

Expression of soluble, active fragments of the morphogenetic protein SpoIIE from *Bacillus subtilis* using a library-based construct screen

Andrea E. Rawlings¹, Vladimir M. Levdikov¹,
Elena Blagova¹, Vicki L. Colledge¹, Philippe J. Mas³,
James Tunaley¹, Ludmila Vavrova², Keith S. Wilson¹,
Imrich Barak², Darren J. Hart^{3,4} and Anthony
J. Wilkinson^{1,4}

¹Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5YW, UK, ²Institute of Molecular Biology, Slovak Academy of Sciences, 845 51 Bratislava 45, Slovakia and ³EMBL Grenoble (UMI3265, UJF-EMBL-CNRS), BP 181, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France

⁴To whom correspondence should be addressed.
E-mail: hart@embl.fr; ajw@ysbl.york.ac.uk

Received July 15, 2010; revised July 15, 2010;
accepted July 21, 2010

Edited by Alan Fersht

SpoIIE is a dual function protein that plays important roles during sporulation in *Bacillus subtilis*. It binds to the tubulin-like protein FtsZ causing the cell division septum to relocate from mid-cell to the cell pole, and it dephosphorylates SpoIIAA phosphate leading to establishment of differential gene expression in the two compartments following the asymmetric septation. Its 872 residue polypeptide contains a multiple-membrane spanning sequence at the N-terminus and a PP2C phosphatase domain at the C-terminus. The central segment that binds to FtsZ is unlike domains of known structure or function, moreover the domain boundaries are poorly defined and this has hampered the expression of soluble fragments of SpoIIE at the levels required for structural studies. Here we have screened over 9000 genetic constructs of spoIIE using a random incremental truncation library approach, ESPRIT, to identify a number of soluble C-terminal fragments of SpoIIE that were aligned with the protein sequence to map putative domains and domain boundaries. The expression and purification of three fragments were optimised, yielding multimilligram quantities of the PP2C phosphatase domain, the putative FtsZ-binding domain and a larger fragment encompassing both these domains. All three fragments are monomeric and the PP2C domain-containing fragments have phosphatase activity.

Keywords: SpoIIE/sporulation/ESPRIT/phosphatase/cell division

Introduction

The formation of spores by *Bacillus subtilis* is an extreme response to starvation and a primitive system of cell

differentiation. During sporulation, the cell divides asymmetrically to produce a larger mother cell and a smaller forespore (Fig. 1A). The mother cell later engulfs the forespore and nurtures it as it develops into a resistant spore. Although the mother cell and the forespore contain identical chromosomes, they follow different programmes of gene expression and have different fates (Errington, 2003; Hilbert and Piggot, 2004). This is brought about by the sequential activation of four compartment-specific sigma factors that direct RNA polymerase to distinct sets of promoters. The forespore-specific σ^F is the first of these to be activated. Although it is present in the pre-divisional cell and it partitions into both compartments, σ^F becomes active only after the formation of the polar septum and only in the forespore.

σ^F (SpoIIAC) activity is governed by a complex interplay with three other proteins, SpoIIAB and SpoIIAA encoded on the three cistron *spoIIA* operon, and SpoIIE. SpoIIAB is an anti-sigma factor that forms a tight complex with σ^F and prevents the latter from combining with core RNA polymerase and initiating transcription. It is also a protein kinase that phosphorylates SpoIIAA on a conserved serine. The phosphorylation state of SpoIIAA, which is an anti-sigma factor antagonist, is the arbiter of cell fate. In the pre-divisional cell and the mother cell following asymmetric septation, SpoIIAA is phosphorylated and σ^F is sequestered as the σ^F :SpoIIAB₂ complex in which it is inactive (Fig. 1A). In the forespore compartment, SpoIIAA~P becomes dephosphorylated, and thereby activated, by a specific phosphatase, SpoIIE. Unphosphorylated SpoIIAA displaces σ^F from the σ^F :SpoIIAB₂ complex (Campbell *et al.*, 2002; Carniol *et al.*, 2004; Masuda *et al.*, 2004). Relieved of SpoIIAB inhibition, σ^F can combine with core RNA polymerase and activate expression of forespore-specific sporulation genes.

SpoIIE is also involved in switching the site of assembly of the cell division apparatus from mid-cell during normal cell division to a polar location during sporulation (Khvorova *et al.*, 1998). This switch involves the migration of the tubulin-like protein, FtsZ, which initially forms a ring (termed the Z-ring) at mid-cell to sites near to the cell poles. Z-ring migration requires a sporulation-specific increase in the levels of FtsZ and expression of *spoIIE* (Khvorova *et al.*, 1998; Ben Yehuda and Losick, 2002). SpoIIE localises to the two polar sites of the cell in the form of rings (the E-rings) (Arigoni *et al.*, 1995; Barak *et al.*, 1996; Levin *et al.*, 1997). SpoIIE helps to stabilise the Z-ring at these positions and one of these E/Z-rings matures into the asymmetric septum. The precise role of SpoIIE in polar Z-ring (re)assembly and asymmetric cell division is not known, although the implied interaction between SpoIIE and FtsZ is supported by data from yeast two-hybrid screens and assays of the purified components (Lucet *et al.*, 2000).

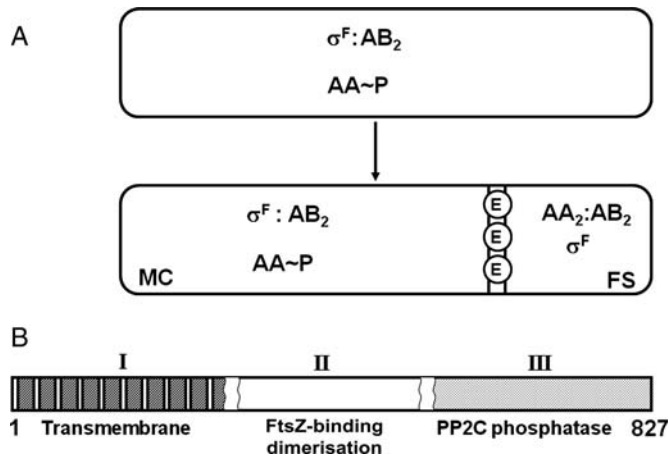


Fig. 1. Asymmetric cell division and the role of SpoIIE. (A) In the pre-divisional cell (upper) and the mother cell (MC) following cell division (lower), SpoIIAA (AA) is phosphorylated and σ^F is in a complex with SpoIIAB (AB_2). In the forespore (FS), σ^F is free and AA and AB are in complex. SpoIIE (E) accumulates at the asymmetric septum, a double membrane structure. (B) The putative three domain structure of SpoIIE. The limits of the putative FtsZ-binding domain are uncertain.

The interplay of SpoIIE, FtsZ and SpoIIAA~P serves to couple the morphological event of polar septation to the activation of σ^F and compartment-specific gene expression. How, having localised to the site of asymmetric septation, SpoIIE brings about activation of σ^F in the forespore, but not in the mother cell, has been a subject of much conjecture, with models based on differential membrane localisation and genetic asymmetry having been proposed (Wu *et al.*, 1998; Frandsen *et al.*, 1999; Pan *et al.*, 2001). It has been shown however that an imbalance in the activity of SpoIIE resulting from the (~10-fold) volume difference between the compartments can account for differential regulation (Clarkson *et al.*, 2004; Iber *et al.*, 2006).

The structures of the SpoIIA proteins are known (Kovacs *et al.*, 1998; Seavers *et al.*, 2001; Campbell *et al.*, 2002; Masuda *et al.*, 2004). These structures define the active and inactive (phosphorylated) forms of SpoIIAA and the interactions of SpoIIAB with σ^F and SpoIIAA. What is missing is structural insight into SpoIIE and its interactions with SpoIIAA and FtsZ. The 827 residues of *B. subtilis* SpoIIE are arranged in three functional domains (Fig. 1B). Domain I at the N-terminus contains 10 putative membrane-spanning segments (Arigoni *et al.*, 1999). A large central domain II is involved in protein–protein interactions between SpoIIE subunits and with the cell division protein FtsZ (Khvorova *et al.*, 1998; Lucet *et al.*, 2000). The C-terminal domain III possesses the SpoIIAA~P phosphatase activity (Duncan *et al.*, 1995).

SpoIIE and fragments thereof have been purified in low quantities (i) using an intein-chitin-binding domain fusion system or (ii) following refolding from inclusion bodies allowing biochemical characterisation of the molecule (Lucet *et al.*, 2000). Increasing the expression levels to the multi-milligram levels required to initiate and sustain structural analyses has proved difficult. We reasoned that the poor definition of the domain structure was a factor militating against the high level expression of stable and soluble protein. Domain I is defined by the pattern of hydrophobic residues

and topology analysis (Arigoni *et al.*, 1999), and domain III is defined by homology to PP2C type phosphatases. However, the domain boundaries are imprecise. Domain II has no recognisable motifs or homology to proteins of known structure. To define the domain boundaries in SpoIIE with greater precision, we adopted a truncation library mutagenesis approach using a recently developed method, ESPRIT (expression of soluble proteins by random incremental truncation; Tarendeau *et al.*, 2007; Yumerefendi *et al.*, 2010). In ESPRIT, a comprehensive, oversampled gene deletion library is constructed in which terminally truncated coding sequences are generated by exonuclease III deletion and fused to a sequence encoding a C-terminal biotin acceptor peptide (Beckett *et al.*, 1999). This tag is efficiently biotinylated *in vivo* by the endogenous *Escherichia coli* BirA enzyme and biotinylation levels can be measured by fluorescent streptavidin binding to robotically prepared colony blots. In the first screening step, ranking of clones permits enrichment of putative soluble clones from the library since the biotinylation reaction is more efficient when the target protein is soluble. In the second step, the solubility and purifiability are confirmed by purification screening using an N-terminal hexahistidine tag.

The ESPRIT method has been successfully applied to a number of challenging targets that had previously resisted soluble expression, most notably the PB2 subunit of influenza virus RNA polymerase (Tarendeau *et al.*, 2007; Guilligay *et al.*, 2008; Tarendeau *et al.*, 2008). Here we describe the truncation mutagenesis and screening of SpoIIE leading to the generation of previously unidentified, highly soluble and enzymatically active domains that are purifiable at a quality and quantity suitable for structural and functional studies.

Materials and methods

Subcloning of the SpoIIE gene

The *spoIIE*-coding sequence was cloned into a pET11a plasmid derivative. A silent mutation of the Ala433 codon from GCA to GCG was introduced by site-directed mutagenesis using the Quikchange (Stratagene) system to remove an *Nsi*I restriction endonuclease cleavage recognition site (ATGCAT). The mutated plasmid was used as template for the PCR amplification of a fragment of the *spoIIE*-coding sequence encompassing base pairs 901–2481. The primers used were spoIIE901F 5'-GATCCT AGGGCG CGCCGC CGGTCT GATGAC AACGCT CTATGA GTCAT-3' and spoIIE2481R 5'-CTAGGA TCATGC ATTTGA AATTTC TTGTTT GTTTTG AAAGAT TGCCGG-3' (where the underlined bases are restriction endonuclease cleavage recognition sequences). The resulting 1612 bp fragment was digested with *Nsi*I and *Asc*I and ligated to the similarly digested pET9a-derived plasmid, pTAR010 (Tarendeau *et al.*, 2008). The ligation products were introduced into competent *E. coli* Novablue cells and kanamycin-resistant transformants harbouring the plasmid pTAR-*spoIIE*, in which codons 301–827 are fused in-frame to sequences encoding a TEV protease cleavable hexahistidine tag (MGHHHHHDYDIPTTENLYFQG) at the N-terminus and a biotin acceptor peptide (SNNGSGGLNDIFEAQKIEW HE) at the C-terminus, were identified.

Construction of the spoIIIE deletion library

Ten micrograms of pTAR-*spoIIIE* plasmid DNA was cleaved upstream of the 5'-end of the inserted *spoIIIE* sequence with *AatII*. Having established by gel electrophoretic analysis that the DNA had been efficiently linearised, the plasmid was further digested with *AscI* yielding a DNA fragment with a 5' *AscI* end, which is sensitive to exonuclease III, and a 3' *AatII* end that is insensitive to the action of this enzyme. Four micrograms of the pTAR-*spoIIIE* fragment was digested with 400 units of exonuclease III (New England Biolabs, NEB) in 120 μl of NEB Buffer 1 supplemented with 50 mM NaCl. The reaction mixture was incubated at 22°C and, to ensure an even distribution of fragments, 2 μl aliquots were withdrawn each minute and quenched by addition to a tube containing 200 μl of ice cold 3 M NaCl. The sample was then heated to 70°C to denature the exonuclease III and salt was removed by passage through a Nucleospin Extract II column (Macherey Nagel). As exonuclease III removes mononucleotides stepwise from recessed 3' termini, the products of the digestion reaction have 5' overhangs of varying lengths. To remove single-stranded overhangs, the DNA in a reaction volume of 50 μl was treated with 5 units of mung bean nuclease in its cognate buffer (NEB) at 30°C for 30 min. The DNA was repurified using the Nucleospin Extract II kit and treated with 1 unit of Pfu polymerase, 2 mM dNTPs in 50 μl of Pfu polymerase buffer for 20 min at 70°C to ensure that the products were blunt ended.

The complete reaction mix was loaded onto a 0.5% agarose gel and the products were resolved by electrophoresis. With reference to suitable molecular weight markers, two slices of the gel were excised containing DNA fragments with expected truncations of approximately 0–750 and 750–1500 bp. These will be referred to as the 'large' and 'small' clone sublibraries, respectively. The linear DNA extracted from the agarose slices, was circularised using T4 DNA ligase and a 2 μl aliquot of the ligation products was mixed with 50 μl competent *E.coli* Omnimax cells (Invitrogen). After incubation on ice for 5 min, 42°C for 30 s and on ice for a further 5 min, the cells were allowed to recover in 1 ml of super optimal broth with catabolite repression medium at 37°C in a shaker incubator. One hundred microlitres of the cell suspension was plated onto solid LB media and kanamycin-resistant transformants were isolated. The plasmids from 12 colonies were purified and analysed for insert size to assess the quality of the library. The remaining portions of the stored transformation mixes were each plated onto six LB-kanamycin plates. Following overnight incubation and colony growth, 1 ml of liquid LB medium was placed on the surface of each plate and a glass spreader was used to gently resuspend the cells. The six cell suspensions from each library were pooled and plasmid DNA was extracted and purified using a Qiagen miniprep kit employing three spin columns per library. The two supercoiled plasmid DNA preparations were used to transform competent *E.coli* BL21 (DE3) RIL for expression screening.

Processing of the spoIIIE truncation library

A volume of the transformation mixture sufficient to generate approximately 4000 colonies per plate was spread onto LB-agar kanamycin (30 $\mu\text{g ml}^{-1}$)/chloramphenicol (100 $\mu\text{g ml}^{-1}$) media prepared in square Q-trays (Genetix,

UK) and colonies were allowed to grow overnight at 37°C. Colonies were picked from the Q-trays using a Kbiosystems colony picking robot and used to inoculate media in 384-well plates containing 70 μl Terrific Broth (TB) medium with 30 $\mu\text{g ml}^{-1}$ kanamycin and 100 $\mu\text{g ml}^{-1}$ chloramphenicol. Sixteen plates were filled from the small subclone library and eight from the large subclone library. Plates were shaken overnight at 300 rpm for 20 h at 30°C.

Using the same robot, samples from the 384-well cultures were arrayed onto 22 cm square nitrocellulose membranes (Amersham) on LB agar plates supplemented with antibiotics as above, according to a 3 \times 3 gridding scheme in which each sample was present in duplicate. Colonies were developed on the membrane during overnight incubation at 30°C. Expression of recombinant protein was induced by transferring the membrane onto identical medium, supplemented with 0.1 mM isopropyl- β -D-thiogalactose and 50 μM biotin with a further 4 h incubation at 30°C.

Screening step 1: enrichment from the library of clones expressing putatively soluble SpoIIIE fragments

The nitrocellulose membrane was lifted from the Q-tray and placed first onto Whatman filter paper saturated with lysis solution (0.5 M NaOH, 1.5 M NaCl) for 10 min and subsequently onto filter paper soaked in neutralisation solution (1.5 M Tris pH 7.5, 1.5 M NaCl) for 3 \times 5 min. Cell debris was washed from the membrane by gentle scraping in 2 \times SSC solution (17.5 g NaCl, 8.8 g trisodium citrate, pH 7.0 per litre of water) and the membranes were incubated overnight at 4°C in Superblock (Pierce). The membranes were washed in PBS-T buffer (NaCl/P_i with 0.2% Tween-20) and incubated for 1 h at 4°C with a monoclonal antibody against hexahistidine (Amersham; 1:3000 in PBS-T buffer). The membrane was extensively washed with PBS-T buffer and then suspended in PBS-T buffer containing a secondary mouse antibody conjugated to Alexa532 (Invitrogen) and streptavidin conjugated to Alexa488 (Invitrogen) at dilutions of 1:1000 and 1:5000, respectively. After a 1 h incubation at 4°C, the membranes were washed in PBS-T buffer and then with distilled water. The fluorescent signals were visualised and recorded with a Typhoon Trio imager (GE healthcare) using lasers tuned to excitation wavelengths of 488 and 532 nm to quantify the biotin and hexahistidine signals, respectively.

Screening step 2: purification analysis of isolated SpoIIIE expression clones

Clones with clearly visible hexahistidine signals were ranked according to their biotinylation signal intensity (as a marker for soluble expression) and the top 96 clones were analysed for their insert size by PCR using T7 forward and T7 reverse primers and 1 μl of cell culture as template. The PCR products were analysed by agarose gel electrophoresis.

Expression testing was also carried out on the 96 clones. Four millilitre cultures in the TB medium, supplemented with antibiotics as before, were grown by shaking in 24-well plates and the expression of recombinant protein was induced by the addition of IPTG to 1 mM at 24°C overnight. As described previously (Angelini *et al.*, 2009), cells were lysed using a sphaeroplast method and a Tecan liquid handling robot facilitated purification using a 96-well filter plate containing Ni²⁺-nitriloacetic acid resin.

After washing the resin, the bound protein was eluted in buffer containing 300 mM imidazole. The eluted protein samples were analysed by SDS-polyacrylamide gel electrophoresis.

Protein purification

The B1' and B2–B1-expressing cells harbour pET-YSB LIC-3C derivative plasmids (Fogg and Wilkinson, 2008) containing the subcloned coding sequences of SpoIIE(590–827) and SpoIIE(375–595), respectively linked to a sequence encoding an N-terminal MSSGHHHHHH tag with a human rhinovirus (HRV) 3C protease cleavage site.

Half litre cultures of *E.coli* BL21 (DE3) harbouring plasmids encoding the B2 (SpoIIE(375–827)), H1 (SpoIIE(412–827)), B1' (SpoIIE(590–827)) and B2–B1 (SpoIIE(375–595)) fragments were grown at 37°C until the OD₆₀₀ reached 0.6–0.8. Production of recombinant protein was induced by the addition of IPTG to 1 mM and growth with shaking was continued at 16°C overnight for the B2, H1 and B2–B1 expressing cells or at 37°C for 4 h for the B1'-expressing cells. The cells were harvested by centrifugation and resuspended in 25 ml of 50–100 mM Tris–HCl pH 8.2, 50 mM NaCl, (buffer A) in the presence of Roche protease inhibitor cocktail. One hundred millimolar urea was included in the buffers used for purification of the B2 fragment. The cells were lysed by sonication and the extracts were fractionated by metal chelation chromatography on a HiTrap column charged with nickel. The SpoIIE fragments were highly enriched in successive fractions eluting from this column at circa 250 mM imidazole in yields of 10–15 mg. The affinity tags were removed by proteolytic cleavage (either TEV or HRV-3C) followed by a second passage through the HiTrap column and fractionation on an Superdex S200 16/60 gel filtration column (GE Biosciences) equilibrated in buffer A.

Phosphatase assay

SpoIIAA-phosphate phosphatase assays of the SpoIIE fragments were performed with recombinant SpoIIAA from *Bacillus sphaericus*. SpoIIAA was prepared from *E.coli* cultures expressing *spoIIAA* and SpoIIAA~P was prepared from cells co-expressing *spoIIAA* and *spoIIAB* as described previously (Seavers et al., 2001). The SpoIIE fragments were mixed with 1.6 nmol SpoIIAA~P at a molar ratio of either 1:100 or 1:400 in a 10 µl volume of 50 mM Tris pH 8.2 and 100 mM NaCl containing 10 mM MnCl₂ and the reaction was allowed to proceed for 30 min at 30°C. Ten microlitres of loading dye was added to the sample and the reaction products were analysed by gel electrophoresis in 8.75% polyacrylamide gels under non-denaturing conditions. Gels were run at 100 V for 120 min and stained with Coomassie blue dye.

SEC-MALLS

For determination of the oligomeric state of the proteins, the SpoIIE fragments were analysed using a HPLC-MALLS apparatus. Protein of 0.1 ml samples at ~1 mg ml⁻¹ were loaded onto a Superdex 200, 10/300 gel filtration column equilibrated at 0.5 ml min⁻¹ with a mobile-phase consisting of 50 mM Tris–HCl and 150 mM sodium chloride at pH 8.0. The eluate was passed successively through a SPD20A UV/Vis detector, a Wyatt Dawn HELEOS-II 18-angle light scattering detector and a Wyatt Optilab rEX refractive index monitor with the system driven by a Shimadzu HPLC system

comprising an LC-20AD pump. The data were processed and molecular masses calculated using the Astra V software (Wyatt).

FtsZ-binding assay

Bacillus subtilis FtsZ was isolated and purified as described previously using the *E.coli* expression strain W3110 (Wang and Lutkenhaus, 1993). Purified SpoIIE(412–827), SpoIIE(590–827) and FtsZ were dialysed overnight at 4°C into buffer D (150 mM NaCl, 50 mM Tris–HCl pH 7.5). FtsZ-binding assays were performed by mixing FtsZ with SpoIIE(412–827) and SpoIIE(590–827) samples. Ten microlitres of SpoIIE(412–827) or SpoIIE(590–827) at 0.7 mg ml⁻¹ was mixed with 10 µl of 0.7 mg ml⁻¹ FtsZ in the presence of 1 mM GTP and 10 mM MnCl₂ and incubated on ice for 1 h. Glycerol was added to the samples to a final concentration of 20% and samples were subjected to non-denaturing 8.75% PAGE. The gel was stained with Coomassie blue to visualise the protein species.

Results

Library construction

An exonuclease III, mung bean nuclease truncation strategy was applied to a cloned 1581 bp fragment of *spoIIE*-encoding residues 300–827 of SpoIIE which comprise the C-terminus of the putative transmembrane domain, the putative oligomerisation and FtsZ-binding region, and the PP2C-type phosphatase domain. This fragment was cloned downstream of a sequence encoding an N-terminal cleavable hexahistidine tag and upstream and in-frame with a 3' sequence encoding a C-terminal biotin acceptor peptide to be used as the solubility reporter. The protocol generates a random library of constructs in which all possible truncations of the 1581 base pair fragment are expected to be approximately four-fold oversampled. Following the deletion steps, agarose gel analysis revealed the expected smear of linear truncation products across the range defined by the *spoIIE* fragment with evidence of some overtruncation (Supplementary Fig. S1, available at PEDS online). To reduce the likely dominance of small expression constructs, prior to DNA re-circularisation, the library was divided into two insert size ranges by excising appropriate slices. The 'small' sublibrary contained vectors with inserts of approximately 0–750 bp and the 'large' sublibrary contained inserts of approximately 750 bp full length. Following plating, 12 randomly selected clones were analysed by PCR of plasmid inserts to assess the quality of each sublibrary. All 12 clones from the large sublibrary gave clear and clean PCR products with a diversity of sizes. In contrast, for the small sublibrary, only six of the clones gave clear PCR products, although with the expected size diversity. In the remainder, the 3'-priming site had been deleted by overtruncation.

The pooled sublibraries were used to transform *E.coli* BL21 (DE3) RIL and 3072 and 6144 colonies, for the large and small sublibraries, respectively, were picked robotically into 384-well plates containing growth media. Taking account of the sublibrary efficiency and the fact that each sublibrary may contain 750 different truncation constructs, this represents approximately four-fold oversampling of both sublibraries.

Colony array screen to enrich putatively soluble constructs from the library

Clones from each sublibrary were robotically arrayed in duplicate onto nitrocellulose membranes on which colonies were subsequently grown and the expression of recombinant protein induced by transferring to agar supplemented with IPTG and biotin. Putative soluble expression clones were identified by hybridising membranes with Alexa488 streptavidin to detect the biotinylated C-terminal tag. Simultaneously, a monoclonal anti-hexahistidine antibody with associated Alexa532 mouse secondary antibody was used to detect the N-terminal hexahistidine tag; comparison of this signal with that from streptavidin helped eliminate degraded protein constructs, or those resulting from internal initiation of translation, from the set selected for further testing.

It was evident that there was a very high number of clones exhibiting signals from binding of the streptavidin conjugate especially in the large clone sublibrary (Fig. 2 and Supplementary Fig. S2A, available at *PEDS* online). In contrast, the number of clones signalling the hexahistidine tag was much lower for both sublibraries (Fig. 2 and Supplementary Fig. S2B, available at *PEDS* online). To quantify this analysis, averages of the duplicate biotin and hexahistidine signal intensities were calculated for each clone. The signal intensity values were divided into bins and frequency histograms were plotted (Supplementary Fig. S2C and D, available at *PEDS* online). The majority of clones (>80%) give negligible hexahistidine tag signals in the first two bins (Supplementary Fig. S2A, available at *PEDS* online) with the remaining clones showing a distribution of intensities that gives rise to a smooth decrease in frequency over the higher intensity bins. For the biotin array, only about a third of the clones fall in the low intensity bins, much less than would be expected given that 66% of the clones are expected to encode out-of-frame fusions (Supplementary Fig. S2B, available at *PEDS* online).

The high number of colonies displaying a biotin signal is likely due to internal translation initiation that has been noted previously for SpoIIE (Lucet *et al.*, 2000) when a contaminating 28 kDa SpoIIE species was observed in preparations of a 55 kDa SpoIIE(320–827) fragment. In that study, Edman degradation revealed a SpoIIE sequence beginning at residue 568 with an amino terminal methionine

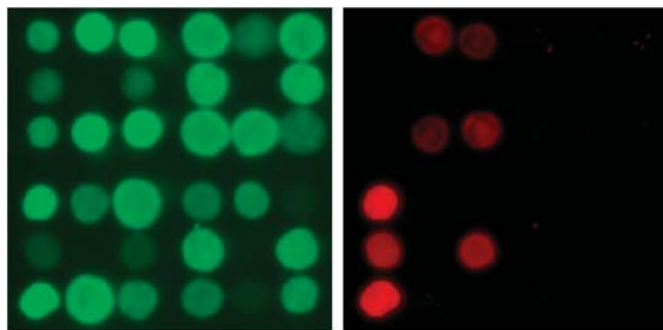


Fig. 2. Screening of the truncation library. The fluorescence signals from a small section of the colony array are shown. The array was probed for the presence of biotin with a streptavidin-Alexa fluor, which gives a green fluorescent signal as an indicator of solubility of the clone (left) and for hexahistidine with conjugated anti-His₆ antibody, which gives a red fluorescent signal (right). Comparison of the two signals permits screening out of degraded or truncated constructs.

replacing the expected Val. The Val568 codon GUG is an alternative start codon in *E.coli*, moreover it is preceded by a purine-rich sequence which may function as a Shine-Dalgarno box. If these elements do constitute an efficient internal ribosome entry site (IRES), it would explain the prevalence of biotinylated protein signals in the large sublibrary designed to express deletants beginning between residues 300 and 563, relative to the small sublibrary designed to express shorter truncated proteins with N-termini between residues 564 and 827.

The background fluorescence intensity values for both biotin and hexahistidine were subtracted from the averaged duplicate signals and those clones having a significant positive hexahistidine signal were then ranked according to their biotin intensity. The top 96 clones, designated A1–H12, according to this ranking were selected for further analysis.

Expression trials

The highest signalling 96 constructs were expressed in small-scale liquid cultures (4 ml) and an aliquot of cells used directly as template in a rapid PCR screen to assess size diversity using the flanking T7 promoter and T7 terminator primers. Gel analysis showed that in the majority of cases (67 of 96 clones) a single clean PCR fragment was observed with an apparent cluster at about 1000 bp. Inserts ranged from 1600 bp down to <100 bp. Twenty-one of the 96 clones produced no PCR product and eight gave rise to more than one product; these were not characterised further. The cells were lysed and the hexahistidine-tagged proteins purified on Ni²⁺-NTA agarose beads using a robotic protocol (Scheich *et al.*, 2003). Imidazole elution fractions were analysed by SDS-PAGE revealing only a few clones giving rise to bands consistent with the overexpression of recombinant protein, notably clone F4 with an intensely staining band at ~25 kDa whose mass was consistent with the PCR fragment size of 700 bp. Retrospective analysis revealed the absence of recombinant protein in the soluble lysis fraction for most clones. A further repeat analysis of total and soluble lysis fractions from a subset of representative clones with high ESPRIT array rankings revealed that, where expressed, the protein was often found in the total fraction and not the soluble fraction. It seemed possible that during cell lysis under the neutral pH conditions required for protein purification, a hitherto soluble but sensitive protein might aggregate into the insoluble fraction. To investigate this, we selected a promising clone B2 (from initial trials), and performed the lysis procedure in 10 different buffers chosen from a sparse-matrix set (Lindwall *et al.*, 2000; Supplementary Table S1, available at *PEDS* online). For lysis buffers 9 and 10 (Supplementary Table S1 and Fig. S3, available at *PEDS* online), which are Tris-buffered at pH 8.2, the soluble fraction was visibly enriched for the recombinant SpoIIE fragment.

Identifying clones for scale-up

In view of the low frequency of purifiable clones from the initial expression trial and indeed the high proportion of clones failing to express any recombinant protein, the sequences of the DNA fragments contained in the highest signalling 48 constructs were determined. This analysis revealed a disproportionate number of clones (36/48) with coding sequences out of frame with the N-terminal

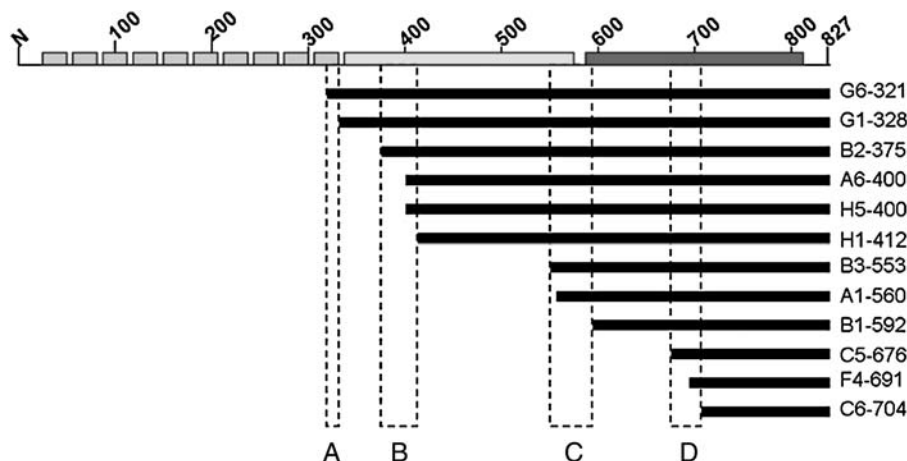


Fig. 3. Mapping soluble SpoIIE constructs onto the primary structure. The sequenced fragments of SpoIIE are represented as black bars. The numbers in bold script represent the residue numbers of full-length SpoIIE. The fragment identifier and position of the first residue are indicated. Dotted boxes indicate clusters of similarly sized fragments, representing possible windows of solubility. The approximate domain organisation of SpoIIE inferred from earlier work is indicated.

hexahistidine tag, despite exhibiting biotinylation in the colony array, consistent with the observations of possible internal initiation of translation (discussed earlier), and explaining the low frequency of purification on Ni^{2+} -NTA agarose. The 12 clones with an in-frame *spoIIE* sequence showed a diversity of sizes and the encoded protein fragments were mapped onto the SpoIIE primary sequence (Fig. 3). When arranged according to fragment size, the fragments appeared to cluster, possibly defining inter-domain segments of the polypeptide.

Site-directed mutagenesis was carried out on the 12 clones to restore the stop codon at the end of the *spoIIE*-coding sequence, thus suppressing the biotin-acceptor peptide. Small-scale expression experiments were performed in which the total cell fraction and the soluble and insoluble lysis fractions were analysed. As a result of these experiments two expressing clones were identified and taken forward for protein purification, B2 and H1 representing fragments expected to encompass the phosphatase domain (III) together with all or part of domain II, respectively. A third promising clone B1 harboured a fragment encoding residues 592–827 of SpoIIE which span the PP2C phosphatase domain. Our experience of a closely related clone expressing SpoIIE(590–810) made in a separate study, prompted us to construct a composite clone, B1', encoding SpoIIE(590–827) which expressed higher levels of soluble recombinant protein than either parent plasmid. Using the data summarised in Fig. 3, we made a further plasmid harbouring an insert encoding a B2–B1 fragment of SpoIIE encompassing residues 375–595. Following its introduction into *E.coli* BL21 (DE3), this plasmid directs expression of soluble protein.

Scale-up expression and purification

Four SpoIIE fragments, B2 (SpoIIE(375–827)), H1 (SpoIIE(412–827)) B1' (SpoIIE(590–827)) and B2–B1 (SpoIIE(375–595)) were expressed in large-scale cultures and purified by metal chelation chromatography and gel filtration (see section Experimental procedures). The B2, H1 and B2–B1 proteins eluted in two distinct peaks (data not shown) from the gel filtration column, the first peak was associated with the void volume of the column indicating the

presence of soluble aggregates. The second peaks suggested the presence of lower molecular weight material either monomers or dimers. For the B2 preparations (data not shown), the ratio of the height of the peaks in the gel filtration traces varied according to protein concentration and the age of the sample. As the tendency of the B2 fragment to aggregate could not be controlled, this fragment was not pursued further. The lower molecular mass fractions of the H1 and B2–B1 fragments in contrast were stable and these proteins were studied further together with the B1' fragment, which showed no signs of aggregation. As shown in Fig. 4A and B, the H1, B2–B1 and B1' proteins could be purified to greater than 95% purity as judged by Coomassie blue staining of SDS-polyacrylamide gels (Fig. 4A). Further analysis of the purified proteins on native polyacrylamide gels showed only a modest degree of heterogeneity, indicating that all three protein fragments are suitable for biophysical and structural investigations (Fig. 4B).

Enzyme activity

To test whether the soluble SpoIIE fragments were able to catalyse the dephosphorylation of phosphorylated SpoIIAA

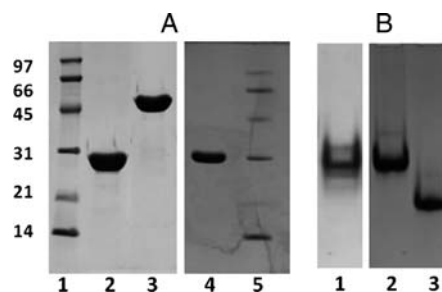


Fig. 4. Gel electrophoretic analysis of samples of the purified SpoIIE fragments. Coomassie blue staining of gels following resolution of the samples by electrophoresis. (A) 12.5% SDS-polyacrylamide gel. Lanes 1 and 5, low molecular markers (Bio-rad) with masses indicated; lane 2, the B1' fragment SpoIIE(590–827); lane 3, the H1 fragment SpoIIE(412–827) and lane 4, the B2–B1 fragment SpoIIE(375–595) and (B) 7.5% non-denaturing polyacrylamide gel lane 1, Fragment B1'; lane 2, Fragment H1 and lane 3, the B2–B1 fragment.

(SpoIIAA~P), we incubated the SpoIIE H1 and B1' fragments with SpoIIAA phosphate. We used SpoIIAA~P from *B. sphaericus* (*Bsp*SpoIIAA~P), prepared by coexpression of *Bsp*SpoIIAA with its cognate kinase *Bsp*SpoIIAB as described previously (Seavers *et al.*, 2001). *Bsp*SpoIIAA and its *B. subtilis* orthologue have 64% sequence similarity and 35% sequence identity with enhanced conservation of the residues that form the surface surrounding the phosphorylatable serine (Seavers *et al.*, 2001). The dephosphorylation assay was carried out in the presence of a large molar excess of *Bsp*SpoIIAA~P in a buffer containing 10 mM MnCl₂. The products were analysed by native polyacrylamide gel electrophoresis. As shown in Fig. 5A for the H1 fragment, the faster migrating SpoIIAA~P (lane 1) species is converted to the slower migrating SpoIIAA species (lane 2). There is greater conversion to product at the 1:100 molar ratio of H1:SpoIIAA~P (lane 3) than at 1:400 (lane 4), though for reasons that are not clear to us, it is not possible to drive the reaction to completion, a phenomenon we also observe with the PP2C domain fragment SpoIIE(590–

827) (Fig. 5A, lanes 5 and 6). This may be a manifestation of the heterologous assay system.

FtsZ Binding

To monitor the binding of the H1 fragment of SpoIIE to FtsZ, we used a native gel electrophoresis technique. As can be seen in Fig. 5B, FtsZ migrates essentially as a single species in 7.5% native polyacrylamide gels and the mobility of a proportion of the protein is increased in the presence of 1 mM GTP (lanes 1 and 2). There is some heterogeneity in the SpoIIE H1 sample (lane 3). When the two proteins are mixed, the pattern of bands resembles the sum of the individual protein bands and there is no convincing evidence of FtsZ-SpoIIE interaction, other than the more intense staining of material that scarcely enters the gel (lane 4). This may be associated with protein–protein interactions since this low mobility staining is less intense when these experiments are repeated with the B1' (PP2C domain) fragment of SpoIIE, which would not be expected to bind to FtsZ (lanes 5–7).

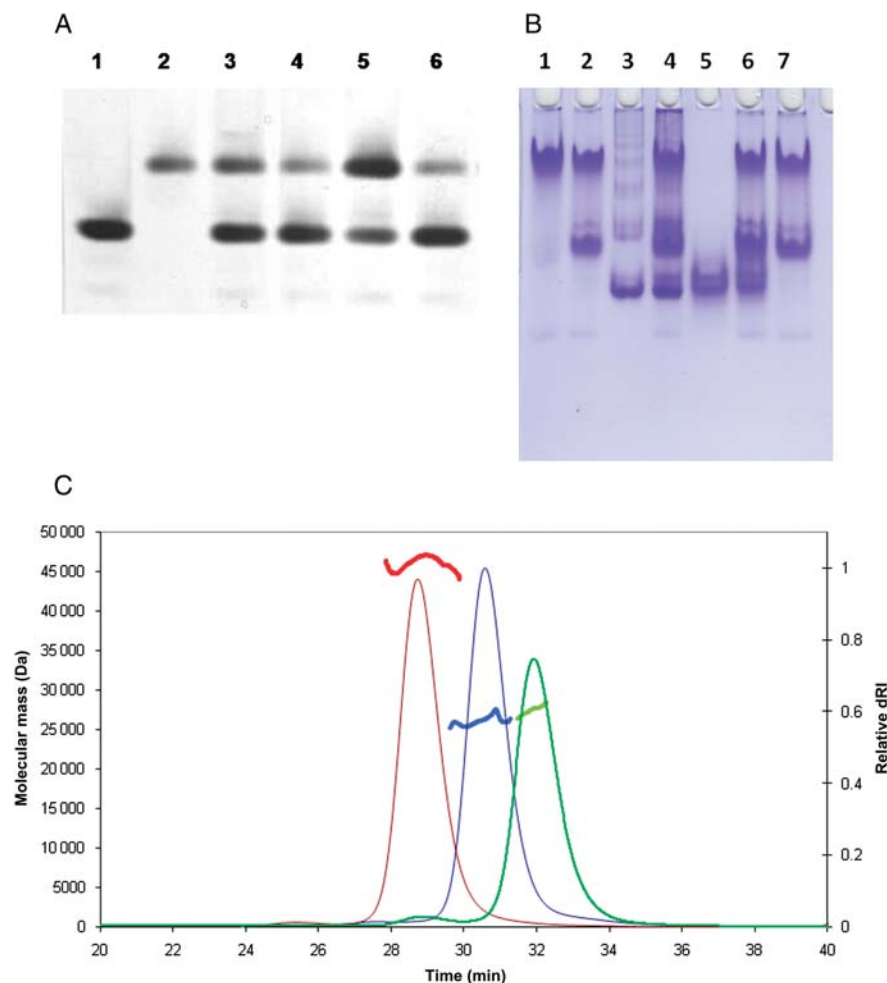


Fig. 5. Assays of SpoIIE fragments. (A) SpoIIAA-phosphate dephosphorylation by the H1 and B1' fragments monitored by native gel electrophoresis. Lane 1, SpoIIAA~P (5 μ g); lane 2, SpoIIAA (5 μ g); lanes 3 and 4, SpoIIAA~P (5 μ g) incubated in the presence of the H1 fragment at 100:1 and 400:1 molar ratios, respectively; lanes 5 and 6, SpoIIAA~P (5 μ g) incubated in the presence of the B1' fragment at 100:1 and 400:1 molar ratios, respectively. The conversion of SpoIIAA~P to the lower mobility SpoIIAA species upon incubation with the SpoIIE fragments is evident. (B) Gel mobility shift assay of FtsZ binding by the H1 and B1' fragments. Lane 1, FtsZ (7 μ g); lane 2, FtsZ (7 μ g) + 1 mM GTP; lane 3, SpoIIE H1 fragment (7 μ g); lane 4, FtsZ (7 μ g) + SpoIIE H1 (7 μ g) + 1 mM GTP; lane 5, SpoIIE B1' fragment (7 μ g); lane 6, FtsZ (7 μ g) + SpoIIE B1' (7 μ g) + 1 mM GTP; lane 7, FtsZ (7 μ g) + 1 mM GTP. There is no mobility shift evident from these gels other than the additional staining of material at the top of lane 4. (C) SEC-MALLS traces of the molecular mass and differential refractive index (dRI) versus time, of the eluate from a Superdex S200 column. The bold lines give molecular mass of the eluting species calculated from measurements of the refractive index and the multi-angle laser light scattering. Three traces for the (i) H1 (red), (ii) B1' (green) and (iii) B2–B1 (blue) fragments are overlaid.

The apparent absence of interaction of the H1 fragment of SpoIIE with FtsZ may be due to the fact that the H1 fragment lacks some of the FtsZ-binding determinants or it may be a limitation of the assay used. In this regard, we were also unable to demonstrate an interaction with FtsZ using cross-linking methods or using a gel filtration chromatography approach described previously (Lucet *et al.*, 2000). The B2–B1 fragment was not assayed as the same authors reported that both the central domain and the phosphatase domain are necessary for FtsZ binding. The assay was however carried out with the B2 fragment of SpoIIE and no interaction was seen. Here the failure to show an interaction may be related to the tendency of this protein to oligomerise.

Quaternary structure determination

The three soluble and stable fragments H1, SpoIIE(412–827); B2–B1, SpoIIE(375–595) and B1', SpoIIE(590–827) were analysed by size exclusion chromatography with MALLS to determine their oligomeric structure (Folta-Stogniew and Williams, 1999). Representative chromatograms have been overlaid in Fig. 5C. Each of the fragments exhibits a clear peak and the molar masses are calculated to be 46.3, 26.3 and 27.5 kDa for SpoIIE(412–827), SpoIIE(375–595) and SpoIIE(590–827), respectively. The calculated masses are 47.0, 25.3 and 26.5 kDa, respectively. We conclude that all three fragments are monomeric.

Discussion

Predicting stable, soluble domains of a protein with little sequence homology to previously characterised domains is a challenge. Conventional bioinformatics-led approaches in which small numbers of individual constructs are cloned and expressed is an expensive and lengthy process which is often unsuccessful. Recently, library-based strategies have provided new routes to well expressing, soluble domains (Hart and Tarendeau, 2006; Prodromou *et al.*, 2007; Savva *et al.*, 2007). The ESPRIT technique has yielded soluble fragments of the human influenza virus polymerase subunit PB2, which resulted in structures from the C-terminal region and the cap-binding domain (Tarendeau *et al.*, 2007; Guilligay *et al.*, 2008; Tarendeau *et al.*, 2008). It has also recently been used to provide stable, protease-resistant domains of CagA from the pathogenic bacterium *Helicobacter pylori* (Angelini *et al.*, 2009).

The presence of a putative IRES (Lucet *et al.*, 2000) made the identification of the best-expressing clones producing SpoIIE more difficult. Screening simultaneously for the C-terminal biotin acceptor peptide and the N-terminal hexahistidine tag reduced but did not eliminate this problem. Production of soluble SpoIIE fragments proved to be highly sensitive to induction and lysis conditions and many promising candidate fragments failed at this stage as well as during purification where latent aggregation problems were encountered. Nevertheless, through employing ESPRIT to identify the constructs coupled to subsequent optimisation of downstream expression, lysis and purification parameters, we were able to achieve the principal goal of the study which was production of pure, soluble SpoIIE domain fragments. Notably, SpoIIE(412–827), SpoIIE(375–595) and SpoIIE(590–827) were obtained in sufficient quantities to initiate and sustain

biophysical and structural studies without the need for detergents or regimes for refolding from inclusion bodies.

Although, only a minority of the highest ranking set of 12 clones expressing soluble protein *in vivo* led to stable purified protein *in vitro*, analysis of this data set is illuminating with respect to the domain structure of SpoIIE. The soluble constructs fall into four clusters A, B, C and D (Fig. 3), according to the location of their N-termini and these can be thought of as defining solubility windows. A truncated mutant that is soluble is more likely to have its N-terminus in a linker region between domains, than within a domain which may expose the hydrophobic core, disrupting folding and hence solubility.

Window A spanning residues 321–328 and defined by fragments G6 and G1, may identify the start of the SpoIIE cytoplasmic domain with residue 321 predicted to be the penultimate residue of a tenth putative transmembrane helix (Arigoni *et al.*, 1999). The 47 residue A–B intervening sequence has a high α -helical propensity according to the PredictProtein server (Rost and Liu, 2003). This segment spans the site of the classic *spoIIE48* mutation (S361F) which blocks σ^F activity (Barak, 1996; Carniol *et al.*, 2004). The A–B region may therefore constitute a small domain or subdomain. Window B, opens at residue 375 and is defined by fragments B2, A6, H5 and H1 the most compact construct beginning at position 412 (Fig. 3). The B–C intervening sequence, residues (412–553), would then constitute a second domain. The ESPRIT analysis indicates that the region assigned as the FtsZ binding and oligomerisation domain in Fig. 1B is partitioned into two soluble domains with a 40 or so residue linker (Fig. 3).

Window C which encompasses residues 553–592 leads into the phosphatase domain (Duncan *et al.*, 1995) which extends to the C-terminus. It is interesting to note that the putative IRES sequence at residue 568 occurs in window C, such that the IRES translation product would constitute a soluble functional domain. This may have functional relevance although it is not known whether the IRES is functional in *B.subtilis*. The fourth window (D) at residues 676–704 is unexpected as it is located within the PP2C phosphatase region which would be expected to constitute a stable domain based on the sequence similarity, albeit low (11%), to the PP2C domain of PstP from *Mycobacterium tuberculosis*, whose structure is known (Pullen *et al.*, 2004). It may be that this window does indeed represent a boundary between subdomains, alternatively, this fragment may simply be an unfolded piece of the target whose sequence composition and size avoids the degradation response in *E.coli*.

In summary, prior to this study, SpoIIE had not been expressed at high yields in recombinant form due to difficulties in defining stable constructs around domain boundaries, largely because no well-characterised homologues exist for constructing multiple sequence alignments. Here we applied an empirical, library-based approach comprising a truncation mutagenesis procedure and a high throughput, two-step expression screen. Despite minor difficulties due to internal initiation of translation and the need to optimise buffers and conditions for downstream steps, the constructs identified by this directed evolution-type methodology led to the production of previously undescribed, highly expressing and active protein domains. These will be of importance for structural and functional studies required for understanding the sporulation mechanism of *B.subtilis*.

Supplementary data

Supplementary data are available at *PEDS* online.

Acknowledgements

We are grateful to Drs Mark Fogg and Andrew Leech for advice on cloning and biophysical method implementation, respectively.

Funding information

This work was supported by the Wellcome Trust (Project Grant 082829), the EU (Contracts: SPINE QLG2-CT-2002-00988 and BaSysBio LSHG-CT-2006-037469) and the Biotechnology and Biological Sciences Research Council, UK as well as grants from the Slovak Academy of Sciences (2/0016/10) and the Slovak Research and Development Agency (Contract: LPP-0218-06).

References

- Angelini,A., Tosi,T., Mas,P., Acajjaoui,S., Zanotti,G., Terradot,L. and Hart,D.J. (2009) *FEBS J.*, **276**, 816–824.
- Arigoni,F., Pogliano,K., Webb,C.D., Stragier,P. and Losick,R. (1995) *Science*, **270**, 637–640.
- Arigoni,F., Guerout-Fleury,A.M., Barak,I. and Stragier,P. (1999) *Mol. Microbiol.*, **31**, 1407–1415.
- Barak,I. and Youngman,P. (1996) *J. Bacteriol.*, **178**, 4984–4989.
- Barak,I., Behari,J., Olmedo,G., Guzman,P., Brown,D.P., Castro,E., Walker,D., Westpheling,J. and Youngman,P. (1996) *Mol. Microbiol.*, **19**, 1047–1060.
- Beckett,D., Kovaleva,E. and Schatz,P.J. (1999) *Protein Sci.*, **8**, 921–929.
- Ben Yehuda,S. and Losick,R. (2002) *Cell*, **109**, 257–266.
- Campbell,E.A., Masuda,S., Sun,J.L., Muzzin,O., Olson,C.A., Wang,S. and Darst,S.A. (2002) *Cell*, **108**, 795–807.
- Carniol,K., Eichenberger,P. and Losick,R. (2004) *J. Biol. Chem.*, **279**, 14860–14870.
- Clarkson,J., Campbell,I.D. and Yudkin,M.D. (2004) *J. Mol. Biol.*, **342**, 1187–1195.
- Duncan,L., Alper,S., Arigoni,F., Losick,R. and Stragier,P. (1995) *Science*, **270**, 641–644.
- Errington,J. (2003) *Nat. Rev. Microbiol.*, **1**, 117–126.
- Fogg,M.J. and Wilkinson,A.J. (2008) *Biochem. Soc. Trans.*, **36**, 771–775.
- Folta-Stogniew,E. and Williams,K.R. (1999) *J. Biomol. Tech.*, **10**, 51–63.
- Frandsen,N., Barak,I., KarmazynCampelli,C. and Stragier,P. (1999) *Genes Dev.*, **13**, 394–399.
- Guilligay,D., Tarendeau,F. and Resa-Infante,P., *et al.* (2008) *Nat. Struct. Mol. Biol.*, **15**, 500–506.
- Hart,D.J. and Tarendeau,F. (2006) *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 19–26.
- Hilbert,D.W. and Piggot,P.J. (2004) *Microbiol. Mol. Biol. Rev.*, **68**, 234–262.
- Iber,D., Clarkson,J., Yudkin,M.D. and Campbell,I.D. (2006) *Nature*, **441**, 371–374.
- Khvorova,A., Zhang,L., Higgins,M.L. and Piggot,P.J. (1998) *J. Bacteriol.*, **180**, 1256–1260.
- Kovacs,H., Comfort,D., Lord,M., Yudkin,M.D. and Campbell,I.D. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 5067–5071.
- Levin,P.A., Losick,R., Stragier,P. and Arigoni,F. (1997) *Mol. Microbiol.*, **25**, 839–846.
- Lindwall,G., Chau,M., Gardner,S.R. and Kohlstaedt,L.A. (2000) *Protein Eng.*, **13**, 67–71.
- Lucet,I., Feucht,A., Yudkin,M.D. and Errington,J. (2000) *EMBO J.*, **19**, 1467–1475.
- Masuda,S., Murakami,K.S., Wang,S., Olson,C.A., Donigian,J., Leon,F., Darst,S.A. and Campbell,E.A. (2004) *J. Mol. Biol.*, **340**, 941–956.
- Pan,Q., Garsin,D.A. and Losick,R. (2001) *Mol. Cell*, **8**, 873–883.
- Prodromou,C., Savva,R. and Driscoll,P.C. (2007) *Drug Discov. Today*, **12**, 931–938.
- Pullen,K.E., Ng,H.L., Sung,P.Y., Good,M.C., Smith,S.M. and Alber,T. (2004) *Structure*, **12**, 1947–1954.
- Rost,B. and Liu,J. (2003) *Nucleic Acids Res.*, **31**, 3300–3304.
- Savva,R., Prodromou,C. and Driscoll,P.C. (2007) *Drug Discov. Today*, **12**, 939–947.

- Scheich,C., Sievert,V. and Büssov,K. (2003) *BMC Biotechnol.*, **3**, doi: 10.1186/1477-6750-1183-1112.
- Seavers,P.R., Lewis,R.J., Brannigan,J.A., Verschueren,K.H.G., Murshudov,G.N. and Wilkinson,A.J. (2001) *Structure*, **9**, 605–614.
- Tarendeau,F., Boudet,J. and Guilligay,D., *et al.* (2007) *Nat. Struct. Mol. Biol.*, **14**, 229–233.
- Tarendeau,F., Crepin,T., Guilligay,D., Ruigrok,R.W., Cusack,S. and Hart,D.J. (2008) *PLoS Pathog.*, **4**, e1000136.
- Wang,X. and Lutkenhaus,J. (1993) *Mol. Microbiol.*, **9**, 435–442.
- Wu,L.J., Feucht,A. and Errington,J. (1998) *Genes Dev.*, **12**, 1371–1380.
- Yumerefendi,H., Tarendeau,F., Mas,P.J. and Hart,D.J. (2010) *J. Struct. Biol.*, in press.