Review

Folding of apomyoglobin:
Analysis of transient intermediate structure during refolding
using quick hydrogen deuterium exchange and NMR

By Chiaki NISHIMURA*1,†

(Communicated by Teruhiko BEPPU, M.J.A.)

Abstract: The structures of apomyoglobin folding intermediates have been widely analyzed using physical chemistry methods including fluorescence, circular dichroism, small angle X-ray scattering, NMR, mass spectrometry, and rapid mixing. So far, at least two intermediates (on sub-millisecond- and millisecond-scales) have been demonstrated for apomyoglobin folding. The combination of pH-pulse labeling and NMR is a useful tool for analyzing the kinetic intermediates at the atomic level. Its use has revealed that the latter-phase kinetic intermediate of apomyoglobin (6 ms) was composed of helices A, B, G and H, whereas the equilibrium intermediate, called the pH 4 molten-globule intermediate, was composed mainly of helices A, G and H. The improved strategy for the analysis of the kinetic intermediate was developed to include (1) the dimethyl sulfoxide method, (2) data processing with the various labeling times, and (3) a new in-house mixer. Particularly, the rapid mixing revealed that helices A and G were significantly more protected at the earlier stage (400 µs) of the intermediate (former-phase intermediate) than the other helices. Mutation studies, where each hydrophobic residue was replaced with an alanine in helices A, B, E, F, G and H, indicated that both non-native and native-like structures exist in the latter-phase folding intermediate. The N-terminal part of helix B is a weak point in the intermediate, and the docking of helix E residues to the core of the A, B, G and H helices was interrupted by a premature helix B, resulting in the accumulation of the intermediate composed of helices A, B, G and H. The prediction-based protein engineering produced important mutants: Helix F in a P88K/A90L/S92K/A94L mutant folded in the latter-phase intermediate, although helix F in the wild type does not fold even at the native state. Furthermore, in the L11G/W14G/A70L/G73W mutant, helix A did not fold but helix E did, which is similar to what was observed in the kinetic intermediate of apoleghemoglobin. Thus, this protein engineering resulted in a changed structure for the apomyoglobin folding intermediate.

Keywords: NMR, apomyoglobin, protein folding, folding intermediate, site-directed mutagenesis, hydrogen deuterium exchange, non-native structure

1. Introduction

Protein folding is important for protein function. An unfolded protein can fold spontaneously into its native structure, based on the primary sequence.1) This suggests that information specifying how a protein should fold into its native tertiary structure is included in the primary sequence. For some proteins between the unfolded and native states, there is a folding intermediate in what is known as a “molten-globule” state; it possesses the compact, native-like secondary structure of the properly folded protein, but it lacks the tertiary structure.2) Models of
the folding process have been considered for many different proteins including the framework model, the nucleation-condensation model, and the hydrophobic collapse models. Structural heterogeneity, achieved via multiple folding routes, has been observed during the protein folding events; this led to the expression by using the "folding funnel" to explain the energy landscapes of protein folding.

Myoglobin (Mb) is a heme protein, and holomyoglobin (holoMb) is composed of eight helices (Fig. 1). In the absence of heme, one of these helices, F-helix, is not folded even at the native state. The heme-bound form has biological activity as a carrier of oxygen for storage in muscle tissue. Carbon monoxide also binds to holoMb. The structure-function relationship and ligand-binding have been intensely studied on holoMb as a typical heme protein.

The formation of a stable folding intermediate can cause protein misfolding, as a misfolded protein, as well as aggregated protein, may arise from the folding intermediate. The apomyoglobin (apoMb) folding and its intermediate have been widely studied with kinetic and equilibrium approaches using fluorescence, circular dichroism, small angle X-ray scattering, NMR, and mass spectrometry. Originally, the folding and unfolding studies on apoMb were initiated as a typical, helical protein in the different groups by using the different approaches. The partially folded apoMb intermediate, in equilibrium, was characterized at the atomic level as being composed of helices A, G and H using NMR and hydrogen deuterium (HD) exchange. Kinetic studies on the folding process were conducted using a combination of pH-pulse labeling and NMR as well as small-angle X-ray scattering. The former kinetic study further revealed that apoMb's kinetic intermediate is composed of helices A, B, G and H, and the latter study indicated that the radius of gyration is 23 Å in the intermediate compared to 34 Å in the unfolded state. Studies of apoMb folding have already generated much information about protein folding. For this reason, apoMb is one of the most well-known model proteins for studying protein folding.

NMR studies have an advantage by being able to analyze the structure at the atomic level, but its sensitivity for the data acquisition is generally so low that it requires long acquisition times. Therefore, the studies on the equilibrium intermediate are widely performed as a model structure for the kinetic intermediate. In this review, the strategy of a pH-pulse labeling and NMR combination method was focused on for the analysis of the kinetic intermediate of apoMb. The data revealed the secondary structure; additionally, when site-directed mutagenesis was further employed with pH-pulse labeling and NMR, the tertiary structure for the transient intermediate was also revealed.

The processes in protein folding are generally classified into two categories, which are characterized by the molecular size. The folding of smaller proteins displays the down-hill folding classification, where folding is caused by local contact between residues. The speed of folding in this classification is very fast, and less folding cooperativity is observed. By contrast, the folding of larger proteins is conducted based on the contact order of not only local contacts but also long-range interactions from the amino acid sequence. The
folding style of apoMb (153 residues) belongs to the latter group, in which the contact order and topology of the protein is closely related to the folding rates. Particularly, the process of the topomer search can be a complicated and time-consuming step. Moreover, as apoMb is a larger protein, there is significant accumulation of a folding intermediate for a short period during its folding pathway after the initiation of refolding owing to the roughness of the energy landscape. Analyses of the intermediate and its structure are important when considering the folding events of apoMb in the energy landscape. In this review, the analysis of the kinetic intermediate of apoMb by NMR and quick HD exchange, which is also called the pH-pulse labeling, is summarized, mainly focusing on the latter-phase kinetic (I2) intermediate (Equation [1], where U = unfolded protein, I1 and I2 = former-phase and latter-phase intermediates, respectively, and N = folded protein at the native state).

\[
\begin{align*}
U & \rightarrow I1 \rightarrow I2 \rightarrow N \\
\text{Folding time} & = 100-500 \mu s \sim 6 \text{ ms}
\end{align*}
\]

For the folding pathway, three kinetic phases for the rate constants have been identified between the unfolded and native states, indicating the existence of two intermediates. In addition, possibilities for sequential, parallel, and dead-end folding pathways that include the formation of I1 and I2 were considered as suitable models for kinetic data analysis. Furthermore, the sequential pathway was also supported by other experiments as indicated in section 2.6. It was also recently shown that the former-phase kinetic intermediate for apoMb (I1) contains one more intermediate. However, for the purpose of simplifying considerations of this folding system, the I1 intermediate is treated as representing a single structural species in this review.

Comparative studies between the kinetic and equilibrium intermediates and mutagenesis studies have been performed for apoMb. From these, it was discussed that the structure of the I2 intermediate is mostly native-like, based on the Go-model, but non-native structure partially exists in the I2 intermediate.

The analysis of the structure in the I1 intermediate is intensely important, because the study contributes to better understanding of the folding mechanism at the early stage; it has also led to the development of a new mixer for the folding experiment, which has advanced the field. However, understanding the I2 intermediate is also important. Protein aggregation is primarily associated with misfolding of the I2 intermediate, because the structure of the I1 intermediate is too flexible and premature.

2. Analysis of the structure for the kinetic intermediate and its improved methodology

2.1 pH-pulse labeling. The structure of the apoMb kinetic intermediate is transient, and this form of the protein accumulates for only the sub-milli-second and millisecond durations. Although NMR reveals the solution structure of the protein, it is difficult to apply it to the transient structure. Therefore, quick HD exchange combined with NMR was employed. Depending on the solution conditions, the HD exchange for the exposed regions in the protein occurs even in a short period, including the milli-second duration, at high pH. The rate constant for the exchange of the amide proton is proportional to the concentration of OH⁻ in the protein solution. We can deduce by pH-pulse labeling which regions in the protein are exposed to the solvent and which regions are formed with hydrogen bonds. By increasing the velocity of the exchange of amide protons, pH-pulse labeling is used for the observation of the snap-shot structure. Typical experiments were performed with 6.4 ms–6 s folding times and 20 ms labeling duration in high pH (pH 10.1) with a deuterated buffer after the initiation of protein folding. The protein is first unfolded in the presence of either 6 M urea or low pH solution (pH 2.2) and after the dilution of the denaturant or acidic solution, the protein folding is initiated.

2.2 Monitoring by mass spectrometry. Analysis by mass spectrometry is important for the identification of different folding species during the process of the protein folding. In the case of apoMb, the pH-pulse labeled sample with the deuterium buffer revealed three different peaks corresponding to three folding species (Fig. 2): the peak with the highest molecular weight corresponds to the unfolded protein; the lowest molecular weight corresponds to the folded protein; and the peak between the two peaks corresponds to the folding species at the I2 intermediate state. As protein folding proceeds, the peak intensity for the native state increases, whereas that for the I2 intermediate state decreases. At the minimum duration after the initiation of the protein folding (6.4 ms) only the peak corresponding to the I2 intermediate is observed. This is evidence that the I2 intermediate observed at 6.4 ms, between the unfolded and native states, is obligatory in the apoMb folding pathway.
2.3 Monitoring by NMR. Typical heteronuclear single quantum coherence (HSQC) spectra for the estimate of the protected amide protons were collected using each pH pulse-labeled sample for their refolding times. Proton occupancies were calculated as a standard of the peak height for the longest refolding time (Fig. 3). This combined strategy with pH-pulse labeling and NMR has been used previously for the kinetic studies on the folding of proteins at the residue level. A limitation of this strategy is the artificial contribution to the data of the proton occupancy. A high pH solution such as pH 10.1 affects the protein structure itself even for a labeling duration of 20 ms. Particularly, the structure of the intermediate is fragile and easily interrupted. Therefore, the extrapolated values were calculated based on the series of values obtained for the different durations of exchange times. The value of A0, which is free from the effects of pH-pulses, represents proton occupancy for 0sec duration (t = 0) of exchange. The amide protons on the side chains are also available for the probe residues. However, the signal assignments for side chains are much more difficult than those for the backbone.

2.4 Dimethyl sulfoxide (DMSO) as an NMR solvent. In the conventional strategy, holoMb is prepared after pH-pulse labeling to increase the protection of amide proton exchange prior to the collection of NMR data for apoMb. The exchange of the amide protons occurs during the pH-pulse labeling with high pH. However, artificial HD exchange is supposed to occur in the destabilized structure of apoMb, even at pH 6 during the sample preparation and data acquisition after the labeling. Originally, the formation of the holo-protein was chosen because it has the more stabilized structure at the native state than apoMb. However, it still possesses a disadvantage that the protons that are located at the exposed and unprotected regions are exchangeable with the protons in solution even in the stabilized holoMb.

By contrast, the protein is unfolded in DMSO. The HD exchange is supposed to more readily occur in the unfolded protein compared to the folded protein; however, DMSO does not contain any exchangeable proton. Therefore, although the protein is unfolded in DMSO, the exchange of amide protons is strongly inhibited in DMSO.

In the new strategy, the samples were quickly frozen in liquid nitrogen and lyophilized after the pH-pulse labeling. The reaction of HD exchange does not occur at the extremely low temperatures of the frozen conditions. DMSO solution was added to the lyophilized sample in a 50 ml tube a few minutes before the NMR data acquisition. In the typical experiments, the acquisition time of 20 min was required for the HSQC in order to get good resolution for the F1 direction of the unfolded protein. Using
DMSO for the NMR solution, data for the amide protons of 94 residues was acquired for the 153 residues of apoMb, compared to the 52 residues probed using the holoMb method (Fig. 4).

2.5 Refined structure for the I2 intermediate.

Using the A0 strategies, in which multi-pulse durations were used for the extrapolation, a new, improved structure was demonstrated.\(^{50}\) In the I2 intermediate, which was observed 6.4 ms after the initiation of refolding by pH-jump, the protection of helices A, G, and H as well as helix B was observed.\(^{50}\) Based on a comparison of the data between the A0 strategy and conventional strategy, it was concluded that the formation of helix B in the I2 intermediate was sensitive to solution pH.

In addition, information about helix E was obtained by the new method. The A0 for L72 and I75 was almost 0.7, indicating that this area is mostly folded.\(^{50}\) The other residues in helix E indicated that their A0 values were near 0.5. This revealed the presence of structural heterogeneity,\(^{41}\) in which half the protein population is folded and the other half is unfolded.\(^{50}\) Helix E region is supposed to constitute a border between the folded and unfolded states in the I2 intermediate, and the structural features may be revealed through the destabilization of the N-terminal part of helix B, as will be shown later.\(^{55}\)

Furthermore, as mentioned above, the DMSO method significantly increased the number of probe residues.\(^{39}\) In particular, new probe residues were identified at both termini of each helix (Fig. 4A, B). Generally, both ends of helices are fragile, whereas the middle parts of helices are stabilized by hydrogen bonds.\(^{56}\)

2.6 Developing the rapid mixer for the analysis of the early stage of the protein folding.

The early stage of protein folding is difficult to observe using the conventional apparatus. The dead time, in which the protein folding cannot be monitored, was 6.4 ms after the initiation of the protein folding when using a Biologic QFM5 Quench Flow instrument. The stopped-flow experiments monitored by circular dichroism (CD), with a dead time of 20 ms when using a DX-17MV Applied Photophysics instrument, revealed that about 70% helical structure was already folded at the first data point. An initial conclusion was that the burst-phase intermediate, corresponding to the I2 intermediate, exists for apoMb.\(^{23}\) Subsequently, two kinetic intermediates for both horse and whale apoMbs were demonstrated as sequential I1 and I2 by the refolding/unfolding studies monitored with fluorescence and CD including double-jump experiments.\(^{33,34,57}\)

For the studies on apoMb refolding, the early stage of the protein folding (sub-millisecond) that includes the I1 intermediate was analyzed by using a new continuous-flow mixer.\(^{25}\) CD and small angle X-ray scattering were used for the studies on horse apoMb. Based on the values of both 1) the radius of gyration from small angle X-ray scattering and
2) helical content from CD, two kinetic intermediates of I1 and I2 were characterized. Between I1 and I2, the values of helical contents were different, whereas the values of the radius of gyration did not change significantly.

Furthermore, the continuous-flow mixer was employed for the studies on the early stage of folding of whale and horse apoMbs. More than two intermediates were identified for apoMb folding based on the refolding/unfolding studies of urea-jump at acidic and neutral pHs, monitored by the fluorescence and CD. Laser-induced oxidative labeling after the initiation of folding by rapid mixing was combined with the mass spectrometry-based peptide mapping for the studies on the folding of horse apoMb. This strategy revealed that helices A and G were already mostly folded after 200 μs in the I1 intermediate.

A new, in-house mixer was developed for the sub-millisecond rapid mixing for NMR studies with site-specific information. The system for the continuous flow was used for the rapid mixing. The dead time was decreased from 6.4 ms to 400 μs. In this system, the data were collected as a function of pH with constant folding time (400 μs or 6 ms) and a subsequent labeling time (3.6 ms). This method revealed that helices A and G are tightly packed compared to helix H in the I1 intermediate of apoMb; helix B does not fold tightly at this stage. In this experiment, the EX2 regime was confirmed for the exchanges of amide protons as a function of pH. A limitation of the quick HD exchange via the pH-pulse labeling method for the folding study is that the pulse-duration is required for sufficient labeling (3.6 ms), which is different from the other kinetic studies including the fluorescence and CD studies. For the system of other proteins, based on the careful studies on Cyt c by the equilibrium and kinetic HD exchange experiments, the stepwise assembly of units was demonstrated.

2.7 Structural change in the helix induced by cis-trans isomerization of chemical linkages. Real-time effects of the breakage of a helix on the structure of apoMb are intended to be observed in the in-house probe on NMR. Azobenzene was used for the bridge linkage in the helical structure. Generally, a linkage with two azobenzenes can be formed between the side chains of the incorporated cysteines between the i and i+11 residues in the helical region of proteins. In the absence of UV light, the conformation of the two azobenzenes is in the trans conformation, whereas it is cis in the presence of UV light. On-off switching via application of UV light was performed in the NMR probe to observe conformational changes.

For preliminary experiments, a peptide segment of helix H from sperm whale myoglobin (G124–L149) was used. The L136 and K147 were mutated to cysteines for the incorporation of azobenzenes. The reason why the helix H peptide was chosen for this experiment was that helix H has the highest propensity for helical formation in the helices of apoMb. Additionally, two alanine mutations (R139A and K140A) were introduced to avoid the steric effects on the helical structure induced by the formation of the azobenzene bridge. Breakage of the helical structure was induced by UV light (inducing the cis conformation in the azobenzenes) and was observed in the NMR probe in real time. It was confirmed that the on-off switch for the helix breakage in the peptide was fully effective in the NMR probe. Based on the peptide experiments, the native protein will be expressed with the azobenzene bridge modification at helix H for future studies of apoMb folding.

2.8 Observation for the equilibrium intermediate of apoMb. The equilibrium molten-globule intermediate of apoMb observed at pH 4.0 was used as a model for the kinetic (I2) intermediate. HD exchange was employed for the analysis of the equilibrium structure. The trapped HD exchange was conducted for the experiment. Briefly, the incubation for the HD exchange was conducted at pH(D) 4.0, and the NMR spectra were collected after formation of the holoMb at pH(D) 5.6. The signal assignments were available at the condition of pH 5.6 for the holoMb. The structure of the equilibrium intermediate of apoMb was indicated as consisting of the helices A, G and H. Subsequently, the structure of the equilibrium intermediate was further demonstrated by using the DMSO strategy (Fig. 5B, D). In this case, to chase the fast-exchange protons, the incubations at pH 4.0, 4 °C were conducted for the short terms at the scale of sec and min. Each data point for HD exchange was collected using individual samples for the different exchange durations. This revealed that the equilibrium intermediate is similar to the I2 intermediate, but that there are also local differences between them (Fig. 5A, B). The more destabilized regions were located at both the ends of helix G and the entire helix B for the equilibrium intermediate, whereas helix D was more stabilized (Fig. 5C, D).
Fig. 5. Difference of the structures between (A) kinetic and (B) equilibrium intermediates. (A) The proton occupancy of the kinetic intermediate (I2) and (B) the protection factor (protected ratio normalized by the intrinsic exchange rate for HD exchange) for the equilibrium intermediate of apoMb. The locations of helices are shown at the top of the figures, represented as boxes. Protections of the (C) I2 (kinetic) and (D) equilibrium intermediates were mapped onto the crystal structure of Mb. High, medium, and low protections for residues are expressed by red (black residue number), green (green residue number), and blue spheres, respectively. The parts indicating the different protections between two intermediates are circled by color rings. These plots and mappings on the structure were reproduced with some modifications from the original paper.39)
The equilibrium intermediate at pH 4.1 was more directly observed under the NMR condition at 50°C and 10% ethanol. Based on the chemical shift and relaxation studies, the secondary structure in the equilibrium intermediate was demonstrated.\textsuperscript{18,65} However, under this condition, there was a difficulty in conducting some other experiments.

Topologies of the equilibrium intermediates of apoMb were determined based on data collected at the acidic condition of pH 2.3 by the residual dipolar coupling\textsuperscript{66} and spin-label experiment.\textsuperscript{67,68} Spin-label experiments reveal the information about the distance between the incorporated probe and each residue. The results obtained for apoMb suggested that the chain cluster is related to the helix formation in the equilibrium intermediate that is expected to comprise helices A, B, G and H. The presence of non-native structure within this equilibrium intermediate was also demonstrated, specifically at helix C region.

In the equilibrium condition, the relaxation dispersion experiments are useful for elucidating 1) the ratio of population of each species, and 2) the structure of a minor component that is usually invisible. Based on the data, the folding events of apoMb were characterized focusing on the existence of the folding intermediates.\textsuperscript{69} Observations of unfolding/refolding kinetics around pH 4.75–4.95 during the relaxation dispersion experiments revealed that the transient molten-globule state is similar to the equilibrium intermediate, and that a hidden intermediate exists between the native and equilibrium intermediate states.

Fast photochemical oxidation was employed for the combined experiments of pulse labeling and mass spectrometry.\textsuperscript{70} It revealed that under the equilibrium condition at pH 4, helix G is highly protected with accompanying partial protections of helices A, B and H. Furthermore, under high-pressure conditions for denaturation, both CD and spin-labeling were used to perform dynamic studies on apoMb folding.\textsuperscript{71} High-pressure conditions make protein structure rapidly fluctuate; as would be expected, the pressure-populated, molten-globule state of apoMb at pH 6.0 displayed the full complement of native-state secondary structures, but a fluctuating tertiary fold.

3. The effects of the mutation of key residues on the structure of the I2 intermediate of apoMb

3.1 Helix B mutations with alanine. On the basis of the HD exchange and pH pulse-labeling experiments, helix B was indicated as a weak point in the equilibrium and kinetic I2 intermediates compared to the regions in helices A, G and H.\textsuperscript{23} The further destabilization of helix B was intended by helix B mutations.\textsuperscript{40} The mutations were tested from hydrophobic bulky residues to alanine residues at the N-terminal part of helix B including I28A, L29A, I30A, and L32A. The effects of the mutation were determined by observing the proton occupancy of the I2 intermediate for helix B residues. Interestingly, the long-range destabilizing effects by L32A were indicated on the residues at helix G including F106, H107, E109, and A110, suggesting that the region around L32 is in close proximity to helix G region in the structure of the I2 intermediate. Since helix B/helix G contact is observed in the native structure, the structure observed in the I2 intermediate is native-like. Furthermore, the I30A mutation stabilized the structure at helix B region as well as that at helix C region in the I2 intermediate. This suggests that the native-like helix B/helix C packing can be enhanced by the I30A mutation. On the other hand, the L29A mutation destabilized the mutated region in helix B while simultaneously stabilizing the C-terminal part of helix E. Since the stabilized region at helix E is not in close contact with the mutated L29 in the native structure, helix B/helix E contact observed in the I2 intermediate is probably non-native. Taken together, these observations suggested that native-like and non-native contacts exist in the I2 intermediate of apoMb (Table 1).

3.2 Helix B stabilized mutants. In order to examine the stabilized effect on the N-terminal part of helix B, two mutants were constructed for the studies on the folding of apoMb\textsuperscript{55} (Fig. 6): The G23A/G25A mutations were intended for the stabilization of the N-terminal part of helix B by introducing helix-favorable alanines.\textsuperscript{72} The H24L/H119F mutations were introduced to stabilize the contacts between the N-terminal part of helix B and the C-terminal part of helix G, so as to form a hydrophobic interaction rather than the normal, pH-dependent salt-bridge.\textsuperscript{73,74} Both G23A/G25A and H24L/H119F mutants stabilized the mutated N-terminal part of helix B in the I2 (Fig. 6A) and equilibrium (Fig. 6B) intermediates. Furthermore, both mutations stabilized the GH-turn in the intermediates. This suggests that the destabilized structure at the N-terminal part of helix B, composed of two helix-unfavorable glycines in the wild type (WT), interrupts the smooth folding of apoMb owing to energetic frustration. Both mutations stabilized helix B in the intermediates and probably supported
the formation of the native-like GH-turn in the intermediates. Furthermore, the hydrophobic contact between L24 and F119 supported helical formation at helix E. This helix B/helix E contact is probably native-like. Both G23A/G25A and H24L/H119F mutations increased the folding rates (4.6 s$^{-1}$ and 4.3 s$^{-1}$, respectively) compared to the WT (3.5 s$^{-1}$).

### 3.3 Distal histidine mutant: H64F.

The distal histidine has an important role in binding heme to carry oxygen. This histidine was mutated to phenyl alanine for the folding studies.\(^75\) The folding rate increased with the mutation more than twofold. This might indicate that the stabilization of helix E in the I2 intermediate increased the folding rate from the I2 intermediate to the native state. The pH-pulse labeling and NMR studies revealed that the structure of the I2 intermediate contained a more protected helix E compared to that of the WT protein. It was concluded that the folding of helix E is critical step after the formation of the I2 intermediate.

### 3.4 Helix H mutation: N132G/E136G.

Helix H has the highest propensity for helical formation\(^64\) and folds in the I2 intermediate of WT. However, helix H mutation N132G/E136G did not show any significant effect on the global folding of apoMb, including the structure of the I2 intermediate, other than a loss of helix H formation.\(^76,77\) This suggests that the role of helix H can be compensated by the other helices in the I2 intermediate.

### 4. Non-native structure in the I2 intermediate

#### 4.1 Observation of the native-like and non-native structure in alanine mutants at each helix.

In order to get the whole picture of the I2 intermediate, helices A, E, G, and H mutations at the core regions were established.\(^40\) The hydrophobic and bulky residues, including leucine, isoleucine, and valine, were replaced with alanine. Mutated regions were targeted to destabilize the I2 intermediate. Interestingly, only for V10A (in helix A) and L115A (in helix G), destabilized regions were also observed remote from the location of mutated residues in the native structure (Table 1). It is possible that a non-native structure exists in the I2 intermediate and that translocation of helix H occurs in the folding transition from the I2 intermediate to the native state. By contrast, the other mutants including V17A, L104A, I111A, L135A, and I142A did not indicate any non-native tendency.\(^40\)

The folding kinetics of some mutants including I28A (in helix B), L69A (in helix E), and L72A (in helix E) were significantly slower than those of the

### Table 1. Native-like and non-native structure in the folding intermediate of apoMb

<table>
<thead>
<tr>
<th>Mutants</th>
<th>Folding rates (Stopped-flow) from I2 to N</th>
<th>Stabilized helix (turn) (Pulse label) in I2</th>
<th>Destabilized helix (Pulse label) in I2</th>
<th>Partial structure (Pulse label) in I2</th>
</tr>
</thead>
<tbody>
<tr>
<td>G23A/G25A(^55)</td>
<td>Fast</td>
<td>B, GH</td>
<td>Native-like</td>
<td></td>
</tr>
<tr>
<td>H24L/H119F(^55)</td>
<td>Fast</td>
<td>B, GH, E</td>
<td>Native-like</td>
<td></td>
</tr>
<tr>
<td>H64F(^75)</td>
<td>Fast</td>
<td>E</td>
<td>Native-like</td>
<td></td>
</tr>
<tr>
<td>I30A(^46)</td>
<td>Fast</td>
<td>B–C</td>
<td>Native-like</td>
<td></td>
</tr>
<tr>
<td>L32A(^46)</td>
<td>Fast</td>
<td>B–G</td>
<td>Native-like</td>
<td></td>
</tr>
<tr>
<td>P88K/S92K(^46)</td>
<td>Slow</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P88K/A90L/S92K/A94L(^60)</td>
<td>Slow</td>
<td>F, GH</td>
<td>B (E)</td>
<td>Non-native</td>
</tr>
<tr>
<td>L28A(^46)</td>
<td>Slow</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L69A(^40)</td>
<td>Slow</td>
<td></td>
<td>B</td>
<td>Non-native</td>
</tr>
<tr>
<td>L72A(^40)</td>
<td>Slow</td>
<td></td>
<td>E</td>
<td>Non-native</td>
</tr>
<tr>
<td>V10A(^40)</td>
<td>Slow</td>
<td></td>
<td>A–H</td>
<td>Non-native</td>
</tr>
<tr>
<td>L29A(^46)</td>
<td>Medium</td>
<td>E</td>
<td>B</td>
<td>Non-native</td>
</tr>
<tr>
<td>L115A(^40)</td>
<td>Medium</td>
<td></td>
<td>G–H</td>
<td>Non-native</td>
</tr>
</tbody>
</table>

* In general, the distance between two residues decreases as the compactness during protein folding.

C. NISHIMURA [Vol. 93,18
WT (Table 1). This suggests that the side chains of I28, L69, and L72 contribute to the formation of the native-like structure in the I2 intermediate.

4.2 Non-native structure observed by Förster resonance energy transfer (FRET) and Cys-quenching. On the basis of the combination of pH-pulse labeling and mutagenesis experiments, it was suggested that some degree of non-native structure exists in the I2 intermediate.40,46 FRET was used for further intermediate studies on apoMb.79 The distance between the donor and acceptor was examined at the equilibrium intermediate and native states. For an instance, the distance between the donor Trp14 (in helix A) and the acceptor probe 5-[[2-[(acetyl)-amino]ethyl]amino]-naphthalene-1-sulfonic acid (AEDANS) coupled at residue position K140C (in helix H) was shorter in the equilibrium intermediate compared to that in the native state. This suggests that the intermediate structure includes some degree of non-native structure between helices A and H. Furthermore, the fluorescence of Trp14 was quenched by the incorpora-
rated Cys135 more efficiently in the equilibrium and I2 intermediates than that in the native state, suggesting the non-native structure induced by helix H translocation exists in the intermediate.\textsuperscript{79} A native-like disulfide bond was formed in the Cys-mutated apoMb protein between C108–C135 in the folding studies.\textsuperscript{79} In this mutant, helix H translocation from the I2 intermediate to the native state was prevented by locking helices G and H with a disulfide bond in the native-like manner.

5. Protein engineering to control the intermediate structure

5.1 Refolding of helix F mutants, in which non-native structures exist in the I2 intermediate. Helix F does not fold even in the native state of apoMb.\textsuperscript{12} Since the NMR signals for helix F residues disappear in the HSQC spectra, the fluctuated structures of helix F, which has multiple conformations, are supposed to exist in the NMR time-scale. Two helix F mutants were established for the experiments: P88K/S92K and P88K/A90L/S92K/A94L\textsuperscript{80} (Fig. 7B). These two mutants of apoMb clearly showed the presence of helix F at the native state, similar to those in holoMb.\textsuperscript{80}

Based on calculations of the Average Area Buried Upon Folding (AABUF) that were generated from the amino-acid sequence,\textsuperscript{81} it may be predicted that helix F in the intermediate state will not fold in the first mutant (P88K/S92K), but that it will fold in the latter mutant (P88K/A90L/S92K/A94L) (Fig. 7C). The AABUF calculation was established based on the idea that the folding initiation sites are not only hydrophobic but are also larger side-chains with the charges.\textsuperscript{82} Interestingly, it was suggested that the prediction by AABUF calculation could be useful in predicting the intermediate structure.\textsuperscript{82} On the other hand, as expected, the formation of helix F was experimentally observed only for the I2 intermediate of P88K/A90L/S92K/A94L mutant (Fig. 7B). Additionally, the refolding rates of the P88K/A90L/S92K/A94L mutant (0.6 s\textsuperscript{-1}) and P88K/S92K mutant (1.8 s\textsuperscript{-1}) were significantly slower than that of the WT protein (3.4 s\textsuperscript{-1}). These suggest that non-native structure in the I2 intermediate was more stabilized by the formation of the central core with helices A, F, G and H, compared to that of the WT. This stabilization of the non-native structure was supported by the observation that the proton occupancies of residues within helices D and E were up and down, respectively. This was particularly true for the P88K/A90L/S92K/A94L mutant, compared to the WT (Fig. 7A, B).

Furthermore, differences in proton occupancy between the P88K/A90L/S92K/A94L mutant and the P88K/S92K mutant were observed not only at helices G and H, but also at helix B and a part of helix E (Fig. 7D). The differences in proton occupancies were observed in the opposite direction between helices G/H and B/E. This suggests that more non-native structure is contained for the former mutant at the regions of helices G and H, and that the native-like folding observed in helix B and a part of helix E was impaired in the P88K/A90L/S92K/A94L mutant by the formation of helix F. The speed of the translocation of helix H, a folding event that occurs during the transition from the I2 intermediate to the native states, was slowed down by this mutation, probably due to the consequent stabilization of helix F in the I2 intermediate. By contrast, it was reported that occupying the empty heme cavity of apoMb with larger side chains as tryptophan has, such as occurs in the H64F/P88A/L89W/V68W mutations of apoMb, increased folding rates.\textsuperscript{85}

5.2 Apooleghemoglobin a (apoLb) vs. apoMb folding. Protein folding of apoLb, derived from soybean, and apoMb, derived from sperm whale were compared. These proteins have a similar native structure and function, but have a different evolutionary origin. It was shown that the structure of the kinetic intermediate (probably I2 intermediate) was different between the two proteins; the structure in apoLb was composed of helices G and H, and a part of helix E (Fig. 8A), whereas that of apoMb was helices A, B, G and H\textsuperscript{84} (Fig. 8B). Comparison between the kinetic and equilibrium intermediates of apoLb indicated that both structures significantly differ, unlike those of apoMb,\textsuperscript{85} and the equilibrium intermediate of apoLb contains helices A, B, G and H similar to that of apoMb. However, it might be possible that the structures of the apoLb intermediates are more sensitive to the solution pH than those of apoMb are. Another point is that the refolding of apoLb contains another slower phase (0.64 s\textsuperscript{-1}) outside of the main phase (2.04 s\textsuperscript{-1}).\textsuperscript{85} Non-native structures, indicated from the protection of the region at the CE-loop in the intermediate, may slow down the folding kinetics of apoLb.\textsuperscript{85}

5.3 Switching the helical structures in the I2 intermediate via protein engineering. The kinetic intermediate (probably I2) for apoLb was composed of helices G and H and a part of helix E.\textsuperscript{84} In order to produce the apoLb-like apoMb intermediate, mutations were designed using AABUF calculation\textsuperscript{81} (Fig. 9). The mutations, including the destabilized
helix A (L11G and W14G) and stabilized helix E (A71L and G73W), were then created (Fig. 9A). Thus, the helix A and helix E switching in the I2 intermediate was created, and compensatory mutations were also included to stabilize the native state. As predicted by AABUF (Fig. 9C), the experimental results of the mutant revealed that the structure of the I2 intermediate was composed of helices B, E, G and H (Fig. 9B) instead of helices A, B, G and H.86) This is an successful example of structurally designing a kinetic intermediate. Similar has already been shown where the folding process of protein G and protein L was switched by mutations of residues at hairpins 1 and 2.87)
6. Conclusions

The combination strategy of pH-pulse labeling and NMR is a powerful tool for elucidating the transient structure of the intermediate during the protein folding at each residue. The improved method was developed to get more images for transient structure of apoMb. In particular, the DMSO methodology combined with pH-pulse labeling, which has more experimental data points, revealed that the structure of the I2 intermediate composed of helices A, B, G and H in apoMb is similar to that of the equilibrium intermediate I2 for apoMb are indicated with high (red sphere), medium (green sphere), and low protections (blue sphere). The locations of differing protections between apoLb and apoMb intermediates are indicated by blue or red rings for helices A and B or for a part of helix E. The residue numbers (black, green, and blue) of the probe residues are indicated for high, medium, and low protections, respectively. Figure 8A<sup>39</sup> and Figure 8B<sup>50</sup> were produced based on the reported data.

It was suggested that the slow folding of apoMb via the I2 intermediate could be due to the formation of the non-native structure induced by helix H translocation. Furthermore, the N-terminal part of helix B can be a weak point in the apoMb folding. The premature structure in helix B interrupts the docking of helix E residues to the core region composed of A, B, G and H helices. In the stabilized helix B mutant, the GH-turn adopts the native-like structure, leading to the faster folding. By contrast, helix F formation observed in the P88K/A90L/S92K/A94L mutant in the I2 intermediate stabilized the non-native contact including the GH-turn, resulting in slower folding. Non-native structures in the I2 intermediate, which might be involved in its misfolding and aggregation, were confirmed by the folding studies on several alanine-scanning mutants at the core of helices A, B, E, G and H; they were also confirmed by FRET and Cys-quenching studies. Both native-like and non-native structures in the I2 intermediate of apoMb are important for the folding process of apoMb (Table 1). This information will contribute to the engineering of new proteins, perhaps someday including a functionally enhanced “super-myoglobin”.

Fig. 8. Difference of the I2 intermediate structure between (A) apoLb and (B) apoMb. Mapping of the proton occupancy derived from the pH-pulse labeling experiment on the crystal structure of Mb and Lb (PDB: 1BIN). (A) The kinetic intermediate for apoLb and (B) the kinetic intermediate I2 for apoMb are indicated with high (red sphere), medium (green sphere), and low protections (blue sphere).
Acknowledgements

I deeply thank Drs. Peter E. Wright and H. Jane Dyson for giving me the long-term opportunity to perform the experiments and for their helpful discussion of apoMb folding at The Scripps Research Institute, La Jolla, CA 92037, U.S.A.

References


9) Dill, K.A. and Chan, H.S. (1997) From Levinthal to...
Folding of apomyoglobin monitored by pH-pulse labeling and NMR

No. 1 | 25


(Received Aug. 9, 2016; accepted Oct. 31, 2016)
Profile

Chiaki Nishimura was born in Tokyo in 1960. He graduated from the Faculty of Pharmaceutical Sciences, the University of Tokyo, in 1984. He got the MS degree by working on the antibiotics for cancer chemotherapy under Prof. Nobuo Tanaka, and received Ph.D. degree for the project of a new inactivating enzyme of antibiotics under Prof. Hideyo Yamaguchi. After such a graduate student’s life in the Institute of Applied Microbiology, the University of Tokyo, he got a position as a research associate in the physical chemistry, Faculty of Pharmaceutical Sciences, the University of Tokyo, under Prof. Yoji Arata. He started his structural biology career with Prof. Yoji Arata. In 1997, he moved to the Department of Chemistry and Biochemistry, University of California, Santa Cruz, and he was working on the protein folding with Prof. Anthony L. Fink. The Uehara Memorial Foundation supported his life in the foreign country as a visiting researcher in 1997. In 1998, he moved to San Diego as a second postdoc career, and he was a research associate, subsequently a research associate senior, and finally a staff scientist (2002) in the Department of Molecular Biology in The Scripps Research Institute (TSRI), La Jolla, California until 2009. His main work that is presented in this review is folding studies on apomyoglobin using NMR with Prof. Peter E. Wright and Prof. H. Jane Dyson. Finally, in 2009 after more than 10 years presence in United States, he moved back to Japan as a professor in the Faculty of Pharmaceutical Sciences, Teikyo Heisei University. He is now interested in the research on the protein folding and misfolding, and continues many collaborations with the researchers.