Native Protein Sequences Are Designed To Destabilize Folding Intermediates †

ARTICLE *in* BIOCHEMISTRY · MARCH 2006

Impact Factor: 3.02 · DOI: 10.1021/bi0523714 · Source: PubMed

1 AUTHOR:

Yasuhiro Isogai
Toyama Prefectural University
59 PUBLICATIONS **808** CITATIONS

Native Protein Sequences Are Designed To Destabilize Folding Intermediates†

Yasuhiro Isogai*

Biometal Science Laboratory, RIKEN Harima Institute, Mikazuki, Sayo, Hyogo 679-5148, Japan, and Biomolecular Characterization Team, Advanced Development and Supporting Center, RIKEN Wako Institute, Hirosawa, Wako-shi, Saitama 351-0198, Japan

Received November 21, 2005; Revised Manuscript Received January 5, 2006

ABSTRACT: Hydrophobic core mutants of sperm whale apomyoglobin were constructed to investigate the amino acid sequence features that determine the folding properties. Replacements of all of the Ile residues with Leu and of all of the Ile and Val residues with Leu decreased the thermodynamic stability of the folded states against the unfolded states but increased the stability of the folding intermediates against the unfolded states, indicating that the amino acid composition of the protein core is important for the protein stability and folding cooperativity. To examine the effect of the arrangement of these hydrophobic residues, mutant proteins were further constructed: 12 sites out of the 18 Leu, 9 Ile, and 8 Val residues of the wild-type myoglobin were randomly replaced with each other so that the amino acid compositions were similar to that of the wild-type protein. Four mutant proteins were obtained without selection of the protein properties. These residue replacements similarly resulted in the stabilization of both the intermediate and folded states against the unfolded states, as compared to the wild-type protein. Thus, the arrangements of the hydrophobic residues in the native amino acid sequence are selected to destabilize the folding intermediate rather than to stabilize the folded state. The present results suggest that the two-state transition of protein folding or the transient formation of the unstable intermediate, which seems to be required for effective production of the functional proteins, has been a major driving force in the molecular evolution of natural globular proteins.

The amino acid sequences of native proteins determine their tertiary structures, thermodynamic stabilities, and physiological functions. Many mutational studies have been performed to elucidate the relationships of the sequences with these protein features. According to these studies, small numbers of amino acid residues within a whole protein sequence are essential to maintain the protein stability and/or function, whereas other regions of the sequence contain informational redundancy for the features. For example, the stabilities and functions of some proteins exhibit significant tolerance to residue substitutions or "mutational robustness" (*1*−*5*). In the cases of myoglobins (Mbs)[1] and hemoglobins (Hbs), which share the common function of reversible O_2 binding with an identical globin fold but exhibit a wide range of sequence variability (*6*, *7*), only two residues (Phe-43 and His-93 in sperm whale Mb) are preserved in all members of the natural globin family and are essential to maintain the protein function. Sequence alignments of the globin family members also indicated that several other residues are well preserved and seem to be important for protein stability and function. In fact, single and double mutations of certain residues in myoglobins affected the heme-related functions and/or the conformational stability and revealed that they are important for the maintenance of these properties of native globins (*8*). Folding intermediates of apomyoglobins (apoMbs) have been observed and characterized by CD, NMR, and small-angle X-ray scattering measurements under equilibrium and transient conditions (*9*−*16*). These studies showed that the folding intermediates have structural properties that are characteristic of the molten globule, which has reduced but still substantial amounts of secondary structures in some multiple conformational states as compared with the folded apoMbs (*17*), and that the formation of the intermediate is an essential step in the protein folding. However, the relationships between the protein sequences and the folding properties remain to be elucidated.

We have computationally designed amino acid sequences to fold into a globin fold, using a knowledge-based structure−sequence compatibility function (*18*, *19*). The synthesized designed globins have α-helix contents and overall molecular shapes that are similar to those of the targeted fold and bind a single heme per molecule as designed. They form an extremely stable folding intermediate and lack the nativelike structural uniqueness, even in the fully folded state. The structural specificity and thermodynamic stability of the intermediate are dependent on the hydrophobic amino acid composition in the protein core (*19*). Thus, the author has been interested in how the native globin sequences modulate the stabilities and achieve the structural uniqueness. In the present work, hydrophobic core mutants of sperm whale Mb were constructed, and the sequence features determining the folding properties of the native protein were investigated.

† This work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Culture, and Sports of Japan.

* To whom correspondence should be addressed at the Biometal Science Laboratory (Wako Branch), Room 101, Main Research Building, RIKEN, Hirosawa 2-1, Wako, Saitama 351-0198, Japan. Phone: 81-48-467-4583. Fax: 81-48-467-9649. E-mail: yisogai@riken.jp.

[1] Abbreviations: CD, circular dichroism; Gdn-HCl, guanidine hydrochloride; Mb, myoglobin.

10.1021/bi0523714 CCC: $33.50
Published on Web 02/03/2006

MATERIALS AND METHODS

The nucleotide sequences of artificial genes encoding sperm whale Mb and its core mutants were written by computational back-translation of their amino acid sequences, using the optimal codons of *Escherichia coli* (*20*, *21*) and mixed codons, (CAG)TT, for Leu, Ile, and Val, in which the bases in parentheses signify an equimolar mixture of the indicated bases. The *Nde*I and *Hin*dIII restriction sites were introduced at the 5′- and 3′-ends, respectively. Each full-length gene was synthesized from several oligonucleotides containing either template or complementary sequences of the gene by the polymerase chain reaction (PCR) with Pyrobest DNA polymerase (Takara Bio Inc.). The PCR products were digested with the restriction enzymes *Nde*I and *Hin*dIII and were ligated into the pRSET-C vector (Invitrogen) digested with the same restriction enzymes using T4 DNA ligase (Takara Bio Inc.).

The vector carrying a Mb mutant gene was incorporated into an *E. coli* strain, SCS1 (Stratagene). The transformed cells were isolated by growing overnight on a Luria−Bertani (LB) agar plate containing 0.1 mg/mL ampicillin. Antibiotic-resistant single colonies were randomly picked, separately transferred into LB media containing 0.1 mg/mL ampicillin, and cultivated for 6−10 h for the amplification of each gene. The vector DNA was purified from the cultures using a Quantum Prep plasmid miniprep kit (Bio-Rad). The Mb mutant genes were sequenced by a dye terminator method, using an ABI PRISM 377XL DNA sequencer (Applied Biosystems). The proteins were synthesized by expression of the genes in an *E. coli* strain, BL21(DE3), and were purified as described previously (*18*). The identity and purity of the proteins were verified by laser desorption mass spectrometry and analytical reverse-phase HPLC. The concentrations of the apoMbs were determined spectrophotometrically using $\epsilon_{280} = 15200\ M^{-1}\ cm^{-1}$.

UV−visible absorption spectra were recorded using a Hitachi U-3000 spectrometer and quartz cuvettes with a 1.0 cm path length. Circular dichroism (CD) spectra were recorded at 20 °C using a JASCO J700 spectropolarimeter and a rectangular quartz cuvette with a 0.2 cm path length. NMR measurements were performed using a Bruker DRX600 spectrometer at 30 °C, with a sample solution containing 0.2 mM protein, 10 mM sodium phosphate (pH 7), 150 mM NaCl, and 9% D_2O.

Denaturation experiments with guanidine hydrochloride (Gdn-HCl) were carried out by monitoring the CD signal intensity at 222 nm and a protein concentration of 5 μM in a buffer solution containing 10 mM Tris-HCl (pH 8), 200 mM NaCl, and 0.1% octyl glucopyranoside. The denaturation data were analyzed using a theoretical curve derived from the three-state model:

$$F \overset{K_1}{\leftrightarrow} I \overset{K_2}{\leftrightarrow} U \quad (1)$$

where F, I, and U represent the folded, intermediate, and unfolded states, respectively, and K_1 and K_2 are the equilibrium constants of F ⇔ I and of I ⇔ U, respectively (K_1 = [F]/[I]; K_2 = [I]/[U]) (*19*). K_1 and K_2 give ΔG_1 and ΔG_2, the free energy of the fully folded state relative to the intermediate and that of the intermediate relative to the unfolded state, respectively. ΔG_1 and ΔG_2 are assumed to depend linearly on the denaturant concentration:

$$\Delta G_1 = G_F - G_I = -RT \ln K_1 = \Delta G^\circ{}_1 + m_1 x \quad (2)$$

$$\Delta G_2 = G_I - G_U = -RT \ln K_2 = \Delta G^\circ{}_2 + m_2 x \quad (3)$$

where $\Delta G^\circ{}_1$ and $\Delta G^\circ{}_2$ are the ΔG_1 and ΔG_2 values in the absence of denaturant, respectively, and m_1 and m_2 are the dependence of ΔG_1 and ΔG_2 on x, the denaturant concentration, respectively. From these relationships, the following formulas were obtained:

$$\alpha = 1/\{1 + \exp A + \exp(-B)\} \quad (4)$$

$$\beta = \exp(-B)/\{1 + \exp A + \exp(-B)\} \quad (5)$$

where α and β are the fractions of the intermediate and the unfolded state, respectively, $A = -(\Delta G^\circ{}_1 + m_1 x)/RT$, and $B = -(\Delta G^\circ{}_2 + m_2 x)/RT$. Then, y, the ratio of the helical content in the transition region to the total helical content of the folded form, is calculated as

$$y = 1 - \alpha - \beta + \gamma\alpha = (\gamma + \exp A)/\{1 + \exp A + \exp(-B)\} \quad (6)$$

where γ is the ratio of the helical content of the intermediate to that of the folded state and the helical content of the unfolded state is assumed to be zero. The theoretical curves derived from eq 6 were fitted to the denaturation data to obtain the thermodynamic parameters $\Delta G^\circ{}_1$, $\Delta G^\circ{}_2$, m_1, and m_2. The sum of $\Delta G^\circ{}_1$ and $\Delta G^\circ{}_2$ ($\Delta G^\circ{}_{1+2} = \Delta G^\circ{}_1 + \Delta G^\circ{}_2$) gives the free energy change from the unfolded state to the folded state in the absence of denaturant and measures the stability of the folded state against the unfolded state irrespective of the stability of the intermediate.

RESULTS

The native globin sequences are rich in Leu, Ile, and Val, which are chemically similar hydrophobic residues. Sperm whale Mb contains 18 Leu, 9 Ile, and 8 Val residues in the sequence of 153 amino acids, in addition to other hydrophobic residues, i.e., 2 Met, 6 Phe, 2 Trp, and 3 Tyr residues (Figures 1 and 2A). The arrangements of the Leu, Ile, and Val residues in the sequence were the focus of the present study. First, two Mb mutants were constructed with different compositions of these hydrophobic residues. In one mutant, Mb-L/I, the 9 Ile residues were all substituted with Leu; in the other mutant, Mb-L/IV, the 9 Ile and 8 Val residues were all substituted with Leu, as shown in Figure 1. The wild-type and mutant proteins were synthesized using recombinant *E. coli* cells and were purified as the apo forms. The α-helix contents of these mutants were almost identical to that of the wild-type apoMb (Table 1). To examine the effects of the residue replacements on the equilibrium among the folded, intermediate, and unfolded states, quantitative disruption of the secondary structures of the apoproteins was carried out using a denaturant Gdn-HCl, and the dependence of the α-helix contents on the Gdn-HCl concentration was measured by CD spectroscopy (Figure 3A). In comparison with the wild-type protein, the two mutants generated broad denaturation curves with midpoint concentrations of Gdn-HCl (C_m) at approximately 1.7 M, which were larger than the

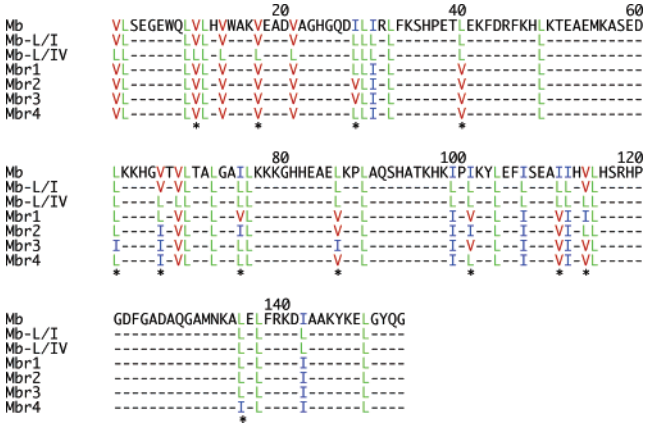

FIGURE 1: Amino acid sequences of sperm whale Mb and its mutants. Leu, Ile, and Val are green, blue, and red, respectively. The other amino acids are constant and indicated by hyphens (-) in the mutant sequences. The asterisks (*) under the sequence of Mbr4 indicate the sites selected for the shuffling among the Leu, Ile, and Val residues in Mbr1−4.

Table 1: Mean Residue Ellipticity and Thermodynamic Parameters of Wild-Type and Mutant Apomyoglobins for Gdn-HCl Denaturation[a]

apoMb	Θ_{222} (deg cm^2 dmol^{-1})	$\Delta G°_1$ (kcal mol^{-1})	m_1 (kcal mol^{-1} M^{-1})	$\Delta G°_2$ (deg cm^2 dmol^{-1})	m_2 (kcal mol^{-1} M^{-1})	$\Delta G°_{1+2}$ (deg cm^2 dmol^{-1})
wild type	19700	−3.6	3.4	−0.11	0.43	−3.7
Mb-L/I	19800	−1.9	1.2	−0.67	0.43	−2.6
Mb-L/IV	20000	−1.2	2.1	−1.9	0.60	−3.1
Mbr1	19500	−2.9	5.2	−3.1	1.4	−5.9
Mbr2	18800	−2.1	3.3	−3.3	1.7	−5.5
Mbr3	18500	−3.6	4.5	−2.5	1.3	−6.1
Mbr4	19000	−2.1	2.9	−3.2	1.5	−5.4

[a] Standard deviations for the data of Θ_{222}, $\Delta G°$, and m were within 1000 deg cm^2 dmol^{-1}, 0.2 kcal mol^{-1}, and 0.2 kcal mol^{-1} M^{-1}, respectively. $\Delta G°_{1+2}$ is the free energy of the fully folded state relative to the unfolded state ($\Delta G°_1 + \Delta G°_2$).

C_m value of 1.0 M for the wild-type apoMb. The denaturation curves were analyzed assuming the three-state transitions and that the folding intermediates had 60% of the α-helix content of the folded states ($\gamma = 0.6$ in eq 6 in Materials and Methods). The obtained values of $\Delta G°_1$ and $\Delta G°_2$, the free energies of the folded states (F) relative to the intermediates (I) and of I relative to the unfolded states (U) in the absence of denaturant, respectively, are listed in Table 1. The denaturant m values, m_1 and m_2, for the transitions from F to I and from I to U, respectively, are also listed in the table. The $-\Delta G°_1$ and $-\Delta G°_{1+2}$ values of Mb-L/I and Mb-L/IV were smaller than those of the wild type; i.e., the folded states were destabilized against both the intermediate and unfolded states by the residue replacements (also see Figure 4). The m_1 values of these mutants were also smaller than that of the wild type, indicating that the folding cooperativity from the intermediate to the folded state was disrupted by the residue replacements. On the other hand, the $-\Delta G°_2$ values were larger than that of the wild type; i.e., the intermediates were stabilized against the unfolded states. These results indicate that the amino acid composition of the protein core was important for the cooperative folding and destabilization of the folding intermediate as well as for the stability of the folded state. The residue replacements may disturb the core packing of the folded state to destabilize it but increase the conformational freedom of the intermediate to stabilize it (also see Discussion).

To examine the site specificity of the hydrophobic residues in the amino acid sequence for the structural properties, mutant proteins were constructed in which four Leu sites (L40, L61, L86, L135), four Ile sites (I28, I75, I101, I111), and four Val sites (V10, V17, V66, V114) were randomly substituted with each other. These 12 sites were selected because they are dispersed throughout the entire sequence (Figure 1) and do not contact each other in the tertiary structure of the folded form (Figure 2B). The four mutant proteins, Mbr1−4, were obtained without selection for any protein properties (see Materials and Methods). Their amino acid compositions are almost identical to that of the wild type, and their sequences are similar to, but different from, the wild type, with an average identity of 94.9 ± 0.6%.

Mbr1−4 generated denaturation profiles that were similar to each other, but apparently different from that of the wild-type protein, as shown in Figure 3B. The thermodynamic data obtained from the denaturation experiments are listed in Table 1. The data show that the residue replacements under the preserved composition in Mbr1−4 increased the stabili-

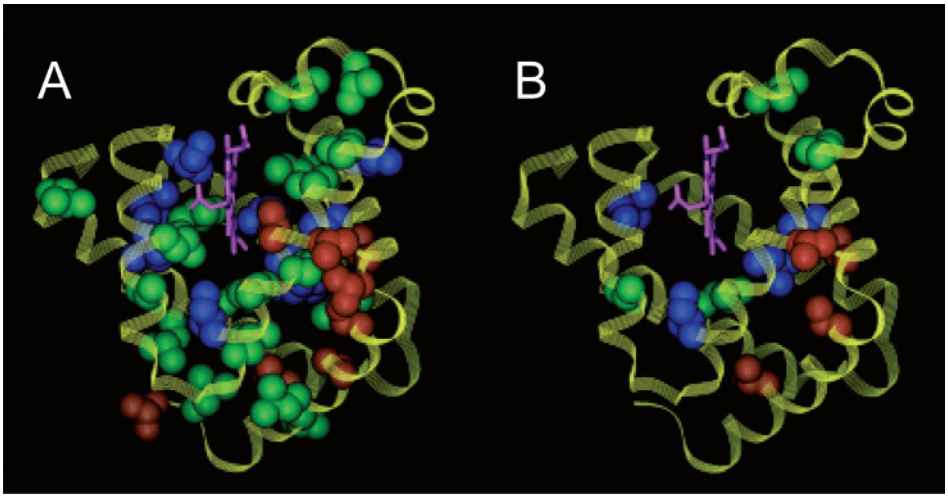

FIGURE 2: Leu (green), Ile (blue), and Val (red) residues in the tertiary structure of sperm whale Mb (1mbd). The side chains of all of these residues (A) and of the sites selected for the residue shuffling in Mbr1−4 (B) are indicated in the CPK rendering with the main-chain backbone (yellow ribbons).

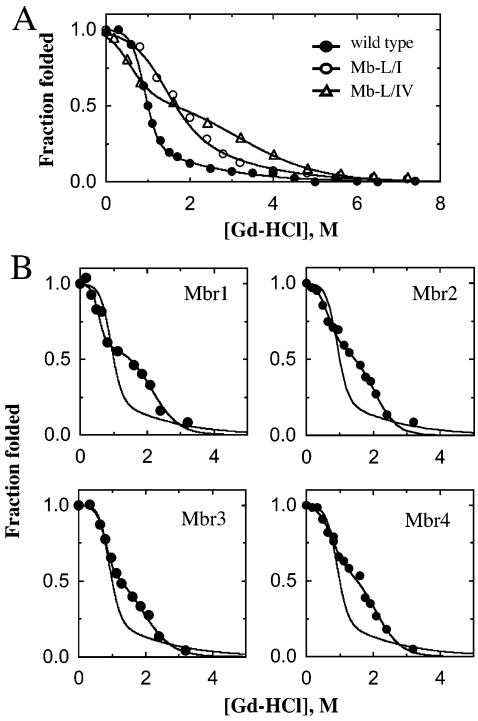

FIGURE 3: Denaturation profiles of native and mutant apoMbs. The folded fractions estimated by the CD signal intensity at 218 nm are plotted against the Gdn-HCl concentration. The data for Mb-L/I (○), Mb-L/IV (●), and wild-type Mb (△) in (A) and those for Mbr1−4 (Mr1, upper left; Mbr2, upper right; Mbr3, lower left; Mbr4, lower right) in (B) are displayed separately for clarification. The data were fitted by theoretical curves for the three-state transitions to obtain the thermodynamic parameters listed in Table 1 (see Materials and Methods). The gray lines without data points in (B) represent the theoretical curve for wild-type Mb in (A) and are shown for comparison.

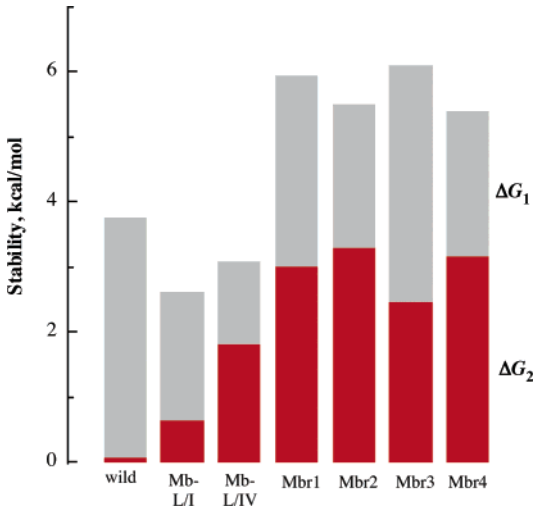

FIGURE 4: Schematic diagram of the thermodynamic stabilities of native and mutant apoMbs. ΔG°_1 and ΔG°_2 are the stabilities of the folded state against the intermediate and of the intermediate against the unfolded state, respectively.

ties of the folded states against the unfolded states ($-\Delta G^\circ_{1+2}$) as well as those of the intermediates against the unfolded states ($-\Delta G^\circ_2$), whereas the stabilities of the folded states against the intermediates ($-\Delta G^\circ_1$) were comparable to that of the wild-type protein in a case of Mbr3 and lower than that of the wild-type protein in the cases of Mbr1, Mbr2, and Mbr4. The m_1 values of Mbr1 and Mbr3 were larger than that of the wild type, but the m_1 values of Mbr2 and Mbr4 were slightly smaller, indicating that these residue replacements have no single tendency for their effects on the folding cooperativity. The free energy differences among the folded, intermediate, and unfolded states of the wild-type and mutant Mbs are schematically displayed in Figure 4. The average values of $-\Delta G^\circ_1$, $-\Delta G^\circ_2$, and $-\Delta G^\circ_{1+2}$ for Mbr1−4 are 2.7 ± 0.7, 3.0 ± 0.4, and 5.7 ± 0.3 kcal/mol, respectively.

In comparison with wild-type Mb, the stabilities of the intermediates ($-\Delta G^\circ_2$) of the four mutants Mbr1−4 were similarly and significantly increased, by approximately 30-fold on average. On the bases of the statistics of the data, the appearance of such lower stability as 0.1 kcal/mol of the wild-type Mb intermediate is quite rare; assuming a Gaussian distribution of the stability against the sequence, the probability of a stability less than 0.1 kcal/mol is estimated to be approximately 5×10^{-15}, which is much smaller than $3^{-12} = 1.8 \times 10^{-6}$, the probability of the native sequence of the mutated sites (either L, I, or V for the 12 sites). Thus, the native placements of the Leu, Ile, and Val residues in the sequence (Figure 2) and also those of the surrounding residues in the tertiary structure are highly selected to destabilize the intermediate.

The proton NMR spectrum of Mbr3, the most stable mutant, was measured and compared with that of wild-type Mb. These spectra reflected the residue substitutions and were distinguishable from each other (not shown). However, the signal quality of the mutant spectrum was comparable to that of the wild-type, indicating that the mutant was well folded and monomeric. Thus, the possibility that the stabilization of the folded states of Mbr3 and the other mutants occurs by self-aggregation can be excluded.

DISCUSSION

In protein tertiary structures, steric interactions between residues are essential to decrease the conformational entropy of the side chains and to achieve the structural uniqueness (*22*−*24*). The effects of these residue−residue contacts depend on both the residue pairs and secondary structures; e.g., the interaction between Ile and Leu residues restricts their side-chain conformations more than the interaction between two Leu residues on an α-helix (*24*). The decrease in conformational entropy itself is disadvantageous for the stability, and hence smaller changes of the conformational entropy during the folding of artificial proteins contribute to their higher stability, in comparison with native proteins (*19*). Thus, the stabilization of the folding intermediates of the Mb mutants constructed here seems to arise from the higher conformational entropy accompanied by the increased side-chain flexibility in their hydrophobic cores. In the folded states, however, the enhanced flexibility or the fluctuation of the side chains with higher conformational entropy may disturb the specific residue−residue interactions that contribute to the stability by enthalpic effects; i.e., these entropic and enthalpic effects can compensate for each other. In the cases of the present Mb mutants Mb-L/I and Mb-L/IV, which exhibited lower stabilities of the folded states relative to their unfolded states than the wild type, the loss of specific residue−residue interactions in the folded states had a more significant effect on the stability than the increase in the conformational entropy. For the other mutants, Mbr1−4,

which exhibited higher stabilities of the folded states against their unfolded states than the wild type, the effect of the entropy increase seems to be significant. The increases in the stabilities of the folded states against the unfolded states approximately coincide with those of the intermediates (Figure 4). Thus, Mbr1−4 in the folded states may also preserve the enhanced flexibility of the intermediates, but the disturbance of the core packing by the residue replacements is not significant, due to the conserved amino acid composition.

The present results demonstrated that native protein sequences encode the structural properties of the folding intermediate, as well as those of the folded states, and are designed to destabilize the intermediate by natural selection. These findings are consistent with inferences from studies on artificial proteins (*18*, *19*, *25*). The formation of the Mb intermediate is an essential step for the protein folding (*13*−*15*); in the hydrophobic core of the formed intermediate, the correct packing of the side chains is effectively searched, which is the rate-limiting step of the folding. However, the intermediate is in a molten globule state and lacks a well-packed protein core, and thus the hydrophobic residues are partially exposed to the surrounding water. This induces nonspecific interactions between molecules to generate misfolded, insoluble aggregates under physiological conditions. Thus, the destabilization of the intermediate is favorable for the effective production of functional proteins in vivo and may be an important selection pressure for protein molecular evolution.

The Mb mutants constructed here bound heme, and these holo-Mbs formed oxygenated derivatives in the ferrous states (not shown). As the folding intermediate of Mb does not bind heme correctly and lacks physiological functions, the stability of the function can be correlated with ΔG_1, rather than ΔG_{1+2}. Thus, the maintenance of substantially negative values of ΔG_1 to retain the protein function is another major driving force in natural selection. The present results depicted in Figure 4 suggest that the stabilities of the folded and intermediate states (ΔG_1 and ΔG_2) can be independently modulated by sequence changes.

ACKNOWLEDGMENT

The author thanks Ms. Yasue Ichikawa and Ms. Rie Nakazawa (Bioarchitect Research Group, RIKEN) for DNA sequencing, Ms. Tomoe Nishimura and Dr. Naoshi Dohmae (Biomolecular Characterization Team, RIKEN) for mass spectrometry measurements, Dr. Ayako Ohno (Biometal Science Laboratory, RIKEN Harima Institute) for NMR measurements, and Prof. Motonori Ota (Tokyo Institute of Technology) for critical discussions.

REFERENCES

1. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A., and Sauer, R. T. (1990) Deciphering the message in protein sequences: tolerance to amino acid substitutions, *Science 247*, 1306−1310.
2. Brown, B. M., and Sauer, R. T. (1999) Tolerance of Arc repressor to multiple-alanine substitutions, *Proc. Natl. Acad. Sci. U.S.A. 96*, 1983−1988.
3. Gassner, N. C., Baase, W. A., and Matthews, B. W. (1996) A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme, *Proc. Natl. Acad. Sci. U.S.A. 93*, 12155−12158.
4. Axe, D. D., Foster, N. W., and Fersht, A. R. (1996) Active barnase variants with completely random hydrophobic cores, *Proc. Natl. Acad. Sci. U.S.A. 93*, 5590−5594.
5. Kuroda, Y., and Kim, P. S. (2000) Folding of bovine pancreatic trypsin inhibitor (BPTI) variants in which almost half the residues are alanine, *J. Mol. Biol. 298*, 493−501.
6. Bashford, D., Chothia, C., and Lesk, A. M. (1987) Determinants of a protein fold. Unique features of the globin amino acid sequences, *J. Mol. Biol. 196*, 199−216.
7. Chothia, C., and Lesk, A. M. (1980) How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins, *J. Mol. Biol. 136*, 225−270.
8. Springer, B. A., Sligar, S. G., Olson, J. S., and Phillips, G. N. (1994) Mechanisms of ligand recognition in myoglobin, *Chem. Rev. 94*, 699−714.
9. Jamin, M., and Baldwin, R. L. (1996) Refolding and unfolding kinetics of the equilibrium folding intermediate of apomyoglobin, *Nat. Struct. Biol. 3*, 613−618.
10. Jamin, M., Yeh, S.-R., Rousseau, D. L., and Baldwin, R. L. (1999) Submillisecond unfolding kinetics of apomyoglobin and its pH 4 intermediate, *J. Mol. Biol. 292*, 731−740.
11. Barrick, D., and Baldwin, R. L. (1993) Three-state analysis of sperm whale apomyoglobin folding, *Biochemistry 32*, 3790−3796.
12. Kay, M. S., and Baldwin, R. L. (1999) Specificity of native-like interhelical hydrophobic contacts in the apomyoglobin intermediate, *Proc. Natl. Acad. Sci. U.S.A. 96*, 2007−2012.
13. Jennings, P. A., and Wright, P. E. (1993) Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin, *Science 262*, 892−896.
14. Eliezer, D., Yao, J., Dyson, H. J., and Wright P. E. (1998) Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding, *Nat. Struct. Biol. 5*, 148−155.
15. Tsui, V., Garcia, C., Cavagnero, S., Siuzdak, G., Dyson, H. J., and Wright, P. E. (1999) Quench-flow experiments combined with mass spectroscopy show apomyoglobin folds through an obligatory intermediate, *Protein Sci. 8*, 45−49.
16. Uzawa, T., Akiyama, S., Kimura, T., Takahashi, S., Ishimori, K., Morishima, I., and Fujisawa, T. (2002) Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of alpha-helical content and compactness, *Proc. Natl. Acad. Sci. U.S.A. 101*, 1171−1176.
17. Arai, M., and Kuwajima, K. (2000) Role of the molten globule state in protein folding, *Adv. Protein Chem. 53*, 209−282.
18. Isogai, Y., Ota, M., Fujisawa, T., Izuno, H., Mukai, M., Nakamura, H., Iizuka, T., and Nishikawa, K. (1999) Design and synthesis of a globin fold, *Biochemistry 38*, 7431−7443.
19. Isogai, Y., Ishii, A., Fujisawa, T., Ota, M., and Nishikawa, K. (2000) Redesign of artificial globins: effects of residue replacements at hydrophobic sites on the structural properties, *Biochemistry 39*, 5683−5690.
20. Springer, B. A., and Sligar, S. G. (1987) High-level expression of sperm whale myoglobin in *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A. 84*, 8961−8965.
21. Sharp, P. M., Cowe, E., Higgins, D. G., Shields, D. C., Wolfe, K. H., and Wright, F. (1988) Codon usage patterns in *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within species diversity, *Nucleic Acids Res. 16*, 8207−8211.
22. Doig, A. J., and Sternberg, M. J. E. (1995) Side-chain conformational entropy in protein folding, *Protein Sci. 4*, 2247−2251.
23. Furukawa, K., Oda, M., and Nakamura, H. (1996) A small engineered protein lacks structural uniqueness by increasing the side-chain conformational entropy, *Proc. Natl. Acad. Sci. U.S.A. 93*, 13583−13588.
24. Isogai, Y., Ota, M., Ishii, A., Ishida, M., and Nishikawa, K. (2002) Identification of amino acids involved in protein structural uniqueness: implication for de novo protein design, *Protein Eng. 15*, 555−560.
25. Scalley-Kim, M., and Baker, D. (2004) Characterization of the folding energy landscapes of computer generated proteins suggests high folding free energy barriers and cooperativity may be consequences of natural selection, *J. Mol. Biol. 338*, 573−583.

BI0523714