

第5章

統計力学モデルとシミュレーション による熱力学量を比較してみよう！

<要旨>

統計力学モデルであるA-W_NILSモデルを3次元格子タンパク質に適用し、理論的に計算された熱力学量と、フォールディングのシミュレーションから得られる熱力学量とを比較・検討している。

ここで採用したモデルは、1990年代後半に米国のShakhnovichらによって、天然構造にフォールドするようにデザインされたアミノ酸配列をもつ（36個及び48個）3次元格子タンパク質である。

さらに、いろいろな温度におけるシミュレーションにおいて、どのような相互作用が、どの程度実現しているかを具体的に解析した結果を示している。

そして、A-W_NILSモデルの妥当性を詳細に検討し、フォールディングの統計力学的描像を議論している。

コンピュータ・シミュレーションを実行するタンパク分子の3次元格子模型の導入

統計力学的モデルである、「A-W_NILS モデル」によって理論的に計算されるフォールディングに関する熱力学量の妥当性を検討するために、3次元格子タンパク質のフォールディング・シミュレーション (Appendix G: 「3次元格子タンパク質のコンピュータ・シミュレーションの方法」参照) によって、フォールディングに関する熱力学量を求めよう。

まず、タンパク質分子の“3次元格子模型”を導入しよう。ここで採用する3次元格子タンパク質模型は、1990年代後半、米国の Shakhnovich らによって導入された。そして、彼らは3次元格子タンパク質模型によるコンピュータ・シミュレーションを本格的に実行した。

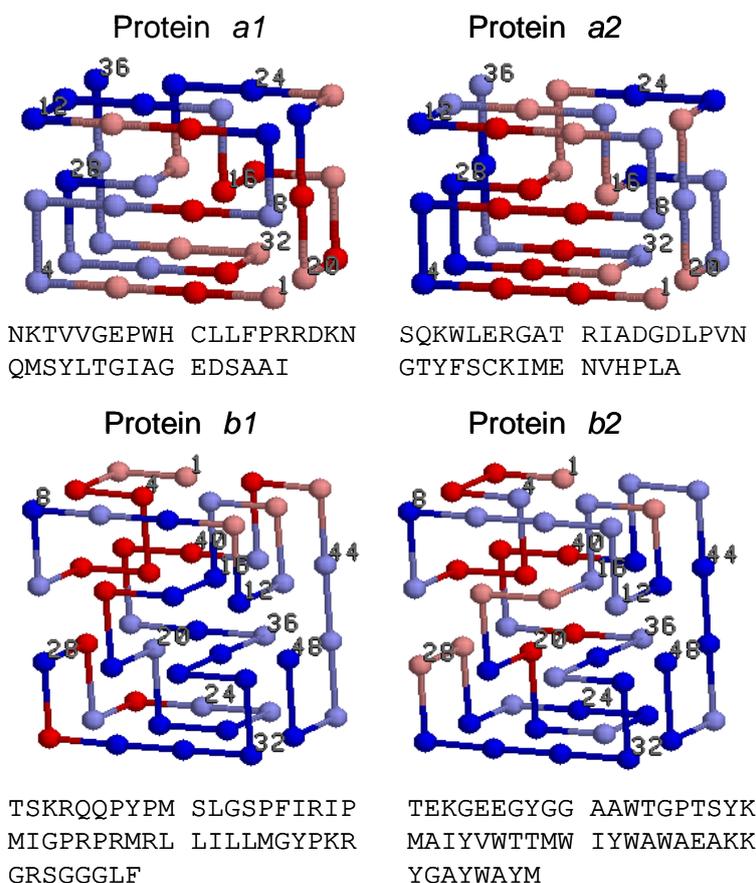
<文献> L. Mirny, V. Abkevich and E. Shakhnovich, Universality and Diversity of the Protein Folding Scenaris: A Comprehensive Analysis with the Aid of a Lattice Model, *Folding and Design*, Vol.2, pp.103-116, 1996.

<文献> Abkevich V.I, Gutin AM, Shakhnovich E.I.: Improved Design of Stable and Fast-Folding Model Proteins, *Folding and Design*, Vol. 1, pp.221-230, 1996..

我々が、フォールディング・シミュレーションを実行する3次元格子タンパク質の天然構造は、Protein a1, Protein a2, Protein b1, Protein b2 の4個である (下図参照)。

3次元格子タンパク質の Protein a1 と Protein a2 は、アミノ酸残基数が36個で、同じ天然構造であるが、アミノ酸配列が異なっている。更に、3次元格子タンパク質の Protein b1 と Protein b2 は、アミノ酸残基数が48個で、同じ天然構造であるが、アミノ酸配列が異なっている。

3次元格子タンパク質の図の下に、それぞれのアミノ酸配列を一文字表示で記している。



図中の青色系, 赤色系は, それぞれ, 疎水性, 及び, 親水基のアミノ酸残基を表している (Appendix C: 「20種類のアミノ酸」参照) :

青色(blue) ; I(Ile), L(Leu), M(Met), F(Phe), W(Trp), Y(Tyr)

空色(light blue) ; A(Ala), G(Gly), P(Pro), V(Val)

ピンク(pink) ; N(Asn), D(Asp), C(Cys), S(Ser), T(Thr)

赤色(red) ; R(Aeg), E(Glu), Q(Gln), H(His), K(Lys)

3次元格子タンパク質のシミュレーションより転移曲線を描く

まず, Protein a1 と Protein a2 の温度毎のフォールディング・シミュレーションの結果より, 次のような手順で転移曲線を描いてみよう :

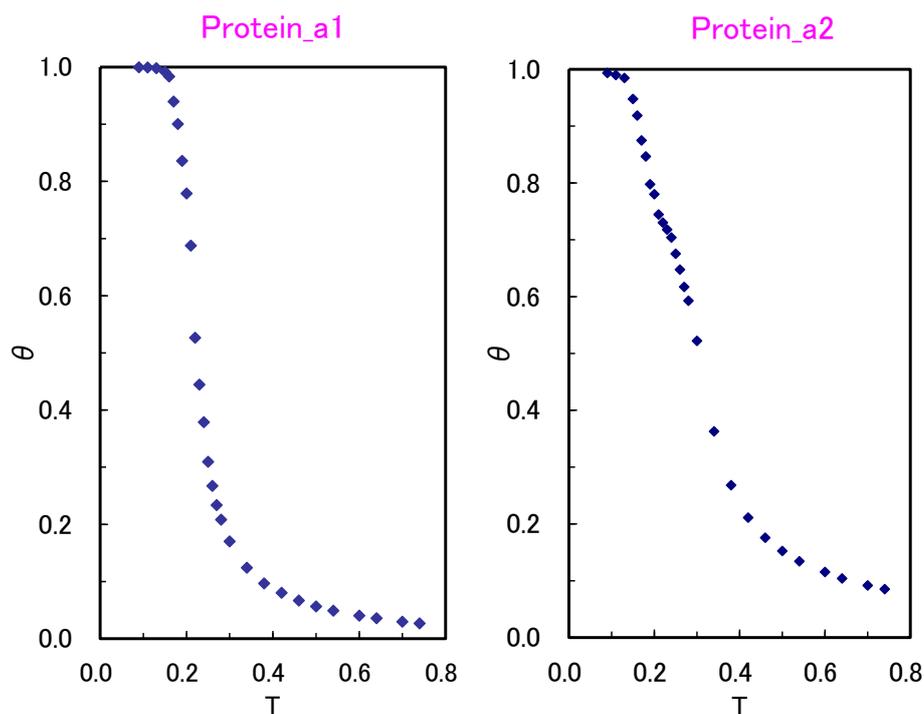
- ① タンパク質分子のフォールディング・シミュレーションの結果から, そこで生成された全てのコンフォメーションのエネルギーの期待値 $\langle E \rangle$ を求める.
- ② エネルギーの期待値 $\langle E \rangle$ を, 天然構造でのエネルギーの値 E_{native} で割った値 θ を求める.
この量は, 生成された全てのコンフォメーションが, 平均として, どれだけ天然構造に近いかを表す指標 (秩序変数) とみなせる ;

$$\theta = \frac{\langle E \rangle}{E_{native}}$$

(天然構造のとき, $\theta = 1.0$, 完全に伸びきった構造のとき, $\theta = 0.0$)

- ③ θ を, 各温度毎に求めて, 転移曲線を描く (図を参照).

図の Protein a1 の転移曲線を眺めると, ほぼ二状態的な転移と見なせるが, Protein a2 の方は, 転移曲線がゆるやかである. このことは, Protein a1 の方が, Protein a2 よりも, より協同的な転移現象であることを示している.





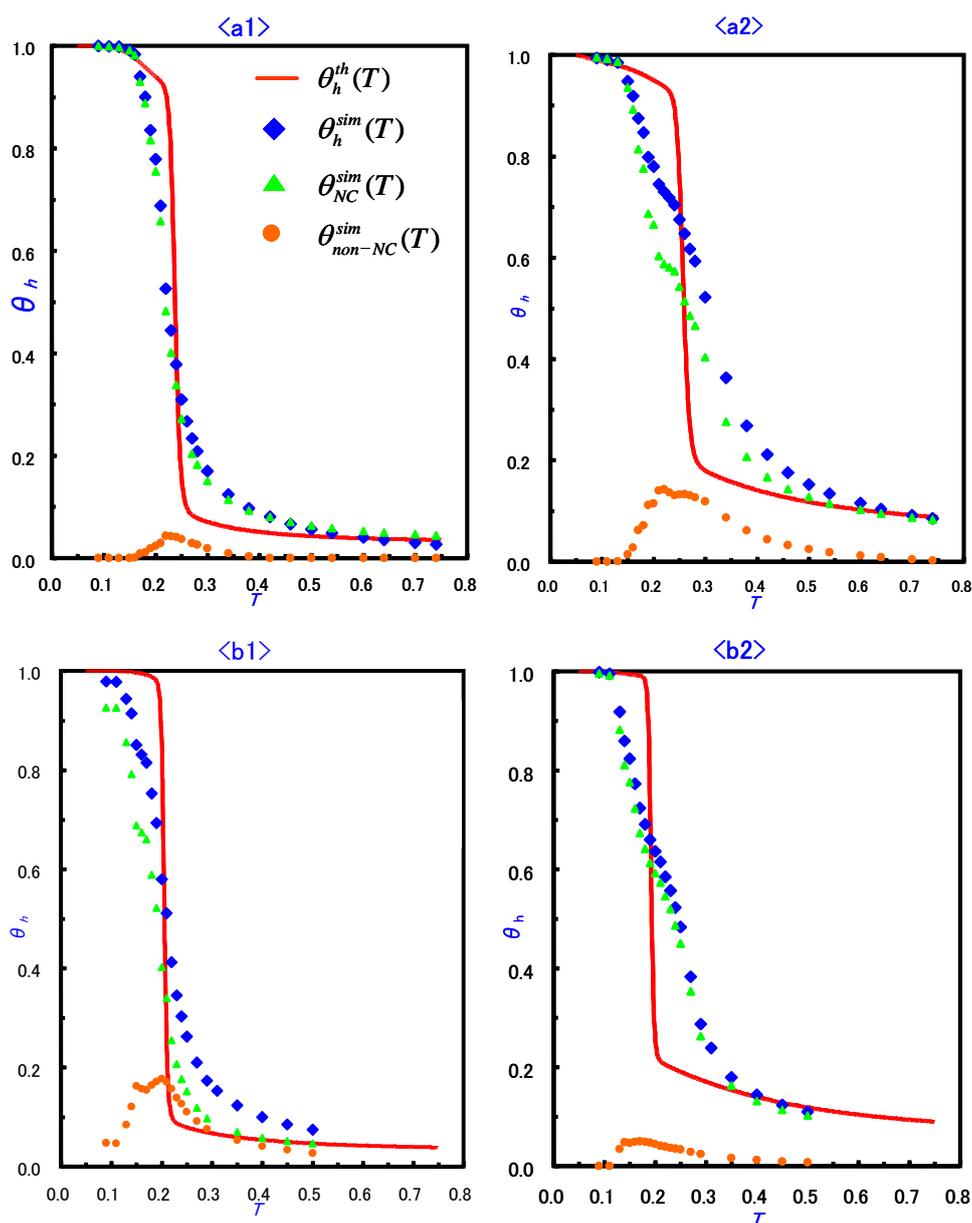
“シミュレーション” と “A-W_NILS モデル” による転移曲線の比較

図は、「A-W_NILS モデル」とコンピュータ・シミュレーションより求めた転移曲線 $\theta_h(T)$ を比較したものである。

<文献> H. Abe and H. Wako, Analyses of simulations of three-dimensional lattice protein in comparison with a simplified statistical mechanical model of protein folding, *Physical Review*, Vol.74, pp. 011913-1, 011913-12, 2006.

<文献> 輪湖 博, 安部 晴男, “タンパク質の立体構造転移”, 複雑系業書 1, 複雑系の構造と予測, 早稲田大学複雑系高等学術研究所編, 共立出版, pp. 59-97, 2006.

<文献> 安部 晴男, 輪湖 博, “タンパク質フォールディングの統計熱力学, II. 格子タンパク質のフォールディング過程”, 西日本工業大学紀要, 第37巻, pp. 49-58, 2007.



上図に描かれている、種々の転移曲線を説明しよう：

○「A-W_NILS モデル」による転移曲線 (— $\theta_h^{th}(T)$ 曲線)

統計力学モデルである「A-W_NILS モデル」より、転移曲線を求める手順は次の通りである：

① 系の分配関数を計算する、② 分配関数より平均のエネルギーを求める、③ 規格化して、温度毎の転移曲線を求める。

この手順を詳細に説明しよう。分配関数は、次式のように書ける；：

$$Z(T) = \sum_h W_1(h) u^h$$

ただし、

$$W_1(h) = \sum_{\eta=0}^{n-3} \Omega(\eta, h) t^\eta, \quad u = \exp(-\beta \varepsilon_0)$$

ここで、 $W_1(h) u^h$ は、エネルギーが $E_h (= h \varepsilon_0)$ である全ての状態の統計重率の和である。上記の分配関数から、系のエネルギーの期待値 $\langle E \rangle$ を求める ($\beta = 1/k_B T$)；

$$\langle E(T) \rangle = \frac{\sum_h E_h W_1(E_h) e^{-\beta E_h}}{Z}$$

$$\theta_h^{th}(T) = \frac{\langle E(T) \rangle}{E_{native}}$$

$\theta_h^{th}(T)$ は温度 T の関数であり、天然構造のエネルギー値 E_{native} で規格化している。この量は、タンパク質分子のフォールディングの程度を表す指標である：

天然構造のとき、 $\theta = 1.0$ 、完全に伸びきった構造のとき、 $\theta = 0.0$ 。

○シミュレーションによる種々の転移曲線

シミュレーションの記録を解析するために、全相互作用のエネルギー E を次の二つのタイプ、 E_{NC}, E_{non-NC} に分類しよう：

$$E \begin{cases} E_{NC} & \text{(天然接触のエネルギーの和)} \\ E_{non-NC} & \text{(非天然接触のエネルギーの和)} \end{cases}$$

当然、次式が成立する： $E = E_{NC} + E_{non-NC}$

シミュレーションでは、次の3種類の結果を示している：

(1) <◆ $\theta_h^{sim}(T)$ 曲線>；

これは、各温度でのフォールディング・シミュレーションから求めた量である。

各点の◆は、その温度での、次の様なシミュレーションの結果である。

シミュレーションの初期構造は、完全にほどけた構造を採用し、試行回数は、 10^8 回である。但し、低温では天然構造のまわりでの揺らぎだけしか起こらないことが自明であるため、初期構造は天然構

造を採用し、試行回数は、 10^7 回とした。シミュレーションの途中では、2つのユニットが隣接する格子点を占有した場合、またそのときのみ、アミノ酸残基間の相互作用が発生するが、そのとき、全てのアミノ酸残基間に相互作用が働くとした。つまり、「天然接触ペア」（天然構造で相互作用しているアミノ酸残基ペア）であろうと「非天然接触ペア」であろうと相互作用が働くとしている。最終的に、各温度でのフォールディングのシミュレーションの結果から、そこで生成された全てのコンフォメーションのエネルギーの期待値 $\langle E \rangle$ を求め、天然構造のエネルギー値 E_{native} で規格化する。従って、この量 $\langle E \rangle$ は、生成された全てのコンフォメーションの平均として、その温度で、どれだけ天然構造に近いかを表す指標（秩序変数）とみなせる。

(2) <▲ $\theta_{NC}^{sim}(T)$ 曲線> ;

これは、シミュレーションの結果を解析して求められる次のような量である：

$$\theta_{NC}^{sim}(T) = \frac{1}{E_{native}} \cdot \langle E_{NC}(T) \rangle$$

各温度でのフォールディングのシミュレーションにおいて生成されたすべてのコンフォメーションについて、すべての「天然接触」（NC, Native Contact）のアミノ酸残基間相互作用のエネルギーの和の平均を規格化した量である。これは、局所構造内の「天然接触」のアミノ酸残基間相互作用のみを考慮している統計力学モデルである「A-W_NILS モデル」から求めた、 $\theta_h^{th}(T)$ 曲線と比較することによって、「A-W_NILS モデル」の妥当性を検討するための量である。

(3) <● $\theta_{non-NC}^{sim}(T)$ 曲線> ;

これは、シミュレーションの結果を解析して求められる次のような量である：

$$\theta_{non-NC}^{sim}(T) = \frac{1}{E_{native}} \cdot \langle E_{non-NC}(T) \rangle$$

各温度でのフォールディングのシミュレーションにおいて生成されたすべてのコンフォメーションについて、すべての「非天然接触」（non-NC, non-Native Contact）のアミノ酸残基間相互作用のエネルギーの和の平均を規格化した量である。これは、局所構造内の「天然接触」のアミノ酸残基間相互作用のみを考慮している統計力学モデルである「A-W_NILS モデル」から求めた、 $\theta_h^{th}(T)$ 曲線と、シミュレーションからもとめた $\theta_h^{sim}(T)$ 曲線との“ずれ”の原因を解明しようとするための量である。

次に、3次元格子タンパク質分子のフォールディング・シミュレーションによって求められた転移曲線と、「A-W_NILS モデル」によって理論的に計算した転移曲線とを比較・検討しよう。

○<転移温度の比較>

図のフォールディングの転移曲線で、転移温度 (T_m : $\theta = 0.5$ に対応する温度) を比較すると、タンパク質 a1, a2, b1 では、理論とシミュレーションから得られる転移温度はほぼ一致している。ただし、タンパク質 b2 では、少し転移温度のずれがある。

調整パラメータは一切使用していない「A-W_NILS モデル」より求めた転移温度が、シミュレーションより求めた転移温度とほぼ等しいことは新鮮な驚きである。簡単な統計力学モデルである

「A-W_NILS モデル」が、第一近似として有力なモデルといえるだろう。

○＜転移曲線の比較＞

シミュレーションから求めた二つの曲線、＜◆ $\theta_h^{sim}(T)$ 曲線＞と＜▲ $\theta_{NC}^{sim}(T)$ 曲線＞は、タンパク質 a1, b2 ではほぼ一致している。タンパク質 a2, b1 では不一致の領域がみられるが、これは、＜● $\theta_{non-NC}^{sim}(T)$ 曲線＞で示しているように、シミュレーションの途中で「非天然接触ペア」の割合が少し増加しているためである。

また、タンパク質 a2, b2 に対する転移曲線＜◆ $\theta_h^{sim}(T)$ 曲線＞は、タンパク質 a1, b1 に対する転移曲線＜◆ $\theta_h^{sim}(T)$ 曲線＞より、より緩やかな曲線になっている。

理論とシミュレーションの転移曲線の形状に関しては、タンパク質 a1, b1 は比較的良好に一致しているが、タンパク質 a2, b2 はかなり異なっている。更に、「A-W_NILS モデル」より求めた転移曲線＜－ $\theta_h^{th}(T)$ ＞は、シミュレーションから求めた転移曲線＜◆ $\theta_h^{sim}(T)$ 曲線＞より急勾配の傾向にある。

この違いを、詳細に次の3つの温度の領域で検討しよう：

(イ) 転移温度 (T_m) 以下 (つまり低温) の領域, (ロ) 転移温度近傍の領域, (ハ) 転移温度 (T_m) 以上 (つまり, 高温) の領域.

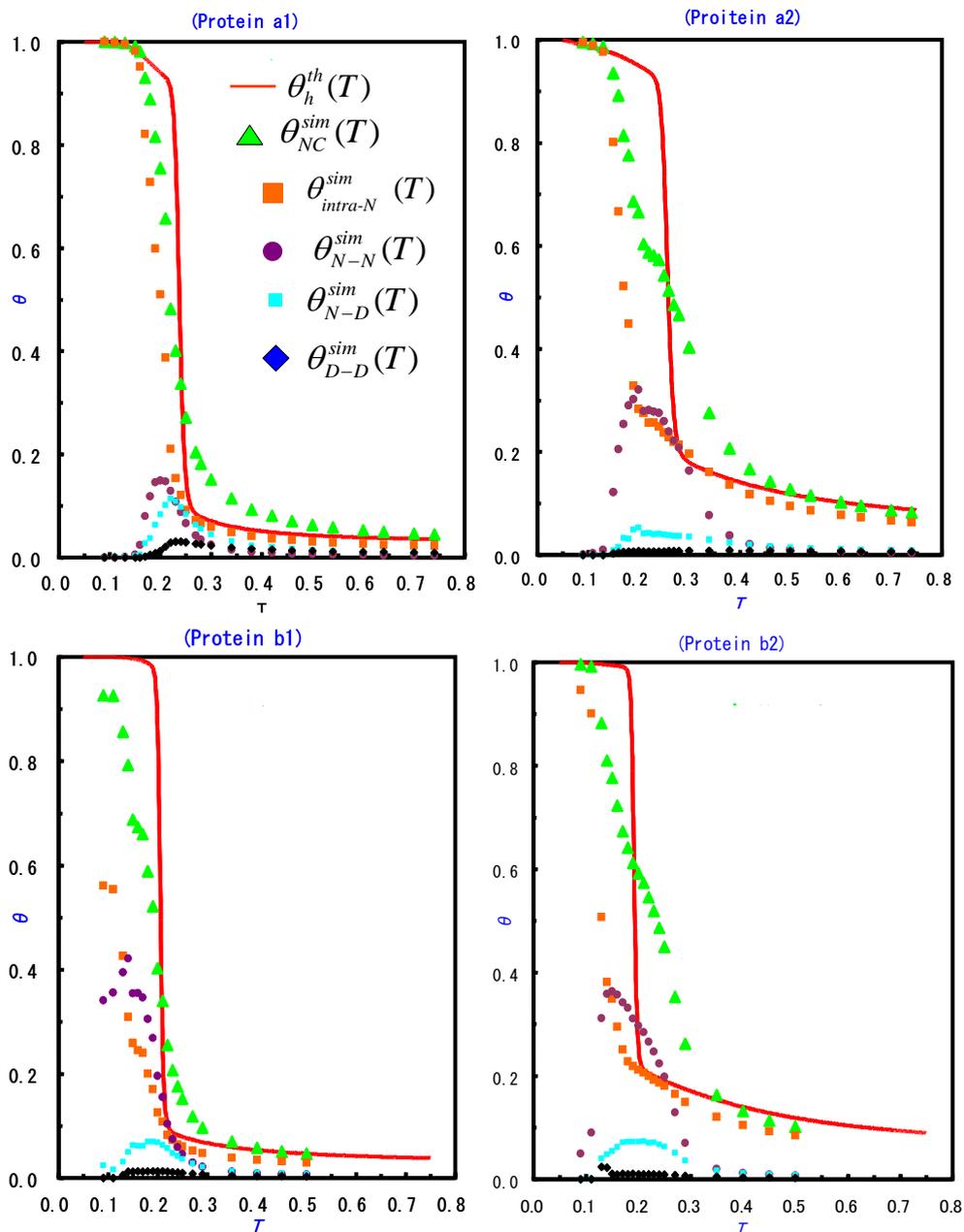
- ・転移温度 (T_m) 以下 (低温) の領域と、転移温度 (T_m) 以上 (高温) の領域では、「A-W_NILS モデル」から得られる結果である＜－ $\theta_h^{th}(T)$ ＞曲線と、シミュレーションから得られる結果である＜◆ $\theta_h^{sim}(T)$ 曲線＞とは、タンパク質 a1, a2, b1, b2 とも、よく一致している。その理由は、＜● $\theta_{non-NC}^{sim}(T)$ 曲線＞が、この温度領域では、4個のタンパク質とも「非天然接触ペア」がほとんど実現していなくて、ほぼ「天然接触」だけが実現しているからである。つまり、この温度領域では、Gō ポテンシャルが主に働いている。
- ・転移温度 (T_m) 近傍の領域では、＜● $\theta_{non-NC}^{sim}(T)$ 曲線＞からわかるように、タンパク質 a1, b2 は、「非天然接触」の割合が小さく、ほぼ「天然接触」だけが実現しているが、タンパク質 a2, b1 では、「非天然接触」が、かなりの割合で存在している結果、タンパク質 a2, b1 での、「A-W_NILS モデル」とシミュレーションとの転移曲線の“ずれ”が発生している。つまり、転移温度 (T_m) 近傍での「非天然接触ペア」の寄与がより大きいことに起因している。

しかしながら、タンパク質 b2 では、＜● $\theta_{non-NC}^{sim}(T)$ 曲線＞からわかるように、「非天然接触」の寄与が小さいにもかかわらず、転移曲線は、「A-W_NILS モデル」とシミュレーションとで“ずれ”がある。この原因は一体どこからくるのだろうか？ ここで、「A-W_NILS モデル」で導入した相互作用の仮定、“天然接触”の相互作用のエネルギーは、局所構造内のみで働くとする仮定”が、どの程度成り立っているかを検証するために、各温度でのシミュレーションにおける、いろいろな相互作用の寄与、特に、「天然接触」の詳細を解析しよう。

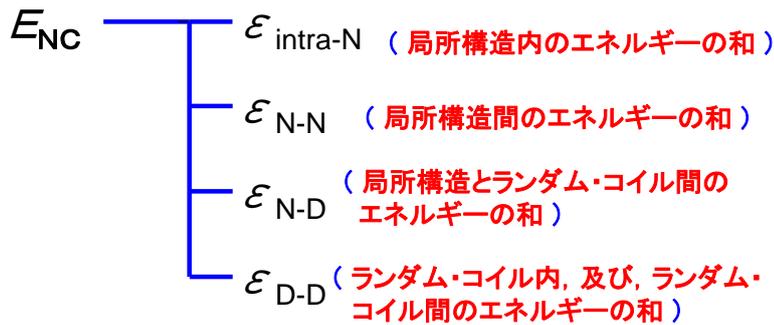


“シミュレーション”におけるいろいろな相互作用の寄与と“A-W_NILSモデル”の検証

「A-W_NILS モデル」で導入した相互作用エネルギーの寄与がどの程度成り立っているかを検証しよう。そのために、シミュレーションにおいて、いろいろな相互作用エネルギーの寄与の詳細を解析した結果が下図である。



シミュレーションの記録を詳細に検証するために、「天然接触」による相互作用のエネルギー E_{NC} を次の4つのタイプ, $\epsilon_{\text{int ra-N}}, \epsilon_{N-N}, \epsilon_{N-D}, \epsilon_{D-D}$ に分類しよう:



(当然, 次式が成立する: $E_{NC} = \epsilon_{\text{int ra-N}} + \epsilon_{N-N} + \epsilon_{N-D} + \epsilon_{D-D}$)

図には, 「A-W_NILS モデル」から得られる転移曲線, $\langle -\theta_h^{\text{th}}(T) \text{ 曲線} \rangle$ と, シミュレーションから, 「天然接触」(NC, native contacts) のアミノ酸残基間相互作用のみの結果を解析した

$\langle \blacktriangle \theta_{NC}^{\text{sim}}(T) \text{ 曲線} \rangle$, つまり, 次のような量:

$$\theta_{NC}^{\text{sim}}(T) = \frac{1}{E_{\text{native}}} \cdot \langle E_{NC}(T) \rangle$$

が描かれている.

この二つの曲線の相違はどこから発生しているのかを検証するために, $\theta_{NC}^{\text{sim}}(T)$ を次の4個のタイプに分類しよう:

(1) $\langle \blacksquare \theta_{\text{int ra-N}}^{\text{sim}}(T) \text{ 曲線} \rangle$;

ある温度のシミュレーションでの全てのコンフォメーションにおいて, 「天然接触」による相互作用のエネルギーの内, 局所構造内の相互作用エネルギーの寄与の割合 (この相互作用エネルギーは「A-W_NILS モデル」で考慮している);

$$\theta_{\text{int ra-N}}^{\text{sim}}(T) = \frac{1}{E_{\text{native}}} \cdot \langle E_{\text{int ra-N}}(T) \rangle$$

(2) $\langle \bullet \theta_{N-N}^{\text{sim}}(T) \text{ 曲線} \rangle$;

ある温度のシミュレーションでの全てのコンフォメーションにおいて, 「天然接触」による相互作用のエネルギーの内, 局所構造間の相互作用エネルギーの寄与の割合 (この相互作用エネルギーは, 「A-W_NILS モデル」で考慮していない);

$$\theta_{N-N}^{\text{sim}}(T) = \frac{1}{E_{\text{native}}} \cdot \langle E_{N-N}(T) \rangle$$

(3) $\langle \blacksquare \theta_{N-D}^{\text{sim}}(T) \text{ 曲線} \rangle$;

ある温度のシミュレーションでの全てのコンフォメーションにおいて, 「天然接触」による相互作用のエネルギーの内, 局所構造とランダム・コイル領域との相互作用エネルギーの寄与の割合 (この相互作用エネルギーは, 「A-W_NILS モデル」で考慮していない);

$$\theta_{N-D}^{sim}(T) = \frac{1}{E_{native}} \cdot \langle E_{N-D}(T) \rangle$$

(4) <◆ $\theta_{D-D}^{sim}(T)$ 曲線> ;

ある温度のシミュレーションでの全てのコンフォメーションにおいて、「天然接触」による相互作用のエネルギーの内、ランダム・コイル領域内、及び、ランダム・コイル領域内の相互作用エネルギーの寄与の割合（この相互作用エネルギーは、「A-W_NILS モデル」で考慮していない）；

$$\theta_{D-D}^{sim}(T) = \frac{1}{E_{native}} \cdot \langle E_{D-D}(T) \rangle$$

これら4つの相互作用エネルギーの寄与を、4個のタンパク質、a1, a2, b1, b2について考察しよう。

○ 転移温度 (T_m) より高い温度（高温）領域では、タンパク質 a1, a2, b1, b2 とも、3つの曲線、<− $\theta_h^h(T)$ 曲線>, <▲ $\theta_{NC}^{sim}(T)$ 曲線>, <■ $\theta_{int\ ra-N}^{sim}(T)$ 曲線>が、ほぼ一致している。これから、次の3つのことがわかる：

- (1) <▲ $\theta_{NC}^{sim}(T)$ 曲線>と<■ $\theta_{int\ ra-N}^{sim}(T)$ 曲線>とがほぼ一致していることから、この高温領域のシミュレーションでは、コンフォメーションの天然接触の相互作用エネルギーは、ほぼ、局所構造内の天然接触エネルギーである。
- (2) <− $\theta_h^h(T)$ 曲線>と<■ $\theta_{int\ ra-N}^{sim}(T)$ 曲線>とが、ほぼ一致していることから、この高温領域では、「A-W_NILS モデル」の仮定（局所構造内の相互作用エネルギーのみを考慮し、他の相互作用エネルギーは無視するという仮定）が、ほぼ成立している。
- (3) タンパク質 a2, b2 に対するこれらの3つの曲線の値は、タンパク質 a1, b1 の曲線の値より大であることから、タンパク質 a2, b2 では、この高温領域でも、いくつかの安定した小さな局所構造が形成されている。一方、タンパク質 a1, b1 では、ほぼランダムコイル状態であることがわかる。

○ タンパク質 a1, a2, b1, b2 のすべてで、温度が減少するにつれて、転移温度 (T_m) 付近から、<● $\theta_{N-N}^{sim}(T)$ 曲線>（局所構造間の相互作用エネルギーの寄与の割合）が上昇している。この曲線の上昇にともなって、<− $\theta_h^h(T)$ 曲線>と、<■ $\theta_{int\ ra-N}^{sim}(T)$ 曲線>との差が増加していくのがわかる。このことは、この温度付近では、「A-W_NILS モデル」で採用している「局所構造内の相互作用エネルギー」の他に、「局所構造間の相互作用エネルギー」が無視できないことを示している。しかしながら、タンパク質 a1 の $\theta_{N-N}^{sim}(T)$ （<●>局所構造間の相互作用エネルギーの寄与の割合）は、タンパク質 a2, b1, b2 よりも小さく、「A-W_NILS モデル」の仮定は、第一近似として成立しているとみなせる。

○ タンパク質 a1, a2, b1, b2 のすべてで、<■ $\theta_{N-D}^{sim}(T)$ 曲線>と<◆ $\theta_{D-D}^{sim}(T)$ 曲線>の値は、小さい。

上記のような、シミュレーションと「A-W_NILS モデル」による転移曲線の詳細な比較から、統計力学モデルである「A-W_NILS モデル」の妥当性に関して何が言えるだろうか？



転移曲線の詳細な検討による，“A-W_NILS モデル”の妥当性と

“A-W_NILS モデル”の限界

1970年代後半，Gōらは，フォールディング過程において，天然構造では接触していない相互作用，“天然非接触相互作用”の効果は正負両方があり，大まかにはそれらは相殺するので，これらの効果はほとんど無視してもよいであろうと考え，天然構造で互いに接触している分子間相互作用，“天然接触相互作用”のみが働く“Gōポテンシャル”を仮定して，コンピュータ・シミュレーションを実行した。

その結果，タンパク質らしい協同的振る舞いを観察した．そして，“天然非接触相互作用”も考慮したシミュレーションでは，フォールディングの協同性が少し崩れることも観測した。

すでに述べたように，当初，Gōポテンシャルは，「天然構造が実現して安定になるように無理やりにバイアスをかけた，物理的ではなく，乱暴な，しかも，人為的なポテンシャルである」と酷評された。

実際のタンパク質分子のフォールディング過程では，天然接触の相互作用ばかりではなく，非天然接触ペアの相互作用も，当然働いているであろう．「Gōポテンシャル」が実際のタンパク質のフォールディング過程でも第一近似として成立しているのだろうか？

1990年代後半には，『実際のタンパク質では，整合性原理が成り立っており，フォールディング過程では，グローバルなファネル様エネルギー地形が実現していて，天然接触相互作用のみが主に働く．』という考えが広く指示され，Gōポテンシャルが実際のタンパク質でも近似として使えるのではと考えられるようになった。

そして，「Gōポテンシャル」を採用したシミュレーションや理論的な統計力学モデルが提案され，検証されるようになった．「A-W_NILS モデル」もこの延長上にあり，Gōポテンシャルを更に理想化して，「天然接触は局所構造内のみで作用する」と仮定したモデルである。

1990年代後半，米国の Shakhnovich らによって，3次元格子モデルタンパク質が導入された．彼らは天然構造へフォールドするようなアミノ酸配列をもつ3次元格子モデルタンパク質をデザインした。

我々は，彼らが導入した，2種類の天然構造である3次元格子モデルタンパク質と，それぞれに対応する2種類のアミノ酸配列を採用して，フォールディングのシミュレーションを行って研究している。

「A-W_NILS モデル」の利点は，局所構造内の「天然接触」(NC) 相互作用のみを採用し，他の相互作用，つまり，「非天然接触」(non-NC) 相互作用，「局所構造間」(inter-N, N-N) 相互作用，「局所構造とランダム・コイル間」(N-D)，及び，「ランダム・コイル内，及び，ランダム・コイル間」(D-D) 相互作用を無視することで，系の分配関数が計算可能である点と，調節パラメータ (adjustable parameter) が含まれていない点である．そして，その分配関数を容易に求めることができ，その分配関数から熱力学量を計算して，その結果とコンピュータ・シミュレーションによる熱力学量と比較・検討することができる。

我々の興味は，フォールディング・シミュレーションにおいて，「天然接触相互作用」(NC) の寄与，特に，「同じ局所構造内の天然接触相互作用 (intra-N) の寄与が“どの程度か？”を知ることである．このことがわかると，「A-W_NILS モデル」の“妥当性”，及び，「A-W_NILS モデル」の“限界”を論じることができるだろう。

これまでのシミュレーションの詳細な解析から，次のような結論が得られた。

○<「A-W_NILS モデル」の妥当性と限界>

タンパク質 a1 における<— $\theta_h^{th}(T)$ 曲線>は，シミュレーションの転移曲線である，<◆ $\theta_h^{sim}(T)$ 曲線>をほぼ再現している。

その理由は、シミュレーションの詳細な解析から、タンパク質 a1 では、non-NC（非天然接触）相互作用がほぼ皆無で、NC（天然接触）相互作用が支配的であり、しかも、そのNC（天然接触）相互作用はほぼ intra-N（局所構造内）相互作用であるからである、つまり、これは、まさに「A-W_NILS モデル」の仮定がシミュレーションでも、ほぼ実現しているとみなしてよいだろう。それ故、タンパク質 a1 は、「A-W_NILS モデル」と適合するようなアミノ酸配列をもったタンパク質と言えるだろう。言い換えると、タンパク質 a1 は、そのフォールディングにおいて、相互作用が段階的に働く、つまり、順に短距離、中距離、そして長距離相互作用が働くような、理想的なタンパク質のモデルと言えるだろう。

一方、他のタンパク質 a2, b1, b2 では、「A-W_NILS モデル」による $\langle -\theta_h^h(T) \text{ 曲線} \rangle$ と、シミュレーションによる $\langle \blacklozenge \theta_h^{sim}(T) \text{ 曲線} \rangle$ とが、かなり異なっている結果が得られた（転移温度では、ほぼ一致しているが）。

その違いの理由として、シミュレーションにおいて、non-NC（非天然接触）相互作用がかなり実現しているか、あるいは、NC（天然接触）相互作用の内、inter-N（局所構造間）相互作用がかなりの寄与をしているか、が考えられる。これらのタイプの相互作用が、タンパク質のフォールディング過程で支配的の場合には、「A-W_NILS モデル」が妥当なモデルとは言えないだろう。



“Autumn Edensor II”

Coffee Break

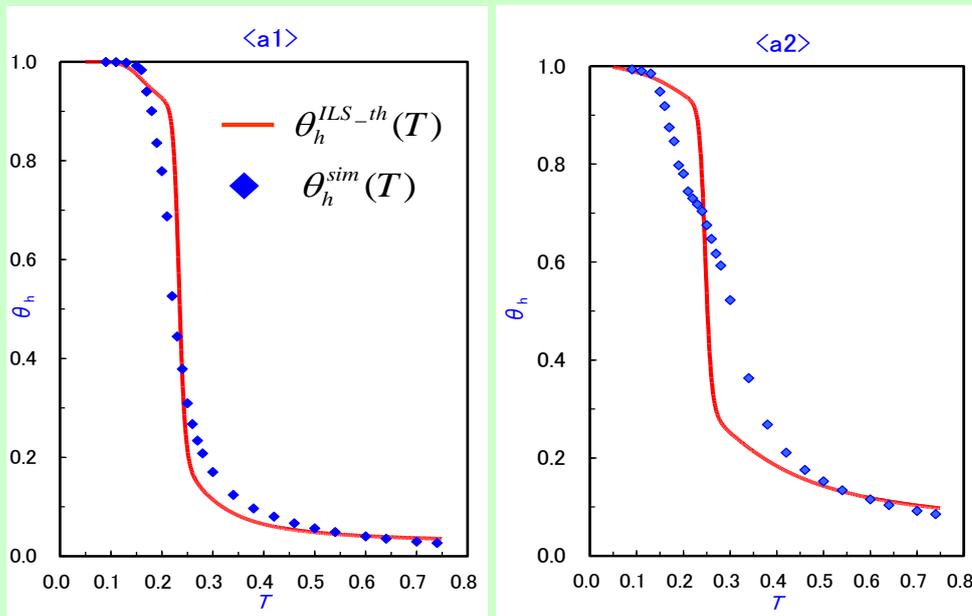
“局所構造間の相互作用”（「A-W_NILS モデル」では、考慮していない）を一部追加して、新たな転移曲線を計算し、シミュレーションの結果と比較してみよう。

図は、天然接触相互作用の内、局所構造内の相互作用と、一部の局所構造間の相互作用を考慮した転移曲線（赤色の曲線）である。

その結果、3次元格子タンパク質、 $a1$, $a2$ とも、シミュレーションによる転移曲線 $\langle \diamond \rangle$ を、より再現する結果が得られた

全ての局所構造間の相互作用を考慮して、さらに新たな転移曲線を描けば、ほぼシミュレーションの結果を再現することが期待される。

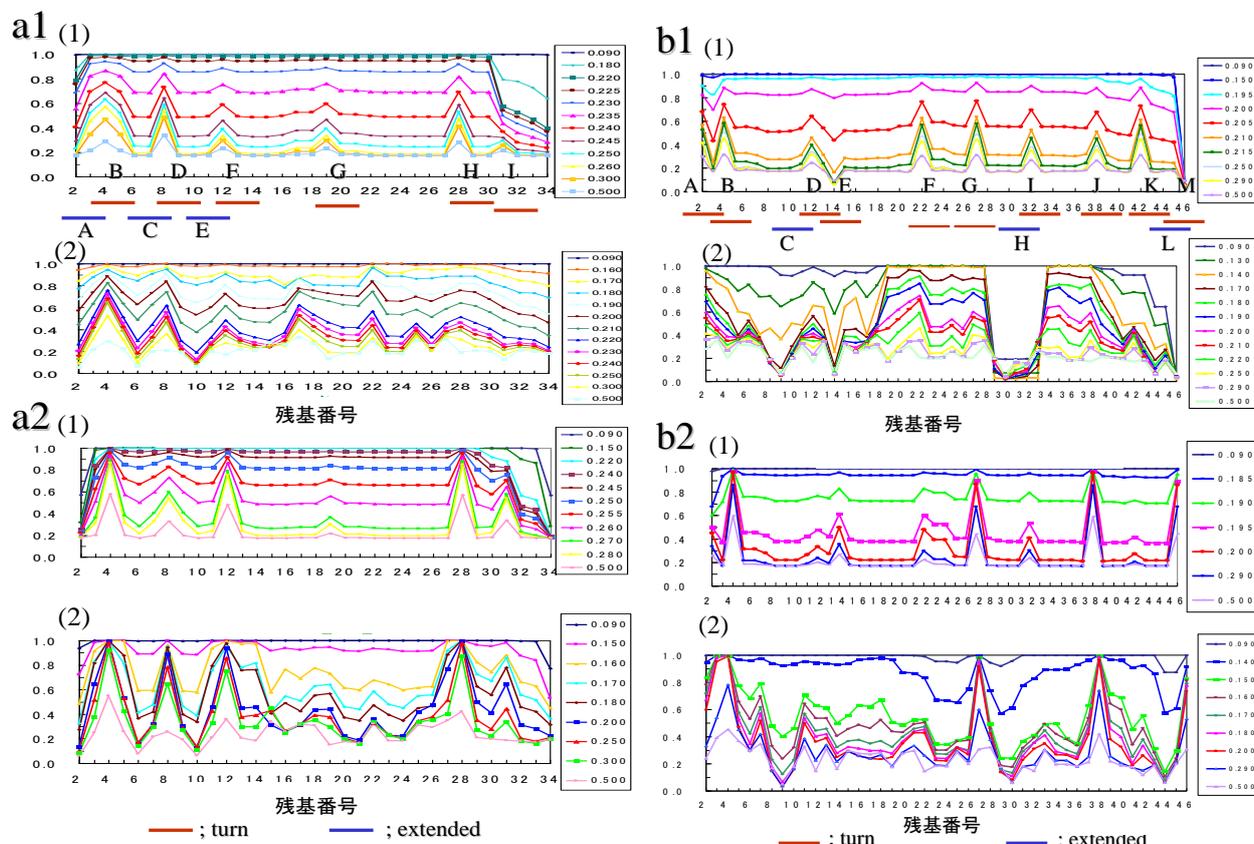
しかしながら、この修正案を実行するには、分配関数を求めるためのアルゴリズムが複雑で、多くの計算時間がかかり、実用的ではない。この修正案は、「簡単な統計力学モデル」とは言い難い。



各ユニット（各アミノ酸残基）が天然状態にある確率

各アミノ酸残基がN状態である確率 $p_i(T)$ を、いろいろな温度に対して、理論とシミュレーションの結果とを図に示した。図は、各温度において、 i -番目のアミノ酸残基が天然状態にある確率を表し、図中の(1)はモデルの結果を、(2) はシミュレーションの結果である。

理論とシミュレーションとを比べると、天然構造でターン部分にピークがある点は、タンパク質 a1, a2, b1, b2 ともよく一致している。つまり、他の領域と比較して、ターン部分は局所構造が形成されやすいことを示している。また、両末端での一致がよりよく、中央付近のアミノ酸残基ほど違いが大きくなる。



タンパク質 a1, a2, 及び、b1, b2 の違いを検証しよう。

タンパク質 a1, b1 では、ピークの高さはタンパク質 a2, b2 ほど高くはなく、各アミノ酸残基の値は温度が下がるにつれ、理論とシミュレーションともに比較的同時に上昇していく。

ところが、タンパク質 a2, b2 では、高温ですでに両末端のターン構造の形成確率がかなり高い。そして温度が下がったとき、理論では、タンパク質 a1, b1 と同様、全体のアミノ酸残基の値が一様に上昇していくのに対して、シミュレーションでは、中央付近のアミノ酸残基の値がなかなか高くない。

これは、両末端のアミノ酸残基にできた、それ自身安定な複数の小さな局所構造（この場合はターン構造）が互いに相互作用して、部分的に非天然構造を形成して、それがさらに大きな局所構造を形成するのを妨げているとみなせるであろう。

特に、タンパク質 b1 のシミュレーションの場合に、アミノ酸残基 29-33（天然構造では extended 領域）がなかなか天然構造を形成しない結果を示しているのが注目される。

その領域の両端、つまり、アミノ酸残基 22-28, 34-38 の領域は、早い段階で天然構造を形成しているにもかかわらず、アミノ酸残基 29-33 の領域は、部分的に非天然構造を形成していて、フォールド

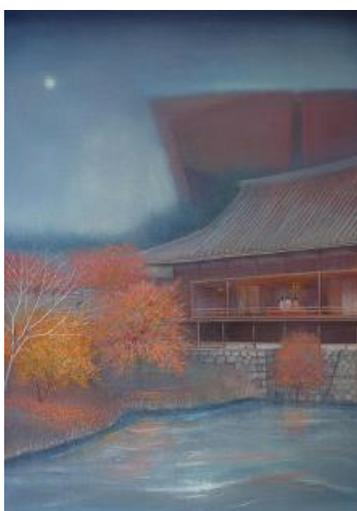
するのを妨げている。

このように、同じ天然構造へと折れたたまるアミノ酸配列でも、そのフォールディング過程のシナリオは、かなり異なっていることが示唆される。

ここまで見てきたように、アミノ酸配列の違いによって、「A-W_NILS モデル」による熱力学量と、シミュレーションによる熱力学量との相違が、かなり異なっている。しかしながら、アミノ酸配列の、どちらが、よりタンパク質らしいのか、という問いは、「A-W_NILS モデル」の物理学的評価とはまた別であり、それは生物学的な問題となる。

たとえば、あるアミノ酸配列の方が、温度を下げていくと最終的には天然構造にフォールドするものの、本来の構造とは異なる構造をとりやすいということが生物学的には不利かもしれない(もちろん、それが有利である局面もあるかもしれない)。

このとは、生物学の難しさである。ここでの議論は平衡論であるが、実際にシミュレーションにおけるフォールディングの様相を見てみると、速度論的にみても、あるアミノ酸配列の方がなかなか天然構造へと折れたたまれず、生物学的には不都合である可能性が高い。すなわち、ここで述べた「A-W_NILS モデル」に、より適合するようなアミノ酸配列が、より生物学的には好ましいということが言えるのかもしれない。



“Autumn Moon in the Garden”

◎ Appendix G

3次元格子タンパク質の コンピュータ・シミュレーションの方法

タンパク質分子の2次元格子模型

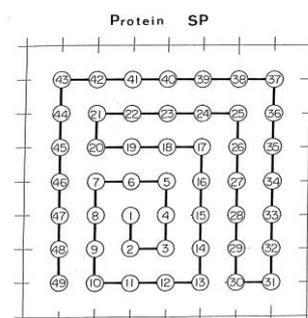
実際のタンパク質分子は非常に複雑な立体構造をしている。この複雑なタンパク質分子の本質を見抜いて、簡単な模型を構築できないだろうか？あまりに単純化することによって、タンパク質の本質を見失ってしまうような簡単な模型では意味がない。また、その模型を用いて得られる結果を現実のタンパク質へと展開するときにも気をつけなければならない。タンパク質分子らしきを残して単純化するにはどうしたらよいのだろうか？

複雑なタンパク質を思い切って簡略化した“格子タンパク質模型”を導入しよう。この模型では、タンパク質分子を構成しているアミノ酸残基を球（ユニット）とみなし、そのユニットが連結していて、折れ線からなるユニットは、格子上的みしか動くことができないとしよう。

1970 代の後半に、Gō らは、はじめて、タンパク質分子の“格子模型”を導入した。彼らは、タンパク質分子のフォールディング転移の本質が1次相転移的（協同的）であるためには、ポリペプチド鎖に沿って測って近いアミノ酸残基間に働く“短距離相互作用”と、ポリペプチド鎖に沿って測って遠いアミノ酸残基間に働く力である“長距離相互作用”の二つが重要な役割を演じていると考え、他の側面は思い切って単純化した格子タンパク質を提案した。

<文献> Taketomi, H., Ueda, Y. & Gō N., Studies on protein folding, unfolding and fluctuations by computer simulation. 1. The effect of specific amino acid sequence represented by specific inter-unit interactions, *In. J. peptide Protein Res.* 7, pp. 445-459, 1975.

図は、Gō らによって導入された、“2次元格子タンパク質模型”の例（Protein SP）を示している。7×7の碁盤の目にビーズ（ユニット数49）を配置し、ある並び方でくまなく占有するひとつつながりの鎖の構造を、2次元格子タンパク質模型の天然構造とした。



その天然構造において、格子上で隣接するユニット間（あるユニットとそのユニットの前後のユニットは共有結合とみなす）を“天然接触ペア”と定義し、その天然接触ペアに働くポテンシャルを仮定した（広く、“Gō ポテンシャル”と呼ばれている）。

2次元格子タンパク質のフォールディング・シミュレーションにおいては、この天然接触ペアが隣接格子点を占有したときのみ相互作用が働き、安定化に寄与するが、天然接触ペアでないペア（天然非接触ペア）が隣接格子点を占有しても相互作用は無視する、つまり、このときは安定化には寄与しないと仮定する。この条件で、いろいろな温度でのコンピュータ・シミュレーションを実行した。このように、Gō ポテンシャルを採用するシミュレーションを実行すると、タンパク質分子らしい協同的なフォールディング／アンフォールディングの二状態転移が観測された。さらに、この Gō ポテンシャルに加えて、天然非接触ペアにもある程度の引力が働くと仮定すると、フォールディング／アンフォールディングの協同性が次第に崩れることも観測した。

Coffee Break

「タンパク質の2次元格子模型」に関する、次のようなエピソードがある：

ある生物物理学会で、「**タンパク質の2次元格子模型**」によるコンピュータ・シミュレーションに関して発表している若い講演者に対して、学会では著名な年配の研究者が突然立ち上がって叫んだ；

『私は、長年、タンパク質分子について研究してきたが、タンパク質分子はもっと複雑だ。君が提案している、そのような玩具のような2次元格子模型で、タンパク質分子に関して何がわかるというのか？そんな2次元格子模型で、複雑なタンパク質分子のことがわかるはずはない。そんなちゃちな玩具の模型はタンパク質分子ではない！』と。

一瞬、会場は静まりかえり、皆、若い講演者を注視した。

そのとき講演者は、「2次元格子タンパク質模型」が描かれたスライドを棒で指し示しながら、一言、次のように言った：

‘**This is my protein !**’

会場は笑いにつつまれ、そして、あちこちで拍手が起こった……

タンパク質分子の格子模型の発展

1970年代から1980年代初頭にかけてGōグループは、“2次元格子タンパク質模型”に対する、タンパク質分子のフォールディング転移の一連のシミュレーションを実行しタンパク質のフォールディング転移についての多くの性質を明らかにし、それらの結果から、Gōは、1983年に「タンパク質は、進化の所産として、ほとんどの全ての相互作用が天然構造で安定になるようにできている」と洞察して“整合性原理”を提案した。

グローバルなファネル様地形を実現するために、天然構造がもっているアミノ酸残基間に働く相互作用だけを取り入れた、いわゆる“Gōポテンシャル”は、天然構造が安定になるように無理矢理バイアスをかけた、一見すると乱暴なポテンシャルにみえたが、今日では、世界中の研究者が“Gōモデル”として採用し、フォールディング・メカニズムの描像の確立に主要の役割をはたしている。

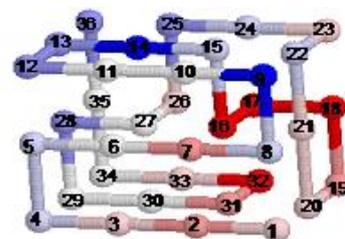
Gōらは、70年代から80年代初頭にかけて、日本で独自に、タンパク質の格子模型を導入して、タンパク質フォールディングの理論研究を展開したが、当時のコンピュータはパワー不足で、しかも、フォールディングに関する実験も未発達で、理論と実験との比較が難しかった。1990年前後になると状況は一変する。欧米の研究者、特に、Shaknovich, Dill, Karplusらによって具体的なアミノ酸配列をもった「タンパク質格子模型」によるコンピュータ・シミュレーションが徹底的に実行された。コンピュータの性能の向上により、タンパク質分子を立体格子上という限られた空間に置き、構造変化のシミュレーションを行い、それからフォールディ

ングに関する統計熱力学量を求めて議論することが容易に行えるようになったのだ。「タンパク質格子模型」の利点は、系を少数のパラメータで制御することが可能な点であり、更に、立体格子上のポリペプチド鎖の取りうる立体構造が有限個であるため、さまざまな場合について網羅的に調べることができる点である。

<コメント> 笹井理生は、著書：『蛋白質の柔らかなダイナミクス』の中で、1980年～1990年代のタンパク質格子模型による研究の現状を次のように述べている：『・・・二状態転移を説明するために導入されたコンシステンシー原理であったが、この考え方は蛋白質のフォールディング転移についてそのほかにも多くの性質を予言した。しかし、1980年前後では計算機のパワーは現在ほど強力ではなく、丸い牛のモデルだとしても、そのモデルの性質を十分明らかにするほどの計算は難しかった。また、フォールディングに関する実験も未発達で理論との比較は難しかった。すなわち、コンシステンシー原理に基づく理論の展開はあまりにも早すぎたのである。実際、郷が総説論文（N. Go, Annu. Rev. Biophys. Bioeng. 12, 183-210, 1983）を書いた後、日本の蛋白質の理論研究者はフォールディングの研究を離れ、そのころ可能になりつつあった全原子レベルの分子動力学計算の蛋白質への応用に興味を移しはじめた。・・・1990年前後に郷らの仕事の内容がアメリカの研究者によって「再発見」され、その重要性が90年代半ば以降に広く認識され・・・』（笹井理生著：『蛋白質の柔らかなダイナミクス』、培風館、2008）

タンパク質分子の3次元格子模型の導入

タンパク質分子の“3次元格子模型”を導入しよう。ここで採用する3次元格子タンパク質模型は、1990年代後半、米国のShakhnovichらによって導入され、3次元格子タンパク質模型によるコンピュータ・シミュレーションを本格的に実行した。図では、例として、アミノ酸残基数が36個の3次元格子モデルタンパク質の模型を示している。



タンパク質分子の3次元格子模型は、次のような特徴をもっている：

- (1) 3次元の立方格子に n 個のユニットからなる枝分かれのない分子鎖を考える。
- (2) 各ユニットは、20種類のアミノ酸の1つとみなし、それを球で表す。各ユニットは、それぞれ別々の格子点を占有し(排除堆積効果)、そして、アミノ酸配列上隣接するユニットは必ず隣接する格子点を占有する。
- (3) アミノ酸の配列上で隣接するとき、つまり、ユニット i, j において、 $i-j=\pm 1$ のときは、共有結合しているので、ユニット間の相互作用は考慮しなくてよい。
- (4) 2つのユニットが隣接する格子点を占有した場合、またそのときのみ、アミノ酸残基間の相互作用が発生する。そのとき、アミノ酸残基間に働く相互作用の大きさは、そのアミノ酸残基ペアの種類に依存するものとし、構造エネルギーは次式で与える：

$$E = \sum_{1 \leq i < j \leq n} U(\xi_i, \xi_j) \delta_{ij}$$

ただし、ユニット i, j が隣接する格子点を占有したとき、 $\delta_{i,j}=1$ 、ユニット i, j が隣接する格子点に無いとき、 $\delta_{i,j}=0$ である。ここに、 ξ_i, ξ_j は残基 i, j のアミノ酸の種類を表している。

また、 $U(\xi_i, \xi_j)$ はアミノ酸残基 ξ_i と ξ_j との相互作用エネルギーの値である。

$U(\xi_i, \xi_j)$ は、Miyazawa & Jerniganによって提案された次のように量である；

『立体構造既知の天然のタンパク質分子について、2種類のアミノ酸残基が接触する確率を求め、それをもとに算出された経験的なエネルギー値である。』

次の表は、彼らによって提案された、20種類のアミノ酸残基間の相互作用のエネルギー値を示し

ている。

	CYS	MET	PHE	ILE	LEU	VAL	TRP	TYR	ALA	GLY	THR	SER	GLN	ASN	GLU	ASP	HIS	ARG	LYS	PRO
CYS	-1.06	0.19	-0.23	0.16	-0.08	0.06	0.08	0.04	0.00	-0.08	0.19	-0.02	0.05	0.13	0.69	0.03	-0.19	0.24	0.71	0.00
MET	0.19	0.04	-0.42	-0.28	-0.20	-0.14	-0.67	-0.13	0.25	0.19	0.19	0.14	0.46	0.08	0.44	0.65	0.99	0.31	0.00	-0.34
PHE	-0.23	-0.42	-0.44	-0.19	-0.30	-0.22	-0.16	0.00	0.03	0.38	0.31	0.29	0.49	0.18	0.27	0.39	-0.16	0.41	0.44	0.20
ILE	0.16	-0.28	-0.19	-0.22	-0.41	-0.25	0.02	0.11	-0.22	0.25	0.14	0.21	0.36	0.53	0.35	0.59	0.49	0.42	0.36	0.25
LEU	-0.08	-0.20	-0.30	-0.41	-0.27	-0.29	-0.09	0.24	-0.01	0.23	0.20	0.25	0.26	0.30	0.43	0.67	0.16	0.35	0.19	0.42
VAL	0.06	-0.14	-0.22	-0.25	-0.29	-0.29	-0.07	0.02	-0.10	0.16	0.25	0.18	0.24	0.50	0.34	0.58	0.19	0.30	0.44	0.09
TRP	0.08	-0.67	-0.16	0.02	-0.09	-0.07	-0.12	-0.04	-0.09	0.18	0.22	0.34	0.08	0.06	0.29	0.24	-0.12	-0.16	0.22	-0.28
TYR	0.04	-0.13	0.00	0.11	0.24	0.02	-0.04	-0.06	0.09	0.14	0.13	0.09	-0.20	-0.20	-0.10	0.00	-0.34	-0.25	-0.21	-0.33
ALA	0.00	0.25	0.03	-0.22	-0.01	-0.10	-0.09	0.09	-0.13	-0.07	-0.09	-0.06	0.08	0.28	0.26	0.12	0.34	0.43	0.14	0.10
GLY	-0.08	0.19	0.38	0.25	0.23	0.16	0.18	0.14	-0.07	-0.38	-0.26	-0.16	-0.06	-0.14	0.25	-0.22	0.20	-0.04	0.11	-0.11
THR	0.19	0.19	0.31	0.14	0.20	0.25	0.22	0.13	-0.09	-0.26	0.03	-0.08	-0.14	-0.11	0.00	-0.29	-0.19	-0.35	-0.09	-0.07
SER	-0.02	0.14	0.29	0.21	0.25	0.18	0.34	0.09	-0.06	-0.16	-0.08	-0.20	-0.14	-0.14	-0.26	-0.31	-0.05	0.17	-0.13	0.01
GLN	0.05	0.46	0.49	0.36	0.26	0.24	0.08	-0.20	0.08	-0.06	-0.14	-0.14	0.29	-0.25	-0.17	-0.17	-0.02	-0.52	-0.38	-0.42
ASN	0.13	0.08	0.18	0.53	0.30	0.50	0.06	-0.20	0.28	-0.14	-0.11	-0.14	-0.25	-0.53	-0.32	-0.30	-0.24	-0.14	-0.33	-0.18
GLU	0.69	0.44	0.27	0.35	0.43	0.34	0.29	-0.10	0.26	0.25	0.00	-0.26	-0.17	-0.32	-0.03	-0.15	-0.45	-0.74	-0.97	0.10
ASP	0.03	0.65	0.39	0.59	0.67	0.58	0.24	0.00	0.12	-0.22	-0.29	-0.31	-0.17	-0.30	-0.15	0.04	-0.39	-0.72	-0.76	0.04
HIS	-0.19	0.99	-0.16	0.49	0.16	0.19	-0.12	-0.34	0.34	0.20	-0.19	-0.05	-0.02	-0.24	-0.45	-0.39	-0.29	-0.12	0.22	-0.21
ARG	0.24	0.31	0.41	0.42	0.35	0.30	-0.16	-0.25	0.43	-0.04	-0.35	0.17	-0.52	-0.14	-0.74	-0.72	-0.12	0.11	0.75	-0.38
LYS	0.71	0.00	0.44	0.36	0.19	0.44	0.22	-0.21	0.14	0.11	-0.09	-0.13	-0.38	-0.33	-0.97	-0.76	0.22	0.75	0.25	0.11
PRO	0.00	-0.34	0.20	0.25	0.42	0.09	-0.28	-0.33	0.10	-0.11	-0.07	0.01	-0.42	-0.18	-0.10	0.04	-0.21	-0.38	0.11	0.26

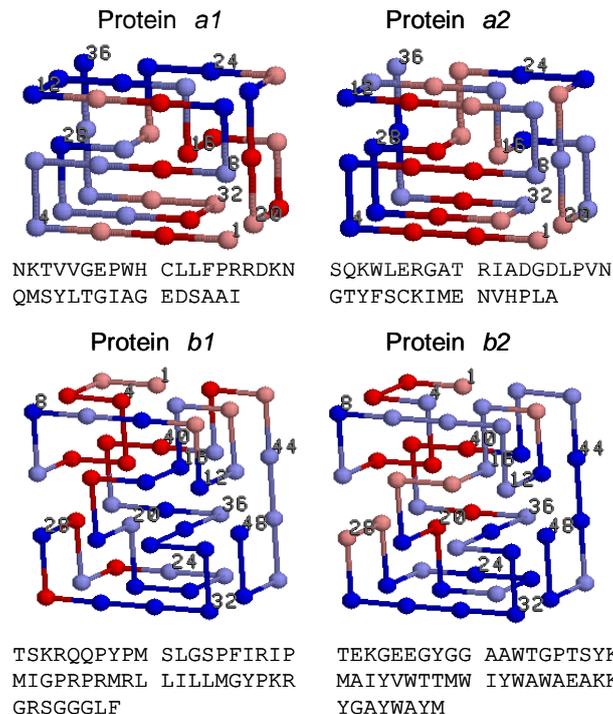
<文献> Miyazawa, S. & Jernigan, R.L. (1985) Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules*, 18, 534-552.

- (5) この分子鎖がとりうる可能な立体構造のうち、もっともエネルギーの低い構造を“天然構造”と呼ぶことにする。



コンピュータ・シミュレーションを実行する、具体的なタンパク分子の3次元格子模型

Shakhnovich らによって導入された、3次元格子タンパク質の4個の天然構造を示そう(下図): Protein a1, Protein a2, Protein b1, Protein b2 である。図の下に1文字表示のアミノ酸配列を示す。



<文献> L. Mirny, V. Abkevich and E. Shakhnovich, *Fold. Des.*, Vol.2, pp.103-116, 1996.

<文献> Abkevich V.I, Gutin AM, Shakhnovich E.I., *Fold. Des.*, Vol. 1, pp.221-230, 1996..

図中の青色系，赤色系は，それぞれ，疎水性，及び，親水基のアミノ酸残基を表している：

青色(blue)；I(Ile), L(Leu), M(Met), F(Phe), W(Trp), Y(Tyr),

空色(light blue)；A(Ala), G(Gly), P(Pro), V(Val),

ピンク(pink)；N(Asn), D(Asp), C(Cys), S(Ser), T(Thr),

赤色(red)；R(Arg), E(Glu), Q(Gln), H(His), K(Lys).

上の，3次元格子モデルタンパク質の Protein a1 と Protein a2 は，アミノ酸残基数が36個で，同じ天然構造であるが，アミノ酸配列が異なっている．更に，3次元格子タンパク質の Protein b1 と Protein b2 は，アミノ酸残基数が48個で，同じ天然構造であるが，アミノ酸配列が異なっている．

これらの，同じ天然構造をもつアミノ酸配列は，Shakhnovich らによって，次のような量を，モンテカルロ・シミュレーションを用いて，最小化することによって求められた：

$$(E_{native} - E_{average}) / \sigma$$

ここで， E_{native} は天然構造のエネルギー $E_{average}$ は，天然構造以外の種々の構造の平均エネルギー σ は標準偏差値を表す．

彼らは，上式の最適化に対して，天然構造のエネルギーが他の構造に比べて際だって低くなるようなアミノ酸配列を求めた結果，同じ天然構造になるような，アミノ酸配列が複数個あることを示した．

モンテカルロ・シミュレーションの方法とは？

3次元格子タンパク質のコンフォメーションを，格子上で，全体的に，あるいは部分的に変化させることを繰り返し行いながら，次の手順によって，Metropolis-Teller 法によるモンテカルロ・シミュレーションを行う：

- ① 適当な3次元格子タンパク質の初期コンフォメーションを決める．
- ② 試行する格子点を，乱数を用いて無作為に選ぶ．
- ③ 選ばれた格子点でのコンフォメーション変化の一つを乱数で選び，コンフォメーション変化（4種類の型を採用する．次節を参照．）を実行して試行コンフォメーションとする．
- ④ 試行コンフォメーションによるエンタルピー（エネルギー）の変化， ΔE を求める．
- ⑤ 次の場合に，試行コンフォメーションを新たに生成されたコンフォメーションとして採択する（試行コンフォメーションの採択）；
 - (イ) $\Delta E \leq 0$ ならば，排除体積効果を考慮して採択
 - (ロ) $\Delta E > 0$ ならば，0と1の間の乱数を r として，

$$e^{-\frac{\Delta E}{kT}} > r \text{ のとき，排除体積効果を考慮して採択}$$

ここで， k はボルツマン定数， T は絶対温度である．

- ⑥ (イ) (ロ) 以外のときは，試行前のコンフォメーションを保持する（試行コンフォメーションの棄却）
- ⑦ 手順②に戻る．

こうしたステップを数多く繰り返すと，生成される構造は“カノニカル分布（正準分布）”に近づくことが知られている．

上記のモンテカルロ・シミュレーションは，完全に伸びきった状態を初期コンフォメーションとして採用し， 10^8 ステップ実行した．ただし，100%近く天然構造となるような低温では，初期コンフォメーションとして天然構造を採用する．その理由は，小さな障壁でも乗り越える確率が小さくなって

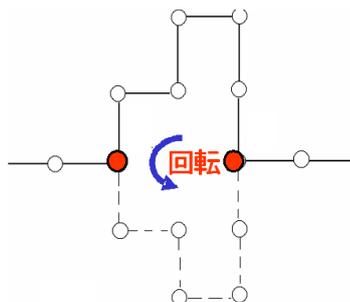
極小値に捕えられやすくなるからである。

いろいろな温度に対してステップを数多く繰り返すことによって、次々に新たなコンフォメーションを生成していく。こうして生成されたコンフォメーションの集合は、温度 T におけるカノニカル集団を構成することになる。その結果から、いろいろな統計熱力学量を求めることが可能である。

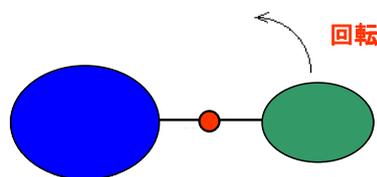
コンフォメーション変化の4種類の型

コンピュータ・シミュレーションは、3次元格子タンパク質の構造を、格子上で、全体的に、あるいは部分的に変化させることを繰り返し行いながら、モンテカルロ法によって進める。あるコンフォメーションから、次のコンフォメーションへの遷移確率はMetropolis-Teller法で定める。我々は、次のような4種類の基本的なコンフォメーションの変化を採用している。

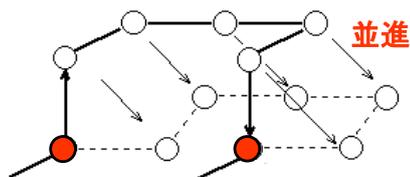
(1) 『局所的な回転』 (Local rotation)



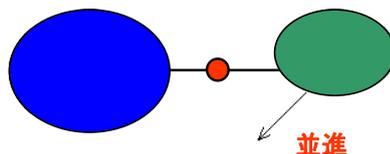
(2) 『全体的な回転』 (Global rotation)



(3) 『局所的な並進』 (Local translation)



(4) 『全体的な並進』 (Global translation)



<文献> Taketomi, H., Ueda, Y. & Gō N., Studies on protein folding, unfolding and fluctuations by computer simulation. 1. The effect of specific amino acid sequence represented by specific inter-unit interactions, *In. J. peptide Protein Res.* 7, pp. 445-459, 1975.

<文献> 安部 晴男, 小副川 博也, “三次元格子タンパク質の変性・再生のモンテカルロ・シミュレーション (1)”, 西日本工業大学紀要, 第30巻, pp. 35-44, 2000.

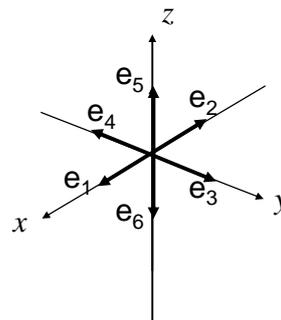
4種類のコンフォメーション変化を高速で実行するための工夫

(a) 単位ベクトルの定義

まず、3次元座標において、次のような単位ベクトル($e_1, e_2, e_3, e_4, e_5, e_6$)を導入し、それらの単位ベクトルを、簡単のために数字で表したベクトルに対応させる：

$$e_1 = (1,0,0) \rightarrow 1, e_2 = (-1,0,0) \rightarrow 2, e_3 = (0,1,0) \rightarrow 3, \\ e_4 = (0,-1,0) \rightarrow 4, e_5 = (0,0,1) \rightarrow 5, e_6 = (0,0,-1) \rightarrow 6$$

3次元格子タンパク質の立体構造は、この数字ベクトル(1,2,3,4,5,6)の並びで表現できる。



(b) 格子点を格子点に移す回転変換（保形変換）

3次元空間において、格子点を格子点に移すような回転変換は、ある座標軸を他の座標軸に移す変換（保形変換と呼ぶ）である。これをベクトルの変換でみると、次の保形変換表のように、23通りの保形変換が可能である。

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	5	5	5	5	6	6	6	6
2	2	2	2	1	1	1	1	4	4	4	4	3	3	3	3	6	6	6	6	5	5	5	5
3	4	5	6	3	4	5	6	1	2	5	6	1	2	5	6	1	2	3	4	1	2	3	4
4	3	6	5	4	3	6	5	2	1	6	5	2	1	6	5	2	1	4	3	2	1	4	3
5	6	4	3	6	5	3	4	6	5	1	2	5	6	2	1	3	4	2	1	4	3	1	2
6	5	3	4	5	6	4	3	5	6	2	1	6	5	1	2	4	3	1	2	3	4	2	1

(c) 格子点を格子点に移す並進変換

3次元空間において、格子点を格子点に移すような並進変換は、右の並進変換表のように、移動ベクトルを5通りの並進変換が可能である。

	1	2	3	4	5
1	2	3	4	5	6
2	1	3	4	5	6
3	1	2	4	5	6
4	1	2	3	5	6
5	1	2	3	4	6
6	1	2	3	4	5

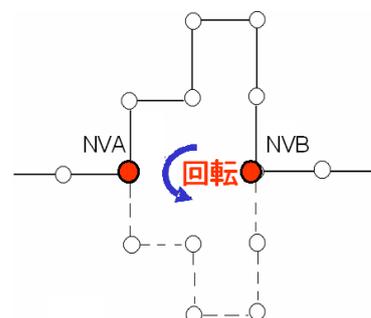
(d) 4種類のコンフォメーション変化の方法

コンピュータ・シミュレーションを実行するために、前もって3次元格子タンパク質の初期構造を設定する。次に、4種類の基本的なコンフォメーションの変化のどれを実行するか乱数を用いて決定する。コンフォメーションの変化が確定した後、元のコンフォメーションから、新たなコンフォメーションへの遷移は、Metropolis-Teller法の遷移確率に従う。ここで、4種類のコンフォメーション変化を高速で実行する方法を述べよう。

(1) 『局所的な回転』(Local rotation)

<手順1>. 回転を実行する支点となるユニットを乱数より決定する (NVA とする)。

<手順2>. 23通りの回転変換のうち、どのタイプの回転にするか



乱数で選ぶ。

<手順3> 手順2で選ばれたタイプの回転を、(NVA+1) ~ (NVA+10)のユニットに対して仮に実行する。(NVA+1) ~ (NVA+10)のユニットの内、回転を実行する前と同じ座標値が存在するときのみ(NVBとする)、“local rotation”の候補が決まる。

<手順4> 同じ座標値が存在しないときは、手順2に戻り繰り返す。

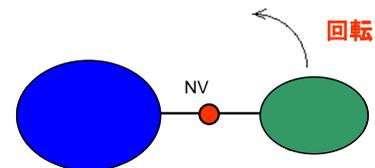
(2) 『全体的な回転』 (Global rotation)

<手順1> . 回転を実行する支点となるユニットを乱数より決定する (NVとする)。

<手順2> . 23通りの保形変換のうち、どのタイプの回転にするか乱数で選ぶ。

<手順3> 手順2で選ばれたタイプの回転を、動かすユニットの全てに対して仮に実行する。全てのユニットで重なり(同じ座標値)がなければ、“Global rotation”の候補が決まる。

<手順4> 手順3を実行した後、全てのユニットで重なり(同じ座標値)があれば、手順2に戻り繰り返す。



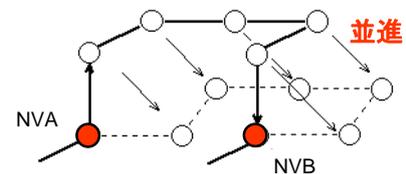
(3) 『局所的な並進』 (Local translation)

<手順1> . 並進を実行する支点となるユニットを乱数より決定する (NVAとする)。

<手順2> . 5通りの保形変換のうち、どのタイプの回転にするか乱数で選ぶ。

<手順3> 手順2で選ばれたタイプの並進を、(NVA+1)のユニットに対して並進を実行し、(NVA+2) ~ (NVA+10)のユニットの座標の並進を仮に実行する。このとき、NVAに応じた移動ベクトルの対ベクトル(例えば、移動ベクトルが1のときの対ベクトルは2)があれば、そこまでのユニット(NVBとする)に対して、“local translation”の候補が決まる。

<手順4> (NVA+2) ~ (NVA+10)のユニットまでに、対ベクトルが存在しないときは、手順2に戻り繰り返す。



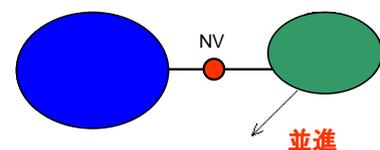
(4) 『全体的な並進』 (Global translation)

<手順1> . 回転を実行する支点となるユニットを乱数より決定する (NVとする)。

<手順2> . 5通りの並進変換のうち、どのタイプの回転にするか乱数で選ぶ。

<手順3> 手順2で選ばれたタイプの並進を、動かすユニットの全てに対して仮に実行する。全てのユニットで重なり(同じ座標値)がなければ、“Global translation”の候補が決まる。

<手順4> 手順3を実行した後、全てのユニットで重なり(同じ座標値)があれば、手順2に戻り繰り返す。

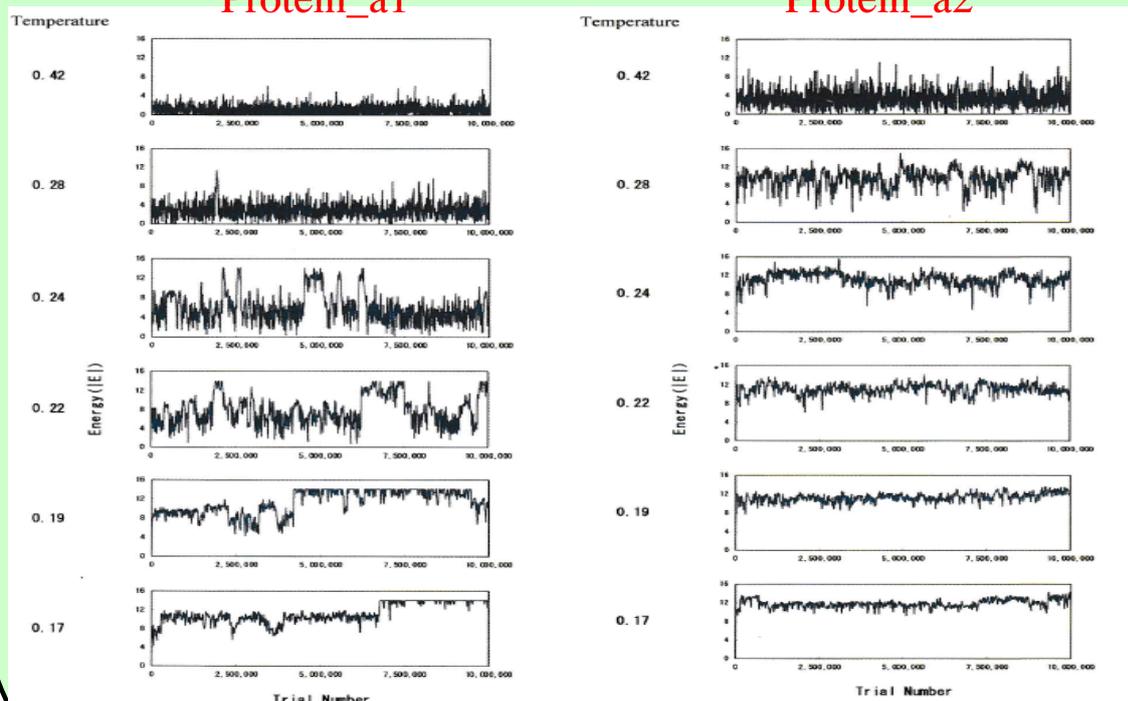


Coffee Break

図は、3次元格子タンパク質の Protein a1 と Protein a2 (天然構造は同じでアミノ酸配列が異なっている) の、それぞれについて、一連の温度 ($T = 0.17 \sim T = 0.42$) におけるコンピュータ・シミュレーションの記録の一部を例として示している。シミュレーションの初期構造は完全にほどけた構造を採用し試行回数は 10^8 回である。但し、低温では天然構造のまわりでの揺らぎだけしか起こらないことが自明であるため初期構造は天然構造を採用し、試行回数は、 10^7 回とした。図の横軸は、モンテカルロ・シミュレーションの試行回数である。縦軸は、系のエネルギーの絶対値 ($|E|$) を表し、下方はほどけた状態の、上方は天然状態のコンフォメーションに対応している。高温におけるシミュレーションでは両方ともほどけた状態で揺らいでいるが、転移温度 ($T = 0.22$) 付近では、Protein a1 の方がほどけた状態から迅速に天然構造が形成されていて、より協同性が強いことがわかる。

Protein_a1

Protein_a2



Coffee Break

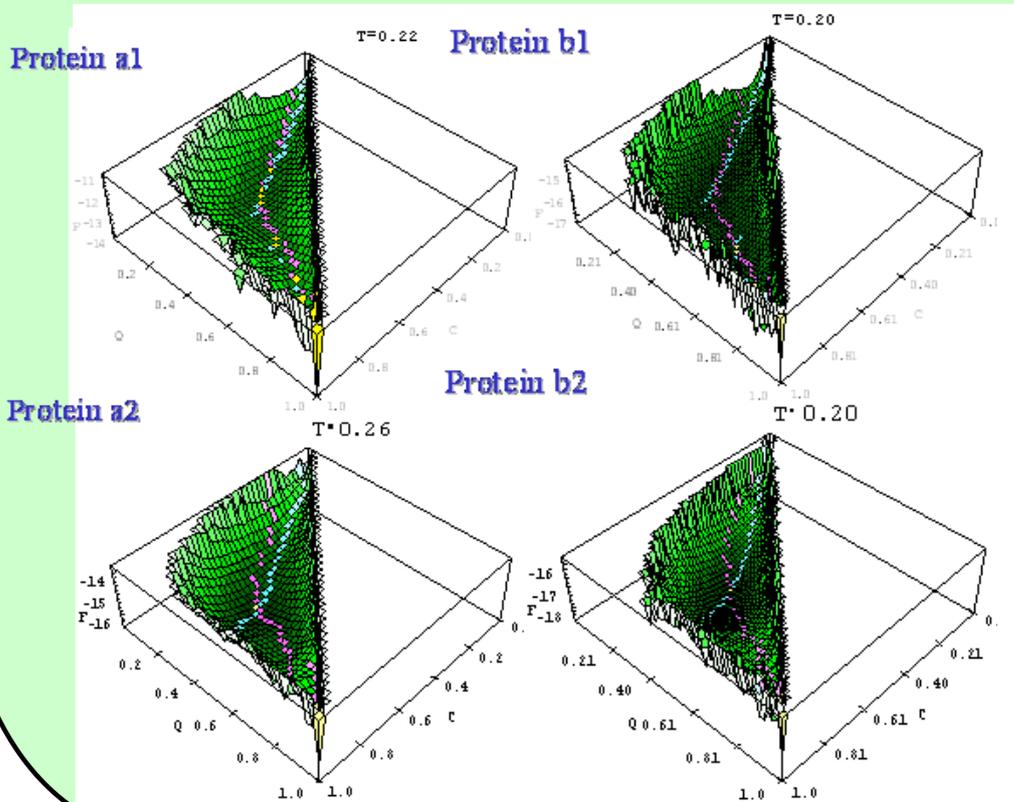
下図は、タンパク質 **a1**, **a2**, **b1**, **b2** の、それぞれの転移温度付近におけるシミュレーションの結果から、2つの秩序変数に対する自由エネルギー曲面を描いたものである： $C = N/N_{total}$ ；全ての接触ペアの割合， $Q = N_{native}/N_{total}$ ；天然接触ペアの割合ここで、 N は相互作用している全てのアミノ酸残基ペア数， N_{native} は相互作用している天然接触ペア数， N_{total} は、天然構造での全接触ペア数を表す。

手前 ($C = Q = 1.0$) に天然構造に対応する細くて深い谷があり、すぐ上が小高い山となっている。その山を越えてその上に広大なD状態（変性状態）の領域が広がっている。

D領域の様子を詳細に見るために、C軸方向での最小値とQ軸方向での最小値を図に示している。タンパク質**a1**, **b1**では両者がほぼ一致し、なだらかな谷を形成しているのがわかる。

一方、タンパク質**a2**, **b2**では、D領域の様子はかなり複雑で、C軸方向での最小値とQ軸方向での最小値は大きく異なり、その交差する付近 ($C=0.6$, $Q=0.3$) に深い窪みが存在し、しかもQ軸方向の最小値はCの値の変化に対してほぼ一定である。

このことは、この深い窪み付近では、上で述べた両末端に安定な局所構造 ($Q=0.3$ の構造) ができたものの、さまざまな、異なる (Cの値が異なる) 非天然構造をとることが可能であり、そのため、その先へフォールドできずに停留している状態であることを表している。



<文献> H. Abe and H. Wako, Analyses of simulations of three-dimensional lattice protein in comparison with a simplified statistical mechanical model of protein folding, *Physical Review*, Vol.74, pp. 011913-1, 011913-12, 2006.

<文献> 安部 晴男, 輪湖 博, “タンパク質フォールディングの統計熱力学, II. 格子タンパク質のフォールディング過程”, 西日本工業大学紀要, 第37巻, pp. 49-58, 2007.



“Bridge Across Stream in Winter”