

2013年8月6日(火)13:00～

戦略分野1「予測する生命科学・医療および創薬基盤」

場所/共催: バイオグリッドセンター関西

SCLS計算機システムの実習 —分子動力学の基礎—

理化学研究所

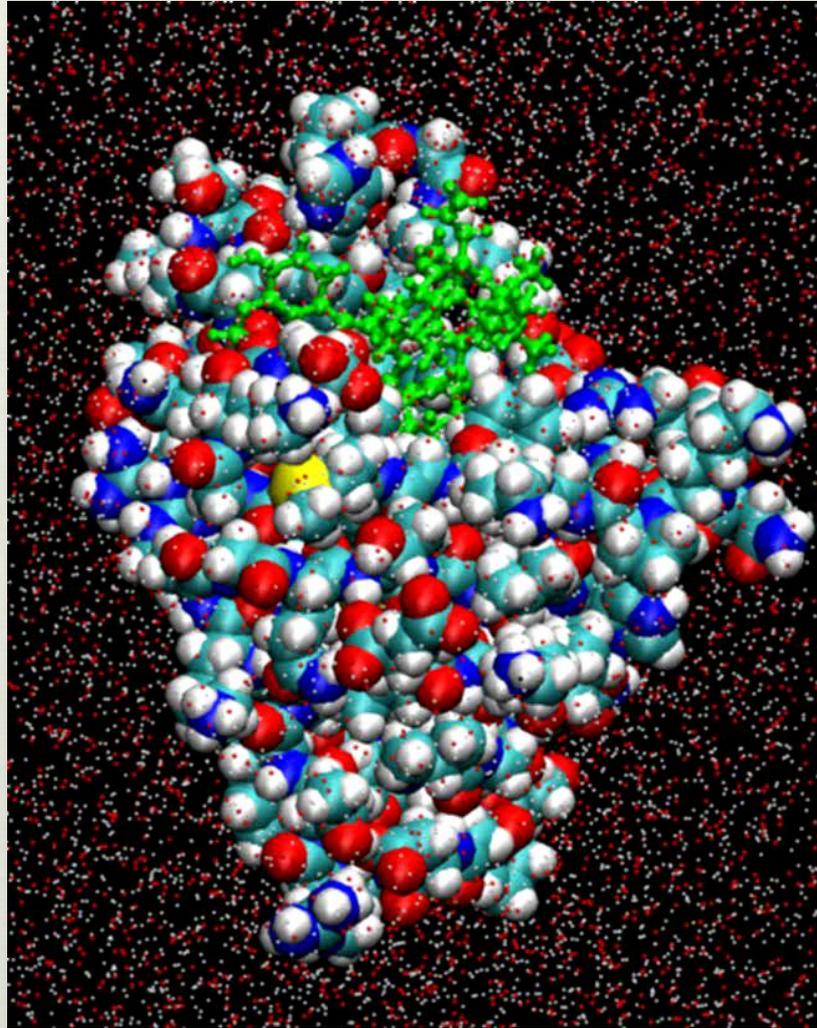
HPCI計算生命科学推進プログラム

副プログラムディレクタ 江口至洋

yeguchi@riken.jp

はじめに

タンパク質の分子動力学計算



分子動力学計算

- ❖ 原子は「電荷と質量を持つ質点」とする
- ❖ ニュートンの運動方程式を解く
- ❖ 力 F やポテンシャル関数 V の構成原理はない
- ❖ ポテンシャル関数は経験的に決められる
- ❖ そこに含まれるパラメータは、分光学などの実験結果を踏まえ、あるいは量子論により、さらには実験結果と整合性が取れるように決められている
- ❖ それらパラメータを用いた分子動力学計算は、広範な実験結果を説明し、かつ新たな知見を生み出している
- ❖ ただし、「分子動力学計算が常に最適な選択」というわけではない
 - ❖ 粗視化モデル(例:1残基を1質点でモデル化)、ブラウン動力学、量子論

分子動力学計算の歴史

- ❖ **McCammon et al. (1977)** Dynamics of folded proteins. *Nature*, **267**, 585-590 初めてタンパク質 (BPTI、58残基) の分子動力学計算を行う。0.978fs刻みで8.8psのシミュレーション。
- ❖ **Brooks et al. (1983)** CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
- ❖ **Pearlman DA et al. (1995)** AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp. Phys. Commun.*, **91**, 1–41.
- ❖ **Berendsen H et al. (1995)** GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56
- ❖ **Kale L et al. (1999)** NAMD2: greater scalability for parallel molecular dynamics. *J. Comput. Phys.* **151**, 283–312.
- ❖ **Lindorff-Larsen K et al. (2011)** How Fast-Folding Proteins Fold. *Science*, **334**, 517-520 10~80残基のタンパク質の分子動力学計算を専用計算機を用いて100 μ sから1ms行い、折り畳み過程を解析して。
- ❖ **Zhao G et al. (2013)** Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*, **497**, 643–646 ペタスケールの計算機 (Blue Waters) を用いてはじめて、1,300以上もの同一のタンパク質 (6,400万原子) からなる巨大なHIVカプシドの原子レベルでの構造を明らかにすることができた。

分子動力学計算の贈り物

- ❖ 計算機顕微鏡として分子の動的動きを見る
- ❖ 自由エネルギーを求める(例:結合の自由エネルギー)
- ❖ 自己拡散係数などの統計量を求める
- ❖ タンパク質の折り畳み過程を見る
- ❖ 生体分子の相互作用に伴う構造変化過程を見る
- ❖ 立体構造の精密化(refinement)を行う

ポテンシャル関数 V

empirical potential energy function

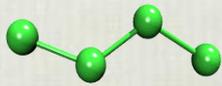
ニュートンの運動方程式？

$$m_i \frac{d^2 \vec{r}_i(t)}{dt^2} = \vec{F}_i(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)$$

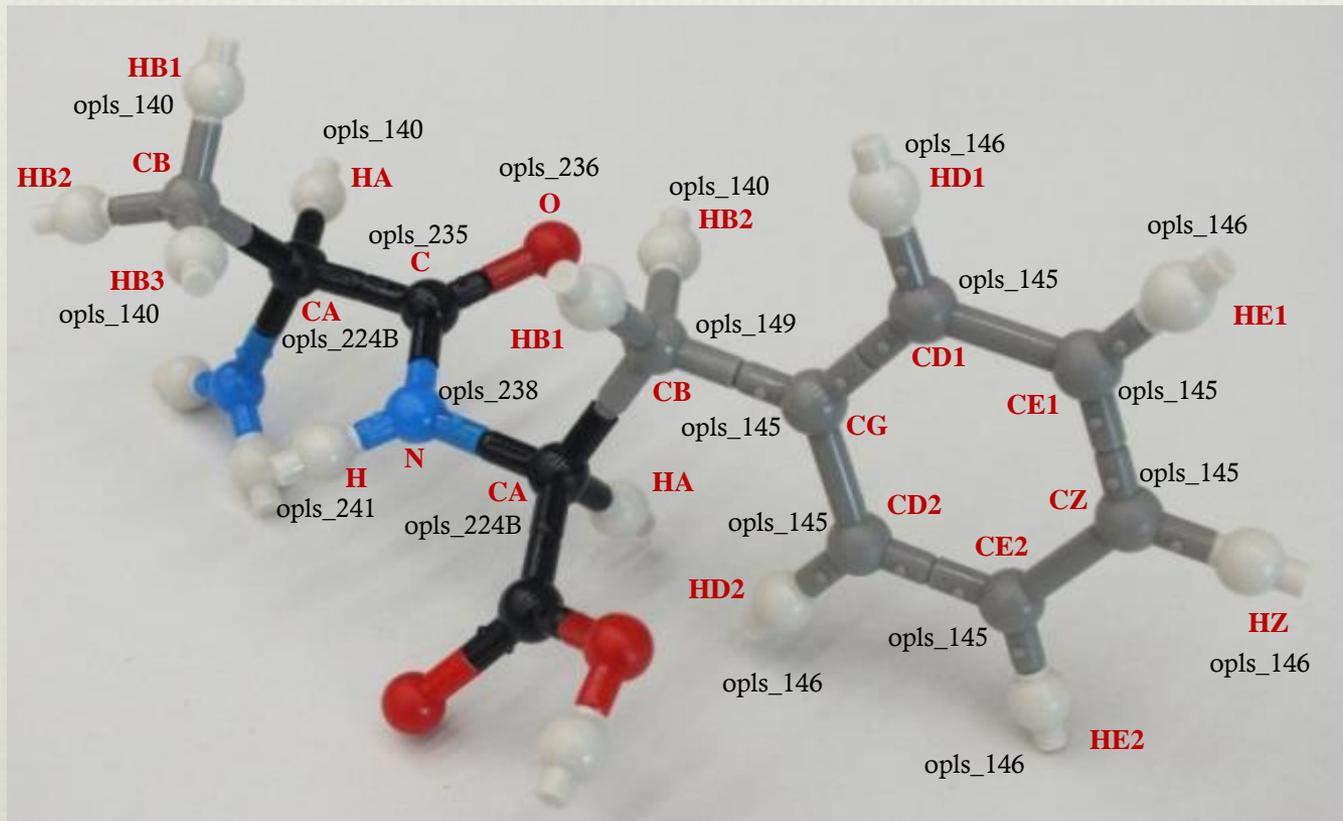
$$\vec{F}_i = - \frac{\partial V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)}{\partial \vec{r}_i}$$

V?

標準的なポテンシャル関数 V

$$\begin{aligned} V = & \sum_{\text{bonds}} \frac{1}{2} K_{ij}^b (r_{ij} - b_{ij})^2 + \sum_{\text{angles}} \frac{1}{2} K_{ijk}^\theta (\theta_{ijk} - \theta_{ijk}^0)^2 \\ & + \sum_{\text{dihedrals}} K_\varphi (1 + \cos(n\varphi - \varphi_0)) \\ & + \sum_{i \neq j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i \neq j} \frac{q_i q_j}{\varepsilon_0 r_{ij}} \end{aligned}$$


原子と原子タイプ



アラニル-フェニルアラニン

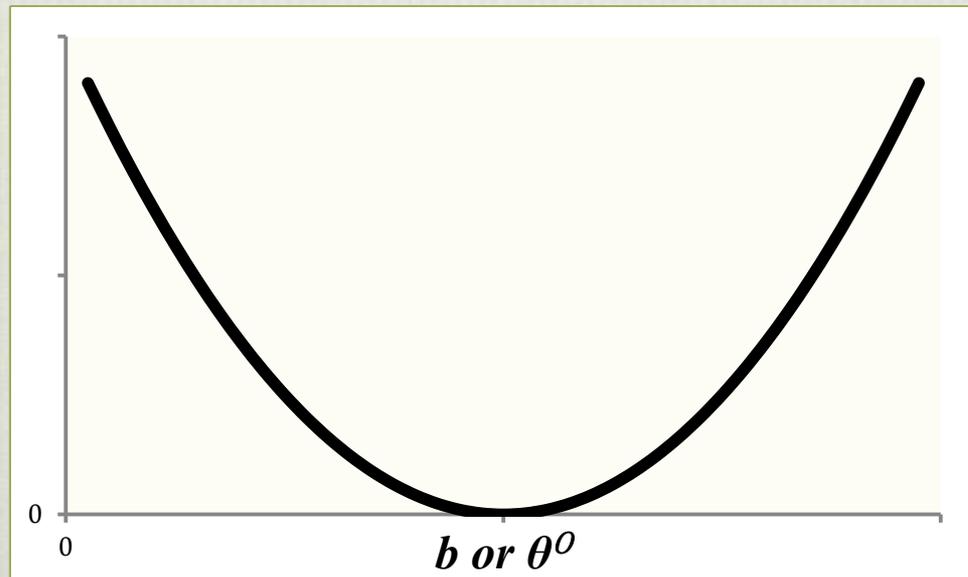
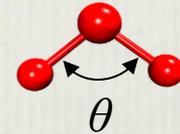
(原子および原子タイプはGROMACSで力場OPLS-AA/Lを用いたLysozyme解析用.top
ファイルから作成)

結合ポテンシャルと結合角ポテンシャル

$$V_{bond} = \sum_{bonds} \frac{1}{2} K_{ij}^b (r_{ij} - b_{ij})^2$$



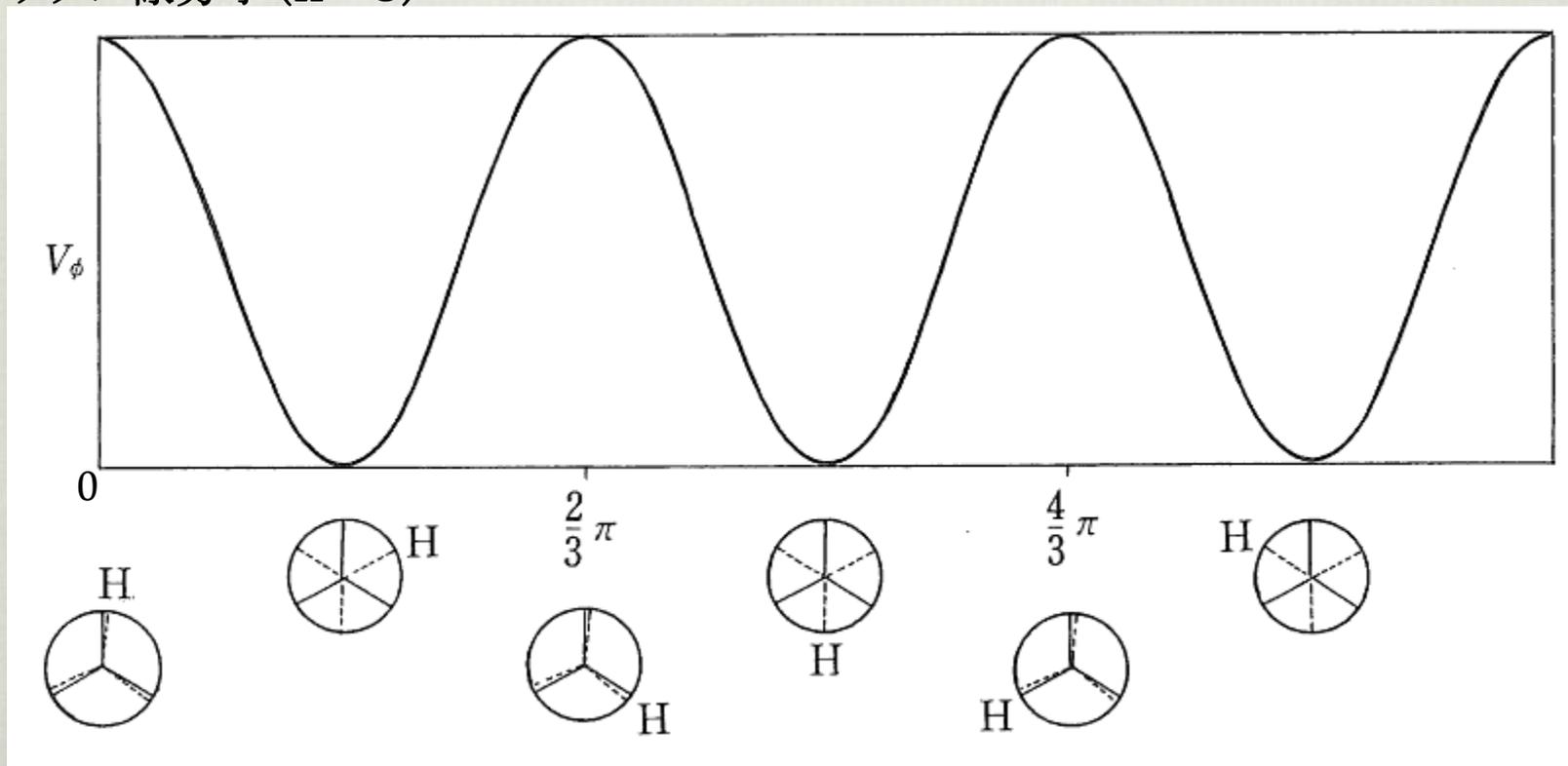
$$V_{angl} = \sum_{angles} \frac{1}{2} K_{ijk}^\theta (\theta_{ijk} - \theta_{ijk}^0)^2$$



分子内束縛回転ポテンシャル(歪ポテンシャル)

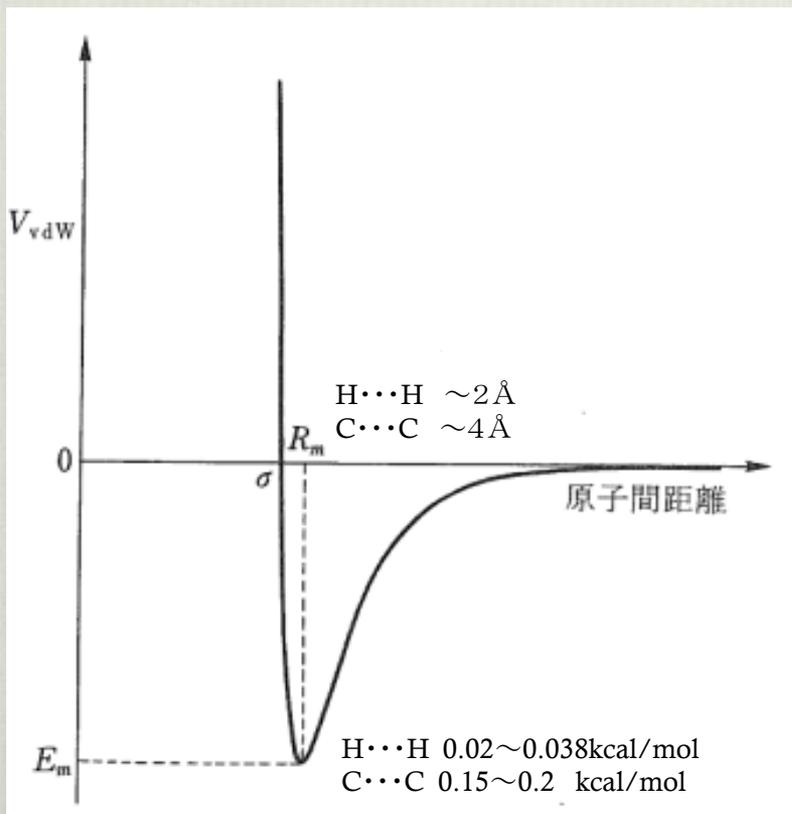
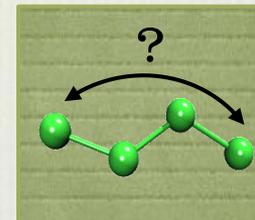
$$V_{\varphi} = \sum_{\text{dihedrals}} K_{\varphi} (1 + \cos(n\varphi - \varphi_0))$$

メタン様分子 ($n=3$)



van der Waalsポテンシャル

$$V_{vdW} = \sum_{i \neq j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$



全ての原子タイプ*i,j*について求めるのは困難なため、例えば、Lorentz-Berthelotの組み合わせ則が用いられる。

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2}$$

$$\epsilon_{ij} = \sqrt{\epsilon_{ii} \epsilon_{jj}}$$

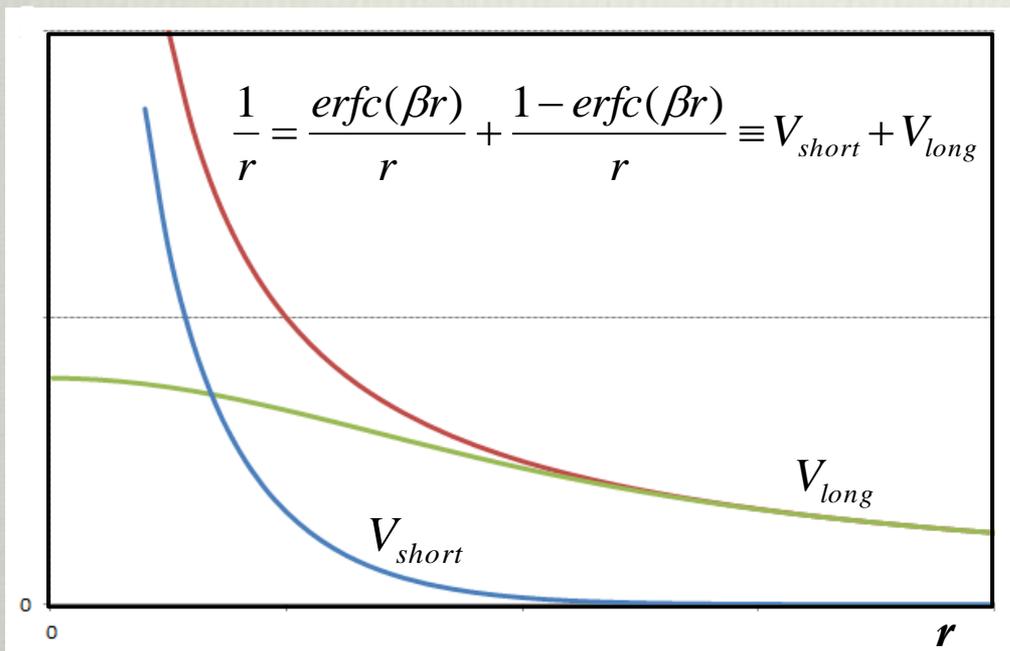
OPLS力場では両方で幾何平均が用いられている。

$$\sigma_{ij} = \sqrt{\sigma_{ii} \sigma_{jj}}$$

$$\epsilon_{ij} = \sqrt{\epsilon_{ii} \epsilon_{jj}}$$

静電ポテンシャル

$$V_{ele} = \sum_{i \neq j} \frac{q_i q_j}{\epsilon_0 r_{ij}}$$

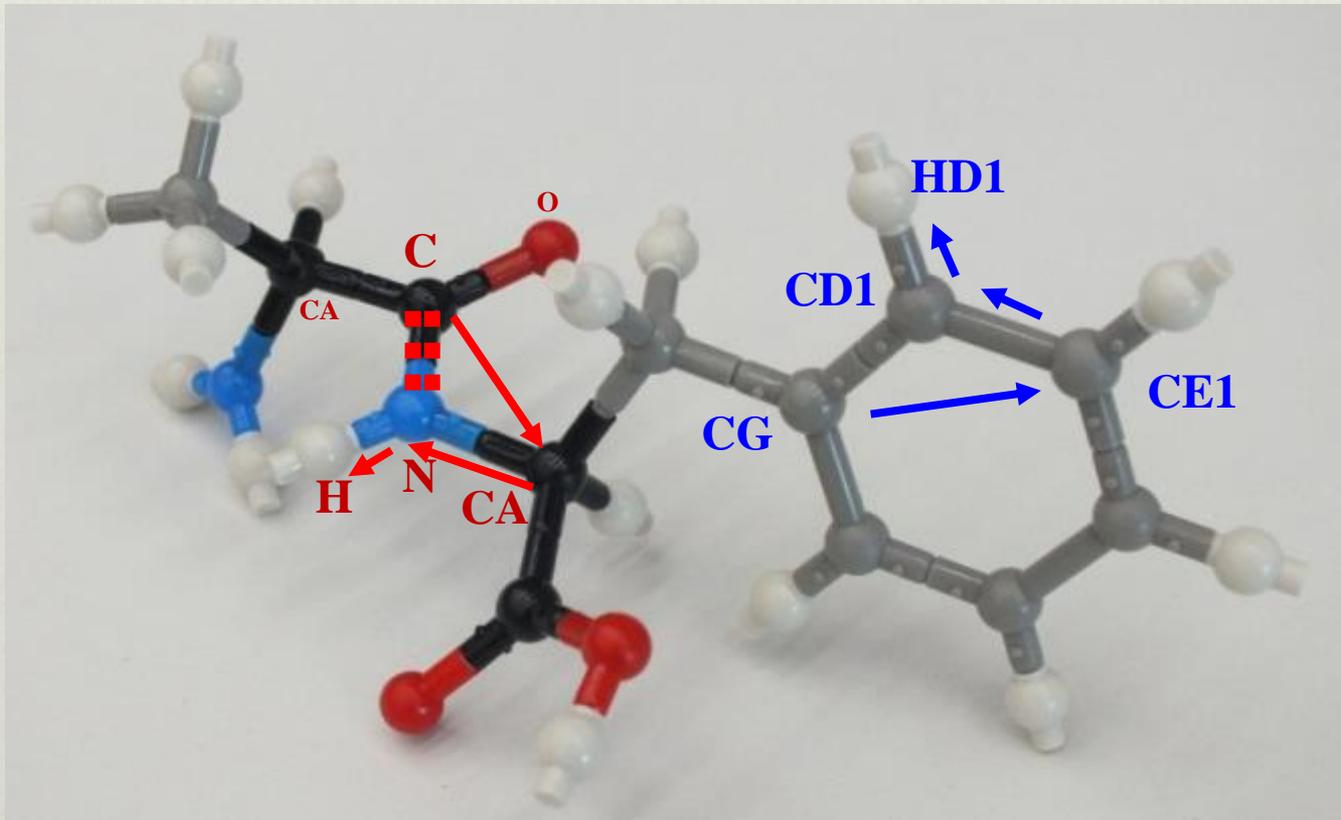


Ewald法

1. V_{short} はカットオフ近似で求める。
2. V_{long} は周期境界条件のイメージセルに関して無限遠のものまで考慮して求める。

広義の回転ポテンシャル

$$V_{improper} = \sum_{dihedrals} K_{\varphi} (1 + \cos(n\varphi - \varphi_0))$$



アラニル-フェニルアラニン

(原子および原子タイプはGROMACSで力場OPLS-AA/Lを用いたLysozyme解析用.topファイルから作成)

運動方程式を解く

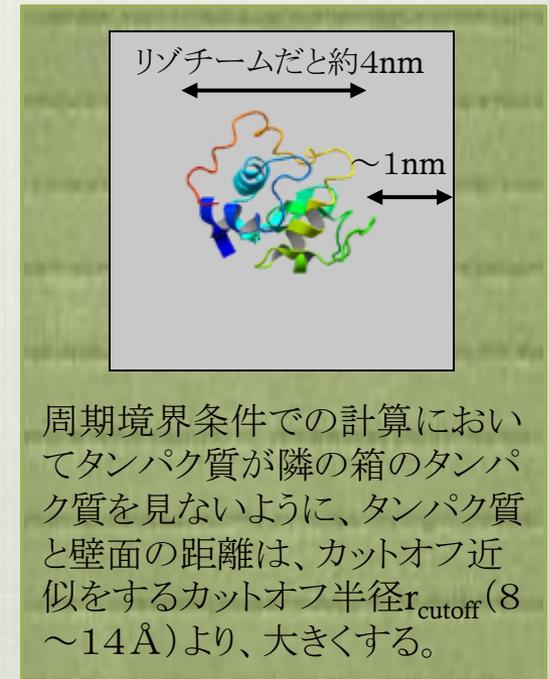
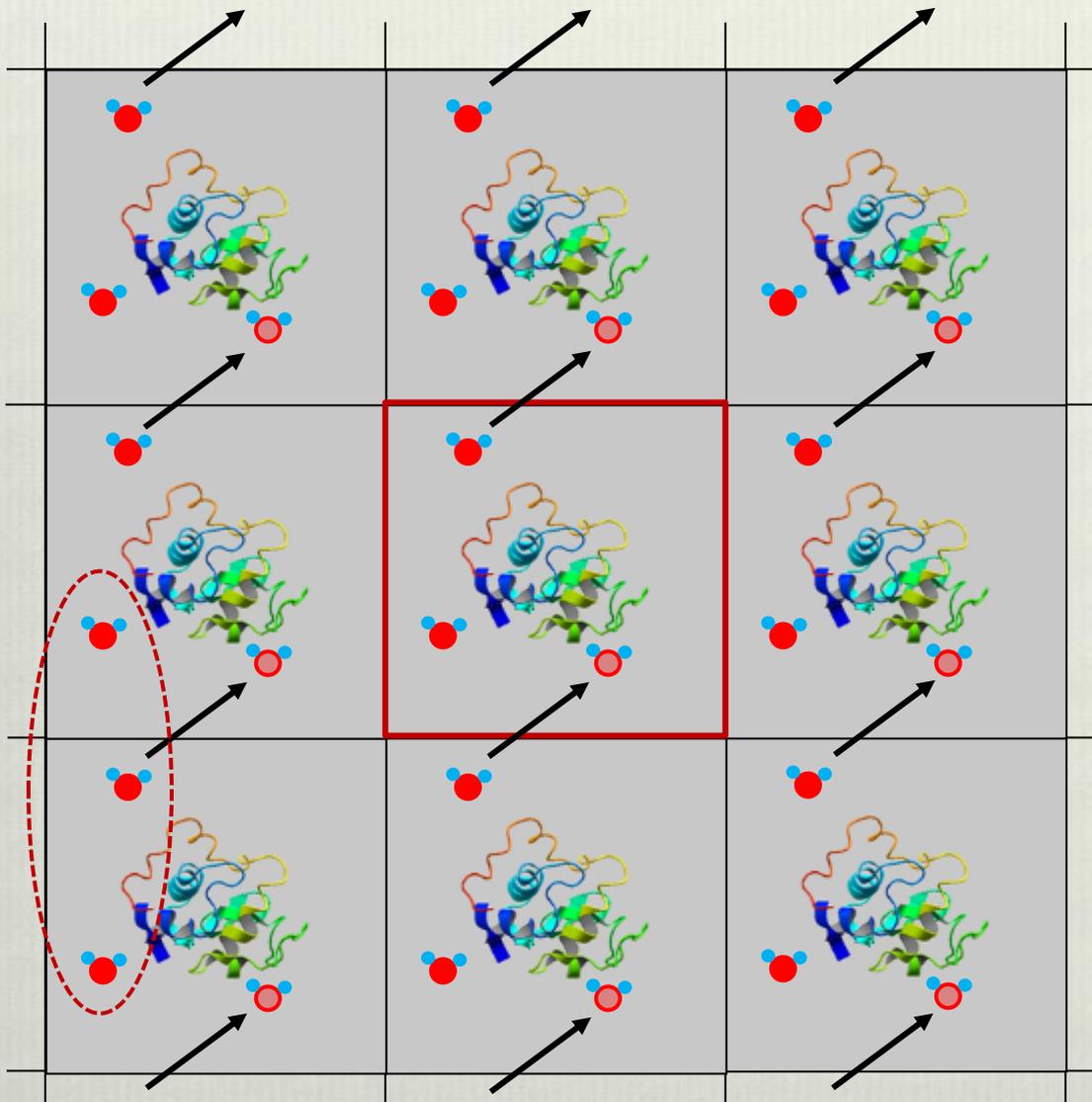
ニュートンの運動方程式の数値解を求める(常微分方程式の初期値問題)

ニュートンの運動方程式を解く

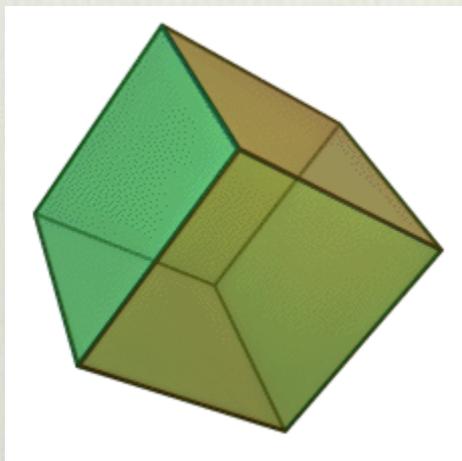
$$m_i \frac{d^2 \vec{r}_i(t)}{dt^2} = \vec{F}_i(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)$$

$$\vec{F}_i = - \frac{\partial V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)}{\partial \vec{r}_i}$$

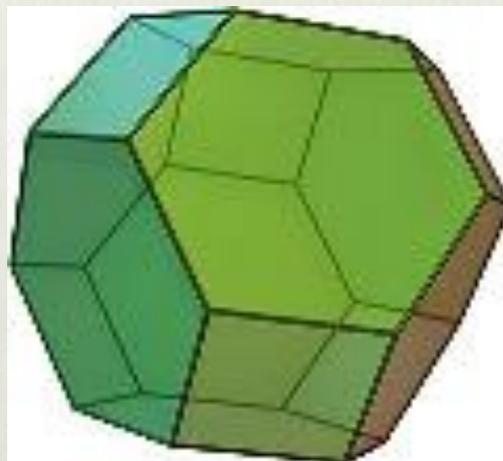
周期境界条件



シミュレーションに用いる箱 (Unit Cell)



正六面体

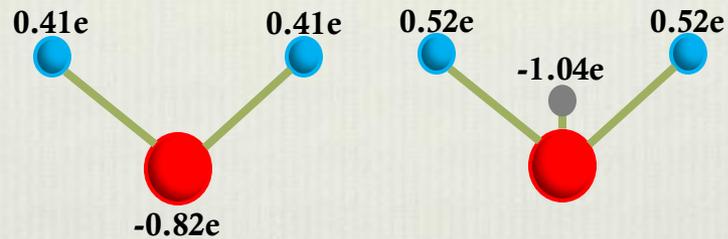


切頂八面体



ほぼ球状のタンパク質には、ほぼ球に近い切頂八面体が適している。
溶媒(水)の数は少ない方がいい。→切頂八面体は正六面体の77%。
(”Lysozyme in Water”の実習では、1辺7.01008nmの正六面体の中にリゾチーム1個、水分子10,824個、Cl⁻イオン8個で出発。NVTアンサンブル、NPTアンサンブルの平衡化終了時は、6.98842nmの正六面体になっている。)
運動量を保存するため、向かい合う面は平行にしておく。

水のモデル



SPC
(他にSPC/Eなど)

TIP4P
(他に TIP4P-Ewなど)

実験値
(25°C、1気圧)

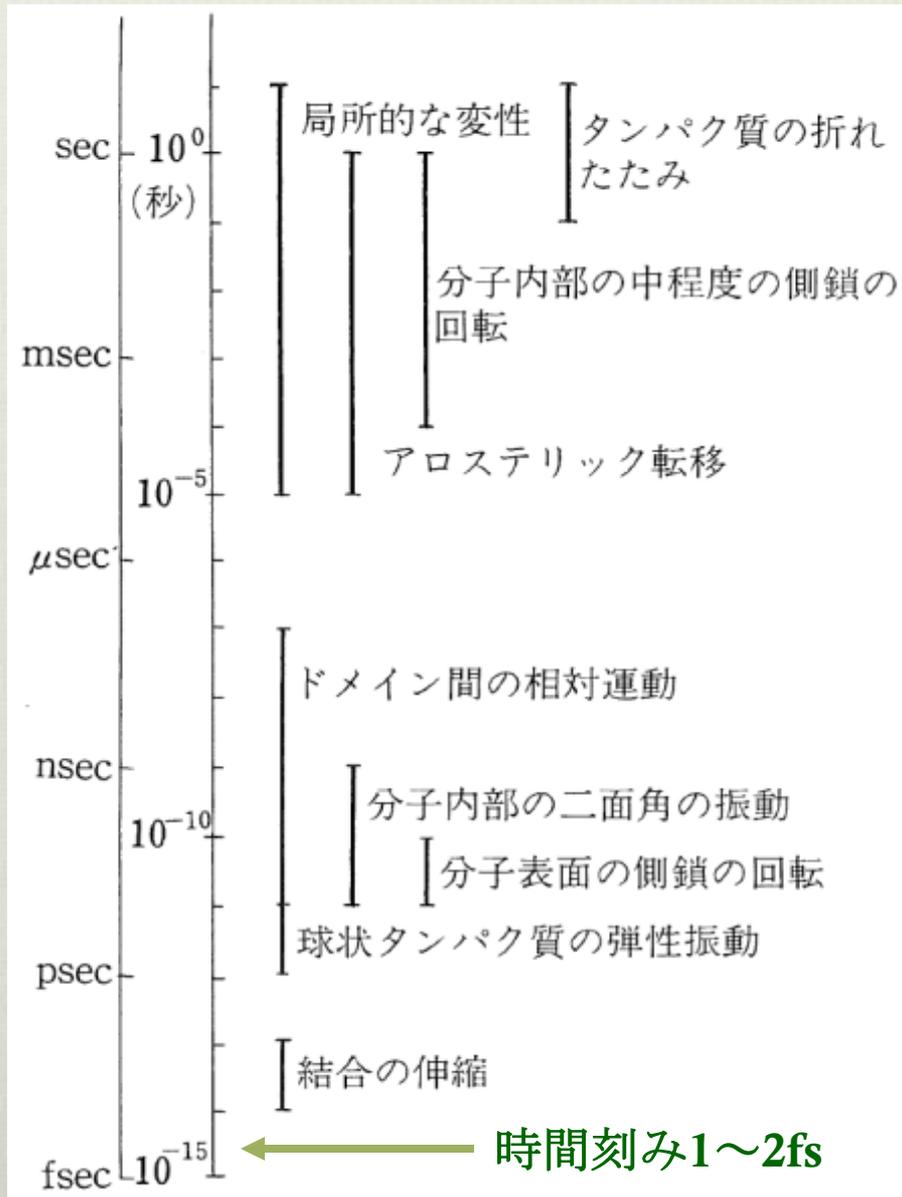
比重 (g/cm ³)	0.971	0.999	0.997
定圧比熱 (cal/mol/deg)	23.4	19.3	17.99
等温圧縮率 (atm ⁻¹)	27 × 10 ⁻⁶	35 × 10 ⁻⁶	45.8 × 10 ⁻⁶

注)

1L=10²⁴nm³に55.4×6.02²³=3.34×10²⁵個の水が存在する。

1nm³には33.4個の水がある。

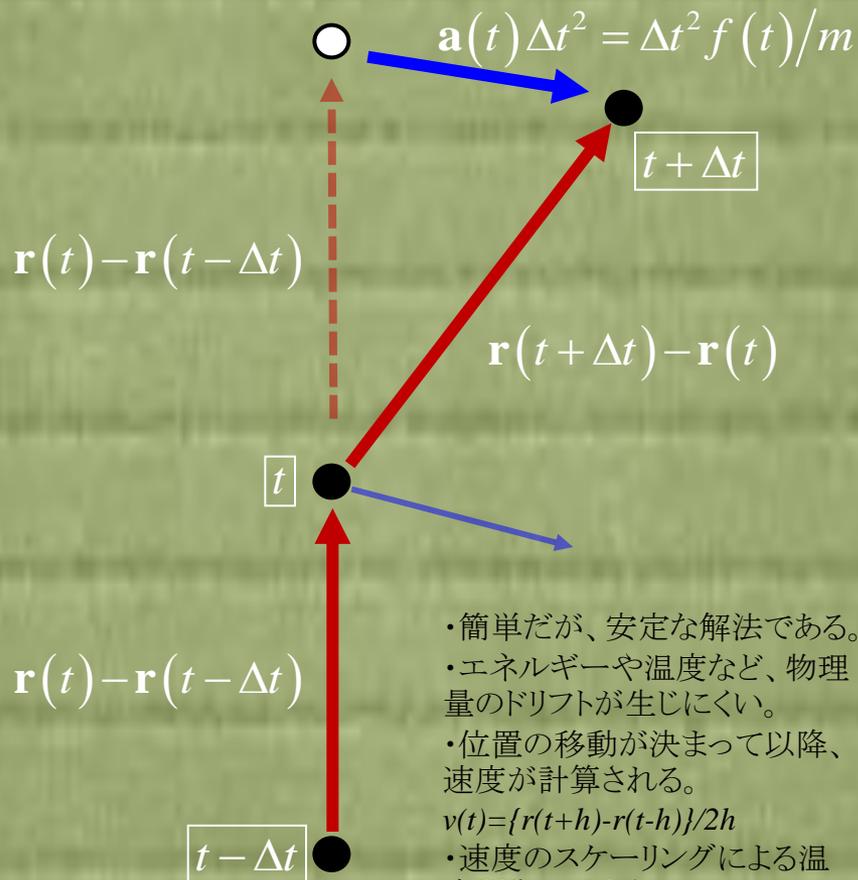
タンパク質の時間構造と時間刻み



ニュートンの運動方程式の数値解法

Verlet法

$$\mathbf{r}(t+\Delta t) - \mathbf{r}(t) = \mathbf{r}(t) - \mathbf{r}(t-\Delta t) + \mathbf{a}(t)\Delta t^2$$



- 簡単だが、安定な解法である。
- エネルギーや温度など、物理量のドリフトが生じにくい。
- 位置の移動が決まって以降、速度が計算される。
- $v(t) = \{r(t+h) - r(t-h)\} / 2h$
- 速度のスケールリングによる温度制御はできない。

leapfrog法

$$v\left(t + \frac{\Delta t}{2}\right) = v\left(t - \frac{\Delta t}{2}\right) + \frac{f}{m} \Delta t$$

$$r(t + \Delta t) = r(t) + v\left(t + \frac{\Delta t}{2}\right) \Delta t$$

$$v(t) = \frac{1}{2} \left\{ v\left(t + \frac{\Delta t}{2}\right) + v\left(t - \frac{\Delta t}{2}\right) \right\}$$

- 解法において速度が陽に現れるため、温度などを速度のスケールリングによって直接制御できる。
- 差を求める操作がないため、数値計算上の誤差は生じにくい

温度と圧力の制御

❖ NVEアンサンブル(ミクロカノニカル・アンサンブル)

- ❖ 普通にNewtonの運動方程式を解く
- ❖ 温度や圧力は計算した結果から求める

$$T(t) = \frac{1}{3Nk_B} \sum_{i=1}^N m_i \vec{v}_i \cdot \vec{v}_i$$
$$P(t) = \frac{Nk_B T}{V} + \frac{1}{3V} \sum_{i=1}^N \vec{r}_i \cdot \vec{F}_i$$
$$= \frac{Nk_B T}{V} + \frac{1}{3V} \sum_{i=1}^N \sum_{j=i+1}^N \vec{r}_{i \leftarrow j} \cdot \vec{F}_{i \leftarrow j}$$

❖ 温度の制御

- ❖ ベレンゼン法(瞬間毎の温度を設定温度 T_0 に徐々に近づける)

- ❖ 速度スケーリング法

$$m_i \frac{d^2 r_i(t)}{dt^2} = F_i(r_1, r_2, \dots, r_N) + \frac{1}{\tau} \left(\frac{T_0}{T(t)} - 1 \right) m_i \frac{dr_i}{dt} \quad \tau = 0.01 \sim 0.5 \text{ps}$$

- ❖ 能勢・フーバー法、能勢・ポアンカレ法(NVTアンサンブルを生成する)

❖ 温度と圧力の制御

- ❖ ベレンゼン法(T) + パリネロ・ラーマン法(P)
- ❖ 能勢(T)・アンダーセン(P)法(NPTアンサンブルを生成する)

分子動力学計算の実行

Lysozyme in waterの実習を例に

典型的なシミュレーションの流れ

- ❖ 構造を得る (通常はPDBから)
- ❖ 欠失領域を埋める (含む、側鎖や水素原子)
- ❖ SS結合の確認、ヒスチジンなどの電荷状態の確認など
- ❖ GROMACSのトポロジーファイルを作る
- ❖ 水分子を付加する (イオンを付加する)
- ❖ 最急降下法などでエネルギーの最小化を行う
- ❖ 平衡化シミュレーションを行う (例: NVT→NPTアンサンブル)
- ❖ プロダクト・ランを行う (タンパク質の重原子を自由にして)
- ❖ シミュレーション結果 (trajectory data) を解析する

PDBのATOMレコードとGROMACSの.topファイル

Record name	Atom Serial No.	Atom name	Res. name	Chain ID	Res. Seq. No.	Orthogonal coordinates			Occupancy	Temperature factor	Element symbol
ATOM	1	N	LYS	A	1	35.365	22.342	-11.980	1.00	22.28	N
ATOM	2	CA	LYS	A	1	35.892	21.073	-11.427	1.00	21.12	C
ATOM	3	C	LYS	A	1	34.741	20.264	-10.844	1.00	16.85	C
ATOM	4	O	LYS	A	1	33.945	20.813	-10.081	1.00	18.94	O
ATOM	5	CB	LYS	A	1	26.872	21.425	-10.206	1.00	20.78	C
ATOM	6	CG	LYS	A	1						
ATOM	7	CD	LYS	A	1						
ATOM	8	CE	LYS	A	1						
ATOM	9	NZ	LYS	A	1						
ATOM	10	N	VAL	A	2						

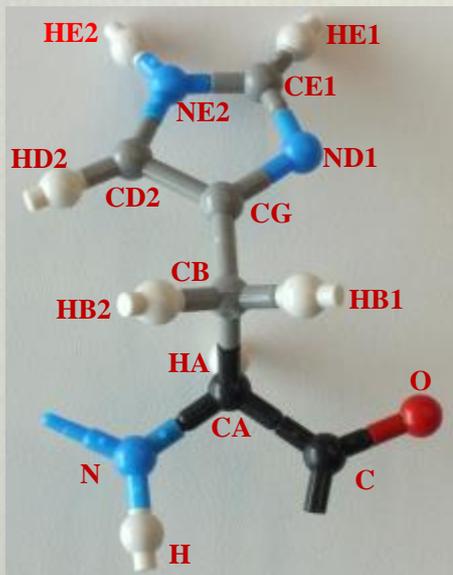
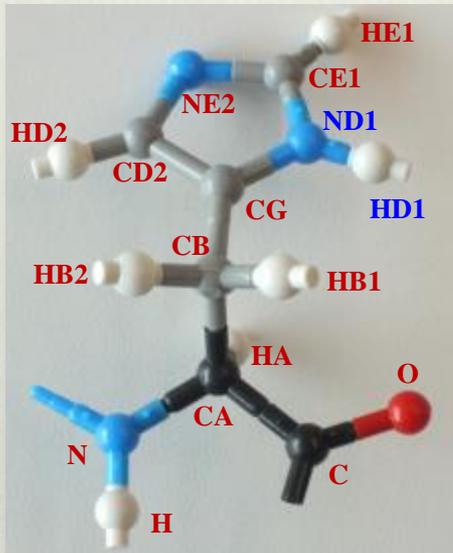
PDBのATOMレコード

```
[ atoms ]
; nr      type  resnr residue  atom  cgnr  charge  mass
; residue 1 LYS rtp LYSH q +2.0
  1  opls_287  1  LYS  N  1  -0.3  14.0067 ; qtot -0.3
  2  opls_290  1  LYS  H1  1  0.33  1.008 ; qtot 0.03
  3  opls_290  1  LYS  H2  1  0.33  1.008 ; qtot 0.36
  4  opls_290  1  LYS  H3  1  0.33  1.008 ; qtot 0.69
  5  opls_293B 1  LYS  CA  1  0.25  12.011 ; qtot 0.94
  6  opls_140  1  LYS  HA  1  0.06  1.008 ; qtot 1
  7  opls_136  1  LYS  CB  2 -0.12  12.011 ; qtot 0.88
  8  opls_140  1  LYS  HB1 2  0.06  1.008 ; qtot 0.94
  9  opls_140  1  LYS  HB2 2  0.06  1.008 ; qtot 1
 10  opls_136  1  LYS  CG  3 -0.12  12.011 ; qtot 0.88
 11  opls_140  1  LYS  HG1  3  0.06  1.008 ; qtot 0.94
 12  opls_140  1  LYS  HG2  3  0.06  1.008 ; qtot 1
 13  opls_136  1  LYS  CD  4 -0.12  12.011 ; qtot 0.88
 14  opls_140  1  LYS  HD1  4  0.06  1.008 ; qtot 0.94
 15  opls_140  1  LYS  HD2  4  0.06  1.008 ; qtot 1
 16  opls_292  1  LYS  CE  5  0.19  12.011 ; qtot 1.19
 17  opls_140  1  LYS  HE1  5  0.06  1.008 ; qtot 1.25
 18  opls_140  1  LYS  HE2  5  0.06  1.008 ; qtot 1.31
 19  opls_287  1  LYS  NZ  6 -0.3  14.0067 ; qtot 1.01
 20  opls_290  1  LYS  HZ1  6  0.33  1.008 ; qtot 1.34
 21  opls_290  1  LYS  HZ2  6  0.33  1.008 ; qtot 1.67
 22  opls_290  1  LYS  HZ3  6  0.33  1.008 ; qtot 2
 23  opls_235  1  LYS  C  7  0.5  12.011 ; qtot 2.5
 24  opls_236  1  LYS  O  7 -0.5  15.9994 ; qtot 2
```

GROMACSの.topファイル

charge group(cgnr)は現在ではneighbor-search groupの役割としてある。

ヒスチジンの構造 (OPLS力場での例)



229	opls_238	15	HIS	N	75	-0.5	14.0067	; qtot 3.5
230	opls_241	15	HIS	H	75	0.3	1.008	; qtot 3.8
231	opls_224B	15	HIS	CA	75	0.14	12.011	; qtot 3.94
232	opls_140	15	HIS	HA	75	0.06	1.008	; qtot 4
233	opls_505	15	HIS	CB	76	-0.297	12.011	; qtot 3.703
234	opls_140	15	HIS	HB1	76	0.06	1.008	; qtot 3.763
235	opls_140	15	HIS	HB2	76	0.06	1.008	; qtot 3.823
236	opls_507	15	HIS	CG	77	0.504	12.011	; qtot 4.327
237	opls_511	15	HIS	ND1	77	-0.564	14.0067	; qtot 3.763
238	opls_508	15	HIS	CD2	78	-0.261	12.011	; qtot 3.502
239	opls_146	15	HIS	HD2	78	0.183	1.008	; qtot 3.685
240	opls_506	15	HIS	CE1	79	0.182	12.011	; qtot 3.867
241	opls_146	15	HIS	HE1	79	0.098	1.008	; qtot 3.965
242	opls_503	15	HIS	NE2	80	-0.291	14.0067	; qtot 3.674
243	opls_504	15	HIS	HE2	80	0.326	1.008	; qtot 4
244	opls_235	15	HIS	C	81	0.5	12.011	; qtot 4.5
245	opls_236	15	HIS	0	81	-0.5	15.9994	; qtot 4

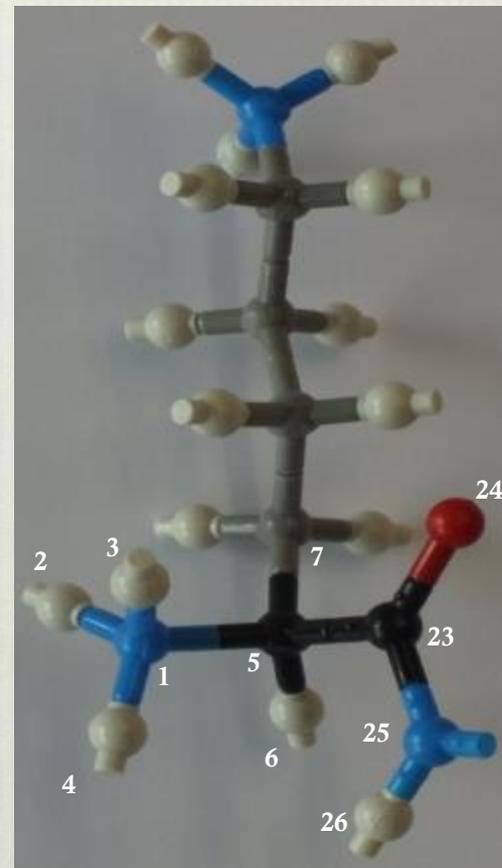
GROMACSの.topファイル



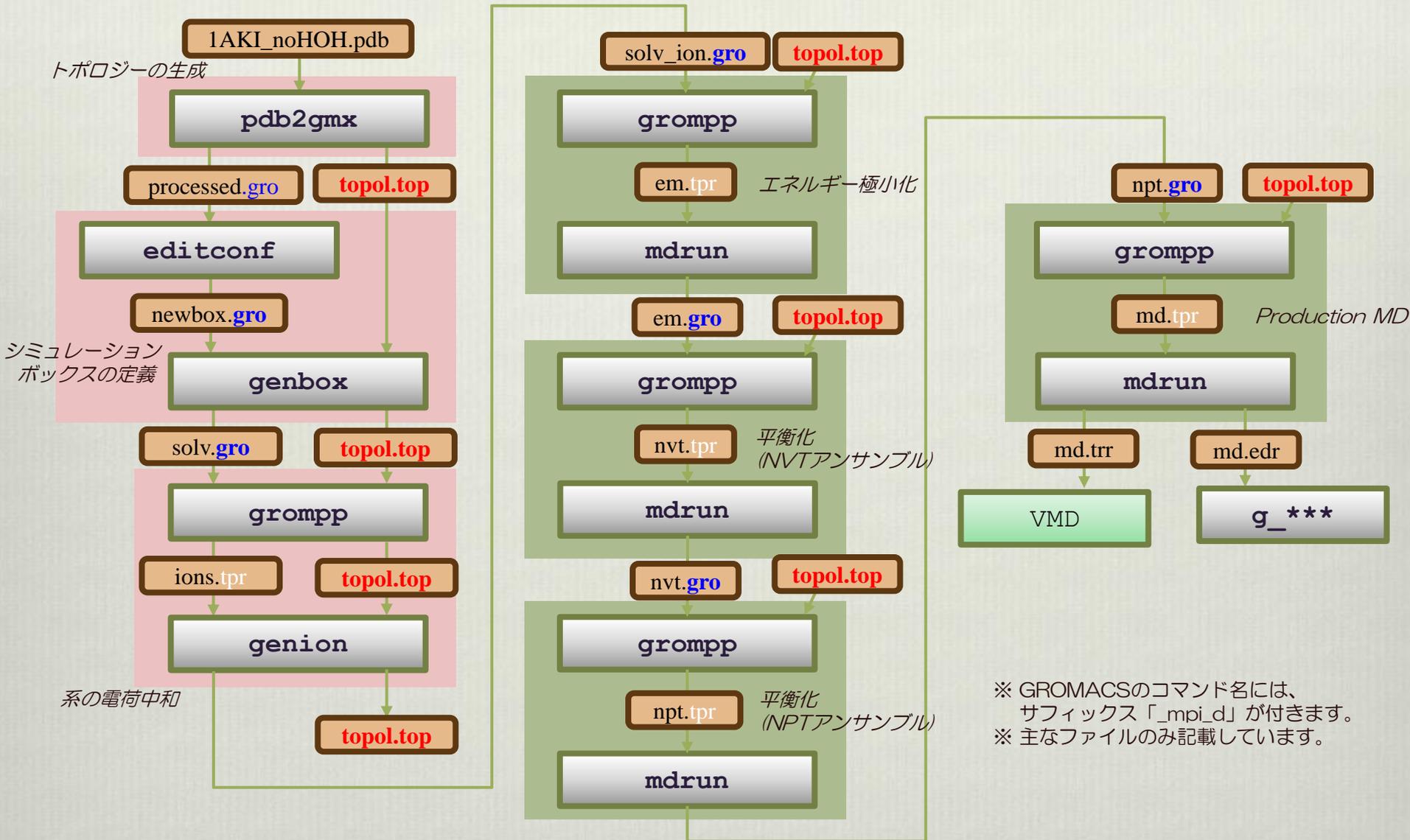
トポロジー(.top)ファイル

```
[ atoms ]
; nr      type  resnr residue  atom  cgnr    charge    mass
; residue 1 LYS  rtp LYSH q +2.0
  1  opls_287  1  LYS    N    1     -0.3    14.0067 ; qtot -0.3
  2  opls_290  1  LYS    H1   1     0.33    1.008   ; qtot 0.03
  3  opls_290  1  LYS    H2   1     0.33    1.008   ; qtot 0.36
  4  opls_290  1  LYS    H3   1     0.33    1.008   ; qtot 0.69
  5  opls_293B 1  LYS    CA   1     0.25   12.011  ; qtot 0.94
  6  opls_140  1  LYS    HA   1     0.06    1.008   ; qtot 1
  7  opls_136  1  LYS    CB   2    -0.12   12.011  ; qtot 0.88
  . . .
 23  opls_235  1  LYS    C    7     0.5    12.011  ; qtot 2.5
 24  opls_236  1  LYS    O    7    -0.5   15.9994 ; qtot 2
; residue 2 VAL  rtp VAL  q 0.0
 25  opls_238  2  VAL    N    8    -0.5   14.0067 ; qtot 1.5
  . . .
[ bonds ]
; ai  aj  funct
  1  2  1
  1  3  1
  1  4  1
  1  5  1
  . . .
[ angles ]
; ai  aj  ak  funct
  2  1  3  1
  . . .
  3  1  5  1
  . . .
  1  5  23 1
  . . .
[ dihedrals ]
; ai  aj  ak  al  funct
  2  1  5  6  3
  2  1  5  7  3
  2  1  5  23 3
  . . .
[ dihedrals ]
; ai  aj  ak  al  funct
  5  25  23  24  1  improper_O_C_X_Y
 23  27  25  26  1  improper_Z_N_X_Y
  . . .
```

GROMACSの.topファイル



実習でのシミュレーションの流れ





戦略分野1「予測する生命科学：医療および創薬基盤」
理化学研究所 HPCI計算生命科学推進プログラム
のみなさんに感謝します