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

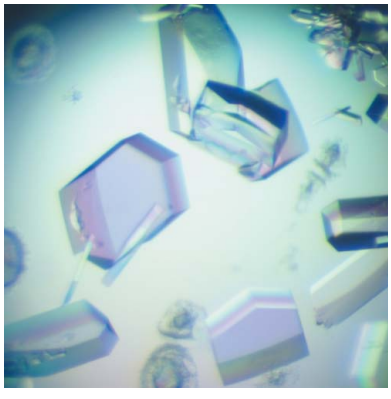
Expression, purification, crystallization and preliminary crystallographic analysis of SpaA, a major pilin from *Corynebacterium diphtheriae*

Hae Joo Kang,[a] Neil G. Paterson[a] and Edward N. Baker[a,b]*

[a]School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand, and [b]Department of Chemistry, University of Auckland, Private Bag 92019, Auckland, New Zealand

Correspondence e-mail: ted.baker@auckland.ac.nz

Received 9 June 2009
Accepted 14 July 2009



Bacterial pili are cell-surface organelles that are critically involved in adhesion to host cells, leading to the colonization of host tissues and the establishment of infections. Whereas the pili of Gram-negative bacteria have been extensively studied, those of Gram-positive bacteria came to light only recently after the discovery and characterization of *Corynebacterium diphtheriae* pili. These newly discovered pili are formed by the covalent polymerization of pilin subunits catalyzed by sortase enzymes, making them fundamentally different from the noncovalent pilin assemblies of Gram-negative bacteria. Here, the expression, crystallization and preliminary crystallographic analysis of SpaA, which forms the shaft of one of the three types of pili expressed by *C. diphtheriae*, are reported. $SpaA_{53–486}$ crystals diffracted to 1.6 Å resolution and belonged to space group $P2_12_12_1$, with unit-cell parameters a = 34.9, b = 64.1, c = 198.7 Å, $\alpha = \beta = \gamma = 90^\circ$.

1. Introduction

Pili are filamentous appendages that extend from the surface of bacteria and have been recognized as important virulence factors in numerous Gram-negative bacterial infections (Duguid *et al.*, 1955; Brinton, 1959; Pizarro-Cerda & Cossart, 2006). More recently, considerable research effort has gone into characterizing the pili of Gram-positive pathogens such as *Corynebacterium diphtheriae*, *Streptococcus agalactiae*, *S. pneumoniae*, *S. pyogenes* and *Bacillus cereus* (Ton-That & Schneewind, 2003; Lauer *et al.*, 2005; Barocchi *et al.*, 2006; Budzik *et al.*, 2007). Similar to their Gram-negative bacterial counterparts, they have been shown to be important in bacterial adhesion and invasion and the pilus subunit proteins, or pilins, have been shown to be promising vaccine candidates (Dramsi *et al.*, 2005; Lauer *et al.*, 2005; Barocchi *et al.*, 2006; Buccato *et al.*, 2006; Abbot *et al.*, 2007). One defining feature of Gram-positive bacterial pili is the covalent polymerization of pilins catalyzed by transpeptidases called sortases, which makes them distinct from the noncovalent pilus assembly seen for Gram-negative bacteria. The covalent polymerization of Gram-positive pili was first discovered for *C. diphtheriae* in 2003 by Ton-That and Schneewind; immunoblotting of *C. diphtheriae* cell-wall extracts using antibodies specific for pilus subunits revealed distinctive high-molecular-weight polymers, indicating the SDS-resistant covalent nature of the linkage between pilus subunits, and the expression of sortase A (SrtA) was found to be required for pilus polymerization (Ton-That & Schneewind, 2003).

The genome of *C. diphtheriae* strain NCTC13129, a toxigenic clinical isolate, produces SpaA, SpaD and SpaH pili and each type of pilus is encoded by a gene cluster containing sortase and pilin genes (Ton-That & Schneewind, 2003; Gaspar & Ton-That, 2006; Swierczynski & Ton-That, 2006). The shaft of the SpaA pilus is formed by the major pilin SpaA; the minor pilins SpaB and SpaC are associated with the shaft. These pili provided the model system through which the mechanisms of covalent pilus assembly were first established. Pilin subunits are covalently joined by sortase action, which catalyzes the formation of an amide linkage between the ε-amino group of a lysine residue and the C-terminal carboxylate of the next pilin subunit (Ton-That & Schneewind, 2003). The lysine residue is widely conserved among major pilins in a sequence motif described as the pilin motif. Indeed, mass-spectral analyses of *S. pyogenes* and

B. cereus pili proved the existence of this proposed linkage between their respective major pilins Spy0128 and BcpA (Kang *et al.*, 2007; Budzik *et al.*, 2008). In addition, a conserved glutamate residue in SpaA, in a sequence motif referred to as the E box, was shown to be critically involved in minor pilin incorporation, although the mechanism has yet to be determined (Ton-That *et al.*, 2004). Finally, the entire pilus is covalently attached to the cell wall *via* a covalent linkage between the basal pilin subunit SpaB and the cell-wall peptidoglycan (Mandlik *et al.*, 2008).

Here, we present the purification and crystallization of SpaA, the archetypal major pilin of *C. diphtheriae*, together with preliminary X-ray crystallographic data. The structure of Spy0128 of *S. pyogenes* has previously been determined and revealed two immunoglobulin-like (Ig-like) domains containing intramolecular amide (isopeptide) bonds between lysine and asparagine residues (Kang *et al.*, 2007). SpaA also contains well conserved lysine and asparagine residues that can form such internal crosslinks. However, its sequence is about 180 amino-acid residues longer than that of Spy0128 and only shares less than 20% sequence identity with Spy0128, suggesting a different domain architecture. Determination of the three-dimensional structure of SpaA will provide important insights into its structural features and the assembly mechanism, which could also be used to infer those of other Gram-positive bacterial pili.

2. Materials and methods

2.1. Cloning and protein expression

The DNA sequence encoding amino acids 53–486 of SpaA (gi:38234577) was amplified from *C. diphtheriae* strain NCTC13129 genomic DNA and cloned using Gateway cloning methods (Moreland *et al.*, 2005). The following primers containing gene-specific sequences (bold) were used for the first round of PCR: forward, 5′-TTCCAAGGTCCG**GAACGTACGAGTATTGCCGTG**-3′, and reverse, 5′-GAAAGCTGGGTGCTA**GTTCTTTTTCTTGTTGTC-GATCTTTAC**-3′. A second round of PCR was performed with the following generic primers containing *attB* sequences for Gateway cloning and the sequence encoding an N-terminal human rhinoviral (HRV) 3C protease site: forward, 5′-GGGGACAAGTTTGTACA-AAAAAGCAGGCTCTCTCGAGGTACTCTTCCAAGGTCCG-3′, and reverse, 5′-GGGGACCACTTTGTACAAGAAAGCTGGGTG-3′. The final PCR product was cloned into the entry vector pDONR221 (Invitrogen) *via* a BP reaction and subsequently into the expression vector pDEST17 (Invitrogen) for expression of a His-tagged construct *via* an LR reaction. The resulting construct was transformed into *Escherichia coli* BL21 (DE3) cells and protein overexpression was carried out in Luria–Bertani (LB) medium supplemented with ampicillin (100 µg ml^{-1}) and chloramphenicol (34 µg ml^{-1}). Each litre of culture was inoculated with 1 ml overnight culture and grown at 310 K for 4 h. Protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 m*M* and then proceeded at 291 K for 24 h.

2.2. Protein purification

Cells were harvested by centrifugation (4000*g* at 277 K for 15 min) and resuspended in lysis buffer (50 m*M* Tris–HCl pH 8.0, 300 m*M* NaCl, 20 m*M* imidazole) supplemented with Mini EDTA-free protease inhibitor (Roche). Cells were lysed with a cell disruptor (Constant Systems) and the supernatant containing soluble proteins was separated from the cell debris by two centrifugation steps at 277 K (5000*g* for 15 min followed by 13 000*g* for 15 min).

The supernatant was loaded onto a 5 ml HiTrap column (GE Healthcare) charged with Ni^{2+} and equilibrated with lysis buffer. After washing the loaded column with lysis buffer, His-tagged SpaA was eluted by a gradient of 20–500 m*M* imidazole. The His tag was removed by incubating the eluted protein with HRV 3C (1 mg per ∼50 mg SpaA), 10 m*M* DTT and 2 m*M* EDTA. The 3C cleavage reaction was carried out by dialyzing against 50 m*M* Tris–HCl pH 8.0, 300 m*M* NaCl overnight at 277 K. Untagged SpaA was separated by passing the dialyzed sample through a charged and equilibrated HiTrap column. SpaA was further purified by size-exclusion chromatography using a Superdex75 10/300 column with 10 m*M* Tris–HCl pH 8.0, 50 m*M* NaCl. Fractions containing SpaA were concentrated to 100 mg ml^{-1} as measured by the absorbance at 280 nm (extinction coefficient 39 545 M^{-1} cm^{-1} at 280 nm, molecular weight 46.8 kDa). The final product contained two additional N-terminal residues, Gly

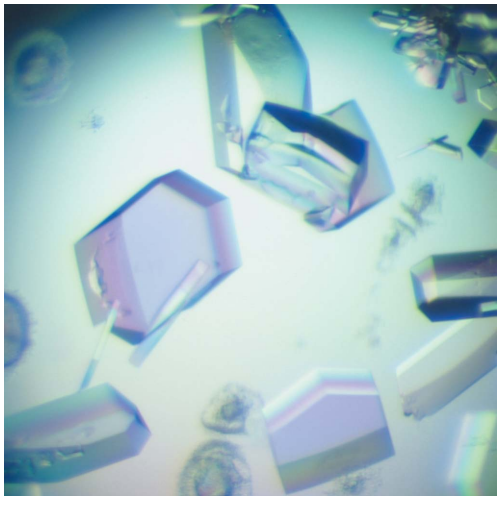


Figure 1
Crystals of SpaA grown from 20% PEG 3350, 0.1 *M* NaI, 0.1 *M* NaF in a 2 µl drop. Crystals grew to approximately 0.5 × 0.3 × 0.1 mm.

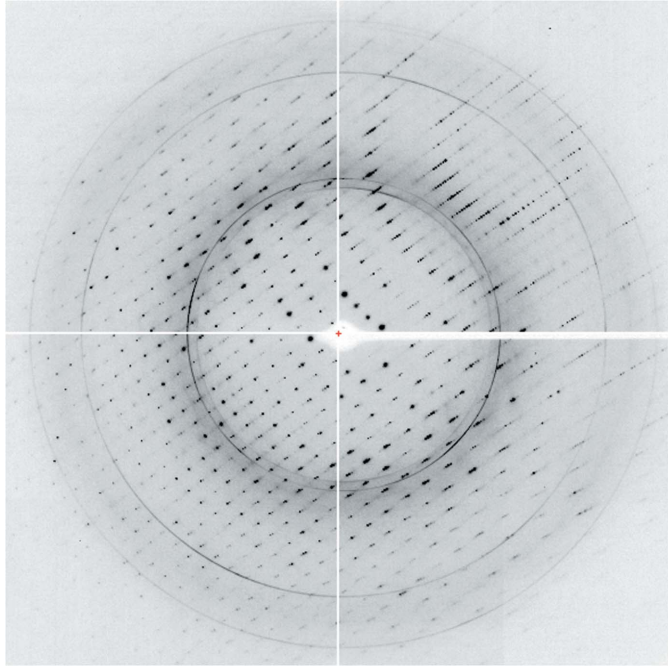


Figure 2
X-ray diffraction pattern of SpaA.

Table 1
Data-collection statistics.

Values in parentheses are for the outermost shell.

Wavelength (Å)	0.77337
Temperature (K)	100
Space group	$P2_12_12_1$
Unit-cell parameters (Å, °)	a = 34.9, b = 64.1, c = 198.7, $\alpha = \beta = \gamma$ = 90.0
Resolution (Å)	40.00–1.60 (1.69–1.60)
Total no. of reflections	370705 (26745)
No. of unique reflections	58478 (7347)
Redundancy	6.3 (6.3)
Completeness (%)	97.3 (85.6)
Average $I/\sigma(I)$	10.0 (1.7)
$R_{\rm merge}$ (%)†	0.095 (0.500)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith measurement and $\langle I(hkl)\rangle$ is the weighted mean of all measurements of $I(hkl)$.

and Pro, after His-tag removal. The typical yield of purified SpaA was 15 mg per litre of culture.

2.3. Crystallization and data collection

The purified proteins were initially screened using 480 different crystallization conditions. Crystallization trials were performed by vapour diffusion at 291 K using sitting drops in 96-well Intelli-Plates (Hampton Research) set up with a Cartesian Honeybee dispensing system (Genomic Solutions). The best diffracting crystals were obtained using a protein concentration of 50 mg ml^{-1} from a mother liquor containing 20%(*v*/*v*) PEG 3350, 0.1 *M* NaI, 0.1 *M* NaF. Crystals took 3–4 d to grow into thick plates in 2 µl drops containing an 1:1 mixture of protein solution and mother liquor (Fig. 1). Diffraction quality was variable, but by screening a large number of crystals it was possible to find several crystals that diffracted isotropically to better than 2.0 Å resolution.

For data collection, SpaA crystals were flash-cooled immediately in liquid nitrogen without any additional cryoprotectant. Data were collected to 1.6 Å resolution (Fig. 2) on beamline PX1 at the Australian Synchrotron (Melbourne, Australia). A total of 360 images were collected with a φ oscillation of 0.5 ° and an exposure time of 1.5 s per image. The crystal-to-detector distance was 230 mm. The data were processed and scaled with *MOSFLM* (Leslie, 1992) and *SCALA* (Collaborative Computational Project, Number 4, 1994).

3. Results

The recombinant SpaA of *C. diphtheriae* contains amino-acid residues 53–486, excluding the signal peptide (residues 1–52) and the C-terminal residues 487–525 which contain the sortase-recognition motif. The protein was highly soluble and crystallized in a number of conditions at concentrations between 50 and 100 mg ml^{-1}. The best crystals diffracted to 1.6 Å resolution at the Australian Synchrotron (Fig. 2). The crystals belonged to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters a = 34.9, b = 64.1, c = 198.7 Å, $\alpha = \beta = \gamma = 90°$. Calculation of the Matthews coefficient ($V_{\rm M}$ = 2.34 Å^3 Da^{-1}) indicated a solvent content of 48%, assuming the presence of one molecule in the asymmetric unit. The diffraction data statistics are given in Table 1.

Initial attempts to obtain phases from the iodide present in the crystallization buffer were unsuccessful. Heavy metal-atom derivatization and production of selenomethionine-substituted SpaA are currently under way.

We would like to thank Dr Ton-That Hung (University of Texas Medical School, Houston, USA) for providing the genomic DNA of *C. diphtheriae* strain NCTC13129. This work was supported by the Maurice Wilkins Centre for Molecular Biodiscovery and a research grant from the Marsden fund to ENB. Use of the synchrotron was supported by the Australian Synchrotron.

References

Abbot, E. L., Smith, W. D., Siou, G. P., Chiriboga, C., Smith, R. J., Wilson, J. A., Hirst, B. H. & Kehoe, M. A. (2007). *Cell. Microbiol.* **9**, 1822–1833.
Barocchi, M. A. *et al.* (2006). *Proc. Natl Acad. Sci. USA*, **103**, 2857–2862.
Brinton, C. C. Jr (1959). *Nature (London)*, **183**, 782–786.
Buccato, S., Maione, D., Rinaudo, C. D., Volpini, G., Taddei, A. R., Rosini, R., Telford, J. L., Grandi, G. & Margarit, I. (2006). *J. Infect. Dis.* **194**, 331–340.
Budzik, J. M., Marraffini, L. A. & Schneewind, O. (2007). *Mol. Microbiol.* **66**, 495–510.
Budzik, J. M., Marraffini, L. A., Souda, P., Whitelegge, J. P., Faull, K. F. & Schneewind, O. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 10215–10220.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dramsi, S., Trieu-Cuot, P. & Bierne, H. (2005). *Res. Microbiol.* **156**, 289–297.
Duguid, J. P., Smith, I. W., Dempster, G. & Edmunds, P. N. (1955). *J. Pathol. Bacteriol.* **70**, 335–348.
Gaspar, A. H. & Ton-That, H. (2006). *J. Bacteriol.* **188**, 1526–1533.
Kang, H. J., Coulibaly, F., Clow, F., Proft, T. & Baker, E. N. (2007). *Science*, **318**, 1625–1628.
Lauer, P., Rinaudo, C. D., Soriani, M., Margarit, I., Maione, D., Rosini, R., Taddei, A. R., Mora, M., Rappuoli, R., Grandi, G. & Telford, J. L. (2005). *Science*, **309**, 105.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Mandlik, A., Das, A. & Ton-That, H. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 14147–14152.
Moreland, N., Ashton, R., Baker, H. M., Ivanovic, I., Patterson, S., Arcus, V. L., Baker, E. N. & Lott, J. S. (2005). *Acta Cryst.* D**61**, 1378–1385.
Pizarro-Cerda, J. & Cossart, P. (2006). *Cell*, **124**, 715–727.
Swierczynski, A. & Ton-That, H. (2006). *J. Bacteriol.* **188**, 6318–6325.
Ton-That, H., Marraffini, L. A. & Schneewind, O. (2004). *Mol. Microbiol.* **53**, 251–261.
Ton-That, H. & Schneewind, O. (2003). *Mol. Microbiol.* **50**, 1429–1438.