

拡張アンサンブル法で探る蛋白質の折り畳み機構

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
e-mail: okamotoy@ims.ac.jp
URL: <http://konf2.ims.ac.jp/>
岡本祐幸

1. はじめに

タンパク質は自然界に存在する最も複雑な分子である。この分子は、小さいものでは数十個から大きいものでは数千個のアミノ酸が鎖状につながった生体高分子であり、DNA 及び RNA と共に生命現象の中樞を担っている。特に、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ことから、その立体構造予測が重要な研究テーマとされてきた。

また、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、多くの実験家及び理論家を魅了してきた。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、配位空間を幅広くサンプルすることができる、拡張アンサンブル法 (generalized-ensemble algorithms) が特に有効なシミュレーション手法である (拡張アンサンブル法のタンパク質の立体構造予測問題及び折り畳み問題への適用に関する解説としては、例えば、文献 [1] を参照。¹)

本講演では、拡張アンサンブル法による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の研究の最近の発展について報告する。

2. シミュレーション手法

2.1 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。本講演では、簡単のため、溶媒の寄与は無視するか誘電率として考慮した場合の結果のみを紹介する。一般的なタンパク質の系の全ポテンシャルエネルギー

¹ 日本語の拡張アンサンブル法という名前は、かなり定着しているようなので、筆者も使わせてもらっているが、英語ではまだはっきりしたものがないようである。筆者らは generalized ensemble という言葉を使っている。[2]

関数 E_{tot} (単位は kcal/mol) は、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。エネルギー関数内の諸々のパラメーターについては、標準的なエネルギー関数のものを使った。特に、モンテカルロシミュレーションでは、ECEPP/2 [3] のものを、また、分子動力学シミュレーションでは、Amber 4.1 [4] のものを使った。

2.2 拡張アンサンブル法

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。

$$W_B(E) = \exp(-\beta E) . \quad (1)$$

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。

$$P_B(T, E) \propto n(E)W_B(E) . \quad (2)$$

ここで、 $n(E)$ は状態密度である。状態密度 $n(E)$ 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。

拡張アンサンブル法 (generalized-ensemble algorithms) 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) [5] 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。

$$P_{mu}(E) \propto n(E)W_{mu}(E) \equiv \text{constant} . \quad (3)$$

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。

$$W_{mu}(E) \propto n^{-1}(E). \quad (4)$$

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[6]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）[7]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。

$$P_B(T, E) = \frac{P_{mu}(E) W_{mu}^{-1}(E) e^{-\beta E}}{\int dE' P_{mu}(E') W_{mu}^{-1}(E') e^{-\beta E'}}. \quad (5)$$

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。

$$\langle A \rangle_T = \int dE A(E) P_B(T, E). \quad (6)$$

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。[8] それ以後、タンパク質や高分子の分野で、いろいろな応用が試みられた（例えば、文献[9]-[18]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。[19]-[21]

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) [22] では、温度空間上の自由酔歩を、また、1/k 法 [23] 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。[2] また、拡張アンサンブル法と徐冷法 (simulated annealing) [24] とを組み合わせると、従来の徐冷法より更に有効に、最低エネルギー状態を求めることも示された。[2]

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

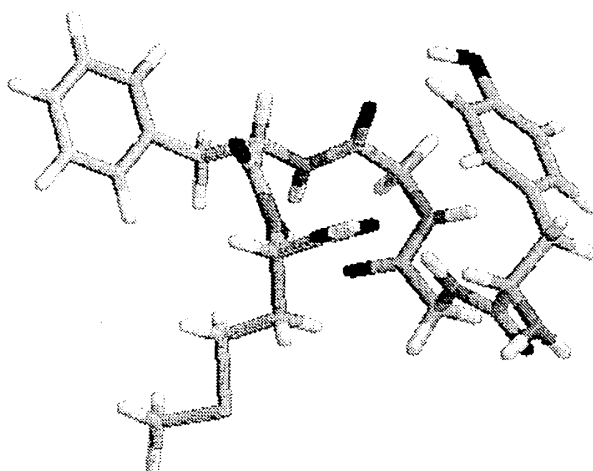


Figure 1: エンケファリンの気相中における最小エネルギー構造

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の開発が強く望まれている。我々は以下のような Tsallis の重み因子に基づいた拡張アンサンブル法を開発した。[25]

$$W(E) \propto \left(1 + \beta \frac{E - E_{GS}}{n_F}\right)^{-n_F}. \quad (7)$$

ここで、 E_{GS} は最小エネルギー値であり、 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）。

レプリカ交換法 (replica-exchange method) [26] も、マルチカノニカル法等より準備が非常に簡単な有望な手法である (レプリカ交換法の歴史については、例えば文献 [27] を参照)。この方法も既にタンパク質の系に導入されている。[28, 29]

3. 結果

ここで、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。

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。[30] そして、このペプチドは ECEPP/2 エネルギーで $E_{tot} \leq -11$ 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。[31, 8] 図 1 にこの最小エネルギー構造を示す。

図 2 には、マルチカノニカルシミュレーションにおける全構造エネルギー E_{tot} を MC

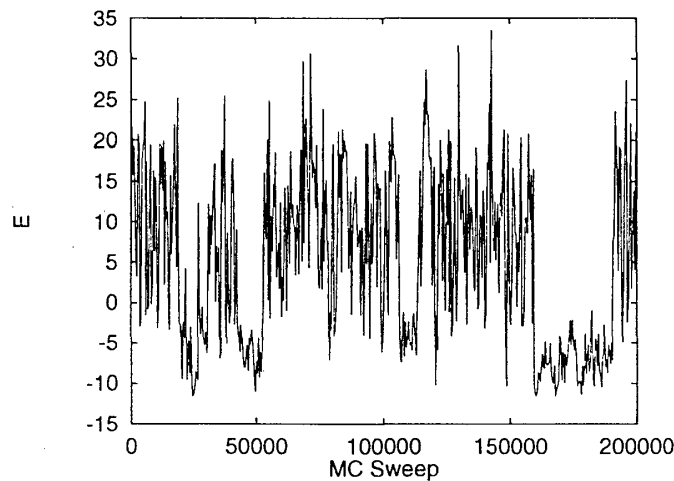


Figure 2: マルチカノニカルモンテカルロシミュレーションにおけるエンケファリンの構造エネルギー E_{tot} (kcal/mol) の「時間発展」.

Sweep の関数として表したものを示す. 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. [2, 25]

拡張アンサンブル法では, 式 (5) と (6) 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 3 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).

$$C \equiv \frac{1}{N k_B} \frac{d(\langle E_{tot} \rangle_T)}{dT} = \beta^2 \frac{\langle E_{tot}^2 \rangle_T - \langle E_{tot} \rangle_T^2}{N}, \quad (8)$$

図 3 は上述の 4 つの拡張アンサンブル法 (マルチカノニカル法, 焼き戻し法, $1/k$ 法及び Tsallis 重みによる手法) の結果 [2, 25] を重ね描きしたものであるが, 良い一致が得られている.

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.

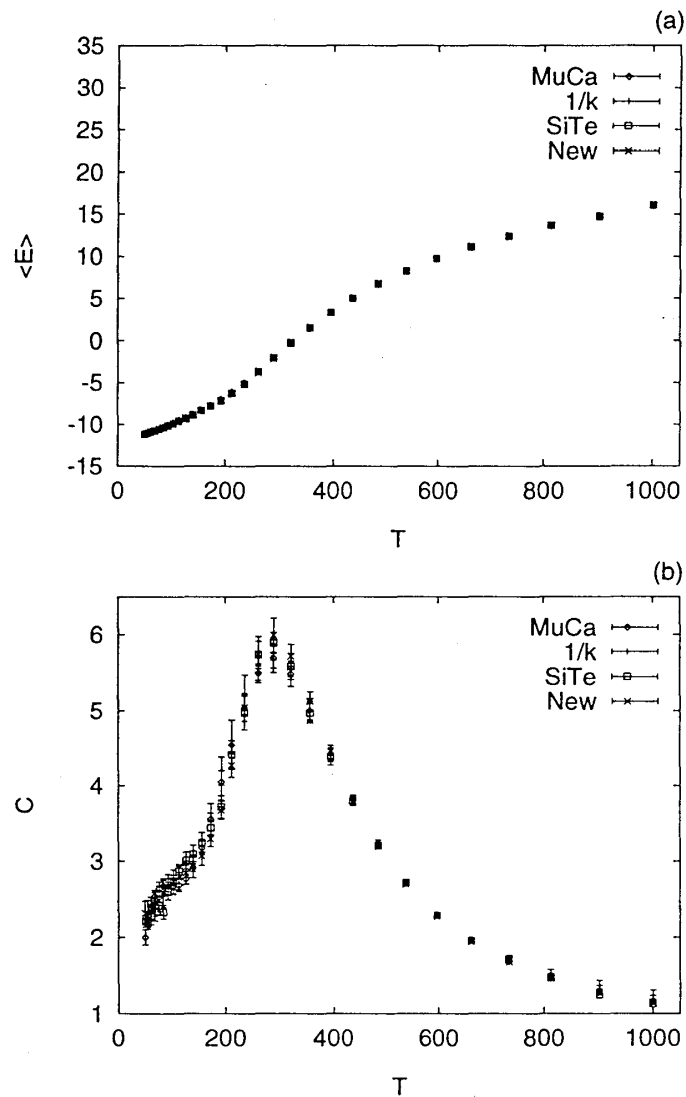


Figure 3: Met-enkephalin の全構造エネルギー E_{tot} の平均値 (kcal/mol) (a) 及び比熱 C (b) を温度 T (K) の関数として表したもの. これらは4つの異なる拡張アンサンブル法の結果を重ね合わせたものである.

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 [25] による計算を実行した. [32] 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4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. [32] 実は, 図3(b)からも分るように, 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.

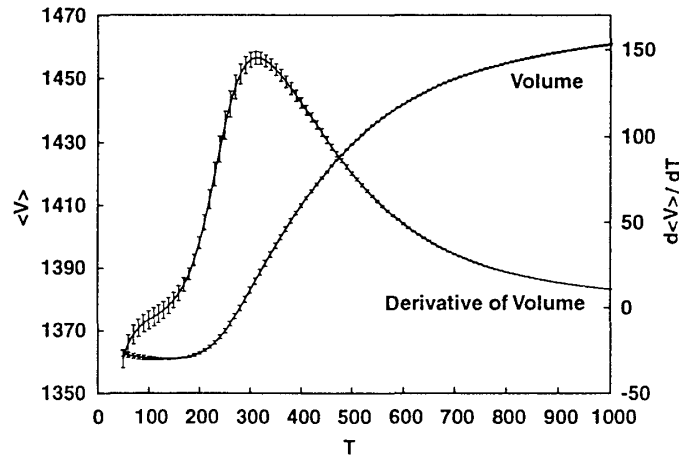


Figure 4: Met-enkephalin の平均体積 $\langle V \rangle_T$ (\AA^3) (a) とその温度 T (K) による微分 (b).

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。

$$O = 1 - \frac{1}{90 n_F} \sum_{i=1}^{n_F} |\alpha_i - \alpha_i^{(GS)}|, \quad (9)$$

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 5 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。[32]

この拡張アンサンブル法による、エンケファリンの折り畳みに関する研究は更に進められ、最近、自由エネルギー曲面がいろいろな温度において直接求められた。[33]

また、エンケファリンにおける最新の結果として、交換分子動力学法の結果を紹介しよう。[29] ここでは、8つのレプリカを用いた。あるレプリカにおける温度の分子動力学シミュレーションの時間ステップに対する変化を図 6 に示す。最低温度 150 K と最高温度 500 K の間を頻繁にランダムウォークしていることが分かる。

図 7 では交換法によって得られた、エネルギーのカノニカル確率分布と、従来のカノニカル分子動力学シミュレーションから得られた確率分布を比較する。低温で、従来のカノニカルシミュレーションがエネルギー極小状態にひっかかってしまうのに対して、交換

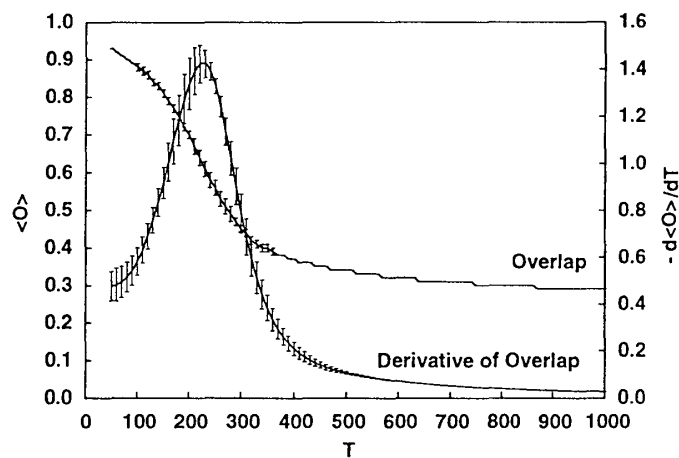


Figure 5: Met-enkephalin の最小エネルギー構造との平均オーバーラップ $\langle O \rangle_T$ (a) とその温度 T (K) による微分 (b).

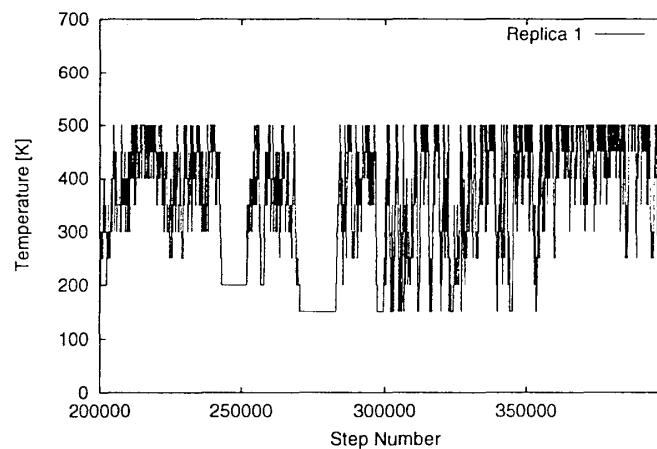


Figure 6: 交換分子動力学法の一つのレプリカにおける温度のランダムウォーク

法は正しい分布が得られることが分かった. [29]

最後に、もう少し大きい系の例として、ribonuclease A の C-peptide における結果を示そう。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 周の α -ヘリックス構造を持つことが知られている。[34] また、孤立した C-peptide においても、NMR 実験 [35] 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。[36] 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

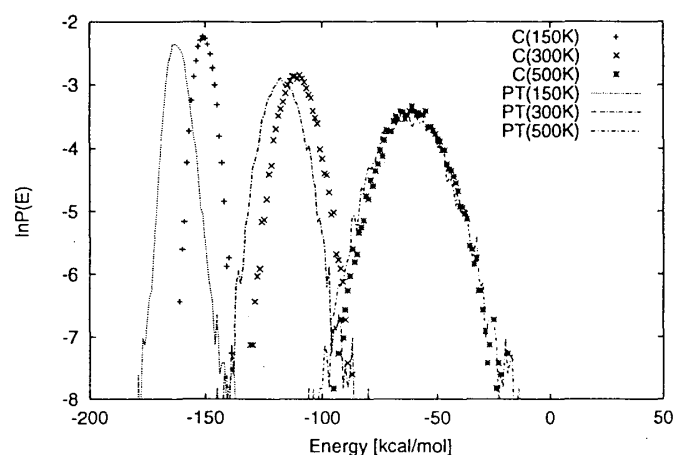


Figure 7: 交換分子動力学法から得られたカノニカル分布 (PT) と従来の一定温度の分子動力学シミュレーションから得られた分布 (C).

た C-peptide の NMR 実験の両方で確認されていた、Glu-2⁻ と Arg-10⁺ 間の塩橋が得られなかった。今回我々は距離に依存した誘電率を取り入れて、マルチカノニカルシミュレーションを実行した。[18] 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。

4. おわりに

この講演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拡張アンサンブルシミュレーションの結果を紹介した。タンパク質の折り畳み問題を理解するには、自然状態ばかりでなく、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、この問題には、配位空間を幅広くサンプルすることができる、拡張アンサンブル法が特に有効であることを主張した。

最後に、本講演で紹介した結果の共同研究者である、Ulrich Hansmann 氏 (元分子研, 現ミシガン工大) 及び杉田有治氏 (分子研) に感謝する。

References

- [1] a) 岡本祐幸 (1996) 日本物理学会誌 51 4 月号, 279–287.
- b) 岡本祐幸 (1998) 物性研究 70 9 月号, 719–742.
- c) U.H.E. Hansmann and Y. Okamoto (1999) in *Annual Reviews of Computational Physics VI*, ed. D. Stauffer (World Scientific, Singapore) pp. 129–157.

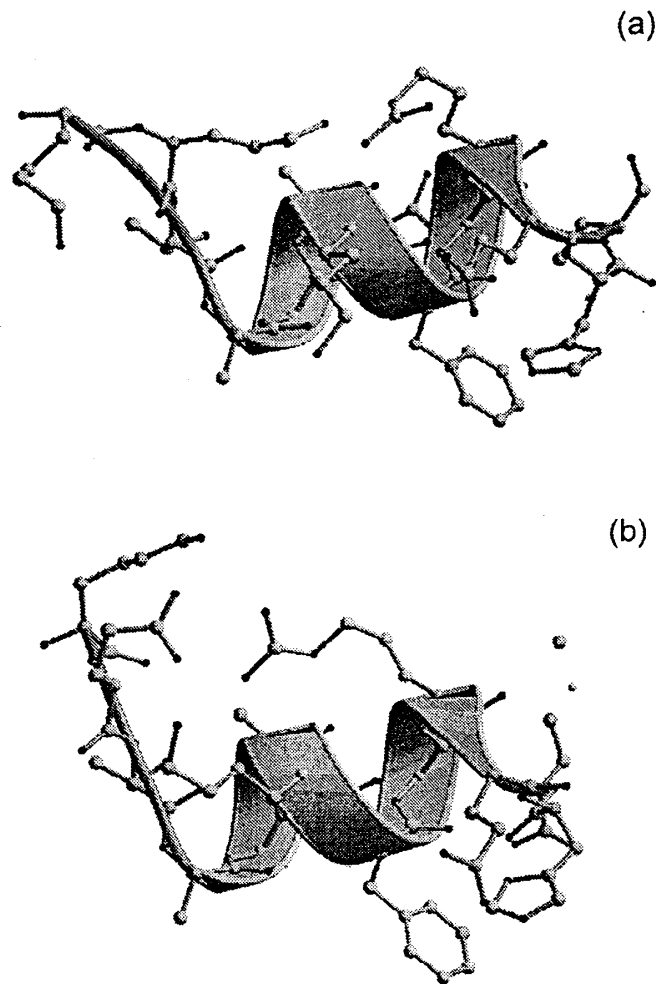


Figure 8: Ribonuclease A の X 線実験で決められた構造の C-peptide の部分 (a) と距離に依存した誘電率を導入したマルチカノニカル法で得られた C-peptide の最小エネルギー構造 (b).

- [2] U.H.E. Hansmann and Y. Okamoto (1997) *J. Comp. Chem.* **18**, 920-933.
- [3] 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.
- [4] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman (1995) *J. Am. Chem. Soc.* **117**, 5179-5197.
- [5] B.A. Berg and T. Neuhaus (1991) *Phys. Lett.* **B267**, 249-253.
- [6] U.H.E. Hansmann and Y. Okamoto (1994) *J. Phys. Soc. Jpn.* **63**, 3945-3949; (1994) *Physica A* **212**, 415-437.

- [7] A.M. Ferrenberg and R.H. Swendsen (1988) *Phys. Rev. Lett.* **61**, 2635–2638; *ibid.* (1989) **63**, 1658.
- [8] U.H.E. Hansmann and Y. Okamoto (1993) *J. Comp. Chem.* **14**, 1333–1338.
- [9] M.H. Hao and H.A. Scheraga (1994) *J. Phys. Chem.* **98**, 4940–4948.
- [10] 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.
- [11] A. Kidera (1995) *Proc. Natl. Acad. Sci. USA* **92**, 9886–9889.
- [12] N. Urakami and M. Takasu (1996) *J. Phys. Soc. Jpn.* **65**, 2694.
- [13] H. Noguchi and K. Yoshikawa (1997) *Chem. Phys. Lett.* **278**, 184–188.
- [14] J. Higo, N. Nakajima, H. Shirai, A. Kidera, and H. Nakamura (1997) *J. Comp. Chem.* **18**, 2086–2092.
- [15] Y. Iba, G. Chikenji, and M. Kikuchi (1998) *J. Phys. Soc. Jpn.* **67**, 3327–2220.
- [16] M. Schaefer, C. Bartels, and M. Karplus (1998) *J. Mol. Biol.* **284**, 835–848.
- [17] U.H.E. Hansmann and Y. Okamoto (1999) *J. Chem. Phys.* **110**, 1267–1276.
- [18] U.H.E. Hansmann and Y. Okamoto (1999) *J. Phys. Chem. B* **103**, 1595–1604.
- [19] U.H.E. Hansmann, Y. Okamoto, and F. Eisenmenger (1996) *Chem. Phys. Lett.* **259**, 321–330.
- [20] N. Nakajima, H. Nakamura, and A. Kidera (1997) *J. Phys. Chem.* **101**, 817–824.
- [21] C. Bartels and M. Karplus (1998) *J. Phys. Chem. B* **102**, 865–880.
- [22] 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.
- [23] B. Hesselbo and R.B. Stinchcombe (1995) *Phys. Rev. Lett.* **74**, 2151–2155.
- [24] S. Kirkpatrick, C.D. Gelatt, Jr., and M.P. Vecchi (1983) *Science* **220**, 671–680.
- [25] U.H.E. Hansmann and Y. Okamoto (1997) *Phys. Rev. E* **56**, 2228–2233.
- [26] a) K. Hukushima and K. Nemoto (1996) *J. Phys. Soc. Jpn.* **65**, 1604–1608; K. Hukushima, H. Takayama, and K. Nemoto (1996) *Int. J. Mod. Phys. C* **7** 337–344.
b) C.J. Geyer (1991) *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, 156–163.
c) M.C. Tesi, E.J.J. van Rensburg, E. Orlandini, and S.G. Whittington (1996) *J. Stat. Phys.* **82**, 155–181.
- [27] 伊庭幸人 (1997) *ISM Research Memorandum* No. 635.
- [28] U.H.E. Hansmann (1997) *Chem. Phys. Lett.* **281**, 140–150.
- [29] Y. Sugita and Y. Okamoto (1999) “Replica-Exchange Molecular Dynamics Algorithm for Protein Folding Simulations,” submitted for publication.

- [30] Z. Li and H.A. Scheraga (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6611–6615.
- [31] Y. Okamoto, T. Kikuchi, and H. Kawai (1992) *Chem. Lett.* **1992**, 1275–1278.
- [32] U.H.E. Hansmann, M. Masuya, and Y. Okamoto (1997) *Proc. Natl. Acad. Sci. USA* **94**, 10652–10656.
- [33] U.H.E. Hansmann, Y. Okamoto, and J.N. Onuchic (1999) *Proteins* **34**, 472–483.
- [34] R.F. Tilton, Jr., J.C. Dewan, and G.A. Petsko (1992) *Biochemistry* **31**, 2469–2481.
- [35] J.J. Osterhout, R.L. Baldwin, E.J. York, J.M. Stewart, H.J. Dyson, and P.E. Wright (1989) *Biochemistry* **28**, 7059–7064.
- [36] Y. Okamoto, M. Fukugita, T. Nakazawa, and H. Kawai (1991) *Protein Eng.* **4**, 639–647.