

Solution structure of a tmRNA-binding protein, SmpB, from *Thermus thermophilus*

Tatsuhiko Someya^{a,1}, Nobukazu Nameki^{a,1,2}, Haruko Hosoi^b, Sakura Suzuki^b, Hideki Hatanaka^b, Michiko Fujii^a, Takaho Terada^b, Mikako Shirouzu^b, Yorinao Inoue^c, Takehiko Shibata^{c,d}, Seiki Kuramitsu^{c,e}, Shigeyuki Yokoyama^{b,c,f,g}, Gota Kawai^{a,b,*}

^aDepartment of Industrial Chemistry, Faculty of Engineering, Chiba Institute of Technology, Chiba 275-0016, Japan

^bProtein Research Group, RIKEN Genomic Sciences Center, Yokohama 230-0045, Japan

^cStructurome Project, RIKEN Harima Institute at SPring-8, Hyogo 679-5148, Japan

^dCellular and Molecular Biology Laboratory, RIKEN, Saitama 351-0198, Japan

^eDepartment of Biology, Graduate School of Science, Osaka University, Osaka 560-0043, Japan

^fCellular Signaling Laboratory, RIKEN Harima Institute at SPring-8, Hyogo 679-5148, Japan

^gDepartment of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan

Received 3 December 2002; accepted 9 December 2002

First published online 2 January 2003

Edited by Thomas L. James

Abstract Small protein B (SmpB) is required for *trans*-translation, binding specifically to tmRNA. We show here the solution structure of SmpB from an extremely thermophilic bacterium, *Thermus thermophilus* HB8, determined by heteronuclear nuclear magnetic resonance methods. The core of the protein consists of an antiparallel β -barrel twisted up from eight β -strands, each end of which is capped with the second or third helix, and the first helix is located beside the barrel. Its C-terminal sequence (20 residues), which is rich in basic residues, shows a poorly structured form, as often seen in isolated ribosomal proteins. The results are discussed in relation to the oligonucleotide binding fold.

© 2002 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.

Key words: Nuclear magnetic resonance; Small protein B; tmRNA; *Trans*-translation; RNA-binding protein; Extended oligonucleotide binding fold

1. Introduction

In bacteria, there is a universal quality control step in protein synthesis mediated by tmRNA possessing both tRNA and mRNA properties (reviewed in [1–3]). The unique translational process is called *trans*-translation, in which a problematic mRNA staying on a ribosome is replaced by the internal coding region of tmRNA including a stop codon, and the normal terminal process occurs so that the stalled ribosome can be rescued.

During the *trans*-translation process when tmRNA acts as tRNA and then as mRNA, various protein factors collaborate with the RNA. After tmRNA is aminoacylated with alanine by alanyl-tRNA synthetase, the aminoacyl-tmRNA can form a ternary complex with elongation factor Tu (EF-Tu) and GTP [4,5]. A unique basic protein, small protein B (SmpB), was identified as an essential component of the *trans*-translation system in *Escherichia coli* [6]. Deletion of the *smpB* gene results in phenotypes similar to those observed in tmRNA-defective cells, including a variety of phage development defects and the absence of tagged proteins translated from defective mRNAs. Purified SmpB binds to tmRNA specifically and with high affinity, and is required for stable association of tmRNA with ribosomes *in vivo*. Several roles of the binding of SmpB to tmRNA have been revealed by biochemical *in vitro* experiments [7,8]. The SmpB binding enhances the aminoacylation of tmRNA, and promotes the binding of tmRNA to ribosomes, preventing tmRNA from degradation in the cell. The nuclear magnetic resonance (NMR) structure of SmpB from a hyperthermophilic marine bacterium, *Aquifex aeolicus*, has been reported, which lacks the last 23 residues (1–133 residues of the full-length 156 residues) [9]. Because the *A. aeolicus* SmpB forms an essentially new fold including an antiparallel β -barrel with three helices packed outside, it is important to elucidate the solution structure of SmpB from a different bacterium to clarify the generality and importance of the structural motif.

In the present study, we determined the solution structure of SmpB from an extremely thermophilic bacterium, *Thermus thermophilus* (HB8 strain), by heteronuclear NMR methods. This RNA-binding protein is composed of 144 residues, of which basic residues, Arg and Lys, account for 22%. To obtain the isotopic labeled sample, we used a refined efficient labeling method in which higher expression yields can be achieved at lower isotope costs [10]. Although the two solution structures from *A. aeolicus* and *T. thermophilus* were similar to each other, several differences were found in secondary structure elements. Comparison of the two structures will show more detailed characteristic features of the isolated SmpB protein.

*Corresponding author. Fax: (81)-47-478 0425.

E-mail address: gkawai@ic.it-chiba.ac.jp (G. Kawai).

¹ These authors equally contributed to this work.

² Present address: Protein Research Group, RIKEN Genomic Sciences Center, Yokohama 230-0045, Japan.

Abbreviations: EF-Tu, elongation factor Tu; SmpB, small protein B; rmsd, root mean square deviation; OB-fold, oligonucleotide binding fold.

2. Materials and methods

2.1. Overexpression and purification of SmpB from *T. thermophilus*

The coding sequence of SmpB from *T. thermophilus* HB8 was amplified by PCR to add a *Nde*I site at the 5' end and a *Bam*HI site at the 3' end. This DNA fragment was cloned into the *Nde*I–*Bam*HI-digested pET-11b expression vector (Novagen, USA).

Expression of ¹⁵N- and ¹³C/¹⁵N-labeled proteins was performed as described by Marley et al. [10]. *E. coli* BL21-CodonPlus[™] (DE3)-RP cells (Stratagene, USA) that carry extra copies of the *argU* and *proL* tRNA genes were transformed with the recombinant plasmid, and the transformant was grown at 37°C in 300 ml of LB medium containing 30 µg/ml ampicillin and 34 µg/ml chloramphenicol. When optical cell densities at 600 nm reached 0.7, the cells were harvested by centrifugation at 4000 × *g* at 4°C for 20 min. The cells were then washed and precipitated using an M9 salt solution, excluding all nitrogen and carbon sources. The cells were resuspended in 1000 ml of isotopically labeled M9 medium including 1 mg/ml ¹⁵NH₄Cl and 2 mg/ml [¹³C]Glucose (Nippon Sanso, Japan), and incubated at 37°C for 1 h to allow the recovery of growth and clearance of unlabeled metabolites. Protein expression was then induced for 4 h in the presence of 1 mM IPTG. The cells were harvested by centrifugation at 4000 × *g* at 4°C for 20 min. Wet cells were dissolved in sonication buffer (50 mM HEPES pH 7.5, 300 mM NaCl, 2 mM dithiothreitol, 2 mM EDTA), and broken on ice by sonication, and centrifuged at 4000 × *g* at 4°C for 30 min. To remove most of the proteins from *E. coli*, supernatant fluid was incubated at 65°C for 20 min, and centrifuged at 14000 × *g* at 4°C for 20 min. The supernatant was dialyzed with 50 mM phosphate buffer (pH 7.4), and the dialysate was purified on a cation exchange column (Hitrap[™] SP Sepharose[™] HP, Amersham Pharmacia Biotech, Sweden) and eluted with a linear gradient of NaCl from 0 to 700 mM in 120 ml of 50 mM phosphate buffer (pH 6.4). The elution pattern was monitored by 17% SDS-PAGE. The SmpB-containing fractions were collected and concentrated by ultrafiltration with a Centricon-10 (Millipore, USA), and the sample was loaded onto Superdex 75 column (26 mm i.d. × 600 mm; Pharmacia Biotech, Sweden) and eluted with 50 mM phosphate buffer (pH 6.5) containing 100 mM NaCl. The SmpB-containing fractions were pooled and concentrated to 2 ml by a Centricon-10.

2.2. NMR spectroscopy

Buffer exchange was done with a solution containing 1 mM HCl and 100 mM NaCl using a Centricon-10, and the [¹⁵N]SmpB and [¹³C/¹⁵N]SmpB samples were concentrated to 2 mM and 0.6 mM in 250 µl, respectively. The final pH of the sample was 3.6 after addition of D₂O to 5%. All NMR experiments were performed at 318 K. NMR experiments for [¹⁵N]SmpB were recorded on a Bruker DRX-600 NMR spectrometer. NMR experiments for [¹³C/¹⁵N]SmpB were recorded on Varian Unity Inova 600 MHz and 800 MHz spectrometers. Main chain assignments were obtained with three-dimensional (3D) ¹⁵N-edited NOESY-HSQC (mixing times: 75 and 200 ms), 3D ¹⁵N-edited HOHAHA-HSQC (mixing time: 50 ms), 3D HNCA, 3D HN(CO)CA, 3D HNCACB, 3D CBCA(CO)NH, 3D HNCO and 3D (HCA)CO(CA)NNH. Assignments of side chains were obtained with 3D HC(C)H-TOCSY (mixing time: 15.6 ms), 3D C(CO)NH, 3D H(CCO)NH and two-dimensional (2D) TOCSY (mixing time: 75 ms). NOEs for distance constraints were derived from 3D ¹⁵N-edited NOESY-HSQC (mixing time: 75 ms) and 3D ¹³C-edited NOESY (mixing time: 75 ms).

2.3. Structure calculation

NOESY peak assignments were obtained using a laboratory-made program for semi-automated NOESY cycles, but for the initial stage we referred to the homology models made from the NMR structures of SmpB from *A. aeolicus*. The upper limits of distance constraints were calculated as $kI^{-1/6}$, where I is the peak intensity and k is a constant adjusted in each NOESY spectrum, and relaxed by 0.5 Å considering mobility. Lower limits of distance constraints were all 1.8 Å. Dihedral angle restraints were produced by the program TALOS [11].

The structures were calculated with the program X-PLOR ver. 3.1 [12]. We did not include the C-terminal 21 residues in the structure calculations because they lacked any chemical shift assignments. Initial coordinates were generated using random ϕ and ψ angles, whereas peptide bonds and side chains took an extended conformation. The

macroprogram sa.inp in X-PLOR ver. 2.1 was used to carry out simulated annealing calculation. The target function minimized during simulated annealing comprised only potential terms for covalent geometry, experimental (distance/dihedral) restraints and van der Waals non-bonded repulsion. No hydrogen bonding, electrostatic or 6–12 Lennard–Jones potential terms were present in the target function.

2.4. Protein Data Bank (PDB) accession number

The atomic coordinates of the 10 converged conformers and a minimized average conformer have been deposited at the RCSB PDB under the PDB accession code 1J1H.

3. Results

The SmpB protein from *T. thermophilus* failed to be efficiently overexpressed in conventional *E. coli* host, probably due to the codon usage bias of the gene [13]. In the case of GC-rich genomes including *T. thermophilus* genomes [14], rare arginine codons (AGG/AGA) and proline codon (CCC) would most frequently affect bacterial expression of genes. In fact, among 144 codons of the *smpB* gene, the numbers of AGG/AGA and CCC codons are 0/5 and 3, respectively. Thus, an *E. coli* strain that carries extra copies of the *argU* and *proL* tRNA genes was used to rescue expression of genes restricted by either AGG/AGA or CCC codons. Consequently, we succeeded in overexpression of the full-length SmpB protein using this strain with LB medium. To obtain SmpB labeled with stable isotopes, we then tried to use the expression system with M9 minimal medium including 1 mg/ml ¹⁵NH₄Cl and 1.4 mg/ml [¹³C]glucose, but the labeled protein was hardly overexpressed. This failure would be due not only to the low concentrations of the carbon/nitrogen sources, but also to relatively low efficiencies of overexpression of SmpB even with rich medium. Thus, we adopted a refined isotopic labeling method, which permits high yields of labeled protein while significantly reducing the cost in both isotopes and production time [10]. The production method generates cell mass using unlabeled rich medium followed by exchange into a small volume of labeled media at high cell density. After a short period for growth recovery and unlabeled metabolite clearance, protein expression is induced by addition of IPTG. Accordingly, we obtained enough amounts of SmpB samples for NMR measurements using low concentrations of isotope sources (see Section 2).

To verify whether this protein could bind to tmRNA, the interaction was analyzed by the optical biosensor assay using an IAsys[™] instrument based on the resonant mirror device (Affinity Sensors, Cambridge, UK) [15]. The results using a tmRNA transcript from *T. thermophilus* show that the labeled SmpB binds to the tmRNA as efficiently as the non-labeled SmpB with a K_d of 20 nM (data not shown), which agrees with the K_d value estimated by RNA gel mobility shift assay in the *E. coli* system [6]. These results confirm that the SmpB overexpressed in *E. coli* is functional.

2D NMR data indicate that many of the resonances are well dispersed, and chemical shift assignments of the main chain were performed from residue 4 to 123 (out of 144). In contrast, a fraction of resonances focuses within a narrow region, suggesting a random coil (Fig. 1), and these highly overlapped resonances made it difficult to complete chemical shift assignments. Similar resonance overlap was observed in full-length SmpB from *A. aeolicus* [9]. In the case of *A. aeolicus*, comparison of resonances between full-length and truncated SmpB proteins shows that such overlapped resonances

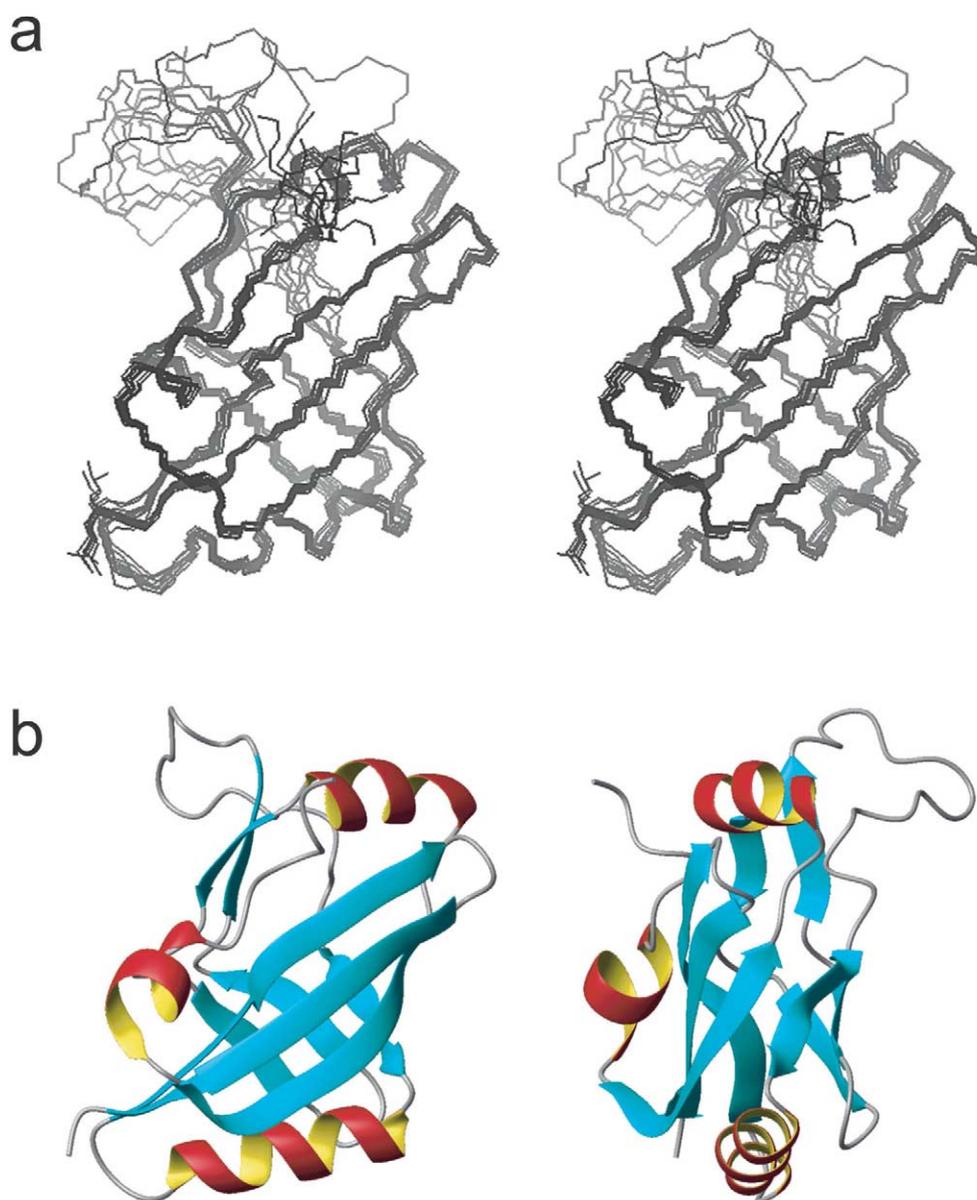


Fig. 2. Solution structures of SmpB from *T. thermophilus*. a: Stereo view of the final ensemble 10 structures superimposed on ordered regions (residues 8–62 and 79–122). Residues 124–144 are not depicted. b: Ribbon diagrams of the average structure of SmpB. The two views differ by a 90° rotation. The diagrams were created using the MOLMOL program [22].

are attributed to the C-terminal residues, and the corresponding region appears structurally independent of the core of SmpB. This would be the case with SmpB from *T. thermophilus*, because the core structure of the full-length protein from *T. thermophilus* is quite similar to the structure of truncated SmpB from *A. aeolicus* as described below.

The distance and torsion angle constraints as well as some characteristics of the obtained SmpB structure are summarized in Table 1. Totally, 287 intraresidual, 171 sequential, 58 medium-range, and 134 long-range interproton distance constraints were used in calculation. A final set of 10 converged structures was selected from 30 calculations on the basis of agreement with the experimental data and van der Waals energy. A mean structure was obtained by averaging the coordinates of the structures that were superimposed in advance to the best converged structure and then minimized

under the constraints. The root mean square deviations (rmsd) from this mean structure were 0.54 ± 0.08 Å for backbone (N, C $^{\alpha}$, C) atoms and 1.21 ± 0.13 Å for all heavy atoms at the well ordered region (residues 8–62 and 79–122).

The results show that the core of SmpB consists of an antiparallel β -barrel containing eight β -strands, and three helices (Fig. 2). The barrel is thus composed of three-stranded antiparallel sheet of β_2 , β_6 and β_7 (residues 16–24, 98–108 and 112–121, respectively) and continuous three-stranded antiparallel sheet of β_3' – β_3 , β_4' – β_4 and β_5 (residues 38–40, 45–48, 53–56, 59–62 and 78–81, respectively). Both NOEs between strands β_3 and β_4 , and NOEs between strands β_3' and β_4' were observed, which yield two putative long strands. Strands β_2 and β_6 include bulges composed of residues 17–18 and 103–104, respectively, which form the twist of the β -sheet. These residues are well conserved, and the bulges are also

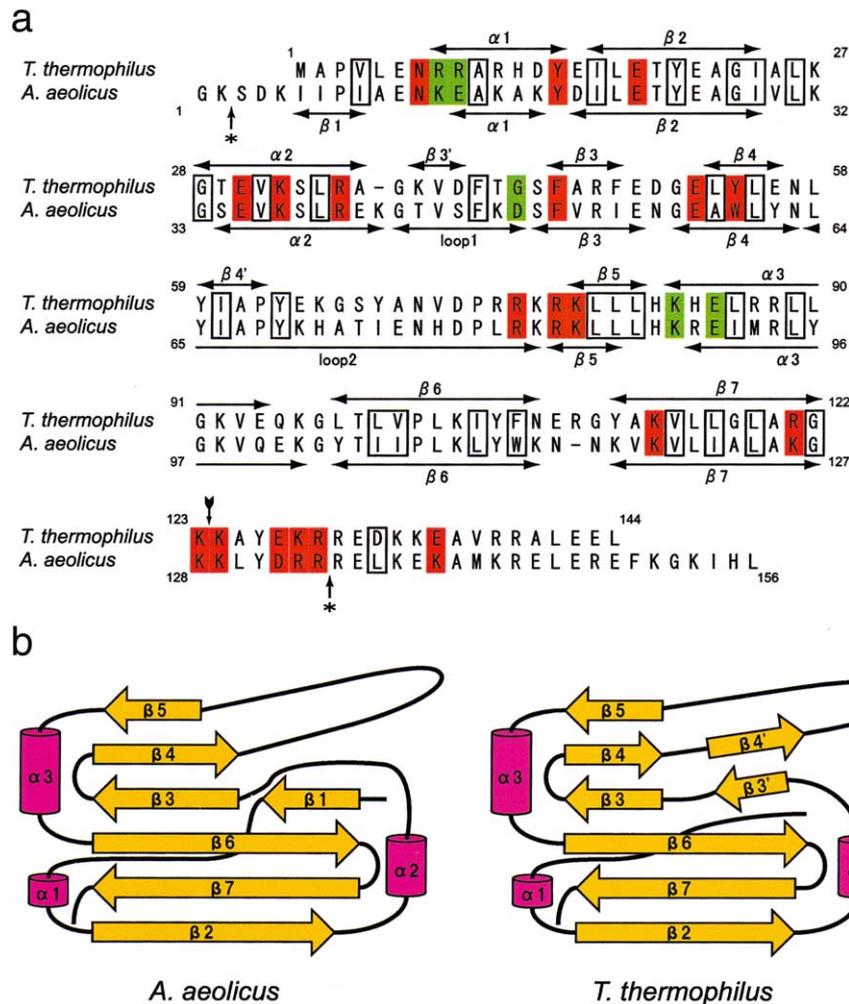


Fig. 3. a: Amino acid sequence alignment for the SmpB proteins from *T. thermophilus* and *A. aeolicus*. According to Dong et al. [9], well conserved hydrophobic, glycine, and proline residues are boxed. Other conserved residues (>90% and 80–90%) are indicated by red and green, respectively. b: The secondary structure elements of SmpB. The secondary structure elements of *A. aeolicus* SmpB are derived from the PDB entry 1K8H. The secondary structure elements are labeled according to the labeling scheme of the structure from *A. aeolicus* so that unnecessary confusion can be avoided.

observed in SmpB from *A. aeolicus* [9]. Both ends of the barrel are capped with two helices $\alpha 2$ and $\alpha 3$ (residues 28–36 and 83–94, respectively), which are almost parallel with each other. Helix $\alpha 1$ (residues 8–14) precedes the barrel. In addition to the C-terminal tail, the loop between strands $\beta 4$ and $\beta 5$ also appears to be flexible.

4. Discussion

We determined the solution structure of SmpB from *T. thermophilus*. The core of the SmpB structure consists of three helices and eight β -strands with which a β -barrel is formed (Fig. 2). Well conserved residues that are most likely important for structural reasons are positioned in almost the same manner in both structures from *T. thermophilus* and *A. aeolicus*, except for the following three conserved residues. In *T. thermophilus*, the side chains of F41 and L58 are directed toward the inside, whereas that of N57 is toward the outside. Interestingly, the reverse state occurs in the NMR structure of SmpB from *A. aeolicus*, although the positions 41 and 58, and position 57 are always occupied by hydrophobic and hydrophilic residues, respectively, in the protein

sequences that have so far been known [9]. It thus appears that strands $\beta 3'$ and $\beta 4'$ can be formed in SmpB from *T. thermophilus* instead of a loop in the counterpart (Fig. 3). The linker regions between strands $\beta 3'$ and $\beta 3$ and between strands $\beta 4$ and $\beta 4'$ differ in length (Fig. 3), and the NOEs between the linker regions were not observed (data not shown). If strands $\beta 3'$ and $\beta 3$, and strands $\beta 4$ and $\beta 4'$ are regarded as one strand, respectively, the barrel is twisted up from six β -strands, as is shown in *A. aeolicus*. A β -strand followed by the first helix $\alpha 1$ in SmpB from *A. aeolicus*, which is not an integral component of the barrel, is absent in this NMR structure. However, since assignment of the first three residues could not be done, it is not known whether the β -strand exists in SmpB from *T. thermophilus*. In contrast to the core, the C-terminal 21 residues could not be assigned, and are apparently unstructured or flexible, as also observed in *A. aeolicus*. These residues are relatively rich in basic residues, but not well conserved in length and sequence (Fig. 3). This flexibility of the tail sequence in isolated SmpB would have been conserved throughout evolution as well as the stable core structure.

It has been pointed out that an oligonucleotide binding

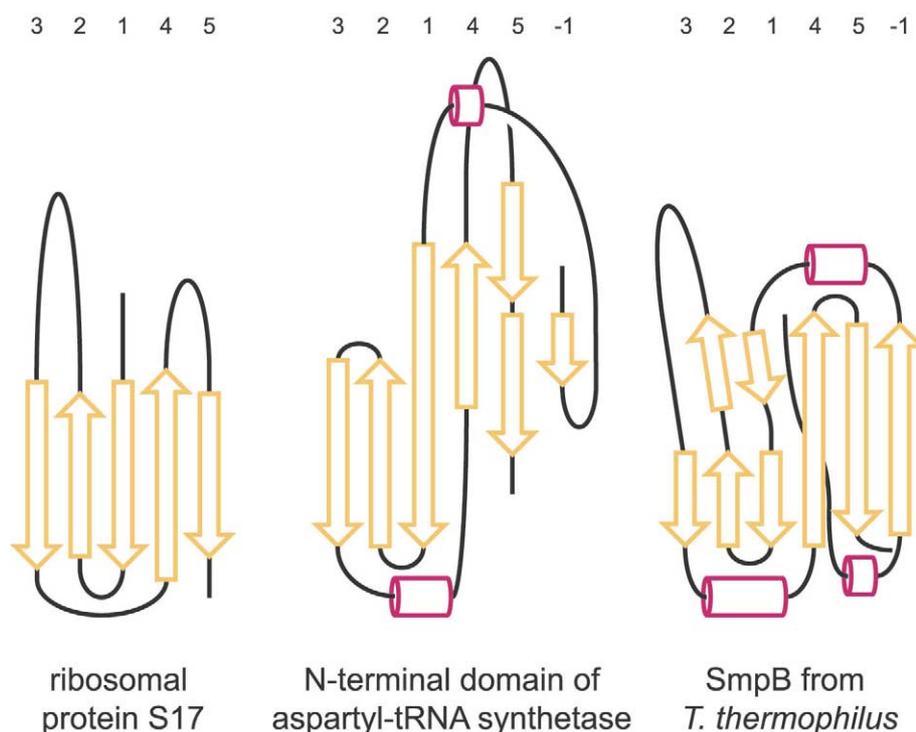


Fig. 4. The secondary structure elements of three different proteins containing an OB-fold or an extended OB-fold. The secondary structure elements of ribosomal protein S17 and the N-terminal domain of aspartyl-tRNA synthetase are derived from the PDB entries 1ASZ and 1RIP, respectively, based on Dong et al. [9].

(OB) fold is embedded in the β -barrel structure of SmpB, although the overall structure appears to be unique [9]. Two structures of SmpB from different bacteria lead to a better understanding of the generality and importance of the structural motif. The minimal common fold of OB-fold proteins comprises a five-stranded β -sheet that is twisted up to form a β -barrel; for instance in ribosomal protein S17 [16] (Fig. 4). In SmpB a helix between the third and fourth strands is added to the OB-fold scheme, with which one end of the barrel is capped, as often found in OB-fold proteins including the N-terminal domain of aspartyl-tRNA synthetase [17]. Furthermore, another β -strand (the -1 st strand) is added prior to the first strand so that it can make networks with both the third and fifth strands to be a component of the barrel, and the other end of the barrel is capped by a helix followed by the -1 st strand (Fig. 4). Consequently, the barrel of SmpB is formed by six β -strands with both ends capped with helices, and this structure can be thought of as an extended OB-fold. Although the N-terminal domain of aspartyl-tRNA synthetase also has the -1 st strand, it is parallel to the fifth strand, and does not interact with the third strand (Fig. 4). The inserted helix between the third and fourth strands is one of the characteristic features of the OB-fold structures of aspartyl-tRNA synthetase as well as SmpB. Because, in the case of aspartyl-tRNA synthetase, this capping helix interacts with the anticodon nucleotides of the cognate tRNA specifically [17], it is interesting to know whether the capping helix in SmpB also interacts with RNA. On the other hand, both SmpB and isolated ribosomal protein S17 have unstructured C-terminal residues and a flexible extended loop between the second and third β -strands (in SmpB, the loop between strands $\beta 4'$ and $\beta 5$) [18]. As for S17, the RNA-protein interactions were already clarified by the crystal structure of the

30S ribosomal subunit from *T. thermophilus* [19]. Within the 30S subunit, the C-terminal extension of S17 becomes helical and wedged in 16S RNA. The extended loop in S17 also interacts with RNA, forming a β -hairpin conformation in the 30S ribosome. Thus, it seems likely that these flexible regions in SmpB are involved in specific interactions with tmRNA via induced fit so that the structures of protein and/or RNA can be changed to be functional [20].

Although biochemical *in vitro* studies show that SmpB binds to the tRNA domain region of tmRNA, little is known about how it interacts with the RNA so that it may not disturb aminoacylation of alanyl-tRNA synthetase and the binding of EF-Tu to the acceptor arm of tmRNA [8,21]. In a tmRNA-SmpB complex mode, it is also unknown whether the SmpB protein interacts with a ribosome. First, studies on the structure of a tmRNA-SmpB complex will be required.

Acknowledgements: We are indebted to Drs. M. Waelchli and Y. Muto for valuable advice on NMR measurements, and to Drs. T. Sakamoto, H. Himeno and A. Muto for helpful suggestions for this work. We also thank Ms. M. Inoue, Ms. K. Hashimoto and Dr. K. Hosono for cloning and purification of the protein. This work was supported by 'Research for the Future' Program (JSPS-RFTF97L00503) from the Japan Society for the Promotion of Science, by a Grant-in-Aid for High Technology Research and Scientific Research on Priority Areas from the Ministry of Education, Science, Sports and Culture, Japan, and by National Project on Protein Structural and Functional Analyses.

References

- [1] Muto, A., Ushida, C. and Himeno, H. (1998) Trends Biochem. Sci. 23, 25–29.
- [2] Karzai, A.W., Roche, E.D. and Sauer, R.T. (2000) Nat. Struct. Biol. 7, 449–455.

- [3] Withey, J.H. and Friedman, D.I. (2002) *Curr. Opin. Microbiol.* 5, 154–159.
- [4] Rudinger-Thirion, J., Giegt, R. and Felden, B. (1999) *RNA* 5, 989–992.
- [5] Barends, S., Wower, J. and Kraal, B. (2000) *Biochemistry* 39, 2652–2658.
- [6] Karzai, A.W., Susskind, M.M. and Sauer, R.T. (1999) *EMBO J.* 18, 3793–3799.
- [7] Shimizu, Y. and Ueda, T. (2002) *FEBS Lett.* 514, 74–77.
- [8] Hanawa-Suetsugu, K., Takagi, M., Inokuchi, H., Himeno, H. and Muto, A. (2002) *Nucleic Acids Res.* 30, 1620–1629.
- [9] Dong, G., Nowakowski, J. and Hoffman, D.W. (2002) *EMBO J.* 21, 1845–1854.
- [10] Marley, J., Lu, M. and Bracken, C. (2001) *J. Biomol. NMR* 20, 71–75.
- [11] Cornilescu, G., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR* 13, 289–302.
- [12] Brünger, A.T. (1993) *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.
- [13] Kane, J.F. (1995) *Curr. Opin. Biotechnol.* 6, 494–500.
- [14] Oshima, T. and Imahori, K. (1974) *Int. J. Syst. Bacteriol.* 24, 102–112.
- [15] Cush, R., Cronin, J.M., Stewart, W.J., Maule, C.H., Molloy, J. and Goddard, N.J. (1993) *Biosens. Bioelectron.* 8, 347–364.
- [16] Murzin, A.G. (1993) *EMBO J.* 12, 861–867.
- [17] Cavarelli, J., Rees, B., Ruff, M., Thierry, J.C. and Moras, D. (1993) *Nature* 362, 181–184.
- [18] Jaishree, T.N., Ramakrishnan, V. and White, S.W. (1996) *Biochemistry* 35, 2845–2853.
- [19] Brodersen, D.E., Clemons Jr., W.M., Carter, A.P., Wimberly, B.T. and Ramakrishnan, V. (2002) *J. Mol. Biol.* 316, 725–768.
- [20] Draper, D.E. (1999) *J. Mol. Biol.* 293, 255–270.
- [21] Barends, S., Karzai, A.W., Sauer, R.T., Wower, J. and Kraal, B. (2001) *J. Mol. Biol.* 314, 9–21.
- [22] Koradi, R., Billeter, M. and Wuthrich, K. (1996) *J. Mol. Graph.* 14, 29–32.