THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 35, Issue of August 27, pp. 37142–37152, 2004
Printed in U.S.A.

N^2-Methylation of Guanosine at Position 10 in tRNA Is Catalyzed by a THUMP Domain-containing, *S*-Adenosylmethionine-dependent Methyltransferase, Conserved in Archaea and Eukaryota*[S]

Received for publication, April 7, 2004, and in revised form, May 28, 2004
Published, JBC Papers in Press, June 20, 2004, DOI 10.1074/jbc.M403845200

Jean Armengaud‡§, Jaunius Urbonavičius¶‖, Bernard Fernandez‡, Guylaine Chaussinand‡, Janusz M. Bujnicki‡‡, and Henri Grosjean¶**

*From the ‡Commissariat à l'Energie Atomique VALRHO, DSV-DIEP-SBTN, Service de Biochimie Post-génomique & Toxicologie Nucléaire, F-30207 Bagnols-sur-Cèze, France, ¶Laboratoire d'Enzymologie et Biochimie Structurales, CNRS, Bld. 34, 1 Avenue de la Terrasse, F-91198 Gif-sur-Yvette, France, and **International Institute of Molecular and Cell Biology, Trojdena 4, Warsaw 02-109, Poland*

In sequenced genomes, genes belonging to the cluster of orthologous group COG1041 are exclusively, and almost ubiquitously, found in Eukaryota and Archaea but never in Bacteria. The corresponding gene products exhibit a characteristic Rossmann fold, *S*-adenosylmethionine-dependent methyltransferase domain in the C terminus and a predicted RNA-binding THUMP (thiouridine synthases, RNA methyltransferases, and pseudouridine synthases) domain in the N terminus. Recombinant PAB1283 protein from the archaeon *Pyrococcus abyssi* GE5, a member of COG1041, was purified and shown to behave as a monomeric 39-kDa entity. This protein (EC 2.1.1.32), now renamed $_{Pab}$Trm-G10, which is extremely thermostable, forms a 1:1 complex with tRNA and catalyzes the adenosylmethionine-dependent methylation of the exocyclic amino group (N^2) of guanosine located at position 10. Depending on the experimental conditions used, as well as the tRNA substrate tested, the enzymatic reaction leads to the formation of either N^2-monomethyl (m^2G) or N^2-dimethylguanosine (m^2_2G). Interestingly, $_{Pab}$Trm-G10 exhibits different domain organization and different catalytic site architecture from another, earlier characterized, tRNA-dimethyltransferase from *Pyrococcus furiosus* ($_{Pfu}$Trm-G26, also known as $_{Pfu}$Trm1, a member of COG1867) that catalyzes an identical two-step dimethylation of guanosine but at position 26 in tRNAs and is also conserved among all sequenced Eukaryota and Archaea. The co-occurrence of these two guanosine dimethyltransferases in both Archaea and Eukaryota but not in Bacteria is a hallmark of distinct tRNAs maturation strategies between these domains of life.

The considerable effort of the last 5 years in sequencing entire genomes from the three domains of life (Bacteria, Archaea, and Eukaryota (1)) has made possible the establishment of a list of clusters of orthologous groups (COGs)[1] of unknown function (2–4) (available on the World Wide Web at www.ncbi.nlm.nih.gov/COG/). This continuously updated list allows functional annotation of newly sequenced genomes and genome-wide evolutionary analyses. Among these COGs, some appear to have great importance, since they are present in most, if not all, members of one or several domains of life. Distinction can be made between those conserved in all three domains (thus defining a universal subset of proteins probably present in LUCA, the last universal common ancestor of the extant life forms), those specific to a single domain, and those shared by two of the three domains of life. Several comparative studies on characterized proteins, aimed at distinguishing organisms of the three biological domains, have indicated that many, although not all, proteins involved in metabolic pathways of Archaea resemble more closely their bacterial than eukaryotic counterparts, whereas proteins involved in the organization or processing of genetic information (structures of ribosome and chromatin, translation, transcription, replication, and DNA repair) suggest a closer relationship between Archaea and Eukaryota (5–8). A series of 32 COGs of unknown function have been identified that are common to Eukaryota and Archaea but not found in Bacteria (9). These groups were named PACEs (for proteins from Archaea conserved in Eukaryota (see, on the World Wide Web, www-archbac.u-psud.fr/projects/pace/paceproteins.html). Interestingly, for some of these proteins, roles in transcription, translation, or replication processes have been postulated (9). For example, polypeptides from the PACE08 group (COG1369) have been identified as one of the subunits of ribonuclease P (RNase P), a ribonucleoprotein complex involved in 5′-end tRNA maturation (10). The SSO0175 and MJ0051 archaeal proteins, as well as the Ybl057c protein from *Saccharomyces cerevisiae* (all from the PACE07 group), have been shown to have peptidyl-tRNA hydrolase activity and are thus involved in recycling peptidyl-tRNA molecules prematurely dissociated from the mRNA template (11, 12).

Among proteins of the PACE groups characterized so far, only those belonging to PACE 11 have been shown not to be related to organization or processing of cellular information. Two of us recently reported that the product of the *Pyrococcus*

* This work was supported in part by a grant from the Centre National de la Recherche Scientifique (Geomex program, 2002–2004) (to H. G.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains two additional figures.

§ To whom correspondence should be addressed. Tel.: 33-4-66-79-68-02; Fax: 33-4-66-79-19-05; E-mail: armengaud@cea.fr.

‖ Supported by a postdoctoral FEBS fellowship.
‡‡ Supported by the EMBO/HHMI Young Investigator Program and by a fellowship from the Foundation for Polish Science.

[1] The abbreviations used are: COG, cluster of orthologous group; AdoMet, *S*-adenosylmethionine; DTT, dithiothreitol; HEPPS, *N*-(2-hydroxyethyl)piperazine-*N*′-(3-propanesulfonic acid); MTase, methyltransferase; PSI-BLAST, position-specific iterated BLAST; TLC, two-dimensional thin layer cellulose; MOPS, 4-morpholinepropanesulfonic acid; bis-tris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol.

abyssi GE5 *PAB0944* gene (COG1019) corresponds to a single domain polypeptide involved in the fourth step of coenzyme A biosynthesis (phosphopantetheine adenylyltransferase), whereas its human ortholog is fused to another domain, dephosphocoenzyme A kinase, which is responsible for catalyzing the last step of the coenzyme A biosynthetic pathway (13). All of these examples clearly indicate that PACE proteins have fundamental cellular functions and that several of them are obviously related to RNA metabolism.

In the present study, we focus our attention on PAB1283 from the hyperthermophilic archaeon *P. abyssi* GE5. This protein belongs to the PACE18 group (9) and is classified as a member of the COG1041 group (2). A search of the NCBI conserved domain data base (available on the World Wide Web at www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) (14) indicates that all sequences from COG1041 encompass two domains: a putative AdoMet-dependent methyltransferase (MTase) domain (pfam01170) at the C terminus and a so-called THUMP (thiouridine synthases, RNA methyltransferases, and pseudouridine synthases) domain (pfam02926 (15)) of about 110 amino acid residues at the N terminus.

The reactions catalyzed by MTases are very diverse and can transfer the AdoMet methyl group onto either nitrogen, oxygen, or carbon atoms in DNA, RNA, proteins, lipids, or small molecules such as catechol (reviewed in Ref. 16). The THUMP domain has been identified in the ThiI enzyme catalyzing the formation of 4-thiouridine in bacterial tRNAs as well as in several other putative RNA modification enzymes, such as MTases and pseudouridine synthases (see Ref. 15 and references therein). For this reason, it is expected to be an ancient RNA-binding domain. Therefore, we hypothesized that COG1041 proteins are as yet uncharacterized RNA MTases for which the target nucleoside within RNA molecules has still to be discovered.

Most stable cellular RNAs (tRNAs, rRNAs, and small nuclear (small nucleolar)RNAs) contain many post-transcriptionally modified nucleosides. To date, at least 96 different chemically distinct modified nucleosides have been identified in the many naturally occurring RNA analyzed (see, on the World Wide Web, medstat.med.utah.edu/RNAmods/). Among them, ribose and/or base methylations are among the most frequently encountered (17) (for a review, see Ref. 18). They have been shown to play a role in stabilization, structural folding, protection from endonucleases, and molecular recognition of the functional RNAs by numerous proteins as well as in RNA-RNA interactions (for a review, see many chapters in Ref. 19).

To verify that proteins belonging to COG1041 are AdoMet-dependent RNA MTases, we purified to homogeneity and characterized a recombinant form of the *PAB1283* gene product from the hyperthermophilic archaeon *P. abyssi*. We identified *in vitro* both the reaction catalyzed by the purified enzyme and the target nucleotide within its RNA substrates. Our results demonstrate for the first time that a THUMP-containing MTase is associated with a tRNA modification function and that the *PAB1283* gene product corresponds to a site-specific tRNA: $m^2{}_2G$ dimethyltransferase acting on the N^2-exocyclic amine group of guanosine at position 10 in tRNA.

MATERIALS AND METHODS

Construction of an N-terminal His_6-tagged PAB1283-overexpressing Plasmid—A construct was designed in order to get overexpression of a slightly modified version of *P. abyssi* PAB1283 open reading frame. The Met-ATG located at 1592098 (numbering refers to whole *P. abyssi* genomic sequence), and not the upstream Leu-TTG codon that genome annotators mentioned in the data base, was taken as the most probable initiator codon. This was the choice because TTG codons are rarely used as initiators in Archaea and also because the resulting N-terminal PAB1283 sequence aligns better with those of the other archaeal homologs. Thirteen amino acid residues (MRGSHHHHHHGMAS) were introduced with the N-terminal His_6 tag. Two synthetic oligonucleotide primers were designed in order to amplify the *PAB1283* gene from *P. abyssi* total DNA: oAD34 (5′-catatggctagcATGTTCTACGTTGAAATC-CTAGGTTTG-3′) that contains two engineered restriction sites (NdeI and NheI) and oAD35 (5′-ggatcctcaTCACCTCGCC TCCATTATGTAG-AAG-3′) that contains an engineered BamHI site just after the stop codon. Nucleotides in lowercase type were not present in the original coding sequence. PCR performed with *Pwo* polymerase (Roche Applied Science) gave a 1.0-kb homogeneous product that was cloned into pCRScript-cam (Stratagene), resulting in plasmid pSBTN-AB89. The 994-bp insert from pSBTN-AB89 was removed by digestion with NheI and BamHI and ligated with T4-DNA ligase with plasmid pSBTN-AB23,[2] a derivative of pCR T7/NT-topo (Invitrogen) containing a T7 promoter and His_6 tag, previously digested with compatible endonucleases (NheI and BglII). The resulting plasmid was sequenced in order to ascertain the integrity of the nucleotide sequence and was named pSBTN-AC18.

Purification of Recombinant PAB1283 Protein—Overexpression of the *PAB1283* construct was achieved with the *Escherichia coli* Rosetta(DE3)pLysS strain (Novagen) transformed with pSBTN-AC18. Large scale liquid cultures were set up at 30 °C and induced with 1 mM isopropyl-γ-D-thiogalactopyranoside as described earlier (13). Cells (63 g of wet material) were resuspended in 300 ml of cold 50 mM Tris/HCl buffer (pH 8.3 at 20 °C) and disrupted by means of a BasicZ cell disrupter (Constant Systems Ltd.) operated at 900 bars. The cold cell extract was then centrifuged at 20,000 × *g* for 20 min at 10 °C to remove cellular debris and aggregated proteins. PAB1283 was purified from 45 ml of this cell extract (corresponding to 7 g of wet cells). The sample was subjected to a 20-min heat treatment at 65 °C and centrifuged at 20,000 × *g* for 15 min at 10 °C. The resulting 40-ml clear supernatant was diluted with an equal volume of 50 mM Tris/HCl buffer (pH 8.3) containing 100 mM KCl and 50 mM imidazole (buffer A). The sample was applied at room temperature onto a 5-ml HiTrap chelating HP column (Amersham Biosciences) at a flow rate of 0.6 ml/min. After a 5-column volume wash with buffer A at a flow rate of 1 ml/min, the N-terminal His_6-tagged PAB1283 protein was eluted over a 15-ml linear gradient comprising 50–300 mM imidazole. The major peak, which eluted from the column at about 150 mM imidazole, was desalted by gel filtration at a flow rate of 1.5 ml/min on an Amersham Biosciences XK26/40 column containing 175 ml of packed G25SF gel (Amersham Biosciences) previously equilibrated with 50 mM Tris/HCl buffer (pH 8.3) containing 50 mM NaCl. The protein was further purified by chromatography either on MonoQ or hydroxyapatite columns. In the first case, samples were injected at a 0.5 ml/min flow rate onto a 1-ml MonoQ HR5/5 column (Amersham Biosciences) previously equilibrated with buffer A, and the bound proteins were eluted over a 40-ml linear gradient comprising 50–550 mM NaCl. The proteins, which eluted between 100 and 150 mM NaCl, were then dialyzed against 50 mM Tris/HCl buffer (pH 8.3) containing 20 mM NaCl and stored at −80 °C. Alternatively, samples were first dialyzed against 100 mM NaH_2PO_4/Na_2HPO_4 buffer at pH 7.2 containing 300 mM NaCl and 2 mM DTT (buffer H1) and then injected at a 0.5 ml/min flow rate onto an HR10/10 column (Amersham Biosciences) packed with 7 ml of hydroxyapatite Bio-gel HT (Bio-Rad) and previously equilibrated with buffer H1. Proteins were resolved with a 3-column volume linear gradient from 100 to 500 mM NaH_2PO_4/Na_2HPO_4 and then dialyzed against 50 mM Tris/HCl buffer (pH 8.3) containing 20 mM NaCl. Protein concentrations were determined by UV spectrophotometry at 280 nm using the molar absorption coefficient of 29,300 M^{-1} cm^{-1} obtained from calculation of the amino acid composition of the recombinant protein (20) (available on the World Wide Web at www.expasy.org/tools/protparam.html).

Estimation of Quaternary Structure of Purified Recombinant Enzyme by Gel Filtration—The native molecular mass of PAB1283 protein was first estimated by gel filtration chromatography on a Superdex 200 gel packed into a HR10/30 column (Amersham Biosciences) with a final bed volume of 24 ml. The column was equilibrated at room temperature at a flow rate of 0.5 ml/min with 50 mM Tris/HCl buffer, pH 8.3, containing 50 mM NaCl and eluted with the same buffer. A second, more accurate estimation was made on a Superdex75 gel packed into a HR10/30 column (Amersham Biosciences) with a final bed volume of 24 ml. This column was equilibrated at room temperature at a flow rate of 0.75 ml/min with 25 mM sodium phosphate buffer, pH 7.2, containing 5 mM $MgCl_2$ and 2 mM DTT. Samples containing the purified recombinant PAB1283 protein (2.3 and 3.7 nmol, respectively) were chromato-

[2] J. Armengaud and V. Chaumont, unpublished data.

graphed on Superdex200 and Superdex75 columns. Complex formation between PAB1283 protein and various samples of tRNAs (bulk yeast tRNAs, bulk *E. coli* tRNAs, or purified yeast $tRNA^{Asp}$) were evaluated by comparing the elution profiles of tRNAs and protein alone with those of tRNA-protein complexes on a Superdex75 gel filtration column. Transfer RNAs from both yeast and *E. coli* were obtained from Sigma, whereas purified yeast $tRNA^{Asp}$ and purified 74-nucleotide RNA that includes part of helix 22 and helices 34 and 35 from *E. coli* 23 S rRNA (described in Ref. 21) were generous gifts of Dr. Anne Theobald-Ditriech (CNRS-IBMC, Strasbourg, France) and Dr. Dominique Fourmy (CNRS-ICSN, Gif-sur-Yvette, France), respectively. Before use, the RNAs were precipitated with ethanol and resuspended in 10 mM Tris/HCl buffer, pH 7.4, containing 10 mM $MgCl_2$ and 150 mM NaCl.

Differential Scanning Calorimetry—The transition temperature, T_m, of purified PAB1283 was determined with a high sensitivity differential scanning microcalorimeter, VP-DSC (MicroCal). The reference buffer was 30 mM HEPPS/NaOH, pH 8.0, containing 50 mM NaCl and 5 mM $MgCl_2$. Prior to the calorimetric analysis, the sample was dialyzed at room temperature against this reference buffer. Recombinant His_6-tagged PAB1283 protein was analyzed at a protein concentration of 5.3 μM. Calorimetric scans against the reference buffer were carried out in 0.5 ml of fixed-in-place cells. A self-contained pressurizing system (0–45 p.s.i.) allowed the scanning of solutions above their boiling point. Scans were recorded between 30 and 110 °C with a heating scan rate of 1.0 °C/min.

tRNA Methylation Assays—To determine the tRNA (G10) MTase activity of PAB1283, various ^{32}P-radiolabeled tRNA transcripts were used as substrates. They were obtained by *in vitro* transcription with T7 polymerase (Promega) of linearized plasmids harboring appropriate synthetic tRNA genes using α-^{32}P-labeled GTP, UTP, or CTP (Amersham Biosciences). Both preparation and purification of resulting transcripts on urea gels have been described elsewhere (22). For testing enzyme activity, an aliquot (a few nmol) of purified recombinant PAB1283 protein in 2.5 μl of 25 mM sodium phosphate buffer (pH 7.2) containing 1 mM $MgCl_2$, 10% glycerol, and 1 mg/ml serum albumin was added to 22.5 μl of the standard reaction mixture consisting in 25 mM sodium phosphate buffer (pH 7.2), 5 mM $MgCl_2$, 2 mM DTT, 40 μg/ml polyuridylic acid (Sigma), 80 μM AdoMet (Sigma), with enough ^{32}P-radiolabeled tRNA to obtain maximum incorporation of the methyl group from AdoMet to the tRNA substrate after 1 h of incubation at 50 °C. AdoMet stock solutions were made at 20 mM in hydrochloric acid, pH 2.0, and stored at −70 °C. Prior to the addition to the reaction mixture, the solutions of tRNA substrates contained in siliconized Eppendorf tubes were renatured by a 3-min incubation in a water bath at 75 °C followed by gradual cooling to room temperature. After incubation, the reaction was stopped by adding 200 μl of cold 0.3 M sodium acetate (pH 5.3) immediately followed by the addition of an equal volume of phenol/chloroform (24:1). Denatured proteins were then removed by centrifugation at 13,000 × *g* for 3 min at room temperature, and nucleic acids present in the upper phase were precipitated with ethanol, collected by centrifugation, washed once with 70% ethanol, dried, and finally completely digested into 5′-monophosphate nucleosides by overnight incubation at 37 °C with an excess (0.4 μg) of nuclease P1 (Roche Applied Science) in 10 μl of 50 mM ammonium acetate/acetic acid buffer at pH 5.3. Alternatively, complete digestion of dried RNA samples into 3′-phosphate nucleosides was performed by overnight incubation at 37 °C with 10 μl of 50 mM ammonium acetate/acetic acid buffer at pH 4.6 containing a home-made mixture of RNase T1 and RNase T2 (23). The resulting hydrolysates were then analyzed by two-dimensional thin layer chromatography on 10 × 10-cm cellulose plates as described elsewhere together with the necessary reference maps (23). Localization of radioactive spots on the thin layer plates and evaluation of their relative radioactivity were performed after exposure of the plates to a PhosphorImager screen followed by scanning with a STORM instrument (Amersham Biosciences). Using this approach, the optimal experimental conditions described above for testing tRNA methylation were obtained by varying the nature of the buffering system, the nature and concentration of monovalent salt and Mg^{2+} ions, and the protein and tRNA stabilizers such as bovine serum albumin and polyuridilic acid. Maximum activity of PAB1283 protein was obtained in 25 mM sodium phosphate buffer (pH 7.2) containing 5 mM $MgCl_2$, 40 μg/ml polyuridylic acid, and 100 μg/ml of bovine serum albumin. For testing the incorporation of the tritiated methyl group into tRNA, the same reaction mixture as described above was used, except that unlabeled bulk *E. coli* tRNA (Sigma) and *methyl*-^{3}H-labeled AdoMet (15 Ci/mmol; Amersham Biosciences) were used as substrate and methyl donor, respectively. Evaluation of tritium incorporation into tRNA was measured after trichloroacetic acid precipitation of the nucleic acid and filtration on a Millipore membrane.

RESULTS

COG1041 Proteins Are Almost Ubiquitous in Archaea and Eukaryota—A search of the nonredundant protein data base from NCBI using the PSI-BLAST tool (24) and PAB1283 sequence as query revealed a set of 20 full-length closely related proteins (hits reported in the first iteration with *E* values below 10^{-14} and sequence identity above 30%), all from Archaea. Among more distantly related proteins, eukaryotic sequences were detected in the second iteration, such as AAQ10284 from *Homo sapiens* (gi33329797) and Yol124c from *S. cerevisiae* (gi6324448). Reciprocal data base searches yielded 41 presumed orthologs from Eukaryota and Archaea grouped in COG1041. No close homologs (*E* value below 10^{-12} in the second iteration) with full-length sequence matching to PAB1283 were found among Bacteria, except closely related proteins from *Bacillus anthracis* A2012 and *Bacillus cereus* ATCC14579. However, reciprocal BLAST searches revealed no further close bacterial homologs of these proteins, suggesting that COG1041 members are not generic to Bacteria and might have been introduced to this domain by horizontal gene transfer.

Remarkably, another related sequence, Vng2242c, was detected in the archaeon *Halobacterium* sp. NRC-1 but matched only the C-terminal part of all of the other COG1041 sequences. After analysis of the corresponding nucleic acid sequence (NC_002607), we found that a 322-amino acid polypeptide exhibiting significant similarity to all COG1041 proteins along the whole sequence would be encoded by the reverse strand from a GTG codon at 1,669,809 (and not from the ATG reported by genome annotators at 1,669,326). Except for the early divergent eukaryon *Encephalitozoon cuniculi* that has an exceptionally small genome, all of the sequenced Archaea and Eukaryota were found to have at least one PAB1283 homolog.

Fig. 1 shows the results of an NCBI conserved domain search with PAB1283 as the query (14). All of the sequences from COG1041 encompass two conserved domains: a THUMP domain (15) in the N terminus (pfam 02926) and an MTase domain in the C terminus (pfam 01170). Fig. 1 also presents a sequence alignment of the COG1041 C-terminal regions from a selection of 10 representative organisms, revealing a pattern of conserved motifs characteristic of an AdoMet-dependent MTase. Both the AdoMet-binding and the catalytic residues are located within a single domain with an α/β Rossmann-like fold. Assuming that the THUMP domain is primarily involved in substrate binding (15), the linear arrangement of structural elements seen in the COG1041 proteins makes them members of the ζ class of Rossmann fold MTases (16, 25).

On the basis of comparisons with other MTases (26, 27), three Asp residues in PAB1283 are probably crucial to cofactor binding (Fig. 1). Asp^{185}, Asp^{208}, and Asp^{235} are predicted to coordinate the methionine moiety via a water molecule (28), the 2′- and 3′-hydroxyl groups of the ribose moiety, and the N^6 and/or N^3 atoms of the adenine moiety, respectively. The predicted substrate-binding pocket comprises motifs IV, VI, VIII, and X. The highly conserved consensus pentapeptide sequence DPPYG (including Asp^{254}) from the β4 region (highlighted in *red* in Fig. 1) comprises a typical motif IV signature, when compared with the consensus (D/N/S)PP(F/Y/W/H) pattern identified in a subset of MTases that methylate exocyclic amines of adenosine, cytidine, or guanosine (29). Asp^{254} is expected to coordinate the methylatable exocyclic amino group, whereas Tyr^{257} stabilizes the target base by stacking interactions (30, 31). In motif VIII (not shown in Fig. 1), an invariant Arg (Arg^{321} in PAB1283) is predicted to participate in the

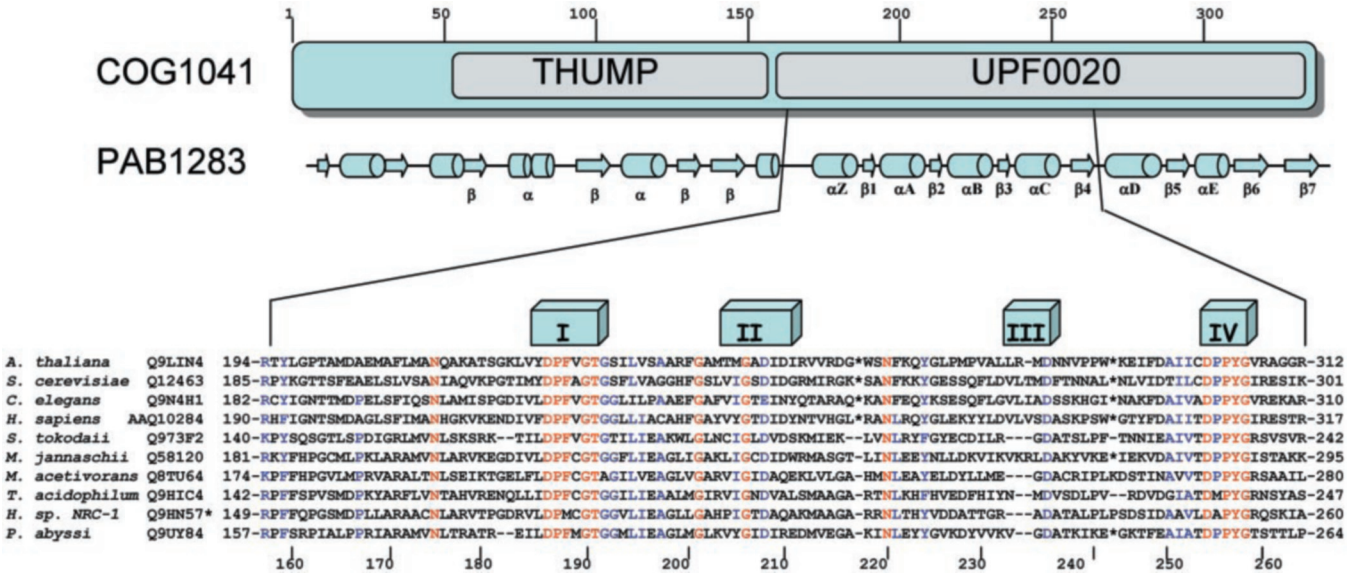


FIG. 1. **General domain organization of COG1041 sequences.** Archaeal and eukaryal homologs were obtained from public data bases (available on the World Wide Web at www.ncbi.nlm.nih.gov/) by PSI-BLAST searches (24). Domains were identified by the NCBI conserved domain search tool (available on the World Wide Web at www.ncbi.nlm.nih.gov). Multiple alignments were performed by ClustalW and manually refined. Predictions of secondary structure elements of the PAB1283 sequence were made through the Psipred Web-interfaced software (available on the World Wide Web at bioinf.cs.ucl.ac.uk/psipred/). Predicted β-sheets and α-helices are represented by *arrows* and *pipes*, respectively. Amino acid numbering refers to PAB1283 sequence. Motifs I, II, IV, and VI, corresponding to AdoMet binding regions, are marked *above* the residues in which they appear. The *asterisks* mark sets of residues that are not shown for clarity. Strictly conserved residues are in *red*, and residues with greater than 70% identity or strictly of the same type are in *blue*. Accession numbers are indicated *beside* each organism name. Based on phylogenetic analysis (data not shown), 10 representative sequences were selected from the 42 COG1041 members. The first four sequences are from Eukaryota, whereas the last six sequences are from Archaea. The Vng2242c sequence from *Halobacterium* sp. NRC-1, labeled *Q9HN57**, has been numbered according to the new annotation proposed under "Results."

binding of the substrate, possibly via direct contacts with the methylated base.

The phylogenetic tree generated from the aligned sequences (data not shown) allows sorting of the MTases belonging to COG1041 into two main groups, depending on their archaeal or eukaryal origin. It is noteworthy that, unlike other Archaea, the pyrococci (*abyssi*, *furiosus*, and *horikoshi*) encode two paralogous subfamilies (32): PAB1283 (this work), PF0400, and PH0338 on the one hand and PAB1719, PF1027, and PH0997 on the other. Proteins from the second group are unique members in that they exhibit a degenerated motif IV signature, with the "classical" DPPY tetrapeptide substituted by an unusual sequence EPYM. A similar replacement (EPPL or GCLT instead of the DPPY motif IV) has already been reported in BAB13453 and Kar4 lineages of proteins related to the MTA70 subunit of human mRNA:m^{6}A MTase (33). Moreover, PAB1719-related proteins contain a unique insertion of about 10 amino acids in motif VI. These deviations from the conserved pattern suggest that they may have evolved some new function not present in the "orthodox" members of the COG1041 family.

The Purified Recombinant PAB1283 Is Hyperthermostable—To characterize the putative MTase activity of COG1041 members, PAB1283 from *P. abyssi* was produced as an N-terminal His$_6$-tagged form in *E. coli*. This polypeptide has an apparent molecular mass of ~38–39 kDa (Fig. 2, *lane 2*). However, only a fraction of PAB1283 was soluble (Fig. 2, compare *lanes 1* and *3*) but remained in this state even when the cell extract was heated to 85 °C. A three-step purification protocol was developed to purify the soluble form. It includes a 20-min heat treatment at 65 °C that eliminated a large proportion of *E. coli* proteins (Fig. 2, *lane 4*), followed by a metal ion affinity chromatography (Fig. 2, *lanes 5* and *6*) and an ion exchange chromatography on either a MonoQ or a hydroxyapatite column (Fig. 2, *lanes 7–9*). This last step was required to remove residual protein contaminants and most of nucleic acids that bound strongly to the recombinant protein. Fig. 2 (*lanes 8* and *9*) shows the high degree of purity obtained (>98%). Yields of ~0.35 and 0.45 mg of pure protein per liter of culture were obtained by the MonoQ and the hydroxyapatite procedures, respectively.

The transition temperature of His-tagged PAB1283 was investigated by differential scanning microcalorimetry. Thermal unfolding of PAB1283 protein was found to be an irreversible process. The differential calorimetric melting curve (Fig. 3) indicates a maximum centered at 102 °C, corresponding to the apparent midpoint melting temperature (T_m). Noticeable unfolding of the protein (>5%) was detected above 86 °C. Although slightly modified by a His$_6$ tag in the N terminus and in an environment quite different from the normal cellular conditions (34), this protein originating from a hyperthermophilic microorganism is therefore highly thermostable.

The Monomeric PAB1283 Protein Interacts with RNAs—The native molecular mass of PAB1283 was determined by size exclusion chromatography. Fig. 4*A* shows the elution profile of the purified enzyme obtained with a Superdex75 column. Results indicate that the enzyme behaves as a monomeric protein with an apparent molecular mass of 40 kDa. This result is supported by the PAB1283 melting profile (Fig. 3), in that most multimeric proteins would be expected to show oligomer dissociation prior to subunit unfolding (35), whereas here a single well defined melting peak was recorded. When chromatography was performed on a Superdex200 and with a less purified enzyme pool obtained after the immobilized metal ion adsorption chromatography and before the hydroxyapatite purification step, an additional minor peak was visible at a position of the elution profile that corresponds to a mean molecular mass of about 70 kDa. Since the ratio A_{260}/A_{280} of this minor peak was 1.60 instead of 0.57 for the main protein pool, the presence of small nucleic acids complexed to the enzyme was suspected.

Since the THUMP domain has been predicted to be a RNA-binding domain (15), we checked whether the purified

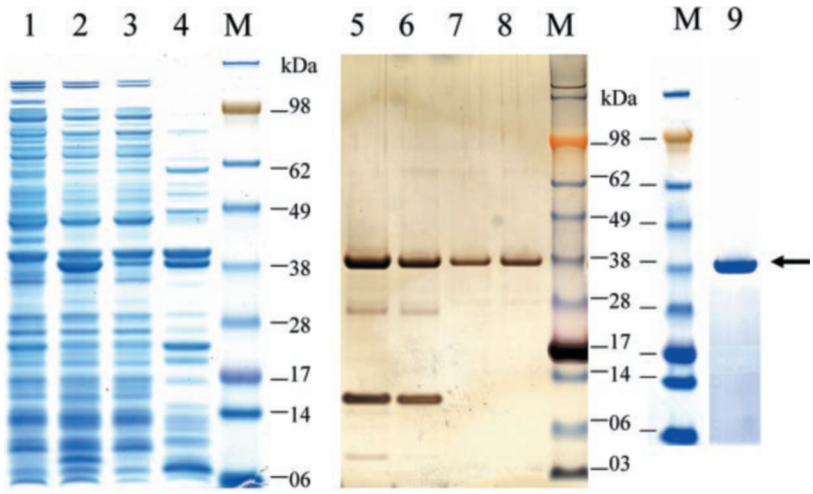


FIG. 2. **SDS-PAGE analysis of expression and purification of recombinant PAB1283.** SDS-PAGE was performed on a 4–12% gradient Novex bis-tris acrylamide gel (Invitrogen) with MOPS buffer. Polyacrylamide gels were stained either with BlueSafe Coomassie solution (*lanes 1–4* and *9*) or silver nitrate staining kit (*lanes 5–8*), both from Invitrogen. *Lane M*, SeeBlue Plus 2 molecular weight markers (Invitrogen); *lane 1*, cell-free extract of *E. coli* Rosetta(DE3)(pLysS)(pSBTN-AC18) uninduced (1.5 μl); *lane 2*, cell-free extract of *E. coli* Rosetta(DE3)(pLysS)(pSBTN-AC18) isopropyl-γ-D-thiogalactopyranoside-induced (1.5 μl); *lane 3*, soluble proteins from cell-free extract (1.5 μl); *lane 4*, soluble proteins after the 20-min heat treatment at 65 °C (10 μl); *lane 5*, proteins eluted from the immobilized metal ion adsorption chromatographic column (2 μg); *lane 6*, proteins eluted from the G25 desalting column (2 μg); *lanes 7* and *8*, MonoQ eluate (1 μg); *lane 9*, hydroxyapatite eluate (2 μg). Bands corresponding to recombinant PAB1283 polypeptide are indicated with the *arrow*.

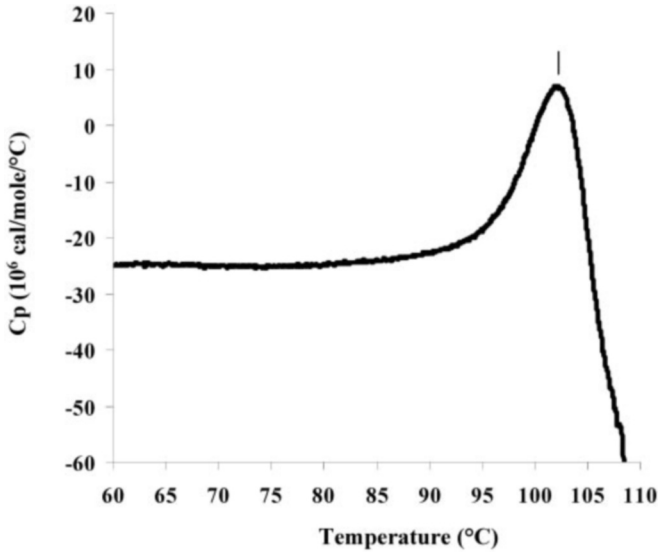


FIG. 3. **Evaluation of PAB1283 thermostability by differential scanning calorimetry.** Thermogram of PAB1283 was recorded at a scan rate of 1 °C/min. The heat capacity (*Cp*) maximum corresponding to the T_m value is indicated. Measurements were carried out at pH 8.0 (at 20 °C) with 30 mM HEPPS/NaOH rather than Tris/HCl buffer. The latter is not adapted for such measurements due to its temperature dependence ($\Delta pH/\Delta T = -0.028$ units/°C). Measurements were performed in fixed-in-place cells with a self-containing pressurizing system, in which temperatures up to 130 °C could be reached.

PAB1283 protein could interact with small nucleic acids such as transfer RNAs. After incubation of the enzyme with yeast bulk (unfractionated) tRNAs, the mixture was subjected to gel filtration on Superdex75, and the elution profile was compared with those obtained under identical conditions with the protein or bulk tRNAs alone. As shown in Fig. 4*A*, the results suggest a strong interaction between the enzyme and tRNAs, since a new peak of compound eluting faster than either the enzyme or tRNAs alone was visible. Identical results were obtained when *E. coli* total tRNAs were mixed with the PAB1283 protein (data not shown). If a stoichiometric amount of purified $tRNA^{Phe}$ or purified $tRNA^{Asp}$ (both from yeast) was mixed with the PAB1283 enzyme, a 1:1 stable complex was clearly visible (Fig. 4*B*). When PAB1283 and a 23 S rRNA fragment of approximately the same size (74 nucleotides) as a tRNA molecule (21) were mixed and subjected to gel filtration, a complex was also seen but eluted over a much larger fraction (Fig. 4*C*), revealing the existence of a 1:1 complex but of lower affinity. These results suggest that the substrate of PAB1283 is more likely to be tRNA than rRNA.

Identification of tRNA Substrates and Type of Methylation—The ubiquitous occurrence of orthologous genes (COG1041) corresponding to MTases present in both Archaea and Eukaryota but not in Bacteria (9), together with the observation that PAB1283 forms strong complexes with bulk or purified tRNAs (see above), prompted us to test whether tRNAs are substrates for methylation and to examine which of the known methylated nucleotides in archaeal and eukaryal tRNAs may be catalyzed by the newly identified putative MTase. Table I lists the positions and types of methylation that are known to occur simultaneously in tRNAs from both archaeon *Haloferax volcanii* (41 sequences known (36, 37)) and eukaryon *S. cerevisiae* (33 sequences known (38); see, on the World Wide Web, www.tRNA.uni-bayreuth.de/) but absent from any of the many bacterial tRNAs sequenced so far (49 sequences known (38); see also Refs. 39–41). Data indicate that, of the many modified nucleosides normally present in tRNAs, only N^2-methylguanosine, N^2,N^2-dimethylguanosine, and C^5-methylcytidine present at positions 10, 26, 40, 48, and/or 49 of a large number of tRNAs fulfill the above criteria. Formation of m^2G and/or $m^2{}_2G$ at position 26 was shown to be catalyzed by a single gene product (*TRM1*) in *S. cerevisiae* (42), *Caenorhabditis elegans* (43), *H. sapiens* (44), and *Pyrococcus furiosus* (45, 46), whereas formation of m^5C at positions 40, 48, and/or 49 in both *S. cerevisiae* and *P. abyssi* is catalyzed by the multisite-specific tRNA:m^5C MTase Trm4p (47).[3] Therefore, only m^2G and $m^2{}_2G$ at position 10 remain unassigned to any characterized tRNA MTases from Archaea and Eukaryota.

When bulk *E. coli* tRNAs (the majority of them have G10, and none contain methylated G10) and bulk tRNAs from yeast (most of them have m^2G10) were incubated *in vitro* at 50 °C with tritiated AdoMet and purified PAB1283, substantial

[3] S. Auxilien and H. Grosjean, unpublished results.

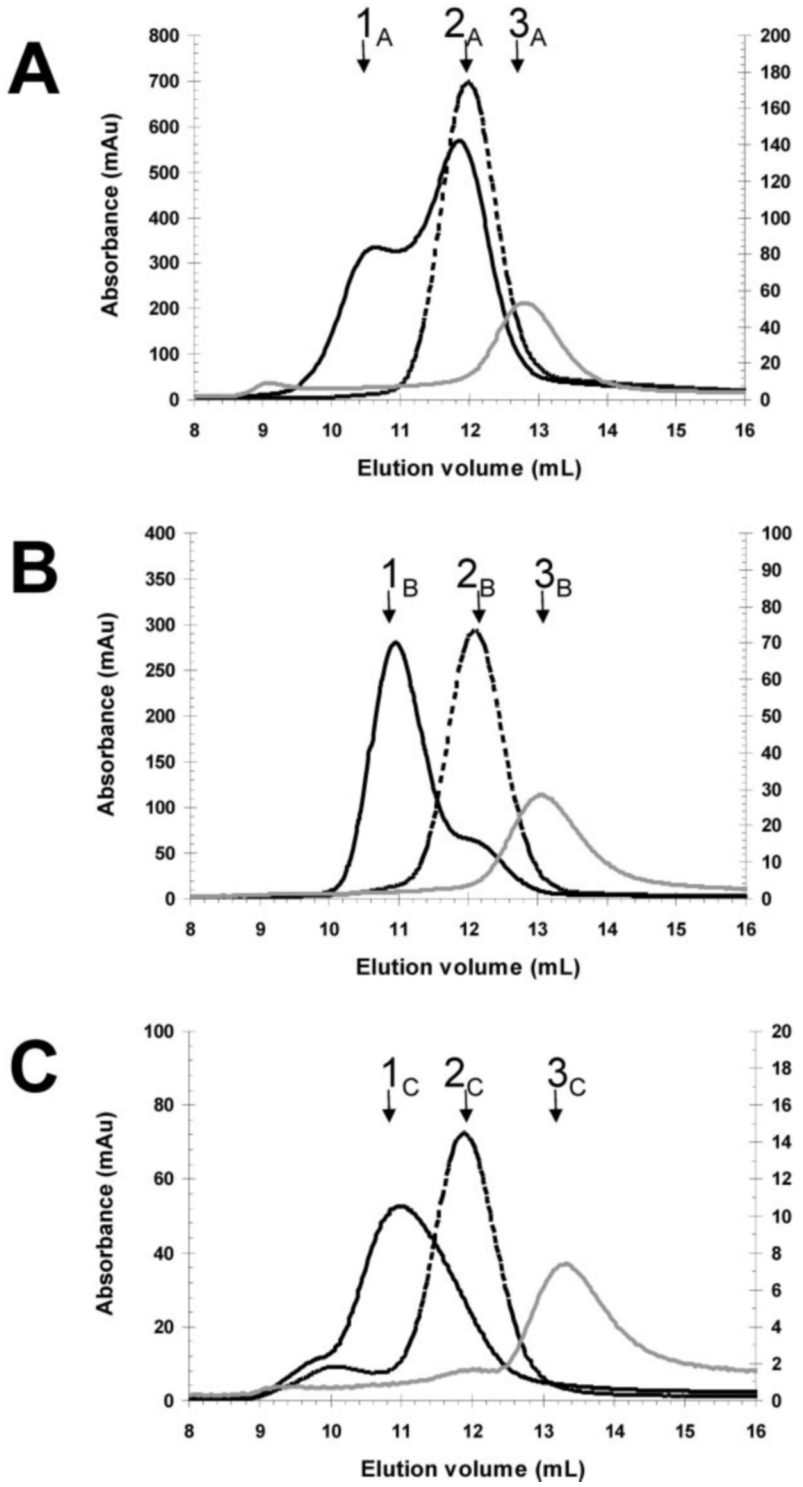


FIG. 4. **Molecular weight evaluation of PAB1283 and its complex with RNAs.** The estimation was performed by gel filtration on a calibrated Superdex75 column (Amersham Biosciences) eluted with 25 mM sodium phosphate buffer, pH 7.2, containing 5 mM $MgCl_2$ and 2 mM DTT. A gel filtration LMW calibration kit (Amersham Biosciences) was used for calibration. Elution volumes of ribonuclease A (14.7 kDa), chymotrypsinogen A (20.2 kDa), ovalbumin (47.2 kDa), albumin (61.6 kDa), and dextran blue 2000 (exclusion limit) are 16.5, 14.5, 12.0, 10.8, and 9.0 ml, respectively. Absorbance was monitored at 280 and 260 nm, but for clarity, only the signal at 280 nm is shown. Samples (60-μl final volume) consisted of the following: 140 μg of PAB1283 (*gray line*), 102 μg of bulk yeast tRNAs (*dotted line*), or a mixture of both (*solid line*) (*A*); 60 μg of PAB1283 (*gray line*), 34 μg of purified yeast $tRNA^{Asp}$ (*dotted line*), or a mixture of both (*solid line*) (*B*); and 23 μg of PAB1283 (*gray line*), 12 μg of purified 23 S rRNA fragment (74 nucleotides) (*dotted line*), or a mixture of both (*solid line*) (*C*). Maxima corresponding to the 1:1 protein-RNA complexes (1_{A-C}), the RNAs (2_{A-C}), and the protein (3_{A-C}) are indicated with *arrows*. At these maxima, calculated $A_{260\ nm}/A_{280\ nm}$ ratios are 1.9 (1_A, 1_B, and 1_C), 2.2 (2_A, 2_B, and 2_C), and 0.53 (3_A, 3_B, and 3_C), and elution volumes are 10.6 ml (1_A), 12.0 ml (2_A), 12.8 ml (3_A), 10.9 ml (1_B), 12.1 ml (2_B), 13.0 ml (3_B), 11.0 ml (1_C), 11.9 ml (2_C), and 13.3 ml (3_C). The *right-side scale* concerns the nucleic acid-free samples, and *left-side scale* concerns the other samples. Only 1:1 complexes were obtained when the enzyme/tRNA ratio was increased to 4.2.

TABLE I

Frequently methylated nucleosides that are exclusively present in tRNAs from Archaea and Eukaryota and never detected in tRNAs from Bacteria

The compilation of frequencies of a given nucleotide (G or C) and its modified derivatives $m^2G/m^2{}_2G$ or m^5C) are from Refs. 36–38.

Position in tRNA	Type of modification	Archaeon *H. volcanii* (41 sequences)	Eukaryon *S. cerevisiae* (33 sequences)	Bacteria *E. coli* + *B. subtilis* (49 sequences)
10	m^2G	4 of 41 G	19 of 33 G	None of 43 G
	$m^2{}_2G$	9 of 41 G	None of 33 G	None of 43 G
26	m^2G	2 of 25 G	1 of 23 G	None of 19 G
	$m^2{}_2G$	17 of 25 G	21 of 23 G	None of 19 G
40	m^5C	1 of 20 C	2 of 19 C	None of 34 C
48	m^5C	27 of 40 C	16 of 28 C	None of 30 C
49	m^5C	17 of 20 C	12 of 14 C	None of 5 C

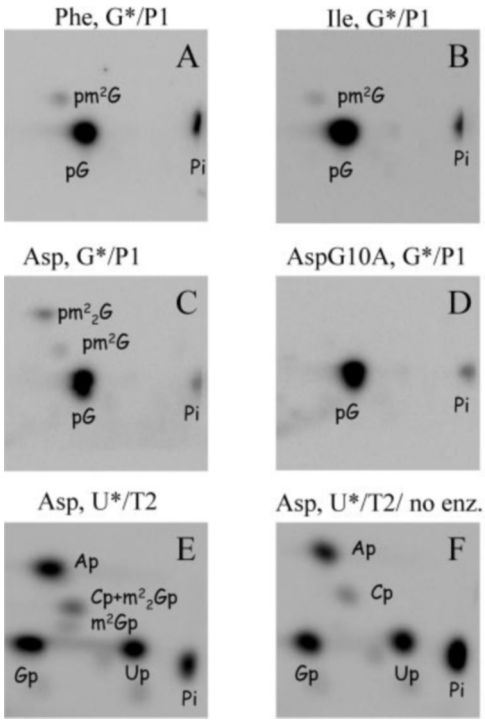


FIG. 5. **Characterization of the PAB1283 activity using different tRNA substrates.** T7 tRNA transcripts of yeast wild-type tRNA^{Phe} (*A*), tRNA^{Ile} (*B*), tRNA^{Asp} (*C* and *E*), and mutant yeast tRNA^{Asp} (G10A) (*D*), uniformly labeled with either [α-^{32}P]GTP or [α-^{32}P]UTP as indicated in the figure (*G** or *U**, respectively) were incubated for 1 h at 50 °C with purified recombinant PAB1283 protein under the experimental conditions described under "Materials and Methods." *F*, a negative control (no enzyme) of the same experiment as in *E*. After incubation, tRNA was phenol-extracted, ethanol-precipitated, and completely digested to monophosphate nucleosides by either nuclease P1 or RNase T2 (as indicated in the figure). Each hydrolysate was analyzed by two-dimensional TLC using chromatography system I (A + B) as in *A–D*, or chromatographic system II (A + C) as in *E* and *F* (for more details, see Ref. 23). Radiolabeled compounds on each TLC plate were detected and quantified by scanning using a PhosphorImager detector.

methyl group incorporation into bulk *E. coli* tRNAs was observed, whereas about half as much was incorporated into bulk yeast tRNAs, as expected if the G10 position in tRNA were the target of PAB1283 (data not shown). However, when incubated with the fragment of 23 S rRNA that binds to the enzyme (Fig. 4), no significant incorporation of tritiated methyl group was detected (data not shown). To demonstrate the catalytic activity of PAB1283 protein at position 10 of tRNAs, several [^{32}P]GMP radiolabeled intronless-tRNA substrates were prepared by *in vitro* transcription of synthetic genes corresponding to tRNA^{Phe}, tRNA^{Ile}, and tRNA^{Asp} of *S. cerevisiae*. In this eukaryon, the naturally occurring tRNA^{Phe} and tRNA^{Ile} both contain m^{2}G10, whereas in tRNA^{Asp} G10 is not modified. Each of the ^{32}P-radiolabeled transcripts was incubated at 50 °C with purified recombinant PAB1283 enzyme in the presence of AdoMet as the methyl donor. Under the assay conditions, we verified that the PAB1283 enzyme remains fully active even after 2 h of incubation at 80 °C (data not shown), although the stability of the tRNA substrates caused problems at higher temperatures. At the end of the incubation period, the modified tRNAs were completely digested by nuclease P1 into 5′-phosphate nucleosides, and the hydrolysates were analyzed by two-dimensional thin layer cellulose (TLC) chromatography. The radiolabeled spots corresponding to ^{32}P-labeled GMP-derivatives were detected by autoradiography and quantified. Results are illustrated in Fig. 5 (*A–F*). Using radiolabeled tRNA^{Phe} or tRNA^{Ile} as substrates, only two radioactive spots were visible on the TLC (in addition to the small amount of [^{32}P]inorganic phosphate; see Fig. 5, *A* and *B*). The heavily radioactive spot corresponds to [^{32}P]GMP (23 G are present in tRNA^{Phe} and 20 G in tRNA^{Ile}), whereas the faint spot corresponds to the fraction of G10 that was enzymatically converted into 5′-[^{32}P]m^{2}GMP (pm^{2}G). With tRNA^{Asp} as substrate (24 G), in addition to the expected [^{32}P]GMP, two other radioactive spots corresponding to pm^{2}G and pm$^2{}_2$G were also identified on the TLC (Fig. 5*C*).

Evidence that the formation of m^{2}G and m$^2{}_2$G in tRNA^{Asp} both occur at position 10 and not at position 26 was obtained after incubation of a tRNA^{Asp} mutant in which the G10 residue (which normally base-pairs in U25) was mutated into A10. Neither m^{2}G nor m$^2{}_2$G was obtained in this case (Fig. 5*D*). A lack of tRNA methylation was also observed with tRNA^{Phe} harboring a C10×G25 base pair in place of G10×C25 (data not shown). Moreover, formation of m^{2}G/m$^2{}_2$G at position 10 (5′ adjacent to a U in the sequence) and not at position 26 (5′ adjacent to a C) of tRNA^{Asp} was also shown by nearest neighbors analysis using [^{32}P]UTP-labeled tRNA^{Asp} and RNase T2 digestion. Under these conditions, the ^{32}P-labeled 3′-monophosphate nucleosides (5 A, 7 G + G derivatives, 5 U, and only 1 C) were detected (mixed with a small amount of CMP; compare with the control experiment in Fig. 5*F*) on the TLC. It is noteworthy that, in yeast tRNA^{Ile}, the nucleotide at position 26 is U (Ψ in naturally occurring yeast tRNA^{Ile} (48)). Therefore, the presence of the radioactive spot pm^{2}G on TLC (Fig. 5*B*) can originate only from modification of G10. As in the case of tRNA^{Asp}, reported above, the "nearest neighbors" analysis, using [^{32}P]CMP-labeled tRNA^{Ile} as substrate, demonstrated the presence of radioactive m^{2}GMP originating from a guanosine at position 10, 5′ adjacent to C11 in the sequence, after incubation with purified recombinant PAB1283 enzyme (data not shown).

Formation of m$^2{}_2$G in tRNA^{Asp} Is a Two-step Process in Which m^{2}G Is the Intermediate—Fig. 6, *A–D*, shows the time course of AdoMet-dependent methylation of guanosine 10 in three different tRNA synthetic substrates. Using transcripts corresponding to intronless yeast tRNA^{Phe} and intronless yeast tRNA^{Ile} (Fig. 6*A*), only m^{2}G10 accumulates during incubation at 50 °C with the purified enzyme, the reaction being more efficient with the tRNA^{Phe} than with tRNA^{Ile}. With transcript corresponding to yeast tRNA^{Asp} as substrate (Fig. 6*B*), m$^2{}_2$G10 accumulates during incubation at 50 °C, whereas the formation of m^{2}G10 appears to be a transient intermediate. The appearance of this intermediate was more obvious when incubation was performed at 40 °C (Fig. 6*C*) or 30 °C (Fig. 6*D*). At 30 °C, m^{2}G was efficiently formed, whereas the formation of m$^2{}_2$G10 was considerably slowed compared with that at higher temperatures. Taken together, these results demonstrate that *P. abyssi* PAB1283 is a AdoMet-dependent tRNA dimethyltransferase, now renamed $_{Pab}$Trm-G10.

DISCUSSION

PAB1283, a COG1041 Member, Is a New tRNA:G10 Dimethyltransferase—In this work, we report the characterization of a new archaeal tRNA MTase responsible for mono- and dimethylation of N^2-exocyclic amino group of guanosine at position 10 of various tRNAs. This enzyme, encoded by the *PAB1283* gene in the archaeon *P. abyssi* and now renamed $_{Pab}$Trm-G10, is a highly thermostable monomeric protein that forms a stable 1:1 complex with its tRNA substrate. It belongs to the COG1041 group of orthologous proteins that are found in all completely sequenced Archaea and almost all Eukaryota. This observation correlates well with the presence of N^2-methylated G10 residues in many of the cytoplasmic tRNAs of eukaryotic and of archaeal origin, whereas no methylated G10 residues have

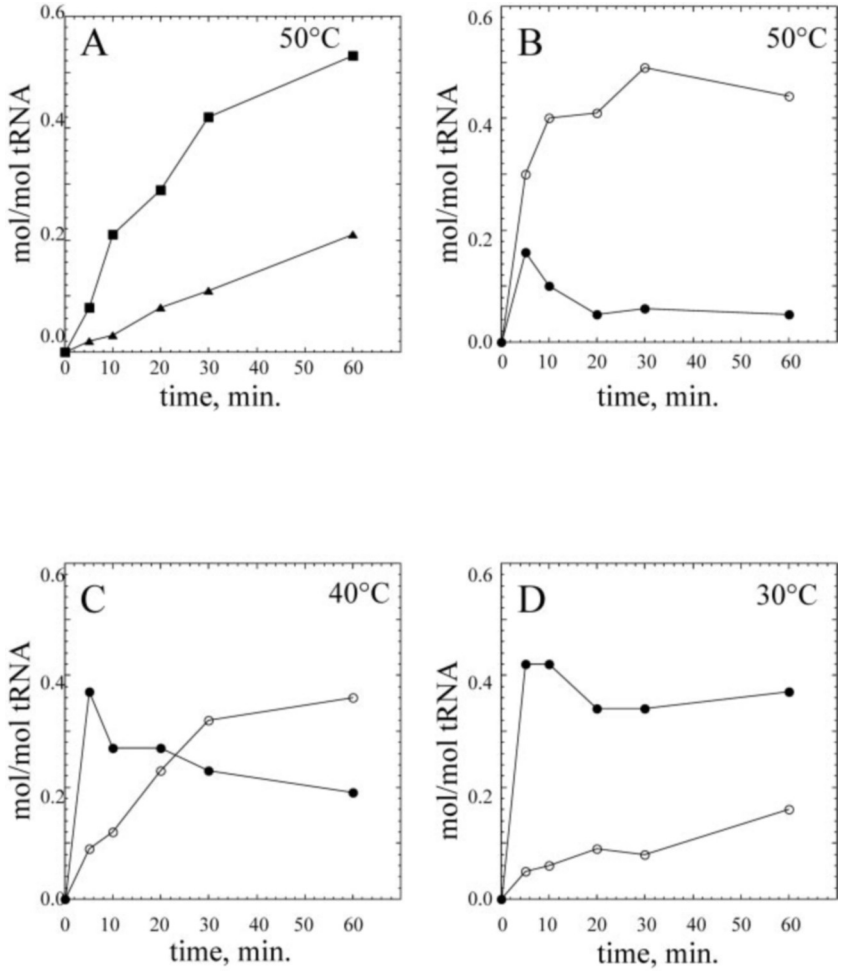


FIG. 6. **Time course of m^2G10/$m^2{}_2G10$ formation catalyzed by PAB1283 protein.** Experiments were carried out as reported in Fig. 5 (*A–D*), with GTP-labeled tRNA transcripts, except that methylation was estimated at different time points. The temperature of incubation is indicated in each *panel*. Estimation of radioactivity in each spot on the TLC plates allowed determination of the molar ratio of modified m^2G10 over total G in transcript of $tRNA^{Phe}$ (*closed squares* in *A*) and in transcript of $tRNA^{Ile}$ (*closed triangles* in *A*). In the transcript of $tRNA^{Asp}$ (*B–D*), the molar ratio of m^2G10 (*closed circles*) and $m^2{}_2G10$ (*open circles*) over total G was evaluated over time at 50, 40, and 30 °C, respectively. Incubation temperature for the experiment in *A* was 50 °C.

been found in any bacterial tRNA sequenced so far (38). Because of the high sequence similarity between the various PAB1283 orthologs and their occurrence pattern related to the phyletic occurrence of the m^2G10 modification, it is reasonable to infer that most, if not all, COG1041 members catalyze the same methylation reaction at G10 of tRNA. During annotation of *Methanosarcina acetivorans* C2A (49), a putative tRNA:m^2G MTase function was attributed to MA0209, an ortholog of $_{Pab}$Trm-G10. Experimental characterization of MJ0710 from *Methanocaldococcus jannaschii* supports this annotation (50). However, no experimental evidence has been reported concerning either the characterization of the corresponding gene product or the identification of the position and the atom within the tRNA molecule supposedly acted upon by any member of COG1041. Our results with the *P. abyssi* enzyme PAB1283 not only support the tRNA:m^2G MTase function but also allow the identification of the target within the tRNA substrate and the final methylation product. As far as the eukaryotic homologs of PAB1283 are concerned, the identification and characterization of an eukaryotic MTase responsible for the formation of m^2G10 in tRNA was demonstrated many years ago using partially purified enzymes originating from different eukaryotic cells (51–55). In these studies, the heterologous tRNAs from *E. coli* harboring an unmethylated G10 residue were used as substrates for the *in vitro* tests. However, the protein sequences and the corresponding eukaryotic genes remain to be fully characterized (56). The ubiquitous occurrence of a MTase acting on the exocyclic amino group of G10 in both Eukaryota and Archaea suggests that the methyl group this enzyme introduces into the tRNAs serves some important, but yet to be determined, physiological function common to these two domains of life.

The Archaeal Trm-G10 Has the Unique Property of Catalyzing Formation of $m^2{}_2G$ at Position 10 of tRNAs—The presence of m^2G10 in many cytoplasmic tRNAs of eukaryotes, including the yeast *S. cerevisiae*, is well documented, and in no case has the presence of $m^2{}_2G10$ been reported (see Ref. 38). In contrast, the few archaeal tRNAs sequenced so far, mostly from the archaeon *H. volcanii* (36, 37), showed a greater prevalence of $m^2{}_2G10$ with respect to m^2G10 (Table I). Moreover, the yeast $tRNA^{Asp}$, which is obviously a good substrate for $_{Pab}$Trm-G10 *in vitro* (Figs. 5 and 6), is not naturally methylated at position 10 within the yeast cell (57). The formation of $m^2{}_2G10$ in yeast $tRNA^{Asp}$ catalyzed by purified $_{Pab}$Trm-G10 was also identified in *P. abyssi* $tRNA^{Asp}$ (data not shown). As observed in this study, dimethylation is a two-step process. It is possible that the ability to carry out the second methylation evolved only in the archaeal Trm-G10 lineage, after radiation of the eukaryotic Trm-G10 lineage. This hypothesis can only be validated by a thorough comparative analysis of the biochemical activities and substrate specificities of members of both these lineages.

Trm-G10 Is Distinct from Trm-G26 (Alias Trm1), Which Also Catalyzes the Formation of $m^2{}_2G$—In all sequenced genomes of Archaea and Eukaryota, an ubiquitous gene for a second tRNA:$m^2{}_2G$ dimethyltransferase is present, whereas in Bacteria, members of the Trm1 family (COG1867) have been identified in only three genomes of 148 examined (27, 58). This pattern suggests that tRNA:$m^2{}_2G26$ methyltransferase, called Trm1 and renamed here as Trm-G26, may be also considered as a "PACE" protein. The specificity of this enzyme has been care-

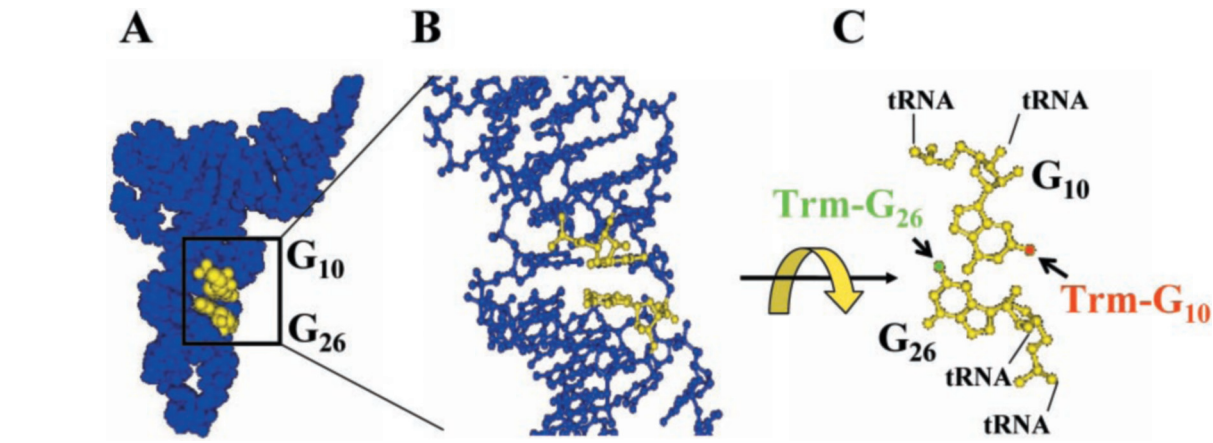
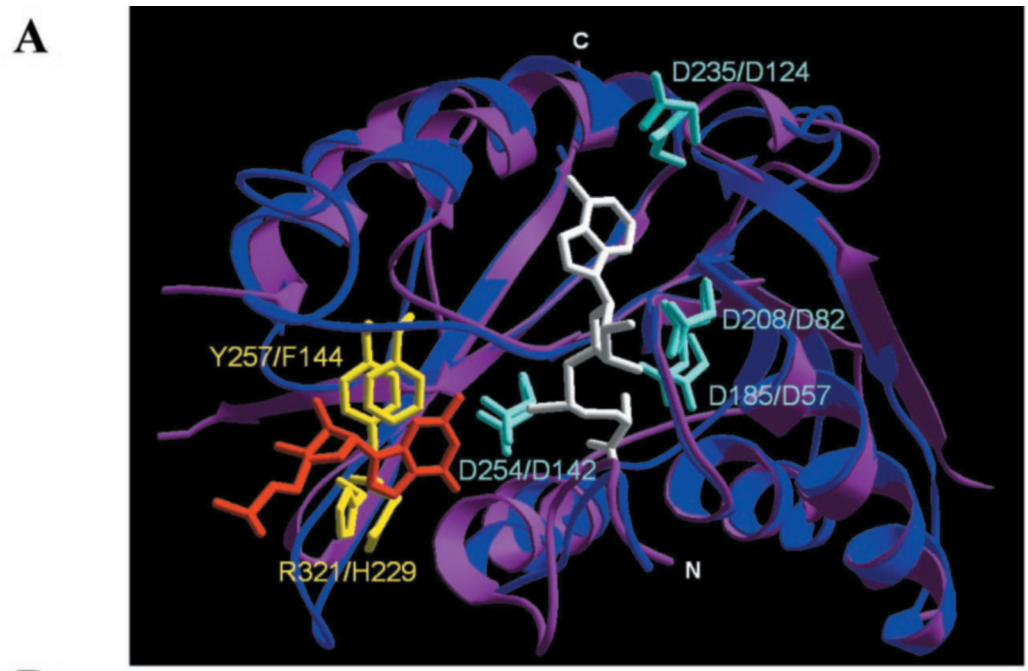


FIG. 7. **Reactions catalyzed by two distinct tRNA:guanosine dimethyltransferases in tRNAs.** Shown on a three-dimensional structure of yeast tRNA^{Asp} (2tra in the Protein Data Bank) is the location of guanosines (*yellow*) at positions 10 and 26 (*A*, whole view; *B*, zoomed view; *C*, rotated view). Note that, in naturally occurring yeast tRNA^{Asp}, guanosines at positions 10 and 26 are unmodified. As indicated by the *arrows*, the exocyclic N^2 atoms of G10 (in *red*), which would be modified by $_{Pab}$Trm-G10 (this work), and G26 (in *green*), which would be modified by $_{Pab}$Trm-G26 (PAB2092) based on data obtained on its closely related ortholog, $_{Pfu}$Trm-G26 (also known as $_{Pab}$Trm1 (41, 42)), point in different directions.



B

	X	I	II	III
PAB2092	MALNRDIAVVLLNVL	RIVLDALSATGIRGIRFAL	EEIWMNDISEDAYNLMKKNVLLNF	KTLVVNHDDANRLMA
PAB1283	PPRIARAMVNLTRAT	REILDPFMGTGGMLIEAGL	LKVYGIDIREDMVEGAKINLEYYG	KDYVVKVGDATKIKE

	IV	V&VI	VII	VIII
PAB2092	FHFIDLDP-FG	SPMEFLDTALRSVKRKGILGITAT	VGVVARYAAK	DHYFRAFIKL
PAB1283	FEAIATDPPYG	LYKRALESMYSVLEGRLAIAFPSD	ALDVAETIGF	SRYFYIMEAR

FIG. 8. **Comparison of the AdoMet-binding and active sites of $_{Pab}$Trm-G10 (PAB1283) and $_{Pab}$Trm-G26 (PAB2092).** *A*, superposition of modeled structures of the catalytic domain of both enzymes. Protein models were built based on the sequence-structure alignments generated by the GeneSilico protein structure prediction metaserver (77). The 1dus.pdb structure was selected as the best scoring template, and the models were built according to the "FRankenstein's monster" protocol (78). Both models passed the evaluation procedure implemented in Verify3D (79). The backbone of PAB1283 (amino acids 161–329) and PAB2092 (amino acids 27–238) is shown in *blue* and *magenta*, respectively. The N and C termini are labeled. Two insertions in PAB2092 (amino acids 100–114 and 173–199) were omitted from the model. The ligand-binding residues are shown and labeled (PAB1283/PAB2092); identical residues (four Asp residues from motifs I, II, III, and IV) are shown in *cyan*, and different residues (from motifs IV and VIII) are shown in *yellow*. AdoMet is shown in *white*, and guanosine is shown in *red*. *B*, sequence alignment of the common MTase motifs, depicting strong differences in all residues except those participating in cofactor binding. Residues shown in *A* are indicated in *red*.

fully tested with different orthologs from several organisms (43, 44, 52, 55, 59–63). This enzyme catalyzes the site-specific formation of m^2G and of $m^2{}_2G$ at position 26 by a mechanism very similar to that reported in this paper for Trm-G10 (46, 60) (reviewed in Ref. 64). It is important to note that the targets G10 and G26 are very close to each other within the three-dimensional structure of the tRNA (in fact they stack on each other; see Fig. 7, *A* and *B*). However, the exocyclic N^2 atoms of G10 and G26 point in different directions (Fig. 7*C*). Therefore, it may be that the two enzymes have to approach the tRNA

substrate from different sides, which could explain why the two enzymes are very different in their amino acid sequences and belong to completely different clusters of orthologous groups (COG1867 for Trm-G26; COG1041 for Trm-G10). The differences in site specificity between archaeal Trm-G10 and Trm-G26 may be explained by several factors such as specific enzyme positioning on the tRNA, the local environment of the active site amino acids, and base composition and sequence of the tRNA. Any of these might facilitate or impede flipping out the guanosine residue into the active site when bound to each enzyme. This question is difficult to tackle, since alternative tertiary structures of bound tRNA may be crucial as, for example, in the structural analysis reported for two other post-transcriptional modification enzymes (65, 66).

To obtain better insight into the potential similarities and/or differences in the active sites of the two m^2_2G dimethyltransferases, structures of both enzymes (PAB1283 (*i.e.* Trm-G10) and PAB2092 (*i.e.* Trm-G26)) from the same organism (*P. abyssi*) were homology-modeled using MJ0882 (1dus in the Protein Data Bank (26)) as a structural template and compared. Fig. 8 shows the structure-based alignment of putative AdoMet-binding and base-binding motifs from both enzymes. It is clear that whereas both enzymes share the three key AdoMet-binding Asp residues, the arrangement of their putative base-binding residues is quite different. Although Trm-G10 and its "orthodox" homologs show a typical "DPPY" catalytic motif, in Trm-G26 this is replaced by a "DPFG" tetrapeptide. Consequently, the aromatic residue (Tyr or Phe) believed to stabilize the methylated base is found in a different position. It is impossible for the DPPY and DPF peptides to position the Asp and Phe/Tyr side chains in the same location and at the same time maintain the same conformation of the backbone. Hence, the shapes of the base-binding pockets of Trm-G10 and Trm-G26 must be at least partially different. Interestingly, both enzymes share a conserved positively charged residue in motif VIII (Arg in nearly all Trm-G10 and its homologs and His in all Archaeal homologs of Trm-G26), which may be involved in substrate recognition (Fig. 8). The comparative model of enzyme-substrate interactions proposed here for these two functionally similar, yet sequentially different, enzymes should stimulate further studies on their sequence-structure-function relationships.

Trm-G10 Contains a THUMP Domain Absent from Trm-G26—The additional domain fused to the MTase domain further differentiates the two tRNA:m^2_2G dimethyltransferases; an N-terminal THUMP is fused to the Trm-G10 MTase domain, whereas another type of domain (putative wHTH relative) is fused to the C terminus of the Trm-G26 catalytic domain (27). No experimentally proven function has been assigned to these additional domains, but they are both predicted to target the catalytic domain onto the tRNA substrate. The presence of the THUMP domain is a conserved trait for all COG1041 proteins. This domain is present in many proteins other than Trm-G10 from the three domains of life (15). To our knowledge, $_{Pab}$Trm-G10 is the third biochemically characterized protein that contains such a domain. The THUMP-containing ThiI protein from *E. coli* is essential for 4-thiouridine formation at position 8 of some bacterial tRNAs (67, 68). Very recently, Tan1 protein from *S. cerevisiae* also harboring a THUMP domain was reported to be required for N^4-acetylcytidine formation in tRNAs (69). This modification is present at position 12 in tRNAs specific for leucine and serine and was shown to be important for tRNA stability. It should be noted that, in agreement with the putative RNA-binding role that may be attributed to the THUMP, the three characterized THUMP-containing proteins, ThiI, Tan1, and Trm-G10, are all involved in site-specific modification within the same region of structured tRNA (at positions 8, 12, and 10, respectively). It is tempting to speculate that the THUMP domain may interact with a specific region of tRNA and targets the modification domain, at least in ThiI, Tan1, and Trm-G10, toward the central three-dimensional core of the tRNA molecule (see Fig. 7).

Dimethylation of G10 as of G26 Probably Has a Structural Role—One characteristic of archaeal tRNAs, especially in hyperthermophilic organisms, is the high G/C content in their stem regions (70), a situation that favors misfolding of tRNAs, as seen with tRNA transcripts produced *in vitro* (71). In eukaryotic tRNAs, G/C-rich stems are also found but less frequently than in archaeal tRNAs. N^2-Methylguanosine base-pairs with either cytidine or uridine, whereas N^2,N^2-dimethylguanosine base-pairs only with uridine. N^2-methylguanosine has been shown to be isoenergetic with guanosine in RNA duplexes, suggesting that its role may be different than modulation of duplex stability (72). A clear correlation between the presence of N^2,N^2-dimethylguanosine at either position 10 or 26 and possible alternative tRNA conformers in various archaeal and eukaryal tRNAs has been well documented (73). Thus, one role of Trm-G10/G26-mediated modification is most probably to help tRNA to fold into a correct cloverleaf/three-dimensional structure or to avoid possible misfolding during one of the numerous steps during maturation and translation processes. Such a structural role has been demonstrated in the case of N^1-methylation of A9 in human mitochondrial tRNA catalyzed by an as yet unidentified MTase (74). This structural function does not exclude the possibility that methylation of the N^2-amine of G10/G26 can play more direct roles such as serving as a cipher to modulate enzyme-protein recognition or to promote accurate translation of mRNA on the ribosomes (see, for example, Refs. 75 and 76). That Trm-G10 and Trm-G26 are both highly conserved "PACE" proteins as reported in this study indicates that the corresponding modifications are definitely ubiquitous and important for archaeal and eukaryal organisms. Although some slight differences have been noted between the proteins from these two domains of life, a clear distinct evolution of tRNAs maturation strategies is shown in this study between these two domains of life and Bacteria.

Acknowledgments—We gratefully acknowledge Anne Theobald-Ditriech (CNRS-IBMC, Strasbourg, France) for the generous gift of purified yeast $tRNA^{Asp}$, Dominique Fourmy and Clothilde Husson (both from CNRS-ICSN, Gif-sur-Yvette, France) for kindly providing a purified 23 S rRNA fragment, Sylvie Auxilien (CNRS-LEBS, Gif-sur-Yvette, France) for preparation of a plasmid carrying $_{Pab}tRNA^{Asp}$, Hannu Myllykallio (CNRS-IGM, Orsay, France) for the generous gift of *P. abyssi* chromosomal DNA, and Patrick Forterre (Université d'Orsay, Orsay, France) for initial discussions of the interest of characterizing proteins belonging to the "PACE18" group. We are indebted to Suresh K. Purushothaman and Bruno Lapeyre (both from CNRS-CRBM, Montpellier, France) for discussions regarding their ongoing project on COG1041 yeast ortholog, Marcus Johansson and Anders Byström (both from Umeå University, Umeå, Sweden) for sharing information on TAN1 prior to publication, and Gary J. Olsen (University of Illinois, Urbana-Champaign, IL) for making us aware, during submission of this manuscript, of their unpublished studies on MJ0710. We thank the following colleagues (all from CEA VALRHO): Isabelle Dany, Jean-Charles Gaillard, and Alain Dedieu for assistance with mass spectrometry controls, Valérie Chaumont for help in overexpression analysis, Charles Marchetti for operating the fermenter facilities, Anne-Hélène Davin for help with differential scanning calorimetry analyses, and Eric Quéméneur for constant support.

REFERENCES

1. Woese, C. R., Kandler, O., and Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 4576–4579
2. Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., and Natale, D. A. (2003) *BMC Bioinformatics* **4,** 41
3. Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997) *Science* **278,** 631–637
4. Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram,

U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D., Koonin, E. V., and Lipman, D. J. (2001) *Nucleic Acids Res.* **29,** 22–28
5. Dennis, P. P. (1997) *Cell* **89,** 1007–1010
6. Olsen, G. J., and Woese, C. R. (1997) *Cell* **89,** 991–994
7. Makarova, K. S., and Koonin, E. V. (2003) *Genome Biol.* **4,** 115
8. Terns, M. P., and Terns, R. M. (2002) *Gene. Expr.* **10,** 17–39
9. Matte-Tailliez, O., Zivanovic, Y., and Forterre, P. (2000) *Trends Genet.* **16,** 533–536
10. Hartmann, E., and Hartmann, R. K. (2003) *Trends Genet.* **19,** 561–569
11. Fromant, M., Ferri-Fioni, M. L., Plateau, P., and Blanquet, S. (2003) *Nucleic Acids Res.* **31,** 3227–3235
12. Rosas-Sandoval, G., Ambrogelly, A., Rinehart, J., Wei, D., Cruz-Vera, L. R., Graham, D. E., Stetter, K. O., Guarneros, G., and Soll, D. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 16707–16712
13. Armengaud, J., Fernandez, B., Chaumont, V., Rollin-Genetet, F., Finet, S., Marchetti, C., Myllykallio, H., Vidaud, C., Pellequer, J. L., Gribaldo, S., Forterre, P., and Gans, P. (2003) *J. Biol. Chem.* **278,** 31078–31087
14. Marchler-Bauer, A., Anderson, J. B., DeWeese-Scott, C., Fedorova, N. D., Geer, L. Y., He, S., Hurwitz, D. I., Jackson, J. D., Jacobs, A. R., Lanczycki, C. J., Liebert, C. A., Liu, C., Madej, T., Marchler-Bauer, A., Marchler, G. H., Mazumder, R., Nikolskaya, A. N., Rao, B. S., Panchenko, A. R., Shoemaker, B. A., Simonyan, V., Song, J. S., Thiessen, P. A., Vasudevan, S., Wang, Y., Yamashita, R. A., Yin, J. J., and Bryant, S. H. (2003) *Nucleic Acids Res.* **31,** 383–387
15. Aravind, L., and Koonin, E. V. (2001) *Trends Biochem. Sci* **26,** 215–217
16. Fauman, E. B., Blumenthal, R. M., and Cheng, X. (1999) in *S-Adenosylmethionine-dependent Methyltransferases: Structures and Functions* (Cheng, X., and Blumenthal, R. M., eds) pp. 1–38, World Scientific Publishing, Singapore
17. Rozenski, J., Crain, P. F., and McCloskey, J. A. (1999) *Nucleic Acids Res.* **27,** 196–197
18. Hopper, A. K., and Phizicky, E. M. (2003) *Genes Dev.* **17,** 162–180
19. Grosjean, H., and Benne, R. (1998) *Modification and Editing of RNA* (Grosjean, H., and Benne, R., eds) American Society for Microbiology Press, Washington, D. C.
20. Gill, S. C., and von Hippel, P. H. (1989) *Anal. Biochem.* **182,** 319–326
21. Lebars, I., Yoshizawa, S., Stenholm, A. R., Guittet, E., Douthwaite, S., and Fourmy, D. (2003) *EMBO J.* **22,** 183–192
22. Grosjean, H., Droogmans, L., Giege, R., and Uhlenbeck, O. C. (1990) *Biochim. Biophys. Acta* **1050,** 267–273
23. Grosjean, H., Keith, G., and Droogmans, L. (2004) *Methods Mol. Biol.* **265,** 357–392
24. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402
25. Cheng, X. (1995) *Curr. Opin. Struct. Biol.* **5,** 4–10
26. Bujnicki, J. M., and Rychlewski, L. (2002) *BMC Bioinformatics* **3,** 10
27. Bujnicki, J. M., Leach, R. A., Debski, J., and Rychlewski, L. (2002) *J. Mol. Microbiol. Biotechnol.* **4,** 405–415
28. Bugl, H., Fauman, E. B., Staker, B. L., Zheng, F., Kushner, S. R., Saper, M. A., Bardwell, J. C., and Jakob, U. (2000) *Mol. Cell* **6,** 349–360
29. Bujnicki, J. M. (2000) *FASEB J.* **14,** 2365–2368
30. Gong, W., O'Gara, M., Blumenthal, R. M., and Cheng, X. (1997) *Nucleic Acids Res.* **25,** 2702–2715
31. Goedecke, K., Pignot, M., Goody, R. S., Scheidig, A. J., and Weinhold, E. (2001) *Nat. Struct. Biol.* **8,** 121–125
32. Fitch, W. M. (2000) *Trends Genet.* **16,** 227–231
33. Bujnicki, J. M., Feder, M., Radlinska, M., and Blumenthal, R. M. (2002) *J. Mol. Evol.* **55,** 431–444
34. Scholz, S., Sonnenbichler, J., Schafer, W., and Hensel, R. (1992) *FEBS Lett.* **306,** 239–242
35. Backmann, J., and Schafer, G. (2001) *Methods Enzymol.* **334,** 328–342
36. Gupta, R. (1984) *J. Biol. Chem.* **259,** 9461–9471
37. Gupta, R. (1986) *System. Appl. Microbiol.* **7,** 102–105
38. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. (1998) *Nucleic Acids Res.* **26,** 148–153
39. Limbach, P. A., Crain, P. F., and McCloskey, J. A. (1994) *Nucleic Acids Res.* **22,** 2183–2196
40. Grosjean, H., Sprinzl, M., and Steinberg, S. (1995) *Biochimie* (*Paris*) **77,** 139–141
41. Motorin, Y., and Grosjean, H. (2004) *Encyclopedia of Life Sciences*, p. A3866, MacMillan Reference Ltd., Basingstoke (Hampshire), England
42. Ellis, S. R., Morales, M. J., Li, J. M., Hopper, A. K., and Martin, N. C. (1986) *J. Biol. Chem.* **261,** 9703–9709
43. Liu, J., Zhou, G. Q., and Straby, K. B. (1999) *Gene* (*Amst.*) **226,** 73–81
44. Liu, J., and Straby, K. B. (2000) *Nucleic Acids Res.* **28,** 3445–3451
45. Constantinesco, F., Benachenhou, N., Motorin, Y., and Grosjean, H. (1998) *Nucleic Acids Res.* **26,** 3753–3761
46. Constantinesco, F., Motorin, Y., and Grosjean, H. (1999) *J. Mol. Biol.* **291,** 375–392
47. Motorin, Y., and Grosjean, H. (1999) *RNA* **5,** 1105–1118
48. Simos, G., Tekotte, H., Grosjean, H., Segref, A., Sharma, K., Tollervey, D., and Hurt, E. C. (1996) *EMBO J.* **15,** 2270–2284
49. Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., Macdonald, P., FitzHugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D., Brown, A., Allen, N., Naylor, J., Stange-Thomann, N., DeArellano, K., Johnson, R., Linton, L., McEwan, P., McKernan, K., Talamas, J., Tirrell, A., Ye, W., Zimmer, A., Barber, R. D., Cann, I., Graham, D. E., Grahame, D. A., Guss, A. M., Hedderich, R., Ingram-Smith, C., Kuettner, H. C., Krzycki, J. A., Leigh, J. A., Li, W., Liu, J., Mukhopadhyay, B., Reeve, J. N., Smith, K., Springer, T. A., Umayam, L. A., White, O., White, R. H., de Macario, E. C., Ferry, J. G., Jarrell, K. F., Jing, H., Macario, A. J. L., Paulsen, I., Pritchett, M., Sowers, K. R., Swanson, R. V., Zinder, S. H., Lander, E., Metcalf, W. W., and Birren, B. (2002) *Genome Res.* **12,** 532–542
50. Graham, D. E. (2000) *Archaeal Gene Identification*. Ph.D. thesis, University of Illinois at Urbana-Champaign
51. Kraus, J., and Staehelin, M. (1974) *Nucleic Acids Res.* **1,** 1455–1478
52. Glick, J. M., Averyhart, V. M., and Leboy, P. S. (1978) *Biochim. Biophys. Acta* **518,** 158–171
53. Izzo, P., and Gantt, R. (1977) *Biochemistry* **16,** 3576–3581
54. Taylor, M. J., and Gantt, R. (1979) *Biochemistry* **18,** 5253–5258
55. Sindhuphak, T., Hellman, U., and Svensson, I. (1985) *Biochim. Biophys. Acta* **824,** 66–73
56. Bujnicki, J. M., Droogmans, L., Grosjean, H., Purushothaman, S. K., and Lapeyre, B. (2004) *Nucleic Acids Mol. Biol.* **15,** 139–168
57. Gangloff, J., Keith, G., Ebel, J. P., and Dirheimer, G. (1972) *Biochim. Biophys. Acta* **259,** 210–222
58. Takeda, H., Hori, H., and Endo, Y. (2002) *Nucleic Acids Res. Suppl.* **2,** 229–230
59. Liu, J., and Straby, K. B. (1998) *Nucleic Acids Res.* **26,** 5102–5108
60. Reinhart, M. P., Lewis, J. M., and Leboy, P. S. (1986) *Nucleic Acids Res.* **14,** 1131–1148
61. Kraus, J., and Staehelin, M. (1974) *Nucleic Acids Res.* **1,** 1479–1496
62. Edqvist, J., Grosjean, H., and Straby, K. B. (1992) *Nucleic Acids Res.* **20,** 6575–6581
63. Edqvist, J., Blomqvist, K., Straby, K. B., and Grosjean, H. (1994) *Biochemistry* **33,** 9546–9551
64. Edqvist, J., Straby, K. B., and Grosjean, H. (1995) *Biochimie* (*Paris*) **77,** 54–61
65. Hoang, C., and Ferre-D'Amare, A. R. (2001) *Cell* **107,** 929–939
66. Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S., and Yokoyama, S. (2003) *Cell* **113,** 383–394
67. Palenchar, P. M., Buck, C. J., Cheng, H., Larson, T. J., and Mueller, E. G. (2000) *J. Biol. Chem.* **275,** 8283–8286
68. Mueller, E. G., Palenchar, P. M., and Buck, C. J. (2001) *J. Biol. Chem.* **276,** 33588–33595
69. Johansson, M. J., and Byström, A. S. (2004) *RNA* **10,** 712–719
70. Marck, C., and Grosjean, H. (2002) *RNA* **8,** 1189–1232
71. Uhlenbeck, O. C. (1995) *RNA* **1,** 4–6
72. Rife, J. P., Cheng, C. S., Moore, P. B., and Strobel, S. A. (1998) *Nucleic Acids Res.* **26,** 3640–3644
73. Steinberg, S., and Cedergren, R. (1995) *RNA* **1,** 886–891
74. Helm, M., Brule, H., Degoul, F., Cepanec, C., Leroux, J. P., Giege, R., and Florentz, C. (1998) *Nucleic Acids Res.* **26,** 1636–1643
75. Roe, B., Michael, M., and Dudock, B. (1973) *Nat. New Biol.* **246,** 135–138
76. Niederberger, C., Graub, R., Costa, A., Desgres, J., and Schweingruber, M. E. (1999) *FEBS Lett.* **464,** 67–70
77. Kurowski, M. A., and Bujnicki, J. M. (2003) *Nucleic Acids Res.* **31,** 3305–3307
78. Kosinski, J., Cymerman, I. A., Feder, M., Kurowski, M. A., Sasin, J. M., and Bujnicki, J. M. (2003) *Proteins* **53,** 369–379
79. Luthy, R., Bowie, J. U., and Eisenberg, D. (1992) *Nature* **356,** 83–85