

## ‘Module’-substituted globins: Artificial exon shuffling among myoglobin, hemoglobin $\alpha$ - and $\beta$ -subunits

Keisuke Wakasugi<sup>1</sup>, Koichiro Ishimori, Isao Morishima<sup>\*</sup>

*Department of Molecular Engineering, Graduate School of Engineering, Kyoto University, Kyoto 606-01, Japan*

---

### Abstract

Based on the detailed structural analysis of proteins, Go [M. Go, *Nature* 291 (1981) 90–92] found that protein structures can be divided into some structural units, ‘modules,’ which correspond to peptides coded by exons. In the present study, to investigate functional and structural roles of modular structures in proteins, we have engineered eight chimera globins, in which the exons are shuffled among human myoglobin, human hemoglobin  $\alpha$ - and  $\beta$ -subunits, in addition to the chimera  $\beta\beta\alpha$ -globin described previously [K. Wakasugi, K. Ishimori, K. Imai, Y. Wada, I. Morishima, *J. Biol. Chem.* 269 (1994) 18750–18756]. Although all of the chimera globins stoichiometrically bound the heme and their  $\alpha$ -helical contents increased by heme incorporation as found for native globins, the  $\alpha$ -helical contents of the chimera globins were significantly lower than those of native globins, suggesting that ‘module’ substitutions seriously affect the protein folding and stability in globins. The comparisons among several chimera globins demonstrated that such structural alterations are mainly attributed to loss of some key intermodular interactions for protein folding. By simultaneous substitution of the modules M1 and M4 from the same globin, the protein structure was stabilized, which indicates that the module packing between modules M1 and M4 would be one of the crucial interaction to stabilize the globin fold. Present results allow us to conclude that module substitutions would be available for designing and producing novel functional proteins if we can reproduce the stable modular packing in the ‘module’-substituted proteins. © 1997 Elsevier Science B.V.

**Keywords:** Heme proteins; Exon shuffling; Molecular evolution; Protein engineering

---

### 1. Introduction

Unlike the genes of prokaryotes whose coding sequences are continuous, the coding sequences of eukaryotic genes were found to be present in blocks, exons, separated by intervening noncoding sequences, introns. Gilbert [1] and Blake [2] hypothesized that exons encode functional and structural units and that new functional proteins have evolved

by selection of various combinations of the units that are produced by unequal crossing-over on introns, exon shuffling. Using a diagonal plot of all the distances between the  $\alpha$ -carbon atoms, Go [3,4] demonstrated that there is a correlation between protein structure and the exon pattern and found that the ‘modules,’ which could be called compact structural units, correspond to the exons. The correspondence of the ‘modular’ boundaries with the position of introns in globins [3], lysozyme [4], cytochrome *c* [5], and other proteins [5–7] suggests that exons may have behaved as evolutionary units to produce new proteins by combining various exons through the mechanism of exon shuffling.

---

<sup>\*</sup> Corresponding author. Tel.: +81-75-753-5921; fax: +81-75-751-7611; e-mail: morisima@mds.moleng.kyoto-u.ac.jp

<sup>1</sup> Present address: The Scripps Institute, La Jolla, CA 92037, USA.

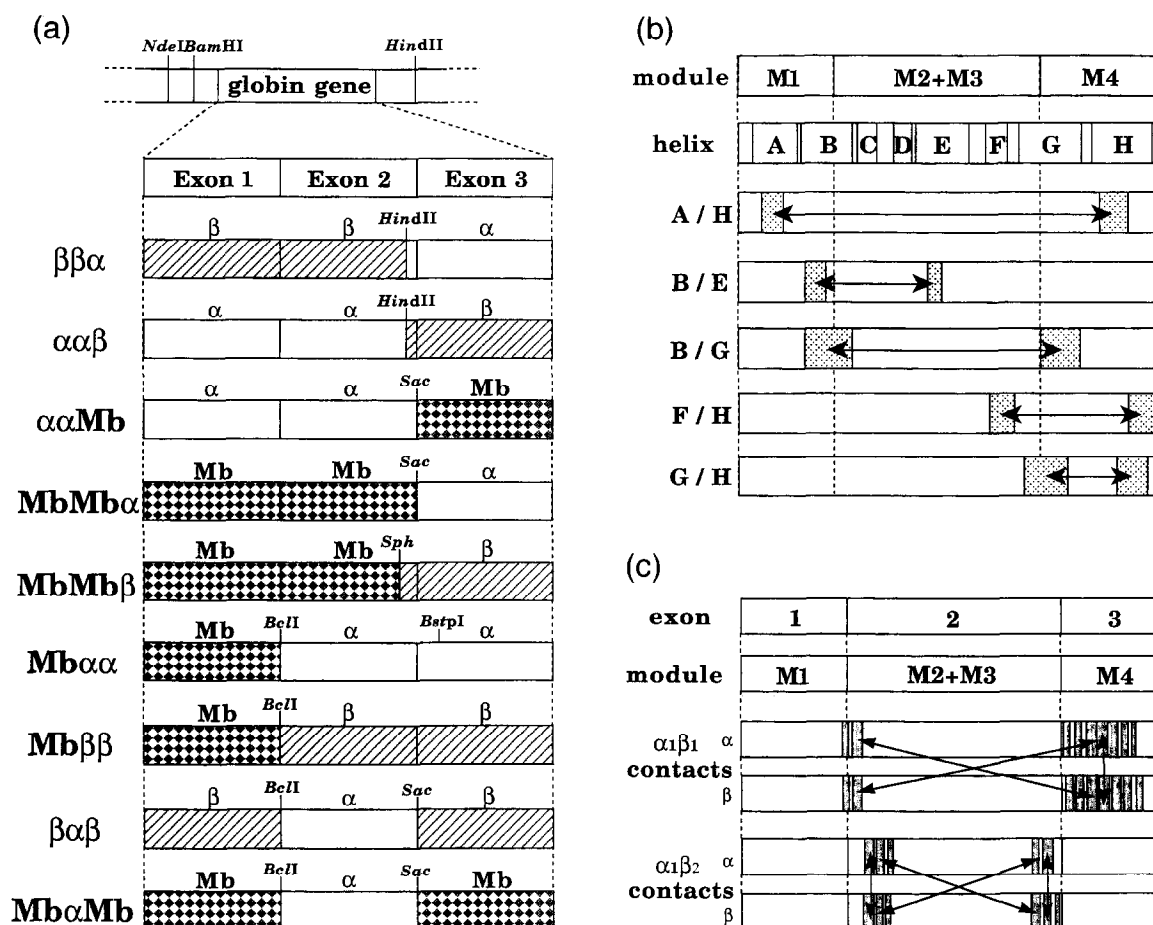


Fig. 1. (a) Construction of expression vectors for the nine chimera globins. (b) Intrasubunit interactions in native globins. The region of the conserved residue-residue interaction for protein packing is denoted by a box. (c) Exon (modular) boundaries and residues with well defined functional roles in native human hemoglobin proposed by Eaton [14]. Main interactions between their residues are denoted by arrows.

Although enormous amount of gene analyses to investigate the traces of exon shuffling have been accumulated [8–10], very few experimental studies to investigate the characteristics of the proteins produced by exon shuffling have been attempted [11,12]. Therefore, we have tried artificial exon shuffling among human myoglobin, human hemoglobin  $\alpha$ - and  $\beta$ -subunits by utilizing gene engineering and investigated the structures and functions of the corresponding ‘module’-substituted globins. The gene of each globin is made up of three exons interrupted by two introns and the exons 1, 2 and 3 correspond to the modules M1, M2 + M3 and M4, respectively [3]. In our previous paper, we have replaced the exon 3

in human hemoglobin  $\beta$ -subunit with that in the  $\alpha$ -subunit by the cassette mutagenesis and showed that we can prepare a novel chimera globin subunit, in which the heme’s environmental structure and ligand-binding properties are similar to those of the  $\beta$ -subunit, whereas the association properties correspond to those of the  $\alpha$ -subunit, by the ‘module’ substitution [13].

In the present study, to gain further insight into the effects of the ‘module’ substitutions on structures and functions of globins, we have engineered eight other chimera globins in addition to the chimera  $\beta\beta\alpha$ -globin, as illustrated in Fig. 1a. By use of absorption, fluorescence and circular dichroism (CD)

spectra and gel filtration, we focused on the effects of the 'module' substitution on the protein structure and subunit association properties. Our results will shed light on the aspects of the relationship between the modules and the structure and function of the globin protein.

## 2. Materials and methods

### 2.1. Preparation of 'module'-substituted globins

The expression vectors of the nine chimera globins were constructed by using the natural or artificial restriction sites, which are shown in Fig. 1a. The chimera globins possess the following additional mutations due to introduction of the restriction sites: chimera  $\beta\beta\alpha$ -globin, three mutations [on His (FG4) to Arg, Glu (G3) to Val, Arg (G6) to Lys] [13]; chimera  $\alpha\alpha\beta$ -globin, three mutations [on Arg (FG4) to His, Val (G3) to Glu, Lys (G6) to Arg] [15]; both chimera  $\alpha\alpha$ myoglobin (Mb)- and Mb $\alpha$ Mb-globins, two mutations [on Lys (G6) to Glu, Phe (G7) to Leu]; chimera MbMb $\beta$ -globin, 13 mutations [on Thr (FG1) to Asp, His (FG3) to Leu, Lys (FG4) to His, Ile (FG5) to Val, Pro (G1) to Asp, Val (G2) to Pro, Lys (G3) to Glu, Tyr (G4) to Asn, Leu (G5) to Phe, Glu (G6) to Arg]; chimera  $\beta\alpha\beta$ -globin, two mutations [on Gly (B11) to Ile, Lys (G6) to Glu]; chimera MbMb $\alpha$ -, Mb $\alpha\alpha$ - and Mb $\beta\beta$ -globins, no mutations. All chimera globins were expressed in *Escherichia coli* and purified as previously reported [13]. We checked whether or not the desired mutations were correctly introduced and no other mutations unexpectedly occurred in the engineered proteins by means of fast-atom-bombardment mass spectrometry on tryptic digests [16].

### 2.2. Measurements

Electronic absorption spectra were recorded on a Hitachi U-3210 UV-VIS spectrophotometer. The experiments were performed in 50 mM tris-HCl, pH 7.4 at 20°C. Each sample concentration was  $\sim 10 \mu\text{M heme}^{-1}$ .

Fluorescence emission spectra of 1-anilino-8-naphthalene sulfonate (ANS) bound to globins were measured at 20°C on a Shimadzu RF-503A spectro-

photometer. The samples were excited at 365 nm, and emission spectra were measured between 400 and 700 nm. Each sample concentration was  $\sim 10 \mu\text{M heme}^{-1}$ . The buffer used for the fluorescence measurements was 50 mM tris-HCl, pH 7.4.

Circular dichroism spectra were recorded on a Jasco J700 spectropolarimeter connected to a personal computer PC98 at 20°C. The buffer used was 10 mM tris-HCl, pH 7.4. A cuvette of 0.1 cm light path was used. Each sample concentration was 5 to 10  $\mu\text{M heme}^{-1}$ . Each molar ellipticity ( $\text{deg cm}^2 \text{dmol}^{-1}$ ) is expressed on a mean residue basis. The  $\alpha$ -helical contents ( $f_H$ ) were calculated by the following equation according to Chen et al. [17].

$$f_H = -([\theta]_{222 \text{ nm}} + 2340)/30,300.$$

The carbon monoxide rebinding kinetic measurements were made by using the millisecond flash photolysis as previously described [18]. The monitoring wavelength was 436 nm. Each sample concentration was  $\sim 20 \mu\text{M heme}^{-1}$ . The experiments were performed in 50 mM tris-HCl, pH 7.4 at 20.0°C.

Association properties of globins were investigated by gel filtration on a Sephacryl S-200 HR column ( $1.0 \times 64.5 \text{ cm}$ ) over the concentration range 1 to 30  $\mu\text{M heme}^{-1}$  [19]. The buffer used for the gel filtration was 50 mM tris-HCl, 0.1 M NaCl and 1 mM  $\text{Na}_2\text{EDTA}$ , pH 7.4 at 10°C and saturated with CO. The eluent was monitored by measuring the absorbance at 420 nm.

## 3. Results and discussion

### 3.1. Structure and function of 'module'-substituted globins

Titration of the chimera globins with ferric hemin showed that the heme bound to the chimera globins in a stoichiometric ratio of 1:1 heme:protein. In order to investigate the hydrophobicity of the heme environment, we utilized the fluorescence of ANS, which can be incorporated into the heme pocket [20], as a sensitive probe. As summarized in Table 1, the emission maxima of ANS bound to the chimera globins (456–461 nm) were significantly shifted from that of the buffer (515 nm), and corresponded to those of native globins (454–457 nm). Table 1 also

Table 1  
Comparison of structure and function among native and chimera globins

	$\alpha$	$\beta$	Mb	$\beta\beta\alpha$	$\alpha\alpha\beta$	$\alpha\alpha Mb$	MbMb $\alpha$	MbMb $\beta$	Mb $\alpha\alpha$	Mb $\beta\beta$	$\beta\alpha\beta$	Mb $\alpha Mb$
<i>Fluorescence emission spectra of globin-bound ANS<sup>a</sup></i>												
$\lambda_{max}$ (nm)	456	457	454	459	461	461	459	459	459	459	459	456
<i>Electronic absorption spectra of carbonmonoxy (CO) derivatives</i>												
Soret (nm)	418	420	423	420	420	420	422	422	422	422	422	421
$\beta$ (nm)	538	539	540	538	540	539	539	541	539	540	540	539
$\alpha$ (nm)	569	569	578	568	568	568	566	570	569	569	569	569
<i>Mean residue ellipticity at 222 nm (<math>\times 10^{-4}</math>) deg cm<sup>2</sup> dmol<sup>-1</sup>)</i>												
heme-free	0.96	0.96	1.86	0.99	0.62	0.70	0.51	0.66	0.70	0.69	0.94	0.96
+ heme (CO)	2.29	2.18	2.25	1.84	0.89	1.38	0.99	1.03	0.89	1.04	1.27	1.14
<i>Kinetic constants for CO rebinding<sup>b</sup> (<math>\mu M^{-1} s^{-1}</math>)</i>												
4.3 (100%)	13.7 (70%)	0.9 (100%)	14.8 (70%)	10.8 (45%)	6.6 (40%)	6.6 (40%)	1.4 (43%)	10.6 (59%)	11.8 (63%)	3.1 (40%)	4.7 (68%)	7.1 (49%)
5.3 (30%)			4.9 (30%)	1.9 (40%)	1.3 (24%)	1.3 (24%)	7.0 (41%)	2.0 (41%)	1.1 (37%)	15.0 (32%)	17.9 (20%)	1.1 (35%)
				0.2 (15%)	0.02 (21%)	0.02 (21%)	0.3 (21%)			0.5 (29%)	0.8 (13%)	0.2 (16%)
						0.2 (16%)						

<sup>a</sup>The molar ratio between ANS and apochimera-subunit was kept below 0.2 to avoid fluorescence contributions arising from fluorophore molecules bound to lower affinity protein sites. The  $\lambda_{max}$  value of free ANS in the buffer was 515 nm.

<sup>b</sup>The values in parentheses denote percentage of the components.

shows that the electronic absorption spectra of the carbon monoxide derivatives of all chimera globins exhibited the same absorption maxima as those of the native globins.

However, in the absence of the exogenous ligand, the absorption spectra of the chimera globins are quite different from that of native globin. The bis-imidazole type spectra (data not shown) were obtained for the chimeric proteins, which implies that the distal histidine is easy to ligate to the heme iron to form the internal hemi- or hemochromogen. The formation of the hemi- or hemochromogen type heme is also encountered for some protease-digested globins [21] and aged native globins [22], in which the protein structures are partially unfolded or denatured, suggesting that folding of the chimeric proteins are significantly perturbed by the 'module' substitution. Oxygen adduct of the chimera globins could not be observed due to their rapid autooxidation.

To investigate the secondary structures of the chimera globins, we measured their CD spectra and the  $\alpha$ -helical contents were estimated on the basis of  $[\theta]_{222\text{ nm}}$ , which are compiled in Table 1 [17]. As illustrated in Fig. 2, the  $\alpha$ -helical contents of the chimera globins are increased by the heme binding. The increases in the  $\alpha$ -helical contents with heme binding are also observed for native globins, suggest-

ing that the heme incorporation into the chimera globins induced structural rearrangements in the globin structure as found for native globins. However, it should also be noted that the  $\alpha$ -helical contents of the chimera globins were drastically reduced as compared with those of native globins in the absence or presence of heme. The reduced  $\alpha$ -helical contents imply that the 'module' substitutions seriously affect the secondary structures of globins, which lead to remarkable decreases in protein stability.

These structural differences are also manifested in the NMR spectra. In native globins, some characteristic NMR signals have been assigned: up-field shifted peaks of the amino acid residues in the heme vicinity, and dispersed peaks from aromatic and histidine  $^1\text{H}^\epsilon$  protons [23]. All these characteristics were present in the chimera  $\beta\beta\alpha$ -globin [13] but not observed for other chimera globins as well as molten-globule-state myoglobin [23]. These NMR resonances indicate that the chimera  $\beta\beta\alpha$ -globin forms almost the same tertiary structures as native globins [13], whereas other chimera globins form unstable tertiary structures and always fluctuate as found for denatured globins.

To examine functions of the chimera globins, the CO rebinding kinetic measurements have been done by using flash photolysis method and the rate con-

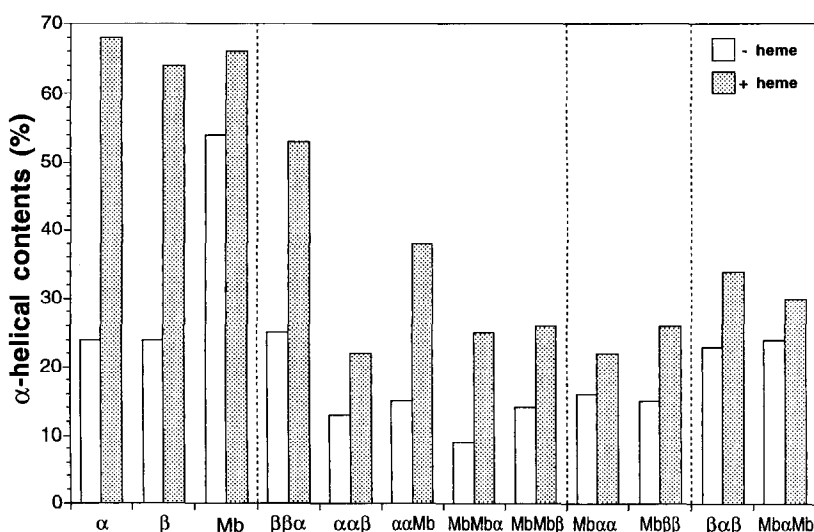


Fig. 2. The  $\alpha$ -helical contents of the nine chimera globins in the absence and presence of heme. The  $\alpha$ -helical contents of all the chimera globins were not altered even by the addition of counterpart globins.

stants were calculated as listed in Table 1. Photolysis of the carbonmonoxy chimera  $\beta\beta\alpha$ -globin gave biphasic rebinding kinetics as found for the  $\beta$ -globin and the rate constants were almost identical to those of the  $\beta$ -globin [13]. In our previous paper, we concluded that the ligand-rebinding kinetics is regulated by the modules M2 + M3, in which the amino acid residues associated with the heme contacts are localized. For the chimera  $\beta\alpha\beta$ -globin, the rate constant of the main fraction (68%) in the CO rebinding ( $4.7 \mu\text{M}^{-1} \text{s}^{-1}$ ) is almost identical to that for the  $\alpha$ -globin ( $4.3 \mu\text{M}^{-1} \text{s}^{-1}$ ), also implying that the modules M2 + M3 of the chimera  $\beta\alpha\beta$ -globin regulate ligand-binding kinetics. The rate constants of other chimera globins, however, are different from those predicted from their modules M2 + M3. As summarized in Table 1, these chimera globins exhibited more multiple phases in the CO rebinding reaction than native globins. These multiphasic kinetics was observed for the CO rebinding process of mini-Mb, which is composed of 108 amino acid residues (32nd–139th) of Mb, and the complicated kinetics has been considered to come from the substantial protein fluctuation upon ligand binding and dissociation [24]. Since large structural destabilization was observed in the chimera globins as described above and the thermal stability was also significantly reduced (results not shown), the multiphasic kinetics for these chimera globins might imply that they cannot dampen the amplitude of the structural changes induced by ligand binding.

### 3.2. Interaction between modules

Recent studies for myoglobin by using NMR spectroscopy, mutagenesis and computer graphics

suggested that the packing among the A (in module M1), G and H helices (in module M4) plays important roles in the protein folding and stability as illustrated in Fig. 1b, since the folding and packing of the A, G, and H helices has been considered to be the first step for the dominant folding pathway of apomyoglobin [25–28]. Based on the amide proton exchange rates, Hughson et al. [26] listed several amino acid residues which are protected from the exchange and supposed to be used for the packing of the helices (Table 2). Although some of the amino acid residues are highly conserved in globin proteins, the positions of A9, B11, G13 and G15 are occupied by the different amino acid residues among human hemoglobin subunits and myoglobin as listed in Table 3. In particular, at the position of A9 and B11, hydrophobicity of the amino acid residues are quite different. Since the positions of A9, B11, G13, and G15 are located on the modules M1 and M4, it is not surprising that the substitution of the modules M1 and M4 disrupted the packing of the helices of globin, resulting in large destabilization in globin structure. In fact, by introducing disruptive mutations into the A–H and G–H helix packing site, Hughson et al. [25] showed that these mutations significantly shifted the transition pH midpoint for acid denaturation of apomyoglobin to neutral pH.

These observation suggest that some key interactions between modules M1 and M4 would be lost in the chimera globins having extremely low  $\alpha$ -helical contents such as the chimera  $\alpha\alpha\beta$ -globin. With the structural analysis of the low  $\alpha$ -helical contents, chimera globins have not been successful due to its low stability. We have not yet elucidated the key interactions or amino acid residues for the stable

Table 2

Amino acid residues having large protection (> 50,000) in apomyoglobin and the corresponding amino acid residues in human hemoglobin subunits

Residue in myoglobin	Position	Protection factor [26]	Residue in $\alpha$ -subunit	Residue in $\beta$ -subunits
Val-10	A8	100,000	Val	Val
Leu-11	A9	> 100,000	Lys	Thr
Trp-14	A12	> 100,000	Trp	Trp
Val-17	A15	80,000	Val	Val
Ile-30	B11	70,000	Glu	Gly
Ile-112	G13	60,000	Leu	Val
Val-114	G15	200,000	Thr	Val
Leu-115	G16	60,000	Leu	Leu

folding of the globins. Instead of the structural analysis of the unstable chimera globins, we engineered and characterized the chimera  $\beta\alpha\beta$ - and Mb $\alpha$ Mb-globin and compared the properties of the corresponding chimera globins, the chimera  $\alpha\alpha\beta$ - and  $\alpha\alpha$ Mb- (Mb $\alpha\alpha$ -) globin to examine the effects of the packing of modules M1 and M4 on the protein stability and subunit association property.

As shown in Table 1 and Fig. 2, the electronic absorption, fluorescence and CD spectra have revealed that the chimera  $\beta\alpha\beta$ -globin bound the heme stoichiometrically at an apolar heme-binding pocket and forms much higher  $\alpha$ -helical structures than the chimera  $\alpha\alpha\beta$ -globin in the absence and presence of heme. The emission maximum (456 nm) of ANS bound to the chimera Mb $\alpha$ Mb-globin also corresponds to that for the  $\alpha$ -subunit (456 nm) and significantly differs from that for the chimera Mb $\alpha\alpha$ - (459 nm) and  $\alpha\alpha$ Mb-globin (461 nm). The  $\alpha$ -helical contents of the chimera Mb $\alpha$ Mb-globin are much higher than those of the chimera Mb $\alpha\alpha$ -globin in the absence and presence of heme (Table 1). The thermal stability of the chimera Mb $\alpha$ Mb-globin was as high as that of the  $\alpha$ -globin, whereas the chimera Mb $\alpha\alpha$ -globin was highly destabilized (results not shown). The stable structures of the chimera  $\beta\alpha\beta$ - and Mb $\alpha$ Mb-globins imply that the interactions between modules M1 and M4 are essential for folding and stability in globins.

### 3.3. Association properties of 'module'-substituted globins

In Fig. 3a–g, several elution patterns of the gel chromatography for the chimera  $\alpha\alpha\beta$ - and  $\beta\alpha\beta$ -globins and their mixture with native globins are shown. Trace D is the elution pattern for the chimera  $\alpha\alpha\beta$ -globin and the elution peak was observed at the position for the tetramer. By addition of the  $\alpha$ -subunit (trace E), two peaks appeared, which correspond to the monomeric  $\alpha$ -subunit and tetrameric  $\alpha\alpha\beta$ -subunit, indicating that the chimera  $\alpha\alpha\beta$ -globin does not associate to the  $\alpha$ -subunit. On the other hand, the chimera  $\alpha\alpha\beta$ -globin gave a single peak as a tetrameric globin in the presence of the  $\beta$ -subunit. Since the  $\beta$ -subunit is equilibrium between the monomer and tetramer under the conditions we used, the single peak at the tetrameric globin corresponds

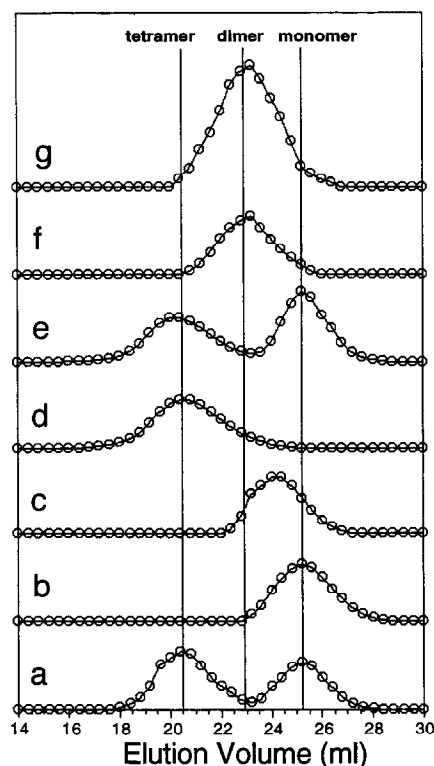


Fig. 3. Chromatography of carbonmonoxy forms of globins on a Sephacryl S-200 HR column (1.0×64.5 cm). (a) the mixture of carbonmonoxy myoglobin (Mb) and cross-linked hemoglobin (XL Hb); (b)  $\alpha$ -globin; (c)  $\beta$ -globin; (d) chimera  $\alpha\alpha\beta$ -globin; (e) the mixture of the  $\alpha$ - and chimera  $\alpha\alpha\beta$ -globins; (f) chimera  $\beta\alpha\beta$ -globin; (g) the mixture of the  $\alpha$ - and chimera  $\beta\alpha\beta$ -globins. Experimental conditions were as follows: Concentration, 1  $\mu$ M on heme basis. The buffer used was 50 mM Tris-HCl, 0.1 M NaCl and 1 mM Na<sub>2</sub>EDTA, pH 7.4 at 10°C and saturated with CO. The eluents were monitored by measuring the absorbance at 420 nm.

to the specific binding of the chimeric  $\alpha\alpha\beta$ -globin to the  $\beta$ -subunit, ( $\alpha\alpha\beta$ )<sub>2</sub> $\beta$ <sub>2</sub> (results not shown).

The elution peak for the chimera  $\beta\alpha\beta$ -globins was observed at the position for the dimeric globins (trace F). Regardless of addition of the monomeric  $\alpha$ -subunit, no additional peak was observed and the elution peak remains at the same position with increased intensity (trace G). The disappearance of the monomeric peak and increased intensity of the peak for the dimeric globins indicates that the chimera  $\beta\alpha\beta$ -globin binds to the  $\alpha$ -subunit to form a dimer,  $\alpha(\beta\alpha\beta)$ . We examined other chimeric globins and the association properties are compiled in Table 3.

Table 3

Association properties of the chimera globins and the mixtures with the  $\alpha$ - or  $\beta$ -globin

	$\alpha$	$\beta$	Mb	$\beta\beta\alpha$	$\alpha\alpha\beta$	$\alpha\alpha\text{Mb}$	MbMb $\alpha$	MbMb $\beta$	Mb $\alpha\alpha$	Mb $\beta\beta$	$\beta\alpha\beta$	Mb $\alpha\text{Mb}$
itself	m	m ~ t	m	t	t	d	@	d	@	t	d	@
+ $\alpha$		d ~ t	—	—	—	—	—	—	—	—	d	—
+ $\beta$	d ~ t		—	t	t	d	d	d	@	—	—	—

Abbreviations: m, monomer; d, dimer; t, tetramer; @, aggregate; —, not associate. The association properties of all chimera globins were independent of their concentrations (1–30  $\mu\text{M}$ ).

Assuming that the chimera globins form the similar tertiary and quaternary structures as native globins, we can estimate the number of the intersubunit hydrogen bonds at the  $\alpha_1\beta_1$  subunit interface which is the most appropriate for the subunit association [29,30]. As listed in Table 4, the number of hydrogen bonds at the  $\alpha_1\beta_1$  surface increases from 3 to 7 by mixing the chimera  $\beta\beta\alpha$ -, MbMb $\alpha$ - or Mb $\alpha\alpha$ -globins with the  $\beta$ -globin, from 4 to 5 by mixing the chimera  $\alpha\alpha\text{Mb}$ -globin with the  $\beta$ -globin. These chimeric globins are found to bind to the  $\beta$ -globin (Table 3). As illustrated in Fig. 3g, the chimera  $\beta\alpha\beta$ - bound to the  $\alpha$ -globin and the hydrogen bonds at the  $\alpha_1\beta_1$  surface also increases from 6 to 7 as listed in Table 4.

However, some 'module'-substituted globins exhibited the complicated association properties. One of such chimeric globins is the chimera  $\alpha\alpha\beta$ -globin. As clearly shown in Fig. 3d, the chimera  $\alpha\alpha\beta$ -globin did not associate with the  $\alpha$ -subunit (Fig. 3e), although the number of the hydrogen bonds for the subunit association can be increased by the binding of the  $\alpha$ -subunit (Table 3). It should be noted here that the structure of the chimera  $\alpha\alpha\beta$ -globin is highly destabilized and the interface could be much more perturbed, compared to that of native globins. In other words, the number of the hydrogen bonds at

the subunit interface might be different from that we estimated for the  $\alpha\alpha\beta$ -chimera globin.

It is quite interesting that the chimera  $\beta\alpha\beta$ -globin preferentially binds to the  $\alpha$ -globin (Fig. 3g). Since the module M1 does not bear so many amino acid residues responsible for the subunit interactions as illustrated in Fig. 1c, the chimera  $\beta\alpha\beta$ -globin would afford the same association property as observed for the chimera  $\alpha\alpha\beta$ -globin. However, the chimera  $\alpha\alpha\beta$ -globin associate with the  $\beta$ -subunit to form a tetramer,  $(\alpha\alpha\beta)_2\beta_2$ , not the  $\alpha$ -subunit (Fig. 3a–g). The different association property of the chimera  $\beta\alpha\beta$ -globin from that of the chimera  $\alpha\alpha\beta$ -globin might be originated from the stability of the globin structure; the chimera  $\beta\alpha\beta$ -globin has much higher  $\alpha$ -helical contents than the chimera  $\alpha\alpha\beta$ -globin (Fig. 2 and Table 1). Thus, we can suggest that the protein folding in the chimera  $\alpha\alpha\beta$ -globin is somewhat different from that in native globin and the structure of subunit interface is highly perturbed, which prevent some key amino acid residues from forming the hydrogen bond between the subunits for the association with the  $\alpha$ -subunit. By the substitution of the module M1, the globin structure would be stabilized by stable packing between the modules M1 and M4 and be able to form the subunit interface for the specific binding to the  $\alpha$ -subunit.

Table 4

The estimated number of intersubunit hydrogen bonds at the  $\alpha_1\beta_1$  subunit surface

	$\alpha$	$\beta$	Mb	$\beta\beta\alpha$	$\alpha\alpha\beta$	$\alpha\alpha\text{Mb}$	MbMb $\alpha$	MbMb $\beta$	Mb $\alpha\alpha$	Mb $\beta\beta$	$\beta\alpha\beta$	Mb $\alpha\text{Mb}$
itself	3	6	4	3	6	4	3	6	3	6	6	4
+ $\alpha$	3	<b>7</b>	4	3	<b>7</b>	4	3	<b>7</b>	3	<b>7</b>	<b>7</b>	4
+ $\beta$	<b>7</b>	6	4	<b>7</b>	6	<b>5</b>	<b>7</b>	6	<b>7</b>	6	6	<b>5</b>

The mixtures indicated by bold-face type means that the association in the chimera globin with the native globin is predicted.



#### 4. Conclusion

By using ‘module’ substitutions corresponding to the artificial exon shuffling, we were able to prepare several novel globins and change subunit association properties of the globins. It was also elucidated that the ‘module’ substitutions often reduce structural stability of the globins due to lack of the interactions between the modules. These results indicate that the packing of the modules is quite essential for stable functional proteins and modules would not be independent structural and functional units. If the interactions between modules can be fine-tuned for stable protein folding, for example, by ‘module’ substitution in combination with site-specific mutagenesis, the ‘module’ substitution based on the artificial exon shuffling would be available as one of the potent strategies to design and produce stable functional proteins.

#### Acknowledgements

We are grateful to Yoshinao Wada (Osaka Medical Center and Research Institute for Maternal and Child Health) for mass spectrometry and Takashi Konno (Kobe University) and Hirokazu Tamamura (Kyoto University) for circular dichroism spectrometry. This work was supported in part by grants from the Ministry of Education, Science, Culture and Sports of Japan (07409003 and 06808058 to I.M. and K.I.). K.W. was supported by the JSPS Research Fellowships for Young Scientists.

#### References

- [1] W. Gilbert, *Nature* 271 (1978) 501.
- [2] C.C.F. Blake, *Nature* 277 (1979) 598.
- [3] M. Go, *Nature* 291 (1981) 90–92.
- [4] M. Go, *Proc. Natl. Acad. Sci. U.S.A.* 80 (1983) 1964–1968.
- [5] M. Go, *Adv. Biophys.* 19 (1985) 91–131.
- [6] W. Gilbert, M. Marchionni, G. McKnight, *Cell* 46 (1986) 151.
- [7] M. Go, M. Nosaka, *Cold Spring Harbor Symp. Quant. Biol.* 52 (1987) 915–924.
- [8] R.L. Dorit, L. Schoenbach, W. Gilbert, *Science* 250 (1990) 1377–1382.
- [9] L. Patthy, *BioEssays* 13 (1991) 187–192.
- [10] R.F. Doolittle, *Science* 253 (1991) 677–680.
- [11] I. Kumagai, S. Takeda, K. Miura, *Proc. Natl. Acad. Sci. U.S.A.* 89 (1992) 5887–5891.
- [12] M. Zhou, C.E. Gomez-Sanchez, D. Xue, M.F. Foecking, *Biochem. Biophys. Res. Commun.* 199 (1994) 130–135.
- [13] K. Wakasugi, K. Ishimori, K. Imai, Y. Wada, I. Morishima, *J. Biol. Chem.* 269 (1994) 18750–18756.
- [14] W.A. Eaton, *Nature* 284 (1980) 183–185.
- [15] K. Inaba, K. Wakasugi, K. Ishimori, I. Morishima, *J. Biol. Chem.* 272 (1997), in press.
- [16] Y. Wada, T. Matsuo, T. Sakurai, *Mass Spectrom. Rev.* 8 (1989) 379–434.
- [17] Y.-H. Chen, J.T. Yang, H.M. Martinez, *Biochemistry* 11 (1972) 4120–4131.
- [18] M. Unno, K. Ishimori, I. Morishima, *Biochemistry* 29 (1990) 10199–10205.
- [19] R. Valdes Jr., G.K. Ackers, *J. Biol. Chem.* 252 (1977) 74–81.
- [20] L. Stryer, *J. Mol. Biol.* 13 (1965) 482–495.
- [21] C.S. Craik, S.R. Buchman, S. Beychok, *Nature* 291 (1981) 87–90.
- [22] E. Bucci, C. Fronticelli, *Biochim. Biophys. Acta* 243 (1971) 170–177.
- [23] S.N. Loh, M.S. Kay, R.L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* 92 (1995) 5446–5450.
- [24] E.E. Di Iorio, W. Yu, C. Calonder, K.H. Winterhalter, G.D. Sanctis, G. Falcioni, F. Ascoli, B. Giardina, M. Brunori, *Proc. Natl. Acad. Sci. U.S.A.* 90 (1993) 2025–2029.
- [25] F.M. Hughson, D. Barrick, R.L. Baldwin, *Biochemistry* 30 (1991) 4113–4118.
- [26] F.M. Hughson, P.E. Wright, R.L. Baldwin, *Science* 249 (1990) 1544–1548.
- [27] P.A. Jennings, P.E. Wright, *Science* 262 (1993) 892–896.
- [28] J.D. Hirst, C.L. Brooks III, *Biochemistry* 34 (1995) 7614–7621.
- [29] G.E.O. Borgstahl, P.H. Rogers, A. Arnone, *J. Mol. Biol.* 236 (1994) 817–830.
- [30] B. Shaanan, *J. Mol. Biol.* 171 (1983) 31–59.