UNDERSTANDING THE MOLECULAR BASIS FOR ADHESION BY HAEMOPHILUS SURFACE FIBRIL

By

Anchal Malik

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds School of Biomedical Sciences
Astbury Centre for Structural and Molecular Biology

December 2022

The candidate confirms that the work submitted is their own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The work in Chapter 3 and 4 of this thesis will result in publication:

(Article under preparation)

Acknowledgement

A Ph.D. degree may be awarded to an individual, but it is a result of cumulative efforts of a team. How can I extend my gratitude with a word-limit to each and every person who helped me in my thesis? Hence, you will find me struggling to convey my feelings with limited words in the following paragraphs.

I feel privileged to have received research training from experts with different biological backgrounds. First of all, I wholeheartedly thank my first supervisor Prof Adrian Goldman for his guidance, support and extreme patience throughout my Ph.D.. I also wish to thank my supervisor Prof Stephen Muench, Prof Lars Jeuken and my co-supervisor Prof Lorna Dougan for all the support they have provided for me to be able to accomplish the thesis successfully. I am grateful to all of my supervisors for believing in my capabilities to meet the demands of the interdisciplinary project and for keeping their patience as well as faith while I was learning new skills unknown to the group, cyclic voltammetry. Here, I would also like to take my opportunity to thank Eluceda for providing me the opportunity for developing this new skillset. I specially thank Dr Ian Eastwood, Dr Carla Solino and the other team members of Eluceda for their support and guidance. I also want to thank Prof Lars Jeuken for being one of my Supervisors at University of Leeds and helping me by introducing to the theory of potentiodynamics and addressing my concerns regarding data evaluation and validation. Again, I would like to acknowledge the consistent encouragement from all of my supervisors for overall development as an independent researcher that includes experimental designing, giving neat, coherent and concise presentations, team-spirit, and scientific writing. You gave

me enough freedom to discuss scientific questions with the experts in and out of the University and were always ready to collaborate to fulfil the project demands. You allowed me to make my own mistakes, encouraged me to come up with new ideas.

All of my supervisors helped me stay critical in designing experiments and data interpretation, teaching me to be resourceful. While working with them on the thesis project, they insisted on independent learning on the subject of structure biology and biophysics. I would like to acknowledge them for contributing to my growth as an independent researcher and to inculcate qualities such as looking beyond the problems, to keep patience and faith. I must mention that all of my supervisors provided me all the help I needed to accomplish this project, from discussing exciting articles in the field to providing me a scientific platform to discuss my thesis work. They have contributed significantly to my growth as researcher by providing exposure to national and international platforms.

I specially want to thank Mrs Pirjo Johnson, project officer, ViBrANT for her special support throughout my PhD and I'll always cherish our coffee chats. She always was a valuable support and first person to share any news that I had. I would also like to thank Goldman Lab members for establishing a healthy bond that not only helped me to conduct experiments in lab but also to shed off all the excess tension and stress during my Ph.D..

I consider my PhD journey a challenging one and I wouldn't have made it through without an exceptionally supportive and loving family. I want to thank them for their constant support, unconditional love and cooperation. My parents have been my greatest support and have always inspired me to gather my strength to accomplish my objectives. During this time, my brother, Uruvansh Malik, as always, was there to comfort me in all possible circumstances.

Most importantly I would like to thank my husband, Dr Vipul Panchal for being an all-weather support system. He has always been there with me during all phases and in every shade of my Ph.D. I am grateful to him for sticking around and for reminding me of life outside the lab.

Abstract

Invasive *Haemophilus influenzae* Infections have remained an epidemiological concern across the globe even after 3 decades of the *H. influenzae* serotype b (Hib) vaccine use (Slack, Cripps, Grimwood, Mackenzie, & Ulanova, 2021). With the emergence of severe invasive *H.influenzae* infections in various parts of the world, it remains crucial to develop novel means of diagnosis and intervention strategies. The Haemophilus surface fibril (Hsf) is a surface adhesin expressed in all encapsulated strains of *H. influenzae* and is therefore believed to be a critical virulence factor. As per the bioinformatic model, Hsf which is comprised of passenger domain and translocator domain is predicted to assume ~200 nm long homo-trimeric morphology. Although the initial biochemical study recently proposed that Hsf assumes a unique twisted hairpin-like architecture at the bacterial surface, this remains to be established in the absence of its molecular structure. This thesis tested this hypothesis by solving the high-resolution structure of the truncated $Hsf_{888\text{-}1337}$ segment containing the putative domain 3 (PD3) and the binding domain 3 (BD3). PD3 and BD3 are unique to Hsf and have been implicated for the bend in the proposed hairpin architecture. Here I show that the $Hsf_{888\text{-}1337}$ assumes only a slight bend which is not significant enough to support the hypothesis of a novel hairpin-like structure. Nevertheless, the work here renders the first-ever structural evidence at atomic resolution for the presence of bend architecture in the Trimeric Autotransporter Adhesins (TAA).

Altogether, the present finding doesn't only refute the earlier proposition of Hsf forming a hairpin structure but provides the first credible evidence for the presence of bending morphology in the TAA protein. In particular, the $Hsf_{888\text{-}1337}$ structure shows that the PD3 and

BD3 domains are linked with a highly flexible neck. Interestingly, the PD3 and BD3 domains are unique to Hsf but also show approximately 60% sequence similarity with the well-characterized binding domain 1 (BD1) of the *Haemophilus influenzae* adhesin (Hia) protein. The BD1 domain of Hia is responsible for its adhesion, hence implying the importance to study the PD3 and BD3 in Hsf.

As the physical interaction between Hsf and the host factor- Vitronectin (abundantly available in the blood) is already well established, for the second part of my thesis, I tested the adhesive property of individual domains of the Hsf with Vitronectin. This part of the study was aimed at demonstrating the utility of Hsf-vitronectin interaction in developing the point of care testing for the Invasive *Haemophilus influenzae* Infections. Using the Cyclic voltammetry, this study contributes towards the development of improved, cost-effective diagnostic tools against *H. influenzae*.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

°C	Degree Celsius
Å	Angstrom
ATP	Adenosine triphosphate
BAM	β-barrel assembly machinery
BD	Binding Domains
bp	Base pair
BSA	Bovine Serum Albumin
CC	Coiled Coil
CD	Circular Dichroisms
Cm	Centimetre
Cryo-EM	Cryo-Electron Microscopy
CV	Cyclic Voltammetry
dH_2O	Distilled Water
DNA	Deoxyribonucleic acid
dNTP	Deoxy nucleotide triphosphate
ECM	Extracellular matrix
g	Gram
h	Hour
Hia	*Haemophilus influenzae* adhesin
Hib	*Haemophilus influenzae* type b
His	Histidine
HRP	Horseradish Peroxidase
Hsf	Haemophilus Surface Fibril
IM	Inner membrane
IPTG	Isopropyl thio-β-D-Galactoside
IS	Insertion sequence
kb	Kilobase
kDa	Kilodalton
L	Liter
LB	Lysogenic Broth
M	Molar
MAC	Membrane attack complex
MFP	Membrane fusion protein
mg	Milligram
Mg2+	Magnesium Ion
min	Minute
ml	Milliliter

mm	Millimeter
mM	Millimolar
MS	Mass Spectrometry
mV	Millivolt
NaCl	Sodium Chloride
NaF	Sodium Fluoride
NaP	Sodium Phosphate
nm	Nanometre
OD	Optical Density or Absorbance at 600 nm wavelength
OM	Outer membrane
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PBST	Phosphate Buffered Saline/Tween20
PCR	Polymerase Chain Reaction
PD	Putative Domains
PDB	Protein data bank
PTM	Post-translational modifications
RE	Restriction endonuclease
Rnase A	Ribonuclease A
RNC	Ribosome nascent chain
rpm	Revolutions Per Minute
RT	Room temperature
RTX	Repeats in Toxins
SD	Standard deviation
SDS	Sodium Dodecyl Sulphate
sec	Second(s)
Sec	Secretory
SR	SRP Receptor
SRP	Signal recognition Particle
T1SS	Type I Secretion System
T2SS	Type II Secretion System
T3SS	Type III Secretion System
T4SS	Type IV Secretion System
T5SS	Type V Secretion System
TAA	Trimeric Autotransporter Adhesins
TAE	Trisbase, Acetic Acid and EDTA
TAT	Twin-arginine translocation
TM	Trans Membrane

TrpRing	Tryptophan ring
UV	Ultra violet
V	Volts
WB	Western Blot
YadA	*Yersinia* Adhesin A
µg	Microgram
µL	Microliter
µM	Micromolar
µmol	Micromole

Chapter 1 Introduction to Bacterial Secretion Systems and Trimeric Autotransporter Adhesins

1.1 Bacterial cell membranes

Bacteria are one of the oldest and the most prevalent form of the life on earth inhabiting a range of ecosystems. They are able to survive a variety of physical and chemical stresses exerted by their environment such as temperature and pH, which has led to intense diversity among bacteria worldwide. This includes bacteria that can resist extreme environments including, acidophiles (pH <3), alkaliphiles (pH >9), halophiles (salt >50 g/L), hyperthermophiles (temperature >80°C and piezophiles (pressure above 10 MPa)(Rampelotto, 2013).

Bacteria can be broadly categorized into Gram-positive and Gram-negative based on their morphology and cell wall structure. The classification is derived based on the Gram staining technique originally developed by Hans Christian Gram in 1882 to identify pneumonia causing microorganisms(BARTHOLOMEW & MITTWER, 1952). The study showed the difference in the colouring of bacteria; those appeared purple-brown due to retention of Gram staining were termed as 'Gram-positive', whereas those appear pink or red were termed as 'Gram-negative'(Baker & Bloom, 1948). Indeed, the property of the Gram-positive bacteria to retain gram stain is attributed to its thick peptidoglycan- a polymer of sugar and amino acids, layer. Regardless of composition the cell wall is crucial to address the myriad of physiological challenges associated with diverse ecosystems(Mueller & Levin, 2020).

Whilst providing a protective barrier around the bacteria, the cell wall must also play a crucial role in interacting and sensing the environment allowing transport both in and out of the cell. Therefore, efficient transport through various compartments of the cell wall is crucial to bacterial survival. Protein secretion in bacteria can be a two-step or one-step mechanism(Rego, Chandran, & Waksman, 2010). In two-step secretion, the proteins are

secreted first to the periplasm through Sec- or Tat- secretion systems followed by a second transport system(Natale, Brüser, & Driessen, 2008). The proteins in a one-step secretion system are secreted through channels that span both the inner and the outer bacterial membrane and do not depend on Sec- or Tat- pathways. Bacteria experience major challenges in homeostasis for which they possess a large variety of protein secretion systems which are described in detail below.

1.2 Secretion systems in bacteria

Protein secretion is a process of transporting proteins from the cytoplasm into other compartments of the cell, the environment, or other prokaryotic and eukaryotic cells. This is an essential function of the cells and is carried out through specific secretion systems. The bacterial secretion systems are composed of protein complexes present on the cell membrane. The secretion systems can be classified into different types based on their specific structure, composition and activity in both Gram-negative and Gram-positive bacteria. While Gram-negative bacteria have at least eight types of secretion systems, Gram-positive bacteria have only four(Costa et al., 2015; Green & Mecsas, 2016; Schneewind & Missiakas, 2012). The secretion systems of Gram-negative bacteria are important for a number of roles including exporting the proteins responsible for causing virulence outside the cell and sometimes directly into the cytoplasm of a target cell which the bacterium is invading(Depluverez, Devos, & Devreese, 2016). Protein secretion in Gram-negative bacteria can be achieved through either a two-step or one-step mechanism. In two-step secretion, the proteins are secreted first to the periplasm through Sec- or Tat- secretion systems followed by a second transport system which takes protein from the periplasm and over the outer bacterial membrane. The

proteins in one-step secretion are secreted through channels that span both the inner and the outer bacterial membrane and do not depend on Sec- or Tat- pathways. Both the one and two-step secretion systems in Gram-negative bacteria are numbered Type I to Type VI, with each system transporting a distinct subset of proteins. However, although this classification covers most transporters, others are yet to be identified and is therefore incomplete. In the following sections, I will focus on the Gram-negative secretion systems as these were the focus of the PhD study.

1.2.1 The Sec Secretion Pathway in one-step transport

The Sec pathway is one of the best-studied protein translocation pathways within bacteria that translocate proteins in their unfolded state to extra cytoplasmic destination. This requires the Sec complex to i) distinguish Sec substrates from other proteins and ii) to discriminate the secretory protein targets from integral membrane proteins without compromising the membrane integrity. Such a high degree of sophistication is achieved through the concerted action of highly conserved Sec pore assembly and non-conserved accessory proteins. The highly conserved pore assembly- the SecYEG translocase, is comprised of a protein targeting component, a motor protein, and a membrane-integrated conducting channel(Thomas, Holland, & Schmitt, 2014), whereas the accessory components of the Sec pathway involves Sec A, D, E and F(Natale et al., 2008; Schatz & Beckwith, 1990). The SecYEG translocon is heterotrimeric protein complex that remains embedded in the inner membrane of Gram-negative bacteria as shown in (Figure 1). The SecY protein is 48 k Da with 10 Trans Membrane (TM) α-helices, with 1-5 and 6-10 comprising the two halves of the protein as revealed by the first X-ray structure of an archaeal homologue(Berg et al., 2004). The 4-6 cytosolic loops of secY are essential for protein secretion as they serve as the docking

site for the ribosomes and the targeting factors(Kuhn et al., 2011),(Chiba, Mori, & Ito, 2002),(Cheng, Jiang, Mandon, & Gilmore, 2005). The two halves of SecY are stabilized and held together by SecE. Mutation studies have shown that SecY mutants unable to interact with SecE are rapidly degraded by membrane protease Fts H(Kihara, Akiyama, & Ito, 1995), showing the stabilising role SecE makes. SecE consists of a single TM domain that is connected to a surface exposed α-helix by a flexible hinge and these domains serve as the major site of interaction between SecE and SecY (Lycklama a Nijeholt, de Keyzer, Prabudiansyah, & Driessen, 2013). SecG is a 12 k Da subunit consisting of two TM helices in *E. coli* placed at the N-terminus of SecY (Satoh, Matsumoto, Mori, & Ito, 2003; Satoh, Mori, & Ito, 2003; van der Sluis, Nouwen, & Driessen, 2002). *In vitro* SecG is shown to stimulate protein translocation at low temperatures in which the movement of protein is compromised(Nishiyama, Hanada, & Tokuda, 1994). It is important to note that SecA and SecE are essential components, whereas SecG is dispensable in bacteria in some instances(Pohlschröder, Hartmann, Hand, Dilks, & Haddad, 2005). Studies have shown that SecG is involved in the transport of SecA dependent secretory proteins, whereas the proteins that are independent of SecA secretion remain unaffected by the presence of SecG (Koch & Müller, 2000; Nishiyama et al., 1994).

Protein translocation by the Sec pathway is initiated by the recognition of a 20 amino acid-long hydrophobic signal sequence at the N-terminus of the secreted protein, that is not a part of the mature protein(Papanikou, Karamanou, & Economou, 2007). This signal sequence contains 3 regions: a positively charged amino terminus, a hydrophobic core, and a polar carboxyl-terminal which cleaves off in the periplasm(Papanikou et al., 2007). Depending on the final destination of the proteins they either contain a SecB- specific signal sequence or the Signal Recognition Particle (SRP) specific signal sequence (described in section 1.2.2). For

proteins with the SecB specific signal sequence, SecB serves as a chaperone by binding to the pre-secretory protein and preventing them from folding. SecB then transfers the protein to the SecA, which not only guides the proteins towards the SecYEG channel, but also serves as an ATPase to provide the energy required for the translocation by Adenosine triphosphate (ATP) hydrolysis. As the final destination of the proteins to be translocated by the Sec system is preriplasmic(Randall & Hardy, 2002), it is crucial to specifically recognize and target them to Sec pathway.

1.2.2 The SRP pathway

As described above proteins can be targeted by the SEC-B or Signal Recognition Particle (SRP) recognition sequences. For the SRP pathway a cytoplasmic protein–RNA complex called Signal Recognition Particle (SRP) is required which is universally conserved(Keenan, Freymann, Stroud, & Walter, 2001; Lütcke, 1995). In Eubacteria, SRP consists of Ffh protein (SRP *F*ifty-*f*our *h*omolog) and 4.5S RNA(Bernstein et al., 1989; Keenan et al., 2001). The SRP RNA interacts with the Ffh with its most conserved domain IV (helix 8 in eukaryotes). Ffh is divided into three domains: the α-helical N-terminal domain, the nucleotide-binding G domain (GTPase) and the methionine-rich, α-helical M domain. The M domain is responsible for RNA interaction and signal peptide binding. In *E. coli*, FtsY functions as the SRP receptor (SR)(Luirink et al., 1994). FtsY consists of three domains, an A domain which is an acidic domain at the N terminus and the other two domains that are homologous to the N and G domains of Ffh, also described as N and G domains of FtsY. The association of FtsY with the plasma membrane and phospholipids contributes to the activation of the GTPase(E De Leeuw et al., 2000; Erik de Leeuw et al., 1997; Luirink et al., 1994).

In order to ensure the efficient insertion of hydrophobic regions of the extra-cytoplasmic proteins in the inner membrane (IM) SRP mediated targeting occurs in a co-translational manner. Briefly, the hydrophobic signal sequence of the nascent protein binds to the SRP as it emerges from the ribosome. This leads to the formation of the SRP/RNC (ribosome nascent chain) complex. The interaction of the SRP/RNC complex with the membrane-bound SRP receptor (SR) initiates the delivery of the complex to the translocation channel- SecYEG channel which in turn translocate nascent peptide to the membrane and thereby, couples the protein translation with the secretion by SecYEG channel(Luirink & Sinning, 2004). Finally, the translocation of the SRP/RNC complex to the translocon leads to the dissociation of the SRP/SR complex in a GTP dependent manner.

1.2.3 The Tat Secretion Pathway

As the Sec pathway is only responsible for the secretion of the unfolded proteins, it is something to keep in mind that not all proteins can be secreted in their unfolded state(Robinson & Bolhuis, 2004). The proteins that contain post-translational modifications (PTM) must be synthesized in the cytoplasm in their fully folded state(Berks, Palmer, & Sargent, 2005). This is due to the unavailability of the PTM enzymes in the periplasm or on the surface of the cell. The Tat pathway plays the primary role in the successful secretion of the folded proteins(Robinson & Bolhuis, 2004).

In Gram-negative bacteria, the Tat pathway consists of TatA, TatB and TatC subunits which interact with the signal sequence of the folded proteins(Sargent, Stanley, Berks, & Palmer, 1999). The signal sequence contains a pair of Arginine in the motif S-R-R at the N-terminus of the folded protein(Müller, 2005). In *E. coli*, Tat B and Tat C first bind to the N-terminal signal

sequence of the target proteins(Müller, 2005). This is then followed by the recruitment of TatA to form a membrane-spanning channel that can transport the protein to the periplasm. The secreted proteins can either remain in the periplasm or can be transported out of the cell by the Type II Secretion System to be reviewed further in this chapter.

1.2.4 The Type I Secretion System

The type I secretion system (T1SS) is a one-step system which is found among a large number of Gram-negative bacteria(Thomas et al., 2014). T1SS spans through both the inner- and the outer- membranes and transports proteins in one-step as shown in Figure 2 (Thomas et al., 2014). T1SS is an assembly of three components- an Inner-membrane ABC transporter, a periplasmic membrane fusion protein (MFP) and a porin in the outer-membrane, as shown in Figure 2 (Morgan, Acheson, & Zimmer, 2017; Symmons, Bokma, Koronakis, Hughes, & Koronakis, 2009). The ABC transporter utilise the energy from ATP hydrolysis for substrate translocation. Along with this, the ABC transporter also interacts with the MFP, and participates in substrate recognition(Kanonenberg, Schwarz, & Schmitt, 2013). The association of MFP with ABC transporter in the inner-membrane allows the complex to associate with the porin in the outer-membrane(Kanonenberg et al., 2013; Lee et al., 2012; Pimenta, Young, Holland, & Blight, 1999). This results in pore formation in the OM to export molecules outside of the cell(Balakrishnan, Hughes, & Koronakis, 2001).

Some Gram-negative bacteria have several T1SSs that resemble ATP-binding cassette (ABC) transporters, dedicated to export one or a few small protein substrates such as toxins and other unfolded substrates out of the cell(Morgan et al., 2017; Symmons et al., 2009). These substrates are proteins that perform distinct functions, such as, digestive enzymes, for example, proteases and lipases, as well as adhesins, heme-binding proteins, and proteins with repeats-in-toxins (RTX) motifs.

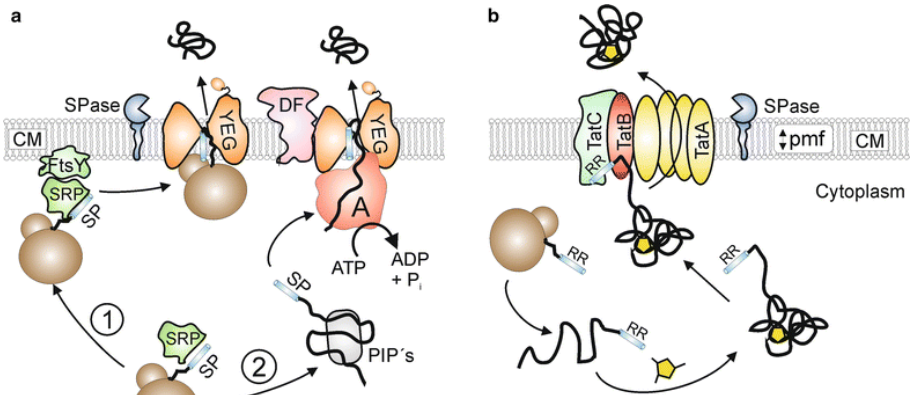


Figure 1 Two major bacterial export pathways: Sec and Tat.

a The general secretion (Sec) protein export pathway. In the cotranslational mode (1), Sec substrates possessing highly hydrophobic signal peptides (SP) are recognized at the ribosome by the signal recognition particle (SRP). Subsequently, the ribosome-nascent chain (RNC)-SRP complex docks to the SRP-receptor FtsY and the RNC is then further transferred to the SecYEG translocation pore such that ribosomal exit site is in close proximity to SecYEG. The energy for translocation in the cotranslational export mode is provided by further elongation of the substrate at the ribosome. In the posttranslational mode (2), Sec-dependent precursor proteins are kept in an export-competent state by posttranslationally interacting proteins (PIP's) such as SecB, the general chaperones GroELS/DnaK–DnaJ–GrpE or the soluble form of SecA. The signal peptide (SP) is recognized by the SecA protein which pushes the protein through the SecYEG protein conducting channel in a stepwise and ATP-dependent manner. In addition, SecDF exerts a proton motive force (pmf)-dependent pulling force on the substrate from the trans-side of the cytoplasmic membrane (CM). During or shortly after translocation, the signal peptide is removed by signal peptidase (SPase) and the mature protein is released on the trans-side of the CM. **b** The twin-arginine translocation (Tat) protein export pathway. After folding and, if required (as shown here), cofactor insertion, preproteins containing a signal peptide with a twin-arginine motif (RR) are recognized by a receptor complex consisting of TatC and TatB. Subsequently, homooligomeric complexes of TatA are recruited to the substrate-loaded receptor complex in a proton motive force (pmf)-dependent manner, followed by the translocation of the substrate across the cytoplasmic membrane (CM). Following to substrate translocation, the signal peptide is cleaved by signal peptidase (SPase) and the mature protein is released on the trans-side of the membrane.

Freudl, R. Signal peptides for recombinant protein secretion in bacterial expression systems. *Microb. Cell Fact.* **17**, 52 (2018).

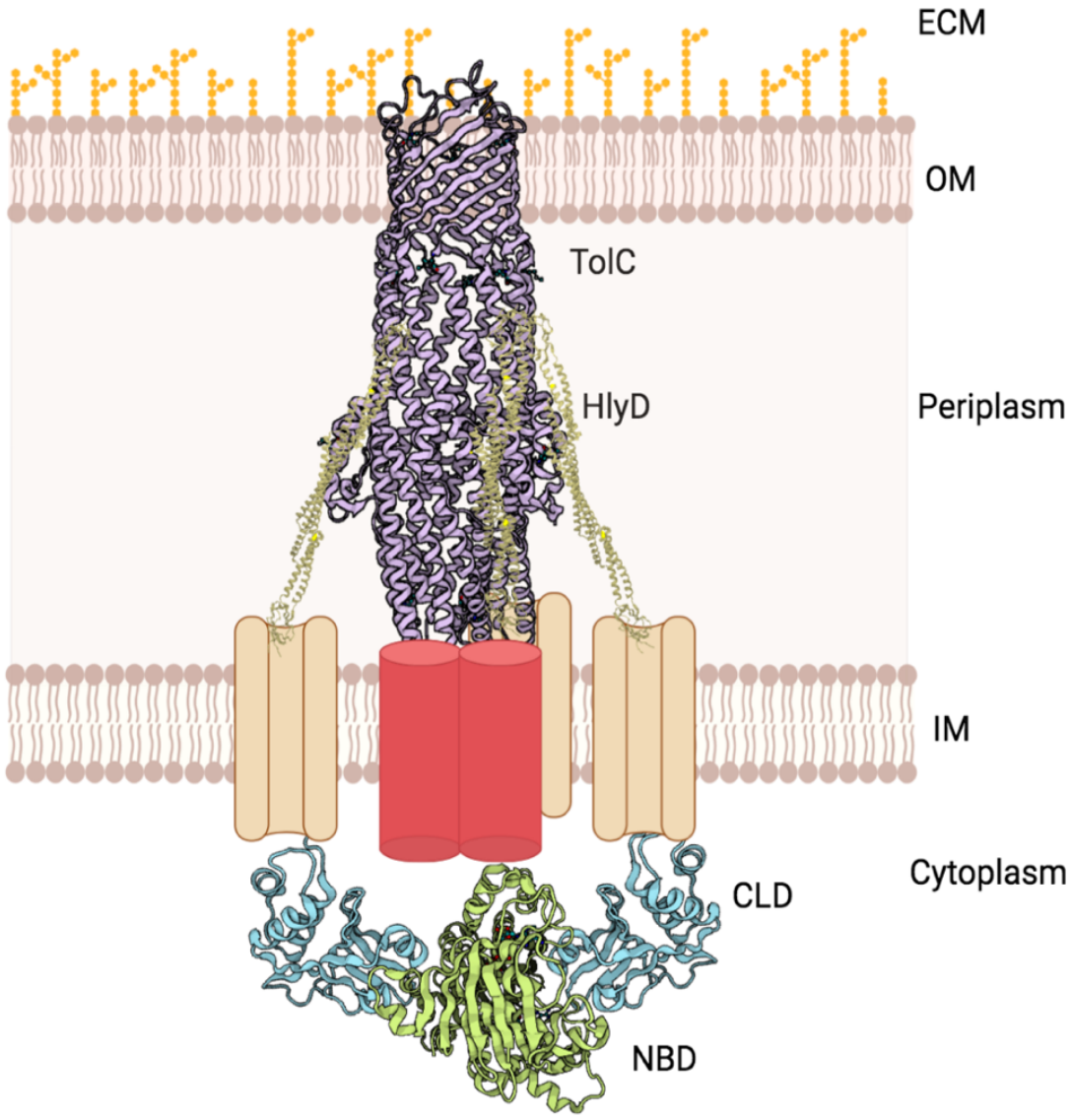


Figure 2 Type I Secretion system of Gram-negative bacteria

Schematic representation of the assembled HlyA T1SS. The trimeric crystal structure of TolC (pdb code 1ek9) in the outer membrane is shown in purple, green and cyan Illustrrations. The crystal structure of the known part of HlyD (pdb code 5c21) is shown as red cyllenders, while symbols represents the unknowns parts, the transmembrane helix and the 60 amino acid large cytoplasmic part. The structures of the CLD (pdb code 3zua) and the NBD (pdb code 1xef) are summarized as green and yellow cartoons, respectively. The TMDs for which a structure is not yet known are highlighted as cylinders.The substrate HlyA contains a secretion signal that has been located to the last 50-60 C-terminal amino acids. A C-terminal secretion signal is generally the case in all substrates of T1SS. Additionally, HlyA belongs to the family of RTX-proteins (“repeats in toxin”), whose characteristic is the presence of a variable number of non peptide repeats (RTX-domains, consensus sequence of the repeat is GGXGXDXUX (X: any amino acid residue, U: large, hydrophobic residue)) that bind calcium ions and trigger folding of the protein in the extracellular space. As the cytosolic concentration of Ca^{2+} is too low to allow binding of Ca^{2+} to the nonapeptide repeats, substrates of T1SS remained unfolded in the cytosol, despite their impressive size (1024 amino acids in the case of HlyA).

1.2.5 The Type II Secretion System

Type II secretion systems (T2SSs) are responsible for transporting proteins into the extracellular environments and are found only in the OM of Gram-negative bacteria. T2SSs work in coordination with the Sec and Tat pathways for delivery of proteins across the IM, described earlier in this chapter (Figure 3). Because of this reason the T2SSs were originally called as the main terminal branch of Sec pathway(Korotkov, Sandkvist, & Hol, 2012). However, as these secretion systems can also transport the Tat secreted proteins their nomenclature was changed to T2SS(Voulhoux et al., 2001). The type II secretion systems are conserved in most Gram-negative bacteria and are capable of secreting a variety of substrates outside the cell. The proteins secreted by the T2SSs have a cleavable Sec/Tat signal sequence and must be folded in periplasm. In some bacteria, T2SSs are responsible for transporting one substrate whereas in others they have been shown to transport multiple substrates(Cianciotto, 2005). The proteins secreted by T2SSs perform various biological functions, some of which also contribute to bacterial virulence. Some of the example include cholera toxin of *V. cholerae,* which causes the watery diarrhoea associated with the disease cholera and exotoxin A of *P. aeruginosa*, which blocks protein synthesis in host cells, leading to lethal infection by this bacterium(Korotkov et al., 2012). Most of the proteins secreted by the T2SSs are enzymes such as proteases, lipases and phosphatases(Korotkov et al., 2012).

T2SSs are a complex of proteins divided into four subassemblies: OM complex, IM platform, secretion ATPase and pseudopilus(Korotkov et al., 2012). Together there can be as many as 15 different proteins in the T2SSs complex(Korotkov et al., 2012). The outer-membrane T2SSs complex is composed of a multimeric protein called secretin which serves as a channel for transporting the folded substrate out of the cell (Figure 3)(Korotkov, Gonen, & Hol, 2011).

Secretin connects to the proteins of the inner-membrane platform by a long N-terminus domain extended through the periplasm (Figure 3)(Korotkov et al., 2011). The Inner-membrane platform plays a crucial role in the secretion process by communicating with the secretin, pseudopilus, and the ATPase to coordinate export of substrates(Korotkov et al., 2012). As its name suggests, the ATPase provides the energy to power the secretion process by ATP hydrolysis in the cytoplasm. Pseudopilus is evolutionarily related and structurally similar to proteins that comprise type IV pili on bacterial cell surfaces(Hobbs & Mattick, 1993). The involvement of pseudopilus is based on the "piston" model of the T2SS. In this model, the pseudophilus retracts to push the folded proteins out of the OM when they come in contact with the periplasmic domain of the secretin(Hobbs & Mattick, 1993; Korotkov et al., 2012; Shevchik, Robert-Baudouy, & Condemine, 1997).

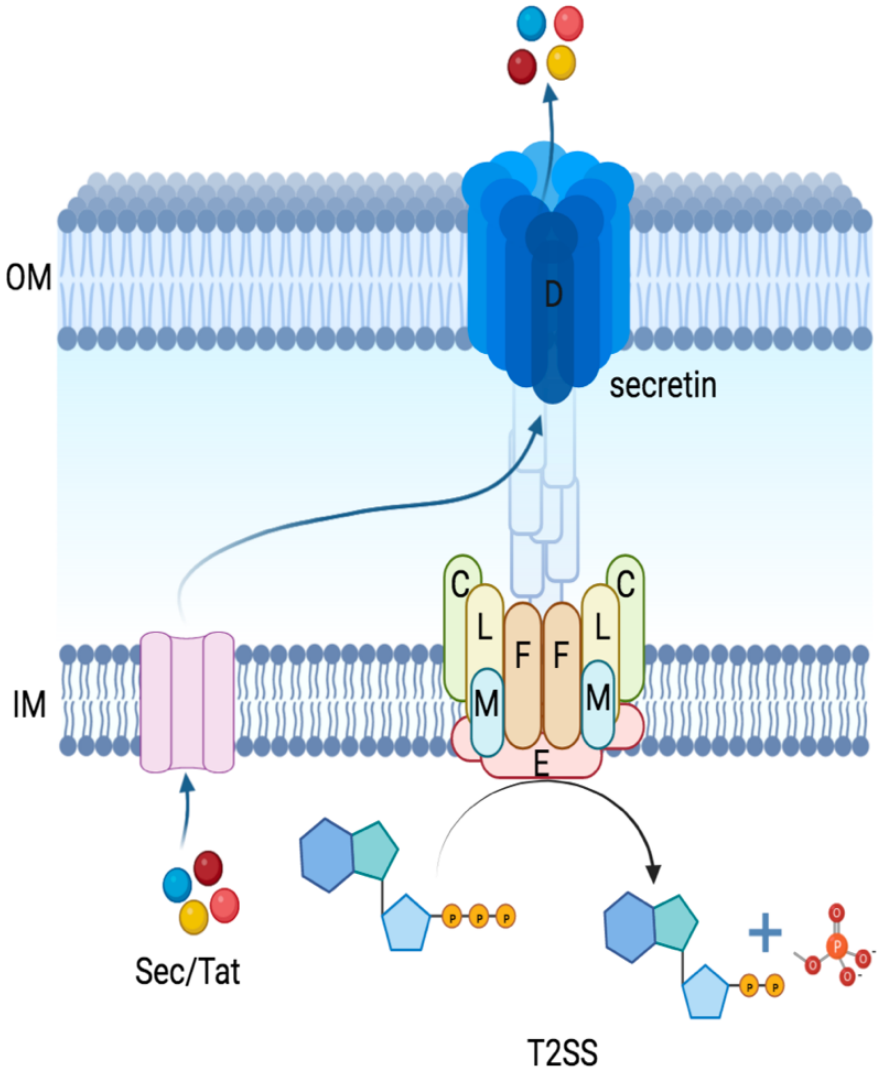


Figure 3 Type II secretion System of Gram-negative bacteria

Model of the Gsp secreton assembly the exported exoproteins depicted as balls in different colours have initially been exported across the IM via either the Sec or Tat machinery . The exoproteins are subsequently recognised by the Gsp machinery and transported across the OM via the secretin, GspD. The secretin GspD is shown as a homomultimeric ring forming a channel with a large central opening. GspM is proposed to direct the location of GspL at a specific site into the cell envelope (cell pole in case of *V. cholerae*). GspL is anchoring the GspE NTPase (traffic ATPase) to the inner face of the cytoplasmic membrane. The GspE NTPase is represented as a homomultimer (hexameric ring). The opening and closure of the GspE central ring is linked to cycles of ATP hydrolysis and promotes the assembly of the pseudopilins. Pseudopilins are thus using the central cavity of the GspE to be incorporated into a pseudopilus structure that pushes exoproteins through the secretin channel. The GspC component is interacting with the periplasmic N-terminal domain of the GspD secretin. The pseudopilus is arbitrarily represented as a succession of different pseudopilins.

1.2.6 The Type III Secretion System

Type III secretion systems (T3SSs) are often essential to the pathogenicity of bacterial pathogens. They are responsible for delivering a wide variety of proteinaceous substrates into eukaryotic cells commonly called effector proteins, as well as to the bacterial flagellum- a key motility orgenelle. As a result, T3SS is believed to play a crucial role in manipulating the host cellular pathways and therefore establishing the infection. Considering the scope of the thesis, the description of the T3SS here is limited to the former scenario termed translocation-associated T3SS.

The translocation-associated T3SSs have “needle and syringe” like overall morphology and are also known as injectosomes (Figure 4)(Buttner, 2012). Typically, they are comprised of an extracellular translocon assembly, a rod-like structure called a needle and a conserved membrane-spanning base complex (Abrusci, McDowell, Lea, & Johnson, 2014; Burkinshaw & Strynadka, 2014; Sauvonnet, Vignon, Pugsley, & Gounon, 2000). The base complex in this T3S system is comprised of ring structures and a centre rod that connects to a translocon assembly in a host plasma membrane via an extracellular needle (Figure 4)(Abrusci et al., 2014; Burkinshaw & Strynadka, 2014; Deane et al., 2006; Demers et al., 2014; Dohlich, Zumsteg, Goosmann, & Kolbe, 2014; Radics, Königsmaier, & Marlovits, 2014). Various structural and mutation studies suggest that structural rearrangement of the needle tip proteins could be contributing to the transmission of the signals such as host cell sensing and subsequent insertion of the translocon and effector proteins in the host (Picking et al., 2005; Price, Cowan, Perry, & Straley, 1991).

Although, a T3SS-mediated protein translocation was initially assumed to be a one-step process where a large number of effector proteins use chaperons to be guided to the T3SSs

in an ATP-dependent manner, recent studies on the effector proteins from *Yersinia* spp. has led to an alternative two-step model as per which the effector proteins are localized to the bacterial OM prior to translocation (Akopyan et al., 2011; Amer et al., 2016; Blocker, Komoriya, & Aizawa, 2003b; Edgren, Forsberg, Rosqvist, & Wolf-Watz, 2012; Hofreuter, Odenbreit, Püls, Schwan, & Haas, 2000; Solomon, Zhang, McCrann, Bliska, & Viboud, 2015; Tejeda-Dominguez, Huerta-Cantillo, Chavez-Dueñas, & Navarro-Garcia, 2017).

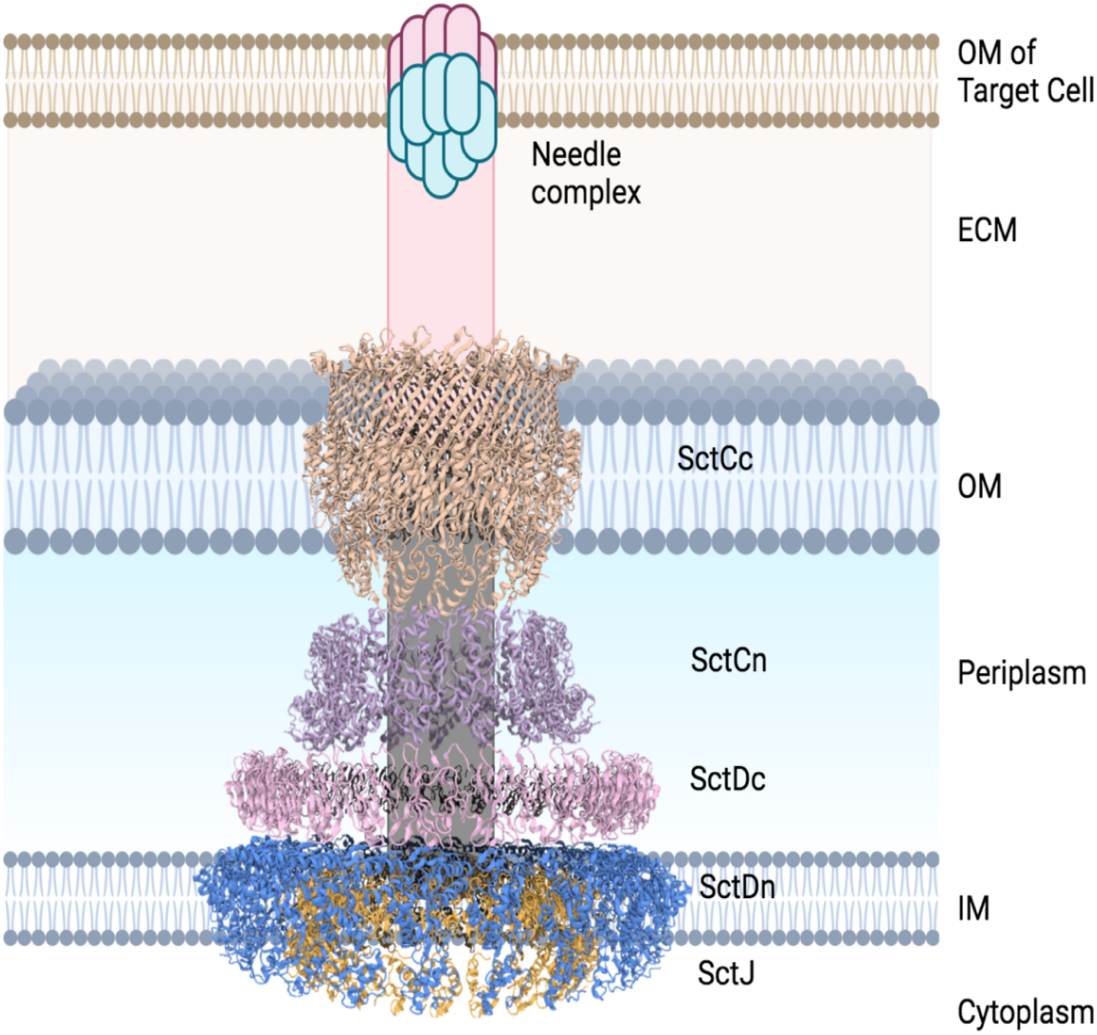


Figure 4 Type III secretion system

The basal body of the T3SSs complex which comprises the inner-membrane and outer-membrane rings, which are assembled using ring-building motifs that are shared between components. The overall architecture of the basal body, the needle and the inner rod in the *Salmonella enterica* subsp. *Enterica* serovar Typhimurium *Salmonella* type III secretion system (T3SS), as determined using cryo-electron microscopy (cryo-EM), is shown. The figure shows a composite of solved structures of individual components of T3SSs from different bacterial species and represents our most current structural understanding of non-flagellar T3SSs. There is no structural information for the T3SS components SctR, SctS, SctT and SctK, and therefore they are not depicted here. The basal body inner membrane and outer membrane ring structures from the *Salmonella* T3SS are shown. The amino-terminal and carboxy-terminal domains in the outer membrane secretin SctC and the inner membrane ring protein SctD form distinct ring structures and are labelled as subscripts ($SctC_N$, $SctC_C$, $SctD_N$ and $SctD_C$).

1.2.7 The Type IV Secretion System

The Type IV secretion systems (T4SSs) are unique among bacterial translocation systems in terms of their functional diversity, which exports both DNA and protein substrates in a cell-contact -dependent and -independent manner. Further, T4SS are not host-specific and target a wide range of species including both prokaryotic and eukaryotic cells(Li, Hu, & Christie, 2019),(Waksman, 2019), including bacteria from the same species(Waksman, 2019).

The double membrane spanning T4SSs of Gram-negative bacteria are categorised into three sub families depending on their biological functions; i) the conjugation systems that delivers DNA to diverse target cells (within phylogenetically diverse bacterial species, fungi, plant and human cells) in a cell-contact dependent manner, ii) the DNA uptake and release systems that function in a cell-contact independent manner and iii) the effector translocator systems that translocates effector molecules through direct cell contact to eukaryotic target cells and is essential for the infection process of various prominent bacterial pathogens. Therefore, the effector translocator T4SS system is reminiscent of the T3SS(Bacon et al., 2000; Blocker, Komoriya, & Aizawa, 2003a; P Bundock, den Dulk-Ras, Beijersbergen, & Hooykaas, 1995; Grohmann, Muth, & Espinosa, 2003; Hofreuter, Odenbreit, & Haas, 2001; Hofreuter et al., 2000).

The most comprehensive understanding of a typical T4SS assembly and architecture is derived from the well-studied VirB/D4 transfer system of *Agrobacterium tumefaciens* (Figure 5)(Paul Bundock, den Dulk-Ras, Beijersbergen, & Hooykaas, 1995; Lessl, Balzer, Pansegrau, & Lanka, 1992; Lessl & Lanka, 1994). This conjugation system is comprised of three unique substructures: a hexameric assembly of the coupling protein- VirD4, a transenvelope complex; and the conjugative pilus. The latter two assemblies are comprised of mating pore

formation (Mpf) proteins- VirB1-VirB11(Fronzes, Christie, & Waksman, 2009). Various genetic and biochemical studies have established the explicit function of the coupling protein assembly in recruiting the conjugal DNA substrates to the transmembrane Mpf structure(Llosa, Gomis-Rüth, Coll, & de la Cruz Fd, 2002). Vir proteins comprising the Mpf structure have different functions and are localised to the different subcellular fractions; 1) a trans envelope channel comprising components- VirB6, VirB8 and VirB10 are localized in the inner-membrane, whereas VirB3, VirB7 and VirB9 are localized in the outer-membrane, 2) two ATPases- VirB4 and VirB11, are localized at the cytoplasmic face of the inner-membrane and 3) the conjugative pilus comprising components VirB1, VirB2, VirB5 and the VirB7 lipoprotein are exocellular(Fronzes et al., 2009). ATP hydrolysis powers the transportation, which is achieved by IM associated VirB4, VirB11, VirD4 and VirB2 (Babic, Lindner, Vulic, Stewart, & Radman, 2008).

VirD4 serves as the first point of interaction with substrate followed by VirB11, also known as the molecular gate to deliver substrate to the IM complex. The substrate then travels through the periplasm to the OM protein complex. The role of VirB1 and VirB5 pilus is restricted to be a host attachment device at present. However, as secretion by T4SSs is still an active area of investigation, some studies have predicted that the pilus may serve as a channel for substrate transfer into the target cell.

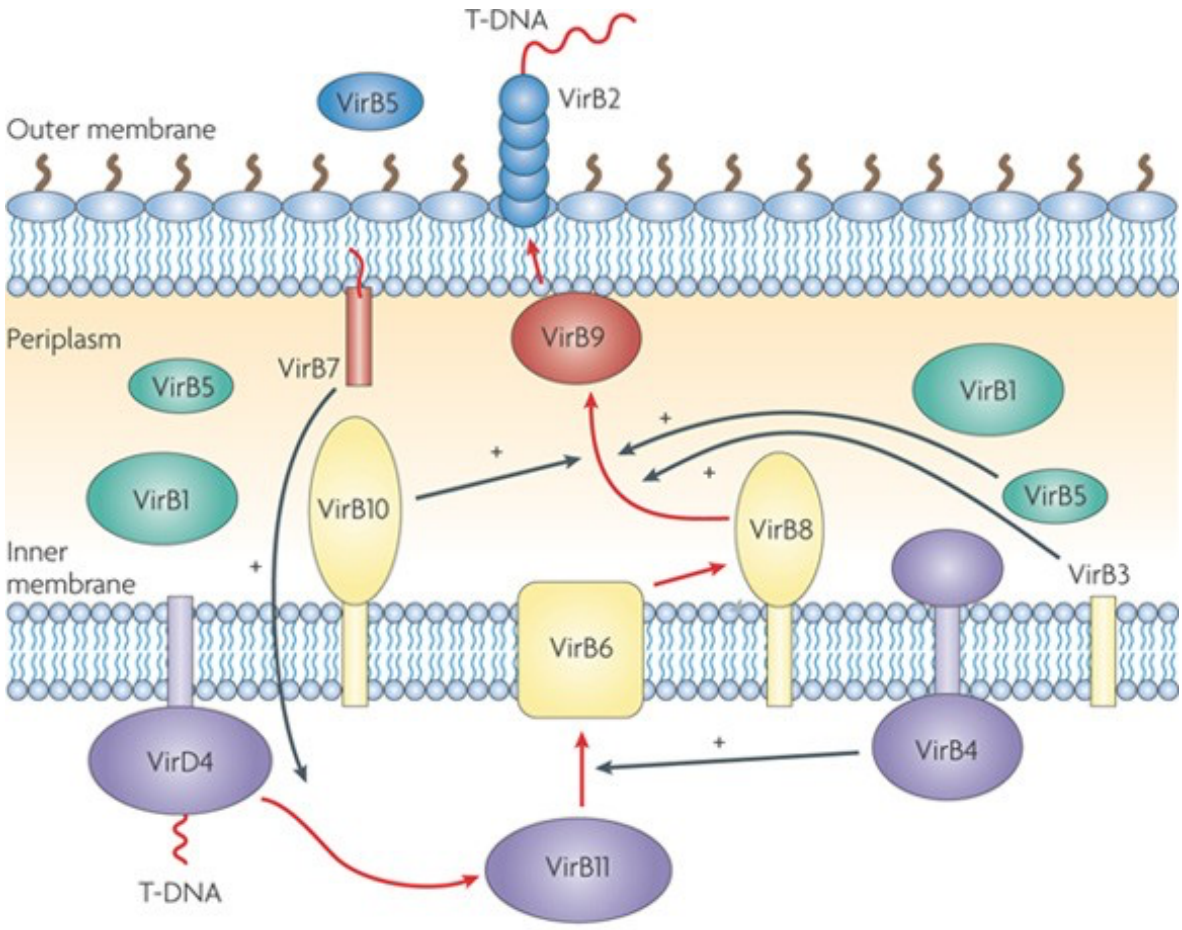


Figure 5 Overview of Type IV secretion system

The figure shows the T-DNA translocation pathway within the type IV secretion apparatus of *Agrobacterium tumefaciens*, as determined by the transfer DNA immunoprecipitation assay. The substrate pathway is shown in red arrows. The T-DNA interacts directly with VirD4, VirB11, VirB6, VirB8 and finally with VirB9–VirB2. The other components influence the transfer of the substrate at different stages, as shown by black arrows.

Fronzes, R., Christie, P. & Waksman, G. The structural biology of type IV secretion systems. *Nat Rev Microbiol* **7**, 703–714 (2009). https://doi.org/10.1038/nrmicro2218

1.2.8 Type V secretion systems

Type V secretion systems (TVSSs) are the simplest among all bacterial secretion systems in the sense that they are typically comprised of a single polypeptide chain. These relatively small assemblies are localized only to the outer-membrane (OM) of the Gram-negative

bacteria and transport its protein substrates from the periplasm across the OM. As the OM does not have any energy source to drive transport, TVSSs are also termed autotransporters. Topologically, autotransporters are comprised of a secreted passenger domain, which harbors its specific function and the transmembrane β-barrel domain, which inserts into the outer cell membrane and forms a channel that can transport secreted protein along with it. Based on the diverse structural features and domain organisation, TVSS are categorised into five sub-classes; type Va -type Ve.

Type Va secretion system

Type Va systems are the classical autotransporters with the first autotransporter of this family described being IgA1 protease from *Neisseria gonorrhoeae*(Emsley, Charles, Fairweather, & Isaacs, 1996). The type Va ATs are comprised of a single polypeptide, wherein the transmembrane β-barrel domain is located at the C-terminus of the secretory passenger domain (Figure 6). Depending on the diversity of the passenger domain function, type Va secretion systems are further classified into SPATE (serine protease autotransporters of Enterobacteriaceae) proteases, SPATE-like and non-SPATE proteases(Ruiz-Perez & Nataro, 2014; Yen, Kostakioti, Henderson, & Stathopoulos, 2008). In some cases, including IgA1 protease, the passenger domain is cleaved off in an autocatalytic process after translocation is complete, releasing the passenger into the extracellular medium but leaving the translocation domain in the outer membrane. Other type Va systems are not auto-processed in this way and thus keep the passenger domain anchored to the cell surface via the translocation domain. Typically, the passengers of non-cleavable type Va systems function as adhesins, while the cleavable ones have enzymatic (protease or lipase) activities(Barnard,

Dautin, Lukacik, Bernstein, & Buchanan, 2007; Charbonneau, Berthiaume, & Mourez, 2006; Dautin, Barnard, Anderson, & Bernstein, 2007).

Type Vb secretion system

Type Vb systems, at first sight, seem like a split variant of type Va systems: the passenger domain and translocation domain are two separate protein chains but are expressed from the same operon structure(Chevalier et al., 2004; Jacob-Dubuisson, Guérin, Baelen, & Clantin, 2013). Type Vb systems are thus also called two-partner secretion systems (TPSSs)(Kajava et al., 2001). The translocator proteins of type Vb systems are collectively called TpsB proteins, and the transported proteins (corresponding to the passenger domains of type Va autotransporters) are referred to as TpsA proteins. Notably, the TpsB protein that forms the translocation pore contains additional periplasmic POTRA (polypeptide transport-associated) domains, not present in classical autotransporters(Kajava et al., 2001). As two domains are made up of distinct polypeptides, unlike in the type Va secretion system, the passenger domain is released after transport without any proteolytic cleavage and therefore, would in principle allow multiple passengers being secreted by the same TpsB protein(Kajava et al., 2001).

Type Vc secretion system

Type Vc systems are the most complex autotransporter systems. Proteins of Type Vc secretion system are expressed as obligate homo-trimers on the surface of Gram-negative bacteria(Linke, Riess, Autenrieth, Lupas, & Kempf, 2006). The fact that most, if not all of them, have a major role in bacterial adhesion has led to their alternative name, trimeric autotransporter adhesins (TAAs).

Structurally, these trimeric fibers are repetitive of a bulkier head domain, a connector domain that mediates the transition from a head region to a stalk region, a slender stalk domain and a C-terminal membrane anchor domain. However, this description is based on simple cases such as YadA and fails to annotate putative TAAs in complex cases such as those with high sequence diversity or the presence of extended regions such as those with unusual coiled coils, or those with low sequence coverage to homologous/orthologous domain(Hoiczyk, Roggenkamp, Reichenbecher, Lupas, & Heesemann, 2000). Further, based on the structural studies, the molecular architecture or topology of TAAs can be classified either as “lollipop ” like- or as “beads-on-a-string” like(Hoiczyk et al., 2000). TAAs, in general, vary dramatically in size with YadA (422 aa) being one of the shorter TAAs with a length of about 23 nm, while BadA (3082 aa) is one of the longest TAAs, with a length of about 240 nm(Okaro, Green, Mohapatra, & Anderson, 2019). The various domains of TAA are described in detail in chapter 1.1.13.

TAAs are comprised of three identical subunits, each containing an N terminal passenger domain, a 70 residue connector/linker and 12 stranded β-barrel domain(Sikdar, Peterson, Anderson, & Bernstein, 2017). The passenger domains of the TAAs have highly diverse sequences that fold into a coiled-coil stalk. Despite its importance in pathogenesis, the secretion of the passenger domain is poorly understood. Earlier it was believed that the C-terminal anchor domain is responsible for the channel formation into the outer membrane that allows the passage of the passenger domain onto the bacterial surface and hence named ‘autotransporter’. However, with the emerging evidence of a β-barrel architecture, with the central aqueous pore of 1 nm, translocating significantly larger folded proteins (~ 2 nm) suggested the role of other cellular players in passenger translocation. Further evidence of

the translocation domain secreted before the complete assembly of the β-barrel together with the barrel assembly machinery A(BamA) dependent folding of the β-barrel of the AT, strongly suggested that the secretion of the passenger domain is mediated by the hybrid channel comprised of β-barrel of the AT and BamA β-barrel in an open conformation(Fleming et al., 2016). As the trimerization of the β-barrel seems to occur in the periplasm, which is devoid of ATP, based on the fairly recent studies it is now accepted that the energy required for the passenger secretion is derived from the vectoral folding of the α helix in the extracellular space.

The type Vd secretion system

Type Vd systems are a more recently discovered family and have been recently described in the literature with a few examples studied in detail. Topologically, type Vd secretion systems resemble the type Va as well as type Vb AT(Casasanta et al., 2017; Salacha et al., 2010); for example, the C-terminal β-barrel domain is similar to the β-barrel domain of type Vb AT protein TpsB and consist of 16 β strands(Leo, Grin, & Linke, 2012). Similar to TpsB, type Vb AT also contains a POTRA domain, albeit one in contrast to two in TpsB. Similar to some of the type Va ATs, such as EstA, the passenger domains of type Vd AT attain α/β-hydrolase fold(da Mata Madeira et al., 2016; Emsley et al., 1996). Type Vd shows an additional feature that these domains are connected via an additional, periplasmic domain homologous to the periplasmic domains of the type Vb translocation pores.

Interestingly, all passenger domains in type Vd ATs have lipase activity which is mediated by a catalytic dyad comprised of Ser and Asp residues. At the moment, it is unclear how the passenger domain of type Vd ATs is secreted or what is the function of the POTRA domain. It

is however hypothesised that the POTRA domain could either function as a chaperon for the passenger or could aid in the recruitment of a protease for passenger. Similarly, it is unclear whether passenger is secreted though autoproteolysis or through cleavage by independent proteases.

Type Ve secretion system

The domain order of the type Ve systems is reversed with the C-terminal part comprising the passenger domain, while the N-terminal part forms the translocation pore. Hence, the type Ve systems are also known as inverse autotransporters. Despite this, the type Ve AT shares striking similarities with type Va system that includes a 12 stranded β-barrel domain and a secretory monomeric passenger domain. An additional, small periplasmic domain is present at the N-terminus of the polypeptide chain, and is not homologous to the periplasmic domains in Vb and Vd systems. The passenger domain of type Ve ATs typically assumes Ig-like and lectin-like folds and remains attached even after translocation. Notably, type Ve ATs are widespread in adhesins from Gram-positive bacteria and in many other cell surface and periplasmic proteins.

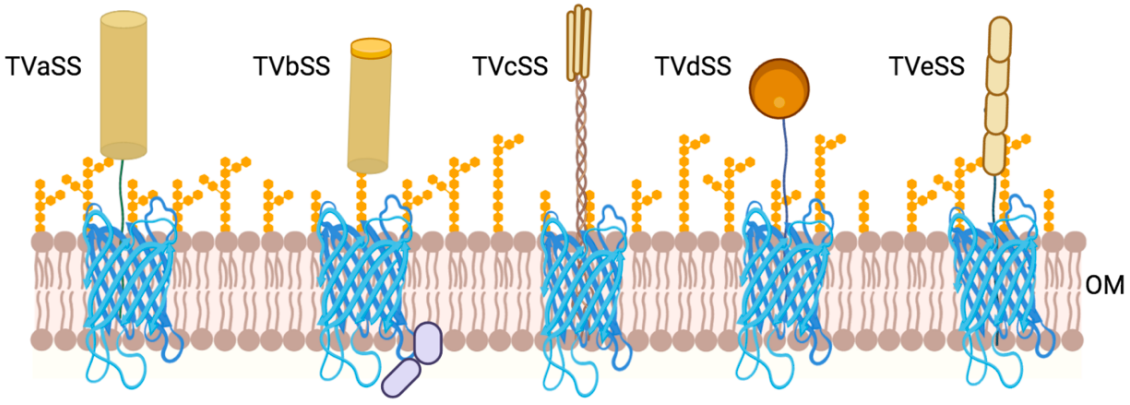


Figure 6 Schematic of type V a-e secretion systems.

1.2.9 The Type VI Secretion System

Type VI secretion systems (T6SSs) are among the most recently discovered specialized secretion systems of Gram-negative bacteria and are implicated in either bacterial pathogenesis, against eukaryotic cells, or bacterial competition, by killing neighbouring bacteria(Bingle, Bailey, & Pallen, 2008; Boyer, Fichant, Berthod, Vandenbrouck, & Attree, 2009). Structurally, this double membrane spanning apparatus is comprised of three major complexes; a tail complex- that is comprised of assembly of the inner tube and sheath, a cytoplasmic baseplate complex- that serves as a platform to form the tube and sheath of the tail complex, and an inner membrane (IM) complex- that anchors the baseplate to the inner membrane and functions as a channel for the passage of the tail tube-spike and maintaining the cell integrity during the process(Hernandez, Gallegos-Monterrosa, & Coulthurst, 2020; Records, 2011). The IM complex is comprised of TssJ, TssL and TssM proteins, ' baseplate is comprised of TssE, TssF, TssG and TssK) proteins, an inner tube inside the sheath is made up of a polymer of hemolysin co-regulated protein (Hcp) (Durand et al., 2015; Nazarov et al., 2018), and the sheath itself is comprised of TssB and TssC proteins(Bröms, Ishikawa, Wai, & Sjöstedt, 2013; Kube et al., 2014; Zhang et al., 2013). Apart from the core components, T6SSs show considerable genetic variation and hence are categorized into six sub-families (1, 2, 3, 4a, 4b, 5).

T6SSs use their contractile-based mechanism to transport effector proteins into the host cells in a single-step process but in two distinct ways(Hernandez et al., 2020). The first single-step delivery method is "cargo" effector wherein effector proteins are recruited to the Hcp tube through non-covalent interaction with (valine–glycine repeat G) VgrG spike and/or PAAR

repeat proteins. However, the second single-step delivery method “specialised” effector has additional homologues of either Hcp, VgrG or PAAR that carry an additional effector domain fused covalently to the core, mostly at the C-terminus(Alcoforado Diniz, Liu, & Coulthurst, 2015; Ma et al., 2017; Pissaridou et al., 2018; Whitney et al., 2014). The secretion by T6SSs is achieved in three steps- extension, contraction and firing, and disassembly collectively known as the firing event. Multiple homologues of Hcp, VgrG and PAAR are present in T6SSs. All of these homologues are linked with their cognate effectors. These homologues are localised within the main T6SS gene cluster as well as dispersed around the genome. This results in the formation of different combinations of Hcp, VgrG and PAAR complexes used for delivering different sets of specialised effectors in different firing events. Recognition of effectors by the T6SS machinery is built-in for specialised effectors but remains incompletely understood for cargo effectors.

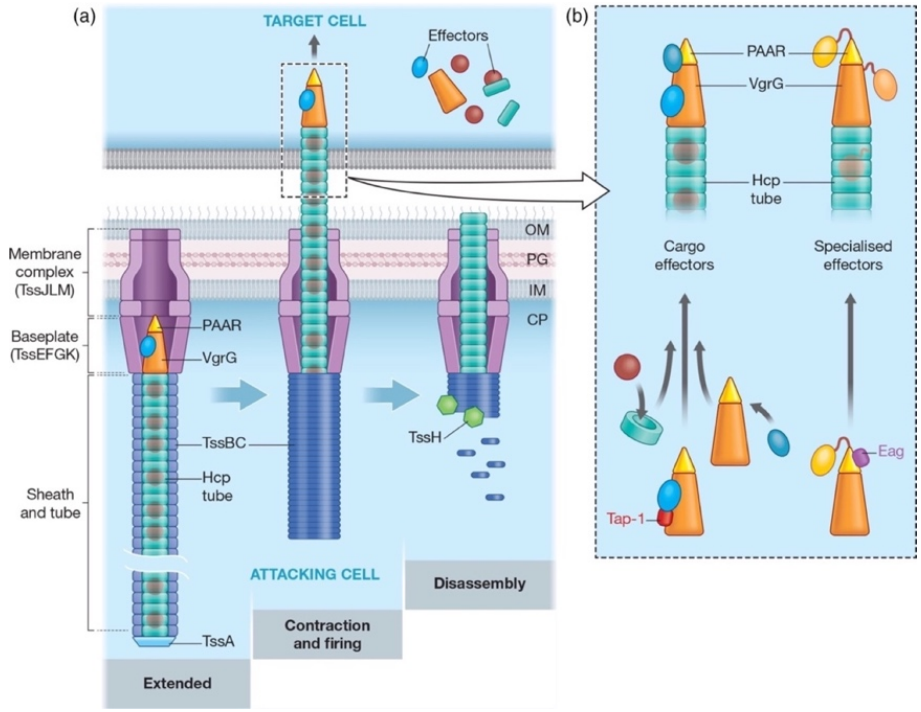


Figure 7 effectors using the same type of core component.

(a) Schematic representation of assembly, firing and disassembly of the T6SS. The T6SS assembles in the "extended" conformation (left). The cytoplasmic baseplate (TssEFGK), containing the VgrG–PAAR spike, docks on the membrane complex (TssJLM) and then acts as a platform for the assembly of the Hcp tube surrounded by the extended TssBC sheath. In a rapid "firing" step (middle), contraction of the TssBC sheath propels the Hcp–VgrG–PAAR structure out of the secreting cell and into an appropriately positioned target cell, where effectors are somehow released. Finally, disassembly (right), involves depolymerisation of the contracted sheath by TssH, allowing the system to partially or fully disassemble, often move to a new location, and begin another round of assembly and firing. CP, cytoplasm; IM, inner membrane; OM, outer membrane; PG, peptidoglycan cell wall.

(b) Schematic representation of the different modes of delivery of T6SS effectors. Effectors delivered by the T6SS can be classified as either "cargo" or "specialised" effectors, based on their interaction with the components of the puncturing structure (Hcp, VgrG or PAAR). Cargo effectors (left) non-covalently interact with the lumen of Hcp hexamers or the exterior of the VgrG–PAAR spike, whereas specialised effectors (right) represent covalent fusions between core Hcp, VgrG or PAAR proteins with an additional effector domain. In some cases, specific adaptor proteins are required for stabilisation and/or loading of particular effectors during T6SS assembly. Tap-1 adaptors (DUF4123, also known as Tec) assist loading of cargo effectors onto VgrG proteins, while Eag adaptors (DUF1795) act as chaperones to stabilise PAAR domains in specialised effectors.

This figure is borrowed from open access article cited below-

Hernandez, RE, Gallegos-Monterrosa, R, Coulthurst, SJ. Type VI secretion system effector proteins: Effective weapons for bacterial competitiveness. *Cellular Microbiology*. 2020; 22:e13241. https://doi.org/10.1111/cmi.13241

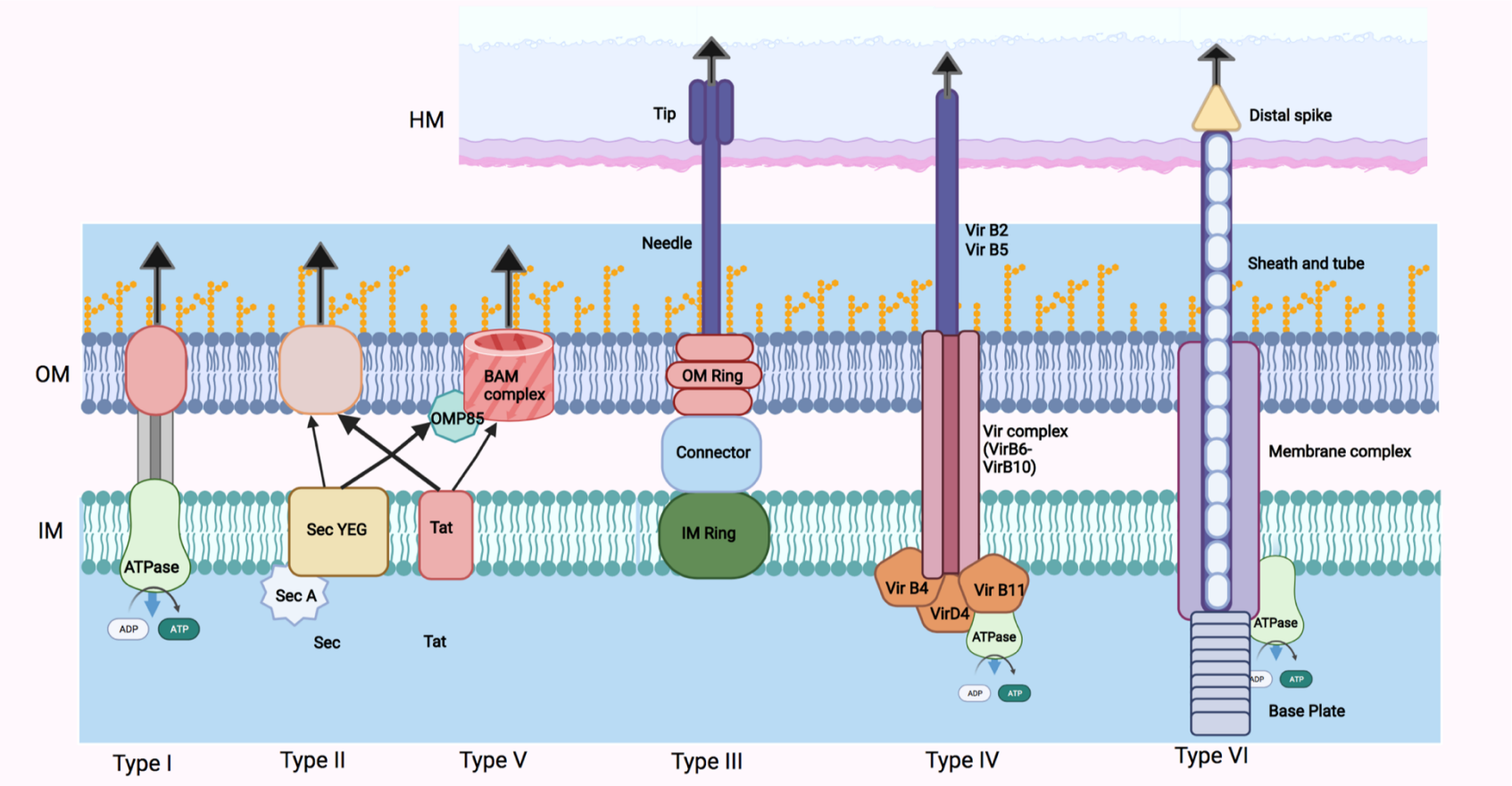


Figure 8 Overall schematic representation of Type I, II, V, III, IV and VI secretion systems of Gram-negative bacteria.

This is a cartoon representation of secretion systems in Gram-negative bacteria spanning through Inner-, outer- membranes of bacteria and host epithelium

1.2.10 Characteristics and functions of TAAs

In Gram-negative bacteria, there are several classes of secretory proteins that either function as virulence factors or provide a defence to the bacteria against the host defence system. One of these classes that not only provides defence to the Gram-negative bacteria but also stimulate their colonization within the host are Trimeric Autotransporter Adhesins (TAAs). There are more than 700 members in the TAA family of proteins and they constitute one of the largest families of Gram-negative bacteria(Ishikawa, Shigemori, & Hori, n.d.). YadA from *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* is one of the well-characterized TAAs(Wollmann, Zeth, Lupas, & Linke, 2006). Hia and Hsf from *H. influenzae*, NadA from *Neisseria meningitidis*, Eib subgroup of proteins from *E. coli* are other prominent examples of TAAs.

In recent years, much effort has been devoted to the study of TAAs and their mechanisms. The highly adhesive nature of TAAs is known to be one of their chief functions. For conquering different niches, TAAs not only adhere to living materials but also can attach to non-living objects. This also helps in bacterial adaptation to changing environments. Considering TAAs are crucial to the colonization of bacteria, defence plays an important role in host immune evasion, and are ubiquitous in bacterial pathogenesis, they are considered potential targets for better diagnosis and drug targets(Hallstrom, Trajkovska, Forsgren, & Riesbeck, 2014; Qin, Wang, & Lei, 2015; Sandt & Hill, 2000).

1.2.11 Structural annotation of TAA domains and their arrangements

As mentioned earlier, TAAs are trimeric fibers with several repeats of a head domain, a connector domain, a slender stalk domain and a C-terminal membrane anchor domain(Bassler, Hernandez Alvarez, Hartmann, & Lupas, 2015; Leo et al., 2011). Each of these domains is described in detail below.

1.2.12 Head

Head domains consist mainly of β-strands and can be divided into two types of structures; transversal, wherein β-strands run perpendicular to the fiber axis, and interleaved, wherein β-strands run parallel. Among transversal heads, a very common motif is the left-handed parallel β-roll (LPBR), as seen in the crystal structure of the recombinant collagen-binding domain of *Yersinia adhesin* YadA, in which two β-strands from each subunit contribute to a single superhelical turn, creating a left-handed solenoid like structure(Bassler et al., 2015). The other type of head structure attains the interleaved topology with each of the three β sheets composed of five β-strands from all three subunits interleaved together. This structural element is present for example in the head domain of *Bartonella henselae* autotransporter BadA(Okaro et al., 2019). Both of these motifs suffice to classify most of the TAA head domains, which is important to determine structural conservation inside each class. The LPBR structure is composed of a 14 residue repeat motif- 7 residues of which form the inner face, whereas the other 7 form the outer face. Although the residue pattern of both faces seems to have a similar pattern, they are distinct in two major aspects; 1) The inner face residues are hydrophobic as they form the inner core whereas the outer face residues are

hydrophilic as they face the solvent and 2) The inner face residues are constrained as they are oriented towards another subunit whereas the outer face residues are not constrained as they are not oriented towards the trimer(Bassler et al., 2015). Among interleaved head structures, the most widespread is the tryptophan ring (TrpRing) domain, named after the presence of highly conserved Trp residue in the first β strand which together with the equivalent Trp residues of the other two subunits forms unique ring geometry. Interestingly, in contrast to the LPBR heads, which transit through connectors to a stalk segment, TrpRing heads do not need connectors, rather they usually connect to the coiled-coil of a stalk through their N terminus.

While the head domains do mediate many molecular interactions ranging from autoagglutination to the attachment to extracellular matrix (ECM) components, the stalk domains also contribute; for instance, the stalk of BadA binds to fibronectin and two different regions of the stalk of EibD bind to IgA and IgG, respectively. Therefore, to identify the interaction partners of a TAA, all regions of the passenger domain, not just the N-terminal head domains, must be considered.

1.2.13 Connectors

YadA and UspA structures reveal that TAA includes a connector domain also known as the Neck domain. It is present in most adhesins and usually occurs in several copies (up to 24) per protein. The neck domain is one of the most highly conserved extracellular sequence elements in Gram-negative bacteria. They connect bulkier heads composed of β-strands and slimmer stalk domains. These conserved neck domains can be classified into two categories α-to-β, going from head to stalk and β-to-α, going from stalk to head(Bassler et al., 2015).

1.2.14 Stalk

The stalk composes the central region of all TAAs. One of the functions of stalks is to project the head out of the bacterial surface. The stalk consists of sub-domains that vary greatly in number among different members of the family. Due to this reason, the stalk domain varies greatly in length in different members of the trimeric autotransporter family. Hsf contains three binding domains and three putative domains which is considerably more than the other well characterised member of the family Hia, which has only two binding domains(St. Geme et al., 1996). The stalk domain appears to assume an extended coiled-coil architecture, a bundle of α-helices interacting through seams of heptad repeats that run through the helices and comprised of hydrophobic residues (h) that are separated by polar residues (p), that is hpphppp, wherein their positions are described by the letters A to G, with hydrophobic residues in A and D positions(Bassler et al., 2015). The heptad coiled coil assumes the left-handed supercoil conformation so as to accommodate 3.5 residues per turn and is the predominant form of the stalk in TAA. Often, variability in residue composition or periodicity of coiled-coil repeats of the TAA stalk segment has been observed. The former includes unusual residue composition in heptad repeats, such as polar residue in the D position or both the A and the D positions (e.g. RxD motif) or even at the cores of the coiled-coil segments. The periodicity other than heptad includes hendecad- repeats of 11 residues and pentadecad-repeats of 15 residues. As the number of residues per turn for hendecad and pentadecad periodicity is either the same as that in an undistorted α-helix (3.63) or slightly higher (3.67 and 3.75, respectively), the former assumes fairly straight supercoiled coil conformation, whereas the latter assumes right handed one. The right-handed helix has a common sequence motif- YxD, which forms a ring of polar residues similar to RxD motif observed in the heptad

repeats of the left-handed coiled-coil stalk. Insertion of a nonhelical motif has also been observed in coiled-coil segments of the stalk and was termed as FGG, which contains the frequent occurrence of LGG. This insertion of FGG motif disrupts the N-terminal segment of the coiled coil to form a 3-stranded β-meander. While FGG motifs are very common among TAAs, a few other non-helical insertions have been also identified so far. The saddle of Eib proteins serve as an example of other nonhelical insertion which connects the upper hendecad stalk region to the lower heptad stalk region.

1.2.15 Anchor

The anchor region is the highly conserved region of TAAs, and is composed of three subunits, each of which is comprised of a 12 stranded β-barrel(Koiwai, Hartmann, Linke, Lupas, & Hori, 2016). This assembly is formed by inserting β-strands into the outer-membrane as soon as TAAs are exported through the IM via the Sec pathway. It is thought to function as a channel for the translocation of the head and passenger domains(Chauhan et al., 2019)(Mühlenkamp, Oberhettinger, Leo, Linke, & Schütz, 2015). This is followed by the trimerization of helices to form a coiled-coil at the centre of the β-barrel to block the passage and stabilize the assembly. It also acts as an anchor as it holds the passenger domain onto the bacterial surface(Chauhan et al., 2019).

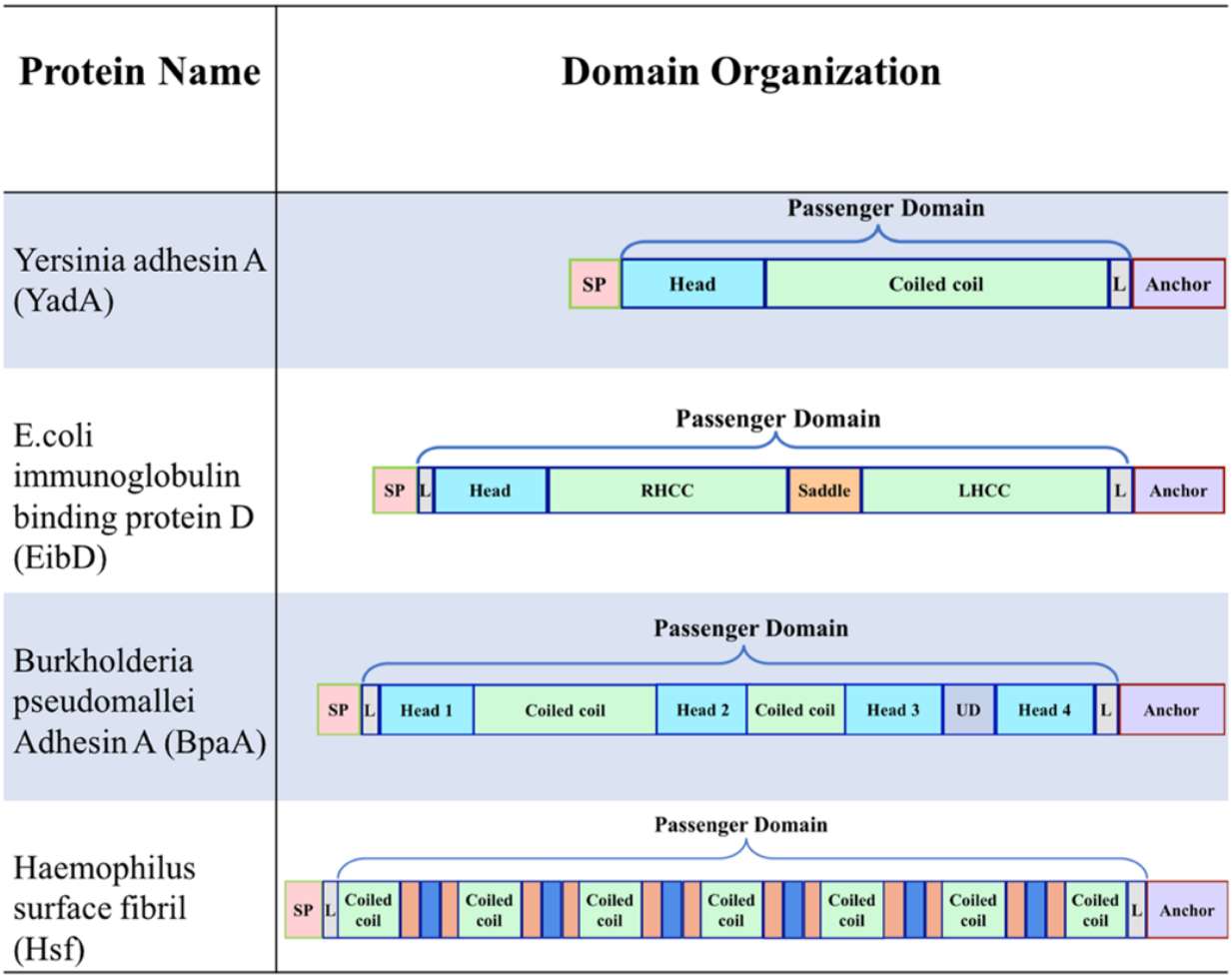


Figure 9 Domain organization in TAAs.

Difference in domain organization of YadA, EibD, BpaA and Hsf. All TAAs contain a signal peptide (SP), passenger domain and anchor domain. The sub-domain organization in passenger domain varies among the proteins of TAA family. Head (cyan): YadA like heads, 'L' (grey): small linker regions; Coiled coil, RHCC (right-handed coil coil), LHCC (left-handed coiled coil) (green): coiled-coil stalk regions that vary in length and handedness. UD (grey): undefined domain. The Trp ring domains are another example of head and are shown by orange colour, whereas the blue colour represents various neck domains present in Hsf.

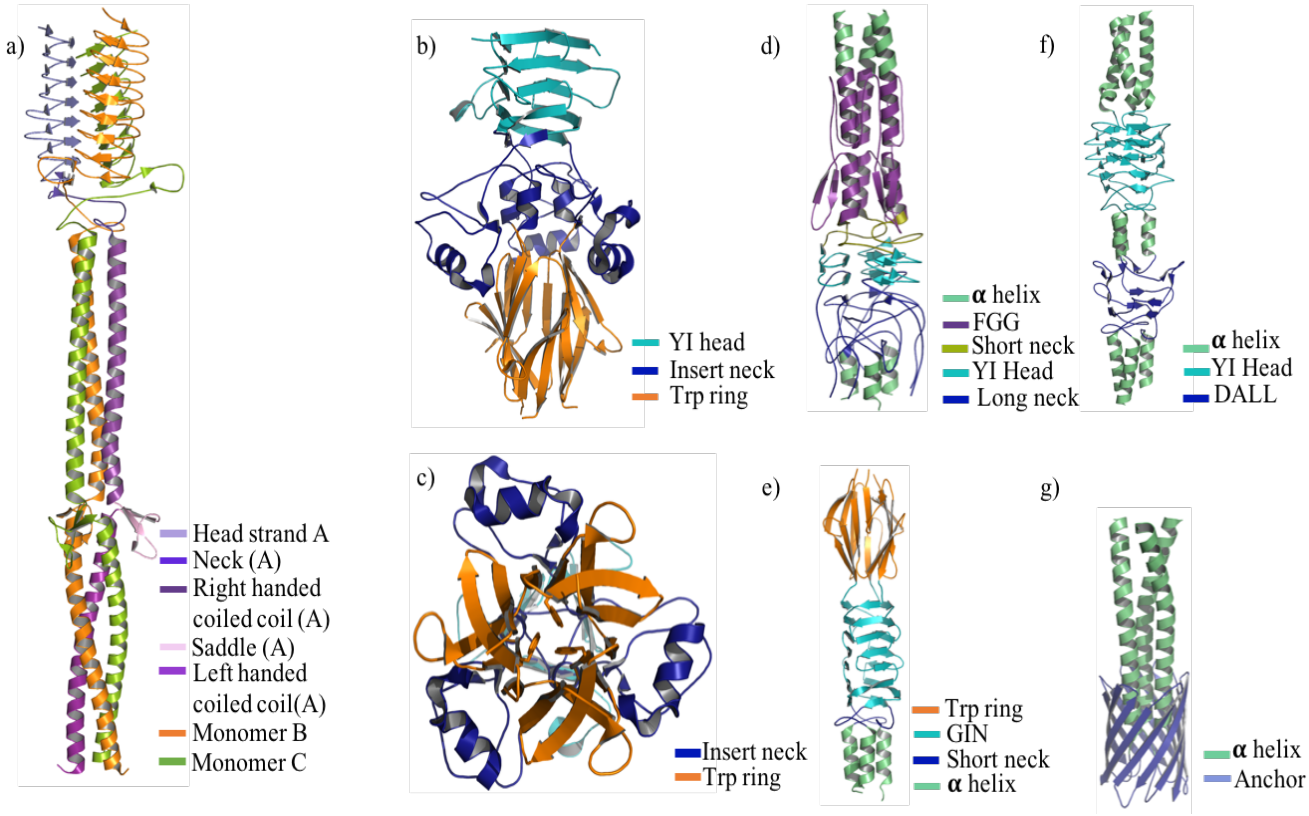


Figure 10 Different domains from TAA family of proteins.

a) The structure of EibD8 (PDB ID: 2XQH) constitute of three identical monomers twinned among themselves. The first monomer is colour coordinated into different shades of purple representing the monomer contribution in head, neck, right- handed coiled coil, saddle and left-handed coiled coil domains of EibD, whereas, the other two monomer chains are represented in orange and green colour. b) The crystal structure of binding domain 1 (BD1) of Hia (PDB ID:1S7M) divided into head (Teal), insert neck (Blue) and Tryptophan (Trp) ring (Orange) domains. c) Top view of Hia BD1(PDB ID:1S7M) representing tryptophan side chain from each monomer. d) Structure of BapA (PDB ID:3LAA) representing FGG (Violet purple), short neck (Olive), head (Teal), long neck (Blue) domains. e) Structure of BadA (PDB ID: 3D9X) representing Trp ring (Orange), GIN (Teal), short neck (Blue) domains. f) Structure of SadA (PDB ID:2YO3) representing DALL (Blue) domain. g) The structure of anchor domain of Hia (PDB ID:2GR7).

1.2.16

1.3 The severity of *H. influenzae*

As per the European Centre for Disease Prevention and Control (ECDC) report, the rate of confirmed cases per 100,000 of invasive *H. influenzae* disease rising steadily since 2011 in some of the European Union countries- such as Germany, France, Spain, Norway, Finland and

also in the United Kingdom (UK). This is in spite of the national coverage of the *H. influenzae* vaccination and was predominantly reported in infants and elderly people.

Table 1 Rate of confirmed cases of invasive *H. influenzae* infection per 100,000 cases

Country	2011	2012	2013	2014	2015	2016	2017	2018
Germany	0.3	0.4	0.5	0.6	0.7	0.7	1.0	1.0
France	1.0	1.0	1.0	0.9	1.1	1.2	1.2	1.2
Spain	0.5	0.4	0.4	0.6	0.6	0.6	0.8	0.9
Norway			1.7	1.4	1.9	1.6	2.3	1.7
Finland	1.2	1.5	0.9	1.1	1.0	1.3	1.3	1.6

Despite the name, *H. influenzae* is not related to the influenza virus and causes the most common type of invasive diseases that includes pneumonia, bacteremia, meningitis, epiglottitis, cellulitis, infectious arthritis, septicemia, and osteomyelitis. All of these diseases are caused by the encapsulated strains of *H. influenzae*, which are categorized into six serotypes- a to f. The most invasive of all is *Haemophilus influenzae* serotype b (Hib). Hib enters through the nasopharynx and spreads from the upper respiratory tract to the lower respiratory tract. In the advanced phase of infection, Hib disseminates to the distant parts of the body through the bloodstream. One of the initial steps in colonizing upper respiratory tract involves the adherence of Hib in the respiratory mucosa, which is mediated by its specialized surface proteins commonly known as adhesins.

1.3.1 Mode of infection

Hib commonly colonizes the respiratory tract and occasionally becomes pathogenic leading to infection. A prerequisite for disease development is epithelial adherence, colonization, evasion of the immune system, and finally penetration of cellular barriers. Several studies have examined Hib respiratory colonization, persistence and invasion, and have identified signalling and trafficking pathways involved in Hib pathogenesis(King, 2012; Slack et al., 2021).

1.3.2 Major Adhesins of Hib

Five major adhesins have been identified in the *H. influenzae* serotype b; namely, *Haemophilus influenzae* adhesin (Hia), Haemophilus surface fibril (Hsf), Haemophilus adherence and penetration protein (Hap), High molecular weight proteins (HMW), and hemagglutinating (HA). Among these, Hia is well characterised and therefore has been considered as a model protein for autotransporter adhesins(Koretke, Szczesny, Gruber, & Lupas, 2006). Notably, Hsf is a large protein of 790 kDa, which is found in all typeable strains of H. *influenzae*. Further, Hsf has been shown to function as an adhesin and interacts with the host factor vitronectin, suggesting its direct role in virulence. Altogether, Hsf is very important target to examine the pathogenicity of Hib.

Hallström *et al*., in 2006 showed direct binding of Hsf-BD1 and Hsf-BD2 domains with vitronectin. Vitronectin inhibits the membrane attack complex of the host complement system through its binding to C5b-7 and C5b-9 (complement system components)(Hovingh, van den Broek, & Jongerius, 2016). Therefore, Hsf is thought to be directly involved in the survival of the *H. influenzae*(Milis, Morris, Sheehan, Charlesworth, & Pussell, 1993).

1.3.3 Available cures of Hib Infection

In 1985, the first vaccine to protect against Hib was introduced, which was replaced by an improved vaccine licensed two years later. Single as well as combined shot vaccinations for Hib are available. But the problem persists with their storage, transportation and majorly the cost of these vaccines. Many under developed and developing countries cannot afford these vaccines and do not have appropriate storage conditions. Apart from the vaccinations, there are antibiotics available to cure Hib infection at an early stage, such as ceftriaxone, cefotaxime, cefuroxime or chloramphenicol in combination with ampicillin, however, emerging resistance against these antibiotics has become alarming.

1.3.4 The model of Hsf

Hsf is 2414 amino-acids long and can be divided into a long N-terminal passenger domain and C-terminal translocator domain. The N-terminal passenger domain is further divided into three putative domains (PD) named PD1, PD2, PD3 and three binding domains (BD) named as BD1, BD2 and BD3. PD domains share more than 90% sequence similarity among each other and so do all the BD domains. Intriguingly, the PD also share more than 70% sequence similarity with each of the BDs. These domains are connected by a small stretch of sequence forming a helical structure, while each of these domains themselves contain two tryptophan rings separated by small necks. The passenger domain is a long stalk-like projection onto the outer membrane of gram-negative Hib. In 2014, Singh *et al*. proposed a unique hairpin model of the N-terminal passenger domain. The aforementioned conclusion was based on the author's claim of the hairpin shape of Hsf on the surface of *H. influenzae* in a transmission electron microscopy (TEM) based experiment. However, no hairpin-like bend was observed in any of the images. Further, for the experiments performed with the purified Hsf, the purity

of the proteins remains unknown. Therefore, in the absence of any significant structural details of Hsf, it becomes highly important to study the structure of either full length Hsf or the passenger domain of Hsf or its central part- PD3BD3. Further, considering their presence in all typeable strains of *H. influenzae*, this could also be helpful for developing new diagnostic tools against Hib.

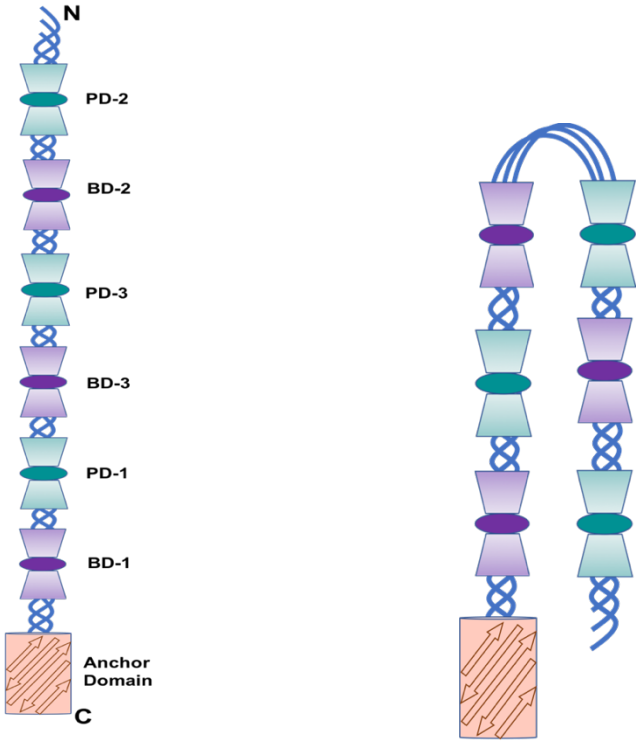


Figure 11 Organisation of domains in Hsf

The figure shows two hypothesised confirmations of Hsf-

a) The first cartoon is based in the bioinformatic model of Hsf, which attains a straight stalk like confirmation. The domain annotations are labelled in the figure.
b) The second cartoon shows the hypothesised modle based on the EM studies, forming a hairpin-like architecture.

1.3.5 Structure of Hsf-PD1 domain

The crystal structure of trimeric PD1 was solved at a resolution of 3.3 Å, which provided the first insight into the molecular arrangement of Hsf to date(Wright et al., 2017). HsfPD1 crystallized in the monoclinic space group *C*2 with nine monomers in the asymmetric unit. The individual monomers of HsfPD1 are comprised of three distinct domains that fold to form well characterized TAA domains. A proposed N-terminal TrpRing domain, a KG domain and a C-terminal TrpRing domain are seen in each PD1 monomer. The N-terminus of HsfPD1 spans 29 amino acids participating in the formation of three β-sheets, $\beta_{W1}1$, $\beta_{W1}2$ and $\beta_{W1}3$ (where 'W' represents tryptophan). Although the sequence identity between these two regions is low (31%; Supplementary) a structural alignment of the 29 N- and C-terminal residues from one HsfPD1 monomer confirms that the N-terminal region is indeed a TrpRing domain. The KG domain is composed of two β-strands, $\beta_{KG}1$ and $\beta_{KG}2$, as well as three α-helices, $\alpha_{KG}1$, $\alpha_{KG}2$ and $\alpha_{KG}3$. The C-terminal TrpRing is composed of five β-strands: $\beta_{W2}1$, $\beta_{W2}2$, $\beta_{W2}3$, $\beta_{W2}4$ and $\beta_{W2}5$. All domains participate in extensive intertwining, where the C-terminal α-helices ($\alpha_{KG}3$) from each monomer come together to create the central core of the trimer interface. The KG and C-terminal TrpRing domains were easily identified by simple structural observation and comparison with other TAAs.

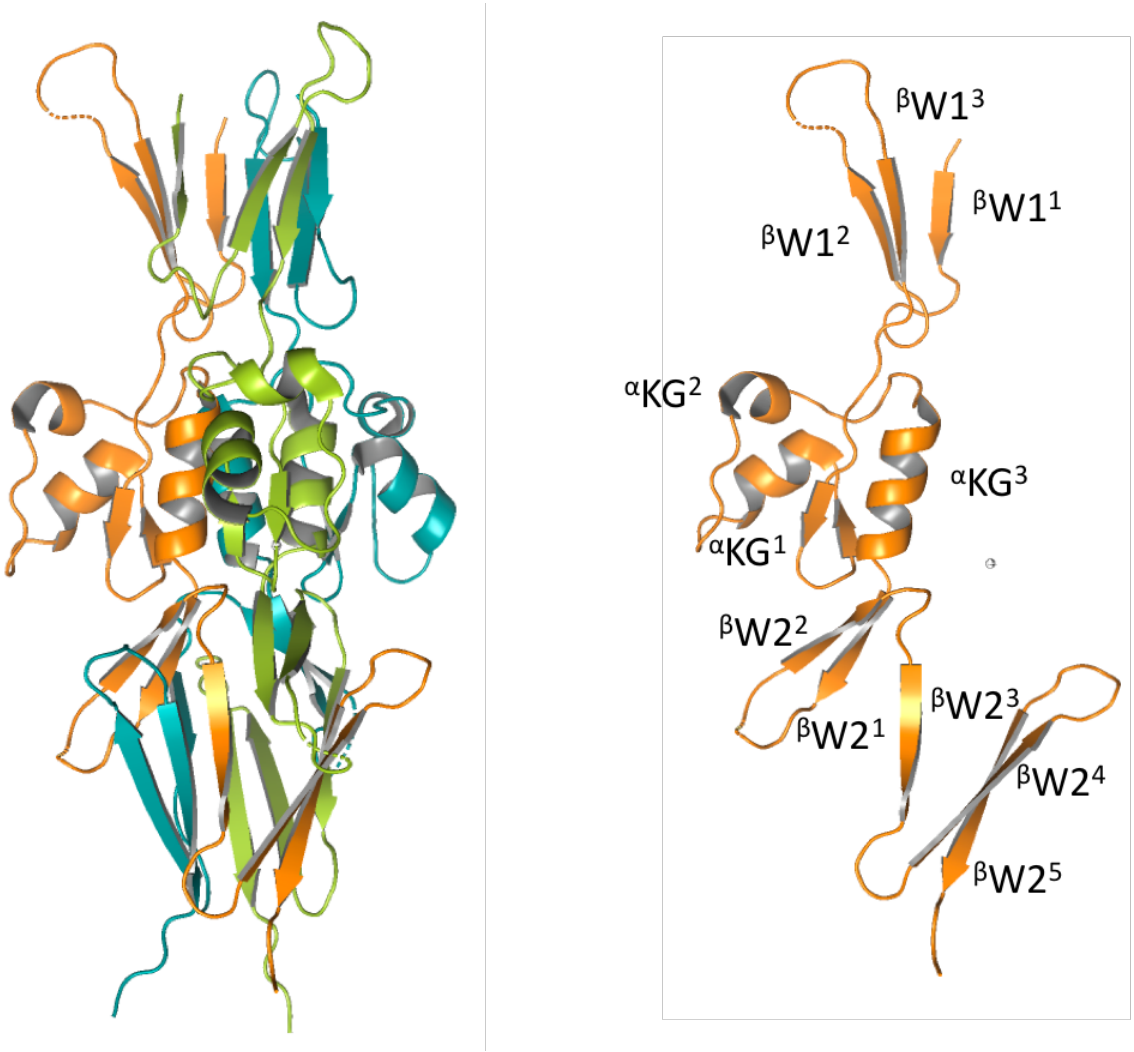


Figure 12 The structure of HsfPD1.

(*a*) HsfPD1 showing a trimeric architecture (three monomer subunits; lime, teal and orange). (*b*) One HsfPD1 subunit with labelled secondary structure showing a novel domain arrangement: N-TrpRing ($^{\beta}$W1):KG ($^{\alpha}$KG):TrpRing-C ($^{\beta}$W2):

1.4 Shiga toxin producing *E. coli*

One of the four widely recognized types of *E. coli* is Shiga toxin-producing *E. coli* (STEC), of which *E. coli* O157:H7 is a prominent cause of bloody diarrhoea and the severity of the illness can occasionally result in kidney failure and even death. Patients under the age of five or older

than sixty-five make up the majority of patients. STEC exposure poses a serious threat to those who have weaker immune systems, are pregnant, or travel to particular nations. *E. coli* begins colonizing the host's gastrointestinal tract as soon as it enters the body through tainted, undercooked, or unprocessed food. Inside host's GI tract TAA subgroups known as *E. coli* immunoglobulin binding proteins (Eibs) produced on the surface of *E. coli* promote autoagglutination and aid *E. coli* in colonizing and adapting to a new environment.

1.4.1 *E. coli* immunoglobulin binding proteins (Eibs)

Eibs is a protein subgroup composed of six proteins: EibA, -C, -D, -E, -F, and -G. The most intriguing function of the Eib family of proteins is that they operate as Fc receptors (FcR) for *E. coli*, allowing them to bind to antibodies in non-immunological ways, such as macrophages, neutrophils, eosinophils, and so on. EibC, -D, and -E bind to both Immunoglobulin G (IgG) and Immunoglobulin A (IgA), whereas the others attach solely to IgG. EibC, -D, and -E bind to both Immunoglobulin G (IgG) and Immunoglobulin A (IgA), whereas the others attach solely to IgG. This characteristic allows E. coli to be recognized as self-cells rather than antigens, allowing them to colonize the host while suppressing its immune system. This emphasizes the significance of studying Eib proteins. I also worked on the structure of the EibD: IgG-Fc complex, as well as a heterotrimeric complex of EibD binding to IgG-Fc and IgA.

1.4.2 *E. coli* immunoglobulin binding protein D (EibD)

EibD conforms to the stereotypical architecture of trimeric autotransporter adhesins with N-terminal domains. The head, neck, stalk, and C-terminal anchor domains are all present. The head domain resembles the well-known and model YadA-like head. The stalk area is separated into two halves, with the upper half forming right-handed coiled coil (RHcc)

secondary structures and the lower half forming left-handed spiral coil (LHcc). A short neck region known as the saddle separates the two parts RHcc and LHcc. LHcc includes the binding region for IgG-Fc, while RHcc has the binding region for IgG-Fc. The stalk domain is followed by the C-terminal anchor domain, which is conserved.

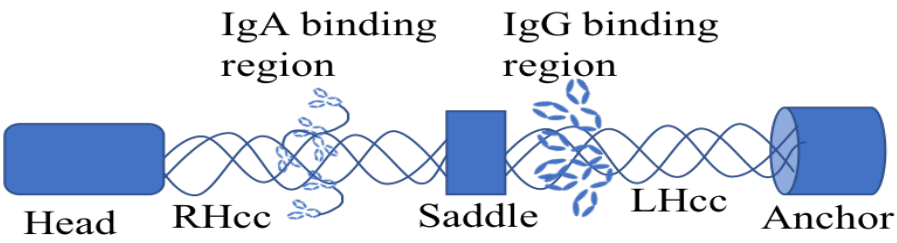


Figure 13 Diagram of EibD showing IgA and IgG binding sites

Scematics of EibD representing head, stalk and anchor architecture from N to C terminals. RHcc and LHcc are right-handed and left-handed coild coils respectively showing the IgA and IgG binding regions.

1.5 Aims and Objectives

The overall aim of this project was to understand the molecular mechanisms behind Hsf's ability to adhere to the proteins present on the host epithelium, focusing predominantly on the adhesive properties of the binding and putative domains. To achieve this, robust overexpression, extraction and purification procedures were established, and the purified Hsf domains were used for subsequent biophysical and structural characterisation. X-ray crystallography, and cyclic voltammetry were the primary techniques used to characterise the purified Hsf domains, to elucidate a potential structural and functional information. Stability and homogeneity of the domains were important for success in X-ray crystallography, and cyclic voltammetry.

Biophysical characterisation of the isolated protein product gave significant and novel insights into the mechanism of adhesion. In addition, the adhesion of different domains with interacting partner proteins such as vitronectin and different surfaces, namely carbon, cobalt and polymers, were explored which may lead to generating a novel method of rapid and cost-effective diagnosis.

Chapter 2 Materials and Methods

2.1 Materials

2.1.1 Chemicals, crystallization kits and biological media

Chemicals used in this study were obtained from various commercial sources. Restriction endonucleases (Fast Digest) and modifying enzymes were procured from New England Biolabs (NEB), U.K. In-fusion HD cloning Kit® was purchased from Takara Bio, USA. PCR cloning kits were obtained from Thermo Scientific, USA, and the expression vector system (pET 28) was available in the lab. The plasmid DNA isolation and purification kit were procured from QIAGEN, Germany. Fine chemicals including those for preparing solutions for protein purification and crystallization were bought primarily from Sigma, USA. Ni-NTA agarose beads were obtained from QIAGEN, Germany. Pre-stained Protein molecular weight marker was obtained from Thermo Scientific, USA. DNA GeneRuler markers were also obtained from Thermo Scientific, USA). Crystallization plates (96 well and 24 well-sitting drop plates) were obtained from Molecular Dimensions (UK). The biological media components were obtained from Merk, Sigma Aldrich and Fisher Scientific. All antibiotics and centrifugation-based concentrators (15 mL) with cutoff of 3 kDa, 10 kDa and 30 kDa were obtained from Sigma Aldrich Centrifugation-based concentrators (0.5 mL and 4 mL) with a cutoff of 3 kDa were purchased from Vivaspin. The syringe filters (0.45 and 0.22 microns) were obtained from Millipore (USA).

2.1.2 Constructs and strains

Multiple truncated constructs produced from full-length Hsf and EibD plasmids were utilized for crystallisation and binding investigations. All of the constructs are listed in Table 2. The

expression of these plasmids was evaluated in several strains to optimize yield. Table 3 shows the final optimised strains for each plasmid.

Table 2 List of constructs and vectors

Hsf Plasmid	**Accession code: p71401_HAEIF**	
Construct	Vector	Insert
$Hsf_{50\text{-}2316}$	pBPT	Hsf Plasmid
PD3BD3	pET 28 c	$Hsf_{50\text{-}2316}$
PD1	pET 28 c	$Hsf_{50\text{-}2316}$
BD1	pET 28 c	$Hsf_{50\text{-}2316}$
PD2	pET 28 c	$Hsf_{50\text{-}2316}$
BD2	pET 28 c	$Hsf_{50\text{-}2316}$
PD3	pET 28 c	$Hsf_{50\text{-}2316}$
BD3	pET 28 c	$Hsf_{50\text{-}2316}$
EibD3	pQE30	EibD: Q9MCI8
EibD8	pET 28	EibD: Q9MCI8

Table 3 List of Proteins and their expression strains

Protein Name	**Strain Used**	**Resistance markers**
$Hsf_{50\text{-}2316}$	C41	Ampicillin
PD3BD3	BL21*	Kanamycin
PD1	BL21*	Kanamycin

BD1	BL21*	Kanamycin
PD2	BL21*	Kanamycin
BD2	BL21*	Kanamycin
PD3	BL21*	Kanamycin
BD3	BL21*	Kanamycin
EibD3	BL21*	Ampicillin
EibD8	BL21*	Kanamycin
IgG-Fc	BL21*	Ampicillin

2.1.3 Stock Preparation

Ampicillin

Ampicillin (Amp) is a penicillin derivative and is bactericidal. A stock solution of 100 mg/mL was prepared by dissolving 1 g of Amp powder in 10 mL of sterile water and was filtered through prewet 0.2 µm syringe filter. For the selection of the transformant colonies in solid and liquid growth media, a stock solution was diluted to 100 µg/mL.

Carbenicillin

Carbenicillin (carb) is a bactericidal antibiotic belonging to the carboxypenicillin subgroup of penicillin. A stock solution of 50 mg/mL was prepared by dissolving 0.5 g of carb powder in 10 mL of sterile water and was filtered through prewet a 0.2 µm syringe filter. For the selection of the transformant colonies in solid and liquid growth media, a stock solution was diluted to 50 µg/mL.

Kanamycin and Streptomycin

Kanamycin (kan) and Streptomycin (strep) belong to the aminoglycoside class of antibiotics. A stock solution of both antibiotics was prepared at 50 mg/mL by dissolving 0.5 g of Kanamycin Monosulfate or Streptomycin Sulphate powder in 10 mL of sterile water and were filtered through prewet a 0.2 µm syringe filter. For the selection of the transformant colonies in solid and liquid growth media, a stock solution was diluted to 50 µg/mL. The Stock solution was kept at -20 ˚C for a maximum of 1 year.

Isopropyl β-D-thiogalactopyranoside

IPTG is a nonmetabolizable analogue of allolactose that functions as an inducer of the Lac-operon by binding to and inhibiting the lac repressor so as to release it from its operator and is used for heterologous protein expression under the control of the Lac operator in *E. coli*. A stock solution of 1 M was prepared by dissolving 2.38 g in 8 mL of autoclaved ultrapure water and bringing the final volume to 10 mL. The solution was filter sterilized by passing through a 0.22 µm syringe filter. The stock solution was stored in aliquots at -20 ˚C and used at different volumes to achieve desired concentration.

Imidazole

Imidazole is a five-membered aromatic molecule containing two annular nitrogen atoms. As imidazole is a functional group of the amino acid histidine, it is used as a competitive agent for elution of histidine-tagged proteins. Exploiting this, imidazole is used at low concentrations (< 50 mM) in the sample and binding buffer to remove contaminant proteins that non-specifically bind to the IMAC column further used in elution buffer at relatively

higher concentrations (100-250 mM) to elute heterologously expressed His-tagged-proteins. The stock solution of Imidazole was prepared at 3 M as per the following, 20.42 g of Imidazole was dissolved in 90 mL of autoclaved ultrapure water, the pH was adjusted to 7.0 with HCl and the final volume was brought to 100 mL with autoclaved ultrapure water. The stock solution was sterilized by autoclaving in a glass bottle and was stable for approx. 2 years at 2-4 ˚C.

Sodium Chloride (NaCl)

5 M NaCl stock solution was prepared as per the following to be used in purification buffer preparation. 292 g of NaCl was dissolved in 700 mL of Autoclaved ultrapure water by stirring with a magnetic stirrer. After NaCl dissolved completely, the volume was brought to 1000 mL with the Autoclaved ultrapure water. The stock solution was sterilized by autoclaving in a glass bottle.

Tris buffer

Tris stands for Tris(hydroxymethyl) aminomethane used as a buffer in wide array of molecular biology experiments as its buffer capacity coincide with the physiological pH of most living system (pH: 7-9). Two different pH Tris buffers, 7.5 and 8.0 were used in different purifications which were carried at different temperatures. The stock solution of 1M Tris Buffers was prepared as per the following:

121.14 g of Tris was dissolved in 700 mL of autoclaved ultrapure water, the pH was adjusted to 7.5 or 8.0 with HCl while constant stirring and the final volume was brought to 1000 mL with autoclaved ultrapure water. The buffer was steriliszed by autoclaving the solution.

Glycerol

Glycerol is a highly viscous colourless and odourless compound. Due to its high viscosity it is is prepared in solution as w/v instead of v/v. A stock solution of 50% glycerol was prepared by dissolving 50 g of glycerol in 50 mL of autoclaved ultrapure water and was sterilized by autoclaving.

PBS

Phosphate buffered saline (PBS) buffer is used for various biological applications such as washing cells and tissues and dilution of cells as it closely mimics the ionic concentration, osmolarity and pH of the human body. A stock solution of 10 X PBS buffer was prepared as per the following. The reagents in the Table 4 were dissolved in 800 mL of autoclaved ultrapure water, its pH was adjusted to 7.4 with HCl and the volume was brought to 1 L. The stock solution was autoclaved. For all experimental work, 1 X working concentration was prepared by diluting the 10 X PBS in autoclaved ultrapure water.

Table 4 Composition for 10 X PBS buffer

Components	**Weight (g) for 1 L**
NaCl	80
KCl	2
Na_2HPO_4	14.4
KH_2PO_4	2.4
$CaCl_2.2H_2O$	1.33
$MgCl_2.6H_2O$	1.0

2.1.4 Media

LB Broth

For preparing 1 L Luria-Bertani media, 10 g Tryptone, 5 g Yeast extract and 10 g NaCl were dissolved in 700 mL of autoclaved ultrapure water and the volume was brought up to 1 L. The media was then equally divided into two 1 L flasks/bottles and autoclaved.

LB Agar

To make 500 mL LB agar, 5 g Tryptone, 2.5 g Yeast extract, 5 g NaCl and 7.5 g agar powder were dissolved in 500 mL autoclaved ultrapure water and autoclaved. In order to make agar plates, the autoclaved solution was allowed to cool down to luke warm and desired antibiotic was added, if needed. Approx. 20 mL LB-agar was added to each plate.

2.1.5 Purification buffers

Lysis buffer

Lysis buffer was used for the resuspension and lysis of the *E. coli* culture pellet. Lysis buffer was comprised of 50 mM Tris, pH 7.5, 300 mM NaCl, 5% Imidazole.

Wash Buffers

Three step washing was performed in every purification with different wash buffers in increasing order of imidazole concentration, the composition of which is described in the Table 5

Table 5 Components of wash buffers used for protein purifications

Buffers → **Stocks ↓**	**Wash-1**	**Wash-2**	**Wash-3**
1 M Tris buffer	(50 mM)	(50 mM)	(50 mM)
5 M NaCl	(300 mM)	(300 mM)	(300 mM)
3 M Imidazole	(10 mM)	(20 mM)	(30 mM)

Dialysis Buffer

Dialysis buffer was used to eliminate excess imidazole and salt from the affinity purified protein eluates. Dialysis buffer consisted of 50 mM Tris, pH 7.5, and 150 mM NaCl.

2.1.6 Gel Electrophoresis buffers

Sample buffer

DNA sample buffer is required to load DNA samples in the wells of the agarose or polyacrylamide gels and tracking its migration during electrophoresis. The DNA sample buffer was prepared in 6X concentration containing 30% (v/v) glycerol, 0.25% (w/v) bromophenol blue and 0.25% (w/v) xylene cyanol FF. The dye was used at working concentration of 1X.

Running buffer

TAE buffer was used as a running buffer for the electrophoresis of nucleic acids in agarose or polyacrylamide gels. The TAE buffer was prepared in 10x concentration as per the following;

48.5 g Tris was dissolved in 700 mL of autoclaved ultrapure water, to which 11.4 mL of glacial acetic acid and 20 mL of pre-made 0.5 M EDTA (pH 8.0) was added and the volume was brought to 1000 mL with MilliQ water.

2.1.7 SDS-PAGE Buffers

Sample buffer (5x SDS-PAGE loading buffer)

The 5X SDS-PAGE loading buffer was comprised of 250 mM Tris·HCl, pH 6.8, 10% SDS, 30% (v/v) Glycerol, 10 mM DTT, 0.05% (w/v) Bromophenol Blue and was stored at 4 ˚C. It was dilute to the appropriate concentration for use.

Running Buffer

A 10X SDS-PAGE running buffer was comprised of 250 mM Tris·HCl, pH 8.3, 1.92 M Glycine, 1.0% SDS and was stored at room temperature. The buffer was used at 1X concentration.

2.1.8 Western blotting buffers

10x TBST buffer

A 10X TBST was comprised of 200mM Tris, pH 7.6, 1.5M NaCl, and 1% Tween® 20 detergent. TBST was used at 1X concentration to wash the blot membrane.

Blocking buffer

The blocking buffer was comprised of 5% BSA and was prepared fresh in 1X TBST.

Ponceau S stain

Ponceau S stain was comprised of 0.1% Ponceau S, Protein dye, prepared in 5% Glacial Acetic Acid.

2.1.9 Commercial screens used in the study

Commercial screens NeXtal JCSG Core I-IV, Crystal Screen HT ™ (HR2-130) and PEG/Ion HT ™ (HR2-139) were purchased from Molecular Dimensions and were used to screen for crystallization conditions using 96-well sitting drop plates. For the optimisation of crystals, the additive screen from Hampton research was used together with the identified crystallization conditions.

2.2 Methods

2.2.1 Cloning of Hsf domains into expression vectors

The bacterial expression plasmid pET 28 c was used in all *E. coli* expression experiments. In total, seven constructs were generated for the study, three of these constructs contained one of the Hsf binding domains (BD1, BD2 and BD3). The other three constructs contained one of the Hsf putative domains (PD1, PD2 and PD3) as mentioned in Table 6. The seventh construct was designed to have the Hsf sequence with both PD3 and BD3 domains along with their joining flexible loop and will be referred as 'PD3BD3' hereafter. The PD3BD3 construct was crystallised to test the Hsf unique hairpin loop structure hypothesis mentioned in the first objective.

Table 6 List of constructs used in this study

Name of construct	Description
BD1	1792-2022 residues of Hsf constitute Binding domain 1
BD2	429-652 residues of Hsf constitute Binding domain 2
BD3	1103-1337 residues of Hsf constitute Binding domain 3
PD1	1579-1549 residues of Hsf constitute Putative domain 1
PD2	272-375 residues of Hsf constitute Putative domain 2
PD3	889-1046 residues of Hsf constitute Putative domain 3
PD3BD3	889-1337 residues of Hsf constitute Binding domain 3 and Putative domain 3

All seven constructs carried Kanamycin resistance for colony selection and an N-terminal 6-His tag for affinity purification. The clones with BD3, PD1, PD2, PD3 and PD3BD3 were

generated previously by Goldman group and were sequenced before use to confirm the correct construct. The clones of Hsf BD1 and BD2 were generated using the IN-fusion cloning method which is a ligation-free and directional molecular cloning method to clone one or multiple DNA fragments in any linearized vector in a single step and is a single-tube reaction. In-Fusion cloning is a simple and efficient method of cloning that allows joining multiple DNA pieces depending upon a 15 or 20 bp homology sequence at their linear ends. The method requires first to design PCR primers that share 15 bp homology with the end sequences of the linearized cloning vector and using them to PCR amplify the insert DNA fragment. The purified PCR product is then combined with the linearized cloning vector in the In-Fusion cloning reaction for 30 min, followed by transformation into desired strain of *E. coli*.

Primer design

For In-Fusion cloning, primer design and the quality of the primers are very important. For a successful In-Fusion cloning, the 5' end of the primers must contain 15 bases that are complementary to the linearized end of the cloning vector and the 3' end must contain bases complementary to the gene of interest(Park, Throop, & LaBaer, 2015). Hence, the set of forward (Fw) and reverse (Rv) primers designed for Hsf BD1 and BD2 PCR amplification contained 15 nucleotides complementary to the insertion sites in vector (red coloured sequence in

Construct	Primer	Sequence of the primer
BD1	Fw primer	GTTTAACTTTAAGAAGGAGATATACCATGAGCTGGACGGCAACT GCTGGTAAAG
	Rv primer	CTCGAGAGCCAAGTGGTGGTGGTGTTCAAAGGTAATCACGCGCG

BD2	Fw primer	GTTTAACTTTAAGAAGGAGATATACCATGAAAGATGGCATTAAA GCAGGTAATAAAG
	Rv primer	CTCGAGATCGCCATGTGGTGGTGGTGCTTTTTCAAGACCGC

Table 7) and region complementary to the genes of interest (GOI) (black coloured sequence in

Table 7) respectively.

Construct	Primer	Sequence of the primer
BD1	Fw primer	GTTTAACTTTAAGAAGGAGATATACCATGAGCTGGACGGCAACT GCTGGTAAAG
	Rv primer	CTCGAGAGCCAAGTGGTGGTGGTGTTCAAAGGTAATCACGCGCG
BD2	Fw primer	GTTTAACTTTAAGAAGGAGATATACCATGAAAGATGGCATTAAA GCAGGTAATAAAG
	Rv primer	CTCGAGATCGCCATGTGGTGGTGGTGCTTTTTCAAGACCGC

After the flanking 15 base sequence the Fw primers contained the start codon to initiate the transcription in the GOIs. The reverse (Rv) primers contained sequence for His-tag (green colored sequence in

Table 7) to be added at the c-terminal of the transcribed protein.

Construct	Primer	Sequence of the primer
BD1	Fw primer	GTTTAACTTTAAGAAGGAGATATACCATGAGCTGGACGGCAACT GCTGGTAAAG
	Rv primer	CTCGAGAGCCAAGTGGTGGTGGTGTTCAAAGGTAATCACGCGCG
BD2	Fw primer	GTTTAACTTTAAGAAGGAGATATACCATGAAAGATGGCATTAAA GCAGGTAATAAAG
	Rv primer	CTCGAGATCGCCATGTGGTGGTGGTGCTTTTTCAAGACCGC

After Primer designing, the GOIs were amplified with PCR chain reaction (Table 9)(Hoseini & Sauer, 2015).

Table 7 List of Primers for BD1 and BD2 PCR amplification

The red sequence of the primers listed above is the flanking region, yellow coloured ATG is

Construct	Primer	Sequence of the primer
BD1	Fw primer	GTTTAACTTTAAGAAGGAGATATACCATGAGCTGGACGGCAACTGCTGGTAAAG
	Rv primer	CTCGAGAGCCAAGTGGTGGTGGTGTTCAAAGGTAATCACGCGCG
BD2	Fw primer	GTTTAACTTTAAGAAGGAGATATACCATGAAAGATGGCATTAAAGCAGGTAATAAAG
	Rv primer	CTCGAGATCGCCATGTGGTGGTGGTGCTTTTTCAAGACCGC

the start codon and green colour sequence is for adding C-terminal His tag

Table 8 Master mix for PCR amplification of BD1 and BD2

Components	Volume
Template Hsf plasmid (40 ng)	1 µl
Forward primer (200 nM)	2 µl
Reverse primer (200 nM)	2 µl
Phusion master mix	50 µl
Autoclaved ultrapure water	45 µl

Table 9 Parameters for PCR optimisation of BD1 and BD2

Steps	Temperature	Time
Initial denaturation	95 ˚C	1 min.
Denaturation	95 ˚C	15 sec.
Annealing	50-62 ˚C	30 sec.
Extension	68 ˚C	45 sec.
Number of cycles		34
Final extension	68 ˚C	10 min.
Volume/tube	10 µl	

Vector digestion

The vector was double digested with restriction enzymes *Nco*I and *Xho*I to be linearized with cohesive ends (Table 10). The digestion of the vector was carried out at 37 ˚C for 3 hours in a 50 µl reaction.

Table 10 Master mix for Vector double digestion

Components	**Volume (50 µl)**
Vector	0.5 µg (12.5 µl)
NcoI-HF	1 µl
XhoI-HF	1 µl
Cut smart buffer	5 µl
Autoclaved ultrapure water	30.5 µl

Agarose gel Electrophoresis

The double digested vector (~ 5.4 kb) and the PCR amplified GOI (BD1, 0.69 kb and BD2, 0.39 kb) were subjected to gel electrophoresis for 30 minutes on 1% and 2% agarose gels, respectively. Electrophoresis was carried out in 1X TAE buffer at 10 V/cm. Commercially available SYBR safe dye from Thermo Scientific, USA was used to visualize the DNA bands under UV radiation using SynGene G:BOX Chemi XX6 system. DNA marker (NEB bioscience) was used to estimate the size of DNA fragments.

PCR extraction and DNA purification from Agarose gel

The PCR products and linearized vector were run on an agarose gel. The bands were observed in UV light and were cut carefully using a sterilised scalpel. The PCR products and the vector were

extracted using Qiagen QIAquick Gel extraction kit following the simple three step 'bind-wash-elute' protocol.

Ligation

The gel purified digested vector (pET 28 c) and GOIs (BD1 and BD2) were subjected to ligation with In-fusion master mix in two separate tubes. The ligation reactions were carried out using 5 ng of pET 28 c and using multiple vector: insert molar ratios- 1:1, 1:3, 1:5, 1:7, at 37 ˚C for 2 hours. The complementary sticky ends in both the vectors and the GOI ensured successful recombination.

Table 11 Components of gene: vector ligation

Reaction component	**Cloning reaction**
Purified PCR fragment	150 ng
Linearized vector	150 ng
5x In-Fusion HD enzyme premix	2 µl
Autoclaved ultrapure water	To 10 µl

Transformation of cloned domains in *E. coli*

Transformation was done for both plasmid propagation and for heterologous expression of proteins. Different *E. coli* strains were used for the two purposes, Oneshot™ OmniMax™ Chemically Competent cells for plasmid isolation and BL21* for the expression standardization. The following protocol was used for the transformation, *E. coli* competent cells were stored in 30 µl aliquots at -80 ˚C and thawed immediately prior to use. When thawed, 1 µl of plasmid DNA (~100 ng/µl) was added to the *E. coli* tube and the mixture was left to incubate on ice for 30 minutes. The transformation mixture was then heat-shocked at 42 ˚C for 45 seconds, before placing back on ice for no longer than 5 minutes. At this point, 200 µl of LB media was added to the tube, before incubating in an orbital shaker at 37 ˚C, 220

rpm, for 1 hour. Samples were then spread onto pre-warmed LB agar plates supplemented with the required antibiotic selection and left to incubate overnight at 37 ˚C. The next day five transformants were picked, inoculated in LB with kan and incubated for 14-16 h at 37 ˚C and 300 rpm. The plasmids were isolated from all the cultures as described below. In order to identify recombinant plasmids, Isolated plasmids were digested with various restriction endonucleases- *Nco*I-HF and *Xho*I-HF and size of released fragments were analysed by agarose gel electrophoresis. Constructs with expected pattern of released fragments were sent for sequencing to verify the sequence of the GOI.

Plasmid Isolation

Plasmid isolation was carried out from 20 mL of overnight growth culture (at 37 ˚C, 220 rpm) of desired transformant. LB media was supplemented with the kan antibiotic. Plasmid isolation was performed using HiYield ™ Plasmid Kit using the following protocol. A 20 mL of logarithmic culture was centrifuged at 4000 X g for 15 min and the supernatant was discarded. To the pellet, 200 µL of PD1 Buffer (RNase A added) was added and the cell pellet was resuspend by vortexing or pipetting. To the solution, 200 µL of PD2 Buffer was added and was mixed gently by inverting the tube 10 times. After 2 min when lysate was clear, 300 µL of PD3 Buffer was added and mixed immediately by inverting the tube 10 times. The mixture was then centrifuged for 2 min at full speed (13,000 rpm) to pellet down cell debris. A clear lysate (supernatant) obtained post centrifugation was applied to the PD Column placed in a 2 mL Collection Tube. The PD column containing the supernatant was centrifuged at full speed (13,000 rpm) for 30 seconds, the flow-through was discarded and the PD column was placed back into the 2 mL Collection Tube. To remove excess salt from the membrane bound DNA molecules, a 400 µl of W1 Buffer was applied to the PD Column, centrifuged at full speed

(13,000 rpm) for 30 seconds, the flow-through was discarded and the PD Column was again placed back to the 2 mL collection tube. The washing was repeated with 600 µl of wash buffer. To remove excess ethanol residue the column matrix was air dried by centrifuging the PD column placed in a 2 mL collection tube at full speed for 3 min. Finally, the dried PD Column was placed in a clean 1.5 mL microcentrifuge tube and plasmid was eluted by applying 100 µl of nuclear free water directly onto the centre of the membrane and then allowing it to stand for 2 min until the nuclease free water is absorbed and centrifuging for 2 min at full speed (13,000 rpm).

2.2.2 Sequence confirmation of all the constructs:

As discussed earlier, constructs with expected pattern of released fragments after digest were sent for sequencing to further verify the sequence of the cloned fragment. Briefly, - sequencing was outsourced to Eurofins Genomics by supplying plasmid samples together with relevant sequencing primers as per the Table 12.

Table 12 Composition for In-Fusion Ligation reaction

Components		**Concentration**	**Volume (µL)**
Plasmid	< 20 kb	100-200 ng/µl	8
	> 20 kb	250-500 ng/µl	
Primers		2-10 pmol/µl	4
Total Volume			12

Table 13 Sequences of Constructs used in this study

BD1
SWTATAGKEGTGEVDPANSAGQEVKAGDKVTFKAGDNLKIKQSGKDFTYSLKKELKDLTSVEFKDANGGTGSESTKITKDGLTITPANGAGAAGANTANTIS VTKDGISAGNKAVTNVVSGLKKFGDGHTLANGTVADFEKHYDNAYKDLTNLDEKGADNNPTVADNTAATVGDLRGLGWVISADKTTGEPNQEYNAQVR NANEVKFKSGNGINVSGKTLNGTRVITFELA

BD2
KDGIKAGNKAITNVASGLRAYDDANFDVLNNSATDLNRHVEDAYKGLLNLNEKNANKQPLVTDSTAATVGDLRKLGWVVSTKNGTKEESNQVKQADEVLF TGAGAATVTSKSENGKHTITVSV
BD3
TISVTKDGISAGNKEITNVKSALKTYKDTQNTAGATQPAANTAEVAKQDLVDLTKPATGAAGNGADAKAPDTTAATVGDLRGLGWVLSAKKTADETQDKEF HAAVKNANEVEFVGKNGATVSAKTDNNGKHTVTID
PD1
AGWNVRGVKPASANNQVENIDFVATYDTVDFVSGDKDTTSVTVESKDNGKRTEVKIGAKTSVIKDHNGKLFTGKELKDANNNGVTVTETDGKDEGNGLVT AKAVIDAVNKAGWRVKTTGANGQNDDFATVASGTNVTFADGNGTTAEVTKANDGSITVKYNVKVADGLKLD
PD2
AGWNIKGAKTAGGNVESVDLVSAYNNVEFITGDKNTLDVVLTAKENGKTTEVKFTPKTSVIKEKDGKLFTGKENNDTNKVTSNTATDNTDEGNGLVTAKAVI DAVNKAGWRVKTTTANGQNGDFATVASGTNVTFESGDGTTASVTKDTNGNGITVKYDAKVGDGLKFD
PD3
AGWNIQGNGNNVDYVATYDTVNFTDDSTGTTTVTVTQKADGKGADVKIGAKTSVIKDHNGKLFTGKDLKDANNGATVSEDDGKDTGTGLVTAKTVIDAV NKSGWRVTGEGATAETGATAVNAGNAETVTSGTSVNFKNGNATTATVSKDNGNINVKYDVNVGDGLKIGDDKKIVADTTTLTVTGGKVSVPAGANSV
PD3BD3
AGWNIQGNGNNVDYVATYDTVNFTDDSTGTTTVTVTQKADGKGADVKIGAKTSVIKDHNGKLFTGKDLKDANNGATVSEDDGKDTGTGLVTAKTVIDAV NKSGWRVTGEGATAETGATAVNAGNAETVTSGTSVNFKNGNATTATVSKDNGNINVKYDVNVGDGLKIGDDKKIVADTTTLTVTGGKVSVPAGANSVNN NKKLVNAEGLATALNNLSWTAKADKYADGESEGETDQEVKAGDKVTFKAGKNLKVKQSEKDFTYSLQDTLTGLTSITLGGTANGRNDTGTVINKDGLTITLA NGAAAGTDASNGNTISVTKDGISAGNKEITNVKSALKTYKDTQNTAGATQPAANTAEVAKQDLVDLTKPATGAAGNGADAKAPDTTAATVGDLRGLGWV LSAKKTADETQDKEFHAAVKNANEVEFVGKNGATVSAKTDNNGKHTVTIDV

2.2.3 Methods of Protein Purification

Expression optimisation for all domains

The goal of expression optimisation is to identify the condition best suited for the heterologous expression of a protein in a soluble fraction. All the constructs used in the present study are listed in Table 3.1. The constructs contain individual domains of Hsf cloned in pET 28 c vector, and have a kanamycin selection marker. The expression optimisation was done for all the Hsf domains according to the protocol mentioned below:

The constructs were transformed into *E. coli* BL21* competent cells and were allowed to grow for 18 hours on agar plates containing appropriate antibiotic selection. A primary inoculum was prepared by inoculating 10 mL of LB media containing either kan or amp antibiotic with

single colony from transformant plates and incubating at 37 ˚C and 200 rpm for the period of 14-16 hr. overnight. Following this 2% of primary cultures were used to inoculate the secondary cultures and were incubated at 37 ˚C and 200 rpm till $O.D._{600}$ reached 0.6. At this point, the cultures were divided and induced with 0.5 mM and 1 mM IPTG and kept at different temperatures (18 ˚C, 25 ˚C and 37 ˚C) for heterologous protein expression. Cultures were harvested at different time points as shown in Figure 14. Cells were then washed with 1x PBS, pH 8.0 and lysed with 30 sec. sonication. Pellets and supernatant were separated by centrifugation at 10,000 g. The results were observed with SDS-PAGE gels. The best induction was achieved with 1 mM IPTG induction at 37 ˚C, 3 hours post induction timepoint.

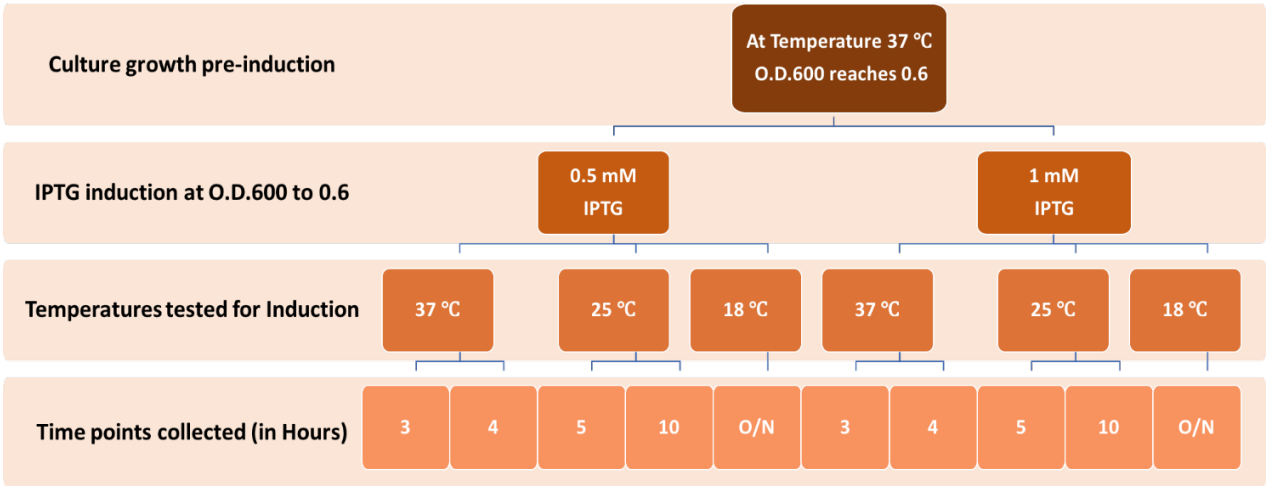


Figure 14 flow chart of expression optimisation timepoints collection

Large-scale purification of Hsf domains

Various Hsf domains were expressed in a large volume (500 mL) as per the condition identified from the expression standardization experiments. They were then harvested and pellets were stored in a deep freezer for later use. At the time of purification experiments, these pellets were allowed to thaw on ice and resuspended in resuspension buffer (50 mM Tris pH 8.0 and 300 mM NaCl for Hsf constructs and 1X phosphate buffer saline (PBS) of pH 8.0 for EibD and

Fc constructs). The lysis of the *E. coli* culture was carefully done by sonication using a cycle of 30 sec on/30 sec off for seven minutes at the 30% amplitude. The suspension was then centrifuged for 20 minutes at 4000 X g for the efficient separation of cell debris. The supernatant containing the protein of interest was transferred to a fresh tube and the debris was discarded.

The supernatant was then poured into a 1 mL Ni-NTA gravity flow column three times to ensure efficient binding. The unwanted impurities and the loosely bound unwanted protein were removed through the step gradient of imidazole. Initial wash steps used 30 mL of 10 mM imidazole followed by 30 mL of 20 mM of imidazole. The elution was then carried out with a 300 mM imidazole and eluates were collected in 0.5 mL aliquots.

Polyacrylamide Gel Electrophoresis (PAGE)

2.2.3.1.1 Denatured Sample Preparation

Prior to loading, 10 µl of sample was routinely mixed with 3 µl of 5X loading dye and made up to a final volume of 15 µl with dialysis buffer. Once mixed, samples were either incubated at 95 ˚C for 5 minutes before loading.

2.2.3.1.2 Sodium-dodecyl Sulphate (SDS)-PAGE

SDS-PAGE was performed using Bio-Rad, U.K. 4-12 % Bis-Tris Plus precast gels (Invitrogen, U.K.). Precast gels were secured in a Bio-Rad, U.K. gel tank and run with 1X Running buffer at 160 V for 50 minutes. 10 µl of the denatured sample was loaded into the wells alongside either 3 µl or 5 µl of PageRuler™ Plus Prestained Protein Ladder (Thermo Scientific, U.K.). Bands were stained with Quick Coomassie (Generon, U.K.) or InstantBlue Protein Stain (Expedeon, U.K.) and imaged with the SynGene G:BOX Chemi XX6 system.

Western Blotting

Western blotting was employed to identify the His-tagged proteins. Purified proteins or cell extracts (10 µg- 40 µg protein/well) were resolved on Bio-Rad precast SDS-PAGE gels and transferred at 20 V for 20 minutes to nitrocellulose membranes (0.45 µm) using a Trans-Blot SD Dry Transfer Cell (Bio-Rad) in presence of 5 mL transfer buffer. After electro-blotting, membranes were stained with Ponceau S to verify the protein transfer. Membranes were washed with Tris buffer saline (TBS) and blocked with 3% Bovine serum albumin (BSA) in TBST (TBS with tween 20) for 1 hour at 24 ˚C or overnight at 4 ˚C. After blocking, the membranes were further incubated for 3 hours at 24 ˚C with HRP-conjugated primary anti-His antibody (ab1187, Abcam, USA) used at a dilution of 1:5,000 diluted in the blocking solution. The membranes were washed four times with TBST for 5 min each and scanned using the SynGene G:BOX Chemi XX6 system.

2.2.4 Structural studies of Hsf PD3-BD3 using X-ray crystallography

Protein sample preparation

Crystallization was done with purified Hsf PD3BD3. The His-Tag affinity purified Hsf fractions were pooled together and were concentrated using the Sigma-Aldrich concentrators with a 3 kDa cut-off limit. The Hsf PD3BD3 were buffer exchanged and concentrated in the concentrators to 8, 10, 12 and 14 mg/mL concentrations. These concentrated protein fractions were collected in microcentrifuge tubes from Eppendorf, Germany. The Hsf PD3BD3 displayed a high propensity to form white precipitates. Therefore, concentrated fractions were centrifuged at 10,000 X g in order to seperate precipitates. The supernatant which contained soluble protein was transferred to a new microcentrifuge tube and immediately used for the crystallization.

Crystallization strategies

The purified and concentrated protein fractions were used for setting up crystallization by the sitting drop vapor diffusion method. Initial crystallization trials were carried out in 96 well plates (Hampton Research, USA) using commercially available crystallization kits mentioned in section 2.1.9. The crystallization screens were set using the Formulatrix NT8 robot. The plates were observed at specific time intervals from 0 hr, 24 hr. for a maximum of one month. After identifying the crystallization conditions from the initial screen, further optimizations were performed by creating hybrid screens in a 96-well plate format using Formulator by Formulatrix, USA and involved varying the concentration, pH and temperature. Some crystallization conditions were expanded in 24 well plate (Hampton Research, USA) using hanging drop method in order to obtain larger crystals .

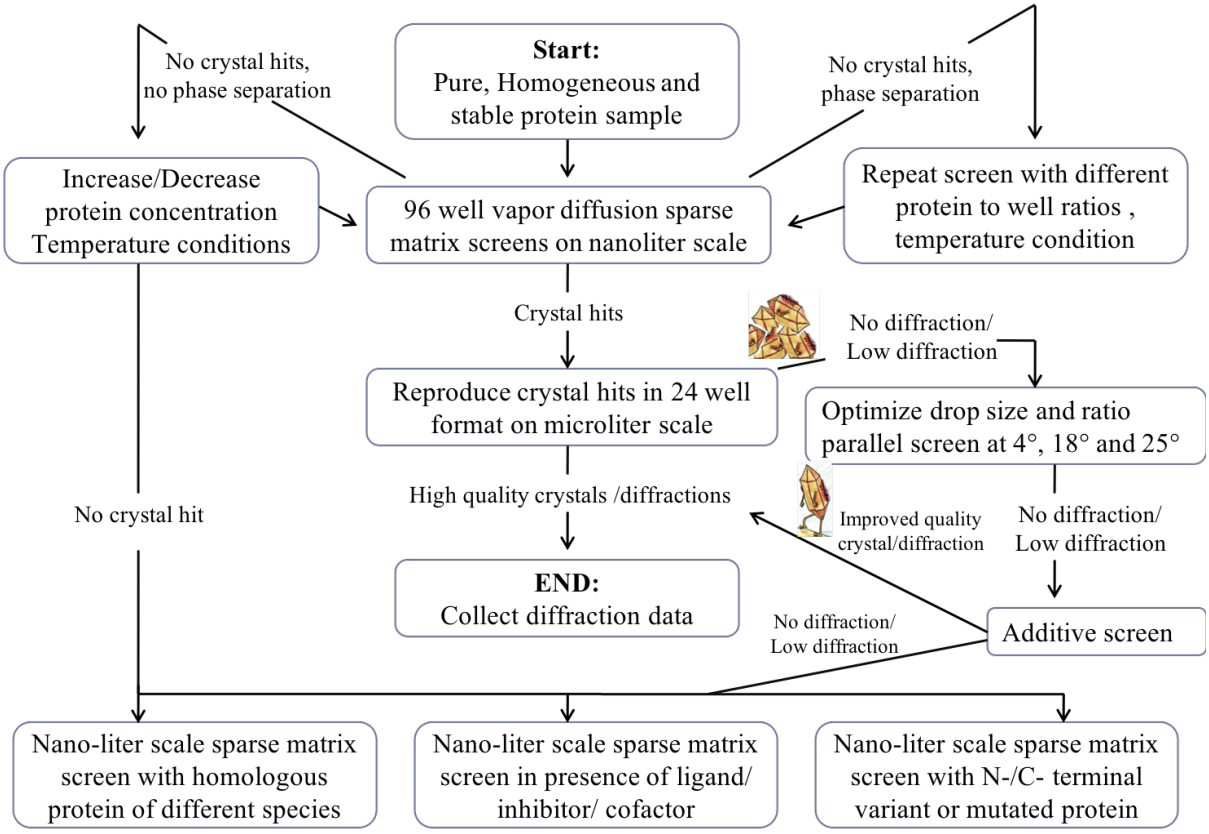


Figure 15 Ideal process for protein crystallization

X-ray Data collection

Prior to data collection, each crystal was mounted on a CryoLoop (Hampton Research, USA) of the appropriate size, soaked in a cryo-protectant solution (mother liquor with 30 % PEG800), and was immediately snap froze in liquid nitrogen before placing in a prechilled SPINE puck In order to collect diffraction data, the Dewar containing liquid nitrogen frozen SPINE pucks was shipped to Diamond Light source (U.K.). All data were collected at 100 K.

2.2.5 Cyclic Voltammetry

Cyclic Voltammetry (CV) is a potentiodynamic electrochemical technique that measures the current that develops in an electrochemical cell under conditions where voltage is in excess of that predicted by the Nernst equation. CV is performed by cycling the potential of a working electrode, and measuring the resulting current(Elgrishi et al., 2018). The CV experiments were designed to create a sandwich of the proteins on the electrodes and was initially conducted on the carbon, cobalt and a binding polymer coated-carbon electrodes. The experiment was conducted in three stages for testing the binding efficiency of Hsf domains on the electrode surface, the binding efficiency of Vitronectin on the electrode surface and testing the interaction of Vitronectin with the Hsf domains.

Hsf: Vn interaction

The electrodes were washed with PBS buffer and each electrode was coated with 10 µl of one of the purified Hsf domains.

The study was intended to test the interaction of Hsf domains with Vitronectin using Cyclic Voltammetry. To achieve this, the surface of electrodes was coated with various Hsf domains- each with three different concentrations; 1.0 ng/µl, 10 ng/µl, and 100 ng/µl. Their binding to different types of electrodes was probed to detect the best working concentration. After

selecting an electrode surface and protein concentration with the optimal response, Vitronectin was deposited on the top of Hsf domains and the response was monitored using Cyclic Voltammetry.

2.2.6 Circular Dichroism (CD) studies

Chiral molecules absorb left and right circularly polarized light to different extents at certain wavelengths. Circular Dichroism is a measure of this difference exerted by chiral centers in a molecule. As proteins have many chiral centers, CD is highly sensitive to the secondary structure of proteins in the far UV region (260-180 nm). Therefore, CD is widely used to study the effect of mutation or change in the protein environment, such as pH, Ionic strength, temperature or the presence of ligands, on its overall structure. CD is measured in a mean residue ellipticity.

CD was used to analyze the secondary structure composition of Hsf-BD1. Far-UV CD spectra for Hsf BD1 was collected on a Jasco-J--810 CD spectropolarimeter flushed with nitrogen at the rate of 9-12 L/min. Data were collected from 260 nm to 200/ 180 nm. Raw ellipticity was converted to mean residue ellipticity according to the equation,

$$[\theta]MRE = \frac{[\theta] \times 100 \times Mr}{C \times 1 \times NA}$$

Where [θ] MRE is the mean residue ellipticity, [θ] is the raw ellipticity in degrees, Mr is the molecular weight of the protein (in Da), C is the concentration of the protein (in mg/mL), 1 is the path length of the cuvette (in cm) and N_A is the number of residues of the protein(Greenfield, 2006).

Chapter 3 Cloning, Expression and Purification of Hsf constructs

3.1 Introduction

This chapter discusses the results of the experiments performed for cloning, expression, purification and crystallization of the truncated sequences from Haemophilus surface fibril (Hsf) of *H. influenzae* Gram-negative bacteria. This chapter is also focused on the expression and purification of complete passenger domain of Hsf. Other proteins discussed in this chapter include *E. coli* immunoglobin binding D (EibD) and truncated Immunoglobulin G-Fc (IgG-Fc) . This chapter will be focused on cloning, purification and crystallization of the above-mentioned proteins.

The purpose behind achieving the purified protein samples of truncated Hsf domains is towards downstream structural and functional characterisation of the Hsf protein. The constructs created from the Hsf protein constitute putative and binding domains of the Hsf full-length protein and are listed in chapter 2, Table 6. Structural studies were carried out on the $Hsf_{888-1342}$ construct here after called as Hsf PD3BD3 domain, via X-ray crystallography. The Hsf PD3BD3 region has a slight bend which may hold importance towards successful protein binding which along with the structure is discussed in chapter 4.

The single-domain constructs (PD1, 2, 3 and BD1, 2, 3) expressed and purified as part of this chapter were used for functional characterisation. The experiments for the functional characterisation are- Cyclic Voltammetry (CV), Impedance spectroscopy and Mass photometry, the results of which are discussed in chapter 5. All of the above-mentioned experiments were conducted to test the difference in adhesive property of the domains. The adhesive properties were studied by testing the binding of the constructs to commercially bought Vitronectin (Vn), which is also known as the biological interacting partner for Hsf(Singh et al., 2014).

This chapter also discusses the optimisation of expression and purification of Hsf_{50-238} construct and Interaction studies of protein *E. coli* immunoglobin binding D (EibD)and IgG-Fc. Hsf_{50-238} contains the long passenger domain of Hsf protein, explained in Chapter 1. The expression optimisationof Hsf $_{50-238}$ was started before lockdown and was later given up due to time limitations.

This chapter is dedicated to the thorough discussion of the results for the cloning of constructs $Hsf_{529-652}$ (Hsf BD1) and $Hsf_{1792-2022}$ (Hsf BD2). The rest of the constructs used in this study were created by former members of the group. All gene sequences relevant to domains in this study were cloned in pET 28 c (+) vectors and transformed in *E. coli*. After cloning, the discussion will be focused towards the expression and purification of all the proteins which includes all the Hsf constructs mentioned above, Hsf_{50-238}, EibD3, EibD8 and IgG-Fc. The protein expression was achieved by IPTG induction and the protein purification by affinity and size-exclusion chromatography techniques for obtaining the purified protein fractions. Ensuring efficient expression and purification of these proteins was vital for downstream applications, such as X-ray crystallography, which required large amounts of highly pure protein.

3.2 Cloning BD1 and BD2 into pET 28 c (+) Vector for Expression in E. coli

In order to clone BD1 and BD2 into pET 28 c (+) vectors, the Hsf sequence and the multiple cloning site of pET 28 c (+) vector were studied to identify appropriate cloning sites. Sites chosen for the cloning were Nco 1 and Xho1. Cultures of Hsf in pBPT vector, and pET 28 c (+) vector in *E. coli* were miniprepped to isolate the plasmids. Quantification for successful plasmid isolation results was done by Nanodrop™.

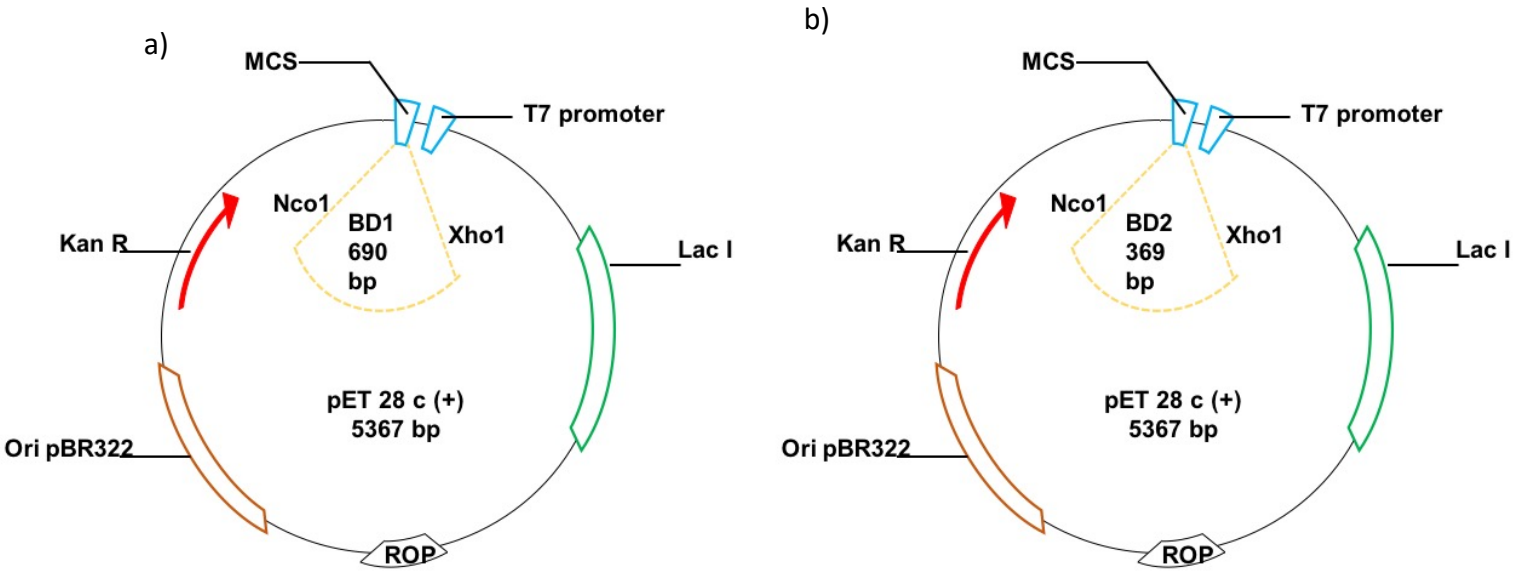


Figure 16 Diagram of pET 28 c (+) vector with BD1 and BD2 inserts

The diagram shows that the Genes of interest are inserted in the linearised vector created by digestion at Xho1 and Nco1 restriction digestion sites to generate cohesive ends. a) PCR amplified BD1 domain of Hsf and b) PCR amplified BD2 domain of Hsf are ligated in the MCS region of the vector between Xho I and Nco I restriction digestion sites.

3.2.1 Modification of the Ends of BD1 and BD2 for in-fusion ligation

In-fusion cloning works on the principle of homologous recombination, for which, first the vector is linearised by digesting at unique restriction sites, leaving cohesive ends. Then both of the primers were substituted with the reverse complements of cohesive ends of the

vectors and the complementary sequences to the insert. The PCR reaction with these primers generates a modified product with the flanking regions. This insert is mixed with the vector and In-fusion master mix for reaction to take place. The In-fusion master mix degrades the complementary ends of the flanking regions added to the insert, leaving the cohesive ends to bind to the vector and thus generating the desired hybrid clone. The 5' and 3' ends of the BD1 and BD2 sequences were then modified with primers (Table 2; chapter 2). The major modifications involved addition of 20 base flanking regions on 5' and 3' ends complementary to the vector required for In-fusion cloning so that homologous recombination can take place on ligation. In addition to the flanking bases, His6 tag at C-terminal and start codon at N-terminal were also added to the final construct. The reactions were first conducted at different temperatures ranging from 52 ˚C to 65 ˚C for optimising the reaction conditions. The reactions were carried out as described in table 2.6 of chapter 2 and the PCR products were analysed with DNA agarose gel electrophoresis (Fig 2). The sizes of the desired bands were 690 bp and 369 bp for BD1 and BD2 genes, respectively. This was analysed in comparison with NEB DNA marker on 2% agarose gel as shown in Figure 17. Due to the presence of consensus sequence repeats in the Hsf, primers ligated at multiple sites resulting in generation of non-specific PCR products. Because of this reason, multiple bands were observed in the gel lanes

with PCR products. The temperatures that resulted in most successful PCR products were at 60 ˚C for BD1 and 62 ˚C for BD2 due to fewest number of impurities.

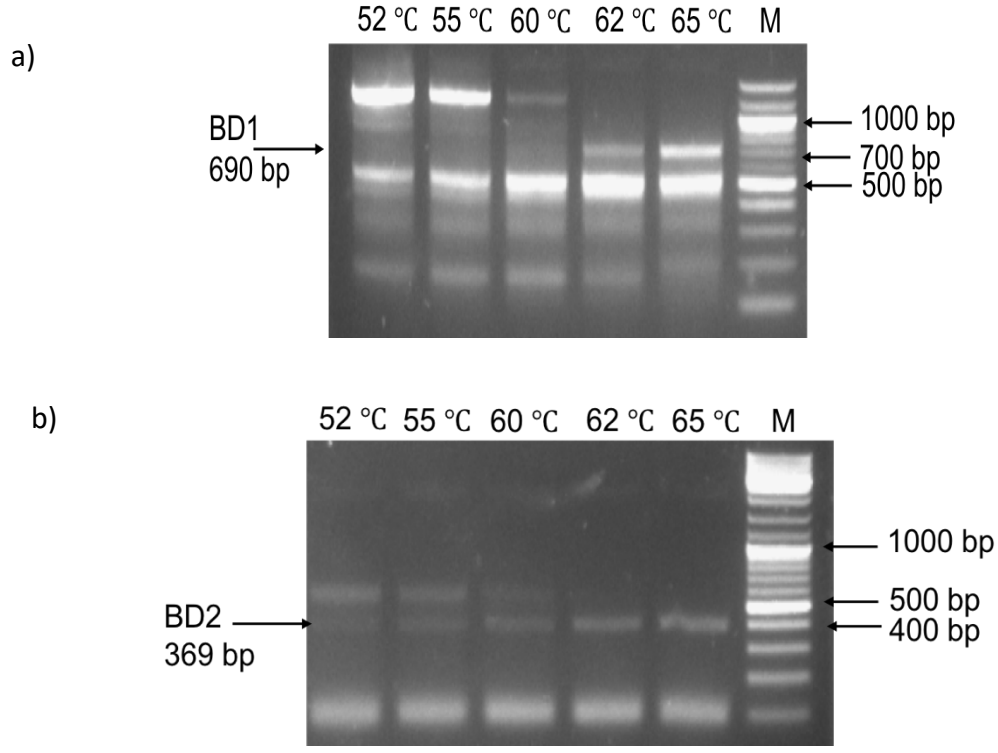


Figure 17 Agarose gel images for PCR standardization

a) PCR standardization of BD1 gene showing the correct band at 690 bp, b) PCR standardization of BD2 gene of size 369 bp. The impurities in both of the

3.2.2 Restriction Digests of Vector

The vector pET 28 c (+) plasmid was purified from Oneshot™ OmniMax™ competent cells as explained in chapter 2. The vector was double digested with *Xho* I and *Nco* I restriction digestion enzymes to generate cohesive ends. The linearization of the vector was observed on 1% agarose and the size of the vector was compared against the NEB DNA Marker (fig 3).

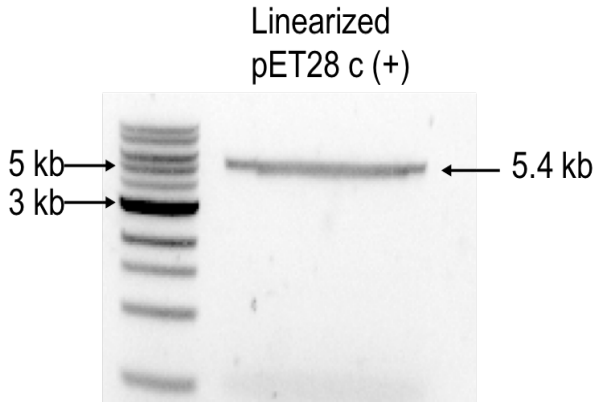


Figure 18 Double digestion of pET 28 c(+) vector with Xho I and Nco I

3.2.3 Cloning Modified BD1 and BD2 into pET 28 c (+) vector

The PCR products obtained for constructs BD1 and BD2 and the double digested pET 28 c (+) vector were gel purified as per the protocol mentioned in Chapter 2. The gel purified genes were sub-cloned into pET 28 c (+) vector with In-fusion ligation master mix and transformed into Oneshot™ OmniMax™ competent cells as described in the material and methods section. Single colonies were selected from LB Kan (100 mg/L) plates and cultured in 3 mL LB Kan (100 mg/L) media. The plasmid was purified from the *E. coli* culture grown for 16 hrs and double digested with *Xho* I and *Nco* I restriction enzymes for confirmation of successful ligation. The digested samples were run on 1.0 % agarose gel (Figure 19). The digested

colonies had a band that corresponded to the size of pET 28 c (+) vector (5367 bp) and another band that corresponded to the size of BD1 (690 bp) and BD2 (369 bp) with addition of 55 base pair including the flanking sequences added by the primers at N- and C- terminal, His6 tag and ATG sequence. This sums up to the sizes BD1 (745 bp) and BD2 (424 bp) shown in the respective lanes in Figure 19 a and b respectively. Aliquots of the plasmid DNA for all 3 constructs were sent for sequencing. Results confirmed the presence of BD1 and BD2 insert in pET 28 c (+) vector in the respective samples. The confirmed clones were further transformed into BL21* competent cells for protein overexpression.

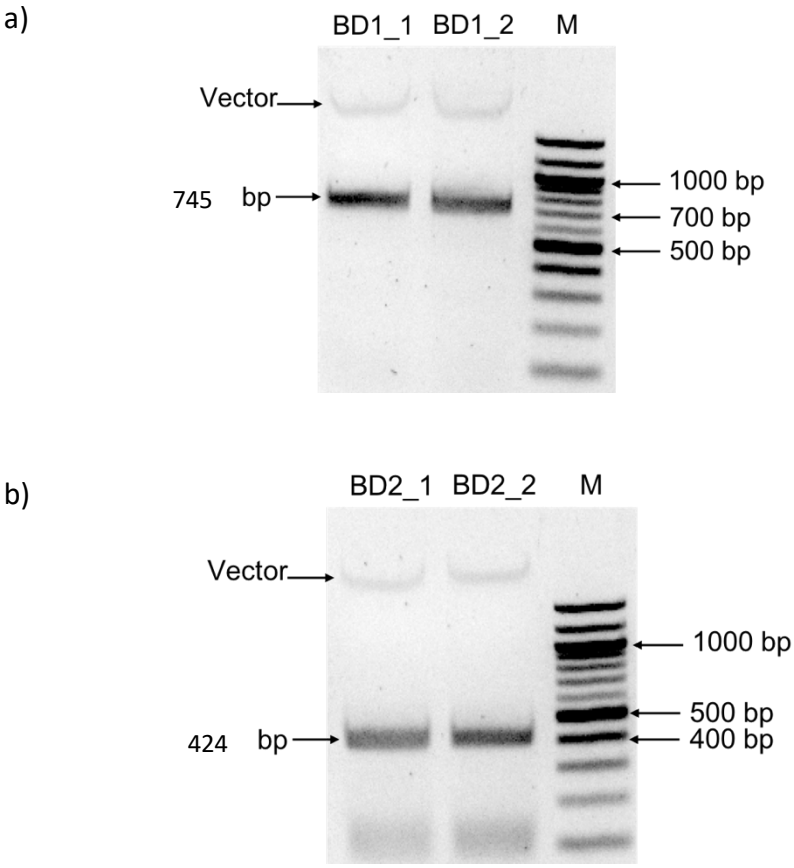


Figure 19 Clone confirmation by Xho I and Nco I double digestion for a) BD1 and b) BD2

3.3 Expression and Purification of Proteins

3.3.1 Purification of Hsf constructs with single domains

The expression and purification of Hsf single domain constructs named PD1, PD2, PD3 for putative domains and BD1, BD2 and BD3 for binding domains were optimised at different temperatures for different time points to improve the yield and purity of the samples, explained in detail in section 2.2.3. The sequences of all the constructs are listed in Table 13. The results for expression were analysed with SDS-PAGE. The expression of BD1 domain was successfully achieved at 0.5 mM IPTG at 37 ˚C and hence 1 L culture pellet was purified by His-tag Ni-NTA affinity chromatography and purified protein fractions were eluted with 300 mM imidazole in 50 mM Tris pH7.5 and 300 mM NaCl buffer. The purity of the eluted fraction was analysed by SDS_PAGE Mini Protean TGX pre-casted gels by Bio-Rad (Figure 20a).

For BD2 domain, successful expression was achieved at 1 mM IPTG 37 ˚C, at 2-hour time point, where the protein was not into the soluble fraction as shown in Figure 20b. At 4-hour time point, less than half of the protein fraction was observed in the supernatant fraction (Figure 20b. The maximum protein in the supernatant fraction was retrieved at 6-hour time point, which was approximately half of the total expressed protein (Figure 20b). The affinity purified BD2 fractions shown in Figure 21, which were used for further experiments. The successful expression of BD3 domain was also optimised with 1mM IPTG induction at 37 ˚C like BD2 domain and the cultures were harvested 4 hours after induction, like other Hsf domains the elutes of purified BD3 were collected with 300 mM Imidazole in 50 mM Tris pH7.5 and 300 mM NaCl buffer. The results of BD3 purification were analysed by SDS-PAGE gel image showing the purified BD3 fractions exist in its multimeric confirmations in the solution (Figure 20c).

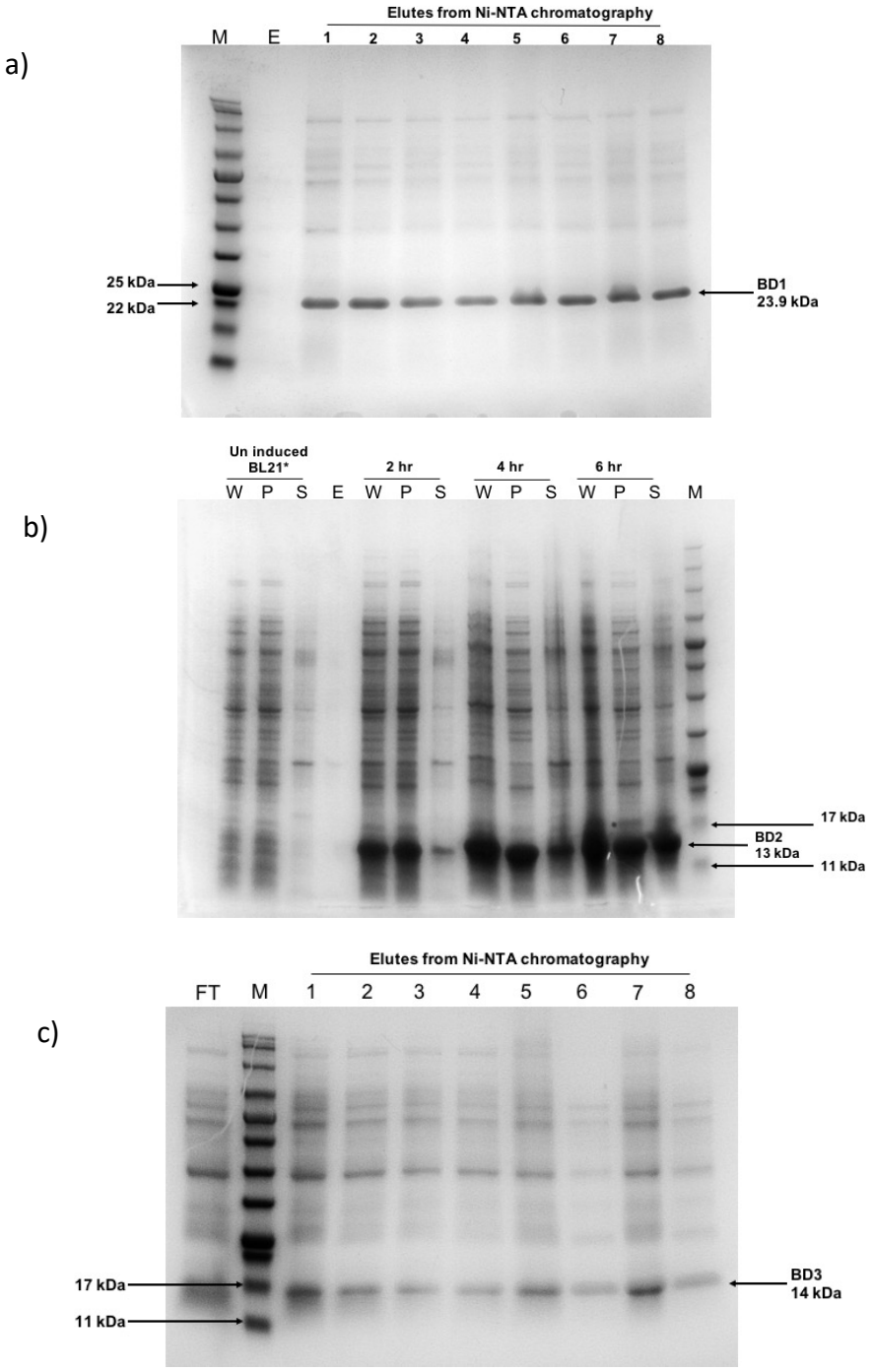


Figure 20 SDS_PAGE gel image of Binding domains

a) Expression optimisation of BD1 showing that the maximum protein in the supernatant fraction was achieved at 6 hr time point. The expected band size was 23.9 kDa. b) The purification profile of BD2 with band expected at 13 kDa showing that the protein exists in multimeric conformations in the solution. c) The purification profile of BD3 with band expected at 14 kDa also shows the multimeric conformations of the protein

The PD1, PD2 and PD3 domains of Hsf were induced and purified using the same protocols. The expression of the mentioned 3 domains was achieved by inducing the *E. coli* cultures with 1 mM IPTG at 37 ˚C and the cultures were harvested 4 hours post-induction. The purification was again done by Ni-NTA affinity chromatography and purified protein fractions were eluted with 300 mM imidazole in 50 mM Tris pH7.5 and 300 mM NaCl buffer. The purity of the proteins were analysed with SDS-PAGE (Figure 21).

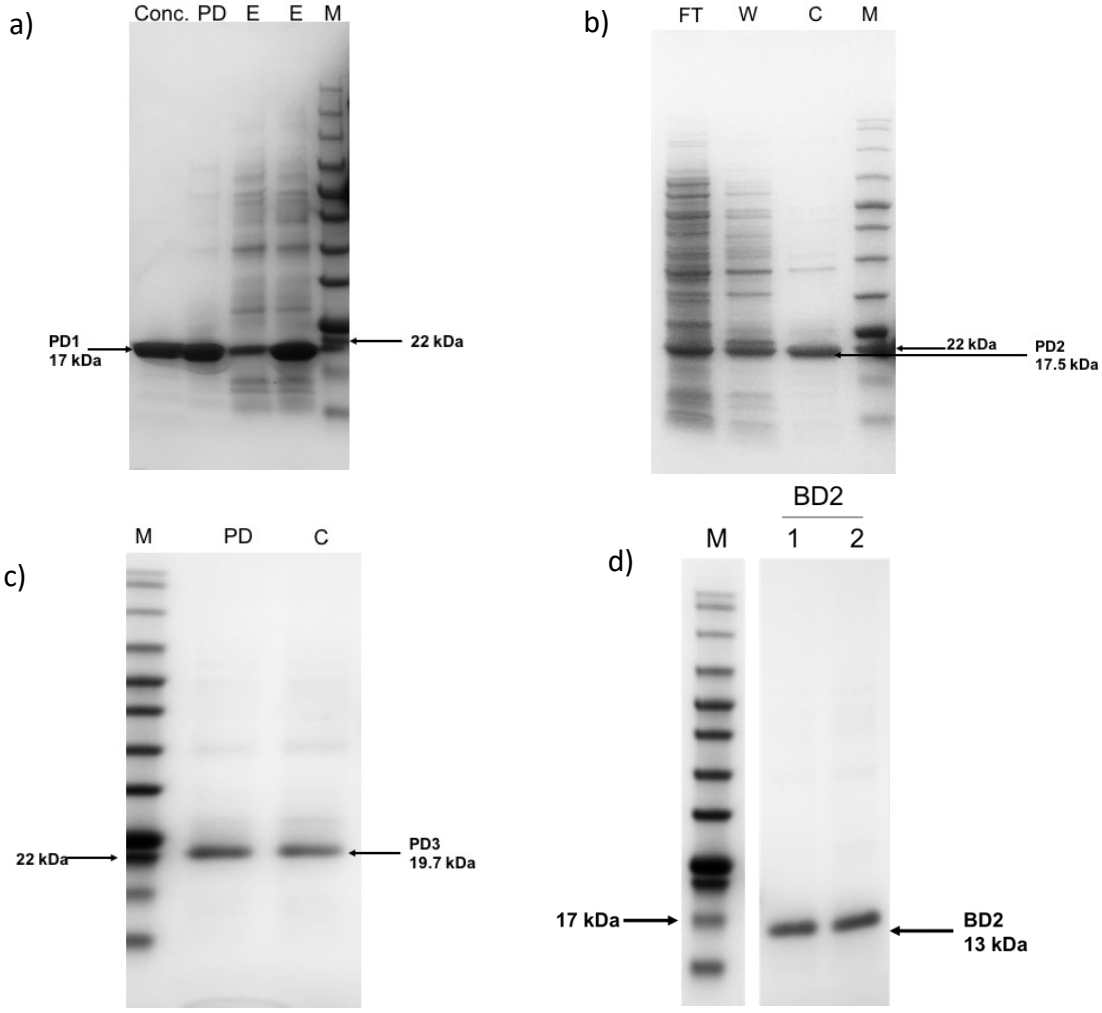


Figure 21 SDS-PAGE gel image for PD1, PD2, PD3 and BD2 purification

a) PD1 Purification, where lane 1 from left is the concentrated purified fraction, PD stands for post dialysis sample and E are the protein elutes from affinity purification. b) Purification image of PD2 with Flow through (FT) fraction in lane 1 from left wash (W) with 30 mM imidazole fraction in 2nd lane followed by the final post dialysis concentrated protein fraction. c) Purification gel image of PD3 domain with Marker in lane 1 from left followed by Post dialysis (PD) and concentrated (C) final protein sample. d) Purification gel image of BD2 domain with Marker corresponding to the post dialysed and concentrated protein fractions.

3.3.2 Purification of Hsf PD3BD3

The Hsf PD3BD3 domains were expressed and purified as described in materials and methods, section2.2.3. In brief, the post cell lysis supernatant fraction was incubated with 500 µL Ni-NTA agarose. The protein was allowed to bind to the resin for 1 hour by passing the supernatant fraction through the gravity flow Ni-NTA column. The flow through was collected after passing the supernatant fraction three times through the column. The protein bound Ni-NTA resin was washed with 10 column volumes of wash buffer I and wash buffer II, containing 10 mM and 30 mM imidazole respectively (Table 5), to remove nonspecific proteins. This was followed by protein elution with elution buffer containing 300 mM Imidazole (section 2.2.3). The purification profile was analysed by SDS-PAGE (Figure 22).

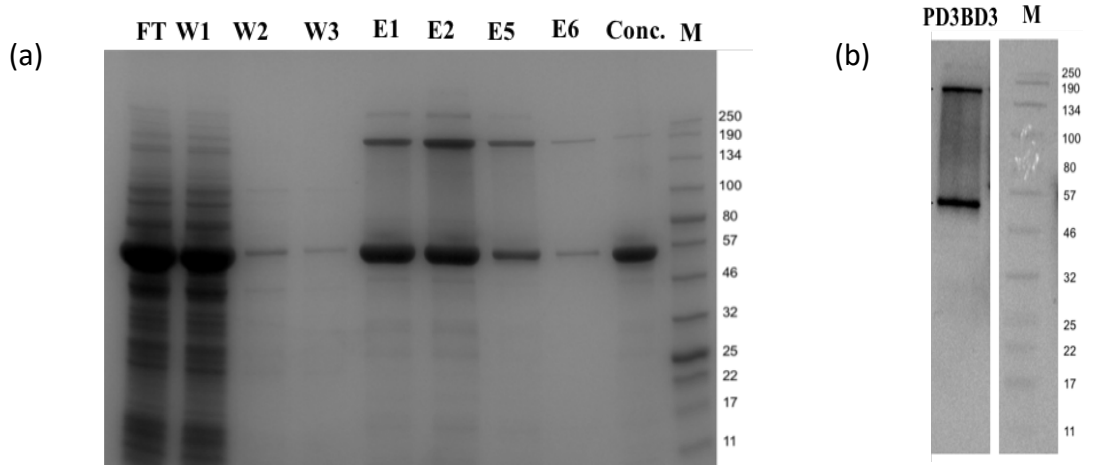


Figure 22 SDS-PAGE analysis of PD3BD3 Purification.

(a) SDS-PAGE analysis of PD3BD3 purification, the expected Mwt of Protein is 50 kDa. (b) Western blot analysis of final purified PD3BD3, further used for crystallization.

A significant proportion of expressed Hsf PD3BD3 was lost in the flow through and the loss of protein during the wash steps was also observed in other constructs as well. I tried to recover these losses by adding 5 mM and 10 mM Imidazole in the protein binding of proteins for 2 hours with Ni-NTA. This was done by adding the post cell lysis supernatant and Ni-NTA in a 50

ml Falcon centrifuge tube. No difference was observed in the two binding protocols . Considering the sufficient quantity of purified protein was obtained in the final fraction with 2 L culture pellet and the protein loss was non-recoverable, more efforts were not directed towards recovering the protein losses.

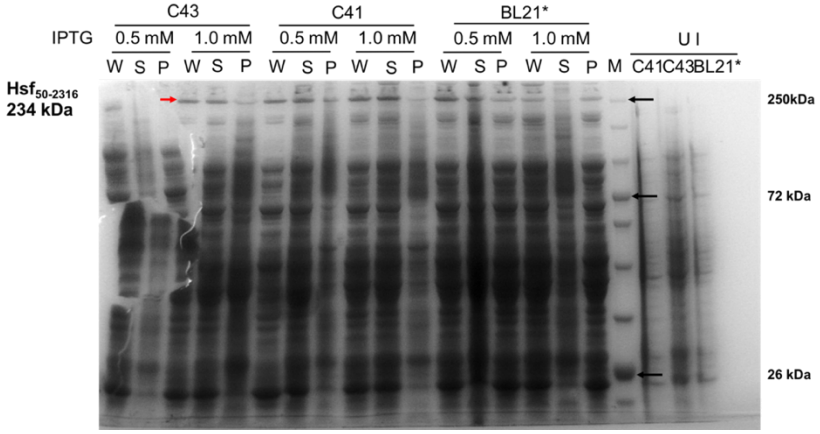
3.3.3 Expression optimisation and Purification of $Hsf_{50-2316}$

The plasmid expressing recombinant $Hsf_{50-2316}$ (~ 145 kDa), comprised of $Hsf_{50-2316}$ along with a C terminal His-tag in pBPT vector, was transformed into *E. coli* strains: BL21*, C41 and C43 competent cells, followed by expression studies. The induction experiments for $Hsf_{50-2316}$ were performed at 37 ^{o}C for 6 hrs and 18 ^{o}C for 16 hrs. This was done to assess the stability and yield of the protein in two different conditions with the lower temperature often being successful for reducing the stress on the bacteria. The cultures were induced with two IPTG concentrations, 0.5 mM and 1.0 mM to assess how well the bacteria grow and express under different stresses of IPTG induction with a lower concentration to reduce the expression level and stress (Figure 14). The protein expression was observed at 37 ^{o}C at 6 hrs timepoint in all three cell lines at both IPTG concentrations. The expression profile at different conditions and time points were analysed by SDS-PAGE (Figure 23a). Protein in the soluble fraction was observed in C41 expression strain thus, for large scale purification, induction of $Hsf_{50-2316}$ was done in C41 competent cells at 37 ˚C with 1.0 mM IPTG induction.

To increase the protein expression, further optimisation was conducted with C41 competent cells induced with 1 mM IPTG and 10 mL samples were collected at 2 hrs, 4 hrs, and 8 hrs and samples were prepared according to the protocol described in chapter 2. The 10 µL samples were run on an SDS-PAGE gel, but SDS_PAGE analysis showed no significant increase in protein production.

The protein bands at the right size were confirmed by Western blot experiment for which the analysis showed multiple bands suggesting the degradation of the protein. Hence, for better protein stability, cell lysis was done by homogenization rather than sonication which can produce local heating and subsequent degradation of the protein. Due to the minimum volume limitations, 5 g culture pellet of $Hsf_{50\text{-}2316}$ resuspended in 50 mL lysis buffer was used for the experiment. The samples were loaded on SDS-PAGE gel where a faint protein band at the right size was observed in whole cell lysate and pellet fractions and no protein bands were visualised in any of lanes with samples from the purification protocol. This required further optimisation of expression at large scale.

a)



b)

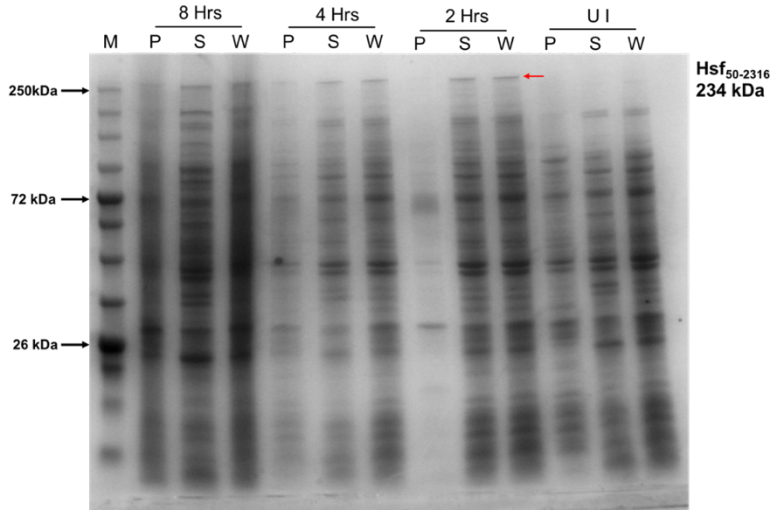


Figure 23 Expression optimisation of $Hsf_{50\text{-}2316}$

a) Results of expression optimisation at 37 ºC at 6 hrs timepoint induced with 0.5 mM and 1.0 mM IPTG in C41, C43 and BL21* expression cell lines of *E. coli*, the band of interest is 250 kDa and marked in red. b) Further optimisation of $Hsf_{50\text{-}2316}$ in C41 cells induced with 1.0 mM IPTG.

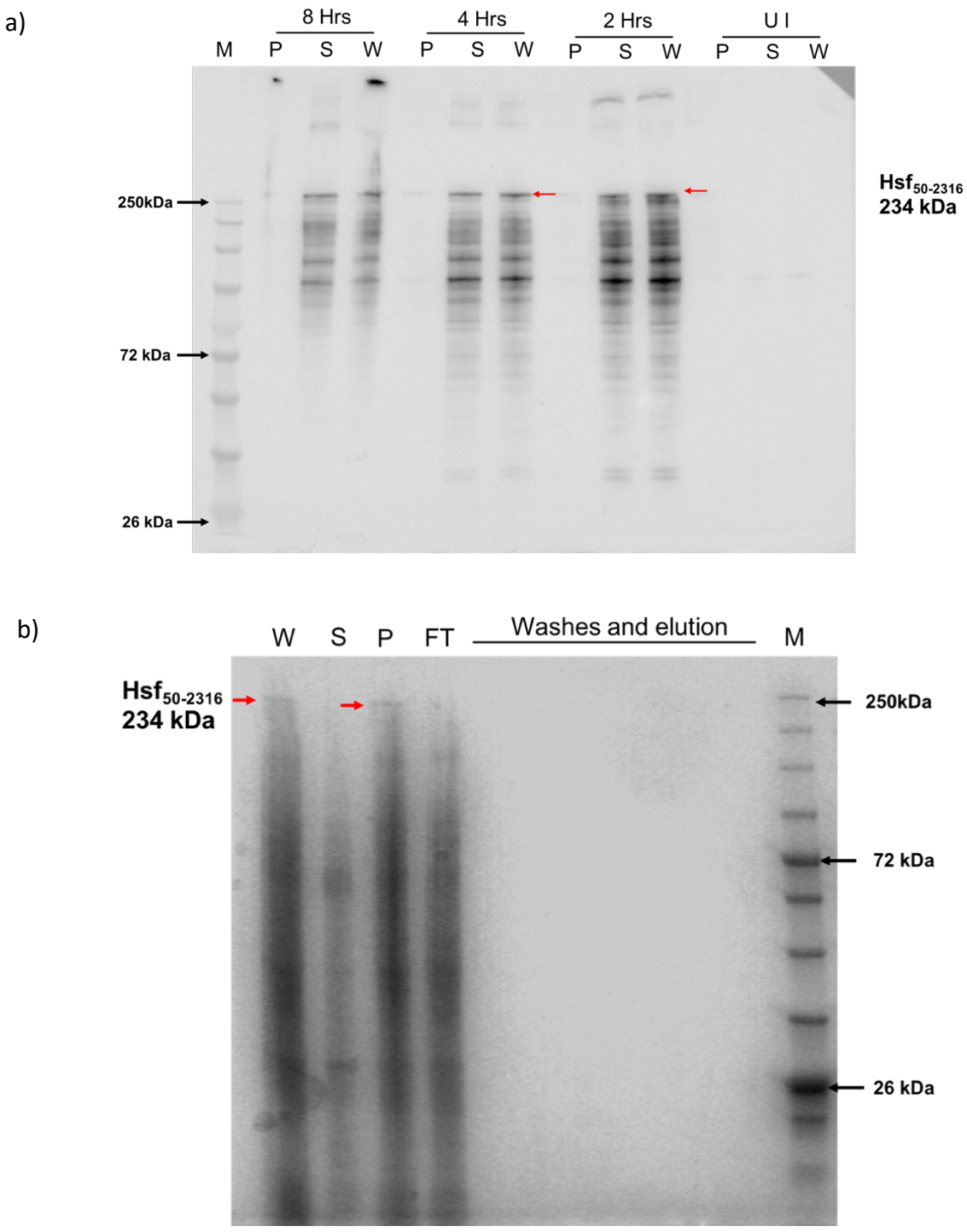


Figure 24 $Hsf_{50\text{-}2316}$ Expression analysis

a) Western blot analysis for expression analysis in C41 *E. coli* expression cell lines. b) His-tag affinity purification trial for Hsf 50-2316 showing bands only in whole cell lysate (W) and pellet (P) fractions.

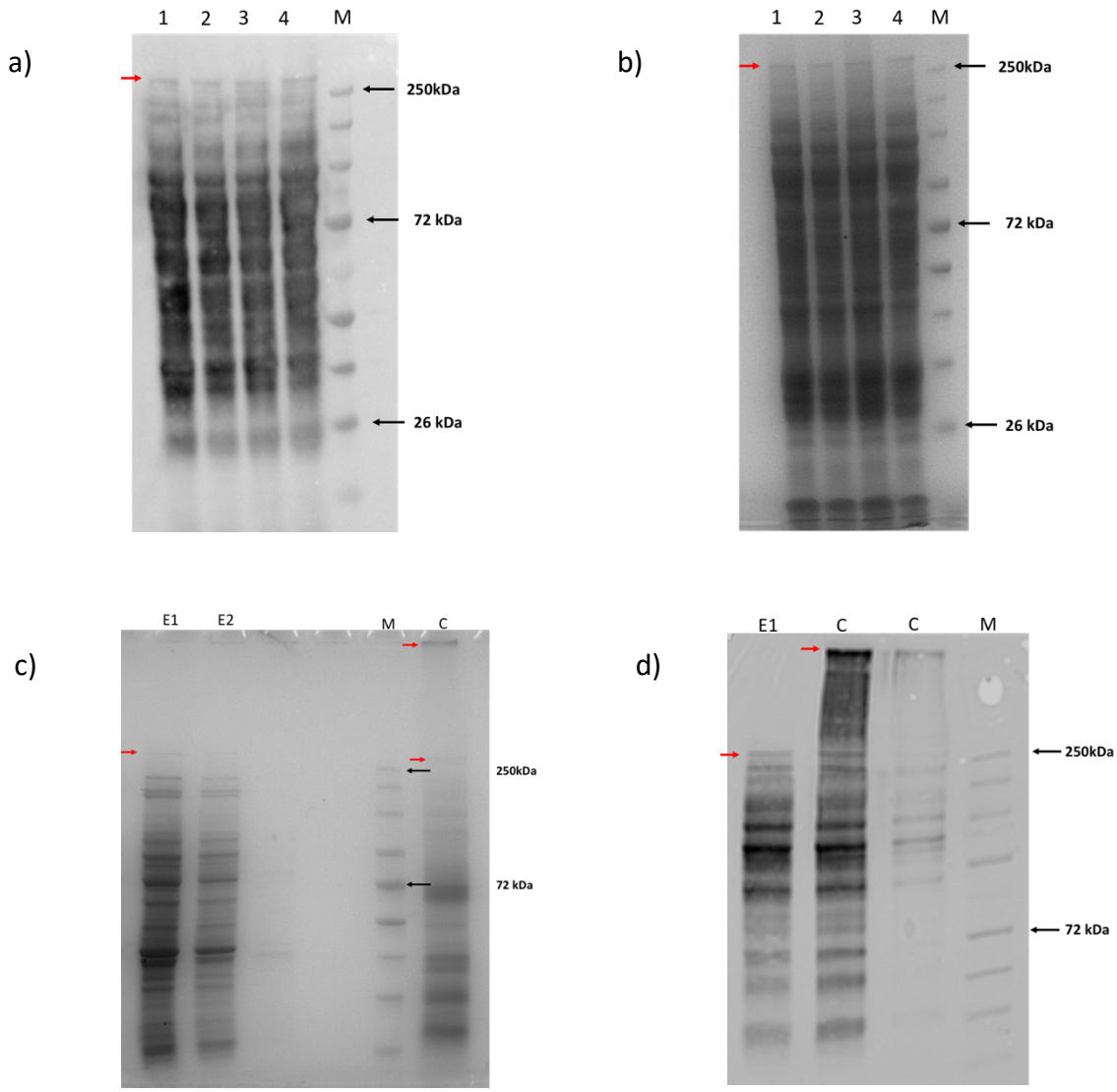


Figure 25 Purification analysis of $Hsf_{50\text{-}2316}$

a) and b) Western blot and SDS-PAGE analysis of $Hsf_{50\text{-}2316}$ purification where 1,2,3,4 are Flow through, wash with 10 mM, 20 mM and 30 mM imidazole in 50 mM Tris buffers pH 7.5 containing 300 mM NaCl respectively and M is marker. The bands corresponding to the size of $Hsf_{50\text{-}2316}$ are pointed by red arrows. c) SDS-PAGE image of gel containing two protein eluates E1 and E2, eluted with 300 mM Imidazole in 50 mM Tris buffers pH 7.5 containing 300 mM NaCl and C is the concentrated sample. d) Western blot image showing the comparison between the protein present in Eluate in the first lane (E1) and concentrated protein sample in the second lane (C).

3.3.4 Purification of E. coli immunoglobulin binding protein D (EibD)

Purification of hybrid EibD constructs was performed to investigate the interaction between EibD and IgG-Fc. According to the EibD: Fc complex model, three Fc monomers can bind to EibD at the same time (Leo & Goldman, 2009). The interaction of EibD with IgG can result in bacterial serum resistance. This was done to determine the role of TAAs in giving serum resistance to pathogens within the host, as well as assisting autoagglutination, a well-known function of TAA proteins. This study was carried out concurrently with the purification of Hsf domains. The results of the purification of constructs used in this study are discussed below; however, because of COVID-19 lockdowns, time constraints, and workplace restrictions, this project was discontinued.

Two hybrid constructs EibD3 (43 kDa) and EibD8 (26.5 kDa) were created by Jack C. Leo, a former member of the group. The two hybrid constructs were created to contain the left handed coiled coil (LHCC), the saddle and the right handed coiled coil (RHCC) (EibD3) and just the RHCC (EibD8) according to the X-ray structure of EibD and sites of IgG binding, shown in literature(Leo et al., 2011). EibD3 and EibD8 constructs were expressed in BL21* *E. coli* strain and the protein purification of was conducted in two steps: His-tag affinity chromatography followed by size exclusion chromatography, results of which are analysed by SDS-PAGE. For EibD3 the results for affinity purification showed that the final protein fraction was pure when eluted in buffer containing with 300 mM imidazole in 50 mM Tris pH7.5 and 300 mM NaCl. EibD3 showed degradation in its post dialysed concentrated protein fraction. For EibD8, the SDS-PAGE gel analysis of affinity purification showed a few impurities in the final elution step, which became more prominent in post dialysis concentrated sample, hence, the size exclusion chromatography was required for EibD8. Another purpose for applying SEC purification method on EibD3 and EibD8 was to confirm that both proteins run at their

trimeric sizes and not monomeric sizes. Hence, a Superdex 200 10/300 GL column was used and the samples were eluted at a flowrate of 0.3 mL/min. On comparison with the standard graph provided by Merck along with the column it was confirmed that both of the proteins were eluted at the expected volume corresponding to their trimeric sizes. However, EibD8 was best purified and showed no degradation when analysed on SDS-PAGE gels, whereas EibD3 was unstable and was heavily degraded.

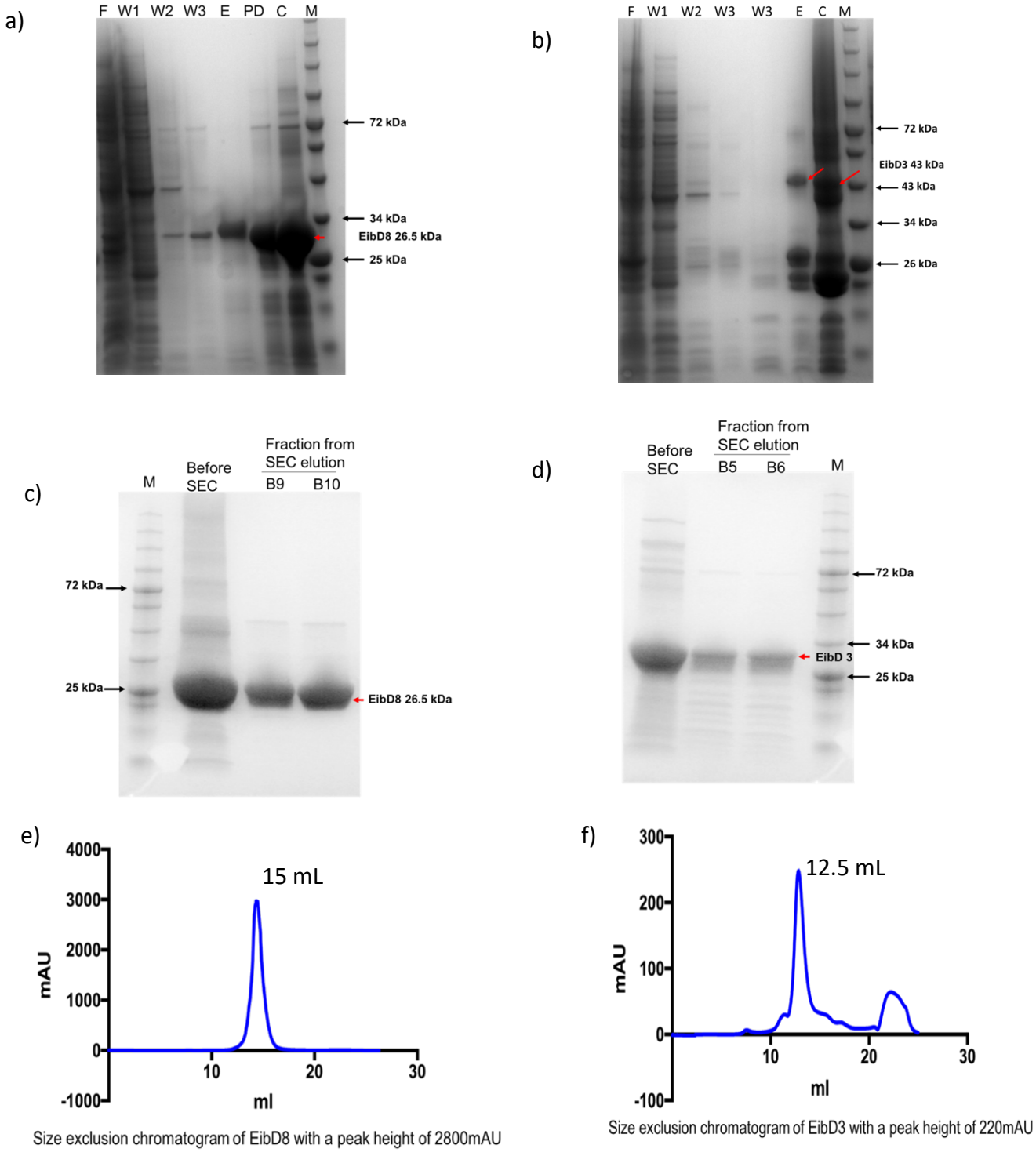


Figure 26 Purification results of EibD8 and EibD3

a) SDS-PAGE gel image of EibD 8 purification b) SDS-PAGE gel image of EibD 3 purification c) SDS-PAGE gel image of EibD 8 protein fractions from Sec purification d) SDS-PAGE gel image of EibD 3 protein fractions from Sec purification e) SEC graph of EibD8 purification depicting the protein elution peak at 15 mL f) SEC graph of EibD3 purification depicting the protein elution peak at 12.5 mL. The additional peak eluting after 20 mL volume is Imidazole.

3.3.5 Purification of Immunoglobulin G-Fc region (IgG-Fc)

EibD binds to the Fc region of IgG antibody, so a truncated IgG-FC molecule was used for the interaction studies. Many complications were faced with these interaction studies due to the fact that both of the monomeric units of the dimeric IgG-Fc molecule have binding sites for the trimeric EibD and all three monomers of the EibD homo-trimeric protein possess binding site for IgG-Fc. Hence, two EibD molecules can bind to three IgG-Fc molecules and those three IgG-Fc can bind to another two EibD and this complex continues to form leading to aggregation of all the protein molecules.

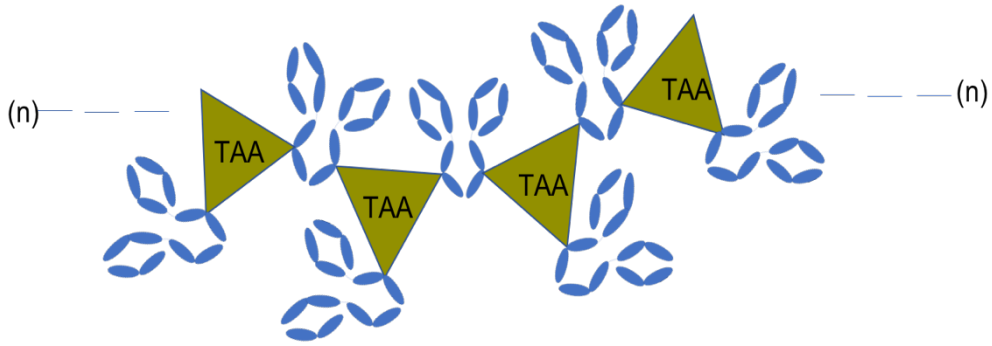


Figure 27 Schematic diagram for IgG-Fc: EibD complex formation.

The IgG antibodies are represented in blue and EibD (TAA) are represented as green coloured triangles binding to the Fc portion of IgG.

To overcome this problem a hybrid IgG-Fc molecule was created with mutation in one of the two monomers of the dimeric IgG-Fc, leaving only one site for EibD binding. Another hurdle that was faced with the IgG-Fc dimer was the dissociation and re-association of the dimer in solution, which means that the mutated IgG-Fc resides in 4 different conformations in the solution that are: two hybrid confirmations with original and mutated monomers, both mutated monomers in one dimer and the original conformation with both non-mutated strands composing the dimer. Hence, to distinguish the mutated strand from the original strand, the mutated strand was tagged with Strep-tag and the original strand was tagged with

His-tag. In order to purify the right constructs with one original and one mutated strand in a dimer, a two-step purification process was done.

The culture of Bl21* competent cells containing mutated IgG-Fc was induced at 25 ^{o}C with 1 mM IPTG for 6 hrs and was purified first by His-tag affinity purification following the protocol mentioned in section 2.2.3. The protein eluted from His-tag purification was then subjected to strep-tag purification. The results of the two-step purification was first analysed by SDS_PAGE, which shows that the IgG-Fc protein was pure and suitable for interaction studies. The confirmation of hybrid dimer was achieved by two-step western blotting, where the samples from both the purifications were detected by anti-His antibody followed by anti-strep antibody. The confirmation of the dimer was achieved by size exclusion chromatography.

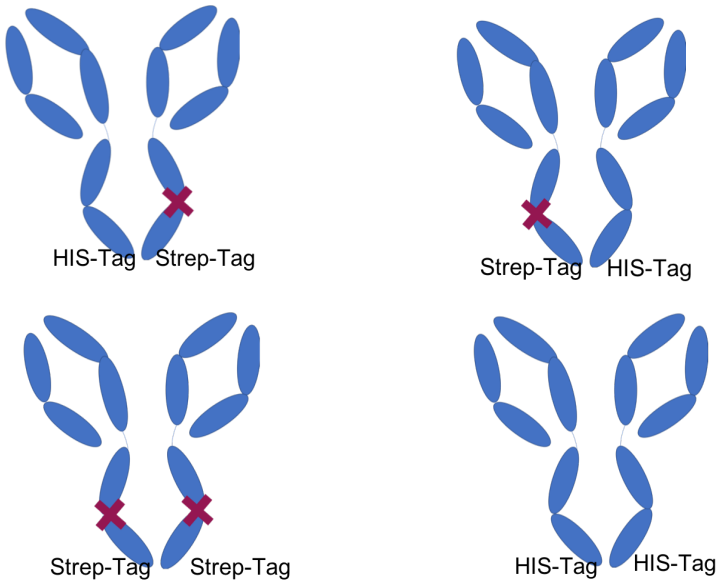
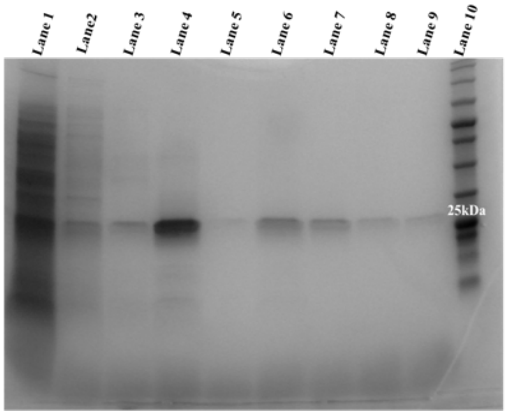


Figure 28 Figure representing the four conformations of IgG in solution

The blue image depicts two strands of the IgG antibody labelled with the either His or Strep tags mentioned at the foot of the strand. The red cross suggests the mutation in the strand with Strep tag.

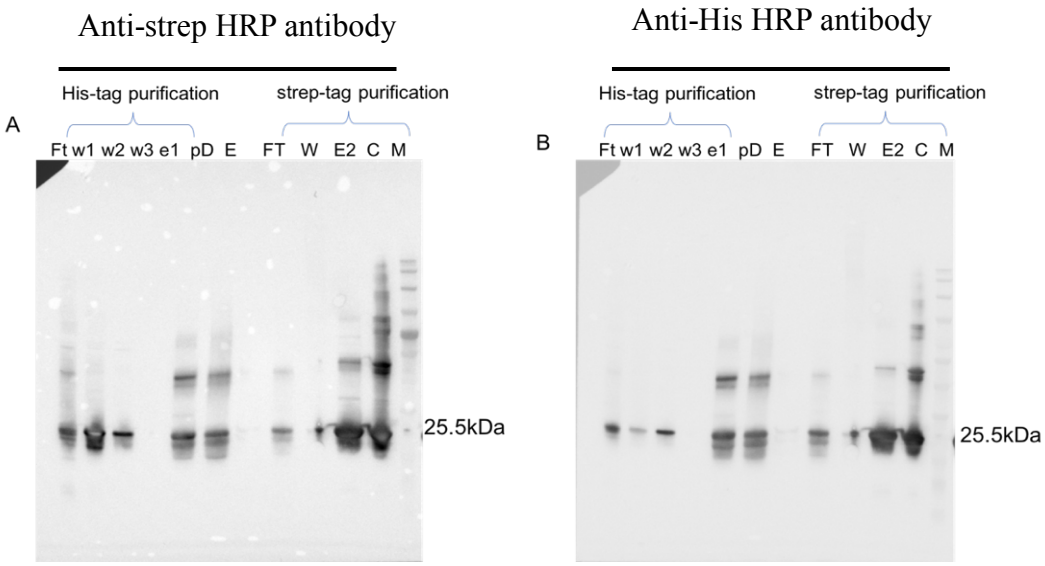
1)



SDS-PAGE analysis of Fc-Fcmutant

His-tag purification samples
Lane1: Flow through
Lane2: Wash with 10mM imidazole
Lane3:Wash with 20mM imidazole
Lane4: Elution with 300mM imidazole
Lane5: Empty
Strep-tag purification samples
Lane6: Flow through
Lane7: Wash with PBS
Lane8: Elution 1 with 2.5mM des-thio-biotin
Lane9: Elution 2 with 2.5mM des-thio-biotin
Lane 10: Marker

2)



3)

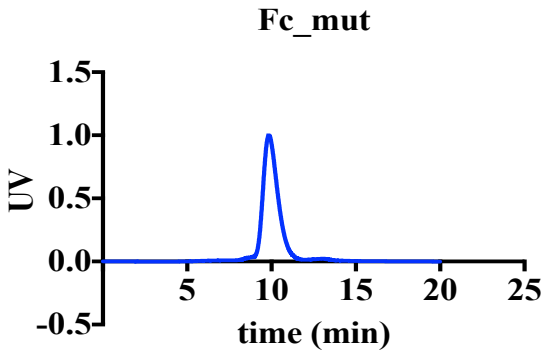


Figure 29 IgG-Fc purification analysis

1) SDS-PAGE gel image showing results of two-step affinity purification. 2) Western blot analysis to confirm the presence of the Hybrid IgG-Fc molecule. 3) Graph from Size Exclusion Chromatography confirming the dimerization of IgG-Fc.

3.4 Crystallisation of Hsf-PD3BD3 construct

3.4.1 Small Scale Crystal Screening

Purified PD3BD3 was concentrated to 5 mg/mL, 10 mg/mL and 14 mg/mL using Amicon Ultra-15 Centrifugal Filter Unit with 10 kDa cut-off (Merck, USA). At these concentrations, initial crystallization screens were performed using commercial kits, as mentioned in section 2.1.8, at nanolitre scale which were performed by NT-8 robot, using sitting drop vapor diffusion techniques in 1:1, 1:2 and 2:1 protein : Mother liquor ratios. Twenty plates were incubated at 4 ˚C and up to 100 Plates were incubated at 25 ˚C. Multiple tiny, fragile but fused crystals were obtained in ~ 10 % conditions at room temperature within 3-14 days of incubation and no crystals were observed at 4 ˚C. Depending on the concentration, the protein samples with 5 mg/mL did not crystallise in any of the conditions, samples with 10 mg/ml crystallised between 10 - 14 days of incubation and samples with 14 mg/mL concentration were crystallised at 3rd – 7th day post incubation in 7 conditions (Table 14). The crystals appeared at 10 mg/mL concentration grew out of precipitates and were not suitable for diffraction. Most of the crystals were from the conditions containing protein of 14 mg/mL concentration.

Table 14 List of crystallization conditions for Hsf PD3BD3 crystals

	salt	buffer	precipitant
1	0.2 M $MgCl_2 \cdot 4H_2O$, pH 6.5		20% PEG 8000
2	0.8 M NaH_2PO_4	0.1 M HEPES 7.5 pH	0.8 M KH_2PO_4
3	0.2 M $LiSO_4$	0.1 M Na Acet 4.5 pH	50 %v/v PEG 400
4	0.2 M NaCl	0.1 M CHES 9.5 pH	10 %w/v PEG 8K
5	0.1 M NaCl	0.1 M BICINE 9 pH	20 %v/v PEG MME 550
6	1 M LiCl	0.1 M BICINE 9 pH	10 %w/v PEG 6K
7	0.2 M $LiSO_4$	0.1 M TRIS 8.5 pH	40 %v/v PEG 400

3.4.2 Small Scale Crystal Optimisation

The small-scale crystal optimisation screens were generated by Rock Maker crystallisation software by Formulatrix. The software generated 96 well screens by optimising the original screen based on molarity, pH, concentration and % solubility. The proper mixing of the components was done by the robot in 96 deep well plates when provided with the core components generated by the rock maker such as the buffers, salts and precipitants. Multiple optimisation plates were generated based on the above mentioned seven screens. Unfortunately, none of the optimisation screens gave crystals of Hsf PD3BD3.

3.4.3 Large Scale Crystal Optimisation

In order to improve the crystal size, crystallization was set at microliter scale, manually, by mixing 1 µl of mother-liquor with 1 µl of PD3BD3 (14 mg/mL) and equilibrated against 1 mL of reservoir buffer (0.2 M $MgCl_2 \cdot 4H_2O$, pH 6.5, 20 %- 30 % w/v PEG 8,000) at RT. Hanging drops were set at molar ratios of 1:1, 1:2 and 2:1 PD3BD3: mother liquor in 24 well crystallization plates. The plates were again incubated at 4 ˚C and 25 ˚C which did not gave crystals, moreover, all the conditions incubated at 25 ˚C resulted in amorphous precipitates, Further refinements on crystallization conditions were tested to obtain single large crystals by changing parameters like pH, salt concentration and precipitant concentration in a similar protein to mother liquor ratios and incubated at the same temperatures as above. Considering the results of drop being clear at 4 ˚C and precipitated at 25 ˚C, the plates were incubated at 4 ˚C for overnight period and then shifted to room temperature to slow down the crystallisation rate in the beginning. Despite all the efforts these crystallisations did not occur in these conditions as well.

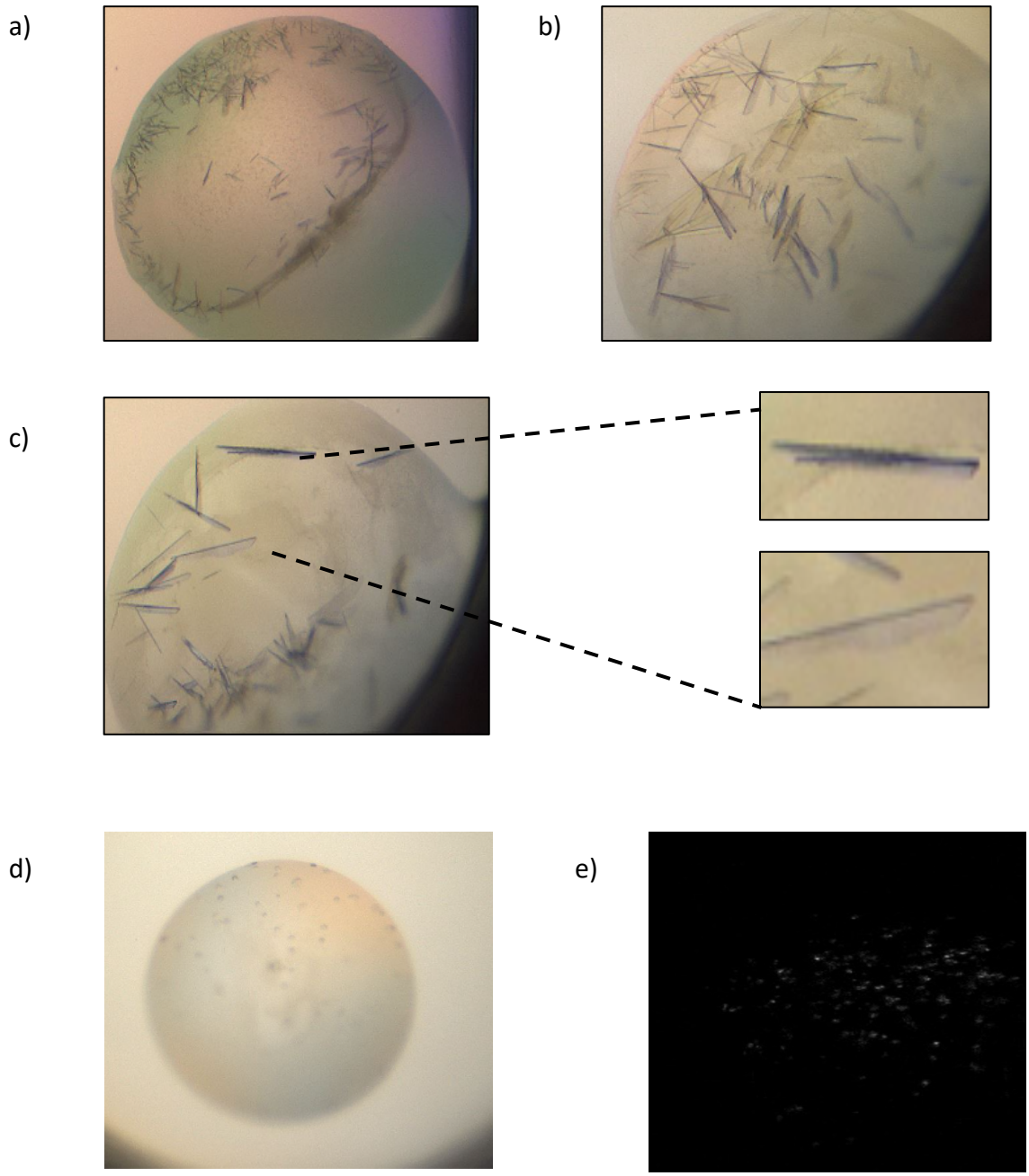


Figure 30 Crystal images of Hsf PD3BD3 domain

a) Picture of Hsf PD3BD3 crystals from 0.2 M $LiSO_4$, 0.1 M TRIS 8.5 pH and 40 %v/v PEG 400 which were harvested with the mesh loop and did not generate solvable data. b) Picture of Hsf PD3BD3 crystals from codition 0.2 M $MgCl_2 \cdot 4H_2O$, pH 6.5 and 20% PEG 8000 c) Picture of Hsf PD3BD3 crystals which were not harvestable and did not grow further. e) Picture depicted the UV signal from the Hsf PD3BD3 crystals.

3.5 Discussion

Hsf and EibD belong to the Trimeric autotransporter family of proteins expressed on the surface of *H. influenzae* and *E. coli* respectively. One of the important functions of the two proteins is to establish host-pathogen interactions. Hsf functions as a bridge between the bacterial cell membrane and the host epithelial lining, whereas, EibD interferes with the host immune response by interacting with IgG antibody. The C-terminal end of the proteins remain connected to the bacterial cell membrane while the N-terminal of the proteins adheres to the specific proteins of the target proteins inside host. This helps the bacteria to survive and colonise inside the host body. This study was aimed to cloning, expression, purification and crystallisation of the truncated segments of Hsf and EibD proteins.

Hybrid clones containing BD1 and BD2 domains of Hsf were very difficult to be cloned due to the high consensus repeats in the whole Hsf sequence. As the sequences of all the Hsf domains were very similar to each other it was very difficult to create primers highly specific to BD1 and BD2. Hence, the ligation of primers to the non-specific sites resulted in multiple non-specific PCR products some of which were of the same size as that of the desired PCR products. The most important step for the in-fusion cloning was to design the 20 bases long primers with flanking region complementary to the ligation sites in the vector. The length of the primers and the presence of consensus sequence in the Hsf gene also resulted in generation of multiple non-specific PCR products. After temperature optimisation at the template ligation step in the PCR reaction, most favourable conditions at 60 ˚C and 65 ˚C were determined for BD1 and BD2, respectively. The vector pET 28 was successfully double digested at the unique restriction sites in the MCS regions with *Nco*I and *Xho*I creating the cohesive ends for the PCR products to ligate in presence of In-fusion ligation buffer. The

clones were confirmed to contain BD1 and BD2 domains with Sanger sequencing method along with the other Hsf domains (PD1, PD2, PD3 and BD3). Several unsuccessful efforts were also made to clone the Hsf-EibD hybrid construct with In-fusion cloning method to join the passenger domain of Hsf with the anchor domain of EibD.

The expression of Hsf domains was achieved at 37 ˚C after inducing with 1 µM IPTG. The proteins were purified with Ni-NTA affinity chromatography using gravity flow columns and > 90 % purified fractions were collected with 300 mM imidazole after three step washing with wash buffers listed in Table 5. Hsf domains were found in trimeric form in solution as determined by SDS-PAGE, where the bands at sizes of trimers were seen along with the monomers. The Hsf domains BD1, BD2, BD3, PD1, PD2 and PD3 were used to test the adhesion of Hsf with Vitronectin, results of which are discussed in chapter 5. Significant efforts were also made towards expression and purification of $Hsf_{50\text{-}2316}$ which consists of the complete passenger domain of Hsf protein. However $Hsf_{50\text{-}2316}$ was successfully expressed at 37 ˚C by inducing with 1 mM IPTG for 6 hours, the protein was highly unstable and showed degradation as observed by SDS-PAGE and confirmed by western blotting.

The PD3BD3 construct containing sequence $Hsf_{888\text{-}1337}$ is important to test the hypothesis if the Hsf protein forms a unique hairpin structure. The Hsf PD3BD3 is the central region of Hsf passenger domain, was purified to gain structural insight into the hypothesis. Hsf PD3BD3 was successfully cloned and expressed at 37 ˚C using 1 mM IPTG induction. The protein was purified to > 90 % of purity. Like all other domains, Hsf PD3DB3 was found in trimeric form in solution as determined by SDS-PAGE and Western blotting. Hsf PD3BD3 formed unstable and fragile crystals at 25 ˚C after setting up over 100 crystallisation screens. The screens were set in both 96 and 24 well sparse matrices with sitting drop and hanging drop vapor diffusion methods respectively. After setting up ample amount of crystallisation screens Hsf PD3BD3

crystallised at 25 ˚C producing highly unstable and fragile crystals, which were shipped for X-ray diffraction. The results of X-ray diffraction and structure solution of Hsf PD3BD3 is discussed in Chapter 4.

Even though it was difficult to clone, express and purify the Hsf domains, the short, truncated constructs containing single domains had higher protein yields producing sufficient quantity for their downstream characterisation. Due to constant degradation, the yield varies among purifications which made it difficult to calculate the yields of the proteins. Despite all difficulties, I think, it is strategically correct to work with the smaller domains than the full length Hsf because the full length Hsf is constantly degrading, highly unstable and low yielding protein.

Chapter 4 Determining Hsf PD3BD3 structure through X-ray crystallography

4.1 Introduction

The aim of the work in this chapter was to solve the structure of the $Hsf_{888-1227}$ region of full-length Hsf, containing both PD3 and BD3 domains connected by 78 residue long neck region. To this end, Hsf-PD3BD3 had been cloned in pET 28 c vector previously by Dr. Kornelia M Mikula (Helsinki). This clone was used in the current study to solve the structure of Hsf PD3BD3 domains. Hsf PD3BD3 was expressed and purified in *E. coli* as part of the work in Chapter 3 and was used to carry out structural studies. The structure of only PD1 domain of Hsf has been previously solved in the Goldman lab(Wright et al., 2017). There were no previous structures of Hsf PD3BD3 domains, and therefore any structural information obtained would provide valuable knowledge about the overall morphology of Hsf protein. Most Importantly, this structure serves as crude evidence against the hypothesis of Hsf forming a unique hairpin like morphology.

Prior to this, the crystal structure of trimeric PD1 was solved at a resolution of 3.3 Å, which provided the first insight into the molecular arrangement of Hsf. The Hsf PD1 shows > 50 % sequence similarity with Hsf PD3 domain which made it a good template for molecular replacements. Another protein is BD1 domain of well-characterised protein Hia of *Haemophilus influenzae* which shares >30 % sequence similarity with Hsf BD3 domain, also serves as second template for molecular replacement.

4.2 Sequence comparison of Hsf PD3BD3

Trimeric Autotransporter Adhesins (TAA) family of proteins have highly conserved considerable sequence identity among the proteins(Wright et al., 2017). These conserved domains are very useful and important for molecular replacement. For Molecular replacement of Hsf PD3BD3, I used two different high-resolution protein structures, one

belongs to the PD1 domain of Hsf (Wright et al., 2017)and the other is the structure of BD1 domain of Hia protein of *H. influenzae*(Wright et al., 2017). The sequence alignment to show the feasibility of these two proteins to be used as templates for Molecular replacement is described in the sections below.

4.2.1 Sequence comparison of Hsf PD3 with Hsf PD1

Hsf encodes for 2413 residues with three putative and three binding domains. All the domains of *H. influenzae* shares more than 60% sequence identity with each other. The sequences shows that the all the domains of Hsf contains two Tryptophan ring domains separated by loops constituting neck domains. The first Trp Ring starts from the third residue Tryptophan (W) to residue 31 Phenylalanine (F) and the second Trp Ring starts from residue 112 Tryptophan (W) to residue 145 Phenylalanine (F)

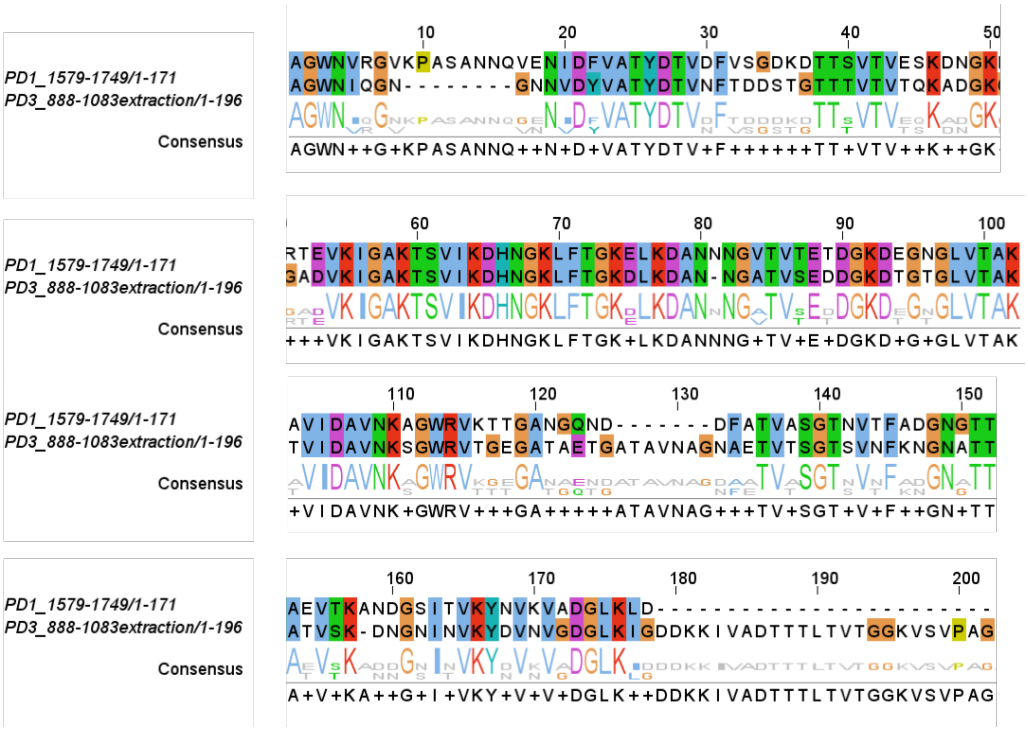


Figure 31 Sequence alignment of Hsf PD1 with Hsf PD3

Sequence comparison between PD1 and PD3 domains of Hsf protein is shown. The extent of conservation is shown by the bigger font size residues in multiple colours. Identical residues are also depicted as consensus. Bigger the font size represents greater conservation.

Hence, I used it as one of the structures to obtain phase information for Hsf PD3BD3 by Molecular Replacement. Hsf domains PD1 and PD3 share more than 90% sequence identity as shown in Figure 31.

4.2.2 Sequence comparison of Hsf BD3 with Hia BD1

Among TAA proteins, the molecular structure of Hia is extensively studied and also serves as one of the model proteins of the family. Figure 32 displays pairwise sequence alignment of Hsf with the structurally characterized *Haemophilus influenzae* Adhesin (Hia), showing the conserved residues highlighted in red. The sequence identity of 64.4% allows to have structure insight as it would facilitate the structure determination of $Hsf_{888\text{-}1342}$ by molecular replacement. More specifically, Binding domain 1 (BD1) of Hia shares 70% sequence identity with Hsf BD3 domain, as shown in Figure 32. Hence, in addition to Hsf PD1, I used Hia BD1 structure for molecular replacement of Hsf PD3BD3.

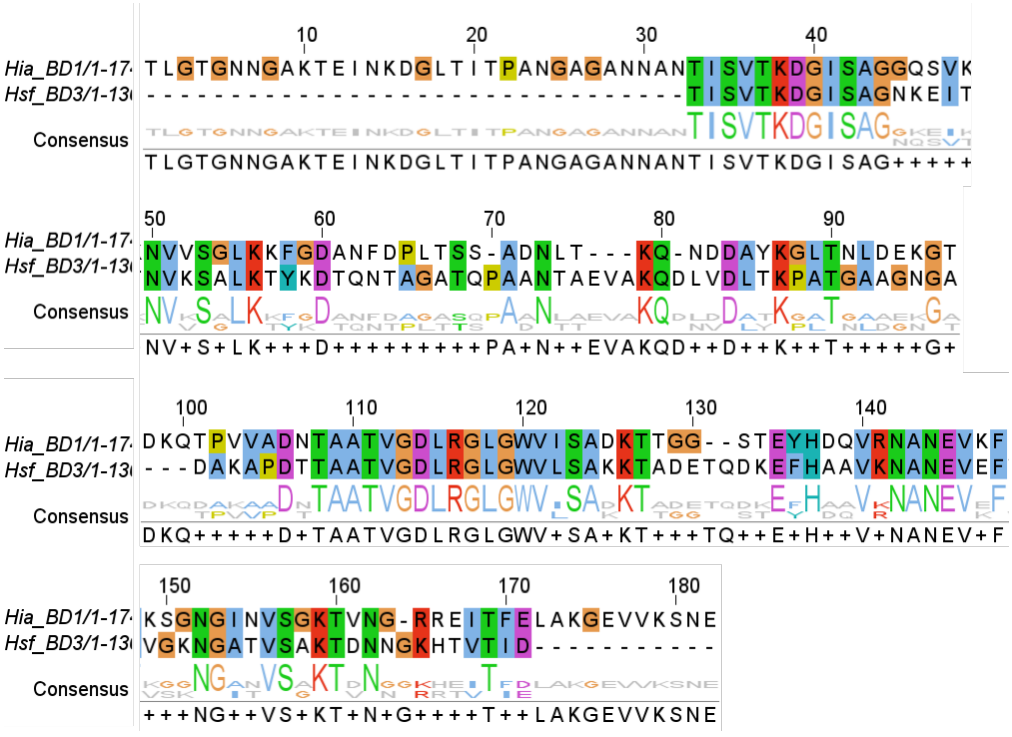


Figure 32 Sequence alignment of Hia BD1 with Hsf BD3

Sequence comparison between Hia BD1 and Hsf BD3 domains is shown. The extent of conservation is shown by the bigger font size residues in multiple colours. Identical residues are also depicted as consensus. Bigger the font size represent greater conservation.

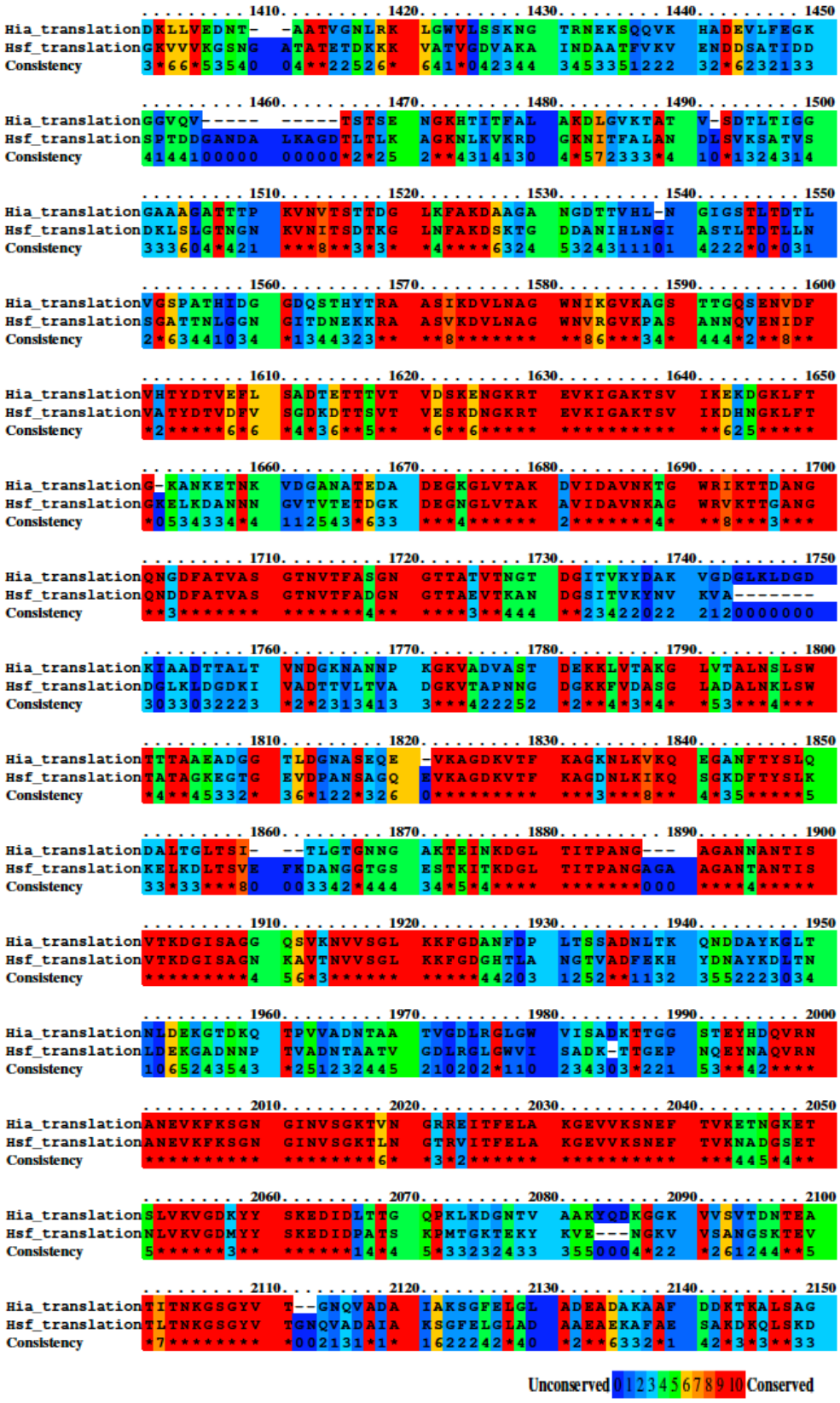


Figure 33 Pairwise Sequence alignment of Hia with Hsf

Partial Sequence comparison between Hia and Hsf protein is shown. The extent of conservation is shown at the bottom, identical residues are depicted red, while least conserved residues are in blue colour. Higher values represent greater conservation.

4.3 Structural study of Hsf using X-ray diffraction

4.3.1 data collection and data processing

Multiple crystals of Hsf PD3BD3 were obtained in condition containing $MgCl_2 \cdot 4H_2O$, $C_2H_{12}AsNaO_5 \cdot 3H2O$ pH 6.5, and PEG 8,000. The crystals obtained were merged and hence, were very difficult to be harvested as a single crystal from the drop. Crystals were carefully harvested under the microscope with cryo-loops (Hampton Research, USA) and soaked in a cryo-protectant solution containing 30% PEG8000 in mother liquor before flash freezing in liquid nitrogen. After flash freezing crystals were transferred for X-ray diffraction. In order to collect X-ray diffraction data, the dewar containing crystals were shipped to Diamond Synchrotron Facility, UK.

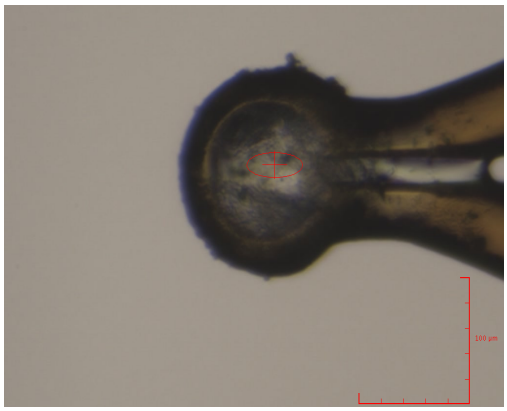
The X-ray data were collected on microfocus beamline I02-VMX at Diamond Light Source (DLS) and the diffraction data parameters are listed in Table 15. An image of the crystal mounted on the beam line is shown in Figure 34. Although radiation damage restricted the data set to the first 1100 images, the crystals diffracted to 3.06 Å resolution and the representative image Figure 34b show the merged diffraction spots. Indexing and integration were performed using XDS(Kabsch, 2010), while scaling and merging statistics were calculated using AIMLESS(Evans & Murshudov, 2013) at DLS.

The data collected from the DLS were added to the PHASER MR as the initial model of PD3BD3 was built by molecular replacement (MR)(McCoy et al., 2007). I used two high resolution structures as templates for Molecular Replacement (MR) which are Hsf PD1 domain (PDB entry 5lnl(Wright et al., 2017)) and Hia BD1 domain (PDB entry 3emf(Meng, St. Geme, & Waksman, 2008)). The trimeric structures of the two proteins are shown in Figure 35 with their domain annotations. Phaser MR found one unique solution in space group $P2_1$.

Table 15 Parameters of X-ray Diffraction data

Diffraction Parameters	*Values*
Diffraction Source	*I02, DLS*
Wavelength (Å)	0.97
Temperature (K)	*100*
Rotation range per Image (°)	*0.1*
Total rotation range (°)	*110*
Exposure time per Image	*0.1*
Space group	$P12_11$
A,b,c (Å)	42.854 53.043 276.189
α,β,γ (°)	91.079
Diffraction limit	2.955 3.465 3.202
High resolution limit	3.067
R_{merge}	0.276
Total number of Reflections	116412
Number of unique reflections	18423
(I/σI)	5.7
Completeness	77.6
Multiplicity	6.3

(a)



(b)

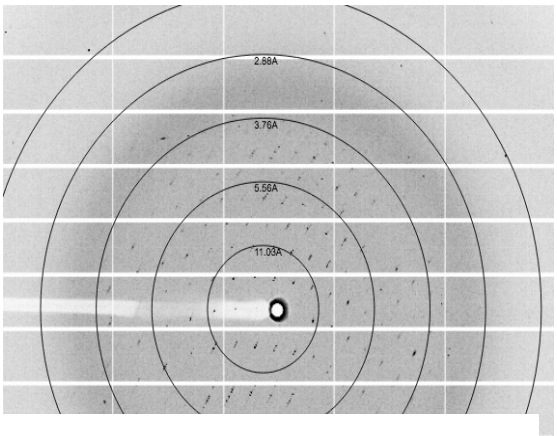


Figure 34 X-ray diffraction image.

(a) Triangle shaped Crystal of PD3BD3 mounted for x-ray diffraction. (b) X-ray diffraction image of PD3BD3 crystals from Diamond beamline.

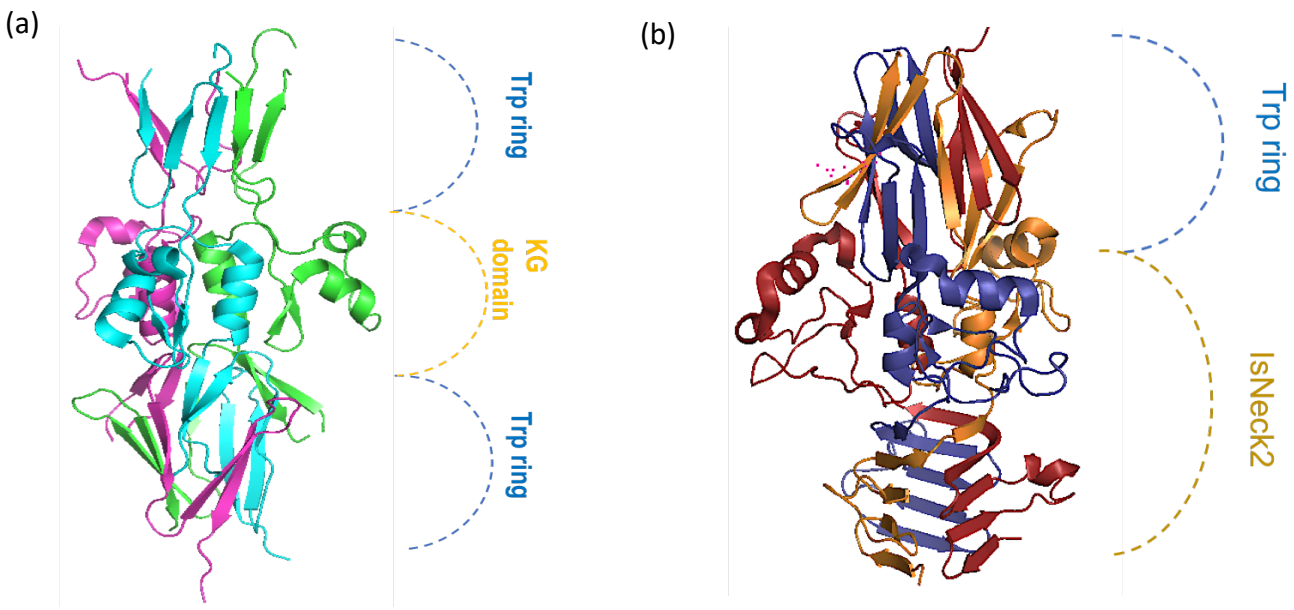


Figure 35 Templates for molecular replacement pf PD3BD3.
(a) Template- Hsf-PD1 (Pdb ID: 5LNL), (b) Template- Hia-BD1 (Pdb ID: 1S7M).

4.3.2 Molecular replacement and Electron density map

The initial structure of Hsf PD3BD3 was determined by Molecular replacement using Phaser MR in PHENIX and Hsf PD1 and Hia BD1 as templates (Figure 35). The initial structure of Hsf PD3BD3 was subjected to rigid body refinement followed by restrained refinement using phenix.refine program and manual investigation and model building using the COOT program. The program PyMOL was used to visualize and analyse the model unless mentioned. Table 15 and 14 summarizes details of data collection and refinement.

The Initial structure of Hsf PD3BD3 had three out of four expected TrpRing domains, Isneck of BD3 and KG domain of PD3 are present but with low electron density, specifically for the side chains. The density of Hsf PD3BD3 structure was not continuous and was very poor for the region connecting the two domains and one of the four Trp ring domains of Hsf; hence, the region was built manually by Dr. James Hillier using COOT program. The electron density was comparatively improved with multiple rounds of refinement but still had missing regions,

especially for the loops as shown in Figure 36. Also, the density for the same regions varies among the chains, hence, each chain requires a different solution for model building.

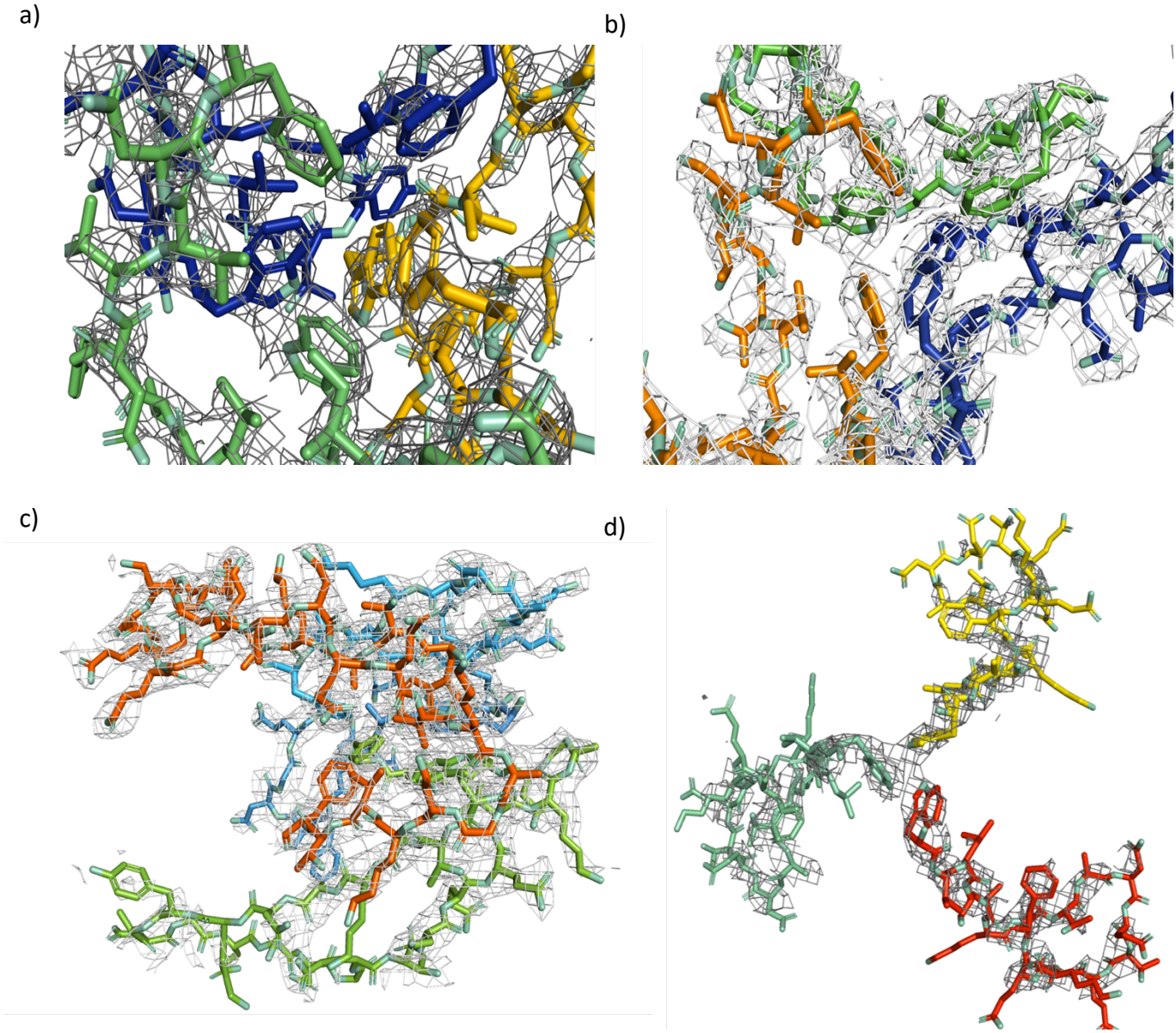


Figure 36 Examples of 2FoFc electron density maps

a) Electron density map around the first Trp Ring of Hsf PD3BD3 shown as an example for good electron density at the amino terminal. b) Electron density map around the second Trp Ring of Hsf PD3BD3. c) Electron density around the third Trp Ring domain of Hsf PD3BD3 showing discontinuous density around the loops, especially in chain B (in green colour). d) Electron density around the fourth Trp Ring domain of Hsf PD3BD3 showing discontinuous density around the loops, especially in chain C (in red colour).

Table 16 Parameters of data refinement

Diffraction Parameters	***Values***
Resolution (Å)	2.8
Completeness (%)	70.8
Number of reflections	18421
R_{work}/R_{free}	0.26/0.31
Average B Factor (Å²)	64.7
Ramachandran Plot	
Most Favored (%)	90.40
Outliers (%)	2.23
R.M.S.D	
Bonds *(Å)*	0.03
Angles(°)	0.12

4.4 Structure comparison of Hsf PD3BD3

4.4.1 Structure comparison of Hsf PD1 with Hsf PD3 domain

The trimeric Hsf PD1 domain has three well-characterised sub-domains of TAA family, N-terminal Trp Ring domain, a KG domain and a C-terminal Trp Ring domain which are also seen in PD3 trimer. A comparison of both Hsf PD1 and PD3 domains is shown in Figure 37a where PD1 is in green colour and PD3 in blue colour Figure 31a. The N-terminus of Hsf PD1 has three β-strands, whereas the N-terminal TrpRing domain of Hsf PD3 have five β-strands, $_{\beta}W1^{1}$, $_{\beta}W1^{2}$, $_{\beta}W1^{3}$, $_{\beta}W1^{4}$ and $_{\beta}W1^{5}$ (where 'W' represents tryptophan) as shown in Figure 31b, whereas PD1 domain only have Three β-strands. In both, PD1 and PD3 domains of Hsf the KG domain is composed of two β-strands, as well as three α-helices, $_{\alpha}KG^{1}$, $_{\alpha}KG^{2}$ and $_{\alpha}KG^{3}$. The C-

terminal Trp Ring in PD1 domain is composed of five β-strands whereas in PD3 it is composed of seven β-strands $_{\beta}W2^{1}$, $_{\beta}W2^{2}$, $_{\beta}W2^{3}$, $_{\beta}W2^{4}$, $_{\beta}W2^{5}$, $_{\beta}W2^{6}$ and $_{\beta}W2^{7}$ Figure 37b.

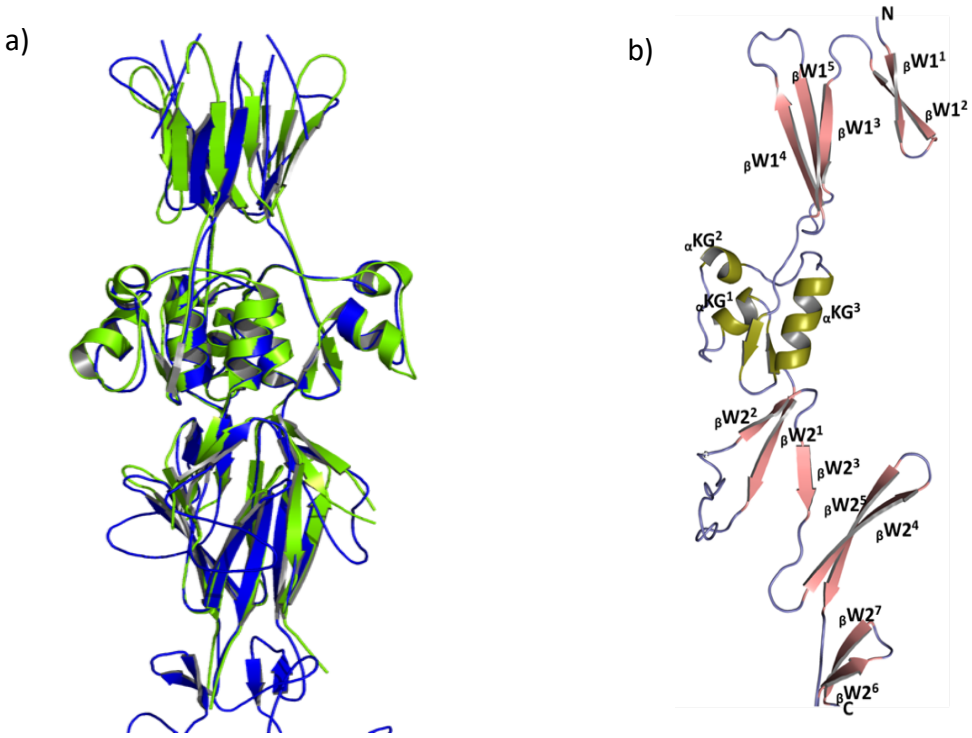


Figure 37 Structure comparison of Hsf PD1 with Hsf PD3

a) Structure comparison of hsf domains PD3 in green and PD1 in blue is shown. b) Single chain of Hsf PD3 domain showing the five β strands $_{\beta}W1^{1}$-$_{\beta}W1^{5}$ of the first Trp ring of PD3 domain, three α helices of KG neck domain: $_{\alpha}KG^{1}$-$_{\alpha}KG^{3}$ and seven β-strands of the $_{\beta}W1^{1}$-$_{\beta}W1^{5}$ of the second Trp ring

4.4.2 Structure comparison of Hia BD1 with Hsf BD3 domain

Hsf and Hia are remarkably similar in their domain arrangement, and share considerable sequence identity in their shared region. The BD1 domain of Hia has been described to have 3 domains. Later with similarity searches among the conserved and well-studied TAA domains, the Hia BD1 domains were described as Trp Ring domain and a long neck with insertion also called as IS Neck. Hence, we can say that Hia BD1 has a contiguous Neck-Trp Ring architecture. The superposition of Hsf BD3 with Hia BD1 (PDB entry 1s7m; Yeo et al.,

2004) reveals modest structural similarity. An overlay of both the structures is shown in Figure 38a, where Hsf BD3 is in green colour and Hia BD1 is in pink colour. However, in addition to the C-terminal Trp ring and IS Neck domain, Hsf BD3 also contains another Trp ring at its N-terminal. The Trp ring at the N-terminal of Hsf BD3 domain is composed of 6 β-strands: $_{\beta}W^{1}$, $_{\beta}W^{2}$, $_{\beta}W^{3}$, $_{\beta}W^{4}$, $_{\beta}W^{5}$ and $_{\beta}W^{6}$ (Figure 38b). Similar to Hia BD1, the IS Neck domain of Hsf BD3 also consists of five β-strands- $_{\beta}ISNeck^{1}$, $_{\beta}ISNeck^{2}$, $_{\beta}ISNeck^{3}$, $_{\beta}ISNeck^{4}$ and $_{\beta}ISNeck^{5}$ and three α helices- $_{\alpha}ISNeck^{1}$, $_{\alpha}ISNeck^{2}$ and $_{\alpha}ISNeck^{3}$. Following to the C-terminal Trp ring domain of Hsf BD3 is composed of 5 β-strands- $_{\beta}W^{1}$, $_{\beta}W^{2}$, $_{\beta}W^{3}$, $_{\beta}W^{4}$ and $_{\beta}W^{5}$, similar to Hia BD1 as shown in Figure 38b.

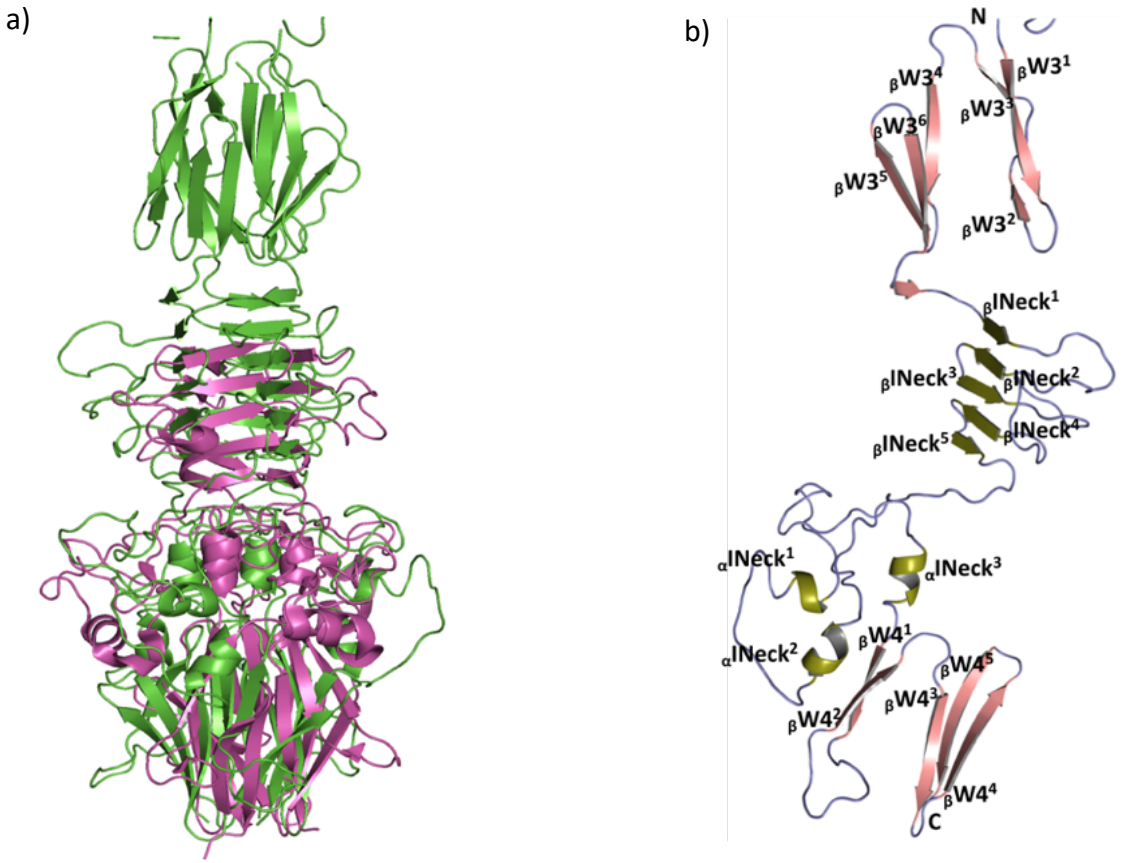


Figure 38 Structure comparison of Hsf BD3 with Hia BD1

a) Structure comparison of hsf domains PD3 in green and PD1 in pink is shown. b) Single chain of Hsf PD3 domain showing the five β strands $_{\beta}W1^{1}$- $_{\beta}W1^{5}$ of the first Trp ring of PD3 domain, three α helices of KG neck domain: $_{\alpha}KG^{1}$- $_{\alpha}KG^{3}$ and seven β-strands of the $_{\beta}W1^{1}$- $_{\beta}W1^{5}$ of the second Trp ring

4.5 Overall structure of Hsf PD3BD3

The final structure of Hsf PD3BD3 was solved at 2.8 Å resolution. The asymmetric units of Hsf PD3BD3 crystal possess three monomers (Figure 40, Figure 41) and as the functional unit is trimer, it exists as homo-trimer in solution. Hsf PD3BD3 attained a stalk-like architecture and the monomeric unit is defined by the presence of two subdomains arranged in stalk-like architecture (Figure 40). The N- and C-terminal subdomains are comprised of residues: 2-215 and residues: 216-459 respectively. Each subdomain contain two Tryptophan rings joined by small, highly flexible Neck domain (Figure 40). These four Trp ring domains form a hydrophobic core which acts as a glue that holds the trimer together as no other interaction are observed among the three monomers of the Hsf PD3BD3 structure. The hydrophobic core runs through the structure and zoomed in image of from the C-terminal end is shown in Figure 39.

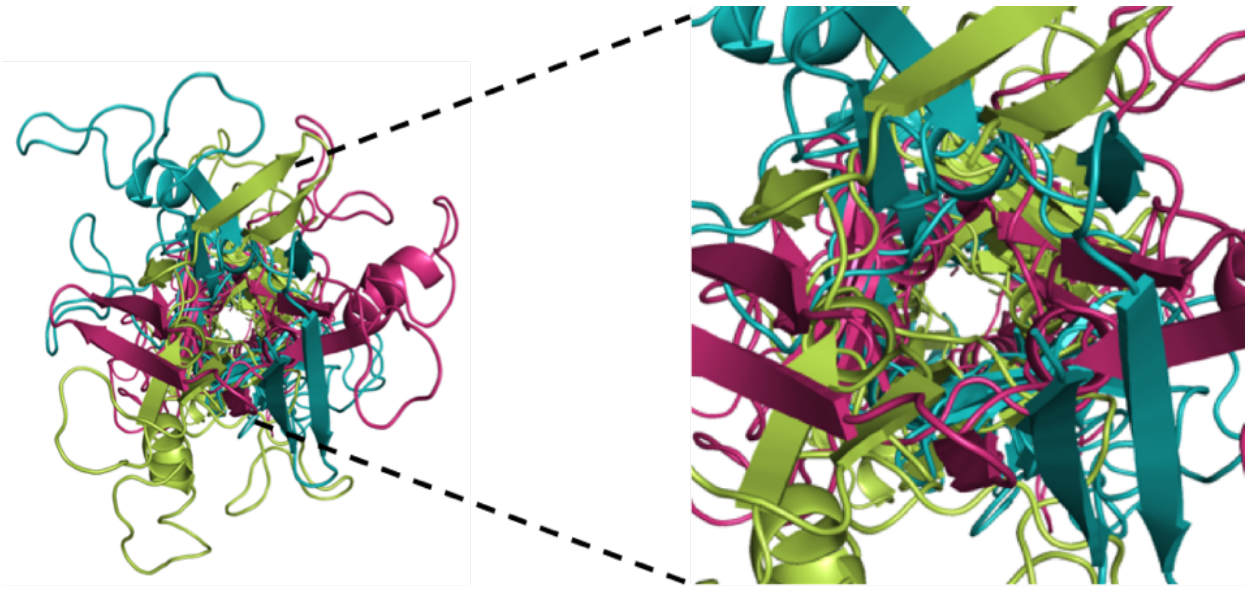


Figure 39 Zoomed in image of the hydrophobic core of Hsf from the C-terminal

Transverse image of the Hsf PD3BD3 structure from the C-terminal end showing zoomed out and zoomed in image of the hydrophobic core of Hsf PD3BD3 trimer structure.

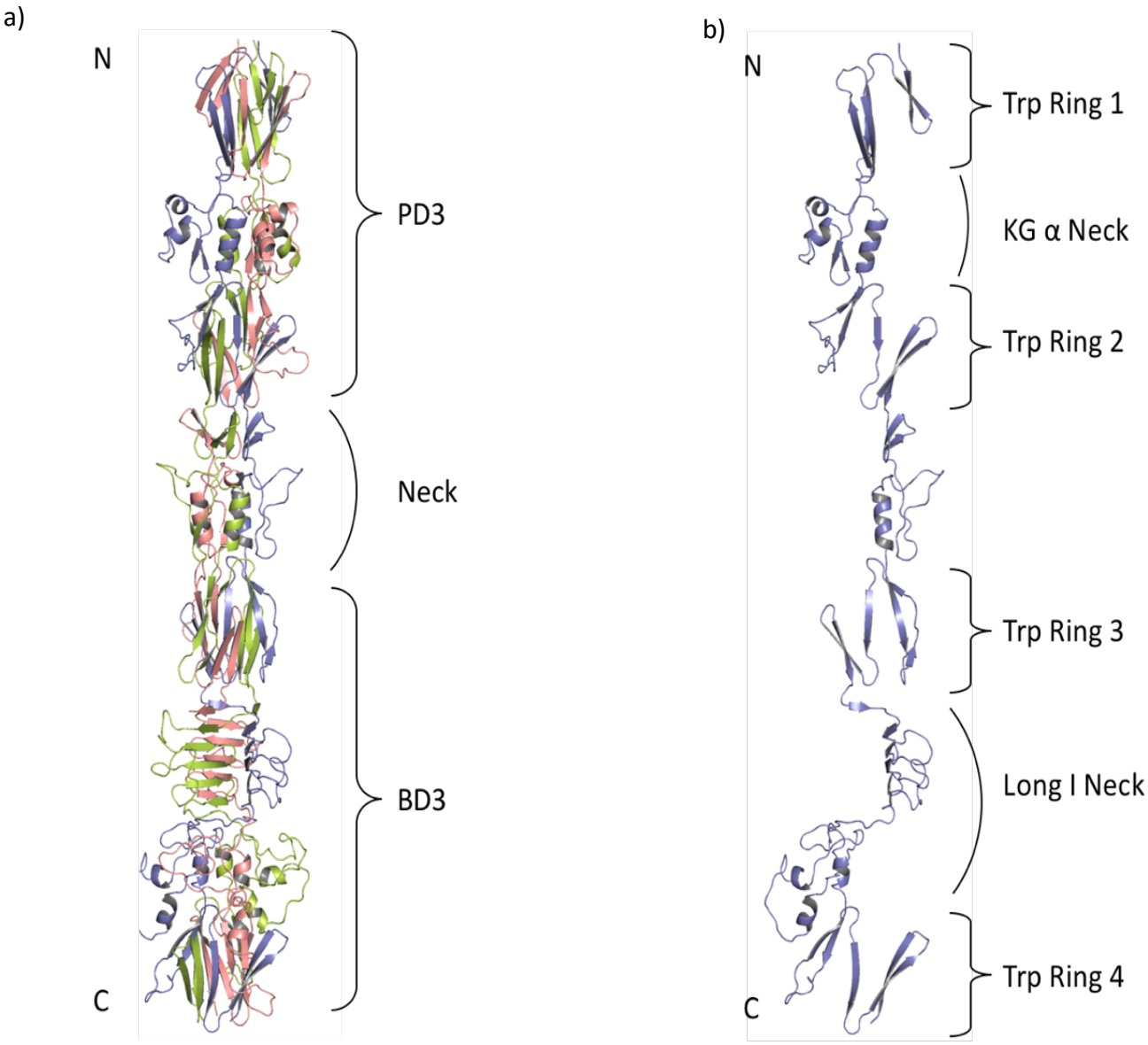


Figure 40 Overall Structure of Hsf PD3BD3

a) Crystal structure of Hsf PD3BD3 domain showing trimeric structure with each monomeric chain depicted in different colour. PD3 domain is at amino-terminal and BD3 domain at carboxy-terminal. b) Monomeric chain of Hsf PD3BD3 structure showing four Trp Ring domains (Trp Ring 1-4) joined by KG α Neck domain in PD3 and long insertion Neck in BD3.

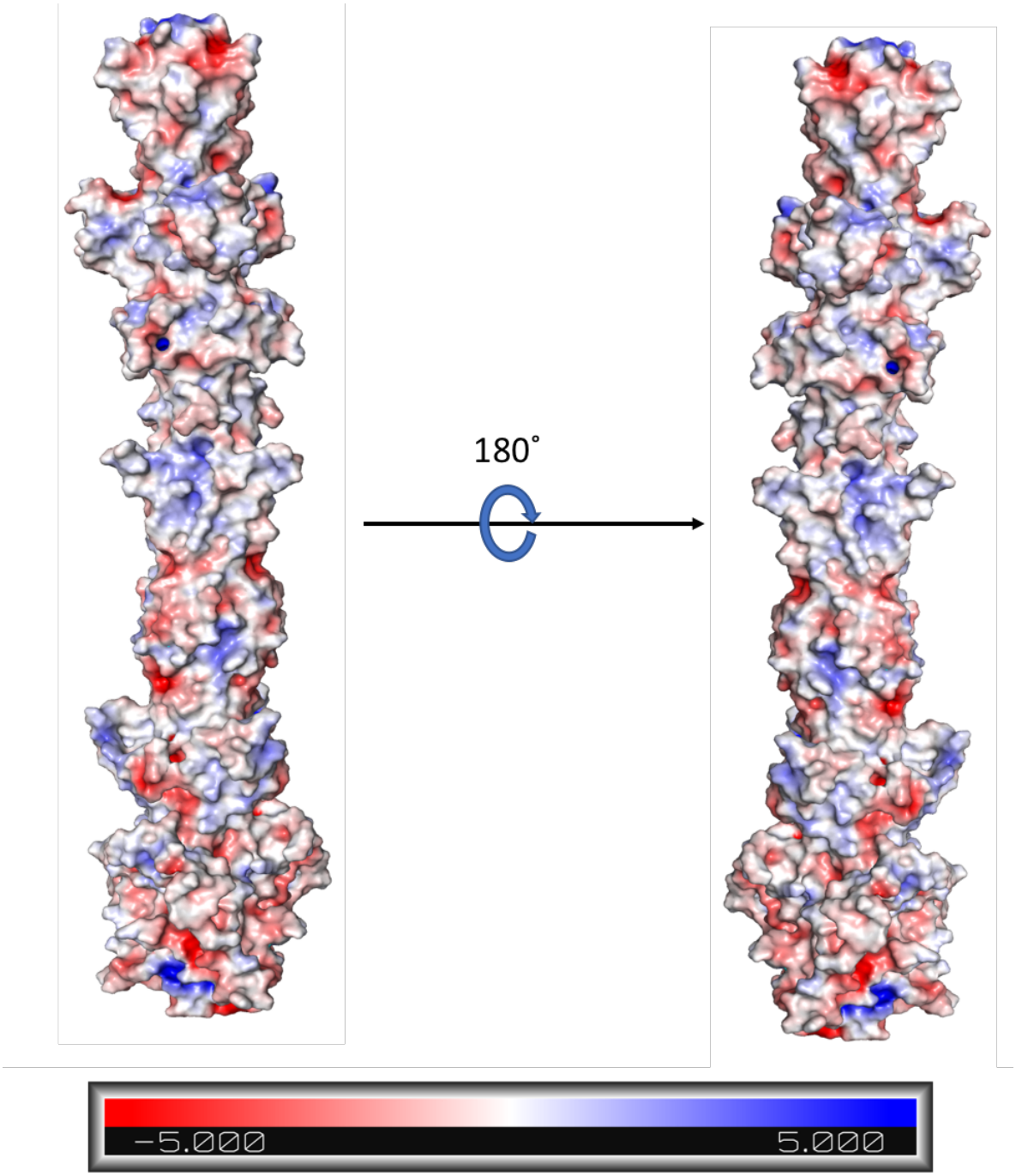


Figure 41 Electrostatic surface potential map of Hsf PD3BD3

4.6 The bend in the Hsf PD3BD3 structure

The objective of the chapter is to test the hypothesis of Hsf forming a novel Hairpin like structure. Based on previous knowledge, if Hsf forms a Hairpin loop the region of complete bend of the protein must be the region containing PD3 and BD3 domains. These domains are half-way through the long passenger domain. The length of the Hsf PD3BD3 structure is 250.5 Å. The solved High-resolution structure of Hsf PD3BD3 domains does not show any U-turn or hair-pin structure as shown in Figure 42. Interestingly, the protein structure shows a slight bend of 23.5° between the connecting neck region and PD3 domain, represented with a red line in Figure 42. The bend in the structure was calculated in PyMOL by selecting the three similar residues, one from each- BD3, KG Neck and PD3 domains and measuring the angle. The angle was measured for all three chains and average was calculated for the three readings.

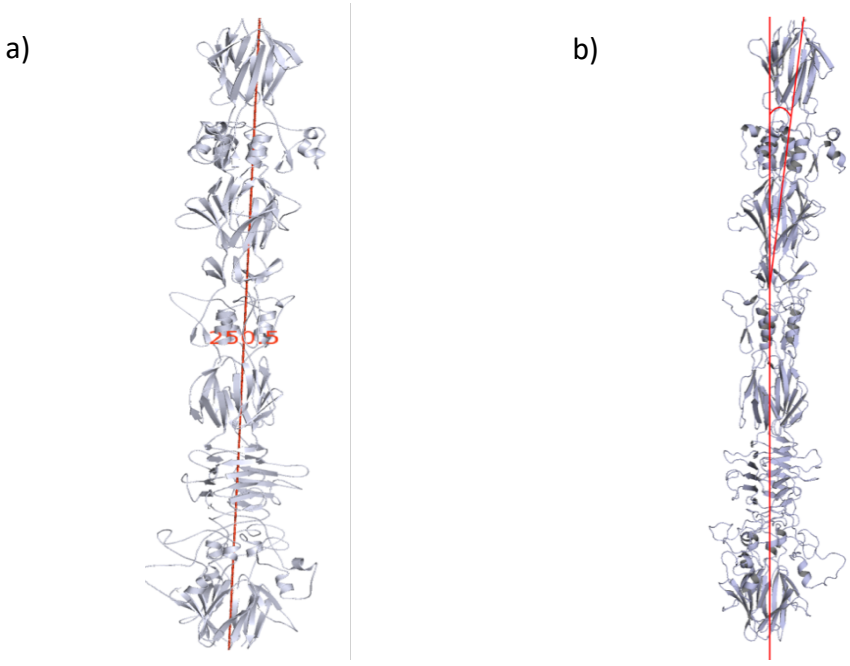


Figure 42 Measurement of the length and angle in the Hsf PD3BD3 structure.

a) The length of the protein was measured to be 250.5 nm by selecting the first and last residues of one of the chains of Hsf PD3BD3 structure. b) The Hsf PD3BD3 structure slightly bend at the junction of Hsf PD3 domain and the joining neck domain forming an obtuse angle of 23.5°.

This slight bend is a novel characteristic of the Hsf protein which has not been observed in any other protein of TAA family. Despite being a novel characteristic, a slight bend of 23.5° does not support the hypothesis of Hsf forming a hairpin structure. Hence, my study does not support the hypothesis of Hsf forming a Hair-pin loop architecture.

4.7 Hydrophobic core of the Hsf PD3BD3

`The functional unit of adhesin proteins is homo-trimer. Therefore, it is crucial to understand the interactions involved in the trimerization of the adhesin group of proteins. As hydrophobic index of adhesins are high, it is quite intuitive to assume that hydrophobic interactions could play an important role in the trimerization. Hence, the hydrophobic core of PD3BD3 may contribute to the stability of the trimeric structure of Hsf as no other interactions are involved in trimerization of the protein. Hsf PD3BD3 contains four TrpRing domains constituting a hydrophobic core of the trimer. These TrpRing domains consist of one tryptophan residue from each of the monomeric subunits of the trimeric Hsf PD3BD3 protein. The sidechains of all three tryptophan residues of the TrpRing domain face each other in the hydrophobic core of the protein represented in Figure 43. The Trp Ring domains are one of the most common interleaved head domains of the Trimeric Autotransporter Adhesins. The sequence of the Trp Ring domains are listed in Table 17, where the distribution of the hydrophobic amino acid residues in all four TrpRing domains is represented in red. The sequence similarity of all four TrpRing domains is 96%. The hydrophobic core of the protein may contribute to the stability of the trimeric structure of the protein as no other interactions are involved in trimerization of the protein.

Table 17 List of sequences of Trp Ring domains of Hsf PD3BD3.

Trp Ring 1 and 2 belong to Hsf PD3 domain and Trp Ring 3 and 4 belong to Hsf BD3 domain. The hydrophobic residues are written in red colour in the sequences.

Trp ring domains	*Sequences*
Trp ring 1	AGWNIQGNGNNVDYVATYDTVNF
Trp ring 2	SGWRVTGEGATAETGATAVNAGNAETVTSGTSVNF
Trp ring 3	LSWTAKADKYADGESEGETDQEVKAGDKVTF
Trp ring 4	LGWVLSAKKTADETQDKEFHAAVKNANEVEF

4.8

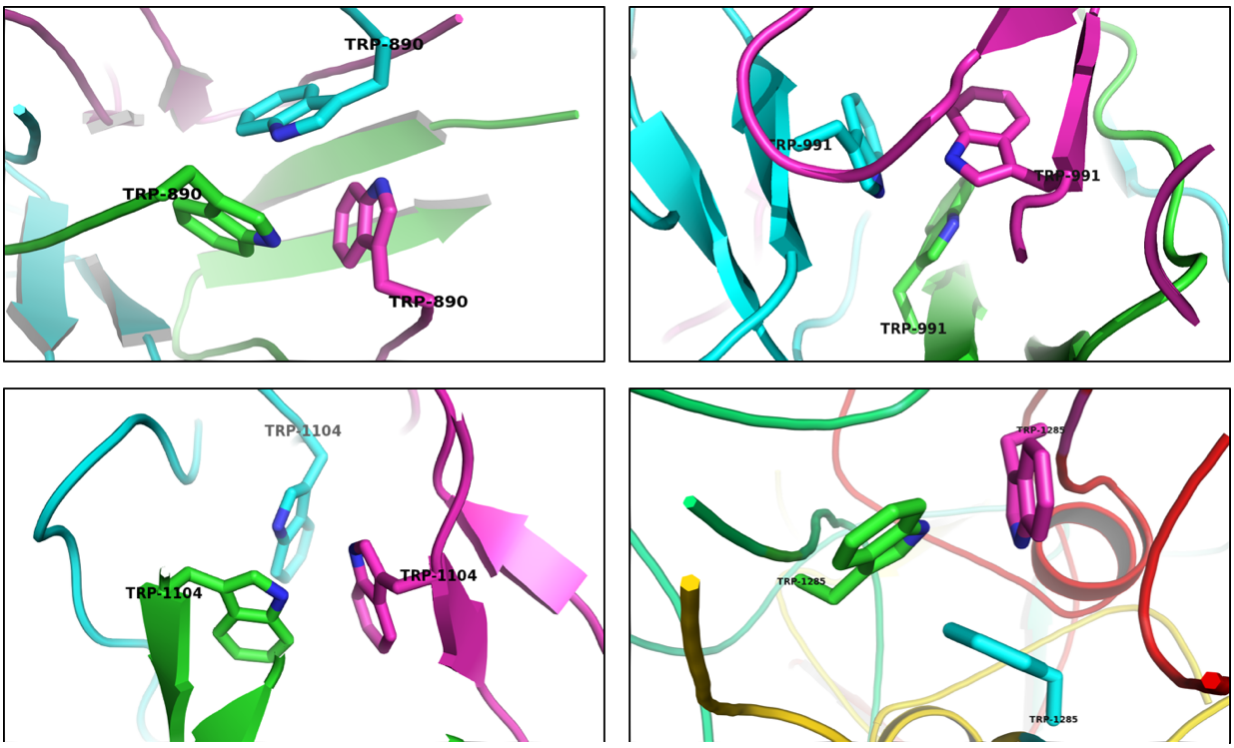


Figure 43 Figure of the four Trp Ring domains of Hsf PD3BD3

The figure represents the alignment of the tryptophan domains forming hydrophobic core of the protein.

4.9 Discussion

My aim was to test the hypothesis of Hsf forming a novel hairpin-like architecture. To test the hypothesis, I carried out the structural studies on trimeric $Hsf_{888-1227}$ containing the putative domain 3 (PD3) and binding domain 3 (BD3), joined together by 78 amino acids long neck domain. The Hsf PD3BD3 constitute the centre of 2413 amino acid long Hsf protein and the region most likely to contain the bend if it forms a hairpin-like architecture.

Hsf PD3BD3 protein existed as a trimer in solution, which was confirmed by SDS-PAGE and Western blot analysis. Model generated by DATAA software suggested that both PD3 and BD3 domains of Hsf consisted of two Trp Ring domains separated by Neck domains(Wright et al., 2017). The presence of the TrpRing domains joined by the neck domain was also observed, high-resolution structure of Hsf PD1 domain. The monomeric unit of Hsf PD3BD3 was 50 kDa big and the size of the protein was 150 kDa as trimer. The big size and highly adhesive nature of the protein made it difficult to work with.

Initial crystallisation attempts of Hsf PD3BD3 produced crystals that were not suitable for diffraction. Through optimisation of conditions, attempts using different types and protein concentrations and drop ratios, I obtained crystals that were very thin plates fused together. These crystals were very fragile and harvested in presence of 30% PEG 8000. These crystals diffracted at 3.2 Å. At this resolution, we had incomplete electron density. The data were processed and the phase information was obtained by molecular replacement. The model obtained from molecular replacement proved the trimeric state of Hsf PD3BD3. It contained nine chins, three chains of each of the domains Hsf PD3, joining neck and Hsf BD3. Molecular replacement generated three out of the four Trp rings and poor density present for the fourth trp ring. The electron density was also poor for the highly flexible loop regions of the neck

domains. So far, it took over 80 refinement cycles to generate the complete structure of Hsf PD3BD3 and structure is now present as three monomeric chains.

For the trimeric Hsf PD3BD3 structure, no interaction or bonds were observed among the monomeric chains to hold the trimer together. The presence of hydrophobic and polar amino acid residues was observed in the sequence, distributed all over the structure, as shown in Figure 44. The hydrophobic residues formed a tunnel that runs through the Hsf PD3BD3 structure responsible for holding the trimer together. The structure of Hsf PD3BD3 trimer showed no signs of a hairpin loop formation and appeared to have a long stalk-like architecture as seen in other well-characterised TAA proteins such as YadA, BadA and EibD. Although, a bend of 23.5° was observed in the Hsf PD3BD3 structure which is most likely to occur in a long protein like Hsf. Another crucial factor is the presence of highly flexible loops in the structure which may also be the reason for the observed bend in the Hsf PD3BD3 structure. Hence, my work proved that Hsf does not form a hairpin loop and have a long stalk-like morphology.

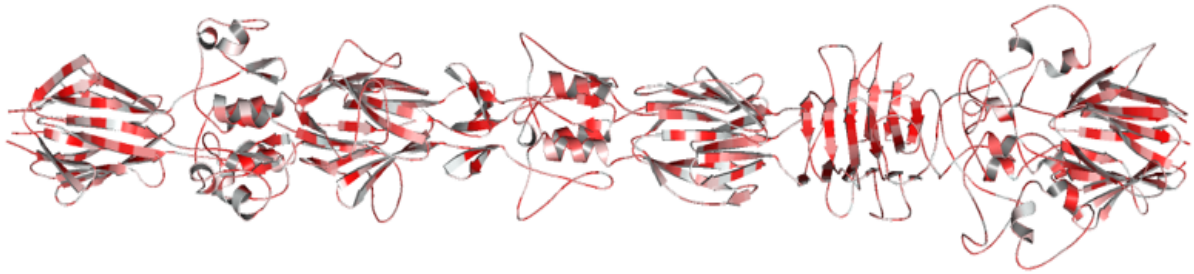


Figure 44 Hsf structure showing the presence of hydrophobic residues in Hsf PD3BD3 structure. Colour code red to white signifies red being the most hydrophobic.

Chapter 5 Elucidating the Adhesive Properties of Hsf Domains

5.1 Introduction

A primary goal of this study was to look at the feasibility of developing a fast, cost effective diagnostic tool to detect *H. infuenzae* infection by employing the adhesive properties of Hsf protein. The first step was to elucidate the molecular mechanisms underpinning the adhesion of Hsf to the host epithelium. *Haemophilus influenzae* type b (Hib) escapes the host immune system by recruitment of the complement regulator Vitronectin, which inhibits the formation of the membrane attack complex (MAC)(Hallstrom et al., 2014; Singh et al., 2014). The binding of Hib with Vitronectin interferes with the C5b-C7 complex formation and C9 polymerization(Milis et al., 1993; Singh, Su, & Riesbeck, 2010). It is been reported that Hib acquires Vitronectin at the surface by using Haemophilus surface fibrils (Hsf).

Here, we studied in detail the interaction between Hsf and vitronectin by employing the adhesion of purified single domains of Hsf to the Vitronectin. The Vitronectin binding region of Hsf was defined at the amino terminus region comprising Hsf residues 429 to 652(Singh et al., 2014). Moreover, the Hsf recognition site on Vitronectin consisted of the carboxy terminus residues 352 to 374(Singh et al., 2014). It has been reported that *H. influenzae* was killed more rapidly in Vitronectin-depleted serum than in normal human serum(Singh et al., 2014). An increase in MAC deposition was also observed at the surface of an Hsf-deficient *H. influenzae* mutant. Moreover, when Vitronectin was bound to Hsf, an increase in bacterial adherence and internalization into epithelial cells were observed. Taking these findings together, it has been hypothesised that a fine-tuned protein-protein interaction between Hsf and Vitronectin may contribute to increased Hib virulence(Singh et al., 2014).

This chapter explores the involvement of binding and putative domains of Hsf in establishing interaction with Vitronectin. Here, I have designed an experiment to study the interaction between the purified protein fractions of Hsf domains and commercially synthesised

molecular grade Vitronectin. Commercially available electrodes from Eluceda pvt. Ltd. were used for the interaction study by employing Cyclic Voltammetry (CV).

5.2 Exploring the difference in adhesion of Hsf domains on different electrode surfaces

Cyclic voltammetry (CV) is a powerful and popular electrochemical technique commonly employed to investigate the reduction and oxidation processes of molecular species. CV is also valuable to study chemical reaction initiated by electron transfer for example catalysis. CV based experiments were designed to study the protein-protein interactions with electrodes of different surfaces for the real-time detection of interactions between Hsf domains and Vitronectin. The proteins were deposited on the electrodes in a layer-by-layer manner and with CV the oxidation and reduction of a redox active reagent was studied. In generic terms, a simplified relevance to the oxidation peak height is that the decrease in oxidation suggested the successful deposition or adhesion of proteins on the electrode surface. In CV, the potential is ramped up from a holding potential to a switching potential and back, usually at a set scan rate which generates a plot of current vs applied potential. The cyclic voltammogram, is typically a “duck-shaped” graph under the effect of a redox reagent. Here, I explore the adhesion of different Hsf domains on different electrode surfaces as well as with the Vitronectin, under the effect of Ferrocene.

5.2.1 Experimental design for the adhesion of Hsf domains on the electrode surfaces

In order to study the interaction of individual `Hsf domains, which are binding domains- BD1, BD2, BD3 and putative domains- PD1, PD2, PD3, to Vitronectin, it is important to test their adhesion on the different electrode surfaces. For this experiment, three different kind of

electrode with Carbon surface, Cobalt surface and the Carbon electrodes coated with the binding enhancer were used. The experimental steps are briefly illustrated in Figure 45. As shown, the electrodes were first washed with 1X PBS buffer. After washing, the electrodes were allowed to dry for 15 minutes and equilibrated with the protein dialysis buffer containing 50 mM Tris, pH 7.5 and 300 mM NaCl. The excess of buffer was removed from the electrode surface before adhesion of the protein. Subsequently, 20 µl of purified Hsf domains were deposited on the different electrode surfaces and allowed to adhere to the surfaces for 60 minutes at room temperature. Please note that each electrode contains only one out of the six domains of Hsf and three electrodes were prepared for each Hsf domain, hence, a total of 21 (18 + 3) electrodes per electrode type were used for the experiment. The excess protein was washed off by rinsing the electrodes with 20 µL of 1X PBS buffer with two minutes incubation each time. Thereafter, the Ferrocene was added on the electrodes covering the entire surface for measuring the transfer of electrons. Please note that the Ferrocene was applied on one electrode at a time right before taking the measurements. The reference or the negative control electrodes were also treated in the same manner. The reference electrodes were equilibrated with only dialysis buffer and did not contain any proteins on them as a negative control for this experiment. Hence, the electrodes with only buffer were also prepared simultaneously in triplicates for each kind of electrode.

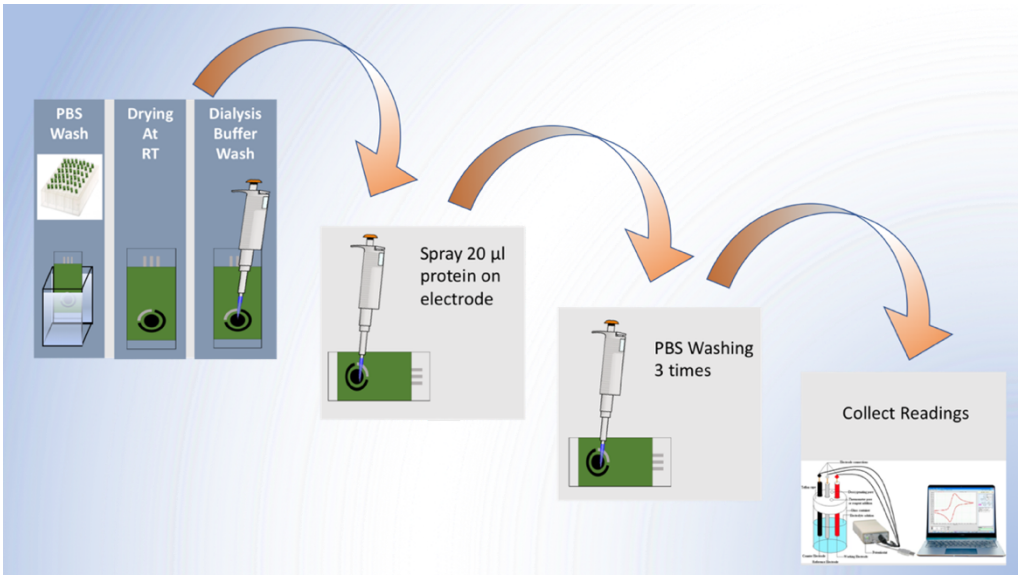


Figure 45 CV experiment design to test domain adhesion on electrodes

5.2.2 The cyclic voltammogram

In cyclic voltammetry, cycle is applied to each electrode with the following parameters: E_{begin} = −0.9 V, E_{vtx1} = −0.5 V, E_{vtx2} = 1.0 V, and scan rate = 0.050 V/s. The measurements were done in triplicate with two scans conducted during one measurement. Background currents, which were measured from the negative control, that is, the electrodes only coated with the dialysis buffer are subtracted out. The readings for the negative controls were also taken in triplicates with the same parameters as of the protein coated electrodes. The decrease in the current during Oxidation reaction is directly proportional to the amount of the proteins adhered on the electrode surface.

When plotted, the x-axis represents the applied potential (E) imposed on the system and the y-axis is the electron transfer as the resulting current (i) passed (Figure 46). A combination of Ferrocenium (Fc^+) and Ferrocene (Fc) was used as the redox active compound and the equilibrium established between Fc^+ and Fc is described by the Nernst equation

$$E = E^{0} + \frac{RT}{nF}\ln\frac{(\mathrm{Ox})}{(\mathrm{Red})} = E^{0} + 2.3026\frac{RT}{nF}\log_{10}\frac{(\mathrm{Ox})}{(\mathrm{Red})}$$

When a solution of Fc^{+} is scanned to negative potentials, Fc^{+} is reduced to Fc locally at the electrode, resulting in the measurement of a current and depletion of Fc+ at the electrode surface. The resulting cyclic voltammogram is presented in Figure 46. Crucially, the concentrations of Fc^{+} vs Fc relative to the distance from the surface of the electrode are dependent on the potential applied and how species move between the surface of the electrode and the bulk solution. These factors all contribute to the "duck"-shaped voltammograms.

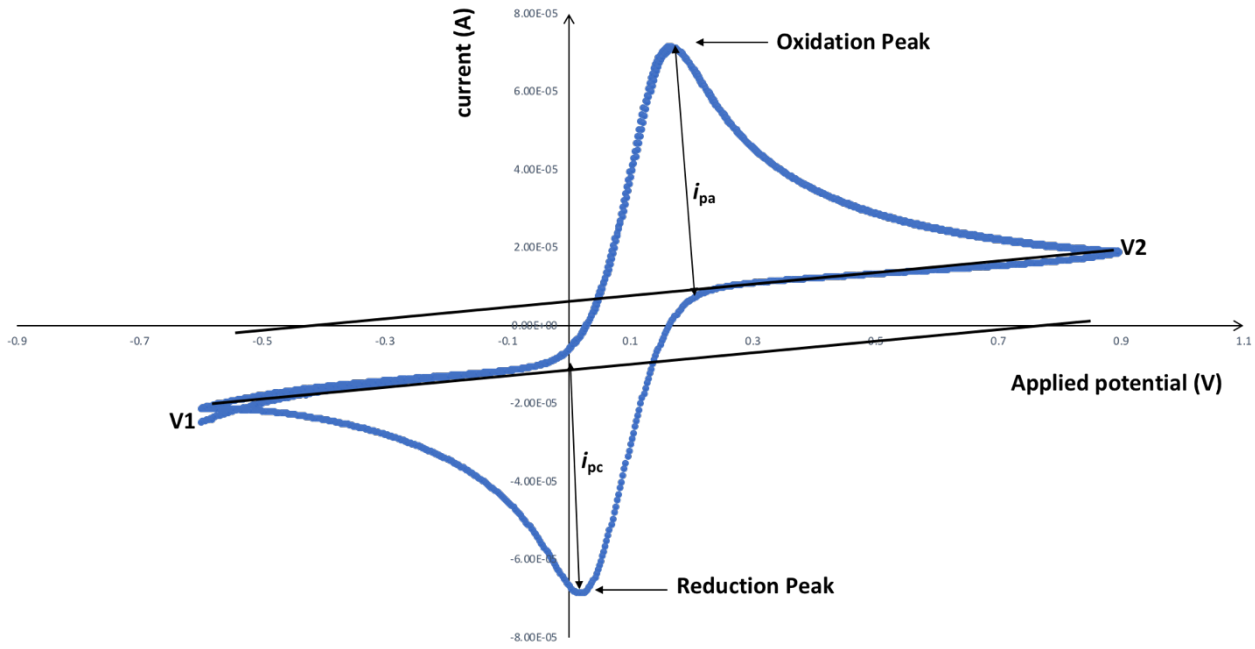


Figure 46 Cyclic Voltammogram

Typical cyclic voltammogram is plotted between applied potential (V) and applied current (A) where i_{pa} and i_{pc} show the peak cathodic and anodic current respectively for a reversible reaction.

5.2.3 Interaction of Hsf domains on Carbon and Cobalt surface Vs the polymer coated electrodes

The adhesive properties of the six domains of Hsf were tested on three different electrode-Carbon, Cobalt and the Carbon electrodes coated with the binding enhancer. Among the binding domains as per the literature, BD1 and BD2 domains of Hsf are known to be the most adhesive domains followed by BD3 domain showing least adhesiveness. Similar results were observed with cyclic voltammetry experiment where BD1 and BD2 showed the most adhesiveness on all three surfaces, followed by BD3 on all the three surfaces. However, the difference in the adhesive properties of each of the domains among the electrodes was also observed. Initial testing was done with 3 different concentrations, (1, 10 and 100 ng/µl). All three domains adhered more strongly in the presence of the binding enhancer.

In addition to the binding domains the objective of the study was to test the contribution of putative domains PD1, PD2 and PD3 towards the adhesion of the Hsf protein. Interestingly, the presence of putative domains is unique to the Hsf and their adhesive nature may establish that the Hsf is one of the most adhesive TAAs with six sites of adhesion. The experiment to test the adhesiveness of the putative domains was done in triplicates. Although binding efficiency of all the three putative domains is less than that of the binding domains, the experiment suggested that PD1 showed signs of adhesion on enhancer coated electrodes. However, PD2 and PD3 did not show any signs of adhesive behaviour on any of the electrodes as shown in Figure 47.

As the binding of the domains is insignificant on the carbon and cobalt electrodes and because of the high variability among these electrodes, I found them unsuitable for the further experiments. Hence, only the binding enhancer coated carbon electrodes were used for further experiments. The adhesion of the six Hsf domains (BD1, BD2, BD3, PD1, PD2, PD3) on the binding enhancer coated electrodes is shown in the Figure 47.

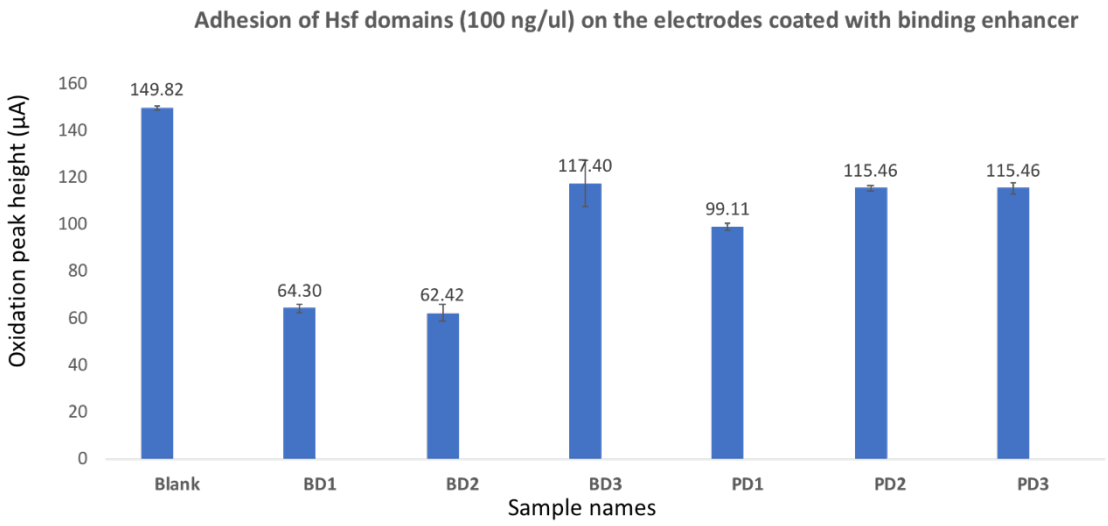


Figure 47 Adhesion of Hsf domains on binding enhancer coated electrodes.

The figure represents the oxidation peak heights on the Y-axis, whereas X-axis represents the domains. The data represented here as bar graph is evaluated in comparison with the negative control labelled as blank. The data shows that BD1 and BD2 are most adhesive domains followed by PD1. BD3, PD2 and PD3 domains shows least adhesion on the electrodes

5.3 Exploring the difference in adhesion of Hsf domains with Vitronectin

5.3.1 Layer-by-layer deposition of proteins to study the interaction of Hsf and Vitronectin

The first step of successful colonisation of the *H. influenzae* is the adherence to the host tissue, which is mediated by the interaction of Hsf with vitronectin (Singh et al., 2014). Here, I examined which of the six domains of Hsf were involved in adhesion to the Vitronectin.

Based on the results of section 5.2.3, it was determined that the electrodes coated with the binding enhancer were best suited for the binding experiments.

The binding experiments with cyclic voltammetry were conducted with layer-by-layer deposition of proteins on the coated electrodes. Preparation of electrodes was done as explained in the section 5.2.1. Briefly, electrodes were washed with PBS, followed by polymer coating and drying for one hour. Coated electrodes were then incubated with dialysis buffer with composition 50 mM Tris pH 7.5 and 300 mM NaCl for equilibration. Purified protein fraction of Hsf domains were deposited on the electrodes with a total of 2000 ng protein per electrode in the volume 20 µl of 100 ng/µl purified Hsf domains as standardised in earlier experiments. The proteins were allowed to dry for 30 minutes followed by 1X PBS washing, three times incubated for two minutes with each wash. After washing, molecular grade Vitronectin resuspended in the same dialysis buffer as of Hsf domains. A cartoon representation of the Hsf-Vitronectin interaction is shown in Figure 47 and the schematics of the binding protocol is illustrated in Figure 49.

A set of experiments were conducted to test the working concentration of Vitronectin with each of the six domains (BD1, BD2, BD3, PD1, PD2 and PD3). The concentrations of Vitronectin was determined by cyclic voltammetry where 20 µL Vitronectin in three different concentrations- 10, 50 and 100 ng/µl was deposited on the electrodes. This initial experiment suggested that 20 µL of 50 ng/µL is the right concentration to work with, which was allowed to bind to the enhancer coated electrodes for 30 minutes at room temperature. The experiments were done in triplicates for all the 100 ng/µL of Hsf domains. This step was again followed by washing with 1X PBS, three times for two minutes each to remove unbounded or excess of the Vitronectin. Washed electrodes were then used to collect Electrochemical Impedance Spectroscopy (EIS) readings. For EIS the resistance is measured to show the

deposition of the proteins. The resistance is directly proportional to the amount of the protein deposited on the electrode surface, hence, higher the resistance value more the protein bound on the electrode surface. Simultaneously, negative control of buffer and protein deposition were also prepared for a comparative analysis significant to ensure the binding of the proteins and comparing the transfer of electrons. Results of EIS experiment are shown in Figure 50 suggested that BD1 and BD2 binds more strongly to the Vitronectin than the other four domains.

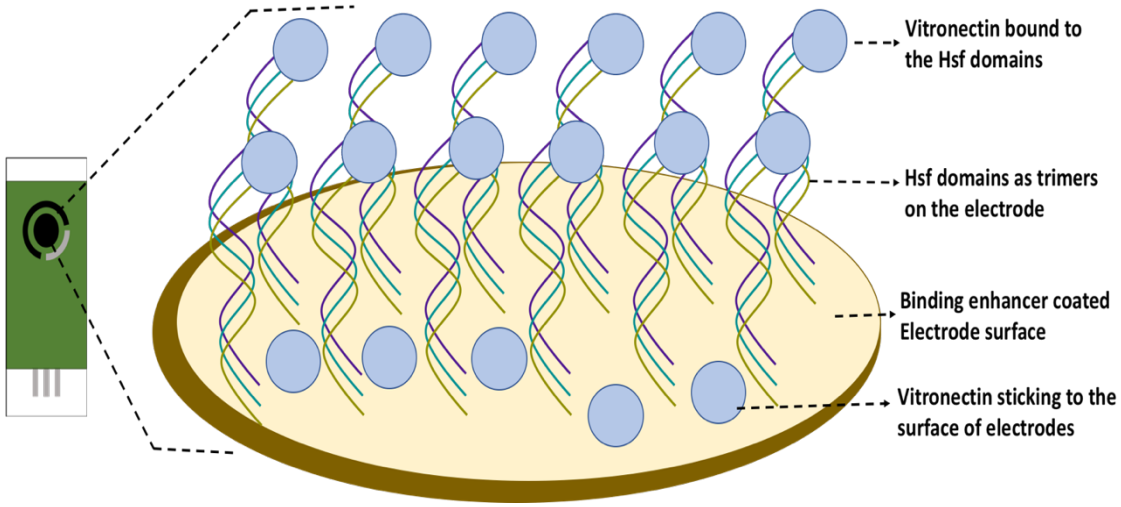


Figure 48 Cartoon Illustration of the Hsf Vn sandwich forming on the electrode

The image shows the magnifying view of the electrode surface coated with binding enhancer depicted as yellow surface. The bound Hsf domains are illustrated as trimers. Please mind that the Hsf domains illustrated in this figure are only in vertical orientation. In experimentation the Hsf domains may adhere in different orientations such as horizontal, Carboxy-terminal up or amino-terminal up. The blue circles represents the Vitronectin (Vn) bound to Hsf domains and the electrode surface.

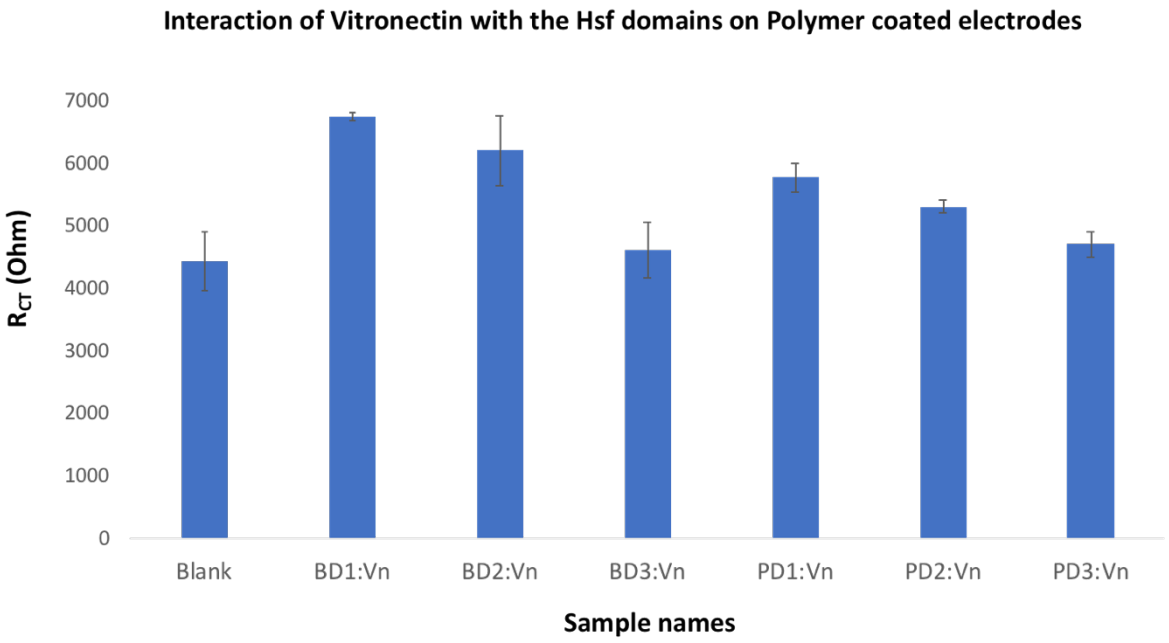


Figure 50 EIS data for the initial Hsf-Vn interaction test

For EIS data, peak heights of the resistance measured in Ohm is plotted on the Y-axis. The readings were calculated against the Buffer. The graph suggests that BD1 and BD2 highly adhesive to the Vitronectin than the other 4 domains.

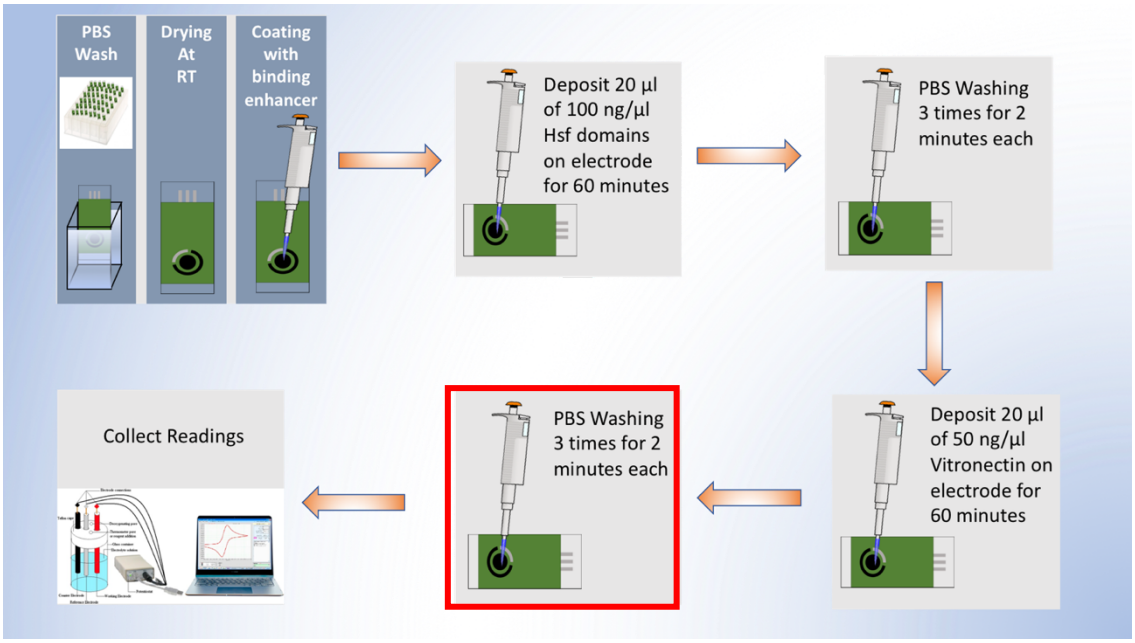


Figure 49 CV experiment design for Vitronectin binding

Schematic representation of the experiment design to test the interaction between Hsf domains and Vitronectin on the electrodes coated with binding enhancer. The steps highlighted with the red box represents the introduction of Vn binding to the previous protocol of electrode coating

5.3.2 Introduction of the BSA blocking step

The results of the previous experiment suggested the involvement of BD1 and BD2 domains of Hsf in interacting with Vitronectin. Intriguingly, a concern was raised against the initial EIS experiment of Hsf-Vitronectin binding, which was the adhesion of Vitronectin to the electrode surface. To address this concern, I added the BSA blocking step to the previous protocol.

Hence, a BSA blocking step post Hsf domain deposition on the electrodes, as shown in Figure 54. The electrodes were blocked with 0.5% of BSA incubated for half an hour. This step was followed three subsequent washes with the 1X PBS buffer to remove excess BSA. The BSA blocking step ensures that the electrode surface was unavailable for the Vitronectin to bind. This allows the Vitronectin only to bind the binding sites of the Hsf domains and not to the surface.

5.3.3 Effect of BSA on adhesion of domains to the electrodes

Another problem that was raised for developing a stable and working protocol for Hsf-Vitronectin binding was the efficient removal of excess of the bound BSA. The presence of BSA was affecting the binding of the Hsf domains, specifically observed for BD1 (Figure 51) . Hence, 1x PBS was not sufficient for removing the BSA from the electrodes. The excess of BSA was not only interfering with the binding of Hsf domains to the electrodes but also masking the Vitronectin binding sites of Hsf domains. This was observed with all the domains. The binding of all the domains on the electrodes was affected by the presence of BSA, as shown in Figure 52.

For removing the excess of the bound BSA efficiently, I optimised the washes in presence of different percentages of Tween 20 detergent as shown in Figure 53. This was followed by addition of 1% Tween to the 1 x PBS for 4 washes, and each wash was 2 minutes long as

illustrated in Figure 54. The results of the final experiments showed that out of the six Hsf domains, only BD1 binds to the Vitronectin (Figure 55).

Another concern here was the degradation of Hsf domains on the electrode surface due to the prolonged waiting time and exposure to different buffer solutions. Hence, I added vitronectin first on the electrodes followed by BSA blocking, Tween washing and Hsf domain adhesion.

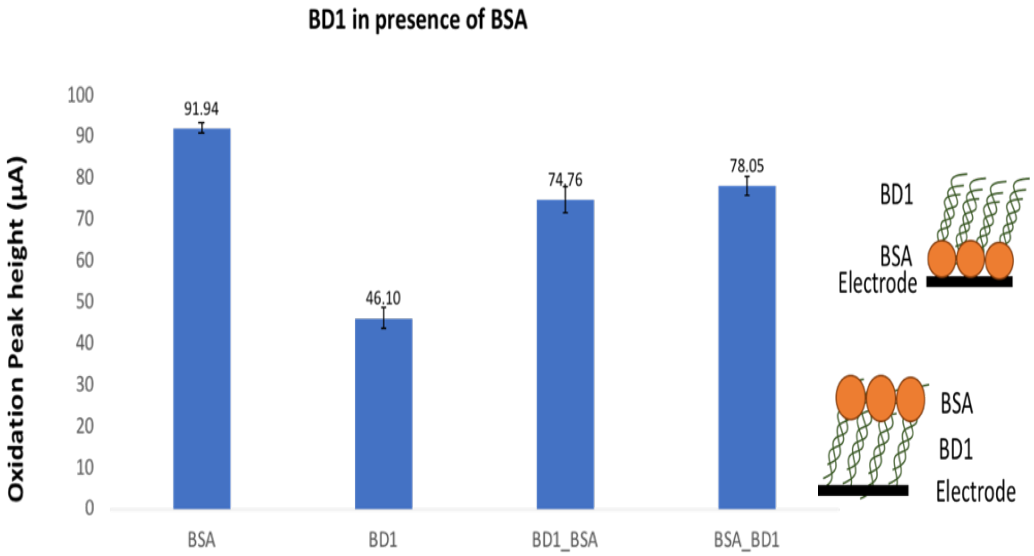


Figure 51 CV graph to show the effect of BSA on BD1 binding to the electrodes

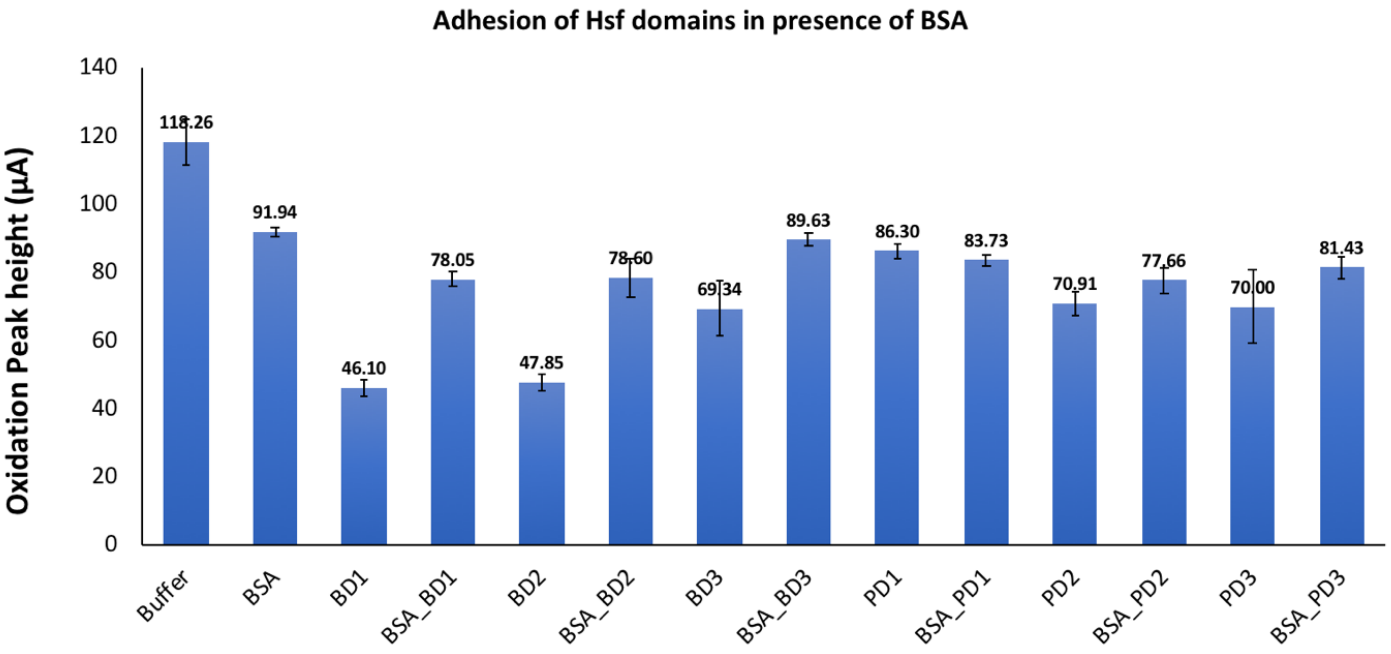


Figure 52 CV graph to show the effect of BSA on binding of all six Hsf domains to the electrodes

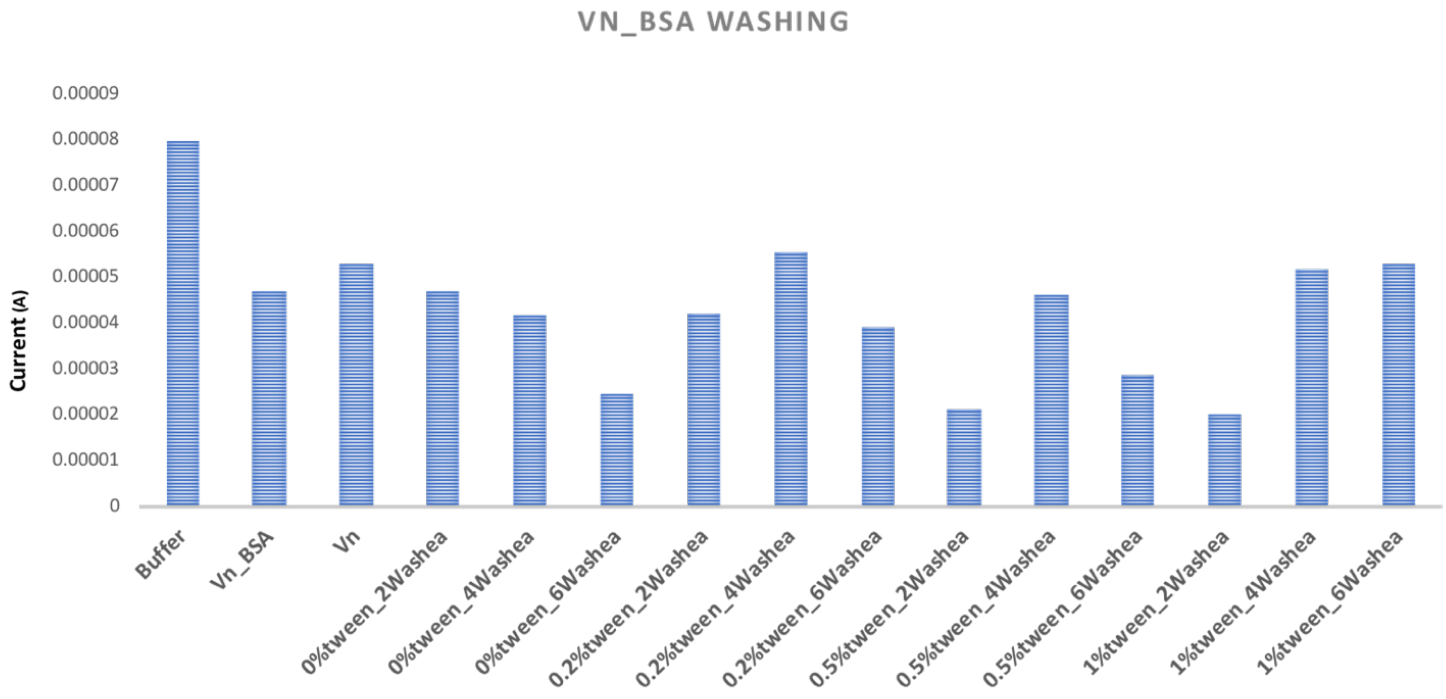


Figure 53 Optimisation of washes with different percentage of Tween 20

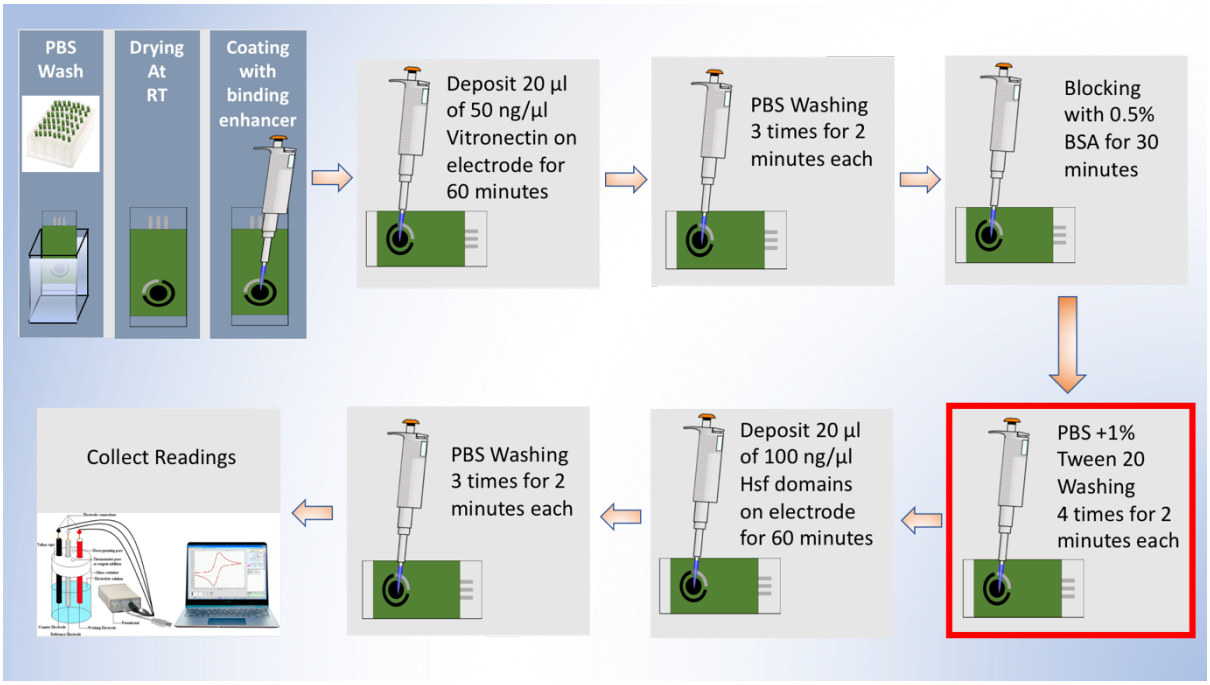
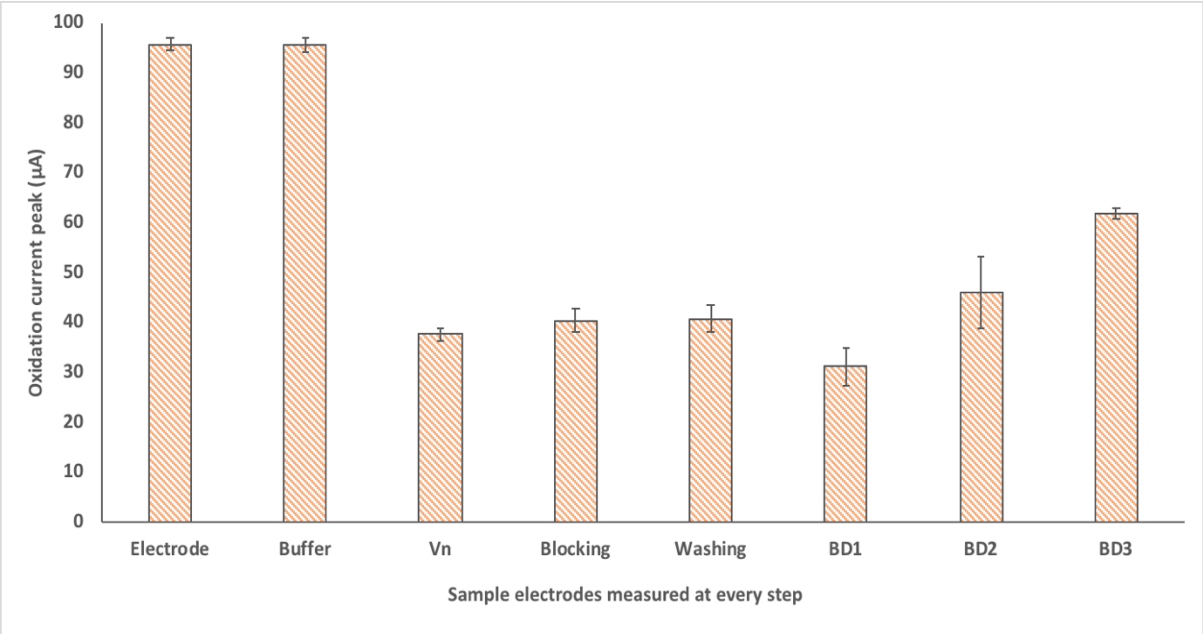


Figure 54 Schematic representation of the experiment design to test the interaction between Hsf domains and Vitronectin on the electrodes coated with binding enhancer.

a)



b)

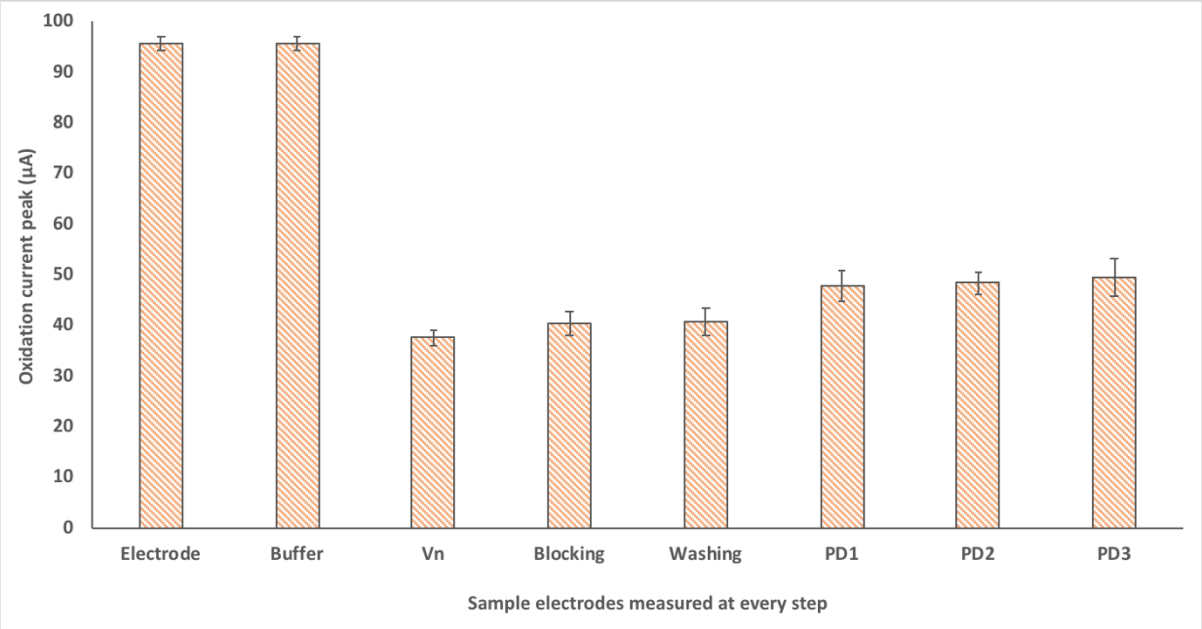


Figure 55 CV data for Hsf-Vitronectin binding after introduction of Tween washing step.

CV bar graph for the adhesion test of a) Binding domains BD1, BD2 and BD3, and b) Putative domains PD1, PD2 and PD3 with Vitronectin. Each bar is the oxidation peak height calculated for electrode sample at each step of the experiment. The first bar from the left is the oxidation peak height reading taken for the empty electrodes coated with binding enhancer. The second bar from the left is for the electrodes coated with the buffer followed by the Vitronectin (Vn), BSA blocking, Tween washing, followed by CV bar graphs of a) or b).

5.3.4 Vitronectin stability test on the electrode surfaces

For developing the diagnostic tools is to it was necessary to test the stability of Vitronectin on the electrodes. For which I conducted a day 1 vs day 2 test. Here, I coated the electrodes with binding enhancer and thereafter deposited 20 µl of 50 ng/µl Vitronectin on them. I incubated Vitronectin for 60 minutes on the electrodes at room temperature and washed off the excess with 1X PBS buffer. The change in oxidation peaks were investigated using Ferrocene. The resulted peak heights were analysed in comparison to peak heights of the reference electrodes named as blank. The blank electrodes only containing the vitronectin resuspension buffer or the dialysis buffer containing 50 mM Tris pH 7.5 and 300 mM NaCl.

The lower the oxidation reflects higher amount of protein adhered to the electrodes. Here, the reading for day one for blank and Vitronectin electrodes were 152.88 µA and 72.07 µA, respectively. This represents the successful binding of Vitronectin on the electrodes. The electrodes were then stored overnight in the fridge at 4 °C and on day 2, same electrodes were again given a 1-minute wash with 1X PBS and the readings of oxidation were measured again. As shown in the Figure 56 the oxidation peak height for blank was 140.05 µA in contrast

to 56.46 µA for Vitronectin electrodes. This suggests that Vitronectin is stable on the binding enhancer coated electrodes.

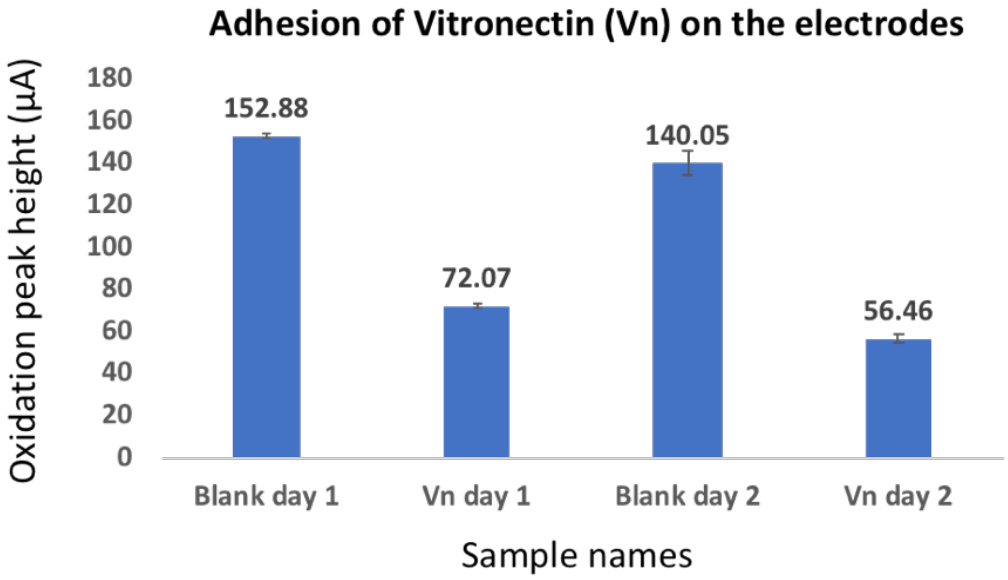


Figure 56 Vitronectin stability test on the electrode surface

5.4 Testing the structural stability of domains with Circular Dichroism

To determine the secondary structure content of Hsf domains, the far-UV CD spectra were collected on a Chirascan plus CD spectrophotometer. Nitrogen was flushed at the rate of 30 L/min during all the experiments using a peak scientific NM-30 generator. The spectra were collected from 260 nm to 200 nm wherever possible.

All of the six Hsf domains were purified in buffer with 50 mM Tris, pH 7.5 and 300 mM NaCl and were not stable in low salt. Hence, the major challenge was buffer exchange of the purified domains. As the domains are not stable and readily precipitated, different methods were applied for buffer exchange. At first, buffer exchange with 3 kDa centrifugal concentrators was carried out, in which 2 mL buffer was added to 2 mL protein and the 4 mL

buffer-protein solution was reduced back to 2 mL. This was to be carried out 5 times to ensure the complete buffer exchange of the protein. However, all the six domains were not stable during this process and completely precipitated in the first cycle. Second method was slow dialysis of the proteins against the buffer of interest at 4 °C for 30 minutes for complete buffer transfer. This method was also unsuccessful as the all of the six domains precipitated after a few minutes of dialysis. As last measure, I decided to try 100 fold dilution of the proteins before injecting the proteins in the cuvette for measurements. Hence, I concentrated the proteins to 10 M and diluted them to 1 µM with the CD suitable buffer. All the domains behaved differently to dilution of the proteins. BD1, BD2, BD3 and PD3 domains precipitated on dilution and hence were not suitable for measuring the CD spectra. However, PD1 and PD2 remained in soluble state and were used to collect far-UV CD spectra.

The far-UV CD spectra for Hsf PD1 and PD2 were taken at a concentration of 1 µM at 4 °C. The precipitates were separated from the protein fractions by centrifuging at 10,000 g for 5 minutes before collecting the spectra. The path length of the cuvette was 0.1 cm and the data was collected at 1 nm bandwidth. Two scanning spectra were averaged with the baseline subtracted, spanning 260 to 200 nm. As shown in Figure 57(a-d) , neither PD1 nor PD2 showed any signature for the presence of secondary structures. This could be attributed to the extremely adhesive nature of the Hsf domains leading to the aggregation of the proteins. Interestingly, 0.2% BSA, which was used as positive control, showed expected spectra under various conditions (in Milli-Q water, PBS and Tris).

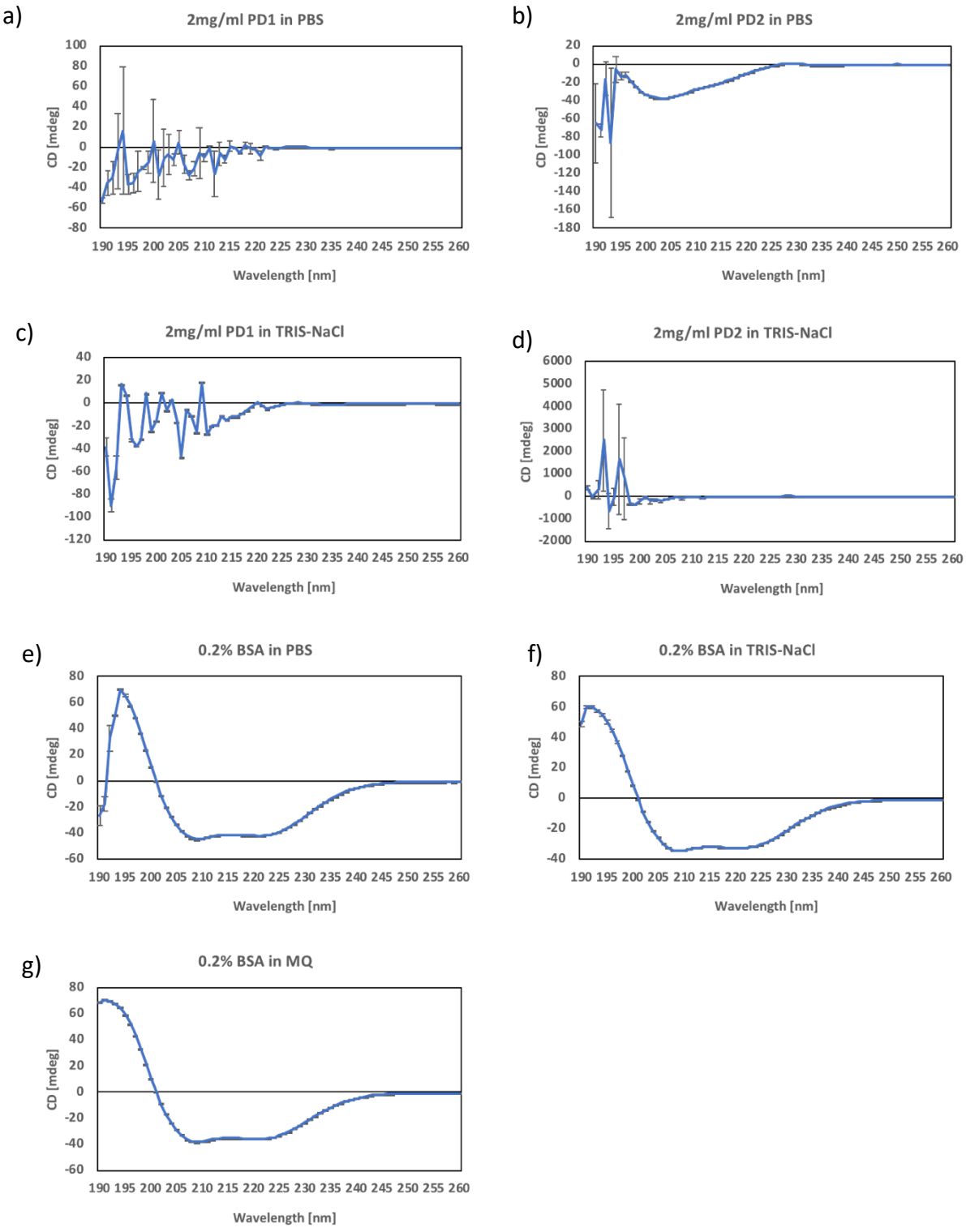


Figure 57 Far-UV-CD spectra of PD1 and PD2 domains of Hsf.

Far-Uv-CD spectra of PD1 in a) PBS and c) 10mM TRIS, pH 7.5, 150mM NaCl and PD2 in b) PBS and d) 10mM TRIS, pH 7.5, 150mM NaCl are shown. Far-UV-CD spectra of BSA in e) PBS and f) 10mM TRIS, pH 7.5, 150mM NaCl and g) MillQ water are shown as a positive control.

5.5 Discussion

The primary goal of the study was to look into the feasibility of developing electrochemical and microfluidics based fast, cost-effective diagnostic tools to detect the *H. influenzae* infection. In order to achieve this, I first tested the adhesiveness of Hsf domains (BD1, BD2, BD3, PD1, PD2 and PD3) on different electrode surfaces- Carbon, Cobalt and Binding enhancer coated. The CV experiment suggested that out of six domains only BD1 and BD2 domains shows adhesion and the best adhesion was observed on the carbon electrodes coated with binding enhancer. Hence, all of the experiments were further carried out on electrodes coated with binding enhancer.

This experiment was followed by testing the interaction between Hsf domains and Vitronectin, which was done by EIS. This experiment raised the concerns of vitronectin binding to the electrode surface which was answered by introduction of the BSA blocking step. Later, I realised that the BSA blocking interferes with the adhesion of Hsf domains and hence, I introduced prolonged washes with PBST buffer. For further experiments, I coated the electrodes with Vitronectin followed by deposition of Hsf domains. The results suggested that out of the six domains only BD1 domain interacts with Vitronectin.

At last, I attempted to inquire the secondary structure content of various Hsf domains. While it was impossible to record CD spectra of BD1, BD2, BD3 and PD3 due to their high propensity to precipitate, the far UV CD spectra of PD1 and PD2 domains didn't show any signature of secondary structure and therefore the secondary structure content of the latter couldn't be calculated. On contrary, BSA which was used as a positive control did show expected far UV CD spectra suggesting, the lack of secondary structures in PD1 and PD2 could be attributed to the possibility of forming soluble aggregates. For In-depth analysis of Hsf: Vn binding, more

robust experiments should be done. In my opinion, SPR, Mass-spec chemical cross-linking and ITC can provide reliable results.

5.6 Future Prospects of Electrochemical diagnostic devices

The utility of electrochemical or microfluidic devices is thoroughly documented in the literature. These approaches offer a once-in-a-lifetime opportunity for the creation of quick, cost-effective, and user-friendly diagnostic instruments. With these procedures in place, there is a high probability of detecting ligands, tiny compounds, and microorganisms quickly. Electrochemical detection technologies are critical for the quantitative investigation of various ligands. The combination of biomarkers and luminophores with disease-specific ligands can be used to create visual early-detection kits for a variety of illnesses. Another detection approach based on chemically synthesized analogues of known biological partners of bacterial proteins can also be developed.

Chapter 6 Overall Discussion, Conclusions and Future Directions

6.1 Overall discussion

Exposure of TAAs on the outer membrane of Gram-negative bacteria is crucial during the infection. The highly adhesive nature of TAAs is indispensable, not only for adhesion and colonisation of bacteria in the host, but also, to provide the ability to resist the immediate immune response. Therefore, it is Important to understand the mechanism of adhesion by these proteins and this knowledge can be used for making fast, user-friendly and cost-effective early diagnostic tools to detect invasive diseases.

My study, targets the concern of developing resistance in *H. influenzae* against the available medication. There is a severe need to develop fast diagnostic tools for early-stage detection of the invasive diseases caused by *H. influenzae*. In this line, my thesis work is step one towards the long achieving goal of developing new diagnostics. For this, I used structure and biophysical techniques to characterise one of the surfaces adhesins of *H. influenzae*, Haemophilus surface fibril (Hsf). Hsf not only plays a crucial role in adhesion of *H. influenzae* to the host epithelium but also, provides protection to the bacteria against the host immune response by binding to Vitronectin protein of the host.

The present work attempts to provide an insight into the binding mechanism of Hsf through its structural and biophysical characterization. There were two goals of this study, one was to develop a robust protocol for protein production and determine the atomic structure of the Hsf to test the hypothesis of Hsf forming a hairpin structure. The second objective was to have biophysical insight into the adhesion mechanism of Hsf.

6.1.1 Analysis of the Hsf Structure

After developing a robust protocol for generating a large amount of pure and concentrated protein the de novo structure of Hsf PD3BD3 domain was determined by X-ray

crystallography. The modest resolution of the structure and poor quality of the map in some areas made model building and refinement challenging. However, after many iterations of refinement a final structure was obtained with good refinement and model statistics. The functional unit of Hsf PD3BD3 is trimer and assumes a novel bend in the structure which has not been observed in any of the TAA family proteins so far. The overall organization of Hsf PD3BD3 is divided into multiple sub-domains conserved through the TAA family proteins. Both PD3 and BD3 domains consist of two TrpRings separated by neck domains. The PD3 and BD3 domains were found to have KG neck and Insertion neck, respectively.

Typically, TAAs obtain a straight stalk-like architecture expressed on the surface of bacteria. While the structure of Hsf PD3BD3 established the bend of 20.5° in Hsf, which is a novel structural element among TAA proteins, it also nullifies the hypothesis of a hair-pin structure. Ostentatiously, for Hsf to form a Hairpin structure, the angles of deflection among the domains have to be much greater. However, the bend in the structure is unique to Hsf, I suspect that it could be a result of multiple necks present in the structure. These necks consist of highly flexible loops, which may impart flexibility in Hsf and hence, could result in bend in the structure.

6.1.2 Insight into the binding mechanism of the Hsf protein

Alongside the structural work, I also studied the adhesive property of each of the Hsf domains using cyclic voltammetry (CV). The initial experiments by cyclic voltammetry provided some insight towards the adhesion of all the six domains of Hsf to different surfaces. I observed that regardless of the biochemical property of the electrode surface, Binding domain (BD) 1 was the most adhesive, followed by BD2 and then BD3, which agrees with the previous studies. Although the binding efficiency of the putative domains (PD) is less than that of the binding domains, the adhesive tendency among PD domains were observed to be the same.

Considering the presence of BD3 and all the three PDs is unique to Hsf, it is noteworthy that Hsf has four extra domains contributing to its adhesiveness, which is more than any other protein of the TAA family.

Followed by domain adhesion experiments on different surfaces, I tested the binding of Hsf domains with Vitronectin. In agreement with the literature BD1 was observed to bind Vitronectin with greater efficiency than the other five domains of Hsf, followed by BD2(Singh et al., 2014). The rest of the four domains which are unique to Hsf were found to have considerably lower binding with vitronectin. Considering that these experiments were performed with purified domains this study implicates possible contribution of individual domains towards overall affinity of Hsf towards Vitronectin.

6.2 Future direction for the project:

The work carried out in this thesis has furthered our understanding on the structure and function of Hsf protein. This covers previously hypothesised hairpin loop architecture and involvement of BD3 and all three PD domains in adhesion . However, in the absence of the full-length protein, our understanding on the overall architecture of Hsf is limited to say the least. The complete structure of the protein could be solved using higher resolution structural data obtained by X-ray crystallography or Cryo-EM and by using alfa-fold structure as reference. While the current study furnished qualitative evidence of adhesiveness of Hsf domains, further biochemical characterization using Surface plasmon resonance and/or Isothermal calorimetry and mass spectrometric chemical cross linking is required to determine the contribution of individual Hsf domains in adhesiveness and identifying the interface residues involved in the adhesion. While this would enhance our mechanistic insight on Hsf adhesive property, it could also be used to design new and improved diagnostic tools against *H. influenzae*.

6.2.1 Determining the overall structure of the full-length Hsf

From the work discussed in chapter 1, it is evident that the over-expression and purification of the full-length Hsf protein is extremely challenging due to its large size and low stability post expression. Although Hsf domains are also difficult to purify and have higher tendency to precipitates, the well-established protocol could readily be used by others in the community. Keeping this into consideration, further structural experiments with new constructs of different lengths could be considered. This would involve expression, purification and crystallization of Hsf variants containing two or three domains together, for example, Hsf-BD1PD1, Hsf-BD2PD2, Hsf-BD1PD1BD3 and so on. The information from different types of Hsf constructs would help to better deduce overall structure/model of the full length Hsf.

6.2.2 Elucidating the underlying mechanism of Hsf mediated adhesion

Higher resolution structural information about Hsf: Vitronectin interaction is imperative to study the adhesion mechanism of Hsf. This would require co-crystallising Hsf with vitronectin, and would require a large screen of co-crystallisation conditions for Hsf binding domains, especially with BD1 domain and Vitronectin. For X-ray crystallography, truncated version of Vitronectin based on the previous literature could also be used. Alternatively, Cryo-EM could be used for the larger complexes, removing the limitation of requiring the protein to crystallise(Kooger, Szwedziak, Böck, & Pilhofer, 2018).

In addition to the structural studies, biochemical studies are required to better decipher the underlying mechanism of Hsf mediated adhesion. While the binding studies in this work using cyclic voltammetry experiments has established the adhesiveness of individual Hsf domains

with Vitronectin, it remains to determine the binding affinities of Hsf domains with Vitronectin and to determine the residues mediating the interaction. To address the former, one could perform Surface Plasmon resonance (SPR) and/or Isothermal calorimetry with Hsf domains and Vitronectin and chemical crosslinking for the latter.

However, the most challenging part based on my work is to establish the native folding of the purified Hsf domains. For my thesis work, I tried CD and Mass photometry experiments to establish the native folding of the Hsf-BD1. However, due to its high adhesiveness it couldn't be performed successfully. The purified Hsf domains were eluted in buffer containing 50 mM Tris, pH 7.5 and 300 mM NaCl.

6.3 Final Conclusions & Significance of the Work to the TAA Field

The initial aims of this project were to try to better understand the structure of complete Hsf protein by providing additional structural information. I have achieved this goal through solving the crystal structure of a truncated Hsf segment containing the BD3 and PD3 domains joined together by a flexible neck. The Hsf-PD3BD3 structure showed the presence of a slight bend of 20.5° which is novel to Hsf and has not been seen in the previous structures of TAA proteins. This observation has led me to reject previous hypothesis of Hsf forming a hairpin loop. The presence of bend in the structure suggest the possibility of highly flexible nature of Hsf which may be essential for its interaction with its biological partner such as vitronectin in the host. Alongside, the work in the thesis also highlights the contribution of individual Hsf domains in binding to the Vitronectin.

While the current study provides invaluable information in the understanding of the structure and function of Hsf, there is still much work to be done on Hsf as highlighted above. Moreover, this structural information can now be used to help drive drug-design programs against pathogenic *H. influenzae*. Lastly, characterising the adhesiveness of the Hsf domains could prove crucial to develop early diagnosis tool for *H. influenzae* infection.

References

Abrusci, P., McDowell, M. A., Lea, S. M., & Johnson, S. (2014). Building a secreting nanomachine: a structural overview of the T3SS. *Current Opinion in Structural Biology*, *25*, 111–117.

Akopyan, K., Edgren, T., Wang-Edgren, H., Rosqvist, R., Fahlgren, A., Wolf-Watz, H., & Fallman, M. (2011). Translocation of surface-localized effectors in type III secretion. *Proceedings of the National Academy of Sciences*, *108*(4), 1639–1644.

Alcoforado Diniz, J., Liu, Y.-C., & Coulthurst, S. J. (2015). Molecular weaponry: diverse effectors delivered by the Type VI secretion system. *Cellular Microbiology*, *17*(12), 1742–1751. https://doi.org/https://doi.org/10.1111/cmi.12532

Amer, A. A. A., Gurung, J. M., Costa, T. R. D., Ruuth, K., Zavialov, A. V, Forsberg, Å., & Francis, M. S. (2016). YopN and TyeA Hydrophobic Contacts Required for Regulating Ysc-Yop Type III Secretion Activity by Yersinia pseudotuberculosis. *Frontiers in Cellular and Infection Microbiology*, *6*, 66. https://doi.org/10.3389/fcimb.2016.00066

Babic, A., Lindner, A. B., Vulic, M., Stewart, E. J., & Radman, M. (2008). Direct visualization of horizontal gene transfer. *Science*, *319*(5869), 1533–1536.

Bacon, D. J., Alm, R. A., Burr, D. H., Hu, L., Kopecko, D. J., Ewing, C. P., … Guerry, P. (2000). Involvement of a plasmid in virulence of Campylobacter jejuni 81-176. *Infection and Immunity*, *68*(8), 4384–4390. https://doi.org/10.1128/IAI.68.8.4384-4390.2000

Baker, H., & Bloom, W. L. (1948). Further Studies on the Gram Stain. *Journal of Bacteriology*, *56*(4), 387–390. https://doi.org/10.1128/jb.56.4.387-390.1948

Balakrishnan, L., Hughes, C., & Koronakis, V. (2001). Substrate-triggered recruitment of the TolC channel-tunnel during type I export of hemolysin by Escherichia coli.

Journal of Molecular Biology, *313*(3), 501–510.

Barnard, T. J., Dautin, N., Lukacik, P., Bernstein, H. D., & Buchanan, S. K. (2007). Autotransporter structure reveals intra-barrel cleavage followed by conformational changes. *Nature Structural & Molecular Biology*, *14*(12), 1214–1220.

BARTHOLOMEW, J. W., & MITTWER, T. (1952). The Gram stain. *Bacteriological Reviews*, *16*(1), 1–29. https://doi.org/10.1128/br.16.1.1-29.1952

Bassler, J., Hernandez Alvarez, B., Hartmann, M. D., & Lupas, A. N. (2015). A domain dictionary of trimeric autotransporter adhesins. *Int. J. Med. Microbiol.* https://doi.org/10.1016/j.ijmm.2014.12.010

Berg, B. van den, Clemons, W. M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C., & Rapoport, T. A. (2004). X-ray structure of a protein-conducting channel. *Nature*, *427*(6969), 36–44. https://doi.org/10.1038/nature02218

Berks, B. C., Palmer, T., & Sargent, F. (2005). Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. *Current Opinion in Microbiology*, *8*(2), 174–181.

Bernstein, H. D., Poritz, M. A., Strub, K., Hoben, P. J., Brenner, S., & Walter, P. (1989). Model for signal sequence recognition from amino-acid sequence of 54K subunit of signal recognition particle. *Nature*, *340*(6233), 482–486.

Bingle, L. E. H., Bailey, C. M., & Pallen, M. J. (2008). Type VI secretion: a beginner's guide. *Current Opinion in Microbiology*, *11*(1), 3–8.

Blocker, A., Komoriya, K., & Aizawa, S.-I. (2003a). Type III secretion systems and bacterial flagella: insights into their function from structural similarities. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(6), 3027–3030. https://doi.org/10.1073/pnas.0535335100

Blocker, A., Komoriya, K., & Aizawa, S. I. (2003b). Type III secretion systems and bacterial flagella: Insights into their function from structural similarities. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(6), 3027–3030. https://doi.org/10.1073/pnas.0535335100

Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y., & Attree, I. (2009). Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC Genomics*, *10*(1), 1–14.

Bröms, J. E., Ishikawa, T., Wai, S. N., & Sjöstedt, A. (2013). A functional VipA-VipB interaction is required for the type VI secretion system activity of Vibrio cholerae O1 strain A1552. *BMC Microbiology*, *13*(1), 96. https://doi.org/10.1186/1471-2180-13-96

Bundock, P, den Dulk-Ras, A., Beijersbergen, A., & Hooykaas, P. J. (1995). Trans-kingdom T-DNA transfer from Agrobacterium tumefaciens to Saccharomyces cerevisiae. *The EMBO Journal*, *14*(13), 3206–3214. https://doi.org/10.1002/j.1460-2075.1995.tb07323.x

Bundock, Paul, den Dulk-Ras, A., Beijersbergen, A., & Hooykaas, P. (1995). Trans-kingdom T-DNA transfer from Agrobacterium tumefaciens to Saccharomyces cerevisiae. *The EMBO Journal*, *14*(13), 3206–3214.

Burkinshaw, B. J., & Strynadka, N. C. J. (2014). Assembly and structure of the T3SS. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1843*(8), 1649–1663.

Buttner, D. (2012). Protein Export According to Schedule: Architecture. *Assembly, And*.

Casasanta, M. A., Yoo, C. C., Smith, H. B., Duncan, A. J., Cochrane, K., Varano, A.

C., … Slade, D. J. (2017). A chemical and biological toolbox for Type Vd secretion: Characterization of the phospholipase A1 autotransporter FplA from Fusobacterium nucleatum. *Journal of Biological Chemistry*, *292*(49), 20240–20254.

Charbonneau, M.-E., Berthiaume, F., & Mourez, M. (2006). Proteolytic processing is not essential for multiple functions of the Escherichia coli autotransporter adhesin involved in diffuse adherence (AIDA-I). *Journal of Bacteriology*, *188*(24), 8504–8512.

Chauhan, N., Hatlem, D., Orwick-Rydmark, M., Schneider, K., Floetenmeyer, M., van Rossum, B., … Linke, D. (2019). Insights into the autotransport process of a trimeric autotransporter, Yersinia Adhesin A (YadA). *Molecular Microbiology*, *111*(3), 844–862. https://doi.org/10.1111/mmi.14195

Cheng, Z., Jiang, Y., Mandon, E. C., & Gilmore, R. (2005). Identification of cytoplasmic residues of Sec61p involved in ribosome binding and cotranslational translocation. *The Journal of Cell Biology*, *168*(1), 67–77.

Chevalier, N., Moser, M., Koch, H.-G., Schimz, K.-L., Willery, E., Locht, C., … Müller, M. (2004). Membrane targeting of a bacterial virulence factor harbouring an extended signal peptide. *Microbial Physiology*, *8*(1), 7–18.

Chiba, K., Mori, H., & Ito, K. (2002). Roles of the C-terminal end of SecY in protein translocation and viability of Escherichia coli. *Journal of Bacteriology*, *184*(8), 2243–2250.

Cianciotto, N. P. (2005). Type II secretion: a protein secretion system for all seasons. *Trends in Microbiology*, *13*(12), 581–588.

Costa, T. R. D., Felisberto-Rodrigues, C., Meir, A., Prevost, M. S., Redzej, A., Trokter, M., & Waksman, G. (2015). Secretion systems in Gram-negative bacteria:

structural and mechanistic insights. *Nature Reviews Microbiology*, *13*(6), 343–359. https://doi.org/10.1038/nrmicro3456

da Mata Madeira, P. V., Zouhir, S., Basso, P., Neves, D., Laubier, A., Salacha, R., … Dessen, A. (2016). Structural basis of lipid targeting and destruction by the type V secretion system of Pseudomonas aeruginosa. *Journal of Molecular Biology*, *428*(9), 1790–1803.

Dautin, N., Barnard, T. J., Anderson, D. E., & Bernstein, H. D. (2007). Cleavage of a bacterial autotransporter by an evolutionarily convergent autocatalytic mechanism. *The EMBO Journal*, *26*(7), 1942–1952.

De Leeuw, E, Te Kaat, K., Moser, C., Menestrina, G., Demel, R., De Kruijff, B., … Sinning, I. (2000). Anionic phospholipids are involved in membrane association of FtsY and stimulate its GTPase activity. *The EMBO Journal*, *19*(4), 531–541.

de Leeuw, Erik, Poland, D., Mol, O., Sinning, I., ten Hagen-Jongman, C. M., Oudega, B., & Luirink, J. (1997). Membrane association of FtsY, the E. coli SRP receptor. *FEBS Letters*, *416*(3), 225–229.

Deane, J. E., Cordes, F. S., Roversi, P., Johnson, S., Kenjale, R., Picking, W. D., … Blocker, A. (2006). Expression, purification, crystallization and preliminary crystallographic analysis of MxiH, a subunit of the Shigella flexneri type III secretion system needle. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, *62*(3), 302–305.

Demers, J. P., Habenstein, B., Loquet, A., Kumar Vasa, S., Giller, K., Becker, S., … Sgourakis, N. G. (2014). High-resolution structure of the Shigella type-III secretion needle by solid-state NMR and cryo-electron microscopy. Nat Commun 5: 4976.

Depluverez, S., Devos, S., & Devreese, B. (2016). The Role of Bacterial Secretion Systems in the Virulence of Gram-Negative Airway Pathogens Associated with

Cystic Fibrosis . *Frontiers in Microbiology* . Retrieved from https://www.frontiersin.org/articles/10.3389/fmicb.2016.01336

Dohlich, K., Zumsteg, A. B., Goosmann, C., & Kolbe, M. (2014). A substrate-fusion protein is trapped inside the type III secretion system channel in Shigella flexneri. *PLoS Pathogens*, *10*(1), e1003881.

Durand, E., Nguyen, V. S., Zoued, A., Logger, L., Pehau-Arnaudet, G., Aschtgen, M.-S., … Dujeancourt, A. (2015). Biogenesis and structure of a type VI secretion membrane core complex. *Nature*, *523*(7562), 555–560.

Edgren, T., Forsberg, Å., Rosqvist, R., & Wolf-Watz, H. (2012). Type III secretion in Yersinia: injectisome or not? *PLoS Pathogens*, *8*(5), e1002669.

Elgrishi, N., Rountree, K. J., McCarthy, B. D., Rountree, E. S., Eisenhart, T. T., & Dempsey, J. L. (2018). A Practical Beginner's Guide to Cyclic Voltammetry. *Journal of Chemical Education*, *95*(2), 197–206. https://doi.org/10.1021/acs.jchemed.7b00361

Emsley, P., Charles, I. G., Fairweather, N. F., & Isaacs, N. W. (1996). Structure of Bordetella pertussis virulence factor P.69 pertactin. *Nature*, *381*(6577), 90–92. https://doi.org/10.1038/381090a0

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D Biological Crystallography*, *69*(7), 1204–1214. https://doi.org/10.1107/s0907444913000061

Fleming, P. J., Patel, D. S., Wu, E. L., Qi, Y., Yeom, M. S., Sousa, M. C., … Im, W. (2016). BamA POTRA domain interacts with a native lipid membrane surface. *Biophysical Journal*, *110*(12), 2698–2709.

Fronzes, R., Christie, P. J., & Waksman, G. (2009). The structural biology of type IV secretion systems. *Nature Reviews Microbiology*, *7*(10), 703–714.

Green, E. R., & Mecsas, J. (2016). Bacterial Secretion Systems: An Overview. *Microbiology Spectrum*, *4*(1). https://doi.org/10.1128/microbiolspec.VMBF-0012-2015

Greenfield, N. J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*, *1*(6), 2876–2890. https://doi.org/10.1038/nprot.2006.202

Grohmann, E., Muth, G., & Espinosa, M. (2003). Conjugative plasmid transfer in gram-positive bacteria. *Microbiology and Molecular Biology Reviews : MMBR*, *67*(2), 277–301, table of contents. https://doi.org/10.1128/MMBR.67.2.277-301.2003

Hallstrom, T., Trajkovska, E., Forsgren, A., & Riesbeck, K. (2014). Haemophilus influenzae Surface Fibrils Contribute to Serum Resistance by Interacting with Vitronectin. *The Journal of Immunology*, *177*(1), 430–436. https://doi.org/10.4049/jimmunol.177.1.430

Hernandez, R. E., Gallegos-Monterrosa, R., & Coulthurst, S. J. (2020). Type VI secretion system effector proteins: effective weapons for bacterial competitiveness. *Cellular Microbiology*, *22*(9), e13241.

Hobbs, M., & Mattick, J. S. (1993). Common components in the assembly of type 4 fimbriae, DNA transfer systems, filamentous phage and protein-secretion apparatus: a general system for the formation of surface-associated protein complexes. *Molecular Microbiology*, *10*(2), 233–243.

Hofreuter, D., Odenbreit, S., & Haas, R. (2001). Natural transformation competence in Helicobacter pylori is mediated by the basic components of a type IV secretion system. *Molecular Microbiology*, *41*(2), 379–391. https://doi.org/10.1046/j.1365-2958.2001.02502.x

Hofreuter, D., Odenbreit, S., Püls, J., Schwan, D., & Haas, R. (2000). Genetic

competence in Helicobacter pylori: mechanisms and biological implications. *Research in Microbiology*, *151*(6), 487–491. https://doi.org/10.1016/s0923-2508(00)00164-9

Hoiczyk, E., Roggenkamp, A., Reichenbecher, M., Lupas, A., & Heesemann, J. (2000). Structure and sequence analysis of Yersinia YadA and Moraxella UspAs reveal a novel class of adhesins. *The EMBO Journal*, *19*(22), 5989–5999.

Hoseini, S. S., & Sauer, M. G. (2015). Molecular cloning using polymerase chain reaction, an educational guide for cellular engineering. *Journal of Biological Engineering*, *9*, 2. https://doi.org/10.1186/1754-1611-9-2

Hovingh, E. S., van den Broek, B., & Jongerius, I. (2016). Hijacking complement regulatory proteins for bacterial immune evasion. *Frontiers in Microbiology*, *7*(DEC), 1–20. https://doi.org/10.3389/fmicb.2016.02004

Ishikawa, M., Shigemori, K., & Hori, K. (n.d.). Application of the adhesive bacterionanofiber AtaA to a novel microbial immobilization method for the production of indigo as a model chemical - Ishikawa - 2013 - Biotechnology and Bioengineering - Wiley Online Library. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/bit.25012/full

Jacob-Dubuisson, F., Guérin, J., Baelen, S., & Clantin, B. (2013). Two-partner secretion: as simple as it sounds? *Research in Microbiology*, *164*(6), 583–595.

Kabsch, W. (2010). *XDS*. *Acta Crystallographica Section D Biological Crystallography*, *66*(2), 125–132. https://doi.org/10.1107/S0907444909047337

Kajava, A. V, Cheng, N., Cleaver, R., Kessel, M., Simon, M. N., Willery, E., … Steven, A. C. (2001). Beta-helix model for the filamentous haemagglutinin adhesin of Bordetella pertussis and related bacterial secretory proteins. *Molecular Microbiology*, *42*(2), 279–292.

Kanonenberg, K., Schwarz, C. K. W., & Schmitt, L. (2013). Type I secretion systems–a story of appendices. *Research in Microbiology*, *164*(6), 596–604.

Keenan, R. J., Freymann, D. M., Stroud, R. M., & Walter, P. (2001). The signal recognition particle. *Annual Review of Biochemistry*, *70*(1), 755–775.

Kihara, A., Akiyama, Y., & Ito, K. (1995). FtsH is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit. *Proceedings of the National Academy of Sciences*, *92*(10), 4532–4536. https://doi.org/10.1073/pnas.92.10.4532

King, P. (2012). Haemophilus influenzae and the lung (Haemophilus and the lung). *Clinical and Translational Medicine*, *1*(1), 10. https://doi.org/10.1186/2001-1326-1-10

Koch, H.-G., & Müller, M. (2000). Dissecting the translocase and integrase functions of the Escherichia coli SecYEG translocon. *The Journal of Cell Biology*, *150*(3), 689–694.

Koiwai, K., Hartmann, M. D., Linke, D., Lupas, A. N., & Hori, K. (2016). Structural basis for toughness and flexibility in the C-terminal passenger domain of an acinetobacter trimeric autotransporter adhesin. *Journal of Biological Chemistry*, *291*(8), 3705–3724. https://doi.org/10.1074/jbc.M115.701698

Kooger, R., Szwedziak, P., Böck, D., & Pilhofer, M. (2018). CryoEM of bacterial secretion systems. *Current Opinion in Structural Biology*, *52*, 64–70. https://doi.org/10.1016/j.sbi.2018.08.007

Koretke, K. K., Szczesny, P., Gruber, M., & Lupas, A. N. (2006). Model structure of the prototypical non-fimbrial adhesin YadA of Yersinia enterocolitica. *Journal of Structural Biology*, *155*(2), 154–161. https://doi.org/10.1016/j.jsb.2006.03.012

Korotkov, K. V, Gonen, T., & Hol, W. G. J. (2011). Secretins: dynamic channels for

protein transport across membranes. *Trends in Biochemical Sciences*, *36*(8), 433–443.

Korotkov, K. V, Sandkvist, M., & Hol, W. G. J. (2012). The type II secretion system: biogenesis, molecular architecture and mechanism. *Nature Reviews Microbiology*, *10*(5), 336–351.

Kube, S., Kapitein, N., Zimniak, T., Herzog, F., Mogk, A., & Wendler, P. (2014). Structure of the VipA/B Type VI Secretion Complex Suggests a Contraction-State-Specific Recycling Mechanism. *Cell Reports*, *8*(1), 20–30. https://doi.org/https://doi.org/10.1016/j.celrep.2014.05.034

Kuhn, P., Weiche, B., Sturm, L., Sommer, E., Drepper, F., Warscheid, B., … Koch, H. (2011). The bacterial SRP receptor, SecA and the ribosome use overlapping binding sites on the SecY translocon. *Traffic*, *12*(5), 563–578.

Lee, M., Jun, S.-Y., Yoon, B.-Y., Song, S., Lee, K., & Ha, N.-C. (2012). Membrane fusion proteins of type I secretion system and tripartite efflux pumps share a binding motif for TolC in gram-negative bacteria. *PloS One*, *7*(7), e40460.

Leo, J. C., & Goldman, A. (2009). The immunoglobulin-binding Eib proteins from Escherichia coli are receptors for IgG Fc. *Molecular Immunology*, *46*(8–9), 1860–1866. https://doi.org/10.1016/j.molimm.2009.02.024

Leo, J. C., Grin, I., & Linke, D. (2012). Type V secretion: mechanism (s) of autotransport through the bacterial outer membrane. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1592), 1088–1101.

Leo, J. C., Lyskowski, A., Hattula, K., Hartmann, M. D., Schwarz, H., Butcher, S. J., … 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*(7), 1021–1030.

https://doi.org/10.1016/j.str.2011.03.021

Lessl, M., Balzer, D., Pansegrau, W., & Lanka, E. (1992). Sequence similarities between the RP4 Tra2 and the Ti VirB region strongly support the conjugation model for T-DNA transfer. *Journal of Biological Chemistry*, *267*(28), 20471–20480.

Lessl, M., & Lanka, E. (1994). Common mechanisms in bacterial conjugation and Ti-mediated T-DNA transfer to plant cells. *Cell*, *77*(3), 321–324.

Li, Y. G., Hu, B., & Christie, P. J. (2019). Biological and structural diversity of type IV secretion systems. *Microbiology Spectrum*, *7*(2), 2–7.

Linke, D., Riess, T., Autenrieth, I. B., Lupas, A., & Kempf, V. A. J. (2006). Trimeric autotransporter adhesins: variable structure, common function. *Trends in Microbiology*, *14*(6), 264–270.

Llosa, M., Gomis-Rüth, F. X., Coll, M., & de la Cruz Fd, F. (2002). Bacterial conjugation: a two-step mechanism for DNA transport. *Molecular Microbiology*, *45*(1), 1–8. https://doi.org/10.1046/j.1365-2958.2002.03014.x

Luirink, J., & Sinning, I. (2004). SRP-mediated protein targeting: structure and function revisited. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1694*(1), 17–35. https://doi.org/https://doi.org/10.1016/j.bbamcr.2004.03.013

Luirink, J., ten Hagen-Jongman, C. M., Van Der Weijden, C. C., Oudega, B., High, S., Dobberstein, B., & Kusters, R. (1994). An alternative protein targeting pathway in Escherichia coli: studies on the role of FtsY. *The EMBO Journal*, *13*(10), 2289–2296.

Lütcke, H. (1995). Signal recognition particle (SRP), a ubiquitous initiator of protein translocation. *European Journal of Biochemistry*, *228*(3), 531–550.

Lycklama a Nijeholt, J. A., de Keyzer, J., Prabudiansyah, I., & Driessen, A. J. M.

(2013). Characterization of the supporting role of SecE in protein translocation. *FEBS Letters*, *587*(18), 3083–3088. https://doi.org/10.1016/j.febslet.2013.07.046

Ma, J., Pan, Z., Huang, J., Sun, M., Lu, C., & Yao, H. (2017). The Hcp proteins fused with diverse extended-toxin domains represent a novel pattern of antibacterial effectors in type VI secretion systems. *Virulence*, *8*(7), 1189–1202. https://doi.org/10.1080/21505594.2017.1279374

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software . *Journal of Applied Crystallography*, *40*(4), 658–674. https://doi.org/10.1107/s0021889807021206

Meng, G., St. Geme, J. W., & Waksman, G. (2008). Repetitive Architecture of the Haemophilus influenzae Hia Trimeric Autotransporter. *Journal of Molecular Biology*, *384*(4), 824–836. https://doi.org/10.1016/j.jmb.2008.09.085

Milis, L., Morris, C. A., Sheehan, M. C., Charlesworth, J. A., & Pussell, B. A. (1993). Vitronectin-mediated inhibition of complement: evidence for different binding sites for C5b-7 and C9. *Clinical and Experimental Immunology*, *92*(1), 114–119. https://doi.org/10.1111/j.1365-2249.1993.tb05956.x

Morgan, J. L. W., Acheson, J. F., & Zimmer, J. (2017). Structure of a Type-1 Secretion System ABC Transporter. *Structure*, *25*(3), 522–529. https://doi.org/10.1016/j.str.2017.01.010

Mueller, E. A., & Levin, P. A. (2020). Bacterial Cell Wall Quality Control during Environmental Stress. *MBio*, *11*(5). https://doi.org/10.1128/mBio.02456-20

Mühlenkamp, M., Oberhettinger, P., Leo, J. C., Linke, D., & Schütz, M. S. (2015). Yersinia adhesin A (YadA) - Beauty & beast. *International Journal of Medical Microbiology*. https://doi.org/10.1016/j.ijmm.2014.12.008

Müller, M. (2005). Twin-arginine-specific protein export in Escherichia coli. *Research*

in Microbiology, *156*(2), 131–136.

Natale, P., Brüser, T., & Driessen, A. J. M. (2008). Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—Distinct translocases and mechanisms. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1778*(9), 1735–1756. https://doi.org/https://doi.org/10.1016/j.bbamem.2007.07.015

Nazarov, S., Schneider, J. P., Brackmann, M., Goldie, K. N., Stahlberg, H., & Basler, M. (2018). Cryo-EM reconstruction of Type VI secretion system baseplate and sheath distal end. *The EMBO Journal*, *37*(4), e97103. https://doi.org/https://doi.org/10.15252/embj.201797103

Nishiyama, K., Hanada, M., & Tokuda, H. (1994). Disruption of the gene encoding p12 (SecG) reveals the direct involvement and important function of SecG in the protein translocation of Escherichia coli at low temperature. *The EMBO Journal*, *13*(14), 3272–3277.

Okaro, U., Green, R., Mohapatra, S., & Anderson, B. (2019). The trimeric autotransporter adhesin BadA is required for in vitro biofilm formation by Bartonella henselae. *NPJ Biofilms and Microbiomes*, *5*(1), 10. https://doi.org/10.1038/s41522-019-0083-8

Papanikou, E., Karamanou, S., & Economou, A. (2007). Bacterial protein secretion through the translocase nanomachine. *Nature Reviews Microbiology*, *5*(11), 839–851.

Park, J., Throop, A. L., & LaBaer, J. (2015). Site-specific recombinational cloning using gateway and in-fusion cloning schemes. *Current Protocols in Molecular Biology*, *110*, 3.20.1-3.20.23. https://doi.org/10.1002/0471142727.mb0320s110

Picking, W. L., Nishioka, H., Hearn, P. D., Baxter, M. A., Harrington, A. T., Blocker, A., & Picking, W. D. (2005). IpaD of Shigella flexneri is independently required for

regulation of Ipa protein secretion and efficient insertion of IpaB and IpaC into host membranes. *Infection and Immunity*, *73*(3), 1432–1440.

Pimenta, A. L., Young, J., Holland, I. B., & Blight, M. A. (1999). Antibody analysis of the localisation, expression and stability of HlyD, the MFP component of the E. coli haemolysin translocator. *Molecular and General Genetics MGG*, *261*(1), 122–132.

Pissaridou, P., Allsopp, L. P., Wettstadt, S., Howard, S. A., Mavridou, D. A. I., & Filloux, A. (2018). The Pseudomonas aeruginosa T6SS-VgrG1b spike is topped by a PAAR protein eliciting DNA damage to bacterial competitors. *Proceedings of the National Academy of Sciences*, *115*(49), 12519–12524. https://doi.org/10.1073/pnas.1814181115

Pohlschröder, M., Hartmann, E., Hand, N. J., Dilks, K., & Haddad, A. (2005). Diversity and evolution of protein translocation. *Annu. Rev. Microbiol.*, *59*, 91–111.

Price, S. B., Cowan, C., Perry, R. D., & Straley, S. C. (1991). The Yersinia pestis V antigen is a regulatory protein necessary for Ca2 (+)-dependent growth and maximal expression of low-Ca2+ response virulence genes. *Journal of Bacteriology*, *173*(8), 2649–2657.

Qin, W., Wang, L., & Lei, L. (2015). New findings on the function and potential applications of the trimeric autotransporter adhesin. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*. https://doi.org/10.1007/s10482-015-0477-4

Radics, J., Königsmaier, L., & Marlovits, T. C. (2014). Structure of a pathogenic type 3 secretion system in action. *Nature Structural & Molecular Biology*, *21*(1), 82–87.

Rampelotto, P. H. (2013). Extremophiles and extreme environments. *Life (Basel,*

Switzerland), *3*(3), 482–485. https://doi.org/10.3390/life3030482

Randall, L. L., & Hardy, S. J. S. (2002). SecB, one small chaperone in the complex milieu of the cell. *Cellular and Molecular Life Sciences CMLS*, *59*(10), 1617–1623.

Records, A. R. (2011). The type VI secretion system: a multipurpose delivery system with a phage-like machinery. *Molecular Plant-Microbe Interactions*, *24*(7), 751–757.

Rego, A. T., Chandran, V., & Waksman, G. (2010). Two-step and one-step secretion mechanisms in Gram-negative bacteria: contrasting the type IV secretion system and the chaperone-usher pathway of pilus biogenesis. *Biochemical Journal*, *425*(3), 475–488.

Robinson, C., & Bolhuis, A. (2004). Tat-dependent protein targeting in prokaryotes and chloroplasts. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1694*(1–3), 135–147.

Ruiz-Perez, F., & Nataro, J. P. (2014). Bacterial serine proteases secreted by the autotransporter pathway: classification, specificity, and role in virulence. *Cellular and Molecular Life Sciences : CMLS*, *71*(5), 745–770. https://doi.org/10.1007/s00018-013-1355-8

Salacha, R., Kovačić, F., Brochier-Armanet, C., Wilhelm, S., Tommassen, J., Filloux, A., … Bleves, S. (2010). The Pseudomonas aeruginosa patatin-like protein PlpD is the archetype of a novel Type V secretion system. *Environmental Microbiology*, *12*(6), 1498–1512.

Sandt, C. H., & Hill, C. W. (2000). Four different genes responsible for nonimmune immunoglobulin-binding activities within a single strain of Escherichia coli. *Infection and Immunity*, *68*(4), 2205–2214. https://doi.org/10.1128/IAI.68.4.2205-2214.2000

Sargent, F., Stanley, N. R., Berks, B. C., & Palmer, T. (1999). Sec-independent protein translocation in Escherichia coli: a distinct and pivotal role for the TatB protein. *Journal of Biological Chemistry*, *274*(51), 36073–36082.

Satoh, Y., Matsumoto, G., Mori, H., & Ito, K. (2003). Nearest Neighbor Analysis of the SecYEG Complex. 1. Identification of a SecY− SecG Interface. *Biochemistry*, *42*(24), 7434–7441.

Satoh, Y., Mori, H., & Ito, K. (2003). Nearest Neighbor Analysis of the SecYEG Complex. 2. Identification of a SecY− SecE Cytosolic Interface. *Biochemistry*, *42*(24), 7442–7447.

Sauvonnet, N., Vignon, G., Pugsley, A. P., & Gounon, P. (2000). Pilus formation and protein secretion by the same machinery in Escherichia coli. *The EMBO Journal*, *19*(10), 2221–2228.

Schatz, P. J., & Beckwith, J. (1990). Genetic analysis of protein export in Escherichia coli. *Annual Review of Genetics*, *24*(1), 215–248.

Schneewind, O., & Missiakas, D. M. (2012). Protein secretion and surface display in Gram-positive bacteria. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1592), 1123–1139. https://doi.org/10.1098/rstb.2011.0210

Shevchik, V. E., Robert-Baudouy, J., & Condemine, G. (1997). Specific interaction between OutD, an Erwinia chrysanthemi outer membrane protein of the general secretory pathway, and secreted proteins. *The EMBO Journal*, *16*(11), 3007–3016.

Sikdar, R., Peterson, J. H., Anderson, D. E., & Bernstein, H. D. (2017). Folding of a bacterial integral outer membrane protein is initiated in the periplasm. *Nature Communications*, *8*(1), 1309. https://doi.org/10.1038/s41467-017-01246-4

Singh, B., Su, Y. C., Al-Jubair, T., Mukherjee, O., Hallström, T., Mörgelin, M., … Riesbeck, K. (2014). A fine-tuned interaction between trimeric autotransporter haemophilus surface fibrils and vitronectin leads to serum resistance and adherence to respiratory epithelial cells. *Infection and Immunity*, *82*(6), 2378–2389. https://doi.org/10.1128/IAI.01636-13

Singh, B., Su, Y., & Riesbeck, K. (2010). Vitronectin in bacterial pathogenesis: a host protein used in complement escape and cellular invasion. *Molecular Microbiology*, *78*(3), 545–560. https://doi.org/10.1111/j.1365-2958.2010.07373.x

Slack, M. P. E., Cripps, A. W., Grimwood, K., Mackenzie, G. A., & Ulanova, M. (2021). Invasive Haemophilus influenzae Infections after 3 Decades of Hib Protein Conjugate Vaccine Use. *Clinical Microbiology Reviews*, *34*(3), e0002821. https://doi.org/10.1128/CMR.00028-21

Solomon, R., Zhang, W., McCrann, G., Bliska, J. B., & Viboud, G. I. (2015). Random mutagenesis identifies a C-terminal region of YopD important for Yersinia type III secretion function. *PloS One*, *10*(3), e0120471. https://doi.org/10.1371/journal.pone.0120471

St. Geme, J. W., Cutter, D., Barenkamp, S. J., St, J. W., Iii, G., Cutter, D., & Barenkamp, S. J. (1996). Characterization of the Genetic Locus Encoding Haemophilus influenzae Type b Surface Fibrils. *Journal of Bacteriology*, *178*(21), 6281–6287. https://doi.org/10.1128/jb.178.21.6281-6287.1996

Symmons, M. F., Bokma, E., Koronakis, E., Hughes, C., & Koronakis, V. (2009). The assembled structure of a complete tripartite bacterial multidrug efflux pump. *Proceedings of the National Academy of Sciences*, *106*(17), 7173–7178.

Tejeda-Dominguez, F., Huerta-Cantillo, J., Chavez-Dueñas, L., & Navarro-Garcia, F. (2017). A Novel Mechanism for Protein Delivery by the Type 3 Secretion System

for Extracellularly Secreted Proteins. *MBio*, *8*(2). https://doi.org/10.1128/mBio.00184-17

Thomas, S., Holland, I. B., & Schmitt, L. (2014). The Type 1 secretion pathway - the hemolysin system and beyond. *Biochimica et Biophysica Acta*, *1843*(8), 1629–1641. https://doi.org/10.1016/j.bbamcr.2013.09.017

van der Sluis, E. O., Nouwen, N., & Driessen, A. J. M. (2002). SecY–SecY and SecY–SecG contacts revealed by site-specific crosslinking. *FEBS Letters*, *527*(1–3), 159–165.

Voulhoux, R., Ball, G., Ize, B., Vasil, M. L., Lazdunski, A., Wu, L.-F., & Filloux, A. (2001). Involvement of the twin-arginine translocation system in protein secretion via the type II pathway. *The EMBO Journal*, *20*(23), 6735–6741.

Waksman, G. (2019). From conjugation to T4S systems in Gram-negative bacteria: a mechanistic biology perspective. *EMBO Reports*, *20*(2), e47012.

Whitney, J. C., Beck, C. M., Goo, Y. A., Russell, A. B., Harding, B. N., De Leon, J. A., … Mougous, J. D. (2014). Genetically distinct pathways guide effector export through the type VI secretion system. *Molecular Microbiology*, *92*(3), 529–542. https://doi.org/https://doi.org/10.1111/mmi.12571

Wollmann, P., Zeth, K., Lupas, A. N., & Linke, D. (2006). Purification of the YadA membrane anchor for secondary structure analysis and crystallization. *International Journal of Biological Macromolecules*, *39*(1–3), 3–9. https://doi.org/10.1016/j.ijbiomac.2005.11.009

Wright, J., Thomsen, M., Kolodziejczyk, R., Ridley, J., Sinclair, J., Carrington, G., … Goldman, A. (2017). The crystal structure of PD1, a Haemophilus surface fibril domain. *Acta Crystallographica Section F Structural Biology Communications*, *73*(2), 101–108. https://doi.org/10.1107/s2053230x17001406

Yen, Y. T., Kostakioti, M., Henderson, I. R., & Stathopoulos, C. (2008). Common themes and variations in serine protease autotransporters. *Trends in Microbiology*, *16*(8), 370–379. https://doi.org/10.1016/j.tim.2008.05.003

Zhang, X. Y., Brunet, Y. R., Logger, L., Douzi, B., Cambillau, C., Journet, L., & Cascales, E. (2013). Dissection of the TssB-TssC Interface during Type VI Secretion Sheath Complex Formation. *PLOS ONE*, *8*(11), e81074. Retrieved from https://doi.org/10.1371/journal.pone.0081074