Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the RNA-processing inhibitor RraA from *Thermus thermophilis*

Peter H. Rehse, Chizu Kuroishi and Tahir H. Tahirov*

Highthroughput Factory, RIKEN Harima Institute, 1-1-1 Kouto, Mikazuki-cho, Sayo-gun, Hyogo 679-5148, Japan

Correspondence e-mail: tahir@spring8.or.jp

Received 18 June 2004
Accepted 27 August 2004

PDB Reference: RraA, 1j3l, r1j3lsf.

The *menG* gene product, thought to catalyze the final methylation in vitamin K_2 synthesis, has recently been shown to inhibit RNase E in *Eschericha coli*. The structure of the protein, since renamed RraA, has been solved to 2.3 Å using the multiple-wavelength anomalous diffraction method and selenomethionine-substituted protein from *Thermus thermophilus*. The six molecules in the asymmetric unit are arranged as two similar trimers which have a degree of interaction, suggesting biological significance. The fold does not support the postulated methylation function. Genomic analysis, specifically a lack of an RNase E homologue in cases where homologues to RraA exist, indicates that the function is still obscure.

1. Introduction

The *menG* gene product was until recently thought to catalyze the last *S*-adenosylmethionine:2-demethylmenaquinone (SAM) dependent methylation step in vitamin K_2 synthesis, based primarily on the proximity of this gene to another member of the menaquinone pathway (Suvarna *et al.*, 1998). This assigned function was called into question when Lee *et al.* (1997) demonstrated that another non-homologous gene product was potentially responsible for this activity. The crystal structure of the *menG* gene product from *Mycobacterium tuberculosis* (Johnston *et al.*, 2003) revealed a fold unlike any known methyltransferase or SAM-binding protein and therefore confirmed the misannotation. Shortly after, an article by Lee *et al.* (2003) demonstrated that the *menG* gene product can bind to and inhibit the RNA-processing activity of RNase E. The exact function of the protein within the cell remains unknown. The crystal structure of the protein from *Escherichia coli* is also available (Monzingo *et al.*, 2003). To reflect the demonstrated activity, the protein had been renamed RraA. In this text, *Ec*RraA and *Mt*RraA will refer to the RraAs from *E. coli* and *M. tuberculosis*, respectively. Both these structures are characterized as a tightly bound planar trimer postulated to be the biological unit.

In this paper, we discuss the purification, crystallization and X-ray structure determination by the multiple-wavelength anomalous dispersion method of a selenomethionine-substituted RraA homologue from *Thermus thermophilis* (*Tt*RraA). We discuss its structural implications and relation to *Ec*RraA and *Mt*RraA, to which it has about 40% identity. Further, an examination of the *T. thermophilus* genome raises some additional questions with respect to the assigned function.

2. Materials and methods

2.1. Protein expression and purification

The polymerase chain reaction (PCR) was used for gene amplification of *T. thermophilus* HB8 genomic DNA. The PCR product was ligated with pT7blue (Novagen) and digested with *Nde*I and *Bgl*II. The fragment was inserted into the expression vector pET-11a made linear by digestion with *Nde*I and *Bam*HI and transformed into *E. coli* strain B834 (DE3) pLysS for SeMet protein. The cells were grown for 4 h at 310 K in 4.5 l medium containing SeMet and 50 μg ml^{-1} ampicillin, after which protein expression was induced by the addition of 1 m*M* IPTG and cultivation was continued for a further 20 h. Cells (17.5 g) were harvested by centrifugation at 6500 rev min^{-1} for 5 min and suspended in 35 ml 20 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl and 5 m*M* 2-mercaptoethanol, disrupted by sonication and finally underwent heat treatment at 343 K for 11.5 min. The cell debris and denatured proteins were removed by centrifugation (14 000 rev min^{-1}, 30 min, 277 K) and the supernatant was desalted using a HiPrep 26/10 column (53 ml, Amersham Biosciences) with 20 m*M* Tris–HCl pH 8.0 (buffer *A*). The elutant was applied onto a SuperQ Toyopearl 650M column (30 ml, Tosoh) equilibrated with buffer *A* and eluted with a 0–0.3 *M* NaCl linear gradient. The main protein peak was desalted using a HiPrep 26/10 column with buffer *A* and applied onto a Resource Q column (6 ml, Amersham Biosciences) equilibrated with buffer *A* and eluted with a 0–0.2 *M* NaCl gradient. The main protein fraction was dialyzed with 10 m*M* sodium phosphate pH 7.0 (buffer *B*), applied onto a CHT20-I (20 ml, Bio-Rad) column equilibrated with buffer *B* and eluted with a 0.01–0.05 *M* sodium phosphate gradient. The main protein peak was then concentrated and applied onto a HiLoad 16/60 Superdex 75 (120 ml, Amersham Biosciences) column equilibrated with 20 m*M* Tris–HCl, 50 m*M* NaCl pH 8.0. The main protein peak was desalted using a HiPrep 26/10 column with 50 m*M* sodium phosphate pH 7.0 (buffer *C*), after which ammonium sulfate was added to 1.05 *M*. The solution was applied onto a Resource Phe6 (6 ml, Amersham Biosciences) column equilibrated in buffer *C* containing 1.05 *M* ammonium sulfate. The protein was eluted with a 1.05–0 *M* ammonium sulfate gradient in buffer *C*, dialyzed with 20 m*M* Tris–HCl, 50 m*M* NaCl pH 8.0 and concentrated using ultrafiltration (Vivaspin) to 20.0 mg ml^{-1}. The purified protein was homogenous on SDS and native PAGE.

2.2. Crystallization

The protein solution described above was used for crystallization trials. Crystals were initially obtained by the microbatch method (Chayen *et al.*, 1990) using a TERA crystallization robot and a screening kit designed for high-throughput protein crystallization (Sugahara & Miyano, 2002). These poorly shaped and poorly diffracting crystals were improved dramatically with the sitting-drop vapour-diffusion method, altering the polyethylene glycol (PEG) polymer MW from 4000 to 1000 and optimizing for pH, PEG and ion concentration. The buffer remained the same. Diffraction-quality crystals were grown using the sitting-drop method, mixing 1 μl protein solution with 1 μl of a reservoir solution containing 100 m*M* Tris–HCl pH 8.3, 50 m*M* $MgCl_2$ and 29% PEG 1000. Similar block-shaped crystals (0.2 × 0.2 × 0.15 mm) were grown within a week over a range of magnesium concentrations (10–100 m*M*); however, the range was restricted with respect to diffraction quality (50 m*M*).

Table 1
Crystallographic data.

Values in parentheses refer to the highest resolution shell (2.38–2.30 Å).

	Peak	Inflection point	High-energy remote
Resolution (Å)	2.3	2.3	2.3
Wavelength (Å)	0.97900	0.97925	0.97000
$R_{\rm merge}(I)$ (%)	8.4 (43.2)	8.3 (44.1)	7.7 (38.0)
$\langle I/\sigma(I)\rangle$	16.0 (3.3)	17.2 (3.2)	17.5 (3.7)
Completeness (%)	90.8 (92.3)	90.6 (94.2)	89.3 (86.0)
Redundancy	5.1 (4.9)	5.1 (2.7)	5.1 (4.8)
Phasing to 2.3 Å			
No. selenium sites expected	12 (sequence position 1 and 20)		
No. selenium sites found	6		
Overall figure of merit			
After *SHARP*	0.28		
After *RESOLVE*	0.78		
Refinement using peak data			
Resolution range (Å)	38.5–2.3		
No. reflections	66983		
No. reflections for $R_{\rm free}$	3996		
No. protein atoms	7291		
No. solvent atoms	398		
No. other atoms	3 (ions)		
R factor (%)	21.7 (27.7)		
$R_{\rm free}$ (%)	27.9 (32.3)		
R.m.s. deviations, bonds (Å)	0.006		
R.m.s. deviations, angles (°)	1.30		

2.3. Data collection

Crystals were briefly passed through a 50/50 mixture of silicone and paraffin oil and flash-frozen at 100 K in a nitrogen stream. Multiwavelength anomalous diffraction (MAD) data were collected at the synchrotron beamline BL-26B1 at SPring-8 (Harima, Japan) using an R-AXIS V detector. Three data sets were collected from a single crystal using wavelengths determined from a selenium absorption spectrum: peak (0.97900 Å), inflection point (0.97925 Å) and high-energy remote (0.97000 Å). All intensity data were indexed, integrated and scaled with *DENZO* and *SCALEPACK* implemented in the *HKL*2000 program package (Otwinowski, 1993; Otwinowski & Minor, 1997). Data collection is summarized in Table 1.

2.4. Structure determination and model refinement

The X-ray structure of the protein was solved using phases derived from the Se-MAD experiment. The crystal space group was $C222_1$, with unit-cell parameters a = 61.87, b = 109.07, c = 270.32 Å, $\alpha = \beta = \gamma = 90^\circ$, and six molecules in the asymmetric unit. The positions of six of a possible 12 Se atoms were found using the Patterson function within the program *SOLVE* (Terwilliger & Berendzen, 1999). The

remaining six sites correspond to the N-terminal methionine and were expected to be disordered. The six sites were directly input into *SHARP* (de La Fortelle & Bricogne, 1997) for phasing and the resultant phases were input into *RESOLVE* (Terwilliger, 2001) in order to utilize the automatic chain-building function. *SHARP* was unable to locate the selenium sites and the automatic chain tracing failed. However, with the use of *SHARP* the quality of maps was much improved, allowing rebuilding using *QUANTA* (Accelrys) to proceed from the 36% of residues placed. The figure of merit produced by *SHARP* phasing was 0.28; density modification within *SHARP* and *RESOLVE* increased the figure of merit to 0.76. Refinement was carried out using *CNS* (Brünger *et al.*, 1998) with phase combination in the initial stages along with non-crystallographic averaging as monomers were identified. The latter was particularly difficult as molecular replacement using partially completed molecules (even the final molecule) was not effective. In fact, the sixth monomer was only identified after examining the packing of the first five and the placement of the sixth selenium. Residues 1–12 and the C-terminal region could be built only after the monomers had been individually refined. Stereochemical analysis of the structure was performed by *PROCHECK* (Laskowski *et al.*, 1993). Figures were generated with *MOLSCRIPT* (Kraulis, 1991), *BOBSCRIPT* (Esnouf, 1999) and *RASTER3D* (Merritt & Bacon, 1997). Sequence alignments were performed using *CLUSTAL W* (Thompson *et al.*, 1994). Homologous proteins were found using a *BLAST* (Altschul *et al.*, 1997) search of sequence and structure databases.

3. Results and discussion

3.1. Model quality

The structure was solved by the MAD method using a SeMet-substituted recombinant protein. The *R* factor for the final model to 2.3 Å is 21.7%. The model consists of six 164-residue monomers, of which four are complete, with three C-terminal residues missing from one monomer and seven from another. The model contains 7291 protein atoms, 398 water molecules, one magnesium and two chloride ions. The quality of the molecule was assessed using *PROCHECK* (Laskowski *et al.*, 1993), with 82.1% of the residues in the most favourable region and 0.4% in disallowed regions. Refinement is summarized in Table 1.

3.2. Monomer structure

The overall fold of all six molecules in the asymmetric unit is very similar and is characterized by a large proportion of random coil and very flexible N-terminal (residues 1–3) and C-terminal (158–164) regions. The core of the monomer structure consists of a four-stranded parallel β-sheet directly facing three α-helices. This arrangement is generated by $\beta3$–$\alpha2$–$\beta4$–$\alpha3$–$\beta5$–$\alpha4$–$\beta6$; see Fig. 1(*a*) for secondary-structure labelling and Fig. 2 for a cartoon of the monomer structure. The β-sheet is extended on one side by a fifth β-strand ($\beta7$) antiparallel to $\beta3$ and joined to $\beta6$ (located on the opposite end of the sheet) by a long random coil. Antiparallel to $\beta6$, a short β-strand ($\beta2$) extends the β-sheet in the other direction, although it appears to be offset beyond the expected twist of the β-sheet. The three α-helices are in contact with each other, in parallel and roughly in plane. The chain direction of the α-helices is opposite to that of the β-sheets.

The long $\beta3$ β-strand bends almost 90° and forms the base of a further four-stranded antiparallel β-sheet generated by $\beta3$–$\beta1$–$\beta10$–$\beta11$ (see Figs. 1*a* and 2; the latter three strands are coloured green). The orientation is such that the directions of the two β-sheets are rotated 90° with respect to each other when viewed in plane. However, when the β-sheets are viewed edge on they diverge from each other by about 50°. The result is an open-faced wedge with the C-terminal portion of $\beta3$ acting as the base with the open 'mouth' exposed to solvent. The random coil preceding $\beta2$ closes one side of the double β-sheet arrangement, while the other side is covered by a further β-sheet made up of two short β-strands ($\beta8$ and $\beta9$) joined by a turn.

The α-helix ($\alpha1$) is found outside the wedge and in contact with the β-sheet-generated $\beta1$–$\beta11$–$\beta10$. The N-terminal

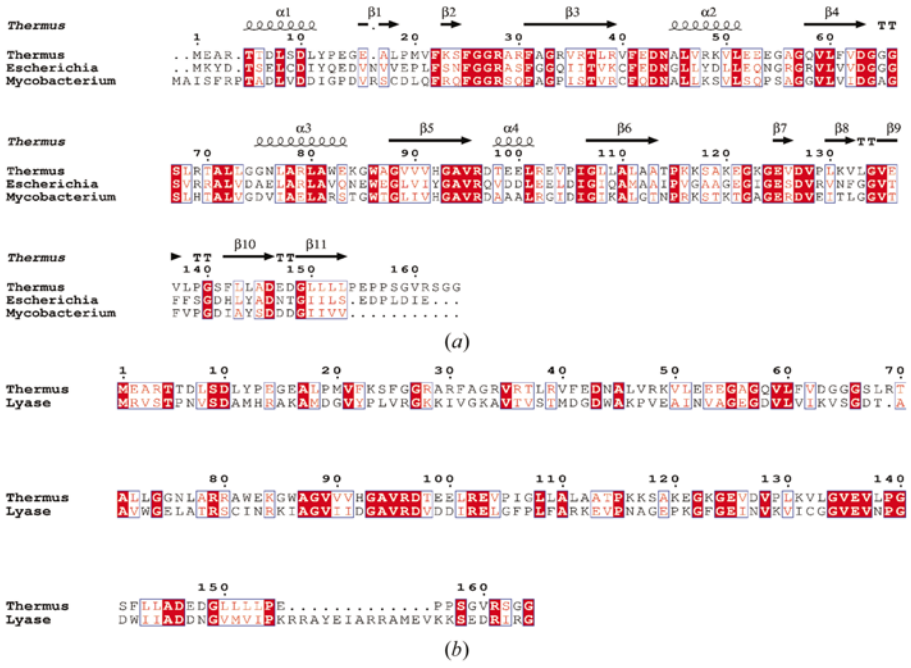

Figure 1
Sequence alignment of *Tt*RraA with (*a*) the sequences of the two other solved RraA structures and (*b*) a lyase. Secondary-structural elements were obtained using *DSSP* (Kabsch & Sander, 1983) and the solved *Tt*RraA structure. The figure was produced using *ESPript* (Gouet *et al.*, 1999). Numbering refers to the *T. thermophilis* sequence. Alignment was performed using *CLUSTAL W* (Thompson *et al.*, 1994).

region including this α-helix is one of the main sources of divergence between the different monomers and in fact only became visible during rebuilding when all molecules were refined independently. Both termini end in proximity to each other, a situation further complicated by both the packing of the two trimers and the crystallographic symmetry. It is here that large variations in main chain occur. With few exceptions, the main chain of the remainder of the molecule is pretty well conserved.

Although the core of the structure described above is a Rossmann fold, a search using the program *DALI* (Holm & Sander, 1993) indicates that the protein does not resemble any other class of folds besides the *Ec*RraA (Monzingo *et al.*, 2003) and *Mt*RraA (Johnston *et al.*, 2003) homologues. More specifically, it does not share features with either methyltransferases or SAM-binding proteins both at the three-dimensional and the sequence levels (Kagen & Clarke, 1994).

Fig. 1(*a*) is an alignment of the sequences from the three crystallographically solved structures; the alignment of the structures is shown in Fig. 3. The *T. thermophilus* sequence shares 41 and 43% identity with the *E. coli* and *M. tuberculosis* sequences, respectively. The corresponding r.m.s. differences of the structures are 1.04 and 0.91 Å, respectively. Most of the deviation occurs outside the core $\beta3$–$\alpha2$–$\beta4$–$\alpha3$–$\beta5$–$\alpha4$–$\beta6$ structure described above, with some variation in the size of the 'wedge' and also the trimer (see below). Johnston *et al.* (2003) provide a detailed alignment of both bacterial and plant species, while Monzingo *et al.* (2003) extend their alignment further to include members of the oxalcitramalate aldolase families. In both cases several residues are conserved throughout, but it is difficult to relate these residues to any sort of function. An examination of conserved residues within the cleft does not suggest which are functionally important. Even when one examines the restricted alignment of the three crystallographic structures (all bacterial sources) shown in Fig. 1(*a*), potential catalytic sites do not present themselves. For example, the region 93–97 corresponds to a loop region which seems to be important in trimer formation (see below), while 22–31 corresponds to a loop blocking one side of the cleft.

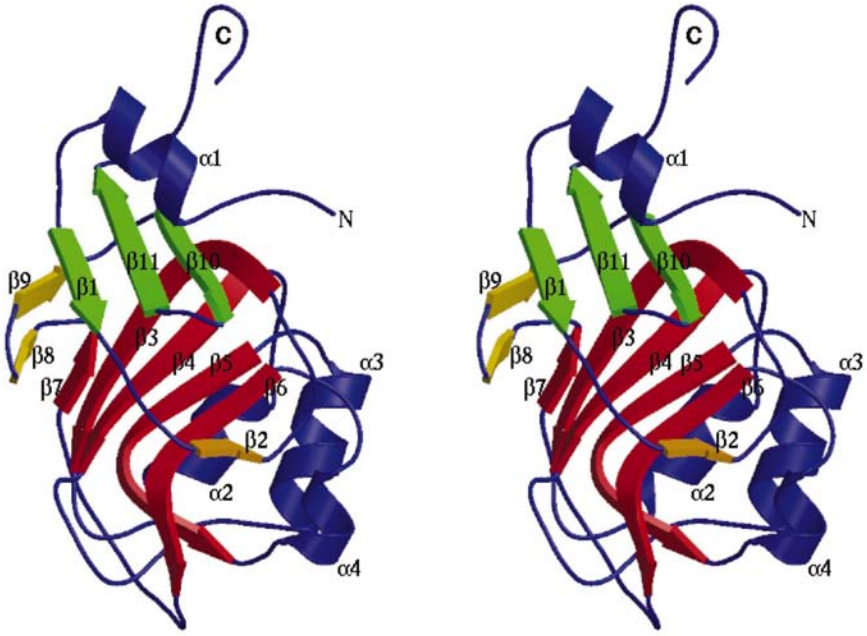

Figure 2
Monomer structure of the *Tt*RraA. The β-sheet Rossmann fold is in red; other β-sheets in the wedge in green and the remaining β-sheets in yellow.

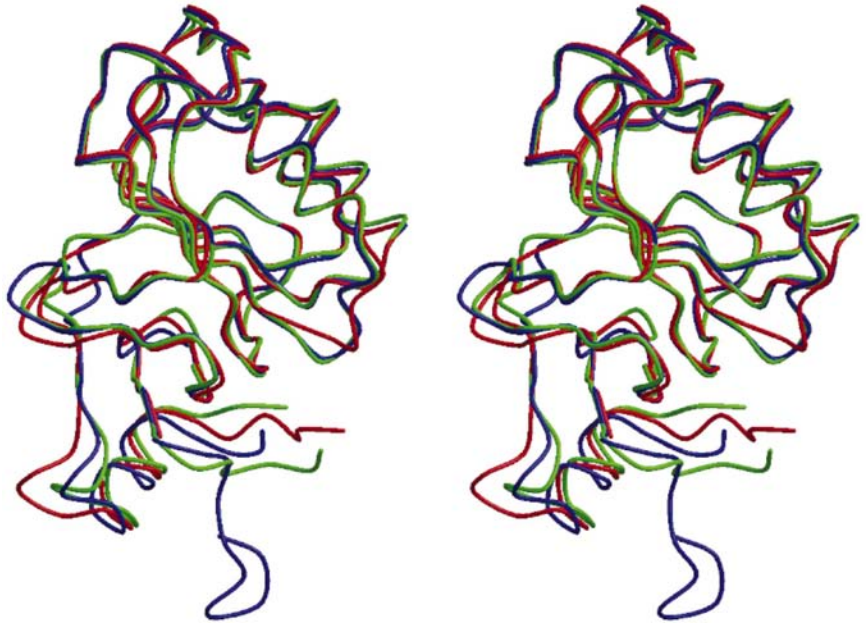

Figure 3
Overlapped *Tt*RraA (blue), *Mt*RraA (red) and *Ec*RraA (green) structures.

3.3. Trimer structure

The six molecules in the asymmetric unit are arranged as two stacked trimers. The monomers of each trimer are arranged in a planar fashion around a threefold axis so that the *AB*, *BC* and *CA* interfaces are nearly the same (see Fig. 4*a*). The main source of interaction is between the $\alpha1$ α-helix and the turn joining β-strands $\beta10$ and $\beta11$ from one monomer and the long loop joining β-sheets $\beta5$ and $\beta6$ and a short stretch of random-coil N-terminal to the α-helix ($\alpha4$) from the other. The excluded surface area calculated using *GRASP* (Nichols *et al.*, 1991) between individual monomers (*A*/*B*, 1139 Å^2; *A*/*C*, 1171 Å^2) is below the minimal value suggested by Janin (1997) for biological significance; however, when the full trimer is considered (*A*/*BC*,

2310 Å^2) the value is well within that expected for a stable interaction and biological significance. This is further demonstrated by a visual inspection of a CPK model of the trimer, which interestingly shows a hole at the threefold interface. This hole is bracketed by the coil/sheet/coil generated by residues 19–30, the turn found at residues 147–148 (also found in contact with another monomer) and the N-terminal portion of β-sheet β6. One can only guess at the significance of this feature, as the residues surrounding this feature are not particularly conserved among members of the RraA family. What are conserved are a number of residues involved in trimer formation. For example, in the *Tt*RraA structure Asp10 and Arg96 forms a salt bridge across the trimer interface. Both these residues are completely conserved among the RraA members. Moreover, a *BLAST* search reveals an *Archaeoglobus* D-aribino 3-heculose 6-phosphate formaldehyde lyase that shares a 31% similarity to the *Tt*RraA when one discounts the first 293 residues of the former. The residues corresponding to Asp10 and Arg96 of the *T. thermophilis* sequence are conserved, suggesting the structural importance of these residues.

Also noteworthy is the different distribution of charged residues on opposite 'sides' of the trimer, suggesting that this structure forms a specific interaction with another molecule. This charge distribution is maintained among other RraA members. The access of solvent to the 'wedge' of the monomer is not hindered by the trimer formation. The relationship of the RraA sequences to the *Archaeoglobus* D-arabino 3-heculose 6-phosphate formaldehyde lyase mentioned above has not previously been considered. The difficulty in assigning catalytic residues, the charge distribution and the linking to a large lyase domain suggests that the RraA has a structural/binding role rather than a catalytic role.

3.4. Hexamer structure

The two trimers are stacked with a 5–10° offset around each of the three axes, with no translation. Therefore, although the threefold axis of the trimers are not perfectly aligned, the holes described above appear to be. Although there is interaction between the trimers, it is not nearly as significant as the interactions of the monomers within the trimers and appears at first glance to be only a function of the crystal packing (see Fig. 4*b*). The large amount of random coil is particularly obvious in this figure. Most of the differences between the six monomers of the asymmetric unit are a consequence of the interaction between trimers and crystallographic packing, not within the trimer. Although the interactions are not as strong as with the trimer, nearly identical hexamer structures are found in both the *Ec*RraA (Monzingo *et al.*, 2003) and *Mt*RraA (Johnston *et al.*, 2003) structures, even though the space groups are *C*2 and $P6_3$, respectively. In the latter two cases, unlike the *Tt*RraA structure, the hexamer requires crystallographic symmetry to be generated. Monzingo *et al.* (2003) do note that the two 4-oxalocitramaltate aldolases characterized so far (Hara *et al.*, 2003; Maruyama, 1990) aggregate as hexamers and require Mg^{2+} for catalysis. One wonders if the hexamer arrangement really does have a biological significance, although the common hexamer motif could be an artifact driven by the properties of the trimer. The excluded surface area between the two trimers of the hexamer (*ABC*/*DEF*, 1560 Å^2) is above the minimal value described by Janin (1997), but not by a large margin. The magnesium found in the *Tt*RraA structure appears to arise from the crystallization conditions rather than from any relation to function.

3.5. Function?

Although its exact role in the cell remains unknown, the ability of RraA to affect RNA processing through its inhibition of RNase E has clearly been shown in *E. coli*. Lee *et al.* (2003) suggest it acts as a global modulator of RNA abun-

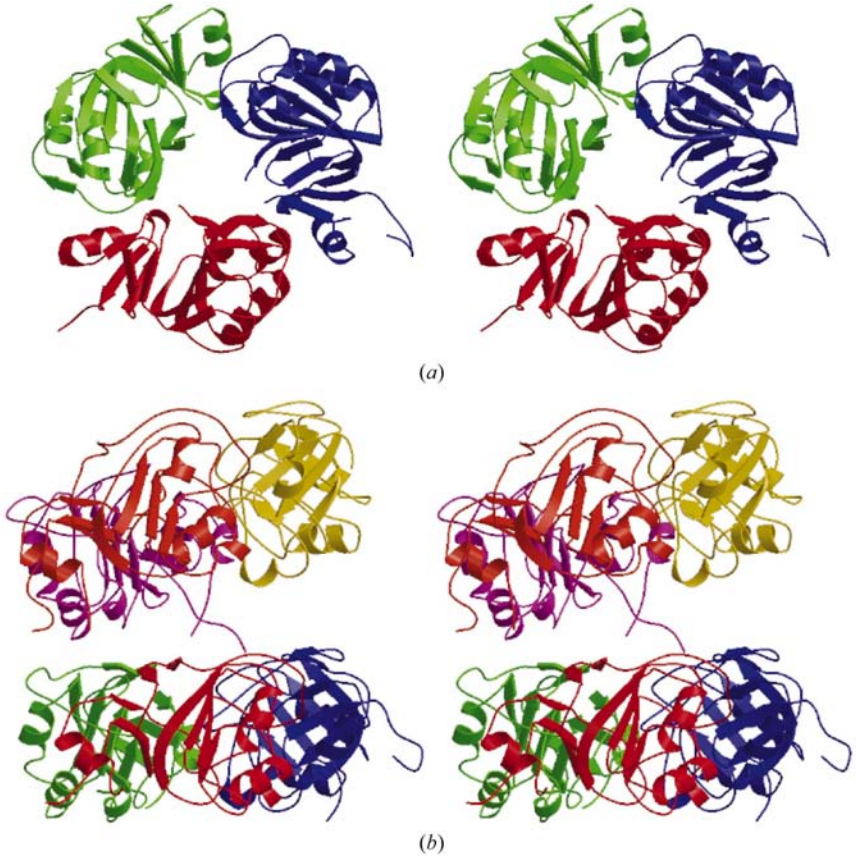

Figure 4
(*a*) Biological unit. The degree of interaction suggests that the trimer structure is the biological unit. The view is directly down a threefold non-crystallographic axis, with the monomers arranged in a planar fashion such that the interfaces are nearly identical. (*b*) Arrangement within the asymmetric unit. The two stacked trimers are offset between 5–10° on each of the three coordinate axes with no translation.

dance. It is clear from the crystal structures obtained so far that the designation as an *S*-adenosylmethionine-dependent methyltransferase is incorrect. However, a *BLAST* search of the genome of both *T. thermophilis* and *M. tuberculosis* using *E. coli* RNase E did not reveal any homologues. This strongly suggests, considering that the relation of the individual RraA sequences is around 40%, that any broad functional assignment of all homologues is still premature. Further the protein's homology with *Archaeoglobus* D-arabino 3-heculose 6-phosphate formaldehyde lyase suggest that the functional role of the motif may be quite broad.

We would like to thank Seiki Kuramitsu and Shigeyuki Yokoyama for the plasmid vector used in the protein expression and Yumi Tokunaga and Masumi Maekawa for technical help during purification. We would also like to thank Mitsuaki Sugahara for the TERA contribution and Mayumi Sugahara for assistance with the homology search. We would like to thank Koji Takio for very useful comments. This work was supported by the National Project on Protein Structural and Functional Analysis funded by MEXT of Japan (Project TT0249/HTPF00080).

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Brünger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chayen, N. E., Shaw Stewart, P. D., Maeder, D. L. & Blow, D. M. (1990). *J. Appl. Cryst.* **23**, 297–302.
Esnouf, R. M. (1999). *Acta Cryst.* D**55**, 938–940.
Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Hara, H., Masai, E., Miyauchi, K., Katayama, Y. & Fukuda, M. (2003). *J. Bacteriol.* **185**, 41–50.
Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
Janin, J. (1997). *Nature Struct. Biol.* **4**, 973–974.
Johnston, J. M., Arcus, V. L., Morton, C. J., Parker, M. W. & Baker, E. N. (2003). *J. Bacteriol.* **185**, 4057–4065.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
Kagen, R. M. & Clarke, S. (1994). *Arch. Biochem. Biophys.* **310**, 417–427.
Kraulis, J. (1991). *J. Appl. Cryst.* **24**, 946–950.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lee, K., Zhan, X., Gao, J., Qiu, J., Feng, Y., Meganathan, R., Cohen, S. M. & Georgiou, G. (2003). *Cell*, **114**, 623–634.
Lee, P. T., Hsu, A. Y., Ha, H. T. & Clarke, C. F. (1997). *J. Bacteriol.* **179**, 1748–1754.
Maruyama, K. (1990). *J. Biochem.* **108**, 327–333.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Monzingo, A. F., Gao, J., Qiu, J., Georgiou, G. & Robertus, J. D. (2003). *J. Mol. Biol.* **332**, 1015–1024.
Nichols, A., Sharp, K. A. & Honig, B. (1991). *Proteins*, **11**, 281–296.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 55–62. Warrington: Daresbury Laboratory.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Sugahara, M. & Miyano, M. (2002). *Tanpakushitsu Kakusan Koso*, **47**, 1026–1032.
Suvarna, K., Stevenson, D., Maganathan, R. & Hudspeth, M. E. S. (1998). *J. Bacteriol.* **180**, 2782–2787.
Terwilliger, T. C. (2001). *Acta Cryst.* D**57**, 1755–1762.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.