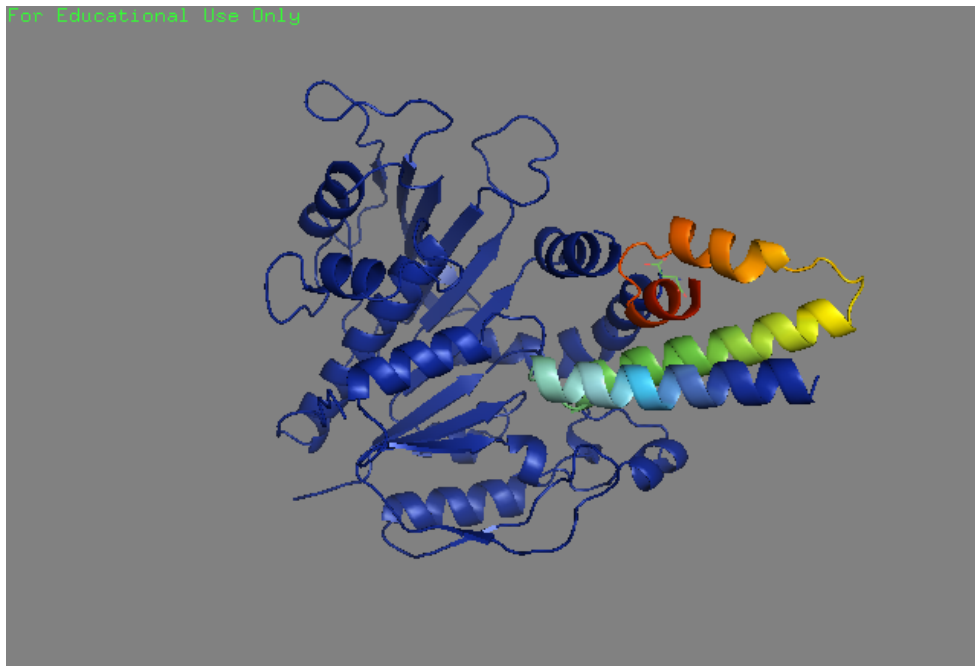


湯井-AMBER使用法 ver. 3.1

タンパク質の分子動力学計算

平成 25 年 7 月
和田野 加筆・訂正



第1章 基本事項

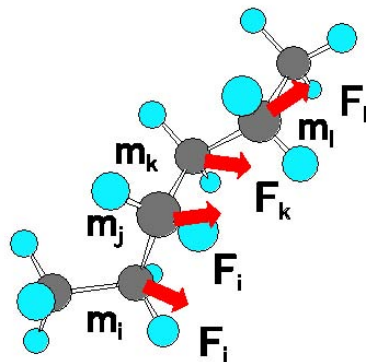
1. 分子動力学の概念

分子動力学法（MD法；Molecular Dynamics method）は，計算機シミュレーションの一種で，物質の諸性質を調べる際に使われる手法の一つである。

物質（固体，液体，気体）は，粒子（原子，分子，イオン）の集合体であるため、物質のマクロな性質（密度，硬さ，内部エネルギー等）は、粒子の配置の仕方・運動の仕方といったミクロな現象によって決定される。計算機の中で全ての粒子の運動を時事刻々と追跡することにより，物質全体としての性質を導き出すのが分子動力学法である。

各粒子の運動はどのようにして決まるか？それは，その粒子が周りの粒子からどのような力を受けるかによって決まる。粒子は，周囲から受ける力の大きさに比例し，その粒子の質量に逆比例した大きさの加速度を受ける（ニュートンの運動の第2法則）。右図のようにシミュレーション対象を分子または分子の集合体とした場合、粒子とは分子を構成する原子となる。原子の質量を m 、原子に働く力と加速度をそれぞれ、 F および a とするとそれらは第二法則より次式の関係となる。

$$m_i \times a_i = F_i \quad (i = 1, 2, 3, \dots, N)$$



個々の原子の運動方程式を積分することで時間変化に対する各原子の位置と速度が決定される。

力も原子の位置によって決定されるが、分子動力学では原子間距離、結合長、結合角、および回転角を変数とするポテンシャル関数によって計算される。各粒子の位置（ポテンシャル）エネルギーと運動エネルギーは粒子の位置と速度に関係する。従って、時間変化とともに粒子のポテンシャルエネルギーと運動エネルギーは変化するが、エネルギー保存則により系全体の全粒子のポテンシャルエネルギーと運動エネルギーの総和は一定となる。

古典力学の世界ではこれら各粒子の速度や位置は連続的に変化することを前提としているが、計

算機内部でこのような物理現象を実現する場合、ある微小時間 (Δt) を設定する。

分子動力学法計算の流れとして、

- 1) 物質を構成する各粒子間に働く力を設定されたポテンシャル関数によって求める,
- 2) ある瞬間における各粒子が受ける加速度を求め,
- 3) 微小時間 Δt 秒後の各粒子の座標 (および速度) を求める,

といった計算を繰り返して行う。 Δt は積分での積分幅 (時間ステップ) に相当し、通常、分子の挙動で周期の短い C-H結合振動周期をもとに1fs (10^{-15} 秒) が指定される。ポテンシャル関数は van der Waals相互作用、静電相互作用、結合伸縮、結合角変角および結合回転、等の項目に対して経験式とパラメータによって構成される。これら式とパラメータを総称して分子力場と呼ばれる。分子力場はシミュレーションプログラムによって異なるが、通常、プログラムのバージョンアップにともない関数式はそのままパラメータが変更・追加される。

もちろん現在のコンピュータの性能をもってしても、実際に我々が観測するマクロな性質と比べてはるかに少ない粒子数とはるかに短い時間を計算機の内部で仮想的に再現している。従って、近似的に温度一定 (すなわち運動エネルギーの平均が一定)、体積一定、あるいは圧力一定の条件を必要に応じて設定することが必要となる。

分子動力学計算は生体高分子 (DNA、蛋白質、糖鎖) に対しては、分子の構造、物性、機能関連の研究に有用である。

2. AMBERの概略

AMBERは、カリフォルニア大学のコールマン教授らのグループによって生体分子のために開発された、モデリングおよび分子力学と動力学計算シミュレーションプログラムのパッケージである。AMBERとはこのプログラムパッケージ名とプログラムで使用される力場の名称の両方に適用される。ここでは最も主要な 4つのプログラムについて説明する。

- ① LEaP 計算の前段階で、力場やトポロジー（原子の結合状態）情報が記載されているprmtopファイルと初期原子座標情報が記載されてinpcrdファイルを作成する。グラフィックス画面で操作できる“xleap”とテキストターミナルから操作する“tleap”の2種類のバージョンがある。
- ② sander 実際のシミュレーション作業で使用する最も主要なプログラム。分子力学法による構造最適化と分子動力学計算を行う。並列計算機の複数cpuには”sander.MPI”を用いる。cpuの数は 2^n に限定される。
- ③ pmemd sanderをさらに並列化にチューニングさせたバージョン。sander.MPIはcpu=4でスケールリングが頭打ちになるが、pmemdはcpu=256でも可能なようである。pmemdはnvidiaのGPUにも対応しcuda環境中で高度な並列計算を実行出来る。
- ④ ptraj 計算結果を解析するためのプログラム。sanderが出力するトラジェクトリーファイル（時間ステップごとの各原子の座標が書かれたファイル）の座標をもちいて、シミュレーション中の座標RMSや構造パラメータ変位や平均値、および水素結合寿命などを計算する。

AMBER HP ; <http://amber.scripps.edu/>

3. sanderの基本的な使い方と関連ファイル

基本的なsander の使い方は次のとおりである：

```
sander [-0] -i mdin -o mdout -p prmtop -c inpcrd -r restrt [-ref refc -x mdcrd -v mdvel  
-e mden -inf mdinfo]
```

“-” 指標でファイルの種類を指定する。ファイル名（斜体文字）は自由に決めてよい。なお[]内の引数はオプションで、指定しなければデフォルトの名称が使われる。ファイルの順序は任意である。

-0 指定すると同じ名前前の出力ファイルが存在している場合、新しい計算結果が上書きされる。

これらのファイル名はデフォルト名を使用する必要はなく、全く任意に命名してよい。後述するように、計算の段階に応じてファイル名にファイルや計算の種類、プロジェクト名の簡略表現を適用する。その結果、個々のファイル名称が長くなり計算の実行コマンド入力が面倒になるため、以下のように計算実行のためのスクリプトファイルを作成し使用してもよい。また、すぐに終了する試験計算でない限り計算の実行はバックグラウンド処理で行う。

デフォルト ファイル 名	指標	入力/ 出力	目的
mdin	-i	入力	計算条件（オプション）を記述したファイル名。
mdout	-o	出力	シミュレーション・構造最適化の過程における、エネルギーその他の計算結果を書き出すファイル名。
prmtop	-p	入力	パラメーター・トポロジーファイル名。
inpcrd	-c	入力	出発構造の座標を記述したファイル名。
restrt	-r	出力	構造最適化あるいはMD計算の最終構造の座標を書き出すファイル名。
refc	-ref	入力	mdin ファイル内で座標束縛オプションが指定されている場合の参照座標ファイル名。
mdcrd	-x	出力	分子動力学計算を実施したときに、トラジェクトリーを書き出すファイル名。
mdvel	-v	出力	分子動力学計算を実施したときに、各原子の速度情報を書き出すファイル名。
mden	-e	出力	分子動力学計算を実施したときのエネルギーの要約ファイル名。
mdinfo	-inf	出力	構造最適化あるいは MD 計算の各ステップが書き込まれる、要約ファイル名。

sander 実行ためのスクリプトファイル例（run_sander）

```
#!/bin/csh
```

```
sander -O -i myModel_min.in ¥
        -o myModel_min.out ¥
        -c myModel_ini.inpcrd ¥
        -p myModel.parm7 ¥
        -r myModel_min.rst
```

次に以下のコマンドによりバックグラウンド処理で計算を起動する。chmod は最初に一回だけ実行すればよい。

```
$chmod +x run_sander
```

```
$. /run_sander >& log & (&はバックグラウンド処理でプログラムを動かす記号)
```

研究室の計算機はすべて並列機なので、実際は単独 cpu を前提とした使い方はしない。さらにジョブスケジューリング（自動的に計算ノードを振り分ける）プログラムである LAVA の制御下で計算を実行するためスクリプトファイルは以下のようにする。

- 並列計算実行のためのスクリプトファイル例 (md_sub、ノード数 : 8)

```
$ mpirun -np 8 sander.MPI -O -i ./Mymodel_md.in -o ./Mymodel_md.out -c ./Mymodel_min.rst  
-p ./Mymodel.parm7 -r ./Mymodel_md.rst -x ./Mymodel_md.mdcrd"
```

mpirun 以降は改行を入れずに一行で記述する。-np は cpu 数であり、計算機がマルチスレッドでなければ同じ値となる。cpu 数は 2^n (2, 4, 8, 16, 32, ...) に限定される。pmemd を使用する場合は、sander.MPI を pmemd で置き換える。pmemd は Sander における溶媒の表現方法の項で後述するように周期境界条件+Particle Mesh Ewald(PME)モデル用に開発されたプログラムであるから真空中のモデルには適応してはいけない。

4. 入力ファイル “mdin” の作成

sanderの実行において mdinファイルの作成が最も重要な作業となる。計算条件を設定する各種のオプションはネームリストで指定される。以下に例を示す。

```
Molecular dynamics run  
  &cntrl  
    IMIN = 1, NCYC = 250, MAXCYC = 500,  
  &end  
  &ewald  
    A = 43.5297, B = 44.3736, C = 41.0257,  
    NFFT1 = 45, NFFT2 = 45, NFFT3 = 45,  
  &end
```

sanderの計算では様々なネームリストのコントロール変数をもとに計算条件が設定されるが、実際はその一部を指定するだけでよい。入力ファイル中で指定されていない変数は、予めプログラムで決められたデフォルトの値が設定される。マニュアル中に使用できるネームリストご

なお、1行目はタイトル行でどんな内容を記述しても計算には影響しない。通常、計算システムや主要な計算条件をコメントする。

生体高分子のシミュレーションは真空中で行われることはめったになく、溶媒分子、すなわち水分子を溶質である生体高分子の周囲に配置した環境で計算を行う。sanderにおける溶媒分子の表現方法は1) “仮想溶媒“による水分子を連続体として表現する方法と2) 周期境界条件により実際に水分子を配置する方法が備えられている。

水分子の持つ誘電率と静電的な遮蔽効果を持つ連続体として水分子による溶媒環境を近似する。周期境界条件による計算と比較して必要な計算量ははるかに削減される。水分子が明示的に設定されないため周期境界条件の計算で必要となる長時間の平衡化計算が不必要となる。周期境界条件で設定される PME 近似（後述）を設定する必要がない。

実際に水分子の集合体の中に溶質分子を置く。但し、無限大のサイズではなく溶媒分子を満たした“Box”を設定する。このままでは溶媒 Boxの周辺部で水分子が蒸発するため、同一の溶媒 Boxのレプリカを周期的に設定する。その結果、溶媒 Boxの一方の境界から外側に移動した水分子は反対側の境界から溶媒 Boxの中に拡散してくるため(右図)、

The diagram shows a 3x3 grid of cells. The central cell is shaded gray and contains five numbered circles (1-5). The other cells (A, B, C, D, E, F, G, H) contain one circle each. A vertical double-headed arrow on the left is labeled 'L', and a horizontal double-headed arrow at the bottom is labeled 'L'.

方法によって近似的に計算される。溶媒 Boxには、一定体積と一定圧力の 2種の設定方法がある。

3) Cap Waterモデル 溶媒Box計算のように溶質分子である生体高分子を完全に溶媒Box中に置くのではなく、ある1点を中心にして球状の水分子クラスターを設定する。通常、その同じ中心点から特定の距離内部の水分子と蛋白質の一部を含む領域をシミュレーション時の可動領域とする。外部空間への水分子の蒸発を防ぐための中心方向に引力を設定する。当然、溶媒 Boxより計算量が少なく酵素タンパクの活性部位の挙動のように、生体高分子全体でなく限られた部分の現象を対象とするときに有効である。

第2章 Glyceraldehyde 3-phosphate dehydrogenase(gapdh)と cp12 complex の 分子動力学計算 AMBER Tutorial

1. “prmtop” と “inpcrd” ファイルの作成

初期 PDB ファイルの作成

LEaP において必要となる対象分子の初期立体構造情報は sequence コマンド等で LEaP で作成する方法と PDB を読み込む方法に 2 種類がある。ここでは阪大工学部の松村先生によって構造決定された gapdh の PDB ファイルと cp12 の部分 PDB ファイルを用いてその結合エネルギーを計算する。

1.2. 孤立分子システムの AMBER ファイル作成

最初に最も簡単な事例として、gapdh分子のみとした孤立分子システムに対する AMBERファイルを作成する。以下の内容の LEaP入力ファイル” leap_vac.in” を作成する。

```
set default writel4scale on
source leaprc.ff99SB

model = loadPdb gapdh.pdb
saveAmberParm model gapdh_vac.parm7 gapdh_vac.inpcrd

quit
```

leaprc.ff99SBにより parm99パラメータを指定する。loadPdbにより初期座標となる PDBファイルを読み込み、これを modelという名称の model unitに定義する。この名称は任意でよく、同じ LEaPセッション内で段階ごとに新しい名前を定義しても良い。PDBファイルに水素原子がなければこの段階で水素原子が付加される。saveAmberParmにより modelに定義された分子システムからトポロジーファイル (parm7) と初期座標ファイル (inpcrd) を生成する。AMBERはファイル種類にかかわらず名称は任意であるが、一般に拡張子にファイルの種類を、残りの部分を分子や状態や計算条件を簡単に表現する名前にする。例えば、 gapdh_vac.parm7AT gapdh分子の真空状態 (vac) の計算に使用するトポロジーファイル (parm7) 拡張子名は慣例として、トポロジーファイルは.prmtopまたは.parm7、計算条件を記載した入力ファイルは.inまたは.inp、入力座標ファイルは.inpまたは.inpcrd、最終座標(リスタート)ファイルが.rst、出力ファイルが.out、そしてMD軌跡ファイルが.crdまたは.mdcrd、等が一般的である。

ファイル名称の前半部について、計算条件等の影響を受けないトポロジーファイルはプロジェクト（対象分子、計算システム）に関する名称だけで良いが、それ以外のファイルは計算段階や条件によって内容が変化するため、それらを記述する名称とする。さらにプロジェクト名称を追加してもよいが、一般にファイル名称が長くなるので避けた方が良い。

mdl.in	第一ステップの MD 計算の入力ファイル
mdl_2.out	第一ステップの MD 計算の 2 回目の出力ファイル

これらのファイル名はルールを決めて組織的に命名するように心がけること。後日になって何の計算か分からなくなるようなことはないように！

次に、leap_vac.in を入力ファイルとして” tleap” コマンドを起動。

```
$ tleap -f leap_vac.in
$ set default writel4scale on
$ source leaprc.ff99SB
logFile leap.log
#
# -----leaprc for loading the ff99SB (Hornak & Simmerling) force field
# -----this file is updated for PDB format 3
#
#     load atom type hybridizations
#
addAtomTypes
    { "H"  "H"  "sp3" }

    (省略)

#
NHIS = NHIE
HIS  = HIE
CHIS = CHIE
$ model = loadPdb prk.pdb
$ saveamberParm model prk_vac.parm7 prk_vac.inpcrd
$ quit
$
```

prk_vac.parm7 と prk_vac.inpcrd の 2 種類のファイルが発生していることを確認。

1.3. 溶媒 Box システムのファイル作成

実際の生体高分子の挙動は水溶液中として観察され、溶質である生体高分子の電荷を中和するように対イオンが周囲に存在するはずである。水溶液中で たんぱく質は陰イオンである場合も陽イオンである場合もあるが、本計算では六面体の溶媒 Boxで タンパク質分子を取り囲み、対イオンとしてナトリウムイオンを配置する。溶媒 Boxシステム用の LEaP入力ファイル” leap_wat.in” を作成する。

```
set default writel4scale on
source leaprc.ff99SB

model = loadPdb prk.pdb
addions model Na+ 0

solvateBox model TIP3PBOX 8.0
saveamberParm model prk_wat.parm7 prk_wat.inpcrd
quit
```

addionsによりmodelシステムにナトリウムイオン (Na⁺) をシステム全体の電荷が中性 (0) となるように配置させる。電荷バランスに基づく 15イオンの位置を PRK周囲の静電ポテンシャルより計算する。solvateBoxにより溶質分子の最外殻から8Åの厚さを持つ六面体溶媒 Boxを設定する。

1.2. と同様に “tleap” を起動し PDBファイルを読み込む。

```
$ tleap -f leap_wat.in (以下、1.2と同様なメッセージが続く)
$ solvateBox model TIP3PBOX 8.0
rgn size: 41.736 41.736 56.099
svt size:   18.774   18.774   18.774
nx,ny,nz: 3 3 3
$ saveamberParm model prk_wat.parm7 prk_wat.inpcrd
$
```

```
$ quit
```

2. 孤立分子（真空）システムの構造最適化と分子動力学計算

2.1. 出発構造の緩和

通常、X線結晶構造解析より得られたPDBファイルには水素原子座標が含まれない。前の過程でLEaPにより自動的に水素原子が付加されるが、これはプログラムが残基ごとのデータベース座標を参照して水素原子座標を生成するため、分子全体では水素原子が原子間接触を発生させている可能性がある。さらに、出発構造座標が粗く安定構造でないこともある。このような場合、分子動力学計算に移行する前に短時間（小ステップ）の構造最適化計算を実施し分子構造を修正しておくとうい。このための入力ファイル（prk_vac_init_min.in）を以下に示す。

最適化計算のためのprk_vac_init_min.inファイル

```
prk: initial minimisation prior to MD
```

```
&cntrl
```

```
    imin = 1,  
    maxcyc = 500,  
    ncyc = 250,  
    ntb = 0,  
    igb = 0,  
    cut = 12
```

```
&end
```

構造最適化指定のために”imin=1“とする。この計算では真の最適化構造に到達する必要はないので、”maxcyc=500“により最適化ステップ数の上限を500とする。sanderは最大勾配法と共役勾配法の2種類の最適化アルゴリズムを提供する。最大勾配法は構造中の大きな歪みを取り除くのに有効であるが、構造が極小点に近づくにつれて収束が遅くなる。一方、極小点近傍では共役勾配法がより有効である。ncyc < maxcyc 時、最初にncycステップの最大勾配法が実施され、残りのステップ（maxcyc-ncyc）において共役勾配法が適用される。本計算では2種の最適化法が同じステップ数（250）実行される。

sanderはデフォルトで周期境界条件を指定するため、”ntb=0“により周期境界条件を無効にする。”igb=0“により仮想（GB）溶媒モデルを使用しない設定とする。但し、この設定はデフォルトであるため、この変数はなくてもよい。

各原子から原子間の非共有結合相互作用を計算するための計算範囲をカットオフ距離と呼ぶ。” cut=12 “により非結合カットオフの距離を12Åに設定する。この値は大きいほど良好な計算結果を与えるが、その一方で計算量を増大させる。

● 計算の実行

以下のスクリプトファイルを作成し、計算を実行する。

```
$mpirun -np 4 sander.MPI -O -i ./prk_vac_init_min.in -o ./prk_vac_init_min.out  
-c ./prk_vac.inpcrd -p ./prk_vac.parm7 -r ./ prk_vac_init_min.rst”
```

このファイル名の指定により、計算結果が prk_vac_init_min.out に、最終座標が prk_vac_init_min.rst に出力される。

最適化計算が終了後、prk_vac_init_min.rst に出力した最適化構造をもとに PDB ファイル “prk_vac_init_min.pdb” を作成する。得られた PDB ファイルを Rasmol のような分子グラフィックスソフトで表示し構造の変化状況を確認する。

```
$ambpdb -p prk_vac.parm7 < prk_vac_init_min.rst > prk_vac_init_min.pdb
```

2.2. 分子動力学計算

比較のために条件の異なる次の 2 種の短時間分子動力学計算を実行する。
カットオフ距離12Å、誘電率 1.0 条件の計算

```
”prk_vac_md1_12Acut.in”
```

```
prk MD in-vacuo, 12 angstrom cut off  
&cntrl  
  imin = 0, ntb = 0,  
  igb = 0, ntp = 100, ntwx = 100,  
  ntt = 1, tautp = 0.5,  
  tempi = 300.0, temp0 = 300.0,  
  nstlim = 100000, dt = 0.001,  
  cut = 12.0  
&end
```

カットオフ設定無し、誘電率 1.0 条件の計算

```
”prk_vac_md1_nocut.in”
```

```
prk MD in-vacuo, infinite cut off
&cntrl
  imin = 0, ntb= 0,
  igb = 0, ntp = 100, ntwx = 100, ntwr= 100,
  ntt = 1, tautp = 0.5,
  tempi = 300.0, temp0 = 300.0,
  nstlim = 100000, dt = 0.001,
  cut = 999
&end
```

今回は分子動力学計算のため、最適化指定を無効にするため“imin=0”とする。真空中の計算のため、溶媒設定に関わる周期境界条件および GB モデルを無効 (ntb=0、igb=0) に設定する。出力ファイル (mdout) およびトラジェクトリーファイル (mdcrd) にシミュレーション経過の情報を積分ステップ 100 ステップごとに出力する (ntpr=100, ntwx=100)。後述するように、この計算における時間ステップは 1fs のため 100 倍の 0.1ps ごとにエネルギー情報、温度、およびトラジェクトリー座標がファイルに出力される。さらに、“ntwr=100”の設定により、リスタートファイル (restrt) に 0.1ps ごとの座標が更新される。これにより、アクシデント等により計算が途中で終了しても、リスタートファイルを入力ファイルに指定することで計算が継続される。(注1) 定温シミュレーションの条件 (ntt=1) のもとで熱源とのカップリング時間を 0.5ps (tautp=0.5) に設定する。カップリング時間が短いほど粒子の運動エネルギーがより一定になるが全エネルギーの変動が激しくなり不自然なシミュレーションとなる。一般に 0.5-2.0 間の値を設定する。プログラムのデフォルト値は 1.0。初期温度と最終温度 (参照温度) いずれも 300K とする (tempi=300、temp0=300)。トラジェクトリーの計算における時間ステップを 1fs (dt=0.001) とし全体で 100000 ステップの計算 (nstlim=100000) を指定する。従って、シミュレーション時間は 100ps となる。以上の条件は 2つの計算で共通であるが、カットオフ距離の設定を 12Å (cut=12) または事実上カットオフなし (cut=999) とすることで互いに条件が異なる。

下記のとおりファイル名を指定して計算を実行する。今回は分子動力学計算に相当するため、“-x” オプションによりトラジェクトリーファイルを指定する。

```
#!/bin/csh
```

12 Å cut off run

```
mpirun -np 8 sander.MPI -O -i ./prk_vac_md1_12Acut.in -o ./prk_vac_md1_12Acut.out  
-c ./prk_vac_init_min.rst -p ./prk_vac.parm7 -x ./prk_vac_md1_12Acut.mdcrd  
-r ./prk_vac_md1_12Acut.rst
```

no cut off run

```
mpirun -np 8 sander.MPI -O -i ./prk_vac_md1_nocut.in -o ./prk_vac_md1_nocut.out  
-c ./prk_vac_init_min.rst -p ./prk_vac.parm7 -x ./prk_vac_md1_nocut.mdcrd  
□r ./prk_vac_md1_nocut.rst”
```

その結果、12Åカットオフ計算は 100psまで計算が実行されたがカットオフなしの計算は 5.6psでエラーメッセージが出されて計算が途中終了となった。

注 1 計算が中断したタイミングによっては、トラジェクトリーファイル (mdcrd) が最後まで出力されないことがある。その場合、リスタートファイルから計算を再開すると、リスタートファイルの時間帯に相当する 1 ステップ分が欠損した状態のトラジェクトリーファイルが生成する。従って、ptraj (trajin、およびtrajout) によりリスタートファイルを変換し、1 ステップ分のトラジェクトリーファイルを生成することが必要となる。

2.3. 結果の解析

まず、出力ファイル (prk_vac_md1_12Acut.out、prk_vac_md1_nocut.out) よりポテンシャルエネルギー値を抽出するため、以下のコマンドを実行する。

```
$grep TIME prk_vac_md1_12Acut.out | awk '{print $6}' > time1.dat  
$ grep EPtot prk_vac_md1_12Acut.out | awk '{print $9}' > epot1.dat  
$ grep EPtot prk_vac_md1_nocut.out | awk '{print $9}' > epot2.dat
```

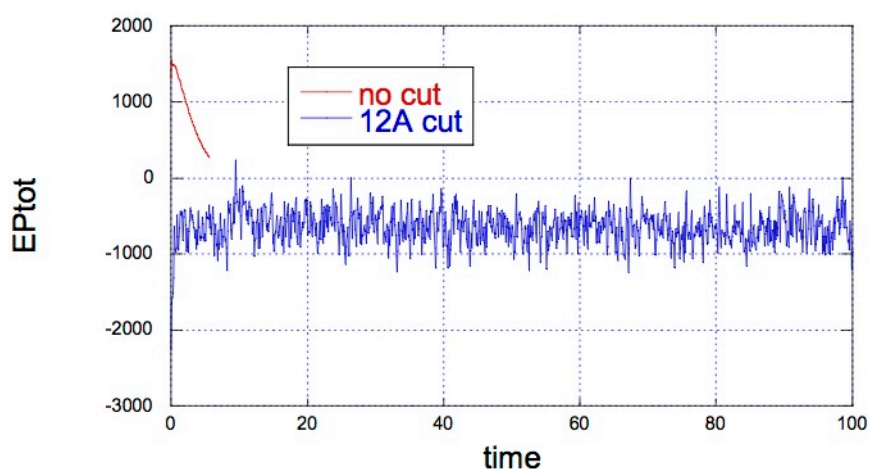
grepやawkコマンドの意味はここでは省略するが、time1.datには時間が、epot1.datとepot2.datにはそれぞれ、12Åカットオフとカットオフ未指定計算のポテンシャルエネルギーが出力される。但し、最後まで計算が終了したprk_vac_md1_12Acut.outファイルから抽出したエネルギーの最後の2つは、平均エネルギーと RMSエネルギーであるため、このファイル最後尾の2行の値を削除する。

出力ファイルからエネルギー値を抽出する別の方法として、20頁 pythonプログラム

process_out.pyを用いてもよい。実行方法を以下に示す。

`$python process_out.py`出力ファイル名

エラーがなければprocess_out.xlsが発生する。これには、時間、温度、全エネルギー、運動エネルギー、および位置エネルギーが出力されている。これらのファイルを PCにダウンロードし、エクセル等を用いて以下のようなグラフを作成する。



12Åカットオフ計算ではポテンシャルエネルギーが一定値付近で振動しているが、カットオフなし計算では12Åカットオフ計算よりエネルギーが約3000kcal/mol高い初期値を示し、その後急激に減少する途中で計算が終了した。

本計算は定温シミュレーションであり、プログラム内部で粒子を仮想的な熱源とカップリングさせ粒子の運動エネルギーの平均値が一定になるように調整している。シミュレーションが進行し、熱平衡が成立するとエネルギー保存則よりポテンシャルエネルギーも一定となるべきである。12Åカットオフ計算においてシミュレーションの初期からポテンシャルエネルギーが定常状態になっているように見えるが、ここではさらに詳細に結果を検討する。

これら 2つの計算の出力ファイルから初期構造エネルギーの内訳を比較すると。

12 Åカットオフ

NSTEP =	0	TIME (PS) =	0.000	TEMP (K) =	298.54	PRESS =	0.0
Etot =	-1701.9827	EKtot =	565.9592	EPtot =	-2267.9419		
BOND =	16.2543	ANGLE =	91.3593	DIHED =	386.7369		
1-4 NB =	162.4709	1-4 EEL =	-351.0609	VDWAALS =	-348.9801		
EELEC =	-2224.7223	EBBOND =	0.0000	RESTRAINT =	0.0000		

カットオフなし”

NSTEP =	0	TIME (PS) =	0.000	TEMP (K) =	298.54	PRESS =	0.0
Etot =	1877.8647	EKtot =	565.9592	EPtot =	1311.9055		
BOND =	16.2543	ANGLE =	91.3593	DIHED =	386.7369		
1-4 NB =	162.4709	1-4 EEL =	-351.0609	VDWAALS =	-351.6488		
EELEC =	1357.7937	EBBOND =	0.0000	RESTRAINT =	0.0000		

12

両者とも各エネルギー項がほぼ等しいが、唯一、静電相互作用エネルギー項（EELEC）がカットオフなしの計算において極めて大きな値となっている。

以上の現象の理由を検討するために、次にトラジェクトリーファイルを使った解析を行う。

AMBER パッケージの解析用プログラム ptraj を使用し、初期構造と各時間帯における構造との RMSd 値（各原子座標の違いの二乗平方根の平均値）を計算する。2 つの計算に対応する ptraj の入力ファイルを作成する。なお、ptraj については 5 章で解説する。

- 12Åカットオフ計算結果の ptraj 入力ファイル prk_vac_md1_12Acut.calc_rms
trajin prk_vac_md1_12Acut.mdcrd
rms first mass out prk_vac_md1_12Acut.rms time 0.1
- カットオフなし計算結果の ptraj 入力ファイル prk_vac_md1_nocut.calc_rms
trajin prk_vac_md1_nocut.mdcrd
rms first mass out prk_vac_md1_nocut.rms time 0.1

1行目にトラジェクトリーファイル名を指定する。2行目の意味は、原子量の重み（mass）をかけた rms計算を初期構造（first）に対し計算し、その結果をmdoutキーワードで指定するファイルに出力する。各トラジェクトリー座標の時間間隔はtimeキーワードで指定する（0.1ps、本計算でトラジェクトリーファイルは 100積分ステップごとに出力された）。

続けて、以下のようにしてトポロジーファイルと入力ファイル名を指定して ptrajを実行する。

```
$ ptraj prk_vac.parm7 < prk_vac_md1_12Acut.calc_rms
```

```
¥-/
-/-PTRAJ: a utility for processing trajectory files
/-¥
¥-/ Version: "AMBER 11.0 integrated" (4/2010)
-/-Executable is: "ptraj"
/-¥ Running on 1 processor(s)
¥-/ Residue labels:
```

```
DA5 DA  DA DA DA DA DA DA DA  DA3
DT5 DT  DT DT DT DT DT DT DT  DT3
```

```
PTRAJ: Processing input from "STDIN" ...
```

```
PTRAJ: trajin prk_vac_md1_12Acut.mdcrd Checking coordinates: prk_vac_md1_12Acut.mdcrd
Rank: 0 Atoms: 638 FrameSize: 15504 TitleSize: 81 NumBox: 0 Seekable 1
```

(省略)

```
ACTIONS 1> RMS to first frame using mass weighting
Dumping RMSd vs. time (with time interval 0.10) to a file named prk_vac_md1_12Acut.rms Atom
selection follows * (All atoms are selected)
```

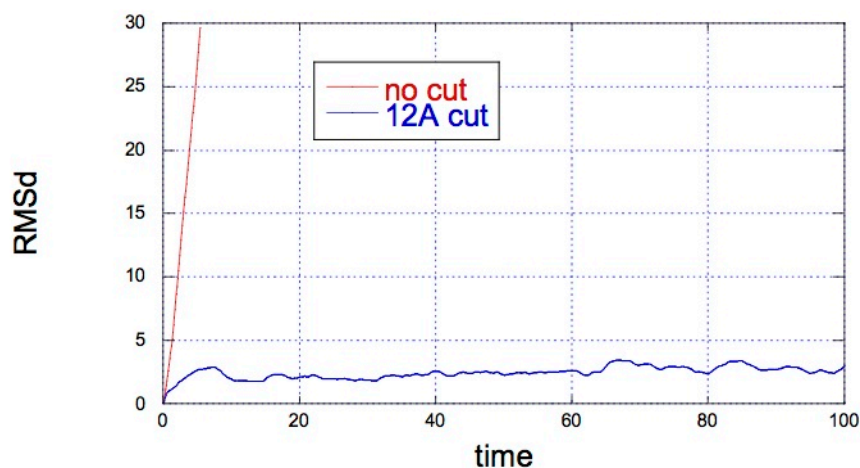
```
Processing AMBER trajectory file prk_vac_md1_12Acut.mdcrd
1% ..... 25% ..... 50% ..... 75% ..... 100%
```

```
PTRAJ: Successfully read in 1000 sets and processed 1000 sets. Dumping
accumulated results (if any) PTRAJ RMS: dumping RMSd vs time data
```

また、同様にして、

```
$ ptraj prk_vac.prmtop < prk_vac_md1_nocut.calc_rms
```

出力ファイルである” prk_vac_md1_12Acut.rms” および” prk_vac_md1_nocut.rms” を PC にダウンロードしエクセル等でプロットを作成する。



2Åカットオフ計算では、RMSd 値が 2Å近傍で増減しているが、カットオフ無し計算 では計算開始と同時に急激に RMSd 値が増大し 30Åまで達している。これは、シミュレーションの進行に伴って初期構造からどんどん構造が変化していることを意味する。 この状態の分子を視覚的に検討する

ために、分子グラフィックスにより分子構造の挙動を 観察すればよい。これには、次の 2 つのアプローチがある。

① トポロジーファイル(parm7)とトラジェクトリーファイル(mdcrd)を PC にダウンロードし、分子動力学計算のグラフィックスソフト VMDを使用して、分子の動きを観察する。その詳細は省略する。

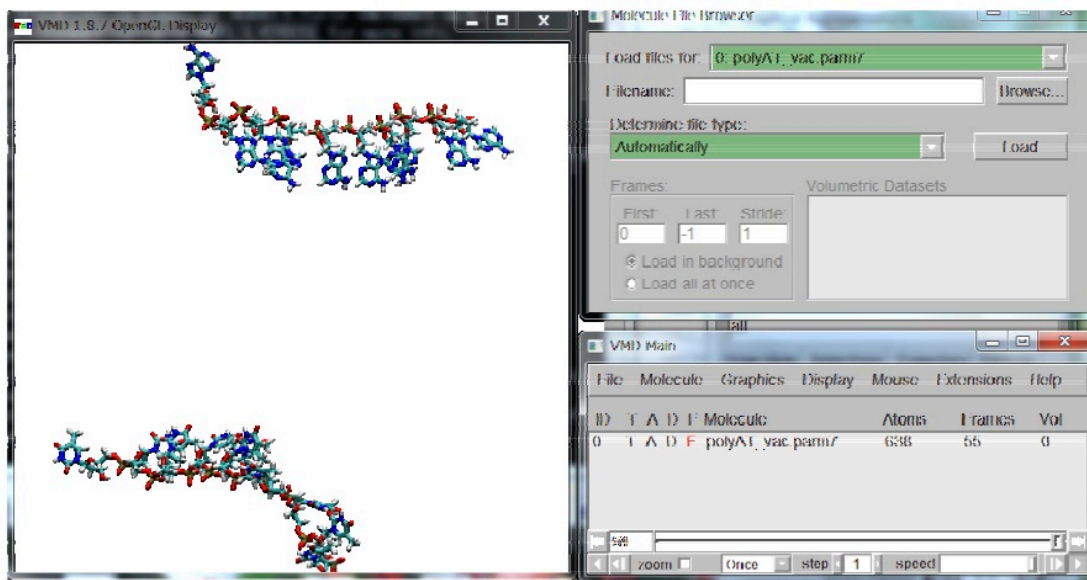
② ambpdbプログラムを実行し、rstrtファイルの最終構造から PDBファイルを作成する。

```
$ ambpdb -p prk_vac.parm7 < prk_vac_md1_12Acut.rst > prk_vac_md1_12Acut.pdb
```

```
$ ambpdb -p prk_vac.prmtop < prk_vac_md1_nocut.rst > prk_vac_md1_nocut.pdb
```

これらの PDBファイルを PCにダウンロードし Rasmolのようなグラフィックスで最終構造を検討する。

これより、12Åカットオフ計算は DNAの 2重らせん構造は維持されているが、カットオフなし計算では、2重らせん構造が壊れて分子鎖が離れていったことが分かる。一見、12Åカットオフ計算は安定な結果を与えたようであるが、今回は真空条件での計算である。一方、実際の DNAは負電荷を帯び、通常、陽イオンの存在下の溶媒中で 2重らせん構造を維持している。真空中ではリン酸骨格の負電荷が互いに反発し、2重らせん構造は壊れることが十分予想され、カットオフなしの計算結果が実際の挙動に近いものと考えられる。もちろん、真空中における DNA分子の状態(言い換えると、気化した DNA分子)は明らかに非現実的な状況であり、DNAに限らず生体高分子を対象とした分子動力学シミュレーションでは可能な限り、生体内に近い環境を設定する。特に、3.以降で解説する水和環境の設定が不可欠となる。



No cut off条件でのMD最終構造 (VMD画面)

amber 出力ファイルからエネルギー値を抽出するプログラム “process_out.py”

```
import sys
mdout = sys.argv
x = open(mdout[1], "r")
time = [ ]
temp = [ ]
etot = [ ]
ektot = [ ]
eptot = [ ]
for line in x.readlines():

    if "NSTEP" in line:
        a = line.split()
        temp.append(a[8])
        time.append(a[5])
        continue
    elif "Etot" in line:
        a = line.split()
        etot.append(a[2])
        ektot.append(a[5])
        eptot.append(a[8])
        continue
    x.close()
y = open("process_out.xls", "w")
print >> y, "time(ps)".rjust(10) +¥
      "Temp".rjust(10) +¥
      "Etot".rjust(14) +¥
      "EKtot".rjust(14) +¥
      "EPtot".rjust(14)

check = ""

for i in range(0, len(time)):

    if time[i] == check:
        continue
    print >> y, time[i].rjust(10) +¥
      temp[i].rjust(10) +¥
      etot[i].rjust(14) +¥
      ektot[i].rjust(14) +¥
      eptot[i].rjust(14)

    check = time[i]
y.close()
```

3. GB仮想溶媒和モデルの構造最適化と分子動力学計算

3.1. 出発構造の緩和

使用する初期構造は前回と同じ“prk_vac.inpcrd”である。以下の mdinファイル (prk_gb_init_min.in) の条件で計算を行う。ここで、GB仮想溶媒和モデルに関する変数“igb”を 1とし有効にする。”ntb”は後述する溶媒和ボックスに関するパラメーターであるが、値が”0”のため、この例では未設定である（省略可）。

最適化計算のための”prk_gb_init_min.in”ファイル

```
polyA-polyT 10-mer: initial minimisation prior to MD GB model
&cntrl imin = 1, maxcyc = 500, ncyc = 250, ntb = 0, igb = 1, cut = 12
&end
```

以下のスクリプトファイルにより計算を実行。

```
#!/bin/csh bsub -K -n 4 -o log/ "$LSF_BINDIR/openmpi_wrapper -np 4
$AMBERHOME/exe/sander.MPI -O -i ./prk_gb_init_min.in -o ./prk_gb_init_min.out
-c ./prk_vac.inpcrd -p ./prk_vac.parm7 -r ./prk_gb_init_min.rst"
```

最適化計算が終了後、”prk_gb_init_min.rst”に出力した最適化構造をもとに PDBファイル“prk_vac_gb_min.pdb”を作成する。得られた PDBファイルを Rasmolのような分子グラフィックスソフトで表示し構造の変化状況を確認する。

```
$ ambpdb -p prk_vac.parm7 < prk_gb_init_min.rst > prk_gb_init_min.pdb
```

igbオプション設定以外は前回の真空計算と同じ条件による最適計算である。両者のエネルギー (kcal/mol) を比較すると；

cd

孤立分子（真空）モデル		
初期構造	$-1.4356 \times 10_3$	$-2.2480 \times 10_3$
最適化構造	$-2.2679 \times 10_3$	$-3.3356 \times 10_3$
変化量	$-1.8323 \times 10_3$	$-1.0876 \times 10_3$

GB溶媒和を設定することで初期構造からエネルギーは非常に低い値となっている。言い換えれば真空条件では DNAの負電荷の静電反発により初期構造から不安定な状態になっている。GB溶媒和は対イオンを設定していないが、溶媒効果により静電反発が緩和されている。

3.2. 分子動力学計算

前回と同様にカットオフ条件の異なる 2種類の mdinファイルを作成。

□カットオフ距離12Å、誘電率 1.0条件の計算” prk_gb_md1_12Acut.in” 10-mer DNA MD Generalised Born, 12A cut off &cntrl imin = 0, ntb= 0, igb = 1, ntp = 100, ntwx = 100, ntwr = 100, ntt = 1, tautp = 0.5, tempi = 300.0, temp0 = 300.0, nstlim = 100000, dt = 0.001, cut = 12.0 &end

□カットオフ設定無し、誘電率 1.0条件の計算” prk_gb_md1_nocut.in” 10-mer DNA MD Generalised Born, no cut off &cntrl imin = 0, ntb= 0, igb = 1, ntp = 100, ntwx = 100, ntwr = 100, ntt = 1, tautp = 0.5, tempi = 300.0, temp0 = 300.0, nstlim = 100000, dt = 0.001,

cut = 999 &end

下記のスクリプトファイルに従って計算を実行。

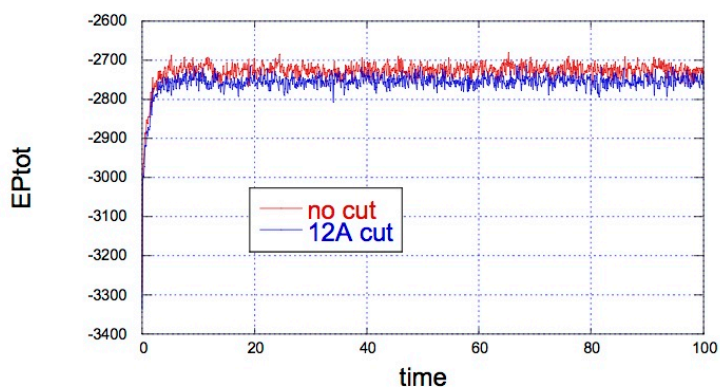
```
#!/bin/csh
```

```
# 12 A cut off run bsub -K -n 4 -o log/ "$LSF_BINDIR/openmpi_wrapper -np 4  
$AMBERHOME/exe/sander.MPI -O -i ./prk_gb_md1_12Acut.in -o ./prk_gb_md1_12Acut.out  
□c ./prk_gb_init_min.rst -p ./prk_vac.parm7 -x ./prk_gb_md1_12Acut.mdcrd  
-r ./prk_gb_md1_12Acut.rst"
```

```
# no cut off run bsub -K -n 4 -o log/ "$LSF_BINDIR/openmpi_wrapper -np 4  
$AMBERHOME/exe/sander.MPI -O -i ./prk_gb_md1_nocut.in -o ./prk_gb_md1_nocut.out  
□c ./prk_gb_init_min.rst -p ./prk_vac.parm7 □x ./prk_gb_md1_nocut.mdcrd  
-r ./prk_gb_md1_nocut.rst"
```


3.3 結果の解析

前回と同様に、mdout ファイル (prk_gb_md1_12Acut.out、prk_gb_md1_nocut.out) よりポテンシャルエネルギー値を抽出する。



明らかに前回の真空計算と異なり、今回はどちらのカット-オフ条件計算結果ともよく似た エネルギー値の定常状態を示している。若干、カット-オフなし設定計算の方でエネルギーが高くなっている。

再び、ptraj により RMSd 値を計算し、2 つの計算結果を比較する。入力ファイルは以下のとおりである。

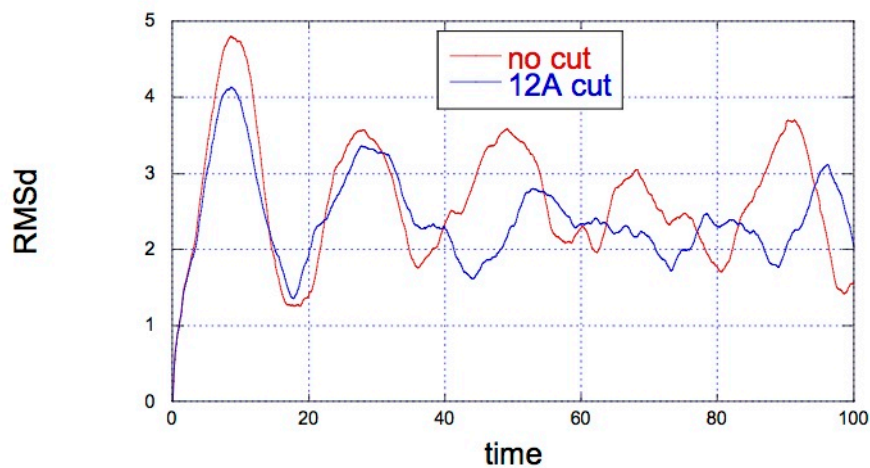
- 12Åカット-オフ計算結果のptraj入力ファイル “prk_gb_md1_12Acut.calc_rms”
trajin prk_gb_md1_12Acut.mdcrd
rms first mass out prk_gb_md1_12Acut.rms time 0.1
- カット-オフなし計算結果のptraj入力ファイル “prk_gb_md1_nocut.calc_rms”
trajin prk_gb_md1_nocut.mdcrd
rms first mass out prk_gb_md1_nocut.rms time 0.1

続けて、ptraj を実行。

```
$ ptraj prk_vac.parm7 < prk_gb_md1_12Acut.calc_rms
```

```
$ ptraj prk_vac.parm7 < prk_gb_md1_nocut.calc_rms
```

出力ファイル “prk_gb_md1_12Acut.rms” と “prk_gb_md1_nocut.rms” を PC にダウンロードしエクセル等で以下のようなグラフを作成する。



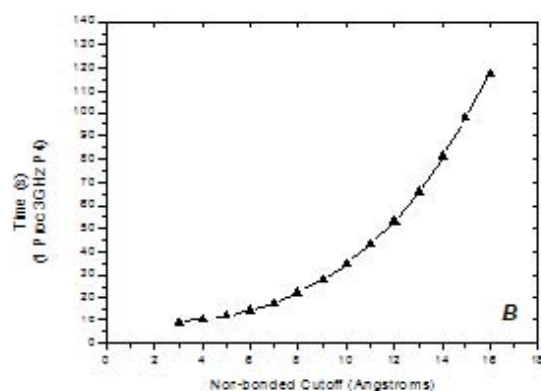
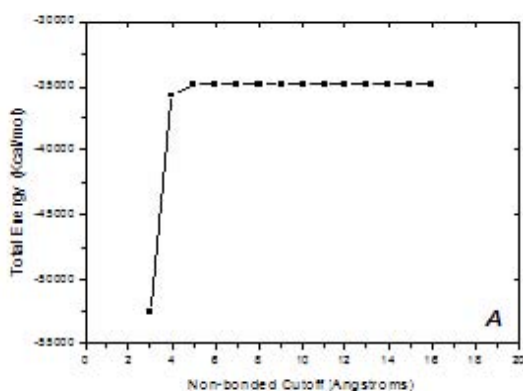
いずれの計算も、真空モデル計算と比較してはるかに安定な挙動を示している。

一般に RMSd値は1.5～2.0Åの間で変動する状態が望ましい。従って、本計算結果は若干大きすぎる。カットオフ設定は小さすぎると合理的な結果を与えないが、これを設定することで計算量が減少する。本計算の場合、12Åカットオフ設定によりカットオフなしと比べて約20%計算時間が短縮される。最後に、最終構造 (rstrt) ファイルまたはトラジェクトリー (mdcrd) ファイルを分子グラフィックスプログラムに読み込ませ、最終構造や構造変化の様子を分子グラフィックスで検討する。

4. 溶媒和Boxモデル計算の構造最適化と分子動力学計算

4.1. 溶媒和Boxシステムにおける平衡化前例の孤立分子システムの計算と異なり、溶媒 Boxは多数の水分子を含むはるかに複雑な計算対象となる。LEaPは機械的に溶質分子やイオンの位置を避けて水分子を生成する。その際、水分子とそれらの分子との相互作用を考慮しないため、当然、溶媒分子を含めた全体の初期構造には原子間相互作用が発生する。逆に、水分子と溶質分子の隙間や溶媒 Boxの周辺部に“真空の泡”が発生することもある。物理的に有意な結果は計算システムを熱平衡状態に設定することが必要であるため、溶媒 Boxシステム生成後、**全体構造の最適化と設定温度を徐々に上昇させる動力学計算を行う**平衡化プロセス計算を慎重に実施することになる。そして、平衡化に到達したことを十分慎重に確認した後、データ収集のための本来のシミュレーション (production run) に移行する。

4.2. カットオフ値の選択真空モデル計算でも示されてように分子動力学計算においてカットオフ値の設定は計算結果に大きな影響を与える。但し、溶媒環境では真空中ほど目立ったものではない。さらに particle mesh Ewald (PME) 法を適用することでカットオフの効果は減少する。下のプロットは異なるカットオフ値について 100ステップの最大勾配法による最適化計算を DNA 10量体に適用したときの全エネルギー値 (A) および計算時間 (B) を示す。これよりカットオフ値が少し増大するだけでエネルギーは急速に一定値に収束している。これより、カットオフが10Åより長い条件では計算時間の増大に見合うだけの結果が期待できないことが分かる。以降、この計算においてカットオフ値10Åを使用する。



4.3. 第一段階の位置束縛条件を伴った構造最適化計算

今回の溶媒 Box システムに対する構造最適化計算は、2 段階で行う。最初は溶質分子である DNA を固定したまま、溶媒分子（水）とナトリウムイオンの位置を最適化する。分子・構造の一部を“凍結”して残りの座標を動かすために AMBER には 2 つの手法が準備されている。

- ①凍結する原子に加わる力をゼロに設定する“belly”オプションを適用する。以前は好まれていた方法であるが、今回のような定圧周期境界 Box に対する計算の場合、不安定な挙動を与えるため、現在ではあまり好まれない。
- ② 参照座標を定義し、凍結する原子がこの参照座標から逸脱しないように位置束縛条件を課する。参照座標から凍結原子に対して強力な“スプリング”を設定し、構造最適化や分子動力学の時に凍結原子が参照座標からずれるとともに戻す力が生じるようにする。

今回の計算では位置束縛条件を適用する。

● 位置束縛条件を伴った構造最適化計算のための入力ファイル“prk_wat_min1.in”

```
polyA-polyT 10-mer: initial minimisation solvent + ions
&cntrl
  imin = 1,
  maxcyc = 1000,
  ncyc = 500,
  ntb = 1,
  ntr = 1,
  cut = 10
Hold the DNA fixed
500.0
RES 1 20
END
END
```

各パラメータの解説。

- ① `imin = 1` : 構造最適化を設定。
- ② `maxcyc = 1000` : 構造最適化ステップの最大ステップ数を 1000 とする。
- ③ `ncyc = 500` : 最初に 500 ステップの最大勾配法を実行する。その後、最大ステップ数 `maxcyc-ncyc` の共役勾配法に移行する。
- ④ `ntb = 1` : 定圧周期境界条件計算を適用。(`ntb > 0` の時は常に PME 法が起動する)
- ⑤ `cut = 10` : カットオフ距離を 10Å に設定。
- ⑥ `ntr = 1` : 位置束縛条件を適用する。位置束縛の内容は”&end” 移行の 5 行に記述される GROUP 入力で定義される。
- ⑦ GROUP 入力の解説: 1 行目はタイトル行。2 行目に束縛の力の定数として $500 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ を設定。3 行目より位置束縛条件を適用する原子として 1 残基から 20 残基の原子

を定義する。この例では溶質分子である DNA を構成する残基に相当する (21 残基以降は水分子となる)。最終 2 行で GROUP 入力の終了を定義。

計算を実行する。位置束縛条件計算のため、“-ref”キーワードで参照座標ファイル (refc) を定義する。この計算では参照ファイルは入力座標ファイル (inpcrd) と同じものを定義する。トポロジーファイルは 1.2. で設定した溶媒 Box に相当するファイルである。

```
#!/bin/csh bsub -K -n 4 -o log/ "$LSF_BINDIR/openmpi_wrapper -np 4
$AMBERHOME/exe/sander.MPI -O -i ./prk_wat_min1.in -o ./prk_wat_min1.out
-p ./prk_wat.parm7 -c ./prk_wat.inpcrd -r ./prk_wat_min1.rst -ref ./prk_wat.inpcrd"
```

4.4. 第二段階の構造最適化計算

第二段階として束縛条件をはずし十分なステップ数を指定した構造最適化計算を行う。

- 通常の構造最適化計算のための入力ファイル” prk_wat_min2.in”
polyA-polyT 10-mer: initial minimisation whole system
&cntrl imin = 1,
maxcyc = 2500,
ncyc = 1000,
ntb = 1,
ntr = 0,
cut = 10
&end

最大ステップ数を 2500 とする。この数が少なすぎると次の分子動力学計算が不安定な結果を与える。逆に、極端に多くとると計算時間が長くなる。

計算を実行する。2 段階の初期構造ファイル (inpcrd) に第一段階の最終構造ファイル (rstprt) を指定する。当然、この段階では参照構造ファイル (refc) は不要となる。

```
#!/bin/csh bsub -K -n 4 -o log/ "$LSF_BINDIR/openmpi_wrapper -np 4
$AMBERHOME/exe/sander.MPI -O -i ./prk_wat_min2.in -o ./prk_wat_min2.out
-p ./prk_wat.parm7 -c ./prk_wat_min1.rst -r ./prk_wat_min2.rst"
```

ambpdb プログラムにより初期構造と最終最適化構造の PDB ファイルを作成し、これらを分子グラフィックスで確認してもよい。

```
$ ambpdb -p prk_wat.parm7 < prk_wat.inpcrd > prk_wat.pdb $ ambpdb -p prk_wat.parm7 < prk_wat_min2.rst > prk_wat_min2.pdb
```

4.5. 位置束縛条件を伴う平衡化のための分子動力学計算

この段階では熱平衡に達するための分子動力学計算を実施する。以下にこの段階に関連する計算技術上の特徴を述べる。

- ① 温度設定を最初から 300Kとするのではなく、シミュレーション初期の 20psにおいて、初期温度を 0Kから 300Kまで温度を上昇させる。同時に溶質分子 (DNA) に” 弱い位置束縛を課する。
- ②データ収集を行う”production run” では定温・定圧条件の計算を実施する。しかし、低温での定圧計算は最初の数 psで正確なシミュレーションが行われないことが知られている。さらに束縛条件が導入されるため、より問題の多い結果を与える。従って、この段階における 20ps平衡化計算では定積条件を適用し、平衡化を確認した後、定圧条件へとスイッチする。
- ③ 溶媒和 Boxで使われる水分子は TIP3P水分子モデルと呼ばれもので、H-O-H結合角を固定することで計算量の削減を可能にしている。さらに TIP3P水分子モデルを用いるときに、水分子と溶質分子 DNAの両方の水素原子の動きを固定する必要がある。AMBERではこれを目的とした” SHAKE” 機能が装備されている。” SHAKE” を設定することで水素原子に関する結合の伸縮が制限される。さらに、水素原子の動きは分子の挙動で最も周期の短い動きに相当するために、時間ステップを 2fsと 2 倍することが可能になる。

もし、計算システムの分子分布状態や構造が偏り分子の歪みが大きい場合はさらに、より短い時間ステップ (0.5fs) で短時間の計算 (2fs) を長時間の平衡化計算に先立って実行する。

- 20ps、0K-300K の平衡化分子動力学計算の入力ファイル” prk_wat_md1.in” polyA-polyT 10-mer: 20ps MD with res on DNA

```
&cntrl imin = 0,
      irest = 0,
      ntx = 1,
      ntb = 1,
      cut = 10,
      ntr = 1,
      ntc = 2,
      ntf = 2,
      tempi = 0.0,
      temp0 = 300.0,
      ntt = 1,
      tautp = 0.2,
      nstlim = 10000, dt = 0.002,
      ntpr = 50, ntwx = 50, ntwr = 50
/
Keep DNA fixed with weak restraints
10.0
```

RES 1 20

END

END

各パラメータの解説。

- `imin = 1` : 構造最適化を無効にする (分子動力学計算を有効にする)。
- “`irest = 0, ntx = 1` : ボルツマン分布に従って各原子の初期速度を発生させる。分子動力学計算の最初の段階で設定するオプションとなる。
- “`ntb = 1` : 定積の周期境界計算を設定する。同時にPMEが設定される。
- “`cut = 10` : カットオフ距離を10Åに指定。
- “`ntr = 1` : GROUP入力に従った位置束縛条件を設定する。4.3. と同じように DNA分子に束縛が適用される。束縛の力の定数として10 kcal mol⁻¹ Å⁻²を設定。
- “`ntc = 2, ntf = 2` : SHAKEを有効にし、水素原子が関係する結合の結合伸縮を制限する。
- “`tempi = 0.0, temp0 = 300.0` : シミュレーションの初期温度設定を0Kとし、定常温度300Kまで昇温する。
- “`ntt = 1, tautp = 0.2` : 熱源とのカップリング時間を 0.2psとして定温シミュレーションを設定する。
- “`nstlim = 10000, dt = 0.002` : 時間ステップを 2fsとしてシミュレーションを 10000ステップ実施する。従って、シミュレーション時間は 20psとなる。
- “`ntpr = 50, ntwx = 50, ntwr = 50` : 出力関係オプション。” `mdout`” ファイルにエネルギー等の情報を 50ステップ (0.1ps) ごと (`ntpr`)、トラジェクトリーファイル” `mdcrd`” に原子座標を 50ステップごと (`ntwx`)、および最終構造ファイル” `restrt`” に原子座標と速度を 50ステップごと (`ntwr`) にそれぞれ、書き出す。” `mdout`” ファイルと” `mdcrd`” ファイルには各情報が追加されてゆくが、” `restrt`” ファイルでは設定時間ごとに内容が更新される。

計算を実行する。この段階では、初期構造ファイル” `inpcrd`” および参照座標ファイル” `refc`” に前段階で求めた最適化構造を用いることに注意すること。

```
#!/bin/csh bsub -K -n 4 -o log/ "$LSF_BINDIR/openmpi_wrapper -np 4 $AMBERHOME/exe/sander.MPI  
-O -i ./prk_wat_md1.in -o ./prk_wat_md1.out -p ./prk_wat.parm7 -c ./prk_wat_min2.rst  
-r ./prk_wat_md1.rst -x ./prk_wat_md1.mdcrd -ref ./prk_wat_min2.rst"
```

4.6. 平衡化のための分子動力学計算

100 psの平衡化計算を実施する。

- 100 ps、定圧・定温（300K）の平衡化分子力学計算の入力ファイルprk_wat_md2.in
polyA-polyT 10-mer: 100ps MD &cntrl imin = 0, irest = 1, ntx = 5, ntb = 2, pres0 = 1.0, ntp = 1, taup = 2.0, cut = 10, ntr = 0, ntc = 2, ntf = 2, tempi = 300.0, temp0 = 300.0, ntt = 1, tautp = 0.2, nstlim = 50000, dt = 0.002, ntpr = 50, ntwx = 50, ntwr = 50 &end

前段階と異なるパラメータ設定について解説。

1. “irest = 1, ntx = 5: 直前の計算を引き継ぐ再出発計算であることを指定し (irst = 1)、原子の出発座標、出発速度および周期境界Boxの大きさを “inpcrd” ファイルからアスキー形式で読み込むことを指定する (ntx = 5)。
2. “ntb = 2, pres0 = 1.0, ntp = 1, taup = 2.0: 参照圧力を 1atm (pres0 = 1.0) とした定圧周期境界条件を設定する (ntb = 2)。圧力緩和時間を 2psとした (taup = 2.0) 等方性の圧力制御を適用する (ntp = 1)。” taup” のデフォルト値は 0.2であり、1.0～5.0が推奨値とされる。
3. “nstlim = 50000” によってシミュレーション時間を100psに設定する。

計算を実行する。ここでは参照ファイルは不要となる。

```
#!/bin/csh bsub -K -n 4 -o log/ "$LSF_BINDIR/openmpi_wrapper -np 4
$AMBERHOME/exe/sander.MPI -O -i ./prk_wat_md2.in -o ./prk_wat_md2.out
-p ./prk_wat.parm7 -c ./prk_wat_md1.rst -r ./prk_wat_md2.rst
-x ./prk_wat_md2.mdcrd"
```

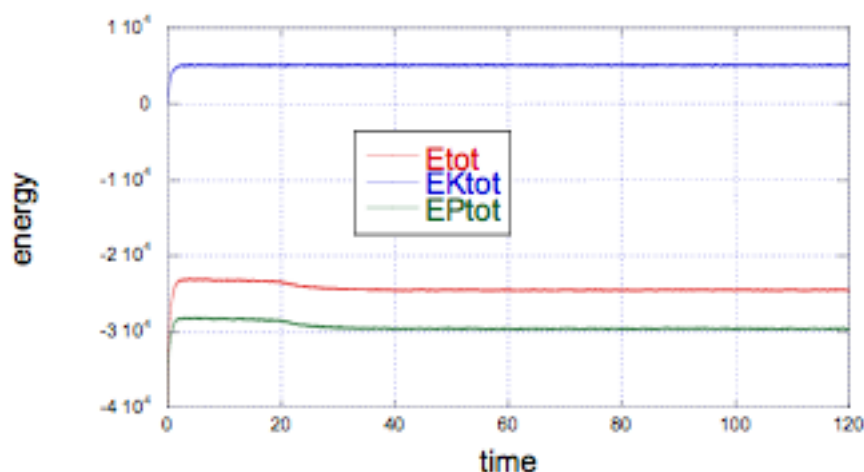

4.7. 平衡化の確認のための結果の解析

平衡化を確認するための以下の情報を確認する。

- ポテンシャルエネルギー、運動エネルギー、全エネルギー
- 温度
- 圧力
- 体積
- 密度
- RMSd

4.7.1. “mdout” 情報の解析 各種エネルギー情報の抽出方法については既に解説した。溶媒和ボックス計算でのみ得られる圧力、体積および密度情報の抽出方法については、各自、工夫下さい (grepコマンドの利用、または process_out.pyを変更する等)。

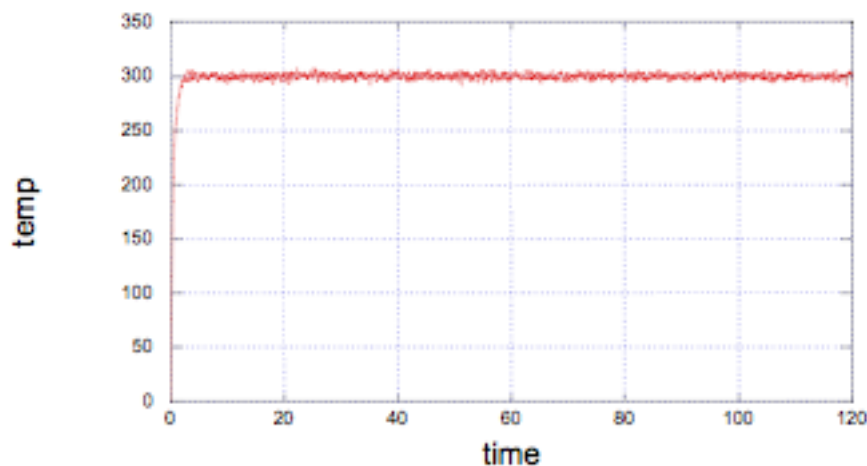
- まず、全エネルギー (Etot)、ポテンシャルエネルギー (EPtot)、運動エネルギー (EKtot) をプロットする。



最初の 0Kから

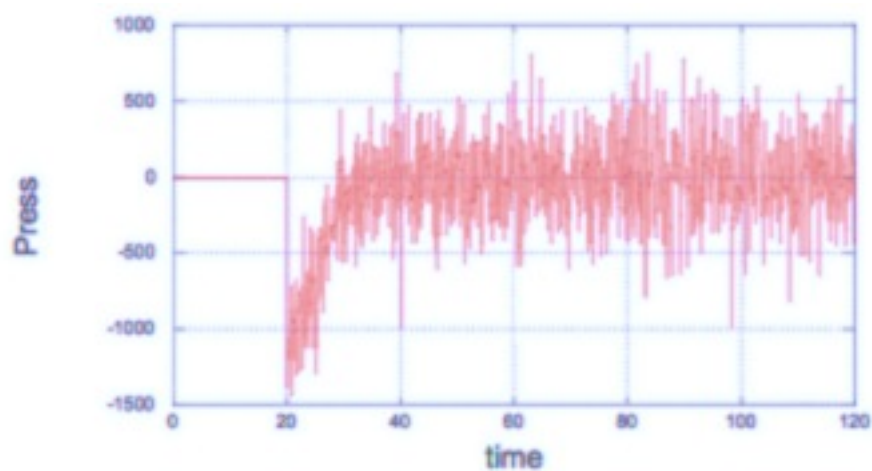
300Kまでの加熱過程ですべての値が上昇している。直後に運動エネルギーは一定値になる。一方、全エネルギー (ポテンシャルエネルギーと運動エネルギーの和) と位置エネルギーは定積過程の 0ps~20psの途中で一定になり、定圧にスイッチし定圧へ移行する間に減少している。これらの値は 40ps以降、定常状態になっている。

- 温度 (Temp) をプロットする。



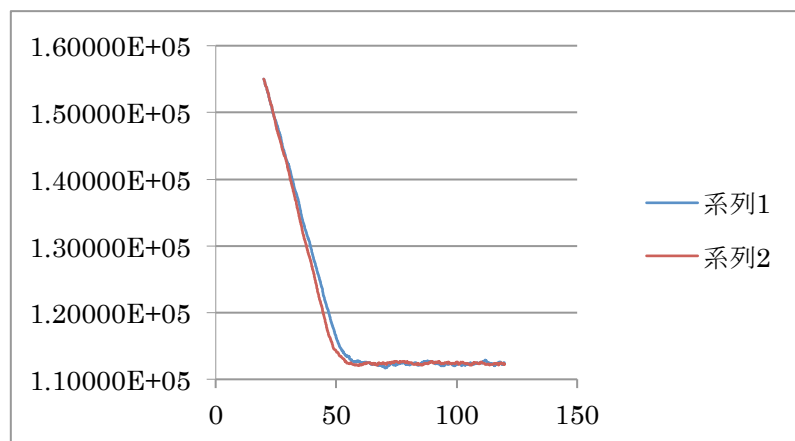
運動エネルギーの挙動と関連して最初の 5psで 300Kの一定値に到達している。プログラムの温度調整機構が機能していることが分かる。

- 圧力 (Press) をプロットする。



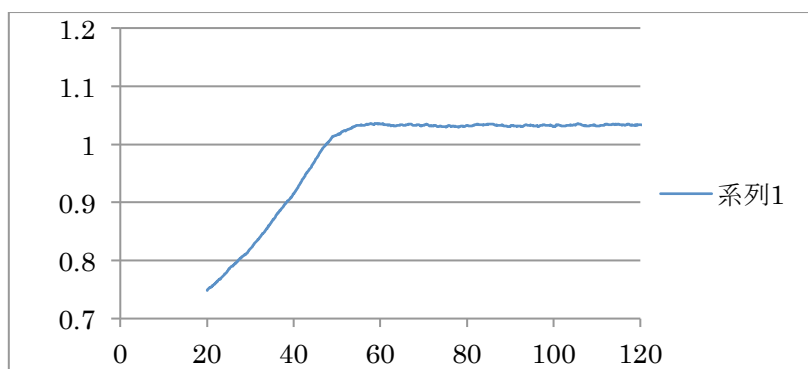
最初の 20psまでは定積シミュレーションのため圧力はゼロである。20psで定圧にスイッチした直後に圧力がマイナスにジャンプしている。マイナス圧力では周期境界 Boxはその大きさを縮小する作用を受ける、一方、プラス圧力は Boxの大きさを増大させる。圧力変化の振幅は極めて大きい。50ps以降からは定常状態に到達している。圧力調整により平均値は 1atmとなっているはずである。

- 体積 (Volume) をプロットする。但し、オリジナルファイルの最初の 20ps のデータは定積計算結果のため体積情報が含まれていないため、最初の 101 行を削除した修正ファイルを作成する。
-



溶媒 Boxが緩和するに従って体積 (\AA^3) は急激に減少した後、一定値を維持している。明らかに平衡状態に到達している。

- 密度 (Density) をプロットする。同様に、オリジナルファイルの最初の 20ps のデータは密度情報が含まれていないため、最初の101行を削除した修正ファイルを作成する。
-



この結果より平衡状態における密度はおおよそ 1.04 gcm^{-3} である。300Kでの純粋の密度は 1.00 gcm^{-3} のため、DNAを溶質として加えること生じた静電相互作用により密度が約4%増加せたと解釈できる。

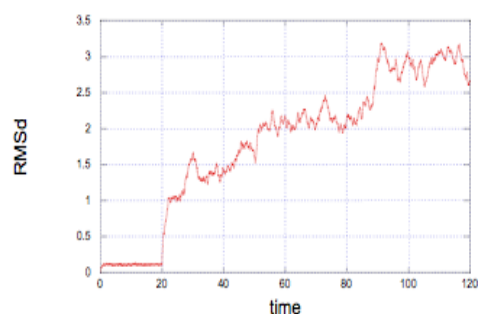
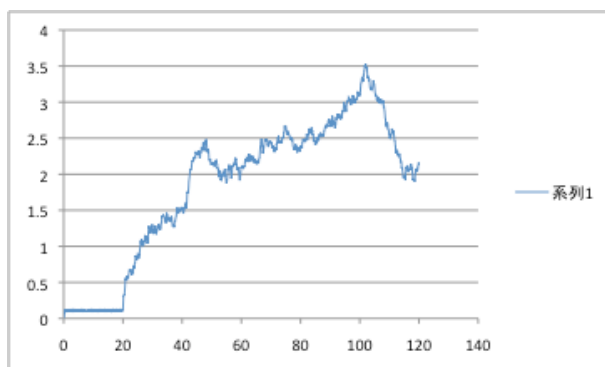
4.7.2 トラジェクトリーファイル “mdcrd” 情報の解析

ptraj により出発構造（最適化計算の最終構造）に対する RMSd を計算し、シミュレーション構造の妥当性を検討する。構造全体ではなく骨格原子（P, O3*, O5*, C3*, C4*および C5*）座標を参照として RMSd を求めた。トラジェクトリーファイルには定積計算と定圧計算の 2 つのファイルを読みこんだ。分子動力学計算の出力オプションで 50 ステップごとに座標を出力したため、時間ステップは 0.1ps となる。入力ファイルを以下に示す。

- ptraj入力ファイル” prk_wat_calc_backbone_rms.in”
trajin prk_wat_md1.mdcrd
trajin prk_wat_md2.mdcrd rms first out prk_wat_backbone.rms
@P,O3',O5',C3',C4',C5' time 0.1

計算を実行する。

```
$ ptraj prk_wat.prmtop < prk_wat_calc_backbone_rms.in
```



ls -al出力ファイル” prk_wat_backbone.rms” に出力された RMSdをプロットする。



溶質に束縛条件が課せられている最初の 20ps までは RMSd 値は低い値のままである。DNA 分子が緩和するに従って、RMSd 値が増加し、やがて定常状態に到達している。RMSd 値より分子構造は破綻していないと考えられる。

ここで得られたトラジェクトリーファイルを VMD 等で観察すると、水分子が周期境界 Box から拡散しているような画面が得られる。このような水分子は隣接の Box に移動しているため計算上、Box 内部にとどまっている。そこで、計算結果を“視覚的に”改良するため ptraj によりトラジェクトリーファイルのすべての座標を基準 Box に移動しなおす re-image 操作を行う。入力ファイルを以下に示す。

- ptraj入力ファイル” prk_wat_reimage.ptraj”
 trajin prk_wat_md1.mdcrd
 trajin prk_wat_md2.mdcrd
 trajout prk_wat_md_reimaged.mdcrd
 center :1-20
 image familiar go

このコマンドにより DNA分子（残基1～20）の重心を原点に平行移動し、それに従って水分子も移動させる（re-image）。オリジナルの 2つのトラジェクトリーファイルは” re-image” 後、” prk_wat_md_reimaged.mdcrd” に書き出される。計算を実行。

```
$ ptraj prk_wat.parm7 < prk_wat_reimage.ptraj
```

トポロジーファイル” prmtop” とトラジェクトリーファイル” mdcrd” を PCにダウンロードし VMD等で分子の挙動を観察する。

4.8 データ収集のための” Production Run” 実行

4.8.1 分子動力学計算の実行

平衡化が確認されたら、続けてデータ収集や現象を解析することを目的とした本格的な” production run” を実施する。計算条件は最後の定圧計算と同じであるが、十分なアンサンブル平均を得るために ns オーダーのシミュレーション時間を設定する。ところが、計算時間が長期化するに伴ってトラブルが発生する可能性が増大する。それは計算の異常終了だけが原因でなく、ディスクやネットワークのトラブルで計算が停止するアクシデントも少なくない。その結果、長時間を費やした計算がすべて無駄になることもある。また、今回のように溶媒 Box 条件の計算では最終的に発生する” mdout” や” mdcrd” ファイルは膨大な量となるため、次のデータ解析においてファイルの取り扱いが困難になる。

そのようなトラブルからの被害を最小限にとどめるために、本計算では 200psの計算を 5 回にわたって実行し、合計 1.0nsの計算を自動的に起動する手段を採用する。

まず、シミュレーション時間を 200psとした入力ファイル” prk_wat_md3_1000ps.in”
示す。

- 200ps、定圧・定温 (300K) 分子動力学計算入力ファイル” prk_wat_md3_1000ps.in”
polyA-polyT 10-mer: 200ps production MD
&cntrl imin = 0, irest = 1, ntx = 5,
ntb = 2, pres0 = 1.0, ntp = 1,
taup = 2.0,
cut = 10, ntr = 0,
ntc = 2, ntf = 2,
tempi = 300.0,
temp0 = 300.0,
ntt = 1, tautp = 0.2,
nstlim = 100000, dt = 0.002,
ntpr = 500, ntwx = 500, ntwr = 500
&end

出力ファイル” mdout” とトラジェクトリーファイル” mdcrd” の抽出ステップを 1psに設定し、前回の平衡化計算と比べて出力量を10%に抑えた。

5回の計算を手動で起動させてもよいが、ここでは、一連の計算を自動的に処理するシェル・スクリプトファイルを作成した。さらに一回計算が終了するたびに” gzip” を実行し、トラジェクトリーファイルを圧縮する。以下にスクリプトファイルを示す。

- 自動計算のためのシェル・スクリプトファイル” run_prk_md_1000ps.x”
#!/bin/csh
set MDSTARTJOB=1

set MDENDJOB=5
set MDCURRENTJOB=\$MDSTARTJOB
set MDINPUT=0
cp prk_wat_md2.rst prk_wat_md3_0.rst

echo -n "Starting Script at: " date echo ""
while (\$MDCURRENTJOB <= \$MDENDJOB)
echo -n "Job \$MDCURRENTJOB started at: "
date @ MDINPUT = \$MDCURRENTJOB - 1

bsub -K -n 4 -o log/ "\$LSF_BINDIR/openmpi_wrapper -np 4 \$AMBERHOME/exe/sander.MPI
-O -i ./prk_wat_md3_1000ps.in -o ./prk_wat_md3_\$MDCURRENTJOB.out

```
-c ./prk_wat_md3_$MDINPUT.rst -r ./prk_wat_md3_$MDCURRENTJOB.rst
-p ./prk_wat.parm7 -x prk_wat_md3_$MDCURRENTJOB.mdcrd"
```

```
gzip -9 -v prk_wat_md3-$MDCURRENTJOB.mdcrd
echo -n "Job $MDCURRENTJOB finished at: "
date
@ MDCURRENTJOB = $MDCURRENTJOB + 1
```

```
end
echo "ALL DONE"
```

最初に実行する場合は” chmod” コマンドで実行ファイルとして定義する。

```
$ chmod +x run_prk_md_1000ps.x
```

Time span (ps)	mdout	rstrt	Mdcrd
120-320	prk_md1.out	prk_md1.rst	prk_md1.mdcrd
320-520	prk_md2.out	prk_md2.rst	prk_md2.mdcrd
520-720	prk_md3.out	prk_md3.rst	prk_md3.mdcrd
720-920	prk_md4.out	prk_md4.rst	prk_md4.mdcrd
920-1120	prk_md5.out	prk_md5.rst	prk_md5.mdcrd

シェルスクリプトをバックグラウンドで実行。

```
$ run_prk_md_1000.x >& log&
```

発生する各出力ファイル名とシミュレーション時間帯との対応は以下のとおり。トラジェクトリーファイルは圧縮ファイル (*.gz) で得られる。

注意) この例はあくまでもトレーニングを前提としたもので、実際の計算の場合、200 ps の分割時間による1 nsのシミュレーション時間は短すぎる。現在の計算機の能力を踏まえれば全体で数十 ns程度の MD計算が要求されるため、計算対象の分子サイズに合わせて1nsかそれ以上で分割すべきである。

4.8.2. 結果の解析

平衡化の処理と同様にエネルギー、温度、圧力などの時間変化をプロットで確認する。これらの値がほぼ定常値を示していれば、シミュレーションは正常に行われたとみなす。

ptrajにより次にリン酸骨格原子の RMSdの計算を行い“prk_1000ps.rmsfit”に結果を出力する。同時にすべてのトラジェクトリーファイルに“re-image”処理を加え水分子の座標を削除した新たなトラジェクトリーファイル“prk_1000ps_nowat.mdcrd”を作成する。

RMSd計算、re-image、および溶媒除去のためのptraj入力ファイル

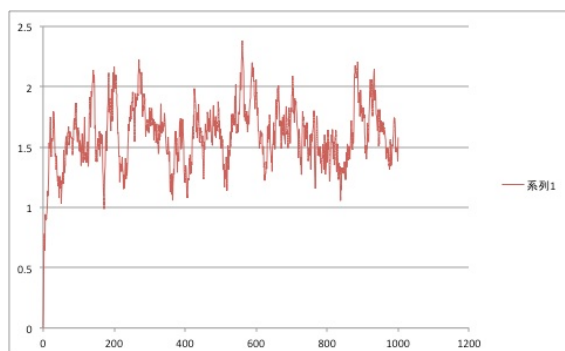
```
"combine_mdcrds_and_strip.ptraj"
```

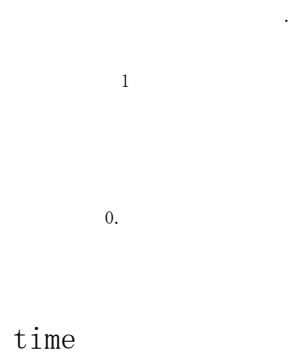
```
trajin prk_wat_md3_1.mdcrd.gz
trajin prk_wat_md3_2.mdcrd.gz
trajin prk_wat_md3_3.mdcrd.gz
trajin prk_wat_md3_4.mdcrd.gz
trajin prk_wat_md3_5.mdcrd.gz
trajout prk_1000ps_nowat.mdcrd nobox
rms first out prk_1000ps.rmsfit @P,O3',O5',C3',C4',C5' time 1.0
center :1-20
image familiar
strip :WAT
```

ptrajを実行。

```
$ ptraj prk_wat.prarm7 < combine_mdcrds_and_strip.ptraj
```

“prk_1000ps.rmsfit” をプロットする。





VMD により “prk_1000ps_nowat.mdcrd” 座標の分子グラフィックスを表示する。但し、水分子を設定していない DNA と対イオンだけのトポロジーファイルが必要となるため、以下の LEaP 入力ファイル “leap_cion.in” から “prk_cion.parm7” を作成する。

```
set default writel4scale on
source leaprc.ff99SB
model = loadPdb nuc.pdb
addions model Na+ 0
saveamberParm model prk_cion.parm7 prk_cion.inpcrd

quit
```

tLeapを実行。

```
[tyui@seneca01 leap_work]$ tLeap -f leap_cion.in
```

“prk_cion.parm7” と “prk_1000ps_nowat.mdcrd” を PCにダウンロードし VMDを用いて分子の挙動を観察する。周囲に水分子が配置された溶媒和環境のもとでも 300Kで polyA-polyT 10量体の立体構造にある程度のゆらぎ（ダイナミクス）が発生することを観察しなさい。

次に、ptrajを用いて平均構造を計算する。

□平均構造作成のための ptraj入力ファイル” prk_calc_av_struct.ptraj”

```
trajin prk_wat_md3_1.mdcrd.gz
trajin prk_wat_md3_2.mdcrd.gz
trajin prk_wat_md3_3.mdcrd.gz
trajin prk_wat_md3_4.mdcrd.gz
trajin prk_wat_md3_5.mdcrd.gz
average prk_wat_1000ps_average.pdb pdb nowrap
```

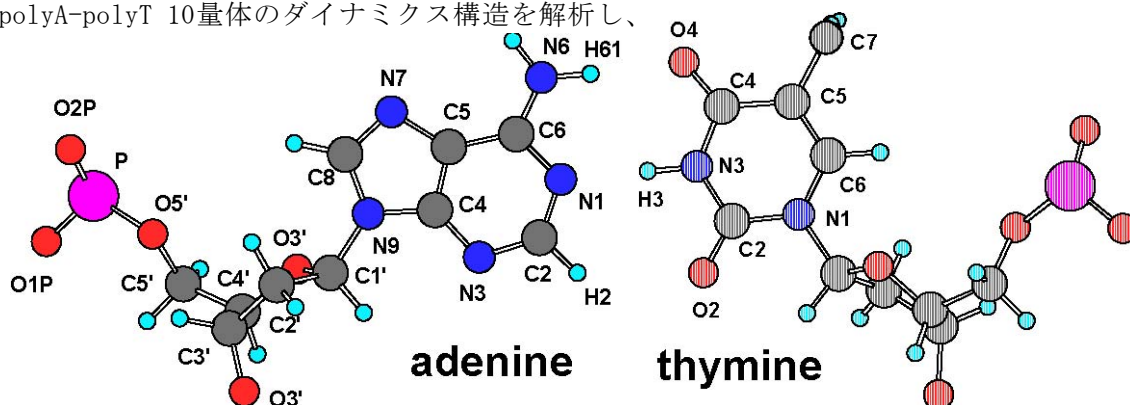
ptrajを実行。

```
$ ptraj prk_wat.parm7 < prk_calc_av_struct.ptraj
```

平均構造の PDBファイル” prk_wat_1000ps_average.pdb” を PCにダウンロードし、グラフィックスプログラムで平均構造を表示させる。

polyA-polyTにおける原子ラベル

polyA-polyT 10量体のダイナミクス構造を解析し、



溶媒和環境における DNAオリゴマーの構造ゆらぎ

を評価しなさい。末端効果の影響が最も小さい真ん中部分の2～3残基を選ぶとよい。また、軌跡ファイルとして、水分子が除去された” prk_1000ps_nowat.mdcrd” を利用すれば ptrajの処理速度が速くなる（同時にトポロジーファイルには” prk_cion.parm7” を使用）。

解析対象とする構造パラメータの例として；

1. 塩基間水素結合 (adenine N6-H6---- O4 thymine、thymine N3-H3---N1 adenine)
2. 糖残基に対する塩基面の回転

$\chi = \text{O4}' - \text{C1}' - \text{N1} - \text{C2}$ (pyrimidines or thymine), $\text{O4}' - \text{C1}' - \text{N9} - \text{C4}$ (purines or adenine)

3. 糖-リン酸骨格の結合回転 $\text{O3}' - \text{P} - \text{O5}' - \text{C5}' - \text{C4}' - \text{C3}' - \text{O3}'$ (但し、 $\text{C4}' - \text{C3}'$ は糖残基の五員環を含めたendocyclic bond)

$\alpha = \text{O3}' - \text{P} - \text{O5}' - \text{C5}'$; $\beta = \text{P} - \text{O5}' - \text{C5}' - \text{O4}'$; $\gamma = \text{O5}' - \text{C5}' - \text{C4}' - \text{C3}'$; $\delta = \text{C5}' - \text{C4}' - \text{C3}' - \text{O3}'$; $\epsilon = \text{C4}' - \text{C3}' - \text{O3}' - \text{P}$; $\zeta = \text{C3}' - \text{O3}' - \text{P} - \text{O5}'$

4. 以下に定義される糖残基の五員環の擬似回転位相角 p

$$\text{ここで、} \nu_0 = \tan(p) = \frac{(\nu_4 + \nu_1) - (\nu_3 + \nu_0)}{2\nu_2(\sin(36) + \sin(72))}$$

$\text{C4}' - \text{O4}' - \text{C1}' - \text{C2}'$; $\nu_1 = \text{O4}' - \text{C1}' - \text{C2}' - \text{C3}'$; $\nu_2 = \text{C1}' - \text{C2}' - \text{C3}' - \text{C4}'$;
 $\nu_3 = \text{C2}' - \text{C3}' - \text{C4}' - \text{O4}'$; $\nu_4 = \text{C3}' - \text{C4}' - \text{O4}' - \text{C1}'$

実際の解析結果の一部を次に示す。

- 水素結合解析の入力ファイル例” hb.in”
trajin prk_1000ps_nowat.mdcrd

```
donor mask :5,6@N1
donor mask :15,16@O4
acceptor mask :5,6@N6 :5,6@H61
acceptor mask :15,16@N3 :15,16@H3
```

```
hbond distance 3.5 angle 120.0
```

水素結合解析コマンドの出力は画面出力がデフォルトなので、リダイレクションで出力ファイル指定する (hb.out)。

```
[tyui@seneca01 analysis]$ ptraj prk_cion.parm7 < hb.in > hb.out
```

出力の最終部分に水素結合の形成状況が表示される。

(省略)

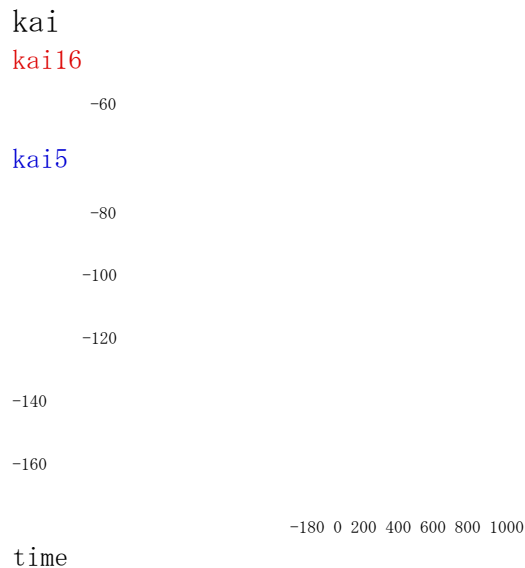
```
DONOR ACCEPTORH ACCEPTOR
atom# :res@atom atom# :res@atom atom# :res@atom %occupied distance angle | 148 :5@N1 |
501 :16@H3 500 :16@N3 | 100.00 2.938 ( 0.11) 16.14 ( 9.20) | 180 :6@N1 | 469 :15@H3 468 :15@N3
| 99.90 2.954 ( 0.11) 15.60 ( 8.33) | 467 :15@O4 | 178 :6@H61 177 :6@N6 | 97.70 2.974 ( 0.16)
16.94 ( 9.26) | 499 :16@O4 | 146 :5@H61 145 :5@N6 | 95.90 3.019 ( 0.18) 17.54 (10.00) | 180 :6@N1
| 501 :16@H3 500 :16@N3 | 2.60 3.325 ( 0.09) 53.59 ( 4.32) | 467 :15@O4 | 146 :5@H61 145 :5@N6
| 1.10 3.238 ( 0.14) 52.31 ( 7.88) | 499 :16@O4 | 178 :6@H61 177 :6@N6 | 0.20 3.248 ( 0.03)
56.46 ( 0.57) | 180 :6@N1 | 146 :5@H61 145 :5@N6 | 0.10 3.455 ( 0.00) 59.21 ( 0.00)
```

□5,16番残基の χ パラメータの出力ファイル例

```
trajin prk_1000ps_nowat.mdcrd

dihedral kai16 :16@O4' :16@C1' :16@N1 :16@C2 out kai16.xls dihedral
kai5 :5@O4' :5@C1' :5@N9 :5@C4 out kai5.xls
```

時間に対する χ 値の時間変化が” out” キーワードで指定した” kai16.xls”
と” kai5.xls” ファイルに出力される。



4. 9. 応用問題－DNAの熱変性－

DNAはある温度以上で塩基間の水素結合が切断され、二重らせん構造が壊れる。定温・定積条件（NVTアンサンブル）、昇温速度 1 K/psのもとで 800Kまで設定温度を上昇させ、DNAの構造変化を観察し、変性時における構造パラメータ変化をモニターしなさい。以下の入力ファイルを参考にして、既に計算が終了した熱平衡化後 120psの座標（poly_AT_wat_md2.rst）を出発構造として昇温 MD計算を継続すること。nmropt=1オプションを指定し、11行目以降に温度制御条件を記載する。

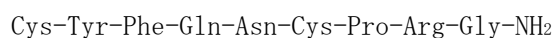
```
z 昇温 MD計算の入力ファイル polyA-polyT 10-mer: heating MD up to 800 K for 500 ps &cntrl
  imin = 0, irest = 1, ntx = 5,   ntb   = 1, cut   = 10, ntr   = 0,   ntc   = 2,
  ntf   = 2,   tempi = 300.0, temp0 = 800.0, ntt   = 1, tautp = 0.2,   nstlim =
  250000, dt = 0.002,   ntp = 500, ntwx = 500, ntwr = 500,   nmropt = 1 &end &wt
  type = 'TEMP0', istep1 = 1, istep2 = 250000,   value1 = 300.0, value2 = 800.0, &end
&wt
  type = 'END' &end
```

5. ペプチドと糖鎖への応用

5.1. ペプチドホルモン

5.1.1. トポロジーファイルの作成

バソプレッシン (Vasopressin) は、ヒトを含む多くの動物で見られるペプチドホルモンである。下垂体後葉から分泌され、抗利尿ホルモン (Antidiuretic hormone, ADH)、血圧上昇ホルモン、バゾプレシン、バゾプレッシンとも呼ばれる。(Wikipediaより) アミノ酸配列を以下に示す。



C末端は NH_2 でキャップされ、さらに 2箇所のシステイン残基の間に SS結合が形成される。

ここでは、バソプレッシンを例として、アミノ酸配列情報をもとにペプチド鎖のトポロジーファイルを作成する方法について述べる。なお、より一般的な PDBファイル読み込みによるトポロジーファイル作成については第4章 “タンパク質分子動力学計算の基礎知識” を参考にすること。

□初期トポロジーファイル作成のための LEaP入力ファイル

```
set default writel4scale on
set default disulfide off source leaprc.ff99SB

model = sequence { NCYX TYR PHE GLN ASN CYX PRO ARG GLY NHE }
bond model.1.SG model.6.SG

saveamberParm model VPN_vac.parm7 VPN_ini.inpcrd

quit
```

この入力ファイルから構造緩和計算に必要な “VPN_vac.parm7” と “VPN_vac.inpcrd” が作成される。N末端とアルギニン側鎖に正電荷が設定され、C末端をキャップする NH_2 基はイオン化していない。但し、この電離状態の正当性については不明である。

この LEaPより得られる初期構造は環状構造に相当しないので、SS結合は異常な長さ (18 Å) となっている。そこで、真空条件の構造最適化により立体構造を緩和させる (バソプレッシンの立体構造は trypsin-vasopressin 複合体の PDBデータ (1YF4) より得られる)。最適化構造から適当な名称の PDBファイル (VPN_opt.pdb) を作成する。N末端システインの残基名称が NCYXに定義されているが、何故かこれでは次の LEaPでエラーが発生するため、CYXに修正する。

修正した” VPN_opt.pdb” をもとに溶媒和分子動力学計算に必要なトポロジーファイルと初期座標ファイルを作成する。

REMARK

ATOM	1	N	NCYX	1	7.691	7.874	4.010	1
ATOM	2	H1	NCYX	1	7.446	7.198	4.724	1
ATOM	3	H2	NCYX	1	6.971	7.862	3.298	1
REMARK								
ATOM	4	H3	NCYX	1	7.740	8.790	4.431	1
ATOM		N	CYX	1	7.691	7.874	4.010	1
(以下、省略)								



ATOM	2	H1	CYX	1	7.446	7.198	4.724	1
ATOM	3	H2	CYX	1	6.971	7.862	3.298	1
ATOM	4	H3	CYX	1	7.740	8.790	4.431	1

□溶媒和分子動力学計算のための LEaP入力ファイル set default write14scale on
 set default disulfide off source leaprc.ff99SB

ATOM 5 CA CYX 1 8.992 7.478 3.422 1
 model = loadPdb VPN_opt.pdb
 (以下、省略) 1. 1. SG model. 6. SG

addions model Cl- 0

savePdb model VPN_aion.pdb

saveamberParm model VPN_aion.parm7 VPN_aion.inpcrd

solvateBox model TIP3PBOX 8.0 saveamberParm model VPN_wat.parm7 VPN_wat.inpcrd

quit

PDBファイルには SS結合の情報が失われるので、2回目の LEaPにおいて、再び SS結合の設定が必要である。” addion” により塩素イオンが2個追加される。後の水分子を削除した軌跡データの解析に必要なペプチド分子と塩素イオンだけのファイル (VPN_aion.parm7、VPN_aion.inpcrd) を同時に作成している。

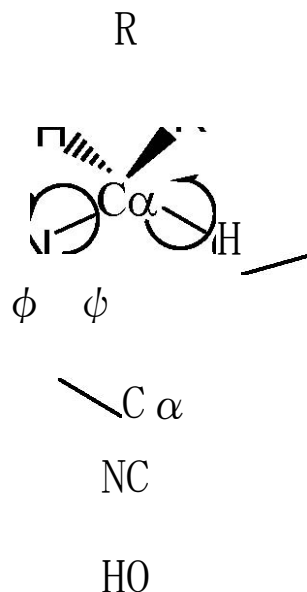
5.1.2. 軌跡ファイルの解析例

DNAの計算と同様な条件で溶媒和 MD計算を実施下さい。得られた 1ns production run の軌跡ファイルから溶媒分子を除去させた軌跡ファイルを作成する。対応するトポロジーファイルは” VPN_aion.parm7” となる。

□溶媒分子除去のための ptraj入力ファイル

```
trajin md3_1.mdcrd.gz trajin
md3_2.mdcrd.gz trajin md3_3.mdcrd.gz
trajin md3_4.mdcrd.gz trajin
md3_5.mdcrd.gz trajout
VPN_1000ps_nowat.mdcrd nobox
center :1-10
image familiar strip :WAT
```

□VPN_1000ps_nowat.mdcrd” を用いて軌跡データを解析する。ポリペプチド骨格の立体配座は、ペプチド結合 (-NHCO-) が trans平面なので C α 炭素を含む 2つの単結合の回転角 ϕ - ϕ で表現される。



ここで、n番目のアミノ酸の ϕ - ψ は ; $\phi = C_{n-1}-N_n-C_{\alpha n}-C_n$ 、 $\psi = -N_n-C_{\alpha n}-C_n-N_{n+1}$ ptraj の dihedral コマンドで ϕ - ψ 値を求める。

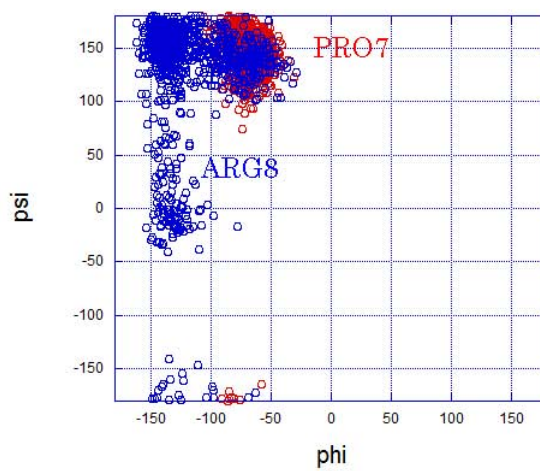
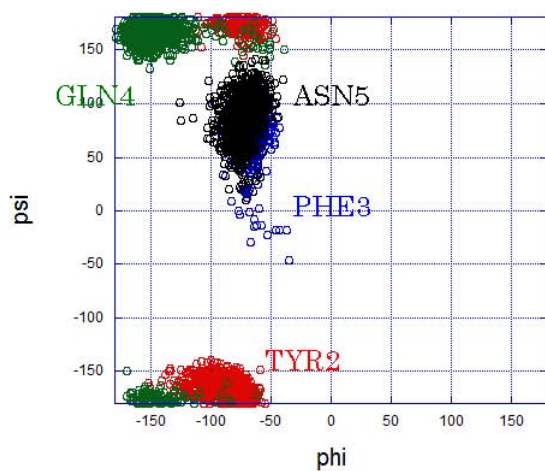
```
dihedral phi2 :1@C :2@N :2@CA :2@C out phi2.xls
dihedral psi2 :2@N :2@CA :2@C :3@N out psi2.xls
```

ϕ - ψ 値を 2次元プロットしたものを Ramachandranプロットとよぶ。以下に SS結合による環状構造内部のアミノ酸 Tyr2-Asn5と環外アミノ酸Pro7、Arg8の Ramachandranプロットを示す。環外 Arg8を除くアミノ酸の ϕ - ψ 値は初期構造の周辺で分布している。Arg8の ϕ

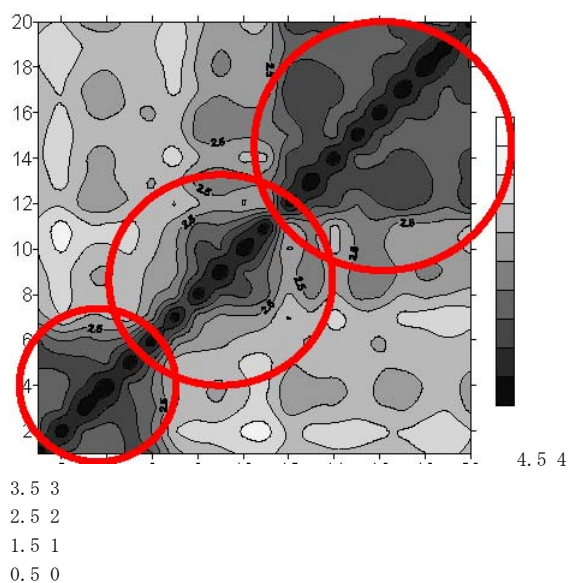
値は、後半の 500 ps以降、 -70° 付近から -140° へと分布中心が移動している。一方、 ϕ 値は2回、瞬間的の値が変化している。これらの分布領域はいずれも β -ストランド構造領域に相当する。なお、Ramachandranプロットは VMDによっても表示できる。構造ゆらぎが生じている MD軌跡から代表構造の抽出を行う。そのため、2D-RMSマップを計算する。これは、MD軌跡の個々の構造をすべて比較し、2次元の RMSd値として表現したものである。そのための ptraj入力ファイルを以下に示す。

z2D-RMSマップ作成のための ptraj入力ファイル

```
trajin VPN_1000ps_nowat.mdcrd 1 1000 50 2drms out 2d_rmsd.xls plotmtv :1-10
```



1 nsの軌跡を 200 ps×50フレームに分割して計算した二次元 RMSd値が” 2d_rmsd.xls”
に出力される。これをエクセル等で表示す



2 4 6 8 10 12 14 16 18 20

円で示されるように 1 nsの軌跡は三種類の立体構造に分類されることが推定される。但し、グループ分類は等高線の定義によって異なることもあり、いくつかのパターンを試したうえで直感的に適切なものを判断する（しかない）。次に、それぞれのグループ領域内で、各行または各列の RMSd値の総和を求め、総和値が最も小さい（残りの立体構造と比較して最も構造のズレが小さい）時間フレームの立体構造がそのグループ内での代表構造となる。

立体構造変化はコンタクトマップと Difference Distance Matrix (DDM) から評価できる。コンタクトマップは全残基間の距離を行列要素とした行列で距離行列 (Distance Matrix) ともよぶ。以下の入力ファイルを用いて ptrajを実行する。

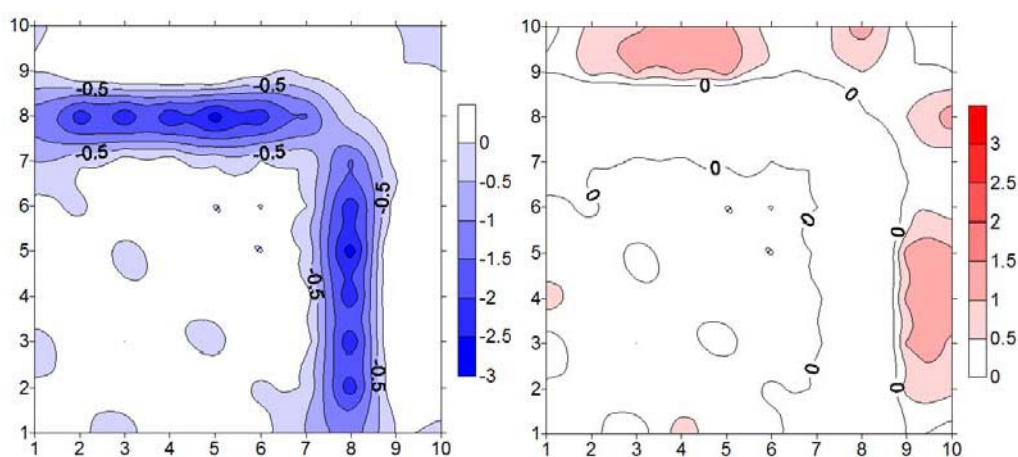
□コンタクトマップ計算のための ptraj入力ファイル

```
trajin VPN_1000ps_nowat.mdcrd
```

```
matrix dist :1-10 out dist_Map_200ps.xls start 200 stop 300 bytes
```

```
matrix dist :1-10 out dist_Map_500ps.xls start 500 stop 600 bytes
```

このより、200-300psの構造からコンタクトマップを計算し、それより求めた平均値が”dist_Map_200ps.xls” に、同様に500-600psの構造の平均値が” dist_Map_500ps.xls” に出力される。次に、500psのマップから 200psのマップの各要素を差し引いて DDMを得る。残基間距離が接近する要素（値がマイナス）と離れる要素（値がプラス）についてそれぞれ 2次元マップで表現したものを以下に示す。



SS結合で連結された環内部の残基はお互いに距離を変化させないが、Arg8が環状部分の Tyr2-Cys6に 3 Å程度近づき、C末端は若干離れる方向に変化している。但し、前の 2D-RMS マップと同様に、バソプレッシン程度の大きさの分子では、立体構造変化に対して構造生物学的に意味のある知見は得られないと思われる。

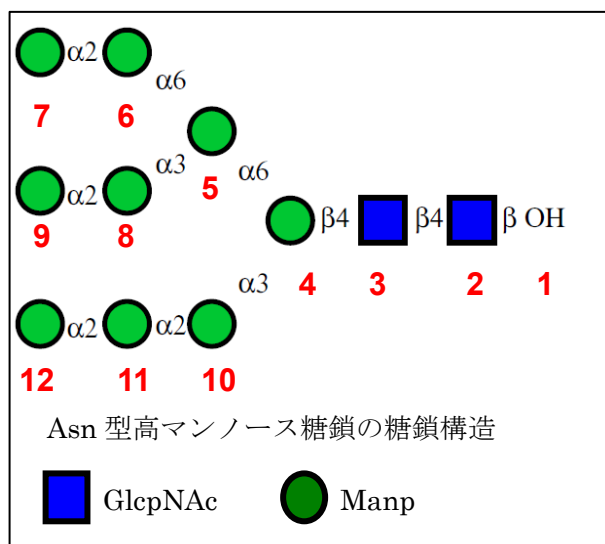
さらに、初期構造構築の段階で LEaPが生成した直鎖状の立体構造からかなり強引に SS結合を形成させ、環状構造にしている。このプロセスにより後続く MD計算では解消できないほどの立体化学的ストレスを内在させている可能性がある。実際は、最初に述べたトリプシンとの複合体構造 (1YF4) 中のバソプレッシン部分を初期構造として利用の方が無難である。もし、X線結晶構造データ等が得られない場合、束縛条件 MD計算を適用し、ペプチド末端を少しずつ短くさせることで環状構造を得るアプローチも可能であるが、いずれにせよ合理的な環状構造をモデリングで求めることは容易でない。また、10残基程度のペプチドであっても今回のような 1nsのシミュレーション時間は、配座アンサンブルを網羅する条件としては全く短すぎることに留意すべきである。

5.2. 分岐状糖鎖 (AmberTools User's Manual ver.1.5 “Building oligosaccharids and lipids” より)

糖鎖がタンパク質に共有結合した分子は糖タンパク質とよばれ単純タンパク質ではもちえない機能をもつ。糖タンパク質糖鎖は、糖鎖とペプチド結合領域の違いからアスパラギン残基に結合するアスパラギン型 (Asn型、N-グリコシド結合型) 糖鎖とセリンやトレオニンに結合するムチン型 (O-グリコシド結合型) 糖鎖に分類される。Asn型糖鎖は、さらに、側鎖構造の違いから高マンノース型、混成型および複合型に分類される。ここでは、高マンノース型糖鎖を例としたトポロジーファイル作成方法について述べる。

- トポロジーファイル作成におけるペプチド鎖と糖鎖の主な違いは以下の2点となる。
- ①糖鎖の力場パラメータとして、ジョージア大学の Robert J. Woodsが開発した Glycamパラメータセット (<http://glycam.ccrc.uga.edu/ccrc/>) を利用する。LEaPにおいてこのパラメータセットをロードする。
 - ②ペプチド鎖と異なり、多くの糖鎖は分岐構造をもつ。そのため、残基の順序に注意を払う必要がある。

5.2.1. “sequence” コマンドによる作成法



右図に高マンノース型糖鎖の糖鎖構造を示す。残基番号 (赤字) は還元末端から非還元末端方向に定義する。還元末端側のアグリコン原子団が1番目の残基となる (この例では、OH基)。グリコシド酸素は還元末端側の残基 (またはアグリコン) に所属し、それに結合する残基の構成原子はアノメリック炭素 (C1) から始まる。Glycamにおける残基名称の一般ルールは、グリコシド結合番号+糖残基のタイプ+アノメリックタイプの3文字で表現する。例え

ば、2番残基は β -1,4結合のN-アセチルグルコサミンなので” 4YB”、6番残基は α -1,2結合のマンノースなので” 2MA” となる。なお、非還元末端側の残基（残基番号7、9、12）の結合番号は” 0”、分岐部の残基は” V” となる。マンノース、N-アセチルグルコサミン以外の糖残基の名称については、AmberマニュアルやGLYCAM_06.prepファイルの中身から確認する。次表に Asn型高マンノース糖鎖の残基番号と残基名称の対応を示す。

番号	1	2	3	4	5	6	7	8	9	10	11	12
名称	ROH	4YB	4YB	VMB	VMA	2MA	OMA	2MA	OMA	2MA	2MA	OMA

ペプチド鎖と同様に” sequence” コマンドで残基配列を定義するが、分岐部で一旦、定義を止める。次に、” tail” コマンドで糖鎖が結合するグリコシド酸素を定義し、それに結合する糖鎖配列を” sequence” で定義する。さらに分岐部がある場合は、再び、その分岐残基で残基配列の定義と止め、別の” tail” と” sequence” コマンドで糖鎖配列を定義してゆく。言い換えると、還元末端残基からいずれかの非還元末端残基までの主鎖配列を定義しそれに側鎖配列を結合するのではなく、分岐部分を境にブロックに分割してブロックごとに定義する。これにより分岐部におけるトラブルを回避できる。

以下に LEaP入力ファイルの例を示す。この例では、糖鎖配列が (1, 2, 3, 4)、(5)、(6, 7)、(8, 9)、(10, 11, 12) のブロックに分割されている。” impose” コマンドによりグリコシド結合周辺の結合回転角を指定している。なお、 ϕ 配座 (H1-C1-O_n-C_n) に対してはエキソアノメリック効果で安定化される値 ($\pm 60^\circ$) がデフォルトで設定される。

z残基配列に基づいた LEaP入力ファイル例

```
set default writel4scale on

source leaprc.GLYCAM_06
glycan = sequence { ROH 4YB 4YB VMB }
set glycan tail glycan.4.O6
glycan=sequence { glycan VMA }
set glycan tail glycan.5.O6
glycan=sequence { glycan 2MA OMA }
set glycan tail glycan.5.O3
glycan=sequence { glycan 2MA OMA }
set glycan tail glycan.4.O3
glycan=sequence { glycan 2MA 2MA OMA }

impose glycan {3 2} { {H1 C1 O4 C4 60.0} }
impose glycan {3 2} { {C1 O4 C4 H4 0.0} }
impose glycan {4 3} { {H1 C1 O4 C4 60.0} }
impose glycan {4 3} { {C1 O4 C4 H4 0.0} }
impose glycan {5 4} { {H1 C1 O6 C6 -60.0} }
impose glycan {5 4} { {O6 C6 C5 O5 60.0} }
impose glycan {5 4} { {C1 O6 C6 C5 180.0} }
impose glycan {10 4} { {H1 C1 O3 C3 -60.0} }
impose glycan {10 4} { {C1 O3 C3 H3 0.0} }
impose glycan {6 5} { {H1 C1 O6 C6 -60.0} }
```

```

impose glycan {6 5} { {06 C6 C5 O5 -60.0} }
impose glycan {6 5} { {C1 O6 C6 C5 180.0} }
impose glycan {8 5} { {H1 C1 O3 C3 -60.0} }
impose glycan {8 5} { {C1 O3 C3 H3 0.0} }
impose glycan {7 6} { {H1 C1 O2 C2 -60.0} }
impose glycan {7 6} { {C1 O2 C2 H2 0.0} }
impose glycan {9 8} { {H1 C1 O2 C2 -60.0} }
impose glycan {9 8} { {C1 O2 C2 H2 0.0} }
impose glycan {11 10} { {H1 C1 O2 C2 -60.0} }
impose glycan {11 10} { {C1 O2 C2 H2 0.0} }
impose glycan {12 11} { {H1 C1 O2 C2 -60.0} }
impose glycan {12 11} { {C1 O2 C2 H2 0.0} }

savePdb glycan n-man9_vac.pdb
saveamberparm glycan n-man9_vac.parm7 n-man9_vac.inpcrd

quit

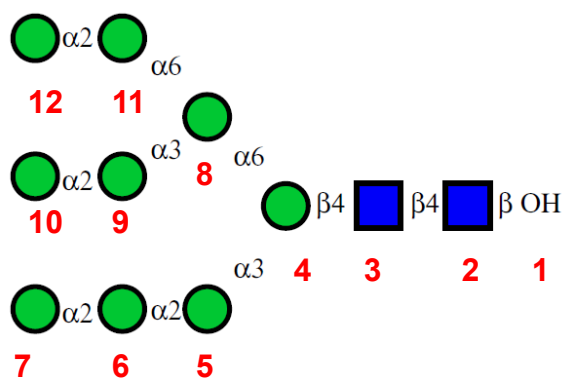
```

5.2.2. PDBファイルの読み込みによる作成法

糖タンパクや糖質分子結晶のように糖鎖立体構造の結晶構造データが得られる場合、それらの結晶構造を分子動力学計算の初期構造として利用の方が好都合である。糖鎖の PDB ファイルを LEaP に読み込ませるが、PDB データにおいて糖鎖はヘテロ分子扱いとなり、タンパク質部分のように明確なルールが定められていないようである。そのため、自動修正機構も整備され、わずかな修正で LEaP が成功するタンパク質と比較して、糖鎖の PDB ファイルに対しては LEaP に読み込ませる前に大幅な修正が必要となることが多い。

以下の主要な修正点を述べる。

- ① IUPAC ルールでは糖鎖は非還元末端から定義されるが、Glycam では還元末端のアグリコン部分を始点とする。そのため、PDB データ内の糖鎖全体について残基の順番を変更する必要がある。還元末端側のアグリコン基 (-OH、-OCH₃、等) は独立の残基として
み
なし、通常、1 番残基となるため、糖鎖全体の残基数は、+ 1 増加する。
- ② 残基名称を Glycam の命名法に変更する。原子名称についても置換基などで一致しないことがある。特に、水素原子は数が多く、不一致によるエラーの原因になり易いので、水素原子座標のみあらかじめ削除する方が便利なこともある。但し、糖質分子結晶の高分解能中性子回折データのように、水酸基水素原子の座標まで確定している場合は、その限りではない。
- ③ 残基内の構成原子の順序については、prep ファイルの定義と完全に一致していなくても、LEaP がある程度までは処理する。但し、先頭原子がアノメリック炭素 (C1)、最後尾の原子が次の残基と結合するグリコシド原子とすることは必須のようである。



以上の PDBファイルの修正に加え、対象糖鎖分子が分岐部を持つ場合、糖鎖の順序に注意を必要とする。実際の Asn型高マンノース糖鎖の例について説明する。分岐残基の prepファイル (VMA、VMB) の最後尾の原子は 03であるため、それに続く残基は 03に結合するものとなる。従って、糖鎖全体の残基番号は右図のように定義される。これに、7 番残基と 8 番残

基、10番残基と 11番残基の間に” TER” を挿入する（無ければエラーが発生する）。

ATOM	138	H62	OMA	7	-10.530	18.374	6.643	1.00	0.00
ATOM	139	O6	OMA	7	-10.735	17.325	4.838	1.00	0.00
ATOM	140	H6O	OMA	7	-10.124	16.726	5.262	1.00	0.00
TER									
ATOM	141	C1	VMA	8	2.862	13.586	4.466	1.00	0.00
ATOM	142	H1	VMA	8	3.097	14.654	4.328	1.00	0.00
ATOM	143	C2	VMA	8	3.943	12.702	3.856	1.00	0.00
(省略)									
ATOM	201	H62	OMA	10	2.628	5.883	5.385	1.00	0.00
ATOM	202	O6	OMA	10	4.305	6.975	6.014	1.00	0.00
ATOM	203	H6O	OMA	10	3.653	7.661	6.146	1.00	0.00
TER									
ATOM	204	C1	2MA	11	3.724	12.607	9.813	1.00	0.00
ATOM	205	H1	2MA	11	2.951	13.313	10.158	1.00	0.00
ATOM	206	C2	2MA	11	5.113	13.055	10.252	1.00	0.00

LEaP入力ファイルにおいては PDBファイルのロードに加え、2箇所の分岐部残基と側鎖糖鎖残基の間に” bond” コマンドで C1-O6結合を定義する必要がある。

PDBファイルの読み込みによる LEaP入力ファイル例

```
set default writel4scale on source leaprc.GLYCAM_06

glycan = loadPdb n-man9.pdb

bond glycan.4.O6 glycan.8.C1
bond glycan.8.O6 glycan.11.C1

savePdb glycan n-man9_vac.pdb
saveamberparm glycan n-man9_vac.parm7 n-man9_vac.inpcrd

quit
```

トポロジーファイルの作成を確認した後、続けて MD計算を実施しても良いが、複雑な糖鎖構造であれば、適切なトポロジーファイルが得られていないことが多い。簡単な検証法として本格的な MD計算に着手する前に短時間の MD計算を実施し、出力結果を確認する。

①出力ファイルの冒頭部分に総電荷値のメッセージが現れることがある（Sum ofcharges from parm topology file = x.xxx）。想定される 0 または整数値から極わずかに異なる場合があるが、それは数値計算誤差によるもので実質的に問題はない。そのような値からはっきりと逸脱している場合は、残基名称の定義ミスにより中間残基が末端部分に指定されている可能性がある（open bondが生じた）。

②”bond” 定義が正しくない、または不完全である場合、MD計算の結果、一部の残基が残りの部分から分離する（分子が分解する）。分子グラフィックス等で最終構造を検証する。

問題が無ければ、DNAの計算例に従って溶媒和の MD計算を実施し、得られた軌跡を解析しなさい。

5.2.3. 軌跡ファイルの解析例

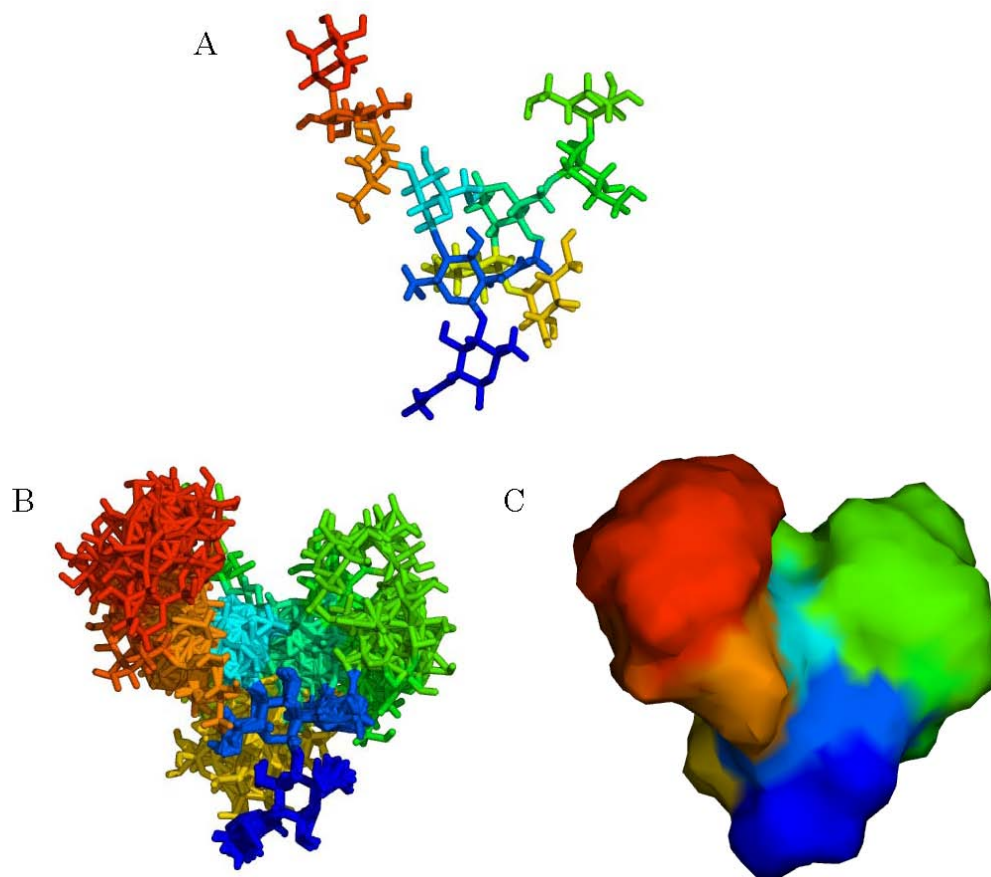
次の ptraj入力ファイルより、1nsの production runの軌跡から水分子を除去し、軌跡ファイル形式（n-man9_1000ps_nowat.mdcrd）と PDB形式（n-man9_1000ps_div.pdb）の2種類の座標ファイルが得られる。RMSd計算も含まれるが、これは還元末端キトビオース部分（残基番号1－3）にすべてフレームの座標をフィッティングさせるのが目的である。このプロセスにより座標ファイルを VMDや分子グラフィックスで表示させるとキトビース部分が空間にほぼ固定された状態になる。

z溶媒分子の除去と座標フィッティングおよび座標ファイル出力のための ptraj入力ファイル

```
trajin wat_md3_1.mdcrd.gz
trajin wat_md3_2.mdcrd.gz
trajin wat_md3_3.mdcrd.gz
trajin wat_md3_4.mdcrd.gz
trajin wat_md3_5.mdcrd.gz

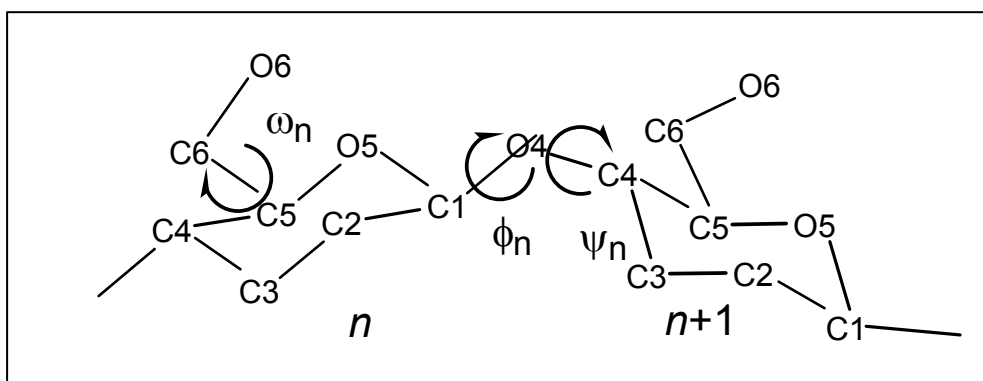
trajout n-man9_1000ps_nowat.mdcrd nobox
trajout n-man9_1000ps_div.pdb pdb
rms first out n-man9_1000ps.rmsfit :1-3 time 1.0
center :1-12
image familiar
strip :WAT
```

PDB形式ファイルの最初のフレーム座標 (A) と 100psごとの 10個のフレーム座標を重ねた図 (B)、およびそれを溶媒接触表面で表示したもの (C) を以下に示す。



タンパク質に結合する還元末端キトビオース部位を起点として、構造ゆらぎにより全体構造が広がっているのが分かる。タンパク質において、活性ループのような領域を除き、特に二次構造部分で局所的に構造がゆらぐ程度である。一方、糖タンパク質の糖鎖部分は溶液中で大幅な立体構造変化が許されるため、化学構造から想定されるより大きな空間を占めるものとみなされる。従って、糖タンパク質によっては、実質的に糖鎖部分がタンパク質部分にほぼ匹敵するほどの空間を占める場合もある。立体構造変化を詳細に解析するために、バソプレッシンと同様に 2D-RMS または DDM 解析を試みなさい。

糖鎖の主要構造パラメータを以下の図に示す。



ペプチドと異なり、糖鎖ではグリコシド酸素を挟んでグリコシド結合配座 $\phi-\phi$ を定義する。2種類の $\phi-\phi$ の定義がある。

$$\phi = \text{O5-C1-O4-C4}, \quad \phi = \text{C1-O4-C4-C5} \quad \phi_{\text{H}} = \text{H1-C1-O4-C4}, \quad \phi_{\text{H}} = \text{C1-O4-C4-H4}$$

立体化学やNMRデータの解析等では $\phi_{\text{H}}-\phi_{\text{H}}$ 定義が利用されるが、分子力学分野ではねじれポテンシャル関数を表現するために $\phi-\phi$ が一般的である。糖鎖において $\phi-\phi$ だけでなくメチル水酸基の配向 (ω) も極めて重要である。特に、分岐状糖鎖では、この配座が分子の全体構造に大きく影響する。

$$\omega = \text{O5-C5-C6-O6}$$

メチル水酸基配向における三種の安定配座には名称がある； gg ($\omega = -60^\circ$)、gt ($\omega = 60^\circ$)、および tg ($\omega = 180^\circ$)。グルコース残基の場合、立体化学的には tg が最安定であるが、水溶液中では溶媒分子との相互作用の結果、gt と gg がほぼ同じ割合で分布している。そして、糖分子の結晶構造中では gt が圧倒的に多い。

Ω

□メチル水酸基配向を求める ptraj入力ファイ

ル trajin n-man9_1000ps_nowat.mdcrd

dihedral pmega4 :4@O5 :4@C5 :4@C6 :4@O6 out omega4.xls dihedral

pmega5 :5@O5 :5@C5 :5@C6 :5@O6 out omega5.xls

omega4

omega5

150 100 50 0 -50 -100 -150

0 200 400 600 800 1000

time

このメチル水酸基の時間変化プロットより、最初の 200ps以降、 ω 値が一定になっているのが分かる。しかしながら、前のペプチドホルモンの場合と同様に 1nsのシミュレーション時間は立体構造変化を発生させるには短すぎる。グルコース残基単独の溶媒和

MD計算においてもggと gt間の配座変換を観察するためには 10ns程度のシミュレーション時間を必要とする。同様にして ϕ - ψ 値を求め、Ramachnadoranマップを作成しなさい。

糖残基に多く存在する水酸基は水素結合において受容体と供与体の両方の機能を持つため、水素結合解析において両方の設定が必要となる。グリコシド酸素は水素結合の設定から除外する。さらに溶媒分子を含むオリジナルの軌跡ファイルを用いてい溶媒分

子との水素結合形成を解析してもよい。

□水素結合解析のための ptraj入力ファイ

ル trajin ../wat_md3_1.mdcrd.gz

trajin ../wat_md3_2.mdcrd.gz

trajin ../wat_md3_3.mdcrd.gz

trajin ../wat_md3_4.mdcrd.gz

trajin ../wat_md3_5.mdcrd.gz

05 atoms for all residues donor mask :2-12@05

4YB donor mask :2, 3@02N, N2, 03, 06 acceptor mask :2, 3@N2, 03, 06 :2, 3@H2N, H30, H60

2MA donor mask :6, 8, 10, 11@03, 04, 06 acceptor
mask :6, 8, 10, 11@03, 04, 06 :6, 8, 10, 11@H30, H40, H60

OMA donor mask :7, 9, 12@02, 03, 04, 06 acceptor
mask :7, 9, 12@02, 03, 04, 06 :7, 9, 12@H20, H30, H40, H60

VMB and VMA donor mask :4, 5@02, 04 acceptor mask :4, 5@02, 04 :4, 5@H20, H40

hbond distance 3.5 angle 120.0 solventneighbor 6 solventdonor WAT 0 solventacceptor WAT 0
H1 solventacceptor WAT 0 H2

以下に水素結合解析の出力結果の一部を示す。

(省略)

DONOR ACCEPTORH ACCEPTOR

atom#	:res@atom	atom#	:res@atom	atom#	:res@atom	%occupied	distance	angle
32	:3@05	18	:2@H30	17	:2@03			
77.70	2.802	(0.16)	23.77	(13.38)	59	:4@05	45	:3@H30
44	:3@03	59.30	2.938	(0.22)	32.59	(14.89)	87	:5@04
182	:9@H60	181	:9@06	41.00	2.888	(0.18)	23.90	(11.20)
51	:3@02N	88	:5@H40	87	:5@04	8.40	2.837	(0.20)
30.90	(11.87)	38	:3@06	18	:2@H30	17	:2@03	6.00
3.312	(0.14)	51.72	(7.56)					

(省略)

74	:4@02	solvent	acceptor	47.00	2.946	(0.21)	27.54	(14.55)	138	:7@06	solvent	acceptor	38.30	
2.932	(0.21)	27.18	(14.65)	193	:10@04	solvent	acceptor	29.80	2.924	(0.21)	26.13	(14.72)	91	:5@05
solvent	acceptor	28.70	2.962	(0.21)	30.43	(15.19)	44	:3@03	solvent	acceptor	27.80	2.954	(0.22)	28.43

(省略)

キトビオース部分の 03-H---05隣接残基間水素結合と Man5-Man8-Man9部分が折れ曲がった
結果、糖残基配列上で離れた残基同士での残基間水素結合が発生している。

第3章 ptrajによる軌跡データの解析

1. コマンドの基本構成

“ptraj” プログラム使用の基本形は下記の形式となる。

ptraj トポロジーファイル < scriptファイル [> 出力ファイル]

入力ファイルの相当する scriptファイルの基本構成は、

①軌跡ファイルのリスト； “trajin” コマンドにより、mdcrdファイル名を指定する。複数ファイル指定しても良く、また mdcrdファイルは” gz” または” Z” 拡張子を持つ圧縮ファイルを指定して良い。

②出力座標ファイルのリスト； ” trajout” コマンドにより座標データを指定された形式とファイル名で出力する。オプションのコマンドである。

③アクションリスト；必要な解析を実行するコマンドのリスト。複数のコマンド群を指定してもよいが、指定された順序で各コマンドが実行される。

例) trajin md1.mdcrd.Z 1 20 trajin md2.mdcrd.Z 1 100 trajout fixed.mdcrd rms first out rms
@CA,C,N center :1-20 image strip :WAT go

最初の 2行が” trajin” で指定される入力軌跡ファイル名、3行目が” trajout” で指定された出力座標ファイル名に相当し、以降、5種類の実行コマンドが続いている。

以後の解説では、コマンドを太字、ユーザーが指定するパラメータ値を斜体、標準字体をそのコマンドの実行オプションを指定するキーワードを意味する。各キーワード、パラメータ値において” []” で囲まれている場合はオプションで、その中で” |” は区切られている場合は” or” の意味となる。一方、” []” で囲まれていないものは必須のキーワードに相当する。

実行コマンドが対象分子全体に適用する必要が無い場合も多く、そのときは“mask”により、実行コマンドが作用する原子団や残基を指定する。”@”は特定の原子（団）を、そして”：”は特定の残基をそれぞれ指定する文字である。原子や残基はそれらのラベル名や通し番号のいずれかで指定される。一連の領域を指定する場合は継続を意味する”-”を用いる。maskについての詳細は5. で解説する。

2. 入力/出力ファイル指定コマンド

trajin 軌跡ファイル名 [初期値 終値 オフセット] デフォルトでは全ての軌跡データを読み込むが、必要に応じて読み込む範囲、およびデータ間隔を指定する。軌跡ファイルとして “.Z” または “.gz” 拡張子を持つ圧縮ファイルを指定できる。

reference 座標ファイル名 座標ファイルの座標 (リスタートファイルまたは軌跡ファイルの最初の座標) を参照座標として登録する。 ” rms ” コマンドと併用する。

trajout 座標ファイル名 [書式] [nobox] [little|big] [dumpq|parse] [nowrap] 重要な書式は、 ” trajectory ” 軌跡ファイル書式でデフォルト設定、 ” restart ” リスタートファイル書式、或いは ” pdb ” PDB書式、等が指定できる。他のオプションはマニュアルを参照。

3. アクションコマンド

angle パラメータ名 mask1 mask2 mask3 [out 出力ファイル名] [time 間隔] 軌跡座標より ” mask ” で指定する 3つの原子で定義された結合角を出力ファイルに出力する。 ” name ” は結合角を定義するパラメーター名を設定する。 ” time ” により各座標の時間間隔を ps で指定する。

atomicfluct [out 出力ファイル名] [mask] [start 初期値] [stop 終値] [offset オフセット値] [byres | byatom | bymask] [bfactor] “mask” で指定された原子の位置変位を計算し、出力ファイルに出力する。 ” byres ” 指定により残基ごとの平均値を出力し、 ” byatom ” は原子ごとに値を出力する (デフォルト)、 ” bymask ” では ” mask ” で指定された原子すべてに対する平均値を出力する。 ” start ” , ” stop ” および ” offset ” により位置変位の平均値の計算対象とする範囲を指定する。 ” bfactor ” を指定すれば、温度因子 (デフォルトの位置変位に $8/3 \times \pi^2$ を掛けた値) が出力される。次の例ではポリペプチド鎖の骨格原子 (C α , N, C) の温度因子を残基ごとの平均値として ” back.apf ” に出力する。

atomicfluct out back.apf @C,CA,N byres bfactor

closest total mask [oxygen | first] [noimage]各座標データにおける” mask” により指定された溶質分子に最も近い” total” 数の溶媒分子（” solvent” コマンドで指定する）を抽出する。” first” 指定により溶媒分子の最初の原子からの距離が評価対象となる。” noimage” 指定により溶媒分子を基準溶媒和ボックスに移動する操作が解除される。これにより image操作が省略され計算負担が大きく軽減される。 次の例では” solvent” コマンドで” WAT” 残基を溶媒分子として定義し、1-20番残基と溶媒残基の最初の原子（酸素原子）間の距離がもっとも近い 50個の溶媒分子を抽出する。

solvent byres :WAT closestwater 50 :1-20 first

dihedralパラメータ名 mask1 mask2 mask3 mask4 [out 出力ファイル名] [time 間隔] 軌跡座標より” mask” で指定する 4つの原子で定義された二面角を出力ファイルに出力する。” name” は結合角を定義するパラメーター名を設定する。” time” により各座標の時間間隔を psで指定する。” mask” が複数の原子を定義さればその重心に対して二面角が計算される。

diffusion mask time_per_frame [average] [filenameroot] “mask” で指定された原子の平均二乗変位量 (\AA^2) を計算する。” time_per_frame” により各座標の時間間隔 (ps) を指定。各時間帯の平均値に加え、” average” 指定により積算値の平均が同時に出力される。” filenameroot” は出力ファイル名の” root” 部分を定義する。平均二乗変位量の出力ファイル名は” filenameroor_r.xmgr”、その x, y, z成分は” filenameroot_x.xmgr” 等、そして総移動行程は” filenameroot_a.xmgr” にそれぞれ、出力される。

distanceパラメータ名 mask1 mask2 [out 出力ファイル名] [image]座標ごとに “mask” で指定した残基または原子団の重心間の距離を主力ファイルに出力する。” image” 指定により座標を基準溶媒和ボックスに移動する。

radialroot-ファイル名 スペース(A) 最大値 (A) solvent-mask [solute-mask] [closest] [density値] [noimage]溶質分子の全原子に対して、” solvent-mask” で指定された溶媒分子の動径分布関数 (RDF) を計算する。“root-ファイル名” は出力ファイル名の先頭部分を指定。0から” 最大値” までの範囲を” スペース” で分割した間隔でヒストグラムを求める。” solute-mask” の指定により、特定の溶質分子に対して RDFを計算する。” noimage” を指定しない限り、デフォルトでは” image” 処理 (セル外の溶媒をセル内に移動する) が行われる。他はマニュアルを参考。

rms mode [mass] [out filename] [time interval] mask [name name] [nofit] “mask” で指定された原子に対して、軌跡にともなう座標の RMSdを計算する。” mode” で指定されるキーワードを意味は以下のとおり。

previous : 直前の時間フレームの座標からの変位量

first : 初期時間の座標からの変位量

reference : ” reference” キーワードで予め指定された座標に対する変位量 “mass” を指定すれば、原子量重みを掛けた変位量を計算。” out” キーワードで出力ファイル名を指定。” time” キーワードで出力ファイル時間軸の間隔を指定。” name” キーワードで RMSdパラメーター名を指定。” nofit” 指定により RMSd計算に先立っての座標のfittingを行わない。

matrix dist | covar | mwcovar | distcovar | correl | idea | ired [name name] [order order] [mask1] [mask2] [out filename] [start start] [stop stop] [offset offset] [byatom | byres | bymask]距離 (dist)、共分散 (covar)、重み付き共分散 (mwcovar)、相関 (correl)、距離—共分散 (distcovar)、等の行列を計算する。結果は” filename” に出力される。” mask” で定義される原子数は、mask1>=mask2となる必要がある。” start” 、” stop” 、および” offset” で軌跡ファイルの処理範囲を指定する。“matrix” コマンドの結果は、さらに” nalyze” コマンドで解析される。

strip mask “mask” で指定された原子を軌跡ファイルなどより削除する。” strip” を指定後の他の ptraj操作は、編集された座標に対して行われる。

watershellmask出力ファイル名 [lower lower upper upper] [solvent-mask] [noimage] “mask” で指定された原子からの一定距離 (shell) にある溶媒分子数を計算する。” lower” および” upper” は第一 shellと第二 shellの距離を指定。デフォルトはそれぞれ、3.4Åと 5.0Å。” solvent-mask” で溶媒分子を指定。” image” キーワードで” image” 処理を行う。デフォルトでは行わない。

4. 水素結合設定

donor resname atomname | mask mask | clear | print 水素結合受容体を指定。残基名-原子名標記、あるいは” mask” 標記で指定する。

```
donor ADE N7 (残基名-原子名標記)
```

```
donor mask :10@N7 (mask標記)
```

“clear” キーワードでこれまでに指定した donor リストを消去する。” print” キーワードで donor リストを表示。

acceptor resname atomname atomnameH | mask mask maskH | clear | print 水素供与体を指定。重原子と水素原子 (atomnameH, maskH) のセットで指定する他は” donor” キーワードと同じ。同じ原子が重原子と水素原子に指定される場合は、結合角を評価しない (イオン性水素結合に適用)。

hbond [distance value] [angle value] [solventneighbor value] [solventdonor donor-spec] [solventacceptor acceptor-spec] [nosort] [time value] [print value] [series name] 水素結合として評価する原子対の幾何学的条件を指定する。” distance”、” angle” により水素結合を評価するカットオフ距離や角度を指定。デフォルトでは角度は未指定で距離は 3.5Å以下。一方の原子について原子 (残基) タイプのみで個々の原子の情報を必要としない場合 (例、溶媒分子との水素結合)” solvent” を付加したキーワードを用いる。” solventneighbor” により供与体-受容体あたりの最大相互作用数を指定。

```
例) hbond distance 3.5 angle 120.0 solventneighbor 6 solventdonor WAT 0 solventacceptor WAT 0 H1 solventacceptor WAT 0 H2
```

```
hbond distance 3.5 angle 0.0 solventneighbor 6 solvnetdonor Na+ Na+ solventacceptor Na+ Na+ Na+
```

5. ambermaskの指定方法 (“Amberによる分子シミュレーション入門“より)

標準的に用いられる記号は以下のとおりである。

： 残基の指定 (残基番号もしくは残基タイプ) @ 原子の指定 (原子番号、原子名称もしくは%原子タイプ)
-数字から数字の間を連続して指定
* すべて (“:*” は全残基、“@*” は全原子を指定) % 原子タイプを指定

(注) 原子名称は元素記号と番号で表現される残基内で特定の原子を定義する記号。残基名称とともに PDBファイルにもみられる。原子タイプは力場によって定義される原子のタイプで原子の混成状態や周囲の原子団の環境によって設定される。prepファイルに記載されている。

ambermaskによる具体的な指定例を次に示す。

- ①残基を指定 :1-7, 5, 7-9 残基番号1-7、5、7-9を指定 :LYS リシン (lys) 残基を指定
- ②原子を指定 @12, 54-85 12、54-85番目の原子を指定 @CA, C, O, N 原子名称がCA、C、O、N (ペプチド骨格) を指定 @%CT amber原子タイプがCTであるものを指定
- ③残基指定と原子指定を組み合わせる:1-10@CA, CB 残基番号 1-10のうち、原子名称が CAと CBの原子を指定

さらに次の記号を組み合わせで論理式や距離制限等を指定する。 A & B and (論理積; 条件 Aと条件 Bを同時に満たすもの) A | B or (論理和; 条件 Aもしくはは条件 Bを満たすもの) ! A not (否定; 条件 Aでないもの) A <: dAからの距離が dÅより小さい残基 A >: dAからの距離が dÅより大きい残基 A <@ dAからの距離が dÅより小さい原子 A >@ dAからの距離が dÅより大きい原子 = wild card (任意の文字列) () 括弧で囲まれたものは優先

これらの記号を組み合わせ、次に示すように複雑な指定が可能となる。

@C= & !@CA,C ペプチド結合を形成する C_α炭素 (CA) とカルボニル炭素 (C) 以外のすべての炭素原子を指定 (:1-3@CA | :5-7@CB) 1番から3番残基までの C_α炭素、および5番から7番残基までの C_β炭素を指定:CYS, ARG & !(:1-10 | @CA, CB) システインおよびアルギニン残基のなかで1番から10番残基以外で C_α炭素でも C_β炭素でもないものを指定 :* & !H= もしくは !:H= 水素原子以外のすべての原子を指定 :5 <@ 4.5 5番残基から 4.5Å以内の距離にあるすべての原子を指定 (:1-35 <: 3.0) & :WAT 1番から35番残基から 3.0Åまでの範囲にある残基のうち水分子をすべて指定

mask指定が想定した原子に対応しているかどうか確認するために、“ambmask”というコマンドがある。次の例は、1番—10番残基から 3.0Å以内にある水分子を抽出するコマンドである。

```
[tyui@seneca01 wat]$ ambmask -p prk_wat.parm7 -c prk_wat_min1.rst -prnlev 1 -out pdb -find "( :1-10 <: 3.0) & :WAT" > stdout
```

- ① “-prnlev” オプションを”0”と指定すれば”-find”で指定された原子のみが出力され、”1”とすれば”0”出力に指定した条件が追加される。
- ② “-out” オプションに出力結果のファイル形式を指定する。”pdb”はデフォルト。他に”amber”や”short”がある。
- ③ “-find” オプションに確認したい maskの内容を記載する。引用符 (” ”) で囲っておくこと。
- ④ “-out” オプションは出力ファイル名。

第4章 タンパク質分子動力学計算の基礎知識

1. PDBファイル基本事項

- 1.1. 生体高分子構造座標データの構成 タンパク質、核酸および糖鎖を含めた生体高分子は残基を構成要素として、それらが特定の配列で連結した構成を持つ。タンパク質、核酸、および糖鎖の残基は、それぞれ、アミノ酸、ヌクレオチド、および単糖である。DNAは 2本のポリヌクレオチド鎖、多くのタンパク質も複数のポリペプチド鎖で構成される。

1.2. PDBファイルについて

タンパク質・核酸などの生体高分子の立体構造は、主にX線結晶構造解析法によって決定される（一部は NMR分光法による）。得られた座標データは多数の原子から構成されるため、膨大な量の座標を学術論文に直接掲載することが現実的でない。Protein Data Bank (PDB)はこのような生体高分子立体構造の座標と解析パラメータなどのデータの編集、公開を WEBと ftpで行われているデータベースである。ほとんどが蛋白質の立体構造で、一部核酸と蛋白質の複合体や核酸の構造も含まれる。

PDB検索の日本ミラーサイト <http://pdb.protein.osaka-u.ac.jp/pdb/index.html>

PDBには一件の登録データに対して、PDB IDと呼ばれる一つの IDが付けられる。この IDは英数字 4文字から構成される。一つの IDに対して、データは一つのファイルが対応し、同じ種の蛋白質でも、実験条件、結果が違ふと異なる PDB IDが与えられます。99年7月から PDB IDは登録された順に、アルファベット順、数字順に付けられている PDBはこの PDB IDだけでなく関連キーワード、タンパク質名、登録された年月日、解析方法などで検索できる。

1. 3. PDBファイルの書式

PDBファイル名は PDB ID.pdbとなっている。データの内容は大まかに分けてファイルの前半の蛋白質の情報や構造解析に関するパラメータ、コメントなどが記入されているヘッダーと呼ばれる部分と後半の座標データの部分との二つに分かれる。PDBファイルにおけるデータの記入方法、すなわち文法は厳密に決まっている。ファイルの行に対し、列はカラムと呼び、一番左のカラムに来る文字列がレコードタイプとよばれるものでその行のデータの内容を決定する。各レコードタイプによってカラムの何番目から何番目までがどの情報を記入するか決まっている。例えば先頭の一番始めの行は HEADERと呼ばれるレコードタイプで、11から 50カラム目までは蛋白質の分類名が入る。51から 59カラム目までが登録された年月日、63から 66カラム目まではこのエントリーの PDB IDが入る。レコードタイプの順番も決まっており、その一部を以下の表に示す。各種の分子グラフィックス、および AMBER等の多くの計算化学プログラムは PDB書式

レコードタイプ	おもな内容
HEADER	蛋白質の分類 トリーの PC
TITLE	このエントリ
COMPND	蛋白質の名前
SOURCE	蛋白質の発現 名
KEYWDS	キーワード
EXPDTA	構造解析手法
AUTHOR	このエントリ
REVDAT	公開履歴
JRNL	このエントリ の人名
REMARK 3	構造解析に関
SEQRES	帰属できなか (missingresi 三文字表記。
HELIX	ヘリックスの
SHEET	シート構造の
LINK	蛋白質以外の 結合している
ATOM	蛋白質の原子 番号、座標
HETATM	蛋白質以外の chain id、残
CONECT	蛋白質以外の 子が結合して

に従ったファイルを認識する。これらのソフトウェアはすべての PDB情報を必要とするのではなく、実際は“ATOM”、“HETATM”レコードタイプに記入される座標データのみを必要とする。“HETATM”は標準のアミノ酸以外のリガンド、コファクターおよび水分子を定義する。

1 行の“ATOM”、“HETATM”には 1 個の原子座標に対応する。タンパク質においてポリペプチド鎖 1 本につき N 末端残基から C 末端残基までの配列順序に従って残基ごとに個々の原子座標がÅを単位とした XYZ 直交座標系で記載されている。PDB 内に複数の分子鎖（ポリペプチド鎖、ポリヌクレオチド鎖）が含まれている時は、異なる分子鎖の“ATOM”レコードタイプの間は“TER”レコードタイプで区切られている。AMBER において“TER”レコードタイプは必要である。ファイルの終端は“END”レコードタイプで示される。以下に実際の PDB ファイルにおける“ATOM”“HETATM”箇所の例を示す。

カラム位置

	1	2	3	4	5	6	7	8
1234567890123456789012345678901234567890123456789012345678901234567890								
ATOM	1	N ILE	1	-1.492	19.511	26.826	1.00 13.46	1PLC 124
ATOM	2	CA ILE	1	-1.280	20.833	27.511	1.00 12.56	1PLC 125
ATOM	3	C ILE	1	-1.342	21.935	26.477	1.00 12.21	1PLC 126
ATOM	4	O ILE	1	-0.827	21.772	25.344	1.00 11.05	1PLC 127
ATOM	5	CB ILE	1	0.032	20.826	28.259	1.00 13.31	1PLC 128

途中省略

ATOM	1542	HA ASN	99	10.482	12.089	31.724	1.00 18.40	1PLC1665
ATOM	1543	1HB ASN	99	10.085	12.549	34.024	1.00 24.02	1PLC1666
ATOM	1544	2HB ASN	99	10.618	13.773	33.221	1.00 22.17	1PLC1667
ATOM	1545	1HD2 ASN	99	8.032	15.280	35.033	1.00 29.75	1PLC1668
ATOM	1546	2HD2 ASN	99	9.582	14.928	35.108	1.00 30.04	1PLC1669
TER	1547	ASN	99	←分子鎖データの切れ目を示す				1PLC1670
HETATM	1548	CU CU	100	7.050	34.960	18.716	1.00 8.78	1PLC1671
HETATM	1549	O HOH	101	17.504	16.825	14.073	1.00 20.28	1PLC1672
HETATM	1550	O HOH	102	18.877	15.088	18.086	1.00 22.16	1PLC1673
HETATM	1551	O HOH	103	11.165	21.823	31.513	1.00 16.99	1PLC1674
HETATM	1552	O HOH	104	8.855	32.590	26.747	0.94 6.45	1PLC1675

以下省略

以下の表に“ATOM”と“HETATM”レコードタイプの各カラムの説明を示す。

カラム	内容
1-6	レコード名
7-11	すべての原子 (HET化合物も含む)につけられる通し番号
13-16	その行の原子の名前 (IUPAC標準書式)
17	一つの原子に複数の立体配座が存在する場合、これを識別するための ID
18-20	残基名 (ATOMレコードにおいては、標準アミノ酸 20種類の三文字表記 (IUPAC標準書式) が対応する)
22	各ポリペプチド鎖に名付けられる ID (chain id)
23-26	残基番号
27	chain idと残基番号双方が同じである二つ以上の残基を識別するのに使う (insertion id)
31-38	直交座標系の X軸の値 (単位はオングストローム)
39-46	直交座標系の Y軸の値 (単位はオングストローム)
47-54	直交座標系の Z軸の値 (単位はオングストローム)
55-60	占有率(0~1.0)
61-66	等方性温度因子

残基の IUPAC三文字略号を以下に示す。

略号	名称	略号	名称
ALA	アラニン	MET	メチオニン
CYS	システイン	ASN	アスパラギン
ASP	アスパラギン酸	PRO	プロリン
GLU	グルタミン酸	GLN	グルタミン
PHE	フェニルアラニン	ARG	アルギニン
GLY	グリシン	SER	セリン
HIS	ヒスチジン	THR	スレオニン
ILE	イソロイシン	VAL	バリン
LYS	リシン	TRP	トリプトファン
LEU	ロイシン	TYR	チロシン

2. LEaPにおけるタンパク質特有の事項

2.1. アミノ酸略号

LEaPは上記の IUPAC略号に従って PDBファイル中のアミノ酸配列を認識するが、タンパク質 PDBファイルをもとに AMBER入力ファイルを作成するときに、LEaPの実行に先立ってそれらのアミノ酸略号を変更する必要がある。

以下に AMBER特有のアミノ酸略号を示す。

略号	アミノ酸の状態
ACE	アセチル基末端
NHE	アミノ基末端
NME	N-メチルアミン末端
ASH	水素化されたアスパラギン酸残基（通常の残基は ASN）
CYX	S-S 結合を形成するシステイン残基（通常の残基は CYS）
GLH	水素化されたグルタミン酸残基（通常の残基は GLN）
HID	側鎖 δ 位に水素を持つヒスチジン
HIE	側鎖 ϵ 位に水素を持つヒスチジン
HIP	水素化されたヒスチジン

注) IUPACのヒスチジン略号、HISは用いない。

PDBファイルをそのまま使う [例 {VAL, GLY, ALA, PHE}]、或いは両末端残基略称に “C” “N” を付記する [例 {NVAL, GLY, ALA, CPHE}] ことにより、LEaPは末端残基を電荷状態にする (COO^- 、 NH_3^+)。或いは、“ACE”、“NHE”、“NME” のキャップ残基を設定する方法もある。

2.2. ジスルフィド (S-S) 結合の設定

PDB ファイル以下のエントリーを見つけた場合。

```
SSBOND 1 CYS A 32 CYS A 35
```

32番残基と 35番残基のシステイン “CYS” がジスルフィド結合を形成していることを示すエントリーである。このことをトポロジーファイル “prmtop” に反映するためにこれらの残基の “ATOM” エントリー部分での残基名称と “CYS” から “CYX” に変更する。

変更後の PDBファイル

```
ATOM 267 N CYX A 32 30.492 36.041 -15.250 1.00 22.22 2TRX 418
ATOM 268 CA CYX A 32 31.749 35.264 -15.067 1.00 20.24 2TRX 419
ATOM 269 C CYX A 32 31.532 33.814 -15.354 1.00 19.15 2TRX 420
ATOM 270 O CYX A 32 30.902 33.016 -14.643 1.00 18.30 2TRX 421
ATOM 271 CB CYX A 32 32.291 35.603 -13.649 1.00 18.86 2TRX 422
ATOM 272 SG CYX A 32 33.555 34.457 -13.062 1.00 17.52 2TRX 423
. . . . .
ATOM 284 N CYX A 35 32.981 31.281 -12.915 1.00 17.05 2TRX 435
ATOM 285 CA CYX A 35 32.187 31.344 -11.645 1.00 16.98 2TRX 436
ATOM 286 C CYX A 35 30.993 30.419 -11.728 1.00 18.73 2TRX 437
ATOM 287 O CYX A 35 30.668 29.740 -10.780 1.00 17.87 2TRX 438
ATOM 288 CB CYX A 35 31.674 32.747 -11.366 1.00 16.47 2TRX 439
ATOM 289 SG CYX A 35 32.923 33.992 -11.125 1.00 17.50 2TRX 440
```

LEaP実行時に次のコマンドを実行し、32番残基と 35番残基間の S-S結合を定義する。

```
> trx = loadpdb trx.pdb
> bond trx.35.SG trx.32.SG
```

□bond” コマンドは同じ” UNIT” 内の 2 つの” ATOM” オブジェクトに結合を形成する。

□trx.35.SG” は UNIT名” trx” の 35番残基の SG原子を定義する。

2.3.” alternative” 構造の選択

PDBファイル中で以下のエントリーを見つけた場合。

```
ATOM 393 OE1AGLU A 48 30.500 15.474 -2.178 0.50 25.49 2 2TRX 544 ATOM 394 OE1BGLU A 48
33.384 14.801 1.307 0.50 26.62 2 2TRX 545 ATOM 395 OE2AGLU A 48 32.049 16.115 -0.781
0.50 25.61 2 2TRX 546 ATOM 396 OE2BGLU A 48 32.104 15.515 -0.318 0.50 27.54 2 2TRX 547
```

残基ラベル” AGLU” と” BGLU” を持つ連続した原子は、2つの可能な構造を意味している。これは X線結晶構造解析において、構造が一部乱れておりデータからこれらの原子座標を一義的に特定できなかったためである。通常、”A” 構造の方がより可能性が高い構造とされるので、PDBファイルから”B” 構造を削除し残基名を修正しておくといよい。

```
ATOM 393 OE1 GLU A 48 30.500 15.474 -2.178 0.50 25.49 2 2TRX 544 ATOM 395 OE2 GLU A 48 32.049
16.115 -0.781 0.50 25.61 2 2TRX 546
```

修正後、LEaPを適用する。

3. リゾチームー阻害剤（アセチルキトトリオース）複合体における基質相互作用解析

3.1. 目的

リゾチーム（Lysozyme）とは、真正細菌の細胞壁を構成する多糖類を加水分解する酵素である。この作用があたかも細菌を溶かしているように見えることから溶菌酵素とも呼ばれる。ヒトの場合涙や鼻汁、母乳などに含まれている。工業的には卵白から抽出したリゾチームが食品や医薬品に応用されている。本稿では、DeppViewのトレーニングで使用したリゾチームー阻害剤複合体の結晶構造データを用いて、タンパク質の MD計算と基質に関わる構造情報や熱力学量について解析する。

3.2. PDBデータの入手と修正

PDBよりリゾチームー阻害剤（アセチルキトトリオース）複合体の座標データ（PDBID “1HEW”）のファイルを手に入る。

以下の “SSBOND” エントリー行にあるように、このリゾチームには4組の SS結合が確認されている。

SSBOND	1 CYS	A	6	CYS	A	127	1555	1555	2.09
SSBOND	2 CYS	A	30	CYS	A	115	1555	1555	2.08
SSBOND	3 CYS	A	64	CYS	A	80	1555	1555	2.11
SSBOND	4 CYS	A	76	CYS	A	94	1555	1555	2.09

この指定に従って、Cys6、Cys30、Cys64、Cys76、Cys80、Cys94、Cys115、および Cys127 の “ATOM” エントリー行にあるすべての残基名称を” CYS” から” CYX” に変更する。

CYS6の例

ATOM	43	N	CYX	A	6	-6.494	19.744	4.929	1.00	31.37	N
ATOM	44	CA	CYX	A	6	-7.927	19.378	5.088	1.00	32.81	C
ATOM	45	C	CYX	A	6	-8.096	18.138	5.922	1.00	19.65	C
ATOM	46	O	CYX	A	6	-8.951	18.087	6.840	1.00	24.72	O
ATOM	47	CB	CYX	A	6	-8.579	19.336	3.697	1.00	25.91	C

ATOM	48	SG	CYX	A	6	-8.763	21.096	3.142	1.00	29.62	S
------	----	----	-----	---	---	--------	--------	-------	------	-------	---

次に、” ATOM” エントリーの残基名称が” NAG” と指定された基質部分を以下の要領で修正する。

- 201番残基の 01原子（1012）を糖鎖部分の先頭行に移動。そして残基番号を 200に変更し、残基名称を” NAG” から” ROH” に変更。
- 201～203番残基の 04をそれぞれの残基部分の最終行（07の後）に移動。そして07を 02N、C7を C2N、C8を CMEに変更。さらに、201番、202番残基名称を” 4YB”、203番残基名称を” 0YB” に変更。

PDBデータにあった結晶水分子を MD計算においても利用するため、204番～306番残基のHOHをWATに変更する。

最後に、座標データが記載されている、” ATOM” と” HETATM” エントリー行以外の領域をすべて削除する。特に、” CONNECT” を削除しないと次の LEaP作業でエラーが発生する。

本例では、修正後の内容を” 1HEW_mod. pdb” の名称でファイルに保存する。

3.3. トポロジーファイルの作成

以下の LEaP入力ファイルを作成し、”sleap” コマンドによってトポロジーファイルを初期座標をファイルを作成する。

- LEaP入力ファイル” leap_wat.in” の例

```
set
default write14scale on source leaprc.ff99SB
source leaprc.GLYCAM_06 set default disulfide off

model = loadPdb 1HEW_mod.pdb

bond model.6.SG model.127.SG bond model.30.SG model.115.SG bond model.64.SG model.80.SG
bond model.76.SG model.94.SG

addions model Cl- 0

solvateBox model TIP3PBOX 8.0 savePdb model lysozyme_nag3_wat.pdb saveamberParm model
lysozyme_nag3.parm7 lysozyme_nag3.inpcrd

quit
```

”tleap” の実行、

```
$ sleap □s □f < leap_wat.in > leap_wat.out
```

トポロジーファイル” model lysozyme_nag3.parm7”、初期座標ファイル” lysozyme_nag3.inpcrd” および初期座標の PDBファイル” lysozyme_nag3_wat.pdb” が生成しているのを確認する。

3.4. MD 計算の実行

4 章の DNA の溶媒和 Box モデル計算の要領に従って、計算システムの最適化、平衡化および MD 計算を実施すること。なお、位置束縛を設定する計算においては、設定範囲はタンパクと基質となる（残基番号 1～133）。また、本例におけるファイル名称は、プロジェクト名称部分を省略し、”min1.in”、”min2.in”、”md1.in” とした。

Production runのみ、DNAの例とは異なり、1nsの計算を 5回実行し、合計 5nsの軌跡を求めること。

- Production runの入力ファイル”
md3.inゆいの例 1HEW: 1000ps production MD

```
&cntrl imin = 0, irest = 1, ntx = 5,
ntb = 2, pres0 = 1.0, ntp = 1,
taup = 2.0,
cut = 10, ntr = 0,
ntc = 2, ntf = 2,
empi = 300.0, temp0 = 300.0,
ntt = 1, tautp = 0.2,
nstlim = 500000, dt = 0.002,
ntpr = 500, ntwx = 500, ntwr = 500 &end
```

MD計算は、従来どおり” sander.MPI” で実行しても良いが、SANDERモジュールでは cpu数が4で、計算速度が頭打ちになる。本格的な計算では、より多くの cpu数が設定可能で、

計算が早く進行する” pmemd.MPI” で実行すべきである。

□PMEMDによるproduction runの実行スクリプト” run_md3_5ns.x” の例 (16 cpu) #!/bin/csh

```
set MDSTARTJOB=1 set MDENDJOB=5 set MDCURRENTJOB=$MDSTARTJOB set MDINPUT=0

cp md2.rst md3_0.rst

echo -n "Starting Script at: " date
echo ""
while ( $MDCURRENTJOB <= $MDENDJOB ) echo -n "Job $MDCURRENTJOB started at: " date @ MDINPUT
    = $MDCURRENTJOB - 1

bsub -K -n 16 -o log/ "$LSF_BINDIR/openmpi_wrapper -np 4 $AMBERHOME/exe/pmemd.MPI -O
-i ./md3.in -o ./md3_$MDCURRENTJOB.out -c ./md3_$MDINPUT.rst -r ./md3_$MDCURRENTJOB.rst
-p ./lysozyme_nag3.parm7 -x ./md3_$MDCURRENTJOB.mdcrd"

gzip -9 -v md3_$MDCURRENTJOB.mdcrd echo -n "Job $MDCURRENTJOB finished at: " date @
MDCURRENTJOB = $MDCURRENTJOB + 1
end echo "ALL DONE"
```

以上の計算が正常に終了すると、” md3_1” から” md3_5” の名称を持つ、出力(out)、軌跡(mdcrd)および最終座標ファイル(rst)が発生する。

1. 5. 結果の解析

2. 5.1. 構造パラメータの解析

以下の ptraj入力ファイルに従って、構造パラメータを解析するために、溶媒分子と塩素イオンを削除した 5nsの軌跡ファイルを作成する。

z溶媒分子と塩素イオンを削除するための ptrajスクリプトファイル” stripWat.csh”

の例 #!/bin/csh

```
Ptraj lysozyme_nag3.parm7 << eof trajin ./md3_1.mdcrd.gz trajin ./md3_2.mdcrd.gz
trajin ./md3_3.mdcrd.gz trajin ./md3_4.mdcrd.gz trajin ./md3_5.mdcrd.gz

rms out rms.xls :1-129 strip :WAT strip :Cl

trajout md3_vac.mdcrd nobox
```

スクリプトを実行すると、タンパク質と基質の座標で構成される” md3_vac.mdcrd” が生成する。この軌跡ファイルに適合するトポロジーファイルは 3.5.2. で作成する (” complex.parm7”)。

次のリストを参考に、構造パラメータを評価する。必要な ptraj入力ファイルは各自、工夫して作成しなさい。

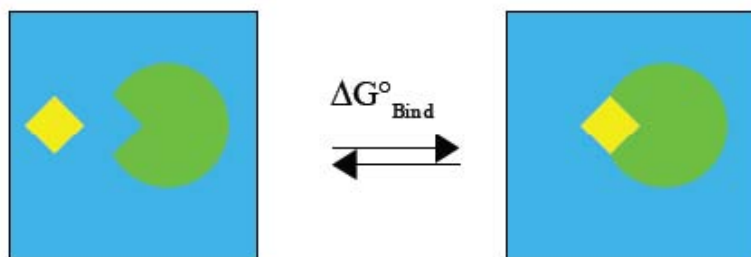
- 1 基質のグリコシド結合回転角 (Φ - Ψ) とメチル水酸基配向 (ω) を評価する。時系列データからパラメータ値の分布を得る。残基によってこれらの構造パラメータが一定のものと、ゆらぎが大きいものがあることを確認し、その理由を考察する。
- 2 基質と基質認識アミノ酸の間の水素結合について解析する。基質認識に関わるアミノ酸の同定には、DeepViewトレーニングを参考にする。
- 3 基質と溶媒分子間の水素結合を解析し、溶媒和との相互作用を基質残基ごとに違いを比較する。
P
- 4 3.5.2. に従って、タンパク質—基質間の相互作用を評価する。
- 5 MM_PBSA計算によるタンパク質—基質相互作用の解析

MM_PBSAとは、タンパク質と基質の間の相互作用または結合エネルギー、 ΔG_{bind} 、を計算する方法である。あらかじめ MD計算によって得られた複合体(complex)の軌跡ファイルをタンパク質(receptor)と基質(ligand)に分割し、時間フレームごとに三つの要素のエネルギー、E、を計算する。

” amber tutorial HP” より

以下の式に従って、真空中の結合エネルギーを計算する。

$$\Delta E_{\text{bind, vacuum}} = E_{\text{complex}} - (E_{\text{receptor}} + E_{\text{ligand}})$$



次に、三つの要素に対して、Poisson-Boltzmann 法(PB)による溶媒和自由エネルギー近似と表面張力による自由エネルギー(SA)の和として溶媒和自由エネルギー、 ΔG_{solv} 、を求め、こ

れらの項目を加えた値を、溶媒和中の結合自由エネルギー、 $\Delta \Delta G_{\text{bind, solv}}$ 、と定義する。

$$\Delta \Delta G_{\text{bind, solv}} = \Delta E_{\text{bind, vacuum}} + \Delta G_{\text{solv, complex}} - (\Delta G_{\text{solv, receptor}} + \Delta G_{\text{solv, ligand}})$$

MM_PBSA法の実施にあたっては、複合体、タンパク質および基質の三つの要素ごとに、トポロジーファイルが必要となる。以下にそれらのトポロジーファイル作成のための LEaP入力

ファイルを示。

```
□MM_PBSAトポロジーファイル作成” leap_pbsa.in” の例 set default writel4scale on source
leaprc.ff99SB source leaprc.GLYCAM_06
set default disulfide off set default PBRadii mbondi2

comp = loadPdb lysozyme_nag3.pdb
bond comp.6.SG comp.127.SG
bond comp.30.SG comp.115.SG
bond comp.64.SG comp.80.SG
bond comp.76.SG comp.94.SG
saveamberParm comp complex.parm7 inpcrd

lig = loadPdb nag3.pdb
saveamberParm lig ligand.parm7 inpcrd

rec = loadPdb lysozyme.pdb
bond rec.6.SG rec.127.SG
bond rec.30.SG rec.115.SG
bond rec.64.SG rec.80.SG
bond rec.76.SG rec.94.SG
saveamberParm rec receptor.parm7 inpcrd

quit
```

これを用いた sleapの実行により、それぞれ、複合体、基質、およびタンパク質に対応する、“complex.parm7”、“ligand.parm7”、および“receptor.parm7”が生成する。

ここでの MM_PBSAの実行には、“MM_PBSA.py”モジュールを使用する。必要なファイルは、既に作成済みの5個の軌跡ファイル(“md3_1.mdcrd.gz”～“md3_5.mdcrd.gz”)とトポロジーファイル(“lysozyme_nag3.parm7”)と上記三種のトポロジーファイルである。さらに、MM_PMSA.pyの入力ファイル“mmpbsa.in”と計算実行のためのスクリプトファイル“MM_PBSA.csh”を作成する(実際の名称は任意)。

- “mmpbsa.in”の例 nput file for running PB &general endframe=1000, interval=10, verbose=1,
/ &pb istrng=0.100,
/ MM_PBSA.csh

1ns軌跡ファイルは1000フレームの座標で構成されるため、最初のフレームからPB/SA計算に供するため、“endframe=1000”となる。計算時間節約のため、10フレームごとにPB/SA計算を実行するため、“interval=10”を設定されている。

- “MMPBSA.ch”の例
#!/bin/csh

```
set INI = 1 set MAX = 5 set NUM = $INI
while ($NUM <= $MAX)
mkdir "$NUM"ns
bsub -K -n 1 -o log/ "$LSF_BINDIR/openmpi_wrapper -np 1 $AMBERHOME/exe/MMPBSA.py -O -i
mmpbsa.in -o FINAL_RESULTS_MMPBSA.dat -sp lysozyme_nag3.parm7 -cp complex.parm7 -rp
receptor.parm7 -lp ligand.parm7 -y ./md3_$NUM.mdcrd.gz"
```

```

mv _MMPBSA_* "$NUM"ns
mv FINAL_RESULTS_MMPBSA.dat "$NUM"ns

@ NUM ++

end

exit

```

- □MMPBSA.ch” を実行する。

```
[tyui@seneca01 mm_pbsa]$ ./MMPBSA.csh >& log &
```

上記のスクリプトの内容に従って、軌跡ファイル □md3_1.mdcrd.gz”

～” md3_5.mdcrd.gz” を順番に読み込み、” MMPBSA.py” モジュールが起動される。計算の

進行に伴い、5 個の軌跡ファイルそれぞれに対して、” 1ns” ～” 5ns” のディレクトリが自動生成してゆく。計算が終了すると、それぞれのディレクトリに下記のファイルが作成されている。

```
[tyui@seneca01 mm_pbsa]$ ls 1ns/ FINAL_RESULTS_MMPBSA.dat _MMPBSA_complex_pb.mdout
_MMPBSA_receptor.mdcrd _MMPBSA_avgcomplex.pdb _MMPBSA_ligand.mdcrd _MMPBSA_receptor_pb.mdout
_MMPBSA_complex.mdcrd _MMPBSA_ligand_pb.mdout [tyui@seneca01 mm_pbsa]$
```

フレームごとの計算結果は” *.mdout” から得られ、軌跡ファイル全体の平均値は”

FINAL_RESULTS_MMPBSA.dat” にまとめられている。必要なのは、このファイルの最終部分である。

z” FINAL_RESULTS_MMPBSA.dat” の最終部分の例(”1ns” ディレクトリより)

VDWAALS	-36.4332		3.0667
EEL	-62.4045		14.4958
EPB	77.8339		12.1157
ECAVITY	-3.3025		0.2765
DELTA G gas	-98.8378		219.5331
DELTA G solv	74.5313		12.1188
DELTA G binding =	-24.3064	+/	4.5800

(上部省略) Differences (Complex - Receptor - Ligand): Energy Component Average Std. Dev. Std. Err. of Mean

WARNINGS: igb=5 should be used with either mbondi2 or bondi pbradii set. Yours are modified Bondi radii (mbondi)

“DELTA Gbinding” の値が結合自由エネルギー ($\Delta \Delta G_{bind, solv}$) の平均値となる。他、” DELTA G gas” が真空中の結合エネルギー ($\Delta E_{bind, vacuum}$)、” DELTA G solv” は複合体形成に伴う溶媒和自由エネルギーの変化量 ($\Delta G_{solv, complex} - (\Delta G_{solv, receptor} + \Delta G_{solv, ligand})$) の、それぞれ、平均値である。それらの値のあとに、標準偏差と偏差平均が示されている。