Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallographic study of a naturally occurring *trans*-splicing intein from *Synechocystis* sp. PCC 6803

Xuehui Chen,[a] Ming-Qun Xu,[b] Yi Ding,[a] Sébastien Ferrandon[b] and Zihe Rao[a]*

[a]Laboratory of Structural Biology, Department of Biological Science and Technology and MOE Laboratory of Protein Science, Tsinghua University, Beijing 100084, People's Republic of China, and [b]New England Biolabs, Inc., 32 Tozer Road, Beverly, MA 01915, USA

Correspondence e-mail: raozh@xtal.tsinghua.edu.cn

Received 26 November 2001
Accepted 11 April 2002

A naturally occurring split intein from the *dna*E gene of *Synechocystis* sp. PCC6803 (*Ssp* DnaE intein) has been purified and crystallized using PEG 8K as precipitant. The crystal belongs to space group $P3_121$ or $P3_221$, with unit-cell parameters $a = b = 58.5$, $c = 70.2$ Å. It has one molecule per asymmetric unit and diffracts to beyond 2.9 Å under cryoconditions (110 K) using a Cu rotating-anode X-ray generator in-house.

1. Introduction

Post-translational autoprocessing involving peptide-bond rearrangement has been observed in a variety of systems, including protein splicing, autoprocessing of hedgehog proteins, autocleavage of amidohydrolases and pyruvoyl enzyme formation (Paulus, 1998). A protein-splicing element was first discovered in the VMA gene of *Saccharomyces cerevisiae* (Kane *et al.*, 1990). The mechanism of protein splicing has been well investigated (Xu & Perler, 1996; Chong *et al.*, 1996; Shao *et al.*, 1996; Southworth *et al.*, 2000). Naturally occurring inteins can be grouped into three categories: inteins that contain a homing endonuclease between the splicing domains, inteins that lack a homing endonuclease region (mini-inteins) and inteins in which the splicing domain is split and are capable of *trans*-splicing (Perler, 2000).

To date, more than 100 inteins have been identified in organisms from the eukaryotic, archaeal and eubacterial kingdoms (Perler, 2000). Only three of them have three-dimensional structure data deposited in the PDB, one of which belongs to a mini-intein (Duan *et al.*, 1997; Ichiyanagi *et al.*, 2000; Klabunde *et al.*, 1998).

The only naturally occurring split intein was found in the *dna*E gene encoding the catalytic subunit of *Synechocystis* sp. PCC6803 DNA polymerase III, which is separated into two coding regions (Wu *et al.*, 1998). These two split genes are located 745 kbp apart on the genome and are found on opposite strands of the DNA. The first split gene encodes the 774 N-terminal amino-acid residues of the DnaE protein fused to the 123 N-terminal amino-acid residues of the *Ssp* DnaE intein sequence, while the second fragment consists of the 36 C-terminal amino-acid residues of the *Ssp* DnaE intein followed by the 423 C-terminal amino acids of the DnaE polymerase. When the two intein halves are coexpressed in *Escherichia coli*, they are capable of *trans*-splicing to produce intact DNA polymerase III (Wu *et al.*, 1998). This naturally occurring split intein can mediate efficient *trans*-splicing *in vivo* and *in vitro* when expressed in a foreign protein context (Evans *et al.*, 2000; Scott *et al.*, 1999). Unlike other artificially split inteins, the *Ssp* DnaE intein does not require a urea-treatment step, which may enhance its general utility for protein semisynthesis (Evans *et al.*, 2000; Martin *et al.*, 2001; Ghosh *et al.*, 2001). Protein *trans*-splicing technology provides many useful applications, including segmental labelling of proteins for NMR analysis (Yamazaki *et al.*, 1998), ordered three-fragment ligation (Otomo *et al.*, 1999) and cyclization of proteins and peptides (Evans *et al.*, 2000; Scott *et al.*, 1999). Interestingly, when the *Ssp* DnaE intein was expressed as a single polypeptide, it exhibited essentially the same splicing efficiency as the split intein segments coexpressed in *E. coli* (Evans *et al.*, 2000).

In this report, we employed an intein-mediated purification scheme to isolate the 20 kDa *Ssp* DnaE intein along with its flanking extein sequences. In order to purify the intein in its pre-splicing form, mutation of the intein catalytic residues was performed. We believe that this new structure may provide insight into the interaction of the two intein halves and the structural basis of the protein *trans*-splicing process.

2. Results and discussion

2.1. Purification of *Ssp* DnaE intein

2.1.1. Construction of the plasmid. The pTWIN vector (Evans *et al.*, 1999) was used for expression and purification of a precursor protein containing the *Ssp* DnaE intein in

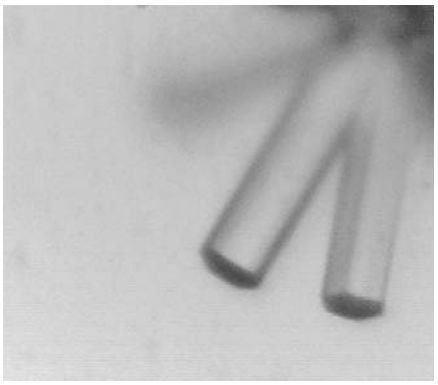

Figure 1
Crystals of intein DnaE mini-intein grown in 10% PEG 4K pH 6.5, with approximate dimensions of 0.3 × 0.1 × 0.1 mm.

E. coli. The DNA sequence encoding the *Ssp* DnaE intein consisting of the 123-residue N-terminal fragment plus the 36-residue C-terminal fragment along with its five native N-extein residues and four native C-extein residues was transferred from pMEB3 (Evans *et al.*, 2000) into the pTWIN vector in frame with the C-terminus of the *Ssp* DnaB intein, which had been modified for cleavage at its C-terminal junction by a pH and temperature shift. The resulting plasmid pBSE uses an IPTG-inducible T7 promoter and directs the expression of a fusion protein consisting of the chitin-binding domain (CBD) from *Bacillus circulans*, the DnaB intein followed by the full-length *Ssp* DnaE intein.

2.1.2. Protein expression and purification. Purification of the *Ssp* DnaE intein precursor was conducted by an intein-mediated cleavage procedure as described previously (Xu *et al.*, 2000). The recombinant plasmid pBSE was transformed into ER2566 cells (New England Biolabs, Inc.) and the resulting transformant was grown at 310 K until an OD_{600} of approximately 0.5 was reached. Protein expression was induced by the addition of IPTG to a final concentration of 0.3 m*M* and cultures were grown overnight at 288 K with shaking. Cells were harvested by centrifugation at 277 K and lysed by sonication (over ice). Cellular debris was removed by centrifugation at 23 000*g* for 30 min. The fusion protein was purified at 277 K using chitin resin (New England Biolabs, Inc.) and cleavage was induced by quickly washing the column with three column volumes of buffer (40 m*M* sodium phosphate pH 6.5, 500 m*M* NaCl) and incubating overnight at 291 K. Fractions were collected after overnight incubation and those containing protein were buffer-exchanged to 20 m*M* Tris–HCl pH 7.5 using a Sephadex G25 column (Amersham Pharmacia Biotech). The desalted solution was applied to a Resource Q column and the DnaE intein precursor was eluted with a linear gradient of 20 m*M* Tris–HCl pH 8.5, 500 m*M* NaCl. The peak containing the DnaE intein precursor was further applied to a Superdex 75 column pre-equilibrated with 10 m*M* Tris–HCl pH 7.5, 25 m*M* NaCl, 10 m*M* β-mercaptoethanol and eluted with the same buffer. Homogeneity was demonstrated by SDS–PAGE.

The protein concentration was determined by absorption spectroscopy using a molar extinction coefficient at 280 nm of $A_{1\%} = 6.82$ mg ml^{-1}. Prior to crystallization, the protein was concentrated to approximately 12 mg ml^{-1} using a 10K ultrafiltration membrane (Amicon).

2.2. Crystallization

Crystallization trials were set up using the hanging-drop vapour-diffusion method on Linbro crystallization plates at 291 K. Initial screening of crystallization conditions using the sparse-matrix method (Jancarik & Kim, 1991) with commercially available buffers (Hampton Research) resulted in small needle-like crystals with 30% PEG 8K, 0.2 *M* sodium acetate, 100 m*M* cacodylate pH 6.5. Crystals large enough for X-ray diffraction experiments were grown in ~10% PEG 8K, 0.2 *M* sodium acetate, 100 m*M* cacodylate pH 6.5 (Fig. 1). A 1 µl droplet of 12 mg ml^{-1} DnaE solution in 10 m*M* Tris–HCl pH 7.5, 25 m*M* NaCl, 5 m*M* DTT was mixed with an equal volume of reservoir solution (100 m*M* cacodylate buffer pH 6.5 containing 8–12% PEG 4K, 0.2 *M* sodium acetate). This droplet was equilibrated against 200 µl of reservoir solution. Rod-like crystals grew to a typical size of 0.2–0.3 mm along one edge in three months.

2.3. X-ray analysis

Data collection was carried out at 100 K using a MAR345 image plate with an in-house Rigaku rotating Cu anode X-ray generator operating at 48 kV and 98 mA (λ = 1.5418 Å). Paraffin oil was used as the cryoprotectant. Prior to data collection, the crystal was soaked in the paraffin oil for a few seconds. The crystal was then flash-cooled in a nitrogen-gas stream at 100 K. The DnaE intein crystal diffracted beyond 2.8 Å and a set of data was collected at 2.9 Å with an oscillation range of 1.5° per image. The exposure time was 60 s per frame (Fig. 2). All diffraction data were processed using the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). For one molecule per asymmetric unit (Z = 1),

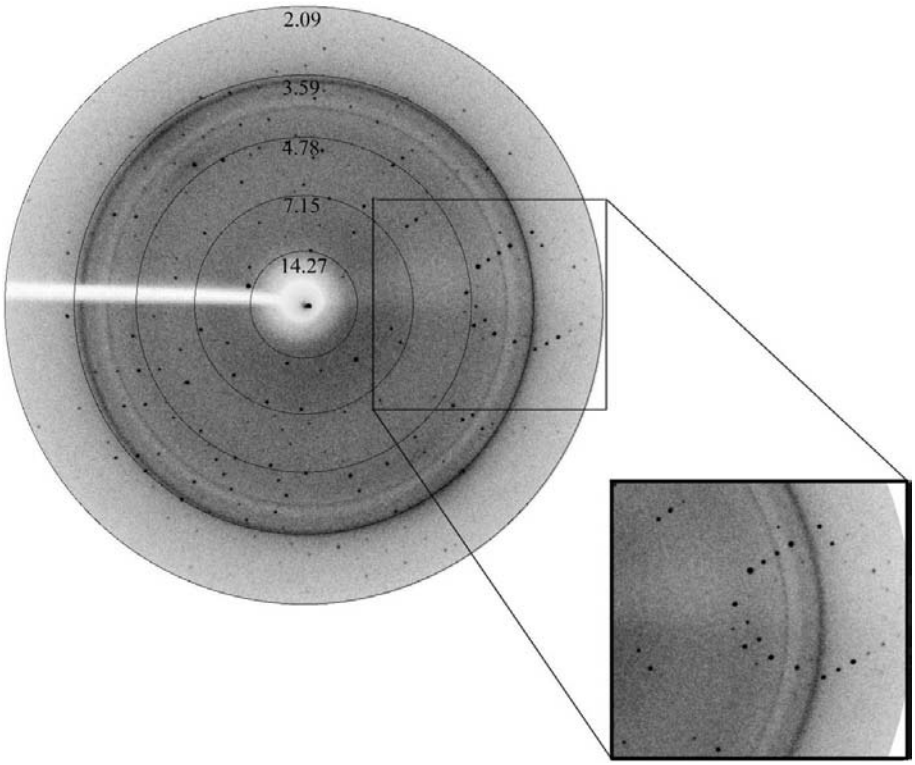

Figure 2
X-ray diffraction image of a DnaE crystal. (*a*) A 1.5° oscillation image taken from a crystal of DnaE mini-intein at 100 K. This image was taken in-house on a MAR345 detector. The resolution is 2.9 Å at the edge of the plates. (*b*) Enlargement of one corner, showing the spots at the outermost resolution.

Table 1
Crystallographic parameters and data-collection statistics.

Values in parentheses are for the last resolution shell (2.99–2.90 Å).

Crystal data	
Space group	$P3_121$ or $P3_221$
Unit-cell parameters (Å)	$a = b = 58.5$, $c = 70.2$
Data collection	
Resolution (Å)	2.9 (2.99–2.90)
No. of observations	29517
No. of unique reflections	3326
R_{merge} (%)	8.4 (27.3)
Completeness (%)	100 (96.8)
Average $I/\sigma(I)$	14.7
Mean redundancy	8.8

the Matthews coefficient is 1.74 Å^3 Da^{-1} and the solvent content is 28% (Matthews, 1968). Crystal data and data-collection statistics are listed in Table 1.

We are grateful to Rong Cao, Feng Xu and Zhicheng Wang for technical assistance, and to Feng Gao for assistance with data collection. New England Biolabs and Dr Donald Comb generously supported this research. We thank Inca Ghosh and Thomas Evans for valuable discussion and reading of the manuscript. This research was supported by the following grants: National Natural Sciences Foundation of China No. 39871074, NSFC Nos. 39870174 and 39970155, Project '973' Nos. G1999075602, G1999011902 and 1998051105.

References

Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, F. B. & Xu, M.-Q. (1996). *J. Biol. Chem.* **271**, 22159–22168.
Duan, X., Gimble, F. S. & Quiocho, F. A. (1997). *Cell*, **89**, 555–564.
Evans, T. C. Jr, Benner, J. & Xu, M.-Q. (1999). *J. Biol. Chem.* **274**, 18359–18363.
Evans, T. C. Jr, Martin, D., Kolly, R., Panne, D., Sun, L., Ghosh, I., Chen, L., Benner, J., Liu, X. Q. & Xu, M. Q. (2000). *J. Biol. Chem.* **275**, 9091–9094.
Ghosh, I., Sun, L. & Xu, M. Q. (2001). *J. Biol. Chem.* **276**, 24051–24058.
Ichiyanagi, K., Ishino, Y., Ariyoshi, M., Komori, K. & Morikawa, K. (2000). *J. Mol. Biol.* **300**, 889–901.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goeble, M. & Stevens, T. H. (1990). *Science*, **250**, 651–657.
Klabunde, T., Sharma, S., Telenti, A., Jacobs, W. R. & Sacchettini, J. C. (1998). *Nature Struct. Biol.* **5**, 31–36.
Martin, D., Xu, M. Q. & Evans, T. C. (2001). *Biochemistry*, **40**, 1393–1402.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otomo, T., Ito, N., Kyogoku, Y. & Yamazaki, T. (1999). *Biochemistry*, **38**, 16040–16044.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Paulus, H. (1998). *Chem. Soc. Rev.* **27**, 375–386.
Perler, F. B. (2000). *Nucleic Acids Res.* **28**, 344–345.
Scott, C. P., Abel-Santos, E., Wall, M., Wahnon, D. C. & Benkovic, S. J. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 13638–13643.
Shao, Y., Xu, M. Q. & Paulus, H. (1996). *Biochemistry*, **35**, 3810–3815.
Southworth, M. W., Benner, J. & Perler, F. B. (2000). *EMBO J.* **19**, 5019–5026.
Wu, H., Xu, M. Q. & Liu, X. Q. (1998). *Biochim. Biophys. Acta*, **1387**, 422–432.
Xu, M. Q., Paulus, H. & Chong, S. (2000). *Methods Enzymol.* **326**, 376–418.
Xu, M. Q. & Perler, F. B. (1996). *EMBO J.* **15**, 5146–5153.
Yamazaki, T., Otomo, T., Oda, N., Kyogoku, Y., Uegaki, K., Ito, N., Ishino, Y. & Nakamura, H. (1998). *J. Am. Chem. Soc.* **120**, 5591–5592.