Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

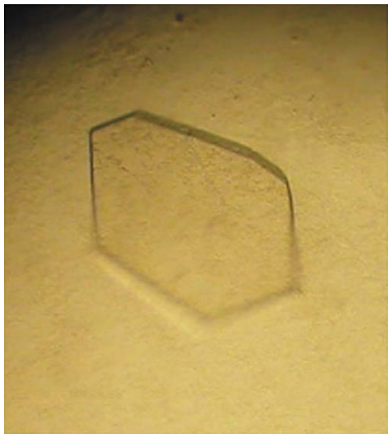
Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of the glutamate-1-semialdehyde aminotransferase from *Bacillus subtilis*

Xinhuai Lv,[a,b] Jun Fan,[a,b] Honghua Ge,[a,b] Yongxiang Gao,[a,b] Xiao Zhang,[a,b] Maikun Teng[a,b]* and Liwen Niu[a,b]*

[a]Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science and Technology of China, 96 Jinzhai Road, Hefei, Anhui 230027, People's Republic of China, and [b]Key Laboratory of Structural Biology, Chinese Academy of Sciences, 96 Jinzhai Road, Hefei, Anhui 230027, People's Republic of China

Correspondence e-mail: mkteng@ustc.edu.cn, lwniu@ustc.edu.cn

Received 20 February 2006
Accepted 11 April 2006



5-Aminolevulinic acid (ALA) is the first committed universal precursor in the tetrapyrrole-biosynthesis pathway. Plants, algae and many other bacteria synthesize ALA from glutamate by a C5 pathway in which the carbon skeleton of glutamate is converted into ALA by a series of enzymes. Glutamate-1-semialdehyde aminotransferase (GSAT) is the last enzyme in this pathway. The gene that codes for GSAT was amplified from the cDNA library of *Bacillus subtilis* and overexpressed in *Escherichia coli* strain BL21(DE3). The protein was purified and crystallized. Well diffracting single crystals were obtained by the hanging-drop vapour-diffusion method. Preliminary X-ray diffraction studies yielded excellent diffraction data to a resolution of 2.0 Å.

1. Introduction

5-Aminolevulinic acid (ALA) is the first committed precursor in the tetrapyrrole-biosynthesis pathway. Tetrapyrroles such as haems and chlorophylls are arguably among the most important pigments in the biosphere and the most versatile cofactors (Astner *et al.*, 2005). In animals and α-proteobacteria, ALA formation is catalyzed by ALA synthase through a one-step condensation reaction of succinyl-CoA and glycine (Kikuchi *et al.*, 1958). In plants and the majority of bacteria (Jahn *et al.*, 1992), *e.g. Escherichia coli* (Avissar & Beale, 1989), *Salmonella typhimurium* (Elliott, 1989) and *Bacillus subtilis* (O'Neill *et al.*, 1989), this compound is made from glutamate-1-semialdehyde (GSA) in a reaction catalyzed by GSAT. ALA is generated from glutamate in a three-step process (Beale & Weinstein, 1990). Firstly, glutamate is ligated to its cognate tRNA by the action of glutamyl-tRNA synthetase (GTS; Schon *et al.*, 1986). Activated glutamate is converted to glutamate-1-semialdehyde (GSA) by glutamyl-tRNA reductase (GTR; Wang *et al.*, 1984). GSA is then rearranged to ALA by GSAT (Hoober *et al.*, 1988).

GSAT is a pyridoxal 5-phosphate (PLP) dependent enzyme and addition of PLP has been shown to enhance the activity of a plant GSAT (Tsang *et al.*, 2003). Enzymes using PLP cofactors provide an excellent example of the evolutionary development of catalytic versatility in homologous enzymes. Thus, the family of PLP-dependent enzymes (Alexander *et al.*, 1994) contains aminotransferases, racemases, decarboxylases, mutases and synthetases. In order to shed light on the mechanism of GSAT and investigate the regulation of the genes of the C5 pathway in plants, the three-dimensional structure of GSAT is required. Such a structure will also contribute to our understanding of how related proteins that catalyze different reactions confer different reaction specificities on the coenzyme. In this paper, we present the purification and crystallization of the GSAT from *B. subtilis*.

2. Materials and methods

2.1. Cloning, expression and purification of GSAT

2.1.1. Cloning. The primers used were 5′-CCGGCTC**CATATG**-AGAAGCTATGAAAAATC-3′ and 5′-TAC**CTCGAG**TTATCTGC-GGCTGATCTCG-3′. They contained *Nde*I and *Xho*I restriction sites, which are depicted in bold. The predicted open reading frame coding for GSAT was amplified by PCR using a cDNA library

Table 1
Data-collection and reduction statistics.

Values in parentheses are for the last shell.

Space group	*C*2
Unit-cell parameters (Å, °)	a = 111.428, b = 56.496, c = 81.028, β = 131.928
Molecules per ASU	1
V_M (Å^3 Da^{-1})	2.0
Resolution limits (Å)	20–2.0 (2.07–2.00)
$I/\sigma(I)$	8.5 (1.6)
Observations	60652
Independent reflections	22786
R_{merge}† (%)	8.5 (32.3)
Completeness‡ (%)	89.3 (67.8)

† $R_{merge} = \sum_h \sum_j |I(h)_j - \langle I(h)\rangle| / \sum_h \sum_j I(h)_j$, where $I(h)_j$ is the observed intensity and $\langle I(h)\rangle$ is the mean intensity of reflection h. ‡ Completeness is the ratio of the number of reflections to the number of possible reflections.

obtained from RT-PCR of total *B. subtilis* RNA. In order to place a hexa-His tag at the N-terminus of the expressed protein, the PCR product was cloned into the p28 vector using the *Nde*I and *Xho*I restriction sites [p28 was derived from pET28a (Novagen) by deleting the sequence AGCAGCGGCCTGGTGCCGCGCGGCAGC between the *Nco*I and *Nde*I restriction sites]. The recombinant plasmid was sequenced to ensure that no mutations had occurred during the polymerase chain reaction.

2.1.2. Expression and purification. The p28-GSAT-encoding plasmid was transformed into *E. coli* strain BL21(DE3) (Stratagene) and the cells were grown at 310 K in LB medium supplemented with 50 µg ml^{-1} kanamycin. Protein overproduction was induced at an OD_{600} of 0.6–0.8 with 1 m*M* IPTG (isopropyl β-D-thiogalactopyranoside) and growth continued for an additional 4 h. Following harvesting, the cells were resuspended in buffer *A* (20 m*M* Tris pH 8.0, 10 m*M* imidazole pH 8.0, 300 m*M* NaCl) and lysed by sonication. The supernatant containing soluble protein was obtained by centrifugation at 25 000*g* for 30 min. The supernatant was loaded onto an Ni^{2+}-charged chelating Sepharose column (Amersham) and washed first with buffer *A* and then with buffer *A* supplemented with 25 m*M* imidazole. The recombinant protein was finally eluted with 20 m*M* Tris pH 8.0, 75 m*M* imidazole and 300 m*M* NaCl. The yield was approximately 20 mg protein per litre of bacterial culture. The recombinant protein was about 98% homogenous as assessed by SDS–PAGE (Fig. 1): samples (10 or 15 µg) were denatured in Laemmli sample buffer (5 min at 373 K) and then applied onto the gel (12 or 15% acrylamide). Purified protein was also analysed for homogeneity by dynamic light scattering (DLS) using a DynaPro-MS800 instrument (DynaPro, USA). DLS results indicated a monomodal distribution of the purified protein, with an observed hydrodynamic radius (R_h) of 2.1 nm (the calculated R_h is 3 nm) and 13.6% polydispersity, which falls within the range (10–15%) indicative of good homogeneity. The pure protein was desalted and concentrated to 9 mg ml^{-1} in 5 m*M* NaCl and 5 m*M* Tris pH 8.0; the concentration was estimated using the BCA Protein Assay Kit (Pierce).

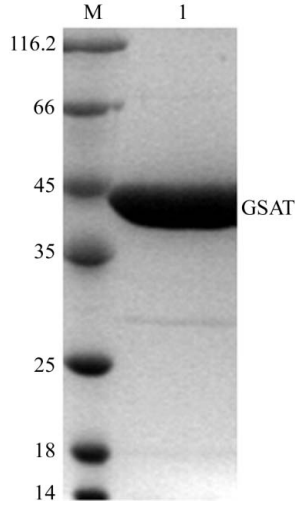


Figure 1
SDS–PAGE of GSAT. Lane 1, purified GSAT; lane M, molecular-weight markers (kDa).

2.2. Crystallization

Crystallization trials were performed using the hanging-drop vapour-diffusion method. Initial crystallization attempts were performed at 295 K using Crystal Screens I and II (Hampton Research), which led to microcrystalline precipitate under several conditions (Nos. 14, 40 and 88) in times ranging from one to two weeks. Optimization of the crystallization conditions was performed varying the pH, protein concentration, ionic strength, precipitant concentration, temperature, cationic compounds, additives and protein reservoir volume ratio. The very thin GSAT crystals were prone to break when transferred into cryoprotectant solutions and so different concentrations of various cryoprotectants were tested as additives in the reservoir solution. 2 µl protein sample and 1 µl reservoir were mixed thoroughly and equilibrated against 400 µl reservoir solution. Crystals (Fig. 2) were obtained within two weeks under the following optimized conditions: reservoir, 5 m*M* sodium chloride, 100 m*M* Bicine pH 8.5, 30%(*w*/*v*) PEG 3350; protein concentration, 8–9 mg ml^{-1}; temperature, 298 K.

2.3. X-ray diffraction data collection and preliminary analysis

Crystals were quickly passed through a cryoprotectant solution [5 m*M* sodium chloride, 100 m*M* Bicine pH 8.5, 30%(*w*/*v*) PEG 3350, 15%(*v*/*v*) glycerol] and directly flash-cooled in a stream of cold nitrogen gas at 105 K using an Oxford Cryosystems cooling device (Oxford Cryosystems Ltd, England). X-ray diffraction data were collected using a MAR 345 dtb imaging-plate detector (MAR

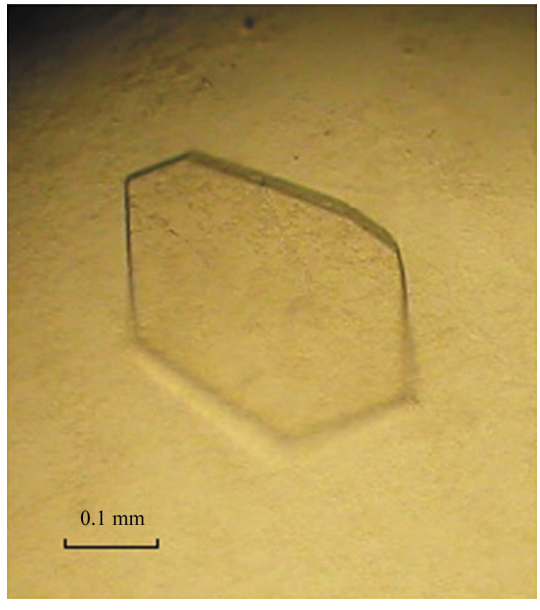


Figure 2
Photomicrograph of a GSAT crystal.

Research, Germany) with a rotating-anode X-ray generator (copper anode; RA-Micro007, Rigaku, Japan). The oscillation angle was 1.0° and the exposure time was 20 min per frame. A total of 129 diffraction images were recorded at a crystal-to-detector distance of 150 mm and the data were processed using *AUTOMAR* v.1.4 (MAR Research GmbH). The data-collection and processing statistics are listed in Table 1. The crystals belong to the monoclinic space group *C*2, with unit-cell parameters $a = 111.428$, $b = 56.496$, $c = 81.028$ Å, $\beta = 131.928°$. Assuming the presence of one molecule per asymmetric unit, the Matthews coefficient (V_M) was calculated to be 2.0 Å^3 Da^{-1}, which corresponds to 39.4% solvent content.

2.4. Results and discussion

GSAT expressed using both the pET28a and p28 vectors was overproduced in *E. coli*, purified and crystallized. Crystals of GSAT obtained from protein expressed using the pET28a vector were beautifully shaped but diffracted to lower resolution (3.0 Å) than GSAT crystals obtained from protein expressed using the p28 vector. GSAT expressed from the p28 vector has nine fewer residues between the N-terminal hexa-His tag and the GSAT sequence than protein expressed from the pET28a vector.

GSATs from *B. subtilis* and *Synechococcus* share 58% amino-acid sequence identity. According to the structure of GSAT from *Synechococcus* (Hennig *et al.*, 1997), the structure of the *B. subtilis* enzyme may also contain three domains. The difference between the two structures should illuminate the structural basis of the function of these enzymes and the functional difference that exists between the *Synechococcus* and *B. subtilis* GSAT enzymes (Smith *et al.*, 1998).

Financial support for this project to LN and MT was provided by research grants from the Chinese National Natural Science Foundation (grant Nos. 30121001, 30025012, 30130080 and 30571066) and the '973' and '863' Plans of the Chinese Ministry of Science and Technology (grant Nos. 2004CB520801 and 2002BA711A13).

References

Alexander, F. W., Sandmeier, E., Mehta, P. K. & Christen, P. (1994). *Eur. J. Biochem.* **219**, 953–960.
Astner, I., Schulze, J. O., Heuvel, J., Jahn, D., Schubert, W. D. & Heinz, D. W. (2005). *EMBO J.* **24**, 3166–3177.
Avissar, Y. J. & Beale, S. I. (1989). *J. Bacteriol.* **171**, 2919–2924.
Beale, S. I. & Weinstein, J. D. (1990). *Biosynthesis of Heme and Chlorophyll*, edited by H. A. Dailey, pp. 287–391. New York: McGraw–Hill.
Elliott, T. (1989). *J. Bacteriol.* **171**, 3948–3960.
Hennig, M., Grimm, B., Contestabile, R., John, R. A. & Jansonius, J. N. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 4866–4871.
Hoober, J. K., Kahn, A., Ash, D. E., Gough, S. P. & Kannangara, C. G. (1988). *Carlsberg Res. Commun.* **53**, 11–25.
Jahn, D., Verkamp, E. & Söll, D. (1992). *Trends Biochem. Sci.* **17**, 215–218.
Kikuchi, G., Kumar, A., Talmage, P. & Shemin, D. (1958). *J. Biol. Chem.* **233**, 1214–1219.
O'Neill, G. P., Chen, M. W. & Soll, D. (1989). *FEMS Microbiol. Lett.* **51**, 255–260.
Schon, A., Krupp, G., Gough, S., Berry-Lowe, S., Kannangara, C. G. & Soll, D. (1986). *Nature (London)*, **322**, 281–284.
Smith, M. A., King, P. J. & Grimm, B. (1998). *Biochemistry*, **37**, 319–329.
Tsang, E. W., Hu, Z., Chang, Q., McGregor, D. I. & Keller, W. A. (2003). *Protein Expr. Purif.* **29**, 193–201.
Wang, W. Y., Huang, D. D., Stachon, D., Gough, S. P. & Kannangara, C. G. (1984). *Plant Physiol.* **74**, 569–575.