Biochem. J. (2013) **449**, 469–477 (Printed in Great Britain) doi:10.1042/BJ20121209

Characterization of *Staphylococcus aureus* EssB, an integral membrane component of the Type VII secretion system: atomic resolution crystal structure of the cytoplasmic segment

Martin ZOLTNER, Paul K. FYFE, Tracy PALMER and William N. HUNTER[1]

College of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, U.K.

The Type VII protein translocation/secretion system, unique to Gram-positive bacteria, is a key virulence determinant in *Staphylococcus aureus*. We aim to characterize the architecture of this secretion machinery and now describe the present study of *S. aureus* EssB, a 52 kDa bitopic membrane protein essential for secretion of the ESAT-6 (early secretory antigenic target of 6 kDa) family of proteins, the prototypic substrate of Type VII secretion. Full-length EssB was heterologously expressed in *Escherichia coli*, solubilized from the bacterial membrane, purified to homogeneity and shown to be dimeric. A C-terminal truncation, EssBΔC, and two soluble fragments termed EssB-N and EssB-C, predicted to occur on either side of the cytoplasmic membrane, have been successfully purified in a recombinant form, characterized and, together with the full-length protein, used in crystallization trials. EssB-N, the 25 kDa N-terminal cytoplasmic fragment, gave well-ordered crystals and we report the structure, determined by SAD (single-wavelength anomalous diffraction) targeting an SeMet (selenomethionine) derivative, refined to atomic (1.05 Å; 1 Å = 0.1 nm) resolution. EssB-N is dimeric in solution, but crystallizes as a monomer and displays a fold comprised of two globular domains separated by a cleft. The structure is related to that of serine/threonine protein kinases and the present study identifies that the Type VII secretion system exploits and re-uses a stable modular entity and fold that has evolved to participate in protein–protein interactions in a similar fashion to the catalytically inert pseudokinases.

Key words: early secretory antigenic target of 6 kDa system 1 (ESX-1), Gram-positive bacterium, protein kinase, protein secretion, pseudokinase, X-ray crystallography.

INTRODUCTION

The T7SS (Type VII secretion system) is a specialized and complex protein-secretion pathway first identified in *Mycobacterium tuberculosis*, the infectious agent that causes tuberculosis [1]. T7SS gene clusters are widespread among Gram-positive bacteria of the phyla Actinobacteria and Firmicutes [2]. A characteristic of these clusters is the presence of genes encoding ESAT-6 (early secreted antigenic target of 6 kDa) proteins. These small highly immunogenic polypeptides represent the prototype substrates for the T7SS [3]. The clusters also encode an integral membrane protein containing domains related to both the FtsK cell division protein from *Escherichia coli* [4] and the stage III sporulation protein SpoIIIE from *Bacillus subtilis* [5]. Other genes in the T7SS genomic loci are less well conserved, although in general they are essential for a functional secretion system to be formed in organisms of each phylum [1,6].

The T7SS is a major virulence determinant in Gram-positive bacteria, indeed it is the loss of T7SS genes that attenuates pathogenicity of the vaccine strain *Mycobacterium bovis* BCG (Bacille Calmette–Guérin) [7]. Deletion of the relevant genomic region in *M. tuberculosis* mimics the BCG attenuation [8] and in the reverse control experiment BCG partially regains virulence upon complementation [9]. In *Staphylococcus aureus*, a versatile opportunistic pathogen and major cause of nosocomial and community-acquired infections [10], mutants compromised in their ability to secrete the ESAT-6-family proteins EsxA and EsxB display a 2–4-log reduction in virulence in a murine pathogenicity model [6]. This is consistent with the observations made in *Mycobacteria*. However, the presence of T7SS gene clusters in non-pathogenic bacteria, for example *Streptomyces coelicolor* and *B. subtilis*, suggests this secretion system is not a specific adaption to virulence itself, but may perhaps contribute to survival fitness.

Little is known about the structure of the T7SS, or individual components or substrates of this secretion system. To address this deficiency we set out to investigate the T7SS system of *S. aureus*. The disruption of individual genes in the *ess* [ESX-1 (ESAT-6 system 1) secretion system] locus has identified three secreted proteins. These are EsaC and the two ESAT-6 family members EsxA and EsxB [6,11]. A total of four integral membrane proteins, EssA, EssB, EsaD and the Ftsk-SpoIII type ATPase EssC, have been shown to be necessary for translocation of ESAT-6 proteins across the bacterial envelope [6,12]. In the present study we targeted EssB, a 52 kDa bitopic membrane-bound protein presenting two segments of similar size on either side of the membrane (Figure 1). The protein is conserved among T7SS gene clusters in Firmicutes, but is of unknown function. A total of four expression constructs were prepared (Figure 1), heterologously expressed, the products purified, quaternary structure investigated and crystallization trials carried

Abbreviations used: BAP, biotin-acceptor peptide; BCG, Bacille Calmette–Guérin; BN-PAGE, Blue native PAGE; CV, column volume; DDM, dodecyl maltoside; DOPC, 1,2-dioleoyl-*sn*-glycero-3-phosphocholine; DTT, dithiothreitol; ESAT-6, early secreted antigenic target of 6 kDa; ESI–Q–TOF-MS, electrospray ionization–quadrupole–time-of-flight MS; ESX-1, ESAT-6 system 1; ess, ESX-1 secretion system; IPTG, isopropyl-β-D-thiogalactopyranoside; LB, Luria–Bertani; MALDI–TOF-MS, matrix-assisted laser-desorption ionization–time-of-flight MS; MWCO, molecular-mass cut-off; PEG3350, poly(ethylene glycol) 3350; rmsd, root-mean-square deviation; SAD, single-wavelength anomalous diffraction; SeMet, selenomethionine; SPR, surface plasmon resonance; TEV, tobacco etch virus; T7SS, Type VII secretion system.

[1] To whom correspondence should be addressed (email w.n.hunter@dundee.ac.uk).

The atomic co-ordinates and structure factors have been deposited in the PDB under code 4ANN.

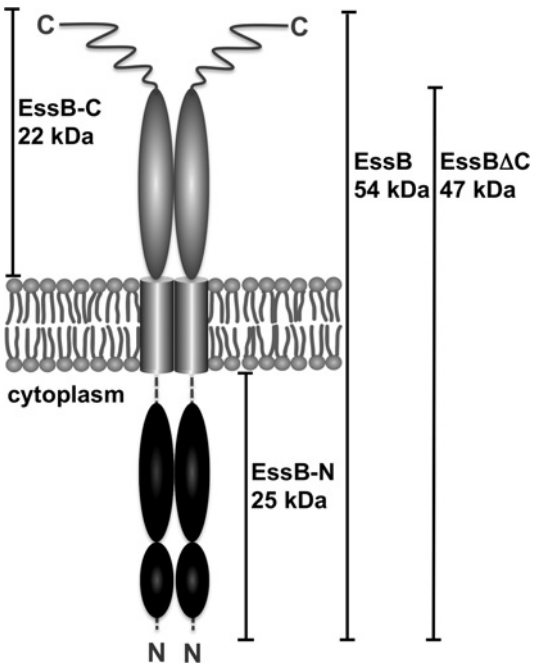

Figure 1 Schematic diagram of EssB predicted topology and domain borders

Summary of constructs used in the present study. EssB is a 52-kDa integral membrane protein with a single predicted transmembrane segment (grey cylinders) spanning Trp^{229}–Ser^{251}. EssBΔC denotes the construct C-terminally truncated by 43 residues, which are predicted to be disordered. EssB-N, the 25-kDa N-terminal cytoplasmic fragment, is predicted to reside in the cytoplasm owing to the absence of a N-terminal signal sequence. Broken lines indicate EssB-N residues not defined by electron density. The 22-kDa C-terminal fragment is predicted to be on the *trans*-side of the cytoplasmic membrane.

out. Of these constructs, three generated crystals, which although of good appearance were poorly ordered. Fortunately, EssB-N, the 25 kDa N-terminal fragment predicted to reside in the cytoplasm, gave well-ordered crystals and the structure was solved using SAD (single-wavelength anomalous diffraction) phasing then refined to 1.05 Å (1 Å = 0.1 nm) resolution. The structural similarity of this EssB-N to serine/threonine protein kinases suggested a functional role in mediating protein–protein interactions. Biophysical and bacterial two-hybrid methods were used to initiate a search for potential binding partners from known components of the T7SS.

EXPERIMENTAL

Expression and purification

The *S. aureus* subsp. *aureus* EssB (UniProt code Q2G185) coding sequence [codon optimized for *E. coli* K12 (Genscript)] was cloned into the SalI/XhoI site of a modified pET27b vector (Novagen) using the forward primer 5′-GCGCGC**GTCGAC**-AAAAAACCATAACCCGAAAAACG-3′ and the reverse primer 5′-GCGCGC**CTCGAG**CTATTTTTTGCGTTCCGCTTCCTG-GCG-3′ (restriction sites are in bold). A gene fragment encoding an EssB construct truncated by 43 C-terminal residues (EssBΔC) was cloned likewise, but using 5′-GCGCGC**CTCGAG**CT-ATTTATCCAGAATATCCTGCAGTTTATCGG-3′ (restriction site is in bold) as the reverse primer. The plasmids produce an N-terminal His_6-tagged protein followed by a TEV (tobacco etch virus) protease cleavage site.

The coding sequence spanning amino acid residues 12–226, the predicted cytoplasmic N-terminal fragment EssB-N, was cloned using the forward primer 5′-GCGCGC**GTCGAC**A-CAGGATATGCTGACCCCGCTGGATG-3′ and the reverse primer 5′-GCGCGC**CTCGAG**CTACACGGTATGGCCCACTT-TGCGCAC-3′ (restriction sites are in bold). The DNA encoding residues 257–444, the C-terminal fragment EssB-C, was cloned using 5′-GCGCGC**GTCGAC**AGAACGCATTGAAAAAGGCT-ATCAGGC-3′ and 5′-GCGCGC**CTCGAG**CTATTTTTTGCGT-TCCGCTTCCTGGCG-3′ as the forward and reverse primers respectively (restriction sites are in bold).

The gene fragments encoding EssB-N and EssB-C were expressed and the recombinant proteins purified using the same protocols. Freshly transformed *E. coli* BL21(DE3) cells were cultivated at 37 °C in 20 ml of LB (Luria–Bertani) medium containing 50 μg/ml kanamycin and 0.2 % glucose, which provided inoculum of the 1 litre main culture (grown at 37 °C in LB medium supplemented with 1 mM $MgCl_2$, 0.5 mM $CaCl_2$ using baffled 5 litre Erlenmeyer flasks). The temperature was lowered to 25 °C when an attenuance of 0.6 at $\lambda = 600$ nm was attained and gene expression was induced with 1 mM IPTG (isopropyl β-D-thiogalactopyranoside). Cells were harvested after 14 h, resuspended in buffer A [50 mM sodium phosphate (pH 8.0), 300 mM NaCl, 1 mM DTT (dithiothreitol) and 10 % (w/v) glycerol, supplemented with one protease inhibitor cocktail tablet (complete, EDTA-free, Roche)] and broken by two passages through a French press (Thermo). The homogenate was centrifuged (150 000 ***g*** for 60 min at 4 °C) and the supernatant was passed through a 0.45-μm filter, supplemented with 25 mM imidazole and then loaded on to a 5 ml CV (column volume) His-Trap HP column (GE Healthcare) previously charged with Ni^{2+}. After a washing step (4 CV) in the same buffer, the samples were eluted applying a linear imidazole gradient (25–250 mM over 18 CV). The buffer was exchanged using a spin concentrator [10 kDa MWCO (molecular-mass cut-off); Sartorius] to buffer B [50 mM Tris/HCl (pH 7.8), 2 mM DTT and 10 % (w/v) glycerol]. After a 14 h incubation step with His-tagged TEV protease (at a molar ratio of 1:20 TEV/recombinant protein) the sample was passed through a HisTrap HP column equilibrated with buffer A to remove the protease, cleaved peptide and non-cleaved material. The cleaved product retains a glycine-alanine-serine sequence at the N-terminus.

The flow through was collected, concentrated and passed through a size-exclusion chromatography column (HR 30/100 GL, Superdex 75, GE Healthcare) equilibrated with buffer C {10 mM Tris/HCl (pH 7.5), 10 mM NaCl and 0.5 mM TCEP [tris-(2-carboxyethyl)phosphine)]}. All size-exclusion chromatography columns were calibrated with molecular mass standards (thyroglobulin, 670 kDa; γ-globulin, 158 kDa; serum albumin, 67 kDa; ovalbumin; 44 kDa, myoglobin, 17 kDa; and vitamin B_{12}, 1 kDa).

The preparation of full-length EssB and EssBΔC followed the same protocol except that *E. coli* LEMO21(DE3) cells [13] were used in medium supplemented with 100 μM L-rhamnose and gene expression was induced with 400 μM IPTG. Membranes were isolated by centrifugation (150 000 ***g*** for 60 min at 4 °C), resuspended in buffer A and solubilized at room temperature (20 °C) with DDM (dodecyl maltoside) at a protein/detergent mass ratio of 1:3 for 2 h. Solid material was removed by centrifugation and the supernatant purified as described above except that all buffers were supplemented with 0.02 % DDM and 1 μg/ml DOPC (1,2-dioleoyl-*sn*-glycero-3-phosphocholine; a gift from Lipoid AG) and for size-exclusion chromatography a Superdex 200 column (GE Healthcare) was used.

An SeMet (selenomethionine) derivative of EssB-N was obtained using the protocol described above with full incorporation achieved by metabolic inhibition [14] and confirmed by

Table 1 Crystallographic statistics

Values in parentheses refer to the highest resolution bin. $R_{merge} = \Sigma h \Sigma i |(h,i) - <I(h)> \Sigma h \Sigma i\, I(h,i)$. $R_{work} = \Sigma hkl||F_o| - |F_c||/\Sigma|F_o|$, where F_o is the observed structure factor amplitude and F_c is the structure–factor amplitude calculated from the model. R_{free} is the same as R_{work} except only calculated using a subset, 5 %, of the data that are not included in any refinement calculations.

Parameter	SeMet	Native
Diffraction data		
Wavelength (Å)	0.9788	0.9814
Resolution (Å)	47.6–1.4 (1.44–1.37)	58.74–1.05 (1.15–1.05)
Unit cell lengths (Å)	38.7, 58.6, 95.3	38.7, 58.7, 95.1
Unique reflections	46296 (6595)	100245 (13554)
Completeness (%)	99.8 (98.8)	99.0 (92.8)
$<I/\sigma(I)>$	46.3 (10.4)	18.0 (3.6)
Redundancy	26.3 (15.0)	7.1 (4.6)
R_{merge}	0.050 (0.236)	0.061 (0.409)
Refinement statistics		
R_{work}/R_{free}		13.8/15.8
Protein residues		176
Waters		376
rmsd from ideal geometry (Å)		
Bond lengths		0.015
Angular distances		0.033
Thermal parameter (B) values (Å^2)		
Wilson B		6.1
Mean B over all atoms/protein/water		18.4/13.8 /25.6
Ramachandran favoured/allowed/outliers (%)		98.4 /1.6/0

MALDI–TOF-MS (matrix-assisted laser-desorption ionization–time-of-flight MS) analysis.

Blue native gel electrophoresis

Linear 4–16 % gradient Native Bis-Tris gels (Novex®, Life Technologies) were run and destained according to the manufacturer's protocol. Samples comprising 3 μg of protein were loaded. Membrane protein samples were supplemented with 0.1 % DDM. Lactate dehydrogenase (146 kDa), BSA (66 kDa), and soya bean trypsin inhibitor (20 kDa) were used as the standard proteins.

MS

ESI–Q–TOF-MS (electrospray ionization–quadrupole–time-of-flight MS) and MALDI–TOF-MS analysis was performed at the University of Dundee 'Fingerprints' Proteomics Facility (Dundee, U.K.) using an Applied Biosystems DE-STR and a Agilent 6520 QTOF spectrometer respectively. Crystals of EssB-N were harvested by centrifugation (18000 ***g*** for 2 min at 20 °C and dissolved in 5 mM Tris/HCl (pH 7.8). Residual PEG3350 [poly(ethylene glycol) 3350] from the mother liquor was removed using a spin concentrator (10 kDa MWCO; Sartorius).

Crystallization and structure determination

Isomorphous crystals of native and SeMet-labelled EssB-N were obtained by hanging-drop vapour diffusion at 20 °C in 3 μl drops of a 1:1 ratio of protein solution [20 mg/ml EssB-N in 10 mM Tris/HCl (pH 7.5), 10 mM NaCl and 0.5 mM TCEP] with the reservoir [0.18 M ammonium citrate and 20 % (w/v) PEG3350]. Orthorhombic blocks (100–300 μm) with space group $P2_12_12_1$ appeared after 2 days. The crystals were transferred to LV-oil (MiTeGen) before flash cooling in a stream of gaseous nitrogen at − 173 °C and then characterized in-house using a Rigaku 007 rotating anode X-ray generator coupled to an RAXIS IV^{++} image plate detector. Suitable samples were stored and subsequently used to measure full datasets on station ID29 at the ESRF (European Synchrotron Radiation Facility; Grenoble, France). Intensity measurements were recorded on an ADSC Q315R Charge Couple Device detector, integrated and processed using XDS [15]. Initial phases were calculated to 1.4 Å by SAD methods in CRANK [16] as implemented in the CCP4 suite of programs [17].

Two selenium sites (out of the three possible) were located and gave an initial figure-of-merit of 0.35, which increased to 0.75 after density modification. An initial model, consisting of residues 31–204, was built using ARP/wARP [18] and gave R_{work}/R_{free} values of 0.178 and 0.211 respectively.

High- and low-resolution data sets, with resolution limits of 1.05 Å and 2.5 Å respectively (Table 1), were collected using different translations of a single native crystal. The datasets were merged using SCALA [19] and the phases determined by molecular replacement with PHASER MR [20] using the SeMet structure as the search model. The refinement was initiated in REFMAC5 [21] and then ported to SHELXL [22]. Several rounds of model-map inspection and re-building in COOT [23] interspersed with least-squares optimization and isotropic thermal parameter refinement gave R_{work} and R_{free} values of 0.217 and 0.246 respectively. The addition of water molecules, hydrogen atoms and inclusion of anisotropic B-factor refinement lowered R_{work}/R_{free} to 0.138 and 0.158 respectively.

Macroscopically different rounded EssB-N crystals were grown using a reservoir condition containing a high phosphate concentration [0.1 M Hepes (pH 7.5), 0.8 M NaH_2PO_4 and 0.8 M KH_2PO_4]. Diffraction data were collected on beam line I03 at the Diamond Light Source (Harwell, U.K.). Processing the data, using XDS and SCALA, indicated that the two crystal forms, although distinct in appearance, are in fact isomorphous. The structure was determined to 1.6 Å (results not shown). A total of three iterative cycles of inspection and manual fitting in COOT and refinement in REFMAC5 gave R_{work}/R_{free} values of 0.167 and 0.199 respectively. Comparisons indicate that this model is essentially the same as that determined at atomic resolution and so no further details are given.

Model geometry and the fit to electron-density maps were monitored with MOLPROBITY [24] and the validation

tools in COOT. Figures were prepared using PyMOL (http://www.pymol.org). The DALI server was used to search the PDB for structural homologues and structural superimpositions were calculated using DALILITE [26]. Multiple sequence alignments were obtained using CLUSTALW2 [27] and edited with ALINE [28]. The conservation of residues at the protein surface was investigated using CONSURF [29]. Potential EssB orthologues were excluded from the analysis if the sequence identities were less than 19 %. Altogether 54 EssB-homologues were used.

Computational analysis of the sequence

A transmembrane segment of EssB was predicted using TMHMM v2.0 [30]. The design of constructs was guided by secondary structure prediction using JPRED [31] and the PHYRE fold-recognition server [32].

Bacterial two-hybrid system

The bacterial two-hybrid screening for EssB-N interactions was performed on MacConkey agar plates containing 1 % maltose, essentially following an established protocol [33]. The *S. aureus* subsp. *aureus* coding sequences for EsxA, EsxB, EsaB and EssB-N [codon optimized for expression in *E. coli* K12 (Genscript)] were cloned into the XhoI/HindIII or BamHI/KpnI sites of a modified pT18 or pT25 plasmid [34] respectively (for primers see Supplementary Table S1 at http://www.biochemj.org/bj/449/bj4490469add.htm). The genes on pT25narJ and pT18narG express proteins that form a heterodimer [34] provided a positive control. *E. coli* strain BTH101 cells were co-transformed with the pT18 and pT25 constructs, plated and cultured for 2 days at 25°C. Colonies were re-streaked on to fresh MacConkey medium and incubated for 36 h at 25°C.

SPR (surface plasmon resonance) spectroscopy

For the SPR experiments EssB-N was immobilized via a C-terminally fused BAP (biotin-acceptor peptide) [35]. Biotin-protein ligase BirA was used for site-specific monobiotinylation of this 15-amino-acid peptide (GLNDIFEAQKIEWHE). The covalently attached biotin binds streptavidin with high affinity [36].

The *E. coli* K12 BirA (P06709) coding sequence stretch was PCR-amplified from genomic copy and cloned into the Nco1/Xho1 site of a modified pET27b vector (Novagen) fusing a C-terminal His_6-tag-coding sequence with the reverse primer. Expression followed the same protocol as described for EssB-N. Harvested cells were resuspended in 50 mM Tris/HCl (pH 7.5), 300 mM NaCl, 1 mM DTT and 10 % (w/v) glycerol and broken by two passages through a French press (Thermo). The homogenate was centrifuged at 150 000 ***g*** for 60 min at 4 °C. The resulting supernatant was passed through a 0.45-μm filter, supplemented with 25 mM imidazole and loaded on to a 5 ml HisTrap HP column (GE Healthcare). After a 4 CV washing step in the same buffer, the recombinant protein was eluted applying a linear imidazole gradient (25–250 mM over 18 CV). The buffer of the late elution fractions was exchanged using a spin concentrator (10 kDa MWCO; Sartorius) to 10 mM Tris/HCl (pH 7.5), 200 mM KCl and 0.5 mM DTT and the protein concentration was adjusted to 10 mg/ml.

EsaB, EsxA and BAP-tagged EssB-N were cloned, expressed and purified essentially as described for EssB-N (for the primers see Supplementary Table S2 at http://www.biochemj.org/bj/449/bj4490469add.htm). BAP-tagged EssB-N (20 μM) was incubated at 37 °C for 2 h in buffer C [10 mM Tris/HCl (pH 7.5), 200 mM KCl, 5 mM $MgCl_2$ and 100 μM d-biotin) containing 500 μM ATP and 1 μM His_6-tagged BirA. The sample was passed through a HisTrap HP column equilibrated with buffer A to remove BirA. The flow through was collected, concentrated and passed through a size-exclusion chromatography column (HR 30/100 GL, Superdex 75 prep grade, GE Healthcare) equilibrated with buffer C to remove free biotin. The incorporation of biotin was monitored by MALDI–TOF-MS.

SPR analysis was performed using a Biacore 3000 instrument (GE Healthcare) using 20 mM Tris/HCl (pH 7.8), 150 mM NaCl, 0.1 mg/ml BSA and 0.005 % Tween 20 as a running buffer. BAP-tagged EssB-N was captured on an immobilized streptavidin surface at a density of 3500 resonance units. Binding of EsxA and EsaB was measured at 25°C and 4°C in the reported concentrations (Supplementary Figure S1 at http://www.biochemj.org/bj/449/bj4490469add.htm). Curves for binding to a blank reference chip were subtracted using the software package Biaevaluation (GE Healthcare).

Data deposition

The atomic co-ordinates and structure factors have been deposited in the PDB under code 4ANN.

RESULTS AND DISCUSSION

Characterization of membrane-bound EssB

Two expression constructs of membrane-bound EssB (Figure 1) were prepared, the encoded proteins characterized and crystallization trials carried out. Full-length EssB was solubilized from the cytoplasmic membrane and purified to homogeneity with a yield of approximately 1–1.5 mg/l of culture. A protein truncated by 43 C-terminal residues, EssBΔC, which retains membrane-binding capacity, was obtained at a similar yield. The deleted residues, from Lys^{401} to the terminal Lys^{444}, form a highly basic segment with a theoretical pI of 9.4. This stretch is highly divergent, indeed absent in some EssB orthologues, as evidenced by a multiple sequence alignment (Supplementary Figure S2 at http://www.biochemj.org/bj/449/bj4490469add.htm). Secondary structure prediction and fold analysis suggested a heptad-repeat on the basis of α-helical propensity from Glu^{421} to Gln^{438} together with disorder in other regions. It appeared sensible therefore to omit the C-terminal tail to improve the chances of crystallization.

The EssBΔC construct has a theoretical mass of 47 kDa, but in gel filtration eluted as a single species with an apparent molecular mass of 198 kDa (Supplementary Figure S3 at http://www.biochemj.org/bj/449/bj4490469add.htm). This suggests a dimeric state taking into consideration the detergent micelle around the transmembrane helices. The average micellar size of the detergent used, DDM, is 50–70 kDa. The 52 kDa full-length EssB elutes with a lower retention volume and an apparent molecular mass of 254 kDa. The shift in retention volume is larger than expected from the mass difference of 5 kDa per monomer. A possible explanation is a change in the overall shape of the protein and thus its hydrodynamic radius, when the potentially flexible C-terminal residues are present. A small peak at 0.38 CV in the chromatogram (Supplementary Figure S3) indicates the presence of aggregated material in the full-length EssB preparation, material that is absent from the EssBΔC chromatogram. The dimeric state of EssB was furthermore corroborated by BN-PAGE (Blue native PAGE) analysis (Supplementary Figure S4 at http://www.biochemj.org/bj/449/bj4490469add.htm).

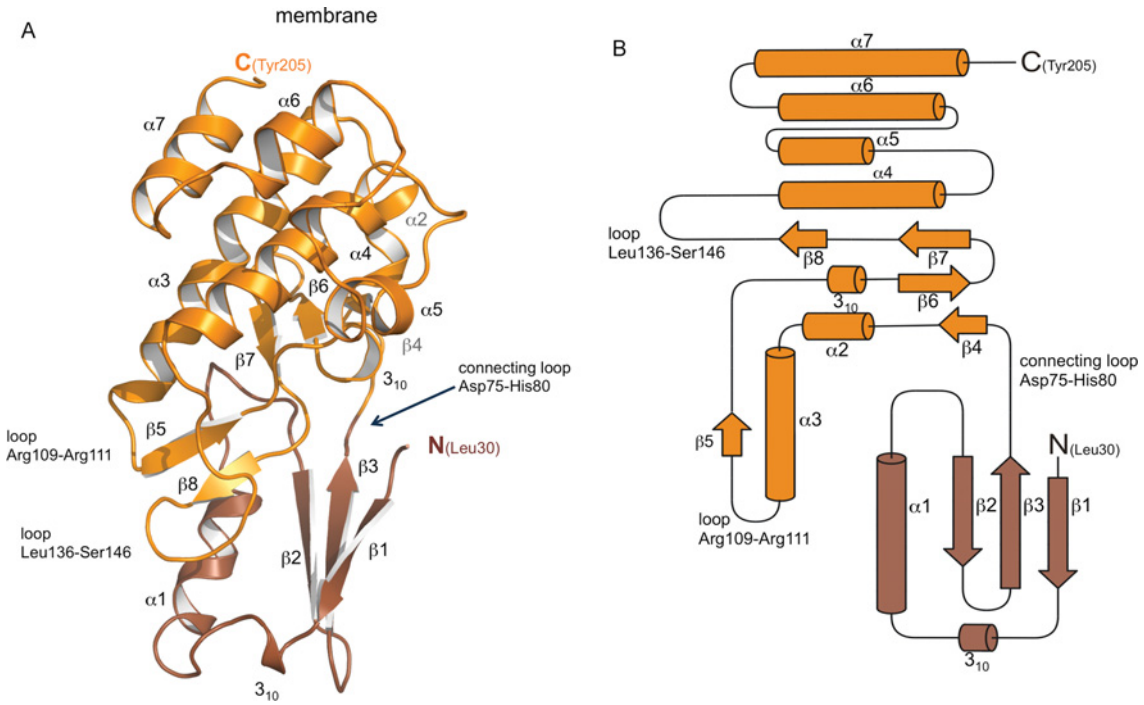

Figure 2 Overall structure of EssB-N

(**A**) Cartoon representation. (**B**) Topology diagram. The N- and C-terminal domains of EssB-N are coloured maroon and orange respectively. Labelled loop regions are discussed in the text.

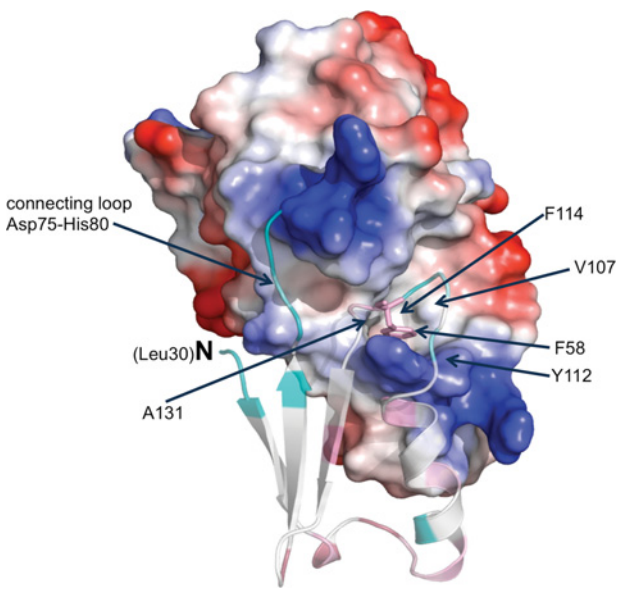

Figure 3 van der Waals surface representation of the EssB-N C-terminal domain

The PyMOL calculated electrostatic potential is mapped on to the EssB-N C-terminal domain. The N-terminal domain is in cartoon representation. Phe[58] is drawn in stick representation. Residues forming a hydrophobic pocket accommodating Phe[58] are labelled.

Extensive crystallization trials were undertaken. EssBΔC formed needles, of length 100 μm, in the presence of DDM, DOPC and DHPC (1,2-diheptanoyl-*sn*-glycero-3-phosphocholine), but no useful diffraction was observed even when tested on a microfocus-beam line (ID23) at the ESRF, merely smears between 20 and 35 Å resolution.

Characterization of soluble EssB fragments

Recombinant proteins corresponding to the N-terminal fragment, EssB-N, predicted to be localized in the cytoplasm and the C-terminal fragment, EssB-C, predicted to reside on the *trans*-side of the membrane, were purified with yields of 3–4 mg/l and 6–7 mg/l of bacterial culture respectively. EssB-N and EssB-C have predicted masses of 25 kDa and 22 kDa respectively. In BN-PAGE analysis the fragments migrate with apparent masses of 58 kDa for EssB-N and 42 kDa for EssB-C (Supplementary Figure S4). This indicates that each forms dimeric species in solution. EssB-N runs as a double band on the BN-PAGE most probably reflecting the 1 kDa C-terminal degradation observed by MS (discussed below).

Although EssB-N eluted at an apparent mass of 54 kDa in analytical gel filtration the observed retention volume for EssB-C was lower than expected for a protein of mass 44 kDa (Supplementary Figure S5 at http://www.biochemj.org/bj/449/bj4490469add.htm). This is in agreement with the mass discrepancy observed when comparing full-length EssB and EssBΔC (see above) and it appears that the highly basic potentially flexible C-terminal segment does indeed influence the protein shape. Expression of a construct encoding a polypeptide truncated at the C-terminus of EssB-C, in a similar fashion to EssBΔC, led only to the formation of inclusion bodies. Rectangular plate crystals of EssB-C, measuring up to 80 μm after optimization, appeared after 3 weeks, but only diffracted to about 7 Å (Supplementary Figure S6 at http://www.biochemj.org/bj/449/bj4490469add.htm) and were not studied further. EssB-N crystallized readily and high-resolution diffraction data were obtained.

Structure of EssB-N

The crystal structure of *S. aureus* EssB-N was solved by SAD methods exploiting a SeMet derivative and produced an

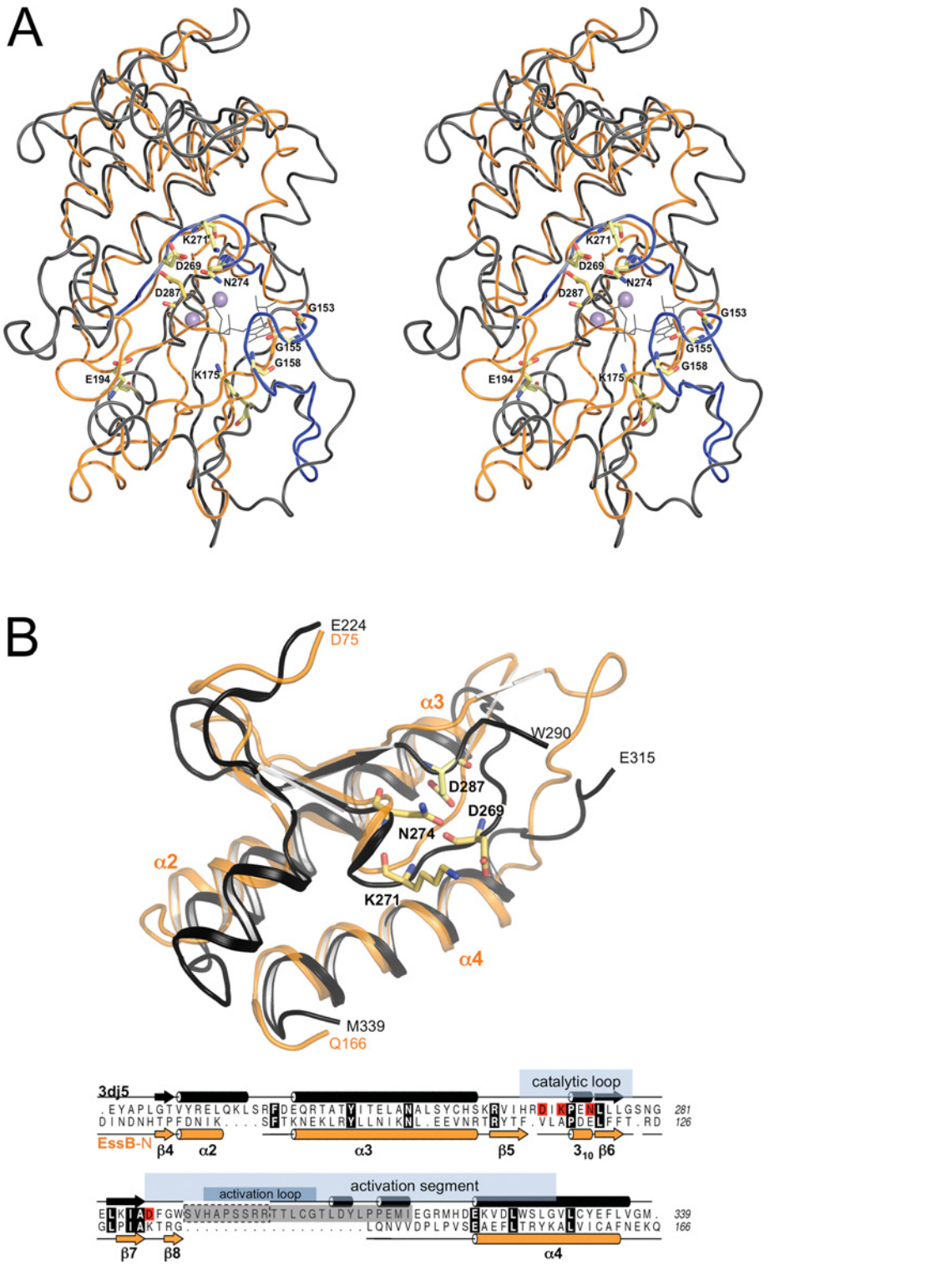

Figure 4 Comparison with kinases

(**A**) Stereo view of rigid body superimposition of EssB-N (orange) and Aurora-A kinase catalytic domain (grey) is depicted in ribbon representation. Prosite serine/threonine protein kinase signature motifs PS00107 ([LIV]-G-{P}-G-{P}-[FYWMGSTNH]-[SGA]-{PW}-[LIVCAT]-{PD}-x-[GSTACLIVMFY]-x(5,18)-[LIVMFYWCSTAR]-[AIVP]-[LIVMFAGCKR]-K) and PS00108 ([LIVMFYC]-x-[HY]-x-D-[LIVMFY]-K-x(2)-N-[LIVMFYCT](3)) are highlighted in black. Highly conserved serine/threonine kinase residues are drawn as sticks. ATP (shown as black lines) and two

atomic-resolution model. The asymmetric unit consists of a monomer and the model comprises 176 residues from Leu[30] to Tyr[205]. All residues display ϕ/ψ combinations within allowed regions of a Ramachandran plot (Table 1). The model is missing 39 residues due to a combination of proteolysis (accounting for about 30 residues) and diffuse electron density at N- and C-termini, which we were unable to interpret. ESI–Q–TOF-MS analysis of the EssB-N sample prior to crystallization revealed a mass difference of 1038 Da (Supplementary Figure S7 at http://www.biochemj.org/bj/449/bj4490469add.htm), which would correspond to nine missing C-terminal residues from Tyr[218] onwards (YVRKVGHTV). EssB-N crystals were subjected to SDS/PAGE and MALDI–TOF-MS analysis after tryptic digest of the excised 24 kDa band. An intact N-terminal fragment was identified and the last C-terminal tryptic fragment covered the sequence from Lys[207] to Ala[217] (KQEQDYSQNYA) (Supplementary Table S3 at http://www.biochemj.org/bj/449/bj4490469add.htm), which is proximal to the peptide identified as missing by the ESI–Q–TOF-MS data. The loss of only nine C-terminal residues, not 39, was thus confirmed, ruling out the possibility that proteolysis occurred during crystallization and indicting therefore that degradation of the C-terminus stretch occurred in the expression host.

The EssB-N structure is cylindrical with a length of approximately 53 Å and diameter of 33 Å, and consists of two domains (Figure 2 and Supplementary Figure S8 at http://www.biochemj.org/bj/449/bj4490469add.htm). The smaller N-terminal domain consists of a three-stranded antiparallel β-sheet with a single turn of 3_{10}-helix and α1, interspersed between β1 and β2. Adjacent to β1 there is diffuse electron density, which may represent an additional N-terminal β-strand, but we were unable to model this satisfactorily. β3 leads into a loop (Asp[75]–His[80]) connecting the domains. The C-terminal domain structure is dominated by a helical bundle formed from two antiparallel helix pairs α3/α4 and α7/α8. The connecting loop leads to β4–α2 and a sequence stretch (Thr[110]–Ser[146]) that forms two short β-strands (β6–β7) that with β4 creates a β-sheet tucked in below the helical bundle. A buried turn of 3_{10}-helix forms an acute bend that changes the direction of the polypeptide from β5 to β6 and supports the alignment of β6 with β4. The model terminates with α7. The amino acid sequence is predicted to form a longer α-helix extending into the outer membrane.

A deep cleft, approximately 30 Å, is formed between the two domains of EssB-N. The cleft is closed off at one side by the interdomain connecting loop (Asp[75]–His[80]). Situated next to this loop, that joins the neighbouring strand β3 of the N-terminal sheet to β4, a further interdomain contact is made by a loop interspersed between α1 and β2. Phe[58] contributes to this structural feature by forming a hydrophobic pocket together with Val[107], Tyr[112], Phe[114] and Ala[131] (Figure 3). A number of conserved residues (Thr[113], Pro[118], Glu[120], Arg[134] and Gly[134]) cluster around the C-terminal side of the cleft (Supplementary Figure S9 at http://www.biochemj.org/bj/449/bj4490469add.htm). This structure is formed by two back-to-back β-sheet motifs between α3 and α4. A two-stranded (β5 and β8) sheet is located at the mouth of the cleft. The 3_{10}-helix, preceding β6, is a key structural element and features the highly conserved residues Pro[118] and Glu[120]. Residues lining the N-terminal domain side of the cleft are less well conserved but this serves to highlight the structural importance of the 3_{10}-helix carrying conserved residues (Ser[38] and Ile[40]) that forms a lining to the cleft.

The core of EssB-N is extremely well-ordered. However, the three loops (Asp[75]–His[80], Arg[109]–Arg[111] and Leu[137]–Ser[145]), which create part of the inter domain cleft, display higher *B*-factor values (26.4 Å^2, 25.1 Å^2 and 21.5 Å^2 respectively) than the overall average (13.8 Å^2). This suggests a greater flexibility of the loops compared with the core of the structure.

EssB-N crystallizes with a single molecule in the asymmetric unit yet forms a stable dimer in solution. Highly ordered crystals with a different morphology were grown in high phosphate concentrations at pH 7.5. We thought it possible that this might represent a different oligomeric state, but the crystals were isomorphous to those obtained in the presence of citrate and gave a virtually identical structural model comprising residues Leu[30]–Gln[204]. In these crystals there are four interfaces formed by EssB-N with symmetry related molecules and these have solvent-accessible surface areas in the range 100–560 Å^2. The surface area of EssB-N is approximately 9890 Å^2 so the maximum interface represents less than 6 % of the monomer surface area. Such a low value is unlikely to be thermodynamically stable in solution and the discrepancy may simply reflect the different conditions under which quaternary structure has been analysed.

The EssB-N structure is reminiscent of a pseudokinase

The structure of EssB-N resembles the bilobal structure typical of protein kinases, i.e. an N-terminal domain or N-lobe with a β-sheet packed against an intersecting α-helix, and a predominantly α-helical C-terminal domain (C-lobe, Figure 2). In serine/threonine protein kinases the cleft formed between the two lobes is occupied by the ATP cofactor and protein substrates. Many protein kinases are regulated by phosphorylation of an activation segment, which produces an extended active state permissive for substrate binding [37]. EssB-N displays closest structural similarity to inactive protein kinases that result from a specific orientation of the kinase activation segment. The most similar orthologue is the catalytic domain, in an inhibited form, of the mammalian serine/threonine kinase Aurora-A (PDB code 3DJ5) with a Z-score 9.4 and rmsd (root-mean-square deviation) of 4.5 Å for the alignment of 170 Cα atoms (Figure 4). Similarity is measured by DALI Z-scores and significant similarities give values greater than 2; such values usually correspond to related folds [38]. The sequence identity is only 12 %, nevertheless these statistics suggest an evolutionary relationship in which a significant degree of structural homology is retained. Comparison with the active conformation of Aurora-A kinase (PDB code 2C6D) gives a lower Z-score of 7.9 and slightly higher rmsd of 4.8 Å compared with the inactive state. The major difference is the orientation of the EssB-N sequence stretch from Lys[132]–Ser[146] (Figure 2A and Supplementary Figure S8) that corresponds to the activation segment of serine/threonine protein kinases (Figure 4). In EssB-N the N-terminal domain is more similar to the conformation observed in inactive protein kinases rather than the extended active conformation.

EssB lacks the protein kinase ATP-binding signature motif situated in the N-lobe (Figure 4A); an observation that is consistent with failure of EssB to bind the ATP-analogue 2′-3′-*O*-(*N*′-methylanthraniloyl) adenosine-5′-*O*-triphosphate

Mn^{2+} ions (lilac spheres) were modelled by superimposition of the ATP-bound structure of protein kinase A (PDB code 1ATP). (**B**) View of the C-terminal side of the cleft, which features the highest degree of structural similarity. A partial structural alignment over 86 residues (EssB Asp[75]–Gln[166] and Aurora-A kinase Glu[224]–M[339]) is shown. Aurora-A kinase invariant active-site residues Asp[269], Lys[271] and Asn[274], and the activation segment residue Asp[287] are in sticks and highlighted in red in the underlying sequence alignment. Aurora-A kinase residues Ser[291]–Glu[315] (grey box in the sequence alignment), including seven residues of the activation loop, which is not resolved in the structure (boxed in a broken line), have been omitted for clarity. Secondary structure elements and known functional serine/threonine protein-kinase segments deduced from [47] are indicated.

(results not shown). EssB also lacks the serine/threonine protein kinase active-site signature, which is located in the C-lobe and includes a catalytic aspartate (Asp^{269} in Aurora-A kinase; Figure 4). However, on the inner surface of the EssB cleft, residues Pro^{118}–Ile^{130}, a remarkable degree of structural similarity is observed (Figure 4). This involves the turn of 3_{10}-helix and the β-sheet motif at the rear of the cleft, a sequence stretch that in the structural alignment corresponds to the localization of the serine/threonine protein kinase active-site signature in Aurora-A. Helices $\alpha 2$, $\alpha 3$ and the central $\alpha 4$ are similar in size and orientation in the two structures and these help to align the two domains with respect to each other. These similarities are reflected in a significantly lower rmsd of 2.9 Å (Z-score 7.5) for a partial structural alignment (Figure 4B) over 86 residues (from Asp^{75}–Gln^{166}) of the C-terminal domain.

The topology of secondary structure elements that facilitate the interdomain contact area is similar and this involves the regions important for polypeptide substrate recognition by protein kinases [39]. Moreover, this part of the EssB-N structure displays the highest degree of sequence conservation in orthologues (Supplementary Figures S2 and S9). The implication is that EssB-N uses the interdomain cleft to bind a polypeptide partner in a similar fashion to serine/threonine kinases binding to substrates. The absence of obvious catalytic residues renders it unlikely that EssB-N exhibits kinase activity itself and it is therefore reminiscent of pseudokinases; a sub-group of the protein kinase family that facilitate protein–protein interactions [40,41]. A single bacterial pseudokinase has been reported, MviN. This multi-domain integral membrane protein forms a complex that controls cell wall synthesis in *M. tuberculosis* upon phosphorylation by the serine/threonine kinase PknB [42]. A structural alignment of 160 Cα positions of the MviN pseudokinase domain (PDB code 3UQC [42]) with EssB-N indicates a structural relationship at a similar level as with active kinases (Z-score 8.2 and rmsd 4.7 Å). This appears sensible since pseudokinases have retained the overall fold, whereas the active-site signature motifs are modified or absent. Eukaryotic pseudokinases participate in various biological processes exerting scaffolding and/or regulatory functions [40] and the structural similarity displayed by EssB-N may reflect a capacity to act as a protein-binding module to provide a scaffold as part of the functional multi-subunit T7SS.

There are also intriguing similarities between EssB and eukaryote-like serine/threonine kinases that occur in bacteria, such as PknB and PknG. *M. tuberculosis* PknB and PknG are localized to the cytoplasmic membrane positioned with the C-terminal lobe of the kinase domain directed into the membrane leading to an extracellular domain [43]. Ligand-induced dimerization is important in these systems and hints that in EssB the fluxional behaviour we observe may be an intrinsic property of the protein and one that could be influenced by the presence of a molecular partner. Such a partner might be substrate for the T7SS or a component of the secretion apparatus itself.

EssB interaction studies

The structure suggested a role in binding partner proteins, and to address the hypothesis that EssB-N is a scaffold or hub of the secretion machine, a bacterial two-hybrid assay was established to test for potential interactions with other proteins encoded within the T7SS gene cluster. The assay depends on reconstitution of adenylate cyclase activity from two non-interacting *Bordetella pertussis* CyaA (bifunctional haemolysin-adenylate cyclase) fragments called T18 and T25 [44,45]. cAMP synthesis triggers the transcriptional activation of maltose and lactose catabolic operons and *E. coli cya* BTH101 colonies fermenting these carbohydrates acidify the medium and produce a red phenotype on MacConkey/maltose indicator plates.

EssB-N was fused to the T18 fragment, and T25 fusions of EsxA, EsxB, EsaB and EsaC were co-expressed (Supplementary Figure S10 at http://www.biochemj.org/bj/449/bj4490469add.htm). Colonies remained white except for EsxA and EsaB, which adopted a faint red colour with a later onset compared with the positive control (Supplementary Figure S10). SPR methods were used to test if purified EsxA, or EsaB, a small soluble protein that appears to regulate the production of the T7SS substrate protein EsaC [11], bind to immobilized EssB-N (Supplementary Figure S1). However, no binding was observed. In fact significant negative responses were obtained after subtracting responses from the reference flow cell due to binding to the reference. These results hint to a non-specific interaction, which is conceivable given the highly polar surface of the EsxA dimer [46] and this would classify the weak signals observed in the bacterial two-hybrid system as false positives.

Numerous factors can complicate detection of protein–protein interactions in a multi-component secretion apparatus. For example, more than two binding partners might be necessary to stabilize protein–protein interactions. A further issue is that the expression levels of membrane-bound proteins may be insufficient for the two-hybrid assay. Also, that isolated fragments might in some cases poorly reflect their protein–protein interaction characteristics, especially if transmembrane domains participate in binding. Furthermore *in vitro* protein–protein interaction studies might be hampered by the presence of a detergent that poorly mimics a lipidic environment. In future studies homologous expression in *Staphylococcus* mutants should allow for the isolation of T7SS protein complexes to shed light on the architecture of the entire secretion system.

Concluding remarks

The monotopic membrane protein EssB from *S. aureus* and constructs encoding the isolated soluble domains have been expressed in recombinant form, the products purified to homogeneity and their quaternary structure determined to be dimeric. The cytoplasmic fragment EssB-N yielded highly ordered crystals and the structure was determined to atomic resolution.

EssB-N exhibits an intriguing structural relationship to serine/threonine protein kinases. The lack of conserved catalytic residues rules out EssB-N as an enzyme, but rather suggests a function as a protein–protein interaction module using the stable modular entity of the protein kinase fold. In effect EssB resembles a membrane inset pseudokinase. The biological role may be as an adaptor protein; to recruit or position other components of the secretion apparatus at the cytoplasmic membrane and in so doing form the functional machine. Future studies in this area will now be required to elucidate the precise contribution of this protein to the T7SS and the virulence of an important human pathogen.

AUTHOR CONTRIBUTION

Tracy Palmer and William Hunter conceived the project. All aspects of sample preparation, characterization and crystallization were carried out by Martin Zoltner. Crystallographic studies involved Martin Zoltner, Paul Fyfe and William Hunter. Martin Zoltner and William Hunter were primarily responsible for interpretation of the results. Martin Zoltner contributed most to preparation of the paper with help from all other authors.

ACKNOWLEDGEMENTS

We thank the DLS (Diamond Light Source) and ESRF (European Synchrotron Radiation Facility) for synchrotron beam time and excellent support, and Mark Agacan, Grant Buchanan and Sarah Coulthurst for assistance and materials.

FUNDING

This work was supported by the Biotechnology and Biological Sciences Research Council [grant number H007571], the Medical Research Council [grant number G117/519] and the Wellcome Trust [grant numbers 082596, 083481 and 094090].

REFERENCES

1 Abdallah, A. M., Gey van Pittius, N. C., Champion, P. A. D., Cox, J., Luirink, J., Vandenbroucke-Grauls, C. M. J. E., Appelmelk, B. J. and Bitter, W. (2007) Type VII secretion: mycobacteria show the way. Nat. Rev. Microbiol. **5**, 883–891
2 Pallen, M. J. (2002) The ESAT-6/WXG100 superfamily: and a new Gram-positive secretion system? Trends Microbiol. **10**, 209–212
3 Champion, P. A. D., Stanley, S. A., Champion, M. M., Brown, E. J. and Cox, J. S. (2006) C-terminal signal sequence promotes virulence factor secretion in *Mycobacterium tuberculosis*. Science **313**, 1632–1636
4 Begg, K. J., Dewar, S. J. and Donachie, W. D. (1995) A new *Escherichia coli* cell division gene, ftsK. J. Bacteriol. **177**, 6211–6222
5 Wu, L. J. and Errington, J. (1994) *Bacillus subtilis* SpoIIIE protein required for DNA segregation during asymmetric cell division. Science **264**, 572–575
6 Burts, M. L., Williams, W. A., DeBord, K. and Missiakas, D. M. (2005) EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus aureus* infections. Proc. Natl. Acad. Sci. U.S.A. **102**, 1169–1174
7 Mahairas, G. G., Sabo, P. J., Hickey, M. J., Singh, D. C. and Stover, C. K. (1996) Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. J. Bacteriol. **178**, 1274–1282
8 Lewis, K. N., Liao, R., Guinn, K. M., Hickey, M. J., Smith, S., Behr, M. A. and Sherman, D. R. (2003) Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette–Guérin attenuation. J. Infect. Dis. **187**, 117–123
9 Majlessi, L., Brodin, P., Brosch, R., Rojas, M.-J., Khun, H., Huerre, M., Cole, S. T. and Leclerc, C. (2005) Influence of ESAT-6 secretion system 1 (RD1) of *Mycobacterium tuberculosis* on the interaction between mycobacteria and the host immune system. J. Immunol. **174**, 3570–3579
10 Lowy, F. D. (1998) *Staphylococcus aureus* infections. New Engl. J. Med. **339**, 520–532
11 Burts, M. L., DeDent, A. C. and Missiakas, D. M. (2008) EsaC substrate for the ESAT-6 secretion pathway and its role in persistent infections of *Staphylococcus aureus*. Mol. Microbiol. **69**, 736–746
12 Anderson, M., Chen, Y. H., Butler, E. K. and Missiakas, D. M. (2011) EsaD, a secretion factor for the Ess pathway in *Staphylococcus aureus*. J. Bacteriol. **193**, 1583–1589
13 Wagner, S., Klepsch, M. M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Högbom, M., van Wijk, K. J., Slotboom, D. J., Persson, J. O. and de Gier, J.-W. (2008) Tuning *Escherichia coli* for membrane protein overexpression. Proc. Natl. Acad. Sci. U.S.A. **105**, 14371–14376
14 Doublie, S. (2007) Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems. Methods Mol. Biol. **363**, 91–108
15 Kabsch, W. (2010) XDS. Acta Crystallogr. Sect. D Biol. Crystallogr. **66**, 125–132
16 Ness, S., de Graaff, R. and Abrahams, J. (2004) Crank: new methods for automated macromolecular crystal structure solution. Structure **12**, 1753–1761
17 Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr. Sect. D Biol. Crystallogr. **50**, 760–763
18 Perrakis, A. and Morris, R. (1999) Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. **6**, 458–463
19 Evans, P. (2006) Scaling and assessment of data quality. Acta Crystallogr. Sect. D Biol. Crystallogr. **62**, 72–82
20 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) Phaser crystallographic software. J. Appl. Crystallogr. **40**, 658–674
21 Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. Sect. D Biol. Crystallogr. **67**, 355–367
22 Sheldrick, G. M. and Schneider, T. R. (1997) SHELXL: high-resolution refinement. Methods Enzymol. **277**, 319–343
23 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr. Sect. D Biol. Crystallogr. **60**, 2126–2132
24 Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. and Richardson, D. C. (2003) Structure validation by Cα geometry: phi, psi and Cβ deviation. Proteins **50**, 437–450
25 Reference deleted
26 Holm, L. and Park, J. (2000) DaliLite workbench for protein structure comparison. Bioinformatics **16**, 566–567
27 Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R. et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics **23**, 2947–2948
28 Bond, C. (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. Acta Crystallogr. Sect. D Biol. Crystallogr. **65**, 510–512
29 Landau, M., Mayrose, I. and Rosenberg, Y. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res. **33**, 299–302
30 Krogh, A., Larsson, B., Heijne, von, G. and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305**, 567–580
31 Cole, C., Barber, J. D. and Barton, G. J. (2008) The Jpred 3 secondary structure prediction server. Nucleic Acids Res. **36**, 197–201
32 Kelley, L. A. and Sternberg, M. J. E. (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. **4**, 363–371
33 Handford, J. I., Ize, B., Buchanan, G., Butland, G. P., Greenblatt, J., Emili, A. and Palmer, T. (2009) Conserved network of proteins essential for bacterial viability. J. Bacteriol. **191**, 4732–4749
34 Karimova, G., Pidoux, J., Ullmann, A. and Ladant, D. (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proc. Natl. Acad. Sci. U.S.A. **95**, 5752–5756
35 Beckett, D., Kovaleva, E. and Schatz, P. J. (1999) A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. Protein Sci. **8**, 921–929
36 Weber, P. C., Ohlendorf, D. D., Wendolowski, J. J. and Salemme, F. R. (1989) Structural origins of high affinity biotin binding to streptavidin. Science **243**, 85–88
37 Rajakulendran, T. and Sicheri, F. (2010) Allosteric protein kinase regulation by pseudokinases: insights from STRAD. Sci. Signaling **3**, pe8
38 Holm, L., Kääriäinen, S., Rosenström, P. and Schenkel, A. (2008) Searching protein structure databases with DaliLite v.3. Bioinformatics **24**, 2780–2781
39 Knighton, D., Zheng, J., Eyck, Ten, L., Ashford, V., Xuong, N., Taylor, S. and Sowadski, J. (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. Science **253**, 407–414
40 Boudeau, J., Miranda-Saavedra, D. and Barton, G. (2006) Emerging roles of pseudokinases. Trends Cell Biol. **16**, 443–452
41 Scheeff, E. D., Eswaran, J., Bunkoczi, G., Knapp, S. and Manning, G. (2009) Structure of the pseudokinase VRK3 reveals a degraded catalytic site, a highly conserved kinase fold, and a putative regulatory binding site. Structure **17**, 128–138
42 Gee, C. L., Papavinasasundaram, K. G., Blair, S. R., Baer, C. E., Falick, A. M., King, D. S., Griffin, J. E., Venghatakrishnan, H., Zukauskas, A., Wei, J. R. et al. (2012) A phosphorylated pseudokinase complex controls cell wall synthesis in *Mycobacteria*. Sci. Signaling **5**, ra7
43 Pereira, S. F. F., Goss, L. and Dworkin, J. (2011) Eukaryote-like serine/threonine kinases and phosphatases in bacteria. Microbiol. Mol. Biol. Rev. **75**, 192–212
44 Karimova, G. and Ullmann, A. (2001) Protein–protein interaction between *Bacillus stearothermophilus* tyrosyl-tRNA synthetase subdomains revealed by a bacterial two-hybrid system. J. Mol. Microbiol. Biotechnol. **3**, 73–82
45 Ize, B., Coulthurst, S. J., Hatzixanthis, K., Caldelari, I., Buchanan, G., Barclay, E. C., Richardson, D. J., Palmer, T. and Sargent, F. (2009) Remnant signal peptides on non-exported enzymes: implications for the evolution of prokaryotic respiratory chains. Microbiology **155**, 3992–4004
46 Sundaramoorthy, R., Fyfe, P. K. and Hunter, W. N. (2008) Structure of *Staphylococcus aureus* EsxA suggests a contribution to virulence by action as a transport chaperone and/or adaptor protein. J. Mol. Biol. **383**, 603–614
47 Pike, A. C. W., Rellos, P., Niesen, F. H., Turnbull, A., Oliver, A. W., Parker, S. A., Turk, B. E., Pearl, L. H. and Knapp, S. (2008) Activation segment dimerization: a mechanism for kinase autophosphorylation of non-consensus sites. EMBO J. **27**, 704–714

Received 30 July 2012/1 October 2012; accepted 25 October 2012
Published as BJ Immediate Publication 25 October 2012, doi:10.1042/BJ20121209

Biochem. J. (2013) **449**, 469–477 (Printed in Great Britain) doi:10.1042/BJ20121209

SUPPLEMENTARY ONLINE DATA

Characterization of *Staphylococcus aureus* EssB, an integral membrane component of the Type VII secretion system: atomic resolution crystal structure of the cytoplasmic segment

Martin ZOLTNER, Paul K. FYFE, Tracy PALMER and William N. HUNTER[1]

College of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, U.K.

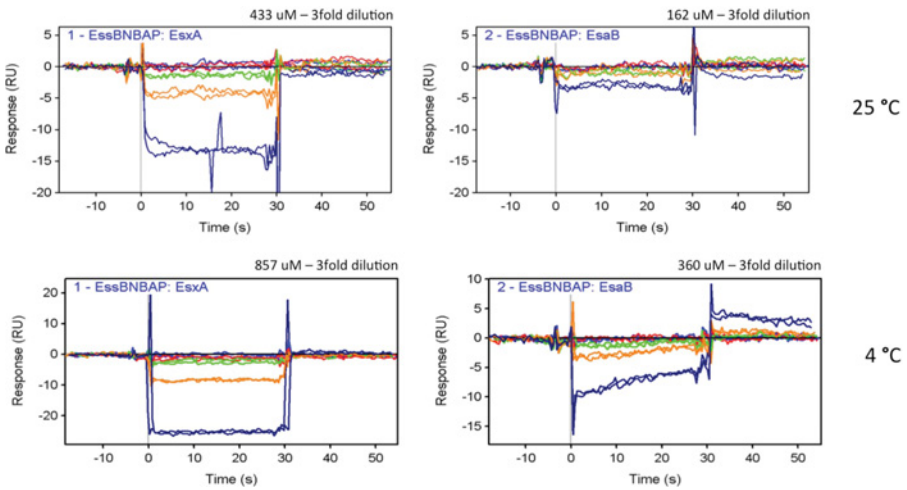

Figure S1 SPR analysis of potential EssB-N interactions

BAP-tagged EssB-N was captured on a streptavidin chip (see the Experimental section of the main text). Binding of EsxA (left-hand panels) and EsaB (right-hand panels) was measured at 25 °C (upper panels) and 4 °C (lower panels). The curves correspond to a dilution series starting from the concentration indicated (dark blue curve). Shown are the binding curves after subtracting binding to the empty reference chip. The negative response signals obtained suggest non-specific binding (see the Results and discussion section of the main text).

[1] To whom correspondence should be addressed (email w.n.hunter@dundee.ac.uk).

The atomic co-ordinates and structure factors have been deposited in the PDB under code 4ANN.

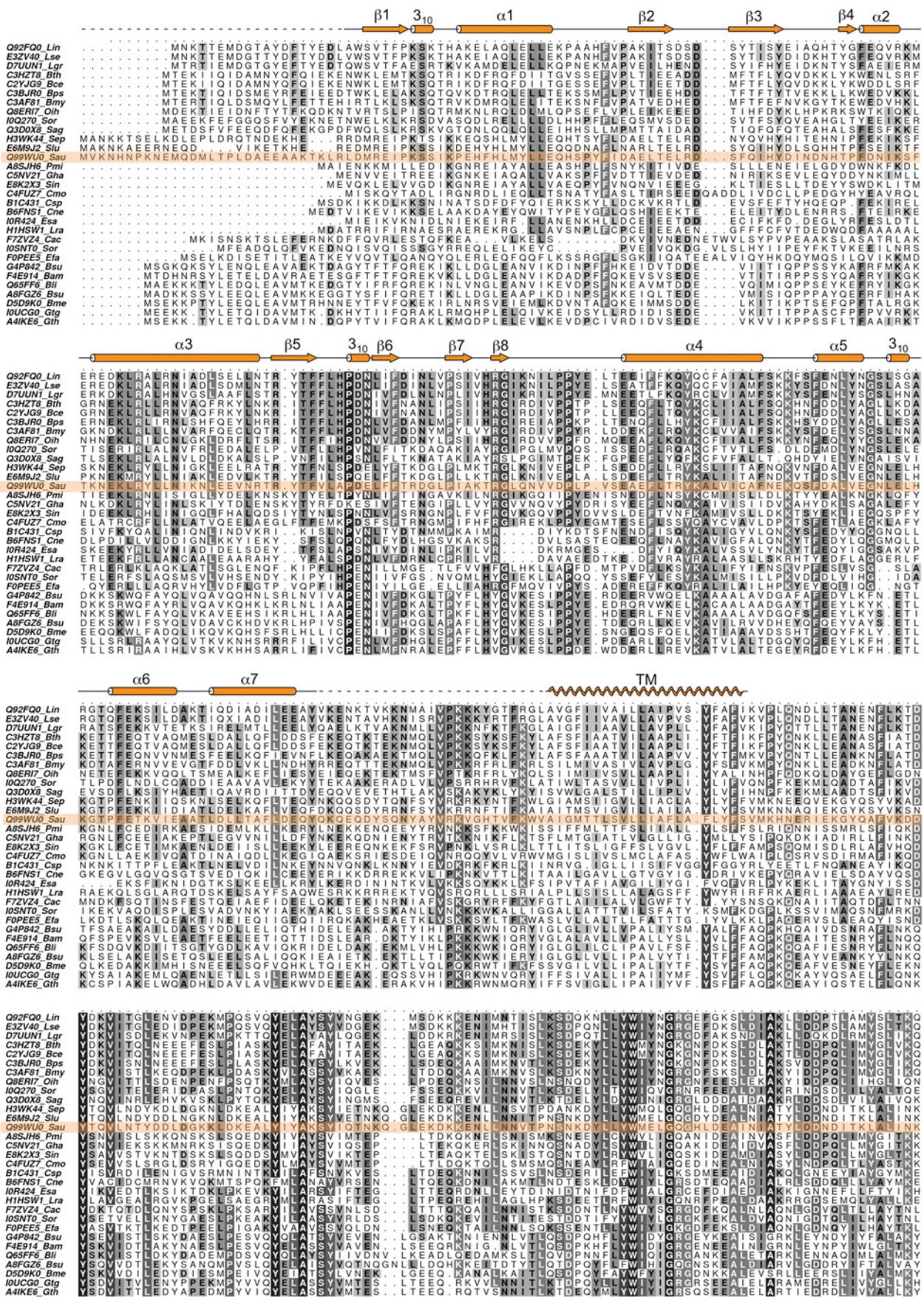

Figure S2 Multiple sequence alignment of EssB-N orthologues

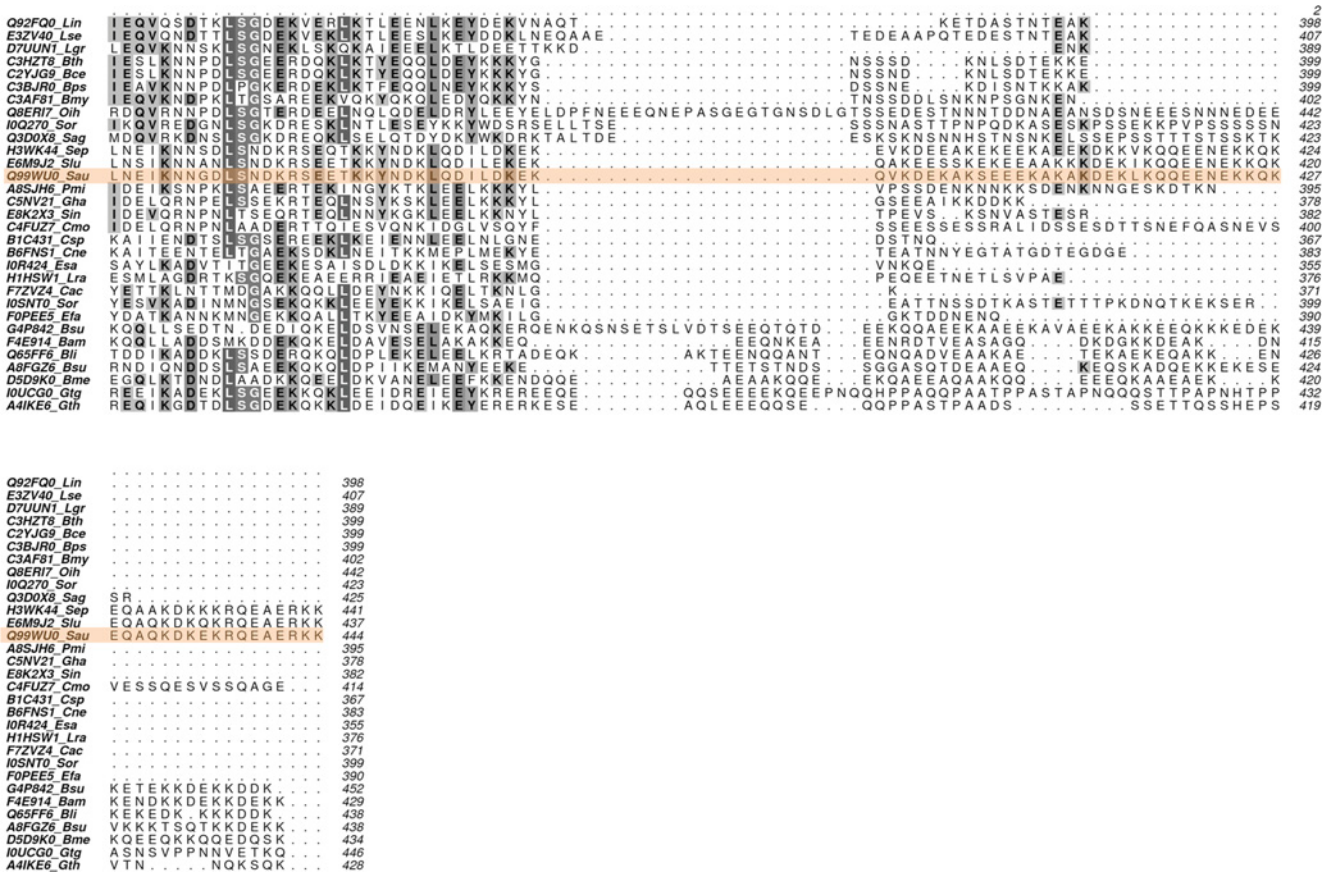

Figure S2 Multiple sequence alignment of EssB-N orthologues (cont.)

EssB-N secondary structure elements and the predicted transmembrane span (TM) are indicated. Selected EssB-N sequences are titled with a UniProt identifier and species abbreviation. Lin, *Listeria innocua*; Lse, *Listeria seeligeri*; Lgr, *Listeria grayi*; Bth, *Bacillus thuringiensis*; Bce, *Bacillus cereus*; Bps, *Bacillus pseudomycoides*; Bmy, *Bacillus mycoides*; Oih, *Oceanobacillus iheyensis*; Sor, *Streptococcus oralis*; Sag, *Streptococcus agalactiae*; Sep, *Staphylococcus epidermis*; Slu, *Staphylococcus lugdunensis*; Sau, *Staphylococcus aureus*; Pmi, *Parvimonas micra*; Gha, *Gemella haemolysans*; Sin, *Streptococcus infantis*; Cmo, *Catonella morbi*; Csp, *Clostridium spiroforme*; Cne, *Clostridium nexile*; Esa, *Eubacterium saburreum*; Lra, *Lachnospiraceae bacterium*; Cac, *Clostridium acetobutylicum*; Efa, *Enterococcus faecalis*; Bsu, *Bacillus subtilis*; Bam, *Bacillus amyloliquefaciens*; Bli, *Bacillus licheniformis*; Bme, *Bacillus megaterium*; Gtg, *Geobacillus thermoglucosidans*; and Gth, *Geobacillus thermodenitrificans*.

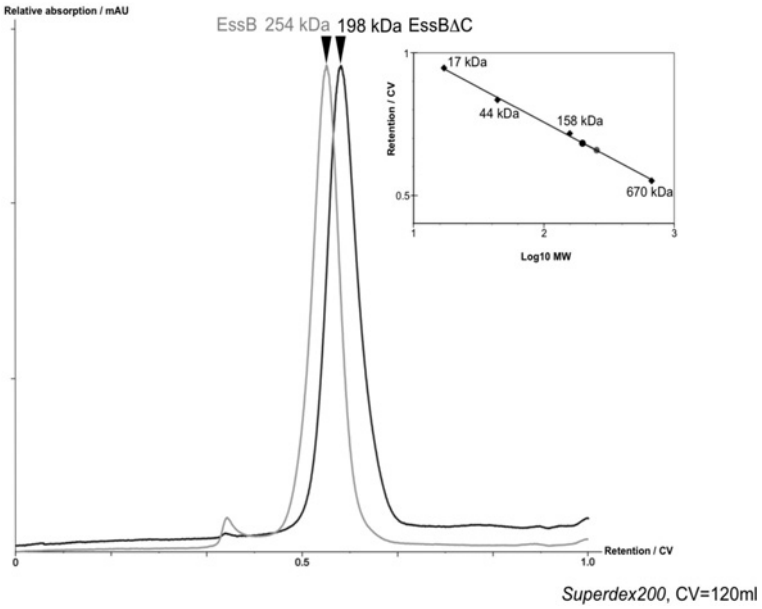

Figure S3 Analytical gel filtration of detergent-solubilized EssB

Overlay of size-exclusion chromatography chromatograms for EssB and truncated EssBΔC. Protein (5 mg) was loaded on to a CV of 120 ml Superdex 200 (prep grade; GE Healthcare) column equilibrated with 10 mM sodium phosphate, 50 mM NaCl, 1 mM DTT, 1 μg/ml DOPC and 0.02 % DDM. Peaks are labelled with apparent molecular masses calculated from a previously determined calibration curve (inset; see the Experimental section of the main text).

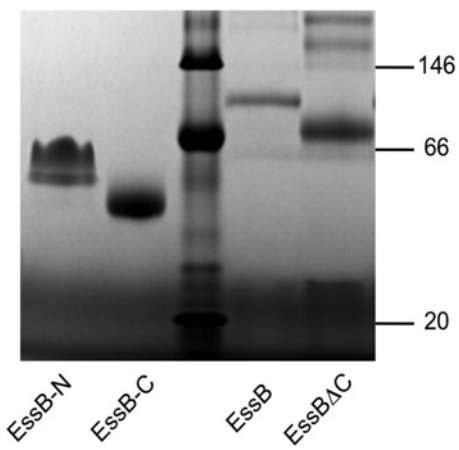

Figure S4 BN-PAGE analysis of the EssB constructs

BN-PAGE of purified EssB-N (far-left lane), EssB-C (near-left lane), EssB (near-right lane) and EssBΔC (far-right lane). All constructs are migrating as dimers with apparent molecular masses of 58 kDa, 42 kDa, 98 kDa and 72 kDa, derived by comparison with the molecular mass standards (centre lane, masses given in kDa on the right-hand side; see the Experimental section of the main text). For EssB-N an additional band at 51 kDa is visible that corresponds to the degradation product (see the Results and discussion section of the main text).

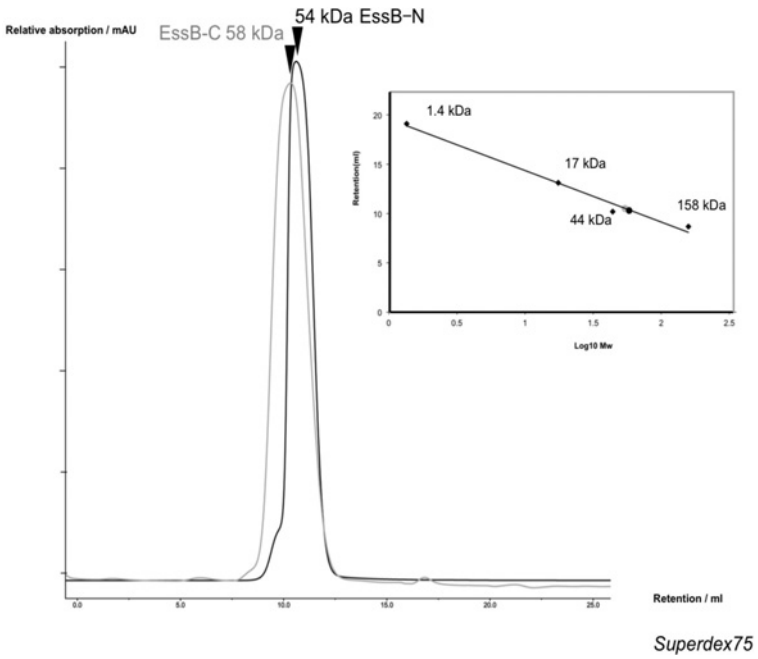

Figure S5 Analytical gel filtration of EssB isolated soluble domains

Overlay of size-exclusion chromatography chromatograms for EssB-N and EssB-C. Protein (8 mg) was loaded on a CV of 24 ml Superdex 75 (prep grade, GE Healthcare) column equilibrated with 10 mM sodium phosphate, 50 mM NaCl and 1 mM DTT. Peaks are labelled with apparent molecular masses calculated from the calibration curve (inset; see the Experimental section of the main text).

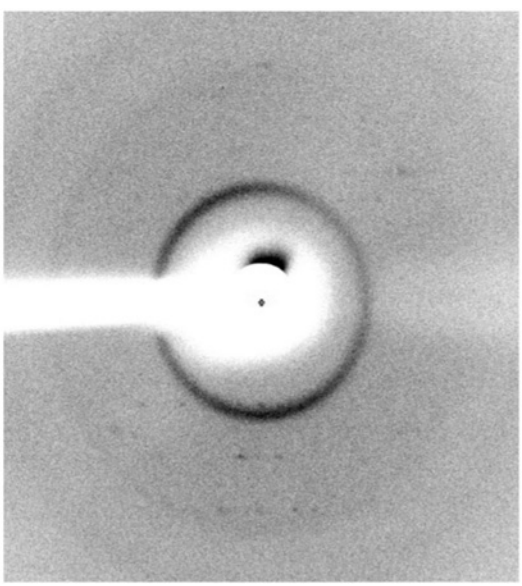

Figure S6 Diffraction pattern from EssB-C crystals

An example diffraction image of EssB-C crystals recorded using in-house facilities (see the Experimental section of the main text).

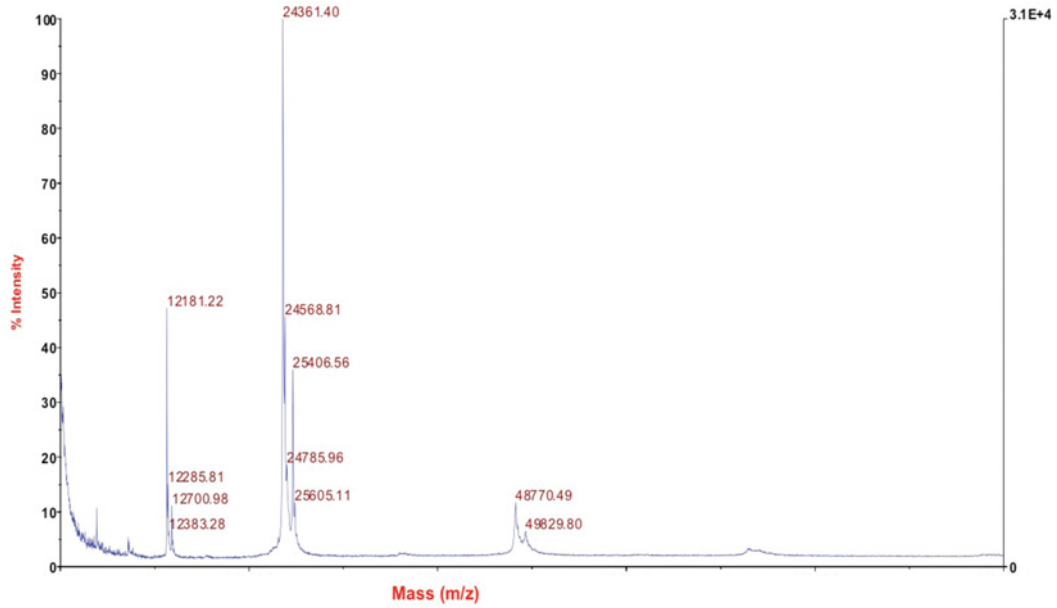

Figure S7 ESI–Q–TOF-MS analysis of purified EssB-N

ESI–Q–TOF-MS analysis of the EssB-N sample prior to crystallization (see the Results and discussion section of the main text).

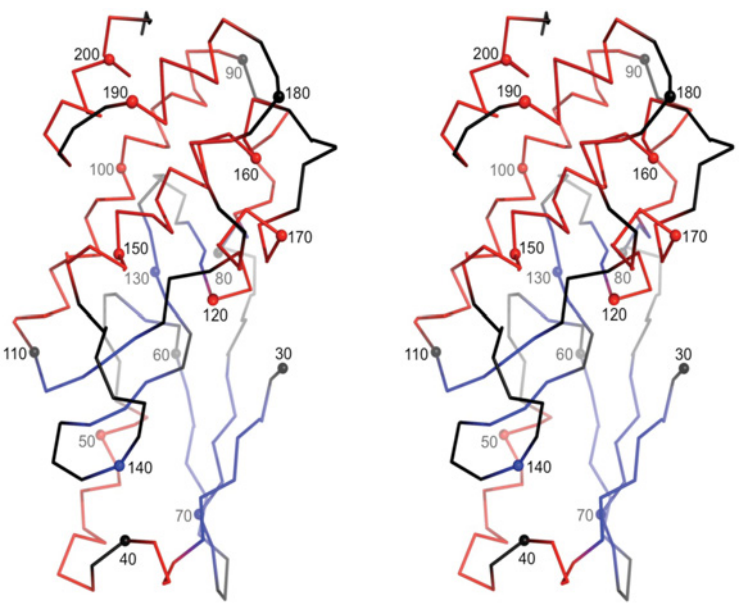

Figure S8 EssB-N structure, stereo view of the Cα trace

The chain is coloured by secondary structure elements (black, loop; red, helix; and blue, strand) and residue numbers are indicated at every tenth Cα (shown as spheres).

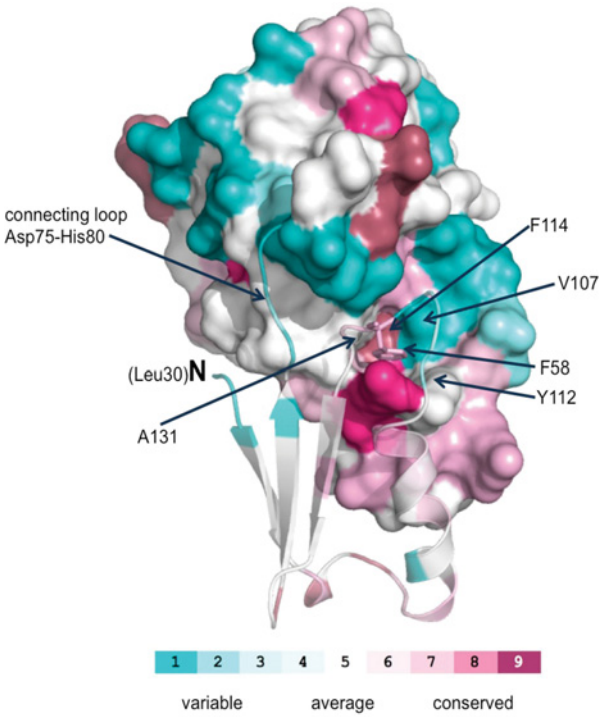

Figure S9 Conservation mapping on to the EssB-N C-terminal domain surface

The continuous conservation scores are normalized and partitioned into nine discrete conservation bins (darkest pink indicates the most conserved), reflecting the relative degree of conservation of each amino acid position (see the Experimental section of the main text).

Table S1 Primers used to generate samples for the bacterial two-hybrid experiments

Restriction sites are in bold.

Construct	Primers (forward/reverse)
pT18EssB-N	5′-GCGC**CTCGAG**GAAAAACCATAACCCGAAAAACGAAATG-3′
	5′-GCGC**AAGCTT**GCCACGGTATGGCCCACTTTGCGCAC-3′
pT25EsxA	5′-GCGC**GGATCC**CGCGATGATTAAAATGAGCCCGGAAG-3′
	5′-GCGC**GGTACC**TTACTGCAGGCCAAAGTTGTTGCTCAGC-3′
pT25EsxB	5′- GCGC**GGATCC**CGGCGGCTATAAAGGCATTAAAGC-3′
	5′-GCGC**GGTACC**TTACGGGTTCACGCGATCCAGGCCCTG-3′
pT25EsaB	5′-GCGC**GGATCC**CAATCAGCACGTAAAAGTAAC-3′
	5′-GCGC**GGTACC**TTACAGCAGTTTCAGAATATCGCCATC-3′
pT25EsaC	5′-GCGC**GGATCC**CAACTTTAACGATATTGAAACGATG-3′
	5′-GCGC**GGTACC**TTAGTTCATCGCTTTGTTAAAATATTCGCTCG-3′

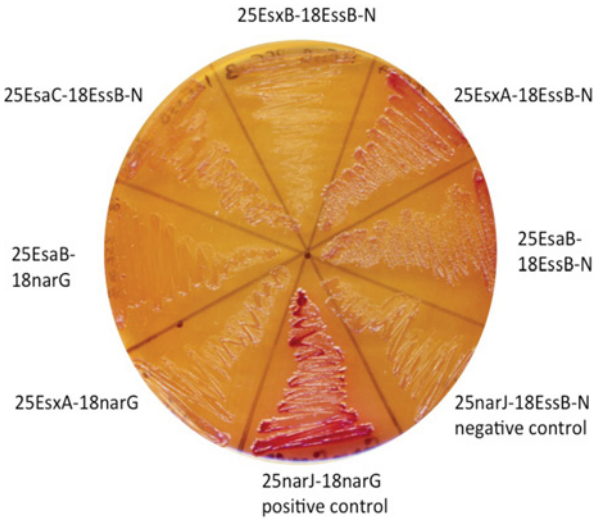

Figure S10 Bacterial two-hybrid assay

E. coli BTH101 cells were co-transformed with pT25 constructs encoding EsxA, EsxB, EsaB and EsaC fusions in combination with pT18EssB or either pT25narJ or pT18narG as a negative control. The combination pT25narJ/pT18narG, producing a stable dimer, served as a positive control. Initial colonies were re-streaked on to fresh MacConkey/maltose plates and incubated for 36 h at 25 °C.

Table S2 Primers used to generate samples for the SPR experiments

Restriction sites are in bold.

Construct	Primers (forward/reverse)
pT18EssB-N	5′-GCGC**CTCGAG**GAAAAACCATAACCCGAAAAACGAAATG-3′
	5′-GCGC**AAGCTT**GCCACGGTATGGCCCACTTTGCGCAC-3′
pT25EsxA	5′-GCGC**GGATCC**CGCGATGATTAAAATGAGCCCGGAAG-3′
	5′-GCGC**GGTACC**TTACTGCAGGCCAAAGTTGTTGCTCAGC-3′
pT25EsxB	5′-GCGC**GGATCC**CGGCGGCTATAAAGGCATTAAAGC-3′
	5′-GCGC**GGTACC**TTACGGGTTCACGCGATCCAGGCCCTG-3′
pT25EsaB	5′-GCGC**GGATCC**CAATCAGCACGTAAAAGTAAC-3′
	5′-GCGC**GGTACC**TTACAGCAGTTTCAGAATATCGCCATC-3′
pT25EsaC	5′-GCGC**GGATCC**CAACTTTAACGATATTGAAACGATG-3′
	5′-GCGC**GGTACC**TTAGTTCATCGCTTTGTTAAAATATTCGCTCG-3′

Table S3 Tryptic fragments identified from analysis of EssB-N crystals

MALDI–TOF-MS analysis of EssB-N after tryptic digest (for details see the Experimental section of the main text). The N- and C-terminal peptides are highlighted in blue and red respectively. M_r(calc), calculated M_r; M_r(expt), expected M_r.

Query	Observed	M_r(expt)	M_r(calc)	p.p.m.	Miss	Score	Expect	Rank	Peptide
367	582.810440	1163.606328	1163.600861	4.70	0	75	3.4e – 08	1	K.ALVICAFNEK.Q
368	582.810440	1163.606328	1163.600861	4.70	0	(56)	2.5e – 06	1	K.ALVICAFNEK.Q
369	582.810440	1163.606328	1163.600861	4.70	0	(37)	0.00018	1	K.ALVICAFNEK.Q
370	582.810440	1163.606328	1163.600861	4.70	0	(44)	4.5e – 05	1	K.ALVICAFNEK.Q
373	583.309937	1164.605322	1163.600861	863	0	(9)	0.12	1	K.ALVICAFNEK.Q
415	614.743591	1227.472630	1227.468002	3.77	0	48	1.5e – 05	1	K.QEQDYSQNYA.- + Gln->pyro-Glu (N-term)
416	614.743591	1227.472630	1227.468002	3.77	0	(28)	0.0016	1	K.QEQDYSQNYA.- + Gln->pyro-Glu (N-term)
420	623.257599	1244.500646	1244.494537	4.91	0	(26)	0.0027	1	K.QEQDYSQNYA.-
421	623.257599	1244.500646	1244.494537	4.91	0	(33)	0.00046	1	K.QEQDYSQNYA.-
422	623.257599	1244.500646	1244.494537	4.91	0	(21)	0.0086	1	K.QEQDYSQNYA.-
423	623.257599	1244.500646	1244.494537	4.91	0	(20)	0.0095	1	K.QEQDYSQNYA.-
425	623.752838	1245.491124	1244.494537	801	0	(10)	0.096	1	K.QEQDYSQNYA.-
426	623.752838	1245.491124	1244.494537	801	0	(12)	0.058	1	K.QEQDYSQNYA.-
607	841.926209	1681.837866	1681.831146	4.00	0	(51)	7.6e – 06	1	K.QSFDALVEGNLELHK.G + glycine-> pyroglutamate (N-terminus)
608	841.926209	1681.837866	1681.831146	4.00	0	(70)	1e – 07	1	K.QSFDALVEGNLELHK.G + glycine-> pyroglutamate (N-terminus)
609	841.926209	1681.837866	1681.831146	4.00	0	(28)	0.0016	1	K.QSFDALVEGNLELHK.G + glycine-> pyroglutamate (N-terminus)
615	850.441223	1698.867894	1698.857681	6.01	0	75	2.9e – 08	1	K.QSFDALVEGNLELHK.G
616	850.441223	1698.867894	1698.857681	6.01	0	(22)	0.0057	1	K.QSFDALVEGNLELHK.G
617	850.441223	1698.867894	1698.857681	6.01	0	(47)	1.9e – 05	1	K.QSFDALVEGNLELHK.G
618	850.441223	1698.867894	1698.857681	6.01	0	(17)	0.022	1	K.QSFDALVEGNLELHK.G
333	567.298391	1698.873345	1698.857681	9.22	0	(12)	0.064	1	K.QSFDALVEGNLELHK.G
334	567.298391	1698.873345	1698.857681	9.22	0	(9)	0.13	1	K.QSFDALVEGNLELHK.G
335	567.298391	1698.873345	1698.857681	9.22	0	(11)	0.079	1	K.QSFDALVEGNLELHK.G
336	567.298391	1698.873345	1698.857681	9.22	0	(26)	0.0023	1	K.QSFDALVEGNLELHK.G
619	567.298391	1698.873345	1698.857681	9.22	0	(54)	4.1e – 06	1	K.QSFDALVEGNLELHK.G
620	567.298391	1698.873345	1698.857681	9.22	0	(33)	0.00053	1	K.QSFDALVEGNLELHK.G
621	567.298391	1698.873345	1698.857681	9.22	0	(39)	0.00014	1	K.QSFDALVEGNLELHK.G
622	567.298391	1698.873345	1698.857681	9.22	0	(22)	0.0063	1	K.QSFDALVEGNLELHK.G
623	567.298391	1698.873345	1698.857681	9.22	0	(26)	0.0025	1	K.QSFDALVEGNLELHK.G
624	567.298391	1698.873345	1698.857681	9.22	0	(21)	0.0072	1	K.QSFDALVEGNLELHK.G
626	567.298391	1698.873345	1698.857681	9.22	0	(12)	0.061	1	K.QSFDALVEGNLELHK.G
337	567.629924	1699.867944	1698.857681	595	0	(19)	0.014	1	K.QSFDALVEGNLELHK.G
338	567.629924	1699.867944	1698.857681	595	0	(23)	0.0056	1	K.QSFDALVEGNLELHK.G
339	567.629924	1699.867944	1698.857681	595	0	(49)	1.3e – 05	1	K.QSFDALVEGNLELHK.G
345	573.634064	1717.880364	1717.871567	5.12	0	(52)	6.1e – 06	1	R.YTFVLAPDELFFTR.D
633	573.634064	1717.880364	1717.871567	5.12	0	(51)	7.1e – 06	1	R.YTFVLAPDELFFTR.D
634	573.634064	1717.880364	1717.871567	5.12	0	(55)	3.1e – 06	1	R.YTFVLAPDELFFTR.D
635	573.634064	1717.880364	1717.871567	5.12	0	(20)	0.0092	1	R.YTFVLAPDELFFTR.D
636	859.948171	1717.881790	1717.871567	5.95	0	61	8e – 07	1	R.YTFVLAPDELFFTR.D
637	859.948171	1717.881790	1717.871567	5.95	0	(51)	8.1e – 06	1	R.YTFVLAPDELFFTR.D
638	859.948171	1717.881790	1717.871567	5.95	0	(45)	3.3e – 05	1	R.YTFVLAPDELFFTR.D
639	859.948171	1717.881790	1717.871567	5.95	0	(17)	0.022	1	R.YTFVLAPDELFFTR.D
640	859.948171	1717.881790	1717.871567	5.95	0	(36)	0.00023	1	R.YTFVLAPDELFFTR.D
641	859.948171	1717.881790	1717.871567	5.95	0	(12)	0.061	1	R.YTFVLAPDELFFTR.D
643	860.451172	1718.887792	1717.871567	592	0	(18)	0.015	1	R.YTFVLAPDELFFTR.D
684	621.960266	1862.858970	1862.856750	1.19	0	(21)	0.0089	1	-.GASTQDMLTPLDAEEAAK.T + oxidation (M)
685	932.437927	1862.861302	1862.856750	2.44	0	87	2.1e – 09	1	-.GASTQDMLTPLDAEEAAK.T + oxidation (M)
686	932.437927	1862.861302	1862.856750	2.44	0	(51)	7.2e – 06	1	-.GASTQDMLTPLDAEEAAK.T + oxidation (M)
741	659.353699	1975.039269	1975.020355	9.58	1	57	1.9e – 06	1	R.TRYTFVLAPDELFFTR.D
742	659.353699	1975.039269	1975.020355	9.58	1	(21)	0.0073	1	R.TRYTFVLAPDELFFTR.D
459	659.687622	1976.041038	1975.020355	517	1	(23)	0.005	1	R.TRYTFVLAPDELFFTR.D
504	695.376800	2083.108572	2083.094971	6.53	0	(37)	0.00021	1	R.GLQNVVDPLPVSEAEFLTR.Y
749	695.376800	2083.108572	2083.094971	6.53	0	(62)	5.8e – 07	1	R.GLQNVVDPLPVSEAEFLTR.Y
750	695.376800	2083.108572	2083.094971	6.53	0	(87)	2.2e – 09	1	R.GLQNVVDPLPVSEAEFLTR.Y
751	695.376800	2083.108572	2083.094971	6.53	0	104	4.4e – 11	1	R.GLQNVVDPLPVSEAEFLTR.Y
752	695.376800	2083.108572	2083.094971	6.53	0	(60)	9.8e – 07	1	R.GLQNVVDPLPVSEAEFLTR.Y
753	695.376800	2083.108572	2083.094971	6.53	0	(40)	0.0001	1	R.GLQNVVDPLPVSEAEFLTR.Y
754	1042.563375	2083.112198	2083.094971	8.27	0	(94)	3.9e – 10	1	R.GLQNVVDPLPVSEAEFLTR.Y
755	1042.563375	2083.112198	2083.094971	8.27	0	(68)	1.7e – 07	1	R.GLQNVVDPLPVSEAEFLTR.Y
756	1042.563375	2083.112198	2083.094971	8.27	0	(70)	9.7e – 08	1	R.GLQNVVDPLPVSEAEFLTR.Y
757	1042.563375	2083.112198	2083.094971	8.27	0	(90)	1e – 09	1	R.GLQNVVDPLPVSEAEFLTR.Y
758	1042.563375	2083.112198	2083.094971	8.27	0	(76)	2.6e – 08	1	R.GLQNVVDPLPVSEAEFLTR.Y
759	1042.563375	2083.112198	2083.094971	8.27	0	(27)	0.0019	1	R.GLQNVVDPLPVSEAEFLTR.Y
505	695.709351	2084.106225	2083.094971	485	0	(39)	0.00011	1	R.GLQNVVDPLPVSEAEFLTR.Y
760	695.709351	2084.106225	2083.094971	485	0	(31)	0.00074	1	R.GLQNVVDPLPVSEAEFLTR.Y
761	1043.061279	2084.108006	2083.094971	486	0	(44)	3.8e – 05	1	R.GLQNVVDPLPVSEAEFLTR.Y

Table S3 Continued

Query	Observed	M_r(expt)	M_r(calc)	p.p.m.	Miss	Score	Expect	Rank	Peptide
762	1043.061279	2084.108006	2083.094971	486	0	(24)	0.0036	1	R.GLQNVVDPLPVSEAEFLTR.Y
801	1141.108002	2280.201452	2280.188904	5.50	0	127	1.8e − 13	1	K.VIEAATLDLLTAFLDEQYQK.Q
802	1141.108002	2280.201452	2280.188904	5.50	0	(104)	3.9e − 11	1	K.VIEAATLDLLTAFLDEQYQK.Q
803	1141.108002	2280.201452	2280.188904	5.50	0	(105)	2.9e − 11	1	K.VIEAATLDLLTAFLDEQYQK.Q
804	1141.108002	2280.201452	2280.188904	5.50	0	(85)	3.4e − 09	1	K.VIEAATLDLLTAFLDEQYQK.Q
805	761.075882	2280.205818	2280.188904	7.42	0	(88)	1.6e − 09	1	K.VIEAATLDLLTAFLDEQYQK.Q
806	761.075882	2280.205818	2280.188904	7.42	0	(89)	1.2e − 09	1	K.VIEAATLDLLTAFLDEQYQK.Q
807	761.075882	2280.205818	2280.188904	7.42	0	(76)	2.5e − 08	1	K.VIEAATLDLLTAFLDEQYQK.Q
808	761.075882	2280.205818	2280.188904	7.42	0	(77)	1.8e − 08	1	K.VIEAATLDLLTAFLDEQYQK.Q
558	761.411194	2281.211754	2280.188904	449	0	(49)	1.2e − 05	1	K.VIEAATLDLLTAFLDEQYQK.Q
559	761.411194	2281.211754	2280.188904	449	0	(42)	6.6e − 05	1	K.VIEAATLDLLTAFLDEQYQK.Q
828	805.104858	2412.292746	2412.272934	8.21	1	23	0.0047	1	R.YTFVLAPDELFFTRDGLPIAK.T
829	805.104858	2412.292746	2412.272934	8.21	1	(23)	0.005	1	R.YTFVLAPDELFFTRDGLPIAK.T
830	609.037720	2432.121776	2432.103302	7.60	0	(14)	0.043	1	R.DSFQIHYDINDNHTPFDNIK.S
831	811.714905	2432.122887	2432.103302	8.05	0	37	0.0002	1	R.DSFQIHYDINDNHTPFDNIK.S
832	1217.069397	2432.124242	2432.103302	8.61	0	(9)	0.13	1	R.DSFQIHYDINDNHTPFDNIK.S
906	911.717468	3642.840768	3642.818039	6.24	0	32	0.0006	1	K.SSIKPEHFHLMYLLEQHSPYFIDAELTELR.D
907	729.577667	3642.851955	3642.818039	9.31	0	(15)	0.033	1	K.SSIKPEHFHLMYLLEQHSPYFIDAELTELR.D
908	729.577667	3642.851955	3642.818039	9.31	0	(15)	0.031	1	K.SSIKPEHFHLMYLLEQHSPYFIDAELTELR.D
910	738.368835	3686.807795	3684.828598	537	0	(2)	0.7	1	K.SSIKPEHFHLMYLLEQHSPYFIDAELTELR.D acetyl

Received 30 July 2012/1 October 2012; accepted 25 October 2012
Published as BJ Immediate Publication 25 October 2012, doi:10.1042/BJ20121209