ELSEVIER

Biophysical Chemistry 50 (1994) 47-61

Biophysical Chemistry

De novo design and creation of a stable artificial protein

Toshiki Tanaka, Mayumi Hayashi, Hiromi Kimura, Motohisa Oobatake, Haruki Nakamura *

Protein Engineering Research Institute, 6-2-3 Furuedai, Suita, Osaka 565, Japan

(Received 5 January 1994)

Abstract

Protein *de novo* design has been performed, as an exercise of the inverse folding problem. A β/α-barrel protein was designed and synthesized using the *Escherichia coli* expression system for the structural characterization. A tertiary model with a two-fold symmetry was built, based upon the geometrical parameters extracted from X-ray crystal structures of several β/α-barrel proteins. Amino acid frequencies at each position on the α- and β-structures were investigated, and an amino acid sequence with 201 residues was designed. The associated gene was chemically synthesized and the fusion protein with human growth hormone was expressed in *Escherichia coli*. The purified protein after being cleaved and refolded was found to be stable and globular with the large amount of secondary structures. However, it has similar characteristics to the molten globules of natural proteins, with loose packing of side-chains. The approach for the tight packing is discussed.

Key words: De novo design; β/α-barrel; Protein engineering; Molten globule

1. Introduction

One of the goals of the protein engineering is to understand the folding mechanism of proteins. For that purpose, the usual path of investigation is the analysis from the amino acid sequence to the tertiary structure. The alternative approach from the tertiary structure to the primary structure is called an inverse folding problem, and it can give us very different information from the usual folding problem. That is, the tertiary structure is first considered and the primary sequence is then designed, so as to be the most suitable to the tertiary structure. By this approach, new methods to predict three-dimensional (3D) protein structures [1,2] and to simulate protein folding [3–5] have been developed.

Actual design of amino acid sequences for proposed 3D structures is called "*de novo*" design [6–9]. For the α-bundle topology, *de novo* design seems to have succeeded [10–12] and several artificial α-bundle proteins have been created [6–9]. Several functions have been produced by adding the functional motif sequence on the structural frames of the artificial α-bundle proteins [11,13,14]. However, despite lots of efforts

* Corresponding author.

SSDI 0301-4622(94)00008-8

of such *de novo* design, new artificial proteins having other topologies especially including β-strands [15,16] have not been fully successful.

Among those *de novo* designs, β/α-barrel proteins, that are composed of a parallel β-barrel with eight β-strands accompanying with the surrounding eight α-helices, have been tried to be constructed because of the symmetrical fashion and stable characteristics of natural proteins.

X-ray crystallographic studies have so far revealed that there are lots of proteins which have the same typical topologies as the β/α-barrel. Since the only weak homologies of the amino acid sequence can be seen among those proteins, arguments are still left whether the topological similarity is the result of the divergence from a unique ancestor or of the convergence to the stable fold [17]. At least, the β/α-barrel topology must be stable, because so many different kinds of amino acid sequence can exist, taking the same β/α-barrel structure.

Either our previous design [18] or the design by Goraj et al. [19] was not completely successful to produce an artificial β/α-barrel protein as stable as the native ones. In the present study, the common geometrical features were extracted from four X-ray crystal structures of spinach glycolate oxidase (GO) [20], chicken breast muscle triose phosphate isomerase (TIM) [21], tryptophane synthase from *Salmonella thyphimurium* (TS) [22] and α-amylase from *Aspergillus oryze* (AMYL) [23]. Based upon these analyses, we tried further *de novo* design of a β/α barrel protein and produced it as a fused protein using our *Escherichia coli* (*E. coli*) overproduction system.

2. Methods

The flow of *de novo* design is shown in Fig. 1, following the concept to solve the inverse folding problem; first, the 3D backbone structure was constructed, and then an amino acid sequence was designed suitable for the structure. The designed protein was produced in the *E. coli* system, and the secondary and tertiary structures were analyzed.

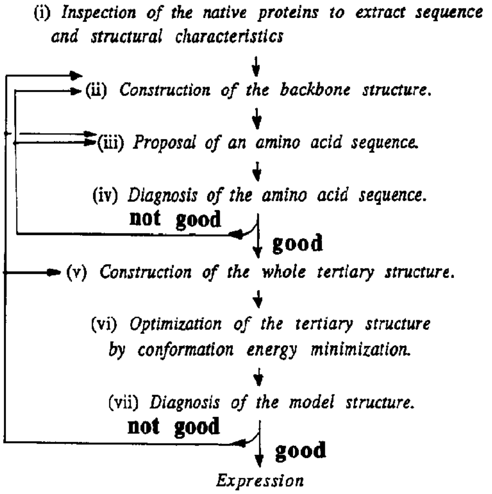
(i) Inspection of the native proteins to extract sequence and structural characteristics
(ii) Construction of the backbone structure.
(iii) Proposal of an amino acid sequence.
(iv) Diagnosis of the amino acid sequence.
not good
good
(v) Construction of the whole tertiary structure.
(vi) Optimization of the tertiary structure by conformation energy minimization.
(vii) Diagnosis of the model structure.
not good
good
Expression

Fig. 1. Flow of *de novo* protein design.

2.1. Analysis of 3D structures of native β / α-barrels

Atomic coordinates of the four typical β/α-proteins, GO (PDB code, 1GOX [20]), TIM (1TIM [21]), TS (1WSY [22]) and AMYL (2TAA [23]) were taken from PDB database. Their secondary structures were analyzed by the program DSSP developed by Kabsch and Sander [24].

At first, the central β-barrels were analyzed by fitting hyperboloid in the same way as that by Lasters et al. [25]. Several geometrical parameters of the β-strands and the surrounding α-helices were extracted based upon the uniquely defined equatorial plane and the perpendicular barrel axis. Next, the loop structures were analyzed following the classification by Thornton et al. [26].

2.2. Analysis of primary structures of native β / α-barrels

Homologous proteins to each of the above four β/α-proteins were collected from the amino acid sequence databases of NBRF and PIR. The multiple alignment program by Barton and Sternberg [27] was used to align those proteins in the corresponding superfamilies. The weight of each sequence in the individual superfamily was calculated using the Monte Carlo algorithm [28]. Finally, the statistical frequencies of the amino acids at each position in the β-barrel were calculated.

2.3. Modelling the designed protein

Following several geometrical parameters of the β/α-barrel, eight α-helix axes and eight β-strand axes were generated. Then, typically eight α-helices and eight β-strands structures were put on those axes, respectively. Typical loop structures were looked for by the inhouse program FRGMNT [29] using the conventional loop search method. Side chains were generated by the molecular graphics program INSIGHT (Biosym, Inc.) on the computer graphics screen (PS390, E & S), with preferential χ angles analyzed statistically by Ponder and Richards [30]. In order to get rid of bad contacts among atoms, the conformation energy was minimized by the conjugate gradient method using a molecular mechanics program PRESTO [31]. The calculation was carried out *in vacuo* with the AMBER-united atom force field [32] with the dielectric constant $2r_{ij}$. Here, r_{ij} is the distance between ith and jth atoms.

2.4. Diagnosis of the designed structure

The secondary structure prediction was performed by several different methods; sequence homology method [33], Ptitsyn–Finkelstein method [34], Gibrat–Garnier–Robson method [35], neural network method [36], Nagano method [37] and Lim method [38]. 3D–1D compatibility scores were calculated by the method developed by Nishikawa and Matsuo [39,40]. The polarity of the designed 3D structure was evaluated using the program "Poldiagnostics" by Baumann et al. [41]. The packing in the designed structure was calculated by the program QPACK [42]. The unfolding free energy was estimated by the method of Ooi and Oobatake [43], comparing the accessible surface area of each atom between the folded and extended structures.

2.5. Synthesis, purification and refolding of the designed protein

The associated gene with the designed amino acid sequence was produced by successive ligation of oligodeoxyribonucleotides, synthesized with an Applied Biosystems Automatic Synthesizer model 380A by the phosphoramidite method. *E. coli* HB101 transformed by the plasmid harboring the synthetic gene was cultured at 37°C in M9 medium containing two fold strength of the salts and 0.5% casamino acids. Protein synthesis was induced by adding 20 μg/ml 3-β-indoleacrylic acid at the Klett unit between 120 to 200. After 20 h of the induction, cells were harvested and suspended in lysis buffer (50 mM TrisHCl pH 7.5, 100 mM NaCl and 1 mM EDTA). After adding lysozyme and stand at 0°C for 30 min, the mixture was sonicated four times for 20 s. The insoluble material was collected by centrifugation and dissolved in 6 M guanidine hydrochloride (GdnHCl) containing 1% 2-mercaptoethanol. After centrifugation at 4°C for 1 h at 5000g, the supernatant was dialized against H_2O. The precipitate was collected and treated with BrCN in 70% HCOOH for 3 h. The mixture was applied to a column of Aquapore C-8 (Applied, 30 nm pore size, 1 cm × 25 cm) with a linear gradient of CH_3CN containing 0.1% TFA at a flow rate of 3.5 ml per min. The fractions were analyzed by SDS-PAGE. The peak fractions containing the desired protein were pooled and lyophilized. The protein was dissolved in 8 M GdnHCl with 10 mM sodium acetate (pH 4), and dialized stepwise against 6 M, 4 M, 2 M and 0 M GdnHCl at pH 4.

2.6. Characterization of the designed protein

The apparent molecular weight of the folded protein was observed from size exclusion chromatography using Sephadex G-75SF gel filtration column (Pharmacia) in 10 mM sodium phosphate (pH 4.0). Sedimentation equilibrium runs were carried out in an analytical ultracentrifuge (Beckman, Spinco Model E) equipped with Schlieren and interference optical system. Double-sector cells with a path-length of 12 mm were used at 14290 rpm in an AnG rotor. The apparent molecular weight (M_{app}) was estimated from the slope of the natural logarithm of the solute concentration against the square of the radial distance [44].

The circular dichroism (CD) spectra were measured using a Jasco J-600 spectropolarimeter. The fluorescence emission spectra excited at 280 nm were measured with a Hitachi F-4000 fluorescence spectrophotometer. The denaturation procedure by GdnHCl was observed by measuring both the CD spectra at 220 nm and the fluorescence emission spectra at 325 nm for the solution with 0 to 6 M GdnHCl concentration. ^{1}H-nuclear magnetic resonance (NMR) spectra were measured by 600 MHz spectrometer AM-600 (Bruker). All the measurements were carried out at 20°C.

3. Results

3.1. Common geometry of native β / α-barrels

The individual β-strands were approximated by straight lines, which were then fitted with a hyperboloid, following the procedure described above. The results are shown in Table 1, essentially the same as the previous analyses [25,45]. Taking the z axis of the hyperboloid as the axis of the whole β/α-barrel, the axis lines of the individual α-helices were analyzed. The results are shown in Table 2. The average distance between the cross section of the helix axis and the center of the barrel axis is 17.8 Å, which is 10.6 Å longer than the average distance between the axis of the β-strand and the barrel center. It should be noticed that the structural similarities of the β-barrels among the four proteins are much better than those of the α-helices. The tilt angles of the α-helices deviate largely around the average tilt angle, which is larger than the average tilt angles of the β-strands. The length of α-helices are also different from each other. Moreover, in TIM and GO, the angles nxy$_\alpha$ are almost perpendicular,

Table 2
Geometrical parameters of α-helices around the parallel β-barrel

Protein	$\langle R_\alpha \rangle$ [a]	$\langle Tz_\alpha \rangle$ [b]	$\langle nxy_\alpha \rangle$ [c]
	(Å)	(deg)	(deg)
GO	17.6 (1.2) [d]	47.0 (9.8)	85.6 (17.0)
TIM	16.9 (1.1)	45.2 (12.2)	89.2 (9.9)
TS	18.1 (1.0)	46.6 (8.3)	64.9 (13.0)
AMYL	18.6 (2.4)	49.9 (11.3)	50.4 (12.4)
average [e]	17.8	47.2	72.5

[a] $\langle R_\alpha \rangle$ is the average distance between the cross section point of each α-helix axis and the barrel axis on the equator plane.
[b] $\langle Tz_\alpha \rangle$ is the average tilt angle of the α-helix axis against the barrel axis.
[c] $\langle nxy_\alpha \rangle$ is the average tangent angle of the α-helix axis projected on the equator plane at the ellipse.
[d] Values in parentheses are the corresponding root-mean-square deviations around the average values for the eight helices.
[e] The average for the four proteins.

Table 1
Geometrical parameters of parallel β-barrels

Protein	a [a]	b [a]	$\sqrt{ab}$	$\langle R_\beta \rangle$ [b]	$\langle Tz_\beta \rangle$ [c]	$\langle nxy\beta \rangle$ [d]
	(Å)	(Å)	(Å)	(Å)	(deg)	(deg)
GO	7.4	7.1	7.3	7.2 (0.2) [e]	38.5 (4.6)	89.9 (5.9)
TIM	8.3	5.9	7.0	7.1 (1.0)	37.9 (4.4)	91.2 (21.6)
TS	8.4	6.5	7.4	7.3 (0.8)	35.3 (6.2)	90.4 (16.3)
AMYL	7.6	7.0	7.3	7.2 (0.4)	35.2 (8.1)	91.1 (8.4)
average [f]			7.2	7.3	36.7	90.6

[a] a and b are the longer and shorter axis of the ellipse on the equator plane (XY plane).
[b] $\langle R_\beta \rangle$ is the average distance between the cross section point of each β-strand axis and the barrel axis on the equator plane.
[c] $\langle Tz_\beta \rangle$ is the average tilt angle of the β-strand axis against the barrel axis (Z axis).
[d] $\langle nxy_\beta \rangle$ is the average tangent angle of the β-strand axis projected on the equator plane at the ellipse.
[e] Values in parentheses are the corresponding root-mean-square deviations around the average values for the eight strands.
[f] The average for the four proteins.

but they are much smaller than 90° in TS and AMYL. It means that in TIM and GO, the diameter of the β/α-barrel does not change much along the barrel axis. On the contrary, in TS and AMYL, the N-termini of the α-helices of the β/α-barrels tend to expand.

One of the common characteristics of the β/α-barrel proteins is that most of them are enzymes and their active sites are located on the loops near the N-termini of the α-helices, although their enzymatic functions are very different each other. The size and sequence of the amino acids of this loop region from β to α are completely different among the β/α-barrel proteins. Usually the loops from β to α (loops(βα)) are very long.

On the contrary, every loop from α to β (loop(αβ)) is short. In Table 3, the loops(αβ) in the four β/α-barrel proteins are classified following Thornton et al. [26]. It should be noticed that many of the loops from odd number α to even number β are "αβ3" types [46], which is composed of the initial Gly residue having a positive ϕ angle and the successive two residues having β- and α-conformations, respectively. On the contrary, the loops from even number α to odd number β, several "αβ1" types are observed instead of "αβ3" types. The "αβ1" type is characterized by a single Gly residue, that has a left-handed helical conformation, giving a sharp chain reversal.

3.2. Backbone design of a β / α-barrel

Based upon the above geometrical features, an ideal backbone structure of the β/α-barrel was constructed. The structural unit in this study was $-\beta_1$–loop(βα)–α_1–loop(αβ)–β_2–loop(βα)–α_2–loop(αβ)–. Previously, Goraj et al. [19] designed the eight periodic structural units of –β–loop(βα)–α–loop(αβ)–. However, since the β-strands tilt with respect to the barrel axis, the side chains of the adjacent β-strands at the same height point to different directions along the β-strand axis [45]. It means that any eight equal structural units cannot produce a tilted parallel β-barrel.

The lengths of both the major and the minor semiaxes of the ellipse of the parallel β-barrel was determined as the average values of 7.3 Å from Table 1. The distance between the cross section of each helix axis and the barrel center was determined as 17.6 Å, taking from the value of GO. Because the cross section of the hyperboloid of GO at the equator is the most circular and those of other three proteins largely deviate from a circle. The tilt angles of the axis of the β-strands, Tz_β and nxy_β, were determined as the

Table 3
Classification of short loops from α-helices to β-strands

Loop	Protein			
	GO	TIM	TS	AMYL
$\alpha_1 \rightarrow \beta_2$	–ĠTI [a]– (αβ1)[b]	–AKLSADT–	–ĠAD– (αβ3) [b]	–ĠFT– (αβ3)
$\alpha_3 \rightarrow \beta_4$	–ĠFK– (αβ3)	–ĠAA– (αβ3)	–ĠVD– (αβ3)	–SNYSID–
$\alpha_5 \rightarrow \beta_6$	–ĠAA– (αβ3)	–VKDWSK–	–GRGY–	–NVMD–
$\alpha 7 \rightarrow \beta_8$	–ĠAA– (αβ3)	–SQHDVD–	–ĠAA– (αβ3)	–NDG–
$\alpha_2 \rightarrow \beta_3$	–GPG–	–LDAK–	–HPTIP– (β-I) [c]	–GM– (αβ1)
$\alpha_4 \rightarrow \beta_5$	–TSLP–	–EĠLG– (αβ1)	–HṄI– (αβ1)	–AGVY–
$\alpha_6 \rightarrow \beta_7$	–QGRIP–	(α-helix) [d]	–YḢAAP– (αβ1)	–CPDSTL–

[a] Amino acid sequence is indicated by single-letter codes. A dot on a letter indicates that the backbone structure of the residue is the left-handed helical structure.

[b] αβ1 and αβ3 are typical loop structures, which were classified by Thornton et al. [46].

[c] Typical type I β-turn [46].

[d] The loop is very long and takes a helical structure.

average values in Table 1, 36° and 90°, respectively. Those of the α-helices, Tz_α and nxy_α were determined from the typical values of GO, 47° and 86°, respectively.

Ideal secondary structures of the β-strands and α-helices were built by using the typical (ϕ, ψ) values with a trans-peptide plane. An N-cap structure [47] was introduced for every α-helix, referring the typical structure found in carboxypeptidase A (PDB code; 5CPA [48]) at residues from 93 to 96 (–Asn–Pro–Ser–Phe–).

The backbone structures of the loops were built, borrowing the loop structures in natural proteins. For the short loops(αβ), the typical loop structures of "αβ1" type and "αβ3" type were taken from the loop between the fourth α-helix and the fifth β-strand of TIM [20] and the loop between the third α-helix and the fourth β-strand of GO [21], respectively. For the loops(βα), we made a model by looking for loop structures by the conventional loop search method. For the loop from odd number β to even number α, a four-residues loop structure was taken from Rhinovirus 14 coat protein (PDB code; 1R08 [49]) at the residues from 88 to 90 (–Thr–Gly–Ile–Asp–). For the loop from even number β to odd number α, another five-residues fragment was taken from lactate dehydrogenase (6LDH [50]) at the residues from 100 to 104 (–Gln–Gln–Glu–Gly–Glu–).

Joining the above secondary structures and loops, the three-dimensional model of the whole backbone structure was constructed.

3.3. *Analysis of amino acid sequences of native β / α-barrels*

The next step is to put amino acids on the individual positions. A few years ago, Richardson and Richardson have found typical tendency of the amino acid frequency depending upon the position in the α-helix [47]. In the present β/α-barrel, similar analysis was carried out for the four proteins. Since the barrel structure has been analyzed, it is easy to decide a residue, which is the nearest of the equator in every β-strand. This position was defined as the center, and the residues in every strand were re-numbered referring to the center position. Since the side-chain points either towards the center of the β/α-barrel or outwards, the residues on the β-strands were further classified by these side-chain directions.

The numbers of the amino acids from the X-ray crystal structures are not enough for a statistical analysis, and we added more proteins. More than 25% homologous proteins to the four β/α-barrel proteins were collected from the amino acids sequence data base; a total of 14 homologous proteins to TIM, 17 to TS, 4 to AMYL and 2 to GO. These proteins in each superfamily were aligned simultaneously [27]. Their structural homologies are expected because the regions of the secondary structures can be well aligned.

Using the Monte Carlo method [28], the weights of the each sequence in the individual superfamily were calculated, and the amino acid frequencies were summarized. Assuming the four superfamilies are independent, we averaged the frequency for the four superfamilies. When we summarized the inner five successive residues composing the inner barrel, very characteristic features appear as shown in Fig. 2. Here, the amino acid frequencies are normalized by the average and standard deviations for 1604 representative proteins [51]. In addition to Val and Ile,

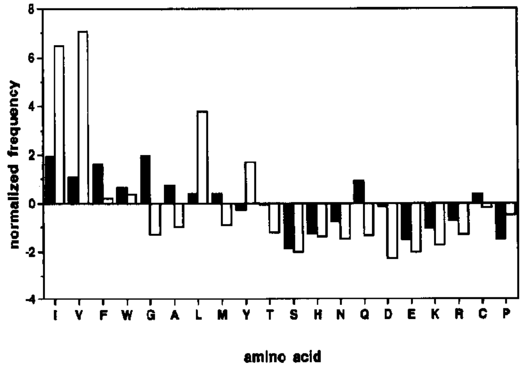

Fig. 2. Frequencies of amino acid constructing the inner β-barrel in 37 β / α-barrel proteins, normalized by the average and standard deviations for 1604 representative proteins [51]. Filled bars are the amino acids towards the barrel center, and the white bars are those interacting with the outer helices. The amino acids are indicated by single-letter codes.

which prefer a β-structure, Gly, Ala, Phe and Trp side chains are frequently used towards the barrel center (filled bars in Fig. 2). On the contrary, very few Gly residues exist in the inverse direction (open bars in Fig. 2). Only Ile, Val and Leu are remarkable in this region.

The distribution of the amino acids in the α-helices was also analyzed for the four crystal structures of GO, TIM, TS and AMYL. The amino acid frequencies at individual positions on helix show that the helices have typical amphiphilic nature with the usual position tendency of each amino acid [47].

3.4. Sequence design of the β / α-barrel

Although the backbone structure has a four-fold symmetry, $4 \times (\beta_1\alpha_1\beta_2\alpha_2)$, the amino acid sequence cannot have the same symmetry. Because the same amino acids would gather at the same position on the four β_1 (or β_2)-strands towards the barrel center in the four-fold symmetry. When Gly or aromatic side chains were designed to put on there, such residues would gather at the position, generating either a hole or a clash in the barrel center. In native β/α-barrels, Gly and aromatic residues are located in a complementary manner. Thus, instead of the four fold symmetry $4 \times (\beta_1\alpha_1\beta_2\alpha_2)$, the structure having the two fold symmetry $2 \times (\beta_1\alpha_1\beta_2\alpha_2\beta_3\alpha_3\beta_4\alpha_4)$ was taken in the current design, where β_1, β_2, β_3 and β_4 are different β-strands expecting complementary packing among the side chains at the barrel center.

Fig. 3a shows the designed amino acid sequence composed of 201 amino acids (Mw = 22, 150). The inside of the β-barrel is composed of Phe and Trp in β_1-strand, Val and Gly in β_2-strand, Ile and Ala in β_3-strand, and Ile and Tyr in β_4-strand, considering the complementary packing as mentioned above. The side chains of the β-strands outwards from the barrel center were designed to be all hydrophobic interacting with α-helices. Each helix has a typical amphiphilic nature with a few putative salt-bridges along the helix. At every N-terminus, the N-cap sequence was added. Loop($\alpha_1\beta_2$) and loop($\alpha_3\beta_4$) have the same "αβ3" type sequence -GAD-, and loop($\alpha_2\beta_3$) and loop($\alpha_4\beta_1$) have the same "αβ1" type sequence -Gly-Leu-Pro-. Loop($\beta_1\alpha_1$) and loop($\beta_3\alpha_3$) have the similar backbone structure, characterized by a single Gly having a left-handed helical conformation and the following hydrophobic residue. The sequence of loop ($\beta_1\alpha_1$) was decided as -His-Gly-Leu-Asp- and loop($\beta_3\alpha_3$) was -Glu-Gly-Ala-Asp-. The last Asp residues can electrostatically compensate the helix dipoles of the following α_1 and α_3 helices. Loop($\beta_2\alpha_2$) and loop($\beta_4\alpha_4$) have the same sequence -Thr-Gln-Pro-Gly-Leu-, which was modified from the original sequence in lactate dehydrogenase from 100 to 104, where Gly103 is in the left-handed helical structure. The N-terminal four residues Arg-Ala-Gly-Tyr- and the C-terminal sequence -Gly-Ala-Gly-Arg were finally added as flexible termini, considering the charge balance after the whole sequence was determined.

Total positive charges of Lys and Arg residues are 30, and the negative ones of Asp and Glu are 26. The whole charge distribution on the model structure is quadrupolar type, in which the top and bottom of the barrel are both negative and the middle of the surface helices surrounding the β-barrel is positive.

3.5. Model building and diagnosis of the designed β / α-barrel

After the side-chains were generated on the computer graphics screen, all the bad contacts between atoms were repaired by minimizing the conformation energy of the model structure. The final three dimensional model is shown in Fig. 3b.

In the last step of the design, the designed sequence was inspected. At first, several methods of the secondary structure prediction were applied to the designed sequence. Table 4 indicates that α- and β-structures were well predicted by almost all the methods. The 3D-1D alignment method developed recently [1,2] can inspect whether the designed amino acid sequence prefer the β/α-barrel structure or not. Using the 3D-1D compatibility function developed by Nishikawa and Matsuo [39,40], the designed sequence was threaded on 128 natural protein structures in lots

(a)

ArgAlaGlyTyr CGTGCTGGTTAC			
VaLPheValTrpLeu GTTTTCGTTTGGCTG	10 HisGlyLeuAsp CACGGTCTAGAT	20 AsnProGluLysLeuLeuLysAlaPheGluLysAsn AACCCGGAAAAGCTGCTGAAAGCGTTTGAAAAGAAC	GlyAlaAsp GGCGCTGAT
30 GlnValValPheGly CAGGTCGTGTTCGGT	ThrGlnProGlyLeu ACCCAACCGGGTCTG	40 50 AsnProGluLysLeuLeuLysAlaLeuGluLysLys AACCCGGAGAAACTGCTGAAGGCGCTCGAGAAGAAA	GlyLeuPro GGTCTGCCG
ValIleLeuAlaIle GTTATCCTGGCAATT	60 GluGlyAlaAsp GAGGGTGCTGAT	70 AsnProGluLysLeuLeuLysAlaPheGluLysAsn AACCCAGAGAAACTGCTTAAGGCGTTCGAAAAGAAC	GlyAlaAsp GGTGCTGAC
80 TyrIleIlePheTyr TACATCATTTTCTAC	ThrGlnProGlyLeu ACGCAACCGGGTCTG	90 AsnProGluLysLeuLeuLysAlaLeuGluLysLys AATCCGGAGAAGCTGCTGAAAGCTTTAGAGAAGAAA	100 GlyLeuPro GGTCTGCCG
ValPheValTrpLeu GTTTTCGTTTGGCTG	110 HisGlyLeuAsp CACGGTCTAGAT	120 AsnProGluLysLeuLeuLysAlaPheGluLysAsn AACCCGGAAAAGCTGCTGAAAGCGTTTGAAAAGAAC	GlyAlaAsp GGCGCTGAT
130 GlnValValPheGly CAGGTCGTGTTCGGT	ThrGlnProGlyLeu ACCCAACCGGGTCTG	140 AsnProGluLysLeuLeuLysAlaLeuGluLysLys AACCCGGAGAAACTGCTGAAGGCGCTCGAGAAGAAA	150 GlyLeuPro GGTCTGCCG
ValIleLeuAlaIle GTTATCCTGGCAATT	160 GluGlyAlaAsp GAGGGTGCTGAT	170 AsnProGluLysLeuLeuLysAlaPheGluLysAsn AACCCAGAGAAACTGCTTAAGGCGTTCGAAAAGAAC	GlyAlaAsp GGTGCTGAC
180 TyrIleIlePheTyr TACATCATTTTCTAC	ThrGlnProGlyLeu ACGCAACCGGGTCTG	190 AsnProGluLysLeuLeuLysAlaLeuGluLysLys AATCCGGAGAAGCTGCTGAAAGCTCTGGAGAAGAAA	200 GlyAlaGlyArg GGTGCGGGTCGT
└ β-strand ┘	└─loop(βα)─┘	└──────── α-helix ────────┘	└loop(αβ)┘

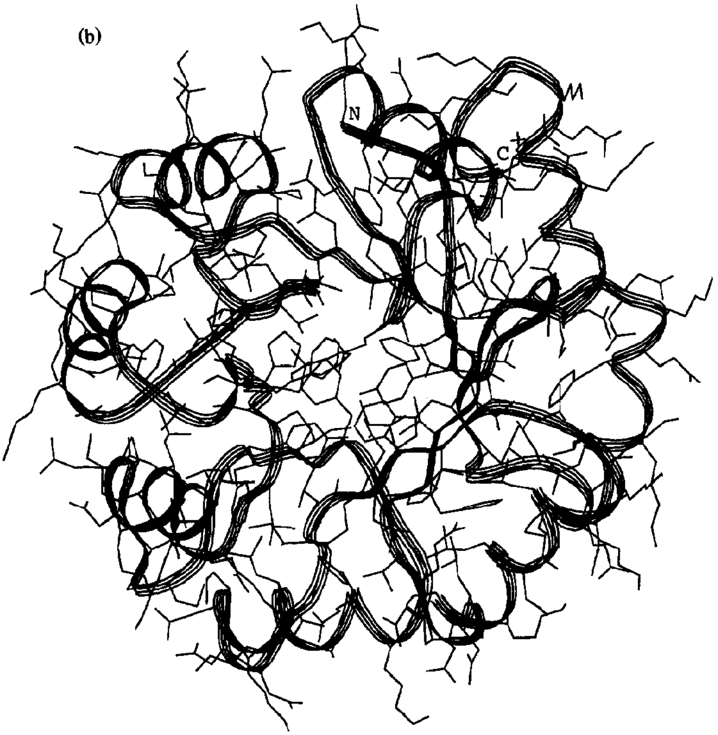

Fig. 3. (a) The amino acid and nucleotide sequence of the designed protein. (b) The tertiary model structure of the designed protein. The backbone is indicated by a ribbon, and the side-chains are by sticks. N and C indicate the N- and C-terminus, respectively.

of different superfamilies. The 3D–1D compatibility scores are shown in Table 5. Here, the scores are normalized, so that the score zero corresponds to the average of a 100 random sequence having the same sequence lengths, and that the score 1 corresponds to the standard

Table 4
Secondary structure prediction by several different methods

	`1       10        20        30        40        50`
	`    .    .    .    .    .    .    .    .    .    .`
sequence [a]	`RAGYVFVWLHGLDNPEKLLKAFEKNGADQVVFGTQPGLNPEKLLKALEKK`
DSSP [b]	`  S EEEE SSSS HHHHHHHHHHHT SEEEE  BTTB HHHHHHHHHHT`
homology [c]	`   eeeeeee    hhhhhhhh     eeeeee      hhhhhhhhhh`
PF [c]	`   eeeeee    hhhhhhhh       eee       hhhhhhhhhhh`
GGR [c]	` h eeeeeh    hhhhhhhhhh    heee        hhhhhhhhhh`
neural [c]	`   eeeeee     hhhhhhh      eee        hhhhhhhhh`
Nagano [c]	`   ehhhhh     hhhhhhhhh     eeee       hhhhhhhhh`
Lim [c]	`   eeeeeee    hhhhhhhhh    eeee        hhhhhhhhhh`
	`        60        70        80        90       100`
	`    .    .    .    .    .    .    .    .    .    .`
sequence	`GLPVILAIEGADNPEKLLKAFEKNGADYIIFYTQPGLNPEKLLKALEKKG`
DSSP	`T  EEE  TT SSHHHHHHHHHHHT  EEEE   TT  HHHHHHHHHHTT`
homology	`  eeeeee     hhhhhhhh    eeeeee       hhhhhhhhhh`
PF	` eeeeeee    hhhhhhhh      eeeeee    hhhhhhhhhhh`
GGR	`  hhhhhhh   hhhhhhhhhh   heeeeee      hhhhhhhhhh`
neural	`   eeeee     hhhhhhhh     eeeee      hhhhhhhhh`
Nagano	`   hhhhhh    hhhhhhhhh    eeeee       hhhhhhhhh`
Lim	`   eeeee     hhhhhhhh     eeeeeee     hhhhhhhhhh`
	`       110       120       130       140        150`
	`    .    .    .    .    .    .    .    .    .    .`
sequence	`LPVFVWLHGLDNPEKLLKAFEKNGADQVVFGTQPGLNPEKLLKALEKKGL`
DSSP	`  EEEE  S S HHHHHHHHHHHT SEEEE  BTTB HHHHHHHHHHTT`
homology	`  eeeeee    hhhhhhhh     eeeeee       hhhhhhhhhh`
PF	`eeeeeee    hhhhhhhh       eee        hhhhhhhhhhh e`
GGR	` hhhehh    hhhhhhhhhh    heee         hhhhhhhhhh`
neural	`  hhhee     hhhhhhh      eee         hhhhhhhhh`
Nagano	` ehhhhh     hhhhhhhhh     eeee        hhhhhhhhh`
Lim	`  eeeeee    hhhhhhhhh    eeee         hhhhhhhhhh`
	`      160        170        180        190        200`
	`    .    .    .    .    .    .    .    .    .    .`
sequence	`PVILAIEGADNPEKLLKAFEKNGADYIIFYTQPGLNPEKLLKALEKKGAGR`
DSSP	` EEEEETT SSHHHHHHHHHHHT SEEEE  BTTB HHHHHHHHHHHTS`
homology	`eeeeee      hhhhhhhh    eeeeee       hhhhhhhhhh`
PF	`eeeeee     hhhhhhhh      eeeeee     hhhhhhhhhhh`
GGR	`hhhhhh     hhhhhhhhhh   heeeeee      hhhhhhhhhh`
neural	` eeeee      hhhhhhhh     eeeee      hhhhhhhhh`
Nagano	` hhhhhh     hhhhhhhhh    eeeee       hhhhhhhhh`
Lim	` eeeee      hhhhhhhhh    eeeeeee     hhhhhhhhhh`

[a] Amino acid sequence is indicated by single-letter codes.

[b] DSSP indicates conformations/secondary structures of the designed tertiary model analyzed by the program DSSP [24]. (H, α-helix; G, 3-to-10-helix; E, β-strand; B, β-bridge; T, 3-, 4- or 5-turn; S, bend).

[c] "Homology", "PF", "GGR", "neural", "Nagano" and "Lim" lines are the results of the secondary structure predictions by the sequence homology method [33], the Ptitsyn-Finkelstein method [34], the Gibrat–Garnier–Robson method [35], the neural network method [36], the Nagano method [37], and the Lim method [38], respectively. e and h indicate the extended and helical structures predicted, respectively.

deviation of the random sequence. The lower scores indicate the better 3D-1D compatibility. It is evident that the designed structure has the significantly lowest score, and that all the proteins with low compatibility scores are the α/β type, including several proteins taking the folds of the β/α-barrel.

In addition, the polarity of the designed protein was investigated [41], and the result is shown in Table 6. It is indicated that the model structure has similar features of the polarity to those of natural proteins. The program QPACK to measure the packing [42] shows that the pair potential sum was −44.0, the number of outsiders was 58, the mean square sphere size was 106.5%, and the standard deviation of the sphere size was 20.8%, for the model structure. These values are very similar to those of natural proteins except the standard deviation, which is about 5–10% larger than natural proteins. The melting temperature and the maximum unfolding free energy were estimated as 51.6°C and 6.5 kcal/mol at 20°C, respectively, by the method of Ooi and Oobatake [43].

Table 5
Compatibility values of the designed sequence, mounted on the designed model structure and natural protein structures [39,40]

Structure [a] (PDB code)	N [b]	Compatibility values	Sequence identity (%)
designed	201	−11.414	100
2LBP	193	−9.019	9.0
5P21	150	−8.622	7.8
2GBP	201	−8.198	11.0
1PFK	194	−8.033	15.9
8ATC	191	−7.810	10.5
8ABP	201	−7.754	10.0
1GKY	185	−7.647	15.0
2AAT	194	−7.530	13.1
1GOX	190	−7.526	9.1
3ENL	200	−7.235	10.9
6XIA	184	−6.895	12.5
5TIM	201	−6.809	8.9
4PTP	181	−6.670	6.2
1ALD	189	−6.443	9.4
3ADK	184	−6.396	14.0

[a] The protein structures, for which N was greater than 150, are listed, sorted by the negatively large compatibility scores. Proteins with italic names are β/α-barrels.
[b] Number of mounted residues.

Table 6
Diagnosis of the polarity of the designed structure [41]

Item	Designed protein	150 natural proteins [a]		
		lowest	mean	highest
max. pol. [b]	0.163	0.163	0.173	0.185
srf. ext. [c]	0.342	0.320	0.400	0.520
int. pol. [d]	0.161	0.158	0.172	0.190
ext. pol. [e]	0.169	0.147	0.172	0.204
sc. int. pol. [f]	0.092	0.074	0.098	0.118
sc. ext. pol. [g]	0.141	0.115	0.143	0.183

[a] The diagnosis is to measure the polarity values indicated by each item, whether they are within the lowest and highest values for 150 typical natural proteins. Precise descriptions of the items below are in ref. [41].
[b] Maximum polar fraction (The unit is e).
[c] External surface area per molecular weight (Å^2/dalton).
[d] Internal polar fraction (e).
[e] External polar fraction (e).
[f] Internal polar fraction for side-chains (e).
[g] External polar fraction for side-chains (e).

These diagnoses suggest that the model structure has similar features to stable natural proteins, and that the designed amino acid sequence is promising for the structure.

3.6. Synthesis and purification of the designed protein

The nucleotide sequence encoding the designed β/α-barrel protein was deduced using the preferred codons of *E. coli*, and it is shown in Fig. 3a under the corresponding amino acid sequence. The synthetic gene was inserted downstream of Bgl II site of hGH gene with the linker sequence composed of Ile–Glu–Asn–Ser–Asp–Pro–Ser–Met. Over-expressed fusion protein was treated with BrCN, so as to be cleaved at the last Met residue in the linker. The cleaved protein was purified by the reversed phase high-performance liquid chromatography (HPLC). The purified protein had the expected amino acid sequence over 10 residues by Edman degradation. The amino acid composition was also consistent with the expected sequence.

3.7. Characterization of the designed protein

The result of size-exclusion chromatography indicates that the apparent molecular weight of the protein is about 30000, when it was refolded at pH 4. Although this value is 1.4 times higher than the calculated one, it shows that the protein is monomeric. The monomeric protein was dimerized over pH 5.

Fig. 4 shows the results of the sedimentation equilibrium in 50 mM acetate buffer (pH 4.2). The lines are almost linear, and M_{app} was estimated to be 24640 ± 810 from the slopes, independent of the concentration. The protein is confirmed to be monomeric in this condition, too.

CD spectra of the protein refolded at pH 4 was shown in Fig. 5. From the spectrum, the protein was estimated to contain 30% α-helix and 40% β-strand using the algorithm of Provencher and Glockner [52]. As seen in Fig. 5, the CD spectrum is similar to that of the rabbit TIM.

The stability of the protein against urea or GdnHCl was measured by monitoring the mean residue ellipticity at 220 nm. Even at 7 M urea, the protein was found to be partially denatured since the negative band around 220 nm in the CD spectrum was still observed. When the protein was denatured by adding GdnHCl, a two-states transition was observed as shown in Fig. 6. The midpoint of the unfolding transition was 3.9 M

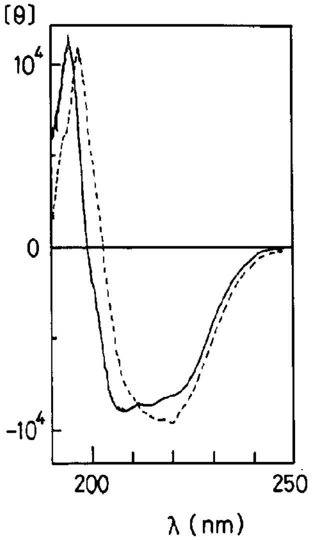

Fig. 5. CD spectra of the designed protein (———) in 10 mM acetate buffer (pH 4.0) and of rabbit triose phosphate isomerase (— — —) in 10 mM phosphate buffer (pH 7.0) at 20°C.

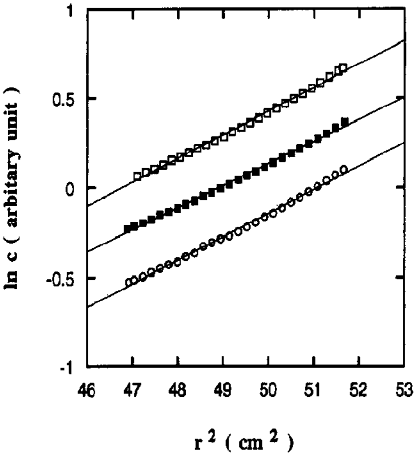

Fig. 4. Results of sedimentation equilibrium at 20°C in 50 mM acetate buffer (pH 4.2) for solutions with three different initial concentrations; (○) 0.5 mg/ml; (■) 1 mg/ml; (□) 2 mg/ml. The solid lines indicate the fitted linear relations between the square of the distance (r^2) and the natural logarithm of the solute concentration (c).

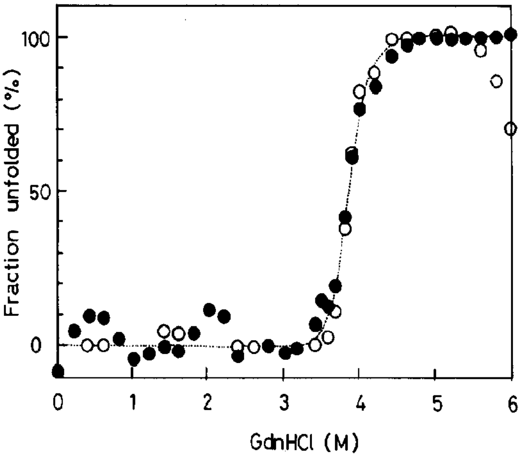

Fig. 6. Unfolding by GdnHCl in 10 mM acetate buffer (pH 4.0) at 20°C. Open circles are far-CD values observed at 220 nm and filled circles are fluorescence emission at 325 nm excited by 280 nm. Those two values were normalized as the fraction unfolded, respectively.

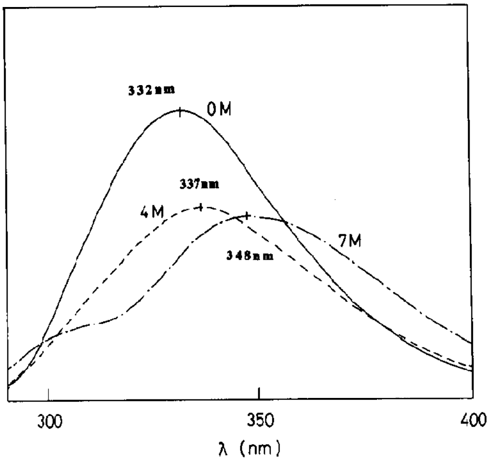

Fig. 7. Fluorescence emission spectra excited at 280 nm, with 0 M (———), 4 M (— — —) and 7 M (· — · —) GdnHCl, in 10 mM acetate buffer (pH 4.0) at 20°C. The wave length at each peak is also indicated.

GdnHCl indicating that the protein is significantly stable.

The protein is designed to have two Trp residues on the β-strands, which should be inside the protein. Without GdnHCl, fluorescence maximum excited at 280 nm due to the two Trp residues was 332 nm. With 7 M GdnHCl, it was shifted to longer wavelength and the intensity was decreased, as shown in Fig. 7. It suggests that the Trp residues are buried in the folded structure of the protein as designed. In Fig. 6, the denaturation curve measured by fluorescence intensity at 325 nm was overlapped with that obtained by CD spectra. This indicates that the secondary and tertiary structure were decomposed simultaneously.

^{1}H-NMR spectrum of the protein solution (1 mg/ml), which was the same as that observed in the sedimentation equilibrium experiment, was measured at 20°C in 50 mM deuterated acetate buffer (pH 4.2) with 100% D_2O. As a result, each resonance peak was too broad for native monomer proteins having the similar molecular weight. Here, the protein was confirmed to be monomeric in this solution from the sedimentation equilibrium study. Further analysis to extract structural information was difficult.

4. Discussion

Summarizing the results, the current protein designed *de novo* is experimentally confirmed to be compact, and to have pronounced secondary structures with a hydrophobic core. It is monomeric around pH 4, meaning good solubility to water. It is very stable against GdnHCl, and when it is denatured by adding GdnHCl, the secondary structures and hydrophobic core are deformed simultaneously.

In spite of these characteristics like a natural protein, the observed broad peaks in ^{1}H-NMR spectra suggest that the designed protein lacks a unique tight packing or it has polymorphic structures. That is, the designed protein has the similar features to natural proteins in the molten globule state, an intermediate state between the folded and unfolded state [53]. 1-anilino-naphthalene-8-sulphonate (ANS), which is a hydrophobic fluorescence dye, is often used as a probe to monitor the existence of the molten globule [54]. ANS rarely binds natural proteins in both the native and unfolded state, but binds significantly those in the molten globule state. Here, we observed that ANS can bind the designed protein even without GdnHCl, as shown in Fig. 8.

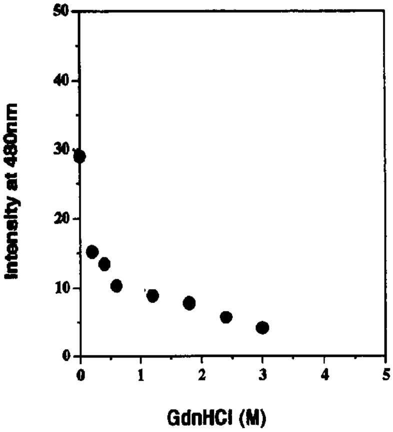

Fig. 8. Binding of 1-anilino-naphthalene-8-sulphonate (ANS) to the designed protein with GdnHCl was monitored by fluorescence at 480 nm, excited at 400 nm, in 10 mM acetate buffer (pH 4.0) at 20°C. The concentrations of the designed protein and ANS were 4.3 μM and 2.1 mM, respectively.

Therefore, it can be concluded that a definite organization of side-chains is absent in the designed protein. Rather, loose packing makes it possible for ANS to bind the designed protein. It is mainly due to a lack of knowledge on the construction of tight packing of the hydrophobic core in natural proteins. In fact, one of the measures of packing diagnosis mentioned above, the standard deviation of sphere size, was significantly larger than natural proteins, indicating that the current model structure does not have as tight a packing as natural proteins.

One of the answers to the question, how such tight packing is formed in natural proteins, has been proposed by Finkelstein and Nakamura [55]. They examined a way of close packing antiparallel β-sandwiches, and show that there are inherent "weak points", which cannot be filled by the surrounding aliphatic side-chains, but are mostly filled either by aromatic side-chains of the same sheet or by the residues of the other sheet of the sandwich. They also show that the defects, sometimes found in the interior of proteins, tend to occupy the positions of the intrinsic "weak points" [55]. Such a defect, a vacant space located in the hydrophobic core of a native protein, certainly destabilizes the protein structure, revealed by a protein engineering study [56]. A distinct cavity has been found in the hydrophobic core of a native protein, *E. coli* ribonuclease HI, and the thermal stability was enhanced remarkably when Val74 was replaced with Leu filling the cavity [56].

Therefore, one way to realize a tight packing of side chains is to analyze the "weak points" specific to the individual folds of natural proteins. In case of β/α-barrel proteins, the largest intrinsic "weak point" is located at the barrel center. The size of the barrel is strictly determined by the hydrogen bonds among the backbones constituent of the barrel, as shown in Table 1. The volume of the inner barrel is too large for any aliphatic residues to be filled. Therefore, several aromatic residues of Trp and Phe should be located there. Once such aromatic side-chains take places, the spatially adjacent residue has to be Gly or Ala with very small side-chains, due to steric clash with the aromatic side-chains. This explains very well the tendency of the amino acid distribution in the β/α-barrels, indicated in Fig. 2.

Interestingly, the designed protein is very stable against heat. Even at 90°C, CD spectrum in the far-UV region did not change with the same amount of secondary structures. This phenomenon has never been observed for natural proteins in the molten globule states. The entropy of the designed protein must be as large as that of a random-coil. In fact, our another designed β/α-barrel protein, recently re-designed in order to compensate the intrinsic "weak points" and to form more tight packing than before, shows a calorimetric thermal denaturation with much smaller changes in enthalpy and entropy than those of natural proteins (Tanaka et al., manuscript in preparation). Therefore, the designed protein in the current study is thermostable, probably because of the hydrophobic interaction with as large entropy as that of a random-coil.

5. Conclusion

Structural features of native β/α-barrel proteins were analyzed, and according to the typical characteristics, a β/α-barrel protein was designed *de novo*. On the ideal backbone structure, a 201 amino acid sequence with a two-fold symmetry was carefully designed one by one referring to the residue preferences of local structures. The proposed amino acid sequence was investigated by the 3D–1D alignment method, whether it prefers to fold to the β/α-barrel or not. The whole 3D model was diagnosed from several view points; the distribution of polar–non polar residues, the packing densities in the protein and the denaturation free energy.

The designed protein was produced in *E. coli* expression system as the fusion protein. The purified protein is monomeric and globular and has a large amount of secondary structures at low pH. The designed protein is very stable against GdnHCl as much as natural proteins, but it is similar to the molten globules of natural proteins, lacking a tight packing. The high thermostability is considered to be due to the large entropy in the globular state with the secondary structures. Much more knowledge on the construction of tight packing of the hydrophobic core is necessarily extracted from natural proteins.

Acknowledgement

The authors thank H. Anaguchi for preparation of the designed protein, E. Kitakuni and S. Okada for experiments of sedimentation equilibrium, Y. Matsuo and K. Nishikawa for their 3D–1D alignment study, C. Sander for the program "Poldiagnostics", F.E. Cohen for the program Qpack, G.J. Barton for the multiple-alignment program, Y. Kuroda and A.V. Finkelstein for discussion, and M. Ikehara for encouragement.

References

[1] J.U. Bowie, R. Luthy and D. Eisenberg, Science 253 (1991) 164.

[2] D.T. Jones, W.R. Taylor and J.M. Thornton, Nature 358 (1992) 86.

[3] E.I. Shakhnovich and A.M. Gutin, Nature 346 (1990) 773.

[4] A.V. Finkelstein and B.A. Reva, Nature 351 (1991) 497.

[5] K.M. Fiebig and K.A. Dill, J. Chem. Phys. 98 (1993) 3475.

[6] J.S. Richardson and D.C. Richardson, TIBS 14 (1989) 304.

[7] C. Sander, Curr. Opin. Struct. Biol. 1 (1991) 630.

[8] C. Sander, G. Vriend, F. Bazan, A. Horovitz, H. Nakamura, L. Ribas, A.V. Finkelstein, A. Lockhart, R. Merkl, J. Perry, S.C. Emery, C. Gaboriaud, C. Marks, J. Moult, C. Verlinde, M. Eberhard, A. Elofsson, T.J.P. Hubbard, L. Regan, J. Banks, R. Jappelli, A.M. Lesk and A. Tramontano, Prot. Struct. Funct. Genet. 12 (1992) 105.

[9] S.F. Betz, D.P. Raleigh and W.F. DeGrado, Curr. Opin. Struct. Biol., 3 (1993) 601.

[10] L. Regan and W.F. DeGrado, Science 241 (1988) 976.

[11] T. Handel and W.F. DeGrado, J. Am. Chem. Soc. 112 (1990) 6710.

[12] D.P. Raleigh and W.F. DeGrado, J. Am. Chem. Soc. 114 (1992) 10079.

[13] K.W. Hahn, W.A. Klis and J.M. Stewart, Science 248 (1990) 1544.

[14] L. Regan and N.D. Clarke, Biochemistry 29 (1990) 10878.

[15] J.S. Richardson, D.C. Richardson, N.B. Tweedy, K.M. Gernert, T.P. Quinn, M.H. Hecht, B.W. Erickson, Y. Yan, R.D. McClain, M.E. Donlan and M.C. Surles, Biophys. J. 63 (1992) 1186.

[16] A. Pessi, E. Bianchi, A. Crameri, S. Venturini, A. Tramontano and M. Sollazzo, Nature 362 (1993) 367.

[17] G.K. Farber and G.A. Petsko, TIBS 15 (1990) 228.

[18] T. Tanaka, K. Fukuhara, H. Nakamura, S. Saito, T. Tanaka, M. Hayashi, Y. Yamamoto, A. Kohara, M. Kikuchi and M. Ikehara, J. Cell Biochem. 14C (1990) 233.

[19] K. Goraj, A. Renard and J.A. Martial, Protein Eng. 3 (1990) 259.

[20] Y. Lindqvist, J. Mol. Biol. 209 (1989) 151.

[21] D.W. Banner, A.C. Bloomer, G.A. Petsko, D.C. Phillips, C.I. Pogson and I.A. Wilson, Nature 255 (1975) 609.

[22] C. Hyde and E.W. Miles, Bio/Technology 8 (1990) 27.

[23] Y. Matsuura, M. Kusunoki, W. Harada and M. Kakudo J. Biochem. 95 (1984) 697.

[24] W. Kabsch and C. Sander, Biopolymers 22 (1983) 2577.

[25] I. Lasters, S.J. Wodak, P. Alard and E. van Custem, Proc. Natl. Acad. Sci. USA 85 (1988) 3338.

[26] J.M. Thornton, B.L. Sibanda, M.S. Edwards and D.J. Barlow, BioEssays 8 (1988) 63.

[27] G.J. Barton and M.J.E. Sternberg, J. Mol. Biol. 212 (1990) 389.

[28] P.R. Sibbald and P. Argos, J. Mol. Biol. 216 (1990) 813.

[29] H. Nakamura, K. Katayanagi, K. Morikawa and M. Ikehara, Nucl. Acids Res. 19 (1991) 1817.

[30] J.W. Ponder and F.M. Richards, J. Mol. Biol. 193 (1987) 775.

[31] K. Morikami, T. Nakai, A. Kidera, M. Saito and H. Nakamura, Comput. Chem. 16 (1992) 243.

[32] S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta Jr. and P. Weiner, J. Am. Chem. Soc. 106 (1984) 765.

[33] K. Nishikawa and T. Ooi, Biochim. Biophys. Acta 871 (1986) 45.

[34] O.B. Ptitsyn and A.V. Finkelstein, Biopolymers 22 (1983) 15.

[35] J.-F. Gibrat, J. Garnier and B. Robson, J. Mol. Biol. 198 (1987) 425.

[36] N. Qian and T.J. Sejnowski, J. Mol. Biol. 202 (1988) 865.

[37] K. Nagano, J. Mol. Biol. 109 (1977) 251.

[38] V.I. Lim, J. Mol. Biol. 88 (1974) 857.

[39] K. Nishikawa and Y. Matsuo, Protein Eng. 6 (1993) 811.

[40] Y. Matsuo and K. Nishikawa, 11th International Biophysics Congress, Budapest (1993) A-5.7.

[41] G. Baumann, C. Frommel and C. Sander, Protein Eng. 2 (1989) 329.

[42] L.M. Gregoret and F.E. Cohen, J. Mol. Biol. 211 (1990) 959.

[43] T. Ooi and M. Oobatake, Proc. Natl. Acad. Sci. USA 88 (1991) 2859.

[44] J.W. Williams, Ultracentrifugation of macromolecules (Academic Press, New York, 1972).

[45] A.M. Lesk, C.-I. Branden and C. Chothia, Prot. Struct. Funct. Genet. 5 (1989) 139.

[46] J.-P. Scheerlinck, I. Lasters, M. Claesseus, M.De Maeyer, F. Pio, P. Delhaise and S.J. Wodak, Prot. Struct. Func. Genet. 12 (1992) 299.

[47] J.S. Richardson and D.C. Richardson, Science 240 (1988) 1648.

[48] D.C. Rees, M. Lewis and W.N. Lipscomb, J. Mol. Biol. 168 (1983) 367.

[49] J. Badger, S. Krishnaswamy, M.J. Kremer, M.A. Oliveira, M.G. Rossmann, B.A. Heinz, R.R. Rueckert, F.J. Dutko and M.A. McKinlay, J. Mol. Biol. 207 (1989) 163.

[50] C. Abad-Zapatero, J.P. Briffith, J.L. Sussman and M.G. Rossmann, J. Mol. Biol. 198 (1987) 445.

[51] H. Nakashima, K. Nishikawa and T. Ooi, Prot. Struct. Funct. Genet. 8 (1990) 173.

[52] S. Provencher and J. Glockner, Biochemistry 20 (1981) 33.

[53] K. Kuwajima, Prot. Struct. Funct. Genet. 6 (1989) 87.

[54] V.N. Uversky, G.V. Semisotnov, R.H. Pain and O.B. Ptitsyn, FEBS Letters 314 (1992) 89.

[55] A.V. Finkelstein and H. Nakamura, Protein Eng. 6 (1993) 367.

[56] K. Ishikawa, H. Nakamura, K. Morikawwa and S. Kanaya, Biochemistry 32 (1993) 6171.