Insertion of foreign random sequences of 120 amino acid residues into an active enzyme

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

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

Received 13 December 1996

Abstract Random sequences of 120–130 amino acid residues were inserted into a surface loop region of *Escherichia coli* RNase HI. This library was screened and about 10% of the clones were found to retain RNase H activity. Subsequent random mutagenesis led to an increase in RNase H activity and solubility of the protein. The inserted regions were found not to contribute to the secondary structure of the mutant protein. The high frequency of insertion of flexible sequences and the increase in the protein's function by further mutagenesis simulate one of the events in protein evolution.

Key words: Insertional mutagenesis; Protein evolution

1. Introduction

Domain insertions, as well as point mutations are thought to have greatly contributed to protein evolution. However, the effects of amino acid insertions on the structure and function of proteins have been studied much less than for point mutations. Several studies indicated that tolerant sites to a few amino acid insertions are observed almost exclusively in loop regions [1–4] and rarely within α-helices [5,6]. So far, the length of insertions at the tolerant sites of water-soluble enzyme has been studied only up to 14–16 residues [3,4]. Thus, it will be more interesting to investigate the process of inserting longer sequences into an active protein.

To study the frequency of the insertion of protein domains into an active enzyme, we constructed a library of mutant proteins and estimated the percent of random sequences with more than 100 amino acid residues which could be inserted into the middle of an enzyme, *E. coli* RNase HI. Furthermore, we made one of the obtained mutants evolve to increase RNase H activity and solubility. The implication of these results for domain insertion in the course of evolution is discussed.

2. Materials and methods

2.1. Library construction and screening

Site-directed insertional mutagenesis was carried out on plasmid pSK760 [7] (from S. Kanaya) bearing the *rnhA* gene whose expression is regulated by its cognate promoter. *Bpm*I, the insertion site (shown in Fig. 1A), is a unique site in the *rnhA* gene but not unique in whole pSK760. Thus, the plasmid was linearized with *Bpm*I in buffer containing ethidium bromide at appropriate concentration for partial digestion [8]. A kanamycin-resistance gene cassette of pBEST501 [9] was amplified by PCR with two primers, one containing a *Hin*dIII site and the other a *Kpn*I site. The PCR product was directly ligated into the blunt-ended *Bpm*I fragment of pSK760. *E. coli* JM109 were transformed with the ligated DNA, kanamycin resistant colonies were selected, and the recovered plasmid DNAs were analyzed by restriction mapping. From the plasmid of interest, the kanamycin-resistance gene cassette was removed by digestion with *Hin*dIII and *Kpn*I. The backbone vector was purified by agarose gel electrophoresis. As shown in Fig. 1B, 1 unit (about 120 bp) and 2 units (about 240 bp) of DNA fragments consisting of random sequences were amplified from the plasmid mixture 'R1MIX' [10] by PCR with P1 and P5 primers (Toyobo). These fragments were digested with pairs of *Hin*dIII-*Xho*I or *Xho*I-*Kpn*I, purified by 10% polyacrylamide gel electrophoresis and ligated into the *Hin*dIII-*Kpn*I backbone vector fragment, yielding a random DNA library encoding mutants of RNase HI in which the 3 units (about 120 amino acids) of random proteins are inserted. Mutant RNase HI genes were screened from this library by their ability to suppress the temperature-sensitive growth phenotype of *E. coli* MIC3037 (*rnhA-339*::*cat recC271*) strain [11]. For the mutant genes thus obtained, the DNA sequences of the inserted random regions were determined by dideoxy sequencing [12].

2.2. Random PCR mutagenesis

Subsequent random mutagenesis of inserted amino acids was conducted using error-prone PCR [13]. The amplified DNA fragments were digested with *Kpn*I and religated into the same sites of the mutant *rnhA* gene. The ligated DNAs were introduced into *E. coli* MIC2028 (*rnhA-339*::*cat zib*::*Tn10 dnaA508 proA2*) for screening RNase HI mutant with increased activity. Strain MIC2028 is an equivalent strain to MIC1041 [11] except the genetic background is strain W3110.

2.3. Overproduction and purification

Mutant RNase HI proteins were overexpressed using the T7 expression system [14] containing the carboxyl-terminal hexahistidine sequence to permit affinity purification with nickel-NTA agarose resin (Qiagen) [15]. By the procedure recommended, the proteins were purified to >95% homogeneity with yields of 2–4 mg per 1 culture broth. The purity of the mutant proteins was confirmed by SDS-PAGE [16]. Since the host *E. coli* MIC2090 (*rnhA-339*::*cat recB270* λ*(DE3)*) has a mutation in the genomic *rnh* gene (M. Itaya, unpublished), contamination of RNase H activity from the host cells would not have been significant.

2.4. Enzymatic activity

RNase H activity was determined by three assays. First, in vivo assay was performed using the two *rnh* mutants of *E. coli* MIC3037 and MIC2028. Strain MIC3037 does not form colonies at non-permissive temperature 42°C: the ratio of colonies at 42°C to those at 30°C is <0.1% as the background level. This temperature sensitive growth phenotype is suppressed when more than 1% of wild-type RNase H activity is supplied [11]. On the other hand, strain MIC2028 lost the ability to form colonies at 42°C in the presence of >10% wild-type RNase H activity [11]. Plasmids having the mutant gene were introduced into these strains and transformants were selected on LB plates containing ampicillin (100 μg/ml) at 42°C. In the second assay, renaturation gel assay was performed for crude extracts and purified mutant proteins as previously described [17] although strain MIC2067 (M. Itaya, unpublished) served as the host for crude assay. Both RNase HI and RNase HII genes of strain MIC2067 are

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

Abbreviations: PCR, polymerase chain reaction; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; CD, circular dichroism

PII S0014-5793(96)01522-0

mutated, providing low RNase H activity levels. In the third assay, the relative activity of purified proteins was determined based on the radioactivity of the acid-soluble digestion product from a DNA/RNA hybrid poly([α-^{32}P]rA)·poly(dT) as described in [17,18]. The assay conditions were the same as those for wild-type RNase HI [18], thus making possible a direct comparison of specific RNase H activity.

2.5. CD spectra

CD spectra were measured on J-600 spectropolarimeter (Japan Spectroscopic) at 5°C. The protein concentrations of samples ranged from 2 to 15 μM, and the light path length used was 1 or 10 mm. The results were expressed as mean residue molar ellipticity, [θ].

3. Results and discussion

3.1. Construction and screening of the insertional mutant library

E. coli RNase HI (EC 3.1.26.4; 155 amino acid residues) is an endonuclease that specifically degrades the RNA moiety of DNA/RNA hybrids. *E. coli* RNase HI was used as a host protein for following three advantages: The three-dimensional structure of *E. coli* RNase HI is well-defined [19–21]; information on active sites [22], substrate binding sites [23], and backbone flexibility [24] of *E. coli* RNase HI is available; and a screening system of RNase H activity has been established [11,17].

Fig. 1A shows a ribbon diagram of the crystal structure of *E. coli* RNase HI, in which the His-62 site was chosen as the target for the insertion of long polypeptides. The His-62 site is located at the turn between α-helix I and β-strand D, in which large internal motion was observed on NMR relaxation analysis [24]. The His-62 site is situated just behind the active center of RNase HI, so that the long inserted residues would not directly inhibit RNase H activity. A preliminary experiment actually indicated that the His-62 site is tolerant to 2–18 amino acid insertions (N. Doi, M. Itaya, and H. Yanagawa, unpublished data).

Using the vector and general strategy shown in Fig. 1B, we constructed a library in which the random DNA sequences encoding 120–130 amino acid residues were inserted into the *Bpm*I site of *E. coli* RNase HI gene. This random sequence [10] contains all 20 naturally occurring amino acids and no stop codon in any of the six frames. Screening of the clones by in vivo assay using strain MIC3037 (see Section 2) showed that about 10% of the insertion mutants in the library retained > 1% of wild-type RNase H activity. Eight positive clones were randomly selected and the plasmids were isolated. The nucleotide sequences of the inserts were determined by DNA sequencing [12]. They certainly consisted of the random sequences that would be expected from the design [10], and several fixed residues at both ends of inserted region and the cassette junctions (Fig. 2).

Table 1
Characterization of insertion mutants of RNase HI

Mutant	RNase H activity		
	In vivo[a]	In gel[b]	In vitro[c]
HR301-5	+++	+++	3.0%
HR301	–	++	n.o.
HR302	+	++	1.0%
HR303	+	++	0.3%
HR304	+++	++	0.2%
HR305	+++	+	0.4%
HR306	–	+++	n.o.
HR307	+++	+	
HR308	+++	+	

[a]Total RNase H activity in vivo estimated based on the temperature-sensitive growth of *E. coli* MIC2028 at 42°C. See also Section 2. +++, no growth (> 10% activity); +, small colony (< 10%); –, large colony (≪ 10%).
[b]Renaturation gel assay performed as described in Section 2. +++, strong band; ++, medium band; +, weak band.
[c]Relative activity determined from radioactivity of acid-soluble digestion product from [^{32}P]poly(rA)·poly(dT). The wild-type specific activity (100%) was 150 units/μg [17]. One unit is defined as generating 1 nmol of acid-soluble nucleotide per 15 min at 37°C. n.o., no overexpression.

3.2. Subsequent random mutagenesis of inserted sequences

The eight mutant clones were analyzed by in vivo assay using strain MIC2028 (see Section 2). As shown in Table 1, positive results were found with four mutants, HR304, HR305, HR307 and HR308; indicating that the mutants retained total RNase H activity in vivo greater than 10% of that of the wild-type. For the rest of four clones, subsequent ran-

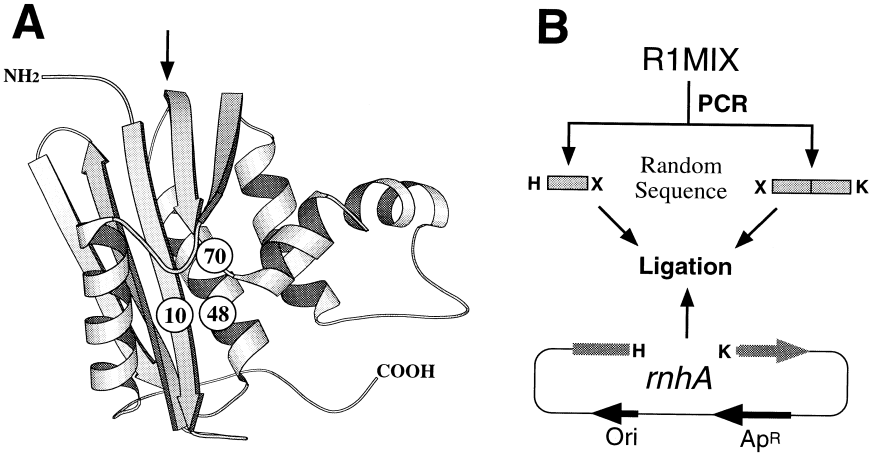

Fig. 1. (A) Ribbon diagram of the crystal structure of *E. coli* RNase HI. The insertion site is marked with an arrow and the three open circles indicate active site residues. The crystal structure of *E. coli* RNase HI was determined by Katayanagi et al. [19,20]. (B) Schematic diagram of the insertions of random sequences into *rnhA* gene. 'R1MIX' [10] is the source of the random sequences. Restriction sites are abbreviated as: H, *Hin*dIII; K, *Kpn*I; X, *Xho*I.

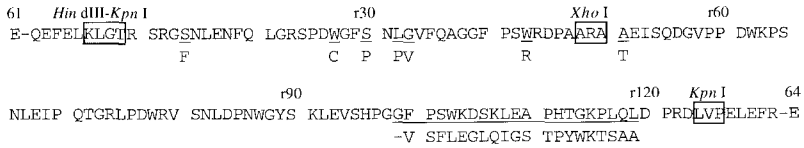

Fig. 2. Predicted amino acid sequence of insertional random region of one of the isolated mutants, HR301. The sequence has been inserted between Glu-61 and Glu-64 in RNase HI. The restriction sites of cassette junctions are boxed. A mutant of HR301, HR301-5, caused 9 base substitution and 3 base deletion mutations, resulting in 7 amino acid substitutions and a frame-shift mutation between r99 and r119. Residues with mutations are underlined, and substituted residues appear below.

dom mutations were introduced into the long inserted regions to isolate mutant with higher RNase H activity. After mutagenesis, several positive clones were obtained from each of three mutants of HR301, HR302, and HR306. One clone HR301-5, from HR301, was further characterized and the amino acid sequence is shown in Fig. 2.

3.3. Overproduction and purification

All eight mutant genes were inserted into a T7 expression vector [14] containing a carboxyl-terminal hexahistidine sequence to permit affinity purification [15]. In this expression system, HR301 and HR306 showed no overexpression. The remaining six proteins were overexpressed but remained insoluble. On the other hand, HR301-5 was overexpressed and soluble. The four insoluble proteins, HR302, HR303, HR304, and HR305, and one soluble protein, HR301-5 were purified.

3.4. Enzymatic activity

Renaturation gel assay of the crude extracts of all eight clones and all five purified proteins gave the white bands indicating RNase H activity (Fig. 3). Their expected sizes (~32 kDa) as compared with the band of the wild-type RNase HI (~18 kDa) indicate that the insertion mutant of RNase HI identified on the gel certainly possessed RNase H activity and was not degraded by cellular proteases. Activities of the purified protein with the same molecular mass confirmed that no degradation occurred during the purification procedure. In addition, HR301-5 gave a more intense band than the parent HR301. Relative activities of five purified proteins ranged from 0.2 to 3% wild-type specific activity (Table 1).

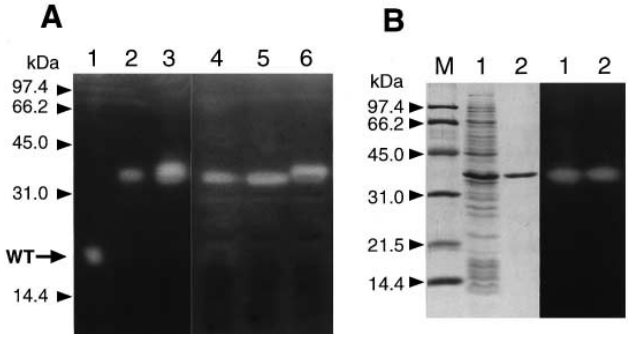

Fig. 3. Renaturation gel assay showing RNase H activity as white bands. On the renaturation gel assay [17], whole cell extracts (A) or purified protein (B) were loaded onto SDS-PAGE, in which the 15% polyacrylamide separating gel contains DNA/RNA hybrid poly([α-^{32}P]rA)·poly(dT). After electrophoresis, the gel was washed and incubated in RNase H assay buffer. The enzymes were renatured by removal of SDS and the degraded ^{32}P-labeled oligoriboadenylates were diffused from the gel. Exposure of the gel to X-ray film led to the appearance of white bands at the site of RNase H activity. (A) Lanes: 1 (control), whole cell extracts of *E. coli* JM109; 2–6, whole cell extracts of *E. coli* MIC2067 having the mutant gene of HR301 (lane 2), HR301-5 (lane 3), HR302 (lane 4), HR303 (lane 5), and HR304 (lane 6). No band was seen for whole cell extracts of *E. coli* MIC2067 (data not shown). (B) Overexpressed HR301-5 protein. (Left) Coomassie brilliant blue staining. (Right) Renaturation gel assay showing RNase H activity as white bands. Lanes: 1, whole cell extract; 2, purified protein.

3.5. Secondary structure

The CD spectra of mutant proteins indicated the presence of 11–16% helical structure (Fig. 4) estimated from the mean residue ellipticity at 222 nm [25]. The contents of α-helix were somewhat lower than that of the contribution from the region of RNase HI. This result indicates that the secondary structure of the mutant proteins mostly attributable to that of RNase HI, and the inserted region adopts a random coil structure. The flexibility of the inserted random sequences accounts for the high frequency of the insertion.

In the present study, we found that about 10% of random sequences of 120–130 residues could be inserted into a surface loop region of an enzyme, *E. coli* RNase HI, with a decreased but significant level of RNase H activity for further selection. This result provided experimental evidence that insertion of long sequences such as more than 100 amino acid residues is not a rare event in the course of protein evolution. The high frequency of insertion is attributed to the flexibility of the inserted sequence. It was shown that a mutant of *E. coli* with lower glutamine synthetase activity than that of the wild type is not always excluded from the population in a chemostat [26]. This fact suggests that a mutant of an enzyme with a long insertion, though decreasing its activity, is not always eliminated but has a chance of further evolution. Hence, at the viewpoint of evolution, we propose that the active proteins may take in flexible sequences which slowly

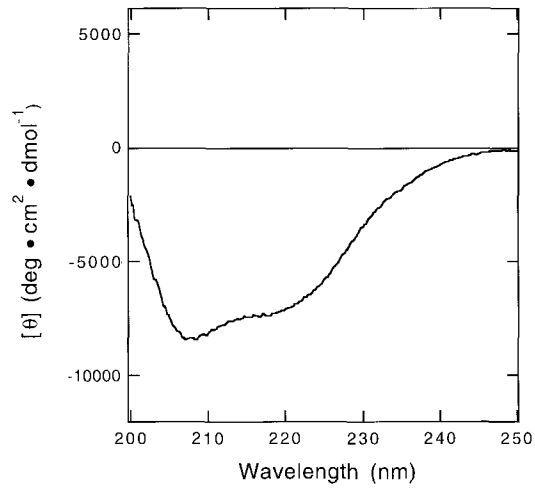

Fig. 4. CD spectrum of HR301-5 (4 μM protein, 5°C). The spectrum was measured as described under Section 2.

fold in during the course of evolution as the activity or function of the newly folded protein is slowly being optimized through spontaneous mutation in nature. In other words, the domain in the present active protein may have arisen from the folding of the flexible sequences in the course of optimization of the function of the protein. Indeed, random mutagenesis led to the increase in activity and solubility of the mutant *E. coli* RNase HI HR301.

Acknowledgements: We thank Dr. S. Kanaya for plasmid pSK760, and Dr. K. Katayanagi for illustration of the crystal structure of *E. coli* RNase HI.

References

[1] Freimuth, P.I., Taylor, J.W. and Kaiser, E.T. (1990) J. Biol. Chem. 265, 896–901.
[2] Sondek, J. and Shortle, D. (1992) Proteins 13, 132–140.
[3] Starzyk, R.M., Burbaum, J.J. and Schimmel, P. (1989) Biochemistry 28, 8479–8484.
[4] Ladant, D., Glaser, P. and Ullmann, A. (1992) J. Biol. Chem. 267, 2244–2250.
[5] Kavanaugh, J.S., Moo-Penn, W.F. and Arnone, A. (1993) Biochemistry 32, 2509–2513.
[6] Heinz, D.W., Baase, W.A., Zhang, X.-J., Blaber, M., Dahlquist, F.W. and Matthews, B.W. (1994) J. Mol. Biol. 236, 869–886.
[7] Kanaya, S. and Crouch, R.J. (1983) J. Biol. Chem. 258, 1276–1281.
[8] Parker, R.C., Watson, R.M. and Vinograd, J. (1977) Proc. Natl. Acad. Sci. USA 74, 851–855.
[9] Itaya, M., Kondo, K. and Tanaka, T. (1989) Nucleic Acids Res. 17, 4410.
[10] 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.
[11] Itaya, M. and Crouch, R.J. (1991) Mol. Gen. Genet. 277, 433–437.
[12] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[13] Leung, D.W., Chen, E.Y. and Goeddel, D.V. (1989) Technique 1, 11–15.
[14] Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Methods Enzymol. 185, 60–89.
[15] Hochuli, E., Döbeli, H. and Schacher, A. (1987) J. Chromatogr. 411, 177–184.
[16] Laemmli, U.K. (1970) Nature 227, 680–685.
[17] Itaya, M. (1990) Proc. Natl. Acad. Sci. USA 87, 8587–8591.
[18] Carl, P.L., Bloom, L. and Crouch, R.J. (1980) J. Bacteriol. 144, 28–35.
[19] Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T. and Morikawa, K. (1990) Nature 347, 306–309.
[20] Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Nakamura, H., Ikehara, M., Matsuzaki, T. and Morikawa, K. (1992) J. Mol. Biol. 223, 1029–1052.
[21] Yang, W., Hendrickson, W.A., Crouch, R.J. and Satow, Y. (1990) Science 249, 1398–1405.
[22] Kanaya, S., Kohara, A., Miura, Y., Sekiguchi, A., Iwai, S., Inoue, H., Ohtsuka, E. and Ikehara, M. (1990) J. Biol. Chem. 265, 4615–4621.
[23] Kanaya, S., Katsuda-Nakai, C. and Ikehara, M. (1991) J. Biol. Chem. 266, 11621–11627.
[24] Yamasaki, K., Saito, M., Oobatake, M. and Kanaya, S. (1995) Biochemistry 34, 6587–6601.
[25] Chen, Y.-H., Yang, J.T. and Martinez, H.M. (1972) Biochemistry 11, 4120–4131.
[26] Xu, W., Kashiwagi, A., Yomo, T. and Urabe, I. (1996) Res. Popul. Ecol. 38, in press.