Characterization of soluble artificial proteins with random sequences

Asao Yamauchi[a], Tetsuya Yomo[a,b], Fumihiro Tanaka[a], Irfan D. Prijambada[a], Seiji Ohhashi[a], Keizo Yamamoto[c], Yasufumi Shima[a], Kyoko Ogasahara[d], Katsuhide Yutani[d], Mikio Kataoka[e], Itaru Urabe[a,*]

[a]*Department of Biotechnology, Faculty of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565, Japan*
[b]*Form and Function Project, PRESTO, JST, 2-1 Yamadaoka, Suita, Osaka 565, Japan*
[c]*Department of Chemistry, Nara Medical University, 840 Shijo, Kashihara, Nara 634, Japan*
[d]*Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565, Japan*
[e]*Department of Earth and Space Science, Graduate School of Science, Osaka University, Toyonaka, Osaka 560, Japan*

Received 24 October 1997; revised version received 8 December 1997

Abstract The structural and catalytic properties of two soluble random proteins, RP3-42 and RP3-45, of 141 amino acid residues were investigated. Although no marked secondary structure was detected by CD spectrum, sedimentation equilibrium and small-angle X-ray scattering studies showed that they form an oligomeric structure and are as compact as the molten globule. The random proteins have low but distinct esterase activity; the values of the second-order rate constant for the hydrolysis of *p*-nitrophenol were 0.78 and 1.39 M^{-1} s^{-1} for RP3-42 and RP3-45, respectively. The differences in the properties of the random and the native proteins are discussed from the evolutionary point of view.

Key words: Artificial protein; Random sequence; Enzyme evolution; Circular dichroism spectrum; Small-angle X-ray scattering; Esterase activity

1. Introduction

Advancement in what we know about proteins has been obtained from the observation of natural proteins. However, in view of the fact that natural proteins have a long history in which some of their properties must have been specialized due to evolutionary constraints, we may have a biased knowledge with regard to the properties of proteins. Therefore, to know proteins well, it is far better to distinguish the general properties of natural proteins from the specialized one in the event of evolution. In our previous work, we prepared a library of 141 amino acid residue proteins with random sequences [1]. As the properties of the artificial random proteins are free from any evolutionary constraints, they are fit to be the standard tool for discriminating the properties of natural proteins. Our previous work showed that the probability of random proteins being soluble is about 20% [1]. Therefore, solubility does not seem to be a specialized property of natural proteins. In this work, we examined the structural and catalytic properties of the two soluble random proteins, RP3-42 and RP3-45 [1]. The random proteins form an oligomeric and compact globular structure without any marked secondary structure, and have low but distinct esterase activity. Based on the results, the possible evolutionary process from a primordial to a natural enzyme is presented.

*Corresponding author. Fax: (81) (6) 879 7428.

Abbreviation: TFE, 2,2,2-trifluoroethanol

2. Materials and methods

2.1. Soluble random proteins

Two soluble proteins, RP3-42 and RP3-45, were purified from the library of constructed random proteins as described previously [1]. Protein concentration was measured spectrophotometrically at 280 nm using the molar absorption coefficients of 44 100 M^{-1} cm^{-1} and 60 000 M^{-1} cm^{-1} for RP3-42 and RP3-45, respectively, which were determined from the amino acid composition (see legend to Fig. 1) and A_{280} value of the purified protein solution according to the procedure described by Suga et al. [2]. Based on the deduced amino acid sequences, the values of the molecular weight of RP3-42 and RP3-45 were estimated to be 14 750 and 14 726, respectively.

2.2. Structural analysis

CD spectra in the far-UV region were recorded on a Jasco J-720 spectropolarimeter. The spectra were scanned eight times at a scan rate of 20 nm/min, using a 0.25 s time constant and a spectral band width of 1 nm. The light pass length of the cell used was 1 mm. For calculation of the mean residue ellipticity, [θ], the mean residue weight was taken as 106.

Sedimentation equilibrium studies were carried out at 25°C in 50 mM potassium phosphate (pH 7.0) at 12 000 rpm using a Beckman XL-A analytical ultracentrifuge. Initial protein concentrations were 10.8 μM and 9.5 μM for RP3-42 and RP3-45, respectively.

Solution X-ray scattering experiments were carried out at the solution scattering station (SAXES) installed at BL-10C, the Photon Factory, Tsukuba, Japan, using synchrotron radiation [3,4]. The experimental conditions were as described previously [5]. The radius of gyration (R_g) and the intensity at zero scattering angle [$I(0)$] were estimated by the Guinier approximation, $I(Q) = I(0)\exp(-R_g^2Q^2/3)$, where Q is the momentum transfer [6]. Q is defined as $Q = 4\pi \sin\theta/\lambda$, where 2θ and λ are the X-ray scattering angle and wavelength (1.488 Å), respectively.

2.3. Esterase activity

The hydrolysis rate of *p*-nitrophenyl acetate was measured at 40°C in 50 mM potassium phosphate (pH 7.0) using a Hitachi 220A spectrophotometer. The reactions were recorded as the increase in absorbance at 410 nm. The concentration of *p*-nitrophenol released by the reaction was calculated using a molar absorption coefficient of 8410 M^{-1} cm^{-1}, which was determined experimentally at pH 7.0.

3. Results and discussion

3.1. Ammonium sulfate precipitation

The solubility of both random proteins decreased in the presence of a high concentration of ammonium sulfate as does that of natural proteins. RP3-45 precipitated at a lower concentration of ammonium sulfate than RP3-42. Half of the RP3-45 and RP3-42 (10 μM) was precipitated at the concentrations of 0.6 and 0.8 M ammonium sulfate, respectively. The difference in solubility may be related to the difference in the hydrophobicity of these proteins. Using the score of each

PII S0014-5793(97)01552-4

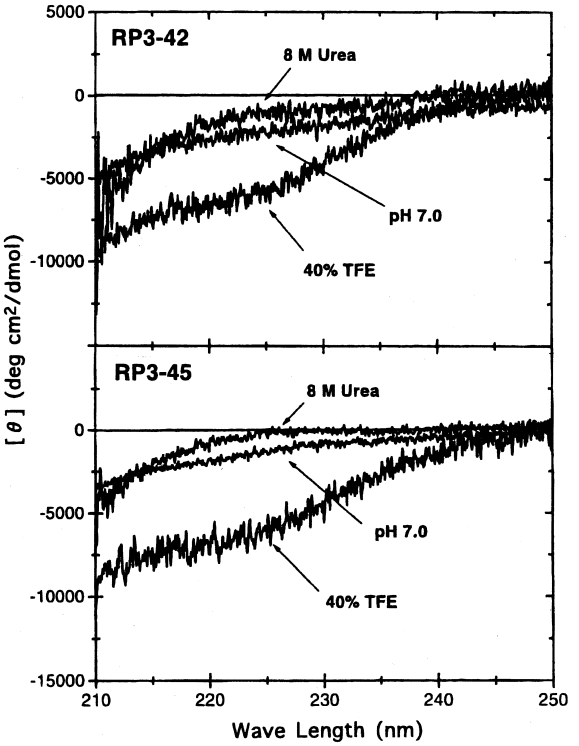

Fig. 1. CD spectra of RP3-42 and RP3-45. The spectra were measured at 25°C in 50 mM potassium phosphate (pH 7.0) in the absence or presence of either 8 M urea or 40% TFE. Protein concentrations used were in the range of 0.05–0.1 mg/ml. The amino acid sequences of the random proteins deduced from their nucleotide sequences are: RP3-42, MASMTGGQQM GRGSRGSILE GPHVGNPPSW GIPKLEKSQA GEPPFWDFCS SSCRNSPVWG SPRLGSSQIG KFPEWDSPNM GDFQLGSKLE TFHVGGIPCW ETSILERSHL GGLPNWENSA AGSTRSGTGL ESTDLASSN; RP3-45, MASMTGGQQM GRGSRGSNLG ESHVGEIPGW GYPSLEVFQL GSPPNWEVPA ARAAGSSRLE GPQVGIPPAW RTSNMEYFQN GGSPGWIQLG DFPCWKLPML EKLQPGSLPI WKDSSLEYSA AGSTRSGTGL ESTDLASSN. It is to be noted that there were some typographical errors in the sequences reported previously [1]. Phe-41, Gly-45, Ser-49, Cys-50, and Leu-78 in the sequence of RP3-42 are Gly-41, Phe-45, Cys-49, Ser-50, and Pro-78 respectively, while Cys-41 and Gly-45 of RP3-04 are Gly-41 and Cys-45, respectively, and Ile-18, Ala-41, and Gly-45 in RP3-65 are Ser-18, Gly-41, and Ala-45, respectively.

amino acid residue reported by Fauchere and Pliska [7] and the amino acid composition, the values of the hydrophobicity of RP3-42 and RP3-45 are calculated to be 50 and 53, respectively. These results suggest that a less hydrophobic protein precipitates at a higher ammonium sulfate concentration.

When the proteins of various concentrations were heated at 65°C for 20 min, a decrease in the solubility was observed. For example, the solubility in 0.61 M ammonium sulfate decreased from 78% to 26% for RP3-42 and from 43% to 18% for RP3-45. These results indicate that at a higher temperature, the chance of the hydrophobic residues being exposed to the surface of the proteins increases thereby decreasing its solubility. Therefore, like a soluble and flexible amphipathic polymer, the soluble random proteins may assume a globular structure such that most hydrophobic residues are located on the inside and hydrophilic residues on the outside.

3.2. Secondary structure

CD spectra of RP3-42 and RP3-45 were measured at pH 7.0 in the presence and absence of 8 M urea. As shown in Fig. 1, the CD spectra of both proteins have similar patterns and the presence of 8 M urea caused no significant change in both spectra. In addition, almost the same spectra were observed for both proteins in the pH region from 2 to 11 (data not shown). These results indicate that the random proteins do not have any marked secondary structure. Accordingly, although the soluble random proteins may have a globular structure with hydrophobic residues on the inside, the structure seems to be highly disordered.

The CD spectra of RP3-42 and RP3-45 were also measured in the presence of various concentrations of trifluoroethanol (TFE), which is known to stabilize the α-helical structure in proteins [8]. For both proteins, the addition of more than 30% TFE (v/v) decreased the value of ellipticity at 222 nm and reached a minimum with more than 40% TFE (Fig. 1). These results suggest that although the random proteins have intrinsically some propensity to form an α-helical structure, the structure is not stable in an aqueous environment.

Artificial proteins composed mainly of a random combination of glutamine, leucine, and arginine (QLR proteins) have been reported to have helical structures [9,10], while RP3-42 and RP3-45, composed of 20 kinds of amino acids, do not have a secondary structure. Therefore, this suggests that proteins with selected kinds of amino acids have the tendency to form a secondary structure. However, it is to be noted that natural proteins are composed of 20 kinds of amino acids, the

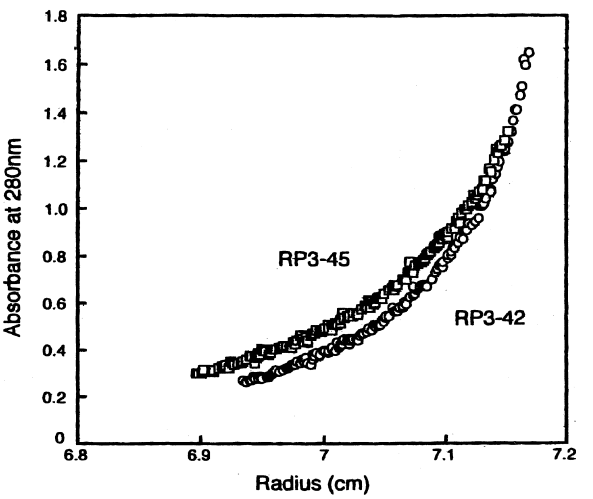

Fig. 2. Distribution profiles of RP3-42 and RP3-45 after equilibrium centrifugation. The data for both proteins were best fitted by a monomer-tetramer function:

$$A_r = A_M \exp[M\omega^2(1-v\rho)(r^2-r_0^2)/(2RT)] + A_T \exp[4M\omega^2(1-v\rho)(r^2-r_0^2)/(2RT)]$$

where A_r is the total solute absorbance at radius r, A_M and A_T are the absorbance of the monomeric and tetrameric species, respectively, at a reference distance r_0, ω is the angular velocity, M is the molecular weight of the protein as a monomer, v is the partial specific volume of the protein, ρ is the density of the solution, R is the gas constant, and T is the temperature. The values of the constants use were as follows: for RP3-42, $r_0 = 6.9355$, $M = 14750$, and $\omega^2(1-v\rho)/(2RT) = 1.8474 \times 10^{-5}$; for RP3-45, $r_0 = 6.8975$, $M = 14726$, and $\omega^2(1-v\rho)/(2RT) = 1.2265 \times 10^{-5}$. The equilibrium constant for tetramer association was calculated as $K = \varepsilon^3 A_T/(4A_M^4)$, where ε is the molar absorption coefficient of the protein.

same as RP3-42 and RP3-45, but each protein has a unique secondary and tertiary structure. Hence, this implies that the folding of natural proteins into the unique structures seems to have been oriented during the course of evolution, hence is a specialized property of natural proteins.

3.3. Oligomeric structure and molecular size

Fig. 2 shows the results of sedimentation equilibrium studies. The data were best fitted in the monomer-tetramer function with equilibrium constants for tetramer association of 1.6×10^{14} M^{-3} for RP3-42 and 1.0×10^{15} M^{-3} for RP3-45. This indicates that most protein molecules are present as monomeric and tetrameric species in a solution.

To confirm the oligomeric structure of the random proteins, RP3-45 was analyzed by small-angle X-ray scattering at a concentration of 180 μM, where almost all molecules are expected to be in the tetrameric form as based on the above data. Based on the Guinier plot, the intensity at zero scattering angle [$I(0)$] is proportional to the molar concentration and the square of the molecular weight of a protein. The value of $I(0)$ of an oligomeric protein with n protomers is then expressed as:

$$I(0) \propto (c/n)(nM)^2 = ncM^2 \quad (1)$$

where c is the molar concentration of the protomer and M is the molecular weight of the protomer. This equation indicates that $I(0)/(cM^2)$ is proportional to n. Hence, the value of n can be estimated as the ratio of the $I(0)/(cM^2)$ values of the oligomeric and a monomeric proteins. The values of $I(0)$ were 25.50 ± 0.09 and 20.16 ± 0.04 for RP3-45 (180 μM) and cytochrome c (635 μM), respectively; the latter is a monomeric protein (i.e. $n=1$) with an M value of 12 384. Accordingly, the n value of RP3-45 ($M=14\,726$) was calculated to be 3.2, indicating a trimeric or tetrameric structure of RP3-45. In addition, most QLR proteins are reported to be in an oligomeric state [9,10]. Therefore, the oligomeric structures often observed in natural proteins seem to be a general property of proteins.

The values of the radius of gyration (R_g) have been measured by small-angle X-ray scattering for many proteins [5,11–16]. Interestingly, as shown in Fig. 3, the R_g values of native proteins correlate well with their molecular weight (nM) with a regression line of

$$R_g = 0.05(nM)^{0.36} \quad (2)$$

It is worth noticing that the value of the exponent, 0.36, indicates that R_g is almost proportional to $(nM)^{1/3}$. This means that native proteins have similar shapes with the same degree of deviation from a sphere, regardless of the difference in sizes. It is known that the R_g value of the denatured or molten globule state of a protein is about 2 or 1.3 times, respectively, as large as that of the native state [5,14,16]. The R_g value of RP3-45 was measured to be 31.9 ± 0.4, which is 1.3 times as large as that of a native protein with the same nM value estimated from the curve in Fig. 3. Therefore, RP3-45 is as compact as a molten globule state, and much more compact than the denatured state, though RP3-45 has no secondary structure (Fig. 1). Compared to the random structure of the denatured state of a natural protein as produced by artificially expanding the native compact structure through the influence of denaturants, RP3-45 by itself has an intrinsically random structure, hence allowing it to be more compact than the denatured state. From these results, we propose that in general, a protein with an intrinsically random structure takes a compact globular form with most hydrophobic residues located on the inside.

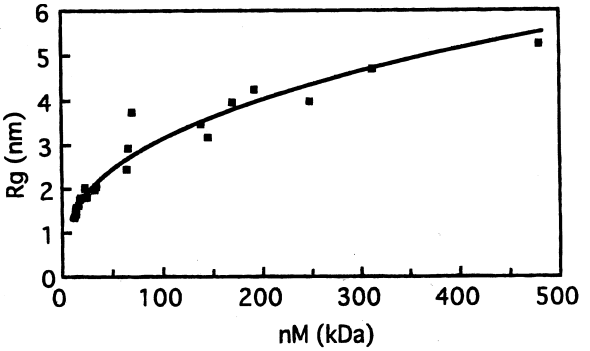

Fig. 3. Correlation between the radius of gyration (R_g) and the molecular weight (nM) of various native proteins. Data are from [5,11–16].

Table 1
Primordial esterase activity

Protein	Observed k_E (M^{-1} s^{-1})	Calculated k_E[d] (M^{-1} s^{-1})
RP3-42	0.78[a]	0.76
RP3-45	1.39[a]	0.32
Metmyoglobin	0.3[b]	0.69
Mercaptalbumin	4[c]	1.82

The observed k_E values were measured at the following temperatures: [a]40°C (Fig. 5); [b]25°C [17]; and [c]25°C [18].
[d]The k_E value of each protein was calculated assuming that the activity is the sum of the activities of the amino acid residues contained in each protein. Esterase activities of various amino acids and their derivatives were measured at 40°C. From the results, only the following k_E values were used for calculations: the values from the *N*-acetyl derivatives were: 0.196 (Cys), 0.0288 (His), 0.0152 (Lys), and 0.0112 M^{-1} s^{-1} (Tyr); for the N-terminal amino group, a k_E value of 0.02 M^{-1} s^{-1} was used. The k_E values of the other amino acid residues were neglected as they are insignificant. Calculations were based on the amino acid compositions of the proteins RP3-42: Cys (3), His (3), Lys (4), and Tyr (0); RP3-45: Cys (1), His (1), Lys (3), and Tyr (3); metmyoglobin [17]: Cys (0), His (12), Lys (19), and Tyr (3); and mercaptalbumin Cys (1), His (17), Lys (59), and Tyr (19).

3.4. Catalytic activity

Many natural proteins have catalytic activities as enzymes. Though artificial random proteins may not have the level of activity of natural enzymes, still possibilities exist that there may be some primordial activity. Here, the level of esterase activity exhibited by the random proteins was examined using *p*-nitrophenyl acetate as a substrate. When the hydrolysis of these esters was monitored in the presence and absence of RP3-42 or RP3-45, a clear increase in the reaction rate was observed in the presence of the random proteins (Fig. 4).

Assuming the first-order kinetics, the hydrolysis rate of the ester, S, in the absence of a random protein, i.e. the background rate, is expressed as:

$$d[S]/dt = -k_B[S] \quad (3)$$

where k_B is the rate constant of the background reaction. In

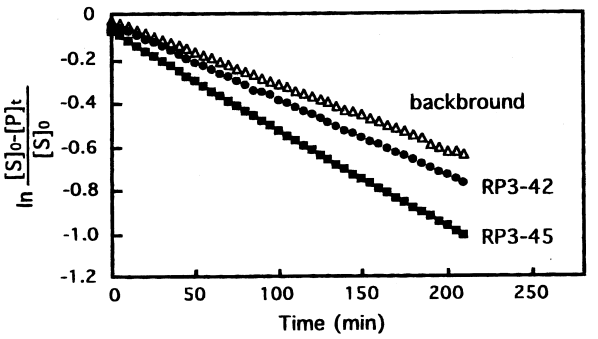

Fig. 4. First-order plot of kinetic data for the hydrolysis of *p*-nitrophenol. Initial *p*-nitrophenol concentration ($[S]_0$) was 287 μM. ● 12.0 μM RP3-42, ■ 19.1 μM RP3-45, and △ background.

the presence of a random protein, the reaction rate is expressed as:

$$d[S]/dt = -(k_E[E] + k_B)[S] \quad (4)$$

where k_E is the rate constant of the catalytic reaction by a random protein (E). Eqs. 3 and 4 are converted, respectively, to

$$\ln\{([S]_0-[P]_t)/[S]_0\} = -k_B t \quad (5)$$

$$\ln\{([S]_0-[P]_t)/[S]_0\} = -(k_E[E] + k_B)t \quad (6)$$

where t is the reaction time, $[S]_0$ is the initial concentration of the ester, and $[P]_t$ is the concentration of *p*-nitrophenol released by the reaction at time t.

Fig. 4 shows that hydrolysis of the ester in the presence of RP3-42 or RP3-45 followed the linear relationship of Eq. 6. This indicates that (k_E[E]+k_B) is constant during the course of the reaction, and hence [E] does not decrease. Therefore, the release of *p*-nitrophenol is not due to the acetylation of the random proteins, but to the catalyzed hydrolysis of the ester. The slopes of the reactions give a k_B value of 4.80×10^{-5} s^{-1} of the background reaction and k_E values of 0.78 and 1.39 M^{-1} s^{-1} for RP3-42 and RP3-45, respectively (Table 1). In addition, at reaction time 210 min, 25.2 and 57.3 μM, respectively, of the product was released due to the presence of 12.0 μM RP3-42 and 19.1 μM RP3-45.

To rule out the possibility that the observed esterase activity in RP3-42 and RP3-45 might be due to contamination of enzyme(s) present in the host cell, a negative control using cell lysate of the host *Escherichia coli* was prepared by the procedures described previously [1] and the esterase activity was measured. The amount of the host lysate used for the assay was equivalent to the amount of lysate used in the preparation of the random proteins. The k_E[E] value of the host lysate (2.36×10^{-4} s^{-1}) decreased (1.12×10^{-5} s^{-1}) upon heating at 65°C for 20 min. The activity decreased further to 5.80×10^{-6} s^{-1} by a second heat treatment. On the other hand, heat treatment did not cause any decrease in the k_E[E] values of RP3-42 and RP3-45, which were 9.40×10^{-6} s^{-1} and 2.67×10^{-5} s^{-1}, respectively. These results clearly show that the esterase activity observed is a catalytic property of the random proteins (Fig. 4).

The results obtained so far suggest that the soluble random proteins work as a catalyst with a very low level of activity. As shown in Table 1, the observed level of activity is of the same order as that calculated by assuming that the activity is the sum of the activities of the amino acid residues contained in each protein. Interestingly, low esterase activity was also found in metmyoglobin [17] and mercaptalbumin [18], though both proteins are better known in their different physiological functions. As shown in Table 1, the level of activity of the two known proteins is similar to that of the random proteins and so with the calculated values. Hence, the random proteins as enzymes are at a primitive level, and the known high and specific activity of natural enzymes seems to have been acquired during the course of evolution. Therefore, the soluble random proteins can be regarded as a model of primordial enzymes. In addition, it was found that the difference in the k_E values of RP3-42 and RP3-45 reflects the difference in their amino acid sequences.

The properties of the soluble random proteins suggest that a primordial enzyme may not take a unique conformation but can still have catalytic activity. Due to high flexibility of the structure, the low activity observed for the random proteins may arise from the average of the many kinds of conformations that a protein may take as a primordial enzyme. As this study shows that the activity changes in accordance with the amino acid sequence, then such a primordial enzyme can evolve by functional selection. The increase in activity due to the changes in the sequence confines the protein to taking the ensemble of the more active conformations rather than the less active ones. Such restriction in the flexibility of the conformation eventually leads to a unique and stable active conformation. Recently, such an evolutionary process was simulated with a simple spin-glass-like model of proteins [19]. The results of the simulation show that selection based only on the local configurations of active site residues is sufficient for the protein to fold into a unique conformation [19]. Therefore, we propose that the unique conformation of a natural protein has been acquired by the evolutionary constraint on catalytic efficiency. The random proteins thereby serve as an important tool for further elucidation of the hypothesis.

Acknowledgements: This work was supported in part by Grants 08272214, 07280101, 09440147 and 09450304 from the Ministry of Education, Science, Sports and Culture, Japan.

References

[1] Prijambada, I.D., Yomo, T., Tanaka, F., Kawama, T., Yamamoto, K., Hasegawa, A., Shima, Y., Negoro, S. and Urabe, I. (1996) FEBS Lett. 382, 21–25.
[2] Suga, Y., Yomo, T. and Urabe, I. (1996) J. Ferment. Bioeng. 81, 259–261.
[3] Ueki, T. (1991) Nucl. Instrum. Methods Phys. Res. A303, 464–475.
[4] Kataoka, M., Head, J.F., Persechini, A., Kretsinger, R.H. and Englman, D.M. (1991) Biochemistry 30, 1188–1192.
[5] Kataoka, M., Nishii, I., Fujisawa, T., Ueki, T., Tokunaga, F. and Goto, Y. (1995) J. Mol. Biol. 249, 215–228.
[6] Guinier, A. and Fournet, B. (1955) Small-Angle Scattering of X-rays, John Wiley, New York.
[7] Fauchere, J.-L. and Pliska, V. (1983) Eur. J. Med. Chem. 18, 369–375.
[8] Shiraki, K., Nishikawa, K. and Goto, Y. (1995) J. Mol. Biol. 245, 180–194.
[9] Davidson, A.R. and Sauer, T. (1994) Proc. Natl. Acad. Sci. USA 91, 2146–2150.

[10] Davidson, A.R., Lumb, K.J. and Sauer, R. (1995) Nature Struct. Biol. 2, 856–864.
[11] Kumosinski, T.F. and Pessen, H. (1985) Methods Enzymol. 117, 154–182.
[12] Kataoka, M., Head, J.F., Seaton, B.A. and Engelman, D.M. (1989) Proc. Natl. Acad. Sci. USA 86, 6944–6948.
[13] Flanagan, J.M., Kataoka, M., Shortle, D. and Engelman, D.M. (1992) Proc. Natl. Acad. Sci. USA 89, 748–752.
[14] Kataoka, M., Hagihara, Y., Mihara, K. and Goto, Y. (1993) J. Mol. Biol. 229, 591–596.
[15] Shi, L., Kataoka, M. and Fink, A.L. (1996) Biochemistry 35, 3297–3308.
[16] Kataoka, M., Kuwajima, K., Tokunaga, F. and Goto, Y. (1997) Protein Sci. 6, 422–430.
[17] Breslow, E. and Gurd, F.R.N. (1962) J. Biol. Chem. 237, 371–381.
[18] Tildon, J.T. and Ogilvie, J.W. (1972) J. Biol. Chem. 247, 1265–1271.
[19] Saito, S., Sasai, M. and Yomo, T. (1997) Proc. Natl. Acad. Sci. USA 94, 11324–11328.