マルチカノニカル動力学シミュレーションによる タンパク質構造解析

東京薬科大学生命科学部*, 科学技術振興事業団 BIRD**

産業技術総合研究所生命情報科学研究センター***

池田和由*, **, ***, 肥後順一*, **

Energy landscapes of three peptides were studied with multicanonical molecular dynamics. The landscapes were different depending on the amino-acid sequence of the peptides. The landscapes, which consisted of conformational clusters, showed the existence of free-energy barriers among the clusters, and provided pathways of conformational changes. Folding mechanism of proteins was discussed from the current results. Some problems, which should be overcome to treat a protein system, are pointed out.

molecular dynamics / energy landscape / folding / multicanonical / replica exchange

1. はじめに

アミノ酸配列だけからタンパク質の立体構造を予測することは、生物物理学における長年の中心的問題である。これまでの旺盛な研究にもかかわらず、原子レベルまたはそれに近い分解能で、自由エネルギー地形 (free-energy landscape) が求められたタンパク質はまだない。我々は、厳密な計算をもとにペプチドの自由エネルギー地形を求めてきた。本稿ではその研究を紹介するが、その前に立体構造予測に関係する研究を概観する。

多くのタンパク質が実験的に構造解析された結果、立体構造に関する知識は大幅に増えた。Protein Data Bank (PDB) に登録された立体構造の数は、2002年9月現在18600個になり、その数は加速度的に増加し続けている。また、その構造情報をもとにタンパク質の立体構造の多様性 (fold) の数を概算すると、おおまかではあるが、 10^3 のオーダーであろうと言われている¹⁾。1次配列の多様性に比べて立体構造のそれがかなり小さいことは、驚きである。もしそれがほんとうなら、すべてのタンパク質の立体構造を網羅的に決定できるかもしれない。

各foldから代表的なタンパク質を選び、その立体構造を実験的に決めてしまえば、fold内の個々のタンパク質のそれは計算機を用いてモデリングできるからである。

ヒトゲノムのドラフトシーケンスが解読されたことも、立体構造予測の研究に拍車をかけている。ヒトのすべてのタンパク質の1次配列がわかれば、次は立体構造の決定である。タンパク質立体構造予測コンテスト CASP (<http://PredictionCenter.llnl.gov/casp4/>) は、その時代背景に呼応している。知られているように、CASPにおいてBakerらはデータベースから導いた経験則を利用して、かなり精度よく立体構造予測を行った。

フォールディング機構についての基礎的な研究も90年代後半に進んだ。実験としてはFershtらの ϕ 値の測定がある²⁾。彼らはフォールディング遷移状態にかかわるアミノ酸残基 (つまりフォールディングの核形成にかかわる残基) が存在することを示した。他方、理論からのアプローチとしてfunnel理論が提案された³⁾。この理論の展開の発端は、郷らによって提唱されたいわゆるGo modelである⁴⁾。funnelとは、日本語で「じょうご」の意味である。この理論では、タンパク質のエネルギー曲面がじょうごの形をしていると仮定する。ランダム

Protein Conformational Analysis by Multicanonical Molecular Dynamics Simulation

Kazuyoshi IKEDA*, **, ***, and Junichi HIGO*, **

* School of Life Science, Tokyo University of Pharmacy and Life Science

** BIRD, JST

***Computational Biology Research Center (CBRC), AIST

状態にあるタンパク質はじょうごのへりの領域で揺らいであり、へりから中心に落ちていく過程が折れたたまりである^{5) 6)}。じょうごを素朴に思い描けばフォールディングは簡単に起こりそうであるが、実際はそうではない。じょうごのへりは実は広大な領域であり、そのエントロピーは大きい(つまり、ランダム状態に含まれる鎖の構造は多様である)。したがって、フォールディングの初期には、鎖はさまざまな立体構造を造っては壊すという試行錯誤を行っている。その過程で偶然安定な核構造が形成されたとき、鎖はエネルギー曲面を滑り降り、天然状態へと折れたたまっていく。ところで、funnel理論ではタンパク質が粗視化されて議論される場合が多い。また、モデルでは、天然構造が最安定になるようにあらかじめ設定されている。仮にfunnel理論が正しいにしても、そこから天然構造を予測することは不可能である。

フォールディングを根本的に理解するとはどういうことだろうか。タンパク質は自由エネルギー最小を割り当てられた状態に折りたたまれる。物理化学的な法則だけを利用してエネルギー地形を描くことが、根本的な問題解決であると我々は考える。以下では、こうした根本的な問題解決への取り組みに議論を移す。

2. 効率的な構造サンプリング法

タンパク質の全原子と周囲の水分子を厳密に計算に取り込み、エネルギー地形が描けたら、フォールディング機構は解明され、天然状態の立体構造も導き出せる。そのためには、サンプリング法(計算機中でさまざまな立体構造を高速で作成し、そのエネルギーを評価する方法)の効率化が必要である。効率化は、ここ5年ほどで大きく進歩した。しかし、その話をする前に、より素朴かつ直接的な方法について考えてみる。

エネルギー計算だけからタンパク質の立体構造を予測しようとする考えは、新しいものではない。しかし、ランダム状態の立体構造から出発して、エネルギーが低くなるように構造変化をさせていっても、天然構造は得られない。エネルギー曲面には複雑な凹凸がある結果、系が正しい立体構造に行き着くはるか手前で穴の1つにトラップされる。また、ランダム構造の多様性の問題も深刻である。ランダム構造から出発して分子動力学(MD)を行っても、鎖はそのまま揺らぎ続け、いっこうに天然状態には近付かない。広大な構造空間の中から天然構造の場所を探し出すのは容易ではない。

その問題を克服するため、いくつかの効率的サンプリング法が提案された⁷⁾。その1つにマルチカノニカル法がある。もともとマルチカノニカル法はスピン系の物理

学を研究するために提案され、モンテカルロ法と組み合わせられて利用された。それをタンパク質に応用したのは日本の研究者である^{8) 10)}。さらに、このアルゴリズムをMD法に組み合わせた^{11) 12)}。そのうち我々は中島らによる方法¹¹⁾を採用している。この方法では、系の状態を記述する自由度は各原子の xyz 座標であり、従来のMD法同様に水分子も計算に含めることが可能である。各原子には通常の力だけでなく、系のdensity of statesから導かれる人工的な力も働く。この人工的な力の効果により、系は高温での状態と室温での状態を行き来するようになる。ここには2つの大きな意味がある。

- (1) 高温状態にあるときには、系はエネルギー局面上に分散する凹凸を乗り越えられる。その過程で折れたたまった鎖はほどけて別の形に折れたたまる。
- (2) シミュレーションから得られたさまざまな立体構造から、任意の温度の立体構造だけを抜き出せる。

したがって、室温で取り得る立体構造をあまねく眺めてその自由エネルギーを議論できる。

杉田と岡本は、レプリカ交換法をペプチドの系に応用した¹³⁾。この方法では、同一の系を多数用意して、それぞれ異なる温度で同時にMD計算を行う。シミュレーションの過程で2つの系のエネルギーが近くなれば、ある遷移確率($\exp[-\Delta\beta\Delta E]$; $\Delta\beta$ と ΔE はそれぞれ2つの系の温度の逆数の差とエネルギー差)で温度の交換を行う。これにより、1つの系は高温から室温までの広いエネルギー状態を経験する。高温で系はエネルギーの凹凸を飛び越える。室温時の立体構造を集めると、室温で許される立体構造の集合が得られる。この方法には、系が大きくなるほど ΔE の値が大きくなり、遷移確率が小さくなるという問題点がある。また、温度交換の過程で系が非平衡状態にさらされる。この問題を軽減するために、マルチカノニカル法とレプリカ交換法の融合が図られている¹⁴⁾。

3. 3つのペプチドのエネルギー地形

我々は、マルチカノニカルMD法を使い、ペプチドのエネルギー地形を研究してきた^{15) 17)}。その結果、全原子モデルで10残基程度のペプチドのエネルギー地形を詳細に得られるようになった。計算結果は、実験結果をうまく再現する^{15) 16)}。ペプチドは3種類で、ここでは便宜的にペプチド1、ペプチド2、ペプチド3とよぶ(詳細は後述)。いずれのペプチドも水分子に満たされた球の中心に置いた(球の大きさは、ペプチド構造が多様に変化するのに十分な大きさを確保している)。力場はAMBER parm 96¹⁸⁾を使用した。ペプチドの初期構造はランダム構造であり、数十ナノ秒オーダーのシミュレ

ーションの後、そこから得られる立体構造を解析に利用した。結果の詳細を以下で見ていく。

ペプチド 1¹⁵⁾(Ace-Lys-Gln-Cys-Arg-Glu-Arg-Ala-Nme. Ace, Nme は、それぞれ N 末端, C 末端に付けられたアセチル基, メチル基) は DNA 結合タンパク質 cMyb の DNA 結合ヘリックス部分から切り出した。したがって、ペプチド 1 はこのタンパク質の機能的に重要な部分である。我々は、シミュレーションによって 300 K における立体構造の集まりを得た。このペプチドは自由度 399 ($= 3N - 6$; N はペプチドを構成する原子数) をもつので、その構造空間は多次元である。得られた多数の立体構造は、この多次元空間に分布する。我々は、主成分分析 (PCA)¹⁵⁾を用いてこの分布を評価した。Fig.1 は第 2 および第 3 成分へ射影された立体構造の分布である。個々の点はそれぞれ立体構造を示していて、点同士が近ければその立体構造も似ている。点が多数集中して

いる領域 (クラスター) は構造分布が高いことを表し、系がそのクラスター中にあるとき自由エネルギーが下がる。よって、この図は単なる構造分布ではなく自由エネルギー地形である。ところで、第 1 成分への射影では、本来分離しているクラスター A と B が重なってしまう。すなわち、第 1 成分への射影では、エネルギー地形の重要な側面が消えてしまう。

ペプチド 1 でのランダム構造の存在確率は約 96 % であった。CD スペクトル測定によって、このペプチドは単独ではランダム構造をとることがわかっている¹⁵⁾、計算と実験は定性的に一致する。2 つの α -helix クラスター A, B のうち、A はこのペプチドが cMyb 中での立体構造に等しいが、B は主鎖内水素結合が若干乱れている。両者とも存在確率は低い (A, B ともに 1 %)。A を安定化している要因はエンタルピーであり、B のそれはエントロピーである。クラスター A, B の分離はその間の

Fig.1 Energy landscape of peptide 1 at 300 K. Conformations are projected on the second and third principal axes. Cluster A consists of conformations (complete helix) that involve three intra-mainchain hydrogen bonds (HB α) to form α -helix, and cluster B does of those (imperfect helix) where the hydrogen bonds are partly broken. Most of conformations in cluster C are random-coil conformations, but small parts of them are β -hairpin (HPN1, HPN2) or conformations (conf. with HB3) that contain a HB α . Conformation appearing in cMyb is included in cluster A.

Fig.2 Energy landscape of peptide 2 at 300 K. Conformations are projected on the first and second principal axes. Conformations in clusters A-C are β -hairpin. Conformations in cluster D have an intra-mainchain hydrogen bond to form α -helix, although they are not helical. Cluster E consists of random-coil conformations with no intra-mainchain hydrogen bond.

自由エネルギー障壁の存在を示すが、この障壁はペプチドと水との水素結合を切断すること起因している。また、クラスターCの中にはおもにランダム構造が分布するが、 β -hairpin（存在確率2%）や 3_{10} helix（0.5%）などの2次構造も存在した。つまり、実験的にはランダム構造であると判定される状態の中に、多様な2次構造が低確率で含まれている。また、エネルギー地形の中に α -helix クラスターが存在したことは、この配列が潜在的に α -helix をとりやすい傾向をもっており、置かれた環境によってその傾向が顕在化することを示す。

ペプチド2¹⁶⁾(Ace-Ile-Thr-Val-Asn-Gly-Lys-Thr-Tyr-Nme)は単独で β -hairpinになるように設計されており、NOE測定によってそれが確かめられている¹⁹⁾。Fig.2は計算から得られた室温でのエネルギー地形である。クラスターA, B, C はいずれも β -hairpin構造に対応するが、水素結合パターンが互いに異なる。Aの中の多くの立体構造ではstrand間水素結合が3本、Bでは2本、Cでは1本含まれている。なお、クラスターDは、 α -helix形

成のための主鎖内水素結合を1本もっているが、形はhelixでない。クラスターE内の構造はランダム構造であった。 β -hairpinの存在確率は約30%で、NOEの測定結果と定性的に一致する。また、ペプチド2に α -helixが含まれないことはペプチド1と対照的であり、配列の違いによるエネルギー地形の違いを示している。なお、ペプチド2の β -hairpinの安定化要因は、Val, Ileの側鎖とLysの側鎖の根元部分が空間的に集まることによる疎水性コアの形成であると、我々は推測している。

次に、タンパク質の一部分の自由エネルギー地形からそのタンパク質全体のフォールディングを議論する。ペプチド3¹⁷⁾はSH3ドメインのdistal β -hairpinに相当する配列(Ace-Ala-His-Ser-Leu-Thr-Thr-Gly-Gln-Thr-Nme)をもつ。この部分は Φ 値解析によってタンパク質全体のフォールディングの核になると言われている²⁰⁾。得られた室温での立体構造の集合を、PCAからの上位3成分による3次元空間に射影した(Fig.3)。クラスターA, Bは α -helix構造からなる(2つのクラスターの存在確

Fig.3 Energy landscape of peptide 3 at 300 K. Conformations are projected on the first, second, and third principal axes. Cluster A and B consist of helical conformations. Conformations in cluster E have the same intra-mainchain hydrogen bonds as that of distal β -hairpin in SH3. Cluster D, which involves a variety of β -hairpins, surrounds cluster E. Cluster C is an intermediate between helical and hairpin clusters. Other conformations in broken line are random-coil.

率の和は18%)。helix形成のための主鎖内水素結合を4本もったクラスターがA、2本もったクラスターがBである。これらのクラスター間には自由エネルギーの障壁が存在する。水素結合を3本もった立体構造も少数あるが、それらはAの中に埋もれている。よって、このペプチドはhelixを形成していく過程で、水素結合を2本もつ構造から4本の構造へと変化する。Aの中の α -helix構造が比較的安定である理由は、疎水性コアの形成である。すなわち、Gly(この残基からC末端側ではhelixが壊れている)からN末端側の領域がhelixを形成するとAla, Leu, さらに最もC末端側にあるThrの側鎖の疎水性面が集まる。PDBで検索すると、 α -helixがGlyによってブレイクした領域にクラスターAに類似した構造が頻繁に見られる。

一方、 β -hairpin構造のクラスターも見つかった。Fig.3でSH3ドメイン中のdistal β -hairpin構造に対応するのはクラスターEであり、それを取り囲んでクラスターDがある。Dの中にはさまざまなstrand間水素結合パターンをもつ β -hairpinが分布している。これら β -hairpin構

造の存在確率の和は約9%であった。さらにEとDは、ランダム構造群(Fig.3の破線で囲まれた部分)と切れ目なく接していた。つまり、ランダム構造と β -hairpinの間は自由エネルギー障壁がなく、構造変化が容易である。一方、この図ではわかりにくい、クラスター(D+E)とBとの間には、中間的なクラスターCが存在する。また、クラスター(D+E)からAへの直接的な構造変化は起こりにくい。

ペプチド3の計算結果は、タンパク質全体のフォールディングへの示唆を与える。このペプチドは β -hairpinをとれるが、より高確率で α -helixやランダム構造もとる。しかし、この領域は最終的には β -hairpinをとることでタンパク質全体として安定になる。フォールディング初期において、この領域は、ランダム、 β -hairpin、 α -helixおよびその中間構造の間を揺らいでいるが、それは鎖がさまざまな立体構造を生成消滅させることで、安定な立体構造を模索していることを意味する。この過程で偶然 β -hairpinが生成され、かつ鎖のほかの領域も適当な立体構造に折れたたまり、さらにそ

れらがエネルギー的に安定な相互配置に置かれたときに、タンパク質全体のフォールディングが進むのであろう。類似したメカニズムがprotein G に対して提出されている²¹⁾。

4. 解析法の問題点

今後、より大きい系(すなわちタンパク質)を取り扱うためには、適切な解析方法も必要である。タンパク質の構造空間は本来多次元空間である。一方、人間が直感的に理解できる空間は3次元までである。PCAでは、多次元空間に分布する点(すなわち立体構造)を3次元空間に射影することで、理解を可能にした。しかし、安易な射影はエネルギー地形の重要な側面を覆い隠す。多次元空間中で互いに分離したクラスターが、射影によって互いに重なり合うこともある。この問題を、最近神谷らは浮き彫りにした²²⁾。すでに見たようにペプチド2には、 β -hairpinとランダム構造のクラスター間に自由エネルギー障壁がある。このエネルギー地形を慣性半径、分子表面積などのよく知られたパラメータを座標軸とした空間に射影してみると、障壁は消失して(すなわち、見落とされて)funnel型に近いエネルギー地形が得られる。

これまで我々の経験したペプチドでは、少数の座標軸への射影によりエネルギー地形はうまく表現できた。しかし、鎖が長くなると構造分布は複雑になり、従来の方法では解析が困難になるだろう。また、単一の座標系で全分布を記述することにも、無理が出てくるだろう。サンプリング法の効率化とあわせて、よりよい解析手法の開発も必要である。

5. 最後に

本稿では省略したが、我々は力場の問題も考察してきた^{15), 23)}。その結果、AMBER parm 94, parm 96, CHARMの力場の比較では、parm 96とCHARMがparm 94よりも妥当であることがわかった²³⁾。最近AMBER parm 99がリリースされたが、我々はこの力場についての経験がまだない。分子研の岡本らがこの力場についての評価を始めたので報告が待たれる。

ペプチド鎖が長くなれば、サンプリングのさらなる効率化が望まれる。ある形に折れたたまった鎖をほかの形に折れたたませるためには、鎖を「ほどく」必要がある。これまでのシミュレーションに立脚した方法では、ほどく操作を高温での系の構造揺らぎに頼っている。しかし、鎖が長くなるとこの操作は困難になる。解決法の1つは計算機の処理速度の向上であり、もう1つはアルゴリズムの効率向上である。

ほどく操作を省略するアルゴリズムとして、千見寺らは「すり抜け」可能な鎖のモデルを提案している²⁴⁾。しかし、このアルゴリズムは粗視化モデル(lattice model)に対しては有効性であるが、原子が大きさをもっている全原子モデルでは問題点が多いであろう。ここを克服する手法が求められる。

文 献

- 1) Chothia, C. (1992) *Nature* **357**, 543-544.
- 2) Fersht, A. (1999) *Structure and mechanism in protein science*, W. H. Freeman & Company, New York.
- 3) Onuchic, J. N., Luthey-Schulten, Z. and Wolynes, P. G. (1997) *Annu. Rev. Phys. Chem.* **48**, 545-600.
- 4) Go, N. (1983) *Annu. Rev. Biophys. Bioeng.* **12**, 183-210.
- 5) 笹井理生, 寺田智樹(2002)蛋白質 核酸 酵素 **47**, 677-683.
- 6) 高田彰二(2002)蛋白質 核酸 酵素 **47**, 684-689.
- 7) Mitsutake, A., Sugita, Y. and Okamoto, Y. (2001) *Biopolymers* **60**, 96-123.
- 8) Hansmann, U. H. E. and Okamoto, Y. (1993) *J. Comp. Chem.* **14**, 1333-1338.
- 9) Okamoto, Y. and Hansmann, U. H. E. (1995) *J. Phys. Chem.* **99**, 11276-11287.
- 10) Kidera, A. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 9886-9889.
- 11) Nakajima, N., Nakamura, H. and Kidera, A. (1997) *J. Phys. Chem. B* **101**, 817-824.
- 12) Hansmann, U. H. E., Okamoto, Y. and Eisenmenger, F. (1996) *Chem. Phys. Lett.* **259**, 321-330.
- 13) Sugita, Y. and Okamoto, Y. (1999) *Chem. Phys. Lett.* **314**, 141-151.
- 14) Sugita, Y. and Okamoto, Y. (2000) *Chem. Phys. Lett.* **329**, 261-270.
- 15) Higo, J., Ito, N., Kuroda, M., Ono, S., Nakajima, N. and Nakamura, H. (2001) *Protein Sci.* **10**, 1160-1171.
- 16) Higo, J., Galzitskaya, O. V., Ono, S. and Nakamura, H. (2001) *Chem. Phys. Lett.* **337**, 169-175.
- 17) Ikeda, K., Galzitskaya, O. V., Nakamura, H. and Higo, J., *J. Comp. Chem.* in press.
- 18) Kollman, P. A., Dixon, R., Cornell, W. D., Fox, T., Chipot, C. and Pohorille, A. (1997) *Computer Simulation of Biomolecular Systems*, KLUWER/ESCON, **3**, 83-96, The Netherlands.
- 19) Ramirez-Alvarado, M., Blanco, F. J. and Serrano, L. (1996) *Nat. Struc. Biol.* **3**, 604-612.
- 20) Grantcharova, V. P., Alm, E. J., Baker, D. and Horwich, A. L. (2001) *Curr. Opin. Struc. Biol.* **11**, 70-82.
- 21) 本田信也(2002)生物物理 **42**, 174-178.
- 22) Kamiya, N., Higo, J. and Nakamura, H. (2002) *Protein Sci.* **11**, 2297-2307.
- 23) Ono, S., Kuroda, M., Higo, J., Nakajima, N. and Nakamura, H. (2002) *J. Comp. Chem.* **23**, 470-476.
- 24) Chikenji, G., Kikuchi, M. and Iba, Y. (1999) *Phys. Rev. Lett.* **83**, 1886-1889.

池田和由

池田和由 (いけだ かずよし)

東京薬科大学生命科学部, 科学技術振興事業団BIRD, 産業技術総合研究所生命情報科学研究センター
2001年東京薬科大学生命科学部環境生命科学科卒業. 卒業研究テーマは甲殻類(オオミジンコ)の形態形成遺伝子の同定であったが, その後現研究室に移り今の研究を始める. また, 現在BLS バイオインフォマティクス分科会 (<http://www.tit-bls.org/bioinfo/index.html>) という若手の会の運営にも携わっている.

連絡先: 〒192-0392 八王子市堀之内1432-1

E-mail: s978003@educ.ls.toyaku.ac.jp

肥後順一 (ひご じゅんいち)

東京薬科大学生命科学部, 科学技術振興事業団BIRD

学部で原子核理論, 大学院修士で超伝導を研究するが, 落ちこぼれる. 博士後期からタンパク質の理論計算をはじめ, 現在に至る. 最近は, そば打ちや釣りに凝っている.

連絡先: 同上

E-mail: higo@ls.toyaku.ac.jp