LUND UNIVERSITY

Haemophilus influenzae surface fibril (Hsf) is a unique twisted hairpin-like trimeric autotransporter

Singh, Birendra; Jubair, Tamim Al; Mörgelin, Matthias; Sundin, Anders; Linse, Sara; Nilsson, Ulf J.; Riesbeck, Kristian

Published in:
International Journal of Medical Microbiology

DOI:
10.1016/j.ijmm.2014.10.004

2015

Link to publication

Citation for published version (APA):
Singh, B., Jubair, T. A., Mörgelin, M., Sundin, A., Linse, S., Nilsson, U. J., & Riesbeck, K. (2015). Haemophilus influenzae surface fibril (Hsf) is a unique twisted hairpin-like trimeric autotransporter. *International Journal of Medical Microbiology*, *305*(1), 27-37. https://doi.org/10.1016/j.ijmm.2014.10.004

Total number of authors:
7

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Haemophilus influenzae surface fibril (Hsf) is a unique twisted hairpin-like trimeric autotransporter

Birendra Singh[1], Tamim Al Jubair, Matthias Mörgelin[2], Anders Sundin[3], Sara Linse[4], Ulf Nilsson[3], and Kristian Riesbeck[1*]

[1]*Medical Microbiology, Department of Laboratory Medicine Malmö, Lund University, Jan Waldenströms gata 59, SE-205 02 Malmö, Sweden,* [2]*Section of Clinical and Experimental Infectious Medicine, Department of Clinical Sciences, Lund University, SE-221 84 Lund, Sweden,* [3]*Organic Chemistry, Lund University, PO Box 124, SE-221 00 Lund, Sweden, and* [4]*Biochemistry and Structural Biology, Lund University, PO Box 124, SE-221 00 Lund, Sweden*

Running Title: Structure of the *Haemophilus* surface fibril

[*]**Corresponding author**: Dr. Kristian Riesbeck
Medical Microbiology, Department of Laboratory Medicine Malmö,
Lund University, Jan Waldenströms gata 59, SE-205 02 Malmö, Sweden
Tel.: +46 40 338494; fax: +46 40 336234
E-mail address: kristian.riesbeck@med.lu.se

Key words: *Haemophilus influenzae* type b, *Haemophilus* surface fibril, Hib, Hsf

Abstract

The *Haemophilus* surface fibril (Hsf) is an extraordinary large (2413 amino acids) trimeric autotransporter, present in all encapsulated *Haemophilus influenzae*. It contributes to virulence by directly functioning as an adhesin. Furthermore, Hsf recruits the host factor vitronectin thereby inhibiting the host innate immune response resulting in enhanced survival in serum. Here we observed by electron microscopy that Hsf appears as an 100 nm long fibril at the bacterial surface albeit the length is approximately 200 nm according to a bioinformatics based model. To unveil this discrepancy, we denaturated Hsf at the surface of Hib by using guanidine hydrochloride (GuHCl). Partial denaturation induced in the presence of GuHCl unfolded the Hsf molecules, and resulted in an increased length of fibres in comparison to the native trimeric form. Importantly, our findings were also verified by *E. coli* expressing Hsf at its surface. In addition, a set of Hsf-specific peptide antibodies also indicated that the N-terminal of Hsf is located near the C-terminal at the base of the fibril. Taken together, our results demonstrated that Hsf is not a straight molecule but is folded and doubled over. This is the first report that provides the unique structural features of the trimeric autotransporter Hsf.

Highlights

- Trimeric autotransporters are important virulence factors of Gram negative bacteria, and have a “lollipop-shape” structure or straight trimeric strand.
- *Haemophilus* surface fibril is involved in adherence to host epithelium and for recruitment of vitronectin.
- Here we show for the first time that *Haemophilus* surface fibril has a twisted hairpin-like structure.

Introduction

Haemophilus influenzae is a Gram-negative respiratory pathogen that are categorised into encapsulated (serotype a to f) and unencapsulated strains, the later group is designated as non-typeable *H. influenzae* (NTHi). *H. influenzae* type b (Hib) causes pneumonia, osteomyelitis, epiglottitis, sepsis, joint infections, and acute meningitis and is hence considered as the most virulent type (Morris et al., 2008; Agrawal and Murphy, 2011). Even though the incidence of Hib infections in developed countries has been significantly reduced after introduction of the Hib conjugate vaccine in the early 1990s (Danovaro-Holliday et al., 2008), Hib remains a major infectious agent in infants and children in developing countries (Fitzwater et al., 2010). Hib infection starts by attachment of the bacteria to the nasopharyngeal and lung epithelial surfaces resulting in epithelial damage followed by penetration of the underlying tissues mediated by various sophisticated mechanisms (Geme and Cutter, 1995; Geme, 1996; Ulanova and Tsang, 2009; Agrawal and Murphy, 2011). Hib is furthermore able to penetrate the blood-brain barrier and thus causes meningitis (Singh et al., 2012). Survival of Hib in the blood is controlled by acquiring complement regulators to the surface of the pathogen for an effective inhibition of the membrane attack complex (MAC) (Winkelstein and Moxon, 1992; Hallström and Riesbeck, 2010; Singh et al., 2010). However, penetration of this pathogen to deep tissues depends on multiple virulence factors (Leroy-Dudal et al., 2004).

Gram-negative pathogens possess a specific group of proteins known as autotransporters, which are translocated to the cell surface by a type V secretion mechanism. Unlike the type I-IV secretory systems that involve multiple proteins, autotransporters are composed of a single protein consisting of an N-terminal signal peptide for secretion, followed by a passenger domain, and a C-terminal translocator (membrane anchoring) domain (Dautin and Bernstein, 2007; Leo et al., 2012). These autotransporters are multifunctional proteins ranging from monomeric to multimeric arrangements (Meng et al., 2011). Recently, the biological role of few autotransporters has been studied in *H. influenzae*. The monomeric autotransporter *Haemophilus* adhesion and penetration protein (Hap) is involved in bacterial aggregation and adherence to host cells (Meng et al., 2011; Hallström et al., 2011). A trimeric autotransporter adhesin (TAA) known as Hia has also been extensively studied for its functional and structural characteristics (Dautin et al., 2007; Spahich and St Geme, 2011). Hia is only present in approximately 25% of the clinical NTHi isolates (St Geme et al., 1998), and cannot be found in encapsulated *H. influenzae*. In contrast, *Haemophilus* surface fibril (Hsf) is present in all typeable strains (St Geme et al., 1996; Rodriguez et al., 2003; Watson et al., 2013). Notwithstanding homologous, Hia and Hsf are relatively different sizewise; Hia has a size of ≈114 kDa (≈342 kDa as a trimer), whereas Hsf is almost twice in size with a monomer of approximately 243 kDa that builds up a ≈750 kDa trimer (Cotter et al., 2005b). These two TAAs are highly homologous at their N- and C-termini with an overall 81% similarity and 72% identity. Moreover, Hia and Hsf are constituted of various repetitive domains, which in parts have been defined according their biological functions. For instance, Hia has two functional host epithelial cell binding domains designated BD1 and BD2. Similar binding domains are present in Hsf, but in addition, Hsf has a third binding domain (BD3). The binding domains are similar in their secondary structures. However, a few key amino acids of the binding pocket of the BD3 of Hsf are different as compared to BD1 and BD2, and therefore BD3 appears not to interact with host cells (Danovaro-Holliday et al., 2008; Spahich and St Geme, 2011).

Previously we reported that Hsf recruits Vn and thus inhibits the lytic pathway of the complement system by inhibiting C5b6-7 complex formation and C9 polymerization (Hallström

et al., 2006). Recently, we studied the Hsf-Vn interaction in detail and found that BD2 of Hsf selectively interacts with the C-terminal end of Vn and thus inhibits MAC formation (Singh et al., 2014). Vn also mediates an increased adherence of Hib to epithelial cells. We proposed that when Vn is bound to BD2 it may function as a bridging molecule between the bacteria and epithelial surface integrins. However, BD1 may directly bind to host epithelium, leading to a stronger Hsf-mediated interaction of Hib with the host cell surface (Singh et al., 2014).

In the present study, we investigated the molecular architecture of Hsf at the bacterial surface. Since the Hsf molecule is an extraordinary large protein and has a repetitive domain structure we modeled this protein by using an *in silico* approach. Our computed model suggested a protein length of approximately 200 nm, whereas the length of Hsf observed by electron microscopy was only 100 nm. Based upon these observations, we analyzed the organisation of Hsf on the bacterial surface by denaturing Hsf using GuHCl in order to unfold the Hsf molecule. In addition, a set of specific anti-Hsf peptide antibodies was included in the analyses to locate the precise regions of the molecule. Our results show that Hsf is not a straight fibre, as reported with several other known bacterial TAAs, but rather consists as a "hairpin-like" twisted molecule.

Material and Methods

Bacterial strains and culture conditions. The type b *H. influenzae* RM804 and mutants (Hallström et al., 2006) were grown in brain heart infusion (BHI) liquid medium containing 10 μg ml^{-1} nicotinamide adenine dinucleotide (NAD) and hemin, or on chocolate agar plates. Cultures were incubated at 37 °C in a humid atmosphere containing 5% CO_2. The *hsf* mutant was grown in BHI supplemented with 18 μg ml^{-1} kanamycin. Luria Bertani (LB) broth or on LB agar plates were used to grow *E. coli* DH5α and *E. coli* BL21 (DE3). *E. coli* harboring expression vectors pET26b-*hsf*(s) and pET16b-hsf^{1-2413} were grown in 50 μg ml^{-1} kanamycin and 100 μg ml^{-1} ampicillin, respectively.

Bioinformatics and protein modeling. An *in silico* model of Hsf was constructed by using a comparative protein structure modeling on domains with a highly homologous amino acid sequence to that of a known three-dimensional structure determined by X-Ray Diffraction Crystallography. Domains with low homology to known three-dimensional structures was built as a trimeric coiled-coil as described elsewhere (Griffiths et al., 2011). The initial alignment from the BLAST/PSI-BLAST homology search was refined using ClustalW alignment and prime STA alignment combined with manual alignment. From the final alignment, an energy-based homo-multimer model was constructed. The resulting model was evaluated to ensure that hydrophilic amino acids were on the surface and hydrophobic amino acids were buried in the protein interior. If these criteria were not sufficiently met, a new model was constructed after sequence re-alignment. The comparative protein structure modeling was performed using Prime version 3.1, which is part of the Schrödinger Suite 2012 (Prime, version 3.1, Schrödinger, LLC, New York, NY, 2012). In addition, UCSF Chimera 1.6 was used for bookkeeping purposes such as the renumbering of amino acid sequences (Pettersen et al., 2004).

Protein expression and purification. Recombinant DNA plasmids were obtained as described elsewhere (Singh et al., 2014). For surface expression of the Hsf protein, pET16-*hsf*54-2300 (Hallström et al., 2006) was freshly transformed into *E. coli* BL21 (DE3) and plated on LB ampicillin plates. Single colonies were inoculated in 10 ml LB containing 100 μg ml^{-1} ampicillin

and incubated at 37°C at 200 rpm shaking. Expression was induced by addition of 0.2 mM IPTG when $OD_{600\ nm}$ reached to 0.4-0.5. Induced cultures were further incubated at 37°C with shaking at 200 rpm for the next 15 h. Expression of Hsf at the surface was verified by using anti-PD antibodies in flow cytometry. For purification purposes, *E. coli* BL21(DE3) containing pET26b-*hsf*(s) was grown in LB medium with kanamycin at 37°C until $OD_{600\ nm}$ reached to 0.8–1. Protein expression was induced by addition of 1 mM IPTG, and protein purification was performed as described elsewhere (Singh and Rohm, 2008). Some Hsf fragments produced inclusion bodies that were purified by using an inclusion body purification protocol (Singh and Rohm, 2008). The purified proteins were dialyzed in PBS and concentrated by using Centricon cartridge (Millipore, Bedford, MA). Purity of the proteins was confirmed by SDS-PAGE and concentrations were estimated by a Nano-drop spectrophotometer (Thermo Scientific, Wilmington, DE) and also verified by a Bicinchoninic acid (BCA) assay (Pierce, Rockford, IL).

Production of antibodies. Specific peptides covering the sequence Hsf 153-164, anti-Hsf 876-881 (anti-KTRAAS), anti-Hsf 1912-1942 were synthesized and conjugated with KLH (Innovagen, Lund, Sweden). For immunization, 500 µl peptides (200 µg) were mixed with 500 µl complete Freunds adjuvant followed by immunization of rabbits. After 4 weeks, the same amount of peptides were injected with aluminium hydroxide as adjuvant. Two boosters were administered after 5 and 6 weeks. Finally, blood was drawn from immunized rabbits and antibodies (Abs) from sera were purified by standard affinity purification protocol (Singh et al., 2013). BD and PD domains were aligned and modeled to identify the conserved surface exposed regions (Fig. 1B-C and Fig. S1). Peptide antibodies against the BD and PD domains were produced in rabbits and purified by affinity purification (Genscript, NJ). Antibodies were concentrated to 1 mg/ml in PBS and stored at -20 °C.

Western blotting. Purified Hsf fragments (50 ng) were mixed with SDS-PAGE loading dye, incubated at 95ºC for 5 min and centrifuged (14,000xg) for 5 min. Supernatants were loaded on 4-12% Bis-tris gels (NuPAGE; Invitrogen, Carlsbad, CA) and resolved at 80 V until the dye front reached the bottom of gels. Proteins were transferred to a PVDF membrane for 15 h followed by blocking with 5% milk in PBS. Blots were incubated with anti-PD, anti-BD, anti-$Hsf^{153-164}$, anti-$Hsf^{876-881}$, anti-$Hsf^{1912-1942}$ Abs or a combination of Abs as indicated. After 1 h of incubation at RT, blots were washed 4 times in PBS containing 0.05% Tween-20 (PBS-T). Thereafter, secondary anti-rabbit swine polyclonal Abs (Dako, Glostrup, Denmark) were added to the blot in PBS containing 5% milk and incubated for 1 h at RT. Finally, blots were washed 4 times in PBS-T and developed using an ECL western blotting kit (Pierce, Rockford, IL).

Circular dichroism (CD). Purified recombinant $Hsf^{54-2300}$ was dialysed against 25 mM phosphate buffer, pH 7.5 containing 100 mM sodium fluoride. Two samples of $Hsf^{54-2300}$ (0.2 mg/ml) were prepared, one in buffer and one in buffer with 6 M GuHCl. Spectra (15x) were recorded for these two samples between 210-250 nm in a Jasco J-815CD spectrophotometer (Jasco, Easton, MD) in a 1 mm quartz cuvette thermostated to 25 °C, and the mean spectrum was plotted. Each spectrum was subtracted from a control spectrum obtained in the absence of protein. The two solutions were mixed in different proportions to prepare a set of solutions containing 0.25-5.75 M GuHCl in steps of 0.25 M. For each these solutions, and the two stock solutions, the ellipticity at 222 nm was recorded during 2 min. The following equation was fitted to the average signal versus GuHCl concentration:

$$y = \frac{y_f^0 + m_f[\mathrm{GuHCl}] + \left(y_u^0 + m_u[\mathrm{GuHCl}]\right)\exp\left(-\frac{1}{RT}\left(\Delta G^0(H_2O) - m[\mathrm{GuHCl}]\right)\right)}{1 + \exp\left(-\frac{1}{RT}\left(\Delta G^0(H_2O) - m[\mathrm{GuHCl}]\right)\right)} \qquad \text{(Equation 1)}$$

The six fitted parameters are the free energy of unfolding in the absence of denaturant, $\Delta G^0(H_2O)$, and its dependence on denaturant concentration, m, the intercept, y_f^0, and slope, m_f, of the naive state baseline, and the intercept, y_u^0, and slope, m_u, of the unfolded state baseline. The GuHCl concentration at which 50% of the protein molecules are denatured, C_M, is calculated from the fitted parameters as

$$C_M = \Delta G^0(H_2O)/m \qquad \text{(Equation 2)}$$

For presentation, the data are normalized to the apparent fraction folded, F_{app}, according to

$$F_{\mathrm{app}} = \frac{y - y_u}{y_f - y_u} \qquad \text{(Equation 3)}$$

Transmission electron microscopy (TEM). Different Abs were labelled with colloidal gold as described (Roth, 1996). The wild type *H. influenzae* RM804 were grown in BHI for 3 h at 37 ºC. *E. coli* were induced with IPTG to express Hsf, and the expression of the protein at the surface was verified by flow cytometry prior to TEM. Bacteria were treated with GuHCl and washed twice in PBS. Thereafter, bacteria were fixed in 2.5% glutaraldehyde, dehydrated and embedded in Epon as described earlier (Oehmcke et al., 2009). For negative staining, bacteria were incubated with 5 nm and 10 nm gold-conjugated antibodies as described, fixed in PBS containing 4% paraformaldehyde and 0.1% glutaraldehyde, and prepared as described (Carlemalm, 1990). TEM was performed as described elsewhere (Bengtson et al., 2008), and specimens were examined in a JEOL JEM 1230 transmission electron microscope (JEOL, Peabody, MA) at 60 kV accelerating voltage. The Images were recorded with a Gatan Multiscan 791 CCD camera (Gatan, Pleasanton, CA). Only particles observed within a distance of 15 nm or less adjacent to the cell surface were counted. This corresponds to the established maximum distance between an IgG and its antigen, or between a protein labeled with gold of this size, and its target.

Results

Hsf is approximately a 200 nm long fibril according to a modelled structure. Autotransporters are very difficult to crystallize, and therefore most structures have been solved in fragments and followed by compilation into full models (Agnew et al., 2011; Hartmann et al., 2012). In order to model the Hsf structure, we first predicted different structural motifs/domains by the DATAA server at Max Planck Institute for developmental Biology, Tubingen, Germany (http://toolkit.tuebingen.mpg.de/dataa/search). Our prediction revealed the presence of an N-terminal signal sequence (1-43 aa) required for protein translocation to the membrane, 14 distinguished Trp-ring domains, 3 GANG domains, 3 Neck/IS-Neck domains, 4 KG motifs, 1 Y head, 1 TTT, and finally a highly conserved membrane anchoring domain (2342-2413 aa) at the C-terminus of Hsf (Fig. 1A). We thereafter modeled various regions of the Hsf molecule by using known templates and finally the full structure was compiled (Fig. 1A).

Previously, on the basis of interactions with Chang conjunctival epithelial cells, three binding domains ($BD2^{529-652}$, $BD3^{1206-1337}$ and $BD1^{1896-2022}$) have been characterized in the Hsf molecule (Cotter et al., 2005b). All binding domains consist of an N-terminal Neck domain and a C-terminal Trp-ring domain "N-Neck : Trp ring-C" (Fig. 1A). The motifs arrangement showed that within the N-terminus of the BDs, an additional Trp ring domain is present and assembled in an "N-Trp ring : Neck : Trp ring-C" triplet arrangement (Fig. 1A). In parallel to Hsf, Hia has a similar N-terminal Trp-ring domain within BD1, whereas it is absent in the BD2. We included these N-terminal Trp-ring domains along with BDs (N-Trp ring : Neck : Trp ring-C) in our study in order to analyze the biological function, that is, Vn binding of these domain triplets. Hence, the domains $BD2^{429-652}$, $BD3^{1103-1338}$, and, $BD1^{1792-2022}$ were expressed and purified. The alignment of the BDs showed 48.7-54.7% identity and 60.8-61.9% similarity (Fig. S1). The three BDs were modeled against the template 1s7M and superimposed (Fig. 1B). Importantly, the conserved region of the Neck domains, the peptide sequence N-TKDGISAGNKAITNVAS-C was identified and selected for Ab production (Fig. 1B and S1).

In addition to BDs, other structural Hsf motifs are organised in a distinguished series that results in three putative domains (PD) without any known biological function, that is, $PD2^{272-375}$, $PD3^{938-1046}$, and $PD1^{1637-1740}$ with 54.1-74.5% identity and 65.8-80.2% sequence similarity, respectively (Fig. S1). The binding domain numbering of Hsf was assigned on the basis of the homology with Hia (Cotter et al., 2005b). Following the previous nomenclature, we named the adjacent PDs with a similar numbering (Fig. 1A). The PDs consist of a KG domain followed by a Trp-ring domain (N-KG : Trp ring-C). A KG domain is also present between amino acids 2063-2112 (Fig. 1A), which has a more variable sequence than the other 3 KG domains.

PDs were aligned for identifying regions with unique amino acids for design of peptides to produce specific Abs that could recognize all three PDs (Fig. S1). The PDs were also modeled against the template (PDB: 3emi) and model structures were superimposed (Fig. 1C). Importantly, the amino acid sequence of the KG domains, N-GLVTAKAVIDAVNKAGWR-C, was conserved, exposed on the surface, and was therefore selected for Ab production. In structural models, highly surface-exposed regions were variable in sequences of the BD and PD, as compared by aligning the models (Fig. 1B-C and Fig. S1). Therefore, it was almost impossible to select highly surface exposed regions for production of common antibodies that would recognize all three sequences on the fibril at the same time. Hence, we selected the conserved and maximally exposed sequences for designing Abs that would recognize the PDs and BDs in the fibrils (Fig. S1).

The amino acid sequence of the C-terminal membrane anchoring domain and the trimeric coiled coil domain connected to it (2308-2413 aa) completely matched the PDB: 3emo sequence. The PD1, PD2 and PD3 have a high sequence identity with PDB:3emi, whereas the BD1, BD2, and BD3 have high sequence identity with PDB:1s7m. In our models 6 out of 14 Trp-rings predicted by the DATAA server were thus modeled using those templates. The remaining 8 Trp-ring domains present in the Hsf molecule were modelled using templates PDB:1s7m, PDB: 3emi, or PDB: 3d9x, with 15-26% sequence identity, and 30-43% sequence similarity. A domain close to the C-terminal (2144-2315 aa) was constructed from PDB:3laa with low identity, but having a high local similarity. The remaining regions that were not matching with known three-dimensional structures or had low identity/similarity to known templates were constructed as trimeric coiled coils (Griffiths et al., 2011). The trimeric models were first refined with Prime side chain prediction, and strained loops were refined using Prime loop prediction, followed by a constrained Prime energy refinement (imperf minimization). Finally, the 23 domain models were

joined to a continuous 0136-2413 aa Hsf-model (Fig. 1A), and the joints were refined with Prime loop prediction and Prime energy refinement. The length of the Hsf-model measured on the symmetry axis was found to be 1675 Å (167.5 nm) from the centroid of the C-alpha of asp136 to the centroid of the C-alpha of gln2371. Nevertheless, this measurement (167.5 nm) did not include the length of 54-135 amino acids of the N-terminal (not modeled) and 2372-2413 amino acids of the C-terminal (membrane anchor). Therefore, we assume that the approximate total length of Hsf (54-2413 amino acids) might reach up to 200 nm.

Partial denaturation suggests a bended and twisted Hsf molecule. We performed TEM to visualise the Hsf molecule on the surface of *H. influenzae* RM804 (Fig. 2A; upper left panel, untreated). At least 50 individual fibrils were measured at the bacterial surface. The mean length of the Hsf^{1-2413} molecule was found to be approximately 100 nm. Theoretically, each Hsf monomer consisted of 2413 amino acids and if monomers trimerize as a straight fibril, the length would be more than 100 nm (Fig. 2A). In support of this, the calculated length of Hsf in our modeled structure was 1675 Å (167.5 nm), *i.e.*, the length of the fibril from Asp136 to Gly2371 (Fig. 1A). The visualized length of the fibre in TEM was less in size than the calculated length of the model. This finding prompted us to in detail investigate this trimeric molecule on the surface of *H. influenzae*. For this purpose, we treated *H. influenzae* RM804 with 0.5-2.5 M guanidine hydrochloride (GuHCl) and visualized the Hsf molecule by using negative staining and TEM. The experiments revealed a stability of the Hsf fibril in up to 1.0 M GuHCl (Fig. 2A). However, bacteria treated with 1.5 M GuHCl showed a partial unwinding of the fibril, whereas the quarternary structure was further open, and more than three fibrils appeared at >2.0 M GuHCl (Fig. 2A). This result clearly indicated that the Hsf molecule at the surface of *H. influenzae* does not consist of three single strands only in the form of straight fibril.

In parallel, we performed an equilibrium unfolding experiment with recombinant $Hsf^{54-2300}$ in the presence of the denaturing agent GuHCl. Circular dichroism (CD) spectra were recorded in the absence (native protein) and presence of 6 M GuHCl. The large change in signal at 222 nm reflected denaturation of the secondary structure (Figure 2B). The ellipticity at 222 nm was measured as a function of GuHCl concentration. The data was fitted by equation 1, allowing us to estimate, and presented in Figure 2C as normalised data (equation 3). C_M, the GuHCl concentration at which 50% of the protein is unfolded is estimated to 2.6 M (equation 2), The denaturation data are in agreement with the TEM data (fig. 2A). By TEM only folded Hsf is detected up to 1 M GuHCl, which corresponds to the pre-transition baseline (Figure 2C). To confirm the above observations, we performed embedded sectioning of *H, influenzae* RM804 WT, RM804Δ*hsf*, *E. coli* expressing Hsf, and an *E. coli* control containing an empty vector. Each sample was incubated in PBS or PBS containing 6M GuHCl, washed twice in PBS and fixed. Bacterial pellets were embedded, and sections were produced in a microtome. Only *H. influenzae* RM804 WT and *E. coli* expressing Hsf showed a layer of Hsf at their surface (Fig. 3A). Interestingly, when RM804 WT and *E. coli* expressing Hsf were treated with 6M GuHCl for 30 min, the surface proteins appeared as a thick fuzzy layer, approximately with a two-fold increase in thickness in comparison to non-treated bacteria. In contrast, treatment of RM804Δ*hsf* and the *E. coli* control with 6M GuHCl did not show any Hsf at the surface (Fig. 3B). In the next step, those EM sections were visualized at a higher magnification to observe the fibrils. Hsf was clearly seen at the surface with an approximate length of 100 nm in *H. influenzae* RM804 WT and Hsf-expressing *E. coli* (Fig. 3C). Treatment with 6M GuHCl revealed an increased fibre length that was more than 100 nm (Fig 3C, lower panel). Further magnification of the bacterial

surface fibrils (*E. coli*-Hsf) showed a clear difference between the length of the fibril in untreated and 6M GuHCl-treated samples (Fig. 3D). These results thus indicated that Hsf is not a straight fibre. It is bended and two trimeric halves may result in a hairpin-like shape with twisting of each half of the molecule.

Anti-Hsf peptide antibodies are highly specific. Hsf has several repetitive sequence domains as demonstrated by bioinformatics and modeling (Fig. 1A). Some of these sequence motifs in a particular combination serve as functional binding domains. The trimeric organisation of these various domains in a large protein like Hsf (2,413 amino acids) has not yet been elucidated or predicted. It is thus interesting to observe how these domains position in the fibril since we suggest that Hsf is a bended and twisted molecule (Fig. 3). The distribution of PDs and BDs were analyzed by using immunogold cryo TEM. Peptide Abs directed against "GLVTAKAVIDAVNKAGWR" (anti-PD Ab), and "TKDGISAGNKAITNVAS" (anti-BD Ab) (Fig. S1), peptide specific anti-Hsf$^{153-164}$, anti-Hsf$^{876-881}$(KTRAAS), and anti-Hsf$^{1912-1942}$ Abs were produced in rabbits. Hsf fragments, BD1-3, PD1-3, Hsf54-608, Hsf608-1351, Hsf1047-1751, Hsf1536-2031, and full length (Hsf54-2300) were tested for Ab cross-reactivity. The anti-Hsf$^{153-164}$ Ab recognized Hsf54-608 and the full length Hsf molecule (Fig. 4A). Blots treated with anti-Hsf$^{153-164}$ + anti-PD Abs recognized all three PDs, but did not react with BDs. Hsf1536-2031 was not recognized by the anti-PD Ab. However, PD1 encompassing amino acids Hsf1637-1740 (*i.e.*, part of the Hsf1536-2031) was easily recognized by anti-PD Abs. This may be due to partial masking of the Ab recognition sequence of Hsf1536-2031 in the blots tested and/or in combination with a low affinity of interaction (Fig. 4A). The anti-Hsf$^{876-881}$Ab crossreacted with Hsf608-1351 and the full-length molecule (Fig. 4B). Incubation of the same blot with anti-Hsf$^{876-881}$ + anti-BD Ab detected all BD domains, and in addition to the other protein fragments. Anti-BD Ab did not cross-react with any of the PD domains (Fig. 4B, Fig S2). Yet another antibody, anti-Hsf$^{1912-1942}$ Ab detected BD1 (Hsf1536-2031) and full length protein (Fig. 4C). These results suggested that the Abs are highly specific for recognizing the corresponding amino acid sequences (Fig. 4D).

Antibody mapping also demonstrated a bending of the Hsf molecule. We used TEM in order to visualize various regions of the Hsf molecules. In the first set of experiments, anti-PD and anti-BD Abs were labeled with 10 nm colloidal gold particles and the other specific Abs were labeled with 5 nm gold particles. Thereafter, *H. influenzae* RM804 WT in parallel with Hsf-expressing *E. coli* were probed with Abs in combination as shown to the left and at the top of Fig. 5A and B. The anti-Hsf$^{153-164}$ Ab recognized the fibril within 10 nm length at the base from the surface. In support of the previous results (Fig. 2A, Fig. 3), we here proved that the N-terminal sequence was located close to the base of the fibril, suggesting that the Hsf trimer was a bended molecule. Of notice, the N-terminus in a straight fibril structure should be located at the tip of the fibril (Fig. 1A). As can be seen in Fig. 5A, the anti-Hsf$^{876-881}$ Ab bound to the protein fibril almost at the tip, that is, 80-90 nm from the base. However, probing Hsf with the anti-Hsf$^{1912-1942}$ Ab revealed an interaction at a 35 nm distance from the base of the fibril, and thus verified the C-terminal end near the base (Fig. 5A-B). The relative Ab-binding as revealed by TEM was plotted against the length of the Hsf, and a diagram representing the overall distribution of Abs recognizing the various sequences is shown in Fig. 5C.

To characterize the precise positions of PD and BD domains in the fibril, we probed *H. influenzae* RM804 WT and Hsf-expressing *E. coli* with both anti-PD and anti-BD Abs (10 and 5

nm gold particles, respectively) at a ratio 1:1. The results showed an interaction of the 10 nm gold particles at three distinguished places on the fibril. Similarly, 5 nm gold particles also interacted at three different locations on the fibril (Fig. 6A-B). Several individual fibrils were measured, and the distribution of gold particles in relation to the full length of the fibril was calculated. The results showed that PD2 and BD1 were present at 30-35 nm, BD2 and PD1 at 60-70 nm, and, finally, PD3 and BD2 were located at a 90-95 nm distance from the base of the fibril (Fig. 6C). Taken together, the position of our Abs suggested approximate parallel distribution of PD and BD on the Hsf molecule, and clearly demonstrated that the fibril is bent back from the middle portion of the protein (Fig. 6C).

Discussion

Hsf is an important virulence factor of *H. influenzae* and is present in all encapsulated clinical isolates. Among several known bacterial trimeric autotransporters, Hsf is unique for its structural appearance (Cotter et al., 2005b). It is not a trimeric straight fiber like protein, rather a folded and twisted molecule at the bacterial surface. A prototypical autotransporter such as *Yersinia* adhesion (YadA) and ubiquitous surface proteins (Usp) A1/A2 of *M. catarrhalis* appear as straight "lollipop" like structures consisting of a membrane-anchoring domain, a stalk and finally a head (Nummelin et al., 2004). On the other hand, some of the autotransporters may appear like a fibril, and thus may not be similar to a prototypical TAA (Cotter et al., 2005a). Previously, we noticed a similar double-fold in a large TAA, *i.e., Moraxella catarrhalis* IgD binding protein (MID). In parallel to Hsf, MID is also an extraordinary large protein (Hallstrom et al., 2008). Interestingly, the Hsf homologue Hia having 70% similarity is present in NTHi, and is only 1000 amino acids in length and not bended and twisted like Hsf (Spahich and St Geme, 2011). Most trimeric autotransporters are straight fibrils regardless of their large sizes. For example, Ishikawa and co-workers reported that the AtaA autotransporter of *Acinetobacter* sp. that is constituted of 3630 amino acids appears like a straight fibril structure (Ishikawa et al., 2014). The BadA protein of *Bartonella henselae* similarly appears as a straight fibril (≈250 nm) that is composed of 3082 amino acids (Muller et al., 2011).

Interestingly, the straight structure of some of the TAAs is dynamically modified once they bind to the host ligands. For example, UspA1 showed a bending after binding to CAECAM-1 and fibronectin (Agnew et al., 2011). Similarly, the *E. coli* autotransporter EibD also bends from a saddle-like structure after binding to its ligand IgG (Leo et al., 2011). This kind of bending has only been observed after interaction with host ligands. On the other hand, the biological significance of bending and twisting of the bacterial autotransporters such as MID and Hsf is currently not understood. The folding might, however, have certain benefits. Firstly, the large Hsf-like fibril (≈200 nM in length) may be very fragile that might easily shear. The winding of two trimeric halves of the fibrils make it stiffer and provide mechanical strength to the fiber. It will be more stable during bacterial mobility and host interaction. Secondly, the trimeric structure of the fibril constitutes functional domains. The double intertwining of the trimeric fibre might generate new intramolecular interactions that may lead to the formation of several new functional domains around the fibril. In Hsf, the core of the repetitive domains is conserved, whereas the surface-exposed regions of these domains are variable as explained in our model (Fig. 1B-C). Thus the conserved core provides a structural stability to the Hsf molecule, while variable surface exposed regions might contribute to different biological functions.

We modeled 136-2413 amino acids of Hsf by using different suitable templates and then finally joined all models together to build the whole fibril (Fig. 1A). The total length of the model was measured from the centroid of the C-alpha of Asp136 to the centroid of the C-alpha of Gln2371 that accounted for 1675 Å (167.5 nm). In this model some important issues have been considered during the comparison of model length with the estimated length in TEM. The Hsf 2372-2413 region is present inside the outer membrane, hence it will not contribute to the visual length at the surface of bacteria during analysis in TEM. Importantly, these membrane-spanning amino acids are not included in the 1675 Å length. In Hsf, the 55-135 aa do not have any suitable crystal structure to be used as a template. Therefore, the 54-135 aa were not included in the model, and these amino acids will definitely contribute to the length of the fibril. In this model, however, only partial structures that have very poor template similarity were modeled as a coiled-coil as described elsewhere (Griffiths et al., 2011). The axial rise for an alpha-helical coiled coil is 1.5 Å per amino acid. The 3emi template (similar to PDs) was found to have an axial rise per amino acid of 0.55 Å. This indicates that Hsf aa 136-2371 is longer than 123 nm even if it is modeled by 3emi template, and less than 335 nm, also if those regions are put as a coiled coil shape. Thus, the approximate visual length of Hsf under natural conditions as revealed by TEM was ≈100 nm, and after complete denaturation this might have a length of ≈200 nm. The modeled Hsf (136-2371 aa) structure is 167 nm that is matching with the length of protein observed in TEM.

Our TEM data showed that bending only does not take place in the Hsf fibre. For instance, in some of the TAAs, interaction of ligands causes a bend in the trimer (Agnew et al., 2011; Leo et al., 2011). Unlike those proteins, the Hsf fibril is twisted after bending and the bended region may have loose or tight interactions with the first half part of the fibre. Twisting of two trimers is evident from our experiments with bacteria incubated in the presence of GuHCl, and here unwinding of two trimers was observed at high concentrations of GuHCl (Fig. 2 and 3). This intertwining between two halfes of the same trimers may thus generate a much more complicated structure.

In the present paper we used protein unfolding and TEM to demonstrate that Hsf is a twisted “hairpin-like” trimeric autotransporter. Our study provides new knowledge on the structural arrangement of the Hsf molecule at the bacterial surface. It will be a challenge to prove the biological significance of this particular folding, and it also remains to study the ultrastructure of this interesting protein to present a full crystal structure model.

Acknowledgements

Potential conflicts of interest: none reported.

Financial support: This work was supported by grants from Foundations of Alfred Österlund, Anna and Edwin Berger, Greta and Johan Kock, Åke Wiberg, Torsten Söderberg, Lars Hierta Foundation, O. E. och Edla Johansson, Kungliga Fysiografiska Sällskapet, as well as the Swedish Medical Research Council (grant number 521-2010-4221 and K2012-66X-14928-09-5, www.vr.se), the Cancer Foundation at the University Hospital in Malmö, the Physiographical Society (Forssman’s Foundation), nanometer structure consortium at Lund University, and Skåne County Council’s research and development foundation.

References

Agnew, C., Borodina, E., Zaccai, N.R., Conners, R., Burton, N.M., Vicary, J.A., Cole, D.K., Antognozzi, M., Virji, M., Brady, R.L., 2011. Correlation of in situ mechanosensitive responses of the *Moraxella catarrhalis* adhesin UspA1 with fibronectin and receptor CEACAM1 binding. Proc. Natl. Acad. Sci. U S A. 108, 15174-15178.

Agrawal, A., Murphy, T.F., 2011. *Haemophilus influenzae* infections in the H. influenzae type b conjugate vaccine era. J. Clin. Microbiol. 49, 3728-3732.

Bengtson, S.H., Eddleston, J., Morgelin, M., Zuraw, B.L., Herwald, H., 2008. Regulation of kinin B(2) receptors by bradykinin in human lung cells. Biol. Chem. 389, 1435-1440.

Carlemalm, E., 1990. Lowicryl resins in microbiology. J. Struct. Biol. 104, 189-191.

Cotter, S.E., Surana, N.K., St Geme, J.W., 3rd, 2005a. Trimeric autotransporters: a distinct subfamily of autotransporter proteins. Trends Microbiol. 13, 199-205.

Cotter, S.E., Yeo, H.J., Juehne, T., St Geme, J.W., 3rd, 2005b. Architecture and adhesive activity of the *Haemophilus influenzae* Hsf adhesin. J. Bacteriol. 187, 4656-4664.

Danovaro-Holliday, M.C., Garcia, S., de Quadros, C., Tambini, G., Andrus, J.K., 2008. Progress in vaccination against *Haemophilus influenzae* type b in the Americas. PLoS Med. 5, e87.

Dautin, N., Barnard, T.J., Anderson, D.E., Bernstein, H.D., 2007. Cleavage of a bacterial autotransporter by an evolutionarily convergent autocatalytic mechanism. Embo J. 26, 1942-1952.

Dautin, N., Bernstein, H.D., 2007. Protein secretion in gram-negative bacteria via the autotransporter pathway. Annu. Rev. Microbiol. 61, 89-112.

Fitzwater, S.P., Watt, J.P., Levine, O.S., Santosham, M., 2010. *Haemophilus influenzae* type b conjugate vaccines: considerations for vaccination schedules and implications for developing countries. Hum. Vaccin. 6, 810-818.

Geme, J.W., 3rd, 1996. Molecular determinants of the interaction between *Haemophilus influenzae* and human cells. Am. J. Respir. Crit. Care. Med. 154, S192-196.

Geme, J.W., 3rd, Cutter, D., 1995. Evidence that surface fibrils expressed by *Haemophilus influenzae* type b promote attachment to human epithelial cells. Mol. Microbiol. 15, 77-85.

Griffiths, N.J., Hill, D.J., Borodina, E., Sessions, R.B., Devos, N.I., Feron, C.M., Poolman, J.T., Virji, M., 2011. Meningococcal surface fibril (Msf) binds to activated vitronectin and inhibits the terminal complement pathway to increase serum resistance. Mol. Microbiol. 82, 1129-1149.

Hallstrom, T., Muller, S.A., Morgelin, M., Mollenkvist, A., Forsgren, A., Riesbeck, K., 2008. The *Moraxella* IgD-binding protein MID/Hag is an oligomeric autotransporter. Microbes Infect. 10, 374-381.

Hallstrom, T., Riesbeck, K., 2010. *Haemophilus influenzae* and the complement system. Trends Microbiol. 18, 258-265.

Hallstrom, T., Trajkovska, E., Forsgren, A., Riesbeck, K., 2006. *Haemophilus influenzae* surface fibrils contribute to serum resistance by interacting with vitronectin. J. Immunol. 177, 430-436.Hallström, T**.**, Singh, B**.**, Resman, F**.**, Blom, A. M**.**, Mörgelin, M**.**, Riesbeck**,** K., 2011. *Haemophilus influenzae* protein E binds to the extracellular matrix by concurrently interacting with laminin and vitronectin. J**.** Infect**.** Dis**.** 204(7):1065-74.

Hartmann, M.D., Grin, I., Dunin-Horkawicz, S., Deiss, S., Linke, D., Lupas, A.N., Hernandez Alvarez, B., 2012. Complete fiber structures of complex trimeric autotransporter adhesins conserved in enterobacteria. Proc. Natl. Acad. Sci. U S A. 109, 20907-20912.

Ishikawa, M., Shigemori, K., Hori, K., 2014. Application of the adhesive bacterionanofiber AtaA to a novel microbial immobilization method for the production of indigo as a model chemical. Biotechnol. Bioeng. 111, 16-24.

Leo, J.C., Grin, I., Linke, D., 2012. Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367, 1088-1101.

Leo, J.C., Lyskowski, A., Hattula, K., Hartmann, M.D., Schwarz, H., Butcher, S.J., Linke, D., Lupas, A.N., Goldman, A., 2011. The structure of E. coli IgG-binding protein D suggests a general model for bending and binding in trimeric autotransporter adhesins. Structure 19, 1021-1030.

Leroy-Dudal, J., Gagniere, H., Cossard, E., Carreiras, F., Di Martino, P., 2004. Role of alphavbeta5 integrins and vitronectin in *Pseudomonas aeruginosa* PAK interaction with A549 respiratory cells. Microbes Infect. 6, 875-881.

Meng, G., Spahich, N., Kenjale, R., Waksman, G., St Geme, J.W., 3rd, 2011. Crystal structure of the *Haemophilus influenzae* Hap adhesin reveals an intercellular oligomerization mechanism for bacterial aggregation. Embo J. 30, 3864-3874.

Morris, S.K., Moss, W.J., Halsey, N., 2008. *Haemophilus influenzae* type b conjugate vaccine use and effectiveness. Lancet Infect. Dis. 8, 435-443.

Muller, N.F., Kaiser, P.O., Linke, D., Schwarz, H., Riess, T., Schafer, A., Eble, J.A., Kempf, V.A., 2011. Trimeric autotransporter adhesin-dependent adherence of *Bartonella henselae*, *Bartonella quintana*, and *Yersinia enterocolitica* to matrix components and endothelial cells under static and dynamic flow conditions. Infect. Immun. 79, 2544-2553.

Nummelin, H., Merckel, M.C., Leo, J.C., Lankinen, H., Skurnik, M., Goldman, A., 2004. The *Yersinia adhesin* YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. Embo J. 23, 701-711.

Oehmcke, S., Morgelin, M., Herwald, H., 2009. Activation of the human contact system on neutrophil extracellular traps. J. Innate. Immun. 1, 225-230.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605-1612.

Rodriguez, C.A., Avadhanula, V., Buscher, A., Smith, A.L., St Geme, J.W., 3rd, Adderson, E.E., 2003. Prevalence and distribution of adhesins in invasive non-type b encapsulated *Haemophilus influenzae*. Infect. Immun. 71, 1635-1642.

Roth, J., 1996. The silver anniversary of gold: 25 years of the colloidal gold marker system for immunocytochemistry and histochemistry. Histochem. Cell. Biol. 106, 1-8.

Singh, B., Al-Jubair, T., Morgelin, M., Thunnissen, M.M., Riesbeck, K., 2013. The Unique Structure of *Haemophilus influenzae* Protein E Reveals Multiple Binding Sites for Host Factors. Infect. Immun. 81, 801-814.

Singh, B., Su, Y. C., Riesbeck, K., 2010. Vitronectin in bacterial pathogenesis: a host protein used in complement escape and cellular invasion. Mol. Microbiol. 78:545-560.

Singh, B., Fleury, C., Jalalvand, F., Riesbeck, K., 2012. Human pathogens utilize host extracellular matrix proteins laminin and collagen for adhesion and invasion of the host. FEMS Microbiol. Rev. 36, 1122-1180.

Singh, B., Rohm, K.H., 2008. Characterization of a *Pseudomonas putida* ABC transporter (AatJMQP) required for acidic amino acid uptake: biochemical properties and regulation by the Aau two-component system. Microbiology 154, 797-809.

Singh, B., Su, Y.C., Al-Jubair, T., Mukherjee, O., Hallstrom, T., Morgelin, M., Blom, A.M., Riesbeck, K., 2014. A fine-tuned interaction between the trimeric autotransporter *Haemophilus* surface fibrils and vitronectin leads to serum resistance and adherence to respiratory epithelial cells. Infect Immun. 82, 2378-89.

Spahich, N.A., St Geme, J.W., 3rd, 2011. Structure and Function of the *Haemophilus influenzae* Autotransporters. Front. Cell. Infect. Microbiol. 1, 5.

St Geme, J.W., 3rd, Cutter, D., Barenkamp, S.J., 1996. Characterization of the genetic locus encoding *Haemophilus influenzae* type b surface fibrils. J. Bacteriol. 178, 6281-6287.

St Geme, J.W., 3rd, Kumar, V.V., Cutter, D., Barenkamp, S.J., 1998. Prevalence and distribution of the hmw and hia genes and the HMW and Hia adhesins among genetically diverse strains of nontypeable *Haemophilus influenzae*. Infect. Immun. 66, 364-368.

Ulanova, M., Tsang, R.S., 2009. Invasive *Haemophilus influenzae* disease: changing epidemiology and host-parasite interactions in the 21st century. Infect. Genet. Evol. 9, 594-605.

Watson, M.E., Jr., Nelson, K.L., Nguyen, V., Burnham, C.A., Clarridge, J.E., Qin, X., Smith, A.L., 2013. Adhesin genes and serum resistance in *Haemophilus influenzae* type f isolates. J. Med. Microbiol. 62, 514-524.

Winkelstein, J.A., Moxon, E.R., 1992. The role of complement in the host's defense against *Haemophilus influenzae*. J. Infect. Dis. 165 Suppl 1, S62-65.

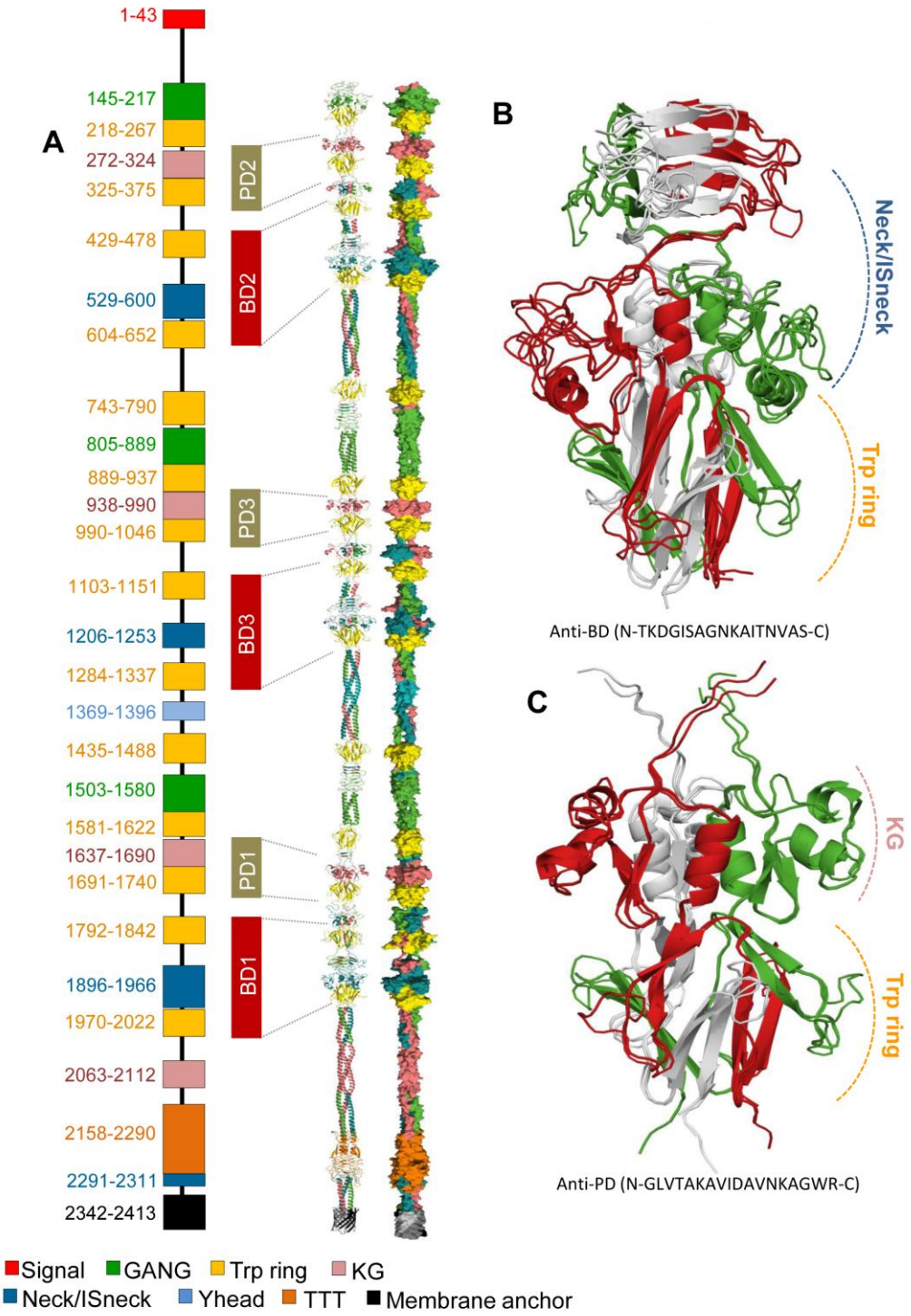

Fig. 1. Modelling of the full length Hsf molecule. (A) Prediction of repetitive motifs in the Hsf molecule by using the domain annotation of trimeric autotransporter adhesins (DATAA) server at Max Planck Institute for Developmental Biology, Tubingen, Germany (http://toolkit.tuebingen.mpg.de/dataa/search). Domains are indicated according to their amino acid numbers, and the BD and PD are named according to the previously described nomenclature (Cotter *et al.*, 2005). (B) A superimposed cartoon model that demonstrates putative domains (3x) present at the regions Hsf 263-376, 938-1050 and 1630-1741 modeled by using template 3emi. (C) Superimposed model that shows binding domains (3x) present at the Hsf regions 500-652, 1173-1338 and 1863-2023 modeled by

using template 1s7m. The amino acid regions encompassing PD and BD domains were aligned (Fig. S1), and a conserved surface exposed region was selected for production of peptides that were used for immunization of rabbits. These conserved sequences are shown at the bottom of panels (B-C). PDB templates, 1S7M, 3emi, and 3 emo; Hia of *H. influenzae*, 2qih; UspA1 of *Moraxella catarrhalis*, 3d9x; BadA of *Bartonella henselae*, 3laa; *Burkholderia pseudomallei*. Homology modeling was performed by using different templates available in the PDB data base. The templates are presented as Hsf amino acids (PDB code of template: identity/similarity, gap), 0136-0269 (1s7m: 23/35, 4), 0263-0376 (3emi: 63/73, 12), 377-0481(3emi: 16/31, 5), 0482-0499 (2qih: 22/38, 0), 0500-0652 (1s7m: 54/68, 2), 0653-0742 (2qih: 15/26, 0), 0743-0858 (3d9x: 26/43, 11), 0859-0876 (2qih: 27/44, 0), 0877-0937 (3emi: 22/31, 7), 0938-1050 (3emi: 45/56, 11), 1051-1155 (3emi: 21/33, 4), 1156-1172 (2qih: 23/35, 0), 1173-1338 (1s7m: 47/58, 8), 1339-1433 (2qih: 15/30, 0), 1434-1558 (3d9x: 17/32, 12), 1559-1578 (2qih: 23/33, 0), 1579-1629 (3d9x: 15/33, 3), 1630-1741 (3emi: 71/77, 3), 1742-1862 (3emi: 21/34, 22), 1863-2023 (1s7m: 73/79, 3), 2024-2143 (2qih: 11/23, 0), 2144-2315 (3laa: 15/22, 21), and 2308-2413 (3emo: 100/100, 0*)*.

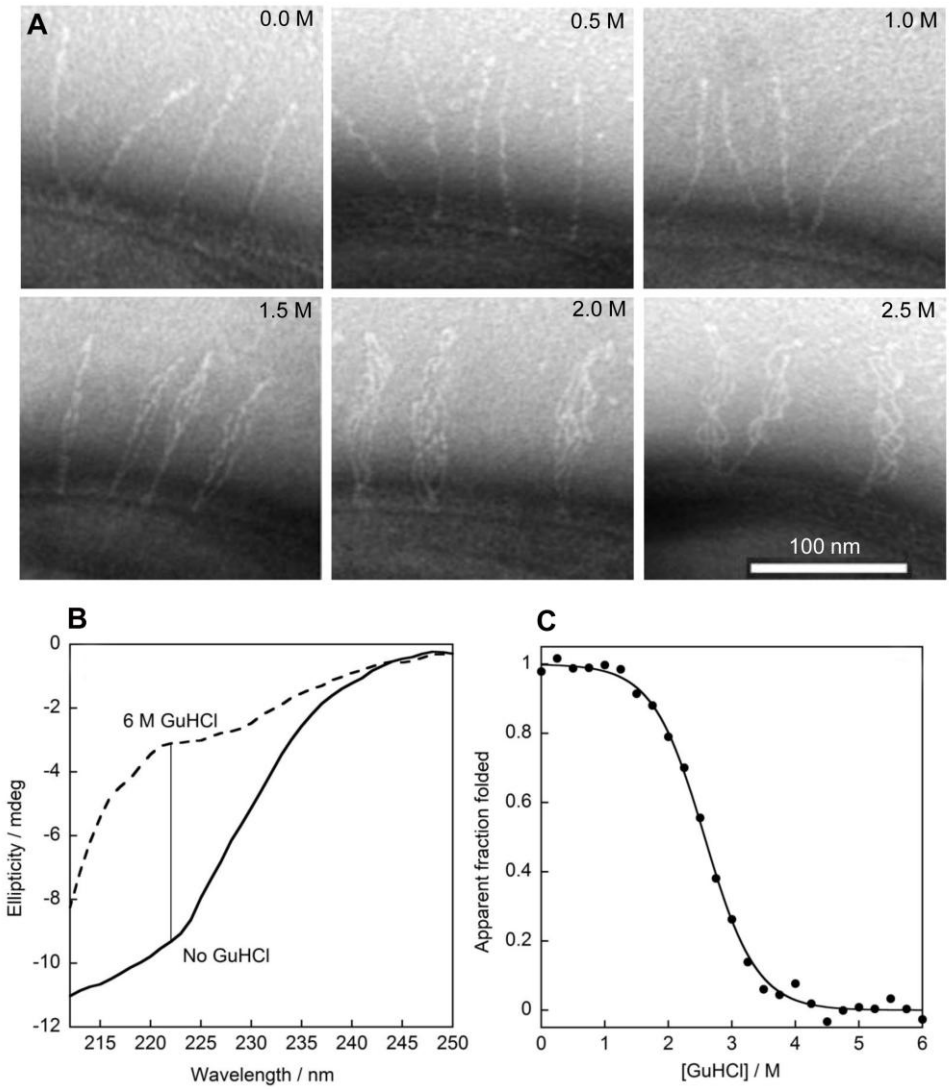

Fig. 2. Hsf is not a straight trimeric autotransporter but a double-folded molecule. (A) *H. influenzae* RM804 expressing Hsf was treated with 0.5 M, 1.0 M, 1.5 M, 2.0 M, and 2.5 M GuHCl as indicated. At high GuHCl concentrations (≥ 1.5 M) the Hsf molecule opened up and the quarternary structure was abolished. (B) Spectra of folded (no GuHCL) and unfolded (6 M GuHCl) recombinant Hsf 54-2300 were recorded using CD spectropolarimetry, with the vertical line placed at 222 nm. (C) The normalized ellipticity at 222 nm is plotted against the GuHCl concentration. The solid line shows the fit using equation 1 with the following parameters: $\Delta G^0(H_2O) = 16.6$ kJ·mol^{-1} and $m = 6.4$ kJ·mol^{-2}.

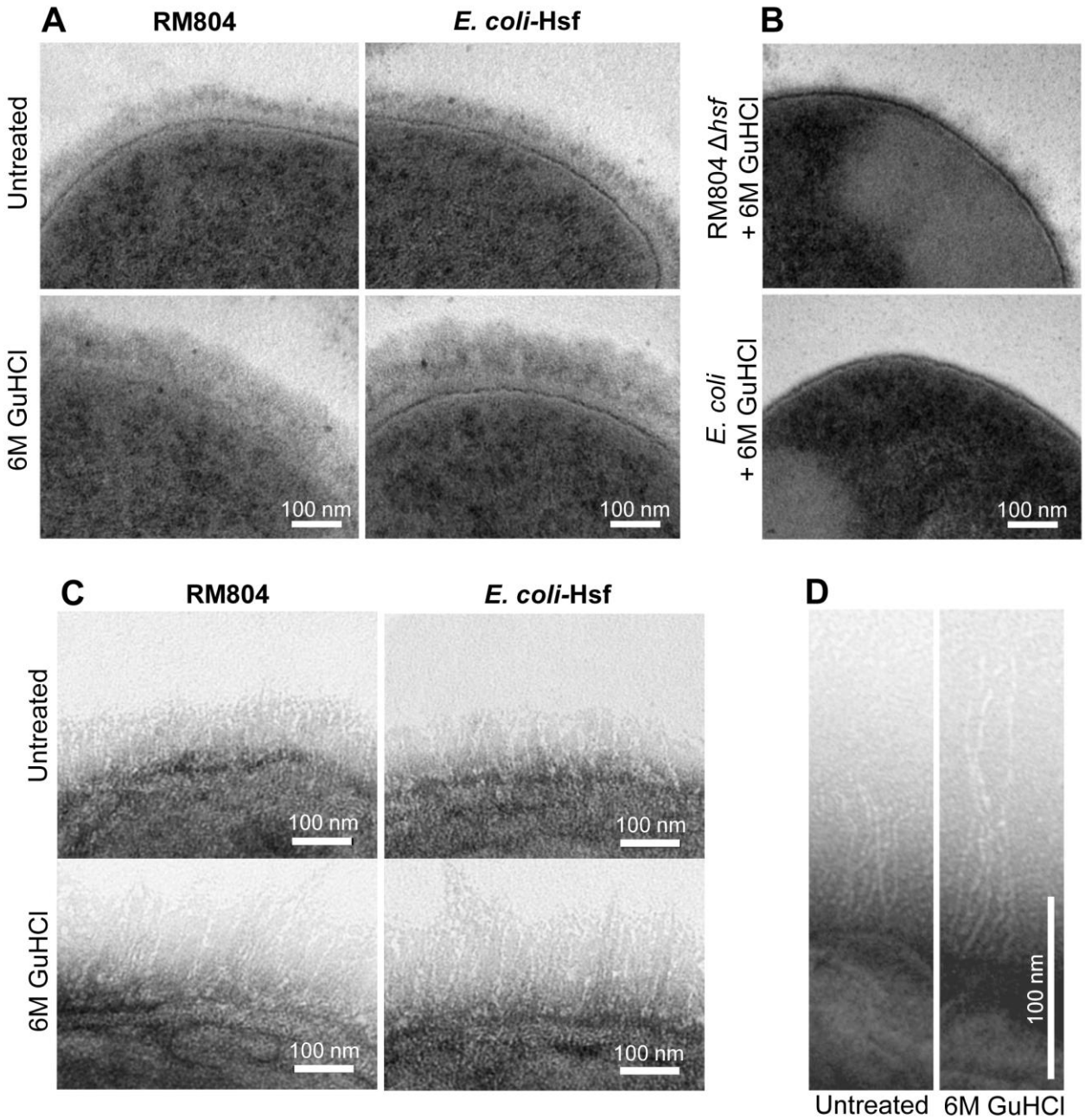

Fig. 3. Unwinding of the Hsf fibril at the bacterial surface by using GuHCl. (A) Hsf appears as a fuzzy layer after GuHCl treatment. *H. influenzae* RM804 WT and *E. coli* expressing Hsf were chosen as untreated, and in another set treated with 6 M GuHCl for 30 min, washed twice in PBS and fixed. Bacterial pellets were embedded in epon and sections were visualised in TEM. (B) Controls including *H. influenzae* RM804 Δ *hsf* and *E. coli* containing an empty vector treated with 6 M GuHCl for 30 min. (C) The sections shown in panel A were enlarged in order to observe Hsf at the bacterial surface. (D) *E. coli* expressing Hsf visualized at higher resolution.

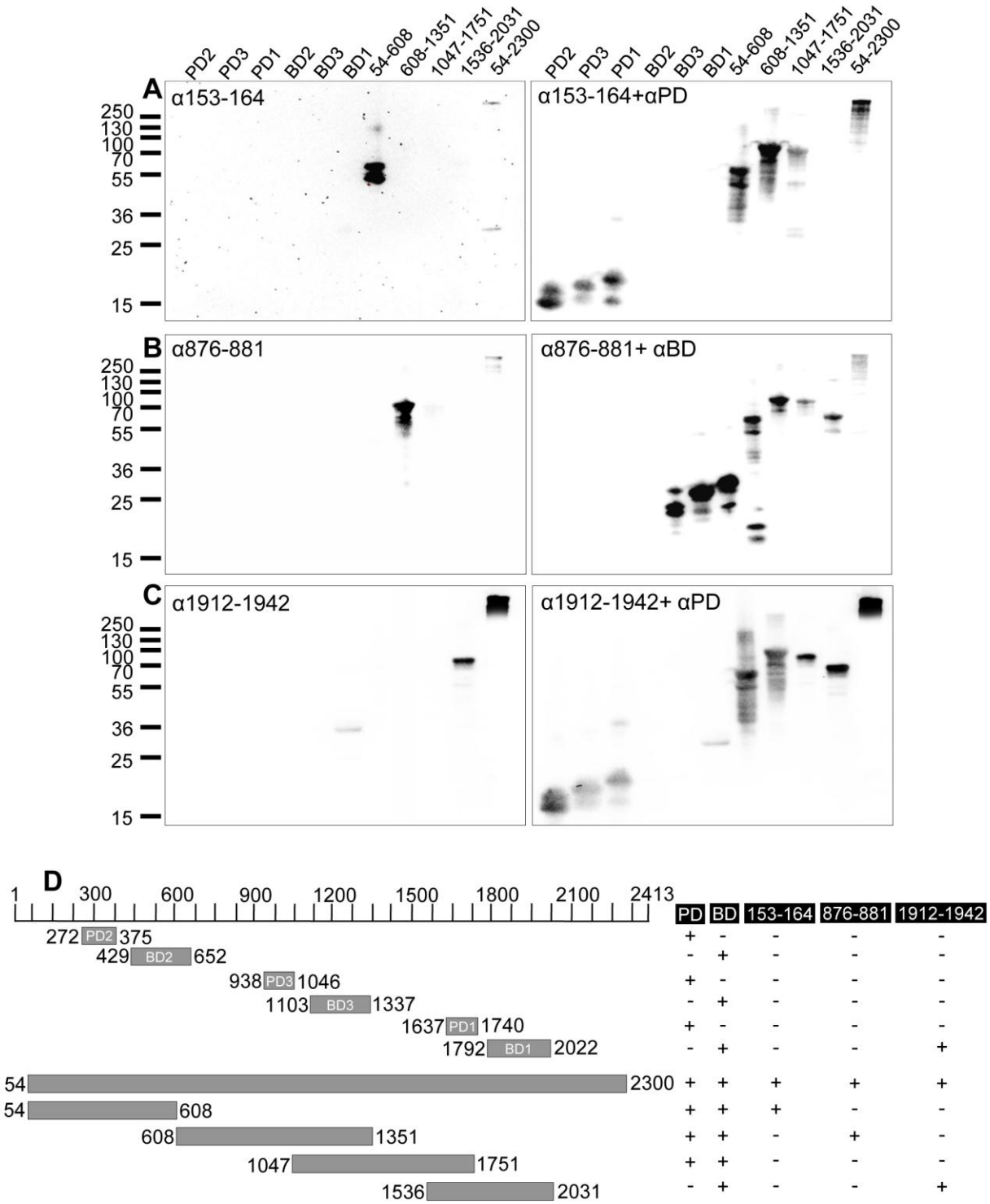

Fig. 4. Quality control of antibodies used in the present study. (A) Western blot that demonstrates the specificity of anti-Hsf$^{153-164}$ pAb (left panel), and cross reactivity of anti-PD antibodies (right panel). The anti-Hsf$^{153-164}$ pAb was added simultaneously with the anti-PD pAb as shown in the right panel. (B) Blots probed with anti-Hsf$^{876-881}$ (left), anti-Hsf$^{876-881}$ and anti-BD pAbs in right. (C) Blots incubated with anti-Hsf$^{876-881}$ (left), anti-Hsf$^{1912-1942}$ and anti-PD pAbs in right. (D) Summary of the results obtained from panels A-C.

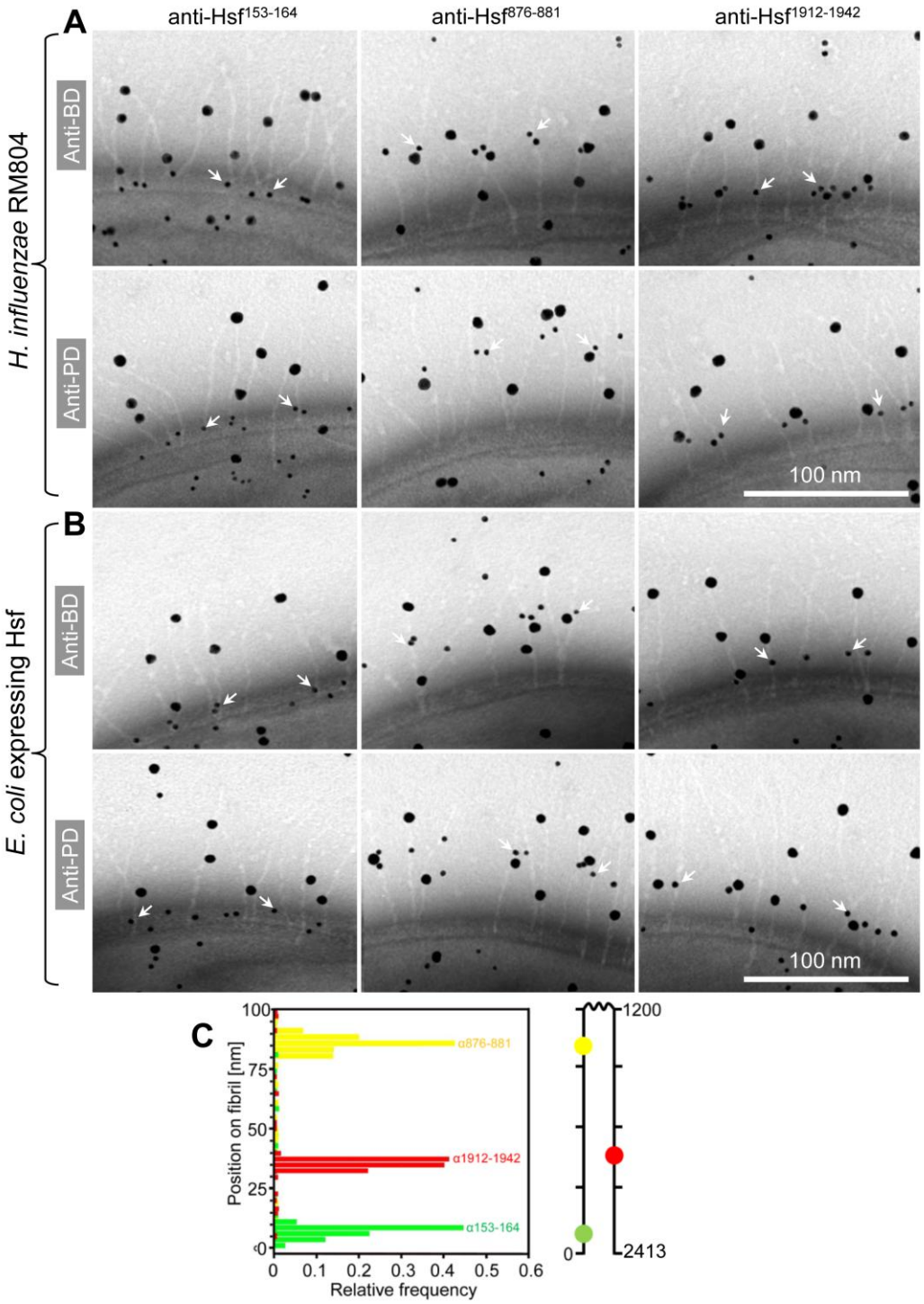

Fig. 5. Mapping of specific amino acid regions of the full length Hsf molecule by anti-BD and anti-PD Abs. (A) The anti-BD and anti-PD pAbs were labelled with gold particles having a size of 10 nm. The other specific pAbs shown at the top of the panels were labelled with 5 nm gold particles. Anti-BD or anti-PD pAb (10 nm) were used to probe *H. influenzae* RM804 with a combination of specific Abs (5 nm). (B) In parallel, a similar strategy as in (A) was used to probe *E. coli* expressing Hsf at its surface. The specificity of the anti-PD and anti-BD pAbs is shown in Figure S2. (C) The relative frequency of gold particles bound by specific pAbs directed against several fibrils (A-B) were plotted against the length of fibrils. Distribution of specific Abs is schematically presented in a cartoon to the right.

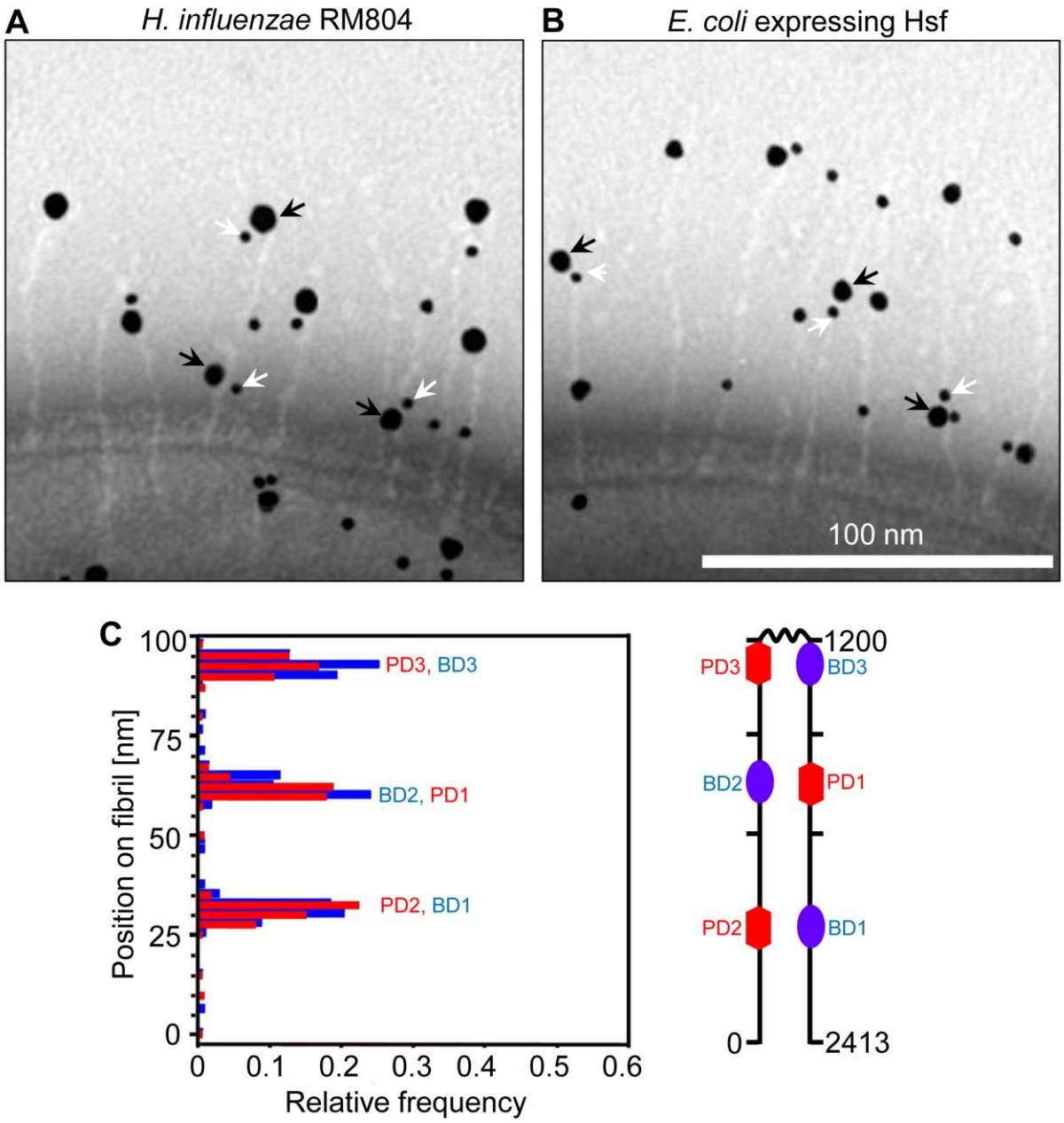

Fig. 6. Hsf PD and BD are parallelly distributed in the double-folded Hsf. (A) The anti-BD and anti-PD pAbs were labeled with 10 nm and 5 nM gold particles, respectively. *H. influenzae* RM804 was probed at a ratio of 1:1. (B) *E. coli* expressing Hsf was also probed with anti-BD (10 nm) and anti-PD (5 nm) pAbs. (C) The distance of the small and large gold particles in several fibrils was measured and their relative frequency was plotted against the length of the fibril. To the right, a cartoon delineates the positions of sequences detected by Abs.

Anti-PD (N-GLVTAKAVIDAVNKAGWR-C)

```
PD2(272-375)     KTSVIKEKDGKLFTGKENNDTN-KVTSNTATDNTDEGNGLVTAKAVIDAVNKAGWRVKTT
PD1(1637-1740)   KTSVIKDHNGKLFTGKELKDANNNGVTVTETDGKDEGNGLVTAKAVIDAVNKAGWRVKTT
PD3(938-1046)    KTSVIKDHNGKLFTGKDLKDAN-NGATVSEDDGKDTGTGLVTAKTVIDAVNKSGWRVTGE
                 ******:::*******: :*:* : .:  : *..* *.******:*******:****.

PD2(272-375)     TANGQN-------GDFATVASGTNVTFESGDGTTASVTKDTNGNGITVKYDA
PD1(1637-1740)   GANGQN-------DDFATVASGTNVTFADGNGTTAEVTKANDGS-ITVKYNV
PD3(938-1046)    GATAETGATAVNAGNAETVTSGTSVNFKNGNATTATVSKD-NGN-INVKYDV
                  *..:.       .:..**:***.*.* .*:.*** *:*  :*. *.***:.
```

Anti-BD (N-TKDGISAGNKAITNVAS-C)

```
BD2(429-652)     SWKAKAEADTDGALE-GISKDQEVKAGETVTFKAGKNLKVKQDGANFTYSLQDALTGLTS
BD3(1103-1337)   SWTAKADKYADGESE-GET-DQEVKAGDKVTFKAGKNLKVKQSEKDFTYSLQDTLTGLTS
BD1(1792-2022)   SWTATAGKEGTGEVDPANSAGQEVKAGDKVTFKAGDNLKIKQSGKDFTYSLKKELKDLTS
                 **.*.*     *  : . :  .*****:.******.***:**.  :*****:. *..***

BD2(429-652)     ITLGGTTNGGNDAKTVINKDGLTITPAG---NGGTTGTN--TISVTKDGIKAGNKAITNV
BD3(1103-1337)   ITLGGTANGRNDTGTVINKDGLTITLANGA-AAGTDASNGNTISVTKDGISAGNKEITNV
BD1(1792-2022)   VEFKDANGGTGSESTKITKDGLTITPANGAGAAGANTAN--TISVTKDGISAGNKAVTNV
                 : : .: .* ..  * *.******* *. .   * : :*  *********.**** :***

BD2(429-652)     ASGLRAYDDANFDVLNNSATDLNRHVEDAYKGLLNLNEKNANKQ-----PLVTDSTAATV
BD3(1103-1337)   KSALKTYKDT-QNTAGATQPAAN-TAEVAKQDLVDLTKPATGAAGNGADAKAPDTTAATV
BD1(1792-2022)   VSGLKKFGDG-HTLANGTVADFEKHYDNAYKDLTNLDEKGADNN-----PTVADNTAATV
                  *.*: : *      . : .  :   : * :.* :* :  :.       . ..*.*****
```

Fig S1 Alignment of Hsf PD and BD protein sequences and identification of conserved sequences for design of peptides (black background) that were used for immunization of rabbits.

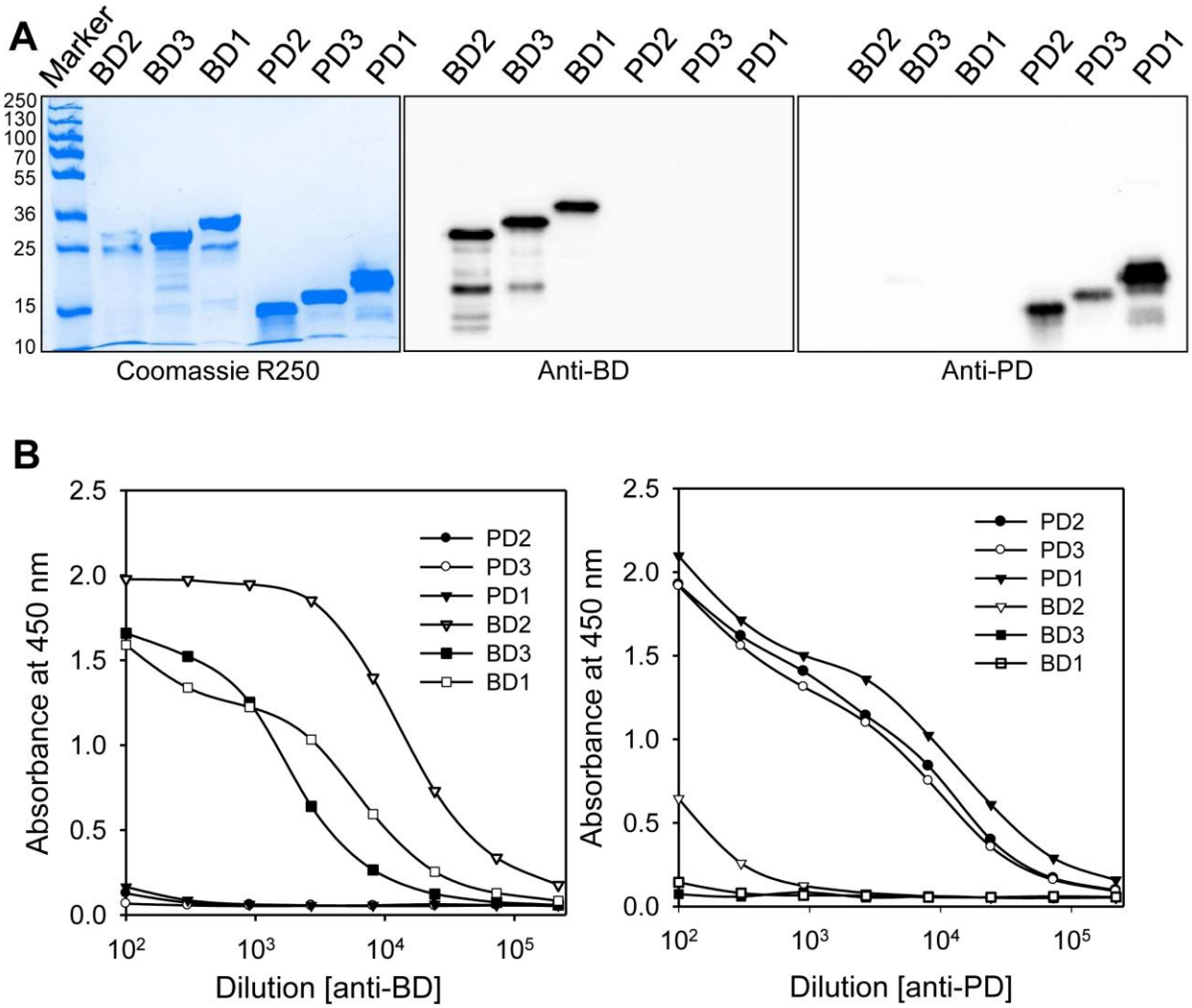

Fig S2 Specificity of anti-BD and anti-PD Abs. *A.* Left panel shows purified recombinant BD and PD proteins. Proteins (5 µg) were separated in a 12% SDS-PAGE and stained with Coomassie blue R250. Middle and right panels indicate blots probed with anti-BD and anti-PD Abs, respectively. Fifty ng of each protein was loaded on gels. *B.* ELISA that demonstrates the specificity of anti-BD and anti-PD Abs. All recombinant proteins (50 nM) were coated on ELISA plates and different dilutions of Abs were used.