Biosynthesis of Wyosine Derivatives in tRNA: An Ancient and Highly Diverse Pathway in Archaea

Valérie de Crécy-Lagard,1,8 Céline Brochier-Armanet,2,3 Jaunius Urbonavičius,1,4 Bernard Fernandez,5 Gabriela Phillips,1 Benjamin Lyons,1 Akiko Noma,6 Sophie Alvarez,7 Louis Droogmans,4 Jean Armengaud,5 and Henri Grosjean1,3,8

1Department of Microbiology and Cell Science, University of Florida
2Université de Provence, Aix-Marseille 1, Marseille, France
3Laboratoire de Chimie Bactérienne (CNRS UPR9043), IFR88, Marseille, France
4Laboratoire de Microbiologie, Institut de Recherches Microbiologiques J.-M. Wiame, Université Libre de Bruxelles, Bruxelles, Belgium
5CEA, DSV, IBEB, Lab Biochim System Perturb, Bagnols-sur-Cèze, France
6Department of Chemistry and Biotechnology, Graduate School of Engineering, University of Tokyo, Bunkyo-ku, Japan
7Donald Danforth Plant Science Center, St. Louis, Missouri
8Institut de Génétique et de Microbiologie, CNRS, UMR8621, Université Paris-Sud, Orsay, France

*Present address: Institut de Biologie et de Médecine Moléculaires, Université Libre de Bruxelles, rue des Professeurs Jeener et Brachet 12, B-6041 Gosselies-Charleroi, Belgium.

*Corresponding author: E-mail: vcrecy@ufl.edu; henri.grosjean@igmors.u-psud.fr.

Associate editor: Hervé Philippe

Abstract

Wyosine (imG) and its derivatives such as wybutosine (yW) are found at position 37 of phenylalanine-specific transfer RNA (tRNAPhe), 3’ adjacent to the anticodon in Eucarya and Archaea. In Saccharomyces cerevisiae, formation of yW requires five enzymes acting in a strictly sequential order: Trm5, Tyw1, Tyw2, Tyw3, and Tyw4. Archaea contain wyosine derivatives, but their diversity is greater than in eukaryotes and the corresponding biosynthesis pathways still unknown. To identify these pathways, we analyzed the phylogenetic distribution of homologues of the yeast wybutosine biosynthesis proteins in 62 archaeal genomes and proposed a scenario for the origin and evolution of wyosine derivatives biosynthesis in Archaea that was partly experimentally validated. The key observations were 1) that four of the five wybutosine biosynthetic enzymes are ancient and may have been present in the last common ancestor of Archaea and Eucarya, 2) that the variations in the distribution pattern of biosynthesis enzymes reflect the diversity of the wyosine derivatives found in different Archaea. We also identified 7-aminocarboxypropyl-demethylwyosine (yW-86) and its N4-methyl derivative (yW-72) as final products in tRNAs of several Archaea when these were previously thought to be only intermediates of the eukaryotic pathway. We confirmed that isowyosine (imG2) and 7-methylwyosine (mimG) are two archaeal-specific guanosine-37 derivatives found in tRNA of both Euryarchaeota and Crenarchaeota. Finally, we proposed that the duplication of the trm5 gene in some Archaea led to a change in function from N1 methylation of guanosine to C7 methylation of 4-demethylwyosines (imG-14).

Key words: tRNA, modification enzymes, methyltransferase, biosynthetic pathway, anticodon loop, mass spectrometry, phylogeny, evolution, Archaea.

Introduction

Posttranscriptional modifications are essential for the activity of transfer RNAs (tRNAs). The most complex modifications are located in the tRNA anticodon loop region and their biosynthesis often requires several enzymatic steps (reviewed in Björk and Hagervall 2005, and Grosjean 2009). The fluorescent and highly hydrophobic wyosine derivatives are among the most hypermodified nucleosides. These derivatives of genetically encoded guanosine (fig. 1A) are located at position 37 (3’ adjacent to anticodon) of phenylalanine-specific transfer RNA (tRNAPhe). They are characterized by an imidazopurine (tricyclic) core structure (1H-imidazo[1,2-a] purine; fig. 1C–I). The first member of the family identified four decades ago in cytoplasmic tRNAPhe of baker’s yeast was wybutosine (abbreviated as yW or simply Y; fig. 1) (Rajbhandary et al. 1967; Thiebe and Zachau 1968; Yoshikami et al. 1968). Wybutosine contains a methyl group on N4 of the imidazopurine core and an α-amino-α-carboxypropyl lateral chain on C7. A simpler version of yW, lacking the C7 lateral chain (designated as wyosine and abbreviated as imG, fig. 1D), was later identified in the yeast Torulopsis utilis (Takemura et al. 1974; Kasai et al. 1976). The absence of yW derivatives at position 37 is the norm in all mitochondrial, chloroplastic, or bacterial tRNAPhe sequenced so far. Instead, m1G (fig. 1B) or derivatives of isopentenyl-adenosine (i6A, ms2i6A or msio6A) are present in these tRNA species (Jühling et al. 2009). In Bacteria, the enzyme TrmD catalyzes...
the insertion of m^1G37 (Byström and Björk 1982), whereas in yeast, the unrelated Trm5 enzyme catalyzes the same reaction (Björk et al. 2001). In Archaea, the sequence of a fully modified tRNA^{Phe} is only available for the euryarchaeon Haloferax volcanii, which harbors m^1G37 (Gupta 1984). The presence of m^1G at this position must, however, be more an exception than a rule. Indeed, many imidazopurine-containing nucleosides were identified by the McCloskey group by analyzing enzymatic digests of unfraccionated (or partially purified) tRNA preparations from diverse Archaea using liquid chromatography–tandem mass spectrometry (LC-MS/MS) (see table 1 for the complete list of organisms and references). These included imG and its isomer isowyosine (imG2, fig. 1E), 4-demethylwyosine (imG-14, fig. 1C), 7-methylwyosine (mimG, fig. 1F), and a fluorescent compound of unknown structure N422 that was identified in bulk tRNA hydrolysate of a few Methanococcales but not of Methanococcus thermolithotrophicus and Methanococcus igneus (McCloskey et al. 2001). The imG2 compound that contains methyl groups on C6 and C7 (fig. 1E) and is of similar molecular weight to imG, was unambiguously identified in Sulfolobus solfataricus P2, Stetteria hydrogenophila, and Pyrobaculum furmarii (three crenarchaeota) (McCloskey et al. 2000; Noon et al. 2003; Zhou, Sitaramaiah, Pomerantz, et al. 2004). By analogy to their eukaryotic counterparts, all these fluorescent archaeal modifications were predicted to be also located at position 37 of tRNA^{Phe} (Urbonavičius et al. 2009).

**Fig. 1.** Guanosine derivatives identified in naturally occurring tRNA^{Phe} of various origins. The derivatives are listed in order of complexity. The symbols for wyosine derivatives are those used in the original papers of McCloskey et al. for Archaea (Zhou, Sitaramaiah, Noon, et al. 2004) or more recently by Noma et al. in Saccharomyces cerevisiae (Noma et al. 2006). The numbering of the atoms is different for guanosine (IUPAC convention) and for wyosine derivatives that follow the convention defined by Blobstein et al. (1973) that was systematically used in recent papers on wyosine nucleosides. In panels G and H, the acp group originating from S-AdoMet is boxed.
Despite this substantial body of work, it is difficult to predict which wyosine derivatives (imG-14, imG, imG2, mimG, and/or N422) are present in a given archaeon. However, because of the obvious structural analogies between the wyosine derivatives found in Eukarya and in Archaea (fig. 1C–I), the recently elucidated yW pathway of yeast (Kalhor et al. 2005; Waas et al. 2005; Noma et al. 2006) is a reasonable starting point to try and decipher the archaeal wyosine pathways.

Wybutosine synthesis in yeast is a multienzymatic process that comprises six sequential enzymatic reactions involving five enzymes. The initial step, the formation of m’G37 in precursor tRNA Phe, is catalyzed by Trm5 (Björk et al. 2001). Next, the tricyclic ring is formed from m’G37 by Tyw1 leading to imG-14 (fig. 1C) (Waas et al. 2005; Noma et al. 2006). Tyw1 is a radical S-Adenosyl-L-Methionine (S-AdoMet)-dependent [4Fe–4S] cluster containing enzyme using FMN as cofactor, but the exact mechanism of the reaction is unknown. In this step, the carbon introduced by Trm5 becomes part of the imidazopurine ring. Next, the alpha-amino-alpha-carboxypropyl (acp) group is transferred from S-AdoMet to the side chain at the C7 position by Tyw2 (also designated Trm12) to produce yW-86 (fig. 1G) (Kalhor et al. 2005; Noma et al. 2006). In the fourth step, an additional S-AdoMet-dependent enzyme Tyw3 methylates the N4 position of the imidazopurine ring to yield yW-72 (fig. 1H) (Kalhor et al. 2005; Noma et al. 2006). Finally, an S-AdoMet-dependent bifunctional

### Table 1. Summary of Previous Work on the Presence of Wyosine Derivatives in Archaeal tRNAs.

<table>
<thead>
<tr>
<th>Subgroup of Archaea Phylum</th>
<th>Organism</th>
<th>imG-14</th>
<th>imG</th>
<th>imG2</th>
<th>mimG</th>
<th>N422</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euryarchaeota</td>
<td>Methanosarcinales</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Methanococcales (subgroup)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Methanococcales (subgroup)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Methanococcus vannielii</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Noon et al. (2003)</td>
</tr>
<tr>
<td></td>
<td>Methanococcus maripaludis</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Edmonds et al. (1991)</td>
</tr>
<tr>
<td></td>
<td>Methanococcus igneus</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>McCluskey et al. (2001);</td>
</tr>
<tr>
<td></td>
<td>Methanococcus thermolithotrophicus</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Noon et al. (2003)</td>
</tr>
<tr>
<td></td>
<td>Methanobacteriales</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Methanothermus fervidus</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Edmonds et al. (1991)</td>
</tr>
<tr>
<td></td>
<td>Methanobacterium thermotartet phicum</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Edmonds et al. (1991)</td>
</tr>
<tr>
<td></td>
<td>Thermococcales (subgroup)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thermococcus sp. (isolate MSB4)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Noon et al. (2003)</td>
</tr>
<tr>
<td></td>
<td>Desulfurococcales</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pyrodiculum occultum</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>McOlskey et al. (1987, 2001);</td>
</tr>
<tr>
<td></td>
<td>Stetteria hydrogenophilus</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Edmonds et al. (1991)</td>
</tr>
<tr>
<td></td>
<td>Thermosolobus thermotartet phicum</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Zhou, Sitaramaiah, Noon, et al. (2004); Zhou, Sitaramaiah, Pomerantz, et al. (2004)</td>
</tr>
<tr>
<td></td>
<td>Methanolobus tindarium</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Edmonds et al. (1991)</td>
</tr>
<tr>
<td></td>
<td>Sulfolobales</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sulfolobus acidocaldarius BC65</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Gupta (1986)</td>
</tr>
<tr>
<td></td>
<td>Su. solfataricus P2</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Gupta (1986)</td>
</tr>
<tr>
<td></td>
<td>Acidianus infernus</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Gupta (1986)</td>
</tr>
<tr>
<td></td>
<td>Thermoproteales</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thermoproteus neutrophilis</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>McOlskey et al. (1987, 2001);</td>
</tr>
<tr>
<td></td>
<td>Pyrolobus fumaris</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Edmonds et al. (1991)</td>
</tr>
<tr>
<td></td>
<td>Pyrolobus fumaris</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Edmonds et al. (1991)</td>
</tr>
<tr>
<td></td>
<td>Pyrobaculum islandicum</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Edmonds et al. (1991)</td>
</tr>
<tr>
<td></td>
<td>From tRNA sequence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Haloferacales</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Haloferacales</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Gupta (1986)</td>
</tr>
</tbody>
</table>

**Note.** — Yes, compound experimentally identified; no, compound lacking (experimentally verified); empty cells, the information was not found in the literature, but the compound is likely to be absent. The only wyosine derivative that has been systematically investigated in all species analyzed is the characteristic archaeal mimG.
enzyme Tyw4 methoxycarbonylates the free alpha-amino group and methylates the free carboxy group to complete yW synthesis (Noma et al. 2006; Suzuki et al. 2009). In this last reaction, the carboxyl group originates from the carbon dioxide. Using the Tyw2, Tyw3, and Tyw4 recombinant proteins, yW was successfully reconstituted from imG-14 (Noma et al. 2006; Suzuki et al. 2009).

In this study, we performed the phylogenetic analysis of homologues of the five yeast yW biosynthetic pathway in Archaea. This led us to propose a model for the origins of wyosine derivatives synthesis in Archaea and for its subsequent evolution in the main archaeal orders. This model both rationalized previous observations made by McCloskey and his collaborators and led to a comprehensive picture of wyosine biosynthesis in Archaea that turned out to be complex and significantly different in the different archaeal orders.

Materials and Methods

Phylogenetic Analysis

The sequences of the 62 archaeal complete genomes available in July 2009 were retrieved from NCBI. From these complete genomes, archaeal homologues of each Saccharomyces cerevisiae enzyme involved in wyosine synthesis (Trm5 = Yhr070w, Tyw1 = YPL207w, Tyw2 = YML005w, Tyw3 = YGL050w, and Tyw4 = YOL141w) were identified using the BlastP program with default parameters (Altschul et al. 1997). Blast outputs were manually checked (no a priori/arbitrary e value cutoff was used). The visual examination of blast outputs showed that for each enzyme, the sequence similarity between homologues was sufficient to discriminate without doubt true positives from false positives or false negatives. The absence of homologues in any archaeal genome was systematically verified by additional searches in the nr database at the NCBI using TBLastN and Psi-Blast with default parameters (Altschul and Koonin 1998). However, no additional archaeal homologues belonging to these five gene families were detected during this step.

Eukaryotic and bacterial homologues of these five enzymes were retrieved from the nr database using the BlastP program. The complete list of the enzymes belonging in the corresponding five gene families is given in supplementary data 1, Supplementary Material online. Distantly related homologues belonging, as Trm5 and Tyw2, to the superfamily of S-adenosylmethionine-dependent methyltransferases (COG1092) were also retrieved from the nr database using Psi-Blast. Among them, a subset of sequences was chosen for a phylogenetic analysis (accession numbers given in supplementary data 4A, Supplementary Material online).

For each enzyme family, homologues were aligned using MUSCLE with default parameters (Edgar 2004). The resulting alignments were then visually inspected and manually refined using ED program from the MUST software (Philippe 1993). At this step, we carefully controlled that our alignments contained only homologous sequences. Regions where homology between sites was doubtful were removed from phylogenetic analyses (all the alignments are available on request). More precisely, we kept only regions bounded by highly conserved positions where gaps (if present) are unambiguously located. The comparison of our manual procedure for site selection with Gblocks (Castresana 2000) showed that in most cases, the number of positions retained was similar (data not shown).

Maximum likelihood analyses were performed using Treefinder, a LG model including a gamma distribution (four categories of sites and an estimated alpha parameter) to modelize among-site rate variation (Jock et al. 2004). The robustness of branches was estimated by a bootstrap analysis as implemented in Treefinder (100 replicates of the original data sets). Bayesian trees were constructed using the program MrBayes 3.1.2 with a mixed substitution model and a gamma distribution (four rate categories) (Ronquist and Huelsenbeck 2003). The Markov chain Monte Carlo search was run with four chains for 1 million generations; with trees being sampled every 100 generations, the first 2,500 trees were discarded as “burn-in.”

Comparative Genomic Analysis

The functional annotations for all predicted archaeal wyosine biosynthesis genes were incorporated in the “Wyosine biosynthesis” subsystem in the SEED database (Overbeek et al. 2005) publicly available at http://theseed.uchicago.edu/FIG/index.cgi. Physical gene clustering was analyzed using the SEED clustering tools. Structure-based alignment of a subset of Trm5a, b, and c sequences were performed using the ESPript platform (Gouet et al. 1999) through the ESPript Web interface (http://escript.ibcp.fr/EScript/EScript/).

Strains and Media

Haloflex volcanii DS70 (Wendoloski et al. 2001) and Halobacterium NRC-1 (ATCC700922) (both gifts of Julie Maupin-Furlow, University of Florida) were grown at 45 °C as described in supplementary data 2, Supplementary Material online. Frozen cell pastes of other archaeal organisms were kindly provided by colleagues in the United States and France: Pyrococcus furiosus and Methanococcus maripaludis S2 (Michael Adams and Barnie Whitman, University of Georgia, Athens, GA), Thermococcus kodakaraensis KDO3, Methanosarcina barkeri, Methanosarcina acetivorans, and Methanobacterium thermoautotrophicum (Tom Santangelo, Joe Krzycki, and Ross C. Larue, Ohio State University, Columbus, OH), Pyroc. abyssi strain GES (Evelyn Marquet and Patrick Forterre, Universite Paris-Sud, Orsay, France), Su. shibatae (Dominique Fourmy, CNRS, Gif-sur-Yvette, France), Pyrob. caldivontis (Todd Lowe, University of California, Santa Cruz, CA). Saccharomyces cerevisiae wild type (BY4742) and corresponding deletion strains (tyw2Δ—strain Y10571, tyw3Δ—strain Y04418, and tyw4Δ—strain Y16650) were obtained from Euroscarf (Frankfurt, Germany). Saccharomyces cerevisiae GBY9 strain harboring a deletion of TRMS (trm5::HIS3), thus
lacking m’G37 in tRNA was obtained from Gunilla Jäger and Glenn Björk (University of Umeå, Sweden). All yeast strains were grown at 30 °C in standard YPD medium and harvested during exponential growth phase. *Salmonella typhimurium* wild-type LT2 and GT5337 (*trmD27*) strains, lacking m’G37 in tRNA, was a gift of Gunilla Jäger and Glenn Björk (University of Umeå, Sweden). These strains were grown at 37 °C in standard L medium.

**Preparation of Bulk tRNA**

As a rule, extreme care was taken when extracting tRNA for wyosine content analysis. Indeed, due to lability of the glycosidic bound, wyosine derivatives can be excised from tRNA under mildly acidic conditions and prolonged incubation (especially under pH 4.5) without breaking the tRNA phosphodiester backbone (Thiebe et al. 1971; Golankiewicz et al. 1985). This is especially important for wyosine derivatives bearing the methyl group at N4 of the imidazolpurine (Golankiewicz et al. 1990). All details about preparation of bulk tRNA are described in supplementary data 2, Supplementary Material online.

**LC-MS/MS Determination of the tRNA Hydrolysates Content**

Four hundred micrograms of bulk tRNA was resuspended in a total volume of 100 μl sterile water. The solution was boiled for 1 min, briefly cooled down on ice, and adjusted to 0.01 M ammonium acetate (pH 5.3) and incubated 12 h at 37 °C in presence of a large excess of nuclease P1 (10 units, Sigma N8630) according to Gehrke et al. (1982). Then, the solution was adjusted to 0.1 M ammonium bicarbonate (pH 7.0), and 0.01 units of Phosphodiesterase I (Sigma P3243) as well as 1 unit of Alkaline phosphatase (Sigma N8630) according to Gehrke et al. (1982). Then, the solution was adjusted to 0.1 M ammonium bicarbonate (pH 7.0), and 0.01 units of Phosphodiesterase I (Sigma P3243) as well as 1 unit of Alkaline phosphatase (Sigma N8630) according to Gehrke et al. (1982). Then, the solution was adjusted to 0.1 M ammonium bicarbonate (pH 7.0), and 0.01 units of Phosphodiesterase I (Sigma P3243) as well as 1 unit of Alkaline phosphatase (Sigma N8630) according to Gehrke et al. (1982).

**tRNA Methylation Assays**

Purification of the recombinant enzymes and preparation of the substrates are described in details in supplementary data 2, Supplementary Material online. The tRNA:m’G37 methyltransferase activity of purified *Pyroc. abyssi* recombinant PAB0505 and PAB2272 (both orthologous to *Sac. cerevisiae* TrrM5) was measured after incubation with the different tRNA substrates as indicated in text and supplementary figures 5 and 6, Supplementary Material online. Routinely, the reaction mixture of 0.1 ml contained 1 μg of purified recombinant enzyme, 100 nCi of [methyl-32C]-S-adenosylmethionine (GE Healthcare), 40 μg of tRNA isolated from the wild-type *Sal. typhimurium* strain LT2 or from the *trmD27* mutant and from yeast wild-type strain BY4742 or *trm5* mutant, or tRNA from yeast wild type or yeast mutants *tyw2Δ* or *tyw3Δ*, in 50 mM Tris–HCl buffer (pH 8.0), 10 mM MgCl2, 0.5 mM DTT, and 5% glycerol. Each mix was incubated for 30 min at 50 °C and then processed as described below. To determine the tRNA (m’G37) MTase activity of PAB0505 and PAB2272 proteins, an alternative method using of synthetic [32P-GTP]-radiolabelled tRNAAsp and tRNAPhe transcripts as substrates was also applied. These radiolabeled substrates were obtained by in vitro transcription with T7-polymerase (Promega) of linearized plasmids harboring synthetic yeast tRNAAsp or tRNAPhe genes. Both the preparation and purification of the resulting transcripts on urea gel have been described previously (Grosjean et al. 1990). Reaction mixtures for testing the enzymatic activity of arcahal PAB0505 and PAB2272 proteins were the same as above, except that ~6,000 cpm [32P-GTP]-radiolabeled
tRNA was used instead of nonradiolabeled substrate, and 80 μM of nonradioactive S-AdoMet (Sigma) was added. In all cases, after incubation, the reactions were 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 cold ethanol, collected by centrifugation, washed once with cold 70% ethanol, dried, and finally completely digested into 5′-monophosphate nucleosides by overnight incubation at 37°C with nuclease P1 (Sigma N8630-1VL, only 1 unit) in 10 μl of 50 mM ammonium acetate/acetic acid buffer at pH 5.3. The resulting hydrolysates containing mostly wyosine dinucleotides pG*pA were then analyzed by two-dimensional thin layer chromatography on 10 × 10-cm cellulose plates as described elsewhere together with the necessary reference maps (Grosjean et al. 2004). Localization of radioactive spots on thin layer plates was performed after overnight exposure (for 32P) or after exposure for several days (for 14C) to a BIOMAX MR film (Kodak).

**Results and Discussion**

**Origin and Evolution of Archaeal Homologues Involved in the Metabolism of Wyosine Derivatives**

Using the sequences of the five yeast enzymes involved in wybutosine biosynthesis (Trm5, Tyw1 to Tyw4) as seeds, we retrieved their homologues from the three domains of life. Depending on the evolutionary scenario considered, LACA may be also the ancestor of Eucarya (see text for more details). The brown, yellow, and green filled circles represent Trm5 subgroups. Blue squares, orange circles, and pink triangles indicate the presence of Tyw1, Tyw2, or Tyw3 homologues, respectively. Evolutionary events that occurred during the evolution of these proteins are indicated by letters: “H” indicates a putative HGT event and “+2” indicates a duplication event. Gene losses are symbolized by empty symbols.

**Fig. 2.** Evolution of the enzymes involved in wyosine biosynthetic pathway in Archaea. The phylogeny shown on the left is a schematic drawing of the relationships between the main archaeal lineages (most of the representative 62 organisms of which fully sequenced genomes were available last July 2009) based on previous published works (Brochier-Armanet et al. 2008). LACA corresponds to the last archaea common ancestor. Depending on the evolutionary scenario considered, LACA may be also the ancestor of Eucarya (see text for more details). The brown, yellow, and green filled circles represent Trm5 subgroups. Blue squares, orange circles, and pink triangles indicate the presence of Tyw1, Tyw2, or Tyw3 homologues, respectively. Evolutionary events that occurred during the evolution of these proteins are indicated by letters: “H” indicates a putative HGT event and “+2” indicates a duplication event. Gene losses are symbolized by empty symbols.
subsequently referred to as Taw1, were identified in all genomes except in three Halobacteriales and in the thaumarchaeon *Cenarchaeum symbiosum*. Archaeal homologues of yeast Tyw2, subsequently referred to as Taw2, were found exclusively in Euryarchaeota, whereas homologues of the eukaryotic Tyw3, here referred as Taw3, were present in all Crenarchaeota in the nanoarchaeon but only in a few Euryarchaeota. By contrast, no Tyw4 homologues were detected in Archaea indicating that the last step of this pathway is specific to eukaryotes. Wysosine derivatives, but not yW, are therefore expected to be present in Archaea (at least in the 62 species analyzed).

To understand the origin and the evolution of the wyosine pathway in Archaea, we performed the phylogenetic analysis of Trm5, Taw1, Taw2, and Taw3 archaeal homologues using *Sac. cerevisiae*, *Homo sapiens*, and *Arabidopsis thaliana* as representatives of eukaryotes. As previously noted by Noma et al. (2006), *Sac. cerevisiae* Trm5 and Tyw2 are homologues and belong to the same Cluster of Orthologous Group (COG2520). A preliminary maximum likelihood phylogenetic analysis of these two enzymes showed that they formed a distinct family within the super family of S-adenosylmethionine-dependent methyltransferases (COG1092) [bootstrap value (BV) = 100%; supplementary data 4A, Supplementary Material online]. Moreover, the resulting topology showed that Tw2a and Trm5 formed two distinct subfamilies each containing archaeal and eukaryotic sequences. This suggested that Trm5 and Taw2 resulted from an ancient gene duplication event. The BVs supporting these two subfamilies are rather weak (BV = 81% supporting the monophyly of Taw2 and BV = 45% for the monophyly of Trm5). This weak statistical support very likely resulted from the poor quality of the alignment linked to the inclusion of very distantly related bacterial homologues (supplementary data 4A, Supplementary Material online). Nevertheless, this analysis clearly confirmed that Trm5 and Taw2 formed a family within the S-adenosylmethionine-dependent methyltransferases. The removal of the distant bacterial S-adenosylmethionine-dependent methyltransferases outgroup led to a phylogeny where Trm5 and Taw2 formed two distinct and well-supported groups (BV = 98%; supplementary data 4B, Supplementary Material online). This is in agreement with the previous result that suggested that these two genes formed distinct but sister subfamilies within the large family of S-adenosylmethionine-dependent methyltransferases. Those two subfamilies were therefore considered separately for further phylogenetic analyses (supplementary data 4C and E, Supplementary Material online). We then performed the phylogenetic analysis of Trm5, Taw1, Taw2, and Taw3 using the maximum likelihood method. The resulting trees should be considered as unrooted because the monophyly of archaea was recently questioned (Pisani et al. 2007; Cox et al. 2008). According to those hypotheses, Eucarya and Archaea are not two sister domains as in classical rooted universal trees. On the contrary, Eucarya are supposed to be related to Crenarchaeota (Cox et al. 2008) or Euryarchaeota (Pisani et al. 2007). Hence, this implies that the last common ancestor of Archaea is also the ancestor of Eucarya. In Trm5, Taw1, Taw2, and Taw3 trees, eukaryotic sequences formed a monophyletic group (BV ≥ 99%; supplementary data 4C–F, Supplementary Material online), but their position relative to specific archaean orders was not robust, reflecting the poor resolution of most basal nodes in those trees. However, this indicated that the presence of these enzymes in archaea and eukaryotes did not result from recent HGTs between these two domains of life. Nevertheless, irrespective of the relationships between Archaea and Eucarya, it is still possible to infer if a given wyosine biosynthesis enzyme was present in the ancestor of Archaea (that may be also the ancestor of Eucarya). Indeed, in this case, the corresponding gene is expected to be widely distributed in Archaea, and more importantly, its phylogeny will be congruent with the phylogeny of Archaea.

To answer this question and further pursue the investigation of the evolutionary history of the archaeal wyosine biosynthesis enzymes, we performed additional phylogenetic analyses without the eukaryotic sequences (fig. 3). Indeed, the eukaryotic sequences are relatively divergent, thus their inclusion limited the number of positions that could be kept for phylogenetic analyses and possibly introduced tree reconstruction artefacts. The resulting trees were congruent but slightly better resolved compare with those containing both archaeal and eukaryotic sequences (fig. 3 and supplementary data 4, Supplementary Material online). In particular, monophyletic groups of sequences corresponding to archaeal orders were recovered and well supported in all trees. More precisely, for Trm5, the monophyly of all archaeal orders but Thermoproteales and Desulfurococcales were recovered and well supported; for Taw1, all but Thermoproteales and Methanosarcinales; for Taw2, all but Methanobacteriales; and for Taw3, all but Thermoproteales and Desulfurococcales. This indicated that the proteins were present in the ancestors of these lineages and that the taxonomic distribution observed for these enzymes did not result from numerous recent HGT. More precisely, the ancestors of each crenarchaeal order likely harbored Trm5, Taw1, and Taw3, whereas the ancestors of most euryarchaeal orders likely harbored Trm5, Taw1, Taw2, and Taw3. This suggested that two different sets of enzymes were present in the ancestor of those two archaeal phyla. If the monophyly of most archaeal orders was well resolved in the four phylogenies, most of the basal (e.g., supra-order) nodes were poorly resolved, certainly because of the small number of positions available for phylogenetic analyses. However, we recovered a number of well-supported supra-order relationships that are in agreement with the archaeal phylogeny. For instance, we recovered the grouping of Sulfolobales and Desulfurococcales in the Trm5, Taw1, and Taw3 trees, and the grouping of Halobacteriales, Methanobacteriales, and Methanosarcinales in the Trm5 tree (fig. 3). Altogether, this suggested that these enzymes might be ancient in Archaea and might have been already present in the ancestor of Archaea (and of Eucarya); even
Fig. 3. Unrooted phylogenies of archaeal enzymes involved in wyosine derivative synthesis. (A) Trm5 (168 unambiguously aligned positions), (B) Taw1 (232 unambiguously aligned positions), (C) Taw2 (177 unambiguously aligned positions), and (D) Taw3 (140 unambiguously aligned positions). The topologies of these trees were computed by a Bayesian approach. Numbers close to nodes represent the posterior probabilities, whereas those correspond to ML BVs computed by Treefinder (for clarity only BVs greater than 50% are shown). The scale bar represents the average number of substitutions per site. Squares correspond to the five archaeal phyla: black, Crenarchaeota; black outlined in gray, Korarchaeota; white outlined in black, Thaumarchaeota; white outlined in gray, Nanoarchaeota; and gray, Euryarchaeota. Group names ending in –ales represent orders.

Biosynthesis of Archaeal Wyosine Derivatives · doi:10.1093/molbev/msq096 MBE

2069 Downloaded from https://academic.oup.com/mbe/article-abstract/27/9/2062/1009405 by guest on 04 February 2019
The hypothesis that the distribution of these enzymes in archaea resulted from a few and ancient HGT events, even if less parsimonious, could not be definitively ruled out. Accordingly, their absence in any archaeal lineage should therefore be interpreted as secondary losses.

One clear case of HGT was, however, observed for Trm5. In the corresponding tree, a few euryarchaeal sequences (i.e., from Thermococcales, a few Methanosarcinales and Methanomicrobiales and *Archaeoglobus fulgidus*), as well as the nanoarchaeon one, formed a monophyletic group (posterior probability = 1.00, BV = 74%; fig. 3A) and emerged from within crenarchaeota, and more precisely, as a sister group of Thermoproteales (posterior probability = 1.00, BV = 79%; fig. 3A). This suggested that an HGT occurred from an ancestor of Thermoproteales to one of these euryarchaeota, followed by secondary HGT between these euryarchaeal lineages (fig. 3A). Some Euryarchaeota, like the Thermococcales, conserved both the Trm5 copy of crenarchaeal origin and their native copy (figs. 2 and 3A) and emerged from within crenarchaeota, and more precisely, as a sister group of Thermoproteales (posterior probability = 1.00, BV = 79%; fig. 3A). This suggested that an HGT occurred from an ancestor of Thermoproteales to one of these euryarchaeota, followed by secondary HGT between these euryarchaeal lineages (fig. 3A). Some Euryarchaeota, like the Thermococcales, conserved both the Trm5 copy of crenarchaeal origin and their native copy (figs. 2 and 3A). These organisms consequently harbor two types of Trm5 coding genes. Others, such as *Archaeoglobus fulgidus* or the nanoarchaeon, have lost their native copy (fig. 2). To distinguish between the two homologues present in these Euryarchaeota species, we proposed to call the native form and by extension all the Trm5 sequences of euryarchaeal origin Trm5b, whereas the copy of crenarchaeal origin will be referred as Trm5a (figs. 2 and 3A). An additional and divergent Trm5 sequence, referred as Trm5c (figs. 2 and 3A), was found in two of the three crenarchaeal orders (i.e., Desulfurococcales and Sulfolobales). Trm5c emerged within the euryarchaeal sequences, as sister group of *Methanopyrus kandleri*, suggesting an HGT from Euryarchaeota to Crenarchaeota. However, because of the poor resolution of basal nodes in the Trm5 phylogeny, the hypothesis of a duplication of Trm5a in the ancestor of Desulfurococcales and Sulfolobales could not be definitively discarded.

This phylogenetic analysis allowed us to firmly establish the distribution of all wyosine biosynthesis genes in sequenced archaeal genomes and was used to make predictions on the function of the archaeal enzymes and also predict that yW modifications are found in a given organism (fig. 4).

**Formation of m′G37 in Archaeal tRNA**

All Archaea without any exception encode at least one Trm5 homologue (fig. 2). They should therefore all have the capacity to produce m′G in G37-containing tRNAs not only in tRNA^{Phe} but also in many others such as tRNA^{Leu}, tRNA^{Pro}, and tRNA^{His}. This was previously confirmed in *Halof. volcanii* (which encodes Trm5b but no enzyme involved in wyosine derivative synthesis—fig. 2) by directly sequencing all the tRNA species (Gupta 1984, 1986). Analysis of nucleoside content of bulk tRNA from *Halof. volcanii* by LC-MS/MS (fig. 5) confirmed this early result as no wyosine derivatives were detected beside the expected m′G (MH+ 298 m/z at 21.0 min). Two other Halobacteria as well as the thaumarchaeon *C. symbiosum* harbor only one Trm5 coding gene, but no other predicted gene involved in wyosine biosynthesis (fig. 2). Accordingly, m′G37 is expected to be found in tRNA^{Phe} in these archaea as final product (pathway 1 in fig. 4).

As mentioned above, Trm5 encoding genes can be found in some archaea in more than one copy. This raises the hypothesis that the distribution of these enzymes in archaea resulted from a few and ancient HGT events, even if less parsimonious, could not be definitively ruled out. Accordingly, their absence in any archaeal lineage should therefore be interpreted as secondary losses.
Fig. 5. Analysis of yW content in bulk tRNA from selected archaeal species. Chromatograms showing the separation of the yW derivatives from archaeal tRNA (Haloferax volcanii, Pyrococcus furiosus, Methanococcus maripaludis, Methanosarcina acetivorans, Pyrococcus abyssi, Halobacterium salinarium, Sulfobatus shibatiae, and Pyrobaculum calidifontis) using LC-MS/MS. The modifications detected and their retention times are indicated in the chromatograms. Next to each ultraviolet trace, the extracted ion chromatograms (XIC) of each yW derivative detected were also indicated. The peak area (PA) indicated next to the peaks was integrated from the XIC of each modification (mG, imG-14, imG, imG2, mimG, yW-86, and yW-72) and normalized using the m^2G modification (MH^- 312 m/z).
the question of their functional role. Therefore, we tested the ability of purified recombinant Pyroc. abyssi Trm5b (PAB0505) and Trm5a (PAB2272) enzymes to form m^1G in tRNA in vitro using [14C-methyl]-S-AdoMet as methyl donor. Bulk tRNA from Sal. typhimurium trmD27 or Sac. cerevisiae trm5 strains, both lacking m^1G37, or from the corresponding wild-type strains (containing fully mature tRNA harboring m^1G37) were used as substrates. As shown from thin layer chromatographies of radiolabeled tRNA hydrolysates (supplementary data 5, panels A–F, Supplementary Material online), both enzymes were able to catalyze the formation of radioactive m^1G-containing tRNA and this only when the tRNA was extracted from the trmD37 or trm5Δ backgrounds, thus lacking m^1G37 (panels A + B and D + E, respectively). However, differences in the enzymatic activities of the two proteins were observed: First, PAB2272 (Trm5a) produced much less radioactive material on prolonged incubation under identical experimental conditions than PAB0505 (Trm5b) (compare panels A + D with panels B + E in supplementary data 5, Supplementary Material online); second, PAB2272 (Trm5a) was only able to catalyze the formation of m^1G in purified [32P] G-radiolabeled yeast tRNA^Phe^ transcript (panel H) and not in purified [32P] G-radiolabeled yeast tRNA^Asp^ transcript (panel K) that also contains a G at position 37, whereas PAB0505 (Trm5b) could methylate both these transcripts (panels G and J in supplementary data 5, Supplementary Material online). These results demonstrated that both Trm5a and Trm5b can catalyze the methylation of G37 in tRNA but that Trm5a was less efficient and appears to be selective for methylating G37-containing tRNA^Phe^.

Formation of the Minimalist Wyosine Derivative (imG-14)
In Sac. cerevisiae, the formation of imG-14 by Tyw1 is the first dedicated step in synthesis of yW. The main difference between the archaeal Taw1 proteins and their yeast homologue is that the former are shorter because they lack the N-terminal FMN binding/flavodoxin domain found in Tyw1 from Sac. cerevisiae (Goto-Ito et al. 2007; Suzuki et al. 2007). Examination of amino acid sequences of Trm5a, Trm5b, Trm5c, Taw2, and Taw3 homologues did not reveal any obvious cryptic flavodoxin domain fused to an archaeal wyosine enzyme. These observations therefore raise the possibility that in Archaea, an additional flavodoxin protein acting in trans with Taw1 is required for wyosine biosynthesis. A survey of the regions encompassing the taw1 genes in all the archaeal genomes in the SEED database revealed one possible candidate, the gene family coding for a flavodoxin family protein (the thioredoxin reductase TrxB, COG0492) located next to taw1 in Aeropyrum pernix and next to trm5c in Metallosphaera sedula and in three Sulfolobus (see Wyosine biosynthesis subsystem in the SEED database: http://theseed.uchicago.edu/FIG/index.cgi). This physical clustering could be a sign of functional association (Osterman and Overbeek 2003). However, as these flavodoxin homologues are found in all Archaea, including the three Halobacteriales and C. symbiosum that lack Taw1, this putative cofactor TrxB, if involved in wyosine biosynthesis, is most certainly involved in other metabolic pathways and this hypothesis absolutely requires further experimental validation. Archaeal Taw1 homologues are found in nearly all the genomes analyzed (fig. 2). Thus, based on the hypothesis that Taw1 and Tyw1 share the same function, imG-14 is expected to be synthesized in all these Archaea at least as an intermediate in the synthesis of more complex wyosine derivatives. Indeed, imG-14 had been previously detected in tRNA of some Crenarchaeota (Sulfolobus sp. and Pyrob. fumarii) and some Euryarchaeota (various Methanococcales and Methanococcoides burtonii); however, only as a minor component compared with the other wyosine derivatives detected in the same tRNA samples (table 1 and references therein). In an independent set of experiments, we were also able to detect imG-14 in bulk tRNA extracted from Pyroc. furiosus, Methanococcus maripaludis, and Methanosarcina acetivorans (MH^+ 322 m/z at 29.2–29.3 min; fig. 5) as well as in the control yeast strain deleted in TYW2 known to accumulate imG-14 (supplementary data 3, Supplementary Material online; see also Noma et al. 2006). Because imG-14 is mainly an intermediate, it may not always accumulate enough for detection as in Pyroc. abyssi (fig. 5). The only case where imG-14 might be a possible end product of wyosine metabolism is in Nitrosopumilus maritimus because its genome encodes only for Trm5a and Taw1 (fig. 2 and pathway 2 in fig. 4).

Formation of yW-86 and yW-72 in Euryarchaeota
As discussed in the phylogenetic analysis presented above, Taw2 proteins are found in the majority of Euryarchaeal genomes analyzed and absent in all Crenarchaeota without exceptions (fig. 2). The yeast homologue Tyw2 catalyzes the formation of yW-86, a compound that had not been formally identified in Archaea until very recently using Pyroc. horikoshii and Methanocaldococcus jannaschii Tyw2 enzymes in vitro (Urnitsu et al. 2009). However, analysis of bulk tRNA from Halobacterium salinarium sp. NRC1 that contains only homologues of the Trm5b, Taw1, and Taw2 coding genes (fig. 2) identified a compound eluting at 24.2 min with a MH^+ of 423 m/z (fig. 5). The MS/MS fragmentation profile of this compound showed similar fragments to what was initially observed for yW-86 MS/MS fragmentation obtained from tRNA of the yeast tyw3Δ strain known to accumulate this intermediate (supplementary data 3, Supplementary Material online; see also Noma et al. 2006). The same compound was also identified in bulk tRNA from Pyroc. furiosus, Pyroc. abyssi, Methanosarcina acetivorans, and Methanococcus maripaludis (fig. 5) and the corresponding MS/MS fragmentation confirmed it was indeed yW-86 (shown for Pyroc. furiosus in supplementary data 3, Supplementary Material online). yW-86 has the same elution profile as the previously identified putative wyosine derivative N422 of unknown structure originally
detected by McCloskey’s group in tRNA of *Methanococcus maripaludis, Methanocaldococcus jannaschii, and Methano-
coccus vannielii* (table 1; McCloskey et al. 2001) that are known to contain a Taw2 homologue (fig. 2). As expected, this compound was absent in tRNA extracted from *Halof. volcanii* that contains only Trm5 and of *Su. shibatae* and *Pyrob. calidifontis* that both lack Taw2 (cf. fig. 2 and fig. 5).

These results suggested that Taw2, like the yeast homologue Tyw2, catalyzes the insertion of an acp side chain at C7 of imG-14, a prediction nicely confirmed recently by the group of Nureki in Japan. They demonstrated that purified recombinant Taw2 of *Pyroc. horikoshii* and of *Methanocal-
dococcus jannaschii* can transfer the acp moiety from S-AdoMet to imG-14 containing yeast tRNA\textsuperscript{Phe} in vitro (Umitsu et al. 2009). The existence of an acp side chain in an archaeal wyosine derivative is therefore not specific to eukaryotic yW derivatives and may actually be one of the most common wyosine derivatives synthesized in Euryarchaeota (pathways 3–6 in fig. 4). In a few cases (pathways 5 and 6 in fig. 4), yW-86 is further methylated on the N4 position of the imidazopurine ring by Taw3 (discussed below) to form yW-72 as detected in *Pyroc. furiosus* and *Pyroc. abyssi* (fig. 5).

**Formation of imG2 and mimG, Two Archaeal-
Specific Modifications**

**Distribution of imG2 and mimG in Archaea**

ImG2 and mimG contain a methyl group at C7 instead of the acp side chain found in eukaryotes and some Euryarchaeota. Both derivatives were previously unambiguously identified in several Crenarchaeota such as *Su. solfataricus* P2, *Su. acidocaldarius,* and *Acidianus infernus* (table 1 and references therein). This observation was confirmed for a fourth representative of Sulfolobales, *Su. shibatae,* where the dimethylated imG-14 (mimG, MH\textsuperscript{+} 350 m/z at 33.0 min) was found to be more abundant than the monomethylated intermediate (imG2, MH\textsuperscript{+} 336 m/z at 31.9 min) (fig. 5). MimG had also been identified in tRNA of Thermoproteales (table 1). We confirmed this for *Pyrob. calidifontis* where the presence of mimG was shown by the MS/MS fragmentogram profile of a MH\textsuperscript{+} 350 m/z compound eluting at 33.0 min (see fig. 5 and supplementary data 3, Supplementary Material online).

**Methylation at N4 of imG-14**

*Pyrobaculum calidifontis* has only three putative yW biosynthetic gene homologues (Trm5a, Taw1, and Taw3—fig. 2), but to account for the presence of mimG, three methyla-
tions must occur in tRNA\textsuperscript{Phe} of this organism: First, the methylation of N1 of G37 to form m1G37, then the methyla-
tions of the C7 and N4 positions of the tricyclic core. Because of the homology with the yeast Tyw3 family, we assumed that Taw3 is responsible for the methylation at N4 of imG-14. We predict, however, that the archaeal Taw3 protein is able to methylate both yW-86 and imG-14 at the N4 position, as attested by the presence of yW-72 and imG in *Pyroc. abyssi* and *Pyroc. furiosus* (fig. 5), whereas in *Sac. cerevisiae,* the Tyw3 substrate spectrum appears restricted to yW-86 (Noma et al. 2006). It is also possible that in *Pyroc. abyssi/Pyroc. furiosus,* both imG and yW-72 are synthesized because of the presence of two copies of Taw3 (fig. 2), each Taw3 protein targeting different components leading to the formation of yW-72 or to mimG (pathway 6 in fig. 4).

**Methylation at C7 of imG-14**

The identity of the enzyme responsible for the methylation at C7 was not as obvious. Our phylogenetic analysis combined with biochemical data obtained on recombinant en-
zymes lead to the proposition that Trm5a catalyzes not only the N1 methylation of G37 but also the C7 methyla-
tion of imG-14. This seemed chemically plausible as the carbon of the newly added methyl group in m1G37 corresponds to the position of C7 in the imidazopurine ring (compare fig. 1B with fig. 1C). The active site of Trm5a could therefore accommodate imG-14 and methylate the ‘C7’ position. This hypothesis was tested in vitro using the purified recombinant *Pyroc. abyssi* Trm5a (PAB2272). When incubated at 50°C in the presence of [1\textsuperscript{4}C]-methyl-S-AdoMet, PAB2272 not only catalyzes the formation of m1G37 in tRNA\textsuperscript{Phe} (see supplementary data 5, Supplementary Material online) but also incorporates a substantial amount of radioactivity into bulk tRNA extracted from a *Sac. cerevisiae* tyw2Δ strain that contains imG-14 instead of the normal yW base. 2D-TLC analyses of the nucleoside content after complete digestion of the radioactive tRNA into nucleotides, using two different types of chromatographic systems (supplementary data 6, panels A and D, respectively, Supplementary Material online), revealed the presence of a radioactive spot characteristic of the wyosine derivatives (possibly pimG2pA) (Droogmans and Grosjean 1987; Grosjean et al. 1990). Control experiment using bulk wild-type yeast tRNA (containing yW nucleotide) gave no trace of such methyl incorporation (supplementary data 6, panels C and F, Supplementary Material online). When tRNA extracted from a *Sac. cerevisiae* tyw3Δ strain harboring yW-86 was used, a faint spot that clearly migrates as pm1G monophosphate was detected and an even fainter spot migrating similarly to the one identified when tRNA from the tyw2Δ strain was used as substrate (supplementary data 6, panels B and E, compare with supplementary data 5, Supplementary Material online). The formation of such small amount of m1G, as well as of the wyosine derivative, is certainly due to the presence in bulk tRNA extracted from the tyw3Δ strain of a small fraction of tRNA\textsuperscript{Phe} harboring G37 as well as imG-14 in addition to the expected yW-86. Additional control experiments performed under identical conditions with *Pyroc. abyssi* Trm5b (PAB0505) gave no traces of wyosine derivative formation when tRNA from tyw2Δ strain harboring imG-14 was used (supplementary data 6, panel G, Supplementary Material online), whereas when tRNA from tyw3Δ strain harboring yW-86 was used, traces of m1G were detected (supplementary data 6, panel H, Supplementary Material online), probably because of the presence of some trace amounts of G37-containing tRNA\textsuperscript{Phe} in samples originating from the
yeast *tyw3A* strain. It is worth noting that the chromatographic migration of the putative pimG2A (supplementary data 6, panel D, Supplementary Material online) was distinct from that of pyW-72pA as well as from pyWpA as attested by control experiments using Tyw3p and Tyw4p and tRNA from *tyw3D* and from *tyw4D* mutants, respectively (see figure at the end of supplementary data 2, Supplementary Material online).

**Comparisons of the Trm5a, Trm5b, and Trm5c Families**

The results of the in vitro analysis suggested that *Pyroc. abyssi* Trm5a but not Trm5b has the ability to methylate the C7 position of imG-14. The Trm5 enzymes are part of COG2520, a family with members that have already been shown to catalyze different types of reactions. The *Sac. cerevisiae* Trm5 and the corresponding archaeal *Methanocaldococcus jannaschii* Trm5b protein have been experimentally shown to be genuine m1G37 methyltransferases (Björk et al. 2001; Christian et al. 2004; Goto-Ito et al. 2008) and we have reproduced this result here for the *Pyroc. abyssi* homologue (PA80505). The Taw2/Tyw2 proteins that are also COG2520 members catalyze a totally different reaction with the transfer of acp from AdoMet to C7 of imG-14 (Kalhor et al. 2005; Noma et al. 2006; Umitsu et al. 2009, and this work).

To investigate if differences in amino acid sequences between the Trm5a and Trm5b proteins could explain potential differences in activity, we build a structure-based multiple alignment of all the Trm5a, Trm5b, and Trm5c proteins using the structure of Trm5b from *Methanocaldococcus jannaschii* (MJ0883; Protein Data Bank ID = 2YX1) as seed (a subset of the alignment is shown in supplementary data 7, Supplementary Material online). This alignment was analyzed in the light of the recent costructure of MJ0883 with its tRNA substrate (Goto-Ito et al. 2009), and the results are summarized in figure 6. Most of the residues involved in cofactor (S-AdoMet) binding in MJ0883 are strictly conserved in all Trm5abc proteins analyzed (such as Phe203 in motif 1, Asn225 in motif 2, and Asp251 in motif 3 that are conserved in every Trm5abc but *Nanoarchaeum equitans*) or replaced by conservative amino acid changes (such as Asp/Glu223 and Ile/Leu/Val224 in motif 2). More variations were observed in the key substrate binding residues. Two of the residues involved in recognition of the guanine ring and ribose moieties in the costructure (Goto-Ito et al. 2009) that are shared between guanosine and imG-14 are conserved (with only conservative changes observed) in all Trm5abc sequences: these are Arg145 and Tyr177 in MJ0883 (supplementary data 7, Supplementary Material online). Among the others (empty circles in supplementary data 7, Supplementary Material online), one residue clearly discriminates between the Trm5a and Trm5b families. All Trm5a have a Pro or Thr residue instead of the Asn residue found in all Trm5b proteins at position 265 (fig. 6 and red arrow in supplementary data 7, Supplementary Material online). Asn265 in the *Methanocaldococcus jannaschii* Trm5b is the first residue of the NPPY motif (fig. 6) known to be

![Fig. 6. Domain organization of COG2520 proteins. A summary of the structure-based alignments presented in supplementary data 7, Supplementary Material online, is shown here. The dashed gray D1 domains signify that some members of subfamily have the domain and some not. Full black D1 domains mean all members of the family have one. Numbering in the motifs is based on the *Methanocaldococcus jannaschii* Trm5b sequence (MJ0883). The one letter conventional symbols for amino acids are used. h means hydrophobic residue and x means any residue, residues that are in more than 70% of the subfamily are underlined and residues conserved in all the subfamilies are boxed.](https://academic.oup.com/mbe/article-abstract/27/9/2062/1009405)
important in positioning the target nitrogen atoms in several aminomethyltransferases (Bujnicki 2000). The side-chain carbonyl group of this specific residue was shown to hydrogen bond with the exocyclic 2-amino group of G37 in Methanocaldococcus jannaschii Trm5b (Goto-Ito et al. 2009). The change from Asn265 to a Pro or Thr residue might disrupt the binding site and might allow the accommodation of the imG-14 substrate by Trm5a proteins. A second major difference is the absence of D1 domain in 16 of 21 analyzed Trm5a proteins (data not shown). This domain was shown to be important for tRNA recognition in Methanocaldococcus jannaschii Trm5b (Goto-Ito et al. 2009), so Trm5a proteins might use another mechanism to recognize tRNA or might be in a complex with other tRNA binding wyosine biosynthesis enzymes. All Trm5c proteins analyzed contained the D1 domain and had an Asn (or Asp) at position 265. They seem more similar to Trm5b than to Trm5a, and we therefore predict that Trm5c, like Trm5b, catalyzes only the methylation of N1 of G37. The presence of Trm5c may have allowed Trm5a to function exclusively at methylating the C7 of imG-14 in all the Desulfurococcales and Sulfolobales (pathway 7 in fig. 4). This would be an interesting case of enzyme evolution that should be further studied by detailed structure function analyses. The N. equitans Trm5a enzyme that from our analysis (fig. 2) is predicted to have dual function is also of obvious interest, and the structure-based alignments shows specific changes in the N. equitans Trm5a protein with two specific insertions and several mutations in strictly conserved residues such as a Tyr residue in motif 3 corresponding to the strictly conserved Asp in all other Trm5 proteins (fig. 6 and supplementary data 7, Supplementary Material online).

Experimental Validating Predictions from the Comparative Genomic Analysis

If Trm5a and Taw3 methylate imG-14 at the C7 and N4 position, respectively, then mimG should be present in tRNA\(^{\text{Phe}}\) of all Crenarchaeota and Korarchaeota because of the presence of both Trm5a and Taw3, whereas yW-86/yW-72 should be absent because of the lack of Taw2 (pathway 7 in fig. 4). Also both the imG-14 and imG2 intermediates could accumulate. As shown in our LC-MS analysis (fig. 5), these predictions were confirmed for both Sul. tokodaii and Pyrob. caldiphilis. Similarly, tRNA\(^{\text{Phe}}\) molecules of the few Euryarchaeota that have both Trm5a and Taw2 homologues but no Taw3 homologues, like Methanosarcina acetocepora, should harbor imG2 and/or yW-86 (pathway 4 in fig. 4) and no mimG nor yW-72. Indeed, as shown in figure 5, both types of wyosine derivatives imG2 and yW-86 were detected by HPLC-MS analysis in Methanosarcina acetivorans. Finally, Thermococcales that have homologues of Taw2 but also of both Trm5a and Taw3 should in addition to imG2 and yW-86 contain mimG and yW-72 (pathway 6 in fig. 4). Indeed, the LC-MS/MS analyses of bulk tRNA hydrolyses from both Pyroc. abyssi and Pyroc. furiosus did reveal the presence of all these compounds (fig. 5). MimG was not detected in the tRNA hydrolysates from the Euryarchaeota that lacked Taw3 analyzed in this work (as Methanosarcina acetivorans and Halob. salinarum) and also in other Methanococcales analyzed by McCloskey and collaborators (table 1). The case of Methanococcus maripaludis is interesting because this organism does harbor a gene coding for Taw3 (in addition to Trm5b, Taw1, and Taw2; fig. 2) and, at variance with the Thermococcales, HPLC-MS analysis (fig. 5) did not reveal the presence of the expected yW-72 (indicated as dashed red arrow in pathway 5 in fig. 4). Thus, either the amount of yW-72 is beyond the limit of its detectability or it does not form under the physiological conditions used. A very small amount of imG was detected attesting that Taw3 was functional (dashed black arrow in pathway 5 in fig. 4). Likewise, imG was detected in Thermococcales (dashed black arrow in pathway 6 in fig. 4). Thus, the in vivo activity of Taw3, and possibly of other enzymes of wyosine metabolism in Archaea, may well depend on the physiological growth conditions.

Conclusions

The combination of phylogenetic analysis, comparative genomic analysis, and analytical and biochemical experimental validations presented here allows to give an overview of the multiple biosynthetic routes leading to the various wyosine derivatives found in Archaea (fig. 4). Several key conclusions can be derived from this work. First, our phylogenetic analyses reveal that a full pathway similar to the eukaryotic one but maybe lacking the last step catalyzed by Tyw4 (fig. 4, pathway of Sac. cerevisiae) might have been present in the archaeal ancestor. It would have then evolved by gene loss, gene duplication, and HGT to the different biosynthetic paths found in extant Archaea. Second, yW-86 and yW-72 are obviously not specific to Eucarya as previously thought but found in several euryarchaeota. Third, imG2 and mimG were confirmed to be specific to archaea and found both in euryarchaeota and in crenarchaeota as previously shown by the McCloskey group (table 1). Fourth, the most complex archaeal yW patterns are found in Thermococcales that contain all possible archaenal yW compounds (pathway 6 in fig. 4). Finally, a combination of biochemical and phylogenetic analyses led us to propose that the archaeal Trm5a family has acquired additional catalytic activity and is now able to methylate the C7 position of imG-14 in addition (or in replacement) to the N1-methylation of G37. The identification of several amino acid differences between the Trm5b and Trm5a families in key substrates binding residues lays the foundation for more detailed structure function studies that are required to understand the change in substrate specificity that could have occurred in the Trm5a subfamily.

Supplementary Material

Supplementary data 1–7 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).
Acknowledgments

This work was supported by the National Science foundation (MCB-0516948 to V.d.C.L.) and the Commissariat à l’Energie Atomique (to J.A.). C.B.A is supported by an “Action Thématique et Initiante sur Programme” of the French Centre National de la Recherche Scientifique (www.cnrs.fr). J.U. was supported by a postdoctoral fellowship from the Fonds pour la Recherche Fondamentale Collective (FRFC) and by a Federation of European Biochemical Societies Distinguished Young Investigator Award. L.D. was supported by a grant of the FRFC and by the Fonds D. and A. Van Buuren. H.G. holds a position of Emeritus scientist in the laboratory of Prof. J. P. Rousset at University of Paris. The authors would like to thank Helen McGuirk and Marc Bailly for help with the tRNA preparations, Guylaine Chausain for PAB0505 and PAB2272 cloning and expression experiments, Charles Marchetti for producing recombinant cells, Ian Blaby for critical reading and editing of the manuscript, and Basma El Yacoubi for editing the figures.

References


Takemura S, Kasai H, Goto M. 1974. Structure of wye (Yt base) and wyosine (Yt) from Torulopsis utilis phenylalanine tRNA. Nucleic Acids Res. 15:683–693.


