Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Purification and initial crystallization studies of a DnaB 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 154-residue mini-intein from the *dna*B gene of *Synechocystis* sp. PCC 6803 (*Ssp* DnaB intein) has been purified and crystallized using PEG 4000 as a precipitant. The crystal belongs to space group $P3_121$ or $P3_221$, with unit-cell parameters $a = b = 58.2$, $c = 70.3$ Å. It has one molecule per asymmetric unit and diffracts to beyond 2.0 Å under cryoconditions (100 K) using a rotating copper anode X-ray generator.

1. Introduction

Protein splicing is a post-translational processing event in which an internal protein domain (termed an intein) is excized from a precursor protein with the concomitant formation of a peptide bond between the adjacent protein sequences (termed exteins; reviewed in Paulus, 1998, 2000, 2001; Perler, 1998; Shao & Kent, 1997). The mechanism of protein splicing has been well studied and has been described previously (Xu & Perler, 1996; Chong *et al.*, 1996). Approximately 100 inteins have so far been identified in organisms from all three kingdoms of life (Perler, 2000). Known inteins share little overall sequence identity, except between closely related inteins found at the same insertion site in homologous proteins of different organisms (Perler *et al.*, 1997).

Ten conserved intein motifs have been identified, six (blocks A–H) of which are found in inteins of all sizes and presumably function in protein splicing, whereas four (blocks C–E and H) occur only in large inteins and function in homing endonuclease activity (Pietrokovski, 1998) Only two residues are absolutely conserved in all inteins, a His in block B and the intein C-terminal Asn. Ser/Thr/Cys are found on the C-terminal side of both splicing sites.

Many inteins are bifunctional elements, possessing protein-splicing activity as well as endonuclease activity (Gimble & Thorner, 1992; Belfort & Roberts, 1997). Recently reported structures of the VMA intein (PI-SceI) from *Saccharomyces cerevisiae* (Duan *et al.*, 1997) and PI-PfuI from *Pyroccocus furiosus* (Ichiyanagi *et al.*, 2000) revealed a two-domain protein, with the endonuclease activity in one domain and the protein-splicing active site embedded in the other. Deleting the endonuclease domain of the *Sce* VMA intein and the *Mtu* RecA intein from *Mycobacterium tuberculosis* has produced mini-inteins that are capable of protein splicing (Chong & Xu, 1997; Derbyshire *et al.*, 1997). The GryA intein from *M. xenopi* naturally lacks an endonuclease domain and was shown to be capable of protein splicing (Telenti *et al.*, 1997).

The *dna*B gene encoding the DNA helicase of the cyanobacterium *Synechocystis* sp. strain PCC6803 (*Ssp*) contains an intein of 429 amino-acid residues. This intein has been recognized as a theoretical intein based on the presence of intein-like sequence motifs (Pietrokovski, 1996). The DnaB intein is capable of efficient splicing, with five native extein residues flanking both splice junctions (Wu *et al.*, 1998). A deletion of the central 275 amino-acid-residue endonuclease domain resulted in a splicing-proficient minimal intein (*Ssp* DnaB mini-intein) consisting of the N-terminal 106 residues and the C-terminal 48 residues. Based on sequence alignments, this functional mini-intein (154 amino acids) corresponds to a major part of domain I (splicing domain) of the *Sce* VMA1 intein (Duan *et al.*, 1997). The extein residues could contribute to the highly coordinated splicing process, as the presence of native N-extein and C-extein residues flanking the *Ssp* DnaB intein appears to affect the splicing and cleavage reactions (Mathys *et al.*, 1999). The crystal structure of this mini-intein with its flanking sequence will definitely aid in the understanding of the splicing process and the subtle role of the flanking sequence in splicing. Furthermore, the splicing reaction of *Ssp* DnaB suggests that this intein is intrinsically more efficient than many other inteins in a heterologous system (Wu *et al.*, 1998). This difference among inteins is both interesting and of practical importance in engineering inteins for various applications. In this study,

we have successfully crystallized the *Ssp* DnaB mini-intein with its native extein residues.

2. Results and discussion

2.1. Purification of DnaB intein

2.1.1. Construction of the plasmid. The pTWIN vector (Evans *et al.*, 1999) was used for expression and purification of the *Ssp* DnaB mini-intein in *Escherichia coli*. This plasmid contains an IPTG-inducible T7 promoter and ampicillin resistance and directs the expression of a fusion protein consisting of the 154-residue *Ssp* DnaB mini-intein along with its five native residue flanking sequences on both sides and the *Mxe* GyrA intein, which has been modified for thiol-induced cleavage of the peptide bond at its N-terminal junction, followed by the chitin-binding domain (CBD) from *Bacillus circulans*. The splicing activity of the *Ssp* DnaB intein was inhibited by replacement of the two catalytic residues, Cys1 and Asn154, with alanine residues.

2.1.2. Protein expression and purification. Purification of the *Ssp* DnaB intein was conducted by an intein-mediated cleavage method as described previously (Xu *et al.*, 2000). Plasmid pTWIN was transformed into *E. coli* 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) in a buffer containing 20 m*M* Tris–HCl pH 8.5, 500 m*M* NaCl, 10 m*M* β-mercaptoethanol (buffer *A*). 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.) in buffer *A*. Cleavage was induced by quickly washing the column with three column volumes of the same buffer containing 30 m*M* DTT and incubating overnight at 277 K. Fractions were collected after overnight incubation and dialyzed twice against 20 m*M* Tris–HCl pH 8.5, 10 m*M* β-mercaptoethanol at 277 K. After dialysis, the protein sample was applied to a Resource Q column (Amersham Pharmacia Biotech) and eluted with a linear gradient of NaCl. The peak containing the DnaB mini-intein protein was further applied to a Superdex 75 column pre-equilibrated with 10 m*M* Tris–HCl, 25 m*M* NaCl pH 7.5, 10 m*M* β-mercaptoethanol and eluted with the same buffer. Purity was monitored by SDS–PAGE. (Fig. 1)

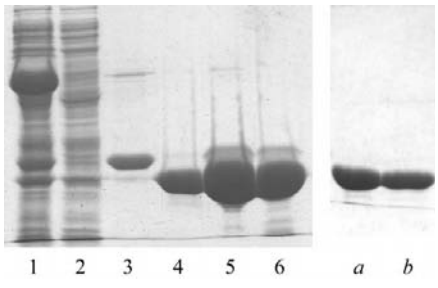

Figure 1
SDS–PAGE analysis of DnaB mini-intein. Lane 1, bacteria lysis supernatant; lane 2, flowthrough of chitin column; lane 3, fusion tag on chitin column after cleavage; lane 4, fractions from chitin column after cleavage; lane 5, peak after Resource Q; lane 6, peak after Superdex 75 column; lane *a*, protein before crystallization; lane *b*, sample from dissolved crystals.

The protein concentration was determined by absorption spectroscopy using a molar extinction coefficient at 280 nm of $A_{1\%} = 11.33$ mg ml^{-1} (theoretical prediction based on the amino-acid sequence of DnaB mini-intein). Prior to crystallization, DTT was added to the protein to a final concentration of 5 m*M* and the protein was concentrated using a 10K ultrafiltration membrane (Amicon).

2.2. Crystallization and data collection

Crystallization trials were set up using the hanging-drop vapour-diffusion method on Linbro crystallization plates at 291 K. Two different crystal forms, tetragonal and trigonal, were obtained. Data collection was carried out at 100 K using a MAR345 image plate with a Rigaku rotating copper anode X-ray generator operating at 48 kV and 98 mA (λ = 1.5418 Å). Determination of unit-cell parameters and integration of reflections were performed using the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

2.2.1. Tetragonal form. Initial screening for crystallization conditions using the sparse-matrix method (Jancarik & Kim, 1991) was set up by mixing equal volumes of 10 mg ml^{-1} protein solution with commercially available buffers (Hampton Research). Small bipyramidal crystals emerged from precipitant containing 10% PEG 4K, 0.2 *M* sodium acetate, 0.1 *M* sodium citrate pH 5.6. Fine tuning of the protein concentration to 4 mg ml^{-1} and the PEG 4K concentration to 8–12% in the pH range 5.3–5.6 resulted in crystals large enough for X-ray diffraction experiments. These bipyramidal morphology crystals grew to a typical size of 0.4–0.5 mm along one edge in a week (Fig. 2*a*). Prior to data collection, the crystal was transferred to a cryoprotectant identical to the mother liquor but containing 15%(*v*/*v*) glycerol for several seconds. The crystal was then flash-cooled in a nitrogen-gas stream at 100 K. The distance between the crystal and the image plate was 300 mm; the exposure time was 90 s per frame. The crystal belonged to space group $P3_121$ or $P3_221$, with unit-cell parameters $a = b = 79.0$, $c = 383.2$ Å. To some extent the poor data quality might be a consequence of the long *c* axis and large unit cell, which resulted in spot overlapping. The highest useful data resolution is 4.2 Å. Data-

Table 1
Data-collection and processing statistics.

Values in parentheses are for the last resolution shell (4.18–4.00 Å for the tetragonal form and 2.01–2.07 Å for the trigonal form).

Crystal form	Tetragonal form	Trigonal form
Space group	$P3_121$ or $P3_221$	$P3_121$ or $P3_221$
Unit-cell parameters (Å)	$a = b = 79.0$, $c = 383.2$	$a = b = 58.2$, $c = 70.3$
Crystal-to-detector distance (mm)	300	120
Maximum resolution (Å)	4.0	2.01
No. of observations	105953	107903
No. of unique reflections	12551	9307
Completeness (%)	100 (100)	97.6 (75.5)
Average $I/\sigma(I)$	9.3	20.5 (7.5)
Mean redundancy	4.7	11.6
$R_{\rm merge}$	0.192 (0.366)	0.046 (0.262)
Mean mosaicity	0.39	0.46

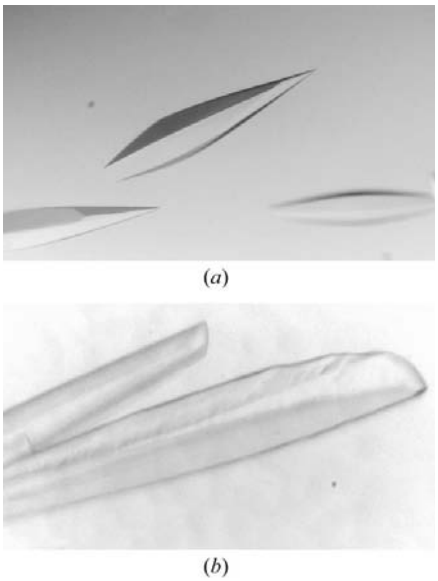

Figure 2
Crystals of the DnaB mini-intein. (*a*) Tetragonal form. Crystals were grown in citrate buffer pH 5.6 with PEG 4K as a precipitant. Maximum dimensions are about 0.5 × 0.3 mm. (*b*) Trigonal form. Crystals were grown in Tris buffer pH 7.5 with PEG 4K as a precipitant. Crystal dimensions are 0.3 × 0.1 × 0.1 mm

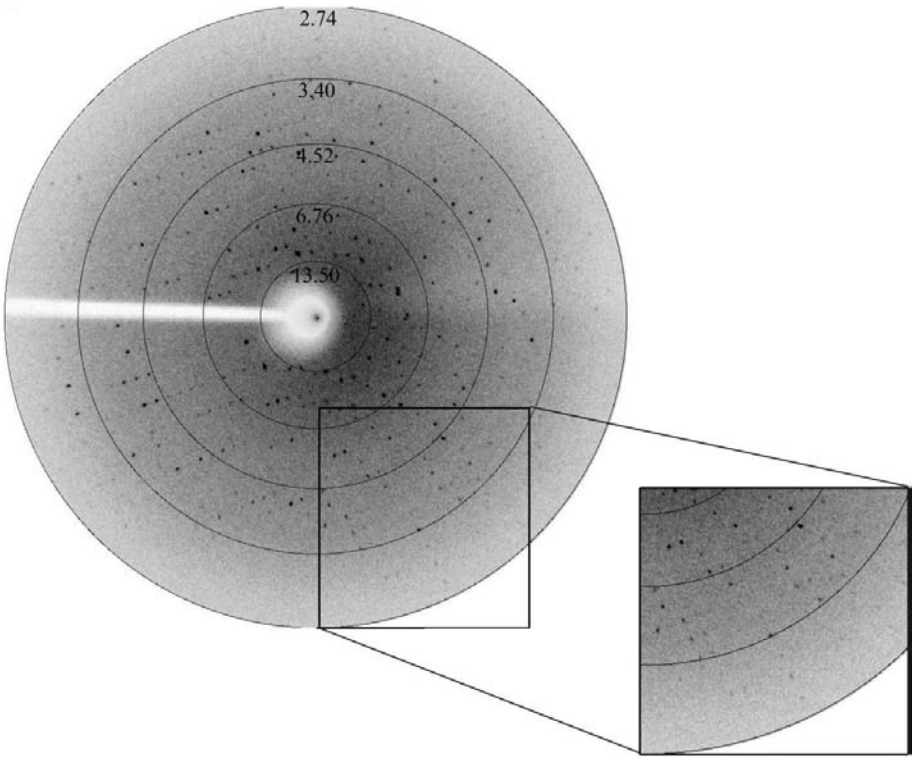

Figure 3
X-ray diffraction image from a trigonal form DnaB mini-intein crystal. (*a*) A 1.5° oscillation image taken from a crystal of DnaB mini-intein at 100 K. This image was taken in-house on a MAR345 detector. The resolution is 2.9 Å at the edge of the plates. (*b*) Enlargement of one corner, showing the spots at the highest resolution.

collection and processing statistics are listed in Table 1.

2.2.2. Trigonal form. Screening of crystallization conditions using PEG 4K as precipitant at different pH intervals gave another crystal form. A 1 µl droplet of 10 mg ml^{-1} protein 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* Tris buffer at pH 7.5–8.0 containing 16–20% PEG 4K). This droplet was equilibrated against 400 µl reservoir solution. Rod-like crystals grew to a typical size of 0.3–0.4 mm along one edge in a month. Increasing the protein concentration to 24 mg ml^{-1}, increasing the drop size to 8 µl and including 4%(*v*/*v*) ethyl glycerol in the reservoir solution gave larger crystals of 0.3 × 0.3 × 2.0 mm in three to four weeks (Fig. 2*b*). Prior to data collection, the crystal was soaked in a cryoprotectant identical to the mother liquor but containing 6%(*v*/*v*) glycerol for several seconds. The crystal was then flash-cooled in a nitrogen-gas stream at 100 K. The distance between the crystal and the image plate is 120 mm; the exposure time was 250 s per frame. The crystal belongs to space group $P3_121$ or $P3_221$, with unit-cell parameters $a = b = 58.2$, $c = 70.3$ Å (Fig. 3). For one molecule per asymmetric unit ($Z = 1$), the Matthews coefficient is 1.70 Å^3 Da^{-1} and the solvent content is 27% (Matthews, 1968). A set of data was collected at 2.0 Å resolution with an oscillation range of 1.0° per image. Crystal data and data-collection statistics are listed in Table 1.

2.3. Future work

Owing to the low homology between this DnaB mini-intein and other proteins of known structure, crystals of a selenomethionine derivative have been prepared. Several potential heavy-atom derivative data sets extending to 2.4 Å resolution have also been obtained. Both anomalous and isomorphous derivative diffraction data will be used to solve the structure.

We are grateful to Rong Cao, Feng Xu, Zhicheng Wang and Mark Bartlam 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 Nos. 39871074, and 39970155, Project '973' Nos. G1999075602, G1999011902 and 1998051105.

References

Belfort, M. & Roberts, R. (1997). *Nucleic Acids Res.* **25**, 3379–3388.
Chong, S., Shao, Y., Paulus, H, Benner, J., Perler, F. B. & Xu, M.-Q. (1996). *J. Biol. Chem.* **271**, 22159–22168.
Chong, S. & Xu, M.-Q. (1997). *J. Biol. Chem.* **272**, 15587–15590.
Derbyshire, V., Wood, D. W., Wu, W., Dansereau, J. T., Dalgaard, J. Z. & Belfort, M. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 11466–11471.
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.
Gimble, F. S. & Thorner, J. (1992). *Nature (London)*, **357**, 301–306.
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.
Mathys, S., Evans, T. C., Chute, I. C., Wu, H., Chong, S., Benner, J., Liu, X.-Q. & Xu, M.-Q. (1999). *Gene*, **231**, 1–13.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Paulus, H. (1998). *Chem. Soc. Rev.* **27**, 375–386.
Paulus, H. (2000). *Annu. Rev. Biochem.* **69**, 447–495.
Paulus, H. (2001). *Bioorg. Chem.* **29**, 119–129.
Perler, F. B. (1998). *Cell*, **92**, 1–4.
Perler, F. B. (2000). *Nucleic Acids Res.* **28**, 344–345.
Perler, F. B., Olsen, G. J. & Adam, E. (1997). *Nucleic Acids Res.* **25**, 1087–1093.
Pietrokovski, S. (1996). *Trends Genet.* **12**, 287–288.
Pietrokovski, S. (1998). *Protein Sci.* **7**, 64–71.
Shao, Y. & Kent, S. B. (1997). *Chem. Biol.* **4**, 187–194.
Telenti, A., Southworth, M., Alcaide, F., Daugelat, S., Jacobs, W. R. & Perler, F. B. (1997). *J. Bacteriol.* **179**, 6378–6382.
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.