

## 解説

# ポストゲノム時代と蛋白質分子熱力学の重要性

曾田邦嗣

(受取日：2005年1月4日，受理日：2005年1月13日)

## Post-genome Era and Importance of Protein Thermodynamics

Kunitsugu Soda

(Received January 4, 2005; Accepted January 13, 2005)

It is described that genomic-DNA sequence analyses of many living organisms have been completed, and studies on structural and functional genomics are now proceeding. The dual nature of natural proteins is remarked that they have two different aspects of technological and scientific entities, which is distinct from industrial functional elements. Characteristic aspects of the molecular thermodynamic quantities of proteins, and their physical and biological implications are explained including the enthalpy-entropy compensation and the large positive heat-capacity change in protein denaturation. Finally, perspectives on the research of protein thermodynamics at the post-genome era are presented and a possibility of the development of protein science in the expanded sequence space is discussed.

### 1. 生命科学研究とゲノム解読

2000年6月，国際ヒトゲノム計画プロジェクトチームと米国のバイオベンチャー企業のセセラ社が，互いに独立に，30億個の文字からなるヒトゲノム塩基配列の概要を明らかにしたと発表した。更に2004年にはその精密解析が完了し，人間の設計図であるヒトゲノムは，約22,000個の遺伝子を含むことが確認された。ヒトゲノムの解読と並んで，チンパンジー，マウス等の哺乳類，鶏，ゼブラフィッシュ，フグ等の脊椎動物，シヨウジョウバエ，線虫等の無脊椎動物，イネ，シロイヌナズナ等の植物から，微生物，細菌類を含む多様な生物のゲノム塩基配列が決定されている。このように多くの生物種についてゲノムの解読が進んだ結果，今や生命科学研究は，狭義のポストゲノム科学 (genomics) の時代，より正確には，ポスト配列ゲノム科学の時代に入りつつある。実際，ゲノムの遺伝情報がコードする蛋白質分子の立体構造を網羅的に決定する研究 - 構造ゲノム科学

(structural genomics) - の重要性が高まり，数千種と推定される蛋白質の基本構造を解明する研究が，日本では，蛋白3000プロジェクトとして推進されている。これが完了すれば，任意の遺伝子の塩基配列について，それがコードする蛋白質の構造を知る，あるいは予測することができる。その構造に基づいて遺伝子の機能を解析する研究 - 機能ゲノム科学 (functional genomics) - が展開できるようになりつつある。更に進んで，塩基配列の情報から直接，それがコードする蛋白質の機能を予測する研究も進んでいる。これ迄の分子生物学の研究では，主に生物機能素子としての単独の，あるいは少数の関連蛋白質を対象として，それらの構造・機能の解析が行われて来た。これに対して，生命の単位である細胞を，多くの分子機能素子で構成される生物機能システムとして捉え，その機能発現の機構を研究するプロテオーム解析が進められている。本稿では，蛋白質分子熱力学の現状を述べると共に，生命科学研究の進展の中で間もなく到来することが予想される，真のポストゲ

ノム時代におけるその有り様を展望する。本論に入る前に、次節では、生命科学分野外の読者のために、予備知識を提供しておこう。

## 2. ゲノムとDNA<sup>1)</sup>

生物は、親の持つ遺伝情報を子孫が正確に受け継ぐことにより種の安定性を保つと共に、その情報を発現することによって生命を維持している。この遺伝情報を担っている細胞内の構造が染色体である。これは、動植物の細胞が分裂する際に現れる棒状の構造であり、塩基性色素で染色されるので光学顕微鏡でも観察できる。細胞内の染色体の本数は、生物種によって決まっている。高等生物の体細胞は通常二倍体であり、基本数個の相同染色体を1対ずつ持っている。周知の通り、人間では基本数が22(常染色体) + 1(性染色体) = 23で、全ての体細胞が23対46本の染色体を持っている。各染色体上には、一群の遺伝子が存在している。生物がその生命を維持するために必要な最小限の遺伝子群を含む、基本数個の染色体の一組を、ゲノムという。

染色体は、主に高分子DNA(デオキシリボ核酸)とヒストン蛋白質からなり、遺伝情報を担っているのがDNAである。二重らせん構造を持つ紐状のDNA分子鎖は、ヒストンに巻き付くことによってヌクレオソーム構造を形成し、それが折り畳まれてソレノイド構造になる。更に折り畳みを繰り返して約1万分の1に小さくなる。伸張されたヒトDNAの長さは約1mに達するが、これが10 $\mu$ mに折り畳まれて細胞内に収まっている。

遺伝子DNA上に核酸塩基A, T, G, Cの4文字の配列として書かれた遺伝情報は、先ずm-RNA(伝令RNA)転写される。次のステップで、m-RNA上の塩基配列が、遺伝コードに従って3文字ずつ読みとられ、特定のアミノ酸に翻訳される。その結果、1個の遺伝子の塩基配列情報から、指定されたアミノ酸配列を持つ1本のポリペプチド鎖=1個の蛋白質分子が生成される。この一方向性の情報の流れをセントラルドグマという。このようにして生体内で紐状の分子として合成された蛋白質は、生理的条件下で、「自発的」に折り畳まれて固有の立体構造を形成する。蛋白質鎖は、この立体構造を形成することによってのみ、特異的な機能を発現できるようになる。またこの折り畳み過程は、基本的に他の物質の支援に依らない物理的な過程である。<sup>2)</sup>従って、蛋白質が機能を発現するために必要な全ての情報は、遺伝子の塩基配列の中に書き込まれていると言える。

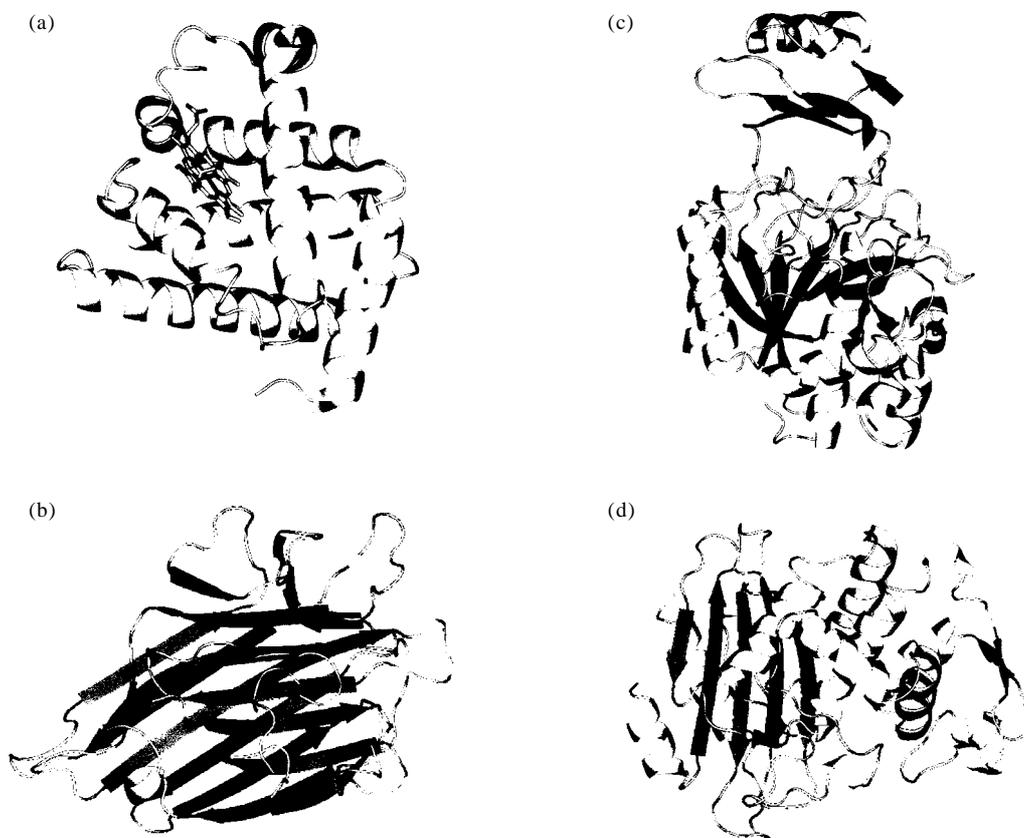
## 3. 生物機能素子としての蛋白質の二重性

前節から、生物は、自己複製によって子孫を残すと共に、遺伝情報を発現して作る数千~数万種の遺伝子産物-主に蛋白質-を利用することにより、自立して生命を維持でき

る、多機能分子機械システムと見ることができる。このシステムの設計図、或いは設計文書の役割を果たすのがゲノムの遺伝情報である。但し工業製品の設計者が人間であるのに対して、蛋白質分子を設計するのは、遺伝子の重複・突然変異などによる遺伝子産物の修正と環境の自然選択による産物の選別からなる、分子進化の過程である。<sup>3)</sup>工業製品と蛋白質では、設計の主体は全く異なるが、評価と修正を繰り返して産物を改良していくという点では、両者の設計過程は非常に良く似ている。即ち、蛋白質を分子進化によって設計された分子機能素子として見ると、それは、「コンピュータの演算素子や記憶素子と同質の存在であり、「工学的な存在」と見ることができる。

その一方で、人工の機能素子と蛋白質は、明確に異なる側面を持っている。前者では、充分大きなS/N比で確実な動作が保証されている単位素子を組み合わせることによって、複雑な機能が設計され、実現される。従って、単位素子には揺らぎによる誤作動の可能性の検討は必要ない。あるいは、そのような可能性のない設計が始めから求められている。一方後者では、上の可能性を考慮することが本質的に重要である。それは、蛋白質に課せられている大きな制約に起因する：生物は、室温近傍の温度で、生体分子の有機化学反応や分子認識過程を通して、自身を構築し、多様な機能を発現し、制御しなければならない。その中で最も基本的な過程が、蛋白質分子の立体構造形成=折り畳み過程である。それは、必要な機能のみを発現できる構造を、歩留まり100%で生成できなければならない。前節で述べた通り、蛋白質の折り畳み過程は自発的に為されなければならないから、機能を発現できる構造=天然構造は、熱力学的に最安定で、かつ熱力学的に安定な唯一の構造でなければならない(これは、アンフィンゼンのドグマと呼ばれている)。換言すると蛋白質は、会合体の形成も含めて、安定な非天然構造を形成せず、かつ不要時には直ちに処分(分解)できなければならない。

多くの蛋白質は、天然(N)、変性(D)両状態の間で、転移的に構造変換することが知られている。N, D両状態の自由エネルギーを $G_N, G_D$ とあくと、天然蛋白質の構造安定性の指標は、両者の差である変性自由エネルギー $\Delta G_d = G_D - G_N$ で与えられる。<sup>4)</sup>これは、与えられた条件下で、N状態にある蛋白質をD状態に変性させる際に、外力がしなければならない仕事に相当する。この値は、蛋白質のサイズに殆ど依らず、 $40 \pm 15 \text{ kJ mol}^{-1}$ 程度である。従って1残基当たりの構造安定性への寄与は $0.4 \text{ kJ mol}^{-1}$ 程度であり、室温での熱エネルギー $2.5 \text{ kJ mol}^{-1}$ の数分の1と極めて小さい。このように天然蛋白質は全て、必要最低限の安定性しか持たないように設計されている。蛋白質分子の構造安定性が“marginal stability”と呼ばれる所以である。



**Fig.1** 3D-structural models of proteins typical of the four structural types, (a) all  $\alpha$ , (b) all  $\beta$ , (c)  $\alpha/\beta$ , (d)  $\alpha + \beta$ . Spiral ribbons, arrow-shaped ribbons, and thin strings represent  $\alpha$ -helices,  $\beta$ -strands and coil segments, respectively. (a) Sperm whale myoglobin, (all  $\alpha$ ). (b) Jack bean concanavalin A, (all  $\beta$ ). (c) Protease, subtilisin carlsberg, ( $\alpha/\beta$ , lower) combined with an inhibitor, eglin-c (upper). (d) Hydrolytic enzyme,  $\beta$ -lactamase, ( $\alpha + \beta$ ).

これは、天然蛋白質に課せられている上述の条件を満たすためには、このような熱力学的特性を持たせることが必要なためと考えられる。Fig.1に示す4種の蛋白質は、天然蛋白質の四つの構造型、(a) all  $\alpha$ , (b) all  $\beta$ , (c)  $\alpha/\beta$ , (d)  $\alpha + \beta$ , からの代表蛋白質である。これらの蛋白質の立体構造は相互に大きく異なるが、構造安定性の差は小さい。このように、蛋白質は生物機能素子として設計された工学的な存在でありながら、熱擾乱が本質的な効果を持つ、勝れて「理学的な存在」であることが分かる。

#### 4. 蛋白質分子熱力学の特徴

上述の通り、天然蛋白質の立体構造安定性は極めて低い。その結果、生理的温度から30 ~ 50 の温度上昇によって、多くの天然蛋白質は協同的に立体構造転移して変性する。

変性自由エネルギー $\Delta G_d$ は、変性エンタルピー $\Delta H_d$ , 変性エントロピー $\Delta S_d$ により、

$$\Delta G_d(T) = \Delta H_d(T) - T\Delta S_d(T) \quad (1)$$

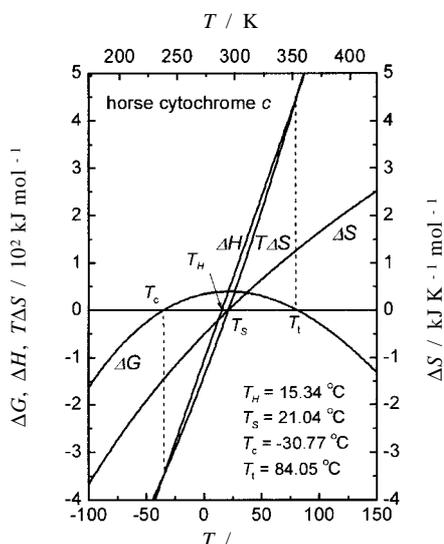
と表される。熱変性温度を $T_i$ と表すと、 $\Delta G_d(T_i) = 0$ であるから、

$$\Delta H_d(T_i) = T_i \cdot \Delta S_d(T_i) \quad (2)$$

が成り立つ。即ち当然のことながら、温度 $T = T_i$ では、変性自由エネルギーに対するエンタルピーとエントロピーの寄与が、完全に相殺する。 $\Delta G_d$ の温度微分は $-\Delta S_d$ であり、

$$\left( \frac{\Delta G_d}{T} \right)_p = -\Delta S_d \quad (3)$$

変性熱容量 $\Delta C_{p,d}$ は、



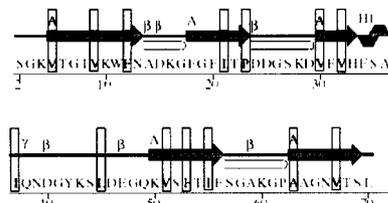
**Fig.2** Temperature dependence of the denaturation-thermodynamic quantities of an electron-transferring protein, horse heart cytochrome *c*.

$$\begin{aligned} \Delta C_{p,d} &= (\Delta H_d / T)_p \\ &= T \cdot (\Delta S_d / T)_p = -T \cdot (\Delta G_d / T^2)_p \quad (4) \end{aligned}$$

で与えられる。上式と実測の  $T_i$ ,  $\Delta H_d(T_i)$ ,  $\Delta C_{p,d}$  の値, 及び  $\Delta C_{p,d}$  = 一定の近似を用いて計算により求めた, 電子伝達系の蛋白質 ウマ cytochrome *c* の変性熱力学量の温度依存性が, Fig.2 に示されている。図から,  $\Delta H_d$ ,  $\Delta S_d$  共に室温近傍で零になることが分かる:  $\Delta H_d$  は  $T = T_H = 15.3$  で零になる。一方  $\Delta S_d$  は  $T = T_S = 21.0$  で零になり, この温度で  $\Delta G_d$  は極大に達し, 構造安定性が最高になる。また変性は大きな熱容量の上昇を伴い,  $\Delta C_{p,d}$  が大きな正値を示す。その結果,  $\Delta H_d$ ,  $\Delta S_d$  が共に大きな温度依存性を示す: 熱変性温度での  $\Delta H_d$  の値は,  $\Delta G_d$  の最大値の 10 倍以上に達する。しかし立体構造安定性  $\Delta G_d$  に対するエンタルピーとエントロピーの寄与,  $\Delta H_d$  と  $T\Delta S_d$  は大部分相殺している。

蛋白質の天然構造は, 分子内残基間に働く多様な相互作用により保たれている。蛋白質が変性すると, これらの相互作用が失われる一方, 分子内部の原子団が溶媒水中に露出して, 水和水との相互作用が生成する。生理的温度に近い室温付近で  $\Delta H_d$  が零になることは, 生理的条件下では, 変性によって失われた大きな分子内相互作用エネルギーの大部分が, 水との相互作用で相殺されることを意味している。また,  $\Delta S_d$  が零になることは, 変性によって分子鎖が解けることによる大きなコンホメーション・エントロピーの増加が, 水和水によるエントロピーの減少によって大部分打

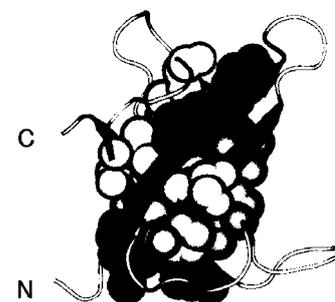
(a) SGKMTGIYK WENADKGF ITPDDGSKDV  
 FVHFSAIQND GYKSLDEGQK VSF~~T~~I~~E~~SGAK  
 GPAAGNVTSL



(b)



(c)



**Fig.3** 3D-structure of cold shock protein (csp A) and its hydrophobic cluster: (a) The underlined residues in the one-letter coded amino-acid sequence (upper) and the residues indicated with vertical bars on the secondary-structure (lower) are hydrophobic-cluster residues. (b) Ribbon model of the 3D-structure of cspA. (c) Hydrophobic cluster of cspA. Hydrophobic residues cluster inside the sandwich-like structure formed of two  $\beta$ -sheets.

ち消されることを示している。また, 温度の上昇に伴ってこの均衡が崩れ,  $\Delta H_d$ ,  $\Delta S_d$  共に大きな正の値を持つようになる。

大部分の天然蛋白質は塊状の天然構造を持ち, その分子内部には疎水性の非極性残基がクラスタを形成している。

大腸菌の cold shock 蛋白質の疎水クラスタの例を Fig.3 に示す。蛋白質分子の大きな正の変性熱容量  $\Delta C_{p,d}$  は、主に変性に伴って露出する非極性原子団の水和に起因することが分かっている。 $\Delta G_d$  の温度依存性を決めるのが  $\Delta C_{p,d}$  であるから、蛋白質の立体構造安定性を決定する上で、水が重要な役割を果たしていることが分かる。これは、立体構造安定化の物理機構を解明する上で、高精度での  $\Delta C_{p,d}$  (とその温度依存性) の測定が極めて重要であることを意味している。

これまで、蛋白質分子の立体構造安定性に寄与する因子を評価する研究が、数多く報告されてきた。その中で、最も包括的な研究は Privalov 等によるものであろう。特に、Makhatadze と Privalov による総説論文<sup>5)</sup>は、彼らの研究の総決算とも言えるものであり、蛋白質熱力学の全体像の現状を知りたい人には一読を勧めたい。Privalov とその共同研究者による精力的な研究によって、蛋白質分子の熱力学的特性の概要が解明された。然し熱力学的描像の背後にある分子機構の詳細に関しては、未だ殆ど未解明である。熱測定精度の向上と共に、他研究分野との連携が必須と思われる。

本節を終えるに当たって、立体構造安定性に寄与する因子の評価に関して付言したい：上述の通り、立体構造安定性は、大きなエンタルピー、エントロピー効果の間の小さな差に起因している。従って、多様な因子からの寄与をどのように分割するかによって、得られる描像が全く異なることが有り得ることに注意しなければならない。一般に構造形成は、相互作用の生成によって起こり、エンタルピーの低下を伴うが、同時に可動空間の減少によるエントロピーの低下も伴う。例えば、極性基間の水素結合形成に対して、蛋白質分子鎖のコンホメーション・エントロピーを独立な因子として解析すると、以下のような描像が得られる：変性状態で水和している分子鎖の極性基が、折り畳まれて分子内水素結合を形成すると、水和水が解放されてエントロピーが上昇する。従って、「立体構造安定化への水素結合の寄与は、水和水のエントロピーの上昇に起因する」という結論に導かれる。しかし分子内水素結合の形成は、蛋白質分子のコンホメーション・エントロピーの低下をもたらすので、水素結合の寄与を評価する際には、このエントロピーの寄与も同時に評価する必要があることが分かる。

##### 5. ポストゲノム時代の蛋白質分子熱力学

ゲノム科学の先駆けとしての分子生物学は、生命現象を担っている生体高分子の構造と機能の解析を中心に発展してきた。その研究スタイルは、分子情報素子としての核酸や、分子機能素子としての蛋白質などを単位として、生物を構成する部品を虱潰しに調べるといったものである。これ

は、コンピュータを知らない異星人が、空から降ってきたコンピュータを解体し、回路素子の構造と機能を解明することによりコンピュータの動作を理解しようとする姿勢に例えられよう。それに対して、ゲノム科学は、生命の設計図を解読し、その設計原理を解読しようとするものである。その研究目標は、個別の機能素子の構造・機能の解析から、分子機械システムとしての生物の理解へと広がりがつつある。配列ゲノム科学 構造ゲノム科学 機能ゲノム科学の流れは、益々加速されている。蛋白質の全基本構造 (folds) が決定されれば、任意の新規遺伝子がコードする蛋白質の立体構造は、X線結晶構造解析などを経ずに、共通の祖先を持つ相同蛋白質の立体構造情報を用いて、相同モデリングによって決定できる可能性が高い。その結果、多くの生物種について、文字配列としての遺伝情報と共に、翻訳されてできる蛋白質の全構造がデータベース化され、利用可能になるだろう。遠くない将来、塩基配列を与えた任意の遺伝子について、その産物である蛋白質の構造と機能を正しく予測することが可能になるだろう。このような状況が、ポストゲノム時代の生命科学・工学研究を取り巻く環境になると予想される。

多くの生物種について全蛋白質の構造・機能データが集積されれば、多種の相同蛋白質について、配列・構造・物性(機能を含む) 相関の包括的な解析が可能になる。その結果「比較蛋白質物性学」とも呼ぶべき研究分野が出現するだろう。その中で、蛋白質の物性を最も客観的に特徴づける分子熱力学は、重要な位置を占める筈である。また、これ迄の蛋白質分子熱力学の研究対象の多くは、水溶性で、可逆的に変性する、比較的小さな蛋白質であった。大きな蛋白質は通常複数個のドメインからなり、各ドメインが互いに独立に構造転移するとされてきた。分子量10万以上の大きな蛋白質の構造データが網羅的に集積されれば、この近似の妥当性がより広範に検定できるようになるだろう。更に、今後構造解析が進むと予想される、膜蛋白質などの非水溶性蛋白質の分子熱力学的研究は、測定法の開発と共に、それらの折り畳みの分子機構とも関連して重要になると思われる。

天然蛋白質の立体構造データと熱力学物性情報の体系的な集積は、天然蛋白質の物性を改変する分子設計に対しても、有用な情報を提供する可能性が高い。部位を指定した人工変異の導入による天然蛋白質の構造安定性、結合親和性、相互作用特異性の変化を予測することは、極めて重要である：それは、変異の導入による、(a) 溶媒水の寄与を含むエンタルピー変化の予測精度を上げると共に、(b) エントロピー変化も予測することにより、(c) 自由エネルギー変化の予測が、定量的に行えるようにすることである。これには、分子動力学模擬計算法などを含む、理論・計算科

学との連携が必要になるだろう。<sup>6)</sup>

### 6. 配列空間を広げた蛋白質科学の展開

多くの天然蛋白質は、1次構造上の部位とアミノ酸残基種をランダムに選んで残基置換を行うと、その大半が構造安定性を下げるか、影響を与えないかの何れかになる。逆に、構造安定性が著しく向上することは、殆どない。この事実は、現存する生物が持っている天然蛋白質は、構造安定性に関して局所的な最適点の近傍にあることを示している。その一方で、前述のように、天然構造の安定性は $40 \text{ kJ mol}^{-1}$ 程度であり、分子のサイズからすると極めて低い値である。人工蛋白質では $100 \text{ kJ mol}^{-1}$ 程度の安定性を持つものが作り出せることを考慮すると、天然蛋白質の安定性は、ぎりぎりの低い値を取るように意図的に設計されていると考えられる。これは多分、安定な非天然構造を生成させず、歩留まり100%の分子機能素子を作るための設計指針によるものと考えられる。上の両方の事実を考慮すると、天然蛋白質の構造は、アミノ酸配列空間において、深さの非常に浅い局所安定点の近傍にあると見なせることが分かる。塩基配列の翻訳によって生成されるアミノ酸配列が、蛋白質の構造・物性に関する全情報を担っていることは、前に述べた。これから、アミノ酸配列空間には、天然構造たり得る配列、即ち上の浅い局所安定点がどの程度あるのかが、生物の進化の機構を考える上で重要な問題になる。<sup>3)</sup>

活性部位にセリン残基を持つchymotrypsinは、241アミノ酸残基からなる蛋白分解酵素である。241アミノ酸残基からなるポリペプチド鎖が取り得る理論的配列数は $20^{241} \approx 10^{314}$ であり、実質的に無限大である。現存する生物が持つ蛋白質分子種の数=一つの生物種が持つ遺伝子の数 $\times$ 生物種の数 $\leq 10^{13}$ が、天然構造を作り得る配列数であるとすると、現存する生物は「奇跡」の産物と言うことになる。「奇跡は起こり得ない」と考えると、生物は地球環境の下で「必然」の結果として誕生したと考えるべきである。そうすると天然構造を取り得るアミノ酸配列も、実質的に無数にあると結論される。どの配列が選ばれたかという「偶然」の要素はあるが、天然蛋白質として存在し得る配列を持つポリペプチド鎖の生成は、「必然」である。そうならば、現存する天然蛋白質と同程度の構造・物性・機能を持つ蛋白質は、人知を以て充分設計可能であるという予測が成り立つ。配列空間を大きく広げた蛋白質分子について、分子熱力学の解析を行うことにより、空間の特異点としての天然蛋白質の科学を、生物進化の機構解析を俎上に載せ得る科学としての拡大できる可能性がある。このとき、配列空間を大きく広げた蛋白質分子の設計において、ポストゲノム情報が有効に活用できるだろう。

### 文 献

- 1) B. Alberts 他著、中村桂子・松原謙一監訳、「細胞の分子生物学」第4版、教育社(2004).
- 2) R. H. Pain 編、崎山文夫監訳、「タンパク質のフォールディング」、シュプリンガー・フェアラーク(2002).
- 3) C. A. Orengo, D. T. Jones & J. M. Thornton, "Bioinformatics: Genes, Proteins & Computers", BIOS Scientific, Oxford (2003).
- 4) T. E. Creighton, "Proteins: Structure and Molecular Properties", Freeman, New York (1993).
- 5) G. I. Makhatadze & P. L. Privalov, "Energetics of Protein Structure", *Adv. Prot. Chem.* **47**, 307 (1995).
- 6) 岡崎 進, 「コンピュータ・シミュレーションの基礎」、化学同人(2000).

### 要 旨

多くの生物についてゲノムDNA配列の解析が完了し、構造、機能ゲノム科学の研究が進行していることを述べる。工業的な機能素子と異なり、工学的な存在であると共に理学的な存在でもあるという、異なる二つの側面を持つ蛋白質の二重性について述べる。蛋白質の分子熱力学量の特性和その物理的及び生物学的意味を、蛋白質変性におけるエンタルピー-エントロピー相殺と大きな正の熱容量変化について説明する。最後に、ポストゲノム時代における蛋白質熱力学研究を展望し、拡大した配列空間での蛋白質科学の展開の可能性を議論する。



曾田邦嗣 Kunitsugu Soda  
 長岡技術科学大学 生物系, Dept. of Bioengineering, Nagaoka Univ. of Technology, TEL.&FAX. 0258-47-9424, e-mail: soda@vos.nagaokaut.ac.jp  
 研究分野: 分子生物物理学, 蛋白質物理学

研究題目: 蛋白質の折り畳み過程・構造安定化の分子機構の解析。分子模擬計算法と溶液X線散乱法による非天然構造の解析。

趣味: モーツァルトの音楽を聴くこと