

Structure and Host-Receptor Recognition  
Studies of Gram-Negative Bacterial Fimbriae  
Assembled *via* the Chaperone/Usher  
Pathway

Saumendra Prasad Roy

*Faculty of Natural Resources and Agricultural Sciences*

*Department of Chemistry and Biotechnology*

Uppsala

Doctoral Thesis  
Swedish University of Agricultural Sciences  
Uppsala 2015

Acta Universitatis agriculturae Sueciae

2015:99

Cover: Fimbria-receptor binding  
(Design: S. Roy and A. Zavialov)

ISSN 1652-6880

ISBN (print version) 978-91-576-8396-0

ISBN (electronic version) 978-91-576-8397-7

© 2015 Saumendra Prasad Roy, Uppsala

Print: SLU Service/Repro, Uppsala 2015

## Structure and host-receptor recognition studies of Gram-negative bacterial fimbriae assembled *via* the chaperone/usher pathway

### Abstract

Enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), and *Yersinia enterocolitica* (YE) are the principle etiological agents of acute and persistent diarrhea worldwide including resource-poor and industrialized regions. Adhesion of these pathogens to host receptor is the first important step to deliver toxins and virulent factors to initiate pathogenicity. Coli surface antigen (CS6) of ETEC, aggregative adherence factor-1 (AAF-1) of EAEC, and mucoid *Yersinia* fimbriae (Myf) of YE are the widely expressed colonization factors (fimbriae or pili) that mediate bacterial adhesion to the small intestinal epithelium. CS6 consists of CssA and CssB, AAF-1 consists of AggA and AggB, and Myf is made of MyfA, which are secreted and assembled on the cell surface via the chaperone/usher pathway. Using X-ray crystallography, we present atomic resolution insight into the fibre forming subunits and show that subunits of these organelles are assembled into a linear polymer by donor strand complementation. Myf is a homo-polymer, AAF-1 is a polymer of AggA subunits decorated by a single copy of the AggB subunit at the tip of the fiber, and CS6 is hetero-polymer of CssA and CssB subunits that are assembled alternately into a single fiber. Using spectroscopic and biochemical studies, we demonstrate that all subunits of these three fimbriae bind to receptors on intestinal epithelial cells, establishing multipoint attachment. The crystal structure of a MyfA-galactose complex reveals the receptor binding site and responsible residues. Structure guided mutagenesis of AggA shows that it binds to fibronectin via a novel electrostatic binding mechanism. Based on our structural and biochemical studies we conclude that CS6, AAF-1, and Myf are adhesive fimbriae of a novel type: they can act as poly-adhesins with rich binding properties. These findings steer the research field forward by revealing insight into receptor-fimbria interactions. This opens a window to search for receptor inhibitor molecules to decrease the infantile death rate and associated malnutrition.

*Keywords:* Gram-negative, bacteria, fimbriae, assembly, chaperone/usher, subunit, binding, receptor, X-ray crystallography, structure.

*Author's address:* Saumendra Prasad Roy, Department of Chemistry and Biotechnology, SLU, P.O. Box 7015, SE-750 07 Uppsala, Sweden

*E-mail:* [saumendra.prasad.roy@slu.se](mailto:saumendra.prasad.roy@slu.se)

## Dedication

*To my loving parents.....* মা ও বাবা



# Contents

<b>List of publications</b>	<b>7</b>
<b>Abbreviations</b>	<b>9</b>
<b>1 Background</b>	<b>11</b>
1.1 General introduction	11
1.2 The chaperone/usher pathway	13
1.3 Fimbrial subunit and architecture of a fibre	13
1.4 Diarrheagenic bacteria and fimbriae	15
1.5 Aim and outline of this thesis	17
<b>2 Methods</b>	<b>19</b>
2.1 Protein engineering, expression, and purification	19
2.2 X-ray crystallography	20
2.2.1 <i>Introduction</i>	20
2.2.2 <i>SeMet incorporation for isomorphous replacement</i>	22
2.2.3 <i>Determining the structures of CS6 and the MyfA-Gal complex</i>	22
2.3 Determination of binding constants	23
2.3.1 <i>Surface plasmon resonance</i>	23
2.3.2 <i>Tryptophan fluorescence spectrophotometry</i>	24
<b>3 Results and discussion</b>	<b>27</b>
3.1 CS6 fimbriae (Paper I)	27
3.1.1 <i>Introduction</i>	27
3.1.2 <i>Modelling and design of self-complemented subunits</i>	27
3.1.3 <i>Hetero constructs are more stable than homo constructs</i>	29
3.1.4 <i>Atomic resolution structure of C<sub>ss</sub>AdsB and C<sub>ss</sub>BdsA</i>	30
3.1.5 <i>Model of CS6 fibre</i>	32
3.1.6 <i>CS6 subunits recognize host receptor(s)</i>	33
3.1.7 <i>Discussion</i>	36
3.2 AAF-1 fimbriae (Paper II)	37
3.2.1 <i>Introduction</i>	37
3.2.2 <i>Atomic model of AAF-1</i>	37
3.2.3 <i>AAF-1 recognizes fibronectin as a receptor</i>	38
3.2.4 <i>The AggAdsA-fibronectin interaction is ionic not hydrophobic</i>	39
3.2.5 <i>Discussion</i>	41

3.3 Myf fimbriae (Paper III)	43
3.3.1 Introduction	43
3.3.2 MyfA binds galactose and lactose	43
3.3.3 MyfA binds isoglobotriose	44
3.3.4 Atomic structure of MyfA-Gal complex reveals the binding site	44
3.3.5 Discussion	47
<b>4 Concluding remarks and future perspective</b>	<b>49</b>
<b>References</b>	<b>51</b>
<b>Acknowledgements</b>	<b>57</b>

## List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

- I Roy SP\*, Rahman MM, Yu XD\*, Tuittila M\*, Knight SD, Zavialov AV (2012). Crystal structure of enterotoxigenic *Escherichia coli* colonization factor CS6 reveals a novel type of functional assembly. *Mol Microbiol.* 86(5): 1100-15. doi: 10.1111/mmi.12044. Epub 2012 Oct 10.
- II Berry AA, Yang Y, Pakharukova N, Garnett JA, Lee WC, Cota E, Marchant J, Roy S, Tuittila M, Liu B, Inman KG, Ruiz-Perez F, Mandomando I, Nataro JP, Zavialov AV, Matthews S (2014). Structural insight into host recognition by aggregative adherence fimbriae of enteroaggregative *Escherichia coli*. *PLoS Pathog.* 18; 10(9): e1004404. Doi: 10.1371/journal.ppat.1004404. eCollection 2014.
- III Saumendra Roy, Jerry Ståhlberg, Torleif Hård, and Anton V. Zavialov. Myf antigen of *Yersinia enterocolitica* recognizes host receptor through galactosyl residue in glycosphingolipid. (manuscript)

Papers I is reproduced with the permission of the publisher. \* denotes shared first authorship.

The contributions of Saumendra Prasad Roy to the papers included in this thesis were as follows:

- I Performed most part of the planning, experimental work (protein expression and purification, crystallography, and biochemical characterization). Also took part along with co-authors in data analysis, data interpretation, and writing of manuscript.
- II Performed the planning, experimental work, data analysis, and data interpretation of biochemical characterization through SPR.
- III Performed the planning, experimental work (protein expression and purification, crystallography, and biochemical characterization). Also took part along with co-authors in data analysis and data interpretation. Was responsible for manuscript writing.

## Abbreviations

3D	Three-dimensional
Ail	Attachment-invasion locus
C	Chaperone
CS	Coli surface
CU	Chaperone/usher
D	Aspartic acid
DS	Donor strand
DSC	Donor strand complemented
<i>E. coli</i>	<i>Escherichia coli</i>
EAEC	Enteroaggregative <i>Escherichia coli</i>
ECM	Extra cellular matrix
ETEC	Enterotoxigenic <i>Escherichia coli</i>
F	Phenylalanine
FGL	F <sub>1</sub> G <sub>1</sub> long loop
Gal	Galactose
G <sub>d</sub>	Donor β strand having name G
H/His	Histidine
I	Isoleucine
Inv	Invasin
K	Lysine
L	Leucine
Lac	Lactose
M	Methionine
Myf	Mucoid Yersinia factor
N	Asparagine
P	Proline
SC	Self complemented
SeMet	Selenomethionine

SPR Surface plasmon resonance  
T Threonine  
U Usher  
V Valine  
W Tryptophan  
WT Wild type  
Y Tyrosine  
YadA Yersinia adhesin A  
YE *Yersinia enterocolitica*

# 1 Background

## 1.1 General introduction

Bacteria are unicellular prokaryotic microorganisms with relatively simple structures. They are available in every habitat on Earth: in multicellular living organisms, soil, rock, oceans, and even in arctic snow. In the human body, most bacteria live in the digestive tract. Bacteria do both good and bad to animals and plants for the cycling of their own nourishment. There are incredible uses of bacteria in the production of yoghurt, cheese, soy sauce and many more food products. But a few bacteria are parasites and pathogens, which cause disease in animals and plants. Temperature and nutrition are the two major aspects of the reproduction of bacteria. With supportive environment some bacterial species like *E. coli* can divide at every twenty minutes and just after 8 hours, the number will have risen to 16, 777, 216. This is the ultimate reason why we get sick quickly after infestation of pathogenic bacteria.

Based on the structural characteristics of their cell wall through a staining method developed by the Danish scientist Christian Gram in 1884, bacteria are of two types, Gram-positive and Gram-negative. Gram-positive bacteria possess a thick and mesh-like cell wall, which is composed of mainly peptidoglycans that act as an exoskeleton and surrounds the cytoplasmic membrane. There are some deadly species in this group, for example anthrax-causing *Bacillus anthracis*, pneumonia-causing *Streptococcus* species, and botulism-causing *Clostridium* species. Gram-negative bacteria possess a thin peptidoglycans layer, which is covered by an outer layer of lipopolysaccharides and proteins. There is a space between this outer membrane and the inner cytoplasmic membrane, which is called the periplasm. The periplasm is very important for Gram-negative bacteria, because it is

where fimbrial and toxin subunits are accumulated (Zavialov et al., 2007). Pathogenically important genera of Gram-negative bacteria are *Vibrio*, *Salmonella*, *Shigella*, *Yersinia*, and *Escherichia*.

Pathogenic bacteria exert deadly impact on human and animal populations and health. For example, in the 14th century almost 50 million people died due to *Yersinia pestis* causing plague worldwide, which was named as the ‘Black death’ at that time (WHO, 2014). 28000-142000 people die from 1.4 – 4.3 million cases of cholera worldwide every year (WHO, 2015). Advanced development in hygiene, vaccination, and antibacterial agents improve the situation dramatically, but still millions are dying due to tuberculosis, pneumonia, and diarrhea. Very recently in a German outbreak, 54 people died due to hemolytic uremic syndrome caused by shiga-toxin producing *E. coli* (Frank et al., 2011).

Antibacterial resistance is not a future myth, but reality now. Severe life threatening problems have already arisen, for example Methicillin-resistant *Staphylococcus aureus* (MRSA). In the USA, MRSA has become deadlier than HIV in account of death (Levy et al., 2004). Gram-negative bacteria become multi-drug resistant (WHO, 2015) due to horizontal resistance-gene transfer to more-scarce pathogens in gut while taking traditional antibiotic against pathogenic bacteria (Keyser et al., 2008, Levy et al., 2004).

Both Gram-positive and Gram-negative bacteria move towards higher concentration of nutrients by using their motile organelles. To get maximum proximity to nutrient source, bacteria attach to the surface of host cells by using their ‘fimbriae or pili’ organelles. Or in another way, by using pili/fimbriae bacteria target and bind to the host tissue and initiate infection. Most pathogenic bacteria express hair-like organelle fimbriae, which are assembled in the periplasmic space from smaller subunits, on their outer surface. Fimbriae in Gram-positive bacteria are formed by covalent polymerization of subunits (Proft & Baker, 2009) whereas fimbriae in Gram-negative bacteria are formed by non-covalent polymerization (homo or heteropolymerization) of subunits (Waksman & Hultgren, 2009; Zavialov et al., 2007).

The fimbriae of Gram-negative bacteria are classified in five different groups: curli, type IV pili, type IV secretion pili, type III secretion needle, and chaperone-usher pili, based on their biogenesis (Fronzes et al., 2008). Chaperone-usher pili often constitute virulence factors and/or adhesins and



these have been most exclusively studied. They are responsible for tissue tropism, evade host-immune defenses, and facilitate invasion and colonization, which all are prerequisites to initiate infection (Wright et al., 2007; Zavialov et al., 2007).

### 1.2 The chaperone/usher pathway

The most common and extensively studied mechanism of assembling subunits into fimbriae is the chaperone/usher pathway (Waksman & Hultgren, 2009; Zavialov et al., 2007; Ofek et al., 2003). In this mechanism of formation of fimbriae, subunits are captured by the periplasmic chaperone and transported to the outer membrane counterpart usher where they are released from the chaperone, form a fibre, and finally translocate to the cell surface (Fig. 1).

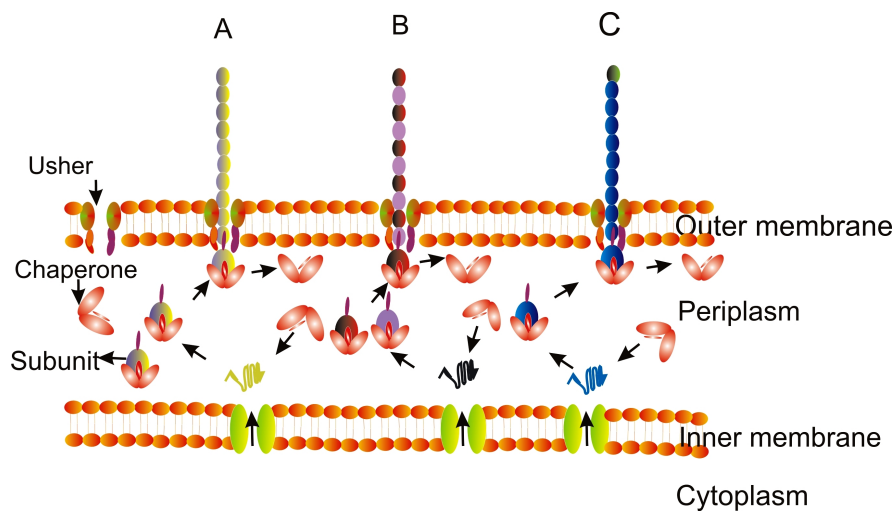


Figure 1. CU pathway of assembling subunits into fibre polymer on the bacterial surface. (A). Polymerization of F1 and Myf. (B). Polymerization of CS6. (C). Polymerization of AAF-1

### 1.3 Fimbrial subunits and architecture of a fibre

Fimbria is made up of single or multiple types of subunits. When a subunit is secreted from the cytoplasm to the periplasmic space through general secretory pathway, it consists of an incomplete immunoglobulin-like fold with six  $\beta$ -strands (A to F) instead of seven. These subunits are unstable in the periplasmic space and unable to assemble themselves since the subunit fold is incomplete (Zavialov et al., 2007; Sauer et al., 2002). In the periplasmic space,

the chaperone readily captures the subunit and forms a chaperone-subunit complex by donating and inserting its G1 strand into the cleft between strand A and F of the subunit to complete the immunoglobulin-like fold in a process called donor strand complementation (Sauer et al., 1999)(Fig. 1). After forming a complex with the chaperone, subunits become stable; otherwise their ultimate fate would be to be degraded by the proteases (Vetsch et al., 2004; Jones et al., 1997). During polymerization of the fibre at the outer membrane protein usher, the G<sub>d</sub> (N terminal extension) strand of subunits replaces the donor strand G1 from the cleft by a mechanism called donor strand exchange (Choudhury et al., 1999) (Fig.1).

Theoretically, it is possible to create a circularly permuted construct for any fimbrial subunit that is capable of DSC polymerization (Anderson et al., 2004; Zavialov et al., 2005). In this type of construct, the G<sub>d</sub> strand is moved to the C terminus of the subunit, enabling self-donor-strand complementation of individual subunit to form a self-complemented monomer with a classical Ig-fold. Such self-complementing design has been followed to design the subunits that are used in this thesis. Previous high resolution studies of self-complemented monomer and fibre assembled subunits have revealed nearly identical structures, for example CafI of *Yersinia pestis* (Zavialov et.al., 2005; Piatek et al., 2009).

Some CU assembled fimbriae are formed with polymerization of only one or two types of subunits, usually 2-3 nm long, thin, and flexible. Such fimbriae are, for example, the F1 antigen that is formed by the polymerization of CafI subunits (Zavialov et al., 2002; Chapman et al., 1999), Myf that is a polymer of one type of subunit (Nataniel et al., 2012) (Fig. 1A), CS6 that is formed by the polymerization of two different types of subunit (Paper I) (Fig. 1B), and AAF-1 that is a polymer of subunits A and B, where B is a single subunit and sits on the top of the fibre (Paper II) (Fig.1C). Type 1 fimbriae are thick and rod like, composed of four different types (FimA, FimF, FimG, and FimH) of subunits where subunit FimA forms the main body (polymer of thousand copies of FimA) followed by one copy of each remaining three where FimH sits on the distal part of the fimbria (Hahn et al., 2002; Jones et al., 1995). A couple of decades ago, it was hypothesized that fimbrial adhesins bind to the host cell receptors through the distal single subunit of the fibre.

In contrast to this hypothesis, recent studies suggested that many fimbrial adhesins have their binding sites on the structural subunits that form the main shaft of the fibre (Zav'yalov et al., 2010). For example, the family of fimbriae

assembled via the FGL-chaperone-usher pathway (Zavialov 2007) have no binding-specialized two-domain subunits, yet they act as adhesins.

In this thesis, three types of fimbriae assembled via the FGL-chaperone-usher pathway in Gram-negative bacteria have been studied: CS6 of enterotoxigenic *E. coli* (paper I), Aaf-1 of enteroaggregative *E. coli* (paper II), and Myf from *Yersinia enterocolitica* (paper III).

#### 1.4 Diarrheagenic bacteria and fimbriae

Diarrhea is the second largest and most common cause of infantile death, malnutrition, and growth stunting at less than five years of age. Annually, diarrhea kills around 760,000 children before seeing their 5<sup>th</sup> birthday from almost 2 billion cases globally. Virus, bacteria, and parasitic pathogens can cause diarrhea (Qadri et al., 2005; WHO, 2013). Travellers from the resourceful part of this planet suffer from traveller's diarrhea and the number is around 20 million per year where 80% travellers have suffered from bacteriogenic diarrhea. Among diarrheagenic bacteria, enterotoxigenic *E. coli* (ETEC) and enteroaggregative *E. coli* (EAEC) are responsible for 50% of the diarrheal episodes worldwide (Herbert L.D., 2005). Another important organism that causes gastrointestinal infection mostly in Europe is *Yersinia enterocolitica* (Fredriksson et al., 2012; Bottone et al., 1997)

Transmission of these organisms occurs through consumption of contaminated food and water, unpasteurized milk and milk products, and direct contact with farm and pet animals. Swallowing lake water while swimming and following no hygienic measure during cooking food put people in risk of getting infection. It is almost a truth that everyone has some risk of getting diarrheagenic infection in his/her lifetime (www.cdc.gov/ecoli/general/index/html as of 2015.09.20).

ETEC causes severe watery diarrhea along with dysentery, abdominal cramps, and sometimes fever. Very often diarrheagenic ETEC infection becomes life threatening and it is the second leading cause of infantile death (due to significant fluid loss and severe dehydration) in developing countries like Bangladesh and Egypt (Fisher and Black, 2010). According to the World Health Organization (WHO, 2006), 300,000 – 500,000 infants die annually from ETEC infection and associated malnutrition. ETEC colonizes the epithelial cell line of the host's small intestine by colonization factors. The bacteria bind host enterocytes by fimbrial adhesins. Adhesion to the

enterocytes allows this bacterium to transfer enterotoxins into the enterocytes and consequently stimulate efflux of water along with electrolytes from the cytoplasm. There are 25 colonization factors/fimbriae that have been identified till to date and *E. coli* surface antigen 6 (CS6) is the most common among these 25 factors (Gaastra and Svennerholm, 1996; Qadri et al., 2005). CS6 is a polymer of two distinct protein subunits CssA and CssB (Wolf et al., 1997) and analysis of purified native CS6 from a clinical isolate demonstrated that both subunits are produced in equal amounts and present in equal stoichiometry in CS6 fimbria (Ghosal et al., 2009).

Enteroaggregative *E. coli* (EAEC) is one of the most common diarrheagenic species of *E. coli* in USA and other parts of the world (Nataro et al., 2006). In some regions of the world, EAEC surpasses the ETEC as the most common diarrheagenic pathogen (Pablo et al., 2010). This bacterium causes persistent diarrhea in children, immunosuppressive patients (for example HIV infected) and travellers returning from the developing countries (Mathewson et al., 1995; Jiang et al., 2002). From the beginning of this decade, a Shiga-toxin producing strain of EAEC has been thought to be a significant threat to public health. The O104:H4 strain of Shiga-toxin producing EAEC was responsible for the German outbreak in 2011 when 3816 cases of gastroenteritis and 845 cases of hemolytic uremic syndrome occurred, and a total 54 deaths were reported (Frank et al., 2011). Like other enteric pathogens, EAEC uses a specific adherence factor called ‘aggregative adherence fimbriae (AAF)’ to recognize receptors prior to colonization in the intestine (Nataro et al., 1993; Czczulin et al., 1997; Bernier et al., 2002; Boisen et al., 2008). There are four known variants of AAF and AAF-1 is one of them. AAF-1 is encoded by the *agg* gene cluster and composed of two protein subunits AggA and AggB (Pakharukova et al., 2013; Paper II)

*Yersinia enterocolitica* (YE) infection is the 3<sup>rd</sup> most frequently occurring enteric disease in Europe and the rest of the world (Fredriksson et al., 2012 and 2003; Bottone et al., 1997). It causes diarrhea after ingestion of raw or uncooked pork (Fredriksson et al., 2003). The initiation of YE infection depends on some surface exposed proteins like Yops, YadA, Inv, Ail, and Myf (Bottone et al., 1997). All these surface proteins facilitate the adhesion of YE to the host enterocytes prior to colonization (Nataniel et al., 2012). Proteolytic effects readily degrade most of the surface proteins when they pass through the stomach and intestine but Myf can survive in these adverse conditions (Janja et al., 2011). Thus only Myf is able to adhere to enterocytes before any surface proteins get to involve in adhesion (Iriarte et al., 1993 and 1995; Levine et al.,

1984). Myf is regarded as a chromosome encoded virulence factor of YE and is composed of only one protein subunit A named MyfA (Nataniel et al., 2012).

### 1.5 Aim and outline of this thesis

Till to date, different types of fimbriae, assembled via the chaperone-usher pathway, have been identified and studied but many more are left unidentified. During the past few decades, plenty of work hours have been devoted into understanding the mechanism of assembly, atomic resolution structure, and adhesion to host tissues of these Gram-negative bacterial fimbriae. The aim of this thesis work was to elucidate the assembly, structure and biochemical properties of three fimbriae (CS6, AAF-1, and Myf) with regard to polyvalent adhesion to host tissues. The major methods used to obtain the results stated in this thesis will be covered in chapter 2. The results of these investigations will be presented and discussed in chapter 3 with individual headings for different fimbriae. The specific objectives of this thesis were

- *To determine the assembly mechanism of fimbrial subunits (Paper I).*
- *To determine the atomic resolution structures of subunit and subunit-ligand complexes (Paper I and Paper III).*
- *To elucidate the binding properties of subunits to galactosyl residues and extracellular matrix proteins (Papers I to III).*



## 2 Methods

### 2.1 Protein engineering, expression, and purification

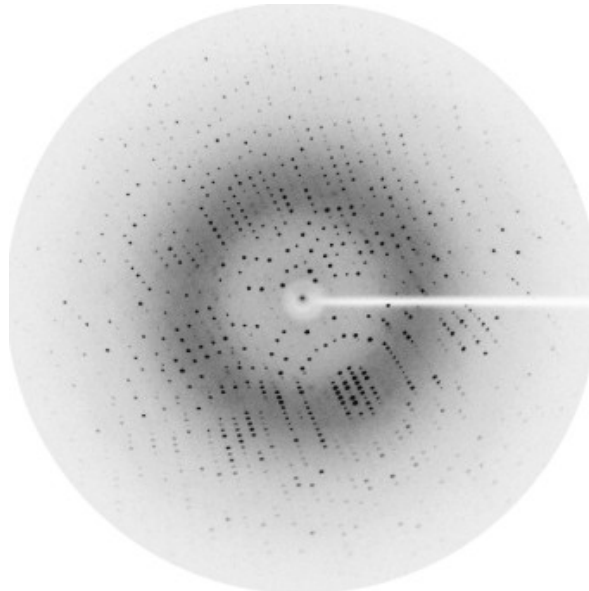
Donor strand complemented subunits, C<sub>ss</sub>AdsA, C<sub>ss</sub>BdsB, C<sub>ss</sub>AdsB, and C<sub>ss</sub>BdsA of ETEC, AggAdsA and AggBdsA of EAEC, and MyfA-sc of YE, were produced by extending the C-terminal sequences with the donor strand sequence from the same (self-complementation: C<sub>ss</sub>AdsA, C<sub>ss</sub>BdsB, AggAdsA, and MyfA-sc) or another subunit (C<sub>ss</sub>AdsB, C<sub>ss</sub>BdsA, AggBdsA). To give enough freedom for the donor strand to insert into the acceptor cleft of the subunit, a short linker (from two to four residues) was introduced between the C-terminus of the mature protein sequence and N-terminus of the predicted donor strand sequence. The original donor strand sequence at the N-terminus was either deleted or replaced by a 6His tag to facilitate protein purification. The signal peptide was left intact to ensure secretion of donor strand-complemented subunits to the periplasm. The synthetic genes were ordered from GenScript or produced by modifying the original genes with reverse PCR. The genes were inserted under the control of the T7 promoter in the pET101D expression vector (Invitrogen) to create expression plasmid(s). These plasmids were transformed into *E. coli* BL21 (DE3) or BL21 (AI) cells for overexpression. The proteins were purified using Ni<sup>+2</sup> chelate chromatography and/or ion exchange and size exclusion chromatography. SDS-PAGE analyses were performed after each step of purifications. Pure proteins were concentrated and stored for crystallization and binding studies.

## 2.2 X-ray crystallography

### 2.2.1 Introduction

The functional evolution of proteins arises through selective pressure in nature. Functional properties of proteins mostly depend on their 3D structures. In order to study the functional details of a given protein, it is therefore important to determine its 3D structure. X-ray crystallography has been used in this thesis to determine the structures of different fimbrial subunits and their complexes with ligand(s).

X-ray crystallography is based on the diffraction of electromagnetic radiation with a wavelength that should not be longer than a few Ångstrom ( $\text{\AA}$ ) to recreate an image at atomic resolution. Electromagnetic waves of such wavelength are called X-rays. The method determines the position of atoms of e.g. protein molecules ordered in a crystal lattice. Electron clouds of atoms scatter the incident X-rays, and due to the repeating pattern of the crystal lattice, constructive interference occur in certain directions, resulting in amplified diffracted X-rays that are recorded as spots on the detector (Fig. 2).



*Figure 2. X-ray diffraction pattern of a protein crystal. Black spots represent the 2D reflection of atomic positions.*



From the resulting diffraction pattern at different orientations of the crystal relative to the incident X-ray beam, obtained by small rotation of the crystal between consecutive images, and Bragg's Law of diffraction (Fig. 3), it is possible to back-calculate the distribution of electrons within the crystal, i.e. to retrieve an electron density map that is used to build an atomic model of the protein molecule (Giacovazzo et al., 2002; McRee, 1999).

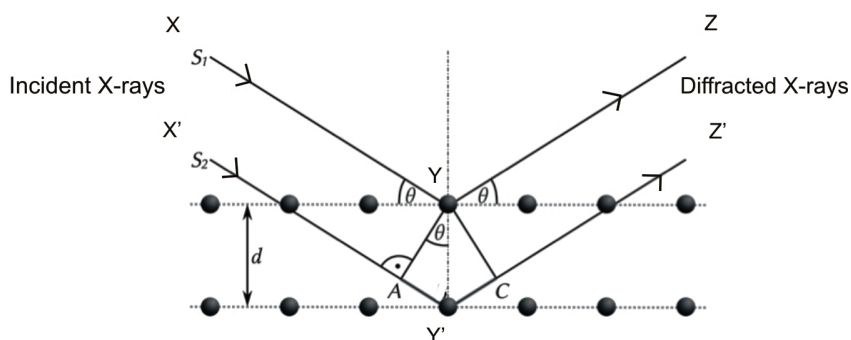


Figure 3. Bragg diffraction, when two beams ( $S_1$  and  $S_2$ ) of identical wavelength and phase travel the paths  $XYZ$  and  $X'Y'Z'$ , the lower beam  $S_2$  travels extra path ( $AY' + Y'C$ ) of  $2d \sin \theta$ . This  $2d \sin \theta$  should be equal to the wavelength multiplied by an integer to get constructive interference. Horizontally parallel black circles represent the planes.

Myoglobin and hemoglobin were the first X-ray crystallographic structures being solved by Perutz and Kendrew in the 1950's. Since then more than 90,000 protein structures, determined by X-ray crystallography, have been deposited in the Protein Data Bank (<http://www.rcsb.org/pdb/statistics/holdings.do> as of 2015-06-18). X-ray crystallography is a continuously evolving science based on the rapid advancement in computer software, synchrotron radiation facilities, detectors and sample handling equipment, large-scale recombinant protein expression systems, and high-throughput crystallisation robotics.

Upon diffraction data collection, the intensities of the spots are recorded, i.e. the amplitudes of the diffracted X-rays relative to each other. However, in order to calculate an electron density map, the phase angle for each diffracted X-ray is also needed, which is absent in the experimental data. There are several ways to overcome this phase angle problem (phase problem); in this thesis we used isomorphous replacement (incorporation of heavy metal into protein) and molecular replacement *in silico*.

### 2.2.2 SeMet incorporation for isomorphous replacement

Selenomethionine (SeMet) labelled donor strand complemented CS6 and His<sub>6</sub>-MyfA-sc subunits were expressed in *E. coli* BL21 (DE3) and BL21 (AI) cells, respectively. They were grown in M9 minimum medium supplemented with 100 µg/ml ampicillin to an optical density of 0.4 to 0.9 at 600 nm at 37°C. 30 min prior to induction, selenomethionine (50 mg/l), lysine (100 mg/l), phenylalanine (100 mg/l), threonine (100 mg/l), isoleucine (50 mg/l), leucine (50 mg/l), and valine (50 mg/l) were added to the growing cells followed by induction. (Adapted from Duyne et al., 1993). Protein harvest and purification protocols were as similar as stated in section 2.1.

### 2.2.3 Determining the structures of CS6 subunits and the MyfA-Gal complex

Crystallization was performed by the hanging-drop vapour-diffusion method at 293 K for CS6 subunits (30 mg/ml in 20 mM Hepes buffer at pH 7.5 + 150 mM NaCl) and at 278 K for MyfA-sc (35 mg/ml in 20 mM Hepes buffer at pH 7.4 + 150 mM NaCl) for both apo and SeMet derived crystals. CssAdsA and CssAdsB crystals were grown in drops with 24-30 % PEG4000 in 0.2 M ammonium acetate, 0.1 M Na acetate at pH 4.6. CsbdsA crystals were grown in drops with 30% PEG4000 in 0.2 M Na acetate, 0.1 M Tris-HCl at pH 8.5. MyfA-sc Crystals were grown in 35% 1,4-dioxane (Hampton, USA). All crystallization drops were of 4 µl initial volume (2µl protein + 2µl mother liquor) and equilibrated against 0.4 to 1 ml of respective precipitant.

Attempts to co-crystallize MyfA-sc with ligands (galactose, lactose, sialic acid, the GM1 ganglioside, and isoglobotriose) were also done at 278 K by both hanging (only for MyfA-sc-galactose complex) and sitting-drop vapor diffusion under the conditions mentioned above. For the MyfA-sc-galactose complex, MyfA-sc apo crystals were soaked in 250-500 mM galactose at 278 K for 2-10 hours.

Co-crystallization of MyfA-sc with other ligands were done with ~12.5 mM final concentration of ligand in the crystallization drops. MyfA-sc crystals grew within a few hours in the presence of sialic acid and in two weeks with the GM1 ganglioside and isoglobotriose, whereas no crystals appeared with lactose. Further, soaking experiments were done by adding lactose to 100 mM final concentration, or a small amount of beta lactose powder, to sitting drops with apo MyfA-sc crystals. Individual crystals were withdrawn using ready-made cryo loops (Hampton Research, CA, USA) followed by a brief (transient)

soaking in cryoprotectant (10-30 s; 5 parts mother liquor : 3 parts 50% PEG 400), flash-cooled and stored in liquid nitrogen for further use.

X-ray diffraction data were collected under cryo (liquid nitrogen) condition at the synchrotron beamlines ID29, BM14 (for CS6 subunits) and ID23-2 (for MyfA-sc complex with ligands) at ESRF, Grenoble, France, and MAX-Lab (for MyfA-sc complex with ligands), Lund, Sweden. The data were indexed, integrated, and scaled using the programs iMOSFLM and SCALA in the CCP4 suite (CCP4, 1994). 5% of the reflections were set aside for calculation of  $R_{\text{free}}$  factors. Heavy atom parameters were obtained, refined, and phases were calculated for CS6 subunits by using PHENIX (Adams et al., 2002). The initial models were built using PHENIX and O (Jones et al., 1991).

MyfA-sc-ligand complex crystal structures were solved by molecular replacement using Phaser\_MR (McCoy et al., 2007) and a previous structure of apo MyfA-sc without ligand (not published) as search model. Structure model refinement was done by alternating cycles of maximum-likelihood refinement with REFMAC5 (Murshudov et al., 2011), and manual rebuilding against  $2F_o - F_c$  and  $F_o - F_c$  electron density maps using Coot (Emsley et al., 2004).

## 2.3 Determination of binding constants

### 2.3.1 Surface plasmon resonance

Surface plasmon resonance (SPR) is an extremely versatile optical technique to study biomolecular binding interactions (Schuck., 1997). A target molecule is immobilized on the gold surface of a sensor chip and sample solutions with an interacting counterpart are injected through a series of flow cells. When polarized light projects on the gold surface (there is a glass surface behind the gold surface) of the chip, a total internal reflection occurs dependent on changes of mass, which causes changes in the excited light angle (refractive index). A detector in the system records these resonance changes in real time. We have used a Biacore system (Biacore X100, GE Healthcare, Sweden) to study binding interaction by SPR.

Two types of sensor chips were used, CM5 and NTA: Fibronectin (Sigma) was immobilized on a CM5 sensor chip by amine coupling using an Amine Coupling Kit (GE Healthcare, Sweden). Samples of subunits or mini-fibers at varying concentrations were injected onto the chip for 3 min to record the association and dissociation curves followed by flushing of the cell with 10

mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% Tween 20 (HBS-EP) for 3 min at a flow rate of 10 ml min<sup>-1</sup>. Identical samples were injected over a control flow cell to determine non-specific binding, which was subtracted from the experimental curves. After each data acquisition cycle, the chip was fully regenerated with 10 mM NaOH in HBS-EP containing 1 M NaCl. The purified His<sub>6</sub>-MyfA-sc protein was loaded onto a NTA sensor chip (GE Healthcare, Sweden). Contact time and capture stabilization periods were in both cases 90 s. Resonance signals were plotted as a function of time for several concentrations (0-1000 nM) of isoglobotriose (Elicityl, France). Association and dissociation times were 150 s and 300 s respectively. The chip was regenerated at each step of experiment with 350 mM EDTA. The equilibrium constants were determined by applying a one- or two-receptor binding model using the Biacore X100 evaluation software.

### *2.3.2 Tryptophan fluorescence spectrophotometry*

Tryptophan fluorescence spectrophotometry is a significant tool in biophysical research because of its robustness and sensitivity. Continuous advancements in optics and electronics have brought the use of fluorescent moieties (fluorophores) for much biophysical research instead of expensive and hazardous radioactive materials (Ladokhin et al., 2000).

Among the aromatic amino acids of protein, tryptophan is the most reliable in intrinsic fluorescent spectrophotometry because of its indole ring which acts as fluorophore and absorbs UV light at ~ 280 nm, can be excited at ~ 295 nm (there is no absorption by tyrosine and phenylalanine at this wavelength), and emits at ~ 350 nm. The photophysical properties of tryptophan (spectrum and intensity) are very sensitive to its local environment, hydrogen bonding, and some other non-covalent interactions. These properties make it unique in studying protein-protein and protein-ligand interactions (Ghisaidoobe et al., 2015). Fluorescence emission spectra were recorded with a Varian Cary Eclipse fluorescence spectrophotometer.

Frozen MyfA-sc was quickly thawed and diluted with buffer (10 mM Hepes pH7.4 + 50 mM NaCl). 2-fold dilutions of galactose and lactose ( $\beta$ -lactose; Sigma) from 150 mM to 9.37 mM and 250 mM to 1.95 mM respectively were prepared with the same buffer. MyfA-sc (43  $\mu$ M) was added to each concentration of galactose and lactose. One tube was kept as a negative control with no galactose and lactose but only MyfA-sc. Fluorescence emission spectra at the wavelength range of 300-500 nm were recorded at room temperature

with an excitation wavelength of 295 nM and a 5 nM slit width. Three spectra were recorded for each sample and data were analyzed with the software Origin7.

Dilution of the GM1 ganglioside (Carbosynth, England) powder, a 150 mM stock solution, and 2-fold dilutions down to 18.75 mM were prepared in 35% Dioxane. Individual spectra were recorded for the GM1 ganglioside in 35% Dioxane, 35% Dioxane, and MyfA-sc in 35% Dioxane. The protocol for measuring spectra was the same as described above for galactose and lactose.



## 3 Results and Discussion

### 3.1 CS 6 fimbriae (Paper I)

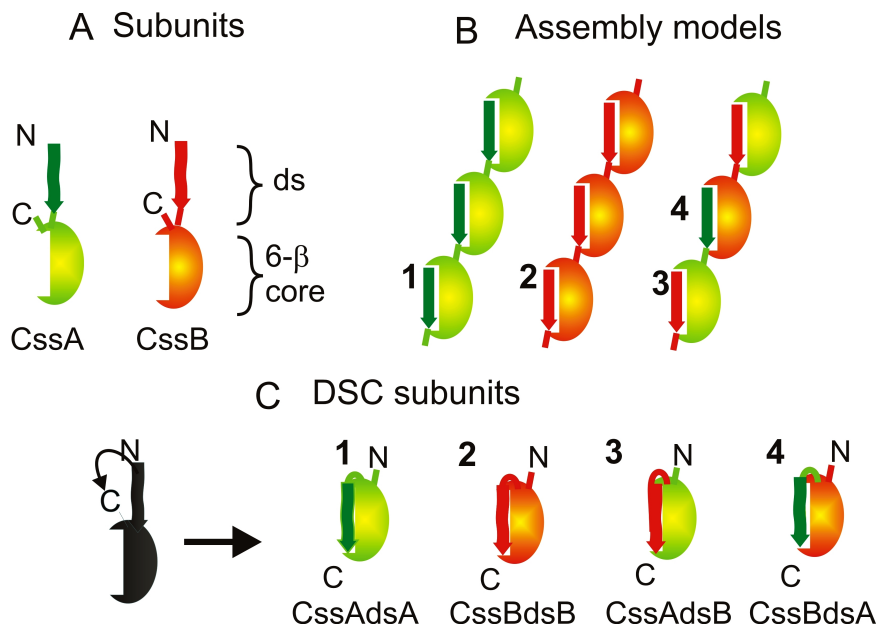
#### *3.1.1 Introduction*

ETEC binds to host receptors using CS6 fimbria. CS6 consists of two subunit, CssA and CssB, that are present in approximately equal amounts. This fact suggests that both of them play a role of major subunits. This is unusual as all previously characterized CU fimbriae consist of only one major subunit that forms the main 'shaft' structure of the pilus. What is the structure of the CS6 fibre? Are these two subunits assembled based on donor strand complementation principle? If yes, how do they become complemented, homo or hetero? Do they adopt the Ig-like fold? How do they appear in the building of the fibre? Which receptors do they recognize to facilitate adhesion? Is it polyvalent or monovalent adhesion? To answer these questions, we proceeded as follows.

#### *3.1.2 Modeling and design of self-complemented CS6 subunits*

CS6 subunits CssA and CssB display no sequence similarity to any other CU assembled fimbrial subunits. CS6 subunits might have a typical Ig-like fold, as indicated by sequence threading of CS6 subunits with known CU assembled fimbriae. Moreover, the N-terminus of both subunits show a pattern of alternating hydrophobic and hydrophilic residues which is a signature pattern of the  $G_d$  donor strand motif (Zavialov et al., 2003). Hence, we hypothesized that both CS6 subunits are able to exchange (donate and accept) inter-subunit N-terminal donor strands ( $G_d$ ) to form monomeric self-complemented (SC) subunits.

We considered two principle modes of subunit polymerization, homo-polymerization (insertion of  $G_d$  strand into the acceptor cleft of same type of subunit) and hetero-polymerization (insertion of  $G_d$  strand into the acceptor cleft of different type subunit) to explain our hypothesis (Fig. 4A, 4B). We engineered four self-complemented subunits (Fig. 4C) based on acceptor cleft-donor strand contacts:  $CssA_{6st} - G_dA$  ( $CssAdsA$ ),  $CssA_{6st} - G_dB$  ( $CssAdsB$ ),  $CssB_{6st} - G_dB$  ( $CssBdsB$ ), and  $CssB_{6st} - G_dA$  ( $CssBdsA$ ). We inserted one four-residue (DNKQ) linker between last residue of the subunit and first residue of the donor strand to provide sufficient conformational freedom to be inserted into the acceptor cleft. All four SC constructs were expressed in the periplasmic space of Expression cell *E. coli* BL21 due to the native signal peptide at the beginning of the amino acid sequence. Hetero-complemented constructs were expressed at a significantly higher level than homo-complemented constructs (Fig. 5A), which suggest that hetero-complemented constructs are more efficient in folding and resistant to local proteases.



*Figure 4. CS6 models with donor strand complementation. A. Schematic diagram of non-complemented subunits CcssA (green) and CcssB (orange). The N-terminal donor sequences ( $G_d$ ) are shown as arrows. The six-stranded ( $6\beta$ ) core structure is an incomplete oval to indicate its incomplete immunoglobulin like fold. B. Three possible versions of donor strand complementation-based CS6 polymers: CcssAn and CcssBn homo-polymers and . . .-CcssA-CcssB-CcssA-CcssB-. . . hetero-polymer. The numbers indicate possible donor strand complementation contacts. C. Possible versions (engineered) of self-complemented (SC) monomeric subunits.*



### 3.1.3 Hetero constructs are more stable than homo constructs

CS6 subunit constructs were analysed using circular dichroism (CD) spectroscopy. CD spectra at the characteristic wavelength of 208-215 nm suggested that the subunits possess  $\beta$  sheets. But exposure to heat and denaturant (guanidium hydrochloride; GdmHCl) changed the spectra dramatically, which is an evidence of losing native conformation. The temperature (up to 90° C) and GdmHCl (up to 4M) dependence of the CD of each SC subunit was recorded at a fixed wavelength to study the temperature and chemical induced conformational changes. All the resulting spectra produced the sigmoidal shape of a cooperative melting transition. Hetero constructs (CssAdsB and CssBdsA) are more stable than homo constructs (CssAdsA and CssBdsB). The transition melting temperature difference ( $\Delta T_m$ ) between CssAdsB and CssAdsA is 12.2° C and difference between CssBdsA and CssBdsB is 22.2° C (Fig.5B, Table 1). Different concentrations of GdmHCl caused a big change in the CD spectra of three constructs but no change up to 4M to CssBdsA (Fig.5C, Table 1). CssAdsA was the least stable and CssBdsA was the most stable.

This high stability of the CssBdsA construct is an indication of correct folding of a donor strand complemented subunit compared to other CU assembled fimbriae structures (Zavialov et al., 2005; Yu et al., 2012). Also, based on the thermal and chemical denaturation results, it is clearly demonstrated that self-complemented hetero constructs of CS6 are more stable than homo constructs suggesting that homo constructs (CssA<sub>6st</sub> - G<sub>d</sub>A and CssB<sub>6st</sub> - G<sub>d</sub>B) are unlikely to be native that eventually supports the hetero-polymerization of CS6 model.

Table 1. Expression and stability of CS6 donor strand complemented subunits

<i>Construct</i>	<i>Level of expression</i>	<i>T<sub>m</sub> (°C)</i>	<i>[GdmHCl] (M)</i>
CssAdsA	+	55.8 ± 0.2	0.67 ± 0.01
CssAdsB	++++	78.0 ± 0.2	2.12 ± 0.04
CssBdsB	++	58.9 ± 0.3	2.30 ± 0.05
CssBdsA	++++	71.1 ± 0.3	>4

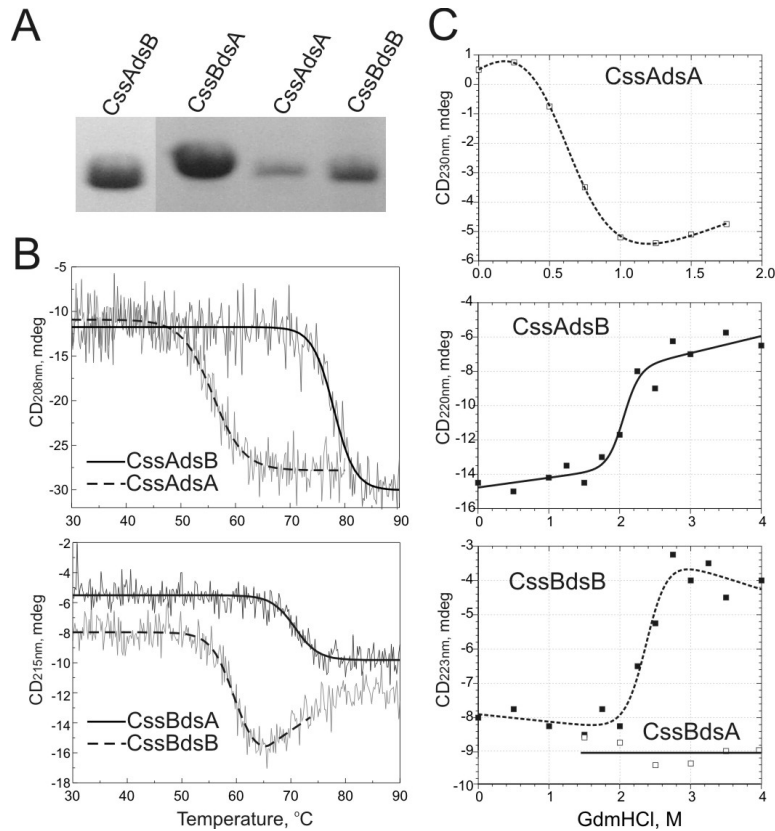


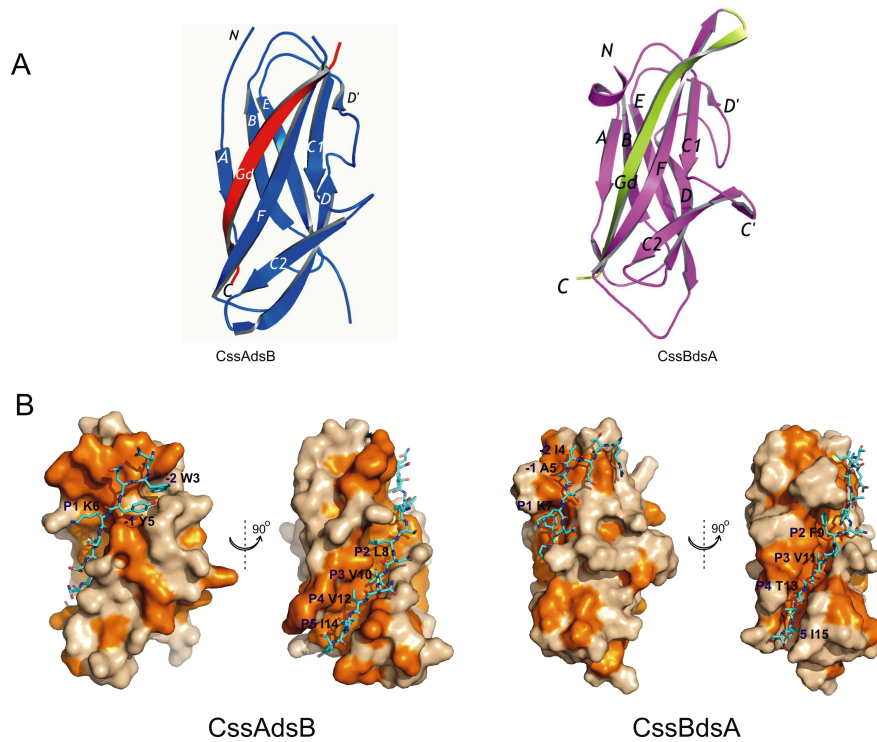
Figure 5. Stability analysis of self-complemented (SC) subunits. A. SDS-PAGE of periplasmic extracts of *E. coli* expressing four different SC subunits (labeled). B. Temperature stability of CssAdsA and CssAdsB (upper panel) and CssBdsA and CssBdsB (bottom panel). Temperature dependence of the CD signal was recorded at a heating rate of  $1^{\circ}\text{C min}^{-1}$  at 208 nm for CssAdsA and CssAdsB and at 215 nm for CssBdsA and CssBdsB. C. GdmHCl-induced denaturation. The CD signal was recorded at different concentrations of GdmHCl at 230 nm for CssAdsA (upper panel), at 220 nm for CssAdsB (middle panel), and 223 nm for CssBdsA and CssBdsB (bottom panel).

### 3.1.4 Atomic resolution structure of CssAdsB and CssBdsA

We determined the crystal structures of CssAdsB (1.5 Å) and CssBdsA (1.0 Å) (Fig. 6) to explain the structural pattern and assembly mechanism of CS6 subunits at atomic level. Both CssAdsB and CssBdsA structures have a classical immunoglobulin like fold that consists of a  $\beta$  sandwich of two  $\beta$  sheets (Fig. 6A). Strands A, B, E, and D form one sheet and strands F, C, and G<sub>d</sub> form the other one to form the sandwich pattern. The subunits have only 23

% sequence similarities but their folding topologies are similar. There is an  $\alpha$ -helix at the beginning of strand A in C<sub>ss</sub>B<sub>ds</sub>A which is absent in C<sub>ss</sub>A<sub>ds</sub>B but otherwise strands A, B, D, E, F, and G<sub>d</sub> are similar in both subunits. However, strands C differ. Strand C is split into C1 and C2 strands which are connected by a loop in C<sub>ss</sub>A<sub>ds</sub>B. In C<sub>ss</sub>B<sub>ds</sub>A, strand C includes an additional small strand C' in the C1, C2 connecting loop which is antiparallel to C2. Hence, the region between strands C1 and C2 represents the main topological difference between these two subunits. This dissimilarity in the C strand could potentially play a vital role in the biological function of both subunits and so for the entire fimbriae. Based on a DALI (Holm and Rosenstrom, 2010) search of the protein data bank, the FGL CU assembled subunits Dra invasin (of *E. coli*) and CafI (of *Yersinia pestis*) were identified as homologs, but they do not display significant sequence similarity to C<sub>ss</sub>A<sub>ds</sub>B and C<sub>ss</sub>B<sub>ds</sub>A subunits.

The atomic resolution structures of C<sub>ss</sub>A<sub>ds</sub>B and C<sub>ss</sub>B<sub>ds</sub>A suggest a model for assembling subunits in CS6. The donor strand complementation for assembling CS6 subunits (Fig. 6B) is inserting five donor strand (G<sub>d</sub>) residues into the respective acceptor cleft of the neighbouring subunit. The binding between donor strand and acceptor cleft is clearly driven by the hydrophobic effect because of the hydrophobic surfaces. There are classical 3 to 5 pockets in the acceptor cleft of the CU assembled subunits (Zav'yalov et al., 2010). Both C<sub>ss</sub>A<sub>ds</sub>B and C<sub>ss</sub>B<sub>ds</sub>A subunits acceptor clefts accommodate the last donor strand residue isoleucine at the last pocket (P5) in their acceptor cleft like Pap, Type 1, and F1 antigen (Sauer et al., 2002; Zavialov et al., 2005). Residue I14 of donor strand G<sub>d</sub>B and I15 of donor strand G<sub>d</sub>A occupied the deep pocket P5 of C<sub>ss</sub>A and C<sub>ss</sub>B respectively. Residue L8 of donor strand G<sub>D</sub>B and F9 of donor strand G<sub>d</sub>A occupied the pocket P2 of C<sub>ss</sub>A and C<sub>ss</sub>B respectively but do not enter deep into the pocket. Residues V10, V12, and I14 of donor strand G<sub>d</sub>B and V11, T13 and I15 of donor strand G<sub>d</sub>A enter deep into the pocket P3, P4 and P5 respectively. Pocket 1 is shallow for both donor strands and is occupied by the aliphatic side chain of lysine. Having a large residue like isoleucine in the pocket 5 is essential for initiation of donor strand exchange (Remaut et al., 2006; Yu et al., 2012).



**Figure 6. Crystal structures of CS6 subunits.** *A. Cartoon diagrams of 3D structures of C<sub>ss</sub>AdsB and C<sub>ss</sub>BdsA. The core structures of C<sub>ss</sub>A and C<sub>ss</sub>B subunits are colored in blue and purple respectively. The G<sub>d</sub>A and G<sub>d</sub>B strands are painted in green and red respectively. The  $\beta$ -strands are labeled. B. DSC contacts in C<sub>ss</sub>AdsB and C<sub>ss</sub>BdsA. Molecular surfaces of core structures are shown in beige with hydrophobic residues painted in orange. Donor strands are shown with sticks and colored by element type. Carbon, oxygen and nitrogen atoms are shown in cyan, red and blue respectively. The donor residues and pockets in the acceptor cleft are labelled.*

### 3.1.5 Model of CS6 fibre

The structures of hetero constructs C<sub>ss</sub>AdsB and C<sub>ss</sub>BdsA were used to model the CS6 fibre (Fig. 7). To model the CS6 fibre, the C<sub>ss</sub>AdsB and C<sub>ss</sub>BdsA subunits were positioned alternately in a line to conform to the hetero-assembly model. The artificial linker residues “DNKQ” which linked the N terminus of donor strand with the C terminus of CS6 subunits were deleted. C-terminus of donor strand of one subunit connected the N-terminus of the alternating subunit by using a native linker sequence of 4 to 5 and 7 residues in C<sub>ss</sub>A and C<sub>ss</sub>B, respectively, to return the mutation in C<sub>ss</sub>AdsB and C<sub>ss</sub>BdsA. The native linker sequences were not shown in the C<sub>ss</sub>AdsB and C<sub>ss</sub>BdsA structures (Fig. 6A). The CS6 fibre was modeled manually by using the program O (Jones et al., 1991) where appropriate stereochemistry was ensured by regularization. Two two-subunit repeats -[C<sub>ss</sub>AdsB-C<sub>ss</sub>BdsA)- and -[C<sub>ss</sub>BdsA-C<sub>ss</sub>AdsB)- with different interfaces between subunits in this CS6

fibres are suggested by the hetero-assembly model (Fig. 7). These two repeats, collectively, contain the entire surface of the CS6 fibre. We engineered two mini-fibres by using the two mentioned repeats to expose all the epitopes of CS6 antigen on SC constructs. The mini-fibres were expressed and accumulated at high level in the periplasmic space of *E. coli* BL21 expression cells (data not shown).

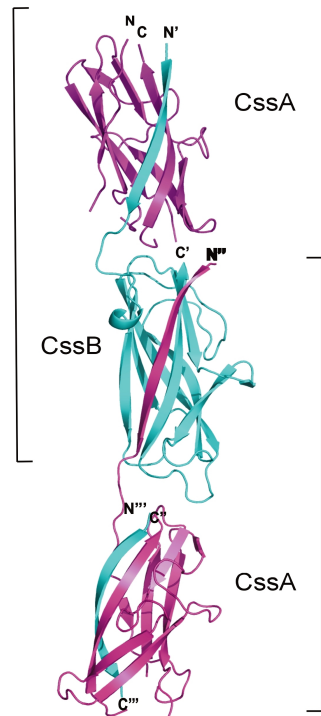


Figure 7. Model of CS6 polymer. 3D reconstruction of a CS6 polymer based on the crystal structures of the hetero constructs *CssAdsB* and *CssBdsA*. *CssA* and *CssB* are shown in purple and cyan respectively. Brackets indicate two distinct two-subunit repeats in the CS6 fiber.

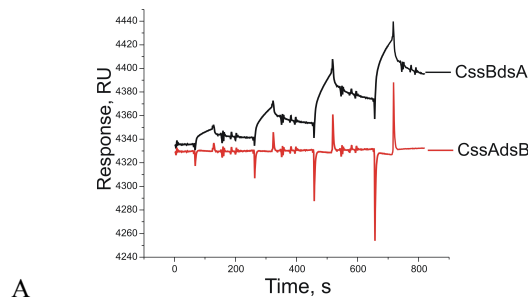
### 3.1.6 CS6 subunits recognize host receptor(s)

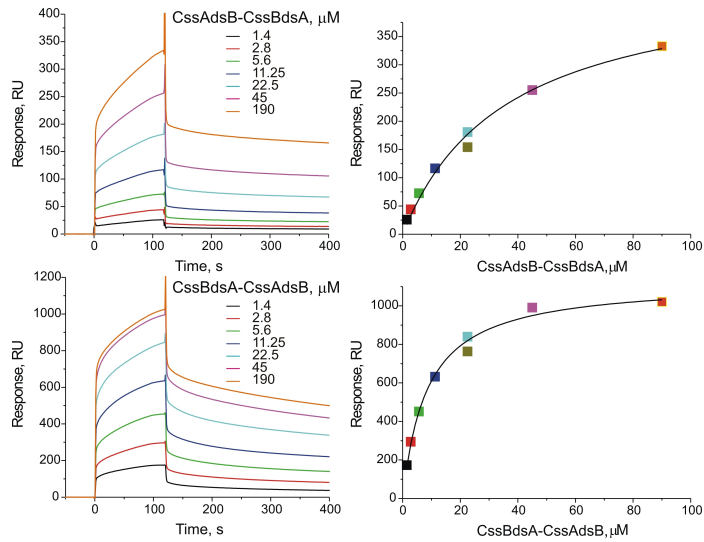
CS6 fibre has been suggested to recognize at least two different receptors; the extracellular matrix protein fibronectin and the glycosphingolipid sulfatide (Ghosal et al., 2009; Jansson et al., 2009). However, since the binding of most proteins depend on their 3D structure, the existing binding studies of CS6 subunits to fibronectin and sulfatide with non-native aggregates (Ghosal et al., 2009) or non-complemented subunits obtained from inclusion bodies (Tobias et al, 2008; Jansson et al., 2009) respectively should be treated with caution.

Hence, we re-examined the receptor recognition properties of CS6 using individual SC subunits and mini-fibres.

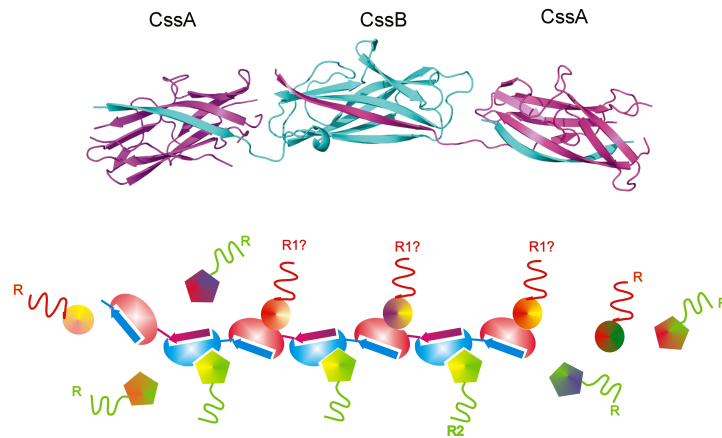
Quantitative characterization of binding between our SC constructs and human plasma fibronectin was performed using SPR. Fibronectin was immobilized covalently onto a sensor chip and the SPR signal was measured after adding purified individual constructs in the liquid phase. C<sub>ss</sub>B<sub>ds</sub>A showed distinct binding to fibronectin with a dissociation constant ( $K_D$ ) of  $9.1 \pm 2.0$  and  $340 \pm 50 \mu\text{M}$  (two binding sites) whereas C<sub>ss</sub>A<sub>ds</sub>B showed no affinity to fibronectin (Fig. 8A). Homo-construct C<sub>ss</sub>B<sub>ds</sub>B showed a similar binding affinity as the hetero-construct C<sub>ss</sub>B<sub>ds</sub>A to fibronectin but C<sub>ss</sub>A<sub>ds</sub>A showed no binding. So, the characterization of binding interactions of SC subunits revealed that the fibronectin binding site is situated on the surface of C<sub>ss</sub>B and not on C<sub>ss</sub>A, which is the opposite what has been suggested by Ghosal et al. 2009. C<sub>ss</sub>A does not bind to fibronectin but binds to cells (Paper I), which suggested that C<sub>ss</sub>A must recognize a receptor other than fibronectin.

Both mini-fibres recognized fibronectin via a single binding site (Fig. 8B). The absence of second or unspecific binding site on mini-fibre suggests that the low affinity second binding site of C<sub>ss</sub>B<sub>ds</sub>A to fibronectin is not native but unspecific and could be buried in the subunit junction. Mini-fibre C<sub>ss</sub>B<sub>ds</sub>A-C<sub>ss</sub>A<sub>ds</sub>B showed stronger affinity ( $8.5 \pm 0.7 \mu\text{M}$ ) to fibronectin than C<sub>ss</sub>A<sub>ds</sub>B-C<sub>ss</sub>B<sub>ds</sub>A ( $30 \pm 5 \mu\text{M}$ ). The order of individual subunit in the mini-fibres and the interface between the subunits may affect the binding site and affinity to fibronectin. The weaker binding of C<sub>ss</sub>A<sub>ds</sub>B-C<sub>ss</sub>B<sub>ds</sub>A to fibronectin and tighter binding to cells (Paper I) suggested that there is a different receptor to be recognized by the C<sub>ss</sub>A subunit of the CS6 fibre. Polyvalent binding of CS6 to fibronectin and other receptor molecules are consistent with our model for the CS6 fibre (Fig. 8C).





B



C

Figure 8. Binding analysis of CS6 SC-hetero-construct and mini-fibres. (A). CcssBdsA, but not CcssAdsB binds to fibronectin. Comparison of binding of CcssAdsB and CcssBdsA to a Biacore sensor chip with immobilized fibronectin. (B). Binding analysis of mini-fibres CcssAdsB-CcssBdsA (top) and CcssBdsA-CcssAdsB (bottom) to fibronectin. SPR sensorgrams recorded for different concentrations of mini-fibres are shown in the left panels. One-site binding saturation curves are shown in the right panels. (C). Polyvalent binding (schematic diagram) of a CS6 fibre to multiple receptors of two different types (R1 and R2). 'R' and '?' stand for receptor and unknown respectively

### 3.1.7 Discussion

A linear polymer of subunits joined ‘head-to-tail’ by donor strand complementation is the basic structure of all fimbriae assembled via the chaperone/usher (CU) pathway. We show that CS6 assembly also involves donor strand complementation despite the non-fimbrial morphology (Ludi et al., 2006). All previously studied CU fimbriae consist of long stretches of homo-polymers made of subunits of the same type. In contrast, the CS6 fibre is built as a hetero-polymer with two different subunits alternating along the entire fibre. The poor fit between donor strands ( $G_dA$  and  $G_dB$ ) and acceptor clefts of *CssA* and *CssB* (data not shown) explains the low stability of the SC homoconstruct (*CssAdsA* and *CssBdsB*). Since the donor strand–acceptor cleft contact determines the order of subunit arrangement in the fibre (Rose et al., 2008), our SC hetero constructs strongly support the hetero assembly model of CS6 fibre.

A number of CU (FGL type) operons are predicted to assemble surface organelles from two different subunits (Zavialov et al., 2007). CU pathway assembled fimbriae can be divided into the two groups monovalent and polyvalent adhesins based on functional properties (Zav’yalov et al., 2010).

We found that only *CssBdsA* specifically binds to the extracellular matrix protein fibronectin whereas *CssAdsB* does not bind to fibronectin but binds to cell (Paper I), which suggested that *CssAdsB* must recognize a receptor other than fibronectin. This is in accord with the observation of large differences in the structure of the loop region between the C and D strands of the two subunits, which is likely to be responsible for their binding properties. Since *CssAdsB* and *CssBdsA* alternate along the fibre, the different binding surfaces must appear at a certain periodicity along the fibre. Such an arrangement of subunits in the formation of fibre would enable binding of the CS6 fibre simultaneously to multiple receptors of two (or more) different types (Fig. 8C). The high flexibility of the CS6 fibre predicted from our modelling would help to establish such multiple contacts. Our structural and biochemical studies suggest that CS6 is a polyvalent adhesin of a novel type: a hetero-polyadhesin.



## 3.2 AAF-1 fimbriae (Paper II)

### 3.2.1 Introduction

AAF-1 is a polymer of the two subunits AggA and AggB. AggA is the major subunit and AggB is the putative minor subunit. By using this AAF-1 fimbria, EAEC adheres to the host receptor(s). What is the structure of AAF-1? Which receptors do they recognize to facilitate adhesion? Is it polyvalent or monovalent adhesion? We performed the following to answer all these questions.

### 3.2.2 Atomic model of AAF-1

We designed donor strand complemented subunits to determine the structure of AAF-1. Since AggA is the major subunit, we created a self-complemented AggA monomer (AggAdsA). AggB was complemented with the donor strand from AggA (AggBdsA). Crystal structures of AggAdsA and AggBdsA were determined to resolution 1.6Å and 2.4Å resolution, respectively. High complementarity between donor strands and acceptor clefts suggested that donor strands were chosen correctly. Hence, we conclude that AggA is the major subunit that is capable of self-polymerization and AggB is the minor subunit that uses the donor strand of AggA to be located at the tip of the AggA polymer to accomplish the formation of AAF-1 fimbria (Fig. 9) as in a mini-fibre model of F1 antigen (Zavialov et al., 2003).

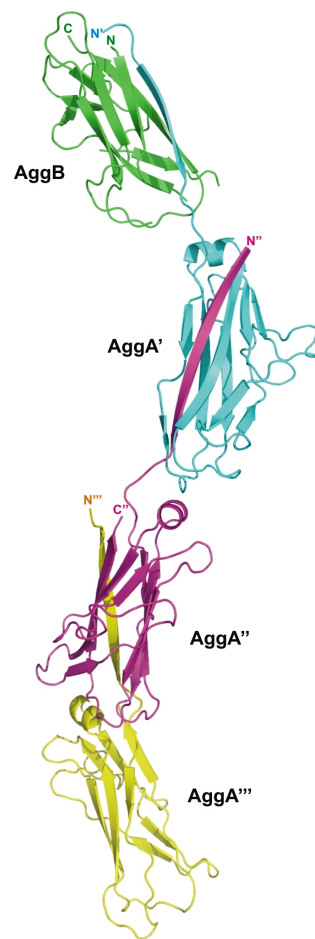


Figure 9. Model of AAF-1 fimbria based on the atomic resolution crystal structures of donor strand complemented subunits, AggAdsA, AggBdsA. A single copy of AggBdsA (green) subunit is located on the tip of the AggAdsA subunits polymer.

### 3.2.3 AAF-1 recognizes fibronectin as a receptor

AggAdsA and AggBdsA recognize fibronectin. We studied the affinity of AggAdsA and AggBdsA to fibronectin using surface plasmon resonance (SPR). The experiment showed that AggAdsA binds to fibronectin with a dissociation constant of  $16 \pm 2 \mu\text{M}$  which is similar to that previously found for AafAdsA (Farfan et al., 2008) whereas AggBdsA shows significantly lower affinity (Fig. 10, Table 2).

### *3.2.4 The AggAdsA-fibronectin interaction is ionic not hydrophobic*

Basic arginine and lysine residues have been identified, by NMR spectroscopy, as functionally important for the binding of the AAF-II major subunit to fibronectin (Paper II, the experiments were performed in our collaborator's lab). AAF-1 subunits are rich in basic residues (Fig. 11) and hence we hypothesized that those basic residues and ionic interactions are significantly important for AggAdsA to bind fibronectin. We reduced the surface positive charge of AggAdsA by mutating basic amino acids (corresponding to the equivalent basic region previously identified in AafA of AAF-II in Paper II) into alanine through site directed mutagenesis and measured the affinity to fibronectin and ionic dependency by SPR (Fig. 10, Table 2). We mutated the proximal pairs of lysine including K51 and K109, K55 and K103, K73 and K76, and closely positioned lysine K73, K76, and K78 (Fig. 12). AggAdsA possesses closely positioned surface W57, W59, and exposed F91 (Fig. 11). We also mutated these three residues to alanine and studied the binding affinity to fibronectin by SPR (Table 2).

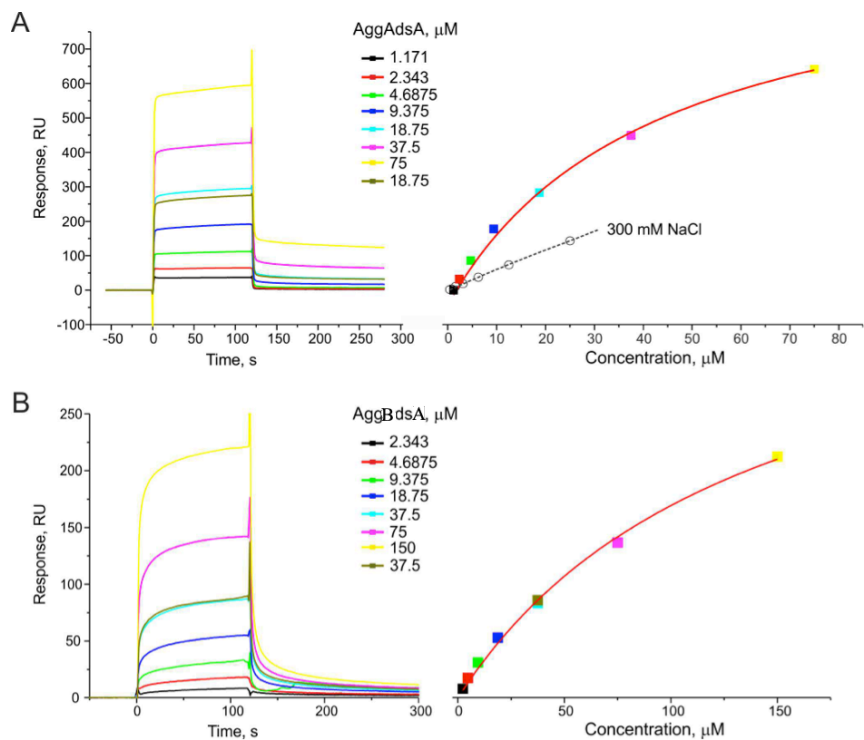


Figure 10. SPR analysis of *AggAdsA* (top) and *AggBdsA* (bottom) binding to fibronectin. Left panels: SPR sensograms recorded for different concentrations of *AggAdsA* and *AggBdsA*. Right panels: Saturation curves (one binding site model). Top right panel: dashed line in black represents experiment at high salt concentration to study ionic interactions.

Table 2. Effects of ionic dependency and basic residue mutations on the dissociation constant of *AggAdsA*-fibronectin interactions.

Subunit/Mutation	$K_D$ , $\mu\text{M}$
WT, 75 mM NaCl	16 $\pm$ 2
WT, 300 mM NaCl	>200
W57	31 $\pm$ 6
W59	17 $\pm$ 3
F91	20 $\pm$ 3
K51 and K109	49 $\pm$ 4
K55 and K103	56 $\pm$ 5
K73 and K76	61 $\pm$ 6
K73, K76, and K78	>200
<i>AggBdsA</i>	100

AggAdsA	ASQQTQTIRLTVTNDPCVTITTPPQTVGVSSTPIG <b>FSAK</b> VTTSDQ <b>CIKAGAKV</b> VLWG 60
AggBdsA	-----AEITLISH <b>K</b> TLG <b>SQ</b> LRDGM <b>K</b> LATG <b>RIACRE</b> PHDGFHIWINA 41
AggAdsA	TGPANK <b>V</b> VLQ <b>HAKVAKQKY</b> TLNPSIDGGAD <b>FVNQ</b> GTDA <b>KIYKK</b> LTSGN <b>KFL</b> NASVSVN 118
AggBdsA	<b>K</b> VGHYIVQNN <b>RETK</b> HEL <b>KVK</b> IGGGG <b>W</b> SSSLIEG <b>Q</b> RG-V <b>YR</b> Q <b>GEEK</b> QAIFDIMS <b>DGN</b> 100
AggAdsA	<b>P</b> K <b>TQ</b> VLIPGE <b>Y</b> TMILHAA <b>VDF</b> --- 139
AggBdsA	--- <b>Q</b> YSAPGE <b>Y</b> IFSVSGE <b>CLIS</b> R <b>G</b> 121

Figure 11. Sequences of AggAdsA and AggBdsA. Basic residues are bold and painted in pink color. Aromatic residues are painted in blue color.

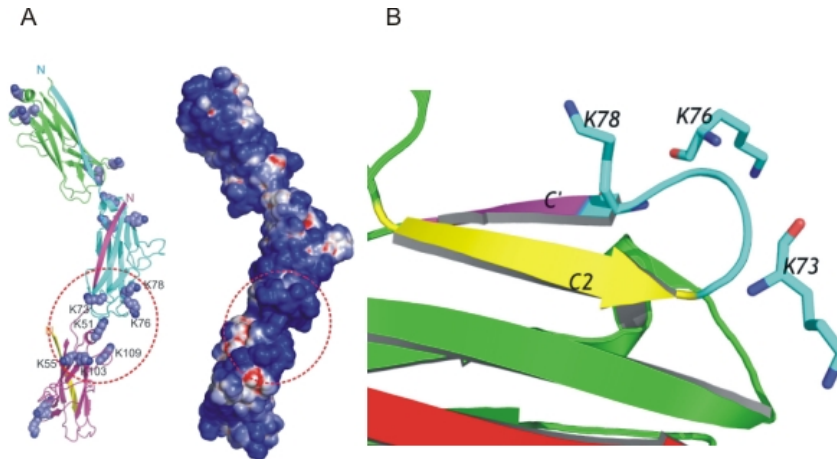


Figure 12. Location of fibronectin binding sites in AAF-1. A. Cartoon presentation (left) and electrostatic potential assessable surfaces (right) of three major subunit polymers. Positively charged residues (interacting with fibronectin) are shown as sphere models. Positive surface potential regions are depicted in blue and negative potential regions in red. The basic residues are concentrated at the interface between adjacent subunits. Seven lysine residues, 51, 55, 73, 76, 78, 103 and 109 in this region are involved in fibronectin binding. B. Important location of three close-positioned lysines. C2 and C' strands and their connecting loop are painted as yellow, purple, and cyan respectively. Lysines are marked with their sequence number.

### 3.2.5 Discussion

High complementarity of donor strand and acceptor clefts of AAF-1 subunits suggest that subunit AggAdsA is self complemented and AggB is complemented with the donor strand from AggA. An AggB subunit complements neither itself nor AggA, which consequently suggests a subunit arrangement in the fibre where AggB is located on the tip of the AggA polymer to form AAF-1.

All the double lysine mutated AggAdsA mutants showed 3-4 fold drops in affinity to fibronectin compared to the wild type AggAdsA monomer. The triple lysine mutated mutant's affinity was below the detection limit of the SPR experiment. By comparing the binding affinity of wild type and mutant variants, it is clear that binding affinity depends on the presence of basic surface residues lysine 51, 55, 73, 76, 78, 103, and 109 and most exclusively 73, 76, and 78, which are concentrated at the interface between adjacent subunits (Fig. 12A, Table 2). Strand C2 and C' along with their connecting loop and three closely positioned lysine residues (Fig. 12B) are found to be the most critical region and residues, respectively, in fibronectin binding. Mutation of these three lysines can abolish the affinity to fibronectin (Table 2).

Wild type AggAdsA's affinity to fibronectin dropped more than ten times when we used 300 mM instead of 150 mM NaCl in the running buffer (10 mM Hepes at pH 7.4). This ionic strength of the running buffer could disrupt the salt bridge between the interacting molecules. On the other hand, mutation of W57, W59, and exposed F91 did not cause any significant change neither in protein expression (Paper II) nor in binding to fibronectin compared to the wild type. So, the hydrophobic interactions by these residues are not essential to bind fibronectin (Table 2).

The involvement of basic residues, ionic strength dependency, and little involvement of hydrophobic surface in affinity to fibronectin reveal that this binding of AggAdsA to fibronectin is electrostatic rather than hydrophobic which is in contrast to the mechanism used by most Gram-positive bacterial fibronectin-binding proteins (MSCRAMMs: microbial surface components recognizing adhesive matrix molecules) (Henderson et al., 2011).

The dissociation constant for the wild type AAF-1 forming subunits is in the micromolar range that can be defined as a weak interaction but polyvalent interactions via the entire AAF-1 fibre may enforce it. Moreover, it can be said that fibronectin is not the only receptor for AggAdsA or the entire AAF-I. A single AggAdsA at the tip of the fimbriae is unlikely to promote attachment of bacteria to enterocytes through fibronectin. However, a polymer of AggA subunits with polyvalent attachment would mediate a tight bacterial adhesion with several fibronectin and/or other ECM proteins or carbohydrate molecules.

### 3.3 Myf fimbriae (Paper III)

#### 3.3.1 Introduction

*Yersinia enterocolitica* assembles Myf fimbria on its surface. Myf is a polymer of single subunit MyfA. The biological importance of having this fimbria on the YE cell surface is still unknown (Nataniel et al., 2012). The closest homologue of Myf, the pH6 antigen functions as carbohydrate-binding adhesin. Hence, we hypothesized that Myf may promote attachment of *Yersinia enterocolitica* to epithelial cells by recognizing carbohydrates. In this study, we decided to determine the structure of Myf and examine its possible carbohydrate-binding properties.

#### 3.3.2 MyfA-sc binds galactose and lactose

MyfA-sc contains four tryptophans at positions W50, W79, W85, and W86. We hypothesized that one or more tryptophan can be located near the active binding site of MyfA. We designed self-complemented MyfA (MyfA-sc) and studied MyfA-sc-galactose and MyfA-sc-lactose interactions using tryptophan fluorescence spectrophotometry. These experiments revealed an affinity of MyfA-sc to galactose and lactose with dissociation constants of  $42 \pm 15$  mM and 9 mM respectively (Fig. 13, Table 3).

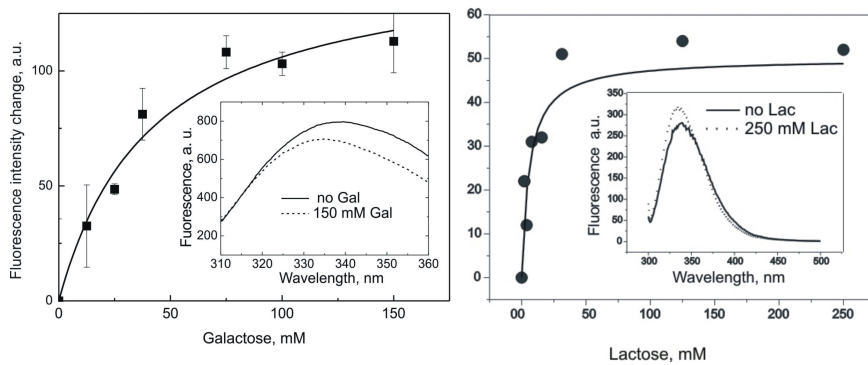


Figure 13. Saturation analysis of galactose and lactose binding to MyfA-sc. Insert: fluorescence emission spectra with and without galactose and lactose are represented by solid lines and dotted lines respectively. Bars and black circles represent estimated standard deviation and average value respectively based on 3 to 5 independent measurements

### 3.3.3 MyfA-sc binds isoglobotriose

The binding affinity of MyfA-sc to isoglobotriose was studied using SPR. MyfA-sc was captured at an NTA chip (GE Healthcare, Sweden) and binding of isoglobotriose was measured by recording the SPR signals after injecting different concentrations of isoglobotriose in the liquid phase. The calculated dissociation constant depicted that MyfA-sc showed very strong binding affinity ( $68 \pm 18$  nM) to isoglobotriose (Fig. 14).

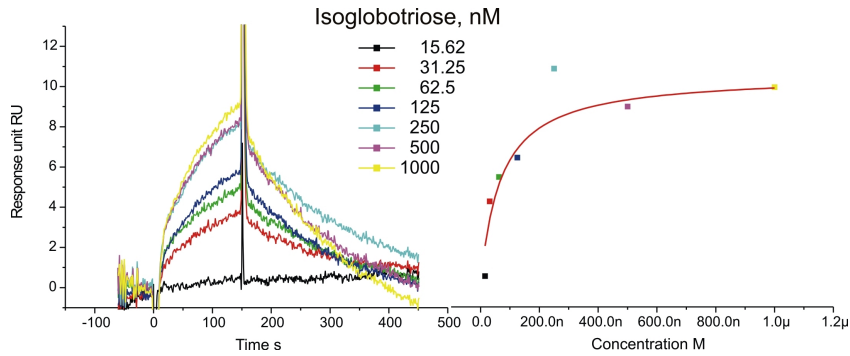


Figure 14. Real time SPR association and dissociation analysis of binding of MyfA-sc to isoglobotriose. Left panel: representative sensorgrams recorded for different concentrations of isoglobotriose. Right panel: Saturation curve.

### 3.3.4 Atomic structure of MyfA-sc –Gal complex reveals the binding site

The recombinantly expressed and purified MyfA-sc was successfully crystallized under similar conditions as previously used to determine the structure of unliganded MyfA-sc (Natalia et al., Manuscript), with the same space group ( $P2_12_12_1$ ), similar cell dimensions and two protein chains in the asymmetric unit. Soaking and co-crystallization experiments were performed with MyfA-sc and galactose, lactose, sialic acid, the GM1 ganglioside, and isoglobotriose and structures were determined. Soaking with galactose was successful and showed clear electron density for a bound ligand, whereas no convincing ligand density could be seen with the other sugars. The MyfA-sc – galactose complex structure (MyfA-sc-Gal) was refined at 1.65 Å resolution and final  $R_{\text{work}}/R_{\text{free}}$  of 0.19/0.21.

The N-terminus of the protein (13 residues including the His<sub>6</sub> tag) is not visible in the electron density. Also the loop between strand F and the relocated donor strand G<sub>d</sub> appear to be disordered, and is not included in the structure



model. All the remaining amino acids could be placed with confidence in the electron density.

There is a clear density for a bound galactose molecule in chain A but not in chain B. The positions of all atoms of the bound galactose are well defined by the electron density map (Fig. 15A), and the refined galactose molecule fits very well. Galactose binds in a shallow pocket formed by  $\beta$ -strands D2, D3, C1 and F near one end of the protein (Fig. 15B), and the pocket seems to show good complementarity in shape to galactose (Fig. 15C). Similar pockets are found in type C carbohydrate-binding modules, and galactose binds in a similar location as in the *Yersinia pestis* PsaA-galactose complex (Bao et al., 2013). In our MyfA-sc-Gal complex, the galactose is surrounded by MyfA-sc residues K48, N51, D81 and H84 that make hydrogen bonds with galactose hydroxyl groups at position 6, 3, 2 and 3, respectively. The W86 and Y120 residues also surround the galactose and provide hydrophobic binding surfaces (Fig. 15D). The hydrophobic face of the galactose ring (C1-C2-C3-C4-C5) lays rather flat on one face of the indole sidechain of W86, which is exposed and makes up one wall of the binding pocket. The Y120 sidechain is facing the hydrophobic C4-C5-C6 edge of galactose. The fact that W86 is directly involved in galactose binding and strongly suggests that W86 is the tryptophan that is responsible for quenching of fluorescence intensity upon sugar binding.

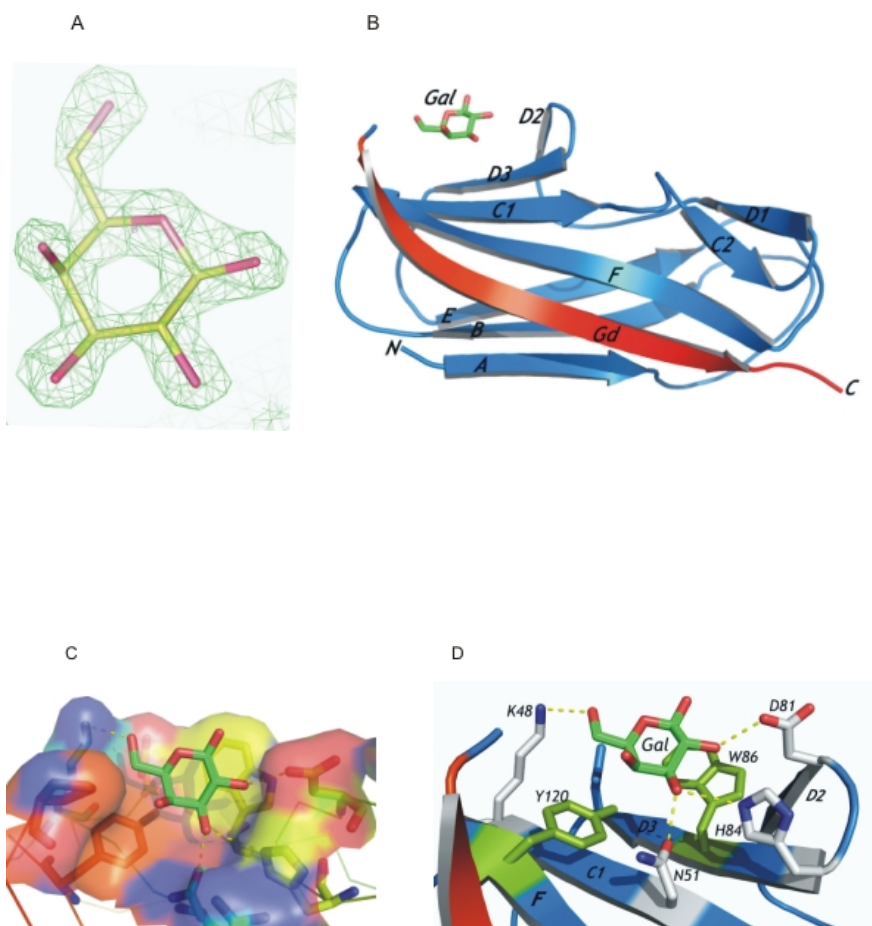


Figure 15. Crystal structure of the MyfA-sc-Gal complex. (A) The galactose molecule fits well into the electron density. (B) Overall backbone secondary structure colored blue (all six strands) to red (donor strand) from N- to C-terminus, showing the position of the bound galactose molecule (stick model; green carbon atoms) near  $\beta$ -strands D2, D3, C1 and F. (C) Close-up view of galactose and the surface of the binding pocket. (D) Amino acid sidechains that surround and interact with the bound galactose molecule. Hydrogen bonds are depicted as dashed yellow lines.

Table 3. Dissociation constants for ligands binding to MyfA-sc

Ligand	Dissociation constant ( $K_D$ )	Method
The GM1 ganglioside	140 mM	Tryptophan fluorescence
Galactose	$42 \pm 15$ mM	Tryptophan fluorescence
Lactose	9 mM	Tryptophan fluorescence
Isoglobotriose	$68 \pm 18$ nM	SPR
Sialic acid	Not detected	Tryptophan fluorescence
Fibronectin, collagen type_4	Not detected	SPR

### 3.3.5 Discussion

We examined the interactions between MyfA-sc and galactose, lactose, and some other sugar analogues of glycosphingolipids that have terminal galactosyl residues. It is already evident that galactose causes a large decrease in fluorescence emission when it binds with cholera toxin subunit B (Jennifer et al., 1996). Here, we find that MyfA-sc recognizes galactose.  $\beta$ -strands D2, D3, C1, F and associated loops of MyfA-sc (Fig. 15D) make a shallow groove in which galactose binds. PsaA of *Yersinia pestis*, a 43% homolog of MyfA-sc, has previously shown the same binding behavior to the galactosyl residue of glycosphingolipids (Payne et al., 1998). The highly variable C and D strands and their associated loops are indeed functionally important to provide binding niches to ligands for most CU assembled fimbrial (such as Psa, F1, Saf, Caf1, DraE) subunits which possess overall similar structures and act as polyadhesins (Bao et al., 2013).

We have also studied MyfA-sc–lactose, MyfA-sc–GM1 ganglioside, and MyfA-sc–isoglobotriose complex and found that lactose and the GM1 ganglioside bind at millimolar levels but isoglobotriose binds with nanomolar affinity. Lactose shows 5 times stronger affinity than galactose to MyfA-sc (Table 3). Alpha configuration of galactose in isoglobotriose elicits stronger binding compare to beta configuration in lactose and in the GM1 ganglioside. Another CU assembled fimbriae Fim-H of *E. coli* selects alpha configuration of D-mannose (Hung et al., 2002). So in selecting primary determinant of host receptor/glycosphingolipids, MyfA-sc recognizes and binds the alpha anomer of galactose more tightly than beta anomer.

The MyfA-sc-galactose complex quenched the emission of fluorescence from W86; but for lactose and the GM1 ganglioside, there was increase of emission. This increase would be the reason of bringing the tryptophan from hydrophobic core into hydrophilic environment by inducing some conformational changes in binding site that we do not observe in the structure. Most often the millimolar range binding (common affinity of sugars to lectins (Bouckaert et al., 2005) is defined as an inactive state but Nature can make great functional use of this weakly binding ligand (Turnbull et al., 2004). However, we have failed to determine atomic resolution structures of MyfA-sc-lactose, MyfA-sc-GM1 ganglioside, and MyfA-sc-isoglobotriose complexes, so further discussion in this regard is not possible. We also attempted to study the potential interactions of MyfA-sc with sialic acid, collagen type\_4, and fibronectin using both X-ray crystallography and biophysical method (SPR) but could not identify any binding to MyfA-sc (data not shown).

We conclude that there is a binding site for a galactosyl residue on MyfA-sc. There is also evidence for binding of MyfA-sc to lactose, isoglobotriose and the GM1 ganglioside. MyfA might therefore act as an adhesin or polyadhesin to adhere host enterocytes.

## 4 Concluding remarks and future perspectives

Molecular insight into bacterial adhesion to host receptors is significantly important to know the mechanism of delivering toxins and virulent factors, which will eventually lead to the development of a new antimicrobial therapy as an effective alternative to antibiotics. Antibiotics against bacterial infection is a must remedy where it is necessary to save life. WHO's 2014 report on global surveillance on antibiotic resistance unveil the truth that antibiotic resistance is no longer a future prediction but reality.

Most of the Gram-negative bacteria already developed resistance against available antibacterial drugs. There is no effective vaccine against enterotoxigenic and enteroaggregative *E. coli*, and *Yersinia* infection. It is already evident that bacterial fimbriae are highly immunogenic and consequently antibody titres raised against these fimbrial structures can protect bacterial colonization (Nuccitelli et al., 2011; Huesca et al., 2000).

CS6, AAF-1 and Myf alone or in combination with other virulence factors can be used in developing vaccines against respective pathogenesis. On the other hand, disarming the bacteria by designing small molecule inhibitors to prevent adhesion could be one of the finest alternatives. An adhesion inhibitor will neither kill the bacteria nor retard the growth, but rather stops adhesion to receptors and thus deprives the bacteria to use their molecular weapons to induce disease conditions. Inhibition or blocking of adhesion leads to no or less chance of evolving and spreading resistance to another species of bacteria (Duncan et al., 2012). So it is high time now to think about how to treat and control these deadly bacterial infections effectively.

The results presented here is a small fraction of the robust work performed by many researchers globally to elucidate the detailed assembly and receptor-binding mechanisms of CU assembled Gram-negative bacterial fimbriae. Many questions regarding assembly and binding still need to be answered. It is predicted in this thesis that CS6, AAF-1, and Myf make polyvalent attachment to the host cells regarding the participation of fimbrial subunit and host-receptors, which could be tighter than monovalent attachment.

Based on the atomic resolution 3D structure of both C<sub>ss</sub>AdsB and C<sub>ss</sub>BdsA, we assumed that strand C and D can make the binding site for biological function like adhesion. But it is not yet established which fragment of these strands or others strands are responsible for recognizing the host receptor(s) and what is the second receptor that is predicted in this thesis. The fibronectin-binding site is located on the variable strands C and their loop of AggA subunit of AAF-1 but AggA may recognize more receptors along with fibronectin. MyfA-sc recognizes lactose, isoglobotriose and the GM1 ganglioside but there are no atomic resolution structures of MyfA-sc with these mentioned molecules till to date. So, my recommendations for future are as follows:

- We need to explore further through NMR and X-ray crystallography together to reveal the binding site(s) of CS6, AAF-1 and Myf along with other receptors.
- Mutagenesis is recommended for CS6 and Myf to know the essential amino acids for recognizing receptors.
- Thorough screening through binding studies of these fimbriae to identify carbohydrate/protein/lipid molecules apart from discovered molecules is very much essential to list the responsible receptors for defining the adhesion mechanism.

## References

- Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., et al. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58: 1948–1954.
- Bao, R., Nair, M. K. M., Tang, W., Esser, L., Sadhukhan, A., Holland, R. L., Schifferli, D. M. (2013). Structural basis for the specific recognition of dual receptors by the homopolymeric pH 6 antigen (Psa) fimbriae of *Yersinia pestis*. *Proceedings of the National Academy of Sciences of the United States of America*, 110(3), 1065–1070. doi:10.1073/pnas.1212431110
- Bernier C, Gounon P, Le Bouguenec C (2002) Identification of an aggregative adhesion fimbria (AAF) type III-encoding operon in enteroaggregative *Escherichia coli* as a sensitive probe for detecting the AAF-encoding operon family. *Infect Immun* 70: 4302–4311.
- Boisen N, Struve C, Scheutz F, Krogfelt KA, Nataro JP (2008) New adhesin of enteroaggregative *Escherichia coli* related to the Afa/Dr/AAF family. *Infect Immun* 76: 3281–3292.
- Bottone EJ: . (1997). *Yersinia enterocolitica*: the charisma continues. *Clin Microbiol Rev*, 10:257-276.
- CCP4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50: 760–763
- Chapman, D.A., Zavialov, A.V., Chernovskaya, T.V., Karlyshev, A.V., Zav'yalova, G.A., Vasiliev, A.M., Dudich, I.V., Abramov, V.M., Zav'yalov, V.P. & MacIntyre, S. (1999). Structural and functional significance of the FGL sequence of the periplasmic chaperone Caf1M of *Yersinia pestis*. *J Bacteriol* 181(8), 2422- 9.
- Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S.L., and Knight, S.D. (1999) X-ray structure of the FimC–FimH chaperone–adhesin complex from uropathogenic *Escherichia coli*. *Science* 285: 1061– 1066.
- Czeczulin JR, Balepur S, Hicks S, Phillips A, Hall R, et al. (1997) Aggregative adherence fimbria II, a second fimbrial antigen mediating aggregative adherence in enteroaggregative *Escherichia coli*. *Infect Immun* 65: 4135–4145.
- Duncan E. McRee, (1999). XtalView/Xfit—A Versatile Program for Manipulating Atomic Coordinates and Electron Density, *Journal of Structural Biology*, Volume 125, Issues 2–3, Pages 156-165, ISSN 1047-8477

- Duncan MC, Linington RG, Auerbuch V. (2012). Chemical inhibitors of the type three secretion system: disarming bacterial pathogens. *Antimicrob Agents Chemother* 56:5433–5441. <http://dx.doi.org/10.1128/AAC.00975-12>.
- Emsley P, Cowtan K (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* 60, 2126–2132.
- Farfan MJ, Inman KG, Nataro JP (2008) The major pilin subunit of the AAF/II fimbriae from enteroaggregative *Escherichia coli* mediates binding to extracellular matrix proteins. *Infect Immun* 76: 4378–4384.
- Fisher-Walker CL and Black RE (2010) “Diarrhea morbidity and mortality in older children, adolescents and adults.” *Epidemiology and Infection* 138:1215–1226
- Frank C, Werber D, Cramer JP, Askar M, Faber M, et al. (2011) Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *N Engl J Med* 365: 1771–1780.
- Fredriksson-Ahomaa M, Korkeala H (2003) Low occurrence of pathogenic *Yersinia enterocolitica* in clinical, food, and environmental samples: a methodological problem. *Clin Microbiol Rev* 16:220–229
- Fronzes, R., Remaut, H. & Waksman, G. (2008). Architectures and biogenesis of non-flagellar protein appendages in Gram-negative bacteria. *EMBO J* 27(17), 2271–80.
- Gaastra, W., and Svennerholm, A.M. (1996) Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC). *Trends Microbiol* 4: 444–452
- GD Van Duyne, RF Standaert, PA Karplus, SL Schreiber, J Clardy. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *Journal of Molecular Biology*, 229, 105–124.
- Ghisaidoobe, Amar B. T., and Sang J. Chung. (2014). “Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Förster Resonance Energy Transfer Techniques.” Ed. Herbert Schneckenburger. *International Journal of Molecular Sciences* 15.12, 22518–22538. PMC. Web. 29 June 2015
- Ghosal, A., Bhowmick, R., Banerjee, R., Ganguly, S., Yamasaki, S., Ramamurthy, T., et al. (2009) Characterization and studies of the cellular interaction of native colonization factor CS6 purified from a clinical isolate of enterotoxigenic *Escherichia coli*. *Infect Immun* 77: 2125–2135.
- Giacovazzo, Carmelo & International Union of Crystallography. (2002). *Fundamentals of crystallography*, 2nd ed, Oxford University Press, Oxford ; New York
- Hahn, E., Wild, P., Hermanns, U., Sebbel, P., Glockshuber, R., Haner, M., Taschner, N., Burkhard, P., Aebi, U. & Muller, S.A. (2002). Exploring the 3D molecular architecture of *Escherichia coli* type I pili. *J Mol Biol* 323(5), 845–57
- Henderson, B., Nair, S., Pallas, J. and Williams, M. A. (2011), Fibronectin: a multidomain host adhesin targeted by bacterial fibronectin-binding proteins. *FEMS Microbiology Reviews*, 35: 147–200. doi: 10.1111/j.1574-6976.2010.00243.x
- Herbert L.D. (2005). Traveler’s diarrhea: antimicrobial therapy and chemoprevention. *Nature clinical practice gastroenterology & hepatology*, 2, 191–198, doi: 10.1038/ncpgasthep0142
- Holm L, Rosenstrom P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Research* 38: W545–549.
- Huesca M, Sun Q, Peralta R, Shivji GM, Sauder DN, et al. (2000). Synthetic Peptide Immunogens Elicit Polyclonal and Monoclonal Antibodies Specific for Linear Epitopes in the



- D Motifs of Staphylococcus aureus Fibronectin-Binding Protein, Which Are Composed of Amino Acids That Are Essential for Fibronectin Binding. *Infection and Immunity* 68: 1156–1163.
- Hung, C.-S., Bouckaert, J., Hung, D., Pinkner, J., Widberg, C., DeFusco, A., Auguste, C. G., Strouse, R., Langermann, S., Waksman, G. and Hultgren, S. J. (2002), Structural basis of tropism of *Escherichia coli* to the bladder during urinary tract infection. *Molecular Microbiology*, 44: 903–915.
- Iriarte M, Cornelis GR (1995) MyfF, an element of the network regulating the synthesis of fibrillae in *Yersinia enterocolitica*. *J Bacteriol* 177:738–744
- Iriarte M, Vanootehem JC, Delor I, Diaz R, Knutton S, Cornelis GR (1993) The Myf fibrillae of *Yersinia enterocolitica*. *Mol Microbiol* 9:507–520
- J. Bouckaert, J. Berglund, M. Schembri, E. De Genst, L. Cools, M. Wuhler, C.S. Hung, J. Pinkner, R. Slattegard, A. Zavialov, D. Choudhury, S. Langermann, S.J. Hultgren, L. Wyns, P. Klemm, S. Oscarson, S.D. Knight, H. De Greve. (2005), Receptor binding studies disclose a novel class of high-affinity inhibitors of the *Escherichia coli* FimH adhesin. *Mol. Microbiol.* 55 : 441–455.
- Janja Trček, Marc F Oellerich, Katy Niedung, Frank Ebel, Sandra Freund and Konrad Trülsch. (2011). Gut proteases target *Yersinia* invasin in vivo. *BMC Research Notes*, 4:129
- Jansson, L., Tobias, J., Jarefjall, C., Lebens, M., Svennerholm, A.M., and Teneberg, S. (2009) Sulfatide recognition by colonization factor antigen CS6 from enterotoxigenic *Escherichia coli*. *PLoS ONE* 4: e4487.
- Jennifer A. Mertz, Jameson A. McCann, William D. Picking (1996). Fluorescence Analysis of Galactose, Lactose, and Fucose interaction with the Cholera Toxin B Subunit. *Biochemical and Biophysical Research Communications*. Volume 226, Issue 1, Pages 140-144.
- Jiang ZD, Greenberg D, Nataro JP, Steffen R, DuPont HL (2002) Rate of occurrence and pathogenic effect of enteroaggregative *Escherichia coli* virulence factors in international travelers. *J Clin Microbiol* 40: 4185–4190.
- Jones, C.H., Danese, P.N., Pinkner, J.S., Silhavy, T.J. & Hultgren, S.J. (1997). The chaperone-assisted membrane release and folding pathway is sensed by two signal transduction systems. *EMBO J* 16(21), 6394-406.
- Jones, C.H., Pinkner, J.S., Roth, R., Heuser, J., Nicholes, A.V., Abraham, S.N. & Hultgren, S.J. (1995). FimH adhesin of type 1 pili is assembled into a fibrillar tip structure in the Enterobacteriaceae. *Proc Natl Acad Sci U S A* 92(6), 2081-5.
- Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 47 (Part 2): 110–119.
- Keyser P, Elofsson M, Rosell S, Wolf-Watz H. (2008). Virulence blockers as alternatives to antibiotics: type III secretion inhibitors against Gram- negative bacteria. *J. Intern. Med.* 264:17–29.
- Alexey S. Ladokhin, Sajith Jayasinghe, Stephen H. White. (2000). How to Measure and Analyze Tryptophan Fluorescence in Membranes Properly, and Why Bother? *Analytical Biochemistry*, Volume 285, Issue 2, Pages 235-245, ISSN 0003-2697
- Levine MM, Ristaino P, Marley G, Smyth C, Knutton S, Boedeker E, Black R, Young C, Clements ML, Cheney C, Patnaik R. (1984). Coli surface antigens 1 and 3 of colonization

- factor antigen II-positive enterotoxigenic *Escherichia coli*: morphology, purification, and immune responses in humans. *Infect Immun* 44:409–420
- Levy SB, Marshall B. (2004). Antibacterial resistance worldwide: causes, challenges and responses. *Nat. Med.* 10:S122–S129.
- Ludi, S., Frey, J., Favre, D., and Stoffel, M.H. (2006) Assessing the expression of enterotoxigenic *Escherichia coli*-specific surface antigens in recombinant strains by transmission electron microscopy and immunolabeling. *J Histochem Cytochem* 54: 473–477.
- M. Fredriksson-Ahomaa & N. Cernela & H. Hächler & R. Stephan. (2012) *Yersinia enterocolitica* strains associated with human infections in Switzerland 2001–2010. *Eur J Clin Microbiol Infect Dis* 31:15–1550 DOI 10.1007/s10096-011-1476-7
- Mathewson JJ, Jiang ZD, Zumla A, Chintu C, Luo N, et al. (1995) HEp-2 cell-adherent *Escherichia coli* in patients with human immunodeficiency virus-associated diarrhea. *J Infect Dis* 171: 1636–1639.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., & Read, R.J. (2007). Phaser crystallographic software. *Appl. Cryst.* 40, 658-674.
- Murshudov G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography* 67, 355-367.
- Natalia Pakharukova, Saumendra Roy, Minna Tuittila, Sarri Paavilainen, Anna-Karin Ingars, Torleif Härd, Susann Teneberg, and Anton V.Zavialov. Structural basis for Myf and pH6 fimbriae mediated tropism of pathogenic strains of *Yersinia*. (Manuscript)
- Nataniel Białas, Katarzyna Kasperkiewicz, Joanna Radziejewska-Lebrecht, Mikael Skurnik. Bacterial Cell Surface Structures in *Yersinia enterocolitica*. *Archivum Immunologiae et Therapiae Experimentalis*, (2012), Volume 60, Number 3, Page 199
- Nataro JP, Mai V, Johnson J, Blackwelder WC, Heimer R, Tirrell S, Edberg SC, Braden CR, Glenn Morris J Jr, Hirshon JM. (2006). Diarrheagenic *Escherichia coli* infection in Baltimore, Maryland, and New Haven, Connecticut. *Clin Infect Dis* 43:402–407
- Nataro JP, Yikang D, Giron JA, Savarino SJ, Kothary MH, et al. (1993) Aggregative adherence fimbria I expression in enteroaggregative *Escherichia coli* requires two unlinked plasmid regions. *Infect Immun* 61: 1126–1131.
- Nuccitelli A, Cozzi R, Gourlay LJ, Donnarumma D, Necchi F, et al. (2011) Structure-based approach to rationally design a chimeric protein for an effective vaccine against Group B *Streptococcus* infections. *Proceedings of the National Academy of Sciences* 108: 10278–10283.
- Ofek, I., Hasty, D. L. and Doyle, R. J. (2003) *Bacterial Adhesion to Animal Cells and Tissues*, ASM Press, Washington.
- Pablo C. Okhuysen and Herbert L. DuPont (2010). Enteroaggregative *Escherichia coli* (EAEC): A Cause of Acute and Persistent Diarrhea of Worldwide Importance. *J Infect Dis.* 202 (4): 503-505 doi:10.1086/654895
- Pakharukova, N., Tuittila, M. & Zavialov, A. (2013). Crystallization and sulfur SAD phasing of AggA, the major subunit of aggregative adherence fimbriae type I from the *Escherichia coli* strain that caused an outbreak of haemolytic-uraemic syndrome in Germany. *Acta Cryst.* F69, 1389-1392.

- Payne, D., Tatham, D., Williamson, E. D., & Titball, R. W. (1998). The pH 6 Antigen of *Yersinia pestis* Binds to  $\beta$ 1-Linked Galactosyl Residues in Glycosphingolipids. *Infection and Immunity*, 66(9), 4545–4548.
- Piatek, R., Bruzdziak, P., Zalewska-Piatek, B., Kur, J., and Stangret, J. (2009) Preclusion of irreversible destruction of Dr adhesin structures by a high activation barrier for the unfolding stage of the fimbrial DraE subunit. *Biochemistry* 48: 11807–11816.
- Proft, T. & Baker, E.N. (2009). Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease. *Cell Mol Life Sci* 66(4), 613-35.
- Qadri, F., Svennerholm, A.M., Faruque, A.S., and Sack, R.B. (2005) Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clin Microbiol Rev* 18: 465–483.
- Remaut, H., Rose, R.J., Hannan, T.J., Hultgren, S.J., Radford, S.E., Ashcroft, A.E. & Waksman, G. (2006). Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted beta strand displacement mechanism. *Mol Cell* 22(6), 831-42.
- Rose, R.J., Verger, D., Daviter, T., Remaut, H., Paci, E., Waksman, G., et al. (2008) Unraveling the molecular basis of subunit specificity in P pilus assembly by mass spectrometry. *Proc Natl Acad Sci USA* 105: 12873–12878
- Rosner BM, Stark K, Werber D (2010) Epidemiology of reported *Yersinia enterocolitica* infections in Germany, 2001–2008. *BMC Publ Health* 10:337
- Sauer, F.G., Futterer, K., Pinkner, J.S., Dodson, K.W., Hultgren, S.J. & Waksman, G. (1999). Structural basis of chaperone function and pilus biogenesis. *Science* 285(5430), 1058-61.
- Sauer, F.G., Pinkner, J.S., Waksman, G. & Hultgren, S.J. (2002). Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation. *Cell* 111(4), 543-51.
- Schuck P. (1997) Use of surface plasmon resonance to probe the equilibrium and dynamic aspects of interactions between biological macromolecules. *Annu. Rev. Biophys. Struct.* 26:541–566).
- Tobias, J., Lebens, M., Kallgard, S., Nicklasson, M., and Svennerholm, A.M. (2008) Role of different genes in the CS6 operon for surface expression of Enterotoxigenic *Escherichia coli* colonization factor CS6. *Vaccine* 26: 5373–5380.
- Turnbull, W. B.; Precious, B. L.; Homans, S. W. (2004), Dissecting the Cholera Toxin–Ganglioside GM1 Interaction by Isothermal Titration Calorimetry. *J. Am. Chem. Soc.* 126, 1047
- Vetsch, M., Puorger, C., Spirig, T., Grauschopf, U., Weber-Ban, E.U. & Glockshuber, R. (2004). Pilus chaperones represent a new type of protein-folding catalyst. *Nature* 431(7006), 329-33
- Waksman, G. & Hultgren, S.J. (2009). Structural biology of the chaperone-usher pathway of pilus biogenesis. *Nat Rev Microbiol* 7(11), 765-74
- WHO (2006) Future directions for research on enterotoxigenic *Escherichia coli* vaccines for developing countries. *Wkly Epidemiol Rec* 81: 97–104.
- WHO 2013 Media centre, Diarrheal disease, Factsheet no. 330
- WHO, 2014, Media centre, Plague, Factsheet no. 267
- WHO, 2015, Media centre, Cholera, Factsheet no. 107
- Wolf, M.K., de Haan, L.A., Cassels, F.J., Willshaw, G.A., Warren, R., Boedeker, E.C., and Gastra, W. (1997) The CS6 colonization factor of human enterotoxigenic *Escherichia coli* contains two heterologous major subunits. *FEMS Microbiol Lett* 148: 35–42.

- Wright, K.J., Seed, P.C. & Hultgren, S.J. (2007). Development of intracellular bacterial communities of uropathogenic *Escherichia coli* depends on type 1 pili. *Cell Microbiol* 9(9), 2230-41.
- [www.cdc.gov/ecoli/general/index/html](http://www.cdc.gov/ecoli/general/index/html) as of 2015.09.20
- [www.rcsb.org/pdb/statistics/holdings.do](http://www.rcsb.org/pdb/statistics/holdings.do) as of 2015-06-18
- Yu, X.D., Fooks, L.J., Moslehi-Mohebi, E., Tischenko, V.M., Askarieh, G., Knight, S.D., MacIntyre, S. & Zavialov, A.V. (2012). Large is fast, small is tight: determinants of speed and affinity in subunit capture by a periplasmic chaperone. *J Mol Biol* 417(4), 294-308.
- Zav'yalov, V., Zavialov, A., Zav'yalova, G., and Korpela, T. (2010) Adhesive organelles of Gram-negative pathogens assembled with the classical chaperone/usher machinery: structure and function from a clinical standpoint. *FEMS Microbiol Rev* 34: 317–378.
- Zavialov, A., Zav'yalova, G., Korpela, T., and Zav'yalov, V. (2007) FGL chaperone-assembled fimbrial polyadhesins: anti-immune armament of Gram-negative bacterial pathogens. *FEMS Microbiol Rev* 31: 478–514.
- Zavialov, A.V., Tischenko, V.M., Fooks, L.J., Brandsdal, B.O., Aqvist, J., Zav'yalov, V.P., MacIntyre, S. & Knight, S.D. (2005). Resolving the energy paradox of chaperone/usher-mediated fibre assembly. *Biochem J* 389(Pt 3), 685-94.
- Zavialov AV, Berglund J, Pudney AF, Fooks LJ, Ibrahim TM, et al. (2003) Structure and biogenesis of the capsular F1 antigen from *Yersinia pestis*: preserved folding energy drives fiber formation. *Cell* 113: 587–596.
- Zavialov, A.V., Kersley, J., Korpela, T., Zav'yalov, V.P., MacIntyre, S. & Knight, S.D. (2002). Donor strand complementation mechanism in the biogenesis of non-pilus systems. *Mol Microbiol* 45(4), 983-95

## Acknowledgements

Since 1<sup>st</sup> of July 2011 to today the 26<sup>th</sup> November 2015, an amazing journey has come to its very end. From the beginning to end, I have had the pleasure to meet a lot of kind people to whom I would like to be grateful.

**Anton Zavalov**, primary supervisor of my graduate study for first two years, inspired and brought me into the fantastic world of protein chemistry and x-ray crystallography. Long hours discussions, sometimes even in late night at your residence, of running projects and future perspectives were not always pleasant but definitely were caring to educate me. Though at the beginning, my biochemistry knowledge was far bellow the acceptable level but, finally, you who made me a Biochemist. It was wonderful experience to work with you. Heartfelt thanks to you, **Anton!**

**Torleif Hård**, primary supervisor of my graduate study for last two years, who cared me a lot even though fimbrial research was not his field of interest. He spent plenty of his valuable time to guide me and focused on my proper education without any delay and compromise. I am very much grateful to you, **Torleif!**

**Mats Sandgren** has taken all the pivotal steps through this entire journey of graduate study to put me in this very last day. Thanks **Mats** for keeping me away from anxiety for funding.

I would like to thank my co-supervisor **Jerry Ståhlberg** whom I found tremendously helpful and caring. Your invaluable guidance and support in writing a research article will always keep me in right track. I would like to offer warm thanks to **Stefan Knight** for the help at the key point of my career.

**Xiaodi Yu, Glareh Askarieh, and Nils Mikkelsen** were always with me since my master degree project. Whenever I needed advice, felt lost or had the opportunity to consult someone seniors, I turned to you. You were my true mentors (Guru) in wet lab and x-ray crystallography. This study would not have been successful without your presence. It has been my good fortune to work with you. I am offering a big hug for you guys. **Glareh**, keep your biggest laugh on to keep science go on.

When I needed a coffee break to share some happy and unhappy moments, **Anna Borisova** was always there with me. Thank you mate for making things easier and thanks for the candy. Wish you a successful career and happiest life ahead. This experience would not have been the same without great colleagues and friends. There is no work in science like teamwork. My greatest thanks to, **Sultana, Majid, Benjamin Schmuck, Jule, Mahafuzur Rahman, Bing, Mikael Gudmundsson, Henrik Hansson, Saied Karkehabadi, Miao Wu, Li Zhang, Wangshu Jiang, Lu Lu, Sanjeevani, Evalena Andersson, Christofer Lendel, Maria Dimarogona, Anatoly Dubnovitsky, Andreas Digre, Roland Bergdahl, Mahmud, Mubin, Minna Tuittila, Natalia P, Tobias, Martin** and all other colleagues from the department of **Chemistry and biotechnology, SLU**.

I am grateful to **Sonja Jansson** and **Ellenor** (rtd) for all the administrative assistance.

I would like to express my great appreciation and warm thanks to **Rajotda, Mamun, Nishat, Ashim, Abdullah Suman, Tuntun, Ashish, Habib, Butchchuvai, Khosruvai, Shantu, Asad, Broadleaf-49, Suman** and the **Someone** who in many ways contributed in my life.

Life is not only for work and science. I am very much fortunate to have Bangladeshi fantastic friends here in abroad in my personal and social life who in many ways stand beside me. I am truly grateful to them, especially to **Shuvra Das. Sunju**, you are the best. God bless you.

I am particularly indebted to all my **Teachers** from Konda Primary School (Bangladesh) to SLU & UU (Sweden) who were and are simply the best.

মা, it would not have been possible without your support. I am proud to have you in my life. My loving **mami, babu kaka,** and **mamami,** your endless love and guidance brought me to this day of success. It would be unfair if I do not appreciate and thank **Ratna** for her cosy hosting in California.

মা, বাবা: my deepest gratitude to you what is impossible to describe in words. I am very much glad to dedicate this thesis to my parents. My father unfortunately passed away two years ago. I wish I could bring my parents in Sweden and they would have been experience the biggest moment of my life! I am extremely grateful to my dearest **brothers, sisters** including **boudi** for always being supportive to complete my doctoral study successfully. **Poly, Moly, Bithi, Juthi, Momo, Arup, Protya, Tusi, Tathoi,** and **Tonu:** your love always encourages me to do better. Kids! I will be with you like the way your parents were with me.

My soul mate, **Mukta Roy,** has sacrificed her academic career with a big smile and taken care of everything without being mad on me when I have been working in lab day and nights long. This achievement is for you and because of you. You are the divine gift for me. I do love you.

**Urmila** and **Chanchal!** Always being right next to me, no matter how far away. Lots of love my kids.

**Rishita Roy,** my daughter, who is the most important than anything else in my life. Your smile makes me energetic when I get home with exhausted soul. Now we will play together all day long, it is a promise.

*Roy Saumendra*

Uppsala, 20<sup>th</sup> October 2015