Redesign of Artificial Globins: Effects of Residue Replacements at Hydrophobic Sites on the Structural Properties†

Yasuhiro Isogai,*,‡ Anna Ishii,§ Tetsuro Fujisawa,|| Motonori Ota,⊥ and Ken Nishikawa⊥

The Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Wako, Saitama 351-0198, Japan, Department of Physics, Faculty of Science, Gakushuin University, Mejiro, Toshima-Ku, Tokyo 170-0031, Japan, The Institute of Physical and Chemical Research (RIKEN), RIKEN Harima Institute, Mikazuki-cho, Sayo, Hyogo 679-5148, Japan, and National Institute of Genetics, Yata, Mishima, Shizuoka 411-8540, Japan

Received November 23, 1999; Revised Manuscript Received February 29, 2000

ABSTRACT: Artificial sequences of the 153 amino acids have been designed to fit the main-chain framework of the sperm whale myoglobin (Mb) structure based on a knowledge-based 3D-1D compatibility method. The previously designed artificial globin (DG1) folded into a monomeric, compact, highly helical and globular form with overall dimensions similar to those of the target structure, but it lacked structural uniqueness at the side-chain level [Isogai, Y., Ota, M., Fujisawa, T., Izuno, H., Mukai, M., Nakamura, H., Iizuka, T., and Nishikawa, K. (1999) *Biochemistry 38*, 7431−7443]. In this study, we redesigned hydrophobic sites of DG1 to improve the structural specificity. Several Leu and Met residues in DG1 were replaced with β-branched amino acids, Ile and Val, referring to the 3D profile of DG1 to produce three redesigned globins, DG2−4. These residue replacements resulted in no significant changes of their compactness and α-helical contents in the absence of denaturant, whereas they significantly affected the dependence of the secondary structure on the concentration of guanidine hydrochloride. The analyses of the denaturation curves revealed higher global stabilities of the designed globins than that of natural apoMb. Among DG1−4, DG3, in which 11 Leu residues of DG1 are replaced with seven Ile and four Val residues, and one Met residue is replaced with Val, displayed the lowest stability but the most cooperative folding−unfolding transition and the most dispersed NMR spectrum with the smallest line width. The present results indicate that the replacements of Leu (Met) with the β-branched amino acids at appropriate sites reduce the freedom of side-chain conformation and improve the structural specificity at the expense of stability.

Structural uniqueness is characteristic of native proteins and is essential to express their biological functions. This view has been highlighted by recent progress in de novo protein design, which attempts to create an artificial amino acid sequence to fold into a desired three-dimensional structure independently of the native protein sequences (*1*−*3*). A number of de novo designed proteins have been reported to retain the targeted overall topology with significant secondary structure, but they have low structural specificity without the fixed conformations of side chains, as demonstrated by several experimental criteria, including limited dispersion of their nuclear magnetic resonance (NMR)[1] spectra and noncooperative folding−unfolding transition (*4*−*9*). These conformational states of the artificial designed proteins are called "gemisch" states, indicating their multiple ground-state conformations without any uniquely folded structure that is seen in native proteins under appropriate conditions (*10*). This difficulty in reproducing the nativelike uniqueness in designed proteins has been overcome by considering the combination of side-chain conformations and improving the packing specificity in hydrophobic cores in a few small structural motifs (*11*−*13*). For a sizable protein consisting of more than 100 amino acids, however, an exhaustive search of the side-chain conformations of various amino acids over all the residue sites requires too much computational task to complete within a reasonable CPU time, even if efficient optimization algorithms are used (*14*). Thus, to the establishment of a design methodology, it is essential to find hidden codes or simplified assumptions for relationships between sequences and unique structures, which will enable us to understand the principles of protein architecture.

In the previous report (*9*), we proposed a new computational method for designing an entire amino acid sequence

† This work was supported in part by the Biodesign Research Program and MR Science Program of RIKEN (to Y.I.), and by Grant-in-Aids for Scientific Research from the Ministry of Education, Science, Culture, and Sports of Japan (to Y.I., to M.O., and to K.N.).

* To whom correspondence should be addressed. Fax: 81-48-462-4660. E-mail: yisogai@postman.riken.go.jp.

‡ The Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa.

§ Department of Physics.

|| The Institute of Physical and Chemical Research (RIKEN), RIKEN Harima Institute.

⊥ National Institute of Genetics.

[1] Abbreviations: CD, circular dichroism; C_m, midpoint concentration of denaturant; 1D, one-dimensional; 3D, three-dimensional; DG, designed globin; Gd-HCl, guanidine hydrochloride; HPLC, high-performance liquid chromatography; Mb, myoglobin; M_r, relative molecular mass; NMR, nuclear magnetic resonance; R_g, radius of gyration; SWMb, sperm whale myoglobin; UV, ultraviolet.

10.1021/bi992687+ CCC: $19.00
Published on Web 04/22/2000

to adopt a desired structure based on a knowledge-based 3D-1D compatibility function (*15*). This function was composed of the following four terms: side-chain packing, hydration, hydrogen bonding, and local conformation potentials, which were normalized referring to the random environmental state (the virtual-denatured state) and were then described by a folding free-energy difference (ΔG) rather than by an energy of the folded state (G) (*16*). Assuming an asymmetric α-helical single-domain structure of sperm whale myoglobin (Mb) as the target, the optimal amino acid sequence to fit the main-chain framework has been searched by recursive generation of the protein 3D profile, which is a $20 \times N$ (sequence length) two-dimensional array representing the fitness of 20 amino acids to each structural site of a protein. The heme-binding site was designed by fixing His64 and His93 at the distal and proximal positions, respectively, and by penalizing residues that protrude into the space for heme with a repulsive function to reserve the space. The artificial sequences of the best-score amino acid at each position, generated during the calculation starting from the native sequence, were converged into the self-consistent sequence (SCS1), which shares 26% of the sequence with the native Mb. The apparent bumps among side chains found in the computer 3D model of SCS1 were removed by replacing some of the bumping residues with smaller ones according to the 3D profile, and finally the designed-globin-1 (DG1) sequence was obtained.

Our design method was validated by synthesis and characterization of DG1 (*9*). The synthesized DG1 folded into a monomeric, compact, highly helical and globular form with an overall molecular shape similar to, but slightly expanded from, that of the target structure in an aqueous solution, as demonstrated by size-exclusion chromatography, circular dichroism spectroscopy, and solution X-ray scattering analyses. Furthermore, DG1 bound a single heme per protein molecule with the global shape compacted, and it exhibited well-defined spectroscopic properties of six-coordinated low-spin heme proteins. However, it exhibited lower folding cooperativity in denaturation experiments and poor chemical-shift dispersion in the NMR spectrum. These properties of DG1 indicate higher conformational diversity of the side-chain structure and are characteristic of the gemisch states of artificial proteins. These results show that our method using an empirical potential function and a simple optimization algorithm is useful for designing an entire sequence to adopt a sizable globular protein structure. However, improvements of the method are needed to increase the structural specificity. DG1 provides a good system for examining trials to create a higher order structure as well as to add a nativelike function.

The highly specific structure of native proteins is formed mainly by the excellent packing of hydrophobic side chains in the protein cores or by the complementary disposition of residues with various shapes and sizes inside the protein (*17*, *18*). On the other hand, the DG1 sequence is rich in Leu and Met residues, as compared with native Mbs (*9*). These preferences for Leu and Met in our sequence selection are mainly due to their higher propensities for an α helix in the compatibility function used. Therefore, DG1 has a only single Ile and no Val, and the hydrophobic core is mainly composed of Leu and Met in the 3D model (see Figure 3). In the X-ray crystallographic structures of nine distantly related globins (1ash, 1eca, 1hlb, 1mbd, 1pbx, 2fal, 2gdm, 2hbg, and 3sdh), Ile and Val residues (nine Ile and eight Val in 1mbd) are distributed dispersedly in the hydrophobic region and constitute the core in combination with Leu and other hydrophobic residues. Theoretical and experimental studies on native protein structures have shown that the side-chain conformations of the β-branched side chains of Ile and Val residues on α-helices are more restricted by their interaction with the neighboring side chains and main chains than those of Leu and Met (*19*−*21*). Thus, we focused on the residue composition of the hydrophobic sites, and Leu and Met residues in DG1 were changed to Ile and Val. The structural properties of the three redesigned globins, DG2−4, were investigated and compared with those of DG1 and also of native Mb.

N	I	H	S	3D profile table
28	L	9	a	L I M F V W Y A C H R Q T E S G D N K P
29	L	9	a	L I M F V W Y A H C T R Q S N E D G K P
31	A	6	a	R L Q W A K M Y E N H I F D C V T G S P
39	L	8	a	L M I V A C R Q H T Y K N E S G D F P W
46	L	9	g	L F I H V W Y M A C S Q D T E N P G R K
68	L	9	a	W M L F I V Y H A C R T Q S E N D G K P
69	L	9	a	L I V M W F A Y C R H T Q S E G N D K P
76	M	9	a	M W F L I Y V H A R Q C S T N D K E G P
86	L	7	a	L I V M W F Y A C R H T Q E K N S D G P
108	M	6	a	M W F L Y R I Q A V H C E D N K T S G P
110	L	7	a	M L A I F H Y V C Q E T D W S G N K R P
112	L	6	a	L M I V F Y E A R Q W C T H D N K S G P
114	L	7	a	L I M W F V A Y C R H T Q S K E N G D P
115	L	8	a	L I F M W V Y A C R H Q T E K S N D G P
131	L	8	a	L I M F V W Y C H A R Q T E N K S D G P
137	L	6	a	L M W I V F R Y Q A C E H T N D S K G P
138	M	9	a	L M I F Y W Y A C H T Q R N E D S G K P
142	L	9	a	M L F I W V A C H Y T R Q N D E S G K P

FIGURE 1: 3D profile of the DG1 sequence for the sperm whale Mb structure at the sites replaced in DG2−4. The table has been excerpted from the complete table consisting of the entire rows for the 153 residue sites. The table contains residue numbers (N), amino acids of the input sequence, i.e., the DG1 sequence (I), hydration classes (H), secondary structures (S), and the 3D-profile table. In the profile tables, the 20 amino acids denoted by their one-letter code are arranged in order of the compatibility score from left to right at each site. The amino acids selected in the DG2−4 sequences are highlighted in black (DG2−4), half tone (DG3, DG4), and by boxes (DG4). For more detailed description of the 3D profile, see ref *15*.

MATERIALS AND METHODS

Construction of 3D Profile and 3D Molecular Model. The 3D profiles of artificial sequences on the sperm whale Mb backbone structure were constructed using the 3D structure coordinates, Protein Data Bank (PDB) code 1mbd (*22*) according to the procedure described previously (*9*, *15*). Molecular model building and molecular mechanics calculation were performed on SGI Octane workstation using Insight II 98 and Discover 3 (MSI). Bumps among the residues were defined as van der Waals overlaps between the residue atoms of more than 10% of the sum of the van der Waals radii of the atoms (*9*).

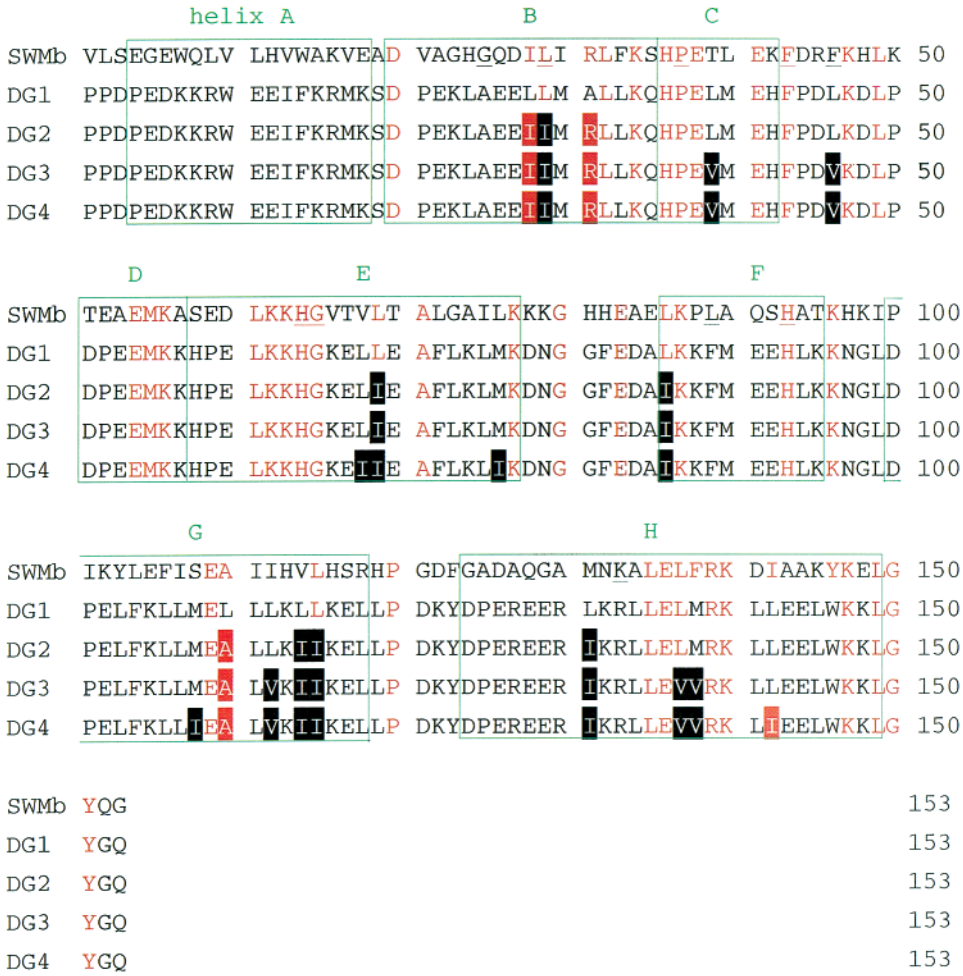

FIGURE 2: Redesigned globin (DG2−4) sequences compared with the first designed globin (DG1) and sperm whale myoglobin (SWMb) sequences. The regions corresponding to helices A−H in the globin tertiary structure are shown in the boxes. The underlined sites on the native sequence indicate the highly conserved residues in natural globins (*18*). The red and red-outlined characters indicate the sites conserved between the native sequence and the artificial sequences. The outlined characters indicate the replaced sites as compared with the DG1 sequence.

Sequence Redesign. The DG1 sequence was modified to construct redesigned globins named DG2−4 by replacements of some of the 34 Leu and eight Met residues of DG1 with β-branched amino acids, Ile and Val. Among these Leu and Met sites, Ile positioned second best at the seven sites of nos. 28, 29, 69, 86, 114, 115, and 131 in the 3D profile of DG1 (Figure 1). There was no Val at the second best positions in the 3D profile. Then the Leu residues at these sites were replaced with Ile in DG2. Besides these replacements, four Leu residues at positions 39, 46, 112, and 137 and one Met residue at a position 138, having Val at the positions within the best five, were replaced with Val in DG3 (Figure 1). In DG4, further replacements were carried out irrespective of the 3D profile: the residual Leu and Met sites sterically close to bulky residues (Phe, Tyr, and Trp) on the helices in the 3D model, i.e., Leu68 and Met76 close to Phe72, Met108 close to Phe104, and Leu142 close to Trp146, were replaced with Ile. These replacements in DG4 were expected to reduce the side-chain conformations further by the strong steric interactions (*20*, *21*). Additionally, Leu110 in DG1 was replaced with Ala in DG2−4, because the replacement of Ala or Cys at this position with Leu in natural Mbs is known to greatly reduce the folding cooperativity (*23*). Accordingly, Ala31 in DG1, which had been changed from Arg in SCS1 (i.e., Arg is ranked best at this site in DG1; see Figure 1) due to bumps with Leu110 in the 3D model, was put back to Arg in DG2−4.

Synthesis and Purification. The designed globins were synthesized as described (*9*) by expression of the synthetic genes encoding the artificial amino acid sequences with the optimal codons in *E. coli*. The synthesized proteins were purified according to a previously described method (*9*). The protein identities were verified by laser-desorption mass spectrometry (MALDI/TOFMS) and N-terminal amino acid sequencing. The DG concentrations were determined spectrophotometrically using $\epsilon_{280} = 14.4$ mM^{-1} cm^{-1} based on 5.8 mM^{-1} cm^{-1} for Trp and 1.4 mM^{-1} cm^{-1} for Tyr at pH 8.

Horse metMb was purchased from Sigma and used to compare structural properties of DGs with those of native Mb. ApoMb was prepared by 2-butanone extraction according to Hapner et al. (*24*) and purified using a Superose-6 gel filtration column (Amersham Pharmacia Biotech). Concentrations of the apoMb were determined spectrophotometrically using $\epsilon_{280} = 14.4$ mM^{-1} cm^{-1} at pH 8.0.

Small-Angle X-ray Scattering (SAXS). The SAXS measurements were carried out at 20 °C with synchrotron radiation at RIKEN structural biology beamline I (BL45XU) at SPring-8, Harima, Japan, as described previously (*9*). Samples were solubilized in 10 mM Tris-HCl (pH 8.0), 200

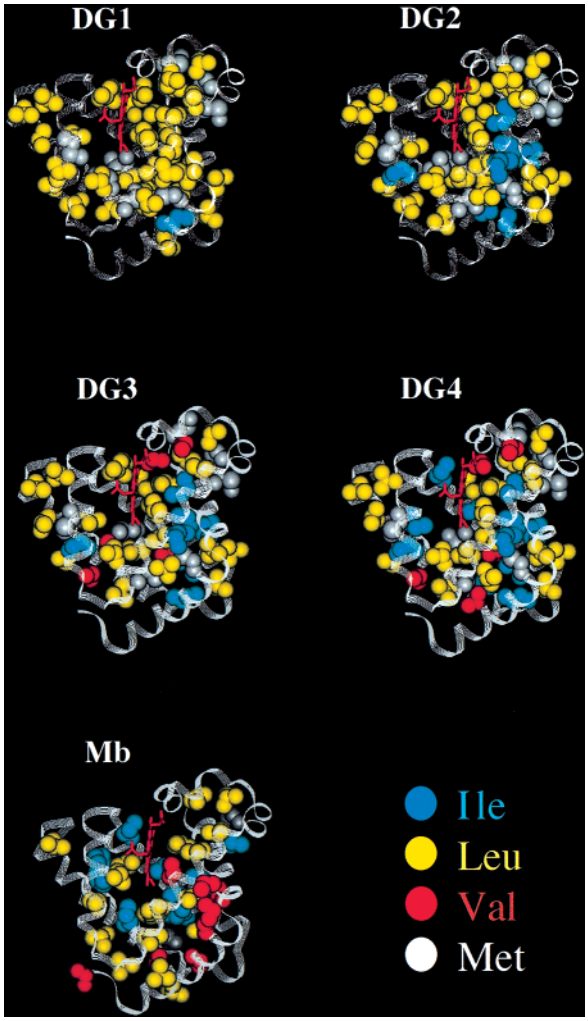

FIGURE 3: Distribution of Leu, Met, Val, and Ile residues in the 3D models of the designed globins and sperm whale Mb (1mbd). The models of the designed globins were constructed by mounting the sequences on the backbone structure of the crystal structure of SWMb in computer graphics and by the molecular mechanics calculation of side-chain conformations.

mM NaCl, and 0.2% octylglucopyranoside at various protein concentrations from 15 to 250 μM. Prior to the measurements, all samples were centrifuged at 15 000 rpm to remove precipitates. The R_g value was determined by the Guinier approximation: $I(S) = I(0) \exp(-4\pi^2 R_g^2 S^2/3)$, where S and $I(0)$ are the momentum transfer and intensity at the zero scattering angle, respectively. S is defined as $S = 2 \sin \theta/\lambda$, where 2θ and λ are the scattering angle and the X-ray wavelength, respectively. The S range used for R_g determination satisfied the condition $2\pi S R_g < 1.3$.

Spectroscopic Analysis. UV−vis spectra were recorded with a Hitachi U-3000 spectrometer using quartz cuvettes of 1.0 cm in path length. Circular dichroism (CD) spectra were recorded at 20 °C with a JASCO J700 spectropolarimeter using a rectangular quartz cuvette of 0.1 or 0.2 cm in path length with protein concentrations between 2 and 20 μM. One-dimensional proton NMR spectra were acquired at 30 °C on a Bruker ARX400 spectrometer with a time domain data size of 8192 points using a spectral width of 6024 Hz. The designed globins and apoMb were prepared at 50−500 μM in a buffer solution containing 50 mM sodium phosphate (pH 7), 200 mM NaCl, and 9% D_2O for the NMR measurements.

Denaturation. Quantitative disruption of protein secondary structures with a denaturant guanidine hydrochloride (Gd-HCl) was monitored at 20 °C by measuring CD signal intensity at 222 nm. The denaturation data were analyzed with a theoretical curve based on a three state model:

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

where F, I, and U indicate folded, intermediate, and unfolded states, respectively; K_1 and K_2 are equilibrium constants of F ⇔ I and of I ⇔ U, respectively ($K_1 = [\mathrm{F}]/[\mathrm{I}]$, $K_2 = [\mathrm{I}]/[\mathrm{U}]$). They give ΔG_1 and ΔG_2, the free energy of the folded state relative to that of the intermediate and the free energy of the intermediate relative to that of the unfolded state, respectively, which are assumed to depend linearly on the denaturant concentration:

$$\Delta G_1 = G_\mathrm{F} - G_\mathrm{I} = -RT \ln K_1 = \Delta G_1^\circ + m_1 x \quad (2)$$

$$\Delta G_2 = G_\mathrm{I} - G_\mathrm{U} = -RT \ln K_2 = \Delta G_2^\circ + m_2 x \quad (3)$$

where ΔG_1° and ΔG_2° are ΔG_1 and ΔG_2 in the absence of denaturant, respectively; 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; $A = -(\Delta G_1^\circ + m_1 x)/RT$; $B = -(\Delta G_2^\circ + m_2 x)/RT$. Then, y, the ratio of the helical contents in the transition region per 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 per that of the folded state, and the helical content of the unfolded state is zero. The theoretical curves derived from formula 6 were fitted to the denaturation data to obtain the thermodynamic parameters: ΔG_1°, ΔG_2°, m_1, and m_2.

RESULTS

Redesign of DG Sequences. The DG1 sequence was modified to restrict side-chain conformations at hydrophobic sites in the folded form, by referring to the 3D profile of DG1 and by considering steric interactions between side chains (Figures 1 and 2; see also Materials and Methods). Seven Leu residues of DG1 were replaced with Ile in DG2; besides these sites, four Leu and one Met residues were replaced with Val in DG3; further, two Leu and two Met residues were replaced with Ile in DG4. By these replacements, the total compatibility score of sequences increased to be less fit, from −168.1 kcal/mol of DG1 to −158.7 kcal/mol of DG4 (Table 1). The compositions of the hydrophobic amino acids in the redesigned globins approached that of Mb, whereas the sequence identities to sperm whale Mb were lowered as compared with DG1. The calculated molecular masses of DG2−4 are almost the same as that of DG1, yet

Table 1: Properties of Native and Designed Globin Sequences

protein	M_r[a] (Da)	identity to native (%)	composition of hydrophobic residues L	I	V	M	F	W	Y	score[b] (kcal/mol)
SWMb	17 198	100	18	9	8	2	6	2	3	−94.54
DG1	18 421	26.1	34	1	0	8	6	2	2	−168.05
DG2	18 464	25.1	26	8	0	8	6	2	2	−165.47
DG3	18 376	24.8	22	8	5	7	6	2	2	−161.00
DG4	18 340	25.5	20	12	5	5	6	2	2	−158.67

[a] Molecular mass calculated from the sequences shown in Figure 2. [b] Total compatibility score calculated from the sequences using the pseudo-energy potentials (*15*).

Table 2: Structural Parameters of the Folded Apo-Forms of Designed Globins and Horse Mb

protein	R_g (Å)[a]	helix content (%)
DG1	20.4 ± 0.1	60.8 ± 2.0
DG2	20.3 ± 0.1	60.0 ± 1.3
DG3	20.2 ± 0.1	61.2 ± 0.7
DG4	19.7 ± 0.1	61.4 ± 2.5
apoMb	18.4 ± 0.3	61.7 ± 1.1

[a] The radii of gyration (R_g) and the helix contents were determined with SAXS and CD measurements at 20 °C.

those of DG3 and DG4 decreased slightly. These properties of the artificial sequences are summarized in Table 1.

As displayed in the 3D models constructed by mounting the DG2−4 sequences on the Mb backbone structure (Figure 3), the replaced sites are not localized and are distributed dispersively in the hydrophobic region of these models. No apparent bumps were observed between nonbonded atoms in the 3D models.

Synthesis and Purification. Artificial genes for DG2−4 were constructed from synthetic oligonucleotides, cloned, and expressed in *E. coli*. The recombinant cells efficiently synthesized the gene products, which are mainly found in the soluble fraction, as in the case of DG1 (*9*). They had apparent molecular masses similar to that of DG1 in SDS−PAGE analysis. The products were purified into homogeneity with more than 95% purity judged by analytical-reversed phase HPLC. The amino-terminal sequence analyses of DGs showed the deletion of their N-terminal Pro residues during the expression and/or the purification. This was confirmed by the mass-spectroscopic analyses which showed that their experimental molecular masses were smaller than their values calculated from the sequences (see Table 1) by approximately 100 Da corresponding to M_r of one Pro residue.

Overall Molecular Dimensions. The radii of gyration (R_g) of DG1−4 monomers and apoMb were determined by the SAXS analyses as listed in Table 2. The value of DG1 was reestimated under the present experimental conditions for comparison and agrees with the data previously reported (*9*). The size-exclusion chromatography and SAXS analyses showed that DG1−4 were in equilibrium between monomer and dimer under the experimental conditions and the monomeric physical parameters were determined by extrapolating to the infinite dilute. The R_g value of DG1 was determined to be 20.4 ± 0.1 Å, which is larger by 2 Å than the R_g of apoMb (18.4 ± 0.3 Å) determined under the same conditions. The residue replacements to produce DG2 and DG3 gave no effects on their R_g within the experimental errors. DG4 has the R_g (19.7 ± 0.1 Å) slightly smaller than that of DG1 but is still larger than apoMb.

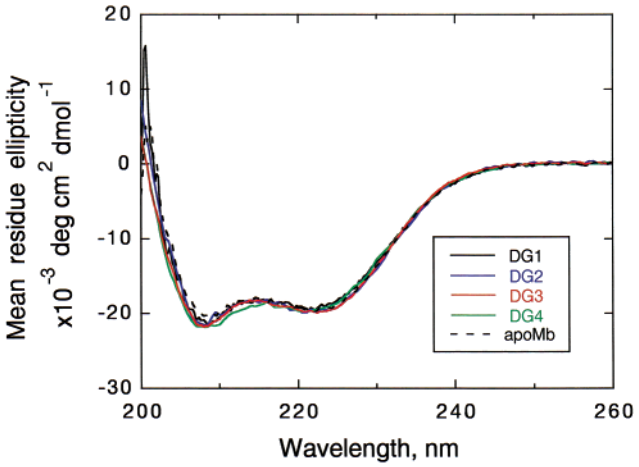

FIGURE 4: CD spectra of designed globins, DG1 (black), DG2 (blue), DG3 (red), DG4 (green), and horse apoMb (black, dashed line). The spectra were recorded at 20 °C in 10 mM Tris-HCl (pH 8.0), 200 mM NaCl, and 0.1% octyl glucopyranoside. Protein concentrations were 5 μM as determined spectrophotometrically. The data of DG1 and apoMb are taken from ref *9*.

Secondary Structure. Many statistical and experimental studies on natural proteins (*25*−*29*) and synthetic polypeptides (*30*−*32*) have shown specific tendencies of various amino acids to form a secondary structure. Ile and Val prefer β-sheets to α-helices, whereas the preference of Leu is opposite. Thus, the present residue replacements in DG2−4 could reduce the α-helix contents. The secondary structures of the DGs were analyzed by far-UV CD measurements. The CD spectra of DG1−4 are indistinguishable from each other and are also similar to apoMb, as shown in Figure 4. The helical contents estimated from the CD measurements were summarized with their R_g values in Table 2. The helical contents were independent of protein concentration in the examined range of 2−20 μM. These results show that the replacements of Leu residues with Ile and Val residues at many sites did not affect the α-helix contents in the absence of denaturant and that DGs have significant robustness of secondary structure against the intense mutagenesis. However, these replacements significantly affected the denaturation process (see below).

Effects on Stability. The denaturation of the natural and designed globins with a denaturant guanidine hydrochloride (Gd-HCl) was measured by monitoring the CD signal intensity at 222 nm to examine the effects of the residue replacements on the thermodynamic stability and the folding cooperativity (Figure 5). DG1 showed a broad denaturation curve with a higher transition midpoint (C_m = 4.9 M) for a 60% fraction of the total helix content, with the other 40% fraction preserved even under the maximum Gd-HCl concentration close to 8.0 M (*9*). In the cases of DG2−4, the C_m values significantly decreased and complete denaturation of the whole fractions was observed in the range of Gd-HCl concentration within 8.0 M. Since the denaturation curves of every redesigned globin were apparently resolved into at least two components, we analyzed these data using a three state model (see Materials and Methods) to obtain the thermodynamic parameters: ΔG_1° and ΔG_2°, the free-energy changes from the folded state (F) to the intermediate (I) and from the intermediate (I) to the unfolded state (U) in the absence of denaturant; and m_1 and m_2, the dependence of the free energy changes on the Gd-HCl concentration,

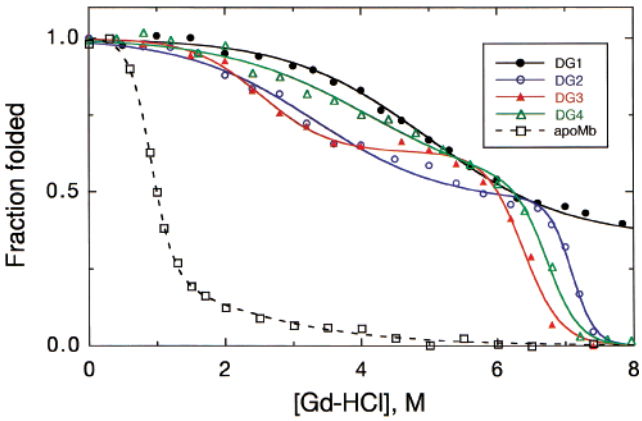

FIGURE 5: Denaturation curves of designed globins and horse apoMb by monitoring their CD signal intensity at 222 nm against Gd-HCl concentration at 20 °C. The data are displayed in different colors: DG1 in black, DG2 in blue, DG3 in red, DG4 in green, and apoMb in black (dashed line). The proteins were denatured by titration with Gd-HCl in 10 mM Tris-HCl (pH 8.0), 200 mM NaCl, and 0.1% octyl glucopyranoside at protein concentration of 5 μM. The data of DG1 are taken from ref *9*.

respectively. The parameter, *m*, measures the cooperativity of the two-state transition (*33*). The analyses revealed much more stable intermediates of the redesigned globins (DG2−4) than that of apoMb. The intermediates of DG1−4 accumulate to form major components during the denaturation, whereas the apoMb intermediate forms transiently (Figure 6). Of all the designed globins (DG1−4), DG3 has the most cooperative F ⇔ I transition, the largest helical content of the intermediate and, possibly, the lowest total stability ($\Delta G^\circ_{total} = \Delta G^\circ_1 + \Delta G^\circ_2$). The data are summarized in Table 3.

NMR Spectroscopy. The conformational specificity of the designed globins, DG1−4, were probed by measuring their one-dimensional proton NMR spectra. Figure 7 shows the spectra in the region of amide and aromatic proton chemical shifts, which are compared with that of natural apoMb. These spectra of DGs show characteristic features of α-helical proteins, i.e., the range of the amide-proton chemical shifts between 6 and 9 ppm and the absence of Cα-H resonances in 5−6 ppm, like the spectrum of apoMb. DG1 exhibits rather broad lines and poor chemical shift dispersion. Thus, it adopts many different thermally accessible conformations that slowly interconvert on the proton chemical shift time scale, while it preserves the overall shape and the α-helical content similar to those of apoMb, as judged by CD and SAXS analyses. DG2 shows the spectrum similar to that of DG1 but slightly narrower proton resonance. On the other hand, DG3 displays clearly more dispersed and resolved signals. The spectrum of DG4, however, shows quite poor signal dispersion and such broad lines that the resonances of tryptophan indol protons in 10−10.5 ppm are unresolved. Similar results of signal qualities were obtained in other ranges of chemical shifts (not shown). These NMR data agree with the results of the denaturation experiments mentioned above (see also Discussion), and they indicate that DG3 has the most specific structure of DG1−4.

The dimerization and further oligomerization of DGs at high protein concentration could broaden the NMR signals. However, both the overall patterns and the line widths of the NMR spectra did not change in the concentration range from 50 to 250 μM, in which DGs were at equilibrium

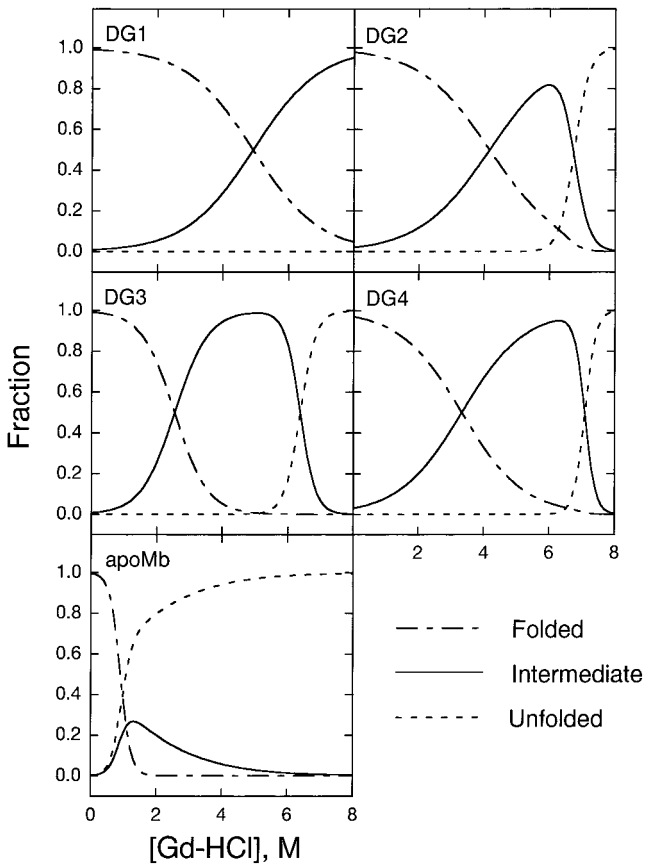

FIGURE 6: Fractions of the folded, intermediate and unfolded states of designed globins and horse apoMb as a function of Gd-HCl concentration. The theoretical curves are drawn based on formulas 4 and 5 in Materials and Methods using the parameters listed in Table 3. In a CD stopped-flow measurement, the helical content of the kinetic intermediate of apoMb was estimated to be 41%, which was assumed to be similar to that of the static intermediate in the denaturation experiment and was used for the present analyses.

between dimer and monomer based on size-exclusion chromatography and SAXS analyses. Thus, the dimerization gives no significant effects on the monomer structure nor on the NMR signal dispersion. The differences of the NMR spectra between the proteins in Figure 7 are not due to those of the oligomerization states but due to those of the structural specificity as described above.

DISCUSSION

The present residue replacements of DG1 to give DG2 and DG3 resulted in no significant changes in their R_g, whereas DG4 has the dimension slightly but clearly reduced from that of DG1 (Table 2), indicating that the side-chain packing in DG4 was improved to compact the molecule by the replacements. However, no improvement was observed in the NMR spectrum of DG4, whereas DG3 with the R_g close to that of DG1 exhibited the best spectrum among DG1−4. These apparently controversial results on the SAXS and NMR experiments may originate from the coexistence of several conformational states in the most compact core of DG4 as well as in the cores of other DGs. Then, the structural specificity is not necessarily measured by the overall molecular dimensions with such slight differences. Even the compactness of DG4 and the NMR signal quality of DG3 were less or worse than those of native apoMb,

Table 3: Thermodynamic Parameters of the Designed Globins and Horse ApoMb for Gd-HCl Denaturation[a]

protein	ΔG_1° (kcal mol^{-1})	m_1 (kcal mol^{-1} M^{-1})	ΔG_2° (kcal mol^{-1})	m_2 (kcal mol^{-1} M^{-1})	ΔG_{total}° (kcal mol^{-1})
DG1	2.8 ± 0.1	0.57 ± 0.01			
DG2	2.0 ± 0.2	0.61 ± 0.06	25.1 ± 3.7	3.5 ± 0.5	27.1 ± 3.7
DG3	2.9 ± 0.5	1.1 ± 0.2	15.5 ± 1.6	2.4 ± 0.3	18.4 ± 1.7
DG4	2.3 ± 0.2	0.54 ± 0.08	16.7 ± 2.9	2.5 ± 0.4	19.0 ± 2.9
apoMb	3.7 ± 0.2	3.4 ± 0.2	0.01 ± 0.08	0.41 ± 0.11	3.7 ± 0.2

[a] The data were obtained using the helical contents of the folded forms listed in Table 2 and those of intermediates: 28% (DG2), 38% (DG3), 30% (DG4), 40.7% (apoMb), based on analyses of the denaturation data as described in Materials and Methods (see also Figure 6). The helix content of the apoMb intermediate was directly determined by CD-stopped flow measurement in a separate experiment. $\Delta G_{total}^\circ = \Delta G_1^\circ + \Delta G_2^\circ$.

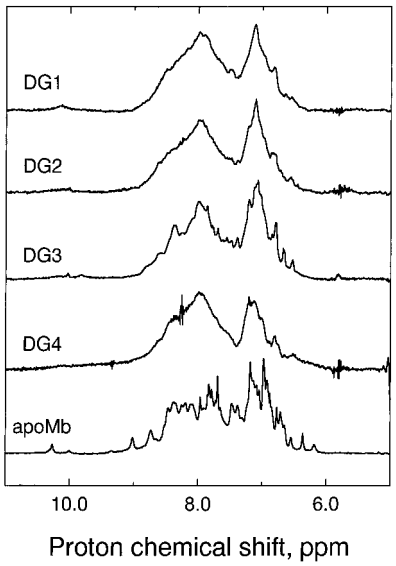

FIGURE 7: One-dimensional NMR spectra of the designed globins and horse apoMb. The protein concentrations used were 250 μM (DG1), 210 μM (DG2), 219 μM (DG3), and 226 μM (DG4), and 500 μM (apoMb) in a buffer solution containing 50 mM sodium phosphate (pH 7), 200 M NaCl, and 9% D_2O.

indicating that the designed globins did not yet gain the structural uniqueness of native proteins.

Side-chain conformations of Ile and Val residues on α helices can be more restricted by steric interactions of their β-branched side chains with neighboring residues than those of Leu and Met residues (*19*−*21*). On the other hand, Ile and Val have lower helical propensities in native protein structures (*25*−*27*) and in small model peptides (*30*−*32*) than Leu and Met. The differences in the helical propensities of these amino acids can be explained partly by larger loss of side-chain conformational entropy during the protein folding for Ile and Val than that for Leu and Met, since the entropy loss ($-\Delta S_{conf}$) itself is apparently disadvantageous in the protein stability according to $\Delta G = \Delta H - T\Delta S$. Native proteins keep delicate balance between ΔH and $T\Delta S$ to form their unique structures with the necessary stabilities, ΔG, under physiological conditions (*34*). The intense sequence modifications of DG1, in which 13 of 34 Leu residues and three of eight Met residues were replaced with Ile and Val to produce DG4, did not affect the α-helix content in the absence of denaturant. This mutational robustness of DG1 seems to be due to rather loose side-chain packing of the protein core, which enables these residues to be replaced without causing change in the secondary structure under the experimental conditions. However, these sequence modifications significantly reduced the stability.

The present residue replacements changed the dependence of the secondary structure on the concentration of Gd-HCl in the denaturation experiments. The denaturation curves were analyzed using a three state model, i.e., F ⇔ I ⇔ U (see Materials and Methods). DG1 is too resistant against the denaturant; thus, unfortunately the I ⇔ U transition of DG1 could not be observed. On the other hand, the F ⇔ I transitions of DG1−4 and the I ⇔ U transitions of DG2−4 were measured and their values of ΔG° and m were determined (Table 3). The folding intermediates of DG2−4, and possibly that of DG1, are highly stable with much larger values of ΔG_2° than their ΔG_1°, and they show cooperative denaturation with the m_2 values larger than their m_1. The stabilities of these intermediates of the designed globins are even much larger than the overall stabilities of apoMb and many other natural proteins. On the other hand, the folding intermediate of apoMb is much less stable and exhibits less cooperative unfolding. The apoMb intermediate is known to preserve the secondary structure, with 20% or higher helical contents, corresponding to the A, G, and H helices of the native state, which is disordered as in a molten globule state (*35*−*38*). In contrast with the apoMb intermediate, the intermediates of DGs should preserve the structural properties similar to those of the state of the corresponding part of the folded proteins. These results indicate contrasting structural properties of the natural and designed globins: each helix of the designed globins is very stable in itself and the interactions between the helices are relatively weak, whereas each helix of apoMb is quite unstable in itself and the helices interact closely with each other to form a moderately stable global structure, which folds and unfolds cooperatively.

The overall stabilities ($\Delta G_2^\circ + \Delta G_1^\circ$) of the natural and designed globins are in the order of apoMb (3.7 kcal/mol) < DG3 (18.4 kcal/mol) ≈ DG4 (19.0 kcal/mol) < DG2 (27.1 kcal/mol) (Table 2), and DG1 may be the most stable since it exhibits the highest values of C_m for Gd-HCl (Figure 5). On the other hand, $m_1 + m_2$, or even the value weighted with the helical contents disrupted in these transitions, namely, $(H_F - H_I)m_1 + H_I m_2$, where H_F and H_I were helical contents of the folded and intermediate forms, respectively, does not measure the overall cooperativity or the global structural specificity because the values of these parameters are lowest for apoMb and are not consistent with the NMR data (Figure 7); rather than these, simply m_1 seems to correlate with the global structural specificity in the present cases (see Table 3). In conclusion, of DG1−4, DG3 has the most specific structure with the lowest global stability, and replacements of Leu with Ile or Val are effective for increasing the structural specificity. The total numbers of the β-branched amino acids of DG1, DG2, DG3, and DG4 are 1, 8, 13, and 17, respectively. The number of Ile and

Val residues of DG4 is the same as that of sperm whale Mb (Table 1), whereas the number of the β-branched residues of DG3 is the closest to 13.6 amino acids, the average number of 75 natural Mbs with more than 100 amino acid residues in the PIR protein sequence database (Release 61, Section 1). Thus, the structural specificity depends on the composition of these hydrophobic amino acids and also on their positions. Our results agree with those of the iterative design of artificial four helix bundles (*13*) and also of mutagenesis studies of native helical proteins (*21*).

The present results suggest that the uniqueness of protein structure can be attained at the expense of stability. Thus, the uniqueness or decrease in conformational freedom during the folding (ΔS_{conf}) as well as the stability (ΔG) should be considered as the index for the design of artificial protein sequences to fold into a nativelike unique structure.

DG1 binds a single heme per protein molecule with a concomitant decrease of the R_g by approximately 1 Å, as previously described (*9*). The heme-bound DG1 showed well-defined UV−vis absorption and resonance Raman spectra indicating low-spin six-coordinated heme iron in the ferric form and the mixture of six- and five-coordinated heme irons in the ferrous form. Also DG2−4 bind a single heme with decreases in their R_g and show spectroscopic features similar to those of DG1 (M. Ishida, M. Mukai, A.I., T.F., and Y.I., unpublished results). However, neither increases in the helix contents of DG1−4 nor significant improvement of their NMR spectra were observed with the association with heme. The affinity of DG2−4 to heme was lower than that of DG1, and the desired five-coordination state, which was detected in the ferrous form of heme-DG1, was destabilized in those of heme-bound DG2−4. These results suggest that the compacted heme-bound forms of DG1−4 still have structural diversity and that the present sequence modifications induced unexpected changes in structure of the heme-binding site to fix conformation unfavorable for the heme binding. Thus, redesign to improve the structural specificity only by considering the local structure, as conducted in the present work, is ineffectual to improve the function. De novo rational design of proteins with nativelike functions will primarily require to produce unique structures with sufficient robustness against the sequence modifications to introduce specific functions, which frequently spoil the stabilities (*39*).

ACKNOWLEDGMENT

We thank Ms. Yasue Ichikawa (Biodesign DNA Sequencing Facility, RIKEN) for DNA sequencing, and Drs. Naoshi Dohmae and Koji Takio (Division of Biomolecular Characterization, RIKEN) for mass spectroscopic and N-terminal sequence analyses of proteins. We also thank Prof. Haruki Nakamura (Osaka University) for his critical discussions.

REFERENCES

1. Desjarlais, J. R., and Handel, T. M. (1995) *Curr. Opin. Biotech. 6*, 460−466.
2. Bryson, J. W., Betz, S. F., Lu, H. S., Suich, D. J., Zhou, H. X., O'Neil, K. T., and DeGrado, W. F. (1995) *Science 270*, 935−941.
3. Cordes, M. H. J., Davidson, A. R., and Sauer, R. T. (1996) *Curr. Opin. Struct. Biol. 6*, 3−10.
4. Hecht, M. H., Richardson, J. S., Richardson, D. C., and Ogden, R. C. (1990) *Science 249*, 884−891.
5. Handel, T. M., Williams, S. A., and DeGrado, W. F. (1993) *Science 261*, 879−885.
6. Tanaka, T., Kimura, H., Hayashi, M., Fujiyoshi, Y., Fukuhara, K., and Nakamura, H. (1994) *Protein Sci. 3*, 419−427.
7. Choma, C. T., Lear, J. D., Nelson, M. J., Dutton, P. L., Robertson, D. E., and DeGrado, W. F. (1994) *J. Am. Chem. Soc. 116*, 856−865.
8. Gibney, B. R., Johansson, J. S., Rabanal, F., Skalicky, J. J., Wand, A. J., and Dutton, P. L. (1997) *Biochemistry 36*, 2798−2806.
9. Isogai, Y., Ota, M., Fujisawa, T., Izuno, H., Mukai, M., Nakamura, H., Iizuka, T., and Nishikawa, K. (1999) *Biochemistry 38*, 7431−7443
10. Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D., and Chan, H. S. (1995) *Protein Sci. 4*, 561−602.
11. Betz, S. F., Liebman, P. A., and DeGrado, W. F. (1997) *Biochemistry 36*, 2450−2458.
12. Schafmeister, C. E., LaPorte, S. L., Miercke, L. J. W., and Stroud. R. M. (1997) *Nat. Struct. Biol. 4*, 1039−1046.
13. Gibney, B. R., Rabanal, F., Skalicky, J. J., Wand, A. J., and Dutton, P. L. (1999) *J. Am. Chem. Soc. 121*, 4952−4960.
14. Dahiyat, B. I., and Mayo, S. L. (1997) *Science 278*, 82−87.
15. Ota, M., and Nishikawa, K. (1997) *Protein Eng. 10*, 339−351.
16. Ota, M., Kanaya, S., and Nishikawa, K. (1995) *J. Mol. Biol. 248*, 733−738.
17. Chothia, C., Levitt, M., and Richardson, D. (1977) *Proc. Natl. Acad. Sci. 74*, 4130−4134.
18. Lesk, A. M., and Chothia, C. (1980) *J. Mol. Biol. 136*, 225−270.
19. Creamer, T. P., and Rose, G. D. (1992) *Proc. Natl. Acad. Sci. 89*, 5937−5941.
20. Nakamura, H., Tanimura, R., and Kidera, A. (1996) *Proc. Jpn. Acad. 72B*, 149−152.
21. Furukawa, K., Oda, M., and Nakamura, H. (1996) *Proc. Natl. Acad. Sci. 93*, 13583−13588.
22. Phillips, S. E. (1980) *J. Mol. Biol. 142*, 531−554.
23. Hughson, F. M., and Baldwin, R. L. (1989) *Biochemistry 28*, 4415−4422.
24. Hapner, K. D., Bradshaw, R. A., Hartzell, C. R., and Gurd, F. R. N. (1968) *J. Biol. Chem. 243*, 683−689.
25. Chou, P. Y., and Fasman, G. D. (1978) *Annu. Rev. Biochem. 47*, 251−276.
26. Horovitz, A., Matthews, J. M., and Fersht, A. R. (1992) *J. Mol. Biol. 227*, 560−568.
27. Blaber, M., Zhang, X. J., Lindstrom, J. D., Pepiot, S. D., Baase, W. A., and Matthews, B. W. (1994) *J. Mol. Biol. 235*, 600−624.
28. Minor, D. L., Jr., and Kim, P. S. (1994) *Nature 367*, 660−663.
29. Smith, C. K., Withka, J. M., and Regan, L. (1994) *Biochemistry 33*, 5510−5517.
30. O'Neil, K. T., and DeGrado, W. F. (1990) *Science 250*, 646−651.
31. Chakrabartty, A., Kortemme, T., and Baldwin, R. L. (1994) *Protein Sci. 3*, 843−852.
32. Kim, C. A., and Berg, J. M. (1993) *Nature 362*, 267−270.
33. Shortle, D., and Meeker, A. K. (1986) *Proteins: Struct., Funct., Genet. 1*, 81−89.
34. Pfeil, W., and Privalov, P. L. (1976) *Biophys. Chem. 4*, 41−50.
35. Hughson, F. M., Barrick, D., and Baldwin, R. L. (1991) *Biochemistry 30*, 4113−4118.
36. Jennings, P. A., and Wright, P. E. (1993) *Science 262*, 892−895.
37. Kataoka, M., Nishii, I., Fujisawa, T., Ueki, T., Tokunaga, F., and Goto, Y. (1995) *J. Mol. Biol. 249*, 215−228.
38. Eliezer, D., Jao, J., Dyson, H. J., and Wright, P. E. (1998) *Nat. Struct. Biol. 5*, 148−155.
39. Ota, M., Isogai, Y., and Nishikawa, K. (1997) *FEBS Lett. 415*, 129−133.

BI992687+