Characterization of random-sequence proteins displayed on the surface of *Escherichia coli* RNase HI

Nobuhide Doi[a,b], Tetsuya Yomo[c], Mitsuhiro Itaya[a], Hiroshi Yanagawa[a,*]

[a] *Mitsubishi Kasei Institute of Life Sciences, Machida, Tokyo 194, Japan*
[b] *Graduate School of Environmental Earth Sciences, Hokkaido University, Sapporo, Hokkaido 060, Japan*
[c] *Department of Biotechnology, Osaka University, Suita, Osaka 565, Japan*

Received 5 March 1998; revised version received 25 March 1998

Abstract In a previous study, random-sequence proteins of 120–130 amino acid residues were inserted into the surface loop region of the enzyme, *Escherichia coli* RNase HI [Doi et al. (1997) FEBS Lett. 402, 177–180]. Here we established that the RNase H activity of the insertion mutants is correlated with their secondary structure contents evaluated by circular dichroism measurement at 222 nm. The random-sequence insert of a mutant enzyme possessing relatively high RNase H activity was detached from the RNase HI scaffold, and its characterization indicated that the random-sequence protein maintains its secondary structure after separation from the scaffold. Thus, the structural features of random-sequence proteins were suggested to be monitored by measuring the activity of the scaffold enzyme into which these proteins have been inserted.

Key words: Directed evolution; Evolutionary engineering; Insertional mutagenesis; Protein folding; Protein scaffold

1. Introduction

For the past decade, several researchers have been designing synthetic gene libraries encoding random sequences with long ORFs [1–5]. Pools of random-sequence polypeptides synthesized from such gene libraries can be used to identify the inherent properties of natural proteins [6]. Random protein libraries are also useful as potential sources of novel proteins or enzymes, if efficient selection methods are available [7]. So far, random peptide libraries have often been displayed on the surface of scaffold proteins (reviewed in [8]). This approach was expected to have the following advantages. Random-sequence polypeptides on a scaffold are conformationally constrained because of the proximity of their N- and C-termini, and thus are stabilized by the reduction of conformational entropy. Further, insertion of a random sequence into the middle of a scaffold protein protects it from exopeptidase digestion in vivo.

What fractions of all random-sequence proteins have folded structures is an important question. However, experimental studies using random protein libraries face the difficulty that no efficient screening system is currently available for selecting folded proteins from expression libraries. Indeed, in a random protein library designed by one of the authors, 20% of the proteins were soluble [4], but no protein with extensive secondary structure was found [6]. In our previous work, random-sequence proteins in the library were inserted into the surface loop region of *Escherichia coli* RNase HI [9]. We hoped to monitor the structural features of the inserted random-sequence proteins by measuring the scaffold RNase H activity. If greater instability or flexibility of inserted random-sequence polypeptides results in more stress on the RNase HI scaffold with a consequent reduction in enzyme activity, insertion mutants with high RNase H activity may be expected to contain random-sequence proteins with high stability and/or well-ordered structure. To establish such a selection system in which folded proteins are screened by monitoring a scaffold enzyme activity, the following criteria must be met. (i) The random protein library must be displayed on the surface of an active enzyme. (ii) The structure of the inserted random-sequence protein must correlate with the activity of the mutant scaffold enzyme. (iii) The random-sequence protein must maintain its structure following detachment from the scaffold. In our previous study, we were able to isolate clones corresponding to roughly 10% of the library [9]. To investigate the latter two points, we characterized in the present study the structure-activity relationship of insertion mutants of RNase HI and we examined whether any change occurred in the secondary structure of a random-sequence protein following its detachment from the RNase HI scaffold.

2. Materials and methods

2.1. Construction and screening of library

Random DNA sequences encoding 120–130 amino acid proteins were inserted into the *rnhA* gene encoding *E. coli* RNase HI as previously described [9]. Mutant RNase HI genes were screened from the library on the basis of their ability to suppress the temperature-sensitive growth phenotype of *E. coli* strain MIC3037 (*rnhA-339*::*cat recC271*) [10]. The inserted random sequences of the isolated mutants were determined by sequencing both strands of DNAs using an ALF DNA sequencer and an AutoCycle sequencing kit (Pharmacia).

2.2. Overexpression and purification

The mutant and wild-type RNase HI genes and the random-sequence region of one of the mutant genes were amplified by PCR with two primers, one containing a *Nde*I site and the other a *Sal*I site. The PCR product was digested with *Nde*I and *Sal*I and then subcloned into the *Nde*I-*Xho*I fragment of pET20b(+) (Novagen). This resulted in overexpression of recombinant proteins controlled by the T7 promoter and addition of a C-terminal hexahistidine sequence to permit affinity purification with nickel-NTA agarose resin (Qiagen). *E. coli* MIC2090 (*rnhA-339*::*cat recB270* λ(*DE3*)) cells transfected with individual recombinant plasmids were grown in 2×YT broth containing 50 μg/ml carbenicillin at 25 or 30°C. When the culture achieved an optical density of 0.4–0.6 at 600 nm, the cells were induced by the addition of isopropylthio-β-D-galactoside to a final concentration of 1 mM. After an additional 4 h incubation, the cells were harvested by centrifugation. The pellets thus obtained were resuspended in lysis buffer and frozen-thawed on ice three times [11]. The centrifuged supernatant was loaded on the affinity column

*Corresponding author. Fax: (81) (427) 24-6317.

Abbreviations: ANS, 1-anilinonaphthalene-8-sulfonic acid; CD, circular dichroism; ORF, open reading frame; PCR, polymerase chain reaction

PII S0014-5793(98)00392-5

Fig. 1. Amino acid sequences of the random-sequence inserts in the nine isolated mutants of *E. coli* RNase HI. The site of insertion was between Glu^{61} and Glu^{64} of the enzyme. Partial consensus sequences corresponding to restriction sites of cassette junctions are shaded. HR301-5 is a mutant of HR301 [9], whose substituted residues are underlined. The somewhat shorter sequence of HR305 was caused by coincidental appearance of a cloning restriction enzyme site in the random-sequence region.

and the recombinant proteins were eluted with a pH gradient. The eluted fractions were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The protein molar concentration was determined from UV absorption at 280 nm and molar absorption coefficients calculated from $\varepsilon = 5690$ M^{-1} cm^{-1} for Trp, 1280 M^{-1} cm^{-1} for Tyr, and 120 M^{-1} cm^{-1} for Cys [12].

2.3. Enzymatic activity

RNase H activity of mutant proteins was determined by means of three assays: in vivo, in vitro, and activity staining in a gel, as previously described [9].

2.4. CD and fluorescence measurements

CD spectra of purified proteins were measured on a J-600 spectropolarimeter (JASCO) as previously described [9] and expressed as total molar ellipticity, [θ]. Fluorescence measurements were performed at 25°C on a Shimazu RF-502 spectrofluorophotometer. The emission spectra of aromatic residues of proteins were measured at an excitation wavelength of 280 nm, and the fluorescence spectra of 50 μM ANS in the absence and presence of 1 μM protein were measured with excitation at 376 nm. All experiments were conducted in 10 mM sodium acetate buffer (pH 5.5) containing 0.1 M NaCl.

3. Results and discussion

3.1. Insertion of random-sequence proteins into RNase HI

In the previous study, we first found that random-sequence proteins of more than 100 amino acid residues can be inserted into an active enzyme, *E. coli* RNase HI [9]. Nine mutant proteins were isolated and characterized; they had various levels of RNase H activity as determined by in vivo and renaturation gel assay [9]. The inserted random amino acid sequences of the nine isolated mutants were determined in this study (Fig. 1) by sequencing both strands of the DNA inserted into the *rnhA* gene; the sequences included codons for all 20 naturally occurring amino acids, but no stop codon. For assessment of the dependence of scaffold RNase H activity on the amino acid composition of the inserted random sequences, we calculated the content of hydrophobic residues (Leu, Ile, Phe, Met, and Val) which are often buried in the protein interior, and the content of charged residues (Asp, Glu, Lys, and Arg) which are found almost exclusively on the surface of natural proteins [13].

As shown in Fig. 2, two mutants, HR301 and HR306, with low RNase H activity in vivo have lower contents of hydrophobic residues in the inserts. They may undergo some degradation in the hydrophilic random-sequence regions by cellular proteases in vivo, which would be consistent with the

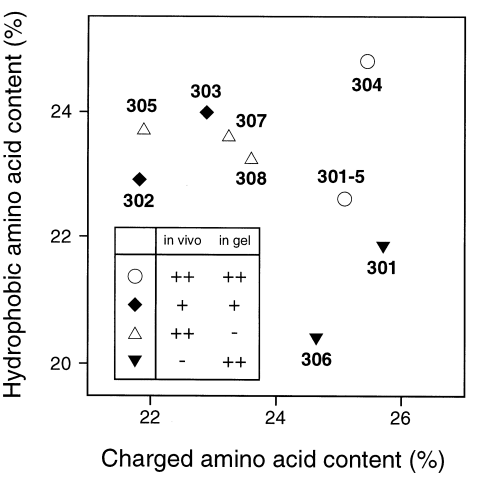

Fig. 2. Contents of hydrophobic vs. charged amino acid residues in the mutant RNase HI proteins. The ordinate and abscissa indicate hydrophobic amino acid content (Ile, Leu, Met, Phe and Val) and charged amino acid content (Asp, Glu, Lys and Arg), respectively. The RNase H activities of the mutant proteins analyzed by in vivo assay and by renaturation gel assay [9] are shown in the lower left-hand inset; symbols correspond to those in the main figure.

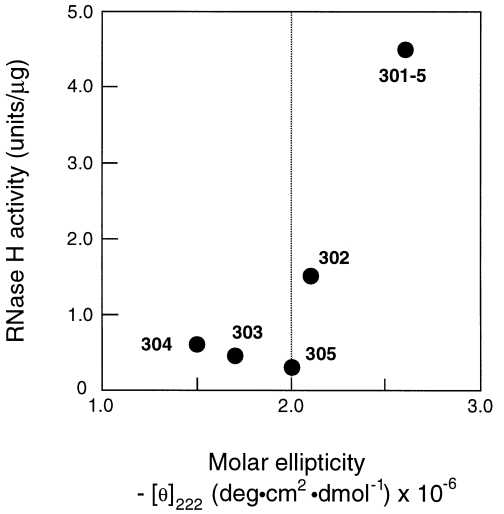

Fig. 3. Structure-activity relationship of mutant RNase HI proteins. The secondary structure (abscissa) was estimated from the CD molar ellipticity at 222 nm. The center line indicates the molar ellipticity of wild-type RNase HI. RNase H activity (ordinate) was determined from the radioactivity of the acid-soluble digestion product formed from ^{32}P-poly(rA)·poly(dT) [9]. Specific activity of the wild-type enzyme was 150 units/μg [19], where one unit is defined as that generating 1 nmol of acid-soluble nucleotide per 15 min at 37°C.

failure of attempts to overexpress these two proteins [9]. On the other hand, three mutants, HR305, HR307 and HR308, with low activity in the renaturation gel assay have relatively high hydrophobic levels and low charged contents. These mutants may have difficulty in refolding in the gel due to a tendency to aggregate. These results suggested that the balance of polar and non-polar amino acid contents of the inserted random-sequence proteins affect the scaffold RNase H activity. In addition, not only amino acid composition but also sequence of the insert seems important in determining molecular properties, because the mutants HR302 and HR303, belonging to different types of activity, have similar amino acid compositions with HR305, HR307 and HR308.

3.2. Structure-activity relationship of mutant RNase HI proteins

To examine whether the structure of the inserted random-sequence proteins also affects the scaffold RNase H activity, five insertion mutants of RNase HI were purified and characterized by CD measurement and in vitro RNase H assay [9]. As shown in Fig. 3, we found a correlation between the specific activity of the mutant RNase HI proteins and their molar ellipticity at 222 nm. The mutants HR301-5 and HR302 with relatively high activity showed slightly higher molar ellipticity than that of the wild-type enzyme, suggesting the presence of a definite helical structure in the inserted random-sequence region. On the other hand, the molar ellipticity of the mutants HR303, HR304, and HR305, with low RNase H activity was comparable to or more less than that of the wild-type enzyme, indicating the secondary structure of the insertion mutant proteins to be due predominantly that of RNase HI, i.e. the inserted region has a random coil structure.

It is somewhat surprising that the mutant RNase HI proteins could fold correctly in spite of the presence of a large internal unfolded region. Yamasaki et al. previously indicated that α-helix I of RNase HI (residues 34–60) is the initial site of folding [14]. Recently, residues 1–36 and 62–121 of RNase HI were also suggested to form small structural units defined by minima of the van der Waals interaction energy [15]. Since our random sequences were inserted into the 61–64 site located between these predicted folding units, the N- and C-terminal RNase HI fragments may independently fold and assemble along the large flexible linker.

3.3. Structural analyses of a random-sequence protein detached from RNase HI

Does the inserted random-sequence protein maintain its structure in the absence of the scaffold RNase HI? The purified mutant, HR301-5, with relatively high activity showed a high helical content among the isolated mutants. Thus, the inserted random sequence of the HR301-5 gene was subcloned, overexpressed in soluble form, purified as a random-sequence protein designated as R301-5, and characterized by CD and fluorescence measurements.

The far-UV CD spectrum of R301-5 revealed secondary structure (Fig. 4), and the helical content was estimated as 8% based on the ellipticity at 222 nm [16]. The spectrum of the random-sequence protein was in agreement with that calculated by subtraction of the spectrum of the scaffold RNase HI from that of the insertion mutant, HR301-5. Thus, the secondary structure of the random-sequence region of HR301-5 would appear to remain essentially intact after detachment of this region from the RNase HI scaffold.

In the fluorescence experiments (Fig. 5), the emission maximum of the four tryptophan residues in the random-sequence protein, R301-5, was at 346 nm in aqueous buffer, suggesting that the almost all the tryptophans are exposed to the solvent [17]. The emission maximum of R301-5 in buffer containing 6 M urea, however, slightly shifted to 350 nm and its intensity decreased by half (Fig. 5A). Thus, some of the tryptophan side chains of the random-sequence protein would appear to be located in a hydrophobic environment under native conditions and to undergo denaturation in 6 M urea. The presence of hydrophobic clusters in the random-sequence polypep-

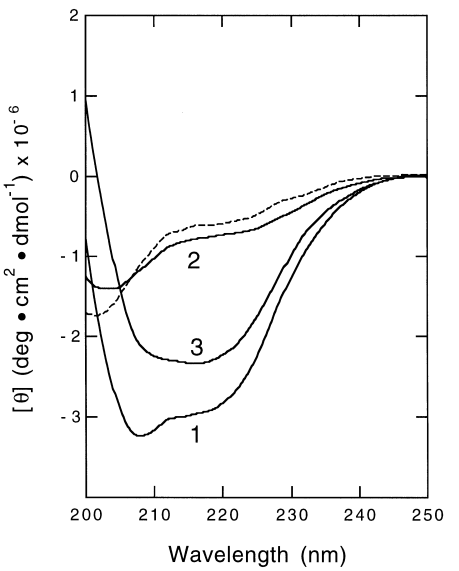

Fig. 4. Far-UV CD spectra of RNase HI mutant HR301-5 (line 1), random-sequence protein R301-5 (line 2), and wild-type RNase HI (line 3). The dashed line shows the calculated spectrum obtained by subtracting line 3 from line 1.

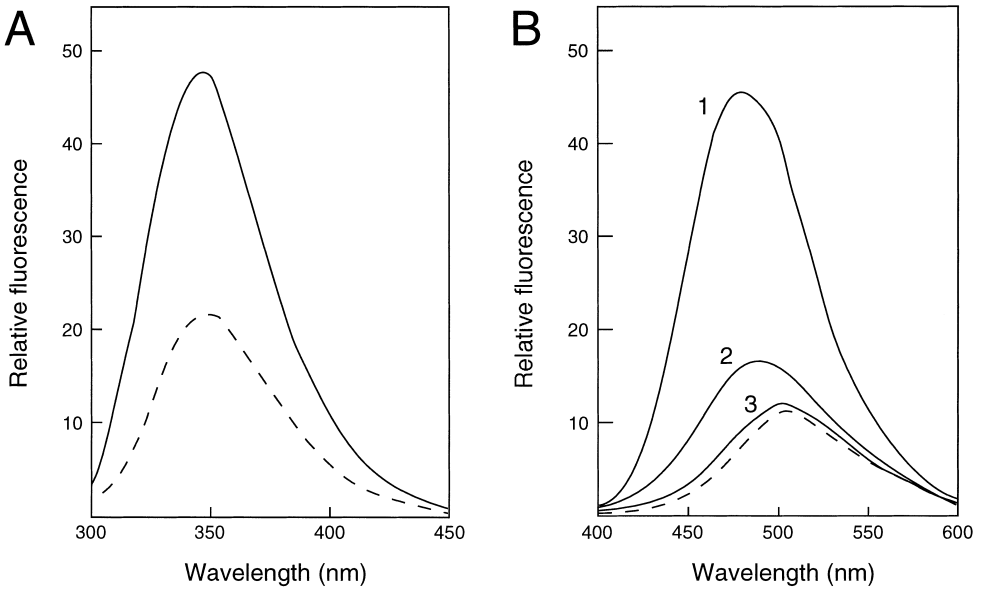

Fig. 5. A: Aromatic fluorescence emission spectra of random-sequence protein R301-5 were measured in 0 M urea (solid line) and in 6 M urea (dashed line). B: Fluorescence spectra of 50 μM ANS in the presence of 1 μM proteins; mutant RNase HI protein HR301-5 (line 1), random-sequence protein R301-5 (line 2), and wild-type RNase HI protein (line 3). The dashed line shows fluorescence of 50 μM ANS in the absence of protein.

tide is also supported by the results of ANS binding experiments. The fluorescence emission spectrum of ANS is known to be enhanced when the dye binds to hydrophobic regions of proteins [18]. As shown in Fig. 5B, ANS fluorescence increased in the presence of the R301-5 random-sequence protein (1.5-fold at maximum intensity), whereas only a small increase was observed for the scaffold RNase HI. Since a larger enhancement was observed for the HR301-5 fusion protein (4-fold), the hydrophobic cluster seems to be partially disrupted in the absence of the scaffold, on which the conformation of random-sequence polypeptide may be stabilized because of the proximity of the N and C-termini.

In summary, random-sequence proteins displayed on the surface of an enzyme were characterized in the present work. The scaffold enzyme activity was found to be correlated with the degree of order in the structure of the random-sequence insert. The CD and fluorescence analyses of a random-sequence protein detached from the scaffold indicated the presence of small but significant amounts of secondary and tertiary structures. It should thus be possible to establish a new method for selecting folded proteins from random protein libraries by monitoring the scaffold enzyme activity. Such an approach to the analysis of artificial evolutionary processes should provide new insights into protein design, folding and evolution.

References

[1] Mandecki, W. (1990) Protein Eng. 3, 221–226.
[2] LaBean, T.H. and Kauffman, S.A. (1993) Protein Sci. 2, 1249–1254.
[3] Davidson, A.R. and Sauer, R.T. (1994) Proc. Natl. Acad. Sci. USA 91, 2146–2150.
[4] 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.
[5] Shiba, K., Takahashi, Y. and Noda, T. (1997) Proc. Natl. Acad. Sci. USA 94, 3805–3810.
[6] Yamauchi, A., Yomo, T., Tanaka, F., Prijambada, I.D., Ohhashi, S., Yamamoto, K., Shima, Y., Ogasahara, K., Yutani, K., Kataoka, M. and Urabe, I. (1998) FEBS Lett. 421, 147–151.
[7] Nemoto, N., Miyamoto-Sato, E., Husimi, Y. and Yanagawa, H. (1997) FEBS Lett. 414, 405–408.
[8] Doi, N. and Yanagawa, H. (1998) Cell. Mol. Life Sci. 54 (in press).
[9] Doi, N., Itaya, M., Yomo, T., Tokura, S. and Yanagawa, H. (1997) FEBS Lett. 402, 177–180.
[10] Itaya, M. and Crouch, R.J. (1991) Mol. Gen. Genet. 277, 433–437.
[11] Johnson, B.H. and Hecht, M.H. (1994) BioTechnology 12, 1357–1360.
[12] Gill, S.C. and von Hippel, P.H. (1989) Anal. Biochem. 182, 319–326.
[13] Miller, S., Janin, J., Lesk, A.M. and Chothia, C. (1987) J. Mol. Biol. 196, 641–656.
[14] Yamasaki, K., Ogasahara, K., Yutani, K., Oobatake, M. and Kanaya, S. (1995) Biochemistry 34, 16552–16562.
[15] Berezovsky, I.N., Tumanyan, V.G. and Esipova, N.G. (1997) FEBS Lett. 418, 43–46.
[16] Chen, Y.-H., Yang, J.T. and Martinez, H.M. (1972) Biochemistry 11, 4120–4131.
[17] Teale, F.W. (1960) Biochem. J. 76, 381–388.
[18] Stryer, L. (1965) J. Mol. Biol. 13, 482–495.
[19] Itaya, M. (1990) Proc. Natl. Acad. Sci. USA 87, 8587–8591.