイオンを発生させる部分が、少し異なる。5章で説明したファイルについては、ファイルの内容は省略した。

6.1 mkmol (BPTI, DNA)

<u>入力</u>	<u>ファイルの内容</u>
molin	
../../data/dbase/amber94/amb94tot.dbas	
../../data/dbase/amber94/amb94nt.dbas	
../../data/dbase/amber94/amb94ct.dbas	
(DNAのみ	
../MG_dat/mg.dat	マグネシウムイオンのトポロジー。mkdbasの出力。 molinの中で定義する。)
<u>実行シェルスクリプト</u>	
mkmol.csh	

出力

molout
moltop

molin のポイント

1. molin 内で、イオンを定義することが必要。イオンは、5.1 で出力された molout の total charge を中和するだけの数を発生させればよい。BPTI の場合は、6 つの Cl⁻ イオン (CIM) を発生させている。これは、AMBER94 のデータベースに標準で入っている。一方、DNA の場合は、6 つの Mg²⁺ イオン (MG) を発生させているが、これは標準のデータベースにはない。よって、molin 内で、mg.dat を読むように指定することが必要。

molout のポイント

1. 系の全電荷がきちんと中和されているか (TOTAL CHARGE OF THE SYSTEM が、0 になっているか)

6.2 mkcor

基本的には 5.2 節と同じだが、ここでは、水を発生させるための設定が必要である。

入力

corin
moltop

../mktop_v/vac.pdb

../.../data/dbase/fSPC/amber94/wattop

../.../data/dbase/fSPC/amber94/watcrd

ファイルの内容

入力用の PDB 形式の座標ファイル。ここでは、5 章で作った座標を入力する。

flexible SPC water の分子トポロジー。

水の座標データ。

実行シェルスクリプト

mkcor.csh

出力

corout
cortop
corcor

runmd で使うための座標ファイル。

corin のポイント

1. "GENION" フラグをたてる。これで、イオンの座標を自動発生することができる。

corout のポイント

1. イオンの座標ができていないか。つまり "GEN" というラベルがついているか。

別々の例を示すため、BPTIでは立方体の、DNAでは直方体の水を、それぞれ発生させているので注意。

Q: イオンの座標はどういう基準で作られたのか？

A: 溶質 + イオン系のクーロンポテンシャルが最小になるような位置を選んで、溶媒分子を置き換える形で発生させている (文献 [1] 2.2.2 節)

6.3 mkpara

入出力とも、5.3 節と全く同様なので、省略。

ここで作成された partop ファイルが、runmd で使用するための最終的なトポロジーファイルである。

6.4 mkpdb (BPTI,DNA)

partop と corcor を使って、PDB 形式のファイルを作る。このファイルは、MD 計算で利用するわけではないが、分子グラフィクスプログラムで分子構造を表示させて、イオンや水などが思ったように発生されているか、チェックするために使う。Figure 6.4 に水とイオンを発生させた、最終的な座標ファイル (solv.pdb) をグラフィクスプログラム RASMOL (Sayle [3]) で表示した。

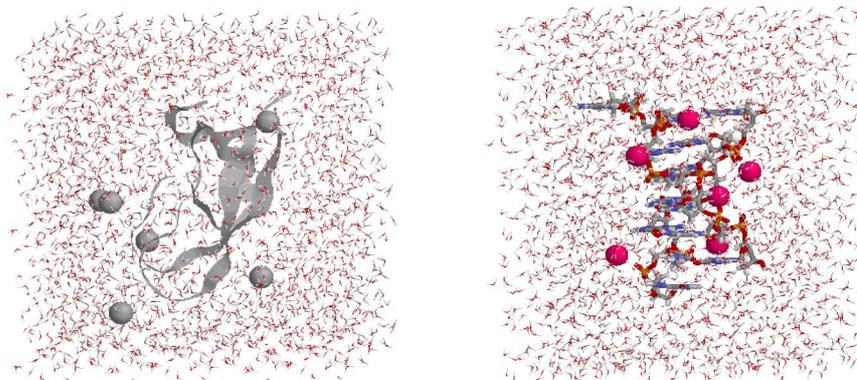


Figure 6.4. Figure 2 に示した、BPTI (左) と DNA (右) の結晶構造に対して、水とイオンを発生させたもの。

Q: 分子トポロジーファイルと座標ファイルができたものの、なぜこんなに作成が面倒なのか？作業が複雑なことにメリットがあるのか？

A: 今は、GUI がついて、簡単にシミュレーション計算が実行できる市販ソフトウェアが多い。PEACH はそれらに比べて原始的であり、ブラックボックスとして使うのには不向きである。ただし、プログラムの中で一体何が行われているのか各段階毎に確認できるし、ユーザーが習熟すれば細かい条件を自在に設定できるメリットがある。

7 エwald 法の条件決め (cond-ewald/)

入力

なし (入力データは、スクリプト中に記載されている)

実行シェルスクリプト

vewald.csh

出力

vewald.log

vewald.out

エwald 法は、箱の大きさや要求される精度により、パラメータを変化させる必要がある。vewald は、精度とパラメータを自動的に算出するプログラムである。ここでは、一応、精度を $=10^{-3}$ になる条件を採用した。Kmax=8, Rcut=12 , $=4.8$ 。ただし、ここで用いた誤差解析方法はかなり厳しい条件を課しているの、これを厳密に守る必要はない (もっと、甘い条件でも精度は保たれる)。以下の EM 計算と MD 計算では、この条件を使った。

8 エネルギー極小化と昇温 (heat/)

このディレクトリでは、まず系全体のエネルギーを極小化し、ついで、温度を 300 K まで上昇させる。

```
./HEAT.csh > & HEAT.log &
```

とジョブを発行する。計算時間が掛かる (約 10 時間強) のので、上記のようにバックグラウンドで走らせるとよい。ログは、HEAT.log に出力される。なお、以下で、座標や速度の「トラジェクトリー」という表現を使うが、これは、「分子を構成する各原子の座標や速度の、時々刻々の値」を記録したファイルのことを示す。

入力

../mktop_bx/partop

../mktop_bx/corcor

min0.in

qd0.in

md*.in

実行シェルスクリプト

HEAT.csh

出力

min0.out

qd0.out

md*.out

atomout

*.rst

*.crd

ファイルの内容

分子トポロジー。

初期座標。

ジョブ制御入力 (EM 用)

" (QD 用、Quenched Dynamics、後述)

" (昇温 MD 用)

runmd プログラムを実行する。

EM の診断出力。

QD の診断出力。

昇温 MD の診断出力。

原子に関する様々な情報のまとめ。

それぞれの段階での最終座標 (と速度) ファイル。

座標のトラジェクトリー (溶媒込み)

*.crdv	" (溶媒なし)
*.vel	速度のトラジェクトリー (溶媒込み)
*.velv	" (溶媒なし)
.ene_	さまざまなエネルギー関係の量の時間変化。
minrep	EM の経過報告。
qdrep	QD の経過報告。
mdrep	MD の経過報告。

HEAT.csh を実行すると、

```
min0 qd0 md2 md4 md6 md8 md10
```

の順に、ジョブが実行され、EM から始まって最終的に 300 K まで温度が上がるようにしてある。各過程について、以下で詳しく説明する。

8.1 min0, qd0

MD の準備としての EM は、

「初期構造が持っている、不適當な歪みを取り除く」

ため だけ にある。よって、(たとえば基準振動解析を行うときのように)凝る必要はない。今回は、Steepest descent (SD) と、Quenched dynamics (QD) で行った (文献 [1] 2.3 節)。SD は収束しにくいアルゴリズムなので、収束した / しないに関わらず、ある程度までエネルギーが下がったところで、QD に切り替えてしまう。今回は、200 ステップで、次の qd0 に切り替えた。QD とは、低温 (今回は、0.1 K) で MD を行って EM の効果を得るための方法。時間積分 (文献 [1] 4 章) の発散を防ぐため時間刻み幅が短い (0.1 fs)。また、系が歪みを持っていた場合に起こる急激な温度変化に対応できるように、Nose の定温アルゴリズム (文献 [1] 5.3 節) の時間定数も極端に短い (0.1 fs)。さらに、もし時間積分が発散してしまった場合は、仮想熱浴粒子の座標と運動量を 0 にリセットするようにしてある。

8.2 md2 - md10

続く昇温過程では、10 ps の間を 5 段階に分けて 10 K から 300 K まで温度を上げている。昇温過程では、Nose の時間定数は、10 fs と短めにしてある。なお、md2, md4, といったネーミングだが、ここでは、

「ジョブが終了した点でのシミュレーション時刻」

をつけている。たとえば、md2 は終了時点で 2 ps の時刻ということになる。ジョブの名前付けは、それぞれのユーザーが自分のわかりやすいように付ければよい。

今回の実行例での温度の上げ方は多少荒っぽいのが、もっと念入りにしたければ「時間を長めに」したり「温度の上げ幅を小刻みに」したりすることで対応できる。

昇温過程でも、それに続く本計算 (9 章) でも、重心運動は凍結している。また、RESPA による多重時間刻みを使っている (文献 [1] 4.2 節)。結合距離、結合角、ねじれ角の力には、0.25 fs、VDW 力、エワルド実空間は、2 fs、エワルド逆格子空間は 4 fs の時間刻み幅を、それぞれ

採用している (Table 2)

Q: エネルギー極小化と昇温過程において、もっとも重要なポイントは何か？

A: QD を行うことである。これは、説明したとおり、極低温 (0.1 - 5 K 程度) で、しかも短い時間幅で (0.1 fs 程度) で MD を行うことである。わずかに数 10- 数 100 ステップ行うだけであるが、これをしないで、いきなり長い時間幅を与えると、温度が急上昇し、時間積分が発散して、ジョブが止まってしまうことが多い。なお、今回は、Nose の定温アルゴリズムでの時定数を、極端に短くすることで、QD 温度の上昇を防いだ。もし、これでも温度が急上昇してしまう場合は、Woodcock のスケーリング法で (文献 [1] 5.2 節) 無理矢理ステップ毎に温度を押さえ込んでしまえばよい。筆者の経験では、QD さえ無事終われば、そのあと時間積分が失敗することは、ほとんどない。

Q: 重心運動を凍結する理由は？

A: 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"問題が生ずる。これは、重心運動にばかり運動エネルギーが分配されてしまい、結果として、系の内部の運動エネルギーが低下して温度が下がってしまう現象 (文献 [1] 2.5.2 節) である。

Q: 多重時間刻み法 RESPA の条件は上記でよいのか？

A: ここでは、経験上まず問題ないと思われる条件をとりあえず採用した。ただし、本来は、MD の対象や力場の種類に応じて、様々な時間幅の組み合わせを試すべきである。目安としては、

結合長、結合角、ねじれ角 :	0.1 - 0.5 fs
VDW 力、クーロン力 (エワルド 実空間):	0.5 - 2.0 fs
エワルド 波数空間 :	2.0 - 5.0 fs

程度の範囲で、精度と速度を考慮して選べば良い。

9 本計算 (mdprod/)

本番の MD 計算 (production run) は、mdprod/の下で、

```
./MDPROD.csh > & MDPROD.log &
```

とジョブを実行した。これも時間が掛かる (約一日) ので、上記のように、バックグラウンドで走らせるとよい。ログは、MDPROD.log に出力される。入出力ファイルは、基本的には 8 章の heat と同じである。

<u>入力</u>	<u>ファイルの内容</u>
../mktop_bx/partop	分子トポロジー。
../heat/md10.rst	時刻 10 ps での座標と速度。
md.in	制御用入力。md20, md30 とともに共通。
実行シェルスクリプト	
MDPROD.csh	runmd プログラムを実行し、md20(10-20 ps) と、

md30 (20-30ps) のジョブを行う。

出力

md*.out	MD の診断出力。
atomout	原子に関する様々な情報のまとめ。
md*.rst	それぞれの段階での最終座標と速度ファイル。
md*.crd	座標のトラジェクトリー (溶媒込み)
md*.crdv	" (溶媒なし)
md*.vel	速度のトラジェクトリー (溶媒込み)
md*.velv	" (溶媒なし)
md*.ene_*	さまざまなエネルギー関係の量の時間変化。
mdrep	MD の経過報告。

Q: runmd を動かすと何種類もの座標ファイルと速度ファイルが作成されるが、なんのためか？

A: md*.rst は、restart 用のファイル。つまり、一つのジョブが終了した時点のシミュレーションでの時刻と各原子の座標と速度が出力される。次のジョブでは、これが入力になる。

一方、md*.crd, crdv, vel, velv には、時々刻々の変化が入っている。これらは、トラジェクトリー解析のためのファイルである。今回のデモでは、*.crd と *.vel では 1ps おきの座標と速度が、また *.crdv と *.velv では 0.2 ps おきの座標と速度が、それぞれ出力されている。

なお、*.crd と *.vel には溶媒を含む全原子のデータが出力されるのに対し、*.crdv と *.velv には溶質とイオンのデータしか出力されない。溶媒まで含むとデータが巨大になるが、溶質とイオンだけなら、小さいファイルで済むからである。普通は溶質とイオンだけが解析の主眼となる。しかし、解析の主眼が、「溶媒分子の挙動」ならば、*.crdv と *.velv は作る必要はなく、*.crd と *.vel に溶媒分子を含むデータを頻繁に出力させるべきである。

10 結果の解析 (analmd/)

このディレクトリでは、8 章と 9 章で作成した出力の解析を行っている。MD の結果の解析は着眼点により様々であるが (文献 [1] 6 章) 標準的なものを例として挙げておいた。PEACH では、解析は analmd と呼ばれるプログラム群で行う。サブディレクトリの名前は、プログラム名と一致させておいた。

analmd/	
anen/	エネルギーの解析
tmpvel/	アミノ酸残基/核酸塩基毎の温度の解析
rmscrd/	初期構造からの根平均二乗変位の解析
gyrcrd/	回転半径の解析

analmd に属するプログラム群は、PEACH の本体 (mkdbas, mkmol, mkcor, mkpara, runmd) と違って、対話式に実行するようになっている (user friendly とは言い難いが)。だが、今回の実行例では、バッチ式のシェルスクリプトで実行した。入力パラメータの意味を知りたい方は、対話式に実行して欲しい。

ここで例を挙げた4つのプログラムは、runmd モジュールの出力のどれかを読み込んで、それらをまとめて、さらに何かの量を計算する、という形になっている。つまり、以下の種類のファイルがされる。

md*.crd	座標のトラジェクトリー (溶媒込み)
md*.crdv	" (溶媒なし)
md*.vel	速度のトラジェクトリー (溶媒込み)
md*.velv	" (溶媒なし)
md*.ene_*	さまざまなエネルギー関係の量の時間変化

10.1 anen/ エネルギーの解析

anen は、エネルギーの時間変化のファイルをいくつも読み込んで、それらをまとめ、平均と標準偏差を計算するプログラムである。

入力

../heat/md2.ene_tot	0-2 ps 間のエネルギーの時間変化。
../heat/md4.ene_tot	2-4 ps "
../heat/md6.ene_tot	4-6 ps "
../heat/md8.ene_tot	6-8 ps "
../heat/md10.ene_tot	8-10 ps "
../mdprod/md20.ene_tot	10-20 ps "
../mdprod/md30.ene_tot	20-30 ps "

実行シェルスクリプト

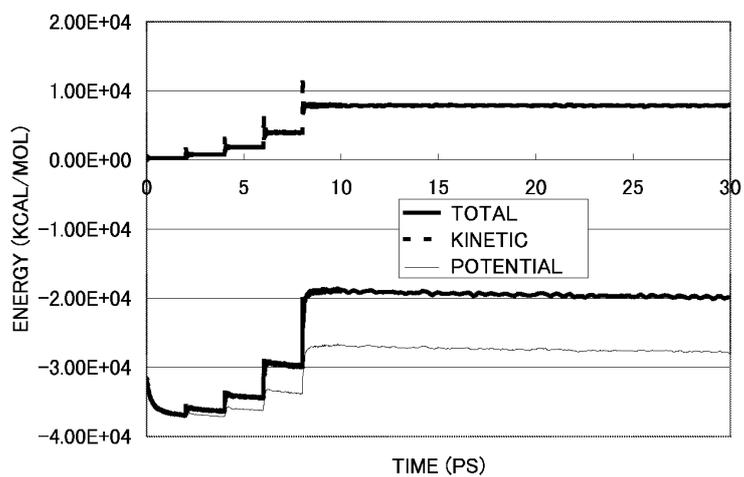
md0_30.csh	anen プログラムを実行する。
------------	------------------

出力

md0_30.log	実行ログ。エネルギーの平均と標準偏差。
md0_30.ene	0-30 ps 間のエネルギーの時間変化。

Figure 10.1 に、BPTI と DNA それぞれの系のエネルギーの時間変化を、全エネルギー (total)、運動エネルギー (kinetic)、ポテンシャルエネルギー (potential) に分けて、示した。昇温過程では運動エネルギーが段階的に上昇し、それに連れてポテンシャルエネルギーも上昇するのがわかる。系全体のエネルギーは、昇温がおわった 8 ps 以降は大体安定だが、わずかに下がる傾向にある。これは、系がまだ緩和の途中であることを示している。

A. BPTI



B. DNA

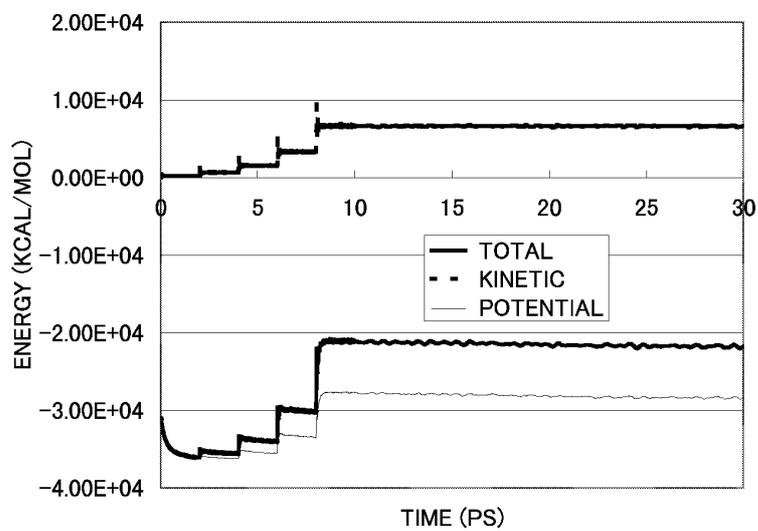


Figure 10.1. エネルギーの時間変化

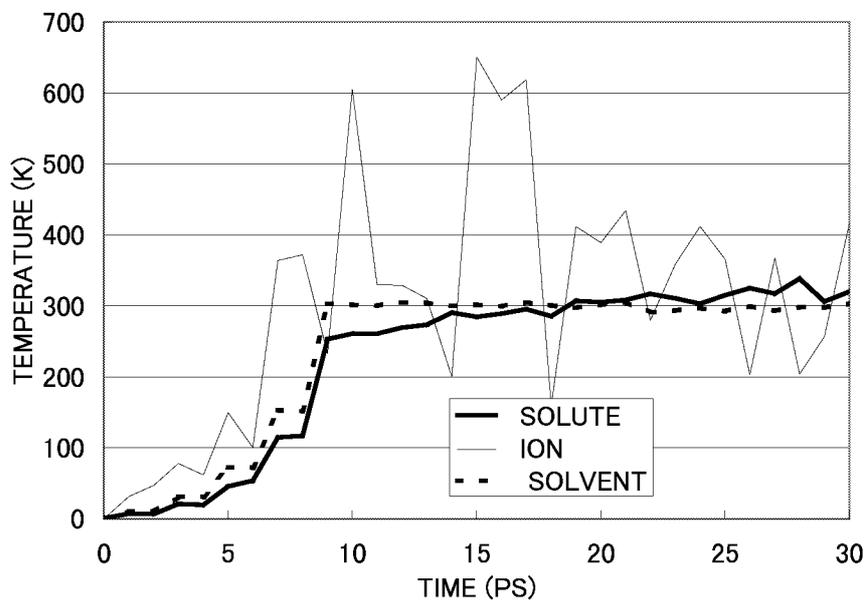
10.2 tmpvel/ 温度分布の解析

tmpvel は、速度の時間変化のファイルをいくつも読み込んで、それらをまとめ、さらに残基の重心運動の温度と内部運動の温度を計算するプログラムである (文献 [1] 6.1.1 節)

<u>入力</u>	<u>ファイルの内容</u>
../heat/md2.vel	0-2 ps 間の速度の時間変化。
../heat/md4.vel	2-4 ps "
../heat/md6.vel	4-6 ps "
../heat/md8.vel	6-8 ps "
../heat/md10.vel	8-10 ps "
../mdprod/md20.vel	10-20 ps "
../mdprod/md30.vel	20-30 ps "
<u>実行シェルスクリプト</u>	
md0_30.csh	tmpvel プログラムを実行する。
<u>出力</u>	
md0_30.log	実行ログ。
md0_30.tmp	0-30 ps 間の溶質、イオン、溶媒の温度の時間変化、平均と標準偏差。
md0_30.tmpres	0-30 ps 間の残基毎の温度の時間変化、平均と標準偏差。

Figure 10.2 は、溶質、イオン、溶媒毎に分けて、温度の時間変化を表示した。温度の揺らぎの大きさは、イオン > 溶質 > 溶媒の順であるが、これは、原子数が少ない順に揺らぎが大きくなる、ということである。30 ps では、完全に熱平衡には至らないので、三者の平均温度は完全には一致しないが、数 100 ps の MD を行ったあと平均をとれば、温度は一致するはずである。ただし、クーロン力に対しカットオフを使った場合は、溶質、イオン、溶媒の間に温度格差が生ずる (文献 [1] 6.1.2 節)

A. BPTI



B. DNA

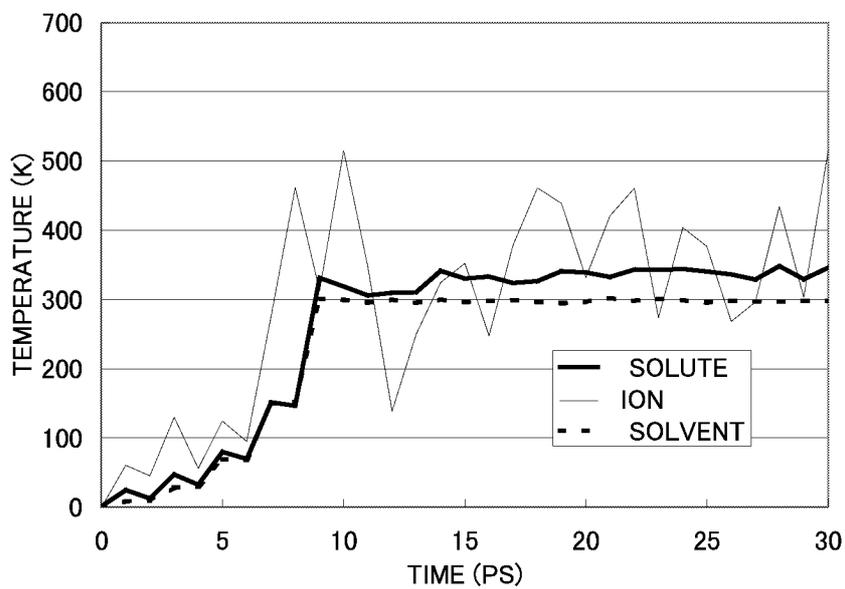


Figure 10.2. 溶質、イオン、溶媒の温度変化

10.3 rmscrd/ 初期構造からの根平均二乗変位の解析

rmscrd は、初期構造からの MD 構造の違い (根平均二乗変位, RMSD) を比較するプログラムである (文献 [1] 6.2.1 節)

入力

.././mktop_bx/partop	トポロジーファイル。
.././mktop_bx/corcor	初期構造の座標ファイル。
.././heat/md2.crdv	0-2 ps 間の座標の時間変化。
.././heat/md4.crdv	2-4 ps "
.././heat/md6.crdv	4-6 ps "
.././heat/md8.crdv	6-8 ps "
.././heat/md10.crdv	8-10 ps "
.././mdprod/md20.crdv	10-20 ps "
.././mdprod/md30.crdv	20-30 ps "

実行シェルスクリプト

md0_30.csh

出力

md0_30.log	実行ログ。
md0_30.avg	全体、溶質、イオン、分子、残基、それぞれの 0-30 ps 間の RMSD の自乗平均。
md0_30.tot	0-30 ps 間の全体、溶質、イオンの RMSD の時間変化。
md0_30.mol	0-30 ps 間の分子毎の RMSD の時間変化。

Figure 10.3 に RMSD の時間変化を示す。溶質分子全体 (ALL)、主鎖 (MAIN)、側鎖 (SIDE) に分けて表示した。MAIN や SIDE は、Main chain、Side chain の略である。

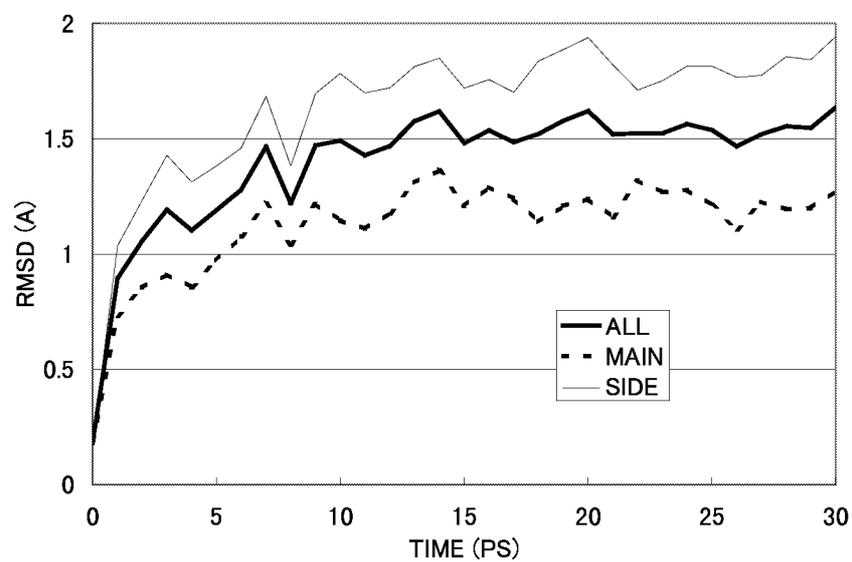
Q: 水素原子を無視するのはなぜか？

A: PDB の結晶構造には水素の座標は入っていない。よって水素原子のずれを RMSD に考慮しても、「結晶構造との比較」にならないため。

Q: BPTI では、主鎖のほうが側鎖に比べて RMSD が小さいのに対し、DNA ではその逆になっている。なぜか？

A: BPTI をはじめ、タンパク質では側鎖は溶媒に面していることが多く、動き易いのが普通である。一方、DNA の場合、主鎖のリン酸基は溶媒に露出しているのに対し、ここで側鎖として扱われている塩基は、内部で塩基対水素結合を作っている。よって、側鎖のほうが動きにくく、結果として、初期構造からのずれも小さくなった。

A. BPTI



B. DNA

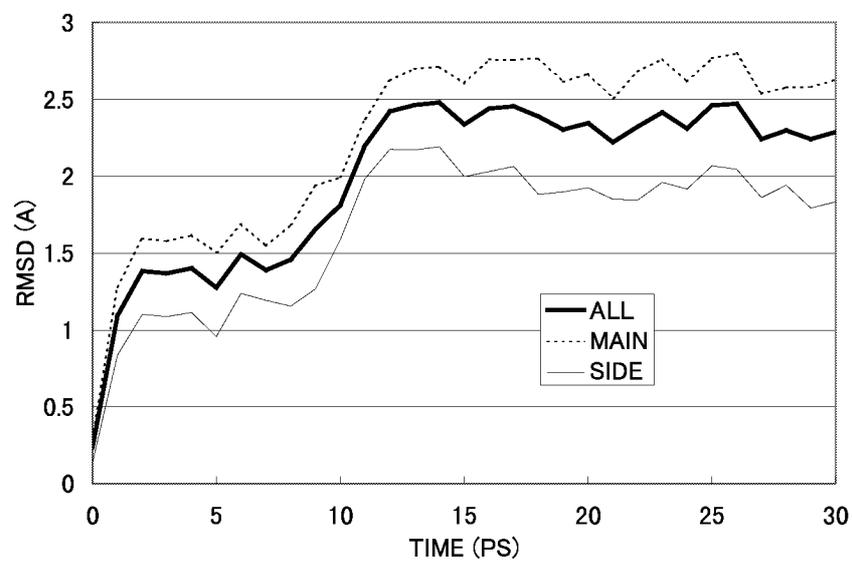


Figure 10.3. 結晶構造からのずれの時間変化

10.4 gyrcrd/ 回転半径の解析

gyrcrd は、分子の回転半径を計算するプログラムである (文献 [1] 6.2.3 節)

<u>入力</u>	<u>ファイルの内容</u>
../mktop_bx/partop	分子トポロジーファイル。
../heat/md2.crdv	0-2 ps 間の座標の時間変化。
../heat/md4.crdv	2-4 ps "
../heat/md6.crdv	4-6 ps "
../heat/md8.crdv	6-8 ps "
../heat/md10.crdv	8-10 ps "
../mdprod/md20.crdv	10-20 ps "
../mdprod/md30.crdv	20-30 ps "
<u>実行シェルスクリプト</u>	
md0_30.csh	gyrcrd プログラムを実行する。
<u>出力</u>	
md0_30.log	ログファイル
md0_30.gyr	0-30 ps 間の溶質全体の回転半径の時間変化、平均、標準偏差 (DNA のみ)
md0_30.mol	二本の鎖それぞれの、0-30 ps 間の回転半径の時間変化、平均、標準偏差。

Figure 10.4 に回転半径の時間変化を示した。目盛りを小さく取っているので激しい変化に見えるが、実際は、BPTI、DNA とともに、0.5 Å 以内の揺らぎに収まっている。ただし、両方とも、初期構造に比べてわずかに広がる傾向にある。これは、結晶構造を初期構造に用いた MD ではごく普通の現象で、結晶環境から溶液環境に出たことで、分子が広がる傾向を示している。

11 おわりに

以上、生体分子の MD について、例を挙げて解説した。計算が終わるのにそれぞれ 30 時間以上掛かるが、それでもわずか 30 ps のトラジェクトリが得られるだけである。MD に詳しい方の中には、runmd モジュールを「遅い」と感じられた方も多いと思う。だが、並列化された runmd モジュールを利用すればもっと早く計算できるので、並列計算機をお持ちの方は、そちらも試して頂きたい。詳しくは、付録を参照のこと。今日では、今回例に挙げた程度の小さい分子なら、数 100 ps から 10 ns 程度のトラジェクトリを作成して解析するのが普通となっている。

実際に MD を行う場合、今回紹介した以外に、様々な工夫が行われている。たとえば、今回はタンパク質や DNA 分子と、溶媒とイオンを、同時に平衡化している。だが、特に DNA の場合は、安定した結果を得るために、まず DNA は固定しておいて、溶媒とイオンを EM と MD で最適化させる、という前処理を行うことも多い。このような工夫は、MD の対象と目的に応じて、各自試みていただきたい。

10 章では、MD トラジェクトリーの解析作業の基本的な例を挙げた。PEACH の analmd と

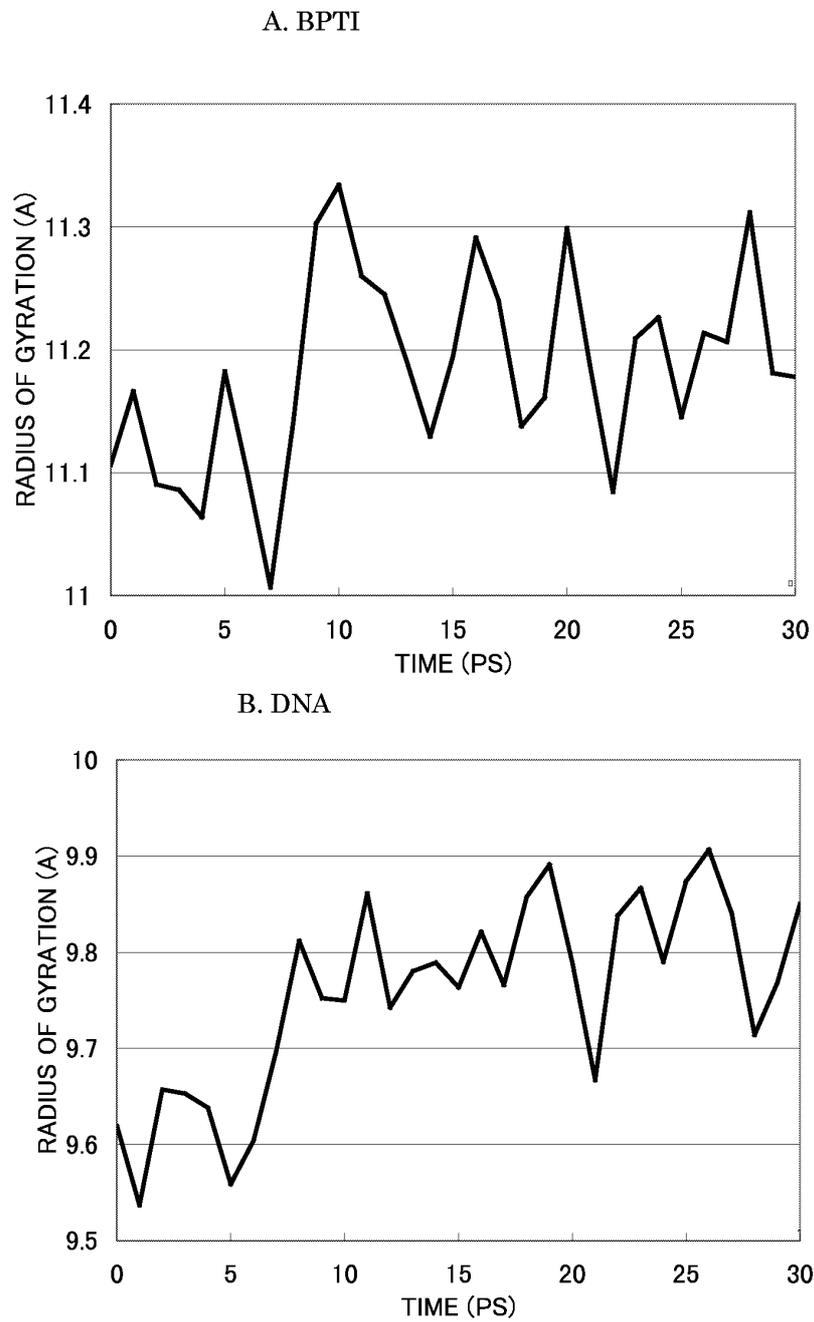


Figure 10.4. 回転半径の時間変化

misc のプログラム群を用いれば、これ以外にも様々な解析を行うことができる。残念ながら、ドキュメントや実行例は完全ではないが、各自試みて下さると幸いである。また、もし PEACH 対応の解析プログラムをユーザー自身が作成し、それを寄託して下さいならば、望外の喜びである。

産業技術融合領域研究所の寺倉清之主席研究官のご指導ご支援に感謝いたします。

参考文献

- [1] 古明地勇人, 上林正巳, 長嶋雲兵, 生体分子の分子動力学シミュレーション (1) 方法, *J. Chem. Software*, **6**, 1-36 (2000).
- [2] Ueno, Y., Asai, K., A new plug-in software architecture applied for a portable molecular structure browser, *Proceedings, Intelligent systems for molecular biology*, **5**, 329-332 (1997).
- [3] Sayle, R., *RASMOL 2.6*, Glaxo Research and Development, Hertfordshire, U. K. (1995).

付録：PEACHのMPIによる並列化

本文で紹介した runmd モジュールによる MD 計算の結果は、単一 CPU を用いたものである。だが、PEACH ver. 3.0 では、runmd モジュールは、MPI ライブラリを用いて並列化されている。Table A に、本文の BPTI のジョブを例にとって、MPI により並列計算を行ったベンチマークの結果を示す。計算条件は Table 2 と同様だが、計算機は、工業技術院情報計算センター (TACC) の IBM RS/6000 SP スカラー計算サーバーを用いた。

Table A. 並列計算に要した計算時間 (Table 2 の BPTI の計算と同一条件)

MD 1 fs あたりの経過時間を秒で表示。カッコ内は並列化効率 (%)

CPU 数	1	4	8	16
エワルド 実空間	3.3	0.83	0.42	0.22
+ VDW 力 の計算	(100)	(98)	(98)	(91)
エワルド 波数空間の計算	6.4	1.60	0.81	0.41
	(100)	(99)	(99)	(97)
MD 全体	10.2	3.02	1.81	1.23
	(100)	(84)	(70)	(50)

現在並列化されているのは、クーロン力 (エワルド法) と VDW 力計算用サブルーチンである。これらの部分は、16 台の場合でも 90% 以上の高い効率で並列化されていることがわかる。MD 計算全体では、並列化されていない部分 (特にペアリスト作成サブルーチン) が目立つようになるため、台数が増えると効率は落ちるが、それでも 16 台で 50% は保っている。現時点では、10 台前後の CPU で計算するのが、効率的だろう。

Molecular Dynamics Simulation of Biological Molecules (2) Practice

Yuto KOMEIJI^{a, b, c*}, Sumie TAJIMA^{d, e}, Makoto HARAGUCHI^d, Nobuyuki TAKAHASHI^f, Masami UEBAYASI^c and Umpei NAGASHIMA^a

^aNational Institute for Advanced Interdisciplinary Research
1-1-4 Higashi, Tsukuba 305-8562, Japan

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

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

^dBestsystems, Inc.
2-1-6 Sengen, Tsukuba 305-0047, Japan

^eOchanomizu University
2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan

^fHokkaido University of Education
1-2 Yawata, Hakodate, Hokkaido, 040-8567

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

Molecular dynamics simulation has been extensively applied to analyses of structure, dynamics, and the function of biological molecules. In this review, the practical procedures of molecular dynamics simulations of a protein and DNA by using the software PEACH (Program for Energetic Analysis of bioCHEmical molecules) are described.

Keywords: Program for Energetic Analysis of bioCHEmical molecules (PEACH), Molecular Dynamics (MD), Protein, DNA