The appearance of pyrrolysine in tRNA^{His} guanylyltransferase by neutral evolution

Ilka U. Heinemann[a], Patrick O'Donoghue[a], Catherine Madinger[b], Jack Benner[b], Lennart Randau[a], Christopher J. Noren[b], and Dieter Söll[a,c,1]

Departments of [a]Molecular Biophysics and Biochemistry and [c]Chemistry, Yale University, New Haven, CT 06520-8114; and [b]New England Biolabs, 240 Country Road, Ipswich, MA 01938-2723

Contributed by Dieter Söll, October 19, 2009 (sent for review October 7, 2009)

tRNA^{His} guanylyltransferase (Thg1) post-transcriptionally adds a G (position −1) to the 5′-terminus of tRNA^{His}. The *Methanosarcina acetivorans* Thg1 (MaThg1) gene contains an in-frame TAG (amber) codon. Although a UAG codon typically directs translation termination, its presence in *Methanosarcina* mRNA may lead to pyrrolysine (Pyl) incorporation achieved by Pyl-tRNA^{Pyl}, the product of pyrrolysyl-tRNA synthetase. Sequencing of the MaThg1 gene and transcript confirmed the amber codon. Translation of MaThg1 mRNA led to a full-length, Pyl-containing, active enzyme as determined by immunoblotting, mass spectrometry, and biochemical analysis. The nature of the inserted amino acid at the position specified by UAG is not critical, as Pyl or Trp insertion yields active MaThg1 variants in *M. acetivorans* and equal amounts of full-length protein. These data suggest that Pyl insertion is akin to natural suppression and unlike the active stop codon reassignment that is required for selenocysteine insertion. Only three Pyl-containing proteins have been characterized previously, a set of methylamine methyltransferases in which Pyl is assumed to have specifically evolved to be a key active-site constituent. In contrast, Pyl in MaThg1 is a dispensable residue that appears to confer no selective advantage. Phylogenetic analysis suggests that Thg1 is becoming dispensable in the archaea, and furthermore supports the hypothesis that Pyl appeared in MaThg1 as the result of neutral evolution. This indicates that even the most unusual amino acid can play an ordinary role in proteins.

amber codon | *Methanosarcina acetivorans* | natural supression

The methanogenic archaeon *Methanosarcina acetivorans* is one of only 11 organisms (among nearly 1,000 for which complete genomic information is available) that genetically encode pyrrolysine, the 22nd amino acid. This unusual amino acid, a lysine in N$^{\varepsilon}$-amide linkage to a pyrroline ring, is cotranslationally inserted in response to a UAG (amber) codon (1). While typically a stop codon in ribosomal protein synthesis, in Pyl-decoding organisms UAG is read through by Pyl-tRNA^{Pyl}, an amber suppressor tRNA acylated by pyrrolysyl-tRNA synthetase (PylRS) (2, 3). The genes encoding tRNA^{Pyl} (*pylT*) and PylRS (*pylS*) are normally present in an operonal organization with three genes (*pylBCD*) responsible for biosynthesis of Pyl (4). According to structural (1) and genetic (5) analysis, Pyl plays an essential catalytic role in the active sites of a set of methylamine methyltransferases that allow the *Methanosarcinales* to use the atypical growth substrates mono-, di-, and trimethylamines as sole carbon sources. To date, the only biochemically characterized Pyl-containing proteins are the monomethylamine, dimethylamine, and trimethylamine methyltransferases (1, 6). Mutagenesis of the Pyl residue in these or any other putative Pyl-proteins has not yet been reported.

Regulation of the methylamine methyltransferases is based on the available growth substrate. They are highly expressed in vivo during growth on methylamine, but weakly or not expressed when *M. acetivorans* is grown on methanol (7). The importance of Pyl in these enzymes initially suggested that the availability of Pyl-tRNA^{Pyl} could also be regulated in a similar fashion. Quantitative PCR data revealed constitutive expression of tRNA^{Pyl} in growth on trimethylamine (TMA) or on methanol in *Methanosarcina mazei* (8); likewise a monomethylamine methyltransferase (MtmB1) expression construct could be overexpressed in *M. acetivorans* during growth on methanol (7). Nevertheless, the fact that Pyl is found in the active sites of the methylamine methyltransferases and that all Pyl-decoding organisms share the Pyl operon and at least one methylamine methyltransferase has led to the belief that Pyl was selectively retained during evolution to support growth on methylamines (9). While this may be correct, it is possible that Pyl has found its way into other proteins, as evidenced by a number of putative genes with in-frame amber codons found in organisms that encode the Pyl operon (10).

Nature evolved the genetic code to 22 amino acids by reassigning two termination codons, UGA for selenocysteine, and UAG for Pyl (11). While selenocysteine insertion relies on a complete reassignment of codon sense that involves a special elongation factor and an RNA signal (12), all studies to date indicate that Pyl insertion does not require a UAG reassignment mechanism in *Methanosarcina* (7) or in a heterologous *Escherichia coli* context (4, 13). Thus, Pyl insertion is reminiscent of nonsense suppression, in which certain aminoacyl-tRNAs compete successfully with release factors and bring about translation read-through of in-frame stop codons (14). Unlike the informational suppressor tRNAs that derive from duplicated tRNA species by anticodon mutations, Pyl insertion is a case of "natural suppression" (ref. 15 and references therein) where a normal tRNA/aminoacyl-tRNA synthetase pair present in wild-type cells achieves the nonsense codon read-through. In *Methanosarcina*, this would be tRNA^{Pyl} and PylRS, which clearly coevolved early in the history of life (16) for the specific incorporation of Pyl in response to UAG codons.

In a study of archaeal tRNA^{His} guanylyltransferases (Thg1), we found that *M. acetivorans* may encode a Thg1 enzyme (MaThg1) in a putative ORF that includes an in-frame amber codon at position 142; this presents an opportunity to investigate the efficiency of natural suppression by tRNA^{Pyl} in *M. acetivorans*. Thg1 is involved in tRNA^{His} biosynthesis. During tRNA maturation the 5′ leader sequence of precurser tRNA is removed by RNase P (17). Most known tRNA^{His} species, however, have an additional guanosine nucleotide appended to their 5′ terminus (18). This G-1 residue is the essential identity element in tRNA^{His} for His-tRNA^{His} formation (reviewed in ref. 19). In many bacteria, the G-1 residue is genome-encoded; RNase P mis-cleaves the precursor tRNA^{His} and leaves G-1 in the mature tRNA (20, 21). In eukaryotes, the essential Thg1 enzyme rescues

Author contributions: I.U.H., P.O., and D.S. designed research; I.U.H., P.O., C.M., J.B., and L.R. performed research; I.U.H., P.O., C.M., J.B., and C.J.N. analyzed data; and I.U.H., P.O., J.B., C.J.N., and D.S. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: dieter.soll@yale.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0912072106/DCSupplemental.

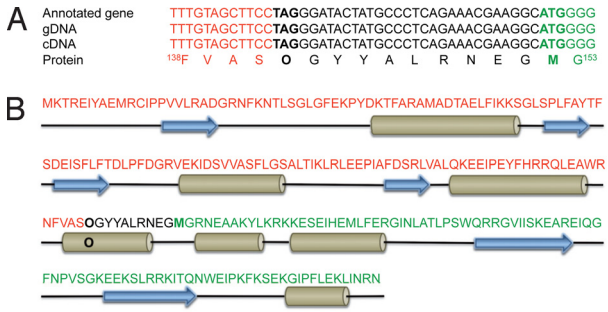

Fig. 1. Encoding of Pyl in MaThg1. Sequences are displayed in green for peptide MA0816 and in red for MA0817. The amber stop codon TAG and the downstream Met codon ATG are marked in bold. The Pyl containing peptide linker and its coding sequence between MA0817 and MA0816 is shown in black. (*A*) Sequencing of genomic DNA (gDNA) and cDNA confirmed in-frame amber codon in the *thg1* gene. (*B*) A secondary structure prediction of MaThg1 places the Pyl residue within an α-helix.

the otherwise deficient $tRNA^{His}$ by posttranscriptional GMP addition to the 5′ end of $tRNA^{His}$ (22–24). Here, we present the characterization of MaThg1 as a Pyl protein, and discuss the possibility that UAG codons in the *Methanosarcinaceae* may be translated by Pyl-$tRNA^{Pyl}$, leading to the use of Pyl as a "normal" amino acid.

Results

In-Frame Amber Codon in MaThg1 Is Translated as Pyl. Bioinformatic analysis of the *M. acetivorans* genome revealed that the single $tRNA^{His}$ gene encodes a G-1 residue. Thus, it is likely that archaeal RNase P will cleave aberrantly and leave the G-1 residue intact, and an active Thg1 enzyme may not be needed. The analysis also exposed a Thg1 sequence made up of two ORFs (MA0816 and MA0817) that appeared to be connected by an in-frame amber codon. It had been reported that this genome includes a number of genes that encode in-frame TAG codons (10). We PCR amplified the putative *M. acetivorans* gene; sequencing of independent clones confirmed the presence of an in-frame TAG codon at amino acid position 142 (Fig. 1*A*). cDNA synthesis from RNA of *M. acetivorans* cells, and successful PCR amplification showed that a transcript of this MaThg1 gene is expressed. *M. mazei* and *Methanosarcina barkeri*, close relatives of *M. acetivorans*, possess TGG (encoding Trp) at the homologous position in their *thg1* genes. We considered that adenosine to inosine RNA editing could convert UAG to UIG; this codon would be read as UGG. Although A-to-I editing has yet to be described in archaea (25), we eliminated this possibility by sequencing three independent cDNA samples, all showing the presence of the UAG codon (Fig. 1*A*).

To determine whether the MaThg1 UAG codon is translated as Pyl, we constructed a N-terminal His_6-tagged Thg1 variant under the control of the constitutive promoter of the methyl coenzyme M reductase gene (PmcrB), which has been used earlier in translational studies of Pyl-proteins in *M. acetivorans* in vivo (7). The *E. coli*/*Methanosarcina* shuttle vector backbone allowed integration of the His-tagged variant in single copy into the *M. acetivorans* chromosome. Immunoblot analysis of this His_6-tagged MaThg1 protein with anti-His antibody revealed read-through of the in-frame amber codon and translation to the next genuine stop codon (UGA); thus full-length Thg1 is produced in *M. acetivorans.* The purified recombinant mutant (rTrp) with Trp encoded at position 142 that was expressed in *E. coli* is of the same size as the amber codon read-through product of the His-tagged Thg1 (Pyl) expressed in *M. acetivorans* (Fig. 2*A*).

The His_6-MaThg1 protein was purified from *M. acetivorans* cells by talon column chromatography and subjected to mass spectroscopic analysis. The calculated average mass for the intact MaThg1 protein containing a Lys-142 is 30,184.7 Daltons, but if a Pyl residue is present the calculated intact protein mass is approximately 109 Da greater at 30,293.6 Da. The mass of the entire protein as determined by ESI-TOF was measured as 30,294.9 ± 3 Da, which corresponds to the theoretical mass of the intact His-tagged Pyl-containing MaThg1 (Fig. 2*B*). Mascot analyses of the MS/MS spectra of the peptides generated from the tryptic digestion of MaThg1 were found to produce multiple peptides that were 109 mass units greater than that calculated for a hypothetical peptide containing Lys at position 142. The MS/MS spectrum of the peptide (Fig. 2*C*), Asn-Phe-Val-Ala-Ser-Pyl-Gly-Tyr-Tyr-Ala-Leu-Arg, demonstrated that the added mass corresponded to the residue at position 142. These data are consistent with the incorporation of a pyrrolysine residue at the site of the UAG codon.

UAG Read-Through Is Independent of the Carbon Source for Growth. Previous studies suggested that Pyl is only essential for growth on methylamines and a *pylT* knockout reveals no phenotype for growth on methanol (5). Incorporation of Pyl might, therefore, be favored on TMA as opposed to methanol. This could result in loss of read-through or down regulation of Pyl-containing proteins. We could show that, in agreement with previous experiments (7, 8), no difference in read-through efficiency on methanol or TMA is observed (Fig. 3*A*, lanes 1 and 2), indicating

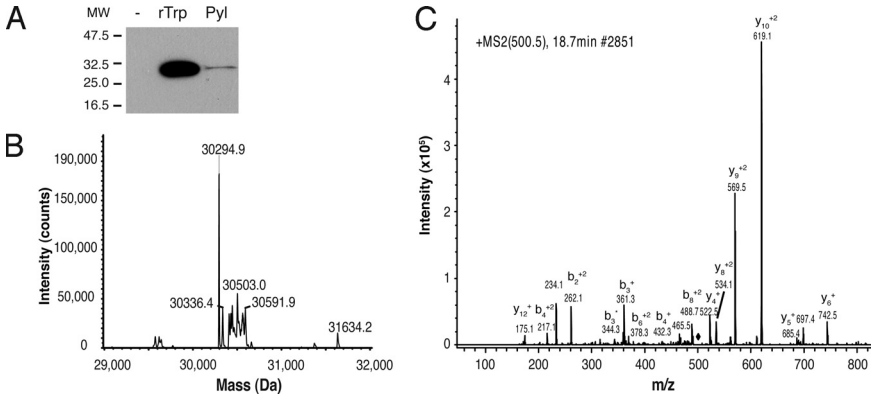

Fig. 2. UAG translation in Thg1 in *M. acetivorans*. (*A*) SDS-solubilized MaThg1 mutant Pyl142Trp purified from *E. coli* (rTrp) and extracts of *M. acetivorans* expressing His-tagged Thg1 (Pyl) were analyzed by anti-His immunoblot. (*B*) Deconvolution of Thg1 protein purified from *M. acetivorans* analyzed by ESI-TOF MS. (*C*) MS/MS spectrum of peptide 137–148 with a Pyl at position 142 (*m/z* of 500.5^{3+}). The location where the precursor ion is indicated by a solid diamond. Observed fragment ions (*b* and *y*) were consistent with the following McLafferty diagram: N F/V A S/O G Y Y/A/L/R where O indicates a Pyl residue.

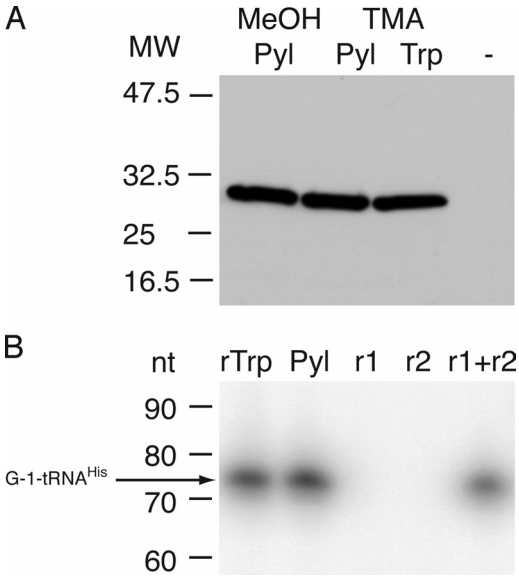

Fig. 3. Role of Pyl in MaThg1. (*A*) SDS-solubilized extracts of *M. acetivorans* expressing His-tagged Thg1 (Pyl) and His-tagged Thg1 mutant Pyl142Trp (Trp) grown on either methanol (MeOH) or TMA were analyzed by anti-His immunoblot. *M. acetivorans* C2A cell extracts were used as control (−). (*B*) Recombinant MaThg1 mutant Pyl142Trp (rTrp) and the two Thg1 halves MA0817 (r1) and MA0816 (r2) were produced and purified from *E. coli*. The Pyl containing MaThg1 (Pyl) variant was purified from *M. acetivorans*. MA0817 and MA0816 were assayed separately and together (r1+r2).

that incorporation of Pyl into proteins is independent of growth substrate and perhaps entirely unregulated. We did not observe formation of any truncated gene product under the used conditions. While it is possible that truncated protein is quickly degraded, the Pyl142Trp mutant shows comparable production of full-length protein to the Pyl-containing version (Fig. 3*A*). Furthermore, we were able to successfully produce a recombinant version of the truncated MaThg1 (MA0817 ending with a TAA instead of TAG) in *E. coli*.

MaThg1 Is Enzymatically Active. To investigate whether MaThg1 is expressed in vivo, we assayed cell extracts from wild-type *M. acetivorans* C2A grown on either methanol or TMA for enzyme activity. A $tRNA^{His}$ (Fig. 4*A*) substrate lacking the G-1 residue and [α-^{32}P]GTP were added to specially prepared cell-free extracts (see *Materials and Methods*). After incubation, the reaction mixture was subjected to gel electrophoresis and the radioactive band corresponding to [α-^{32}P]G-1-$tRNA^{His}$ was visualized by phosphorimaging. Thg1 activity could be shown in extracts from cells grown on either TMA or methanol (Fig. 4*B*), indicating that MaThg1 expression and read-through are not differentially regulated on methanol versus TMA in vivo.

MaThg1 was tested for enzymatic function on different $tRNA^{His}$ΔG-1 substrates. MaThg1 containing the Pyl residue was over expressed and purified from *M. acetivorans*. Activity assays were carried out using substrate $tRNA^{His}$ΔG-1 from *Methanothermobacter thermautotrophicus*, *Pyrobaculum aerophilum*, *E. coli*, and *M. acetivorans*. All four $tRNA^{His}$ species share significant sequence identity (Fig. 4*C*) and were competent substrates for MaThg1 (Fig. 4*B*).

In the absence of a Thg1 crystal structure, the role of the amino acid (Pyl) in position 142 is unknown, but a secondary structure prediction (Fig. 1*B*) places the residue within an α-helical context. The complete Thg1 protein sequence was initially annotated as two independent ORFs (MA0816 and MA0817), each encoding approximately half of the full-length Thg1 protein (Fig. 1*B*). The Pyl-containing peptide of 10 amino acids was annotated as untranslated sequence. To evaluate whether Pyl plays a role in the catalytic activity of the enzyme, we mutated residue 142 to Trp. The MaThg1 Pyl142Trp mutant was expressed in *E. coli* and purified. In addition, His_6-tagged proteins corresponding to the MA0816 and MA0817 ORFs were also separately expressed in *E. coli*. Lastly, Pyl-containing His_6-tagged MaThg1 was produced in *M. acetivorans*. Each variant was assayed for Thg1 activity, and the activity of an equimolar mixture of the MA0816 and MA0817 gene products was also determined. Both variants, the Pyl-containing and the Trp-containing protein, displayed activity in vitro (Fig. 3*B*). Neither MA0816 nor MA0817 gene product displayed any detectable activity on their own, and both proteins proved to be instable after purification. Protein stability was increased when both halves were combined immediately after elution. The protein complex of the two MaThg1 halves revealed Thg1 activity in trans, which shows that MaThg1 can even support activity upon deletion of the Pyl-containing peptide (Fig. 3*B*). As know from

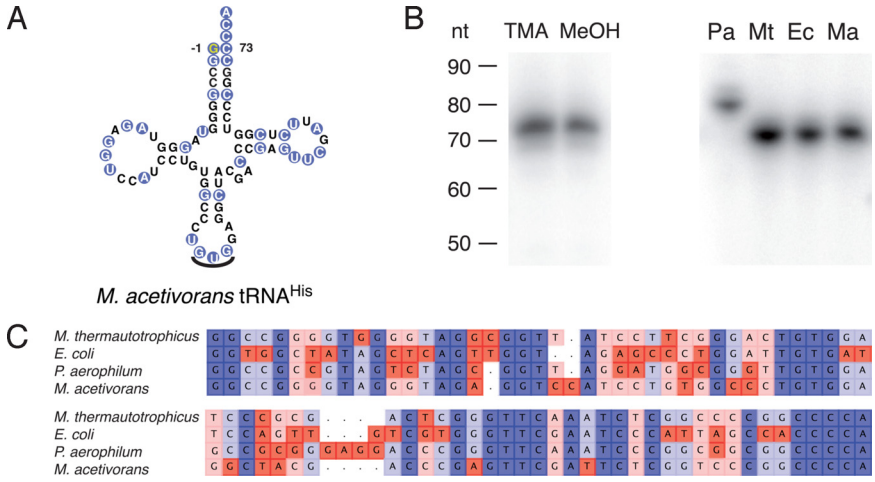

Fig. 4. Enzymatic activity and substrate specificity of MaThg1. (*A*) *M. acetivorans* $tRNA^{His}$. Residues conserved among *M. thermautotrophicus*, *P. aerophilum* and *E. coli* $tRNA^{His}$ are marked in blue. The G-1 residue is highlighted. (*B*) Activity of Thg1 was assayed by addition of *M. acetivorans* $tRNA^{His}$ΔG-1 and [α-^{32}P]GTP to cell-free extracts from *M. acetivorans* grown on TMA or methanol (MeOH). $tRNA^{His}$ΔG-1 from *P. aerophilum* (78 nt), *M. thermautotrophicus* (75 nt), *E. coli* (75 nt), and from *M. acetivorans* (75 nt) were assayed with Thg1 purified from *M. acetivorans*. (*C*) Sequence alignment of $tRNA^{His}$ from *M. thermautotrophicus*, *E. coli*, *P. aerophilum*, and *M. acetivorans*.

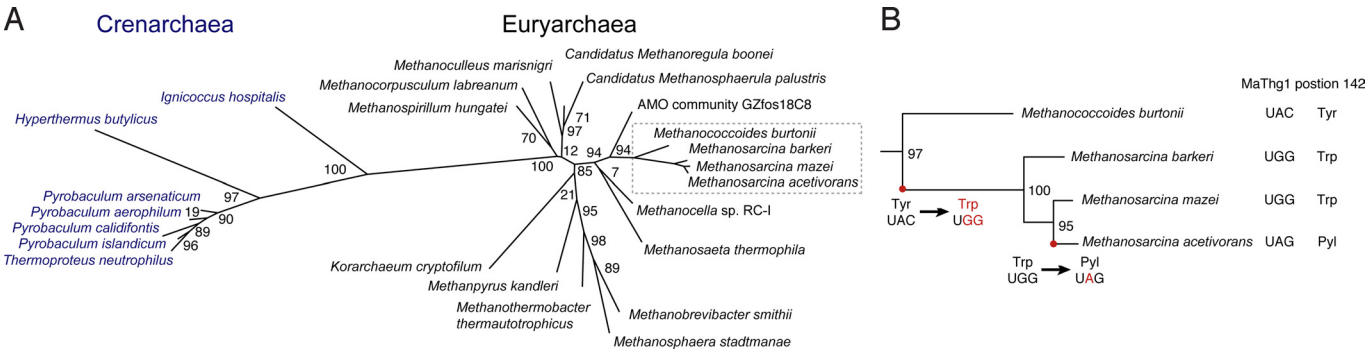

Fig. 5. Phylogeny of archaeal Thg1. All archaea known to encode Thg1 are included in the maximum likelihood phylogeny (*A*). Magnification (*B*) of the *Methanosarcinaceae* Thg1 phylogenetic pattern (gray box in *A*). The codon and amino acid at the position homologous to Pyl142 in MaThg1 are shown adjacent to the branches. The most parsimonious sequence changes at position 142 are shown beneath. Bootstrap support is given at each node.

other studies (26) there are dispensable regions in proteins that can be deleted, yet the protein fragments can be functional in trans.

Evolution of Thg1 in Archaea. Although Thg1 is found only in a minority of the archaea, the enzyme has homologs in both major archaeal phyla and also representatives in many methanogens. Phylogeny of these Thg1 sequences (Fig. 5*A*) indicates a deep divide between the crenarchaeal and euryarchaeal versions. Thg1-containing archaea group essentially in accord with accepted taxonomy; this indicates that vertical inheritance and gene loss best explain archaeal Thg1 evolution. Thg1 is only found in one-third of sequenced archaeal genomes. Seven of the 24 archaeal *thg1* genes are likely not essential, since the G-1 of $tRNA^{His}$ is genome-encoded (Fig. S1). This situation is further evidence that Thg1 is becoming a dispensable gene in the archaea, even though its role may still be critical in some species.

A particularly illuminating example is the *Ignicoccus hospitalis* Thg1 (IhThg1), which is interrupted by a TGA stop codon after codon 162 (homologous to MaThg1 position 183). Since no TGA codon read-through system is known to exist in *I. hospitalis*, the IhThg1 gene must lead to a Thg1 protein truncated by 40–60 amino acid residues compared to its archaeal counterparts. As the G-1 of *I. hospitalis* $tRNA^{His}$ is genome-encoded, the truncated IhThg1 likely represents a transitional form and may support what appears to be a progressive evolutionary loss of this enzymatic activity in the archaea.

Evolutionary Appearance of Pyl in Thg1. Examination of the evolution of codon 142 (MaThg1 numbering) in the *Methanosarcinaceae* family suggests the most parsimonious pathway for Pyl introduction into Thg1 (Fig. 5*B*). Most euryarchaea encode a Tyr (UAC codon) in this position, as does *Methanococcoides burtonii*, the most distantly related *Methanosarcinaceae* to *M. acetivorans*. Because the next most distant relative in the family (*M. barkeri*) and the closest relative (*M. mazei*) to *M. acetivorans* encode a Trp (UGG codon) in this position, it is most likely that the UAC codon evolved to UGG before the divergence of the three *Methanosarcina* species. Even though all four species in the *Methanosarcinaceae* include the Pyl operon and express Pyl-proteins, only the MaThg1 includes the amber codon calling for Pyl; the phylogenetic tree indicates that Pyl was introduced recently into the MaThg1 by a simple point mutation. This appears to have occurred without the coevolution of a downstream recoding signal, since a stem loop structure similar to that proposed to recode UAG from stop to Pyl for MtmB1 (7, 27) cannot be detected.

Discussion

Pyl as a Normal Amino Acid. Although, in *M. acetivorans*, interruption of a gene with the amber codon still leads to full-length read-through, having Pyl in particular at position 142 seems likely the result of neutral evolution that did not confer a selective advantage to Thg1 in *M. acetivorans*. While it may be common place for the canonical amino acids to occupy nonessential positions in proteins, filling replaceable roles, the observation of Pyl in such a position has not been previously characterized. The finding shows a broader biochemical role for Pyl and represents a departure from the initial view of Pyl as a critical active site constituent possibly selectively evolved for the enzymes that enable growth on methylamines. Even though Pyl has not taken on functional significance in Thg1, the fact that Pyl can appear essentially by chance in a protein means that in Pyl-decoding organisms the 22nd amino acid is not only selectively maintained in certain methyltransferases. Rather, Pyl is also expanding the evolutionary potential of the proteome by appearing in enzymes that have not previously experimented with the functional properties of this rare and unusual amino acid.

Expanding the Pyl Proteome. The meaning of UAG codons in mRNAs of *M. acetivorans* and other Pyl-decoding organisms is at present unclear. In these organisms, UAG may be interpreted as Pyl and translationally read-through with near 100% efficiency as is the case for translation of canonical amino acids. Using anti-His immunoblots we demonstrated conclusively that the N-terminal His-tagged Thg1 protein is translated to full-length in *M. acetivorans*, and that Pyl is incorporated in response to its in-frame amber codon with no detectable truncated protein produced. MaThg1 is only the second native *Methanosarcina* Pyl-protein for which the efficiency of translational read-through has been visualized. The first example, *M. barkeri* monomethylamine methyltransferase (MtmB1) visualized by anti-MtmB1 immunoblot, displayed similarly efficient translation of the Pyl codon (7). No detectable truncated MtmB1 protein was produced, even though replacement of the in-frame amber codon with TAA yielded significant truncated protein. Although, the meaning of UAG codons in Pyl-decoding organisms is not yet clearly known, the above results show that native *Methanosarcina* sequences with in-frame amber codons are indeed read through with remarkably high efficiency.

Thg1 and the methylamine methyltransferases may only be examples of many more, potentially hundreds (10, 28) of Pyl-proteins. For instance, *M. acetivorans* has >200 genes with in-frame UAG codons. There has been a general assumption that UAG must, at least at times, act as a terminator in Pyl-decoding organisms (10, 28), but incorporation of Pyl may not require a recoding apparatus and UAG may indeed encode

Pyl wherever it appears in transcripts of Pyl-decoding organisms. Future experiments with other putative Pyl-proteins will define the extent of the Pyl proteome, and may reveal how new amino acids invade the genetic code and propagate through genomes.

Materials and Methods

Oligonucleotide synthesis was carried out by the Keck Foundation Biotechnology Research Laboratory at Yale University and DNA sequencing was performed at DNA Analysis Facility on Science Hill at Yale University. [α-^{32}P]GTP (3,000 Ci/mmol) was from GE Healthcare.

Bacterial Strains and Cloning. For expression of recombinant proteins in *E. coli*, *M. acetivorans* full-length *thg1* was amplified from genomic DNA and cloned between *Eco*RI and *Sal*I restriction sites in pET20b vector (Novagen) with a C-terminal His-tag. A point mutation to Trp was introduced for the Pyl codon at position 142 with the QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions to enable read-through for expression in *E. coli*. Furthermore, since *M. acetivorans* Thg1 was originally annotated as two ORFs MA0816 (Met-152–243) and MA0817 (Met-1–141), both halves were cloned into separate plasmids. For expression of Thg1 variants in *M. acetivorans*, the N-terminal His-tagged Thg1 gene from a pET15b construct (PCR amplified *thg1* from genomic DNA with primers allowing cloning into the *Nde*I and *Xho*I sites) was fused to the *M. barkeri* PmcrB promoter [PCR amplified from pMP44 (29)] via overlap extension PCR (30) with *Kpn*I restriction sites at the 5′ and 3′ ends. The constructs where then cloned into pMP42 (29) at the *Kpn*I site. A point mutation to Trp was introduced for the Pyl codon at position 142 with the QuikChange site-directed mutagenesis kit (Stratagene). All clones and mutations were verified by DNA sequencing.

cDNA Synthesis. RNA was harvested form *M. acetivorans* C2A cells grown to stationary phase with the RNeasy kit (Qiagen). According to the manufacturer's instructions, cDNA was synthesized using random hexameric primers from 1 μg of total RNA with the SuperScript II reverse transcriptase (Invitrogen). cDNA was then amplified with Taq DNA polymerase (New England Biolabs) and specific forward (5′-ATGAAAACCCGGGAAATATATGCT-3′) and reverse (5′-TCAGTTTCTATTAATAAGTTTCTCTAAAAACG-3′) primers for MaThg1.

***Methanosarcina* Growth and Transformation.** All *M. acetivorans* manipulations were carried out under strictly anaerobic conditions in an anaerobic chamber (Coy Laboratory Products Inc.). *M. acetivorans* was grown as single cells in disaggregating high salt media (prepared as described in refs. 31 and 32) with 40 mM sodium acetate and either 50 mM trimethylamine (TMA) or 0.5% methanol by volume. The liposome mediated transformation protocol was applied as described in ref. 33.

Analysis of UAG Translation by Immunoblotting. For analysis of UAG translation, 1 mL *M. acetivorans* cells carrying a His-tagged copy of Thg1 (with either a UAG or UGG codon at amino acid position 142) integrated into the *hpt* locus were harvested in the logarithmic growth phase at an $A_{600} = 0.75$. Cells were suspended in SDS loading buffer and proteins were separated by SDS/PAGE and blotted onto a nitrocellulose membrane. The *thg1* gene products in equivalent amounts of lysates were then analyzed by using mouse anti-His antibody (GE Healthcare). Broad-range Prestained Protein Marker (New England Biolabs) was used as molecular size standard.

Protein Purification. For purification of a Pyl containing MaThg1, 2 L of the *M. acetivorans* strain carrying a His-tagged copy of Thg1 integrated into the *hpt* locus were inoculated with 50 mL preculture and grown for 4 days. Cells were harvested and lysed by addition of 20 mL distilled water. After 5 min, 5 mL of 5× Thg1 buffer were added. After sonication and centrifugation, His-tagged MaThg1 was purified via Talon affinity chromatography (Clontech) according to the manufacturer's instructions. A 2-L culture yielded approximately 250 μg His-tagged MaThg1.

Plasmids containing Thg1 halves or the Pyl142Trp mutant of MaThg1 in pET20b were transformed into *E. coli* BL21 (DE3) codon plus (Stratagene). A preculture was used to inoculate 1 L LB broth with 100 μg/mL ampicillin and 34 μg/mL chloramphenicol. The cells were grown at 37 °C until $A_{600} = 0.6$, and protein expression was induced by addition of IPTG to a final concentration of 250 μM and continued at 20 °C for 14–16 h. The cells were pelleted and resuspended in 1× Thg1 buffer [50 mM Tris·HCl (pH 7.0), 200 mM NaCl, 20 mM $MgCl_2$, 5% glycerol, and 3 mM DTT]. After sonication and centrifugation, the cell lysates were applied to Ni-NTA metal affinity resin (Qiagen) and purified according to the manufacturer's instructions. The eluted enzymes were dialyzed into 1× Thg1 buffer. SDS/PAGE electrophoresis followed by staining with Coomassie Blue revealed >90% purity.

MS TOF Anaylsis of Intact Thg1. Samples of MaThg1 protein purified as described above were analyzed by reverse phase liquid chromatography (LC) and electrospray ionization time-of-flight mass spectrometery (ESI-TOF MS). Purified MaThg1 was injected onto an 1100 Series Capillary LC System (Agilent Technologies) and separated on a PLRP-S RP column (0.5 × 150 mm, 5-μm particles, 1,000-Å pore size; Higgins Analytical) using a 10-min 15–90% FB gradient (FA = 0.1% formic acid, FB = CH_3CN, 0.1% formic acid) at a flow rate of 10 μL/min and analyzed online by a 6210 series ESI-TOF MS (Agilent Technologies). Mass spectra were acquired from 150 to 3200 *m/z*, one cycle/s and 10,000 transients per scan. A capillary voltage of 4,000 V was used with a gas temperature of 300 °C and a drying gas flow rate of 7.0 L/min. The acquired spectra were extracted and the protein spectra deconvoluted with Bioconfirm A 2.0 software (Agilent Technologies) using a mass range of 10 k to 75 kDa and a step mass (Da) of 1.01.

Ion Trap MS/MS and Data Analysis of Trypsin-Digested Thg1 Protein. MaThg1 protein samples were mixed with Trypsin Reaction Buffer and digested overnight at 37 °C with Trypsin (New England Biolabs) at a protein to protease ratio of 20:1. Online LC-MS/MS analyses of digested fractions using an Agilent 6330 Ion Trap mass spectrometer with an integrated C18 Chip/nanoESI interface were performed as described in ref. 34. The nanoESI-MS/MS data were analyzed with the Mascot (Version 2.2.04, Matrix Science Ltd.) search engine (35) using parameters described in ref. 34 with minor modifications. Data were searched against a database containing the MaThg1 protein sequence. The Mascot search criteria were set to allow two missed cleavages by a tryptic digest with all Unimod protein modifications permitted. Peptides scoring 40 or better and with a Significance Threshold (*P* value) of <0.01 were then manually examined.

tRNA Purification. The genes encoding *M. thermautotrophicus* $tRNA^{His}\Delta G$-1, *P. aerophilum* $tRNA^{His}\Delta G$-1, and *E. coli* $tRNA^{His}\Delta G$-1 with the preceding sequence of the T7 promoter were cloned into the pUC19 plasmid. Plasmid DNA was purified using the HISpeed Plasmid Maxi kit (Qiagen). The purified plasmid was digested with *Bst*NI for runoff transcription as described in ref. 36. The transcript was purified by electrophoresis on a 12% denaturing polyacrylamide gel. The tRNA transcripts were refolded at a concentration of 10 μM by heating for 5 min at 70 °C in buffer containing 10 mM Tris·HCl (pH 7.0), followed by addition of 5 mM $MgCl_2$ and immediate cooling on ice. An RNA oligonucleotide encoding *M. acetivorans* $tRNA^{His}\Delta G$-1 was obtained from Dharmacon RNAi Technologies. The 5′ end was phosphorylated according to manufacturer's instructions using T4 polynucleotide kinase (New England Biolabs).

Thg1 Assay. Activity of Thg1 proteins was assayed monitoring the incorporation [α-^{32}P]GTP into $tRNA^{His}\Delta G$-1 of as described in refs. 22 and 24. Reactions were incubated at 37 °C for 30 min to plateau level. All reactions were carried out in 1× Thg1 buffer (see above). Protein concentrations were adjusted to 0.1 mg/mL, tRNA was added to a final concentration of 0.2 mg/mL, giving an approximate 2:1 molar ratio of tRNA to enzyme. Thg1 halves were combined in equal stoichiometry and preincubated for 1 h at room temperature. Reactions analyzed via polyacrylamide gel electrophoresis containing 12% urea and visualized by phosphorimaging. The RNA Decade marker (Ambion) was used.

M. acetivorans cell extracts for assaying native Thg1 activity were prepared as following. Ten-milliliter cultures grown either on methanol or TMA were harvested at $A_{600} = 0.75$. Cells were broken by the addition of 200 μL distilled water. After 5 min 50 μL of 5× Thg1 buffer were added [250 mM Tris·HCl (pH 7.0), 1 mM NaCl, 100 mM $MgCl_2$, 25% glycerol, and 15 mM DTT]. Cell suspensions were incubated with RNaseA (Qiagen) and DNase I (New England Biolabs) for 1 h at 37 °C. RNaseA was inhibited by addition of Superase-In (Ambion) as described by the manufacturer. Cell extracts were centrifuged and the supernatant was subjected to activity assays as described below using *M. acetivorans* $tRNA^{His}\Delta G$-1.

Bioinformatic and Phylogenetic Analysis. The secondary structure of MaThg1 was predicted using Jpred (37). Sequences were downloaded from National Center for Biotechnology Information and from the Integrated Microbial Genomes database (38). Muscle (39) and the Multiseq alignment editor from VMD 1.8.7 (40) were used for sequence alignment and editing. Phyml (41) was used to determine the maximum likelihood phylogeny for the archaeal Thg1 sequences. The starting tree was generated with BioNJ, and the tree space was search with the SPR followed by the NNI algorithm. The JTT+Γ model with four rate categories for SPR and eight for NNI was applied. Likelihood parameters were initially estimated from the alignment.

ACKNOWLEDGMENTS. We thank Dan Su for helpful suggestions, and William W. Metcalf for providing *M. acetivorans* cells, shuttle vectors, and invaluable advice. We are grateful to Richard J. Roberts for encouragement. This work was supported by grants from the National Institute of General Medical Sciences (to D.S.) and the National Science Foundation (to D.S.). I.U.H. is a Postdoctoral Fellow of the Deutsche Forschungsgemeinschaft (HE5802/1-1).

1. Hao B, et al. (2002) A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* 296:1462–1466.
2. Polycarpo C, et al. (2004) An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc Natl Acad Sci USA* 101:12450–12454.
3. Blight SK, et al. (2004) Direct charging of $tRNA_{CUA}$ with pyrrolysine in vitro and in vivo. *Nature* 431:333–335.
4. Longstaff DG, et al. (2007) A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. *Proc Natl Acad Sci USA* 104:1021–1026.
5. Mahapatra A, et al. (2006) Characterization of a *Methanosarcina acetivorans* mutant unable to translate UAG as pyrrolysine. *Mol Microbiol* 59:56–66.
6. Soares JA, et al. (2005) The residue mass of L-pyrrolysine in three distinct methylamine methyltransferases. *J Biol Chem* 280:36962–36969.
7. Longstaff DG, Blight SK, Zhang L, Green-Church KB, Krzycki JA (2007) In vivo contextual requirements for UAG translation as pyrrolysine. *Mol Microbiol* 63:229–241.
8. Veit K, Ehlers C, Schmitz RA (2005) Effects of nitrogen and carbon sources on transcription of soluble methyltransferases in *Methanosarcina mazei* strain Go1. *J Bacteriol* 187:6147–6154.
9. Krzycki JA (2004) Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases. *Curr Opin Chem Biol* 8:484–491.
10. Zhang Y, Baranov PV, Atkins JF, Gladyshev VN (2005) Pyrrolysine and selenocysteine use dissimilar decoding strategies. *J Biol Chem* 280:20740–20751.
11. Ambrogelly A, Palioura S, Söll D (2007) Natural expansion of the genetic code. *Nat Chem Biol* 3:29–35.
12. Yoshizawa S, Böck A (2009) The many levels of control on bacterial selenoprotein synthesis. *Biochim Biophys Acta* 1790:1404–1414.
13. Namy O, et al. (2007) Adding pyrrolysine to the *Escherichia coli* genetic code. *FEBS Lett* 581:5282–5288.
14. Eggertsson G, Söll D (1988) Transfer ribonucleic acid-mediated suppression of termination codons in *Escherichia coli*. *Microbiol Rev* 52:354–374.
15. Beier H, Barciszewska M, Krupp G, Mitnacht R, Gross HJ (1984) UAG readthrough during TMV RNA translation: Isolation and sequence of two tRNAs with suppressor activity from tobacco plants. *EMBO J* 3:351–356.
16. Kavran JM, et al. (2007) Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc Natl Acad Sci USA* 104:11268–11273.
17. Altman S, et al. (1980) in Transfer RNA: Biological Aspects, eds Söll D, Abelson J, Schimmel P (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp 71–82.
18. Jühling F, et al. (2009) tRNAdb 2009: Compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* 37:D159–162.
19. Giege R, Sissler M, Florentz C (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res* 26:5017–5035.
20. Burkard U, Willis I, Söll D (1988) Processing of histidine transfer RNA precursors. Abnormal cleavage site for RNase P. *J Biol Chem* 263:2447–2451.
21. Orellana O, Cooley L, Söll D (1986) The additional guanylate at the 5′ terminus of *Escherichia coli* tRNAHis is the result of unusual processing by RNase P. *Mol Cell Biol* 6:525–529.
22. Cooley L, Appel B, Söll D (1982) Post-transcriptional nucleotide addition is responsible for the formation of the 5′ terminus of histidine tRNA. *Proc Natl Acad Sci USA* 79:6475–6479.
23. Pande S, Jahn D, Söll D (1991) Histidine tRNA guanylyltransferase from *Saccharomyces cerevisiae*. I. Purification and physical properties. *J Biol Chem* 266:22826–22831.
24. Gu W, Jackman JE, Lohan AJ, Gray MW, Phizicky EM (2003) $tRNA^{His}$ maturation: An essential yeast protein catalyzes addition of a guanine nucleotide to the 5′ end of $tRNA^{His}$. *Genes Dev* 17:2889–2901.
25. Conticello SG (2008) The AID/APOBEC family of nucleic acid mutators. *Genome Biol* 9:229.
26. Shiba K, Schimmel P (1992) Functional assembly of a randomly cleaved protein. *Proc Natl Acad Sci USA* 89:1880–1884.
27. Namy O, Rousset JP, Napthine S, Brierley I (2004) Reprogrammed genetic decoding in cellular gene expression. *Mol Cell* 13:157–168.
28. Deppenmeier U, et al. (2002) The genome of *Methanosarcina mazei*: Evidence for lateral gene transfer between bacteria and archaea. *J Mol Microbiol Biotechnol* 4:453–461.
29. Pritchett MA, Zhang JK, Metcalf WW (2004) Development of a markerless genetic exchange method for *Methanosarcina acetivorans* C2A and its use in construction of new genetic tools for methanogenic archaea. *Appl Environ Microbiol* 70:1425–1433.
30. Bose A, Pritchett MA, Rother M, Metcalf WW (2006) Differential regulation of the three methanol methyltransferase isozymes in *Methanosarcina acetivorans* C2A. *J Bacteriol* 188:7274–7283.
31. Metcalf WW, Zhang JK, Shi X, Wolfe RS (1996) Molecular, genetic, and biochemical characterization of the *ser*C gene of *Methanosarcina barkeri* Fusaro. *J Bacteriol* 178:5797–5802.
32. Sowers KR, Boone JE, Gunsalus RP (1993) Disaggregation of *Methanosarcina* spp. and growth as single cells at elevated osmolarity. *Appl Environ Microbiol* 59:3832–3839.
33. Metcalf WW, Zhang JK, Apolinario E, Sowers KR, Wolfe RS (1997) A genetic system for Archaea of the genus *Methanosarcina*: Liposome-mediated transformation and construction of shuttle vectors. *Proc Natl Acad Sci USA* 94:2626–2631.
34. Swaim CL, Anton BP, Sharma SS, Taron CH, Benner JS (2008) Physical and computational analysis of the yeast *Kluyveromyces lactis* secreted proteome. *Proteomics* 8:2714–2723.
35. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20:3551–3567.
36. Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* 15:8783–8798.
37. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* 36:W197–201.
38. Markowitz VM, Kyrpides NC (2007) Comparative genome analysis in the integrated microbial genomes (IMG) system. *Methods Mol Biol* 395:35–56.
39. Edgar RC (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
40. Roberts E, Eargle J, Wright D, Luthey-Schulten Z (2006) MultiSeq: Unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* 7:382.
41. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704.