

平成22年6月14日

分子モデリングと分子シミュレーション

分子シミュレーションの応用

東京大学大学院農学生命科学研究科

アグリバイオインフォマティクス

教育研究プログラム

寺田 透

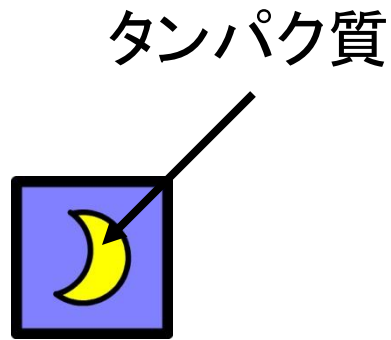
本日の講義内容

- 実験データとの比較
- 分子シミュレーションの応用事例
 - RecA・DNA複合体のシミュレーション
 - Simulated annealing
 - フォールディングシミュレーション
- ドッキングシミュレーション実習
 - 課題

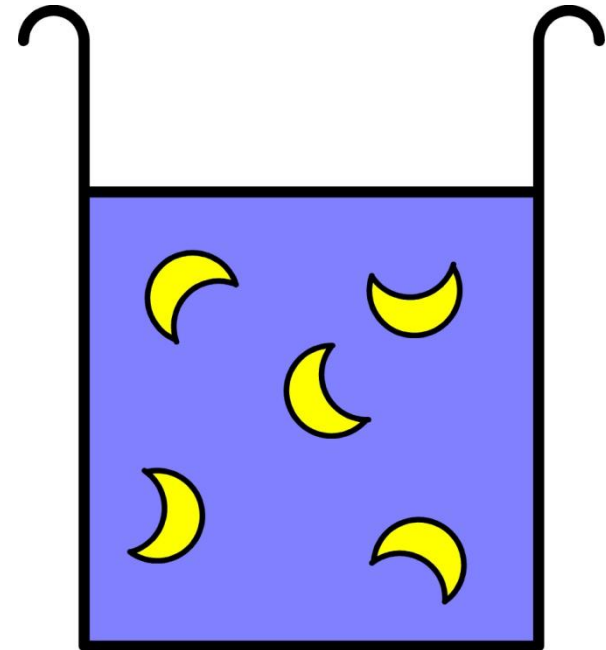
実験データとの比較

- 単にニュートン方程式を解くと、その解は、周辺環境とエネルギーのやり取りのない「孤立系」のシミュレーションを行ったことになる
- 実際の実験データは、分子が 10^{23} 個程度存在する系に対する平均的な値として得られる
- 分子シミュレーションで得られる結果は、実験データと比較できるのだろうか？

現実のシステム

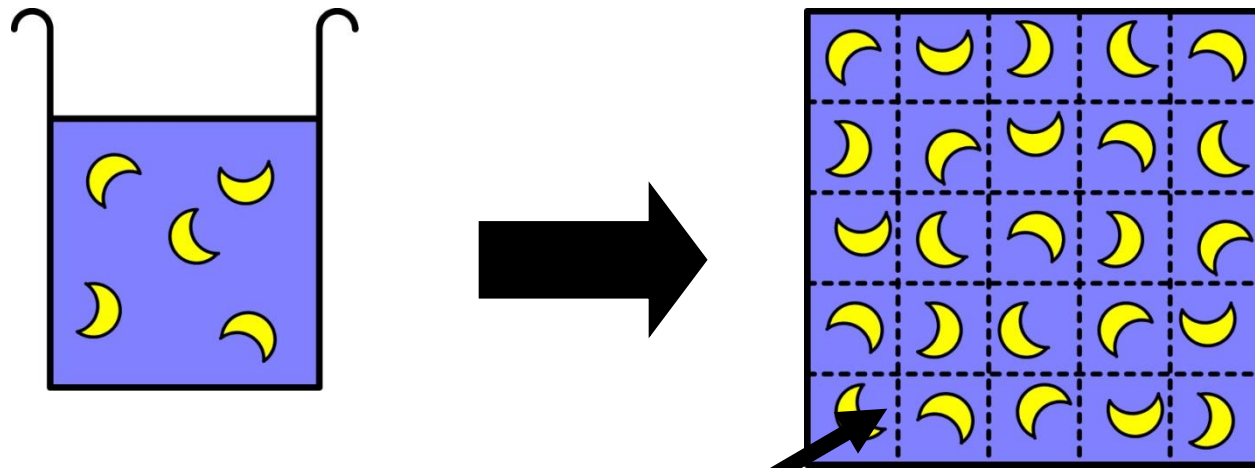


タンパク質1分子
を含む孤立系
(NVE 一定)



タンパク質を 10^{23} 個程度含む温
度・圧力(または体積)一定の系
(NVT 、 NPT 一定)

温度・体積一定のシステム(1)



熱だけが交換できる仕切り

- タンパク質を1分子と周辺の水分子を1つの系と考える
- 全系は、多数の同じ系から構成されている
- 各系は熱だけが交換できる仕切りを介して接触している
- 系の数、全エネルギーは一定

温度・体積一定のシステム(2)

- 実験データは、全系を構成する各系(タンパク質とその周辺環境)の状態によって決まる値の平均値(状態の出現確率での重みつき平均)

$$\langle A \rangle = A_i \rho_i, \quad \sum_i \rho_i = 1 \quad \rho_i: \text{状態 } i \text{ の出現確率}$$

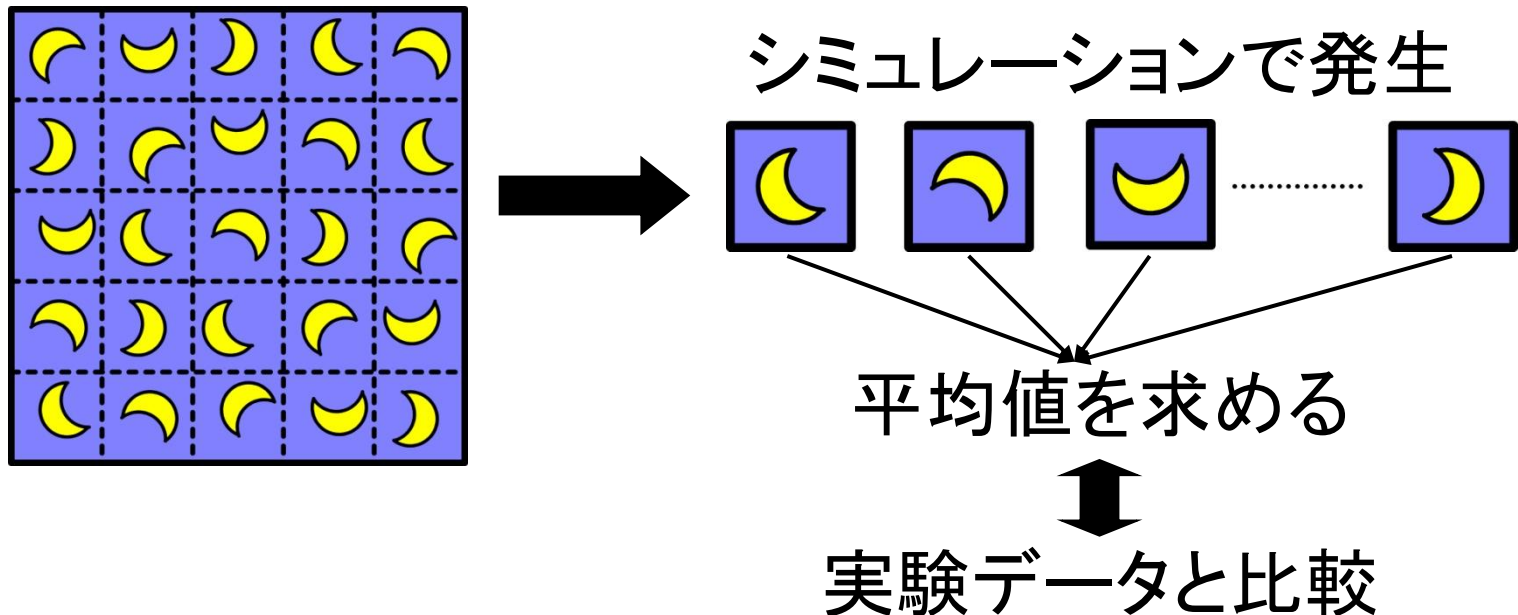
- 系の数、全エネルギーは一定の条件の下で、エントロピー(場合の数)が最大となる、各状態の出現確率 = **カノニカル分布**

$$\rho_i = Z^{-1} \exp(-e_i/k_B T) \quad e_i: \text{状態 } i \text{ のエネルギー}$$

$$Z = \sum_i \exp(-e_i/k_B T) \quad Z: \text{分配関数}$$

温度・体積一定のシステム(3)

- 分子シミュレーションでは、全系を構成する各状態を逐次的に発生させ、各状態での値を出現確率で加重平均することで、実験データに相当する値を求める



平均値計算の例

- ばねにつながった玉の系を例に全エネルギーの平均値を計算する(この場合は解析的に計算できる)

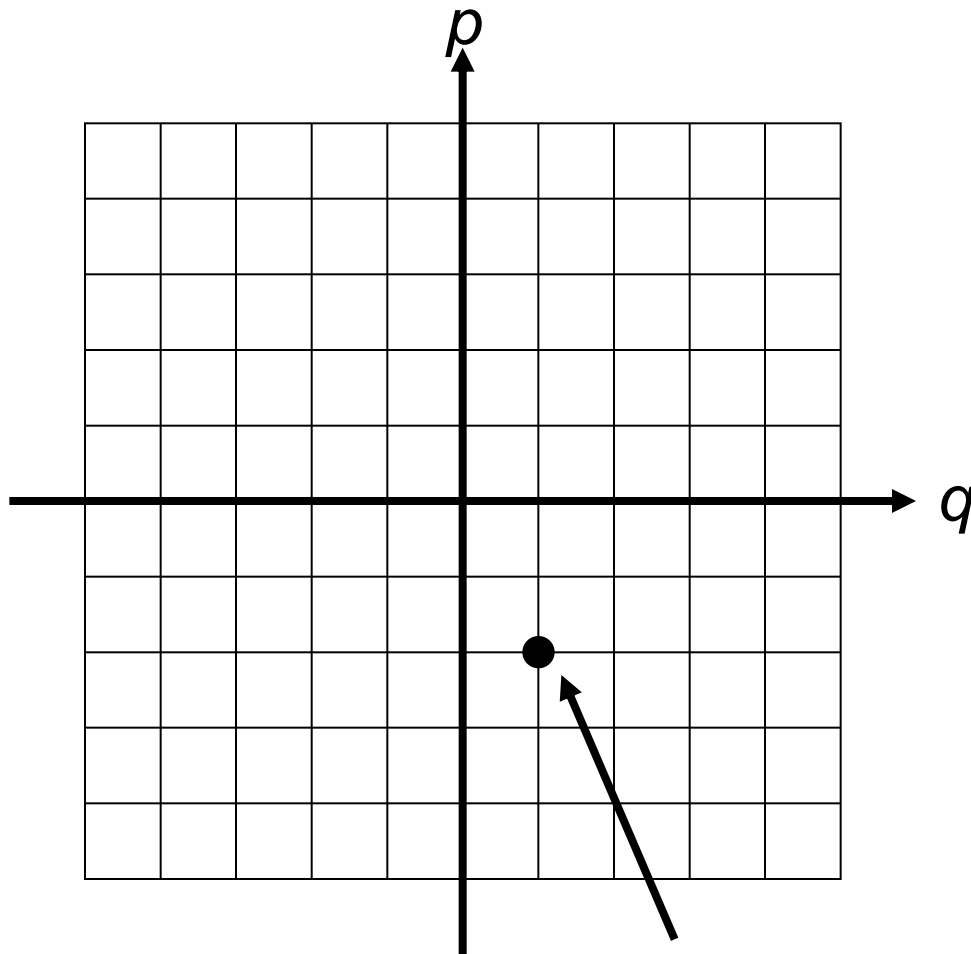
$$H = \frac{p^2}{2m} + \frac{k}{2}q^2, \quad \langle H \rangle = \frac{\int \left(\frac{p^2}{2m} + \frac{k}{2}q^2 \right) \exp \left[-\frac{1}{k_B T} \left(\frac{p^2}{2m} + \frac{k}{2}q^2 \right) \right] dq dp}{\int \exp \left[-\frac{1}{k_B T} \left(\frac{p^2}{2m} + \frac{k}{2}q^2 \right) \right] dq dp} = k_B T$$

- 方法1: グリッドに切る
- 方法2: モンテカルロ積分
- 方法3: importance sampling

確率 ρ



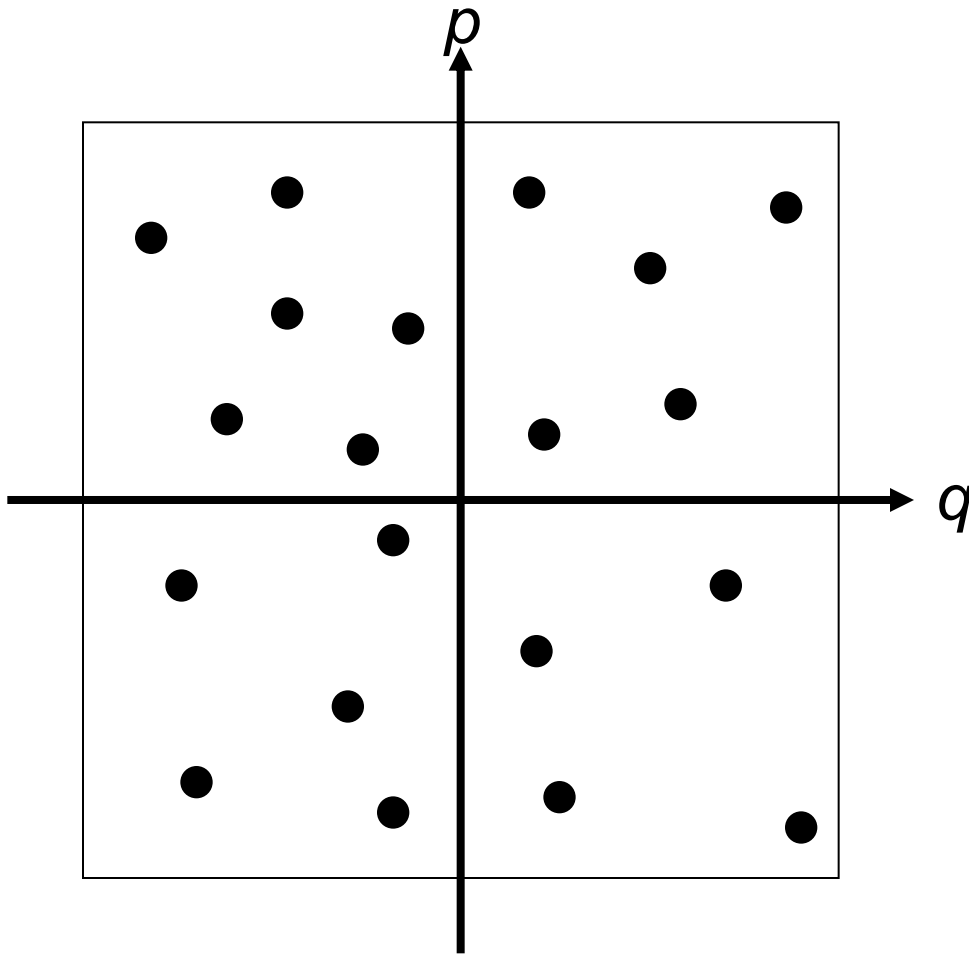
方法1: グリッドに切る



- q - p 平面を等間隔のグリッドに切り、各点で重み因子 $\exp(-H/k_B T)$ を計算し、この和と H との積の和を求める
- 本来は $(-\infty, \infty)$ の区間で計算する必要があるが、それは不可能なので q, p とも $[-10, 10]$ の区間で計算
- グリッドを細かくしていき、上記の和の比 $\langle H \rangle$ の収束を見る

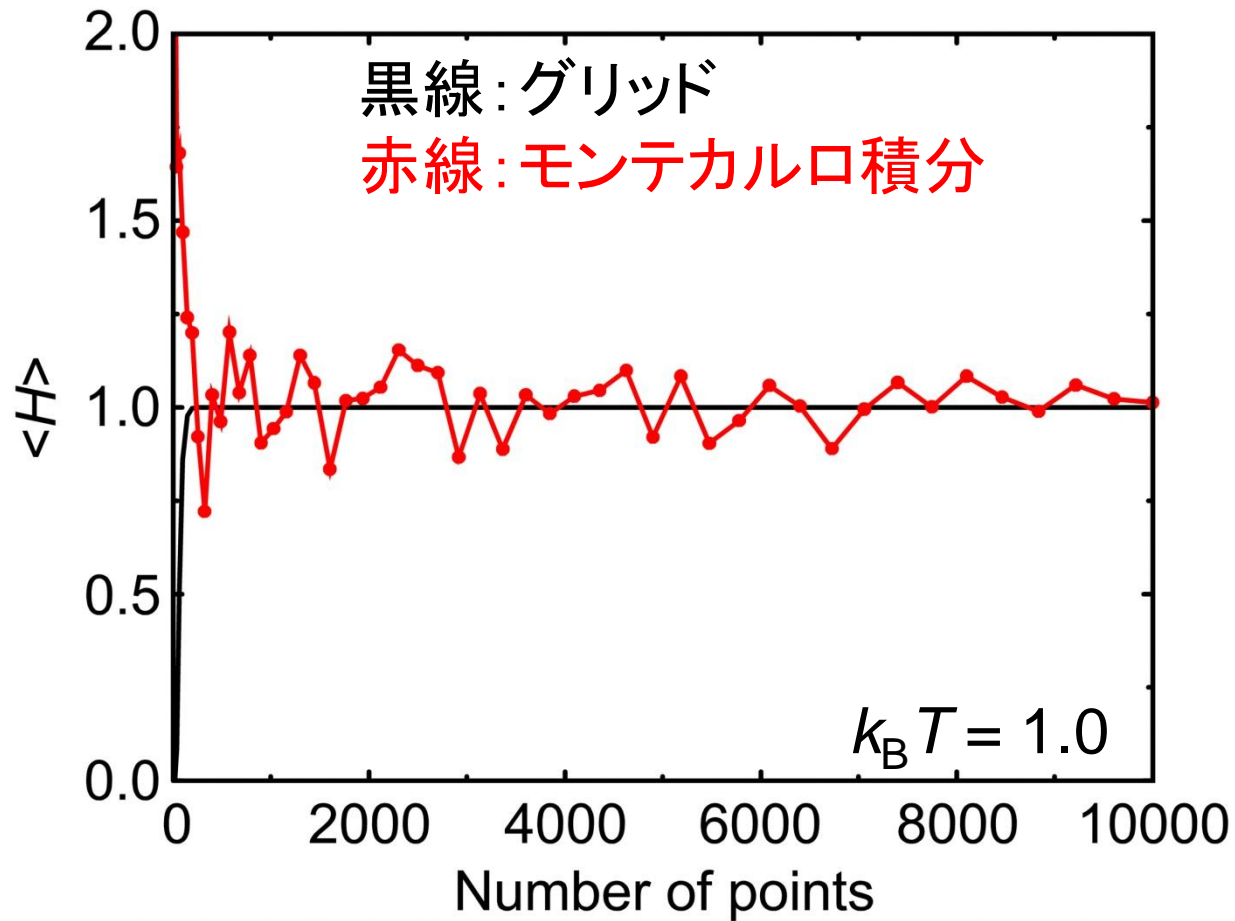
この点における $\exp(-H/k_B T)$ を計算

方法2:モンテカルロ積分



- q 、 p とも $[-10, 10]$ の区間で一様乱数を発生
- 各点で重み因子 $\exp(-H/k_B T)$ を計算し、この和と H との積の和を求める
- 発生させる点の数を増やしていき、上記の和の比($\langle H \rangle$)の収束を見る

結果の比較(1)

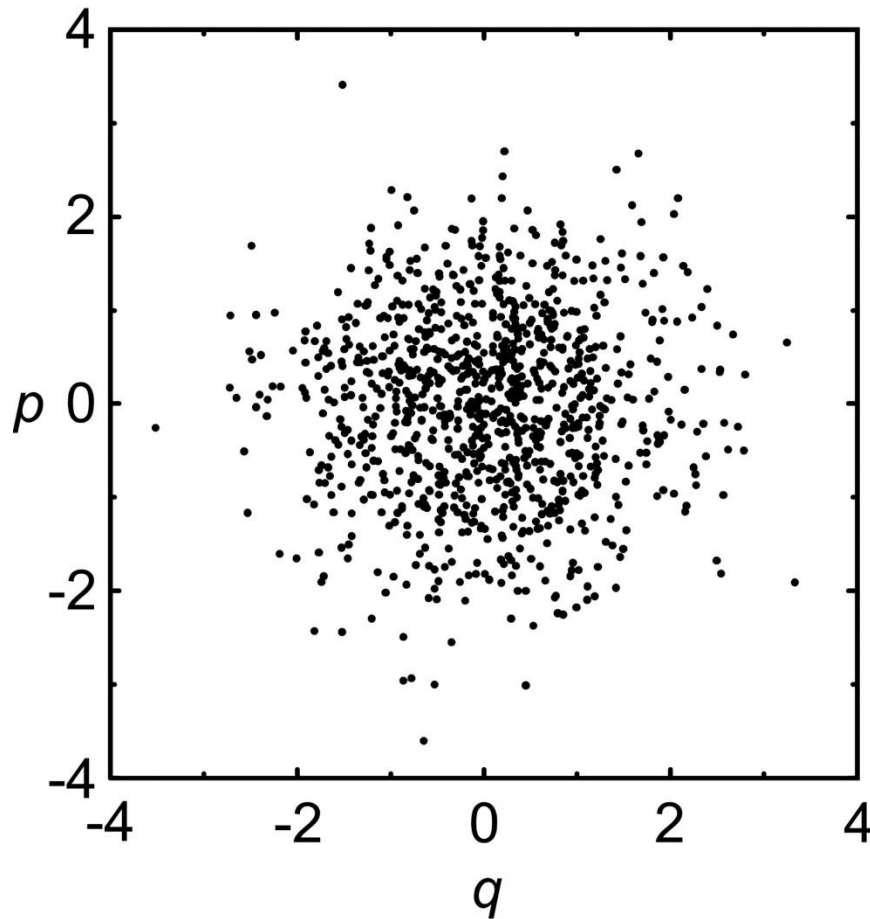


モンテカルロ積分はなかなか収束しない

Importance sampling

- モンテカルロ積分では、 H の値に重み $\exp(-H/k_B T)$ をかける
- $(q, p) = (0, 0)$ では、重みは1であるが、 $(q, p) = (10, 10)$ では、 3.7×10^{-44}
- 標本点ごとに平均への寄与が異なるため、計算に無駄がある
- 重みの大きい領域から集中的に標本点をとると無駄を減らせる→個々の標本点の寄与が等しくなるときが最も効率的
- 標本点の密度が重み $\exp(-H/k_B T)$ に比例するようにすれば良い→importance sampling

方法3: Importance sampling



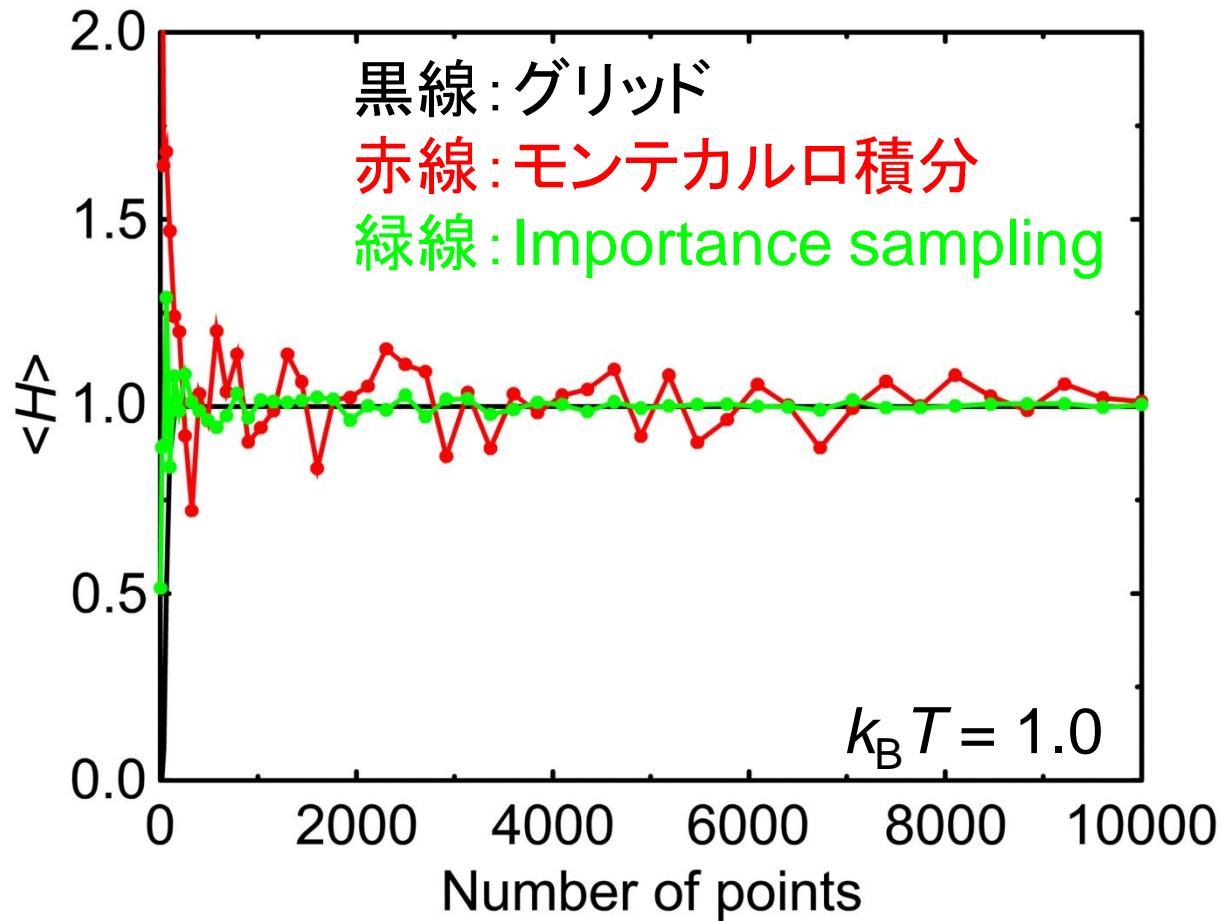
- 標本点の密度が重み $\exp(-H/k_B T)$ に比例するように標本点を発生(実際には分散1のGauss分布)
- 各点で H を計算し、この平均値を求める
- 発生させる点の数を増やしていき、 $\langle H \rangle$ の収束を見る

方法3: プログラム例

```
$kT=1.0;
$pi=atan2(1.0,1.0)*4.0;
$max_npt=100;

for($npt=2;$npt<=$max_npt;$npt+=2) {
    $val1=0.0;
    for($i=0;$i<$npt**2;++$i) {
        $x1=rand;
        $x2=rand;
        #一様乱数を正規分布に変換
        $q=sqrt(-2.0*$kT*log($x1))*cos(2.0*$pi*$x2);
        $p=sqrt(-2.0*$kT*log($x1))*sin(2.0*$pi*$x2);
        $H=0.5*$q**2+0.5*$p**2;
        $val1+=$H;           #標本点における全エネルギーを平均
    }
    printf("%d %f¥n", $npt**2, $val1/($npt**2));
}
```

結果の比較(2)



Importance samplingによって精度と効率が向上

Markov連鎖の利用

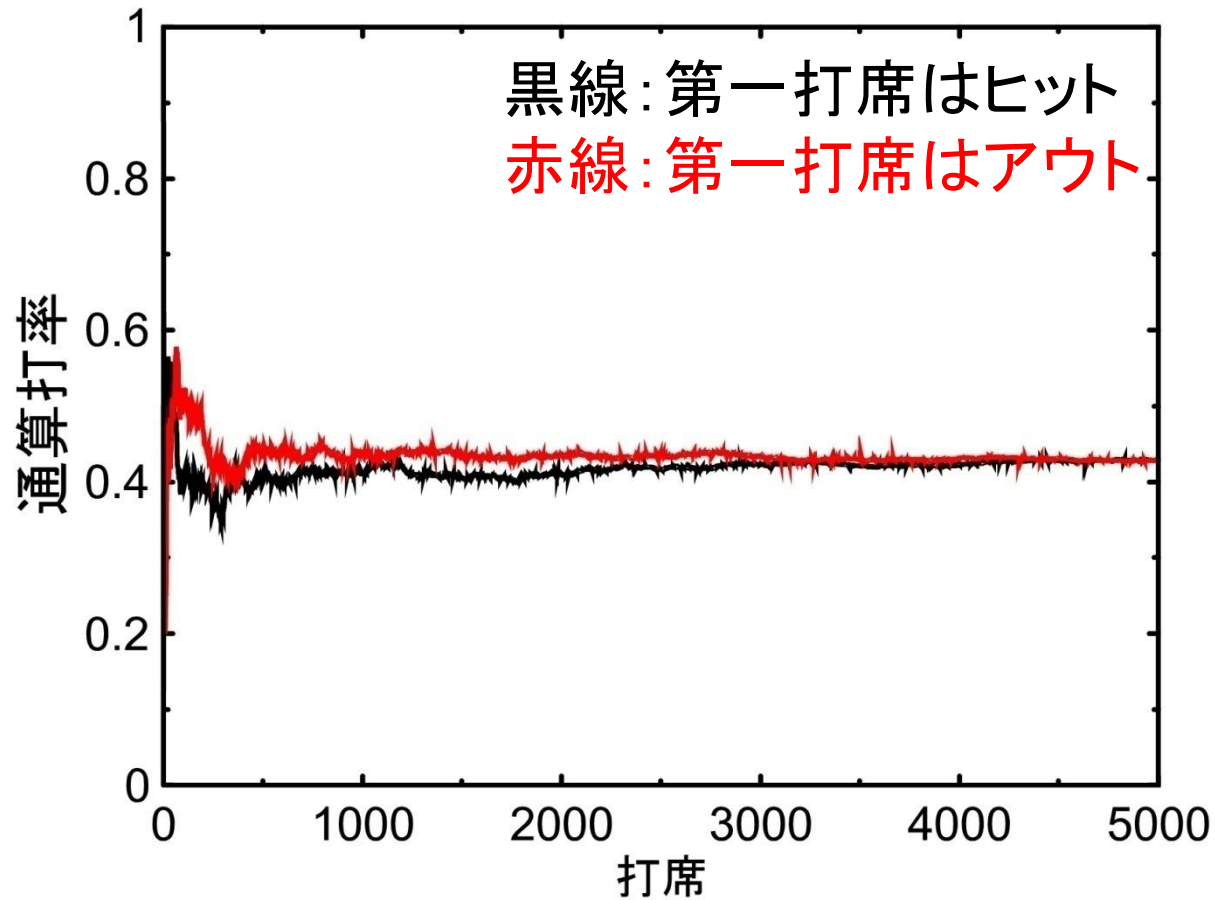
- 生体高分子の系の場合、構造とエネルギーの関係は複雑なので、 $\exp(-H/k_B T)$ に従う標本点(すなわち構造)を発生させる関数 f を求めることはできない
- このような場合はMarkov連鎖を利用する
- Markov連鎖とは？
 - 状態は離散的で、全部で M 個の状態がある
 - ある状態 i から別の状態 j に遷移する時、遷移確率 π_{ij} はそれ以前の状態に依存しない
- 遷移確率 π_{ij} を適切に設定することで、決まった出現確率に従う標本点を発生できる

Markov連鎖の例(1)

- 状態が2つで、遷移確率 π_{ij} をあらかじめ与えた場合の状態の出現頻度の挙動を考える
- ある野球選手がヒットを打つ確率を求める
 - ヒットを打った(状態1)次の打席は60%の確率でヒットを打つ
 - ヒットを打てなかった(状態2)次の打席は70%の確率でヒットを打てない
- 遷移確率(行列)は以下の通り

$$\pi = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} = \begin{pmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{pmatrix}$$

Markov連鎖の例(2)



初期状態(第一打席の成績)に関係なく一定の値に収束

Markov連鎖の例(3)

- 遷移確率 π_{ij} で決まる確率 ρ は以下で与えられる
→ π_{ij} を用いて確率 ρ に従う標本点が生成できる

$$\rho^{(1)} = (\rho_1^{(1)}, \rho_2^{(1)}) = (1, 0)$$

$$\rho^{(2)} = \rho^{(1)} \pi = (1, 0) \begin{pmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{pmatrix} = (0.6, 0.4)$$

$$\rho^{(3)} = \rho^{(2)} \pi = (0.6, 0.4) \begin{pmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{pmatrix} = (0.48, 0.52)$$

⋮

$$\rho = \lim_{N \rightarrow \infty} \rho^{(N)} = \lim_{N \rightarrow \infty} \rho^{(N-1)} \pi = \rho \pi, \quad \rho = (0.3/0.7, 0.4/0.7)$$

Metropolisの方法(1)

- 各状態の出現確率が収束した段階では方程式 $\rho = \rho\pi$ を満たす
- 出現確率 ρ は、遷移行列 π の固有ベクトル
→ 遷移行列が決まれば一意的に決まる
- Importance sampling では各状態の出現確率 ρ はわかっているので、ここから遷移行列 π を求めたい(以下で e_i は状態 i のエネルギー)

$$\rho_i = \frac{1}{Z} \exp\left(-\frac{e_i}{k_B T}\right), \quad Z = \sum_{i=1}^M \exp\left(-\frac{e_i}{k_B T}\right)$$

Metropolisの方法(2)

- 詳細釣り合いの式

$$\rho_i \pi_{ij} = \rho_j \pi_{ji}$$

- これが成り立てば $\rho = \rho \pi$ が満たされる

$$\sum_i \rho_i \pi_{ij} = \sum_i \rho_j \pi_{ji} = \rho_j \sum_i \pi_{ji} = \rho_j$$

- Metropolisらは詳細釣り合いを満たす遷移行列として以下を提案

$$\begin{cases} \pi_{ij} = \alpha_{ij} & \text{if } \rho_j \geq \rho_i \text{ and } i \neq j \\ \pi_{ij} = \alpha_{ij} (\rho_j / \rho_i) & \text{if } \rho_j < \rho_i \text{ and } i \neq j \end{cases}$$

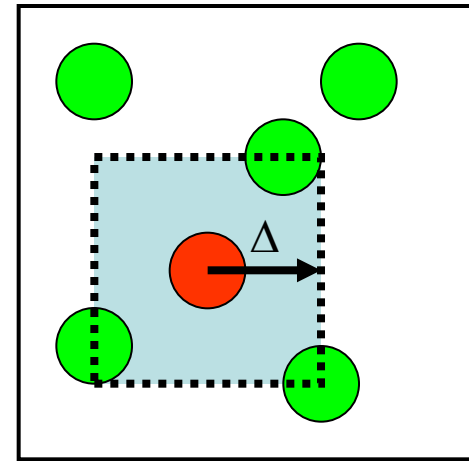
$$\pi_{ii} = 1 - \sum_{j \neq i} \pi_{ij}, \quad \alpha_{ij} = \alpha_{ji}, \quad \sum_{j=1}^M \alpha_{ij} = 1$$

Metropolisの方法(3)

- 注目している原子を、その原子を中心とした、1辺 2Δ の立方体の範囲内でランダムに遷移させる

$$\alpha_{ij} = \frac{1}{N_{\Delta}} \quad \text{遷移先が立方体の中}$$

$$\alpha_{ij} = 0 \quad \text{遷移先が立方体の外}$$



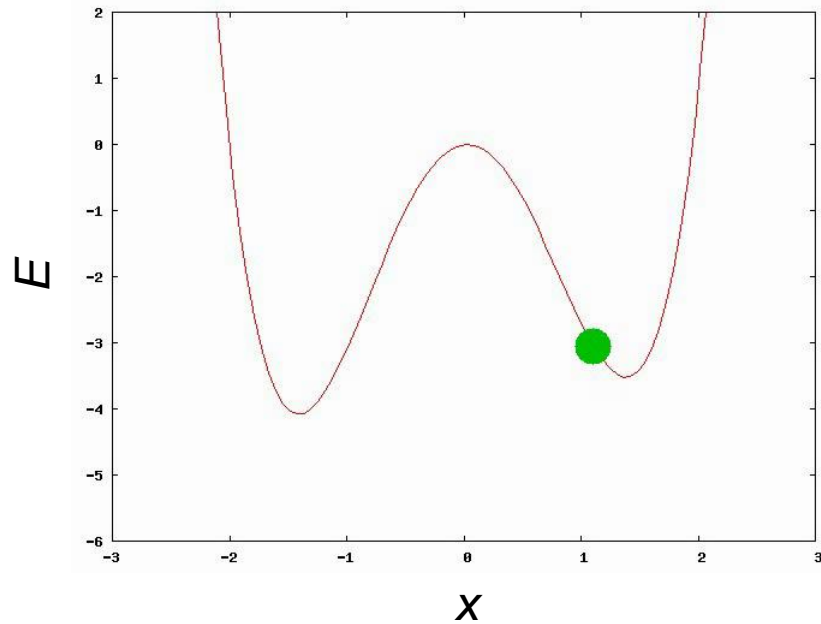
- 遷移先のエネルギー e_j が、もとの状態のエネルギー e_i よりも小さければ、その遷移を採択し、大きければ、以下の確率で採択する

$$\rho_j / \rho_i = \exp[-(e_j - e_i) / k_B T] = \exp(-\Delta e_{ji} / k_B T)$$

- 採択されなかった場合は、もとの状態を新しい状態とする

計算例

- 2つの極小値を持つ関数について、確率密度分布を求める
- $X = 1$ からスタートし、 10^7 ステップ実行
- 変位の最大値 $\Delta = 0.1$ 、温度 $kT = 3.0$ とする
- 得られた確率密度分布を理論式 $\exp(-E(x)/kT)$ と比較する



$$E(x) = \left[(x+1)^2 - 1 \right] \left[(x-1)^2 - 0.9 \right]$$

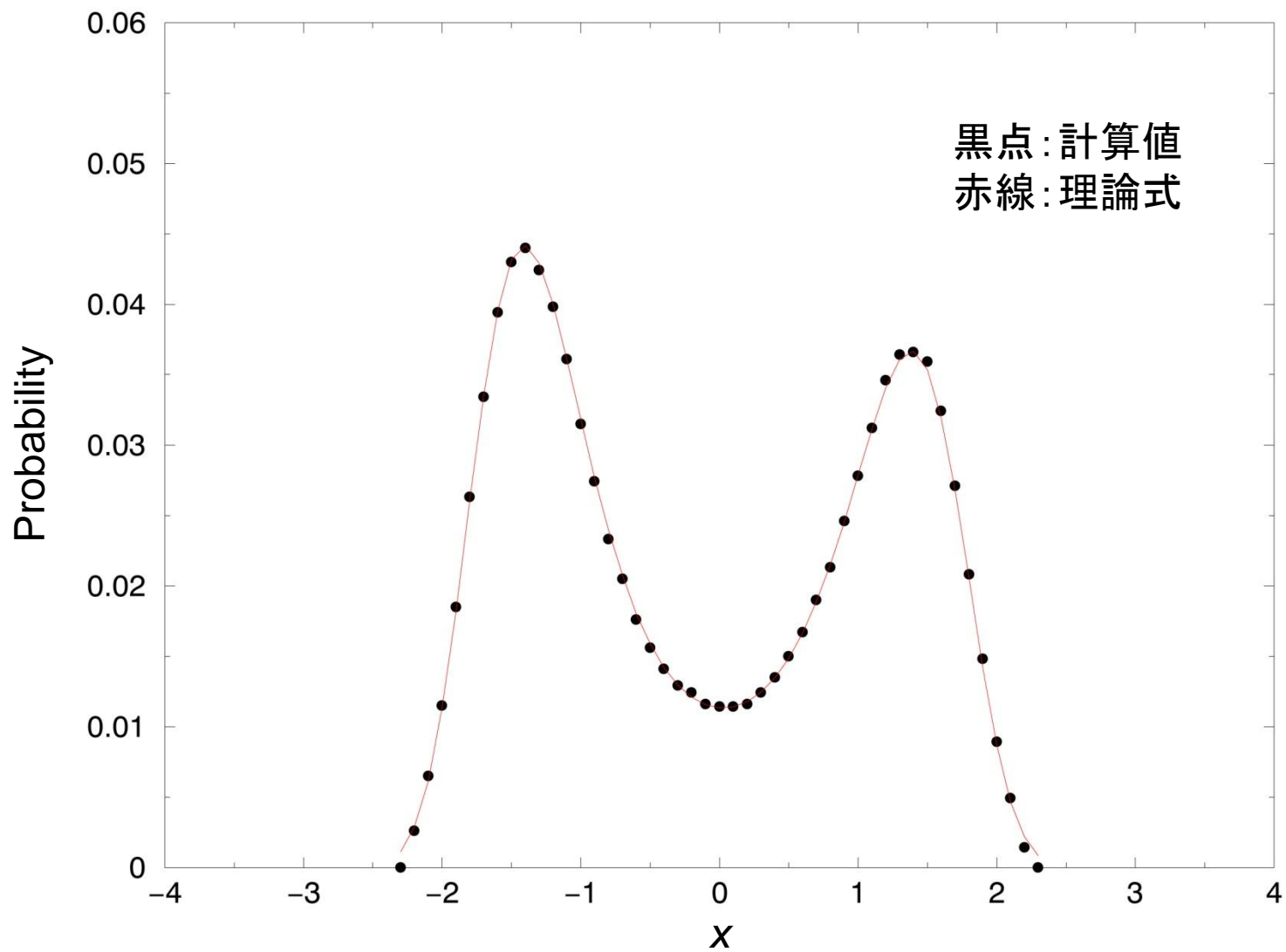
プログラム例

```
$nstep=1000000; #ステップ数
$x=1.0; #初期位置
$ene=&calc_ene($x); #初期エネルギー
$delta=0.1; #変位の最大値
$kT=3.0;

for($i=0;$i<$nstep;$i++) {
    $x_new=$x+2.0*$delta*(rand()-0.5); #試行
    $ene_new=&calc_ene($x_new);
    $p=exp(($ene-$ene_new)/$kT);
    if($p >= 1.0 || $p >= rand()) { #試行が採択される条件
        $x=$x_new;
        $ene=$ene_new;
    }
    printf("%d %f %f¥n", $i+1, $x, $ene);
}

sub calc_ene { #エネルギー関数
    my ($x)=@_;
    return (($x+1.0)*($x+1.0)-1.0)*((($x-1.0)*($x-1.0)-0.9));
}
```


結果



生体高分子への適用(1)

- 生体高分子では、「状態」は立体構造に対応
- Metropolisの方法は生体高分子から原子を任意に選び、その位置を動かすことで実現可能
- ただし、原子を動かすと大抵の場合共有結合長を変えることになり、エネルギーが増加するが多い→棄却される確率が高い
- このため、共有結合長や共有結合角を固定し、二面角のみを動かすのが一般的だが困難も多い
 - 複数分子の扱いが難しい
 - タンパク質のコアのように密にパックしている部分では、二面角を回転すると原子が衝突する可能性が高い

生体高分子への適用(2)

- 生体高分子の系では、canonical分布に従う立体構造分布を生成する目的で定温分子動力学法を用いる
- 平均値を計算する際には時間平均として計算する
- 実際にcanonical分布を生成するためには、そのために特に工夫された方法を用いる必要がある
 - 能勢-Hoover法、能勢-Hoover chain法
 - 束縛法
 - Langevin dynamics法
- Berendsenのcoupling法ではcanonical分布は生成されないことに注意

分子シミュレーションの応用事例

- RecA•DNA複合体のシミュレーション
- Simulated annealing
- フォールディングシミュレーション

RecA•DNA複合体(1)

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 284, NO. 44, pp. 30230–30239, October 30, 2009
© 2009 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in the U.S.A.

A Non-canonical DNA Structure Enables Homologous Recombination in Various Genetic Systems^{*[S]}

Received for publication, July 13, 2009, and in revised form, August 28, 2009 Published, JBC Papers in Press, September 3, 2009, DOI 10.1074/jbc.M109.043810

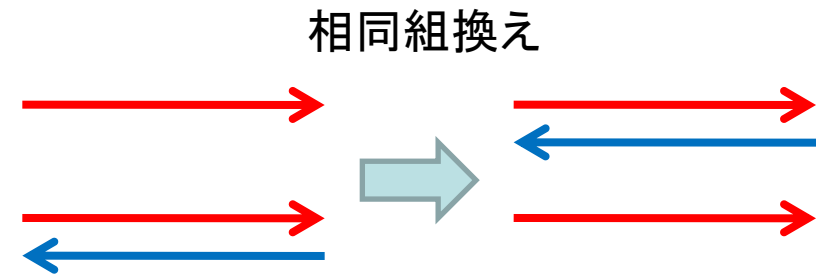
Tokiha Masuda^{‡S}, Yutaka Ito[¶], Tohru Terada^{||}, Takehiko Shibata^{‡S}, and Tsutomu Mikawa^{‡S#¶1}**

*From the [‡]Graduate School of Nanobioscience, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, the ^SRIKEN Advanced Science Institute and the ^{**}RIKEN Computational Science Research Program, 2-1 Hirosawa, Wako, Saitama 351-0198, the [¶]Department of Chemistry, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji, Tokyo 192-0397, the ^{||}Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, and the ^{#¶}RIKEN SPring-8 Center, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan*

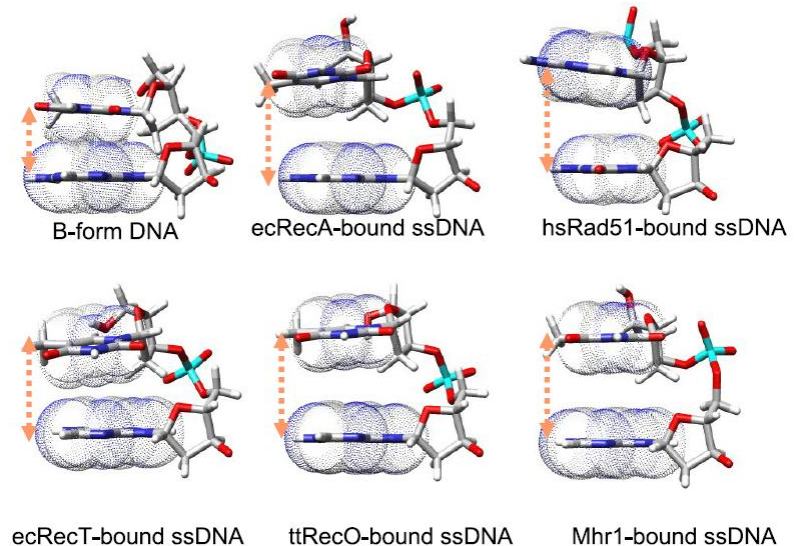
Masuda *et al.* *J. Biol. Chem.* **284**, 30230 (2009).

RecA・DNA複合体(2)

- RecAは大腸菌で、DNAの相同組換えを担うタンパク質
- 1本鎖DNAはRecAに結合すると、伸張した特異な立体構造をとる
- 本論文では、他の生物に存在する、別の相同組換えタンパク質に結合したDNAも同様な立体構造をとることを示した
- タンパク質同士に進化的な類縁関係はないことから、この構造をとることが、相同組換えに重要であると考えられる



組換えタンパク質に結合したDNAの構造



RecA・DNA複合体(3)

- では、この構造をとるとなぜ相同組換えに有利なのか？



- RecA・DNA複合体の結晶構造を基に、分子動力学シミュレーションを用いてDNAの運動性を解析した



PDB ID: 3CMW

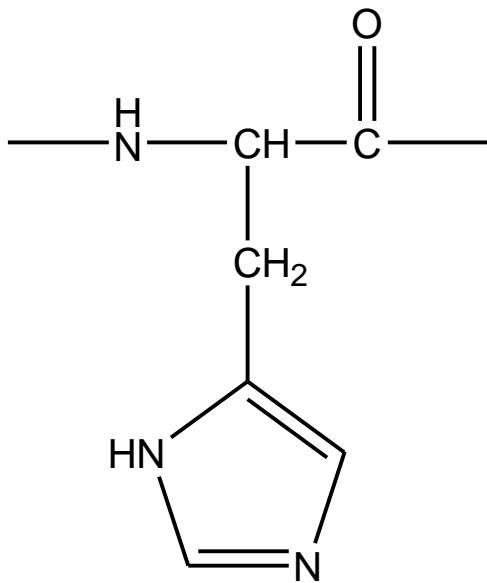
シミュレーションの手順

1. 初期構造の作成
 - 立体構造の取得
 - 欠失残基への対応
 - 水素原子付加
 - リガンドのモデリング
 - 力場パラメータの取得
 - 水分子の配置
2. 立体構造最適化
3. 初期速度の割り当て
4. 平衡化
 - 座標の束縛
 - 水分子配置の最適化
5. プロダクションラン

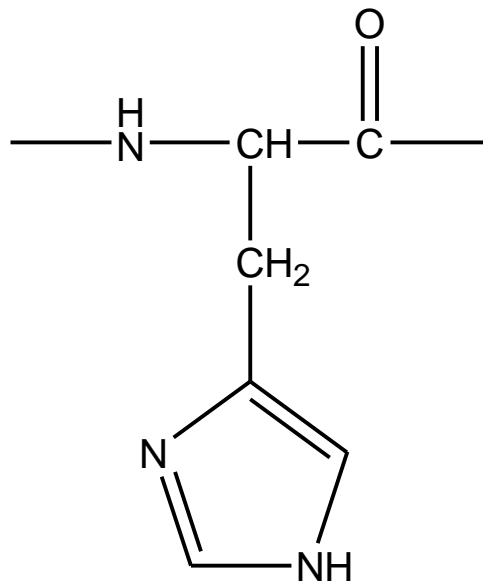
初期構造の作成(1)

- 立体構造の取得
 - PDBのサイト (<http://www.rcsb.org/pdb/>) からダウンロード
 - 非対称単位に2つの複合体が含まれているが、一方だけで機能していることは明らか→一方の複合体のみを選択
- 欠失残基への対応
 - 結晶構造に含まれる欠失残基はモデリングなどで補う必要がある
 - ここでは、欠失残基は人為的に付加されたリンカ配列であるので、欠失残基の前後の残基をacetyl基、N-methyl基でブロック
- 水素原子付加
 - 基本的に自動的に付加できる
 - SS結合の有無、Hisのプロトン化状態に注意

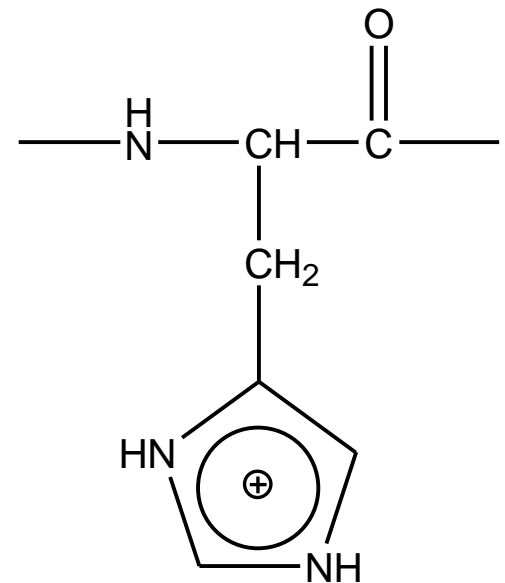
Hisのプロトン化状態



δ 位にプロトン化



ϵ 位にプロトン化



δ, ϵ 位にプロトン化

- His側鎖の pK_a は中性付近であるため2つの窒素原子とも水素原子が結合した状態も十分にとりうる
- 基本的には、His周りの水素結合ネットワークからプロトン化状態を決定する

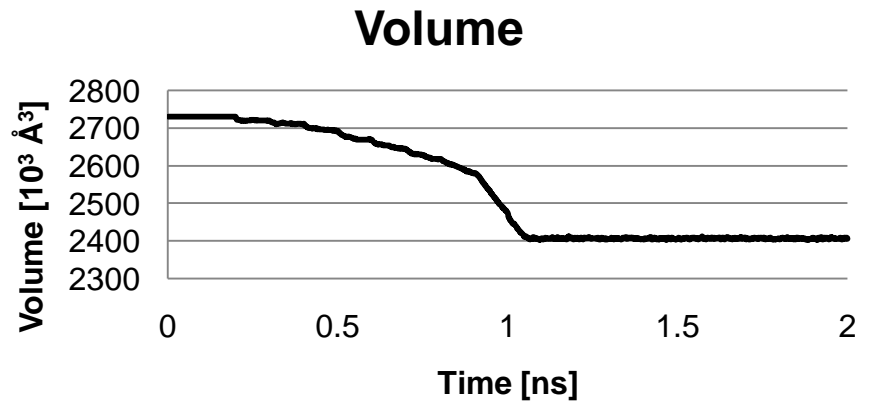
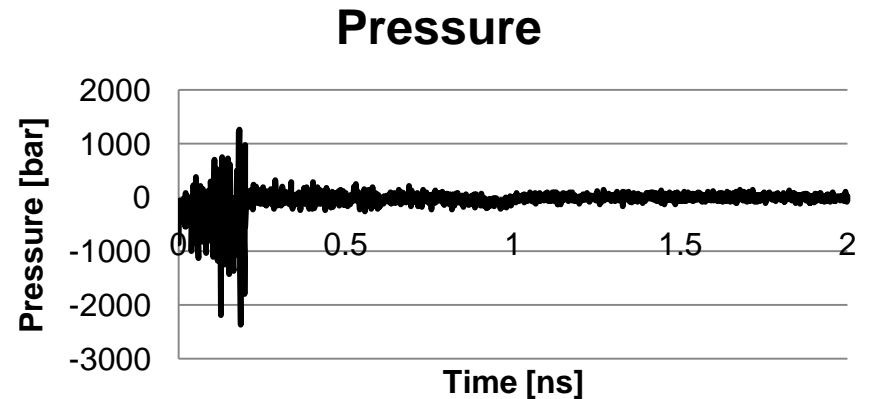
初期構造の作成(2)

- リガンドのモデリング
 - RecAにはATPが結合するが、ここでは反応中間体アナログADP·AlF₃が結合している
 - アナログを本来のATPに戻すモデリングを行う
- 力場パラメータの取得
 - リガンドの力場パラメータは分子動力学ソフトウェアに含まれていないので、自分で作成するか、Amber Parameter Database*等から取得する
- 水分子の配置
 - PMEを利用して高精度かつ高速にシミュレーションを行うため水分子を直方体状に配置する
 - 電荷を中性にするためにカウンターイオンを配置する

*<http://www.pharmacy.manchester.ac.uk/bryce/amber>

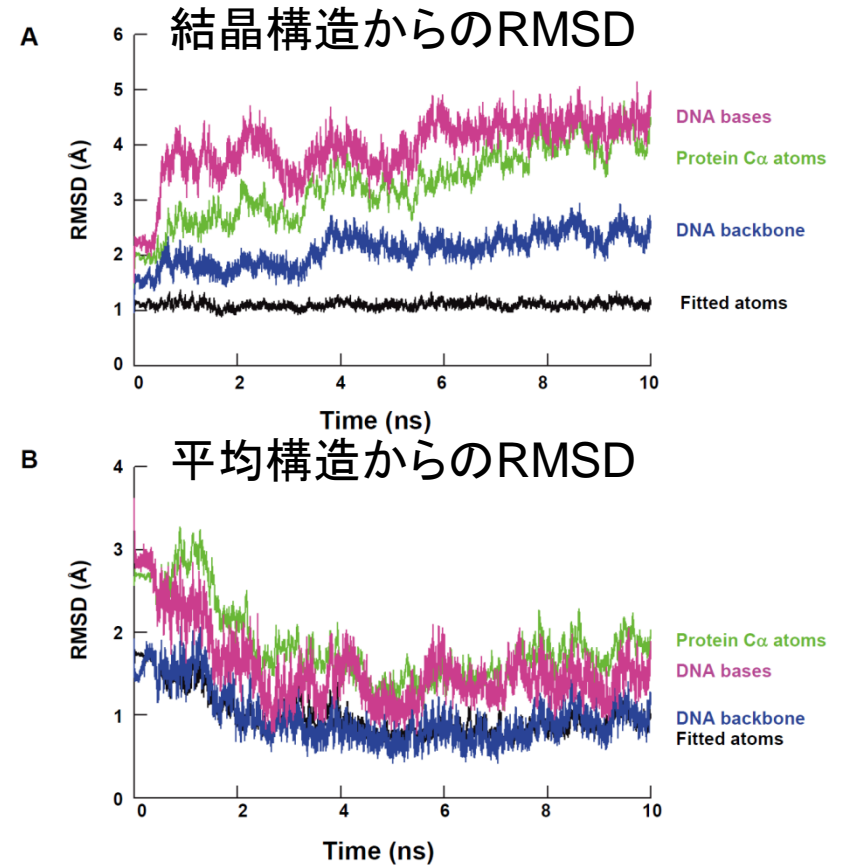
平衡化

- 初期構造では、配置した水分子とタンパク質の間に隙間がある
- 定温定圧シミュレーションを行い、水分子の配置を最適化する
- その際、タンパク質の原子が初期位置からあまり動かないように束縛する



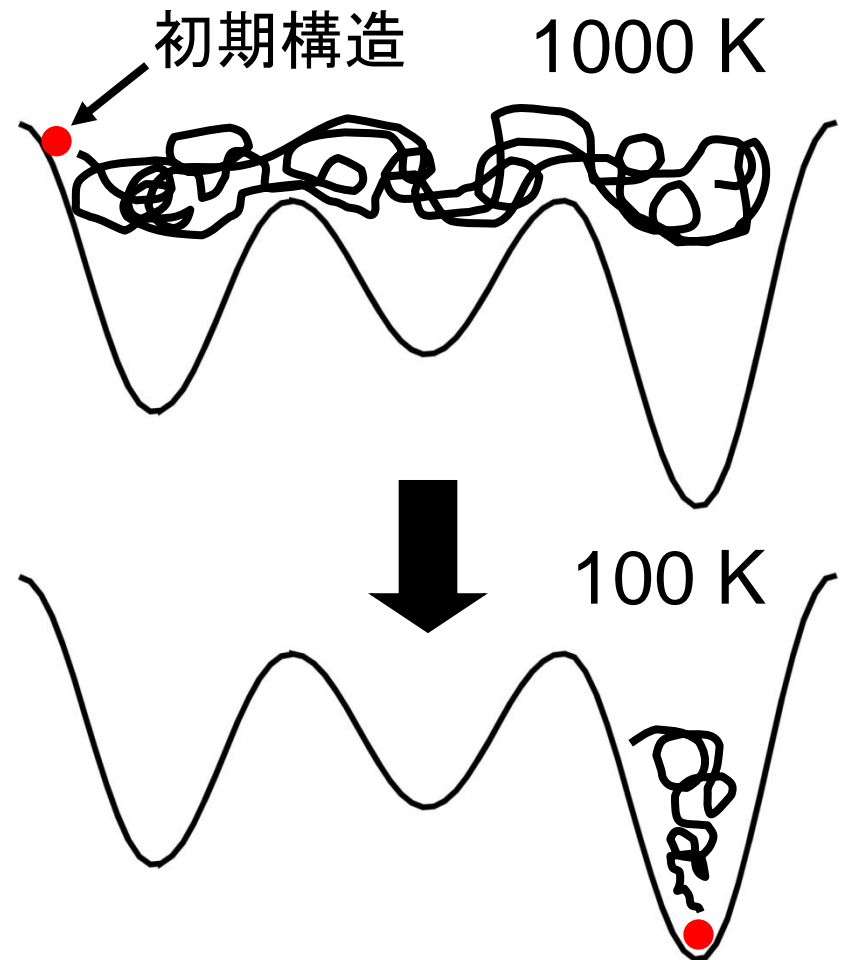
結果の解析

- DNAのbackboneに対してbaseのRMSDは大きく、baseの構造が大きく揺らいでいることが明らかとなった
- これは、RecAに結合したDNAが伸長した構造をとることにより、base間の相互作用が弱くなることによる
- この運動性が相同性探索に有利に働く可能性がある



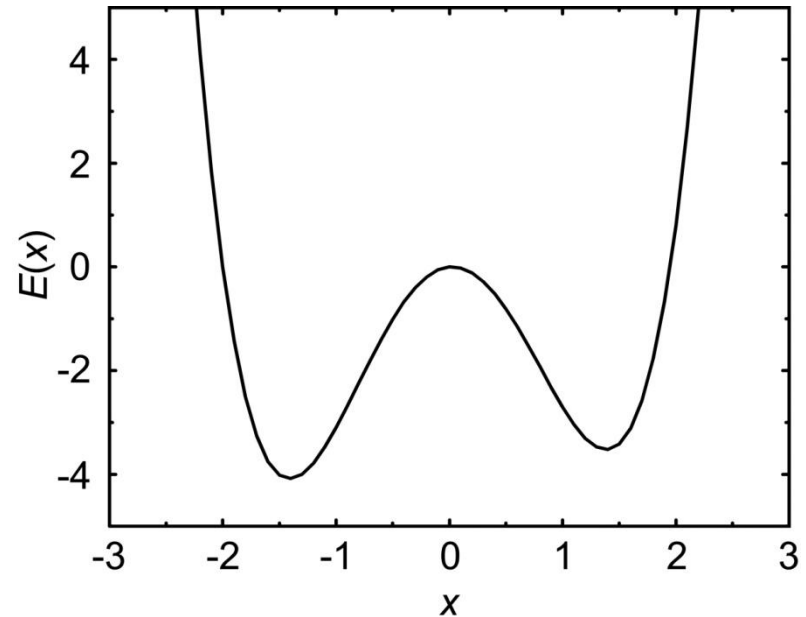
Simulated annealing

- 生体高分子のエネルギー関数はエネルギー極小状態と最小状態が高いエネルギー障壁で隔てられていることがしばしばある
- 300 K程度の定温分子動力学シミュレーションでは、エネルギー障壁を越えられない
- 高温(1000 K程度)から低温まで徐々に温度を下げていくことで、エネルギー最小状態に到達する確率を上げることができる



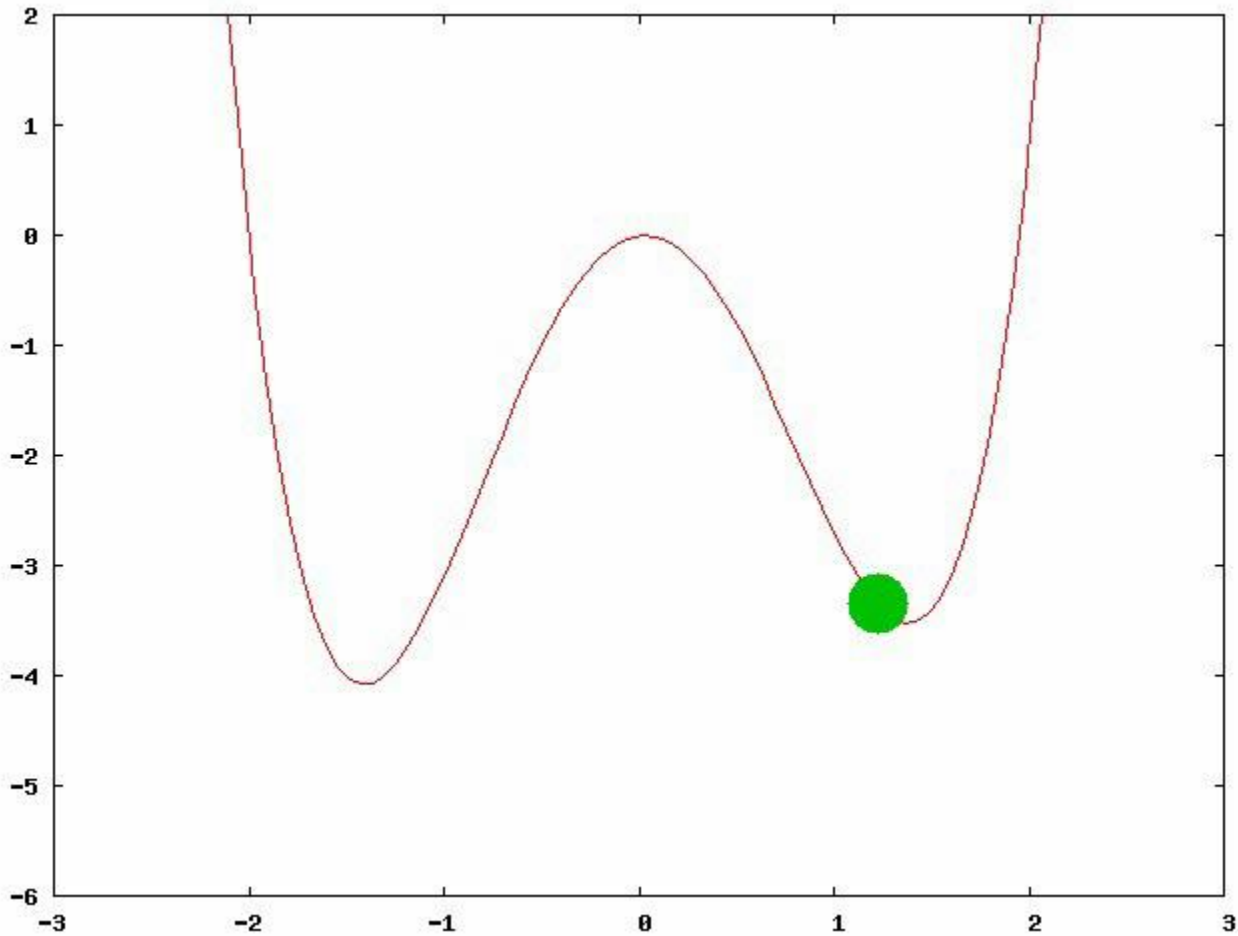
計算例

- 2つのエネルギー極小状態を持つエネルギー関数
- $kT=0.3$ の定温シミュレーションと $kT=3$ から $kT=0.3$ に徐々に下げる simulated annealing を実施
- いずれも $x=1$ から開始し、エネルギー最小状態に到達できるか比較

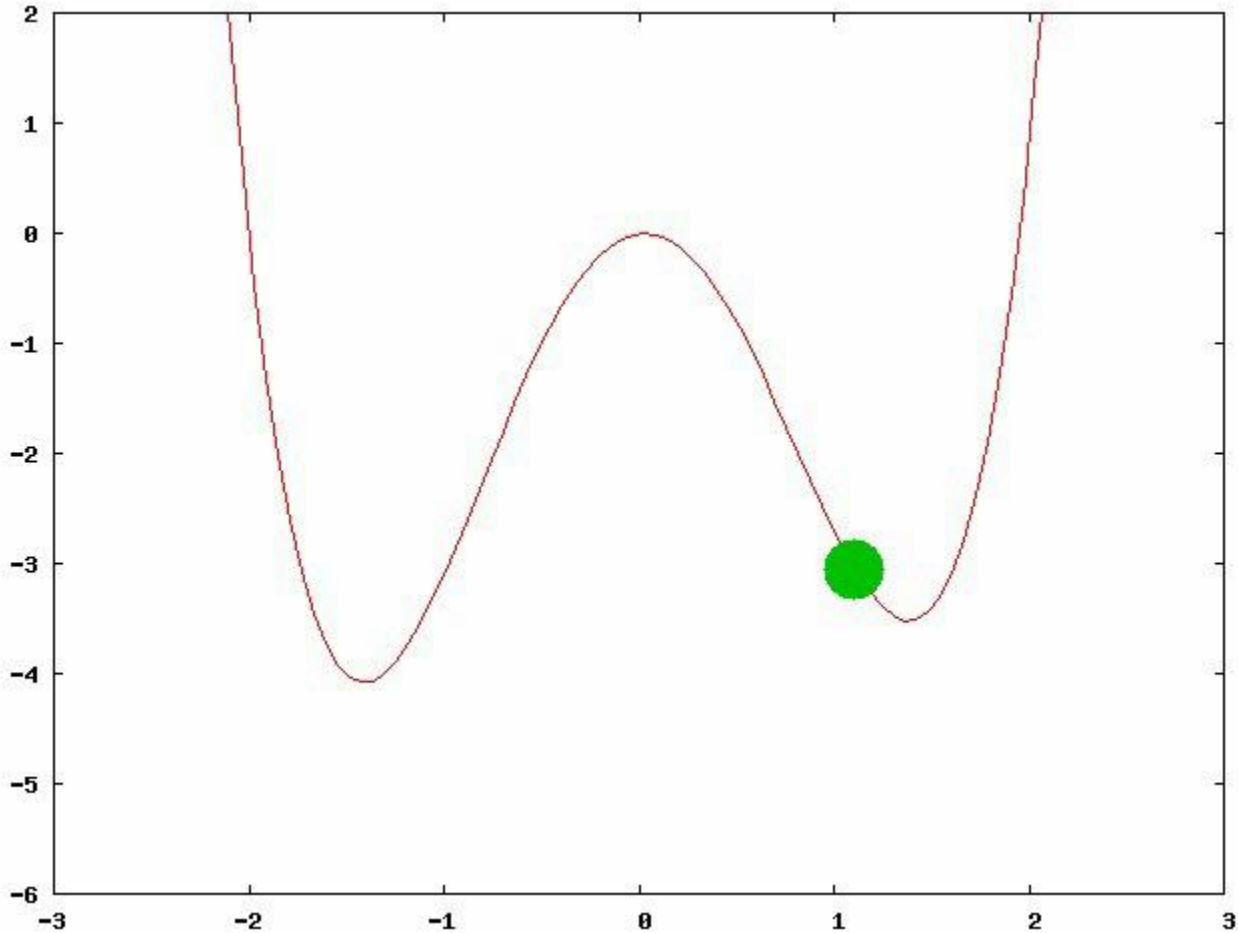


$$E(x) = \left[(x+1)^2 - 1 \right] \left[(x-1)^2 - 0.9 \right]$$

定温シミュレーション ($kT = 0.3$)

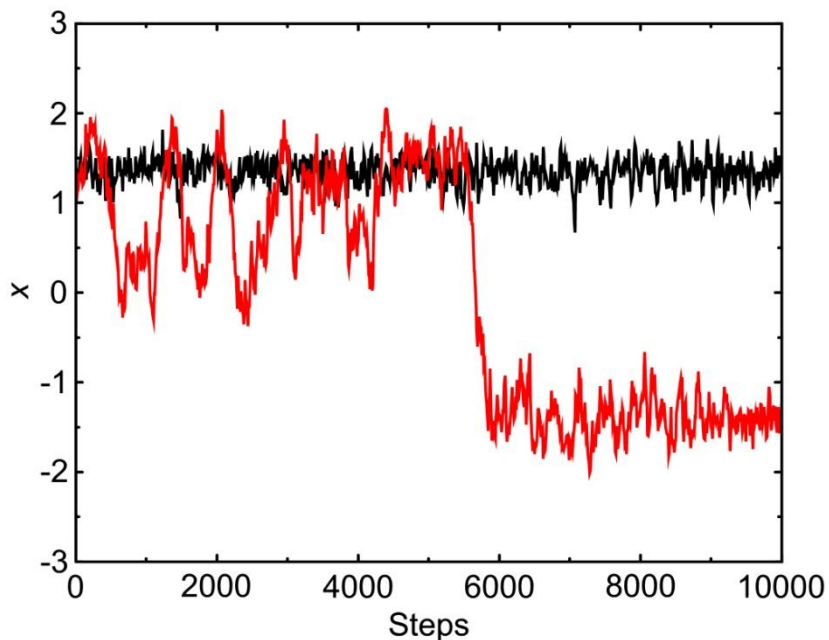


Simulated annealing ($kT = 3 \rightarrow 0.3$)

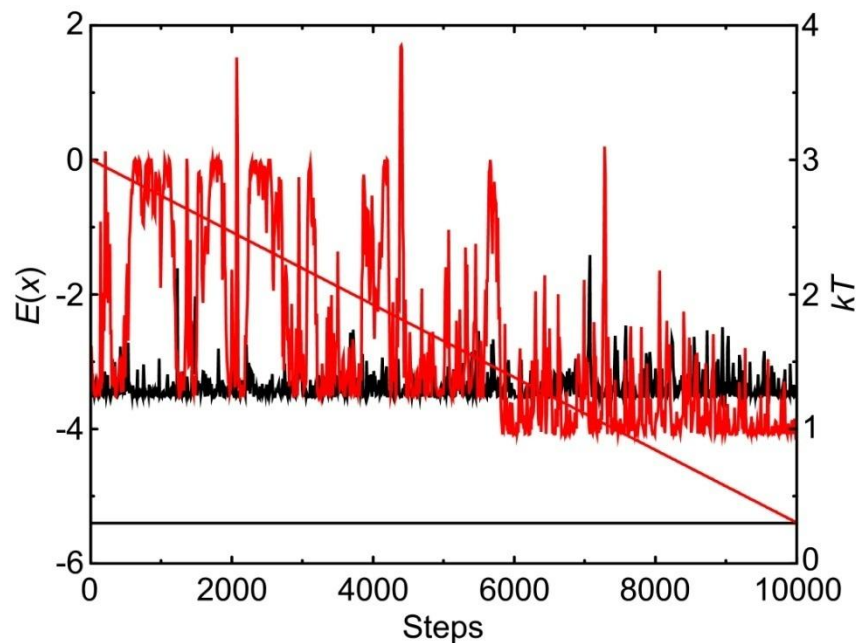


計算結果

座標の時間変化



エネルギー・温度の時間変化



黒: 定温シミュレーション、赤: Simulated annealing

温度は右目盛り

Simulated annealingを用いることによってエネルギー
最小状態に到達することが可能になっている

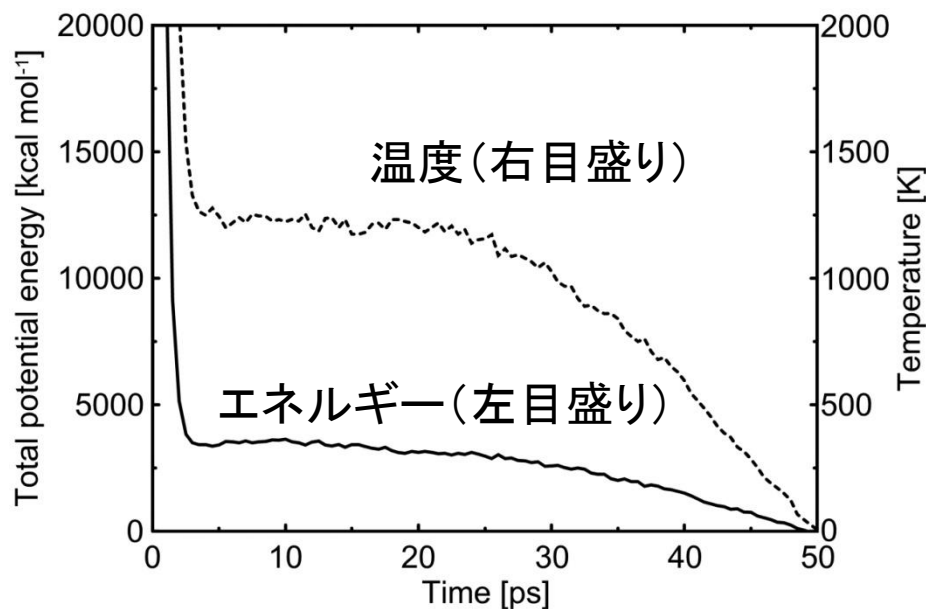
NMR構造計算(1)

- NMRでは水素原子核間の距離が測定できる
(r^6 に比例するNOEシグナルが観測される)
- 測定で得られた水素原子核間距離を満たす立体構造を計算によって求める
→水素原子核間距離が実験値と近くなると値が小さくなるような関数を、ポテンシャルエネルギー関数に加えて分子動力学シミュレーションを行う

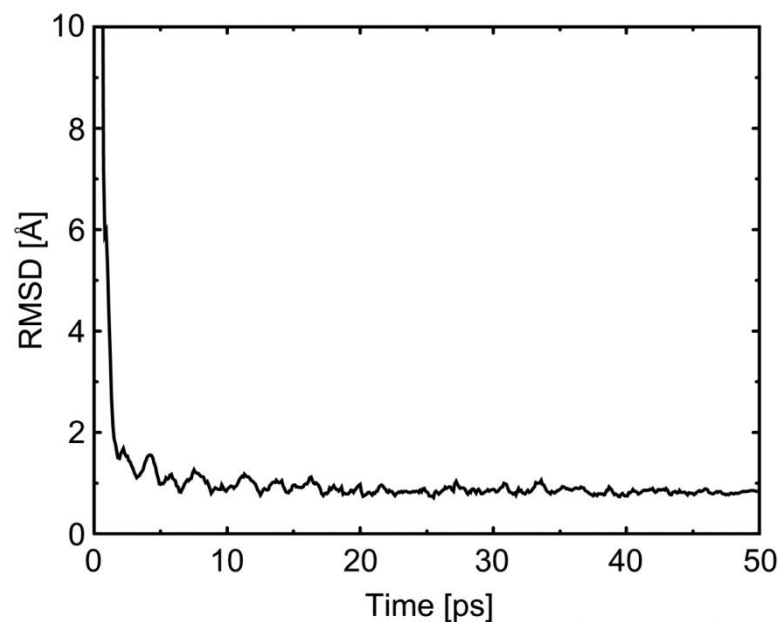
$$E_{\text{total}}(\mathbf{r}) = E(\mathbf{r}) + \sum_i k_i (r_i - r_i^{\text{exp}})^2$$

NMR構造計算(2)

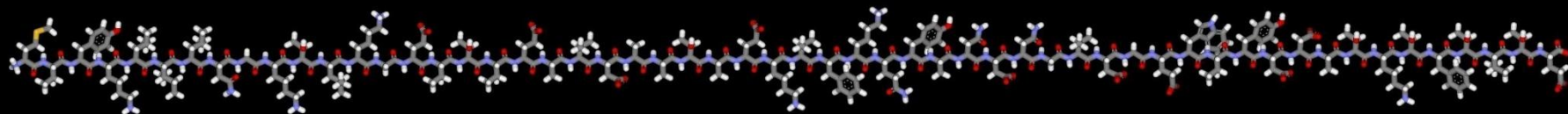
ポテンシャルエネルギーと温度



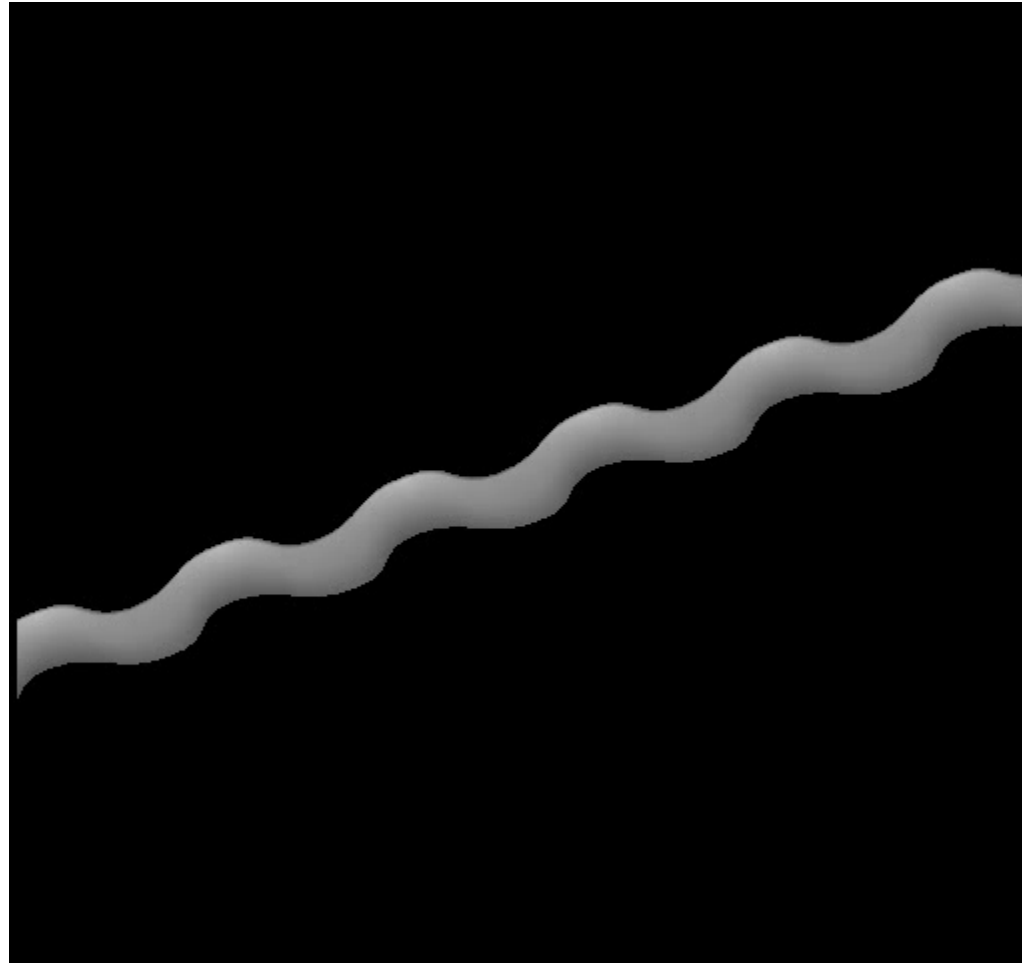
C α RMSD



初期構造は完全に伸展した構造



NMR構造計算(3)



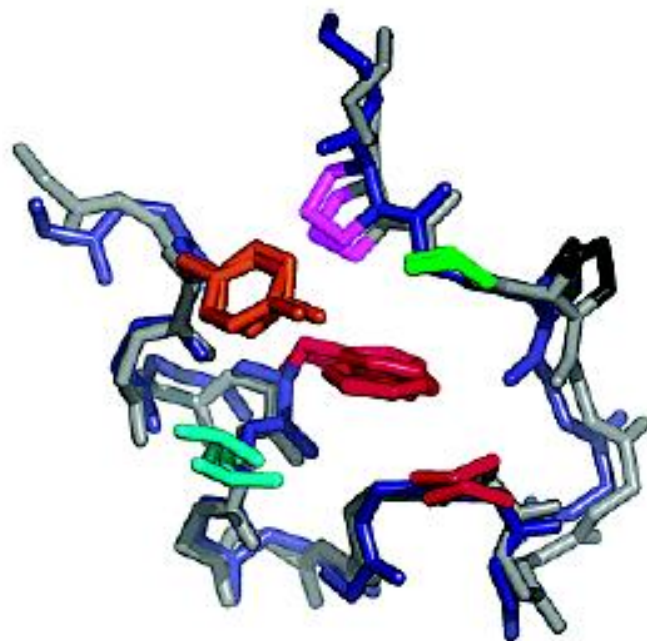
Protein G (PDB ID: 1GB1)

フォールディングシミュレーション

- 熱力学仮説
 - タンパク質の天然構造は自由エネルギー最小構造である
 - 天然構造は原子間相互作用の総和で決まる
→ アミノ酸配列で決まる
- Levinthal paradox
 - 各残基が3つのコンフォメーションをとりうるだとすると、100残基では $3^{100} \simeq 10^{48}$ 種類
 - 1 psごとに別のコンフォメーションに遷移すると考えると、すべて探索するのに 10^{28} 年かかる
 - 実際のタンパク質は秒のオーダーで天然構造に折り畳む
- 分子シミュレーションによるフォールディング問題の解決が期待されている

Trp-cage

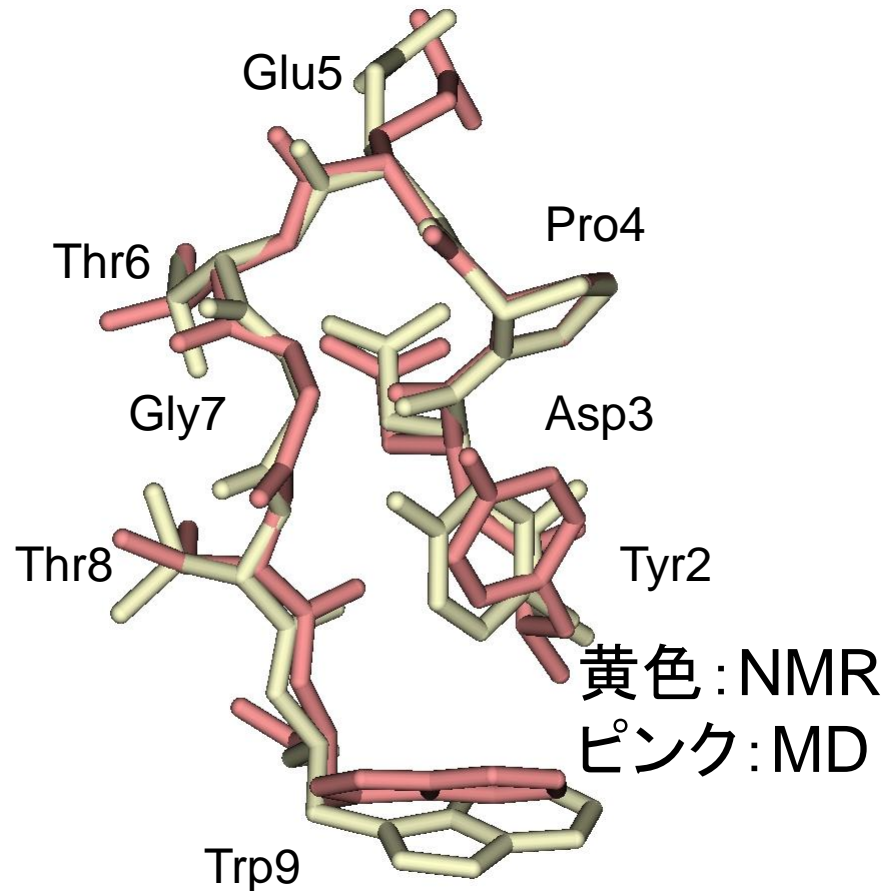
- 折り畳み構造をとるように人工的にデザインされた20残基の小ペプチドTrp-cage
- 配列
NLYIQWLKDGGPSSGRPPPS
- 溶媒和自由エネルギーを一般化Bornモデルで近似した325 Kにおける分子動力学シミュレーションによって伸展構造から天然構造に折り畳むことが示された



灰色：NMR構造、青色：計算

Chignolin

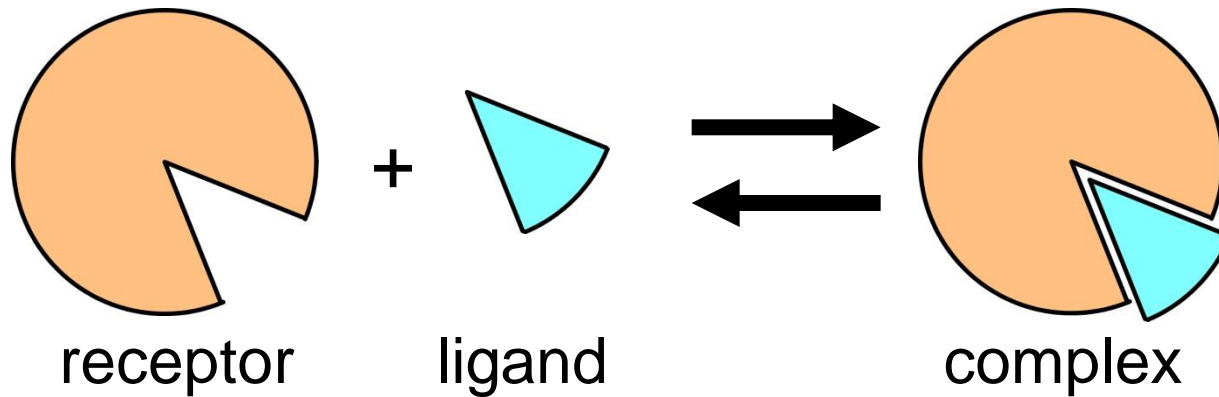
- 産総研の本田らによって設計された10残基のペプチド (GYDPETGTWG)
- 水溶液中で安定な β ヘアピンを形成し、協同的に熱転移を起こす「世界最小のタンパク質」
- マルチカノニカル分子動力学シミュレーションによりフォールディング自由エネルギー地形を計算
- 自由エネルギー最小構造が天然構造に一致



複合体モデリング

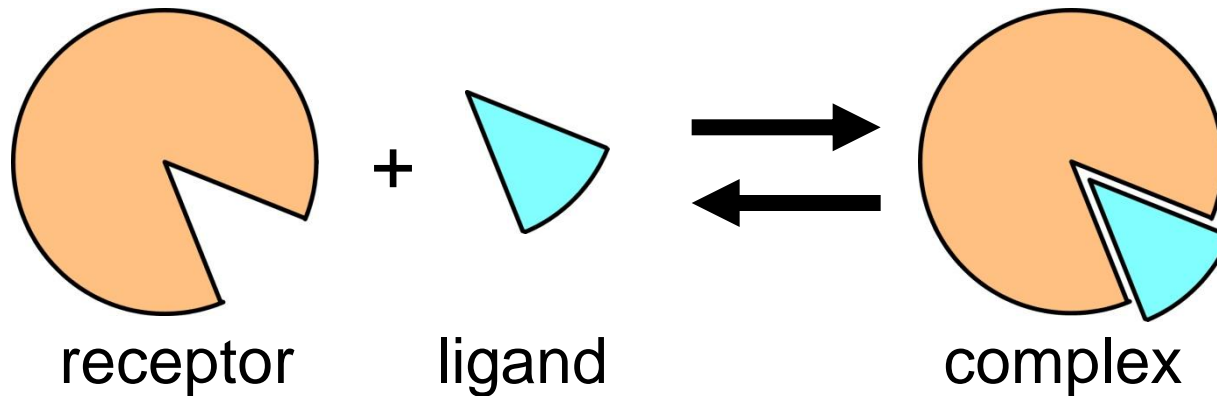
- タンパク質とタンパク質を含む他の分子との複合体の立体構造を予測する
- 類似した複合体の立体構造が利用できる場合
 - ホモロジーモデリング
 - 立体構造の重ね合わせ
- 類似した複合体の立体構造が利用できる場合
 - ドッキングシミュレーション

ドッキングシミュレーション



- タンパク質 (receptor) の表面にあるligand結合サイトにligandを結合させてみる
- Ligandが、タンパク質か低分子化合物かで異なる方法が用いられる

結合自由エネルギー



$$\begin{aligned}\Delta G_{\text{bind}} &= G_{\text{complex}} - (G_{\text{receptor}} + G_{\text{ligand}}) \\ &= G_{\text{complex}}^{\circ} - (G_{\text{receptor}}^{\circ} + G_{\text{ligand}}^{\circ}) + RT \ln \frac{[\text{complex}]}{[\text{receptor}][\text{ligand}]} \\ &= \Delta G_{\text{bind}}^{\circ} - RT \ln K_D = 0\end{aligned}$$

$K_D = \exp(\Delta G_{\text{bind}}^{\circ}/RT)$ 結合自由エネルギーは解離定数と関係づけられる

結合自由エネルギーの成分

- 自由エネルギーはポテンシャルエネルギー項、圧力項、エントロピー項からなる
 - タンパク質-リガンド間相互作用 ΔE_{int} は負→安定化
 - タンパク質およびリガンドの脱水和 ΔE_{desolv} は正→不安定化
 - 構造固定によるエントロピー損失 ΔS_{conf} は負→不安定化
 - 水和水の解放によるエントロピー利得 ΔS_{wat} は正→安定化

$$G = E + PV - TS$$

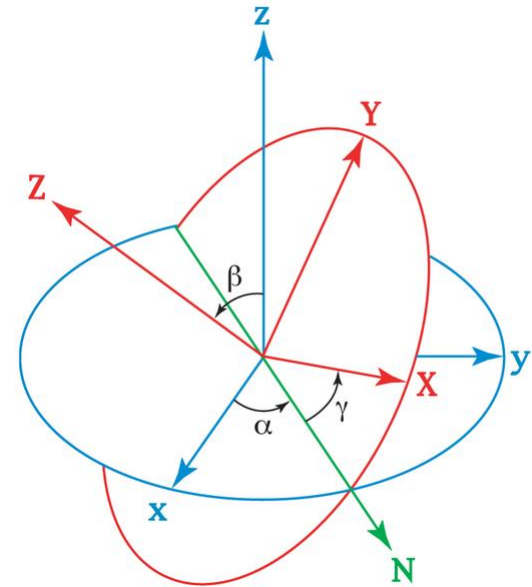
$$\Delta G_{\text{bind}}^{\circ} \approx \Delta E - T\Delta S = \Delta E_{\text{int}} + \Delta E_{\text{desolv}} - T(\Delta S_{\text{conf}} + \Delta S_{\text{wat}})$$

結合自由エネルギーの計算

- エネルギー計算
 - ポテンシャルエネルギー値をそのまま使う
 - 溶媒効果や構造エントロピーの効果を見逃している
- MM-PB/SA法
 - ポテンシャルエネルギー値に、Poisson-Boltzmann方程式と溶媒接触表面積から得た溶媒和自由エネルギーと振動解析から求める構造エントロピーを加える
- 自由エネルギー摂動法、熱力学的積分法
 - 基準となる化合物に置換基を導入したときと自由エネルギー変化を計算する
 - 精度は高いが、構造が異なる化合物を比較できない
- スコア関数の利用

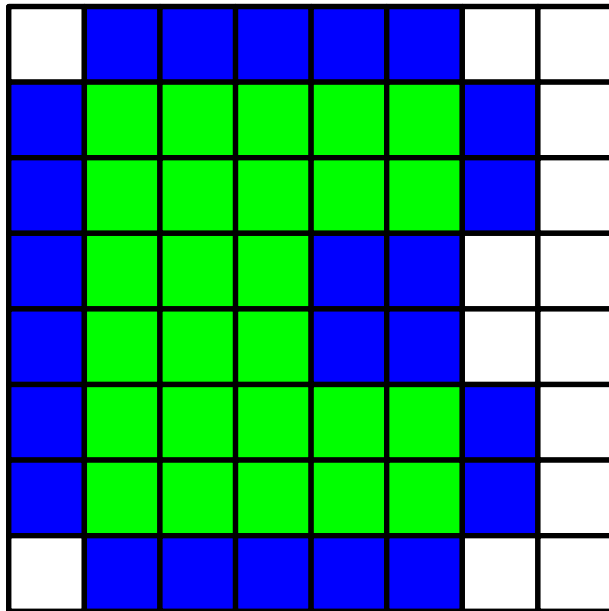
タンパク質・タンパク質ドッキング

- Receptor、ligandともに剛体とみなし、複合体形成による立体構造変化は考慮しない
- Receptorは原点に固定し、ligandの並進3自由度、回転3自由度の計6自由度のみを考慮
 - 回転はEuler angleで記述
- 形の相補性が特に重要

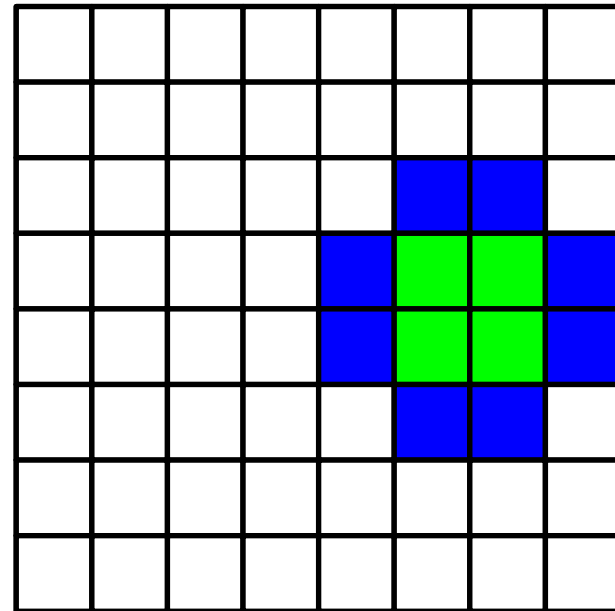




形の相補性計算(1)

Receptor

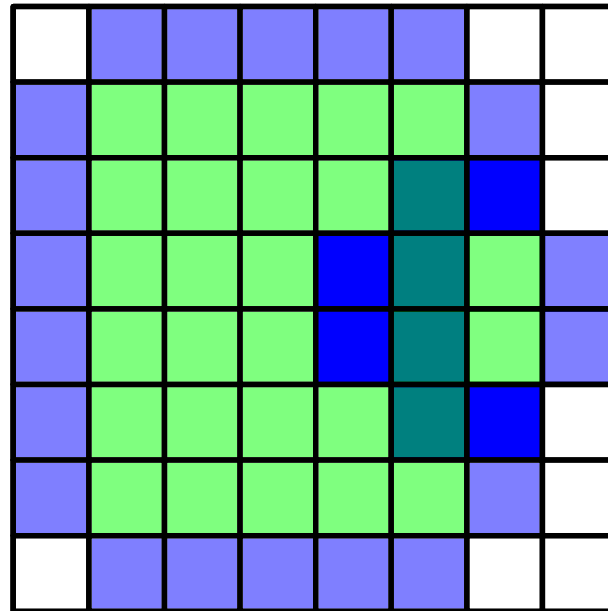


Ligand



-  = 1 (solvent accessible surface layer)
-  = $9i$ (solvent excluding surface layer)

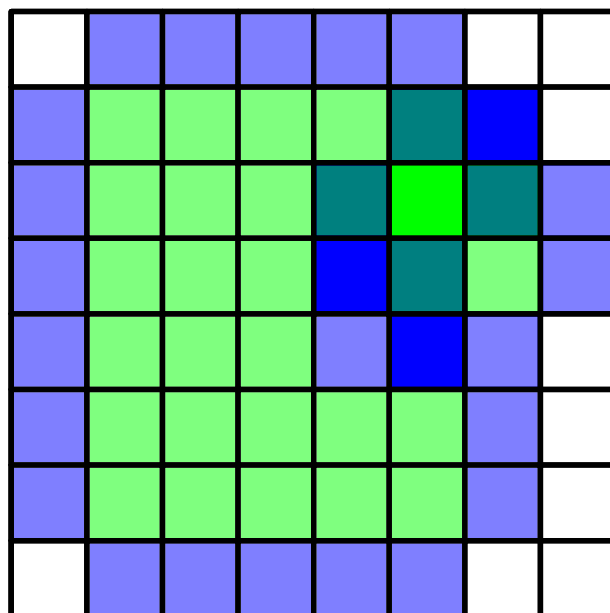
形の相補性計算(2)




重ね合わせてグリッドごとにスコアの積を計算する

スコア積の和の実部 = ドッキングスコア = 4

形の相補性計算(3)



 = -81

重ね合わせてグリッドごとにスコアの積を計算する

スコア積の和の実部 = ドッキングスコア = $3 - 81 = -78$

計算の高速化

- 計算の一般化

$$S(a,b,c) = \sum_{x,y,z} f(x,y,z)g(x+a,y+b,z+c)$$

スコア S を最大にする ligand の並進位置 (a, b, c) を求める

- この計算は fast Fourier transform (FFT) を用いて高速化できる

$$\tilde{S}(h,k,l) = \tilde{f}(h,k,l)\tilde{g}(h,k,l)$$

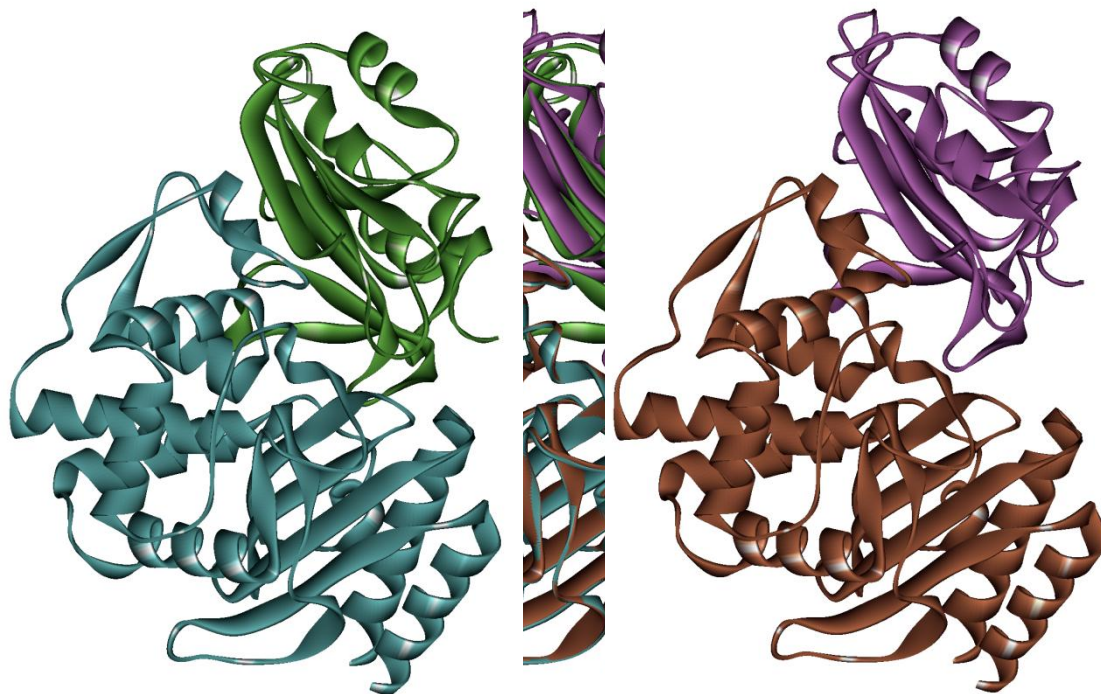
- これを ligand のいろいろな向きについて計算する
- 静電相互作用など、他の相互作用も同様に高速に計算できる

ソフトウェアの例

- **FTDock**
<http://www.bmm.icnet.uk/docking/ftdock.html>
- **ZDock**
<http://zlab.bu.edu/zdock/index.shtml>
- **HEX**
<http://www.loria.fr/~ritchied/hex/>
- **DOT**
<http://www.sdsc.edu/CCMS/DOT/>
- **GRAMM-X**
<http://vakser.bioinformatics.ku.edu/resources/gramm/grammx>

ZDockを用いた計算例

- TEM-1 β -lactamaseとinhibitorの複合体
 - β -lactamase: 1ZG4 (receptor)
 - Inhibitor: 3GMU (ligand)



スコア1位のドッキング構造

正解構造 (1JTG)

タンパク質・低分子化合物ドッキング

- タンパク質 (receptor) の表面にあるリガンド結合をあらかじめ探し、そこにリガンドを結合させる
- リガンドは、回転・並進に加えて、回転可能な結合の二面角をすべて回転させて自由エネルギー(またはスコア)が最小となる構造 (poseと呼ばれる)を探索
- Receptorの原子は通常動かさず、剛体として扱うことが多い

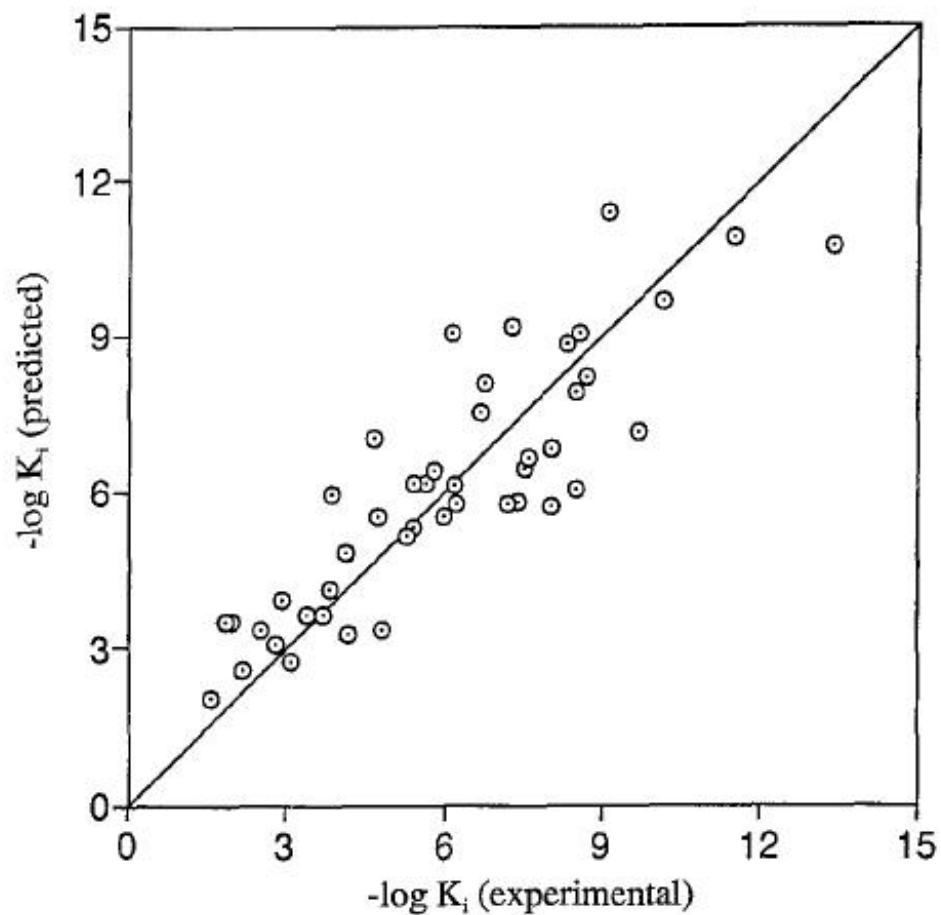
経験的スコア関数(1)

- Ludi

$$\Delta G_{\text{bind}}^{\circ} = \Delta G_0 + \Delta G_{\text{hb}} \sum_{\text{h-bonds}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{ionic}} \sum_{\text{ionicint.}} f(\Delta R, \Delta \alpha) \\ + \Delta G_{\text{lipo}} A_{\text{lipo}} + \Delta G_{\text{rot}} N_{\text{rot}}$$

- 結合自由エネルギー変化を、水素結合、イオン結合、疎水相互作用、リガンドの構造固定によるエントロピー損失の項の和で表す
- 45種類のタンパク質-低分子化合物複合体について、実験で得られる結合自由エネルギー変化と、立体構造から得られる、水素結合長、イオン結合長、疎水相互作用表面積、リガンドの回転可能結合数から上式で計算される値が合うように係数 ΔG_x を決める

経験的スコア関数(2)

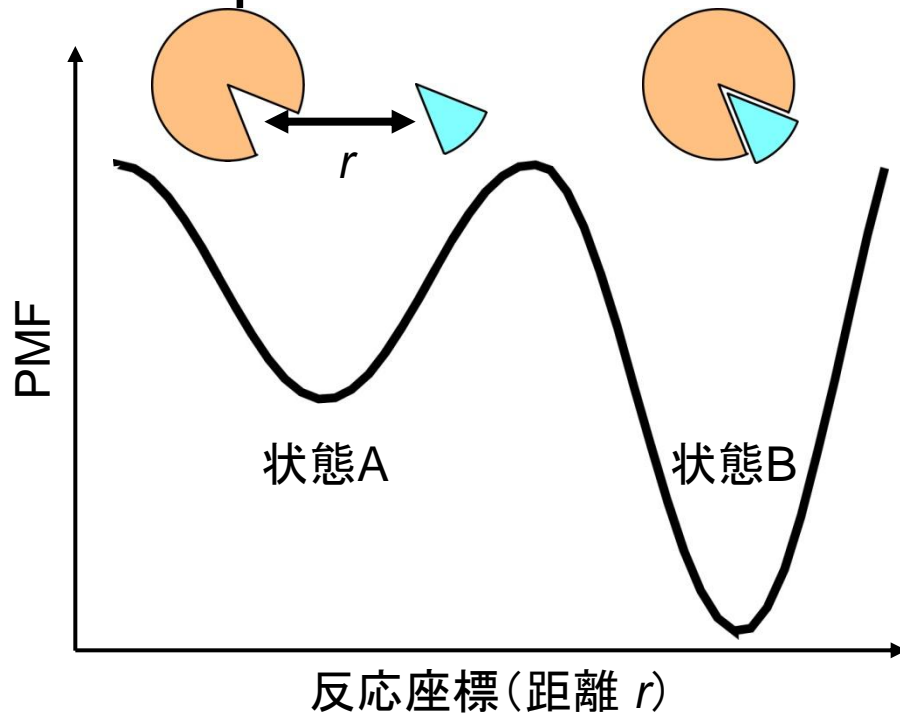


Böhm (1994) *J. Comput.-Aided Mol. Des.* **8**, 243.

統計ポテンシャル

- Potential of mean force (Pmf)

- 自由エネルギーを反応座標に沿ってプロットしたものはpotential of mean force (PMF)と呼ばれる



$$\Delta G = \Delta G_{\text{bind}}^{\circ} + RT \ln \frac{[B]}{[A]} = 0$$

$$\begin{aligned} \Delta G_{\text{bind}}^{\circ} &= -RT \ln \frac{[B]}{[A]} \\ &= -RT \ln p_B / p_A \end{aligned}$$

統計ポテンシャル

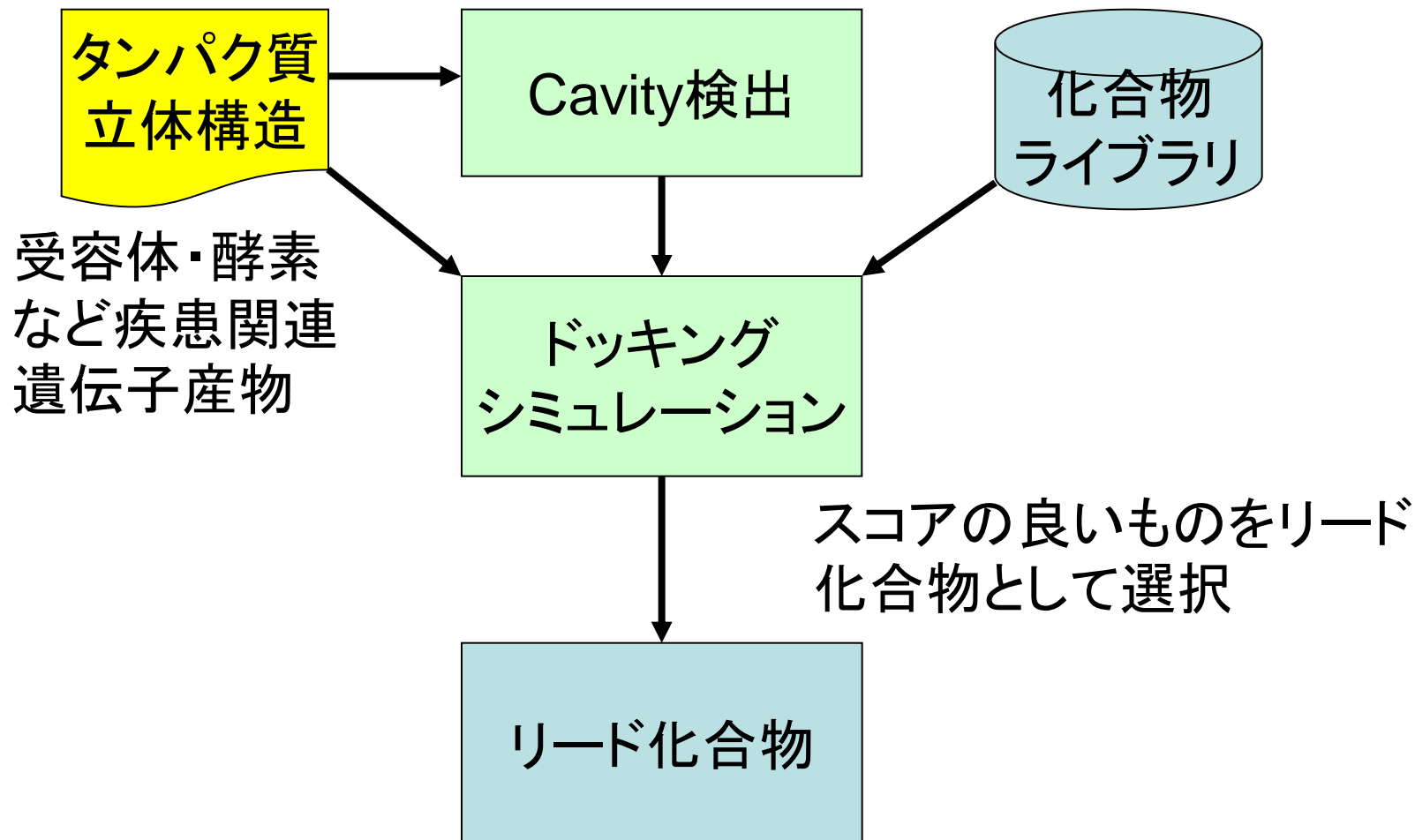
- Potential of mean force (Pmf)
 - 自由エネルギーを反応座標に沿ってプロットしたものはpotential of mean force (PMF)と呼ばれる
 - PMFは確率密度分布と対応付けられる
 - リガンドとタンパク質の原子間距離に対する確率密度分布を77個の複合体立体構造から計算し、原子種ペアごとにまとめて関数 $p_{ij}(r)$ を決める

$$\Delta G_{\text{bind}}^{\circ}(r) = -RT \ln p(r) / p_{\text{bulk}}(r) \approx -RT \sum_{k,l} \ln p_{ij}(r_{kl}) / p_{\text{bulk}}^{ij}(r)$$

ドッキングの創薬への応用

- 創薬の分野では薬剤候補化合物の探索に、化合物のライブラリから、標的タンパク質に強く結合する化合物を、大規模かつ効率的に探し出す high-throughput screening (HTS) がよく用いられる
- 化合物のライブラリの構築、結合のアッセイ系の確立には膨大なコストがかかる
- 化合物の標的タンパク質への結合をコンピュータの中で再現する(=ドッキングシミュレーション)ことで、親和性の評価が可能→virtual screening

Virtual screening



化合物ライブラリ

- Available Chemicals Directory (ACD)
 - 商用化合物データベース
 - <http://www.symyx.com/products/databases/sourcing/acd/>
 - 約1,100,000の化合物を収録
- ZINC
 - USCFが運営するfreeの化合物データベース
 - <http://zinc.docking.org/>
 - 約13,000,000の化合物を収録
- PubChem
 - NCBIが運営するfreeの化合物データベース
 - <http://pubchem.ncbi.nlm.nih.gov/>
 - 約27,000,000の化合物を収録

Cavity検出

- 酵素の基質ポケットや受容体のリガンド結合部位は、タンパク質分子表面のくぼみ (cavity) にあることが多い
- SURFNET
 - <http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html>
 - タンパク質分子表面の”gap region”を検出
- PASS
 - <http://www.ccl.net/cca/software/UNIX/pass/overview.shtml>
 - タンパク質分子表面のcavityを検出しランク付け
- Q-SiteFinder
 - <http://www.bioinformatics.leeds.ac.uk/qsitefinder/>
 - CH₃プローブのエネルギー値に基づいてランク付け

ドッキングソフトウェア

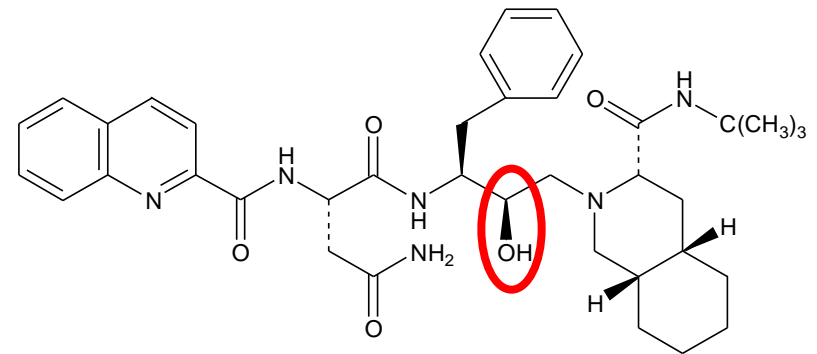
- DOCK
 - <http://dock.compbio.ucsf.edu/>
 - Cavityを特徴付ける球に化合物原子をフィット
- AutoDock
 - <http://autodock.scripps.edu/>
 - Genetic Algorithm (GA)による経験的結合自由エネルギースコアの最適化
- GOLD
 - http://www.ccdc.cam.ac.uk/products/life_sciences/gold/
 - GAによるスコア関数の最適化
- いずれも化合物の並進・回転と二面角の自由度のみを考慮し、タンパク質は剛体として扱う

ドッキングシミュレーション実習

- Discovery Studio Clientを用いてHIV proteaseに阻害剤をドッキングする
 1. HIV proteaseの結晶構造の取得
 2. Cavity検出
 3. 阻害剤構造データの取得
 4. ドッキングシミュレーション
 5. 結果の解析

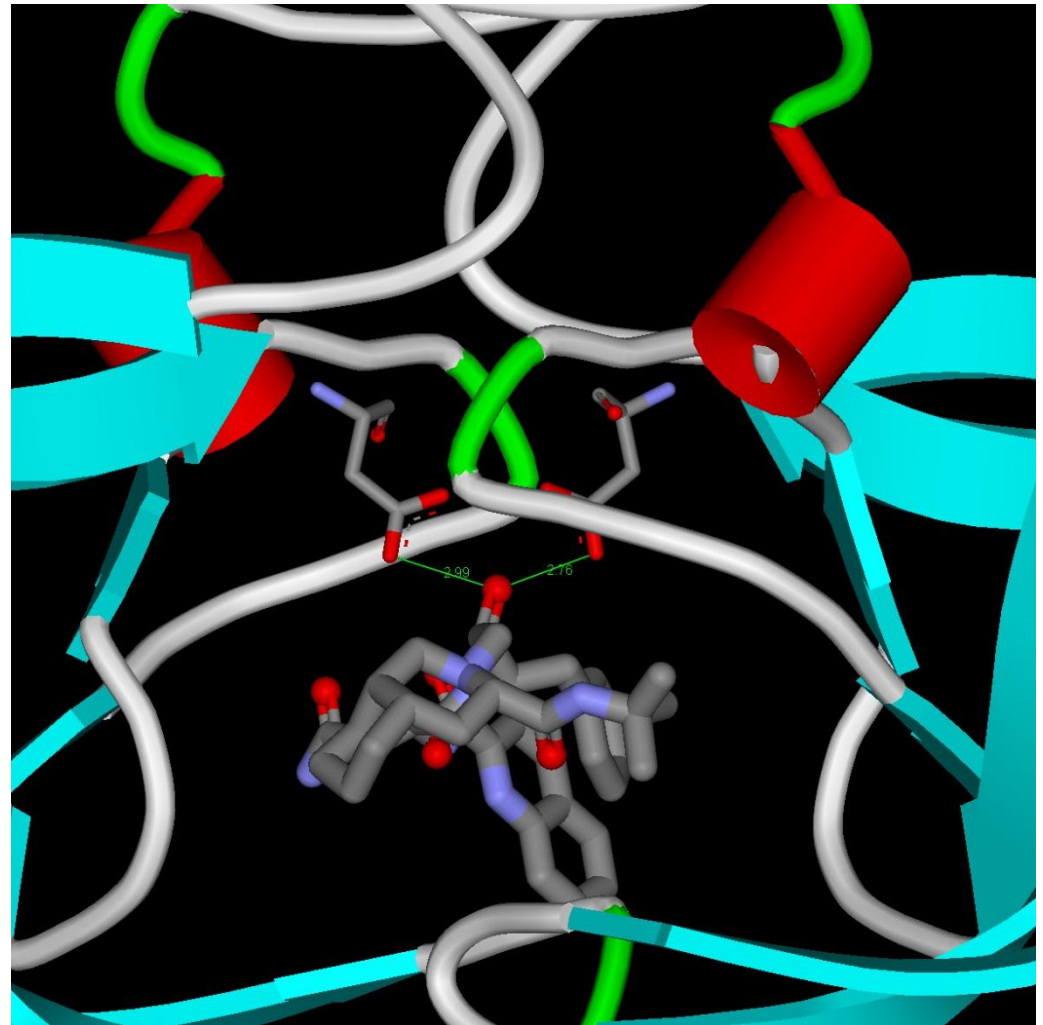
1. 結晶構造の取得

1. PDB ID 1HXBの構造を開く
2. この結晶構造には Rocheが開発し、HIVプロテアーゼ阻害抗エイズ薬として最初にFADに承認された、サキナビル (saquinavir) が結合している。この分子と水分子を削除する



参考：阻害のメカニズム

- 阻害剤のOH基がHIV proteaseの活性中心 Asp25と水素結合を形成している
- 阻害剤のOH基は加水分解反応の中間体(四面体構造)を模している
- 水分子が攻撃する位置に水素が付いているため反応は起こらない



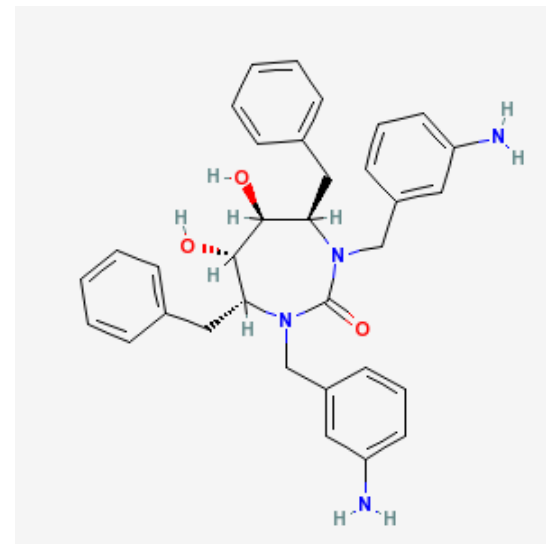
2. Cavity検出

1. Toolsタブの「Simulate Structures」でForcefieldに charmm22を指定し「Apply Forcefield」
2. Molecule Windowで「Ctrl」キーと「A」キーを同時に押し、タンパク質全体を選択
3. Toolsタブの「Define and Edit Binding Site」を展開し、「Define Selected Molecule as Receptor」を左クリック
4. 「Find Sites from Receptor Cavities」を左クリック
→Cavityが表示される
5. Hierarchy WindowでSite 1を選択して、Toolsタブ「Define and Edit Binding Site」の「Define Sphere from Selection」を左クリック
→Cavityを覆う球が表示される



3. 阻害剤構造データの取得(1)

- ここではDu Pontによって開発された阻害剤mozenavirのドッキングを行う
- PubChem (<http://pubchem.ncbi.nlm.nih.gov/>)にアクセスし、テキストボックスに「mozenavir」と入力し「GO」
- 1件ヒットするので、このID (CID: 154044) をクリック
- 構造データは、SDファイルとして保存できるが、ここではSMILESを使う
 1. Descriptors Computed from StructureにあるIsomeric SMILESをコピー
 2. Discovery Studioのメニューから「File」→「New」→「Molecule Window」で新しいMolecule Windowを生成
 3. 「File」→「Insert From」→「SMILES」を選択し、SMILES Stringに貼り付け



3. 阻害剤構造データの取得(2)

- 以下の通り名前の変更とエネルギー最小化を行う
 1. Hierarchy Windowで「Molecule 1」を左クリックした後右クリックして「Attributes of Molecule 1」を選択
 2. Nameを「mozenavir」に変更
 3. ToolsタブのSimulate Structuresを展開し、Forcefieldに「CHARMm」を指定して「Apply Forcefield」
 4. Protocolsタブから「Simulation」→「Minimization」を選択しダブルクリック
 5. Input Typed Moleculeに「Molecule:mozenavir」を指定して「Run」(エラーとなる場合は、ステップ3で「Clear Forcefield」してから再度「Apply Forcefield」)
 6. 計算が終了したら結果を表示する

4. ドッキングシミュレーション

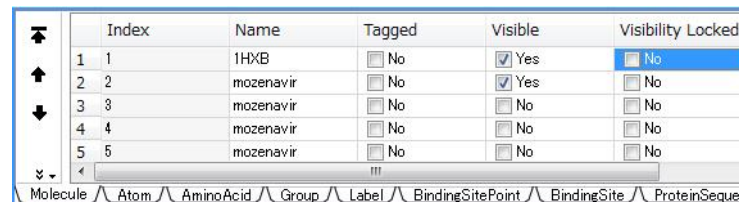
1. Protocolsタブの「Receptor-Ligand Interactions」→「Dock Ligands (CDOCKER)」を選択しダブルクリック
2. Input Receptorに「1HXB:1HXB」を選択
3. Input Ligandsに「mozenavir:Visible」を選択
4. Input Site Sphereに1つだけ候補が表示されるのでそれを選択し「Run」(6分ほどかかる)
5. 計算が終了したら結果を表示する

CDOCKER

- 開発者
 - C. L. Brooks III, M. Viethら
 - Wu *et al.* *J. Comput. Chem.* **24**, 1549 (2003).
- エネルギー関数
 - CHARMM
- 最適化法
 - Simulated annealing (SA)とエネルギー最小化
 - SAではグリッドベースの相互作用エネルギー計算
 - エネルギー最小化では全原子ポテンシャルエネルギー関数に基づくエネルギー計算

5. 結果の解析

1. Output filesにある「View Results」をクリック
2. 新しく表示されるMolecule WindowのData Tableで、1HXBの行のVisibility Lockedの列のチェックをはずす
3. Hierarchy Windowを表示し、結合サイト (Site 1およびSBD_Site_Sphere)のチェックをはずし非表示にする
4. メニューの「Chemistry」→「Hydrogens」→「Hide」を選択すると、水素原子が非表示となり見やすくなる
5. Data Tableの2行目以降は、ドッキング結果 (pose) がCDOCKER_ENERGYの大きい順に並んでおり、Visibleの行をチェックすると表示できる




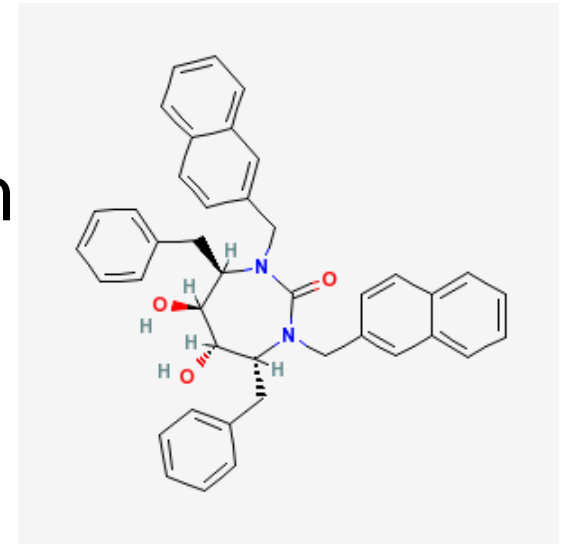
	Index	Name	Tagged	Visible	Visibility Locked
1	1	1HXB	<input type="checkbox"/> No	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No
2	2	mozenavir	<input type="checkbox"/> No	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No
3	3	mozenavir	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
4	4	mozenavir	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
5	5	mozenavir	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No

正解構造との比較


1. メニューの「File」→「Insert From」→「URL」を開き、PDB IDに「1DMP」を指定し「Open」
2. メニューの「Sequence」→「Show Sequence」を選択し、1HXBと1DMPの配列を表示する
3. Sequence Windowで1HXBを選択した後、ドッキング結果が表示されているMolecule Windowに移動し、「Structure」→「Superimpose」→「By Sequence Alignment」を選択、Molecules to Superimposeに「1DMP」を指定して「OK」
4. 1DMPのタンパク質(A鎖およびB鎖)を非表示にして、リガンドの構造を比較せよ

課題

- Mozenavirの構造を元に、Sketch ツールを用いて1HVRに結合している阻害剤XK2を作成せよ
- 1位のドッキング構造に1HVRのリガンドの構造を重ね合わせた図を寺田宛tterada@iu.a.u-tokyo.ac.jpに提出せよ
- -CDOCKER_ENERGYを比較し、XK2と mozenavirではどちらがHIV proteaseと強く結合すると考えられるか考察せよ
- その際件名は「分子モデリング課題」とし、本文に氏名と学生証番号、考察を必ず明記すること



Exercise

- Build a model of XK2 using the model of mozenavir and “Sketch tool.” 
- Compare the first-ranked structure with the crystal structure, 1HVR, by superimposing the protein atoms of 1HVR on those of the receptor and create the image file.
- Discuss which of XK2 and mozenavir binds to the HIV protease stronger based on “-CDOCKER_ENERGY.”
- Send an email attaching the image file to tterada@iu.a.u-tokyo.ac.jp. The subject of the email should be “Molecular modeling.” Write your name, ID card number, and discussion in the mail body.

