

第4章

タンパク質のフォールディング過程の統計力学モデルの導入

<要旨>

タンパク質フォールディングの統計力学モデルである“A-W NLS モデル”の詳細を述べている。このモデルでは、『フォールディング過程におけるアミノ酸残基間の接触相互作用は局所構造内だけに働き、他の相互作用は無視する』と仮定しているが、この仮定によって、系の分配関数が簡単な漸化式によって正確に求められることを示している。

この統計力学モデルは、Wako & Saitô（輪湖・斉藤）による“Island mode（島モデル）”（1978年）、Gō & Abe（郷・安部）による“NLS モデル”（Non-Interacting Local Structure model, 1981年）の拡張版である。

“A-W NLS モデル”は、3次元格子タンパク質に特化したモデルで、格子上のポリペプチド鎖における排除体積効果を考慮して、ある長さのポリペプチド鎖が、3次元格子で取りうる全てのコンフォメーションの数を具体的に求めてエントロピー項を見積もることによって、パラメータを含まない統計力学モデルであることを強調している。



自然現象の本質をついた“理想化された、しかも簡単なモデル”とは？

一般に、物理学では、自然のいろいろな複雑な現象に対して、その現象の本質をついた、しかも、理想化された、より簡単なモデルを構築し、そのモデルを用いて求まるいろいろな物理量と、実際の自然現象によって観測される物理量とを比較・検討し、そのモデルの妥当性を吟味して自然を理解する手法が、しばしば採用されている。

簡単なモデルより導かれた物理量が、観測や実験によって求まる物理量を、ある程度再現できれば、理想化されたモデルは複雑な自然現象の本質をついていると理解できる。

簡単な物理化学のモデルの例として、“はじめに”で述べているように、理想気体モデルがある。実在気体分子には体積があり、分子間力も働いているが、次のような仮想的な気体（理想気体とみなす）を考える：『気体を構成する個々の粒子の体積は無視できるほど小さく、しかも、構成粒子間には引力が働かない。』いかなる気体も厳密には、このような理想気体ではないが、実在気体は、この理想気体からの“ずれ”とみなすことができるだろう。理想気体を仮定すれば、気体分子は独立に運動するとみなせるので、1分子に着目してその運動を解析し、また、理想気体分子全体の集団の挙動や性質を詳細に議論することができる。

実際のタンパク質分子の“フォールディング過程”も複雑であるが、その現象の本質をついた、理想化された、しかも、“簡単に計算可能な統計力学モデル”を構築できないだろうか？

その簡単に計算可能な統計力学モデルが構築できると、フォールディングに関する、いろいろな熱力学量を求めて、実験によって観測される熱力学量と比較・検討することができる。



タンパク質フォールディングの“統計力学モデル”の導入

いま、同一のアミノ酸の重合体である“ホモポリペプチド鎖”を考えよう。このようなホモポリペプチド鎖の構造転移は、“ヘリックス・コイル転移”と呼ばれている。この転移では、隣接した数残基間の相互作用である“近距離相互作用”が重要な役割を演じていて、転移中点では、1分子内の半分がヘリックス状態（折れたたみ状態）で、残りの半分がコイル状態（ほどけている状態）にある。つまり、ホモポリペプチド鎖のヘリックス・コイル転移は、“協同的な二状態転移ではない”。

一方、これから取り上げる“タンパク質分子”は、非対称性、不均一性、非周期性がその本質に関わっており、それが“情報”を生み出している高分子であるといえる。タンパク質分子のフォールディング過程における転移の中点では、集団中の半分のタンパク質分子は折れたたみ状態にあり、残りの半分のタンパク質分子はほどけている状態にある。つまり、ホモポリペプチド鎖の振る舞いと異なり、タンパク質分子のフォールディング過程の転移現象は“協同的な二状態性”を示す。これは、タンパク質のアミノ酸残基の配列に沿って近くに存在する原子間に働く“近距離相互作用”とともに、アミノ酸残基の配列で遠く離れたアミノ酸残基間の相互作用である“長距離相互作用”が重要な働きをするからである。

“タンパク質分子のフォールディング過程を記述する「統計力学モデル」を構築するにはどうすればよいだろうか？ 長距離相互作用をもち、しかも、ヘトロな要素からなるタンパク質分子のフォールディング過程を記述する統計力学モデルを構築することは、なかなか難しい問題である。

1978年、Wako & Saitô（輪湖・斉藤）は、タンパク質のフォールディング過程において重要な役割を演じる“長距離相互作用”を考慮した、合理的と思われる統計力学モデルである“Island モデル”（「島モデル」、*Island model*）をはじめて提案した。この“Island model（島モデル）”は、郷のタンパク質フォールディングの統計力学的描像とヘリックス・コイル転移モデルに、新たに長距離相互作用を導入したモデルを構築しようと意図した過程で生まれたものであった。

<文献>N. Gō, Theory of Reversible Denaturation of Globular Proteins, *Int. J. Peptide Protein Res.* Vol.7, pp. 313-323, 1975.

<文献>H. Wako and N. Saitō, Statistical Mechanical Theory of the Protein Conformation. 1. General Considerations and the Application to Homopolymers, *J. Phys. Soc. Jpn.* Vol.44, pp.1931-1938, 1978.

<文献>H. Wako and N. Saitō, Statistical Mechanical Theory of the Protein Conformation. 2. Folding Pathway for Protein, *J. Phys. Soc. Jpn.*, Vol.44, pp.1939-1945, 1978.

1981 年, Gō & Abe (郷・安部) は, Wako & Saitō (輪湖・斉藤) の “Island mode (島モデル)” を拡張した “NILS モデル” (「自由な局所構造モデル」, *Non-Interacting Local Structure model*) を提案した.

<文献>N. Gō and H. Abe, M. Mizuno, H. Taketomi, Local Structures in the Process of Protein Folding, in *Protein Folding*, R. Jaenicke ed., Elsevier, Amsterdam, 1980.

<文献>N. Gō and H. Abe, Noninteracting Local-Structure Model of Folding and Unfolding Transition in Globular Proteins. 1. Formulation, *Biopolymers*. Vol.20, pp. 991-1011, 1981.

<文献>H. Abe and N. Gō, Noninteracting Local-Structure Model of Folding and Unfolding Transition in Globular Proteins. 2. Application to Two-dimensional Lattice Proteins, *Biopolymers*. Vol.20, pp. 1013-1031, 1981.

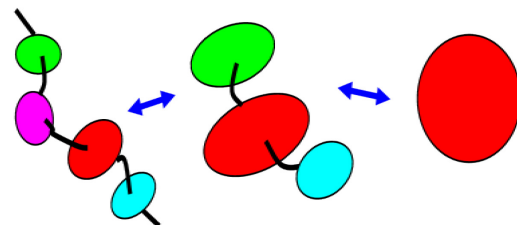
<コメント> 統計力学のもっとも基本的な話題に “磁性体の状態転移” を表現するための “Ising モデル” がある. これは, 最隣接間の相互作用のみを考慮するモデルであるが, この単純化されたモデルにより, 1 次元, 及び, 外部磁場の無い 2 次元においては厳密な解析が可能である. それより離れたペア間の相互作用も考慮すると途端に難しくなり, 平均場近似などの何らかの近似法を使う必要がある. 一般に, 長距離相互作用を考慮して解析的に解ける何らかの統計力学モデルを作ることは極めて難しい. ここで取り上げている “Island モデル” と “NILS モデル” は, そうした観点から眺めると, 3 次元の現象に長距離相互作用もモデルに組み込んで, さらに解析的に解くことのできる極めて珍しいモデルの一つといえる.

系の分配関数を計算するために, “Island モデル” では “マトリックス法” を, “NILS モデル” では簡単な “漸化式” を用いているが, これらの二つの統計力学モデルは, 基本的には同じ統計力学モデルである. そのフォールディング過程の描像は次の通りである:

『タンパク質分子は, 細胞内の生理的環境下において, ランダムコイル構造から天然構造へと自立的にフォールドする. そのフォールディング過程は, まず “近距離相互作用” によって小さな “局所構造*” が構築され, 順次, より “中距離相互作用” が働いて, これらの局所構造が成長し, 更に, “長距離の相互作用” が働いて局所構造が融合され, より大きな局所構造となる. 最終的に, 短距離相互作用や長距離相互作用などの分子内相互作用の間に互いに矛盾が無く整合的である天然構造が形成される.』(図のイラスト参照)

これらの統計力学モデルのキーポイントは, 「フォールディング過程において, アミノ酸残基間の接触相互作用は, 局所構造内だけで働く」と仮定している点である.

* 局所構造: ポリペプチド鎖の局所的な部分構造と対応する天然構造の部分構造が同じ構造 (輪湖らの “Island モデル” では, “島” と呼んでいる) をいう.



ここで, タンパク質のフォールディング過程の統計力学モデルである “NILS モデル” (“Island モデル”) の要点をまとめておこう:

- ・短距離→中距離→長距離と相互作用が順次有効となると仮定する.
- ・タンパク質分子のフォールディング過程においては, “天然構造接触ペア” のみに相互作用が働き, “非天然接触ペア” の相互作用は無視できると仮定する (この天然構造において接触しているペア

のみに働くとは仮定した相互作用は、一般に、“**Gō ポテンシャル**”と呼ばれている)。さらに、この“天然構造接触ペア”の相互作用は、“局所構造内”のみで働き、局所構造間 (図(a)), 局所構造とランダムコイル構造間 (図(b)), 及び、ランダムコイル間の相互作用は、無視できると仮定する。



<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.

<コメント>タンパク質の最大の特徴の一つは、単一の種類のユニットからなるホモポリマーではなく、それぞれがそれ固有のアミノ酸配列をもち、それによって決まる特異的な立体構造を形成することである。したがって、すべてのタンパク質を代表する平均的な立体構造というものはない。この点は、統計力学の教科書などに出てくる、均一な構成要素からなり、対称性に富んだ、あるいは周期性をもった構造を扱う問題とは本質的に異なっている。

“フォールディング過程において、天然接触ペア (天然構造で接触しているペア) のみが接触したときに相互作用する”という仮定は、天然接触ペアを指定することによってそのタンパク質の特異的な立体構造をモデルに組み込んでいると捉えることができる。

したがって、それによって得られた結果はそのタンパク質についてのみ成り立つ性質ということになる。実際、接触している残基が与えられれば立体構造をかなり正確に再現できることから、“接触している残基ペアの情報”と“タンパク質の立体構造の情報”はほぼ等価の情報といってもよいであろう。



タンパク質フォールディングの“統計力学モデル”の合理的な観点

NILS モデルで、非天然接触ペアは無視し、天然接触ペアが“局所構造内”のみで働き、他の相互作用はすべて無視するという仮定は、次のような合理的な考えからである；

『タンパク質分子のフォールディング過程において、天然接触ペアが“遠距離相互作用”するためには、つまり、アミノ酸配列上で離れた天然接触ペアが相互作用するためには、これらの二つのアミノ酸残基が空間的に近接せねばならない。このとき、両者を繋ぐ部分のポリペプチド鎖は、特定の構造をとる必要がある。アミノ酸配列上で離れた天然接触ペアどうしが、両者を繋ぐ部分のポリペプチド鎖がランダムコイル状態であるとき、両者が偶然に出会って相互作用をする確率は極めて低いと思われる。つまり、両者を繋ぐ部分のポリペプチド鎖は、部分的に天然構造と類似の構造である局所構造を形成している確率が極めて高いだろう。』

この NILS モデルで採用している仮定、つまり、『タンパク質分子のフォールディング過程では、天然接触ペアの相互作用だけが働き、その相互作用の範囲は、局所構造内だけに限定する』という仮定は、『限定的な“Gō ポテンシャル”の仮定』と見なすことができる。

NILS モデルは、天然接触ペアのうち、“長距離相互作用”をする天然接触ペアを“陽に”組み込んだモデルを定式化したものであるが、その天然接触ペアの相互作用の範囲を局所構造内だけに限定することによって、「系の分配関数が、簡単に、しかも、正確に求められる」というメリットがある。

天然接触相互作用を、局所構造内だけでなく、他の全ての場合を考慮する計算は、現在のコンピュ

ータの性能では不可能である（第 12 章 “あとがき” にかえて を参照のこと）。

そのメリットのために、『局所構造内だけに天然接触ペアの相互作用が働く』という仮定を導入したという側面もある。

ところで、タンパク質分子のフォールディング過程では、上に述べた描像の他に、次のような、“フォールディング過程で、アミノ酸残基間相互作用が階層的ではない描像” が考えられる：

『フォールディングの初期過程で、アミノ酸配列中の疎水性アミノ酸残基の部分が内部に、親水性アミノ酸残基の部分が表面に、漠然と分布した、大まかな構造がまずできあがり、つまり、NILS モデルの描像と異なり、“短距離相互作用” が働く前か、ほぼ同時に、“長距離相互作用” が働いて、大まかな構造ができ、その後、徐々に、天然構造と部分的に同じ構造を持つ“局所構造” が形成される。そして最終的に、短距離相互作用や長距離相互作用などの分子内相互作用の間に互いに矛盾が無く整合的であるような天然構造を形成する』

このような描像に対する、簡単な統計力学モデルを構築することは困難であろうか？それでも将来的にチャレンジする問題であろう！（天然接触相互作用を、局所構造内だけでなく、他の全ての場合を考慮する、新たな、拡張した統計力学モデルについて、「第 12 章 “あとがき” にかえて：フォールディング・メカニズムの統一的スキーム」に述べている。）

実際のタンパク質分子のフォールディング過程では、局所構造内だけでなく他の相互作用、及び、非天然接触ペアの相互作用も、当然働いているであろう。

NILS モデルの描像には合理性があるが、果たして妥当であろうか？理想化された、しかも簡単な統計力学モデルである NILS モデルから求められる熱力学量が、実際のタンパク質分子のフォールディング転移の実験によって求められた熱力学量の“ずれ”として把握されないだろうか？「NILS model」の妥当性は、実験との比較の中で検討されなければならない。

Coffee Break

1981年、安部・郷は、“2次元モデル格子タンパク質”のフォールディング・シミュレーションの結果から得られる熱力学量と、統計力学モデルである“NILS model”を用いて計算した熱力学量の結果とを比較し・検討した。

その結果、彼らは、“2次元モデル格子タンパク質”のフォールディング・シミュレーションにおいて、“天然接触ペア”の相互作用のみが働き、“非天然接触ペア”は無視すると仮定した、いわゆる、“Gō ポテンシャル”を採用したシミュレーションから得られる熱力学量の結果と、“NILS model”より計算された熱力学量とは、ほぼ一致することを示した。天然接触ペアは、局所構造内だけに働くとする“NILS model”の仮定は、“Gō ポテンシャル”よりかなり制限された仮定であるが、ある程度、妥当なモデルであることを示したことになる。

更に、彼らは、“Gō ポテンシャル”の他に、適当な“非天然接触ペア”の相互作用を加えて、フォールディング・シミュレーションを実行した結果から得られる熱力学量と、“NILS model”から計算された熱力学量とは、少しばかり“ずれ”が生じることも示した。

<文献> H. Abe and N. Gō, Biopolymers Vol.20, pp.1013-1031, 1981.)



統計力学モデルである“NLS モデル”を、3 次元格子タンパク質分子のフォールディングに適用してみよう！

“実際のタンパク質分子”の立体構造空間の様相を理解するために、そのすべての構成原子を考慮して、取り得る全てのコンフォメーション網羅的に計算することは、現在のところ、実質不可能であろう。

次善の策の1つとして、非常に単純化された3次元格子タンパク質を導入する。そして、3次元格子タンパク質のフォールディング過程における熱力学量を、“統計力学モデル”を構築して理論的に計算し、その結果と、コンピュータ・シミュレーションから得られる結果を比較・検討をするというアプローチを考えることにする。その後、そこで得られであろう知見をもとに、“実際のタンパク質”に適応する戦略をとろう。

3次元格子タンパク質のキーポイントは、タンパク質を構成している各アミノ酸残基を、それぞれ、一つのユニット（球）とみなし、そのユニットを立体格子点上に置いて、格子の上だけをユニットが動くとするのである。3次元格子モデルを採用する利点は、格子上の鎖のとりうる立体構造が有限個であるため、それらを様々な場合について網羅的に調べることができる点であり、また、系を少数のパラメータで制御することができ、しかも、その物理的な意味がはっきりしているため解析がしやすい点にある。

タンパク質分子のフォールディング過程に、統計力学モデルであるNLSモデルを適用するために、次のような仮定を採用して定式化しよう：

① 「アミノ酸残基をユニット（球）で表す」

タンパク質分子を構成している各アミノ酸残基を、それぞれ、一つのユニット（球）とみなす。

② 「各アミノ酸残基（ユニット）は天然状態か、変性状態かのどちらかをとる」

3次元格子タンパク質において、個々のアミノ酸残基であるユニットは、天然状態か、変性状態（ランダム・コイル状態）かの、どちらかの状態をとるものとする。

③ 「局所構造の定義」

アミノ酸配列で、部分的に連続したユニットの領域が、タンパク質の天然構造と同じ構造を取り、その両端に連結するアミノ酸残基がともに非天然構造をとっているとき、その領域を“局所構造”と定義する。局所構造内では、天然接触エネルギー*だけを考慮する。

(*) 天然接触エネルギー；天然構造において格子上で隣接しているユニットのペア（天然接触ペア）に働く相互作用のエネルギーのこと。ただし、あるユニットにつながっている両側のユニットは、格子上で隣接しているが、これらは共有結合をしているとみなして、天然接触ペアには含まない。

④ 「最小の局所構造は連続した4個のユニットからなる」

3次元格子タンパク質では、1つのユニットだけでは局所構造を定義できない。最小の局所構造は4個のユニットからなる構造とみなす。そして、 $(i-1)$ 番目ユニットから、 $(i+2)$ 番目のユニットの4残基が局所構造をとったとき、 i 番目のユニットが天然状態にあるものとする。結局、3次元格子タンパク質の天然構造では、天然状態にあるユニットの総数は、 $(n-3)$ 個である（ $i=2,3,4,\dots,n-2$ ）。

⑤ 「フォールディング過程では“局所構造内の天然接触ペア”だけを考慮し、他の天然接触ペア、及び、非天然接触ペアは全て無視する」

タンパク質のフォールディング過程でのコンフォメーションの相互作用は、“局所構造内の天然接触エネルギー”だけを考慮し、次のような天然接触エネルギー、及び、非天然接触エネルギーは全て無視する：(a) 局所構造間のエネルギー、(b) 局所構造と変性状態領域との全ての相互作用エネルギー、(c) 変性状態領域での全ての相互作用エネルギー

⑥「ポリペプチド鎖のエントロピーの定義」

排除体積効果を考慮して、鎖が取り得るすべてのコンフォメーション数を、格子タンパク質分子の鎖エントロピー(chain entropy)とみなす(詳細は後述する)。

“NILS モデル”の重要なキーポイントは、上記の⑤の仮定、つまり、『“局所構造内”での天然接触エネルギーのみを考慮し、他の、天然接触エネルギー、及び、天然非接触エネルギーは全て無視するという仮定である。』

このことを、具体的な図で説明しよう。

図は、例として、3次元格子タンパク質のフォールディング途中の、局所構造内の相互作用のみが存在しているコンフォメーションのスナップショットを示している。

点線で囲っている領域は、局所構造を表し、3角マップとの対応を表している(同色で)。局所構造内の天然接触ペアは、次の通りである(全部で4個、3角マップ上では○印)：[2-7]，[3-6]，[18-21]，[27-30]

一方、下図は、局所構造内の相互作用だけではなく、他の相互作用が存在しているコンフォメーションのスナップショットを示している。ここでは、3個の局所構造と、3角マップでの対応を表している。この場合のコンフォメーションは、“天然接触ペア”の相互作用として、次の4種類の相互作用が存在している：

(a) 局所構造内の相互作用；[7-10]，[18-21]

(b) 局所構造間の相互作用；[19-32]

(c) 局所構造とランダムコイル構造との相互作用；[2-7]

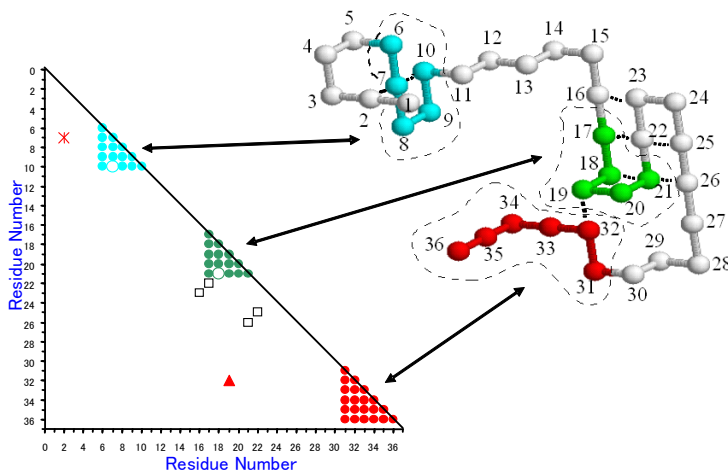
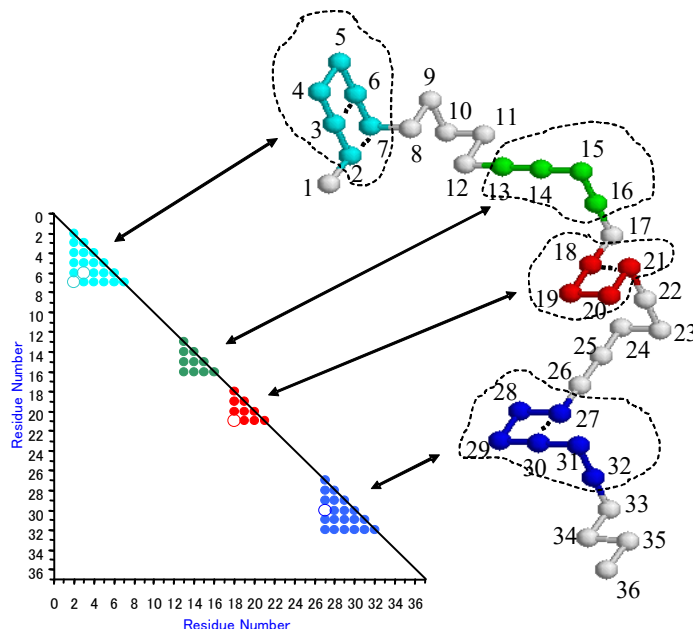
(d) ランダムコイル構造間の相互作用；図のコンフォメーションには、この相互作用は無い

このコンフォメーションの例では、次のような4個の“天然非接触ペア”(□印)の相互作用も存在している：[16-23]，[17-22]，[21-26]，[22-25]

結局、NILS モデルでは(a)の相互作用のみを考慮し、(b)～(d)の相互作用、及び、天然非接触ペアの相互作用は無視する立場をとっている。

<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.



NILS モデルでは、局所構造内の天然接触ペアの相互作用のみを考慮しているが、3次元格子タンパク質のフォールディングのコンピュータ・シミュレーションでは、(a)を含めた、他の天然接触ペアの相互作用（(b)～(d)）はもちろん、非天然接触ペアの相互作用も考慮されており、実際には働いているはずである。

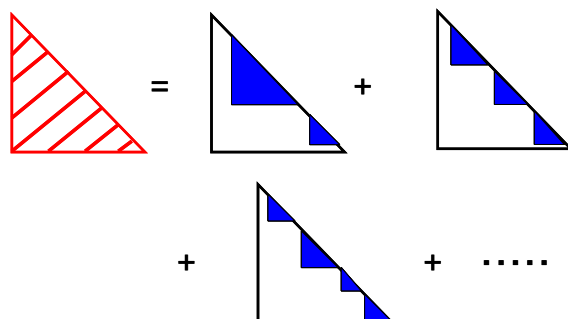
この NILS モデルにおける仮定は、果たして妥当であろうか？ NILS モデルの仮定の妥当性は、3次元格子タンパク質のフォールディング・シミュレーションによる熱力学量との比較の中で検討すべきであろう。



3次元格子タンパク質分子のフォールディングの“NILS モデル”による分配関数を求めるための漸化式

タンパク質分子のフォールディングに関する統計力学モデルである、NILS モデルにより、3次元格子タンパク質のフォールディングの分配関数を求めるには、どうすればよいだろうか？

この系での NILS モデルによる分配関数は、図のような、ポリペプチド鎖に沿って局所構造のすべての可能な配置を考慮して得られる。

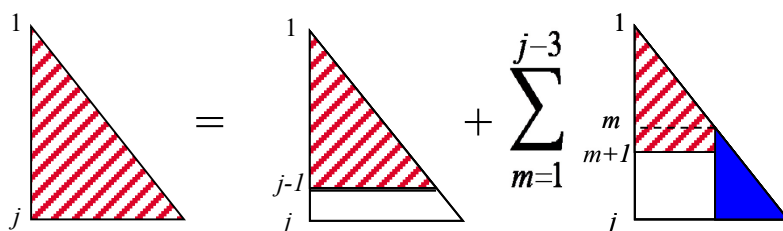


タンパク質分子フォールディングの、NILS モデルにおける分配関数を求めるための、漸化式を定式化しよう。いま、1番目～ j 番目のユニットからなる仮想的なポリペプチド鎖を考え、その分配関数を $Z_{1,j}$ とする。すでに1番目～ m 番目 ($m = 1, 2, \dots, j-1$) までの分配関数は求められているとし、次のような簡単な漸化式が成立する（最小の局所構造は、4個のユニットから成ることを考慮する）：

$$Z_{1,j} = \alpha Z_{1,j-1} + \alpha \sum_{m=1}^{j-3} e^{-\beta E(m,j)} Z_{1,m+1} \quad (j = 4, 5, 6, \dots, n-1, n)$$

ただし、 $Z_{1,1} = 0$, $Z_{1,2} = 1$, $Z_{1,3} = \alpha$, $\beta \equiv \frac{1}{k_B T}$, k_B はボルツマン定数, T は絶対温度である。

図は、上の漸化式をイラストで表している。



図の3角マップの赤色の斜線部分は分配関数を示し、青色の3角形の部分は、 m 番目～ j 番目のユニットからなる局所構造を表している。この漸化式によって求められる最終的な系全体に対する分配関

数 Z は, $Z \equiv Z_{1,n}$ となる.

ここで, 漸化式中の変数の内容を確認しておこう:

$E(m, j)$: 次式のような, m 番目 ~ j 番目のユニットから成る局所構造内のエネルギーの値である;

$$E(m, j) = \sum_{m \leq k, l \leq j} U(\xi_k, \xi_l) \Gamma_{k,l}$$

$U(\xi_k, \xi_l)$: アミノ酸残基 ξ_k と ξ_l との相互作用エネルギーの値であり, これは, アミノ酸残基のタイプ ξ_k と ξ_l に依存する. アミノ酸残基間の相互作用エネルギーの値は, 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.』

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

$\Gamma_{k,l}$: 次式で示すように, 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;

$\Gamma_{i,j} = 1$: ユニット i, j が隣接する格子点を占有したとき

$\Gamma_{i,j} = 0$: ユニット i, j が隣接する格子点に無いとき

$E(m, j)$: タンパク質分子の 3 次元立体格子模型における, 天然接触ペアの相互作用のエネルギー値より前もって求められる量である.

漸化式に表れる, 鎖のエントロピー項である α は, 連続する 4 個のユニットからなる鎖のとりうるコンフォメーションの数に対応する量である. それ故, $-k_B T \ln \alpha$ は, ランダムコイル状態にあるセグメントの鎖エントロピーを表す (詳細は次節で考察する).

タンパク分子中のアミノ酸残基が, ランダムコイル状態にあるときを基準状態 (the reference state) とみなす場合, つまり, アミノ酸残基のランダムコイル状態での統計重率を 1 とみなす場合の, NILS モデルによる分配関数を求めるための漸化式については, Appendix F: 「分配関数を求めるためのもう一つの漸化式」にまとめている.

エントロピー項を考慮した、新たな統計力学モデルである、「A-W_NILS モデル」の提案

あるコンフォメーションに対するエントロピーを、具体的には、どのようにして求めればよいのだろうか？上記に述べているように、鎖のエントロピー項である α は、連続する4個のユニットからなる鎖のとりうるコンフォメーションの数であり、一般には、 i 個のセグメントからなる鎖のとりうるコンフォメーションの数は、 α^{i-3} ($i \geq 4$) と書けるが、1つの格子点を同時に複数のアミノ酸残基が占めることはできないという“排除体積効果”を考慮すると、この表式はよい近似とはいえないだろう。従来は、この α を“調整可能なパラメーター” (adjustable parameter) として、シミュレーションや実験に合うように決められることが多かった。

我々は、この量を、次のような方法で具体的に決定し、従来の「NILS モデル」を発展させよう。3次元格子上のオリペプチド鎖における排除体積効果を考慮して、ある長さのポリペプチド鎖が、3次元格子で取りうる全ての立体構造の数をコンピュータで求めることが可能である。

実際、 $4 \leq i \leq 17$ の範囲について、とりうるすべてのコンフォメーション数を求めた（右の表参照）。ただし、ユニット数が大になると、膨大な計算時間が必要になる。

そこで、 $i \geq 18$ については、 $4 \leq i \leq 17$ の範囲で求めた数値から、次のような近似式を外挿した：

$$\alpha^{i-3} = 1.4084 \times (4.750)^{i-3}$$

我々は、“NILS モデル”を発展させたモデル、シミュレーションや実験に合うように決められる、“調整可能なパラメーター” (adjustable parameter) を一切含まない、新たなモデルを、“A-W_NILS モデル” (Abe-Wako_NILS model) と呼ぶことにしよう。

現在提案されている多くのタンパク質フォールディングの理論的なモデルには、調節パラメーター (adjustable parameter) が含まれている。我々が提案している、この統計力学モデルである、「A-W_NILS モデル」には、調節パラメーター (adjustable parameter) が含まれていないことを、改めて強調しておきたい。

i	α^{i-3}
4	7
5	31
6	148
7	706
8	3,392
9	16,166
10	77,246
11	367,079
12	1,747,257
13	8,285,115
14	39,332,272
15	186,204,654
16	882,297,753
17	4,171,744,833

<文献> H. Abe and H. Wako, Application of a Statistical Mechanical Model for Protein Folding to a Three-Dimensional Lattice Protein, *J. Phys. Soc. Jpn.* Vol.73, pp. 1143-1146, 2004.

<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, I. 定式化, 西日本工業大学紀要, 第36巻, pp. 103-109, 2006.



“Morning Light”

Coffee Break

統計力学モデルにおいて、排除体積効果を考慮した、ペリペプチド鎖のエントロピーを見積もることは、極めて困難な問題である。それ故、鎖エントロピー項は、次のようにして取り扱うのが一般的である：

『統計力学モデルに表れる変数を未知のパラメーター（調整可能なパラメーター, adjustable parameter）として導入し、それから求まる熱力学量が、実験（シミュレーション）で求めた熱力学量とできるだけ一致するように、そのパラメーターを決定する。』

実際、我々は、最初は「NILS model」において、 n 個のセグメントからなる鎖エントロピーに対して、 α^n の形を採用していた。そして、「NILS model」から得られる転移曲線と、シミュレーションから得られる転移曲線とがベストフィットするようにパラメーター α を調整しようと試みた。しかしながら、この試みはうまくいかなかった。任意の α に対して、二つの転移曲線における転移温度はどうしても近づかなかったのだ。

それ故、我々は、次のようにして、排除体積効果を考慮して鎖エントロピーを概算した：

『それぞれのポリペプチド鎖のセグメントが取り得る可能なすべてのコンフォメーションをカウントする。大きなセグメントに対しては、カウントするのに計算時間が膨大になるので、小数のセグメントで求めた数値による近似式から鎖エントロピーを概算する。』

この調整パラメーターを含まない、新たな統計力学モデルを「A-W_NILS model」(Abe-Wako_NILS model)と呼ぶことにしよう。



漸化式による具体例

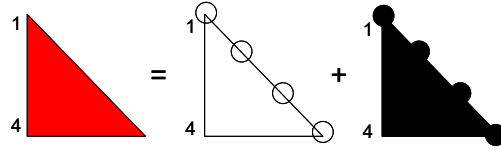
漸化式による具体例を書き下そう：

○ $j=4$ のとき；

$$\begin{aligned} Z_{1,4} &= \alpha Z_{1,3} + \alpha e^{-\beta E(1,4)} Z_{1,2} \\ &= \alpha^2 + \alpha e^{-\beta E(1,4)} \end{aligned}$$

実際に4個のユニットから成る分配関数；

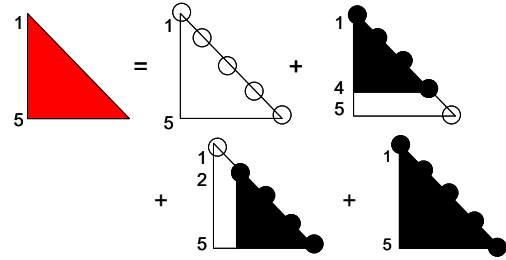
$$\rightarrow Z_{1,4} = \alpha + e^{-\beta E(1,4)}$$



	0.0	$E(1:4)$
$\alpha(\alpha^0)$	0.0	1.0
$\alpha^2(\alpha)$	1.0	0.0

○ $j=5$ のとき；

$$\begin{aligned} Z_{1,5} &= \alpha Z_{1,4} + \alpha \sum_{m=1}^2 e^{-\beta E(m,5)} Z_{1,m+1} \\ &= \alpha \{ \alpha^2 + \alpha e^{-\beta E(1,4)} \} \\ &\quad + \alpha \{ e^{-\beta E(1,5)} Z_{1,2} + e^{-\beta E(2,5)} Z_{1,3} \} \end{aligned}$$



$$= \alpha^3 + \alpha^2 \{ e^{-\beta E(1,4)} + e^{-\beta E(2,5)} \} + \alpha e^{-\beta E(1,5)}$$

実際に5個のユニットから成る分配関数；

$$\rightarrow Z_{1,5} = \alpha^2 + \alpha \{ e^{-\beta E(1,4)} + e^{-\beta E(2,5)} \} + e^{-\beta E(1,5)}$$

	0.0	$E(1:4)$	$E(2:5)$	$E(1:5)$		0.0	
$\alpha(\alpha^0)$	0.0	0.0	0.0	1.0	\rightarrow	$\alpha(\alpha^0)$	1.0
$\alpha^2(\alpha)$	0.0	1.0	1.0	0.0		$\alpha^2(\alpha)$	2.0
$\alpha^3(\alpha^2)$	1.0	0.0	0.0	0.0		$\alpha^3(\alpha^2)$	1.0

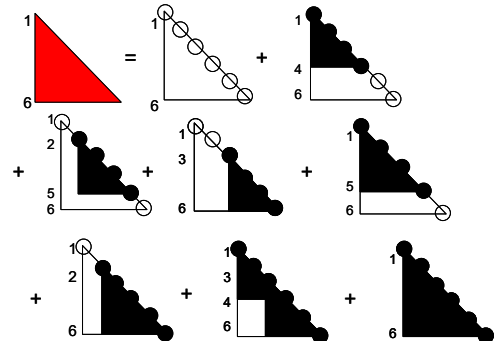
但し、 $E(1:4)=E(2:5)=E(1:5)=0.0$ のとき

○ $j=6$ のとき；

$$\begin{aligned} Z_{1,6} &= \alpha Z_{1,5} + \alpha \sum_{m=1}^3 e^{\beta E(m,6)} Z_{1,m+1} \\ &= \alpha [\alpha^3 + \alpha^2 \{ e^{-\beta E(1,4)} + e^{-\beta E(2,5)} \} + \alpha e^{-\beta E(1,5)}] \\ &\quad + \alpha [e^{-\beta E(1,6)} Z_{1,2} + e^{-\beta E(2,6)} Z_{1,3} + e^{-\beta E(3,6)} Z_{1,4}] \\ &= \alpha^4 + \alpha^3 \{ e^{-\beta E(1,4)} + e^{-\beta E(2,5)} + e^{-\beta E(3,6)} \} \\ &\quad + \alpha^2 [e^{-\beta E(1,5)} + e^{-\beta E(2,6)} + e^{-\beta \{ E(1,4)+E(3,6) \}}] \\ &\quad + \alpha e^{-\beta E(1,6)} \end{aligned}$$

実際に6個のユニットから成る分配関数；

$$\begin{aligned} \rightarrow Z_{1,6} &= \alpha^3 + \alpha^2 \{ e^{-\beta E(1,4)} + e^{-\beta E(2,5)} + e^{-\beta E(3,6)} \} \\ &\quad + \alpha [e^{-\beta E(1,5)} + e^{-\beta E(2,6)} + e^{-\beta \{ E(1,4)+E(3,6) \}}] \\ &\quad + e^{-\beta E(1,6)} \end{aligned}$$





3 次元格子タンパク質分子のフォールディングの“NILS モデル”による分配関数

前節で定式化した漸化式より，最終的に，次の式で表現される分配関数が求まる．この分配関数は，温度 T の関数であり，二つの変数， t と， u の多項式で与えられる；

$$Z = \sum_{\eta} \sum_h \Omega(\eta, h) t^{\eta} u^h$$

ここで， $u = \exp(-\beta \varepsilon_0)$ ， $\beta = \frac{1}{k_B T}$ ， k_B は，ボルツマン定数， T は，絶対温度である．

この系では，エネルギーとエンタルピーには，差が無い．それ故，コンフォメーション・エネルギーをエンタルピーとみなす．また，計算上の便宜のために，系のエネルギー（エンタルピー） E_h を次式のように表現する；

$$E_h = h \varepsilon_0 \quad (\text{ここで，} h \text{ は整数を表し，} \varepsilon_0 \text{ は } \varepsilon_0 = 0.01 \text{ である})$$

変数 t は，天然状態にあるアミノ酸残基の数 η ($0 \leq \eta \leq n-3$) をカウントするために導入した量で，最終的に 1 と置く．

係数 $\Omega(\eta, h)$ は， η と h の値で決まる全てのコンフォメーション数であり，漸化式を用いて最終的に決定される．

この分配関数は，次式のようにも書ける；：

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

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

である．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，系のエネルギー期待値 $\langle E \rangle$ を求めると，

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

また，分配関数は，次の式のようにも書ける；

$$Z = \sum_{\eta=0}^{n-3} W_2(\eta) t^{\eta}$$

ただし， $u = \exp(-\beta \varepsilon_0)$

である． $W_2(\eta) t^{\eta}$ は，天然状態にあるアミノ酸残基（ユニット）の数が η である全ての状態の統計重率の和である．

天然状態にあるアミノ酸残基（ユニット）の数が η であるときの自由エネルギーは，次式で与え

られる (パラメター t を 1 を置く) ;

$$F(\eta) = -k_{\text{B}}T \ln W_2(\eta)$$

次章で, 「A-W_NILS モデル」によって理論的に計算される, 格子タンパク質のフォールディングに関する熱力学量と, 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よって求められた熱力学量とを比較・検討し, 統計力学モデルである, 「A-W_NILS モデル」の妥当性を検討しよう.



“Two Trees in Autumn”

分配関数を求めるための もう一つの漸化式

ここでは、タンパク分子中のアミノ酸残基が、ランダムコイル状態にあるときを基準状態 (the reference state) とみなす場合、つまり、アミノ酸残基のランダムコイル状態での統計重率を 1 とみなす場合の、NLS モデルによる分配関数を求めるための漸化式を記述しよう。



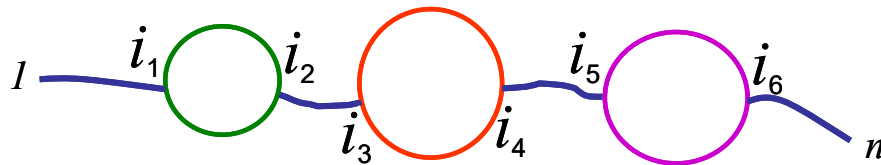
3 次元格子タンパク質分子のフォールディングの“NLS モデル”による、ランダム・コイル状態を基準状態と見なしたときの分配関数

タンパク質分子のフォールディングに関する統計力学モデルである NLS モデルにより、3 次元格子タンパク質模型の場合の分配関数を求めるには、どうすればよいだろうか？

いま、あるタンパク質のポリペプチド鎖において、アミノ酸残基の i_1 番目から、 i_2 番目までが 1 つの局所構造を形成しているとする。そして、その局所構造の自由エネルギーを $F(i_1, i_2)$ と表現しよう。いま、この局所構造がランダムコイル状態にあるときの自由エネルギーは、この中に含まれるアミノ酸残基の数 ($i_2 - i_1 - 1$) に比例していると仮定しよう。更に、そこに含まれるアミノ酸残基のすべてがランダムコイル状態にあるとき、そのような状態を基準状態 (the reference state) とみなそう。これは、ランダムコイル状態での統計重率を 1 と見なすことに等しい。

このような立場に立って、NLS モデルによる、3 次元格子タンパク質のフォールディングの分配関数を書き下そう。

まず、例として、ポリペプチド鎖がフォールディングの途中で、下図のような 3 個の局所構造からなるコンフォメーションを形成したときの分配関数への寄与を考えよう；



ここで、 n は、タンパク質分子のアミノ酸残基数である。ボルツマン定数を k_B ，絶対温度を T で表し、 $\beta = 1/k_B T$ とする。また、ランダムコイル状態にある一個のアミノ酸残基の自由エネルギーを、 $-k_B T \ln \sigma$ とすると、上の例の場合の分配関数への寄与は次式となる；

$$Z \leftarrow \sigma^{i_1-1} e^{-\beta F(i_1, i_2)} \sigma^{i_3-i_2-1} e^{-\beta F(i_3, i_4)} \sigma^{i_5-i_4-1} e^{-\beta F(i_5, i_6)} \sigma^{n-i_6} / \sigma^n$$

ここで、

$$\sigma^n = \sigma^{i_2+(i_4-i_2)+(i_6-i_4)+(n-i_6)}$$

より,

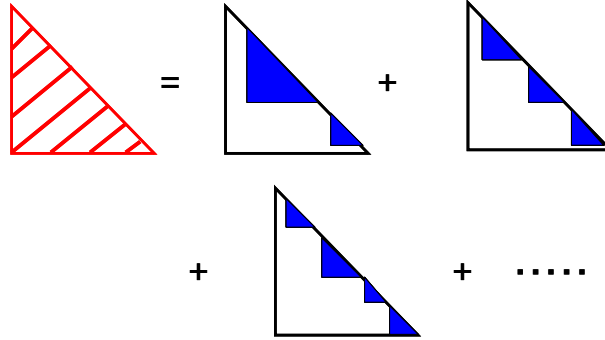
$$\begin{aligned}
 Z &\leftarrow \frac{\sigma^{i_1-1}}{\sigma^{i_2}} e^{-\beta F(i_1, i_2)} \frac{\sigma^{i_3-i_2-1}}{\sigma^{i_4-i_2}} e^{-\beta F(i_3, i_4)} \frac{\sigma^{i_5-i_4-1}}{\sigma^{i_6-i_4}} e^{-\beta F(i_5, i_6)} \frac{\sigma^{n-i_6}}{\sigma^{n-i_6}} \\
 &= (\sigma^{i_2-i_1+1})^{-1} e^{-\beta F(i_1, i_2)} (\sigma^{i_4-i_3+1})^{-1} e^{-\beta F(i_3, i_4)} (\sigma^{i_6-i_5+1})^{-1} e^{-\beta F(i_5, i_6)} \\
 &= \prod_{k=1,3,5} (\sigma^{i_{k+1}-i_k+1})^{-1} e^{-\beta F(i_k, i_{k+1})}
 \end{aligned}$$

となる. ここで, 上式において σ^n で割り算をした理由は, n 個のアミノ酸残基が全て完全にランダムコイル状態にあるとき, その統計重率を 1 とするという選択をしたことに対応している.

結局, 分配関数は, ポリペプチド鎖に沿って局所構造のすべての可能な配置にわたって統計重率を加え合わせることによって得られる:

$$Z = \sum_{\Omega} \prod_{k=1,3,5} (\sigma^{i_{k+1}-i_k+1})^{-1} e^{-\beta F(i_k, i_{k+1})}$$

ここで, Ω は, 図にイラストしているように, ポリペプチド鎖に沿って局所構造 (図の青色) のすべての可能な配置を表している.



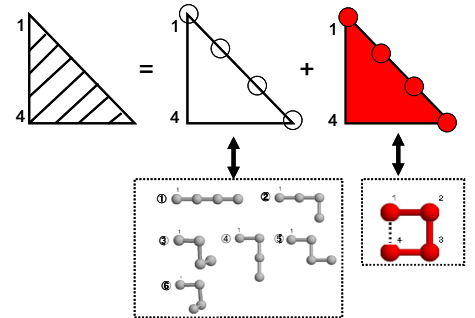
4 個のアミノ酸残基からなるポリペプチド鎖の 3 次元立体格子模型での分配関数

いま, 最も簡単な, 4 個のアミノ酸残基からなるポリペプチド鎖の 3 次元格子模型を取り上げよう. タンパク質分子の 3 次元立体格子模型の場合, 各アミノ酸残基を“ユニット”で表すことにしよう. この場合の, “最小の局所構造”は, 4 個のユニットから成る構造である. このとき, 4 個のユニットのすべてがランダムコイル状態のあるとき, そのような状態を基準状態とみなしたときの, 分配関数を書き下そう.

この系の分配関数は次式となる;

$$\begin{aligned}
 Z &= (\sigma^4 e^{-\beta E_0} + \sigma^0 e^{-\beta E_1}) / \sigma^4 = (\sigma^4 + e^{-E_1/kT}) / \sigma^4 \\
 &= 1 + \frac{e^{-\beta E_1}}{\sigma^4}
 \end{aligned}$$

ここで, $\beta \equiv \frac{1}{k_B T}$,



$E_0 = 0.0$ (4 個のユニットがランダムコイル状態にあるときのエネルギー),

E_1 は, 4 個のユニットが局所構造を形成しているときのエネルギーを表す.

上記の分配関数から、系のエネルギーの期待値 $\langle E \rangle$ を求めると、

$$\langle E \rangle = \sum_{j=1}^2 p_j E_{j_j} = \frac{E_0 \times 1 + E_1 \times \frac{e^{-\beta E_1}}{\sigma^4}}{Z} = \frac{E_1 \frac{e^{-\beta E_1}}{\sigma^4}}{1 + \frac{e^{-\beta E_1}}{\sigma^4}} = \frac{E_1 e^{-\beta E_1}}{\sigma^4 + e^{-\beta E_1}}$$

このエネルギーの期待値 $\langle E \rangle$ を規格化した次の量 θ を導入する：

$$\theta = \frac{\langle E \rangle}{E_{\min} - E_{\max}} = \frac{e^{-\beta E_1}}{\sigma^4 + e^{-\beta E_1}} = \frac{e^{-\beta E_1}}{\alpha + e^{-\beta E_1}}$$

ここで、 $E_{\min} = E_1, E_{\max} = E_0$ を用いた。 σ^4 は、4 個のユニットがランダムコイル状態にあるときの取り得るコンフォメーションの数で、 $\alpha \equiv \sigma^4$ と置換している。この場合の α の値は、 $\alpha = 6$ である。

この θ は、エネルギーの期待値 $\langle E \rangle$ を規格化した量で、温度 T の関数である。これは、フォールディングにおける、転移曲線とみなせる。



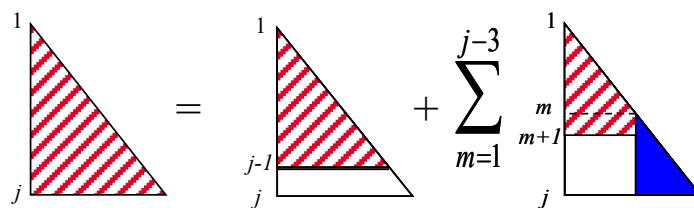
“NILS モデル” による分配関数を求めるためのもう一つの漸化式

いま、1 番目～ j 番目のユニットからなる仮想的なポリペプチド鎖を考え、その分配関数を $Z_{1,j}$ とする。最小の局所構造は、4 個のユニットから成ることを考慮すると、次のような簡単な漸化式が成立する：

$$Z_{1,j} = Z_{1,j-1} + \sum_{m=1}^{j-3} f(m, j)^{-1} e^{-\beta E(m, j)} Z_{1, m+1} \quad (j = 4, 5, 6, \dots, n-1, n)$$

ただし、 $Z_{1,1} = 0$, $Z_{1,2} = Z_{1,3} = 1$, $\beta = \frac{1}{k_B T}$, k_B はボルツマン定数, T は絶対温度である。

図は、上の漸化式をイラストで表している。



図の 3 角マップの赤色の斜線部分は分配関数を示し、青色の 3 角形の部分は、 m 番目～ j 番目のユニットからなる局所構造を表している。

この漸化式によって求められる最終的な系全体に対する分配関数 Z は、 $Z \equiv Z_{1,n}$ となる。

<文献> H. Abe, and H. Wako, Application of a Statistical Mechanical Model for Protein Folding to a Three-Dimensional Lattice Protein, *J. Phys. Soc. Jp.*, Vol. 73, PP. 1143-1146, 2004.

<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, I. 定式化”, 西日本工業大学紀要, 第36巻, pp. 103-109, 2006.

$E(m, j)$ は、次式のような、 m 番目～ j 番目のユニットから成る局所構造内のエネルギーの値である；

$$E(m, j) = \sum_{m \leq k, l \leq j} U(\xi_k, \xi_l) \Gamma_{k,l}$$

ここで、 $U(\xi_k, \xi_l)$ はアミノ酸残基 ξ_k と ξ_l との相互作用エネルギーの値であり、これはアミノ酸残基のタイプ ξ_k と ξ_l に依存する。

$\Gamma_{k,l}$ は、次式で示すように、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；

$\Gamma_{i,j} = 1$: ユニット i, j が隣接する格子点を占有したとき
 $= 0$: ユニット i, j が隣接する格子点に無いとき

$E(m, j)$ は、タンパク質分子の 3 次元立体格子模型における、天然接触ペアの相互作用のエネルギー値より前もって求められる量である。

漸化式に表れる $f(m, j)$ は、 m 番目のアミノ酸残基から、 j 番目のアミノ酸残基までのポリペプチド鎖に含まれる全ての残基がランダムコイル状態にあるときの、可能なコンフォメーション数に対応する量である。それ故、 $k_B \ln f(m, j)$ は、ランダムコイル状態にあるセグメントの鎖エントロピーを表す。また、 $k_B \ln f(m, j)^{-1}$ は、そのセグメントが局所構造を形成したときのエントロピー損失を表している。

あるコンフォメーションに対するエントロピーを、具体的には、どのようにして求めればよいのだろうか？

鎖のエントロピー項である $\alpha (= \sigma^4)$ は、連続する 4 個のユニットからなる鎖のとりうるコンフォメーション数であり、一般には、 $f(m, j) = \alpha^{i-3}$

($i = j - m + 1, i \geq 4$) と書けるが、1 つの格子点を同時に複数のアミノ酸残基が占めることはできないという排除体積効果を考慮すると、この表式はよい近似とはいえないだろう。従来は、この α を調整可能なパラメーター (adjustable parameter) として、シミュレーションや実験に合うように決められることが多かった。

我々は、この量を、次のような方法で具体的に決定し、従来の「NILS モデル」を発展させよう。そして、シミュレーションや実験に合うように決められる、調整可能なパラメーター (adjustable parameter) を一切含まない、新しい「NILS モデル」を、「A-W_NILS モデル」(Abe-Wako NILS model) と呼ぶことにしよう。

3 次元格子上のオリペプチド鎖における排除体積効果を考慮して、ある長さのポリペプチド鎖が、3 次元格子で取りうる全ての立体構造の数をコンピュータで求めることが可能である。実際、 $4 \leq i \leq 17$ の範囲について、とりうるすべてのコンフォメーション数を求めた (表参照)。

ユニット数が大になると、膨大な計算時間が必要になる。そこで、 $i \geq 18$ については、 $4 \leq i \leq 17$ の範囲で求めた数値から、次のような近似式を外挿した：

$$\alpha^{i-3} = 1.4084 \times (4.750)^{i-3}$$

$i \geq 18$ については、この式を用いることにすると、「A-W_NILS モデル」には、調節パラメーター (adjustable parameter) は含まれていないことになる。

i	$\alpha^{(i-3)}$
4	7
5	31
6	148
7	706
8	3,392
9	16,166
10	77,246
11	367,079
12	1,747,257
13	8,285,115
14	39,332,272
15	186,204,654
16	882,297,753
17	4,171,744,833



“NILS モデル” による分配関数

最終的に、分配関数（温度 T の関数）は、二つの変数、 t と、 u の多項式で与えられる；

$$Z = \sum_{\eta} \sum_h \Omega(\eta, h) t^{\eta} u^h$$

ここで、 $u = \exp(-\beta \varepsilon_0)$ ， $\beta = \frac{1}{k_B T}$ ， k_B はボルツマン定数， T は絶対温度である．

変数 t は、天然状態にあるアミノ酸残基の数 η ($0 \leq \eta \leq n-3$) をカウントするために導入した量で、最終的に 1 と置く．

係数 $\Omega(\eta, h)$ は、漸化式を用いて最終的に決定される、 η と h の値で決まる全てのコンフォメーション数である．

系の分配関数は、次式のように書ける；：

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

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 $\langle E \rangle$ を求めると、

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

また、分配関数は、次の式のようにも書ける；

$$Z = \sum_{\eta=0}^{n-3} W_2(\eta) t^{\eta}$$

ただし、

$$W_2(\eta) = \sum_h \Omega(\eta, h) u^h, \quad u = \exp(-\beta \varepsilon_0)$$

$W_2(\eta) t^{\eta}$ は、天然状態にあるアミノ酸残基（ユニット）の数が η である全ての状態の統計重率の和である．

天然状態にあるアミノ酸残基（ユニット）の数が η であるときの自由エネルギーは、次式で与えられる（パラメター t を 1 を置く）；

$$F(\eta) = -k_B T \ln W_2(\eta)$$



“Red House in Winter”