A dual-specificity aminoacyl-tRNA synthetase in the deep-rooted eukaryote *Giardia lamblia*

Shipra Bunjun*[†], Constantinos Stathopoulos*[†], David Graham[‡], Bokkee Min[†], Makoto Kitabatake[†], Alice L. Wang[§], Ching C. Wang[§], Christian P. Vivarès[¶], Louis M. Weiss[‖], and Dieter Söll[†][††]**

[†]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520-8114; **Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06511; [‡]Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL 61801; [§]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-0446; [¶]Laboratoire de Parasitologie Moleculaire et Cellulaire, Université Blaise Pascal, 63177 Aubiere Cedex, France; and [‖]Department of Pathology, Albert Einstein College of Medicine, Bronx, NY 10461

Contributed by Dieter Söll, September 15, 2000

Cysteinyl-tRNA (Cys-tRNA) is essential for protein synthesis. In most organisms the enzyme responsible for the formation of Cys-tRNA is cysteinyl-tRNA synthetase (CysRS). The only known exceptions are the euryarchaea *Methanococcus jannaschii* and *Methanobacterium thermoautotrophicum*, which do not encode a CysRS. Deviating from the accepted concept of one aminoacyl-tRNA synthetase per amino acid, these organisms employ prolyl-tRNA synthetase as the enzyme that carries out Cys-tRNA formation. To date this dual-specificity prolyl-cysteinyl-tRNA synthetase (ProCysRS) is only known to exist in archaea. Analysis of the preliminary genomic sequence of the primitive eukaryote *Giardia lamblia* indicated the presence of an archaeal prolyl-tRNA synthetase (ProRS). Its *proS* gene was cloned and the gene product overexpressed in *Escherichia coli*. By using *G. lamblia*, *M. jannaschii*, or *E. coli* tRNA as substrate, this ProRS was able to form Cys-tRNA and Pro-tRNA *in vitro*. Cys-AMP formation, but not Pro-AMP synthesis, was tRNA-dependent. The *in vitro* data were confirmed *in vivo*, as the cloned *G. lamblia proS* gene was able to complement a temperature-sensitive *E. coli* cysS strain. Inhibition studies of CysRS activity with proline analogs (thiaproline and 5′-*O*-[*N*-(L-prolyl)-sulfamoyl]adenosine) in a *Giardia* S-100 extract predicted that the organism also contains a canonical CysRS. This prediction was confirmed by cloning and analysis of the corresponding *cysS* gene. Like a number of archaea, *Giardia* contains two enzymes, ProCysRS and CysRS, for Cys-tRNA formation. In contrast, the purified *Saccharomyces cerevisiae* and *E. coli* ProRS enzymes were unable to form Cys-tRNA under these conditions. Thus, the dual specificity is restricted to the archaeal genre of ProRS. *G. lamblia's* archaeal-type prolyl- and alanyl-tRNA synthetases refine our understanding of the evolution and interaction of archaeal and eukaryal translation systems.

Aminoacyl-tRNA synthetases (AARSs) are essential for the faithful translation of the genetic code. They ensure the fidelity of protein synthesis by correctly acylating a tRNA species with its cognate amino acid (1). This crucial family of enzymes is divided into two distinct classes (I and II) based on characteristic signature motifs. These and other conserved structural features allow facile recognition of orthologous enzymes in many organisms by sequence similarity searches of the available databases. They also allow comparison between the AARSs and thus classification into subtypes with respect to their phylogenetic origin (2). These enzymes have exquisite specificity for their substrates (amino acid and tRNA), a process mediated in some cases by intricate editing mechanisms (3–5). It is commonly accepted that each cell contains 20 such enzymes, one for each canonical amino acid. This assumption was supported by the description of 20 AARSs found in some bacteria (e.g., *E. coli*) and in the eukaryotic cytoplasm. However, recent discoveries arising from functional genomics studies in bacteria and archaea have overturned this concept and revealed that most organisms do not use a full complement of 20 canonical AARSs (6–8). The major exception is the route to Gln-tRNA or Asn-tRNA formation, which in many organisms involves an amidation of a mischarged Glu-tRNA or Asp-tRNA (8–10). Most surprisingly, it was demonstrated *in vitro* and *in vivo* that the ProRS of *M. jannaschii*, *M. thermoautotrophicum*, and *Methanococcus maripaludis* is able to catalyze the formation of cysteinyl-tRNA (Cys-tRNA; ref. 7). The fact that the two former organisms lack a canonical cysteinyl-tRNA synthetase (CysRS) implies that these organisms use the dual-specificity prolyl-cysteinyl-tRNA synthetase (ProCysRS) for the synthesis of both Cys-tRNA and Pro-tRNA required for protein synthesis. The fact that a single AARS is capable of and required for supplying two different aminoacyl-tRNAs for protein synthesis challenged the accepted view of these enzymes (1).

As we showed earlier, the genome of the lower eukaryote *Giardia lamblia* genome contains genes encoding the archaeal genre of ProRS and alanyl-tRNA synthetase (AlaRS; ref. 2). *G. lamblia* is a parasitic, amitochondrial protist that diverged early in the eukaryotic lineage (11, 12). By comparing genes from this organism with homologous eukaryotic and archaeal genes it may be possible to infer the events that led to the formation of the modern eukaryal cell type. The parasite *G. lamblia* is found in every region of the United States and throughout the world and infects about 200 million people annually worldwide (13). It has become recognized as one of the most common causes of waterborne disease in humans in the United States, causing diarrhea, abdominal cramps, and nausea. Further knowledge of the essential enzymes of protein synthesis may reveal new targets for development of antiparasitic agents. Here we report the existence of the dual-specificity ProCysRS and its activity in *G. lamblia*.

Materials and Methods

Cloning of the *G. lamblia proS, alaS,* and *cysS* Genes. Partial sequence data of the *G. lamblia* genes *proS*, *alaS*, and *cysS*, obtained from the *Giardia lamblia* genome project (supported by the National Institute of Allergy and Infectious Diseases) at the Josephine Bay Paul Center web site at the Marine Biological Laboratory in Woods Hole, MA (http://hermes.mbl.edu/baypaul/Giardia-HTML/index2.html), were used to design primers that enabled cloning of

---

Abbreviations: AARSs, aminoacyl-tRNA synthetases; Cys-tRNA, cysteinyl-tRNA; CysRS, cysteinyl-tRNA synthetase; ProCysRS, prolyl-cysteinyl-tRNA synthetase; ProRS, prolyl-tRNA synthetase; AlaRS, alanyl-tRNA synthetase.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF245445, AF299082, and AF245446).

*S.B. and C.S. contributed equally to the work.

[††]To whom reprint requests should be addressed at: Department of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 208114, 266 Whitney Avenue, New Haven, CT 06520-8114. E-mail: soll@trna.chem.yale.edu.

complete gene sequences from a genomic library of *G. lamblia* WB (14). The PCR products were cloned into the pCR-TOPO vector (Invitrogen) and sequenced to confirm the identity of the gene. Upon digestion with the restriction enzymes *Nde*I and *Bam*HI, the gene was ligated into pET15b (Invitrogen) for subsequent expression of an N-terminal His-tagged protein and into pCBS1 (15) for subsequent complementation tests. The sequences were deposited in GenBank (accession numbers: *alaS*, AF245445; *cysS*, AF299082; *proS,* AF245446). A plasmid for tRNA expression, pTECH, constructed by the introduction of a multiple cloning site, was cloned into the *Hinc*II-*Eco*RV site of a mutant pACYC184 (New England Biolabs), which lacks the *Eco*RI site in the chloramphenicol resistance gene. A tRNA gene cloned in the multiple cloning site is expressed constitutively from the *lpp* promoter, and the transcription is terminated by an *rrn* terminator.

Overexpression and Purification of *G. lamblia, S. cerevisiae,* and *E. coli* ProRS. For overexpression of *G. lamblia proS*, clone pET15b-*proS* was used to transform the *E. coli* strain BL21-Codon Plus-RIL (Stratagene). Transformants were grown in 5 ml LB medium containing 100 μg/ml ampicillin and 34 μg/ml chloramphenicol at 30°C for 15 h. After inoculation of 500 ml of LB medium with the addition of ampicillin (100 μg/ml), the expression of the His$_6$-tagged *proS* gene was induced for 5 h with the addition of isopropyl-β-D-thiogalactoside to a final concentration of 1 mM. Harvesting of cells, lysis, and purification of the cell extract by Ni-nitrilotriacetic acid-agarose chromatography (Qiagen, Chatsworth, CA) were performed as previously described (16). The *G. lamblia* His$_6$-ProCysRS was >95% pure, as judged by Coomassie brilliant blue staining after SDS/PAGE. Active fractions were pooled, concentrated with solid polyethylene glycol 20,000, dialyzed against aminoacylation buffer containing 20% (vol/vol) glycerol, and stored at −20°C.

Both *S. cerevisiae* cytoplasmic and mitochondrial *proS* genes, YHR020w (sp | P38708) and YER087w (sp | P39965), respectively, were cloned by PCR, using a genomic library as DNA template. The pBAD-TOPO/TA system (Invitrogen), which allows the expression of the gene of interest with a His$_6$ tag at the C terminus, was used for expression in *E. coli*. Expression of both genes was induced in the presence of 0.02% L-arabinose in the growth medium. The *E. coli proS* gene was a gift from T. Li (Yale University). All three enzymes were purified as described above for *G. lamblia* His$_6$-ProCysRS.

***G. lamblia* tRNA-Free Cell Extract Preparation.** *G. lamblia* WB cells were cultured as described (17) and harvested as a confluent sheet formed on the surface of the culture flask. Cells were washed once in 100 vol of 0.1 M NaCl, 10 mM Tris·HCl (pH 7.5) and centrifuged at 3,000 × *g* for 15 min. The pellet was resuspended in 4.5 ml of 50 mM Hepes (pH 7.0), 50 mM KCl, 15 mM MgCl$_2$, 5 mM DTT and a mixture of protease inhibitors (10 μg/ml each of PMSF, pepstatin, leupeptin, aprotinin, and 16 μg/ml benzamidine). After sonication (Fisher Sonic Dismembrator 550) at 4°C, the lysate was centrifuged at 4°C in a Beckman TL-100 centrifuge at 100,000 × *g* for 1 h. The supernatant was diluted with an equal volume of glycerol and stored at −70°C. For the aminoacylation assay this solution was dialyzed overnight at 4°C against the reaction buffer of 50 mM Hepes (pH 7.0), 50 mM KCl, 15 mM MgCl$_2$, and 5 mM DTT.

Aminoacylation and ATP-PP$_i$ Exchange Assays. Cys-tRNA and Pro-tRNA synthesis was assayed by measuring the acid-precipitable aminoacyl-tRNA as described (7), after incubation at 37°C. The reaction mixture contained 50 mM Hepes (pH 7.0), 50 mM KCl, 15 mM MgCl$_2$, and 5 mM DTT in the presence of 10 mM ATP, 20–50 μM [^{35}S]cysteine (1075 Ci/mmol; NEN DuPont), or [^{3}H]proline (103 Ci/mmol; Amersham). The substrate used was unfractionated *M. jannaschii* tRNA (1 mg/ml final concentration) unless indicated otherwise. Also used was total tRNA from *E. coli* (Roche Molecular Biochemicals), *S. cerevisiae* (Roche Molecular Biochemicals), and *G. lamblia* (prepared by standard methods). The enzyme concentration ranged from 0.05 to 0.5 μM. Assays for inhibition of enzyme activity were performed in the presence of 2 mM thiaproline (Sigma) or 1 mM 5′-*O*-[*N*-(L-prolyl)-sulfamoyl]-adenosine (gift of S. Cusack, European Molecular Biology Laboratory, Grenoble Outstation).

ATP-PP$_i$ exchange assays were performed at 37°C in the presence or absence of total *M. jannaschii* tRNA with [^{32}P]PP$_i$ (NEN DuPont; 4.6 Ci/mmol). The reaction mixture also contained 2 mM proline or cysteine and 1 mM ATP in a total volume of 200 μl. Samples (40 μl) were taken at various time points, and the reaction was stopped by adding 200 μl of 1% activated carbon in 0.4 M sodium pyrophosphate and 15% (vol/vol) perchloric acid. The mixture was filtered through glass microfiber filter disks (GF/C, 24 mm diam.; Whatman), and radioactivity was measured by liquid scintillation counting.

Complementation of *E. coli* Strain UQ818 (*cysS*^{ts}). The *G. lamblia proS* gene was cloned into the plasmid pCBS1 to yield pCBS-*proS*. The *E. coli* cysS^{ts} strain UQ818 (18), with an additional pTech plasmid for expression of the *M. jannaschii* tRNA^{Cys} gene (pTech-*Mj-tRNA*^{Cys}), was transformed with the compatible plasmid pCBS-*Gl-proS*. For use as positive and negative controls, UQ818 was transformed with the plasmid pCBS-*Ec-cysS* and the empty vector pCBS1, respectively; both of these, as well as the pTech-*Mj-tRNA*^{Cys}, were gifts from T. Li (Yale University). The resulting transformants were tested for growth on LB agar supplemented with ampicillin (100 μg/ml), chloramphenicol (34 μg/ml), and cysteine (5 mM) at 30°C and 42°C, as has been previously described (7). Transformants were also tested for growth in the absence of pTech-*Mj-tRNA*^{Cys}, without chloramphenicol in the growth medium.

Protein Sequence Alignments. The following amino acid sequences for previously described AlaRS proteins were obtained from the nonredundant protein database at the National Center for Biotechnology Information: *Aeropyrum pernix* (sp | Q9Y9 × 3), *Aquifex aeolicus* (sp | O67323), *Arabidopsis thaliana* (sp | P36428), *Archaeoglobus fulgidus* (sp | O28029), *Bacillus subtilis* (sp | O34526), *Borrelia burgdorferi* (sp | O51238), *Caenorhabditis elegans* (gb | AAB52339.1), *Chlamydia trachomatis* (sp | O84754), *Dictyostelium discoideum* (gb | AAF05592.1), *Drosophila melanogaster* (gb | AAF05593.1), *Escherichia coli* (sp | P00957), *Homo sapiens* (sp | P49588), *Methanobacterium thermoautotrophicum* (sp | O27718), *Methanococcus jannaschii* (sp | Q57984), *Mycobacterium tuberculosis* (sp | O07438), *Pyrococcus horikoshii* OT3 (sp | O58035), *Rickettsia prowazekii* (sp | Q9ZCA4), *Saccharomyces cerevisiae* (sp | P40825), *Sulfolobus solfataricus* (sp | P96041), *Synechocystis* sp. (sp | P74423), *Thermotoga maritima* (sp | Q9 × 1B6), *Thermus aquaticus* (sp | P74941), and *Treponema pallidum* (sp | O83980). The following amino acid sequences for previously described ProRS proteins were also obtained from the nonredundant protein database at the National Institute for Biotechnology Information: *Aeropyrum pernix* (dbj | BAA81340.1), *Aquifex aeolicus* (gb | AAC06648.1), *Archaeoglobus fulgidus* (sp | O28664), *Bacillus subtilis* (emb | CAB13530.1), *Borrelia burgdorferi* (gb | AAC66767.1), *Caenorhabditis elegans* (cytoplasmic: gb | AAA50660.1, mitochondrial: emb | CAB04884.1), *Chlamydia pneumoniae* (gb | AAD18640.1), *Chlamydia trachomatis* (sp | P36431), *Deinococcus radiodurans* (gb | AAF10837.1), *Dictyostelium discoideum* (dbj | C24346.1), *Escherichia coli* (sp | P16659), *Homo sapiens* (mitochondrial: emb | CAB55948.1), *Methanobacterium thermoautotrophicum* (sp | O26708), *Methanococcus jannaschii* (sp | Q58635), *Mycobacterium leprae* (emb | CAB36573.1), *Mycobacterium tuberculosis* (sp | O05814), *Mycoplasma genitalium* (sp | P47525), *Mycoplasma* *pneumoniae* (sp | P75382), *Pyrococcus horikoshii* OT3 (dbj | BAA30103.1), *Pyrococcus abyssi* (emb | CAB49884.1), *Rickettsia prowazekii* (sp | Q9ZDE7), *Saccharomyces cerevisiae* (cytoplasmic: sp | P38708; mitochondrial: sp | P39965), *Schizosaccharomyces pombe* (cytoplasmic: emb | CAA19574.1; mitochondrial: emb | CAA21147.1), *Sulfolobus solfataricus* (gi | 6015904), *Synechocystis* sp. (sp | P73942), *Thermotoga maritima* (gb | AAD35599.1), *Treponema pallidum* (sp | O83195), and *Ureaplasma urealyticum* (gb | AAF30864.1). Sequence data from partial genome sequences for *Clostridium acetobutylicum* were from http://www.genomecorp.com, *Porphyromonas gingivalis* and *Trypanosoma brucei* were obtained from http://www.tigr.org, *Pyrococcus furiosus* was from http://www.genome.utah.edu, and *Streptomyces coelicolor* and *Plasmodium falciparum* were from http://www.sanger.ac.uk. 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, Univ. of California, Los Angeles). Thirty-three amino acid sequences from AlaRS were aligned using the CLUSTALW (v.1.7.4) program (19). Forty-three amino acid sequences from ProRSs were separated into groups of archaeal- and bacterial-type enzymes and then aligned using CLUSTALW. These alignments were combined manually using the AE2 alignment editor (T. Macke, Ribosomal Database Project). From the human and *Drosophila* bifunctional Gluprolyl-tRNA synthetase only regions homologous to other ProRS proteins were used in this analysis.

Phylogenetic Inference. From the alignment of 33 AlaRS proteins, 717 positions were deemed to be confidently aligned. These were analyzed as previously described (2), by protein maximum parsimony methods using a heuristic search algorithm (PAUP* 4.0 beta 2; D. Swofford, Sinauer). The 1,000 shortest trees were evaluated by maximum likelihood criteria, using the PROTML program (v 2.2) in the MOLPHY package (20) with the JTT model for amino acid substitutions. Bootstrap percentages for each node in the tree were estimated by the resampling estimated log-likelihood (RELL) method (21), using the PROTML program to compare the 1,000 most parsimonious trees. The CONSENSE program [PHYLIP (phylogeny inference package), v. 3.5c; J. Felsenstein, Department of Genetics, University of Washington, Seattle, 1993] was used to construct a consensus tree from the RELL weightings. Phylogenetic trees were viewed and edited with the TREEVIEW program (v. 1.5.2; ref. 22). In the alignment of 43 ProRSs, 579 positions were analyzed by the same techniques.

Results

Genomic *Giardia lamblia* Sequence Data Suggest the Presence of Archaeal ProRS and AlaRS Enzymes and of a Canonical CysRS. Our recent analysis of aminoacyl-tRNA synthetase genes in a large number of organisms revealed the presence of an archaeal ProRS and AlaRS in *Giardia lamblia* (2). This finding piqued our interest for two reasons: analysis of these genes may provide additional data on the evolutionary position of *Giardia*. More intriguingly, it suggested that even eukaryotes may possess the novel dual-specificity ProCysRS, which to date was biochemically demonstrated to be present in at least three archaeal genomes, even though sequence data suggested that the same enzyme activity will be found in other archaea (ref. 7; C.S., unpublished work). Thus, based on the partial sequence information (see *Materials and Methods*), we cloned the *G. lamblia proS*, *alaS*, and *cysS* genes. Their sequences were determined (GenBank accession nos. AF245446, AF245445, AF299082); the deduced AlaRS and ProRS sequences are of the archaeal genre (see below), whereas the CysRS showed much similarity to other known CysRS proteins (15).

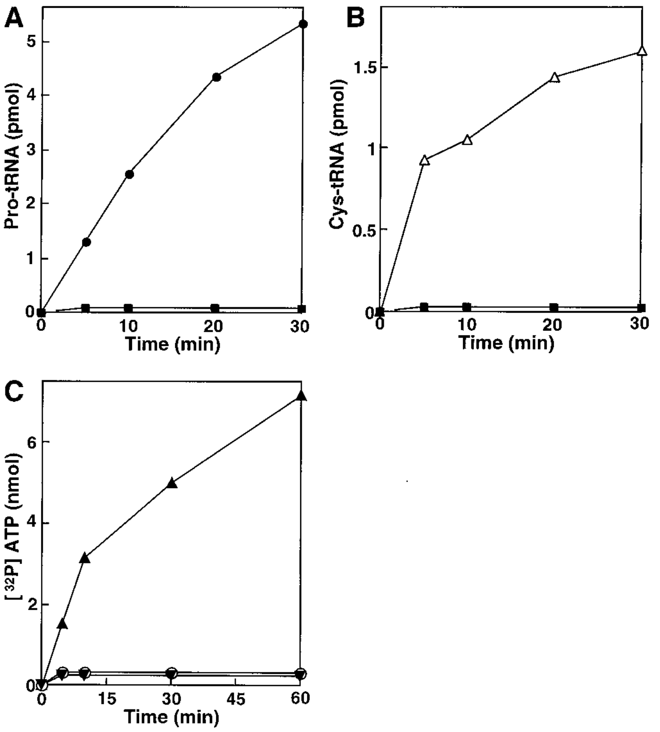

Fig. 1. Aminoacyl-tRNA and Cys-AMP formation by *G. lamblia* ProCysRS. Aminoacylation was performed as described in *Materials and Methods* with 20 μM [^{3}H]proline (*A*, ●) or 20 μM [^{35}S]cysteine (*B*, △). No activity was observed in the absence of either enzyme or tRNA (■). (*C*) Cys-AMP synthesis as measured in the ATP-PP$_i$ exchange reaction. The assay conditions are described in *Materials and Methods*. Cys-AMP was measured in the presence of 1 mg/ml of (▲) or the absence of (▼) total *M. jannaschii* tRNA. ○, The background level (absence of enzyme from the reaction mixture).

***G. lamblia* ProRS Exhibits Dual Specificity *in Vitro* and Requires tRNA for Cysteine Activation.** To determine whether *G. lamblia* ProRS possesses dual aminoacylation specificity, the *G. lamblia proS* gene was recloned with a 5′-terminal His$_6$ epitope tag. Overexpression in *E. coli* and purification over a Ni-nitrilotriacetic acid column led to a His$_6$-tagged protein of the expected molecular mass (61 kDa) and purity (>95%). This enzyme charged unfractionated *M. jannaschii* tRNA almost as well as *Giardia* tRNA; we continued experiments with the archaeal tRNA, as it was easier to obtain. The gene product of the *G. lamblia proS* gene did aminoacylate unfractionated *M. jannaschii* tRNA with proline (Fig. 1*A*) and with cysteine (Fig. 1*B*); thus the enzyme is a prolyl-cysteinyl-tRNA synthetase. In addition to *G. lamblia*, *M. jannaschii*, and *M. maripaludis* tRNAs, the *G. lamblia* ProCysRS also charged *E. coli* tRNA with cysteine.

Earlier it was shown that *M. jannaschii* ProCysRS activates proline in the absence of tRNA, whereas cysteine could only be activated in the presence of tRNA (7; C.S. *et al.,* manuscript submitted for publication), providing a plausible discriminating mechanism in the recognition of the two amino acids. Although the existence of this mechanism is disputed (23), it was interesting to find that the *Giardia* ProCysRS, like the archaeal enzymes, requires tRNA for cysteine activation (Fig. 1*C*), whereas proline activation proceeds in the absence of tRNA (data not shown). Thus, all ProCysRS enzymes may rely on tRNA-dependent cysteine recognition as a mechanism to distinguish between the two cognate tRNAs.

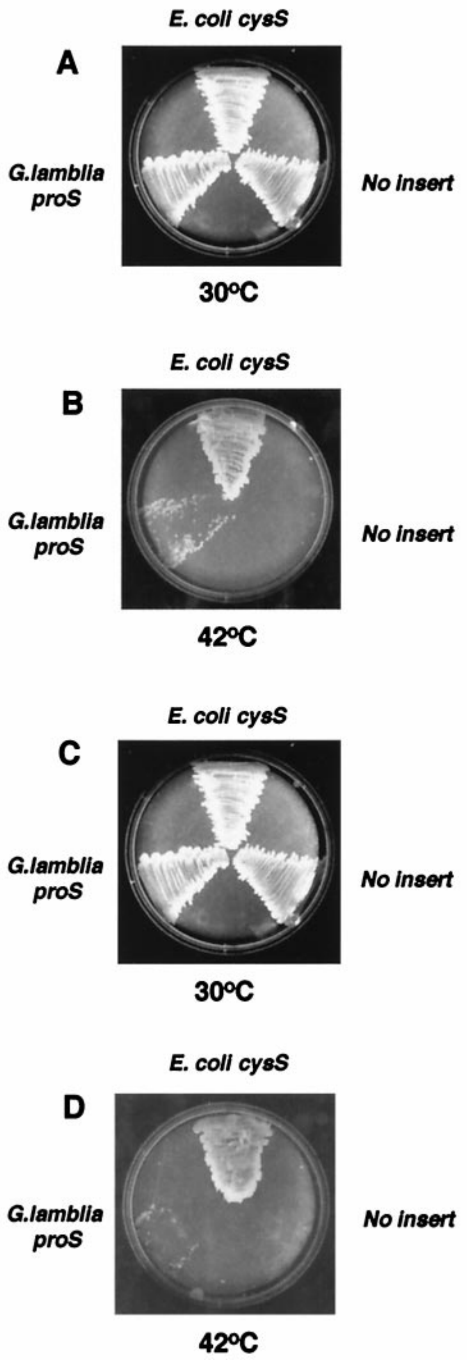

Fig. 2. Complementation of a temperature-sensitive *cysS* mutation in *E. coli* strain UQ818 with the *G. lamblia proS* gene. Each plate was streaked with UQ818 transformed with pCBS1 containing the *G. lamblia proS* gene, the *E. coli* cysS gene, or no insert. Plates were incubated for 3 days at the permissive growth temperature of 30°C (*A* and *C*) and at the nonpermissive growth temperature of 42°C (*B* and *D*). Stronger complementation by the *G. lamblia proS* gene was achieved in the presence of a plasmid for the expression of *M. jannaschii* tRNA^{Cys} (*A* and *B*), whereas in the absence of this plasmid (*C* and *D*) a weaker complementation was observed.

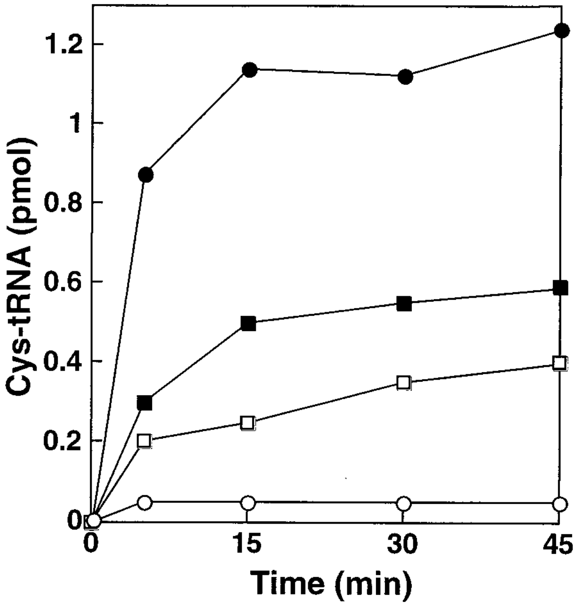

Fig. 3. Cysteinyl-tRNA synthetase activity in *G. lamblia* S-100 cell free extract with homologous tRNA (●), the addition of 2 mM thiaproline (■) or 1 mM 5′-*O*-[*N*-(L-prolyl)-sulfamoyl]adenosine (□), or in the absence of either cysteine or tRNA (○).

G. lamblia **ProRS Synthesizes Cys-tRNA *in Vivo.*** To examine the ability of *G. lamblia* ProRS to synthesize Cys-tRNA *in vivo*, we attempted to rescue growth at a restrictive temperature of a temperature-sensitive *E. coli* cysS strain with the *Giardia proS* gene. As the *M. jannaschii* tRNA was a better substrate for the *Giardia* ProRS than *E. coli* tRNA (see above), coexpression of the genes encoding *M. jannaschii* tRNA and *Giardia* ProRS restored growth of *E. coli* strain UQ818 (18) at 42°C, indicating that ProRS can synthesize Cys-tRNA *in vivo* (Fig. 2 *A* and *B*). The slower growth of the rescued transformants is attributed to the enzyme's preference for homologous tRNA. This conclusion can be tested when the *Giardia* tRNA^{Cys} gene becomes known. *In vivo* complementation proceeds even in the presence of the *E. coli* tRNA^{Cys} gene (Fig. 2 *C* and *D*); however, the poor growth indicates that *G. lamblia* ProRS cannot charge *E. coli* tRNA efficiently. The ability of *G. lamblia* ProRS to synthesize Cys-tRNA both *in vitro* and *in vivo* indicates that this enzyme can specify two amino acids during protein synthesis.

ProCysRS Is Active in *G. lamblia.* In an effort to determine whether both the ProCysRS and the CysRS activities contribute to Cys-tRNA formation in this organism, a *G. lamblia* S-100 cell-free extract was tested for CysRS activity in the presence of *G. lamblia* total tRNA (Fig. 3). In the absence of antibodies against the two enzymes, we attempted to inhibit cysteinylation with the proline analogs, which are not expected to inhibit the canonical CysRS enzyme that may also be expressed in *G. lamblia*. As can be seen, under the conditions tested, the major CysRS activity derives from ProCysRS, as it can be inhibited by thiaproline (24) and 5′-*O*-[*N*-(L-prolyl)-sulfamoyl]adenosine (25). However, the inhibition is not as good as with the pure ProCysRS; thus, there may be a small amount of canonical CysRS activity in *G. lamblia.*

G. lamblia proS **and *alaS*** **Show Archaeal Origin.** ProRS enzymes can be phylogenetically resolved into two main groups: a bacterial group and an archaeal/eukaryal group, which has several bacterial members as well (Fig. 4). Both bacterial and archaeal/eukaryal groups are homologous class IIa aminoacyl-tRNA synthetases, sharing three class-defining motifs in a universally conserved region (26). Surrounding those motifs, however, are significantly diverging primary sequences (2).

The *G. lamblia* ProRS protein described here is specifically related to its archaeal homologs rather than to the eukaryal cytoplasmic ProRS enzymes (Fig. 4; ref. 2). Despite the *Giardia* enzyme's similarity to archaeal ProRS, unmodified *Giardia*

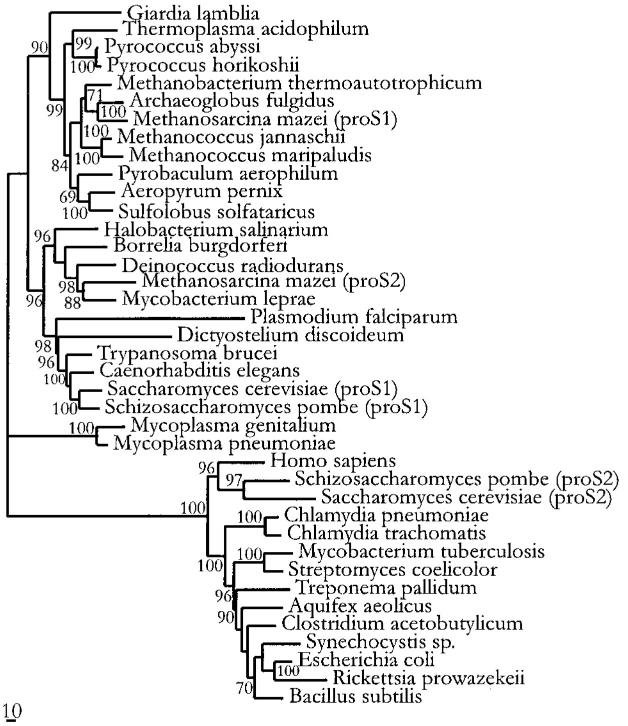

Fig. 4. Phylogenetic tree of ProRS sequences inferred by protein ML analysis. This tree is rooted using threonyl-tRNA synthetase sequences as an outgroup (not shown). Some eukaryotic sequences are distinguished as mitochondrial (mito) or cytoplasmic (cyt). Bootstrap probabilities for each node are estimated by the resampling estimated log-likelihood method. The scale bar represents 10 substitutions per 100 amino acid positions.

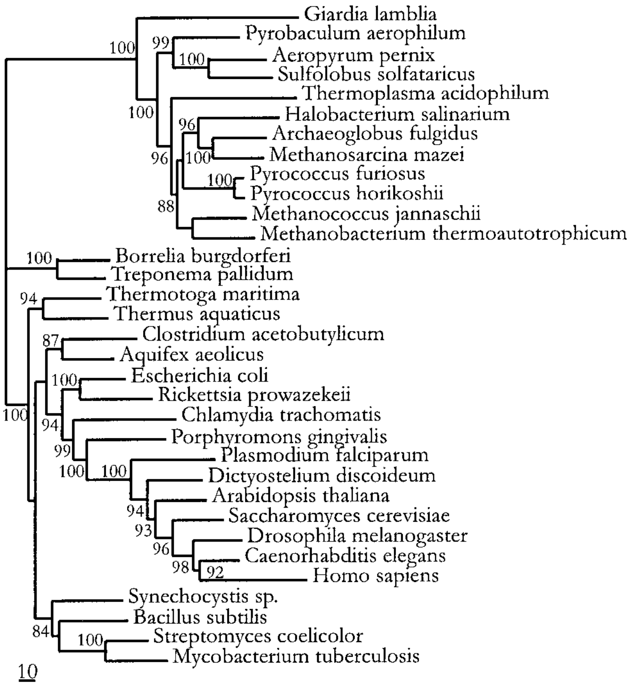

Fig. 5. Phylogenetic tree of AlaRS sequences inferred by protein ML analysis. This tree is rooted using paralogous AlaS2 sequences as an outgroup (not shown). The scale bar represents 10 substitutions per 100 amino acid positions.

tRNA^{Pro} species are 88–92% identical to their human cytoplasmic homologs, differing only in their acceptor stem regions. Conversely, these *Giardia* tRNAs are only 65% identical to archaeal versions (unpublished data). Lateral gene transfer has apparently played a significant role in the evolution of the prolyl-tRNA synthetases. *T. pallidum*, *C. acetobutylicum*, and *M. tuberculosis* all contain bacterial-type ProRS, yet the related species *B. burgdorferi*, *C. sticklandii*, and *M. leprae* each employ archaeal/eukaryal versions. Despite the obvious differences between the two mycobacterial enzymes, unmodified tRNA^{Pro} sequences are identical in *M. tuberculosis* and *M. leprae* (http://www.sanger.ac.uk). Therefore the specific tRNA^{Pro} recognition elements required by the bacterial ProRS are apparently no barrier to gene displacement by an archaeal/eukaryal-type enzyme. Nevertheless, there may be a barrier to displacement of an archaeal/eukaryal-type ProRS by a bacterial ProRS: no archaeon or eukaryote has been found to contain a cytoplasmic bacterial-type ProRS. *M. mazei* contains two *pro*S genes, both archaeal-type, as well as a functional *cys*S gene (15). One *pro*S gene (*proS1* in Fig. 4) from that organism is phylogenetically related to the *A. fulgidus* gene, congruent with the small subunit ribosomal RNA phylogeny. The second *pro*S gene (*proS2*) is related to bacterial members of the archaeal/eukaryal group. It is possible that the two enzymes discriminate among different tRNA species, although their specificities are currently uncharacterized.

In contrast to the archaeal-type ProCysRS found in *G. lamblia*, ProRS from the highly diverged microsporidian *Encephalitozoon cuniculi* is similar to the eukaryotic version. Most aminoacyl-tRNA synthetases from that parasite are related to fungal homologs, consistent with recent phylogenetic studies of other microsporidial genes (27).

The family of alanyl-tRNA synthetases consists of homologous class II enzymes divided into archaeal and bacterial/eukaryal subgroups. The *G. lamblia* AlaRS is an archaeal-type enzyme, which forms an outgroup to the archaeal enzymes in a protein phylogeny (Fig. 5; ref. 2). Two histories would explain these observations. Either the *G. lamblia* AlaRS represents an ancestral eukaryal enzyme, and other eukarya have acquired a bacterial-type AlaRS, or an archaeal-type AlaRS has displaced the eukaryal version in *Giardia*. Additional *alaS* sequence data from diverse eukarya may help in the elucidation of these two hypotheses.

Discussion

Among *Giardia*'s complement of aminoacyl-tRNA synthetases, only AlaRS and ProRS are of archaeal origin, whereas the remaining 18 enzymes bear close resemblance to their eukaryotic relatives. Although there is currently no obvious explanation for the presence of these archaeal synthetases in *G. lamblia,* it is clear that the ProRS, being of the archaeal genre, can provide Cys-tRNA for the organism in addition to that generated by the canonical CysRS. This improved capacity to form Cys-tRNA may be needed in certain circumstances during the life cycle of *G. lamblia*. Cysteine is known to play an important role in this organism; *G. lamblia* trophozoites are coated with cysteine-rich proteins, which, among other roles, function to protect against the hostile gut environment to which *Giardia* is subjected (28, 29). Unlike most eukaryotes, *G. lamblia* produces energy by fermentation, in keeping with its lack of mitochondria and mitochondrial enzymes. The low-oxygen environment of *G. lamblia,* where alanine is the major metabolic product (30), is partially maintained by cysteine and proteins containing cysteine. The pathway of cysteine biosynthesis and its efficiency in *Giardia* are unknown; the currently available genomic sequence data contain only an incomplete fragment of one of the eukaryotic enzymes (cystathionine-β-synthase). On the other hand, *Giardia* may take up cysteine from its environment. Whether the increased capacity for Cys-tRNA formation is reflected in a larger amount of cysteine in *Giardia* proteins will only be known when the genome sequence is complete. However, in archaea that contain both CysRS activities, the cysteine content of the organisms' proteins is not different from those that contain only the canonical CysRS.

Why do some organisms contain two CysRS activities (CysRS and ProCysRS)? The presence of two enzymes for the same activity endows the organism with certain advantages for optimal growth. For instance, it is known for some well-regulated enzymes that their K_M value is within an order of magnitude of the physiological substrate concentration (31). If there were changes in cysteine concentration during the life cycle of *Giardia*, and if ProCysRS and CysRS had different K_M values for cysteine, then Cys-tRNA could be formed by one or the other enzyme, based on the prevailing cellular cysteine concentration, an optimal situation. Whereas the K_M values of *M. jannaschii* ProCysRS for cysteine and proline are comparable to those from ProRS or CysRS enzymes (C.S., unpublished work), the *M. maripaludis* ProCysRS appears to have a lower affinity for cysteine than does the organism's CysRS (C.S., unpublished observation). RNA modifications are sometimes essential for tRNA recognition. Should the two CysRS activities require different RNA modifications for substrate recognition, then *Giardia* could deal better with changes in modification levels caused by different growth rates or gene loss (for modification enzymes) during genome reduction. Future studies of the properties of the enzymes will shed light on some of these questions. In any case, the dual-specificity ProCysRS is an efficient enzyme, as *M. jannaschii* and *M. thermoautotrophicum* are vigorous organisms in which this enzyme is the only means of Cys-tRNA synthesis (32, 33). On the other hand, ProRS activity is usually not duplicated; only one organism (*M. mazei*) is known where a canonical ProRS coexists with the dual-specificity ProCysRS.

ProCysRS is currently the only known dual-specificity AARS. Did this enzyme precede the generation of the canonical CysRS or was the CysRS activity woven into the *proS* context after loss of the *cysS* gene? It is uncertain whether the contemporary ProRSs evolved from the dual-specificity type. However, it is clear that the canonical CysRS enzymes, which are of the class I type and are structurally unrelated to the class II ProRS enzymes, evolved from an ancestor different from the ProCysRS gene. If ProCysRS were the last remaining dual-specificity AARS active in organisms that lack the canonical CysRS, this may suggest that CysRS was the last AARS to have evolved into a single-specificity enzyme. If, however, there was a common ancestor of class II and class I synthetases, proposed as the progenitor of CysRS (34), then it is tempting to speculate that ProCysRS may have had a role in this process.

The existence of two phylogenetically divergent groups of ProRS, bacterial and archaeal/eukaryal types, raises the possibility that the two groups may be functionally divergent as well. Comparisons of the $tRNA^{Pro}$ recognition properties of bacterial-type (*E. coli*) and archaeal/eukaryal-type (human and *Thermus thermophilus*) ProRS enzymes suggest significant differences in tRNA identity elements required for substrate binding and aminoacylation. Both types recognize G(35) and G(36) anticodon bases in $tRNA^{Pro}$ (35) and are capable of efficiently aminoacylating unmodified RNA transcripts *in vitro*. The bacterial-type ProRS also recognizes the conserved nucleotides A(73) and C(1)–G(72) in the $tRNA^{Pro}$ acceptor stem (36), whereas human ProRS requires undetermined elements in both the D- and anticodon stem loops for efficient $tRNA^{Pro}$ aminoacylation (26). Such broadly defined identity elements suggest that the eukaryal ProRS recognizes gross architectural features of its tRNA substrates in addition to the anticodon bases. Broadly defined identity elements would be advantageous for dual-specificity enzymes such as *G. lamblia* and archaeal ProCysRS proteins.

Did the first eukarya have archaeal-type ProRS enzymes that were also responsible for charging $tRNA^{Cys}$, or is this enzyme a unique part of the *Giardia* design fabric? This design fabric, the product of all evolutionary commitments that the lineage has made (37), may constrain major changes in the cell, but is surprisingly open to new invention or gene recruitment when the relevant systems are modular and peripheral to the cell's other systems. The specific charging of tRNAs with their cognate amino acids is imperative for a cell, even though the cell may be indifferent to the mechanism (1, 2).

We thank M. Ibba and C. R. Woese for constructive discussions. We are indebted to S. Cusack for a sample of 5′-O-[*N*-(L-prolyl)-sulfamoyl]adenosine, and to H. D. Becker and T. Li for experimental advice. This work was supported by grants from the National Institute of General Medical Sciences and from National Aeronautics and Space Administration. M.K. was a postdoctoral fellow of the Japanese Ministry for Education.

1. Ibba, M. & Söll, D. (2000) *Annu. Rev. Biochem.* **69,** 617–650.
2. Woese, C. R., Olsen, G., Ibba, M. & Söll, D. (2000) *Microbiol. Mol. Biol. Rev.* **64,** 202–236.
3. Ibba, M. & Söll, D. (1999) *Science* **286,** 1893–1897.
4. Nomanbhoy, T. K., Hendrickson, T. L. & Schimmel, P. (1999) *Mol. Cell* **4,** 519–528.
5. Hong, K-W., Ibba, M., Weygand-Durasevic, I., Rogers, J. M., Thomann, H-U. & Söll, D. (1996) *EMBO J.* **15,** 1983–1991.
6. Ibba, M., Becker, H. D., Stathopoulos, C., Tumbula, D. L. & Söll, D. (2000) *Trends Biochem. Sci.* **25,** 311–316.
7. Stathopoulos, C., Li, T., Longman, R., Vothknecht, U. C., Becker, H., Ibba, M. & Söll, D. (2000) *Science* **287,** 479–482.
8. Tumbula, D. L., Becker, H. D., Chang, W-Z. & Söll, D. (2000) *Nature (London)* **407,** 106–110.
9. Curnow, A. W., Hong, K. W., Yuan, R., Kim, S. I., Martins, O., Winkler, W., Henkin, T. M. & Söll, D. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 11819–11826.
10. Curnow, A. W., Tumbula, D. L., Pelaschier, J. T., Min, B. & Söll, D. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 12838–12843.
11. Upcroft, J. & Upcroft, P. (1998) *BioEssays* **20,** 256–253.
12. Sogin, M. L., Gunderson, J. H., Elwood, H. J., Alonso, R. A. & Peattie, D. A. (1989) *Science* **243,** 75–77.
13. Warren, K. S. (1989) in *New Strategies in Parasitology*, ed. McAdam, K. P. W. J. (Churchill Livingstone, New York), pp. 217–231.
14. Sommer, J. M., Ma, H. & Wang, C. C. (1996) *Mol. Biochem. Parasitol.* **78,** 185–193.
15. Li, T., Graham, D. E., Stathopoulos, C., Haney, P. J., Kim, H.-S., Vothknecht, U., Kitabatake, M., Hong, K.-W., Eggertsson, G., Curnow, A. W., *et al*. (1999) *FEBS Lett.* **462,** 302–306.
16. Ibba, M., Bono, J. L., Rosa, P. A. & Söll, D. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 14383–14388.
17. Keister, D. B. (1983) *Trans. R. Soc. Trop. Med. Hyg.* **77,** 487–488.
18. Bohman, K. & Isaksson, L. A. (1979) *Mol. Gen. Genet.* **176,** 53–55.
19. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680.
20. Adachi, J. & Hasegawa, M. (1996) *Comput. Sci. Monogr.* **28,** 1–150.
21. Kishino, H., Miyata, T. & Hasegawa, M. (1990) *J. Mol. Evol.* **31,** 151–160.
22. Page, R. D. M. (1996) *Comput. Appl. Biosci.* **12,** 357–358.
23. Lipman, R. S., Sowers, K. & Hou, Y. M. (2000) *Biochemistry* **39,** 7792–7798.
24. Busiello, V., di Girolamo, M., Cini, C. & De Marco, C. (1979) *Biochim. Biophys. Acta* **564,** 311–321.
25. Yaremchuk, A., Kriklivyi, I., Cusack, S. & Tukalo, M. (2000) *Acta Crystallogr. D* **56,** 197–199.
26. Stehlin, C., Burke, B., Yang, F., Liu, H., Shiba, K. & Musier-Forsyth, K. (1998) *Biochemistry* **37,** 8605–8613.
27. Van de Peer, Y., Ali, A. B. & Meyer, A. (2000) *Gene* **246,** 1–8.
28. Nash, T. E. (1992) *Parasitol. Today* **8,** 229–234.
29. Gillin, F. D., Hagblom, P., Harwood, J., Aley, S. B., Reiner, D. S., McCaffery, M., So, M. & Guiney, D. G. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 4463–4467.
30. Adam, R. D. (1991) *Microbiol. Rev.* **55,** 706–732.
31. Fersht, A. (1985) *Enzyme Structure and Mechanism* (Freeman, New York), pp. 327–330.
32. 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., *et al.* (1996) *Science* **273,** 1058–1073.
33. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., *et al.* (1997) *J. Bacteriol.* **179,** 7135–7155.
34. Avalos, J., Corrochano, L. M. & Brenner, S. (1991) *FEBS Lett.* **286,** 176–180.
35. Yaremchuk, A., Cusack, S. & Tukalo, M. (2000) *EMBO J.* **19,** 4745–4758.
36. McClain, W. H., Schneider, J. & Gabriel, K. (1994) *Nucleic Acids Res.* **22,** 522–529.
37. Graham, D. E., Overbeek, R., Olsen, G. J. & Woese, C. R. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 3304–3308.