A Kinetic Study of Folding Mechanisms of Globular Proteins Using Ultrarapid Mixing Methods

高速溶液混合法を用いた球状蛋白質のフォールディング機構の速度論的研究

Takuya MIZUKAMI

水上　琢也

Cellular Signaling Biophysics laboratory; K-Lab,

Division of Material Science (Physics), Graduate School of Science, Nagoya University

名古屋大学　大学院理学研究科　物質理学専攻（物理系）細胞情報生物物理研究室（K 研）

Dissertation submitted to Nagoya University in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

March, 2013

2013 年 3 月

PREFACE

The protein folding problem is the question of how a protein's amino acid sequence dictates its three-dimensional atomic structure. Understanding the mechanism by which the amino acid sequence of a protein directs its rapid and efficient folding to a unique native conformation involves the efforts of physicists, mathematicians, engineers, as well as those of the biochemists, biophysicists and chemists who have pursued this problem since the seminal work of Anfinsen over fifty years ago. The rapidly expanding interest in the protein folding problem stems from the major impact that its solution is expected to have on our understanding of the molecular basis for biology and medicine and on the opportunities for the design of new materials.

The protein folding problem has come to be regarded as three different problems: (a) the folding code: the thermodynamic question of what balance of interatomic forces dictates the structure of the protein, for a given amino acid sequence; (b) protein structure prediction: the computational problem of how to predict a protein's native structure from its amino acid sequence; and (c) the folding process: the kinetic question of what routes or pathways some proteins use to fold so quickly. The focus of this thesis is to elucidate the kinetic mechanism of the protein folding process from an experimental point of view.

Very significant advances have been made in understanding the mechanism of protein folding by experiments as well as both by computational and theoretical studies. New and improved experimental techniques for mixing solutions and for triggering and detecting folding reactions have led to breaking the millisecond barrier for kinetics, thus opening up the earliest phases of folding to experimental investigation.

The present thesis contains studies carried out at Department of Physics, Division of Material Science, Graduate School of Science, Nagoya University and partly at Fox Chase Cancer Center during April 2008 – December 2012, and is divided into four parts. CHAPTER I provides background of protein folding phenomena, and includes significance and motivation of these studies. CHPATER II deals with a kinetic study on a model protein staphylococcal nuclease using a fluorescence resonance energy transfer-detected continuous-flow method which is a recently developed powerful technique to investigate progress in residue specific structure formation. The author discusses the early structural events based on the site specific information thus obtained. CHAPTER III contains studies on the folding dynamics of horse skeletal muscle apomyoglobin. The author systematically measured the folding/unfolding kinetics, and investigated the best kinetic pathway, which can solo explain the folding/unfolding kinetics and the equilibrium unfolding under various conditions covering the native, urea-induced unfolding and acid-induced unfolding. The author discusses the nature of the folding intermediates on the basis of the kinetic results. Finally, the general conclusions deduced from the present thesis are described in CHAPTER IV.

March 2013

Takuya MIZUKAMI

ACKNOWLEDGMENTS

The author would like to express his grateful acknowledgment to Associate Prof. Kosuke Maki (Nagoya University) for his kind guidance and generous supports throughout this study and for his review of this thesis. This research could not have been completed without his appropriate guide and helpful advices. The author would like to give his deeply heartily gratitude to Prof. Heinrich Roder (Fox Chase Cancer Center, PA) for his giving the great opportunities for the author to study his laboratory and his valuable advices. The author would like to acknowledge Assistant Prof. Naoya Suzuki (Nagoya University) for his supports.

It should be emphasized that the studies in this thesis was carried out with the cooperation of many people.

The author would like to express his sincere gratitude and thanks to Prof. Satoshi Takahashi (Tohoku University) and Dr. Uzawa (RIKEN) for their important comments and suggestions.
The author would like to express his heartfelt gratitude and thanks to Prof. Marc Jamin (European Molecular Biology Laboratory) for his valuable advice.

The author wishes to express his gratitude to Associate Prof. Yasuo Sugiyama (Nagoya University) to kind permission to use his circular dichroism spectropolarimeter.

This work would not have been possible without help of the members in Roder laboratory (Fox Chase Cancer Center). The author is deeply grateful especially to Dr. Ming Xu and to MS. Ruzaliya Fazlieve for their helpful technical supports, valuable advices and suggestions.

The development of the continuous flow devices with a key role in the thesis has required the cooperation of the members of Machine Shop of Nagoya University Technical Center and of Glass Shop of Nagoya University Technical Center. The author would like to heartily acknowledge them.

The author would like to give special thanks to “Nagoya University Program for Leading Graduate Schools Integrative Graduate Education and Research Program in Green Natural Sciences”, “Nagoya University International Academic Exchange Scholarship for Overseas Study Program 2011” and “Japan Society for the Promotion of Science (JSPS) Institutional Program for Young Researcher Overseas Visits” for their providing him financial supports.

I would like to thank all members of Cellular signaling biophysics Laboratory (K-lab) of Nagoya University.

Finally, I wish to thank my family for their support of all my life choices and their love.

CONTENTS

CHAPTER I

General Introduction

I-1. Protein folding

Protein is an important biomolecule involved in virtually every biological process. Proteins are synthesized on the basis of information encoded in DNA as linear polypeptide chains of typically several hundred amino acid residues in living systems. In many cases, the synthesized polypeptide chains exert their biological activities only after they form the unique structures of the individual proteins. Protein folding is the process by which unstructured polypeptide chains obtain their specific native structures. Thus, elucidating the protein folding mechanisms is a critical step toward understanding of the principles of biology and biophysics. One of the major milestones in protein folding study is the finding by Christian Anfinsen and his colleagues.[1] He showed that unfolded and reduced ribonuclease can fold to the native state with the native disulfide bonds even *in vitro*, suggesting that all the information needed for protein folding is encoded in amino acid sequences of protein molecules. In addition, Anfinsen postulated that the newly synthesized polypeptide chain is folded by the free energy gradient and the protein reaches its thermodynamically most stable conformation (thermodynamic hypothesis). Based on the thermodynamic hypothesis, protein folding was widely appreciated to be a physicochemical process governed by thermodynamics.

Although it was revealed that protein folding is governed by thermodynamics, there still remained a question; how the unstructured polypeptide chains fold into the specific native structures. Cyrus Levinthal pointed out that it takes at least 10^9 years for a polypeptide chain with 100 amino acid residues to find out the most stable conformation by random conformational search because there are nearly 10^{30} possible conformations even if individual amino acid residues can adopt only two conformations.[2] Actually, proteins can reach their native state within seconds to minutes. In order to reconcile the temporal discrepancy, an idea of folding pathways was proposed. The idea assumes that protein folding is represented by a series of interconversions from unstructured polypeptide chains to the thermodynamically most stable, specific native structures with accumulation of folding intermediates, i.e., they are assumed to be ensembles of metastable, partially folded states, which are converted into the native state. Since the accumulation of the partially folded intermediates significantly facilitates the folding into the native state by significantly reducing the number of conformations to search, the folding intermediates were regarded as being productive by acting as step-stones to the native structures. The folding intermediates were found by the beginning of the 1970's after the Anfinsen's experiments.[3]

Almost at the same time, partially unfolded states were found to be accumulated under moderately denaturing equilibrium conditions for some proteins.[4,5] Although they are "intermediate states", which are neither the native nor the unfolded sates, their role in the protein folding was unclear. As the properties of these intermediates were characterized in several proteins, they were found to share common characteristics; a considerable amount of native-like secondary structures, a compact molecular size, a solvent-exposed hydrophobic core or hydrophobic surface, and loosely packed tertiary structures. Based on the common structural characteristics of the equilibrium intermediates, Ohgushi and Wada, and Dolgikh *et al*. have proposed that the equilibrium

intermediates may belong to a common physical state of proteins (the molten globule state).[6,7] On the other hand, kinetic studies using stopped-flow methods coupled with various techniques provided us with deeper insights into the conformational change occurring in the early stages of the folding.[8-10] The structure of the equilibrium intermediates of some proteins, as judged by the properties of the secondary structure, was found to be very close to that of their kinetic folding intermediates. For example, Kuwajima *et al*. and Ikeguchi *et al*. have revealed that the equilibrium intermediate of α-lactalbumin is identical to the folding intermediate in terms of the structural properties and the stability by stopped-flow CD measurements.[11,12] Furthermore, the kinetics associated with the native-state formation was similar between the folding intermediates and the equilibrium intermediates for several proteins.[13-15] Taken these similarities between the folding and equilibrium intermediates into considerations, the equilibrium intermediates have been considered to be an equilibrium counterpart of the folding intermediates. The native-like features of both types of the partially folded intermediates indicated that the intermediates are obligatory in folding reactions, which leads to the idea that the folding intermediates play an important role in protein folding.

More recent theoretical studies proposed a new scenario for protein folding, so called a folding funnel.[16,17] The concept of this model is illustrated in Figure I-1 for a shorthand language for communicating the statistical mechanism of protein folding. Figure I-1 represents effective energy as a function of conformation, energy landscapes, where native structures correspond to the energy minima at the center. For the simple and smooth funnel-shaped energy landscape (Figure I-1a), unstructured polypeptide chains folds into the native state quickly without passing through a specific pathway just like a drop of water going down on a funnel. By the way, the result of the Levinthal's thought experiment corresponds to blind putting on a flat energy landscape (Figure I-1b). The discovery that some small proteins with less than 100 amino acid residues fold into their native states without accumulation of any detectable intermediate was taken to provide support for the funnel-shaped energy landscape model.[18] A question then arises: why do many proteins with more than 100 amino acid residues accumulate intermediates during folding? One of the simplest answers to this question is that larger proteins have more rugged and complicated energy landscapes than the smaller proteins (Figure I-1c). Based on the funnel model, the intermediates may be regarded as being non-productive and kinetically trapped by misfolded conformations.[19] Therefore, it is necessary to reconsider the role of intermediates in protein folding through the analyses of the structure, thermodynamics, and kinetics of protein folding in more detail. However, the folding intermediates are usually formed rapidly within milliseconds in contrast to the formation of the native state occurring in a time range from seconds to minutes. Conventional mixing methods such as stopped-flow with a dead time of a few milliseconds were unable to observe the rapid formation of the folding intermediates. In contrast to the growing knowledge on the late stages of the folding, the understanding of the early folding events still remains scarce due to the limit of time resolutions on the experimental techniques.

Recent development in experimental techniques allowed us to directly observe the early stages

of protein folding. For example, ultrarapid mixing methods such as continuous-flow methods have extended the time resolution to tens of microseconds.[20,21] Combining the continuous- and stopped-flow methods allows us to cover the time window over four-to-five orders of magnitudes from ~10 μs to ~100 s. Furthermore, laser-induced temperature jump methods have extended the time resolution to nanosecond time scale under the heat-denaturing conditions.[22,23] In this thesis, the author focused on two issues associated with the early events of protein folding as described below.

I-2. Nonspecific collapse versus specific folding intermediates (CHAPTER II)

Previous kinetic studies on many proteins by means of conventional techniques such as stopped-flow coupled with fluorescence, circular dichroism (CD) and small-angle X-ray scattering (SAXS) have observed missing amplitudes occurring within the dead time of the measurements, typically a few milliseconds, during folding (the burst phase), indicating that a significant conformational change occurs long before the rate-limiting step of folding.[24] However, what conformational change causes the burst phase is a matter of controversy. Particularly, it is unclear whether the early events represent the formation of folding intermediates from the unfolded states driven by specific interactions or whether they reflect a solvent-dependent readjustment of the unfolded state ensemble driven by a non-specific hydrophobic effect.[24-27] For example, Englander and his coworkers had challenged the former scenario on the basis of some measurements on heme-containing peptide fragments of cytochrome *c* (cyt *c*) and on disulfide-reduced ribonuclease A (RNase A). The intact proteins exhibit the burst phase when the folding reactions are measured by the stopped-flow CD and fluorescence, indicating that the developments in the secondary and tertiary structures within milliseconds. Although the fragments of cyt *c* are too unable to assume a folded structure, they nevertheless exhibit similar changes in signal intensity to those of the intact protein within the dead time of the measurements when the denaturant concentration is rapidly reduced. In addition, by assuming that the fragments remain fully unfolded under all conditions, it was concluded that the burst phase reflect a rapid response of the polypeptide chain to the change in solvent conditions.[28,29] The burst phase for fully reduced RNase A, which was assumed to remain in a random denatured state even under non-denaturing solvent conditions, closely matches the initial kinetic amplitude for the protein with all four disulfide bonds intact.[27] These behaviors are supported by some theoretical studies that have proposed that protein molecules, often at the beginning of folding, very rapidly form a compact collapsed ensemble that consists of a large number of kinetically trapped misfolded conformations with non-native structural features rather than the folding intermediates.[30] On the other hand, the structural properties of the polypeptide chains in the early folding (in a few milliseconds) were investigated by pulsed hydrogen/deuterium exchange experiments to reveal the formation of native-like secondary structure. Furthermore, recent development of the continuous-flow methods allowed us to monitor early folding events in a submillisecond time scale. The findings of a barrier-limited rapid condensation of cyt *c* and single-exponential fluorescence decays during early folding of a single-tryptophan variant of RNase A indicate that the ensemble of states formed on the

submillisecond time scale are distinct from the unfolded state.[31,32] Thus, it is essential toward understanding the mechanisms of protein folding to address the fundamental issue whether the early conformational events during folding are driven by specific or by non-specific conformational changes. For this purpose, site-specific information, especially on distances between residues, is valuable because it makes it possible to characterize the initial compaction occurring in the early stage of folding at the residue level.

For this purpose, the author investigated the time-dependent development of tertiary structure during folding of staphylococcal nuclease (SNase) by monitoring distances between pairs of residues using fluorescence resonance energy transfer coupled with continuous-flow and stopped-flow techniques (CHAPTER II). The native structure of this protein consists of two subdomains, the *N*-terminal β-barrel domain and the *C*-terminal α-helical domain. SNase forms a hydrophobic core mainly within the β-barrel domain on the 100-μs time scale of folding. However, it remains unclear whether the chain condensation occurring in the early stages of folding is driven by specific tertiary interactions or non-specific hydrophobic effects (Figure I-2). Variants containing intra- and inter-domain fluorescence donor/acceptor pairs (tryptophan and cysteine-linked fluorescent dye or quencher) were prepared to probe the intra- and inter-domain structural evolution during the folding. The change in the efficiency of the energy transfer during folding and unfolding was measured over a time range from 30 μs to ~10 s at various urea concentrations. By converting the efficiency of the energy transfer into the donor-acceptor distance, the development of the tertiary structure during folding was investigated. The results gave semi-quantitative information on average donor-acceptor distances at different stages of the folding process. The average distance for the donor/acceptor pairs in the β-barrel domain decreases to nearly native values whereas that of the inter-domain donor/acceptor pairs remains unchanged within the first 300 μs of folding. This indicates a rapid non-uniform collapse resulting in an ensemble of heterogeneous conformations in which the central region of the β-barrel domain is well developed while the C-terminal α-helical domain remains disordered. The distance between the inter-domain donor/acceptor pair decreases to native values on the 100-ms time scale, indicating that the α-helical domain docks onto the preformed β-barrel at late stages of the folding. In addition, the unfolded state is found to be more compact under native conditions, suggesting that changes in solvent conditions may induce a non-specific hydrophobic collapse.

I-3. The kinetic mechanism of formation of two types of intermediates (CHAPTER III)

Despite the similarity in the structure and thermodynamics properties of the folding and equilibrium intermediates, it remains unclear whether the formation/unfolding of the folding and equilibrium intermediates occur with a kinetically common mechanism. This is mainly because the temporal resolution of the conventional kinetic techniques has been limited to a few milliseconds. However, recent development of continuous-flow methods coupled with various techniques made it possible to observe the conformational change occurring on submilliseconds time scale.[20]

The author chose apomyoglobin (apoMb) as a model protein for this purpose. ApoMb transiently accumulates folding intermediates long before the native state is accumulated at equilibrium during the folding under strongly native conditions such as at pH 6. ApoMb accumulates equilibrium intermediate states (molten globule states) under moderately denaturing equilibrium conditions such as at pH 4. These intermediates are structurally similar to each other. Both intermediates exhibit similar overall size with native-like secondary structures in the A-, G- and H-helix regions and part of the B-helix region in the native structure. However, since these intermediates are formed under different conditions, it still remains unclear whether these intermediates are identical not only in the structural properties but in terms of their kinetic mechanisms of the formation (Figure I-3). To address the question, the folding and unfolding kinetics and equilibrium unfolding of horse skeletal muscle apoMb were studied over wide ranges of urea concentration and pH at 8°C (CHAPTER III). The fluorescence- and circular dichroism-detected urea-induced equilibrium unfolding and the tryptophan fluorescence-detected pH-induced equilibrium unfolding confirmed that the equilibrium unfolding was consistent with a three-state model involving an equilibrium intermediate (I_{eq}). The urea concentration and pH dependence of folding and unfolding kinetics was measured by continuous- and stopped-flow fluorescence over a time range from 40–135 μs to 10–100 s. The folding kinetics was quantitatively described by a five-state sequential folding scheme regardless of the denaturing factors (urea and pH). The formation of a folding intermediate (I) was observed over a wider range (from 0 M to 2.7 M urea and from pH 5.9 to pH 2.7 for urea- and pH-induced folding, respectively) than the formation of the native state (N) (from 0 M to 1.6 M urea and from pH 6.6 to pH 5.0 for urea- and pH-induced folding, respectively). Based on the difference in stability between I and N, it was found that the stability of N relative to I determines whether I either transiently accumulates during folding or is populated as I_{eq}. In addition, I accumulates as an unfolding intermediate at acidic pH. As a conclusion, the folding/unfolding intermediates, I, and equilibrium intermediate, I_{eq}, consist of a single molecular species. Two other intermediates, I' and M, are observed in the early stages of folding (I') and unfolding (M) reactions.

REFERENCES

1. Anfinsen CB, Haber E, Sela M, White FH, Jr. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci USA **47**:1309-1314.
2. Levinthal C, How to fold graciously. In: P. Debrunner, J.-C. Tsibris,E. Münck, Ed. (1969) In Mössbauer Spectroscopy in Biological Systems University of Illinois Engineering Publications, Urbana, pp. 22-24.
3. Wong KP, Tanford C (1973) Denaturation of bovine carbonic anhydrase B by guanidine hydrochloride. A process involving separable sequential conformational transitions. J Biol Chem **248**:8518-8523.

4. Kuwajima K (1977) A folding model of alpha-lactalbumin deduced from the three-state denaturation mechanism. J Mol Biol **114**:241-258.
5. Holladay LA, Hammonds RG, Jr., Puett D (1974) Growth hormone conformation and conformational equilibria. Biochemistry **13**:1653-1661.
6. Ohgushi M, Wada A (1983) 'Molten-globule state': a compact form of globular proteins with mobile side-chains. FEBS Lett **164**:21-24.
7. Dolgikh DA, Kolomiets AP, Bolotina IA, Ptitsyn OB (1984) Molten globule state accumulates in carbonic anhydrase folding. FEBS Lett **165**:88-92.
8. Kuwajima K, Yamaya H, Miwa S, Sugai S, Nagamura T (1987) Rapid formation of secondary structure framework in protein folding studied by stopped-flow circular dichroism. FEBS Lett **221**:115-118.
9. Baldwin RL (1993) Pulsed H/D-exchange studies of folding intermediates. Curr Opin Struct Biol **3**:84-91.
10. Kataoka M, Goto Y (1996) X-ray solution scattering studies of protein folding. Fold Des **1**:R107-114.
11. Kuwajima K, Hiraoka Y, Ikeguchi M, Sugai S (1985) Comparison of the transient folding intermediates in lysozyme and α-lactalbumin. Biochemistry **24**:874-881.
12. Ikeguchi M, Kuwajima K, Mitani M, Sugai S (1986) Evidence for identity between the equilibrium unfolding intermediate and a transient folding intermediate: A comparative study of the folding reactions of α-lactalbumin and lysozyme. Biochemistry **25**:6965-6972.
13. Raschke TM, Marqusee S (1997) The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. Nat Struct Biol **4**:298-304.
14. Jamin M, Baldwin RL (1998) Two forms of the pH 4 folding intermediate of apomyoglobin. J Mol Biol **276**:491-504.
15. Enoki S, Maki K, Inobe T, Takahashi K, Kamagata K, Oroguchi T, Nakatani H, Tomoyori K, Kuwajima K (2006) The equilibrium unfolding intermediate observed at pH 4 and its relationship with the kinetic folding intermediates in green fluorescent protein. J Mol Biol **361**:969-982.
16. Leopold PE, Montal M, Onuchic JN (1992) Protein folding funnels: a kinetic approach to the sequence-structure relationship. Proc Natl Acad Sci USA **89**:8721-8725.
17. Dill KA, Ozkan SB, Shell MS, Weikl TR (2008) The protein folding problem. Annu Rev Biophys **37**:289-316.
18. Jackson SE (1998) How do small single-domain proteins fold? Folding & Design **3**:R81-R91.
19. Guo ZY, Thirumalai D (1995) Kinetics of Protein Folding: Nucleation Mechanism, Time Scales, and Pathways. Biopolymers **36**:83-102.
20. Shastry MCR, Luck SD, Roder H (1998) A continuous-flow capillary mixing method to monitor reactions on the microsecond time scale. Biophys J **74**:2714-2721.

21. Matsumoto S, Yane A, Nakashima S, Hashida M, Fujita M, Goto Y, Takahashi S (2007) A rapid flow mixer with 11-μs mixing time microfabricated by a pulsed-laser ablation technique: observation of a barrier-limited collapse in cytochrome *c* folding. J Am Chem Soc **129**:3840-3841.
22. Ballew RM, Sabelko J, Gruebele M (1996) Observation of distinct nanosecond and microsecond protein folding events. Nat Struct Biol **3**:923-926.
23. Huang CY, Klemke JW, Getahun Z, DeGrado WF, Gai F (2001) Temperature-dependent helix-coil transition of an alanine based peptide. J Am Chem Soc **123**:9235-9238.
24. Roder H, Colón W (1997) Kinetic role of early intermediates in protein folding. Curr Opin Struct Biol **7**:15-28.
25. Baldwin RL (1996) On-pathway versus off-pathway folding intermediates. Fold Des **1**:R1-R8.
26. Creighton TE (1997) How important is the molten globule for correct folding? Trends Biochem Sci **22**:6-11.
27. Qi PX, Sosnick TR, Englander SW (1998) The burst phase in ribonuclease A folding and solvent dependence of the unfolded state. Nat Struct Biol **5**:882-884.
28. Sosnick TR, Mayne L, Englander SW (1996) Molecular collapse: the rate-limiting step in two-state cytochrome c folding. Proteins **24**:413-426.
29. Sosnick TR, Shtilerman MD, Mayne L, Englander SW (1997) Ultrafast signals in protein folding and the polypeptide contracted state. Proc Natl Acad Sci USA **94**:8545-8550.
30. Dill KA, Chan HS (1997) From Levinthal to pathways to funnels. Nat Struct Biol **4**:10-19.
31. Shastry MCR, Roder H (1998) Evidence for barrier-limited protein folding kinetics on the microsecond time scale. Nat Struct Biol **5**:385-392.
32. Welker E, Maki K, Shastry MC, Juminaga D, Bhat R, Scheraga HA, Roder H (2004) Ultrarapid mixing experiments shed new light on the characteristics of the initial conformational ensemble during the folding of ribonuclease A. Proc Natl Acad Sci USA **101**:17681-17686.

FIGURES

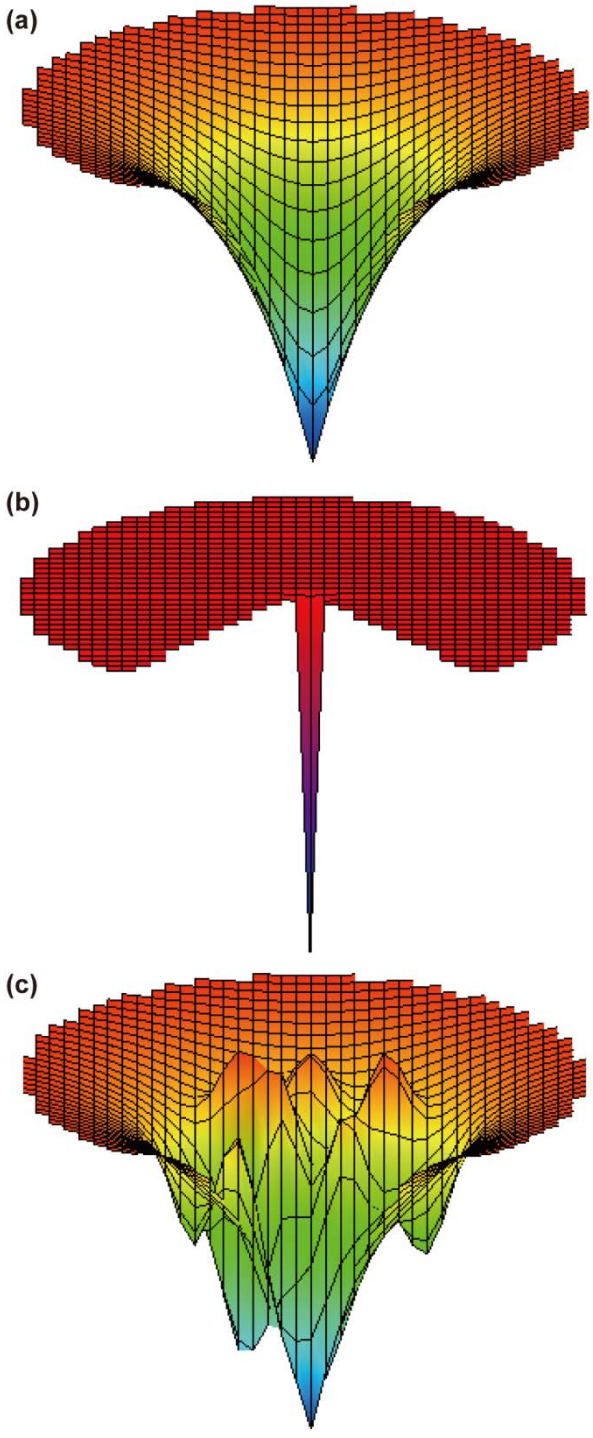


Figure I-1: Cartoons of energy landscape. The effective energy is represented as a function of conformation. (a) A funnel-like smooth energy landscape for a fast-folding protein, (b) a golf course-like energy landscape in which folding is dominated by random search for conformation, and (c) a rugged energy landscape with kinetic traps.

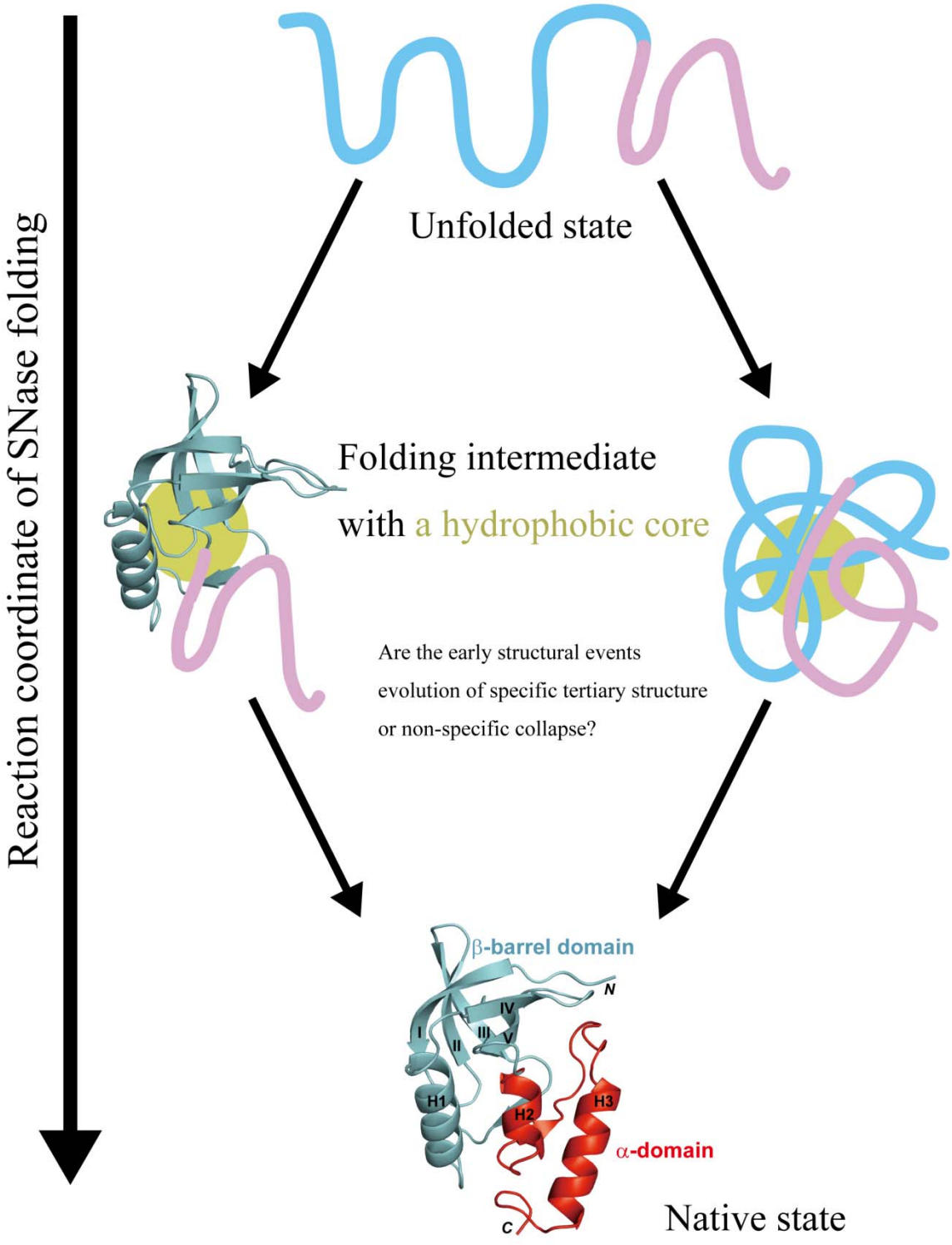


Figure I-2: Two types of scenarios of protein folding . It remains unclear whether the chain condensation duinrng the early stages of folding is driven by specific tertiary interactions (left) or by non-specific hydrophobic effects (right).

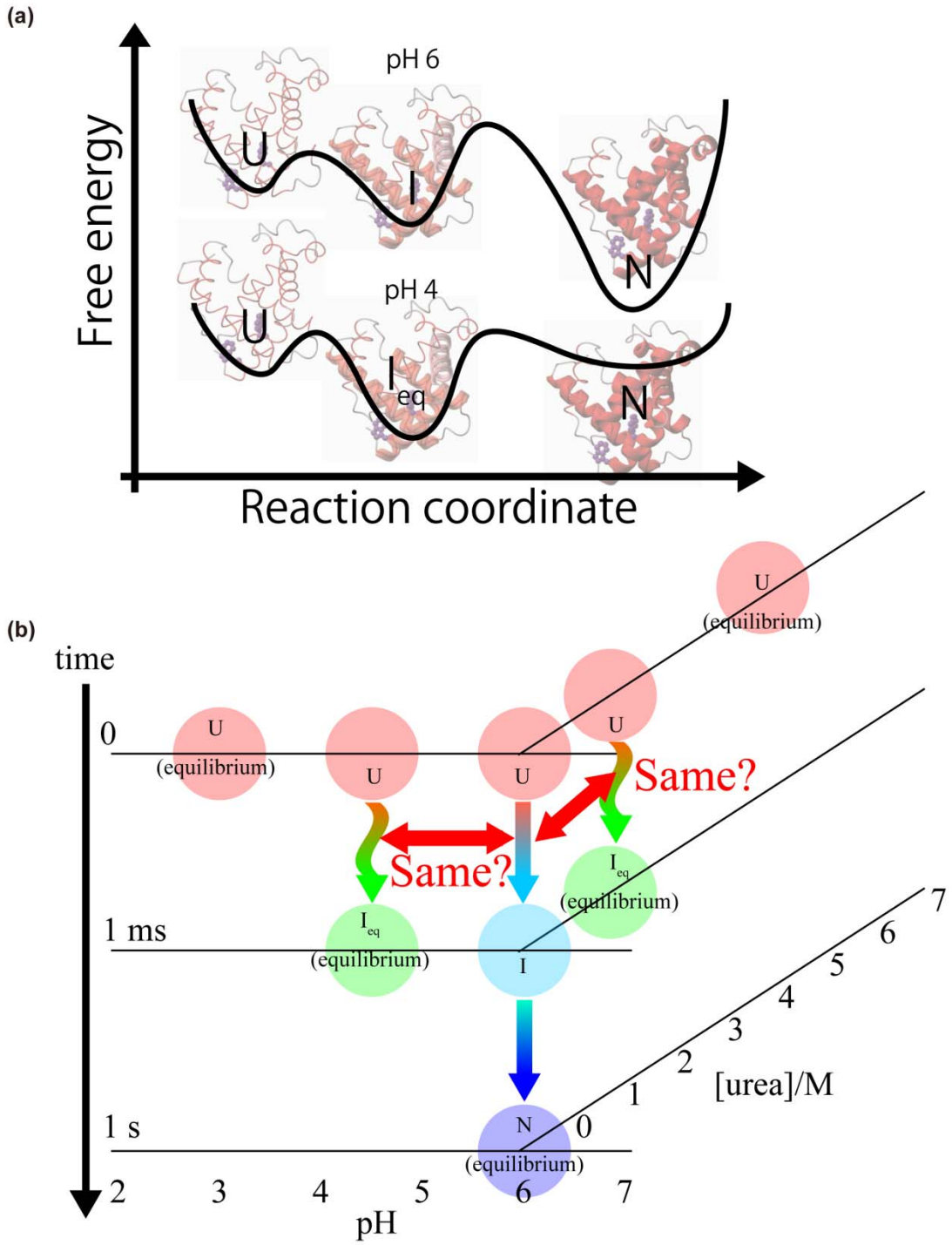


Figure I-3: (a) Free energy diagrams of apoMb under the native (pH 6) and mildly denaturing (pH 4) conditions. U, I, and N are the unfolded, folding intermediate and native states, respectively. I_{eq} is the equilibrium intermediate state with the similar structural properties to those of I. (b) Time-course of structural evolution under various conditions. The folding mechanism within 1 ms is not fully understood.

CHAPTER II

Non-Uniform Chain Collapse during Early Stages of Staphylococcal Nuclease Folding Detected by Fluorescence Resonance Energy Transfer and Ultrarapid Mixing Methods

ABBREVIATIONS

1,5-IAEDANS, 5-((((2-iodoacetyl)amino)ethyl)amino)naphthalene-1-sulfonate;
AEDANS, 5-(((acetylamino)ethyl)amino)naphthalene-1-sulfonate;
ANS, 1-anilinonaphthalene-8-sulfonate;
Bc-Csp, *Bacillus caldolyticus* cold shock protein;
CD, circular dichroism;
CF, continuous-flow;
Cys, cysteine;
DTT, dithiothreitol;
FRET, fluorescence resonance energy transfer;
H/D, hydrogen/deuterium;
NATA, *N*-acetyl-$_{L}$-tryptophanamide;
SF, stopped-flow;
SNase, staphylococcal nuclease;
TNB, 2-nitro-5-thiobenzoic acid;
Trp, tryptophan;
Trp140 SNase, P47G/P117G/H124L variant of SNase;
Trp140/Cys64 SNase, K64C variant of Trp140 SNase;
Trp140/Cys64-TNB SNase, Trp140/Cys64 SNase with modification with TNB at Cys64;
Trp27 SNase, Y27W/W140H variant of Trp140 SNase;
Trp27/Cys64 SNase, Y27W/K64C/W140H variant of Trp140 SNase;
Trp27/Cys64-AED SNase, Trp27/Cys64 SNase with modification with AEDANS at Cys64;
Trp27/Cys97 SNase, Y27W/K97C/W140H variant of Trp140 SNase;
Trp27/Cys97-AED SNase, Trp27/Cys97 SNase with modification with AEDANS at Cys97;
Trp76 SNase, F76W/W140H variant of Trp140 SNase.

INTRODUCTION

How and why unstructured polypeptide chains acquire the specific native structures are important questions to understand the mechanisms of protein folding. Many proteins undergo major conformational changes during early stages of folding resulting in accumulation of compact folding intermediates with well developed secondary structure prior to the rate-limiting step of the folding reactions.[1-3] The initial collapse of the polypeptide chains is mainly driven by hydrophobic effect, which would result in the reduction of the net conformational entropy, facilitating the polypeptide chains to reach the native state.[4] The formation of secondary structure is observed in the regions with the secondary structure elements in the native structure, suggesting its important role to direct the polypeptide chains to the specific native structure. Thus, the formation of the intermediates accompanying the large conformational change occurring in the early folding is expected to be a crucial step for unstructured polypeptide chains to obtain the specificity of the native structure. It follows that a detailed analysis of the structural, thermodynamic and kinetic properties of these intermediates is critical toward understanding the physicochemical principles of protein folding. Kinetic small-angle X-ray scattering (SAXS) and circular dichroism (CD) measurements have provided us with information on the overall size and secondary structure formation during folding. Hydrogen/deuterium (H/D) exchange coupled with NMR has been the primary source of site-resolved information on hydrogen bond formation and secondary structure acquisition during folding.[5-8] However, information on how individual side chains make long-distance interactions remains scarce. For example, it is still an open question whether the initial collapse of polypeptide chain is driven by specific or by non-specific local and long-range interactions. Further, an ensemble of heterogeneous collapsed-states has been observed within the first 1 ms of the folding,[9-12] making it technically challenging to obtain detailed dynamic structural information on these early folding events.

A number of folding studies have used fluorescence resonance energy transfer (FRET) to obtain structural and dynamic information on long-range interactions.[13-24] FRET is a process by which radiationless transfer of energy occurs from an excited state fluorophore (donor) to a second chromophore (acceptor) in the ground state in close proximity through dipole-dipole interactions.[25] The efficiency of energy transfer depends on the extent of spectral overlap of the emission spectrum of the donor with the absorption spectrum of the acceptor, the quantum yield of the donor, the relative orientation of the donor and acceptor transition dipoles, and the separation distance between the donor and acceptor. Especially, the efficiency of energy transfer is extremely sensitive to the separation distance between the donor and the acceptor, and thus FRET measurements have been a valuable tool for probing not only intra-molecular distances but also probing inter-molecular distances or the association/dissociation in macromolecular systems.[13-24] A combination of the mutational method and fluorescence resonance energy transfer is thus a powerful technique to obtain site-specific information of the chain compaction or to follow development of tertiary structure.

Typically a protein is covalently labeled with a donor and an acceptor. In studies of protein structure and folding, a tryptophan (Trp) residue is often used as the intrinsic donor while a fluorophore linked to cysteine (Cys) residue as the acceptor.

Staphylococcal nuclease (SNase) is a 149-residue protein with the intrinsic single-Trp at 140 and no Cys whose structure can be divided into two sub-domains: an N-terminal domain consisting of a five-stranded β-barrel and an α-helix, and a C-terminal domain consisting of two α-helices (Figure II-1)[26]. Prior studies have shown that SNase folds into the native state *via* several partially structured intermediates that accumulate over the 10–500 ms time range along multiple parallel pathways.[27-32] More recent ultrafast mixing measurements showed that additional states accumulate within the first millisecond of folding, including a compact state with a partially structured β-barrel domain.[33,34] NMR-detected H/D exchange studies revealed protected amide protons in strands II and III of the β-barrel domain early during folding (τ ~10 ms) while most other amide protons became protected only during the final stages of folding.[30,32] Previous studies using stopped-flow (SF) CD indicated that significant helical secondary structure is formed on the 10-ms time scale.[28] Walkenhorst *et al.*[29] showed that the state populated at a folding time of ~10 ms is an obligatory intermediate on a pathway directing from the denatured ensemble toward the native state. Taken together, these results indicate that at least two structural events precede the rate-limiting step in SNase folding: a localized chain collapse in the β-barrel region at ~100 μs and the formation of a stable β-hairpin and other elements of secondary structure on the 10-ms time scale. However, it remains unclear (i) whether the early (100 μs) conformational event is driven by non-specific hydrophobic collapse or by specific tertiary interactions, (ii) whether the contacts formed are native-like or non-native, and (iii) whether early structure formation is confined to the β-barrel domain.

In order to observe the development of specific non-local (tertiary) interactions within the β-barrel domain and between the two domains of SNase at various stages of folding, the author carried out a detailed kinetic analysis of a series fluorescence-labeled variants using a FRET-detected continuous-flow (CF) mixing technique[9,20,35] in combination with conventional SF experiments. Two types of single-Trp/dye-labeled Cys variants were prepared: the donor and acceptor sites of one type reside within the β-barrel domain, whereas the other type contains a donor in the α-helical domain and an acceptor in the β-barrel domain. The results indicate that (i) the average distances for all donor/acceptor pairs in the unfolded state of the protein are shorter under folding conditions (low urea concentration) than under strongly unfolding conditions (high urea concentration), suggesting a more compact unfolded state under native condition, (ii) the central region of the β-barrel domain assumes native-like compactness and tertiary interactions within 300 μs of folding, and (iii) the α-helical domain docks onto the β-barrel domain at longer times (>100 ms).

RESULTS

Design of FRET variants

Single-Trp/Cys SNase variants were constructed by substituting appropriate residues of the parent protein containing P47G/P117G/H124L mutations (Trp140 SNase; designated as WT* SNase in previous studies[29,33,34]), which are Y27W/K97C/W140H variant of Trp140 SNase (Trp27/Cys97 SNase), Y27W/K64C/W140H variant of Trp140 SNase (Trp27/Cys64 SNase), and K64C variant of Trp140 SNase, (Trp140/Cys64 SNase). The Cys residue in each variant was chemically modified by either 5-(((acetylamino)ethyl)amino)naphthalene-1-sulfonate (AEDANS) or 2-nitro-5-thiobenzoic acid (TNB)[36,37], resulting in AEDANS-labeled Trp27/Cys97 and Trp27/Cys64 SNase (Trp27/Cys97-AED and Trp27/Cys64-AED SNase, respectively) and TNB-labeled Trp140/Cys64 SNase (Trp140/Cys64-TNB SNase). In the native structure of SNase (Figure II-1), residue 27 is located in the turn between strands II and III, residue 64 in the middle of helix H1, and residue 97 between strand V and helix H2. All three residues are within the β-barrel domain. In contrast, residue 140 is at the loop next to the C-terminal helix H3 in the α-helical domain. Therefore, both donor and acceptor are within the β-domain in the case of the Trp27 variants, but in different domains in the case of the Trp140/Cys64-TNB SNase. The side chains of Lys64 and Lys97 are fully exposed to the solvent in the native structure and expected to be mobile.

Figure II-2 compares fluorescence spectra of the unmodified proteins (selective excitation of Trp at 295 nm) with absorption spectra of the dye-labeled proteins measured at 15°C under the native (100 mM sodium acetate at pH 5.3) and acid-unfolding (~20 mM phosphoric acid at pH 2.0) conditions. Under both conditions the unmodified proteins exhibit fluorescence spectra identical to the parent single-Trp variants, i.e., the quasi-wild type (Trp140 SNase) and the Y27W/W140H variant (Trp27 SNase). The dye-labeled proteins exhibit absorption spectra with dye absorption bands at ~260 nm and ~340 nm for the AEDANS-labeled Trp27 variants or TNB band at ~340 nm for Trp140/Cys64-TNB SNase in addition to the aromatic band at ~280 nm with minor differences under the native and the acid-unfolding conditions. The overlap of Trp fluorescence and dye absorbance bands between 300 nm and 400 nm allows energy transfer to occur from the donor (Trp) to the acceptor (AEDANS or TNB). The Förster distance (R_0), which is the donor-acceptor distance at 50% FRET efficiency, may be changed between under native (R_{0N}) and acid-unfolding (R_{0U}) conditions because the spectroscopic properties of donor are dependent on the conformation of the polypeptide chains. The R_{0N} and R_{0U} values were calculated on the basis of the absorption spectra of the dye-labeled proteins and the Trp fluorescence spectra of the unmodified proteins from 310 nm to 400 nm by Equations (II-1) and (II-2). The calculated R_0 values are listed in Table II-1 and within a range from 21.7 Å to 29.8 Å, which is comparable to C_β-C_β distances between the donor and acceptor sites in the native state (~20 Å) as well as those in the unfolded state (~30 Å; see below). Therefore, these FRET pairs are excellent probes for following the conformational changes associated with SNase folding because even a small change in donor-acceptor distance is detectable

as a significant difference in FRET efficiency (*FE*) near R_0 as represented by Equation (II-3).

Folding kinetics monitored by FRET efficiency

Figure II-3 shows kinetic traces associated with folding of the dye-labeled and unmodified proteins induced by six-fold dilution of the acid-unfolded protein (in ~20 mM phosphoric acid at pH 2.0) with a refolding buffer (100 mM sodium acetate at pH 5.3) at 15°C, resulting in a pH jump from 2.0 to 5.2. This protein is fully in the native state at pH 5.2. The kinetic traces obtained by CF and SF fluorescence experiments under matching conditions (excitation wavelength at 295 nm) are combined to cover the wide time range from 30 μs to 20 s. The dead times of the CF and SF fluorescence measurements were 30 μs and 1.4 ms, respectively. The combined kinetic traces were fitted by a sum of three to five exponential functions. The apparent rate constants and relative amplitudes are listed in Table II-2 comparing with those previously reported. The folding kinetics of the proteins used in this study is similar to that of the parent single-Trp proteins (Trp27 and Trp140 SNase) in terms of the apparent rate constants of the major phases, indicating that the overall folding mechanism is conserved among these variants. However, a few but critical differences are found in terms of the relative amplitudes of the fastest phases. The Trp27Cys97 and Trp27/Cys64 SNase show an *increase* in fluorescence in phases 1 and 3 in the present study while Trp27 SNase has shown a *decrease* in the previous study; an *increase* in fluorescence is observed in phase 1 for Trp140/Cys64 SNase while a *decrease* has been previously reported for Trp140 SNase. These discrepancies can be attributed to the differences in experimental settings, such as the cutoff wavelength of the emission filter. The fluorescence intensities extrapolated to $t = 0$ are consistent with those in unfolded states under the matching native condition, which are obtained by a systematic analysis of the kinetic data (see below), indicating that the whole folding processes from the unfolded states to the native state are detected in the present study.

The fluorescence intensities of Trp27/Cys97-AED and Trp27/Cys64-AED SNase are clearly reduced within the submillisecond time window while those of Trp27/Cys97 and Trp27/Cys64 SNase are enhanced, indicating that the fluorescence which should be enhanced by nature is quenched by FRET to the acceptors approaching. The fluorescence intensities of these variants are quenched within the millisecond–second time window regardless of the acceptors. Thus, this decrease in fluorescence of Trp27/Cys97-AED and Trp27/Cys64-AED SNase is expected to result from the changes in fluorescence properties of donor rather than in the donor-acceptor distance. In contrast to the variants with Trp27, only a slight change in fluorescence intensities of Trp140/Cys64 and Trp140/Cys64-TNB SNase is detected within the submillisecond time window, indicating that the distance between Trp140 and Cys64 remains almost unchanged on this time scale. The fluorescence intensities of Trp140/Cys64 and Trp140/Cys64-TNB SNase are enhanced and reduced, respectively, only after 6 ms of folding. These kinetic behaviors indicate that the strong quenching in donor fluorescence occurs on the millisecond–second time scale by FRET to the acceptor approaching.

The time-courses of FRET efficiency (*FE*) calculated by Equation (II-4) shows an increase on the submillisecond time scale and a following decrease on the millisecond–second time scale for Trp27/Cys97-AED and Trp27/Cys64-AED SNase while only an increase on the millisecond–second time scale for Trp140/Cys64-TNB SNase (Figure II-3). Table II-3 lists the donor-acceptor distances at 0 μs (R(0 μs)), and 300 μs (R(300 μs)) after initiation of folding, and those of the native and unfolded states under the native condition (R_N and R_{UN}, respectively) calculated using Equation (II-3) comparing with those in a random coil (R_{coil}) and in the crystal structure ($R_{crystal}$). The distances for all donor/acceptor pairs in the unfolded state are at least 5 Å shorter under the folding condition in the absence of urea than under the strongly unfolding condition at pH 2.0, indicating that the change in solvent from the unfolding buffer to the refolding buffer makes the protein compact. After the initiation of folding, the intra-domain donor-acceptor distances between Trp27 and the acceptors at positions 64 and 97 approach their native values at t = 300 μs and remain constant throughout the folding reaction, while the inter-domain donor-acceptor distance between Trp140 and the acceptors at positions 64 remains unchanged within 300 μs of folding and approaches the native value only after the native state is formed on the millisecond–second time scale. This critical contrast suggests an idea that the central region of the β-barrel domain acquires native-like compactness and tertiary interactions within 300 μs of folding, and then the α-helical domain docks onto the β-barrel domain at a longer time scale (>100 ms).

Systematic analysis of SNase folding

For a systematic analysis, the folding reactions of the six variants were measured at a various urea concentrations. The reactions were initiated by six-fold dilution of the acid-unfolded proteins (in ~20 mM phosphoric acid at pH 2.0) with a refolding buffer (100 mM sodium acetate at pH 5.3) containing appropriate concentration of urea at 15°C, resulting in a jump from pH 2.0 to pH 5.2 and various urea concentrations. The kinetic traces obtained by CF and SF fluorescence experiments under matching conditions (excitation wavelength at 295 nm) were combined and fitted by a sum of three to five exponential functions. The unfolding reactions of these variants were systematically measured using the SF fluorescence device with the excitation wavelength at 295 nm as well. The unfolding reactions were initiated by six-fold dilution of the proteins dissolved in the refolding buffer (100 mM sodium acetate at pH 5.3) with an unfolding buffer (100 mM sodium acetate at pH 5.3 containing an appropriate concentration of urea) at 15°C, resulting in urea concentration jump from 0 M to various concentrations of urea. The kinetic traces thus obtained were fitted by a single-exponential function.

Figure II-4 compares the urea concentration dependence of the apparent rate constants between the dye-labeled proteins and unmodified proteins, plotting the apparent rate constants logarithmically (chevron plot). In the unfolding region (>1.2 M urea), the apparent rate constants of the four variants with Trp27 show faster values than those of the two variants with Trp140, indicating that the double mutation (Y27W/W140H) results in the destabilization of SNase in the native state. However, the

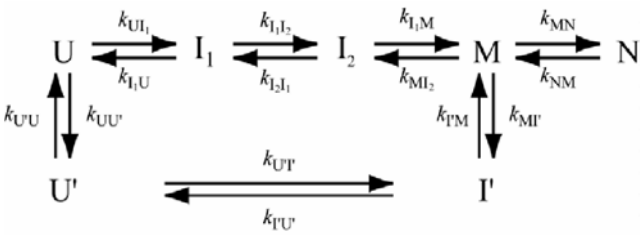
apparent rate constants of all variants show similar behaviors in the folding region (0–1.2 M urea), indicating that these variants fold to the native state by the same mechanism. In addition, the apparent rate constants are almost unchanged by the dye-labeling, indicating that attaching the bulky dye molecules to the variants little affects to the mechanism of SNase folding.

Figure II-5 compares the urea concentration dependence of the fluorescence intensities at the zero time and at a long time after reactions between the dye-labeled proteins and unmodified proteins. For all variants, the fluorescence intensities at the zero time of folding and unfolding show the baselines in the unfolded and native states, respectively. The fluorescence intensities at a long time after reactions show single-transitions between the baselines and are fitted to a two-state equilibrium unfolding model, yielding the thermodynamic parameters (mid-point concentrations, c_m, and m-values for the transitions; Table II-4) and spectroscopic parameters (fluorescence extrapolated to zero urea concentration; Figures II-3 and II-5; and the corresponding slopes of the native and the unfolded states). All variants are somewhat destabilized by the mutations and/or the side chain modifications than Trp140 SNase but are still fully folded into the native state under the native conditions.

DISCUSSION

Folding kinetics of SNase variants is consistent with that of the parent proteins

The folding kinetics of SNase is described by Scheme II-1 (excluding the slow processes reflecting isomerization of peptidyl prolyl bonds), which was proposed on the basis of previous studies on Trp140 SNase and a series of variants, including a proline-free variant and seven single-Trp variants.[29,33,34,38]



Scheme II-1

In Scheme II-1, U and U' represent the unfolded species on the major and minor folding pathways, respectively, while I_1, I_2, I' and M are intermediate ensembles, and N is the native state. I_1 is required to explain the fastest phase (phase 1), which accumulates on the submillisecond time scale under native conditions. In the case of Trp140 SNase, the close similarity of the fluorescence between I_1 and I_2 accounts for a lag phase (phase 3) observed at ~10 ms. Accumulation of I_2 also explains the curvature in the fourth and dominant phase (phase 4) of the chevron plot. Phase 4 is the rate-limiting step of the folding reaction, and the corresponding rate constant, λ_4, is strongly dependent on denaturant concentration. A high-energy intermediate with native-like properties, M, is introduced to account for the curvature on the unfolding limb of the chevron plot, which is the result of a switch in the rate-limiting step for unfolding.[29,39,40] An intermediate on the secondary folding pathway, I', is

required to account for the second fastest phase (phase 2) while a minor slow phase (phase 5) comes from the conversion from I' to M. In addition to these processes, a much slower and largely denaturant-independent phase (phase 6) arises from the isomerization of the peptidyl prolyl bonds (not shown in Scheme II-1) but is not considered further in this study.

The urea dependence of the (un)folding reaction gives further insight into the folding mechanism of SNase. Figure II-4 shows chevron plots for the SNase variants studied. The fastest phase for the four variants with Trp27 is assigned to phase 1 since its rate constants are similar to that of the Trp27 SNase (λ_1 ~ 5,500 s^{-1}). The slowest phase of Trp27/Cys97-AED SNase with an apparent rate constant of ~ 4 s^{-1} and the second slowest phases (4.1 s^{-1} to 6.0 s^{-1}) of Trp27/Cys97, Trp27/Cys64-AED, and Trp27/Cys64 SNase are assigned to phase 4. In terms of relative amplitude this is the dominant phase for variants with Trp27. The rate constants of the second fastest phase of Trp27/Cys97-AED (115 s^{-1}), Trp27/Cys97 (136 s^{-1}) and Trp27/Cys64 (130 s^{-1}) SNase are an order of magnitude larger than λ_3 of Trp27 SNase (11 s^{-1}). However, this process resembles phase 3 of Trp27 SNase in terms of urea dependence and the fact that it precedes the rate-limiting step (phase 4). The slowest phase of Trp27/Cys97, Trp27/Cys64-AED, and Trp27/Cys64 SNase (phase 5) may correspond to the late steps along the secondary folding pathway (phases 5 and 6), the rate constants of which became too close to each other to be resolved.[34]

Similarly, the fastest phase observed for the Trp140/Cys64-TNB and Trp140/Cys64 SNase is comparable to that of Trp140 SNase (12,700 s^{-1}) and can be assigned to phase 1. In addition, the second fastest phase of the unmodified protein, with a rate constant comparable to λ_3 of Trp140 SNase (150 s^{-1} *vs*. 26 s^{-1}), represents a lag phase and was assigned to phase 3. Both Trp140/Cys64 and Trp140/Cys64-TNB SNase exhibit major phases with rates comparable to the rate-limiting folding phase for Trp140 SNase (λ_4 = 11 s^{-1}). The slower phases of Trp140/Cys64 SNase and its TNB derivative are assigned to phases 5 and 6. Trp27/Cys64-AED and Trp140/Cys64-TNB SNase lack phase 3, probably due to its low amplitude (see below). Some of the fluorescence changes observed in this study differ from a previous study in terms of the sign of the amplitude of phases 1 and 3.[34] These discrepancies in the sign can be attributed to the differences in experimental settings, such as the cutoff wavelength of the emission filter, and show no evidence for change in the folding mechanism by the mutations and modifications.

As shown in Figure II-4, for all variants, the rate constant of the unfolding reaction (λ_u) decreases with decreasing urea concentration, and coincides with that of the dominant folding phase 4, λ_4, in the transition region (Figure II-5), indicating that phase 4 is the rate-limiting step. The chevron plots of Trp140/Cys64 and Trp140/Cys64-TNB SNase (Figure II-4) are consistent with that of Trp140 SNase while those of the variants with Trp27 are qualitatively similar to that of Trp76 SNase, a destabilized single-Trp variant.[33] The rate constants of phases 1 (λ_1) and 3 (λ_3) are virtually independent of urea concentrations below 1 M. The roll-over (downward curvature of the chevron plot) of the folding limb of λ_4 of Trp140/Cys64 and Trp140/Cys64-TNB SNase can be attributed to accumulation of I_2 on the major pathway (Scheme II-1), and can be approximated as λ_4 =

$k_{I1I2}/(k_{I1I2}+k_{I2I1})\times k_{I2M}$. The kink in the unfolding limb of the chevron plot for the variants with Trp27 is attributed to a high-energy native-like intermediate (M), as reported previously for Trp140 SNase and other variants.[29,33] Qualitatively, the kinetic behavior of the all variants used in this study is similar to that of the parent protein, Trp140 SNase, and can thus also be described by Scheme II-1.

The initial signal intensities of the kinetic folding and unfolding data (extrapolated to $t = 0$) are fully consistent with the fluorescence values of the unfolded and the native states at equilibrium, respectively (colored lines in Figure II-5). Direct comparison between the equilibrium transitions and the (un)folding kinetics under identical final conditions shows no evidence for kinetically unresolved changes. FE_{UN} is comparable to the initial FRET efficiency of the pH-induced folding reaction, consistent with the absence of a burst phase. The variants with Trp27 in the native state are less stable than the Trp140/Cys64 variants due to the double mutation (Y27W/W140H), since the naturally occurring Trp140 is important for the stability of the native state.[41] In most cases, the replacements of Lys by Cys (at positions 64 and 97) as well as addition of an AEDANS or TNB group destabilize the protein. Although the destabilization occurs for the all variants used here, the rate constants observed during folding remain unchanged from the parent protein. Thus, the variants used in the present study fold to the native state *via* the pathways represented by Scheme II-1, conserving the folding mechanisms from the parent proteins.

Evolution of tertiary structure accompanying SNase folding

For the quantitative analysis of the FRET efficiencies, donor quenching (Trp emission changes) is used but not enhancement in acceptor (AEDANS) fluorescence because of the following reason. In order to estimate FRET efficiency, two reference fluorescence values are usually required, for example those at 0% and 100% FRET efficiencies. However, it is difficult to obtain the acceptor fluorescence at 100% FRET efficiency in a protein because it is achieved when the acceptor is in contact with the donor. In contrast, the donor fluorescence intensity at 100% FRET efficiency should be zero and that at 0% efficiency should be consistent with that in the protein without the acceptor, which is easily and accurately measured. Thus, the FRET efficiency estimated by donor quenching is expected to be more reliable than that by enhancement in acceptor fluorescence.

The overall change in time-dependent FRET efficiency during folding is different for the variants with Trp27 and with Trp140. Specifically, for Trp27/Cys97-AED and Trp27/Cys64-AED SNase, *FE* increases from 0.39 to 0.53 whereas it remains essentially unchanged (0.33 to 0.35) for the Trp140/Cys64-TNB within the first 300 μs. This is an indication that the β-barrel domain forms independently from the α-helical domain. In contrast, *FE* increases significantly during the late stages of the folding for the Trp140/Cys64-TNB while a decrease is observed for Trp27/Cys97-AED and Trp27/Cys64-AED SNase. It should be noted that the time-dependent traces of *FE* were not fitted to Equation (II-5) because $FE(t)$ is given by Equation (II-4), where the $F_D(t)$ and $F_{DA}(t)$ are given by Equation (II-5), in addition to the small difference in rate constants of corresponding phases between the dye-labeled and the unmodified proteins.

The R_U is smaller than R_{coil} and the difference between R_U and R_{coil} increases with increasing distance along the sequence. This trend was found in other proteins,[21] and was attributed to the residual structure in the acid-unfolded state. R_N is ~2 Å larger than $R_{crystal}$ probably because of potential errors in estimating R_0 values: (i) the overlap integral can be measured only above 310 nm because of aromatic absorbance and scattering backgrounds, (ii) the assumption that donors and acceptors rotate freely (yielding an orientation factor $\kappa^2 = 2/3$) may not be valid under native conditions (Trp27 and Trp140 are largely buried in the native structure), and (iii) the uncertainty in determining the position of the bulky dye may reflect a shift in the position of the acceptor compared to that in the crystal structure. For example, the aromatic region in the AEDANS-labeled histidine residue is at least 10.0 Å far away from its C_β atom in the crystal structure of the modified ribonuclease A.[42]

A previous study suggested that 65–80% of the molecules on the major pathway are converted into I_1 at $t = 1/\lambda_1$. Because the time constant of phase 1 ranges from 60 μs to 230 μs for the dye-labeled proteins, R(300 μs) approximates the donor-acceptor distance in I_1. The side chain of Trp27 and 140 are solvent-exposed in the unfolded state under native condition, and probably also in I_1 as expected from the only slight increase in fluorescence intensity from that in the unfolded state, justifying the use of R_{0U} as a Förster distance to determine donor-acceptor distances (R(0 μs), R(300 μs) and R_{UN}).

The distances between Trp27 and the acceptors at positions 64 and 97 are within 1–2 Å of R_N at $t = 0$. These donor/acceptor pairs span the β-barrel domain with Trp27 located at one end and Cys64 and Cys97 (on helix H1 and the loop between strand V and helix H2, respectively) adjacent to the opposite end (Figure II-1). In both cases, R(0 μs) is much smaller than R_U, indicating that the β-barrel domain already adopts a compact conformation at $t = 0$. The fact that two different probes separated by over 30 residues along the chain approach the donor to within 2 Å of the final distance is consistent with an overall compactness and arrangement of the β-strands and helix H1 resembling the native structure. The donor-acceptor distances approach native values at $t = 300$ μs and remain constant throughout the folding reaction (Table II-3). The observation that Trp27/Cys97 and Trp27/Cys64 SNase undergo only a slight change in fluorescence during phase 1 suggests that the Trp27 side chain is still exposed and relatively mobile at this stage. This further supports the conclusion that the observed changes in FRET efficiency reflect a major conformational change rather than changes in donor environment or mobility. A previous study on SNase folding probed by 1-anilinonaphthalene-8-sulfonate (ANS) binding showed that solvent-accessible hydrophobic pockets appear within ~100 μs of folding.[33] Since there is no evidence for large hydrophobic cores elsewhere in the protein, the binding site(s) for ANS should be located within the β-barrel domain. Taken together, these results indicate that the β-barrel domain becomes nearly as compact as the native state and forms native-like tertiary interactions within 300 μs of folding.

In contrast to Trp27/Cys97-AED and Trp27/Cys64-AED SNase, the donor-acceptor distance for Trp140/Cys64-TNB SNase decreases substantially as the native state is formed (Table II-3). R(300

μs) is ~4 Å larger than R_N, whereas the corresponding distances for the variants Trp27/Cys97-AED and Trp27/Cys64-AED SNase are very close to the native values. The distance between Trp140 and TNB attached to Cys64 at the beginning of the folding reaction, $R(0\ \mu s) = 26.3 \pm 0.4$ Å, is close to R_{UN} (25 ± 1 Å), remains essentially constant throughout the early stages (~10 ms) and decreases to its native value (22.0 Å) only during the final folding phase. A slight enhancement in donor fluorescence during phase 1 observed for the unlabeled protein (Trp140/Cys64 and Trp140 SNase) due to a change in the local environment of Trp140 is probably responsible for the small increase in apparent FRET efficiency at $t \sim 100$ μs (Figure II-3). Therefore, the C-terminal region close to helix H3, where Trp140 is located, is not yet docked onto the β-barrel domain, but may form some local structure around Trp140 on the 100-μs time scale. The final folding phase with a rate of ~10 s^{-1} is accompanied not only by a major enhancement in Trp140 fluorescence of the unlabeled protein, but also a significant quenching of the donor in the presence of the TNB acceptor (Figure II-3) consistent with a ~4 Å decrease in donor-acceptor distance. Taken together, the results for all three FRET pairs indicate that during the rate-limiting step of folding the C-terminal helix docks onto the preformed β-barrel domain. Since information on intra-domain collapse within the α-helical domain still lacks, the possibility is not ruled out that both domains acquire some structure at an early stage and dock to each other during the rate-limiting step of folding, as reported for dihydrofolate reductase, an α/β-type protein.[24]

For all three donor/acceptor pairs, the average distance in the acid-unfolded state at pH 2.0 is substantially larger (by at least 5 Å) than that of the unfolded state at pH 5.2 (based on extrapolation from the urea-unfolded state at equilibrium) and the kinetically observed state at the beginning of the folding reaction (based on extrapolation of the kinetics to $t = 0$). The similarity among the variants indicates that under folding conditions (pH 5.2, low denaturant concentration) the polypeptide chain is uniformly more compact than under unfolding conditions (pH 2 or high urea concentration). It appears that this state, which can be considered the unfolded state under native conditions, has undergone a general collapse of the chain driven by a combination of hydrophobic and electrostatic interactions. The low affinity for ANS suggests that this conformational ensemble is mobile and lacks a persistent hydrophobic core.[33] In contrast, the donor-acceptor distances differ significantly among the variants in the transient state populated at 300 μs, indicating that specific side chain interactions are formed within the first few 100 μs of folding. If this process was a uniform compaction, one would expect similar donor-acceptor distances for the Trp27/Cys97 and Trp140/Cys64 pairs, since their separation along the chain is similar (71 and 77 residues, respective). These findings indicate that the earliest intermediate ensemble formed on the 100-μs time scale is structurally bimodal: the core structure of the β-barrel domain is formed and has native-like compactness while the C-terminal helix is still unfolded and/or detached; the two subdomains associate only during the final stages of folding.

Magg *et al*. suggested on the basis of detailed kinetic experiments that a partially collapsed state of *Bacillus caldolyticus* cold shock protein (*Bc*-Csp) represents the unfolded state under native

conditions rather than a barrier-limited intermediate.[18,21] Although the overall topology is different, both SNase and *Bc*-Csp contain a β-barrel structure, and the partially collapsed initial state might be a common feature. One possibility to consolidate these findings is that the starting conformation of SNase at $t = 0$ is analogous to the compact unfolded ensemble of *Bc*-Csp, whereas the I_1 state populated on the 100-μs time scale represents a partially structured intermediate not present in the smaller cold-shock protein.

The FRET probes studied here, along with prior observations, provide qualitative information on the structural properties of the I_1 intermediate. The microsecond folding kinetics of an SNase variant was previously studied which was containing a Trp at position 76,[34] located at the C-terminal end of β-strand IV on the same face of the β-barrel as Trp27. The observation that the fluorescence of both Trp27 and Trp76 is partially enhanced during phase 1 is consistent with the assembly of a loosely packed cluster of hydrophobic side chains at one end of the β-barrel. The presence of a dynamic clustering of hydrophobic side chains is supported by the fact that ANS binds preferentially to intermediate populated on the 100-μs time scale.[33] The near-native distance for the Trp27/Cys64-AED pair suggests that helix H1 is already formed and interacting with the structured region of the I_1-state. The Cys97-AED FRET probe also places the loop following strand V within the collapsed core of the intermediate. In addition, earlier H/D exchange labeling studies showed that amide protons in β-strands II and III are already protected from exchange at this early stage of folding, suggesting the presence of a hydrogen-bonded β-hairpin.[30,32] Thus, the chain compaction observed here appears to encompass most of the β-barrel as well as helix H1. While the combined results are consistent with native-like topology, more detailed structural information will be needed to confirm this conclusion (e.g., we have no probes reporting on the structure of β-strand I). On the other hand, the observation that the distance between Trp140 and Cys64 remains unchanged in phase 1 demonstrates that helix H3 is still detached from the structural core of the I_1 intermediate. The observation of a bimodal $P(r)$ function in SAXS measurements on a C-terminally truncated SNase variant[43] further supports the conclusion that an early event in SNase folding gives rise to a partially structured state with a compact β-barrel/H1 subdomain and a more disordered C-terminal region.

CONCLUSIONS

It has been a longstanding question on protein folding whether the initial collapse of polypeptide chain is driven by specific or by non-specific local and long-range interactions. In order to address the question for the folding of SNase, the author investigated its folding kinetics at the residue level. SNase exhibits a localized chain collapse in the β-barrel region at ~100 μs and the formation of a stable β-hairpin and other elements of secondary structure on the 10-ms time scale. However, it remains to be understood: (i) whether the early conformational event is driven by non-specific hydrophobic collapse or by specific tertiary interactions, (ii) whether the contacts formed are native-like or non-native, and (iii) whether early structure formation is confined to the β-barrel

domain. The folding kinetics of two types of single-Trp/fluorescent dye-labeled Cys variants, where the donor and acceptor sites of one type reside within the β-barrel domain while the other type contains a donor in the α-helical domain and an acceptor in the β-barrel domain, were measured using a FRET-detected CF and SF mixing technique. The results indicate that (i) the average distances for all donor/acceptor pairs in the unfolded state of the protein are shorter under folding conditions than in strongly unfolding conditions, suggesting a more compact unfolded state under native condition, (ii) the central region of the β-barrel domain assumes native-like compactness and tertiary interactions within 300 μs of folding, and (iii) the α-helical domain docks onto the β-barrel domain at longer times (>100 ms). These conclusions are summarized in Figure III-6.

MATERIALS AND METHODS

Protein expression, modification and purification

Site-directed mutagenesis was carried out by QuickChange (Agilent, CA, USA). The single-Trp/Cys variants of SNase (Trp27/Cys64, Trp27/Cys97 and Trp140/Cys64 SNase) constructed were expressed and purified according to a previous report.[34] Prior to the modification of Trp140/Cys64 SNase with TNB, 100 mM Tris-HCl / 7 M urea / 2 mM dithiothreitol (DTT) (pH 8.0) was purged by Argon gas for two hours to remove oxygen in the buffer, followed by addition of lyophilized Trp140/Cys64 SNase and concentrated 5,5'-dithiobis-(2-nitrobenzoic acid) solution with 157-fold molar excess to the protein. The reaction was performed in dark at room temperature for two hours. The unmodified protein and side-reaction products were removed by ion exchange chromatography (Bio-Rex70, Bio-Rad, CA, USA). The fraction of modified protein was dialyzed against ~50 mM ammonium bicarbonate (pH ~ 7), followed by lyophilization. The labeling efficiency (percent of the protein molecular labeled) was estimated by the ratio of Trp140/Cys64 SNase to TNB attached to the protein through independently determined concentrations of TNB and Trp140/Cys64 SNase. Specifically, the labeled protein in Tris-HCl (pH 8.0) was reduced by ~1 mM DTT, resulting in a mixture of free TNB and Trp140/Cys64. The concentrations were then determined spectrophotometrically by the molar extinction coefficients of free TNB[37] ($\varepsilon^{TNB}_{412\ nm}$ = 14,150 M^{-1} cm^{-1} and $\varepsilon^{TNB}_{280\ nm}$ = 2,900 M^{-1} cm^{-1}) and Trp140/Cys64 SNase ($\varepsilon^{Trp140/Cys64}_{280\ nm}$ = 14,700 M^{-1} cm^{-1}). AEDANS labeling of Trp27/Cys64 and Trp27/Cys97 SNase was carried out in a similar way of TNB-labeling with minor change in the procedure. Specifically, a 100-fold molar excess of 5-((((2-iodoacetyl)amino)ethyl)- amino)naphthalene-1-sulfonate (1,5-IAEDANS) was used for the modification. The labeled protein was purified by Sephadex G-50 and dialyzed against ~50 mM ammonium bicarbonate (pH ~ 7), followed by lyophilization. The labeling efficiency was estimated by measuring the ratio of each variant to AEDANS attached to the protein. Their concentrations were determined spectrophotometrically according to the molar extinction coefficients of free 1,5-IAEDANS[36] ($\varepsilon^{AEDANS}_{336\ nm}$ = 5,700 M^{-1} cm^{-1} and $\varepsilon^{AEDANS}_{280\ nm}$ = 1,310 M^{-1} cm^{-1}), Trp27/Cys64 ($\varepsilon^{Trp27/Cys64}_{280\ nm}$ = 12,700 M^{-1} cm^{-1}), and Trp27/Cys97 ($\varepsilon^{Trp27/Cys97}_{280\ nm}$ = 12,700 M^{-1} cm^{-1}). The

overall labeling efficiency was about 100%.

Determination of R_0 and R values

The Förster distance, R_0, which is defined as the donor-acceptor distance when FRET efficiency is 50%, was calculated according to Equation (II-1):

$$R_0 = (8.79 \times 10^{-5} (\kappa^2 \Phi_D n^{-4} J))^{\frac{1}{6}} \quad \text{(II-1)}$$

where κ^2 is the orientation factor as a function of the relative orientation of the transition dipoles of the donor and the acceptor in space, which was assumed to rotate freely ($\kappa^2 = 2/3$), Φ_D is the quantum yield of the donor, which was determined by estimating the fluorescence intensity relative to *N*-acetyl-$_L$-tryptophanamide (NATA) assuming a quantum yield of 0.14 for NATA,[44] n is the refractive index of the medium, and J is the overlap integral of the emission of the donor and the absorption of the acceptor, which was calculated by the normalized fluorescence spectrum of the protein with only the donor and the absorbance spectrum of the acceptor according to Equation (II-2):

$$J = \frac{\int F_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda}{\int F_D(\lambda) d\lambda} \quad \text{(II-2)}$$

The R_0 values under the native and the acid-unfolding conditions (R_{0N} and R_{0U}) were calculated. The distance between the donor and the acceptor sites R was calculated from the apparent FRET efficiency, FE, and R_0 values by Equation (II-3):

$$FE = \frac{1}{1 + \left(R / R_0\right)^6} \quad \text{(II-3)}$$

FE was calculated using fluorescence intensity according to Equation (II-4):[25]

$$FE = 1 - \frac{F_{DA}}{F_D} \quad \text{(II-4)}$$

where F_D and F_{DA} are the fluorescence intensity in the absence and presence of the accepter, respectively.

Kinetic measurements using CF and SF fluorescence

The folding reaction was initiated by six-fold dilution of the acid-unfolded protein solution in ~20 mM phosphoric acid (pH 2.0) with 100 mM sodium acetate (pH 5.3) with varying concentrations of urea, giving a final pH of 5.2, while the unfolding reaction was initiated by six-fold dilution of the native protein solution in 100 mM sodium acetate (pH 5.3) with 100 mM sodium acetate at (pH 5.3) with varying concentrations of urea. The solutions for the measurements on the unmodified proteins contained 2 mM DTT for keeping the proteins reduced. The protein concentration was 10 μM for all the experiments. All the experiments were conducted at 15°C. The time-dependent change in

fluorescence was recorded with the combination of a 305-nm long pass filter and a 287/366-nm band pass filter with the excitation at 295 nm. The dead time of the CF device was 30 μs, which was estimated by measuring the quenching of *N*-acetyl-$_{L}$-tryptophan fluorescence by *N*-bromosuccinmide[45]. The SF measurements were conducted on an SX-20 (Applied Photophysics, UK) with a dead time of ~1.4 ms. The kinetic traces were fitted by nonlinear least-squares method to the following equation:

$$F(t) = F_{eq} + \sum_i A_i \exp(-\lambda_i t) \qquad \text{(II-5)}$$

where F (t) and F_{eq} are the observed values of the fluorescence at time t and infinite time at the equilibrium, respectively, while A_i and λ_i are the fluorescence change and the apparent rate constant, respectively, of the i-th kinetic phase.

The equilibrium unfolding transition curves were obtained by plotting the F_{eq} values versus urea concentrations. The baselines of the unfolded state and the native state were obtained by extrapolating the folding and the unfolding traces to zero time. The equilibrium unfolding transition curves of the modified and unmodified variants were fitted to a two-state model using IGOR Pro software package (Wavemetrics, Inc., OR).

REFFERENCES

1. Baldwin RL, Rose GD (1999) Is protein folding hierarchic? II. Folding intermediates and transition states. Trends Biochem Sci **24**:77-83.
2. Arai M, Kuwajima K (2000) Role of the molten globule state in protein folding. Adv Protein Chem **53**:209-282.
3. Roder H, Maki K, Cheng H (2006) Early events in protein folding explored by rapid mixing methods. Chem Rev **106**:1836-1861.
4. Dill KA, Ozkan SB, Shell MS, Weikl TR (2008) The protein folding problem. Annu Rev Biophys **37**:289-316.
5. Udgaonkar JB, Baldwin RL (1988) NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. Nature **335**:694-699.
6. Roder H, Wüthrich K (1986) Protein folding kinetics by combined use of rapid mixing techniques and NMR observation of individual amide protons. Proteins **1**:34-42.
7. Roder H, Elöve GA, Englander SW (1988) Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. Nature **335**:700-704.
8. Miranker A, Robinson CV, Radford SE, Aplin RT, Dobson CM (1993) Detection of transient protein folding populations by mass spectrometry. Science **262**:896-899.
9. Shastry MCR, Roder H (1998) Evidence for barrier-limited protein folding kinetics on the microsecond time scale. Nat Struct Biol **5**:385-392.

10. Hagen SJ, Eaton WA (2000) Two-state expansion and collapse of a polypeptide. J Mol Biol **297**:781-789.
11. Akiyama S, Takahashi S, Kimura T, Ishimori K, Morishima I, Nishikawa Y, Fujisawa T (2002) Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering. Proc Natl Acad Sci USA **99**:1329-1334.
12. Uzawa T, Akiyama S, Kimura T, Takahashi S, Ishimori K, Morishima I, Fujisawa T (2004) Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of α-helical content and compactness. Proc Natl Acad Sci USA **101**:1171-1176.
13. James E, Wu PG, Stites W, Brand L (1992) Compact denatured state of a staphylococcal nuclease mutant by guanidinium as determined by resonance energy transfer. Biochemistry **31**:10217-10225.
14. Wu P, Brand L (1994) Conformational flexibility in a staphylococcal nuclease mutant K45C from time-resolved resonance energy transfer measurements. Biochemistry **33**:10457-10462.
15. Lillo MP, Szpikowska BK, Mas MT, Sutin JD, Beechem JM (1997) Real-time measurement of multiple intramolecular distances during protein folding reactions: a multisite stopped-flow fluorescence energy-transfer study of yeast phosphoglycerate kinase. Biochemistry **36**:11273-11281.
16. Ratner V, Sinev M, Haas E (2000) Determination of intramolecular distance distribution during protein folding on the millisecond timescale. J Mol Biol **299**:1363-1371.
17. Navon A, Ittah V, Landsman P, Scheraga HA, Haas E (2001) Distributions of intramolecular distances in the reduced and denatured states of bovine pancreatic ribonuclease A. Folding initiation structures in the C-terminal portions of the reduced protein. Biochemistry **40**:105-118.
18. Magg C, Schmid FX (2004) Rapid collapse precedes the fast two-state folding of the cold shock protein. J Mol Biol **335**:1309-1323.
19. Sinha KK, Udgaonkar JB (2005) Dependence of the size of the initially collapsed form during the refolding of barstar on denaturant concentration: evidence for a continuous transition. J Mol Biol **353**:704-718.
20. Teilum K, Maki K, Kragelund BB, Poulsen FM, Roder H (2002) Early kinetic intermediate in the folding of acyl-CoA binding protein detected by fluorescence labeling and ultrarapid mixing. Proc Natl Acad Sci USA **99**:9807-9812.
21. Magg C, Kubelka J, Holtermann G, Haas E, Schmid FX (2006) Specificity of the initial collapse in the folding of the cold shock protein. J Mol Biol **360**:1067-1080.
22. Bilsel O, Matthews CR (2006) Molecular dimensions and their distributions in early folding intermediates. Curr Opin Struct Biol **16**:86-93.
23. Orevi T, Ben Ishay E, Pirchi M, Jacob MH, Amir D, Haas E (2009) Early closure of a long loop in the refolding of adenylate kinase: a possible key role of non-local interactions in the initial folding steps. J Mol Biol **385**:1230-1242.

24. Arai M, Iwakura M, Matthews CR, Bilsel O (2011) Microsecond subdomain folding in dihydrofolate reductase. J Mol Biol **410**:329-342.
25. Lakowicz JR (2006) Principles of fluorescence spectroscopy. Springer, New York.
26. Truckses DM, Somoza JR, Prehoda KE, Miller SC, Markley JL (1996) Coupling between trans/cis proline isomerization and protein stability in staphylococcal nuclease. Protein Sci **5**:1907-1916.
27. Schechter AN, Chen RF, Anfinsen CB (1970) Kinetics of folding of staphylococcal nuclease. Science **167**:886-887.
28. Sugawara T, Kuwajima K, Sugai S (1991) Folding of staphylococcal nuclease A studied by equilibrium and kinetic circular dichroism spectra. Biochemistry **30**:2698-2706.
29. Walkenhorst WF, Green SM, Roder H (1997) Kinetic evidence for folding and unfolding intermediates in Staphylococcal nuclease. Biochemistry **63**:5795-5805.
30. Jacobs MD, Fox RO (1994) Staphylococcal nuclease folding intermediate characterized by hydrogen exchange and NMR spectroscopy. Proc Natl Acad Sci USA **91**:449-453.
31. Ikura T, Tsurupa GP, Kuwajima K (1997) Kinetic folding and cis/trans prolyl isomerization of staphylococcal nuclease. A study by stopped-flow absorption, stopped-flow circular dichroism, and molecular dynamics simulations. Biochemistry **36**:6529-6538.
32. Walkenhorst WF, Edwards JA, Markley JL, Roder H (2002) Early formation of a beta hairpin during folding of staphylococcal nuclease H124L as detected by pulsed hydrogen exchange. Protein Sci **11**:82-91.
33. Maki K, Cheng H, Dolgikh DA, Shastry MC, Roder H (2004) Early Events During Folding of Wild-type Staphylococcal Nuclease and a Single-tryptophan Variant Studied by Ultrarapid Mixing. J Mol Biol **338**:383-400.
34. Maki K, Cheng H, Dolgikh DA, Roder H (2007) Folding kinetics of staphylococcal nuclease studied by tryptophan engineering and rapid mixing methods. J Mol Biol **368**:244-255.
35. Shastry MCR, Luck SD, Roder H (1998) A continuous-flow capillary mixing method to monitor reactions on the microsecond time scale. Biophys J **74**:2714-2721.
36. Hudson EN, Weber G (1973) Synthesis and characterization of two fluorescent sulfhydryl reagents. Biochemistry **12**:4154-4161.
37. Riddles PW, Blakeley RL, Zerner B (1979) Ellman's reagent: 5,5'-dithiobis(2-nitrobenzoic acid)--a reexamination. Anal Biochem **94**:75-81.
38. Maki K, Ikura T, Hayano T, Takahashi N, Kuwajima K (1999) Effects of proline mutations on the folding of staphylococcal nuclease. Biochemistry **38**:2213-2223.
39. Colón W, Elöve GA, Wakem LP, Sherman F, Roder H (1996) Side chain packing of the N- and C-terminal helices plays a critical role in the kinetics of cytochrome *c* folding. Biochemistry **35**:5538-5549.
40. Sanchez IE, Kiefhaber T (2003) Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. J Mol Biol **325**:367-376.

41. Kato S, Kamikubo H, Hirano S, Yamazaki Y, Kataoka M (2010) Nonlocal interactions are responsible for tertiary structure formation in staphylococcal nuclease. Biophys J **98**:678-686.
42. Baudet-Nessler S, Jullien M, Crosio MP, Janin J (1993) Crystal structure of a fluorescent derivative of RNase A. Biochemistry **32**:8457-8464.
43. Flanagan JM, Kataoka M, Shortle D, Engelman DM (1992) Truncated staphylococcal nuclease is compact but disordered. Proc Natl Acad Sci USA **89**:748-752.
44. Eftink MR, Jia Y, Hu D, Ghiron CA (1995) Fluorescence Studies with Tryptophan Analogues: Excited State Interactions Involving the Side Chain Amino Group. J Phys Chem **99**:5713-5723.
45. Peterman BF (1979) Measurement of the dead time of a fluorescence stopped-flow instrument. Anal Biochem **93**:442-444.
46. McCarney ER, Werner JH, Bernstein SL, Ruczinski I, Makarov DE, Goodwin PM, Plaxco KW (2005) Site-specific dimensions across a highly denatured protein; a single molecule study. J Mol Biol **352**:672-682.

TABLES

Table II-1. Förster distances between the pair of the donor and acceptor of the three variants of SNase under the acid-unfolding (~20 mM phosphoric acid at pH 2.0) and the native (100 mM sodium acetate at pH 5.3) conditions.

	Trp27/Cys97-AED	Trp27/Cys64-AED	Trp140/C64-TNB
R_{0U} (Å)	21.7	22.6	23.4
R_{0N} (Å)	21.1	21.5	29.8

Table II-2. Apparent rate constants (λ_i) and relative amplitudes (A_i) of the folding kinetics of SNase variants induced by a pH jump from 2.0 to 5.2 at 15°C obtained by multi-exponential least-squares fitting of the combined CF and SF traces.

	Trp27/Cys97 -AED	Trp27/Cys97	Trp27/Cys64 -AED	Trp27/Cys64	Trp140/Cys64 -TNB	Trp140/Cys64	Trp140[a]	Trp27[a]	Trp76[a]
F_{eq}	0.47645 ± 0.00008	0.8670 ± 0.0002	0.5654 ± 0.0001	0.9375 ± 0.0008	0.7119 ± 0.0002	5.1296 ± 0.0005	-	-	-
A_1	0.123 ± 0.002	-0.098 ± 0.001	0.0442 ± 0.0005	-0.0793 ± 0.0008	-0.0158 ± 0.0006	-0.072 ± 0.002	0.066 ± 0.011	0.051 ± 0.005	-0.30 ± 0.02
λ_1 (s^{-1})	**(1.62 ± 0.03) ×10^4**	**(8.5 ± 0.2) ×10^3**	**(4.86 ± 0.07) ×10^3**	**(4.7 ± 0.2) ×10^3**	**(7.3 ± 0.2) ×10^3**	**(1.59 ± 0.07) ×10^3**	**(1.27 ± 0.14) ×10^4**	**(5.50 ± 0.54) ×10^3**	**(1.35 ± 0.10) ×10^4**
A_2	-	-	-	-	-	-	-	-	-0.196 ± 0.005
λ_2 (s^{-1})	-	-	-	-	-	-	-	-	**3,870 ± 200**
A_3	0.0563 ± 0.0003	-0.0250 ± 0.0005	-	-0.0475 ± 0.0003	-	0.172 ± 0.002	0.69 ± 0.31	0.26 ± 0.11	0.28 ± 0.04
λ_3 (s^{-1})	**115 ± 2**	**136 ± 5**	**-**	**130 ± 2**	**-**	**150 ± 30**	**26 ± 4**	**11 ± 2**	**59 ± 7**
A_4	0.0600 ± 0.0002	0.3458 ± 0.0003	0.0772 ± 0.0002	0.3209 ± 0.0004	0.217 ± 0.003	-1.534 ± 0.002	-2.16 ± 0.25	0.27 ± 0.10	-1.69 ± 0.04
λ_4 (s^{-1})	**5.26 ± 0.04**	**5.70 ± 0.01**	**5.98 ± 0.03**	**4.15 ± 0.01**	**5.6 ± 0.1**	**6.6 ± 0.2**	**11 ± 1.2**	**4 ± 1.1**	**9.9 ± 0.5**
A_5	-	0.0504 ± 0.0003	0.0401 ± 0.0002	0.1002 ± 0.0006	0.033 ± 0.003	-1.468 ± 0.006	-0.50 ± 0.06	0.05 ± 0.01	-0.34 ± 0.05
λ_5 (s^{-1})	-	**0.283 ± 0.004**	**0.298 ± 0.003**	**0.338 ± 0.005**	**0.65 ± 0.06**	**0.349 ± 0.002**	**2.1 ± 0.4**	**0.35 ± 0.11**	**1.8 ± 0.3**
A_6	-	-	-	-	-	-0.598 ± 0.007	-0.29 ± 0.03	-	-0.10 ± 0.02
λ_6 (s^{-1})	-	-	-	-	-	**0.0801 ± 0.0007**	**0.28 ± 0.07**	-	**0.015 ± 0.006**

[a] Values from ref 34

Table II-3. Distances between the pair of donor and acceptor under various conditions.

	U at pH 2.0	U at pH 2.0 (Spectra)[d]	U at 0 M urea	$t = 0$ μs	$t = 300$ μs	N at pH 5.2 ($t = \infty$)[c]	N at pH 5.3 (spectra)[d]	coil	Crystal structure
FRET efficiency	FE_U		FE_{UN}[a]	FE(0 μs)[b]	FE(300 μs)	FE_N		-	-
Distance (Å)	R_U		R_{UN}	R(0 μs)	R(300 μs)	R_N		R_{coil}	$R_{crystal}$
Trp27/Cys97 -AED	0.160 ± 0.001	0.206	0.37 ± 0.02	0.388 ± 0.002	0.530 ± 0.001	0.469 ± 0.001	0.407	-	-
	28.6 ± 0.1	**27.2**	**23.6 ± 0.4**	**23.4 ± 0.1**	**21.3 ± 0.1**	**21.5 ± 0.1**	**22.5**	**62.6**	**19.0**
Trp27/Cys64 -AED	0.092 ± 0.001	0.258	0.38 ± 0.02	0.406 ± 0.002	0.464 ± 0.001	0.397 ± 0.001	0.403	-	-
	33.1 ± 0.1	**26.9**	**24.5 ± 0.3**	**24.1 ± 0.1**	**23.1 ± 0.1**	**23.1 ± 0.1**	**23.0**	**43.0**	**21.1**
Trp140/Cys64 -TNB	0.173 ± 0.001	0.190	0.38 ± 0.06	0.33 ± 0.02	0.346 ± 0.009	0.861 ± 0.001	0.813	-	-
	30.4 ± 0.1	**29.8**	**25 ± 1**	**26.3 ± 0.4**	**26.0 ± 0.2**	**22.0 ± 0.1**	**23.3**	**65.7**	**20.7**

The distances between the donor and the acceptor were calculated according to Equation (II-3). The Förster distances used were R_{0U} for R_U, R_{UN}, R(0 μs) and R(300 μs) and R_{0N} for R_N. R_{coil} was estimated according to McCarney *et al.*[46] $R_{crystal}$ was the distance between the C_β atoms of the donor (Tyr27 / Trp140) and acceptor (Lys64 / Lys97) sites in the crystal structure of Trp140 SNase.

[a] The energy transfer efficiency in U under the native condition (at 0 M urea) (FE_{UN}) was estimated by fluorescence intensities obtained by the baselines of the unfolded state of the dye-labeled and the unmodified proteins linearly extrapolated to 0 M urea.

[b] The energy transfer efficiency at 0 μs (FE(0 μs)) was estimated by fluorescence intensities extrapolated to zero time of the kinetic traces of the pH-induced folding from 2.0 to 5.2.

[c] The energy transfer efficiency at long time after the (un)folding reactions was estimated by fluorescence intensities after the pH-induced folding reaction from 2.0 to 5.2 was fully relaxed to the native state.

[d] The energy transfer efficiency and distance between donor and acceptor were also estimated based on emission spectra measurements at equilibrium.

Table II-4. Thermodynamic parameters describing the urea-induced equilibrium unfolding of SNase variants in ~100 mM sodium acetate at pH 5.2 and 15°C monitored by Trp fluorescence emission with the excitation wavelength at 295 nm.

	c_m (M)	m (kcal mol^{-1} M^{-1})	ΔG^{H2O} (kcal mol^{-1})
Trp140*[a]	3.45	2.30	7.92
Trp27	1.38	2.94	4.06
Trp27/Cys97	1.22	2.67	3.26
Trp27/Cys97-AED	1.27	2.60	3.30
Trp27/Cys64	1.17	2.58	3.08
Trp27/Cys64-AED	1.20	3.49	4.18
Trp140/Cys64	3.10	1.20	3.72
Trp140/Cys64-TNB	3.07	0.74	2.26

[a] Values from ref 33

FIGURES

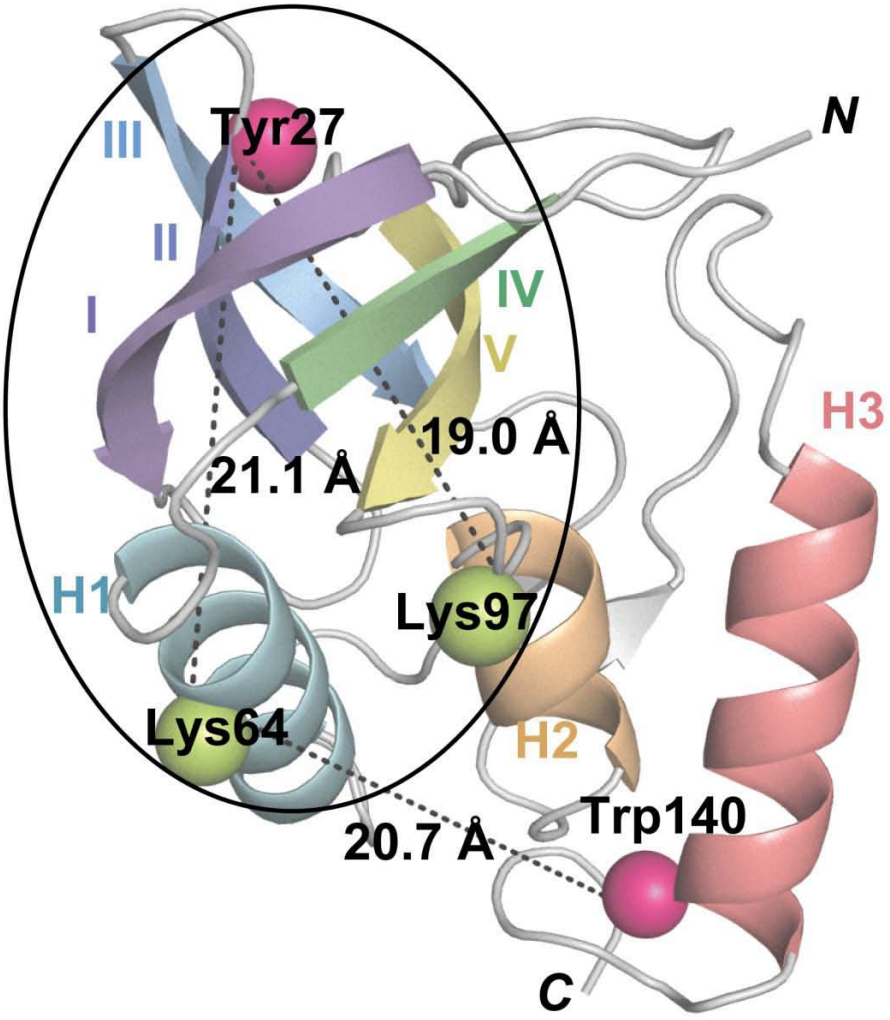


Figure II-1: Ribbon diagram of Trp140 SNase (P47G/P117G/H124L) based on the crystal structure (PDB 1SNQ)[26]. The C_β atoms of the mutation sites are shown as spheres; Tyr27 and Trp140 (red), Lys64 and Lys97 (green). The C_β-C_β distances of Tyr27-Lys64, Tyr27-Lys97 and Trp140-Lys64 are shown in dotted lines. The β-barrel domain is encircled.

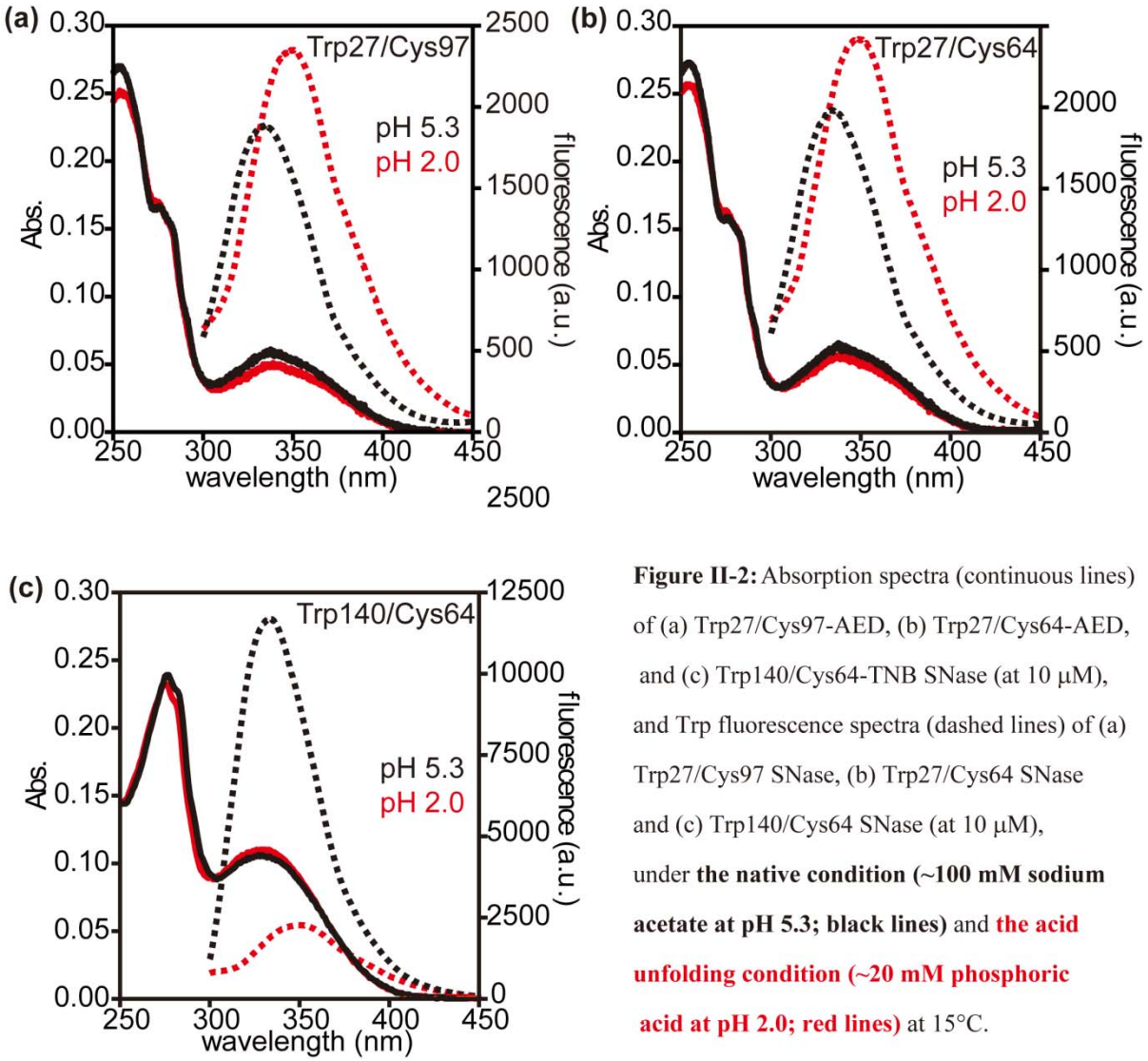


Figure II-2: Absorption spectra (continuous lines) of (a) Trp27/Cys97-AED, (b) Trp27/Cys64-AED, and (c) Trp140/Cys64-TNB SNase (at 10 μM), and Trp fluorescence spectra (dashed lines) of (a) Trp27/Cys97 SNase, (b) Trp27/Cys64 SNase and (c) Trp140/Cys64 SNase (at 10 μM), under **the native condition (~100 mM sodium acetate at pH 5.3; black lines)** and **the acid unfolding condition (~20 mM phosphoric acid at pH 2.0; red lines)** at 15°C.

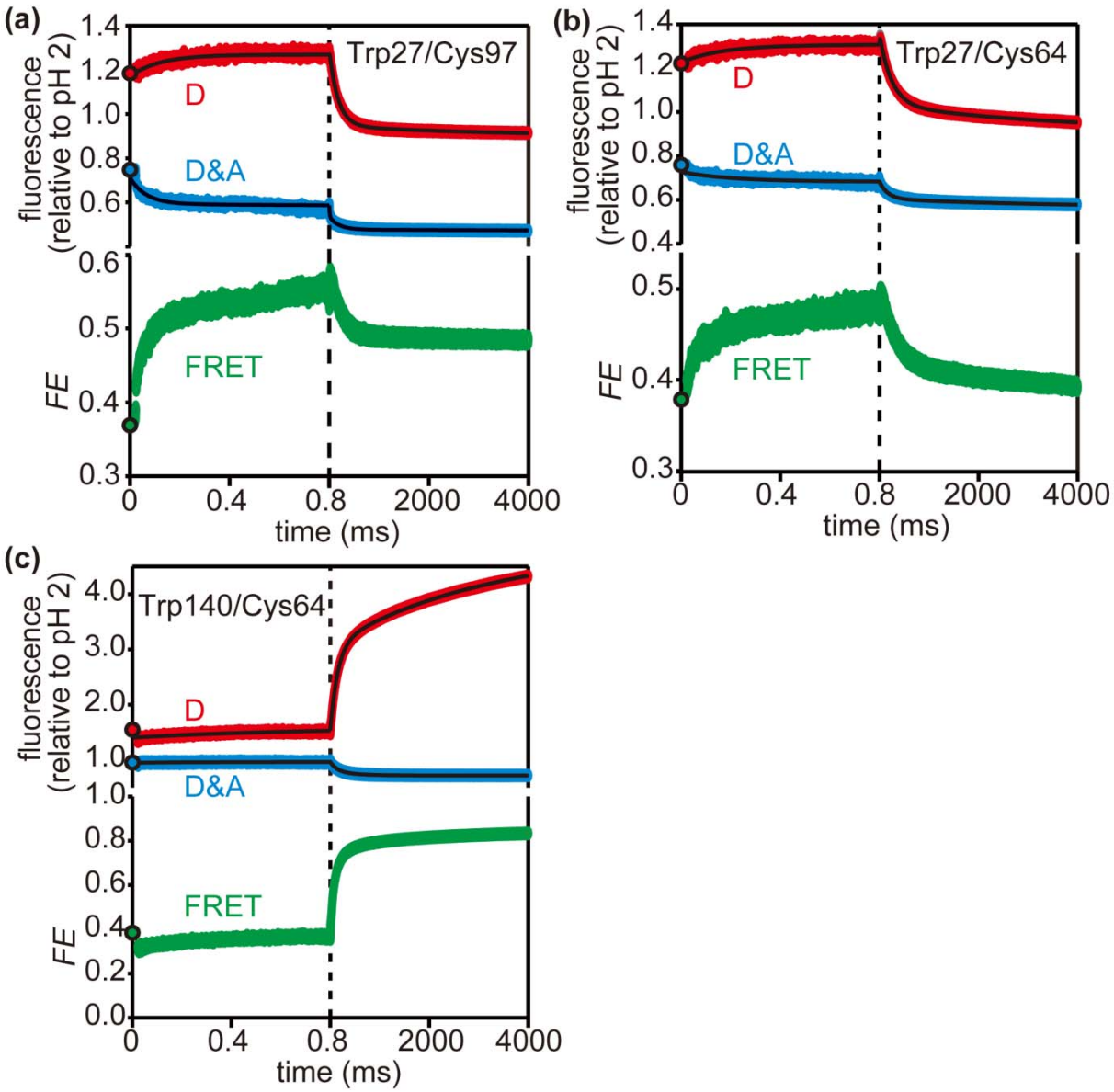


Figure II-3: Time-dependence of Trp fluorescence intensity and the energy transfer efficiency during the folding reactions of (a) Trp27/Cys97, (b) Trp27/Cys64, and (c) Trp140/Cys64 variants induced by a pH-jump from 2.0 to 5.2 (~100 mM sodium acetate at pH 5.2) at 15°C. The upper panels show the kinetic traces of the pH-induced folding reaction of **the unmodified (D; red)** and **the dye-labeled (D&A; blue)** proteins. The extrapolated fluorescence intensities of unfolded states at 0 M urea based on equilibrium measurements for each variant are shown in circles. Fluorescence was monitored with a combination of a 305-nm long-pass filter and a 278/366-nm band-pass filter. The lower panels show **the energy transfer efficiency (FRET; green)** during folding calculated by Equation (II-4). The apparent FRET efficiency calculated using fluorescence intensities of the unmodified and the dye-labeled proteins extrapolated to 0 M urea is in circle.

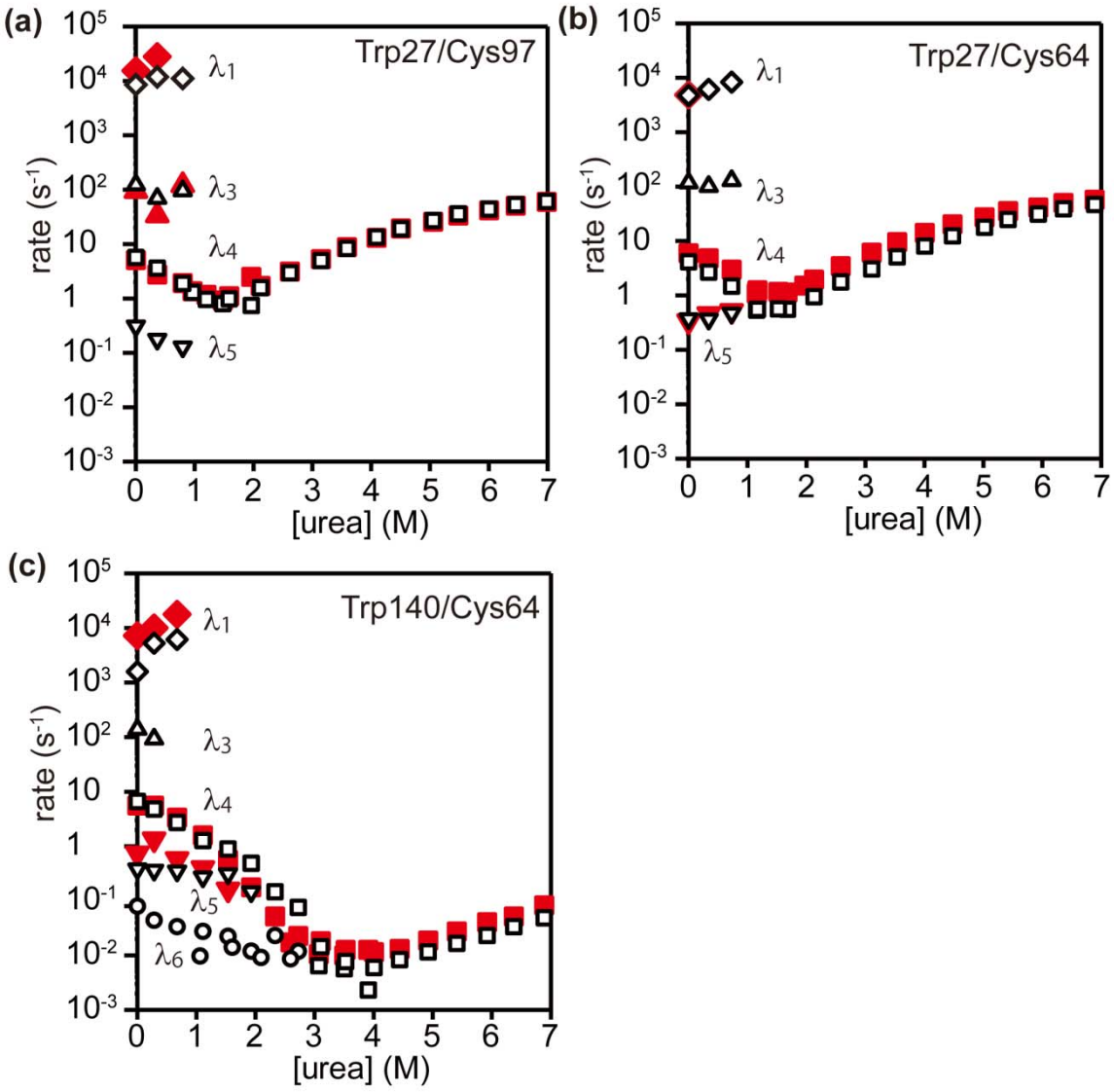
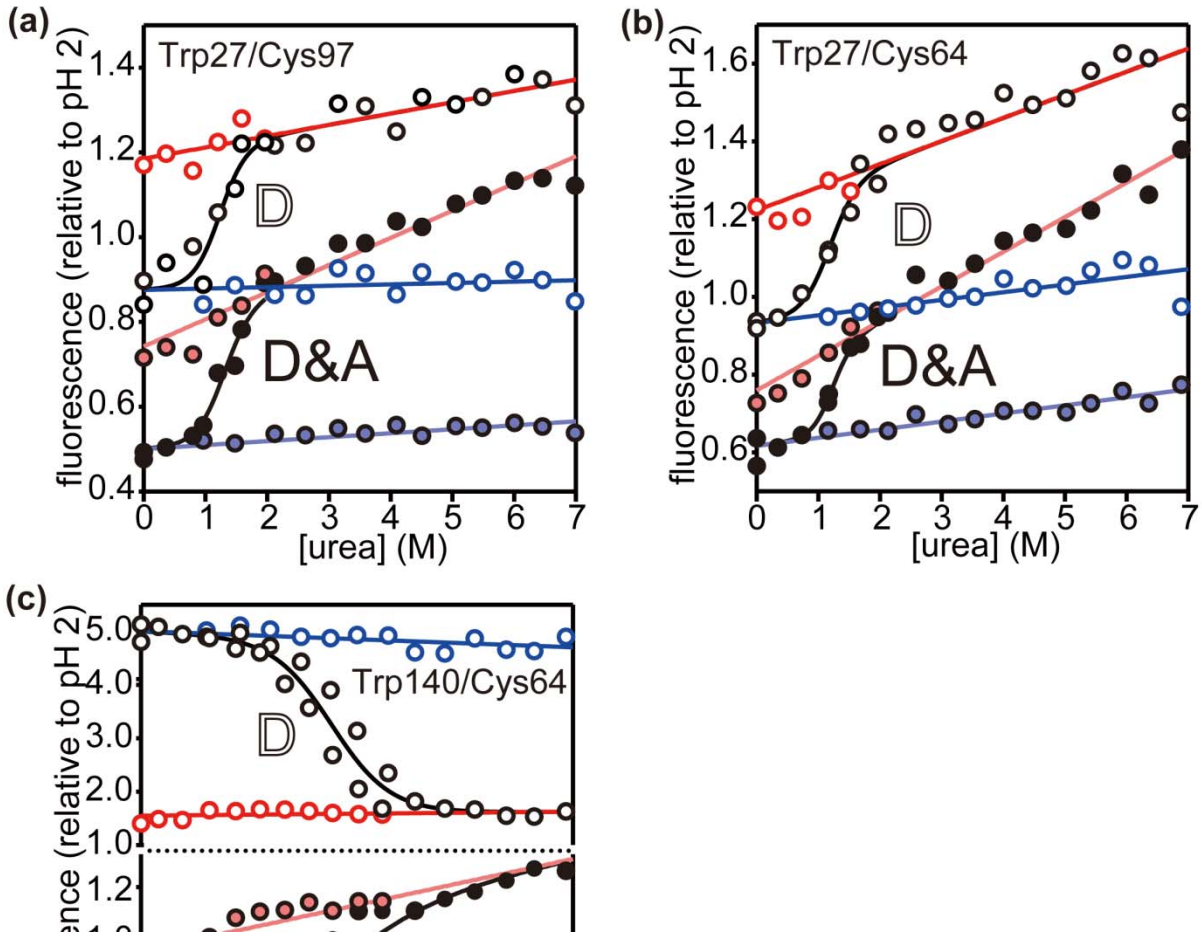


Figure II-4: Chevron plot for **the unmodified (open symbols: black)** and **the dye-labeled (filled symbols: red)** (a) Trp27/Cys97, (b) Trp27/Cys64 and (c) Trp140/Cys64 variants measured in ~100 mM sodium acetate at pH 5.2 and 15°C. The rate constants of the fastest phases (λ_1 (♦)) was obtained by the continuous-flow experiments while those of the other phases (λ_3 (▲), λ_4 (■), λ_5 (▼) and λ_6 (●)) were obtained by the stopped-flow experiments.


(a)
Trp27/Cys97
D
D&A
fluorescence (relative to pH 2)
[urea] (M)
(b)
Trp27/Cys64
D
D&A
fluorescence (relative to pH 2)
[urea] (M)
(c)
Trp140/Cys64
D

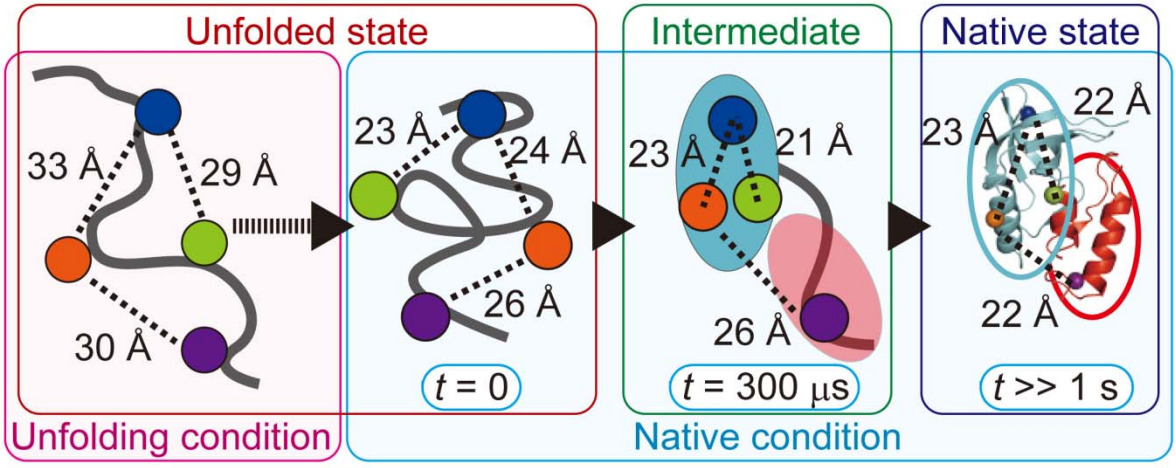


Figure II-6: The summary of results.

(i) The average distances for all donor/acceptor pairs in the unfolded state of the protein are shorter under folding conditions than those under strongly denaturing conditions, suggesting a more compact denatured state under the native condition.

(ii) The central region of the β-barrel domain assumes native-like compactness and tertiary interactions within 300 μs of folding.

(iii) The α-helical domain docks onto the β-barrel domain at longer times (>100 ms).

CHAPTER III

A Kinetics Study on the Early Structural Events during Folding/Unfolding of Horse Apomyoglobin Using Ultrarapid Mixing Methods

ABBREVIATIONS

apoMb, apomyoglobin;
CD, circular dichroism;
CF, continuous-flow;
f_H, helical content;
hapoMb, horse skeletal muscle apomyoglobin;
H/D, hydrogen/deuterium;
NATA, *N*-acetyl-$_L$-tryptophanamide;
NBS, *N*-bromosuccinimide;
p$K_{a\,COOH}^{i}$, pK_a for carboxyl groups in *i* state;
p$K_{a\,His}^{i}$, pK_a for histidine side chains in *i* state;
R_g, gyration radius;
SAXS, small-angle X-ray scattering;
SF, stopped-flow;
swapoMb, sperm whale apomyoglobin;
Trp, tryptophan;
UVRR, UV resonance Raman;
λ_{max}, emission maximum.

INTRODUCTION

Protein folding is the complex process during which protein molecules find out and form unique three-dimensional structures among a great number of possible conformers. Nevertheless, many proteins fold into their specific native structures on a second to minute time scale. Proteins with more than 100 amino acid residues often accumulate folding intermediates during the early stages of folding within a millisecond.[1] Many of the folding intermediates are partially folded with a compact size, native-like secondary structures and even specific tertiary interactions for some side chain pairs. Based on the native-like features of the folding intermediates, they have been regarded as being productive, that is, playing an important role in guiding the unstructured polypeptide chains toward the specific native structures.[1] Thus, elucidating the thermodynamics, kinetics, and structure of intermediates, especially focusing on the formation process, is essential to physicochemically understand the mechanism of protein folding. However, it has been extremely difficult to directly measure the formation of the folding intermediates because it occurs typically within a millisecond of folding, which is too fast to be directly detected by conventional experimental techniques such as stopped-flow (SF) methods. Thus, the early events during folding still remain one of the most important and challenging issues in the study of protein folding.[2,3] Recent development of experimental techniques such as continuous-flow (CF) methods and laser-induced temperature jump made it possible to directly observe the early stages of folding occurring within a millisecond and can be used to investigate the early folding events and the role of the kinetic intermediates.[4-7]

Under moderately denaturing equilibrium conditions, some proteins accumulate equilibrium intermediates that are structurally and thermodynamically similar to the folding intermediates.[1] The equilibrium intermediates have been investigated as a counterpart of the folding intermediates in order to overcome the difficulty in characterizing the short-lived folding intermediates in detail.[8-16] Previous studies on the equilibrium intermediates have actually provided us with detailed properties of the intermediates, for example, the topology (fold) of the intermediates[17] and the cooperativity of formation of the intermediates,[18,19] which would not be obtainable by kinetic experiments. However, the properties obtained for the equilibrium intermediates can be taken as those of the folding intermediates only if the equilibrium intermediates are equivalent to the folding intermediates. The equivalence is confirmed by directly comparing the kinetic processes associated with the formation of the folding and equilibrium intermediates in addition to the structural properties. Although the recent development of experimental techniques made it possible to extend the time resolution of kinetic measurements to the range of nanoseconds to microseconds, to date there have been only a few studies in which the equivalence has been confirmed by direct comparison of the formation of the intermediates.

Myoglobin (Figure III-1), is an α-helical heme protein with 153 amino acid residues, consisting of eight helices (A–H) formed in the globin fold.[20] Apomyoglobin (apoMb), the apo-form without the heme group, is one of the proteins whose folding mechanisms have been most extensively

investigated. This protein also forms a specific native structure close to myoglobin with the F-helix, the N-terminus of the G-helix and some loops unstructured under physiological conditions.[21] Previous studies on the folding of horse skeletal muscle apoMb (hapoMb) and sperm whale apoMb (swapoMb) have indicated that apoMb folds to the native state *via* a sequential folding pathway accumulating a series of folding intermediates.[22,23] At least two intermediates accumulate during the folding reaction of apoMb with a significant amount of helical structures in the A-, G- and H-helix regions and part of the B-helix in the native structure on the millisecond to second time scale of the folding reaction under native conditions.[11] The intermediate ensemble is compact with a hydrophobic core surrounded by the native-like secondary structures involving some specific interactions between side chains.[24] ApoMb has two tryptophan (Trp) residues, Trp7 and Trp14, located in A-helix, which are surrounded by the hydrophobic core consisting of the A-, G- and H-helix regions. Thus, when the folding kinetics is measured by monitoring the Trp fluorescence, a large enhancement accompanying the structure formation of the folding intermediates occurs within a few millisecond of folding. The structure formation of the fast folding intermediate was detected on submillisecond time scales using CF methods by a few groups.[23,25,26] However, the mechanism of the intermediate formation is not fully understood. In addition to the folding intermediates, apoMb accumulates equilibrium intermediates under mildly denaturing equilibrium conditions, which share many structural properties with the folding intermediate transiently accumulated during the folding.[10-12,27,28] The formation process of the equilibrium intermediate has been detected previously whereas the direct observation of the formation of the folding intermediate is scarce. The relationship between the folding and equilibrium intermediates is obtained by direct comparison of the formation process between the two types of the intermediates but it is not yet achieved due to the technical difficulty in the time resolution.

Here the author aimed to address the issues associated with the relationship between the folding and equilibrium intermediates by systematically investigating the folding/unfolding kinetics as well as the equilibrium unfolding of hapoMb as a function of urea concentration and pH. By fluorescence-detected CF and SF measurements, the author directly observed the formation of a series of intermediates in wide ranges of urea concentrations and pH. A quantitative modeling[29] of the folding/unfolding kinetics and equilibrium unfolding revealed that the folding of this protein is consistent with a five-state sequential scheme, where the three intermediates are placed on the folding pathway. The results indicated the equivalence between the transient folding/unfolding intermediates and the equilibrium intermediate. The apparent difference in the kinetic and equilibrium behavior of these intermediates is attributed to the relative stability of the intermediate to other states which varies with the conditions such as urea concentration and pH. A very fast unfolding phase is assigned to the partial unfolding of the native state to a native-like unfolding intermediate. The unfolding intermediate is located after the rate-limiting step along the reaction coordinate of the folding reactions, and thus corresponds to a "hidden intermediate" in folding reactions.

RESULTS

(A) Urea-induced folding/unfolding experiments

Circular dichroism and fluorescence spectra of hapoMb

Figures III-2a and III-2b show far- and near-UV circular dichroism (CD) spectra of hapoMb at 8°C, respectively. The far-UV CD spectrum at pH 6.0 (in 12 mM sodium citrate) shows a negative band with two local minima at 208 nm and 222 nm, reflecting the α-helical structure, while the near-UV CD spectrum shows an aromatic band under the same conditions, confirming the specific native structure.[30-32] The intensities in ellipticity in the far- and near-UV regions are significantly reduced at 0 M urea and pH 4.0 as well as at 1.2 M urea and pH 6.0 (in 12 mM sodium citrate), which is attributed to the accumulation of an equilibrium intermediate ensemble.[30] The far- and near-UV CD spectra at 0 M urea and pH 2.0 and at 8.0 M urea and pH 6.0 show characteristics typical of the unfolded state.[30]

Figure III-2c shows the fluorescence spectra of hapoMb at 8°C upon selective excitation of the Trp residues at 295 nm. The fluorescence spectrum at 0 M urea and pH 6.0 exhibits the emission maximum (λ_{max}) at 333 nm and a yield of 1.28 (relative to pH 2.0), reflecting the specific native structures around the side chains of two Trp (Trp7 and Trp14) that come into contact with quenching partners. The fluorescence spectra at 0 M urea and pH 4.0 and at 1.2 M urea and pH 6.0 exhibit a λ_{max} at 334 nm and 336 nm, and yields of 1.67 and 1.61, respectively. The enhanced fluorescence with the slightly red-shifted λ_{max} compared with the native state confirms the accumulation of the equilibrium intermediate ensemble, where the Trp side chains are largely shielded from solvent in hydrophobic environments.[22] The fluorescence spectra under the unfolding conditions at 0 M urea and pH 2.0 and at 7.8 M urea and pH 6.0 exhibit a λ_{max} at 348 nm and 351 nm, respectively, close to the λ_{max} of free Trp under these conditions (355 nm), indicating that the Trp side chains are exposed to solvent under these unfolding conditions.[33,34]

Figure III-2d shows temperature dependence of the ellipticity at 222 nm at pH 6.0 (in 12 mM sodium citrate). The ellipticity is decreased with lowering temperature until it begins to increase at 4°C, at which cold denaturation occurs.[27,35] The results in turn confirm that hapoMb fully remains in the native state at pH 6.0 and 8°C in the absence of urea. All the experiments in this study were carried out at 8°C.

Urea-induced equilibrium unfolding

Figures III-3a and III-3b show the urea-induced equilibrium unfolding transition of hapoMb measured by monitoring the ellipticity at 222 nm and Trp fluorescence emission spectra ranging from 300 nm to 450 nm (excitation at 295 nm), respectively. Since the urea concentration dependence of the fluorescence intensity at each emission wavelength can be regarded as an urea-induced equilibrium unfolding transition curve monitored by fluorescence at that wavelength[14], as illustrated in Figure III-3b (black symbols at 360 nm), the series of fluorescence emission spectra shown in

Figure III-3b represent a collection of unfolding transition curves monitored by fluorescence intensities at 149 wavelengths (300–450 nm). The unfolding transition curves at representative wavelengths thus obtained are shown in Figure III-3c. These unfolding transition curves monitored by fluorescence intensities and ellipticity at 222 nm were fitted by non-linear least squares fitting with a global fitting algorithm to an equilibrium three-state model consisting of the native (N_{eq}), intermediate (I_{eq}) and unfolded (U_{eq}) states:

$$N_{eq} \rightleftharpoons I_{eq} \rightleftharpoons U_{eq} \ , \qquad \text{Scheme III-1}$$

which yielded a set of global parameters (midpoint urea concentrations of $N_{eq} \rightleftharpoons I_{eq}$ and $I_{eq} \rightleftharpoons U_{eq}$ transitions, $c_{mNeqIeq}$ and $c_{mIeqUeq}$, and the corresponding *m*-values, m_{NeqIeq} and m_{IeqUeq}, for each transition) and local parameters (the fluorescence intensities and ellipticity of N_{eq}, I_{eq} and U_{eq} extrapolated to 0 M urea, and the slope of the baselines). The thermodynamic parameters thus obtained are listed in Table III-1, and are consistent with previous reports.[26,28,36-38] Based on these equilibrium parameters, fractions of the three species are calculated as a function of urea concentration (Figure III-4a). A tiny fraction of the molecules are populated as I_{eq} even in the absence of urea. As urea concentration increases, the fraction of I_{eq} increases at the expense of the fraction of N_{eq} until it reaches the maximum (54%) at 1.3 M urea, followed by a decrease with increasing in the fraction of U_{eq} (Figure III-4a).

A collection of fluorescence intensities over the range of wavelengths for each species, obtained by the curve fitting, represents the fluorescence spectrum linearly extrapolated to 0 M urea[14] as shown in Figure III-4b. The fluorescence spectrum and the ellipticity at 222 nm for N_{eq} are essentially identical to those at 0 M urea with only a slight difference due to a tiny fraction of I_{eq}. The fluorescence spectrum and the ellipticity at 222 nm for I_{eq} are similar to those observed at 1.2 M urea and pH 6.0 and at 0 M urea and pH 4.0. The differences in the ellipticity, the fluorescence intensity and the λ_{max} values between the fitting results and the results of the experiments at 1.2 M urea and pH 6.0 may be attributed to N_{eq} and U_{eq} populated under the experimental conditions. The fluorescence spectrum for U_{eq} exhibits no wavelength shift and reduced intensity compared with that observed at 7.8 M urea and pH 6.0. The ellipticity at 222 nm for U_{eq} is more intense than that observed at 8.0 M urea and pH 6.0. The difference in the ellipticity and fluorescence intensities extrapolated to 0 M urea and at 8.0 M/7.8 M urea is expected from the slope in the unfolding regions.

Folding kinetics initiated by a combined pH and urea concentration jump

The folding and unfolding kinetics of hapoMb were measured by monitoring fluorescence at various urea concentrations at pH 6.0 and 8°C. The initial conditions of the folding experiments were 0 M urea and pH 2.0, or 0.8 M urea and pH 6.0 whereas the initial condition of the unfolding experiments was 0 M urea and pH 6.0. CF and SF experiments were combined to monitor folding initiated at pH 2.0 whereas SF experiments were carried out to monitor folding initiated at 0.8 M urea and pH 6.0 as

well as unfolding. The dead times of the CF and SF measurements were 135 μs and 5.3 ms, respectively. Figure III-5a shows representative kinetic traces of the folding reaction. For the folding reaction initiated at acid-unfolded protein at pH 2.0, the kinetic traces of the folding reaction were fitted to a double (0–1.6 M urea) or a single-exponential function (1.8–2.7 M urea). When the folding reaction is initiated at 0.8 M urea and pH 6.0, the kinetic trace is fitted to a single-exponential function (0.13–0.53 M urea). The folding reaction initiated at 0.8 M urea and pH 6.0 is dominated by the conversion from I_{eq} to N_{eq} since I_{eq} is predominantly populated under the condition. The kinetic traces of unfolding reaction are fitted to a single-exponential function. Representative kinetic traces are shown in Figure III-5b.

Urea concentration dependence of the rate constants (chevron plot) of (un)folding reactions is summarized in Figure III-6a. Cumulative amplitudes (F_{01}, F_{02} and F_{0U}) are exploited instead of the amplitudes (F_i for the *i*-th phase) for the analysis of fluorescence intensity (Figure III-6b).[16] They are defined as $F_{02} = F_{eq} + F_2$ and $F_{01} = F_{eq} + F_2 + F_1$ for the folding reaction initiated at pH 2.0 and $F_{0U} = F_{eq} + F_1$ for the unfolding reaction, which give the fluorescence intensity at $t = 0$ in the absence of the preceding phases (F_{eq} represents the fluorescence intensity at equilibrium). The author first focuses on the folding reaction initiated at pH 2.0. The slower decreasing phase (phase 2) corresponds to the rate-limiting step during the folding,[23,39] whose rate constant (λ_2 ~ 2–5 s^{-1}) exhibits typical chevron behavior with the minimum at ~1.0 M urea. As the urea concentration increases, F_{02} approaches to F_{eq} to merge, i.e., phase 2 disappears, at ~1.6 M urea and coincides with F_{eq} ($F_{02} = 0$) at > 1.6 M urea. The chevron plot for λ_1 is shifted toward higher urea concentration (the minimum at ~2.0 M urea) with three-to-four orders of magnitude larger rate constant than that of λ_2. The curvature found at ~0.8 M urea indicates the accumulation of a folding intermediate within the dead time of the CF measurements. This is also indicated by the discrepancy between F_{01} (fluorescence intensity extrapolated to $t = 0$ of CF measurements) and the fluorescence expected for the unfolded state (burst phase) over a range from 0 M to 3.0 M urea. For the folding reaction initiated at 0.8 M urea and pH 6.0, the only decreasing phase (phase 2') exhibits the rate constant ($\lambda_{2'}$) close to λ_2 under the matching conditions. This is also the case for the rate constant of the unfolding reaction (λ_u), further indicating that phase 2 is the rate-limiting step of the overall folding/unfolding reactions. As urea concentration increases from ~2.0 M, λ_u increases with the curvature at ~3.3 M urea, which results from a switch in stability between the native and an unfolding intermediate, M, for unfolding (see below). F_{0U} exhibits a small but distinct transition at ~1.5 M urea, which is attributed to unfolding of I_{eq} populated at 0 M (Figure III-4a).

Quantitative modeling of the folding/unfolding kinetics

A folding scheme is modeled to quantitatively reproduce the rate constants (λ_1, λ_2, and λ_u) and amplitudes (F_{01}, F_{02}, and F_{0U}). As reported previously, the kinetic behavior of the (un)folding reactions is consistent with a sequential mechanism rather than a parallel-pathway mechanism.[10,34,39,40] For example, a folding scheme with phases 1 and 2 located on respective

parallel pathways would not account for enhancement of fluorescence during folding under strongly native conditions. The author finds that the minimal sequential scheme requires at least five states including the native (N) and the unfolded (U) states. At least one intermediate (I') is responsible for the burst phase of CF folding measurements and the curvature of λ_1 at ~0.8 M urea, a second intermediate (I) for phase 1, and a third intermediate (M) for the curvature of λ_2 (λ_u) at ~3.3 M urea.

The model assumes five thermodynamically distinct states interconverting according to the following mechanism:

$$\mathrm{U} \underset{k_{\mathrm{I'U}}}{\overset{k_{\mathrm{UI'}}}{\rightleftharpoons}} \mathrm{I'} \underset{k_{\mathrm{II'}}}{\overset{k_{\mathrm{I'I}}}{\rightleftharpoons}} \mathrm{I} \underset{k_{\mathrm{MI}}}{\overset{k_{\mathrm{IM}}}{\rightleftharpoons}} \mathrm{M} \underset{k_{\mathrm{NM}}}{\overset{k_{\mathrm{MN}}}{\rightleftharpoons}} \mathrm{N} \quad \text{Scheme III-2}$$

The elementary rate constants, k_{ij}, are defined by Equation (III-5). The elementary rate constants in the absence of urea at pH 6.0, ${k^{\mathrm{H2O}}}_{ij}$, and the urea concentration dependence, ${m^{\ddagger}}_{ij}$ were manually varied to reproduce the urea concentration dependence of the rate constants using the rate-matrix approach (see Appendix A). The predicted kinetic parameters are listed in Table III-2 and plotted in Figure III-6a. The λ_1 and λ_2 with curvature are approximated as $\lambda_1 \sim k_{\mathrm{UI'}}/(k_{\mathrm{UI'}} + k_{\mathrm{I'U}}) \times k_{\mathrm{I'I}}$ and as $\lambda_2 \sim k_{\mathrm{NM}}/(k_{\mathrm{NM}} + k_{\mathrm{MN}}) \times k_{\mathrm{MI}}$, respectively (see Appendix B). On the other hand, the unfolding limb of λ_1 and the folding limb of λ_2 are used as a guide for determining $k_{\mathrm{II'}}$ and k_{IM}, respectively. The model is further characterized by the fluorescence intensity of each state. Fluorescence intensities of U and N are determined according the corresponding baselines experimentally obtained. F_{01} is the fluorescence intensity of the burst phase while F_{02} approximates fluorescence intensity transiently saturated after phase 1, at which I is most populated during folding. Thus, F_{01} and F_{02} under stabilizing conditions are used as guides to determine the fluorescence intensity of I' and I, respectively. The fluorescence of M is assigned based on a result of pH-induced folding/unfolding experiments (see below). The fluorescence intensity of each state is listed in Table III-3 and plotted in Figure III-6b.

In order to test the validity of the model, the kinetic traces of (un)folding reactions at representative urea concentrations were reproduced by the kinetic parameters and the fluorescence intensities (Figure III-5). The kinetic traces are in good agreement with those experimentally obtained, especially in terms of the rate constants. Deviations are found on the submillisecond time scale of the kinetic traces of folding at 1.5 and 2.0 M urea, which arises from the fluctuation in F_{01} values (Figure III-6b). In addition, the thermodynamic parameters calculated by using the kinetic parameters are in good agreement with those obtained by the fitting of the equilibrium unfolding (Table III-1). This is illustrated as the reproduced fraction of the equilibrium states, N_{eq}, I_{eq} and U_{eq} in Figure III-4a. Because some of the species in Scheme III-2 were not resolved in the equilibrium unfolding, the author combined U and I' into U_{eq}, and M and N into N_{eq}, and thus I corresponds to I_{eq}, based on the fluorescence properties of each state. Taking these results together, the quantitative modeling fully accounts for the folding/unfolding kinetics and equilibrium unfolding experimentally obtained. A free energy diagram in this system is shown in Figures III-7.

(B) pH-Induced folding/unfolding experiments

Acid-induced equilibrium unfolding monitored by Trp fluorescence

Figure III-8 shows fluorescence emission spectra of hapoMb at pH ranging from 2.0 to 7.1 (in HCl at pH 2.0 or in 12 mM sodium citrate at pH 2.2–7.1) and 8°C. The two Trp residues were selectively excited at 295 nm. As was previously reported,[28] two transitions were observed at pH ~ 5.5 and ~3.0, confirming that at least three states are required to represent the pH-induced equilibrium unfolding of hapoMb; the native state (N_{eq}) is predominantly populated at pH > 5.7, the acid unfolded state (U_{eq}) at pH < 2.5, and an intermediate state (I_{eq}) at 3.5 < pH < 5.3. The fluorescence spectra at pH < 2.5 exhibit the emission maximum (λ_{max}) (348 nm) comparable to that of free Trp (355 nm), indicating that the two Trp side chains are fully exposed to solvent in U_{eq}. The spectra exhibit significantly blue-shifted λ_{max} (~334 nm) and enhanced intensities at pH 3.5–5.3, which is consistent with two Trp side chains within solvent-shielded hydrophobic core of the equilibrium intermediate I_{eq}. The λ_{max} is further blue-shifted (333 nm) and the intensity is reduced by specific contact of Trp side chains with specific quenching partners in N_{eq} at pH > 5.7.[41]

pH Dependence of folding and unfolding reaction

The folding and unfolding kinetics of hapoMb were systematically measured by monitoring fluorescence (excitation at 295 nm) as a function of pH at 8°C. The folding reaction was initiated at pH 2.0 (in HCl; U_{eq}) or pH 4.0 (in 2 mM sodium citrate; I_{eq}). The final pH value ranges from 2.2 to 6.2 for the folding reactions initiated at pH 2.0 and from 4.0 to 6.6 for those initiated at pH 4.0. The unfolding reaction was initiated at pH 4.0 (in 2 mM sodium citrate; I_{eq}) or at pH 6.0 (in 2 mM sodium citrate; N_{eq}). The final pH value ranges from 2.6 to 4.0 for the unfolding reactions initiated at pH 4.0 and from 2.8 to 6.7 for those initiated at pH 6.0. Combination of CF and SF methods covered a time window ranging from 40 μs to 100 s. The dead times of the CF and SF measurements were 40 μs and 3.5 ms, respectively.

The kinetic traces of folding and unfolding reactions thus obtained were fitted by non-linear least squares fitting. The fitting functions were a single-exponential function (pH 2.2–2.7), a two-exponential function (pH 3.0–4.8) and a three-exponential function (5.0–6.2) for folding reactions initiated at pH 2.0 whereas they were a single-exponential function (pH 4.0–4.8) and a two-exponential function (pH 5.0–6.6) for folding reactions initiated at pH 4.0 (Figure III-9a). On the other hand, the fitting functions were a three-exponential function (pH 2.6–3.4 and 3.8–4.5) and a two-exponential function (pH 3.6 and 4.6–6.7) for unfolding reactions initiated at pH 6.0 whereas they were a single-exponential function (pH 3.3–4.0) and two-exponential functions (pH 2.6–3.1) for unfolding reactions initiated at pH 4.0 (Figure III-9b). The pH dependence of the rate constants and amplitudes thus obtained are summarized in Figure III-10. Among the kinetic phases, a slow minor phase (phase D) with rate constant ~ 1 s^{-1} observed under almost all the conditions was assigned to dissociation of dimer or oligomer of hapoMb because this phase is detected when protein solution at pH 4.0 is diluted (Figures III-10a and III-10d). Thus, phase D is not considered further here to focus

on the folding/unfolding reactions.

(i) pH Dependence of folding reactions initiated at pH 2.0 and 4.0.

The author first focuses on the folding reactions initiated at pH 2.0 and 4.0 (Figures III-10a and III-10b). A decreasing phase (phase R2) with the rate constant around 3.6 s^{-1} was observed only under the native conditions above pH 5.0 for folding reactions initiated at pH 2.0. Phase R2 is the rate-limiting step of the folding reaction. The amplitude was reduced from 0.4 at pH 6.2 until the phase disappeared at pH 5.0. A faster increasing phase (phase R1) with the rate constant ranging from 1.5×10^4 s^{-1} at pH 5.9 to 620 s^{-1} at pH 3.2 was observed for folding reaction initiated at pH 2.0. The amplitude remained almost unchanged between pH 5.9 and 4.4 followed by a substantial reduction as pH decreased from pH 4.4 until the phase disappeared at pH 2.7. The only kinetic phase (excluding phase D) observed for folding initiated at pH 4.0 exhibits the pH dependences of the rate constant and the amplitude identical to those of phase R2, and thus is assigned to phase R2 irrespective of the initial conditions.

In addition to conventional amplitudes (F_i for the i-th phase), cumulative amplitudes (F_{0R2} and F_{0R1}) are exploited for the analysis of fluorescence intensity (Figures III-11a and III-11b) as described above. They are defined as $F_{0R2} = F_{eq} + F_{R2}$ and $F_{0R1} = F_{eq} + F_{R2} + F_{R1}$ for the folding reaction, which give rise to the fluorescence intensity at $t = 0$ in the absence of the preceding phases (F_{eq} represents the fluorescence intensity at equilibrium). Since phase R2 disappears at pH 5.0, F_{0R2} approaches to F_{eq} to merge toward pH 5.0. F_{0R1} (fluorescence intensity extrapolated to $t = 0$) exhibits the sigmoidal feature as a function of pH with deviation from the fluorescence expected for the unfolded state at pH > 3.5, indicating the accumulation of a folding intermediate within the dead time of CF measurements as the burst phase (phase R0).

(ii) pH Dependence of unfolding reaction initiated at pH 6.0 and 4.0

For the quantitative modeling of folding/unfolding reaction, the kinetic phases detected for unfolding reactions initiated at pH 6.0 (excluding phase D) are categorized into three groups. A minor fast increasing phase (phase U1) with a rate constant of 4,800–9,000 s^{-1} is detected only at pH 3.8–4.5 because the rate constant becomes too large to resolve as pH decreases and because the amplitude becomes too small as pH increases, as shown in Figures III-10a and III-10c. The second increasing phase (phase U2) exhibits strong pH dependence of the rate constant ranging from 11,000 s^{-1} at pH 2.6 to 3.4 s^{-1} at pH 5.7. The overlap of the rate constants of phase U2 and R2 above pH 5.0 under matching conditions indicates that phases R2 and U2 are the rate-limiting step of the overall folding and unfolding. The amplitude is also strongly dependent on pH; as pH increases from pH ~ 3.5, it is enhanced at the expense of the amplitude of phase U1 until it reaches the highest intensity at pH 5.1, followed by a reduction until it disappears at pH 6.3. The third decreasing phase (phase U3) with the rate constant ranging from 1,100 s^{-1} to 240 s^{-1} is observed from pH 2.6 to pH 3.4, whose amplitude decreases from -0.74 at pH 2.6 until it disappears at pH 3.6. The only kinetic phase (excluding phase

D) observed for unfolding initiated at pH 4.0 exhibits pH dependence of the rate constant and the amplitude identical to that of phase U3, which leads to assigning the phase to phase U3 irrespective of the initial conditions.

The cumulative amplitude for unfolding reactions are defined as $F_{0U3} = F_{eq} + F_{U3}$, $F_{0U2} = F_{eq} + F_{U3} + F_{U2}$, and $F_{0U1} = F_{eq} + F_{U3} + F_{U2} + F_{U1}$ (Figures III-11b and III-11c). The overlap between the equilibrium values obtained by the folding and unfolding kinetics confirms the reversibility of the folding reaction of hapoMb. Since phase U3 disappears above pH 3.6, F_{0U3} approaches to F_{eq} to merge toward pH 3.6 ($F_{U3} = 0$ at pH > 3.6). F_{0U2} deviates from the fluorescence intensity expected from the baseline of the native state at pH below 5.5. The deviation is fully accounted for by F_{U1} between pH 4 and 5 since F_{0U1} and the native baseline overlap each other. Phase U1 is detected as the burst phase, whose fluorescence intensity is represented by F_{0U2}, at pH below 4 since it becomes too fast to resolve.

Quantitative modeling of the pH dependence of the folding/unfolding kinetics

The author quantitatively models the folding scheme to reproduce the kinetic behavior of pH-induced folding and unfolding reactions of hapoMb. It is assumed that the pH-induced folding/unfolding is consistent with Scheme III-2 since the combined pH- and urea-induced folding in the absence of urea is identical to the folding induced by a pH jump from pH 2.0 to 6.0.

$$\mathrm{U} \underset{k_{\mathrm{I'U}}}{\overset{k_{\mathrm{UI'}}}{\rightleftharpoons}} \mathrm{I'} \underset{k_{\mathrm{II'}}}{\overset{k_{\mathrm{I'I}}}{\rightleftharpoons}} \mathrm{I} \underset{k_{\mathrm{MI}}}{\overset{k_{\mathrm{IM}}}{\rightleftharpoons}} \mathrm{M} \underset{k_{\mathrm{NM}}}{\overset{k_{\mathrm{MN}}}{\rightleftharpoons}} \mathrm{N} \qquad \text{Scheme III-2}$$

I' is responsible for the burst phase of the CF folding measurements at pH 3.5–5.9, I for phase R1 and U2, and M for phase U1 and the burst phase of the CF unfolding measurements at pH 2.5–3.8 (see below). The elementary rate constants, k_{ij}, are defined in Equation (III-6) (see Appendix C). The elementary rate constants at pH 6.0 (in the absence of urea), $k^{\mathrm{H2O}}{}_{ij}$, are taken from those obtained in the urea-induced folding and unfolding reactions under the matching condition (0 M urea and pH 6.0). The pK_a values of ionizable groups (pK_{aHis} for imidazole group of histidine side chains and pK_{aCOOH} for carboxyl groups), and the number of the ionizable groups associated with the transition were manually varied to reproduce the pH dependence of the rate constants using the rate-matrix approach (see Appendix A). The predicted kinetic parameters are listed in Tables III-2 and III-4, and plotted in Figure III-10a.

The rate-limiting step under the native condition (pH ~ 5–6) (phase R2) is assigned to the I ⇌ M interconversion, as described in the urea-induced (un)folding reactions. Phase R2 is connected to phase U2 through the approximation $\lambda_{U2} \sim k_{NM} / (k_{NM} + k_{MN}) \times k_{MI}$ (λ_{U2}; the rate constant of phase U2), which increases with decreasing pH to reach ~10^5 s^{-1} at pH 2.0. The rate constant of phase R1 (λ_{R1}) is approximated as $\lambda_{R1} \sim k_{UI'} / (k_{UI'} + k_{I'U}) \times k_{I'I}$, which corresponds to the rate constant of phase U3 ($\lambda_{U3} \sim 10^3$ s^{-1}) at pH 2–3. The rate-limiting step of the unfolding reaction is phase U3 at pH below ~3.5 since λ_{U2} is larger than λ_{U3}, allowing the accumulation of an unfolding intermediate (I)

during the unfolding reactions. Phase U1 is assigned to the M ⇌ N interconversion, which is observed in the unfolding reactions in a limited pH range from pH 3.8 to pH 4.5 because the rate constant increases with decreasing pH in the unfolding reactions. In the folding reactions, the M ⇌ N interconversion is undetectable because it is faster than the rate-limiting step of the folding reaction (phase R2).

The model is further characterized by the fluorescence intensity of each state. The fluorescence intensity at pH 6.0 in the absence of urea was taken from the value obtained by the urea-induced (un)folding experiments. The pH dependence of the fluorescence intensities of U and N is determined based on the corresponding baselines experimentally obtained (F_{eq}). F_{0R1} is the fluorescence intensity of the burst phase while F_{0R2} and F_{0U3} approximates the fluorescence intensity transiently saturated after phase R1 and U2, respectively (I is most populated during folding). Thus, F_{0R1} under stabilizing conditions are used as a guide to determine the fluorescence intensity of I', and F_{0R2} at pH above 4 and F_{0U3} at pH below 4 to determine the fluorescence intensity of I. The fluorescence intensity of M is experimentally determined by F_{0U1} through the accumulation of M as the burst phase in the unfolding reaction. The fluorescence intensity at pH 6.0 for each state is listed in Table III-3. The kinetic traces of (un)folding reactions (Figure III-9) were reproduced by using the kinetic parameters and the fluorescence intensities (Figure III-10) especially in terms of the rate constants, indicating the validity of the model. Figures III-12a and III-12b show a free energy diagram at representative pH values and the pH dependence of the free energy of each state, respectively.

Acid-induced unfolding equilibrium reproduced by the kinetic parameters

To further test the validity of the modeling, the fluorescence spectrum for each state is compared with that obtained in the urea-induced equilibrium unfolding. The fluorescence emission spectra at various pH values shown in Figures III-8 and III-13a are converted into a collection of unfolding transition curves monitored by fluorescence intensities at 149 wavelengths (301–449 nm) (Figures III- 13a (black symbols at 360 nm) and III-13b). The pH dependence of the fraction of the five states (U, I', I, M and N) is calculated by using the kinetic parameters obtained by the modeling (Figure III-13c). With the pH dependence of the fraction of each state as a known parameter, the equilibrium unfolding transition curves monitored by fluorescence emission at each wavelength ranging from 301 nm to 449 nm are fitted by linear least-squares fitting to Equation (III-2). The fluorescence intensity for each state is assumed to be linearly dependent on pH. Because some of the species in Scheme III-2 were not resolved in the equilibrium unfolding, U and I' are combined into U_{eq}, and M and N into N_{eq}, and thus I corresponds to I_{eq} for the analysis. The fluorescence emission of N_{eq}, I_{eq} and U_{eq} at 149 wavelengths (301–449 nm) thus obtained gives the fluorescence spectra of N_{eq}, I_{eq} and U_{eq} (Figure III-13d). The fluorescence spectrum of U_{eq} obtained from this analysis is more intense than that obtained from the urea-induced equilibrium unfolding, probably due to the long extrapolation of the unfolding baseline from the unfolding conditions to the native condition.

However, the fluorescence spectra of N_{eq} and I_{eq} are in good agreement with the corresponding fluorescence spectra obtained by the urea-induced unfolding equilibrium under the same condition, indicating that the model used here is consistent with the folding and unfolding of hapoMb.

DISCUSSION

Relationship between the folding intermediate I and the equilibrium intermediate I_{eq}

Previous H/D exchange NMR studies on swapoMb revealed that the two types of intermediates, the folding and the equilibrium intermediates, exhibit a similar pattern of protection factors despite the slight difference, i.e., both types of intermediates have a native-like secondary structure in the A-, G- and H-helix regions and part of the B-helix region of the native structure.[10-12,42,43] The time-resolved CD and small-angle X-ray scattering (SAXS) experiments revealed that the helix content and overall molecular size of these intermediates are also similar to each other for hapoMb.[23] In addition, folding kinetics on the native-state formation from the folding intermediate and from the pH 4 intermediate exhibited similar rate constants and amplitudes, indicating that these intermediates are converted into the native state by a common kinetic mechanism.[22]

Evidence of the equivalence between I and I_{eq} was found from a kinetic viewpoint by virtue of directly observing the formation of I over range of urea concentrations and pH using the CF method (Figures III-6a and III-10a). I transiently accumulates during folding at pH 6.0 and 0 M urea, which is a common final condition of the urea- and pH-induced folding reactions. The formation of I is continuously observed in both urea- and pH-induced folding as phase 1 up to 2.0 M urea (urea-induced folding) and as phase R1 down to pH 3 (in pH-induced folding). The observations indicate that I consists of a single species whether it is formed by a pH-induced folding or by a urea-induced folding (see below). Thus, I is not distinguished by denaturing factors to induce the folding reactions (pH and urea), and designated as I irrespective of these denaturing factors.

In the folding reaction, the formation of N is observed only at < 1.5 M urea and pH 5.0–7.0 due to relatively poor stability (ΔG^{H2O}_{NeqIeq} ~ 1.0 kcal/mol at pH 6.0 and 0 M urea). In contrast, the formation of I is observed at < 2.0 M urea and at pH > 3.0 as long as λ_1 (for urea-induced folding kinetics) and λ_{R1} (for pH-induced folding kinetics) are detected; in other words, F_{02} and F_{0R2} are larger than the baseline of U. It follows that I accumulates as an equilibrium intermediate, I_{eq}, at 1.5–2.0 M urea and pH 3.0–5.0, in urea concentration and pH ranges at which N is no longer populated at equilibrium.[28] Therefore, the intermediates I and I_{eq} consist of a single molecular species, and the stability of N relative to I determines whether I transiently accumulates during folding under strongly native conditions (~0 M urea and pH ~6.0–7.0) or is stably accumulated as I_{eq} at equilibrium under moderately denaturing conditions (1.5–2.0 M urea or pH 3.2–5.5). A few lines of evidence further support this idea. The finding that the urea and pH dependences of the apparent rate constant of the formation of N are independent of the initial conditions (λ_2 and $\lambda_{2'}$ in urea-induced folding are identical to each other and the folding kinetics initiated at pH 2.0 and 4.0 are identical to each other)

indicates that I and I_{eq} are converted into N by the same kinetics.[22] The quantitative modeling revealed that the thermodynamic stabilities of I and I_{eq} are very close to each other (Table III-1). The fluorescence intensity for I expected from the (un)folding kinetics (1.90) is close to that of I_{eq} estimated by the equilibrium unfolding (1.72–1.83) (Table III-3). This is partially demonstrated by the fact that F_{02} and F_{0R2} share the transition regions with F_{eq} at ~1.6 M urea and at pH < 5.0, respectively, because the urea concentration dependence of F_{02} and the pH dependence of F_{0R2} represent the preequilibrium unfolding of I by urea and pH, respectively. These observations are consistent with the differences in urea concentrations, where the apparent rate constants go through a minimum (Figure III-6a); at 0.6–2.0 M urea, folding from I' as well as unfolding from M is preferred because, for example, λ_1 and λ_2 (λ_u) as functions of urea are found on the folding and unfolding limbs, respectively, resulting in accumulation of I in comparison with the other states at equilibrium. Essentially similar kinetic behavior is observed in pH-induced folding. At pH 3.2–5.5, the slowest phase prefers the unfolding reaction (phase U2) whereas the fast phase R1 drives the formation of I, which results, again, in the accumulation of I at equilibrium. The energetics of the (un)folding reaction is schematically summarized by a free energy diagram, which shows that I is one of the most stable species at 2.0 M urea or at pH 4.0–5.0, leading to the accumulation of I as an equilibrium intermediate (Figures III-7 and III-12a). The equivalence between pH- and urea-induced equilibrium intermediates was previously reported.[28] Combined pH- and urea-induced equilibrium unfolding (urea-induced unfolding at pH ranging from 1.5 to 8.6 and pH-induced unfolding at 0–7.6 M urea) was consistent with a three-state model; a single equilibrium intermediate accounted for the accumulation of a pH- and urea-induced intermediate. Considering these facts together, the author concludes that I and I_{eq} are equivalent to each other irrespective of the denaturing factors (urea and pH), with respect to their kinetic mechanisms as well.

At pH below 3, the conversion into U (unfolding) is dominantly observed in the folding/unfolding reactions, in contrast to the formation of I observed at pH above ~4. During the unfolding reaction, an unfolding intermediate transiently accumulates in phase U2, which is found to be equivalent to I with the assignment of phase U2 to the M ⇌ I interconversion by quantitative modeling.[34] Because phase U2 (or R2 in folding reaction) is the rate-limiting step of the overall (un)folding reactions at pH above ~4, the transient accumulation of I during unfolding indicates the pH-dependent switch of the rate-limiting step of the reaction, which is previously reported in the folding/unfolding of swapoMb (N ⇌ Ib at mildly acidic pH and Ia ⇌ U at strongly acidic pH such as pH 2.7).[34] This was found by the observation that the unfolding kinetics is monophasic at mildly acidic pH and biphasic at strongly acidic pH. The quantitative modeling provides a mechanistic account for the observations. The combination of the differences in pK_{aHis} and pK_{aCOOH} between M and the transition state between I and M (TS3) brings about the strong pH-dependent decrease in k_{MI} from ~10^5 s^{-1} at pH~2 to ~2 s^{-1} at pH~6. In contrast, virtually no difference in pK_a values associated with the I → the transition state between I' and I (TS2) transition gives rise to the elementary rate constant ($k_{II'}$) almost independent of pH. At pH above 4, the M → I transition, which is slower than

the I → I' transition, is the rate-limiting step. At pH below ~3.5, in contrast, the M → I transition is faster than the I → I' transition, resulting in the accumulation of I in the unfolding. Therefore, I accumulates not only as the folding intermediate under native conditions, the equilibrium intermediate under moderately denaturing conditions but also as an unfolding intermediate under acidic conditions.

Previous studies on apoMb have revealed some differences in the properties of the folding and equilibrium intermediates, although there are many more similar properties.[26,37,44] For example, previous H/D exchange studies on swapoMb found that the folding intermediate at pH ~ 6.0 is more stable than the equilibrium intermediate at pH ~ 4.0 (with the overall patterns of labeling and secondary structure formation being similar to each other).[26] These differences would be accounted for by the pH-dependent stability of the folding intermediate, according to the equivalence between the folding and equilibrium intermediates for pH-induced (un)folding. The pH-dependent change in the stability would affect the protection of amide protons as is the case for the denaturant dependence of the stability estimated by native-state H/D exchange experiments.[45] A previous study has shown that apoleghemoglobin exhibits different patterns of labeling between the folding and equilibrium intermediates in contrast to the intermediate of swapoMb.[46] The difference in the labeling patterns would arise either from the pH dependence of a single pH-sensitive intermediate or from two distinct intermediates preferentially populated at each pH. Kinetic experiments on the (un)folding of the intermediate over a range of pH would address the question by estimating the number of relevant intermediate(s) and its stability as a function of pH.

Early folding intermediate, I'

A folding intermediate, I', is indicated to accumulate during the folding reaction within the dead time of CF measurements based on the curvature in λ_1 as a function of urea as well as by the missing amplitude of the folding experiments at low urea concentrations and at pH > 3.0. A very rapid structure formation has been reported for apoMb which occurs within a 100-μs time scale of the folding reaction initiated at an acid-unfolded state at low pH. Previous studies on pH-induced folding under native conditions revealed the accumulation of a burst phase intermediate in this time range by means of infrared spectroscopy (at pD 6.2) and UV resonance Raman (UVRR) scattering (pH 5.6), combined with the CF measurements.[25,39] A recent study on the structure formation and unfolding of the pH 4 intermediate ensemble of swapoMb revealed two faster phases prior to the formation of the pH 4 intermediate ensemble by the CF fluorescence method, which is consistent with the accumulation of I' as a burst phase intermediate at pH 6.0.[26]

The fluorescence intensity for I' is estimated to be lower than that of U (0.80 *vs*. 1.25) based on the quantitative modeling. The low fluorescence intensity appears to be unexpected because the fluorescence is expected to be fully quenched by solvent due to the exposure of the Trp residues to solvent even in U. It should be noted here that the fluorescence intensity of U at pH 6.0 and 0 M urea, which is obtained by linearly extrapolating the baseline of the unfolded state down to 0 M urea at pH

6.0, is expected to be larger than that at pH 2.0 and 0 M urea (1.25 *vs*. 1.00). This discrepancy in the fluorescence intensities of the unfolded states under acid-unfolding and native conditions has been observed for other proteins.[47] The apparently unusual quenching of the fluorescence intensity of I' could be explained by a blue-shifted fluorescence spectrum along with the low-level enhancement in the intensity in comparison with U. The modeling indicates that the α-value of I' is closer to that of I than to that of U (Figure III-7). The pK_{aCOOH} values in I' (p$K_a^{I'}{}_{COOH}$) and in I (p$K_a^{I}{}_{COOH}$) states for five ionizable carboxyl groups are closely similar to each other. In addition, the significant contraction in size (~6 Å) occurs within ~300 μs of the folding reaction.[23] These facts suggest that the overall structure of I' is already rather compact and that at least part of the core consisting of the A-, G- and H-helix regions are formed in I'. Thus, the λ_{max} of I' would be close to that of the subsequently accumulated I_{eq} (equivalent to I) (~335 nm), and significantly blue-shifted compared with the λ_{max} of U_{eq} (351 nm). On the other hand, the infrared spectroscopy and UVRR scattering studies have indicated that some amounts of solvated helix is present as well as buried helix, but the side chains of the Trp residues are partially solvent-shielded on the 100-μs time scale.[25,39] These considerations suggest that I' would exhibit a partially quenched fluorescence spectrum with a blue-shifted λ_{max}, which could have resulted in the apparently reduced fluorescence intensity after passing the optical filter in the experiments due to the absorption.

Early unfolding intermediate, M

A native-like high energy unfolding intermediate, M, is required to account for the curvature of the unfolding limb observed in the unfolding kinetics of some proteins. It is regarded as a "hidden intermediate" which accumulates neither as a transient unfolding intermediate nor as an equilibrium intermediate. This is not only because the intermediate is too unstable to be populated even transiently but also because the activation free energies associated with this intermediate is lower than those of the preceding kinetic steps. The pH-induced unfolding of hapoMb exhibits very fast phase (phase U1 at pH above 3.8 and the burst phase at pH below 3.8) prior to accumulation of an unfolding intermediate I, indicating the accumulation of an additional unfolding intermediate, which is designated as M. Quantitative modeling of the combined pH- and urea-induced (un)folding revealed that M accumulates in the urea-induced unfolding as the burst phase (Figure III-6b). The M state of hapoMb is different from that of other proteins in that M is not a high energy intermediate for hapoMb because it is stable enough to be transiently populated.[48,49] M is only 1.6 kcal/mol less stable than N at pH 6 and becomes more stable than N above 2.8 M urea. Furthermore, it accumulates even as an equilibrium intermediate as a minor species (Figure III-13c).

The M state of hapoMb is, however, indicated to exhibit native-like structural properties as were reported for the M state of other proteins. The fluorescence intensity of M is closer to that of N than to that of I. Since the two Trp residues as a source of the fluorescence are located in A-helix as described above, the fluorescence for M close to N suggests that the Trp residues are located in the native-like hydrophobic core consisting of the A-, G- and H-helix regions. In addition, M exhibits the

same abnormal pK_{aHis} value (p$K_a^{M}{}_{His}$) as that of N (p$K_a^{N}{}_{His}$) for an imidazole group. Previous studies indicated that His24 (in B-helix) have an abnormal p$K_a^{N}{}_{His}$ value.[50] In the native state, the side chains of His24 and His119 (in G-helix) are stacked to each other buried inside the hydrophobic core consisting of A-, G-, and H-helix regions and part of B-helix region to play an important role in stabilizing the native state. Thus, the imidazole group of His24 is the most probable histidine residue with an abnormal p$K_a^{M}{}_{His}$. By Taking the fluorescence and protonation properties into consideration, it is concluded that M exhibits native-like structure properties. Recently, ^{15}N, $^{1}H^{N}$, and ^{13}CO NMR R_2 dispersion relaxation experiments revealed a native-like intermediate formed by local unfolding of F-helix and undocking from the protein core.[51] The intermediate also exhibits the abnormal pK_{aHis} value and is an on-pathway intermediate. The intermediate would correspond to M taking the structural similarity into consideration.

The native and unfolded states

The fluorescence for N is quenched in comparison with that for I while the λ_{max} is blue-shifted compared with that of U (333 nm *vs.* 348 nm), indicating that the Trp residues are fully buried inside the molecule (Figure III-4b).[34] The mean residue ellipticity at 222 nm for N is -21,000 deg cm^2 dmol^{-1} (Figures III-2a and III-3a), and thus the helix content (f_H) for N is 62%, which is consistent with the f_H estimated by the crystal structure of myoglobin taking into consideration that the F-helix region and N-terminal region of G-helix are unstructured (61%).[21] Two imidazole groups exhibit the abnormal p$K_a^{N}{}_{His}$ value (Table III-4). Based on the previous studies, one of the histidine residues is His24.[50] His24 contacts with His119 through a specific hydrogen bond and is buried inside the molecule in N. Thus, the exposure of the side chain to solvent would be required for the protonation. Another possible mechanism for the protonation is the proton transfer from a neighboring residue such as Lys118 (or Arg118 for swapoMb). His64 and His113 also exhibit pK_{aHis} values lower than the normal value in N and thus are candidates for the other histidine residue with the abnormal p$K_a^{N}{}_{His}$. Although the mechanisms of their low pK_a values are not fully understood, the low p$K_a^{N}{}_{His}$ of His113 may result from a specific contact with Arg31.

The fluorescence spectrum for U exhibits the λ_{max} (348-351 nm) close to that of free Trp (355 nm)[33], indicating that the Trp residues are fully exposed to solvent in U, consistent with the previous reports. In addition, the far-UV CD spectrum at pH 2.0 exhibits structural features of unstructured polypeptide chains. However, the pK_{aCOOH} value in unfolded state (p$K_a^{U}{}_{COOH}$ = 3.6) is slightly lower than those of free aspartate and glutamate (4.0) while the pK_{aHis} value in unfolded state (p$K_a^{U}{}_{His}$ = 6.0) is very close to that of free histidine (6.0). The discrepancy between the pK_{aCOOH} values in U and free amino acids may arise from residual structure in U. In fact, acid-unfolded barnase exhibits pK_a values lower than that of the free amino acids.[52]

Folding mechanism of hapoMb

In the present study, a series of folding and unfolding kinetics and unfolding equilibrium were

systematically measured as a function of urea concentration and pH with CF and SF devices. It was found that Scheme III-2 fully accounts for the results of the folding/unfolding kinetics and unfolding equilibrium of hapoMb obtained in this study. In Scheme III-2, the three intermediates are assumed to be on-pathway intermediates, i.e., productive intermediates. Here the author considers the possibility that an intermediate is a dead-end product under the assumption of a five-state sequential scheme. In this case, the intermediate is not productive and should be at least partially unfolded to reach the native state. The author first asks whether I' could be an off-pathway intermediate by examining Scheme III-3.

$$\mathrm{I'} \underset{k_{\mathrm{UI'}}}{\overset{k_{\mathrm{I'U}}}{\rightleftharpoons}} \mathrm{U} \underset{k_{\mathrm{IU}}}{\overset{k_{\mathrm{UI}}}{\rightleftharpoons}} \mathrm{I} \underset{k_{\mathrm{MI'}}}{\overset{k_{\mathrm{I'M}}}{\rightleftharpoons}} \mathrm{M} \underset{k_{\mathrm{NM}}}{\overset{k_{\mathrm{MN}}}{\rightleftharpoons}} \mathrm{N} \qquad \text{Scheme III-3}$$

The U ⇌ I' interconversion is so fast that the formation of I' could not be observed as a kinetic phase in the folding experiments. Thus, the elementary rate constants for the U ⇌ I' interconversion are not uniquely determined, but only the preequilibrium constant can be estimated. Actually, the results on the kinetics as well as the equilibrium are reproduced by the quantitative modeling based on Scheme III-3. Another possibility is also considered that I is a dead-end product as shown in Scheme III-4.

$$\mathrm{I} \underset{k_{\mathrm{UI}}}{\overset{k_{\mathrm{IU}}}{\rightleftharpoons}} \mathrm{U} \underset{k_{\mathrm{I'U}}}{\overset{k_{\mathrm{UI'}}}{\rightleftharpoons}} \mathrm{I'} \underset{k_{\mathrm{MI'}}}{\overset{k_{\mathrm{I'M}}}{\rightleftharpoons}} \mathrm{M} \underset{k_{\mathrm{NM}}}{\overset{k_{\mathrm{MN}}}{\rightleftharpoons}} \mathrm{N} \qquad \text{Scheme III-4}$$

The behavior predicted by Scheme III-4 is essentially indistinguishable from that of Scheme III-2 in the folding reactions. This is because the preequilibrium among U, I' and I is established very rapidly (within ~ms), and then the conversion to M becomes significant on a time scale of seconds. The difference between these schemes could be kinetically distinguished if the U ⇌ I' interconversion is not only experimentally observed but also strongly coupled with the I' ⇌ I interconversion, as reported for the folding of Immunity protein 7.[53] Nevertheless, the α-values of I' and I are expected to be ~70% and ~50%, respectively, according to Schemes III-3 and III-4, which indicates that I' or I should be globally unfolded prior to the conversion into N on a submillisecond time scale. This is very unlikely when considering that the native-like substructure is already formed even on the 400-μs time scale of folding.[10] Thus the author proposes that Scheme III-2 represents the folding of hapoMb, but further improvements in time resolution will provide more insight into the folding of this protein.

The author observed the formation of a single intermediate (I) on submillisecond–millisecond time scale during folding and unfolding, whereas previous kinetic studies on apoMb often reported that two folding intermediates (designated such as I_1/I_2 or Ia/Ib) accumulated.[22,23,25,34] These two intermediates exhibited secondary structures and overall sizes similar to each other according to the previous CF-combined CD, SAXS and infrared absorption-detected CF experiments.[23,39] In addition, the fluorescence intensities were also similar to each other for the intermediates of swapoMb (Ia and Ib), although the UVRR CF measurements on hapoMb indicated distinct local environments around the two Trp side chains between the folding intermediates.[25] Thus, even if the folding intermediate ensemble (I) observed in this study consisted of multiple forms, such as Ia and Ib, the fluorescence

intensities of the folding intermediates might be very similar to each other for hapoMb, which would give rise to too small an amplitude corresponding to the interconversion to detect by Trp fluorescence under the conditions employed in this study.

CONCLUSIONS

The author systematically investigated the urea concentration and pH dependences of folding/unfolding kinetics and unfolding equilibrium so as to quantitatively model the results. The author found that the folding of this protein is fully consistent with a five-state sequential folding scheme (Scheme III-2) containing three intermediates, I', I and M. The formation of I is observed over wide ranges of urea concentrations (from 0 M to 2.74 M urea) and pH (from 3.0 to 6.2) than the formation of N (from 0 M to 1.6 M urea and from pH 5.0 to pH 6.6). Based on the difference in the urea concentration and pH dependences of the stability between I and N, the author found that I_{eq} is equivalent to I and that the stability of N relative to I determines whether I is only transiently accumulated during folding or is stably accumulated as I_{eq}. I also accumulates as an unfolding intermediate during pH-induced unfolding reaction as a result of pH-dependent switch of the rate-limiting step at pH ~ 2.7. I', which exhibits fluorescence intensity slightly less than that of U, accumulates within the dead time of the CF device (40 μs) as well as at equilibrium under moderately denaturing conditions as a minor species. The unfolding intermediate, M, is transiently accumulated on the submillisecond time scale of unfolding below pH 5.0 and is expected to also be accumulated during unfolding in the presence of urea on a submillisecond time scale.

MATERIALS AND METHODS

Chemicals

A specially prepared reagent-grade urea for biochemical use was purchased from Nakalai Tesque Inc. (Kyoto, Japan). The urea solutions were used within a few days. An Atago 3T refractometer was used to determine the concentration of urea with a refractive index of 589 nm. All other chemicals were either specially prepared or guaranteed reagent-grade chemicals. Horse skeletal muscle myoglobin was purchased from Sigma (St. Louis, MO, USA). The heme group of myoglobin was removed according to Teale[54] as well as Uzawa *et al.*[23] to prepare apo-form of myoglobin. HapoMb thus obtained was lyophilized and stored at -20°C.[39] The concentration of hapoMb was determined spectrophotometrically by using an extinction coefficient of ε_{280} = 14,300 M^{-1} cm^{-1}.[23] The protein concentration of the equilibrium fluorescence and far-UV CD measurements was 5 μM, whereas those of the near-UV CD measurements were 22–30 μM. For the kinetic measurements using CF and SF fluorescence, the concentrations were 20–40 μM and 5 μM, respectively. All measurements were performed in 12 mM sodium citrate containing appropriate concentration of urea at pH 6.0, in 12 mM sodium citrate at pH 2.2–7.1 (in the absence of urea) or in HCl at pH 2.0. All the solutions used

in the measurements were filtered by membrane filters (pore size of 0.20 or 0.45 μm) before the measurements.

Equilibrium CD measurements

The sample solutions at pH 6.0 and 4.0 for the far-UV CD measurements were prepared by mixing a protein stock solution (12 mM sodium citrate pH 6.0) with 12 mM sodium citrate containing an appropriate concentration of urea. Lyophilized hapoMb was dissolved in HCl at pH 2.0. The protein solution was passed through a Sephadex G-25 column (PD-10; GE Healthcare) equilibrated with 12 mM sodium citrate at pH 6.0. The sample solutions containing various concentrations of urea ranging from 0 M to 8.0 M in steps of 0.05–1.0 M at pH 6.0 and 4.0 were prepared by six-fold diluting the protein stock solution with 12 mM sodium citrate containing an appropriate concentration of urea. The sample solution at pH 2.0 was prepared essentially in the same way albeit with a small modification. The protein solution obtained by dissolving the lyophilized protein in HCl at pH 2.0 was passed through a Sephadex G-25 column equilibrated with HCl at pH 2.0, followed by diluting it with HCl at pH 2.0.

The sample solutions at pH 6.0 and 4.0 for the near-UV CD measurements were prepared in a different way. The lyophilized hapoMb was dissolved in HCl at pH 2.0. The protein solution was mixed with 1.0 M sodium citrate to give 25 mM sodium citrate at pH 6.0, followed by centrifugation to remove precipitates. The solution thus obtained was again mixed with the solution containing appropriate concentrations of sodium citrate and urea. The sample solution contained an appropriate concentration of hapoMb in 12 mM sodium citrate / 0–6.7 M urea at pH 6.0 or in 12 mM sodium citrate at pH 4.0. The sample solutions at pH 2.0 were prepared in the same way as the equilibrium fluorescence and far-UV CD measurements.

Equilibrium CD measurements were carried out with a Jasco J-600S spectropolarimeter (Tokyo, Japan) at 0–20°C with a Peltier device. The pathlengths of the sample cuvettes for the far- and near-UV CD measurements were 1.0 and 10.0 mm, respectively.

Equilibrium fluorescence measurements

The sample solutions for the urea-induced equilibrium unfolding measurements were prepared by the same way to the equilibrium far-UVCD measurements. The sample solutions for the pH-induced equilibrium unfolding measurements were prepared in 12 mM sodium citrate at a pH value ranging from 2.2–7.1 or in HCl at pH 2.0. The protein solution obtained by dissolving the lyophilized protein in HCl at pH 2.0 was passed through a Sephadex G-25 column equilibrated with HCl at pH 2.0, followed by two-fold dilution with HCl at pH 2.0 or with 24 mM sodium citrate buffers at appropriate pH. The fluorescence emission spectra from 300 nm to 450 nm were recorded on a spectrofluorometer, Jasco FP-777 (Tokyo, Japan), using an excitation wavelength of 295 nm. The measurements were carried out at 8°C with circulating water.

Kinetic fluorescence measurements

(i) *Urea-induced kinetic measurements using CF fluorescence*

Lyophilized protein was dissolved in HCl at pH 2.0. The protein solution was dialyzed with deionized water, followed by centrifugation to remove the precipitates. Acid-unfolded protein was obtained by titration of the protein solution thus obtained with 1.0 M HCl to give a pH value of 2.0. Folding reactions were initiated by mixing of the acid-unfolded protein solution and 24 mM sodium citrate containing an appropriate concentration of urea at a mixing ratio of 1:1 to give 40 μM hapoMb / 12 mM sodium citrate / 0–2.74 M urea at pH 6.0. Kinetic fluorescence measurements were carried out using a CF device[55] at 8°C with air-cooling. The emission was detected by a charge-coupled device using an optical cutoff filter (50% transmittance at 305 nm) with an excitation wavelength at 295 nm. The dead time of the CF measurements was 135 μs, as calibrated by measuring a bimolecular quenching reaction of *N*-acetyl-$_{L}$-tryptophanamide (NATA) fluorescence by *N*-bromosuccinimide (NBS).[56]

(ii) *Urea-induced kinetic measurements using SF fluorescence*

Lyophilized protein was dissolved in HCl at pH 2.0. The protein solution was passed through the Sephadex G-25 columns equilibrated with HCl at pH 2.0 and 12 mM sodium citrate at pH 6.0 for the folding and unfolding reactions, respectively. The folding reactions from the unfolded state were initiated by mixing the protein solution at pH 2.0 and 14.4 mM sodium citrate containing an appropriate concentration of urea at a volume ratio of 1:5 to give 5 μM hapoMb / 12 mM sodium citrate / 0–7.6 M urea at pH 6.0 whether those from the equilibrium intermediate state were initiated by mixing the protein in 12 mM sodium citrate / 0.8 M urea at pH 6.0 and 12 mM citrate containing an appropriate concentration of urea at a volume ratio of 1:5 to give 5 μM hapoMb / 12 mM sodium citrate / 0.13–7.6 M urea at pH 6.0. The unfolding reactions were initiated by mixing the protein solution at pH 6.0 and 12 mM sodium citrate containing an appropriate concentration of urea at pH 6.0 at a volume ratio of 1:5 to give 5 μM hapoMb / 12 mM sodium citrate / 0–7.6 M urea at pH 6.0. Kinetic fluorescence measurements were carried out using an Applied Photophysics SX-17 SF device (Surrey, UK) at 8°C with circulating water. The emission was detected by a photomultiplier tube using an optical cutoff filter (50% transmittance at 305 nm) with the excitation wavelength at 295 nm. The dead time of the SF device was 5.3 ms, as calibrated by the bimolecular quenching of NATA fluorescence by NBS.[56]

(iii) *pH-Induced kinetic measurements using CF fluorescence*

Lyophilized protein was dissolved in HCl at pH 2.0. The acid-unfolded protein solution (in HCl pH 2.0; 40 μM), the pH 4 intermediate protein solution (in 2 mM sodium citrate pH 4.0; 80 μM) and the native protein solution (in 2 mM sodium citrate pH 6.0; 40 μM) were prepared by passing the protein solution thus obtained through the Sephadex G-25 columns equilibrated with HCl at pH 2.0, 2 mM sodium citrate at pH 4.0 and 2 mM sodium citrate at pH 6.0, respectively. (Un)folding reactions were initiated by mixing of the initial material with sodium citrate buffer at a mixing ratio of 1:1 to give

20 or 40 μM hapoMb / 12 mM sodium citrate at a pH value ranging from 2.6–6.0 with CF devices. The measurements were carried out at 8°C with air-cooling. Fluorescence emission was detected by the charge-coupled device using the optical cutoff filter (50% transmittance at 305 nm) with excitation at 295 nm. The dead time of the CF devices were 42–63 μs, as calibrated by the bimolecular quenching of NATA fluorescence by NBS.[56]

(iv) *pH-Induced kinetic measurements using SF fluorescence*

Lyophilized protein was dissolved in HCl at pH 2.0. The protein solution was passed through the Sephadex G-25 column equilibrated with 2 mM sodium citrate at pH 6.0. The solution thus obtained was seven-fold diluted with HCl pH 2.0, 2 mM sodium citrate at pH 3.7 or 2 mM sodium citrate at pH 6.0 to give the acid-unfolded (in HCl pH 2.0), pH 4 intermediate (in 2 mM sodium citrate at pH 4.0) and the native (in 2 mM sodium citrate at pH 6.0) protein solutions, respectively. The protein concentration was 10 μM. (Un)folding reactions were initiated by mixing of the initial material with sodium citrate buffer at a mixing ratio of 1:1 to give 5 μM hapoMb / 12 mM sodium citrate at a pH value ranging from 2.3–6.7 with the SF device. The measurements were carried out at 8°C with circulating water. Emission was detected by the photomultiplier tube using the optical cutoff filter with excitation at 295 nm. The dead time of the SF device was 3.7 ms, as calibrated by the bimolecular quenching of NATA fluorescence by NBS.[56]

Data analysis

(i) *Fitting of the equilibrium unfolding transition curve to a three state model.*

A three-state equilibrium unfolding transition consisting of N_{eq}, I_{eq} and U_{eq} are represented as follows:

$$N_{eq} \overset{K_{NeqIeq}}{\rightleftharpoons} I_{eq} \overset{K_{IeqUeq}}{\rightleftharpoons} U_{eq}, \qquad \text{Scheme III-1}$$

where K_{NeqIeq} (=[I_{eq}]/[N_{eq}]) and K_{IeqUeq} (=[U_{eq}]/[I_{eq}]) are the equilibrium constants of the $N_{eq} \rightleftharpoons I_{eq}$ and $I_{eq} \rightleftharpoons U_{eq}$ transitions, respectively. The fraction of N_{eq}, I_{eq} and U_{eq} (f_{Neq}, f_{Ieq} and f_{Ueq}, respectively) is represented as follows:

$$f_{Neq} = \frac{1}{1 + K_{NeqIeq} + K_{NeqIeq}K_{IeqUeq}},$$
$$f_{Ieq} = \frac{K_{NeqIeq}}{1 + K_{NeqIeq} + K_{NeqIeq}K_{IeqUeq}}, \qquad \text{(III-1)}$$
$$f_{Ueq} = \frac{K_{NeqIeq}K_{IeqUeq}}{1 + K_{NeqIeq} + K_{NeqIeq}K_{IeqUeq}}$$

The observed intensity in fluorescence or CD at a urea concentration, c, and pH, $I_{obs}(c, \text{pH})$ is represented by;

$$I_{\mathrm{obs}}(c,\mathrm{pH}) = I_{\mathrm{Neq}}(c,\mathrm{pH})f_{\mathrm{Neq}}(c,\mathrm{pH}) + I_{\mathrm{Ieq}}(c,\mathrm{pH})f_{\mathrm{Ieq}}(c,\mathrm{pH}) + I_{\mathrm{Ueq}}(c,\mathrm{pH})f_{\mathrm{Ueq}}(c,\mathrm{pH}) \quad \text{(III-2)}$$

where $I_{\mathrm{Neq}}(c,\mathrm{pH})$, $I_{\mathrm{Ieq}}(c,\mathrm{pH})$, and $I_{\mathrm{Ueq}}(c,\mathrm{pH})$ are the intensity in fluorescence or CD of N_{eq}, I_{eq} and U_{eq}, respectively, at c and pH, which are assumed to be linearly dependent on urea concentration and pH. The Gibbs free energy difference, ΔG_{ij}, between i and j states (i, j: N_{eq}, I_{eq} and U_{eq}) is approximated to be linearly dependent on urea concentration, c, at a fixed pH. Thus, the Gibbs free energy difference is represented as:

$$\begin{aligned}\Delta G_{\mathrm{NeqIeq}} &= -2.303 \times RT \log K_{\mathrm{NeqIeq}} = m_{\mathrm{NeqIeq}}\left(c_{\mathrm{mNeqIeq}} - c\right)\\ \Delta G_{\mathrm{IeqUeq}} &= -2.303 \times RT \log K_{\mathrm{IeqUeq}} = m_{\mathrm{IeqUeq}}\left(c_{\mathrm{mIeqUeq}} - c\right)\end{aligned} \quad \text{(III-3a)}$$

$$\begin{aligned}\Delta G^{\mathrm{H2O}}{}_{\mathrm{NeqIeq}} &= m_{\mathrm{NeqIeq}} c_{\mathrm{mNeqIeq}}\\ \Delta G^{\mathrm{H2O}}{}_{\mathrm{IeqUeq}} &= m_{\mathrm{IeqUeq}} c_{\mathrm{mIeqUeq}}\end{aligned} \quad \text{(III-3b)}$$

where $\Delta G^{\mathrm{H2O}}{}_{ij}$ is the free energy difference between i and j states in the absence of urea, m_{ij} is the coefficient (m-value), which indicate the cooperativity of the transition, $c_{\mathrm{m}ij}$ is the midpoint urea concentrations of the $i \rightleftharpoons j$ transitions. R is the gas constant and T is the absolute temperature. The urea-induced equilibrium unfolding transition curves were fitted by non-linear least-squares fitting using Equation (III-2) in addition to Equations (III-1) and (III-3a) to obtain the thermodynamic parameters and the ellipticity and fluorescence intensity of each state. The pH-induced equilibrium unfolding transition curves were fitted by linear least-squares fitting to obtain the fluorescence intensity of each state, using the fraction of the three states (f_{Neq}, f_{Ieq} and f_{Ueq}) numerically estimated by quantitative modeling and Equation (III-2).

(ii) *Fitting of the kinetic traces obtained by the folding and unfolding reactions.*

The kinetic traces obtained by the CF and SF measurements under matching conditions were combined and then fitted by non-linear least squares fitting to the equation below:

$$F_{\mathrm{obs}}(t) = F_{\mathrm{eq}} + \sum_i F_i \, \mathrm{e}^{-\lambda_i t} \quad \text{(III-4)}$$

where F_{eq} is the fluorescence intensity at long times after the reactions have reached equilibrium, F_i and λ_i are the amplitude and the rate constant of the i-th phase.

(iii) *Kinetic modeling*

Standard numeric methods were used to solve the system of linear differential equations in matrix form describing the kinetic scheme, using IGOR software (Wavemetrics, Inc., Lake Oswego, OR).

The urea concentration dependence of the microscopic rate constants, $k_{ij}(c)$, crossing a transition state ($\ddagger ij$) from i state to j state, was assumed to follow Equation (III-5):

$$\log k_{ij}(C) = \log k_{ij}^{\mathrm{H2O}} + \frac{m_{ij}^{\ddagger}}{2.303 \times RT} C \quad \text{(III-5)}$$

where $k^{H2O}{}_{ij}$ is the elementary rate constant in the absence of urea at pH 6.0, $m^{\ddagger}{}_{ij}$ is the corresponding slope. The elementary rate constants and the slopes were explored systematically to model the apparent rates and amplitudes observed in the kinetic measurements as well as the equilibrium unfolding experiments, based on the associated rate matrix. Free energy diagrams were calculated from the kinetic parameters. The free energy of the relevant states and the transition states was calculated as a function of the α-value as the reaction coordinate at representative urea concentrations. The α-value was defined by a sum of $m^{\ddagger}{}_{ij}$ from N to a given state normalized by the sum of all $m^{\ddagger}{}_{ij}$, which is the measure of the change in solvent-accessible surface area relative to N.

The pH dependence of the elementary rate constants of transition crossing a transition state (‡*ij*) from state *i* to state *j*, k_{ij}(pH), is represented by Equation (III-6):

$$k_{ij}(pH) = \frac{\left(\frac{1+10^{\mathrm{p}Ka^{\ddagger ij}_{\mathrm{His}}-\mathrm{pH}}}{1+10^{\mathrm{p}Ka^{\mathrm{U}}_{\mathrm{His}}-\mathrm{pH}}}\frac{1+10^{\mathrm{p}Ka^{\mathrm{U}}_{\mathrm{His}}-6}}{1+10^{\mathrm{p}Ka^{\ddagger ij}_{\mathrm{His}}-6}}\right)^{n^{\ddagger}_{\mathrm{His}}}\left(\frac{1+10^{\mathrm{p}Ka^{\ddagger ij}_{\mathrm{COOH}}-\mathrm{pH}}}{1+10^{\mathrm{p}Ka^{U}_{\mathrm{COOH}}-\mathrm{pH}}}\frac{1+10^{pKa^{U}_{COOH}-6}}{1+10^{pKa^{\ddagger ij}_{COOH}-6}}\right)^{n^{\ddagger}_{COOH}}}{\left(\frac{1+10^{\mathrm{p}Ka^{i}_{\mathrm{His}}-\mathrm{pH}}}{1+10^{\mathrm{p}Ka^{\mathrm{U}}_{\mathrm{His}}-\mathrm{pH}}}\frac{1+10^{\mathrm{p}Ka^{\mathrm{U}}_{\mathrm{His}}-6}}{1+10^{\mathrm{p}Ka^{i}_{\mathrm{His}}-6}}\right)^{n^{i}_{\mathrm{His}}}\left(\frac{1+10^{\mathrm{p}Ka^{i}_{\mathrm{COOH}}-\mathrm{pH}}}{1+10^{\mathrm{p}Ka^{\mathrm{U}}_{\mathrm{COOH}}-\mathrm{pH}}}\frac{1+10^{\mathrm{p}Ka^{\mathrm{U}}_{\mathrm{COOH}}-6}}{1+10^{pKa^{i}_{COOH}-6}}\right)^{n^{i}_{COOH}}} k^{\mathrm{H2O}}_{ij}. \quad \text{(III-6)}$$

In Equation (III-5), ionizable groups associated with the (un)folding reactions under the conditions (pH < 7) are assumed to be limited to imidazole group of histidine residues (His), carboxyl group of the aspartate and glutamate residues, and the C-terminal α-carboxyl group (COOH). Based on the assumption, $\mathrm{p}K_{\mathrm{a}}{}^{\ddagger ij}{}_{\mathrm{His}}$, $\mathrm{p}K_{\mathrm{a}}{}^{i}{}_{\mathrm{His}}$ and $\mathrm{p}K_{\mathrm{a}}{}^{\mathrm{U}}{}_{\mathrm{His}}$ are the pK_{a} values of the imidazole group of His in the transition state, *i* state and the unfolded state, respectively, while $n^{\ddagger ij}{}_{\mathrm{His}}$ and $n^{i}{}_{\mathrm{His}}$ are the number of histidine residues associated with the $i \rightarrow j$ transition in the transition state, *i*-state and the unfolded state, respectively. Similarly, $\mathrm{p}K_{\mathrm{a}}{}^{\ddagger ij}{}_{\mathrm{COOH}}$, $\mathrm{p}K_{\mathrm{a}}{}^{i}{}_{\mathrm{COOH}}$ and $\mathrm{p}K_{\mathrm{a}}{}^{\mathrm{U}}{}_{\mathrm{COOH}}$ are the pK_{a} values of the carboxyl group in the transition state, *i*-state and the unfolded state, respectively, while $n^{\ddagger}{}_{\mathrm{COOH}}$ and $n^{i}{}_{\mathrm{COOH}}$ are the number of carboxyl groups associated with the $i \rightarrow j$ transition in the transition state and *i*-state, respectively. The $\mathrm{p}K_{\mathrm{a}}{}^{\mathrm{U}}{}_{\mathrm{His}}$ and $\mathrm{p}K_{\mathrm{a}}{}^{\mathrm{U}}{}_{\mathrm{COOH}}$ values are 6.0 and 3.6 (close to the normal pK_{a} values), respectively. $k^{\mathrm{H2O}}{}_{ij}$ is the elementary rate constant for the $i \rightarrow j$ transition at pH 6.0. For simplicity, the ionizable groups are assumed to be independent of each other, i.e., the ionization of an ionizable group does not affect the ionization of the other ionizable groups. It is also assumed that the ionizable group (the imidazole or carboxyl groups) has the same abnormal $\mathrm{p}K_{\mathrm{a}}{}^{\ddagger ij}$ and $\mathrm{p}K_{\mathrm{a}}{}^{i}$ values when there are more than one ionizable groups associated. The elementary rate constants of each transition and the fluorescence intensity of each state as a function of pH were manually varied to reproduce the pH dependence of the rate constants and amplitudes of the (un)folding kinetics.

The activation free energy for crossing the barriers between *i* and *j* states at urea concentration, *c*, and a pH value, $\Delta G^{\ddagger}{}_{ij}(c, \mathrm{pH})$, is calculated by the following equation:

$$\Delta G^{\ddagger}_{ij}(c, pH) = -2.303 \times RT \log \frac{k_{ij}(c, pH)}{A}, \quad \text{(III-7)}$$

where the pre-exponential factor, *A*, corresponds to the rate constant in a reaction with the zero-activation free energy, and is here assumed to equal 10^6 s^{-1}. The free energy diagram was drawn

based on the Equation (III-7).

REFERENCES

1. Arai M, Kuwajima K (2000) Role of the molten globule state in protein folding. Adv Protein Chem **53**:209-282.
2. Baldwin RL, Rose GD (1999) Is protein folding hierarchic? I. Local structure and peptide folding. Trends Biochem Sci **24**:26-33.
3. Baldwin RL, Rose GD (1999) Is protein folding hierarchic? II. Folding intermediates and transition states. Trends Biochem Sci **24**:77-83.
4. Shastry MCR, Roder H (1998) Evidence for barrier-limited protein folding kinetics on the microsecond time scale. Nat Struct Biol **5**:385-392.
5. Huang CY, Klemke JW, Getahun Z, DeGrado WF, Gai F (2001) Temperature-dependent helix-coil transition of an alanine based peptide. J Am Chem Soc **123**:9235-9238.
6. Roder H, Maki K, Cheng H (2006) Early events in protein folding explored by rapid mixing methods. Chem Rev **106**:1836-1861.
7. Matsumoto S, Yane A, Nakashima S, Hashida M, Fujita M, Goto Y, Takahashi S (2007) A rapid flow mixer with 11-μs mixing time microfabricated by a pulsed-laser ablation technique: observation of a barrier-limited collapse in cytochrome *c* folding. J Am Chem Soc **129**:3840-3841.
8. Kuwajima K, Hiraoka Y, Ikeguchi M, Sugai S (1985) Comparison of the transient folding intermediates in lysozyme and α-lactalbumin. Biochemistry **24**:874-881.
9. Ikeguchi M, Kuwajima K, Mitani M, Sugai S (1986) Evidence for identity between the equilibrium unfolding intermediate and a transient folding intermediate: A comparative study of the folding reactions of α-lactalbumin and lysozyme. Biochemistry **25**:6965-6972.
10. Uzawa T, Nishimura C, Akiyama S, Ishimori K, Takahashi S, Dyson HJ, Wright PE (2008) Hierarchical folding mechanism of apomyoglobin revealed by ultra-fast H/D exchange coupled with 2D NMR. Proc Natl Acad Sci USA **105**:13859-13864.
11. Jennings PA, Wright PE (1993) Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. Science **262**:892-895.
12. Hughson FM, Wright PE, Baldwin RL (1990) Structural characterization of a partly folded apomyoglobin intermediate. Science **249**:1544-1548.
13. Raschke TM, Marqusee S (1997) The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. Nat Struct Biol **4**:298-304.
14. Latypov RF, Cheng H, Roder NA, Zhang J, Roder H (2006) Structural characterization of an equilibrium unfolding intermediate in cytochrome c. J Mol Biol **357**:1009-1025.
15. Roder H, Elöve GA, Englander SW (1988) Structural characterization of folding

intermediates in cytochrome c by H-exchange labelling and proton NMR. Nature **335**:700-704.

16. Arai M, Kuwajima K (1996) Rapid formation of a molten globule intermediate in refolding of α-lactalbumin. Fold Des **1**:275-287.
17. Wu LC, Peng Z-y, Kim PS (1995) Bipartite structure of the α-lactalbumin molten globule. Nat Struct Biol **2**:281-286.
18. Schulman BA, Kim PS (1996) Proline scanning mutagenesis of a molten globule reveals non-cooperative formation of a protein's overall topology. Nat Struct Biol **3**:1-6.
19. Luo Y, Kay MS, Baldwin RL (1997) Cooperativity of folding of the apomyoglobin pH 4 intermediate studied by glycine and proline mutations. Nat Struct Biol **4**:925-930.
20. Kendrew JC, Dickerson RE, Strandberg BE, Hart RG, Davies DR, Phillips DC, Shore VC (1960) Structure of myoglobin: A three-dimensional Fourier synthesis at 2 A. resolution. Nature **185**:422-427.
21. Eliezer D, Wright PE (1996) Is apomyoglobin a molten globule? Structural characterization by NMR. J Mol Biol **263**:531-538.
22. Jamin M, Baldwin RL (1998) Two forms of the pH 4 folding intermediate of apomyoglobin. J Mol Biol **276**:491-504.
23. Uzawa T, Akiyama S, Kimura T, Takahashi S, Ishimori K, Morishima I, Fujisawa T (2004) Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of α-helical content and compactness. Proc Natl Acad Sci USA **101**:1171-1176.
24. Ha JH, Loh SN (1998) Changes in side chain packing during apomyoglobin folding characterized by pulsed thiol-disulfide exchange. Nat Struct Biol **5**:730-737.
25. Haruta N, Kitagawa T (2002) Time-resolved UV resonance Raman investigation of protein folding using a rapid mixer: characterization of kinetic folding intermediates of apomyoglobin. Biochemistry **41**:6595-6604.
26. Xu M, Beresneva O, Rosario R, Roder H (2012) Microsecond folding dynamics of apomyoglobin at acidic pH. J Phys Chem B **116**:7014-7025.
27. Griko YV, Privalov PL, Venyaminov SY, Kutyshenko VP (1988) Thermodynamic study of the apomyoglobin structure. J Mol Biol **202**:127-138.
28. Barrick D, Baldwin RL (1993) Three-state analysis of sperm whale apomyoglobin folding. Biochemistry **32**:3790-3796.
29. Khorasanizadeh S, Peters ID, Roder H (1996) Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. Nat Struct Biol **3**:193-205.
30. Nishii I, Kataoka M, Tokunaga F, Goto Y (1994) Cold denaturation of the molten globule states of apomyoglobin and a profile for protein folding. Biochemistry **33**:4903-4909.
31. Kamatari YO, Ohji S, Konno T, Seki Y, Soda K, Kataoka M, Akasaka K (1999) The compact and expanded denatured conformations of apomyoglobin in the methanol-water solvent.

Protein Sci **8**:873-882.

32. Herskovits TT, Solli NJ (1975) Studies of the conformation of apomyoglobin in aqueous solutions and denaturing organic solvents. Biopolymers **14**:319-334.
33. Eftink MR, Jia Y, Hu D, Ghiron CA (1995) Fluorescence Studies with Tryptophan Analogues: Excited State Interactions Involving the Side Chain Amino Group. J Phys Chem **99**:5713-5723.
34. Jamin M, Yeh SR, Rousseau DL, Baldwin RL (1999) Submillisecond unfolding kinetics of apomyoglobin and its pH 4 intermediate. J Mol Biol **292**:731-740.
35. Sabelko J, Ervin J, Gruebele M (1998) Cold-denatured ensemble of apomyoglobin: Implications for the early steps of folding. J Phys Chem B **102**:1806-1819.
36. Nishii I, Kataoka M, Goto Y (1995) Thermodynamic stability of the molten globule states of apomyoglobin. J Mol Biol **250**:223-238.
37. Nishimura C, Dyson HJ, Wright PE (2010) Energetic frustration of apomyoglobin folding: role of the B helix. J Mol Biol **396**:1319-1328.
38. Scott EE, Paster EV, Olson JS (2000) The stabilities of mammalian apomyoglobins vary over a 600-fold range and can be enhanced by comparative mutagenesis. J Biol Chem **275**:27129-27136.
39. Nishiguchi S, Goto Y, Takahashi S (2007) Solvation and desolvation dynamics in apomyoglobin folding monitored by time-resolved infrared spectroscopy. J Mol Biol **373**:491-502.
40. Tsui V, Garcia C, Cavagnero S, Siuzdak G, Dyson HJ, Wright PE (1999) Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through and obligatory intermediate. Protein Sci **8**:45-49.
41. Ballew RM, Sabelko J, Gruebele M (1996) Direct observation of fast protein folding: the initial collapse of apomyoglobin. Proc Natl Acad Sci USA **93**:5759-5764.
42. Eliezer D, Chung J, Dyson HJ, Wright PE (2000) Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin. Biochemistry **39**:2894-2901.
43. Loh SN, Kay MS, Baldwin RL (1995) Structure and stability of a second molten globule intermediate in the apomyoglobin folding pathway. Proc Natl Acad Sci USA **92**:5446-5450.
44. Nishimura C, Dyson HJ, Wright PE (2006) Identification of native and non-native structure in kinetic folding intermediates of apomyoglobin. J Mol Biol **355**:139-156.
45. Bai Y, Sosnick TR, Mayne L, Englander SW (1995) Protein folding intermediates: native-state hydrogen exchange. Science **269**:192-197.
46. Nishimura C, Dyson HJ, Wright PE (2008) The kinetic and equilibrium molten globule intermediates of apoleghemoglobin differ in structure. J Mol Biol **378**:715-725.
47. Maki K, Cheng H, Dolgikh DA, Shastry MC, Roder H (2004) Early Events During Folding of Wild-type Staphylococcal Nuclease and a Single-tryptophan Variant Studied by Ultrarapid Mixing. J Mol Biol **338**:383-400.

48. Colón W, Elöve GA, Wakem LP, Sherman F, Roder H (1996) Side chain packing of the N- and C-terminal helices plays a critical role in the kinetics of cytochrome *c* folding. Biochemistry **35**:5538-5549.
49. Walkenhorst WF, Green SM, Roder H (1997) Kinetic evidence for folding and unfolding intermediates in Staphylococcal nuclease. Biochemistry **63**:5795-5805.
50. Geierstanger B, Jamin M, Volkman BF, Baldwin RL (1998) Protonation behavior of histidine 24 and histidine 119 in forming the pH 4 folding intermediate of apomyoglobin. Biochemistry **37**:4254-4265.
51. Meinhold DW, Wright PE (2011) Measurement of protein unfolding/refolding kinetics and structural characterization of hidden intermediates by NMR relaxation dispersion. Proc Natl Acad Sci USA **108**:9078-9083.
52. Oliveberg M, Arcus VL, Fersht AR (1995) pKA values of carboxyl groups in the native and denatured states of barnase: the pKA values of the denatured state are on average 0.4 units lower than those of model compounds. Biochemistry **34**:9424-9433.
53. Capaldi AP, Shastry RMC, Kleanthous C, Roder H, Radford SE (2001) Ultra-rapid mxing experiments reveal that Im7 folds *via* an on-pathway intermediate. Nat Struct Biol **8**:68-72.
54. Teale FW (1959) Cleavage of the haem-protein link by acid methylethylketone. Biochim Biophys Acta **35**:543.
55. Shastry MCR, Luck SD, Roder H (1998) A continuous-flow capillary mixing method to monitor reactions on the microsecond time scale. Biophys J **74**:2714-2721.
56. Peterman BF (1979) Measurement of the dead time of a fluorescence stopped-flow instrument. Anal Biochem **93**:442-444.
57. Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph **14**:51-55, 29-32.

TABLES

Table III-1: Thermodynamic parameters for urea-induced equilibrium unfolding of hapoMb at pH 6.0 and 8°C.

i j	m_{ij} (kcal/mol/M)	c_{mij} (M)	ΔG^{H2O}_{ij} (kcal/mol)
Equilibrium[a]			
N_{eq} I_{eq}	1.67 ± 0.01	0.99 ± 0.01	1.65 ± 0.01
I_{eq} U_{eq}	1.35 ± 0.01	1.61 ± 0.01	2.17 ± 0.01
N_{eq} U_{eq}	3.01 ± 0.01[b]	1.27 ± 0.01	3.82 ± 0.01[b]
Kinetics[c]			
N_{eq} I_{eq}	1.41	0.61	0.86
I_{eq} U_{eq}	1.41	1.59	2.23
N_{eq} U_{eq}	2.81	1.10	3.09

[a] Thermodynamic parameters obtained by globally fitting Scheme III-1 to a collection of the equilibrium unfolding transition curves. Error estimates for equilibrium parameters are based on goodness-of-fit (± one standard deviation).

[b] ΔG^{H2O}_{NeqUeq} and m_{NeqUeq} were calculated by adding the corresponding thermodynamic parameter for $N_{eq} \rightleftharpoons I_{eq}$ and $I_{eq} \rightleftharpoons U_{eq}$ transitions.

[c] Thermodynamic parameters were calculated from the elementary rate constants and kinetic *m*-values (Table III-2) estimated by the quantitative modeling. N_{eq} is assumed to consist of N and M, U_{eq} is assumed to consist of I´ and U, and I_{eq} corresponds to I.

Table III-2: Kinetic parameters estimated by monitoring the folding kinetics of hapoMb based on Scheme III-2 in the absence of urea at pH 6.0 and 8°C.

i j	k^{H2O}_{ij} (s^{-1})[a]	$m^{‡}_{ij}$ (kcal/mol/M)[b]
U I′[c]	9.57×10^{4}	-0.76
I′ U[c]	2.83×10^{4}	0.24
I′ I	1.60×10^{4}	0.00
I I′	5.13×10^{2}	0.57
I M	3.00	-0.70
M I	1.74×10^{1}	0.20
M N[d]	4.32×10^{3}	-0.20
N M[d]	1.63×10^{2}	0.35

[a] k^{H2O}_{ij} indicates the microscopic rate constant for a given process in the absence of urea at pH 6.0.

[b] Kinetic *m*-values, $m^{‡}_{ij}$, were estimated according to Equation (III-2).

[c] Only the ratio $k_{UI'}/k_{I'U}$ is uniquely determined.

[d] Only the ratio k_{NM}/k_{MN} is uniquely determined

Table III-3: Fluorescence intensity of each species (relative to that at pH 2.0) in the absence of urea at pH 6.0 and 8°C.

i	Equilibrium (urea)[a]	Equilibrium (pH)[b]	Kinetics[c]
U	1.13	1.22	1.25
I´	-	-	0.80
I	1.83	1.72	1.90
M	-	-	1.58
N	1.24	1.28	1.32

[a] Values were obtained by globally fitting Scheme III-1 to a collection of the urea-induced unfolding transition curves.

[b] Values were obtained by fitting a collection of the pH-induced unfolding transition curves.

[c] Values were obtained by modeling the folding/unfolding kinetics based on Scheme III-2.

Table III-4: Kinetic parameters used for the modeling the folding/unfolding kinetics of hapoMb based on Scheme III-2 at pH 6.0 and 8°C.

State (*i*)	$\Delta G^{H2O}{}_{Ni}$ [a] (kcal/mol)	$n^{i}{}_{His}$	$pK_a{}^{i}{}_{His}$	$n^{i}{}_{COOH}$	$pK_a{}^{i}{}_{COOH}$
U	3.8	0	6.0	0	3.6
TS1 [b]	5.1 [d]	0 [d]	6.0 [d]	4 [d]	2.9 [d]
I´	3.1	0	6.0	5	2.7
TS2	5.4	1	5.8	5	2.7
I	1.0	1	5.8	5	2.6
TS3	7.9	1	5.4	5	2.6
M	1.6	1	3.5	7	2.6
TS4 [c]	4.9	1	3.5	7	2.6
N	0	2	3.5	8	2.6

[a] The Gibbs free energy in an state, *i*, relative to N at pH 6.0 in the absence of urea at pH 6.0.

[b] The transition state between the U and I´ states.

[c] The transition state between the M and N states.

[d] These values are not uniquely determined.

FIGURES

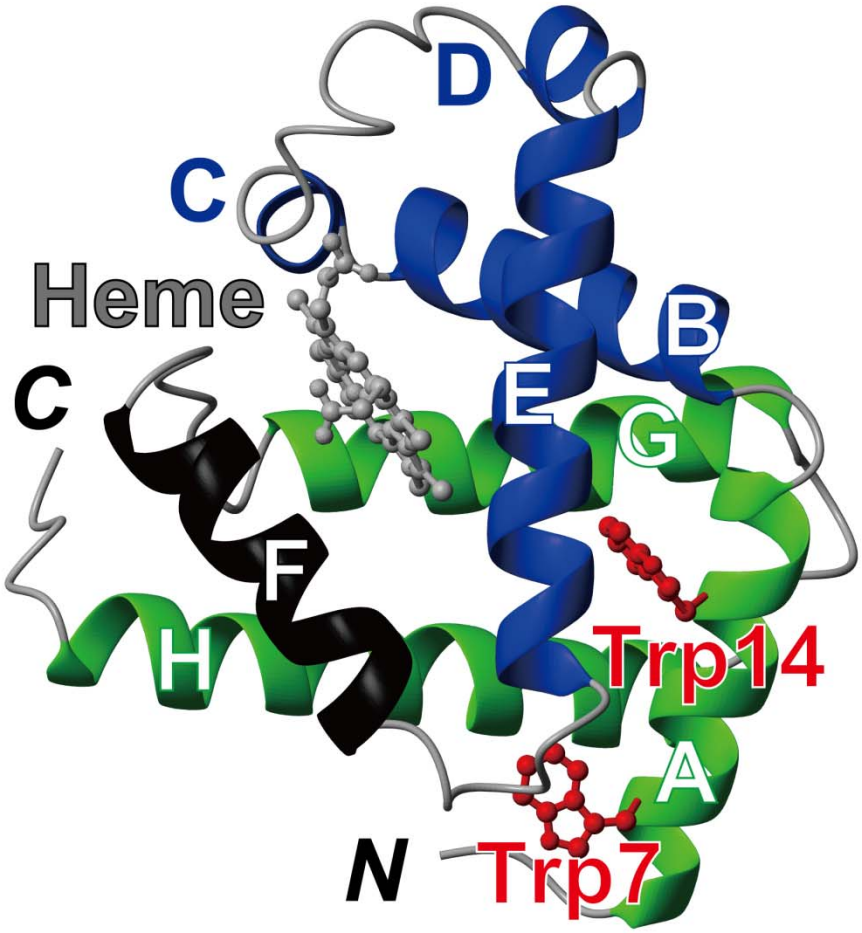


Figure III-1: Ribbon diagram of horse skeletal muscle myoglobin on the crystallographic structure (PDB: 1AZI). The **heme group (light gray)** and **two tryptophanyl side chains (red)** are shown in the ball-and-stick model. The native state of myoglobin contains eight helices (A–H), while the native state of hapoMb contains seven helices (A–E, G, and H (**blue** and **green**)). The intermediate of apoMb exhibits helical structure in **the A-, G- and H-helix regions (green)**. The **F-helix is shown in black**. The figure was prepared using the program MolMol.[57]

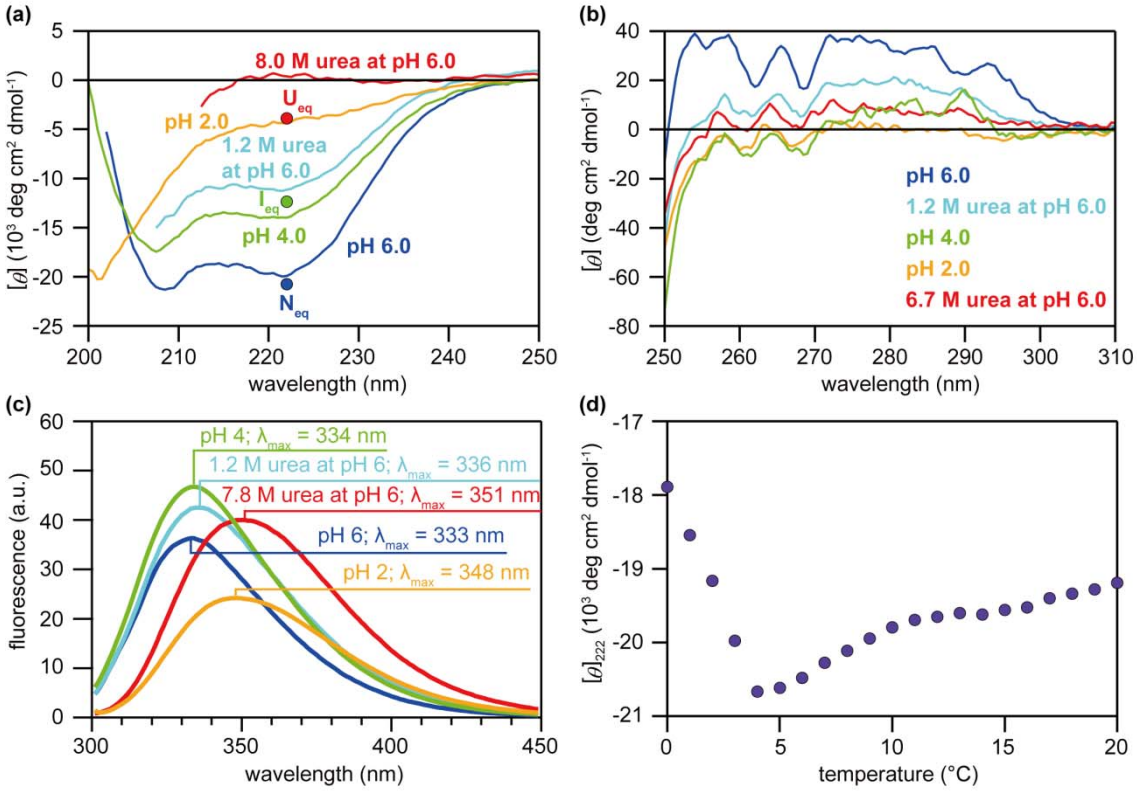


Figure III-2: (a) CD spectra of hapoMb in far-UV (b) near-UV regions as well as (c) fluorescence spectra at 8°C. The spectra under various conditions are denoted by colored lines: **blue (pH 6.0)**, **cyan (1.2 M urea and pH 6.0)**, **red ((a) 8 M urea and (b) 7.8 M urea and pH 6.0))**, **green (pH 4.0)**, and **orange (pH 2.0)**. The circles show the mean residue ellipticity of **N_{eq} (blue)**, **I_{eq} (green)**, and **U_{eq} (red)** at 222 nm obtained by urea-induced equilibrium unfolding. λ_{max} is also explicitly shown in the fluorescence spectra. (d) The temperature dependence of mean residue ellipticity at 222 nm and pH 6.0.

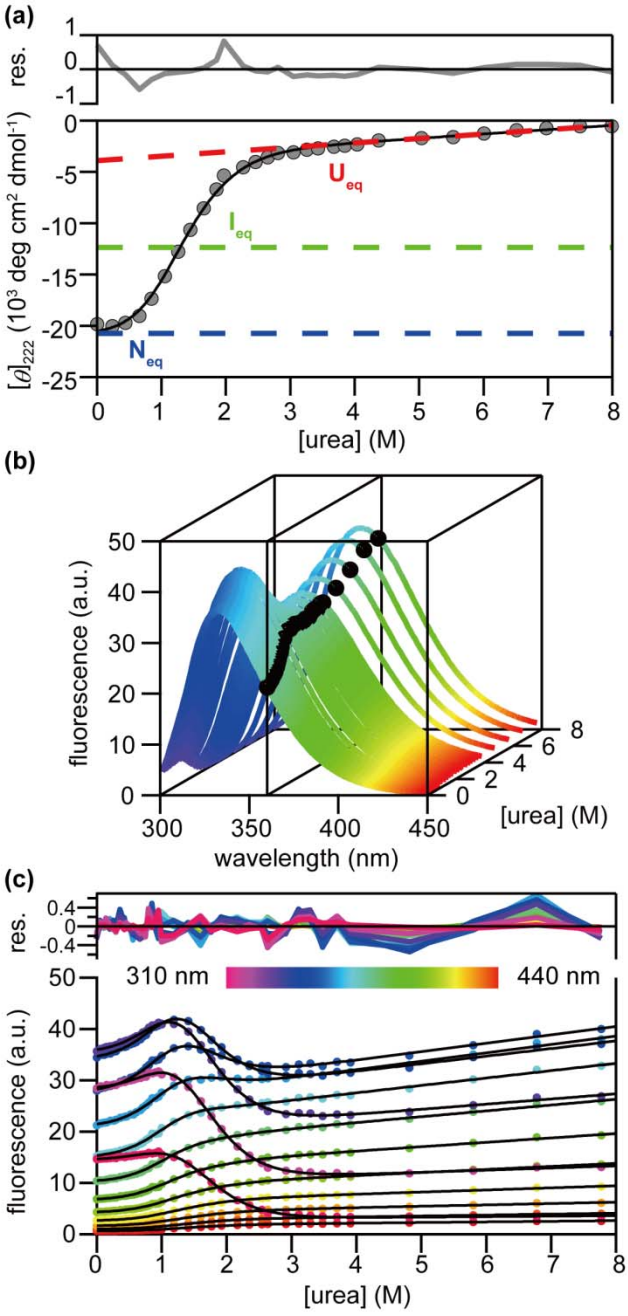


Figure III-3: (a) A urea-induced equilibrium unfolding transition curve of hapoMb monitored by ellipticity at 222 nm at pH 6.0 and 8°C. The solid and broken lines show the unfolding transition curve and the baselines of the **N_{eq} (blue)**, **I_{eq} (green)** and **U_{eq} (red)**, respectively, predicted by global fitting of a three-state model (Scheme 1) to data.

(b) A collection of fluorescence spectra of hapoMb ranging from **310** nm to **440** nm as a function of urea concentration at pH 6.0 and 8°C. Black circles represent an unfolding transition curve monitored by the fluorescence emission at **360 nm**.

(c) Urea-induced equilibrium unfolding transition curves of hapoMb monitored by fluorescence emission at wavelengths ranging from **310** to **440** nm at pH 6.0 and 8°C. The solid lines are the unfolding transition curves predicted by global fitting of a three-state model (Scheme 1) to data. The transition curves in (c) are obtained by transposing the results from (b). Colors represent the emission wavelength. The upper panel shows the residuals of curve fittings.

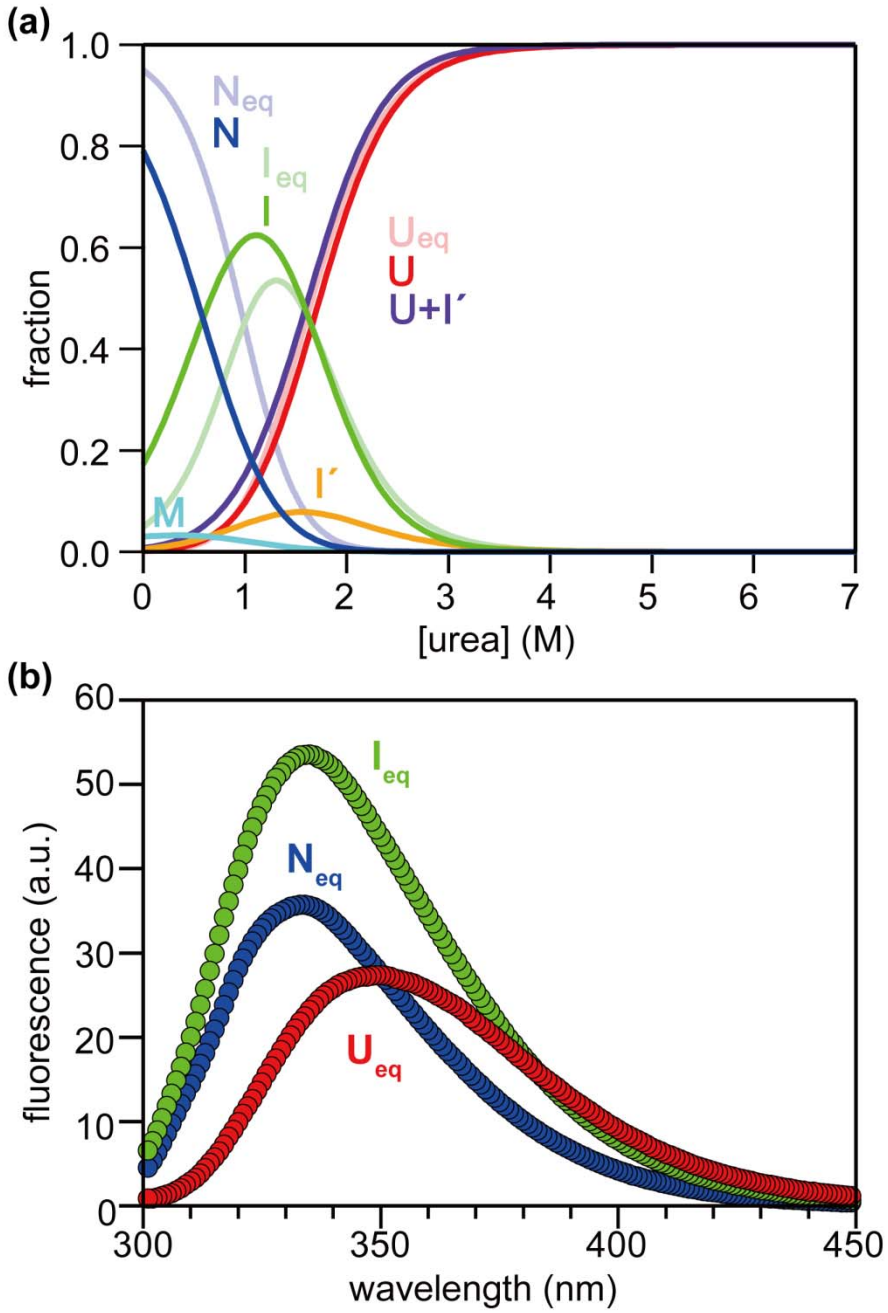


Figure III-4: (a) The fractions of each species as a function of urea concentration. The fractional N_{eq}, I_{eq} and U_{eq} were predicted based on unfolding equilibrium while the fractional **N**, **M**, **I**, **I'** and **U** were predicted based on kinetic experiments. Color codes: **light blue (N_{eq})**, **light green (I_{eq})**, **pink (Ueq)**, **blue (N)**, **cyan (M)**, **green (I)**, **orange (I')**, **red (U)**, and **purple (U and I')**. (b) The deconvoluted fluorescence spectra of Neq, Ieq and Ueq in the absence of urea predicted by global fitting of a three-state model (Scheme 1) to the equilibrium unfolding data. Color codes: **blue (N_{eq})**, **green (I_{eq})** and **red (U_{eq})**.

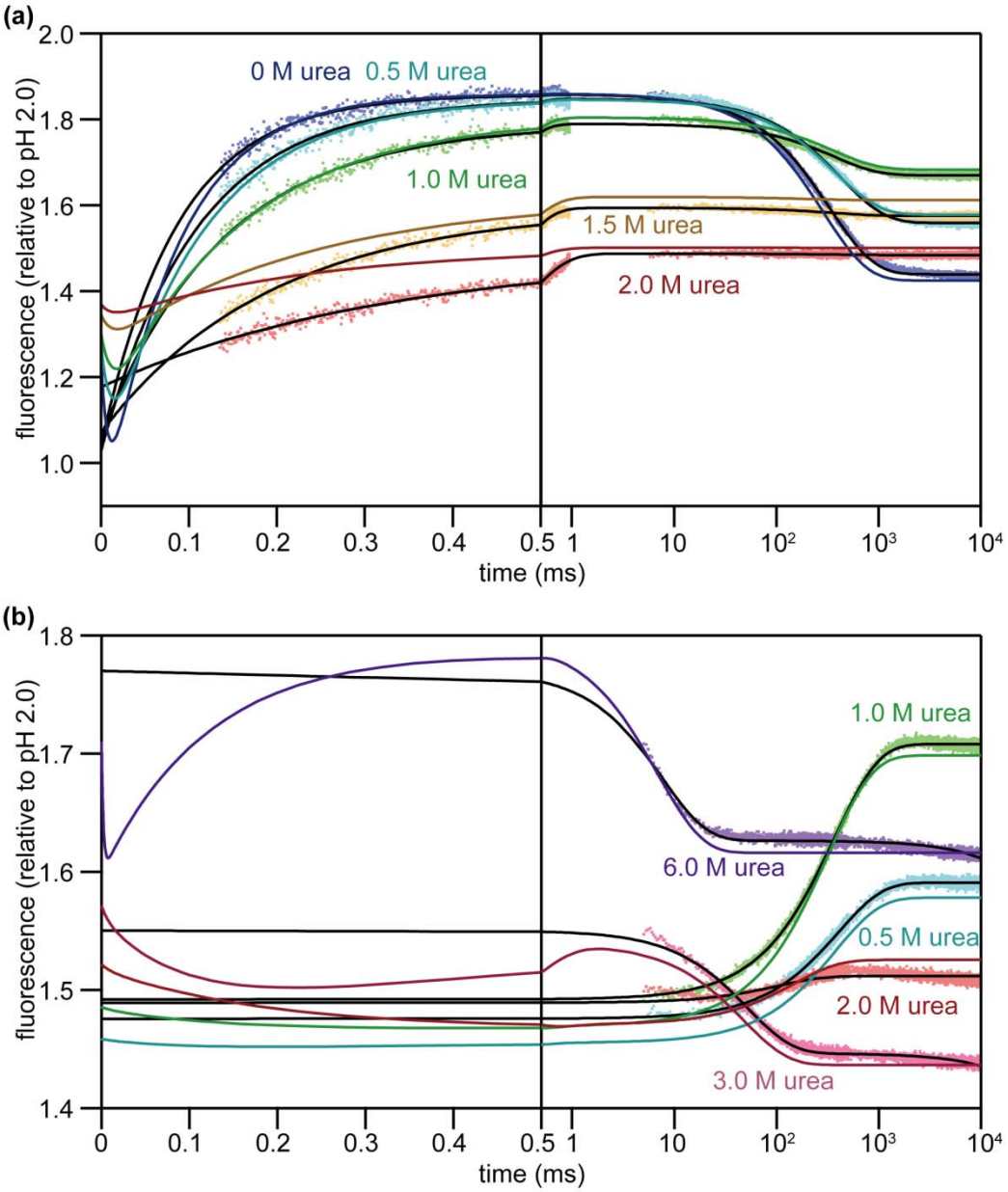


Figure III-5: The kinetic traces of (a) folding and (b) unfolding reactions at various urea concentrations at pH 6.0 and 8°C measured by the CF and SF fluorescence methods. The fluorescence intensity is scaled relative to the fluorescence at pH 2.0. Black lines show the kinetic traces predicted by non-linear least squares fitting of (a) a sum of two exponential functions and (b) a single-exponential function to data. Colored lines show the kinetic traces reproduced by the elementary rate constants (Table III-2) and the fluorescence intensities of each state (Table III-3). The color codes are explicitly shown in the figure.

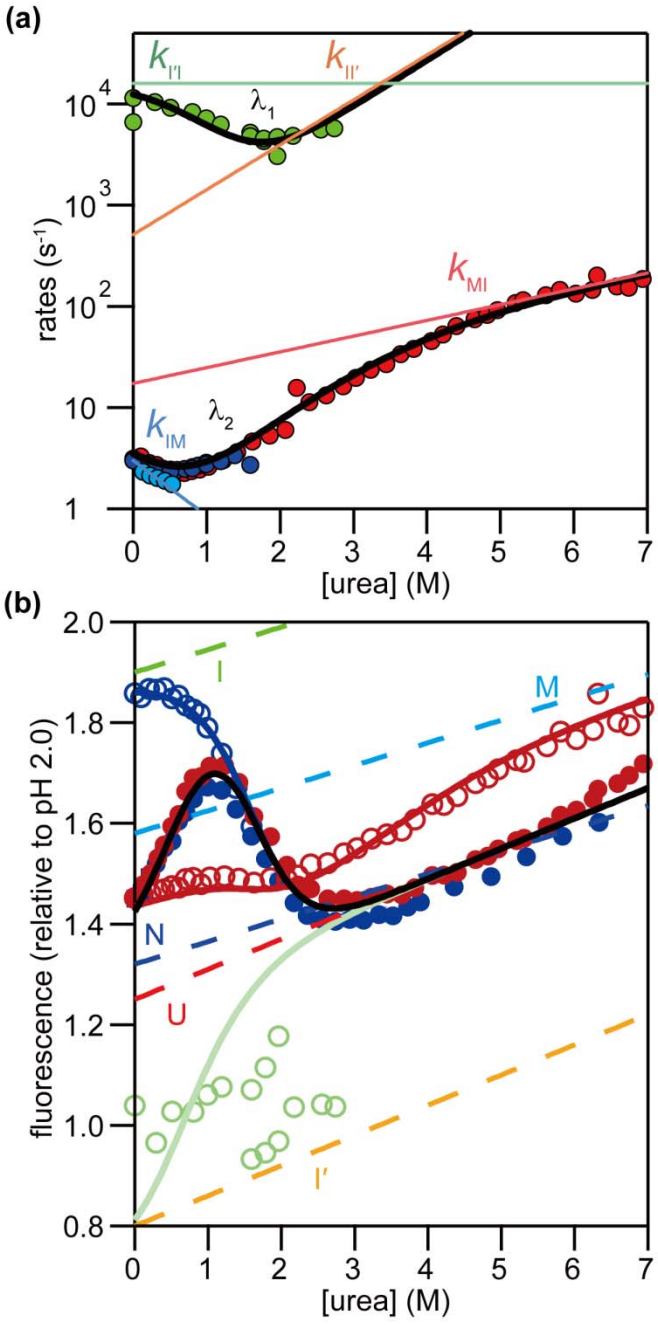


Figure III-6:

(a) Urea concentration dependence of the rate constants of folding/unfolding reactions obtained by the CF and SF experiments (circles). Color codes: λ_1 **(green)**, λ_2 **(blue)**, λ_{1eqN} **(cyan)** and λ_u **(red)**. Black and colored solid lines show the rate constants (λ_1, λ_2 and λ_u) and the elementary rate constants, respectively, predicted by quantitative modeling. The color codes for the elementary rate constants are explicitly shown in the figure.

(b) The amplitudes obtained by the CF and SF experiments (circles). Open circles: F_{01} **(=** F_{eq} **+** F_1 **+** F_2**) (green)**, F_{02} **(=** F_{eq} **+** F_2**) (blue)** for the folding and F_{0U} **(=** F_{eq} **+** F_1**) (red)** for the unfolding. Filled circles: The fluorescence intensity at a long time after **the folding (blue)** or **unfolding (red)** reactions reach the equilibrium **(**F_{eq}**)**. The black solid line represents F_{eq} as a function of the urea concentration obtained by quantitative modeling. The colored solid line represents F_{01} **(green)**, F_{02} **(blue)** and F_{0U} **(red)** as a function of urea concentration predicted by quantitative modeling. Dashed lines represent the urea-concentration dependence of the fluorescence intensities of **U (red)**, **I' (orange)**, **I (green)**, **M (cyan)** and **N (blue)** obtained by quantitative modeling.

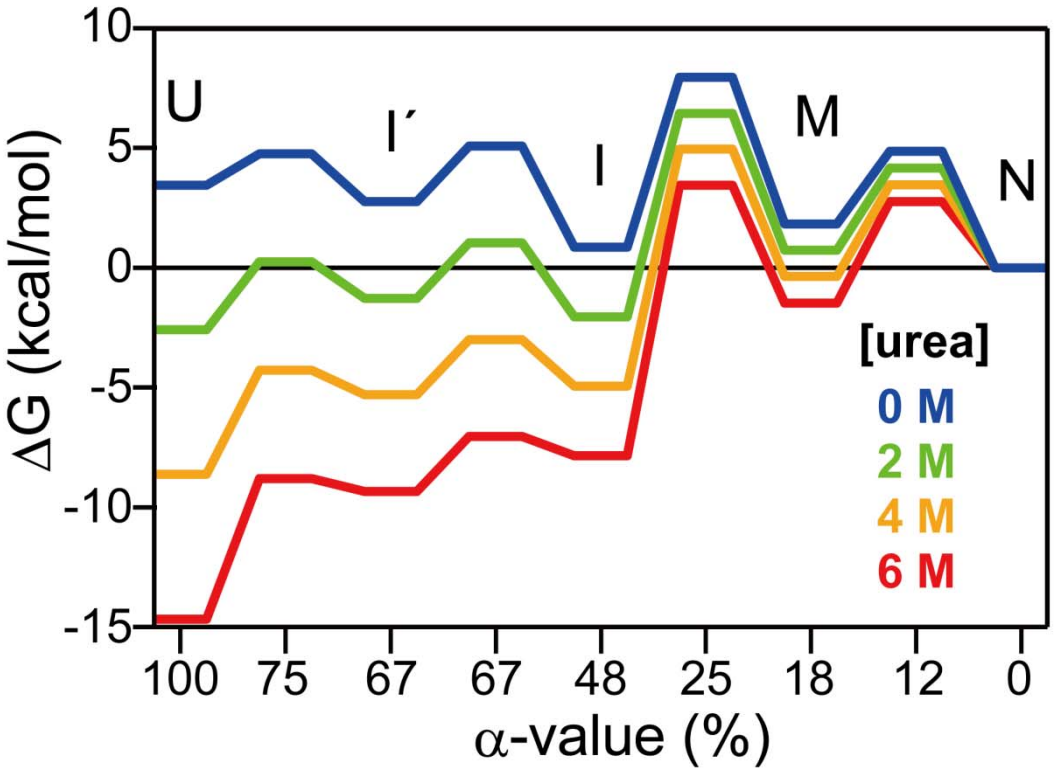


Figure III-7: The free energy diagram of apoMb at pH 6.0 and 8°C showing the effects of urea on the energy levels and the transition states predicted by quantitative analysis of the folding and unfolding kinetics. The reaction coordinate (α) indicates the change in the solvent-accessible surface area relative to N.

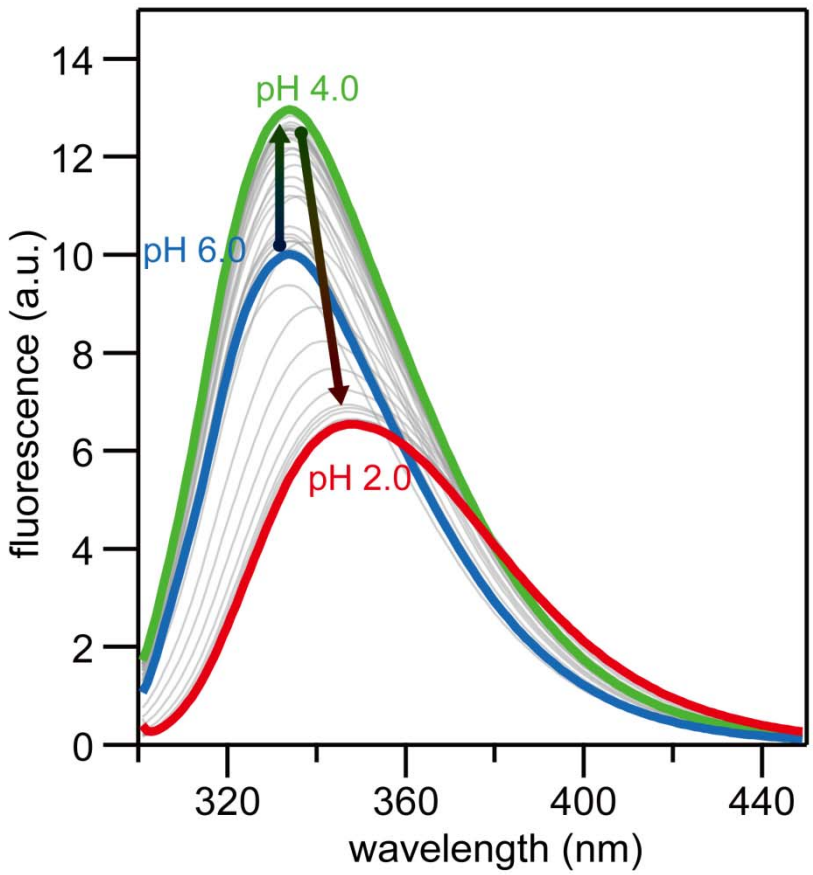


Figure III-8: pH Dependence in fluorescence spectra of hapoMb (in 12 mM sodium citrate buffer at pH 7.1–2.2 or in HCl at pH 2.0) at 8°C, associated with the pH-induced unfolding transition. The fluorescence spectra at **pH 6.0**, **4.0** and **2.0** are shown in **blue**, **green** and **red**, respectively. The fluorescence intensity is enhanced as pH decreases from 6.0 to 4.0, and then quenched as pH decreases from 4.0 to 2.0. The λ_{max} value is red-shifted as pH decreases; **333 nm at pH 6.0**, **334 nm at pH 4.0** and **348 nm at pH 2.0**.

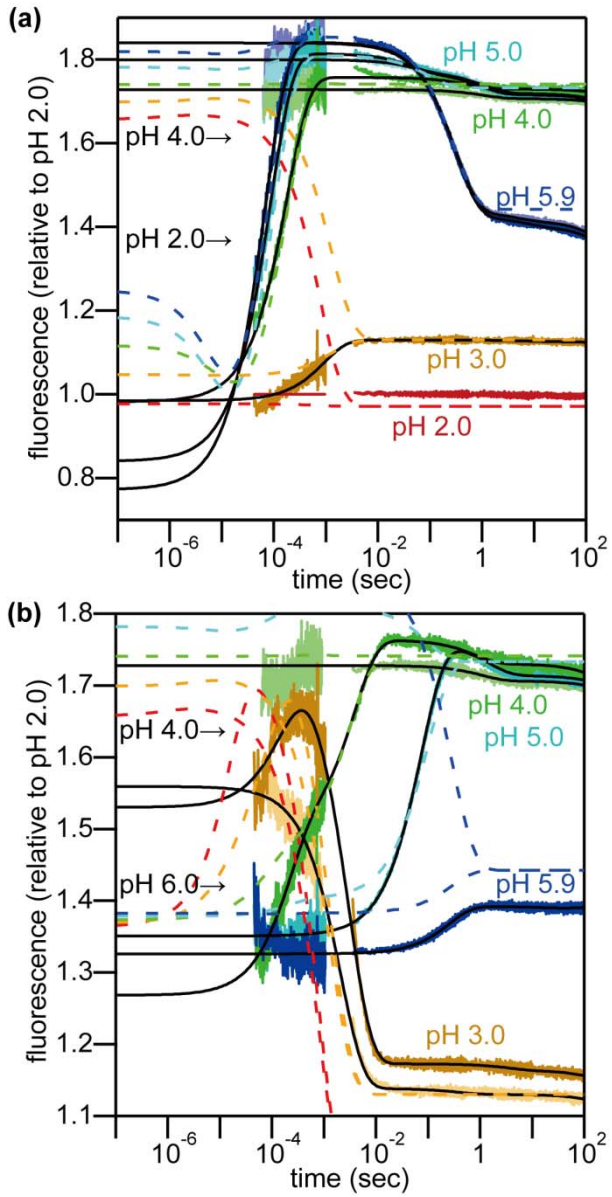


Figure III-9: (a) The folding kinetics at various pH and 8°C measured by the CF and SF methods. The traces of the reactions initiated at pH 2.0 are shown by **blue (at pH 5.9)**, **cyan (at pH 5.0)**, **green (at pH 4.0)**, **ocher (at pH 3.0)** and **red (at pH 2.0)** solid lines. The traces of the reactions initiated at pH 4.0 are shown by **light blue (at pH 5.9)**, **light cyan (at pH 5.0)** and **light green (at pH 4.0)** solid lines. Black lines indicate the result of curves fitting by an exponential function. The dashed line show predicted traced by quantitative modeling.

(b) The unfolding kinetics at various pH and 8°C measured by the CF and SF methods. The traces of the reactions initiated at pH 6.0 are shown by **blue (at pH 5.9)**, **cyan (at pH 5.0)**, **green (at pH 4.0)** and **ocher (at pH 3.0)** solid lines. The traces of the reactions initiated at pH 4.0 are shown by **light green (at pH 4.0)** and **yellow (at pH 3.0)** solid lines. Black lines indicate the result of curves fitting by an exponential function. The dashed line show predicted traced by quantitative modeling.

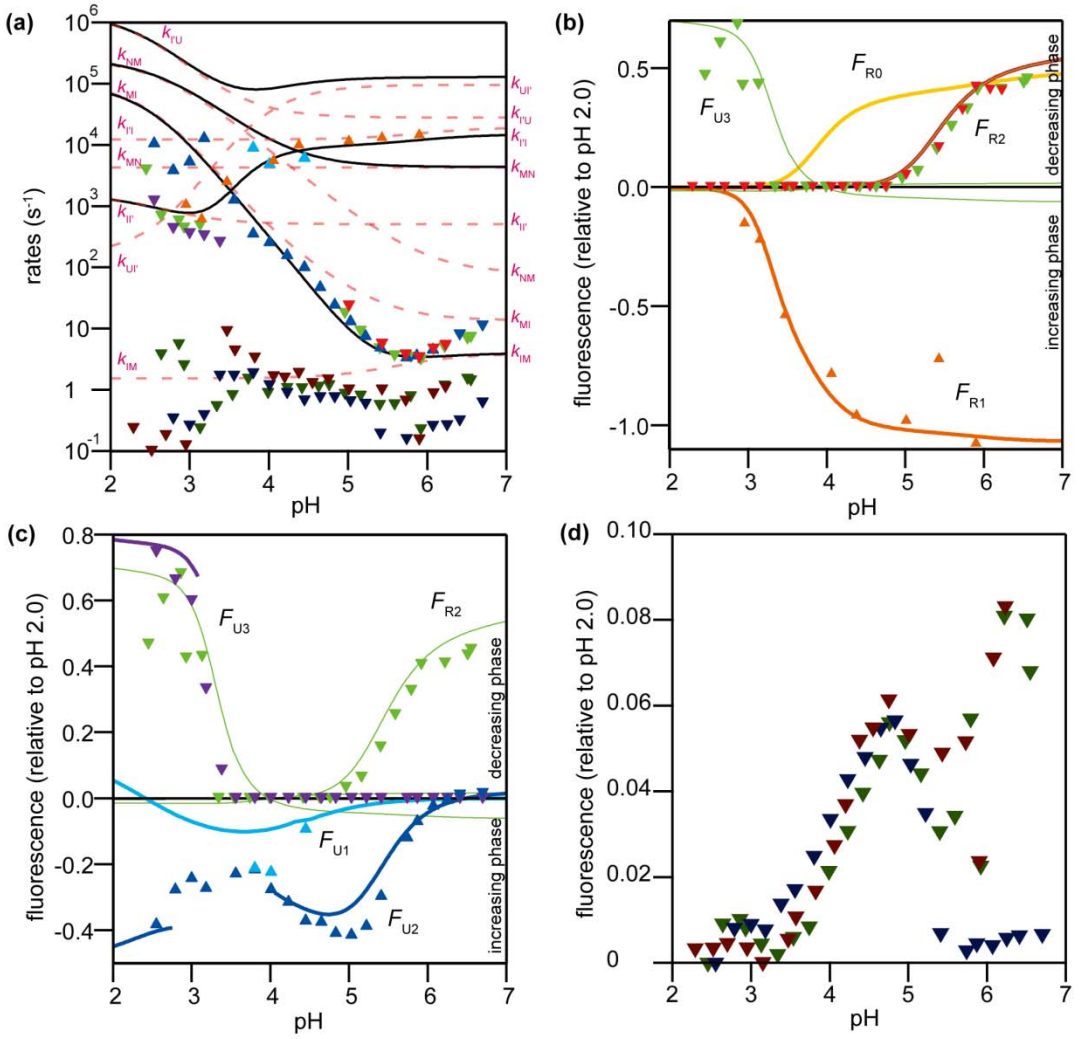


Figure III-10: (a) pH dependence of the rate constants of folding/unfolding reactions obtained by the CF and SF experiments. The upper and lower triangles show increasing and decreasing phases, respectively. Color codes: **R1 (orange)**, **R2 initiated at pH 2.0 (red)**, **U1 (cyan)**, **U2 (blue)**, **U3 initiated at pH 6.0 (purple)**, **folding/unfolding initiated at pH 4.0 (green)**, **D initiated at pH 2.0 (dark red)**, **4.0 (dark green)** and **6.0 (dark blue)**. Black solid lines and red dashed lines show the apparent rate constants and the elementary rate constants, respectively, predicted by quantitative modeling. (b) The amplitudes obtained by the folding experiments initiated at pH 2.0 and at pH 4.0. Lines show the predicted amplitudes by quantitative modeling. Color codes: F_{R0} **(yellow)**, F_{R1} **(orange)**, F_{R2} **initiated at pH 2.0 (red)**, **the amplitude of reactions initiated at pH 4.0 (green)**. (c) The amplitudes obtained by the unfolding experiments initiated at pH 6.0 and at pH 4.0. Lines show the predicted amplitudes by quantitative modeling. Color codes: F_{U1} **(cyan)**, F_{U2} **(blue)**, F_{U3} **initiated at pH 6.0 (purple)**, **the amplitude of reactions initiated at pH 4.0 (green)**. (d) The amplitudes of phase D initiated at **pH 2.0 (dark red)**, **4.0 (dark green) and 6.0 (dark blue)**.

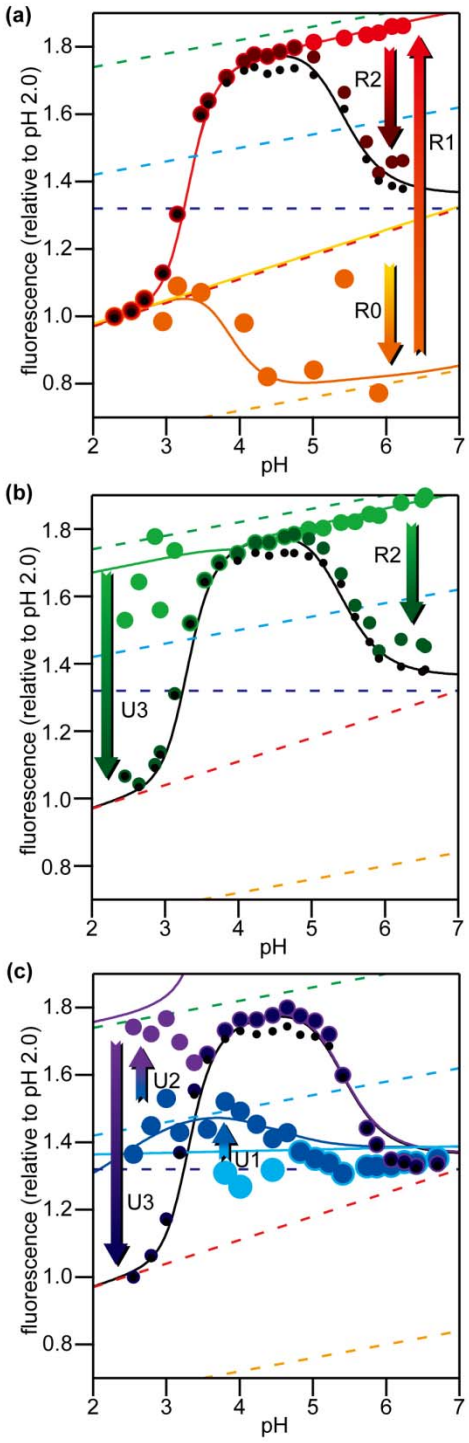


Figure III-11: (a) pH Dependence of cumulative amplitudes of folding initiated at pH 2.0. Dashed lines show the fluorescence intensities of **U (red)**, **I′ (orange)**, **I (green)**, **M (cyan)** and **N (blue)** used for the quantitative modeling. Solid lines show the cumulative amplitudes predicted by quantitative modeling. **The black circles and black solid line** show fluorescence at equilibrium and that predicted by quantitative modeling, F_{eq}. **The dark red circles** show **cumulative amplitudes of F_{eq} and F_D. The red circles and red solid line** show F_{0R2} obtained by the experiment and the simulation, respectively. **The orange circles and orange solid line** show F_{0R1} obtained by the experiment and the simulation, respectively. **The yellow solid line** shows **the fluorescence intensity at the zero time** predicted by the quantitative modeling.

(b) pH Dependence of cumulative amplitudes of folding/unfolding initiated at pH 4.0. Dashed lines show the baseline of each state with the same color codes as in (a). Solid lines show the cumulative amplitudes predicted by quantitative modeling. T**he black circles and black solid line** show fluorescence at equilibrium and that predicted by quantitative modeling, F_{eq}. **The dark green circles** show cumulative amplitudes of F_{eq} **and** F_D. **The green circles and green solid line** show **the cumulative amplitude of F_{eq} and the amplitudes corresponding with R2 and U3 phases**.

(c) pH Dependence of cumulative amplitudes of unfolding initiated at pH 6.0. Dashed lines show the baseline of each state with the same color codes as in (a). Solid lines show the cumulative amplitudes predicted by quantitative modeling. T**he black circles and black solid line** show fluorescence at equilibrium and that predicted by quantitative modeling, F_{eq}. **The dark purple circles** show **cumulative amplitudes of F_{eq} and F_D. The light purple circles and light purple solid line** show F_{0U3} obtained by the experiment and the simulation, respectively. **The blue circles and blue solid line** show F_{0U2} obtained by the experiment and the simulation, respectively. **The cyan circles and cyan solid line** show F_{0U1} obtained by the experiment and the simulation, respectively.

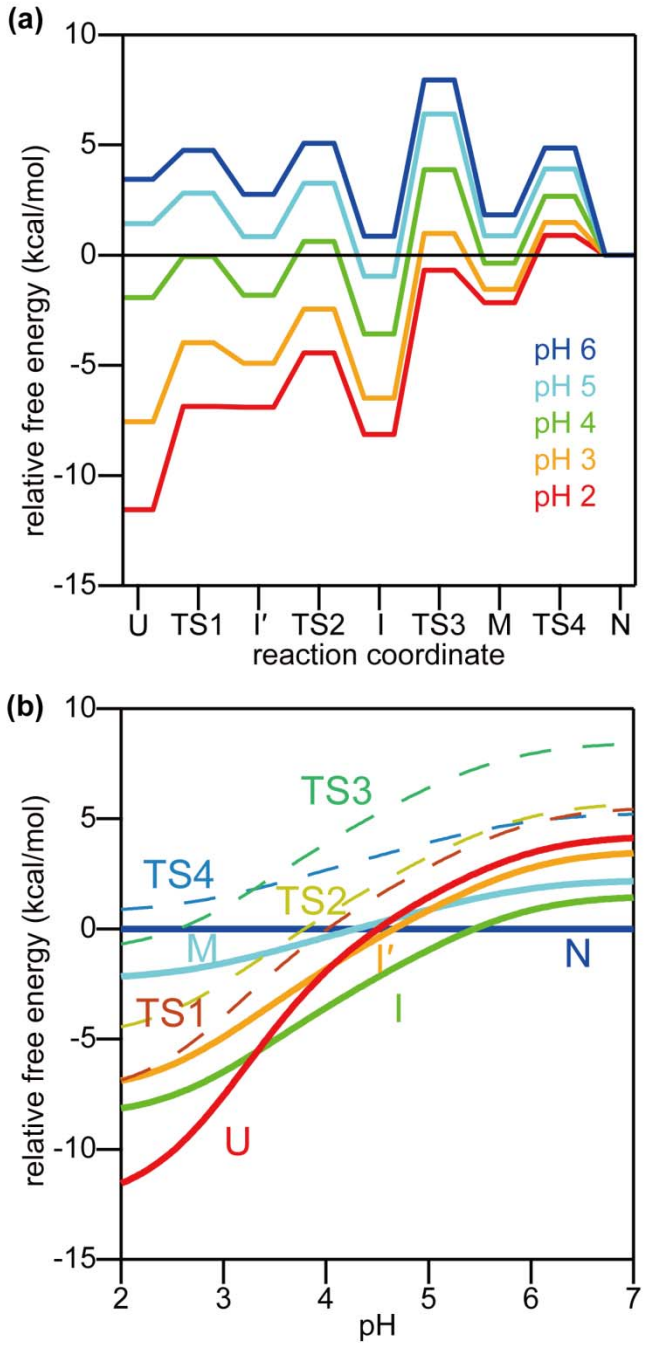


Figure III-12: (a) The free energy diagram of apoMb at representative pH values and 8°C in the absence of urea showing the effects of pH on the energy levels of each state and the transition states predicted by quantitative analysis of the refolding and unfolding kinetics. The color codes are shown in the figure. (b) pH Dependence of the free energy of the each state including the transition states. The color codes are shown in the figure.

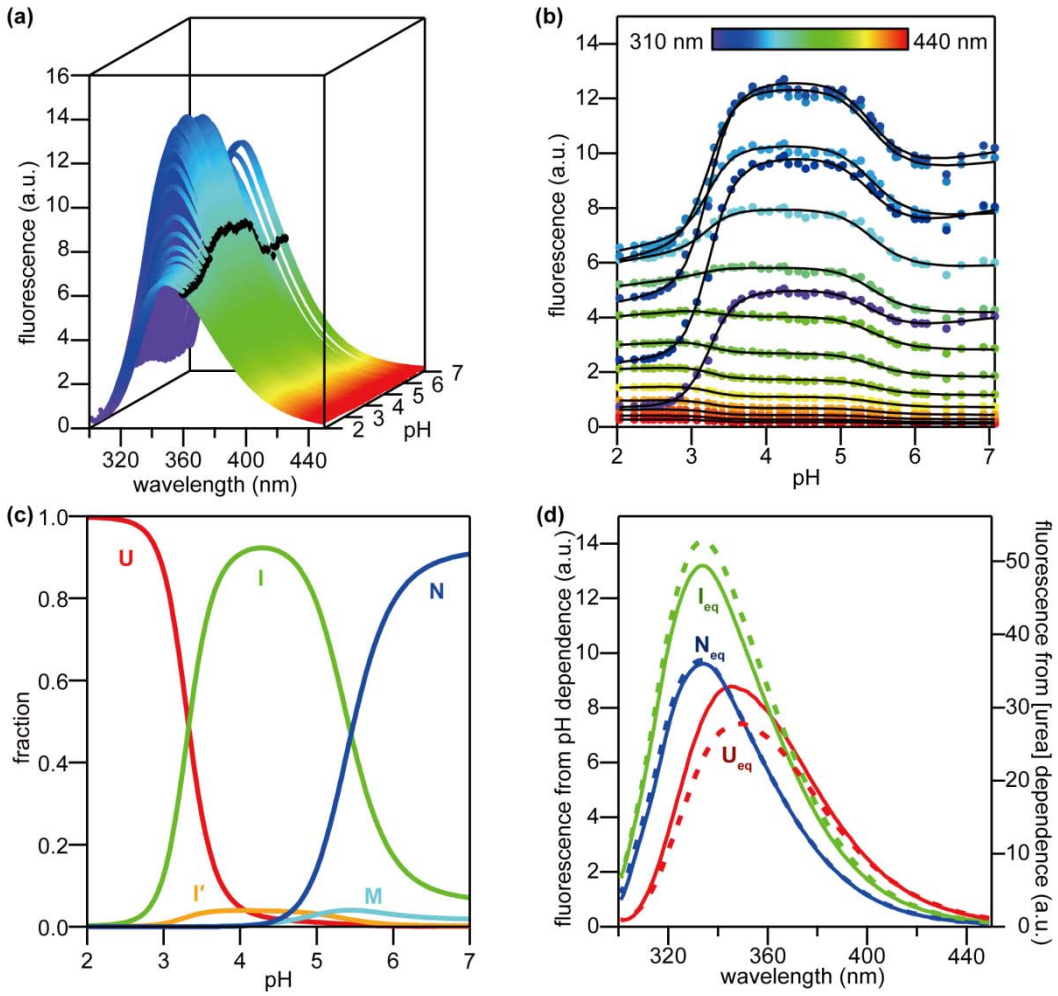


Figure III-13: (a) The collection of fluorescence spectra of hapoMb shown in Figure III-8 ranging from **310** nm to **440** nm as a function of pH at 8°C. **Black** circles represent an unfolding transition curve monitored by the fluorescence emission at **360 nm**. (b) pH-Induced equilibrium unfolding transition curves of hapoMb monitored by fluorescence emission at wavelengths ranging from **310** to **440** nm at 8°C. The solid lines are the unfolding transition curves predicted by the fitting based on the fraction shown in (c). The transition curves in (b) are obtained by transposing the results from (a). Colors represent the emission wavelength. (c) Populations of hapoMb as a function of pH predicted by the quantitative modeling. The lines show the populations of **N (blue)**, **M (cyan)**, **I (green)**, **I′ (orange)**, and **U (red)** as function of pH predicted on the basis on the kinetic experiments. (d) Fluorescence spectra obtained by curve fitting to a series of the unfolding transition curves. The solid and dashed lines show fluorescence spectra of each state, N_{eq} (a mixture of N and M; **blue**), I_{eq} (consists of I; **green**) and U_{eq} (a mixture of U and I′; **red**), at pH 6.0 in the absence of urea estimated by the pH- and urea-induced equilibrium unfolding data, respectively.

CHAPTER IV

General Conclusions

For the purpose of understanding folding mechanisms of globular proteins, the author studied folding kinetics of two proteins, staphylococcal nuclease (SNase) and horse skeletal muscle apomyoglobn (hapoMb), especially focusing on the early events in this thesis. In CHAPTER II, the development of tertiary structure during the SNase folding was studied with variants with a single fluorescence resonance energy transfer pair using a continuous-flow fluorescence device. In CHAPTER III, the relationship between equilibrium and kinetic folding intermediates of hapoMb was studied from a kinetic viewpoint.

SIGNIFICANCE AND ROLE OF FOLDING INTERMEDIATES

The partially folded intermediates have been regarded as being productive for unstructured polypeptide chains to fold into the specific native structure. However, several groups raised a question on the significance of the folding intermediates, a possibility that the folding intermediates are caused by kinetic traps including misfolding or aggregation, which may correspond to off-pathway intermediates. In fact, several experimental studies indicated the transient accumulation of misfolded and aggregated species during the folding processes.[1,2] The finding of two-state folding proteins without accumulation of detectable folding intermediates during the folding further supported the negative idea on the role of the folding intermediates.[3,4] In this thesis, the author examined possible off-pathway folding schemes of hapoMb in addition to the on-pathway five-state sequential folding scheme by the quantitative modeling. The results from CHAPTER III, along with the structural properties reported previously, indicated that hapoMb folding under the native, acid-unfolding and urea-unfolding conditions is consistent with the folding scheme involving on-pathway intermediates.

Recent theoretical and experimental studies suggest that even the on-pathway intermediates could include non-native structures.[5,6] Bacterial immunity protein, Im7, accumulates an on-pathway intermediates during the folding. A mutational approach revealed that the folding intermediate not only contains native-like interactions but also non-native interactions.[5] On-pathway folding intermediates but with non-native structure were also indicated by a theoretical study of a three-dimensional cubic lattice with two types of amino acid residues, i.e., hydrophobic and polar residues. The results proposed that the role of the intermediate is not to act as an obstacle but to restrict the folding route.[6] The heterogeneous chain condensation during SNase folding observed in the present study indicates that the early events in SNase folding are driven by site-specific native-like tertiary interactions. However, it might be necessary for our further understanding of the protein folding mechanisms to detect non-native contacts during protein folding as well as the native-like contacts because even the non-native intermediates are expected to be used as a sort of landmark rather than obstacles in folding processes. Thus, development of novel techniques monitoring both the native-like and non-native contacts is needed to elucidate the protein folding mechanisms in more detail.

CONCLUDING REMAERKS

Finally, the author would like to discuss the recent computational works. David E. Shaw and his coworkers had developed a specialized supercomputer along with modified force fields to accelerate the execution of atomistic molecular dynamics simulations. Their recent studies using molecular dynamics simulation observed the folding/unfolding and dynamics over periods of 100 μs – 1 ms for small proteins folded within a millisecond of folding.[7,8] The combined knowledge obtained from the experiments and simulations allows us to represent the folding/dynamics of small proteins at the atomic level.[7-14]

On the other hand, larger proteins fold to the native state in a much more complicated manner on a much longer time scale (typically 1–10 s) than the small proteins, and there still remain many issues to be understood.[15,16] However, the larger proteins, especially those with more than ~100 amino acid residues, often accumulate folding intermediates within a few millisecond of folding.[17] These intermediates are regarded as playing an important role in guiding the unstructured polypeptide chains toward the specific native structures because of the native-like features.[17] The combination of the recent development of the experimental techniques such as the continuous-flow methods and the advance in the computational techniques will allow us to reach understanding of the early stage of protein folding in detail. Thus, the use of the continuous-flow methods will open a new age of the protein folding studies.

In summary, the author studied the folding of SNase and hapoMb using continuous-flow devices. The proteins fold to their native states through on-pathway intermediates, indicating that the intermediates act as stepping stones in finding their native conformation. Continuous-flow methods have also been applied for folding of some globular proteins. However, information on early kinetic events of protein folding is not enough yet to lead general conclusions of significance of folding intermediates. The author believes that the developed continuous-flow techniques should be helpful to elucidate the significance of folding intermediates, and hopes that the results and discussion in this thesis will contribute to the better understanding of protein folding mechanisms.

REFERENCES

1. Eliezer D, Chiba K, Tsuruta H, Doniach S, Hodgson KO, Kihara H (1993) Evidence of an associative intermediate on the myoglobin refolding pathway. Biophys J **65**:912-917.
2. Yeh S-R, L. RD (1998) Folding intermediates in cytochrome *c*. Nat Struct Biol **5**:222-228.
3. Jackson SE (1998) How do small single-domain proteins fold? Fold Des **3**:R81-R91.
4. Sosnick TR, Mayne L, Hiller R, Englander SW (1994) The barriers in protein folding. Nat Struct Biol **1**:149-156.
5. Capaldi AP, Kleanthous C, Radford SE (2002) Im7 folding mechanism: misfolding on a path to the native state. Nat Struct Biol **9**:209-216.

6. Chikenji G, Kikuchi M (2000) What is the role of non-native intermediates of beta-lactoglobulin in protein folding? Proc Natl Acad Sci USA **97**:14273-14277.
7. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y et al. (2010) Atomic-level characterization of the structural dynamics of proteins. Science **330**:341-346.
8. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. Science **334**:517-520.
9. Bowman GR, Voelz VA, Pande VS (2011) Atomistic folding simulations of the five-helix bundle protein λ_{6-85}. J Am Chem Soc **133**:664-667.
10. Liu F, Du D, Fuller AA, Davoren JE, Wipf P, Kelly JW, Gruebele M (2008) An experimental survey of the transition between two-state and downhill protein folding scenarios. Proc Natl Acad Sci USA **105**:2369-2374.
11. Prigozhin MB, Gruebele M (2011) The fast and the slow: folding and trapping of lambda6-85. J Am Chem Soc **133**:19338-19341.
12. Koulgi S, Sonavane U, Joshi R (2010) Insights into the folding pathway of the Engrailed Homeodomain protein using replica exchange molecular dynamics simulations. J Mol Graph Model **29**:481-491.
13. Religa TL, Johnson CM, Vu DM, Brewer SH, Dyer RB, Fersht AR (2007) The helix-turn-helix motif as an ultrafast independently folding domain: the pathway of folding of Engrailed homeodomain. Proc Natl Acad Sci USA **104**:9272-9277.
14. Jager M, Zhang Y, Bieschke J, Nguyen H, Dendle M, Bowman ME, Noel JP, Gruebele M, Kelly JW (2006) Structure-function-folding relationship in a WW domain. Proc Natl Acad Sci USA **103**:10648-10653.
15. Baldwin RL, Rose GD (1999) Is protein folding hierarchic? I. Local structure and peptide folding. Trend Biochem Sci **24**:26-33.
16. Baldwin RL, Rose GD (1999) Is protein folding hierarchic? II. Folding intermediates and transition states. Trends Biochem Sci **24**:77-83.
17. Arai M, Kuwajima K (2000) Role of the molten globule state in protein folding. Adv Protein Chem **53**:209-282.

APPENDICES

Appendix A: Simulation of (un)folding kinetics by matrix method

The time courses of a series of reactants and products are obtained by resolving the series of rate equations in differential forms when a kinetic scheme along with the elementary rate constants for each reaction step and an initial condition are given. However, it is difficult to analytically solve the rate equations for complex kinetic systems such as multistate folding reactions as a function of urea concentration or pH. Thus, we used a matrix approach to numerically solve a series of kinetic equations of the (un)folding reactions. Unless association and dissociation of oligomeric complexes are involved in the reaction, i.e., the reaction is limited to the interconversion of monomer molecules, the rate equation is given by a set of first-order linear differential equations with constant coefficients, where the unknown functions are the concentration of *i* state, X_i, and the constant coefficients consists of the elementary rate constants of the *i* state ⇌ *j* state transition (k_{ij}). If there are *n* species, the general set of the kinetic equations are;

$$\begin{aligned}
\frac{dX_1}{dt} &= -(\sum_{i=1}^{n} k_{1i})X_1 + k_{21}X_2 + k_{31}X_3 + \cdots + k_{n1}X_n \\
\frac{dX_2}{dt} &= k_{12}X_1 - (\sum_{i=1}^{n} k_{2i})X_2 + k_{32}X_3 + \cdots + k_{n2}X_n \\
&\vdots \\
\frac{dX_n}{dt} &= k_{1n}X_1 + k_{2n}X_2 + k_{3n}X_3 + \cdots - (\sum_{i=1}^{n} k_{ni})X_n
\end{aligned} \qquad \text{Eq. A1}$$

which is also represented by using the rate matrix as follows:

$$\frac{d\mathbf{X}}{dt} = \mathbf{K}\mathbf{X} \qquad \text{Eq. A2}$$

where **X** is the state vector

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \qquad \text{Eq. A3}$$

and **K** is the rate matrix

$$\mathbf{K} = \begin{bmatrix} -\sum_{i=1}^{n} k_{1i} & k_{21} & \cdots & k_{n1} \\ k_{12} & -\sum_{i=1}^{n} k_{2i} & \cdots & k_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ k_{1n} & k_{2n} & \cdots & -\sum_{i=1}^{n} k_{ni} \end{bmatrix}. \quad \text{Eq. A4}$$

The solution to Eq. A2 is represented as follows:

$$\mathbf{X}(t) = \exp(\mathbf{K}t)\mathbf{X}(0) \quad \text{Eq. A5}$$

where exp(**K***t*) is designated as the state transition matrix:

$$\exp(\mathbf{K}t) = \mathbf{E} + \mathbf{K}t + \frac{(\mathbf{K}t)^2}{2!}\cdots. \quad \text{Eq. A6}$$

By diagonalizing the state transition matrix by using a set of its eigenvectors ($\mathbf{P}_i$), the solution is represented as follows:

$$\mathbf{Y}(t) = \exp(\mathbf{\Lambda}t)\mathbf{Y}(0). \quad \text{Eq. A7}$$

Y (*t*) and exp (*Λt*) are as follows:

$$\mathbf{Y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{bmatrix} = \mathbf{P}^{-1}\mathbf{X}(t) \quad \text{Eq. A8}$$

and

$$\exp(\mathbf{\Lambda}t) = \mathbf{P}^{-1}\exp(\mathbf{K}t)\mathbf{P}, \quad \text{Eq. A9}$$

where **P** is a matrix consisting of the eigenvectors:

$$\mathbf{P} = (\mathbf{P}_1, \mathbf{P}_2, \cdots, \mathbf{P}_n), \quad \text{Eq. A10}$$

and **Λ** is a diagonalized matrix consisting of the eigenvalues, λ_i:

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}. \quad \text{Eq. A11}$$

For a closed system, one of the eigenvalues is equal to zero. Thus, the solution is now represented as follows:

$$\mathbf{X}(t) = \mathbf{P}\begin{bmatrix} \exp(\lambda_1 t) & & 0 \\ & \ddots & \\ 0 & & \exp(\lambda_n t) \end{bmatrix}\mathbf{P}^{-1}\mathbf{X}(0) \qquad \text{Eq. A12}$$

$$= \sum_{i=1}^{n} \mathrm{P}_i y_i(0) \exp(\lambda_i t)$$

The initial concentrations $X_i(0)$ (i = 1, ..., n) are given by experiments. The eigenvalues and eigenvectors are numerically calculated at each urea concentration or pH by using a standard library package for solving the most commonly occurring problems in numerical linear algebra when the rate matrix is given.

The calculation using the rate matrix approach is applied to a (un)folding reaction represented by a five-state sequential folding scheme:

$$\mathrm{U} \underset{k_{\mathrm{I'U}}}{\overset{k_{\mathrm{UI'}}}{\rightleftharpoons}} \mathrm{I'} \underset{k_{\mathrm{II'}}}{\overset{k_{\mathrm{I'I}}}{\rightleftharpoons}} \mathrm{I} \underset{k_{\mathrm{MI}}}{\overset{k_{\mathrm{IM}}}{\rightleftharpoons}} \mathrm{M} \underset{k_{\mathrm{NM}}}{\overset{k_{\mathrm{MN}}}{\rightleftharpoons}} \mathrm{N} \qquad \text{Scheme A1}$$

where U is the unfolded state, N is the native state, and I', I and M are the intermediates. k_{ij} is the elementary rate constant for the transition from i state to j state. The state vector and the rate matrix are represented as follows:

$$\mathbf{X}(t) = \begin{bmatrix} \mathrm{U}(t) \\ \mathrm{I'}(t) \\ \mathrm{I}(t) \\ \mathrm{M}(t) \\ \mathrm{N}(t) \end{bmatrix}, \mathbf{K} = \begin{bmatrix} -k_{\mathrm{UI'}} & k_{\mathrm{I'U}} & 0 & 0 & 0 \\ k_{\mathrm{UI'}} & -k_{\mathrm{I'U}} - k_{\mathrm{I'U}} & k_{\mathrm{II'}} & 0 & 0 \\ 0 & k_{\mathrm{I'U}} & -k_{\mathrm{II'}} - k_{\mathrm{IM}} & k_{\mathrm{MI}} & 0 \\ 0 & 0 & k_{\mathrm{IM}} & -k_{\mathrm{MI}} - k_{\mathrm{MN}} & k_{\mathrm{NM}} \\ 0 & 0 & 0 & k_{\mathrm{MN}} & -k_{\mathrm{NM}} \end{bmatrix} \qquad \text{Eq. A13}$$

The rank of this rate matrix is less than five because this system is closed. Thus, the number of eigenvalues, which correspond to the rate constants, are less than or equal to four. An eigenvalue (λ) is obtained by solving the characteristic equation of **K**.

Appendix B: Urea-induced folding/unfolding kinetics

The elementary rate constant of (un)folding reactions, k_{ij}, is approximated to linearly depend on the denaturant concentration, c.

$$\ln k_{ij}(c) = \ln k_{ij}^{\mathrm{H2O}} + \frac{m_{ij}^{\ddagger} c}{RT}, \qquad \text{Eq. B1}$$

where k_{ij} (c) is the elementary rate constant for the $i \rightarrow j$ transition at a denaturant concentration, c, $k^{\mathrm{H2O}}{}_{ij}$ is the elementary rate constant in the absence of denaturant, $m^{\ddagger}{}_{ij}$ is the coefficient, and R and T are the gas constant and the absolute temperature, respectively. For folding reactions, $m^{\ddagger}{}_{ij}$ is in most cases negative reflecting the fact that the folding becomes slower as the denaturant concentration increases. On the other hand, $m^{\ddagger}{}_{ij}$ is positive reflecting the fact that the unfolding becomes faster as the denaturant concentration increases. $m^{\ddagger}{}_{ij}$ also represents the relative change in the accessible surface area between i state to the transition state between i and j states. The free energy difference between the i state and the transition state, $\Delta G^{\ddagger}{}_{ij}$, is calculated according to the following equation:

$$\Delta G_{ij}^{\ddagger} = -RT \ln \frac{k_{ij}}{A}, \qquad \text{Eq. B2}$$

where A is the preexponential factor, which corresponds to the elementary rate constant without the kinetic barrier.

For a simple (un)folding without detectable intermediates during (un)folding, the (un)folding scheme is represented as follows:

$$\mathrm{U} \underset{k_{\mathrm{NU}}}{\overset{k_{\mathrm{UN}}}{\rightleftharpoons}} \mathrm{N}. \qquad \text{Scheme B1}$$

In this scheme, U and N represent the unfolded and the native states, respectively. The rate constant observed in the (un)folding kinetics is represented by the sum of the two elementary rate constants k_{UN} and k_{NU}.

$$\lambda(c) = k_{\mathrm{UN}}(c) + k_{\mathrm{NU}}(c) \qquad \text{Eq. B3}$$

Under strongly native conditions, which in most cases corresponds to low urea concentrations, $\lambda(c)$ is approximated by $k_{\mathrm{UN}}(c)$ because k_{UN} is much larger than k_{NU}.

$$\ln \lambda(c) \approx \ln k_{\mathrm{UN}}^{\mathrm{H2O}} + \frac{m_{\mathrm{UN}}^{\ddagger} c}{RT} \qquad \text{Eq. B4}$$

In contrast, under strongly denaturing conditions, which in most cases corresponds to high denaturant concentrations, $\lambda(c)$ is approximated by $k_{\mathrm{NU}}(c)$ because $k_{\mathrm{NU}}(c)$ is much larger than $k_{\mathrm{UN}}(c)$.

$$\ln \lambda(c) \approx \ln k_{\mathrm{NU}}^{\mathrm{H2O}} + \frac{m_{\mathrm{NU}}^{\ddagger} c}{RT} \qquad \text{Eq. B5}$$

Thus, $\ln\lambda(c)$ undergoes the minimum at a denaturant concentration, c_{min}, which is represented as follows:

$$c_{\min} = \frac{RT}{m^{\ddagger\ddagger\ddagger}_{UN} - m_{NU}} \ln\left(-\frac{m^{\ddagger}_{NU} k^{H2O}_{NU}}{m_{UN} k^{H2O}_{UN}} \right). \qquad \text{Eq. B6}$$

The equilibrium constant of (un)folding reactions in Scheme B1 is represented by

$$K_{NU} = k_{NU} / k_{UN} \qquad \text{Eq. B7}$$

The urea concentration of the midpoint of the equilibrium unfolding transition, c_m, is represented as follows:

$$c_m = \frac{RT}{m^{\ddagger\ddagger}_{UN} - m_{NU}} \ln\left(\frac{k^{H2O}_{NU}}{k^{H2O}_{UN}} \right). \qquad \text{Eq. B8}$$

Thus, c_{min} is close to c_m because the difference between $m^{\ddagger}_{NU}$ and $m^{\ddagger}_{UN}$ is in general within several fold at most. Taken these together, a plot of c *vs.* logλ (c) exhibits a "V-shape", or chevron as a function of c, and thus is sometimes designated as the chevron plot.

For proteins with accumulation of a folding intermediate, the kinetic behavior can be more complicated than that of the two-state proteins. The folding scheme for the three-state proteins is represented as follows:

$$\mathrm{U} \underset{k_{IU}}{\overset{k_{UI}}{\rightleftharpoons}} \mathrm{I} \underset{k_{NI}}{\overset{k_{IN}}{\rightleftharpoons}} \mathrm{N} \qquad \text{Scheme B2}$$

As described in Appendix A, the folding reactions represented by Scheme B2 exhibit two rate constants, λ_1 and λ_2 ($\lambda_1 > \lambda_2$), which are analytically obtained by solving the rate equation.

$$\lambda_{1,2} = \frac{1}{2}\left(k_{UI} + k_{IU} + k_{IN} + k_{NI} \pm \sqrt{\left(k_{UI} + k_{IU} + k_{IN} + k_{NI}\right)^2 - 4\left(k_{UI}k_{IN} + k_{UI}k_{NI} + k_{IU}k_{NI}\right)} \right) \qquad \text{Eq. B9}$$

Here it is assumed that the U ⇌ I interconversion is much faster than the I ⇌ N interconversion, i.e., k_{UI} and k_{IU} >> k_{IN} and k_{NI}. Under the assumption, λ_1 is reduced as:

$$\lambda_1 = k_{UI} + k_{IU}, \qquad \text{Eq. B10}$$

which demonstrates that the U ⇌ I interconversion is well represented as a two-state folding reaction since rapid preequilibrium is achieved in the U ⇌ I interconversion, which is faster than the rate-limiting step (λ_2). In addition, λ_2 are reduced under native conditions as:

$$\lambda_2 = \frac{k_{UI}}{k_{UI} + k_{IU}} k_{IN} + k_{NI}. \qquad \text{Eq. B11}$$

The denaturant concentration dependence of λ_2 exhibits a curvature in the folding region if the midpoint concentration of the U ⇌ I interconversion, c_{mUI}, is lower than the midpoint concentration of the I ⇌ N interconversion, c_{mIN}. At a low denaturant concentration ($c < c_{mUI}$), where k_{UI} >> k_{IU}, λ_2 ~ k_{IN} + k_{NI}. As denaturant concentration increases, k_{UI} approaches k_{IU}, and thus λ_2 ~ 1/(1+K_{IU})×k_{IN} + k_{NI}. For example, λ_2 ~ 1/2×k_{IN} + k_{NI} at c ~ c_{mUI}. At a denaturant concentration higher than c_{mUI}, λ_2 ~ 1/K_{IU}×k_{IN} + k_{NI}. Since ln(k_{IN}) is linearly dependent on the denaturant concentration, ln(λ_2) deviates

from linear denaturant concentration dependence expected from ln(k_{IN}) by a factor $k_{UI}/(k_{UI} + k_{IU})$, which gives rise to a curvature in λ_2. Similar curvatures can be observed in the unfolding kinetics under denaturing conditions.

The urea-induced (un)folding kinetics of horse skeletal muscle apomyoglobin (hapoMb) observed two phases (phases 1 and 2 in CHAPTER III). The urea concentration dependence of log(λ_1) exhibits a curvature at low urea concentrations (~0.8 M). Since phase 1 corresponds to the formation/unfolding of a folding intermediate, I, the curvature in the folding limb of λ_1 indicates the accumulation of at least one additional folding intermediate (I') within the dead time of the continuous-flow measurements. The other curvature is found in the unfolding limb of log(λ_2) at ~3.3 M urea, which is attributed to the accumulation of the third intermediate, M. Along with the results obtained by the urea-induced (un)folding kinetics, the folding scheme of hapoMb requires five states including the native (N) and the unfolded (U) states.

$$\mathrm{U} \underset{k_{\mathrm{I'U}}}{\overset{k_{\mathrm{UI'}}}{\rightleftharpoons}} \mathrm{I'} \underset{k_{\mathrm{II'}}}{\overset{k_{\mathrm{I'I}}}{\rightleftharpoons}} \mathrm{I} \underset{k_{\mathrm{MI}}}{\overset{k_{\mathrm{IM}}}{\rightleftharpoons}} \mathrm{M} \underset{k_{\mathrm{NM}}}{\overset{k_{\mathrm{MN}}}{\rightleftharpoons}} \mathrm{N} \qquad \text{Scheme B3}$$

Appendix C: pH-Induced folding/unfolding kinetics

pH Dependence of equilibrium unfolding and folding/unfolding kinetics is associated with protonation and deprotonation of ionizable groups. For the purpose of making points clear, the (un)folding without any detectable intermediate, i.e., the two-state model is considered first. It is also assumed that only a single ionizable group is associated with the (un)folding for simplicity. In this model, both the native (N) and the unfolded (U) states can be protonated with specific dissociation constants, K_a^N and K_a^U, respectively. The pK_a in N (pK_a^N) is smaller than that in U (pK_a^U) when the ionizable group is buried inside the molecule in the native state because protons can hardly access the buried ionizable group in the native state in comparison with the solvent-exposed ionizable group in the unfolded state. Thus, the (un)folding scheme is represented as:

$$\begin{array}{ccccc} & & K_{NU}^0 & & \\ & U+H^+ & \rightleftharpoons & N+H^+ & \\ K_a^U & \Updownarrow & & \Updownarrow & K_a^N \\ & UH^+ & \rightleftharpoons & NH^+ & \end{array}$$

Scheme C1

In this scheme, N and U are the native and the unfolded states without protonation (deprotonated native and unfolded states; N is assumed to be more stable than U) whereas NH^+ and UH^+ are the native and the unfolded states with the ionizable group protonated, respectively. K^0_{NU} is the equilibrium constants of the N ⇌ U unfolding transitions. NH^+ ⇌ N + H^+ and UH^+ ⇌ U + H^+ (vertical direction) represent the equilibrium of the protonation/deprotonation in the native and the unfolded states, respectively. On the other hand, U ⇌ N and UH^+ ⇌ NH^+ represent the equilibrium (un)folding of the deprotonated and protonated species, respectively. Here, we do not consider the base-induced denaturation. The equilibrium constant of the unfolding transition, K_{UN}, is defined as follows:

$$K_{NU} = \frac{[U]+[UH^+]}{[N]+[NH^+]}. \qquad \text{Eq. C1}$$

Meanwhile, the dissociation constants of protonation/deprotonation of the ionizable group in the native and the unfolded states are represented as:

$$K_a^U = \frac{[U][H^+]}{[UH^+]}$$
$$K_a^N = \frac{[N][H^+]}{[NH^+]} \qquad \text{Eq. C2}$$

By using Eq. C2, K_{NU} is represented as follows:

$$K_{NU} = \frac{[U]}{[N]}\frac{1+[H^+]/K_a^U}{1+[H^+]/K_a^N}, \qquad \text{Eq. C3}$$

which is rewritten by using K^0_{NU} = [U]/[N], pK_a^U, pK_a^N, and pH as follows:

$$K_{\mathrm{NU}} = K^{0}_{\mathrm{NU}} \frac{1+10^{(\mathrm{p}K_a^{\mathrm{U}}-\mathrm{pH})}}{1+10^{(\mathrm{p}K_a^{\mathrm{N}}-\mathrm{pH})}}. \qquad \text{Eq. C4}$$

K^0_{NU} approximates the equilibrium constant of the unfolding transition at high pH, where deprotonated species are dominant. The equilibrium unfolding is dominated by the U ⇌ N transition at pH sufficiently higher than $\mathrm{p}K_{\mathrm{a}}^{\mathrm{N}}$ and $\mathrm{p}K_{\mathrm{a}}^{\mathrm{U}}$, for example, at an alkaline pH, whereas the equilibrium unfolding is dominated by the UH^+ ⇌ NH^+ transition at pH sufficiently lower than $\mathrm{p}K_{\mathrm{a}}^{\mathrm{N}}$ and $\mathrm{p}K_{\mathrm{a}}^{\mathrm{U}}$, for example, at an acidic pH. This makes sense when taking into considerations that the $\mathrm{p}K_{\mathrm{a}}$ values correspond to the pH values where the ratio of the protonated form and the deprotonated form is 1:1. K_{NU} can be written in a different way by taking a reference pH ($\mathrm{pH}_{\mathrm{ref}}$) as:

$$K_{\mathrm{NU}} = K_{\mathrm{NU}}^{\mathrm{pH}_{\mathrm{ref}}} \frac{1+10^{(\mathrm{p}K_a^{\mathrm{U}}-\mathrm{pH})}}{1+10^{(\mathrm{p}K_a^{\mathrm{N}}-\mathrm{pH})}} \frac{1+10^{(\mathrm{p}K_a^{\mathrm{N}}-\mathrm{pH}_{\mathrm{ref}})}}{1+10^{(\mathrm{p}K_a^{\mathrm{U}}-\mathrm{pH}_{\mathrm{ref}})}} \qquad \text{Eq. C5}$$

where $K^{\mathrm{pHref}}_{\mathrm{NU}}$ is the equilibrium constant of the unfolding transition at a reference pH ($\mathrm{pH}_{\mathrm{ref}}$). The Gibbs free energy difference between the unfolded state and the native state is:

$$\Delta G_{\mathrm{NU}} = -RT \ln K_{\mathrm{NU}} = \Delta G^{0}_{\mathrm{NU}} - RT \ln \frac{1+10^{(\mathrm{p}K_a^{\mathrm{U}}-\mathrm{pH})}}{1+10^{(\mathrm{p}K_a^{\mathrm{N}}-\mathrm{pH})}} \qquad \text{Eq. C6}$$

where ΔG^0_{NU} is the Gibbs free energy difference between N and U, R and T are the gas constant and the absolute temperature, respectively. Based on Eq. C5, Eq. C6 can be written in a different way by taking a reference pH ($\mathrm{pH}_{\mathrm{ref}}$) as follows:

$$\Delta G_{\mathrm{NU}} = \Delta G_{\mathrm{NU}}^{\mathrm{pH}_{\mathrm{ref}}} - RT \ln \left[\frac{1+10^{(\mathrm{p}K_a^{\mathrm{U}}-\mathrm{pH})}}{1+10^{(\mathrm{p}K_a^{\mathrm{N}}-\mathrm{pH})}} \frac{1+10^{(\mathrm{p}K_a^{\mathrm{N}}-\mathrm{pH}_{\mathrm{ref}})}}{1+10^{(\mathrm{p}K_a^{\mathrm{U}}-\mathrm{pH}_{\mathrm{ref}})}} \right] \qquad \text{Eq. C7}$$

where $\Delta G^{\mathrm{pHref}}_{\mathrm{NU}}$ is the Gibbs free energy difference at the reference pH, $\mathrm{pH}_{\mathrm{ref}}$. The pH dependence of the activation Gibbs free energy is obtained in a similar way. According to the transition theory, the equilibrium is assumed between the transition state (‡) and the native or unfolded state (N ⇌ ‡ ⇌ U). Therefore, the free energy differences between the native state and the transition state ($\Delta G^{‡}_{\mathrm{NU}}$) and between the unfolded state and the transition state ($\Delta G^{‡}_{\mathrm{UN}}$) are represented as follows:

$$\Delta G^{‡}_{\mathrm{NU}} = \Delta G_{\mathrm{NU}}^{‡\mathrm{pH}_{\mathrm{ref}}} - RT \ln \left[\frac{1+10^{(\mathrm{p}K_a^{‡}-\mathrm{pH})}}{1+10^{(\mathrm{p}K_a^{\mathrm{N}}-\mathrm{pH})}} \frac{1+10^{(\mathrm{p}K_a^{\mathrm{N}}-\mathrm{pH}_{\mathrm{ref}})}}{1+10^{(\mathrm{p}K_a^{‡}-\mathrm{pH}_{\mathrm{ref}})}} \right] \qquad \text{Eq. C8a}$$

$$\Delta G^{‡}_{\mathrm{U‡N}} = \Delta G_{\mathrm{U}}^{‡\mathrm{pH}_{\mathrm{ref}}} - RT \ln \left[\frac{1+10^{(\mathrm{p}K_a^{‡}-\mathrm{pH})}}{1+10^{(\mathrm{p}K_a^{\mathrm{U}}-\mathrm{pH})}} \frac{1+10^{(\mathrm{p}K_a^{\mathrm{U}}-\mathrm{pH}_{\mathrm{ref}})}}{1+10^{(\mathrm{p}K_a^{‡}-\mathrm{pH}_{\mathrm{ref}})}} \right] \qquad \text{Eq. C8b}$$

where $\Delta G^{‡\mathrm{pHref}}_{\mathrm{NU}}$ and $\Delta G^{‡\mathrm{pHref}}_{\mathrm{UN}}$ are the free energy differences between the native state and the transition state, and between the unfolded state and the transition state at a reference pH ($\mathrm{pH}_{\mathrm{ref}}$), respectively. $\mathrm{p}K_{\mathrm{a}}^{‡}$ is the $\mathrm{p}K_{\mathrm{a}}$ value in the transition state.

The analysis of the unfolding transition represented by a three-state model (with a single ionizable group) is an extension of the analysis of the two-state model and straightforward. The scheme is represented as follows:

$$\begin{array}{ccccc} & & K^0_{\mathrm{IU}} & & K^0_{\mathrm{NI}} & & \\ & \mathrm{U+H^+} & \rightleftharpoons & \mathrm{I+H^+} & \rightleftharpoons & \mathrm{N+H^+} & \\ K^{\mathrm{U}}_a & \Updownarrow & & \Updownarrow K^{\mathrm{I}}_a & & \Updownarrow & K^{\mathrm{N}}_a \\ & \mathrm{UH^+} & \rightleftharpoons & \mathrm{IH^+} & \rightleftharpoons & \mathrm{NH^+} & \end{array}$$

Scheme C2

where K^0_{IU} and K^0_{NI} are the equilibrium constants of the I ⇌ U and N ⇌ I unfolding transitions, respectively. In addition to the parameters found in the two-state model, a K_{a} value for I ($K_{\mathrm{a}}^{\mathrm{I}}$) is required ($\mathrm{p}K_{\mathrm{a}}^{\mathrm{N}} < \mathrm{p}K_{\mathrm{a}}^{\mathrm{I}} < \mathrm{p}K_{\mathrm{a}}^{\mathrm{U}}$).

$$K_a^{\mathrm{I}} = \frac{[\mathrm{I}][\mathrm{H^+}]}{[\mathrm{IH^+}]} \qquad \text{Eq. C9}$$

The equilibrium constants for the I ⇌ U and N ⇌ I transitions (K_{IU} and K_{NI}) are defined in a similar way to Eq. C1.

$$\begin{aligned} K_{\mathrm{IU}} &= \frac{[\mathrm{U}]+[\mathrm{UH^+}]}{[\mathrm{I}]+[\mathrm{IH^+}]} = K^0_{\mathrm{IU}} \frac{1+10^{(\mathrm{p}K_a^{\mathrm{U}}-\mathrm{pH})}}{1+10^{(\mathrm{p}K_a^{\mathrm{I}}-\mathrm{pH})}}, \\ K_{\mathrm{NI}} &= \frac{[\mathrm{I}]+[\mathrm{IH^+}]}{[\mathrm{N}]+[\mathrm{NH^+}]} = K^0_{\mathrm{NI}} \frac{1+10^{(\mathrm{p}K_a^{\mathrm{I}}-\mathrm{pH})}}{1+10^{(\mathrm{p}K_a^{\mathrm{N}}-\mathrm{pH})}} \end{aligned} \qquad \text{Eq. C10}$$

Eq. D10 is rewritten by taking the equilibrium constants at a reference pH as:

$$\begin{aligned} K_{\mathrm{IU}} &= K_{\mathrm{IU}}^{\mathrm{pH_{ref}}} \frac{1+10^{(\mathrm{p}K_a^{\mathrm{U}}-\mathrm{pH})}}{1+10^{(\mathrm{p}K_a^{\mathrm{I}}-\mathrm{pH})}} \frac{1+10^{(\mathrm{p}K_a^{\mathrm{I}}-\mathrm{pH_{ref}})}}{1+10^{(\mathrm{p}K_a^{\mathrm{U}}-\mathrm{pH_{ref}})}}, \\ K_{\mathrm{NI}} &= K_{\mathrm{NI}}^{\mathrm{pH_{ref}}} \frac{1+10^{(\mathrm{p}K_a^{\mathrm{I}}-\mathrm{pH})}}{1+10^{(\mathrm{p}K_a^{\mathrm{N}}-\mathrm{pH})}} \frac{1+10^{(\mathrm{p}K_a^{\mathrm{N}}-\mathrm{pH_{ref}})}}{1+10^{(\mathrm{p}K_a^{\mathrm{I}}-\mathrm{pH_{ref}})}} \end{aligned} \qquad \text{Eq. C11}$$

where $K^{\mathrm{pHref}}_{\mathrm{IU}}$ and $K^{\mathrm{pHref}}_{\mathrm{NI}}$ are the equilibrium constants at a reference pH ($\mathrm{pH_{ref}}$). The Gibbs free energy differences of the I ⇌ U and N ⇌ I unfolding transitions (ΔG_{IU} and ΔG_{NI}, respectively) are represented as follows:

$$\begin{aligned} \Delta G_{\mathrm{IU}} &= -RT \ln K_{\mathrm{IU}} = \Delta G^0_{\mathrm{IU}} - RT \ln \frac{1+10^{(\mathrm{p}K_a^{\mathrm{U}}-\mathrm{pH})}}{1+10^{(\mathrm{p}K_a^{\mathrm{I}}-\mathrm{pH})}}, \\ \Delta G_{\mathrm{NI}} &= -RT \ln K_{\mathrm{NI}} = \Delta G^0_{\mathrm{NI}} - RT \ln \frac{1+10^{(\mathrm{p}K_a^{\mathrm{I}}-\mathrm{pH})}}{1+10^{(\mathrm{p}K_a^{\mathrm{N}}-\mathrm{pH})}} \end{aligned} \qquad \text{Eq. C12}$$

where ΔG^0_{IU} and ΔG^0_{NI} are the Gibbs free energy difference between I and U, and between N and I, respectively. Eq. C12 is rewritten by using the Gibbs free energy differences at a reference pH ($\mathrm{pH_{ref}}$) ($\Delta G^{\mathrm{pHref}}_{\mathrm{IU}}$ and $\Delta G^{\mathrm{pHref}}_{\mathrm{NI}}$) as:

$$\Delta G_{IU} = \Delta G_{IU}^{pH_{ref}} - RT \ln\left[\frac{1+10^{(pK_a^U - pH)}}{1+10^{(pK_a^I - pH)}}\frac{1+10^{(pK_a^I - pH_{ref})}}{1+10^{(pK_a^U - pH_{ref})}}\right],$$
$$\Delta G_{NI} = \Delta G_{NI}^{pH_{ref}} - RT \ln\left[\frac{1+10^{(pK_a^I - pH)}}{1+10^{(pK_a^N - pH)}}\frac{1+10^{(pK_a^N - pH_{ref})}}{1+10^{(pK_a^I - pH_{ref})}}\right] \qquad \text{Eq. C13}$$

The activation Gibbs free energy of the (un)folding reactions is very similar to that for the two-state model. In the three-state model, two transition states (‡1 and ‡2) located between I and U and between N and I, respectively, are found. The pH dependence of the activation free energies for the I → U and N → I transitions ($\Delta G^{\ddagger 1}{}_{IU}$ and $\Delta G^{\ddagger 2}{}_{NI}$) are represented as:

$$\Delta G_{IU}^{\ddagger 1} = \Delta G_{IU}^{\ddagger 1 pH_{ref}} - RT \ln\left[\frac{1+10^{(pK_a^{\ddagger 1} - pH)}}{1+10^{(pK_a^I - pH)}}\frac{1+10^{(pK_a^I - pH_{ref})}}{1+10^{(pK_a^{\ddagger 1} - pH_{ref})}}\right] \qquad \text{Eq. C14a}$$

$$\Delta G_{NI}^{\ddagger 2} = \Delta G_{NI}^{\ddagger 2 pH_{ref}} - RT \ln\left[\frac{1+10^{(pK_a^{\ddagger 2} - pH)}}{1+10^{(pK_a^N - pH)}}\frac{1+10^{(pK_a^N - pH_{ref})}}{1+10^{(pK_a^{\ddagger 2} - pH_{ref})}}\right] \qquad \text{Eq. C14b}$$

where $\Delta G^{\ddagger 1pHref}{}_{IU}$ and $\Delta G^{\ddagger 2pHref}{}_{NI}$ are the activation free energies for the I → U and N → I transitions at a reference pH (pH_{ref}). The derivation of the rest of the possible activation free energies ($\Delta G^{\ddagger 1pHref}{}_{UI}$ and $\Delta G^{\ddagger 2pHref}{}_{IN}$) is straightforward.

The two-state equilibrium unfolding of proteins with two ionizable groups (sites 1 and 2) is represented in the scheme below:

$$\begin{array}{ccccccc}
 & \overset{K_{a2}^U}{\rightleftharpoons} & UH^+ + H^+ & \rightleftharpoons & NH^+ + H^+ & \overset{K_{a2}^N}{\rightleftharpoons} & \\
 & & \Updownarrow K_{a1}^U & & \Updownarrow K_{a1}^N & & \\
H^+UH^+ & & U + 2H^+ & \overset{K_{NU}^0}{\rightleftharpoons} & N + 2H^+ & & H^+NH^+ \rightleftharpoons H^+UH^+ \\
 & \underset{K_{a1}^U}{\rightleftharpoons} & \Updownarrow K_{a2}^U & & \Updownarrow K_{a2}^N & \underset{K_{a1}^N}{\rightleftharpoons} & \\
 & & H^+U + H^+ & \rightleftharpoons & H^+N + H^+ & &
\end{array} \qquad \text{Scheme C3}$$

where $K^0{}_{NU}$ is the equilibrium constant of the N ⇌ U transition. In this scheme, $K^N{}_{a1}$ and $K^N{}_{a2}$ are the K_a values in the native state for site 1 and 2, respectively. On the other hand, $K^U{}_{a1}$ and $K^U{}_{a2}$ are the K_a values in the unfolded state for site 1 and 2, respectively. H^+N and H^+U are the native and the unfolded species with site 1 protonated while NH^+ and UH^+ are the native and the unfolded species with site 2 protonated. The ionization is assumed to occur independently.

$$K_{a1}^{\mathrm{U}}=\frac{[\mathrm{U}][\mathrm{H}^+]}{[\mathrm{UH}^+]}=\frac{[\mathrm{H}^+\mathrm{U}][\mathrm{H}^+]}{[\mathrm{H}^+\mathrm{UH}^+]}$$
$$K_{a2}^{\mathrm{U}}=\frac{[\mathrm{U}][\mathrm{H}^+]}{[\mathrm{H}^+\mathrm{U}]}=\frac{[\mathrm{UH}^+][\mathrm{H}^+]}{[\mathrm{H}^+\mathrm{UH}^+]}$$
$$K_{a1}^{\mathrm{N}}=\frac{[\mathrm{N}][\mathrm{H}^+]}{[\mathrm{NH}^+]}=\frac{[\mathrm{H}^+\mathrm{N}][\mathrm{H}^+]}{[\mathrm{H}^+\mathrm{NH}^+]}$$
$$K_{a2}^{\mathrm{N}}=\frac{[\mathrm{N}][\mathrm{H}^+]}{[\mathrm{H}^+\mathrm{N}][\mathrm{H}^+]}=\frac{[\mathrm{NH}^+][\mathrm{H}^+]}{[\mathrm{H}^+\mathrm{NH}^+]}$$

Eq. C15

The equilibrium constant of the unfolding transition (K_{NU}) is represented as:

$$K_{\mathrm{NU}}=\frac{[\mathrm{U}]+[\mathrm{UH}^+]+[\mathrm{H}^+\mathrm{U}]+[\mathrm{H}^+\mathrm{UH}^+]}{[\mathrm{N}]+[\mathrm{NH}^+]+[\mathrm{H}^+\mathrm{N}]+[\mathrm{H}^+\mathrm{NH}^+]}=K_{\mathrm{NU}}^0\left(\frac{1+10^{(\mathrm{p}K_{a1}^{\mathrm{U}}-\mathrm{pH})}}{1+10^{(\mathrm{p}K_{a1}^{\mathrm{N}}-\mathrm{pH})}}\right)\left(\frac{1+10^{(\mathrm{p}K_{a2}^{\mathrm{U}}-\mathrm{pH})}}{1+10^{(\mathrm{p}K_{a2}^{\mathrm{N}}-\mathrm{pH})}}\right) \qquad \text{Eq. C16}$$

This is applied to the calculation of the activation Gibbs free energy in the (un)folding kinetics.

In more general case with multistate unfolding transitions and multiple ionizable groups, the equilibrium constant of the *i* state ⇌ *j* state transition, K_{ij}, is represented as:

$$K_{ij}=K_{ij}^0\prod_{l=1}^{l_N}\left(\frac{1+10^{\mathrm{p}K_{al}^{j}}}{1+10^{\mathrm{p}K_{al}^{i}}}\right) \qquad \text{Eq. C17}$$

where K^0_{ij} is the equilibrium constant of the deprotonated species of *i* and *j* states, $\mathrm{p}K^{\mathrm{j}}_{\mathrm{al}}$ and $\mathrm{p}K^{i}_{\mathrm{al}}$ are the $\mathrm{p}K_{\mathrm{a}}$ values of the *l*-th ionization group, l_{N} is the number of the ionizable groups.

The (un)folding reaction of (horse skeletal muscle apomyoglobin) hapoMb is represented by a five-state sequential scheme. In the pH range used in this study (pH 2.0-7.1), the acid-induced equilibrium unfolding is induced by protonation/deprotonation of the imidazole and carboxyl groups because the $\mathrm{p}K_{\mathrm{a}}$ value of the other ionizable groups in this protein is well above 7, and always protonated. The difference in the $\mathrm{p}K_{\mathrm{a}}$ values between the imidazole and carboxyl groups (~6 *vs.* ~4) is responsible for the three-state equilibrium unfolding. It is practically impossible to resolve the difference in the $\mathrm{p}K_{\mathrm{a}}$ values among different sites of a single ionizable group (imidazole or carboxyl group). For example, 11 and 12 histidine residues are found in hapoMb and sperm whale apomyoglobin (swapoMb), respectively, while there are eight aspartate and 13 glutamate residues in hapoMb and seven aspartate and 14 glutamate residues in swapoMb. Thus, it is assumed that individual imidazole and carboxyl groups exhibit either normal $\mathrm{p}K_{\mathrm{a}}$ values or single abnormal $\mathrm{p}K_{\mathrm{a}}$ values in each state. The equilibrium constant for the *i* ⇌ *j* transition (*i*, *j*; N, M, I, I', and U) is represented as:

$$K_{ij}=K_{ij}^0\left(\frac{1+10^{\mathrm{p}K_{a\mathrm{His}}^{U}-\mathrm{pH}}}{1+10^{\mathrm{p}K_{a\mathrm{His}}^{i}-\mathrm{pH}}}\right)^{n_{\mathrm{His}}^{i}}\left(\frac{1+10^{\mathrm{p}K_{a\mathrm{His}}^{j}-\mathrm{pH}}}{1+10^{\mathrm{p}K_{a\mathrm{His}}^{U}-\mathrm{pH}}}\right)^{n_{\mathrm{His}}^{j}}\left(\frac{1+10^{\mathrm{p}K_{a\mathrm{COOH}}^{U}-\mathrm{pH}}}{1+10^{\mathrm{p}K_{a\mathrm{COOH}}^{i}-\mathrm{pH}}}\right)^{n_{\mathrm{COOH}}^{i}}\left(\frac{1+10^{\mathrm{p}K_{a\mathrm{COOH}}^{j}-\mathrm{pH}}}{1+10^{\mathrm{p}K_{a\mathrm{COOH}}^{U}-\mathrm{pH}}}\right)^{n_{\mathrm{COOH}}^{j}} \qquad \text{Eq. C18}$$

where $\mathrm{p}K_{\mathrm{a}}{}^{i}{}_{\mathrm{His}}$, $\mathrm{p}K_{\mathrm{a}}{}^{j}{}_{\mathrm{His}}$ and $\mathrm{p}K_{\mathrm{a}}{}^{\mathrm{U}}{}_{\mathrm{His}}$ are the $\mathrm{p}K_{\mathrm{a}}$ values of the imidazole groups in *i*, *j* and U states, respectively, $\mathrm{p}K_{\mathrm{a}}{}^{i}{}_{\mathrm{COOH}}$, $\mathrm{p}K_{\mathrm{a}}{}^{j}{}_{\mathrm{COOH}}$ and $\mathrm{p}K_{\mathrm{a}}{}^{\mathrm{U}}{}_{\mathrm{COOH}}$ are the $\mathrm{p}K_{\mathrm{a}}$ values of the carboxyl groups in *i*, *j* and U

states, respectively, n^{i}_{His} and n^{j}_{His} are the number of the imidazole groups with the abnormal pK_a values in *i* and *j* states, respectively, and n^{i}_{COOH} and n^{j}_{COOH} are the number of the carboxyl groups with the abnormal pK_a values in *i* and *j* states, respectively. K^{0}_{ij} is the equilibrium constant for the transition between *i* and *j* species with the imidazole and carboxyl groups deprotonated. The p$K_a{}^{\mathrm{U}}_{\mathrm{His}}$ and p$K_a{}^{\mathrm{U}}_{\mathrm{COOH}}$ values are 6.0 and 3.6 (normal pK_a values), respectively. Since the Gibbs free energy for each state at pH 6.0 and 8°C is estimated by urea-induced (un)folding reactions, pH value of 6.0 is chosen as the reference pH for the analysis. The pH dependence of the Gibbs free energy in *i* state (*i*: U, TS1, I', TS2, I, TS3, M, TS4 or N, where TS is the transition state) relative to U ($\Delta G_{i\mathrm{U}}$) is represented as:

$$\Delta G_{i\mathrm{U}} =$$

$$\Delta G^{\mathrm{H2O}}_{i\mathrm{U}} - RT\left\{\ln\left[\frac{1+10^{(\mathrm{pK}^{\mathrm{U}}_{a\mathrm{His}}-\mathrm{pH})}}{1+10^{(pK^{i}_{a\mathrm{His}}-\mathrm{pH})}}\frac{1+10^{(pK^{i}_{a\mathrm{His}}-6)}}{1+10^{(pK^{\mathrm{U}}_{a\mathrm{His}}-6)}}\right]^{n^{i}_{\mathrm{His}}} - \ln\left[\frac{1+10^{(\mathrm{pK}^{\mathrm{U}}_{a\mathrm{COOH}}-\mathrm{pH})}}{1+10^{(\mathrm{pK}^{i}_{a\mathrm{COOH}}-\mathrm{pH})}}\frac{1+10^{(\mathrm{pK}^{i}_{a\mathrm{COOH}}-6)}}{1+10^{(\mathrm{pK}^{\mathrm{U}}_{a\mathrm{COOH}}-6)}}\right]^{n^{i}_{\mathrm{COOH}}}\right\}$$

Eq. C19

where $\Delta G^{\mathrm{H2O}}_{i\mathrm{U}}$ represents the relative free energy in *i* state to in U state at pH 6. n^{i}_{His} and n^{i}_{COOH} are the number of the imidazole and carboxyl groups with the abnormal pK_a values in *i* state, respectively. p$K_a{}^{i}_{\mathrm{His}}$ and p$K_a{}^{i}_{\mathrm{COOH}}$ are the abnormal pK_a values of the imidazole and carboxyl groups in *i* state, respectively, while p$K_a{}^{\mathrm{U}}_{\mathrm{His}}$ and p$K_a{}^{\mathrm{U}}_{\mathrm{COOH}}$ are the pK_a values of the imidazole and carboxyl groups in U state, respectively, which correspond to the normal pK_a values. The free energy difference between *i* state and *j* state is represented by;

$$\Delta G_{ij} = \Delta G_{i\mathrm{U}} - \Delta G_{j\mathrm{U}} \qquad \text{Eq. C20}$$

In addition, the elementary rate constant is represented as follows:

$$k_{ij} = \frac{\left[\frac{1+10^{(pK^{\mathrm{U}}_{a\mathrm{His}}-\mathrm{pH})}}{1+10^{(pK^{\ddagger\ddagger}_{a\mathrm{His}}-\mathrm{pH})}}\frac{1+10^{(\mathrm{pK}^{\ddagger}_{a\mathrm{His}}-6)}}{1+10^{(pK^{\mathrm{U}}_{a\mathrm{His}}-6)}}\right]^{n^{\ddagger}_{\mathrm{His}}}\left[\frac{1+10^{(pK^{U}_{aCOOH}-pH)}}{1+10^{(pK_{aCOOH}-pH)}}\frac{1+10^{(pK^{\ddagger}_{aCOOH}-6)}}{1+10^{(pK^{U}_{aCOOH}-6)}}\right]^{n^{\ddagger}_{COOH}}}{\left[\frac{1+10^{(pK^{\mathrm{U}}_{a\mathrm{His}}-\mathrm{pH})}}{1+10^{(\mathrm{pK}^{i}_{a\mathrm{His}}-\mathrm{pH})}}\frac{1+10^{(\mathrm{pK}^{i}_{a\mathrm{His}}-6)}}{1+10^{(\mathrm{pK}^{\mathrm{U}}_{a\mathrm{His}}-6)}}\right]^{n^{i}_{\mathrm{His}}}\left[\frac{1+10^{(\mathrm{pK}^{\mathrm{U}}_{a\mathrm{COOH}}-\mathrm{pH})}}{1+10^{(\mathrm{pK}^{i}_{a\mathrm{COOH}}-\mathrm{pH})}}\frac{1+10^{(\mathrm{pK}^{i}_{a\mathrm{COOH}}-6)}}{1+10^{(\mathrm{pK}^{\mathrm{U}}_{a\mathrm{COOH}}-6)}}\right]^{n^{i}_{\mathrm{COOH}}}} k^{\mathrm{H2O}}_{ij} \qquad \text{Eq. C21}$$

$$k^{\mathrm{H2O}}_{ij} = A\exp\left(-\frac{\Delta G^{\ddagger\mathrm{H2O}}_{ij}}{RT}\right)$$

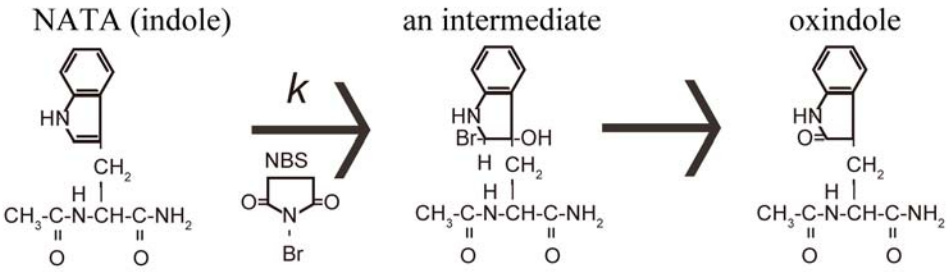
where k^{H2O}_{ij} and $\Delta G^{\ddagger\mathrm{H2O}}_{ij}$ are the elementary rate constant and the activation free energy of the $i \rightarrow j$ transition at pH 6.0, respectively.

Appendix D: Rapid mixing devices

To fully understand the mechanisms of a reaction in chemistry and biology, it is needed to characterize the relevant species and to determine the rate constants for their interconversions. Thus, kinetic measurements are critical for elucidating the reaction mechanisms. For kinetic measurements of protein folding/unfolding reactions, the reactions are often initiated by a rapid change in solvent conditions such as denaturant concentrations and pH using mixing methods. The mixing of a solution containing the native protein with the second solution containing a high concentration of a chemical denaturant such as urea gives rise to the unfolding reaction. Similarly, the folding reaction can be initiated by reducing the denaturant concentration by mixing an unfolded protein solution with a high concentration of a denaturant with a refolding buffer (for example, in the absence of denaturant). In manual mixing experiments, it typically takes a few seconds to completely mix the solutions. Since the kinetics is unable to be monitored until the mixing of solutions is completed, the time required for the mixing is designated as the dead time. Since significant conformational change occurs within milliseconds in typical protein folding reactions, better mixing techniques, which mixes solutions more rapidly and efficiently, are required to observe the folding/unfolding kinetics. Continuous- and stopped-flow methods exhibit much shorter dead times (typically tens of microseconds and a few milliseconds, respectively), which makes it possible to explore the early events occurring during folding/unfolding reactions of proteins. The rapid mixing methods are usually combined with various spectroscopic detection methods or with rapid quenching techniques.

The schematic diagram of typical continuous-flow fluorescence devices is shown in Figure D-1. In folding reactions, two solutions, for example acid unfolded protein solution and refolding buffer, are delivered at a constant flow rate (0.6–1.0 mL/s) from pneumatically driven syringes to the mixer, where the two solutions are efficiently and rapidly mixed to initiate the reaction. The mixed solution is delivered to the flow channel of the quartz flow cell. The progress of the reaction is followed by monitoring the fluorescence profile as a function of the distance from the mixer along the flow channel under steady-state flow conditions. Since the flow rate is constant during observation, the distance from the mixer corresponds to the reaction time. The continuous-flow method allows the solution to flow during the observation, and thus is free from artifacts arising from stop of flow, which are inevitable for stopped-flow experiments (see below). For this reason, the dead time of the continuous-flow experiments is much shorter (almost two orders of magnitudes) than that in stopped-flow experiments. The detection of fluorescence signal is facilitated by using a charge-coupled device (CCD) camera as a detector. The fluorescence intensity along the flow channel is simultaneously collected. An example of a kinetic experiment monitoring the time-dependent fluorescence quenching of *N*-acetyl-$_{L}$-tryptophanamide (NATA) by *N*-bromosuccinimide (NBS) is illustrated in Figure D-2. The concentrations of NATA and NBS are 20 μM and at 8 mM, respectively. In the early stage of the reaction, the indole group in a NATA molecule is brominated by an NBS molecule and finally is converted into an oxindole or brominated

oxindole. A proposed mechanism can be simplified as shown below:



Scheme D1

Thus, when NATA is rapidly mixed with NBS in large excess, the bromination processes are approximated pseudo-first order reactions. Since the brominated intermediates and oxindole are non-fluorescent, the fluorescence of NATA decreases to zero single-exponentially with a rate constant in proportion to the NBS concentration.

The raw data includes contribution from the background signal and the spatial variation of the incident light. Thus, in addition to observing the kinetic reaction, two measurements are needed; a background measurement to estimate the contribution from the background, and a control measurement where the NATA solution is just diluted with the buffer for correcting the spatial variation in the incident light. The corrected fluorescence, *F*, is calculated according to;

$$F = \frac{I_e - I_b}{I_c - I_b} \qquad \text{Eq. D1}$$

where I_e, I_c and I_b are the raw intensities from experiment, control and background measurements, respectively. The time axis is calibrated on the basis of the flow velocity and the dead time calibration. The dead time of continuous-flow fluorescence experiments is typically 10–200 μs. The upper limit of the time scale that can be measured is determined by the length of the flow-cell and the flow velocity.

The schematic diagram of the stopped-flow fluorescence devices (SX-17 and SX-20, Applied Photophysics, UK) is shown in Figure D-3. In stopped-flow experiments, two reagents rapidly mixed in a mixer are delivered to the observation cell, and then ‘stopped’ in the observation cell by using the stop syringe or by closing a valve. The SX series adopt the stop syringe, which stops the flow abruptly when the stop syringe hits the stopping block. The typical flow rates range from 5 to 10 mL/s. The change in signal such as a fluorescence signal and the absorbance typically at a specific wavelength is recorded as a function of time as the reaction proceeds in the observation cell. Typical dead time of the stopped-flow device is a few milliseconds. The time axis is calibrated on the dead time calibration.

The determination of the dead time of the instruments is critical for the kinetic measurements as is also the case for the continuous-flow measurements. Thus, the following conditions should be satisfied for test reactions to calibrate the dead time of continuous- and stopped-flow devices: (i) the reaction is simple (for example a simple order reaction); (ii) the reaction accompanies a significant change detectable by the detector; (iii) the reaction rate is under the control, for example, by

adjusting the concentration of reagents; (iv) a time constant less than 1 ms can be easily achieved especially for kinetic measurements of protein folding; (v) the reagents of high purity is stable, readily available and can be handled without difficulty.

In this study, the fluorescence quenching of NATA by NBS was used for estimating the dead time of the continuous- and stopped-flow fluorescence devices. A 40 μM NATA solution was mixed with an NBS solution (4–32 mM) in the continuous-flow device with the mixing ratio of 1:1 resulting in the final conditions of 20 μM NATA / 2–16 mM NBS. The excitation wavelength was 295 nm and the emission was detected by using a 305-nm cutoff filter. The fluorescence intensity decreased exponentially to zero. Figure D-4 shows the kinetic traces corrected according to Eq. D1. Each kinetic trace was fitted by a single-exponential function. In addition to the quenching reaction, the fluorescence intensity obtained by diluting 40 μM NATA with the buffer (in the absence of NBS) is also recorded (straight lines in Figure D-4) as a reference for the initial fluorescence intensity of the reaction. Thus, the point of the intersection between the reference intensity and the extrapolated theoretical curve of the kinetic trace corresponds to the zero time of the reaction. The dead time is estimated by the time period from the zero time thus defined to a time where the experimentally obtained kinetic trace is well accounted for by the theoretical curve. As can be seen from Figure D-4a the dead time of the continuous-flow experiment was 97 μs. In the case of the dead time estimation of the stopped-flow device, the final experimental condition was 5 μM NATA / 200–800 μM NBS to reduce the rate constant of the quenching reaction to milliseconds time scale (Figure D-4b). The dead time of the stopped-flow experiment was 3.5 ms.

FIGURES

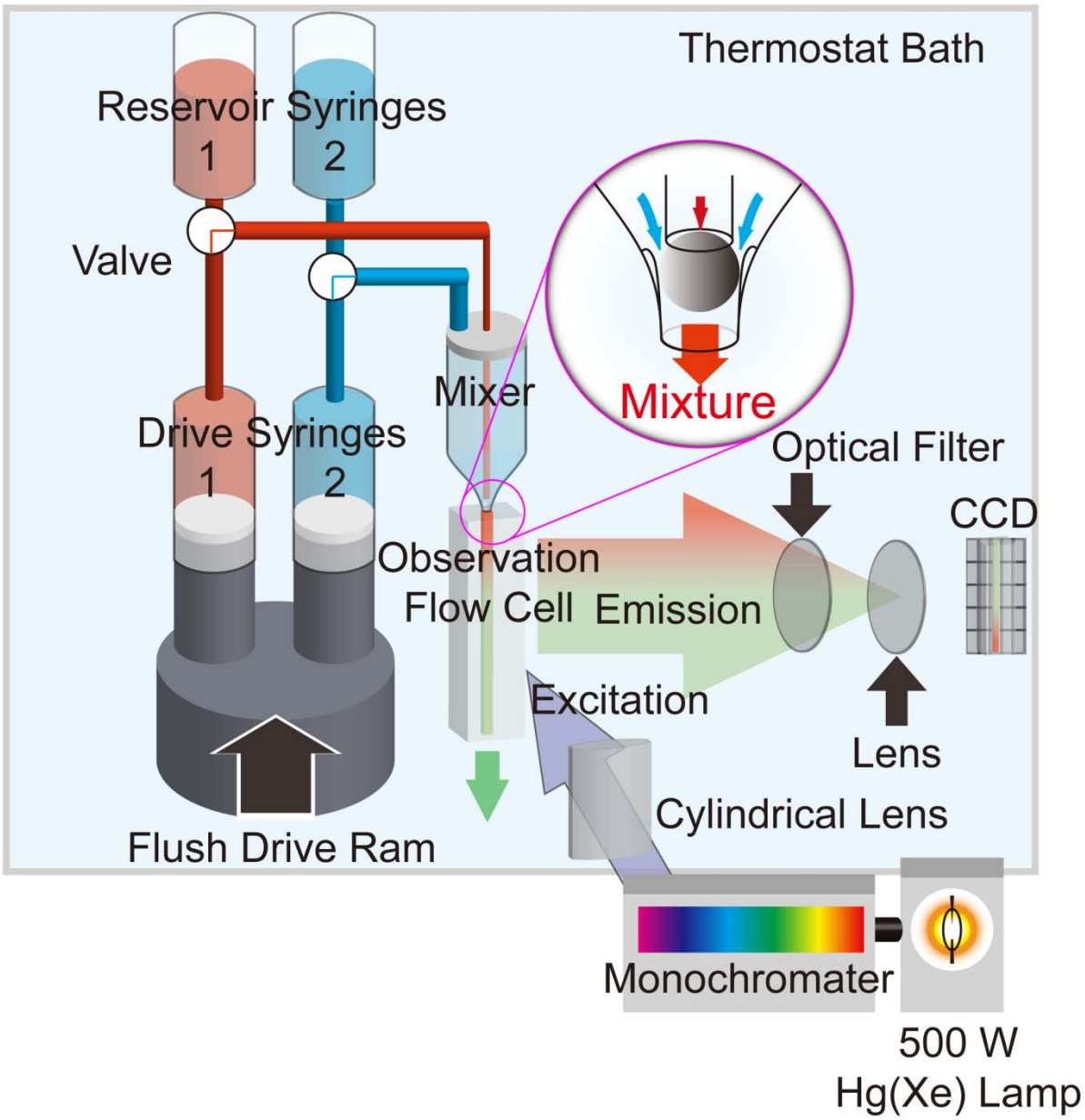
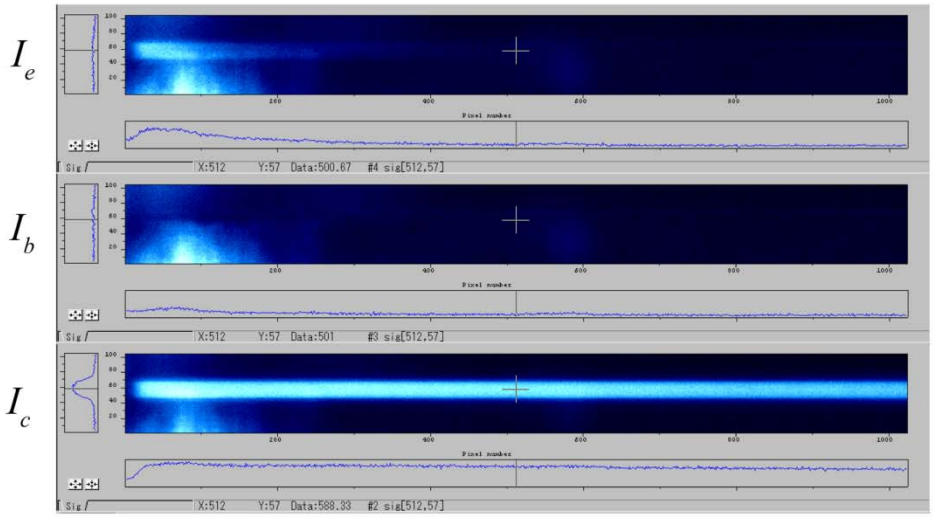
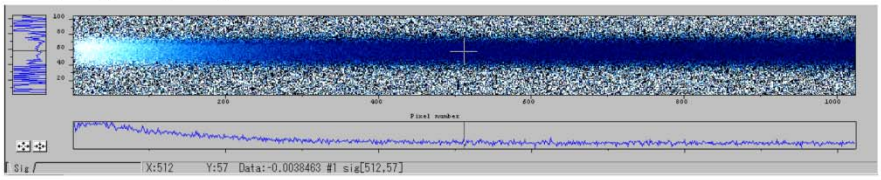


Figure D-1: Continuous-flow fluorescence device. Schematic diagram of the solution delivery system, mixer, observation cell and optical arrangement. The **inset** represents the expanded view of the mixer assembly.



$$F = (I_e - I_b)/(I_c - I_b)$$



Average intensity across the flow channel

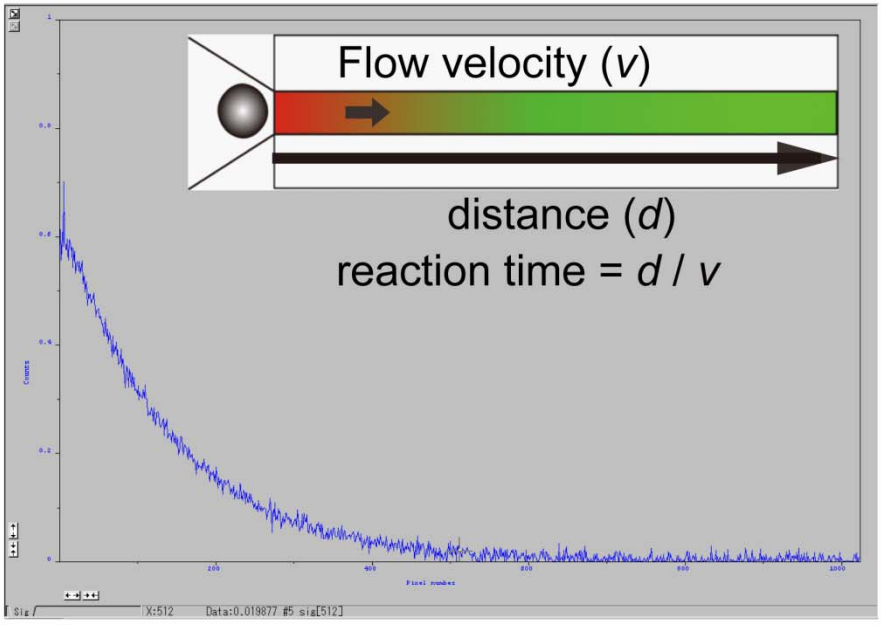


Figure D-2: A typical continuous-flow experiment. I_e) The image of raw experimental data. I_b) Background. I_c) Fluorescence control. F) The corrected data by Eq.D1.
The lowest panel shows an average intensity across the flow channel. As shown in the inset, the time axis is calibrated on the basis of the flow velocity.

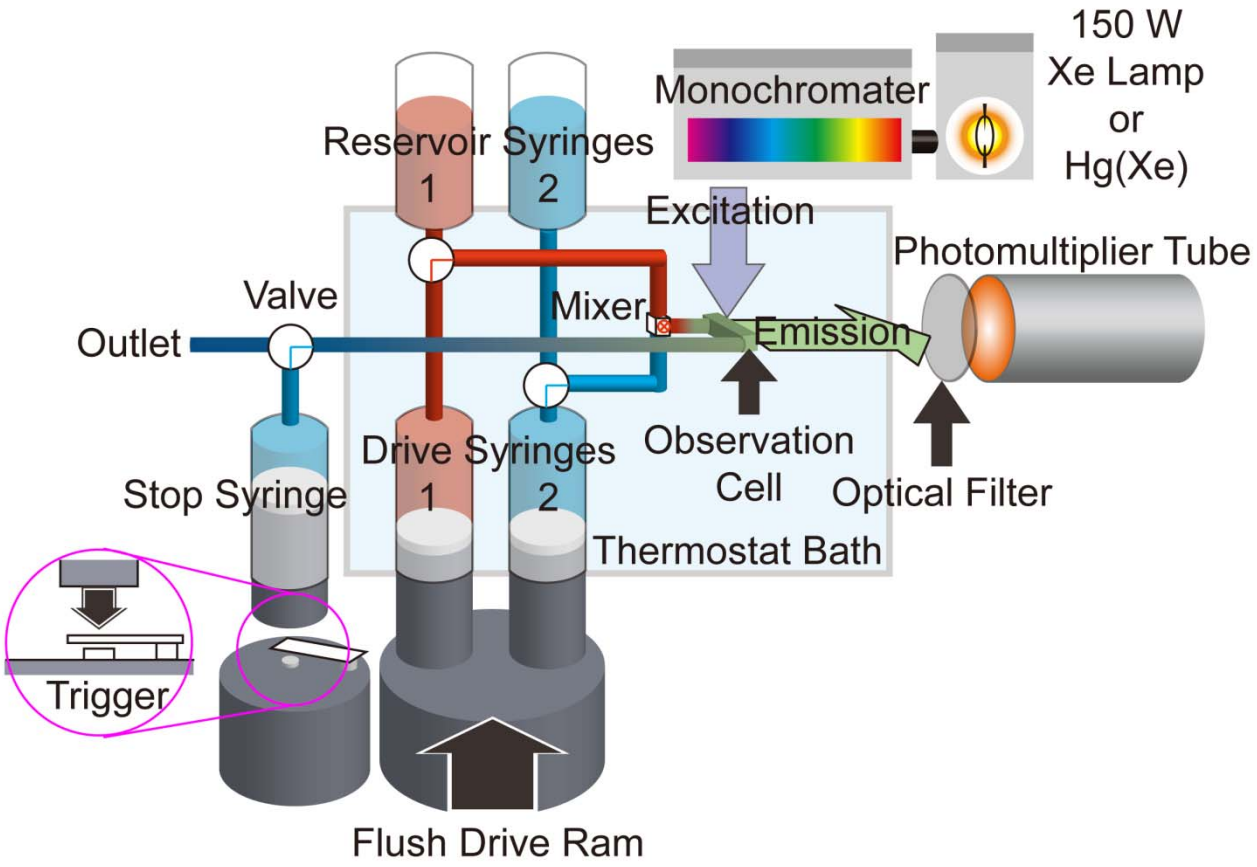


Figure D-3: Stopped-flow fluorescence device. Schematic diagram of the solution delivery system, mixer, observation cell and optical arrangement. The **inset** represents an expanded view of the stopping block. For SX series, recording of reaction is triggered by hitting the stopping block.

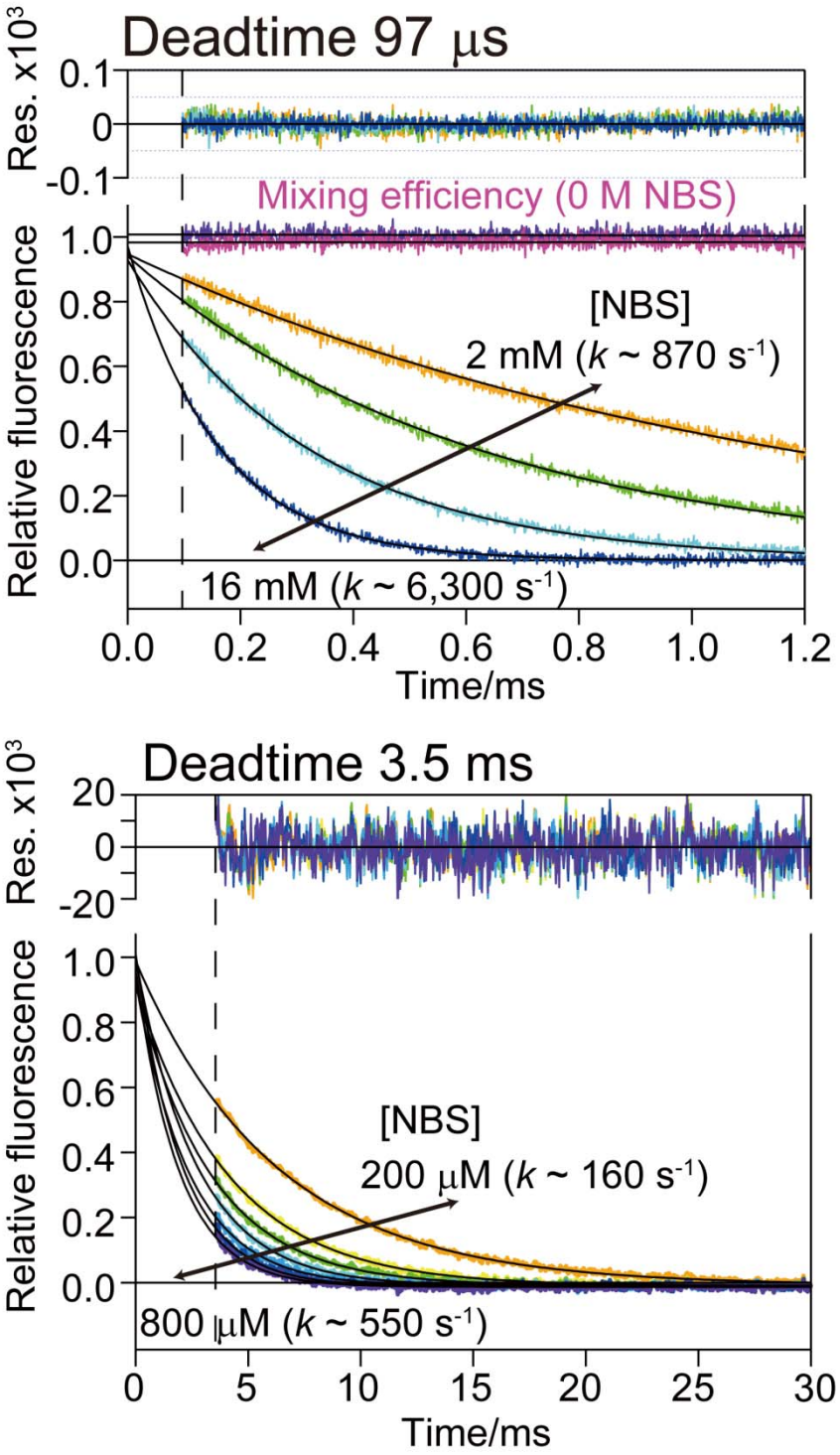


Figure D-4: Estimation of the dead time of (a) continuous-flow and (b) stopped-flow fluorescence devices, using fluorescence quenching of NATA by NBS. The black solid lines represent the fitting curves to a singel-exponential function: $y(t) = \exp(-k\,t)$, where t is the time and k is the apparent rate constant. The dashed lines indicate the dead times.