Structure of Spo0M, a sporulation-control protein from *Bacillus subtilis*

Yo Sonoda,* Kimihiko Mizutani and Bunzo Mikami*

Laboratory of Applied Structural Biology, Division of Applied Life Sciences, Graduate School of Agriculture, Kyoto University, Uji, Kyoto 611-0011, Japan. *Correspondence e-mail: sonoda.yo.68v@st.kyoto-u.ac.jp, mikami@kais.kyoto-u.ac.jp

Received 22 July 2015

Accepted 4 November 2015

Edited by A. Nakagawa, Osaka University, Japan

Keywords: sporulation-control protein; Spo0M; arrestin; FP domain; PI31.

PDB reference: Spo0M, 5cl2

Supporting information: this article has supporting information at journals.iucr.org/f

Spo0M is a sporulation-control protein that is thought to play an essential role in the early stage of endospore formation. While little is known about the functions of Spo0M, a recent phylogenetic study suggests that, based on its amino-acid sequence, Spo0M might belong to the arrestin clan. The crystal structure of the Spo0M protein was determined at a resolution of 2.3 Å. Ten amino acids at the end of the N-terminus were removed to improve the thermal stability of the purified Spo0M protein and the crystal structure of Spo0M was determined by SAD. Spo0M has a well conserved N-terminal domain with an arrestin-like fold, which consists of a β -strand sandwich structure. Surprisingly, the C-terminal domain of Spo0M, which has no structural homology to arrestin-clan proteins, bears significant structural similarity to the FP domain of the human PI31 protein. In addition, Spo0M harbours a potential polar-core structure connecting the N- and C-terminal domains with several salt bridges, as seen in the crystal structures of arrestin and VPS26. The structure reported here constitutes the first structural information on a bacterial protein that shares significant structural homology to members of the arrestin clan and the FP domain.

1. Introduction

Responses to changes in the external environment are an essential function in the survival of all organisms, and a wide range of strategies are adopted for this purpose. In the bacterial kingdom, cellular responses to environmental changes include stress responses to heat, organic solvents or high salt concentration and emergency responses to the depletion of particular nutrient sources. The bacterial species *Bacillus subtilis* forms endospores to survive once it senses nutrient limitation. Endospores are metabolically inactive and are extremely resistant to heat, toxic chemicals and even radiation (Nicholson *et al.*, 2000). Since *B. subtilis* lends itself to genetic manipulation and thus is a commonly used bacterial model organism, the sporulation mechanism of *B. subtilis* has been well studied (Piggot & Hilbert, 2004).

Once a *B. subtilis* cell senses a sporulation-initiation signal, it stops its normal cell division (stage 0) and forms a septum (stage II), a new cell wall between the dividing daughter cells that is observed only in the endospore-formation phase (Stragier & Losick, 1996). At the initiation of sporulation, the Spo0A protein is activated through phosphorylation (Sonenshein, 2000). Phosphorylated Spo0A (Spo0A-P) is an active transcriptional factor that regulates a group of genes essential for subsequent endospore formation. Spo0A-P directly and

© 2015 International Union of Crystallography

indirectly regulates more than 500 genes (Fawcett *et al.*, 2000; Molle *et al.*, 2003). For example, the expression of various σ factors is upregulated by Spo0A-P. Some of the σ factors are individually distributed into the endospore or mother cell, resulting in distinct gene expression in each cell (Phillips & Strauch, 2002; Piggot & Hilbert, 2004). Spo0A-P also represses *arbB* gene expression (Strauch *et al.*, 1990). The ArbB protein acts as a repressor of the *spo0H* gene, which encodes the σ^H protein (Weir *et al.*, 1991). σ^H induces expression of the σ^F protein, which in turn is an essential factor for stage II of sporulation (Wu *et al.*, 1992). σ^H also plays an important role in transcription of the *spo0A* gene (Siranosian & Grossman, 1994).

Spo0M is one of the sporulation-controlling proteins; it is made up of 258 amino acids and is transcriptionally regulated by σ^H . The *spo0M* gene is not essential for cell viability, but disruption of the *spo0M* gene results in sensitivity to lysis and a defect in endospore formation from stage 0 to stage II (Han *et al.*, 1998). Overexpression of the Spo0M protein also negatively affects the sporulation process, indicating that both too little and too much Spo0M protein in the cells will influence the initial step of sporulation (Han *et al.*, 1998).

The FtsH protein of *B. subtilis* is an ATP- and Zn^{2+} -dependent membrane metalloprotease that degrades both membrane and cytoplasmic proteins. The *ftsH* null-mutant strain has a heat-sensitive pleiotropic phenotype that includes sporulation blockage between stage 0 and stage II (Lysenko *et al.*, 1997; Deuerling *et al.*, 1997; Wehrl *et al.*, 2000). Proteome analysis of the *ftsH* null strain revealed that disruption of the *ftsH* gene induces a fourfold higher expression of Spo0M, indicating that Spo0M is one of the substrates of FtsH (Thi Nguyen & Schumann, 2012). FtsH may control the amount of Spo0M protein to maintain a level appropriate for sporulation for the reasons described above.

Although the Spo0M protein is thought to play an important role in the initiation of sporulation, only little is known about the functions of this protein. A recent phylogenetic study revealed that the N-terminal half of the Spo0M protein showed significant homology to the consensus sequence of eukaryotic arrestins, which indicates that Spo0M might be a member of the arrestin clan (Alvarez, 2008). The arrestin clan consists of β -arrestins and visual arrestins, α -arrestins and vacuolar protein sorting-associated protein 26 (VPS26) (Alvarez, 2008; Kang *et al.*, 2014; Patwari & Lee, 2012). The β -arrestins and visual arrestins (β/v -arrestins) are the best characterized members and are known to be multifunctional scaffolding proteins that regulate G protein-coupled receptor (GPCR)-dependent signalling and membrane trafficking (Tian *et al.*, 2014). The β/v -arrestins have structurally well conserved two-lobe, ω -shaped, immunoglobulin-like, β -strand sandwich structures (Hirsch *et al.*, 1999; Han *et al.*, 2001; Sutton *et al.*, 2005). In addition, the crystal structures of other arrestin-clan members, such as VPS26 (Shi *et al.*, 2006), thioredoxin-interacting protein (TXNIP; Polekhina *et al.*, 2013; Hwang *et al.*, 2014) and the N-terminal domain of arrestin domain-containing 3 (ARRDC3; Qi *et al.*, 2014), which are distantly related to β/v -arrestins on the basis of their

amino-acid sequences ($\sim 20\%$ identity), have a β -strand sandwich domain that resembles β/v -arrestins. These structural studies indicate that the three-dimensional structures of mammalian members of the arrestin clan share highly and structurally conserved N- and C-terminal domains in spite of their poor amino-acid homologies. Although Spo0M is thought to be a bacterial member of the arrestin clan, no structural information on any bacterial member of the arrestin clan has been reported to date. Furthermore, in contrast to the N-terminal half, there has been no reliable information or even speculation as to the structure and function of the C-terminal half of the Spo0M protein.

In the present study, we determined the crystal structure of the truncated form Spo0M^{11–258}, lacking the ten amino acids at the N-terminus, at a resolution of 2.3 Å. On the basis of this structure, we compared the structures of the N-terminal half of Spo0M and the other proteins belonging to the arrestin clan. In addition, using the C-terminal half of the Spo0M structure as a query, we identified potential homologues of the C-terminal half of the Spo0M protein.

2. Materials and methods

2.1. Protein expression and purification

The DNA fragments encoding the full-length (258 amino acids; Spo0M^{FL}) and truncated (Spo0M^{1–255}, Spo0M^{1–245}, Spo0M^{1–235}, Spo0M^{6–258}, Spo0M^{11–258}, Spo0M^{16–258} and Spo0M^{21–258}) versions of the *spo0M* gene were amplified by PCR and cloned into pET-15b vector using the NdeI and BamHI restriction sites, in which an N-terminal His tag is appended to the coding sequence. *Escherichia coli* (DE3) CodonPlus cells harbouring the recombinant plasmid were grown in LB medium at 37°C to an optical density of 0.6 at 600 nm. 0.1 mM isopropyl β -D-1-thiogalactopyranoside was added prior to further culture at 25°C for 16 h. The cells were harvested and resuspended in buffer A (20 mM Tris–HCl pH 7.5, 50 mM NaCl, 1 mM EDTA, 0.5 mM DTT, 25 mM imidazole). DNase and 3 mM MgCl₂ were added to the suspension. The cells were passed through an Avestin EmulsiFlex-C3 cell disruptor at a pressure of 103 MPa. The cell lysate was loaded onto a HisTrap FF crude column (GE Healthcare) equilibrated with buffer A. The columns were washed with ten column volumes of buffer A. The proteins were eluted using a linear gradient between buffer A and buffer B (20 mM Tris–HCl pH 7.5, 50 mM NaCl, 1 mM EDTA, 0.5 mM DTT, 500 mM imidazole). The fractions containing Spo0M proteins were pooled and dialyzed overnight at 4°C in the presence of thrombin protease against buffer C (20 mM Tris–HCl pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM DTT). The dialyzed sample was applied onto a Q Sepharose column (GE Healthcare) equilibrated with buffer C and eluted using a linear gradient between buffer C and buffer D (20 mM Tris–HCl pH 7.5, 1 M NaCl, 1 mM EDTA, 1 mM DTT). The fractions containing Spo0M proteins were pooled and were concentrated to ~ 10 mg ml^{−1} by ultrafiltration using a Vivaspin 6 concentrator (Sartorius). Finally, the concentrated samples were purified

by gel filtration using a Superdex 200 10/30 column (GE Healthcare) in buffer *E* (20 mM Tris–HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM DTT) at a flow rate of 0.4 ml min^{−1}. The fractions containing Spo0M proteins were confirmed by SDS–PAGE, pooled and concentrated to 15–20 mg ml^{−1} by ultracentrifugation using a Vivaspin 2. The molecular mass of purified Spo0M^{FL} was estimated by gel filtration using a Superdex 75 10/30 column (GE Healthcare) with a calibration curve obtained using bovine serum albumin (66 kDa), chicken ovalbumin (44 kDa), horse myoglobin (17 kDa) and vitamin B₁₂ (1.35 kDa) as molecular-mass standards.

2.2. Crystallization

An attempt was made to crystallize the purified Spo0M proteins using the sitting-drop vapour-diffusion method. An initial crystallization screening was performed using Crystal Screen and Crystal Screen 2 (Hampton Research) by mixing 0.4 µl protein solution with 0.4 µl reservoir solution. The droplets were equilibrated against 75 µl reservoir solution at 4 and 20°C. To obtain diffraction-quality crystals, the crystallization conditions were further refined. The optimal crystallization conditions for Spo0M^{FL} were defined as 0.1 M MES–NaOH pH 6.5, 0.9–1.0 M Li₂SO₄ at 20°C. The crystallization drops were prepared by mixing 2 µl protein solution with 2 µl buffer. Crystals of Spo0M^{FL} appeared within one week. The crystals were transferred into a solution containing 25% glycerol as a cryoprotectant prior to flash-cooling in a cold nitrogen-gas stream (100 K). The optimal crystallization conditions for the N-terminally ten-amino-acid truncated version, Spo0M^{11–258}, were defined as 0.1 M Tris–HCl pH 7.0, 1.0–1.1 M sodium citrate at 4°C. The crystallization drops were prepared by mixing 2 µl protein solution with 2 µl buffer solution. The crystals of Spo0M^{11–258} appeared within 2 d. The crystals were transferred into a solution containing 15% ethylene glycol as a cryoprotectant prior to flash-cooling in a cold nitrogen-gas stream (100 K).

2.3. Heavy-atom derivatization

To screen heavy atoms, the Spo0M^{11–258} crystals were soaked in a heavy-atom solution and X-ray diffraction data were then collected using an in-house Bruker Hi-Star detector coupled with a Bruker M18XHF rotating-anode X-ray generator. K₂PtCl₄ was selected for heavy-atom derivatization. A platinum-derivatized Spo0M^{11–258} crystal was obtained by soaking crystals in a solution consisting of 0.1 M Tris–HCl pH 7.0, 1.1 M sodium citrate, 10 mM K₂PtCl₄ for 15 min. The crystals were then transferred into a solution containing 15% glycerol as a cryoprotectant prior to flash-cooling in a cold nitrogen-gas stream (100 K).

2.4. Data collection and structure determination

All X-ray diffraction data were collected at SPring-8, Hyogo, Japan after an in-house diffraction check. Diffraction data for Spo0M^{FL} were collected on the BL44XU beamline. The crystal of Spo0M^{FL} was mounted on a nylon loop (Hampton Research) to collect diffraction data using a MAR

Table 1

Data-collection and refinement statistics.

	Spo0M ^{11–258} , native (PDB entry 5cl2)	Spo0M ^{11–258} , K ₂ PtCl ₄ derivative
Data collection		
Source	BL44XU, SPring-8	BL26B1, SPring-8
Wavelength (Å)	0.9000	1.0700
Detector	MAR 300HE CCD	Saturn A200 CCD
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Unit-cell parameters (Å)	<i>a</i> = 66.00, <i>b</i> = 74.23, <i>c</i> = 133.08	<i>a</i> = 64.70, <i>b</i> = 79.06, <i>c</i> = 131.33
Resolution limits (Å)	50–2.3 (2.34–2.30)	50–2.8 (2.85–2.80)
Total reflections	588370	405995
Unique reflections	29843 (1428)	17097 (833)
Completeness (%)	99.6 (96.8)	100 (100)
Multiplicity	7.6 (7.1)	8.7 (8.6)
<i>R</i> _{merge}	0.056 (0.510)	0.069 (0.505)
Mean <i>I</i> /σ(<i>I</i>)	48.99 (4.90)	52.09 (7.80)
Structure determination		
Figure of merit		
Prior to density modification		0.25
After density modification		0.77
No. of Pt sites per asymmetric unit		7
Refinement		
Program used	<i>phenix.refine</i>	
Resolution range (Å)	40–2.3 (2.38–2.30)	
No. of reflections used	29788 (2837)	
Completeness (%)	99.5 (96.2)	
Protein molecules per asymmetric unit	2	
Contents of asymmetric unit		
Residues	491	
Waters	43	
Average <i>B</i> factor (Å ²)		
Protein	62.9	
Waters	54.8	
Bond-length r.m.s.d. (Å)	0.01	
Bond-angle r.m.s.d. (°)	1.39	
<i>R</i> _{work}	0.23 (0.31)	
<i>R</i> _{free}	0.27 (0.38)	

CCD scanner. The distance between the crystal and detector was set to 200 mm, with an oscillation range of 1°. The data were processed using *HKL-2000* (Otwinowski & Minor, 1997), with a final *R*_{merge} of 6.0% in the resolution range 50–3.2 Å. The space group was estimated to be *P*1. Although we tried to solve the structure of Spo0M^{FL} by the molecular-replacement method using published structures of arrestin-clan proteins as search models, no reasonable solution was obtained, which is not surprising owing to the extremely poor sequence homology. In addition, the crystal quality of Spo0M^{FL} was not reproducible, which made optimization difficult. To obtain crystals of better quality, truncation mutants were screened and Spo0M^{11–258} was obtained as a stable mutant (see below).

Diffraction data for Spo0M^{11–258} were also collected on the BL44XU beamline with the settings described above. The data were processed with a final *R*_{merge} of 5.6% in the resolution range 50–2.3 Å. The space group was estimated to be *P*2₁2₁2₁. As seen for Spo0M^{FL}, the phase could not be determined by the molecular-replacement method. To solve the phase problem, a Pt-derivatized Spo0M^{11–258} crystal was prepared for single-wavelength anomalous diffraction (SAD) phasing. The wavelength was set at 1.07 Å from X-ray absorption

Table 2

Summary of the results obtained from the thermal shift assay.

	T_m (°C)	No. of experiments
Spo0M ^{FL}	41.7 ± 0.76	3
Spo0M ^{6–258}	46.3 ± 0.42	3
Spo0M ^{11–258}	50.2 ± 0.56	3
Spo0M ^{16–258}	46.7 ± 1.44	3
Spo0M ^{1–255}	32.7	1

fine-structure (XAFS) measurements. The diffraction data were processed with a final R_{merge} of 6.9% in the resolution range 50–2.8 Å. The data-collection statistics for native and Pt-derivatized Spo0M^{11–258} are summarized in Table 1.

The crystal structure of the Pt derivative of Spo0M^{11–258} was determined by the SAD method. We identified two Spo0M molecules and seven platinum-binding sites in the asymmetric unit. The initial phase was determined by *AutoSol* and the initial model was obtained using *AutoBuild* in the *PHENIX* suite (Adams *et al.*, 2010). The crystal structure of native Spo0M^{11–258} was solved by molecular replacement with *Phaser* from the *PHENIX* suite using the coordinates of the Pt-derivative structure as a starting model. Further model building and refinement were performed using *Coot* (Emsley *et al.*, 2010) and *PHENIX*, resulting in a 2.3 Å resolution structure of Spo0M^{11–258}. Noncrystallographic symmetry (NCS) restraints were used in all stages of refinement. Structure analysis was performed using *MolProbity* in the *PHENIX* suite (Chen *et al.*, 2010). Molecular-graphics images were prepared using *PyMOL* (Schrödinger; <http://www.pymol.org>). Protein interfaces in the structures were analyzed using *PDBEPIA* (Krissinel & Henrick, 2007).

2.5. Thermal shift assay

The Spo0M proteins at 25 µM in buffer *C* containing Protein Thermal Shift Dye (Life Technologies) were incubated in the wells of RT-PCR devices (Life Technologies, StepOnePlus). The samples were heated from 25 to 90°C. Fluorescence intensities were plotted as a function of temperature, and accurate fitting to the Boltzmann equation by nonlinear regression was performed using *GraphPad Prism*. The inflection point of the transition curve, T_m , was calculated using the Boltzmann equation $y = LL + (UL - LL) / \{1 + \exp[(T_m - x)/\text{Slope}]\}$, where LL and UL are the values of the minimum and maximum intensities, respectively.

3. Results and discussion

3.1. Expression and purification

All Spo0M proteins (Spo0M^{FL}, Spo0M^{1–255}, Spo0M^{1–245}, Spo0M^{1–235}, Spo0M^{6–258}, Spo0M^{11–258}, Spo0M^{16–258} and Spo0M^{21–258}) were overexpressed as N-terminally His₆-tagged proteins in *E. coli*. The expression levels of these proteins were almost the same. However, the Spo0M^{1–245}, Spo0M^{1–235} and Spo0M^{21–258} proteins were not obtained in soluble fractions of the cell lysate, indicating that these mutant proteins were misfolded and aggregated. The proteins that were

successfully overexpressed in soluble fractions, Spo0M^{FL}, Spo0M^{1–255}, Spo0M^{6–258}, Spo0M^{11–258} and Spo0M^{16–258}, were purified to homogeneity by Ni-NTA, anion-exchange and gel-filtration column chromatography, resulting in proteins of greater than 95% homogeneity as confirmed by SDS-PAGE. All proteins eluted at similar retention times using gel-filtration column chromatography. As shown in Fig. 1, after comparison with the gel-filtration markers, the molecular mass of the purified Spo0M^{FL} protein in solution was estimated to be approximately 51 kDa.

3.2. Thermal stability of Spo0M truncation mutants

Although we first attempted to determine the structure by using crystals of the Spo0M^{FL} protein, only crystals that diffracted poorly, *i.e.* up to 3.2 Å resolution, were obtained and the phase could not be determined by the molecular-replacement method. In addition, the poor reproducibility of the crystal quality led to difficulties in optimizing the crystallization conditions as well as heavy-atom derivatization. Since the thermal stability of a purified protein can be a good indicator of the crystal quality (Ericsson *et al.*, 2006; Sonoda *et al.*, 2011), we tried to obtain more thermally stable proteins by using N/C-terminus truncations, and the thermal stabilities of each truncation were evaluated using the differential scanning fluorimetry (DSF) technique. The melting temperatures (T_m) of each protein are shown in Table 2. The T_m of the full-length Spo0M was 41.7°C. In terms of C-terminus truncations, the truncation of five or more amino acids from the end completely abolished stable expression of the Spo0M protein in solution. Although an Spo0M^{1–255} protein that lacks only three amino acids at the C-terminal end was overexpressed in the soluble fraction and successfully purified, the T_m of the purified Spo0M^{1–255} protein decreased to 32.7°C, which is 9°C lower than that of Spo0M^{FL}. The N-terminus truncation

Figure 1

Gel-filtration chromatography of Spo0M^{FL} and molecular-mass standards. The elution profile of the Spo0M^{FL} protein from a Superdex 75 10/30 gel-filtration column is indicated. The elution volumes of the protein molecular-mass standards, bovine serum albumin (BSA; 66 kDa), chicken ovalbumin (OVA; 44 kDa), horse myoglobin (17 kDa) and vitamin B₁₂ (VB₁₂; 1.35 kDa), are indicated at the bottom.

mutants Spo0M^{6–258}, Spo0M^{11–258} and Spo0M^{16–258}, but not Spo0M^{21–258}, were successfully overexpressed and purified. In contrast to the C-terminus truncations, which resulted in unstable proteins, the N-terminus truncation mutants exhibited greater stability than the Spo0M^{FL} protein, with a 5–10°C

higher T_m , indicating that the N-terminus of Spo0M was flexible and negatively influences the thermal stability after purification. Among these mutants, the crystals of the Spo0M^{11–258} protein, which showed the highest T_m among the mutants tested, diffracted to significantly higher resolution

Figure 2
Overall structure of the *B. subtilis* Spo0M^{11–258} protein. (a) Cartoon representation of *B. subtilis* Spo0M^{11–258}. (b) Topology diagram of the Spo0M^{11–258} protein. (c) Two distinct dimeric interactions in the Spo0M^{11–258} crystal. Dimers formed through the N-interface and C-interface are shown in the left and right panels, respectively.

than the crystals of Spo0M^{FL}, indicating that screening for a thermal stable mutant is a reliable strategy for obtaining good diffraction-quality crystals, as described elsewhere.

3.3. Overall structure

The structure of *B. subtilis* Spo0M^{11–258} was determined by X-ray crystallography using the SAD technique with Pt-soaked crystals. The structure of Spo0M^{11–258} in the *P*₂₁₂₁₂₁ crystal form was refined to 2.3 Å resolution. The Ramachandran plot of the final model shows 99.4, 0.6 and 0% of the residues in the preferred, allowed and disallowed regions, respectively. Two molecules, molecules *A* and *B*, were found in the asymmetric unit. In the final model, the residues of four amino acids derived from the expression vector and two amino acids of the N-terminus of Spo0M^{11–258} in both molecules and a glutamate residue at the C-terminal end of molecule *B* could

not be assigned. One Na⁺ ion was added to form bonds to the side chains of Asp128 and Asp130 and the main chain of Thr20 in each molecule. The two monomers in an asymmetric unit are essentially identical, with a root-mean-square deviation (r.m.s.d.) of 0.289 Å for the corresponding main-chain atoms.

The overall structure of Spo0M is shown in Fig. 2. The Spo0M^{11–258} molecule consists of two domains and contains 14 β-strands and six α-helices. The N-terminal domain (N domain) contains residues 13–134, which form a β-sandwich structure composed of two antiparallel β-sheets of three (β1a, β2 and β5) and four (β3, β4a, β6 and β7) β-strands. In contrast, the C-terminal domain (C domain) contains residues 138–258, which adopt an α/β fold involving an antiparallel β-sheet of five β-strands (β8–β12) and six α-helices (α1–α6) surrounding the β-sheet. There is a short loop consisting of residues 135–137 which connects the N and C domains (Figs. 2*a* and 2*b*).

Figure 3
Comparison of the N domain of Spo0M^{11–258} with other arrestin-clan proteins. (a) Superposition of the N domain of Spo0M^{11–258} (green) with that of human TXNIP (PDB entry 4gej, magenta) and ARDDC3 (PDB entry 4r7x, blue). (b) Sequence alignment of the N domain of *B. subtilis* Spo0M with that of human TXNIP and ARDDC3. β-Sheet structure elements are indicated in red and by cartoons above the sequences. Conserved amino acids are indicated by asterisks below the sequences.

Two distinct dimeric arrangements of the Spo0M^{11–258} protein were observed in the crystal lattice, both of which were composed of molecules *A* and *B* in the asymmetric unit (Fig. 2*c*). In these dimers, two protomers interact with each other through substantial intermolecular interactions. In the dimeric arrangements shown on the left and right in Fig. 2(*c*), the intermolecular interface is formed between the two N domains (referred to as the N-interface) and between the two C domains (referred to as the C-interface), respectively. There are four and six potential hydrogen bonds in the N- and C-interfaces, respectively. Both the N-interface and the C-interface possess sufficiently large buried surface areas to be the potential dimer interface in solution: approximately 1000 and 800 Å² (8.1 and 6.4% of the whole surface area), respectively. The Δ^iG *P*-values (Krissinel & Henrick, 2004) of the N- and C-interfaces were 0.587 and 0.671, respectively, which indicates that both of the interfaces might be artifacts of crystal packing. As a positive correlation between the radius of gyration (R_g) and the molecular size of a protein has been reported (He & Niemeyer, 2003; Lobanov *et al.*, 2008), the R_g values of the proteins used for the estimation of molecular size in the gel-filtration study were calculated using *HYDROPRO* (Ortega *et al.*, 2011). The R_g values for ovalbumin (PDB entry 1ova; Stein *et al.*, 1991), the Spo0M^{11–258} monomer, the

Spo0M^{11–258} dimer with an N-interface, the Spo0M^{11–258} dimer with a C-interface and BSA (PDB entry 4f5s; Bujacz, 2012) were calculated as 22.5, 22.6, 26.6, 26.1 and 27.6 Å, respectively. The R_g of ovalbumin is equivalent to that of the Spo0M monomer, indicating that the estimation of molecular size of Spo0M by gel filtration described above might be inaccurate. Thus, we could not come to a conclusion about the oligomeric state of Spo0M in solution in this study.

3.4. Structural homology of the N domain of Spo0M to proteins with the arrestin-like fold

The N domain of Spo0M is composed of a seven-stranded β -sandwich (Fig. 2*a*), which closely resembles that of the other arrestin-clan members, especially those belonging to the α -arrestins. Human α -arrestins are composed of ARRDC1–5 and TXNIP. Among the members of the α -arrestins, only the crystal structures of the N domain of ARRDC3 and TXNIP and the overall structure of TXNIP have been determined to date (Polekhina *et al.*, 2013; Hwang *et al.*, 2014; Qi *et al.*, 2014). Structural and sequence alignments of the N domains of Spo0M, TXNIP and ARRDC3 are shown in Fig. 3. In spite of their extremely poor sequence similarities, the N domain of Spo0M^{11–258} superimposed onto those of TXNIP (PDB entry

Figure 4

Comparison of the C domain of Spo0M^{11–258} with the FP domain of PI31. (*a*) Superposition of the N domain of Spo0M^{11–258} (green) with the FP domain of human PI31 (PDB entry 2vt8, orange). In the PI31 structure, the first 19 amino acids at the N-terminus containing an α -helix were removed to clarify the view. (*b*) The C-interface of Spo0M^{11–258} (left) and the β -interface of PI31 (right) forming a dimer in the crystal lattice.

4gej; Polekhina *et al.*, 2013) and ARRDC3 (PDB entry 4r7x; Qi *et al.*, 2014) with r.m.s.d.s of 2.09 and 2.04 Å for 108 and 109 C α atoms, respectively (Figs. 3a and 3b). In contrast to the β /v-arrestins, the α -arrestins are thought to share conserved amino-acid sequences, being found in many eukaryotic genomes (Alvarez, 2008). The structure of the Spo0M N domain revealed in this study indicates that the α -arrestin-like β -sandwich fold is structurally conserved from bacteria to mammals. One of the best studied functions of eukaryotic arrestin is the regulation of endocytosis of membrane proteins such as GPCRs (Tian *et al.*, 2014). Although there is no endocytosis event on bacterial cell membranes, Spo0M might play a role in membrane trafficking to initiate sporulation that is triggered by environmental change. Eukaryotic arrestins are also known to act as scaffolding proteins facilitating signal transduction and to interact with multiple binding partners. For instance, a number of proteins related to signal transduction, which include c-Src, MAPK, Mdm2, ASK1 and JNK3, have been reported to bind to β -arrestin (Lefkowitz & Shenoy, 2005). In *B. subtilis* cells, a study using interactome analysis

has reported that the Spo0M protein might directly or indirectly bind to catabolite-control protein A (CcpA; Wünsche *et al.*, 2012), the master regulator of carbon metabolism and catabolite repression, indicating that the Spo0M protein might play a role in the regulation of gene expression related to catabolite metabolism during the initiation step of sporulation by interacting with the CcpA protein. In contrast to the N domain, which has significant structural similarity to other arrestin-clan proteins, no protein with the arrestin-like fold that bears significant structural homology to the C domain of Spo0M was identified by the DALI server (Holm & Rosenström, 2010) or the PDBeFold server (Krissinel & Henrick, 2004). The eukaryotic arrestin structures determined so far comprise two-lobe β -sandwich domains that are related by a pseudo-twofold rotation axis. The Spo0M structure reported here is the first example of an arrestin-clan protein that has only one half of the two-lobe domain, which is linked to a non-arrestin domain. Indeed, a potential protein with an arrestin-like fold that is assumed to have only half of the two-lobe domain like Spo0M has been predicted in the *Dictyostelium discoideum* genome (Aubry *et al.*, 2009).

3.5. Structural homology between the C domain of Spo0M and the FP domain of PI31

Surprisingly, the dimerization domain, called the FP domain, of the human proteasome inhibitor protein subunit PI31, also known as PSMF1, has been shown to bear significant structural homology to the C domain of Spo0M (Fig. 4a). The C domain of Spo0M^{11–258} superimposed onto the FP domain of PI31 (PDB entry 2vt8; Kirk *et al.*, 2008) with an r.m.s.d. of 2.09 Å for 124 C α atoms (Fig. 4a). PI31 mediates the maturation of immunoproteasomes, which are an essential proteasome subtype in the mammalian immune system (Kirk *et al.*, 2008). Only two proteins, PI31 and F-box only protein 7 (Fbxo7), have been shown to harbour the FP domain so far.

Figure 5
Potential polar-core structure of Spo0M^{11–258}. (a) Polar-core residues in Spo0M^{11–258}. Two separated parts of the polar-core structure are indicated by red squares. (b, c) Hydrogen-bond network between the N domain (green) and C domain (magenta) in each part indicated in (a) shown in stereo.

From the crystal structure and biochemical studies of the PI31 protein, PI31 is suggested to form a homodimer in solution *via* the α -helices of its FP domains (α -interface). In addition, PI31 forms a homodimer in the crystal *via* the β -strands of its FP domains (β -interface). The β -interface has also been suggested to contribute to the formation of a heterodimer with Fbxo7 in cells. Interestingly, while the α -helix of PI31, which forms an α -interface, was not conserved between PI31 and Spo0M, the β -strand forming a β -interface in the PI31 crystal did seem to be structurally conserved in the C-interface in the Spo0M^{11–258} crystal (Fig. 4b). These findings suggested that the interface is a conserved dimerization interface between *B. subtilis* Spo0M and human PI31, that a common evolutionary origin might exist and that Spo0M might have similar functions to mammalian FP domains.

3.6. Potential polar-core structure of Spo0M

The polar core is a structurally and functionally important feature of arrestin, and is found in the crystal structures of β /v-arrestins and VPS26 (Shi *et al.*, 2006) but not in those of TXNIP (Hwang *et al.*, 2014). As shown in Fig. 5, Spo0M was assumed to harbour a polar-core structure located at the interface between the N and C domains, which is represented by charge–charge interactions. The charge–charge interactions include the N-domain residues Glu67 (β 4), Arg69 (β 4) and Asp102 (loop between β 5 and β 6) and the C-domain residues Asp158 (β 8), Arg168 (loop between α 2 and β 9), Gln173 (β 9) and Arg216 (loop between β 11 and α 4). The potential polar core of Spo0M^{11–258} was derived into two parts as reported in the crystal structure of the VPS26 protein (Fig. 5a; Shi *et al.*, 2006). The first part of the network involves the side chains of Glu67 and Arg69 of the N domain and Asp158 and Arg216 of the C domain, and the main chain of Val60 of the N domain and Cys159 and Gln161 of the C domain (Fig. 5b). The second part of the network involves the side chains of Asp102 of the N domain and Arg168 and Gln173 of the C domain and the main chain of Ile74 and Ile105 of the N domain (Fig. 5c). As demonstrated in other arrestin-clan members, this potential polar-core structure of Spo0M may play a critical role in the function of the protein.

4. Conclusion

The Spo0M proteins were successfully expressed, purified and crystallized. Although initial attempt to solve the structure of the Spo0M^{FL} protein failed owing to phasing problems and poor diffraction, truncation-mutant screening using the DSF method resulted in a more stable Spo0M^{11–258} protein lacking ten amino acids at the N-terminus which had a T_m that was 10°C higher than that of Spo0M^{FL}. The crystal structure of Spo0M^{11–258} was successfully determined by the SAD method. The N domain of Spo0M adopts a typical arrestin-like fold that particularly resembles α -arrestins, which include human ARRDC1–5 and TXNIP, while sharing an extremely poor level of sequence identity. This finding indicates that the β -sandwich fold is a structurally conserved feature in the

arrestin clan from bacteria to mammals and that Spo0M may share some functions with mammalian arrestin-clan members. In contrast to the N domain, the C domain of Spo0M had no structural homology to arrestin-clan proteins. These findings about the Spo0M N domain indicate that Spo0M is the first bacterial member found to belong to the arrestin clan and also that there are arrestin-clan members that have only one β -sandwich domain. Since a potential arrestin-clan protein with one half of the two-lobe domain has also been predicted in other organisms (Aubry *et al.*, 2009), the ancient arrestin might have a monodomain structure. Surprisingly, we found that the FP domain of human PI31 had significant structural homology to the C domain of Spo0M, suggesting that there may be a common evolutionary origin between the Spo0M C domain and mammalian FP domains. A functional study will be needed to clarify the role of each domain and to determine whether there may also be functional similarities between bacterial Spo0M and mammalian arrestin-clan proteins as well as proteins containing the FP domain.

Acknowledgements

X-ray diffraction data were collected at SPring-8 (proposal Nos. 2013B6849, 2014A1308, 2014A1279, 2014A1279, 2014B1321, 2014B6948 and 2014B1274) with assistance from the beamline scientists.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Alvarez, C. E. (2008). *BMC Evol. Biol.* **8**, 222.
- Aubry, L., Guetta, D. & Klein, G. (2009). *Curr. Genomics*, **10**, 133–142.
- Bujacz, A. (2012). *Acta Cryst.* **D68**, 1278–1289.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Deuerling, E., Mogk, A., Richter, C., Purucker, M. & Schumann, W. (1997). *Mol. Microbiol.* **23**, 921–933.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N. & Nordlund, P. (2006). *Anal. Biochem.* **357**, 289–298.
- Fawcett, P., Eichenberger, P., Losick, R. & Youngman, P. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 8063–8068.
- Han, M., Gurevich, V. V., Vishnivetskiy, S. A., Sigler, P. B. & Schubert, C. (2001). *Structure*, **9**, 869–880.
- Han, W. D., Kawamoto, S., Hosoya, Y., Fujita, M., Sadaie, Y., Suzuki, K., Ohashi, Y., Kawamura, F. & Ochi, K. (1998). *Gene*, **217**, 31–40.
- He, L. & Niemeyer, B. (2003). *Biotechnol. Prog.* **19**, 544–548.
- Hirsch, J. A., Schubert, C., Gurevich, V. V. & Sigler, P. B. (1999). *Cell*, **97**, 257–269.
- Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
- Hwang, J., Suh, H.-W., Jeon, Y. H., Hwang, E., Nguyen, L. T., Yeom, J., Lee, S.-G., Lee, C., Kim, K. J., Kang, B. S., Jeong, J.-O., Oh, T.-K., Choi, I., Lee, J.-O. & Kim, M. H. (2014). *Nature Commun.* **5**, 2958.
- Kang, D. S., Tian, X. & Benovic, J. L. (2014). *Curr. Opin. Cell Biol.* **27**, 63–71.
- Kirk, R., Laman, H., Knowles, P. P., Murray-Rust, J., Lomonosov, M., Meziane, E. K. & McDonald, N. Q. (2008). *J. Biol. Chem.* **283**, 22325–22335.
- Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* **D60**, 2256–2268.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

- Lefkowitz, R. J. & Shenoy, S. K. (2005). *Science*, **308**, 512–517.
- Lobanov, M. I., Bogatyreva, N. S. & Galzitskaya, O. V. (2008). *Mol. Biol.* **42**, 623–628.
- Lysenko, E., Ogura, T. & Cutting, S. M. (1997). *Microbiology*, **143**, 971–978.
- Molle, V., Fujita, M., Jensen, S. T., Eichenberger, P., González-Pastor, J. E., Liu, J. S. & Losick, R. (2003). *Mol. Microbiol.* **50**, 1683–1701.
- Nicholson, W. L., Munakata, N., Horneck, G., Melosh, H. J. & Setlow, P. (2000). *Microbiol. Mol. Biol. Rev.* **64**, 548–572.
- Ortega, A., Amorós, D. & García de la Torre, J. (2011). *Biophys. J.* **101**, 892–898.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Patwari, P. & Lee, R. T. (2012). *Trends Endocrinol. Metab.* **23**, 216–222.
- Phillips, Z. E. & Strauch, M. A. (2002). *Cell. Mol. Life Sci.*, **59**, 392–402.
- Piggot, P. J. & Hilbert, D. W. (2004). *Curr. Opin. Microbiol.* **7**, 579–586.
- Polekhina, G., Ascher, D. B., Kok, S. F., Beckham, S., Wilce, M. & Waltham, M. (2013). *Acta Cryst. D* **69**, 333–344.
- Qi, S., O’Hayre, M., Gutkind, J. S. & Hurley, J. H. (2014). *Protein Sci.* **23**, 1708–1716.
- Shi, H., Rojas, R., Bonifacino, J. S. & Hurley, J. H. (2006). *Nature Struct. Mol. Biol.* **13**, 540–548.
- Siranosian, K. J. & Grossman, A. D. (1994). *J. Bacteriol.* **176**, 3812–3815.
- Sonenshein, A. L. (2000). *Curr. Opin. Microbiol.* **3**, 561–566.
- Sonoda, Y., Newstead, S., Hu, N.-J., Alguet, Y., Nji, E., Beis, K., Yashiro, S., Lee, C., Leung, J., Cameron, A. D., Byrne, B., Iwata, S. & Drew, D. (2011). *Structure*, **19**, 17–25.
- Stein, P. E., Leslie, A. G. W., Finch, J. T. & Carrell, R. W. (1991). *J. Mol. Biol.* **221**, 941–959.
- Stragier, P. & Losick, R. (1996). *Annu. Rev. Genet.* **30**, 297–341.
- Strauch, M., Webb, V., Spiegelman, G. & Hoch, J. A. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 1801–1805.
- Sutton, R. B., Vishnivetskiy, S. A., Robert, J., Hanson, S. M., Raman, D., Knox, B. E., Kono, M., Navarro, J. & Gurevich, V. V. (2005). *J. Mol. Biol.* **354**, 1069–1080.
- Thi Nguyen, H. B. & Schumann, W. (2012). *Res. Microbiol.* **163**, 114–118.
- Tian, X., Kang, D. S. & Benovic, J. L. (2014). *Handb. Exp. Pharmacol.* **219**, 173–186.
- Wehrli, W., Niederweis, M. & Schumann, W. (2000). *J. Bacteriol.* **182**, 3870–3873.
- Weir, J., Predich, M., Dubnau, E., Nair, G. & Smith, I. (1991). *J. Bacteriol.* **173**, 521–529.
- Wu, J.-J., Schuch, R. & Piggot, P. J. (1992). *J. Bacteriol.* **174**, 4885–4892.
- Wünsche, A., Hammer, E., Bartholomae, M., Völker, U., Burkovski, A., Seidel, G. & Hillen, W. (2012). *FEBS J.* **279**, 2201–2214.