FEBS Letters 462 (1999) 302–306 FEBS 22982

Cysteinyl-tRNA formation: the last puzzle of aminoacyl-tRNA synthesis

Tong Li[a,1], David E. Graham[e,1], Constantinos Stathopoulos[a], Paul J. Haney[e], Hyun-soo Kim[a], Ute Vothknecht[a], Makoto Kitabatake[a], Kwang-won Hong[a], Gudmundur Eggertsson[c], Alan W. Curnow[a], Winston Lin[d], Ivana Celic[a], William Whitman[d], Dieter Söll[b,*]

[a] *Department of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 208114, 266 Whitney Avenue New Haven, CT 06520-8114, USA*
[b] *Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06511, USA*
[c] *Institute of Biology, University of Iceland, Reykjavik, Reykjavik, Iceland*
[d] *Department of Microbiology, University of Georgia, Athens, GA 30602-2605, USA*
[e] *Department of Microbiology, University of Illinois, Champaign-Urbana, IL 61801, USA*

Received 11 October 1999

Edited by Horst Feldmann

Abstract With the exception of the methanogenic archaea *Methanococcus jannaschii* and *Methanobacterium thermoautotrophicum* ΔH, all organisms surveyed contain orthologs of *Escherichia coli* cysteinyl-tRNA synthetase (CysRS). The characterization of CysRS-encoding (*cysS*) genes and the demonstration of their ability to complement an *E. coli cysS*^{ts} mutant reveal that *Methanococcus maripaludis* and *Methanosarcina barkeri*, two other methanogenic archaea, possess canonical CysRS proteins. A molecular phylogeny inferred from 40 CysRS sequences indicates that the CysRS of *M. maripaludis* and *Methanosarcina* spp. are specific relatives of the CysRS of *Pyrococcus* spp. and *Chlamydia*, respectively. This result suggests that the CysRS gene was acquired by lateral gene transfer in at least one euryarchaeotic lineage.

Key words: Cysteinyl-tRNA synthetase; Aminoacyl-tRNA

1. Introduction

Cysteine is the third least-frequently occurring amino acid in all *Methanococcus jannaschii* proteins, yet its thiol moiety has disproportionate importance as a nucleophilic catalyst, a metal ligand and electron-carrying thiol. Thus, it was surprising that the first two known archaeal genome sequences (for *M. jannaschii* [1] and *Methanobacterium thermoautotrophicum* ΔH [2]) lacked recognizable genes for enzymes in cysteine biosynthesis and the canonical cysteinyl-tRNA synthetase. As both autotrophic archaea grow in minimal media supplemented with sulfide, sulfite, thiosulfate, elemental sulfur or mercaptans [3,4], they must have the capacity to synthesize and incorporate cysteine. The subsequently sequenced archaeal genomes (*Archaeoglobus fulgidus*, *Pyrococcus* spp., and *Aeropyrum pernix*) encode recognizable CysRS proteins; this makes this puzzle all the more intriguing.

Cysteinyl-tRNA synthetase is a class I synthetase with the conserved class-defining HIGH and KMSKS motifs. It is the last synthetase whose primary structure was determined [5]. All bacteria and eukarya examined to date contain a CysRS similar to the monomeric 52 kDa enzyme found in *Escherichia coli*. Although extensive studies on tRNA-recognition were done with the *E. coli* enzyme [6,7], altogether there is much less known about CysRS than about the other aminoacyl-tRNA synthetases [8]. However, this should change in the near future with the expected availability of a crystal structure [9].

Several alternative mechanisms of Cys-tRNA^{Cys} synthesis can be considered for *M. jannaschii* and *M. thermoautotrophicum* and possibly other methanogens. If Cys-tRNA were formed by direct acylation of tRNA^{Cys} with the cognate amino acid, the reaction may be carried out by a highly diverged CysRS, as observed for the methanogen seryl-tRNA synthetase [10]. Alternatively, there may be a non-canonical, unrecognized form of CysRS, as is the case for the euryarchaeal class I lysyl-tRNA synthetase [11]. Otherwise, these organisms may employ a tRNA-dependent amino acid transformation pathway [12]. Similar to selenocysteine formation [13], this route would entail misacylation of tRNA^{Cys} with serine by seryl-tRNA synthetase to form Ser-tRNA^{Cys} which would be thiolated to Cys-tRNA^{Cys}. A scenario involving this indirect thiolation pathway is unlikely to operate in *M. thermoautotrophicum*, as purified SerRS from this organism will not in vitro charge homologous tRNA^{Cys} with serine [10]. In addition, it was recently shown that *M. jannaschii* cell extracts can acylate cysteine directly to tRNA^{Cys} [14].

We therefore investigated the formation of Cys-tRNA in a variety of organisms either biochemically, or genetically by complementation of an *E. coli cysS*^{ts} mutant strain and isolation of *cysS* genes. The results, combined with data from ongoing genome sequencing projects, allowed us to examine the evolutionary history of Cys-tRNA formation.

2. Materials and methods

2.1. Methanococcus maripaludis CysRS purification and assay

CysRS activity was measured at 37°C in a reaction mixture containing 30 mM HEPES-KOH (pH 7.5), 15 mM $MgCl_2$, 25 mM KCl, 5 mM DTT, 20 μM [^{35}S]cysteine, 1 mM ATP and 40 μg/ml of *M. maripaludis* total tRNA prepared by standard methods. After preincubation for 1 min the reaction was initiated by the addition of the enzyme. Acid-precipitable radioactivity was measured as described [15]. The CysRS preparation was purified by anion exchange chromatography from *M. maripaludis* S-100 extract.

2.2. Complementation of an E. coli cysS^{ts} strain

The *E. coli* strain UQ818 contains a thermolabile CysRS [16] and therefore cannot grow at the non-permissive temperature of 42°C.

*Corresponding author. Fax: (1)-203-432-6202.
E-mail: soll@trna.chem.yale.edu

[1] Both authors contributed equally to this work.

PII: S0014-5793(99)01550-1

Competent cells of this strain were transformed with libraries or clones. Transformants were selected for growth at 42°C. The *Methanosarcina barkeri* library [17] was kindly provided by D. Jahn (University of Freiburg, Germany). The other *cysS* genes were inserted into a pCBS1, a pBR322-derived vector with an *E. coli trpS* promoter.

2.3. Protein sequence alignment

Amino acid sequences for previously described CysRS proteins were obtained from the non-redundant protein NCBI database: *A. fulgidus* VC-16 (AE001076), *Pyrococcus* sp. OT3 (AP000003), *Pyrococcus abyssi* (AJ248287), *A. pernix* (AP000062), *Aquifex aeolicus* (AE000721), *Bacillus subtilis* (L14580), *Thermotoga maritima* (AE001743), *E. coli* (X59293), *Synechocystis* sp. (D90914), *Helicobacter pylori* (AE000598), *Rickettsia prowazekii* (AJ235270), *Chlamydia trachomatis* (AE001350), *Mycobacterium tuberculosis* (Z92774), *Borrelia burgdorferi* (AE001161), *Treponema pallidum* (AE001193), *Mycoplasma genitalium* (U39704), *Mycoplasma pneumoniae* (AE000047), *Azospirillum brasilense* (X99587), *Arabidopsis thaliana* (AC006593), *Drosophila melanogaster* (AF132160), *Caenorhabditis elegans* (AF077541), *Saccharomyces cerevisiae* (X96722) and *Homo sapiens* (L06845). A second, paralogous gene from *M. tuberculosis* (*cysS2*) (Z97559) was also included in the analysis. Sequence data from partial genome sequences for *Chlorobium tepidum*, *Deinococcus radiodurans* and *Porphyromonas gingivalis* were obtained from http://www.tigr.org, *Pyrococcus furiosus* from http://www.genome.utah.edu, *Campylobacter jejuni* from http://www.sanger.ac.uk, *Clostridium acetobutylicum* from http://www.genomecorp.com, *Pseudomonas aeruginosa* from http://www.pseudomonas.com, *Neisseria gonorrhoeae* from http://www.genome.ou.edu and *Rhodobacter capsulatus* from http://wit.mcs.anl.gov/WIT. Sequences not yet publicly available were from *Thermoplasma acidophilum* (courtesy of A. Ruepp, Max-Planck-Institut für Biochemie, Martinsried, Germany), *Halobacterium salinarum* (courtesy of S. Schuster and D. Oesterhelt, Max-Planck-Institut für Biochemie, Martinsried, Germany), *Methanosarcina mazei* Gö1 (courtesy of T. Hartsch, Göttingen Genomics Laboratory, Germany), and *Pyrobaculum aerophilum* (courtesy of S. Fitz-Gibbon, UCLA). Forty amino acid sequences from CysRS proteins were aligned using the CLUSTALW (v.1.7.4) program [18].

2.4. Phylogenetic inference

From the alignment, positions corresponding to 404 of the 461 amino acids in the *E. coli* CysRS sequence were deemed to be confidently aligned. These were analyzed by protein maximum parsimony methods using a heuristic search algorithm (PAUP* v4 beta 2, D. Swofford, Sinauer Associates). The 1000 shortest trees were evaluated by maximum likelihood (ML) criteria using the PROTML program (v2.2) in the MOLPHY package [19] with the JTT model for amino acid substitutions. The TreeCons program (v1.0) [20] standardized and exponentially weighted the trees using ML scores and the Kishino-Hasegawa test for significance ($P \leq 0.01$). The CONSENSE program (J. Felsenstein, PHYLIP (phylogeny inference package). Version 3.5c. Department of Genetics, U. Washington, Seattle, WA, USA, 1993) constructed a consensus tree from the weightings. The neighbor-joining (NJ) consensus tree was constructed from the same data set using the PAUP* program to evaluate 1000 bootstrap samples. Phylogenetic trees were viewed and edited with the TreeView program (v. 1.5.2) [21].

3. Results and discussion

3.1. Cloning of the M. barkeri cysS gene by heterologous complementation

To identify functional archaeal *cysS* genes, we transformed libraries from *M. barkeri*, *Haloferax volcanii* and *Methanopyrus kandleri* into the temperature-sensitive *E. coli* UQ818 strain (carrying a *cysS*^{ts} allele) and selected for growth at the non-permissive temperature (42°C). DNA sequencing revealed that colonies obtained from this selection with the *M. barkeri* library contained a plasmid encoding a canonical *cysS* (GenBank accession number AF164201) (Fig. 1). The lack of complementation with the other two libraries may be due to the halophilic or hyperthermophilic nature of the *H. volcanii* and *M. kandleri* proteins, or may reflect incompatible tRNA recognition elements between those archaea and *E. coli* tRNA^{Cys} [7]. Alternatively, these organisms might use a novel CysRS requiring unique cofactors or consisting of multiple heterologous subunits.

3.2. M. maripaludis charges cysteine directly and contains a canonical CysRS

M. maripaludis is a mesophilic relative of the hyperthermophilic *M. jannaschii*. Because *M. maripaludis* can be easily grown and genetically manipulated [22], we biochemically tested whether this organism attaches cysteine directly to tRNA. A partially purified *M. maripaludis* CysRS preparation charged cysteine efficiently to unfractionated homologous tRNA (Fig. 2) to a level of about 68 pmol/A_{260} unit. This represents a much more efficient charging than that reported for *M. jannaschii* [14]. Independently, we identified a clone with similarity to the canonical *cysS* gene from a small insert library of *M. maripaludis* genomic DNA (GenBank accession number AF163997). This *cysS* ORF encodes a 55 kDa protein homologous to *E. coli* CysRS. Thus unlike *M. jannaschii*, *M. maripaludis* has a canonical cysteinyl-tRNA synthetase.

3.3. Are diverse cysS genes active in E. coli?

In order to demonstrate in vivo activity of diverse *cysS* gene products, we complemented the *E. coli cysS*^{ts} mutant with *cysS* genes from four archaea (*A. pernix, P. aerophilum, M. maripaludis, M. barkeri*) and the bacterium *R. marinus*. While the cloned *cysS* genes from *M. maripaludis* and *M. barkeri* gave rise to colonies that grew very well at the non-permissive temperature, the *A. pernix* transformants grew very slowly, possibly due to reasons discussed above for the hyperthermophile *M. kandleri*. The *P. aerophilum* and *Rhodothermus marinus* transformants grew slowly, but only in the presence of added cysteine; thus, the affinity of their CysRS enzymes for cysteine may be lower, a case also seen in the complementation of an *E. coli lysS*^{ts} mutant with the *M. maripaludis lysS* gene [23]. From these results, we conclude that all of the genes tested have CysRS activity in vivo. Thus, there is no prima facie barrier to the lateral transfer and gene displacement of evolutionarily diverse *cysS* genes into *E. coli*.

3.4. Cloning and sequencing the R. marinus cysS gene

R. marinus, a moderately thermophilic, aerobic, marine bacterium of interest in Iceland [24], is distantly related to the Flexibacter/Cytophaga/Bacteroides group [25]. A clone encoding the organism's *cysS* gene was obtained and sequenced (GenBank accession number AF162864). The deduced 504 amino acid protein is similar to that of other bacterial *cysS* gene products (Fig. 1).

3.5. Evolutionary relationship of CysRS proteins

Amino acid sequences from 40 CysRS enzymes were aligned and then analyzed by several methods to infer that protein's phylogeny. The unrooted consensus ML tree (Fig. 3A) supported the presence of multiple ancient lineages, the branching order of which could not be clearly resolved. The eukaryal CysRS proteins appeared specifically related to one another, although the *Arabidopsis* gene was highly divergent, and a NJ tree placed it in another group with low bootstrap support (Fig. 3B). Bacterial CysRS proteins typically grouped with their closest relatives predicted by the SSU rRNA phy-

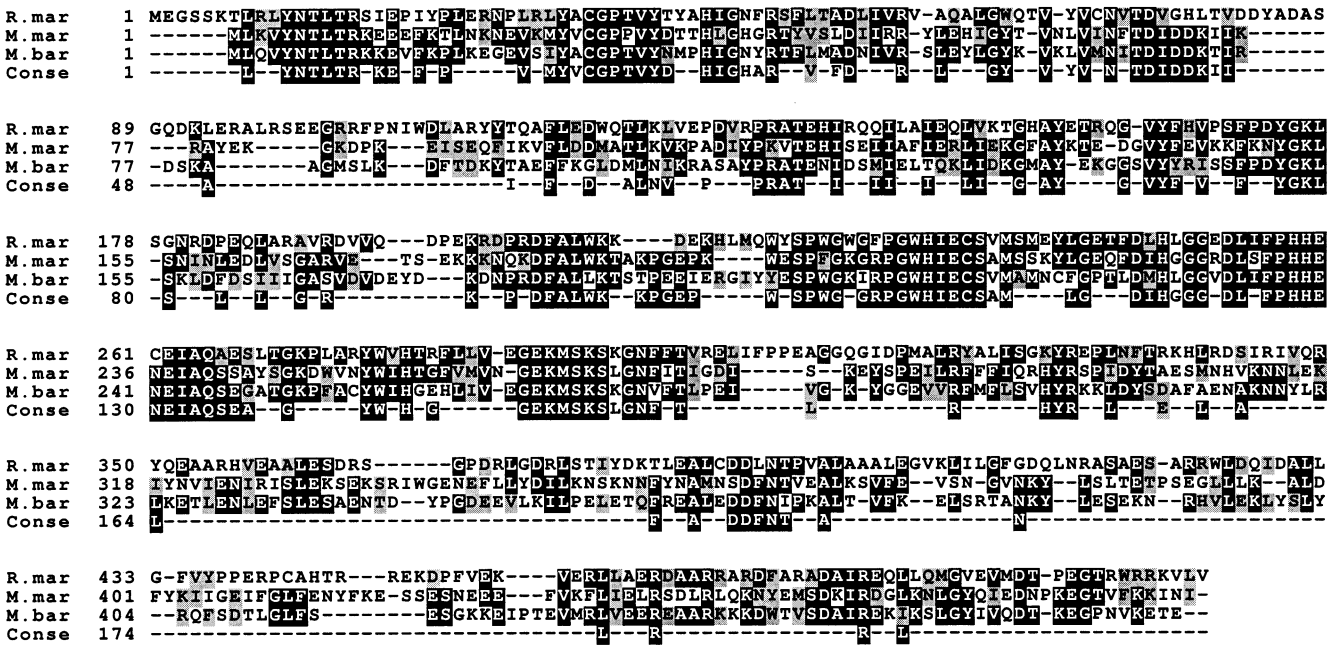

Fig. 1. Sequence alignment of the CysRS protein sequences from *M. maripaludis*, *M. barkeri*, *R. marinus* and a consensus sequence of 43 CysRS sequences from eukaryotic (5), archaeal (8) and bacterial (30) origin. At least 50% amino acid identity was required for the consensus.

logeny [26]. However, more distant relationships were not clearly resolved, as indicated by the low relative likelihood support (RLS) for the deep branches and the lack of congruence between the ML and NJ trees (Fig. 3). Among euryarchaeota, CysRS from *Pyrococcus* spp. appeared closely related to the enzyme described here from *M. maripaludis*. This relationship was robust, i.e. it was strongly supported by high RLS scores and it was present in both ML and NJ trees. In contrast, the *Methanosarcina* spp. CysRS appeared closely related to the enzyme from *C. trachomatis*, suggesting that one of these lineages may have acquired CysRS by lateral gene transfer. A tentative relationship between the euryarchaeal CysRS from *H. salinarum* and the bacterial protein from *D. radiodurans* may be influenced by the biased amino acid composition observed in halophiles. A tentative relationship was also detected between the *T. acidophilum* CysRS and the enzyme from the low G+C Gram-positive lineage of bacteria. The ML tree also grouped the two crenarchaeal sequences with CysRS proteins from the hyperthermophilic bacteria *A. aeolicus* and *T. maritima.* Despite a high RLS value for the *Pyrobaculum*/*Thermotoga* group, this relationship was not strongly supported in the NJ trees and was considered tentative. If the associations of archaeal enzymes with different bacterial lineages were caused by lateral gene transfer(s), these were not recent events because the mol percentage G+C contents of all genes in this study closely match each organism's chromosomal composition.

Two of the methanogenic euryarchaeota, *M. jannaschii* and *M. thermoautotrophicum*, are known to lack a canonical CysRS. Therefore, we did not expect to discover these proteins in the closely related *M. maripaludis* and *Methanosarcina* spp. Although the highly divergent nature of the CysRS sequences makes it difficult to resolve deep evolutionary relationships, the *M. maripaludis* and *Methanosarcina* CysRS appear to have very different histories, suggesting that lateral transfer occurred in these lineages. To fully reconstruct the evolutionary history of CysRS in euryarchaea will require knowledge of the alternative cysteinyl-tRNA charging systems from *M. jannaschii* and *M. thermoautotrophicum*, combined with better sampling of charging systems from euryarchaeota, especially other members of the *Methanococcales*, the *Methanomicrobiales* and *M. kandleri*.

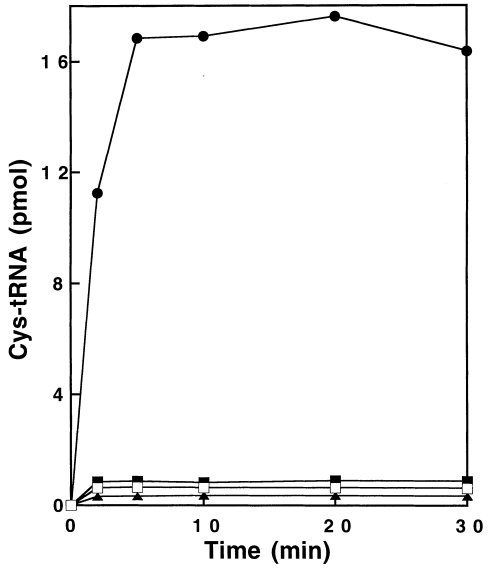

Fig. 2. Direct attachment of cysteine to tRNA by partially purified *M. maripaludis* CysRS. Aminoacylation reactions were performed as described in the presence of 2 μg of CysRS partially purified (see Section 2) in a reaction volume of 120 μl. Samples (20 μl) were periodically removed and analyzed. Reactions were performed in the presence of 20 μM [^{35}S]cysteine (●), 20 μM [^{35}S]cysteine in the absence of CysRS (■), 20 μM [^{35}S]cysteine in the absence of tRNA (□), and 20 μM [^{35}S]cysteine in the presence of 0.8 mM non-radiolabelled cysteine (▲).

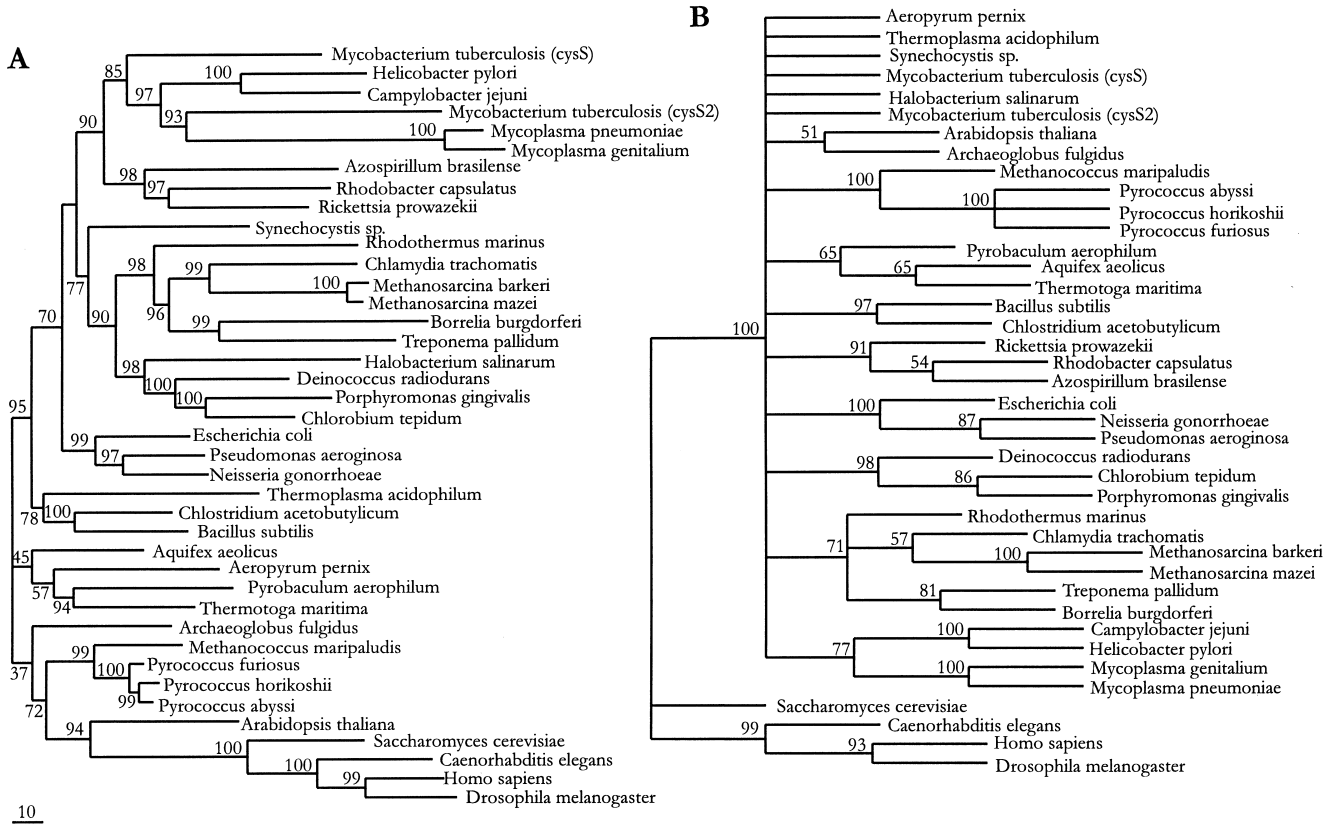

Fig. 3. Consensus phylogenetic trees of CysRS proteins, inferred by protein ML analysis (A) or the NJ method (B). Trees are arbitrarily rooted using the eukaryal CysRS proteins as an outgroup. Numbers near each node in (A) are the relative-likelihood support values for each branching. Bootstrap probabilities estimated by the resampling estimated log-likelihood method are comparable to RLS values and produce an identical consensus tree (data not shown). The scale bar in (A) represents 10 substitutions per 100 amino acid positions along each branch. Numbers for each branching in (B) are percent bootstrap confidence values calculated from 1000 samplings. Lineages without majority bootstrap support are represented as polychotomies and may be considered unresolved by this method.

3.6. Conclusion

In this study we have identified CysRS enzymes in the archaea *M. maripaludis* and *M. barkeri*. Both protein sequences are similar to *E. coli* CysRS and can complement the *E. coli* *cysS*^{ts} mutation. The existence of these canonical proteins in two methanogenic archaea makes their absence from *M. jannaschii* and *M. thermoautotrophicum* all the more bizarre. As the pathways for the synthesis of all other aminoacyl-tRNAs are now clear [27], the discovery of the nature of Cys-tRNA formation in *M. jannaschii* and *M. thermoautotrophicum* will be the last step in solving the puzzle of aminoacyl-tRNA formation.

Acknowledgements: We thank S. Fitz-Gibbon, T. Hartsch, D. Oesterhelt, A. Ruepp and S. Schuster for sharing unpublished sequence data. M.K. was a postdoctoral fellow of the Japan Society for the Promotion of Science. D.E.G. was supported by an NIH Cellular and Molecular Biology Training grant. This work was supported by grants from the National Institute of General Medical Sciences (to D.S.) and the U.S. Department of Energy (to D.S. and W.B.W.).

References

[1] Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S.M. and Venter, J.C. (1996) Science 273, 1017–1140.

[2] Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prabhakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrokovski, S., Chruch, G., Daniels, J., Mao, J., Rice, P., Nölling, J. and Reeve, J.N. (1997) J. Bacteriol. 179, 7135–7155.

[3] Daniels, L., Belay, N. and Rajagopal, B.S. (1986) Appl. Environ. Microbiol. 51, 703–709.

[4] Rajagopal, B.S. and Daniels, L. (1986) Curr. Microbiol. 14, 137–144.

[5] Eriani, G., Dirheimer, G. and Gangloff, J. (1991) Nucleic Acids Res. 19, 265–269.

[6] Pallanck, L., Li, S. and Schulman, L.H. (1992) J. Biol. Chem. 267, 7221–7223.

[7] Lipman, R.S. and Hou, Y.-M. (1998) Proc. Natl. Acad. Sci. USA 95, 13495–13500.

[8] Newberry, K.J., Kohn, J., Hou, Y.-M. and Perona, J.J. (1999) Acta Crystallogr. D Biol. Crystallogr. 55, 1046–1047.

[9] Meinnel, T., Mechulam, Y. and Blanquet, S. (1995) in: tRNA: Structure, Biosynthesis, and Function (Söll, D. and RajBhandary, U., Eds.), pp. 251–292, ASM Press, Washington, DC.

[10] Kim, H.S., Vothknecht, U.C., Hedderich, R., Celic, I. and Söll, D. (1998) J. Bacteriol. 180, 6446–6449.

[11] Ibba, M., Morgan, S., Curnow, A.W., Pridmore, D.R., Vothknecht, U.C., Gardner, W., Lin, W., Woese, C.R. and Söll, D. (1997) Science 278, 1119–1122.

[12] Ibba, M., Curnow, A.W. and Söll, D. (1997) Trends Biochem. Sci. 22, 39–42.

[13] Commans, S. and Böck, A. (1999) FEMS Microbiol. Rev. 23, 335–351.

[14] Hamann, C.S., Sowers, K.R., Lipman, R.S. and Hou, Y.M. (1999) J. Bacteriol. 181, 5880–5884.
[15] Sherman, J.M. and Söll, D. (1996) Biochemistry 35, 601–607.
[16] Bohman, K. and Isaksson, L.A. (1979) Mol. Gen. Genet. 176, 53–55.
[17] Vorholt, J.A., Vaupel, M. and Thauer, R.K. (1996) Eur. J. Biochem. 236, 309–317.
[18] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[19] Adachi, J. and Hasegawa, M. (1996) Comput. Sci. Monogr. 28, 1–150.
[20] Jermiin, L.S., Olsen, G.J., Mengersen, K.L. and Easteal, S. (1997) J. Mol. Evol. 14, 1296–1302.
[21] Page, R.D.M. (1996) Comput. Appl. Biosci. 12, 357–358.
[22] Tumbula, D.L. and Whitman, W.B. (1999) Mol. Microbiol. 33, 1–7.
[23] Ibba, M., Losey, H.C., Kawarabayasi, Y., Kikuchi, H., Bunjun, S. and Söll, D. (1999) Proc. Natl. Acad. Sci. USA 96, 418–423.
[24] Halldorsdottir, S., Thorolfsdottir, E.T., Spilliaert, R., Johansson, M., Thorbjarnardottir, S.H., Palsdottir, A., Hreggvidsson, G.O., Kristjansson, J.K., Holst, O. and Eggertsson, G. (1998) Appl. Microbiol. Biotechnol. 49, 277–284.
[25] Andresson, O.S. and Fridjonsson, O.H. (1994) J. Bacteriol. 176, 6165–6169.
[26] Maidak, B.L. (1999) Nucleic Acids Res. 27, 171–173.
[27] Tumbula, D.L., Vothknecht, U.C., Kim, H-S., Ibba, M., Min, B., Li, T., Pelaschier, J., Stathopoulos, C., Becker, H.D. and Söll, D. (1999) Genetics 152, 1269–1276.