Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Cloning, crystallization and preliminary X-ray analysis of a nucleotide-diphospho-sugar transferase spsA from *Bacillus subtilis*

Simon J. Charnock and Gideon J. Davies*

Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, England

Correspondence e-mail: davies@yorvic.york.ac.uk

Received 14 September 1998
Accepted 7 October 1998

Nucleotide-diphospho-sugar transferases represent, in terms of quantity, one of the most important groups of enzymes on Earth, yet little is known about their structure and mechanism. Such a transferase, the spsA gene product involved in the synthesis of the bacterial spore coat in *Bacillus subtilis*, has been cloned and over-expressed in an *Escherichia coli* expression system. Crystals have been grown, using PEG 8000 as a precipitant, in a form suitable for high-resolution X-ray analysis. They belong to space group $C222_1$, with unit-cell dimensions $a = 42.4$, $b = 142.0$, $c = 81.4$ Å and with one molecule of spsA in the asymmetric unit. The crystals diffract beyond 1.5 Å using synchrotron radiation.

1. Introduction

Glycosyltransferases (E.C. 2.4.*x*.*y*), catalyse the formation of glycosidic bonds. They use a donor sugar activated by the presence of a nucleotide-diphospho group and catalyse the reaction

$$\mathrm{NDP-Sugar} + \mathrm{HOR} \longleftrightarrow \mathrm{NDP} + \mathrm{Sugar-O-R}. \quad (1)$$

They are responsible for the production of di-, oligo- and polysaccharides and complex carbohydrates such as lipopolysaccharides and glycosylated proteins. They thus play a central role in food storage, structure and cellular signalling. Despite the wealth of sequences that are available for these enzymes (Campbell *et al.*, 1997), little is known about their structures or mechanisms (reviewed in Davies *et al.*, 1997).

The *spsA* gene encodes a glycosyltransferase involved in the production of the mature spore-coat during the bacterial spore response (Glaser *et al.*, 1993; Roels & Losick, 1995; Stragier & Losick, 1996). It encodes a protein of 256 amino acids whose exact substrate specificity is as yet unclear. The endospore cell wall from *B. subtilis* contains a complex peptidoglycan featuring *N*-acetylglucosamine, *N*-acetylmuramic acid and muramic acid δ-lactam (Atrih *et al.*, 1996). As such, a wide spectrum of NDP-sugars and acceptor molecules are potential substrates. Analogy with other bacterial systems suggests that a UDP- or TDP-linked sugar is the most likely donor, but the acceptor species is particularly difficult to define.

It is possible, however, to derive considerable information form the amino-acid sequence alone. A classification of nucleotide-diphospho-sugar transferases based on amino-acid sequence similarity placed spsA in family 2 (Campbell *et al.*, 1997). SpsA thus displays high sequence similarity with other glycosyltransferases involved in spore synthesis such as cgeD (Roels & Losick, 1995) and with a large number of open reading frames identified during genomic sequencing of many organisms. The substrates for these enzymes also remain elusive. The closest similarity, for an enzyme whose substrates are well characterized, is with one of the root-nodulation factors NodC, a UDP-*N*-acetylglucosamine transferase, which displays approximately 21% sequence identity with spsA. Family 2 of the glycosyltransferases contains enzymes which act with net inversion of anomeric configuration, utilizing α-linked NDP-sugars to generate products with β-anomeric configuration. The spsA structure should therefore provide a simple model system for family 2 transferases, many of which are complex membrane-bound enzymes with multiple catalytic domains. In this paper, we present the cloning and over-expression of the *spsA* gene in an *E. coli* expression system, together with the crystallization and preliminary X-ray analysis. Crystals of spsA diffract to beyond 1.5 Å using synchrotron radiation.

2. Materials and methods

2.1. Cloning, expression and purification

Vent DNA polymerase (New England Biolabs) and the oligonucleotide primers CATATGCCTAAAGTATCAGTCATT and CTCGAGGGCCGACAAGCTCTC were used to generate a PCR fragment from the genome of *B. subtilis* containing the *spsA* gene flanked by *Nde*I and *Xho*I restriction sites, respectively. The 0.79 kbp synthetic product was cloned blunt-ended into pCR-Blunt

(Invitrogen) and the DNA sequence confirmed using the manual Sequenase version 2.0 kit (United States Biochemicals). The insert was subcloned into the *Nde*I and *Xho*I restriction sites of the pET22b expression vector (Promega), generating plasmid pBSGT1. The recombinant plasmid was transformed into *Escherichia coli* strain BL21 (DE3), which had been made competent by the method of Cohen (Cohen *et al.*, 1972) and the bacteria were grown on LB agar supplemented with 50 μg ml^{-1} ampicillin at 310 K for 16 h. Cultures for induction were inoculated from a single colony and incubated at 301 K in LB supplemented with ampicillin. IPTG was added to a final concentration of 0.5 m*M* when the optical density at 600 nm had reached 0.5, after which growth was allowed to continue until the stationary phase. The bacteria were harvested by centrifugation and sonicated for 1 min in 50 m*M* Tris–HCl buffer pH 8.0. Cell debris was removed by centrifugation at 30000*g* for 30 min and buffer exchanged into 25 m*M* Tris–HCl buffer pH 8.0, using a Sephadex PD-10 G-25M column (Pharmacia Biotech).

The crude extract was loaded onto a 2.6 × 15 cm Q-SOURCE anion-exchange column (Pharmacia Biotech) and proteins eluted in a 0–250 m*M* linear gradient of NaCl. Fractions containing spsA were identified by SDS–PAGE, pooled and concentrated using a 10 kDa molecular-weight cut-off membrane (Amicon). After buffer exchange into 25 m*M* Tris–HCl buffer pH 8.0, the concentration of ammonium sulfate was slowly raised to 1.5 *M* and the spsA solution was loaded onto a 2.6 × 15 cm phenyl Sepharose hydrophobic interaction column (Pharmacia Biotech). Chromatography was achieved using a decreasing concentration of ammonium sulfate. After buffer exchange into 50 m*M* Tris–HCl buffer pH 8.0 containing 100 m*M* NaCl, the protein at 10 mg ml^{-1} was loaded, in 0.5 ml fractions, onto a HiLoad 16/60 Superdex 75 gel filtration column (Pharmacia Biotech). The purified protein was washed five times with H_2O prior to crystallization. MALDI–MS analysis of a tryptic digest of purified spsA confirmed the identity of the protein and indicated the absence of the N-terminal methionine residue (data not shown).

2.2. Crystallization, data collection and processing

Crystals of spsA were grown by vapour-phase diffusion using the hanging-drop method. The protein concentration was 15 mg ml^{-1} in 100 m*M* HEPES buffer pH 8.0, with the addition of $MgCl_2$ to a final concentration of 200 m*M*. The precipitant was 15%(*w*/*v*) polyethylene glycol 8000. A single crystal of spsA was transferred to a cryoprotectant stabilizing solution [100 m*M* HEPES buffer pH 8.0, 15%(*w*/*v*) PEG 8000, 200 m*M* $MgCl_2$, 20%(*v*/*v*) glycerol] for 5–10 s, mounted in a rayon-fibre loop and immediately placed in a stream of N_2 gas at 100 K. 100° of data were collected with an oscillation angle of 0.5° per image on beamline PX 9.6 (λ = 0.870 Å) at the Daresbury Synchrotron Radiation Source. An ADSC CCD device was used as detector. Data were processed with *MOSFLM* (A. Leslie, Medical Research Council Laboratory for Molecular Biology, Cambridge) and scaled and reduced with *SCALA* from the *CCP*4 suite of programs (Collaborative Computational Project, Number 4, 1994).

3. Results and discussion

Expression of spsA using the pET system achieved a soluble protein yield of approximately 5–10% total cell protein. Protein was purified to apparent homogeneity in three chromatographic steps, as judged by SDS–PAGE. The theoretical molecular weight from the DNA sequence is 30184 Da; MALDI–MS reveals a protein of 30056.5 Da, indicating that the N-terminal methionine residue has been processed following expression in *E. coli* strain BL21 (DE3).

Crystals of spsA appeared after approximately 3 d at 290 K. They belong to space group $C222_1$ with unit-cell dimensions a = 42.4, b = 142.0, c = 81.4 Å. A single molecule of spsA in the asymmetric unit gives a packing density of 2.04 Å^3 Da^{-1} with a corresponding solvent content of approximately 39% (Matthews, 1968). X-ray diffraction data to 1.5 Å resolution (outer resolution shell, 1.58–1.5 Å, in parentheses) are 96.1% (86%) complete with an $R_{\rm merge}$ ($\sum_{hkl}\sum_i |I_{hkl} - \langle I_{hkl}\rangle| / \sum_{hkl}\sum_i \langle I_{hkl}\rangle$) of 0.037 (0.27), a multiplicity of observation of 3.7 (3.1) observations per reflection and a mean $I/\sigma I$ of 22.5 (3.5). Sequence searches reveal no similar proteins for which three-dimensional coordinates are available and so a screen for heavy-metal derivatives is currently underway. Further crystals of spsA have been successfully grown using both micro- and macroseeding techniques.

In addition to providing information both on the mechanism of glycosidic bond formation and spore-wall synthesis in *B. subtilis*, it is important to note that family 2 of the glycosyltransferases also contains a vast array of different glycosyltransferases including both cellulose and chitin synthases. The three-dimensional structure determination of spsA should thus provide important insights into the production of the world's most abundant biopolymers.

The authors would like to thank the staff of the Daresbury Synchrotron Source for provision of data-collection facilities and Dr J. R. Startin of the Central Science Laboratory for MALDI-MS measurements. The Structural Biology Laboratory at the University of York is supported by the BBSRC. GJD is a Royal Society University Research fellow.

References

Atrih, A., Zöllner, P., Allmaier, G. & Foster, S. J. (1996). *J. Bacteriol.* **178**, 6173–6183.
Campbell, J. A., Davies, G. J., Bulone, V. & Henrissat, B. (1997). *Biochem J.* **326**, 929–942.
Cohen, S. N., Chang, A. C. Y. & Hsu, L. (1972). *Proc. Natl Acad. Sci. USA*, **69**, 2110–2114.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Davies, G., Sinnott, M. L. & Withers, S. G. (1997). *Comprehensive Biological Catalysis*, edited by M. L. Sinnott, Vol. 1, pp. 119–209. London: Academic Press.
Glaser, P., Kunst, F., Arnaud, M., Coudart, M.-P., Gonzales, W., Hullo, M.-F., Ionescu, M., Lubochinsky, B., Marcelino, L., Moszer, I., Presecan, E., Santana, M., Schneider, E., Schweizer, J., Vertes, A., Rapoport, G. & Danchin, A. (1993). *Mol. Microbiol.* **10**, 371–384.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Roels, S. & Losick, R. (1995). *J. Bacteriol.* **177**, 6263–6275.
Stragier, P. & Losick, R. (1996). *Annu. Rev. Genet.* **30**, 297–341.