doi:10.1016/j.jmb.2003.09.011 J. Mol. Biol. (2003) 333, 931–949

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Insights into Catalysis by a Knotted TrmD tRNA Methyltransferase

Patricia A. Elkins[1], Joseph M. Watts[2], Magdalena Zalacain[1,3] Adam van Thiel[2], Patrik R. Vitazka[2,4], Maria Redlak[2] Cecile Andraos-Selim[5], Fraydoon Rastinejad[6] and Walter M. Holmes[2]*

[1]*GlaxoSmithKline, 709 Swedeland Road, UE0447 King of Prussia, PA 19406 USA*

[2]*Institute for Structural Biology and Drug Discovery and Departments of Biochemistry and Microbiology/Immunology Medical College of Virginia campus of Virginia Commonwealth University 800 East Leigh Street, Suite 212, Richmond, VA 23298 USA*

[3]*GlaxoSmithKline, 1250 South Collegeville Rd, Collegeville PA 19426, USA*

[4]*Department of Human Genetics, MCV campus of Virginia Commonwealth University, Richmond, VA 23298, USA*

[5]*Department of Biology Hampton University, Hampton VA 23668, USA*

[6]*Departments of Pharmacology Biochemistry and Molecular Genetics, Box 800735 University of Virginia, 1300 Jefferson Park Avenue Charlottesville, VA 22908, USA*

**Corresponding author*

The crystal structure of *Escherichia coli* tRNA (guanosine-1) methyltransferase (TrmD) complexed with *S*-adenosyl homocysteine (AdoHcy) has been determined at 2.5 Å resolution. TrmD, which methylates G37 of tRNAs containing the sequence G36pG37, is a homo-dimer. Each monomer consists of a C-terminal domain connected by a flexible linker to an N-terminal AdoMet-binding domain. The two bound AdoHcy moieties are buried at the bottom of deep clefts. The dimer structure appears integral to the formation of the catalytic center of the enzyme and this arrangement strongly suggests that the anticodon loop of tRNA fits into one of these clefts for methyl transfer to occur. In addition, adjacent hydrophobic sites in the cleft delineate a defined pocket, which may accommodate the GpG sequence during catalysis. The dimer contains two deep trefoil peptide knots and a peptide loop extending from each knot embraces the AdoHcy adenine ring. Mutational analyses demonstrate that the knot is important for AdoMet binding and catalytic activity, and that the C-terminal domain is not only required for tRNA binding but plays a functional role in catalytic activity.

Keywords: tRNA; methylation; crystal structure; knot; mutagenesis

Introduction

Cellular tRNAs have now been shown to contain over 80 different modified nucleotides.[1] Modifications are to be found in several domains located in remarkably different tRNA tertiary structural milieu. The fact that over one percent of the bacterial genome is dedicated to modifying RNA,

Present addresses: A. van Thiel, University of Western England, Bristol, UK; M. Redlak, Department of Surgery, Box 980629, MCV campus of Virginia Commonwealth University, Richmond, VA, USA.

Abbreviation used: AdoHcy, *S*-adenosyl homocysteine.

E-mail address of the corresponding author: holmes@hsc.vcu.edu

and that these modifications are conserved across different species, suggests that they must play an important role in cellular viability.[2]

All cellular forms contain enzyme(s) that catalyze the *S*-adenosyl-methionine (AdoMet) dependent methylation of tRNA molecules at G37, a modification that, at least in bacteria and yeast, prevents frame shifting.[3] In eubacteria, TrmD, the product of the *trmD* gene, carries out this reaction. No mechanisms have been reported for this reaction but it has been suggested that the enzyme might act through general base catalysis by deprotonating the N1 group of guanosine and allowing it to nucleophilically attack the methyl group of AdoMet.[4] It has been shown that only a subset of tRNAs containing G at positions 36 and 37 are methylated by bacterial TrmD. Thus, other than histidine tRNA species,[1] only tRNAs that recognize codons beginning with C are methylated. It has also been shown that the entire tRNA structure is required for maximal catalytic activity[5] and that the positional sequence G36pG37 is in fact the key identity element required for recognition.

Other studies have more precisely identified sequences in tRNA which make contact with *Escherichia coli* TrmD.[6] Cleavage of $tRNA_1^{Leu}$–TrmD complexes by chemical or enzymatic probes is deterred by tRNA–protein binding especially in the region of the anticodon stem-loop, and in residues located in or near the core structure of tRNA.

The crystal structure of over 26 AdoMet dependent methyltransferases modifying DNA, RNA, small molecules, or proteins[7] have been determined. Many of these methyltransferases share strongly conserved structural features indicative of their possible evolution from a primitive "Rossmann fold-containing" protein ancestor. Martin & McMillan,[8] provide a comprehensive review covering the methyltransferase fold. Recently, the structure of a number of methyltransferases with a special folded domain have been reported.[9] The first proteins to be identified, bacterial 2′O methyltransferases, display three highly conserved sequence motifs of which motif II, (hV/L/I/M-h-G-X-E-G-V/L/I/M; h, hydrophobic, X is any residue), is thought to be the AdoMet-binding domain.[10] Proteins containing all of these motifs have been designated the SpoU family of methyltransferases. An extensive phylogenetic analysis indicates that the SpoU family might share a common ancestry with bacterial TrmD proteins and some other, less well characterized, protein families detected in Archeae and thermophilic bacteria.[9] It has been proposed that proteins of this group, showing strong homology in motif II, represent a new class of AdoMet-dependent methyltransferases designated SPOUT.

More recently, the structures of a hypothetical 2′-methyltransferase from *Thermus thermophilius* (RrmA)[11] and of RlmB rRNA methyltransferase from *Escherichia coli*[10] have been reported. Both structures appear to belong to the SPOUT methyltransferase class and both contain a "non-canonical" AdoMet-binding site and an unusual peptide knot structure. Based on the positioning of conserved primary sequence motifs, it has been proposed that the knot might be actually involved in the catalytic site.[12] Another two deeply knotted structures have been published to date albeit of proteins with no characterized function, the MT1 gene product of *Methanobacterium thermoautotrophicum*[13] and *Haemophilus influenzae* YibK.[12] A eukaryotic lysine methyltransferase contains a knotted AdoMet-binding domain as well.[14] Only recently, and while the manuscript was under review, the crystal structure of *H. influenzae* TrmD in complex with several ligands has been published.[15] Clearly, a structurally defined novel class of knotted methyltransferases is emerging but there have been no reports so far demonstrating the involvement of this structural characteristic in the enzymatic activity of methyltransferases.

Here, we report both the structure of a biochemically well-characterized methyltransferase, crystallized with cofactor bound, and the supporting mutagenesis work that demonstrates the importance of the deep trefoil knot to the enzymatic activity as well as indicating a role for the C-terminal domain in catalytic activity.

Results

Structure of TrmD

The electron density of the 3.2 Å MAD and solvent flattened maps was exceptionally well connected and traceable in space group *P*3221 with unit cell dimensions of $a = b = 97.25$ Å and $c = 62.66$ Å. Even in these low-resolution maps, clear density for many side-chains was evident throughout the map. Connectivity around the knot was unambiguous and the beta strands and connecting loops were completely built before the knot feature was recognized. Statistics for data collection and refinement can be found in Table 1. Residues in the knot surrounding the cofactor-binding site have the lowest *B*-factors in the structure. While *B*-factors for atoms in the well-ordered AdoHcy adenine ring average 51 Å^2, *B*-factors for the flexible methionine end of the cofactor are substantially higher. The cofactor in this structure has been modeled as AdoHcy, since the electron densities did not permit an unequivocal assignment of a methyl group attached to the methionyl sulfur.

The monomer of TrmD consists of two discrete domains connected by an extended flexible linker region of 11 residues (Figure 1(a) and (b)). The N-terminal domain (residues 1–161) has a compact globular structure with dimensions of approximately 49 Å × 36 Å × 41 Å and a fold that is related to the Rossmann fold. Five major parallel beta strands and a sixth short strand extending the

Table 1. Data collection and refinement statistics

	MAD data			
	Peak	Inflection	Remote	High-resolution data
Data collection (highest shell in parentheses)				
Resolution (Å)	3.2 (3.31–3.20)	3.2 (3.31–3.20)	3.2 (3.31–3.20)	2.5 (2.59–2.50)
Wavelength (Å)	0.97920	0.97031	0.95373	1.00000
Unique reflections	11,286 (1152)	11,305 (1144)	11,314 (1116)	12,013 (1127)
Completeness (%)	99.8 (100.0)	99.8 (100.0)	99.8 (100.0)	98.9 (96.6)
I/σ[1]	14.5 (4.6)	13.6 (4.4)	13.4 (4.1)	23.7 (3.3)
R_{merge} (%)[1]	16.2 (43.9)	14.1 (44.1)	12.4 (41.0)	9.6 (37.3)
Refinement statistics				
Resolution (Å)	50.0–2.5 (2.57–2.50)			
Reflections				
Total	11,974			
R_{free}	604			
R_{cryst}/R_{free}	19.6/24.7			
r.m.s. deviations				
Bonds (Å)	0.007			
Angles (deg.)	1.7			
Water molecules	37			
Average B-factor (Å^2)				
Protein	63			
Solvent	60			
Ramachandran plot statistics				
Most favored	185 (94.4%)			
Additional allowed	9 (4.6%)			
Generously allowed	1 (0.5%)			
Disallowed	1 (0.5%)			

sheet on one edge are sandwiched between layers of helices and make up the core of the domain. The cofactor-binding site is found at one end of the beta sheet and is an integral part of a deep trefoil knot structure. While all methyltransferases share a similar core fold, the TrmD structure has a unique loop formation along the edge of the beta sheet opposite the knot structure. Almost all methyltransferase structures, knotted or unknotted, have beta sheets that are built from a central N-terminal beta strand β1, with β2 lying along side of β1, and β3 as the next strand in the sheet forming the outer edge of the sheet. In TrmD, β3 has folded away from the core beta sheet to form a loop structure called "the halo" which extends over the end of the beta sheet towards the cofactor-binding site (Figure 1(a)). The halo appears to have a significant role in monomer–monomer interactions. More than one-half of the enzyme is involved in the knot structure, which forms the cofactor-binding site.

The alpha helical C-terminal domain (residues 173–246) is small, cylindrical (roughly 20 Å in diameter and 37 Å long), and consists of a long peptide coil (14 residues). This coil, emanating from the disordered linker region, proceeds through a short beta hairpin, β7 and β8, and short helix α7, into a long helix, α8 (19 residues). After α8, the chain forms a small helix, α9, and then immediately extends into the last long α10 helix of 15 residues. This C-terminal helix is perpendicular to, and crosses over, helix α8.

The two monomers form a compact dimer across a crystallographic twofold axis with intimate association between the N-terminal catalytic domains, the N-terminal domain of monomer 1, and the C-terminal domain of monomer 2 (Figure 2(a)). The large flat monomer–monomer interface is formed by helices α6 and α1, augmented by the adenine-binding loop, and the beginning of helix α4. This helix runs the length of one side of the beta sheet. The flat faces of the AdoMet-binding domains are packed against each other burying a surface area of about 1480 Å (including the contribution of AdoHcy) or about 18% of the total surface area of a monomer. Figure 2(b) illustrates a ribbon view of the dimeric structure. Although crystals contain only one monomer per asymmetric unit, the interface between monomers is extensive and suggests that the dimer is a physiological entity. In addition, we have shown that the protein is a dimer in solution *via* gel filtration and by dynamic light-scattering where molecular mass values of 62 kDa and 64,000 kDa, respectively are observed. We have never observed monomers during these determinations, indicating that subunits may not be easily disassociated.

Due to the long interdomain linker-strand, and the close association between the catalytic domains, there are no protein–protein interactions between the N-terminal and C-terminal domains of an individual monomer. A disordered peptide linker spans the 23 Å distance between residues

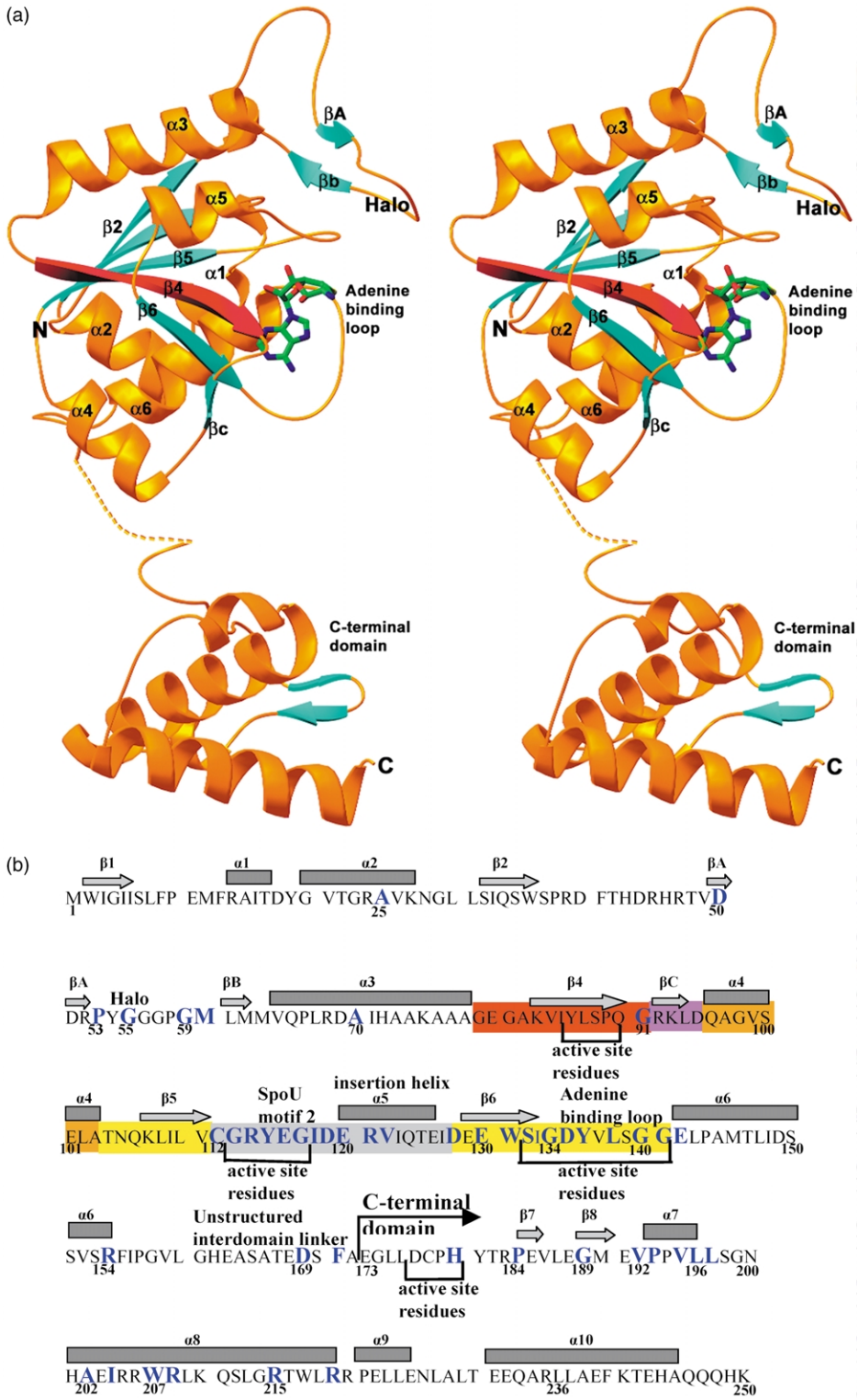

Figure 1. (a) Stereo view ribbon diagram of the TrmD monomer with bound cofactor generated *via* the RIBBONS[23] program. Bound AdoHcy is green, β6 strand is aqua blue, and the remaining strands are lighter blue. Unstructured residues 162 through 172 joining the C-terminal and AdoMet-binding domains are illustrated by the dotted line. (b) Sequence of *E. coli* TrmD annotated to indicate secondary structure and important areas of the structure. Residues that were mutated are shown in bold purple.

161 and 173. This is a reasonable distance to be connected by an extended chain of 11 residues.

The association of the C-terminal domains with the N-terminal domains creates a channel that curves down the side of the dimer, deepening along the cofactor-binding pocket with the AdoHcy buried at the bottom on the channel (Figure 2(a)). The disordered linker peptide that

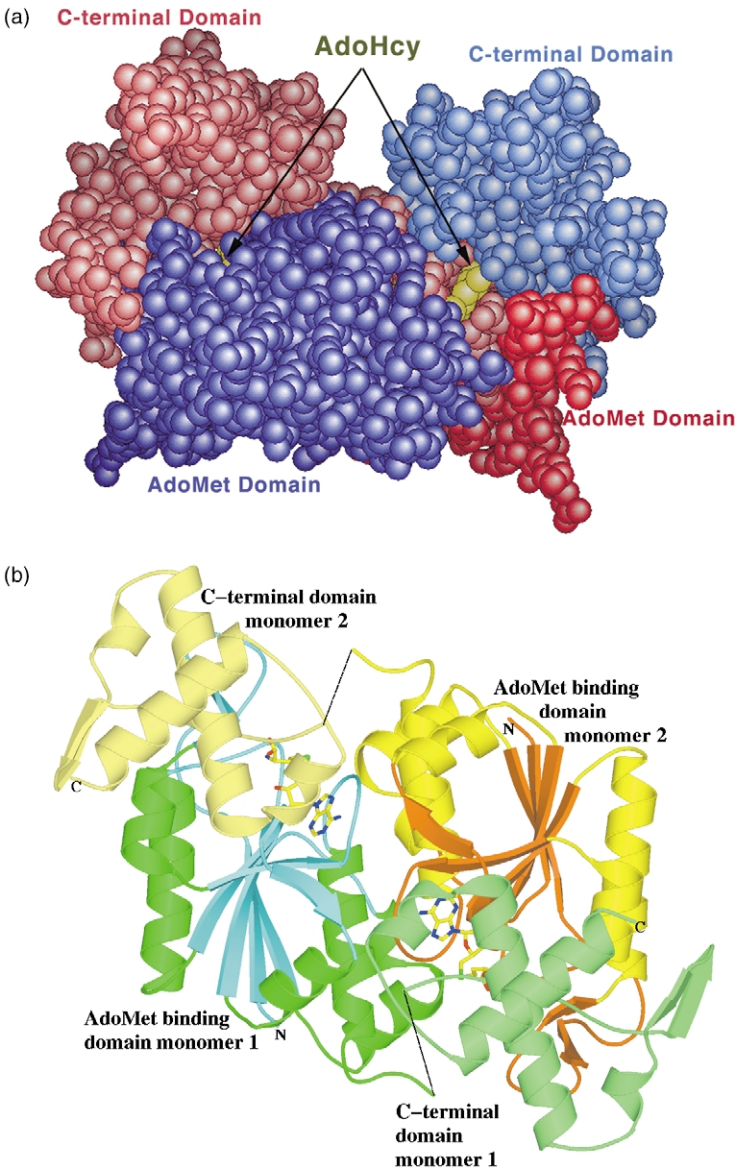

Figure 2. (a) Space filling representation of the dimer. Monomers are shown in red and blue. In this view, the putative tRNA-binding groove runs between the two blue domains of monomer 1 over the buried AdoHcy. (b) The TrmD dimer. The disordered linker peptide between domains has been indicated with a dark broken line and AdoHcy is shown in a ball and stick representation.

connects the two domains is expected to be in an extended conformation to span the available distance and must lie in and along one edge of the groove passing over the AdoHcy-binding site.

The knot

One-half of the core beta sheet structure of the protein is an integral part of the knot structure. Figure 3(a) illustrates how three of the five major beta strands in the central beta sheet make up the deep trefoil knot. The components of this type of overhand knot (illustrated by Nureki *et al.*[11]) can be defined as the knotting loop, the knotting point, and the knotting chain. The knot is initiated with the initial β4 strand in the knotting loop, which leads into the knot. The chain curves around in a loop through strand βC and helix α4. This circular loop forms the outer edge of the catalytic domain. The chain then crosses behind the initial knotting strand β4 to form β5, the central strand of the beta sheet. This strand leads into helix α5, termed the insertion helix, a small helical region that forms a circular insertion to the knot and plays an important role in catalytic activity. The knotting loop terminates in the knotting point, at the beginning of the outermost long strand of the beta sheet, strand β6. β6 initiates the knotting chain and crosses through the knotting loop under strand βC. The knotting chain then encompasses the adenine-binding loop, which encircles the adenine ring of the cofactor. Strands β4 and β5 are splayed to accommodate the ribose moiety and the methionine tail of the cofactor. Figure 3(b) shows an electron density map for AdoHcy in the knot region.

Active site

The cofactor-binding site is buried at one end of the N-terminal catalytic domain in the interface between the N and C-terminal domains. The

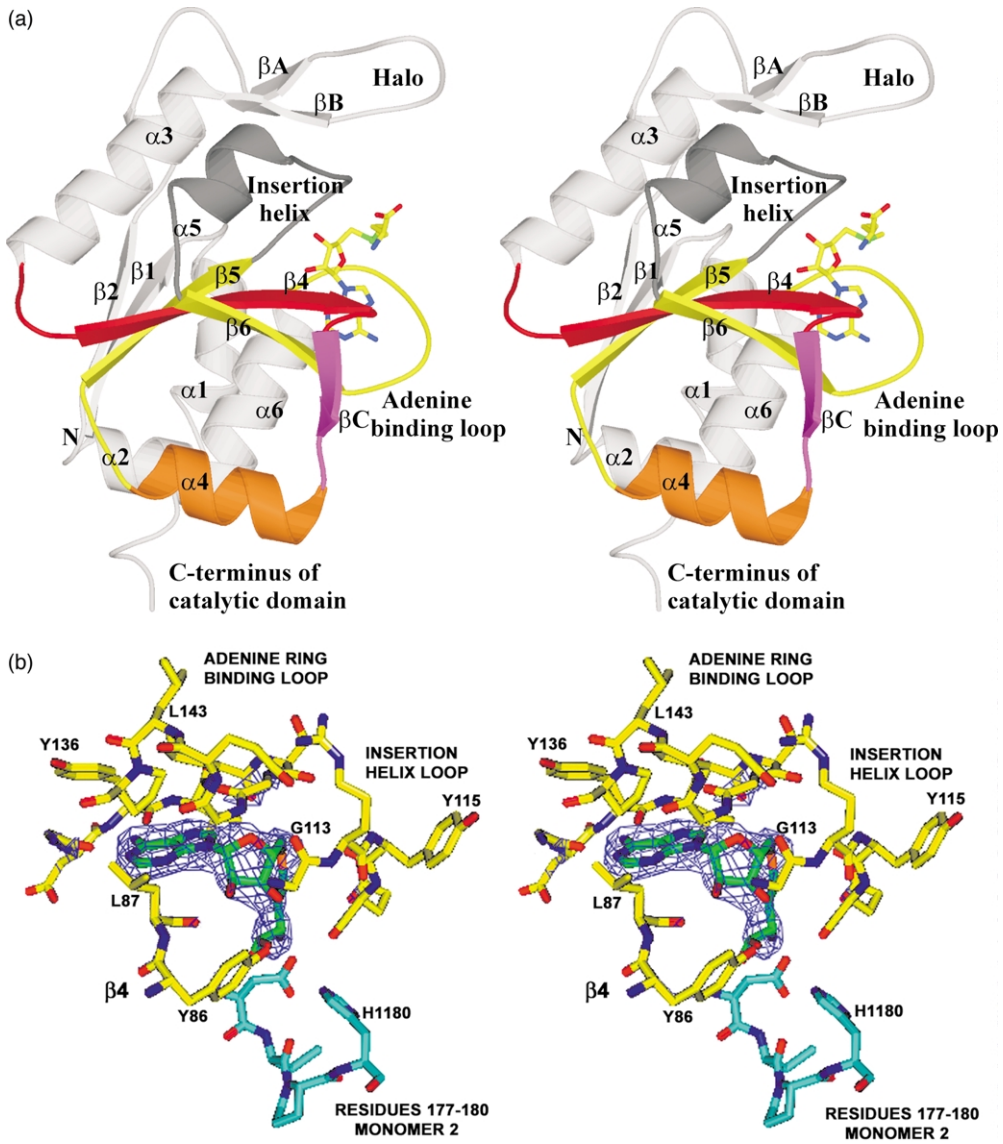

Figure 3 (*legend opposite*)

cofactor is cupped in a deep pocket surrounded by the adenine-binding loop, the C-terminal end of the initial knotting strand β4, and the loop leading into helix α5. In addition, the leading strand of the C-terminal domain, residues 173–184, caps the pocket as it lies across the cofactor-binding site. This suggests interactions across the dimer interface, although precise interactions cannot be ascertained due to the high *B*-factors of atoms in the methionine portion of the cofactor (71–91Å^2). In particular, the adenine moiety of AdoHcy is buried in the pocket. In addition to specific hydrogen-bonding interactions, the adenine six-membered ring is also positioned directly between P89, which stacks face-on to the adenine ring, and P144, which points into the center of the ring. Each proline is about 3.5 Å from the adenine ring. Electron density for the adenine and ribose portions of the cofactor is excellent while that of the methionine end of the cofactor is more diffuse.

AdoHcy assumes a cupped or bent conformation in the binding pocket. The adenine-binding loop, residues 132–141, wraps around the adenine ring. Side-chains of residues in the adenine-binding loop point to either side leaving backbone nitrogen and carbonyl oxygen atoms pointing toward the cofactor (Figure 3(a)). Hydrogen bonds between the backbone nitrogen atoms of I133 and L138 and AdoHcy atoms (3.1 Å and 3.0 Å, respectively) as well as two hydrogen bonds from backbone carbonyl groups to AdoHcy N6 (G134 and Y136, 3.0 Å and 2.8 Å, respectively) tether the cofactor in place. The ribose moiety of the AdoHcy adopts a $^3T_2$ twist conformation that results in an orientation for the methionine portion of the cofactor that is almost perpendicular to the adenine ring. Hydrogen bonds to ribose oxygen atoms fix the middle portion of the cofactor in place.

Residues Y86-Q90 make intimate interactions with the cofactor. AdoHcy is cupped around P89,

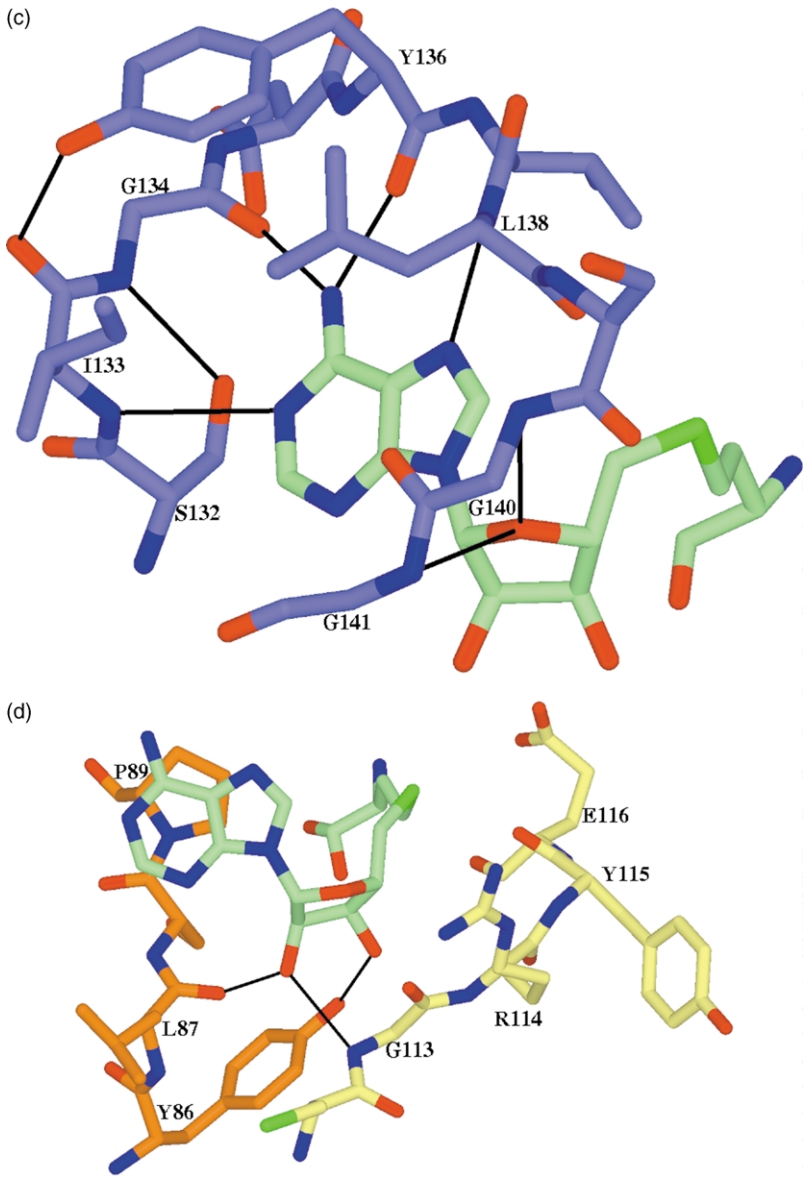

Figure 3. The knot and AdoHcy density. (a) Ribbon diagram of the TrmD cofactor-binding domain. Parts of the structure in the knot are highlighted. The knot can be traced from the N-terminal red initial knotting strand β4, to the magenta strand βC, the orange helix α4, the first yellow strand β5, through the insertion helix-loop in dark grey and crossing through the knot *via* the second yellow strand β6. AdoHcy is shown in a capped-stick representation behind the knot. (b) Stereo view of the knot with the AdoHcy $2F_oF_c$ electron density to 2.5 Å. Contoured at 1.2σ. (c) and (d) Cofactor binding in the active site. AdoHcy is shown in green. (c) Hydrogen-bonding interactions between the cofactor and knot residues are shown with a black line. The adenine-binding loop is shown in blue. (Shown here are residues 132–141.) (d) Residues 86–89 at the end of strand β4 are depicted in orange. Residues 112 to 116 on the insertion helix loop that line the active site groove and represent the SpoU sequence motif II are colored yellow.

which lies face on to the adenine ring. Residues Y86 and L87 make hydrogen-bonding interactions with the cofactor while residues S88-Q90 are in close contact, adjacent and roughly parallel with the methionine portion of the cofactor. (Figure 3(b)). Y86 OH makes a hydrogen bond with the O3* atom of the ribose (2.5 Å), and the carbonyl oxygen of L87 hydrogen bonds to the ribose O2* atom (2.4 Å). Other hydrogen bonding interactions with the ribose are from the backbone nitrogen of G113 in the loop preceding the insertion helix to the ribose O2* (3.3 Å) and from the backbone nitrogen atoms of G140 and G141 to the O4* atom (3.3 Å each).

Comparison with TrmD from *H. influenzae*

Four structures of *H. influenzae* TrmD have been solved[15] in the apo form as well as bound to AdoMet, AdoHcy, or AdoHcy in conjunction with a phosphate ion found near the active site. The sequences of TrmD from the two species are 86% identical in their N-terminal domains and 78% identical in the C-terminal domains and their structures are almost completely conserved (Figure 4). Superposition of the N-terminal domain of *E. coli* TrmD with that of the *H. influenzae* protein bound to AdoHcy gives an r.m.s. difference of 0.38 Å, which is within the range of variability seen in pairwise comparison of the four *H. influenzae* structures (0.10–0.41 Å). The C-terminal domain of *E. coli* TrmD superimposes with the C-terminal domain of the *H. influenzae* TrmD/AdoHcy structure with an r.m.s. difference of 0.75 Å, again within the range of variability of the *H. influenzae* structures (0.25 to 1.31 Å). The only significant area of structural differences between the structure of TrmD from both species is in the halo region. The *E. coli* TrmD halo lies substantially closer to the C-terminal domain of

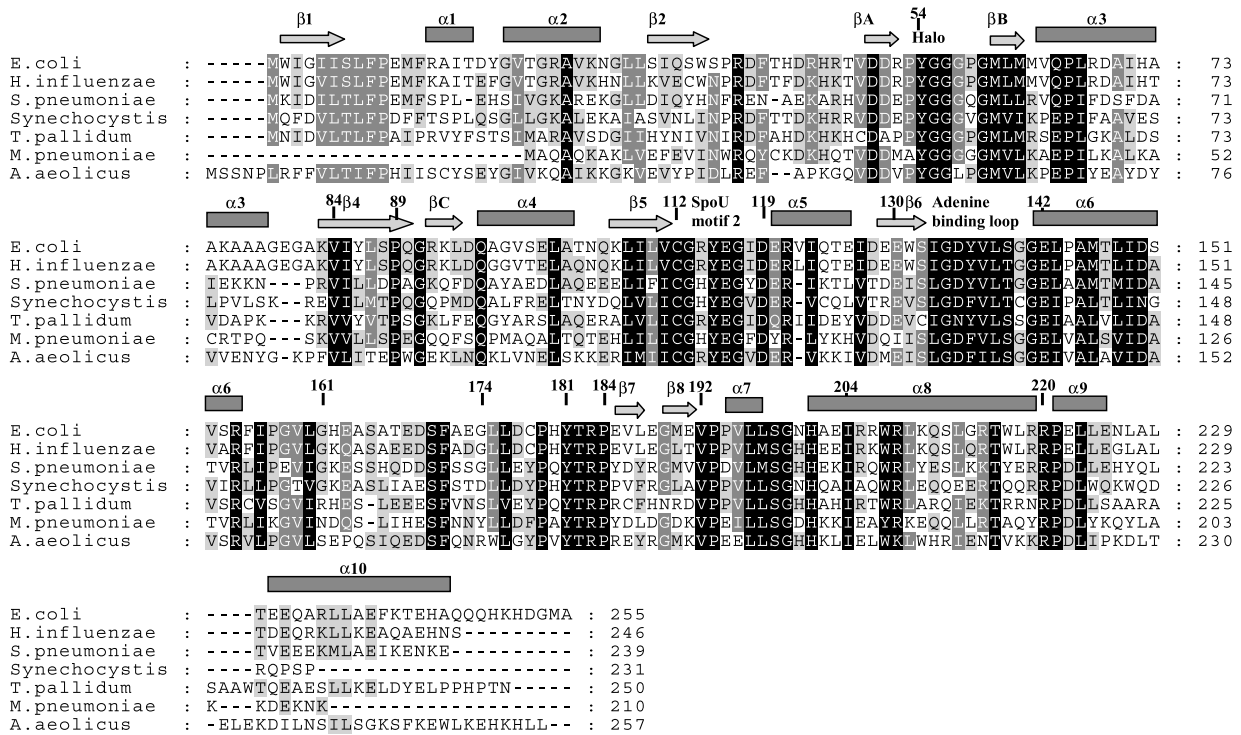

Figure 4. Sequence alignment of diverse TrmD proteins. Multiple sequence alignments of orthologous TrmD sequences from *E. coli* (NP_417098), *H. influenzae* (NP_438371), *Streptococcus pneumoniae* (NP_358281), *Synechocystis sp* (BAA16843), *Treponema pallidum* (AAC65860), *Mycoplasma pneumoniae* (AAB95831) and *Aquifex aeolicus* (AAC07418). According to the percent conservation, columns of residues are shaded dark (100%), medium (80%), light (60%) or not at all. The secondary structure of the *E. coli* TrmD is shown and residue numbers for key residues are indicated.

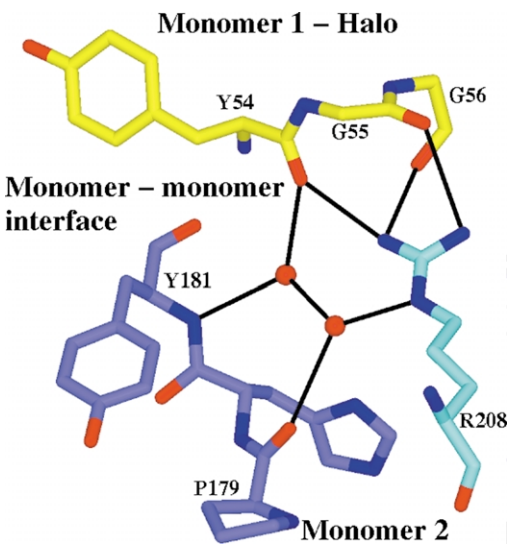

Figure 5. Extensive hydrogen bonding network between the N-terminal domain of the halo region of monomer 1, in yellow, and the C-terminal domain of monomer 2 in cyan and blue in *E. coli* TrmD. In the *H. influenzae* TrmD structures, only the Y181-water-Y54 and P179-water-R208 hydrogen bonds are possible as this interface is slightly more than 1 Å wider in these structures.

monomer 2 across the monomer–monomer interface than occurs in any of the *H. influenzae* structures. The proximity of the halo to monomer 2 results in an extensive network of hydrogen bonds across the interface as shown in Figure 5. The side-chain of R208 of monomer 2, hydrogen bonds to backbone carbonyl groups of residues 54–56 in monomer 1. R208 is also hydrogen bonded to Wat (water) 405 and Wat420, which are in turn hydrogen bonded to the backbone at P179 and Y181 as well as to Y54 in the case of Wat405. In spite of the fact that: (1) the sequences are identical in this region of the *E. coli* and *H. influenzae* structures; (2) R208 is in a similar extended conformation in all of the TrmD structures; and (3) the positions of the ordered water molecules in this part of the interface are approximately conserved, the halo loop in the *H. influenzae* crystal structures is not in a position to allow such an extended hydrogen bonded network to form between the monomers. In these structures, the halo is tethered to monomer 2 *via* one "through-water" hydrogen bond that spans the interface from the carbonyl of Y54 to the backbone amide of Y181.

The *H. influenzae* structures with the three bound cofactors, AdoMet, AdoHcy, and AdoHcy influenced by a nearby phosphate ion, cluster into two classes depending on the conformation of the ribose moiety. For the binary complexes with AdoMet and AdoHcy, the ribose is found in a ^{2}E envelope conformation while in the ternary

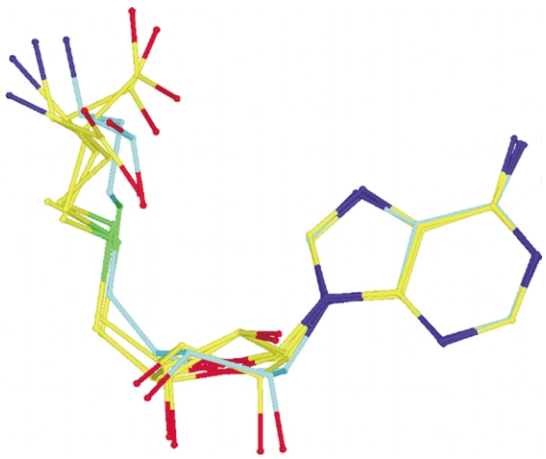

Figure 6. Superposition of cofactors from *E. coli* TrmD (cyan) and *H. influenzae* structures (yellow). *E. coli* TrmD/AdoHcy has the same ribose conformation as the *H. influenzae* TrmD/AdoHcy with phosphate ion. The conformation of the methionine tail portion of the *E. coli* TrmD/AdoHcy is most closely related to that of *H. influenzae* TrmD/AdoMet.

complex with AdoHcy and phosphate, the ribose is in a $^{3}T_{2}$ twist conformation, suggesting that the phosphate ion had a significant effect on the structure of the cofactor in the complex. Interestingly, the ribose in AdoHcy in *E. coli* TrmD is in the $^{3}T_{2}$ twist conformation although there are no phosphate ions apparent in the structure. Figure 6 shows a superposition of these four cofactor structures.

Mutational analysis

One-half of the residues within the active site have been mutated to alanine. Table 2 lists relevant site-directed mutants constructed and tested for effects on catalytic activity. Two obvious structural factors could contribute to dramatic effects of mutation on catalytic activity. First, residues in the adenine-binding loop and throughout the active site make backbone interactions *via* their peptide nitrogen or carbonyl oxygen atoms directly with the cofactor or with side-chains of residues. These interactions help to maintain the tertiary structure of the cofactor-binding pocket. Second, many residues in the active site have side-chains that are involved in hydrophobic or ionic interactions across the monomer–monomer interface and may therefore affect catalytic activity *via* dimer destabilization. Given the apparent central role in AdoMet binding for the adenine-binding loop, extensive mutagenesis of residues in or near this region was informative. The adenine loop, extending approximately from residues 132 to 141 (Figure 1(b)), is very strongly conserved in all bacterial TrmD proteins (Figure 4). Alanine mutagenesis has been carried out on seven out of ten of these residues of which five (G134, Y136, L138, G140, and G141) display hydrogen bonding with AdoHcy *via* backbone interactions (Figure 3(c)). Another residue, D135, makes a salt-bridging interaction with R220 across the monomer–monomer interface. All mutants tested in this region give measurable effects on AdoMet K_m values (Table 2). Mutagenesis of residues S132, G134, D135, Y136 or G141 are particularly detrimental, and either produces an inactive enzyme or yield extremely high K_m values for AdoMet. Mutagenesis of S132 or G134 would not be expected to make extreme alterations to the conformation of the adenine-binding loop. Serine to alanine is a conservative change although S132 OG hydrogen bonds to the backbone nitrogen of G134. Changing G134 to an alanine introduces a methyl group into an open area that appears to have room to accommodate it, although a backbone conformational change would be required. Y136A and G141A mutations result in the highest K_m values for AdoMet measured in these experiments. Mutation of Y136 would result in the loss of the tyrosine side-chain from a hydrophobic pocket in addition to the disruption of a hydrogen bond between the hydroxyl group of tyrosine and the carbonyl oxygen of I133 of monomer 2. This hydrogen bond, like the salt bridge between D135 and R220, would be expected to assist in maintaining the conformation of the adenine loop. Finally, with respect to G141A, some rearrangement would clearly be required to accommodate an additional methyl group in the vicinity of G141. It is clear that the structural conformation of the adenine-binding loop is essential for optimal interaction of enzyme with AdoMet. Notably, the K_m values for tRNA, when they can be measured, are relatively unaffected in these mutants.

Other parts of the active site are residues 113–117 and residues 86–90 in monomer 1 and residue 154 and residues 177–180 in monomer 2 (Figure 1(b)). Residues 113–117 connect strand β5 to the insertion helix α5, (residues 119–125). These and other residues surrounding the SpoU motif II sequence (residues 110–118) have been examined with saturating alanine mutagenesis from C112 to V122. Many of these residues are conserved in all TrmD proteins and SpoU family members implying the crucial nature of this region. Mutations of these residues to alanine have effects on activity ranging from a threefold increase on AdoMet K_m values to outright inactivation (R114, G117, D119, R121). The leading residues in this area have obvious structural roles and appear to be important in cofactor binding (Figure 3(d)). G113 makes a backbone hydrogen bond to the cofactor and R114, Y115, and E116 have large side-chains that line the putative tRNA-binding cleft that runs down the monomer–monomer interface over the AdoHcy-binding pocket. Mutagenesis of G113 has only a modest effect on the K_m of cofactor binding but mutagenesis of R114 inactivates the protein and that of Y115 and E116 significantly decreases cofactor binding. Residues in this region are also involved in two out of the set of three distinct salt bridges (R114:E142, D119:R121, and D135:R220)

Table 2. Mutagenesis table

Mutant	Location	Notes on mutation	tRNA K_m (μM)	tRNA V_{max} (pmol/s)	tRNA K_m/V_{max}	AdoMet K_m (μM)	AdoMet V_{max} (pmol/s)	AdoMet K_m/V_{max}
Mutants implicated in mechanism								
Wild-type			5	0.33	15	8	0.17	48
D169A	Unstructured loop	Proposed general base catalyst responsible for deprotonation of N1 guanine 37	Inactive					
D169E	Unstructured loop	Replacement of base catalyst responsible for deprotonation of N1 guanine 37	18	0.85	21			
R154A	α7	May be required for binding of guanine 37 *via* two hydrogen bonds, one with guanine position 6 oxygen, one with position seven nitrogen	Inactive					
E116A	β5-loop-α5	May be required for binding of guanine 37 *via* two hydrogen bonds with amine group on purine ring position 2	20	0.67	30	215	3.30	65
Active site—adenine-binding loop								
S132A	β6	In active site, S132, guanine oxygen hydrogen bonds to G134N	11	0.50	22	159	3.70	43
G134A	β6-loop-α6	G134 carbonyl hydrogen bonds to adenine ring N6 (3.0 Å)	22	0.22	102	165	0.70	236
D135A	β6-loop-α6	Salt bridge across interface to R220 (2.6 Å)	Inactive					
Y136A	β6-loop-α6	In hydrophobic patch in interface with L138 of monomer 1 and L94, I133, T147 and V99 of monomer 2	22	0.20	110	544	0.70	777
L138A	β6-loop-α6	Packs agains V136 in interface	13	0.50	25	66	1.90	35
G140A	β6-loop-α6	Alanine may disrupt salt bridge between R114 and E142	1	0.01	127	40	0.90	44
G141A	β6-loop-α6					655	0.01	65,500
Active site—SpoU motif II and insertion helix								
C112A	β5-loop-α5	Points out of active site into core of protein	16	1.67	10	50	2.50	20
G113A	β5-loop-α5					56	1.20	47
R114A	β5-loop-α5	Salt bridge with E142	Inactive					
Y115A	β5-loop-α5	Points away from cofactor	4	0.03	114	198	0.35	566
G117A	β5-loop-α5		Inactive					
I118A	β5-loop-α5	Points into a hydrophobic depression on surface of monomer	10	0.13	80	75	0.60	125
D119A	β5-loop-α5	Salt bridge with R121	Inactive					
E120A	α5	In monomer–monomer interface				62	3.80	16
R121A	α5	Hydrogen bond across interface to the carbonyl of P184, salt bridge to D119	Inactive					
V122A	α5	In a hydrophobic pocket adjacent to I118				72	6.30	11

(*continued*)

Table 2 (*Continued*)

Mutant	Location	Notes on mutation	tRNA K_m (μM)	tRNA V_{max} (pmol/s)	tRNA K_m/V_{max}	AdoMet K_m (μM)	AdoMet V_{max} (pmol/s)	AdoMet K_m/V_{max}
Active site—other N-terminal domain								
G91A	β4-loop-βC	Located in the turn between β4 and βC. The backbone is tethered by a hydrogen bond between the S88 carbonyl and the amide nitrogen of G91. CA points towards the backbone of E130 with no room for an additional CB atom	Inactive					
Halo								
D50A	βA	In halo	10	0.17	60	8	0.13	60
P53A	βA-loop-βB	In halo	10	0.33	30	8	0.07	120
G55A	βA-loop-βB	In halo	12	0.17	72	20	0.05	400
G59A	βA-loop-βB	In halo on the opposite side from the loop that is adjacent to the C-terminal domain of monomer 2	Inactive					
M60A	βA-loop-βB	In halo on the opposite side from the loop that is adjacent to the C-terminal domain of monomer 2	10	0.25	40			
C-terminal domain								
F171	unstructured loop		Inactive					
H180A	Linker-loop-α7	In interface next to the SpoU motif 2, the beta strand which interactes with non-adenine portion of the cofactor. H180 points directly towards the methionine tail of cofactor	5	0.07	75	27	0.08	324
P184A	Linker-loop-α7	Extensive hydrogen bonding interactions link the extended stretch of residues from 175–185 of monomer 1 to monomer 2. P184 carbonyl hydrogen bonds to R121 of monomer 2. Mutation should produce structural perturbations	Inactive					
G189A	Linker-loop-α7	Points towards solvent	10	0.08	120			
V192A	Linker-loop-α7	Part of hydrophobic pocket at core of C-terminal domain	Inactive					
P193A	Linker-loop-α7	Structural proline packs in a cluster of F181, W207 and P179	Inactive					
V195A	α7	Packs between two helices	Inactive					
L196A	α7	Side-chain packs in a hydrophobic group of side-chains against the peptide backbone between Y54 and G55 of monomer 2	Inactive					
L197A	α7	Forms part of hydrophobic pocket filled by Y54 of monomer 2	Inactive					

(*continued*)

Table 2 (*Continued*)

Mutant	Location	Notes on mutation	tRNA K_m (μM)	tRNA V_{max} (pmol/s)	tRNA K_m/V_{max}	AdoMet K_m (μM)	AdoMet V_{max} (pmol/s)	AdoMet K_m/V_{max}
A202S	α8	Points into solvent	5	0.17	30			
I204A	α8	Packs against the αC of G55 with L196	Inactive					
W207A	α8	Fills core of C-terminal domain	Inactive					
W207F	α8	Fills core of C-terminal domain	5	0.08	60	10	0.08	120
W207H	α8	Fills core of C-terminal domain	7	0.08	84	15	0.07	225
R208A	α8	Hydrogen bonds to backbone carbonyl groups of Y54 of monomer 2 (3.0 Å) and of G56 (3.1 Å)	Inactive					
R215A	α8		5	0.07	75			
R219A	α8	Salt bridge to E130 of monomer 2	5	0.08	60	16	0.07	240
Other mutations								
A25S	α2	Points from helix α2 towards the beginning of the linker peptide. There appears to be room to accommodate a mutation	5	0.12	43			
G59A	βA-loop-βB	In halo on the opposite side from the loop that is adjacent to the C-terminal domain of monomer 2	Inactive					
M60A	βA-loop-βB	In halo on the opposite side from the loop that is adjacent to the C-terminal domain of monomer 2	10	0.25	40			
A70S	α3	Points at P38	5	0.08	60			
D128A		Points into solvent α5 loop-β6	Inactive					
E130A	β6	Salt bridge across interface to R219	10	0.35	29	43	3.30	13
W131A	β6	Hydrophobic pocket	5	0.25	20	41	0.13	315
D135A	β6-loop-α6	Salt bridge across interface to R220 (2.6 Å)	Inactive					
Y136A	β6-loop-α6	In hydrophobic patch in interface with L138 of monomer 1 and L94, I133, T147 and V99 of monomer 2	22	0.20	110	544	0.70	777
E142A	α6	Salt bridge to R114	8	0.17	46			

between highly conserved residues that appear to be important for AdoMet binding (Figure 7). Disruption of these salt bridges by site-directed alanine mutagenesis at position R114, D119, R121, and D135, completely inactivates the enzyme. In addition, D119 and R121 form a hydrogen-bonding network of interactions between the insertion helix and residues 182–184 of the C-terminal domain of monomer 2. D119OD2 is positioned to make hydrogen bonding interactions with T182N, R121NH1 and, in a through-water interaction, the carbonyl oxygen atoms of 182–184. R121NH1 also interacts with the same water molecule and makes a hydrogen bond with P184O. The other major portion of the active site, residues Y86-Q90 on the C terminus of β4, has not been mutated, however, mutation of G91 to alanine resulted in an inactive protein. Possibly G91 is critical for maintaining the conformation of the preceding loop which is in intimate contact with the cofactor.

R154 in monomer 2 extends into the monomer 1 active site. This residue is near the C terminus of the monomer 2 AdoMet-binding domain reaching into the monomer–monomer interface but not involved in specific ionic interactions. The aliphatic portion of the side-chain has a hydrophobic packing role between the side-chains of V137 and L160. *B*-factors for side-chain residues are high (>70Å^2), indicating flexibility. Mutation of R154 to alanine results in inactivation of the protein. As

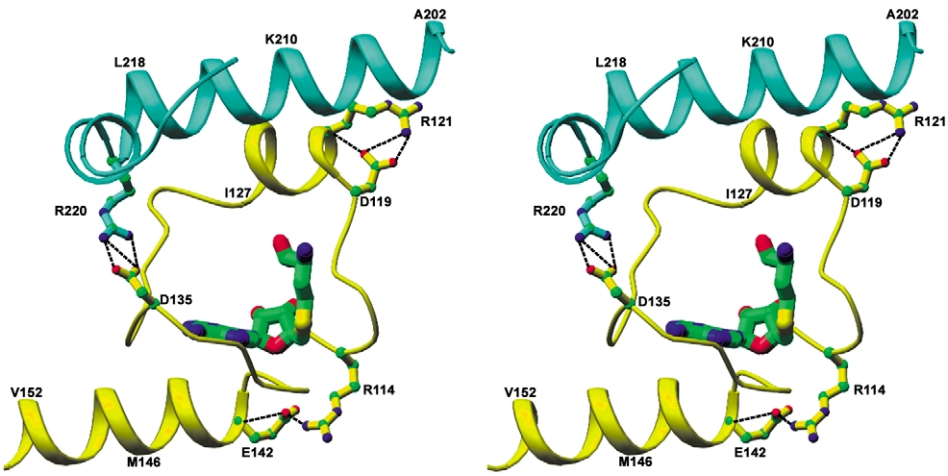

Figure 7. Essential salt bridges between conserved residues defining the AdoMet-binding pocket. Key ionic interactions of the salt bridges are shown as dotted lines.

will be discussed this may be due to crucial interactions with the incoming G37 (Figure 8).

Residues 177–180 of the C-terminal domain of monomer 2 close the active site lying adjacent to the end of the methionine part of AdoHcy. Although these residues are quite close to the cofactor, no specific interactions have been assigned between them and the AdoHcy, since $F_o - F_c$ omit electron density along with high *B*-factors for this portion of the cofactor suggests flexibility. The only mutation carried out in this region, H180, had little effect on cofactor binding.

The C-terminal domain contains another conserved region hypersensitive to alanine mutation. In fact, alanine mutations in residues P184, V192, P193, V195, L196, L197, I204, W207 and R208 lead to complete loss of enzyme activity, while none of them abolish or detectably affect tRNA binding as measured by Plasmon surface resonance. In these experiments (done in triplicate), wild-type enzyme displayed a K_d value of 6.8×10^{-8} M. None of the alanine mutants tested displayed a K_d for tRNA significantly different from this value. These residues, for the most part, form an extended hydrophobic cluster packing against residues 178–180 which are adjacent to the cofactor-binding site. In addition, these hydrophobic residues interact extensively with the halo between Y54 and G57 across the monomer–monomer interface. Significantly, the substitutions W207F or W207H result in

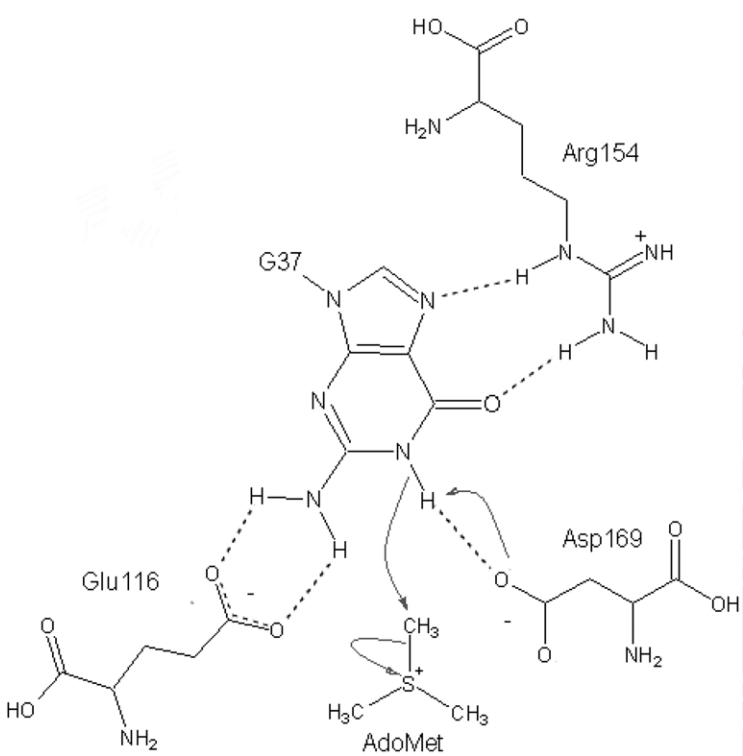

Figure 8. Model of deprotonation of the N1 position of G37 and configuration of G37 bound in the catalytic site.

only a small diminishment of activity, indicating that the effect of W207A is probably to destabilize this hydrophobic pocket. The critical nature of this C-terminal hydrophobic cluster can be illustrated by comparison with a similar tryptophan to alanine mutation of W131 which does not have a crucial effect on activity despite filling a hydrophobic pocket between β6 and α4 in the center of the knot.

With respect to tRNA binding, several other catalytically inactive mutants, G91A, R121A, D135A, and R154A showed no effect on tRNA binding. This strongly suggests that only events after tRNA binding are affected by these mutations.

The C-terminal domain includes four arginine residues, R208, R215, R219, and R220, that are each in an extended conformation directed across the monomer–monomer interface. In every case, the atoms of these arginine side-chains are well determined with *B*-factors for all of them below the average for the protein (<50Å^2 for R208, R215, R219, <41Å^2 for R220). R208 is involved in an extended network of hydrogen bonding with residues 54–56 of the halo (Figure 5) and an R208A mutation inactivates the protein. Interestingly, in the *H. influenzae* structure, this highly conserved R208 does not make any interactions across the interface. R215 makes a through-water hydrogen bond with Q90. The water molecule involved is in the active site within hydrogen bonding distance of the flexible methionine tail of the cofactor. R219 makes a salt bridge with E130, while R220 makes two salt bridges to D135 (NH1 to OD2 and NH2 to OD1). Mutation of R215 or R219 has a modest effect on activity. R220 was not mutated in this study but D135A results in an inactive enzyme.

We have previously indicated that an unstructured peptide linker exists spanning residues 162 to 172 and running over the cofactor-binding site in the extended cleft formed by the monomer–monomer interface. tRNA must contact this peptide region if G37 is to be placed near the catalytic center containing AdoMet. Furthermore, given its location it is possible that this sequence is important for catalysis. Site-directed alanine mutagenesis on one of the conserved residues in this site, F171, shows that catalytic activity is completely eliminated with the loss of the phenylalanine. Although alanine mutagenesis of a large hydrophobic residue could be expected to disrupt important hydrophobic interactions, this phenylalanine is in a disordered loop region in the absence of tRNA. This result clearly demonstrates the essential nature of this region and implies that key base stacking interactions might occur once tRNA is bound over buried AdoMet. Also potentially important in the unstructured region, is residue D169 which may be involved in the deprotonation of N1 of G37 in some enzymes. Residue 169 is conserved in most bacterial genes as either an aspartate or a glutamate residue. Alanine mutagenesis of D169 leads to complete inactivation of enzyme activity; however, mutagenesis to glutamate has very little effect on catalytic activity. These results are consistent with the idea that D169 may be the key residue in the catalytic center of *E. coli* TrmD which is required for N1G37 deprotonation. A number of alanine substitutions around the strongly conserved consensus sequence PYGGG were made which resulted in little or no TrmD protein production. These included P10A, Asp51Ala, Y54A, G56A, and G57A. All except Pro10 are in or near the Halo region that appears to be involved in subunit–subunit interactions. We suggest that this may indicate that when dimer formation is disrupted *in vivo*, protein might be rapidly degraded and/or sequestered in inclusion bodies.

Finally, biochemical characterization of C-terminal truncation mutants shows that at least a portion of the C-terminal domain is required for tRNA binding. Deletion of residues from P194 to the C terminus removes a large part of the C-terminal domain including the large hydrophobic cluster of residues 194–208. This completely inactivates the enzyme as well as eliminating tRNA binding. The loss of tRNA binding is shown by measuring competitive inhibition with wild-type enzyme methylation of tRNA. In these experiments, a tenfold excess (40 μg) of truncated protein (Δ194–255) has no measurable effect on a standard assay, whereas the same amount of protein from two other catalytically inactive mutants, P184A and L196A retains tRNA-binding activity and produces profound competitive inhibition. In contrast, deletion of only 21 residues from the carboxy terminus of the enzyme, residues 236–255, has no measurable effect on activity. This deletion removes one-half of the final alpha helix in the monomer, indicating that it does not play a critical role in enzymatic activity. This is consistent with the observation that TrmD sequences after residue 220 are not conserved in bacteria (Figure 4).

Additional significant alanine mutations are G59A, and D128A. G59 lies on the side of the halo that is distant from the monomer–monomer interface. D128 is found at the N-terminal end of β6 and makes both backbone and side-chain hydrogen bonds to the backbone amide and carbonyl of V84. While not located near the cofactor-binding site nor directly involved in the monomer–monomer interface, both of these mutations inactivate the enzyme.

Discussion

Modification of tRNA is an integral part of its posttranscriptional maturation and involves a large number of enzymes of which only a limited number have been identified and characterized. Currently, little structural information is available concerning tRNA modification enzymes and only four crystal structures have been determined to

date,[15–18], including that of *H. influenzae* TrmD. The current work represents a step further in the knowledge of knotted tRNA methyltransferases as we have determined the crystal structure of *E. coli* TrmD with bound AdoHcy at 2.5 Å resolution and, for the first time, report the results on an extensive mutational analysis of critical sites in the enzyme.

Like *H. influenzae* TrmD,[15] the *E. coli* protein is a homo-dimer in which each monomer contains two distinct domains that do not interact with each other, connected by a flexible linker. Both monomers interact in an antiparallel manner to create the catalytic site of the enzyme and contain deep trefoil peptide knots that form the AdoMet-binding pocket.

A number of knotted structures from members of the SPOUT family have been recently determined (*H. influenzae* YibK[12]), *M. thermoautotrophicum* MT1,[13] *T. thermophilius* RrmA,[11] *E. coli* RlmB[10] and *H. influenzae* TrmD.[15] They all have in common the presence of a non-canonical AdoMet-binding site within their peptide knots and the fact that they are dimeric structures. Based on these common features, Ahn *et al*.[15] have proposed a "SPOUT class of methyltransferase fold", very distinct from the consensus fold. Beyond these two shared features, the structures from members of this class vary considerable and little homology with TrmD exists. Only one of these, *H. influenzae* YibK was co-crystallized with AdoHcy.[12] The YibK enzyme is dimeric but, unlike TrmD, does not display a separate tethered C or N-terminal domain and it has been proposed that this missing domain might be provided by another protein. In addition, the subunit interfaces appear quite different. A second member of this family, MT1, a conserved hypothetical protein,[13] consists of a main domain and a small auxiliary domain of which only the former has structural similarities to the N-terminal region of TrmD. Its structural similarity to CspA, an RNA chaperone, and to ribosomal protein S17,[13] together with the presence of a SPOUT methyltransferase fold[15] suggests that it could function as a RNA methyltransferase. Two other proteins from the same class, RlmB[10] and RrmA,[11] are rRNA methyltransferases and display separate, N-terminal putative rRNA-binding domains, structurally quite different from the C-terminal domain of TrmD. These domains show structural homology to *H. marimortui* L7/L12,[19] and L30 from *Saccharomyces cereviciae*.[20] Thus, it appears that this RNA-binding domain may have evolved from rRNA proteins and later fused to a primitive AdoMet binding catalytic domain. These differences between TrmD and other RNA methyltransferases from the SPOUT family may reflect structural requirements for interaction with different RNA targets.

The dimeric structure of TrmD seems to be fundamental to the formation of the catalytic site of the enzyme. Both subunits participate in forming the deep cleft into which AdoMet is buried. The anticodon loop of tRNA must at some stage be precisely positioned at the bottom of the cleft so that G37 can ultimately be methylated.

The AdoMet-binding region is exquisitely sensitive to mutation. It is easy to rationalize the profound effects of substitutions in the adenine-binding loop, since these residues provide key ionic interactions for AdoMet binding to the cofactor. Equally critical are the residues that make important salt bridging interactions with residues outside the knot structure stabilizing the loop defining the adenine-binding pocket. Clearly, the specific conformation of the loop must be maintained for optimal activity. A more recent mutagenesis study on the AdoMet-binding pocket of ErmC, an example of the canonical methyltransferase fold, revealed that only two substitutions out of eight studied significantly affected catalytic efficiency.[21] This suggests that the AdoMet-binding domain of this class of methytransferases may be much less sensitive to structural changes.[22]

Also important are the effects of mutations in the SpoU motif II-insertion helix region that precedes the β6 strand. Perhaps these residues are critical for the knot or adenine loop conformation, but certainly the mutagenesis results suggest that the hydrogen-bonding network from the insertion helix to residues across the dimer interface, and thus formation of the dimer, is required for activity. Residues 180 to 208 of the C-terminal domain are, like the α5 insertion helix region, hypersensitive to mutation. With the exception of R208, these residues are, in general, quite hydrophobic and constitute the structural core of the C-terminal domain. The fact that deletion of this region severely affects tRNA binding strongly suggests that this domain of TrmD is directly involved in the primary interaction with tRNA, unlike the ErmC protein where alanine mutagenesis showed the C-terminal domain may not contain key residues required for tRNA binding.[22] Alternatively, deletion of the hydrophobic core could destabilize the domain fold and destroy key tRNA binding determinants in the tRNA-binding cleft. Interestingly, none of the single alanine mutations in this region abolish tRNA binding but do abolish catalytic activity. Therefore, we propose that this domain must also play an important role in some post-tRNA binding event required for catalysis.

It is possible that the carboxy terminal region may alone bind tRNA. In this context, we have recently determined that the TrmD enzyme binds several non-substrate tRNA species with affinities equal to that of $tRNA_1^{Leu}$ (our unpublished results). The C-terminal domain might be a general tRNA-binding protein, which makes the initial binding event perhaps at the strongly conserved hinge or core region of tRNA.

We are struck by the unusual positioning of the unstructured 11 residues, which connect the C-terminal domains with the AdoMet-binding

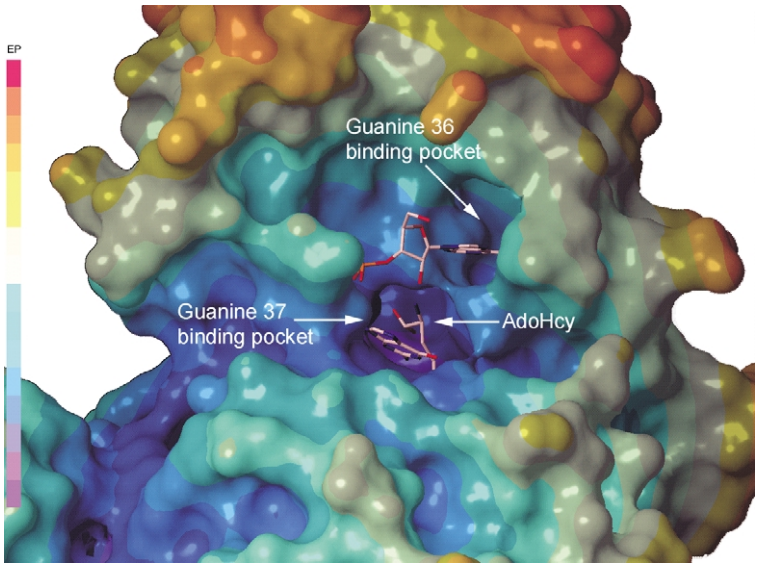

Figure 9. G36pG37-binding pockets. MOLCAD electrostatic potential Figure representing Gasteiger–Huckel charges. GpG was extracted from PDB entry 1EHZ (yeast tRNA Phe[16] position 18 and 19), and merged into TrmD surface. Scale (left) represents charges ranging from negative (blue) to positive (red). The Figure was created with the SYBYL† graphics program.

domains. Given the restraints of distance and the likely path of the peptide, these residues must be located very near, if not directly over, the bound AdoMet during catalysis. Four of the residues (S165, D169, S170, and F171) are conserved in TrmD proteins so far sequenced. It is therefore likely that these residues might play an active role in the catalytic center. The conserved phenylalanine residue could, for example, provide key stacking interactions required for the final positioning of G37. The importance of this sequence is underscored by the finding that alanine mutagenesis of either D169 or F171 results in total inactivation of enzymatic activity. An examination of position 169 in many species of bacterial TrmD genes reveals all exhibit a glutamate or aspartate residue at this position. Given the similar pK_a of aspartate and glutamate functional carbonyl groups, we proposed that this residue might serve well in position 169 as a general base catalyst. This idea is supported by our finding that enzyme with glutamate at this position is quite active.

Why should such an important region be relatively mobile, and therefore unstructured? Perhaps this is an essential characteristic of the G37-binding pocket, facilitating the formation of structures required for catalysis, since the GpG bases apparently must be moved out of the anticodon loop towards bound AdoMet at the bottom of the cleft. N1 deprotonation requires a charged amino acid residue close to AdoMet during catalysis. Several strongly charged residues are located near the bound AdoMet(hcy). Two (E116 and R154) are positioned near AdoHcy in the active site and are attractive candidates for ionic interaction with incoming G37 residues (Figure 8). Alanine mutagenesis of R154 completely inactivates enzyme activity, whereas the E116A mutation has a measurable effect on AdoMet K_m values but does not inactivate the enzyme. This indicates that this residue may not be involved directly in G deprotonation but could play a role in stabilizing G37 during deprotonation. Although the position of D169 was not revealed by the crystal structure, the mutant D169A is completely inactive while D169E is active (Table 2). Given the distance restraints of the linker peptide it is likely that the D169 residue can be positioned near N1 of G37 during catalysis. In addition, Ahn *et al.*[15] report that this residue is indeed near bound AdoMet and propose a role for this residue in catalysis. Thus, the conserved aspartate residue at position 169 could take part in the deprotonation of the N1 position of G37. Another residue could be a histidine residue which often are involved in such reactions in enzyme catalysis. Only H180 is near the catalytic center but does not play a critical role in activity since proteins containing an alanine substitution at H180 remain active. Figure 8 presents a model (similar to that of Ahn *et al.*[15]) for deprotonation of the N1 position of G by aspartate or glutamate based on the mutagenesis data and the position of these residues in the structure of TrmD. Given the potential structural lability of this region, a tRNA–enzyme co-crystal may ultimately be required to determine the exact structure of the catalytic complex.

Both YibK and TrmD display an adenine-binding loop which extends from the central knotted structure and embraces deeply buried AdoHcy residues. This raises questions about how AdoMet might efficiently enter such a deep pocket and be effectively removed during catalysis. It is possible that these enzymes may undergo isomerization during catalysis in order to cycle AdoMet in and out of the catalytic center. Perhaps the unique knotted catalytic center is required for this isomerization. In the case of YibK, a structure with and

† Tripos-Software. SYBIL 6.8, 1699 South Hanley Rd, St Louis, MO 63144, USA.

without bound AdoHcy has been determined and substantial changes in two loop structures near the catalytic center were observed,[12] suggesting allosteric changes may accompany catalysis. However, in the *H. influenzae* structures, the apo form is virtually identical with the cofactor bound forms except in the region of the halo and parts of the C-terminal domain.

An examination of the cleft region between the AdoMet-binding domain and the C-terminal domain of the monomer reveals another hydrophobic depression adjacent to the buried AdoHcy. Figure 9 illustrates GpG modeled into these pockets on the dimer surface. Given the structure of tRNA and the depth of the cleft, we imagine that the bases of G36pG37 must move into the correct position for catalysis. Perhaps adjacent residues are involved in this movement and may even function as a special-binding site for the G36 "recognition" residue. Alternatively, this hydrophobic region might be essential for interaction with other crucial site(s) in tRNA also required for catalysis.

We imagine that the overall reaction might proceed as follows; tRNA is first non-specifically bound to enzyme, perhaps involving surface C-terminal sequences. We propose, based on earlier biochemical evidence,[2] that the C-terminal domain may recognize and bind the core (hinge) structure of the general tRNA molecule. Next, tRNA might be positioned in the cleft and, if the identity element G36pG37 is present, G36pG37 might be moved or perhaps "flipped" into the dual pocket catalytic center for subsequent methylation by bound AdoMet. If a G is not found at position 36, then a stable complex of G37 near AdoMet cannot form and catalysis does not proceed.

Clearly much has to be learned about this interesting group of enzymes that act on tRNA. These studies represent a thorough attempt at understanding how this novel class of knotted methyltransferases may actually work, by providing data that define the residues involved in catalytic activity, AdoMet and tRNA binding. A model that explains how the enzymatic reaction proceeds is proposed (Figure 8); the stage is set for more complete studies aimed at understanding how tRNA is bound and how G37 N1 methylation occurs.

Materials and Methods

Data collection

Crystals were flash cooled by submersion in liquid nitrogen after a 30-second soak in a cryo solution of 20% (v/v) glycerol made up in the reservoir solution. Data were collected at beam line 17-ID in the facilities of the Industrial Macromolecular Crystallography Association Collaborative Access Team (IMCA-CAT) at the Advanced Photon Source. High resolution data were collected to 2.5 Å resolution and MAD data from selenomethionyl TrmD crystals were collected at three wavelengths (see Table 1 for statistics) to 3.2 Å resolution. Data were processed and scaled using the DENZO and HKL2000[23] software packages.

Structure solution

The positions of the selenium atoms were identified by using the software package SOLVE.[24] SOLVE found seven of the eight selenium sites with a Z score of 28 and a mean figure of merit of 0.38. Solvent flattening was applied using the program RESOLVE.[25] SOLVE and RESOLVE were run in space groups $P321$, $P3_121$, and $P3_221$ to determine the correct space group of the crystal. Exceptionally well-connected and traceable maps indicated the correct space group, $P3_221$ that contains one molecule in the asymmetric unit.

Model building and refinement

The structure was built into the 3.2 Å MAD solvent flattened map using the molecular graphics program O.[26] Electron density was excellent and virtually all of the structure could be traced at this resolution. Density for the cofactor was also obvious. CNX (Accelrys Inc.) was used for all refinement. Rigid body refinement and simulated annealing at 3.2 Å without the cofactor resulted in a free R-factor of 41%. High-resolution data were then added to the refinement and cycles of refinement alternated with model building. Final refinement statistics are summarized in Table 1. Structural Figures were prepared with the following programs: Bobscript[27] or Molscript[28] and Raster3D,[29,30] SYBIL,† and MOLCAD.[31]

Sequence alignment

Translated open reading frames (ORFs) and complete DNA sequences were searched using the program BLASTP.[32] Individual protein datasets were aligned using the program CLUSTALW v1.8 with default settings.[33]

Enzyme cloning and expression

The initial expression, and purification of the *E. coli* TrmD protein has been reported elsewhere.[2] Yields have been subsequently improved by the insertion of *E. coli trmd* into the pet28 T7 expression vector (Novagen) using *Eco* RI and *Nde* I restriction endonuclease sites for directional cloning. Constructs were transformed into the BL21 (DE3) host for T7 expression. All subsequent studies have been done using this construction.

Enzyme purification

Typically, 2 l of pre-warmed L-broth were inoculated with the appropriate strain outlined above and grown at 37 °C to mid log($A_{600} = 0.6$). Cells were induced with 1 mM IPTG, then shaken at 18 °C overnight. Cells were harvested, resuspended in 50 mM phosphate buffer (pH 7.5), 250 mM NaCl, 10 mM imidazole, then lysed in a French press. After centrifugation, soluble crude extract was passed over a 5 ml bed of Qiagen® Ni-NTA nickel substituted agarose equilibrated with the above buffer. After extensive washing with the above buffer, the

† Tripos-Software. SYBIL 6.8, 1699 South Hanley Rd, St Louis, MO 63144, USA.

column was washed in a buffer with lower salt (50 mM phosphate buffer (pH 7.5), 50 mM NaCl, 10 mM imidazol), and eluted with a low salt buffer (50 mM phosphate buffer (pH 7.5), 50 mM NaCl, 250 mM imidazol).

Eluted enzyme was further purified through a MonoQ column using a Beckman High Performance Liquid Chromatography (HPLC) system. Enzyme was eluted using a linear salt gradient (50 mM to 1.2 M NaCl in 50mM Tris–HCl, pH 7.8) at 2 ml/minute. Fractions containing purified protein were pooled, concentrated to 10 mg/ml and stored under nitrogen in 40% glycerol at −20 °C.

Preparation of selenomethionine substituted TrmD

Two 30 ml starter cultures were grown overnight with kanamycin 50 μg/ml, pelleted, resuspended in 30 ml pre-warmed M9 minimal media, and allowed to grow for 15 minutes at 37 °C before being added to 2 l of pre-warmed M9 minimal media.[34] Cells were grown to mid log phase after which the following amino acids were added: 100 mg L-Lys, 100 mg L-Phe, 100 mg L-Thr, 50 mg L-Ile, 50 mg L-Leu, 50 mg L-Val, and 50 mg Selenomethionine from Sigma®. Cells were shaken for 15 minutes at 37 °C, then induced with 1 mM IPTG, and shaken at 18 °C overnight. Selenomethionyl protein was bound to a nickel affinity column and purified essentially as outlined above.

Mutagenesis

Alanine site-directed mutagenesis was carried out using the Stratagene "Quickchange" kits and protocols. All oligonucleotides for mutagenesis were prepared with appropriate pairing stability and were usually 24 to 30 nt in length. tRNA synthesis and enzyme assays were carried out as described.[2,6]

Gel filtration

A TSK HPLC gel filtration column (Tohaus) was calibrated using gel filtration molecular mass markers from Sigma (MW-GF-200). The enzyme was dialyzed against 50 mM Tris–HCl (pH 7.5) 100 mM KCl, and run at 1 ml/minute with the same buffer through the column. A Beckman Gold HPLC system was employed.

Enzyme-tRNA binding experiments using plasmon surface resonance

A solution of a synthetic DNA oligonucleotide (5′AAAATTCGAGCTCG GTACCCGGGGATCC*3′) wherein the final 3′nucleotide(C*) was biotinylated was pumped slowly (5 μl/minute) across a BIAcoresensor chip-SA surface in buffer A (10 mM Hepes (pH 7.4), 150 mM NaCl, 10 mM $MgCl_2$) at 37 °C.

Next, a solution of $tRNA_1^{Leu}$ (3.8 μg/ml, in buffer A) was pumped over the Biocore™ streptavidin chip at 2 μl per minute for 30 minutes using the PharmaciaX "Biocore" instrument. This tRNA contained a 24-nt 3′ RNA extension sequence complementary to the first 24 bases of the bound DNA oligonucleotide. Buffer A was pumped over the chip until all unbound tRNA had been removed as evidenced by a stable RU (relative units) reading. In order to assess the kinetics of binding, a TrmD enzyme solution (3 μg/ml in buffer A) was pumped over the streptavidin chip with bound tRNA at 20 μl/minute for 60 minutes, then buffer A was pumped over the chip until no further changes in RU values were obtained. K_d values were calculated using the software of the Pharmacia X instrument. The DNA substituted DNA chips could be reused several times after washing with 0.1N NaOH.

Protein Data Bank accession numbers

A PDB file has been posted and has been assigned the accession code: 1P9P.pdb.

Acknowledgements

Data were collected at beamline 17-ID in the facilities of the Industrial Macromolecular Crystallography Association Collaborative Access Team (IMCA-CAT) at the Advanced Photon Source. These facilities are supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with Illinois Institute of Technology (IIT) executed through ITT's Center for Synchrotron Radiation Research and Instrumentation. The US Department of Energy, Basic Energy Sciences, Office of Science supported use of the Advanced Photon Source. Portions of this work were supported by grants to W.M.H. from the National Science foundation and by GlaxoSmithKline.

References

1. Grosjean, H. & Benne, B. (1998). *Modification and Editing of RNA*, ASM Press, Washington, DC.
2. Redlak, M., Andraos-Selim, C., Giege, R., Florentz, C. & Holmes, W. M. (1997). Interaction of tRNA with tRNA (guanosine-1)methyltransferase: binding specificity determinants involve the dinucleotide G36pG37 and tertiary structure. *Biochemistry*, **36**, 8699–8709.
3. Li, J. N. & Bjork, G. R. (1999). Structural alterations of the tRNA(m1G37)methyltransferase from *Salmonella typhimurium* affect tRNA substrate specificity. *RNA*, **5**, 395–408.
4. Cheng, X. & Blumenthal, R. M. (1999). S-*Adenosylmethionine-Dependent Methyltrasferases: Structures And Functions*, World Scientific, New Jersey.
5. Holmes, W. M., Andraos-Selim, C., Roberts, I. & Wahab, S. Z. (1992). Structural requirements for tRNA methylation. Action of *Escherichia coli* tRNA-(guanosine-1)methyltransferase on tRNA(1Leu) structural variants. *J. Biol. Chem.* **267**, 13440–13445.
6. Gabryszuk, J. & Holmes, W. M. (1997). tRNA recognition for modification: solution probing of tRNA complexed with *Escherichia coli* tRNA (guanosine-1) methyltransferase. *RNA*, **3**, 1327–1336.
7. Cheng, X. & Roberts, R. J. (2001). AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucl. Acids Res.* **29**, 3784–3795.
8. Martin, J. L. & McMillan, F. M. (2002). SAM (dependent) I AM: the *S*-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* **12**, 783–793.
9. Anantharaman, V., Koonin, E. V. & Aravind, L.

(2002). SPOUT: a class of methyltransferases that includes spoU and trmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. *J. Mol. Microbiol. Biotechnol.* **4**, 71–75.
10. Michel, G., Sauve, V., Larocque, R., Li, Y., Matte, A. & Cygler, M. (2002). The structure of the RlmB 23 S rRNA methyltransferase reveals a new methyltransferase fold with a unique knot. *Structure (Camb)*, **10**, 1303–1315.
11. Nureki, O., Shirouzu, M., Hashimoto, K., Ishitani, R., Terada, T., Tamakoshi, M. *et al.* (2002). An enzyme with a deep trefoil knot for the active-site architecture. *Acta Crystallog. sect. D*, **58**, 1129–1137.
12. Lim, K., Zhang, H., Tempczyk, A., Krajewski, W., Bonander, N., Toedt, J. *et al.* (2003). Structure of the YibK methyltransferase from *Haemophilus influenzae* (HI0766): a cofactor bound at a site formed by a knot. *Proteins: Struct. Funct. Genet.* **51**, 56–67.
13. Zarembinski, T. I., Kim, Y., Peterson, K., Christendat, D., Dharamsi, A., Arrowsmith, C. H. *et al.* (2003). Deep trefoil knot implicated in RNA binding found in an archaebacterial protein. *Proteins: Struct. Funct. Genet.* **50**, 177–183.
14. Jacobs, S. A., Harp, J. M., Devarakonda, S., Kim, Y., Rastinejad, F. & Khorasanizadeh, S. (2002). The active site of the SET domain is constructed on a knot. *Nature Struct. Biol.* **9**, 833–838.
15. Ahn, H. J., Kim, H. W., Yoon, H. J., Lee, B. I., Suh, S. W. & Yang, J. K. (2003). Crystal structure of tRNA(m1G37)methyltransferase: insights into tRNA recognition. *EMBO J.* **22**, 2593–2603.
16. Romier, C., Ficner, R., Reuter, K. & Suck, D. (1996). Purification, crystallization, and preliminary X-ray diffraction studies of tRNA-guanine transglycosylase from *Zymomonas mobilis*. *Proteins: Struct. Funct. Genet.* **24**, 516–519.
17. Hoang, C. & Ferre-D'Amare, A. R. (2001). Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell*, **107**, 929–939.
18. Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S. & Yokoyama, S. (2003). Alternative tertiary structure of tRNA for recognition by a post-transcriptional modification enzyme. *Cell*, **113**, 383–394.
19. Leijonmarck, M. & Liljas, A. (1987). Structure of the C-terminal domain of the ribosomal protein L7/L12 from *Escherichia coli* at 1.7 Å. *J. Mol. Biol.* **195**, 555–579.
20. Mao, H. & Williamson, J. R. (1999). Local folding coupled to RNA binding in the yeast ribosomal protein L30. *J. Mol. Biol.* **292**, 345–359.
21. Maravic, G., Feder, M., Pongor, S., Flogel, M. & Bujnicki, J. M. (2003). Mutational analysis defines the roles of conserved amino acid residues in the predicted catalytic pocket of the rRNA:m6A methyltransferase ErmC′. *J. Mol. Biol.* **332**, 99–109.
22. Maravic, G., Bujnicki, J. M., Feder, M., Pongor, S. & Flogel, M. (2003). Alanine-scanning mutagenesis of the predicted rRNA-binding domain of ErmC′ redefines the substrate-binding site and suggests a model for protein–RNA interactions. *Nucl. Acids Res.* **31**, 4941–4949.
23. Otwinowski, Z. & Minor, W. (1992). Processing of X-ray Diffraction Data Collected in Oscillation Mode, *Methods in Enzymology*, Volume 276: *Macromolecular Crystallography* (Carter, C. W. Jr & Sweet, R. M., eds.), part A, pp. 307–326, Academic Press, New York.
24. Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallog. sect. D*, **55**, 849–861.
25. Terwilliger, T. C. (2001). Map-likelihood phasing. *Acta Crystallog. sect. D*, **57**, 1763–1775.
26. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
27. Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 112–3, 132–4.
28. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
29. Merritt, E. A. & Bacon, D. J. (1997). Raster3D photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524.
30. Merritt, E. A. M. & Murphy, M. E. P. (1994). Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallog. sect. D*, **50**, 869–873.
31. Brickmann, J. G. T., Heiden, W. & Waldherr-Techner, M. (1992). Examining and manipulating molecular surfaces: generation and manipulation of solid models with the MOLCAD molecular graphics system. *J. Mol. Graph.* **10**, 35.
32. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402.
33. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
34. Maniatis, T., Fritch, E. F. & Sambrook, J. (1982). *Molecular Cloning*, 1st edit., vol. 1, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Edited by J. Doudna

(Received 29 May 2003; received in revised form 28 August 2003; accepted 8 September 2003)