Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA

Gayathri Srinivasan, Carey M. James, Joseph A. Krzycki*

Pyrrolysine is a lysine derivative encoded by the UAG codon in methylamine methyltransferase genes of *Methanosarcina barkeri*. Near a methyltransferase gene cluster is the *pylT* gene, which encodes an unusual transfer RNA (tRNA) with a CUA anticodon. The adjacent *pylS* gene encodes a class II aminoacyl-tRNA synthetase that charges the *pylT*-derived tRNA with lysine but is not closely related to known lysyl-tRNA synthetases. Homologs of *pylS* and *pylT* are found in a Gram-positive bacterium. Charging a $tRNA_{CUA}$ with lysine is a likely first step in translating UAG amber codons as pyrrolysine in certain methanogens. Our results indicate that pyrrolysine is the 22nd genetically encoded natural amino acid.

In *Methanosarcina* species, specific methyltransferases initiate methanogenesis and carbon assimilation from substrates such as trimethylamine (TMA), dimethylamine (DMA), or monomethylamine (MMA). The highly expressed, nonhomologous genes encoding these methyltransferases have in-frame UAG (amber) codons that do not stop translation during synthesis of the full-length proteins (*1*–*3*). Nearly identical copies of the DMA and MMA methyltransferase genes with conserved single in-frame amber codons occur in the same genome (*2*, *4*). There is no evidence for transcript editing (*2*, *4*), and, unlike many other stop codon readthrough events (*5*), readthrough of the amber codons is highly efficient (*3*). Amber serves as a sense codon within the methylamine methyltransferase genes (*3*), previously unknown in any other group of archaeal genes. In the accompanying manuscript, Hao *et al.* have shown that in intact MtmB the amber-encoded residue is pyrrolysine, whose structure is proposed as lysine with its epsilon nitrogen in amide linkage with (4*R*,5*R*)-4-substituted-pyrroline-5-carboxylate. Here, we describe a specialized $tRNA_{CUA}$ and lysyl-tRNA synthetase (LysRS) that underlie amber codon translation as pyrrolysine in certain methane-producing Archaea.

An unannotated gene, *pylT* (Fig. 1), whose predicted tRNA product has a CUA anticodon, was identified in the *Methanosarcina barkeri* Fusaro genomic database (GenBank accession number NC_002724) using tRNAScanSE (*6*, *7*). Sequencing of *M. barkeri* MS DNA (*7*) also revealed *pylT*, as well as the three following open reading frames, *pylS*, *pylB*, and *pylC* (GenBank AY064401). Northern blots of the RNA pool from MMA-grown *M. barkeri* MS revealed an RNA of the size expected for the $tRNA_{CUA}$ product of *pylT* (Fig. 1C). The predicted secondary structure of $tRNA_{CUA}$ has unusual properties compared with typical tRNAs (*8*). Even though the structure has the expected sizes for the acceptor, D, and T stems and D, T, and anticodon loops, the anticodon stem could form with six, rather than five, base pairs. This would constrain the variable loop to only three, rather than four, bases. However, if the anticodon stem has five base pairs, two bases are found between the D and anticodon stems, also atypical of most tRNA structures. The predicted secondary structure of $tRNA_{CUA}$ has only one base, rather than the typical two bases, between the acceptor and D stems. Many of the conserved bases in tRNAs are found, but not the almost universally conserved GG sequence in the D loop or the TψC sequence in the T loop.

A 4.2-kilobase transcript was detectable with probes for *pylT*, *pylS*, *pylB*, and *pylC* (Fig. 1D), indicating possible cotranscription of these genes. The *pylS* gene has a predicted product similar to the core catalytic domains of class II aminoacyl-tRNA synthetase (AARS) enzymes containing motifs 1 through 3 (*9*, *10*). An RPS-BLAST (*11*) search for conserved domains found in the Pfam database (*12*) maintained at the National Center for Biotechnology Information produced alignments of different portions of the predicted PylS sequence with class II AARS from three different subclasses (*13*, *14*). To test the activity of PylS as an AARS, the *M. barkeri pylS* gene was expressed in *Escherichia coli* (*7*) as a 49-kD protein with an NH_2-terminal hexahistidine tag, which was then isolated by nickel-affinity chromatography (Fig. 2A). The recombinant PylS was tested for LysRS activity, as pyrrolysine is a lysine derivative. PylS ligated [^{14}C]lysine,

Department of Microbiology, Ohio State University, Columbus, OH 43210, USA.

*To whom correspondence should be addressed. E-mail: Krzycki.1@osu.edu

References and Notes

1. N. P. Kristensen, in *The Hierarchy of Life. Molecules and Morphology in Phylogenetic Analysis*, B. Fernholm, K. Bremer, H. Jörnvall, Eds. (Elsevier, Amsterdam, 1989), pp. 295–306.
2. _____, in *The Insects of Australia*, CSIRO, Ed. (Melbourne Univ. Press, Carlton, ed. 2, 1991), vol. 1, chap. 5.
3. A. H. Staniczek, *Zool. Anz.* **239**, 147 (2000).
4. In June 2001, J. Marshall (Natural History Museum, London) showed O.Z. a male insect from Tanzania, which had been submitted for an opinion 16 years ago. Shortly afterward O.Z. received from F. Kernegger a male Baltic-amber insect (subsequently described as *Raptophasma kerneggeri*), whose close similarity to the Tanzania specimen was immediately obvious. In July 2001, O.Z. discovered in the unsorted alcohol collection of Phasmatodea in the Museum für Naturkunde (Berlin) an adult female of a similar insect from Namibia.
5. According to one school of thought among contemporary systematists, the naming of higher taxa that only contain a single genus is "empty formalism"; we accept the logical merits of this stand. Pragmatically we believe, however, that any recognized genus should be assigned to a "family" and an "order," because these categories play an important role in how biologists communicate and how biological knowledge is systematized.
6. The description includes many characters currently known only from a single specimen, because genitalic characters can only be observed in a specimen of the respective sex, muscle characters can only be observed in the ethanol-preserved female, and only a few characters can be judged in the *Raptophasma* fossils.
7. O. Zompro, *Mitt. Geol.-Pal. Inst. Hamburg* **85**, 229 (2001).
8. A. Arillo, V. M. Ortuño, A. Nel, *Bull. Soc. Entomol. Fr.* **102**, 11 (1997).
9. W. C. Wheeler, M. Whiting, Q. D. Wheeler, J. M. Carpenter, *Cladistics* **17**, 113 (2001).
10. R. G. Beutel, S. N. Gorb, *J. Zool. Syst. Evol. Res.* **39**, 177 (2001).
11. E. H. Tilgner, T. G. Kiselyova, J. V. McHugh, *Dtsch. Entomol. Z.* **46**, 149 (1999).
12. J. Sellick, *Ital. J. Zool.* **64**, 97 (1997).
13. K.-D. Klass, *Zool. Anz.* **239**, 231 (2000).
14. _____, *Dtsch. Entomol. Z.* **46**, 3 (1999).
15. E. M. Walker, *Ann. Ent. Soc. Am.* **46**, 681 (1943).
16. K.-D. Klass, *Zool. J. Linn. Soc.* **131**, 251 (2001).
17. E. M. Walker, *Can. J. Res.* **27**, 309 (1949).
18. K.-D. Klass, *Zool. Anz.* **237**, 15 (1998).
19. K. Günther, *Jena Z. Naturwiss.* **68**, 403 (1933).
20. Since the completion of the work described here, additional (including live) material of further mantophasmatodean taxa has been observed and collected in Namibia by O.Z. and colleagues during an international expedition supported by the Max-Planck-Institut für Limnologie (Plön, Germany), the National Museum of Namibia (Windhoek, Namibia), and Conservation International (Washington, D.C.).
21. L. Chopard, *La Biologie des Orthoptères* (Lechevalier, Paris, 1938).
22. H. E. Hinton, *Biology of Insect Eggs* (Pergamon, Oxford, 1981).
23. We thank the Museum für Naturkunde, Humboldt-University, Berlin (M. Ohl, I. Dorandt), the Natural History Museum, London (J. Marshall), and the Zoological Museum, University of Lund (R. Danielsson) for assistance and permission to study the *Mantophasma* specimens; G. Brovad and R. Meier (Zoological Museum, Copenhagen) for help in photographic documentation; and O. Kraus (Zoologisches Institut und Zoologisches Museum, Universität Hamburg), W. Dohle (Institut für Zoologie, Freie Universität Berlin), S. Golovatch (Russian Academy of Sciences, Moscow), H. Enghoff (Zoological Museum, Copenhagen), and T. L. Erwin (Smithsonian Institution, Washington, D.C.) for comments on the manuscript. This publication is dedicated to Professor W. J. Junk (Max-Planck-Institut für Limnologie, Plön).

Supporting Online Material
www.sciencemag.org/cgi/content/full/1069397/DC1
SOM text
References

27 December 2001; accepted 10 April 2002
Published online 18 April 2002;
10.1126/science.1069397
Include this information when citing this paper.

but not [^{3}H]phenylalanine or [^{14}C]histidine, to tRNA in the cellular pool isolated from *M. barkeri* (*7*, *15*). Addition of [^{12}C]lysine, but not the other 19 amino acids, diluted [^{14}C]lysine ligation to tRNA (Fig. 2B). A protein with LysRS activity could not be isolated by nickel-affinity chromatography of extracts of *E. coli* not carrying *pylS*.

In the presence of a saturating level of the tRNA pool from *M. barkeri* , the apparent K_m of PylS for lysine was 2.2 μM, and the apparent k_{cat} was 1.6 min^{-1}. Most measured K_m values for class II LysRS are equivalent or higher, but k_{cat} values typically have ranged from 12 to 7800 min^{-1} (*16*). The k_{cat} value is closer to those for recombinant class I LysRS from *Borrelia burgdorferi* and *Methanococcus maripaludis* that are 4 min^{-1} and 47 min^{-1}, respectively (*17*). To test whether PylS could use $tRNA_{CUA}$ as a substrate, an 89–base pair double-stranded oligonucleotide with a T7 polymerase promoter spaced in front of the *pylT* gene was used as a template to produce a $tRNA_{CUA}$ transcript. PylS charged $tRNA_{CUA}$ with lysine at a rate of 0.3 pmol/min per pmol PylS. Controls also indicated this reaction was specific for lysine (Fig. 2C).

LysRS enzymes are the only type of AARSs with nonhomologous representatives that are either class I or class II enzymes (*18*). PylS adds to this diversity, because its sequence does not indicate a ready grouping with the known class II LysRS, all of which are in subclass IIb (*13*). An alignment of the catalytic core of PylS with other class II LysRS (Fig. 3A) reveals that PylS contains those residues binding Mg–adenosine triphosphate (MgATP) in class IIb LysRS, but not those binding solely lysine (*19*). BlastP alignments with the PylS sequence and the NH_2-terminal tRNA binding domain of LysRS from different sources did not produce any significant alignments. If the oligosaccharide/oligonucleotide binding (OB) fold responsible for binding the anticodon of the tRNA in known class II LysRS (*20*) is present in PylS, the primary sequence is highly divergent. PylS may represent a new subclass of class II AARS and also represents the third LysRS identified in *M. barkeri*. Class I and class II LysRS genes (GenBank AF337056 and AF337055, respectively), which produce active LysRS enzymes (*21*), have also been identified.

A search of the available genome sequences revealed an *mttB* homolog, the gene encoding TMA methyltransferase in *M. barkeri* (*2*), in the Gram-positive bacterium *Desulfitobacterium hafniense*. The *mttB* genes from both organisms have corresponding single in-frame amber codons. The genome also encodes homologs of the *M. barkeri* $tRNA_{CUA}$ and its cognate LysRS (*22*). The bacterial $tRNA_{CUA}$ shares the unusual features of the *M. barkeri* $tRNA_{CUA}$, including the ability to form a six–base pair acceptor stem, as well as one base between the D and acceptor stems (fig. S1). However, in this bacterium, two distinct reading frames encode homologs of the NH_2- and COOH-termini of PylS (figs. S1 and S2 and Fig. 3). The *D. hafniense pylSc* gene follows the *pylT*

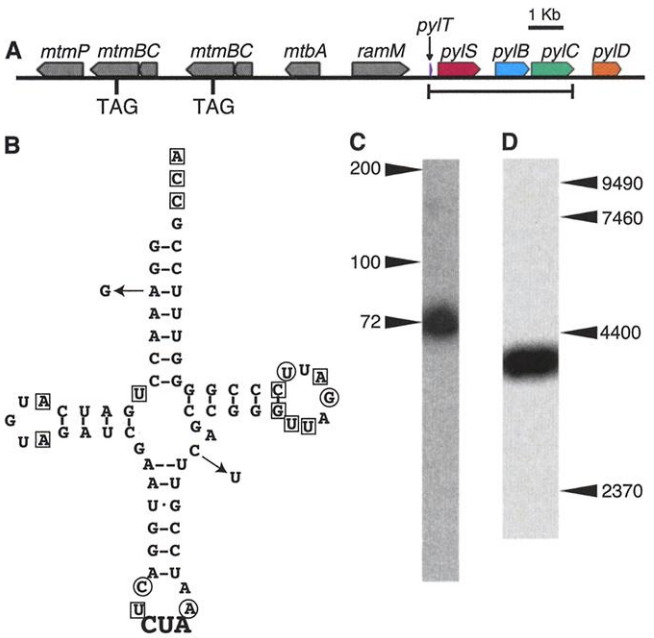

Fig. 1. The *pyl* gene cluster and RNA products in *M. barkeri*. (**A**) Location of the *pyl* cluster in *M. barkeri* Fusaro. The genes were found on contig 1956, GenBank NC_002724. The functions of the various *mtm* genes are in Burke *et al.* (*1*). The *ramM* gene was recently found to be involved in activation of methylamine methyltransferase corrinoid proteins (*28*). The location of in-frame amber codons (TAG) in each *mtmB* gene copy is indicated. The line underneath the *pylT* gene cluster indicates the corresponding sequence obtained from *M. barkeri* MS. (**B**) The deduced secondary structure of the amber suppressing tRNA from *M.* barkeri Fusaro *pylT*. Arrows indicate base changes found in *M. barkeri* MS *pylT* gene product. Bases that are highly conserved (boxed) or semiconserved (circled) in most tRNAs are indicated. The potential sixth base pair in the anticodon stem is denoted by two dashes between the bases. (**C**) Northern blot of 4 μg of the total tRNA pool from *M. barkeri* MS. The RNA was electrophoresed through an 8% polyacrylamide-8 M urea gel, then blotted onto Nytran membrane (Schleicher and Schuell, Keene, NH) before probing. (**D**) Northern blot of total RNA pool made from a 1% agarose electrophoretic gel showing larger transcript from *pylT* gene cluster. Both northern blots were made by probing with a 72-base synthetic DNA oligonucleotide probe complementary to the predicted RNA transcribed from *pylT* following the techniques in (*2*). To the side of each blot, the migration of size standards of indicated base sizes is shown.

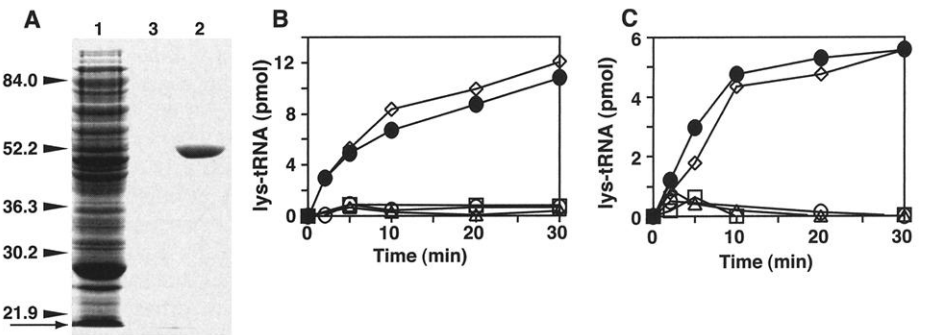

Fig. 2. LysRS activity of the *pylS* gene product from *M. barkeri* MS. (**A**) Coomassie-stained SDS-polyacrylamide gel (*29*) of (lane 1), cell extract (15 μg protein) of *E. coli* transformed with pXRS carrying *pylS* (*7*); (lane 2), recombinant purified PylS (5 μg protein) found in imidazole eluate of the HiTrap column of the same *E. coli* cell extract; and (lane 3), an equivalent volume of the imidazole eluate fraction, but from the same *E. coli* strain transformed with plasmid lacking *pylS* (*7*). Numbers to left of gel indicate molecular mass of marker proteins in kDa, arrow at bottom indicates location of the dye front. (**B**) Charging of tRNA in the total tRNA fraction from *M. barkeri* MS using [^{14}C]lysine. The complete reaction mixture (100 μl volume) contained 20 nM PylS, 50 mM KCl, 10 mM $MgCl_2$, 5 mM ATP, 5 mM dithiothreitol (DTT), 20 μM UL-^{14}C-labeled lysine (717 dpm/pmol) in 100 mM Hepes buffer, pH 7.2, with 330 μg crude *M. barkeri* tRNA preparation. At the indicated timepoints [^{14}C]lys-tRNA formation was assayed by filter assay as described in (*18*). Reactions illustrated are complete (diamonds), complete minus PylS (open circles), minus tRNA (open squares), plus 1 mM [^{12}C]lysine (open triangles), or plus a mixture of 1 mM each of the 20 canonical nonradioactive amino acids except lysine (filled circles). To illustrate isotope dilution of the product, the specific activity of 717 dpm/pmol lysine was used to calculate pmol lys-tRNA formed in all reactions. (**C**) Charging of $tRNA_{CUA}$ with lysine. The conditions and symbols are the same as in the previous figure, but with cellular tRNA replaced with 40 μM in vitro transcribed $tRNA_{CUA}$.

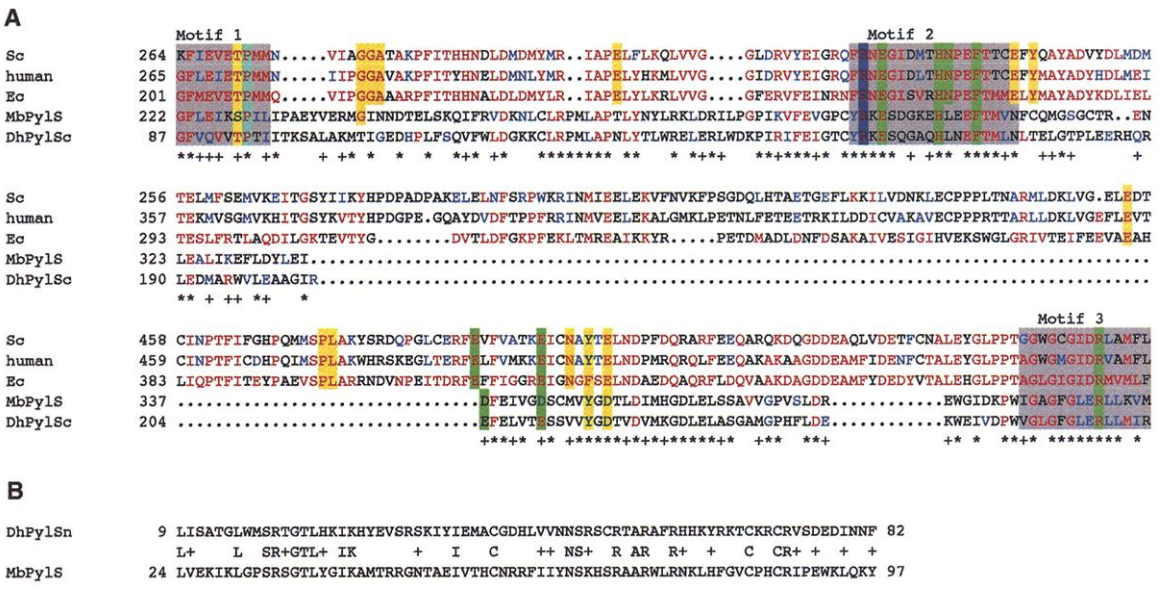

Fig. 3. Sequence alignments with PylS. (**A**) Alignment by the program ClustalW (*30*) using the Blossum62 matrix (*31*) of the core catalytic domains (spanning motif 1 to 3) of LysRS from *E. coli* (Ec, GenBankaccession number AAG58018.1, LysS), human (AAH04132.1), and *Saccharomyces cerevisiae* (Sc, NP_010322.1), as well as PylS from *M. barkeri* MS (MbPylS) and PylSc from *Desulfitobacterium hafniense* (dhPylSc). The blue letters indicate positive scores in the Blosum matrix; the red letters indicate identical residues, as individual sequences are compared with the Ec sequence as reference. In order to highlight similarity between only PylS and PylSc, conserved (+) and identical (*) residues between those two sequences are indicated below the alignment. Residues in motif 1, 2, or 3 are indicated by gray shading. Other shadings of residues indicate conserved resi-dues involved in: dimerization(light blue); lysine binding (yellow); ATP binding (green); or both lysine and ATP binding (dark blue) as seen from the *E. coli* LysRS crystal structures (*19*). (**B**) BLASTP alignment of portions of the predicted NH_2-terminal sequence of PylS from *M. barkeri* MS with a portion of *D. hafniense* PylSn (dhPylSn).

gene and encodes a predicted product homologous to the COOH-terminus of PylS, containing the core catalytic domain of the protein. A putative protein homologous to the NH_2-terminus of PylS is encoded by a short open reading frame (*pylSn*) found downstream of *pylSc*. This apparent splitting of the NH_2-terminus of PylS from its catalytic domain in *D. hafniense* is reminiscent of the encoding of the two domains of methionyl-tRNA synthetase (*23*) by separate genes in certain Eucarya and Crenarchaeota (*24*).

Three predicted *D. hafniense* genes between *pylSc* and *pylSn* are homologs of the *M. barkeri pylB*, *pylC*, and *pylD* genes (figs. S1 and S2). The presence of these genes near *pylT* in two phylogenetically distinct organisms may reflect conserved gene function associated with decoding amber codons. A bioinformatics approach suggests potential roles for *pyl* genes in the synthesis of pyrrolysine (*25*).

The presence of $tRNA_{CUA}$ and an unusual cotranscribed cognate LysRS indicates a mechanism in which pyrrolysine is inserted into MtmB during translation. Because PylS acts as a high-affinity LysRS, lys-$tRNA_{CUA}$ is a likely first intermediate in formation of pyrrolysine. Testing of pyrrolysine itself as a PylS substrate must await the chemical synthesis of this amino acid, but the low apparent K_m of PylS for lysine would render the charging of $tRNA_{CUA}$ with pyrrolysine problematic under cellular conditions where the two amino acids would be in competition. The lysyl group on $tRNA_{CUA}$ could be modified to pyrrolysine either before or during incorporation into the protein. However, the co-translational modification of lysine following incorporation into MtmB as signaled by a stop codon would be an unprecedented phenomenon, possibly involving recruitment of modification enzymes to the ribosome during translation. Rather, synthesis of pyrrolysyl-$tRNA_{CUA}$ could be achieved by condensation of the epsilon nitrogen of lys-$tRNA_{CUA}$ and carboxyl group of (*4R*,*5R*)-4-substituted-pyrroline-5-carboxylate, allowing direct translation of the UAG codon as pyrrolysine. This would have precedent in the biosynthesis of asparagine, glutamine, formylmethionine, or selenocysteine on different tRNA species, as recently reviewed in (*26*). A strong analogy can now be made between pyrrolysine and selenocysteine. Both of these noncanonical amino acids are found at positions encoded by canonical stop codons. Decoding UGA as selenocysteine involves $tRNA_{UCA}$ species with marked deviations from typical tRNA structure, such as an elongated acceptor stem (*27*). Taken together, the present data suggest that pyrrolysine represents the 22nd genetically encoded amino acid to be identified in nature.

References and Notes

1. S. A. Burke, S. L. Lo, J. A. Krzycki, *J.Bacteriol.* **180**, 3432 (1998).
2. L. Paul, D. J. Ferguson, J. A. Krzycki, *J. Bacteriol.* **182**, 2520 (2000).
3. C. M. James, T. K. Ferguson, J. F. Leykam, J. A. Krzycki, *J. Biol. Chem.* **276**, 34252 (2001).
4. G. Srinivasan, S. Burke, J. Krzycki, unpublished data. See also GenBank accession numbers AF013713 and AF230870.
5. R. F. Gesteland, J. F. Atkins, *Annu. Rev. Biochem.* **65**, 741 (1996).
6. T. M. Lowe, S. R. Eddy, *Nucleic Acids Res.* **25**, 955 (1997).
7. Materials and methods are available as supporting material on *Science* Online.
8. R. Giege, J. D. Puglisi, C. Florentz, *Prog. Nucleic Acid Res. Mol. Biol.* **45**, 129 (1993).
9. G. Eriani, M. Delarue, O. Poch, J. Gangloff, D. Moras, *Nature* **347**, 203 (1990).
10. R. W. Alexander, P. Schimmel, *Prog. Nucleic Acid Res. Mol. Biol.* **69**, 317 (2001).
11. S. F. Altschul *et al.*, *Nucleic Acids Res.* **25**, 3389 (1997).
12. A. Bateman *et al.*, *Nucleic Acids Res.* **28**, 263 (2000).
13. D. Moras, *Trends Biochem. Sci.* **17**, 159 (1992).
14. RPS-BLAST aligned three pfam groups with the predicted sequence of PylS. These include AARSs specific for phenylalanine (pfam01409 with PylS residues 254 to 403, which excludes motif 1 and 2 but includes motif 3, with an expect value of 5e-6); lysine, aspartate, or asparagine (pfam00152 with PylS residues 210 to 312, which includes motif 1 and 2 but excludes motif 3, with an expect value of 3e-7); and glycine, histidine, proline, serine, or threonine (pfam00587 with PylS residues 210 to 339, which includes motif 1 and 2 but excludes motif 3, with an expect value of 3e-13).
15. A. W. Curnow, D. L. Tumbula, J. T. Pelaschier, B. Min, D. Söll, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12838 (1998).
16. W. Freist, D. H. Glauss, *Biol. Chem. Hoppe-Seyler* **276**, 451 (1995).
17. M. Ibba *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 418 (1999).
18. M. Ibba *et al.*, *Science* **278**, 1119 (1997).
19. G. Desogus, F. Todone, P. Brick, S. Onesti, *Biochem.* **39**, 8418 (2000).
20. S. Cusack, A. Yaremchuk, M. Tukalo, *EMBO J.* **15**, 6321 (1996).
21. G. Srinivasan, L. Paul, T. Lienard, G. Gottschalk, J. Krzycki, in preparation.
22. The *D. hafniense mttB* homolog with the amber codon is found on contig 3110, available in GenBank DOE_49338, and its predicted product aligns with the entire *M. barkeri mttB* gene product (GenBank AF102623) at 63% similarity. Several other *mttB* homologs are also present in the *D. hafniense* genome that contain no amber codons, but have leucine or isoleucine at the amber codon position. These predicted products of the *mttB* homologs that do not contain amber codons average 50% similarity with *M. barkeri* MttB, and their biochemical function is unknown. The *pylT* gene cluster is found on contig 3273 in GenBank DOE_49338. Similarities for the *M.*

barkeri Fusaro and *D. hafniense pyl* gene products not shown in the figures are as follows: PylB, 63%; PylC, 49%; and PylD, 45% (see fig. S2). In each case, >90% of the two genes were involved in the alignment, and the Blosum62 matrix was used to establish similarity of residues in alignments.

23. A. J. Gale, P. Schimmel, *Biochemistry* **34**, 8896 (1995).
24. C. R. Woese, G. J. Olsen, M. Ibba, D. Söll, *Microbiol. Mol. Biol. Rev.* **64**, 202 (2000).
25. BLASTP searches at the National Center for Biotechnology Information Web site against the nonredundant databases demonstrated a strong similarity of the predicted *pylB* gene product to biotin synthase from a number of sources (alignments at 49% similarity and expect values of 3e-29), which may reflect involvement in formation of five-membered heterocyclic rings. An RPS-BLAST search indicated a fainter but significant relationship to D-alanine:D-alanine ligases in pfam01820 for the *M. barkeri* (expect value 0.005) and *D. hafniense* (expect value 10e-6) PylC proteins, suggestive of a potential role in forming the amide bond in pyrrolysine.
26. M. Ibba, D. Söll, *Annu. Rev. Biochem.* **69**, 617 (2000).
27. S. Commans, A. Bock, *FEMS Microbiol. Rev.* **23**, 335 (1999).
28. T. Ferguson, T. Lienard, G. Gottschalk, J. Krzycki, unpublished.
29. U.K. Laemmli, *Nature (London)* **227**, 680 (1970).
30. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994).
31. S. Henikoff, J. G. Henikoff, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10915 (1992).

32. The authors wish to thank C. Daniels, F. Grundy, T. Henkin, M. Ibba, and J. Reeve for helpful discussions, and the DOE Joint Genome Institute (www.jgi.doe.gov/tempweb/JGI_microbial/html/index.html) for making available the preliminary sequence data from the genomes of *M. barkeri* Fusaro and *D. hafniense*. This work was supported by the National Science Foundation (MCB-9808914).

Supporting Online Material

www.sciencemag.org/cgi/content/full/296/5572/1459/DC1

Materials and Methods

figs. S1 and S2

4 January 2002; accepted 27 March 2002

A New UAG-Encoded Residue in the Structure of a Methanogen Methyltransferase

Bing Hao,[1] Weimin Gong,[1] Tsuneo K. Ferguson,[2] Carey M. James,[2] Joseph A. Krzycki,[2]* Michael K. Chan[1]*

Genes encoding methanogenic methylamine methyltransferases all contain an in-frame amber (UAG) codon that is read through during translation. We have identified the UAG-encoded residue in a 1.55 angstrom resolution structure of the *Methanosarcina barkeri* monomethylamine methyltransferase (MtmB). This structure reveals a homohexamer comprised of individual subunits with a TIM barrel fold. The electron density for the UAG-encoded residue is distinct from any of the 21 natural amino acids. Instead it appears consistent with a lysine in amide-linkage to (4*R*,5*R*)-4-substituted-pyrroline-5-carboxylate. We suggest that this amino acid be named L-pyrrolysine.

The catabolism of methylamines by methanogens involves a conserved arrangement of proteins. A specific monomethylamine (MMA), dimethylamine (DMA), or trimethylamine (TMA) methyltransferase activates the substrate for methyl transfer to a cognate corrinoid protein (*1*, *2*). A second methyltransferase catalyzes the transfer of the methyl group from the methylated corrinoid cofactor to coenzyme M (CoM). Methyl-CoM is subsequently used to generate methane by methyl-CoM reductase (*3*).

All known methanogen methylamine (MMA, DMA, or TMA) methyltransferase genes contain a single in-frame amber (UAG) codon that does not appear to stop translation during protein synthesis (*4*, *5*). This strict conservation contrasts with the lack of sequence similarity between different MMA, DMA, and TMA methyltransferase gene families. Analysis of tryptic fragments of MMA methyltransferase (MtmB) by mass spectrometry and Edman degradation suggested that the amber codon serves as a sense codon that corresponds to lysine (*6*). The harsh conditions of peptide isolation, however, left open the possibility that the amber codon signals a modified lysine residue. The use of what is normally a stop codon to signal the incorporation of an unusual amino acid has precedent in the use of UGA to encode selenocysteine, the 21st natural amino acid found in Bacterial, Eucaryal, and Archaeal proteins (*7*, *8*).

The structure of the *Methanosarcina barkeri* MS monomethylamine methyltransferase was determined to clarify the identity of the UAG-encoded residue (Tables 1 and 2). Two forms of the enzyme were obtained from crystallization conditions that differed only in the precipitating salt used [NaCl for form 1, and $(NH_4)_2SO_4$ for form 2] and were solved to 1.55 Å and 1.7 Å resolution, respectively. The global conformations for these two forms are virtually identical, except for the region around the UAG-encoded amino acid. The overall MtmB structure consists of a homohexamer arranged into a dimer of trimers with overall D_3 symmetry (Fig. 1A). Each subunit adopts an α/β TIM barrel fold (*9*, *10*) (Fig. 1B) that is reminiscent of other corrinoid-cofactor associated proteins: tetrahydrofolate:corrinoid/iron-sulfur protein methyltransferase (*11*), diol dehydratase (*12*), and racemases/mutases (*13*, *14*), although no significant sequence similarity is found between MtmB and these proteins. The eight-stranded β barrel is formed from strands β6 to β13 (*15*). Consistent with other methyltransferases, this barrel forms a deep cavity that for the MtmB is negatively charged (*15*).

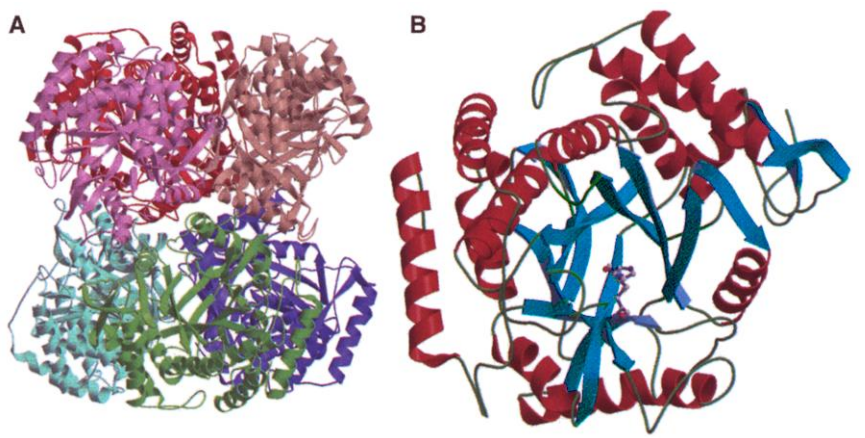

Fig. 1. (**A**) Ribbon diagram of the MtmB hexamer. The subunits forming the two trimers are shaded in red and blue hues, respectively. (**B**) Ribbon diagram of one MtmB subunit colored by secondary structure element (α helices, red; β sheets, cyan; random coil, green). The atoms of the UAG-encoded residue are shown as ball-and-stick models and are colored by their elements, with carbon as gray, nitrogen as blue, and oxygen as red. These figures were prepared with the programs MOLSCRIPT, and Raster3D (*22*, *23*).

[1]Departments of Biochemistry and Chemistry, and [2]Department of Microbiology, The Ohio State University, 484 West 12th Avenue, Columbus, OH 43210, USA.

*To whom correspondence should be addressed: E-mail: chan@chemistry.ohio-state.edu and krzycki.1@osu.edu