Acquisition of a bacterial RumA-type tRNA(uracil-54, C5)-methyltransferase by Archaea through an ancient horizontal gene transfer.

Jaunius Urbonavicius, Sylvie Auxilien, Hélène Walbott, Kalliopi Trachana, Béatrice Golinelli-Pimpaneau, Céline Brochier-Armanet, Henri Grosjean

Molecular Microbiology (2008) 67(2), 323–335 ■
doi:10.1111/j.1365-2958.2007.06047.x
First published online 9 December 2007

Acquisition of a bacterial RumA-type tRNA(uracil-54, C5)-methyltransferase by Archaea through an ancient horizontal gene transfer

Jaunius Urbonavičius,[1‡] Sylvie Auxilien,[1‡] Hélène Walbott,[1] Kalliopi Trachana,[1] Béatrice Golinelli-Pimpaneau,[1] Céline Brochier-Armanet[2,†‡] and Henri Grosjean[1,3*‡]

[1] *Laboratoire d'Enzymologie et Biochimie Structurales, CNRS, 1 avenue de la Terrasse, 91190, Gif-sur-Yvette, France.*

[2] *EA 3781 EGEE (Evolution Génome, Environnement), Université de Provence, Aix-Marseille I, Marseille, France.*

[3] *Univ Paris-Sud, IGM, UMR8621, F-91405 Orsay, France.*

Summary

The *Pyrococcus abyssi* genome displays two genes possibly coding for S-adenosyl-L-methionine-dependent RNA(uracil, C5)-methyltransferases (PAB0719 and PAB0760). Their amino acid sequences are more closely related to *Escherichia coli* RumA catalysing the formation of 5-methyluridine (m^5U)-1939 in 23S rRNA than to *E. coli* TrmA (tRNA methyltransferaseA) methylating uridine-54 in tRNA. Comparative genomic and phylogenetic analyses show that homologues of PAB0719 and PAB0760 occur only in a few Archaea, these genes having been acquired via a single horizontal gene transfer from a bacterial donor to the common ancestor of Thermococcales and Nanoarchaea. This transfer event was followed by a duplication event in Thermococcales leading to two closely related genes. None of the gene products of the two *P. abyssi* paralogues catalyses *in vitro* the formation of m^5U in a *P. abyssi* rRNA fragment homologous to the bacterial RumA substrate. Instead, PAB0719 enzyme (renamed $_{Pab}$TrmU54) displays an identical specificity to TrmA, as it catalyses the *in vitro* formation of m^5U-54 in tRNA. Thus, during evolution, at least one of the two *P. abyssi* RumA-type enzymes has changed of target specificity. This functional shift probably occurred in an ancestor of all Thermococcales. This study also provides new evidence in favour of a close relationship between Thermococcales and Nanoarchaea.

Accepted 24 October, 2007. *For correspondence. E-mail henri.grosjean@igmors.u-psud.fr; Tel. (+33) 1 69154637; Fax (+33) 1 69154629. †Present address: Laboratoire de Chimie Bacterienne, CNRS, Institut de Biologie Structurale et de Microbiologie, Marseille, France. ‡The first two and last two authors contributed equally to this work.

Introduction

Naturally occurring RNAs always contain a number of nucleosides that are post-transcriptionally modified by specific enzymes acting on their base and/or their 2′-hydroxyl group of the ribose backbone. To date, 107 structurally distinguishable modified nucleosides originating from different types of RNAs of diverse organisms have been reported; most of them (more than 90) were found in tRNAs [for a complete list of naturally occurring modified nucleosides and their conventional symbols, see http://library.med.utah.edu/RNAmods/ (Limbach *et al.*, 1994; McCloskey and Crain, 1998)].

One frequently encountered modified nucleoside in RNA is 5-methyluridine (m^5U, also designated rT for ribothymidine). In transfer RNAs of most organisms examined so far, m^5U is invariably found at position 54 (m^5U-54), in the so-called T-loop (loop of tRNA usually harbouring ribothymidine) (see http://www.staff.uni-bayreuth.de/~btc914/search/index.html; Edmonds *et al.*, 1991; Kowalak *et al.*, 1994; Sprinzl and Vassilenko, 2005). The enzyme catalysing the formation of m^5U-54 in tRNA was first discovered in *Escherichia coli* (Fleissner and Borek, 1962; Svensson *et al.*, 1963; Greenberg and Dudock, 1980) and initially named RUMT for RNA Uridine MethylTransferase but is now renamed TrmA for tRNA methyltransferase A. This enzyme (EC.2.1.1.35) uses S-adenosyl-L-methionine (S-AdoMet) as methyl donor. The corresponding *E. coli trmA* gene was identified years later, cloned and sequenced (Ny and Bjork, 1980; Gustafsson *et al.*, 1991). Search in fully sequenced bacterial genomes reveals that *trmA* orthologues are present in many but not all bacteria analysed to date (essentially in Beta-, Epsilon- and Gammaproteobacteria, see below). Indeed, in Alpha- and Deltaproteobacteria, Cyanobacteria and Deinococci, one finds instead a *trmFO* gene coding for an evolutionary

unrelated tRNA(uracil-54, C5)-methyltransferase, using tetrahydrofolate instead of S-AdoMet as the methyl donor (Urbonavicius *et al*., 2005; 2007).

A eukaryotic homologue of *trmA*, called *TRM2*, was identified in the genome of *Saccharomyces cerevisiae* (Nordlund *et al*., 2000). Its gene product Trm2p is responsible for the same S-AdoMet-dependent formation of m^5U-54 in both cytoplasmic and mitochondrial tRNAs (Hopper *et al*., 1982). Sequence analysis shows that the genes coding for the yeast Trm2p and the bacterial TrmA proteins belong to the same cluster of orthologous groups (COG2265, http://www.ncbi.nlm.nih.gov/COG/new; Tatusov *et al*., 2001; 2003). They are found in almost all sequenced eukaryotic genomes. It has been suggested that the ancestor of the *TRM2/trmA* genes emerged early in bacterial evolution and was subsequently transferred to eukaryotes through mitochondrial endosymbiosis (Anantharaman *et al*., 2002).

In the genome of *E. coli*, two additional homologous genes (*rumA* and *rumB*) are found beside the *trmA*. They encode two distinct site-specific S-AdoMet-dependent rRNA(uracil, C5)-methyltransferases, RumA and RumB, targeting uridine at positions 1939 and 747 of the *E. coli* 23S rRNA respectively (Agarwalla *et al*., 2002; Madsen *et al*., 2003). *RumA* homologues are found in the genomes of various bacteria and *RumB* homologues seem restricted to Gammaproteobacteria (see below). Both RumA and RumB belong to the same cluster of orthologous groups as TrmA and Trm2p (COG2265), suggesting that this large family of S-AdoMet-dependent RNA(uracil, C5)-methyltransferases probably emerged from a single common ancestor (Anantharaman *et al*., 2002). Moreover, the presence of up to three paralogues in some bacterial genomes (as *E. coli*) suggests that duplication events occurred during the evolution of the bacterial domain, some of them being probably followed by changes of target specificity.

The modified nucleoside m^5U has never been detected in bulk tRNA from Archaea analysed to date, except for the closely related members of the Thermococcales order (i.e. *Thermococcus* and *Pyrococcus* genera). In these hyperthermophilic organisms, m^5U is partially or totally hypermodified into a thiolated derivative (2-thio-5-methyluridine, s^2m^5U), depending on the temperature at which the cells are grown (Edmonds *et al*., 1991; Kowalak *et al*., 1994). In agreement with this observation, a genomic survey of the completely sequenced archaeal genomes reveals that only members of the Thermococcales and Nanoarchaea orders contain homologues of the *trmA/TRM2/rumA/rumB* family (Urbonavicius *et al*., 2007). Interestingly, all members of the Thermococcales lineage contain two homologous genes while a single homologue is present in the genome of *Nanoarchaeum equitans*, the only representative of the Nanoarchaea lineage. The amino acid sequences of these archaeal enzymes are more similar to *E. coli* RumA than to *E. coli* TrmA or *S. cerevisiae* Trm2p suggesting they might be involved in rRNA rather than tRNA modification. However, as the activity of an S-AdoMet-dependent tRNA(uracil-54, C5)-methyltransferase was clearly demonstrated *in vitro* in cell extract of *Pyrococcus furiosus* (Constantinesco *et al*., 1999), we wanted to determine whether one of the two potential open reading frames (ORFs) detected in the Thermococcales genomes may code for a site-specific tRNA(uracil-54, C5)-methyltransferase. Indeed, the possibility exists that an enzyme phylogenetically unrelated to TrmA-like enzyme may catalyse the C5-methylation of uracil-54 in tRNAs from Archaea.

To shed light on this problem, the genes from *Pyrococcus abyssi* (PAB0719 and PAB0760) were cloned, and the enzymatic activity of each of the two purified recombinant proteins was tested *in vitro* with transcripts of tRNAs and rRNA fragment as potential substrates. Here, we show that only one of the two candidates, PAB0719 protein, catalyses the S-AdoMet-dependent site-specific formation of m^5U-54 in tRNA and none of the candidates can methylate an rRNA fragment homologous to a substrate of *E. coli* RumA. Detailed phylogenetic analyses reveal that this S-AdoMet-dependent tRNA(uracil-54, C5)-methyltransferase was acquired by the common ancestor of Thermococcales and Nanoarchaea through a single horizontal gene transfer (HGT) of a *rumA*-type gene from a bacterial donor. This strongly suggests that a functional shift occurred in the archaeal lineage.

Results

Putative genes coding for RNA(uracil, C5)-methyltransferases in archaeal genomes

A genomic survey of 410 complete genome sequences available at NCBI enabled us to retrieve 456 homologues of RNA(uracil, C5)-methyltransferases (31 RumB-type, 303 RumA-type, 49 Trm2-type, 64 TrmA-type and 9 archaeal sequences). In Archaea, two homologues were detected only in members of the Thermococcales group (*P. abyssi*, *P. furiosus*, *Pyrococcus horikoshii* and *Thermococcus kodakarensis*), while only one homologue is found in the genome of *N. equitans*, the single representative of the recently proposed new archaeal phylum Nanoarchaea (Huber *et al*., 2002).

Figure 1 schematically shows the conserved motifs/domains identified in the COG2265 family of RNA(uracil, C5)-methyltransferases (parts 1–9), including the proteins from *P. abyssi* and *N. equitans* (part 9, for detailed sequence alignment, see Fig. S1). Sequence alignment of all the RumA, RumB, TrmA, Trm2p homologues

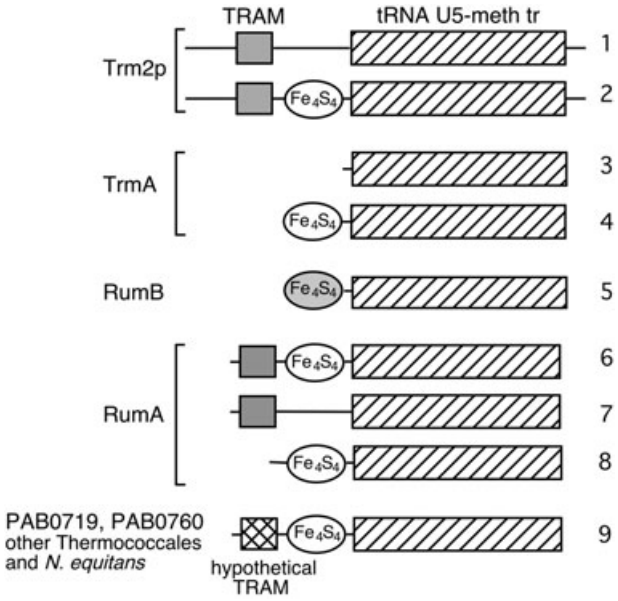

Fig. 1. Schematic view of the organization of conserved regions of the sequences of the S-AdoMet-dependent RNA(uracil, C5)-methyltransferases family. The signature sequence for the catalytic core domain of the RNA(uracil, C5)-methyltransferases family (annotated as 'tRNA U5-meth tr' in the sequence database) encompassing S-AdoMet-binding motifs and adopting the Rossmann-like fold is indicated by dashed boxes, the detailed sequence information being provided in Fig. S1. In some sequences a conserved 10- to 12-amino-acid-long motif is found at the N-terminal part. Together with a fourth cysteine, a cysteine triad of this motif probably forms a characteristic iron-sulphur cluster [4Fe-4S] (illustrated by a Fe_4S_4 encircled symbol), as demonstrated in the crystal structure of RumA (Lee *et al.*, 2004). The sequence of this motif is usually C*XX*(FYA)*XX*CGGC (parts 2, 4, 6, 8 and 9; for details see Figs S2 and S3), except for the RumB homologues for which a slightly different motif exists (C*XX*(FY)*XXXX*C(RTH)SC) (part 5, encircled grey symbol). In the other sequences, the motif is absent or degenerated in such a way that it may no longer form an iron-sulphur cluster (parts 1, 3 and 7). Finally, the presence of a 50- to 80-amino-acid-long TRAM motif as defined by Anantharaman *et al.* (2001) for the yeast Trm2p and *E. coli* MiaB domain is symbolized by a grey rectangle. The presence of a 'hypothetical TRAM' in archaeal enzymes is indicated by an open dashed rectangle (for definition of the 'hypothetical TRAM', see text). Parts 1 through 9 correspond to the following enzymes. Part 1 corresponds to the domain organization of the experimentally verified Trm2p of *S. cerevisiae* (569 amino acids) and homologues of most eukaryotes; part 2 to Trm2p homologues of all fungi (except *S. cerevisiae* and its close relative *Candida glabrata*), *Entomoeba* and land plants; part 3 to the experimentally verified TrmA of *E. coli* (366 amino acids) as well as homologues of Beta- and Gammaproteobacteria and a few Epsilonproteobacteria; part 4 to homologues of all other Epsilonproteobacteria; part 5 to all RumB-type enzymes (present only in Gammaproteobacteria, including the genuine *E. coli* RumB with 375 amino acids); part 6 to the majority of RumA-type homologues (including *E. coli* with 433 amino acids); part 7 to RumA homologues found in some bacteria such as Deinococci, certain Firmicutes and certain Actinobacteria; part 8 to RumA homologues of Chlamydia and certain Spirochetes; part 9 to all RumA-like proteins present in the archaeal Thermococcales (including PAB0719, 405 amino acids, and PAB0760, 410 amino acids) and in *N. equitans* (NEQ053, 447 amino acids).

shows a conserved C-terminal part annotated as tRNA(uracil, C5)-methyltransferase domain in the Pfam database (PF05958.1 – about 340 ± 30 amino acids, shaded boxes in Fig. 1). This corresponds to the conserved catalytic core domain encompassing the catalytic residues responsible for the RNA(uracil, C5)-methyltransferase activity and the six characteristic motifs (X, I, II, IV, VI and VIII) responsible for the binding of S-AdoMet (reviewed in Fauman *et al.*, 1999; Martin and McMillan, 2002; Schubert *et al.*, 2003; Bujnicki *et al.*, 2004). From the crystal structure of *E. coli* RumA, this conserved domain was demonstrated to adopt the typical class I S-AdoMet-dependent methyltransferase fold (also designated Rossmann-like fold; Lee *et al.*, 2004).

In contrast, the N-terminal part of the sequences is more variable. In most sequences, a 10- to 12-amino-acid sequence encompassing a cysteine triad is found (Fig. 1, parts 2, 4–6, 8 and 9). With a fourth cysteine, this motif may form a characteristic iron-sulphur cluster [4Fe-4S], as demonstrated in *E. coli* RumA (Agarwalla *et al.*, 2004; Lee *et al.*, 2004). Interestingly, in several sequences this motif is absent or degenerated in such a way that it may no longer form an iron-sulphur cluster (Fig. 1, parts 1, 3 and 7). This suggests that such iron-sulphur cluster was present in the ancestral protein, and has been lost along the evolution of the Trm2p/TrmA/RumA. Therefore, it may not be essential for the catalytic formation of ribothymidine in certain RNAs.

In addition to the iron-sulphur cluster, all the eukaryotic Trm2p and most bacterial RumA homologues contain a conserved motif of 50–80 amino acids called TRAM domain (grey box in Fig. 1) in the Pfam database (PF01938.10; Anantharaman *et al.*, 2001; Lee *et al.*, 2004). Although no TRAM domain is detected in the archaeal sequences, their N-terminal sequence displays some similarity with the genuine TRAM domain. In fact, the N-terminal part of the archaeal sequences is only four amino acids shorter than that of *E. coli* RumA (Fig. 1, part 6). Moreover this deletion seems to be located in a variable segment of TRAM as *S. cerevisiae* Trm2p has an insertion of seven amino acids in the same segment (see alignment of Fig. S1). This strongly suggests that the archaeal sequences nevertheless contain a TRAM-like domain (indicated as 'hypothetical TRAM' in Fig. 1, part 9).

In summary, comparison of the members of the S-AdoMet-dependent RNA(uracil, C5)-methyltransferase family revealed no apparent link between their target specificity and their domain organization. Consequently, one cannot unambiguously guess which one of the two archaeal sequences PAB0719 or PAB0760 may correspond to the tRNA(uracil-54, C5)-methyltransferase, whose activity was detected in *P. furiosus* cell extracts (Constantinesco *et al.*, 1999).

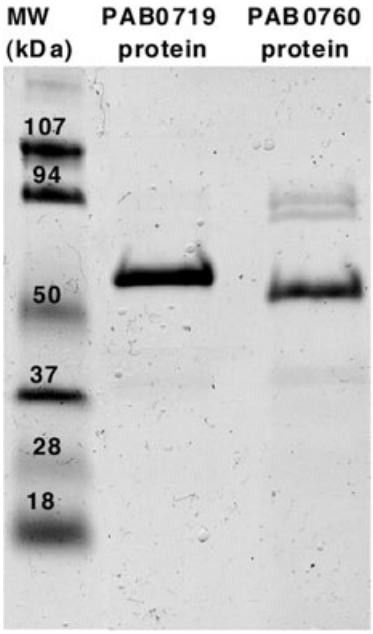

Fig. 2. Electrophoretic analysis of the purified recombinant PAB0719 and PAB0760 proteins. An SDS-PAGE was performed using 10% acrylamide gels, and the proteins were stained by Coomassie blue. Lane 1, molecular weight markers in kDa; lane 2, PAB0719 protein (49 kDa); lane 3, PAB0760 protein (48.6 kDa). Both proteins were purified by Ni-affinity chromatography.

Recombinant PAB0719 protein catalyses methylation at C5 of uridine in tRNA

The recombinant PAB0719 and PAB0760 proteins, both tagged with six histidine residues at the N-terminus, were successfully expressed in *E. coli* BL21 (DE3) CodonPlus™ and purified to near homogeneity by affinity chromatography (Fig. 2). The enzymatic activity of both purified enzymes was tested at 50°C, using [methyl-^{3}H]-AdoMet as methyl group donor and yeast tRNA^{Asp} transcript as substrate. The methyl group transfer activity was clearly shown with the recombinant PAB0719 enzyme, while under identical experimental conditions the PAB0760 protein did not show any significant incorporation of the radioactive methyl group into tRNA (Fig. 3).

Confirmation that only recombinant PAB0719 and not PAB0760 catalyses the formation of ribothymidine in tRNA and verification that the product of methyl transfer is m^{5}U were obtained using synthetic [α^{32}P]-UTP-labelled tRNA transcripts corresponding to the nucleotide sequence of a naturally occurring tRNA of *P. abyssi*. One of the two substrates used is the full-length tRNA^{Asp} (Fig. 4A), and the other one is its D-arm-truncated version, thus lacking the characteristic three-dimensional (3D)-fold generated by the T-loop/D-loop (loop of tRNA usually harbouring dihydrouridine) interactions (Fig. 4B).

[α^{32}P]-UTP-labelled $_{Pab}$tRNA^{Asp} lacking the D-arm was incubated with either of the two enzymes in the presence of S-AdoMet, and then hydrolysed into 5′-[^{32}P]-nucleosides by nuclease P1. The resulting RNA digests were then separated by two-dimensional thin-layer chromatography (2D-TLC). The results indicate that no significant methylation occurs using the recombinant PAB0760 enzyme (compare Fig. 4E with Fig. 4C), while the formation of m^{5}U is clearly visible with the PAB0719 enzyme (Fig. 4D). The activity of the PAB0719 enzyme was then examined as a function of time for the two different tRNA substrates and at two different temperatures (50°C and 80°C; Fig. 4F and G). At 50°C, the tRNA lacking the D-loop is a better substrate than full-length tRNA^{Asp} (Fig. 4F), whereas, at 80°C, the full-length tRNA becomes the better substrate (Fig. 4G). This result can probably be explained by the unfolding of the truncated tRNA substrate at 80°C due to lack of self-stabilization through the D-loop/T-loop interaction that occurs in the full-length molecule (Grosjean and Oshima, 2007). At 50°C, uracil-54 in the truncated substrate is more accessible to the enzyme than in the compact entire molecule, which probably explains why the methylation reaction occurs more easily in the D-loop-truncated substrate. The same observation has been made for 1-methyladenosine and pseudouridine formation in tRNAs at positions 57/58 and 55 respectively (Grosjean *et al.*, 1996; Becker *et al.*, 1997; Gu *et al.*, 1998; Constantinesco *et al.*, 1999; Hur *et al.*, 2006).

Same experiments at 50°C and 80°C were performed with the recombinant enzymes PAB0719 and PAB0760 but tested with [α-^{32}P]-UTP-labelled fragment of *P. abyssi* 23S rRNA (56 nucleotides in length; Fig. 5). This putative substrate sequence is homologous to a fragment (56 nucleotides) of *E. coli* 23S rRNA used previously to test the ability of *E. coli* RumA to site-specifically methylate C5

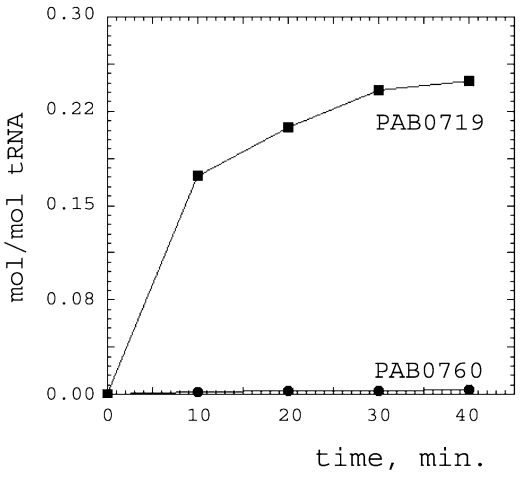

Fig. 3. tRNA methyltransferase activity of the PAB0719 and PAB0760 proteins. Kinetics at 50°C of incorporation of [^{3}H]-methyl group from radiolabelled S-AdoMet into yeast tRNA^{Asp} by PAB0719 (filled squares) and PAB0760 (filled circles) enzymes (for details see *Experimental procedures*).

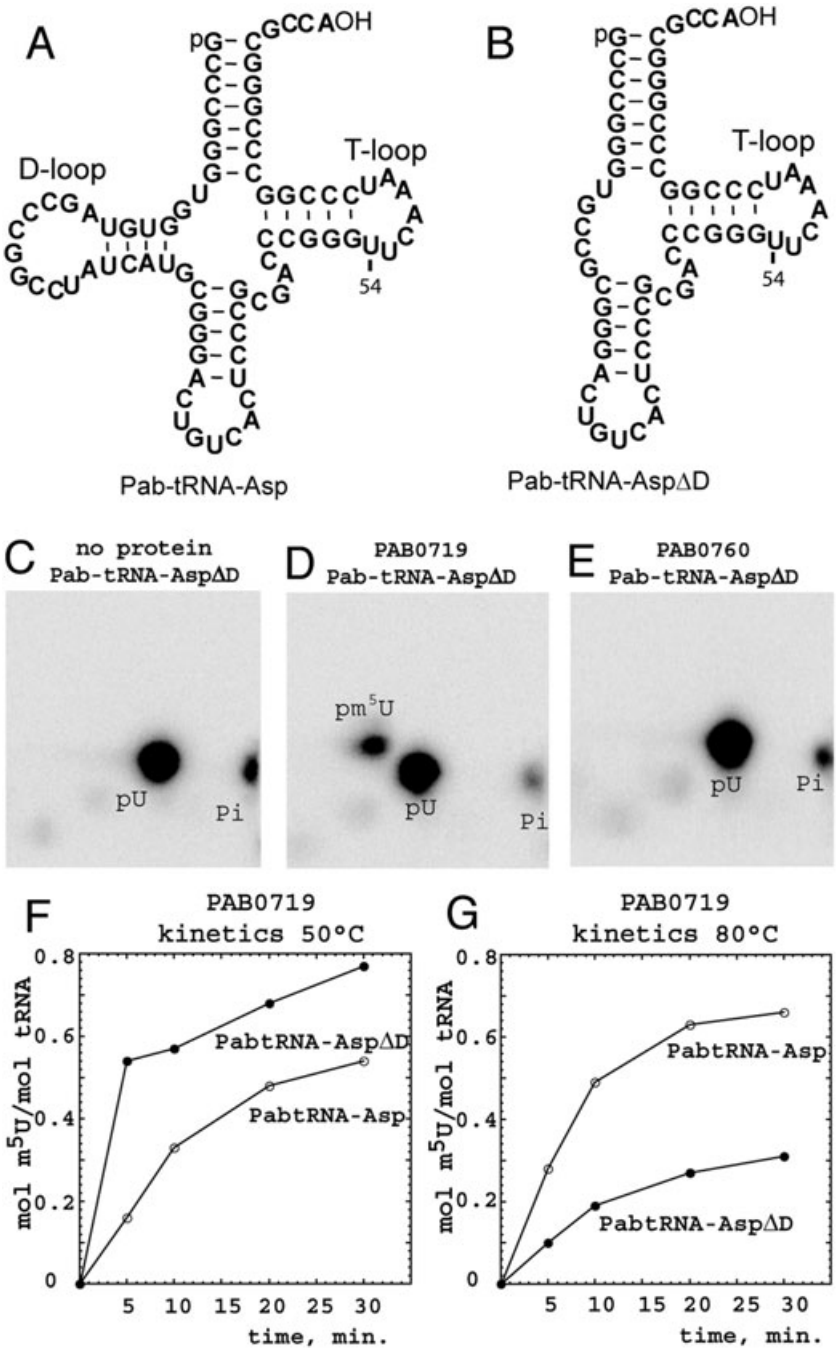

Fig. 4. Enzymatic activity of the PAB0719 and PAB0760 proteins on *P. abyssi* $tRNA^{Asp}$.
A. Sequence of the $_{Pab}tRNA^{Asp}$ substrate.
B. Sequence of the $_{Pab}tRNA^{Asp}$, lacking the D-arm (Pab-tRNA-ΔD).
C–E. Autoradiograms of the TLC plates (2D-TLC) corresponding to the Pab-tRNA-ΔD incubated for 30 min at 50°C without protein (C), with PAB0719 protein (D) or PAB0760 protein (E) and digested with nuclease P1.
F. Time-course of m^5U-54 formation at 50°C catalysed by the PAB0719 protein. The enzyme was incubated in the presence of S-AdoMet and [α-^{32}P]-UTP-labelled $_{Pab}tRNA^{Asp}$ (open circles) or $_{Pab}tRNA^{Asp}$ lacking the D-arm (filled circles).
G. Same as in (F) but time-course of m^5U-54 formation was performed at 80°C (more details are in *Experimental procedures*).

of uridine at position 1939 (indicated by an arrow in Fig. 5; Madsen *et al*., 2003). Results from 2D-TLC analysis of the nuclease P1 digestion products indicate that none of these two enzymes can methylate any uridine in this rRNA fragment (data not shown). These results favour the idea that PAB0719 enzyme is specific for tRNA and has no dual specificity. Concerning PAB0760 protein, preliminary results using other rRNA fragments indicate that the enzyme is active but the exact uridine target remains to be identified (S. Auxilien *et al*., in preparation).

The methylated uridine product of PAB0719 is m^5U at position 54 in tRNA

The position of the methylation in tRNA catalysed by PAB0719 enzyme was easily identified. Indeed, in $_{Pab}tRNA^{Asp}$, a tandem of U exists only at positions 54 and 55 (Fig. 4A and B). Therefore, complete digestion of [α-^{32}P]-UTP-labelled tRNA with RNase T2 generates 3′-[^{32}P]-nucleosides with all the [^{32}P]-UMP originating exclusively from position 54 (nearest neighbour analysis). Figure 6

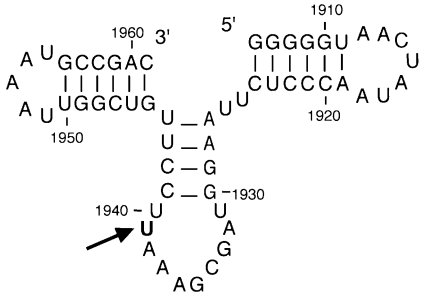

Fig. 5. *Pyrococcus abyssi* 23S ribosomal RNA fragment tested as potential substrate for the PAB0719 and PAB0760 enzymes. This rRNA fragment corresponding to domain IV sequence 1906–1961 (56-mer) of 23S ribosomal RNA is homologous to the substrate of *E. coli* RumA that has previously been used *in vitro* (Madsen *et al.*, 2003).

shows the autoradiograms of 2D-thin-layer plates resulting from the chromatographic analysis of the T2-nucleotide hydrolysates of D-loop-lacking $tRNA^{Asp}$ pre-incubated at 50°C for 30 min in the absence of enzyme (Fig. 6A, control experiment) or with the recombinant PAB0719 protein (Fig. 6B). The presence of radiolabelled 3′-[^{32}P]-m^5U is demonstrated only upon incubation with the PAB0719 enzyme (Fig. 6B), the fraction of non-methylated uridine remaining in the RNase T2 hydrolysates (indicated as Up in Fig. 6B) being less than 10%. Based on these evidences, we renamed the PAB0719 protein into the acronym name $_{Pab}$TrmU54.

Enzymatic formation of m^5U-54 in tRNA is dependent on the presence of Mg^{2+}

During the initial optimization of the experimental conditions to test the PAB0719 activity, we noticed that the yield of m^5U-54 absolutely depends on the presence of Mg^{2+}, with an optimal concentration of about 5 mM. To better illustrate the effect of Mg^{2+}, full-length or D-arm-lacking *P. abyssi* $tRNA^{Asp}$ was incubated at 50°C in the absence of Mg^{2+} and in the presence of 0.5 mM of chelating agent EDTA. After 30 min, the magnesium concentration was adjusted to 5 mM and the enzymatic activity was followed for additional 45 min. Figure 7 demonstrates that m^5U-54 can be formed only in the presence of Mg^{2+}.

The purified recombinant PAB0719 protein contains an iron-sulphur cluster

The freshly purified PAB0719 protein is yellowish and its spectrum shows a peak at 405 nm (Fig. 8, continuous line). The yellow colour and the spectrum is typical for [4Fe-4S] cluster-containing proteins such as purified recombinant *E. coli* RumA (Agarwalla *et al.*, 2004). Both properties disappear progressively upon storage of the protein for only a few hours at room temperature (Fig. 8, dotted line), showing oxidation of the [4Fe-4S] cluster of the protein.

TrmU54 gene emerged in Archaea after a single HGT from Bacteria

Maximum likelihood (ML) phylogenetic analyses of the 456 homologues composing the S-AdoMet RNA(uracil, C5)-methyltransferases family were performed with different evolutionary models which led to very similar trees. According to these phylogenies, a sample of 126 sequences (6 RumB-type, 76 RumA-type, 16 Trm2p-type, 19 TrmA-type and the 9 archaeal sequences) representative of the entire diversity of the S-AdoMet RNA(uracil, C5)-methyltransferases family was chosen for further phylogenetic analyses (Fig. 9 and Fig. S3). Bayesian and ML analyses (see *Experimental procedures*) provided very similar phylogenetic topologies (not shown). However, because of the restricted number of unambiguously aligned positions available for phylogenetic analyses (altogether 152), these phylogenies are globally poorly resolved [weak posterior probabilities (PP); and bootstrap values (BV), Fig. 9 and not shown]. Nevertheless, it is interesting to point out some peculiar well-supported groups of sequences (high PP and high BV).

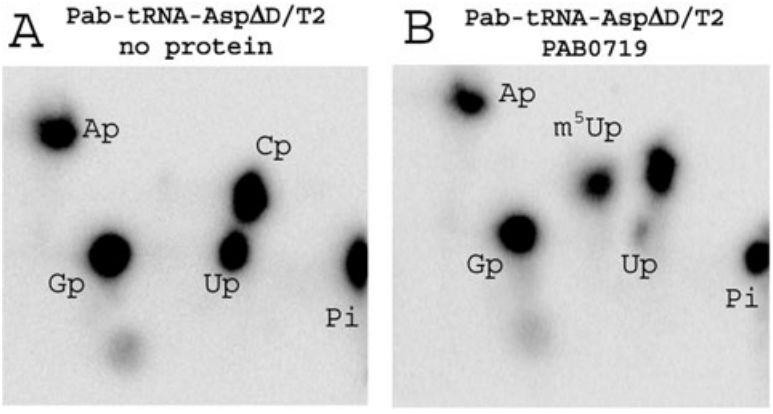

Fig. 6. Formation of m^5U-54 catalysed by the PAB0719 protein. 2D-TLC of RNase T2 digest of $_{Pab}tRNA^{Asp}\Delta D$ incubated for 30 min at 50°C without protein (control; A) or with PAB0719 protein (B). Experimental procedures are detailed in Grosjean *et al.* (2007).

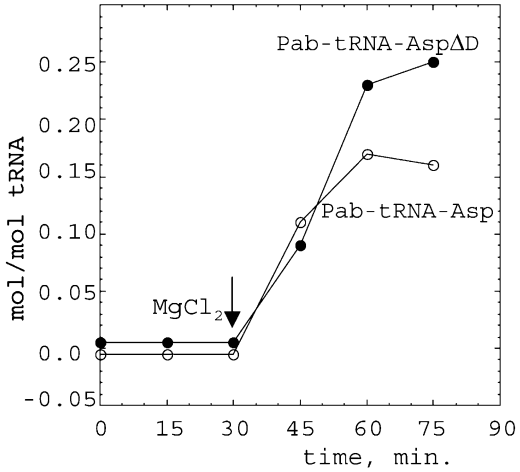

Fig. 7. Influence of Mg^{2+} on enzymatic formation of m^5U-54. PAB0719 protein was incubated with $_{Pab}tRNA^{Asp}$ or $_{Pab}tRNA^{Asp}\Delta D$ at 50°C in the standard reaction mixture, but containing 0.5 mM EDTA instead of 5 mM $MgCl_2$. For each tRNA, samples were withdrawn after 15 and 30 min of incubation, and $MgCl_2$ was added to 5 mM final concentration. Three more samples were withdrawn after an additional 15, 30 and 45 min of incubation, and analysed by 2D thin-layer chromatography after digestion with nuclease P1.

A first well-defined group of proteins, referred to as TrmA cluster (see upper part of Fig. 9, with yellow background), includes all the TrmA-type sequences (PP = 1.0 and BV = 91%). These are found in some Epsilon-, Beta- and Gammaproteobacteria, including that of *E. coli*. A second group of sequences, referred as RumB cluster (lower part of Fig. 1, with pink background, PP = 0.92 and BV = 87%), encompasses all the RumB-type sequences. These are present only in a few Gammaproteobacteria, mainly Enterobacteriales and Pasteurellales, including *E. coli* RumB. In contrast with these two monophyletic groups restricted to Proteobacteria (and mainly Gammaproteobacteria), the RumA-type sequences are widespread among Bacteria (Fig. 1, white background). These are present in almost all of bacterial phyla, such as Acidobacteria, Actinobacteria, Aquificales, Chlorobi/Bacteroidetes, Planctomycetales/Verrucomicrobiales/Chlamydia, Chloroflexi, Cyanobacteria, Deinococci/Thermus, Firmicutes, Fusobacteria and Proteobacteria except in the Epsilonproteobacterial subgroup (Fig. 1; Figs S2 and S3). This strongly supports the possibility that the ancestor of all living Bacteria had a RumA-type sequence that was mainly vertically transmitted during bacterial evolution and that the TrmA-type and RumB-type sequences originated secondarily by gene duplication in the Proteobacterial lineage from RumA-type ancestors.

As far as the archaeal sequences are concerned, they are detected only in the Thermococcales lineage and in *N. equitans*. Interestingly, these sequences form a well-defined monophyletic group (Fig. 9, with dark grey background; PP = 1.0 and BV = 78%) that clearly emerges from the RumA cluster. This strongly suggests that the ancestor of Thermococcales and *N. equitans* acquired a RumA-type sequence by a single HGT from a bacterial donor. Unfortunately, given the poor resolution of our phylogenetic tree, the bacterial donor cannot be identified with confidence. Moreover, our tree shows that a duplication event occurred in the Thermococcales lineage after its divergence from *N. equitans*, leading to two groups of paralogues P1 and P2 (see position of white circle at the node in Fig. 9). Group P1 contains the *P. abyssi* sequence (PAB0719) involved in the formation of m^5U-54 in tRNA (this work), and group P2 contains the PAB0760 sequence most probably involved in rRNA base modification (S. Auxilien *et al.*, in preparation).

The case of the Trm2p family of enzymes is surprising. Indeed, because of the possible mitochondrial origin of the Trm2-type sequences (Anantharaman *et al.*, 2002), together with the fact that contemporary *S. cerevisiae* Trm2p and *E. coli* TrmA enzymes have identical catalytic functions (they both catalyse the site-specific formation of m^5U-54 in tRNA), we expected that the Trm2p-type sequences would be closely related to the TrmA-type sequences and form a monophyletic group including sequences from Alphaproteobacteria. At variance with these expectations, our phylogenetic analyses show that: (i) genomes of Alphaproteobacteria do not contain TrmA-type sequences, (ii) Trm2p-type sequences are not closely related to the TrmA cluster, and (iii) Trm2p-type sequences do not form a monophyletic group but, on the contrary, are mixed with RumA-type sequences (Fig. 9, blue background). Thus, together with the fact that the domain organization of the Trm2p-type sequences is more similar to that of RumA-type enzymes than to the

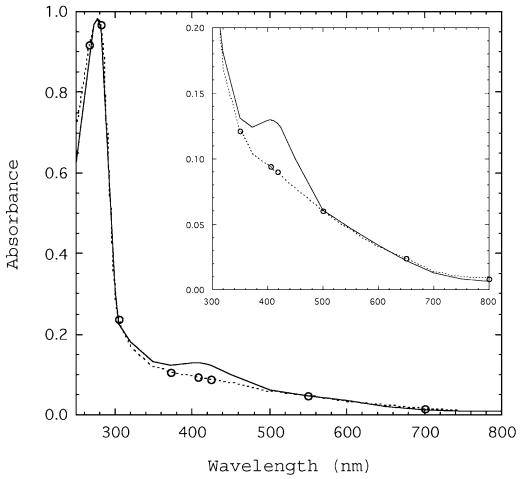

Fig. 8. UV-visible spectra of 20 μM freshly purified PAB0719 protein (continuous line) or PAB0719 protein stored for 1 day at 18°C (dotted line).

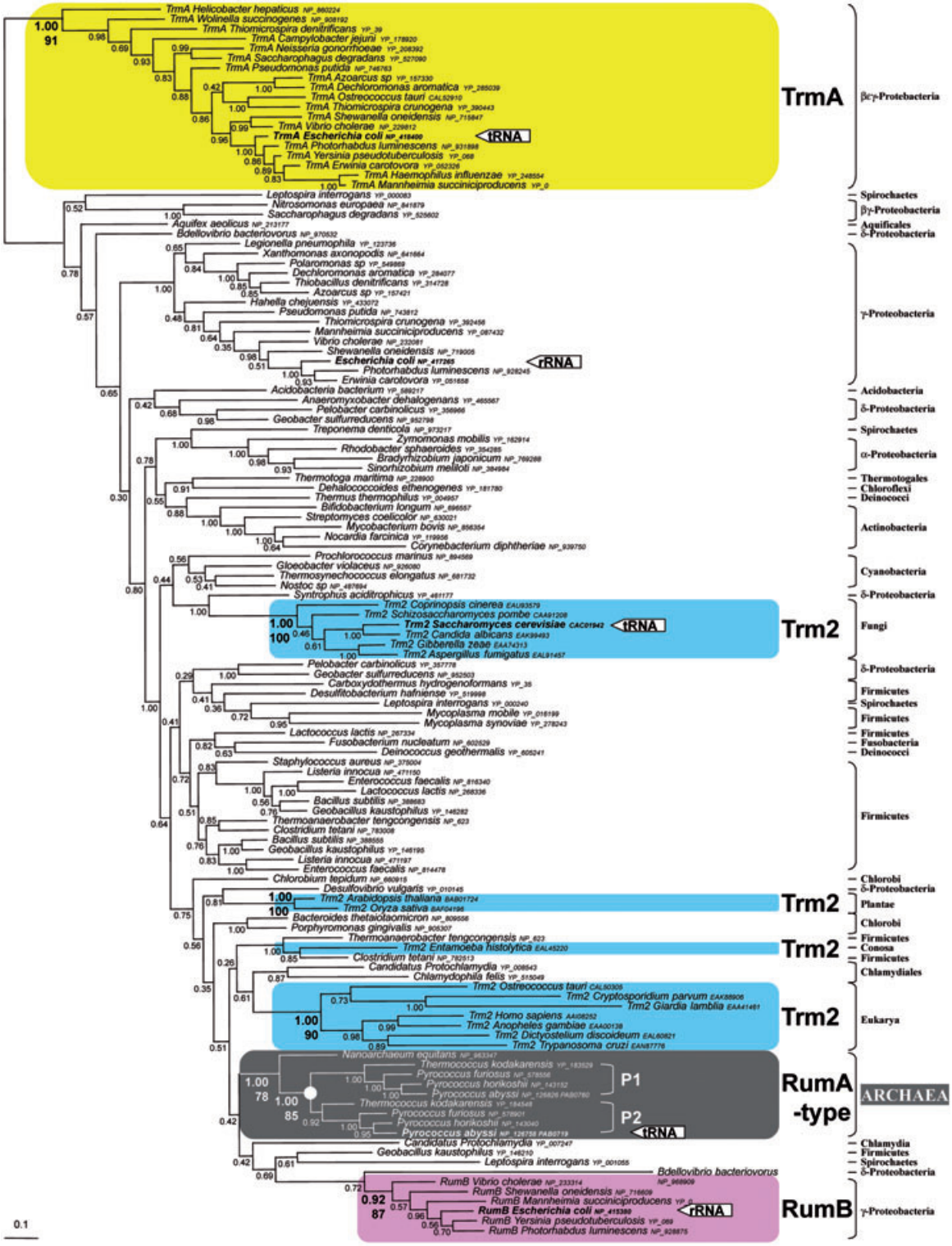

Fig. 9. Bayesian unrooted phylogenetic tree of 126 sequences representative of the diversity of the S-AdoMet-dependent RNA(uracil, C5)-methyltransferases. A total of 456 sequences were computed by MrBayes using mixed models and a Γ-correction to take into account the differences between evolutionary rates (see *Experimental procedures*). Numbers below or equal to 1.0 at nodes represent posterior probabilities (PP). For important clusters, bootstrap values (BV) are also provided. The scale bar represents the average number of substitutions per position. Enzymes with known functions are indicated by white arrows; all other enzymes should be considered as putative. The small white circle in the dark blue box indicates a duplication event.

TrmA-type sequences (Fig. 1), we propose that the eukaryotic *TRM2* genes derived from bacterial *rumA*-type sequences (and not from TrmA-type sequences) and were acquired through HGT.

Discussion

The PAB0719 (not PAB0760) protein is the genuine tRNA(uracil-54, C5)-methyltransferase in P. abyssi

In Archaea, m^5U has been experimentally detected in bulk tRNAs isolated from only the hyperthermophiles *P. furiosus* and *Thermococcus* sp. (Edmonds *et al*., 1991; Kowalak *et al*., 1994). Contrary to the situation in *E. coli*, nothing is known about the presence of m^5U in archaeal 23S rRNA. An enzymatic activity corresponding to a S-AdoMet-dependent tRNA(uracil-54, C5)-methyltransferase was previously detected in crude extracts of *P. furiosus* (Constantinesco *et al*., 1999). Using sequences comparison coupled with *in vitro* enzymatic verification, the gene encoding the tRNA(uracil-54, C5)-methyltransferase was identified in *P. abyssi*. In this genome, two homologous genes belong to the same cluster of orthologous groups (COG2265) including the genuine *E. coli trmA* and *S. cerevisiae TRM2* (involved in m^5U formation at position 54 in tRNA), as well as *E. coli rumA* and *rumB* (involved in m^5U formation at positions 1939 and 747 in 23S rRNA). Here we show that only the product of the PAB0719 gene (and not that of the PAB0760 gene) possesses the activity of a S-AdoMet-dependent tRNA(uracil-54, C5)-methyltransferase *in vitro*. We suggest to rename this enzyme $_{Pab}$TrmU54.

This newly identified archaeal enzyme is site-specific for methylating C5 in uridine 54 of the T-loop and works on D-loop-truncated tRNA. As bacterial TrmA and eukaryotic Trm2p, $_{Pab}$TrmU54 is not dependent on the L-shaped 3D conformation of the tRNA molecule (Gu *et al*., 1996; Becker *et al*., 1997; 1998). Moreover, the activity of this archaeal enzyme is strictly dependent on the presence of magnesium ions. Similarly, the *E. coli* TrmA activity was shown to be stimulated in the presence of magnesium ions, but no clear-cut dependence was demonstrated (Ny and Bjork, 1980).

Also, like most fungal Trm2p proteins, TrmA of a few Epsilonproteobacteria, all RumB and a majority of RumA enzymes, the *P. abyssi* TrmU54 protein possesses a [4Fe-4S] cluster (see Fig. 1). Recent structural studies of non-oxidized *E. coli* RumA complexed with an RNA substrate showed that the [4Fe-4S] cluster is involved in RNA binding (Lee *et al*., 2005). The iron-sulphur cluster may also be important for the regulation of the enzyme activity under oxidative stress (Agarwalla *et al*., 2004), a property that could be important in the case of *P. abyssi* (possibly all Thermococcales) that normally grow under strictly anaerobic conditions with an optimal temperature of 100°C.

tRNA(uracil-54, C5)-methyltransferase in Thermococcales and Nanoarchaea derived from a RumA-type coding gene acquired by a single HGT from a bacterial donor

The phylogenetic analyses of the 456 sequences (including nine sequences from Archaea) composing the RNA(uracil-54, C5)-methyltransferases family suggest that a common ancestor of the Thermococcales and *N. equitans* acquired a RumA-type coding gene from a non-identified bacterial donor by a single HGT event. This possible relationship between the archaeal sequences and the RumA-type sequences is strengthened by similar domain organization (Fig. 1) and high sequence similarity (Fig. S1). Taking into account that we show that PAB0719 catalyses the formation of m^5U-54 in tRNA, and that PAB0760, which does not work on tRNA, is probably a genuine rRNA(uracil, C5)-methyltransferase (S. Auxilien *et al*., in preparation), two scenarios can be proposed for the emergence of distinct RNA site-specific archaeal RumA-like methyltransferases (Fig. 10). First, if the transferred ancestral bacterial gene coded for an enzyme acting on tRNA rather than on rRNA, then a functional shift from a tRNA to an rRNA modification activity should have occurred in the P1 lineage (as indicated by the circled symbol 'F' in Fig. 10A). In this case, the unique *N. equitans* enzyme, as well as the paralogues P2 of Thermococcales, has probably conserved the ancestral function (i.e. m^5U-54 formation in tRNA, Fig. 10A). On the contrary (alternative scenario), if the ancestral function of the transferred bacterial gene was the modification of rRNA (as the present-day *E. coli* RumA), then the functional shift from an rRNA to a tRNA modification activity should have occurred in the P2 lineage (as indicated by the square boxed symbol 'F' in Fig. 10B). In this case, there should be a good chance that the unique present-day RumA-like enzyme found in *N. equitans* as well as the paralogues P1 of Thermococcales has kept the same specificity for rRNA as the genuine original bacterial RumA. However, the possibility also exists that the ancestral RumA-type enzyme had originally a more relaxed specificity than the present-day *E. coli* RumA having for example a 'dual specificity' for both rRNA and tRNA. In this case, functional specializations for tRNA or rRNA might have occurred after the duplication event in the Thermococcales lineage. Finally to complete the evolutionary scenario, it remains to determine the function of the present-day RumA-like of *N. equitans*.

Evolutionary implications of the presence of a bacterial-type RNA methyltransferase in N. equitans

Based on phylogenies of the RNA of the small subunit of the ribosome (16S rRNA), it was initially proposed that

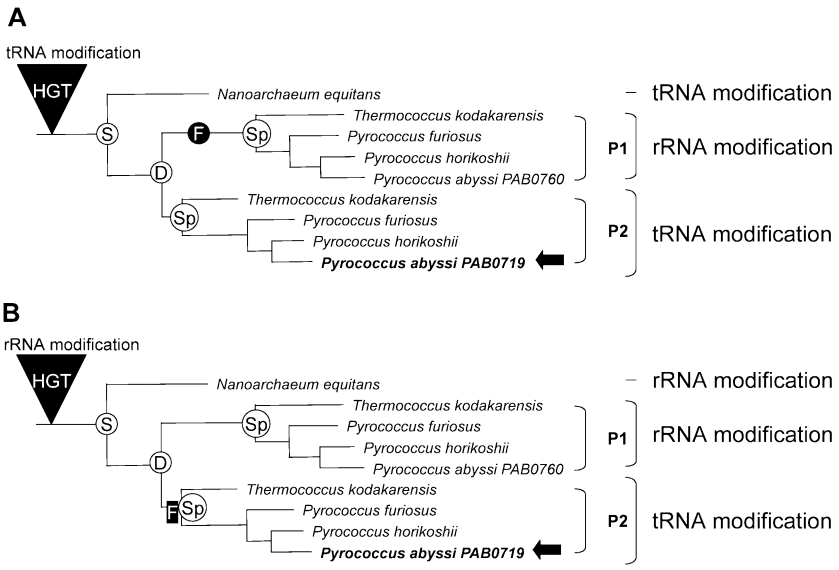

Fig. 10. Probable scenario for the functional evolution of the archaeal genes coding for S-AdoMet-dependent RNA(uracil, C5)-methyltransferases. The black triangles with HGT indicate when a bacterial RumA-type enzyme was acquired by HGT. The encircled symbol 'S' is used for the gene speciation of the Thermococcales and *N. equitans* lineages; the encircled symbol 'D' for a duplication event; the encircled symbol 'Sp' for the diversification of the Thermococcales lineage. The symbol 'F' indicates the functional shift from the m^5U-54 formation in tRNAs to m^5U formation in rRNAs (black circle) or from the m^5U formation in rRNA to m^5U-54 formation in tRNAs (black rectangle).

N. equitans was possibly the representative of a third archaeal phylum (Nanoarchaeota) that branches deeply in the archaeal phylogeny [i.e. before the split between the Euryarachaea and the Crenarchaea phyla (Huber *et al*., 2002; Waters *et al*., 2003)]. However, later phylogenetic and genomic analyses suggested that *N. equitans* is rather a fast evolving Euryarchaea possibly related to Thermococcales (Brochier *et al*., 2005; Makarova and Koonin, 2005). If *N. equitans* corresponds to a deeply branching archaeal phylum, as initially suggested by Huber *et al*. (2002), this implies that the RumA-like enzyme was acquired by an ancestor common to all living Archaea. To account for the absence of RumA-like enzymes in all archaeal lineages (except for Thermococcales and *N. equitans*), massive independent gene losses should have occurred in the archaeal domain. On the contrary, if *N. equitans* is related to Thermococcales only, this implies that a single HGT occurred in the ancestor of these two lineages, and no multiple gene losses need to be involved. The presence of genes coding for enzymes catalysing the formation of the m^5U modification in the Thermococcales and *N. equitans* genomes is thus additional and strong evidence supporting the second possibility, i.e. the close relationship between the two lineages Thermococcales and Nanoarchaeota.

Conclusion

In conclusion, the widespread presence of a given modified nucleoside in RNA of the organisms of the three domains of life is generally taken as evidence that the family of corresponding homologous (orthologous/paralogous) modification enzymes has a very ancient common evolutionary origin, before the separation of the three domains (Cermakian and Cedergren, 1998). This assumption must be taken with caution. Indeed, we show that the S-AdoMet-dependent m^5U formation in RNA probably appeared first in the bacterial domain and was transferred by HGT only later to the other domains. Based on this functional information, it is not possible to predict the type of specificity of the primordial enzyme for tRNA, rRNA or both. An interesting hypothesis is that the primordial enzyme could have displayed a 'relaxed/dual' specificity. Only during evolution of a few bacterial groups (i.e. Proteobacteria), distinct activities for site-specific RNA(uracil, C5)-methyltransferases could have appeared after duplication or HGT events, followed by subfunctionalization of genes. This would give rise to the present bacterial *trmA*- and *rumB*-type gene subfamilies in addition to the existing *rumA*-type genes, each coding for a site-specific enzyme catalysing the formation of m^5U at different locations in tRNA and rRNA. Concerning the ancestors of the eukaryotic *TRM2* genes and of the archaeal genes (two paralogues in Thermococcales and unique gene in *N. equitans*), our results clearly show that they all were acquired from ancestor *rumA*-type genes via HGT from unidentified bacterial donors.

Experimental procedures

Construction of N-terminal His_6-tagged PAB0719- and PAB0760-overexpressing plasmids

The PAB0760 and PAB0719 ORFs were amplified by PCR from *P. abyssi* genomic DNA, using the Vent DNA polymerase (Biolabs) and the following primers (sequence in small characters corresponds to genome sequence): AAAACATatgaggggaatcataaagggag (PAB0760fw) and AAAAGAATtcaccttctcctgagcttaacg (PAB0760rev) or AAAACATatgaggggagtcataagaaagc (PAB0719fw) and AAAAGAATTCagactagcttcgcaacgagctc (PAB0719rev). The

'forward' (fw) and 'reverse' (rev) primers created an NdeI site at the ATG start codon and an EcoRI site overlapping the TGA stop codon respectively. The resulting PCR products were digested with NdeI and EcoRI restriction enzymes and cloned into the pET28b vector (Novagen), cleaved with the same enzymes. The resulting vectors were used to transform the *E. coli* BL21 (DE3) Codon-Plus™ (Novagen) expression cells.

Purification of the recombinant proteins

Expression cells were grown at 37°C in 1 l of Luria–Bertani medium (Invitrogen) containing 50 μg ml^{-1} kanamycin and 30 μg ml^{-1} chloramphenicol until OD_{600} = 0.8. After induction of the PAB0760 or PAB0719 proteins expression by isopropyl β-D-thiogalactopyranoside (IPTG, Eurobio) at final concentration of 1 mM, the cultures were further grown at 20°C for 3 h. After harvesting the cells by centrifugation, the pellets were re-suspended in 5 volumes of lysis buffer containing 50 mM sodium phosphate pH 8.0, 1 M NaCl, 5 mM β-mercaptoethanol, 10 mM imidazole and 1% (v/v) protease inhibitor cocktail for the purification of histidine-tagged proteins (Sigma). Cells were broken by ultrasonication and centrifuged for 30 min at 20 000 *g* at 4°C. Supernatants were loaded onto an Ni-NTA superflow resin (Qiagen) and washed with lysis buffer containing 300 mM NaCl and 20 mM imidazole. PAB0719 or PAB0760 proteins were eluted with lysis buffer containing 300 mM NaCl and 250 mM imidazole. Fractions containing the PAB0760 or PAB0719 proteins (431 and 425 amino acids in length respectively) were pooled. The absorption spectrum of the purified PAB0719 protein was immediately measured. Proteins were diluted in lysis buffer containing 300 mM NaCl and 250 mM imidazole and 50% (v/v) glycerol, aliquoted and stored at –20°C.

Preparation of the RNA substrates

Non-labelled and [α-^{32}P]-UTP (GE Healthcare) labelled full-length or D-arm-lacking $tRNA^{Asp}$ transcripts from *P. abyssi* or yeast were prepared by *in vitro* transcription with T7 RNA polymerase (Promega) of linearized plasmids harbouring the corresponding synthetic genes while their purification by urea gels electrophoresis were performed as described elsewhere (Grosjean *et al*., 2007). The rRNA fragment radiolabelled with [α-^{32}P]-UTP was T7-transcribed *in vitro* from single-stranded DNA template corresponding to the minus sequence terminated on 3′ by T7 promoter and containing a 2′-*O*-methyl guanosine in position 2 of 5′ end (Kao *et al*., 1999).

RNA methylation assays

To determine the potentiality of the purified recombinant PAB0760 and PAB0719 proteins to catalyse the formation of m^5U in tRNA, two methods were used. In the first one, transfer of [^{3}H]-radiolabelled methyl group from [^{3}H]S-AdoMet (15 Ci $mmole^{-1}$, GE Healthcare) into non-radioactive tRNA was tested. Enzyme kinetics experiments were performed at 50°C in 25 mM sodium phosphate buffer pH 7.2, 25 mM KCl, 5 mM $MgCl_2$, 2 mM dithiotreitol, 5% glycerol, with 2.6 μM [^{3}H]S-AdoMet, 0.9 μM yeast $tRNA^{Asp}$ transcript and 0.9 μM of PAB0760 or PAB0719 protein in a final volume of 50 μl. After incubation, the reactions were stopped by addition of cold 5% (w/v) trichloracetic acid, and the precipitates were collected by vacuum filtration through GF/C filters (Whatman). The filters were washed twice with cold 5% (w/v) trichloracetic acid, dried, and the radioactivity was measured by liquid scintillation counting.

In the second method, non-radiolabelled S-AdoMet and [α-^{32}P]-UTP-labelled RNA transcripts were used to determine the RNA(uracil, C5)-methyltransferases activity of purified recombinant proteins. Three types of substrates were used: transcripts of a full-length *P. abyssi* $tRNA^{Asp}$ with the T-loop presumably interacting with the D-loop in solution, a D-arm-truncated version of the same $tRNA^{Asp}$, thus unable to generate the classical L-shaped 3D conformation, or a fragment of 23S rRNA (nucleotides 1906–1960) from *P. abyssi*. To test the enzymatic activity, about 1 μg of PAB0760 or PAB0719 proteins (about 20 pmol) were incubated in 25 μl of 25 mM sodium phosphate buffer pH 7.2 containing 25 mM KCl, 5 mM $MgCl_2$ (or 0.5 mM EDTA, when the importance of Mg^{2+} for the enzymatic activity was examined), 2 mM dithiotreitol, 5% glycerol, 0.1 mg ml^{-1} bovine serum albumin, 80 μM S-AdoMet (Sigma) and about 50 fmol of [^{32}P]-radiolabelled RNA at 50°C or 80°C. After incubation, the reactions were stopped by addition of phenol, the RNA was recovered by ethanol precipitation, and the molar content of ribothymidine monophosphate was evaluated after complete digestion of the recovered RNA by nuclease P1 or RNase T2 and subsequent analysis by 2D-TLC as described in detail elsewhere (Grosjean *et al*., 2007).

Bioinformatic analyses

We retrieved all the PAB0719 and PAB0760 homologous sequences from 410 completely sequenced archaeal, bacterial and eukaryotic genomes available at the NCBI (http://www.ncbi.nlm.nih.gov/). Sequences were aligned by CLUSTALW (Thompson *et al*., 1994) and MUSCLE (Edgar, 2004). The resulting alignments were manually refined using '*ed*' from the MUST package (Philippe, 1993). Regions where homology was doubtful were manually removed from further analyses. ML phylogenetic trees were inferred using PHYML (Guindon and Gascuel, 2003) with various models (JTT, WAG and VT) and a Γ-correction to take into account the heterogeneity of the evolutionary rates between sites (four discrete categories, an estimated alpha parameter and an estimated proportion of invariant sites). The robustness of each node was estimated by bootstrap analyses (1000 replicates of the original data set) using PHYML (Guindon and Gascuel, 2003). Bayesian trees were inferred using MrBayes v.3.0b4 (Ronquist and Huelsenbeck, 2003) with a mixed model of amino acid substitution and a Γ-correction (four discrete categories and a proportion of invariant sites). MrBayes was run with four chains for one million generations and trees were sampled every 100 generations. To construct the consensus tree, the first 1500 trees were discarded as 'burnin'. To identify known functional domains, we performed searches in the Pfam database (http://www.sanger.ac.uk/Software/Pfam/). Alignments against the Pfam profiles having *E*-values lower than 0.1 were considered significant.

Acknowledgements

We thank Richard Giegé from University of Strasbourg, France for a plasmid carrying the yeast $tRNA^{Asp}$ gene and Clotilde Husson and Dominique Fourmy from CNRS-ICSN (Gif-sur-Yvette, France) for the synthetic *P. abyssi* rRNA fragment. Kalliopi Trachana, presently working at EMBL/Heidelberg, Germany was an Erasmus student from the University of Thessaly (Larissa, Greece). We thank Stephen Douthwaite from the University of Southern Denmark (Odense, Denmark) for useful advice and critical reading of the manuscript and Anne-Lise Haenni (University Paris-5, Institut J. Monod, France) for considerable improvements of the text. J.U. acknowledges the financial support from FEBS (long-term postdoctoral fellowship, 2003-06). H.W. is supported by a fellowship from the Association pour la Recherche sur le Cancer. H.G. benefits from a research grant from the CNRS (GEOMEX programme).

References

Agarwalla, S., Kealey, J.T., Santi, D.V., and Stroud, R.M. (2002) Characterization of the 23S ribosomal RNA m^5U1939 methyltransferase from *Escherichia coli*. *J Biol Chem* **277:** 8835–8840.

Agarwalla, S., Stroud, R.M., and Gaffney, B.J. (2004) Redox reactions of the iron-sulfur cluster in a ribosomal RNA methyltransferase, RumA: optical and EPR studies. *J Biol Chem* **279:** 34123–34129.

Anantharaman, V., Koonin, E.V., and Aravind, L. (2001) TRAM, a predicted RNA-binding domain, common to tRNA uracil methylation and adenine thiolation enzymes. *FEMS Microbiol Lett* **197:** 215–221.

Anantharaman, V., Koonin, E.V., and Aravind, L. (2002) Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res* **30:** 1427–1464.

Becker, H.F., Motorin, Y., Sissler, M., Florentz, C., and Grosjean, H. (1997) Major identity determinants for enzymatic formation of ribothymidine and pseudouridine in the T psi-loop of yeast tRNAs. *J Mol Biol* **274:** 505–518.

Becker, H.F., Motorin, Y., Florentz, C., Giege, R., and Grosjean, H. (1998) Pseudouridine and ribothymidine formation in the tRNA-like domain of turnip yellow mosaic virus RNA. *Nucleic Acids Res* **26:** 3991–3997.

Brochier, C., Gribaldo, S., Zivanovic, Y., Confalonieri, F., and Forterre, P. (2005) Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? *Genome Biol* **6:** R42.

Bujnicki, J.M., Droogmans, L., Grosjean, H., Purushothaman, S.K., and Lapeyre, B. (2004) Bioinformatics-guided identification and experimental characterization of novel RNA methyltransferases. In *Nucleic Acids and Molecular Biology*, Vol. 15. Bujnicki, J.M. (ed.). Heidelberg: Springer-Verlag, pp. 139–168.

Cermakian, N., and Cedergren, R. (1998) Modified nucleosides always were: an evolutionary model. In *Modification and Editing of RNA*. Grosjean, H., and Benne, R. (eds). Washington, DC: American Society for Microbiology Press, pp. 535–541.

Constantinesco, F., Motorin, Y., and Grosjean, H. (1999) Transfer RNA modification enzymes from *Pyrococcus furiosus*: detection of the enzymatic activities *in vitro*. *Nucleic Acids Res* **27:** 1308–1315.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792–1797.

Edmonds, C.G., Crain, P.F., Gupta, R., Hashizume, T., Hocart, C.H., Kowalak, J.A., *et al.* (1991) Posttranscriptional modification of tRNA in thermophilic archaea (Archaebacteria). *J Bacteriol* **173:** 3138–3148.

Fauman, E.B., Blumenthal, R.M., and Cheng, X. (1999) Structure and evolution of AdoMet-dependent methyltransferases. In *S-Adenosylmethionine-Dependent Methyltransferases: Structures and Functions*. Cheng, X., and Blumenthal, R.M. (eds). Singapore: World Scientific, pp. 1–38.

Fleissner, E., and Borek, E. (1962) A new enzyme of RNA synthesis: RNA methylase. *Proc Natl Acad Sci USA* **48:** 1199–1203.

Greenberg, R., and Dudock, B. (1980) Isolation and chracterization of m^5U-methyltransferase from *Escherichia coli*. *J Biol Chem* **255:** 8296–8302.

Grosjean, H., and Oshima, T. (2007) How nucleic acids cope with high temperature. In *Physiology and Biochemistry of Extremophiles*. Gerday, C., and Glansdorff, N. (eds). Washington, DC: American Society for Microbiology Press.

Grosjean, H., Edqvist, J., Straby, K.B., and Giege, R. (1996) Enzymatic formation of modified nucleosides in tRNA: dependence on tRNA architecture. *J Mol Biol* **255:** 67–85.

Grosjean, H., Droogmans, L., Roovers, M., and Keith, G. (2007) Detection of enzymatic activity of transfer RNA modification enzymes using radiolabeled tRNA substrates. *Methods Enzymol* **425:** 55–101.

Gu, X., Ivanetich, K.M., and Santi, D.V. (1996) Recognition of the T-arm of tRNA by tRNA (m^5U54)-methyltransferase is not sequence specific. *Biochemistry* **35:** 11652–11659.

Gu, X., Yu, M., Ivanetich, K.M., and Santi, D.V. (1998) Molecular recognition of tRNA by tRNA pseudouridine 55 synthase. *Biochemistry* **37:** 339–343.

Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52:** 696–704.

Gustafsson, C., Lindstrom, P.H., Hagervall, T.G., Esberg, K.B., and Bjork, G.R. (1991) The *trmA* promoter has regulatory features and sequence elements in common with the rRNA P1 promoter family of *Escherichia coli*. *J Bacteriol* **173:** 1757–1764.

Hopper, A.K., Furukawa, A.H., Pham, H.D., and Martin, N.C. (1982) Defects in modification of cytoplasmic and mitochondrial transfer RNAs are caused by single nuclear mutations. *Cell* **28:** 543–550.

Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., and Stetter, K.O. (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417:** 63–67.

Hur, S., Stroud, R.M., and Finer-Moore, J. (2006) Substrate recognition by RNA 5-methyluridine methyltransferases and pseudouridine synthases: a structural perspective. *J Biol Chem* **281:** 38969–38973.

Kao, C., Zheng, M., and Rudisser, S. (1999) A simple and

efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5:** 1268–1272.
Kowalak, J.A., Dalluge, J.J., McCloskey, J.A., and Stetter, K.O. (1994) The role of posttranscriptional modification in stabilization of transfer RNA from hyperthermophiles. *Biochemistry* **33:** 7869–7876.
Lee, T.T., Agarwalla, S., and Stroud, R.M. (2004) Crystal structure of RumA, an iron-sulfur cluster containing *E. coli* ribosomal RNA 5-methyluridine methyltransferase. *Structure* **12:** 397–407.
Lee, T.T., Agarwalla, S., and Stroud, R.M. (2005) A unique RNA Fold in the RumA–RNA-cofactor ternary complex contributes to substrate selectivity and enzymatic function. *Cell* **120:** 599–611.
Limbach, P.A., Crain, P.F., and McCloskey, J.A. (1994) Summary: the modified nucleosides of RNA. *Nucleic Acids Res* **22:** 2183–2196.
McCloskey, J.A., and Crain, P.F. (1998) The RNA modification database – 1998. *Nucleic Acids Res* **26:** 196–197.
Madsen, C.T., Mengel-Jorgensen, J., Kirpekar, F., and Douthwaite, S. (2003) Identifying the methyltransferases for m(5)U747 and m(5)U1939 in 23S rRNA using MALDI mass spectrometry. *Nucleic Acids Res* **31:** 4738–4746.
Makarova, K.S., and Koonin, E.V. (2005) Evolutionary and functional genomics of the Archaea. *Curr Opin Microbiol* **8:** 586–594.
Martin, J.L., and McMillan, F.M. (2002) SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr Opin Struct Biol* **12:** 783–793.
Nordlund, M.E., Johansson, J.O., von Pawel-Rammingen, U., and Bystrom, A.S. (2000) Identification of the TRM2 gene encoding the tRNA(m^5U54)methyltransferase of *Saccharomyces cerevisiae*. *RNA* **6:** 844–860.
Ny, T., and Bjork, G.R. (1980) Cloning and restriction mapping of the *trmA* gene coding for transfer ribonucleic acid (5-methyluridine)-methyltransferase in *Escherichia coli* K-12. *J Bacteriol* **142:** 371–379.
Philippe, H. (1993) MUST, a computer package of Management Utilities for Sequences and Trees. *Nucleic Acids Res* **21:** 5264–5272.
Ronquist, F., and Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572–1574.
Schubert, H.L., Blumenthal, R.M., and Cheng, X. (2003) Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem Sci* **28:** 329–335.
Sprinzl, M., and Vassilenko, K.S. (2005) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* **33:** D139–D140.
Svensson, I., Boman, H.G., Eriksson, K.G., and Kjellin, K. (1963) Studies on microbial Rna. I. Transfer of methyl groups from methionine to soluble Rna from *Escherichia coli*. *J Mol Biol* **7:** 254–271.
Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., *et al.* (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29:** 22–28.
Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., *et al.* (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4:** 41.
Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22:** 4673–4680.
Urbonavicius, J., Skouloubris, S., Myllykallio, H., and Grosjean, H. (2005) Identification of a novel gene encoding a flavin-dependent tRNA:m5U methyltransferase in bacteria:evolutionary implications. *Nucleic Acids Res* **33:** 3955–3964.
Urbonavicius, J., Brochier-Armanet, C., Skouloubris, S., Myllykallio, H., and Grosjean, H. (2007) *In vitro* detection of the enzymatic activity of folate-dependent tRNA (Uracil-54,-C5)-methyltransferase: evolutionary implications. *Methods Enzymol* **425:** 103–119.
Waters, E., Hohn, M.J., Ahel, I., Graham, D.E., Adams, M.D., Barnstead, M., *et al.* (2003) The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc Natl Acad Sci USA* **100:** 12984–12988.

Supplementary material

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2958.2007.06047.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.