Volume 325, number 3, 183–186 FEBS 12611 July 1993
© 1993 Federation of European Biochemical Societies 00145793/93/$6.00

Sequence, overproduction and crystallization of aspartyl-tRNA synthetase from *Thermus thermophilus*

Implications for the structure of prokaryotic aspartyl-tRNA synthetases

Arnaud Poterszman[a,b], Pierre Plateau[a], Dino Moras[b], Sylvain Blanquet[a], Marie-Hélène Mazauric[c], Roland Kreutzer[d] and Daniel Kern[c]

[a]*Laboratoire de Biochimie, URA 240 CNRS, École Polytechnique, 91128 Palaiseau Cédex, France,* [b]*Laboratoire de Biologie Structurale, UPR 9004 CNRS, 15 rue René Descartes, 67084 Strasbourg Cédex, France,* [c]*Unité 'Structure des Macromolécules Biologiques et Mecanismes de Reconnaissance', 15 rue René Descartes, 67084 Strasbourg Cédex, France* and [d]*Laboratorium für Biochemie, Universität Bayreuth, Universitätsstrasse 30, 8580 Bayreuth, Germany*

Received 9 March 1993; revised version received 4 May 1993

The genes of aspartyl-tRNA synthetase (AspRS) from two *Thermus thermophilus* strains VK-1 and HB8, have been cloned and sequenced. Their nucleotidic sequences code for the same protein which displays the three characteristic motifs of class II aminoacyl-tRNA synthetases. This enzyme shows 50% identity with *Escherichia coli* AspRS, over the totality of the chain (580 amino acids). A comparison with the eukaryotic yeast cytoplasmic AspRS indicates the presence in the prokaryotic AspRS of an extra domain between motifs 2 and 3 much larger than in the eukaryotic ones. When its gene is under the control of the *tac* promoter of the expression vector pKK223-3, the protein is efficiently overexpressed as a thermostable protein in *E. coli*. It can be further purified to homogeneity using a heat treatment followed by a single anion exchange chromatography. Single crystals of the pure protein, diffracting at least to 2.2 Å resolution (space group $P2_12_12_1$, $a = 61.4$ Å, $b = 156.1$ Å, $c = 177.3$ Å) are routinely obtained. The same crystals have previously been described as crystals of threonyl-tRNA synthetase [1].

Aminoacyl-tRNA synthetase; Aspartyl-tRNA synthetase; *Thermus thermophilus*; Crystallization

1. INTRODUCTION

Aminoacyl-tRNA synthetases (aaRS) are ubiquitous enzymes that play a key role in the translation process by specifically charging tRNAs with their cognate amino acids. Based on mutually exclusive sets of signature sequences, the twenty aaRS can be partitioned into two classes [2]. So far, five crystal structures of aaRS in their free state or complexed with their cognate tRNA have been determined. Three of them belong to class I (MetRS, $tRNA^{Gln}$-GlnRS and TyrRS [3–5]), two of them are members of class II (SerRS and $tRNA^{Asp}$-AspRS [6,7]). Their tertiary structures have revealed that the binary classification might be associated to structural and functional characteristics of their ATP binding site and of the tRNA acceptor stem recognition mode. However, whether the recognition pattern of these substrates is or is not conserved among aaRS from different origins, specific for a given amino acid, remains to be established.

The 3D structure of the complex $tRNA^{Asp}$-AspRS from *Saccharomyces cerevisiae* is known at 2.9 Å resolution and five primary AspRS sequences are available: *Escherichia coli* (ec) [8], *S. cerevisiae* cytoplasmic (y) [9], and mitochondrial (ym) [10], *Homo sapiens* [11] and *Rattus norvegicus* (rat) [12]. In addition, an open reading frame likely to correspond to AspRS was recently evidenced in the genome of *Caenorhabditis elegans* (EMBL accession number Z19152). Sequence comparisons indicate that eukaryotic and prokaryotic AspRS form two sub-groups of higher similarities, each with higher similarities than the other one. Has this partition a structural and/or functional basis? The study of the crystal structure of the *E. coli* AspRS has been undertaken to answer such a question (Boeglin et al., in preparation).

Meanwhile, during parallel investigation of the crystalline ThrRS from *Thermus thermophilus* [1], we found that the crystallized protein behaved in fact as an AspRS. The high diffracting power as well as the sta-

Correspondence address A. Poterszman, Laboratoire de Biologie structurale, UPR 9004 CNRS, 15 rue René Descartes, 67004 Strasbourg Cedex, France.

Abbreviations. aaRS, aminoacyl-tRNA synthetase, AsnRS, AspRS, GlnRS, LysRS, MetRS, ThrRS and TyrRS, asparaginyl-, aspartyl-, glutaminyl-, lysyl-, methionyl-, threonyl- and tyrosyl-tRNA synthetase, respectively; ec, *Escherichia coli*, tt, *Thermus thermophilus*; y, yeast cytoplasmic; ym, yeast mitochondrial; IPTG, isopropyl-β-thiogalactopyranoside; bp, basepair(s); kb, kilobasepair(s).

bility of the crystals in the X-ray beam prompted us to continue the study of this thermophilic system. We report here the cloning and sequencing of *T. thermophilus* AspRS gene (tt *aspS*) from two strains (VK1 and HB8). The encoded protein was overexpressed in *E. coli*, purified and single crystals of this protein, suitable for high resolution crystallographic investigation, were obtained.

2. MATERIALS AND METHODS

2.1. *Enzymes and chemicals*

RNase A was purchased from Worthington (UK), T4 polynucleotide kinase from New England Biolabs, T7 DNA polymerase, exonuclease III, nuclease S1, Taq DNA polymerase and unlabeled deoxyribonucleotide triphosphates from Pharmacia LKB (Uppsala, Sweden). Bovine alkaline phosphatase and IPTG were obtained from Appligene. Restriction endonucleases, lysozyme, unfractionated ec tRNA and T4 DNA ligase were from Boehringer Mannheim. [^{14}C]L-Aspartic acid was from CEA (France), [α-^{32}P]dATP and [^{35}S]dATPαS were from Amersham. The ion exchange trimethylaminoethyl EMD column was from Merck and the Centricon apparatus from Millipore.

2.2. *Selection of a probe*

The sequence of the first 52 amino acids of the protein was determined from dissolved crystals of ttAspRS, previously believed to be ThrRS. Two oligonucleotides derived from this sequence were synthesized taking into account the codon usage in tt (GGAATTC-ATGCG(C,G)CG(C,G)AC(C,G)CACTACGC(C,G)GG(C,G)(A,T)-(C,G)C and GGAATTC(C,G)GGGTG(C,G)GC(C,G)AC(C,G)AGC-TG(C,G)AC(C,G)AG(C,G)CCCTC). Genomic DNA were obtained from the tt strains VK1 and HB8 (kindly provided to us respectively by Dr. M. Springer (Paris) and Prof. M. Sprinzl (Bayreuth)) and was prepared by the method of Harris-Warrick et al. [13]. Then, a polymerase chain reaction (PCR) experiment was performed with 5 μg of each DNA and 25 pmol of the above oligonucleotides (30 cycles of amplification; 94°C for 1 min, 55°C for 2 min, 72°C for 2 min). The amplified fragment was further cloned into M13mp18 [14] and sequenced by the dideoxy chain termination method [15].

2.3. *Cloning of the gene*

The tt *aspS* genes, encoding ttAspRS, were cloned using similar approaches. For simplicity, we will describe only the strategy used to clone the VK1 gene.

From the sequence obtained from the amplified fragment, a 35-mer oligonucleotide was synthesized, labeled with [γ-^{32}P]ATP [16] and used as a hybridizing probe. When genomic tt DNA was analysed by Southern hybridization, a strong signal was reproducibly obtained from a 5 kb *Hin*dIII fragment. Consequently, the *Hin*dIII fragments with a size of approximately 5 kb were partially purified by gel filtration according to [17], and inserted into the *Hin*dIII site of plasmid pBluescript BS SK$^+$. After transformation of *E. coli* strain JM101TR [18], positive clones were screened by colony hybridization using the above DNA probe. From the plasmid DNA harboured by one selected clone, various restriction fragments were subcloned into M13mp18 or M13mp19, and sequenced using an automatic Pharmacia ALF sequencer. Remaining gaps were filled by manual sequencing with synthetic oligonucleotides and [^{35}S]dATPαS.

2.4. *Construction of overproducing strains*

To obtain an efficient production of ttAspRS in *E coli*, the tt *aspS* gene was placed under the control of the *tac* promoter from the expression vector pKK223-3 (Pharmacia, Sweden). For this purpose, 4 bases were mutagenized to create an *Eco*RI site 1 bp before the *aspS* initiator codon. Then, the 1.75 kbp *Eco*RI–*Bam*HI fragment carrying the tt *aspS* gene was subcloned between the *Eco*RI and *Hin*dIII sites of pKK223-3.

2.5. *Purification*

After transformation of *E coli* strain JM101TR with the resulting plasmid, the recombinant thermostable protein could be easily purified to homogeneity by a heat treatment followed by a single chromatographic step, as follows: *E. coli* cells carrying the recombinant plasmid were grown at 37°C in 5 × LB medium containing 200 μg/ml ampicillin. When the optical density of the culture reached 0.6 at 650 nm, IPTG was added at a final concentration of 0.2 mM and the culture continued for 12 h. After centrifugation, the cell pellet was suspended in a 50 mM Tris-HCl buffer (pH 8.0) containing 150 mM NaCl and 10 mM $MgCl_2$ (10 ml of buffer per g of wet cells). Cells were disrupted by ultrasonication and centrifuged for 25 min at 60,000 × *g*. Incubation of the supernatant at 65°C for 10 min, followed by a centrifugation at 60,000 × *g* for 25 min, removed most *E. coli* proteins. The heat-treated extract was applied to a trimethylaminoethyl ion exchange column (2 cm^2 × 15 cm, flow rate 1 ml/min) equilibrated in 50 mM Tris-HCl (pH 8.0) buffer containing 15 mM KCl. Proteins were eluted with a linear 15–500 mM KCl gradient in the same buffer tRNA aminoacylation activity of AspRS in the various fractions was assayed and analysed by SDS-PAGE gels as described previously [8]. Fractions eluting at 220 ± 15 mM KCl contained the activity. They were pooled and concentrated to 20 mg/ml in a 20 mM Tris-HCl buffer, pH 7.5, containing 100 mM KCl, using a Centricon apparatus.

2.6. *Crystallization*

Crystallization was carried out using standard vapour diffusion technique at 15°C, with 15 μl drops of protein solution equilibrated against a 400 μl reservoir of precipitant solution

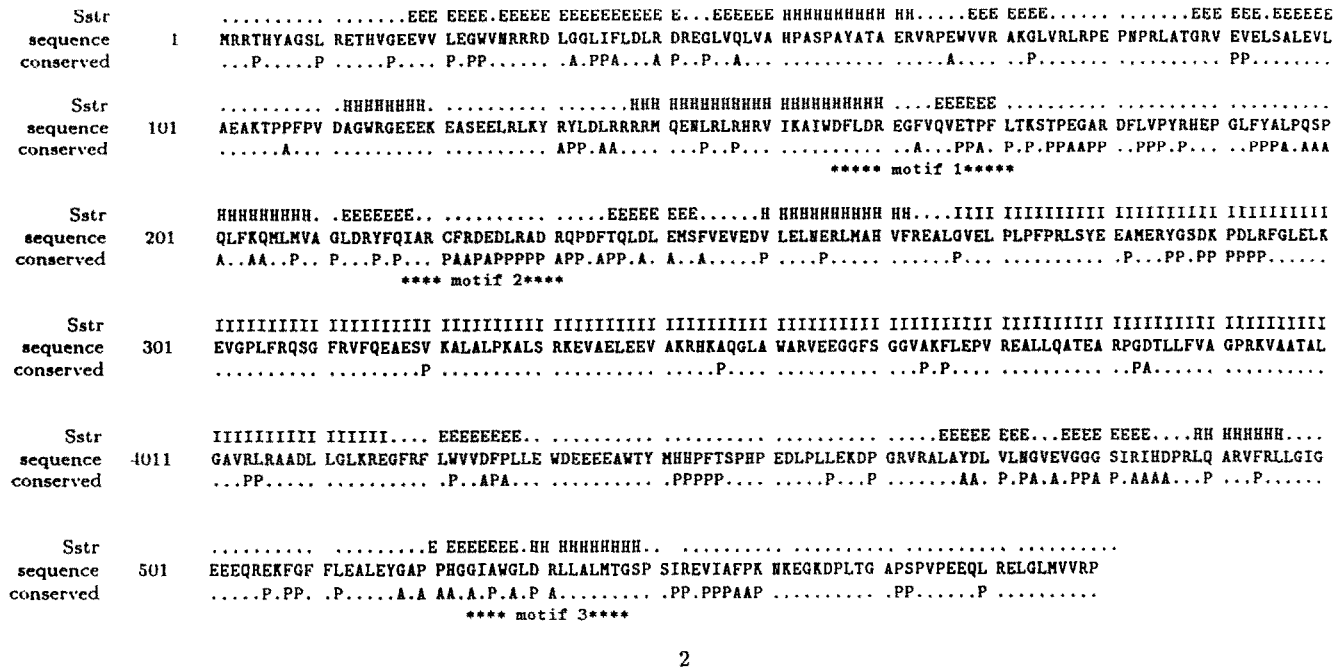
3. RESULTS

3.1. *Sequence*

The comparison of the nucleotide sequences of the genes cloned from both strains (VK1 and HB8) revealed only two differences (GAG in HB8 to GAA in VK1) without modification of the encoded amino acid sequence. An open reading frame encoding a polypeptide of 580 amino acids with a calculated mass of 66,030 daltons was found within a 1,761-bp *Hin*dIII–*Bam*HI fragment. The first amino acids of this polypeptide exactly corresponded to those of the N-terminal sequence analysis of crystalline ttAspRS. The DNA sequences were deposited in the EMBL database with accession number X70943.

3.2. *Purification and crystallization*

The purification procedure described in section 2 produced 20 mg of pure AspRS for cells obtained from one liter of the *E. coli* culture. As judged from SDS-PAGE analysis, the recovered protein was homogeneous (data not shown). While new crystals could be obtained using ammonium sulfate and PEG/NaCl as precipitant, for the sake of continuity with respect to the already engaged structure determination, the crystallization was attempted in sodium formate. In formate, large crystals are obtained using the macroseeding technique, after equilibrating drops containing 5 mg/ml protein, 2 M sodium formate, 50 mM Tris-HCl (pH 7.5) against 2 M sodium formate for 5 days. A single crystal, washed 3 times in a 2 M solution of precipitant without protein, is introduced in the drop while the concentration of



2

Fig. 1. Amino acid sequence of ttAspRS. The sequence from ttAspRS was compared with those of other AspRS using the UWGCG software package. An 'A' under a residue indicates its presence in both prokaryotic (tt, ec and ym) and eukaryotic AspRS; a 'P' indicates that it is specifically prokaryotic. Putative secondary structure elements are indicated; 'H' stands for helix and 'E' for β sheet; residues from the insertion domain are indicated by a 'I'.

formate in the well is raised to 3.4 M. Crystals grow for 3 weeks and their final size reaches $400 \times 200 \times 150\ \mu m$. These crystals belong to orthorhombic space group $P2_12_12_1$, with cell parameters $a = 61.4$ Å, $b = 156.1$ Å, $c = 177.3$ Å and one dimeric molecule per asymmetric unit. They are very stable under an X-ray beam and diffract to a resolution beyond 2.2 Å.

4. DISCUSSION

Sequence analysis indicated that ttAspRS actually belongs to the sub-group formed by prokaryotic enzymes. Among prokaryotic enzymes (tt, ec and ym AspRS), 24% of the residues are strictly conserved and 50% identity is observed along the sequences of ttAspRS and ecAspRS. These conservations suggest that prokaryotic

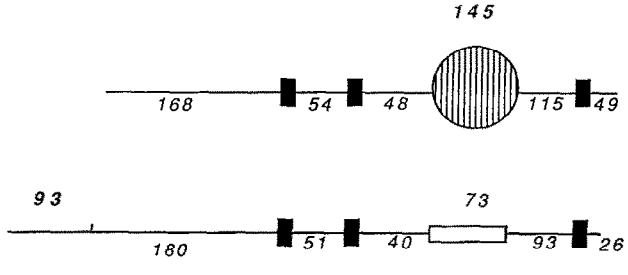


Fig. 2. Comparison of the organisation of a eukaryotic (y) and of a prokaryotic (tt) AspRS primary sequences. Dark rectangles indicate the three conserved motifs, the shaded region represents the prokaryotic extra domain and the italicized numbers the length of the polypeptide sequences in the proteins.

AspRS share identical 3D structures. Significant similarities between prokaryotic and eukaryotic enzymes are observed over two-thirds of the sequence (Fig. 1). Regions of high conservation correspond: (i) to an N-terminal β barrel, found in the structure of yAspRS and detected by structure prediction in the other aaRS of the same subclass (AsnRS and LysRS); and (ii) to the active site common to class II aaRS. An insertion sequence, located between motifs 2 and 3 appears to be characteristic of prokaryotic AspRS. This connecting peptide of approximately 155 amino acids (residues 265–420) has the size of a structural domain. In yAspRS and other eukaryotic AspRS, the corresponding module having only 75 amino acids is located between residues 365 and 440 (Figs. 2 and 3).

Most residues involved in tRNA recognition and in catalysis are strictly conserved among prokaryotic and eukaryotic AspRS [19]. Amino acids, identified as important for anticodon recognition in the structure of yAspRS have their counterparts in ttAspRS (F36 and E66). This suggests a common recognition mode of the anticodon, via a β barrel, for all AspRS. The three motifs characteristic of class II dimeric aminoacyl-tRNA synthetases are present. Motif 1 contains an invariant structural proline at position 169. Invariant arginines at positions 223 and 531 from motifs 2 and 3 interact with ATP [19]. Residues of the variable loop from motif 2 (E225 and F235) as well as the glycine-rich strand of motif 3, which are both involved in ATP binding, are present.

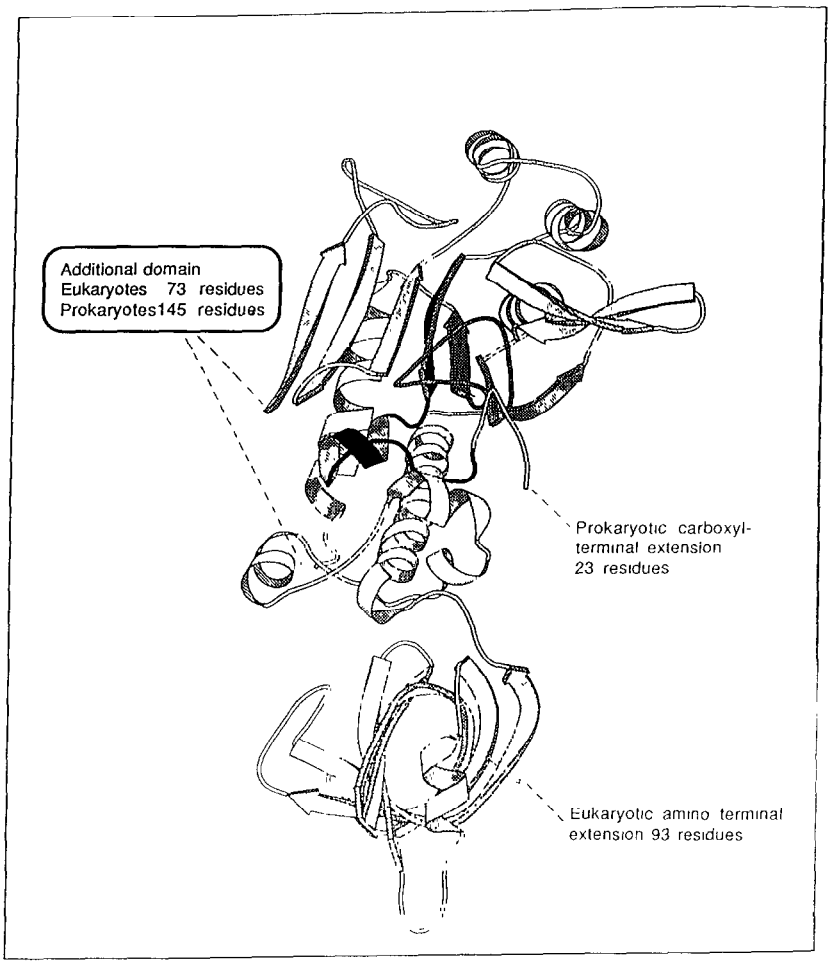


Fig. 3. Model of the structures of ttAspRS based on that of yAspRS. The location of the insertion domain between motifs 2 and 3 is indicated by a circle Motifs 1, 2 and 3 are represented in black.

Acknowledgements We thank Dr. J. Reinbolt for the N-terminal sequence analysis of crystalline AspRS, Prof R Giegé and M. Sprinzl for facilitating some steps of this work and for helpful discussions, Dr P. Dessen for pointing out the existence of the sequence of AspRS from *Ceanorhabditis elegans* and Dr. J. Cavarelli for help in drawing Fig. 2 using MOLSCRIPT (Kraulis, 1991). This work was supported in part by the Fondation pour la Recherche Médicale and the Interface Chimie Biologie program of the CNRS.

REFERENCES

[1] Garber, M.B., Yaremchuck, A D., Tukalo, M.A , Egorova, S.P., Fomenkova, N.P. and Nikonov, S V. (1990) J. Mol Biol. 214, 819–820
[2] Eriani, G., Delarue, M., Poch, O., Gangloff, J. and Moras, D. (1990) Nature 347, 203–206.
[3] Brunie, S., Zelwer, C. and Risler, J.L (1990) J. Mol. Biol. 216, 411–424.
[4] Rould, M.A., Perona, J.J., Soll, D. and Steitz, T.A. (1989) Science 246, 1135–1142.
[5] Brick, P., Bhat, T.N. and Blow, D.M. (1989) J. Mol Biol. 208, 83–98.
[6] Cusack, S., Berthet-Colominas, C., Härtlein, M., Nassar, N and Leberman, R. (1990) Nature 347, 249–255.
[7] Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B , Thierry, J.-C. and Moras, D. (1991) Science 252, 1682–1689
[8] Eriani, G., Dirheimer, G. and Gangloff, J. (1990) Nucleic Acids Res. 18, 7109–7117.
[9] Sellami, M., Fasiolo, F., Dirheimer, G , Ebel, J P. and Gangloff, J. (1986) Nucleic Acids Res 14, 1657–1666.
[10] Gampel, A. and Tzagoloff, A (1989) Proc. Natl. Acad. Sci. USA 86, 6023–6027.
[11] Jacobo-Molina, A., Peterson, R. and Yang, D C.H. (1989) J Biol. Chem. 264, 16608–16612.
[12] Mirande, M and Waller, J P. (1989) J Biol Chem 264, 842–847.
[13] Harris-Warrick, R.M., Elkana, Y., Ehrlich, S.D. and Lederberg, J. (1975) Proc. Natl. Acad. Sci USA 72, 2207–2211.
[14] Vieira, J. and Messing, J. (1982) Gene 19, 259–268.
[15] Sanger, F., Nicklen, S. and Coulsen, A.R (1977) Proc Natl. Acad. Sci. USA 74, 5463–5467.
[16] Sambrook, J., Fritsch, E F. and Maniatis, T. (1989) Molecular Cloning, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p. 777
[17] Beauvallet, C., Hountondji, C. and Schmitter, J.M. (1988) J. Chromatogr. 438, 347–357.
[18] Hirel, P.H., Levêque, F., Mellot, P., Dardel, F , Panvert, M., Mechulam, Y. and Fayat, G. (1988) Biochimie 70, 773–782.
[19] Cavarelli, J., Rees, B , Ruff, M., Thierry, J C and Moras, D. (1993) Nature 362, 181–184.