PROTEINS: Structure, Function, and Bioinformatics 64:1095–1097 (2006)

STRUCTURE NOTE

Solution Structure of TA1092, a Ribosomal Protein S24e from *Thermoplasma acidophilum*

Byoung-Young Jeon,[1] Jinwon Jung,[1] Dae-Won Kim,[1] Adelinda Yee,[2] Cheryl H. Arrowsmith,[2] and Weontae Lee[1]*

[1]*Department of Biochemistry and National Research Laboratory, Yonsei University, Seoul, Korea*

[2]*Division of Molecular and Structural Biology, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada*

Introduction. The ribosome is a large ribonucleoprotein complex that catalyzes protein synthesis. In *archaea*, the ribosome is composed of a small 30S subunit and a large 50S subunit. The 30S subunit of the *archaea* ribosome is composed of a 16S rRNA and 28 ribosomal proteins. Genes that encode ribosomal proteins consist of multiple processed pseudogenes, which are dispersed through the genome. Recent advances in ribosome structure[1] suggest close molecular interactions between rRNA and ribosomal protein, within ribosomal subunits. However, the structural role of ribosomal components during various stages of translation remains to be characterized.

Here we report the solution structure of TA1092, a member of the S24e protein family from *Thermoplasma acidophilum*.[2] Members of the S24e protein family are components of the 30S subunit in both *archaea* and *eukaryotes*. Because the detailed structure of ribosomal protein S24e is not yet available for model organisms, the structure of TA1092 will expand our knowledge and understanding of 30S ribosomal protein structure–function for different species.

Methods and Materials. The TA1092 gene was cloned into the expression vector, pET15b, and overexpressed in *Escherichia coli* strain BL21 (DE3) (Novagen Inc., Madison, WI). The cells were grown on a minimal M9 medium, with $^{15}NH_4Cl$ and ^{13}C-glucose to obtain uniformly labeled ^{15}N/ ^{13}C- protein. The protein was purified by immobilized metal affinity chromatography, and the N-terminal His-Tag was cleaved by tobacco etch virus (TEV) protease, for use in NMR experiments. The final protein sample was concentrated to 1.5 mM, in 50 mM potassium phosphate, 0.1 mM NaN_3, and 90% H_2O/10% ^2H_2O buffer solution.

2D-[^{1}H-^{15}N] HSQC, HNCA, HNCACB, CBCA(CO)NH, H(CC)(CO)NH, HCCONH, 3D ^{15}N-edited NOESY (t_{mix} = 200 ms) and 3D ^{13}C-edited NOESY[3] (t_{mix} = 200 ms) spectra were obtained using Bruker DRX 500, 800 and Varian INOVA 500 MHz spectrometers. The resulting data were processed using NMRPipe.[4] Backbone assignments of TA1092 were obtained using Sparky,[5] and side-chain proton resonances were identified using H(C-C)(CO)NH and HCCH-TOCSY spectra. Dihedral angle restraints were derived from chemical shift analysis of Cα, Cβ, CO, and Hα, which were performed using the TALOS program.[6]

Structure calculations were performed by combined use of CYANA[7] and CNS[8] programs. Additionally, PROCHECK-NMR,[9] MOLMOL,[10] and PyMOL[11] programs were used to analyze the final structures.

Results and Discussion. NMR structures were determined from the following: 1141 NOEs [307 intraresidues, 377 sequentials, 134 medium ranges ($2 = |i - j| = <4$), 323 long ranges ($|i - j| > 4$)], 58 distance restraints (for backbone hydrogen bonds) and 129 dihedral angle restraints. The low energy structures were optimized using a water refinement protocol, to obtain final structures. The root-mean-square deviation (RMSD) values of the 10 structures with the lowest energy were determined to be 0.81 ± 0.26 Å, for the structural regions (Table I). The solution structure shows that the N-terminal region forms a compact folded structure, while the C-terminal end (residues 85–98) is very flexible [Fig. 1(A)]. The structure is classified as an α + β fold, with a ββαββα order [Fig. 1(B)]. The first helix (α1) adjoins its counterpart strands (β1:β2, β3:β4) and is anchored to one side of the β-sheet [Fig. 1(C)]. The C-terminal helix (α2) is flanked by β-sheets and moves independently.

Previous reports have suggested that most ribosomal proteins have common structural features, such as N-terminal globular domains and extension of the C-terminal disordered region.[12] The structural data base search by DALI[13] provided eight structural homologues (Z-score > 2.0). The best score was for the S24E protein, which is derived from *Methanosarcina mazei* (1XN9, Z-

Grant sponsor: NRL program of MOST NRDP; Grant number: M1-0203-00-0020, to W.L; Grant sponsor: the Korea Research Foundation Grant (funded by the Korean Government (MOEHRD); Grant number: KRF-2004-005-C00129, to J.J.

*Correspondence to: Weontae Lee, Department of Biochemistry and National Research Laboratory, Yonsei University, 134 Shinchon-Dong, Seodaemoon-Gu, Seoul, Korea 120-749. E-mail: wlee@spin.yonsei.ac.kr

Received 16 March 2006; Accepted 30 March 2006

Published online 19 June 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21045

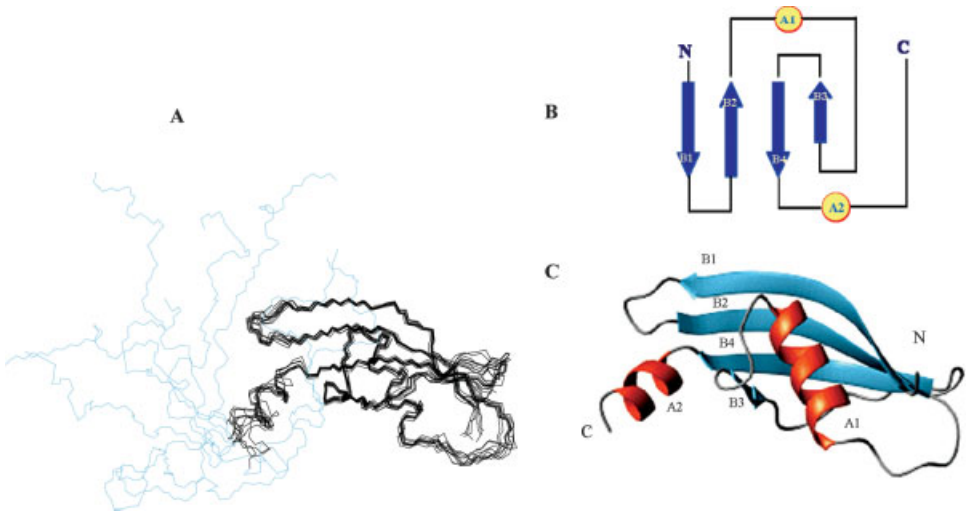

Fig. 1. NMR solution structure of TA1092. (**A**) Superposition of the final 10 structures over the energy-minimized average structure. (**B**) Topological diagram demonstrating the secondary structural arrangement of TA1092. Secondary structures are represented as circles (helices) and arrows (strands), respectively. (**C**) Ribbon diagram of TA1092. The β-strands and α-helices are displayed in cyan and orange color, respectively.

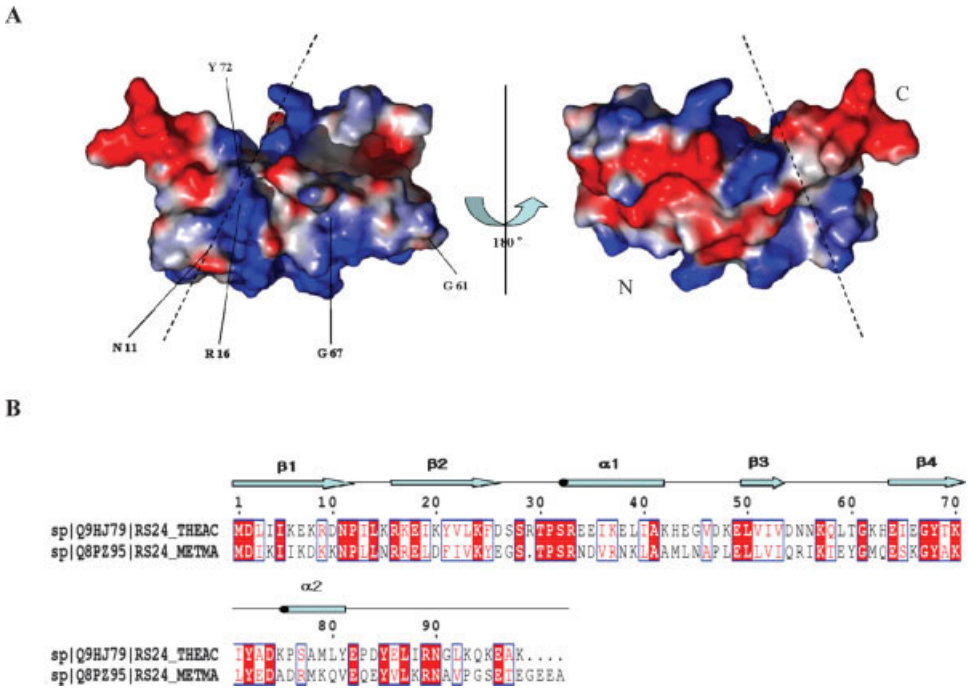

Fig. 2. Mapping the conserved amino acids and surface electrostatic potential of TA1092. (**A**) Red and blue colors represent negative and positive electrostatic potential, respectively. Broken lines indicate a clustered layer of conserved residues on β-sheets. (**B**) Sequence alignment of TA1092 from *Thermoplasma acidophilum* with homologous protein, S24e derived from *Methanosarcina mazei*. The secondary structure of TA1092 determined by NMR is displayed on the top of the sequences.

score 6.6), and has a 31% sequence identity with TA1092. The S24e protein from *M. mazei* showed a common molecular topology with TA1092, specifically the relative orientation of the β-sheet and α1, which are very similar to those of *M. mazei*. The major structural differences are found in β3 and the α2 helix [Fig. 2(B)]. Even though β3 of TA1092 is much shorter than that of *M. mazei*, the overall molecular shape of TA1092 is very similar to that of *M. mazei*.

Highly conserved residues [Asn11 (β1), Arg16 (β2), and Tyr72 (β4)] among the S24e protein family are all found as solvent exposed forms that generate a hydrophilic paddle [Fig. 2(A)]. We hypothesize that these residues have an

TABLE I. Structural Statistics for the Ensemble of 10 Lowest Energy Structures of TA1092

Distance restraints		
All	1141	
Intraresidue	307	
Sequential ($\|i-j\|=1$)	377	
Medium range ($2 \le \|i-j\| \le 4$)	134	
Long range ($\|i-j\|>4$)	323	
Hydrogen bonds	58	
Dihedral angle restraints		
All	129	
ϕ	65	
ψ	64	
CNS energy (kcal/mol)		
Total	−3203.24	
Van der Waals	−192.15	
Electrostatic	−4366.62	
Pairwise RMSD[a]		
	Backbone atoms	All heavy atoms
All residues[b]	1.41 ± 0.33Å	2.33 ± 0.43 Å
Ordered regions[c]	0.81 ± 0.26Å	1.67 ± 0.51 Å
Ramachandran plot (%)[d]		
Residues in most favored regions	84.6	
Residues in additional allowed regions	11.7	
Residues in generously allowed regions	2.8	
Residues in disallowed regions	0.9	

[a]Structural ensemble of the 10 lowest energy structures.
[b] Rmsd values for residues 1–84.
[c] Residues for secondary structural region.
[d] Determined from PROCHECK_NMR[9].

important effect on protein function. This implies that Asn11, Arg16, and Tyr72 residues might interact directly with rRNA, via a hydrogen bonding network. In addition, our structure suggests that hydrophilic residues, located in the β-sheet region, could be involved in the stabilizing the protein/rRNA complex, by intercalating with the rRNA molecule. The conserved basic residues (Arg89, Lys93, Lys95, Lys98), located near the flexible C-terminal tail, could allow an intermolecular interaction with the phosphate group of rRNA, thus providing extra selectivity for binding of TA1092 with the rRNA molecule.

Note: Both the backbone and side-chain chemical shifts for TA1092 have been deposited in the BioMagResBank (accession code BMRB-6989). Coordinates for the 10 structures, and the average energy minimized structure, have been deposited in the RCSB PDB with accession code 2G1D.

REFERENCES

1. Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, Noller HF. Crystal structure of the ribosome at 5.5 Å resolution. Science 2001;292:883–896.
2. Darland G, Brock TD, Samsonoff W, Conti SF. A thermophilic acidophilic mycoplasm isolated from a coal refuse pile. Science 1970;170:1416–1418.
3. Kay LE, Keifer P, Saarinen T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J Am Chem Soc 1992;144:10663–10065.
4. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995;6:277–293.
5. Goddard TD, Kneller DG. SPARKY 3. San Fransicso, CA: University of California; 2004.
6. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 1999;13:289–302.
7. Güntert P, Mumenthaler C, Wüthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 1997;273:283–298.
8. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998:54(Pt. 5);905–921.
9. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;8:477–486.
10. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 1996;14:51–55.
11. DeLano WL. The PyMOL molecular graphics system. San Carlos, CA: DeLano Scientific; 2002.
12. Ramakrishnan V, Moore PB. Atomic structures at last: the ribosome in 2000. Curr Opin Struct Biol 2001;11:144–154.
13. Holm L, Sander C. DALI: a network tool for protein structure comparison. Trends Biochem Sci 1995;20:478–480.