Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Gene cloning, expression, crystallization and preliminary X-ray analysis of *Thermus thermophilus* arginyl-tRNA synthetase

Atsushi Shimada,[a] Osamu Nureki,[a,b] Naoshi Dohmae,[c] Koji Takio[c] and Shigeyuki Yokoyama[a,b]*

[a]Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan, [b]Cellular Signaling Laboratory, RIKEN Harima Institute, 1-1-1 Kouto, Mikazuki-cho, Sayo-gun, Hyogo 679-5148, Japan, and [c]Division of Biomolecular Characterization, The Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

Correspondence e-mail: yokoyama@biochem.s.u-tokyo.ac.jp

Received 28 August 2000
Accepted 7 November 2000

The gene encoding the highly thermostable arginyl-tRNA synthetase (ArgRS) from *Thermus thermophilus* was cloned and overexpressed in *Escherichia coli* under the control of the T7 promoter. The recombinant ArgRS was purified by two chromatographic steps and was crystallized by the hanging-drop vapour-diffusion method using PEG 8000 and ethylene glycol as precipitants. The crystals belong to the hexagonal space group $P6_5$, with unit-cell parameters $a = b = 156.04$ (7), $c = 87.17$ (4) Å. X-ray data to 2.8 Å resolution were collected at room temperature from a native crystal using an in-house X-ray source. Uranium, platinum and selenomethionine derivatives were found to be useful for phasing by the multiple isomorphous replacement method with anomalous scattering. The flash-frozen crystals diffracted beyond 2.3 Å resolution using synchrotron radiation from the beamline 41XU at SPring-8 (Harima).

1. Introduction

Aminoacyl-tRNA synthetases (aaRSs) form a highly diversified enzyme family that catalyzes the esterification of cognate amino acids with their specific tRNAs, which decode trinucleotide sequences (anticodons) into amino acids strictly according to the genetic code. In the first step of the aminoacylation reaction, the amino acid and adenosine triphosphate (ATP) form an aminoacyl-adenosine monophosphate as an active intermediate. Then, in the second step, the aminoacyl moiety is transferred to the 3′-terminal adenosine of the cognate tRNA. To discriminate tRNAs, aaRSs recognize a small number of characteristic nucleotides, called identity elements, which are unique to a subset of tRNAs specific for a given amino acid (Giegé *et al.*, 1998). Most of the identity elements are located in the two extremities of the L-shaped tRNA; namely, the anticodon loop and the acceptor end. The tRNA recognition by arginyl-tRNA synthetase (ArgRS) is unique compared with other systems. In $tRNA^{Arg}$, the major identity elements have been identified as A20 and C35, which reside in the D loop and the anticodon loop, respectively (McClain & Foss, 1988; Schulman & Pelka, 1989). A20 is located at the outside corner of the tRNA, which is far from both the anticodon and the acceptor end. No tRNAs specific for the other amino acids have a major identity element at this position. The mechanism of A20 recognition by ArgRS has not yet been elucidated. Although the crystal structure of the yeast ArgRS has already been reported (Cavarelli *et al.*, 1998), yeast is an exceptional organism in which the $tRNA^{Arg}$ has either C20 or U20 in place of A20 (Sprinzl *et al.*, 1998). The species-specific differences in the crystal structures between the yeast and *T. thermophilus* ArgRSs may highlight the structural elements that recognize A20. In an attempt to elucidate the mechanism of A20 recognition by ArgRS, we report here the gene cloning, purification and preliminary X-ray diffraction analysis of *T. thermophilus* ArgRS.

2. Experimental and results

2.1. Purification and gene cloning of *T. thermophilus* ArgRS

T. thermophilus HB8 cells (250 g) were suspended in 50 m*M* Tris–HCl buffer pH 7.9 containing 10 m*M* magnesium acetate, 5 m*M* β-mercaptoethanol, 0.5 m*M* phenylmethylsulfonyl fluoride and 500 m*M* NH_4Cl and were homogenized by sonication. The crude lysate was centrifuged at 30 000*g* for 30 min. The supernatant fraction was dialyzed against 50 m*M* Tris–HCl buffer pH 7.9 containing 10 m*M* magnesium acetate, 5 m*M* β-mercaptoethanol, 0.1 m*M* phenylmethylsulfonyl fluoride (buffer *A*) and was loaded onto a DEAE-Sephacel column (Pharmacia) equilibrated with buffer *A*. The protein was eluted with a linear gradient of 0–0.25 *M* NH_4Cl. The ArgRS fractions were pooled and dialyzed against 50 m*M* potassium phosphate buffer pH 7.0 containing 5 m*M* β-mercaptoethanol (buffer *B*). Buffer *C* [buffer *B* containing 4 *M* $(NH_4)_2SO_4$] was added so that the final concentration of $(NH_4)_2SO_4$ of the protein

solution reached 0.8 *M*; the resultant solution was then loaded onto a Butyl-Toyopearl column (Tosoh). The protein was eluted with a linear gradient of 0.8–0 *M* $(NH_4)_2SO_4$. The protein fractions were next dialyzed against buffer *A* and were loaded onto an AF-Red Toyopearl column (Tosoh) equilibrated with buffer *A*. The ArgRS showed a significantly high affinity for AF-Red Toyopearl, which probably served as an analogue of ATP or tRNA, and the protein could not be eluted by the initial trial with a linear gradient of 0–1 *M* KCl. The ArgRS was finally eluted with 2 *M* KCl. At this point, the ArgRS was purified to approximately 90% homogeneity, as analyzed by SDS–PAGE. The protein fractions were equilibrated with buffer *B* and were mixed with buffer *C* so that the final concentration of $(NH_4)_2SO_4$ reached 0.8 *M*. The protein solution was then loaded onto a Phenyl-Superose HR 10/10 column (Pharmacia), which was eluted with a linear gradient of 0.8–0 *M* $(NH_4)_2SO_4$. Throughout the purification process, the fractions containing ArgRS were traced by measuring the arginylation activity of each fraction (Kohda *et al.*, 1987). The reaction mixture contained 100 m*M* Tris–HCl buffer pH 7.5, 5 m*M* magnesium acetate, 0.3 m*M* tRNA mixture from *T. thermophilus*, 0.02 m*M* [^{14}C]-arginine, 10 m*M* KCl, 2 m*M* ATP and 20%(*v*/*v*) of the protein fraction.

The amino-acid sequence of the amino-terminus of the purified ArgRS was determined to be MLRRALEEAIAQALKEMGVPVRLKVAR. In addition, we obtained two polypeptide fragments by limited digestion of the purified ArgRS with *Achromobacter* protease I (lysyl endopeptidase) and determined the partial amino-acid sequences of the amino-termini of the two fragments: DRLPLPEFVEEAVPVGGYLNFRLRTEALLREALRPKAPFPRRPGVVLVEHTSVNPNKELHVGHLRNIALGDAI and GLAVSVDEVLEEATRRARAIVEEKNPDHPDKEEAARMVALGAIRF. The first fragment was found to contain an ATP-binding signature sequence specific for the class-I aaRSs (Eriani *et al.*, 1990), HVGH, and showed significant similarity with the corresponding region of the ArgRSs from various organisms. In contrast, the second fragment showed less similarity with the sequences of the ArgRSs from a limited number of organisms. Based on the partial amino-acid sequences of the N-terminus and the first proteolytic fragment of the protein, a mixed 36-oligonucleotide probe with a *Bam*HI site at the 5′ end, 5′-d[GCGCGGATTCGC(G/C)(C/T)T(G/C)GAGGAGGC(G/C)ATCGC(G/C)CAGGC]-3′, and a mixed 40-oligonucleotide antisense probe with a *Hin*dIII site at the 5′ end, 5′-d[GCGCAAGCTTGTG(G/C)CC(C/G)ACGTG(G/C)A(G/A)CTCCTTGTT(G/C)GGGTT]-3′, were synthesized and used for polymerase chain reaction (PCR) amplification with *T. thermophilus* chromosomal DNA as the template. An amplified DNA fragment was cloned into the appropriate sites of pUC119 and was sequenced. The DNA fragment was found to encode the proteolytic peptides. The PCR fragment was then labelled with ^{32}P and used for Southern hybridization to *T. thermophilus* chromosomal DNA digested with several restriction enzymes. The probe hybridized remarkably well to an *Sph*I fragment of about 6.5 kbp. This 6.5 kbp *Sph*I fragment was cloned into the *Sph*I site of plasmid pUC118 by colony hybridization. The insert was sequenced and was found to include the entire *argS* gene. The nucleotide sequence of the gene has been deposited in the EMBL Nucleotide Sequence Database (accession number: AJ278815). The *argS* gene encodes a 592-residue polypeptide with an M_r of 66 212.

2.2. Purification of the recombinant ArgRS

The *T. thermophilus argS* gene was PCR amplified with artificial *Nde*I and *Sal*I sites and was ligated into the appropriate sites of the expression vector pK7, which has a potent T7 promoter (Kigawa *et al.*, 1995). The wild-type ArgRS was expressed in the *E. coli* strain JM109 (DE3), which was grown at 310 K in Luria–Bertani broth. The cells were centrifuged at 7000*g* for 15 min. The pellet was resuspended in 50 m*M* Tris–HCl buffer pH 7.9 containing 10 m*M* magnesium acetate, 10 m*M* β-mercaptoethanol, 1 m*M* phenylmethylsulfonyl fluoride and 10 m*M* NH_4Cl and was homogenized by sonication. The crude lysate was centrifuged at 30 000*g* for 30 min and the supernatant fraction was heated for 30 min at 348 K to denature the *E. coli* proteins. The heat-treated fraction was centrifuged again at 30 000*g* for 30 min. The supernatant fraction was dialyzed against 50 m*M* Tris–HCl buffer pH 7.9 containing 10 m*M* magnesium acetate, 5 m*M* β-mercaptoethanol (buffer *D*) and was loaded onto a DEAE-Sephacel column (Pharmacia) equilibrated with buffer *D*. The protein was eluted with a linear gradient of 0–0.25 *M* NH_4Cl. The protein fractions were further purified on a Phenyl-Superose HR 10/10 column (Pharmacia) by exactly the same procedure used to purify ArgRS from *T. thermophilus* cells.

2.3. Crystallization

Prior to crystallization, the ArgRS fraction was equilibrated with 10 m*M* Tris–HCl buffer pH 7.8 containing 5 m*M* $MgCl_2$ and was concentrated to approximately 20 mg ml^{-1} by ultrafiltration using a Centricon concentrator (Amicon) with a 30 kDa cutoff. The initial screening of crystallization conditions was carried out at 293 K using the hanging-drop vapour-diffusion method based on the conditions described by Jancarik & Kim (1991). Crystals were obtained using ammonium sulfate or sodium potassium tartrate as precipitants. The initial crystallization conditions were further refined by changing the pH, the concentrations of precipitants and the temperature and by screening additives. The most promising crystals (form I) were obtained by mixing equal volumes of the protein solution and 100 m*M* HEPES buffer pH 8.0 containing 0.4 *M* sodium potassium tartrate and by equilibrating the mixture with 100 m*M* HEPES buffer pH 8.0 containing 0.8 *M* sodium potassium tartrate. The crystals grew to maximum dimensions of 0.15 × 0.15 × 0.3 mm in a week. The form I crystals belong to the hexagonal space group $P6_3$ and have unit-cell parameters $a = b = 205.0$ (4), $c = 97.1$ (2) Å. However, they diffracted to a relatively low resolution (6.0 Å) at 293 K on a Rigaku R-AXIS IV imaging-plate detector mounted on a rotating-anode X-ray source and showed extremely high mosaicity along the *c* axis. Therefore, these crystals were considered to be unsuitable for the structure determination. Refinement of the other conditions that produced quasi-crystals in the initial screening was attempted by screening the additives; finally, the best crystals (form II) were obtained by mixing 2 µl of protein solution, 2 µl of 100 m*M* HEPES buffer pH

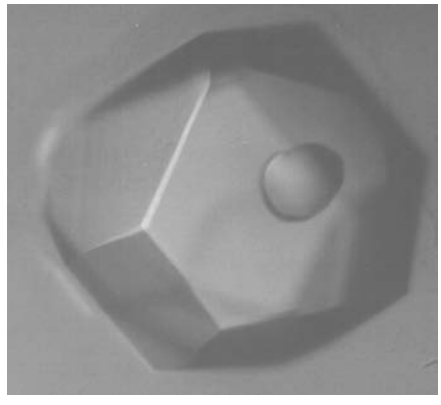

Figure 1
A typical hexagonal crystal of ArgRS from *T. thermophilus*. The dimensions of the crystal are about 0.4 × 0.4 × 0.4 mm.

7.0 containing 10%(*v*/*v*) PEG 8000 and 8%(*v*/*v*) ethylene glycol (reservoir solution) with 1 µl of 2.5% polyvinylpyrrolidone K15 as an additive and equilibrating them with the reservoir solution (Fig. 1). The form II crystals appeared six months after setting the hanging drop at 293 K. Only the macroseeding method can reproduce the crystals, which grow in a week to average dimensions of 0.4 × 0.4 × 0.4 mm.

2.4. Native and derivative crystal data collection

The crystals diffracted beyond 2.8 Å resolution at 293 K on a Rigaku R-AXIS IV imaging-plate detector mounted on a rotating-anode X-ray source (Table 1). They belong to the space group $P6_1$ or $P6_5$ [unit-cell parameters $a = b = 156.04\,(7)$, $c = 87.17\,(4)$ Å] and contain one monomer per asymmetric unit, according to the Matthews coefficient ($V_M = 2.3$ Å^3 Da^{-1}; Matthews, 1968). The selenomethionine (SeMet) substituted protein and the platinum and uranium derivatives were also prepared in order to solve the phase problem by the multiple isomorphous replacement method with anomalous scattering (MIRAS; Table 1). The SeMet-substituted protein was obtained by growing the *E. coli* methionine-auxotroph B834 (DE3) strain (Novagen) transformed with the overexpression vector in minimal medium in which the methionine was replaced by SeMet. Platinum and uranium derivatives were prepared by soaking crystals in reservoir solutions containing K_2PtCl_4 (1 m*M*, 8 h) and $UO_2(CH_3COO)_2$ (2 m*M*, 1 h), respectively. Data were collected on an R-AXIS IV imaging-plate detector in the same way as for the native crystals (Table 1). The native crystals were flash-frozen at 100 K in reservoir solution containing an increased concentration (15%) of ethylene glycol as a cryoprotectant; the high-resolution data were collected to 2.3 Å using synchrotron radiation at beamline 41XU of SPring-8 (Harima, Japan). The flash-frozen crystals belong to the same space group $P6_1$ or $P6_5$ and have unit-cell parameters $a = b = 154.74\,(9)$, $c = 84.58\,(8)$ Å. A total of 58 imaging frames were recorded at a wavelength of 0.708 Å and a 400 mm crystal-to-imaging plate distance using a 3° oscillation angle (Fig. 2 and Table 1).

Table 1
Crystallographic data of *T. thermophilus* arginyl-tRNA synthetase and its derivatives.

Values in square brackets are for the highest resolution shell.

	Native (form II)	Native (form II)	$UO_2(CH_3COO)_2$	K_2PtCl_4	SeMet
Diffraction data					
Temperature (K)	293	100	293	293	293
Unit-cell parameters (Å)					
$a = b$	156.04 (7)	154.74 (9)	156.02 (16)	156.17 (18)	156.22 (10)
c	87.17 (4)	84.57 (8)	86.94 (31)	86.71 (8)	87.28 (13)
Wavelength (Å)	1.5418	0.708	1.5418	1.5418	1.5418
Resolution (Å)	50–2.8	50–2.3	50–3.0	50–2.8	50–2.8
Unique reflections	29482	49236	23379	29766	29923
Total reflections	274773	408324	187033	190958	273715
$I/\sigma(I) > 3$ (%)	83.5 [55.6]	83.9 [65.5]	68.8 [29.4]	58.8 [13.2]	80.2 [48.4]
R_{merge}† (%)	9.0 [33.2]	5.1 [18.6]	18.9 [48.7]	14.4 [45.1]	11.4 [40.2]
Completeness (%)	98.3 [96.3]	95.8 [94.7]	96.3 [91.2]	99.1 [96.5]	99.8 [99.3]
Phasing statistics (3.0–50 Å)					
R_{der}‡ (%)			16.4	17.2	9.6
Phasing power§					
Acentric			0.84	0.64	1.16
Centric			0.62	0.63	0.92
R_{Cullis}¶					
Acentric			0.87	0.92	0.80
Centric			0.83	0.83	0.73
Mean overall figure of merit			0.4292		

† $R_{merge} = \sum_h \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i |I_{h,i}|$, where h refers to unique reflection indices and i indicates symmetry-equivalent indices. ‡ $R_{der} = \sum ||F_{PH}| - |F_P|| / \sum |F_P|$, where $|F_P|$ and $|F_{PH}|$ refer to the measured structure-factor amplitudes of the native and the derivative, respectively. § Phasing power = $f_{r.m.s.}/E_{r.m.s.}$, where $f_{r.m.s.} = [(\sum f_H^2)/n]^{1/2}$ and $E_{r.m.s.} = [\sum (F_{PH} - |F_P + f_H|)^2/n]^{1/2}$. ¶ $R_{Cullis} = \sum [|F_H| - (|F_{PH}| - |F_P|)] / \sum (|F_{PH}| - |F_P|)$, where $|F_H|$ represents the calculated heavy-atom structure factor.

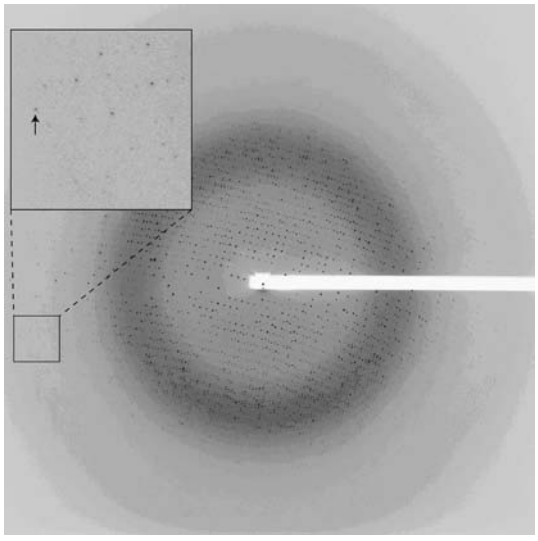

Figure 2
A 3° oscillation image of the crystal of ArgRS from *T. thermophilus* obtained using synchrotron radiation at beamline 41XU of SPring-8. The small arrow indicates the diffraction limit, which corresponds to 2.1 Å.

2.5. Phasing

All data were processed using the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Subsequent phase calculations were carried out with the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). Two uranium sites were determined with the program *RSPS* (Collaborative Computational Project, Number 4, 1994) from the isomorphous and anomalous difference Patterson maps. The initial phase from the uranium derivative was used to locate the positions of the Pt and Se atoms by difference Fourier analysis. Heavy-atom parameters were refined using the program *MLPHARE* (Collaborative Computational Project, Number 4, 1994) and the resulting MIRAS map was of excellent quality, showing a clear solvent boundary. MIR phasing finally determined the space group of the crystals as $P6_5$. Diffraction data for each heavy-atom derivative and phasing statistics are summarized in Table 1.

We thank Masayoshi Nakasako, Michiko Konno and Masami Horikoshi for data collection and helpful discussions. We are greatly indebted to Nobuo Kamiya (RIKEN) and Masahide Kawamoto (JASRI) for their help in data collection at SPring-8. We thank Kiyotaka Shiba, Shun-ichi Sekine, Takaho Terada and Kyoko Saito for helpful discussions. This work was supported in part by Grants-in-Aid for Science Research on Priority Areas (09278101 and 11169204 to SY and ON, respectively) from the Ministry of Education, Science, Sports and Culture of Japan

and by a grant from the JSPS Research Fellowships for Young Scientists (AS).

References

Cavarelli, J., Delagoutte, B., Eriani, G., Gangloff, J. & Moras, D. (1998). *EMBO J.* **17**(18), 5438–5448.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990). *Nature (London)*, **347**, 203–206.

Giegé, R., Sissler, M. & Florentz, C. (1998). *Nucleic Acids Res.* **26**, 5017–5035.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Kigawa, T., Muto, Y. & Yokoyama, S. (1995). *J. Biomol. NMR*, **6**, 129–134.

Kohda, D., Yokoyama, S. & Miyazawa, T. (1987). *J. Biol. Chem.* **262**, 558–563.

McClain, W. H. & Foss, K. (1988). *Science*, **241**, 1804–1807.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Schulman, L. H. & Pelka, H. (1989). *Science*, **246**, 1595–1597.

Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. & Steinberg, S. (1998). *Nucleic Acids Res.* **26**, 148–153.