Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary diffraction analysis of *Escherichia coli* cysteinyl-tRNA synthetase

Kate Juliet Newberry,[a] Jonathan Kohn,[a] Ya-Ming Hou[b] and John J. Perona[a]*

[a]Department of Chemistry and Interdepartmental Program in Biochemistry and Molecular Biology, University of California at Santa Barbara, Santa Barbara CA 93106-9510, USA, and [b]Department of Biochemistry and Molecular Pharmacology, Thomas Jefferson University, 233 South 10th St, Philadelphia PA 19107, USA

Correspondence e-mail: perona@chem.ucsb.edu

Received 23 November 1998
Accepted 19 January 1999

Crystals of the 52 kDa monomeric *Escherichia coli* cysteinyl-tRNA synthetase complexed with ATP and cysteine have been grown by hanging-drop vapor diffusion from solutions containing ammonium sulfate as the precipitating agent. The crystals form long hexagonal rods in the space group $P321$ with unit-cell dimensions $a = b = 82.3$, $c = 168.9$ Å. There is one enzyme molecule in the asymmetric unit. A complete native data set has been collected from a rotating-anode source to a resolution of 2.7 Å at 103 K, with an $R_{\rm merge}$ of 6.7%.

1. Introduction

Aminoacyl-tRNA synthetases are central players in the transmission of genetic information and are responsible for attaching amino acids to the 3′-acceptor ends of specific transfer-RNA molecules (Cusack, 1997; Arnez & Moras, 1997; Carter, 1993). Each synthetase uses ATP as a high-energy cofactor in order to activate a specific amino acid to form an enzyme-bound aminoacyl adenylate species. In a second step, the enzyme then transfers the activated amino acid to either the 2′- or 3′-terminal hydroxyl group of the 3′ A76 nucleotide of tRNA. Because all tRNAs have a common L-shaped structure, the basis for discrimination between different sets of isoacceptors relies on subtle sequence-dependent differences in conformation. Thus, aminoacyl-tRNA synthetases must be simultaneously capable of a very high degree of selectivity against each of two structurally diverse classes of substrates.

Primary sequence and tertiary structural information on synthetases has resolved the enzymes into two classes of ten enzymes each (Cusack *et al.*, 1990; Eriani *et al.*, 1990). Class I synthetases possess a Rossmann-fold active-site domain, aminoacylate tRNA on the 2′-OH of the A76 ribose and bind tRNA on the minor-groove side of the acceptor stem. By contrast, class II synthetases have active-site domains which are constructed from a different antiparallel β-sheet architecture. These enzymes approach tRNA on the major-groove side and (with one exception) aminoacylate the 3′-OH of the terminal tRNA ribose. The clear functional and structural distinctions between the two classes almost certainly reflects an early evolutionary divergence in this ancient family of enzymes. However, despite the common functions, there is only limited primary sequence similarity within each class and no conservation of tertiary or quaternary structure outside the active-site domains.

The extreme diversity and functional importance of the aminoacyl-tRNA synthetase family makes structure determination of each of the 20 enzymes an important undertaking. Structures are now available for 15 of the 20 enzymes (Cusack, 1997; Cavarelli *et al.*, 1998). No structural information is as yet available for the class I valyl-, cysteinyl- or leucyl-tRNA synthetases or for the class II threonyl and alanyl enzymes. Here, we report crystals suitable for high-resolution structure determination of the 52 kDa monomeric *Escherichia coli* cysteinyl-tRNA synthetase (CysRS). This enzyme possesses at least two unique features of outstanding interest. Firstly, the mechanism for discrimination against serine poses an intriguing dilemma owing to its chemical similarity to cysteine. Early work has suggested that no editing mechanism may be present in CysRS (Fersht & Dingwall, 1979). If this is in fact the case, then the basis by which the slightly smaller serine hydroxyl group is excluded from the amino-acid binding site may reveal novel principles for molecular selectivity. Secondly, $tRNA^{Cys}$ possesses a unique G15–G48 Levitt base pair in the tertiary core region of the molecule (Sprinzl *et al.*, 1998). This base pair is an identity determinant for *E. coli* CysRS (Hou *et al.*, 1993), and biochemical data suggest that it functions by facilitating formation of a specific tertiary structure in the RNA (Hamann & Hou, 1997*a*,*b*). Identity determinants in the tertiary core region of tRNA are relatively rare (Giegé *et al.*, 1993). Thus, elucidation of the mechanism of Levitt-pair recognition should expand our understanding of RNA–protein interactions. The only class I synthetase–tRNA complex solved at high resolution, the GlnRS–$tRNA^{Gln}$ complex, does not rely on tertiary RNA

Table 1
Native cysteinyl-tRNA synthetase data-collection statistics.

Resolution (Å)	R_{merge}†	$\langle I\rangle/\sigma$	Total number of observations	Number of unique reflections	Completeness (%)	Multiplicity
8.54	0.035	18.0	3000	679	98.9	4.4
6.04	0.039	16.6	5620	1151	99.8	4.9
4.93	0.057	12.1	7199	1445	99.8	5.0
4.27	0.052	13.2	8500	1685	99.8	5.0
3.82	0.058	11.8	9592	1901	99.9	5.0
3.49	0.075	9.6	10529	2085	100.0	5.0
3.23	0.117	6.4	11425	2263	99.8	5.0
3.02	0.164	4.6	10848	2397	99.9	4.5
2.85	0.205	3.7	7457	2472	97.0	3.0
2.70	0.237	3.2	5448	2316	87.2	2.4
Overall	0.067	9.9	79618	18394	97.6	4.3

† $R_{merge} = \sum_{hkl}\sum_i(|I_{hkl} - \langle I_{hkl}\rangle|)/\sum_{hkl,i}\langle I_{hkl}\rangle$, where $I_{hkl,i}$ is the intensity of an individual measurement of the reflection with indices hkl and $\langle I_{hkl}\rangle$ is the mean intensity of that reflection.

determinants for selectivity (Rould *et al.*, 1989; Hayase *et al.*, 1992; Frugier *et al.*, 1994).

2. Purification, crystallization and diffraction data collection

E. coli CysRS was expressed in *E. coli* strain JM109 containing the plasmid pKK107 (Hamann & Hou; 1997*b*; Hou *et al.*, 1991). 3 l of cells were cultured in 2× YT media, induced with 0.3 m*M* IPTG at $A_{600} = 0.5$ and grown further for 5 h at 310 K. Cells were harvested and pellets washed in cold 20 m*M* potassium phosphate (pH 7.4), 50 m*M* NaCl, 20 m*M* β-mercaptoethanol, 10% glycerol (buffer *A*) and frozen at 253 K. Cells were thawed, resuspended in buffer *A* with the addition of 1 m*M* PMSF and lysed by sonication. The cell lysate was loaded onto a DEAE cellulose column (Sigma) and washed with buffer *A*. CysRS was eluted with a linear NaCl gradient (50–500 m*M*); the enzyme eluted at approximately 80 m*M* NaCl. Fractions containing CysRS were identified by SDS–PAGE, pooled and dialyzed into buffer *A*. The CysRS was then loaded onto a hydroxyapatite column (BioRad) and eluted by washing with buffer *A*. CysRS does not bind to hydroxyapatite under these low-salt conditions at pH 7.4. The hydroxyapatite wash containing CysRS was analyzed by SDS–PAGE for purity, pooled and concentrated to approximately 10–15 mg ml^{-1} using an Amicon ultrafiltration cell. Purified concentrated CysRS was dialyzed into 20 m*M* potassium phosphate (pH 7.4), 50 m*M* NaCl, 50% glycerol and stored at 253 K. A final yield of approximately 80 mg of homogeneous protein, estimated at 99% purity on the basis of SDS–polyacrylamide gel electrophoresis, was obtained beginning from a 3 l growth of cells.

Crystals of CysRS were obtained at 300 K using the hanging-drop vapor-diffusion method. An initial screen with ammonium sulfate produced bundles of needles, which were resolved into large single crystals upon supplementation of the mother liquor with 3% ethylene glycol. Large crystals were obtained by mixing 4 µl of protein solution, containing 8–12 mg ml^{-1} CysRS, 10 m*M* HEPES (pH 7.4), 50 m*M* NaCl, 1 m*M* DTT, 5 m*M* ATP, 10 m*M* cysteine, with 2 µl of reservoir solution containing 2.1 *M* ammonium sulfate, 100 m*M* Tris (pH 8.2), 3.2% ethylene glycol. Hexagonal rods of maximum dimensions 0.5 × 0.25 × 0.25 mm can be reproducibly grown within one week at 300 K. Both ATP and cysteine are required for crystal growth.

The CysRS crystals were characterized on an R-AXIS IIc image-plate detector mounted on a Rigaku RU-200 rotating-anode source operating at 50 kV, 100 mA with Cu $K\alpha$ radiation. Data were collected at 100 K by flash-freezing crystals in a nitrogen-gas stream (Oxford Cryosystems Cryostream). Crystals were stabilized in an artificial mother liquor containing all components of the growth medium, with the addition of further ammonium sulfate to a concentration of 2.6 *M*, and were cryoprotected in a solution containing 2.6 *M* ammonium sulfate, 100 m*M* Tris (pH 8.2), 20% ethylene glycol. Data were collected at a crystal-to-detector distance of 175 mm, with 1° oscillation images and an exposure time of 2400 s per frame. A complete native data set was collected in 89 frames from one frozen crystal. The oscillation frames were indexed and integrated with *MOSFLM* and scaled with *SCALA* from the *CCP*4 software package (Collaborative Computational Project, Number 4, 1994). The data are 97.6% complete to 2.7 Å with $R_{merge} = 6.7\%$ (Table 1). The $I/\sigma(I)$ value in the highest 0.1 Å resolution shell is 3.2. The space group is *P*321 with unit-cell dimensions $a = b = 82.3$, $c = 168.9$ Å. The calculated Matthews coefficient of 3.27 Å^3 Da^{-1} indicates one molecule per asymmetric unit (M_r = 52201 Da; Hou *et al.*, 1991) with a solvent content of 62% (Matthews, 1968). Crystals are being soaked in a variety of heavy-atom solutions in search of useful derivatives. The presence of both ATP and cysteine in the crystals suggests that this structure may immediately yield insight into the mechanism for the first step of aminoacylation.

We thank Tim Bullock and Nancy Horton for assistance in data collection and processing. This work was partly supported by NIH grant GM47935 to Y-MH.

References

Arnez, J. G. & Moras, D. (1997). *Trends Biochem. Sci.* **22**, 211–216.
Carter, C. W. (1993). *Annu. Rev. Biochem.* **62**, 715–748.
Cavarelli, J., Delagoutte, B., Eriani, G., Gangloff, J. & Moras, D. (1998). *EMBO J.* **17**, 5438–5448.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cusack, S., Berthet-Colominas, C., Hartlein, M., Nassar, N. & Leberman, R. (1990). *Nature (London)*, **347**, 249–255.
Cusack, S. (1997). *Curr. Opin. Struct. Biol.* **7**, 881–889.
Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990). *Nature (London)*, **347**, 203–206.
Fersht, A. R. & Dingwall, C. (1979). *Biochemistry*, **18**, 1245–1249.
Frugier, M., Söll, D., Giegé, R. & Florentz, C. (1994). *Biochemistry*, **33**, 9912–9921.
Giegé, R., Puglisi, J. D. & Florentz, C. (1993). *Prog. Nucleic Acids Res. Mol. Biol.* **45**, 129–206.
Hamann, C. S. & Hou, Y.-M. (1997*a*). *Biochemistry*, **36**, 7967–7972.
Hamann, C. S. & Hou, Y.-M. (1997*b*). *Bioorg. Med. Chem.* **5**, 1011–1019.
Hayase, Y., Jahn, M., Rogers, M. J., Sylvers, L. A., Koizumi, M., Inoue, H., Ohtsuka, E. & Söll, D. (1992). *EMBO J.* **11**, 4159–4165.
Hou, Y.-M., Shiba, K., Mottes, C. & Schimmel, P. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 976–980.
Hou, Y.-M., Westhof, E. & Giege, R. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 6776–6780.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Rould, M. A., Perona, J. J., Söll, D. & Steitz, T. A. (1989). *Science*, **246**, 1135–1142.
Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. & Steinberg, S. (1998). *Nucleic Acids Res.* **26**, 148–153.