Cysteinyl-tRNA synthetase: determination of the last *E.coli* aminoacyl-tRNA synthetase primary structure

Gilbert Eriani*, Guy Dirheimer and Jean Gangloff
Institut de Biologie Moléculaire et Cellulaire du CNRS and Université Louis Pasteur, Laboratoire de Biochimie, 15 rue René Descartes, 67084 Strasbourg, France

Received November 7, 1990; Revised and Accepted December 14, 1990 EMBL accession no. X56234

ABSTRACT

The gene coding for *E.coli* cysteinyl-tRNA synthetase (*cysS*) was isolated by complementation of a strain deficient in cysteinyl-tRNA synthetase activity at high temperature (43°C). Sequencing of a 2.1 kbp DNA fragment revealed an open reading frame of 1383 bp coding for a protein of 461 amino acid residues with a Mr of 52,280, a value in close agreement with that observed for the purified protein, which behaves as a monomer. The sequence of CysRS bears the canonical His-Ile- Gly -His (HIGH) and Lys-Met-Ser-Lys-Ser (KMSKS) motifs characteristic of the group of enzymes containing a Rossmann fold; furthermore, it shows striking homologies with MetRS (an homodimer of 677 residues) and to a lesser extent with Ile-, Leu-, and ValRS (monomers of 939, 860, and 951 residues respectively). With its monomeric state and smaller size, CysRS is probably more closely related to the primordial aminoacyl-tRNA synthetase from which all have diverged.

INTRODUCTION

The aminoacyl-tRNA synthetases (aaRS) are a class of enzymes which play a crucial role in the accurate translation of genetic information : they promote the attachment of the appropriate amino acid to each tRNA molecule in a highly specific two step catalytic process. Despite the fact that they catalyse the same reaction (the formation of an aminoacyl bond), these enzymes differ widely in size and oligomeric state, and exhibit relatively low homologies at the level of primary structure. The amino acid sequences of nearly 50 aminoacyl-tRNA synthetases from different origins are known, however only four crystallographic structures have been determined: they include tyrosyl-tRNA synthetase from *B.stearothermophilus* (1,2), methionyl-tRNA synthetase from *E.coli* (3,4), glutaminyl-tRNA synthetase from *E.coli* complexed with its cognate tRNA (5), and more recently the structure of the seryl-tRNA synthetase from *E.coli* (6). The first three share a common structural feature called the Rossmann fold, which contains the two signature sequences HIGH and KMSKS which are involved in ATP and tRNA binding respectively (7,8). Solution of the crystallographic structure of SerRS (6), an enzyme whithout a Rossmann fold and built around a beta-sheet made of seven antiparallel strands, resumed interest after the recent sequence analysis which led to the classification of the nineteen known aaRS into two families (9). This partition in two classes, established on the basis of two mutually exclusive sets of sequence motifs, is found to be strongly correlated on the functional level with the acylation process, occurring either on the 2′-OH (class I, bearing the HIGH and KMSKS motifs) or 3′-OH (class II, bearing the three new motifs) of the ribose of the 3′-end of the tRNA (9). In order to complete the partition, the gene coding for CysRS from *E.coli* was cloned and its primary structure determined. The sequence of CysRS shows strong homologies with MetRS, and to a lesser extent with the group of enzymes charging amino acids encoded by XUX codons composed of Ile-, Leu- and ValRS.

MATERIALS AND EXPERIMENTAL PROCEDURES

Strains

Strain UQ818 (cysS818, aroE, metA, rpoB, (Rif^r), lac, thi, nalA) is a temperature-sensitive cysteinyl-tRNA synthetase mutant of *E.coli* K12 (10). Strain TGI (Δ (lac-proAB), supE, endA, sbcB15, hsdR4, F′ (traD36, proAB, lacIq, LacZΔM15)), and strain JM103 (Δ (lac-proAB), thi, strA, supE, endA, sbcB15, hsdR4, F′ (traD36, proAB, lacIq, lacZΔM15)) were used as recipients for plasmid DNA transformation, and bacteriophage M13 DNA transfection respectively.

Enzymes, chemicals, nucleic acids, and standard procedures

Restriction enzymes, T4 DNA ligase, T4 polynucleotide kinase, T4 DNA polymerase and terminal deoxynucleotidyl transferase were from Boehringer Mannheim. Modified T7 DNA polymerase was from Pharmacia and Moloney Leukaemia Virus reverse transcriptase from BRL. All enzymes were used according to the manufactor's instructions. [^{35}S] α-dATP, [^{32}P] γ-ATP, [^{32}P] inorganic pyrophosphate and [^{35}S] L-cysteine were from Amersham. DEAE Sephacel was from Pharmacia, TSK HW65S was from Merck. Standard procedures were used for plasmid amplification, isolation and ligation of DNA fragments (11).

* To whom correspondence should be addressed

Cloning and sequencing of *cysS*

The UQ818 mutant was grown in 50 ml LB medium at 30°C and transformed by the $CaCl_2$ procedure with 1 μg of a recombinant plasmid library of partially digested Sau3A-HpaII fragments (3–5 kbp) of *E.coli* nuclear DNA ligated into the BamHI-ClaI cloning sites of pBR322 (12). Incubation was performed at the non-permissive temperature of 43°C.

The 0.35kpb and 1.75 kbp HindIII DNA fragments of pCysS2 were cloned in both orientations into the bacteriophage M13mp18. The complete nucleotide sequence was determined by the dideoxy-DNA sequencing method of Tabor & Richardson (13) after subcloning of the inserts by the procedure of Dale et al. (14). DNA and protein sequence were analysed by using the programs of the UWGCG (version June 1990).

Mapping of the 5′ terminus mRNA

Total bacterial RNA was extracted as described (14). A 19-mer oligonucleotide (5′-GATCGTAAACGGTGATTCC-3′), complementary of the *cysS* mRNA sequence was synthesized on an Applied Biosystem 381A apparatus, and 5′- end labelled with [^{32}P] ATP using T4 polynucleotide kinase. Primer extension reactions were carried out on 100 μg total RNA as described (15).

Purification, amino terminal sequencing and molecular weight determination of cysteinyl-tRNA synthetase

Purification of CysRS was performed starting from 2 liters LB medium containing 200 μg ampicillin per ml. Cell growth, disruption by sonication, supernatant clarification and DEAE Sephacel chromatography were performed as previously described (16). The active fractions were pooled, and precipitated in 80% saturation of ammonium sulphate at 0°C. Precipitated protein were recovered by centrifugation at 10,000×g for 10 min, dissolved in a minimal volume of 50 mM Tris-HCl, pH 7.5 and loaded on a TSK HW65S column (1.6×32 cm) equilibrated in 50 mM Tris-HCl, pH 7.5, 2.4 M ammonium sulphate.

The protein eluted with a linear gradient of 2.4 to 0 M ammonium sulphate in the same Tris-HCl buffer, was estimated pure enough (more than 95%) for N-terminal protein sequencing; this was performed by sequential Edman degradation on a 470A protein sequencer coupled to a 3120A PTH analyzer (Applied Biosystems). This allowed identification of the first ten amino acids of the protein.

The molecular weight was determined by HPLC gel filtration on a TSK 3000 SW column (0.75 cm×30 cm). Elution was conducted at 25°C with a 50 mM potassium phosphate buffer pH 7.5, 100 mM NaCl, at a flow rate of 0.4 ml per min. Amounts of 40 μg of CysRS and of each standard proteins were loaded on the column. The effluents were continuously monitored for the presence of proteins by measurement of their absorbance at 280 nm.

Measurements of kinetic parameters

Aminoacylation reaction: the reaction mixture contained 50 mM Tris-HCl pH7.5, 30 mM KCl, 10 mM $MgCl_2$, 5 mM DTT, 0.1 mM [^{35}S]L-cysteine (25,000 cpm/nmole), 2 mM ATP, 6 mg/ml unfractionated *E.coli* tRNA (it contains about 1% $tRNA^{Cys}$), 5 mM dithio-erythritol, 0.1 mg/ml bovine serum albumin and appropriate amounts of enzyme. After various incubation times at 37°C, [^{35}S]-cysteinyl-$tRNA^{Cys}$ was assayed in 40 μl of the reaction mixture by the method previously described (16) except that glass fiber discs were used to reduce the background level.

ATP-PPi exchange reaction: The standard reaction mixture contained 100 mM sodium Hepes pH 7.2, 10 mM $MgCl_2$, 2 mM ATP, 2 mM [^{32}P]-PPi (1–2 cpm/pmole), 5 mM L-cysteine, 10 mM KF (to inhibit pyrophosphatase activity), and enzyme. After various incubation times at 37°C, the [^{32}P]-ATP was assayed in 50 μl aliquots as described previously (16) and the initial rates were determined.

The specific activities are defined here as the number of

```
-143                                          TCAACCCAGTTCGGGTCATATAT
-120  AGGGTGGTGTTATAGCATAACCGCACGATCGGATCATCACGCAATGTATGCTGATTCGCG
 -60  CGGGAAATATGGGTATTATACGCAACTCAATTACCCACACATGTCTAAACGGAATCTTCG

   1  ATGCTAAAAATCTTCAATACTCTGACACGCCAAAAAGAGGAATTTAAGCCTATTCACGCC
      M  L  K  I  F  N  T  L  T  R  Q  K  E  E  F  K  P  I  H  A     20

  61  GGGGAAGTCGGCATGTACGTGTGTGGAATCACCGTTTACGATCTCTGTCATATCGGTCAC
      G  E  V  G  M  Y  V  C  G  I  T  V  Y  D  L  C  H  I  G  H     40

 121  GGGCGTACCTTTGTTGCTTTTGACGTGGTTGCGCGCTATCTGCGTTTCCTCGGCTATAAA
      G  R  T  F  V  A  F  D  V  V  A  R  Y  L  R  F  L  G  Y  K     60

 181  CTGAAGTATGTGCGCAACATTACCGATATCGACGACAAAATCATCAAACGCGCCAATGAA
      L  K  Y  V  R  N  I  T  D  I  D  D  K  I  I  K  R  A  N  E     80

 241  AATGGCGAAAGCTTTGTGGCGATGGTGGATCGCATGATCGCCGAAATGCACAAAGATTTT
      N  G  E  S  F  V  A  M  V  D  R  M  I  A  E  M  H  K  D  F    100

 301  GATGCTTTGAACATTCTGCGCCCGGATATGGAGCCGCGCGCGACGCACCATATCGCAGAA
      D  A  L  N  I  L  R  P  D  M  E  P  R  A  T  H  H  I  A  E    120

 361  ATTATTGAACTCACTGAACAACTGATCGCCAAAGGTCACGCTTATGTGGCGGACAACGGC
      I  I  E  L  T  E  Q  L  I  A  K  G  H  A  Y  V  A  D  N  G    140

 421  GACGTGATGTTCGACGTCCCGACCGATCCAACTTATGGCGTGCTGTCGCGTCAGGATCTC
      D  V  M  F  D  V  P  T  D  P  T  Y  G  V  L  S  R  Q  D  L    160

 481  GACCAGCTGCAGGCAGGCGCGCGCGTTGACGTGGTCGACGACAAACGCAACCCAATGGAC
      D  Q  L  Q  A  G  A  R  V  D  V  V  D  D  K  R  N  P  M  D    180

 541  TTCGTTCTGTGGAAGATGTCGAAAGAGGGCGAACCGAGCTGGCCGTCTCCGTGGGGCGCG
      F  V  L  W  K  M  S  K  E  G  E  P  S  W  P  S  P  W  G  A    200

 601  GGTCGTCCTGGCTGGCACATTGAATGTTCGGCAATGAACTGCAAGCAGCTGGGTAACCAC
      G  R  P  G  W  H  I  E  C  S  A  M  N  C  K  Q  L  G  N  H    220

 661  TTTGATATCCACGGCGGCGGTTCAGACCTGATGTTCCCGCACCACGAAAACGAAATCGCG
      F  D  I  H  G  G  G  S  D  L  M  F  P  H  H  E  N  E  I  A    240

 721  CAGTCCACCTGTGCCCATGATGGTCAGTATGTGAACTACTGGATGCACTCGGGGATGGTG
      Q  S  T  C  A  H  D  G  Q  Y  V  N  Y  W  M  H  S  G  M  V    260

 781  ATGGTTGACCGCGAGAAGATGTCCAAATCGCTGGGTAACTTCTTTACCGTGCGCGATGTG
      M  V  D  R  E  K  M  S  K  S  L  G  N  F  F  T  V  R  D  V    280

 841  CTGAAATACTACGACGCGGAAACCGTGCGTTACTTCCTGATGTCGGGCCACTATCGCAGC
      L  K  Y  Y  D  A  E  T  V  R  Y  F  L  M  S  G  H  Y  R  S    300

 901  CAGTTGAACTACAGCGAAGAGAACCTGAAGCAGGCGCGTGCGGCGCTGGAGCGTCTCTAC
      Q  L  N  Y  S  E  E  N  L  K  Q  A  R  A  A  L  E  R  L  Y    320

 961  ACTGCGCTGCGCGGCACAGATAAAACCGTTGCGCCTGCCGGTGGCGAACGCTTTGAAGCG
      T  A  L  R  G  T  D  K  T  V  A  P  A  G  G  E  R  F  E  A    340

1021  CGCTTTATTGAAGCGATGGACGACGATTTCAACACCCCGGAAGCCTATTCCGTACTGTTT
      R  F  I  E  A  M  D  D  D  F  N  T  P  E  A  Y  S  V  L  F    360

1081  GATATGGCGCGTGAAGTAAACCGTCTGAAAGCAGAAGATATGGCAGCGGCGAATGCAATG
      D  M  A  R  E  V  N  R  L  K  A  E  D  M  A  A  A  N  A  M    380

1141  GCATCTCACCTGCGTAAACTTTCCGCTGTATTGGGCCTGCTGGAGCAAGAACCGGAAGCG
      A  S  H  L  R  K  L  S  A  V  L  G  L  L  E  Q  E  P  E  A    400

1201  TTCCTGCAAAGCGGCGCGCAGGCAGACGACAGCGAAGTGGCTGAGATTGAAGCGTTAATT
      F  L  Q  S  G  A  Q  A  D  D  S  E  V  A  E  I  E  A  L  I    420

1261  CAACAGCGTCTGGATGCCCGTAAAGCGAAAGACTGGGCGGCGGCGGATGCGGCGCGTGAT
      Q  Q  R  L  D  A  R  K  A  K  D  W  A  A  A  D  A  A  R  D    440

1321  CGTCTTAACGAGATGGGGATCGTGCTGGAAGATGGCCCGCAAGGGACCACCTGGCGTCGT
      R  L  N  E  M  G  I  V  L  E  D  G  P  Q  G  T  T  W  R  R    460

1381  AAGTAATTGCGCTATTGCCGGATGCGAGTTTTCGCATCCGGTTATCGTCTGCGCCACCAC
      K  *                                                          461

1441  AACATTCCCATCAGTAGCATCCCCGGCAACCACACCCACATCAATTCAGAAATAATCACC
1501  TGATGCCCGTACGGCGTGGTGTAACGAGACAATGCAAACGGCGCGACTTTTATCACCTGC
1561  CAGGGAGCGAAAAAGCGTTCATCTGACCACGCCACAGCCAGCCAACGCCTTTACCGCCAG
1621  TGGTTACCGAATCCAGCAAGCTGTGCGATAGCAACGAGACGGTTAAAAACAGCCAGCAGC
1681  GAATCAGCCCAGCCCTGAACCATCGGCGTCCAATAAACACACATAACAGCGGGACAACAA
1741  ACGCAAACACCAGCGAATGGGTAAACCCGCGATGACCAAAAACATTGCCGTAAGCAACGC
1801  CAAATTTAAACGACAATACGTCGGCGTCGGGCAGCATCGCCAGGATGATTCCGGCAAATA
1861  ACAGACGCGGAGGGATGACTTTCGAACCCAACCCTAAACCAATGCATAGGGGAACGGCGG
1921  CGTGCGTAATAACGGTTGGCATGATGGTCGCTTCGGCAAAATGTCGATGCTATCAGCATG
1981  GATGAACGGGGCGTAGAGGGCAAAAGTCTGAAAAGAGAACCGGCCTGTTGATACAGGCCG
2041  GGAAAGGATC  2051
```

Figure 1: Nucleotide sequence and deduced amino acid sequence of the *cysS* gene. Numbering of nucleotides starts at the first residue of the ATG codon encoding the initiating methionine. Standard single-letter amino acid code is used. The vertical arrow indicate the location of the 5′-end of the *cysS* transcript; underlining shows possible regulatory sequences like −10/−35 transcription control elements and a possible stop transcription signal.

exchanged pyrophosphate or charged $tRNA^{Cys}$ molecules in one minute, per mg of pure CysRS and under the above assay conditions.

RESULTS AND DISCUSSION

Cloning and sequencing of *cysS*, characteristics of the CysRS

A pBR322 library of E.coli DNA was used to complement the strain UQ818 which is defective in CysRS activity at 43°C. Fifteen colonies grew on plates after transformation (see Materials and Methods); plasmid DNA prepared from these clones showed, after PstI restriction analysis, that they all contained overlapping DNA fragments (of 5 kbp and more) corresponding to the same chromosomal region. Crude extracts prepared from these clones exhibited more than a 20-fold increase in cysteine dependent ATP-pyrophosphate exchange activity compared to the recipient UQ818. We isolated the plasmidic DNA from one of these clones and submitted it to limited Sau3A digestion. Partial digestion products were fractionated by agarose gel electrophoresis, the 2.0–3.5 kbp DNA fragments were recovered and ligated into the BamHI site of the vector pUC18 and used to re-transform strain UQ818. Selection of ampicillin resistant clones at 43°C led to the isolation of plasmids with inserts of different sizes, the smallest of which being 2.1 kbp long. We isolated the 0.35 and 1.75 HindIII DNA fragments from this plasmid, which were subcloned in both orientations into M13mp18. The recombinant bacteriophages were submitted to the procedure of partial exonuclease digestion described by Dale et al. (14) in order to generated a set of overlapping fragments for sequencing with modified T7 DNA polymerase (13). The DNA fragments which have been completely sequenced on both strands revealed the presence of a long open reading frame on one strand (Fig.1); the deduced amino acid sequence starting from the first in-frame methionine is 461 residues long; its N-terminal end was independently confirmed by sequencing the ten first N-terminal residues of the purified enzyme.

The molecular weight of the CysRS calculated from its primary structure (Mr= 52,280) is in good agreement with the value determined by SDS polyacrylamide gel electrophoresis (data not shown). Analysis of the pure enzyme by gel filtration under non-denaturating conditions clearly shows that CysRS behaves as a monomer (Fig. 2). The extinction coefficient of CysRS at 280 nm, established from its aminoacid composition, is 1.09 $mg^{-1}.ml^{-1}.cm^{-1}$. Its specific activity values relative to the acylation and the ATP-PPi exchange reactions are respectively 1560 and 6880.

Analysis of the *cysS* mRNA terminus

The initiation site of transcription was determined by the primer extension method using Moloney Leukaemia Virus reverse transcriptase, total cellular RNA and a 5′-end labelled synthetic oligonucleotide. Extension products were fractionated on an urea sequencing gel; figure 3 shows one major reverse transcriptase stop located at the position −34. Figure 1 shows a TATA sequence 7 nucleotides upstream of the mRNA 5′-terminus and a TTCGC sequence 28 nucleotides upstream of the mRNA initiation site (TTGAC-like sequence); these distances strongly suggest that both sequences correspond to transcription control elements.

Analysis of the DNA region downstream of the TGA stop signal revealed a G/C rich sequence of hyphenated dyad symmetry, centered on position +1411, and followed by several T. This structural feature could correspond to a rho-independant termination signal for *E.coli* RNA polymerase.

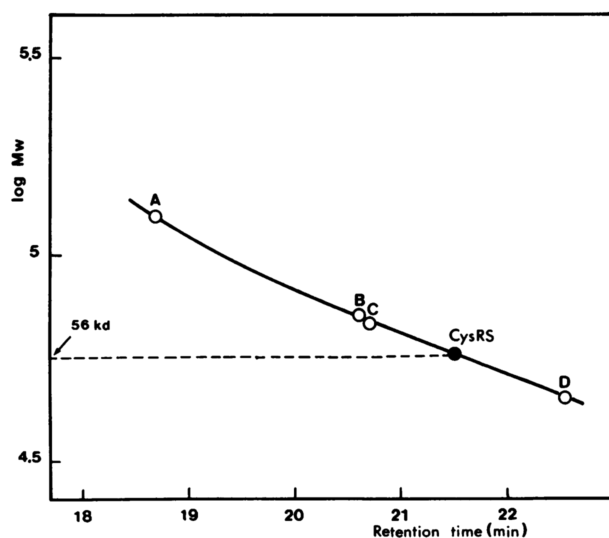

Figure 2: Molecular weight determination of native CysRS. The HPLC gel filtration on the TSK 3000 SW column was achieved as described in the text. The calibration curve was established with the following standard proteins: A, aspartyl-tRNA synthetase from *E.coli*, dimer (Mr= 122,000); B, arginyl-tRNA synthetase from *E.coli* (Mr= 70,000); C, bovine serum albumin, monomer (Mr= 68,000); D, ovalbumin (Mr= 45,000).

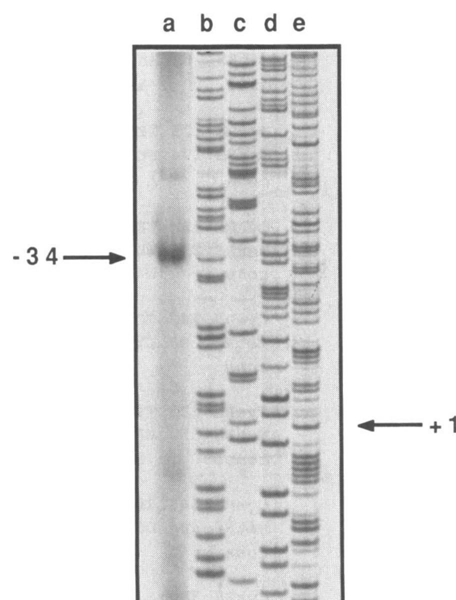

Figure 3: Determination of the 5′-ends of *cysS* transcript. Analysis of the 5′-ends of the *cysS* transcript by primer extension was carried out as described in 'Experimental Procedures'. After MLV-reverse transcriptase extension, the products were fractionated on a 6% polyacrylamide-urea sequencing gel. Lane a, primer extension of total RNA in the presence of [^{32}-P]phosphorylated primer; lanes b–e, dideoxy-sequencing pattern of *cysS* in the presence of the primer (phosphorylated) used for 5′-mapping (ACGT respectively). The numbers refer to positions in the DNA sequence (Fig.1).

Codon Usage in *cysS* and activity of the enzyme

A strong correlation has been observed between the degree of codon bias and the level of gene expression (17): very highly expressed genes contain almost exclusively those codons corresponding to the most abundant tRNAs, whereas codons recognized by rare tRNAs are rarely present. We used the index P2′ defined by Sharp and Li (18) to estimate the degree of expression of the CysRS gene. The calculated value of P2′ (= 0,5) corresponds to a moderately expressed gene with a low codon bias. We also detected an abundant usage of the Ser codon UCG (32% of the Ser codons; in strongly expressed genes: 8% (18)) and the Gly codon GGG (16%; 3% in strongly expressed genes). These observations are in good agreement with the moderate accumulation of the CysRS protein we observed in crude extracts (about 5% of the total cytoplasmic proteins versus 20% for ArgRS (15) and 30% for ProRS (9) .

Sequence comparisons

CysRS bears the canonical HIGH and KMSKS sequences

We found in the primary structure of CysRS the signature sequences HIGH (residues 37−40) and KMSKS (residues 266−270) which are common to nine other aaRS specific for the amino acids arginine, glutamine, glutamate, isoleucine, leucine, methionine, tryptophane, tyrosine and valine (15, 19−26)(enzymes of class I; Figure 4). A third region of similarity was found in the characteristic DWCISRQ region (26) conserved in six other aaRS, but not found in Gln-, Glu-, Tyr-and TrpRS.

Homologies with other aminoacyl-tRNA synthetases

The sequence comparison of the CysRS sequence with other aaRS shows extensive homologies with MetRS (Fig. 5). 95 amino acids residues are conserved between the two proteins. The overall identity is 22% of the aligned positions, a value which leads us

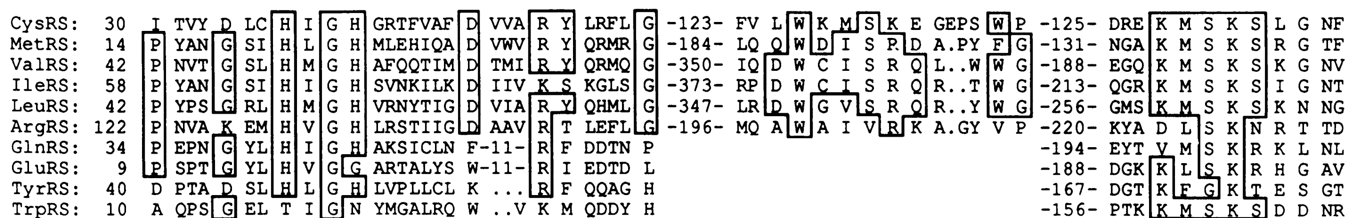

Figure 4: Alignment of the signature sequences of aminoacyl-tRNA synthetases of class I. The ten *E.coli* aaRS of class I are represented. The position of the first residue shown is indicated as well as the distance between the consensus sequences; conserved residues are boxed.

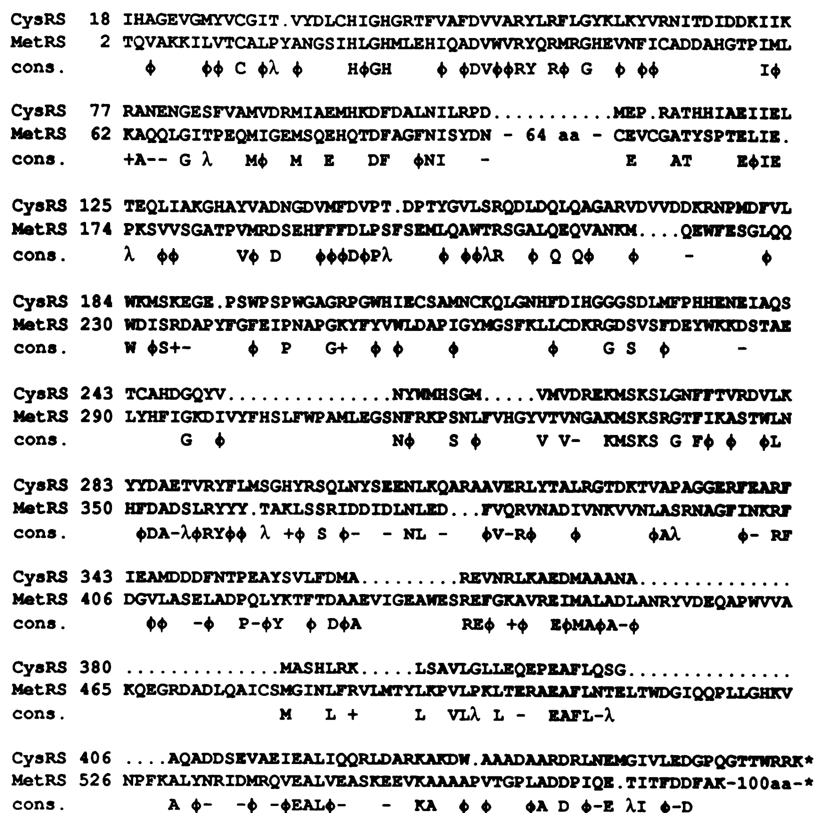

Figure 5: Alignment between *E. coli* CysRS and MetRS , obtained by the UWGCG Bestfit program; the alignment of the two proteins was then manually adjusted. Conserved residues in the three sequences are indicated by a single letter code for the invariant residue, 'λ' for a small amino acid -P, G, S, T-, '−' for a negative charge or homologous polar residues -D, E, N, Q-, '+' for a positive charge -H, R, K- and 'ϕ' for an hydrophobic residue -F, Y, W, I, L, V, M or A.

to expect similarities at the tertiary level. The homologies are distributed along the entire length of the two molecules, broken occasionally by insertions in the MetRS sequence: we notice the presence of a large insertion from residue number 95 to 158 which could correspond to a part of the variable connective peptide as defined by Starzyk et al. (27). In MetRS and GlnRS, the crystallographic data have shown that the two halves of the Rossmann fold are interrupted by a structural domain corresponding to this insertion, and in the case of GlnRS it is folded into 3 alpha-helices and 5 beta-strands (4,5). The function of this region, in GlnRS, is to distort the acceptor end of the tRNA : it corresponds to the tRNA acceptor binding domain (5). In MetRS its structure is less regular and its function is not clearly established; in CysRS the size of this domain seems to be greatly reduced. In the second part of the molecule six major insertions are found in MetRS when compared to CysRS. In MetRS, this region is organized in an independent domain, which is mainly helical and involved in the binding of $tRNA^{Met}$ through a specific interaction of Trp461 with the C residue of the anticodon CAU (28). Such variability between the two enzymes in this region may suggest a different mode of binding of $tRNA^{Cys}$ by CysRS in order to ensure specific aminoacylation. The remaining similarities are found at the carboxy terminal part of the tryptic fragment of MetRS, in the region folded towards the first domain and involved in the formation of a bridge between the N- and C-terminal domains; the indispensable character of this peptide has been demonstrated by site directed mutagenesis (29–30). A 100-residues long extension is found at the C- terminal end of MetRS; it corresponds to the oligomerization domain (31) and is not present in CysRS. Additional homologies are found with Ileu-, Leu- and ValRS according to the alignment proposed with MetRS by Heck and Hatfield (26). Large additional domains are found in these three enzymes which may be the result of gene fusion and duplication from the ancestral gene during evolution. This original molecule would have possessed the characteristic Rossmann fold and would have evolved in a divergent evolutionary process to give rise to the present class I aaRS, kinases and dehydrogenases.

ACKNOWLEDGEMENTS

We would like to thank M. Springer (IBPC, Paris) for the gift of the strain UQ818, S. Gasparini for having carryied out the N-terminal analysis of the CysRS, and F. Wittinghofer (MPI, Heidelberg) for the gift of the gene bank. The skilful technical assistance of M.L. Gangloff is gratefully aknowledged. We are gratefull to L. Pallank for helpful discussion and correction of the manuscript. This work was supported by the Association pour la Recherche sur le Cancer and the Fondation pour la Recherche Médicale.

REFERENCES

1. Bhat, T.N., Blow, D.M. and Brick, P. (1982) J. Mol. Biol. **158**, 699–710.
2. Brick, P., Bhat, T.N. and Blow, D.M. (1988) J. Mol. Biol. **208**, 83–98.
3. Risler, J.L., Zelwer, C. and Brunie, S. (1981) Nature **292**, 384–386.
4. Zelwer, C., Risler, J.-L. and Brunie, S. (1982) J. Mol. Biol. **155**, 63–81.
5. Rould, M.A., Perona, J.J., Söll, D. and Steitz, T.A. (1989) Science **246**,1135–1142.
6. Cusack, S., Berthet-Colominas, C., Härtlein, M., Nassar, N. and Leberman, R. (1990) Nature **347**, 249–255.
7. Webster, T.A., Tsai, H., Kula, M., Mackie, G.A. and Schimmel, P. (1984) Science **226**, 1315–1317.
8. Hountondji, C., Dessen, P. and Blanquet, S. (1986) Biochimie **68**, 1071–1078.
9. Eriani, G., Delarue, M., Poch, O., Gangloff, J. and Moras, D. (1990) Nature **347**, 203–206.
10. Bohman, K. and Isaksson, L.A. (1979) Molec. Gen. Genet. **176**, 53–55.
11. Maniatis, T., Fritsh, E.F. and Sambrook, J. (1989). Molecular cloning: a laboratory manual. Second edition. Cold Spring Harbour Laboratory, Cold Spring Harbor, New York.
12. Brune, M., Schumann, R. and Wittinghofer, F. (1985) Nucleic Acids Res. **13**, 7139–7147.
13. Tabor, R. and Richardson, C.C. (1987) Proc. Natl. Acad. Sci. USA **84**, 4767–4771.
14. Dale, R.M.K., McClure, B.A. and Houchins, J.P. (1985) Plasmid **13**, 31–40.
15. Eriani, G., Dirheimer, G. and Gangloff, J. (1989) Nucleic Acids Res. **17**, 5725–5735.
16. Eriani, G., Dirheimer, G. and Gangloff, J. (1990) Nucleic Acids Res. **18**, 1475–1479.
17. Ikemma, T. and Ozeki, H. (1982) Cold Spring Harbor Symposia on Quantitative Biology,vol. XLVIII, 1087–1097.
18. Sharp, P.M. and Li, W.-H. (1986) Nucleic Acids Res. **14**, 7737–7749.
19. Hoben, P., Royal, N., Cheung, A., Yamao, F., Bieman, K. and Söll, D. (1982) J. Biol. Chem. **257**, 11644–11650.
20. Breton, R., Sanfaçon, H., Papayannopoulos, I., Bieman, K. and Lapointe, J. (1986) J. Biol. Chem. **261**, 10610–10617.
21. Webster, T.A., Tsai, H., Kula, M., Mackie, G.A. and Schimmel, P. (1984) Science **226**, 1315–1317.
22. Härtlein, M. and Madern, D. (1987) Nucleic Acids Res. **15**, 10199–10210.
23. Dardel, F., Fayat, G. and Blanquet, S. (1984) J. Bacteriol. **160**, 1115–1122.
24. Hall, C.V. and Yanofsky, C. (1982) J. Bacteriol. **148**, 941–949.
25. Barker, D.G., Bruton, C.J. and Winter, G. (1982) FEBS Lett. **150**, 419–423.
26. Heck, J.D. and Hatfield, G.W. (1988) J. Biol. Chem. **263**, 868–877.
27. Starzyk, R.M., Webster, T. and Schimmel, P. (1986) Science **237**, 1614–1618.
28. Ghosh, G., Pelka, H. and Schulman, L.H. (1990) Biochemistry **29**, 2220–2225.
29. Brunie, S., Mellot, P., Zelwer, C., Risler, J.-L., Blanquet, S. and Fayat, G. (1987) J. Mol. Graphics **5**, 18–21.
30. Mellot, P., Mechulam, Y., Le Corre, D., Blanquet, S. and Fayat, G. (1989) J. Mol. Biol. **208**, 429–443.
31. Koch, G.L.E. and Burton, J. (1974) FEBS Lett. **40**, 180–182.