Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

An enzyme with a deep trefoil knot for the active-site architecture

Osamu Nureki,[a,b,c] Mikako Shirouzu,[b,c] Kyoko Hashimoto,[c] Ryuichiro Ishitani,[a,b] Takaho Terada,[b,c] Masatada Tamakoshi,[d] Tairo Oshima,[d] Masao Chijimatsu,[e] Koji Takio,[e] Dmitry G. Vassylyev,[c] Takehiko Shibata,[c,f] Yorinao Inoue,[c] Seiki Kuramitsu[b,c,g] and Shigeyuki Yokoyama[a,b,c]*

[a]Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan, [b]RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan, [c]RIKEN Harima Institute at SPring8, 1-1-1 Kohto, Mikazuki-cho, Sayo, Hyogo 679-5148, Japan, [d]Department of Molecular Biology, Tokyo University of Pharmacy and Life Science, 1432 Horinouchi, Hachioji, Tokyo 192-0392, Japan, [e]Division of Biomolecular Characterization, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan, [f]Cellular and Molecular Biology Laboratory, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan, and [g]Department of Biology, Graduate School of Science, Osaka University, Japan

Correspondence e-mail: yokoyama@biochem.s.u-tokyo.ac.jp

Received 21 February 2002
Accepted 12 April 2002

PDB Reference: *T. thermophilus* RrmA, 1ipa, r1ipasf.

Knots in polypeptide chains have been found in very few proteins. Only two proteins are considered to have a shallow 'trefoil' knot, which tucks a few residues at one end of the chain through a loop exposed on the protein surface. Recently, another protein was found by a mathematical algorithm to have a deep 'figure-of-eight' knot which had not been visually identified. In the present study, the crystal structure of a hypothetical RNA 2′-*O*-ribose methyltransferase from *Thermus thermophilus* (RrmA) was determined at 2.4 Å resolution and a deep trefoil knot was found for the first time. The present knot is formed by the threading of a 44-residue polypeptide chain through a 41-residue loop and is better defined than the previously reported knots. Two of the three catalytic residues conserved in the 2′-*O*-ribose methyltransferase family are located in the knotting loop and in the knotted carboxy-terminal chain, which is the first observation that the enzyme active site is constructed right on the knot. On the other hand, the amino-terminal domain exhibits a geometrical similarity to the ribosomal proteins which recognize an internal loop of RNA.

1. Introduction

In formal knot theory, knots are only defined for closed paths (Adams, 1994). Therefore, the knot state of open chains, such as polypeptides, is determined from their Alexander polynomials (Adams, 1994) after joining both ends of the protein to make a closed path. A few proteins, such as *S*-adenosylmethionine synthetase (MAT; Takusagawa & Kamitori, 1996) and carbonic anhydrase B (CAB; Kannan *et al.*, 1975), were evaluated by their Alexander polynomials and were found to be knotted (Mansfield, 1994, 1997). However, their knots are shallow trefoil knots which are formed by tucking a few (at most 15) residues at one end of the chain through a loop exposed on the protein surface (Takusagawa & Kamitori, 1996; Kannan *et al.*, 1975); such weak knots can disappear if the structure is viewed from a different angle (Taylor, 2000). An alternative approach was made by a new algorithm which smooths the protein structure by contracting the protein chain with the termini fixed at the exterior (Taylor, 2000). This new approach could evaluate the depth of various protein knots and detected a deep figure-of-eight knot in plant acetohydroxy acid isomeroreductase (AIR; Taylor, 2000; Biou *et al.*, 1997). However, in the knotted structure of AIR a helical domain passes through a long loop consisting of more than 100 residues (Biou *et al.*, 1997), which makes it difficult to visually identify the knot in the overall structure of the protein. Here, we report the first deep trefoil knot in an enzyme. Moreover,

two of the three putative catalytic residues are located on the knot.

The enzyme is a hypothetical RNA 2′-*O*-ribose methyltransferase (MTase) from *T. thermophilus* (RrmA). The RNA 2′-*O*-ribose MTases form a large enzyme family that catalyze methyl transfer from *S*-adenosyl-L-methionine (AdoMet) to the 2′-OH group of the backbone ribose (Cavaillé *et al.*, 1999). To date, three 2′-*O*-ribose MTases have well been characterized: SpoU, which is essential for tRNA (Gm18)methyltransferase activity (Persson *et al.*, 1997), *Streptomyces* Tsr, which confers resistance to the antibiotic thiostrepton by methylating a single adenosine in 23S rRNA (Thompson *et al.*, 1982), and yeast Pet56p, which methylates G2270 in mitochondrial 21S rRNA (Sirum-Connolly & Mason, 1993). These three 2′-*O*-ribose MTases share three common short amino-acid motifs, motifs 1, 2 and 3, which define the active site of this enzyme family (SpoU family; Cavaillé *et al.*, 1999; Fig. 1). The *T. thermophilus* RrmA protein clearly contains the three motifs and is therefore likely to belong to the SpoU family (Fig. 1). It should be noted here that *T. thermophilus* contains the authentic *spoU* gene, the product of which shows no sequence similarity with RrmA in the putative RNA-binding domain, in spite of its striking similarity in the catalytic domain. Similarly, RrmA shows no sequence similarity with *Streptomyces* Tsr and yeast Pet56p other than in the catalytic domain. Therefore, compared with the other well characterized 2′-*O*-ribose MTases, RrmA may have a different specificity for the RNA substrate and is probably involved in a different biological function.

2. Materials and methods

2.1. Gene cloning and protein preparation

The ORF of RrmA was amplified by PCR using *T. thermophilus* HB8 genomic DNA as the template and two oligonucleotide primers with artificial *Nde*I and *Bgl*II sites. The 839 base-pair PCR product was digested with *Nde*I and *Bgl*II and was ligated into the pET11b expression vector (Novagen), which was transformed into *Escherichia coli* DH5α cells. The plasmid DNA extracted from a positive colony was sequenced to confirm the identity of the *rrmA* gene. The RrmA overexpression vector, pET11b/rrma, was then transformed into *E. coli* BL21(DE3) cells. For overexpression, cells were grown to an OD_{600} of 0.8 and induced with 0.2 m*M* isopropyl-β-D-thiogalactoside (IPTG) for 4 h. Cells were harvested and sonicated in 20 m*M* Na_2HPO_4 buffer pH 7.0 containing 30 m*M* boric acid and 50 m*M* NaCl (PBN buffer); an equal volume of pre-heated (343 K) PBN buffer was then immediately added to the supernatant. The mixture was heated at 338 K for 10 min to denature the *E. coli* proteins. The heat-treated mixture was centrifuged at 40 000*g* for 20 min and the supernatant was applied onto a 50 ml column of DEAE-Sephacel (Pharmacia Biotech) equilibrated with 20 m*M* Tris–HCl buffer pH 8.0 containing 2 m*M* DTT. RrmA was eluted with a linear gradient of 0–0.3 *M* NaCl. The fractions containing RrmA were pooled, dialysed against 20 m*M* Tris–HCl buffer pH 8.0 containing 2 m*M* DTT and loaded onto a 10 ml column of AF-Heparin Toyopearl (Tosoh). The enzyme was eluted with linear gradient of 0–1 *M* NaCl. RrmA was further purified on a Resource Q column mounted on a Pharmacia FPLC system, being eluted with a linear gradient of 0–0.3 *M* NaCl. Finally, RrmA was purified by gel filtration on a Superdex-75 column equilibrated with 20 m*M* Tris–HCl buffer pH 8.0 containing 2 m*M* DTT and 0.1 *M* NaCl. The estimated yield was 20 mg of purified protein per litre of culture.

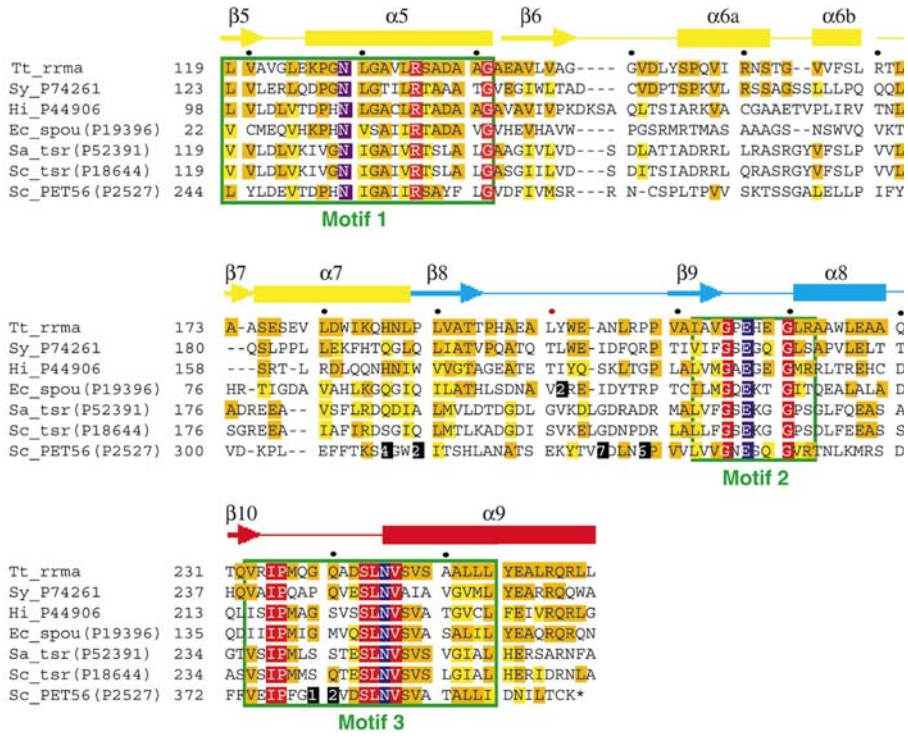


Figure 1
Primary sequence alignment of the putative catalytic domains of *T. thermophilus* RrmA and RNA 2′-*O*-ribose MTases. The secondary-structural elements of RrmA are shown, in which the 'knotting loop' and the carboxy-terminal 'knotted chain' are coloured blue and red, respectively, as in Fig. 4(*c*). Insertions are denoted as black boxes with the number of inserted residues. Identical amino acids and conservative replacements are coloured dark and light orange, respectively. Completely conserved residues are coloured red. The three conserved motifs are indicated as green boxes, in which candidates for the catalytic residues corresponding to Asn129 in motif 1, Glu217 in motif 2 and Asn245 in motif 3 of RrmA are coloured blue. These three motifs have recently been used as probes for detecting homologues in the genomic database from the three primary kingdoms and have identified more than 50 ORFs encoding hypothetical 2′-*O*-ribose MTases (Cavaillé *et al.*, 1999). Of these, *T. thermophilus* RrmA has the highest homology with P74261 from *Synechocystis* sp. and P44906 from *Haemophilus influenzae*.

Table 1
Summary of data collection and refinement statistics.

Values in parentheses are for the highest resolution shell (2.75–2.70 Å for crystal 1 and 2.44–2.40 Å for crystal 2).

	Crystal 1 SeMet				Crystal 2 SeMet			
	Peak	Edge	Remote 1	Remote 2	Peak	Edge	Remote 1	Remote 2
Data collection								
Beamline	BL41XU, SPring-8				BL44B2, SPring-8			
Wavelength (Å)	0.97931	0.97947	0.98726	0.97164	0.9787	0.9791	0.9869	0.9710
Resolution (Å)	40–2.7	40–2.7	40–2.7	40–2.7	40–2.33	40–2.33	40–2.33	40–2.33
Unique reflections	9110	9036	9074	9163	12923	12905	12643	13149
Redundancy	11.9	12.2	12.1	11.6	11.6	10.1	11.4	7.6
Completeness (%)	99.1 (99.6)	98.7 (99.6)	98.5 (99.5)	99.6 (99.3	97.3 (97.9)	98.3 (98.6)	97.5 (98.3)	97.8 (98.9)
$I/\sigma(I)$	14.0 (4.0)	15.8 (6.1)	15.8 (5.3)	10.3 (2.2)	11.1 (2.7)	10.7 (3.3)	10.4 (3.0)	10.6 (2.9)
R_{sym}† (%)	7.2 (23.8)	6.6 (17.4)	5.7 (17.9)	7.1 (36.2)	5.0 (24.5)	5.1 (20.4)	4.6 (22.5)	4.8 (19.8)
Phasing statistics								
No. of sites	3	3	3	3	3	3	3	3
Phasing power, iso	0.34	—	1.39	1.32	0.34	—	1.31	1.84
R_{cullis}‡, iso	0.97	—	0.76	0.75	0.97	—	0.71	0.63
R_{cullis}‡, ano	0.70	0.82	0.98	0.72	0.59	0.71	0.57	0.94
FOM		0.58				0.77		
Refinement								
Resolution (Å)		19–2.7				19–2.4		
Reflections		9128				11584		
Protein atoms		2136				2136		
Water O atoms		286				151		
R_{work} (%)		24.0				22.6		
R_{free} (%)		30.9				28.6		
R.m.s.d. bond length (Å)		0.007				0.006		
R.m.s.d. bond angles (°)		1.30				1.30		
R.m.s.d. dihedrals (°)		23.0				23.1		
R.m.s.d. impropers (°)		0.989				0.937		

† $R_{sym} = \sum |I_{avg} - I_i| / \sum I_i$. ‡ $R_{cullis} = \sum \big||F_{PH} - F_P| - F_H(\text{calc})\big| / \sum |F_{PH} - F_P|$.

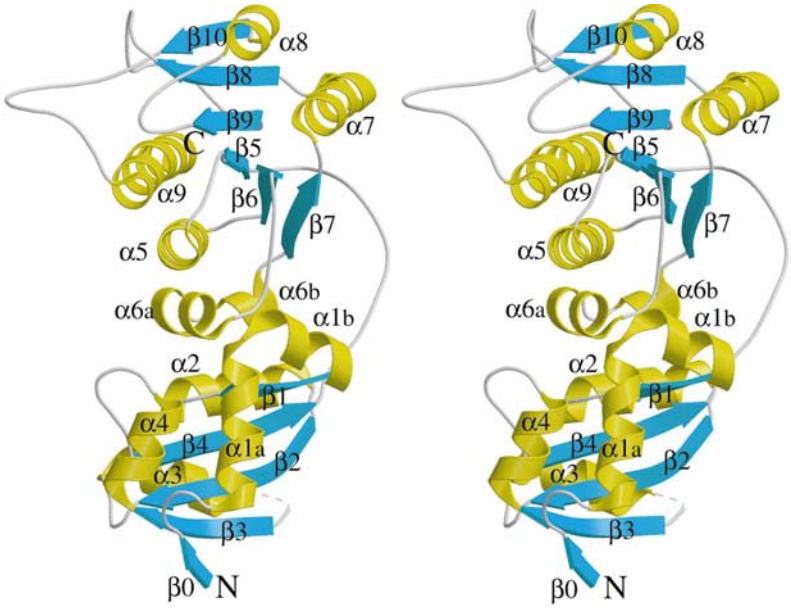


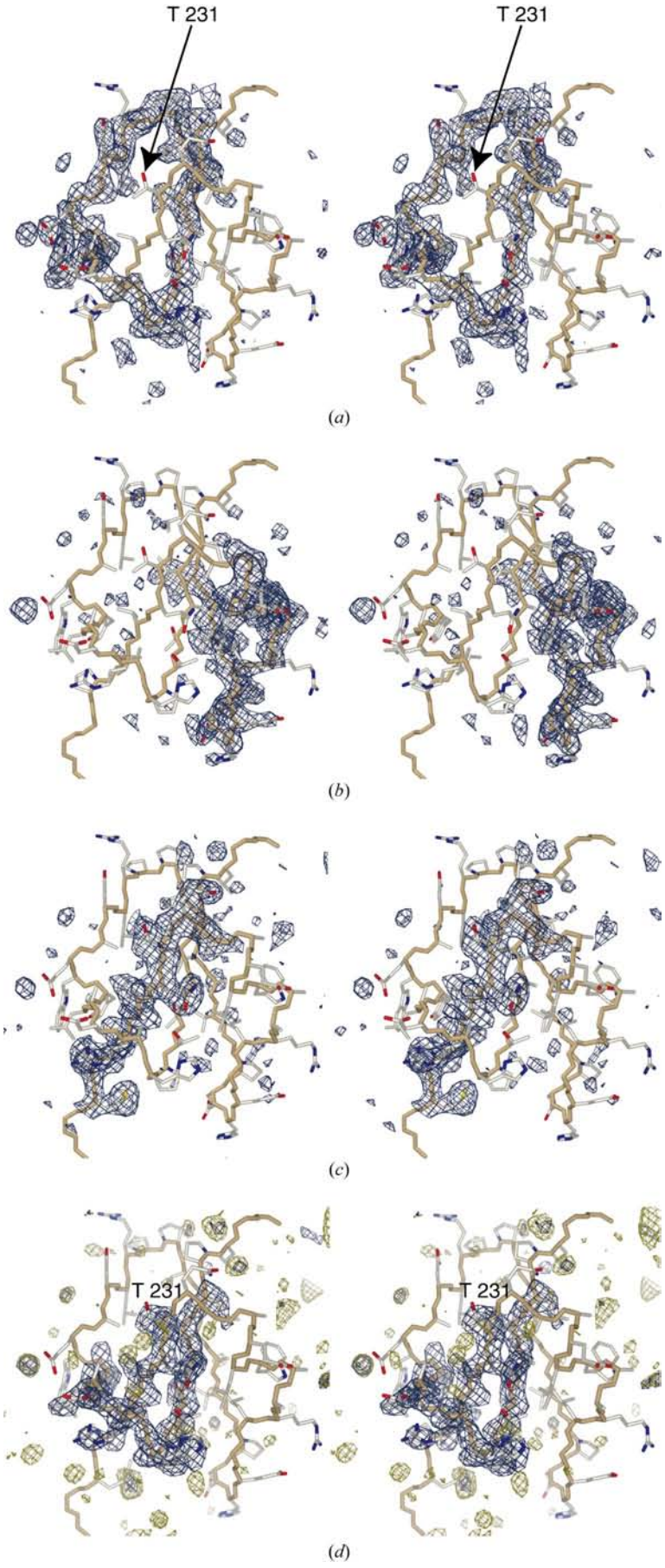
Figure 2
A stereoview of the overall structure of *T. thermophilus* RrmA. The α-helices are shown in yellow, the β-strands are in cyan and the random coils are in grey.

2.2. Crystallization

For crystallization, the hanging-drop vapour-diffusion method was used by mixing 2 µl of the protein solution (15 mg ml^{-1}) with 1 µl reservoir solution and equilibrating against 500 µl of the reservoir solution at 293 K. Two tetragonal crystal forms (forms 1 and 2) were grown in 3 d to dimensions of approximately 0.7 × 0.5 × 0.5 mm. The form 1 crystals (space group $P4_32_12$; unit-cell parameters $a = b = 69.06$, $c = 130.2$ Å) were grown with a reservoir solution containing 0.7 *M* ammonium sulfate with 1.7% 2-propanol. The form 2 crystals (space group $P4_32_12$; unit-cell parameters $a = b = 66.73$, $c = 125.7$ Å) were grown with a reservoir solution containing 6.0% dioxane. Preliminary X-ray data were collected at room temperature on a Rigaku R-AXIS IV imaging-plate system using Cu $K\alpha$ radiation from a Rigaku RU-300 rotating-anode generator operated at 40 kV and 100 mA. In-house experiments revealed both crystal forms to diffract to 3.0 Å resolution. To collect the data under cryoconditions, form 1 and form 2 crystals were stabilized in a solution containing 1.75 *M* ammonium sulfate, 3.75% 2-propanol and 25% ethylene glycol and in a solution containing 12% dioxane and 30% ethylene glycol, respectively.

2.3. Structure determination and refinement

A SeMet derivative was prepared after retransformation of the RrmA gene into met$^-$ *E. coli* B834(DE3). MAD data from the form 1 and form 2 crystals were collected at beamlines



BL41XU and BL44B2 at SPring-8 (Harima) to 2.7 and 2.4 Å, respectively. For both crystal forms, the data were processed and scaled with *HKL*2000 (Otwinowski & Minor, 1997). Three selenium sites were found from the anomalous difference Patterson map using the *RSPS* program (Collaborative Computational Project, Number 4, 1994). The selenium parameters were refined and the phases were calculated using *SHARP* (de La Fortelle & Bricogne, 1997). Phases were improved by density modification using *SOLOMON* (Abrahams & Leslie, 1996). The resultant map was of sufficient quality to trace the entire backbone and to assign the side chains in the carboxy-terminal domain for both crystal forms. In contrast, in some parts of the amino-terminal domain the resulting electron-density map was unable to be interpreted for both crystal forms. Although crystals of form 1 and form 2 belong to the same space group, the unit-cell parameters differ substantially (by 3.5% in a and b and 3.6% in c), which results in an essentially high R factor of 56.8% between the data collected for the form 1 and form 2 crystals, as calculated by the program *SCALEIT* (Collaborative Computational Project, Number 4, 1994). Such a large difference between the structure factors of the form 1 and form 2 crystals allowed us to effectively use the multi-crystal averaging procedure for phase improvement. Indeed, in a course of the multi-crystal averaging, as implemented in the program *DMMULTI* (Collaborative Computational Project, Number 4, 1994), the correlation coefficients between the electron density produced for the two crystal forms increased from 0.202 to 0.958, and the electron-density map was dramatically improved, especially in the amino-terminal domain, enabling unambiguous chain tracing. Basically, the crystal structures of the two crystal forms are the same. An atomic model was fitted into the electron-density map using the graphics program *O* (Jones *et al.*, 1991). Crystallographic positional and simulated-annealing refinements were carried out against the 2.7 Å data set of the form 1 crystals and the 2.4 Å data set of the form 2 crystals using *CNS* (Brünger *et al.*, 1998). The phasing and refinement statistics are summarized in Table 1.

2.4. Analytical ultracentrifugation

Sedimentation-velocity and sedimentation-equilibrium experiments were performed in a

Figure 3
Stereoviews of simulated omit $F_o - F_c$ maps contoured at 2.5σ. Residues 189–211 (*a*), 212–228 (*b*) and residues 229–243 (*c*) are omitted. In (*a*), the 'knotting point' (Thr231) is shown. (*d*) An omit $F_o - F_c$ map (residues 190–201 and 231–235 omitted) (blue) and the residual $F_o - F_c$ map (light green) are presented. Residues 190–201 and 231–235 are in direct contact with each other owing to the knotting. The contour level for both maps is 3.0σ.

Table 2
B factors and stereochemistry.

	Whole molecule	Knot region (190–201 and 231–235)
B factor (Å^2)		
Average	39.2	34.0
Minimum	16.1	24.1
Maximum	72.8	46.9
Overall		
R.m.s.d. bond length (Å)	0.006	0.006
R.m.s.d. bond angles (°)	1.30	1.44
R.m.s.d. dihedrals (°)	23.1	26.5
R.m.s.d. impropers (°)	0.937	0.855
Ramachandran plot		
Most favoured regions (%)	93.3	100
Additional allowed regions (%)	6.7	0
Generously allowed regions (%)	0	0
Disallowed regions (%)	0	0
Main chain		
ω angle s.d.	1.1	1.6
Bad contacts per 100 residues	0.7	0.0
ζ angle s.d.	0.9	1.0
Hydrogen-bond energy s.d.	0.7	0.9
Overall *G* factor	0.2	0.3
Side chain		
χ_1 *gauche*$^-$ s.d.	16.4	23.2
χ_1 *trans* s.d.	10.0	11.4
χ_1 *gauche*$^+$ s.d.	13.6	13.1
χ_1 pooled s.d.	13.2	17.5
χ_1 *trans* s.d.	1	7.0

Beckman XL-A analytical ultracentrifuge at 40 000 and 15 000 rev min^{-1}, respectively, and at 293 K. The RrmA protein was dialyzed against 20 m*M* Tris–HCl buffer pH 8.0 containing 2 m*M* DTT and 0.2 *M* NaCl and was concentrated to 5.7 mg ml^{-1}. In the sedimentation-velocity experiment, 1.1 mg ml^{-1} of RrmA was applied to the ultracentrifuge and the radial position was measured every 5 min. In the sedimentation-equilibrium experiment, 0.68, 0.53 and 0.34 mg ml^{-1} RrmA was loaded and the sedimentation reached equilibrium in 18 h. The empirical specific volume, 0.73 ml g^{-1}, was used to estimate the molecular weight of RrmA.

2.5. Disruption of *rrmA* gene in *T. thermophilus* cells

A fragment containing the kanamycin-resistance gene (*kat*) encoding a thermostabilized kanamycin nucleotidyltransferase (Hoseki *et al.*, 1999) was amplified by PCR with two primers, f-HTK1 (5′-GGTGAGCAATAACTAGGAGGACCATATGAAAGGACCAATAATAAT-3′) and r-HTK1 (5′-GGATATCGTTCAAAATGGTATGCGTTT-3′). The forward primer f-HTK1 contains three termination codons for each different reading frame upstream of the initiation codon of the *kat* gene, while the reverse primer r-HTK1 contains the termination codon of the *kat* gene. The amplified fragment was cloned in the *Eco*RV site of pBluescript II SK+ (Stratagene) in the opposite direction to the *lacZ* gene (pBHTK1). Two gene fragments encoding the amino- and carboxy-terminal regions of the RrmA protein were amplified by PCR with two sets of primers (5′-ATGCGTATCACCAGCACCGCTAAC-3′ and 5′-GGTTTCTCCAGCCCCACCGCTA-3′ for the amino-terminal region, and 5′-GCTGCGCTCAGCAGACGCCG-3′ and 5′-TTATAAAGTTGAGTGCGTCTTAGTC-3′ for the carboxy-terminal region). The amplified fragments were cloned in the *Sma*I site of pUC18 in the same direction as the *lacZ* gene. The *Eco*RI–*Bam*HI fragment encoding the carboxy-terminal region of the RrmA protein was cloned in the corresponding sites of pBHTK1 after disruption of the *Kpn*I site derived from pUC18 by *Kpn*I digestion, blunt-ending and self-ligation reaction (pHTK-RMT34K1). The *Kpn*I–*Hin*dIII fragment of the amino-terminal region of RrmA protein was then cloned into the corresponding sites of pHTK-RMT34K1. The resulting plasmid, in which the *kat* gene was inserted within the *rrmA* gene, was used to disrupt the *rrmA* gene of *T. thermophilus* HB8 *via* insertion of a thermostable kanamycin-resistance marker gene. Transformation of *T. thermophilus* was performed as described

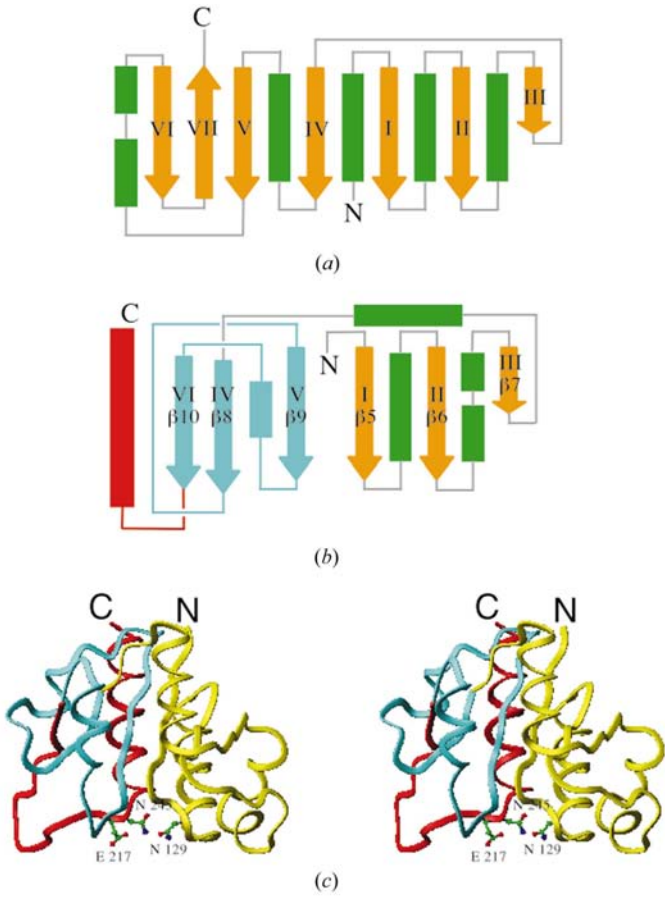


Figure 4
The catalytic fold of *T. thermophilus* RrmA compared with the consensus catalytic fold of 15 other MTases. (*a*) The topology of the consensus fold of the previously reported MTase catalytic domains. The α-helices and β-strands are coloured green and orange, respectively. The β-strands are numbered from the amino-terminus to the carboxy-terminus with roman numbers. (*b*) The topology of the RrmA catalytic domain. The 'knotting loop' and the 'knotted carboxy-terminal chain' are shown in blue and red, respectively. The β-strands are numbered with roman numbers above and arabic numbers below which correspond to Fig. 2. (*c*) A stereoview of the RrmA carboxy-terminal catalytic domain. The 'knotting loop' and the 'knotted chain' at the carboxy-terminus are coloured as in Fig. 1. The three putative catalytic residues Asn129, Glu217 and Asn245 are shown.

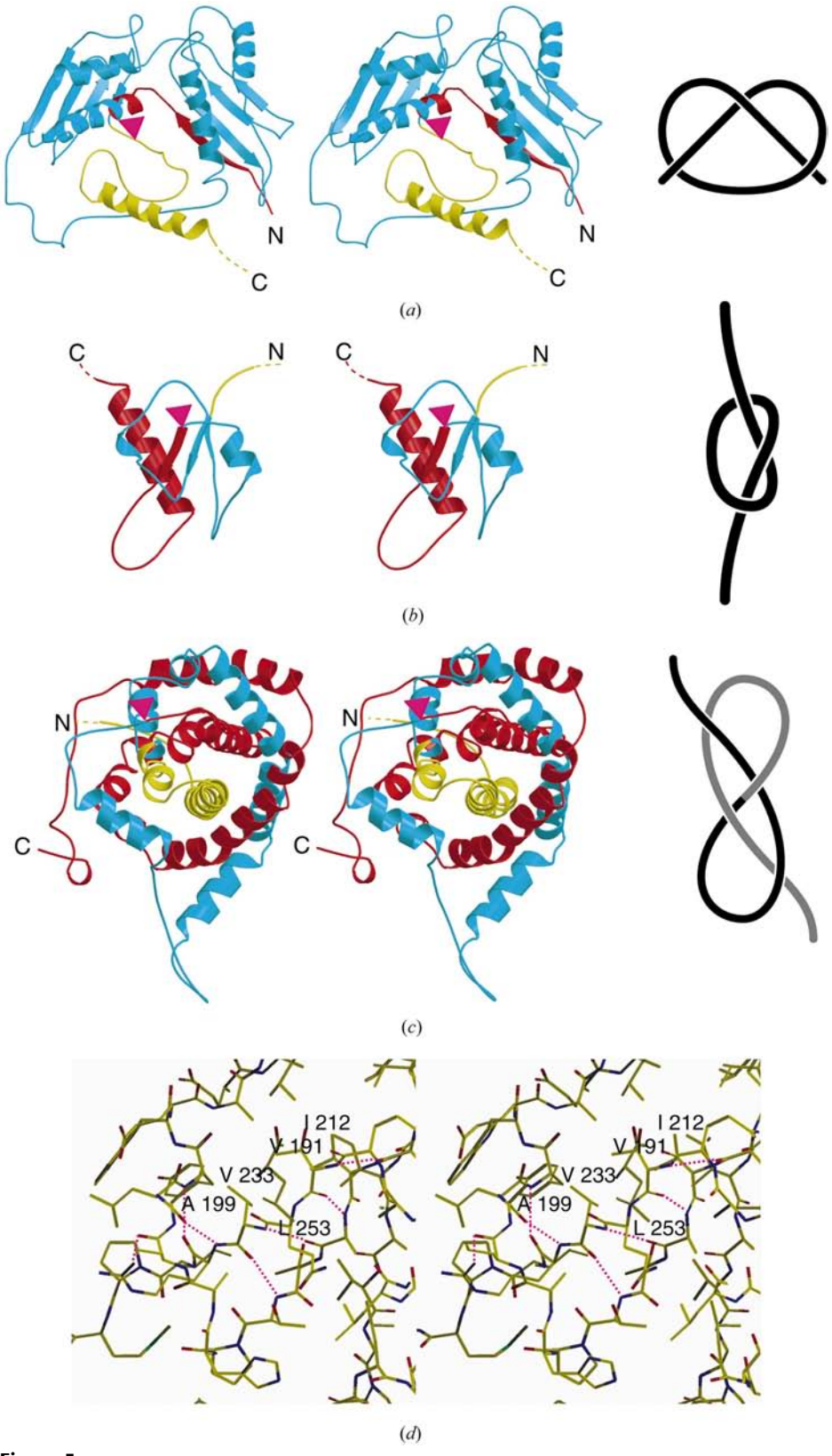


Figure 5
(*a*)–(*c*) Protein knots (stereoviews) and their topologies. The 'knotting loop' and the 'knotted chain' are coloured blue and red, respectively. The 'knotting point' is indicated with a magenta rectangle. (*a*) The shallow trefoil knot of MAT (Takusagawa & Kamitori, 1996). (*b*) The deep trefoil knot of RrmA (this work). (*c*) The deep figure-of-eight knot of AIR (Taylor, 2000). (*d*) Interactions between the threaded polypeptide chain and the 'knotting loop'. Hydrogen-bonding interactions are shown by dotted magenta lines. The amino-acid residues that form a hydrophobic core are indicated.

previously (Hoseki *et al.*, 1999). The transformation mixture was spread on a nutrient medium agar plate (0.8% polypeptone, 0.4% yeast extract, 0.2% NaCl, 0.35 m*M* $CaCl_2$, 0.4 m*M* $MgCl_2$ and 1.5% agar) containing 50 mg ml^{-1} kanamycin and was incubated at 321 K for at least one week.

3. Results and discussion

3.1. Overall structure

RrmA (M_r = 29 980) has an elongated dumbbell shape (Fig. 2), with molecular dimensions of approximately 36 × 46 × 72 Å. The structure consists of two domains, which are connected by a long extended linker (residues 104–115) (Fig. 2). The amino-terminal domain (residues 1–103) folds into a three-layer $\alpha/\beta/\alpha$ sandwich. Amino-acid residues 69–73 and the 11 carboxy-terminal residues are structurally disordered. The carboxy-terminal domain, including the knot region, is structurally well ordered, as shown by the simulated omit map in Fig. 3. The final 2.4 Å model from crystal 2 has an R factor of 22.6% and an $R_{\rm free}$ of 28.6% and shows very good geometry, as examined by a Ramachandran plot: all the residues are observed in the most favoured and additionally allowed regions (93.3 and 6.7%, respectively).

3.2. A knot structure in the catalytic domain

The carboxy-terminal domain (residues 116–275) contains all three of the motifs conserved in the SpoU family (Fig. 1) and therefore is probably the catalytic domain. This domain has a novel fold fundamentally different from that of the catalytic domains of previously reported MTases. Thus far, the crystal structures of 21 AdoMet-dependent MTases have been reported. Of the 21 MTases, 15 MTases, including six DNA MTases (Hha I, Taq I, Hae III, Pvu II, DpnM and RsrI MTases; Cheng *et*

al., 1993; Labahn *et al.*, 1994; Reinisch *et al.*, 1995; Gong *et al.*, 1997; Tran *et al.*, 1998; Scavetta *et al.*, 2000), three RNA MTases (ErmC′, VP39 and FtsJ; Bussiere *et al.*, 1998; Hodel *et al.*, 1996; Bugl *et al.*, 2000), two protein MTases (CheR and PRMT3; Djordjevic & Stock, 1997; Zhang *et al.*, 2000) and four small-molecule MTases (catechol *O*-MTase, glycine *N*-MTase, serine hydroxyltransferase and isoflavone-*O*-MTase; Vidgren *et al.*, 1994; Fu *et al.*, 1996; Scarsdale *et al.*, 2000; Zubieta *et al.*, 2001), have in common a seven-stranded α/β-fold core in their catalytic domains, which is reminiscent of the Rossmann fold or nucleotide-binding fold (Tran *et al.*, 1998; Fig. 4*a*). In the present RrmA structure, the first half of the carboxy-terminal domain folds into a three-stranded α/β topology, like the consensus catalytic domain of the previously reported MTases (Fig. 4*b*). In contrast, in the second half, the three α/β components are arranged in a completely different manner and the last α/β component (β10 and α9) forms a clear well defined knot that has never before been observed in any protein structures (Figs. 2, 4*b* and 4*c*). To confirm that the knot is a real knot, we generated symmetry-related molecules to see if there is a possibility of segment swapping between molecules, which could untangle the knot in the protein fold. However, this possibility could be ruled out, as shown in the simulated omit electron-density map (Fig. 3). In Table 2, the *B* factors and stereochemical parameters are compared between the whole molecule and the knotted regions that have intramolecular contacts arising from the knotting. These values are compatible between the two. Therefore, it can be concluded that the present knot of the RrmA protein is a real knot.

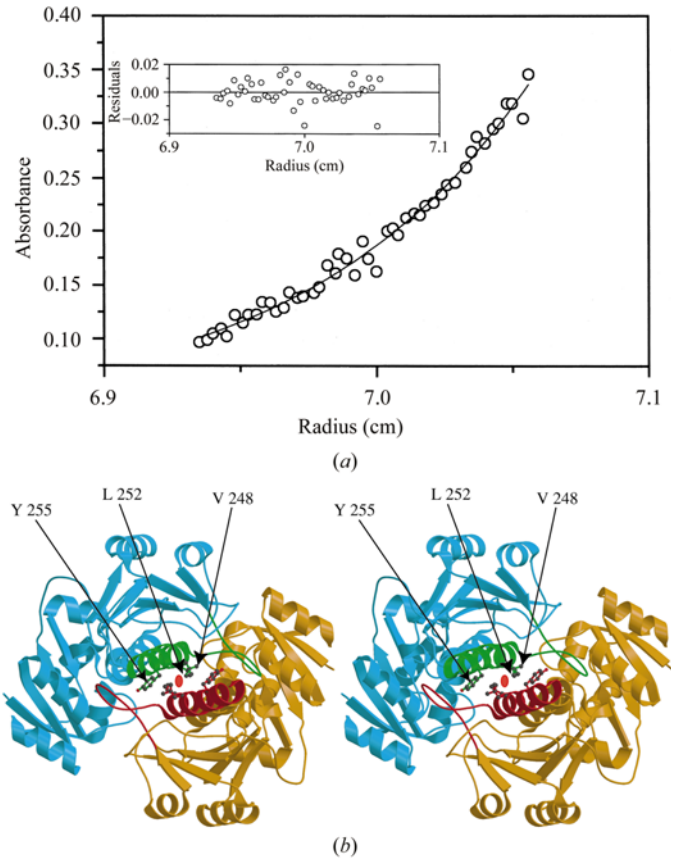


Figure 6
(*a*) Concentration distribution of RrmA (0.34 mg ml^{-1}) during the sedimentation-equilibrium experiment at 15 000 rev min^{-1}. (*b*) The structure of the RrmA dimer. The 'knotted carboxy-terminal chain' and the remainder of one subunit are coloured red and orange, respectively, and the corresponding chains of the other subunit are coloured green and cyan, respectively. The amino-acid residues involved in the molecular dimerization are indicated as ball-and-stick models. The twofold axis is denoted by a red ellipse.

3.3. A deep trefoil knot of RrmA

The present knot of RrmA is a deep trefoil knot that is formed by threading 44 residues (a loop and helix α9) through a circular loop connecting β8 and β9 (Fig. 2). This knotted structure is apparently deeper than the previously reported shallow trefoil knot observed in *S*-adenosylmethionine synthetase (MAT), which was formed by tucking the amino-terminal 15 amino-acid residues through a large loop consisting of 225 residues (Takusagawa & Kamitori, 1996; Fig. 5*a*). Such weak knots may disappear if the structure is viewed from a different angle, whereas the present knot of RrmA can be recognized from any angle. On the other hand, compared with the figure-of-eight knot in AIR (Taylor, 2000), the RrmA knot is much simpler and has a basically different topology (Figs. 5*b* and 5*c*): threading the polypeptide chain through an untwisted loop gives a trefoil knot, while threading the chain through a twisted loop gives a figure-of-eight knot (Figs. 5*b* and 5*c*).

Here, we defined the 'knotting loop' of protein knot structures by selecting two amino-acid residues that cross each other within a minimum distance and, simultaneously, at a point closest to the terminus (termed the 'knotting point'). The 'knotting loop' can thus be defined as being encompassed by the two selected amino-acid residues. In Fig. 5, the 'knotting loop' and 'knotted chain' at the amino- or carboxy-terminus are coloured blue and red, respectively, with the 'knotting point' indicated by a magenta triangle. In AIR, the figure-of-eight knot is formed by threading 137 residues through a large 'knotting loop' consisting of 104 residues, while in the present knot of RrmA the trefoil knot is formed by threading 41 residues through a small 'knotting loop' composed of 44 residues (Figs. 5*b* and 5*c*). Therefore, the present knot of RrmA is comparable to the figure-of-eight knot in AIR with respect to its 'depth', but is quite unique compared with the previously reported protein knots.

The knot of RrmA may be formed by threading either the amino-terminal 190-residue region or the carboxy-terminal 41-residue region through the 44-residue 'knotting loop'. The C-terminal-region threading is possible only after release of the polypeptide from the tRNA. It is more likely that the shorter carboxy-terminal region is threaded through the loop, as the amino-terminal region is much longer and better folded. In the present structure, the 'knotting loop' is too short to thread the knotted helix α9. Therefore, the parallel β-sheet (β9, β8 and β10) around the 'knotting loop' (residues 191–231) or the carboxy-terminal knotted helix α9 should be denatured during the threading.

In the present knotted area, the 'knotting loop' directs its main-chain amide and carbonyl groups inward, while placing the side chains outward. Therefore, the threaded polypeptide chain (β10) forms hydrogen bonds extensively with the inner side of the 'knotting loop', which contains β8 (Fig. 5*d*). Furthermore, Val233 on the threaded chain forms a hydrophobic core with Val191, Ala199, Ile212 and Leu253 (Fig. 5*d*). Thus, these interactions may maintain the structure of the present deep trefoil knot.

3.4. The active site constructed on the knot

An amino-acid sequence alignment of the SpoU MTase family, mapped onto the present structure of the RrmA catalytic domain, reveals that the three conserved motifs 1, 2 and 3 are located on regions β5–α5, β9–α8 and β10–α9, respectively (Fig. 1). Intriguingly, motifs 2 and 3 are located on the 'knotting loop' and the 'knotted chain' at the carboxy-terminus, respectively (Fig. 1), indicating that the knotted structure is conserved in the SpoU MTase family. In the present structure, three strictly conserved residues, Asn129 in motif 1, Glu217 in motif 2 and Asn245 in motif 3, closely approach each other and may form the AdoMet-binding site and the catalytic site (Fig. 4*c*). Kinetic studies of the closely related SpoU enzyme [tRNA (Gm18)methyltransferase] using conventional filter-binding and gel-autoradiography assays suggested that mutation of the Glu residue in motif 2 (corresponding to Glu217 in RrmA) and the Asn residue in motif 3 (corresponding to Asn245 in RrmA) completely impairs the methylation activity (Hori *et al.*, personal communication). In the RrmA structure, Glu217 and Asn245 are located only 14 residues upstream and downstream, respectively, of the 'knotting point' (Thr231) (Fig. 4*c*). Therefore, the present knotted structure is likely to be essential for the construction of the enzyme active site.

While the asymmetric unit of the crystals contains only one molecule, RrmA forms extensive interactions with a symmetry-related molecule. This suggests that RrmA forms an α_2 dimer. The sedimentation-velocity experiment on RrmA gave a sedimentation coefficient of 3.742 S, from which the molecular weight of RrmA was estimated to be about 50 kDa. Furthermore, the sedimentation-equilibrium experiment of RrmA at concentrations of 0.68, 0.53 and 0.34 mg ml^{-1} gave molecular weights of 56 941, 55 454 and 60 409 Da, respectively (Fig. 6*a*). Therefore, we conclude that the RrmA protein behaves as a dimer under the experimental conditions. In the crystal structure, the twofold axis exists almost along the knotted helix α9 and the dimer formation involves hydrophobic interactions between amino-acid residues on the two symmetry-related α9 helices, such as Val248, Leu252 and Tyr255 (Fig. 6*b*). These residues are conserved or replaced with equivalent amino acids in the SpoU MTase family (Fig. 1). Therefore, the knotted structure also has a role in molecular dimerization and this feature may be conserved in the methyltransferase family.

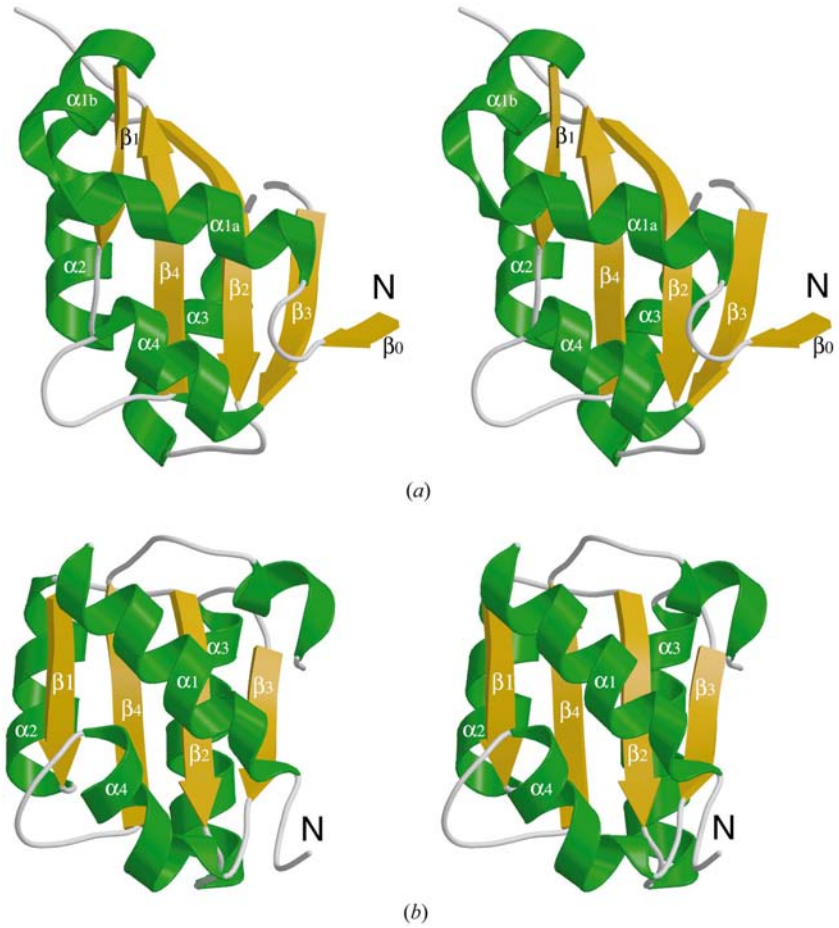


Figure 7
Structural similarity between the RrmA amino-terminal domain (*a*) and the eukaryotic ribosomal protein eL30. (*b*). The architecture of the archaeal ribosomal protein L7ae is also the same. The disordered five-residue loop of RrmA is shown as a dotted line.

3.5. RNA-recognition domain mimicking ribosomal proteins

Most of the AdoMet-dependent DNA/RNA MTases consist of two separate domains: the common catalytic domain and a variable DNA/RNA-recognition domain (Tran *et al.*, 1998). Consequently, the amino-terminal domain of RrmA may function in recognition of the substrate RNA. The question then arises as to what the substrate of RrmA is. The amino-terminal domain has a three-layer $\alpha/\beta/\alpha$ sandwich topology (Fig. 7*a*), which is strikingly similar to those of the eukaryotic ribosomal protein L30 (eL30) (Mao *et al.*, 1999) and the archaeal ribosomal protein L7ae in the large ribosomal subunit (Ban *et al.*, 2000) (Fig. 7*b*), although RrmA shows no sequence similarity to either ribosomal protein. Both ribosomal proteins recognize a helix–loop–helix RNA. Therefore,

the amino-terminal domain of RrmA may recognize the internal loop of a ribosomal RNA (rRNA). This is the first example that an RNA-modification enzyme, possibly for an rRNA, has a domain that mimics a ribosomal protein.

Methylation activity of RrmA against *T. thermophilus* 23S rRNA and 50S and 70S intact ribosomes was not detected (data not shown), which can be ascribed to the fact that some rRNA MTases, such as FtzJ/RrmJ, methylates only the modification-free ribosomes prepared from the bacterial strain deficient in the MTase (Caldas *et al.*, 2000). We then tried to select an *rrmA*-deficient strain of *T. thermophilus* by genetics as described in §2, but we finally found that disruption of the *rrmA* gene was lethal. Therefore, although we could not prepare the modification-free ribosomes that may be necessary for the methylation assay, the RrmA protein is suggested to be essential for the bacteria.

3.6. Conclusions

The present knotted structure of RrmA is not only topologically unique but also quite essential to the active-site architecture and the molecular dimerization. For the figure-of-eight knot in AIR, neither the substrate-binding site nor the catalytic site is adjacent to the 'knotting point' (Gly458; Biou *et al.*, 1997). Therefore, the present structure is the first to show a deep trefoil knot essential for the enzymatic function.

This work was supported in part by Grants-in-Aid for Science Research on Priority Areas (09278101 and 11169204, respectively) to SY and ON from the Ministry of Education, Culture, Sports, Science and Technology of Japan. We are greatly thankful to Dr A. Murzin (MRC) for his helpful discussion. We are greatly indebted to Drs M. Kawamoto (JASRI), N. Kamiya (RIKEN) and S. Adachi (RIKEN) for their help in data collection at SPring-8.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Adams, C. C. (1994). *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots.* New York: Freeman.
Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). *Science*, **289**, 905–920.
Biou, V., Dumas, R., Cohen-Addad, C., Douce, R., Job, D. & Pebay-Peyroula, E. (1997). *EMBO J.* **16**, 3405–3415.
Brünger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, P., Pannu, N., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Bugl, H., Fouman, E. B., Staker, B. L., Zheng, F., Kushner, S. R., Saper, M. A., Bardwell, J. C. & Jakob, U. (2000). *Mol. Cell*, **6**, 349–360.
Bussiere, D. E., Muchmore, S. W., Dealwis, C. G., Schluckebier, G., Nienabev, V. L., Edalji, R. P., Walter, K. A., Ladror, U. S., Holzman, T. F. & Abad-Zapatero, C. (1998). *Biochemistry*, **37**, 7103–7112.
Caldas, T., Binet, E., Bouloc, P., Costa, A., Desgres, J. & Richarme, G. (2000). *J. Biol. Chem.* **275**, 16414–16419.
Cavaillé, J., Chetouani, F. & Bachellerie, J.-P. (1999). *RNA*, **5**, 66–81.
Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W. & Roberts, R. J. (1993). *Cell*, **74**, 299–307.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Djordjevic, S. & Stock, A. M. (1997). *Structure*, **5**, 545–558.
Fu, Z., Hu, Y., Kanishi, K., Takata, Y., Ogawa, H., Gomi, T., Fujioka, M. & Takusagawa, F. (1996). *Biochemistry*, **35**, 11985–11993.
Gong, W., O'Gara, M., Blumenthal, R. M. & Cheng, X. (1997). *Nucleic Acids Res.* **25**, 2702–2715.
Hodel, A. E., Gersham, P. D., Shi, X. & Quiocho, F. A. (1996). *Cell*, **85**, 247–256.
Hoseki, J., Yano, T., Koyama, Y., Kuramitsu, S. & Kamamiyama, H. (1999). *J. Biochem. (Tokyo)*, **126**, 951–956.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kannan, K. K., Notstrand, B., Fridberg, K., Lovgren, S., Ohlsson, A. & Petef, M. (1975). *Proc. Natl Acad. Sci. USA*, **72**, 51–55.
Labahn, J., Granzin, J., Schluckebier, G., Robinson, D. P., Jack, W. E., Schildkraut, I. & Saenger, W. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 10957–10961.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Mansfield, M. L. (1994). *Nature Struct. Biol.* **1**, 213–214.
Mansfield, M. L. (1997). *Nature Struct. Biol.* **4**, 166–167.
Mao, H., White, S. A. & Williamson, J. R. (1999). *Nature Struct. Biol.* **6**, 1139–1147.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Persson, B. C., Jagerand, G. & Gustafsson, C. (1997). *Nucleic Acids Res.* **25**, 3969–3973.
Reinisch, K. M., Chen, L., Verdine, G. L. & Lipscomb, W. N. (1995). *Cell*, **82**, 143–153.
Scarsdale, J. N., Radaev, S., Kazanina, G., Schirch, V. & Wright, H. T. (2000). *J. Mol. Biol.* **296**, 155–168.
Scavetta, R. D., Thomas, C. B., Walsh, M. A., Szegedi, S., Joachimiak, A., Gumport, R. I. & Churchill, M. E. (2000). *Nucleic Acids Res.* **28**, 3950–3961.
Sirum-Connolly, K. & Mason, T. L. (1993). *Science*, **262**, 1886–1889.
Takusagawa, F. & Kamitori, K. (1996). *J. Am. Chem. Soc.* **118**, 8945–8946.
Taylor, W. R. (2000). *Nature (London)*, **406**, 916–919.
Thompson, J., Schmidt, F. & Cundliffe, E. (1982). *J. Biol. Chem.* **257**, 7915–7917.
Tran, P. H., Korszun, Z. R., Cerritella, S., Springhorn, S. S. & Lacks, S. A. (1998). *Structure*, **6**, 1563–1575.
Vidgren, J., Svensson, L. A. & Liljas, A. (1994). *Nature (London)*, **368**, 354–358.
Zhang, X., Zhou, L. & Cheng, X. (2000). *EMBO J.* **19**, 3509–3519.
Zubieta, C., He, X.-Z., Dixon, R. A. & Noel, J. P. (2001). *Nature Struct. Biol.* **8**, 271–279.