temperature and filtered through a 0.45-μm cellular acetate membrane and stored at 4 °C. We diluted concentrated solutions of DNA in the gelatin solution to keep the gelatin concentration >0.17%. Unless otherwise specified, final plasmid DNA concentrations were 0.033 μg μl^{-1}. For the Genestorm clones, the final DNA concentration was 0.040 μg μl^{-1}.

Reverse transfection of microarrays

We used the Effectene transfection kit (301425; Qiagen) as follows. In a 1.5-ml microcentrifuge tube, 16 μl enhancer was added to 150 μl EC buffer, mixed, and pre-incubated for 5 min at room temperature. 25 μl Effectene lipid was added, mixed and the entire volume pipetted onto a 40 × 20-mm cover well (PC200; Grace Bio-Labs). A slide with the printed side down was placed on the cover well such that the solution covered the entire arrayed area while also creating an airtight seal. After a 10–20-min incubation, the cover well was prised off the slide with forceps and the transfection reagent removed carefully by vacuum aspiration. The slide was placed printed side up in a 100 × 100 × 10 mm square tissue culture dish and 1 × 10^7 actively growing HEK293T cells in 25 ml medium (DMEM with 10% IFCS (inactivated fetal calf serum), 50 units ml^{-1} penicillin and 50 μg ml^{-1} streptomycin) were poured into the dish. Three slides can be transfected side by side in this fashion. The cells grew on the slide for 40 h before fixing for 20 min at room temperature in 3.7% paraformaldehyde/4.0% sucrose in PBS. We have also tested other commonly used mammalian cells lines, such as HeLa and A549 cells, and obtained similar results but with transfection efficiencies of 30–50% of those obtained with HEK293T cells.

Immunofluorescence

For immunofluorescence staining, the cells were fixed as above, permeabilized in 0.1% Triton X-100 in PBS for 15 min and probed with primary and secondary antibodies as described[21]. Primary mouse monoclonal antibodies (500 μl per slide) were used at the following concentrations: 1:500 anti-HA ascites fluid (BaBCo), 2 μg ml^{-1} anti-Myc 9E-10 (Calbiochem), 2 μg ml^{-1} anti-V5 (Invitrogen), and 10 μg ml^{-1} 4G10 anti-phosphotyrosine (Upstate Biotechnologies). Primary rabbit polyclonal antibodies (500 μl per slide) were used at the following concentrations: 0.7 μg ml^{-1} anti-p38 pTGpY (Promega), 0.7 μg ml^{-1} anti-JNK pTPpY (Promega), and 0.5 μg ml^{-1} anti-ERK1/2pTEpY (NEB). Secondary antibodies were Cy3-labelled anti-mouse or anti-rabbit antibodies produced in donkeys (Jackson Immunoresearch) and were used at 3.1 μg ml^{-1}.

Laser scanning and fluorescence microscopy

We imaged microarrays at a resolution of 5 μm with a laser fluorescence scanner (ScanArray 5000; GSI Lumonics). GFP and Cy3 emission was measured separately after sequential excitation of the two fluorophores. To obtain images at cellular resolution, cells were photographed with a conventional fluorescence microscope. To measure fluorescence intensity, cell clusters were photographed and the signal intensity quantified with Image Quant (Fuji). All images were pseudocoloured and superimposed using Photoshop 5.5 (Adobe Systems).

FK506 and SCH23390 autoradiography

We added 5 nM dihydro-FK506 [propyl-^{3}H] (NEN) to the cell medium containing the microarrays for 1 h before rinsing once with PBS at room temperature and fixation. When used, rapamycin was added at 1 μM for 30 min before FK506 addition. Before autoradiography, slides were processed for anti-Myc immunofluorescence, scanned at 5-μm resolution and photographed using a fluorescence microscope. We then exposed slides to tritium-sensitive film (Hyperfilm, Amersham) for 4 days. Autoradiographical emulsion was performed as described by the manufacturer (EM-1; Amersham). Image Quant (Fuji) was used to quantify 0.01-mm^2 areas of autoradiograms. For the dopamine receptor experiments, 1 nM SCH23390 [*N*-methyl-^{3}H] (NEN) was added to the cell medium containing the microarrays for 1 h before rinsing once with PBS at room temperature and fixation. Before processing for autoradiography as above, we photographed GFP fluorescence using a fluorescence microscope.

Received 13 November 2000; accepted 26 March 2001.

1. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270,** 467–470 (1995).
2. Lockhard, D. J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* **14,** 1675–1678 (1996).
3. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403,** 623–627 (2000).
4. Kino, T. *et al.* FK-506, a novel immunosuppressant isolated from a Streptomyces. II. Immunosuppressive effect of FK-506 *in vitro*. *J. Antibiot.* (*Tokyo*) **40,** 1256–1265 (1987).
5. Harding, M. W., Galat, A., Uehling, D. E. & Schreiber, S. L. A receptor for the immunosuppressant FK506 is a cis-trans peptidyl-prolyl isomerase. *Nature* **341,** 758–760 (1989).
6. Siekierka, J. J., Hung, S. H., Poe, M., Lin, C. S. & Sigal, N. H. A cytolsolic binding protein for the immunosuppressant FK506 has peptidyl-prolyl isomerase activity but is distinct from cyclophilin. *Nature* **341,** 755–757 (1989).
7. Schreiber, S. L. Chemistry and biology of the immunophilins and their immunosuppressive ligands. *Science* **251,** 283–287 (1991).
8. Jacinto, E., Werlen, G. & Karin, M. Cooperation between Syk and Rac1 leads to synergistic JNK activation in T lymphocytes. *Immunity* **8,** 31–41 (1998).
9. Gavrieli, Y., Sherman, Y. & Ben-Sasson, S. A. Identification of programmed cell death *in situ* via specific labeling of nuclear DNA fragmentation. *J. Cell Biol.* **119,** 493–501 (1992).
10. Walczak, H. *et al.* TRAIL-R2: a novel apoptosis-mediating receptor for TRAIL. *EMBO J.* **16,** 5386–5397 (1997).
11. Tandon, N. N., Kralisz, U. & Jamieson, G. A. Identification of glycoprotein IV (CD36) as a primary receptor for platelet-collagen adhesion. *J. Biol. Chem.* **264,** 7576–7583 (1989).
12. Simonsen, H. & Lodish, H. F. Cloning by function: expression cloning in mammalian cells. *Trends Pharmacol. Sci.* **15,** 437–441 (1994).
13. Strausberg, R. L., Feingold, E. A., Klausner, R. D. & Collins, F. S. The mammalian gene collection. *Science* **286,** 455–457 (1999).
14. Kawai, J. *et al.* Functional annotation of a full-length mouse cDNA collection. *Nature* **409,** 685–690 (2001).
15. Yudate, H. T. *et al.* HUNT: launch of a full-length cDNA database from the Helix Research Institute. *Nucleic Acids Res.* **29,**185–188 (2001).
16. Wiemann, S. *et al.* Toward a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs. *Genome Res.* **11,** 422–435 (2001).
17. MacBeath, G. & Schreiber, S. L. Printing proteins as microarrays for high-throughput function determination. *Science* **289,** 1760–1763 (2000).
18. Lueking, A. *et al.* Protein microarrays for gene expression and antibody screening. *Anal. Biochem.* **270,** 103–111 (1999).
19. de Wildt, R. M., Mundy, C. R., Gorick, B. D. & Tomlinson, I. M. Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nature Biotechnol.* **9,** 989–994 (2000).
20. Morgenstern, J. P. & Land, H. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res.* **18,** 3587–3596 (1990).
21. Sabatini, D. M. *et al.* Interaction of RAFT1 with gephyrin required for rapamycin-sensitive signaling. *Science* **284,** 1161–1164 (1999).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank C. Wilson and T. Volkert for advice on use of the microarrayer, and K. McKernan and P. McEwan for the neurotransmitter receptor constructs. We also thank R. Young, N. Hacohen and members of the Sabatini laboratory for their support and suggestions; P. Kim and M. Gerlach for help with the manuscript; and Invitrogen for the full-length clones from the Genestorm collection. This research was supported by the G. Harold and Leila Y. Mathers Charitable Foundation, the Whitehead Institute Fellows Program and Corning. Microscopy work was conducted at the W. M. Keck Foundation Biological Imaging facility at the Whitehead Institute.

Correspondence and requests for materials should be addressed to D.M.S. (e-mail: sabatini@wi.mit.edu).

An aminoacyl tRNA synthetase whose sequence fits into neither of the two known classes

Carme Fàbrega*, Mark A. Farrow*, Biswarup Mukhopadhyay†, Valérie de Crécy-Lagard*, Angel R. Ortiz‡ & Paul Schimmel*

** The Skaggs Institute for Chemical Biology, The Scripps Research Institute, Beckman Center, 10550 North Torrey Pines Road, La Jolla, California 92037, USA*
† Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
‡ Department of Physiology & Biophysics, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1218, New York, New York 10029, USA

Aminoacyl transfer RNA synthetases catalyse the first step of protein synthesis and establish the rules of the genetic code through the aminoacylation of tRNAs. There is a distinct synthetase for each of the 20 amino acids and throughout evolution these enzymes have been divided into two classes of ten enzymes each[1,2]. These classes are defined by the distinct architectures of their active sites, which are associated with specific and universal sequence motifs[1–5]. Because the synthesis of aminoacyl-tRNAs containing each of the twenty amino acids is a universally conserved, essential reaction, the absence of a recognizable gene for cysteinyl tRNA synthetase in the genomes of Archae such as *Methanococcus jannaschii* and *Methanobacterium thermoautotrophicum*[6–8] has been difficult to interpret. Here we describe a different cysteinyl-tRNA synthetase from *M. jannaschii* and

***Deinococcus radiodurans* and its characterization *in vitro* and *in vivo*. This protein lacks the characteristic sequence motifs seen in the more than 700 known members of the two canonical classes of tRNA synthetase and may be of ancient origin. The existence of this protein contrasts with proposals that aminoacylation with cysteine in *M. jannaschii* is an auxiliary function of a canonical prolyl-tRNA synthetase[9,10].**

The ten class I enzymes have a Rossmann nucleotide-binding fold with highly conserved motifs such as the 11-amino-acid signature sequence that ends in HIGH and, separately, the KMSKS pentapeptide[3–5]. Class II enzymes have three distinct motifs that are part of a seven-stranded β-structure with three α-helices[1,2]. All known synthetases fit strictly into one of these two classes, so that they are easily recognizable by sequence homology searches with programs such as BLAST or PSI-BLAST[11]. For example, all cysteinyl tRNA synthetases (CysRSs) can be identified as close homologues that fall into class I. However, the genomes of some Archaea lack recognizable genes for cysteinyl-tRNA synthetase, even though CysRS activity is present in crude extracts[12].

It has been proposed[9] that another tRNA synthetase replaces cysteinyl tRNA synthetase in *M. jannaschii*. In particular, it was reported[9,10] that a canonical class II prolyl-tRNA synthetase (termed a ProCysRS) can catalyse aminoacylation with both cysteine and proline. (Interestingly, *Giardia lamblia* encodes a ProCysRS but nonetheless retains a canonical CysRS[13]). Because the presence of ProCysRS raises questions about how one enzyme can accurately maintain aminoacylation with two stereochemically distinct amino acids and two different tRNAs, we searched for an unconventional gene that might encode cysteinyl tRNA synthetase in *M. jannaschii*. Considering that previous methods had failed, we developed a new computational technique for orthologue annotation.

First, we analysed the genome of *M. jannaschii* by comparison with the well-characterized genomes of *Escherichia coli*, *Bacillus subtilis* and *Haemophilus influenza* (see Methods). The idea was to start with well-annotated genes in *M. jannaschii* (about 350 or 20% of the complete genome) and find their most likely orthologues in the other three organisms. The degree of functional matching was evaluated with the 'μ-score', a measure of sequence similarity between two sequences in two different genomes designed so that the higher the μ-score, the higher the likelihood of orthology (see Methods). Conversely, low μ-scores are mainly associated with protein pairs that have different functions. A cutoff (μ_c) of about 2.5 was found reliably to separate pairs that had the same function from those that had different functions (Fig. 1). A μ-score of about 2.5 would therefore correspond to a protein pair having on average the lowest possible sequence similarity compatible with keeping the same function in two given genomes. For pairs of distant organisms, this threshold value can potentially relate sequence pairs falling well below the upper limit of the 'twilight zone', the point where sequence similarity is so weak that the trace back to a common ancestor becomes unreliable[14,15].

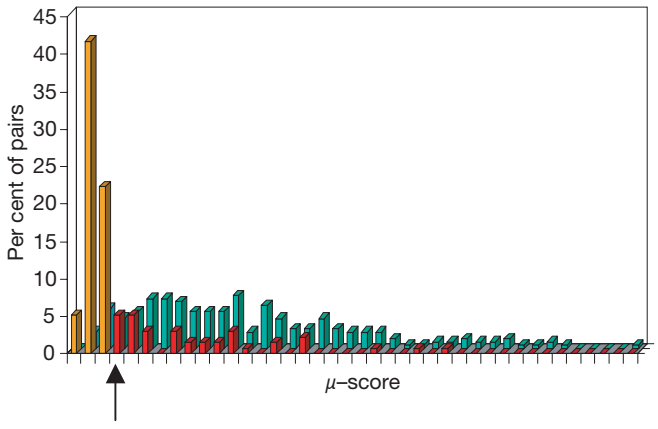

Figure 1 Calibration of the orthologue detection procedure for the *M. jannaschii* genome (see Methods). Pairs of proteins with the same function are represented with green histograms; protein pairs with different functions are in red and orange histograms. The cutoff value of $\mu_c = 2.5$ (arrow) identifies the threshold in the μ-score that can eliminate the bulk of the true negatives (orange bars) while keeping most of the true positives (green bars) in the functional annotation. Complete genome sequences and their functional annotations were downloaded from the European Bioinformatics Institute web site.

Next, we compared the remaining (about 1,300) genes of *M. jannaschii* with each gene in the whole genomes of the three aforementioned organisms. The idea was to find one or more sequences in *M. jannaschii* that had a μ-score of about $\mu_c = 2.5$ when compared against sequences in the other three organisms. A match of this sort would be a candidate for the missing gene for cysteinyl tRNA synthetase. We found a single match of $\mu = 2.4$ between open reading frame MJ1477 in *M. jannaschii* and the cysteinyl tRNA synthetase in the genome of *B. subtilis* (SYC_BACSU). The MJ1477 coding sequence has 346 amino acids, and the open reading frame did not have significant matches with any other genes in the three genomes.

The gene for MJ1477 (*mj1477*) was cloned and its encoded protein was expressed and purified from *E. coli* (see Methods). We then assayed the CysRS activity of MJ1477 in two ways. Typically, aminoacylation occurs in two steps. In the first step, the amino acid (AA) is condensed with ATP to form the enzyme-bound aminoacyl adenylate (E-AA~AMP). Next, the adenylate reacts with tRNA to form aminoacyl-tRNA (AA-tRNA). In one assay, we monitored the cysteine-dependent exchange of $^{32}PP_i$ into ATP and in the second we investigated overall aminoacylation (formation of Cys-tRNA).

The *in vitro* ATP-PP_i exchange assay (Fig. 2a) showed that MJ1477 can activate cysteine in the presence of ATP to yield the corresponding Cys-adenylate (Cys-AMP). The activity is specific for cysteine, because activation does not occur when the isosteric amino acid serine is used instead (Fig. 2a). Activation occurs in the presence as well as in the absence of tRNA. (In contrast, the ProCysRS characterized in ref. 9 activates cysteine only in the presence of tRNA.) We then investigated the specificity of the reaction with respect to the tRNA. *Escherichia coli* $tRNA^{Cys}$ purified by high-performance liquid chromatography (HPLC) was efficiently aminoacylated (Fig. 2b). This aminoacylation was specific for $tRNA^{Cys}$, because HPLC fractions that were not charged by *E. coli* CysRS were similarly not aminoacylated by MJ1477. Next, and to rule out the possibility of contamination from *E. coli*, we looked at similarities and differences between the *E. coli* and *M. jannaschii* proteins (Fig. 2c). *Escherichia coli* CysRS has some activity with *M. jannaschii* tRNA at 37 °C (considerably less than with its natural substrate), but cannot aminoacylate *M. jannaschii* tRNA at 70 °C. In contrast, purified MJ1477 can aminoacylate *M. jannaschii* tRNA at both 37 and 70 °C (as might be expected for an enzyme from a thermophile). Together, these results strongly support the proposal that MJ1477 is the enzyme catalysing the synthesis of Cys-$tRNA^{Cys}$.

Attempts to demonstrate the activity of MJ1477 in *E. coli* by complementation of UQ818 cells (ref. 16) containing a temperature-sensitive allele of *CysS* were unsuccessful because of the toxicity of the expressed protein in the test strain (data not shown). To investigate the activity of MJ1477 further, we determined the kinetic parameters for the activation of cysteine and for the aminoacylation of *M. jannaschii* tRNA. We assayed the rate of synthesis of Cys-AMP by ATP-PP_i exchange (at 70 °C) to give a catalytic constant (k_{cat}) of 0.96 s^{-1} and a Michaelis constant (K_m) (for cysteine) of 125 μM. (The calculated k_{cat} for aminoacylation at 70 °C under the conditions used was greater than 0.09 s^{-1}.) These values compare with reported initial rates of 0.132 s^{-1} for Cys activation and 0.0055 s^{-1} for aminoacylation (at 70 °C) with the ProCysRS from the same organism[10].

We used the protein-coding sequence of MJ1477 to search for

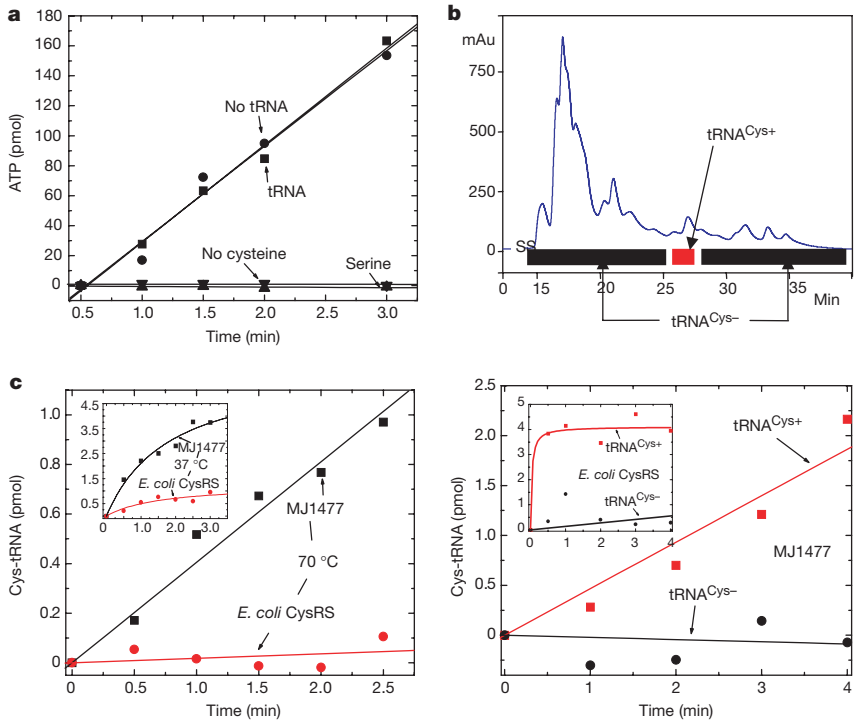

Figure 2 Experimental characterization of the enzymatic activity of MJ1477. **a**, Specific activation of Cys by *M. jannaschii* CysRS at pH 7.0, 70 °C in the presence of 10 mM cysteine and 56 μM *M. jannachii* tRNA (squares); in the absence of *M. jannaschii* tRNA (circles); in the presence of 10 mM serine (up triangles; no enzymatic activity detected); and in presence of the enzyme alone (down triangles; no enzymatic activity detected). **b**, Top, HPLC purification of *E. coli* $tRNA^{Cys}$, with identification of fractions charged with cysteine by *E. coli* CysRS. Bottom, aminoacylation by MJ1477 or *E. coli* CysRS (insert) of *E. coli* $tRNA^{Cys}$ ($tRNA^{Cys+}$) (squares) and *E. coli* tRNA which had been depleted of $tRNA^{Cys}$ ($tRNA^{Cys-}$) (circles). **c**, Comparison of aminoacylation activity of *E. coli* and *M. jannaschii* CysRS with natural, modified *M. jannaschii* tRNA (56 μM) at 70 °C. Aminoacylation of *M. jannaschii* $tRNA^{Cys}$ was performed in the presence of MJ1477 (squares) or of *E. coli* CysRS (circles). Insert, same experiment at 37 °C, in the presence of MJ1477 (squares) or *E. coli* CysRS (circles).

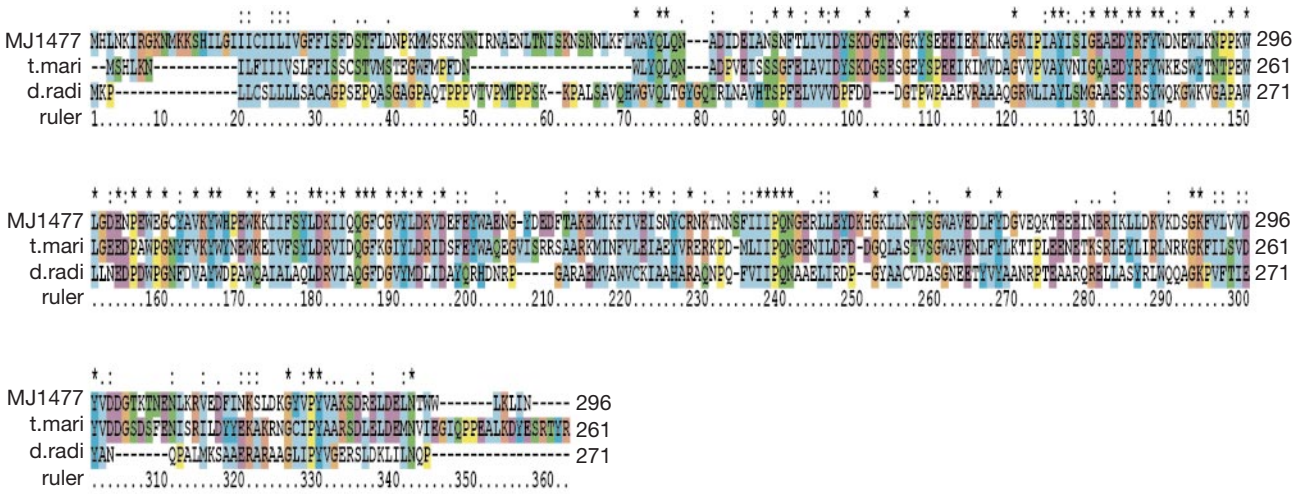

Figure 3 Multiple sequence alignment of MJ1477 with the close homologues detected by PSI-BLAST[11]. The alignment was obtained with CLUSTALW[17] using default parameters. No class I or class II tRNA synthetases were detected during the database search (see text). Colours follow CLUSTALW convention[17]. t.mari, *T. maritima*; d.radi, *D. radiodurans*.

homologous sequences in a non-redundant sequence database using PSI-BLAST[11]. We found two statistically significant hits in the genomes of the eubacteria *Thermotoga maritima* (a thermophile) and *D. radiodurans* (a mesophile). We carried out a multiple sequence alignment (MSA) of the three sequences using the program CLUSTALW[17] with default parameters (Fig. 3). (From this alignment, MJ1477 and the *T. maritima* enzyme appear to be the most similar in sequence of the three apparent homologues.) Using polymerase chain reaction (PCR) and a sample of genomic DNA from *D. radiodurans*, we cloned the MJ1477 homologue from *D. radiodurans* (designated DR0705) and tested it for complementation of *E. coli* strain UQ818. Complementation of the temperature-sensitive lethal *CysS* mutation was observed (Fig. 4). Thus, DR0705 is active *in vivo* in protein synthesis as a cysteinyl tRNA synthetase.

The orthologue detection approach using known sequences of

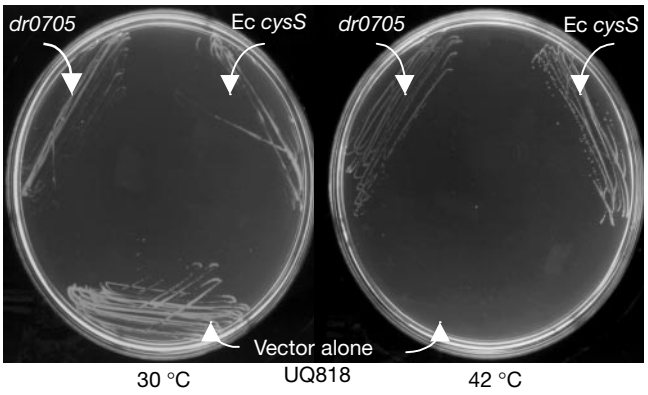

Figure 4 Complementation of the growth defect at 42 °C of *E. coli* strain UQ818. Cells were transformed with vector alone or with plasmids bearing *E. coli cysS* or *D. radiodurans dr0705*, and streaked onto LB plates containing IPTG (1 mM) and ampicillin (100 μg ml^{-1}). Plates were incubated for 24 h at 30 or 42 °C.

class I enzymes has led to the identification of a tRNA synthetase whose sequence bears little or no resemblance to those of known class I enzymes. (It is possible that a three-dimensional structure of MJ1477 or DR0705 would reveal an overall fold similar to that of class I enzymes, but with substantial changes in the active site.) The average per cent identity of MJ1477, for example, with the canonical CysRSs is 17% (not shown), falling well below the twilight zone, conventionally situated in the region of 25–30%[14]. This percentage of identity (17%) is significantly lower than the 38% pairwise identity between sequences of canonical CysRSs (not shown). Thus, the alignments provide no support for MJ1477 or its homologues being class I enzymes. More importantly, the hallmarks of class I tRNA synthetases[3–5], including the 11-amino-acid signature sequence ending in HIGH and the KMSKS pentapeptide, are missing from the coding sequences of the new CysRSs. These enzymes also lack signatures of the known class II enzymes. Therefore, by the conventional criteria of sequence, MJ1477 and its homologues are the first synthetases to fit into neither of the two known classes of these enzymes.

Thermotoga maritima and *D. radiodurans* present both types of cysteinyl-tRNA synthetase. Small-subunit ribosomal RNA phylogeny has placed *Thermotoga*, together with *Aquifex*, as the deepest and most slowly evolving branch in the tree of life[18]. It is possible that the canonical CysRS was acquired later in evolution, and the slow evolutionary pace of *Thermotoga* is the reason why the organism still has both copies. It is also noteworthy that archae such as *M. thermoautotrophicum* lack the novel CysRS. Perhaps the newly identified CysRS is an ancient relic that was gradually eliminated from most organisms. Regardless of the detailed explanation that will eventually emerge, MJ1477 and its homologues are, to our knowledge, the first tRNA synthetases with sequences distinct from those of every other synthetase described. □

Methods

Computational method for orthologue annotation

We used the following protocol to develop the annotation method. (1) Selection of a training set of well-annotated sequences in *M. jannaschii, E. coli, B. subtilis* and *H. influenzae*. As described in the text, we used about 350 genes, or about 20% of the genome of *M. jannaschii*. (2) For each gene in *M. jannaschii*, the most likely orthologue in the other three genomes was assigned to the sequence with the highest μ-score (see below). (3) The matches obtained were classified into true positives (same function) and true negatives (different function). (4) The cutoff value in the μ-score (μ_c) that optimizes discrimination of true positives versus true negatives for this training set was determined (Fig. 1) using discriminant analysis[19].

The μ-score is obtained as follows: during each of the comparisons described above, sequences were aligned using the Needleman and Wunsch algorithm[20] with zero end gaps, using a normalized Gonnet matrix and a threshold ensuring structural similarity[21]. From the sequence alignment score, we developed an orthology likelihood score (μ-score) on the basis of two premises: first, for two proteins (in two different organisms) to have an orthological relationship they have to be at least structurally related (in addition to having amino acids dictating the same fold, they should have an additional set of similar residues that allows them to carry out a similar function); and second, we assumed the orthological relationship to be biunivocal (only one pair of orthologues exists between two different organisms). With these premises, we defined the μ-score simply as the maximum value of the σ_A score that a protein in one organism can have when compared with all proteins in another organism. In turn, the σ_A score is defined as the number of times the sequence similarity between the two sequences exceeds the minimum value that ensures, on average, a statistically significant structural similarity between both sequences, as derived from training sets of sequence-structure matches[21]. The 'excess' of sequence similarity (over the minimum that provides structural similarity[21]) needed for the pair to have the same function can then be obtained by discriminant analysis (as described above) to yield μ_c.

Cloning and expression of MJ1477 and DR0705

We obtained a clone of the putative CysRS gene from *M. jannaschii* (AMJEC64) (encoding MJ1477) from the American Type Culture Collection (ATCC). *mj1477* was cloned by PCR; for expression and purification of MJ1477, it was subcloned into the maltose-binding protein (MBP) modified expression vector pGEX (Pharmacia) to generate plasmid pMMJ1477. *Escherichia coli* TG1[22] was transformed with pMMJ1477, and grown to mid-logarithmic phase at 37 °C in 1 l of Lennox broth containing 100 μg ml^{-1} ampicillin. The cells were then induced by the addition of 0.5 mM isopropyl 1-thio-β-*D*-galactopyranoside and incubated for an additional 90 min. Following cell lysis the fusion protein was cleaved using factor Xa and the enzyme purified by a modification of established procedures[23]. We determined enzyme concentrations by active site titration[24].

We used similar procedures for cloning by PCR of the putative CysRS (Fig. 3) from *D. radiodurans*, using a sample of chromosomal DNA (ATCC13939) from ATCC. The PCR product was cloned into plasmid pTRC99A[25] to give plasmid pVDC564, where the *D. radiodurans* gene is under control of an IPTG-inducible promoter. We confirmed the primary structure of the cloned gene (*dr0705*) by direct sequencing methods.

Complementation analysis

Complementation at 42 °C of the growth defect of the temperature-sensitive *CysS* allele of *E. coli* strain UQ818 (ref. 16) (*lacZ4 gyrA222*(nal^R) *aroE24 metB rpoB*(rif^R) *cysS818*^{ts}) was done by standard procedures on LB/ampicillin plates with 1 mM IPTG, using UQ818 cells transformed with the vector (pTRC99A) alone, the vector bearing *dr0705* (pVDC564) and a positive control with the gene for *E. coli* CysRS cloned into a pUC18 vector (gift from G. Eriani).

Enzyme assays and materials

Methanococcus jannaschii tRNA was obtained by phenol extraction from *M. jannaschii* cells[26], followed by ion-exchange chromatography (Nucleobond) and desalted with a Centricon 10 concentrator. Total *E. coli* tRNA (Roche Molecular Biochemicals) was purified by HPLC[27]. Fractions containing tRNA that could be aminoacylated with cysteine (by *E. coli* CysRS) were pooled and desalted using a Centricon-10 concentrator. The remaining tRNA was pooled and desalted. The RNA concentration was determined by absorbance at 260 nm. Cysteinyl adenylate synthesis was assessed with the cysteine-dependent ATP-PP$_i$ exchange assay[28] at 70 °C in 25 mM HEPES (pH 7.0), 10 mM KF, 25 mM KCl, 3.75 mM $MgCl_2$, 2 mM dithiothreitol (DTT), 1 mM ATP, 1 mM [^{32}P]NaPP$_i$. The kinetic parameters for amino-acid activation were determined using Cys concentrations of 50–2,000 μM. Aminoacylation assays[29] were performed in in 25 mM Hepes (pH 7.0), 25 mM KCl, 7.5 mM $MgCl_2$, 2.5 mM DTT, 1 mM ATP, 50 μM [^{35}S]-cysteine, 250 nM MJ1477 or 250 nM *E. coli* CysRS.

Received 31 January; accepted 2 March 2001.

1. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347,** 203–206 (1990).
2. Cusack, S., Berthet-Colominas, C., Hartlein, M., Nasar, N. & Leberman, R. A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. *Nature* **347,** 249–255 (1990).
3. Webster, T., Tsai, H., Kula, M., Mackie, G. A. & Schimmel, P. Specific sequence homology and three-dimensional structure of an aminoacyl transfer RNA synthetase. *Science* **226,** 1315–1317 (1984).
4. Hountondji, C., Dessen, P. & Blanquet, S. Sequence similarities among the family of aminoacyl-tRNA synthetases. *Biochimie* **68,** 1071–1078 (1986).
5. Ludmerer, S. W. & Schimmel, P. Gene for yeast glutamine tRNA synthetase encodes a large amino-terminal extension and provides a strong confirmation of the signature sequence for a group of the aminoacyl-tRNA synthetases. *J. Biol. Chem.* **262,** 10801–10806 (1987).
6. Bult, C. J. *et al.* Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273,** 1058–1073 (1996).
7. Smith, D. R. *et al.* Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J. Bacteriol.* **179,** 7135–7155 (1997).
8. Reeve, J. N. Archaebacteria then . . . Archaes now (are there really no archaeal pathogens?). *J. Bacteriol.* **181,** 3613–3617 (1999).
9. Stathopoulos, C. *et al.* One polypeptide with two aminoacyl-tRNA synthetase activities. *Science* **287,** 479–482 (2000).
10. Lipman, R. S., Sowers, K. R. & Hou, Y. M. Synthesis of cysteinyl-tRNA^{Cys} by a genome that lacks the normal cysteine-tRNA synthetase. *Biochemistry* **39,** 7792–7798 (2000).
11. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
12. Hamann, C. S., Sowers, K. R., Lipman, R. S. & Hou, Y. M. An archaeal aminoacyl-tRNA synthetase missing from genomic analysis. *J. Bacteriol.* **181,** 5880–5884 (1999).

13. Bunjun, S. *et al.* A dual-specificity aminoacyl-tRNA synthetase in the deep-rooted eukaryote *Giardia lamblia*. *Proc. Natl Acad. Sci. USA* **97,** 12997–13002 (2000).
14. Chung, S. Y. & Subbiah, S. A structural explanation for the twilight zone of protein sequence homology. *Structure* **4,** 1123–1127 (1996).
15. Rost, B. Twilight zone of protein sequence alignments. *Protein Eng.* **12,** 85–94 (1999).
16. Bohman, K. & Isaksson, L. A. Temperature-sensitive mutants in cysteinyl-tRNA ligase of *E. coli* K-12. *Mol. Gen. Genet.* **176,** 53–55 (1979).
17. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680 (1994).
18. Woese, C. R. Interpreting the universal phylogenetic tree. *Proc. Natl Acad. Sci. USA* **97,** 8392–8396 (2000).
19. Johnson, R. A. & Wichern, D. W. *Applied Multivariate Statistical Analysis* (Prentice Hall, Upper Saddler River, New Jersey, 1992).
20. Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48,** 443–453 (1970).
21. Abagyan, R. A. & Batalov, S. Do aligned sequences share the same fold? *J. Mol. Biol.* **273,** 355–368 (1997).
22. Sambrook, J., Fitsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).
23. Glasfeld, E., Landro, J. A. & Schimmel, P. C-terminal zinc-containing peptide required for RNA recognition by a class I tRNA synthetase. *Biochemistry* **35,** 4139–4145 (1996).
24. Fersht, A. R. *et al.* Active site titration and aminocyl adenylate binding stoichiometry of aminoacyl-tRNA synthetases. *Biochemistry* **14,** 1–4 (1975).
25. Amann, E., Ochs, B. & Abel, K. J. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene* **69,** 301–315 (1988).
26. Mukhopadhyay, B., Johnson, E. F. & Wolfe, R. S. Reactor-scale cultivation of the hyperthermophilic methanarchaeon *Methanococcus jannaschii* to high cell densities. *Appl. Environ. Microbiol.* **65,** 5059–5065 (1999).
27. Hale, S. P., Auld, D. S., Schmidt, E. & Schimmel, P. Discrete determinants in transfer RNA for editing and aminoacylation. *Science* **276,** 1250–1252 (1997).
28. Calender, R. & Berg, P. Purification and physical characterization of tyrosyl ribonucleic acid synthetases from *Escherichia coli* and *Bacillus subtilis*. *Biochemistry* **5,** 1681–1690 (1966).
29. Shepard, A., Shiba, K. & Schimmel, P. RNA binding determinant in some class I tRNA synthetases identified by alignment-guided mutagenesis. *Proc. Natl Acad. Sci. USA* **89,** 9964–9968 (1992).

Acknowledgements

This work was supported by grants from the National Institutes of Health (P.S. and to R. S. Wolfe) and the Department of Energy (to R. S. Wolfe) and by a fellowship from the National Foundation for Cancer Research (P.S.). Mount Sinai School of Medicine start-up funds are acknowledged (A.R.O.). We thank Y.-M. Hou for *E. coli* CysRS protein and R. Turner for advice. We thank S. Cusack for reading the manuscript before submission.

Correspondence and requests for materials should be addressed to P.S. (e-mail: schimmel@scripps.edu) or A.R.O. (e-mail: ortiz@inka.mssm.edu).