Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Role of domains within the autotransporter Hbp/Tsh

Kaoru Nishimura,[a] Young-Ho Yoon,[a]‡ Atsushi Kurihara,[a] Satoru Unzai,[a] Joen Luirink,[b] Sam-Yong Park[a] and Jeremy R. H. Tame[a]*

[a]Yokohama City University, Suehiro 1-7-29, Tsurumi, Yokohama 230-0045, Japan, and [b]Department of Molecular Microbiology, Faculty of Earth and Life Sciences, Vrije Universiteit, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands

‡ Present address: Trans-Membrane Trafficking Unit, Okinawa Institute of Science and Technology, 1919-1 Tancha, Onna, Kunigami, Okinawa 904-0412, Japan.

Correspondence e-mail: jtame@tsurumi.yokohama-cu.ac.jp

Received 8 July 2010
Accepted 15 September 2010

PDB Reference: Hbp, 3ak5.

The autotransporter Tsh (temperature-sensitive haemagglutinin) secreted by avian pathogenic *Escherichia coli* was reported in 1994 and the almost identical Hbp (haemoglobin protease) was discovered some years later in isolates from patients suffering from peritoneal abscesses. However, the function of the protein remains uncertain. The crystal structure of Hbp shows that the protein carries a serine protease domain (domain 1) and a small domain of 75 residues called domain 2 which is inserted into the long β-helix characteristic of autotransporter passenger proteins. In this paper, domain 1 is shown to bind calcium, although metal ions binding to this site do not seem to regulate protease activity. Tsh has been reported to bind red cells and components of the extracellular matrix, but it is demonstrated that these properties are not a consequence of the presence of domain 2.

1. Introduction

Autotransporters are a very large group of virulence proteins secreted by pathogenic Gram-negative bacteria (Dautin & Bernstein, 2007). They carry out a variety of functions to promote disease, such as helping bacteria adhere to host cells or enter them (Wells *et al.*, 2007). Autotransporter genes have a common organization, encoding a pre-protein with a signal peptide at the N-terminus which is often much longer than normal signal peptides but which is removed as the protein enters the periplasm (Jacob-Dubuisson *et al.*, 2004; Henderson *et al.*, 2004; Desvaux *et al.*, 2004*b*). The C-terminal domain forms a β-barrel domain which inserts into the outer membrane with assistance from the Bam complex (Bernstein, 2007; Voulhoux *et al.*, 2003). The role of this barrel in moving the passenger domain across the outer membrane remains unclear. The majority of autotransporter passenger proteins are characterized by a long β-helix capped at the C-terminal end with a so-called junction region or autochaperone domain which is believed to initiate folding of the protein outside the cell (Ohnishi *et al.*, 1994; Oliver *et al.*, 2003). The unique mechanism of crossing the outer membrane is the subject of intense scrutiny and has been reviewed several times (Dautin & Bernstein, 2007; Jacob-Dubuisson *et al.*, 2004; Desvaux *et al.*, 2004*a*,*b*). Autotransporters continue to be discovered at a rapid rate and are extremely interesting from a medical point of view. It is known that some of the most copiously secreted proteins from pathogenic bacteria are autotransporters and that they are required for maximum pathogenicity. A number of autotransporters carry a serine protease domain at the N-terminus and are known as SPATEs (serine protease

autotransporters of Enterobacteriaceae), although other bacterial species may also secrete such proteins (Dautin, 2010). For example, immunoglobulin A protease (IgAP) is secreted by *Haemophilus influenzae* and serves to cut antibodies at a specific proline-rich site to protect the bacteria from the host's immune response. SepA is the most highly secreted protein from *Shigella* and is required for pathogenicity, but its functional role is unknown (Benjelloun-Touimi *et al.*, 1998). EspC is another highly studied member of this family. Both SepA and EspC strongly resemble IgAP and Hbp (Nishimura *et al.*, 2010) and homology models can be built from the experimentally determined crystal structures of these two proteins. Both IgAP and Hbp have a small domain called domain 2 projecting from the β-helix and in the case of IgAP this domain is believed to bind to substrate antibodies which are cleaved by the protease domain (Johnson *et al.*, 2009). Not all autotransporter passengers contain β-helical regions or possess enzyme activity, but even members which resemble each other structurally may have very different functions. The serine protease subfamily of autotransporters includes Pet (plasmid-encoded toxin), which is known from extensive studies to enter epithelial cells and cause vacuolization (Dutta *et al.*, 2003). Pic, another member of the same subfamily, is suggested to be a mucinase and has been shown to promote intestinal colonization by enteroaggregative *Escherichia coli* (Harrington *et al.*, 2009; Parham *et al.*, 2004). Pic also prevents the killing of *E. coli* by human serum, apparently by cleaving a protein involved in the early steps of complement activation (Henderson *et al.*, 1999), but does not have an equivalent of domain 2.

Although Hbp was the first serine protease autotransporter crystal structure to be solved (Otto *et al.*, 2005), the true role of the protein in disease remains unclear (Fig. 1). Hbp is almost identical to Tsh, which was identified in avian pathogenic *E. coli* (Provence & Curtiss, 1994). It has been reported that Hbp can bind to dried red blood cells (Kostakioti & Stathopoulos, 2004) and that the protein is able to bind haem and to digest haemoglobin (Otto *et al.*, 1998). It has also been shown that Hbp is required for microbial synergy between *E. coli* and *Bacteroides fragilis* in the formation of peritoneal abscesses. The importance of Hbp to pathogenicity was demonstrated by immunizing mice with Hbp before challenging them with the bacteria and showing that this conferred protection against abscess formation (Otto *et al.*, 2002). The picture which emerges from these studies, of a protein devoted to haem capture by disruption of red cells and digestion of haemoglobin, has yet to be matched to the structural work, which provides a molecular view of the protein but offers little direct clue to its substrates. In this paper, we have attempted further characterization of Hbp through mutation, direct binding studies and crystallography. In particular, we examine the role of a small side domain of Hbp called domain 2. We show that Hbp is an active protease and is a calcium-binding protein, but we are unable to demonstrate any binding activity of domain 2.

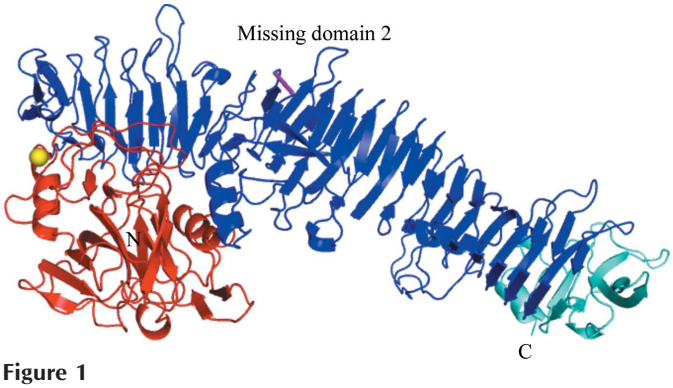

Figure 1
The overall structure of Hbp Gdel-2. A C^{α} trace of the *A* subunit of the protein, with helices shown as ribbons and β-strands as arrows. Domain 1 (residues 53–309) is shown in red, the β-helix is shown in blue and the C-terminal autochaperone domain (residues 1000–1100) is shown in cyan. The calcium ion is shown as a yellow sphere. The break in the main chain of the model from Tyr532 to Gly611 is shown as a pink line. The figure was produced with *PyMOL* (http://www.pymol.org).

2. Materials and methods

2.1. Expression and purification

Hbp has UniProt reference code O88093. The deletion of domain 2 was carried out by standard PCR procedures as described previously, linking Ala533 to Asn608 with a single glycine residue (Otto *et al.*, 2005). The resulting Hbp Gdel-2 was expressed in *E. coli* and purified according to published procedures (Otto *et al.*, 2005). The sample was purified by size-exclusion column chromatography (HiLoad 16/60 Superdex 200 prep grade, Amersham Bioscience) equilibrated with 20 m*M* Tris pH 7.5, 150 m*M* NaCl. The protein was concentrated to 4.5 mg ml^{-1} before crystallization. The Hbp Gdel-2 protein has 975 residues with a molecular weight of 103.6 kDa.

2.2. Crystallization

Crystals of Gdel-2 were formed in a 1:1 mixture (4 µl final volume) of protein (4.5 mg ml^{-1} in buffer consisting of 20 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl, 1 m*M* $CaCl_2$) and reservoir solution consisting of 9% PEG 1K, 10% PEG 8K, 0.1 *M* sodium acetate pH 5.0 and 20% glycerol using the hanging-drop vapour-diffusion method at 293 K. Crystals grew as rods of up to 300 µm in length over a period of about a month.

2.3. Data collection and processing

The crystal was directly flash-frozen in a cold stream of nitrogen gas and stored in liquid nitrogen. Gdel-2 diffraction data were collected using a single crystal at 100 K on beamline BL41XU at SPring-8. 250 images were collected with an oscillation angle of 1° using incident radiation with a wavelength of 1.000 Å. The data sets were indexed and integrated using *HKL*-2000 and scaled with *SCALEPACK* to 2.2 Å resolution. The crystal belonged to space group *C*2, with unit-cell parameters a = 292.8, b = 53.9, c = 335.9 Å, β = 103.0° and four molecules per asymmetric unit, giving a Matthews coefficient of 2.95 Å^3 Da^{-1} and a solvent content of 58%.

Table 1
Refinement statistics.

Values in parentheses are for the outer shell.

Refinement resolution (Å)	50.0–2.20 (2.28–2.20)
Measured reflections	853639
Unique reflections	239062
Completeness (%)	93.3 (71.9)
R_{merge}† (%)	8.1 (21.3)
Multiplicity	3.7 (2.5)
$\langle I/\sigma(I)\rangle$	14.1 (2.2)
σ cutoff/reflections used	0.0/227076
R_{cryst}/R_{free}‡ (%)	22.5/28.0
R.m.s.d. bond lengths/bond angles (Å)	0.021/1.949
No. of water molecules	1485
Average B factor (protein/water) (Å^2)	32.8/30.8
Average B factor of calcium ions (Å^2)	20.9
Ramachandran plot	
Residues in most favourable regions (%)	84.6
Residues in additional allowed regions (%)	14.4
Residues in additional allowed regions (%)	0.8

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an observation, $\langle I(hkl)\rangle$ is the mean value for that reflection and the summations are over all reflections. ‡ R factor $= \sum_{hkl} ||F_{obs}| - |F_{calc}||/\sum_{hkl}|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively. The free R factor was calculated with 5% of data that were excluded from refinement.

2.4. Structure determination

The structure was solved by molecular replacement using *MOLREP* (Vagin & Teplyakov, 2000) and *EPMR* (Kissinger *et al.*, 1999), with the wild-type structure (PDB code 1wxr; Otto *et al.*, 2005) as the search model after deleting domain 2. The calculations resulted in one clear solution with a correlation coefficient of 0.52, an R factor of 0.54 and crystal packing with no severe clashes between molecules. Manual model building was performed using *Coot* (Emsley & Cowtan, 2004) and refinement was performed using *REFMAC*5 (Murshudov *et al.*, 1999) from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). The free R factor was calculated using a randomly selected 5% of reflections. Data to the highest resolution were used from the beginning of refinement. Noncrystallographic symmetry restraints were applied throughout refinement and isotropic temperature factors were refined for each atom. Data-collection and refinement statistics are summarized in Table 1. Seven residues in the final model were found to lie outside the expected regions of the Ramachandran plot, all of which were located in flexible surface loops that were not visible in every copy of the molecule. Regions close to the deletion site also showed signs of disorder, so that Asp531 and Asn632 refined poorly, indicating that the domain deletion destabilizes the structure locally. Other regions of unusual geometry included residues 966–969 and 991–993 in some chains; both regions form a turn at the surface. Residues 1015–1025 are also poorly ordered, forming a loop that projects from the remainder of the protein and has no contact with it. These regions within the C-terminal domain are also poorly ordered in the model of the wild-type protein. Atoms belonging to several surface side chains were set to zero occupancy in order to reflect their absence from the electron-density map and six water molecules were set to occupancy 0.5.

2.5. Oligopeptide-cleavage assay

5 µg of Hbp (wild type or Gdel-2) dissolved in 45 µl 50 m*M* Tris–HCl pH 7.5, 100 m*M* NaCl were added separately to 0.5 m*M* pNA-conjugated elastase substrate IV, Suc-Ala-Ala-Pro-Abu-*p*NA (Calbiochem), in the same buffer. Reactions were performed at 310 K in a volume of 100 µl. Absorbance readings at 405 nm were taken at various time points up to 20 h. The maximum absorbance reading obtained was set to 100% and other readings were normalized accordingly with each protease. All reactions were performed in triplicate to determine the standard deviation. To test the effects of divalent metal ions on Hbp, the enzyme activity was assayed in the presence of 1 m*M* EDTA, $MgCl_2$, $CaCl_2$, $MnCl_2$, $CoCl_2$, $NiCl_2$, $CuCl_2$, $ZnCl_2$, $SrCl_2$ or $BaCl_2$. A comparative assay was conducted with trypsin using a suitable substrate, 50 m*M* Tris–HCl, 100 m*M* NaCl, under the same conditions with no divalent metal ions present. To test the effects of two protease inhibitors [AEBSF and elastase inhibitor II (MeOSuc-Ala-Ala-Pro-Ala-CMK)] on Hbp, the protein solution was incubated for 1 h at 310 K with 5 m*M* of each inhibitor before the activity assay.

2.6. Haem-binding assay

400 µl of haemin–agarose beads (Sigma) was washed with 20 m*M* HEPES–NaOH pH 7.5. Washing was performed five times by suspension of the agarose in 2 ml buffer. Purified Hbp and Gdel-2 (100 µg in 500 µl buffer) were mixed with the haemin–agarose at room temperature. The agarose was washed five times with 2 ml of the same buffer plus 0.5 *M* NaCl and thereafter with 1.0 *M* NaCl; the first wash with each NaCl solution was run on SDS–PAGE.

2.7. Analytical ultracentrifugation

Sedimentation-velocity measurements were carried out using a Beckman XL-I analytical ultracentrifuge (Beckman Coulter, California, USA). Cells with a standard Epon two-channel centrepiece and sapphire windows were used. In each case the buffer used was 50 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl and the temperature was 293 K. The rotor temperature was equilibrated at 293 K in the vacuum chamber for 1–2 h prior to starting each run. The domain 2–GFP fusion protein has a strong intrinsic yellow colour and sedimentation was monitored by absorption at 490 nm. Absorbance scans were collected at 10 min intervals during sedimentation at 50 000 rev min^{-1}. The fusion protein was mixed with the binding substrate to be tested in a 1:1 molar ratio, each being present at a concentration of 6 m*M*. The partial specific volume of the protein, solvent density and solvent viscosity were calculated from standard tables using the program *SEDNTERP*. The resulting scans were analyzed using the continuous distribution [$c(s)$] analysis module in the program *SEDFIT* (Schuck, 2000). Sedimentation-coefficient increments of 100 were used in the appropriate range for each sample and the weight-average frictional ratio (f/f_o) was allowed to float during fitting.

3. Results and discussion

3.1. Overview

Wild-type Hbp is expressed as a 1377-residue pre-protein with a signal sequence extending from residues 1 to 52 and a C-terminal domain from residues 1101 to 1377. The passenger protein, residues 53–1100, is autocatalytically cut from the C-terminal domain after it has been transported to the cell surface. The passenger is a monomer in solution (Scott *et al.*, 2002) and its crystal structure has been solved (Otto *et al.*, 2005). This model (PDB code 1wxr) was numbered from 1, so that the residue numbering is out of step by 52 with the pre-protein. Most functional studies of Hbp/Tsh have been described numbering the initiator methionine as residue 1 and here we have used the same numbering for ease of comparison. The numbering scheme for this model therefore differs from that of PDB entry 1wxr. The crystal structure of the full-length Hbp passenger shows clearly defined domains within the protein: a trypsin-family serine protease domain (domain 1) connected to a long β-helix from which various loops emerge. Domain 2 is a separate self-folding structure built from 75 residues, with the chain trace leaving and returning to the β-helix at the same point. Domain 2 possesses a binding site of some sort, but not apparently for haem. Deleting domain 2 by linking Ala533 to Asn608 with a glycine residue does not hinder protein expression or secretion into the medium (Otto *et al.*, 2005). It has been suggested that this domain may be required for correct folding of the passenger (Brockmeyer *et al.*, 2009), so we began by determining the structure of this mutant lacking domain 2, called 'Gdel-2'. The protein crystallized in a different space group to the wild-type passenger, as expected from the crystal contacts formed by domain 2 in the previous model. The new crystal form in space group *C*2 contained four molecules in the asymmetric unit and diffracted to 2.2 Å resolution. Refinement proceeded smoothly, but a number of surface loops showed disorder and several residues were missing at the site of the deletion between Tyr532 and Gly611. Each copy of the molecule showed only minor differences from the others, although some of the more mobile loops were poorly represented in the electron-density map. In some cases, these loops were omitted from the model. NCS restraints were initially applied separately to the protease domain and to the remainder of the protein, but it was soon clear that each copy of the protein showed only small deviations from the wild-type structure, so subsequent refinement was carried out with one set of NCS restraints applied to the whole molecule. Domain 2 is clearly not required for folding of the passenger. In the original model (PDB code 1wxr), the C-terminal region shows some flexibility and several loops in this region were not visible in the electron-density map. In the new crystal form, these loops are visible in chains *A* and *B* so that the entire autochaperone domain can be traced. The four independent copies of the molecule indicate that the C-terminal region is structurally uniform. The r.m.s. differences between main-chain atoms of chains *A* and chains *B*, *C* or *D* were 0.26, 0.60 and 0.55 Å, respectively. Compared with the wild-type Hbp structure (PDB entry 1wxr), chain *A* showed an r.m.s. difference of 0.985 Å.

3.2. Calcium ion-binding site

It quickly became apparent on refinement that a peak of high electron density is present at the same position in each copy of the protein, liganded by Asp145, Asn147 and Glu306 (Fig. 2). These residues are found at the junction between domain 1 and the β-helix, not far from the active-site serine Ser259. The only divalent metal present in the crystallization conditions is calcium, which matches the coordination pattern of the structure. Calcium binding was not observed in the model of the wild-type protein as it was crystallized in the presence of ammonium sulfate, which will sequester any calcium ions present. Calcium-ion concentration is known to be substantially higher outside epithelial cells than inside, raising the possibility that the protease activity may be controlled by calcium or other metal binding. Protease assays were therefore carried out to determine whether activity was subject to regulation by metal ions.

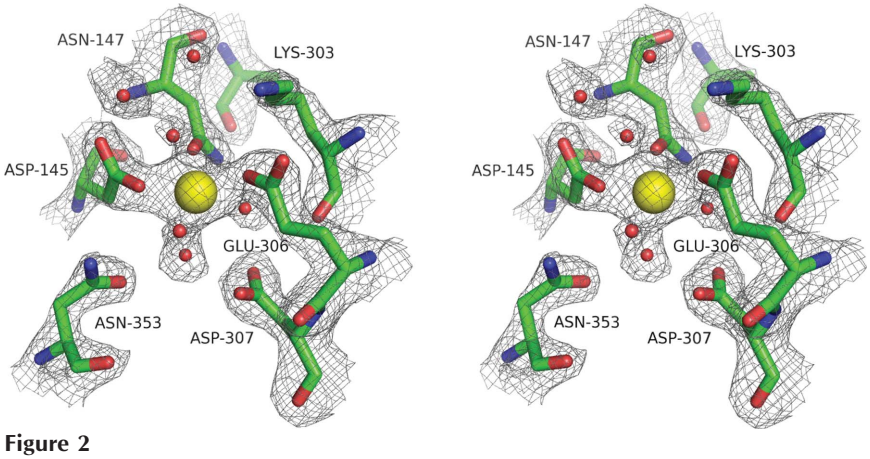

Figure 2
The calcium-binding site. A stereoview of the calcium-binding site, with the final electron-density map ($2mF_o - DF_c$) contoured at 1.5σ. The calcium ion is shown as a yellow sphere and water molecules are shown as smaller red spheres. Residues binding the calcium ion are shown as stick models, with O atoms coloured red, N atoms blue and C atoms green.

3.3. Protease assays

Protease assays were carried out using chromogenic substrate peptides, an approach that has been used previously for other SPATEs (Dutta *et al.*, 2002). Initially, Hbp protease activity was compared with that of trypsin using two different substrates: elastase substrate IV-pNA and benzoyl-arginine-pNA. In each case the reaction was monitored by absorption at 405 nm to follow the appearance of *p*-nitrophenol. The results show that Hbp is a highly efficient protease against elastase substrate IV but much less so against substrates preferred by trypsin (Fig. 3*a*). Wild-type Hbp and the Gdel-2

mutant show identical protease activity against each substrate tested. The two inhibitors tested completely abolished the protease activity. Further assays were carried out to determine whether the protease activity was affected by metal ions including calcium, but no significant effect was found (Fig. 3*b*). The metal-binding site found in the new crystal structure probably therefore acts to stabilize the protein structure but not to control the enzyme.

3.4. Analytical ultracentrifugation assays

The structure of domain 2 suggests a binding site for a hydrophobic molecule containing a number of aromatic residues which are presented at the surface of the protein. The crystal structure of IgAP shows that it possesses a very similar domain, but an insertion of 12 residues at the presumed binding site completely changes the nature of any substrate which binds to the protein in this region (Johnson *et al.*, 2009). Since IgAP is known to cleave antibodies, it has been suggested that domain 2 helps to bind the substrate protein and to position it for cleavage. This raises the possibility that domain 2 of other ATs may function in the same way and help to determine the role of the protein in promoting disease. We therefore cloned domain 2 of Hbp and expressed it as a fusion with green fluorescent protein (GFP) attached to its C-terminus and a six-residue histidine tag at the C-terminus of the GFP. This protein construct proved to be stable and could be readily purified, so domain 2 appears capable of folding independently of the remainder of the Hbp passenger protein. The fusion protein was then used to test binding of domain 2 to three different substrates from the extracellular matrix: collagen, fibronectin and fibrinogen (Sigma). Since all of these substrates have a substantial molecular mass, over 100 kDa in the cases of fibronectin and fibrinogen, binding was assayed by sedimentation velocity in an analytical ultracentrifuge, measuring the absorption of light at 490 nm to monitor the movement of the domain 2 fusion protein. No indication was seen of any interaction between domain 2 and the three substrates used (Fig. 4). A further test was made using the fusion protein and fresh red blood cells, but again no binding was seen (data not shown). It has been reported that Tsh is able to bind to fibronectin, collagen IV and dried red blood cells (Kostakioti & Stathopoulos, 2004), but domain 2 is apparently not responsible for this property.

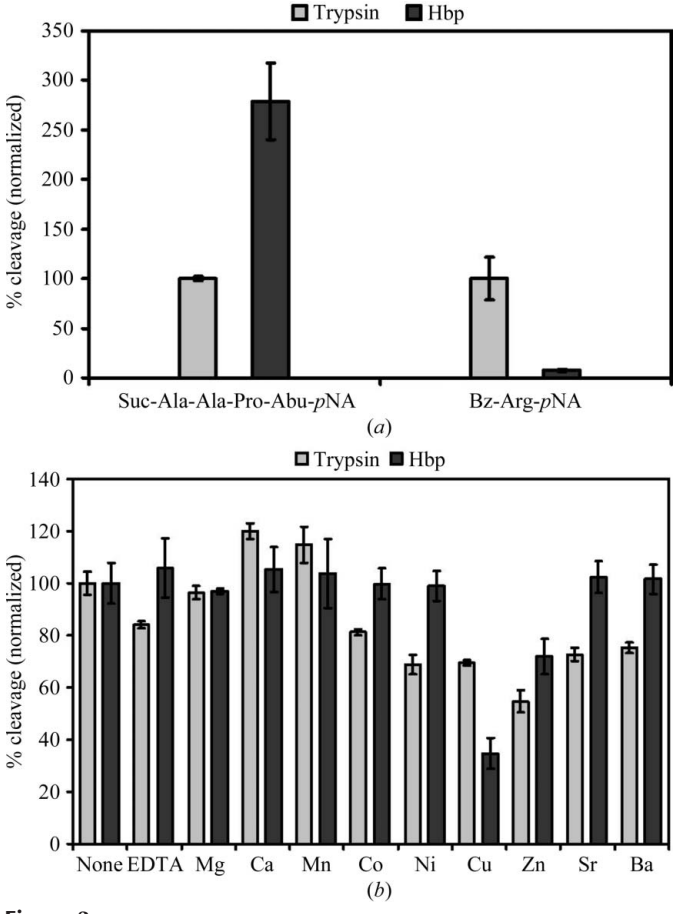

Figure 3
Protease assays. (*a*) Trypsin and Hbp Gdel-2 activity against elastase substrate IV (left) and benzoyl arginine (right). In each case the activity of trypsin is set to 100%. (*b*) Metal dependence. No significant change in Hbp protease activity against elastase substrate IV was seen on adding a variety of divalent metal ions to a 1 m*M* concentration. Copper ions appear to have a mild inhibitory effect, but 1 m*M* is an unphysiologically high concentration for this metal. Copper ions have rather different coordination preferences to calcium ions, so it seems unlikely that copper is acting at the calcium ion-binding site; it possibly interferes directly with the protease active site.

3.5. Haem-binding assay

Previous studies by Otto and coworkers have shown that Hbp is retained by haemin–agarose, which supported the suggestion that the protein is involved in haem transport (Otto

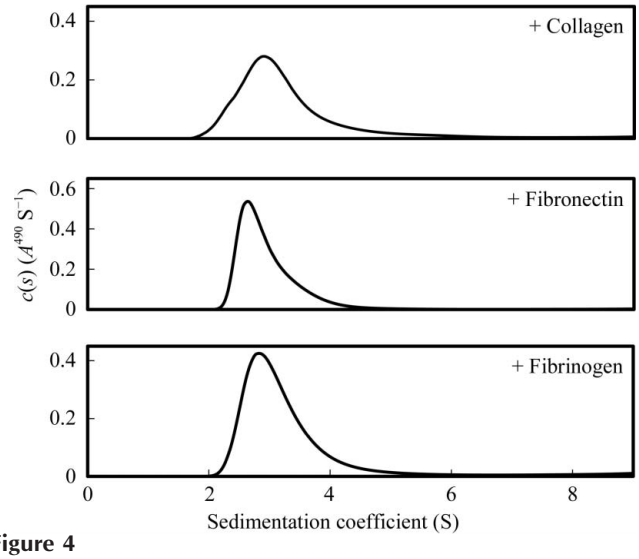

Figure 4
Binding assays. Domain-2–GFP fusion protein was sedimented with a variety of potential binding partners and its movement in the analytical centrifuge was monitored using absorption at 490 nm. In the cases of fibronectin and fibrinogen the sedimentation distribution appeared to be unchanged from control runs with domain-2–GFP alone. In the case of added collagen the shape of the distribution is slightly altered, but strong binding to the collagen fragments, with a size distribution of 25–100 kDa, would be expected to produce a much more marked increase in the sedimentation velocity.

et al., 1998). We repeated this assay using both wild-type protein and Gdel-2. Both showed very similar behaviour, binding strongly to the resin and being eluted between 0.5 and 1 *M* NaCl. Haem is a highly apolar molecule and there is always a possibility that retention on haemin–agarose may be a consequence of nonspecific hydrophobic interactions rather than specific haem binding. No obvious haem-binding site is readily identifiable from the structure of the protein and the possibility cannot be discounted that the interaction observed between Hbp and hemin agarose is in fact nonspecific. We have demonstrated that domain 2 is not required for this interaction, in agreement with Otto and coworkers, who used tetramethylbenzidine staining to observe haem binding of Hbp (Otto *et al.*, 2005).

4. Conclusions

Despite testing a range of substrates suggested to bind to Tsh, we failed to find a role for domain 2 of Hbp/Tsh. This region of the protein is not required for protein secretion or folding, yet its structure is conserved among a number of SPATE-family members, including IgAP. No experimental structure of SepA has been determined, but a homology model for domain 2 of SepA can readily be generated from the model of Hbp. It has yet to be tested whether deletion of domain 2 of Hbp results in an attenuation of pathogenicity, but the domain 2 regions from different SPATEs may prove to be of value as components of vaccines targeting different diseases. Similarly, the β-helical Hbp passenger protein is a stable platform that is suitable for the expression of different domains on the bacterial surface and domain 2 can clearly be replaced with other sequences of interest. We have demonstrated that Hbp is a calcium-binding protein and that it is proteolytically active against similar substrates to elastase. Given the wide range of substrates that Hbp/Tsh is reported to bind, care must clearly be taken to exclude the possibility of nonspecific hydrophobic interactions. Sequence differences between Pic, SepA and Hbp suggest they may proteolyze quite different targets, but each contributes to the survival of pathogenic bacteria within the host by effects such as countering the immune response or digesting the mucin layer protecting epithelial cells. Further work is required in order to understand the precise roles of these proteins and their domains.

References

Benjelloun-Touimi, Z., Si Tahar, M., Montecucco, C., Sansonetti, P. J. & Parsot, C. (1998). *Microbiology*, **144**, 1815–1822.

Bernstein, H. D. (2007). *Trends Microbiol.* **15**, 441–447.

Brockmeyer, J., Spelten, S., Kuczius, T., Bielaszewska, M. & Karch, H. (2009). *PLoS One*, **4**, e6100.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Dautin, N. (2010). *Toxins*, **2**, 1179–1206.

Dautin, N. & Bernstein, H. D. (2007). *Annu. Rev. Microbiol.* **61**, 89–112.

Desvaux, M., Parham, N. J. & Henderson, I. R. (2004*a*). *Curr. Issues Mol. Biol.* **6**, 111–124.

Desvaux, M., Parham, N. J. & Henderson, I. R. (2004*b*). *Res. Microbiol.* **155**, 53–60.

Dutta, P. R., Cappello, R., Navarro-Garcia, F. & Nataro, J. P. (2002). *Infect. Immun.* **70**, 7105–7113.

Dutta, P. R., Sui, B. Q. & Nataro, J. P. (2003). *J. Biol. Chem.* **278**, 39912–39920.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Harrington, S. M., Sheikh, J., Henderson, I. R., Ruiz-Perez, F., Cohen, P. S. & Nataro, J. P. (2009). *Infect. Immun.* **77**, 2465–2473.

Henderson, I. R., Czeczulin, J., Eslava, C., Noriega, F. & Nataro, J. P. (1999). *Infect. Immun.* **67**, 5587–5596.

Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C. & Ala'Aldeen, D. (2004). *Microbiol. Mol. Biol. Rev.* **68**, 692–744.

Jacob-Dubuisson, F., Fernandez, R. & Coutte, L. (2004). *Biochim. Biophys. Acta*, **1694**, 235–257.

Johnson, T. A., Qiu, J., Plaut, A. G. & Holyoak, T. (2009). *J. Mol. Biol.* **389**, 559–574.

Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.

Kostakioti, M. & Stathopoulos, C. (2004). *Infect. Immun.* **72**, 5548–5554.

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.

Nishimura, K., Tajima, N., Yoon, Y. H., Park, S. Y. & Tame, J. R. H. (2010). *J. Mol. Med.* **88**, 451–458.

Ohnishi, Y., Nishiyama, M., Horinouchi, S. & Beppu, T. (1994). *J. Biol. Chem.* **269**, 32800–32806.

Oliver, D. C., Huang, G., Nodel, E., Pleasance, S. & Fernandez, R. C. (2003). *Mol. Microbiol.* **47**, 1367–1383.

Otto, B. R., Sijbrandi, R., Luirink, J., Oudega, B., Heddle, J. G., Mizutani, K., Park, S. Y. & Tame, J. R. H. (2005). *J. Biol. Chem.* **280**, 17339–17345.

Otto, B. R., van Dooren, S. J., Dozois, C. M., Luirink, J. & Oudega, B. (2002). *Infect. Immun.* **70**, 5–10.

Otto, B. R., van Dooren, S. J., Nuijens, J. H., Luirink, J. & Oudega, B. (1998). *J. Exp. Med.* **188**, 1091–1103.

Parham, N. J., Srinivasan, U., Desvaux, M., Foxman, B., Marrs, C. F. & Henderson, I. R. (2004). *FEMS Microbiol. Lett.* **230**, 73–83.

Provence, D. L. & Curtiss, R. (1994). *Infect. Immun.* **62**, 1369–1380.

Schuck, P. (2000). *Biophys. J.* **78**, 1606–1619.

Scott, D. J., Grossmann, J. G., Tame, J. R. H., Byron, O., Wilson, K. S. & Otto, B. R. (2002). *J. Mol. Biol.* **315**, 1179–1187.

Vagin, A. & Teplyakov, A. (2000). *Acta Cryst.* D**56**, 1622–1624.

Voulhoux, R., Bos, M. P., Geurtsen, J., Mols, M. & Tommassen, J. (2003). *Science*, **299**, 262–265.

Wells, T. J., Tree, J. J., Ulett, G. C. & Schembri, M. A. (2007). *FEMS Microbiol. Lett.* **274**, 163–172.