Protein Science (1994), *3*:419–427. Cambridge University Press. Printed in the USA.
Copyright © 1994 The Protein Society

Characteristics of a de novo designed protein

TOSHIKI TANAKA, HIROMI KIMURA, MAYUMI HAYASHI, YOSINORI FUJIYOSHI, KEN-ICHI FUKUHARA, AND HARUKI NAKAMURA
Protein Engineering Research Institute, Osaka 565, Japan

(RECEIVED August 31, 1993; ACCEPTED January 5, 1993)

Abstract

A series of 204 amino acid proteins intended to form TIM (triose phosphate isomerase) barrel structures were designed de novo. Each protein was synthesized by expression of the synthetic gene as a fusion protein with a portion of human growth hormone in an *Escherichia coli* host. After BrCN treatment, the protein was purified to homogeneity. The refolded proteins are globular and exist as monomers. One of the designed proteins is stable toward guanidine hydrochloride (GuHCl) denaturation, with a midpoint of 2.6 M determined from CD and tryptophan fluorescence measurements. The GuHCl denaturation is well described by a 2-state model. The NMR spectra, the thermal denaturation curves, and the 1-anilino-8-naphthalene sulfonic acid binding imply that the stability of the protein arises mainly from hydrophobic interactions, which are probably of a nonspecific nature. The protein has a similar shape to that of rabbit triosephosphate isomerase, as determined by electron microscopy.

Keywords: de novo designed protein; molten globule; secondary structure; tertiary structure; TIM barrel

The de novo design of a protein is one of the methods used to evaluate current ideas of protein organization. It may be suitable for the first trial to produce tertiary structural motifs that are often seen in natural proteins. These motifs may include a 4-helical bundle protein, an immunoglobulin-like protein, or a TIM (triose phosphate isomerase) barrel protein. Until recently, only a few de novo designed proteins were reported. They are mainly proteins composed of α-helical structures, because the formation and stability of α-helices have been extensively studied. Regan and DeGrado (1988) reported the design of a 4-helical bundle protein by a minimalist approach where 4 identical α-helices were linked by 3 identical loops. They also succeeded in generating a metal-binding site within the protein (Handel & DeGrado, 1990; Regan & Clarke, 1990). Hecht et al. (1990) also reported the design of a 4-helical bundle protein in which the amino acid sequence is nonrepetitive. Proteins with β-sheet structures were reported by Moser et al. (1987) and Pessi et al. (1993). A structure consisting of 2 α-helices on 4 β-stranded antiparallel β-sheets by a repeated α-β-β unit was designed and synthesized, although a detailed structural analysis was not performed (Fedorov et al., 1992). The TIM barrel structure, typified by triosephosphate isomerase, consists of a parallel β-barrel core of 8 strands surrounded by 8 α-helices, $(\beta\alpha)_8$. This type of structure can be seen in many proteins, although the amino acid sequences are not related, even within a given protein. Therefore, it is still in dispute as to whether this structure is formed by divergent evolution from a single common ancestor, or by convergent evolution to a stable fold (Farber & Petsko, 1990). The design of a TIM barrel structure was reported by Goraj et al. (1990), and the protein seems to be folded, albeit loosely. Based upon the concept that the secondary structures are formed first and are assembled by hydrophobic interactions, we designed a TIM barrel structure. The designed protein was synthesized and characterized using physicochemical measurements, including electron microscopy (Fujiyoshi et al., 1980).

Reprint requests to: Toshiki Tanaka, Protein Engineering Research Institute, 6-2-3 Furuedai, Suita, Osaka 565, Japan.

Results and discussion

Design of amino acid sequence

The TIM barrel structure is composed of 8 fold repeats of a loop-β-loop-α motif, $(\beta\alpha)_8$. There are 2 types of β-strand, which are discriminated by the orientation of the amino acid side chains on the β-strand. For example, on the first β-strand, the side chain is directed toward the center, and the second is toward the outside, with the rest alternating in and out accordingly. On the second β-strand, the associated side chain is directed toward the outside, and the next is toward the center, with the rest alternating out and in accordingly. Thus, the odd-numbered β-strands have the same orientation, and the even-numbered β-strands have the same orientation. Therefore, we designed the 4 periodic structural units of $\beta_1\alpha_1\beta_2\alpha_2$.

The β-strand of a TIM barrel structure is considered to be a hyperboloid (Lasters et al., 1988; Lasters, 1990). We defined the equatorial plane of the hyperboloid as 0 and the parallel planes are numbered as −2, −1, 0, +1, and +2 from the NH_2-terminus of the β-strand. Table 1 shows the frequency of amino

Table 1. *Number of amino acids appearing at a given plane of the β-strand*[a]

	Y	F	L	V	I	A	G	T	M	C	P	S	R	K	E	D	Q	N	W
A. The plane having outward side chain																			
−2				3	1								1						
−1		1	2	4	4	1			1	1	1	1							
0		2	4	4	2	1		1		1									
+1	2	1	2	3	1	2	2	1			1						1		
+2	1		5	2	4							1							
B. The plane having inward side chain																			
−2	1	1				3	3				2		1				1		
−1	1	2	2	2		2	4		1			1							1
0		1	1	1	6	2	3	1	1										
+1	1		3		3	2	1						1	1	1	1	1		1
+2		2	1					1		1		2				1	1	1	

[a] The number is calculated using 4 species of natural TIM barrel proteins. Therefore, each plane has 16 amino acids. Some planes have less than 16 amino acids due to ambiguity of the direction of the amino acid side chain.

acids appearing at a given position on the β-strand from the 4 natural proteins, triosephosphate isomerase, tryptophan synthase α-subunit, glycolate oxidase, and Taka-amylase. Frequently used amino acids toward the outside of the β-strand are Val at −2, Val and Ile at −1, Leu and Val at 0, Val at +1, and Leu and Ile at +2 position. For the inside positions, Ala and Gly at −2, Gly at −1, Ile at 0, Leu and Ile at +1, Phe and Ser at +2 are often used. Subsequently the amino acid sequence of the designed TIM barrel protein was chosen using these data. One of the β-strands with side chains directed in the in/out pattern is β_1, and the amino acid sequence is Gly-Val-Ile-Tyr-Phe. On the other β-strand, β_2, the amino acid sequence is Val-Val-Leu-Gly-Ile. Although Gly at an interior +1 position is not predominant, we chose Gly because the amino acid on the neighboring β-strand is bulky Phe. Because these amino acid sequences lack Trp, we designed an alternate amino acid sequence, Ala-Ile-Ile-Val-Thr for β_1 and Val-Phe-Val-Trp-Leu for β_2.

The amino acid sequences of 11 residues for the α-helix should be amphiphilic and were chosen based on the statistical results reported by Richardson and Richardson (1988). The N-cap residue is Asn and the N1 position is Pro. Asn at the N-cap stabilizes a short α-helical peptide (Lyu et al., 1992). Glu and Lys were chosen for the hydrophilic helical residues and they are placed at the N2, N3 and C2, C1 positions, respectively. Lys is often used at the C5 position in natural proteins, and it may form an ion pair with Glu at the N2 position to stabilize the α-helix. At the C-cap position, Gly is the most favored. At the remaining positions, N4, N5, C4, and C3, Leu was selected because their side chains are directed toward the inside of the protein on a wheel model. It was considered that Leu at the C3 position could be substituted by Ala.

The loop from the α-helix to the β-strand is short and usually composed of 3 or 4 amino acids. One of the loops was designed to form a tight turn and therefore we selected Pro. Arg has a guanidino group with a long side chain and can make an ion pair with an acidic residue. Arg may be rare on a turn. However, Regan and DeGrado (1988) used a Pro-Arg-Arg sequence for the turn to construct a 4-helical bundle protein. The other turn from an α-helix to a β-strand is rather flexible, and the amino acid sequence Ser-Ala-Asp-Thr was chosen from the loop between α_1 to β_2 of triosephosphate isomerase because it has an Asp that can form an ion pair with Arg on the neighboring turn. Loops from a β-strand to an α-helix in natural proteins are relatively long, ranging from 3 to 12 amino acids in length, and may possess active sites. This loop may not be important for protein folding and stability (Urfer & Kirschner, 1992). Therefore, the amino acid composition of this loop was not critical, and we chose the amino acid sequences Gly-Leu-Asp-Gly from β_5 to α_5 of Taka-amylase and Gly-Ser-Val-Thr-Gly-Gly from β_7 to α_7 of triosephosphate isomerase. The amino acid sequences for $\beta_1\alpha_1\beta_2\alpha_2$, designated as L, A, and βL, are shown in Figure 1 along with the nucleotide sequences of the artificial gene. Secondary structure predictions using the sequence homology method (Nishikawa & Ooi, 1986), the Ptitsyn–Finkelstein method (Ptitsyn & Finkelstein, 1983), and the Gibrat–Garnier–Robson method (Gibrat et al., 1987) are in agreement with the designed structure.

To construct the TIM barrel structure, the amino acid sequence for L, as an example, is repeated 4 times as L-L-L-L. The protein A-A-A-A, where 4 Leu residues were substituted with 4 Ala residues on the odd-numbered α-helix, and the proteins L-βL-L-L, which contained Trp to analyze the environment of the indole side chain, were also constructed by replacing the $\beta_1\alpha_1\beta_2\alpha_2$ unit.

Gene design and expression

The amino acid sequence of 1 unit for $\beta_1\alpha_1\beta_2\alpha_2$, as well as its nucleotide sequence, are shown in Figure 1. The nucleotide sequence was deduced using the preferred codon usage of *Escherichia coli* (Ikemura, 1981). In order to prepare an expression plasmid, the gene for $\beta_1\alpha_1\beta_2\alpha_2$, which is preceded by 10 amino acids and followed by 2 amino acids, was constructed from 12 oligonucleotides with lengths between 48 and 55 nucleotides. The oligonucleotides were ligated into the *Eco* RI-*Sal* I site of pUC 119 and sequenced. The short *Bgl* II-*Sal* I fragment and the large *Bam* HI-*Sal* I fragment of the resultant plasmid were ligated to produce the plasmid containing the gene for $(\beta_1\alpha_1\beta_2\alpha_2)_2$. This procedure was repeated to construct the plasmid pUC-TIM, which has the gene encoding $(\beta_1\alpha_1\beta_2\alpha_2)_4$. The short *Cla* I-*Sal* I fragment of pUC-TIM and the large *Bam* HI-*Sal* I fragment of pGH-L9 (Ikehara et al., 1984; Tanaka et al., 1990) were then ligated with a linker to construct the plasmid, pGH-TIM. Then the gene was expressed as a fusion protein with a portion of human growth hormone (Met-1 to Ile-122) and a heptapeptide spacer (Glu-Asn-Ser-Asp-Pro-Ser-Met) under the control of the *trp* promoter. After BrCN treatment, the synthetic protein has an additional Ile-Glu-Gly-Arg sequence at the NH_2-terminus and a Ser-Ala sequence at the COOH-terminus (Fig. 2). A plasmid for direct expression was also constructed by digestion of the plasmid pGH-TIM with *Cla* I, followed by self-ligation.

Direct expression of the designed protein in *E. coli* was first tried. However, the protein was in the insoluble fraction after disruption of the cell. Therefore, a fusion protein overexpression system was used for the preparation. The fusion protein was abundantly expressed and was sequestered in inclusion bodies in the cell. The insoluble protein was collected after cellular disruption. Then the fusion protein was treated with BrCN and the designed protein was purified by reverse-phase HPLC (Fig. 2).

```
        N  S  D  P  S  M  I  E  G  R
        GAATTCCGACCCATCGATGATTGAAGGTAGA
        Eco RI       Cla I         Bgl II

L:      loop          β1            loop             α1
        S A D T       G V I Y F     G L D G          N P E E L L K L L K K G
        TCTGCTGACACT  GGTGTTATTTACTTC GGTCTAGATGGT   AACCCAGAAGAACTGTTAAAGCTACTGAAGAAAGGT
                                      Xba I

        loop          β2            loop             α2
        P R G         V V L G I     G S V T G G      N P E E L L K L L K K G
        CCGCGTGGC     GTAGTTTTAGGTATC GGTTCTGTAACTGGTGGC AACCCGGAAGAATTACTGAAACTGCTTAAGAAAGGA
                                                                              Bam HI

                                                     S A * *
                                                     TCCGCATAATAGTTAACGTCGAC
                                                                     Sal I

A:                                                   α1
                                                     N P E E L L K L A K K G
                                                     AACCCAGAAGAACTGTTAAAGCTAGCAAAGAAAGGT

βL:                   β1
                      A I I V T
                      GCAATCATTGTTACC

                      β2
                      V F V W L
                      GTATTCGTTTGGCTG
```

Fig. 1. Designed amino acid sequences for $\beta_1\alpha_1\beta_2\alpha_2$. In the case of L, the designed amino acid sequence is shown in large letters. Under the amino acid sequence, only the upper strands of the nucleotide sequences are shown. Amino acid sequences written in bold letters come from the plasmid construction and asterisks show stop codons. The restriction sites are indicated in italics. The gene between the *Eco* RI and *Sal* I sites was constructed, cloned into pUC119, and sequenced. For A and βL, only the different amino acids and DNA sequences in the α-helices or β-strands are shown.

The purified proteins have the expected amino acid sequences in the first 15 residues. The amino acid compositions are also consistent with the expected sequences.

All the designed proteins were refolded by stepwise dialysis toward 50 mM Tris-HCl at pH 7. In aqueous buffer, they are soluble in millimolar scale and exist as globular monomers by size exclusion chromatography. The apparent molecular masses (28.8 kDa) of the proteins A-A-A-A, L-L-L-L, and L-βL-L-L, are slightly larger than the calculated monomeric molecular masses (21.6, 21.7, and 21.8 kDa, respectively).

CD spectra

CD spectra of the 3 proteins were measured in phosphate buffer at pH 7 (Fig. 3). All the proteins exhibited similar shapes in the CD spectra, with minima at 206 and 220 nm, suggesting that all the proteins may have similar structures. These CD spectra are somewhat different from that of native triosephosphate isomerase (Manavalan & Johnson, 1983). The minimum at 206 nm suggests that the designed proteins may have more β-strand content than that predicted from the design. The CD spectra did not change between pH 1.6 and 7.

The stability of the designed proteins toward denaturation by guanidine hydrochloride (GuHCl) was measured by monitoring the mean residue ellipticity at 222 nm as a function of GuHCl concentration (Fig. 4). The midpoint of the unfolding transition was 1.8 M GuHCl for the protein L-L-L-L. The protein A-A-A-A was slightly less stable than the protein L-L-L-L. When Trp and bulky amino acids were used in the β-strand, such as in the protein L-βL-L-L, it was much more stable than the protein L-L-L-L, as shown by the midpoint of 2.6 M GuHCl.

Fluorescent spectra

The protein L-βL-L-L is relatively stable toward denaturation by GuHCl and contains only 1 Trp in the protein. Therefore we undertook further characterization using this protein. The fluorescence spectra of the protein L-βL-L-L in aqueous buffer and in the presence of 6 M GuHCl were measured (Fig. 5). The maximum at 327 nm in the aqueous buffer was shifted to 351 nm in the denaturing condition, indicating that the Trp was buried in the folded protein.

The stability of the protein toward denaturation by GuHCl was measured by monitoring the intensity of fluorescence at 328 nm as a function of GuHCl concentration (Fig. 6). The denaturation curve obtained was identical to that obtained from the CD measurement, indicating a 2-state transition.

NMR spectra

The ^{1}H-NMR spectrum of the protein L-βL-L-L, measured at pH 5 in D_2O, is too broad for a monomer form with this molecular weight (Fig. 7). The aromatic protons gave a very broad spectrum and the amide proton signals around 8–9 ppm were not observed. This broad signal is not due to aggregation of the protein because the spectrum was not changed even at a low con-

Fig. 2. A: Strategy to obtain the designed protein using the expression plasmid, pGH-TIM. pGH-TIM was constructed by ligation of the *Cla* I–*Sal* I fragment of pUC-TIM, the *Bam* HI–*Sal* I fragment of pGH-L9, and a linker. The nucleotide sequences of the linker are 5′-GATCGAGAATTCCGACCCAT-3′ (upper strand) and 5′-CGATGGGTCGGAATTCTC-3′ (lower strand). Underlined are *Eco* RI sites. The *Bam* HI site between pGH-L9 and the linker becomes inactivated after ligation (ΔBa). The numbered boxes indicate the 4 repeats of the $\beta_1\alpha_1\beta_2\alpha_2$ unit. The following abbreviations are used in this figure: C, *Cla* I; Ba, *Bam* HI; Bg, *Bgl* II; E, *Eco* RI; H, *Hin* dIII; S, *Sal* I; A^r, ampicillin resistance; *trp*, the *trp* promoter. **B:** SDS-polyacrylamide gel (15%) electrophoresis analysis of the expressed designed protein L-βL-L-L. Lane 1, total protein; lane 2, soluble fraction after sonication; lane 3, insoluble fraction including the fusion protein; lane 4, fusion protein cleaved with BrCN; lane 5, HPLC-purified designed protein; lane M, protein molecular weight markers: from top to bottom, 92.5, 66.2, 45, 31, 21.5, and 14.4 kDa.

centration of the protein (0.2 mg/mL). The NMR spectrum of β-lactamase is broader in molten globule state than in native state, and almost no amide protons are detected (Ptitsyn et al., 1990). It is likely that the protein contains a mixture of many structural species and/or is in a molten globule state.

ANS binding

1-Anilino-naphthalene-8-sulfonate (ANS) is useful as a probe for the molten globule state of proteins. ANS itself has no fluorescence in aqueous buffer, but in a hydrophobic environment

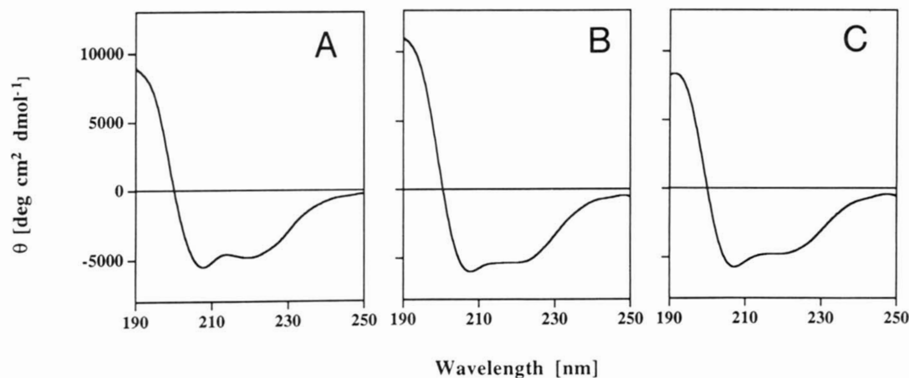

Fig. 3. CD spectra of the designed proteins. The protein concentration was 10 μM. Spectra were obtained at 20 °C in 1 mM sodium phosphate, pH 7. **A:** A-A-A-A. **B:** L-L-L-L. **C:** L-βL-L-L.

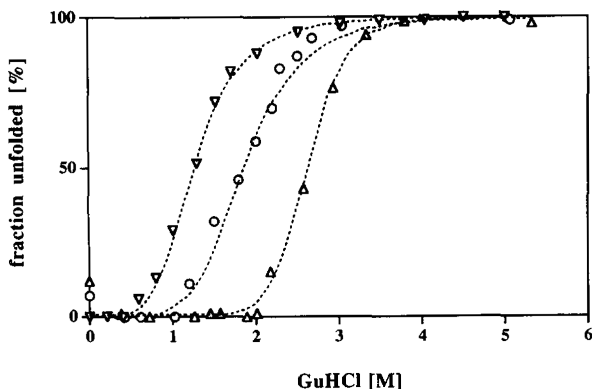

Fig. 4. GuHCl denaturation curves of the designed proteins. The ellipticity at 222 nm was monitored as a function of GuHCl concentration. ▽, A-A-A-A; ○, L-L-L-L; △, L-βL-L-L.

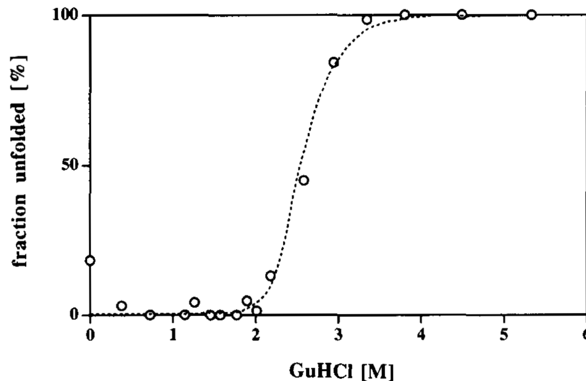

Fig. 6. GuHCl denaturation curve of the designed protein (L-βL-L-L), monitored by fluorescence. The fluorescence intensity at 328 nm was monitored as a function of GuHCl concentration. The excitation wavelength was 280 nm.

it fluoresces. Therefore, the fluorescent intensity of ANS is much increased in the molten globule state of proteins than in the native and denatured states (Semisotnov et al., 1991; Brchkova et al., 1992; Uversky et al., 1992). The fluorescence of ANS at 480 nm was measured when ANS and the designed protein were mixed together in various GuHCl concentrations (Fig. 8). Even in the absence of GuHCl, fluorescence of ANS was observed. At concentrations of GuHCl less than 1.5 M, when the protein seems to be folded by CD and Trp fluorescence measurements, ANS can bind to the protein. ANS sometimes can bind to folded natural proteins if they have a hydrophobic hollow, such as an active site (Semisotnov et al., 1991). However, it seems to be reasonable to speculate that the designed protein is not packed well internally and that there is enough space for ANS to bind to the protein.

Thermal denaturation

We measured the stability of the protein toward thermal denaturation. The ellipticity at 222 nm was monitored as a function of temperature (Fig. 9). At 0 M GuHCl, no transition is observed and the protein is stable between 4 and 90°C. With 1.5 M GuHCl, where the protein is still folded by CD and fluorescence measurements, the protein is stable at high temperature. However, the protein is less stable at low temperature. With 2.5 M GuHCl, at the midpoint of denaturation, cold denaturation is observed. At high temperature, the protein is again stable and forms a similar structure to that observed in the absence of GuHCl. At 3.5 M GuHCl, cold denaturation was also observed.

Natural proteins are often destabilized under acidic conditions. Therefore, we analyzed the thermal denaturation of the protein at pH 3.5. However, the protein is stable and no transition occurred as at pH 7.

Electron microscopy

Rabbit triosephosphate isomerase and one of the proteins L-L-L-L) were observed by electron microscopy after staining

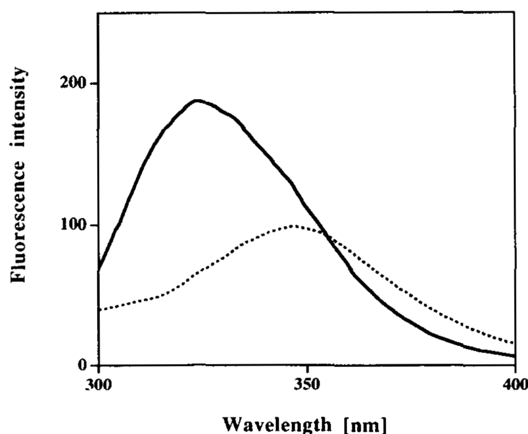

Fig. 5. Fluorescence emission spectra of the protein (L-βL-L-L). The protein was dissolved to 20 μM in 50 mM Tris-HCl, pH 7, at 20 °C in the presence (dotted line) and absence (solid line) of 6 M GuHCl. The excitation wavelength was 280 nm.

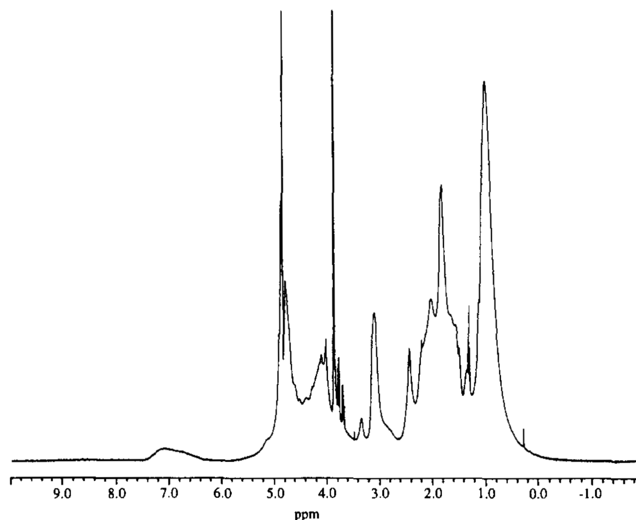

Fig. 7. ^{1}H-NMR spectrum of the protein, L-βL-L-L. The protein was dissolved at a concentration of 5 mg/mL in D_2O containing 10 mM CD_3COONa, pH 5.5 at 20 °C.

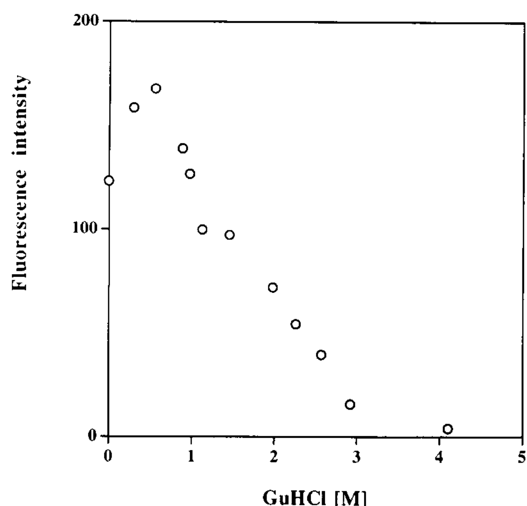

Fig. 8. Change of fluorescence intensity at 480 nm of ANS in the presence of the protein (L-βL-L-L). Fluorescence of ANS was measured at a protein concentration of 15 μM and an ANS concentration of 1.9 mM in 50 mM Tris-HCl, pH 7, at 20 °C, containing the indicated GuHCl concentration.

with uranyl acetate. Rabbit TIM is a protein dimerized by hydrophobic interaction. We could observe both the dimeric and monomeric shapes by electron microscopy, which is often the case when a dimeric protein is formed by hydrophobic interaction (Fujiyoshi, unpubl. data). In Figure 10A, only the doughnut-shaped image of the monomeric protein is shown in comparison with our designed protein because we designed a monomeric TIM barrel structure. The images of the designed protein were similarly shaped. In the same photograph, white oval-shaped images were observed, as indicated by the arrowhead in Figure 10B. The former may correspond to a top view of the protein and the latter to a side view. The size of the protein calculated from the image was about 5 nm in diameter and 3.5 nm in height. This size is very similar to the model structure. This result suggests

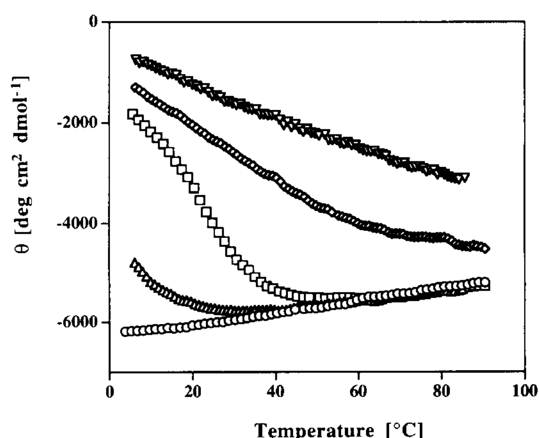

Fig. 9. Thermal denaturation curves of the protein (L-βL-L-L). Ellipticity at 222 nm was monitored as a function of temperature at the indicated GuHCl concentrations. ○, 0 M; △, 1.5 M; □, 2.5 M; ◇, 3.5 M; and ▽, 7.2 M.

the designed protein folded into a compact form. In the case of the protein L-βL-L-L, the same shapes were observed in negatively stained images. Such small size proteins have not, to date, been observed as clear shapes by electron microscopy. For example, human growth hormone, a 4-helical bundle protein, did not exhibit the same shape, although the size was very similar to that of the designed protein.

As shown in Figure 10C, 2 circles close to each other were observed in the double TIM barrel protein, in which the sequence for L-L-L-L was tandemly linked with a flexible linker, $(\text{Ser-Gly-Gly-Gly-Gly-Ser})_2$ (Fig. 11). The double TIM barrel protein has exactly the same CD spectra, and twice the molecular weight by size-exclusion chromatography, as the monomeric protein, L-L-L-L. These results suggest that each domain refolds independently.

Although the designed protein gives a similar image to that of natural TIM by electron microscopy, we cannot conclude that the designed protein has the same tertiary structure as natural TIM. To analyze the structure of the designed protein at an atomic resolution, NMR or X-ray crystallography should be employed.

Conclusion

We designed and synthesized a de novo TIM barrel structure protein. One of the problems of de novo designed proteins is low solubility in aqueous solution, which makes further characterization difficult. The proteins synthesized here are, however, quite soluble in aqueous solution. In addition, the proteins are relatively stable against GuHCl, heat, and pH change. Two-state transitions were observed in the GuHCl denaturation curves. Electron microscopy suggests that our de novo proteins have a similar shape to native TIM, but the TIM-barrel folding has not been confirmed. The peaks of the NMR spectra of the protein are rather broad, and ANS can bind inside the protein without GuHCl, showing that (1) the side chains of the amino acids inside the protein may be flexible, (2) the protein may have many conformational species, and (3) there is enough space inside the protein for ANS to bind with the protein.

The behavior of our designed protein is similar to that of natural proteins in their molten globule state. However, our designed protein is much more stable toward GuHCl and thermal denaturations than natural proteins in the molten globule state. The CD spectra of the protein between pH 1.6 and 7 did not change much, even at high temperature. It is likely that ion pair interactions have little effect and the stability of the protein may arise primarily from hydrophobic interactions. The protein in aqueous buffer is stable toward heat denaturation. On the other hand, in the presence of GuHCl, cold denaturation was observed. This suggests that strong hydrophobic interactions make the protein stable. Because the NMR signals are broad and ANS binds inside the protein, the hydrophobic interactions are probably nonspecific.

Some problems with the present design are: (1) There are 11 Pro residues, which should accompany the *cis–trans* isomerization during refolding. (2) There are 40 Gly residues out of 204 amino acid residues. Gly is a flexible amino acid. This may facilitate the adoption of a molten globule state. (3) We did not carefully choose the amino acid sequence for the loops in this design. For future de novo protein design, the most important point is to create a specific side-chain–side-chain interaction.

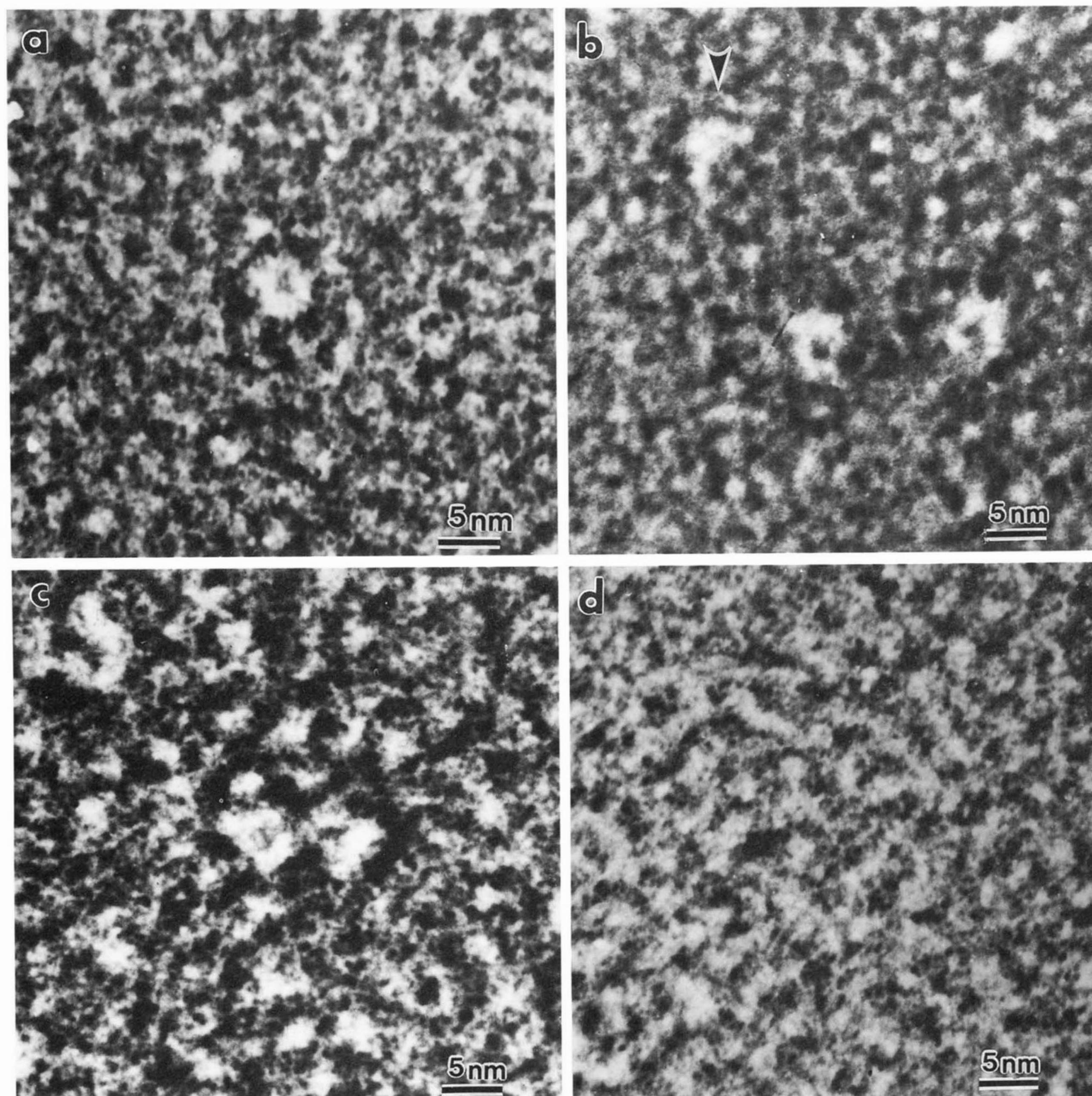

Fig. 10. The negatively stained image of the proteins observed by electron microscopy. **a:** Rabbit triose phosphate isomerase. **b:** The designed protein L-L-L-L. **c:** The double designed protein. **d:** Without protein.

Materials and methods

Expression of the designed proteins

E. coli strain HB101, transformed by the plasmid (pGH-TIM), was cultured at 37 °C in M9 medium containing a 2-fold strength of the salts and 0.5% casamino acids. When the culture had attained a density between 120 to 200 Klett units, protein synthesis was induced by the addition of 3-β-indoleacrylic acid to 10 μg/mL final concentration. After 20 h of induction, cells were harvested and disrupted by sonication. The insoluble material was collected by centrifugation and dissolved in 20 mM Tris-HCl, pH 8, containing 6 M GuHCl and 1% 2-mercaptoethanol. After the insoluble materials were removed by centrifugation, the proteins were precipitated by dialysis against 50 mM Tris HCl at pH 7. The precipitate was dissolved in 70% formic acid containing 0.3 M BrCN and treated for 2 h. The mixture was subjected to reverse-phase HPLC (Aquapore Prep 10, C-4, 300-Å pore size, Applied Biosystems), the protein was eluted by a linear gradient of CH_3CN (30–70%) containing 0.1% trifluoroacetic acid at a flow rate of 3 mL/min, and the fractions containing the desired protein were collected and lyophilized. About 15 mg of the protein was routinely obtained from a 200-mL culture. The purified protein was dissolved in 50 mM Tris-HCl at

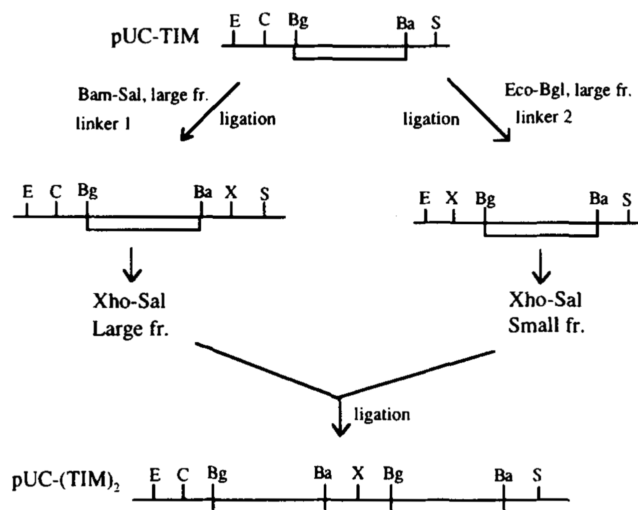

Fig. 11. Strategy to construct the gene for the double TIM. The linkers having nucleotide sequences of 5′-GATCCTCTGGTGGCGG TGGCTCGAGCTTAACG-3′ (linker 1) and 5′-AATTCCGGCTCGAG CGGTGGCGGTGGCTCTA-3′ (linker 2) were inserted into the *Bam* HI-*Sal* I and the *Eco* RI-*Bgl* II site of pUC-TIM, respectively. Only nucleotide sequences of upper strand for the linkers were shown. Underlined are *Xho* I sites. Open boxes indicate the gene for the designed monomeric protein. Ligation of the large *Xho-Sal* fragment of the former plasmid and the small *Xho-Sal* fragment of the latter plasmid yielded the plasmid, pUC-$(TIM)_2$. Amino acid sequences between 2 designed proteins are Ser-$(SerGlyGlyGlySer)_2$-Arg. Ligation of the small *Eco-Sal* fragment of pUC-$(TIM)_2$ and the large *Eco-Sal* fragment of pGH-TIM yielded the expression plasmid, pGH-$(TIM)_2$. Expression and purification of double TIM was exact same as the case of the monomeric protein. The following abbreviations are used in this figure: C, *Cla* I; Ba, *Bam* HI; Bg, *Bgl* II; E, *Eco* RI; H, *Hin* dIII; S, *Sal* I; X, *Xho* I.

pH 7 containing 6 M GuHCl and was refolded by stepwise dialysis against 4 M, 3 M, 2 M, 1 M, and 0 M GuHCl in 50 mM Tris-HCl at pH 7.

The NH_2-terminal sequences of the purified proteins were determined with a gas-phase protein sequencer and HPLC system (models 477A and 120A, Applied Biosystems). The amino acid compositions were analyzed with an amino acid analyzer (Beckman system 6300E), using 4 nmol of the protein, after hydrolysis at 110 °C for 22 h in 6 M HCl.

Molecular size estimation by size-exclusion chromatography

The designed proteins dissolved in 10 mM Tris-HCl (pH 7) containing 0.1 M NaCl were applied to a column of Sephadex G-75 (1.5 × 94 cm), preequilibrated with the same buffer as that of the sample solution, and developed with the same buffer at 0.2 mL/h. The molecular weights were determined from the calculation curve obtained with the standard proteins: bovine serum albumin, 67 kDa; ovalbumin, 43 kDa; chymotrypsinogen, 25 kDa; ribonuclease A, 13.7 kDa.

CD spectra of the protein

The designed proteins were dissolved at 0.3 mg/mL (10 μM) in 1 mM sodium phosphate, pH 7, in a cuvette with a 1-mm path-length. Spectra were measured at 20 °C on a J-600 spectropolarimeter (Japan Spectroscopic).

Fluorescence spectra of the designed proteins in folded and unfolded forms

The designed proteins were dissolved at 0.4 mg/mL (20 μM) in 50 mM Tris-HCl, pH 7, in the presence or absence of 6 M GuHCl. Spectra were measured at 20 °C on a Hitachi F400 fluorescence spectrophotometer at an excitation wavelength of 280 nm.

NMR measurement

The designed protein was dissolved at 5 mg/mL in D_2O containing 10 mM CD_3COONa, pH 5.5. The ^1H-NMR spectrum was recorded at 20 °C on a Bruker AM 600 spectrometer. Chemical shifts are expressed in ppm to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS).

Fluorescence measurement of ANS

The designed proteins were dissolved at 0.3 mg/mL (15 μM) in 50 mM Tris-HCl at pH 7 in the presence of different concentrations of GuHCl. A stock solution of ANS, in 50 mM Tris-HCl, was added to the protein solutions to give final ANS concentration of up to 2 mM. Fluorescence of ANS at 480 nm was measured at 20 °C with the Hitachi fluorescence spectrophotometer.

Electron microscopy

A small droplet (2 mL) of specimen solution was applied to a carbon-coated EM-mesh, and then the excess solution was removed by capillary action using the edges of filter paper. Next the crystals were stained with 2% uranyl acetate solution on the carbon support. Images were taken with an electron microscope, JEM 100CX. It was usually operated at 100 kV, and the image was recorded on Fuji FG film with a dose of about 5,000 electrons/nm^2. The image was developed with a Kodak D-19 developer at 20 °C for 7 min. Magnification was 40,000, where 1 nm of the specimen corresponds to 40 mm on the film.

Acknowledgments

We thank Dr. T. Tanaka for help with initial plasmid construction and Dr. Y. Oda for NMR measurements. We also thank Drs. Y. Kuroda and M. Oobatake for helpful suggestions.

References

Brchkova VE, Berni R, Rossi GL, Kutyshenko VP, Ptitsyn OB. 1992. Retinol-binding protein is in the molten globule state at low pH. *Biochemistry 31*:7566–7571.

Farber GK, Petsko GA. 1990. The evolution of α/β barrel enzymes. *Trends Biochem Sci 15*:228–234.

Fedorov AN, Dolgikh DA, Chemeris VV, Chernov BK, Finkelstein AV, Schulga AA, Alakhov YB, Kirpichnikov MP, Ptitsyn OB. 1992. *De novo* design, synthesis and study of albebetin, a polypeptide with a predetermined three-dimensional structure: Probing the structure at the nanogram level. *J Mol Biol 225*:927–931.

Fujiyoshi Y, Kobayashi T, Ishizuka K, Uyeda N, Ishida Y, Harada Y. 1980. A new method for optimal-resolution electron microscopy of radiation-sensitive specimens. *Ultramicroscopy 5*:459–468.

Gibrat JF, Garnier J, Robson B. 1987. Further developments of protein sec-

ondary structure prediction using information theory: New parameters and consideration of residue pairs. *J Mol Biol 198*:425–443.

Goraj K, Renard A, Martial JA. 1990. Synthesis, purification and initial structural characterization of octarellin, a *de novo* polypeptide modelled on the α/β-barrel proteins. *Protein Eng 3*:259–266.

Handel T, DeGrado WF. 1990. De novo design of a Zn^{2+}-binding protein. *J Am Chem Soc 112*:6710–6711.

Hecht MH, Richardson JS, Richardson DC, Ogden RC. 1990. De novo design, expression, and characterization of Felix: A four-helix bundle protein of native-like sequence. *Science 249*:884–890.

Ikehara M, Ohtsuka E, Tokunaga T, Taniyama Y, Iwai S, Kitano S, Miyamoto T, Ohgi T, Sakuragawa Y, Fujiyama K, Ikari T, Kobayashi M, Miyake T, Shibahara S, Ono A, Ueda T, Tanaka T, Baba H, Miki T, Sakurai A, Onishi T, Chisaka O, Matubara K. 1984. Synthesis of a gene for human growth hormone and its expression in *Escherichia coli*. *Proc Natl Acad Sci USA 81*:5956–5960.

Ikemura T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J Mol Biol 151*:389–409.

Lasters I. 1990. Estimating the twist of β-strands embedded within a regular parallel β-barrel structure. *Protein Eng 4*:133–135.

Lasters I, Wodak SJ, Alard P, van Cutsem E. 1988. Structural principles of parallel β-barrels in proteins. *Proc Natl Acad Sci USA 85*:3338–3342.

Lyu PC, Zhou HX, Jelveh N, Wemmer DE, Kallenbach NR. 1992. Position-dependent stabilizing effects in α-helices: N-terminal capping in synthetic model peptides. *J Am Chem Soc 114*:6560–6562.

Manavalan P, Johnson WC Jr. 1983. Sensitivity of circular dichroism to protein tertiary structure class. *Nature 305*:831–832.

Moser R, Frey S, Munger K, Hehlgans T, Kiauser S, Langer H, Winnacker EL, Mertz R, Gutte B. 1987. Expression of the synthetic gene of an artificial DDT-binding polypeptide in *Escherichia coli*. *Protein Eng 1*:339–343.

Nishikawa K, Ooi T. 1986. Amino acid sequence homology applied to the prediction of protein secondary structure, and joint prediction with existing methods. *Biochim Biophys Acta 871*:45–54.

Pessi A, Bianchi E, Crameri A, Venturini S, Tramontano A, Sollazzo M. 1993. A designed metal-binding protein with a novel fold. *Nature 362*: 367–369.

Ptitsyn OB, Finkelstein AV. 1983. Theory of protein secondary structure and algorithm of its prediction. *Biopolymers 22*:15–25.

Ptitsyn OB, Pain RH, Semisotnov GV, Zerovnik E, Razgulyaev OI. 1990. Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Lett 262*:20–24.

Regan L, Clarke ND. 1990. A tetrahedral zinc(II)-binding site introduced into a designed protein. *Biochemistry 29*:10878–10883.

Regan L, DeGrado WF. 1988. Characterization of a helical protein designed from first principles. *Science 241*:976–978.

Richardson JS, Richardson DC. 1988. Amino acid preferences for specific locations at the ends of α helices. *Science 240*:1648–1652.

Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas AF, Gilmanshin RI. 1991. Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers 31*:119–128.

Tanaka T, Shiratori Y, Saitoh S, Yamamoto Y, Kohara A, Nakamura H, Tanaka T, Kikuchi M, Ikehara M. 1990. Synthesis of a new helical protein: The effect of secondary structure rearrangement on structure formation. *Biochem Biophys Res Commun 170*:104–110.

Urfer R, Kirschner K. 1992. The importance of surface loops for stabilizing an eightfold $\beta\alpha$ barrel protein. *Protein Sci 1*:31–45.

Uversky VN, Semisotnov GV, Pain RH, Ptitsyn OB. 1992. "All-or-none" mechanism of the molten globule unfolding. *FEBS Lett 314*:89–92.