

2013 年度（平成 25 年度）

博士論文

側鎖の効果を含んだ CαGō モデルの開発と  
タンパクフォールディング機構解析への応用

立命館大学大学院

理工学研究科 総合理工学専攻

杉田昌岳

## 目次

第 1 章 序論	
1.1 フォールディングに関する一般的知見	5
1.2 実験的アプローチ	6
1.3 理論的アプローチ	7
1.4 適切な計算手法を用いる事の必要性	8
1.5 多状態タンパクの計算機的な研究の必要性	8
1.6 研究目的	9
参考文献	10
第 2 章 側鎖の効果を取り込んだ C $\alpha$ G $\beta$ モデルの開発	
2.1 背景	17
2.2 方法	19
2.3 結果	24
2.4 考察	35
2.5 結論	36
参考文献	37
第 3 章 ferredoxin-like fold タンパクのフォールディング機構の解析： 自由エネルギープロファイルの形とフォールディングコアとの関係	
3.1 背景	41
3.2 方法	43
3.3 結果	46
3.4 考察	57
3.5 結論	59
参考文献	60
第 4 章 多状態タンパクのフォールディング機構の解析：フォールドンの探索と その特徴付け	
4.1 背景	67
4.2 方法	68
4.3 結果	71
4.4 考察	79
4.5 結論	82
参考文献	83

第 5 章 結語 87

謝辞 91



## 第1章 序論

### 1.1 フォールディングに関する一般的知見

遺伝子の情報は DNA から RNA に転写された後、RNA からポリペプチド鎖（タンパク）に翻訳される。しかし大抵のタンパクは伸びた鎖として機能する訳ではなく、翻訳後は特有の立体構造（天然構造）へ折り畳まれる。この折りたたみの過程をフォールディングと呼ぶ。タンパクの機能はその構造と大きく関わっているため、フォールディング機構の理解はゲノムの配列情報とプロテオームの機能とを結びつけるために非常に重要である。他方、アルツハイマーなどの原因とされているアミロイド繊維はタンパクのミスフォールドが原因で生じる凝集体である事が知られている。その予防策や治療法を開発する上でもフォールディング機構の理解は重要な役割を担っている。フォールディングは生体を構成する基本的な機構であるため、その理解は生理学的に大きな意義を持つ。

他方、フォールディングは物理的にも非常に興味深い特徴を持つ。フォールディング研究の発端となったアンフィンセンの実験<sup>1</sup>は、タンパクを変性剤で変性させた後に生理学的な条件へ戻すと、天然構造へ自発的にリフォールドする事を示した。この実験はタンパクの天然構造は自由エネルギー的に安定な構造であり、タンパクはポリペプチド鎖の取り得る無数の構造の中からその天然構造を自力で見つけ出す事が出来る事を示している。この実験事実はアンフィンセンのドグマと呼ばれる。しかしその一方で非常に大きな自由度を持つタンパクがランダムな探索の結果天然構造を見つけ出すためには天文学的な時間が必要である事がレビンタールによって示された。<sup>2</sup>これは一見タンパクが自発的に素早くフォールドするという実験事実と矛盾しているように感じるため、レビンタールのパラドクスと呼ばれる。アンフィンセンのドグマとレビンタールのパラドクスは、全てのフォールドしうるタンパクには天然構造へ素早く到達するための物理的な機構が存在することを示唆する。その機構を明らかにする事がフォールディング研究の中核である。

これまで、実験と理論の両者が協調して研究を行う事で、フォールディングの大まかな特徴が明らかになってきた。実験的な研究の役割は、フォールディングが実際にどのような特徴を持っているかを出来るだけ詳細な形で示す事である。一方、理論的な研究の役割は、実験的に示された振る舞いを無理なく説明できるようなモデルを立てる事であり、数値解やシミュレーションの結果を

基にそのモデルを評価する。以下に実験と理論の両面からこれまでに明らかにされてきた事の概要を説明する。

## 1.2 実験的アプローチ

フォールディング機構を明らかにする上で重要な実験的なアプローチは、その速度論的な特徴を知る事である。反応速度はその自由エネルギーランドスケープの形に関する情報を与える。これまで、ストップフロー装置<sup>3</sup>や連続フロー装置<sup>4</sup>と分光学的手法を用いる事でその特徴が調査されてきた。その結果、変性状態から一気に天然状態へフォールドする2状態タンパクや、一度準安定な中間状態へ遷移した後に天然状態へ遷移する多状態タンパクなど、<sup>5,6</sup> そのフォールディング機構は多様である事が示された。また、フォールディングにかかる時間や遷移の回数はタンパクごとで大きく異なるものの、鎖が長くなる程、遷移の回数が増え、フォールディングにかかる時間が長くなるなど一定の規則性が見られる事が示されてきた<sup>7-9</sup>。

一方、フォールディングの順序、すなわち、経路上で形成されている構造の特徴を知る事も重要である。その役割は上述の方法に加え、NMR<sup>10</sup>や $\phi$ 値解析<sup>11,12</sup>が担ってきた。フロー装置と分光学的な手法を用いた実験では、経路上で形成されている2次構造要素の割合や特定残基の埋没度合いなどを知る事が出来る。平衡状態で中間体が存在するタンパクに対しては直接その特徴をNMRにより計測する事ができる。<sup>13</sup> 速度論的な中間体は水素重水素交換と組み合わせる事でその構造情報を得る事が出来る。<sup>14-17</sup> この方法の応用範囲は広く、これまで多数のタンパクの中間体の構造情報が示されてきた。その結果、中間体の段階で部分的に天然構造をとるタンパクが多数存在する事が示された。その中でも天然構造と近い構造を取るが全体的にゆるくパッキングした状態であるモルテングローブルに関する研究が有名であろう<sup>18</sup>。

中間状態が存在せず、2状態でフォールドするタンパクに関してはそのフォールディング径路に関する情報を得る事は困難であったが、1996年にFershtらによって提案された $\phi$ 値解析<sup>11,12</sup>を用いる事でその特徴を間接的に知る事が可能となった。この手法は2状態タンパクのフォールディング機構を知る上で決定的な役割を果たした。<sup>19,20</sup> 特に小さなタンパクにおいて立体構造とフォールディング径路に関する事示す上で $\phi$ 値解析のデータは欠かせないものであった<sup>21-25</sup>。

1997年に Plaxco らがコンタクトオーダーとフォールディング速度との相関関係を示して以降、<sup>26</sup> タンパクの立体構造とその速度論的な特徴には関連性がある事が示されてきた。<sup>8,9,27-29</sup> また、天然構造とフォールディング径路に大きな関係がある事が示されてきた。<sup>21-25,30,31</sup> これらの研究は天然構造の形によってそのフォールディング機構がある程度決まっている事を示唆しており、近年のフォールディングの理解を大きく進めた。

### 1.3 理論的アプローチ

理論的な研究は、Wolynes らのグループを中心に研究が進められてきたエネルギーランドスケープ理論の確立をきっかけに大きく発展した。<sup>32-34</sup> この理論は、タンパクの局所的な構造と全体の構造は矛盾無く全体として安定化しているという、郷の提案した整合性原理<sup>35</sup> をより定量的に示したものと考えて良いだろう。エネルギーランドスケープ理論は小さな速くフォールドするタンパクがエネルギーのフラストレーションの少ないファネル型のエネルギーランドスケープを持つ事を示す。エネルギーのフラストレーションとは、エネルギー的に競合する準安定な構造の数や競合する構造の安定性を示す値であり、フラストレーションが大きいほどタンパクが局所的な構造にトラップされる温度であるガラス転移温度  $T_g$  の値を高くする。エネルギーランドスケープ理論では、 $T_g$  がフラストレーションの度合いに、 $T_f$  が変性状態のアンサンブルと天然状態間のエネルギーの差に比例し、フォールディング温度  $T_f$  が  $T_g$  よりも大きくなる場合にタンパクはフォールディング可能である事を示した。また、 $T_f$  が  $T_g$  と比較してより大きい程フォールディングが速くなる事を示した。

タンパクがファネル型のエネルギーランドスケープを持つ場合、天然構造に近づく程エネルギーは低くなる。しかし、それに伴い構造の自由度が失われていくため、その構造エントロピーは失われていく。その構造エントロピーの減少がエネルギーの減少を上回った場合、自由エネルギーバリアが生じると考えられる。タンパクの構造が複雑であるほど、フォールディングの早い段階でエントロピーが大きく失われるはずであるため、複雑な構造を持つタンパクほどフォールディングが遅くなるはずである。実際に、タンパクの構造が複雑になるほど、フォールディング速度が遅くなる事が知られており、<sup>8,9,26-29</sup> エネルギーランドスケープ理論の正当性を示す根拠の一つとされている。

理論の発展と平行して、計算機の能力の飛躍的な発展に伴い、シミュレーション

ョンを用いた研究も大きく発展してきた。1990年代以前は格子模型を用いたシミュレーションが主流であったが<sup>36-40</sup>、現在は連続空間上でのシミュレーションが主流となっている。<sup>41-44</sup>特に近年は、対象の大きさは限られているものの、水を含めた全原子シミュレーションが可能となっており、<sup>43,45-47</sup> 計算機を用いる事で実験では見る事の出来ないより詳細な描像を仮想的にはあるが観察する事が可能になりつつある。

#### 1.4 適切な計算手法を用いる事の必要性

計算機的能力は向上し続けているものの、計算機上でフォールディングを正確に再現する事は今なお大変な困難を伴う作業である。そのため、計算機上でフォールディング機構を解析する場合、いくつかの情報を捨てる事で計算量を減らす事が必要となる。その一つのアプローチは、タンパクのフォールディングは可逆であるという仮定のもと、タンパクの変性過程のみをシミュレートし、その逆過程をフォールディングの過程として解析を行う方法である。<sup>48,49</sup> もう一つの代表的なアプローチは粗視化モデルを用いる事である。<sup>44,50-52</sup> 粗視化の方法は多種多様であるが、使用するモデルは対象の大きさと得たい結果の解像度に依存する。また、人工の統計集団に基づいて計算を行う事でサンプリングの効率を高める手法も数多く考案されている。<sup>53-55</sup> 上述の通り、限られた計算能力の中でフォールディング機構を解析するための方法は多数提案されているが、状況に合わせて適切な手法を用いる必要がある。

#### 1.5 多状態タンパクの計算機的研究の必要性

これまでの研究の蓄積により、フォールディングに対する概念的な理解は大きく進んできたが、フォールディングの過程で具体的に起きていることは依然として明らかになっていない。<sup>56</sup> それは、タンパクが自由度の高い系である事に加え、フォールディングの過程を実験的に観察することは大変困難である事に起因している。その困難を打開するため、先に述べたような、複数回の遷移を経てフォールドするタンパク（多状態タンパク）の中間体の特徴付けが研究行われてきたが、その中間状態の間の遷移の過程で起きている事は不確かなま



までである。実験での中間状態の特徴付けに加え、シミュレーションで各構造間の遷移に関する得る事でフォールディングの全体的な描像を推測する事が可能である事が示唆されるが、複数回の遷移を経てフォールドするタンパクは大きなタンパクが多いため、多状態タンパクの計算機を用いた研究はまだあまり行われていない。そのため、フォールディングの個々の遷移の過程で何が起きているのかに関しては、これまでに多数の仮説がたてられてきたものの、8,9,57-60未だどの仮説がもっとも正しいのか、はっきりとした結論は得られていない。

## 1.6 研究目的

そこで、本研究では典型的に用いられている粗視化モデルに新たな効果を加える事で、少ない計算コストでより正確なフォールディング機構を再現する事を試みた。また、複数回の遷移を経てフォールドするタンパクに注目し、それぞれの遷移の過程で何が起きているのかを計算機シミュレーションを用いて明らかにする事を目的とした。

第二章では、本研究のために開発したモデルの詳細を述べた。また、そのモデルを小さなタンパクのシミュレーションへ応用した結果を示し、典型的に用いられるモデルとの違いを述べた。第三章では、フォールディングのコアとなる領域が複数存在する事が実験的に示唆されている *ferredoxin-like fold* タンパクのフォールディング機構を詳しく調査し、フォールディングコアの存在の有無を確認した。第四章では、複数のバリアを経てフォールドすることが示唆されている多数のタンパクに対して計算を行い、フォールディングコアの数と重なり具合でどの程度多くのタンパクのフォールディング機構を説明する事が出来るかを調査した。また、フォールディングコアのもつ一般的な特徴を明らかにする事を試みた。第五章では、本研究の結果から推察される、タンパクのフォールディングにおける自由エネルギーバリアの数を決める、分子的なメカニズムを述べた。また、今後解決すべき問題を提示した。

## 参考文献

1. Anfinsen CB, Haber E, Sela M, White FH, Jr. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A* 1961;47:1309-1314.
2. Levintha.C. Are There Pathways for Protein Folding. *J Chim Phys Pcb* 1968;65(1):44-&.
3. Roughton FJW. The kinetics of haemoglobin VII - Some notes on the reactivity of freshly reduced haemoglobin. *P R Soc Lond B-Conta* 1934;115(795):495-503.
4. Hartridge H, Roughton FJW. A method of measuring the velocity of very rapid chemical reactions. *P R Soc Lond a-Conta* 1923;104(726):376-394.
5. Li J, Shinjo M, Matsumura Y, Morita M, Baker D, Ikeguchi M, Kihara H. An alpha-helical burst in the src SH3 folding pathway. *Biochemistry-Us* 2007;46(17):5072-5082.
6. Batey S, Scott KA, Clarke J. Complex folding kinetics of a multidomain protein. *Biophys J* 2006;90(6):2120-2130.
7. Garbuzynskiy SO, Ivankov DN, Bogatyreva NS, Finkelstein AV. Golden triangle for folding rates of globular proteins. *Proc Natl Acad Sci U S A* 2013;110(1):147-150.
8. Kamagata K, Arai M, Kuwajima K. Unification of the folding mechanisms of non-two-state and two-state proteins. *Journal of Molecular Biology* 2004;339(4):951-965.
9. Kamagata K, Kuwajima K. Surprisingly high correlation between early and late stages in non-two-state protein folding. *Journal of Molecular Biology* 2006;357(5):1647-1654.
10. Wuthrich K. Nmr - This Other Method for Protein and Nucleic-Acid Structure Determination. *Acta Crystallogr D* 1995;51:249-270.
11. Fersht AR, Matouschek A, Serrano L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *Journal of Molecular Biology* 1992;224(3):771-782.
12. Matouschek A, Kellis JT, Jr., Serrano L, Fersht AR. Mapping the

- transition state and pathway of protein folding by protein engineering. *Nature* 1989;340(6229):122-126.
13. Hughson FM, Wright PE, Baldwin RL. Structural characterization of a partly folded apomyoglobin intermediate. *Science* 1990;249(4976):1544-1548.
  14. Nishimura C, Wright PE, Dyson HJ. Role of the B helix in early folding events in apomyoglobin: evidence from site-directed mutagenesis for native-like long range interactions. *Journal of Molecular Biology* 2003;334(2):293-307.
  15. Khan F, Chuang JI, Gianni S, Fersht AR. The kinetic pathway of folding of barnase. *Journal of Molecular Biology* 2003;333(1):169-186.
  16. Hu WB, Walters BT, Kan ZY, Mayne L, Rosen LE, Marqusee S, Englander SW. Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc Natl Acad Sci U S A* 2013;110(19):7684-7689.
  17. Raschke TM, Marqusee S. The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nat Struct Biol* 1997;4(4):298-304.
  18. Kuwajima K. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins* 1989;6(2):87-103.
  19. Matthews JM, Fersht AR. Exploring the energy surface of protein folding by structure-reactivity relationships and engineered proteins: observation of Hammond behavior for the gross structure of the transition state and anti-Hammond behavior for structural elements for unfolding/folding of barnase. *Biochemistry-US* 1995;34(20):6805-6814.
  20. Villegas V, Martinez JC, Aviles FX, Serrano L. Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *Journal of Molecular Biology* 1998;283(5):1027-1036.
  21. Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, Baker

- D. Experiment and theory highlight role of native state topology in SH3 folding. *Nat Struct Biol* 1999;6(11):1016-1024.
22. Martinez JC, Serrano L. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat Struct Biol* 1999;6(11):1010-1016.
  23. Chiti F, Taddei N, White PM, Bucciantini M, Magherini F, Stefani M, Dobson CM. Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nat Struct Biol* 1999;6(11):1005-1009.
  24. Clarke J, Cota E, Fowler SB, Hamill SJ. Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway. *Structure* 1999;7(9):1145-1153.
  25. Hamill SJ, Steward A, Clarke J. The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. *Journal of Molecular Biology* 2000;297(1):165-178.
  26. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of Molecular Biology* 1998;277(4):985-994.
  27. Plaxco KW, Simons KT, Ruczinski I, Baker D. Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry-U S A* 2000;39(37):11177-11183.
  28. Ivankov DN, Garbuzynskiy SO, Alm E, Plaxco KW, Baker D, Finkelstein AV. Contact order revisited: influence of protein size on the folding rate. *Protein Sci* 2003;12(9):2057-2062.
  29. Baker D. A surprising simplicity to protein folding. *Nature* 2000;405(6782):39-42.
  30. Nickson AA, Clarke J. What lessons can be learned from studying the folding of homologous proteins? *Methods* 2010;52(1):38-50.
  31. Lappalainen I, Hurley MG, Clarke J. Plasticity within the obligatory folding nucleus of an immunoglobulin-like domain. *Journal of Molecular Biology* 2008;375(2):547-559.
  32. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels,

- pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 1995;21(3):167-195.
33. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* 1997;48:545-600.
  34. Onuchic JN, Wolynes PG. Theory of protein folding. *Curr Opin Struct Biol* 2004;14(1):70-75.
  35. Go N. Theoretical studies of protein folding. *Annu Rev Biophys Bioeng* 1983;12:183-210.
  36. Taketomi H, Ueda Y, Go N. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int J Pept Protein Res* 1975;7(6):445-459.
  37. Ueda Y, Taketomi H, Go N. Studies on Protein Folding, Unfolding, and Fluctuations by Computer-Simulation .2. 3-Dimensional Lattice Model of Lysozyme. *Biopolymers* 1978;17(6):1531-1548.
  38. Go N, Taketomi H. Studies on Protein Folding, Unfolding and Fluctuations by Computer-Simulation .4. Hydrophobic Interactions. *Int J Pept Protein Res* 1979;13(5):447-461.
  39. Go N, Taketomi H. Studies on Protein Folding, Unfolding and Fluctuations by Computer-Simulation .3. Effect of Short-Range Interactions. *Int J Pept Protein Res* 1979;13(3):235-252.
  40. Abkevich VI, Gutin AM, Shakhnovich EI. Specific Nucleus as the Transition-State for Protein-Folding - Evidence from the Lattice Model. *Biochemistry-Ussr* 1994;33(33):10026-10036.
  41. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *Journal of Molecular Biology* 2000;298(5):937-953.
  42. Shimada J, Kussell EL, Shakhnovich EI. The folding thermodynamics and kinetics of crambin using an all-atom Monte Carlo simulation. *Journal of Molecular Biology* 2001;308(1):79-95.

43. Day R, Daggett V. Direct observation of microscopic reversibility in single-molecule protein folding. *Journal of Molecular Biology* 2007;366(2):677-686.
44. Liwo A, Oldziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. A united-residue force field for off-lattice protein-structure simulations .1. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J Comput Chem* 1997;18(7):849-873.
45. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y, Wriggers W. Atomic-level characterization of the structural dynamics of proteins. *Science* 2010;330(6002):341-346.
46. Piana S, Lindorff-Larsen K, Shaw DE. Atomic-level description of ubiquitin folding. *Proc Natl Acad Sci U S A* 2013;110(15):5915-5920.
47. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. *Science* 2011;334(6055):517-520.
48. Wong KB, Clarke J, Bond CJ, Neira JL, Freund SM, Fersht AR, Daggett V. Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *Journal of Molecular Biology* 2000;296(5):1257-1282.
49. Day R, Daggett V. Ensemble versus single-molecule protein unfolding. *Proc Natl Acad Sci U S A* 2005;102(38):13445-13450.
50. Hills RD, Jr., Brooks CL, 3rd. Insights from coarse-grained go models for protein folding and dynamics. *Int J Mol Sci* 2009;10(3):889-905.
51. Clementi C. Coarse-grained models of protein folding: toy models or predictive tools? *Curr Opin Struct Biol* 2008;18(1):10-15.
52. Kmiecik S, Kolinski A. Folding pathway of the b1 domain of protein G explored by multiscale modeling. *Biophys J* 2008;94(3):726-736.
53. Berg BA, Neuhaus T. Multicanonical Algorithms for 1st Order Phase-Transitions. *Phys Lett B* 1991;267(2):249-253.
54. Lyubartsev AP, Martsinovski AA, Shevkunov SV, Vorontsovvelaminov PN. New Approach to Monte-Carlo Calculation

- of the Free-Energy - Method of Expanded Ensembles. *Journal of Chemical Physics* 1992;96(3):1776-1783.
55. Hukushima K, Takayama H, Nemoto K. Application of an extended ensemble method to spin glasses. *Int J Mod Phys C* 1996;7(3):337-344.
  56. Sosnick TR, Barrick D. The folding of single domain proteins--have we reached a consensus? *Curr Opin Struct Biol* 2011;21(1):12-24.
  57. Lindberg MO, Oliveberg M. Malleability of protein folding pathways: a simple reason for complex behaviour. *Curr Opin Struct Biol* 2007;17(1):21-29.
  58. Gianni S, Guydosh NR, Khan F, Caldas TD, Mayor U, White GW, DeMarco ML, Daggett V, Fersht AR. Unifying features in protein-folding mechanisms. *Proc Natl Acad Sci U S A* 2003;100(23):13286-13291.
  59. Xu M, Beresneva O, Rosario R, Roder H. Microsecond folding dynamics of apomyoglobin at acidic pH. *J Phys Chem B* 2012;116(23):7014-7025.
  60. Vu DM, Brewer SH, Dyer RB. Early Turn Formation and Chain Collapse Drive Fast Folding of the Major Cold Shock Protein CspA of *Escherichia coli*. *Biochemistry-U S* 2012.





## 第2章 側鎖の効果を取り込んだ C $\alpha$ G $\ddot{o}$ モデルの開発

### 2.1 背景

ファネル型のエネルギーランドスケープを再現する事でフォールディングのプロセスを簡便にシミュレートする事のできるモデルを G $\ddot{o}$  モデルと呼ぶ。G $\ddot{o}$  モデルはフォールディングのシミュレーションに用いるモデルとして大きな役割を担ってきた。G $\ddot{o}$  モデルは2次元の格子模型でタンパクのフォールディングをシミュレートする際、郷らによって初めて用いられたが、<sup>1</sup>その後 Clementi らの3次元空間におけるシミュレーションによって多数のタンパクのフォールディングメカニズムを再現できる事が示されてきた。<sup>2</sup>これまで、単純な G $\ddot{o}$  モデルを用いて多数のタンパクのフォールディング機構を再現する事が出来る事に加え、トポロジーとフォールディング速度との関係、<sup>3,5</sup> 中間体の有無<sup>2,6,7</sup> などが解析されてきた。G $\ddot{o}$  モデルを用いたフォールディングシミュレーションがうまく行く事は、実際のタンパクがファネル型のエネルギーランドスケープを持つことを意味する。

G $\ddot{o}$  モデルを用いる大抵の場合、残基間相互作用の強さは一定として扱われる。しかし、残基間相互作用の強さのばらつきをモデルに含め、アミノ酸の種類を区別して計算する事で、計算の正確さが向上する事が Brooks らによって示されてきた。<sup>7,8</sup> このようなモデルは度々“フレーバーモデル”と呼ばれる。<sup>9</sup> 大抵のフレーバーモデルは天然構造のデータベースの統計情報を用いてパラメータが決定される。<sup>8,10</sup> また、すべての重原子を陽に取り扱う G $\ddot{o}$  モデルも存在し、全原子 G $\ddot{o}$  モデルと呼ばれている。<sup>11-14</sup> 全原子 G $\ddot{o}$  モデルの場合は原子間のポテンシャルを考慮するため、残基間の相互作用エネルギーの多様性は定義に含まれている。但し、全原子 G $\ddot{o}$  モデルの場合は相互作用の強さに加え、側鎖の形や向きなどの多様性もモデルに含まれる。

より詳細なモデルの利用はより実験値に近い値を再現するが、モデルが複雑になるほど計算コストは拡大する。従って、対象とする問題に合わせて適切なモデルを選択しなければ必要な情報を得る事ができないだろう。適切なモデルの選択を可能とするためにはそれぞれのモデルに含まれている特徴がどのような効果を持つのかを明らかにしておく必要がある。

興味深い事に、Clementi らは C $\alpha$ G $\ddot{o}$  モデルや全原子 G $\ddot{o}$  モデルの結果を比較する事で、フォールディングにおける側鎖の役割を二つ示した。一つの役割は残基間相互作用の強さのばらつきに依存してフォールディング経路に偏りを持

たせる事である。もう一つは相互作用形成時の鎖の構造エントロピーの減少量を増大させ、フォールディングの協同性を上昇させる事である。構造エントロピーの減少量の増大は、相互作用形成時に  $C\alpha$  原子の位置が拘束されるだけでなく、鎖の配向も固定されるために生じる。これら2つの効果は通常の  $C\alpha G\ddot{o}$  モデルには含まれていないが、フォールディング機構をより正確に知るためには重要であると考えられる。

いくつかの研究では、粗視化  $G\ddot{o}$  モデルにおいて残基間相互作用のばらつきを考慮する事は、より正確なフォールディング経路を再現するが自由エネルギーバリアの高さを下げる事が示されている。<sup>10</sup> 他方、側鎖の配向に起因した鎖エントロピーの減少の効果はこれまで議論されていない。通常の  $C\alpha G\ddot{o}$  モデルを用いた場合、実験的に予想されるより低い協同性を示す事が報告されているため、<sup>4,15</sup> より正確にフォールディング機構を再現するためには側鎖の配向も考慮するモデルを用いる必要があるだろう。

そこで本研究では残基間相互作用の強さのばらつきだけではなく、側鎖の配向を効果的に区別する拘束を  $C\alpha$  原子間に付加した。そしてプロテインG、プロテインLのフォールディング機構を解析し、全原子  $G\ddot{o}$  モデルの結果と比較した。その結果、全原子  $G\ddot{o}$  モデルと類似した結果を得た。また、残基間相互作用エネルギーのばらつきはもっともらしいフォールディング経路を再現するが自由エネルギーバリアの高さを下げる事を、側鎖の方向を考慮した拘束は協同性を高める事を確認した。同様に、all- $\beta$ タンパクであるsrc-SH3 domain、all- $\alpha$ タンパクであるprotein Aにも応用し、相互作用エネルギーのばらつきと、側鎖の配向を考慮した拘束の効果がトポロジーに依存して変化する事を明らかにした。

## 2.2 方法

### モデル

本研究では結合長を  $3.8 \text{ \AA}$  に固定した  $C\alpha$  モデルを用いた。モデルの概要を図 1 に示す。

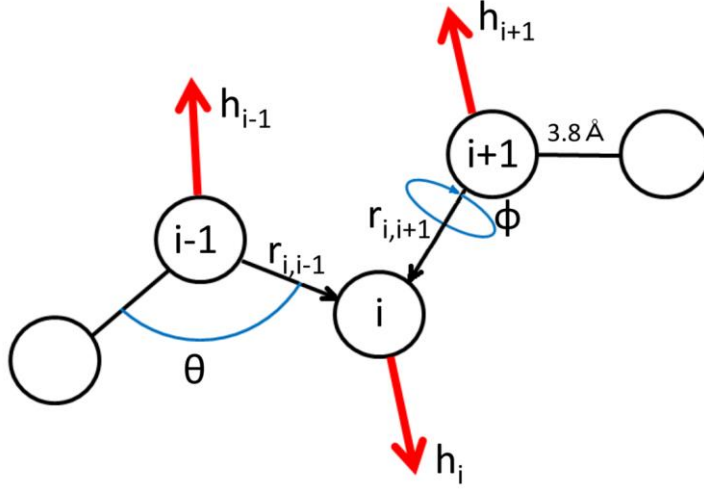


図 1: 本研究で用いたモデルの概要。本研究ではタンパクを  $C\alpha$  原子の連なった鎖で表現し、それらの間の結合角、回転角に自由度を持たせた。 $C\alpha$  原子間の距離は  $3.8 \text{ \AA}$  とした。側鎖の相対的な配向を効果的に区別するために、 $\mathbf{r}_{i,i-1} + \mathbf{r}_{i,i+1}$  の 2 つのベクトルから定義されるベクトル  $\mathbf{h}_i$  を用いた。

構造  $\Gamma$  の持つエネルギーは式(1)にて定義される。

$$\begin{aligned}
 E(\Gamma, \Gamma_0) = & \sum_{\text{angles}} K_\theta (\theta_i - \theta_{i0})^2 \\
 & + \sum_{\text{dihedral}} \{K_\phi^1 [-\cos(\phi_i - \phi_{i0})] + K_\phi^3 [-\cos 3(\phi_i - \phi_{i0})]\} \\
 & + \sum_{ij}^{NC} \varepsilon C_{ij} \left[ 5 \left( \frac{r_{ij0}}{r_{ij}} \right)^{12} - B_{ij} \cdot 6 \left( \frac{r_{ij0}}{r_{ij}} \right)^{10} \right] \\
 & + \sum_{ij}^{NNC} \varepsilon \left( \frac{4}{r_{ij}} \right)^{12} \tag{1}
 \end{aligned}$$

この数式は典型的な  $G\ddot{o}$  ポテンシャルに改良を加えたものである。第一、第二、第三、第四項目はそれぞれ、結合角、回転角、天然相互作用、非天然相互作用に関するエネルギーであり、 $\theta$ 、 $\phi$ 、 $r_{ij}$ 、NC、NNC は結合角、回転角、残基間距離、天然コンタクト、非天然コンタクトを示す。下付の 0 は天然構造での値を意味する。 $K_\theta = 20 \epsilon$ 、 $K_\phi^1 = \epsilon$ 、and  $K_\phi^3 = 0.5 \epsilon$  とした。重原子間距離が  $4 \text{ \AA}$  以内のペアを 1 組以上持つ残基ペアを天然コンタクトを形成している残基ペアとした。三項目の  $C_{ij}$ 、 $B_{ij}$  は本研究のモデルに固有のパラメータである。 $C_{ij}$  は重原子間コンタクトの数をその平均値で割ったものであり、スケールされた残基間相互作用の強さを示す。 $B_{ij}$  は側鎖の方向を効果的に区別する事で  $C\alpha$  原子を拘束する変数であり、以下の式(2)、(3)で定義される。

$$B_{ij}(\Theta_{ij}) = \begin{cases} 1 - (\Theta_{ij} - \Theta_{ij0})^2 / a_\Theta^2 & , \quad \text{if } \Theta_{ij0} - a_\Theta < \Theta_{ij} < \Theta_{ij0} + a_\Theta \\ 0 & , \quad \text{otherwise} \end{cases} \quad (2)$$

$$\Theta_{ij} = \arccos\left(\frac{\mathbf{h}_i \cdot \mathbf{h}_j}{|\mathbf{h}_i| |\mathbf{h}_j|}\right) \quad (3)$$

$\mathbf{h}_i = \mathbf{r}_{i,i-1} + \mathbf{r}_{i,i+1}$  であり、 $\mathbf{r}_{i,i-1}$  は  $i$  番目の残基と  $i-1$  番目の残基間のベクトルである。 $i$  番目と  $j$  番目の残基の側鎖間の相対的な角度は  $\mathbf{h}_i$  と  $\mathbf{h}_j$  の相対的な角度である  $\Theta_{ij}$  を用いて区別される。 $\mathbf{h}_i$  は  $C\beta$ - $C\alpha$  間のベクトルを表現するために  $\mathbf{r}_{i,i-1} \times \mathbf{r}_{i,i+1}$  と組み合わせられて用いられる。 ${}^{16}B_{ij}(\Theta_{ij})$  は  $i$  番目と  $j$  番目の残基の相対的な角度が天然構造にどの程度近いかを示すパラメータである。 $\Theta_{ij}$  が天然構造に近い場合は、 $B_{ij}$  は 1 になり、 $i$  と  $j$  が相互作用可能となる。 $\Theta_{ij}$  がカットオフ値より大きい場合は  $B_{ij}$  は 0 となり、残基  $i$  と  $j$  は相互作用できなくなる。 $\Theta_{ij}$  のカットオフ値は  $0.6\pi$  とした。この値は **protein L** の自由エネルギープロファイルにおける天然状態の位置と変性状態の位置が全原子  $G\ddot{o}$  モデルでの位置と対応するように決定した。末端の残基は  $\mathbf{h}_i$  を定義する事が出来ないため  $B_{ij} = 1$  とした。また、 $\mathbf{h}_i = 0$  となる事を防ぐため、常に  $\theta < \pi$  となるようにした。

相互作用エネルギーの多様性と側鎖の配向を効果的に考慮した拘束のフォールディングに及ぼす効果を確認するため、各タンパクにおいて次の 4 種類のシミュレーションを行い、結果を比較した。(1)  $C_{ij} = 1$  かつ  $B_{ij} = 1$  : 通常の  $C\alpha G\ddot{o}$  モデル。以後 CA と表記する。(2)  $0 \leq B_{ij} = B_{ij}(\Theta_{ij}) \leq 1$  かつ  $C_{ij} = 1$  : 側鎖の配向のみを区別したモデル。以後 SO (Side chain Orientation) と表記する。(3)  $B_{ij} = 1$  かつ  $C_{ij} =$  スケーリングされたコンタクト数 : 相互作用エネルギーの多様性のみを考慮したモデル。以後 CH

(Contact Heterogeneity)と表記する。(4)  $0 \leq B_{ij} = B_{ij}(\Theta_{ij}) \leq 1$  and  $C_{ij}$  = スケールリングされたコンタクト数: 側鎖の配向と相互作用エネルギーの多様性の両方を考慮したモデル。以後 SOCH と表記する。

## シミュレーション

本研究では構造空間をサンプリングするため、それぞれのタンパクにて一連の温度にて等温モンテカルロシミュレーションを行った。Protein L では、温度パラメータ ( $k_B T / \epsilon$ ) 0.750-1.000 の間で 27-30 個の異なる温度で、protein G では温度パラメータ 0.650-0.850 の間で 27 個の異なる温度で、protein A では、温度パラメータ 0.650-0.890 の間で 25-29 個の異なる温度で、src-SH3 domain では温度パラメータ 0.680-0.890 の間で 24-26 個の異なる温度でシミュレーションを行った。1 ステップにはランダムに選んだ残基に対する pivot move  $M$  回、ランダムに選んだ領域に対する crankshaft move  $M$  回、計  $2M$  回のメトロポリステストを含む。crankshaft move にて動かす領域の大きさは配列長の半分を超えないようランダムに選択した。 $M$  は配列長とした。 $10^6$  ステップの平衡化の後、 $5 \times 10^7$  ステップの計算を行った。一連のシミュレーションの後、Weighted Histogram Analysis Method (WHAM 法)<sup>17-19</sup> を用いて自由エネルギーを計算した。最も高い自由エネルギーバリアを生じた protein L の SO モデル、SOCH モデルにおいて、異なるランダムシードを用いて 5 回のシミュレーションを行った結果から得られた自由エネルギープロファイルの標準誤差の平均は  $1.5 \times 10^{-2}$ 、 $5.4 \times 10^{-3}$  であり、非常に小さな値であったため、エラーバーは記載しなかった。他の結果は遷移温度での計算のトラジェクトリから計算した。遷移温度は熱容量曲線のピークとした。

## ターゲットタンパク

Clementi らの全原子 Gō モデルシミュレーションの結果<sup>11</sup>と比較するため、本研究では immunoglobulin binding domain of peptostreptococcal protein L (protein L)、immunoglobulin binding domain of streptococcal protein G (protein G) のフォールディング機構を解析した。これらのタンパクは実験的にも計算機的にもよく研究されている。また、同じトポロジーを持つが異なる経路を経てフォールドする事が知られている。また、異なるトポロジーを持つタンパクでも解析を行うた

め、all- $\alpha$ タンパクである immunoglobulin binding B domain of *Staphylococcus aureus* protein A (protein A)、all- $\beta$ タンパクである src-tyrosine kinase SH3 domain (src-SH3 domain)のフォールディング機構も調査した。対象タンパクの実験的に示されているフォールディングの特徴を簡単に記載する。

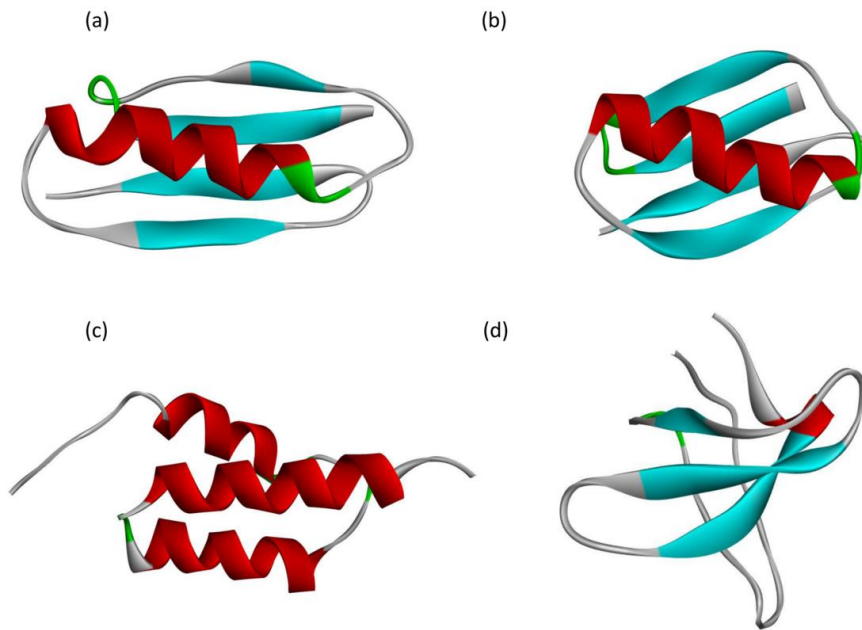


図 2: 本研究の対象タンパクの構造。(a)、(b)、(c)、(d)はそれぞれ protein L、protein G、protein A、src-SH3 domain の構造を示す。

### protein L

protein L の構造を図 2(a)に示す。protein L は配列の中心に位置する $\alpha$ ヘリックスとその両側の $\beta$ ヘアピンから構成される。また、2つの $\beta$ ヘアピンは4本のストランドから成る $\beta$ シートを構成する。各二次構造要素の位置は Clementi らに合わせ<sup>11</sup>、21-25 残基目をストランド 1、31-35 残基目をストランド 2、42-53 残基目をヘリックス、51-54 残基目をストランド 3、72-76 残基目をストランド 4 とした。protein L は実験的に、N 末端の $\beta$ ヘアピンのフォールディングを起点に 2 状態でフォールドする事が示されている。<sup>20,21</sup>

## protein G

protein G の構造を図 2(b)に示す。protein G は protein L と同じトポロジーをもつが、protein L と異なり C 末端からフォールドする事が示されている。<sup>22,23</sup> 2 次構造の割当は 2-7 残基目をストランド 1、14-19 残基目をストランド 2、22-26 残基目をヘリックス、42-46 残基目をストランド 3、51-55 残基目をストランド 4 とする。また、C 末端のヘアピンは単体で安定な構造を取る事が示されている。

24,25

## protein A

図 2(c)に protein A の構造を示す。protein A は 3 本の $\alpha$ ヘリックスから成る。本研究では N 末端から順番にヘリックス 1、ヘリックス 2、ヘリックス 3 と呼ぶ。2 次構造の割当は PDB の定義に従い、10-19 残基目をヘリックス 1、25-37 残基目をヘリックス 2、42-55 残基目をヘリックス 3 とする。protein A は 2 状態でフォールドし、遷移状態にてヘリックス 2 とヘリックス 3 の一部が形成されている事が示されている。<sup>26</sup>

## src-SH3 domain

src-SH3 domain の構造を図 2(d)に示す。src-SH3 domain は 3 本のストランドから成る 2 組の $\beta$ シートがパッキングして形成されている。一方の $\beta$ シートはストランド 1、ストランド 2、ストランド 5 から、もう一方の $\beta$ シートはストランド 2、ストランド 3、ストランド 4 で構成されている。2 次構造の割当は PDB の定義に従い、10-13 残基目をストランド 1、31-35 残基目をストランド 2、42-47 残基目をストランド 3、52-57 残基目をストランド 4、61-63 残基目をストランド 5 とした。ストランド 1、ストランド 2 はそれぞれ RT-loop<sup>27</sup> と呼ばれる領域の N 末端と C 末端である。ストランド 2 の C 末端とストランド 3 の N 末端は n-src loop<sup>27</sup> と呼ばれている。同様にストランド 3 の C 末端とストランド 4 の N 末端は distal loop<sup>27</sup> と呼ばれている。src-SH3 domain は distal loop を中心とした領域を起点に 2 状態でフォールドする事が示されている。<sup>27,28</sup> しかし、いくつかの研究は中間体の存在を示唆している。<sup>29</sup>

## PDB 構造

シミュレーションには PDB に登録されている NMR 構造を用いた。Protein L は 2PTL の 1 番目の構造を、protein G は 2GB1 を、protein A は 1BDD を、src-SH3 domain は 1SRL を天然構造として用いた。

## 2.3 結果

### protein L の遷移状態付近における自由エネルギープロファイル

まず、本研究ではポテンシャル関数の各パラメータが protein L のフォールディング機構へ及ぼす影響を調査した。主な結果の解析には反応座標として、シミュレーション中に形成されている天然コンタクトの割合を表す  $Q$  値を用いた。シミュレーション中の構造において、天然構造でコンタクトしている残基ペアが天然距離の 1.2 倍以下になったとき、その残基ペアを天然コンタクトとみなした。すなわち  $r_{ij} \leq r_{ij0} \times 1.2$  とした。 $Q = 1$  はすべての天然コンタクトが形成されている事を意味し、 $Q = 0$  は天然コンタクトが全く形成されていない事を意味する。



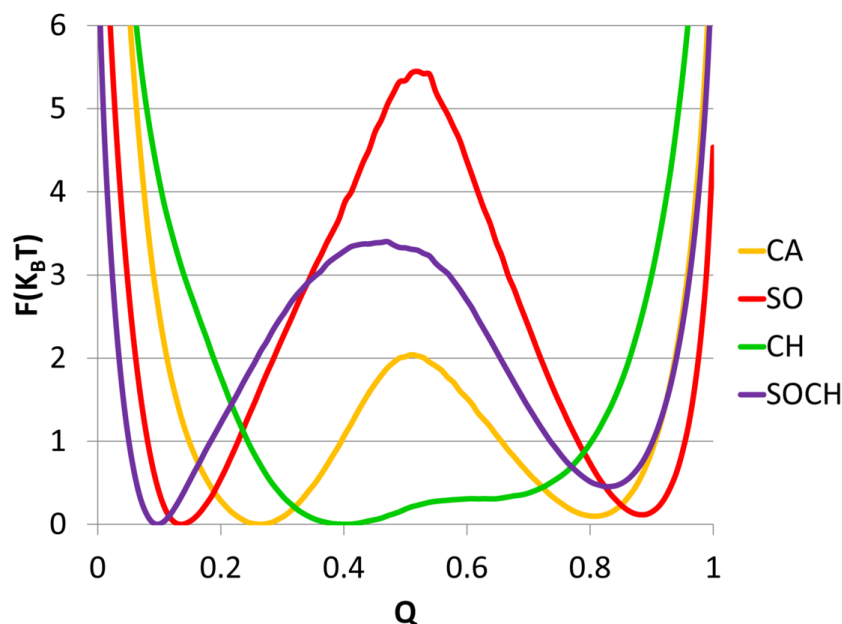


図 3: 遷移温度  $T_t$  での protein L の自由エネルギープロファイル。横軸は  $Q$  値を示し、縦軸は自由エネルギーを  $k_B T$  で割った値を示す。CA、SO、CH、SOCH のそれぞれのモデルでは 2、5.5、0、3  $k_B T$  の高さをもつ自由エネルギーバリアを再現した。それぞれのモデルの遷移温度( $k_B T_t$ )は 0.855 (CA)、0.842 (SO)、0.912 (CH)、0.897 (SOCH)であった。

遷移温度  $T_f$  における 4 つのモデルから計算された自由エネルギープロファイルを図 3 に示した。CA、SO、CH、SOCH のそれぞれのモデルでの遷移温度  $T_f$  は 0.855、0.842、0.912、0.896 であった。自由エネルギープロファイルの形はそれぞれのモデルごとに大きく異なったため、各モデルのパラメータとの対応関係を明確にする。 $C_{ij}$  の定義が異なる CA モデル、CH モデルの結果を比較すると、自由エネルギーバリアの高さと、変性状態・天然状態の位置が大きく異なった。CA モデルでは 2  $k_B T$  程度のバリアが観察されたが、CH モデルではバリアはほぼ観察されなかった。これは Cho らによって指摘されている通り<sup>10</sup>、残基間相互作用のエネルギーに大きなばらつきがある場合は自由エネルギーバリアの高さが減少する事を示す。一方、CA モデルと  $B_{ij}$  の定義が異なる SO モデルでは CA モデルより大きな 5.5  $k_B T$  の自由エネルギーバリアを生じた。また、SO モデルの変性状態の  $Q$  値は CA モデルの値より低く、SO モデルの天然状態の  $Q$  値は CA モデルよりも高くなった。この結果は側鎖の配向を考慮した  $C\alpha$  原

子に対する拘束が協同性を高めた事を示す。SOCH モデルは SO モデル、CH モデルの組み合わせであるため、自由エネルギーバリアの高さは SO モデルと CH モデルの間であった。SOCH モデルは残基間相互作用に強いばらつきがあったとしても一定の協同性を保ってフォールドする。

本研究の CH モデルにおける天然コンタクトの定義は Clementi らの全原子 Gō モデルと同じである。しかし、全原子 Gō モデルでは  $4 k_B T$  のバリアをもつのにに対し、本研究の CH モデルの自由エネルギーバリアはほぼ 0 であり、両者の自由エネルギーバリアの高さは異なっている。この違いには、全原子 Gō モデルには含まれていて CH モデルには含まれていない、側鎖の配向に関するエントロピーの変化が大きく関わっていると考えられる。天然コンタクトの定義が CH と同じであるが側鎖の配向を考慮した拘束を含んでいる SOCH モデルにて自由エネルギーバリアが回復した事も側鎖の配向が自由エネルギーバリアの形成に関わっている事を強く示唆する。

## protein L の 2 次構造形成順序

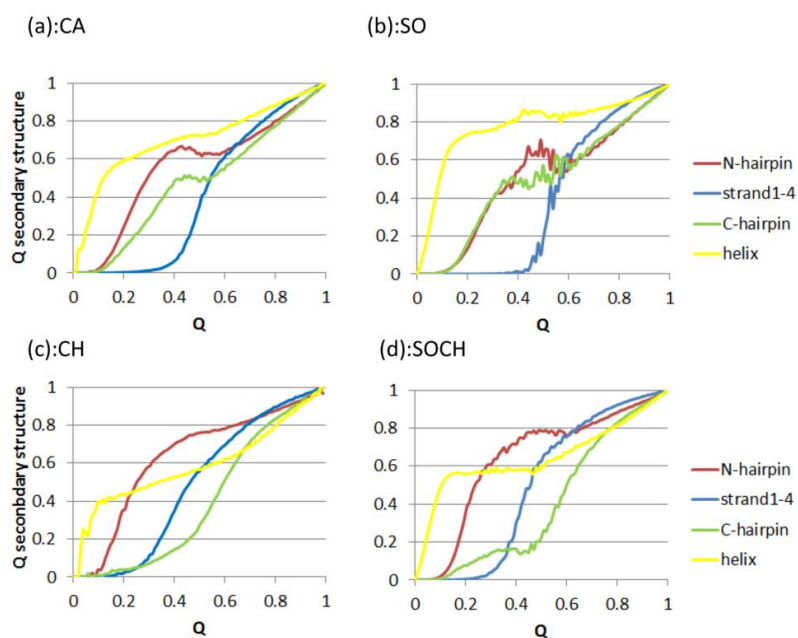


図 4: 遷移温度における protein L のシミュレーション結果から計算した二次構形成度合いのプロファイル。

(a)、(b)、(c)、(d)はそれぞれ CA、SO、CH、SOCH モデルの結果を示す。

図 4(a-d)に protein L の  $Q$  値に対する各 2 次構造内、2 次構造間コンタクトの形成度合いを示した。このプロファイルは任意の  $Q$  値における各コンタクトの形成度合い  $\langle Q_i(Q) \rangle$  の 2 次構造要素内における平均を示している。 $\langle Q_i(Q) \rangle$  は式(4)により定義される。

$$\langle Q_i(Q) \rangle_T = \langle \delta(Q_i - 1) \delta \left( \sum_{i=1}^M \frac{Q_i}{M} - Q \right) \rangle \quad (4)$$

あるスナップショットでコンタクト  $i$  が形成されている場合、 $Q_i$  は 1 である。コンタクトが形成されていない場合  $Q_i$  は 0 となる。 $M$  は全天然コンタクト数を示す。このデータを基に各モデルにおけるフォールディング経路を解析し、その違いを比較した。

図 4 では全てのデータにおいてヘリックスと N 末端のヘアピンが  $Q$  値の低い領域、すなわち、フォールディングの初期にフォールドする事を示す。しかし、他の領域においては各モデルの間でいくつかの違いが見られる。 $C_{ij}$  が全ての残基間相互作用において等しい CA モデル、SO モデルでは、ストランド 1-4 間の相互作用が形成される前に C 末端ヘアピンが形成される。しかし、相互作用の強さ  $C_{ij}$  が各残基ペアごとに異なる CH モデル、SOCH モデルではそれらのフォールドの順序は逆転し、C 末端のヘアピンが形成される前にストランド 1-4 間の相互作用が形成される。

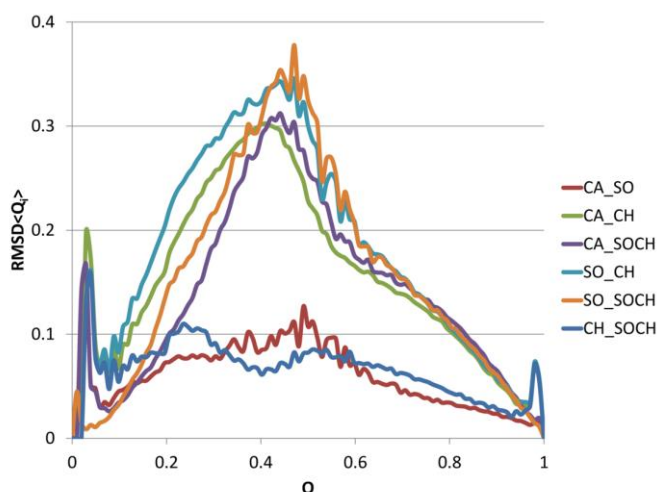


図 5: 各モデルの間で計算した  $\langle Q_i(Q) \rangle$  の RMSD のプロファイル。この値が小さい場合は類似した経路を辿ってフォールドしている事を意味する。

フォールディング経路に強い影響を与えているパラメータを明確にするために、モデル間のフォールディング経路のずれの大きさを定量した。各 Q 値における2つのモデルの間の構造形成度合いの違いの大きさを図 5 に示した。図 5 の各プロファイルは式(5)に示されるように $\langle Q_i(Q) \rangle$ の RMSD から計算される。

$$RMSD(Q)_{model\_a,model\_b} = \sqrt{\sum_i^M \left( \langle Q_i(Q) \rangle_{T_{fa,model\_a}} - \langle Q_i(Q) \rangle_{T_{fb,model\_b}} \right)^2 / M} \quad (5)$$

M は天然相互作用の総数を示す。図 5 より、CA モデルと SO モデル間の RMSD と CH モデル、SOCH モデル間の RMSD は他のプロファイルと比較して明らかに低い値を取る。CA モデルと SO モデルは相互作用エネルギーにばらつきが無いモデルであるが、CH と SOCH モデルは相互作用エネルギーにばらつきがあるモデルである。この結果は protein L のフォールディング経路の違いは相互作用エネルギーのばらつきによって生じている事を示す。

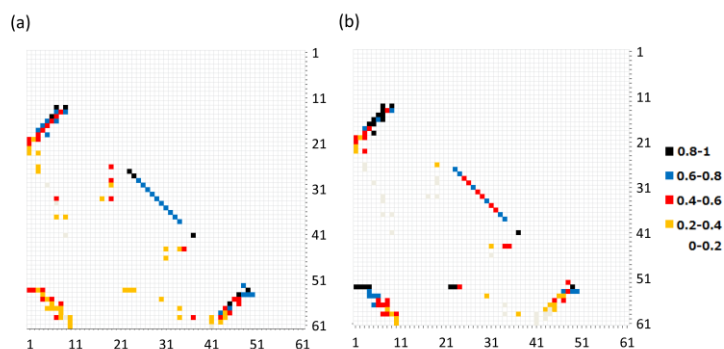


図 6: CA モデル(a)、SOCH モデル(b)から計算された protein L の遷移状態アンサンブルのコンタクト頻度のマップ。縦軸と横軸は残基番号を示す。各点の色はそのコンタクトが遷移状態で形成されている確率を示す。

図 6 のコンタクトマップは CA モデルと SOCH モデルにおける遷移状態アンサンブルの特徴を示す。図 3 に示された自由エネルギープロファイルのピークを遷移状態として定義した。点(i,j)における色の違いは遷移状態のアンサンブル中で残基 i と j の間の相互作用が形成されている確率を示す。

CAモデルにおけるコンタクトマップである図 6(a)では対角線に近い領域に色の濃い点が集中しているのに対し、SOCHモデルにおけるコンタクトマップである図 6(b)では対角線からはなれた位置にも色の濃い点が存在していた。これらは、CAモデルは配列上の距離の短い相互作用から順番に形成されているが、SOCHモデルでは配列上の距離の遠い相互作用も、それらが強い相互作用である場合は、フォールディングの初期に形成されている事を示す。protein L はシンメトリな構造を持つタンパクであるが、そのフォールディング経路はシンメトリではなく N 末端に偏った領域から形成される事が実験的に示されている。<sup>20,21</sup>CAモデル、SOモデルでは N 末端ヘアピンと C 末端ヘアピンがほとんど同時に形成されるが、CHモデル、SOCHモデルでは N 末端ヘアピンのフォールドが C 末端ヘアピンより早く起きる。CHモデル、SOCHモデルの方がより実験的に得られているデータと近い結果を再現している。また、本研究で用いた相互作用エネルギーのばらつきの強さは protein L のフォールディング経路に偏りを持たせるために十分な強さである事が示された。

我々の SOCHモデルに取り込んだ効果が全原子 Gōモデルの特徴をどの程度再現しているかを明らかにするために、我々の SOCHモデルの計算結果と全原子 Gōモデルを比較した。文献 21 の図 3(b)では  $Q < 0.1$  の段階でヘリックスの構造が約 60%、 $Q = 0.1-0.3$  の間に N 末端ヘアピンの構造が約 80%、 $Q = 0.4-0.6$  の間でストランド 1-4 間の相互作用が約 80%形成され、 $Q = 0.4$  以降に徐々に C 末端ヘアピンが形成されてくことが示されている。一方、SOCHモデルでは  $Q < 0.1$  の段階でヘリックスの構造が約 50%、 $Q = 0.1-0.3$  の間で N 末端ヘアピンが約 70%、 $Q = 0.4-0.6$  の間でストランド 1-4 間の相互作用が約 80%形成され、 $Q = 0.4$  以降に C 末端ヘアピンの構造が徐々に形成される事が示された。両者は類似したフォールディング経路を示した。図 6(b)に示された遷移状態構造のコンタクト頻度マップと本研究の SOCHモデルで得られた遷移状態のコンタクト頻度マップを比較しても、Phe23-Tyr53 間の相互作用や N 末端のクラスタが共通して強い相互作用を形成しており、非常に類似した結果を示した。これらは本研究で取り入れた特徴が全原子 Gōモデルの性質を再現する上で十分な効果を持つ事を意味する。

しかし、本研究の結果と文献 21 の結果ではいくつかの違いが見られた。例えば、本研究の SOCHモデルの結果では C 末端ヘアピンがフォールディングの初期に全原子 Gōモデルの結果と比較してやや多くの部分構造を形成していた。ま

た、後述するパーシャルコンタクトオーダー (CO(Q)) やルートメジャー(R(Q)) のプロファイルも異なる形であった。

但し、これらの違いは残基 49-52 で定義される 2 面角を調整する事で取り除く事が出来た。残基 50、51 付近のターンは天然構造において、やや不安手な左巻きヘリックスの 2 面角を形成している。しかし、CaGō モデルではその特徴は考慮されず、天然構造で左巻きヘリックスの 2 面角を取っていたとしても他の領域と同じ安定性を示す。その違いを解消するため、49-52 の 4 残基で定義される 2 面角では  $\phi_0$  の値を右巻きヘリックスの値とし、左巻きヘリックスの構造がやや不安定になるよう調整した。その結果、図 7(a) に示されるように、低い Q 値での C 末端ヘアピンの部分的な形成がなくなり、それぞれの 2 次構造形成のプロファイルが通常の SOCH モデルの結果より協同的な遷移を示すようになった。

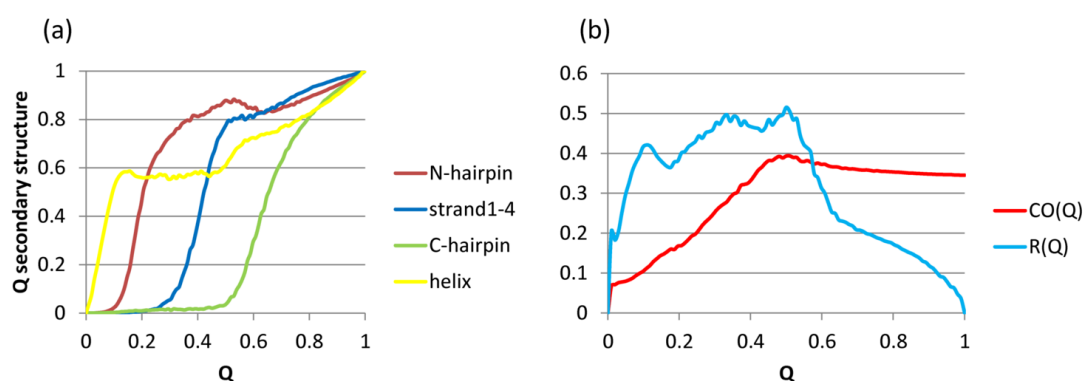


図7: SOCH モデルの 49-52 残基目で定義される 2 面角の天然の値を右巻きヘリックスの値に置き換えた時の (a) 2 次構造形成度合い、(b) CO(Q)、R(Q) のプロファイル。図 3 と比較すると、個々の 2 次構造の形成はより協同的になり、 $Q < 0.5$  の領域で C 末端ヘアピンの部分構造の形成がほぼ観察されなくなった。CO(Q) の値は  $Q = 0.5$  付近で  $Q = 1$  の値より大きな値を取り、R(Q) の値は 3 つのピークを持つ。

Clementi らは protein L の全原子 Gō モデルシミュレーションにおいて、パーシャルコンタクトオーダー (CO(Q)) と呼ばれる値が  $Q = 0.4$  付近で天然での CO(Q) の値を上回る、すなわち、 $CO(0.4) > CO(1)$  となる事を示した。CO(Q) の値は任意の Q 値をもつアンサンブルにおいて形成されている相互作用のもつ平均の配列上距離を示し、式(6)によって定義される。

$$CO(Q) = \frac{\sum_{i=1}^M L_i \langle Q_i(Q) \rangle}{NM} \quad (6)$$

$L_i$  は相互作用している残基間の配列上距離を示す。ただし、オリジナルの定義とは異なり、本研究では CO(Q) の値は配列長で割り、規格化した値を用いている。同様に、Clementi らはフォールディング経路の偏りを定量するルートメジャー (R(Q)) という値を式(7)を用いて定義した。

$$R(Q) = \frac{\sum_{i=1}^M \langle (\langle Q_i \rangle - Q)^2 \rangle}{MQ(1-Q)} \quad (7)$$

彼らの全原子 Gō モデルの結果では protein L におけるルートメジャーの値は 3 つのピークを持つ。本研究において、残基 49-52 で定義される 2 面角の値を調整したモデルを用いた CO(Q)、R(Q) の計算結果を図 7(b) に示した。図 7(b) の CO(Q) のプロファイルは遷移状態付近で天然の値より大きな値を取り、R(Q) の値は 3 つのピークを持つ。従って、本研究の SOCH と全原子 Gō モデルとの間に見られた不一致も、49-52 間の残基で定義される 2 面角を調整する事で解消された。

### protein G、protein L のフォールディング経路の比較

方法の項で記載した通り、protein G は protein L と同じトポロジーを持つが、protein L とは異なる経路を経てフォールドする事が示唆されている。<sup>22,23</sup> これらの違いは通常の Gō モデルでは区別する事が出来ない。本モデルがトポロジーのみでは識別できない特徴を再現する事が出来るかどうかを確認するため、我々のモデルが protein L、protein G のフォールディング機構の違いを識別できるかどうかを確認した。

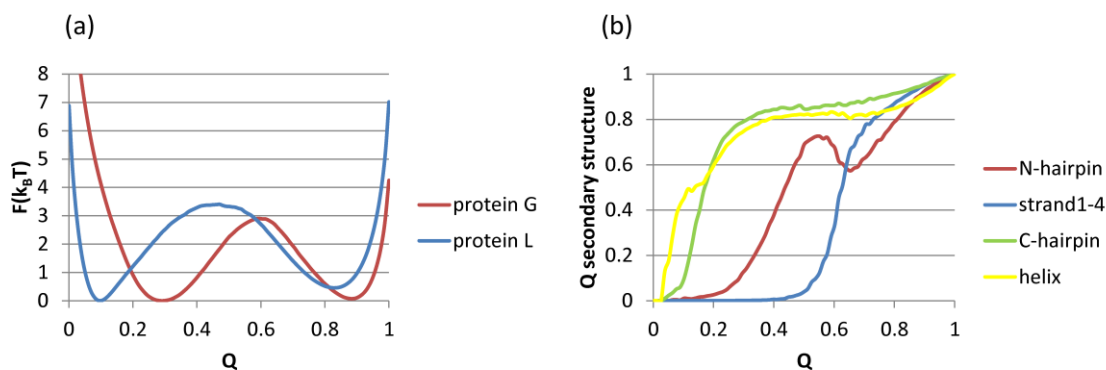


図 8: (a) SOCH モデルを用いて計算した protein G と protein L の自由エネルギープロファイル。自由エネルギーバリアの高さはほとんど等しいが、protein G の遷移状態や変性状態の位置は protein L と比較して高い Q 値を持つ。(b) protein G の二次構造形成度合いのプロファイル。

図 8(a)には protein G、protein L の遷移温度の自由エネルギープロファイルを、図 8(b)には遷移温度のトラジェクトリから計算した protein G のフォールディング経路を記載した。図 8(b)より、protein G は protein L とは異なり C 末端ヘアピンが N 末端より早くフォールドする事がわかる。この結果は実験的な報告<sup>22,23</sup>と一致する。図 8(a)の自由エネルギープロファイルにおいて protein G の変性状態の位置は protein L の変性状態と比較して高い Q 値に位置する。また、図 8(b)のプロファイルと比較すると、protein G の変性状態に対応する Q 値では protein G の C 末端はほぼフォールドしていることが示された。この結果も、protein G の C 末端ヘアピンは単体でも安定に存在できる事<sup>24,25</sup>と矛盾しない。

### protein A、src-SH3 domain のフォールディング機構

ここまでの結果では、protein L、protein G を用いて、本研究で C $\alpha$ G $\delta$  モデルに導入したパラメータである B<sub>ij</sub>、C<sub>ij</sub> の効果を確認してきた。しかし、protein L も protein G も  $\alpha$ + $\beta$ タンパクである。ここでは他のトポロジーをもつタンパクのフォールディング経路に B<sub>ij</sub> と C<sub>ij</sub> がどのような影響を及ぼすかを明らかにするため、all- $\alpha$ タンパクである protein A と all- $\beta$ タンパクである src-SH3 domain のフォールディング機構を CA、SO、CH、SOCH の 4 つのモデルを用いて解析した。



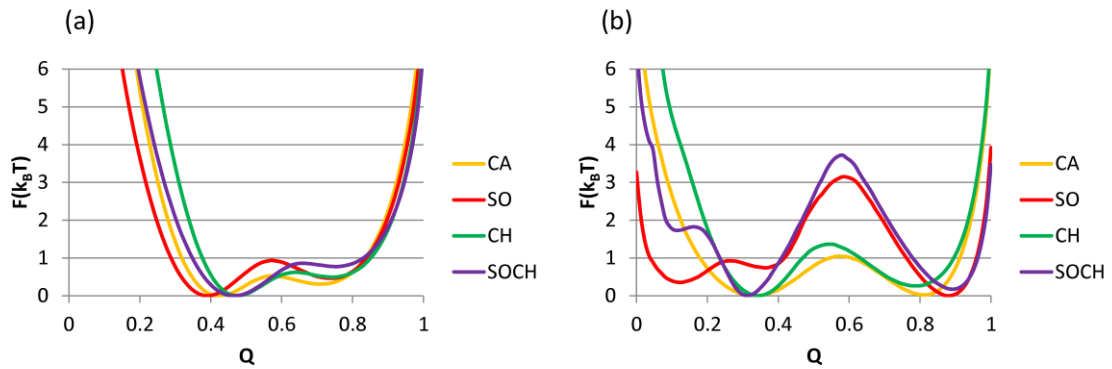


図 9: (a) 4つのモデルを用いて計算した protein A の自由エネルギープロファイル。それぞれの遷移温度 ( $k_B T$ ) は 0.766 (CA)、0.798 (SO)、0.766 (CH)、0.70 (SOCH) である。(b) 4つのモデルを用いて計算した src-SH3 domain の自由エネルギープロファイル。それぞれの遷移温度 ( $k_B T$ ) は 0.800 (CA)、0.784 (SO)、0.790 (CH)、0.768 (SOCH) である。横軸は Q 値、縦軸は自由エネルギーを  $k_B T$  で割ったものとした。

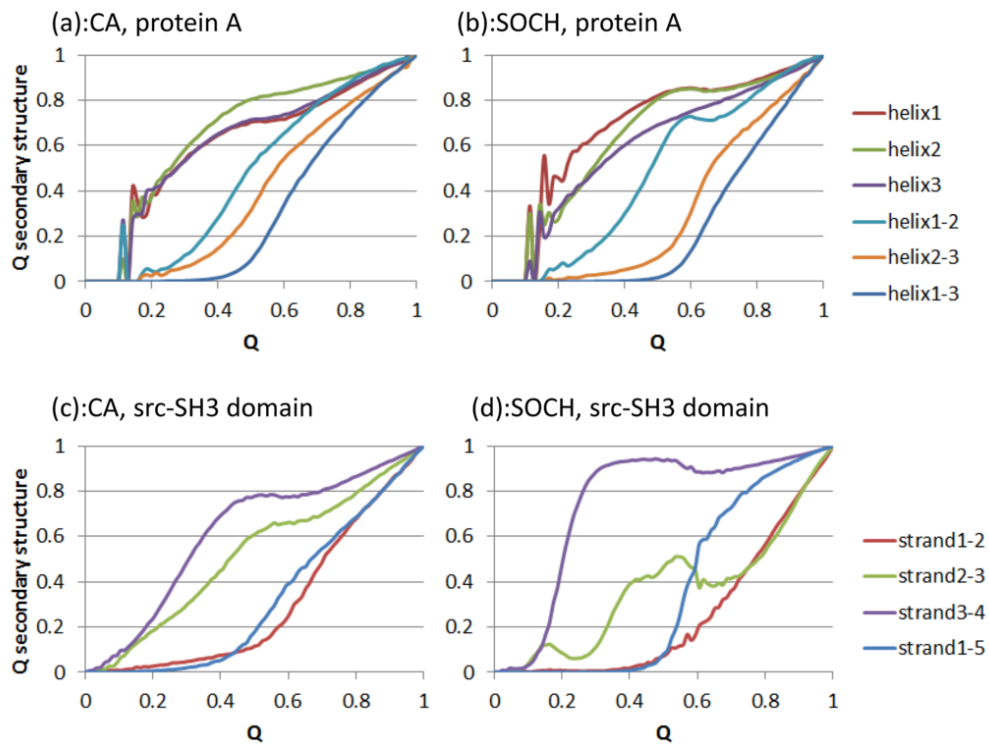


図 10: (a)CA モデルと(b)SOCH モデルで計算した protein A の 2 次構造形成度合いのプロファイル。(c)CA モデルと(d)SOCH モデルで計算した src-SH3 domain の 2 次構造形成度合いのプロファイル。

図 9(a)に示された通り、protein A ではどのモデルを用いても自由エネルギープロファイルにはほとんど変化が無い事が示された。Q = 0.4-0.5 の領域の領域ではどのプロファイルも変性状態に位置しており、Q = 0.4-0.5 の領域では全てのヘリックスが 70-80%形成されている事が図 10(a, b)より示された。また、遷移状態付近のアンサンブルでは、どのモデルの結果においてもヘリックス 1 とヘリックス 2 のパッキングが確認された。遷移状態構造の形成にヘリックス 1 とヘリックス 2 が関与している点は実験的に示されている protein A のフォールディング経路<sup>26</sup>と一致する。他方、図 9(b)の src-SH3 domain の自由エネルギープロファイルは、側鎖の配向を仮定した拘束の有無で大きく変化した。CA モデル、CH モデルの結果とは異なり、SO モデル、SOCH モデルでは新たなバリアが生じた。CA モデルと SO モデルの結果、CH モデルと SOCH モデルの結果を比較すると、拘束の付加により自由エネルギーバリアの高さが高くなることが示された。ストランド 2-4 で構成されるβシートは Q = 0.3-0.4 の領域で形成されており、この結果はこの領域が遷移状態の形成に大きく寄与するという実験データと一致する。<sup>27,28</sup> また、src-SH3 domain のデータでは変性状態付近に新たなバリアが観察された。βシートはαヘリックスと比較して配列上離れた残基間で相互作用を形成するため、αタンパクと比較して短い構造でもバリアを経てフォールドする事が示唆される。本研究の結果は側鎖の配向を考慮した拘束がフォールディングの途中の構造エントロピーの減少量を高めたため新たなバリアが生じたと考えられる。これは、all-βタンパクのフォールディング機構の複雑さの原因となっているかもしれない。<sup>30</sup>

## 2.4 考察

本研究では全原子 G $\ddot{o}$  モデルの特徴を 2 つの効果に分けて C $\alpha$ G $\ddot{o}$  モデルへ取り込む事を試みた。本研究で取り入れた効果は(1)残基間相互作用エネルギーの多様性と(2)側鎖の配向の違いを効果的に区別する C $\alpha$ 原子の方向の拘束の 2 点である。ここではそれら 2 つの効果は側鎖の効果を模倣するために十分であったかどうかを考察する。

本研究では残基間相互作用の強さの違いをモデルに含めるため、係数 C $_{ij}$  を定義した。そして C $_{ij}$  を含むモデル (CH、SOCH) と含まないモデル (CA、SO) を比較する事で C $_{ij}$  の効果をあきらかにした。図 4、図 5、図 6、図 10 にて C $_{ij}$  の有無はフォールディング径路に大きく影響を与える事を示した。そして図 4 で示された protein L のフォールディング径路や、図 6 で示された protein L の遷移状態付近のアンサンブルのコンタクトマップは全原子 G $\ddot{o}$  モデルによって計算された結果と類似している事を示した。<sup>11</sup> また、protein L とフォールディング径路が異なる protein G でも全原子 G $\ddot{o}$  モデルと同様、C 末端からのフォールディングを観察する事が出来た。従って、全原子 G $\ddot{o}$  モデル上での側鎖の存在に依存した残基間相互作用エネルギーの強さの多様性は、C $\alpha$ 原子間の相互作用に重み付けする事で十分再現される事がわかった。他方、残基間相互作用の強さのばらつきは自由エネルギーバリアの逆さを下げ、フォールディングの協同性を下げる事が示された。同様の結果は他の研究でも示されている。ただし、全原子 G $\ddot{o}$  モデルの協同性は C $\alpha$ G $\ddot{o}$  モデルよりも高い事が示されているため、側鎖の存在は残基間相互作用の強さの多様性意外に、協同性を高める効果を持つ事が示唆された。

C $\alpha$ 原子が同じ位置にあったとしても、その側鎖の配向は様々である。その配向が少し変わると側鎖間の相互作用は形成されない上、側鎖どうしが衝突する場合もある。すなわち、全原子 G $\ddot{o}$  モデルのように側鎖が存在する場合は相互作用の形成と共に、C $\alpha$ 原子間の距離だけではなく側鎖の配向も固定される。本研究ではそのような効果を B $_{ij}$  という変数を用いて効果的にモデルに導入した。そして B $_{ij}$  の効果を含むモデル (SO、SOCH) と含まないモデル (CA、CH) を比較することでその効果を明らかにした。図 3、図 4 からは B $_{ij}$  の有無で協同性が大きく変化させるが、フォールディング径路にはほとんど変化が無い事が示された。そして図 3 の SOCH モデルを用いて計算した自由エネルギーバリアの高さは全原子 G $\ddot{o}$  モデルと同等の結果を得た。従って、側鎖の配向の固定に依存

したエントロピーの減少の効果は  $B_{ij}$  を用いて代用可能である事が示唆された。

## 2.5 結論

本研究では、 $C\alpha G\ddot{o}$  モデルの計算の精度を上げるため、相互作用エネルギーの強さの多様性と、仮想的な側鎖の配向に依存した  $C\alpha$  原子間の幾何学的な拘束を含む、新たな  $C\alpha G\ddot{o}$  モデルを開発した。相互作用エネルギーの強さの多様性はフォールディング経路を決定した。仮想的な側鎖の配向に依存した  $C\alpha$  原子間の拘束は協同的な遷移をもたらした。また、その効果は複雑なトポロジーを持つタンパクほどより協同的な遷移をもたらす事が示された。それらの効果を導入した SOCH モデルは protein L、protein G において全原子  $G\ddot{o}$  モデルと類似の結果を再現した。2つの効果の導入により、同じトポロジーを持つが、異なるフォールディング径路をもつタンパクのフォールディング機構も解析する事が可能と成った。3章以降はより複雑なタンパクに本モデルを応用する。

## 参考文献

1. Go N. Theoretical studies of protein folding. *Annu Rev Biophys Bioeng* 1983;12:183-210.
2. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *Journal of Molecular Biology* 2000;298(5):937-953.
3. Takada S. [Protein folding and its structural topology]. *Tanpakushitsu Kakusan Koso* 2001;46(2):148-153.
4. Ferguson A, Liu Z, Chan HS. Desolvation barrier effects are a likely contributor to the remarkable diversity in the folding rates of small proteins. *Journal of Molecular Biology* 2009;389(3):619-636.
5. Chavez LL, Onuchic JN, Clementi C. Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J Am Chem Soc* 2004;126(27):8426-8432.
6. Larriva M, Prieto L, Bruscolini P, Rey A. A simple simulation model can reproduce the thermodynamic folding intermediate of apoflavodoxin. *Proteins* 2010;78(1):73-82.
7. Karanicolas J, Brooks CL, 3rd. Improved Go-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *Journal of Molecular Biology* 2003;334(2):309-325.
8. Karanicolas J, Brooks CL, 3rd. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci* 2002;11(10):2351-2361.
9. Hills RD, Jr., Brooks CL, 3rd. Insights from coarse-grained go models for protein folding and dynamics. *Int J Mol Sci* 2009;10(3):889-905.
10. Cho SS, Levy Y, Wolynes PG. Quantitative criteria for native energetic heterogeneity influences in the prediction of protein folding kinetics. *Proc Natl Acad Sci U S A* 2009;106(2):434-439.
11. Clementi C, Garcia AE, Onuchic JN. Interplay among tertiary contacts, secondary structure formation and side-chain packing in the

- protein folding mechanism: all-atom representation study of protein L. *Journal of Molecular Biology* 2003;326(3):933-954.
12. Whitford PC, Noel JK, Gosavi S, Schug A, Sanbonmatsu KY, Onuchic JN. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins* 2009;75(2):430-441.
  13. Luo A, Wang W, Sima N, Lu Y, Zhou J, Xu G, Yu H, Wang S, Ma D. Short hairpin RNA targeting c-FLIP sensitizes human cervical adenocarcinoma HeLa cells to chemotherapy and radiotherapy. *Cancer Lett* 2008;271(2):323-332.
  14. Shimada J, Kussell EL, Shakhnovich EI. The folding thermodynamics and kinetics of crambin using an all-atom Monte Carlo simulation. *Journal of Molecular Biology* 2001;308(1):79-95.
  15. Kaya H, Chan HS. Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: how adequate is native-centric topological modeling? *Journal of Molecular Biology* 2003;326(3):911-931.
  16. Sulkowska JI, Cieplak M. Selection of optimal variants of Go-like models of proteins through studies of stretching. *Biophys J* 2008;95(7):3174-3191.
  17. Ferrenberg AM, Swendsen RH. New Monte Carlo technique for studying phase transitions. *Phys Rev Lett* 1988;61(23):2635-2638.
  18. Ferrenberg AM, Swendsen RH. Optimized Monte Carlo data analysis. *Phys Rev Lett* 1989;63(12):1195-1198.
  19. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J Comput Chem* 1992;13(8):1011-1021.
  20. Kim DE, Fisher C, Baker D. A breakdown of symmetry in the folding transition state of protein L. *Journal of Molecular Biology* 2000;298(5):971-984.
  21. Scalley ML, Nauli S, Gladwin ST, Baker D. Structural transitions in the protein L denatured state ensemble. *Biochemistry-U*

- 1999;38(48):15927-15935.
22. McCallister EL, Alm E, Baker D. Critical role of beta-hairpin formation in protein G folding. *Nat Struct Biol* 2000;7(8):669-673.
  23. Frank MK, Clore GM, Gronenborn AM. Structural and dynamic characterization of the urea denatured state of the immunoglobulin binding domain of streptococcal protein G by multidimensional heteronuclear NMR spectroscopy. *Protein Sci* 1995;4(12):2605-2615.
  24. Blanco FJ, Rivas G, Serrano L. A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. *Nat Struct Biol* 1994;1(9):584-590.
  25. Lewandowska A, Oldziej S, Liwo A, Scheraga HA. beta-hairpin-forming peptides; models of early stages of protein folding. *Biophys Chem* 2010;151(1-2):1-9.
  26. Sato S, Religa TL, Daggett V, Fersht AR. Testing protein-folding simulations by experiment: B domain of protein A. *Proc Natl Acad Sci U S A* 2004;101(18):6952-6956.
  27. Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, Baker D. Experiment and theory highlight role of native state topology in SH3 folding. *Nat Struct Biol* 1999;6(11):1016-1024.
  28. Grantcharova VP, Riddle DS, Santiago JV, Baker D. Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat Struct Biol* 1998;5(8):714-720.
  29. Li J, Shinjo M, Matsumura Y, Morita M, Baker D, Ikeguchi M, Kihara H. An alpha-helical burst in the src SH3 folding pathway. *Biochemistry-U S* 2007;46(17):5072-5082.
  30. Vu DM, Brewer SH, Dyer RB. Early Turn Formation and Chain Collapse Drive Fast Folding of the Major Cold Shock Protein CspA of *Escherichia coli*. *Biochemistry-U S* 2012.





## 第3章 ferredoxin-like fold タンパクのフォールディング機構の解析：自由エネルギープロファイルの形とフォールディングコアとの関係

### 3.1 序論

ここ20年の間に小さな速くフォールドするタンパクのフォールディング機構は実験<sup>1-7</sup>と理論<sup>8-13</sup>の両面から広く解析されてきた。このプロセスにおいてGōモデルは、フォールディング速度の多様性や<sup>11,14,15</sup>中間体の有無、<sup>9,16,17</sup>フォールディング経路<sup>9,14,17,18</sup>が天然トポロジーから解釈できる事を示唆してきた。その結果、小さな速くフォールドするタンパクはファネル型のエネルギーランドスケープを持つ事により天然構造の探索を効率的に行っている事が示唆されてきた。<sup>19-21</sup>しかしフォールディングの過程でいつ何が起きているのかは依然として詳細に理解されていない。

フォールディングの過程で起きている事を理解する事が困難であるのは、その過程の構造は不安定であり、実験的に直接観察する事が困難である事に起因している。そのため、フォールディングにおける中間状態の役割や、それぞれの遷移の過程で起きる事を明らかにする事は重要な問題として取り組まれてきた。<sup>22</sup>複数回の遷移を経てフォールドするタンパクのそれぞれの遷移の過程で起きている事を明らかにする事はフォールディングの過程で起きている事の順序を明らかにする事につながるだろう。

しかし、複数回の遷移の過程で何が起きているかに関して、これまで多数の仮説が提案されてきたものの、我々は未だその詳細を理解していない。代表的なものとしてはトポロジーの探索の後に側鎖のパッキングが起きることを仮定しているものや、<sup>23,24</sup>別々のサブドメインのフォールディングが複数回の遷移に対応しているという仮説<sup>25</sup>や、2次構造、3次構造が別々に起こる事を仮定しているもの<sup>26</sup>などが挙げられるが、それ以外にも多数提案されている。<sup>28,29</sup>その理解が進まない原因としては、実験では遷移の過程を直接見る事が出来ない事に加え、複数回の遷移を経てフォールドするタンパクは大きいものが多いため、全原子シミュレーションのような計算の対象にする事が難しい事にも起因している。実際、複数回の遷移を経てフォールドするタンパクのフォールディング機構に関して計算を用いてアプローチしている論文は少ない。<sup>9,27</sup>計算機を用いて複数回の遷移を経てフォールドするタンパクのフォールディングメカニ

ズムを明らかにするためには、その中でもシンプルなものを対象とするべきであろう。

近年、実験技術の向上により、これまで検出されていなかったような中間状態が検出されるようになってきた。<sup>28</sup> また、2状態で遷移すると考えられてきたタンパクでも中間状態が存在する可能性がある事も示されてきた<sup>29,30</sup> そのようなタンパクは小さい上、遷移の回数も少ないため、シミュレーションのターゲットとして適しているだろう。一方、以前より **chevron plot**<sup>31,32</sup> にカーブが観察されるタンパクは自由エネルギーの高い領域に中間状態が存在する事が示唆されてきた。<sup>26,32</sup> このようなタンパクでは実際に複数回の遷移は観測されないものの、実験的に観測されるかどうかは単にその安定性の差であると考えられるため、<sup>26,27,33</sup> これらのタンパクも複数回の遷移を示すがシンプルな構造を持つタンパクであると言えるだろう。自由エネルギーの高い領域に複数のバリアが存在するタンパクに関する計算からの知見は未だ少ない。<sup>34,35</sup>

本研究では、**ferredoxin-like fold** タンパクを複数回の遷移を経てフォールドする単純なタンパクとして研究対象とする。**Ferredoxin-like fold** タンパクの一つである **U1A** は自由エネルギーの高い領域に中間状態を持つ事が示唆されている。<sup>25,33,36-38</sup> **U1A** と同じトポロジーを持つ **S6** の一残基変異体<sup>38</sup> と円順列変異体<sup>41,42</sup> はカーブした **chevron plot** を示し、自由エネルギーの高い領域に中間体を持つ、もしくは変性剤濃度に依存して基底状態が移動する事が示唆されている。対照的に、**S6** や **ADA2h** の野生型は2状態でフォールドする事が示唆されている。<sup>39-41</sup> **ferredoxin-like fold** タンパクにおいて中間体の有無を左右するメカニズムを明らかにする事は複数回の遷移を経てフォールドするタンパクのフォールディングメカニズムの理解につながると考えられる。これまで、**ferredoxin-like fold** タンパクはいくつかのグループによって理論的な研究の対象とされてきたが、<sup>17,42-44</sup> 大抵の研究は2状態で遷移する **S6** を研究対象とし、中間状態の有無に注目した研究は少ない。<sup>35</sup>

そこで、本研究では第2章にて導入したモデル<sup>45</sup> を用いて **U1A**、**ADA2h**、**S6** のフォールディング機構を解析した。また、構造の解かれている **S6** の円順列変異体のフォールディング機構も解析した。これらの結果を基に部分構造の協同的なフォールディングがタンパクの自由エネルギープロファイルの形を決定している事や、**ferredoxin-like fold** タンパクの一般的なフォールディング機構を提案する。

## 3.2 方法

### モデル

本研究では第 2 章で用いた SOCH モデルと同じモデルを用いてシミュレーションを行った。

### シミュレーション

シミュレーションの方法も第 2 章と共通である。ここでは対象タンパクによって異なる温度パラメータや、シミュレーションの efficiency を示す数値のみを記載する。U1A では温度パラメータ ( $k_B T / \epsilon$ ) 0.760-0.950 の間で 29 個の異なる温度で、ADA2h は 0.730-0.820 の間で 22 個の異なる温度で、S6 は 0.610-0.750 の間で 21 個の異なる温度でシミュレーションを行った。異なるランダムシードを用いて 5 回のシミュレーションを行った結果から得られた自由エネルギープロファイルの標準誤差の平均は U1A では  $1.1 \times 10^{-2}$ 、ADA2h では  $5.1 \times 10^{-3}$ 、S6 では  $1.8 \times 10^{-2}$  と非常に小さな値となったため、エラーバーは記載しなかった。

### chevron plot

2 状態でフォールドするタンパクにおいて、変性剤濃度の変化に対する速度の変化をプロットすると、その値は以下の式によく従う事が知られている。<sup>31,32</sup>

$$\ln k_{obs} = \ln \left[ k_f^{H_2O} \exp(-m_{k_f} [\text{denaturant}]) + k_u^{H_2O} \exp(m_{k_u} [\text{denaturant}]) \right]$$

$k_{obs}$  は実際に観測される速度定数を示し、 $k_f^{H_2O}$ 、 $k_u^{H_2O}$  はそれぞれ、変性剤を除いた水溶液中でのフォールディング速度、アンフォールディング速度を示す。この直線は低変性剤濃度の領域では傾き  $m_{k_f}$  の直線となり、高変性剤濃度領域では傾き  $-m_{k_u}$  の直線となる。 $m_{k_f} = \sum_i \alpha_{i\ddagger} \delta g_i$  であり、 $\alpha_{i\ddagger}$  は原子団  $i$  の変性状態と遷移状態の間での露出表面積の差であり、 $\delta g_i$  は原子団  $i$  の水中から変性溶媒中への移相自由エネルギーの大きさを示す。従って、変性剤濃度に依存して遷移状態の位置が変化しない場合は、 $m_{k_f}$  も  $m_{k_u}$  も一定の値をとるため、一定となる。2 状態遷移でフォールドするタンパクは通常  $m_{k_f}$ 、 $m_{k_u}$  共に一定であり、このプロットは V 字型のプロットとなる。その形からこのプロットはたびたび chevron plot と呼ばれる。但し、複数回の遷移を経てフォールドするタンパクの場合は、変性剤の濃度が変化する過程でその律速段階の位置が変化するため、変性剤の

濃度に依存して、 $m_{k_f}$ あるいは $m_{k_u}$ の値が変化する。そのため、chevron plot にカーブが見られるときは中間体の存在が示唆される。

### ターゲットタンパク

本研究では、the RNA-binding domain of the U1A spliceosomal protein (U1A)、 the activation domain of human procarboxypeptidase A2 (ADA2h)、 and ribosomal protein S6 (S6)と名付けられた3つの ferredoxin-like タンパクのフォールディングメカニズムを計算した。構造を図1に示す。実験的に示されているフォールディングプロセスの概要を以下に記載する。

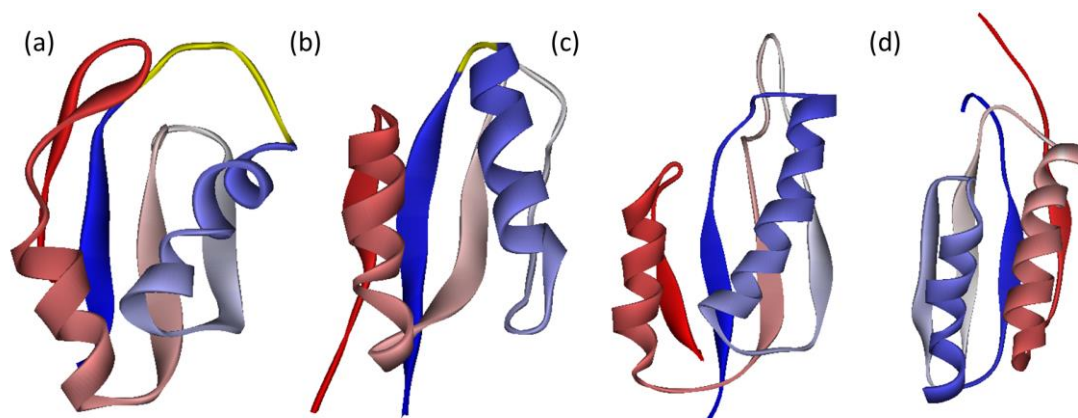


図1: (a) U1A、(b) ADA2h、(c) S6、(d) S6\_p54-55の天然構造。N末端からC末端にかけて青から赤のグラデーションで色付けした。U1AとADA2hの $\beta 1$ と $\alpha 1$ をつなぐループは黄色で色付けした。

### U1A

U1Aは自由エネルギープロファイルの形とchevron plotのカーブとの関係を調査するためによく用いられてきた。U1Aは野生型、一残基変異体共にchevron plotにカーブが生じる。<sup>36-38</sup> chevron plotのカーブは、自由エネルギーの高い領域に凹凸が存在する事に起因して遷移状態の移動する事で生じる事が示唆されているが、その凹凸の数や大きさに関しては意見が分かれている。いくつかの研究結果は、U1Aの遷移状態付近に多数の小さな凹凸が存在するために、chevron plotにカーブが生じる事を示唆しているが、<sup>35,46</sup> 他の研究では3個のはっきりとしたバリアの存在を仮定している。<sup>33</sup> フォールディング径路に関しては、 $\beta$ スト

ランド 2 ( $\beta 2$ )、3 ( $\beta 3$ )、 $\alpha$ ヘリックス 1 ( $\alpha 1$ )が遷移状態の形成に大きな役割を果たしている事が示唆されている。<sup>36</sup>

### ADA2h

ADA2h は 2 状態で天然構造へ遷移する事が示唆されている。<sup>39,41</sup> いくつかの一残基変異体は chevron plot にカーブを生じる事が示唆されている。<sup>39</sup> ADA2h のフォールディングにおいては、 $\alpha 2$ 、 $\beta 1$ 、 $\beta 3$  が中心的な役割を果たしている事が示唆されている。<sup>39</sup>

### S6

S6 の野生型は 2 状態でフォールドするが、一残基変異体、<sup>38,40</sup> 円順列変異体<sup>47,48</sup> の chevron plot にはカーブが生じる事が示されている。カーブの原因としては遷移状態の移動や<sup>38</sup> 基底状態の移動が示唆されている。<sup>40</sup> 野生型の S6 では $\beta 1$ 、 $\beta 3$ 、 $\alpha 1$  が遷移状態の形成に重要な役割を果たしている事が示唆されているが、いくつかの円順列変異体では $\beta 1$ 、 $\beta 4$ 、and、 $\alpha 2$  が遷移状態の形成に重要な役割を果たしていた。<sup>47,48</sup> 従って、S6 には 2 つのフォールディングコアが存在する事が示唆されている。また、 $\beta 2$  はフォールディングにはほとんど寄与しない事が示されている。<sup>49</sup>

### PDB 構造

本研究では PDB に記録されている NMR 構造の 1 番目の構造を用いてシミュレーションを行った。但し、揺らいでいる末端残基は削除したため、実際には 1FHT の残基番号 8-86 番目、1O6X の 8-81 番目、2KJV の 1-93 番目の構造を天然構造として用いた。ただし、U1A の完全な構造には通常の ferredoxin-like fold には存在しないループとヘリックスが存在するため、同じトポロジーを用いて議論をする事を目的として、1FHT の C 末端領域に存在するループとヘリックスを削除した。末端領域の削除によりバリアの高さは低くなったものの、バリアの数には変化は生じなかった。また、本研究では構造の解かれている S6 の円順列変異体である S6p\_54-55 の計算も行った。使用した構造は 2KJW の 1-93 残基目までの構造である。それぞれの構造は図 1 に示した。

### 3.3 結果

U1A は自由エネルギーの高い領域に複数のバリアが存在する事が示唆されているが、S6 や ADA2h は2状態でフォールドする事が示唆されている。本研究のモデルを用いて上述のプロファイルを再現できるかどうかを確認するため、U1A、ADA2h、S6 の自由エネルギープロファイルを計算し、図 2 に記載した。反応座標には、シミュレーション中に形成されている天然コンタクトの割合を表す Q 値を用いた。シミュレーション中の構造において、天然構造でコンタクトしている残基ペアが天然距離の 1.2 倍以下になったとき、その残基ペアを天然コンタクトとみなした。すなわち  $r_{ij} \leq r_{ij0} \times 1.2$  とした。

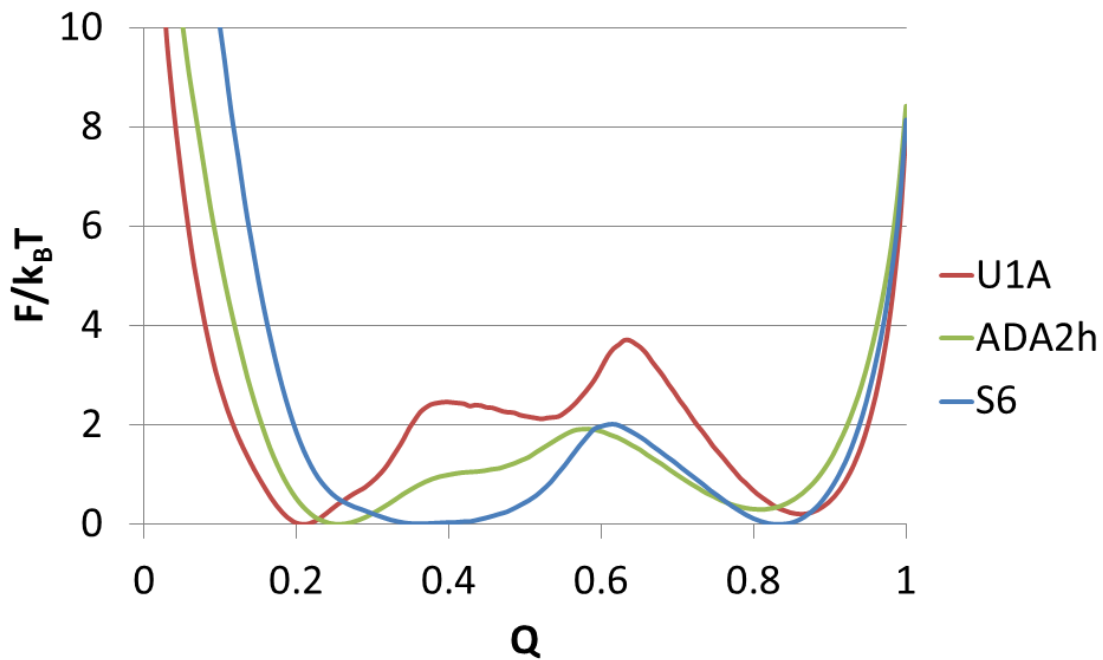


図 2: U1A、ADA2h、S6 の自由エネルギープロファイル。自由エネルギーの値は  $k_B T$  でスケールされている。

図 2 より、U1A は自由エネルギーの高い領域に準安定なアンサンブルが生じ、2つのバリアを経てフォールドする事が示されたが、ADA2h や S6 ではそのようなアンサンブルは観察されず、2状態でフォールドする事が示された。このプロファイルは実験的に示されているデータとよく一致する。但し、ADA2h と

S6は2状態でフォールドするものの、S6はADA2hと比較して広い変性状態のアンサンブルを得た。

続いて、実験的に得られているようなカーブした chevron plot が図 2 の U1A の自由エネルギープロファイルから得られる事を確認するために、図 2 の自由エネルギープロファイルから理想的な chevron plot を計算し、図 3 に記載した。U1A のプロットには中間状態の存在を仮定した式(4)を用いた。<sup>50,51</sup> また、2つのバリアのうち片方のバリアが律速段階となる過程をおいた上で導出された近似式である式(5)、式(6)も用いて計算した。<sup>50</sup> 比較のため、2状態遷移を仮定した S6 の理想的な chevron plot も図 3 に示した。すなわち、 $\lambda = k_f + k_u$  となる $\lambda$ の値を示した。

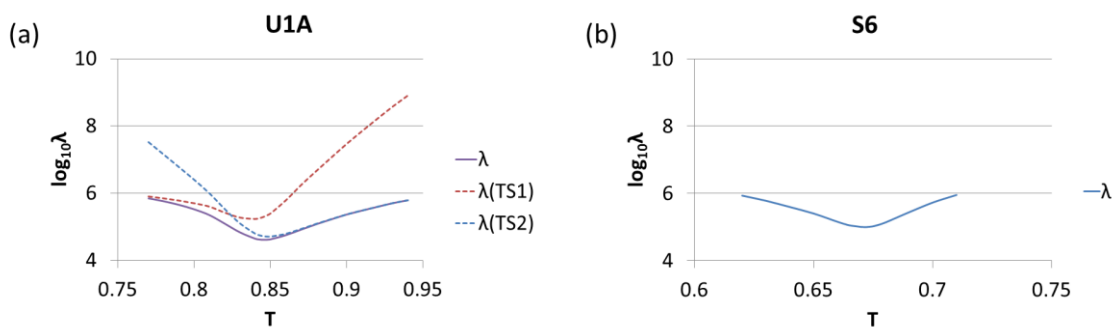


図3: (a) U1A の自由エネルギープロファイルから予測された Chevron plot.  $\lambda$  は中間体の存在を仮定してたてられた式(4)から計算された。 $\lambda(TS1)$  は中間体の存在を仮定しているが、変性状態側のバリアが律速段階になる事を仮定してたてられた式(5)を用いて計算された。 $\lambda(TS2)$  は中間体の存在を仮定しているが、天然状態側のバリアが律速段階になる事を仮定してたてられた式(6)を用いて計算された。(b) 2条対千位を仮定した式 ( $\lambda = k_f + k_u$ ) を用いて計算された S6 の Chevron plot. 速度定数はループ形成速度の極限として提案されている値を係数として用いた、すなわち、 $k = 10^6 \exp(-\Delta G/k_B T)$  とした。<sup>52</sup> ただし、この値はプロットの形には影響しない。

$$\lambda = \frac{-B - \sqrt{B^2 - 4C}}{2} \quad B = -(k_{UI} + k_{IU} + k_{IN} + k_{NI}) \quad C = k_{UI}(k_{IN} + k_{NI}) + k_{IU}k_{NI} \quad (4)$$

$$\lambda(TS1) = k_f(TS1) + k_u(TS1) \quad k_f(TS1) = k_{UI} \quad k_u(TS1) = \frac{k_{NI}}{k_{IN}} k_{IU} \quad (5)$$

$$\lambda(TS2) = k_f(TS2) + k_u(TS2) \quad k_f(TS2) = \frac{k_{UI}}{k_{IU}} k_{IN} \quad k_u(TS2) = k_{NI} \quad (6)$$

図 3(a)では、グラフ左端の温度の低い領域において、式(4)を用いて計算したプロファイルと式(5)を用いて計算したプロファイルが重なった。同様に、グラフ右端の温度の高い領域において、式(4)を用いて計算したプロファイルと式(6)を用いて計算したプロファイルが重なった。これは温度の変化に従って、律速段階が変化している事を示す。従って、カーブした chevron plot が準安定な中間体によってもたらされる事を確認する事ができ、U1A の自由エネルギープロファイル上に準安定な中間体が存在する事は実験と定性的に一致する事が示された。

36-38

但し、U1A に関する過去の研究においては、3つのバリアの存在を仮定した上で実験的に得られる chevron plot のカーブを近似していたが、<sup>33</sup> 我々の研究では2つのバリアのみが観察された。これらの結果が一致しない原因は明らかではない。本研究において、U1A の C 末端を除いた事が不一致の原因になると考える事が出来るが、U1A の完全な構造を用いてもバリアの数が変化しない事は確認している。他方、本研究で用いているモデルは側鎖や水を含んでいない。側鎖の回転や体積、水の排出が遷移の回数を増やす可能性があるため、今後の研究でそれらの効果を調査する予定である。

続いて、フォールディング径路を解析し、自由エネルギープロファイルとの関係を調査した。また、計算したフォールディングプロセスを実験と比較し、一致点や相違点を調査した。フォールディングプロセスの比較を容易にするため、本研究では鎌形らによって提案されたノンローカルクラスタを図 4 (a)、(b)、(c)のコンタクトマップ上に定義した。<sup>23</sup> また、図 4 (d)、(e)、(f)に、Q 値の変化に依存した各ノンローカルクラスタの形成度合いを記載した。



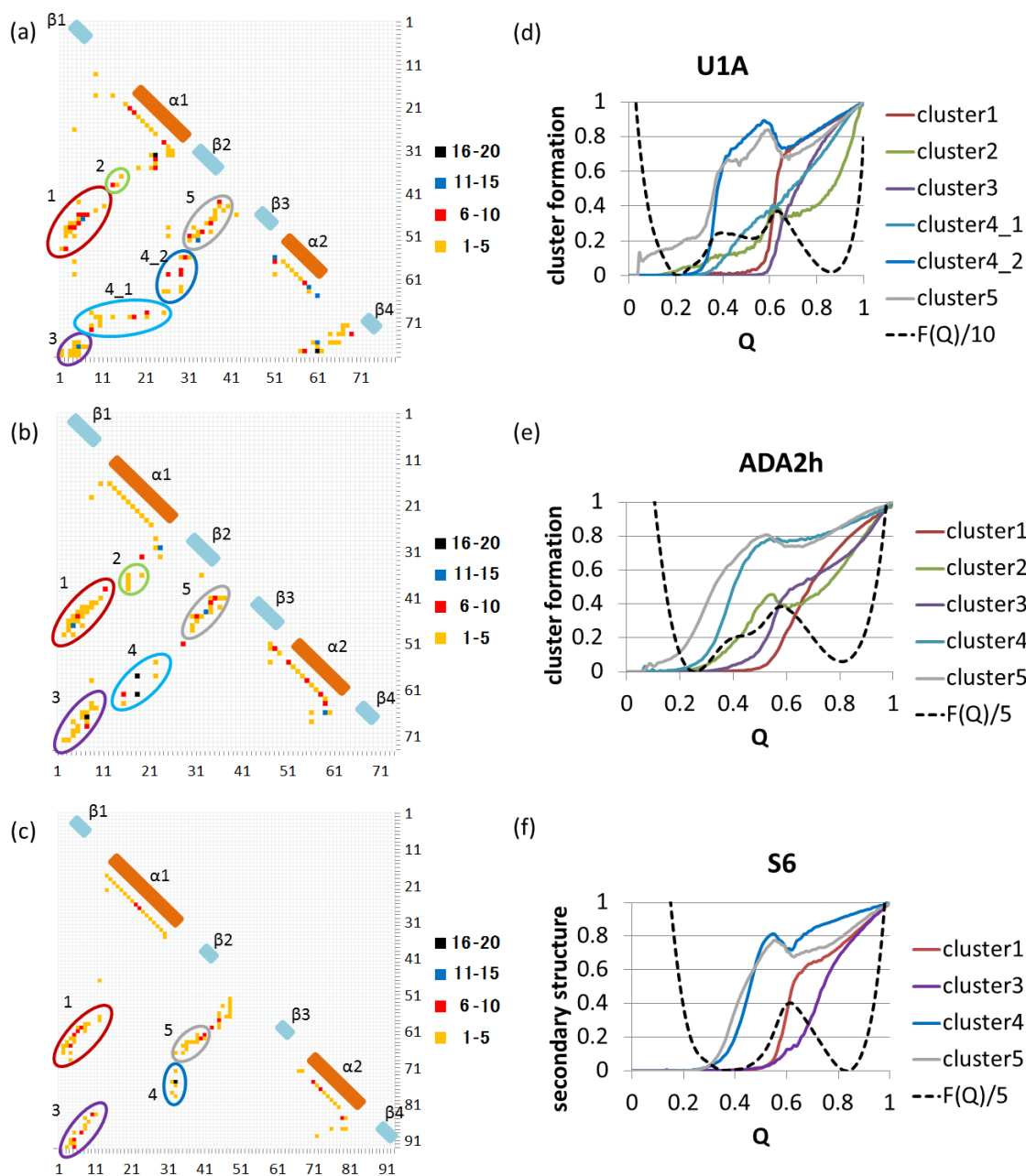


図4: (a) U1A、(b) ADA2h、(c) S6の天然構造のコンタクトマップ。それぞれの点の色は残基ペアの間に形成されている重原子間コンタクトの数を示す。マップ上にはノンローカルクラスタ<sup>23</sup>を楕円で囲む形で示した。また、二次構造要素を対角線上の青とオレンジのリボンとして示した。(d) U1A、(e) ADA2h、(f) S6のQ値に対するノンローカルクラスタの形成度合い。それぞれのプロファイルは各Q値におけるクラスタに含まれる残基間コンタクトの形成度合いの平均を示す。コンタクトマップ上のクラスタの色とプロファイルの色はそろえてある。自由エネルギープロファイルとの比較を容易にするため、(d)-(f)にはスケールした自由エネルギープロファイルも記載した。

ノンローカルクラスタは配列上の距離が4を超えない残基ペア、すなわち、 $|i - k| + |j - l| \leq 4$  という条件を満たすコンタクトで形成される。但し、一つのクラスタに2個以下のコンタクトしか含まない場合は無視する。また、配列上の距離が16以下の残基ペアもクラスタの計算からは除く。図4(a)、(b)、(c)のマップ内の一つ一つのスポットは一つの残基間コンタクトを示し、その色は残基間に存在する重原子間コンタクトの数を示す。U1A、ADA2h、S6はそれぞれ類似したトポロジーをもっているものの、コンタクトの密度や分布が少しずつ異なっている。その違いに依存して、U1Aでは5つのクラスタを、ADA2hやS6では4つのクラスタを定義する事が出来た。 $\beta 1\beta 3$ 、 $\alpha 1\beta 2$ 、 $\beta 1\beta 4$ 、 $\alpha 1\alpha 2$ 、 $\beta 2\beta 3$ 、間の相互作用をクラスタ1、2、3、4、5とした。すべてのタンパクはクラスタ1、3、4をもっている。U1Aは $\alpha 1\alpha 2$ 間の相互作用に関わるクラスタを2つ定義する事が出来たため、それらのクラスタをクラスタ4\_1、クラスタ4\_2とした。S6はクラスタ2を定義する事が出来なかった。U1AとADA2hは $\beta 2\beta 3$ 間の相互作用が配列上の距離が近い相互作用で形成されており、クラスタ5を定義する事が出来なかったが、 $\beta$ ヘアピンはバリアの形成に関与する事が知られているため、<sup>53</sup>  $\beta 2\beta 3$ 間の相互作用をクラスタ5として扱った。

上述のクラスタの形成と自由エネルギープロファイルとの関係を解析する。図4(d)より、U1Aでは、クラスタ4\_2、クラスタ5に含まれる相互作用が $Q = 0.3-0.4$ 付近で形成される。この領域はU1Aの2つある自由エネルギーバリアのうち一方の形成のタイミングと一致する。また、クラスタ1、3が $Q = 0.6$ 付近でフォールドする。この領域はU1Aのもう一方の自由エネルギーバリアの形成のタイミングと一致する。従って、クラスタ4\_2、クラスタ5に含まれる相互作用と、クラスタ1、クラスタ3に含まれる相互作用は別々のバリアの形成に寄与している事が示唆される。

S6では $Q = 0.3-0.5$ 付近でクラスタ4、クラスタ5が、 $Q = 0.5-0.6$ 付近でクラスタ1がフォールドする(図4(d))。S6のクラスタ4、クラスタ5とU1Aのクラスタ4\_2、クラスタ5は ferredoxin-like fold の中の類似した構造要素である。同様に、S6とU1Aのクラスタ1は類似した構造要素である。S6とU1Aは共通した構造要素が類似した順序でフォールドしているため、類似した経路を辿ってフォールドする事が示唆される。しかし、U1Aと異なり、S6のクラスタ4、クラスタ5のフォールディングの際、バリアは生じなかった。従って、U1AとS6

は類似した経路を通してフォールドするが、S6 のクラスタ 4、クラスタ 5 のフォールディングはバリアの形成に寄与しないため、S6 は U1A と異なり、2 状態でフォールドする事が示唆される。

ADA2h も U1A や S6 と同じ順序でフォールディングしているが、ADA2h は  $Q = 0.3-0.5$  付近でのクラスタ 4 のフォールディングが完了する前にクラスタ 1、クラスタ 2、クラスタ 3 がフォールディングし始める。ADA2h は S6 とは異なり、クラスタ 1、クラスタ 2、クラスタ 3、クラスタ 4 の全てがバリアの形成に関与している。また、U1A と異なり、2 つのバリアは生じない。

実験データより、U1A においては  $\beta$  ストランド 2 ( $\beta 2$ )、3 ( $\beta 3$ )、 $\alpha$  ヘリックス 1 ( $\alpha 1$ ) が、<sup>36</sup> ADA2h においては  $\alpha 2$ 、 $\beta 1$ 、 $\beta 3$  が、<sup>39</sup> また、S6 においては  $\beta 1$ 、 $\beta 3$ 、 $\alpha 1$  が<sup>38,40</sup> 遷移状態の形成に中心的な役割を果たす事が示されている。本研究において U1A の 1 つめのバリアの形成に関与していたクラスタは  $\beta 2\beta 3\alpha 1$  間の相互作用を含んでいた。また、ADA2h や S6 においても、 $\beta 2\beta 3$  は速くフォールドするものの、遷移状態付近でフォールドする領域は  $\beta 1\beta 3$  を含む相互作用であり、実験と一致した。

続いて、それぞれのタンパクのフォールディングの熱力学的な特徴を明らかにするために  $CO(Q)$  のプロファイルを図 5 に、 $TdS/dE$  のプロファイルを図 6 に示した。M 個の天然コンタクトを持つ N 残基のタンパクの  $CO(Q)$  のプロファイルは式(7)により定義される。

$$CO(Q) = \sum_{i=1}^M L_i \langle Q_i(Q) \rangle / NM \quad (7)$$

$L_i$  は  $i$  番目の残基ペアにおけるループの長さを示す。この値はシミュレーション中に生じたアンサンブルにおける平均のコンタクトオーダーを示している。コンタクトオーダーの大きな上昇は配列上遠くにある鎖同士が多数拘束される事を意味しており、鎖エントロピーの減少の大きさと関係付けられる。また、 $TdS/dE$  はエネルギーの変化に対する  $k_B T \ln \Omega(E)$  の変化の大きさを示し、WHAM<sup>54-56</sup> によって計算される。 $TdS/dE = 1$  になる点でグラフの傾きが正である点は基底状態を示し、 $TdS/dE = 1$  になる点でグラフの傾きが負になる点は自由エネルギープロファイルのピークを示す。

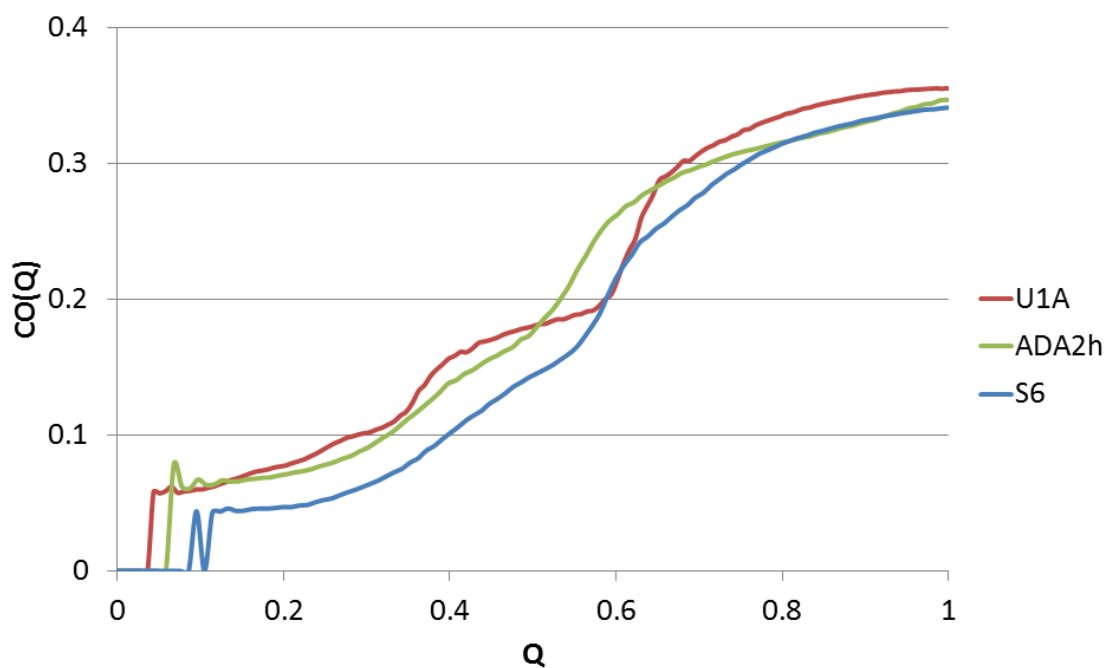


図 5: U1A、ADA2h、S6 の Q 値に対するパーシャルコンタクトオーダーCO(Q)のプロファイル。

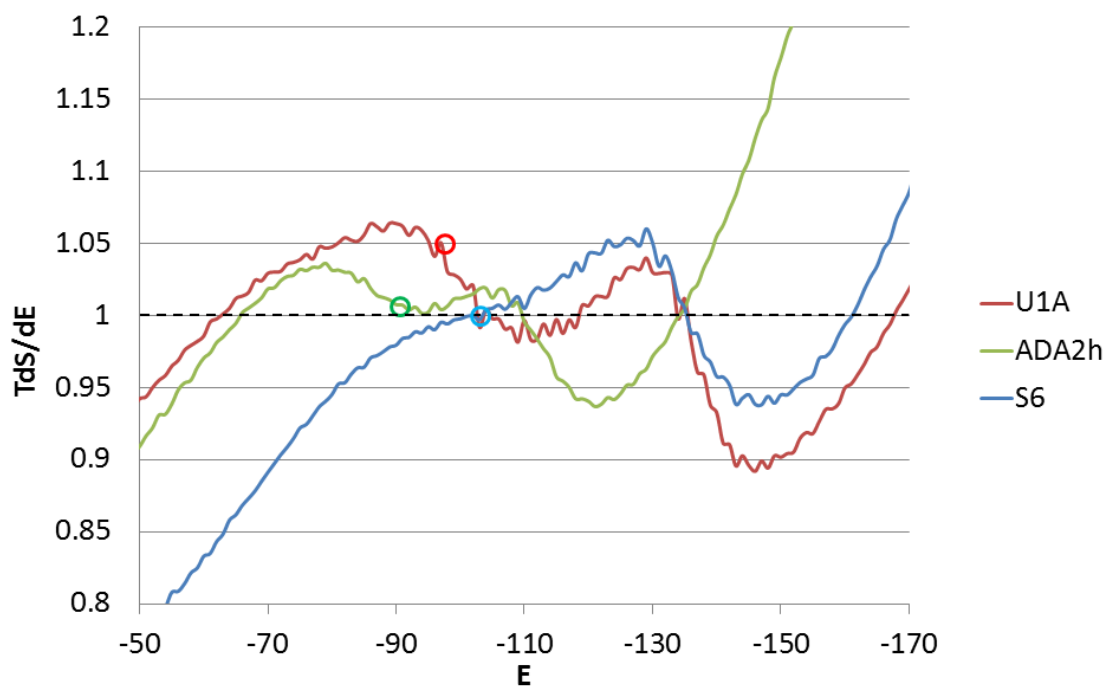


図 6: WHAM から計算された U1A、ADA2h、S6 における Tds/dE のプロファイル。各点の値は  $k_B T \ln \Omega(E)$  の単位エネルギーあたりの鎖を示している。Q = 0.4 での値に丸印をつけた。

まず、 $Q = 0.4$  付近における **S6** と他の 2 つの熱力学的な特徴にタンパク間の違いに注目する。ここまでの結果では  $Q = 0.4$  付近ではすべてのタンパクがクラスタ 4 (**U1A** はクラスタ 4\_2)、クラスタ 5 を形成しており、構造的な特徴が類似したアンサンブルをもっていた。しかし、自由エネルギープロファイルに注目すると、**S6** は基底状態に位置するのに対し、**U1A** や **ADA2h** は自由エネルギーの高いアンサンブルである事を示していた。この領域の  $CO(Q)$ 、 $TdS/dE$  のプロファイルに注目する。図 5 の  $CO(Q)$  のプロファイルにおいて  $Q < 0.6$  の領域では **S6** のみ他と比較して 0.02 小さい値を取っていた。この結果は **S6** は他のタンパクと比較してロングレンジのコンタクトが少ない事を意味する。この違いはこの領域の鎖エントロピーの減少の大きさに影響すると考えられる。図 6 の  $TdS/dE$  のプロファイルにおいても、 $Q = 0.4$  付近のアンサンブルは **U1A** と **ADA2h** は 1 を超える値を取っていたのに対し、**S6** は 1 を下回っていた。これらの結果より、 $Q = 0.4$  付近の領域のバリアの有無はロングレンジ相互作用の不足による鎖エントロピーの減少の少なさに起因している事が示唆された。式(8)で示される  $Q = 0.4-0.45$  間のアンサンブルにおける残基間距離の RMSD を計算した図 7 からも同様の結論が示唆された。

$$\sigma_{ij} = \sqrt{\langle (r_{ij} - \langle r_{ij} \rangle)^2 \rangle} \quad (8)$$

図 7 (a)、(b)における **U1A** と **ADA2h** のプロファイルでは、ほとんどの領域が 5 Å 以下の揺らぎを示す黄色で色付けされているのに対し、**S6** は  $\alpha 1$ 、 $\alpha 2$  自身の残基間距離や  $\beta 2$ - $\beta 3$  間の揺らぎ意外の領域では 10Å 近く揺らいでいる結果となった。

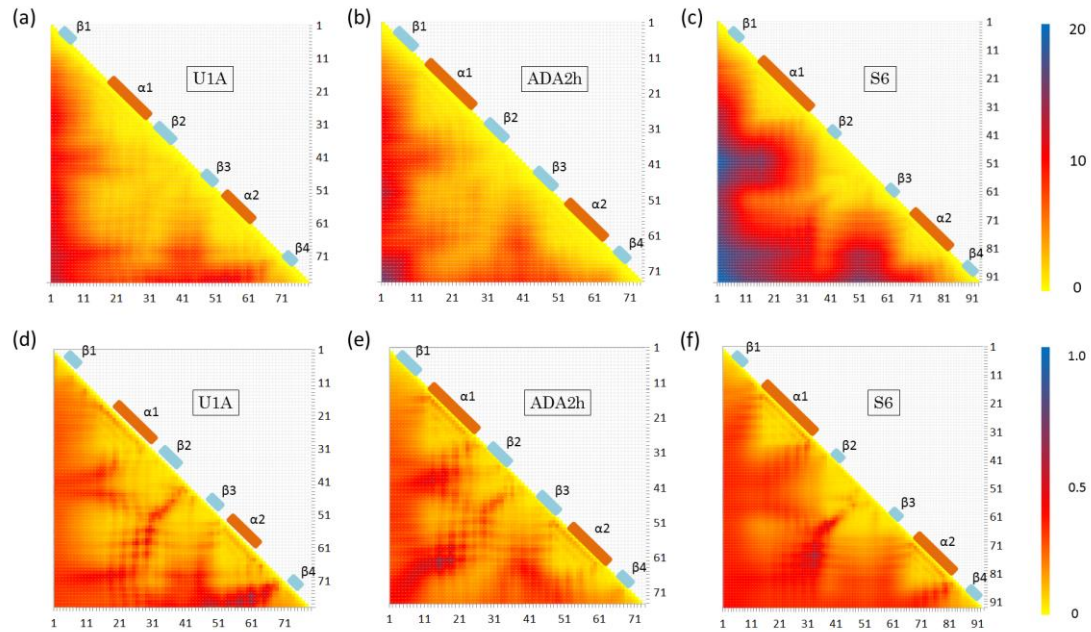


図 7:  $Q = 0.4-0.45$  のアンサンブルにおける(a) U1A, (b) ADA2h, and (c) S6 の残基間距離の RMSD。各値の定

義は以下の式を用いて計算した:  $\sigma_{ij} = \sqrt{\langle (r_{ij} - \langle r_{ij} \rangle)^2 \rangle}$ 。それぞれの点は揺らぎの大きさを色付けて

おり、色が青に近づく程大きく揺らいでいる。対角線上の青のリボンは $\beta$ ストランドを、オレンジのリボンは $\alpha$ ヘリックスを示す。(d)、(e)、(f)は U1A、ADA2h、S6 における are the RMSD をその平均値で割ったものである。

続いて U1A と ADA2h の自由エネルギープロファイルの違いが何に起因しているのかを解析する。図 5 の U1A の  $CO(Q)$  のプロファイルでは、 $Q = 0.4$  付近、 $Q = 0.65$  付近にて 2 回の急激な増加が見られる。このプロファイルは反応座標上の 2 カ所で多数のノンローカルな相互作用が一気に形成され、 $Q = 0.4-0.6$  の領域ではローカルな相互作用が形成された事を意味する。一方、ADA2h においては  $CO(Q)$  の値は連続的に増加しており、ノンローカルな相互作用が段々と増えている事を示した。この違いは  $Q = 0.55-0.60$  のアンサンブルでの残基間距離の RMSD と  $Q = 0.4-0.45$  のアンサンブルの値の差を示した図 8 から確認する事が出来た。

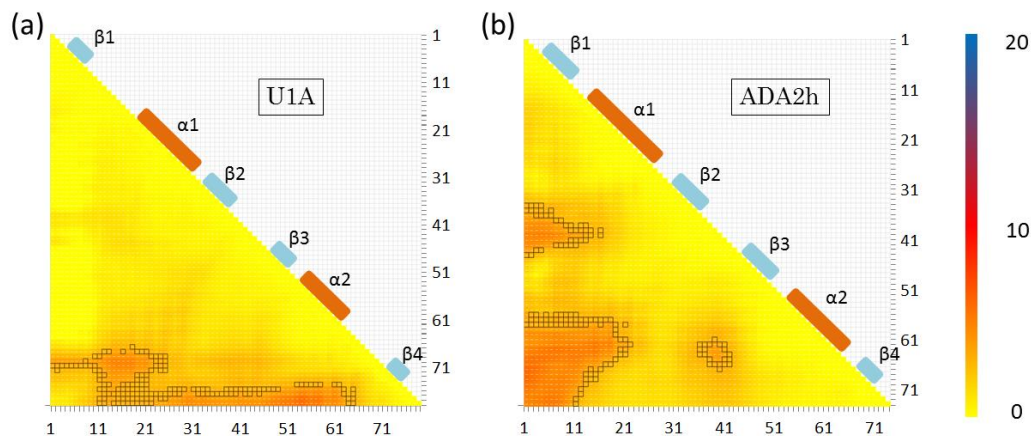


図 8: (a)U1A と(b)ADA2h の  $Q = 0.55-0.60$  と  $Q = 0.40-0.45$  における残基間距離の RMSD の差を示した。枠で囲われている点は RMSD の差が 3-3.5 の残基ペアを示す。

図 8 にてオレンジで色付けされた点は  $Q = 0.4-0.6$  の間にフォールドした領域を強調しており、ADA2h はグラフの対角線から最も離れた領域、すなわちノンローカルな相互作用が、U1A は対角線付近のローカルな相互作用を含む領域が  $Q = 0.4-0.6$  の間でフォールドした事を示す。ノンローカルな相互作用の増加は鎖のエントロピーを大きく下げる。従って、U1A は 2 度の大きなエントロピーの減少が離散的に生じるのに対し、ADA2h は連続的にエントロピーが減少し続ける事を示唆する。この違いが、U1A と ADA2h の間での準安定なアンサンブルの有無に影響しているだろう。

上述の違いは U1A と ADA2h の  $\alpha$  ヘリックスの配向やループの長さの違いに起因していると考えられる。図 1 より、ADA2h は  $\beta 1$  と  $\alpha 1$  の間、 $\alpha 2$  と  $\beta 4$  の間のループは非常に短いですが、U1A は  $\beta 1$  と  $\alpha 1$ 、 $\alpha 2$  と  $\beta 4$  の間に長いループを持つ。U1A のこの長いループがフォールディングの前半と後半を離散的にしていると考えられる。実際、2 次構造をつなぐループの長さの最適化がファネル型のエネルギーランドスケープを形成するために非常に重要な役割を果たしている事が示唆されている。<sup>57</sup>

最後に、円順列変異体と野生型のフォールディング機構の違いを明らかにするために、S6 の円順列変異体 (S6\_p54-55) のフォールディング機構を計算し、



そのコンタクトマップ、自由エネルギープロファイル、フォールディング経路を図9に示した。

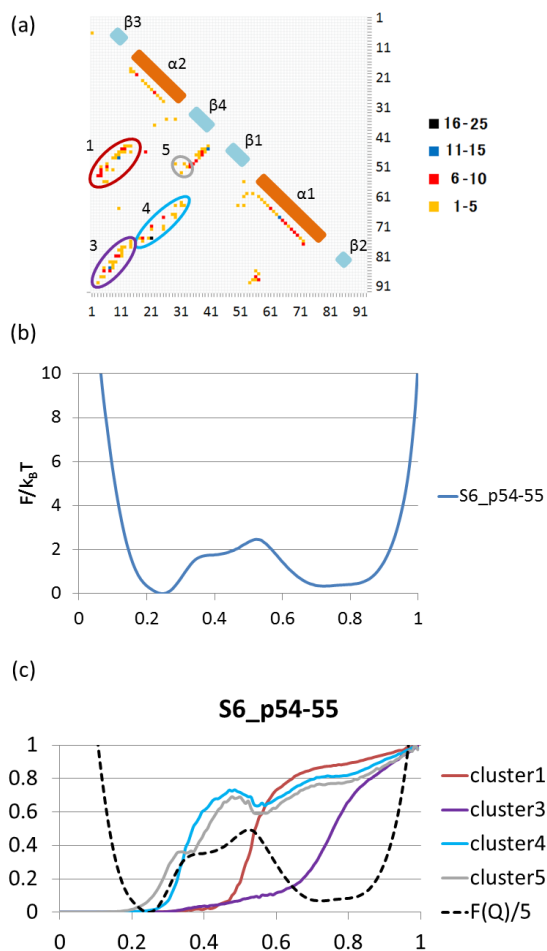


図9：S6\_p54-55の(a)コンタクトマップ、(b)自由エネルギープロファイル、(c)フォールディング経路。各2次構造要素は野生型と対応するように記載した

S6\_p54-55のコンタクトマップにおけるクラスタの位置は野生型と類似していたが、クラスタに含まれる構造要素が異なっていた。例えば、野生型のクラスタ5はβ2-β3間の相互作用を含むが、円順列変異体のクラスタ5は(野生型の配列における)β1-β4間の相互作用を含んでいた。円順列変異体のコンタクトマップ上の二次構造要素は野生型と一致するように表記した。ただし、S6\_p54-55におけるクラスタ3、クラスタ4は定義上一つのクラスタとして定義されたが、野



生型との比較を容易にするために二つに分けた。

S6\_p54-55 の自由エネルギープロファイルは ADA2h と類似しており、2 状態遷移でフォールドする事を示唆したが、 $Q = 0.4$  付近に小さなふくらみが生じた。フォールディング経路はクラスタ 4 とクラスタ 5 が  $Q = 0.3-0.4$  の間にフォールドし、クラスタ 1 が  $Q = 0.5-0.6$  の間にフォールドした。S6\_p54-55 はそのフォールディング経路も ADA2h と類似した結果となった。但し、円順列変異体のクラスタ 5 は、野生型では最後にフォールドする  $\beta 4$ 、 $\beta 1$  の配列で構成されている。S6\_p54-55 の天然状態のアンサンブルが広いウェルを形成しており、そのウェルは  $\beta 2$  がフォールドした状態とアンフォールドした状態の両方を含んでいた。これらの結果は、 $\beta 1$ 、 $\beta 4$ 、 $\alpha 2$  が S6\_p54-55 の遷移状態の形成に大きく寄与しており、 $\beta 2$  を削除してもフォールドするという実験的なデータと一致する。この結果は S6\_p54-55 の chevron plot のカーブは天然構造付近の基底状態の移動で起きている事も示唆する。我々は、配列に注目した場合は S6 の野生型と円順列変異体の間でフォールディング機構が変化する、すなわち、アミノ酸ごとのフォールドの順序は異なる事を確認した。しかし、トポロジーに注目した場合は野生型と円順列変異体は同じ構造要素から順番にフォールドしている事を確認した。

### 3.4 考察

本研究では U1A は準安定な中間体を経てフォールドするが、ADA2h、S6 は 2 状態で遷移する事が示された。一方で、そのフォールディング経路に注目すると、多少の違いはあるものの、各 2 次構造要素間の相互作用はほとんど同じ順序で形成される事が示された。これらの 3 つのタンパクの熱力学的なパラメータの変化を解析していったところ、いくつかの違いが見られた。S6 はフォールディング初期のクラスタ 4、クラスタ 5 の形成時における鎖のエントロピーの減少量が小さく、この領域は協同的にフォールドしづらくなっている事が示唆された。U1A は鎖エントロピーの大きな減少が 2 回に分けて起きている事が示された。これは  $\beta 1$  と  $\alpha 1$  の間、 $\alpha 2$  と  $\beta 4$  の間に長いループがある事に起因していると考えられる。ADA2h は全体的に密な構造を取っておりループも短いため連続的に鎖のエントロピーが下がっていく事が示唆された。

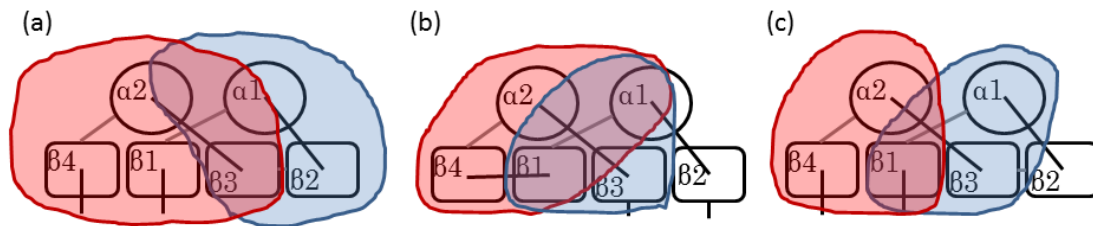


図 10 : (a)本研究で示唆された Ferredoxin-like fold タンパクが潜在的に持つフォールディングコア。(b)本研究で示唆された S6\_p54-55 のフォールディングコア。(c)Oliveberg らによって提案された S6 のフォールドン。

本研究の結果は、ferredoxin-like fold タンパクにはフォールディングコアとなり得る領域が 2 個存在する事を仮定するとうまく説明が出来る。一つ目は $\beta 2$ 、 $\beta 3$  を中心とした配列上連続した領域である。もう一方は $\beta 3$ 、 $\beta 1$ 、 $\beta 4$  からなる $\beta$ シートを含む領域である(図 10 (a))。U1A は 2 つのフォールディングコアのそれぞれが協同的に別々のタイミングでフォールドしたため、複数のバリアが生じたと考えられる。一方 ADA2h は 2 つのコアをつなぐループの長さが短いため両方のコアのフォールディングが連動し、2 状態でフォールドしたと考えられる。このループの長さの違いはそれぞれのヘリックスの配向に大きな影響を与えている。S6 は 2 つのコアのうちの一方が不安定になり協同的にフォールドする能力を失ってしまったため、1 つのコアのみがバリアを形成して協同的にフォールドしたため 2 状態でフォールドすると考えられる。S6 の円順列変異体は同じトポロジーをとっているが、配列がつながり変わっているため、 $\beta 1$  と  $\beta 4$  を含む $\beta 1$ ヘアピンからなる領域と、 $\beta 1$ 、 $\beta 3$ 、 $\beta 2$  から成る $\beta$ シートを含む領域を 2 つのコアとする(図 10 (b))とすると、ADA2h と同様のフォールディング機構でフォールドすると考えられる。ただし、どちらのコアにどのヘリックスを含むかははっきりとした法則性は見られなかったため、ここでは言及しなかった。

近年、Oliveberg らのグループが、自由エネルギーバリアの数や中間体の有無は独立して協同的にフォールディングする事の出来る部分構造であるフォールドン<sup>58,59</sup>の数やその重なり具合によって決定されるという事を主張している。<sup>25</sup> 彼らは、一連の実験から、 $\beta 1$ 、 $\alpha 1$ 、 $\beta 3$  と  $\beta 1$ 、 $\alpha 2$ 、 $\beta 4$  が S6 の 2 つのフォールドンを構成すると結論づけている(図 10 (c))。この結論は、S6 の一残基変異実験では $\beta 1$ 、 $\alpha 1$ 、 $\beta 3$  が、<sup>47</sup> P13-14、P33-34、P54-55 などの円順列変異体では $\beta 1$ 、 $\alpha 2$ 、

$\beta 4$  が<sup>47,48</sup> フォールディングに重要な役割を果たしていることから導かれた。彼らは円順列変異の有無に関わらず、S6 の配列上の同じ領域がコアとして働く事を仮定しているが、我々はトポロジーが変化した場合は、コアとして働く領域は配列上異なるがトポロジー的に対応する領域となると結論づけた。彼らは過去に、U1A のフォールディング機構も調べており、 $\beta 2$ 、 $\beta 3$ 、 $\alpha 1$  がフォールディングの初期に、 $\beta 1$ 、 $\alpha 2$ 、 $\beta 4$  がフォールディングの後半にフォールドする事を結論づけている。<sup>36</sup> また、S6、U1A、ADA2h のヘリックスの配向の違いがフォールディング機構の違いと関連しているかもしれない事を示唆していた。<sup>36</sup> しかし、U1A や ADA2h におけるフォールドンや、ヘリックスの配向とフォールディング機構との関連は曖昧なままであった。本研究では、S6 のフォールディングと同じメカニズムで U1A や ADA2h のフォールディング機構を説明しうる事を示した。また、ヘリックスの配向の違いではなく、ヘリックスとストランドをつなぐループの長さが本質的に重要である事を示唆した。

### 3.5 結論

本研究では我々オリジナルの G $\bar{5}$  モデルシミュレーションを用いて ferredoxin-like fold タンパクの U1A、ADA2h、S6 のフォールディング機構を解析した。そしてそのフォールディング機構の違いを生み出す原因の特定を試みた。その結果、U1A では自由エネルギープロファイルに準安定な中間体が観察されるのに対し、ADA2h と S6 は 2 状態でフォールドする事が示された。その違いは、ferredoxin-like fold タンパクには協同的にフォールドすることが可能なフォールディングコアが潜在的に 2 つある事を仮定するとうまく説明する事が出来る。U1A はそれぞれのコアが離散的にフォールドするため準安定な中間体を生じる。しかし、ADA2h はコアの間をつなぐループの長さが短いため、それぞれのコアのフォールドが連動してしまうため、2 状態でフォールドする事が示唆される。S6 の場合は片方のコアの安定性が下がり、協同的にフォールドできなくなっているため、一方のコアのみがバリアを生じてフォールディングする事が 2 状態での遷移をもたらす事が示唆された。これらの結果は協同的にフォールドする部分構造の単位であるフォールドンの数とその重なり具合がフォールディングの自由エネルギーランドスケープの形を決定するという仮説とよく一致し、ferredoxin-like fold タンパクのフォールディング機構をより詳細に示した。

## 参考文献

1. Plaxco KW, Simons KT, Ruczinski I, Baker D. Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry-US* 2000;39(37):11177-11183.
2. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of Molecular Biology* 1998;277(4):985-994.
3. Martinez JC, Serrano L. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat Struct Biol* 1999;6(11):1010-1016.
4. Chiti F, Taddei N, White PM, Bucciantini M, Magherini F, Stefani M, Dobson CM. Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nat Struct Biol* 1999;6(11):1005-1009.
5. Nickson AA, Clarke J. What lessons can be learned from studying the folding of homologous proteins? *Methods* 2010;52(1):38-50.
6. Matouschek A, Kellis JT, Jr., Serrano L, Fersht AR. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* 1989;340(6229):122-126.
7. Zarrine-Afsar A, Larson SM, Davidson AR. The family feud: do proteins with similar structures fold via the same pathway? *Curr Opin Struct Biol* 2005;15(1):42-49.
8. Onuchic JN, Wolynes PG. Theory of protein folding. *Curr Opin Struct Biol* 2004;14(1):70-75.
9. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *Journal of Molecular Biology* 2000;298(5):937-953.
10. Shimada J, Kussell EL, Shakhnovich EI. The folding thermodynamics and kinetics of crambin using an all-atom Monte Carlo simulation. *Journal of Molecular Biology* 2001;308(1):79-95.

11. Chavez LL, Onuchic JN, Clementi C. Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J Am Chem Soc* 2004;126(27):8426-8432.
12. Day R, Daggett V. Direct observation of microscopic reversibility in single-molecule protein folding. *Journal of Molecular Biology* 2007;366(2):677-686.
13. Karanicolas J, Brooks CL, 3rd. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci* 2002;11(10):2351-2361.
14. Koga N, Takada S. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *Journal of Molecular Biology* 2001;313(1):171-180.
15. Ferguson A, Liu Z, Chan HS. Desolvation barrier effects are a likely contributor to the remarkable diversity in the folding rates of small proteins. *Journal of Molecular Biology* 2009;389(3):619-636.
16. Larriva M, Prieto L, Bruscolini P, Rey A. A simple simulation model can reproduce the thermodynamic folding intermediate of apoflavodoxin. *Proteins* 2010;78(1):73-82.
17. Karanicolas J, Brooks CL, 3rd. Improved Go-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *Journal of Molecular Biology* 2003;334(2):309-325.
18. Clementi C, Garcia AE, Onuchic JN. Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: all-atom representation study of protein L. *Journal of Molecular Biology* 2003;326(3):933-954.
19. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* 1997;48:545-600.
20. Baker D. A surprising simplicity to protein folding. *Nature* 2000;405(6782):39-42.
21. Dill KA, MacCallum JL. The protein-folding problem, 50 years on. *Science* 2012;338(6110):1042-1046.
22. Sosnick TR, Barrick D. The folding of single domain proteins--have we

- reached a consensus? *Curr Opin Struct Biol* 2011;21(1):12-24.
23. Kamagata K, Kuwajima K. Surprisingly high correlation between early and late stages in non-two-state protein folding. *Journal of Molecular Biology* 2006;357(5):1647-1654.
  24. Kamagata K, Arai M, Kuwajima K. Unification of the folding mechanisms of non-two-state and two-state proteins. *Journal of Molecular Biology* 2004;339(4):951-965.
  25. Lindberg MO, Oliveberg M. Malleability of protein folding pathways: a simple reason for complex behaviour. *Curr Opin Struct Biol* 2007;17(1):21-29.
  26. Gianni S, Guydosh NR, Khan F, Caldas TD, Mayor U, White GW, DeMarco ML, Daggett V, Fersht AR. Unifying features in protein-folding mechanisms. *Proc Natl Acad Sci U S A* 2003;100(23):13286-13291.
  27. Ferreira DU, Cho SS, Komives EA, Wolynes PG. The energy landscape of modular repeat proteins: topology determines folding mechanism in the ankyrin family. *Journal of Molecular Biology* 2005;354(3):679-692.
  28. Xu M, Beresneva O, Rosario R, Roder H. Microsecond folding dynamics of apomyoglobin at acidic pH. *J Phys Chem B* 2012;116(23):7014-7025.
  29. Vu DM, Brewer SH, Dyer RB. Early Turn Formation and Chain Collapse Drive Fast Folding of the Major Cold Shock Protein CspA of *Escherichia coli*. *Biochemistry-US* 2012.
  30. Park SH, Shastry MC, Roder H. Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing. *Nat Struct Biol* 1999;6(10):943-947.
  31. Tanford C. Protein denaturation. *Adv Protein Chem* 1968;23:121-282.
  32. Tanford C. Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* 1970;24:1-95.
  33. Sanchez IE, Kiefhaber T. Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *Journal of Molecular Biology* 2003;325(2):367-376.
  34. Morrone A, Giri R, Toofanny RD, Travaglini-Allocatelli C, Brunori M,

- Daggett V, Gianni S. GB1 Is Not a Two-State Folder: Identification and Characterization of an On-Pathway Intermediate. *Biophys J* 2011;101(8):2053-2060.
35. Shen T, Hofmann CP, Oliveberg M, Wolynes PG. Scanning malleable transition state ensembles: comparing theory and experiment for folding protein U1A. *Biochemistry-U S A* 2005;44(17):6433-6439.
  36. Ternstrom T, Mayor U, Akke M, Oliveberg M. From snapshot to movie: phi analysis of protein folding transition states taken one step further. *Proc Natl Acad Sci U S A* 1999;96(26):14854-14859.
  37. Silow M, Oliveberg M. High-energy channeling in protein folding. *Biochemistry-U S A* 1997;36(25):7633-7637.
  38. Otzen DE, Kristensen O, Proctor M, Oliveberg M. Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots. *Biochemistry-U S A* 1999;38(20):6499-6511.
  39. Villegas V, Martinez JC, Aviles FX, Serrano L. Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *Journal of Molecular Biology* 1998;283(5):1027-1036.
  40. Otzen DE, Oliveberg M. Conformational plasticity in folding of the split beta-alpha-beta protein S6: evidence for burst-phase disruption of the native state. *Journal of Molecular Biology* 2002;317(4):613-627.
  41. Scalley-Kim M, Baker D. Characterization of the folding energy landscapes of computer generated proteins suggests high folding free energy barriers and cooperativity may be consequences of natural selection. *Journal of Molecular Biology* 2004;338(3):573-583.
  42. Matysiak S, Clementi C. Optimal combination of theory and experiment for the characterization of the protein folding landscape of S6: how far can a minimalist model go? *Journal of Molecular Biology* 2004;343(1):235-248.
  43. Hubner IA, Oliveberg M, Shakhnovich EI. Simulation, experiment, and evolution: understanding nucleation in protein S6 folding. *Proc Natl Acad Sci U S A* 2004;101(22):8354-8359.
  44. Wu L, Zhang J, Wang J, Li WF, Wang W. Folding behavior of ribosomal

- protein S6 studied by modified Go-like model. *Phys Rev E Stat Nonlin Soft Matter Phys* 2007;75(3 Pt 1):031914.
45. Sugita M, Kikuchi T. Incorporating into a Calpha Go model the effects of geometrical restriction on Calpha atoms caused by side chain orientations. *Proteins* 2013.
  46. Oliveberg M. Characterisation of the transition states for protein folding: towards a new level of mechanistic detail in protein engineering analysis. *Curr Opin Struct Biol* 2001;11(1):94-100.
  47. Lindberg MO, Haglund E, Hubner IA, Shakhnovich EI, Oliveberg M. Identification of the minimal protein-folding nucleus through loop-entropy perturbations. *Proc Natl Acad Sci U S A* 2006;103(11):4083-4088.
  48. Haglund E, Lindberg MO, Oliveberg M. Changes of protein folding pathways by circular permutation. Overlapping nuclei promote global cooperativity. *J Biol Chem* 2008;283(41):27904-27915.
  49. Haglund E, Danielsson J, Kadirvel S, Lindberg MO, Logan DT, Oliveberg M. Trimming down a protein structure to its bare foldons: spatial organization of the cooperative unit. *J Biol Chem* 2012;287(4):2731-2738.
  50. Bachmann A, Kiefhaber T. Apparent two-state tendamistat folding is a sequential process along a defined route. *Journal of Molecular Biology* 2001;306(2):375-386.
  51. Kiefhaber T, Kohler HH, Schmid FX. Kinetic coupling between protein folding and prolyl isomerization. I. Theoretical models. *Journal of Molecular Biology* 1992;224(1):217-229.
  52. Hagen SJ, Hofrichter J, Szabo A, Eaton WA. Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding. *Proc Natl Acad Sci U S A* 1996;93(21):11615-11617.
  53. Lewandowska A, Oldziej S, Liwo A, Scheraga HA. beta-hairpin-forming peptides; models of early stages of protein folding. *Biophys Chem* 2010;151(1-2):1-9.
  54. Ferrenberg AM, Swendsen RH. New Monte Carlo technique for studying phase transitions. *Phys Rev Lett* 1988;61(23):2635-2638.



55. Ferrenberg AM, Swendsen RH. Optimized Monte Carlo data analysis. *Phys Rev Lett* 1989;63(12):1195-1198.
56. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J Comput Chem* 1992;13(8):1011-1021.
57. Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, Montelione GT, Baker D. Principles for designing ideal protein structures. *Nature* 2012;491(7423):222-227.
58. Panchenko AR, Luthey-Schulten Z, Cole R, Wolynes PG. The foldon universe: a survey of structural similarity and self-recognition of independently folding units. *Journal of Molecular Biology* 1997;272(1):95-105.
59. Panchenko AR, Luthey-Schulten Z, Wolynes PG. Foldons, protein structural modules, and exons. *Proc Natl Acad Sci U S A* 1996;93(5):2008-2013.



## 第4章 多状態タンパクのフォールディング機構の解析：フォールドンの探索とその特徴付け

### 4.1 背景

複数回の遷移を経てフォールドするタンパクのそれぞれの遷移の過程で何が起きているかに関しては、いくつかの仮説がたてられているものの、<sup>1-6</sup> その詳細は明らかになっていない。いくつかのグループによる研究結果は、タンパクの構造は協同的にフォールドする事の出来るサブドメインに分割する事ができ、それぞれの遷移は個々のサブドメインがフォールドする過程に対応する事を示唆している。<sup>3,7,8</sup> 本論文の第3章では **ferredoxin-like fold** タンパクにおいてそのような部分構造の存在を示唆する結果を得た。

タンパクの協同的にフォールドする事の出来る部分構造の単位はフォールドンと呼ばれる事が多い。<sup>8,9</sup> この概念は **Wolynes** らによって提案された。彼らはエネルギーランドスケープ理論<sup>10-12</sup> を基に、タンパク内のフォルダビリティの高い部分構造がフォールドンであると定義し、ある領域でアミノ酸配列を区切ったときに、その N 末端側の領域、C 末端側の領域のフォルダビリティの和が極大となる点をフォールドンの境界として定義した。彼らの定義したフォールドンは、天然構造中の密な領域を指すモジュール<sup>13</sup> と一定の対応を示したため、フォールディングの単位と成る領域は天然構造でもコンパクトな構造を取っていると考えられた。

その後、**Oliveberg** や **Englander** らの2つのグループによって、いくつかのタンパクにおいてフォールドンして機能していると考えられる領域が実験的に特定されてきた。<sup>3,7,14,15</sup> しかし、その過程でフォールドンの特徴に関する新たな解釈が生じてきた。**Oliveberg** らは **ferredoxin-like fold** タンパクの実験を基に、複数のフォールドンはその一部の部分構造を共有する、配列上の離散的な領域がフォールドンを形成する場合もある、という2つの新たな考え方を提示した。<sup>3,14,15</sup> 他のタンパクのフォールディングにおいても同様の解釈が出来るかどうかははっきりとしていない。また、**Oliveberg** らは個々のフォールドンが並行してフォールディングする事を提案しているが、<sup>14</sup>**Englander** らのデータは、フォールドンは並行してフォールドしている訳ではなく個々のフォールドンが順番にフォールドしているという事を示唆している。<sup>7</sup> このように、フォールドンの性質には定義によってあいまいな点が存在する。

そこで、本研究では複数回の遷移を経てフォールドする事が示唆されている

タンパクを対象に粗視化シミュレーションを行い、バリアの数と実験的に報告されている遷移の回数を比較した。本研究では計算した自由エネルギープロファイルにおける個々の極小の間でフォールドする領域をフォールドンとした。そして、複数のフォールドンは構造の一部を共有しているかどうか、配列上の離散的な領域でフォールドンを形成するかどうかを調査した。また、個々のフォールドンは並行してフォールドするのか、順番にフォールドするのかを調査した。それらの結果からフォールドンのもつ一般的な特徴とフォールディングにおける役割を考察した。

## 4.2 方法

### モデル

本研究では第二章で用いた SOCH モデルと同じモデルを用いてシミュレーションを行った。

### シミュレーション

シミュレーションの方法は第 2 章とほとんど共通である。ただし、本研究では効率よくサンプリングを行うためレプリカ交換法を用いた。ここでは、対象タンパクによって異なる温度パラメータのみ記載する。RNase H では温度パラメータ ( $k_B T / \epsilon$ ) 0.500-1.050 の間で 32 個の異なる温度で、Barnase は 0.500-1.050 の間で 32 個の異なる温度で、IFABP では 0.500-0.950 の間で 26 個の異なる温度で、hILBP では 0.600-0.950 の間で 32 個の異なる温度でシミュレーションを行った。

### レプリカ交換法

レプリカ交換法<sup>16</sup>は焼き戻し法<sup>17</sup>の改良版として開発された方法であり、温度等の異なる  $M$  個の独立なシミュレーション (レプリカ) を並行して行い、一定の間隔でレプリカ同士のパラメータを変化させながらシミュレーションを行う方法である。ある状態  $X$  の重み因子は以下の式で与えられる。

$$H(q, p) = K(p) + E(q) \quad (1)$$

$$W_{REM}(X) = \exp \left\{ - \sum_{i=1}^M \beta_{m(i)} H(q^i, p^i) \right\} \quad (2)$$

ここでは  $i$  はレプリカのインデックス、 $m$  は温度のインデックスとする。このとき、ある状態  $X$  から一組のレプリカの温度を入れ替えた  $X'$  への遷移を考える。

$$X = \{\dots, x_m^i, \dots, x_n^j, \dots\} \rightarrow X' = \{\dots, x_m^{j'}, \dots, x_n^{i'}, \dots\} \quad (3)$$

ここでは、 $i, j$  はレプリカのインデックス、 $m, n$  は温度とする。ただし、分子動力学計算を行う際は、温度の変化と共に、運動量が以下のように変化する。

$$\begin{cases} p^{i'} \equiv \sqrt{\frac{T_n}{T_m}} p^i \\ p^{j'} \equiv \sqrt{\frac{T_m}{T_n}} p^j \end{cases} \quad (4)$$

このとき、詳細釣り合いを仮定するとレプリカ同士を交換する際の遷移確率  $w$  は以下のように定義される。

$$W_{REM}(X)w(X \rightarrow X') = W_{REM}(X')w(X' \rightarrow X) \quad (5)$$

以上の数式を整理すると、以下の数式が得られる。

$$\frac{w(X \rightarrow X')}{w(X' \rightarrow X)} = \exp(-\Delta) \quad (6)$$

$$\Delta = (\beta_m - \beta_n)\{E(q^j) - E(q^i)\} \quad (7)$$

実際のレプリカ対の遷移確率は次式のメトロポリステストで与えられる。

$$w(X \rightarrow X') = \begin{cases} 1 & \Delta \leq 0 \\ \exp(-\Delta) & \Delta > 0 \end{cases} \quad (8)$$

## フォールドン

背景でも述べたように、フォールドンの概念はオリジナルの定義以降、新たな解釈が提案されており、やや曖昧なものとなっている。それらの全ての解釈で共通しているのは、協同的にフォールドする事が出来る部分構造であるという点である。従って、本研究では自由エネルギープロファイルの極小と極小の間で形成される相互作用に参与している2次構造要素の集合をフォールドンとして、その特徴を解析する。

## 対象タンパク

### RNase H

Ribonuclease H1 (RNase H) は RNA-DNA 複合体の RNA を切断する酵素であり 5 本の $\beta$ ストランドと 4 本の $\alpha$ ヘリックスから構成される。<sup>18</sup> そのフォールディング機構は詳細に研究されてきた。RNase H は速度論的な中間体への早い遷移を経た後に、天然構造へ折れ畳まれる事が示されており、中間体では 45-120 残基目の領域が形成されている事が示唆されている。<sup>7,19-21</sup> また、中間体で形成されている領域の安定性を減少させた変異体は 2 状態でフォールドする事が示唆されている。<sup>19</sup>

### Barnase

Barnase はリボヌクレアーゼの一種であり 5 本の $\beta$ ストランドから成る逆平行 $\beta$ シートと 3 本の $\alpha$ ヘリックスから構成される。<sup>22</sup> Barnase は中間状態を経てフォールドする事が示されており、<sup>23</sup> 中間状態では $\alpha 1$  と $\beta 1$  から $\beta 4$  で構成される $\beta$ シートが形成されている事が示唆されている。<sup>24</sup> また、変性状態のアンサンブルにおいて、 $\alpha 1$  と 86-99 残基目のヘアピンを形成している事が示されているが、<sup>25</sup> この状態も中間体である可能性も示唆されている。

### iLBPs

intracellular Lipid Binding Protein family (iLBPs) は両親媒性の分子と結合するタンパクであり、1 本の $\beta$ ストランドと 2 本の短い $\alpha$ ヘリックスから構成される。<sup>26</sup> 10 本の $\beta$ ストランドは直行する 2 層の逆平行 $\beta$ シートを形成し、その $\beta$ シートの間にはリガンドとの結合部位を持つ。<sup>26</sup> iLBPs は類似した構造を取るが、その配列は多様であり、フォールディング機構も多様である事が示されている。<sup>27-29</sup> 多くの iLBPs は中間体を経てフォールドする事が示唆されているが、human ileal lipid binding protein (hILBP) は 2 状態でフォールドする事が示されている。<sup>28</sup> 従って、本研究ではフォールディングメカニズムの違いを調査するために中間体を経てフォールドする intestinal fatty acid binding protein (IFABP) と 2 状態でフォールドする hILBP の 2 つに対してシミュレーションを行った。

### 4.3 結果

本研究のモデルが実験的に得られている速度論的な特徴を再現しているかどうかを確認するため、自由エネルギープロファイルを計算した。また、個々の遷移の過程でフォールドする構造要素を明らかにするため、それぞれのタンパクにおいてフォールディング経路を解析した。反応座標には第2章、第3章と同じく  $Q$  値を用いた。対象タンパクごとにその特徴を記載していく。

#### RNase H

RNase H の熱容量曲線、自由エネルギープロファイルを図1に示した。RNase H は実験的に示されている通り、平衡では2状態で遷移するが、自由エネルギープロファイルでは、2つの自由エネルギーバリアを再現した。このプロファイルからは、変性状態から中間状態まで低いバリアを経て素早く遷移し、中間状態から天然状態の間の高いバリアを経てゆっくりと天然状態へ遷移する事が示唆される。

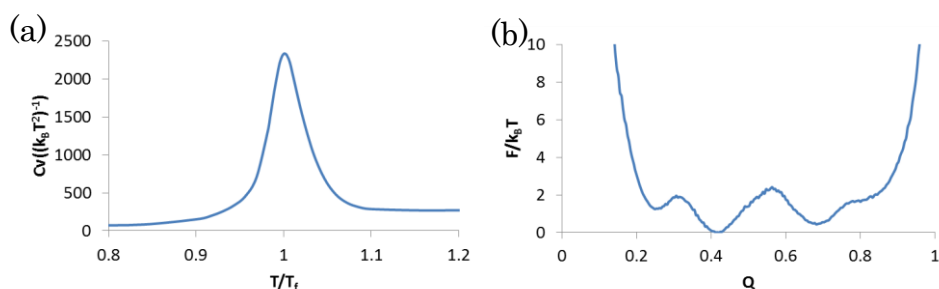


図1： RNase H の熱容量曲線(a)と自由エネルギープロファイル(b)。

続いて、RNase H の中間体で形成されている領域、中間体以後に形成される領域をコンタクトマップの形式で図2(a)、(b)に示した。このマップは中間体では $\alpha 1$  から $\beta 5$  までの領域が安定な構造を形成している事を示す。これは実験的に示されている傾向と一致する。<sup>7</sup>中間体の形成以後は $\beta 1$  から $\beta 4$  までの領域がフォールドする事が示された。この結果も実験的に示されているデータと一致する。<sup>7</sup>本研究では $\alpha 1$  から $\beta 5$  までの領域をRNase H のフォールドン A、 $\beta 1$  から $\beta 4$  までの領域をフォールドン B とした。フォールドン A とフォールドン B は $\alpha 1$  と $\beta 4$  を共有している。フォールドン A を青、フォールドン B を赤、両者が共有している領域を紫で色付けし、図2(c)に示した。

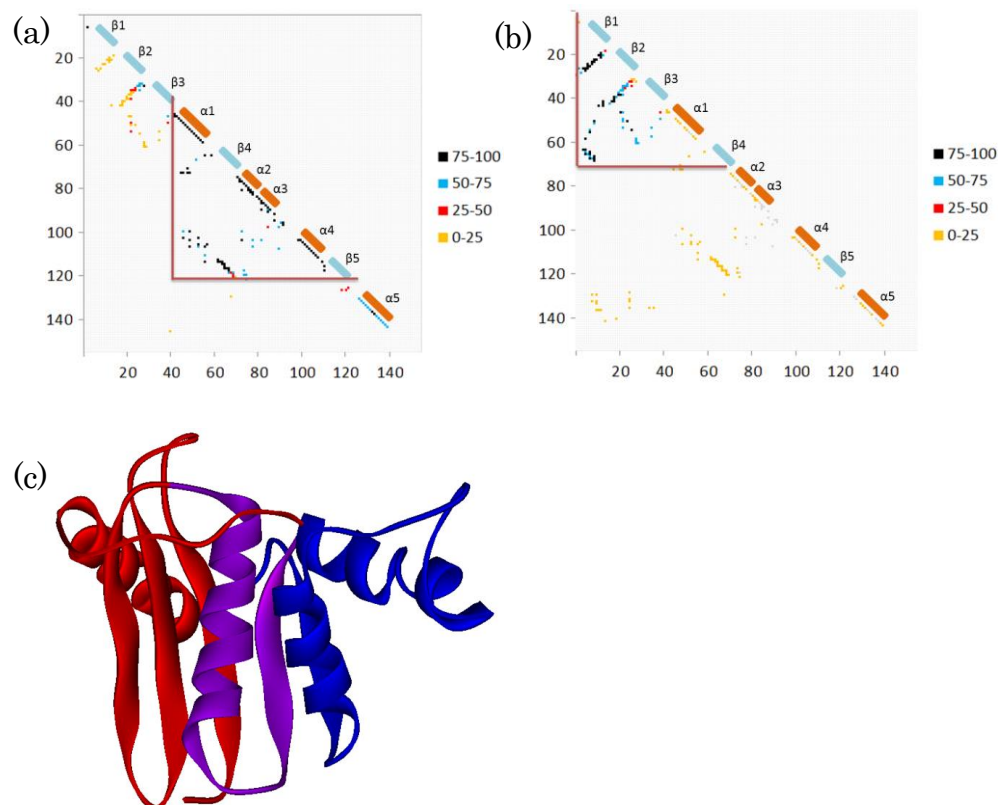


図 2 : (a)RNase H の中間体 ( $Q=0.4$ ) のコンタクト頻度マップ。各点の色は  $Q=0.4$  のアンサンブル中におけるコンタクトの形成頻度を示す。対角線上のリボンは 2 次構造要素を示す。赤線と対角線で囲まれた領域は  $Q=0.4$  のアンサンブル中で安定な領域である。(b)RNase H の中間体から天然構造へフォールドする間に形成される相互作用を示したマップ。各点の値は天然構造 ( $Q=0.7$ ) でのコンタクトの形成頻度と中間体 ( $Q=0.4$ ) のコンタクト頻度の差である。(c)RNase H の構造。中間体で安定な領域と中間体から天然構造への遷移の過程で安定と成る領域を青色と赤色で色分けした。但し、 $\alpha 1$  と  $\beta 4$  は両方の構造の安定化に関与していたため紫で色付けした。

但し、天然状態には準安定なアンサンブルが存在する事が示唆された。この準安定なアンサンブルは C 末端のヘリックスと N 末端の領域との相互作用を形成した状態であり、本研究の結果では C 末端のヘリックスが揺らいでいる構造が遷移温度で最安定となる事が示唆された。この結果は実験的には示されていないため、モデルのアーティファクトである可能性がある。

フォールドとなる領域は天然構造でコンパクトな構造を取っている事が示唆される。その仮説を検証するため、部分配列のコンタクト数を調査し、図 3



にプロットした。このプロファイルには 2 つのピークが観察され、それぞれのピークの示す領域は、30 残基目から両側 35 残基程度、90 残基目から両側 35 残基程度の領域であり、先ほど生じた 2 つのフォールドときれいに一致した。従って、RNase H には 2 つのフォールドが存在し、立体構造からもそれを識別出来る事が示唆された。

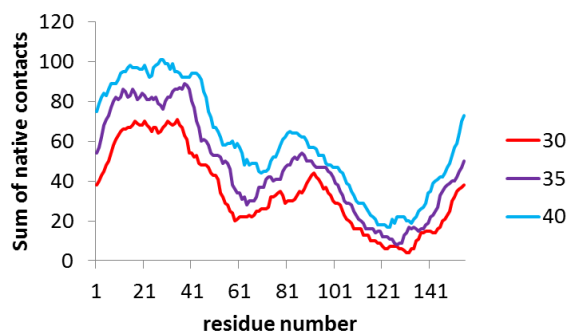


図 3: RNase H の部分配列のコンタクト数を示したプロファイル。配列の任意の残基の両側 30 残基を含む 61 残基の配列内における残基間コンタクトの数を示したプロファイルを赤で、71 残基の配列内における残基間コンタクト数のプロファイルを紫で、81 残基の配列内における残基間コンタクト数のプロファイルを青で示した。配列の末端同士でコンパクトな構造を形成している場合もあるため、N 末端と C 末端がつながっているものとして末端付近のコンタクト数を計算した。

RNase H では 2 つのフォールドを見つける事が出来たため、それぞれの特徴を表 1 に示した。そのコンタクト密度<sup>1</sup>の高さや RCO<sup>30</sup>の高さより、フォールド B の方がフォールド A より高いバリアを生じてフォールドする事が示唆され、実際の計算結果と一致した。但し、フォールド A とフォールド B のフォールディングシミュレーションを独立に行った結果、フォールド A の方が高い安定性を持つことが示された。この安定性の違いはシミュレーション中でフォールド A が先にフォールドした点、またフォールド A は実験的にも安定である事が示唆されている点と一致する。

表 1 : RNase H のフォールドンの持つ特徴

	A	B
配列長	84	73
コンタクト数	106	106
コンタクト数/配列長	1.26	1.42
RCO	0.25	0.31
遷移温度 ( $k_B T_f$ )	0.75	0.70

## Barnase

Barnase の自由エネルギープロファイルを図 4 に示した。RNase H と同様、Barnase の自由エネルギープロファイルでも 2 つのバリアが観察された。実験的には変性状態付近と、その後のフォールディングの途中の自由エネルギーの高い領域に中間体の存在が示唆されている。<sup>23</sup> 変性状態付近の中間体は安定な部分構造をほとんどもたないが、フォールディングの中程で形成される中間体では  $\beta$  シートのほとんどの部分が高い  $\phi$  値を持つため、<sup>24</sup> 本研究で示された中間体はフォールディングの中程で形成される中間体と対応すると考えられる。

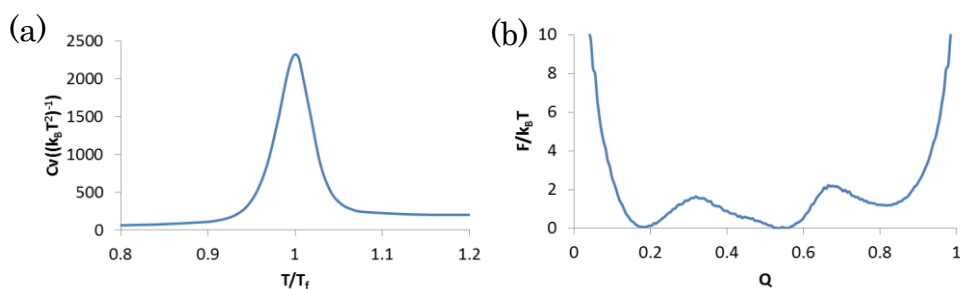


図 4 : Barnase の熱容量曲線(a)と自由エネルギープロファイル(b)。

続いて、Barnase の中間体のコンタクト頻度マップを図 5 に示した。このマップからは、中間体では  $\beta 1$  から  $\beta 5$  までの領域が安定化されている事を示す。実験的には  $\alpha 1$  と  $\beta 1$  から  $\beta 4$  で構成される  $\beta$  シートが中間状態で部分的に安定な構造を形成している事が示されており、その結果と対応する。また、中間体から天然状態の間に形成される領域のコンタクト頻度マップを図 5 に示した。本研究

の結果では中間体を足場にしてその周りの領域がフォールドするという結果を再現した。これは、後半にフォールドする領域はそれ自身が安定な訳ではなく、中間体の構造と合わせてひとつの協同的にフォールドする構造であることを示唆する。従って、本研究では $\beta 1$  から $\beta 5$  までの領域を **Barnase** のフォールドン A、タンパク全体をフォールドン B とした。フォールドン A を紫に、フォールドン A 以外の領域を赤に色付けして図 5 に示した。

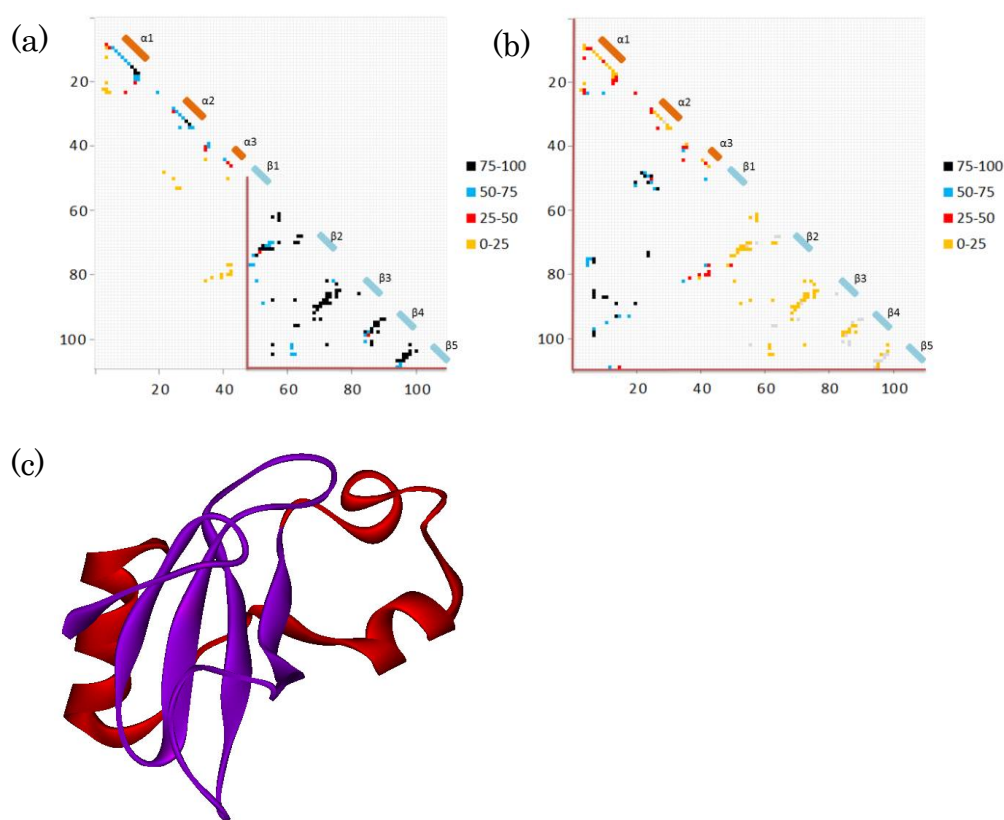


図 5 : (a)Barnase の中間体 ( $Q=0.5$ ) のコンタクト頻度マップ。各点の色は  $Q=0.55$  のアンサンブル中におけるコンタクトの形成頻度を示す。対角線上のリボンは 2 次構造要素を示す。赤線と対角線で囲まれた領域は  $Q=0.55$  のアンサンブル中で安定な領域である。(b)Barnase の中間体から天然構造へフォールドする間に形成される相互作用を示したマップ。各点の値は天然構造 ( $Q=0.8$ ) でのコンタクトの形成頻度と中間体 ( $Q=0.55$ ) のコンタクト頻度の差である。(c)Barnase の構造。中間体で安定な領域と中間体から天然構造への遷移の過程で安定と成る領域を紫色と赤色で色分けした。

フォールドと立体構造との関係を明らかにするために、RNase H と同様に、部分配列のコンタクト数を調査し、図 6 にプロットした。その結果、RNase H とは異なり一つのピークのみが観察され、80 残基目を中心とした部分領域が高いコンタクト数を持つことが示された。

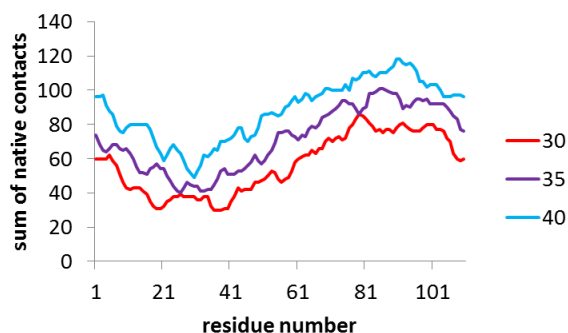


図 6 : Barnase の部分配列のコンタクト数を示したプロファイル。配列の任意の残基の両側 30 残基を含む 61 残基の配列内における残基間コンタクトの数を示したプロファイルを赤で、71 残基の配列内における残基間コンタクト数のプロファイルを紫で、81 残基の配列内における残基間コンタクト数のプロファイルを青で示した。配列の末端同士でコンパクトな構造を形成している場合もあるため、N 末端と C 末端がつながっているものとして末端付近のコンタクト数を計算した。

## LBP<sub>s</sub>

LBP<sub>s</sub> は類似したトポロジーを持つにも関わらず、多様なフォールディング機構を持つ事が示されている。LBP に関しては、個々のタンパクのフォールディング機構だけではなく、この違いは何によって決まっているかに注目した。

hILBP と IFABP の自由エネルギープロファイルを図 7 に示す。本研究では hILBP では 1 つのバリアを、IFABP は 2 つのバリアを再現した。この違いは実験と一致するが、<sup>27,28</sup> 本研究における IFABP の中間状態は過度に安定化されていた。IFABP の野生型は速度論的な中間体をもつものの、平衡では 2 状態での遷移が観察される。しかし、いくつかの 1 残基変異体は 2 状態でフォールドせず平衡状態で安定な中間体が存在する事が示唆されている。<sup>31</sup> 本研究の結果とはなんらかの関係があるかもしれない。

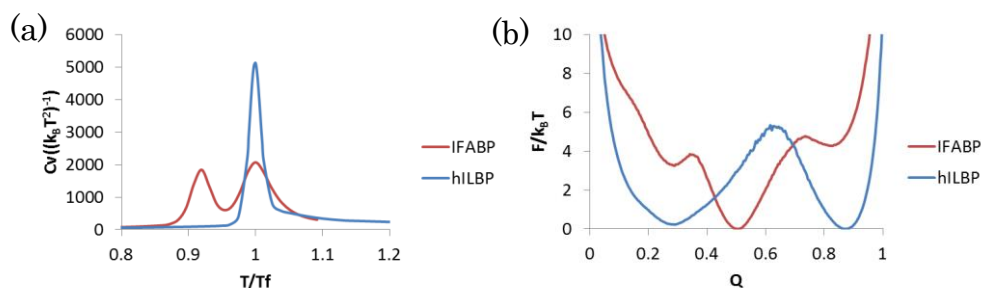


図 7： hILBP と IFABP の熱容量曲線(a)と自由エネルギープロファイル(b)。

続いて、IFABP の中間体のコンタクト頻度マップを図 8(a)に、中間体以前に形成される領域のコンタクト頻度マップを図 8(b)に示した。IFABP の中間状態では $\beta 8$ から $\beta 10$ の間の相互作用と N 末端と C 末端間の相互作用が形成されていた。この結果は、ターンの変異による速度論的な変化を解析した実験とはあまり対応しない。しかし、同じ LBP s である CRABP1 では  $\phi$  値解析により本研究で示された領域と類似した領域が中間体でフォールドしている事が示されている。遷移状態以降は $\beta 2$ から $\beta 7$ までの相互作用が、中間体でフォールドしていた領域と相互作用を形成しながらフォールドする事が示された。IFABP も、Barnase と同様に、中間体を足場にしてその周りの領域がフォールドする結果となったため、IFABP の中間体の構造をフォールドン A、タンパク全体をフォールドン B とした。フォールドン A を紫に、フォールドン A 以外の領域を赤に色付けして図 8(e)に示した。IFABP と同様に、hILBP の遷移状態のコンタクト頻度マップと、遷移状態以降に形成される領域のコンタクト頻度マップを図 8(c)、(d)に示した。hILBP の遷移状態では IFABP とは対照的に、 $\beta 2$ から $\beta 7$ までの領域がフォールドしていた。遷移状態以降は $\beta 1$ 、 $\beta 2$ 、 $\beta 8$ - $\beta 10$ が中間体との相互作用を形成しながらフォールドした。この結果はフォールディングの順序の違いと遷移の回数に関係がある事を示唆する。

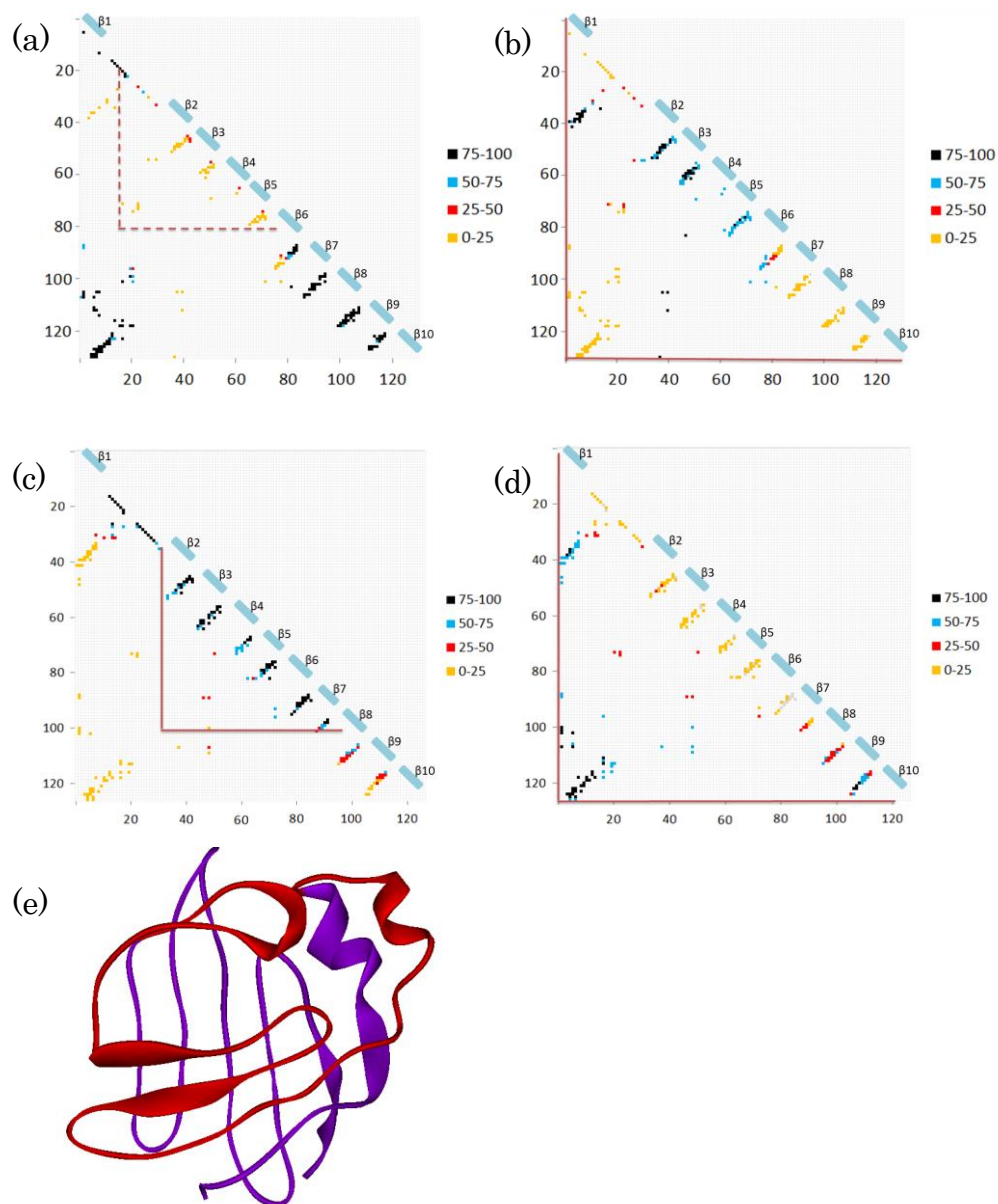


図 8: (a)IFABP の中間体 ( $Q=0.5$ ) のコンタクト頻度マップ。各点の色は  $Q=0.5$  のアンサンブル中におけるコンタクトの形成頻度を示す。対角線上のリボンには 2 次構造要素を示す。赤い点線と対角線で囲まれた領域の外側は  $Q=0.5$  のアンサンブル中で安定な領域である。(b)IFABP の中間体から天然構造へフォールドする間に形成される相互作用を示したマップ。各点の値は天然構造 ( $Q=0.85$ ) でのコンタクトの形成頻度と中間体 ( $Q=0.5$ ) のコンタクト頻度の差である。(c)hILBP の遷移状態 ( $Q=0.5$ ) のコンタクト頻度マップ。赤い線と対角線で囲まれた領域の外側は  $Q=0.5$  のアンサンブル中で安定な領域である。(d)hILBP の遷移状態から天然構造へフォールドする間に形成される相互作用を示したマップ。各点の値は天然構造 ( $Q=0.85$ ) でのコンタクトの形成頻度と遷移状態 ( $Q=0.5$ ) のコンタクト頻度の差である。(e)IFABP の構造。中間体で安定な領域と中間体から天然構造への遷移の過程で安定と成る領域を紫色と赤色で色分けした。



フォールドンと立体構造との関係を明らかにするために、hILBP と IFABP における部分配列のコンタクト数を調査し、図 9 にプロットした。その結果、IFABP は hILBP と比較して C 末端の領域にコンタクトを多く持つことが示された。また、IFABP は 110 残基目を中心とした領域が他の領域と比較して密な構造を取っているが、hILBP には、はっきりとしたピークが見られず、特別密な領域は見当たらなかった。この結果は中間体の形成には相互作用の密度に偏りが必要である事が示唆される。

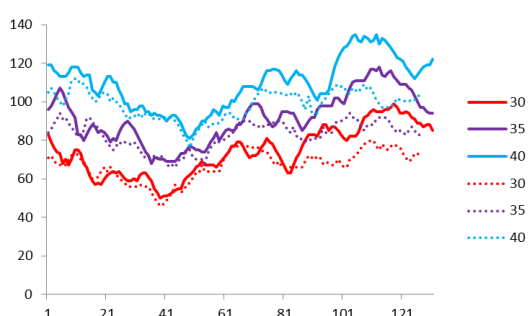


図 9： IFABP の部分配列のコンタクト数（実線）、hILBP の部分配列のコンタクト数（点線）を示したプロファイル。配列の任意の残基の両側 30 残基を含む 61 残基の配列内における残基間コンタクトの数を示したプロファイルを赤で、71 残基の配列内における残基間コンタクト数のプロファイルを紫で、81 残基の配列内における残基間コンタクト数のプロファイルを青で示した。配列の末端同士でコンパクトな構造を形成している場合もあるため、N 末端と C 末端がつながっているものとして末端付近のコンタクト数を計算した。

#### 4.4 考察

##### フォールドンの特性：先行研究との関係

まずこれまでに示されているフォールドンの特徴を改めて言及する。フォールドンとは協同的にフォールドする事のできる構造単位であり、高いフォルダビリティを持つ部分構造として Wolynes らによって定義された。その後、Oliveberg らは、オリジナルの定義とは異なり、実際のフォールドンは配列の一部を共有している可能性や、配列上の離散的な領域が一つのフォールドンを形成する可能性を示した。本研究では、シミュレーションから得られた自由エネルギープロファイルの極小の間で相互作用を形成する領域をフォールドンとみ

なし、すでに提案されているフォールドンの特徴との関係を調査した。

RNase H の自由エネルギープロファイルとフォールディング径路の結果を解析すると、 $\alpha 1$  と  $\beta 4$  を共有する 2 つのフォールドンを示す事が出来た。他方、Barnase や IFABP は RNase H とは異なる特徴を示した。Barnase では  $\beta 1$  と  $\beta 5$  の間の領域が中間状態で形成され、一つのフォールドンを形成する事が示された。しかし、その後は、中間体を足場にしてその周りの領域がフォールドするという結果を得ため、もう一つのフォールドンは中間体の構造を含むタンパク全体である事が示された。IFABP も Barnase と同様の結果を示した。従って、いくつかのパターンはあるものの、Oliveberg らの提案しているように、フォールドン同士は重複を許す方が自然であると考えられる。

RNase H は基本的には  $\alpha 1$  から  $\beta 5$  までの領域 (フォールドン A) と  $\beta 1$  から  $\beta 4$  までの領域 (フォールドン B) が 2 つのフォールドンであることが示唆された。しかし、本研究では天然構造付近に付加的な準安定状態が示され、その間の遷移の過程で C 末端のヘリックスとフォールドン B 間の相互作用が形成されている事が示唆された。実験的にはそのような遷移の過程は特定されていないため、フォールドン B と、C 末端のヘリックスを合わせた構造が本来のフォールドンであるのかもしれない。その場合はタンパクの N 末端と C 末端の領域が一つのフォールドンを形成する事となる。同様に、IFABP でも中間状態ですでに N 末端と C 末端の相互作用がみられた。それらは配列上離れた領域が合わさって一つのフォールドンを形成している可能性を強く示唆する。IFABP の結果では、N 末端と C 末端の間の相互作用はフォールディングの初期に形成されていたが、RNase H では、中間体や他の領域が相互作用を形成することで N 末端と C 末端の物理的な距離が近づき、相互作用を可能にしているように見える。従って、N 末端と C 末端のような配列上離れた領域も、その間にある領域のフォールディングが先に起こる事で、容易に相互作用が可能となり、フォールドンとして機能する事が可能となるだろう。実際、N 末端と C 末端で相互作用しているタンパクは多く、配列の末端同士で協同的にフォールドする部分構造を形成する場合があっても不思議ではない。従って、配列上の離散的な領域も一つのフォールドンとして機能すべきであろう。



## フォールドンの特性：新たな示唆

本研究ではこれまでに示されてきたフォールドンの特徴に加えて、フォールドンに新たな特徴を仮定した場合に実際のフォールディング機構をよく説明出来る事を示唆した。本研究にて示された重要な特徴は、複数のフォールドン間の安定性の違いである。ここでは、複数のフォールドン間の安定性の違いを考慮することで、**Oliveberg** らと **Englander** らのグループの意見の違いを説明できる事を示す。

**RNase H** では、2つのフォールドンを別々にシミュレーションすると、フォールドン **A** の遷移温度 ( $k_B T$ ) は **0.75**、フォールドン **B** の遷移温度は **0.70** であり、両方で安定性が大きく異なる事が示された。しかし、**RNase H** 全体のフォールディングでは、速度論的な中間体の存在は示唆されるものの、熱容量カーブは温度パラメータ **0.774** の点で一つのピークを示し、遷移温度にて構造のポピュレーションが変性状態から天然状態へ移動する事が示唆された。この結果は、フォールドン同士が重なっているため、安定なフォールドンがフォールドすることで、不安定な側のフォールドンが安定化される事を示唆している。**Barnase** 等でも、中間体の形成に関与していない部分は単体では非常に不安定であり、中間体の構造が形成される事で、他の領域のフォールディングが誘発される事が示唆される。

この結果を考慮すると、タンパクの個々のフォールドンの安定性に依存して、**Oliveberg** らが主張しているように複数のフォールドンが平行してフォールドする場合と、**Englander** らの主張するように個々のフォールドンが連続的にフォールディングする場合がある事が示唆される。すなわち、互いの一部を共有している複数のフォールドンがあり、それらの安定性が異なる場合は、必ず安定なフォールドンが先にフォールドし、そのフォールディングに誘発されて不安定なフォールドンがフォールドすると考えるのが妥当であろう。一方、複数のフォールドンの安定性が等しい場合は、それらが別々のタイミングでフォールドする理由が無い場合、アンサンブルとして見た場合はそれらのフォールディングが平行して起きる事が示唆される。但し、そのような場合は本研究の **hILBP** で示されたように、複数回の遷移は起こらず、2状態でフォールドすると考えられる。実際 **RNase H** の実験でも1残基変異によりタンパクの安定性を変化させると中間体が観測できなくなる事が示されている。<sup>19</sup>従って、安定性の異なる複数のフォールドンが存在する場合のみ、複数回の遷移が実験的に観測

される事になると考えられる。

また、本研究の結果ではトポロジーが単純である場合、フォールドンは相互作用エネルギーの偏りに大きく依存する事が示唆される。RNase Hのようにトポロジーのみでタンパクを相互作用が密な領域と粗な領域に分ける事が出来るタンパクではフォールドンとして機能し得る領域はトポロジー上非常に限られている。従って、多少の変異により各残基間相互作用の強さのバランスが変化したとしても、フォールドンの境界は変化しづらいと考えられる。一方、LBPのような単調な構造をもつタンパクでは、エネルギー的に安定性の偏った領域がフォールドンとして機能するため、残基間相互作用の強さのバランスが少し変化しただけで、最初にフォールドする領域が変化すると考えられる。実際LBPは遷移の回数もフォールディングの経路も多様であることが示唆されている。<sup>27-29</sup> また、単調な構造のリピートで形成されているタンパクの研究で類似の現象が示唆されている。<sup>32</sup> 従って、単調な構造を持つタンパク程ファミリー内でそのフォールディング機構が変化しやすくなるだろう。

#### 4.5 結論

本研究では、複数回の遷移を経てフォールドするタンパクのフォールディング機構を  $\text{CaG}\bar{5}$  モデルを用いて調査した。特に、協同的にフォールドする部分構造であるフォールドンの特徴とそのフォールディングにおける役割を明らかにする事を試みた。その結果、フォールドンはそのオリジナルの定義とは異なり、複数のフォールドンは構造の一部を共有していることが示唆された。また、配列上離散的な領域で一つのフォールドンを形成する場合もある事が示唆された。これらの結果はフォールドンの定義を更新する必要がある事を示唆する。他方、タンパク内の部分構造の安定性の違いが、フォールドンの連続的なフォールディングをもたらしている事が新たに示され、フォールドン同士の安定性の違いを考慮する事でそれらが並行してフォールドするのか、連続的にフォールドするのかを識別する事が出来る事が示唆された。フォールドン同士が並行してフォールドする場合は中間体が観測されなくなると考えられる。また、トポロジーが単純である程、そのフォールディング機構は部分構造の安定性の違いに影響を受けやすくなる事が示唆された。独立してフォールドするフォールドンの存在と、それらの連鎖的なフォールディングは、フォールディング機構を決定する大きな要因の一つであると考えられる。

## 参考文献

1. Kamagata K, Arai M, Kuwajima K. Unification of the folding mechanisms of non-two-state and two-state proteins. *Journal of Molecular Biology* 2004;339(4):951-965.
2. Kamagata K, Kuwajima K. Surprisingly high correlation between early and late stages in non-two-state protein folding. *Journal of Molecular Biology* 2006;357(5):1647-1654.
3. Lindberg MO, Oliveberg M. Malleability of protein folding pathways: a simple reason for complex behaviour. *Curr Opin Struct Biol* 2007;17(1):21-29.
4. Gianni S, Guydosh NR, Khan F, Caldas TD, Mayor U, White GW, DeMarco ML, Daggett V, Fersht AR. Unifying features in protein-folding mechanisms. *Proc Natl Acad Sci U S A* 2003;100(23):13286-13291.
5. Xu M, Beresneva O, Rosario R, Roder H. Microsecond folding dynamics of apomyoglobin at acidic pH. *J Phys Chem B* 2012;116(23):7014-7025.
6. Vu DM, Brewer SH, Dyer RB. Early Turn Formation and Chain Collapse Drive Fast Folding of the Major Cold Shock Protein CspA of *Escherichia coli*. *Biochemistry-U S* 2012.
7. Hu WB, Walters BT, Kan ZY, Mayne L, Rosen LE, Marqusee S, Englander SW. Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc Natl Acad Sci U S A* 2013;110(19):7684-7689.
8. Panchenko AR, Luthey-Schulten Z, Wolynes PG. Foldons, protein structural modules, and exons. *Proc Natl Acad Sci U S A* 1996;93(5):2008-2013.
9. Panchenko AR, Luthey-Schulten Z, Cole R, Wolynes PG. The foldon universe: a survey of structural similarity and self-recognition of independently folding units. *Journal of Molecular Biology* 1997;272(1):95-105.
10. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels,

- pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 1995;21(3):167-195.
11. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* 1997;48:545-600.
  12. Onuchic JN, Wolynes PG. Theory of protein folding. *Curr Opin Struct Biol* 2004;14(1):70-75.
  13. Go M. Protein structures and split genes. *Adv Biophys* 1985;19:91-131.
  14. Haglund E, Lindberg MO, Oliveberg M. Changes of protein folding pathways by circular permutation. Overlapping nuclei promote global cooperativity. *J Biol Chem* 2008;283(41):27904-27915.
  15. Lindberg MO, Haglund E, Hubner IA, Shakhnovich EI, Oliveberg M. Identification of the minimal protein-folding nucleus through loop-entropy perturbations. *Proc Natl Acad Sci U S A* 2006;103(11):4083-4088.
  16. Hukushima K, Takayama H, Nemoto K. Application of an extended ensemble method to spin glasses. *Int J Mod Phys C* 1996;7(3):337-344.
  17. Lyubartsev AP, Martsinovski AA, Shevkunov SV, Vorontsovveliyaminov PN. New Approach to Monte-Carlo Calculation of the Free-Energy - Method of Expanded Ensembles. *Journal of Chemical Physics* 1992;96(3):1776-1783.
  18. Fujiwara M, Kato T, Yamazaki T, Yamasaki K, Nagayama K. NMR structure of ribonuclease HI from *Escherichia coli*. *Biol Pharm Bull* 2000;23(10):1147-1152.
  19. Connell KB, Miller EJ, Marqusee S. The Folding Trajectory of RNase H Is Dominated by Its Topology and Not Local Stability: A Protein Engineering Study of Variants that Fold via Two-State and Three-State Mechanisms. *Journal of Molecular Biology* 2009;391(2):450-460.
  20. Raschke TM, Kho J, Marqusee S. Confirmation of the hierarchical folding of RNase H: a protein engineering study. *Nat Struct Biol* 1999;6(9):825-831.

21. Raschke TM, Marqusee S. The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nat Struct Biol* 1997;4(4):298-304.
22. Bycroft M, Ludvigsen S, Fersht AR, Poulsen FM. Determination of the three-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy. *Biochemistry-US* 1991;30(35):8697-8701.
23. Khan F, Chuang JI, Gianni S, Fersht AR. The kinetic pathway of folding of barnase. *Journal of Molecular Biology* 2003;333(1):169-186.
24. Matthews JM, Fersht AR. Exploring the energy surface of protein folding by structure-reactivity relationships and engineered proteins: observation of Hammond behavior for the gross structure of the transition state and anti-Hammond behavior for structural elements for unfolding/folding of barnase. *Biochemistry-US* 1995;34(20):6805-6814.
25. Wong KB, Clarke J, Bond CJ, Neira JL, Freund SM, Fersht AR, Daggett V. Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *Journal of Molecular Biology* 2000;296(5):1257-1282.
26. Kurz M, Brachvogel V, Matter H, Stengelin S, Thuring H, Kramer W. Insights into the bile acid transportation system: the human ileal lipid-binding protein-cholytaurine complex and its comparison with homologous structures. *Proteins* 2003;50(2):312-328.
27. Dalessio PM, Ropson IJ. Beta-sheet proteins with nearly identical structures have different folding intermediates. *Biochemistry-US* 2000;39(5):860-871.
28. Ropson IJ, Boyer JA, Schaeffer BA, Dalessio PM. Comparison of the folding mechanism of highly homologous proteins in the lipid-binding protein family. *Proteins* 2009;75(4):799-806.
29. Burns LL, Ropson IJ. Folding of intracellular retinol and retinoic acid binding proteins. *Proteins* 2001;43(3):292-302.

30. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of Molecular Biology* 1998;277(4):985-994.
31. Chattopadhyay K, Zhong S, Yeh SR, Rousseau DL, Frieden C. The intestinal fatty acid binding protein: the role of turns in fast and slow folding processes. *Biochemistry-Us* 2002;41(12):4040-4047.
32. Ferreiro DU, Cho SS, Komives EA, Wolynes PG. The energy landscape of modular repeat proteins: topology determines folding mechanism in the ankyrin family. *Journal of Molecular Biology* 2005;354(3):679-692.

## 第5章 結語

本研究では第2章にて側鎖の効果をとり込んだ  $\text{CaG}\ddot{o}$  モデルを開発し、小さなタンパクのフォールディング機構を解析し、全原子  $\text{G}\ddot{o}$  モデルの結果<sup>1</sup>と比較した。第3章では第2章にて開発したモデルを用いて ferredoxin-like fold タンパクのフォールディング機構を解析し、複数回の遷移を経てフォールドする U1A と2状態でフォールドする S6 や ADA2h とのフォールディング機構の違いを解析した。第4章では複数回の遷移を経てフォールドするタンパクのフォールディング機構を調査し、協同的にフォールドする部分構造であるフォールドンの有無と、その性質を調査した。本章では、本研究にて開発した  $\text{G}\ddot{o}$  モデルの特徴と、当該モデルを複数回の遷移を経てフォールドするタンパクのフォールディング機構解析に応用する事で得られた知見を述べる。

第2章では、単純な  $\text{CaG}\ddot{o}$  モデルに、残基間相互作用エネルギーの強さのばらつきと、側鎖の配向に依存した鎖の幾何学的な拘束といった効果を加える事で、通常の  $\text{CaG}\ddot{o}$  モデルとは異なり、<sup>1,3</sup>protein L と protein G において全原子  $\text{G}\ddot{o}$  モデルと類似の結果を再現する事を示した。また、3章と4章では2章で用いたタンパクと比較してより複雑なタンパクにおいても、本論文で紹介したタンパクの範囲では実験とよく対応する結果を示した。これらの結果から、以下の結論を導く事が出来る。第一に、本研究では  $\text{G}\ddot{o}$  モデルを用いて研究を行っているため、本研究の結果が実験的に得られているデータと対応する事は、本研究の対象としたタンパクはエネルギーのフラストレーションの少ないタンパクである事を示唆する。第二に、フラストレーションが少ないタンパクに関しては、側鎖の特徴は上述の2つの効果で代用する事ができる事を示唆する。エネルギーのフラストレーションが少ないタンパクは天然構造と競合する準安定な構造を取りにくい<sup>2</sup>ため、天然構造と異なる位置での側鎖同士の接触を考慮する必要がないことに起因していると考えられる。一方、本研究で開発したモデルは  $\text{CaG}\ddot{o}$  モデルとほとんど同じ計算コストで、より正確なフォールディング機構を再現できる事が期待される。実際、通常の  $\text{CaG}\ddot{o}$  モデルでは protein L、protein G のフォールディング機構の違いは再現できない。また、3種類の ferredoxin-like fold タンパクのフォールディング機構の違いも、通常の  $\text{CaG}\ddot{o}$  モデルでは再現する事が出来ない(未発表データ)。これらの結果は、本研究で開発したモデルは、多くの計算コストが必要なより大きなタンパクのフォールディング機構を解析する事が可能である事に適しているだろう。

一方、本研究の後半では複数回の遷移を経てフォールドするタンパクにおいて、個々の遷移の過程で何が起きているかという問題に注目して研究を行った。その中でも鎖のエントロピーと残基間相互作用のエネルギーの差し引きで議論できる、フォールドと自由エネルギープロファイルとの関係<sup>4</sup>を調査した。その結果、本研究で対象とした ferredoxin-like fold タンパク、RNase H、Barnase、LBPs では確かにフォールド<sup>5,6</sup>として独立して協同的にフォールドする領域の存在が示唆された。また、Oliveberg らが提案しているように<sup>4</sup>、複数のフォールドは部分的に重複しており、配列上離散的な領域同士で一つのフォールドを形成する場合がある事も示唆された。これらの結果はフォールドを新たに定義しなおす必要がある事を提示する。また、本研究では新たに、それらの安定性の違いが、フォールドの連続的なフォールディング<sup>7</sup>が起きるために重要な役割を果たしており、複数のフォールドが存在する場合でも、それらの安定性が等しければそれらは並行してフォールドすることが示唆された。ただし、複数のフォールドが並行してフォールドする場合は 2 状態で遷移すると考えられる。以上より、複数回の遷移を経てフォールディングするタンパクのフォールディング機構は、フォールドの数とその安定性で議論する事が出来る事が示された。

ただし、本研究はフォールディングに関わっている物理的な要素の中から鎖のエントロピーの効果のみを抜き出し、その役割を調査したものである。本研究で提案したメカニズムは多くのタンパク中で機能していると考えられるが、本来のフォールディングはもっと複雑であろう。特に、エネルギーのフラストレーションに起因した非天然の相互作用や、水の効果はフォールディング機構をより複雑にし、遷移の数を増やす方向に働くかもしれない。実際、データとしては記載していないが、Myoglobin<sup>8</sup>などいくつかのタンパクでは実験的に示されている複数回の遷移を再現する事が出来なかった。さらに詳細なフォールディングメカニズムを知るためには、本研究では議論する事の出来なかった要素がフォールディングにどのような効果を及ぼすのかを明らかにする事が必要となるだろう。



1. Clementi C, Garcia AE, Onuchic JN. Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: all-atom representation study of protein L. *Journal of Molecular Biology* 2003;326(3):933-954.
2. Onuchic JN, Wolynes PG. Theory of protein folding. *Curr Opin Struct Biol* 2004;14(1):70-75.
3. Koga N, Takada S. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *Journal of Molecular Biology* 2001;313(1):171-180.
4. Lindberg MO, Oliveberg M. Malleability of protein folding pathways: a simple reason for complex behaviour. *Curr Opin Struct Biol* 2007;17(1):21-29.
5. Panchenko AR, Luthey-Schulten Z, Cole R, Wolynes PG. The foldon universe: a survey of structural similarity and self-recognition of independently folding units. *Journal of Molecular Biology* 1997;272(1):95-105.
6. Panchenko AR, Luthey-Schulten Z, Wolynes PG. Foldons, protein structural modules, and exons. *Proc Natl Acad Sci U S A* 1996;93(5):2008-2013.
7. Hu WB, Walters BT, Kan ZY, Mayne L, Rosen LE, Marqusee S, Englander SW. Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc Natl Acad Sci U S A* 2013;110(19):7684-7689.
8. Xu M, Beresneva O, Rosario R, Roder H. Microsecond folding dynamics of apomyoglobin at acidic pH. *J Phys Chem B* 2012;116(23):7014-7025.



## 謝辞

本研究を遂行するにあたり直接ご指導いただいた立命館大学生命科学部教授 菊地武司博士に厚く御礼申し上げます。また、日頃から研究に関する多数の助言を頂いた、現・東北大学東北メディカル・メガバンク機構助教 河合洋介博士、立命館大学生命科学部客員教授 平田文男博士、現・Schrödinger, Inc. Daniel Sindhikara 博士に心から感謝の意を申し上げます。

計算生命化学研究室にてたくさんの議論を行った松岡雅成氏や、計算生命科学研究室の皆様に深く感謝いたします。

最後に、本論文をまとめるにあたり、ご助言を賜りました立命館大学生命科学部教授 高橋卓也博士、立命館大学薬学部教授 加藤稔博士に心から感謝いたします。