

モンテカルロシミュレーションで探るタンパク質の折り畳み機構

岡崎国立共同研究機構分子科学研究所理論研究系
総合研究大学院大学数物科学研究科機能分子科学専攻
URL: <http://konf2.ims.ac.jp/>
岡本祐幸

(1998年7月10日受理)

1. はじめに

タンパク質分子の立体構造とその生化学的機能の間には密接な関係がある。よって、タンパク質の立体構造の研究は、その機能発現の仕組みを解明することを目指しており、究極的には新薬品や特定の機能を持った人工タンパク質の開発ばかりでなく、狂牛病やアルツハイマー病などのタンパク質の誤った折り畳みに起因する病理の発現機構の解明へと直接つながっている。

タンパク質の自然の立体構造（自由エネルギーが最小の状態）をそのアミノ酸配列及びタンパク質分子を取り巻く環境因子の情報のみから予測することを、第一原理からの立体構造予測と呼ぶ。この問題はタンパク質が細胞内で合成される事実を考えると、環境因子が複雑すぎて、一見、理論家の挑戦を絶望的に退けているように思われる。しかし、1960年代初頭の所謂 Anfinsen の実験 [1] は、試験管内でタンパク質の変性状態から自然の立体構造を再構築するものであり、細胞内のもろもろの複雑な環境因子を忘れて良いことを示唆した。よって、この実験は、計算機シミュレーションによってタンパク質の立体構造予測が可能であるという希望を多くの理論家に与えたが、第一原理からの構造予測問題は現在に至るまで、依然未解決な難問である。

一方、タンパク質の立体構造がどのような物理的原理や機構によって構築されるかを理解しようとする問題はタンパク質の折り畳み問題 (protein folding problem) と呼ばれ、これも Anfinsen の実験以来、多くの実験家及び理論家を魅了してきた。この問題はタンパク質を自由に折り畳んだり変成したりできる実験家が長い間研究をリードしてきたが、最近やっと理論家もこの問題に取り組み始めている（初期の理論的考察としては、例えば、文献 [2] がある）。タンパク質の折り畳み問題を理解するには、タンパク質の自然の立体構造ばかりでなく、変性状態や折り畳みの中間状態等、配位空間全体の情報が必要となる。よって、本来、理論家にとっては、まず、第一原理からの立体構造予測問題を片づけてから、その延長上で、折り畳み機構を解明しようとするのが自然であるが、第一原

* 本稿は、編集部の方から特にお願いして執筆していただいた記事である。

理からの構造予測が不可能であると多くの研究者が信じてしまっている現在、むしろ、多くの理論家がまず折り畳み問題に挑戦しつつあるというのが、世界的な理論的研究の動向の一つといえるであろう。

タンパク質の理論的研究の方法は主に三種類に分類できる。一つは「物理学的手法」で、格子模型で代表されるように、エネルギー関数をできるだけ簡略化した、所謂「最小模型」に基づいた理論である。この手法ではオングストローム単位の精度の立体構造予測は不可能であるが、折り畳み問題の研究の最近の発展はこのアプローチによるところが大きい。(日本語の解説としては、例えば、文献 [3] を参照。) 二つ目は「分子科学的手法」で、全原子模型に基づいて、ミクロの詳細を取り入れたエネルギー関数によるものである。この方法は要求される計算時間が膨大である難点を除けば、精度の良い立体構造予測と折り畳み問題の研究の両方を扱える、オーソドックスな手法である。(日本語の解説としては、例えば、文献 [4] を参照。) 三つ目は「博物学的手法」で、X線回折実験等で決定された、タンパク質の立体構造のデータベースを基にして、未知の構造をアミノ酸配列の情報から予測しようというものである。この手法は系のエネルギー関数に依っていないため、折り畳み問題の研究には適さないが、現在のところ、アミノ酸数が百以上のタンパク質の立体構造予測がある程度の確率で成功する数少ない方法といえることができる。(日本語の解説としては、例えば、文献 [5] を参照。)

タンパク質の系のエネルギー関数を与え、その立体構造を分子動力学法やモンテカルロ法による計算機シミュレーションで予測しようとするとき、我々は大きな困難に突き当たる。それは、系の自由度が多いことに起因する。特に、1個のタンパク質分子につき、考慮しなければならない溶媒分子の個数が膨大であるので、溶媒の寄与を精度良く見積もるのに多くの計算時間を要することになる。また、たとえ溶媒を忘れて、タンパク質分子1個に注目したとしても、エネルギー極小状態に対応する準安定構造が無数に存在する。よって、従来の手法に依ってはいは、シミュレーションが、それらエネルギー極小状態に留まってしまって、最小状態に対応する自然の構造に到達するのが至難の業となる。

上の困難を克服するための我々の戦略は二つの要素から成る。一つはタンパク質の系の正確なエネルギー関数を定義することであり、もう一つはシミュレーションがエネルギー極小状態に留まらない強力な手法を導入することである。本総説では、分子科学的手法に基づいて、タンパク質の第一原理からの立体構造予測問題及び折り畳み問題について議論する。

2. タンパク質の系のエネルギー関数

ここでは、タンパク質の系として、タンパク質分子1分子とそれを取り巻く多数の溶媒分子の系を考える。系の全エネルギー関数 E_{tot} としては、タンパク質分子の構造エネルギー E_P と溶媒和の自由エネルギー E_S の和で与えることにする（単位は kcal/mol）。

$$E_{tot} = E_P + E_S \quad (1)$$

ここで、溶媒の効果については、溶媒のいろいろな状態からの寄与をまとめて取り入れるという意味で、自由エネルギーをとった。タンパク質分子の構造エネルギー E_P は、次のように、静電相互作用項 E_C 、12-6 レナード・ジョーンズ項 E_{LJ} 、水素結合項 E_{HB} の分子内の全ての原子対についての和に、全てのボンドの周りの回転角についての和である、ねじれエネルギー項 E_{tor} を足したもので与えられる。

$$\begin{aligned} E_P &= E_C + E_{LJ} + E_{HB} + E_{tor} , \\ E_C &= \sum_{(i,j)} \frac{332 q_i q_j}{\epsilon r_{ij}} , \\ E_{LJ} &= \sum_{(i,j)} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) , \\ E_{HB} &= \sum_{(i,j)} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) , \\ E_{tor} &= \sum_i U_i \left(1 \pm \cos(n_i \chi^i) \right) . \end{aligned} \quad (2)$$

ここで、 r_{ij} は i 番目の原子と j 番目の原子の間の距離（単位は Å）である。また、 ϵ は誘電率、 χ^i はボンド i における二面角である。各原子ではその重心に相互作用の中心があるとし、部分電荷 q_i （単位は電子電荷）がそこに集中しているとする。更に、 E_C 中の 332 という因子はエネルギーを kcal/mol の単位で表すための係数である。上のエネルギー関数内の諸々のパラメータ及び分子の幾何情報については、標準的なエネルギー関数の一つである、ECEPP/2 [6] のものを使った。気相中のシミュレーションでは誘電率 ϵ を 2 とし、 $E_S = 0$ とした。一回の Monte Carlo (MC) Sweep において、全ての回転角を一回ずつ変更し、変更の度ごとにメトロポリス判定 [7] を課した。

溶媒の寄与については、精度のレベルによって、いろいろな取り入れ方がある。例えば、最も単純なものでは、距離に依存した誘電率 $\epsilon(r)$ を構造エネルギーの静電相互作用項 E_C で使う方法がある（レベル 1 の溶媒効果）。ここでは、文献 [8] で使われた S 字型関数

による結果を紹介する。この関数は距離 r の関数として、以下のような振る舞いをする。

$$\begin{cases} \epsilon(r) \rightarrow 2, & \text{as } r \rightarrow 0, \\ \epsilon(r) \rightarrow 78, & \text{as } r \rightarrow \infty. \end{cases} \quad (3)$$

つまり、距離がゼロに近づく極限で、タンパク質内部における誘電率 (2) に近づき、距離が無限大に近づく極限で、バルク水の誘電率 (78) に近づく関数である。この近似は溶媒による静電相互作用のみを取り入れていて、他の重要な寄与 (例えば、疎水効果等) を無視している。二つ目の例としては、タンパク質分子の溶媒接触表面積に比例した項を溶媒和自由エネルギー E_S として、以下のように導入するやり方がある (レベル 2 の溶媒効果)。

$$E_S = \sum_i \sigma_i A_i. \quad (4)$$

ここで、 A_i は i 番目の種類の原子の溶媒接触表面積であり、 σ_i は対応する比例定数である。この比例定数は小分子における実験データをもとに決定されている。この方法は溶媒の効果を全て含んでいて、上の距離に依存した誘電率による方法よりは精度が高い。我々は提案されている 5 種類のパラメーターセットの是非を考察し、文献 [9] のものが比較的良い近似を与えると結論した。[10] 最後に、最も厳密な溶媒の寄与の取り入れ方としては、水分子を陽にシミュレーションに取り込む方法が広く使われているが、本稿では、液体の統計力学に基づいた RISM 理論 [11] による結果を紹介する (レベル 3 の溶媒効果)。

3. シミュレーション手法

タンパク質の系のエネルギー関数の準備の後には、エネルギー極小状態に留まらないシミュレーション手法の幾つかを紹介しよう。我々は徐冷モンテカルロ法 (Monte Carlo simulated annealing) [12] 及び拡張アンサンブル法 (generalized-ensemble algorithms) [13]-[16] の適用を提案してきた。

3.1 徐冷法

絶対温度 T における従来のカノニカルアンサンブルに基づく統計力学系では、エネルギー E を持つ各状態の出現確率は以下のボルツマン因子に比例する。

$$W_B(E) = \exp(-\beta E). \quad (5)$$

ここで、 $\beta \equiv 1/k_B T$ はボルツマン定数と温度の積の逆数である。すると、エネルギーの確率分布は次で与えられる。

$$P_B(T, E) \propto n(E)W_B(E). \quad (6)$$

ここで、 $n(E)$ はエネルギー E を持つ状態の数 (状態密度) である。状態数 $n(E)$ は E の増加関数であるが、ボルツマン因子 $W_B(E)$ は E と共に指数関数的に減少するから、確率分布 $P_B(T, E)$ は一般にベル型の形をしている。高温では β が小さいので、 $W_B(E)$ は E と共にゆっくり減少する。よって、 $P_B(T, E)$ は幅広いベル型をしている。一方、低温では β が大きく、 $W_B(E)$ が E と共に急速に減少する。よって、 $P_B(T, E)$ は幅の狭いベル型をしている。そして、 $T \rightarrow 0$ K の極限で、 $P_B(T, E)$ は $\delta(E - E_{GS})$ に比例する。ここで、 E_{GS} は最小エネルギー値である。しかし、従来のシミュレーション法では低温におけるカノニカル分布を得るのは非常に困難である。これは、低温では熱的揺らぎが小さいためにシミュレーションがエネルギー極小状態に留まってしまうからである。

徐冷法は物質の結晶を作る過程に基づいている。すなわち、(融解温度より) 十分高温でシミュレーションを始め、シミュレーションを通じて徐々に温度を下げて行き、最終的に最小エネルギー状態 (結晶) に到達しようとする方法である。もし、温度の下げ方が熱平衡を保つ程緩やかならば、(最終温度が 0 K の時) 最小エネルギー状態のみが得られる。しかし、急冷すれば、初期状態の近傍のエネルギー極小状態に留まってしまう訳である。

徐冷法はいろいろな複雑系の問題で広く使われているが、タンパク質の立体構造予測問題へは約 10 年前に幾つかのグループによって提唱された。[17]-[20] 以後、タンパク質のフォールディング問題や NMR・X 線実験等における構造精密化問題の標準的手法の一つになっている (解説としては、例えば文献 [21] を参照)。我々の研究グループは徐冷法の有効性を主に小ペプチド系で確かめてきた。我々の方法では初期構造は乱数で各二面角をランダムに準備したものを使う。徐冷の仕方としては初期温度 $T_I = 1000$ K から最終温度 $T_F = 250$ K (場合によっては $T_F = 100, 50$, 及び 1 K も使った) まで、指数関数的に下げるものを採用した。[22] すなわち、 n 番目の MC Sweep での温度は次で与えられる。

$$T_n = T_I \gamma^{n-1}. \quad (7)$$

ここで、 γ は T_I , T_F , 及び一つの試行の全 MC sweep 数で決まる定数である。また、一回の試行は $10^4 \sim 10^6$ 回の MC sweep から成り、異なった初期構造からの試行を普通 10 回から 20 回繰り返した。

3.2 拡張アンサンブル法

拡張アンサンブル法によるモンテカルロ法は、エネルギー空間上の一次元酔歩を実現する手法である。よって、シミュレーションがどんなエネルギー障害も乗り越えることができ、エネルギー極小状態に留まるのを避けることができる。

マルチカノニカル法 (multicanonical algorithm) [13] は拡張アンサンブル法でも最も知られた方法の一つであろう。マルチカノニカルアンサンブルでは、エネルギーの確率分布が次で定義される。

$$P_{mu}(E) \propto n(E)W_{mu}(E) \equiv \text{constant} . \quad (8)$$

よって、マルチカノニカル法の重み因子はボルツマン因子ではなく、次で与えられる。

$$W_{mu}(E) \propto n^{-1}(E) . \quad (9)$$

一般に、この重み因子は前もって分っていないので、数回の試行モンテカルロシミュレーションの繰り返しで、逐次的に求めなければならない (求め方の詳細については、例えば文献 [23] 参照)。重み因子 $W_{mu}(E)$ が求まれば、それを使って、長いシミュレーションを一回だけ実行すれば良い。マルチカノニカル法の利点はこの一回のシミュレーションの結果から、最小エネルギー状態ばかりでなく、任意の温度 $T(= 1/k_B\beta)$ におけるカノニカル分布 $P_B(T, E)$ が得られることである。後者については再重法 (ヒストグラム法) [24] を用いる。すなわち、 $P_B(T, E)$ は重み因子 $W_{mu}(E)$ と最終的に得られたエネルギー分布 $P_{mu}(E)$ から次の式で与えられる。

$$P_B(T, E) = \frac{P_{mu}(E) W_{mu}^{-1}(E) e^{-\beta E}}{\int dE' P_{mu}(E') W_{mu}^{-1}(E') e^{-\beta E'}} . \quad (10)$$

すると、物理量 A の温度 T における期待値は以下で得られる。

$$\langle A \rangle_T = \int dE A(E) P_B(T, E) . \quad (11)$$

マルチカノニカル法及びその類似法のタンパク質の立体構造予測問題への適用は数年前に提唱された。[25] それ以後、いろいろな応用が試みられた (例えば、文献 [26]–[28] 参照)。また、マルチカノニカル法の分子動力学法版も開発された。[29, 30]

マルチカノニカル法ではエネルギー空間上の自由酔歩を実現するが、焼き戻し法 (simulated tempering) [14] では、温度空間上の自由酔歩を、また、1/k 法 [15] ではエントロピー空間上の自由酔歩をそれぞれ実現する。これらの拡張アンサンブル法では同時にエネルギー空間上の (重みのついた) 酔歩も実現するため、シミュレーションがエネルギー極小状態から逃れることができる。これら三つの手法のタンパク質の問題における有効性の比較が最近なされた。[31]

上に述べたように、拡張アンサンブル法における重み因子は、短いシミュレーションによって逐次的に決めなければならないが、系によっては、なかなか重み因子の決定が困難である場合がある。よって、簡単な重み因子に基づく手法が望まれる。我々は以下のような単純な重み因子に基づいた拡張アンサンブル法を開発した。[16]

$$W(E) \propto \left(1 + \beta \frac{E - E_0}{n_F}\right)^{-n_F} \quad (12)$$

ここで、 E_0 は最小エネルギー値の推定値、 n_F は系の自由度の数である（タンパク質のモンテカルロシミュレーションでは、 n_F は回転角の数である）。

タンパク質の問題へのこれらの拡張アンサンブル法の適用においては、我々はまず短いシミュレーションで重み因子を逐次的に求めた。そして、その後ランダムな初期構造から始めて、MC sweep が 200,000 回から 1,000,000 回の長い試行を一回だけ実行した。

4. 結果

ここで、徐冷モンテカルロ法及び拡張アンサンブル法による、筆者等の計算結果を紹介する。全てのシミュレーションはランダムな初期構造から始められた（すなわち、初期構造は変数である各二面角を乱数で完全にランダムに準備したものを使う）。

4.1 第一原理からのタンパク質の立体構造予測問題

最初の例は Met-enkephalin というペプチドである。このペプチドは 5 個のアミノ酸からなり、Tyr-Gly-Gly-Phe-Met というアミノ酸配列を持つ。このペプチドの ECEPP/2 エネルギー関数に基づいた気相中 ($\epsilon = 2$) における最小エネルギー構造はいろいろな方法から得られている。[32] そして、このペプチドは ECEPP/2 エネルギーで $E_P \leq -11$ kcal/mol の領域で、最小エネルギー構造をしていることが分っている。[33, 25] 図 1 にこの最小エネルギー構造を示す。

図 2 には従来のカノニカルアンサンブル上のモンテカルロシミュレーションの結果として、全構造エネルギー E_P (式 (1) 参照) を MC Sweep の関数として表したものを示す。図 2(a), (b), (c) はそれぞれ $T = 1000, 300, 50$ K におけるものである。図 2(c) の $T = 50$ K における試行では、熱的揺らぎがとて小小さく、このシミュレーションがエネルギー値が ~ -7 kcal/mol の極小状態に引っ掛かってしまったことが分る（なんとすれば、以下の図 5 にも示したように、50 K での平均エネルギーは ~ -11 kcal/mol であるからである [25]) 一方、図 3 には徐冷モンテカルロ法によるシミュレーションの同様の結果を示した。この試行では温度がシミュレーションを通じて、1000 K から 50 K まで下げら

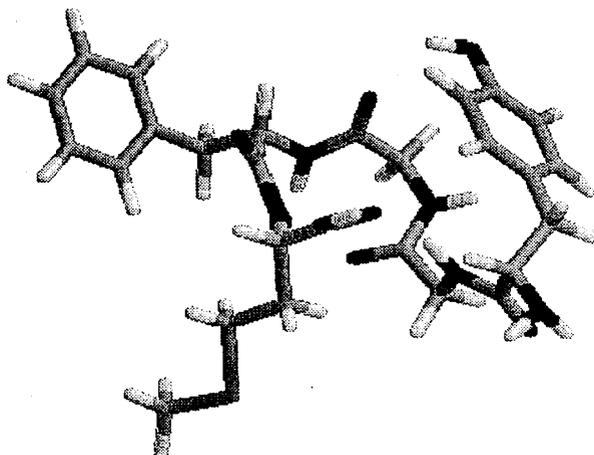


Figure 1: エンケファリンの気相中における最小エネルギー構造

れたが、最小エネルギー領域 ($E_P \leq -11$ kcal/mol) に到達している。更には、図4にマルチカノニカル法によるシミュレーションの同様の結果を示した。この試行は最小エネルギー領域と高エネルギー領域の間を実際に酔歩していることが良く分る。他の拡張アンサンブル法による結果もエネルギー空間上の類似の酔歩を実現する。[31, 16]

拡張アンサンブル法では、式(10)と(11)の再重法を使って、唯一回のシミュレーションの結果から、いろいろな熱力学量を温度の関数として求めることができる。その例として、1,000,000回のMC sweepのシミュレーションから得られた、エンケファリンの系の平均エネルギーと比熱を温度の関数として表したものを図5に示す。ここで、比熱は次の式で定義される ($N (=5)$ はこのペプチドのアミノ酸数である)。

$$C \equiv \frac{1}{N k_B} \frac{d(\langle E_P \rangle_T)}{dT} = \beta^2 \frac{\langle E_P^2 \rangle_T - \langle E_P \rangle_T^2}{N}, \quad (13)$$

図5は上述の4つの拡張アンサンブル法の結果[31, 16]を重ね描きしたものであるが、良い一致が得られている。

図6には構造エネルギーの各成分の平均値を温度 T (K) の関数として表したものを示す(式(2)参照)。この図から分ることは、各項の中では、レナード・ジョーンズ項が温度とともに最も大きく変化することであり、気相中のエンケファリンにおいては、このエネルギー項が折り畳みに一番効いていることである。

上に見たように、エンケファリンの気相中のシミュレーションで得られる最小エネルギー構造は β -ターンに折り畳まれている(図1参照)が、水中のNMR実験[34]の結果

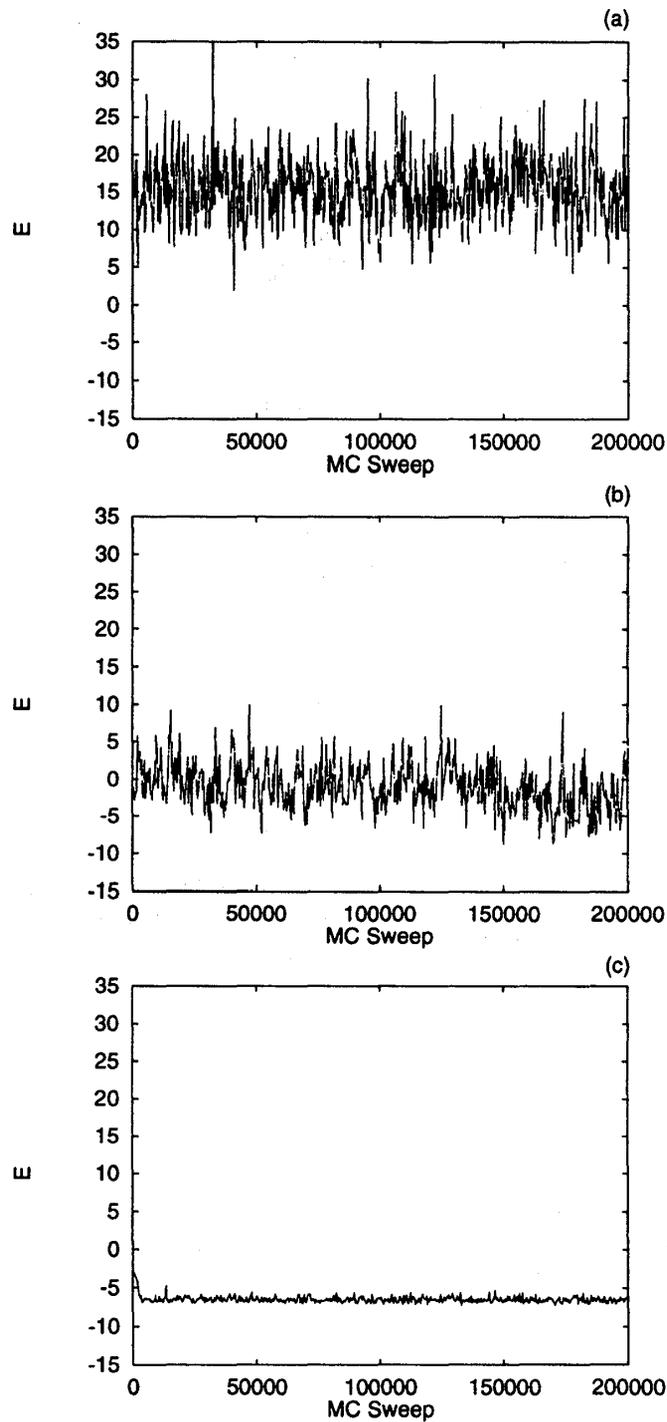


Figure 2: 従来のカノニカルアンサンブル上のモンテカルロシミュレーションにおけるエンケファリンの構造エネルギー E_P (kcal/mol) の「時間発展」(a) $T = 1000$ K, (b) 300 K, (c) 50 K.

は伸びた構造を示唆している。ここでは、RISM 理論による、最も厳密な溶媒効果（レベ

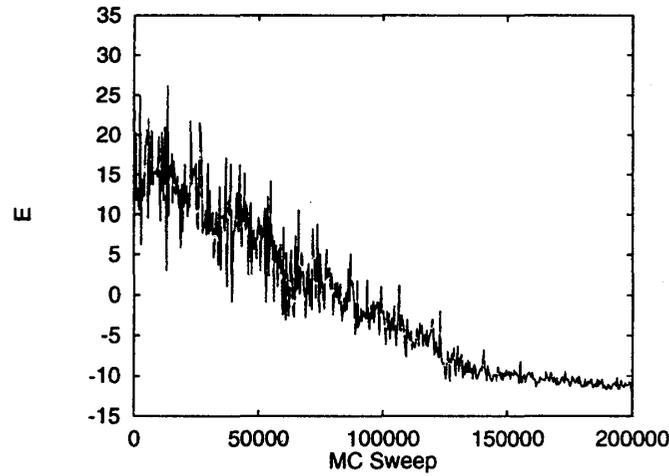


Figure 3: 徐冷モンテカルロシミュレーションにおけるエンケファリンの構造エネルギー E_P (kcal/mol) の「時間発展」 (200,000 回の MC Sweep の間に温度は $T_I = 1,000$ K から $T_F = 50$ K まで下げられた)。

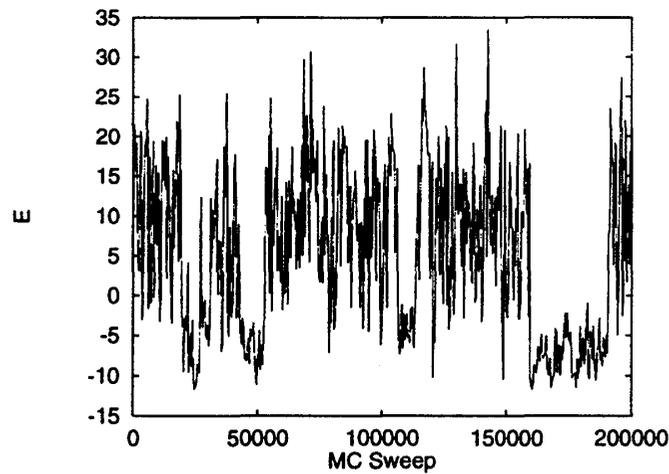


Figure 4: マルチカノニカルモンテカルロシミュレーションにおけるエンケファリンの構造エネルギー E_P (kcal/mol) の「時間発展」。

ル 3) を取り入れた計算結果を紹介する。 [35]

まず、表 1 に溶媒和自由エネルギー差 ΔE_S , 構造エネルギー差 ΔE_P , 及び全エネルギー差 $\Delta E_{tot} = \Delta E_P + \Delta E_S$ を 4 つの準備された構造について示した。 [35] 構造 1 は気相中の最小エネルギー構造, 構造 2 と 3 は他のコンパクトな構造, そして構造 4 は伸びた構造である。これらの数値は構造 1 を基準とした相対値を表す。この表から分るように, 伸びた構造 (構造 4) において溶媒和自由エネルギーが最小であると共に, 全エネルギー

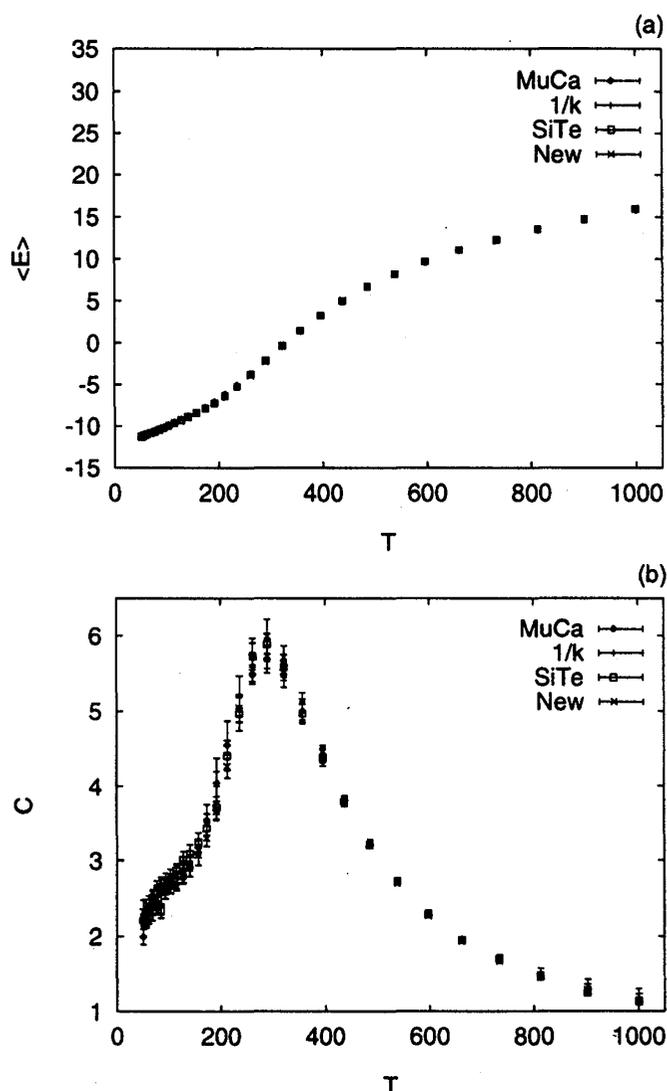


Figure 5: Met-enkephalin の全構造エネルギー E_P の平均値 (kcal/mol) (a) 及び比熱 C (b) を温度 T (K) の関数として表したもの. これらは 4 つの異なる拡張アンサンブル法の結果を重ね合わせたものである.

も最小であり, NMR 実験と一致している. [35]

この結果を更に確認するため, 我々は実際に RISM で溶媒効果を取り入れて, 徐冷シミュレーションを行った. [35] 図 7 に NMR 実験の結果から示唆された 5 つの構造を重ね合わせたもの (図 7(a)) と, 8 回の独立な徐冷法で得られた構造を重ね合わせたもの (図 7(b)) とを比較した. 両者が非常に似ていることは自明であろう. また, この結果は溶媒効果を取り入れると, 一定構造への収束が加速されることも示唆している. [35]

このように, レベル 3 の RISM 理論に基づけば厳密な溶媒効果を取り入れられるが,

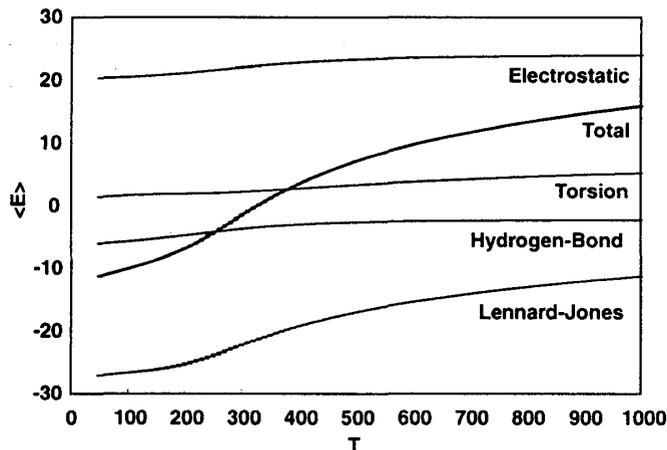


Figure 6: Met-enkephalin の構造エネルギーの各成分の平均値 (kcal/mol).

Table 1: Met-enkephalin の準備された 4 つの構造の溶媒和自由エネルギー差 ΔE_S , 構造エネルギー差 ΔE_P , 及び全エネルギー差 $\Delta E_{tot} = \Delta E_P + \Delta E_S$ (全て kcal/mol). 溶媒接触表面積 A (\AA^2), 溶媒排除体積 V (\AA^3), 及び根平均 2 乗距離 (\AA) も与えた. 溶媒和自由エネルギーの計算は RISM による.

Conf.	A	V	ΔE_S	ΔE_P	ΔE_{tot}	rms
1	738	1366	0	0	0	0
2	779	1413	-18.8	24.2	5.4	2.9
3	708	1349	5.9	9.5	15.4	2.2
4	849	1456	-20.0	12.8	-7.2	3.2

レベル 1 や 2 の溶媒効果に比べて, 計算時間が格段に多くかかってしまう難点がある. この意味で RISM 計算の高速化は重要な問題である. (その一つの試みとしては, 文献 [36] を参照.) 以下では, 主にレベル 1 や 2 の溶媒効果を取り入れたシミュレーションの結果を紹介する.

次の例は ribonuclease A の C-peptide である. このペプチドは 13 個のアミノ酸からなるが, 酵素全体の X 線回折実験データから, Ala-4 から Gln-11 までが約 3 週の α -ヘリックス構造を持つことが知られている. [37] また, 孤立した C-peptide においても, NMR 実験 [38] 等で, α -ヘリックスの形成が観測されている. 我々はまず, 気相中の徐冷シミュレーションによって得られた最小エネルギー構造が, X 線と同じ場所に α -ヘリックスを持つことを示した. [22] しかし, この計算では側鎖の構造までの一致は得られなかった. 特に, 酵素全体の X 線実験及び孤立した C-peptide の NMR 実験の両方で確認されていた,

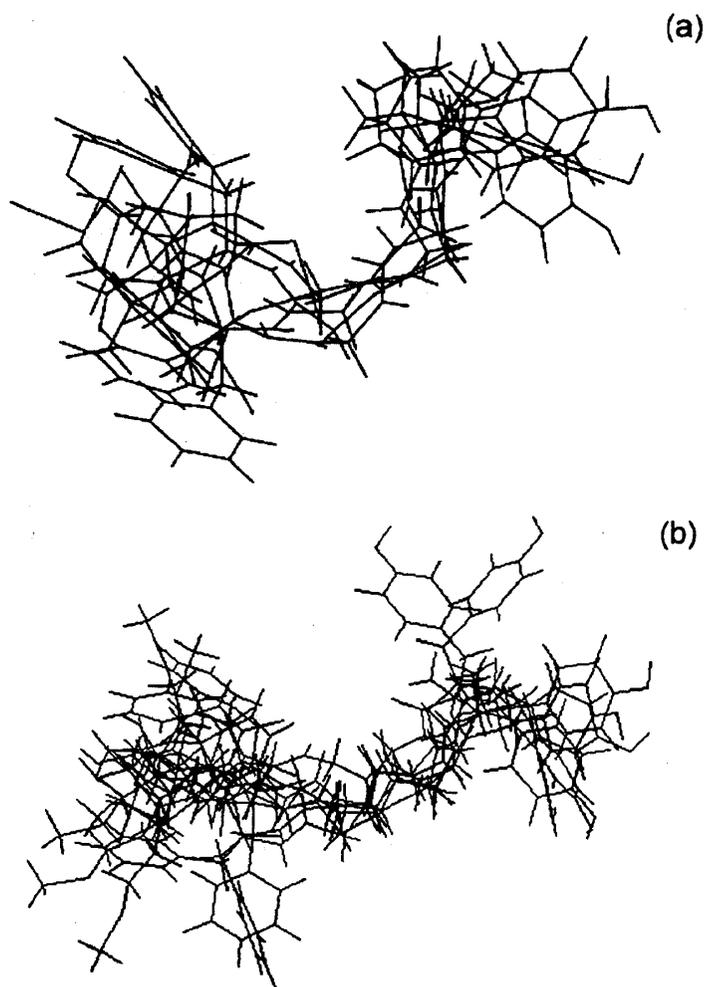


Figure 7: Met-enkephalin の NMR 実験のデータを精密化した 5 つの構造を重ね合わせたもの (a) と RISM によって溶媒効果を取り入れた徐冷シミュレーションの 8 回の試行で得られた 8 つの最小エネルギー構造を重ね合わせたもの (b).

Glu-2⁻と Arg-10⁺ 間の塩橋が得られなかった。今回我々はレベル 1 の溶媒効果を取り入れて、マルチカノニカル法で計算した。[39] 図 8 に X 線構造とこの計算から得られた最小エネルギー構造を比較する。両者の α -ヘリックスの位置が完全に一致していると共に、上述の Glu-2⁻と Arg-10⁺ 間の塩橋も今回は得られていることに注目されたい。

次の例はウシ膵臓トリプシンインヒビターのフラグメント BPTI(16-36) である。このペプチドは 21 個のアミノ酸からなるが、タンパク質全体の X 線回折実験データから、フラグメント全体が β -シート (β -ヘアピン) 構造を持つことが知られている。[40] 我々はまず、気相中の徐冷計算によって、 α -ヘリックスができずに、 β -ストランドばかりができることを示した。また、得られた最小エネルギー構造が、一部 β -シート構造をしていること

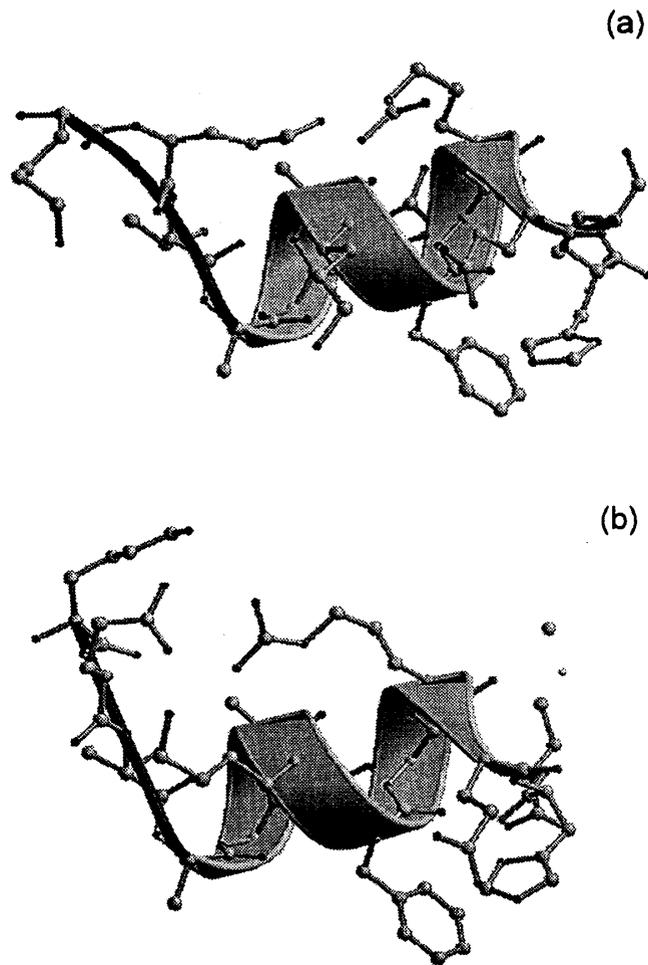
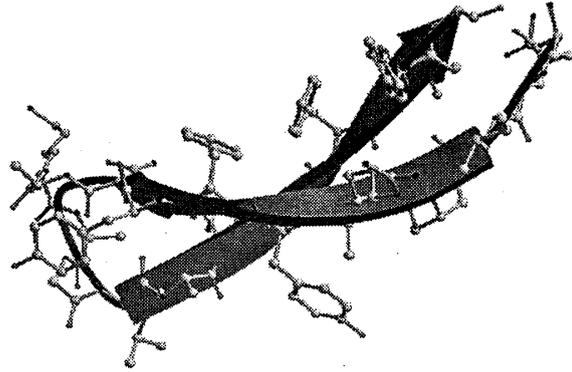


Figure 8: Ribonuclease A の X 線実験で決められた構造の C-peptide の部分 (a) と距離に依存した誘電率を導入したマルチカノニカル法で得られた C-peptide の最小エネルギー構造 (b).

も示した。[41] 但し、この構造と X 線構造は似ているとは言えないものであった。そこで、我々はレベル 2 の溶媒効果（溶媒接触表面積に比例する項）を導入して、徐冷シミュレーションを実行した。その結果、今度はきれいな β -シート構造ができることを確認した。[42] 図 9 に X 線構造とこの計算から得られた最小エネルギー構造を比較する。ストランドのねじれ具合等も含めて両者が似ていることは自明であろう。

ここまでは単純な二次構造要素（一つの α -ヘリックスまたは一つの β -シート）を持った小ペプチド系における第一原理からの構造予測の結果を紹介した。しかし、タンパク質は一般に複数の二次構造要素を持っているので、以下にもう少し複雑な系における結果を述べよう。

(a)



(b)

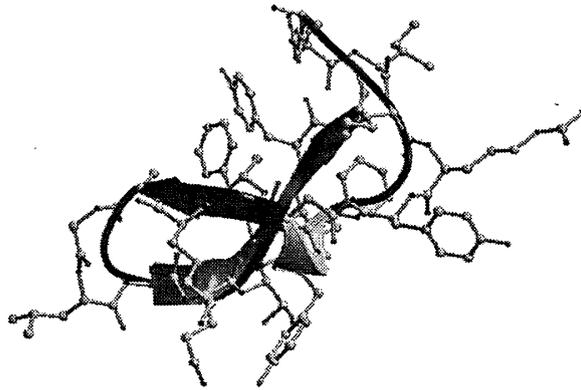


Figure 9: BPTI の X 線実験で決められた構造の残基 16-36 の部分 (a) と溶媒接触表面積に比例する項を導入した徐冷法で得られた BPTI(16-36) の最小エネルギー構造 (b).

最初の例はヒト副甲状腺ホルモンのフラグメント PTH(1-37) である。このフラグメントの NMR 実験は二本の α -ヘリックスの存在を示唆した。[43] 例えば、精密化された構造の一つは Gln-6 から His-9 までと Ser-17 から Lys-27 までの二本の α -ヘリックス構造を持っている。[43] 我々は、まず PTH(1-34) において気相中の徐冷計算を実行した。[44] 得られた最小エネルギー構造は Val-2 から Asn-10 までと Met-18 から Glu-22 までの二箇所に α -ヘリックス構造を持っていて NMR 実験の示唆と合致した。但し、二本目のヘリックスは短めになっている。我々は更にレベル 2 の溶媒効果を取り入れて今度は PTH(1-37) の徐冷シミュレーションを実行した。[45] 得られた最小エネルギー構造は Gln-6 から His-9 までと Gly-12 から Glu-22 までの二箇所に α -ヘリックス構造を持っている。よって、二本目のヘリックスの位置が少しずれているが、今度は実験と同じようにこのヘリックスの方が一本目のものより長い。図 10 に NMR 実験から得られた構造 [43] とこのシミュレーションで得られた最小エネルギー構造 [45] を示した。

次の例は streptococcal protein G の免疫グロブリン結合ドメインである。これは 56 個のアミノ酸からなるが、X 線実験 [46] 等から一本の α -ヘリックスと四本のストランドからなる β -シートを持った構造をしていることが知られている。この構造を図 11(a) に示す。この系において、我々はレベル 1 の溶媒効果（距離に依存した誘電率）を取り入れた徐冷シミュレーションを実行した。得られた最小エネルギー構造を図 11(b) に示す。この構造は四本の α -ヘリックス構造のみを持ち、 β -シート構造は得られなかった。この予測の失敗は主に溶媒効果の表現が不十分であるためと思われる。図 11(a) から分かるように X 線で得られた構造にはヘリックスとシートの間に疎水的コアが存在するので、静電相互作用のみを考慮に入れて疎水効果を見捨てるレベル 1 の溶媒効果では不十分なものも納得できる。現在、レベル 2 の溶媒効果を取り入れたシミュレーションが進行中である。

4.2 タンパク質の折り畳み問題

拡張アンサンブル法に基づくシミュレーションは最小エネルギー状態ばかりでなく、配位空間全体を幅広くサンプルすることによって、唯一回のシミュレーションから任意の温度における熱力学量を計算できるので、タンパク質の折り畳み問題を研究するのに特に適している。

タンパク質が折り畳まれるとき、最低二段階を経ると考えられる。第一段階はランダムコイル状態からコンパクトなグロビュール状態に移行する段階であり、第二段階はコンパクトな構造の中から自然の立体構造（最小エネルギー構造）に最終的に折り畳まれる段

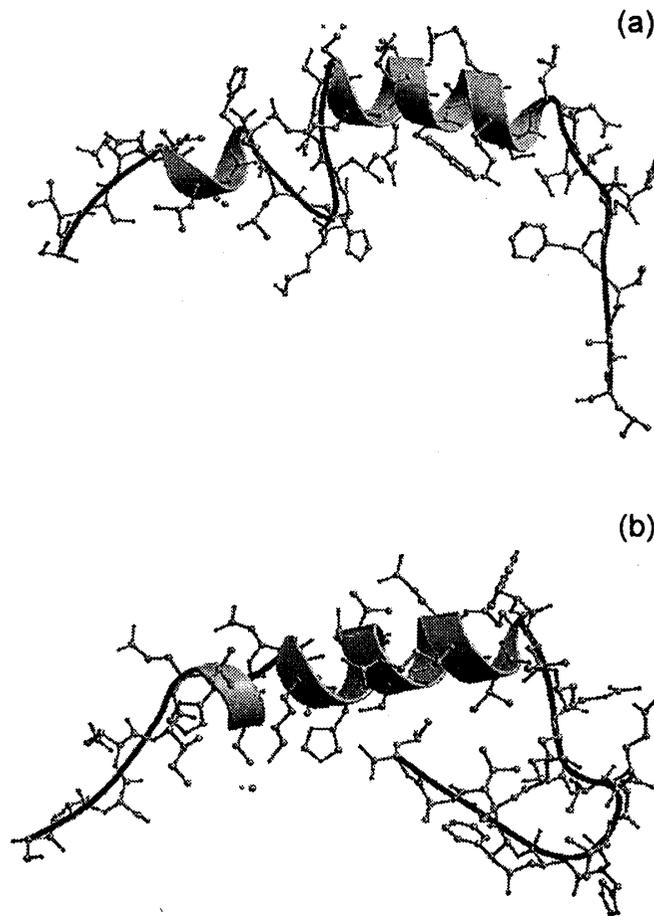


Figure 10: NMR 実験で決められた PTH(1-37) の構造の一つ (a) と溶媒接触表面積に比例する項を導入した徐冷法で得られた PTH(1-37) の最小エネルギー構造 (b).

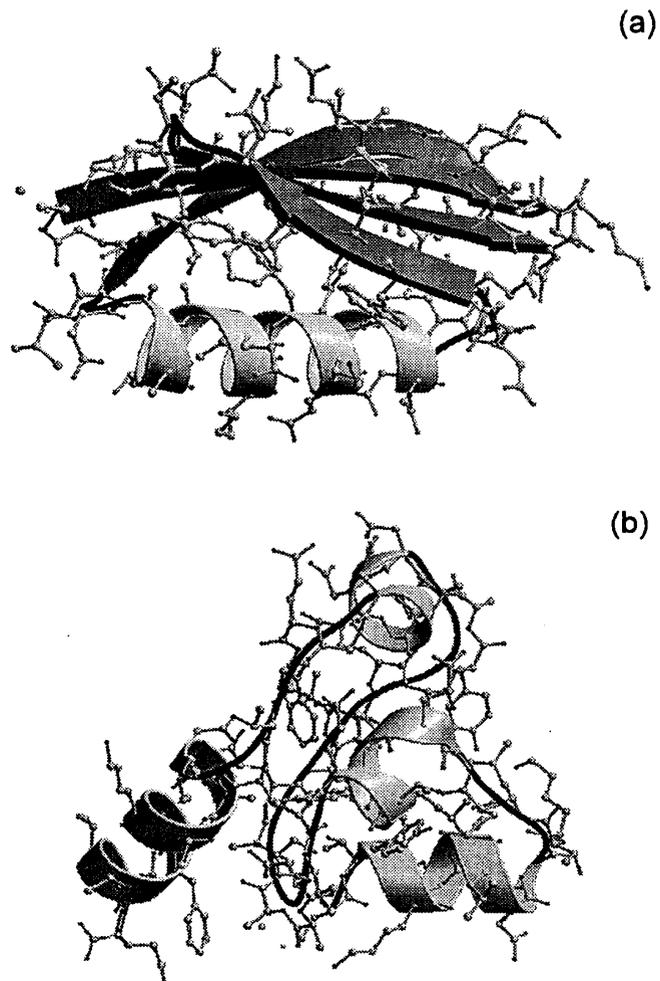


Figure 11: Protein G の X 線実験で決められた構造 (a) と距離に依存した誘電率を導入した徐冷法で得られた最小エネルギー構造 (b).

階である。これら二つの転移に対応してそれぞれ転移温度 T_θ と T_f が存在する。前者は高分子の分野で広く知られるコイル・グロビュール転移温度 (θ 温度) である。後者は「折り畳み温度」と呼ばれる。

これらの温度の計算の例として、我々は気相中のエンケファリンにおいて、新しい拡張アンサンブル法 [16] による計算を実行した。[47] まず、 T_θ は温度の関数としてタンパク質の平均体積が最も急速に変化する温度として定義した。すなわち、平均体積の温度微分が最大になる温度を求めた訳である。図 12 にこれらの量を温度の関数として表したものを示す。この結果から、 $T_\theta = 280 \pm 20$ K を得た。[47] 実は、図 5(b) から分るように、この温度は比熱の最大値を取る温度に対応している。この量からは、 $T_\theta = 310 \pm 20$ K が得られ、上の値と誤差の範囲内で一致している。

一方、折り畳み温度 T_f の方は、最小エネルギー構造とのオーバーラップの平均値の温度微分の絶対値が最大となる温度で定義した。ここで、最小エネルギー構造とのオーバーラップとは問題にしている構造がどれくらい最小エネルギー構造に似ているかを表す量であり、最小エネルギー構造と完全に一致するとき 1、完全にランダムな構造のとき 0 の値を取る。定義式は以下で与えられる。

$$O = 1 - \frac{1}{90 n_F} \sum_{i=1}^{n_F} |\alpha_i - \alpha_i^{(GS)}|, \quad (14)$$

ここで、 α_i と $\alpha_i^{(GS)}$ は、それぞれ、問題にする構造と最小エネルギー構造の (n_F 個の) 対応する二面角である。図 13 にこれらの量を温度の関数として示した。この結果から $T_f = 230 \pm 30$ K を得た。[47]

このようにタンパク質の折り畳みには二つの転移があるが、ここで、これらの転移を引き起こす要因 (driving force) は何かということを考えよう。タンパク質のランダムコイルからコンパクトな構造への転移は溶媒の効果が引き起こしていると考えられる人が多い。しかし、上述のエンケファリンの結果は溶媒を含まない気相中で得られたものである。よって、構造エネルギー E_p そのものがこの転移を引き起こしていると言わざるを得ない。実際、図 5(a) に見られるように、全構造エネルギーの平均値 $\langle E_p \rangle$ は T_θ (≈ 300 K) の前後で大きく変化している。そして、図 6 から分るように、この変化の大部分はレナード・ジョーンズエネルギー項から来ている。これは、タンパク質がコンパクトになるとき、ファンデアワールス接触が大幅に増えることに起因する。

それでは、溶媒の効果がエンケファリンの構造形成にどのように効いているかと考え

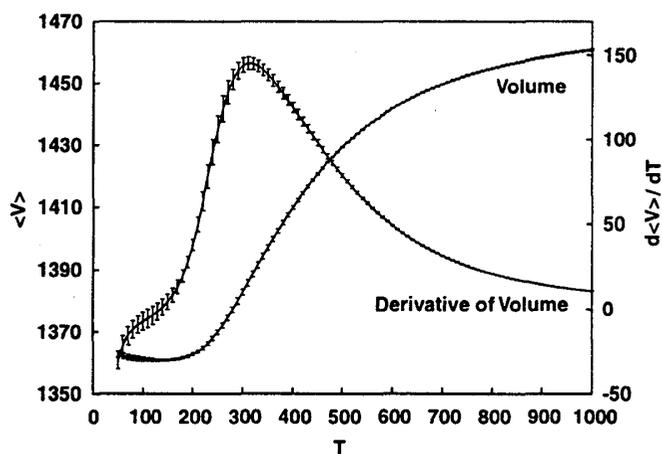


Figure 12: Met-enkephalin の平均体積 $\langle V \rangle_T$ (\AA^3) (a) とその温度 T (K) による微分 (b).

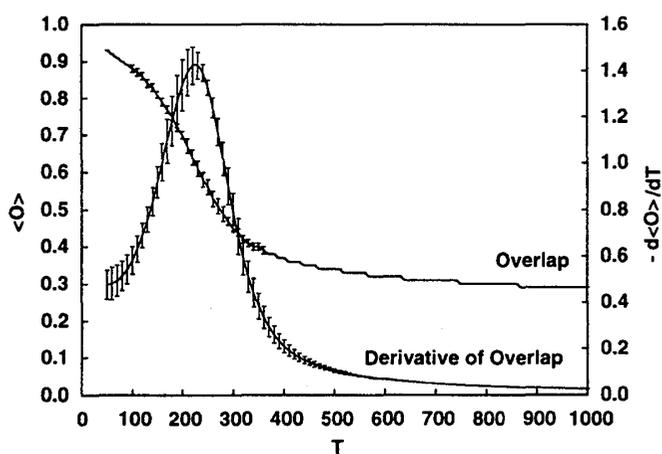


Figure 13: Met-enkephalin の最小エネルギー構造との平均オーバーラップ $\langle O \rangle_T$ (a) とその温度 T (K) による微分 (b).

Table 2: Protein G の準備された 10 個の構造の溶媒和自由エネルギー差 ΔE_S , 構造エネルギー差 ΔE_P , 及び全エネルギー差 $\Delta E_{tot} = \Delta E_P + \Delta E_S$ (全て kcal/mol). 溶媒接触表面積 A (\AA^2), 溶媒排除体積 V (\AA^3), 及び根平均 2 乗距離 (\AA) も与えた. 溶媒和自由エネルギーの計算は溶媒接触表面積に比例する項による.

Conf.	A	V	ΔE_S	ΔE_P	ΔE_{tot}	rms
1	3644	11350	0	0	0	0
2	3806	11842	29.4	-20.6	8.8	10.7
3	4033	12216	25.3	52.8	78.1	9.6
4	4089	12142	45.4	19.9	65.3	10.4
5	3927	12093	36.3	22.3	58.6	10.1
6	3944	11994	13.0	45.4	58.4	10.0
7	3997	12052	18.4	45.2	63.6	9.8
8	4046	12139	29.9	35.7	65.6	10.0
9	3708	11877	38.3	31.6	69.9	11.7
10	7612	14833	-88.5	237.2	148.7	21.1

ると, その答えは表 1 に示されている. 用意した 4 つの構造 (構造 1 は気相中の最小エネルギー構造, 構造 4 は伸びた構造) のうち, 伸びた構造が溶媒和自由エネルギー E_S が一番低いのである. すなわち, 水はエンケファリンを引き伸ばそうとしている.

エンケファリンは 5 個のアミノ酸からなる小ペプチドである. 次に, 前述の 56 個のアミノ酸からなるタンパク質 protein G についての結果を述べる. 表 2 に溶媒和自由エネルギー差 ΔE_S , 構造エネルギー差 ΔE_P , 及び全エネルギー差 $\Delta E_{tot} = \Delta E_P + \Delta E_S$ を 10 個の準備された構造について示した. 溶媒和自由エネルギーはレベル 2 の溶媒接触表面積に比例する項 [9] で表した. ここで, 構造 1 は X 線実験で得られた自然の構造 (図 11(a) 参照) であり, 構造 2 から 9 まではレベル 1 の (不十分な) 溶媒効果を取り入れて実行した徐冷計算から得られたエネルギー極小構造 (構造 2 は図 11(b) の最小エネルギー構造) であり, また, 構造 10 は (エネルギーを極小化した) 伸びた構造である.

この表から分ることは, まず, エンケファリンのときと同様に, 溶媒和自由エネルギーが伸びた構造 (構造 10) において最小であるということである. 一方, 構造エネルギーは伸びた構造が最大値を取るので, この両方の和である全エネルギーはこの構造で最大となる. この事実がエンケファリンのところで述べた, 「構造エネルギーがタンパク質のコイルからコンパクトな構造への転移を引き起こす」という推測を支持する.

表 2 から更に分ることは, コンパクトな構造の中では, 自然の構造 (構造 1) が最小の溶媒和自由エネルギーを持つということである. これは, 自然の構造では, 疎水基がタ

ンパク質の内部に埋もれ、親水基が溶媒に接している事実を反映している。

以上のことから、タンパク質の折り畳み機構について、次の推測を得た。一般に溶媒和自由エネルギーはタンパク質を引き伸ばそうとする。一方、構造エネルギーはタンパク質をコンパクトな構造にしようとする。(エンケファリンのような)小ペプチドでは伸びた構造とコンパクトな構造との構造エネルギー差が、溶媒の伸ばそうとする傾向に打ち勝つ程大きくないので、伸びた構造が得られる。しかし、タンパク質においては、伸びた構造とコンパクトな構造との構造エネルギー差が溶媒和自由エネルギー差に打ち勝って、コンパクトな構造になる。よって、タンパク質の折り畳みの最初の段階(ランダムコイルからコンパクトなグロビュールへの転移)は構造エネルギーによって引き起こされ、次の段階(コンパクトな構造の中から自然の構造が得られる転移)は溶媒和自由エネルギーによって誘導される。

5. おわりに

この総説では筆者等のタンパク質の第一原理からの立体構造予測及び折り畳み問題における研究を紹介した。すべてのシミュレーションはランダムな初期構造から実行されたことをここでもう一度強調したい。本稿の終りにタンパク質の折り畳み機構についての筆者の推測を述べたが、この推測の是非を最終的に決定するためには、RISM等のような厳密な溶媒和理論に基づく計算が必要であろう。最後に私の共同研究者の方々に感謝する。

References

- [1] C.J. Epstein, R.F. Goldberger, and C.B. Anfinsen (1963) *Cold Spring Harbor Symp. Quant. Biol.* **28**, 439-449.
- [2] N. Gō (1983) *Ann. Rev. Biophys. Bioeng.* **12**, 183-210.
- [3] 笹井理生 (1997) 日本物理学会誌 **52** 12月号, 901-908.
- [4] a) 岡本祐幸 (1996) 日本物理学会誌 **51** 4月号, 279-287.
b) 岡本祐幸 (1997) JCPE (Japan Chemistry Program Exchange) Newsletter **9**, No. 3, 3-18.
- [5] 西川建 (1997) 医療情報学 **17**, 421-432.
- [6] a) F.A. Momany, F.A., R.F. McGuire, A.W. Burgess, and H.A. Scheraga (1975) *J. Phys. Chem.* **79**, 2361-2381.
b) G. Némethy, M.S. Pottle, and H.A. Scheraga (1983) *J. Phys. Chem.* **87**, 1883-1887.
c) M.J. Sippl, G. Némethy, and H.A. Scheraga (1984) *J. Phys. Chem.* **88**, 6231-6233.

- [7] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller (1953) *J. Chem. Phys.* **21**, 1087–1092.
- [8] Y. Okamoto (1994) *Biopolymers* **34**, 529–539.
- [9] T. Ooi, M. Oobatake, G. Némethy, and H.A. Scheraga (1987) *Proc. Natl. Acad. Sci. USA* **84**, 3086–3090.
- [10] M. Masuya and Y. Okamoto (1998) in preparation.
- [11] a) D. Chandler and H.C. Andersen (1972) *J. Chem. Phys.* **57**, 1930–1937.
b) F. Hirata and P.J. Rossky (1981) *Chem. Phys. Lett.* **83**, 329–334.
- [12] S. Kirkpatrick, C.D. Gelatt, Jr., and M.P. Vecchi (1983) *Science* **220**, 671–680.
- [13] B.A. Berg and T. Neuhaus (1991) *Phys. Lett.* **B267**, 249–253.
- [14] a) A.P. Lyubartsev, A.A. Martinovski, S.V. Shevkunov, and P.N. Vorontsov-Velyaminov (1992) *J. Chem. Phys.* **96**, 1776–1783.
b) E. Marinari and G. Parisi (1992) *Europhys. Lett.* **19**, 451–458.
- [15] B. Hesselbo and R.B. Stinchcombe (1995) *Phys. Rev. Lett.* **74**, 2151–2155.
- [16] U.H.E. Hansmann and Y. Okamoto (1997) *Phys. Rev. E* **56**, 2228–2233.
- [17] S.R. Wilson, W. Cui, J.W. Moskowitz, and K.E. Schmidt (1988) *Tetrahedron Lett.* **29**, 4373–4376.
- [18] M. Nilges, A.M. Gronenborn, A.T. Brünger, and G.M. Clore (1988) *Protein Eng.* **2**, 27–38.
- [19] H. Kawai, T. Kikuchi, and Y. Okamoto (1989) *Protein Eng.* **3**, 85–94.
- [20] C. Wilson and S. Doniach (1989) *PROTEINS: Struct. Funct. Genet.* **6**, 193–209.
- [21] S.R. Wilson and W. Cui (1994) In: *The Protein Folding Problem and Tertiary Structure Prediction*, K.M. Merz, Jr. and S.M. Le Grand, eds. (Birkhäuser) pp. 43–70.
- [22] Y. Okamoto, M. Fukugita, T. Nakazawa, and H. Kawai (1991) *Protein Eng.* **4**, 639–647.
- [23] U.H.E. Hansmann and Y. Okamoto (1994) *J. Phys. Soc. Jpn.* **63**, 3945–3949; (1994) *Physica A* **212**, 415–437.
- [24] A.M. Ferrenberg and R.H. Swendsen (1988) *Phys. Rev. Lett.* **61**, 2635–2638; and erratum (1989) **63**, 1658.
- [25] U.H.E. Hansmann and Y. Okamoto (1993) *J. Comp. Chem.* **14**, 1333–1338.
- [26] M.H. Hao and H.A. Scheraga (1994) *J. Phys. Chem.* **98**, 4940–4948.
- [27] a) Y. Okamoto, U.H.E. Hansmann, and T. Nakazawa (1995) *Chem. Lett.* **1995**, 391–392.
b) Y. Okamoto and U.H.E. Hansmann (1995) *J. Phys. Chem.* **99**, 11276–11287.
- [28] A. Kidera (1995) *Proc. Natl. Acad. Sci. USA* **92**, 9886–9889.
- [29] U.H.E. Hansmann, Y. Okamoto, and F. Eisenmenger (1996) *Chem. Phys. Lett.* **259**, 321–330.

- [30] N. Nakajima, H. Nakamura, and A. Kidera (1997) *J. Phys. Chem.* **101**, 817–824.
- [31] U.H.E. Hansmann and Y. Okamoto (1997) *J. Comp. Chem.* **18**, 920–933.
- [32] Z. Li and H.A. Scheraga (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6611–6615.
- [33] Y. Okamoto, T. Kikuchi, and H. Kawai (1992) *Chem. Lett.* **1992**, 1275–1278.
- [34] W.H. Graham, E.S. Carter II, and R.P. Hicks (1992) *Biopolymers* **32**, 1755–1764.
- [35] a) M. Kinoshita, Y. Okamoto, and F. Hirata (1997) *J. Chem. Phys.* **107**, 1586–1599.
b) M. Kinoshita, Y. Okamoto, and F. Hirata (1998) *J. Am. Chem. Soc.* **120**, 1855–1863.
- [36] a) M. Kinoshita and D.R. Bérard (1996) *J. Comp. Phys.* **124**, 230–241.
b) M. Kinoshita, Y. Okamoto, and F. Hirata (1997) *J. Comp. Chem.* **18**, 1320–1326.
- [37] R.F. Tilton, Jr., J.C. Dewan, and G.A. Petsko (1992) *Biochemistry* **31**, 2469–2481.
- [38] J.J. Osterhout, R.L. Baldwin, E.J. York, J.M. Stewart, H.J. Dyson, and P.E. Wright (1989) *Biochemistry* **28**, 7059–7064.
- [39] U.H.E. Hansmann and Y. Okamoto (1998) *J. Phys. Chem. B* **102**, 653–656.
- [40] J. Deisenhofer and W. Steigemann (1985) *Acta Crystallogr.* **B31**, 238–250.
- [41] T. Nakazawa, H. Kawai, Y. Okamoto, and M. Fukugita (1992) *Protein Eng.* **5**, 495–503.
- [42] Y. Okamoto, M. Masuya, M. Nabeshima, and T. Nakazawa (1998) submitted for publication.
- [43] U.T. Marx, S. Austermann, P. Bayer, K. Adermann, A. Ejchart, H. Sticht, S. Walter, F.-X. Schmid, R. Jaenicke, W.-G. Forssmann, and P. Rösch (1995) *J. Biol. Chem.* **270**, 15194–15202.
- [44] Y. Okamoto, T. Kikuchi, T. Nakazawa, and H. Kawai (1993) *Int. J. Peptide Protein Res.* **42**, 300–303.
- [45] M. Masuya and Y. Okamoto (1998) in preparation.
- [46] A. Achari, S.P. Hale, A.J. Howard, G.M. Clore, A.M. Gronenborn, K.D. Hardman, and M. Whitlow (1992) *Biochemistry* **31**, 10449–10457.
- [47] U.H.E. Hansmann, M. Masuya, and Y. Okamoto (1997) *Proc. Natl. Acad. Sci. USA* **94**, 10652–10656.