Isolation and characterization of the gene coding for *Escherichia coli* arginyl-tRNA synthetase

Gilbert Eriani, Guy Dirheimer and Jean Gangloff

Institut de Biologie Moléculaire et Cellulaire du CNRS and Université L. Pasteur, Laboratoire de Biochimie, 15 rue René Descartes, 67084 Strasbourg, France

Received April 24, 1989; Revised and Accepted June 23, 1989 EMBL accession no. X15320

ABSTRACT

The gene coding for *Escherichia coli* arginyl-tRNA synthetase (*argS*) was isolated as a fragment of 2.4 kb after analysis and subcloning of recombinant plasmids from the Clarke and Carbon library. The clone bearing the gene overproduces arginyl-tRNA synthetase by a factor 100. This means that the enzyme represents more than 20% of the cellular total protein content. Sequencing revealed that the fragment contains a unique open reading frame of 1734 bp flanked at its 5' and 3' ends respectively by 247 bp and 397 bp. The length of the corresponding protein (577 aa) is well consistent with earlier Mr determination (about 70 kd). Primer extension analysis of the ArgRS mRNA by reverse transcriptase, located its 5' end respectively at 8 and 30 nucleotides downstream of a TATA and a TTGAC like element (CTGAC) and 60 nucleotides upstream of the unusual translation initiation codon GUG; nuclease S1 analysis located the 3'-end at 48 bp downstream of the translation termination codon. *argS* has a codon usage pattern typical for highly expressed *E.coli* genes. With the exception of the presence of a HVGH sequence similar to the HIGH consensus element, ArgRS has no relevant sequence homologies with other aminoacyl-tRNA synthetases.

INTRODUCTION

Aminoacyl-tRNA synthetases are a group of enzymes which play a crucial role in the accurate translation of the genetic message since they catalyze the aminoacylation reaction with a very high specificity (1). These enzymes, although having common functions, exhibit large diversity in both their size and oligomeric structures. Therefore, accumulation of important structural informations is required to shed light onto the molecular mechanisms governing substrate binding and the two step aminoacylation process. The primary structures of more than 20 aminoacyl-tRNA synthetases originating from prokaryotes and eukaryotes are so far known, most of them being derived from the nucleotide sequence of the corresponding structural gene (2). Actually, the full tertiary structures are solved for tyrosyl-tRNA synthetase from *Bacillus stearothermophilus* (3) and for the tryptic fragment of the Met enzyme from *E.coli* (4); crystals of yeast AspRS and its complex with tRNA<sub>Asp</sub> (5) as well as cristals of the GluRS-tRNA<sub>Glu</sub> complex are actively investigated by X-ray diffraction (6).

ArgRS purified from *E.coli* consists of a single polypeptide chain of about Mr 70,000 (7,8).
This enzyme, as well as glutamyl- and glutaminyl-tRNA synthetases, differs from the majority of aminoacyl-tRNA synthetases in that it requires the cognate tRNA to catalyse the ATP-PPi exchange reaction (8). To date none of the numerous enzymatic investigations or discussions could bring clear arguments in favour of one of the two most proposed mechanisms: (a) the synthesis of the activated amino acid proceeds as for the other synthetases (9,10) or (b) by a concerted mechanism without formation of the aminoacyl adenylate intermediate (11,12).

Several E.coli K12 mutants having low ArgRS activity were described (13) and mating experiments could locate the structural gene coding for ArgRS at position 40 min on the E.coli chromosome.

This paper reports the isolation and primary structure determination of the gene coding for ArgRS; furthermore it describes the characterization of the mRNA corresponding to \textit{argS} as well as a highly efficient method for purification of its overproduced expression product.

**MATERIALS AND METHODS**

**Chemicals and Enzymes**

[\(\alpha-^{35}\text{S}\)]-dATP, [\(\gamma-^{32}\text{P}\)]-ATP and [\(\alpha-^{32}\text{P}\)]-ddATP were purchased from Amersham. \(^{125}\text{I}\)-protein A was from Du Pont New England Nuclear and \(^{14}\text{C}\) L-arginine from the CEA (Saclay, France). HA-ultrogel was from LKB and DEAE Sephadel from Pharmacia. Oligonucleotides were synthesized on a "Applied biosystems" apparatus.

Restriction enzymes, bacterial alkaline phosphatase, T4 DNA ligase, T4 DNA polymerase, T4 polynucleotide kinase, \textit{E.coli} DNA polymerase I (Klenow fragment) and lysozyme were all from Boehringer Mannheim. Modified T7 DNA polymerase was from United States Biochemical Corporation. All these enzymes were used as recommended by the manufacturer. If not specified, DNA manipulations were performed as described (14).

**E.coli strains and plasmids**

BJ5183 (F, recBC, sbc B, endo I, gal, met, thi, bio, res, mod\(^+\), str\(^R\)).

JM83 (ara, \(\Delta\) (lac-proAB), strA, \(\phi 80\), lacZ\(\Delta\)M15).

JM103 (\(\Delta\) (lac-proAB), thi, strA, supE, endA, sbcB15, hsdR4, F'(traD36, proAB, lacI\(\Delta\)Z\(\Delta\)M15)).

JA200 (F+\(\Delta\) trpE5, recA, thr, leu, lacY) carrying ColE1 and four hybrid ColE1 plasmids (pLC24-15, 15-17, 15-32 and 24-16) was obtained from the Clarke and Carbon library (15).

Cloning of the \textit{E.coli} \textit{argS} gene was performed in the plasmid vectors pBR322 and pUC18. M13mp18 and M13mp19 were produced in strain JM103.

**Preparation of Colicine E1 crude extract**

Colicine E1 was prepared as described by Schwartz and Helinski (16). It was titrated on agar

5726
plates containing LB medium.

Transformation and nucleic acids preparation

Transformation of *E. coli*, large scale preparation of bacterial plasmids, gel electrophoresis and recovery of DNA fragments from agarose were carried out by the standard procedures (14). RNA was prepared according to Summers (17).

Nucleotide sequence determination

The bacteriophages M13mp18 and M13mp19 containing *argS* gene (fragment HindIII-Aval of 2.4 kb) were linearized by EcoRI or HindIII hydrolysis and deleted for various length by partial digestion with DNA polymerase I according to Dale et al (18); after ligation, vectors containing inserts of various length were selected by cloning and sequenced by the dideoxy chain termination method, using modified T7 DNA polymerase (19).

Mapping of the mRNA termini

The 3'-terminus of the *argS* transcript was determined as described (20) using S1 nuclease and a 3' end labelled double stranded DNA fragment prepared in two steps: a 555 bp Rsal-HindIII fragment located at the 3' end of the gene was isolated and labelled at its 3' end using [α-32P]-ddATP in the presence of terminal transferase; after Dral digestion and electrophoretic fractionation on a 5% polyacrylamide gel, a 345 bp Rsal- Dral probe hanging 182 nucleotides over the UAA stop codon could be obtained; it was hybridized to 100 μg of total RNA and submitted to S1 digestion. The DNA protected against nuclease S1 was fractionated by electrophoresis on a polyacrylamide sequencing gel.

Mapping of the 5' terminus was performed as described by Berger et al (21): cDNA was synthesized in the presence of reverse transcriptase using a DNA primer GCACCTGCTCTGCTAATTGTCGCGGTG labelled at its 5' end and hybridized to total RNA. The elongation products were analysed by electrophoretic fractionation on a 6% polyacrylamide sequencing gel.

Preparation of crude extracts

*E. coli* strains were grown in 100 ml LB containing 100 μg per ml ampicillin at 37°C to stationary phase. Cells were harvested by centrifugation, suspended in 1 ml of 100 mM Tris-HCl, 10 mM MgCl2, 1 mM EDTA pH 8 and sonicated during 4 x 20 seconds at 45 V with an Ultrasons-Annemasse apparatus (type 250TS20K); cell debris were eliminated by centrifugation at 12,000 g during 10 min. Protein concentration of the lysate (about 750 μl) was determined as described (22).

Aminoacylation assay with arginyl-tRNA synthetase

To determine arginyl tRNA synthetase activity, enzyme was added to the following reaction mixture: 100 mM Tris-HCl pH 7.5, 30 mM KCl, 2 mM ATP, 2 mM MgCl2, 2mg/ml...
Fig 1: Restriction maps of plasmid \textit{pArgS1} and derived \textit{pArgS2}

The boxes indicate \textit{pBR322} and \textit{pUC18} DNA, chromosomal DNA inserts are drawn as a single line. The numbers refer to size (in kilobase). The arrow designates the DNA region encoding the \textit{argS} structural gene. Abbreviations for restriction enzymes are A = AvaI; EI = EcoRI; EV = EcoRV; H = HindIII; P = PstI; S = Sall.

Unfractionated \textit{E.coli} tRNA, 0.1 mg/ml BSA and 0.1 mM \textit{[^14]C}-L-arginine (25,000 cpm/nmole). After various incubation intervals at 37°C, aliquots were spotted onto Whatman paper discs and quenched with 5% trichloracetic acid. Filters were washed, dried and counted.

**Purification of arginyl-tRNA synthetase**

All manipulations were performed between 0 and +4°C. JM83-\textit{pArgS2} was cultured at 37°C in 3 liters LB containing 200 μg ampicillin per ml. After a 15 hours incubation time, the bacteria were harvested, washed with 100 ml of 10 mM Tris-HCl pH 7.5, 1 mM EDTA, resuspended in 60 ml of 100 mM Tris-HCl pH 8, 10 mM MgCl2, 1 mM EDTA and finally lysed (15 g wet cell paste) by sonication 6 times for 20 s at 100 V. The lysate was cleared by centrifugation successively at 12,000 g for 30 min and 105,000 g for 2 hours. To the collected supernatant, solid (NH₄)₂SO₄ was added, under stirring over a 30 min period to reach 114 g/l. After further 30 min stirring the precipitate was eliminated by centrifugation at 40,000 g for 30 min. Additional (NH₄)₂SO₄ was given to the supernatant fraction to obtain 189 g/l and centrifugation was repeated for 30 min. The pellet was suspended in 20 ml of 10 mM potassium phosphate pH 7.5, 0.1 mM EDTA (buffer A), dialysed twice for 8 hours against 5 liters of the same buffer and loaded on a HA Ultrogel column (id = 4.5 cm, h = 12 cm) equilibrated with buffer A. The column was washed with 1 volume of buffer A and developed with a 1000 ml linear gradient from buffer A to buffer B (50 mM potassium phosphate pH 7.5, 0.1 mM EDTA). ArgRS activity was eluted at approximately 25 mM potassium phosphate pH 7.5. Active fractions were pooled and directly applied to a DEAE-Sephacel column (id = 2.5 cm, h =10 cm) equilibrated with 20 mM potassium phosphate pH 7.5 0.1 mM EDTA (buffer C). The column was washed with 1 volume of buffer C and developed with a 500 ml linear gradient from buffer C to buffer D (250 mM potassium phosphate...
Fig 2A: Expression of the ArgRS gene in the E.coli JM83 transformant harbouring pArgS2
Crude extracts were prepared by sonication of cells from stationary phase cultures and separated on a SDS-10% PA gel (Laemmli et al., 1970). After electrophoresis, the gel was stained with Coomassie Blue: lane a: 10 μg from E.coli JM83; lane b: 10 μg from E.coli JM83 transformed by pArgS2; lane c: 4 μg pure ArgRS.

Fig 2B: Western blot analysis of crude extracts from recipient and transformed strains
Proteins were fractionated on a SDS-10% PA gel, transferred to nitrocellulose and incubated with antiserum. Bound antibody was revealed by secondary binding of 125I-protein A and exposure to Kodak XAR-5 film. Lane a: 0.03 μg pure ArgRS; lane b: 10 μg from E.coli JM83; lane c: 0.5 μg from E.coli JM83 transformed by pArgS2; lane d: 10 μg from E.coli JM83 transformed by pArgS2.

pH 6.5, 1 mM EDTA). ArgRS activity was eluted at 80% of buffer D; at this step the enzyme (30 mg) was approximately 95% pure as appears from polyacrylamide gel electrophoresis in the presence of SDS.

Antibodies
Rabbits were immunized at 15 day intervals by repeated intramuscular injections of 300 μg enzyme dissolved in 500 μl 10 mM potassium phosphate pH 7.4, 150 mM NaCl and emulsified with 500 μl incomplete Freund's adjuvant. One week after the last injection, the rabbits were bled, the serum adjusted to 50% glycerol and conserved at -20°C.

Detection of ArgRS in crude extracts by immunoblotting
Proteins were run on a 10% polyacrylamide gel in the presence of 0.1% SDS (23) and transferred electrophoretically onto nitrocellulose paper previously treated with a 1/250 dilution of rabbit anti-serum raised against purified arginyl-tRNA synthetase. Antigen detection was performed with [125I] protein A (30 μCi/μg) as described by Young and Davis (24).
Determination of the N-terminal sequence of arginyl-tRNA synthetase

The N-terminal sequence was determined by sequential Edman degradation on an 470A Protein Sequencer coupled to a 3120A PTH Analyser (Applied Biosystems). This led us to the identification of the 19 first amino acids of the protein.

RESULTS AND DISCUSSION

Isolation of the gene coding for \textit{E.coli} arginyl-tRNA synthetase (ArgRS)

Early, Cooper \textit{et al.} (13) isolated mutants of \textit{E.coli} K12 having reduced ArgRS activity. From mating experiments carried out on these strains, the authors concluded that the gene coding for ArgRS is located at position 40 min on the \textit{E.coli} linkage map (25).

Likewise, the gene protein index of \textit{E.coli} K12 established by Neidhardt \textit{et al.} (26) also locates the \textit{argS} at position 40 min. Four recombinant plasmids containing large inserts corresponding to the 40 min region were selected out of the Clarke and Carbon library (15); among the corresponding JA200 transformant cells, only one (JA200 - pLC24-15) produced ArgRS activity at a 10 times higher level than the untransformed cells.

Subcloning of the ArgRS gene present in pLC24-15

pLC24-15 was digested by restriction enzyme BglIII; the resulting fragments were ligated in the BamHI site of pBR322 and introduced into BJ5183. Among the transformants selected for ampicillin resistance, 80\% produced ArgRS activity at a much higher level than that observed for JA200 pLC24-15. They contained recombinant plasmids having incorporated the smallest BglII fragment (8kb). One of them (pArgS1) was submitted to restriction analysis (Fig. 1); a restriction fragment Aval-HindIII (2.4 kb) originating from the 8 Kb insert was prepared and inserted between the Sall-HindIII sites of pUC18; JM83 was transformed by the resulting plasmid (pArgS2) and tested for the presence of expression products corresponding to ArgRS.

Overproduction of ArgRS in \textit{E.coli}

SDS PAGE fractionation of a crude extract corresponding to JM83 transformed by pArgS2 reveals at the migration level of \textit{E.coli} ArgRS, a very strong band (Fig. 2A lane b) not present in the case of JM83 (Fig. 2A lane a). As shown by Western blot analysis, this band reacts with anti ArgRS (Fig. 2B lanes a-d). Furthermore, the transformant JM83-pArgS2 overproduces

\textbf{Fig 3 : Nucleotide sequence of argS and derived amino acid sequence}

The complete nucleotide sequence of the \textit{argS} gene is listed. Numbering of the DNA sequence starts with GTG, the first codon for the gene product. Black arrows indicate the location of the 5'-end and 3'-end of the \textit{argS} transcript; possible regulatory sequences like -35 and -10 transcription control elements; ribosome binding site and stop of transcription signal are underlined.
Fig 4A: Mapping of the 5'-termini
Total RNA (100 μg) and 5.10³ cpm of [³²P] phosphorylated primer (see Materials and Methods) were denatured by heating to 75°C and hybridized for 14 hours at 42°C in 80% formamide, 400 mM NaCl, 1 mM EDTA, 100 mM PIPES (pH 6.4). The mixture was ethanol precipitated and resuspended in 50 μl of 50 mM Tris-HCl (pH 8.3), 6 mM MgCl₂, 40 mM KCl, 10 mM DTT, 1 mM dNTP. Primer extension was performed at 37°C during 30 min with 20 U of reverse transcriptase. Extension products were analysed after ethanol precipitation on a 6% sequencing polyacrylamide gel. Lane a: primer extension of total RNA; lanes b-e: Dideoxy-sequencing pattern of argS in the presence of the phosphorylated oligonucleotide primer used for 5'-termini mapping (T, G, C and A respectively).

Fig 4B: S¹ nuclease mapping of the 3'-termini
A 3'end labelled Rsal-Dral (position +1650 to +1112 in fig. 3) fragment was heat denatured at 75°C and hybridized to total RNA (100 μg) in 50% formamide 0.4 M NaCl, 40 mM PIPES (pH
aminoacylation activity more than 100 times compared to the native strain. All these results strongly suggested that the whole argS gene is present in the HindIII-AvaI insert.

Purification of ArgRS was very easily achieved starting from the overproducing transformant JM83-pArgS2 (see Materials and Methods). Only two chromatographic steps were necessary to prepare, from about 15 g wet cell paste, 30 mg of 95% pure ArgRS (Fig. 2A lane c).

Sequencing of the ArgRS gene

The 2.4 kb fragment present in pArgS2 and containing the ArgRS structural gene was cut out of the plasmid and inserted into M13mp18/19. Several truncated overlapping inserts were generated in the recombinant plasmid by E.coli DNA polymerase I digestion according to the strategy of Dale et al (18) and sequenced by the dideoxy chain termination method. As shown in Fig. 3, the nucleotide sequence of the 2.4 kb fragment contains a large open reading frame of 1734 nucleotides corresponding to a 577 amino acid long protein of Mr 70,000 which is in good agreement with that measured for the purified enzyme. The N-terminal sequence of the protein was confirmed by independent protein sequencing by sequential Edman degradation and identification of the 19 first amino acids residues. We can notice that the polypeptide coding sequence starts with an unusual codon GUG. Like UUG this triplet is found more scarcely than AUG : 10- and 100-times respectively (27). It has been suggested that the use of codons other than AUG may be coupled to a low level of gene expression (28). Results of Reddy et al. (28) and Looman et al. (29) favour this hypothesis ; but on the other hand these triplets are also known to be involved in the initiation of synthesis of major polypeptides in E.coli, e.g. ribosomal proteins and elongation factor Tu (29). Concerning ArgRS no particular problem of expression was observed since overproduction can reach more than 20% of the cytoplasmic protein content. Deeper investigations on the GUG initiator codon by site directed mutagenesis could give an answer to the use of this particular triplet.

Codon usage

It has been well established that there is a positive correlation between the degree of codon bias and the level of gene expression (30-31); very highly expressed genes in E.coli contain almost exclusively those codons corresponding to the most abundant tRNAs whereas codons recognized by rare tRNAs are rarely present; argS exhibits a pattern of non random codon usage similar to that of 6.4), 1 mM EDTA, 0.1% SDS. Hybridization was performed for 14 hours at 42°C. The mixture was then diluted in S1 buffer and treated with S1 nuclease (20). The nuclease S1 protected fragments were fractionated on a 6% polyacrylamide sequencing gel. Lane a : 347 bp-Rsal-DraI fragment [32P]-labelled at the RsaI site used as DNA probe ; lane b : DNA-RNA hybrids formed at 42°C after nuclease S1 digestion (the major protected end is indicated by an arrow) ; lane c-d : Maxam and Gilbert sequencing reactions of the DNA probe (C+T and A+G respectively) ; lane e : DNA probe after treatment with S1 nuclease.
The amino acid sequence of the HIGH region of *E. coli* ArgRS is aligned with homologous regions of other synthetases (38-40, 42-45). The number indicates the distance from the NH2 terminus. Conserved amino acids are boxed.

of strongly expressed genes (32) with the exception of the Val codon bias (GUU<GUG).

**Termini corresponding to ArgRS mRNA and consensus elements**

The position of mRNA synthesis initiation site was determined by the primer extension strategy using reverse transcriptase and a 5’ end labelled synthetic oligonucleotide (See Materials and Methods) complementary to position 161 to 187 of the ArgRS structural gene. Total RNA and the oligonucleotide were hybridized at 42°C for 14 hours and after elongation, polymerization products were fractionated on a polyacrylamide sequencing gel. The results shown in Fig.4 indicate that the mRNA starts mainly at about 60 nucleotides upstream of the translation initiation codon (Fig.3). Furthermore a TATA and a TTGAC like sequence (CTGAC) (33) can be detected at respectively 8 and 30 nucleotides upstream of the transcription initiation site; these distances strongly suggest that both sequences correspond to transcription control elements. At position - 8 there is a UAAGG sequence which is complementary to the 3’ end of the 16s rRNA and thus can correspond to a ribosome binding site (34).

The mRNA 3’ end was determined by S1 nuclease protection studies using a 3’ end labelled double stranded DNA probe and total RNA. After hybridization at 42°C and S1 digestion, the protected DNA fragments were fractionated on a polyacrylamide sequencing gel. From Fig.4 it appears that the major transcription termination site is mapped 48 bp downstream of the translation termination codon UAA. It is interesting to notice that between position 20 to 47 downstream of the stop codon, there is a region of dyad symmetry followed by an U-rich sequence. Furthermore, it may be noticed the presence of another band corresponding to a messenger ending in the open reading frame, however, as we measured, its intensity is lower than 10% of the major signal.

**Sequence homologies with other aminoacyl-tRNA synthetases**

It appears that aminoacyl-tRNA synthetases are a group of enzymes which, although catalysing the same reaction, exhibit very low similarities in their primary structures, except those specific for
the same amino acid. So far the 19.2% of direct identity between valyl- and isoleucyl-tRNA synthetase of *E.coli* is the highest homology reported for two enzymes having different amino acid specificities (35).

Nevertheless, primary structure comparison of several aminoaacetyl-tRNA synthetases revealed for most of them a small number of short regions having limited but significant homologies e.g., the KMSKS sequence, not present in *E.coli* ArgRS, which as was demonstrated for *E.coli* MetRS and *E.coli* TyrRS is implicated in maintaining the CCA end of the specific tRNA near the enzyme activation site (36,37). Near the amino terminal end of several aminoaacetyl-tRNA synthetases is located a consensus sequence HIGH probably involved in the binding of ATP (38), the common substrate of these enzymes. In Fig. 5 we have aligned the putative mononucleotide binding fold of *B.stearothermophilus* TyrRS (39) and *E.coli* MetRS (40), where specific residues required in the binding of ATP have been identified (41), with the corresponding regions of several other synthetases. We can notice that the HIGH consensus tetrapeptide is also present in *E.coli* ArgRS (HVGH); no other relevant homologies were found with other aminoaacetyl-tRNA synthetases. To date no primary structure of an ArgRS other than from *E.coli* has been described.

**Acknowledgements:**

We thank M.L. Gangloff for skilful technical assistance; M. Le Ret, M.H. Metz of the J. Reinbolt and Y. Boulanger's laboratory for carrying out the N-terminal analysis of the enzyme.

*To whom correspondence should be addressed

**REFERENCES**

Nucleic Acids Research