RimO, a MiaB-like enzyme, methylthiolates the universally conserved Asp88 residue of ribosomal protein S12 in *Escherichia coli*

Brian P. Anton*†, Lana Saleh*, Jack S. Benner*, Elisabeth A. Raleigh*, Simon Kasif‡§, and Richard J. Roberts*¶

*New England Biolabs, Ipswich, MA 01938; and †Bioinformatics Program, ‡Department of Biomedical Engineering, and §Center for Advanced Genomic Technology, Boston University, Boston, MA 02215

Edited by Eugene V. Koonin, National Institutes of Health, Bethesda, MD, and accepted by the Editorial Board December 13, 2007 (received for review September 11, 2007)

Ribosomal protein S12 undergoes a unique posttranslational modification, methylthiolation of residue D88, in *Escherichia coli* and several other bacteria. Using mass spectrometry, we have identified the enzyme responsible for this modification in *E. coli*, the *yliG* gene product. This enzyme, which we propose be called RimO, is a radical-*S*-adenosylmethionine protein that bears strong sequence similarity to MiaB, which methylthiolates tRNA. We show that RimO and MiaB represent two of four subgroups of a larger, ancient family of likely methylthiotransferases, the other two of which are typified by *Bacillus subtilis* YqeV and *Methanococcus jannaschii* Mj0867, and we predict that RimO is unique among these subgroups in its modification of protein as opposed to tRNA. Despite this, RimO has not significantly diverged from the other three subgroups at the sequence level even within the C-terminal TRAM domain, which in the methyltransferase RumA is known to bind the RNA substrate and which we presume to be responsible for substrate binding and recognition in all four subgroups of methylthiotransferases. To our knowledge, RimO and MiaB represent the most extreme known case of resemblance between enzymes modifying protein and nucleic acid. The initial results presented here constitute a bioinformatics-driven prediction with preliminary experimental validation that should serve as the starting point for several interesting lines of further inquiry.

methylthiotransferase | posttranslational modification | radical-SAM protein

The translational apparatus undergoes numerous posttranscriptional and posttranslational modifications of its RNA and protein components, respectively, and the enzymes responsible for many of these modifications have been identified in recent years (1, 2). One modification for which the responsible enzyme is currently unknown is the methylthiolation of the β-carbon of residue D88 of ribosomal protein S12 in *Escherichia coli* (Fig. 1*A*) (3). Intriguingly, D88 is strictly conserved in all S12 homologs from bacteria, archaea, and eukaryotes. No spontaneous D88 mutants have been identified, and attempts to alter the residue by site-directed mutagenesis have failed where mutation of surrounding residues succeeded (4), suggesting that this aspartic acid residue serves an essential function. The methylthiol modification has also been observed in S12 from *Rhodopseudomonas palustris* and *Thermus thermophilus* (3, 5, 6), but it appears not to be present in cytoplasmic ribosomes from rat and human cells (7, 8). In *T. thermophilus*, S12 P90R and P90W mutants have been shown to be deficient in the D88 modification, presumably due to steric hindrance of the modifying enzyme by the bulky sidechains (4). Thus, although D88 itself appears to be universal and essential, the methylthiol modification is not.

The function of this modification remains unclear, but its location suggests a possible role in decoding or translocation. The solved crystal structure of S12 from *T. thermophilus* reveals that D88 sits within one of two highly conserved loops that project toward the acceptor site of the ribosome (9). Mutations in surrounding residues K87, L89, P90, G91, and R93 have all been shown to confer resistance to or dependence on streptomycin, and these mutations often lead to less active but hyperaccurate ribosomes (10). Mutations affecting decoding accuracy are believed to do so by altering the stability of the interactions that S12 makes with important structural elements of the 16S rRNA. S12 is also known to serve as a check on spontaneous (independent of EF-G/GTP) translocation by the ribosome (11), and mutation at K87, adjacent to the modified residue, abolishes this function (12).

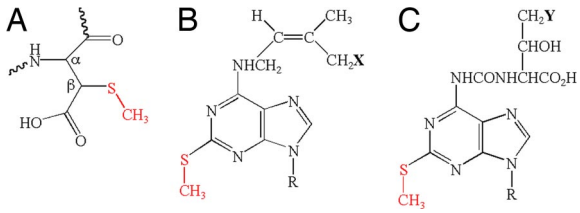

Fig. 1. Schematic structures of the five known naturally occurring methylthiol residues from proteins and nucleic acids. Methylthiol moieties are shown in red. R, ribose. (*A*) β-methylthio-aspartic acid. (*B*) ms^2i^6A when X = H and ms^2io^6A when X = OH. (*C*) ms^2t^6A when Y = H and ms^2hn^6A when Y = CH_3.

The modification of S12 resembles certain tRNA modifications. The second step in 2-methylthio-N^6-isopentenyladenosine (ms^2i^6A; Fig. 1*B*) biosynthesis is the methylthiolation of C^2 of N^6-isopentenyladenosine (i^6A). This reaction has been shown to be catalyzed by the protein MiaB in *E. coli* and *Thermotoga maritima* (13–15). MiaB is a radical-*S*-adenosylmethionine (SAM) protein, and like most members of this superfamily, it uses a reducing equivalent from a $[4Fe\text{-}4S]^{1+}$ cluster to cleave SAM to generate methionine and a 5′-deoxyadenosyl radical (5′-dA·) (13, 14, 16). In the case of MiaB, 5′-dA· is believed to abstract a hydrogen atom from adenosine C^2, generating a reactive substrate radical that is amenable to sulfur insertion. In addition to the radical-generating [4Fe-4S] cluster, a second essential [4Fe-4S] cluster, speculated to be involved with the thiolation reaction, has been identified near the N terminus of MiaB (17). Thiolation is followed by methylation, with a presumed 2-thio-N^6-isopentenyladenosine (s^2i^6A) intermediate that

Author contributions: B.P.A., L.S., E.A.R., S.K., and R.J.R. designed research; B.P.A., L.S., and J.S.B. performed research; B.P.A. and L.S. analyzed data; and B.P.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. E.V.K. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

¶To whom correspondence should be addressed. E-mail: roberts@neb.com.

This article contains supporting information online at www.pnas.org/cgi/content/full/0708608105/DC1.

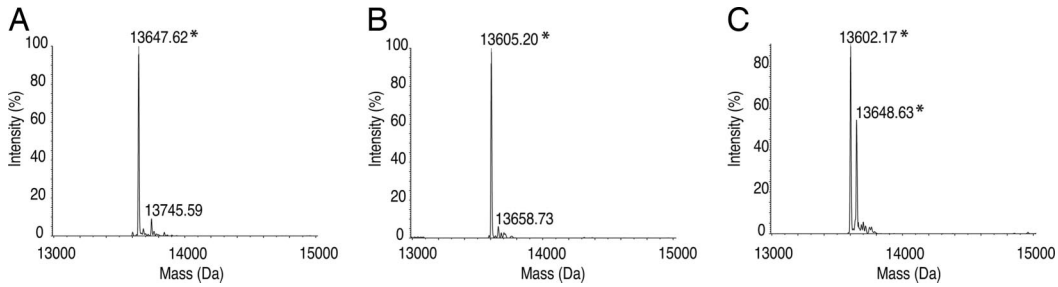

Fig. 2. Deconvoluted ESI-TOF MS of ribosomal protein S12 from various *E. coli* strains. S12 peaks are indicated by asterisks (*). (*A*) Protein from wild-type strain MG1655. (*B*) Protein from mutant strain FB23662 (*yliG*$^-$). (*C*) Protein from mutant strain ER3051/pLIT-yliG.

has yet to be isolated (18). Evidence suggests that the source of the methyl group in the final product, ms^{2}i^{6}A, is SAM (presumably a distinct SAM molecule from that used for the radical generation) and that the methyltransferase (MTase) activity resides in MiaB (16) despite the fact that MiaB shares no protein sequence motifs characteristic of SAM-dependent MTases.

The family containing MiaB and its relatives, often referred to as the "MiaB-like proteins" (13, 19, 20), forms a subclass of the radical-SAM superfamily. At least one MiaB-like protein is encoded in most sequenced genomes from all three domains of life, indicating an origin before the last universal common ancestor (21). It has been assumed that the proteins closely related to MiaB are all involved with tRNA methylthiolation, because ms^{2}i^{6}A is one of several related methylthiol-containing tRNA modifications that differ only in the substituent at the N^6 position (Fig. 1 *B* and *C*) (22). However, none of these paralogous proteins has been functionally characterized.

Here we report the identification and preliminary experimental validation of a novel member of the MiaB-like family, encoded by the *E. coli* ORF *yliG*, showing that it is responsible for β-methylthiolation of the S12 D88 residue. We also analyze the MiaB-like protein family phylogenetically and show it contains only four apparent subfamilies, and we offer specific functional predictions for the remaining uncharacterized subfamilies. These predictions account for all experimentally observed methylthiol modifications, and we refer to the MiaB-like family hereinafter as the methylthiotransferase (MTTase) family. The degree of similarity between the *yliG* product and tRNA MTTases is remarkable, and we speculate on the origins of the apparently unique S12 modification. We suggest that the locus name *yliG* be renamed *rimO*, joining others involved with ribosomal modification.

Results

***yliG* Encodes the S12 D88-Modifying Enzyme.** Methylthiolation of tRNA by MiaB involves sulfur insertion into an unreactive C–H bond, a difficult reaction that is achieved by using a 5′-deoxyadenosyl radical to abstract the hydrogen atom from the tRNA substrate (16). We reasoned that S12 methylthiolation probably proceeds by a similar mechanism to the MiaB reaction and so the enzyme responsible should bear detectable sequence similarity to MiaB. We identified a single *E. coli* protein of unknown function, the product of *yliG*, which exhibited strong similarity (BLASTP E = 6e-36) to MiaB along its entire length. Although similarity of this degree would suggest *yliG*, like *miaB*, encoded an RNA-modifying enzyme, no other *E. coli* tRNA or rRNA methylthiolation other than that catalyzed by MiaB has been reported (23). Furthermore, we identified putative *yliG* homologs in the genomes of *T. thermophilus* and *R. palustris*, the only two other organisms in which β-methylthiolation of S12 has been observed (5, 6). Therefore, we considered the possibility that *yliG* encoded the S12 MTTase.

We purified ribosomal proteins from several *E. coli* strains and determined the mass of S12 in each strain using mass spectrometry (MS) (Fig. 2). S12 from the wild type (wt) strain MG1655 was completely modified as evidenced by a *m*/*z* shift of + 46 relative to the expected mass of the unmodified protein (Fig. 2*A*), in agreement with an earlier study (24). In the two *yliG* mutant strains, FB23662 (Fig. 2*B*) and ER3051 (not shown), S12 appeared as a single peak with a mass close to that expected of the unmodified form. To confirm that the loss of modification in these strains was due specifically to the inactivation of *yliG*, we attempted to rescue the phenotype by reintroducing *yliG* to ER3051 on a plasmid, pLIT-yliG, encoding a contitutively expressed His-tagged derivative of *yliG*. In this strain, we observed peaks corresponding to both the modified and unmodified forms of S12 (Fig. 2*C*). The restoration of S12 modification in this strain confirms that *yliG* indeed encodes the S12 D88 MTTase, and we will hereinafter refer to *yliG* as *rimO*. The presence of some unmodified S12 in ER3051/pLIT-yliG suggests that the *rimO* activity in the rescued strain is lower than that in the wt strain, due possibly to the presence of the His-tag, misfolding of the protein or malformation of the required FeS clusters as a result of overexpression, or some combination of these issues, and at this point we have not attempted to distinguish between them.

Growth Properties of the *rimO* Null Mutant. We compared the growth of *E. coli* MG1655 (*rimO*$^+$) and ER3051 (*rimO*$^-$) in liquid culture in Rich medium at 37°C and found that ER3051 grew slightly but reproducibly more slowly than the wt strain [see supporting information (SI) Fig. 4*A*]. However, both strains reached a similar density in stationary phase with a similar number of viable cells remaining after overnight growth (data not shown). Consistent with this, we found that ER3051 formed smaller colonies on solid Rich medium than the wt MG1655 strain. The effect of RimO was accentuated in complementation with a plasmid-borne copy: *rimO*$^-$ (carrying an empty vector) displayed a significant lag in addition to slower growth than *rimO*$^+$ (complementing plasmid with *rimO*) (SI Fig. 4*B*). Taken together, these results suggest that *rimO* confers a selective growth advantage on *E. coli* cells. The high-copy plasmid (a pUC derivative) and/or growth in ampicillin may impose a stress that is particularly poorly compensated for in the absence of *rimO*.

We also examined the growth of ER3051 in the presence of streptomycin, because streptomycin binds the ribosome in close proximity to the D88 residue of S12. A previous study showed no correlation between D88 modification and streptomycin resistance and dependence phenotypes in *T. thermophilus* (4), and we wished to confirm this in *E. coli* with a wild-type S12 background. We compared the growth of *E. coli* MG1655 and ER3051 on a streptomycin gradient on Rich agar and found the two strains comparable, although ER3051 tolerated a slightly higher concentration of streptomycin than the parental strain (data not shown). We suggest this may be due to low-level phosphorylation of streptomycin by the kanamycin phosphotransferase encoded

BIOCHEMISTRY

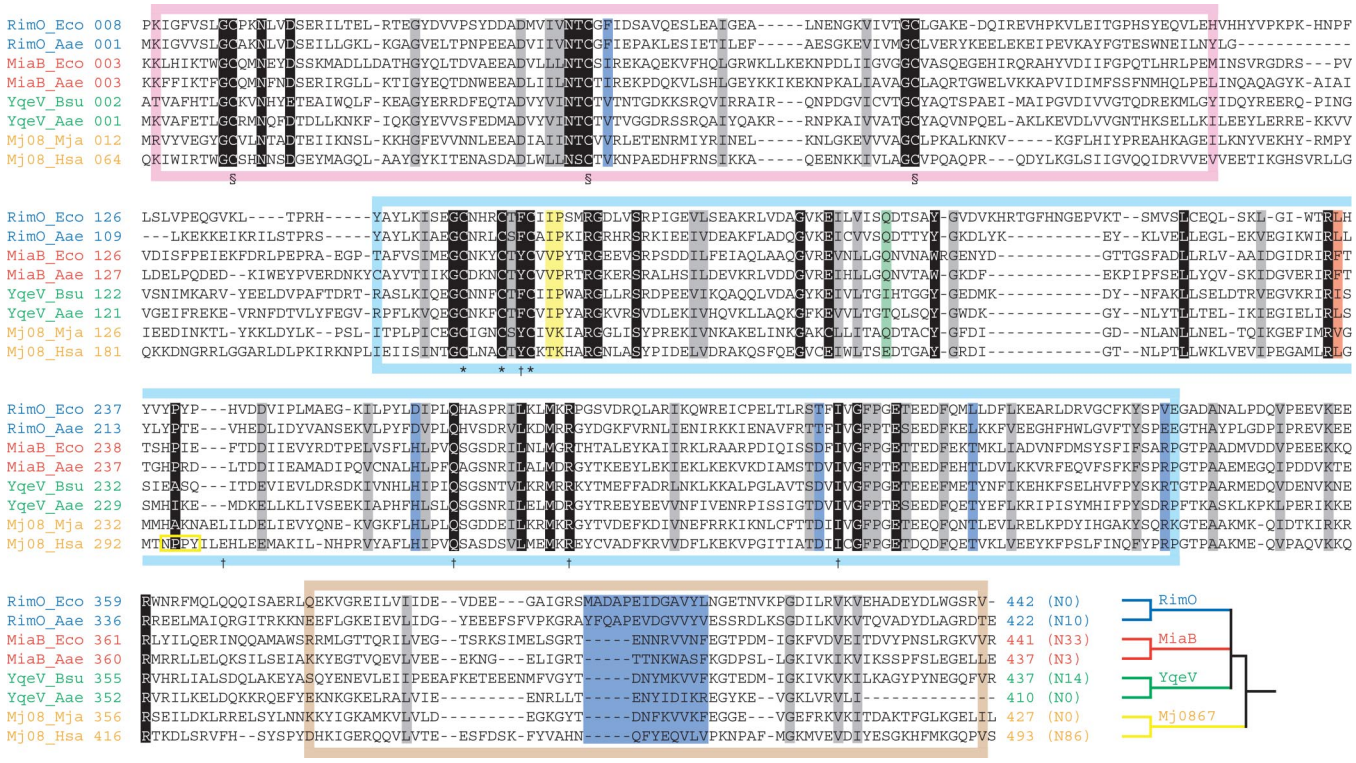

Fig. 3. Multiple sequence alignment of eight diverse MTTases (two from each of the four subfamilies): RimO_Eco (YliG/RimO from *E. coli* K-12; NP_415356), RimO_Aae (aq_849 from *Aquifex aeolicus*; NP_213577), MiaB_Eco (MiaB from *E. coli* K-12; NP_415194), MiaB_Aae (aq_284 from *A. aeolicus*; NP_213198), YqeV_Bsu (BSU25430/YqeV from *B. subtilis*; NP_390421), YqeV_Aae (aq_474 from *A. aeolicus*; NP_213332), Mj08_Mja (Mj0867 from *M. jannaschii*; NP_247862), Mj08_Hsa (CDKAL1 from *Homo sapiens*; NP_060244). The three domains as defined by Pfam are boxed, inclusive of edges: UPF0004 (pink), radical-SAM (cyan), and TRAM (brown). Residues conserved in the larger alignment, listed in SI Table 2, are boxed in black, and additional residues conserved in all eight sequences in this alignment are boxed in gray. (For this purpose, the following residues are treated as equivalent: S = T, E = D, K = R, F = W = Y, and I = L = V.) Subfamily-specific residues as defined in SI Table 3 are boxed in the color of the subfamily that differs from the remaining three: blue for RimO, red for MiaB, green for YqeV, and yellow for Mj0867. The three cysteines that coordinate the [4Fe-4S] cluster in the radical-SAM domain of MiaB are indicated by *, the three cysteines that coordinate the [4Fe-4S] cluster in the UPF0004 domain of MiaB are indicated by §, and residues that may be involved with SAM binding (see SI Table 2) are indicated by †. The NPPY motif conserved among eukaryotic Mj0867 members is boxed in yellow. At the lower right is a schematic phylogenetic tree, based on SI Fig. 5, showing the relationships of these sequences to one another.

by the insertion cassette in ER3051, and not to lack of D88 modification, although this remains to be proven.

Phylogenetic Analysis of the MTTase Family. A Bayesian inference phylogenetic tree of selected MTTase family sequences from all three domains of life is shown in SI Fig. 5. All of the sequences appear to belong to one of four strongly supported clades, three of which contain exclusively bacterial (and in one case, mitochondrial) members and the fourth of which contains exclusively archaeal and eukaryotic members. A neighbor-joining tree on the same sequences also supported the monophyly of each of these four clades as well as a higher-order grouping of the three bacterial clades (data not shown). The first bacterial clade contains *E. coli* RimO as well as sequences from *R. palustris* and *T. thermophilus*, the two other organisms where the S12 D88 modification has been observed. The second bacterial clade contains characterized MiaB sequences from both *E. coli* and *T. maritima*. This group also contains a subclade of sequences from eukaryotic genomes, most of which contain N-terminal signal sequences predicted to direct localization to the mitochondrion (25). The third bacterial clade contains no characterized members, and we designate it the "YqeV" clade based on its *Bacillus subtilis* member. Likewise, the archaeo-eukaryotic clade contains no characterized members, and we designate it the "Mj0867" clade based on its *Methanococcus jannaschii* member.

Using protein sequence motifs characteristic of each clade as a guide, we examined all completely sequenced genomes available at the time of this writing (389 bacterial, 29 archaeal, and 40 eukaryotic) to obtain more complete phylogenetic profiles of the four subfamilies (SI Table 1 and SI Fig. 6). Several trends are evident from the phylogenetic tree and profile results. First, no genome had more than one member of any clade. Second, the tree topology within each of the four clades is broadly consistent with accepted species phylogeny (SI Fig. 5). Third, members of each of the three bacterial clades are present in multiple, divergent phyla (SI Table 1). Finally, none of the four clades has a member in every genome of any domain. Taken together, these results suggest that vertical inheritance with periodic and ongoing gene loss appears to be the dominant process in the evolution of the MTTase family, with horizontal gene transfer (aside from the migration of genes of mitochondrial origin to eukaryotic nuclear genomes) and gene duplication (subsequent to the diversification into the four observed subfamilies) playing little or no role.

Sequence Analysis of the MTTase Family. All four MTTase subfamilies exhibit the same tripartite domain structure (26): N-terminal domain UPF0004 (PF00919; Fig. 3, pink box), central radical-SAM domain (PF4055; Fig. 3, cyan box), and C-terminal TRAM domain (PF01938; Fig. 3, brown box). In the model of Nicolet and Drennan (27), the central domain corresponds to the three-quarter TIM barrel core $[(\beta\alpha)_6]$ responsible for radical generation, and the TRAM domain, predicted to be involved with RNA binding (19) and not shared with other radical-SAM

proteins outside the MTTase family, corresponds to the substrate-specific C-terminal variable region.

To identify functionally important residues in these three domains, we aligned a representative sample of 60 protein sequences from across the MTTase family, eight of which are shown in Fig. 3. A total of 31 residues were conserved among at least 54 (90% of the) sequences (Fig. 3 black boxes and SI Table 2). These likely play a role in MTTase protein structure or catalysis, as they are common to all four subfamilies. Six of these conserved residues were cysteines, which in MiaB coordinate the two [4Fe-4S] clusters (17). Six were glycines, four of which appear to delimit secondary structure elements in the radical-SAM domain, a feature not uncommon in TIM barrel proteins (27). The two glycines in the N-terminal domain may perform a similar function, as it cannot be ruled out that this domain forms an additional $(\beta\alpha)_2$ TIM barrel unit. As many as five additional residues appear to be analogous to BioB residues shown to bind the SAM molecule used for radical generation (27).

Using an approach based on mutual information, we also searched for residues that appeared to be characteristic of one or more MTTase subfamilies. Residues of this type, located in several clusters throughout the alignment, may play a role in determining substrate specificity. Such clusters of subfamily-specific residues include 18–19 and 55–56 in the N-terminal domain, near two of the conserved cysteines, and 190–197, 235–239, and 301–303 in the radical-SAM domain. In addition to these smaller clusters, there is a somewhat larger region in the TRAM domain, approximately residues 401–413, where RimO sequences do not align well with the remaining MTTases (Fig. 3). We were particularly interested in cases where three subfamilies shared the same conserved residue at a particular position, whereas the fourth displayed a different conserved residue at the same position, a pattern that may be indicative of substrate-binding elements unique to one subfamily. Fig. 3 and SI Table 3 show the nine positions fitting this pattern. In the majority of these cases (five of nine), it was RimO that displayed the unique residue or related group of residues relative to the other three subfamilies.

Discussion

Prediction of MTTase Subfamily Functions. Five naturally occurring methylthiol modifications have been observed: one on a protein (Fig. 1*A*) (3) and four on tRNA (Fig. 1 *B* and *C*) (22). We can reasonably predict the function of all four MTTase subfamilies based on the phylogenetic distribution of these observed modifications and clade membership (SI Fig. 5). Based on experimental evidence in this work and in studies of MiaB (13, 15), we hypothesize that all members of the RimO clade modify S12 D88 and all members of the MiaB clade modify i^6A to ms^2i^6A [which can be further modified to 2-methylthio-N^6-(*cis*-hydroxyisopentenyl) adenosine (ms^2io^6A) by hydroxylation (28)]. The modifications ms^2i^6A and ms^2io^6A occur in plant chloroplasts but not in plant cytoplasm (23), consistent with our observation that the MiaB clade contains eukaryotic organellar members.

MiaB recognizes the N^6-isopentenyl group and a portion of the tRNA anticodon stem-loop of its substrate (16), and one might intuitively expect other MTTases to recognize analogous elements of their substrates. The N^6 subtituents of 2-methylthio-N^6-threonyl-carbamoyladenosine (ms^2t^6A) and 2-methylthio-N^6-hydroxynorvalyl-carbamoyladenosine (ms^2hn^6A) differ from one another by only a single methyl group but are structurally dissimilar to the isopentenyl moiety of ms^2i^6A. It is therefore likely that the same MTTase recognizes both the N^6-threonyl and N^6-hydroxynorvalyl moieties, or that ms^2hn^6A is formed by methylation of ms^2t^6A. In either case, the same MTTase subfamily, distinct from that of MiaB, is probably responsible for both modifications. Among bacteria, ms^2t^6A has been observed in *B. subtilis* (29) and ms^2hn^6A has been observed in *T. maritima* (30), both of whose genomes encode a YqeV subfamily member. The ms^2t^6A modification is widespread in archaea and eukaryotes, and ms^2hn^6A also occurs in archaea. Archaea and eukaryotes encode, at most, a single nonorganellar MTTase, always from the Mj0867 subfamily. Neither rat nor human ribosomal protein S23 (the homolog of bacterial S12) is modified at the D88 position (7, 8), suggesting that Mj0867 is not functionally analogous to RimO. Therefore, we predict that both the YqeV and Mj0867 subfamilies catalyze formation of ms^2t^6A and ms^2hn^6A, the former in bacteria and the latter in archaea and eukaryotes. The phylogenetic profiles in SI Table 1 should aid in the validation of these predictions.

A published report indicates the possible existence of 2-methylthio-N^6-methyladenosine (ms^2m^6A) in tRNA from *Thermodesulfobacterium commune* (31). Because formation of ms^2m^6A by methylthiolation of N^6-methyladenosine (m^6A) would probably necessitate the existence of another MTTase subfamily in *T. commune*, our results suggest that this is unlikely. However, *T. commune* does contain other unique tRNA modifications (22), and we await the complete genome sequence of this organism.

Functions of MTTase Domains. The structure of *E. coli* RumA, a TRAM-containing enzyme that methylates 23S rRNA to form 5-methyluridine (m^5U) 1939, shows that the TRAM domain forms a five-strand β-barrel characteristic of an oligosaccharide/oligonucleotide binding (OB) fold (32, 33). This domain makes extensive contact with the RNA loop containing U1939, and upon binding to RNA swings by $\approx 10°$ to lock the substrate in a cleft between the OB and catalytic domains of RumA (33). In MTTases, a similar mechanism may trap the substrate against the TIM barrel opening of the radical-SAM domain during radical activation.

The N-terminal domain of RimO, shared with MiaB and other MTTases, contains three conserved cysteine residues that in MiaB coordinate a [4Fe-4S] cluster distinct from that used for radical generation (17) (Fig. 3). The presence of a second FeS cluster, although not common among radical-SAM enzymes in general, is shared among the three radical-SAM proteins shown definitely to perform sulfur insertion reactions: MiaB, biotin synthase (BioB), and lipoyl synthase (LipA). In each of these three cases, the second cluster has been implicated in the thiolation reaction and proposed to be the sulfur donor (17, 34–37). Although this has yet to be shown conclusively, it suggests the MTTase N-terminal domain likely plays a role in the sulfur insertion reaction that precedes methylation.

Methyl Transfer Activity in MTTases. *E. coli* RimO is the second distinct MTTase to be characterized, after MiaB. *In vitro* studies with purified MiaB have shown that it is responsible for two sequential reactions: thiolation of i^6A to a hypothetical intermediate s^2i^6A, then methylation of the sulfur atom to ms^2i^6A (16). Our *in vivo* studies with RimO cannot rule out the possibility of a second protein acting downstream of RimO to perform the methylation reaction in D88 modification. However, the similarity between RimO and MiaB at the sequence and domain structure levels suggests that RimO, like MiaB, encompasses both thiolation and methylation activities, and we have referred to it as a MTTase throughout this work.

The thiolation and methylation activities of MiaB have not been successfully uncoupled via mutagenesis. However, one *in vivo* study using an *E. coli* (*rel met cys*) mutant demonstrated that starvation with methionine but not cysteine resulted in the accumulation of an uncharacterized, cytokinin-active intermediate believed to be s^2i^6A; furthermore, incubation of this intermediate with *E. coli* crude extract and [^{14}C]-SAM resulted in incorporation of the radiolabel (18). Intriguingly, MiaB MTTase activity *in vitro* has been shown to depend on its [4Fe-4S] cluster, which undergoes rapid oxidative degradation to

inactive forms on exposure to air, with nearly complete loss of the active configuration by 30 min (16). The crude extract preparation used by Agris *et al.* (18) to methylate their tRNA intermediate would almost certainly have undergone this degradation. This suggests that the methylation activity of MiaB, and we presume, the other MTTases such as RimO, is independent of the presence of an active [4Fe-4S] cluster but dependent on the prior insertion of sulfur.

The methyl transfer reaction catalyzed by MTTases is therefore likely to proceed via nucleophilic attack by lone-pair electrons of the inserted sulfur atom on the methyl group of SAM. Most MTases contain one or more widely conserved motifs involved with SAM binding, but MiaB and RimO lack any of these motifs. Interestingly, thiopurine MTase (TPMT), a SAM-dependent MTase that methylates purine sulfur substituents, also lacks any conserved motifs despite adopting a classical SAM-dependent MTase fold (38). Using PSI-BLAST (39), we were unable to identify significant similarity between TPMT and either RimO or MiaB (data not shown), so there do not appear to be sequence motifs unique to sulfur-modifying MTases.

In RumA, methyl transfer is carried out by a dedicated SAM-dependent MTase domain, with no participation by the TRAM domain, so there is no reason to suppose that the TRAM domain of MTTases performs this function. However, MTTases do not appear to contain a dedicated MTase domain as does RumA, so the methyl transfer apparatus must be contained within the N-terminal or radical-SAM domains. In the reaction model proposed for MiaB (16), which we extrapolate to RimO and other MTTases, these enzymes use two molecules of SAM: one for radical generation, yielding the products 5′-dA and methionine, and a second for methyl transfer, yielding the product *S*-adenosylhomocysteine (SAH). Residues involved with methyl transfer, including those involved with binding the second SAM molecule, are likely to be conserved throughout the MTTase family.

Of the conserved residues in SI Table 2 with unspecified functions, residue P240 is of particular interest. This proline forms the second residue of an NPPY motif that is strictly conserved in the eukaryotic members of the Mj0867 subfamily, but not found in any other MTTases (Fig. 3). This motif is a hallmark of MTases modifying nitrogen atoms such as DNA N^6-adenine MTases and the protein N^5-glutamine MTase, PrmC (aka HemK). In these enzymes, the NPPY element serves to orient the planar substrate such that its lone pair electrons are directed toward the reactive methyl group of SAM (40, 41). The purine-bound sulfur atom of a thiolated intermediate would also have a lone pair available for a nucleophilic attack. Furthermore, in MTTases as in *N*-MTases, the NPPY motif is located in a loop region immediately following a β-strand, and we suggest this motif may perform a similar function in those MTTases where it is found. Should this prediction be borne out, this would be the first example of such a motif involved with sulfur methylation, and indeed methylation of any atom besides nitrogen. (We note that TPMT proteins share a strictly conserved GPPF motif, not previously noticed to our knowledge, which may perform a similar function in those *S*-MTases.) Interestingly, this motif is not found in other subfamilies of MTTases or in archaeal Mj0867 members. The conservation of P240 within this motif and R234 nearby (SI Table 2) suggests that this region is involved with substrate orientation regardless of whether the NPPY motif is present, so it appears that MTTases may have evolved more than one solution to the problem of substrate binding and orientation.

Five known protein folds are capable of supporting SAM-dependent methyl transfer (42). None of these folds resembles a TIM-barrel, so MTTase proteins typified by RimO and MiaB represent a possible sixth SAM-dependent MTase architecture. Furthermore, it is interesting that the NPPY motif so characteristic of Rossmann-like *N*-MTases appears to have convergently evolved within this novel MTTase context. Whether methyl transfer itself is a convergently evolved function to enzymes that originally functioned solely as thiotransferases is an open question.

Evolution of RimO. It appears from our functional predictions above that RimO is the only MTTase subfamily that modifies protein, or indeed, anything other than tRNA. Perhaps the most surprising aspect of RimO is its extreme resemblance to the other three subfamilies given the difference in substrate type. Examples of protein- and nucleic acid-modifying enzymes resembling one another are known (43), but the similarity is not to this extent. Furthermore, the resemblance is not due to lack of evolutionary time for divergence: based on their phylogenetic footprints, RimO diverged from the other three MTTase subfamilies at least one billion years ago, before the speciation events that gave rise to most modern bacterial phyla.

Given the demonstrated role of the TRAM domain in RumA, it is likely that tRNA stem-loop recognition by MTTases is accomplished by interaction with the TRAM domain. The substrate elements required for recognition by RimO are at present unknown, but as its substrate is a protein rather than tRNA, one might expect the greatest sequence divergence of RimO from the other three subfamilies to have occurred in the TRAM domain. However, aside from the poorly aligning region at RimO residues 401–413 (Fig. 3), this does not appear to be the case, which is astonishing given that OB folds typically exhibit a high degree of sequence plasticity despite structural similarity. At least two hypotheses may explain this phenomenon.

The first is that the RimO TRAM domain binds an RNA stem-loop proximal to or in contact with S12 in the assembled ribosome rather than S12 itself. One possible candidate for such a structure is the 530 stem-loop, one of the most highly conserved regions of the 16S rRNA molecule (44) and one whose secondary structure is influenced by interaction with S12 (45). Cross-linking studies have shown that this loop contacts S12 in *E. coli* (46), where *rimO* is present, yet intriguingly, the equivalent loop in 18S rRNA does not appear to contact S23 in humans (47), where *rimO* is absent. This hypothesis implies that only a stable S12-rRNA complex such as the assembled 30S ribosomal subunit (in which the surface to which S12 is bound is exposed to solution), and not free S12, can serve as an efficient substrate for RimO.

The second hypothesis is that the RimO TRAM domain indeed binds some portion of S12 itself rather than RNA. The extreme similarity between the TRAM domains of RimO and the tRNA-modifying subfamilies suggests that, if this were the case, the recognized portion of S12 may strongly resemble tRNA in some way. This would not be unprecedented, as several instances in which translation factor proteins have been shown or are believed to act as structural mimics of tRNA have come to light in recent years (48, 49). Although the modification sites of both RimO and MiaB sit at the ends of loop structures, a more general resemblance of S12 to tRNA is not immediately apparent from their structures.

This work suggests several avenues of further study. First, identification of the S12 modifying enzyme and the viability of *rimO* mutants should facilitate determination of the function of this unique posttranslational modification. Second, further biochemical characterization of RimO or MiaB should illuminate the mechanisms of the thiolation and methylation reactions, either or both of which may have novel aspects. [While this manuscript was in review, an independent study was published that performed a phylogenetic analysis and domain structure prediction of MiaB-related proteins (50). Although the authors reach essentially identical phylogenetic conclusions to ours, they postulate the possible involvement of cobalamin with the methylation reaction based on the predicted structure of the N-

terminal domain, providing yet another hypothesis to examine.] Finally, the strong resemblance of RimO to MiaB and other tRNA-modifying MTTases suggests that solving and comparing protein structures of these two enzymes should prove an interesting study of the evolution of substrate specificity of modifying enzymes.

Materials and Methods

Bacterial Strains. Strain FB23662, a derivative of *E. coli* K-12 MG1655 bearing a Tn*5* minitransposon insertion at position 376 in the *yliG* ORF, was from the laboratory of Frederick Blattner (51). The minitransposon carries a kanamycin-resistance marker, and the strain also contains plasmid pKD46, which carries an ampicillin-resistance marker (52). FB23662 was cured of the plasmid by overnight growth at 43°C on LB agar with 50 μg/ml kanamycin, creating ampicillin-sensitive strain ER3051.

Plasmid Construction. The gene *yliG* was amplified from *E. coli* NEB 5-alpha (New England Biolabs, Ipswich, MA) by PCR using Phusion DNA polymerase (New England Biolabs). Full details of the construction of plasmids pET28-yliG and pLIT-yliG are provided in *SI Text*. In these plasmids, the *yliG* coding sequence is fused to eight vector-derived codons encoding LEHHHHHH such that the expressed protein contains a C-terminal hexahistidine tag (His-tag).

Plasmid pDB1282, encoding several *E. coli* genes involved in FeS cluster biosynthesis cloned behind an arabinose-inducible promoter, was a kind gift of S. J. Booker (Pennsylvania State University, State College, PA), and was described in ref. 53.

Purification of Ribosomal Proteins. *E. coli* ribosomal proteins were purified by using the method of Hardy *et al.* (54) with minor modifications. Full details are described in *SI Text*.

Mass Spectrometry. Samples of ribosomal proteins purified as described above were analyzed by reverse phase liquid chromatography (LC) and electrospray ionization time-of-flight (ESI-TOF) MS. Full details are described in *SI Text*.

Phylogenetic and Sequence Analysis. Protein sequences were aligned by using MUSCLE (55). Full details are described in *SI Text*.

ACKNOWLEDGMENTS. We thank Mern Sibley for assistance with bacterial strain maintenance and growth, Elmar Weinhold for very helpful discussions, Squire Booker and Dennis Dean (Virginia Polytechnic Institute and State University, Blacksburg, VA) for the kind gift of plasmid pDB1282, and the laboratory of Frederick Blattner for *E. coli* strain FB23662. S.K. was supported in part by National Science Foundation Grant ITR-048715 and National Genome Research Institute Grant R01 HG003367-01A1. We also thank Don Comb and New England Biolabs for support.

1. Gott JM, ed (2007) *Methods In Enzymology: RNA Modification* (Academic, San Diego), Vol 425.
2. Polevoda B, Sherman F (2007) *Mol Microbiol* 65:590–606.
3. Kowalak JA, Walsh KA (1996) *Protein Sci* 5:1625–1632.
4. Carr JF, Hamburg DM, Gregory ST, Limbach PA, Dahlberg AE (2006) *J Bacteriol* 188:2020–2023.
5. Strader MB, Verberkmoes NC, Tabb DL, Connelly HM, Barton JW, Bruce BD, Pelletier DA, Davison BH, Hettich RL, Larimer FW, Hurst GB (2004) *J Proteome Res* 3:965–978.
6. Suh MJ, Hamburg DM, Gregory ST, Dahlberg AE, Limbach PA (2005) *Proteomics* 5:4818–4831.
7. Louie DF, Resing KA, Lewis TS, Ahn NG (1996) *J Biol Chem* 271:28189–28198.
8. Vladimirov SN, Ivanov AV, Karpova GG, Musolyamov AK, Egorov TA, Thiede B, Wittmann-Liebold B, Otto A (1996) *Eur J Biochem* 239:144–149.
9. Brodersen DE, Clemons WM, Jr, Carter AP, Wimberly BT, Ramakrishnan V (2002) *J Mol Biol* 316:725–768.
10. Kurland CG, Hughes D, Ehrenberg M (1996) in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ed Neidhardt FC (ASM Press, Washington, DC), 2nd Ed, pp 979–1004.
11. Gavrilova LP, Kotelianskii VE, Spirin AS (1975) *Mol Biol (Moscow)* 9:609–621.
12. Asatryan LS, Spirin AS (1975) *Mol Gen Genet* 138:315–321.
13. Pierrel F, Hernandez HL, Johnson MK, Fontecave M, Atta M (2003) *J Biol Chem* 278:29515–29524.
14. Pierrel F, Bjork GR, Fontecave M, Atta M (2002) *J Biol Chem* 277:13367–13370.
15. Esberg B, Leung HC, Tsui HC, Bjork GR, Winkler ME (1999) *J Bacteriol* 181:7256–7265.
16. Pierrel F, Douki T, Fontecave M, Atta M (2004) *J Biol Chem* 279:47555–47563.
17. Hernandez HL, Pierrel F, Elleingand E, Garcia-Serres R, Huynh BH, Johnson MK, Fontecave M, Atta M (2007) *Biochemistry* 46:5140–5147.
18. Agris PF, Armstrong DJ, Schafer KP, Soll D (1975) *Nucleic Acids Res* 2:691–698.
19. Anantharaman V, Koonin EV, Aravind L (2001) *FEMS Microbiol Lett* 197:215–221.
20. Sofia HJ, Chen G, Hetzler BG, Reyes-Spindola JF, Miller NE (2001) *Nucleic Acids Res* 29:1097–1106.
21. Anantharaman V, Koonin EV, Aravind L (2002) *Nucleic Acids Res* 30:1427–1464.
22. Limbach PA, Crain PF, McCloskey JA (1994) *Nucleic Acids Res* 22:2183–2196.
23. Ajitkumar P, Cherayil JD (1988) *Microbiol Rev* 52:103–113.
24. Arnold RJ, Reilly JP (1999) *Anal Biochem* 269:105–112.
25. Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) *J Mol Biol* 300:1005–1016.
26. Eddy SR (1998) *Bioinformatics* 14:755–763.
27. Nicolet Y, Drennan CL (2004) *Nucleic Acids Res* 32:4015–4025.
28. Persson BC, Bjork GR (1993) *J Bacteriol* 175:7776–7785.
29. Vold BS, Keith DE, Jr, Buck M, McCloskey JA, Pang H (1982) *Nucleic Acids Res* 10:3125–3132.
30. Reddy DM, Crain PF, Edmonds CG, Gupta R, Hashizume T, Stetter KO, Widdel F, McCloskey JA (1992) *Nucleic Acids Res* 20:5607–5615.
31. Crain PF, Hashizume T, Reddy DM, McCloskey JA (1991) *Nippon Iyo Masu Supekutoru Gakkai Koenshu* 16:15–26.
32. Lee TT, Agarwalla S, Stroud RM (2004) *Structure (London)* 12:397–407.
33. Lee TT, Agarwalla S, Stroud RM (2005) *Cell* 120:599–611.
34. Ugulava NB, Gibney BR, Jarrett JT (2001) *Biochemistry* 40:8343–8351.
35. Cosper MM, Jameson GN, Hernandez HL, Krebs C, Huynh BH, Johnson MK (2004) *Biochemistry* 43:2007–2021.
36. Jameson GN, Cosper MM, Hernandez HL, Johnson MK, Huynh BH (2004) *Biochemistry* 43:2022–2031.
37. Cicchillo RM, Lee KH, Baleanu-Gogonea C, Nesbitt NM, Krebs C, Booker SJ (2004) *Biochemistry* 43:11770–11781.
38. Scheuermann TH, Lolis E, Hodsdon ME (2003) *J Mol Biol* 333:573–585.
39. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) *Nucleic Acids Res* 25:3389–3402.
40. Goedecke K, Pignot M, Goody RS, Scheidig AJ, Weinhold E (2001) *Nat Struct Biol* 8:121–125.
41. Schubert HL, Phillips JD, Hill CP (2003) *Biochemistry* 42:5592–5599.
42. Schubert HL, Blumenthal RM, Cheng X (2003) *Trends Biochem Sci* 28:329–335.
43. Bujnicki JM, Radlinska M (1999) *IUBMB Life* 48:247–249.
44. Woese CR, Fox GE, Zablen L, Uchida T, Bonen L, Pechman K, Lewis BJ, Stahl D (1975) *Nature* 254:83–86.
45. Stern S, Powers T, Changchien LM, Noller HF (1988) *J Mol Biol* 201:683–695.
46. Alexander RW, Muralikrishna P, Cooperman BS (1994) *Biochemistry* 33:12109–12118.
47. Malygin AA, Dobrikov MI, Repkova MN, Shishkin GV, Ven'yaminova AG, Karpova GG (1999) *RNA* 5:1656–1664.
48. Nissen P, Kjeldgaard M, Thirup S, Polekhina G, Reshetnikova L, Clark BF, Nyborg J (1995) *Science* 270:1464–1472.
49. Selmer M, Al-Karadaghi S, Hirokawa G, Kaji A, Liljas A (1999) *Science* 286:2349–2352.
50. Kaminska KH, Baraniak U, Boniecki M, Nowaczyk K, Czerwoniec A, Bujnicki JM (2007) *Proteins* 70:1–18.
51. Kang Y, Durfee T, Glasner JD, Qiu Y, Frisch D, Winterberg KM, Blattner FR (2004) *J Bacteriol* 186:4921–4930.
52. Datsenko KA, Wanner BL (2000) *Proc Natl Acad Sci USA* 97:6640–6645.
53. Cicchillo RM, Iwig DF, Jones AD, Nesbitt NM, Baleanu-Gogonea C, Souder MG, Tu L, Booker SJ (2004) *Biochemistry* 43:6378–6386.
54. Hardy SJ, Kurland CG, Voynow P, Mora G (1969) *Biochemistry* 8:2897–2905.
55. Edgar RC (2004) *Nucleic Acids Res* 32:1792–1797.