



ジャック・モノーは、その著書『偶然と必然』で、"生命の神秘はタンパク質分子のアミノ酸配列 の順序の中にある"と次のように述べている:『・・・生物の合目的構造と機能の究極の答えは、・・・ 球状タンパク質のポリペプチド鎖のなかのアミノ酸残基の配列順序のなかに埋め込まれているので ある.現実的な意味において、生命の神秘 —そういうものがあるとして— が、かくれているのは、 このレベルの化学的構成の中においてである.たんに、球状タンパク質のポリペプチド鎖のなかのア ミノ酸の配列順序を記述するだけでなく、さらにそれらがどのような法則によって折れたたまれてい るかを明瞭にすることができたら、そのときこそ、生命の神秘が暴かれ、究極の答えが発見されたと 言うことができるであろう.・・・』(J.モノー著;『偶然と必然』渡辺格・村上光彦共訳、みすず書 房、1972年)

タンパク質分子のきわだった特徴の1つは、"アンフィンゼン・ドグマ"と言われているように、"生 理的な条件下で、そのアミノ酸配列によって決まる特異的な立体構造を自発的に形成する"ことであ る.タンパク質は固有の立体構造を形成することによって特定の基質分子だけを認識し、特有の機能 を発揮する.タンパク質分子が、そのアミノ酸配列で決まるそれぞれに固有の立体構造へフォールド する機構とはどのようなものなのだろうか? 特定のタンパク質のアミノ酸配列が与えられたとき、 その天然立体構造を予測するという問題は、非常に難問で、今日、依然として未解決の問題として残 されている.この問題に関しては、近年、人工知能による高精度の予測が可能になりつつある.しか し、人工知能による予測は、その推論過程がわからないため、アミノ酸配列から天然立体構造が決ま る論理構造については依然として謎のままである.

タンパク質のフォールディング過程の特質を明らかにする、つまり、フォールディング・メカニズムを明瞭にする問題もまだ明確な解答が得られていないのが現状である. 我々は、タンパク質のフォールディング・メカニズムを、統計力学的立場から解明したいと思っている.

ところで、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、理想化 された、しかも、より簡単なモデルを構築し、そのモデルを用いて理論的に求まる物理量と、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.理想化されたモデルで、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、この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.

例えば、理想気体モデルがある.実在の気体分子には体積があり分子間力も働いているが、これを、 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 と仮定すれば、つまり、理想気体とみなすと、気体分子は独立に運動するとみなせるので1分子に着 目してその運動を追跡したり、理想気体分子全体の集団の挙動や性質を統計力学を用いて詳細に解析 することができる.いかなる気体も厳密には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.

大沢文夫は,彼の著書:『大沢流手づくり統計力学』で、"タンパク質分子1個でも統計力学が構築 できる"と次のように述べている:『非常に多くの分子が集まった系を巨視的に扱うのが熱力学で、 その中の1つ1つの分子の状況を扱いながら、それらの集まった系の性質を理解しようとするのが統 計力学です.大学の一般の統計力学の講義では「非常に多くの分子の集まりでは・・・」とやってい ます.しかし、思い切り少ない、わずかな個数の、それこそ数個の集まりでやってみても、統計力学 の本質はわかるんです.・・・わずかな個数の分子、分子というか対象で、結構ちゃんと統計力学の 本質がわかってしまう.・・・1個の分子機械の1回の動作で、自由エネルギー変換をするということ が見つかりつつあります.・・・こういう少数のシステムが自由エネルギーの変換をどうやっている かという統計力学を改めてつくらなくてはいけないんです.・・・』(大沢文夫著;『大沢流手づくり 統計力学』、名古屋大学出版会、2011) 永山國昭は、2001 年の生物物理学会誌の中で、"タンパク質分子の統計力学モデルの困難さ"を次 のように述べている:『タンパク質の統計力学モデルの困難さは、1 個のタンパク質のとりえる立体構 造の数の膨大さにある.例えば、タンパク質のアミノ酸残基数が 100 個のとき、可能なコンフォメー ションの数は 10⁹⁵ となる.すると、1 フェムト(10⁻¹⁵)秒に1回この立体構造を探索するとして、 すべてを探索(悉皆探索)して安定構造を見出すには 10⁹⁵/10¹⁵ =10⁶⁰宇宙年かかる.・・・このよう な可能な立体構造の膨大さが、現在、タンパク質の立体構造予測を困難にし、統計力学的扱いを難し くしている.』 さらに彼は "タンパク質分子の格子模型について" 次のように述べている:『この構 造数(微視的状態数)を激減させるものとして、格子モデルが提案された.格子モデル自体は歴史が 古く、2 次元格子モデルを用いて熱力学量を計算した郷らの研究が最初である.・・・この格子モデル は、現在きわめて広い範囲で応用されている.・・・Gō like model として評価されているのは、格子モ デルだけではなく、"相互作用を天然状態原子間コンタクトに絞る"という卓見であった.・・・』(永 山國昭:『タンパク質フォールディングの物理』生物物理 41(4),196-200, 2001).

タンパク質のフォールディング・メカニズムの解明は、険阻な絶壁をもつ難攻不落の高山であろう と思われる.一方,この高山への登頂過程の道筋には、いたるところに美しい花が咲き乱れる花園が 開けているかもしれない.タンパク質フォールディングの統計力学的解明は、この本の最後の頁から、 真の問題が展開すると思っている.この分野を解明しょうとする人々にとって、我々の提案が、何ら かの手引きに少しでもなるなら、著者らにとって、それは望外の幸せであろう.

余白の全ての絵画は,英国の画家 "Thomas Lamb" 氏の作品である. 読者の心を少しでも和らげる ために,氏の同意を得て挿入した. ここに,厚く感謝の意を表したい.

> 2022年5月 安部晴男,輪湖博



"Path beside the Sea study"



はじめに・・・ii

<内容> "生命の神秘はタンパク質分子のアミノ酸配列の順序の中にある"というジャック・モノーの言葉を引用している.そして、タンパク質のフォールディング問題である『どのような機構で、タンパク質分子はそのアミノ酸配列の情報だけで固有の立体構造を形成するのか?』という未解決の問題の解明のアプローチとして、『タンパク質フォールディングの統計力学』や、『タンパク質分子の格子模型による研究』が重要ではないかと述べている.

第1章 タンパク質のフォールディングの研究は何故重要か?・・・1

Appendix A:分子生物学誕生の記念碑的論文・・・7

Appendix B: 遺伝コード表・・・11

Appendix C: 20 種類のアミノ酸・・・12

<内容> 分子生物学が誕生した記念碑的論文(ワトソン・クリック,1953年)を紹介している. 更に,遺伝情報の流れに言及した『セントラル・ドグマ』について図示し,その流れの中の"第2 次遺伝情報解読問題",つまり、"タンパク質のフォールディング問題"は未解決の問題であるこ とを述べている.特に、「フォールディング異常病」の研究や、「免疫系での膨大な多様性を生み 出す機構」の研究などを紹介し、タンパク質フォールディングの研究の重要性を述べている.

第2章 タンパク質分子のフォールディング問題とは?・・・13

<内容> "タンパク質分子のフォールディング問題"には3つの側面があることを紹介している. そして、この本の主たるテーマは、"タンパク質分子のフォールディング・メカニズムの問題"を「統 計力学的観点」から解明することであると強調している.さらに、「整合性原理」や「極小フラス トレーションの原理」について述べ、この原理の展開からから導かれた、フォールディングにおけ る "ファネル描像"や、タンパク質フォールディングにおける "天然接触相互作用"の重要性を指 摘している.

第3章 タンパク質分子はボルツマン分布など聞いたこともない!・・・25

Appendix D: ボルツマン分布の導出・・・32

Appendix E: ラグランジュの未定乗数法とは?・・・35

<内容> 平衡状態で、タンパク質分子のコンフォメーションが、あるエネルギーをもつ確率は"ボルツマン分布"で記述されることを、ラグランジュの未定乗数法を用いて導いている.この導出過程で表れる、系の"分配関数"(状態和)から、いろいろな熱力学量を理論的に求めることができることを示している.その際、最も簡単な例として、4個のアミノ酸残基からなる3次元格子模型を用いている.

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・・・37

Appendix F:分配関数を求めるためのもう一つの漸化式・・・51

<内容> 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_NILS 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いては、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.

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!・・・57

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・・72 <内容> 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。更に、いろ いろな温度における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、A-W_NILS モデルの妥当性を詳細に検討している。

第6章 アミノ酸置換によって熱力学量はどのように変化するのか?・・・83

<内容> 3次元格子タンパク質の天然構造の各アミン酸残基を他の19種類のアミノ酸残基に置換することによって生じる、2つの熱力学量(転移温度、エネルギー)の変化は強い相関をもっていることを示している.さらに、反応座標の各点で、各々のアミノ酸残基がどの程度天然状態であるかを求めて、フォールディングの統計的道筋を議論し、フォールディング過程が協同的な現象であることを示している.

第7章 フォールディングのキネティクスにおけるレートと緩和時間の関係は?・・93 Appendix H:フォールディング・レートの計算方法・・・108

<内容> 簡単な2状態系の構造変化を表す速度定数(レート)と緩和時間との関係を求めている. 次に、多状態系をもつ系のキネティクスを記述する「マスター方程式」より、フォールディング・ レート と、アンフォールディング・レート を求める方法を定式化している.そして、シェブロ ン・プロットを描き、その相違から、新たな、コンタクト・オーダー を考慮すべきであると提案 している.

第8章 タンパク質フォールディングの遷移状態の構造を推定するΦ値解析とは?・・・123 <内容>タンパク質フォールディングにおける遷移状態を考察し、この遷移状態での構造的特徴を 推定できる、実験による"Φ値解析法"を紹介している.一方、理論的にΦ値を求める式を導き、3 次元格子タンパク質の遷移状態でのフォールディング核を推定している.その結果、同じ天然構造を もつが、アミノ酸配列が異なるタンパク質では、それぞれ、異なるセグメント領域のフォールディン グ核をもち、異なるフォールディング過程であることを示している.

第9章 統計力学モデルを用いたΦ値計算とフォールディング・メカニズム・・・135

Appendix I: 実験におけるΦ値解析とは? ・・・<u>147</u>

<内容>実験による"Φ値解析"の手法を紹介している.更に、「A-W_NILS モデル」を用いて、"実際のタンパク質"の各アミノ酸残基のΦ値を理論的に計算する方法を示し、具体的に、2 つのタンパク質、プロテインA (1SS1) とキモトリプシンインヒビター2 (3CI2) に適用している.そして、1SS1 に対するΦ値の分布から、フォールディング・メカニズムは、実験では、"核形成-凝縮モデル"と推論しているが、理論では、"フレームワーク モデル"が妥当であると推論している.3CI2の実験によるΦ値の分布から、フォールディング・メカニズムは、"核形成-凝縮モデル"であると推論し、このタンパク質に対する「A-W_NILS モデル」の限界を述べている.

第10章 理論と実験によるΦ値の相関の検証_その1・・・151

<内容> 27 個の"実際のタンパク質"に対して、"A-W_NILS モデル"を適用して計算したΦ値と、 実験的に観測されたΦ値との相関係数の値を表 (Table 1) で示している. そして、相関係数の値を 3 つのグループに分類し、この分類は、フォールディング・メカニズムのモデルである、"フレームワ ーク モデル"、及び、"核形成-凝縮モデル"のいずれかに対応していると推定している. 更に、 具体的に、天然構造が類似の、SH3 ドメインファミリーに属する4個のタンパク質 1SRM, 1SHG, 1FYN, 1BF4, 及び, 2個のタンパク質, 1PGB, 2PTL を取り上げ、Φ値に対する、実験と理論の結果を詳細に比較している. その結果、フォールディング・メカニズムは、天然構造のトポロジーが本質的だが、アミノ酸配列の情報も重要であると指摘している.

第11章 理論と実験によるΦ値の相関の検証_その2・・・165

<内容>天然構造が類似のタンパク質 1ENH, 1IDY と, タンパク質 3CHY, 1RNB, 1FKB, 及び, 1RIS の理論と実験によるΦ値を比較している. その結果, 1ENH, 1IDY, 3CHY のフォールディン グ・メカニズムは"フレームワーク モデル"と推定している. 一方, 1RNB の場合は, "フレーム ワーク モデル"か, "核形成-凝縮モデル"のどちらかであり, 1FKB, 1RIS の場合は, "核形成-凝縮モデル"であると推定している. しかしながら, 1RIS の 3 個の円順列変異体の場合は, "フレ ーム ワーク モデル"の方が妥当であることを示している.

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』・・・179

<内容>採用されている3次元格子タンパク質のアミノ酸配列を用いると、フォールディング・ シミュレーションで、ランダムコイル状態から天然構造へフォールドする.その際、天然接触ペ アの情報を前もって陽に与えていない.このことは、「このアミノ酸配列の中には、"天然接触ペア に関する情報"が"暗に"隠されているのではないか」、そして、「"Φ値解析"によるフォールディン グ核を推定できることがその傍証ではないか」と述べている.

『"実際のタンパク質"のアミノ酸配列には、進化の過程で獲得した天然構造における"天然接触 相互作用"に関する情報が含まれていて、その情報にしたがって、ランダムコイル状態から天然構 造へフォールドすることができる.』という"フォールディングの統一的スキーム"を提案している. 現在において個々のタンパク質のフォールディング・メカニズムを理論的に解明するためには、 天然構造が既知のタンパク質分子の、"全ての天然接触相互作用"を考慮した場合の統計力学モデル を構築して系の分配関数を計算し、その分配関数から、いろいろな熱力学量を求める必要があるだ ろうと指摘している.

著者紹介・・・187 索引・・・189



" Autumn Edensor "



く要旨>

ワトソンとクリックによる「分子生物学の誕生 の記念碑的論文」(1953, Nature)を紹介している. 更に,フランシス・クリックが提唱した「生物の 遺伝情報の流れに関するセントラル・ドグマ」を 図示し,その流れの中の"第2種遺伝情報解読問題", つまり, "タンパク質のフォールディング問題" は未解決の問題であることを述べている. タンパク質のフォールディング研究が何故重要

かの例として、二つの例を取り上げている:① 「フォールディング異常病」、つまり、「どちら も同じアミノ酸配列であるにもかかわらず、どの ようなメカニズムで、正常プリオンタンパク質が 異常プリオンタンパク質に変化するのか?」② 「免疫系では、部品のランダムな組み合わせで膨 大な多様性を生み出す機構をもつが、タンパク質 が特定の立体構造をとるために満たさなければな らないアミノ酸配列の条件とは何か?」

↓ 分子生物学の誕生

19世紀の生物科学における最大の発見といわれる DNA*の"二重らせん構造"の発表が 1953 年, 英国ケンブリッジ大学の若き生物学者ジェームズ・ワトソンと物理学者フランシス・クリックによっ てなされた. "親のもっている遺伝情報が, どのような機構で子に正確に伝えられるのか"という問い に対する答えが与えられたのである.分子生物学の誕生である.

彼らは,1953年にNature 誌に掲載されたわずか1ページの論文**の冒頭で,次のように述べている: 『デオキシリボ核酸 (DNA)の塩の構造を提案したい.この構造は生物学的にかなり興味深い新しい 特徴を持つ.』そして,次のような,画期的な構造を提案している:

『・・・特定の塩基対同士のみが結合することができる. これらの組み合わせは,アデニンとチミン, グアニンとシトシンである.』



論文の最後は、次のような言葉で締めくくっている:『われわれの主張する特定のペアリング(塩基対) が遺伝物質の複製機構を直ちに示唆することには誰でもが気付くだろう.』

* DNA: Deoxyribose Nucleic Acid (デオキシリボ核)

** Appendix A:「分子生物学誕生の記念碑的論文」(参照)

彼らの論文により, DNAは, ヌクレオチド***を基本単位として, これが 多数重合してできた高分子であり, ヌクレオチドの塩基部分は次のような 4 種類の塩基のどれかであることが確認された:

アデニン (A), グアニン (G), シトシン (C), チミン (T) ただし, RNA [Ribonucleic acid, リボ核酸] では, 次のようになる: チミン (T) →ウラシル (U)

この4種類の塩基には、塩基どうしに相補性があり、AとTとで二本の 水素結合を、GとCとで三本の水素結合を形成して、全体として二重らせ ん構造を構築できる化学的構造になっていることも確認された.

*** ヌクレオチド;骨格部分はリン酸と糖からなる



💶 生物におけるセントラル・ドグマ

フランシス・クリックは、1958年に、生物の遺伝情報の流れに関する次のような"セントラル・ド グマ"(中心教義)を提唱した;

『情報は、つねに、「 DNA → RNA→タンパク質 」の方向に流れ、タンパク質から核酸に向けて、 逆に流れることはない. 』

ところが、1970年、レトロウイルスから、遺伝情報の逆向きの流れの例が発見された. mRNA(メ ッセンジャーRNA)を鋳型として DNA を合成する逆転写酵素の発見である. それでも、大筋では、 セントラル・ドグマが成立していると見なして良いだろう.



タンパク質の特異的立体構造

生物のセントラル・ドグマ

第1章 タンパク質のフォールディングの研究は何故重要か?

🎩 『第1次遺伝情報解読問題』は完全に解読されている!

DNA から転写された mRNA の塩基配列が指定するアミノ酸配列をもったタンパク質を合成する過程を翻訳という. 翻訳の過程で, mRNA の連続する3個の塩基をコドン*というが, これらのコドンが, 20種類のアミノ酸のどれに対応するかは, 『第1次遺伝情報解読問題』と呼ばれた.

1960年代に, mRNA の塩基配列とタンパク質のアミノ酸配列との対応関係が完全に解読された.生物に共通である"遺伝コード表**"としてまとめられている(ごくわずかの例外がある).

* コドン;4 種類の塩基を3 個並べる場合の数は,4×4×4=64 通り

** Appendix B:「遺伝コード表」(参照)

DNA の塩基配列は4種類の文字で書かれた文章であるとみなせる. ヒトの DNA は約 30 億塩基対からなる. この生命の設計図である全遺伝情報をゲノムといい,そのうち,DNA上の各タンパク質に対応する領域を一般に"遺伝子"と呼ぶ. DNA の塩基配列がもつ遺伝情報には、タンパク質が必要な場所で、必要な時に、必要な量を合成するように調節するための塩基配列や、tRNA(トランスファーRNA)の塩基配列などの情報も含まれている.

特定のタンパク質のアミノ酸配列が与えられたとき、その天然状態での特異的立体構造を予測する 問題は、『第2次遺伝情報解読問題』、または、『フォールディング問題』と呼ばれているが、これは未 解決の問題である.

タンパク質分子のきわだった特徴の1つは、生理的な条件下で、そのアミノ酸配列によって決まる "特異的な立体構造"(タンパク質特有の天然構造)を自発的に形成することである.タンパク質分子 が、どのようなメカニズムで、そのアミノ酸配列で決まるそれぞれに固有の天然構造へフォールドす るのだろうか?

この「タンパク質フォールディング問題」は、次のような3つの側面がある:

天然構造の予測の問題;

特定のタンパク質のアミノ酸配列が与えられたとき,その天然状態での特異的立体構造を予測す る問題

② フォールディング・メカニズムの問題;

タンパク質のフォールディング過程の性質である反応レートや,変性状態・中間状態・遷移状態 の構造を推定して,フォールディング・メカニズムを明らかにする問題.

③ 天然構造の特徴の問題;

「タンパク質の天然構造は,静的にも,動的にも,どのように特徴づけられるのか?」を明らか にする問題

これらの問題に対して、いろいろと非常に興味深い提案がなされている.しかしながら、明確な解 答が得られていない.特に、①「天然構造の予測の問題」、②「フォールディング・メカニズムの問題」 は、依然として、未解決の難問として残されている.

我々の目標は,第2の問題「フォールディング・メカニズムの問題」を統計力学的観点から解明することである.

□ タンパク質分子は複雑で高度に組織された生命現象を演じている!

生命現象は,遺伝情報担体物質としての DNA と,その情報の直接的な発現分子であるタンパク質 分子を基盤とした高度に組織化された化学反応系であるといえる.

この生命現象の主役ともいえるタンパク質分子は、長い生命進化の過程で、"20 種類のアミノ酸" (Appendix C:「20 種類のアミノ酸」参照)を1次元的なアミノ酸配列とすることによって天然構 造を生成し、様々な機能をもつ高分子として生み出されてきた.タンパク質の活動の場は、生命現象 のいたるところに及んでいる. DNA の塩基配列というたった4文字で構成された文字列の翻訳であ るアミノ酸配列という1次元情報が、如何にして立体構造という3次元情報へと変換されて様々な機 能をもち得るのであろうか?

タンパク質のフォールディング研究の重要性を示す幾つかの例を示そう.

💶 タンパク質のフォールディング異常病

タンパク質のフォールディング異常病の発見は衝撃的であった. 感染症である結核や赤痢などは細菌が原因であり、また、エイズや鳥インスルエンザなどはウイルス が原因である. 細菌やウイルスはいずれも固有の遺伝情報をもち、核酸(DNA, RNA など)を有 しているが、これらの細菌やウイルスではなく、異常になったタンパク質が感染症の原因物質である という全く新しい概念の病気の発見が米国のプルミナー博士(1997年、ノーベル医学生理学賞を受賞) によってなされた.彼はこの新たな病原菌を「プリオン」と命名した.

プリオンの構成物質は異常プリオンタンパク質であり, DNA のような核酸は含まれていない. 異常プリオンタンパク質は、元来、身体の中に存在する正常プリオンタンパク質の立体構造(下図は、 正常なヒトのプリオンタンパク質の立体構造のリボン図である. PDB ID: "1QLZ")が変化したもの である. つまり、プリオン病は、脳内に存在する正常型プリオンタンパク質が、何らかの原因で、α ヘリックスが減少してβシート構造に富んだ構造である異常型プリオンタンパク質に変換されて、ア ミロイドと呼ばれる繊維状の構造形成が起こり、それらが蓄積すると、神経細胞を障害することによ って発病し、進行性・致死性である病気であることがわかっている. その発病のメカニズムも、また、 根本的治療法も未だに確立していない. 同一のアミノ酸配列でありながら、どのようなメカニズムで 正常プリオンタンパク質が異常プリオンタンパク質に変化するのだろうか? これは、「フォールディ ング・メカニズムの問題」であり、問題の解明が急務である.



正常なヒトのプリオンタンパク質

第1章 タンパク質のフォールディングの研究は何故重要か?

耳 タンパク質分子の免疫系における働き

2020年より今日, "新型コロナウイルス感染症" (Corona Virus Disease 2019, COVID 2019) で,未 曾有のパンデミックが起こっている. 今回のパンデミックは, "新型コロナウイルス" に対抗するた めの免疫システムが正常に作用しないために発生している.

その対策のため、現在のところ、"メッセンジャーRNA ワクチン"(mRNA ワクチン)が有効であ るとみなされている.この"mRNA ワクチン"は、"新型コロナウイルス"の表面にある"スパイク タンパク質"を形成する遺伝情報をもつ"スパイクタンパク質指示体"であり、これを接種すると、 ワクチン中の細胞とヒト細胞とが融合した後、細胞の中のリボソームで、mRNA の遺伝コードが翻訳 されて、"スパイクタンパク質"が生成され、細胞の外に漏れ出すことになる.我々の身体の免疫系 は、この"スパイクタンパク質"を異物として認識し、その抗体を作成する.

ところで,我々は常にバクテリアや細菌,ウイルスなどの攻撃にさらされている.そうした攻撃からどのようにして体を守っているのだろうか?

我々の体は、外来からのバクテリアや細菌、ウイルスに対抗するための免疫システムをもっている. 自己と非自己を認識し、非自己を"異物"として排除すべく攻撃を仕掛けるのが免疫作用である. "異 物"は、バクテリアや細菌、ウイルスだけでなく、花粉や塵などのアレルギー反応を引き起こすもの や、ガン細胞やウイルスに感染した細胞なども含まれる. 生体は、如何にして自己と非自己とを認識 しているのだろうか?

自己と非自己とを認識にする働きは、"MHC^{**}"と呼ばれる複数のたんぱく質の複合体が担っている ことがわかっている.皮膚や臓器の移植を行う場合も、提供する側と提供される側の MCH が一致し ていないと、移植された組織は "異物"とみなされ、免疫系に指令が回り、移植された組織は免疫の 力で破壊されてしまう.

** MCH:主要組織適合抗原(Major Histocompatibility Complex)

ヒトの主要組織適合抗原は、とくに HLA (Human Leukocyte Antigen, ヒト白血球抗原)と呼ばれ、 赤血球を除くほぼ体内のすべての細胞の表面に存在する特別な複数のタンパク質である.人それぞれ 構造の微妙な違いがあり、免疫システムが「自己」と「非自己」を区別するための目印として働いて いる.

免疫系は、膨大な多様性を生み出す機構をもっている.親から受け継いだ DNA に、初めから SARS, MERS, 麻疹, おたふく風邪, エイズ, インフルエンザ, 新型コロナなどのために指定された抗体が あるのではなく、感染してから初めて、部品の組み合わせで創り出された膨大な候補の中から、それ ぞれのウイルスに対して「鍵と鍵穴」の関係で結合できる抗体がそれぞれ選択される.

この抗体は"免疫グロブリン"という4本のポリペプチド鎖からなるタンパク質である.その立体 構造はY字型をしていて,抗原を見分ける二つの先端部分の可変領域において,何種類もの部品をラ ンダムに組み合わせて,膨大な種類の免疫グロブリンを創り出す機構をもっている.この機構は,利 根川進(1987年,ノーベル医学生理学賞)によって解明された.

このように、免疫系は膨大な多様性を生み出す機構をもっているが、こうした機構は、タンパク質 が特定の立体構造をとるために満たさなければならないアミノ酸配列の条件がある中で、部品のラン ダムな組み合わせでも立体構造を構築できることを意味する.これは「フォールディング・メカニズ ムの問題」であり、タンパク質のフォールディング研究の重要性を示している.



1953 年、英国ケンブリッジ大学の若き生物学者ジェームズ・ワトソンと物理学者フランシス・クリックによって、DNA (デオキシリボ核酸, Deoxyribose Nucleic Acid)の"二重らせん構造"に関する わずか1000 字の論文が、科学専門誌ネイチャーに掲載された. "親のもっている遺伝情報が、どのような機構で子に正確に伝えられるのか"という問いに対する答えが与えられたのである.分子生物学 の誕生である.ワトソンは、彼の著書:『The Double Helix』(1968年)の中で、次のように述べている; 「・・・When I asked what the pattern was like, Maurice went into the adjacent room to pick up a print of the new form they called "B"structure.

The instant I saw the picture my mouth fell open and my pulse began race. $\cdot \cdot \cdot$ the black cross of reflections which dominated the picture could arise only from a helical structure. $\cdot \cdot \cdot$

「その*X*線写真模様はどんなふうなのかと質問すると、モーリスは隣の部屋から、彼らが「B型」構造とよんでいる新形態を示す写真のプリントをもってきた.その写真を見たとたん、私は唖然(あぜん)として胸が早鐘のように高鳴るのを覚えた.・・・写真のなかでいちばん印象的な黒い十字の反射はらせん構造からしか生じえないものだった・・・」(ジェームス・D・ワトソン『二重らせん』江上、中村訳)



写真は, Franklin のB-型 DNA 結晶のX線回折写真 (1952年)で,「世界で最も重要な回折像」と呼ばれている.

彼らの論文で,冒頭次のように述べている:

 \lceil We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest. \rfloor

(デオキシリボ核酸 (DNA) の塩の構造を提案したい。この構造は生物学的にかなり興味深い新しい 特徴を持つ.)

次のような, 画期的な構造を提案している:

[If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).]

(塩基が最も妥当な互変異性型(すなわちエノール型ではなくケト型)の構造でのみ存在すると仮定 すれば、特定の塩基対同士のみが結合することができる. これらの組み合わせは、アデニン(プリン) とチミン(ピリミジン)、グアニン(プリン)とシトシン(ピリミジン)である.)



Watson-Crick 塩基対: DNA 中のチミン(T)とアデニン(A), シトシン(C)と グアニン(G) 間の水素結合

第1章 タンパク質のフォールディングの研究は何故重要か?

論文の最後は、次のような言葉で締めくくっている:

 \lceil It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material. \rfloor

(われわれの主張する特定のペアリング(塩基対)が遺伝物質の複製機構を直ちに示唆することには 誰でもが気付くだろう.)

彼らは、この論文において、2本のDNA 鎖が互いに逆方向に走行していることを示している.こ のことが、以降の重要な展開、つまり"DNA 複製のメカニズム"や"岡崎フラグメント"の発見をも たらすことになる.DNA の二重らせんは、互いに他を写した対構造をしていて、二重らせんが解け るとちょうど ポジとネガの関係となる.つまり、ポジをもとに新しいネガがつくられ、もとのネガか ら新しいポジがつくられると、そこには2組の新しいDNA 二重らせんが誕生するという、遺伝情報 による生命の"自己複製"システムの発見であった.この瞬間、新しい分子生物学が誕生したといえ るであろう.



" Chatsworth at Dawn "

- "ワトソンとクリックの論文" (1953 年) ——

(No. 4356 April 25, 1953 NATURE Vol. 171 737-738) MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey¹. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the Xray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van Der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate dieter groups joining β -Ddeoxyribofuranose residues



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

with 3', 5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's ² model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3.4 Å in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 Å. The distance of a phosphorus atom from the fibre axis is 10 Å. As the phosphates are on the outside cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being

hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

第1章 タンパク質のフォールディングの研究は何故重要か?

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given then the sequence on the other chain is automatically determined.

It has been found experimentally^{3,4} that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data^{5,6} on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material. Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donahue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. Watson

F. H. C. Crick

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge.

April 2.

- 1) Pauling, I., and Corey, R. B., Nature, 171, 346 (1953); Proc. U.S. Nat. Acad. Sci., 39, 84 (1953).
- 2) Furberg, S., Acta Chem. Scand., 6, 634 (1952).
- Chrgaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Riochim. Et Biophys. Acta*, 9, 402 (1952).
- 4) Wyatt, G., R., J. Gen. Physiol., 36, 201 (1952).
- 5) Astbury, W., T., Symp. Soc. Exp. Bio. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).
- 6) Wilkins, M. H. F., and Randall, J., T., *Biochim. et Biophys. Acta*, 10, 192 (1953).



DNA から転写された mRNA の塩基配列が指定するアミノ酸配列をもったタンパク質を合成する過程を翻訳という. 翻訳の過程で, mRNA の連続する3個の塩基をコドンという. 4 種類の塩基を3個並べる場合の数は,4×4×4=64 通りある. これらのコドンが,20 種類のアミノ酸のどれに対応するか,つまり, mRNA の塩基配列のコドンと,タンパク質のアミノ酸配列との対応関係を示したのが, "遺伝コード表"である. この表は,ごくわずかの例外(ミトコンドリアの DNA)を除いて,生物に共通であるとしてまとめられている.

		U		C		Α		G			
		コドン	アミノ酸	コドン	アミノ酸	コドン	アミノ酸	コドン	アミノ酸		
		UUU	Phe(F)	UCU	Ser (S)	UAU	Tyr(Y)	UGU	Cys (C)	U	
第	U	UUC		UCC		UAC		UGC		С	第
1		UUA	Leu (L)	UCA		UAA**	終止	UGA**	終止	Α	3
文		UUG		UCG		UAG**		UGG	Trp (W)	G	文
字		CUU	Leu (L)	CCU	Pro(P)	CAU	His (H)	CGU	Arg (R)	U	字
目	С	CUC		CCC		CAC		CGC		С	目
Ø		CUA		CCA		CAA	Glu(Q)	CGA		Α	の
塩		CUG		CCG		CAG		CGG		G	塩
基		AUU	Ile (I)	ACU	Thr (T)	AAU	Asn (N)	AGU	Ser (S)	U	基
	Α	AUC		ACC		AAC		AGC		С	
		AUA		ACA		AAA	Lys (K)	AGA	Arg (R)	Α	
		AUG*	Met (M)	ACG		AAG		AGG		G	
		GUU	Val (V)	GCU	Ala (A)	GAU	Asp (D)	GGU	Gly (G)	U	
	G	GUC		GCC		GAC		GGC		С	
		GUA		GCA		GAA	Glu (E)	GGA		Α	
		GUG		GCG		GAG		GGG		G	

第2文字目の塩基

*;AUGは「開始」コドンでもある.

***;「終止」コドン(停止コドン)は、3個有り:UAA(オーカー)、UAG(アンバー)、UGA(オパール)

上の遺伝コドン表の中には、アミノ酸以外のものを指定するものもある:

- ○「AUG」; metionine(メミオニン, Met, M)というアミノ酸を指定するコドンであるが, 同時に, 「開始コドン」(タンパク質への翻訳を開始するコドン) でもある.
- ○「UAA, UAG, UGA」; アミノ酸を指定せず, 翻訳の終了を指定する「終止コドン」
 (または,「停止コドン」ともいう)である.



💶 タンパク質に対する、20 種類のアミノ酸の寄与は?

タンパク質は、"20種類のアミノ酸"が重合して一次元的につながったポリペプチド鎖である.タンパク質の天然構造において、次のようなアミノ酸の性質がどのように寄与しているのだろうか?

- ・疎水性アミノ酸残基か、親水性アミノ酸残基か
- ・極性アミノ酸残基か, 非極性アミノ酸残基か
- ・酸性アミノ酸残基か、塩基性アミノ酸残基か、あるいは中性アミノ酸残基か
- ・側鎖の水素結合形成能は?
- ・二次構造形成能は?
- ・側鎖の大きさは?
- ・そのアミノ酸が立体構造の内部にあるか、表面にあるか、
- ・ 空間的な近傍に極性のアミノ酸があるか,
- ・水素結合能のある側鎖をもつアミノ酸があるか,
- ・機能部位かどうか,
- ・ α ヘリックスを形成しやすいアミノ酸が配列の前後に多いか.

特に,個々のアミノ酸のもつ意味は,「アミノ酸配列という文脈に依存する」という次の例が興味深い:「部分配列がまったく同じ5個のアミノ酸残基からなる領域が,あるタンパク質ではαヘリックス を形成し,あるタンパク質ではβ構造を形成する事例」

次のページの表は、アミノ酸の物理化学的指標である"疎水性度"の大きい順に、"20 種類のアミノ酸"を並べたものである.



"Trees in Autumn"

分類	アミノア	骏 名		疎水性度	二次構造形成能		成 能	体積	測鎖の
	名称	3文字	1文字		α	β	ターン	(Å ³)	pK值
		表示	表示		ヘリックス	構造			
	phenylalanine	Phe	F	-2.27	1.16	1.33	0.59	135	
	(シェニールアラニン)	Tree	W	2.12	1.02	1.25	0.65	162	
志	(トルプトファン)	Trp	vv	-2.15	1.02	1.55	0.65	105	
山木		Lou	т	1.92	1.24	1.22	0.57	124	
-14	Teucine	Leu	L	-1.82	1.54	1.22	0.57	124	
水	(ロイシン)								
1.4	isoleucine	lle	I	-1.82	1.09	1.67	0.47	124	
作生	(イソロイシン)	_							
	tyrosine (fraid)	Tyr	Y	-1.47	0.74	1.45	0.76	141	10.1
	valine	Val	V	-1 30	0.90	1.87	0.41	105	
↑	(バリン)	v ui	•	1.50	0.20	1.07	0.11	105	
	proline	Pro	Р	-0.99	0.34	0.31	1.32	90	
	(プロリン)								
	methionine	Met	М	-0.96	1.30	1.14	0.52	124	
	(メチオニン)	Alo	Δ	0.20	1.41	0.72	0.82	67	
	(アラニン)	Ala	A	-0.39	1.41	0.72	0.82	07	
	cysteine	Cys	С	-0.25	0.66	1.40	0.54	86	8.2
	(システイン)	-							
	glysine	Gly	G	0.00	0.43	0.58	1.77	48	
	(グリシン) histiding	II.	11	0.64	1.05	0.80	0.91	110	()
	(トスチジン)	HIS	н	0.64	1.05	0.80	0.81	118	6.0
	threonine	Thr	Т	1.00	0.76	1.17	0.90	93	13.6
	(トレオニン)								
	serine	Ser	S	1.24	0.57	0.96	1.22	73	13.6
	(セリン)								
	glutamine	Gln	Q	1.30	1.27	0.98	0.84	114	
	(グルタミン)								
	asparagine	Asn	Ν	1.91	0.76	0.48	1.34	96	
¥	(アスパラギン)								
	lysine	Lys	K	2.77	1.23	0.69	1.07	135	10.5
親	(リジン)								
	glutamic acid	Glu	Е	2.91	1.59	0.52	1.01	109	4.3
7K	(クルタミン酸)	Asp	D	3.81	0.00	0.30	1.24	01	37
性	aspartic actu (アスパラギン酸)	Asp	D	5.61	0.99	0.59	1.24	91	3.7
1-	arginine	Arg	R	3.95	1.21	0.84	0.9	148	12.5
	(アルギニン)								

20種類のアミノ酸を特徴づける物理化学的指標

<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.



第2章 タンパク質分子の フォールディング問題とは?

<要旨>

"タンパク質分子のフォールディング問題"に は3つの側面があることを紹介している.そして, その中の"タンパク質のフォールディング・メカ ニズムの問題"を「統計力学的観点」から解明す ることがこの本の主たるテーマであると述べてい る.

郷信広が提案した「整合性原理」,つまり, 「タンパク質分子は進化の所産として,天然構造 を安定に保つのに寄与している様々な分子内相互 作用の間に互いに矛盾が無く整合的である』と, これと等価であると言われている,Wolynesらの 「極小フラストレーションの原理」について考察 している.

更に、これらの原理の展開から、タンパク質の フォールディングにおける"エネルギーランドス ケープ理論"から導かれる"ファネル描像"の重要性 が広く認識されていると述べている. 第2章 タンパク質分子のフォールディング問題とは?

□ タンパク質分子のフォールディング問題とは?

タンパク質分子のきわだった特徴の1つは、生理的な条件下で、そのアミノ酸配列によって決まる 特異的な立体構造を自発的に形成することである.タンパク質分子が、どのようなメカニズムで、そ のアミノ酸配列で決まる、それぞれに固有の天然構造へフォールドするのだろうか?

このタンパク質フォールディング問題は、第1章で述べたように、次のような3つの側面がある: ① 天然構造の予測の問題:

『特定のタンパク質のアミノ酸配列が与えられたとき,その天然立体構造を予測するという問題』 ② フォールディング・メカニズムの問題;

『タンパク質のフォールディング過程の性質(レート,中間状態,遷移状態など)を明らかにす る、つまり、フォールディングのメカニズムを明瞭にする問題』

③ 天然構造の特徴の問題; 『"タンパク質の天然構造は,静的にも,動的にも,どのように特徴づけられ,それが生物学的にどのように意味をもつのか?"を明らかにする問題』

これらの"フォールディング問題"に関して、いろいろと非常に興味深い提案がなされているが、 依然として、明確な解答が得られていない未解決の難問として残されている。

我々の目標は、「フォールディング・メカニズムの問題」を"統計力学的観点"から解明することである.

↓ ポリペプチド鎖の可能な全配列のうち、タンパク質分子となるアミノ 酸配列は、ごくごく少数である!

タンパク質は、20種類のアミノ酸が枝分かれすることなく重合してで きた鎖状高分子である. つまり、アミノ酸2分子が脱水縮合するとアミ ド結合が形成されジペプチドになり、それらが、数10個から数千個、ペ プチド結合でつながって構成された生体高分子である.





ポリペプチド

"アミノ酸配列の可能な場合の数"を見積もってみよう.

○アミノ酸配列が2個の場合; 20種類のアミノ酸の重複順列なので,

202 (=400) 通りのアミノ酸配列の場合の数

- ○アミノ酸配列が3個の場合; トリペプチドのアミノ酸配列の可能な場合の数は, 20³(=8,000)通りのアミノ酸配列の場合の数
- ○アミノ酸配列が4個の場合; テトラペプチドのアミノ酸配列の可能な場合の数は, 20⁴(=160,000)通りのアミノ酸配列の場合の数
- ○アミノ酸配列が100個の場合;そのアミノ酸配列の可能な場合の数は,

20¹⁰⁰ (≃10¹³⁰) 通りのアミノ酸配列の場合の数で,莫大な数のポリペプチド鎖を作成できる. この莫大なポリペプチド鎖の内,実際に,特異的立体構造を形成してタンパク質として機能できる のは、ごくごく限られたポリペプチド鎖だけであり、ほとんどのポリペプチド鎖は立体構造ができず に沈殿するといわれている.

アミノ酸配列の全空間を考えよう.これは、それぞれのアミノ酸配列がその空間上の1点として表 され、似たアミノ酸配列は互いにその近傍に位置するとする.



一般に、ある1つの天然構造をとることのできるアミノ酸配列は複数ある.そうしたアミノ酸配列 は、配列空間でお互いに近傍にあり、1つの領域を構成する.しかしながら、まったく異なるアミノ 酸配列(配列空間の異なる領域)が同じ天然構造(構造空間の同じ領域)に対応することも、現実の タンパク質でもみることができる.そして、天然構造にはさまざまなタイプが存在しており、それら がアミノ酸配列空間の異なる領域を占めていると考えられる.一方、どんなアミノ酸配列でもすべて が何らかの特異的な立体構造へと折れたたまれるわけではないので、こうした配列空間の中で、フォ ールド可能な領域はまばらにしか存在しないであろう.

地球上には約200万種の生物が存在するだろうと言われているので、10⁸~10¹⁰個位のタンパク質 種が存在すると推定すると、現実に地球上にある生物によって使用されているタンパク質は、全ポリ ペプチド鎖10¹³⁰個のごくごく少数である.

残りの約 10¹²⁰ 個のポリペプチド鎖は,タンパク質として使用できないのだろうか?

生命誕生から今日までに,遺伝子が,すべてのポリペプチド鎖の配列に対して,タンパク質分子として機能するかどうかをチェックしたとは到底思えない.その数を見積もってみよう.

40億年前地球上に生命が誕生して以来,遺伝子が,もし1秒間1個の割合で100個のアミノ酸配列 を合成したポリペプチド鎖がタンパク質として機能するかどうかをチェックできたと仮定すると,現 在までに10¹⁷個のポリペプチド鎖をチェックしたことになるが,残りの約10¹¹⁰個のポリペプチド鎖は チェックされないままであろう.残りのポリペプチド鎖の中に,いろいろな機能をもった新たなタン パク質が発見できるかもしれない.それは,タンパク質工学という学問がめざす目標の1つでもある.

タンパク質分子のアミノ酸配列に書き込まれた特異的な立体構造形成の情報を読み解く問題の解 答は、次に述べるようなアンフィンゼンの実験によって、物理学によって必然的に与えられる性質の ものであるにもかかわらず、今日まで、未解決の難問として残されている. 第2章 タンパク質分子のフォールディング問題とは?

↓ "アンフィンゼン・ドグマ":『タンパク質分子は自発的に、熱力学的に、
 に最も安定な特異的立体構造を形成する。

アンフィンゼン (Anfinsen)らは、1950 年代から 1960 年代はじめの一連の実験、"RNA 分解酵素リ ボヌクレアーゼ A (RNaseA)"を対象にした変性・再生の実験の結果から、生命科学において重要な、 次のような"アンフィンゼン・ドグマ"を提唱した(1972 年、アンフィンゼンは、この研究成果によ りノーベル化学賞を受賞):

『タンパク質分子は自発的に,熱力学的に安定な特異的立体構造を形成する,言い換えると,タンパク質分子の天然立体構造は, そのアミノ酸配列によって一意的に決定される.』

彼らの実験の概要は次の通りである:

彼らは、溶液中で酵素活性をするリボヌクレアーゼA(右図は天 然構造,124残基,PDBID: "5RSA")に、変性剤である高濃度の 尿素と還元剤であるメルカプトエタノールを加えると、天然構造 が壊れ、酵素活性が失われることを観察した.これは、4 個ある ジスルフィド結合(Cys26-Cys84, Cys40-Cys95, Cys58-Cys110, Cys65-Cys72)の全てが切断されて、天然立体構造が壊れ、完全な 変性状態になったと考察した.その後、彼らは、この状態のリボ ヌクレアーゼAを、酸素の存在下で、尿素濃度の低い溶液に戻す と、つまり、溶媒条件を元に戻す(透析する)と、再び、酵素活 性が復活することを確認した.結局、もとの天然立体構造が形成 されたことを確認した.

この実験により,次のことを提案した;

『タンパク質分子の天然立体構造は,タンパク質分子と水を含ん だ系での熱力学的安定状態,つまり,溶液条件によって指定され た自由エネルギー最小状態である.言い換えると,タンパク質分

子のフォールディングは、熱平衡状態(自由エネルギー最小状態)へ向かっての緩和過程である.』 <文献> C. B. Anfinsen, Principles that Govern the Folding of Protein Chains, *Science*, **181**, 223-230, 1973.

アンフィンゼン・ドグマの提唱により、タンパク質のフォールディング問題は、次のような物理・ 化学の問題であると考えられるようになった:

『タンパク質分子は、与えられた溶媒条件下で、自発的に特異的天然立体構造へ折れたたまれていく が、このとき、アミノ酸配列と環境条件に含まれる情報以外に、新たな情報を必要としない.すなわ ち、タンパク質分子の天然立体構造形成に関する情報は、そのアミノ酸配列の中にすべて書き込まれ ている.』最近、アンフィンゼンらが行った試験管でのプロセスと異なり、細胞中では、フォールディ ングを助ける様々な仕組み(分子シャペロンの関与等)が明らかになってきているが、細胞中のフォ ールディングも基本的にタンパク質分子が自己組織化する能力をもっていることにより初めて可能に なると考えられていて、アンフィンゼン・ドグマの重要性は変わらない.

更に最近では、次のような、アンフィンゼン・ドグマがそのままでは成立しない例が知られていて、 大変興味深い:

① タンパク質分子が、複数の異なる秩序構造を取る:

フォールディング病の原因となるプリオンタンパク質のアミノ酸配列は,健常なヒト体内で発 現される正常型タンパク質と完全に一致するにも関わらずその立体構造には大きな違いがある. これは、アミノ酸配列によってタンパク質構造が一意に決定されない好例といえる.この場合、天然立体構造を、周囲の環境も考慮した立体構造のいくつかの準安定構造のひとつと解釈すればアンフィンゼン・ドグマは成り立つ.

② タンパク質分子が、生理的条件下で秩序構造を取らない、つまり、天然変性状態タンパク質の 存在が確認されている.

しかしながら,基本的には、アンフィンゼン・ドグマをタンパク質分子のフォールディングを理解 するための出発点としよう.

🎞 タンパク質分子は如何にして自らの天然構造を見出すのだろうか?

タンパク質のとりえる可能な全コンフォメーションをエネルギー曲面で表すことを考えよう. つま り、全コンフォメーションを表す多次元空間中で、各コンフォメーションを点として表し、その点で のタンパク質分子がもつエネルギーの値からなるエネルギー曲面を考えよう. タンパク質は、このエ ネルギー曲面上で、如何にして、自らの天然構造を、ミリ秒から分の単位で敏速に見出すことができ るのだろうか?

ここで、ゴルフコースで色々な面をもつグリーン面をエネルギー曲面と考え、グリーン面の縁から パッティングをしてボールをカップに沈めることを、タンパク質が折れたたまれて天然構造に到達す ることとみなして考えてみよう. 典型的な例として、次の3つの場合が考えられるだろう:

(a) グリーン面 (エネルギー曲面) が, 広大で平面の場合

グリーン面が平らで広大である場合,ゴルファーが 目隠しされて,グリーン面の端から,ランダム方向に パッティングを試みるように命じられたら,なかなか カップに入らないであろう.偶然にカップに入るのに どれくらい時間が必要だろうか?

全てのコンフォメーションの中から、ランダムに、 しかも、網羅的に天然構造を探索するのにどのくらい 時間がかかるか概算してみよう:

いま,100 個のアミノ酸残基からなるタンパク質分 子を考えよう.



各アミノ酸残基あたりの可能なコンフォメーションの数は、少なくとも3通りはあるので、鎖全体の可能なコンフォメーションの数は、 $3^{100} \approx 5 \times 10^{47}$ 通りのである.

1つの構造を探索するのに、仮に、10⁻¹⁵秒必要であると仮定すると、すべてのコンフォメーション をランダムに探索するのに 5×10^{47} (通り)×10⁻¹⁵ (秒/通り) = 5×10^{32} (秒) = 1.6×10^{25} (年) という莫大な時間が必要になる.

鎖を構成するアミノ酸残基の数が増えるとともに、解決に必要な時間は、指数関数的に伸びること になる.これではタンパク質はほとんど永久に天然構造を探し当てることが出来ないだろう.

しかし、実際のタンパク質分子は、ミリ秒から分の単位で天然構造へと折れたたまれる.これは、 "Levinthal のパラドックス"として知られている.

このパラドクスが生じた原因は、構造の探索に何もガイドラインがなく、すべてのコンフォメー ションの生起確率が等しいとし、ランダムサンプリングによって天然構造を探索する状況を想定した ことにある. 第2章 タンパク質分子のフォールディング問題とは?

(b) グリーン面 (エネルギー曲面) が, 広大で平面であるが, カップに向かって溝がある場合

これは、Levinthalのパラドックスを回避するしくみの1つである.

グリーン面が平面で広大でも、グリーンの一端からホールに 向かって溝がある場合、その溝を通してパッティングを試みれ ばボールは容易にカップに入るであろう.

っまり、ランダムコイルから天然構造へのフォールディング には、非常に限られた道筋があり、道筋に含まれない領域のコ ンフォメーション・エネルギーはきわめて高く、ほとんど実現 しないとみなしているので、容易に天然構造に到達するだろう. タンパク質は、選ばれた順路をたどってフォールディングする のではないかという提案である.

このアイデア,つまり、"フォールディングの道筋は精密に設 「 計されている"というアイデアは、1980年代に主流となり、フォールディングの途中に表れる中間体 の探索の研究が重要であると考えられ、多くの実験が行われた.

タンパク質分子がランダムコイル状態から天然状態へ折れたたむ過程で、実際に形成するであろう 途中の構造(中間体)を検出することが目標になった.

1980年代後半から,特に球状タンパク質では,モルテン・グロビュール中間体を経由して天然構造を形成するとの提案がなされ,注目された.

1990年代に入ると、100残基以下の小さなタンパク質において、中間体を経ずに二状態的に天然構造に折れたたむ例が次々に報告され、特異的なフォールディングの道筋や中間体は必須ではないのではと考えられるようになった.

(c) グリーン面 (エネルギー曲面) が, ファネル状 (漏斗状) になっている場合

グリーン面がカップに向かってファネル状(漏斗状)になっている,つまり,ホールに向かって傾 斜している様なエネルギー曲面を考えよう.

この場合,グリーンの端からパッティングを試みれば, 重力がボールをホールへと送り込み,容易にカップに入る であろう.つまり,立体構造空間が作るエネルギー曲面が 大域的に天然構造を底とするファネル(漏斗)状になって いれば,グリーン面がいろいろと起伏に富んでいても最終 的には天然構造に折れたたむであろう.

このとき、タンパク質のフォールディングでは、特異的 な経路や中間体は必須ではなく、様々なトラジェクトリを 経て、最終的には天然構造へと到達するという考えである.

つまり、エネルギー曲面が全体的に天然状態にバイアス がかかったファネル状の形状をしているため、特異的立体 構造(天然構造)へと折れたたむという考え方である.

今日、このエネルギーランドスケープ理論に基づく捉え

方(ファネル理論)を, new view (新しい捉え方)と呼び, 従来の,「フォールディングには, アミノ酸配列によって規定された, 特異的な経路が存在し, その経路上には, 定まった中間体が在在する」という見方を, old view (古典的捉え方)と呼ぶ.

「タンパク質のフォールディング過程は,ファネルを上から下へ落ちていくようなものである」というファネル理論は,1990年代に Wolynes らによって提案された.

<文献> J. D. Bryngelson, J. N. Onuchic, N. D. Socci and P. G. Wolynes. Funnels, patheways, and the Energy Landscape of Protein Folding: A Synthesis, *PROTEINS: Structure, Function, and Genetics*, **21**, 167-195, 1995.





ファネル理論のルーツは、1983年、Annual Reviewに発表されたGōによる次のような「整合性原理」に まで遡ることができる:『球状タンパク質の天然構造を安定化しているいろいろな相互作用項は、第一 近似としては、互いに整合的である. すなわち、この性質を満たすポリペプチド鎖のみが、進化の過 程を通して、タンパク質分子として、選択されてきた.』

<文献> N. Gō, Theoretical Studies of Protein Folding, Annu. Rev. Biophs. Bioeng. 12 183-210, 1983.

<文献> 郷信広,『蛋白質の立体構造予測と折れたたみ過程』,生物物理, Vol. 23 No.3, 11-19, 1983.

<文献> 郷 信広:『バイオ科学へのチャレンジ』,物理学の挑戦(日本物理学会編),日本評論社,10章, pp.180-196,2006年.

耳 タンパク質のフォールディング転移の協同性

タンパク質のフォールディング転移の協同性について考察しよう. 小さな球状タンパク質分子のもっとも顕著な特性の1 つは,フォールディング における"ニ状態転移"である.タンパク質のフォールディングは、ランダム コイル状態からフォールドされた状態へ一気に起こり、部分的に変性した中間 体は存在しないか、せいぜい数個しか観察されない.同様に、フォールドされ た状態からランダムコイル状態への構造変化も協同的に進む.

平衡状態におけるフォールディング転移の中点においては,分子集団中の半数の分子は完全に折れたたまれた状態にあり,残りの半分は,ほどけた,ランダムコイル状態にある.

つまり、タンパク質分子は、転移温度($T = T_m$)においては、外部条件(温度や溶媒など)の変化にともない、ほとんど天然構造にフォールドした分子か、ほどけたランダムコイル状態にある分子かの二者択一的であるという特徴がある。一つの鎖のなかで、ランダムコイル状態にある部分と局所的に天然状態にある構造が混在したタンパク質分子はまれであることを示している(図参照、N:天然状態、D:ランダムコイル状態).





この小さい球状タンパク質分子のフォールディングの二状態性は、特異的な構造をとることによる エントロピー損失と、長距離相互作用によるエンタルピー(エネルギー)獲得の拮抗が協同的な二状 態転移を引き起こしている.

<コメント> 物理学の研究目的の一つは、自然現象の根本要素がどうなっているかを解明することである.しかし、自然現象を理解するためにはもうひとつ重要な側面がある.それは、協同現象とよばれるもので、簡単な要素でも、それらが多数集まり、互いに相互作用することによって巨視的に特徴ある状態(秩序構造)を生み出す現象である.たとえば、水の分子に注目すると、分子自身は変化しないが、温度によって氷、水、水蒸気とマクロな形態が温度とともに不連続に変化する.このように、構成要素自身が変化しないのに、マクロに特異性が現れる現象を「相転移」というが、その特徴は、ある環境条件では一部の要素が秩序構造を作りエネルギー的に安定化しようとしても、それによって失うエントロピーの方が大きく、すぐに不安定化してしまうが、異なる環境条件では、一斉に秩序化することで相互作用エネルギーの総和が失うエントロピーに打ち克つようになるというものである.この一斉に秩序化する点がポイントであり、協同現象の名前の由来となる.タンパク質が温度や溶媒の変化によって引き起す変性・再生(フォールディング・アンフォールディング)という可逆的な構造変化もこの「相転移」の一種とみなすことができる.相転移の機構は統計力学の中心課題のひとつである.

耳 ファネル理論のルーツは、「整合性原理」である

Gō らは、今日、"Gō ポテンシャル"という量を導入して、2次元格子モデルタンパク質のフォール ディングのコンピュータ・シミュレーションを実行した.

彼らは,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")に着目し,フ オールディング過程においては,その天然接触相互作用のみが働くと仮定して,シミュレーションを 実行した結果,タンパク質らしい協同的振る舞いを観察した.

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、グローバルなファネル様地形では、これらの効果は、ほとんど無視してもよいであろうと彼らは考察した。天然接触相互作用と天然非接触相互作用の両方を考慮したコンピュータ・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。

この Go ポテンシャルの影響下のとき、つまり、理想的に天然接触相互作用のみが働くと仮定する と、次のようなフォールディング過程の描像が考えられる:

『フォールディングの初期段階では、アミノ酸配列上で近い分子間相互作用(近距離相互作用)が働いて、いろいろな部位で部分構造を形成するが、この部分構造は比較的不安定で、壊れたり、また形成されたりする.しかし、部分構造が形成されている間に、たまに、アミノ酸上で距離の離れたアミノ酸残基間相互作用(遠距離相互作用)が働くと、最終的に、部分構造とともに全体構造が安定になって、天然構造が形成される.』

Gō は、一連のシミュレーションの結果から、タンパク質分子のフォールディングにおけるこの 2 状体転移が可能であるための条件は何かという観点から、1983 年、次のような"整合性原理"を提案 した;『タンパク質分子は進化の所産として、天然構造を安定に保つのに寄与している近距離相互作用 や遠距離相互作用などの様々な分子内相互作用の間に、互いに矛盾が無く、整合的である』

その後, Wolynes らは、低温のスピングラスにおいてスピンの向きが凍り付いて動かなくなるガラ ス転移との類推から、タンパク質分子がガラス転移を避けてフォールディングを可能にする条件は何 かという観点から、次のような"極小フラストレーションの原理"を提案した;『タンパク質の天然構 造においては、どのエネルギー項をとっても他のエネルギー項の犠牲になってフラストレーションを 感じることなく、フラストレートした相互作用が十分小さくなるように、進化を通じて、精巧に、そ のアミノ酸配列をデザインしてきた.』

現在では、"Gō ポテンシャル"の成立はタンパク質の基本的な性質であると捉えられている. Wolynes らにより提案された"エネルギーランドスケープ理論"も、本質的な部分は Gō の「整合性 原理」と等価だとされる.つまり、同じ物理的内容を表している.ところで、一体どのようなアミン 酸配列をデザインすれば、これらの原理が満たされるのだろうか?

<コメント>高田彰二は、これらの二つの原理に関して 2010年の生物物理学会誌で次のように述べている:『・・・郷氏の整合性原理と Wolynes のフラストレーション最小原理とは明らかに似ており、 "同じだ"という人も多くいますが、私の解釈は少し違います.対比的にいうと、郷氏のものは、天 然構造の理想的な整合性を主張したものであり、Wolynes のものは、理想的な整合性と不可避的な不 整合性(フラストレーション)の比を最大化する、という主張です.比喩的には、郷氏のものは理想 極限の1項を議論し、Wolynes のものは1項と2項の大きさの比を議論しています.2つの項を比べるこ とにより、Wolynes らは、天然構造の折れたたみ能を明示的な数式にしました.・・・』(高田彰二、『郷 モデルの35年-郷信広先生に応えて-』、生物物理50(4)、158-159、2010)

Gō モデルは現実のタンパク質のフォールディングでも成立しているのだろうか?

当初, Gō ポテンシャルは, 次のように酷評された:『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, 物理的ではない(Gō 's potential are not physical), 乱暴な(rough), しかも, 人為的(artificial)なポテンシャルである.』

1970 年代以降に行われたフォールディングの実験結果は、「小さな球状タンパク質のフォールディングが二状態的に起きる」ということであった.つまり、『小さな球状タンパク質では、変性状態から 天然状態への折れたたたみの途中に寿命の短い中間体が観測されない』という実験事実は、フォール ディングの途中の構造変化を追跡することが原理的に大変難しいと思われた.

Fersht らは, 1990 年代のはじめに, "Φ値解析法"という手法を新たに提案し, フォールディング 遷移状態での構造的特徴を推定することを可能にした.

<文献> A. Fersht, Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding.

Freeman, New York, 1999. [桑島邦博・有坂文雄・熊谷泉(訳),タンパク質の構造と機能, 医学出版(2006)] Gō ポテンシャルは,構造の選択性が非常に高い特性をもっている.すなわち, Gō ポテンシャルは, タンパク質分子が,余計な構造探索を行わずに,間違った構造を回避して短時間内に天然構造を作る. Levinthal のパラドックスは,実際のタンパク質で Gō ポテンシャルが 成立していれば,解決するこ とは自明であるともいえる.

1990年代後半に、『タンパク質は、整合性原理が成り立っており、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、フォールディング過程では、天然接触相互作用のみが主に働く.』という Gō ポテンシャルが、現実のタンパク質にも近似として使えるのではと考えられ、実験との検証が行 われた. 驚くことに、実験的なΦ値解析の結果と、Gō ポテンシャルを採用したモデル計算との結果 が定性的に一致することが確認された. (<文献> N. Koga and S.Takada, Roles of Native Topology and Chain-length Scaling in Protein Folding: A Simulation Study with a Go-like Model, J. Mol. Biol., 313, 171-180, 2001.)

2000 年頃には、タンパク質フォールディングを記述する基礎理論として、 Gō の「整合性原理」や ウォリス (Wolynes)らの「極小フラストレーションの原理」を基にした、"エネルギーランドスケー プ理論"、および、"ファネル描像"の重要性が広く認識された.

Gō の整合性理論が実験的に検証されたことは、科学史に残る大いなる成果であろう. 郷は、2019 年、"Snake cube puzzle"(SC模型)を用いて、自らが提案した「整合性原理」の具体例を示す論文を 発表した(Coffee Break::『パズルによる"整合性原理"の具体例』参照).

しかし、タンパク質のフォールディング問題がこれで解決したわけではない. なぜ、現実のタンパ ク質分子のアミノ酸配列では整合性原理が成立し、フォールドしないポリペプチド鎖のアミノ酸配列 では沈殿してしまって、整合性原理が成立しないのだろうか? 更に、この整合性原理を用いて、任 意のポリペプチド鎖のアミノ酸配列が特異的立体構造を形成するかどうかの判断を行うことが可能で あろうか?

タンパク質フォールディングにおけるエネルギー曲面を描くにはどうすれば良いだろうか?

今日,非常に小さな球状タンパク質の場合か,天然構造の周辺に限定して,エネルギー曲面を描く ことが行われている.タンパク質のフォールディング過程を理解するためには,すべての構成原子を 考慮して,その様々なコンフォメーションについてコンフォメーション・エネルギーを計算し,エネ ルギー曲面を具体的に描くことがもっとも直接的な方法である.

だが, 現時点では, 実際のタンパク質分子で, すべての構成原子を考慮して網羅的な計算を実行し, 詳細なエネルギー曲面を描くことは難しい. そこで, 我々は, 次善の策として、3 次元格子タンパク 質という単純化した模型を用いてエンルギー曲面を描くことを試みよう.



は極めて有効に働く、つまり、異なるエネルギー項が整合的に働くことが、構造を一意的に決めるのにどのように寄与するかを、定量的に示している.』と郷は述べている. (<文献>Go, N. :Snake cube puzzle and protein folding, Biophys, Physicobiol. 16, 256-263, 2019. 郷信広:

『Snake Cube Puzzle とタンパク質フォールディング』,生物物理 61,005-011,2021.)



そして,平衡状態で,そのコンフォメーショ ンが,あるエネルギー値をもつ確率は,"ボル ツマン分布"で記述されることを,Dillらの 方法に準じて,ラグランジュの未定乗数法を用 いて導いている.

この導出過程で表れる,系の"分配関数" (状態和)から,いろいろな熱力学量を理論的 に求めることができることを示している.その 際,具体的な最も簡単な例として,4個のアミ ノ酸残基からなる3次元格子模型を用い,熱力 学量として,転移曲線や比熱曲線を求めている.

➡ タンパク質分子は、"ボルツマン分布"なんて知るはずがない!

生命活動の大部分は、タンパク質分子の働きによって成り立っているとみなして良いであろう.しかしながら、タンパク質分子は、与えられた環境の中で、自分の"形"や"構造"を自発的に"揺らがせる"ことだけが唯一の仕事であり、タンパク質には自覚がなく、単に、物理・化学の法則に従う高分子にすぎない. つまり、タンパク質分子は、構造的揺らぎをもった分子機械であり、我々は、タンパク質分子において観測される振る舞いの統計的な量を記述するのみである. 従って、タンパク質分子は、自分が"ボルツマン分布"で記述されることなど知らないし、聞いたこともないであろう.

Zuckerman は,著書「生体分子の統計力学入門」の第1章の"タンパク質は生物学を知らない"の中で,次の様に述べている:

『・・・生物学は、その大部分が、タンパク質による驚くべきトリックによって成り立っている. し かし、タンパク質は、物理と化学の法則に従う単純な分子にすぎない、タンパク質を機械と考えるこ ともできるが、その中には魂は入っていない、タンパク質は命の吹き込まれていない物体であり、そ の主な役割は配置(つまり形や構造)を揺るがせることだけである.進化によって、生物学は、リガ ンド結合,移動,触媒反応のような有用な構造揺らぎを選択してきた.しかし,1個もしくは複数の 分子の高度に進化した機能を理解するには、物理学者リチャード・ファイマンの言葉を借りると、分 子がどのように自発的に「揺れ動く(wiggling and jiggling)」か考えるのがよい.少し違う言い方をす ると、分子スケールでの生物学とは化学と物理なのだ、したがって、分子生物物理学の原理は、最小 限の暗記のみで明快に理解できる.「平衡」という仰々しい言葉のせいで生物学や化学や物理の一部が 静的であると考えたのなら、それは違っている、事実はそのイメージからかけ離れている、科学的に 興味深い現象はすべて、本質的に、動いたり揺らいだりしているのだ. どんな平衡もそれは見せかけ だけのもので、複数の動きが統計的に釣り合っている結果である.この原理を別の言葉で表すと、「自 然は統計的な計算ができない」となる.いろんな過程が何回も起こって、相反するような現象を統計 平均した結果、われわれが簡単に理解できるようになるのである。タンパク質と同様に、自然も「マ ヌケ」である. 自然は計算機やコンピュータを使うことができないので、確率や平均や標準偏差のこ となど分からない.・・・』(Daniel M. Zuckerman 著:『生体分子の統計力学入門-タンパク質の動きを理解す るために一』、藤崎弘士・藤崎百合訳、共立出版、2014年)

タンパク質分子は「平衡」など聞いたこともないだろうが,我々は,次のことが知りたい:『タンパ ク質分子が,平衡状態で,あるエネルギーをもつコンフォメーションが実現する確率はどのような式 で表されるのだろうか?』

平衡状態で、あるエネルギーをもつコンフォメーションが実現する確率は、 ボルツマン分布で表される

いま, *n* 個のアミノ酸残基の配列から構成されている1個のタンパク質分子が,生理的環境条件 に置かれているとしよう.このタンパク質分子は,動いたり揺らいだりして,多くのコンフォメーシ ョンを形成する.この多くのコンフォメーションからなる系を微視的な系(microscopic system)と呼 ぶことにする.我々は,タンパク質分子が,あるエネルギーをもつコンフォメーションが実現する確 率分布(*p*₁,*p*₂,*p*₃,…,*p_i*,…)を,"ラグランジュ(Lagrange)の未定乗数法"を用いて求めよう(詳細は, Appendix D:「ボルツマン分布の導出」参照): 1個のタンパク質分子が生理的環境条件に置かれているような系を考える. この系の温度 *T*,体積 *V*は一定とする. いま,1個のタンパク質分子は、多くのコンフォメーション(ミクロな状態)をと る.取り得るコンフォメーションの総数を *N*(一定)とし、それぞれのコンフォメーション・エネ ルギーを $E_j(j=1,2,3,\dots,N)$ とする. 我々が知りたいのは、平衡状態で、エネルギーが E_j であるコン フォメーション*j*が存在する確率 P_i は、どのような式で表されるのかである.

系の自由エネルギーをFとすると、F = U - TS である.ここで、Uは、熱力学的な内部エネル ギーで、微視的な立場からすると、ミクロなエネルギーの期待値であり、Sは、統計力学的エントロ ピーである.また、すべてのコンフォメーションが存在する確率の和は1であることより次式が成立 する:

$$\sum_{j=1}^{N} p_j = 1$$

結局,平衡状態で,あるコンフォメーションjがエネルギー $E_j(j=1,2,3,\cdots,t_N)$ をもつとき,その存在する確率を求める問題は,次のように定式化できる:

『系の自由エネルギーをFとすると、平衡条件は、dF = 0であるから、結局、条件式 $\sum_{i=1}^{N} p_{i} = 1$ のもとで、自由エネルギーFの極値を求める問題』

ラグランジュ(Lagrange)の未定乗数法(Appendix D:「ボルツマン分布の導出」参照)を用いて、平衡状態で、エネルギー E_i であるコンフォメーションjが存在する確率 p_i は、次式で与えられる:

$$p_j = \frac{e^{-E_j/k_BT}}{Z}$$
 $(j = 1, 2, 3, ..., t_N, Z = \sum_{j=1}^N e^{-E_j/k_BT})$

これを"ボルツマンの分布則" (Boltzmann distribution law) という.ここで、*T* は絶対温度を表し、 k_B はボルツマン定数である; $k_B = 1.380662 \times 10^{-23} JK^{-1}$. この式の分子の項 e^{-E_j/k_BT} は、"ボルツマン 因子" (Boltzmann factor) と呼ばれる量で、次のような重要な特徴をもっている:

- ・より低いエネルギーをもつコンフォメーションが実現しやすい,つまり,エネルギーの高い コンフォメーションほど存在する確率は,指数関数的に減少する.
- ・ある温度で存在する可能性が低いコンフォメーションでも、温度の上昇とともに、存在する 可能性が高くなる.

ボルツマン分布則の分母の項 $Z(=\sum_{j=1}^{N} e^{-E_j/k_BT})$ は、系の分配関数(状態和)と呼ばれる量である.こ

の分配関数Zは、ポリペプチド鎖が取り得るコンフォメーション全てについての和である.

更に、分配関数Zは、一般的に、温度Tの関数であり、しかも、微視的な個々の状態と、巨視的な 熱力学的量とを結ぶ重要な関数である.つまり、分配関数はいろいろな物理量を計算する際の基点に なる.系の分配関数を求めることができると、その分配関数を用いて、自由エネルギー、内部エネル ギー、エントロピーは次のように計算することができる(温度の関数として表現している).

- ・自由エネルギー : $F(T) = -k_B T \ln Z(T)$
- ・内部エネルギー : $U(T) = \langle E(T) \rangle = = -\frac{\partial \ln Z(T)}{\partial (1/k_B T)}$
- ・エントロピー : $S(T) = \frac{1}{T} \{ U(T) F(T) \}$

次に,具体的に,4個のアミノ酸残基からなるポリペプチド鎖からなる簡単な系での分配関数を求め,あるエネルギー値をもつコンフォメーションが存在する確率や熱力学的量を求めてみよう.

▲ 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

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、この系の 分配関数、あるエネルギー値をもつコンフォメーションの存在確率、及び、熱力学的量を求めよう.

4 個のアミノ酸残基の配列を、具体的に Thr-Val-Val-Gly とする. 図は、この 4 個のユニットが、立 体格子上で取り得る 7 個のコンフォメーションを図示している. ①~⑦のコンフォメーションがもつ エネルギーを、それぞれ、 e_1, e_2, \dots, e_7 とする. この場合、立体格子上で取り得る 7 個のコンフォメ ーション①~⑥がもつエネルギー値は、 $e_1 = e_2 = \dots = e_6 = 0.0$ であり、 コンフォメーション⑦のエネルギー値 e_7 は、アミノ酸残基 Thr と Gly の接触エネルギー値 $e_7 = -0.26\varepsilon_0$ (この経験的なエネルギー値は 第5章で与えられている) となる. ただし、 ε_0 はエネルギー単位を表

この系の分配関数は次のように表現できる:

す.

$$Z = \sum_{i=1}^{7} e^{-e_i/k_B T} = e^{-e_1/k_B T} + e^{-e_2/k_B T} + e^{-e_3/k_B T} + e^{-e_4/k_B T} + e^{-e_5/k_B T} + e^{-e_6/k_B T} + e^{-e_7/k_B T}$$
$$= e^0 + e^{0.26\varepsilon_0/k_B T}$$
$$= e^{0.26\varepsilon_0/k_B T} + 6$$

ところで、7 個のコンフォメーションを二つのエネルギー準位

$$E_1(=-0.26\epsilon_0), E_2(=0.0)$$
で分類すると、右図のように描ける.
この二つのエネルギー準位 E_1, E_2 をマクロ状態のエネルギーと呼び、
 E_1 のマクロ状態には 1 個のミクロ状態、 E_2 のマクロ状態には 6 個の
ミクロ状態が対応しているとみなす. このとき、エネルギー準位数をt、
及び、マクロ状態のエネルギー E_j をもつコンフォメーション(ミクロ
状態)数を $\Omega(E_i)$ で表すと、この系の分配関数は次のようにも書ける:

$$Z = \sum_{i=1}^{1} \Omega(E_i) e^{-E_i/k_B T} = \Omega(E_1) e^{-E_1/k_B T} + \Omega(E_2) e^{-E_2/k_B T} = e^{-E_1/k_B T} + 6e^{0.0} = e^{-E_1/k_B T} + 6e^{0.0$$

ただし, $t = 2, \Omega(E_1) = 1, \Omega(E_2) = 6, E_1 = -0.26\varepsilon_0, E_2 = 0.0$ である.

一般に, t 個の離散的なマクロなエネルギー準位 $E_i(i=1,2,...,t)$ が存在し, エネルギー E_j を もつコンフォメーション(ミクロ状態)数を $\Omega(E_i)$ とすると(縮退しているという), 系の分配 関数は次式のように表現できる:

$$Z = \sum_{i=1}^{t} \Omega(E_i) e^{-E_i/kT}$$

平衡状態で、エネルギー E_1 であるコンフォメーションが存在する確率 p_1 、及び、エネルギー E_2 であるコンフォメーションが存在する確率 p_2 は、それぞれ、
次のように表される:

$$p_1 = \frac{\Omega(E_1)e^{-E_1/k_BT}}{Z} = \frac{e^{-0.26/T}}{e^{-0.26/T} + 6}$$

$$p_2 = \frac{\Omega(E_2)e^{-E_2/k_BT}}{Z} = \frac{6}{e^{-0.26/T} + 6}$$

上式において、次のような無次元量、 $k_B T / \varepsilon_0 = T^* を導入し、$ 改めて、 $T^* \rightarrow T$ とした、また、 $p_1 + p_2 = 1$ が成立している.

右図は、エネルギー E_1 、及び E_2 であるコンフォメーションの 存在確率 p_1, p_2 の温度に対する曲線である.



図より,エネルギー準位で $E_1(=-0.26\varepsilon_0)$ にあるコンフォメーションは,低温では存在する可能性が高いが,温度の上昇とともに存在する可能性が低くなることがわかる.

一方,エネルギー準位で *E*₂(=0.0)にある6個のコンフォメーションは,低温では存在する可能性が低いが,温度の上昇とともに存在する可能性が高くなることを示している.

□ 分配関数から熱力学的量を求める

上記の分配関数から、エネルギーの期待値 < E > (内部エネルギーU)は、次式となる.

$$U = \langle E \rangle = \sum_{j=1}^{t} E_{j} \Omega(E_{j}) P_{j} = \frac{1}{Z} \sum_{j=1}^{t} E_{j} \Omega(E_{j}) e^{-E_{j}/k_{B}T}$$
$$= \frac{1}{Z} \left\{ E_{1} \Omega(E_{1}) e^{-E_{1}} + E_{2} \Omega(E_{2}) e^{-E_{2}/k_{B}T} \right\} = \frac{E_{1} e^{-E_{1}/k_{B}T}}{e^{-E_{1}/k_{B}T} + 6}$$

このエネルギーの期待値 < E > を規格化した次の量 θ を導入する: $\theta = \frac{\langle E \rangle}{E_{\min} - E_{\max}}$

上の系では、 $E_{\min} = E_1, E_{\max} = 0.0$ より、 $\theta = \frac{e^{-E_1/k_B T}}{e^{-E_1/k_B T} + 6}$

第3章 タンパク質分子はボルツマン分布など聞いたこともない!

上のθの式は温度 Tの関数であり、平衡における転移曲線とみなせる.この曲線を描いてみよう(結 果的に *p*₁ と同じ曲線である).

 $E_1 = -0.26\varepsilon_0$ (ε_0 :エネルギー単位)であり,無次元量 $k_B T / \varepsilon_0 = T^* を導入して,改めて. <math>T^* \rightarrow T$ とおくと, θ は次式となる:

$$\theta = \frac{e^{-0.26/T}}{e^{-0.26/T} + 6}$$

この転移曲線は、図のように、協同的な構造転移を表しており、S字状曲線 (sigmoidal transition curve)となる.

ここで、 $\theta = 0.5$ に対応する温度(転移温度) T_m を求める:

$$\frac{e^{-0.26/T_m}}{e^{-0.26/T_m}+6} = \frac{1}{2}$$

より、
$$T_m = \frac{0.260}{\ln 6} = 0.145$$
 となる.

次に、比熱 C_v を求めよう.

$$C_{V} = \left(\frac{\partial \langle E \rangle}{\partial T}\right)_{V} = \frac{6E_{1}^{2}}{kT^{2}} \cdot \frac{e^{-E_{1}/k_{B}T}}{(e^{-E_{1}/k_{B}T} + 6)^{2}}$$

同様に, $E_1 = -0.26\varepsilon_0$, $k_B T / \varepsilon_0 = T^*$ を導入し, 改めて $T^* \rightarrow T$ とすると, 上式の比熱 C_V は, 次式であらわさ れる (図の比熱曲線参照);

$$C_V / k_B = \frac{6(0.26)^2}{T^2} \cdot \frac{e^{-0.26/T}}{(e^{-0.26/T} + 6)^2}$$

転移温度付近での大きな比熱の変化を観測することがで きる.

<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卷, pp. 103-109, 2006.


3 次元格子タンパク質を採用し、フォールディングに対する「統計力学モ デル」を導入しよう!

タンパク質分子のフォールディング過程の様相を理解しようと思えば,全ての構成原子を考慮して, さまざまなコンフォメーションのポテンシャル・エネルギーを求め,そのポテンシャル面を描くこと が王道であろう.

しかしながら、その自由度の大きさ故、実際のタンパク質分子で網羅的な計算を実行することは現 在のコンピュータの性能では実現不可能であろう(将来、"量子コンピュータ"が実用化されると、こ のことが実現できるだろう).

そこで次善の策として、複雑なタンパク質を思い切って簡略化することを考えよう.タンパク質分子を構成しているアミノ酸残基をユニット(球とみなす)のつながりとし、しかもそれぞれのユニットは、3次元の格子上のみしか動くことができないとする模型を導入しよう.タンパク質分子を3次元格子上という限られた空間におき、模型へ組み込む要素をいろいろと考慮して、格子タンパク質分子のコンフォメーション変化のシミュレーションを実行して、統計熱力学量を求めて議論する方法を考えよう.

このアプローチの仕方は、系を少数のパラメータで制御することができ、しかも、物理的な意味が 明瞭であり、さらに、あらゆる場合について網羅的に検討することが可能であるなどの利点がある.

次章で,格子タンパク質のフォールディングに関する統計力学モデルを構築し,それによって,フ ォールディングに関する熱力学量を理論的に求める方法を考察しよう.

そして,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る熱力学量と比較・検討し,その統計力学モデルの妥当性を検討した後,実際のタンパク質にその統計力学モデルを 適用して実験結果と比較・検討するのが,我々の戦略(strategy)である.



" Chatsworth at Dawn "



□ 平衡状態で、タンパク質分子があるエネルギー値をもつコンフォメーションをとる確率は、ボルツマン分布で表される!

タンパク質分子は、与えられた環境の中で、自分の"形"や"構造"を、自発的に揺らがせている 物理・化学の法則に従う分子機械である.平衡状態で、タンパク質分子が、あるエネルギーをもつコ ンフォメーションを実現する確率は、ボルツマン分布で記述されることを Dill らに準じて示そう.

<文献> Dill, K.A. & Bromberg, S., Molecular Driving Forces, Statistical Thermodynamics in Chemistry and Biology. Garland Science, New York, 2003.

1個のタンパク質分子が生理的環境条件に置かれているような系を考える.この系の温度 *T*,体積 *V*は一定とする.いま、1個のタンパク質分子は、多くのコンフォメーション(ミクロな状態)をと る.取り得るコンフォメーションの総数を *N*(一定)とし、それぞれのコンフォメーション・エネ ルギーを $E_j(j=1,2,3,\dots,N)$ とする.我々が知りたいのは、平衡状態で、エネルギーが E_j であるコン フォメーション*j*が存在する確率 P_i は、どのような式で表されるのかである.

いま,系の自由エネルギーをFとすると,F = U - TS である.ここで,Uは,熱力学的な内部 エネルギーで,ミクロな立場からすると,エネルギー $E_i(j = 1, 2, 3, \dots, t_N)$ の期待値である:

$$U = \langle E \rangle = \sum_{j=1}^{N} E_{j} p_{j}$$
.

Sは、統計力学的エントロピーであり、次式のように表される: $S = -k_B \sum_{j=1}^{N} p_j \ln p_j$

この式は、エントロピーの定義式: $S = k_B \ln W$ から導かれる(章末の「Coffee Break」参照). ここで、W は微視的状態の場合の数、 k_B は、ボルツマン定数である; $k_B = 1.380662 \times 10^{-23} J K^{-1}$. また、すべてのコンフォメーションが存在する確率は1であることより、

$$\sum_{j=1}^{N} p_j = 1$$

結局,平衡状態でエネルギー E_j ($j = 1, 2, 3, \dots, N$)をもつコンフォメーションが存在する確率を求める問題は,次のように定式化できる:

『系の自由エネルギー系をFとすると、平衡条件は、dF=0であるから、結局、条件式:

$$\sum_{j=1}^{N} p_{j} = 1$$
のもとで、自由エネルギー F の極値を求める.』

"ラグランジュの未定乗数法"を用いて、平衡状態でエネルギー E_j をもつコンフォメーションの存在確率を求めよう(Appendix E:「ラグランジュの未定乗数法とは?」参照):

未定乗数 α を用いて、 $f = F + \alpha(\sum_{j=1}^{N} p_j - 1)$ とおくと、f は次式となる:

$$f = U - TS + \alpha(\sum_{j=1}^{N} p_j - 1) = \sum_{j=1}^{N} E_j p_j + k_B T \sum_{j=1}^{N} p_j \ln p_j + \alpha(\sum_{j=1}^{N} p_j - 1)$$

極値条件より,

$$df = \sum_{j=1}^{N} (E_j dp_j + p_j dE_j) + k_B T \sum_{j=1}^{N} (1 + \ln p_j) dp_j + \alpha \sum_{j=1}^{N} dp_j = 0$$

一般的に、微視的エネルギー E_j は、エントロピー S と温度 T に依存しないが、体積 V とコンフォメーションの数 N に依存するので、次式が成立する:

$$dE_{j} = \left(\frac{\partial E_{j}}{\partial V}\right) dV + \left(\frac{\partial E_{j}}{\partial N}\right) dN$$

この系では、体積 V , コンフォーションの数 N は一定であると仮定しているので、結局、 $dE_j = 0$ となる.よって、上式のdfの式をまとめると、

$$df = \sum_{j=1}^{N} \{ E_j + k_B T (1 + \ln p_j) + \alpha \} dp_j = 0$$

上式は、jの各項について、 $E_i + k_B T (1 + \ln p_i) + \alpha = 0$ となる. この式より、

$$\ln p_j = -\frac{E_j}{k_B T} - \frac{\alpha}{k_B T} - 1 \qquad \therefore \quad p_j = e^{-E_j/k_B T} e^{(-\alpha/k_B T) - 1}$$

ここで、
$$\sum_{j=1}^{N} p_j = 1$$
 を用いると、 $e^{(-\alpha/k_B T)-1} = \frac{1}{\sum_{j=1}^{N} e^{-E_j/k_B T}}$

故に, 次のような"ボルツマンの分布則"(Boltzmann distribution law)が得られる:

$$p_j = \frac{e^{-E_j/k_BT}}{Z}$$
 $(j = 1, 2, 3, \dots, t_c, Z = \sum_{j=1}^N e^{-E_j/k_BT})$

この式の分子の項 e^{-E_j/k_BT} は、"ボルツマン因子" (Boltzmann factor) と呼ばれる量で、次のような重要な特徴をもっている:

第3章 タンパク質分子はボルツマン分布など聞いたこともない!

- ・より低いエネルギーをもつコンフォメーションが実現しやすい,つまり,エネルギーの高い コンフォメーションほど存在する確率は,指数関数的に減少する.
- ある温度で存在する可能性が低いコンフォメーションでも、温度の上昇とともに、存在する 可能性が高くなる。

また,ボルツマン分布則の分母の項 Z は系の"分配関数"(状態和)と呼ばれる量で,この分配 関数は,ポリペプチド鎖が取り得るコンフォメーション全てについての和である.

分配関数 Z は,温度 T の関数であり、しかも、微視的な個々の状態と、巨視的な熱力学的量とを結 ぶ重要な関数で、いろいろな物理量を計算する際の基点になる.ちなみに分配関数が求まると、これ を用いて、熱力学的量である、自由エネルギー、内部エネルギー、エントロピーが、次のように計算 することができる(温度の関数として表現している).

・自由エネルギー : $F(T) = -k_B T \ln Z(T)$

・内部エネルギー : $U(T) = \langle E(T) \rangle = - = -\frac{\partial \ln Z(T)}{\partial (1/k_B T)}$

・エントロピー : $S(T) = \frac{1}{T} \{ U(T) - F(T) \}$ Coffee Break <エントロピーの式: $S = k_B \ln W$ から、 $S = -k_B \sum_{i=1}^{N} p_i \ln p_i$ を導く> いま、区別できないN個の粒子をN個のエネルギー準位に配置する.1番目の準位にはn,個、2 番目の準位にはn,個,・・・, N 番目の準位にはn_N個配置するとき, N 個の粒子をこのように配置 する場合の数Wは次のようになる: $W = \frac{N!}{(n_1!)(n_2!)....(n_N!)}$ ここで、スターリング (Stirling) の近似式 $x! \approx (\frac{x}{e})^x$, 確率の式 $p_i \approx \frac{n_i}{N}$ を用いると、上式は 次のようになる: $W = \frac{(N/e)^{N}}{(n_{1}/e)^{n_{1}}(n_{2}/e)^{n_{2}}...(n_{N}/e)^{n_{N}}}$ $=\frac{(N)^{N}}{(n_{1})^{n_{1}}(n_{2})^{n_{2}}...(n_{N})^{n_{N}}}=\frac{1}{p_{1}^{n_{1}}p_{2}^{n_{2}}...p_{N}^{n_{N}}}$ 上式の対数をとると $\ln W = -\sum_{i=1}^{N} n_i \ln p_i$ 両辺をNで割ると, $\frac{1}{N}\ln W = -\sum_{i=1}^{N} p_i \ln p_i$ $S = -k_B \sum_{i=1}^{N} p_i \ln p_i$



□ ラグランジュの未定乗数法とは?

ラグランジュ(Lagrange)によって開発された,関数の条件付極値を求める方法である.次のような, 簡単な具体例で,ラグランジュの未定乗数法による解法を示そう.

例題:条件 $x^2 + y^2 = 1$ のもとで,関数xy の極値(最大値,最小値)を求めよ. 解:(ラグランジュの未定乗数法による解法)

 $F = xy + \alpha(x^2 + y^2 - 1)$ とおく. 極値条件より,

$$\frac{\partial F}{\partial x} = y + 2\alpha x = 0$$
 $\frac{\partial F}{\partial y} = x + 2\alpha y = 0$

上式より、y を消去すると、 $x+2\alpha(-2\alpha x)=0$ これより、

x=0 (このときには, y=0)のほかに, x と y が解をもつためには, 未定係数 α が,

 $1-4\alpha^2=0$, $\therefore \alpha=\pm\frac{1}{2}$ orther orthogram.

したがって極値は、 $y \pm x = 0$ の上にあり、これと、条件式 $x^2 + y^2 = 1$ とを連立させて求める ($x^2 + y^2 = 1 \ge z = xy$ の等高線の下図を参照).

 $(1/\sqrt{2}, 1/\sqrt{2}), (-1/\sqrt{2}), -1/\sqrt{2})$ で 最大値 1/2

$$(1/\sqrt{2},-1/\sqrt{2}),(-1/\sqrt{2}),1/\sqrt{2})$$
 で
最小値 $-1/2$





"Autumn Tree under Moon"

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>

第4章

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_NILS 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"NILS モデル"(Non-Interacting Local Structure model, 1981年)の拡張版である.

"A-W_NILS 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 ラメターを含まない統計力学モデルであることを強調し ている. 第4章 タンパク質フォールディング過程の統計力学モデルの導入

耳 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、"簡単で計算可能な統計力学モデル"を構築できないだろうか?

その簡単で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.

耳 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か関わっており、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.

38

- <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. Saitô, 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. Saitô, 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 & Saitô (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 第4章 タンパク質フォールディング過程の統計力学モデルの導入

のみに働くと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 をとる必要があろう.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"Go 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 オ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.



第4章 タンパク質フォールディング過程の統計力学モデルの導入

↓ 統計力学モデルである"NILS 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 動くとすることである.3次元格子模型を採用する利点は、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 パラメター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 NILS 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:

- 「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とり、 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

 $(i = 2, 3, 4, \cdot \cdot \cdot, n - 2).$

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、非天然接触エンルギーは全て 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 オ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

だけではなく,他の相互作用が存在して Residue Number No 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卷, pp. 49-58, 2007.



第4章 タンパク質フォールディング過程の統計力学モデルの導入

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, … *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):

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

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;

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

- <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.
- Γ_{k1}: 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;

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

 $\Gamma_{i,i} = 0$: 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* ≥ 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≤*i*≤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).だだし、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、 $i \ge 18$ については、 $4 \le i \le 17$ 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卷, 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ター(調整可能なパ ラメター, adjustbale 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, αⁿ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;

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

実際に4個のユニットから成る分配関数;

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

○ j=5のとき; $Z_{1,5} = \alpha Z_{1,4} + \alpha \sum_{m=1}^{2} e^{-\beta E(m,5)} Z_{1,m+1}$ $= \alpha \left\{ \alpha^{2} + \alpha e^{-\beta E(1,4)} \right\}$ + $\alpha \left\{ e^{-\beta E(1,5)} Z_{1,2} + e^{-\beta E(2,5)} Z_{1,3} \right\}$



0.0

 $\alpha(\alpha^0)$ 0.0

 $\alpha^2(\alpha)$ 1.0

E(1:4)

1.0

0.0

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

実際に5個のユニットから成る分配関数; 2 (- $-\beta E(1,4)$ BE(2,5)BE(1.5)

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

$$0.0 \quad E(1:4) \quad E(2:5) \quad E(1:5) \qquad 0.0$$

$$\alpha(\alpha^{0}) \quad 0.0 \quad 0.0 \quad 1.0 \quad 1.0 \quad 0.0 \quad 1.0$$

$$\alpha(\alpha^{0}) \quad \alpha(\alpha^{0}) \quad 1.0 \quad 1.0 \quad 0.0 \quad 1.0 \quad 0.0$$

]

$$\bigcirc \mathbf{j} = 6 \ \mathcal{O} \succeq \stackrel{*}{\succeq} ; \qquad \alpha^{3}(\alpha^{2}) \ 1.0$$

$$Z_{1,6} = \alpha Z_{1,5} + \alpha \sum_{m=1}^{3} e^{\beta E(m,6)} Z_{1,m+1}$$

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

$$+ \alpha \left[e^{-\beta E(1,6)} Z_{1,2} + e^{-\beta E(2,6)} Z_{1,3} + e^{-\beta E(3,6)} Z_{1,4} \right]$$

$$= \alpha^{4} + \alpha^{3} \left\{ e^{-\beta E(1,4)} + e^{-\beta E(2,5)} + e^{-\beta E(3,6)} \right\}$$

$$+ \alpha^{2} \left[e^{-\beta E(1,5)} + e^{-\beta E(2,6)} + e^{-\beta \left\{ E(1,4) + E(3,6) \right\}} \right]$$

実際に6個のユニットから成る分配関数;

$$\overrightarrow{Z}_{1,6} = \alpha^{3} + \alpha^{2} \left\{ e^{-\beta E(1,4)} + e^{-\beta E(2,5)} + e^{-\beta E(3,6)} \right\}$$

$$+ \alpha \left[e^{-\beta E(1,5)} + e^{-\beta E(2,6)} + e^{-\beta \left\{ E(1,4) + E(3,6) \right\}} \right]$$

$$+ e^{-\beta E(1,6)}$$



 $\alpha^{3}(\alpha^{2})$ 1.0

のとき

但し、E(1:4)=E(2:5)=E(1:5)=0.0

0.0

0.0

0.0

□□ 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、エン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;

 $E_h = h\varepsilon_0$ (ここで, h は整数を表し, ε_0 は $\varepsilon_0 = 0.01$ である)

変数 t は、天然状態にあるアミノ酸残基の数 η ($0 \le \eta \le n-3$) 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, $W_1(h) = \sum_{n=0}^{n-3} \Omega(\eta, h) t^{\eta}$, $u = \exp(-\beta \varepsilon_0)$

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<E >を求めると、

$$< E >= rac{\displaystyle \sum_{h} E_{h} W_{1}(E_{h}) e^{-\beta E_{h}}}{Z}$$

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, $u = \exp(-\beta \varepsilon_0)$

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

第4章 タンパク質フォールディング過程の統計力学モデルの導入

られる (パラメター t を 1 を置く);

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

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"



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、NILS 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 "NILS 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 NILS 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、NILS 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、 $\beta \sim n$ 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、 $\beta = 1/k_BT$ 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、 $-k_BT\ln\sigma$ 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;

$$Z \leftarrow \sigma^{i_{1}-1} e^{-\beta F(i_{1},i_{2})} \sigma^{i_{3}-i_{2}-1} e^{-\beta F(i_{3},i_{4})} \sigma^{i_{5}-i_{4}-1} e^{-\beta F(i_{5},i_{6})} \sigma^{n-i_{6}} / \sigma^{n}$$

 $\sigma^{n} = \sigma^{i_{2} + (i_{4} - i_{2}) + (i_{6} - i_{4}) + (n - i_{6})}$

第4章 タンパク質フォールディング過程の統計力学モデルの導入

より,

$$Z \leftarrow \frac{\sigma^{i_{1}-1}}{\sigma^{i_{2}}} e^{-\beta F(i_{1},i_{2})} \frac{\sigma^{i_{3}-i_{2}-1}}{\sigma^{i_{4}-i_{2}}} e^{-\beta F(i_{3},i_{4})} \frac{\sigma^{i_{5}-i_{4}-1}}{\sigma^{i_{6}-i_{4}}} e^{-\beta F(i_{5},i_{6})} \frac{\sigma^{n-i_{6}}}{\sigma^{n-i_{6}}}$$
$$= (\sigma^{i_{2}-i_{1}+1})^{-1} e^{-\beta F(i_{1},i_{2})} (\sigma^{i_{4}-i_{3}+1})^{-1} e^{-\beta F(i_{3},i_{4})} (\sigma^{i_{6}-i_{5}+1})^{-1} e^{-\beta F(i_{5},i_{6})}$$
$$= \prod_{k=1,3,5} (\sigma^{i_{k+1}-i_{k}+1})^{-1} e^{-\beta F(i_{k},i_{k+1})}$$

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;

$$\begin{split} Z &= (\sigma^4 e^{-\beta E_0} + \sigma^0 e^{-\beta E_1}) / \sigma^4 = (\sigma^4 + e^{-E_1/kT}) / \sigma^4 \\ &= 1 + \frac{e^{-\beta E_1}}{\sigma^4} \\ \vdots \vdots \vdots \vdots , \quad \beta &= \frac{1}{k_B T} \end{split}$$



 $E_0 = 0.0$ (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の表す. 上記の分配関数から,系のエネルギーの期待値 < E > を求めると,

$$< E >= \sum_{j=1}^{2} p_{j} E_{j_{j}} == \frac{E_{0} \times 1 + E_{1} \times \frac{e^{-\beta E_{1}}}{\sigma^{4}}}{Z} = \frac{E_{1} \frac{e^{-\beta E_{1}}}{\sigma^{4}}}{1 + \frac{e^{-\beta E_{1}}}{\sigma^{4}}} = \frac{E_{1} e^{-\beta E_{1}}}{\sigma^{4} + e^{-\beta E_{1}}}$$

このエネルギーの期待値 < E > 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 < E > 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:

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

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, $Z \equiv Z_{l,n}$ 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卷, pp. 103-109, 2006.

E(m,j) は、次式のような、m 番目 $\sim j$ 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; $E(m, j) = \sum_{m \le k, l \le j} U(\xi_k, \xi_l) \Gamma_{k, l}$

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;

 $\Gamma_{i,i} = 1$: 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、 $f(m, i) = \alpha^{i-3}$

 $(i = j - m + 1, i \ge 4)$ 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 *a* を調整可能なパラメ 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.

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

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

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

 $i \ge 18$ 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、調節パラメター (adjustable parameter) は含まれていないことになる.

💶 "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 ; $Z = \sum_{n} \sum_{h} \Omega(\eta, h) t^{\eta} u^{h}$

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 ($0 \le \eta \le n-3$) 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 オ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

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

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 <E> を求めると,

$$< E > = rac{\sum_{h} E_{h} W_{1}(E_{h}) e^{-\beta E_{h}}}{Z}$$

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,

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

 $W_2(\eta)t^{\eta}$ 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);

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



"Red House in Winter"



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よう!

□ 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 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.



NKTVVGEPWH CLLFPRRDKN QMSYLTGIAG EDSAAI



Protein a2



SQKWLERGAT RIADGDLPVN GTYFSCKIME NVHPLA



TSKRQQPYPM SLGSPFIRIP MIGPRPRMRL LILLMGYPKR GRSGGGLF

TEKGEEGYGG AAWTGPTSYK MAIYVWTTMW IYWAWAEAKK YGAYWAYM

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 al 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:

- タンパク質分子のフォールディング・シミ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 < E > を求める.
- ② エネルギーの期待値 < $E > \delta$, 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 al 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卷, 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:

O「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,

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

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 < E > を求める ($\beta = 1/k_BT$);

$$\langle E(T) \rangle = rac{\sum_{h} E_{h} W_{1}(E_{h}) e^{-\beta E_{h}}}{Z}$$

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

〇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:

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、10⁸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 < E > 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< E > 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.

〇く転移温度の比較>

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 al, a2, bl 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.

〇<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, < $\diamond \theta_h^{sim}(T)$ 曲線>と< $\diamond \theta_{NC}^{sim}(T)$ 曲線>は, タンパク 質 a1, b2 でほぼ一致している. タンパク質 a2, b1 では不一致の領域がみられるが, これは,

< *θ*^{sim}_{non-NC}(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< $igoplus heta_h^{sim}(T)$ 曲線>は、タンパク質 a1,b1 に対する転移曲線< $igoplus heta_h^{sim}(T)$ 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 al, bl は比較的よく一致しているが、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 <一 $\theta_h^{th}(T)$ >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 るく ◆ $\theta_h^{sim}(T)$ 曲線>とは、タンパク質 al, a2, bl, 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、 <● θ_{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, <● θ_{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. 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.



64

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、 $\varepsilon_{int ra-N}, \varepsilon_{N-N}, \varepsilon_{N-D}, \varepsilon_{D-D}$ に分類しよう:



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

図には、「A-W_NILS モデル」から得られる転移曲線、 $<-\theta_h^{th}(T)$ 曲線>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 $< \blacktriangle \ \theta_{NC}^{sim}(T)$ 曲線>、つまり、次のような量:

$$\theta_{\scriptscriptstyle NC}^{\scriptscriptstyle sim}(T) = \frac{1}{E_{\scriptscriptstyle native}} \cdot < E_{\scriptscriptstyle NC}(T) >$$

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、 $\theta_{NC}^{sim}(T)$ を次の4個のタイプに分類しよう:

(1)< $\theta_{int ra-N}^{sim}(T)$ 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、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);

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

(2)<● $\theta_{N-N}^{sim}(T)$ 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,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);

$$\theta_{\scriptscriptstyle N-N}^{\rm sim}(T) = \frac{1}{E_{\scriptscriptstyle native}} \cdot < E_{\scriptscriptstyle N-N}(T) >$$

(3)< $\theta_{N-D}^{sim}(T)$ 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、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. 〇 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、 < $-\theta_h^{th}(T)$ 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) <▲ θ^{sim}_{NC}(T)曲線>と<■ θ^{sim}_{int ra-N}(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) <---θth_h(T)曲線>と<■ θ^{sim}_{int ra-N}(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 al, 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, <- $\theta_h^{th}(T)$ 曲線>と, <■ $\theta_{int m-N}^{sim}(T)$ 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 al 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. 〇<「A-W_NILS モデル」の妥当性と限界>

タンパク質 al 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.

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よう!

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 al 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 al 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 al 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< $-\theta_h^{th}(T)$ 曲線>と,シ ミュレーションによる< $\phi_h^{sim}(T)$ 曲線>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"



👪 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*pi(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 al, 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らない.

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 bl 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れ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が).

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"



💶 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さを残して単純 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,そのユニイトが連結してい て,折れ線からなるユニットは,格子上の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簡単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.



💶 タンパク質分子の格子模型の発展

1970代年から1980年代初頭にかけてGo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、Go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 代年から 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 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よう!

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次元の立方格子上に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*=±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:

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

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、 $U(\xi, \xi)$ はアミノ酸残基 *ξ* と *ξ* との相互作用エネルギーの値である。

 $U(\xi_i, \xi_i)$ 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 al, Protein bl, 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. 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よう!

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(Aeg), 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 al 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.
- ③ 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、

 $e^{-kT} > r$ 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、10⁸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.



- <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卷, 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.

(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.

(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



 e_4 e_4 e_6 e_3

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.









- <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卷, pp. 49-58, 2007.



"Bridge Across Stream in Winter"

アミノ酸置換によって熱力学量は どのように変化するのだろうか?

<要旨>

第6章

2個の天然構造と、それぞれ2個のアミノ酸配列をも つ3次元格子タンパク質のそれぞれで、野生型の各アミ ン酸残基を他の19種類のアミノ酸残基に置換する(変異 体と呼ぶ)ことによって生じる、2つの熱力学量の変化、 フォールディングの転移温度の変化とコンフォメーショ ン・エネルギーの変化とが、強い相関をもっていること を示している.

さらに、アミノ酸配列上の各部位は、アミノ酸置換に 対してある種の感受性 (susceptibility)をもっていると すると、各部位の感受率は天然構造によって基本的に決 まっていることも示している.

また,野生型と変異体に対する,フォールディングの 度合いをみるための秩序度(反応座標)に対する自由エ ネルギー関数の差から,天然状態にあるアミノ酸数の割 合を,秩序度の関数として,各アミノ酸残基毎に求める 方法を示している.この量は,ある秩序度において,各 アミノ酸残基が天然状態と同じ構造を形成しているかど うかを表す量とみなせる.

具体的に、3次元格子タンパク質に対して、各アミノ 酸残基毎にこの量を計算して、反応座標の各点で、各々 のアミノ酸残基がどの程度天然状態であるかを求め、 フォールディングの統計的道筋を議論している.

📮 変異体タンパク質とは?

生物の進化において, DNA の塩基配列がある確率で変化(変異)することは避けがたいだろう. そして,その結果として,タンパク質のアミノ酸配列も変化する.生物の生存にとって有利な変化を 獲得することが,生物の進化における基本的な出来事であろう.しかしそれだけではなく,自然選択 に有利でも不利でもない中立的な変化も起こることが知られており,それがアミノ酸配列の多様化を 引き起こしている.

一方,物理学的観点からすれば、1つのアミノ酸を他のアミノ酸に置換するという変化は、タンパク質に加えられた一種の"摂動"とみなすことができる.この"摂動"による変化を詳細に調べることによって、ある系の本質を理解する手助けになるだろう.

もとの3次元格子タンパク質(野生型タンパク質と呼ぶ)の,ある部位のアミノ酸残基を他のアミノ酸残基に置換して,新たな3次元格子タンパク質(これを変異体タンパク質と呼ぶ)を作成する. このとき,3次元格子タンパク質のフォールディングに関する熱力学量がどのように変化するかを調べてみることは大変興味深い.

しかし,野性型タンパク質の,ある部位のアミノ酸残基を他のアミノ酸残基に置換して作成した変 異体タンパク質に対して,その都度コンピュータ・シミュレーションを実行して,フォールディング に対する熱力学量を求めて,その量がどのように変化するか,つまり摂動に対する応答をチェックす ることは,膨大な計算時間が必要となり,現実的ではない.

それ故, すべての変異体タンパク質のフォールディングに対する熱力学量を, 統計力学モデルである「A-W_NILS モデル」を用いて敏速に計算しよう. 特に, 2 つの熱力学量の変化である,「エンタルピーの変化」と「転移温度の変化」を計算する. ただし, ある部位を他のアミノ酸残基で置換した 新たなタンパク質(変異体タンパク質)が, 確実に天然構造へとフォールドするかどうかのチェック までは踏み込まず, ある部位を他のアミノ酸残基で置換した新たなタンパク質(変異体タンパク質) は, もとの野性型のタンパク質の小さな摂動とみなして, つまり, 変異体タンパク質も野生型タンパ ク質と同じ天然構造へとフォールドするとみなして, アミノ酸置換によって引き起こされる熱力学量 の変化を, とりあえず定性的な結果として眺めてみることにしよう.

<文献> H. Wako and H. Abe, Single Amino Acid Substitutions in Lattice Proteins Using Statistical Mechanical Model for Protein Folding, J. Phys. Soc. Jpn. Vol.76, pp. 104801-1, 10480-2, 2007.

<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, Ⅲ. アミノ酸置換による熱力学 量の変化とフォールディング過程",西日本工業大学紀要,第38卷, pp.93-101, 2008.



"First Sunlight"

🎞 アミノ酸置換による、エンタルピー変化と転移温度変化

下図の左側の上段は、タンパク質 al, a2, bl, b2 に対して、それぞれの部位で、そのアミノ酸残 基を他の "19 種類のアミノ酸残基に置換"したときの天然構造全体のエンタルピーの変化(ΔE) を表し、下段は、その時の転移温度の変化 ΔT_m を示している (T_m ; 転移温度).

右側の立体構造の図は,天然立体構造上で,変化が大きいアミノ酸残基に色をつけたものである. 濃い赤色程,変化が大であることを示している.

エンタルピーの変化 ΔE , 及び, 転移温度の変化 ΔT_m に対する, アミノ酸残基部位の依存性をみて みよう. タンパク質 *al* と *a2* では, かなり異なっている.



(a1)





第6章 アミノ酸置換によって熱力学量はどのように変化するのだろうか?

タンパク質 $b1 \ge b2$ も、かなり異なっている.これは、タンパク質 $a1 \ge a2$ 、 $b1 \ge b2$ は、同じ天然構造であるが、それぞれのアミノ酸配列が異なるので、同じ部位における相互作用の値が異なることになり、その結果、エンタルピーの変化 ΔE 、及び、転移温度の変化 ΔT_m も異なっていると考えられる.

GI Y - CYS MFT PHF ILE --LEU -VAL - TRP -- TYR ALA THR PRO SER GLN ASN GLU ASP HIS ARG -LYS 6.0 4.0 $\Delta E(\xi_i)$ 2.0 0.0 -2.0 8 12 16 20 24 28 32 36 0 4 40 44 48 0.00 0.04 (يَّرُ -0.08 -0.12 8 12 16 20 24 28 32 36 40 44 48 0 4 Resdidue Number (b2)I FU - CYS MET PHF ILE VAI TRP TYR ALA GLY THR SER GLN ASN GLU ASP - HIS ARG LYS PRO 6.0 4.0 $\Delta E(\xi_i)$ 2.0 0.0 -2.0 0 8 12 16 20 24 28 32 36 40 44 48 4 0.00 0.04 (بَيْ) سلس (كَنْ -0.08 -0.12 0 8 12 20 24 28 32 36 40 44 48 4 16 Residue Number

(b1)

💶 エントロピー変化と転移温度変化の相関

下図上段は、タンパク質 a1, a2, b1, b2 に対する AE とAT_m の相関を表している.

多くの部位で ΔT_m と ΔE は相関している. つまり,転移温度の変化は,基本的に天然構造のエンタルピー(内部エネルギー)変化とよい相関がある. しかも,タンパク質 *a1* と *a2* の,両方の,ほとんどの部位でよく一致していることがわかる.

ただし、両末端など一部の部位では、 ΔT_m と ΔE は比例せず ΔE がある程度以上大きくなると、 ΔT_m が一定値となる傾向がみられる.

アミノ酸残基部位毎の $\Delta E \ge \Delta T_m$ との回帰係数 rをプロットしたのが、図の中段であるが、 ΔT_m が ΔE に比例する部位は、その傾きから、アミノ酸残基を 3 つのグループに分類することができる.

アミノ酸残基部位を上の3つに分類し、それを実際の立体構造上に表示してみると(図の下段), タンパク質 *a1* と *a2* は、きわめてよく似ていて、しかもドメインを形成していることがわかる.明 らかに両者で異なる部位は、3,13,14,18,25,31の6個の残基のみである.



第6章 アミノ酸置換によって熱力学量はどのように変化するのだろうか?

一方, タンパク質 $b1 \ge b2$ の場合, グループ $1 \ge 7$ では, 48 個のアミノ酸残の内 20 個の残基が異なるグループに属していることがわかる. しかしながら, グループ $1 \ge 7$ のグ ループ分けはかなり微妙で, $\Delta E \ge \Delta T_m$ との回帰係数 r の分類基準を少し変更して, グループ 2 の 15, 17, 18, 21, 31 をグループ 1 に分類してみると, タンパク質 $b1 \ge b2$ のグループ分けもかなり一致してくるのがわかる.



これらの結果から次のことが推察できる:

『アミノ酸配列上の各部位は、アミノ酸置換に対してある種の感受性 (susceptibility)をもっているとすると、各部位の感受率は天然構造によって基本的に決まっていることを示している.』

ただし、同じ天然構造をもっていても、アミノ酸配列の違いによって、その感受率が異なる部位も ある.

現実のタンパク質でも、アミノ酸置換による疎水性の変化(ΔE に対応)と安定性の変化(ΔT_m に対応)によい相関があるという報告もある一方、そうした相関を見出すことができなかった例も報告されており、一般に、アミノ酸置換による安定性の変化を、正確に予測することはかなり困難であるというのが現状である.

□ フォールディング過程において,各アミノ酸残基は,どの程度,天然状態を 形成しているのだろうか?

タンパク質が取り得るすべてのコンフォメーションを天然状態にある残基数 n で分類し,温度 T での,天然状態にある残基数 n に対する自由エネルギー F(n,T) を考える.

ここで、天然状態にあるアミノ酸残基数の割合 η (=n/(N-3), $0 \le \eta \le 1$)を導入する. η は、 フォールディングの度合いを見るためのある種の秩序変数(反応座標)とみなすことができる.

いま,野生型タンパク質と,その変異体タンパク質の i番目のアミノ酸残基のタイプを,それぞれ, w_i , ξ_i とするとき,次のような量を導入する;

$\phi(\eta,\xi_i) = \frac{F(\eta,\xi_i) - F(\eta,w_i)}{F(1,\xi_i) - F(1,w_i)}$

ここで、上式の分子の項、 $F(\eta, \xi) - F(\eta, w_i)$ は、*i* 番目のアミノ酸残基のタイプを w_i (野性型の残基) から、 ξ_i (変異体の残基)へ置換したときの、与えられた η における野生型と変異体の自由エネル ギーの差を表している.上式の分母の項、 $F(1, \xi) - F(1, w_i)$ は、*i* 番目のアミノ酸残基のタイプを野生 型 w_i から、変異型 ξ_i へ置換したときの、 $\eta = 1$ 、つまり天然状態における野生型と変異型の自由 エネルギーの差を表している.

下図は,例として,秩序変数 ŋに対する次のような量の曲線を描いている;

・野生型で、あるアミノ酸残基がヒスチジン(His)のときの自由エネルギー曲線(赤色)

・このアミノ酸残基を、システイン(Cys) に置換した変異型の自由エネルギー曲線(青色)



第6章 アミノ酸置換によって熱力学量はどのように変化するのだろうか?

ϕ(*η*, *ξ*) 値を使えば遷移状態ばかりでなく,反応座標 *η* の任意の値で,各アミノ酸残基が部分的に 天然状態と同じ構造を形成せているかどうかを見積もることが可能であり,タンパク質のフォールデ ィング過程全般を追跡できる量とみなすことができる.

 $\phi(\eta, \xi)$ が1に近ければ近いほど,残基*i*周辺は,反応座標 η で,部分的に天然状態に近いと考えられ, $\phi(\eta, \xi)$ が0に近ければ近いほど,反応座標 η で,部分的に天然状態とはほど遠い状況にあることになる.

この Ø値は、Fershtらによって導入された、遷移状態での各アミノ酸残基が天然状態かどうかを 推定する量、いわゆる、" Φ値 "に対応する量を拡張したものになっている. この " Φ値 "は、タ ンパク質のフォールディング過程を検証するのに重要な量である. 第8章で、Φ値の定義と3次元格 子タンパク質に対するΦ値の結果を示している. 実際のタンパク質に対するΦ値の結果は、第9章~ 第11章で取り上げている.

耳 φ値により求められたタンパク質のフォールディング過程

タンパク質 al, a2, b1, b2 に対して, 各アミ ノ酸残基が, 反応座標 η で, 部分的に天然状態と 同じ構造を形成している かどうかを推定する量 $\phi_i(\eta)$ を計算しよう. そ の結果が次の図である. これにより, それぞれの タンパク質のフォールデ ィング過程を検証しよう.

 η が小さいとき,一部 のアミノ酸残基を除いて, $\phi_i(\eta)$ は一般にかなり小 さく,タンパク質はD状 態にあることを示してい る.

しかし,一部のアミノ 酸残基部位については, $\phi_i(\eta)$ は小さいながらも はっきりとしたピークを もち,完全なランダムで はなく,N状態がときお り出現していることがわ かる.

そして η が大きくな り, 0.6 から 0.8 付近を



越えると、φ_i(η) は一気に1 へと近づき、ほとんどのアミノ酸残基がN 状態へと移行する.

η が 0.6 から 0.8 付近 (遷移状態とみなせる) で起こるこの急激な $\phi_i(\eta)$ の変化は, N 状態への 折れたたみが協同的な現象であることを示している.





"Autumn Tree Night View"

フォールディングのキネティクスに おけるレートと緩和時間の関係は?

く要旨>

第7章

タンパク質のフォールディング過程を1次元の自由 エネルギープロファイルに沿ったキネティクス (kinetics,反応速度論)と見なしている.

まず,簡単な2状態系のキネティクスを取り上げ, 2状態系の構造変化を表す速度定数(レート)と緩和 時間との関係を求めている.

次に、多状態系をもつ系のキネティクスを記述する「マスター方程式」より、初期値に依存する二つのレート、つまり、タンパク質のフォールディング・レートと、アンフォールディング・レートを求める方法を定式化している(Muñoz & Eaton, 1999, 及び, Henry & Eaton, 2004).

そして、転移温度近傍で直線近似した、温度の関数である、シェブロン・プロット(chevron plot)を描いている.

更に、フォールディング/アンフォールディング・ レートの相違から、従来の、鎖に沿ってのアミノ酸 残基間の距離のみを考慮しているコンタクト・オー ダー (CO) ではなく、新たな、アミノ酸残基間の天 然接触エネルギーを考慮したコンタクト・オーダー を考慮すべきであると提案している. 第7章 フォールディングのキネティクスにおけるレートと緩和時間の関係は?

耳 タンパク質フォールディングのキネティクス

球状タンパク質は、フォールディングによって特異的立体構造を形成して、酵素として作用する などの重要な生命活動の役割を担っている. 我々は、この章で、タンパク質フォールディングのキネ ティクス(kinetics、反応速度論)を考察しよう. そして、フォールディングのキネティクスにおけ る "レート"(rate constant k,速度定数)と緩和時間(τ)との関係を明らかにしよう.

- <文献> H. Abe and H. Wako, Folding/unfolding kinetics of lattice proteins studied using a simple statistical mechanical model for protein folding, 1: Dependence on native structures and amino acid sequences, *Physica A*, Vol.383, pp. 3442-3454, 2009.
- <文献> 安部 晴男,山内 経則,輪湖 博, "タンパク質フォールディングのキネティクス, I. 定式化",西 日本工業大学紀要,第39卷, pp. 93-101, 2009.

まず,簡単な2状態系のキネティクスを取り上げ,2状態系の構造変化を表す"レート"(rate constant k,速度定数)と緩和時間(τ)との関係を求めよう.

次に、多状態系をもつ系のキネティクスを記述する「マスター方程式」を導き、タンパク質フォー ルディングの自由エネルギープロファイルから、タンパク質のフォールディング・レート($\ln k_f$) と、 アンフォールディング・レート($\ln k_u$) を求める方法を定式化しよう.そして、これらのフォール ディング・レートとアンフォールディング・レートを温度(の逆数:1/*T*))の関数として描こう("シ ェブロン・プロット"、chevron plot という).

□□ 簡単な"可逆的2状態転移反応"における、レートと緩和時間の関係

タンパク質のフォールディング反応は、様々の異なった機構で進行する.

最も簡単な場合は、図に示すような、状態1と状態2の可逆的2状態転移である.この可逆的2状態転移の構造変化を表す "レート方程式" (rate equation)を導き、これから、 "レート" ((rate constant k,速度定数) と緩和時間 (τ) との関係を求めよう.



今,状態1にある反応物の単位時間あたりの変化量は,反応物自身の量(濃度,確率を表す) p_1 に 比例し,また,状態2にある生成物の単位時間あたりの変化量は,生成物自身の量(濃度,確率を表 す) p_2 に比例するとみなす. $p_1 \ge p_2$ は,時刻 tの関数で,確率条件 $p_1(t) + p_2(t) = 1$ を満たし ている.ここで,前向きの反応のレート(速度定数)を $k_{1,2}(\equiv w_1)$,逆向きの反応のレート(速度定 数)を $k_{2,1}(\equiv u_1)$ とすると,次のレート方程式(rate equation)が成立する:

$$\frac{dp_1}{dt} = -w_1p_1 + u_1p_2$$
$$\frac{dp_2}{dt} = w_1p_1 - u_1p_2$$

上式は、次のように行列で表現できる:

$$\frac{d}{dt} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} -w_1 & u_1 \\ w_1 & -u_1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}$$

上式右辺の2×2の行列は転移行列と呼ばれる.上式の解が、 $p_1 = \alpha e^{\lambda t}, p_2 = \beta e^{\lambda t}$ (α, β, λ は定数) であると推測し、これらを上式に代入して、共通の因子 $e^{\lambda t}$ で割ってまとめると次式を得る:

 $\{(-w_1) - \lambda\}\alpha + u_1\beta = 0$ $w_1\alpha + \{(-u_1) - \lambda\}\beta = 0$

 α, β, λ がすべて 0 でないためには次式が成立せねばならない:

$$\begin{vmatrix} \{(-w_1) - \lambda\} & u_1 \\ w_1 & \{(-u_1) - \lambda\} \end{vmatrix} = 0 \quad \text{id}, \quad \begin{vmatrix} \lambda + w_1 & u_1 \\ w_1 & \lambda + u_1 \end{vmatrix} = 0$$

上式は λ に関する 2 次方程式(転移行列の固有方程式)で、これを解くと、2 個の固有値、 $\lambda_1 = 0, \lambda_2 = -(w_1 + u_1)$ が得られる.

① $\lambda_1 = 0$ とすると、 $w_1 \alpha = u_1 \beta$ となり、 $\alpha : \beta = u_1 : w_1$ である. 従って、解: $p_1 = u_1, p_2 = w_1$ ② $\lambda_2 = -(w_1 + u_1)$ とすると、 $\alpha = -\beta$ となり、 $\alpha : \beta = 1 : -1$ である. 従って、

解:
$$p_1 = e^{-(w_1+u_1)t}, p_2 = -e^{-(w_1+u_1)}$$

故に、一般解は、積分定数 (c_1, c_2) を用いて次のように書ける:

一般解:
$$p_1 = c_1 u_1 + c_2 e^{-(w_1 + u_1)t}, p_2 = c_1 \omega_1 + \{-c_2 e^{-(w_1 + u_1)t}\}$$

確率条件 $p_1(t) + p_2(t) = 1$ を用いて、初期条件 (t = 0のとき、 $p_1(0) = 1, p_2(0) = 0$ とする)より、 $1 = -c_1u_1 + c_2, 0 = c_1w_1 - c_2$ から積分定数 (c_1, c_2) は次のように求まる:

$$c_1 = \frac{1}{w_1 + u_1}, c_2 = \frac{w_1}{w_1 + u_1}$$

結局,初期条件(t=0のとき, $p_1(0)=1, p_2(0)=0$)を満たす,速度方程式(rate equation)の解は次式で表される:

$$p_1 = \frac{u_1}{w_1 + u_1} + \frac{w_1}{w_1 + u_1} e^{-(w_1 + u_1)t}, p_2 = \frac{w_1}{w_1 + u_1} - \frac{w_1}{w_1 + u_1} e^{-(w_1 + u_1)t}$$

上式より, $p_1(t), p_2(t)$ のグラフの略図

(例として, $w_1 = u_1 = 0.5$ の場合)を描くと図 のようになる.系は, $p_1(t), p_2(t)$ ともに平衡状 態に向かって指数関数的に緩和することがわか る.



第7章 フォールディングのキネティクスにおけるレートと緩和時間の関係は?

緩和時間 (τ , relaxation time) と反応のレート (速度定数) との関係を求めよう. 緩和時間の定義は, 時刻 t での反応物自身の量 (濃度, 確率を表す) が,時刻 $t + \tau$ で 1/e だけ減少するまでの時間 (τ) であるから,

$$\frac{1}{e}e^{-kt} = e^{-k(t+\tau)} \quad \text{for } k \equiv w_1 + u_1$$

故に,緩和時間(τ)と反応のレート(速度定数)との関係は次式で表わされる:

$$\tau = \frac{1}{k} \qquad (\text{trill}, \ k \equiv w_1 + u_1)$$

尚,平衡状態のとき, $dp_1/dt = 0, dp_2/dt = 0$ であるから,そのとき, $p_1 \rightarrow p_{eq}(1), P_2 \rightarrow p_{eq}(2)$ と表現 すると,次の"詳細釣合 (detailed balance)"の式が成立する:

$$w_1 p_{eq}(1) = u_1 p_{eq}(2)$$

多状態をもつ系のキネティクスを記述する"マスター方程式"を導こう.

一般に、状態 i = 1, 2, 3, ..., m とレート定数 $k_{i,j}$ に対して、"マスター方程式"は、次のように表現できる:

$$\frac{dp_i(t)}{dt} = \sum_{j \neq i} p_j(t)k_{j,i} - \sum_{j \neq i} p_i(t)k_{i,j}$$

この微分方程式系の解は、固有方程式 $|\lambda I - A| = 0$ (I:単位行列、A:微分方程式の右辺の転移 行列)の根として求まる固有値 $\lambda_{\nu}(\nu = 1, 2, 3, ..., m)$ を用いて、 $e^{\lambda_{i}t}$ の線形結合

$$p_i(t) = \sum_{\nu} a_{i,\nu} e^{\lambda_{\nu} t}$$
 $(i = 1, 2, 3, ..., m)$

の形に求められる.最大固有値は0で、その他の固有値は負である。

平衡状態への緩和時間 (τ , relaxation time) は、次式のように求まる:

$$\tau = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} + \dots + \frac{1}{\lambda_m}$$

結局,フォールディング・レート (k_f) は,次式で求まる.

$$k_f = \frac{1}{\tau} \qquad (\not \perp \, , \quad \tau = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} + \dots + \frac{1}{\lambda_m})$$

タンパク質のフォールディング過程を,次のような"反応座標"で考察しよう:『"天然状態にある アミノ酸残基の数"』 この反応座標において,"隣同士の状態間でのみ転移する"という制限を加え ることで,"マスター方程式"を具体的に書き下そう.

図に示すように、"可逆連続的な m 個の状態反応"を考える:



ここで、状態 *i* にある反応物の単位時間あたりの変化量は、反応物自身の量(濃度、確率を表す) p_i に比例し、また、状態 *i*+1 にある生成物の単位時間あたりの変化量は、生成物自身の量(濃度、確 率を表す) p_{i+1} に比例するとみなす. さらに、前向きの反応(状態*i*→状態*i*+1)のレート(速度定数) を $k_{i,i+1}(\equiv w_i)$ 、逆向きの反応(状態*i*+1→状態*i*)のレート(速度定数)を $k_{i+1,i}(\equiv u_i)$ とする "マ スター方程式"(反応速度方程式、rate equation)は、次式のように書ける:

フォールディング・レート(k_f)の計算式

タンパク質のフォールディング過程のキネティクスは、1999年、Muñoz & Eaton によって、1 次元 の自由エネルギープロファイルに沿った反応過程として提案され、2004年には、Henry & Eaton によ ってさらに拡張され、定式化された.彼らは、反応座標の隣接の離散値が可逆的に転移することを記 述する"マスター方程式"から導かれる"固有方程式の根"より、"緩和時間"(τ , relaxation time、そ の逆数がフォールディング・レート; k_f)を求める式を導いている(Appendix H:「フォールディン グ・レートの計算方法」参照).

- <文献> Muñoz, V. & Eaton, W.A. (1999). A simple model for calculating the kinetics of protein folding from three-dimensional structures. Proc. Natl Acad. Sci. USA, 96, 11311–11316.
- <文献> Henry, E.R. & Eaton, W.A. (2004). Combinatorial modeling of protein folding kinetics: free energy profiles and rates. Chem. Phys., 307, 163–185.
- <文献> Abe, H. & Wako, H. (2009). Folding/unfolding kinetics of lattice proteins studied using a simple statistical mechanical model for protein folding, I: Dependence on native structures and amino acid sequences. Physica A, 383, 3442–3454.

我々は、彼らの式を3次元格子タンパク質に適用した.フォールディング過程の緩和時間(τ , relaxation time, その逆数がフォールディング・レート; k_f)は、次式で計算できる:

$$\begin{split} \frac{1}{k_f} &\propto \tau \equiv \int_0^\infty \ \frac{<\eta>_{eq} - <\eta(t)>}{<\eta>_{eq} - <\eta(0)>} dt \\ &= \{<\eta>_{eq} - <\eta>_0\}^{-1} \sum_{j=0}^{n-4} \left\{ \frac{1}{p_{eq}(j)k_{j,j+1}} \sum_{i=j+1}^{n-3} \{p_{eq}(i) - p_0(i)\} \sum_{\eta=0}^j p_{eq}(\eta)(<\eta>_{eq} -\eta) \right\} \end{split}$$

この式は、3次元格子タンパク質において、初期条件(時刻0のとき)として、すべてランダム状態であったものがやがて平衡分布になる、その緩和時間 τを求める式である.反応座標として、天然

状態にあるアミノ酸残基数 η を採用している. 3 次元格子タンパク質のおける最小の局所構造は 4 個 のユニットであるから、 η の値の範囲は $0 \le \eta \le n-3$ である. また、 $<\eta_{eq} >$ は、平衡状態での η の 期待値 $<\eta>_{eq} = \sum_{\eta=0}^{n-3} \eta p_{eq}(\eta)$ を表す. ここで、 $p_{eq}(\eta)$ は、遷移状態で、天然状態にあるアミノ酸残基が η であるときの存在確率を表していて、次式のように、自由エネルギー関数 $F(\eta/\eta_N)$ と分配関数 Zより求めることができる:

$$p_{eq}(\eta) = \frac{e^{-F(\eta/\eta_N)/k_BT}}{Z}$$

(ただし, Z は分配関数, $Z = \sum_{\eta=0}^{n-3} e^{-F(\eta/\eta_N)/k_BT}$, k_B は Boltzmann 定数である)

また、初期条件(t=0のとき)は、すべてのアミノ酸残基はランダムコイル状態と仮定すると、次式 が成立する:

$$p_0(\eta) = \delta_{\eta,0}$$
 つまり、 $\eta = 0$ のとき $p_0(0) = 1$
 $\eta = 1, 2, 3..., n-3$ のとき $p_0(\eta) = 0$

$$\label{eq:loss_optimal_states} \begin{split} & \texttt{L} \supset \texttt{T}, \quad <\eta>_0 = \sum_{\eta=0}^{n-3} \eta p_0(\eta) = 0 \quad \texttt{ETSS}. \end{split}$$

前向きの反応(状態 $j \rightarrow$ 状態 j+1)のレート(速度定数) $k_{i,i+1} (\equiv w_i)$,及び、逆向きの反応(状態 $j+1 \rightarrow$ 状態 j)のレート(速度定数) $k_{i+1,i} (\equiv u_i)$ は、それぞれ次式で表される:

$$k_{j,j+1} = \gamma(\frac{p_{eq}(j+1)}{p_{eq}(j)})^{\kappa} \quad , \qquad k_{j+1,j} = \gamma(\frac{p_{eq}(j+1)}{p_{eq}(j)})^{\kappa-1}$$

ここで, κ と γ は, adjustable parameter である. Henry & Eaton は, 結果にあまり影響されないの で, $\kappa = 0.5$, $\gamma = 1$ を採用している.

Coffee Break ここまで、 Eaton らによって定式化された方法を、3次元格子タンパク質に適応 した.そして、初期条件として、時刻 t=0 で、すべてがランダムコイル状態である とし、その後、自由エネルギー・プロファイルで与えられる平衡分布までに達する 緩和時間 τ より, フォールディング・レート を求める式を示した. ここで、新たな疑問が生じる.つまり、平衡への緩和過程と捉えたとき、緩和時 間の初期値依存性はないのだろうか? アンフォールディング過程では,時刻 t=0で,すべてが天然状態であるとし,そ の後、自由エネルギー・プロファイルで与えられる平衡分布までに達する緩和時間 τ より, アンフォールディング・レート を求める式を導いてみよう. そして, フォールディング・レート を求める式と一致するかどうかを検討しよう.

アンフォールディング・レート (k_u) の計算式

前節ではフォールディング・レート k_f を「時刻 t=0 で、すべてがランダムコイル状態である」 とし、その後、自由エネルギー・プロファイルで与えられる平衡分布までに達する緩和時間 τ より求 めている.ここで、次のような疑問が生じる:『平衡への緩和過程と捉えれば、緩和時間の値に対して、 初期値依存性はないのだろうか?』

"アンフォールディング過程での初期条件"は、「時刻 t=0 で、すべてのアミノ酸残基は天然状態 である」と仮定する. その後、平衡に達する緩和時間 τ (逆数がアンフォールディング・レート k_u) を求める式が必要である. 時刻 0 ですべて天然構造にあったものが、やがて平衡分布になるときの、 "緩和時間 τ " (その逆数がアンフォールディング・レート; k_f)は、次式で表される:

$$\begin{split} \frac{1}{k_u} &\propto \tau \equiv \int_0^\infty \ \frac{<\eta>_{eq} - <\eta(t)>}{<\eta>_{eq} - <\eta(0)>} dt \\ &= \{<\eta>_{eq} - <\eta>_0\}^{-1} \sum_{j=0}^{n-4} \left\{ \frac{1}{p_{eq}(j+1)k_{j+1,j}} \sum_{i=0}^j \{p_{eq}(i) - p_0(i)\} \sum_{\eta=j+1}^{n-3} p_{eq}(\eta)(<\eta>_{eq} - \eta) \right\} \end{split}$$

アンフォールディング過程の初期条件(時刻 t=0)は、アミノ酸残基はすべて天然状態にあるので、 $p_0(\eta) = \delta_{n,n-3}$ となる ($p_0(\eta)$ は、初期条件における η の確率分布である).

つまり、
$$\eta = n - 3$$
 のとき
 $\eta = 0, 1, 2, 3..., n - 2$ のとき
 $p_0(n - 3) = 1$
 $p_0(\eta) = 0$
たって、 $<\eta >_0 = \sum_{n=3}^{n-3} \eta p_0(\eta) = n - 3$ となる.

ここで求められた,フォールディング・レート k_f とアンフォールディング・レート k_u は,絶対温度 T の関数である.これらの曲線を,絶対温度の逆数 1/T で描いてみよう(シェブロン・プロットという).

💶 シェブロン・プロット

図は、3 次元格子タンパク質である、タンパク 質 al の場合を例として、フォールディング・レー ト $\ln k_f$ と、 アンフォールディング・レート $\ln k_u$ を絶対温度(の逆数 1/T)の関数として 描いた.挿入図は、転移温度(T_m)付近を拡大 した図である(図の転移温度近傍の点線は、直線 近似を示している).

これら二つの曲線(赤と青の曲線)は,著しく 異なっている.



そこで、近似として、与えられた温度に対するレートは、両方のレートの小さい方の曲線を採用することにしよう.つまり、図の二つの曲線で、転移温度 T_m より小、つまり、 $1/T_m$ より大は、 $\ln k_f$

第7章 フォールディングのキネティクスにおけるレートと緩和時間の関係は?

(赤色の曲線)を採用し、転移温度 T_m より大、つまり、 $1/T_m$ より小は、 $\ln k_u$ (青色の曲線) を採用し、転移温度 T_m 近傍では、直線近似を行う. その交点が転移温度であり、次式が成立している:

 $\ln k_f(T_m) = \ln k_u(T_m)$

このように、 T_m 近傍で直線近似した新たな曲線を、シェブロン・プロット(chevron plot) として、採用することにしよう.

<コメント> シェブロンとは、紋章や階級章に使われる図形の一種で, 三角形あるいは逆三角形の底辺のない図形をいう.右図はパイロットのシェブロン.



101

💶 3 次元格子タンパク質の自由エネルギー曲線とシェブロン・プロット

タンパク質, al, a2, b1, b2 に対して, 具体的に"シェブロン・プロット"(Chevron plot)を描いて みよう.

図の (a), (b) は, タンパク質, *a1*, *a2*, *b1*, *b2* に対して, 転移温度 (T_m) における, 自由エネルギー 曲線 ($F(\eta/\eta_N)$)を描いている.

<文献> H. Wako & H. Abe, (2011). Study of Folding/unfolding kinetics of lattice proteins by applying a simple statistical mechanical model for protein folding, *in Protein Folding*, E. C. Walters ed., Nova Science. Publishers, 349–3376.



D;変性状態, ‡;遷移状態,

N;天然状態

を示す.

横軸は,規格化された反応座標 (η/η_N)である($0 \le \eta/\eta_N \le 1$). ただし,

- η : 天然状態にあるアミノ酸残基数,
- η_N ; η の最大値
- (a) タンパク質 a1, a2 の転移温度
 (T_m)における自由エネルギー曲線 a1;赤色,T_m=0.240, a2;青色,T_m=0.257.
- (b) タンパク質 b1, b2 の転移温度
 (T_m)における自由エネルギー曲線
 b1;赤色,T_m=0.206,
 b2;青色,T_m=0.193.



図の (c), (d) は、フォールディング/アンフォールディング・レート ($\ln k_f \ge \ln k_u$) から求まる "シェブロン・プロット" (Chevron plot) を描いている (横軸:温度の逆数 1/T). (c) タンパク質 a1, a2 の "シェブロン・プロット"

a1;赤色, a2;青色
(d) タンパク質 b1, b2 の "シェブロン・プロット"
b1;赤色, b2;青色

これらの結果から何がわかるのだろうか? 図から一見してわかるように、3次元格子タンパク質、a1, a2, b1, b2 に対する自由エネルギー曲線(図の(a), (b))も"シェブロン・プロット"(図の(c), (d))も、それぞれのタンパク質に対して異なった曲線であるように見える.これらの相違はどこからくるのだろうか?

ここで次のことを確認しよう: 『タンパク質 al, a2 は,同じ天然構造をもっているが,それらのア ミノ酸配列は異なっている. タンパク質 bl, b2 も同様に同じ天然構造をもっているが,それらのアミ ノ酸配列は異なっている. 』 第7章 フォールディングのキネティクスにおけるレートと緩和時間の関係は?

結局,同じ天然構造をもっていても,それらのアミノ酸配列が異なっていれば,それらのタンパク 質のフォールディング過程のキネティクスも異なっているのだろうか? このことを考察してみよう.

図の (a), (b) の, タンパク質, *a1, a2, b1, b2* に対する自由エネルギー曲線 (*F*(η/η_N) から, 次のような表を得る:

自由エネルギーの差	3次元格子タンパク質								
	a 1	a 2	<i>b</i> 1	<i>b</i> 2					
$F(\eta_{+}/\eta_{N}) - F(\eta_{D}/\eta_{N})$	1.3	1.7	1.1	2.3					
$F(\eta_{+}/\eta_N) - F(\eta_M/\eta_N)$	1.4	1.8	1.2	2.6					

ここで,

・ $F(\eta_{*}/\eta_{N})$;遷移状態領域での極大値,

・ $F(\eta_D/\eta_N)$;変性状態領域での極小値,

- ・F(η_M/η_N); 天然状態領域での極小値,
- ・ $F(\eta_{+}/\eta_{N}) F(\eta_{D}/\eta_{N})$; 遷移状態と変性状態との自由エネルギーの差,つまり, タンパク質分子が,変性状態から天然状態へフォールドするための活性化 エネルギー(activation energy,障壁を乗り越えるためエネルギー),
- ・ $F(\eta_{\dagger}/\eta_{N}) F(\eta_{M}/\eta_{N})$; 遷移状態と天然状態との自由エネルギーの差,つまり, タンパク質分子が,天然状態から変性状態へ向かうために乗り越えねばならないエ ネルギー.

上の表で、タンパク質 *a1*、*a2*を比較すると、自由エネルギーの差、 $F(\eta_{+}/\eta_{N}) - F(\eta_{D}/\eta_{N})$ 、 $F(\eta_{+}/\eta_{N}) - F(\eta_{M}/\eta_{N})$ の値は、どちらとも、タンパク質 *a2*の方が、タンパク質 *a1*より大きな値で ある. タンパク質 *b1*、*b2*の場合も比較すると、自由エネルギーの差、 $F(\eta_{+}/\eta_{N}) - F(\eta_{D}/\eta_{N})$ 、 $F(\eta_{+}/\eta_{N}) - F(\eta_{M}/\eta_{N})$ の値は、どちらとも、タンパク質 *b2*の方が、タンパク質 *b1*より、かなり大 きな値である.

これらの結果から、シェブロン・プロット"(図の (c)、(d)は、温度(の逆数)の関数としてのフ オールディング/アンフォールディング・レート、つまり反応の速度定数とみなせる)の相違が次の ように説明できる:

図の (c) から, タンパク質 *a2* のフォールディング・レート (ln *k*_f, 図の 1/*T*_mより高い領域) も, アンフォールディング・レート (ln *k*_u, 図の 1/*T*_mより低い領域) でも, どちらも, タンパク質 *a1* の 場合よりも小である理由は, 自由エネルギーの差, $F(\eta_{+}/\eta_{N}) - F(\eta_{D}/\eta_{N})$, $F(\eta_{+}/\eta_{N}) - F(\eta_{M}/\eta_{N})$ の値が, タンパク質 *a2* の方が大きいからである.

結局,タンパク質 *a*2 の方が障壁を乗り越えるためエネルギーの値が大きいため,反応速度が遅くなる(平衡に達する緩和時間が大となる). つまり,フォールディング・レート($\ln k_f$)も,アンフォールディング・レート($\ln k_u$)も,より小さな値をもつ.

簡単に言えば,『タンパク質 a2 の方が, タンパク質 a1 より, フォールディングもアンフォールディ ングも手間がかかる過程』と言えるだろう.

図の (d) から, タンパク質 *b2* のフォールディング・レート ($\ln k_f$, 図の 1/ T_m より高い領域) も, アンフォールディング・レート ($\ln k_u$, 図の 1/ T_m より低い領域) でも, どちらも, タンパク質 *b1* の
場合よりも,かなり小である理由は,タンパク質 *a1*, *a2* と同様の理由からである. タンパク質 *b1*, *b2* の自由エネルギーの差がかなりあるので,レートがより小となっている.

簡単に言えば, 『タンパク質 b2 の方が, タンパク質 b1 より, フォールディングもアンフォールディングも, より手間がかかる過程』と言えるだろう.

転移温度の値 (*T*_m) はどうだろうか?

タンパク質 a2 の転移温度 の値($T_m = 0.257$)は、タンパク質 a1 の転移温度の値($T_m = 0.240$) より、少し高い. 結局、『タンパク質 a2 の方が、タンパク質 a1 より、少しばかり安定である』とい えるだろう.

一方, タンパク質 b1 の転移温度 の値($T_m = 0.206$)は, タンパク質 b2 の転移温度の値($T_m = 0.193$)より, 少し高い. 結局, 『タンパク質 b1 の方が, タンパク質 b2 より, 少しばかり安定である』といえるだろう.

更に、"シェブロン・プロット"の直線からのズレ("chevron rollover"という)を指摘しておこう. この"chevron rollover"は、フォールディング・キネティキスにおける実験で、溶媒条件を変化した ときに多く観測されている.

図の (c), (d) から, "シェブロン・プロット"は, 極端な低温領域(強いフォールディング条件の 領域)と極端な高温領域(強いアンフォールディング条件の領域)で, 直線からズレていることがわ かる. 我々の3次元格子タンパク質でも観測される理由は, 未だ明確ではないが, 自由エネルギー・ プロファイルにおいて, 変性状態領域で"平ら"であることが影響していると思われる.

ところで、次のような新たな疑問が生じる.

従来, "タンパク質のフォールディング/アンフォールディング・レートは, 鎖に沿ってのアミノ酸 残基間の距離のみを考慮したコンタクト・オーダー (CO) と相関がある"という議論されているが, いままで確認したように, 3 次元格子タンパク質 *a1*, *a2* は(あるいは, *b1*, *b2* は), 同じ天然構造を もっていてコンタクト・オーダー (CO) が同じ値であるにもかかわらず, フォールディング/アンフ ォールディング・レートは相違している. このことは, 何を意味しているのだろうか?

つまり、フォールディング/アンフォールディング・レートと相関する量として、鎖に沿ってのアミノ酸残基間の距離のみを考慮したコンタクト・オーダー(CO)ではなく、他の、相関する量を考慮すべきではないだろうか?

我々は、次のような提案をしたいと思う:

「フォールディング/アンフォールディング・レートは,『鎖に沿ってのアミノ酸残基間の距離のみを 考慮したコンタクト・オーダー (CO)』ではなく,『アミノ酸残基間の天然接触エネルギーを考慮した コンタクト・オーダー 』との相関を考えるべき」ではないだろうか? このことを考察しよう.

↓ アミノ酸残基間の天然接触エネルギーを考慮したコンタクト・オーダーの 提案

いま, $\rho_i \varepsilon$,アミノ酸残基間の距離(残基数)がiのときの天 然接触数を表すと,天然接触数の累積度数 C_k は次式となる;

$$c_k = \sum_{i=1}^k \rho_i$$

ここで,具体的な例として,タンパク質 al に対する天然接触数の数値を右の表に示す.

当然, タンパク質 a2 も天然構造が同じなので, 全く同じ値をもっている.

下図は、例として、タンパク質 al の場合の「天然接触数の累 積度数分布」曲線を描いた.



さて、天然接触数の累積度数 Ck から次の式が導かれる:

$$\sum_{k=1}^{n-1} c_k = \sum_{m=1}^{n-1} (\sum_{i=1}^m \rho_i)$$

= $\rho_1 + (\rho_1 + \rho_2) + (\rho_1 + \rho_2 + \rho_3) + \dots + (\rho_1 + \rho_2 + \dots + \rho_{n-1})$
= $\sum_{k=1}^{n-1} (n-k)\rho_k$
= $nc_{n-1} - \sum_{k=1}^{n-1} k\rho_k$

ここで、 c_{n-1} は、全天然接触数である、上式の左辺は、天然接触数の累積度数分布を表しているが、 これは、上図では、右下の面積を表わしている.

また、上式の右辺の第1項は、全面積を表している.したがって、上式の右辺の第2項は、図では左上の面積を表していることになる.

この上式の右辺の第2項が、タンパク質のコンタクト・オーダーと関係があることを次に示す. 上式の右辺の第2項は次式のように書き換えることができる:

$$\sum_{k} k \rho_{k} = \sum_{k} k \left(\sum_{i} \Gamma_{i,i+k} \right) = \sum_{i < j} \sum_{j} \left| j - i \right| \Gamma_{i,j}$$

ここで, i, j は, アミノ酸残基を表し, また, $\Gamma_{i, i}$ は, 次の量である;

○ アミノ酸残基 *i*, *j* が天然構造において, 互いに接触しているとき,

 $\Gamma_{i,i} = 1$

○ アミノ酸残基 *i*, *j* が天然構造において, 互いに接触していないとき,

 $\Gamma_{i,i} = 0$

タンパク質のコンタクト・オーダー (CO)の定義は次式の通りである:

$$CO = \frac{1}{A} \sum_{i < j} \sum \left| j - i \right| \Gamma_{i,j}$$

ここで, Aは, 次のように定義する:

絶対コンタクト・オーダー (ACO) のとき; A = n

相対コンタクト・オーダー (RCO) のとき; $A = n \times c_{n-1}$

故に,
$$\sum_{k} k \rho_{k}$$
は, コンタクト・オーダー*CO* と関係があることになる.

"アミノ酸残基間の天然接触エネルギーを考慮したコンタクト・オーダー"を考慮すると、同じ天然 構造をもつタンパク質でも、そのアミノ酸配列が異なると、異なるフォールディング/アンフォールデ ィング・レートをもつことを次に示そう.

- <文献> H. Abe and H. Wako, Folding/unfolding kinetics of lattice proteins studied using a simple statistical mechanical model for protein folding, 1: Dependence on native structures and amino acid sequences, Physica A, Vol.383, pp. 3442-3454, 2009.
 - <文献> 安部 晴男,山内 経則,輪湖 博,"タンパク質フォールディングのキネティクス, II.フォール ディング速度とコンタクト・オーダーとの関係",西日本工業大学紀要,第40卷, pp. 51-57, 2010.
- アミノ酸残基間の距離(残基数)が k である時の天然接触エネルギーの和は、次式で表される:

$$\varepsilon_k = \sum_{i} U(\xi_i, \xi_{i+k}) \Gamma_{i,j+k}$$

また、その接触エネルギーの累積度数分布は、次式のようになる:

$$e_k = \sum_{i=1}^k \varepsilon_i$$

図は、タンパク質 a1, a2, b1, b2 に対する、次の ϵ_k と e_k , 及び、 c_k を描いている: ϵ_k ; アミノ酸残基間の距離(残基数)が k である時の天然接触エネルギーの和 e_k ; 天然接触エネルギーの累積度数分布 第7章 フォールディングのキネティクスにおけるレートと緩和時間の関係は?

C_m; 天然接触数の累積度数分布.

ただし、それぞれ、次の量で規格化して描いている:

- *ε*total; 全接触エネルギーの値
- en-1 ; 接触エネルギーの最小値
- cn-1 ; 天然接触数の最大値

図の(a)は、タンパク質 a1、a2 に対する次の量を描いている: ϵ_k ; $\blacktriangle \rightarrow$ タンパク質 a1、 $* \rightarrow$ タンパク質 a2 e_k ; - (赤色の曲線) \rightarrow タンパク質 a1, - (青色の曲線) \rightarrow タンパク質 a2 C_k ; - (緑色の点線) \rightarrow タンパク質 a1, a2 共通 図の(b) は、タンパク質 b1, b2 に対する次の量を描いている: ϵ_k ; $\blacktriangle \rightarrow$ タンパク質 b1, $* \rightarrow$ タンパク質 b2 e_k ; - (赤色の曲線) \rightarrow タンパク質 b1, - (青色の曲線) \rightarrow タンパク質 b2 C_k ; - (緑色の点線) \rightarrow タンパク質 b1, b2 共通

図より,次のことがわかる:

①図の(a)より, タンパク質 *a1*, *a2* に対して, 3 曲線(2つの 赤・青色曲線 e_k と緑色点線 c_k)は, ほぼ一致しているが, 詳細に見ると, *k* が小さいとき, つまり短距離, 及び中距離相互作用のときは, タンパク質 *a2* の e_k 曲線が, タンパク質 *a1* の e_k 曲線よりやや大である. このことは, フォールデ ィングの初期段階で, タンパク質 *a2* の方が, タンパク質 *a1* 小より, 小さな局所構造が, より形成 されやすいことを示唆している.



②図の(b)より, タンパク質 b1 では, $e_k \ge c_k$ 曲線 は, k が小のとき, ほぼ一致している. k が大のとき, つまり, 長距離相互作用が働くときに, タンパク質 b2 の e_k 曲線のみが大きく異なっていることがわかる. k = 21 のときの, 3 個の天然接触ペアである, 6 と 27,7 と 28,12 と-33 の長距離相互作用の天然接触エネルギーの和(絶対値)は, タンパク質 b1 の方が, タンパク質 b2 より大である.

逆に, k = 35 のときの, 3 個の天然接触ペアである, 5 と 40, 13 と 48 の長距離相互作用の天然接触エネルギーの和(絶対値)は、タンパク質 b2 の方が、タンパク質 b1 より大である.

結局,タンパク質 b2 は、5 と 40、13 と 48 の長距離相互作用がフォールディングの最終段階で作用して安定な天然構造へと折れたたまれることを示唆している.

このことは, 『タンパク質 b2 のフォールディング速度がタンパク質 b1 より小である』ことを示す 理由と考えられる.

接触エネルギーの累積度数分布式 $e_k = \sum_{i=1}^k \varepsilon_i$ より、次式が導かれる:

$$\sum_{k=1}^{n-1} e_k = \sum_{m=1}^{n-1} \left(\sum_{i=1}^m \varepsilon_i \right)$$
$$= \varepsilon_1 + (\varepsilon_1 + \varepsilon_2) + (\varepsilon_1 + \varepsilon_2 + \varepsilon_3) + \dots + (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_{n-1})$$
$$= \sum_{k=1}^{n-1} (n-k)\varepsilon_k$$
$$= ne_{n-1} - \sum_{k=1}^{n-1} k\varepsilon_k$$

上式の左辺は、天然接触エネルギーの累積度数分布であり、図で右下の面積に対応している(図の縦軸の値は、規格化されている).

また、上式の右辺の第1項は、図の全面積に対応していて、上式の右辺の第2項は、図の左上の面積に対応している.この右辺の第2項は、天然接触エネルギーを考慮したコンタクト・オーダー(the energy- weighted contact order)とみなせる.

図の(b)より, 左上の面積は, 明らかに, タンパク質 b2 の場合のほうが, タンパク質 b1 より大である. 結局, 『フォールディング・レートと, 天然接触エネルギーを考慮したコンタクト・オーダー (the energy- weighted contact order, EWCO) とが対応している』ことが示唆される.

従来,「フォールディング・レートは, 鎖に沿ってのアミノ酸残基間の距離(残基数)のみを考慮したコンタクト・オーダー(ACO, 及び RCO)との相関に関して議論されている」が, コンタクト・オーダーに関して, 我々は, 次のような提案をしよう:

『コンタクト・オーダー は, "鎖に沿ってのアミノ酸残基間の距離のみを考慮した量"ではなく, む しろ, "フォールディング・レート", 及び, "天然接触エネルギー"との相関を議論すべきである.』



多状態をもつ系の速度論的記述には、「マスター方程式」がある.ここで、n個の状態 i=1,2,3,....,n からなる系の速度論を考察しよう.

いま,前向きの反応 (状態*i*→状態*j*)のレート (速度定数)を $k_{i,j}$,及び,逆向きの反応 (状態*j*→状態*i*) のレート (速度定数)を $k_{j,i}$,状態*i* での状態のポピュレーションを p_i (確率を表す),状態*j* での状態のポピュレーションを p_j (確率を表す)とすると,状態*i*から状態*j*への確率の流れの量は $p_i(t)k_{i,j}$, 状態*j*から状態*i*への確率の流れの量は $p_j(t)k_{j,i}$ であるから,この系の「マスター方程式」は次のよう に表現できる:

$$\frac{dp_i(t)}{dt} = \sum_{i \neq i} p_j(t)k_{j,i} - \sum_{i \neq i} p_i(t)k_{i,j}$$

この式は、状態のポピュレーションが、入ってくる確率と出ていく確率の差によって変化すること を意味している.

「マスター方程式」を"生体分子系の離散化された記述に用いる"ことを考えよう.

タンパク質フォールディングのキネティクスおいて、「天然状態にあるアミノ酸残基の数」を"反応 座標"とし、次のような制限を加える:『反応座標の隣同士の状態間でのみ離散的に転移する.』

図のような可逆離散的なn 個の状態反応に対する「マスター方程式」を具体的に書き下そう 状態iにある反応物の単位時間あたりの変化量は、反応物自身の量(濃度、確率を表す) p_i に比例し、また、 状態i+1にある生成物の単位時間あたりの変化量は、生成物自身の量(濃度、確率を表す) p_{i+1} に比例すると みなす. さらに、前向きの反応(状態 $i \rightarrow$ 状態i+1)のレート(速度定数)を $k_{i,i+1}$ ($\equiv w_i$)、逆向きの反応(状 態 $i+1 \rightarrow$ 状態i)のレート(速度定数)を $k_{i+1,i}$ ($\equiv u_i$) とする.



マスター方程式は、次式のようなレート方程式 (rate equation) で記述される:

ここで、この微分方程式の右辺の転移行列 (transition matrix) をA で表し、I を単位行列とすると、 微分方程式系の解は、固有方程式 $|\lambda I - A| = 0$ の根として求まる固有値 $\lambda_{\nu}(\nu = 1, 2, 3, ..., n)$ を用いて、 次式のように、 $e^{\lambda_{\nu} t}$ の線形結合の形に求められる:

$$p_i(t) = \sum_{\nu} a_{i,\nu} e^{\lambda_{\nu} t}$$
 (*i* = 1,2,3,....,*n*)

結局,固有値 $\lambda_{\nu}(\nu = 1,2,3,...,n)$ を計算することができれば,平衡状態への緩和時間 (τ , relaxation time)は,次式のように求まる:

$$\tau = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} + \dots \dots \frac{1}{\lambda_n}$$

転移のレート(k)は、次式で計算できる:

$$k = \frac{1}{\tau} \qquad (\{ \underline{\square} \, \, \cup, \, \tau = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} + \dots + \frac{1}{\lambda_n} \}$$

Cheng らは、固有方程式 $|\lambda I - A| = 0$ の根を求める計算方法を定式化し、Eaton 及び Henry & Eaton らは、その計算式を、タンパク質フォールディングのキネティクス(kinetics、反応速度論)に適用 した. 彼らは、タンパク質フォールディングを自由エネルギープロファイルに沿った反応過程とみな し、フォールディング・レートを求める式を提案している. Cheng ら、及び、Henry らによる導入さ れた、緩和時間(τ , relaxation time、その逆数がフォールディング・レート; k_f)を求める計算式を 彼らに準じて求めよう.

- <文献> X. Z. Cheng, M. B. A. Jalil, H.K. Lee & Y. Okabe. A new time quantifiable Monte Carlo method in simulating magnetization reversal process, Phys. Rev. B,72, 094420, 2005.
- <文献> Muñoz V. & Eaton WA. A simple model for calculating the kinetics of protein folding from three-dimensional structures, Proc. Natl. Acad. Sci., USA, 96: 11311–11316, 1999.
- <文献> Henry, E.R. & Eaton, W.A. Combinatorial modeling of protein folding kinetics: free energy profiles and rates, Chem. Phys., 307, 163–185, 2004.

📮 固有方程式の行列式は, n 次の多項式で表される

固有方程式 $|\lambda I - A| = 0$ の根である,固有値 $\lambda_{\nu}(\nu = 1, 2, 3, ..., n)$ を変数 x で表すと,固有方程式 は |xI - A| = 0 と書ける. この左辺を $J_n(x)$ とする:

 $J_{n}(x) = \begin{vmatrix} x + w_{1} & u_{1} \\ w_{1} & x + w_{2} + u_{1} & u_{2} \\ w_{2} & \dots & \dots \\ & & & \dots & \dots \\ & & & & \dots & u_{n-1} \\ & & & & & w_{n-1} & x + u_{n-1} \end{vmatrix}$

ところで、平衡状態では、 $dp_i(t)/dt = 0$ (i = 1, 2, 3, ..., n) であるから、「マスター方程式」の右 辺は 0 となる. それ故, $p_i(t) = 0$ (i = 1, 2, 3, ..., n) 以外の解をもつためには, |A| = 0 よって, x=0 は、固有方程式は |xI - A| = 0の解であることがわかる.

ここで、次のような新たな行列式 $K_n(x)$ を定義する:

この行列式 $K_n(x)$ は、行列式 $J_m(x)$ の構成要素 (n,n) の成分に w_n を加えたものである. 行列式 の性質から,次の式が成立する:

> $K_n(x) = J_n + w_n K_{n-1}$ (1)

一方,行列式の定義から, $J_m(x)$ と $K_n(x)$ の間に,次のような関係式が成立する:

$$J_{n}(x) = (x + u_{n-1})K_{n-1} - w_{n-1}u_{n-1}K_{n-2}$$
(2)

(2) 式に, (1) 式を代入すると,

$$J_{n}(x) = (x + u_{n-1})K_{n-1} - u_{n-1}(K_{n-1} - J_{n-1})$$

故に,次式が成立する:

$$J_{n}(x) = xK_{n-1} + u_{n-1}J_{n-1}$$
(3)

式 (3) に,式 (1) を代入すると,

$$J_{n}(x) = x(J_{n-1} + w_{n-1}K_{n-2}) + u_{n-1}J_{n-1}$$
$$= (x + u_{n-1})J_{n-1} + w_{n-1}(xK_{n-2})$$
$$= (x + u_{n-1})J_{n-1} + w_{n-1}(J_{n-1} - u_{n-2}J_{n-2})$$

故に, 次のような "差分方程式" (difference equation) が成立する:

$$J_{n}(x) = (x + u_{n-1} + w_{n-1})J_{n-1} - w_{n-1}u_{n-2}J_{n-2} \quad (\text{ttil}, J_{0} = 0, J_{1} = x \text{til}) \quad (4)$$

特に, n=2のとき, $J_2(x) = x(x+w_1+u_1)$ より, $J_2(x) = 0$ を満たす固有値 x は, $x=0,-(w_1+u_1)$ である (この第7章の「簡単な"可逆的2状態転移反応"における, レートと 緩和時間の関係」の項参照のこと).

また, n=3のときは,

$$J_3(x) = x\{x^2 + (w_1 + w_2 + u_1 + u_2)x + (w_1w_2 + w_1u_1 + w_1u_2 + u_1u_2)\}$$

である.

次に, 次のような, 多項式 $f_{\mu}(x)$ を定義する:

$$f_n(x) = J_n(x) / x \tag{5}$$

結局, 平衡状態への緩和時間 τ (relaxation time) は, $f_n(x) = 0$ の解である 固有値 $x_1, x_2, ..., x_{n-1}$ の逆数の和として求めることができる:

$$\tau = \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_{n-1}}$$

転移のレート(k)は、次式で計算できる:

$$k = \frac{1}{\tau}$$
 ((11), $\tau = \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_{n-1}}$)

固有値 $x_1, x_2, ..., x_{n-1}$ の逆数の和は、 Vieta's theorem (根と係数の関係)より、 近似的に、次式で求めることができる:

$$\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_{n-1}} = -\frac{f_n(0)}{f_n(0)}$$

結局, $f_n(0), f'_n(0)$ を求めることができれば. 平衡状態までの緩和時間 τ (relaxation time) が, 次式で計算できる:

$$\tau = \left| -\frac{f_n'(0)}{f_n(0)} \right| \tag{6}$$

 $f_n(0), f'_n(0)$ を求める計算式を導こう:

◎ 式(6)の分母である *f_n*(0)の計算式を導こう

式(4), (5)より, 次式が成立する:

$$f_n(x) = (x + u_{n-1} + w_{n-1})f_{n-1}(x) - w_{n-1}u_{n-2}f_{n-2}(x)$$
(7)

ただし, $f_0(x) = 0, f_1(x) = 1$

ここで、上式に x=0 を代入すると、

第7章 フォールディングのキネティクスにおけるレートと緩和時間の関係は?

$$f_n(0) = (u_{n-1} + w_{n-1})f_{n-1}(0) - w_{n-1}u_{n-2}f_{n-2}(0)$$

となる. この式より.

$$f_{n}(0) - u_{n-1}f_{n-1}(0) = w_{n-1}\{f_{n-1}(0) - u_{n-2}f_{n-2}(0)\}$$
$$= w_{n-1}w_{n-2}w_{n-3}...w_{2}w_{1}\{f_{1}(0) - f_{0}(0)\}$$
$$= w_{n-1}w_{n-2}w_{n-3}...w_{2}w_{1}$$
$$= \prod_{i=1}^{n-1} w_{i}$$
(8)

ここで、次のような2つの量 g_n, p_i を定義しよう:

$$g_{n} = \frac{f_{n}(0)}{\prod_{i=1}^{n-1} u_{i}}$$
(9)
$$p_{1} = 1$$
$$p_{i} = \prod_{j=1}^{i-1} \frac{w_{j}}{u_{j}}$$
(1 < i ≤ n) (10)

 g_n と g_{n-1} の差は,式 (8),(10) を用いると,次のようになる:

$$g_{n} - g_{n-1} = \frac{f_{n}(0)}{\prod_{i=1}^{n-1} u_{i}} - \frac{f_{n-1}(0)}{\prod_{i=1}^{n-2} u_{i}}$$
$$= \frac{1}{u_{1}u_{2}...u_{n-2}} \{ \frac{f_{n}(0)}{u_{n-1}} - f_{n-1}(0) \}$$
$$= \frac{1}{u_{1}u_{2}...u_{n-2}} u_{n-1} \{ f_{n}(0) - u_{n-1}f_{n-1}(0) \}$$
$$= \frac{\prod_{i=1}^{n-1} w_{i}}{\prod_{i=1}^{n-1} u_{i}} = \prod_{i=1}^{n-1} \frac{w_{i}}{u_{i}} = p_{n}$$

故に、 g_n は、次式で表される:

$$g_{n} = \sum_{i=1}^{n} p_{i}$$
(11)

(9)式より,

$$f_{n}(0) = \left(\prod_{i=1}^{n-1} u_{i}\right) g_{n} = \left(\prod_{i=1}^{n-1} u_{i}\right) \left(\sum_{j=1}^{n} p_{j}\right)$$
(12)

◎ 式(6)の分子である f_n['](0)の計算式を導こう

f'(0)の計算式を導くために式(7)からスタートしよう:

 $f_n(x) = (x + u_{n-1} + w_{n-1})f_{n-1}(x) - w_{n-1}u_{n-2}f_{n-2}(x) \qquad (ただし, f_0(x) = 0, f_1(x) = 1)$ この式を微分すると、

$$f'_{n}(x) = f_{n-1}(x) + (x + u_{n-1} + w_{n-1})f'_{n-1}(x) - w_{n-1}u_{n-2}f'_{n-2}(x)$$

上式に, $x = 0$ を代入すると次のような式が得られる:

$$f'_{n}(0) = f_{n-1}(0) + (u_{n-1} + w_{n-1})f'_{n-1}(0) - w_{n-1}u_{n-2}f'_{n-2}(0)$$
(13)

次に,式(9)を微分すると,

$$g'_{n} = \frac{f'_{n}(0)}{\prod_{i=1}^{n-1} u_{i}}$$
(14)

上式より g_n と g_{n-1} の差は次のようになる:

$$g'_{n} - g'_{n-1} = \frac{f'_{n}(0)}{\prod_{i=1}^{n-1} u_{i}} - \frac{f'_{n-1}(0)}{\prod_{i=1}^{n-2} u_{i}}$$
$$= \frac{1}{\prod_{i=1}^{n-2} u_{i}} \left[\frac{f'_{n}(0)}{u_{n-1}} - f'_{n-1}(0) \right]$$
$$= \frac{1}{\prod_{i=1}^{n-1} u_{i}} \left[f'_{n}(0) - u_{n-1}f'_{n-1}(0) \right]$$
$$= \frac{1}{\prod_{i=1}^{n-1} u_{i}} \left[f'_{n}(0) - u_{n-1}f'_{n-1}(0) \right]$$

ここで、式(13)を変形して、 $f_{n}'(0) - u_{n-1}f_{n-1}'(0) = f_{n-1}(0) + w_{n-1}\{f_{n-1}'(0) - u_{n-2}f_{n-2}'(0)\}$ この式を代入すると.

$$=\frac{1}{\prod_{i=1}^{n-1}} \left[f_{n-1}(0) + w_{n-1} \{ f_{n-1}(0) - u_{n-2} f_{n-2}(0) \} \right]$$

第7章 フォールディングのキネティクスにおけるレートと緩和時間の関係は?

$$\begin{split} &= \frac{1}{\prod_{i=1}^{n-1} u_i} \left[g_{n-1} \left(\prod_{i=1}^{n-2} u_i \right) + w_{n-1} \{ f_{n-1}(0) - u_{n-2} f_{n-2}(0) \} \right] \\ &= \frac{g_{n-1}}{u_{n-1}} + \frac{w_{n-1}}{\prod_{i=1}^{n-1} u_i} \{ f_{n-1}(0) - u_{n-2} f_{n-2}(0) \} \\ &= \frac{g_{n-1}}{u_{n-1}} + \frac{w_{n-1}}{\prod_{i=1}^{n-1} u_i} \left(\prod_{i=1}^{n-2} u_i \right) \left(g_{n-1} - g_{n-2} \right) \\ &= \frac{g_{n-1}}{u_{n-1}} + \frac{w_{n-1}}{u_{n-1}} \left(g_{n-1} - g_{n-2} \right) \\ &= \frac{g_{n-1}}{u_{n-1}} + \frac{w_{n-1}}{u_{n-1}} \left(g_{n-1} - g_{n-2} \right) \\ &= \sum_{i=1}^{n-1} \frac{p_{n-1}}{u_{n-1}} = \frac{w_{n-1}}{u_{n-1}} \quad \text{inder equation in the set of t$$

$$g_{n}^{'} - g_{n-1}^{'} = \frac{g_{n-1}}{u_{n-1}} + \frac{p_{n}}{p_{n-1}} \left(g_{n-1}^{'} - g_{n-2}^{'}\right)$$
$$= \frac{g_{n-1}}{u_{n-1}} + \frac{p_{n}}{p_{n-1}} \left(\frac{g_{n-2}}{u_{n-2}} + \frac{p_{n-1}}{p_{n-2}} \{g_{n-2}^{'} - g_{n-3}^{'}\}\right)$$
$$= \frac{p_{n}}{p_{n}} \frac{g_{n-1}}{u_{n-1}} + \frac{p_{n}}{p_{n-1}} \frac{g_{n-2}}{u_{n-2}} + \dots + \frac{p_{n}}{p_{2}} \frac{g_{1}}{u_{1}}$$

更に, 次の量 c_n を定義する:

$$g'_{n} - g'_{n-1} = c_{n}$$

これより,次式が導かれる:

$$g'_{n} = c_{n} + g'_{n-1} = c_{n} + (c_{n-1} + g_{n-2}) = \dots = c_{n} + c_{n-1} + \dots + (c_{1} + g'_{0}) = \sum_{i=1}^{n} c_{i}$$

(だだし、 $g_0' = 0$) ここで、 $g_n' - g_{n-1}' = c_n$ を用いると、 $c_n, c_{n-1}, c_{n-2}, \dots c_2, c_1$ は具体的に次のように表される: $c_n = \frac{p_n}{p_n} \frac{g_{n-1}}{u_{n-1}} + \frac{p_n}{p_{n-1}} \frac{g_{n-2}}{u_{n-2}} + \frac{p_n}{p_{n-2}} \frac{g_{n-3}}{u_{n-3}} \dots + \frac{p_n}{p_2} \frac{g_1}{u_1}$

$$c_{n-1} = \frac{p_{n-1}}{p_{n-1}} \frac{g_{n-2}}{u_{n-2}} + \frac{p_{n-1}}{p_{n-2}} \frac{g_{n-3}}{u_{n-3}} \dots + \frac{p_{n-1}}{p_2} \frac{g_1}{u_1}$$

$$c_{n-2} = \frac{p_{n-2}}{p_{n-2}} \frac{g_{n-3}}{u_{n-3}} \dots + \frac{p_{n-2}}{p_2} \frac{g_1}{u_1}$$
.....

$$c_3 = \frac{p_3}{p_3} \frac{g_2}{u_2} + \frac{p_3}{p_2} \frac{g_1}{u_1}$$

$$c_2 = \frac{p_2}{p_2} \frac{g_1}{u_1}$$

$$(\because c_1 = g_1 = \frac{f_1(0)}{u_0} = 0)$$

0

それ故,

 $c_1 =$

$$g_{n}' = \sum_{i=1}^{n} c_{i} = \sum_{i=1}^{n-1} \left[\frac{1}{p_{i+1}u_{i}} \left(g_{i}\right) \left(\sum_{k=i+1}^{n} p_{k}\right) \right]$$
(15)

結局, $f_n(0)$ は, 式(11), 式(14), 式(15)より次式となる:

$$f_{n}^{'}(0) = \left(\prod_{i=1}^{n-1} u_{i}\right) g_{n}^{'}$$
$$= \left(\prod_{i=1}^{n-1} u_{i}\right) \sum_{i=1}^{n-1} \left[\frac{1}{p_{i+1}u_{i}} \left(\sum_{j=1}^{i} p_{j}\right) \left(\sum_{k=i+1}^{n} p_{k}\right)\right]$$
(16)

◎ 式(12)の f_n(0)と、式(16)の f'_n(0)より、緩和時間 τ を求める式を導こう
 結局、式(12)の f_n(0)と、式(16) の f'_n(0)より、平衡状態までの緩和時間 τ は次式で求まる:

$$\tau = \left| -\frac{f_{n}(0)}{f_{n}(0)} \right|$$

$$= \left(\prod_{i=1}^{n-1} u_{i} \right) \sum_{i=1}^{n-1} \left[\frac{1}{p_{i+1}u_{i}} \left(\sum_{j=1}^{i} p_{j} \right) \left(\sum_{k=i+1}^{n} p_{k} \right) \right] / \left(\prod_{i=1}^{n-1} u_{i} \right) \left(\sum_{j=1}^{n} p_{j} \right)$$

$$= \frac{1}{\sum_{j=1}^{n} p_{j}} \sum_{i=1}^{n-1} \left[\frac{1}{p_{i+1}u_{i}} \left(\sum_{j=1}^{i} p_{j} \right) \left(\sum_{k=i+1}^{n} p_{k} \right) \right]$$
(17)

Henry & Eaton は、上式(17)をさらに誘導して最終的に次式を提案している:

$$\frac{1}{k_{f}} \propto \tau \equiv \int_{0}^{\infty} \frac{\langle \eta \rangle_{eq} - \langle \eta(t) \rangle}{\langle \eta \rangle_{eq} - \langle \eta(0) \rangle} dt$$

$$= \{\langle \eta \rangle_{eq} - \langle \eta \rangle_{0}\}^{-1} \sum_{j=0}^{n-4} \left\{ \frac{1}{p_{eq}(j)k_{j,j+1}} \sum_{i=j+1}^{n-3} \{p_{eq}(i) - p_{0}(i)\} \sum_{\eta=0}^{j} p_{eq}(\eta)(\langle \eta \rangle_{eq} - \eta) \right\}$$
(18)

第7章 フォールディングのキネティクスにおけるレートと緩和時間の関係は?

式(17)から, Henry & Eaton がどのようにして式(18)を導いたかを推測してみよう.

<参考1> 式(17)を反応座標 η を用いた式へ

上式(17)の緩和時間 τ を求める式を、反応座標 η (天然状態にあるアミノ酸残基数)を用いた式に 誘導することを試みよう.ここで、 η の範囲は、4 個のユニットが最小の局所構造であることから $0 \le \eta \le n-3$ である.特に、 $\eta=0$ 、及び、 $\eta=n-3$ は次の状態を表している:

η=0; アミノ酸残基はすべてランダムコイル状態にある

 $\eta = n - 3$; アミノ酸残基はすべて天然状態にある.

(1) 予備知識

(a) "詳細釣合 (detailed balance)"の式

平衡状態のとき、 $dp_{\eta}/dt = 0, (\eta = 0, 1, 2, \cdot \cdot \cdot, n - 3)$ より、次の"詳細釣合"(detailed balance)の式 が成立する:

$$w_{\eta}p_{eq}(\eta) = u_{\eta}p_{eq}(\eta+1)$$
 ($\eta = 0,1,2,3, \cdot \cdot \cdot, n-3, p_{eq}(0) = 1$)

ここで,

$$w_n$$
:前向きの反応 (状態 $\eta \rightarrow$ 状態 $\eta+1$)のレート (速度定数)

 u_n : 逆向きの反応 (状態 $\eta + 1 \rightarrow$ 状態 η) のレート (速度定数)

 $p_{eq}(\eta)$: 平衡状態において、天然状態にあるアミノ酸残基数が η である確率 ($0 \le \eta \le n-3$) (b) < $\eta(t)$ > の時間変化である反応速度方程式を用いる

< η(t) > の時間変化は,参考 2:「反応速度方程式」より,次のように表される:

$$\frac{d}{dt} < \eta(t) >= -k_{+} < \eta > +k_{-}(\eta - < \eta >)$$

ここで、 k_{\perp},k_{\perp} は遷移確率速度である.

いま,時刻 t において,状態の分子数が n である確率を P(n,t)で表すと,上式は, n の期待値の

式,
$$< n(t) >= \sum_{n=0}^{N} nP(n,t)$$
の時間変化を表す方程式とみなせる

平衡状態のときは、 $\frac{d}{dt} < \eta(t) >= 0$ である.このとき、天然状態にあるアミノ酸残基数が η であるときの遷移確率速度をそれぞれ、 $k_+ \rightarrow w_n, k_- \rightarrow u_n$ で表すと、次式が成立する:

$$w_{\eta} < \eta >_{eq} = u_{\eta}(\eta - < \eta >_{eq})$$

ここで、平衡状態において、天然状態にあるアミノ酸残基が η である期待値 < η_{eq} > は、平衡状態において、天然状態にあるアミノ酸残基が η である存在確率 $p_{eq}(\eta)$ を用いて、次式で表される;

$$<\eta>_{eq}=\sum_{\eta=0}^{n-3}\eta p_{eq}(\eta)$$

(c) 式(10)で定義した p_i を,反応座標 η を用いた式へ

 η の範囲は $0 \le \eta \le n - 3$ であるから、定義式(10)を次のように変換する:

$$p_{1} = 1 \qquad \rightarrow \qquad p_{0} = 1$$

$$p_{i} = \prod_{j=1}^{i-1} \frac{w_{j}}{u_{j}} \qquad (1 < i \le n) \qquad \rightarrow \qquad p_{\eta} = \prod_{j=0}^{\eta-1} \frac{w_{j}}{u_{j}} \qquad (0 < \eta \le n-3)$$

上式を平衡状態のときの式に誘導しよう. 右図の自由エネルギー曲線のイラストにおいて,



$$p_{\eta}$$
 は次のように誘導できる

$$p_{\eta} = \prod_{j=0}^{\eta-1} \frac{w_{j}}{u_{j}} = \frac{w_{0}}{u_{0}} \frac{w_{1}}{u_{1}} \frac{w_{2}}{u_{2}} \dots \frac{w_{j}}{u_{j}} \dots \frac{w_{\eta-1}}{u_{\eta-1}}$$
$$= \frac{e^{-\beta\Delta F_{0,1}^{+}}}{e^{-\beta\Delta F_{1,0}^{+}}} \frac{e^{-\beta\Delta F_{1,2}^{+}}}{e^{-\beta\Delta F_{2,1}^{+}}} \dots \frac{e^{-\beta\Delta F_{j,j+1}^{+}}}{e^{-\beta\Delta F_{j+1,j}^{+}}} \dots \frac{e^{-\beta\Delta F_{\eta-1,\eta}^{+}}}{e^{-\beta\Delta F_{\eta,\eta-1}^{+}}} = e^{-\beta\Delta F_{0,1}} e^{-\beta\Delta F_{1,2}} \dots e^{-\beta\Delta F_{j,j+1}} \dots e^{-\beta\Delta F_{\eta-1,\eta}} = e^{-\beta\Delta F_{0,\eta}}$$

$$= p_{eq}(\eta) - p_0(\eta)$$

ここで、 $p_0(\eta)$ は、時刻 t=0 における状態 η の確率を表す.

(2) 緩和時間 τ を求める式(17)を、反応座標 ηを用いた式へ

平衡状態までの緩和時間 τ を求める式(17)を,次の4つの項に分けて,それぞれの項毎に反応座標 η を用いた式へ誘導しよう:

(A)
$$\sum_{i=1}^{n-1} \sum_{j=1}^{i} p_{j}$$
, (B) $\sum_{i=1}^{n-1} \sum_{k=i+1}^{n} p_{k}$, (C) $\sum_{i=1}^{n-1} \frac{1}{P_{i+1}u_{i}}$, (D) 分母の式: $\sum_{j=1}^{n} p_{j}$
(A) $\sum_{i=1}^{n-1} \sum_{j=1}^{i} p_{j}$ を誘導しよう:

反応座標 η ($0 \le \eta \le n-3$)を採用すると、式の $\sum \sum$ の添え字は次のようになる:

$$\sum_{i=1}^{n-1}\sum_{j=1}^{i}p_{j} \quad \rightarrow \sum_{j=0}^{n-4}\sum_{\eta=0}^{j}p_{\eta+1}$$

第7章 フォールディングのキネティクスにおけるレートと緩和時間の関係は?

ここで, (a), (b)より次式が得られる:

$$p_{eq}(\eta+1) = \frac{w_{\eta}}{u_{\eta}} p_{eq}(\eta) = \frac{p_{eq}(\eta)(\eta - \langle \eta \rangle_{eq})}{\langle \eta \rangle_{eq}}$$

項(A)は、上式を用いると次式のように誘導できる:

$$\sum_{j=0}^{n-4} \sum_{\eta=0}^{j} p_{\eta+1} = \sum_{j=0}^{n-4} \sum_{\eta=0}^{j} p_{eq}(\eta+1) = \sum_{j=0}^{n-4} \sum_{\eta=0}^{j} \frac{p_{eq}(\eta)(\eta-\langle\eta\rangle_{eq})}{\langle\eta\rangle_{eq}} = \frac{1}{\langle\eta\rangle_{eq}} \sum_{j=0}^{n-4} \sum_{\eta=0}^{j} p_{eq}(\eta)(\eta-\langle\eta\rangle_{eq})$$

(B) $\sum_{i=1}^{n-1} \sum_{k=i+1}^{n} p_k$ を誘導しよう:

反応座標 η ($0 \le \eta \le n-3$)を採用すると、次式の $\sum \sum$ の添え字は次のようになる.

$$\sum_{i=1}^{n-1} \sum_{k=i+1}^{n} p_k \quad \rightarrow \sum_{j=0}^{n-4} \sum_{i=j+1}^{n-3} p_i$$

項(B)は、(c)より次式のように誘導できる:

$$\sum_{j=0}^{n-4} \sum_{i=j+1}^{n-3} p_i = \sum_{j=0}^{n-4} \sum_{i=j+1}^{n-3} \left\{ p_{eq}(i) - p_0(i) \right\}$$

(C)
$$\sum_{i=1}^{n-1} \frac{1}{P_{i+1}u_i}$$
を誘導しよう:

反応座標 η ($0 \le \eta \le n-3$)を採用すると、分母の式の \sum の添え字は次のようになる:

$$\sum_{i=1}^{n-1} \frac{1}{p_{i+1}u_i} \to \sum_{j=0}^{n-3} \frac{1}{p_{j+1}u_j}$$

項(C)は、(c)より次式のように誘導できる(但し、 $p_0(j) = 0$ とする):

$$\sum_{j=0}^{n-3} \frac{1}{p_{j+1}u_j} = \sum_{j=0}^{n-3} \frac{1}{(\prod_{j=0}^j \frac{w_j}{u_j})u_j} = \sum_{j=0}^{n-3} \frac{1}{(\frac{w_0}{u_0} \frac{w_1}{u_1} \frac{w_2}{u_m} \dots \frac{w_j}{u_j})u_j} = \sum_{j=0}^{n-3} \frac{1}{(\prod_{j=0}^{j-1} \frac{w_j}{u_j})w_j} = \sum_{j=0}^{n-3} \frac{1}{p_j w_j} = \sum_{j=0}^{n-3} \frac{1}{p_{eq}(j)k_{j,j+1}}$$
(25)

(D) 分母の式:
$$\sum_{j=1}^n p_j$$
を誘導しよう:

反応座標 η ($0 \le \eta \le n-3$)を採用すると、式の \sum の添え字は次のようになる:

$$\sum_{j=1}^n p_j \quad \longrightarrow \sum_{\eta=0}^{n-3} p_\eta$$

平衡状態での $0 \le \eta \le n-3$ における存在確率の和は1であることを用いると、

分母の項(D)は、次式のように誘導できる(但し、 $p_0(\eta) = 0$ とする):

$$\sum_{\eta=0}^{n-3} p_{\eta} = \sum_{\eta=0}^{n-3} p_{eq}(\eta) = p_{eq}(0) + p_{eq}(1) + p_{eq}(2) + \dots + p_{eq}(n-3) = 1$$

結局,式(17)の平衡状態までの緩和時間 τ (その逆数がフォールディング・レート: k_f)を求める式 は次式のようになる:

$$\begin{split} \frac{1}{k_f} &\propto \tau \equiv \int_0^\infty \frac{\langle \eta \rangle_{eq} - \langle \eta(t) \rangle}{\langle \eta \rangle_{eq} - \langle \eta(0) \rangle} dt \\ &= \frac{1}{\langle \eta \rangle_{eq}} \sum_{j=0}^{n-4} \left\{ \frac{1}{p_{eq}(j)k_{j,j+1}} \sum_{i=j+1}^{n-3} \{p_{eq}(i) - p_0(i)\} \sum_{\eta=0}^j p_{eq}(\eta)(\eta - \langle \eta \rangle_{eq}) \right\} \end{split}$$

Henry & Eaton は、上式をさらに誘導して最終的に次式を提案している:

$$\begin{split} \frac{1}{k_f} &\propto \tau \equiv \int_0^\infty \ \frac{\langle \eta \rangle_{eq} - \langle \eta(t) \rangle}{\langle \eta \rangle_{eq} - \langle \eta(0) \rangle} dt \\ &= \{\langle \eta \rangle_{eq} - \langle \eta \rangle_0\}^{-1} \sum_{j=0}^{n-4} \left\{ \frac{1}{p_{eq}(j)k_{j,j+1}} \sum_{i=j+1}^{n-3} \{p_{eq}(i) - p_0(i)\} \sum_{\eta=0}^j p_{eq}(\eta)(\langle \eta \rangle_{eq} - \eta) \right\} \end{split}$$

<コメント>

- 式(27)は絶対値である.
- $\bigcirc \langle \eta \rangle_0 = 0$ $\forall bar{\delta} \delta$.
- <理由>フォールディング過程の初期条件は、t = 0のとき、すべてランダムコイル状態にあり $\eta = 0$ と仮定しているので、時刻 t = 0 における状態の確率 $p_0(\eta)$ は次のようになる: $p_0(\eta) = \delta_{\eta,0}$ つまり、 $\eta = 0$ のとき $p_0(0) = 1$ 、 $\eta = 1, 2, 3..., n-3$ のとき $p_0(\eta) = 0$ よって、フォールディング過程では、 $<\eta >_0 = \sum_{n=0}^{n-3} \eta p_0(\eta) = 0$ である.
- (3) 遷移確率密度 $k_{j,j+1}, k_{j+1,j}$ を,存在確率 $p_{eq}(\eta)$ で表現しよう
 - ここで,遷移確率速度 $k_{j,j+1},k_{j+1,j}$ は次式で表されることを導こう:

$$k_{j,j+1} = \left(\frac{p_{eq}(j+1)}{p_{eq}(j)}\right)^{\kappa}$$
; 状態 j から状態 $j+1$ の遷移確率

$$k_{j+1,j} = \left(\frac{p_{eq}(j+1)}{p_{eq}(j)}\right)^{\kappa-1}$$
; 状態 $j+1$ から状態 j の遷移確率

まず、図の自由エネルギー曲線のイラストにおいて、次のように仮定する:

$$\Delta F_{j,j+1}^{+} = \kappa(\Delta F_{j,j+1})$$

(ただし, κ : adjustable parameter)

ここで、 $k_{j,j+1}, k_{j+1,j}$ は、右図より次式で表される:

$$\begin{split} k_{j,j+1} &\approx e^{-\beta \Delta F_{j,j+1}^{\ddagger}} \\ k_{j+1,j} &\approx e^{-\beta \Delta F_{j+1,j}^{\ddagger}} = e^{-\beta \{\Delta F_{j,j+1}^{\ddagger} - \Delta F_{j,j+1}\}} = e^{-\beta(\kappa-1)\Delta F_{j,j+1}} \\ \vdots & \supset \checkmark, \end{split}$$

$$\kappa = \left(\frac{\Delta F_{j,j+1}^{+}}{\Delta F_{j,j+1}}\right) = \frac{\ln(k_{j,j+1})}{\{\ln(k_{j+1,j})\}/(\kappa-1)}$$

 $\Delta F_{j,j+1}^{\dagger} \begin{pmatrix} k_{j,j+1} & k_{j+1,j} \\ k_{j,j+1} & k_{j+1,j} \end{pmatrix} \Delta F_{j+1,j}^{\dagger}$ $j \text{ the } \Delta F_{j,j+1}$

これより、 $(\kappa - 1) \ln(k_{j,j+1}) = \kappa \ln(k_{j+1,j})$ 故に、 $(k_{j,j+1})^{\kappa-1} = (k_{j+1,j})^{\kappa}$

ここで、"詳細釣合(detailed balance)"の式 $k_{j,j+1}p_{eq}(j) = k_{j+1,j}p_{eq}(j+1)$ を用いると、上式は次式のように誘導できる.

$$(k_{j,j+1})^{\kappa-1} = \left(k_{j,j+1} \frac{p_{eq}(j)}{p_{eq}(j+1)}\right)^{\kappa}, \quad \text{tit}, \quad (k_{j+1,j})^{\kappa} = \left(k_{j+1,j} \frac{p_{eq}(j+1)}{p_{eq}(j)}\right)^{\kappa-1}$$

これより,次式が導かれる:

$$k_{j,j+1} = \left(\frac{p_{eq}(j+1)}{p_{eq}(j)}\right)^{\kappa}, \qquad k_{j+1,j} = \left(\frac{p_{eq}(j+1)}{p_{eq}(j)}\right)^{\kappa-1}$$

Henry & Eaton らは、新たに次式を提案している:

$$k_{j,j+1} = \gamma \left(\frac{p_{eq}(j+1)}{p_{eq}(j)} \right)^{\kappa} \quad (状態 \ j \ h \delta t \% \ j+1 \ \sim \mathcal{O} 遷移確率)$$
$$k_{j+1,j} = \gamma \left(\frac{p_{eq}(j+1)}{p_{eq}(j)} \right)^{\kappa-1} \quad (t \% \ j+1 \ h \delta t \% \ j \ \sim \mathcal{O} 遷移確率)$$

ただし, κ と γ は adjustable parameter である. Henry & Eaton は, κ と γ の値は, 結果にあ まり影響されないので, $\kappa = 0.5$, $\gamma = 1$ を採用している.



"Winter Morning by the Stable"

<参考2> 「反応速度方程式」について

単分子反応を考える. 図のように、分子が独立に 2 つの状態 A と B を取り得るとする. 状態 A に ある分子を A 分子、状態 B にある分子を B 分子と呼ぶことにする. 時刻 t において A 状態のある分 子が時間 Δt の後に B 状態に見出される確率を $p_+(\Delta t)$ 、逆に、時刻 t において B 状態のある分子 が時間 Δt の後に A 状態に遷移している確率を $p_-(\Delta t)$ とする. 遷移をする分子は全体で N 個あるも のとして、個々の分子は独立に A と B の間を遷移しているものとする.



いま,時刻 t において A 状態のある分子数が n である確率 P(n,t) の時間変化を考えよう. 時刻 $t + \Delta t$ に A 分子が n 個であるのは,次の 3 通りのいずれかである:

- ① 時刻 t では n+1 個の A 分子があって、そのうちのどれか1 個が Δt 時間に $A \rightarrow B$ の遷移 を起こす $\rightarrow p_+(\Delta t)(n+1)P(n+1,t)$
- ② 時刻 *t* では n-1 個の A 分子があって, N-(n-1) 個の B 分子のどれか 1 個が Δt 時間に $B \rightarrow A$ の遷移を起こす $\rightarrow p_{-}(\Delta t)(N-(n-1))P(n-1,t)$
- ③ 時刻 t でもA分子の数は n 個であって、 Δt 時間にはまったく遷移が起こらない

 $\rightarrow \{1 - p_+(\Delta t)n - p_-(\Delta t)(N - n)\}P(n, t)$

ただし,時間間隔 Δt が十分短くて,その間に2個以上の分子が反応を起こす確率は無視できるもの としよう.この仮定のもとで,時刻 t においてA状態のある分子数が n である確率 $P(n,t+\Delta t)$ は ①~③より次式で表される:

 $P(n,t+\Delta t)$

 $= p_+(\Delta t)(n+1)P(n+1,t) + p_-(\Delta t)(N-(n-1))P(n-1,t) + \left\{1 - p_+(\Delta t)n - p_-(\Delta t)(N-n)\right\}P(n,t)$ $\exists n \downarrow 0,$

$$\frac{P(n,t + \Delta t) - P(n,t)}{\Delta t}$$

$$= \frac{p_{+}(\Delta t)}{\Delta t}(n+1)P(n+1,t) + \frac{p_{-}(\Delta t)}{\Delta t}(N-(n-1))P(n-1,t) - \frac{p_{+}(\Delta t)}{\Delta t}nP(n,t) - \frac{p_{-}(\Delta t)}{\Delta t}(N-n)P(n,t)$$
ここで、 $\Delta t \to 0$ の極限として次式が得られる:

$$\frac{dP(n,t)}{dt} = k_{+}(n+1)P(n+1,t) + k_{-}(N-(n-1))P(n-1,t) - k_{+}nP(n,t) - k_{-}(N-n)P(n,t)$$
(1)

だだし、 $0 \le n \le N$ である. k_+, k_- は遷移確率速度で、次式で定義される:

$$\lim_{\Delta t \to 0} \frac{p_+(\Delta t)}{\Delta t} = k_+ \quad , \quad \lim_{\Delta t \to 0} \frac{p_-(\Delta t)}{\Delta t} = k_- \tag{2}$$

式(1)の両辺に n を掛けて, n = 0,1,2,3,...,N について和をとって整理すれば次式のような反応速度方程式が導かれる:

$$\frac{d}{dt} < n(t) >= -k_{+} < n > +k_{-}(N - < n >)$$
(3)

上式は, nの期待値 $< n(t) >= \sum_{n=0}^{N} nP(n,t)$ の時間変化を表す方程式とみなせる.ここで, P(n,t)は時

刻 t において,ある状態の分子数が n である確率を表す. <文献> 寺本 英:『ランダムな現象の数学』,吉岡書店,1990年.

Henry & Eaton は、タンパク質のフォールディング過程を、天然状態にあるアミノ酸残基数 ηを反応座標とした1次元の自由エネルギープロファイルに沿った反応過程として考察している.

いま.フォールディング過程の途中で、天然状態にあるアミノ酸残基数が η であるとき、 η の 平均値 < $\eta(t)$ > の時間変化を表す式は、式(3)の類推より次式のように表される:

 $\frac{d}{dt} < \eta(t) >= -k_{+} < \eta > +k_{-}(\eta - < \eta >)$ (4)

平衡状態のとき、 $\frac{d}{dt} < \eta(t) >= 0$ である.このとき、天然状態にあるアミノ酸残基数が η であるときの遷移確率速度をそれぞれ、 $k_+ \rightarrow w_\eta, k_- \rightarrow u_\eta$ とすると、次式が成立する:

$$w_{\eta} < \eta >_{eq} = u_{\eta} (\eta - < \eta >_{eq}) \tag{5}$$

ここで、3 次元格子タンパク質での η の範囲は $0 \le \eta \le n-3$ である. また、平衡状態において、 天然状態にあるアミノ酸残基が η である期待値 $<\eta_{eq} >$ は、平衡状態において、天然状態にあるア ミノ酸残基が η である存在確率 $p_{eq}(\eta)$ を用いて、次式で表される;

$$<\eta>_{eq}=\sum_{\eta=0}^{n-3}\eta p_{eq}(\eta)$$
 (6)



"Sunlight Autumn"

タンパク質フォールディングの遷移状態の 構造を推定するΦ値解析とは?

く要旨>

第8章

タンパク質フォールディング過程を反応座標で表現 した自由エネルギー関数に、何故ピークが表れるかを 考察している. ピークは、フォールディングにおける 遷移状態 (transition state) に対応していて、この遷移 状態での構造的特徴を推定できる、組み換えDNA技 術によるアミノ酸置換の変異体を利用した"Φ値解析 法"を紹介している.

一方, 3次元格子タンパク質に対して, 野生型タン パク質のフォールディング・レートと, 変異体のアン フォールディング・レートから, 理論的にΦ値を求め る式を導いている.

そして,具体的に3次元格子タンパク質の各アミノ 酸残基のΦ値を求め,遷移状態でのフォールディング 核を推定している.その結果,3次元格子タンパク質 の同じ天然構造であっても,アミノ酸配列が異なるタ ンパク質では,それぞれ,異なるセグメント領域の フォールディング核をもち,それ故,異なるフォール ディング過程であると推定している. 第8章 タンパク質フォールディングの遷移状態の構造を推定するΦ値解析とは?

耳 タンパク質・フォールディングの"遷移状態"とは?

1970年代以降に行われた、タンパク質のフォールディングの実験において、次のことが観測された:『小さな球状タンパク質では、変性状態から天然状態へのフォールディングの途中で、準安定状態である寿命の短い中間体は観測されない、二状態転移である.』

タンパク質のフォールディング過程を反応座標で表現した"自由エネルギー関数"の特徴は、天然 状態(N状態)とランダムコイル状態(D状態)に対応する反応座標のところに、それぞれ極小値を もち、その中間の位置にピークがあることである.このピークに対応する状態を"遷移状態"(transition state)と呼ぶ.この遷移状態を乗り越えるときに形成される部分構造は多様であろう.遷移状態での 多様な構造を"遷移状態アンサンブル"と呼ぶことにする.しかしながら、遷移状態アンサンブルは タンパク質の天然構造の形に応じて、ある程度制限は受けるであろう.

タンパク質のフォールディング過程のおける"変性状態の構造"や、"中間体の構造"、そして"天 然状態の構造"については、X線やNMRなどにより直接的な観測が可能である.だが、X線やNMR などを用いて、この"遷移状態の構造"を直接観測することは、その性質上、ほぼ不可能であろうと 思われる.二状態性を示す小さな球状タンパク質のフォールディング過程の"遷移状態の構造"を推 定することは原理的に大変難しいと思われた.この"遷移状態の構造"を推定する方法があるだろう か? "遷移状態で何が起こっているのか"を調べることが可能だろうか?

今日,遺伝子工学が発展した結果,タンパク質のアミノ酸配列において,任意のアミノ酸残基の置換・欠失・挿入などを自在に行うことが出来るようなった.これを"変異体タンパク質"と呼ぶ.タンパク質フォールディングに関する研究にも、タンパク質工学的に創製した"変異体タンパク質"を利用することができるようになり、タンパク質のフォールディング過程の"遷移状態の構造を推定する"新たな方法が可能になった.

1990年代のはじめに, Fersht らは、"Φ値解析法"という新たな手法により、フォールディング過程の"遷移状態での構造的特徴を推定する"ことを可能にするような画期的な方法を提案した.彼らは、組み換え DNA 技術によるアミノ酸置換の変異体を利用して、フォールディングの遷移状態において各アミノ酸残基の状態が、部分的に、天然状態かどうかを知る方法を考えた.

- <文献> Fersht, A.R., Matouschek, A. & Serrano, L.: The folding of an enzyme. 1. Theory of protein engineering analysis of stability and pathway of protein folding, *J. Mol. Biol.*, 224:771-782, 1992..
- <文献> A. Fersht, Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding (Freeman, New York, 1999). (日本語訳:桑島邦博・有坂文雄・熊谷泉(訳),タンパク質の構造と 機能, 医学出版(2006))

我々は,彼らによって導入されたΦ値解析という方法によって,"3 次元格子タンパク質フォールデ ィングの遷移状態での構造"を推定しよう.そして,次章では,Φ値解析の方法を"実際のタンパク 質"に適用しよう.

まず,一般に,二状態転移が観測されるタンパク質のフォールディング過程を表現している自由エ ネルギープロファイルについて考察しよう.

右図は、タンパク質フォールディングの転移温度($T = T_m$)での二状態転移における自由エネルギーを秩序度(フォールドの程度を表す量)の関数として模式的に描いたものである.図中で、N:天然状態、 $\ddagger: 遷移状態, D: ランダムコイル状態(変性状態)を示している.$

この自由エネルギー関数の特徴は、天然状態(N状態) とランダムコイル状態(D状態)に対応する秩序度のところに、それぞれ極小値をもっていることである.

N 状態での極小点付近は、その構造がほぼ一つの特異的 立体構造であることを反映して幅が狭いのに対して、D 状



秩序度

態での極小点は、この付近での秩序度に対して広範囲の状態の集まりを反映して幅広い.これらの二つの極小点の間には、自由エネルギー曲線でピークがある.このピークに対応する状態が"遷移状態"である.

タンパク質のフォールディング過程の描像は次の様であろう:

『最初は、ペプチド鎖がランダムコイル状態で揺らいでいて、あちこちの部位で局所構造を形成する が、相互作用するアミノ酸残基数が少ないため獲得するエネルギーが小さく、構造が制限されたこと により失うエントロピーを上回ることができないでので、ほとんどの場合、もとのランダムコイル状 態に戻ってしまうであろう.たまに、短距離相互作用によって形成されたいくつかの局所構造同士が ぶつかって合体し、中距離・長距離相互作用によって、ある程度の大きさをもった局所構造が形成さ れると、自由エネルギーのピークを越え、天然構造へとフォールドする.』

局所構造がある程度の大きさをもつことが必要であり、そのような局所構造を"核"と呼ぶ. 核が 形成され、ひとたび自由エネルギーの山を越えると、タンパク質分子は、天然構造へと一気にフォー ルドすることになる.

郷信広は,1983年の日本生物物理学会誌「生物物理」で,タンパク質分子が変性状態から天然状態 へのフォールディング過程を,上で示したような自由エネルギー曲線を用いて,次のように詳細に解 説している:

『・・・小さな球状蛋白質の変性・再生転移のもっとも基本的な統計物理学的特徴は、その二状態性にある. 転移中点におけるいろいろな立体構造の分布をみると、いわゆる変性状態と生の状態における特異的立体構造(再生状態)のところにピークをもつことを、これは意味する. 中間状態の分布密度は低い.・・・それから計算し得る量 $-k_BT \ln(\partial \pi \alpha \pi g)$ は、系が、反応座標(折れたたみを記述する適当な変数)のある与えられた値で特徴付けられる状態にあるときの自由エネルギーである(ここで、 k_B はボルツマン定数、T は絶対温度である). 状態分布の2つのピークに対応して、自由エネルギー曲線には、2 つの極小が存在し、その間には山がある. この山は変性・再生転移の活性化状態に対応している. 活性化状態は一種の臨界状態で、この状態より再生状態側にある分子はさらに折れたたまれる傾向があり、変性側にある分子は、折れたたみはほどける傾向がある. 言い換えれば、活性化状態から遠ざかる方向の統計的力をうける. この力は統計的であるから揺らぎとして活性化状態の山の方向に近づく分子もあるであろうが、平均としては山から遠ざかる方向に動く.・・・変性状態にある分子は、自由エネルギー曲線の変性状態側の極小点の周りを行ったり来たりしている. 時には活性状態に近づくであろうが、大体はもとに戻ってしまう. こうして分子は長らく変性状態にとどま

125

第8章 タンパク質フォールディングの遷移状態の構造を推定するΦ値解析とは?

る.極稀には、大きく揺らいで活性化状態の山を越えることもあろう.一旦超えれば、折れたたみが 進行する方向に統計的力が働いてほとんど確実に再生状態に至る.・・・分子が変性状態の極小点の周 りを行ったり来たりしているときには、実は、短距離相互作用によって形成された不安定な局所構造 ができたり壊れたりしていると考える.そのような局所構造を、相転移の統計力学の用語にならって、 "胚"と呼ぶ.胚はそれ自身では不安定である.・・・胚は揺らぎによって自分自身に近接している鎖 を取り込んで成長したり、他の胚とぶつかってうまくすれば合体して成長することもあるだろう.・・・ 胚が成長してある臨界的な大きさに達したものを"核"と呼ぶ.・・・核とは変性・再生途上の活性化 状態にある分子中にみられる秩序構造である.・・・』(郷信広、『蛋白質の立体構造予測と折れたたみ過程』、 生物物理、Vol. 23 No.3, 11-19, 1983)

天然構造とランダム構造の差は,エンタルピー(H)およびエントロピー項(ST)が 100kcal/molのオーダーであるのに対して,その差である自由エネルギーは5~15kcal/molほどである.1つの疎水性相互作用,水素結合,塩結合などの自由エネルギーへの寄与が1~3kcal/molであることを考えると,きわめて微妙なバランスの上に立体構造の安定性は保たれていることになる.しかも,一部のアミノ酸置換が安定性に大きく影響を及ぼすこともあれば,さほど影響しない場合もあり,タンパク質立体構造の安定性は、個々のアミノ酸の性質の単純な総和では議論することはできない.より大きな二つ以上のドメインからなるタンパク質の場合は、そのフォールディングは二状態的ではなく、中間状態を経ることが多いが、やはり協同的であり、現象の基本は、小さい球状タンパク質のフォールディングの示す二状態性にある.

□□ なぜ自由エネルギー曲線に極大点が存在するのか?

自由エネルギー曲線のピークに対応する遷移状態が、如何にして表れるのだろうか?

図で、横軸は、フォールディングの進行変数である"秩序度"であり、0.0:ランダムコイル状態、 1.0:天然状態に対応している.

自由エネルギーは、エンタルピー(H)曲線とエントロピー項 (-*ST*)の曲線の和で表される: F = H + (-ST)

秩序度が 0.0 である D 状態では,獲得エネルギー(H) は 0 である.

D 状態は無数のランダムコイル状態の構造を含むアンサンブルで あり、その鎖の立体配置エントロピー($S = k_B \ln W$,ここで、 W は、微視的状態数を、 k_B はボルツマン定数を表す)は非常に 大きい、したがって、(-ST) は小さな値となる、

変性状態からタンパク質のフォールディングが進みN状態に近づくにつれて, H曲線は, 負の立体構造エネルギーを獲得するので徐々に減少する.



一方,フォールディングが進むにつれて,微視的状態数Wが減少し,鎖の立体配置エントロピー(S) はだんだん減少するため(-ST)曲線は徐々に上昇していく.N状態に近づくにつれてエントロピーは 急激に減少する,つまり,(-ST)曲線は急激に増加する.

その結果,自由エネルギー曲線は上に凸となり,D状態からN状態への途中で極大となる.秩序度

が 1.0 である N 状態では、タンパク質の天然構造はほぼユニークに決まっている. $W \approx 1$ であり、 $S = k_B \ln W \approx 0$ となるので、(-ST) 曲線も 0 に近づく.

したがって, *H*曲線と(-*ST*)曲線の和である自由エネルギー曲線(*F* = *H* + (-*ST*))は,図のよう に,D状態とN状態で極小点が存在することになる.(文献: V. Daggett and A. R. Fersht;「タンパク質フォ ールディングの遷移状態」(第7章),『タンパク質のフォールディング』,第2版, R. H. Pain 編,河田・桑島訳, 丸善出版,2000,笹井理生著;『蛋白質の柔らかなダイナミクス』,培風館,2008,)

□ アミノ酸残基を置換すると自由エネルギーが変化する!

Fersht らが提案した "Ф値解析"という手法の特徴は, "野生型タンパク質"のアミノ酸配列中の, ある部位のアミノ酸残基を, "組み換え DNA 技術"を用いて, 他の残基に置換して新たな, "変異体 タンパク質"を作成することにある.

この,ある部位のアミノ酸残基を他のアミノ酸残基に置換するという変化は、物理的観点からみる と、野生型タンパク質に加えられた"摂動"とみなせる.元の系と"摂動"による新たな系を比較・ 検討することによって、元の系を理解する上で格好の情報が得られることが期待される.

上記のようなアミノ酸置換によって、自由エネルギープロファイルはどのように変化するのだろうか?

いま,右の模式図ように,特定の部位のアミノ酸残基を,例え ばより小さなアミノ酸残基に置換することによって,置換前のア ミノ酸残基との相互作用が消滅すると,構造エネルギー(H)が減 少することになり,遷移状態が不安定になる.

結果的に, 遷移状態での自由エネルギーのピークが d₁だけ上昇 しタンパク質フォールディング・レート(反応速度)は, 置換前 の野生型タンパク質より遅くなる.ここで, アミノ酸置換による エントロピーの影響は無視できるほど小さいとみなす.

もし、この置換によって天然状態が同じように不安定になって、 天然状態での自由エネルギーのレベルが d₀ だけ上昇すると、 次 のような量が定義できる(Φ値の定義):

$$\Phi = \frac{d_1}{d_0}$$

ここで、 $d_0 = d_1$ つまり、 $\Phi = 1$ のとき(上図)は、遷移状態において、その部位を取り巻く空間的な環境が天然状態と同じであり、アミノ酸残基間相互作用も天然状態と同じ相互作用をしているとみなせる、つまり、遷移状態においてその部位は局所的に天然状態と同じ構造を形成していると推定できる.

一方,下の模式図のように,その部位でのアミノ酸置換によって 天然状態が不安定になり天然状態での自由エネルギーのレベルが do だけ上昇したにもかかわらず,フォールディングの反応速度が 変化しない場合を考えよう:



127

第8章 タンパク質フォールディングの遷移状態の構造を推定するΦ値解析とは?

 $d_1 = 0$, つまり, 遷移状態での自由エネルギーでの変化がない場合, $\Phi = 0$ となる. 結局, アミノ酸置換したその部位は, 遷移状態では局所的に天然状態と同じ構造を形成しておらず, まだ大きく揺らいでいる, つまり, 変性状態と同じ環境にあるといえる. 更に, $0 < \Phi < 1$ の場合, つまり, 中間的な Φ 値のときは, その部位でのアミノ酸残基間相互作用が 形成途上であるとみなせるであろう.

Ц "Φ値解析法"による、タンパク質フォールディングの"遷移状態"の

構造の推定

**
 Φ**値は、タンパク質の特定のアミノ酸残基が、フォールディングの遷移状態で、どの程度天然状態
 と同じ構造を形成しているかを、0から1の数値で示す指標であると

見なせる.その理由は、すでに前節の"Φ値の定義"の項で述べているが、改めて詳細に確認しよう(略図参照):

(a) Φ=1 の場合

"野生型タンパク質"のアミノ酸配列の特定の部位のアミノ酸を, 他のアミノ酸に置換したことにより,天然状態を不安定化させたと しよう.

っまり,天然状態での自由エネルギーが上昇する.一方,置換に よって, "変異体タンパク質"のフォールディングの反応速度 (k_f^{mut})が, "野生型タンパク質"のフォールディングの反応速度 (k_f^{wild})より大になれば,遷移状態での自由エネルギーも上昇する



ことになる.

このことは、置換によって削除された相互作用が遷移状態も不安定化していることになり、その 相互作用は遷移状態においてすでに天然状態と同じ環境にある、つまり、このアミノ酸残基を含む 部分構造は、天然状態での部分構造に似ていると推測される.

(b) Φ=0 の場合

"野生型タンパク質"のアミノ酸配列の特定の部位のアミノ酸を 他のアミノ酸に置換したことにより、天然状態が不安定になったに もかかわらず、"変異体タンパク質"のフォールディングの反応速 度(k_f^{mut})は変化しないで、"野生型タンパク質"のフォールディ ングの反応速度(k_f^{wild})と同じならば、遷移状態での相互作用は天 然状態での相互作用と同じではなく、むしろ、ランダムコイル状態 (変性状態)と同じ環境にあるとみなせる.

つまり,このアミノ酸残基が遷移状態でとる構造は,天然状態で とる構造ではなく変性状態のままであると推測される.



(c) 0<Φ<1 の場合</p>

中間的なΦ値をとる場合は、解釈するのが難しいが、次の二つの可能性が考えられる:

(i)相互作用が形成途中で、部分的にほどけた構造である場合

(ii) 多様な遷移状態での平均的な構造をみていると考えられる場合

対象とするタンパク質に対して、そのアミノ酸残基の多くの部位でアミノ酸置換を行って、多くの 変異体を作成し、その変異体毎のΦ値を求め、Φ値が大きいアミノ酸残基を調べることによって、フ ォールディング遷移状態で天然構造と類似の構造を作る部位("フォールディング核")を推定することが可能である.

ここでは, "3 次元格子タンパク質"に対して, 統計力学モデルである "A-W_NILS モデル"を用い てΦ値を計算し, "3 次元格子タンパク質フォールディングの遷移状態での構造"を推定しよう.

次章では、Φ値解析の方法を"実際のタンパク質"に適用してΦ値を理論的に計算し、その結果と、 実験的に観測されたΦ値とを比較・検討して、"実際のタンパク質のフォールディング遷移状態での核 形成部位"を推定しよう.

<文献> H. Wako and H. Abe, Study of Folding/Unfolding Kinetics of Lattice Proteins by Applying a Simple Statistical Mechanical Model for Protein Folding, in Protein Folding, E.C. Walters ED, pp. 349-376, 2011.

<文献> 安部 晴男,山内 経則,輪湖 博, "タンパク質フォールディングのキネティクス,Ⅲ. Φ値解析に よるフォールディング核の推定",西日本工業大学紀要,第41卷, pp. 75-81, 2011.

□□フォールディング/アンフォールディング・レートとΦ値の関係

図は、"野生型の自由エネルギー曲線"(赤色)と"変異体の自由エネルギー曲線"(青色)を模式的 に描いている.

いま,タンパク質の"フォールディングにお けるキネティクス"を考察しよう.

野生型において、"遷移状態([‡])のでの自由 エネルギーの値とD状態での自由エネルギーの 値との差"を $\Delta F_f^{wild_{\pm}}$ とすると、"野生型タンパ ク質のフォールディング・レート k_f^{wild} " は、 次式で表される(k_0 ;任意の定数, k_B ;Boltzmann 定数):

$$k_f^{\text{wild}} = k_0 e^{-\Delta F \frac{\text{wild}}{f} + k_B T}$$

また、変異体において、"遷移状態([‡])で 秩序度 の自由エネルギーの値とD状態での自由エネルギーの値との差を ΔF_{f}^{mut} " とすると、"変異体タンパ ク質のフォールディング・レート k_{f}^{mut} "は、 次式で与えられる: $k_{f}^{mut} = k_{0}e^{-\Delta F_{f}^{mut} \ddagger/k_{B}T}$

遷移状態での、"変異体タンパク質の自由エネルギーの値と野生型タンパク質の自由エネルギーの値 との差 $\Delta\Delta F_{\ddagger - D}$ "は、次式のようになる:

 $\Delta\Delta F_{_{_+-D}^{_+}} = \Delta F_f^{\textit{mut}_+^+} - \Delta F_f^{\textit{wild}_+^+}$

 $= -k_B T (\ln k_f^{mut} - \ln k_f^{wild})$

 $= -k_B T \Delta \ln k_f$

(ただし、 $\ln k_f^{mut} - \ln k_f^{wild} \equiv \Delta \ln k_f$ とおく)

次に、タンパク質の"アンフォールディングにおけるキネティクス"を考えよう.

野生型タンパク質において,遷移状態([‡])のでの"自由エネルギーの値と D 状態での自由エネルギ ーの値との差"を ΔF_{u}^{wild} とすると、"野生型タンパク質のフォールディング・レート k_{u}^{wild} "は、 次式で表される:

$$k_u^{wild} = k_0 e^{-\Delta F \frac{wild}{u}/k_B T}$$

また,変異体タンパク質において,遷移状態([‡])での"自由エネルギーの値とD状態での自由エ ネルギーの値との差を ΔF_u^{mut} " とすると、"変異体タンパク質のフォールディング・レート k_u^{mut} " は、次式で与えられる:

$$k_u^{mut} = k_0 e^{-\Delta F_u^{mut} / k_B T}$$

タンパク質のアンフォールディングにおいて、N状態での"変異型の自由エネルギーの値と、野生 型の自由エネルギーの値との差 $\Delta\Delta F_{N-D}$ "は、次式のようになる:

$$\Delta \Delta F_{N-D} = (\Delta F_u^{wild} + \Delta \Delta F_{+-D}) - \Delta F_u^{mul}$$

$$= (\Delta F_u^{wild} - \Delta F_u^{mut}) + \Delta \Delta F_{\frac{1}{2} - D}$$
$$= -k_B T ((\ln k_f^{mut} - \ln k_f^{wild}) - (\ln k_u^{mut} - \ln k_u^{wild}))$$
$$= -k_B T (\Delta \ln k_f - \Delta \ln k_u)$$

(ただし、 $\ln k_f^{mut} - \ln k_f^{wild} \equiv \Delta \ln k_f$, $\ln k_u^{mut} - \ln k_u^{wild} \equiv \Delta \ln k_u$ とおく) 故に、Φ値は次のように表される;

$$\Phi = \frac{\Delta \Delta F_{\ddagger\text{-D}}}{\Delta \Delta F_{\text{N-D}}}$$

 $=\frac{\Delta \ln k_{\rm f}}{\Delta \ln k_{\rm f} - \Delta \ln k_{\rm u}}$

■ 転移温度におけるΦ値の計算

図は、フォールディング・レート (k_t) とアンフォールディング・レート (k_u) の温度 (の逆数) 依存曲線(シェブロン・プロット, Chevron plot)のイラストである.

転移温度(Tm)においては、図に示し ているように, 次式が成立する;



故に,

$$\Delta \Delta F_{N-D} = -k_B T_m ((\ln k_f^{mut}(T_m) - \ln k_f^{wild}(T_m)) - (\ln k_u^{mut}(T_m) - \ln k_u^{wild}(T_m)))$$
$$= -k_B T_m (\ln k_f^{mut}(T_m) - \ln k_u^{mut}(T_m))$$

さらに、 $\Delta\Delta F_{\pm -D} = -k_B T_m \Delta \ln k_f(T_m)$

結局,転移温度(Tm)におけるΦ値は次のように表される;

$$\Phi = \frac{\Delta \Delta F_{\ddagger\text{-D}}}{\Delta \Delta F_{\text{N-D}}}$$

 $=\frac{\Delta \ln k_f(T_m)}{\ln k_f^{mut}(T_m) - \ln k_u^{mut}(T_m)} \qquad (\quad \ \ \check{\mathcal{T}}\check{\mathcal{T}}\check{\mathcal{L}}\,, \ \ \Delta \ln k_f(T_m) = \ln k_f^{mut}(T_m) - \ln k_f^{wild}(T_m) \quad)$

このように、各アミノ酸残基での Φ 値は、転移温度(T_m)における、次の3個の値から計算することができる(前ページのイラスト参照):

 $k_u^{mut}(T_m)$ ・・・転移温度における、変異体タンパク質のアンフォールディング・レートの値 $k_f^{wild}(T_m)$ ・・・転移温度における、野生型タンパク質のフォールディング・レートの値 $k_f^{mut}(T_m)$ ・・・転移温度における、変異体タンパク質のフォールディング・レートの値



"Path through Snowfall"

図は、具体的な例として、3 次元 格子タンパク質 al の野生型(wild type, W.T.)と、16番目のアミノ酸残 基である Arginine を Methionine に置 換した変異体(mutant type, mut.,) に対する、(a)自由エネルギー曲線 と、(b)シェブロン・プロットを描 いたものである.

ここで,**D**;変性状態,[‡];遷移 状態, N;天然状態,*T*_m;転移温 度(*T*_m=0.240)



📮 "Φ値"より、3 次元格子タンパク質値の"フォールディング核"の推定

図は、4個のタンパク質(Protein al, a2, b1, b2) それぞれに対して、*i*-番目のアミノ酸残基を他の19種類のアミノ酸残基に置換した場合の、アミノ酸残基iのΦ_i値(平均値)を描いている.

図の大きい赤丸は、 $\Phi_i \ge 0.8$ 、小さい 赤丸は、 $0.3 \le \Phi_i < 0.8$ 、に対応するア ミノ酸残基を示している.

大きい赤丸に対応している残基は, 遷移状態で,天然構造とほぼ類似の環 境にあること,つまり,それらのアミ ノ酸残基は,遷移状態でほぼ天然構造 と類似の構造を形成している確率が高 いことを示している(フォールディン グ核とみなせる).

Protein a1 (図の(a1)) では,

"セグメント 15-27 と 31-36 の二つの 領域"が,"フォールディング核"と推 定される.

一方, Protein a2 (図の(a2)) でのフ オールディング核としては、"セグメン ト 8-27 の領域"が推定される.

ここで,残基 19-20 でのΦ値は小さい. その理由として,遷移状態では,



残基 20 と残基 31 との長距離相互互作用,残基 19 と残基 32 との長距離相互互作用が形成されていない為だとみなせる.

結局,タンパク質 Protein a2 の天然構造を考慮すると,この残基 19-20 も含めて,"セグメント 8-27 の領域"をフォールディング核と見なすことが妥当であろう.

タンパク質 al と a2 は、同じ天然構造でもっているが、それぞれのアミノ酸配列が異なっている. 上記のことから、それぞれの遷移状態での構造は異なっている、つまり、フォールディング核も異なっていることがわかる.

詳細にみると、フォールディング核は、一部のセグメント 15-27 は共通であるが、そこから、 *al* の方は、共通のセグメントから C-末端へ延びていて Φ 値も大きく、これを中心にしてフォールドして いるとみなせる.

一方, a2 では, 共通のセグメントから N-末端へ延びていて, しかも Φ 値は, a1 の場合の値と較べ て小さく, 分散的であることがわかる.このことは, a2 では, 一次配列上で離れた残基間の相互作 用(長距離相互作用)をたくさん準備し, これを主な駆動力としてフォールドしているとみなせる.

上記のことは、タンパク質 b1 と b2 の場合にも顕著に見られる.

b1 (図の(b1))のフォールディング核は、"セグメント 19-37 の領域"と推定される.ここで、残基 32-33 でのΦ値は小さい.その理由は、残基 32 と残基 47、及び、残基 32 と残基 47 との長距離相互作 用が遷移状態では形成されていないで、最終段階で形成される為だとみなせる.しかしながら、天然 構造を考慮すると、この残基 32-33 もフォールディング核に含むことが妥当であろう.

一方, b2 (図の(b2))のフォールディング核は、"セグメント 4-31 の領域"と推定される.ここで、 残基 5-6, 12-13, 23, 25 でのΦ値は小さい.その理由は、残基 5 と残基 40、残基 6 と残基 39、残基 12 と 残基 33、残基 13 と残基 44、残基 23 と残基 46、及び、残基 25 と残基 38 との長距離相互作用が、遷移 状態では形成されていないで、最終段階で形成される為だとみなせる.しかしながら、天然構造を考 慮すると、この残基 32-33 もフォールディング核に含むことが妥当であろう.

*b1*の方は, Φ値が大きく, フォールディング核を中心にしてフォールドし, *b2*の方は, Φ値は*b1*の場合の値と較べて小さく, 分散的であり, *a2*の場合と同様に, 一次配列上で離れた残基間の相互作用(長距離相互作用)をたくさん準備し, これを主な駆動力としてフォールドするとみなせるだろう.



"Lambs under Blossom Tree"



"Morning Light"



第9章 統計力学モデルによるΦ値計算とフォールディング・メカニズム

➡ 実験によるプロテインAの遷移状態のΦ値による核形成部位の推定

我々は、フォールディング・メカニズムを明瞭にする問題、つまり、『タンパク質は、ランダムコイル状態から如何にして天然構造へフォールドするのだろうか?』という問題に取り組んでいる。そのための第一歩として、タンパク質のフォールディング過程の"遷移状態の構造"を明らかにしたいが、 遷移状態はきわめて寿命が短く、しかも不安定のため、通常の構造解析手法は使えない。

多くの研究グループが興味のある挙動を示すタンパ ク質である、"プロテイン A (1SS1)" (B domain of protein A) のコンピュータ・シミュレーションを行い、 そのフォールディング過程の解明に取り組んでいる. "プロテイン A" (「黄色ブドウ状球菌プロテイン A の

プロテインA (1SS1) のアミノ酸配列

1 10
Gly – Ser – Thr – Ala – Asp – Asn – Lys – Phe – Asn – Lys – Glu – Gln – Gln -
Asn – Ala – Phe – Trp – Glu – Ile – Leu – His – Leu – Pro – Asn – Leu – Asn -
Glu - Glu - Gln - Arg - Asn - Gly - Phe - Ile - Gln - Ser - Leu - Lys - Asp -
Asp – Pro – Ser – Gln – Ser – Ala – Asn – Leu – Leu – Ala – Glu – Ala – Lys -
Lys – Leu – Asn – Asp – Ala – Gin – Ala – Pro – Lys - Ala

B ドメイン)は、アミノ酸残基数は62 個で、3 個のα-ヘリックス

(α1, α2, α3)をもつ小さなタンパク質である.右の表に、"プロテイン A"のアミノ酸配列を示 している.

このタンパク質の特徴は、ある条件下では、フォールディング速度が 10 μ 秒程度と非常に速いこと である. これは、現在のコンピュータ・シミュレーションのタイムスケールに近い. タンパク質フォ ールディングのコンピュータ・シミュレーションは、フォールディングの道筋に関する情報を提供で きる強力な方法である. 実際、Wolynes は、"プロテイン A"に対する、全原子レベルの分子動力学 (Molecular dynamics; MD)によるコンピュータ・シミュレーションを行い、次のような結果を得た: 『"プロテイン A"は、まずα3 が形成され、その後、α1、α2、あるいはターンが形成される.』

<文献> P.G. Wolynes, Proc. Natl Acad. Sci. USA 101, 6837-6838 (2004).

この結果が妥当であるのか、実験による解明が切望されていた.

タンパク質のフォールディング過程の"遷移状態の構造"を、実験的に明らかにしようとする試み として、Fersht らによって導入された"Φ値解析"がある.彼らは、遺伝子工学的手法を用いて、も との"野生型タンパク質"のフォールディング反応と、任意の部位のアミノ酸残基を他のアミノ酸残 基で置換した"変異体タンパク質"のフォールディング反応とにおける、"反応速度の相違"と"自由 エネルギーの変化"を観測することによって、フォールディング過程の遷移状態の構造を推定しよう とする解析方法を提案し、"実際のタンパク質"に対して実行している(Appendix I:「実験における Φ値解析とは?」参照).

彼らは、具体的に、"プロテインA"のフォールディングの遷移状態におけるアミノ酸残基のΦ値 を、組み換え DNA 技術によるアミノ酸置換の変異体を利用して、系統的に、そして網羅的に測定す る実験を行った.そして、コンピュータ・シミュレーションにおける遷移状態での構造と、実験によ るΦ値解析によって得られた遷移状態の構造とを比較した.

- <文献> Sato, S., Religa, T.L., Daggett, V. & Fersht, A.R. Testing protein-folding simulations by experiment: B domain of protein A. *Proc. Natl Acad. Sci. USA* **101**, 6952 6956 (2004).
- < 文献> 佐藤 聡;「アミノ酸変異導入による蛋白質折れたたみ遷移状態の解析」, 蛋白質・核酸・酵素 Vol.49,No.14(2004).

次のページの図は, Fersht らによる "1SS1" (プロテイン A) の実験によるΦ値の結果を, アミノ酸 配列上と天然構造に重ね合わせて, 3 色で分類して描いている.

- この結果によって推定される遷移状態での構造形成度の考察は次の通りである:
- "α1-ヘリックスの領域"のC末端側の残基でのΦ値は高いが、他の残基でのΦ値は0.3~0.5 と 低い、このことは、遷移状態においてはα1-ヘリックス領域自体の構造形成度は低く、"構造形成の 途上である"とみられる.
- (2) "α2-ヘリックス領域"でのΦ値は全般的に高い値である.このことは、α2-ヘリックスの領域は、"遷移状態においてほぼ天然構造を形成している"とみなせる.

- (3) " α 3-ヘリックスの領域"のN 末端側の Φ 値は 0.4~0.7 という比較的大きい値を示すものの、C 末端側に向かって徐々に減少している.このことは、遷移状態においてはα3-ヘリックスのN末端 側ではある程度天然構造を形成しているとみられるが、この領域自体の"構造形成度は低い"とみ られる.
- (4) "ターン領域 (T1)" (α 1-ヘリックスと α 2-ヘリックスを結ぶ領域)の Φ 値は低く,遷移状態に おいては、"ほぼ変性状態"とみなせる。
- (5) "ターン領域 (T2)" (α 2-ヘリックスと α 3-ヘリックスを結ぶ領域) の Φ 値は中間的な値であり、 遷移状態においては、"ある程度天然構造を形成している"とみなせる。



プロテインA(1SS1)のΦ値(実験)

(無色は実験値が無い残基を示す)

◎実験による、『プロテインAの遷移状態での核形成部位の推定』(まとめ)

- ・"プロテイン A (1SS1)"の遷移状態では、"α2-ヘリックス領域がほぼ天然構造を形成"してい て、そのα2-ヘリックスが、α1-ヘリックスとα3-ヘリックスに位置する多数の疎水性アミノ酸 による相互作用によって安定化されている.このことは、シミュレーションによる結果である、 「まずα3が形成され,その後,α1,α2,あるいはターンが形成される」との予測はみられな かったことになる.
- ・具体的には、α2-ヘリックス領域の疎水性アミノ酸残基である、"Ile-34 と Leu-37"(ともに高い Φ 値を示している)と相互作用している α 1-ヘリックス領域の疎水性アミノ酸残基 "Ile-19 と

第9章 統計力学モデルによるΦ値計算とフォールディング・メカニズム

Leu-20"(ともに高いΦ値を示している)が天然構造形成過程の引き金となる重要な核であると みなせる.

佐藤らは,実験で求められた"プロテイン A (1SS1)"(プロテイン A) のΦ値の分布から,フォー ルディングのメカニズムは、"核形成-凝縮モデル"に近いと,次のように推論している:

『遷移状態で, α2-ヘリックス領域のアミノ酸残基のΦ値は高く, 2 次構造を形成しているものの, α1-ヘリックスとα3-ヘリックスのアミノ酸残基のΦ値は中間的な値を示していて,これらの領域の 2 次構造形成度は低い. "フレームワーク モデル"が該当するためには,α1-ヘリックスとα3-ヘリ ックスも高度に 2 次構造を形成しているはずである.プロテイン A (1SS1)では,むしろ,"核形成-凝縮モデル"に近い,つまり,α2-ヘリックス領域を含む拡大したフォールディング核が,多数の相 互作用で安定化され,その核を中心に残りの 2 次構造と 3 次構造が協同的に形成されているとみなせ る.』

<コメント> "フレームワーク モデル"と"核形成-凝縮モデル"については、第10章の「タンパク 質のフォールディング・メカニズムのモデル」を参照のこと)

「「実際のタンパク質分子"のフォールディング過程に、"A-W_NILS モデル"を適用するにはどうすればよいだろうか?

我々が提案している,タンパク質フォールディングの"統計力学的モデル"である「A-W_NILS モ デル」を、"実際のタンパク質"に適応するにはどうすればよいのだろうか?

3次元格子タンパク質と比較して、"実際のタンパク質分子"は、あまりにも複雑な系である.たと えアミノ酸残基を単純な球(ユニット)とみなしても、そのポリペプチド鎖の取り得るコンフォメー ションの数は莫大で天文学的である.この複雑な系の分配関数を簡単なモデルによって近似的に求め るには、どうすればよいだろうか? 「A-W_NILS モデル」を"実際のタンパク質分子"に適用する ために、新たに次のような"定義"を採用しよう:

(1) 2 個のアミノ酸残基(ユニット)の間の"接触"(contact)の定義

2 個のアミノ酸残基(ユニット)を構成している原子の,少なくとも1つの原子ペアの距離が, あらかじめ定めた距離 Dc (接触判断距離, cutoff distance)以内ならば,これらの2 個のアミノ酸 残基(ユニット)は "接触"(contact)しているとみなす. Dc の具体的な値は,次の4 通り; Dc = 4.0, 5.0, 5.5, 6.0 (Å)

(2) 野生型タンパク質のアミノ酸残基間の"接触エネルギー"(contact energy)の値の定義 野生型タンパク質において, *m* 番目から*j* 番目のアミノ酸残基(ユニット)からなる"局所構造 内のエネルギーの値"は,次のようになる:

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

ここで、 $\Gamma_{k,l}$ は、次式で示すように、2つのユニットが"接触"している場合に、 アミノ酸残基間の相互作用が発生する;

 $\Gamma_{kl} = 1$: ユニット k, l が "接触" しているとき

 $\Gamma_{kl} = 0$: ユニット k, lが "接触" していないとき

また、3次元格子タンパク質に倣って、ユニット*k*、*l*が $|k-l| \leq 2$ の場合; $\Gamma_{k,l} = 0$

 $U(\xi_k,\xi_l)$ 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る.

以下の"実際のタンパク質"を扱う問題では、アミノ酸残基のタイプ ξ, と ξ, によらず、

一定の値を ε と仮定する: $U(\xi_k,\xi_l) = \varepsilon$. 具体的には, $\varepsilon = -0.10$ とする (homogeneous contactenergy approximation).
<コメント> 3 次元格子タンパク質の場合は、アミノ酸残基のタイプ ξ_k と ξ_l で決まるエネルギー値

(Miyazawa & Jernigan [1985] による値) を採用している (第5章, Appendix G を参照のこと).

(3) 変異体タンパク質のアミノ酸残基間の"接触エネルギー"(contact energy)の値の定義

野生型のタンパク質から、変異体タンパク質を作成する.これは、野生型タンパク質のアミノ酸 配列の、ある一カ所のアミノ酸残基(ξ_k)を他のアミノ酸残基(ω_k)に置換したもので、この とき、変異体タンパク質のアミノ酸残基間の"接触エネルギー"(contact energy)の値は、次のよ うに変化する: $U(\xi_k,\xi_l) \rightarrow U(\omega_k,\xi_l)$

 $U(\omega_{\iota},\xi_{\iota})$ の値は、次の二つの"接触エネルギー"と取ると仮定する:

 $U(\omega_k, \xi_l) = \varepsilon + 0.01, \quad U(\omega_k, \xi_l) = \varepsilon - 0.01$

変異体タンパク質では、これらの二つの"接触エネルギー"対してΦ値を求める.そして、それ らのΦ値の平均値を、変異体タンパク質のアミノ酸残基(置換したアミノ酸残基)に対するΦ値 とする.

(4) "実際のタンパク質"のコンフォメーションに対するエントロピーの定義

現在のところ,実際のタンパク質のコンフォメーションに対するエントロピー項, *f*(*m*, *j*)の正確な関数形は確立していない. 我々は,3次元格子タンパク質に対して採用した次のような関数形を,そのまま採用しよう(最善の関数形であるとは言い難いが・・・):

$f(m, j) = 1.4084 \times (4.750)^{j-m-2}$

(1)~(4)の定義を採用すると、調節パラメータ (adjustable parameter) を含まない統計力学モデルで ある「A-W_NILS モデル」を、"実際のタンパク質"のフォールディング過程にも適用できる. その 際の手順に従って、分配関数を求め、その分配関数より、フォールディング・レート k_f と、アン フォールディング・レート k_u を計算し、最終的に、 Φ 値を求めることができる.

Coffee Break 一般に、タンパク質の天然構造のコンフォーメション・ネルギーは最小値を とるであろう. "実際のタンパク質"の天然構造において、2個のアミノ酸残基間の"接触"の 情報から、"接触"しているアミノ酸残基ペアを求め、アミノ酸残基のタイプ で決まるエネルギー値(Miyazawa & Jernigan [1985] による値)を採用し て, 天然構造のコンフォーメション・エネルギー値を求めた. その結果は, 天然構造で"接触"しているアミノ酸残基ペアの数は最大にもかかわらず、天 然構造のコンフォーメション・エネルギー値が最小であるという条件を満た さなかった. 新たな試みとして,アミノ酸残基のタイプ によらず,"接触"しているアミ ノ酸残基ペアの接触エネルギー値を、一定の値 & とすることを試みた.そ して,予備的に, -0.05 ~ -0.40 の範囲で幾つかの *ε* の値を決定し, Φ値 計算を実行した. その結果は €の値には、ほぼ依存しないことが明らかになった. 最終的に, $\varepsilon = -0.10$ として, Φ 値を計算することにした.

各アミノ酸残基のΦ値は、タンパク質フォールディングの転移温度において、次のような二つの 量が計算できれば理論的に求められる:

- ・野生型タンパク質の自由エネルギーのプロファイルから求まるフォールディング・レート; $k_f^{wild}(T_m)$
- ・ 変異体タンパク質の自由エネルギーのプロファイルから求まるアンフォールディング・レート; $k_n^{mut}(T_m)$

図は、フォールディング・レート (k_f) とアンフォールディング・レート (k_u) の温度 (の逆数) 依存曲線 (シェブロン・ プロット, Chevron plot) のイラストであ る.

転移温度(T_m)においては、図に示しているように、次式が成立する;

$$\ln k_f^{wild}(T_m) = \ln k_u^{wild}(T_m)$$

故に,

$$\Delta \Delta F_{N-D} = -k_B T_m ((\ln k_f^{mut}(T_m) - \ln k_f^{wild}(T_m)) - (\ln k_u^{mut}(T_m) - \ln k_u^{wild}(T_m)))$$

$$= -k_{B}T_{m}(\ln k_{f}^{mut}(T_{m}) - \ln k_{u}^{mut}(T_{m}))$$

さらに,

$$\Delta \Delta F_{_{_{\pm}-D}} = -k_{B}T_{m}\Delta \ln k_{f}(T_{m})$$

結局,転移温度(Tm)におけるΦ値は次のように表される;

<コメント>: 詳細は, 第7章:『フォールディングのキネティクスにおけるレートと緩和時間の関係は?』, 第8章:『タンパク質フォールディングの遷移状態の構造の推定するΦ値解析とは?』, 及び, Appendix H:「フォールディング・レートの計算方法」参照.



➡ 理論的に求めたプロテインAの遷移状態のΦ値による核形成部位の推定

統計力学的モデルである「A-W_NILS モデル」を用いて、具体的に、"実際のタンパク質"である"プロテインA(1SS1)"に対して、そのアミノ酸残基のΦ値を理論的に計算した結果を示そう.そして、この理論的に求めたΦ値と、実験的に観測されたΦ値とを比較しよう.両方のΦ値の相関を調べることで、プロテインAのフォールディング・メカニズムとして、「フレームワークモデル」が良いか、あるいは、「核形成-凝縮モデル」が適当かを検討しよう.

下図は、プロテインAの理論による結果を、アミノ酸配列上に3色でアミノ酸残基を分類して描いている.更に、二つのアミノ酸残基間の接触判断距離 *D*c =5.5(Å)の場合に、理論によるΦ値を天然構造に重ね合わせて分類した色で描いている.

• : $0.75 \le \Phi \le 1.0$ • : $0.25 \le \Phi \langle 0.75$ • : $0.0 \langle \Phi \langle 0.25 \rangle$



(Dc=5.5Å).

理論的に求めたプロテインA(1SS1)対するアミノ酸残基毎のΦ値から推定される"遷移状態での 構造形成度"を考察しよう:

- (1) "α1-ヘリックスの領域"でのΦ値は"中間的な値"であり,N 末端側に向かって徐々に低い値 になっている.このことは,遷移状態においてはα1-ヘリックスは,"構造形成の途上である"とみ られる.
- (2) "α2-ヘリックス領域"で、"N 末端側のΦ値は高い値"である. このことは、α2-ヘリックスの 領域は、遷移状態においてほぼ"天然構造を形成している"とみなせる.
- (3) "α3-ヘリックスの領域"のΦ値は、α1-ヘリックスの領域でのΦ値と同様に"中間的な値"で

第9章 統計力学モデルによるΦ値計算とフォールディング・メカニズム

あり, C 末端側に向かって徐々に低い値になっている.このことは,遷移状態においてはα3-ヘリ ックスは,"構造形成の途上にある"とみなせる.

- (4) "ターン領域 (T1)" (α1-ヘリックスとα2-ヘリックスを結ぶ領域)のΦ値は高く,遷移状態に おいては、"ほぼ構造形成している"とみなせる.
- (5) "ターン領域 (T2)" ($\alpha 2$ -ヘリックスと $\alpha 3$ -ヘリックスを結ぶ領域)の Φ 値は"中間的な値"で あり,遷移状態においては、"ある程度天然構造を形成している"とみなせる.

◎理論による『プロテインAの遷移状態での核形成部位の推定』(まとめ)

- ・プロテイン A (1SS1)の遷移状態では、"α2-ヘリックス領域"がほぼ天然構造を形成して、この領域が"核形成部位"となり、両側のα1-ヘリックスとα3-ヘリックスに位置する多数の疎水 性側鎖による相互作用によって安定化されて、最終的に天然構造へフォールドするとみなせる。
- ・ 具体的には、α2-ヘリックス領域のアミノ酸残基である、"Arg-28, Asn-29, Gly-30, Ile-32, Gln-33, Ser-34"と、ターン領域(T1)のアミノ酸残基である、"Glu-25, Glu-26"(これらのアミノ酸残基は高いΦ値を示している)が、天然構造形成過程の引き金となる重要な核部位であるとみなせる.

コーティンA(1SS1)に対する、Φ値解析の実験と理論の比較

プロテイン A (1SS1) Φ値に関して,実験の結果と理論的に求めた結果とを比較・検証ししよう. そして,そのフォールディング・メカニズムを考察しよう.

次ページの図では、実験的に Φ 値解析によって調べられたプロテインAのフォールディング遷移状態構造と、それを「A-W_NILS モデル」を用いて理論的に求めた結果の両方を並べて示している.ただし、理論計算では、二つのアミノ酸残基間の接触判断距離 D_c (Cutoff distance)を、 $D_c = 5.5$ Åとしている.

この図を一瞥すると、"理論と実験との結果は、細かい差異は認められるが、ほぼ良好な一致を示している.実際、プロテインAに対して、理論的に計算した Φ 値と、実験的に観測された Φ 値との相関係数 (*CC*; Correlation Coefficient)は、*CC* = 0.71 であった.この CC の高い値は、プロテインAは、我々が採用している統計力学モデルである「A-W_NILS モデル」で仮定したスキームに沿ってフォールドしているとみなせることを示唆している.

理論計算では、タンパク質立体構造データベース(PDB; Protein Data Bank)から得られた天然構造 での天然接触ペアの情報を用いていて、天然接触ペアの相互作用エネルギーの値は、アミノ酸残基の 種類にかかわらず、近似として、ある一定値を割り当てる.つまり、アミノ酸残基ごとの個性をもっ ていることを省略している.また、置換したアミノ酸残基と他のアミノ酸残基間の相互作用エネルギ ーの変化量は、正または負の小さい値を採用している.このことは、理論計算における結果への天然 構造のトポロジーの寄与が多きいこと、つまり、その重要性を示している.しかも、「A-W_NILS モデ ル」という非常に簡単化された統計力学モデルにもかかわらず、プロテインAに対する実験で観測さ れたΦ値がかなり再現されていることは、統計力学モデルである「A-W_NILS モデル」の仮定が妥当 であることを示唆していると見なせるであろう.



佐藤らは、プロテインA(1SS1)のフォールディングのメカニズムが、"核形成-凝縮モデル"に近いと推論した際に、『フレームワーク モデルが該当するためには、α1-ヘリックスとα3-ヘリックスも高度に2次構造を形成しているはずである』と述べているが、我々の「A-W_NILS モデル」を用いた結果が実験結果をほぼ再現していることから、プロテイン Aのフォールディングのメカニズムは、「核形成-凝縮モデル」よりも、「フレームワーク モデル」(「A-W_NILS モデル」)の方が妥当であるとみなしてよいであろう.

□□ キモトロプシンインヒビター2(3CI2)の実験によるΦ値とそのフォールデ

ィング・メカニズム

Fersht らは、"キモトリプシンインヒビター2(3CI2)"(chymotripsin inhibitor)のフォールディング 過程のキネテクスの実験を行った.キモトロプシンインヒビター2では、中間体は観測されず、変性 状態(D状態)から天然状態(N状態)への二状態転移であることを観測した.その後、キモトロプ シンインヒビター2と同程度の大きさをもつタンパク質が集中的に実験され、多くが、同様にフォー ルディング過程では中間体を作らない、二状態転移を示すことが明らかになった.

中間体を形成しないタンパク質ではそのフォールディング過程の途中で何が起こっているのかを調べるために、キモトロプシンインヒビター2のアミノ酸変異体を 100 個以上作り Φ値解析を実行して アミノ酸残基のΦ値を求め、遷移状態の構造が推定された.

下図は、Fersht らによるキモトロプシンインヒビター2の実験による Φ 値の結果(実験条件; "4M GdmCl concentration")を、アミノ酸配列上と天然構造に重ね合わせて、3色で分類して描いている.

<文献> Itzhaki, L.S., Otzen, D.E. & Fersht, A.R. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* **254**, 260-288 (1995).



実験による "3CI2"の Φ 値は, ほとんどの部位で,おおむね, 0.20~0.5 の間に集中している.より 大きな Φ 値は, α 1-ヘリックスのアミノ酸残基にみられる.特に,『16番目のアミノ酸残基, "Ala_16" が大きな Φ 値をもっている』ことがわかる.彼らは,『"Ala_16"が, "フォールディングの核形成部位" である』と提案している.まとめると、 Φ 値が 0.5 以上のアミノ酸残基は、 α 1-ヘリックス中の N 末 端側の数残基と、このヘリックスにドッキングしている β シート中の数残基のみで、他のアミノ酸残 基はおおむね 0.2~0.5 に収まっている.

Φ値をその大きさで分類して、図のように天然構造上に表示した. 特にアミノ酸残基 "Ala_16" に 近接していてΦ値が 0.5 以上の残基である、"Glu_15"、"Lys_18"、"Leu_49" を天然状態の構造に重ね てその構成原子をファンデルワース半径で表示している.

彼らは、このキモトリプシンインヒビター2 (3CI2)の遷移状態では、幾つかのアミノ酸残基が互い に近づいて、しかも、アミノ酸の一次配列上で離れたアミノ酸残基も含んで"フォールディング核" を形成していると推定した、そして、そのフォールディング過程を、次のように描いている:

『キモトロプシンインヒビター2のフォールディングの初期では変性状態(D状態)にあり,何らかの揺らぎ運動を行っている.ある瞬間,特定の数個のアミノ酸残基が衝突して合体し"フォールディング核"が形成され,さらに,その周りのアミノ酸残基が"凝集"を起こして成長し,最終的に天然構造(N構造)を形成する."フォールディング核"は,特定のアミノ酸残基"Ala16"から形成されるが,多くのアミノ酸残基の部位で独立にフォールディングが進むことはない.つまり,遷移状態では,各部位で独立に安定な2次構造が形成されることはない.』

彼らは、キモトロプシンインヒビター2のフォールディング過程の考察より、小さいタンパク質の フォールディングの一般的なメカニズムとして、次のような、"核形成-凝縮モデル" (nucleation-condensation model)を提案した:『小さいタンパク質のフォールディングは、"核形成" (nucleation)と呼ぶ遷移状態に至るまでの過程と、"凝縮"(codensation)というそれ以降の過程から 成る.』

■キモトリプシンインヒビター2(3012)の理論によるΦ値

我々は、キモトリプシンインヒビター2のΦ値を、統計力学モデル「A-W_NILS モデル」を用いて、 理論的に計算した. その結果、キモトリプシンインヒビター2の、理論的に計算されたΦ値と、実験 的に観測されたΦ値との相関係数 (CC; Correlation Coefficient)の値は、CC = 0.45 (ただし、理論: 二つのアミノ酸残基間の接触判断距離 $D_c = 4.0$ Å 、実験条件: 4M GdmCI)である。中間的なケース で、タンパク質のアミノ酸残基のある領域 (α1-ヘリックス)では実験を再現しているが、他の領域 ではほとんど再現していない、結局、キモトロプシンインヒビター2のフォールディング・メカニズ ムは、「A-W_NILS モデル」が基盤とする"フレームワーク モデル"(framework model)」ではなく、 むしろ、"核形成-凝縮モデル" (nucleation-condensation model)と推定できる。



次章で、我々は、27個の"実際のタンパク質"に対して、統計力学モデル"A-W_NILS モデル"を 適用して理論的に計算したΦ値と、実験的に観測されたΦ値と比較した結果をまとめて表(Table 1) に示そう.その際、二つのアミノ酸残基間の接触判断距離(Cutoff distance, D_c)を次のような4通り の場合に対してΦ値を理論的に計算している: $D_c = 4.0, 5.0, 5.5, 6.0$ (Å).

27 個の"実際のタンパク質"に対して,理論的に計算されたΦ値と,実験的に観測されたΦ値との 相関係数 (CC; Correlation Coefficient)の値を求めているが,この CC の値によって, "A-W_NILS モデ ル"で仮定したスキームに沿ってタンパク質がフォールドしているかどうかを推察しよう.



□ 実験における "Φ値解析法"とは?

Fersht らが提案した実験による "Φ値解析"という手法は、どのようにして、フォールディング過程の遷移状態での構造を推定することができるのだろうか?自由エネルギー曲線との関係を考察しよう.

下図は,野生型タンパク質の自由エネルギー曲線(赤色)と変異体タンパク質の自由エネルギー曲線(青色)を模式的に描いている.



秩序度

実験における、"Φ値解析法"は、次の2つの熱力学的実測からなる:

○ 熱力学的実測1;アミノ酸変異によって、"平衡論的な安定性に与える変化量"の実測

○ 熱力学的実測2;アミノ酸変異によって、"遷移状態の自由エネルギーに与える変化量" の実測

それぞれの熱力学的実測の手順を示そう.

◎熱力学的実測1:

<手順1>

遷移状態の構造的特徴をアミノ酸残基レベルで観測するタンパク質を選ぶ(そのタンパク質を "野生型タンパク質"と呼ぶ).

<手順2>

第9章 統計力学モデルによるΦ値計算とフォールディング・メカニズム

熱力学測定から、野性型タンパク質の変性状態(D状態)と天然状態(N状態)との自由エネ

ルギー差 $(\Delta F_{N-D}^{wild} = F_D - F_N^{wild})$ を測定する.

<手順3>

"野生型タンパク質"のアミノ酸配列の特定の部位のアミノ酸残基を選び、その残基をグリシン (Gly) や、アラニン (Ala) などの体積の小さなアミノ酸残基に置き換える.

<コメント> 組み換え DNA 技術を用いてアミノ酸置換したタンパク質を"変異体タンパク質"と呼ぶ. 置換するアミノ酸がグリシンやアラニンである理由については章末の <Coffee Break> を参照のこと.

<手順4>

この"変異体タンパク質"の、変性状態(D 状態)と天然状態(N 状態)との自由エネルギー 差 ($\Delta F_{N-D}^{mut} = F_D - F_N^{mut}$)を測定する.

<手順5>

手順2と手順4の結果から、アミノ酸変異によって"平衡論的な安定性に与えた変化量"、 $\Delta\Delta F_{N-D} = \Delta F_{N-D}^{wild} - \Delta F_{N-D}^{mut}$ を求める.

◎熱力学的実測2;

<手順1>

"野生型タンパク質"のフォールディングの反応速度 (k_f^{wild}) を測定する. この反応速度 k_f^{wild} と、遷移状態の自由エネルギー (F_{\downarrow}^{wild}) と変性状態の自由エネルギー (F_D) の差 $(\Delta F_{\downarrow}^{wild})$ との関係は、次式で表される (Arrhenius's equation).

$$k_{f}^{wild} = k_{0} e^{-\Delta F_{+}^{wild}/k_{B}^{wild}}$$

(ここで, k_0 ;任意の定数, k_B ;Boltzmann 定数,T;絶対温度)

<手順2>

"変異体タンパク質"のフォールディングの反応速度 (k_f^{mut}) を測定する. この 反応速度 k_f^{mut} と, 遷移状態の自由エネルギー (F_{\uparrow}^{mut}) と変性状態の自由エネルギー (F_D) の差 $(\Delta F_{\uparrow}^{mut})$ との関係は, 次式で表される (Arrhenius's equation).

$$k_f^{mut} = k_0 e^{-\Delta F_{+-D}^{mut} / k_B T}$$

(ここで, k_0 ; 任意の定数, k_B ; Boltzmann 定数, T; 絶対温度) <手順3 >

"野生型タンパク質"のフォールディングの反応速度(k_f^{wild})と、"変異体タンパク質"のフォールディングの反応速度(k_f^{mut})から、"遷移状態の自由エネルギーに与えた変化量"、 $\Delta\Delta F_{\ddagger -D}$ はを、次式のように求めることができる:

$$\Delta\Delta F_{+,-D} = \Delta F_{+,-D}^{mut} - \Delta F_{+,-D}^{wild} = k_B T \ln(\frac{k_f^{wild}}{k_f^{mut}})$$

<手順4>

熱力学的実測1から求まる、"遷移状態の自由エネルギーに与えた変化量"、 $\Delta\Delta F_{\ddagger,-D}$ と、熱力学的実測2から求まる、"アミノ酸変異によって平衡論的な安定性に与えた変化量"、 $\Delta\Delta F_{N-D}$ との比

から、 Φ 値次式のように求めることができる:

$$\phi = \frac{\Delta \Delta F_{+}}{\Delta \Delta F_{N-D}}$$

この Φ 値は、タンパク質フォールディングの遷移状態における構造形成度を0から1の数値で示す $ものである (<math>0 \le \phi \le 1$).

Coffee Break Fersht らは、"Φ値解析"において、どのようなアミノ酸置換を考慮し たのだろうか? 彼らは、対象とするタンパク質において、『アミノ酸残基間の特定の相互作用を排 除するようなアミノ酸置換』を行っている. つまり、タンパク質の天然構造を変化させずに、特定のアミノ酸残基間相互作用に のみ影響を与えるようなアミノ酸変異を実行している. このとき、新たなアミノ酸残基間相互作用が生じないように、大きなアミノ酸残基 を小さなアミノ酸残基に置換することになる. D状態においてはアミノ酸置換の影響 は無視できるほど小さいとみなしている. "理想的なアミノ酸置換"として、次のような具体例を考慮している: イソロイシン (Ile) → バリン (Val) アラニン (Ala) \rightarrow グリシン (Gly) トレオリン (Thr) \rightarrow セリン (Ser) これらの置換では、単一のメチル基が削除され、影響を受ける相互作用の数は少ない. 更に、タンパク質内部の疎水性相互作用に影響を与え、天然構造の形成度を反映す る置換として,次のような具体例も採用している:, バリン (Val) → アラニン (Ala), イソロイシン (Ile) \rightarrow アラニン (Ala), ロイシン (Leu) \rightarrow アラニン (Ala)



"Autumn Morning"

理論と実験による中値の相関の検証_その1

く要旨>

第10章

27個の"実際のタンパク質"に対して,統計力学モデル"A-W_NILS モデ ル"を適用して,4種類のアミノ酸残基間の接触距離 D_e(Å)に対して理論的に 計算したΦ値と,実験的に観測されたΦ値との相関係数 (CC)の値を表で示し ている.そして,その相関係数の値から,27個のタンパク質を3つのグループ に分類し,それぞれのグループのタンパク質のフォールディング・メカニズ ムが,"フレームワーク モデル",あるいは,"核形成一凝縮モデル" のどちらかを検討している.ここで,これらの,タンパク質のフォールディ ング・メカニズムのモデルと" A-W NILS モデル"の要点をまとめている.

更に、27個の"実際のタンパク質"の中から、4個のSH3ドメインファミ リーのタンパク質、1SRM、1SHG、1FYN、1BF4、及び、天然構造が類似の 1PGB と2PTL を取り上げ、Φ値に対する、実験と理論の結果を詳細に比較・ 検討している.

その結果,たとえアミノ酸配列の類似性が小さくとも,天然構造が類似していれば,フォールディング・メカニズムは,天然構造のトポロジーによって決定されているとみなせるだろうと推定している.しかしながら,タンパク質1BF4のΦ値は,3個のSH3ドメインファミリータンパク質とは異なる結果が得られた.フォールディング・メカニズムは,立体構造のトポロジーが本質的だが,アミノ酸配列の情報も重要であると指摘している.

天然構造が類似の1PGBと2PTLのΦ値の分布は明らかに相違していて、この場合も、フォールディング・メカニズムは個々のタンパク質のアミノ酸配列に依存している例であると指摘している.

■ 27 個の"実際のタンパク質"に対して理論的に求めたΦ値と実験的に

求められたΦ値との相関

27 個の"実際のタンパク質"に対して,統計力学モデル"A-W_NILS モデル"を適用して理論的に 計算した Φ 値と,実験的に観測された Φ 値と比較しよう.結果は"Table 1"に示している.

その際、二つのアミノ酸残基間の接触判断距離(Cutoff distance, D_c)を次のような4種類の場合に 対して計算した: $D_c = 4.0, 5.0, 5.5, 6.0$ (Å). ほとんどのタンパク質では、 D_c の依存性はみられないが、 いくつかのタンパク質では D_c の値によっては、理論的に計算した Φ 値と、実験的に観測された Φ 値と の相関係数 (Correlation Coefficient, CC)の値に変化がみられる.

相関係数(CC)の値の大小は、『フォールディング・メカニズムにおけるバリエーションを示している』と考えられる.相関係数(CC)の値によって、タンパク質分子のフォールディング・メカニズムの特徴付け、つまり、"フレームワーク モデル"か、"核形成一凝縮モデル"かを考察しよう.

Table 1 の相関係数 (CC)の値)から,次のように3つのグループに分類し,そのグループに該当するタンパク質の"フォールディング・メカニズム"を次のように推定した:

(i) CC≧0.6 **の場合・・・**『"フレームワーク モデル"が妥当であろう!』

理論的に計算された Φ 値は、実験で観測された Φ 値をよく再現している. "A-W_NILS モデル" の仮定が妥当であるとみなせる.

次のタンパク質が該当する(9個):

1:1ENH, 2:1IDY, 4:1SRM, 5:1SHG, 7:1SS1, 17:1TEN, 19:1URN, 26:1AZU, 27:3CHY

(ii) 0.3≦CC<0.6 の場合・・・『"フレームワーク モデル"か, "核形成-凝縮モデル"かの,

どちらでも説明可能 !』

中間的なケースで、タンパク質のアミノ酸残基のある領域では実験を再現しているが、他の領域 ではほとんど再現していない.

次のタンパク質が該当する(11個):

3:1PGB, 6:1FYN, 8:1BF4, 9:3CI2, 11:1CSP, 15:1TIU, 18:1TTF, 20:1N88, 22:2ACY, 24:1RNB, 25:2VIL

(iii) CC<0.3 の場合・・・『核形成-凝縮モデル" が妥当であろう! 』

理論的に計算された Φ 値は、実験で観測された Φ 値と一致していない. "A-W_NILS モデル"の 仮定そのものが破綻しているとみなせる.

次のタンパク質が該当する(7個);

- 10:2PTL, 12:1UBQ, 13:1AYE, 14:2ABD, 16:1BTB, 21:1RIS, 23:1FKB
- <文献> H. Wako and H. Abe, Calculation of Free-Energy Profiles, Folding Rates and Φ Values by Means of a Simple Statistical-Mechanical Model of Protein Folding, *Advances in Protein Folding Research*, M. Hale ed., Nova Sci. Pub. Inc., pp. 19-63, 2016.
- <文献> H. Wako and H. Abe, Calculation of protein folding by a Φ-value calculation with a statistical-mechanical model, *Biophysics and Physicobiology*, Vol. 13, pp. 263-279, 2016.

			I able .	I. 天	愛てい	$\nabla \Psi_1$	風と埋	「ヨー」	rOA	個との相関
番	タンパク質	残基	PDBの	アミノ酸	天然構造	相関係数(CC)				タンパク質名
号	(PDB code)	数	残基番号	置換数	の型	4.0Å	5.0Å	5.5Å	6.0Å	
1	1ENH	54	3-56	13	All α	0.28	0.43	0.62	0.59	Engrailed homeodomain
2	1IDY	54	140-193	18	All α	0.69	0.62	0.65	0.58	cMyb-transforming protein
3	1PGB	56	1-56	25	α+β	0.36	0.31	0.09	-0.02	B1 IgG-binding domain of protein G
4	1SRM	56	9-64	34	All β	0.63	0.62	0.64	0.65	Src SH3 domain
5	1SHG	57	6-62	14	All β	0.23	0.75	0.75	0.74	α-spectrin SH3 domain
6	1FYN	59	84-142	9	All β	0.35	0.42	0.34	0.39	Fyn SH3 domain
7	1SS1	62	-1-60	31	All α	0.28	0.33	0.71	0.63	B domain of protein A
8	1BF4	63	2-64	21	All β	0.31	0.36	0.35	0.32	DNA binding protein Sso7d
9	3CI2	64	20-83	40	α+β	0.45	0.37	0.3	0.19	chymotrypsin inhibitor 2
10	2PTL	64	1-64	46	α+β	0.12	0.25	0.28	-0.36	B1 IgG-binding domain of protein L
11	1CSP	67	1-67	20	All β	0.49	0.58	0.54	0.5	Cold-shock protein
12	1UBQ	76	1-76	20	α+β	-0.03	0.03	0.01	-0.01	Ubiquitin
13	1AYE	78	4A-83A ^{注1)}	18	α+β	-0.22	-0.22	-0.21	-0.25	Procarboxypeptidase A2 active domain
14	2ABD	86	1-86	16	All α	-0.76	-0.73	-0.75	-0.74	Acyl-coenzyme A binding protein
15	1TIU	89	1-89	26	All β	0.46	0.45	0.51	0.44	Titin I27 domain
16	1BTB	89	1-89	28	α/β	-0.09	-0.18	-0.10	-0.18	Barstar
17	1TEN	90	802-891	26	All β	0.61	0.68	0.65	0.66	TNfn3 domain (tenascin)
18	1TTF	94	1-94	20	All β	0.46	-0.21	-0.10	-0.02	FNfn10 domain (fibronectin)
19	1URN	96	2-97	10	α+β	0.89	0.94	0.94	0.94	U1A
20	1N88	96	1-96	16	α+β	0.49	0.59	0.59	0.47	Ribosomal protein L23
21	1RIS	97	1-97	20	α+β	0.17	0.1	0.03	0.03	Ribosomal protein S6
22	2ACY	98	1-98	22	α+β	0.37	0.45	0.42	0.42	Acylphosphatase
23	1FKB	107	1-107	22	α+β	-0.47	-0.40	-0.50	-0.47	FKBP12
24	1RNB	109	2-110	28	α+β	0.39	0.42	0.4	-0.21	Barnase
25	2VIL	126	1-126	24	α+β	0.46	0.55	0.47	0.45	Villin 14T
26	1AZU	126	3-128	17	All β	0.65	0.65	0.64	0.59	Azurin
27	3CHY	128	2-129	19	α/β	0.81	0.15	0.77	0.83	CheY

Table 1. 実験によるΦ値と理論によるΦ値との相関

注1)4A-34A,34B,34C,35A-42A,47A-83A

🎩 タンパク質のフォールディング・メカニズムのモデル

我々は、タンパク質フォールディングに対する統計力学モデルを用いてΦ値を理論的に求め、実験 によるΦ値を比較・検討することによって、フォールディング・メカニズムを推定することに取り組 んでいる.

タンパク質のフォールディング・メカニズムのモデルに関して、主に、次のような二つモデルが提案されている. 「フレームワーク モデル」(the framework model) と、「核形成-凝縮モデル」(the nucleation-condensation model) である.

◎ 『フレームワーク モデル』 (the framework model)

「フレームワーク モデル」のフォールディング・メカニズムの要点は次の通りである:

『まず短距離相互作用が主要である2次構造が形成され,更に,中・長距離相互作用により,それらの2次構造同士が合体して,最終的に天然構造へとフォールドする.つまり,短距離相互作用→中距離相互作用→長距離相互作用の順に階層的に作用するとみなす.』

我々の「A-W_NILS モデル」は、この「フレームワーク モデル」に属しているが次の点が異なっ ている:「フレームワーク モデル」は、"フォールディングには決まった道筋があり、αヘリックスや βシートなどの2次構造やドメインなどの特別な局所構造が階層的に形成される"という仮定を前提 にしている.「A-W_NILS モデル」の方は、もっと"確率的"なモデルで、αヘリックスやβシート などの2次構造(局所構造)が、それらの統計重率によって、任意の順序で形成することができると みなすモデルである.両方のモデルの重要な共通点は、『長距離相互作用が、短距離相互作用に前に働 く事はない』と仮定している点である.このモデルは、特に、ヘリックス・タンパク質(アミノ酸残 基間の短距離相互作用が主要とみなせるヘリックスが数個から構成されている)のフォールディング 過程を説明するのに魅力的である.

ここで,統計力学モデルである "A-W_NILS モデル"を "実際のタンパク質"に適用する際に, "次のような仮定"をしていることを確認しておこう:

- ・「"アミノ酸残基間の相互作用は、アミノ酸残基のタイプを考慮しないで、一定の値が割り振ら れている.」
- ・「タンパク質のコンフォメーションに対するエントロピー項を,3次元格子タンパク質に対して 採用した関数形をそのまま採用する.」

これらの仮定にもかかわらず、多くのタンパク質に対する実験的に観測されたΦ値がかなり再現されているが、しかしながら、再現されなかったタンパク質もある(特に、(iii) CC<0.3 の場合).理論と実験のΦ値が明らかに不一致の場合は、統計力学的モデル「A-W_NILS モデル」が妥当ではないと思われる.

"フレームワーク モデル"の場合のΦ値の分布はどのようになるのだろうか?

Fersht らは、この「フレームワーク モデル」モデルと「 Φ 値の分布」との関係は、次のようである と推論している:『天然構造で α ヘリックスや β シートなどの2次構造を形成している領域に含まれる 多くのアミノ酸残基の Φ 値が 1.0 に近い値を示す一方、中間的な Φ 値をとるアミノ酸残基が多数分布 している.』

◎『核形成-凝縮モデル』(the nucleation-condensation model)

"核形成-凝縮モデル"のフォールディング・メカニズムの要点は、次の通りである:

『アミノ酸配列上のある領域で,短距離相互作用が主要である2次構造と,長距離相互作用による3次構造が,いくつかの疎水性アミノ酸残基を核として,2次構造と3次構造が同時に形成されるか,あるいは,先に長距離相互作用により大まかな3次構造が形成され,その後,短距離相互作用が主要である2次構造が形成されて細部の微調整が行われ,最終的に天然構造へとフォールド(凝縮)する

というモデルである.』

より大きな、いくつかのドメインからなるタンパク質凝縮の場合は、複数の核形成が存在し、それ ぞれで、核形成一凝縮によるフォールディングが起こるとみなす.この"核形成-凝縮モデル"の重要 な点は、『短距離相互作用が作用する前か、あるいは、ほぼ同時に、長距離相互作用が働く』と仮定し ている点である.理論と実験によるΦ値の相関係数が、CC<0.3 の場合に相当するタンパク質はこの モデルが該当していると推定される.

I SH3 ドメインファミリーである 3 個のタンパク質(1SRM, 1SHG, 1FYN)のΦ 値の理論と実験の比較

Table 1 の相関係数 (CC)の値)から、タンパク質を3つのグループに分類しているが、その中から、"「フレームワーク モデル」framework model)が妥当であろうと考えられるタンパク質"として、SH3 (Src*-Homolgy 3) ドメインファミリーの3 個のタンパク質、(1) "1SRM" (表の番号:4)、(2) "1SHG" (表の番号:5)、(3) "1FYN" (表の番号:6)を取り上げ、Φ値に対する実験と理論の結果を比較しよう.

*Src (サーク); sarcoma (サルコーマ, 肉腫)の短縮形

これらの SH3 ドメインファミリーのタンパク質は、天然構造が類似しているが、はたしてフォール ディング・メカニズムも同じであろうか? 統計力学モデル "A-W_NILS モデル"では、アミノ酸配 列の相違を考慮していないので、理論によるΦ値はほぼ同じであろうと予想されるが、実験によるΦ 値の方はどうであろうか? 実験の方も同じようなΦ値の結果ならば、フォールディング・メカニズ ムは、"天然構造のトポロジーによって決定されている"とみなせるだろう.このことを検証しよう.

"SH3 ドメイン"の特徴は次の通りである:

・アミノ酸残基数が約60個からなるドメインである.

・約300種類が存在している.

・立体構造は、5本、ないしは6本の β ストランドからなり、そのうち2本が反平行の β シートを形成している.

・細胞内シグナル伝達に関与するタンパク質に広く見いだされている.

・ 次のような機能をもっている:『タンパク質のプロリンに富む領域と特異的に結合して,タンパク 質とタンパク質との間の相互作用を制御する.』



"Church at Dawn"

第10章 理論と実験によるΦ値の相関の検証_その1

- (1) "1SRM" (Scr SH3 domain, 表の番号: 4) の Φ 値
- ・二つのアミノ酸残基間の接触判断距離 $D_c = 6.0$ (Å)のとき,相関係数は,CC = 0.65である.CCが 0.6 より大であるので、タンパク質 "ISRM"のフォールディング・メカニズムは "フレームワークモデル"が妥当であると推定される.理論的に計算されたΦ値は、実験で観測されたΦ値をよく再現している.
- ・実験による Φ 値の分布から,遷移状態での"フォールディング核"は、『 β 3 と β 4 を結ぶ"the distal β -hairpin"の領域』であると推定される.特に、大きな Φ 値をもつアミノ酸残基は次の通りである: Val-27, Ala-37, Ser-39, Thr-42, Gly-43

下図は、"ISRM"の実験による Φ 値と、理論による Φ 値を比較したものである。アミノ酸配列上と天然構造に重ね合わせて、3 色で分類して描いている。

<文献> Riddle, D.S., Grantcharova, V. P., Santiago, J.V., Alm, E., Ruczinski, I. & Baker, D. Experiment and Theory Highlight Role of Native State Topology in SH3 Folding. *Nat. Struct. Biol.***6**, 1016-1024, 1999.



(2) "1SHG" (α-spectrin SH3 domain, 表の番号:5) のΦ値

・二つのアミノ酸残基間の接触判断距離 $D_c = 5.0$ (Å) のとき,相関係数は, CC = 0.75 である. CC が 0.6 より大であるので,タンパク質 "1SHG"のフォールディング・メカニズムは "フレームワーク モデル"が妥当であると推定される.理論的に計算されたΦ値は,実験で観測されたΦ値をよく再現している.

・遷移状態での"フォールディング核"は、『 β 3 と β 4を結ぶ"the distal β -hairpin"の領域』である と推定される。特に、大きな Φ 値をもつアミノ酸残基は次の通りである: Val-41

下図は、"1SHG"の実験による Φ 値と、理論による Φ 値を比較したものである。アミノ酸配列上と天然構造に重ね合わせて、3 色で分類して描いている。

<文献> Martinez, J.C. & Serrano. L. The Folding Transition State between SH3 Domains is Conformationally Restricted and Evolutionarily Conserved. *Nat. Struct. Biol.* **6**, 1010-1016, 1999.



(3) "1FYN" (Fyn SH3 domain, 表の番号: 6) の Φ 値

- ・二つのアミノ酸残基間の接触判断距離 $D_c = 5.0$ (Å)のとき,相関係数は, CC = 0.42である. CCの値が 0.3~0.6の範囲であるが,実験データな少ないからと考えられ, 1SRM, 1SHG と同様に, "フレームワークモデル"が妥当であると推定される.
- ・ 遷移状態での"フォールディング核"は、次の大きなΦ値をもつアミノ酸残基であると推定され、 これらは、コアを形成している疎水性残基である:『Ile-28、 Ala-39、 Ile-50 』

下図は、"1FYN"の実験による Φ 値と、理論による Φ 値を比較したものである。アミノ酸配列上と天然構造に重ね合わせて、3 色で分類して描いている。

<文献> Northey, J.G.B., Di Nardo, A.A. & Davidson, A.R. Hydrophobic Core Packing in the SH3 Domain Folding Transition State. *Nat. Struct. Biol.* **9**, 126-130 (2002).



↓ タンパク質のフォールディング・メカニズムは、天然構造のトポロジーで 決定されている?・・・"SH3 ドメインファミリー"の例

SH3 ドメインファミリーである,3 個のタンパク質(1SRM, 1SHG, 1FYN)のΦ値の実験と理論の結果を下図に並べて表示している.アミノ酸配列上と天然構造に重ね合わせて,3 色で分類して描いている.

これらは、細かい差異はあるももの、ほぼ類似していることから、これらの3個のファミリーのフ オールディング・メカニズムは、ほぼ同じで、"フレームワーク モデル"が妥当であろうと推定され る.結局、たとえアミノ酸配列の類似性が小さくとも、天然構造が類似していれば、同じフォールデ ィング・メカニズムであると推定される.つまり、"フォールディング・メカニズムは、天然構造のト ポロジーによって決定されている"とみなせるだろう.

• ; $0.75 \le \Phi \le 1.0$ • ; $0.25 \le \Phi \langle 0.75$ • ; $0.0 \langle \Phi \langle 0.25 \rangle$

Src SH3 Domain (1SRM)



 α -spectrin SH3 domain (1SHG)



Fyn SH3 domain (1FYN)



₍₈₄₎ β1





β^{*}5 (142)

B4

α1

"Trees beside River, Autumn"

■ タンパク質"1BF4"のΦ値の理論と実験の比較

タンパク質 "1BF4" (DNA binding protein Sso7d, 表の番号:8) は、3 個のタンパク質 (1SRM, 1SHG, 1FYN) と同様に SH3 ドメインファミリーで、2 次構造は、『 β 1- β 2- α 1- β 3- β 4- α 2- β 5- α 3 』であ る. 二つのアミノ酸残基間の接触判断距離 $D_c = 5.0$ (Å) のとき、相関係数は CC = 0.36 で、CC は 3 個のタンパク質より小さい値である.理論と実験のΦ値の顕著な相違は、『 β 4- α 2- β 5』の領域であ る. 理論でのこの領域のΦ値は、3 個のタンパク質とほぼ類似して高い値(遷移状態で核形成をして いると推定される領域)であるが、タンパク質 "1BF4"の実験のΦ値は低い値になっている. タンパ ク質 1BF4 の実験のΦ値は、『 β 3 領域と α 3 領域』で高い値をもっていて、この領域が遷移状態では、 核形成していると推定される.

このことから、タンパク質 "1BF4"のフォールディング・メカニズムは "核形成一凝縮モデル"と 推定される. つまり、3 個のタンパク質 SH3 ドメインファミリーと天然構造が類似しているが、フォ ールディング・メカニズムは異なっていると推定される. このことは、フォールディング・メカニズ ムは、立体構造のトポロジーが本質的だが、アミノ酸配列の情報も重要であることを示している. 下図は、"1BF4"の実験によるΦ値と、理論によるΦ値を比較したものである. アミノ酸配列上と天 然構造に重ね合わせて、3色で分類して描いている.

<文献> Guerois, R. & Serrano, L. The SH3-fold family: Experimental Evidence and Prediction of Variations in the Folding Pathways. J. Mol. Biol. **304**, 967-982 (2000).



1BF4 (DNA binding protein Sso7d)の中値(理論と実験)

(Dc=5.0 Å)

α1

<u>β</u>1

B2

a2

α1

↓ 天然構造が類似していても、タンパク質のフォールディング・メカニズム は異なる?・・・"Protein G"と "Protein L"の例

天然構造が類似しているが,フォールディング・メカニズムは異なっていると推定される,次の二つのタンパク質, "1PGB", "2PTL"を取り上げよう:

O"1PGB"(B1 IgG-binding domain of protein G, 表の番号:3)

- アミノ酸残個数・・・56 個
- フォールディング・タイプ・・・α+β
- 2次構造・・・β1-β2-α1-β3-β4

<文献> McCallister, E.L., Alm, E. & Baker, D. Critical Role of β -Hairpin Formation in Protein G Folding.

Nat. Struct. Biol. 7, 669-673 (2000).

○ "2PTL"(B1 IgG-binding domain of protein L, 表の番号:10)

- ・ アミノ酸残個数・・・64 個 (PDB のデータ;15-78)
- 2次構造・・・β1-β2-α1-β3-β4
- <文献> Kim, D. E., Fisher, C. & Baker, D. A Breakdown of Symmetry in the Folding Transition State of Protein L. *J. Mol. Biol.* **298**, 971-984 (2000).

これら二つのタンパク質, "1PGB" (protein G), "2PTL" (protein L) は, 天然構造が類似している ので, フォールディング・メカニズムも同じであろうか? 統計力学モデル "A-W_NILS モデル"で は, アミノ酸配列の相違を考慮していないので, 理論によるΦ値は, ほぼ同じであろうと予想される が, 実験によるΦ値の方はどうであろうか?

<実験によるΦ値>

二つのタンパク質, "1PGB" (protein G)と "2PTL" (protein L) のΦ値による遷移状態の構造推定で, 次のような決定的な相違がある:

- 〇 "1PGB" (protein G)・・・ β 3 と β 4 のつなぎの領域であるターン領域 (Turn 2) での Φ 値が大き く, Turn 2 が, 遷移状態での, フォールディング核であると推定される.
- "2PTL" (protein L)・・・β1とβ2のつなぎの領域であるターン領域(Turn 1) でのΦ値が大きく, Turn 1 が, 遷移状態での, フォールディング核であると推定される.

<理論によるΦ値>

統計力学モデル "A-W_NILS モデル"では、アミノ酸配列の相違を考慮していないので、天然構造 が類似の、二つのタンパク質、"1PGB" (protein G)と "2PTL" (protein L) で、ほぼ同じ結果が得ら れた:

- α1-ヘリックスは形成されている.
- β2ストランドは形成されていて、β3-ストランドも、ほぼ形成されている。
- 〇 二つのターン領域 ($\beta_1 \ge \beta_2$ のつなぎの領域, $\beta_3 \ge \beta_4$ のつなぎの領域) は、形成されていな くて、変性状態である.
- "フォールディング核"は、β2-ストランドとα1-ヘリックスの領域と推定される.

結局, "1PGB" (protein G) と "2PTL" (protein L) の実験による Φ 値の結果は,明らかに相違している. これらの二つのタンパク質は,天然構造は類似しているが,"フォールディング・メカニズムは異なっている (それぞれのタンパク質のアミノ酸配列に依存している)例である.

次ページの図は、"1PGB"、及び、"2PTL"の実験による Φ 値と、理論による Φ 値を比較したものである。アミノ酸配列上と天然構造に重ね合わせて、3色で分類して描いている。



1PGB (Protein G) の Φ 値 (理論と実験)





第11章 理論と実験による
 の相関の検証_その2



第11章 理論と実験によるΦ値の相関の検証_その2

➡ 天然構造が類似している二つのタンパク質,"1ENH"と"1IDY"の フォールディング・メカニズムは?

天然構造が類似している2個のタンパク質, "1ENH"と"1IDY"について, 理論と実験のΦ値の値 を検証しよう.

- O "1ENH" (Engrailed homeodomain, 表の番号:1)
 - ・ アミノ酸残個数・・・54 個 (PDB のアミノ酸残基番号;3-56)
 - ・ フォールディング・タイプ・・・All α
 - 2次構造・・・α1-α2-α3
 - 相関係数 (CC) の値; CC = 0.62 (ただし、二つのアミノ酸残基間の接触判断距離
 D_c = 5.5 Å ・・・理論と実験の相関が良い)
- O "1IDY" (c-Myb, 表の番号:2)
 - ・ アミノ酸残個数・・・54 個 (PDB のアミノ酸残基番号;140-193)
 - ・ フォールディング・タイプ・・・All α
 - 2次構造・・・α1-α2-α3
 - ・ 相関係数 (CC) の値; CC = 0.69 (ただし, 二つのアミノ酸残基間の接触判断距離 $D_c = 4.0$ Å・・・ (理論と実験の相関が良い)
 - <文献> Gianni, S., Guydosh, N.R., Khan, F., Caldas, T.D., Mayor, U., White, G.W.N, *et al.* Unifying features in protein-folding mechanisms. *Proc. Natl Acad. Sci. USA* **100**, 13286-13291 (2003).

これら二つのタンパク質"1ENH, 1IDY"は、天然構造が類似しているので、フォールディング・ メカニズムも同じであろうか?

<実験によるΦ値>

"1ENH"の方は、遷移状態で、α1ヘリックス領域と、ターン領域(α1とα2を繋いでいる領域) がフォールディング核を形成していると推定される.一方、"1IDY"の方は、遷移状態で、2次構造 (3 個のヘリックス)の形成が明確ではなく、形成途中の段階のように推定される. Gianni らは論文 の中で、"1ENH"のフォールディング・メカニズムは、「フレームワークモデル」で、"1IDY"の方 は、「フレームワークモデル」と「核形成一凝縮モデル」の混合型であると推定している.つまり、 "1IDY"の方は、2次構造と3次構造が、ほぼ同時に形成されると推定している.

<理論によるΦ値>

天然構造が類似している2個のタンパク質, "1ENH"と"1IDY"の理論による Φ 値の分布は,アミノ酸配列の相違を考慮していないので,ほぼ同じであり,遷移状態で, α 2 ヘリックス近傍の領域がフォールディング核であると推定される.

実験によるΦ値との相関は、両方のタンパクク質とも相関係数 (CC)が 0.6 以上であり、フォールディング・メカニズムは、「フレームワーク モデル」に該当していると推定される.

次のページの図は、"1ENH",及び、"1IDY"の実験によるΦ値と、理論によるΦ値を比較したもので ある、アミノ酸配列上と天然構造に重ね合わせて、3色で分類して描いている.



En-HD(Engrailed HomeoDomain,1ENH)の中値(理論と実験)

• ; $0.75 \le \Phi \le 1.0$ • ; $0.25 \le \Phi \langle 0.75$ • ; $0.0 \langle \Phi \langle 0.25 \rangle$





(*D*_c=5.5 Å)



C-Myb (c-Myb-transforming protein ,1IDY)の中値(理論と実験)

ロタンパク質 "3CHY" (CheY) の Φ 値の, 理論と実験の比較

タンパク質 "3CHY"(CheY, 表の番号: 27)の天然構造の特徴とΦ値による遷移状態での構造の 推定しよう.

・アミノ酸残基:128 個,フォールディング・タイプ; α/β

- ・次のような2個のサブドメインからなる:
 - N-末端側のサブドメイン;(β1-α1-β2-α2-β3)
 - C-末端側のサブドメイン;(β3-α3-β4-α4-β5-α5)
 - ・・・(β3シートは,2個のサブドメインに共通)
- ・理論で、二つのアミノ酸残基間の接触判断距離 $D_c = 6.0$ (Å)のとき、実験との相関係数はCC = 0.83で、非常に高いので、フォールディング・メカニズムは、「フレームワークモデル」に該当していると推定できる。
- ・実験で大きな Φ 値をもつアミノ酸残基はN-末端側のサブドメイン領域である.従って,遷移状態 での"フォールディング核"は、次の領域であると推定される:『N-末端側のサブドメインの領域 " β 1- α 1- β 2"』
- ・遷移状態では、C-末端側のサブドメインは形成されていない.

下図は、"3CHY"の実験による Φ 値と、理論による Φ 値を比較したものである。アミノ酸配列上と天然構造に重ね合わせて、3色で分類して描いている。

<文献> Lopez-Hernandez, E. & Serrano, L. Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. *Fold. Des.* **1**, 43-55 (1996).



第11章 理論と実験によるΦ値の相関の検証_その2

👪 タンパク質"1RNB"のΦ値の, 理論と実験の比較

タンパク質 "1RNB" のΦ値の理論と実験の相関係数 (CC) の値は, CC = 0.42 (ただし, $D_c = 5.0$ Å) で、 $0.3 \leq CC < 0.6$ の場合に相当し、"フレームワーク モデル"か、"核形成—凝縮モデル"かの、どちらでも説明可能である.

タンパク質 "1RNB" (barnase, 表の番号:24)の特徴:

- ・ アミノ酸残個数・・・109 個 ・ フォールディング・タイプ・・・ $\alpha + \beta$
- ・ 2 次構造・・・『 α1-β1-α2-α3-β2-β3-β4-β5-β6 』
- ・ 相関係数 (CC) の値; CC = 0.42
- (ただし、二つのアミノ酸残基間の接触判断距離*D_c* = 5.0 Å)
 <実験でのΦ値>
 - ・ 遷移状態での Φ 値は $\beta 2$ シートと $\alpha 1$ ヘリックスのC末端領域で 1.0 近傍の値をとり,
 - この領域の2次構造が形成されていると推定できる.→『<mark>核形成—凝縮モデル</mark>』と推定.

<理論でのΦ値>

・遷移状態での Φ 値は、 β 3シートと β 4シートの間のターン領域と、 β 4シート~ β 5シートの 領域で 1.0 に近い値をもち、この領域がフォールディング核とみなせる.

下図は、"1RNB"の実験による Φ 値と、理論による Φ 値を比較したものである。アミノ酸配列上と 天然構造に重ね合わせて、3色で分類して描いている。

<文献> Serrano, L., Matouschek, A. & Fersht, A.R. The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J. Mol. Biol.* **224**, 805-818 (1992).



1RNB (barnase)の Φ 値 (理論と実験)

■ タンパク質"1FKB"のΦ値の、理論と実験の比較

タンパク質 "1FKB" (EFKBP12, 表の番号;23) は,理論的に計算された Φ 値と,実験で観測された Φ 値との相関が低く(相関係数:CC<0.3), "A-W_NILS モデル"の仮定そのものが破綻していて, 我々の統計力学モデルが適用できない典型的な例である.その主な理由は,長距離相互作用が要求さ れる多くの β シート構造をもつタンパク質であることが考えられる.

"1FKB"の特徴:

- アミノ酸残個数・・・107 個
- ・ フォールディング・タイプ・・・ $\alpha + \beta$
- 2 次構造・・・『 β1-β2-β3-β4-α1-β5-α2-β6 』
- ・ 相関係数 (CC) の値; CC = -0.40
- (ただし,二つのアミノ酸残基間の接触判断距離*D_c*=5.0 Å)

<実験でのΦ値>

・遷移状態でのΦ値は 0.4 近傍で、2次構造は完全には形成されていない。

→ 『核形成—凝縮モデル』(the nucleation-condensation model) とみなせる.

<理論での0値>

・遷移状態での Φ 値は、 β シートと α 1ヘリックス近傍で 1.0 に近い値をもちフォールディング核 とみなせる.

<特記>

タンパク質 "1FKB"で、理論的な Φ 値を求めるために、相互作用を強めて($U = \varepsilon + 0.01$)計算するとき、あるアミノ酸残基の部位で Φ 値の異常値($\Phi > 1.0$ や $\Phi < 0.0$)が発生することがある.

実験でも、置換するアミノ酸基の部位や、置換するアミノ酸残基の種類によってΦ値の異常値が出ることがあり、置換したアミノ酸残基が何らかの影響を与えていると説明されるが、理論計算では、 アミノ酸残基を置換しても、元の立体構造は変化しないと仮定しているので、このような説明は成立しない.理論的にΦ値を求める手法そのものに、不安定化させる要因が内在しているかもしれない.

<文献> Fulton, K.F., Main, E.R.G., Daggett, V. & Jackson, S.E. Mapping the interactions present in the transition state for unfolding/folding of FKBP12. *J. Mol. Biol.* **291**, 445-461 (1999).

次のページの図は、"1FKB"の実験による Φ 値と、理論による Φ 値を比較したものである、アミノ酸 配列上と天然構造に重ね合わせて3色で分類して描いている.



"Summer Evening beside Lake"



1FKB (FKBP12) の Φ 値 (理論と実験)

Residue number



C末端 β6 β2 実験 β4 β1 N末端 α2

• ; $0.75 \le \Phi \le 1.0$

; 0.25 ≤ Φ(0.75

▲ ; 0.0⟨Φ⟨0.25

(Dc=5.0 Å)

■ タンパク質 "1RIS"のΦ値に対する、実験と理論の比較

タンパク質"1RIS"(ribosomal protein S6,表の番号;21)は、理論的に計算された Φ 値と、実験で 観測された Φ 値との相関が低く(相関係数:CC<0.3)、"A-W_NILS モデル"の仮定そのものが破 綻していて、我々の統計力学モデルが適用できない典型的な例である。その主な理由は、長距離相互 作用が要求される多くの β シート構造をもつタンパク質であることが考えられる。

タンパク質"1RIS"の特徴:

- アミノ酸残個数・・・97 個
- フォールディング・タイプ・・・ $\alpha + \beta$
- 2次構造・・・『β1-α1-β2-β3-α2-β4』
- ・ 相関係数 (CC) の値; CC = 0.17
- (ただし,二つのアミノ酸残基間の接触判断距離 D_c = 4.0 Å)

<実験での0値>

- ・ 実験による Φ値から 遷移状態では、両末端部分(N末端とC末端)の構造が形成されている と推定される. アミノ酸配列の中央部は構造形成されていない.
- ・ "フォールディング核は、(β 1- α 1- β 3)からなる " α 1-nucleus" と (β 1- α 2- β 4)からなる " α 2-nucleus"であると推定される.
- ・ これらの二つのフォールディング核が競合してフォールドすると推定される.

<文献> Otzen, D.E. & Oliveberg, M. Conformational Plasticity in Folding of the Split β - α - β Protein S6: Evidence for Burst-Phase Disruption of the Native State. *J. Mol. Biol.* **317**, 613-627, 2002.

<理論での0値>

- ・ 理論での Φ 値は、 α 1-領域と β 1-領域で大きいので、これらの部位がフォールディング核と 推定される.
- 「A-W_NILS モデル」を用いて理論的に計算されたΦ値は、実験的に観測されたΦ値と ほとんど一致しない結果が得られた.フォールディングに対する"統計力学モデルの仮定"そ のものが破綻している、つまり、1RISのフォールディング・メカニズムは"核形成→凝縮モデ ル"に該当すると推定される.

次のページの図は、"1RIS"の実験によるΦ値と、理論によるΦ値を比較したものである. アミノ 酸配列上と天然構造に重ね合わせて3色で分類して描いている.



"Trees on the Hill At Night"



• ; $0.75 \le \Phi \le 1.0$ • ; $0.25 \le \Phi \langle 0.75$ • ; $0.0 \langle \Phi \langle 0.25 \rangle$





 $(D_{\rm C}=5.0{\rm \AA})$
2008年, Haglundらは, 次のような, タンパク質 "1RIS" (ribosomal protein S6) の5個の円順列変異 体を作成して, それぞれのΦ値をもとめて, フォールディング・キネティクスの変化を議論している: P¹³⁻¹⁴ P³³⁻³⁴ P⁵⁴⁻⁵⁵ P⁶⁸⁻⁶⁹ P⁸¹⁻⁸²

彼らは, 次のような手順で"1RIS"の円順列変異体を作成している:

- ① まず、タンパク質のアミノ酸配列のN末端とC末端とを適当なリンカー配列で共有結合する.
- ② 次に、もとのアミノ酸配列の適当な部位(特に、2次構造のつなぎ目)を意識して、ペプチド結合 を一カ所を切断し新たな末端(N末端とC末端)とする.例えば、P¹³⁻¹⁴は、β1とα1のつなぎ の部分である、アミノ酸残基13番目と14番目を切断して、その部位を両末端とする.

この円順列変異体の隣り合うアミノ酸の配列は、もとの野性型と同じで完全に保存されている(ただし、末端部分だけが異なっている).しかしながら、野生型と円順列変異体の両方のアミノ酸配列をN末端側から並べて比べてみると、ほぼ異なるアミノ酸配列である.両方のアミノ酸配列間の同一性は、たかだか約5%程度である.

次に、我々は、Haglund らが実験した、"IRIS"の5個の円順列変異体の Φ 値を理論的に計算した. その結果、 Φ 値の理論と実験の間の相関係数は、次のようになった:

タンパク質	野生型(wt)	P^{13-14}	P^{33-34}	P^{54-55}	P^{68-69}	P^{81-82}
相関係数(CC)	0.20	0.89	0.47	0.57	-0.16	-0.20

- <文献> Haglund, E., Lindberg, M.O. & Oliveberg, M. Changes of protein folding pathways by circular permutation. Overlapping nuclei promote global cooperativity. *J. Biol. Chem.* **283**, 27904-27915, 2008.
- <文献> H. Wako and H. Abe, Calculation of Free-Energy Profiles, Folding Rates and Φ Values by Means of a Simple Statistical-Mechanical Model of Protein Folding, in Advances in *Protein Folding Research*, M. Hale ed., Nova Sci. Pub. Inc., pp. 19-63, 2016.
- <文献> H. Wako and H. Abe, Characterization of Protein Folding by a Φ -value calculation with a Statistical-Mechanical Model. *Biophysics and Physicobiology*, Vol. 13, pp. 263-279, 2016.

これは、何を意味するのだろうか?

我々の"簡単な統計力学的モデル"より求めた Φ 値が,野生型"1RIS"では,ほとんど一致しなかったが、3 個の円順列変異体 ($P^{13-14}, P^{33-34}, P^{54-55}$)では,相関係数 (CC)の改善がみられた.このことは、フォールディング・メカニズムは、"核形成一凝縮モデル"ではなく、"フレーム ワーク モデル"が適合するとみなしてよいのでないだろうか.つまり、アミノ酸配列のおける、2 次構造の順序を入れ替えただけで、フォールディング・メカニズムが変更するのではないかと推論できる.

野生型,及び,3個の円順列変異体 P^{13-14} , P^{33-34} , P^{54-55} の2次構造の順序は次の通りである: 野生型: $\beta 1 - \alpha 1 - \beta 2 - \beta 3 - \alpha 2 - \beta 4$

- P^{13-14} : $\alpha 1 \beta 2 \beta 3 \alpha 2 \beta 4 \beta 1$
- P^{33-34} : β 2-β 3-α 2-β 4-β 1-α 1

3 個の円順列変異体で, 共通の2次構造の順序が, 『β3-α2-β4-β1』のとき, フォールディング・ メカニズムは, 「フレーム ワーク」モデルに適合すると推論できる.

例として、2個の円順列変異体 (P^{13-14}, P^{33-34})の Φ 値の実験と理論の結果を下図と次のページ に示す.アミノ酸配列上と天然構造に重ね合わせて3色で分類して描いている.



1RIS円順列変異体[P¹³⁻¹⁴]のΦ値(理論と実験)

• $0.75 \le \Phi \le 1.0$: $0.25 \le \Phi \langle 0.75$ Δ : 0.0(Φ(0.25





• ; $0.75 \le \Phi \le 1.0$ • ; $0.25 \le \Phi \langle 0.75$ • ; $0.0 \langle \Phi \langle 0.25 \rangle$





"Autumn Night "

"あとがき"にかえて: 『フォールディングの統一的スキーム』

く要旨>

第12章

ここで採用されている、3次元格子タンパク質のアミノ酸配列は、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にデザインされてい る.そのようなアミノ酸配列を用いると、フォールディング・シミュレー ションで、ランダムコイル状態から天然構造へフォールドする.その際、天 然接触ペアの情報を前もって陽に与えていない.このことは、見方を変えれ ば、「このアミノ酸配列の中には、"天然接触ペアに関する情報"が"暗に"隠さ れているのではないか」と述べている.

Fersht らの"Φ値解析"による、実際のタンパク質のフォールディング過程の 遷移状態でのフォールディング核(部分的に天然構造と同じ構造を形成して いるアミノ酸残基)を推定できることが、その傍証ではないかと述べている. "タンパク質フォールディングの統一的スキーム"を次のように提案してい

る:『"実際のタンパク質"のアミノ酸配列には,進化の過程で獲得した,天然 構造における"天然接触相互作用"に関する情報が含まれていて,その情報にし たがって,ランダムコイル状態から,自らの天然構造へフォールドすること ができる.』

現在のところ、まだ、実際のタンパク質のアミノ酸配列の中に書き込まれ ているであろう"タンパク質の天然接触相互作用に関する情報"は解読されてい ないが、もし解読されれば、アミノ酸配列だけの情報から、その天然立体構 造を予測するという問題は解決されるだろうと述べている.

現在において個々のタンパク質のフォールディング・メカニズムを理論的 に議論するためには、天然構造が既知のタンパク質分子の、"全ての天然接 触相互作用"を考慮した場合の統計力学モデルを構築して系の分配関数を計 算し、その分配関数から、いろいろな熱力学量を求める必要があるだろうと 指摘している.このことは、現在のコンピュータの性能では難しいが、将来、 "量子コンピュータ"が実用化されれば可能になるだろうと述べている.

🎞 タンパク質フォールディングの統一的スキームについて

タンパク質の"フォールディングの統一的スキーム"について、Fersht は次のように述べている: 『タンパク質の多彩な構造様式,生体内でフォールディングする際のさまざまな環境,および,進化 においてタンパク質に課せられるさまざまな束縛を考慮すれば,フォールディングに単一の機構とい うのはありそうもない.それにもかかわらず,個々の特性を超えたフォールディングの一般的特徴が 存在して欲しいという期待がある.すなわち,"フォールディングの統一的スキーム"が存在し,そ れを若干変えるだけで数多くの経路を記述できるという期待である.・・・』

- <文献> Fersht, A., (1999) Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding. W. H.freeman, New York, Chapter 19.
- <文献> Mechanics of Protein Folding, 2nd ed.(2001), Edited by R.H. Pain, Chapter 7, Transition State of Protein Folding, V.R. Daggett & A.R. Fersht. (「タンパク質のフォールディング」第2版, R.H. Pain 編集, 2002, 日本語訳;崎山・桑島・河田, 第7章, 「タンパク質フォールディングの遷移状態」, V.R. Daggett & A.R. Fersht.)

我々が、タンパク質分子の多彩な天然構造を眺め、更に、個々のタンパク質に依存した個別的で多 様性のあるフォールディング過程を考察すると、タンパク質分子のフォールディング過程の単一の機 構というのは、とてもありそうには思えない.しかしながら、それでも、タンパク質の"フォールデ ィングの統一的なスキーム"が考えられないだろうか? 果たして、タンパク質の"フォールディング 転移の普遍的なメカニズム"は、どのようなものなのだろうか?

ここで、タンパク質の"フォールディングの統一的なスキーム"について考察してみよう. ジャック・モノーは、"はじめに"で引用しているように、彼の著書:『偶然と必然』の中で、"生命の 神秘の謎は、タンパク質のアミノ酸残基の配列順序のなかに封じ込められている"と述べている.そ して今日、我々は、タンパク質のフォールディングに関して第2章で紹介しているように、次のよう な"アンフィンゼン・ドグマ"を確信している:『タンパク質分子は、与えられた溶媒条件下で、自発 的に特異的天然立体構造へフォールドしていくが、このとき、アミノ酸配列と環境条件に含まれる情 報以外に、新たな情報を必要としない、つまり、"タンパク質分子の天然立体構造形成に関する情報は、 そのアミノ酸配列の中に、すべて書き込まれている".』

タンパク質分子は、与えられた環境の中で、その"アミノ酸配列の中に書き込まれた情報"に従っ て(!?)、自発的に揺いで"形"や"構造"を変化させながら、最終的に天然構造へフォールドして いくと考えられる.しかしながら、この"アミノ酸配列の中に書き込まれた情報"の中に隠されてい ると思われる"天然構造形成機構の原理"は、未だに解読されていない.『アミノ酸配列だけの情報か ら、その天然立体構造を予測するという問題』は、未解決の難問として残されているのだ!

□ アミノ酸配列の中に書き込まれた情報とは?

ここで、『アミノ酸配列の中に書き込まれた情報の中には、天然接触相互作用(タンパク質の天然構 造で互いに接触している分子間相互作用)に関する情報"が含まれている』という大胆な提案をしよ う.これは、『未だに解読されていない、とても受け容れ難い提案である』と思われるが、このことを 考察してみよう.

第2章で詳細に考察しているように、100個のアミノ酸配列からなるタンパク質において、20種類 のアミノ酸を順番に並べる配列の可能な場合の数は 20¹⁰⁰ (≃10¹³⁰) 通りで、莫大な数のポリペプチド 鎖が存在する.この莫大なポリペプチド鎖の内、今日の生命体で特異的立体構造を形成して実際に機 能しているのは、ごくごく限られた、一握りのポリペプチド鎖だけであり、それらを、我々は"タン パク質"と呼んでいる.生命誕生から今日までに、遺伝子が、すべてのポリペプチド鎖に対して、"タ ンパク質分子として機能するかどうか?"を、チェックしたとは到底思えない.だが、遺伝子は(自然は)、非常に長い生命の進化の過程で、莫大なポリペプチド鎖の中から、わずか一握りの、"生命現象で機能するタンパク質分子を選びだした"と言えるだろう.

Gōは、フォールディング過程におけるアミノ酸残基間の相互作用は、第一近似として、"天然接触 相互作用"(Gōポテンシャル)のみが主に働き、他の相互作用は無視できるだろうと仮定した.そし て、2次元格子タンパク質に対して、天然接触相互作用しているアミノ酸残基ペアの情報を"陽に" 与えてコンピュータ・シミュレーションを実行し、タンパク質らしい協同的振る舞いが起こることを 観察した.天然非接触相互作用(天然構造では接触していない相互作用)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、グローバルなファネル様地形では、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、フォールディング過程では"Gōポテンシャル"だけが主に作用すると仮 定したのである.結局、Gōポテンシャルは、タンパク質分子が、余計な構造探索を行わずに、間違 った構造を回避して短時間内に天然構造を形成する理想的なポテンシャルであり、また、Levinthalの パラドックスを回避するポテンシャルであると言えるだろう.

この Go ポテンシャルの影響下では、次のようなフォールディング過程の描像が考えられる: 『フォールディングの初期段階では、アミノ酸配列上で近い、近距離相互作用の天然接触相互作用が 働いて、いろいろな部位で部分構造を形成するが、この部分構造は比較的不安定で、壊れたり、また 形成されたりする.しかし、部分構造が形成されている間に、たまに、アミノ酸配列上で距離の離れ た、長距離相互作用の天然接触相互作用が働き、いくつかの局所構造が合体して、最終的に、天然構 造が形成される.』

そして、Gō は、次のような"整合性原理"を提案した;『球状タンパク質の天然構造を安定化しているいろいろな相互作用項は、第一近似としては、互いに矛盾が無く整合的である.この性質を満たすポリペプチド鎖のみが、進化の過程を通して、タンパク質分子として選択されてきた.』その後、Wolynes らは、本質的に"整合性原理"と等価である、次のような"極小フラストレーションの原理"を提案した;『タンパク質の天然構造においては、どのエネルギー項をとっても他のエネルギー項の犠牲になってフラストレーションを感じることなく、フラストレートした相互作用が十分小さくなるように、進化を通じて、精巧に、そのアミノ酸配列をデザインしてきた.』

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、 乱暴な、人為的な、しかも、物理的でないポテンシャルであると見なされていた.しかしながら、タ ンパク質の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 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 過程にも第一近似として使えるのではないかと考えられ、実験による検証が必要だと考えられるよう になった.

Fersht らは、遺伝子操作を応用して、"Φ値解析"という新たな手法を開発し、"実際のタンパク質" のフォールディング過程の遷移状態でのフォールディング核(部分的に天然構造と同じ構造を形成し ているアミノ酸残基)を推定することを可能にした.そして、実験によるΦ値解析の結果は、驚くべ きことに、Gō ポテンシャルを採用したモデル計算の結果と定性的に一致することが確認された.つ まり、フォールディング過程の遷移状態で、天然接触相互作用が働いているアミノ酸残基を推定でき たことになる.

181



3次元格子タンパク質のアミノ酸配列に書き込まれている(!?) "天 然接触相互作用"の情報

米国の Shakhnovich らによって、3 次元格子タンパク質によるコンピュータ・シミュレーションが 1990 年代後半から本格的に実行された.その際、彼らは、まず、コンパクな3 次元モデル格子タンパ ク質の天然構造をデザインし、そのデザインした天然構造にフォールド可能なアミノ酸配列を次のよ うな方法で決定した:『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に定義し た目的関数が極小値をもつように、モンテカルロ・シュミレーションを実行してアミノ酸配列を決定 する.』 その結果、彼らは、与えられた天然構造に対して、フォールド可能な複数のアミノ酸配列を 決定することが出来た.第5章で、実際に我々が採用した、Protein a1、Protein a2(36 個のアミノ酸残 基数)と、Protein b1、Protein b2(48個のアミノ酸残基数)は彼らによって求められたもので、それぞれ、天然構造が同じだがアミノ酸配列が異なっている3次元格子タンパク質である.

これらの3次元格子タンパク質に対する、それぞれのアミノ酸配列を用いて、フォールディング・ シミュレーションを実行すると、アミノ酸配列の中に、天然接触相互作用しているアミノ酸残基ペア の情報を"陽に"加えなくても、つまり、"Gō ポテンシャル"を"陽に"仮定しなくても、いずれの タンパク質も、ランダムコイル状態から天然構造へフォールドすることが確認されている.

このことから,次のようなことが言えるだろう:『Shakhnovich らの方法によって求められた3次元 格子タンパク質のアミノ酸配列の中には、"暗に",天然接触相互作用しているアミノ酸残基の情報が 含まれているとみなしてよい.』

タンパク質のフォールディング・メカニズムに関するモデルをまとめておこう!

長い年月の生命進化を通じて、そのアミノ酸配列をデザインしてきたであろう"実際のタンパク質" のアミノ酸配列の情報の中にも、『天然構造における"天然接触相互作用"に関する情報』が、"暗に" 含まれているとみなせないだろうか?

ここで,タンパク質のフォールディング・メカニズムのモデルをまとめていこう.

我々は、第9章~第11章で、27個の"実際のタンパク質"に対して、簡単な統計力学モデルを適 用して理論的に計算したΦ値と、実験的に観測されたΦ値と比較・検討した.そして、"A-W_NILS モ デル"で仮定したスキームが妥当である、つまり、『短距離相互作用、中距離相互作用、長距離相互作 用が、順に、階層的に、しかも相互作用は局所構造内だけに作用し、最終的に、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へフォールドする』というスキームを満たしている と考えられる"実際のタンパク質"が幾つか判明した.

タンパク質のフォールディング・メカニズムのモデルである"フレームワーク モデル"は、"フォ ールディングには決まった道筋があり、αヘリックスやβシートなどの2次構造やドメインなどの特 別な局所構造が階層的に形成される"という仮定を前提にしているが、「A-W_NILS モデル」の方は、 もっと"確率的"なモデルで、αヘリックスやβシートなどの2次構造が、それらの統計重率によっ て、任意の順序で形成することができるとみなすモデルである。両方のモデルの重要な共通点は、『短 距離相互作用が働く前に、長距離相互作用が働く事はない』と仮定している点である。ところで、

「A-W_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, "Gō ポテンシャル"が局所構造内だけに作用するとする,より理想化された仮定であるとみなされる.

一方、タンパク質のフォールディング・メカニズムのモデルである"核形成-凝縮モデル"のスキー ムに沿ってフォールドしていると考えられる"実際のタンパク質"も幾つか同定された.このモデル は、『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が内部に、親水性アミノ 酸残基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、"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』という モデルである.この"核形成-凝縮モデル"の重要な点は、『短距離相互作用が作用する前か、あるい は、ほぼ同時に、長距離相互作用が働く』と仮定している点である. 第12章 "あとがき"にかえて:『フォールディングの統一的スキーム』

↓ "実際のタンパク質"のフォールディングの統一的スキーム

前節でまとめているフォールディング・メカニズムのスキームに沿った"実際のタンパク質"の例 が、それぞれ存在することをどのように考えたら良いのだろうか? 異なるフォールディング・メカニ ズムのスキームから、タンパク質の"フォールディングの統一的スキーム"が導けるだろうか?

これらのフォールディング・メカニズムのモデルには、共通点がある.それは、"タンパク質の天然 接触相互作用"の重要性である."実際のタンパク質"は、アミノ酸配列に書き込まれている天然接触 相互作用の情報により、天然構造を形成できるのではないだろうか?タンパク質分子のフォールディ ング過程を明らかにすることは、そのアミノ酸配列中に書き込まれているであろう特異的天然構造に 関する暗号を、自然が解読する過程を明らかにすることに他ならない.

タンパク質分子自身の立場になって、フォールディング過程を考えてみよう:

『タンパク質分子は、生体内でフォールディングする際のさまざまな環境に応じて自発的に揺いで "形"や"構造"を変化させている。その際、アミノ酸配列の中に書き込まれている"天然接触相互 作用に関する情報"にしたがって、あるときは階層的な"フレームワークモデル"で、あるときは 非階層的な"核形成-凝縮モデル"で、どちらのスキームによっても、最終的な天然構造へフォールド することができる。』

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"Go ポテンシャル"の本 質は、天然構造で接触しているペアを指定することによってそのタンパク質の特異的立体構造をモデ ルに組み込んでいると捉えることができる.実際、接触している残基の情報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.

ここで, 我々は, 次のような "タンパク質フォールディングの統一的スキーム"を提案しよう: 『"実際のタンパク質"のアミノ酸配列には,進化の過程で獲得した, 天然構造における "天然接触相 互作用"に関する情報が含まれていて,その情報にしたがって,ランダムコイル状態から自らの天然 構造へフォールドする.』

現在のところ,まだ,実際のタンパク質のアミノ酸配列の中に書き込まれているであろう "タンパ ク質の天然接触相互作用に関する情報"に関しては,我々は何も知らないし,解読されるかどうかも わからない.それ故,アミノ酸配列だけの情報から,その天然立体構造を予測するという問題も解決 されるかどうかもわからない.

これについては近年,大きな進歩があった.人工知能によって高精度で予測が可能となったのだ. このことはもちろん,アミノ酸配列だけの情報からその天然立体構造を予測することが可能であるこ とを示しているが,一方で、"人工知能がいかなる論理で予測したのか?"を現時点では知ることがで きないため、われわれがそれを理解した、という段階にはまだまだ至っていないのが現状である.

(<文献>J. Jumper et al., Highly accurate protein structure prediction with AlphaFold, Nature, 596, 583-589, 2021.)

ところで,現在までにすでに決定されているタンパク質の天然構造を利用してフォールディング 機構が解明できないだろうか?

天然構造既知のタンパク質では、当然、"タンパク質の天然接触相互作用に関する情報"が"陽に" 得られる.この情報をもっと積極的に用いれば、"タンパク質フォールディング問題"の第2の命題で ある、フォールディング・メカニズムの問題、つまり、『タンパク質のフォールディング過程の性質(レ ート、中間状態、遷移状態など)を明らかにする問題』は、解明できるのではないだろうか?

↓↓"天然接触相互作用によるフォールディングモデル"の提案

"A-W_NILS モデル"は、"天然接触相互作用が局所構造内のみで働く"と仮定した統計力学モデルであるが、この仮定を拡張した新たな統計力学モデル、"天然接触相互作によるフォールディングモ

デル"を構築することを考えよう:

- ① アミノ酸残基をユニット(球)で表す
- ② 各のアミノ酸残基(ユニット)は、天然状態か("1"で表す)、変性状態か("0"で表す)、のどちらかの状態とると仮定する
- ③ 局所構造の定義

あるコンフォメーションにおける,各アミノ酸残基の状態を,"1"と"0"の並びで表す. 今、例として、100個のアミノ酸残基からなるタンパク質を考えよう:

あるコンフォメーションにおける各アミノ酸残基の状態を,"1"と"0"の並びで表すとき,100個 のアミノ酸残基からなるタンパク質の,"1"と"0"の並び方の場合の数(コンフォメーションの数) は、2¹⁰⁰≈10³⁰ 個である.このとき,部分的に連続した"1"の領域(両端は"0")を、タンパク 質の天然構造と類似の構造であるとみなして,その領域を"局所構造"と定義する.また,部分的 に連続した"0"の領域は、変性状態領域と定義する.

当然,局所構造内では天然接触エネルギーを考慮するが,次のような場合の"天然接触相互作用 のエネルギー"も採用する;

① 局所構造間の相互作用・・・図(a)

② 局所構造と変性状態領域との間の相互作用・・・図(b)

③ 変性状態領域間の相互作用(・・・図には示していない)



④実際のタンパク質のコンフォメーションに対するエントロピー項の正確な関数形を確立する必要 がある(第9章参照).

この新たな統計力学モデル、"天然接触相互作によるフォールディングモデル"を用いて、フォール ディング過程の遷移状態におけるΦ値を理論的に求めれば、実験によるΦ値解析の結果をかなり再現 するのではないかと期待できる.つまり、莫大な費用と時間をかける必要がある遺伝子操作によるΦ 値解析のかわりに、理論的に計算したΦ値を求めて、タンパク質のフォールディング遷移状態の構造 の推定が容易にできると思われるが、まだ機が熟していない.これら全てのコンフォメーションに対 して、全ての天然接触相互作用を考慮した場合の分配関数を求めるのは、現在のコンピュータの性能 では難しいと思われる.将来、"量子コンピュータ"が実用化されれば可能になるだろう.

<コメント>:第5章の「Coffee break」で、「局所構造内の全ての天然接触相互作用のエネルギー」 と「局所構造間の一部の天然接触相互作用のエネルギー」を考慮した場合の計算結果を示している。 そこでは、「天然接触相互作用が局所構造内だけ」と仮定した「A-W_NILS モデル」による転移曲線 より、コンピュータ・シミュレーションによる転移曲線の結果に、より接近した曲線が得られている。

🚨 "多様性こそ自然である!"

第6章と第8章で、"3次元格子タンパク質"に対して、同じ天然構造をもつ、異なる2個のアミノ酸配列を採用して、フォールディングに関する統計力学モデル(「A-W_NILS モデル」)を用いて求めた熱力学量を様々な角度から検討し、天然構造が同じであることによる共通点と、アミノ酸配列が異

第12章 "あとがき"にかえて:『フォールディングの統一的スキーム』

なるがゆえの相違点は、タンパク質の立体構造とアミノ酸配列の関係に関して様々な示唆を与えてく れたが、次のような疑問が生じる:『異なった天然構造の3次元格子タンパク質を採用すれば、また 異なった結果が引き出されてくるのではないか・・・?』

更に, 第9章~第11章で, いくつかの"実際のタンパク質"で, 実験と理論によるΦ値の分布から, フォールディングの遷移状態での構造形成を推定してフォルールディング・メカニズムを検討したが, ここでも次のような疑問が生じる: 『他の"実際のタンパク質"では, また違った結果になるので は・・・?』

このような不安が常に残ってしまう.しかしながら,その個別性を無視した総括的な議論だけでは 意味がない.むしろ,『フォールディング転移の普遍的なメカニズムは何か?』と,『個々のタンパク 質に依存した個別的で多様性のあるフォールディング・メカニズムとは何か?』の狭間でゆれながら, そのような両面性をもった研究対象であるがゆえの面白さを,タンパク質フォールディングの研究に おいて見出すことが重要であろう.

輪湖博は,著書:『遺伝情報の科学』の「はしがき」で"多様性こそ自然である"と次のように述べている:『・・・多くの社会科学がそうであるように,生物学もまた,ここ 35 億年の間に起こった 事象の研究というケーススタディにすぎないのである.普遍性をもった法則を見出そうなどとうそぶ いている研究者にとって,このケーススタディということは大きな困惑を生じさせるものなのである. しかし一方で,物理科学の方も変化してきた.自然の多様性ということを少しづつではあるが認識し 始めたのである.物理学が対象とする,単純で決定論的な機構によって変化する系の中にも,その未 来が予測不可能で,混沌とした様相を示すものがあることがわかってきた.最適解が無数にあって, 解析的にはそれらをすべて求めることができないような系がたくさんあることがわかってきた.・・・ こうした自然の描像は,自然は単純な美しさをもつと信じられていたこれまでの自然科学に対して, 複雑であるがゆえの美しさこそ自然の本質である,と主張し始めているように私には思えてならない. 多様性こそ自然なのである,と・・・』(輪湖博著:『遺伝情報の科学』,成分堂,1990年)

『はじめに』で述べているように、タンパク質のフォールディング・メカニズムの解明は、この本 の最後の頁から展開すると思っている.この分野を解明しょうとする人々にとって、我々の提案が、 何らかの手引きに少しでもなることを切に願いつつ、この分野に興味をもつ若い研究者に、次のよう な、武者小路実篤の言葉を贈って筆を置こう(いや、パソコンの日本語のソフトを閉じよう).

(武者小路実篤)



"この道より,我を生かす道なし.この道を歩く"

"Path beside the sea II"

186

安 部 晴 男(Haruo Abe) 著者略歴

1944 年 福岡県古賀市に生まれる. 1968 年 熊本大学理学部物理学科卒業.1970 年 熊本大学大学院理学研究科物理学専攻修士課程修了.1970 年 西日本工業大学教養 教室助手.1986 年 理学博士(九州大学).1987 年 米国国立衛生研究所(NIH) 留学(1 年間).1988 年 西日本工業大学教養教室教授.2009 年 西日本工業大学定年 退任,名誉教授.

主な書籍等出版物

- N. Gō and H. Abe, M. Mizuno, H. Taketomi, Local Structures in the Process of Protein Folding, in Protein Folding, pp. 167-181, R. Jaenicke ed., Elsevier, Amsterdam, 1980.
- ・ 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(分担執筆), 早稲田大学複雑系高等学術研究所編, 共立出版, 59-97, 2006.
- H. Wako and H. Abe, Study of Folding/Unfolding Kinetics of Lattice Proteins by Applying a Simple Statistical Mechanical Model for Protein Folding, in Protein Folding, E.C. Walters ed., pp. 349-376, 2011.
- H. Wako and H. Abe, Calculation of Free-Energy Profiles, Folding Rates and Φ Values by Means of a Simple Statistical-Mechanical Model of Protein Folding, in Advances in *Protein Folding Research*, M. Hale ed., Nova Sci. Pub. Inc., pp. 19-63, 2016.

輪湖博(Hiroshi Wako)

著者略歴

1951年 長野県松本市に生まれる. 1973年 早稲田大学理工学部物理学科卒業. 1978年 早稲田大学大学院理工学研究科物理学及応用物理学専攻博士課程修了(理学博士). 1978年 コーネル大学博士研究員(~1981年). 1982年 早稲田大学助手. 1986年 早稲田大学専任 講師. 1993年 早稲田大学社会科学総合学術院社会科学部教授. 2021年 早稲田大学定年退任,名誉教授.

主な書籍等出版物

- ・ 『BASICによる生物』: (共著) 共立出版, 1987年.
- ・ 輪湖 博, 郷 信広, "蛋白質の立体構造シミュレーション", 『計算物理学と計算化学』(分担執筆), 海文堂, 1988年.
- 『遺伝情報の科学』:輪湖博,成分堂,1990年.
- ・ 輪湖 博, "タンパク質という複雑系へのアプローチ", 『タンパク質のかたちと物性』(分担執筆), 共立出版, 164-176, 1997年.
- ・ 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(分担執筆), 早稲田大学複雑系高等学術研究所編, 共立出版, 59-97, 2006年.
- H. Wako and H. Abe, Study of Folding/Unfolding Kinetics of Lattice Proteins by Applying a Simple Statistical Mechanical Model for Protein Folding, in Protein Folding, E.C. Walters ed., Nova Sci. Pub., pp. 349-376, 2011.
- H. Wako and H. Abe, Calculation of Free-Energy Profiles, Folding Rates and Φ Values by Means of a Simple Statistical-Mechanical Model of Protein Folding, in Advances in *Protein Folding Research*, M. Hale ed., Nova Sci. Pub. Inc., pp. 19-63, 2016.







Memorandum



あ

Ising モデル, 39
Island モデル(島モデル), 38, 39
アデニンとチミン, 2
アミノ酸置換, v, 83, 84, 85, 89, 123, 124, 126, 127, 128, 136, 148, 149
アミノ酸変異導入, 136
アミノ耐変定異導入, 136
アシロイド, 5
アレルギー反応, 6
アンフィンゼン・ドグマ, ii, 17, 18, 180
アンフォールディング・レート, v, 93, 94, 98, 99, 102, 103, 123, 130, 139, 140

い

異常型プリオンタンパク質,5 1次元情報,5 1次相転移的(協同的),72 遺伝コード表,iv,4,11 遺伝子工学,14,124,136 遺伝子操作,14,181,185 遺伝子編集技術,14

え

A-W_NILS モデル, iv, v, 37, 46, 50, 54, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 71, 84, 129, 145, 146, 151, 152, 154, 155, 162, 173, 183, 184, 185
S 字状曲線 (sigmoidal transition curve), 30
SH3 ドメインファミリー, vi, 151, 155, 159, 161
HP 模型, 24
エネルギー曲面, 18, 19, 20, 23, 81
エネルギーの期待値, 27, 29, 53, 55, 59, 61, 62
エネルギーランドスケープ理論, 15, 20, 23
塩基性アミノ酸残基, 12
遠距離相互作用, 22, 40
円順列変異体, vi, 165, 175, 176
エンタルピー(エネルギー) 獲得, 21
エントロピー損失, 21, 54

お

old view (古典的捉え方), 20 岡崎フラグメント, 8 折れたたみ過程, 20, 126

か

Karplus, 73 回帰係数, 87, 88 鍵と鍵穴, 6 核形成-凝縮モデル, v, vi, 135, 138, 141, 143, 145, 152, 154, 155, 183, 184 活性化エネルギー, 102 活性化状態, 125, 126 カノニカル分布(正準分布), 76 感受性(susceptibility), 83, 89 緩和時間, v, 93, 94, 96, 97, 98, 99, 102, 109, 111, 115, 116, 117, 119, 140

き

キネティクス, v, 93, 94, 96, 97, 102, 105, 108, 109, 129, 140, 175
キモトリプシンインヒビター2 (3CI2), v, 135, 144, 145
逆転写酵素, 3
協同現象, 21
極小フラストレーションの原理, iv, 15, 22, 23, 181
局所的な回転(Local rotation), 77
局所的な並進(Local translation), 77
極性アミノ酸残基, 12
近距離相互作用, 22, 38, 39, 181

く

グアニンとシトシン,2 偶然と必然,ii,180 組み換え DNA 技術,136 CRISPR-Cas9 (クリスパー キャスナイン),14

_

Gō & Abe (郷・安部), 37, 39 郷信広, 15, 20, 22, 125, 126 郷モデル, 22, 182 Gō like model, iii Gō ポテンシャル, 22, 23, 40, 63, 67, 72, 181, 183, 184 コドン, 4, 11 固有値, 95, 96, 109, 110, 111 固有方程式, 95, 96, 97, 109, 110 コンシステンシー原理, 74 コンフォメーション変化, 31, 76, 77, 78

さ

笹井理生, 74, 127

Zuckerman, 26

3 次元格子タンパク質, v, vi, 23, 31, 37, 42, 43, 44, 49, 51, 57, 58, 59, 62, 69, 74, 75, 76, 77, 78, 80, 83, 84, 90, 97, 98, 99, 101, 103, 122, 123, 124, 129, 132, 138, 139, 140, 154, 179, 182, 183, 185 酸性アミノ酸残基, 12

l

ジェームズ・ワトソン,2,7 ジェニファー・ダウドナ,14 シェブロン・プロット, v, 93, 94, 99, 100, 101, 102, 103, 130, 132, 140 chevron rollover, 103 自己組織化,18 指数関数的に, 19, 27, 34, 95 ジスルフィド結合(SS 結合), 18 磁性体の状態転移,39 自然の多様性,186 島モデル、37 Shakhnovich, 57, 58, 67, 73, 74, 75, 76, 182, 183 ジャック・モノー, ii, 180 自由エネルギー最小状態,18 自由エネルギープロファイル, 93, 94, 97, 109, 122, 124, 127 自由な局所構造モデル,39 主要組織適合抗原(Major Histocompatibility Complex), 6 詳細釣合(detailed balance), 96, 116, 120

新型コロナウイルス感染症,6 人工知能,ii,184 親水性アミノ酸残基,12,41,183

す

Snake cube puzzle, 23, 24 スパイクタンパク質, 6 スピングラス, 22

せ

整合性原理, iv, 15, 20, 21, 22, 23, 67, 73, 181, 182 整合性原理の具体例, 24 正常型プリオンタンパク質, 5 生理的環境条件, 26, 27, 32 絶対コンタクト・オーダー, 105 遷移確率, 77, 78, 116, 119, 120, 121, 122 漸化式, iv, 37, 39, 44, 45, 48, 49, 51, 53, 54, 55 全体的な回転 (Global rotation), 77 全体的な並進(Global translation), 77 セントラル・ドグマ, iv, 1, 3

そ

相関係数, v, 142, 145, 146, 151, 152, 155, 156, 157, 158, 161, 165, 166, 169, 170, 171, 173, 175 相対コンタクト・オーダー, 105 相転移, 21, 126 疎水性アミノ酸残基, 12, 41, 137, 154, 158, 183 疎水相互作用, 24

た

第1次遺伝情報解読問題,4 第2次遺伝情報解読問題,iv,1 第ゼロ近似,182 高田彰二,22,182

ち

秩序構造, 18, 21, 126 秩序度, 83, 125, 126 秩序変数, 59, 62, 81, 89 中距離相互作用, 39, 106, 154, 183 中性アミノ酸残基, 12

て

DNA (デオキシリボ核酸, Deoxyribose Nucleic Acid), 7
Dill, 25, 32, 73
寺本 英, 122
転移行列, 95, 96, 109
転移曲線, 25, 30, 47, 53, 59, 60, 61, 62, 63, 65, 66, 67, 69, 185
天然構造形成機構の原理, 180
天然構造のトポロジー, vi, 142, 151, 155, 159
天然接触相互作用, iv, vi, 21, 22, 23, 41, 67, 69, 179, 180, 181, 182, 183, 184, 185
天然接触ペア, vi, 40, 41, 42, 43, 44, 45, 54, 62, 63, 67, 72, 81, 107, 142, 179
天然非接触相互作用, 21, 67, 181
天然非接触ペア, 43, 72

と

統計重率, 45, 49, 51, 52, 55, 61, 154, 183 統計的に釣り合っている, 26 統計物理学的特徴, 125 統計力学モデル, iii, iv, v, vi, 31, 37, 38, 39, 40, 41, 42, 44, 46, 47, 50, 51, 57, 61, 62, 66, 67, 69, 84, 129, 139, 142, 145, 146, 151, 152, 154, 155, 162, 165, 171, 173, 179, 183, 184, 185 Thomas Lamb, iii

特異的な立体構造, ii, 4, 16, 17, 40 利根川進, 6

な

永山國昭, iii

に

2 次元格子モデル, iii, 21
二重らせん構造, 2, 7
20 種類のアミノ酸, iv, 4, 5, 11, 12, 16, 59, 74, 180
二状態転移, 21, 38, 72, 74, 124, 125, 144
new view(新しい捉え方), 20
NILS モデル, 37, 39, 41, 42, 44, 46, 51, 54, 62, 63, 66, 67, 68, 145, 151, 154, 183

の

Non-Interacting Local Structure model, 37, 39

は

胚, 126 排除体積効果, 37, 43, 46, 47, 54, 74, 76 Haglund, 175 反応座標, v, 83, 89, 90, 96, 97, 98, 101, 108, 116, 117, 118, 122, 123, 124, 125

ひ

PDB(Protein Data Bank), 142 非極性アミノ酸残基, 12 微視的状態の場合の数, 32 比熱曲線, 25, 30

રુ

Φ値解析法, v, 22, 123, 124, 128, 147 Fersht, 22, 90, 124, 127, 135, 136, 144, 147, 149, 154, 170, 179, 180, 181 van der Waals 力, 24 ファネル状(漏斗状),20 ファネル描像, iv, 15, 23 ファネル理論, 20, 21 フォールディング・シミュレーション, vi, 31, 41, 44, 50, 58, 59, 61, 62, 67, 179, 183 フォールディング・レート, v, 93, 94, 96, 97, 98, 99, 101, 102, 103, 105, 107, 109, 119, 123, 127, 129, 130, 131, 139, 140 フォールディング異常病, iv, 1,5 フォールディング核, v, vi, 123, 129, 132, 133, 138, 145, 156, 157, 158, 162, 166, 169, 170, 171, 173, 179, 181 フォールディング転移, 21, 41, 72, 73, 74, 180, 186 フォールディングの統一的スキーム, vi, 179, 180, 184 フォールディングの統計的道筋, v, 83 フォールディングの統計力学的描像,38 フォールディング問題, iv, 1, 4, 15, 16, 18, 23, 184 Franklin, 7 フランシス・クリック, 1, 2, 3, 7 プリオン病、5 プルミナー.5

フレームワーク モデル, v, vi, 135, 138, 141, 143, 145, 154, 155, 165, 166, 169, 183, 184
プロテインA (1SS1), v, 135, 136, 137, 138, 141, 142, 143
protein L, 162
protein G, 162
分子機械, ii, 26, 32
分子シャペロン, 18
分子生物学, iv, 1, 2, 7, 8
分子生物物理学, 26
分子動力学, 74, 136
分配関数, iv, vi, 25, 27, 28, 29, 34, 37, 39, 41, 44, 45, 48, 49, 51, 52, 53, 55, 61, 67, 69, 98, 138, 139, 179, 185
分布密度, 125

~

平衡状態, iv, 18, 21, 25, 26, 27, 29, 32, 33, 95, 96, 98, 109, 110, 111, 115, 116, 117, 118, 119, 122 並進変換, 78, 79 ヘリックス・コイル転移, 38 変異体タンパク質, 84, 89, 124, 127, 128, 129, 130, 131, 136, 139, 140, 147, 148 変性・再生の実験, 17 Henry & Eaton, 93, 97, 98, 109, 115, 119, 120, 122

ほ

保形変換, 78, 79 ホモポリマー, 40 ボルツマン因子 (Boltzmann factor), 27, 33 ボルツマン分布, iv, 25, 26, 27, 32, 34 翻訳, 4, 5, 6, 11

ま

マクロ状態のエネルギー,28 マスター方程式,v,93,94,96,97,108,110

み

Muñoz & Eaton, 93, 97

む

武者小路実篤,186

め

mRNA(メッセンジャーRNA),3 Metropolis-Teller 法, 76, 77, 78 免疫グロブリン, 6

も

モルテン・グロビュール, 20 Molecular dynamics; MD, 136

や

野生型タンパク質, 84, 89, 123, 127, 128, 129, 130, 131, 136, 138, 139, 140, 147, 148, 165

6

ラグランジュの未定乗数法, iv, 25, 26, 27, 33, 35 乱数, 76, 78, 79 ランダムコイル状態, vi, 20, 21, 40, 42, 45, 51, 52, 53, 54, 66, 98, 99, 116, 119, 124, 125, 126, 128, 136, 179, 183, 184 ランダムサンプリング, 19

ŋ

理想気体モデル, ii, 38 リチャード・ファイマン, 26 リボヌクレアーゼ A, 18 量子コンピュータ, 31, 179, 185

る

累積度数分布, 104, 105, 106, 107

れ

レート方程式(rate equation), 94 レトロウイルス, 3 Levinthal のパラドックス, 19 わ

Wako & Saitô(輪湖・斉藤), 37, 38, 39 輪湖博, 186



"Two Trees in the Mist at Dusk"

Memorandum