タンパク質立体構造形成における溶媒効果

京都大学エネルギー理工学研究所 木下正弘

分子科学研究所・総合研究大学院大学 岡本祐幸, 平田文男

We have developed a new methodology for predicting tertiary structure of protein in solvent from the first principles, which combines the RISM theory for incorporating the solvent effects and a powerful method for sampling the structure space of protein. Roles of water in stabilizing protein structure and alcohol effects have been analyzed, and subjects to be treated in further studies have been discussed.

protein folding / tertiary structure / solvent effects / RISM theory / water / alcohol

1. はじめに

タンパク質は、生体内で種々の高度な機能を担っている。タンパク質の機能は、その立体構造と密接に関連している。自然界に存在するタンパク質のアミノ酸配列を人為的に改変して、新たな機能をもつ人工タンパク質を設計しようという一つのトレンドがあるが、それを試行錯誤的ではなく論理的に行うためには、少なくともアミノ酸配列と立体構造の関係が把握できている必要がある。しかし、次の二つの難問が大きく立ちはだかる：(1) タンパク質が取り得る立体構造の数が天文学的数字にのぼる（アミノ酸残基数を N とすると、少なめに見積もっても 10^N 通りはある）；(2) 立体構造形成には、溶媒環境が大きく影響し、溶媒効果をできる限り精密に取り込まねばならない。

上記二つの難問のため、タンパク質の立体構造を第一原理から（アミノ酸配列と溶媒環境のみから、経験的要素に頼ることなく）予測することは一般に不可能であるとされている。我々は、独自の複合アプローチによってそれらの解決を図り、不可能を可能にするための方法論の開発を目指している。本稿では、上記の(2)に力点を置いて、我々の方法論の概要を述べるとともに、これまでの研究で得られた主な成果を要約する。

2. 溶媒中におけるタンパク質立体構造予測のための方法論

2.1 方法論の概要

タンパク質の立体構造は、一組の主鎖（ペプチド結合の部位は平面構造を取ると仮定する）と側鎖の二面角を指定することで記述する。研究の第一段階では、タンパク質自身の構造エントロピーは除外して考えることが多い。その場合、気相中では、構造エネルギー E_C を最低にする立体構造を特定することが問題になり、徐冷モンテカルロ (MC) 法^{1), 2)}などの強力な手法が有効となる。しかし、溶媒中では事情が大きく異なる。タンパク質は、 E_C ができるだけ低い構造を取ろうとするが、溶媒は、溶媒和自由エネルギー $\Delta\mu_s$ ができるだけ低くなる構造をタンパク質に取らせようとする。ただし、溶媒和自由エネルギーは、溶液中に微量の溶質を入れた場合に生じる自由エネルギー変化の過剰量（溶質1 mol当りに換算したもの）であり、ここでは溶質が無限希釈で存在する場合を考える。溶媒中におけるタンパク質の最安定構造は、 E_C を最低にする構造ではなく、 E_C と $\Delta\mu_s$ の和（これを E_T で表し、全エネルギーと呼ぶ）を最低にする構造である。我々の方法論の概要は次の通りである。

- (1) タンパク質のある初期立体構造から出発し、徐冷MC法のアルゴリズムにしたがって、二面角を試行的に変えてみることにより、多くの立体構造を

Solvent Effects on Formation of Tertiary Structure of Protein

Masahiro KINOSHITA*, Yuko OKAMOTO** and Fumio HIRATA**

* Institute of Advanced Energy, Kyoto University

** Institute for Molecular Science, The Graduate University for Advanced Studies

サンプリングする。ただし、 E_C ではなく、 E_T を評価関数とする。

- (2) 各構造に対する $\Delta\mu_S$ は、積分方程式論の一つであるRISM理論³⁾ (分子性流体用の精密な統計力学理論)の最新型⁴¹⁻⁶⁾を用いて計算する。すなわち、各構造をもつタンパク質と平衡にある溶媒の配位・配向構造のアンサンブル平均を理論的に計算することによって $\Delta\mu_S$ を得る。MC法や分子動力学法に基づく通常のシミュレーションでは、多くの溶媒分子の配位・配向を示すスナップショットが得られる。平衡状態に達した後、膨大な数(〜アボガドロ数個)のスナップショットを平均したものが、ここで言うアンサンブル平均と考えてよい。
- (3) サンプリングした多くの立体構造のうち、 E_T を最低にする構造を選び出す。

最安定構造のみではなく、タンパク質の構造エントロピーをも考慮した平均構造を特定したい場合、徐冷MC法をマルチカノニカル法^{13,2)}などの拡張アンサンブル法に変えてサンプリングを実行する。ただし、評価関数はやはり E_T とする。我々の方法論の特徴の一つは、タンパク質と溶媒の両方を原子レベルで扱う点にある。溶媒は水でもアルコールでも何でもよい。連続体モデル(溶媒を、大きさをもつ分子の集まりとして扱わず、特定の誘電率をもつ連続媒体とみなすモデル)を用いるなど、溶媒をなおざりに扱うと失敗することを強調したい。

ここで、RISM理論についてごく簡単に言及しておく。1個の分子は、複数の原子が化学結合してできたものである。各原子核に相互作用サイトを配置し、各サイトに電荷とレナード・ジョーンズ(LJ)相互作用パラメーター(ϵ と σ)を割り当てる。分子1が N_1 個の原子、分子2が N_2 個の原子からなる場合、分子1と分子2の間の相互作用 $u(1,2)$ を

$$u(1,2) = \sum_{A=1}^{N_1} \sum_{B=1}^{N_2} v_{AB}(r_{AB}) \quad (1)$$

で与える。ただし、 v_{AB} は分子1中のサイトAと分子2中のサイトBの間の「クーロン+LJ」相互作用(通常、LJ相互作用パラメーターは、 $\epsilon_{AB} = (\epsilon_A \epsilon_B)^{1/2}$ 、 $\sigma_{AB} = (\sigma_A + \sigma_B)/2$ から計算する)、 r_{AB} はサイト間距離である。同一分子内の任意の二つのサイト間(A-A'またはB-B')距離は固定しておく。以上の相互作用サイトモデルに基づき、統計力学の基本原則にしたがって、サイトAとサイトB間の種々の相関関数を定義しつつ、それらが満たすべき関係式を近似的に導出する。こうして得られた基礎式を数値的に解くと、分子集団系

のミクロな構造やマクロな量が計算可能となる(詳細は、文献3を参照されたい)。

2.2 方法論の利点

通常のシミュレーション法では、タンパク質と溶媒分子全体を同時に扱う。その場合、必要な溶媒分子の数はきわめて多くなり、計算時間の問題が深刻化する。我々の方法論は、溶媒効果を理論的に取り込んでおり、次の利点を有する：(i) タンパク質に対してのみMC法を適用すればよく、系の自由度の数が大幅に減少している；(ii) 無限個の溶媒分子を扱える；(iii) クーロン相互作用の取扱いに何の困難も生じない。ただし、 $\Delta\mu_S$ を計算するには、非線形性の強い膨大な連立方程式(RISM方程式)を解かねばならない。従来の数値解法では、長大な計算時間を要する上、収束解が得られないこともしばしばであり、溶媒効果を理論的に取り込むことの意味がまったくなくなってしまう。我々は、従来法と比較にならないほど安定であるのみならず、実に100倍以上高速の新数値解法^{7),8)}の開発に成功した。結果として、本方法論は、通常のシミュレーション法よりも遥かに効率的なものになった。

3. 水中におけるペプチドの立体構造解析

3.1 方法論のエンケファリンへの適用

8通りの異なる初期構造から出発し、気相中($\Delta\mu_S = 0$ であり、RISM理論を適用する必要はない)、単原子溶媒(斥力相互作用系)中および水中のエンケファリンに本方法論を適用した⁵⁾。エンケファリンは、脳内ホルモンの一種として知られ、「Tyr-Gly-Gly-Phe-Met」というアミノ酸配列をもつ。ペプチドのエネルギー関数とパラメーターにはECCEP/2(コーネル大学のScheragaのグループが開発したもので、クーロン項、LJ項、水素結合項、ねじれエネルギー項からなる)に基づくものを用い、水の分子モデルにはSPC/Eを用いた。SPC/Eでは、正四面体の重心に酸素原子サイト(電荷 $= -0.8476e$ ； e は電気素量)、4つの頂点のうちの2つに水素原子サイト(電荷 $= 0.4238e$)が配置される。酸素原子サイトと水素原子サイトの距離は0.1 nmである。徐冷MC法により、気相中および単原子溶媒中では100-4000 MCスイープ、水中では100-400 MCスイープの計算を進めた。このとき、着目する二面角(エンケファリンの場合には全部で19個)を順に試行的に変えてみるのであるが、その一回りを1MCスイープと呼ぶ(エンケファリンの場合、1MCスイープ当り19通りの異なる構造をサンプリングする)。スイープ数が例えば400の場合、RISM方程式を実に7600回も解くことになる。それでも、我々の数値解法^{7),8)}を用いた

場合、一度も収束に失敗することなく、しかも高速で解を得ることができた。これは、従来の解法では到底考えられないことである。

気相中、単原子溶媒中、水中において、各々8通りの構造が最安定構造（サンプリングした構造の中では最低の E_T 値を持つが、真の最安定構造とは限らない）として得られるが、それらを主鎖構造が根平均二乗距離に関して最もよくフィットするように重ね合わせたものを図1に比較する。NMRによる実測結果も添付する。水中の場合の方がスイープ数がずっと少ない(1/10)にもかかわらず、8通りの構造の主鎖構造は遥かによく収束している。8通りの構造の E_T の値もほぼ同じである。実測構造は、「伸びた主鎖構造+比較的大きく揺らぐ側鎖構造」で特徴づけられ、本方法論で得られた結果と良好に一致する。気相中や単原子溶媒中で得られた8通りの構造はまちまちであり、 E_T の値も大きく異なっている。真の最安定構造（あるいはそれに近い構造）を探し出すには、 10^5 程度のスイープ数が必要であった。

気相中と単原子溶媒中の最安定構造はほぼ同じであるが、水中のそれは大きく異なることもわかった。溶媒としての機能は、単原子溶媒（極性の低い有機溶媒の単純なモデル）と水では大きく異なる。 $\Delta\mu_s$ が、単原子溶媒中ではペプチド構造によってあまり変化しないのに、水中では大きく変化する。なお、本方法論を水中におけるより大きなC-ペプチド（アミノ酸配列：Lys-Glu-Thr-Ala-Ala-Ala-Lys-Phe-Leu-Arg-Gln-His-Met）にも適用し、容易に最安定構造を特定できた⁶⁾ことを

付記しておく。

3.2 タンパク質立体構造形成における水の役割

エンケファリンとC-ペプチドを対象にした我々の研究により、タンパク質立体構造形成において、水が決定的な役割を果たしているものと推定できた。第一に、多くのタンパク質構造は高い $\Delta\mu_s$ の値をもち、水中では強く拒絶される。その結果、水中では、タンパク質の取り得る構造の数が激減する⁵⁾。さらに、 $\Delta\mu_s$ の値が充分低く、水に比較的受け入れられやすい構造のみを取り上げても、次のことが言える^{6), 9)}。

- (1) E_C が相対的に高い構造の多くは、以下の特徴をもつ：側鎖の同符号に帯電した部位が互いに接近している；カルボニル酸素と水の水素、アミド水素と水の酸素が静電結合している。このような場合、 $\Delta\mu_s$ は相対的に低くなる。
- (2) E_C が相対的に低い構造の多くは、以下の特徴をもつ：側鎖の同符号に帯電した部位は充分に離れている；カルボニル酸素とアミド水素の静電結合によって形成される分子内水素結合が多い。このような場合、 $\Delta\mu_s$ は相対的に高くなる。
- (3) 結局、 E_C が高い場合には $\Delta\mu_s$ が低く、 E_C が低い場合には $\Delta\mu_s$ が高くなる傾向がある。その結果、水中では全エネルギー表面がより平坦化し、ある安定構造から隣の安定構造に移るために越えねばならない障壁が低くなる。

以上により、タンパク質は、水のおかげで安定構造に素速く到達できる^{5), 6), 9)}ものと考えられる。

4. アルコール中におけるペプチドの立体構造解析

溶媒効果および二次構造形成機構の両方の解明において、非常に注目されているアルコール効果の解析結果¹⁰⁾について述べる。 β -ラクトグロブリンのいくつかのフラグメントやメリチンは、水溶液環境中では伸びた構造を取るが、アルコールを加えると α -ヘリックス構造への転移を起こすという実験事実がある。転移後の構造はアルコールの種類に依らないが、アルコールの添加効果は、アルコール分子中の炭化水素基の大きさとともに増加する¹¹⁾。我々は、RISM理論を用いてアルコール中のペプチドの立体構造解析を行い、上記の現象のメカニズムを初めて分子論的に解明した。解析では、メタノールおよびエタノール中のエンケファリン、メタノール中のC-ペプチドを取り上げた。アルコールの分子モデルにはJorgensenのOPLSを用いた。アルコール分子は水分子より大きい。水、メタノール、エタノールの数密度の比は、3.2 : 1.4 : 1.0である。以下

図1 気相中、単原子溶媒中および水中の各場合に、最安定構造として得られた8通りの構造を重ね合わせたものとNMRによる実測結果との比較

で、単にHあるいは溶媒分子中のHと書いた場合、それらは、水酸基中の水素原子を表す。

4通りの異なる構造をもつエンケファリンの、水とメタノール中における $\Delta\mu_s$ を表1に示す。“-0”は、エンケファリンのサイト電荷をすべて強制的にゼロにした場合を表す。メタノール中の $\Delta\mu_s$ は、水中のそれよりもずっと小さい。サイト電荷をゼロにしても、メタノール中の $\Delta\mu_s$ の増加量は、水中のそれよりも小さい（70%程度に過ぎない）。メタノールをエタノールに置き換えると、これらの傾向が少し強まる。

伸びた構造（Conf.4）をもつエンケファリンについて、代表的な4つの原子（図2参照）からの $\Delta\mu_s$ への寄与を表2に示す。主鎖中の典型的な親水原子からの寄与は、水中では負になるが、メタノール中ではより高くなり、エタノール中ではさらに高くなる。酸素原子Oより露出度の低い窒素原子Nからの寄与は、アルコール中では正になる。ペプチド中の原子Aの周りの、溶媒分子中の原子Bの配位数 N_B は、

$$N_B = 4\pi\rho_B \int_0^R r^2 g_{AB}(r) dr \quad (2)$$

から計算できる（ $g_{AB}(r)$ はペア分布関数、 ρ_B はBの数密度、 R は $g_{AB}(r)$ の第一最小位置）。A=O, N；B=H

表1 エンケファリンの4通りの構造（Conf.1-Conf.4）についての水およびメタノール中の溶媒和自由エネルギー（kcal/mol）

Conf.	H ₂ O	H ₂ O-0	CH ₃ OH	CH ₃ OH-0
1	197	216	73	87
2	178	209	58	82
3	203	229	75	93
4	177	201	59	77

“-0”は、エンケファリンのサイト電荷を強制的にすべてゼロにした場合を表す。

図2 伸びた構造（Conf.4）をもつエンケファリン
“23 N”、“43 C”、“47 C”、“56 O”は、各々、Gly-2中の“23 N”、Phe-4中の“43 CD1”、Phe-4中の“47 CZ”、Phe-4中の“56 O”を表す。

の場合の計算結果を表3に示す。“(Ratio)”は、メタノールの場合の値で割ったものを表す。まず、 $N_B(\text{H}_2\text{O}) > N_B(\text{CH}_3\text{OH}) > N_B(\text{C}_2\text{H}_5\text{OH})$ が成立する。これは、 $\rho_B(\text{H}_2\text{O}) > \rho_B(\text{CH}_3\text{OH}) > \rho_B(\text{C}_2\text{H}_5\text{OH})$ であることに起因する。また、 $N_B(A=N) < N_B(A=O)$ が成立する。これは、Nの露出度がOのそれよりも低いからである。 $N_B(\text{Ratio})$ の溶媒間の差は、 $A=N$ の場合の方が $A=O$ の場合よりも大きい。これは、アルコール分子中の炭化水素基の立体障害により、溶媒分子中のHとペプチド分子中のNまたはOとの静電結合が妨げられる効果が、露出度の低いNに対して大きく出るためである。この効果は、エタノールの場合のほうが大きい。以上の結果は、表2にも反映されている。また、表2における側鎖中の典型的な疎水原子からの寄与は、水中では正であるが、メタノール中ではより低くなり、エタノール中ではさらに低くなる。これは、アルコール分子が水分子より大きく（アルコールの数密度<水の数密度）、キャビティー生成に要する仕事はアルコールの場合のほうが小さくなるためである。また、アルコール分子中には、アルコール分子間の水素結合に参加できない炭化水素基があり、これとペプチド中の疎水原子とのコンタクトが安定化されるためである。これらの効果は、エタノールの場合のほうが大きい。実際、 $A=\text{"47 CZ Phe-4"}$ の周りの $B=\text{CH}_3, \text{CH}_2$ の配位数 N_B を計算すると、 $\rho_B(\text{CH}_3\text{OH}) > \rho_B(\text{C}_2\text{H}_5\text{OH})$ であるのに、 $N_B(B=\text{CH}_3 \text{ in } \text{CH}_3\text{OH}) < N_B(B=\text{CH}_3 \text{ in } \text{C}_2\text{H}_5\text{OH}) < N_B(B=\text{CH}_2 \text{ in } \text{C}_2\text{H}_5\text{OH})$ となった。

表2 構造4（Conf.4）のエンケファリンの水、メタノールおよびエタノール中での溶媒和自由エネルギーに対する代表的な原子からの寄与（kcal/mol）

Atom		H ₂ O	CH ₃ OH	C ₂ H ₅ OH
23 N	Gly-2	-1.62	0.96	1.41
56 O	Phe-4	-5.98	-3.56	-3.21
43 CD1	Phe-4	1.91	-0.23	-0.76
47 CZ	Phe-4	2.21	-0.35	-1.00

上から順に、アミド窒素、カルボニル酸素（以上は親水原子）、フェニルアラニン側鎖中の二つの炭素（これらは疎水原子）を表す。

表3 構造4（Conf.4）のエンケファリンにおける原子Aの周りの溶媒水素原子Bの配位数 N_B

A	B	N_B	$N_B(\text{Ratio})$
23 N Gly-2	H in H ₂ O	0.196	2.20
	H in CH ₃ OH	0.089	1.00
	H in C ₂ H ₅ OH	0.072	0.81
56 O Phe-4	H in H ₂ O	0.883	1.64
	H in CH ₃ OH	0.538	1.00
	H in C ₂ H ₅ OH	0.463	0.86

エンケファリンの伸びた構造 (Conf.4; 分子内水素結合をもたない) の E_T は、真空中の最安定構造 (Conf.1; 3つの分子内水素結合をもつ) の E_T よりも、水中では7 kcal/mol 低い。しかし、メタノール中では1 kcal/mol 低くなるに過ぎない。エタノール中では、2つの E_T はほぼ同じ値になる。C-ペプチドの伸びた構造 (分子内水素結合をもたない) の E_T は、水中では α -ヘリックス構造の E_T より87 kcal/mol も低いが、メタノール中では15 kcal/mol 低くなるに過ぎない。分子内水素結合を有する構造は、アルコール中では水中における場合よりもずっと安定化されることがわかる。

ペプチド中の疎溶媒原子の「疎溶媒性」は、アルコール中では水中よりも低くなる。親溶媒原子の「親溶媒性」は、アルコール中では水中よりも低くなる (原子の露出度が低い場合に特にそうである)。アルコール中では、疎溶媒原子が露出するペプチド構造が許されやすくなり、親溶媒原子が露出する構造を取ることはそれほど重要でなくなる。ペプチドは、構造エネルギー的に最も安定化できる構造を自ら取ろうとする。この傾向は、アルコール分子内の炭化水素基の大きさとともに強まる。 α -ヘリックスは、ペプチド自身が、分子内水素結合で最も安定化できる構造の典型例であり、アルコール中ではペプチドはこの構造を取りやすい (ただし、この構造はアルコールの種類に依らない)。

5. 今後の展望

現在のところ、RISM理論と徐冷MC法およびマルチカノニカル法との統合を終了している^{5), 6), 12)}。次の段階として、最も有力な拡張アンサンブル法の一つと考えられるレプリカ交換法¹³⁾のMC法バージョンとRISM理論を統合する。RISM理論を用いた場合、塩 (NaCl など) の効果の解析¹⁴⁾⁻¹⁶⁾も可能になる。その場合、水をも原子レベルで扱うことが重要になる。無論、RISM理論には近似が入っているが、理論自体の改良も進んでおり、数値解法のさらなる高速化と併せて、今後の発展が大いに期待される。しかも、最近の計算機能力の進展には目を見張るものがあり、ここ数年のうちに、 $N=100$ 程度のタンパク質の水溶液中における立体構造

が、上記の統合を行った本方法論を用いて第一原理から特定できるようになると期待される。

さらに、本方法論は、溶媒中において両親媒性分子が自己組織化によって形成する、ミセルなどの分子集合体にも拡張されており¹⁷⁾、それらの形態やサイズ分布を第一原理から予測することも可能になりつつある。

文 献

- 1) 岡本祐幸 (1998) 物性研究 **70**, 719.
- 2) Hansman, U.H.E. and Okamoto, Y. (1999) *Curr. Opin. Struct. Biol.* **9**, 177.
- 3) 平田文男 (1995) 化学総説 No.25, 溶液の分子論的描像, pp.147-165, 日本化学会編, 東京.
- 4) Kinoshita, M., Okamoto, Y. and Hirata, F. (1997) *J. Chem. Phys.* **107**, 1586.
- 5) Kinoshita, M., Okamoto, Y. and Hirata, F. (1998) *J. Am. Chem. Soc.* **120**, 1855.
- 6) Kinoshita, M., Okamoto, Y. and Hirata, F. (1999) *J. Chem. Phys.* **110**, 4090.
- 7) Kinoshita, M., Okamoto, Y. and Hirata, F. (1997) *J. Comput. Chem.* **18**, 1320.
- 8) Kinoshita, M., Okamoto, Y. and Hirata, F. (1998) *J. Comput. Chem.* **19**, 1724.
- 9) Kinoshita, M., Okamoto, Y. and Hirata, F. (2000) *J. Mol. Liq.* in press.
- 10) Kinoshita, M., Okamoto, Y. and Hirata, F. (2000) *J. Am. Chem. Soc.* **122**, 2773.
- 11) Hirota, N., Mizuno, K. and Goto, Y. (1998) *J. Mol. Biol.* **275**, 365.
- 12) Mitsutake, A., Kinoshita, M., Okamoto, Y. and Hirata, F. (2000) *Chem. Phys. Lett.* **329**, 295.
- 13) Sugita, Y. and Okamoto, Y. (1999) *Chem. Phys. Lett.* **314**, 141.
- 14) Kinoshita, M. and Hirata, F. (1997) *J. Chem. Phys.* **106**, 5202.
- 15) Kinoshita, M., Okamoto, Y. and Hirata, F. (1998) *Chem. Phys. Lett.* **297**, 433.
- 16) Imai, T., Kinoshita, M. and Hirata, F. (2000) *Bull. Chem. Soc. Jpn.* **73**, 1113.
- 17) Kinoshita, M. and Sugai, Y. (1999) *Chem. Phys. Lett.* **313**, 685.