

**Structure - Function Studies of
Organelle Assembly and Receptor
Recognition in Organelles Assembled
via the Chaperone/Usher Pathway**

Jenny Berglund

*Department of Molecular Biology
Uppsala*

**Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2004**

Acta Universitatis Agriculturae Sueciae
Agraria 441

ISSN 1401-6249
ISBN 91-576-6492-7
© 2004 Jenny Berglund, Uppsala
Tryck: SLU Service/Repro, Uppsala 2004

Abstract

Berglund, J., 2004. *Structure-function studies of organelle assembly and receptor recognition in organelles assembled via the chaperone/usher pathway*. Doctoral dissertation.

ISSN 1401-6249, ISBN 91-576-6492-7.

Adhesion is an important first step in infection, where the microorganism attaches to a host cell. In many cases adhesion is mediated by fimbriae, or pili; hairlike organelles composed of a large number of protein subunits, protruding out from the bacterial surface. In this thesis, the adhesion and assembly of two such organelles has been studied: type-1 pili from *Escherichia coli* and the capsular F1 antigen from *Yersinia pestis*.

The adhesin molecule of type-1 pili is FimH, and the structure of the FimH lectin domain was determined to 2.3 Å. High structural similarity to the same domain in the FimH:FimC adhesin:chaperone structure shows rigidity and structural independency of the lectin domain. In the crystal structure a butyl mannoside was discovered in the FimH binding site. Binding studies of alkyl mannosides and aryl mannoside show that *E. coli* FimH recognizes these two classes of compounds with high affinity. Using a series of trimannosides corresponding to structures present in *N*-linked high-mannose glycoproteins, the binding properties of FimH from two different UPEC and one fecal *E. coli* strains were investigated. Our results suggest that the differences in adhesion phenotype mediated by these different adhesins are caused by differences in adhesin presentation rather than by affinity differences.

The antiphagocytic capsule around *Yersinia pestis* is constructed from multiple copies of the CafI subunit assembled into thin fibres. The structure of the CafIM:CafI chaperone:subunit binary complex and the CafIM:CafI:CafI ternary complex from *Y. pestis* was determined. Comparison of the chaperone bound CafI subunit with the CafI fibre subunit revealed that the CafIM chaperone jams the folding of CafI in a high-energy conformation with a poorly packed hydrophobic core. When the chaperone dissociates and is replaced by the donor strand from the next subunit, folding is allowed to continue to completion. The folding energy released in this step is proposed to drive fibre assembly.

Keywords: carbohydrate binding, lectin, urinary tract infection, plague, antiphagocytic capsule, pathogenesis, donor strand complementation, donor strand exchange, immunoglobulin fold, vaccine, protein crystallography.

Author's address: Jenny Berglund, Department of Molecular Biology, SLU, Uppsala Biomedical Center, PO-Box 590, 751 24 Uppsala, Sweden

Contents

1. Background	7
1.1 General introduction	7
1.2 Secretion pathways	7
1.3 The chaperone/usher pathway	9
1.3.1 Introduction	9
1.3.2 Chaperone structure	10
1.3.3 Donor strand complementation	11
1.4 Urinary tract infections	14
1.4.1 Type-1 pili	14
1.5 Plague and F1 antigen	15
1.6 Aim and outline of this thesis	17
2 Methods	18
2.1 X-ray crystallography	18
2.1.1 Introduction	18
2.1.2 Determining the structure of the Caf1:Caf1M binary complex	18
2.2 Dot-blot	20
2.3 Docking	21
2.4 Mass spectrometry	21
2.5 Determination of binding constants	22
3 Results and discussion	24
3.1 FimH lectin domain (Paper III)	24
3.1.1 Introduction	24
3.1.2 The structure of FimH lectin domain	24
3.1.3 Butyl mannoside	26
3.1.4 Binding studies of alkyl mannosides	28
3.1.5 Aromatically substituted mannosides	29
3.1.6 Different strains of <i>E. coli</i> show different binding phenotypes	31
3.1.7 FimH affinity to trimannosides	32
3.1.8 Why does shear force promote monomannose high binding?	33
3.2 F1 antigen (Paper I and II)	35
3.2.1 The Caf1M:Caf1 binary complex	35
3.2.2 The Caf1M:Caf1:Caf1 ternary complex	37
3.2.3 Folding energy is preserved by the chaperone	39
3.3 The immunoglobulin fold in organelle subunits	42
3.3.1 Introduction	42
3.3.2 The immunoglobulin fold in organelle subunits	42
3.3.3 Aligning the pilin subunits	44
3.3.4 The lectin domains	47

4. Future perspectives	49
4.1 <i>Structure and sequence comparison of the organelle subunits</i>	49
4.2 <i>The usher</i>	49
4.3 <i>Binding phenotypes of FimH</i>	49
5. References	51
6. Acknowledgments	55

Appendix

Papers I-III

This thesis is based on the following papers, which in the thesis will be referred to by their Roman numerals:

- I. Zavialov, A., Berglund, J., and Knight, S. D. (2003a). Overexpression, purification, crystallization and preliminary X-ray diffraction analysis of the F1 antigen Caf1M-Caf1 chaperone-subunit pre-assembly complex from *Yersinia pestis*. *Acta Crystallogr D Biol Crystallogr* 59, 359-362.
- II. Zavialov, A. V., Berglund, J., Pudney, A. F., Fooks, L. J., Ibrahim, T. M., MacIntyre, S., and Knight, S. D. (2003b). Structure and biogenesis of the capsular F1 antigen from *Yersinia pestis*: preserved folding energy drives fiber formation. *Cell* 113, 587-596.
- III. Berglund, J., Schembri, M., Klemm, P., Zavialov, A., Choudhury, D., Oscarsson S., Knight, S.D. Structure and receptor binding properties of the FimH lectin domain from uropathogenic and fecal strains of *Escherichia coli*. *In manuscript*.

Article I and II are reprinted by permission of the journal concerned.

Related publications

- i. Knight S.D., Berglund J. Choudhury D. (2000). Bacterial adhesions: structural studies reveal chaperone function and pilus biogenesis. *Curr Opin Chem Biol*. 4 (6), 653-60.
- ii. Berglund J., Knight S.D., (2003). Structural basis of bacterial adhesion in the urinary tract. *Glycoimmunology* 3 (Advances in Experimental Medicine and Biology Series), JS Axford, ed., pp 33-52, Kluwer Academic/Plenum Publishers, New York.

1. Background

1.1 General introduction

Bacteria require nutrients in order to survive, and they have evolved diverse strategies in order to facilitate their search for these nutrients. One example is organelles for motility, allowing the microorganism to move towards a higher concentration of the nutrient. The ability to attach to various receptors on surfaces and/or other cells is another important development since this permits a more permanent attachment close to a nutrient source.

Attachment of bacteria to a surface is mediated by adhesins displayed on the surface of the bacteria, and the importance of this step is revealed in the number of adhesins expressed by both Gram-positive and Gram-negative bacteria. Gram-positive bacteria have a peptidoglycan cell wall on the outside of the cytoplasmic membrane and adhesins are anchored to the cell surface either by attachment to the peptidoglycan matrix, or by anchoring to the cytoplasmic membrane. Gram-negative bacteria instead have two membranes, the inner and outer membrane separated by the periplasmic space, with a peptidoglycan cell wall on the inside of the outer membrane. The anchoring of adhesins is therefore different and the most common adhesion structure for Gram-negative bacteria is the fimbrium, or pilus (Ofek et al., 2003), a hairline organelle assembled from many protein subunits, protruding out from the bacterial surface.

Most fimbriae attach to carbohydrate receptors on the surface of a host cell, either by the subunit making up the bulk of the fimbriae or by a specialized lectin at the tip. Different surfaces of a bacterial host have different types of receptors, which are recognized by distinct types of fimbriae. In this way the type of adhesin expressed on the bacterial surface can provide an important tool for the bacteria when selecting a tissue for invasion - the bacteria show tissue tropism.

When assembling an organelle on the bacterial surface, the bacteria are faced with a problem of transporting the subunits to be assembled across the cell membrane. For this purpose a number of secretion systems have evolved, each one responsible for the assembly of a certain type of organelle.

1.2 Secretion pathways

Secretion systems developed by Gram-negative bacteria are used for secretion of surface organelles, toxins as well as nucleic acids. The secretion systems can be

roughly divided into two groups, the general secretory pathway (GSP) and Sec-independent secretion.

Sec translocase independent secretion pathways are exemplified by type I, type III and in part type IV secretion. Type I secretion systems, also known as ATP binding cassette (ABC) exporters (Binet et al., 1997, Stathopoulos et al., 2000), allow translocation from the cytoplasm and through the two membranes without a periplasmic intermediate. It consists of three parts: an inner membrane (IM) spanning ABC exporter, an IM anchored protein that extends out through the periplasm called membrane fusion protein (MFP), and an outer membrane (OM) channel-forming protein (OMP). The proteins to be exported by this pathway contain a carboxy-terminal peptide that targets the ABC exporter.

The type III secretion system is a complex apparatus (Hueck, 1998) which enables the bacteria to translocate antihost factors directly into the cytosol of the target eukaryotic cell. It is built up from approximately 20 different types of proteins, most of which are located at the IM, and show high resemblance to the flagellar export system. The type III apparatus extends from the IM, over the periplasmic space, through the OM to the outside of the bacteria where it forms a hollow organelle that can connect the bacteria to the eukaryotic cell. Secretion of *Yersinia* outer proteins (Yops) represents the prototypical type III export pathway.

The general secretory pathway (GSP) utilises the Sec secretion machinery to transport protein over the inner membrane. The protein sequence is targeted for Sec secretion by an amino-terminal signal peptide. Proteins are then transported across the IM in an unfolded state, and the signal peptide is cleaved off upon entry into the periplasm. Three terminal branches finish the GSP and are responsible for the transport over the outer membrane: type II secretion, the autotransporter pathway and the chaperone/usher pathway.

The main terminal branch of the GSP is the type II secretion pathway (Stathopoulos et al., 2000, Thanassi and Hultgren, 2000b, Nunn, 1999), which is closely related to the biogenesis of type IV pili. This is a quite complex pathway, which requires between 12 and 16 accessory proteins that together form a secreton. The secreton spans from the IM to the OM, with a pore formed through the latter. Type II secretion is ATP dependent and utilises energy from the cytosolic side of the IM to export proteins through the OM pore.

The autotransporter pathway (Henderson et al., 1998, Thanassi and Hultgren, 2000b) has everything needed for transport packaged into one protein. This protein consists of three parts: an amino-terminal signal peptide, an internal passenger

domain and a carboxy-terminal β -domain. The amino-terminal signal peptide targets the protein to the Sec machinery to ensure IM passage. The β -domain is predicted to fold into a beta-barrel, which inserts into the OM to form a pore through the membrane. The passenger domain is then believed to pass through this pore, where it is either retained at the surface, or cleaved off and released. The process requires no input of external energy.

The third terminal branch of the GSP is the chaperone/usher pathway, which utilises two accessory proteins for membrane transport, a chaperone and an usher (Thanassi et al., 1998a, Thanassi and Hultgren, 2000a, Zavialov et al., 2001, Berglund and Knight, 2003, Ofek et al., 2003). The chaperone/usher pathway is used for assembly of a number of surface organelles, two of which have been studied in this thesis.

1.3 The chaperone/usher pathway

1.3.1 Introduction

Gram-negative bacteria express fimbrial organelles on their surface, which are typically used for attachment to host cells. The most common mechanism for fimbrial biogenesis is the chaperone/usher pathway (Thanassi et al., 1998a, Sauer et al., 2000, Thanassi and Hultgren, 2000a, Ofek et al., 2003). In this pathway, the protein subunits are transported from the inner to the outer membrane, where they are joined together to form a growing chain, which is subsequently translocated through the outer membrane to the surface of the bacteria. The subunits have to be protected from premature aggregation in the periplasm, as well as inserted in the correct order into the organelle. A chaperone/usher pair directs this process.

The protein subunits of the chaperone/usher pathway are transported over the inner membrane by the Sec-machinery, where the signal peptide is cleaved off. In the periplasm, the subunits remain associated with the membrane in a semi-unfolded state. The chaperone retrieves the subunits from the membrane, and is thought to aid in their folding and simultaneously cap the interactive surface to prevent aggregates from being formed (Figure 1.1). If the chaperone is absent the subunits cannot fold properly and form aggregates in the periplasm, which are later degraded by proteases.

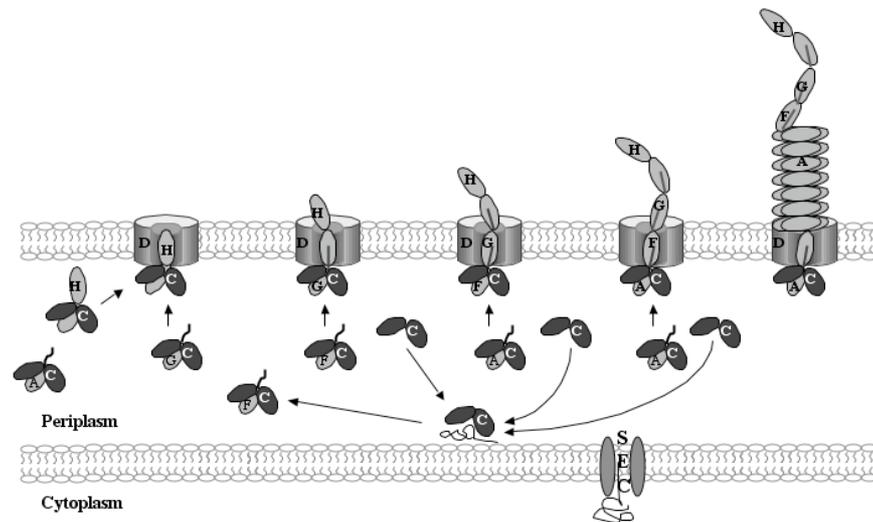


Figure 1.1: Schematic picture of type-1 pilus assembly by the chaperone/usher pathway. The chaperone FimC picks up the subunits at the inner membrane, assists in their folding and delivers them to the usher, where they are incorporated into the growing pilus.

The chaperone:subunit complexes are then targeted to the usher, located in the outer membrane. The usher is thought to form a pore-like structure, through the membrane, with an inner diameter of 2-3nm, large enough for the folded subunits to pass through (Thanassi et al., 1998b). The subunits are assembled at the usher, and the growing chain translocated through the usher pore (Figure 1.1). The process does not require input of external energy (Jacob-Dubuisson et al., 1994).

1.3.2 Chaperone structure

PapD is the prototypic chaperone of a large chaperone family, and is part of the assembly machinery for P pili from *Escherichia coli*. The structure was solved in 1989 (Holmgren and Branden, 1989), and has two immunoglobulin-like beta barrels joined together at an approximate 90° angle (Figure 1.2A). The arrangement of the two domains creates a cleft between them, in which two invariant, positively charged amino acids, Arg8 and Lys112, are located (Figure 1.2B). All of the chaperones contain a conserved inter domain hydrogen-bonding network, and a highly conserved beta sheet in the N-terminal domain (Hung et al., 1996).

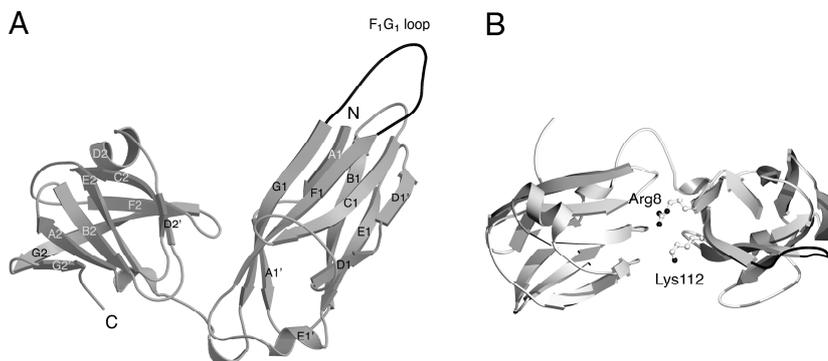


Figure 1.2: Chaperone PapD. (A) Front view of the two domains of PapD. (B) Top view with PapD rotated 90°, and the conserved residues Arg8 and Lys112 shown in ball-and-stick. Figure prepared using Molscript (Kraulis, 1991).

Sequence analysis of the chaperone family (Hung et al., 1996) revealed high sequence similarity between the 26 members of the family known at that time (25-56% sequence identity). It also demonstrated that the chaperone family could be divided into two groups depending on features of the loop connecting the F₁ strand with the G₁ strand. In the first group this loop is quite short, giving the chaperones the name FGS (FG loop short) chaperones. The second group has a long F₁-G₁ loop, and is thereby given the FGL (FG loop long) epithet. Interestingly, these two groups of chaperones seem to assemble two structurally distinct types of organelles. The FGS chaperones assemble rigid, complex pili, consisting of several different types of subunits, with one ultimate subunit dedicated to adhesion. The FGL chaperones on the other hand assemble more simple, non-pilus structures, often consisting of one or at most two types of subunits, which share the dual role of being both a structural subunit and an adhesin.

The structures of several chaperones have been determined after PapD: FimC (type-1) (Choudhury et al., 1999, Pellecchia et al., 1998), SfaE (S pili) (Knight et al., 2002), and one FGL chaperone from a non-pilus system, Caf1M from the F1 capsule (Paper II). The structures reveal that the sequence similarity as expected also results in a high structural similarity.

1.3.3 Donor strand complementation

Fimbrial subunits are not stable on their own, but if co-expressed with the chaperone they form stable, soluble complexes. In this way two chaperone:subunit complexes could be purified and crystallised, FimC:FimH chaperone:adhesin

complex from type-1 pili and PapD: PapK chaperone: subunit complex from P pili. The two structures were solved in 1999 (Sauer et al., 1999, Choudhury et al., 1999). The structures revealed both the basis for the strong interactions of a chaperone: subunit complex, and also provided a model for how the subunits in a pilus are linked together.

FimH is a two-domain protein, with an amino-terminal lectin domain and a carboxy-terminal pilin domain. The lectin domain is an 11-stranded beta barrel responsible for carbohydrate binding, and the pilin domain is a smaller 6-stranded beta barrel, which makes contacts with the chaperone in the complex, and is later thought to mediate contacts with the next subunit in the pilus. PapK is a linker protein from the flexible tip of P pili. It is a one-domain protein with a similar fold to the pilin domain of FimH.

PapK and the FimH pilin domain both consist of an immunoglobulin-like beta barrel, except the 7th strand is missing. This creates a cleft in the barrel between strand A and F, exposing part of the hydrophobic core. In order to “repair” the barrel, the chaperone donates its G₁-strand to the subunit by inserting this strand into the cleft. This complements the immunoglobulin fold in a mechanism termed donor strand complementation, DSC (Sauer et al., 1999, Choudhury et al., 1999) (Figure 1.3).

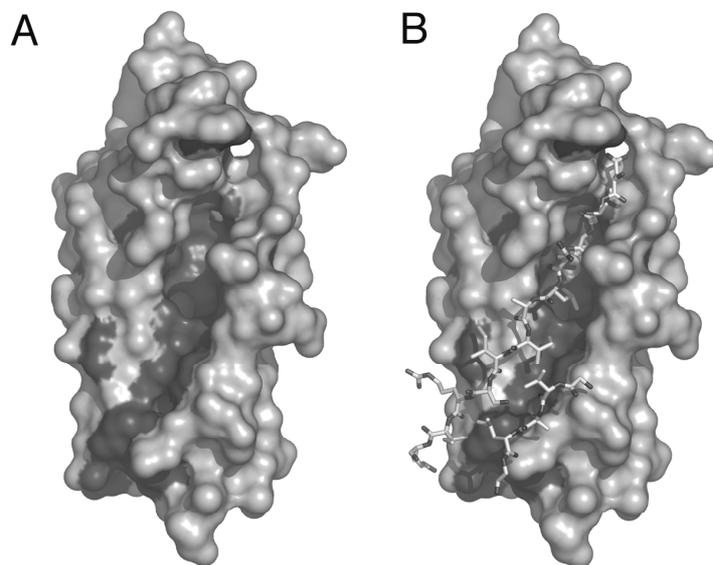


Figure 1.3 Donor strand complementation. The hydrophobic side chains of the core are marked in dark grey. (A) Figure of the cleft in the barrel between strand A and F, exposing part of the hydrophobic core. (B) The chaperone G₁ and A₁ strand are inserted into the cleft, complementing the fold.

The cleft is part of the subunit assembly surface (see the following part) and by capping the cleft of the subunits the chaperone is able to prevent premature aggregation, and can guide subunits to the scene of assembly, the usher. The incomplete fold and the exposure of part of the hydrophobic core also explain why the subunits are unstable without the chaperone.

The discovery of donor strand complementation led to a hypothesis on how the subunits in a pilus are linked together. All pilin subunits have an N-terminal extension with a conserved pattern of alternating hydrophobic residues. This extension was disordered in the PapD: PapK complex, and is not part of the beta-barrel. The extended strand was suggested to replace the chaperone G_1 strand of the preceding subunit in the pilus, thereby complementing the immunoglobulin fold of this subunit in a mechanism termed donor strand exchange (DSE) (Figure 1.4).

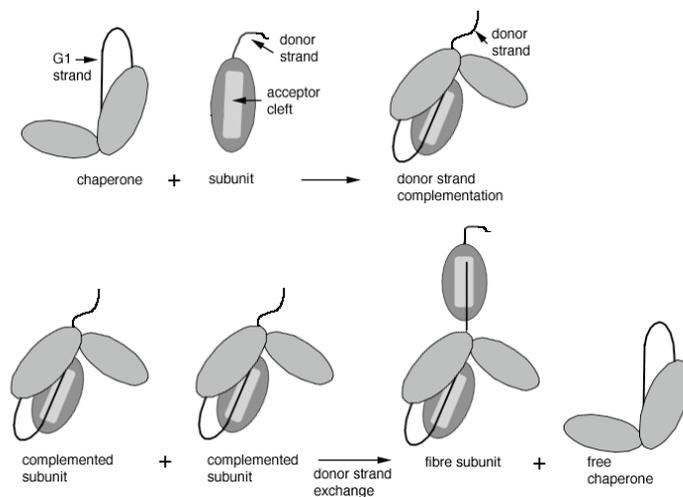


Figure 1.4 Schematic figure of donor strand exchange, where the chaperone G_1 strand is replaced by the subunit G_{donor} strand.

FimH does not have an N-terminal donor strand extension attached to its pilin domain, but instead has a whole domain at this position, the lectin domain. Following from the model described above, FimH can never complement another subunit, but can only hold one ultimate position at the tip of the pilus.

In this thesis subunits from two types of organelles have been studied, type-1 pili from uropathogenic *E. coli* (an FGS system) (Paper III), and the capsular F1 antigen from *Yersinia pestis*, the causative agent of plague (an FGL system) (Paper I and Paper II).

1.4 Urinary tract infections

Attachment of bacteria to their host cells is a first and crucial step in a number of diseases, for example urinary tract infections (UTI:s). Although not lethal, UTI:s are unpleasant diseases, and are responsible for a medical cost of around \$2 billion in the US alone (Foxman, 2002). 50% of all women are affected at some time in their life (Foxman, 2002), and in many cases recurrent infections occur. Uropathogenic *E. coli* (UPEC) are the most common cause of infection, responsible for 80% of the reported cases (Ronald, 2002). UTI:s can be treated with antibiotics, but resistant strains and recurrent infections are becoming an increasing problem. Presently no vaccine is available.

A number of UPEC adhesive organelles have been reported as virulence factors for UTI (Table 1.1). Different organelles show specificity for receptors in different tissues of the host, and are therefore involved in different classes of UTI. Type-1 pili, present on 90% of all *E. coli* strains, have been shown to be critical for establishment of cystitis (Bahrani-Mougeot et al., 2002, Connell et al., 1996). Type-1 pili bind to Uroplakin 1a, a mannose containing receptor on the uroepithelium of the bladder (Wu et al., 1996, Zhou et al., 2001, Min et al., 2002). P pili bind to receptors in the upper urinary tract causing pyelonephritis. F1C and S pili, which are related to each other and also show similarities to the type-1 system, bind to galactosyl ceramide and globotriaosyl ceramide containing receptors (Backhed et al., 2002, Khan et al., 2000) and to sialic-acid-containing receptors (Korhonen et al., 1984, Hanisch et al., 1993). They are associated with ascending UTI. The Dr family of adhesive organelles (Nowicki et al., 2001) do not appear to recognise carbohydrate structures, but bind to the Dr(a+) antigen on the surface of the decay-accelerating factor (DAF;CD55). Dr adhesins are associated with cystitis and pyelonephritis.

Table 1.1 UTI-associated adhesion organelles

Organelle	UTI	Adhesin	Receptor	System
Type-1 pilus	Cystitis	FimH	Mannose	FGS
P pilus	Pyelonephritis	PapG	Galabiose	FGS
F1C pilus	Ascending UTI	FocH	Galactose	FGS
S pilus	Ascending UTI	SfaS, SfaH?	Sialic acid, Galactose?	FGS
Dr adhesins (non-pilus)	Cystitis, Pyelonephritis	DraE	DAF, type IV collagen	FGL

1.4.1 Type-1 pili

Type-1 pili and P pili, are the most studied pili systems. Type-1 pili mediate binding to the glyco-protein Uroplakin 1a, which is glycosylated in one position

with a high mannose (Wu et al., 1996, Zhou et al., 2001, Min et al., 2002). Uroplakin 1a is a membrane protein in the asymmetric membrane covering the uroepithelium in the bladder. Attachment of type-1 pili to the uroepithelium triggers internalisation of the bacteria (Mulvey, 2002, Martinez et al., 2000, Mulvey et al., 2001), and UPEC can thereby find a sheltered niche for intracellular replication. The epithelial cells respond by exfoliation, in a suicidal attempt to clear the bladder from bacteria. Bacteria that manage to escape from the dying cells can invade the exposed underlying layer of cells. Infection by UPEC stimulates production of cytokines and influx of neutrophils, as well as the expression of a number of pro-inflammatory molecules.

The chaperone:usher pair FimC:FimD is responsible for the assembly of type-1 pili. The FimA subunit makes up the main part of the pilus, which consists of many thousand copies of FimA linked into a fibril, wound up in a tight, right-handed helix. A flexible tip is attached to this quite rigid rod-like structure, and the tip consists of the two linker proteins FimF and FimG, and the adhesin protein FimH, responsible for mannose binding (Figure 1.1).

1.5 Plague and F1 antigen

In sharp contrast to urinary tract infections, the plague is a highly invasive and lethal disease, although fortunately it is not very common in our times. *Yersinia pestis* is the causative agent of plague, which in the middle ages killed 17-28 million people in Europe alone, approximately 30-40% of the population at that time (Drancourt and Raoult, 2002). There are, however, still outbreaks of plague in e.g. India and Madagascar (Mansotte, 1997, Boisier et al., 2002, Chanteau et al., 2000, Ramalingaswami, 1995).

In bubonic plague, *Yersinia pestis* has an infection cycle that passes from infected rats, via flea, and from the bite of the flea to humans (Titball et al., 2003, Perry and Fetherston, 1997). The infection results in swelling of the lymph nodes (bubos), presenting the classical symptom of bubonic plague. The infection can occasionally spread to the bloodstream, leading to infection of the lungs and thereby becoming pneumonic plague. Pneumonic plague is spread via aerosols from human to human, mediating an extremely efficient invasion process which can kill a human within a couple of days (Titball et al., 2003, Titball and Williamson, 2001, Perry and Fetherston, 1997). Bubonic plague can be treated with antibiotics, but pneumonic plague is difficult to treat because of the rapid development, and even with antibiotic treatment the outcome is often fatal.

Vaccines against plague exist, both in the form of killed whole cells and attenuated live bacteria, although none of them approved (Titball and Williamson, 2001). A vaccine based on killed cells does not show effective protection against pneumonic plague, and an attenuated vaccine retains some virulence and is therefore not suitable for humans. A vaccine based on protein has shown promising results, and the most efficient seems to be a mixture of the F1 antigen and the V-antigen. V-antigen, or LcrV, is a protein secreted by the type III secretion systems, and is involved in regulating Yop expression and secretion (Price et al., 1991, Pettersson et al., 1999). The F1 antigen, Caf1, forms a capsule around *Y. pestis*, and has been studied in this thesis (Zavialov et al., 2003a, Zavialov et al., 2003b).

The F1 capsule is built up from one type of subunit, Caf1, linked together in thin, fimbriae-like organelles forming a thick gel-like capsule around the bacteria (Zavialov et al., 2002). The capsule is assembled by the FGL chaperone:usher pair Caf1M:Caf1A, and expression regulated by Caf1R. The production of the capsule is induced at 37°C. Around 4 hours after induction the capsule can be observed at the outside of the bacteria but full encapsulation takes up to two days (Du et al., 2002, Perry and Fetherston, 1997).

The F1 capsule is antiphagocytic, and encapsulation of *Y. pestis* has been shown to reduce the number of bacteria interacting with macrophages. A knockout of Caf1M lowered the ability to prevent uptake by J774 cells (a macrophage-like cell line) (Du et al., 2002), and F1 is therefore thought to act together with the Yersinia outer proteins (Yops) to prevent phagocytosis. Since the F1 capsule is induced only at 37°C, it is thought to be most important in the late stage of the infection. The capsule has not (yet) been shown to mediate binding, which makes it an unusual member of the chaperone/usher family where most organelles are adhesins.

The Yersinia family consists of three members, *Yersinia pestis*, *Yersinia pseudotuberculosis* and *Yersinia Enterocolitica*. All three species share a common virulence plasmid encoding the type III secretion system of Yop virulence effector proteins. *Y. pestis* has two additional plasmids, pPla (or pPCP1) and pFra, both unique to *Y. pestis* (Perry and Fetherston, 1997). pPla encodes the Pla protease which has adhesive properties to extracellular matrix components, and is also a possible invasin (Cowan et al., 2000). The second unique plasmid is pFra, encoding both a murine toxin necessary for survival in the rat flea midgut (Hinnebusch et al., 2002) and the operon to produce the F1 capsule.

1.6 Aim and outline of this thesis

Many different types of organelles assembled by the chaperone:usher pathway are known today, and many more presumably exist. A great deal of work has been put into understanding the mechanism of assembly and adhesion of these organelles.

In this thesis two organelle systems have been studied in greater detail, the structure and binding properties of the adhesin of type-1 pili from *E. coli*, FimH, and the structure and assembly of the prototypic FGL system organelle, the F1 antigen from *Yersinia pestis*. A structure:sequence comparison of the known 3D structures of domains from subunits from the chaperone:usher pathway organelles has also been conducted. The methods used to obtain the results will be covered in chapter 2, and the results and discussion of this investigation will be presented in chapter 3, which is divided into three parts:

Section 3.1: To obtain a better understanding of FimH mediated adhesion, the structure of the FimH lectin domain has been determined, and binding studies of this domain to a variety of mannosides conducted. A comparison of the binding properties of FimH from three different *E. coli* strains has given new insights into the molecular basis of this adhesion.

Section 3.2: The first structure from a FGL system has been determined, the Caf1M:Caf1 chaperone: subunit binary complex from the F1 capsule of *Yersinia pestis*. The structure of the F1 minimal fibre, the Caf1M:Caf1:Caf1 ternary complex, provided the first direct evidence of donor strand exchange. A comparison of the Caf1 subunits in the binary and the ternary complex provide insights into the driving force behind chaperone/usher mediated subunit folding and organelle assembly.

Section 3.3: All domains from the chaperone:usher pathway with a known 3D structure share an immunoglobulin fold, despite very low sequence similarity. A comparison of the structures available has been performed in order to find sequence similarities and structural patterns important for function of the subunits.

2 Methods

2.1 X-ray crystallography

2.1.1 Introduction

Proteins are one of the fundamental building blocks of life, participating in a wide variety of functions. The biological function of a protein depends on its three-dimensional (3D) structure, which has evolved through selective pressure to optimise the protein for its specific task. In order to understand the details of the function of a particular protein it is therefore important to determine its 3D structure.

The method used to determine the 3D structures of proteins in this thesis is x-ray crystallography, a method based on the fact that electromagnetic radiation can interact with the electrons in a protein. If x-rays hit a protein crystal (many ordered copies of a protein molecule), scattering is enhanced in certain directions, allowing it to be recorded on a detector. From the resulting diffraction pattern the electron density can be retrieved, and a model of the protein built (Giacovazzo et al., 2002, Drenth, 1994, McRee, 1999). The field of x-ray crystallography has been under steady development since the first structural work on myoglobin and hemoglobin by Perutz and Kendrew in the 1950's. Since then 24444 structures have been deposited in the Protein Data Bank (Feb. 24, 2004). The development of the method is mainly due to advances in three fields: new and stronger sources of x-rays in the form of synchrotrons, a rapid increase in computer hardware as well as software, and the recombinant DNA technique making expression of large amounts of protein possible. Synchrotrons with tuneable wavelengths also led to the development of new phasing methods: MAD/SAD (Multiple/Single wavelength Anomalous Dispersion), which is now a common method for the determination of an unknown protein structure.

2.1.2 Determining the structure of the Caf1M:Caf1 binary complex

The structure of the Caf1M:Caf1 complex from *Y. pestis* capsule was determined using a combination of selenium MAD and platinum SIRAS. Structure determination was not entirely straightforward, and special care had to be applied in order to get correct phases. Since this was not trivial, the procedure used will be described in some detail.

A native dataset of Caf1M:Caf1 complex was collected to 1.8Å resolution on beamline id14:2, ESRF, France. The spacegroup was determined to be $P2_1$ with the cell dimensions $a = 36\text{Å}$, $b = 36\text{Å}$, $c = 69\text{Å}$, $\beta = 93^\circ$ and 1 molecule in the asymmetric unit. In order to determine the phases, seleno-methionine substituted

protein was expressed, purified and crystallised, and a three-wavelength MAD data set was collected at beamline id14:4, ESRF, France.

In the Caf1M:Caf1 complex there are 2 methionines per protein, which means that with 100% incorporation there is 1 selenium per 96 amino acid residues. A good MAD signal is estimated to be 3% for Bijvoet differences and 2% for dispersive differences, which corresponds to maximum 90 residues per selenium (http://www.chess.cornell.edu/Publications/Newsletter_1995/gomad.html).

Although structures have been solved with lower Bijvoet differences, this means extremely accurate, high redundancy data need to be collected. This proved difficult since the crystals suffered from radiation damage after the three-wavelength MAD dataset, and the remote wavelength had to be discarded.

The Selenium positions were located using the program RSPS (Knight, 2000, CCP4, 1994) and preliminary phases were obtained using the program SHARP (La Fortelle and Bricogne, 1997). The initial maps, however, were noisy and hard to interpret. The crystals had a solvent content of 33%, which is quite low and makes solvent flattening less powerful. In addition there was only one molecule/asymmetric unit, which means the electron density cannot be improved by NCS averaging.

To improve phasing power, platinum derivative datasets were collected near the L_{II} and the L_{III} absorption edges (beamline 7:11, MaxLab Lund and beamline id14:2, ESRF, respectively). Using all available data in SHARP, parameter refinement was not robust, and small differences in input parameters resulted in unexpected fluctuations throughout refinement. This was especially true for the values of f' and f'' , for which we only had rough estimates. In order to stabilise the process, and to get reasonable initial values for the input parameters, a thorough boot strapping procedure was applied using the program SHARP. By refining the parameters carefully for one heavy atom compound at the time, and alternating the runs with or without the use of external phases, reasonable estimates of the heavy atom positions, occupancy, B-factors, f' and f'' was obtained. Finally all the data were put together in SHARP: 2 wavelengths of selenium data (peak and inflection point), 2 platinum datasets plus native data. The values for f' and f'' were locked to the values obtained in earlier runs, and not refined any further. This proved to be a successful strategy, and nice maps could be calculated and the model built. The conclusion drawn from this is to be adamant that beamline scientists help in getting good experimental values of f' and f'' , in difficult cases this will prove important.

2.2 Dot-blot

FimH is the mannose binding protein located at the ultimate tip of type-1 pili. In order to determine whether it retained its mannose binding capabilities after expression and purification, and thereby also most probably the correct fold, a dot blot assay was developed. The idea behind the assay is the fact that FimH binds horseradish peroxidase, which is a mannosylated protein. Horseradish peroxidase catalyses the reduction of H_2O_2 to O_2 , and this conversion can be linked to the oxidation of 4-chloro-1-naphthol into an insoluble blue-coloured product. H_2O_2 in the presence of 4-chloro-1-naphthol can thus be used to detect peroxidase by the appearance of a blue colour.

4 different concentrations of FimH were blotted onto nitrocellulose filter. Blocking buffer containing 2% tween was applied to block further binding to the filter. After washing, peroxidase at a concentration of $50\mu\text{g/ml}$ was added, and allowed to bind to FimH. Excess peroxidase was washed away. Finally 4-chloro-1-naphthol was added in the presence of 0.01% H_2O_2 to detect bound peroxidase.

The assay was performed with Concanavalin A as a positive control, and Lysozyme as a negative control, both in the absence and presence of mannose. The results clearly show blue dots at the places where FimH lectin domain or FimH:FimC complex were blotted on the cellulose filter (Figure 2.1). The absence of blue dots, and thereby absence of bound peroxidase, when mannose was present in the sample indicates that binding is mannose specific.

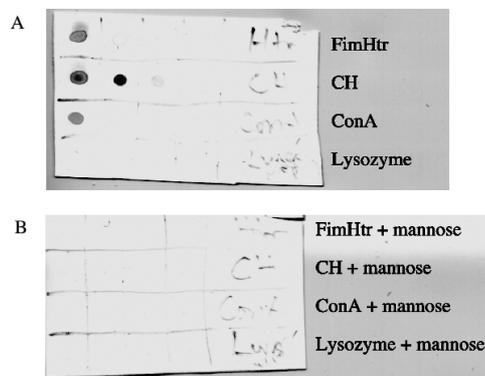


Figure 2.1. (A) Clear binding of peroxidase to FimH lectin domain (FimHtr), FimC:FimH complex and ConcanavalinA, but no binding to Lysozyme. The difference in staining is probably due to differences in protein concentration. (B) In presence of mannose, peroxidase does not bind any of the proteins.

2.3 Docking

When crystallographic studies of ligand binding proved impossible, the method of docking was used. Docking is a tool to computationally simulate interaction of a small ligand to a macromolecule of known 3D structure. All docking studies were performed with the program Autodock3.0 (Morris et al., 1996).

Autodock keeps the structure of the macromolecule fixed, while searching the 6-dimensional space created of three rotation angles and three translation axes for the ligand. Autodock also allows the ligand to rotate around selected internal torsion angles.

A grid of affinity potentials is created around the binding site of the macromolecule, one grid for each atom type present in the ligand. Each grid represents the interaction energy of the particular type of atom at every grid point. Also, an electrostatic grid is calculated, using a point charge of +1 as a probe. At every conformation and new position of the ligand, the interaction energy and the internal energy of the ligand are calculated, searching for the energy minima.

In the docking studies performed in this thesis, the Lamarckian algorithm was used for sampling the space of possible ligand conformations. Each simulation consisted of 100 independent runs, with a population size of 200, 500 generations and a maximum of 25000000 energy evaluations. The numbers were chosen based on the work by Hetenyi and van der Spoel (Hetenyi and van der Spoel, 2002), and by personal communication with Dr. D. Choudhury. Solutions were ranked based on their docking energies, and similar solutions were clustered. The top solutions in each cluster were visually inspected using the program O (Jones et al., 1991).

2.4 Mass spectrometry

Mass spectrometry is a technique to determine the mass of protein or smaller molecules to very high accuracy. It has become increasingly popular because of the simplicity of using the method combined with highly accurate results, and is now a standard technique in many laboratories.

The analyte of interest is converted into gas phase ions by various techniques (Glish and Vachet, 2003), the most common being matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI). ESI was the technique used in this thesis. In ESI a solution containing the sample is passed through a small capillary, where a high voltage is applied to the outlet spray tip.

This produces a fine spray of charged droplets containing the sample and its solvent. Eventually the ions are separated from the solvent, and are transported to the analyser, where the mass of the ion is determined.

ESI is very sensitive to salts, and also using an organic buffer in too high concentration relative to the protein may mask the signal. The samples were therefore dialysed against large volume of water over night. The sample was kept at a concentration around 10^{-6} M, and diluted 50/50 with methanol to facilitate ionization.

2.5 Determination of binding constants

Many binding studies have been done on type-1 pili in the past, but most of them were performed on entire, pilated bacteria, and not on the carbohydrate binding protein FimH alone. To determine the dissociation constant (K_d) of a series of mannose compounds binding to FimH, we developed the binding assay described below (Figure 2.2).

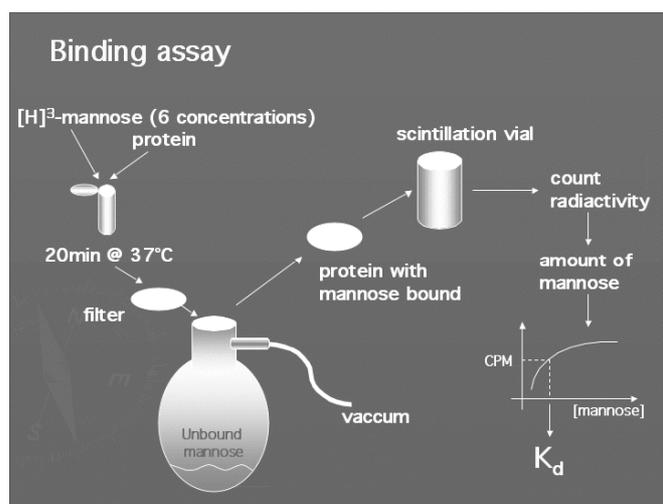


Figure 2.2: Binding assay to determine the dissociation constant to α -D-mannose. K_d is defined as the concentration of ligand when half the sites are occupied. By plotting the counts per minute (CPM) against the concentration of mannose, K_d can be obtained from the graph.

$[^3\text{H}]$ -mannose of 6 different concentrations, 0 - $43.5\mu\text{M}$, was mixed with FimH protein. The mixtures were incubated for 20 min at 37°C to allow equilibration. After incubation, the mixture was applied to a filter (Portran BA 85 Cellulosenitrate filter, Schleicher & Schuell Germany), with pores large enough to

allow passage of free mannose but small enough to retain protein, and thereby mannose bound to protein. When a vacuum is applied, free mannose goes through the filter. The filter was washed once with ice-cold 1ml PBS (phosphate buffer saline) to wash away free mannose, and then put into scintillation liquid. The radioactivity was counted for 3 min per sample.

By plotting radioactivity in the form of counts per minute (CPM) against the concentration of mannose, a hyperbola can be fitted to the data, and K_d determined from the mannose concentration half way to equilibrium (Figure 2.2). To get K_d of different mannose derivatives, the same experiments were performed with the concentration of mannose kept fixed at $43.5\mu\text{M}$ and instead varying the concentration of the inhibitor. The amount of mannose bound in the presence of the inhibitor was measured, and could be plotted against the concentration of the inhibitor. The $[I]_{0.5}$ value of the inhibitor (inhibitor concentration displacing 50% of the ligand, IC_{50}) can be determined from this plot, and the K_d of the inhibitor is given by the Cheng and Prusoff (Cheng and Prusoff, 1973) equation:

$$K_i = \frac{[I]_{0.5}}{\frac{[L]}{K_L} + 1}$$

$[I]_{0.5}$ is the IC_{50} value of the inhibitor, $[L]$ is the concentration of α -D-mannose and K_L is the dissociation constant of α -D-mannose. This equation was used when both the concentration of the radioactive ligand (L) and the displacing agent (I) are in excess over the protein ($L_T \gg R_T$; $I_T \gg R_T$, T indicates total concentration), (Cheng and Prusoff, 1973). For very strong inhibitors, when I_T is no longer in excess over R_T , another version of the equation was used:

$$K_i = \frac{I_T}{\frac{1 - Y}{Y} \cdot \frac{L_T}{K_L} - 1} \cdot \frac{R_T \cdot K_L \cdot Y}{L_T}$$

(Horovitz and Levitzki, 1987) where Y is the fraction of the ligand bound in presence of the inhibitor. A plot of $I_T / ((1 - Y) \cdot (L_T / K_L) - Y)$ against $1/Y$ gives a straight line with a slope of K_i .

3 Results and discussion

3.1 FimH lectin domain (Paper III)

3.1.1 Introduction

One FimH carbohydrate binding protein is situated at the tip of each type-1 pilus. FimH is a two-domain protein consisting of one lectin domain and one pilin domain, and the structure of FimH in complex with its chaperone FimC was solved previously in our laboratory (Choudhury et al., 1999). In this thesis the lectin domain of FimH has been extensively studied in order to understand the molecular details of the carbohydrate binding mechanism. The aim was to investigate if the lectin domain is structurally and/or functionally independent of the pilin domain, to determine the binding specificity to various mannose-compounds and to verify whether this specificity is maintained between FimH variants originating from different strains of *E. coli*.

3.1.2 The structure of FimH lectin domain

A truncated FimH, FimH_{tr}, consisting of the FimH lectin domain can be expressed separately from the pilin domain and has been shown to be stable and soluble on its own (Schembri et al., 2000). A construct was made consisting of the first 158 amino acids of FimH, with a carboxy-terminal 6-his-tag attached (Schembri et al., 2000). The protein was purified using Ni chelate chromatography and crystallised in 65% MPD, 100mM Cacodylate buffer pH 6.5 using hanging drop vapour diffusion. Data were collected to 2.5Å resolution at beamline 7-11, MaxLab, Lund, and later to 2.3Å resolution at id29 at the ESRF, Grenoble, France. The structure was solved with molecular replacement using Molrep (CCP4, 1994) with the lectin domain of the FimH:FimC complex as a search model. All 158 amino acids could be modelled into the density, with the exception of the 6-his-tag, which was not visible.

The FimH lectin domain is an 11-stranded beta-barrel with a jelly-roll like topology, built from 3 beta sheets (Figure 3.1). The back sheet is rather large and runs along the whole back part of the beta barrel while the front of the barrel is divided into 2 sheets, related by a pseudo two-fold axis.

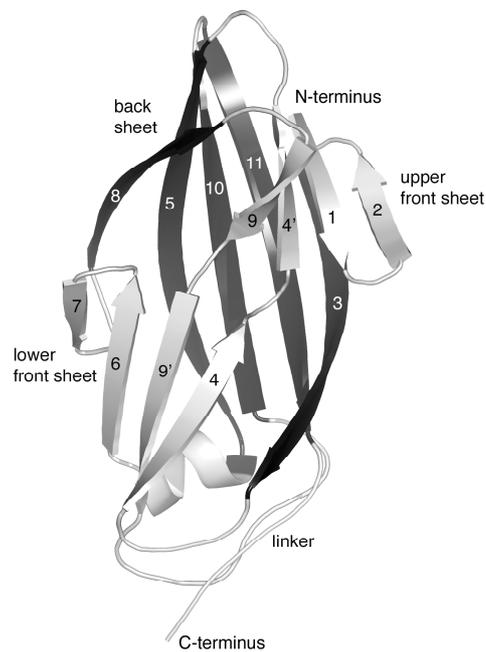


Figure 3.1 The structure of FimH_{truncate} with the back sheet coloured black, the front sheets in grey.

The mannose binding-site is located at the tip of the barrel, identified both crystallographically (Choudhury et al., 1999, Hung et al., 2002), and by mutagenesis (Hung et al., 2002). The binding site forms a deep cave made up from three rather short loops, and is large enough to totally enclose a mannose-ring. The back of the cave is made up from loops β 3- β 4 and β 10- β 11, while the lower front consist of loop β 2- β 3, and the floor is constituted by the amino-terminus. The binding cavity is situated at the side of the tip of the domain, shielded from the top by the two back-wall loops, and particularly by two tyrosine rings, Tyr48 and Tyr137, one on each loop (Figure 3.2).

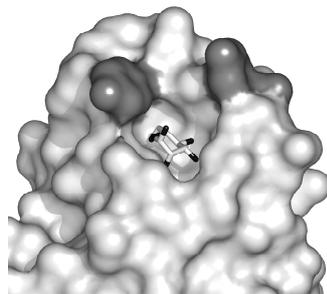


Figure 3.2 Tyrosine gate coloured grey, and a bound mannose shown in ball-and-stick in the monomannose binding pocket.

The structure of FimH_{truncate} is very similar to the lectin domain in the FimC:FimH complex (an overall r.m.s.d. of 0.55Å for 158 C α atoms). The main difference can be seen in the α 9- α 10 loop, which stabilises the linker-region between the lectin domain and the pilin domain. In the FimC:FimH complex, the α 9- α 10 loop packs against the chaperone, while in the FimH_{tr} structure it packs against the linker region. The mannose-binding site, which is located at the opposite end of the domain, is practically unchanged, and amino acids making up the binding pocket have virtually identical side chain conformations in the two structures. The high structural similarity between FimH_{tr} and the lectin domain in the FimH:FimC complex indicates that the lectin domain is structurally independent of the pilin domain.

The affinity of α -D-mannose for FimH was determined using [³H]-labelled mannose (see chapter 2) and a value of $K_d = 4.1\mu\text{M}$ was obtained. This is a quite strong binding for a mono-carbohydrate binding to a lectin, normally in the mM range (Rini, 1995), and confirms that the tight fit of the mannose-ring in the binding site is also reflected in strong binding. In order to examine whether a binding constant determined for the FimH lectin domain is a reasonable estimate for FimH binding, binding of α -D-mannose to full length FimH in complex with the chaperone FimC was also measured. A $K_d = 5.3\mu\text{M}$ was obtained, which corresponds well to the measured value for FimH lectin domain. Hence, removal of the pilin domain does not seem to affect the binding properties of the lectin domain (Paper III).

3.1.3 Butyl mannoside

When all 158 amino acids of the FimH lectin domain had been modelled, a strong F_o-F_c density still remained in the mannose-binding site, although no mannose was present in the crystallisation setup or in the protein buffer. A mannose unit could nicely be modelled into this density (Figure 3.3A). After refinement, an additional difference density turned up at the O1 oxygen, extending out of the binding pocket (Figure 3.3B). An alkyl tail of four carbon atoms could be modelled into this extra density, suggesting that the bound ligand was a butyl mannoside (Figure 3.3C) (Paper III).

The position of the mannose ring superimposes well with the structure of α -D-mannose bound to FimH (Hung et al., 2002), making direct contacts via hydrogen bonds to the side chains of residues Asp54, Gln133, Asn135 and Asp140, and to the main chain of Phe1 and Asp47. Additional, water mediated hydrogen bonds are observed between the mannose ring and the side chain of residue Glu133 as well as the main chain of Phe1 and Gly14. The alkyl tail extends out of the

pocket, packing against the rings of Tyr48 and Tyr137, situated on the two back-wall loops of the binding cavity (Figure 3.3).

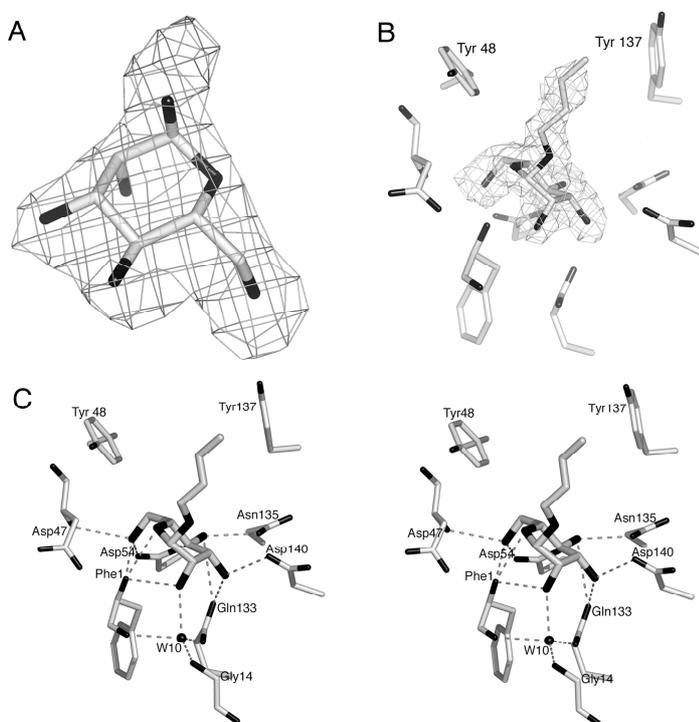


Figure 3.3. (A) A mannose modelled into F_O-F_C density. (B) Butyl-mannoside modelled into the binding site with the $2F_O-F_C$ map shown. (C) Stereo figure of the interactions between FimH and the butyl-mannoside.

To verify the identity of the ligand, electro-spray ionization mass spectrometry (ESI-MS) was used. A reference sample of butyl-mannoside was synthesised, which showed a strong peak at 259D in the MS, consistent with a butyl-mannoside plus a Na-ion (Figure 3.4A inset). When repeated with a sample of the protein, also here a peak of 259D could be measured (Figure 3.4 A). The fact that the ligand has the exact molecular weight of a butyl mannose and with electron density matching such a compound, strongly supports the crystallographic identification of the ligand as a butyl mannose.

Hypothesising that the LB-media used for growing the bacteria was the origin of the ligand, the protein was re-expressed in bacteria grown in M9-minimal media and purified according to the same protocol. When running a sample of this protein through the MS, no peak could be detected at 259D (Figure 3.4B), suggesting that the LB media is indeed the origin of the butyl mannose.

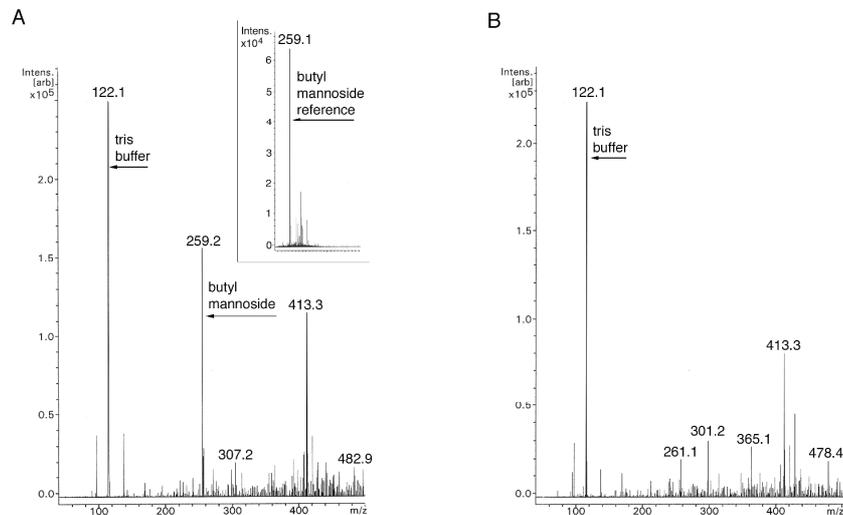


Figure 3.4. (A) ESI-MS on FimH expressed in *E. coli* grown in LB-media, with a reference butyl-mannoside spectrum as an inset. A peak of 259D, consistent with a butyl-mannoside, is visible in both spectra. (B) In protein expressed in bacteria grown in minimal media, the 259D peak is no longer present.

3.1.4 Binding studies of alkyl mannosides

The butyl mannoside discovered in the binding site of FimH suggested a new class of FimH inhibitors, alkyl-substituted mannosides. The butyl mannoside resides in the binding-pocket throughout purification and extensive dialysis against mannose-free buffer, which suggests that binding between FimH and alkyl mannosides is quite strong.

The dissociation constant to a number of alkyl mannosides was determined using displacement studies (see chapter 2). The added binding strength mediated by the hydrophobic tail can be nicely demonstrated by the binding series of methyl up to octyl mannoside (Table 3.1).

As can be seen in the table, the binding strength increases by a factor of two for every methyl group added to the mannose ring in a near linear decrease (Figure 3.5), and heptyl and octyl mannoside both proved to be very strong binders (Table 3.1). The fact that alkyl mannosides are easily synthesised, in addition to being highly soluble in water make them interesting potential blocking agents of FimH mediated adhesion.

Table 3.1: Binding constants of a series of alkyl mannosides

ligand	Kd (M)	ΔG^0 (kcal/mol)
α -D-mannose	4.1E-6	-7.6
methyl mannoside	1.8E-6	-8.1
ethyl mannoside	7.4E-7	-8.7
propyl mannoside	4.0E-7	-9.1
butyl mannoside	1.5E-7	-9.7
pentyl mannoside	2.0E-7	-9.5
hexyl mannoside	1.0E-7	-9.9
heptyl mannoside	3.2E-8	-10.6
octyl mannoside	2.8E-8	-10.7

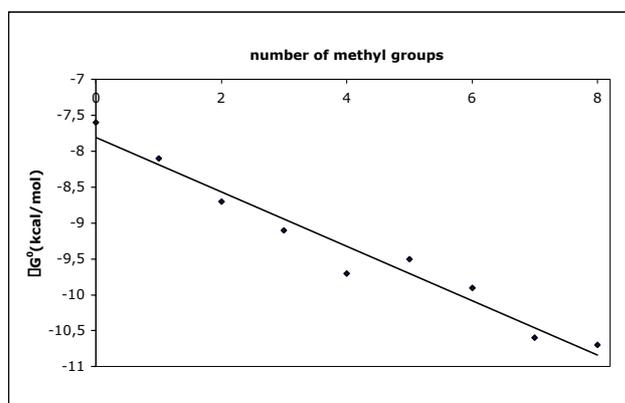


Figure 3.5. Binding energy for each of the alkyl mannosides. A near linear increase of binding strength can be seen for each methyl group added to the mannoside ring.

3.1.5 Aromatically substituted mannosides

Aromatically substituted α -glycosides of mannose have previously been shown to bind strongly to FimH, indicating that the binding site includes a hydrophobic region next to the mannose-binding pocket (Firon et al., 1987, Firon et al., 1984). The strongest binders in the study done by Firon et al. were 4-methylumbelliferyl- α -mannoside (MeUmb α Man) and p-nitro-o-chlorophenyl- α -mannoside (pNoCIP α Man), binding up to 1000 and 70 times stronger than methyl- α -D-mannoside respectively (Firon et al., 1987, Firon et al., 1984). We determined the dissociation constant to 12nM for MeUmb α Man and 26nM for pNoCIP α Man, which is around 220 and 90 times stronger than the determined binding for α -D-mannose.

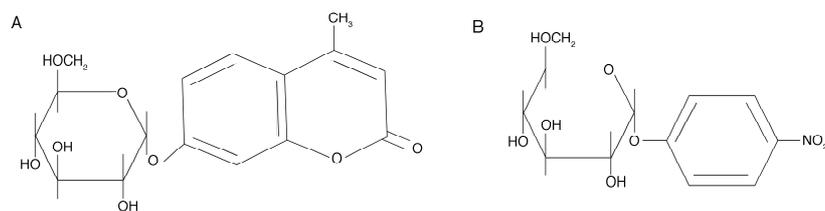


Figure 3.6: Schematic structure of (A) 4-methylumbelliferyl- β -mannoside and (B) p-nitro- α -chlorophenyl- β -mannoside.

Despite extensive trials no crystals could be obtained with FimH in complex with either of these compounds, possibly due to restrictions of crystal packing. To nevertheless get an indication of the binding mode, the compounds were docked to FimH using the program Autodock3 (Morris et al., 1996) (chapter 2).

Both compounds dock with their mannose ring in the same conformation as the crystallographically determined mannose (Figure 3.7). For the MeUmb β Man compound, basically all solutions cluster in one group, with the umbelliferyl rings inserted between the two tyrosine rings, Tyr48 and Tyr137, making tight stacking interactions (Figure 3.7A). These interactions could easily be imagined to provide the extra binding strength. The lowest docking energy obtained for MeUmb β Man was $E_{\text{doc}} = -10.9$ kcal/mol. pNP β Man docking results are also individually very similar and pNP β Man orients its phenyl ring towards the two tyrosines, with a docking energy of $E_{\text{doc}} = -10.4$ kcal/mol. The phenyl ring does not quite reach in between the two tyrosines, but instead hydrogen bonds with the nitro group to the carboxyl group of Tyr137 (Figure 3.7B).

Both the docking experiments and the crystallographically determined butyl-mannoside structure points towards the importance of the tyrosine gate. The alkyl tail of the butyl-mannoside packs against the two aromatic rings, and the umbelliferyl aromatic ring-system inserts between them. These data suggest that the tyrosine gate corresponds to the extended hydrophobic region of the mannose-binding site proposed by Firon et al. (Firon et al., 1984).

Carbohydrate rings frequently interact with aromatic side chains, and the tyrosine gate could be imagined to be part of a trimannose binding-site. Docking attempts with trimannoses (unpublished) show a tendency to dock with the non-reducing end in the mannose pocket and the reducing end inserted in the tyrosine gate, again pointing towards an important role of the two tyrosines.

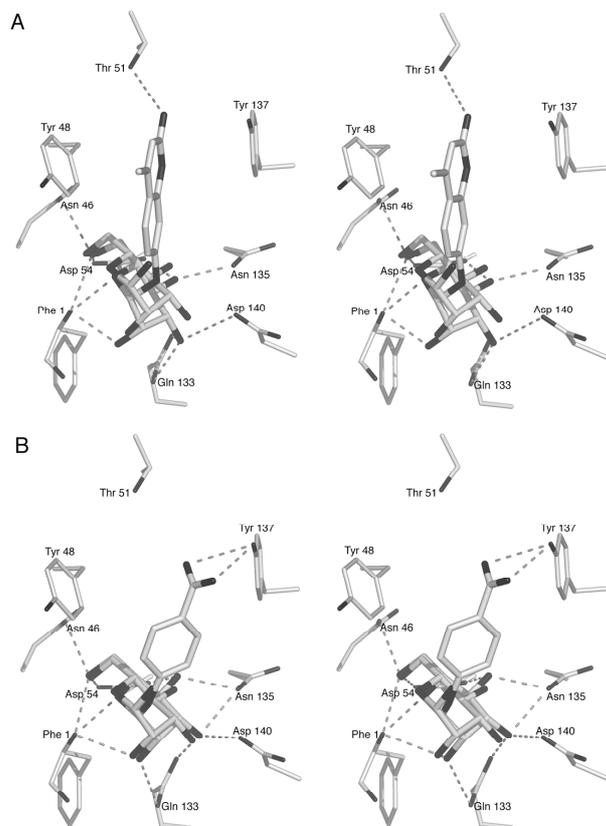


Figure 3.7. Stereo-figure showing docking results from (A) 4-methylumbelliferyl- α -mannoside and (B) p-nitro-o-chlorophenyl- α -mannoside, with a crystallographically determined mannose-ring shown for comparison.

3.1.6 Different strains of *E. coli* show different binding phenotypes

Type-1 pili are present on over 90% of all *E. coli* strains, pathogenic as well as commensal. For a long time it remained unclear if type-1 pili should be considered an UPEC virulence factor since they are also present on non-pathogenic strains. As more results accumulated, an apparent pattern was revealed. Although FimH proteins in different strains have very conserved sequences (Hung et al., 2002, Abraham et al., 1988), differences exist that give rise to differences in binding pattern. Sokurenko and co-workers have divided the different strains into two groups with distinct phenotypes: members of both groups bind well to trimannosyl residues, but only one group shows tight binding to monomannosyl residues (Sokurenko et al., 1992, Sokurenko et al., 1997, Sokurenko et al., 1998, Sokurenko et al., 1995). Here trimannosyl residues refer to compounds with a free trimannoside, mostly RnaseB or BSA-linked trimannosides (man(1,3)-man(1,6)-man). Monomannose high binding refers to strong binding to structures with a

terminal mannose, such as yeast mannan. A correlation between strong binding to yeast mannan and binding to BSA-linked α -D-mannose has been shown (Sokurenko et al., 1997). UPEC strains seem to typically belong to the monomannose high binding group, whereas fecal, commensal strains are monomannose low binders. Binding well to monomannose correlates with capability to agglutinate red blood cells and with high binding to the uroepithelium (Sokurenko et al., 1997).

Interestingly, the sequence differences between the two phenotypes are not typically located in the identified mannose-binding pocket, but rather at the opposite end of the domain, often in or close to the linker region between the lectin domain and the pilin domain.

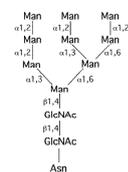
3.1.7 *FimH* affinity to trimannosides

Binding studies on *E. coli* have traditionally been done on entire, piliated bacteria. We decided to determine whether the above results would be reflected in measurements done on the FimH protein. The FimH lectin domain from three different clinical isolates were chosen for this study: J96, CI#4 and F18. CI#4 and J96 are both high monomannose binding UTI strains (Sokurenko et al., 1995) and F18 is a low monomannose binding fecal strain, that does not agglutinate RBC and does not bind J92 human bladder epithelial cells (Sokurenko et al., 1995, Sokurenko et al., 1997). The sequence of the FimH_{J96} lectin domain differs from the two others by mutations V27A, N70S and S78N. FimH_{CI#4} differs from FimH_{F18} and FimH_{J96} in mutation G73E, located at the lower part of the lectin domain on the opposite side to the linker region.

The binding constants of a series of different trimannosides to the FimH lectin domain from the three strains were determined and compared to α -D-mannose binding. The trimannosides chosen correspond to structures present in N-linked high-mannose glycoproteins.

Table 3.2. Results from binding studies with a series of tri-mannosides to FimH from three different *E. coli* strains. The tri-mannosides are all branches from the high-mannose tree, shown to the right of the table.

Ligand	Kd J96 (nM)	Kd CI#4 (nM)	Kd F18 (nM)	ΔG° J96 (kcal/mol)	ΔG° CI#4 (kcal/mol)	ΔG° F18 (kcal/mol)
α -D-mannose	4100	10700	9800	-7,6	-7,0	-7,1
man(1,2)man(1,2)man	1600	3950	3250	-8,2	-7,7	-7,8
man(1,2)man(1,3)man	1800	3650	3050	-8,1	-7,7	-7,8
man(1,2)man(1,6)man	830	2200	1800	-8,6	-8,0	-8,1
man(1,3)man(1,6)man	350	1030	730	-9,2	-8,5	-8,7
man(1,6)man(1,6)man	1400	7500	5900	-8,3	-7,3	-7,4



In contrast to results showing differences in monomannose binding between these three strains, our results instead point towards conform binding (Table 3.2), with a

virtually identical binding phenotype for FimH_{F18} and FimH_{C1#4}. For monomannose binding, a K_d of 10 μ M for FimH_{F18} and 11.4 μ M for FimH_{C1#4} was determined, and correspondingly similar binding constants to the trimannose series (Table 3.2). This suggests that the G73E mutation has no effect on affinity, and hence that affinity of FimH to mannose is not responsible for the differences in binding observed between the strains.

FimH_{J96} seems to have a significantly higher affinity for all compounds tested in this study, suggesting that the V27A, N70S and S78N mutations are of importance. Even though this significant difference in affinity exists, FimH_{J96} binding follows the same trend as the other two, suggesting that the shape of the binding site is very similar in all three proteins. Differential binding of the trisaccharides presumably reflects differences in the fit to an extended binding site, where the stronger binders are able to fit the additional mannose residues more or less well, giving rise to an increased number of interactions.

3.1.8 Why does shear force promote monomannose high binding?

Thomas et al. (Thomas et al., 2002) show in an interesting experiment how shear force influences adhesion of piliated bacteria, and moreover that different strains are differently affected by shear force. Monomannose low binding strains of *E. coli* are not capable of agglutinating red blood cells (RBC) under static conditions. When these strains are subjected to shear, agglutination of RBC:s is observed. Binding is reversible, so when shear is released the RBC:s are released. Mono mannose high binding strains bind to RBC:s with roughly the same strength with or without shear.

Mutations influencing shear dependence are again not found in the mannose-binding pocket, but close to the linker region between the two domains. Thomas et al. show in the same article that if this linker region is stabilised, shear is required for RBC agglutination – the bacteria become shear dependent. If this region instead is kept flexible, RBC agglutination is possible independently of shear force.

Several suggestions for how mutations around the linker region would influence specificity have been proposed. The main hypothesis involves exposure of cryptic binding sites upon shear, or conformational changes propagating from the linker region to the binding site. Based on our results we instead propose that the presentation of the mannose-binding site is crucial for binding, which would be facilitated by a flexible linker region. The binding site is not situated at the very tip of the domain but rather displaced to one side, thus shielded from the tip by

the back wall of the cavity. Flexibility between the two domains could therefore be imagined to generate a productive presentation of this site to a surface of receptors. If one pilus displaying a FimH binds to its receptor, the flexibility might enhance binding of another pilus nearby, and yet another pilus nearby the second (Figure 3.8). The added binding strength from many pili would enlarge any such effect on binding, and the accumulated effect would show up as a high binding strain.

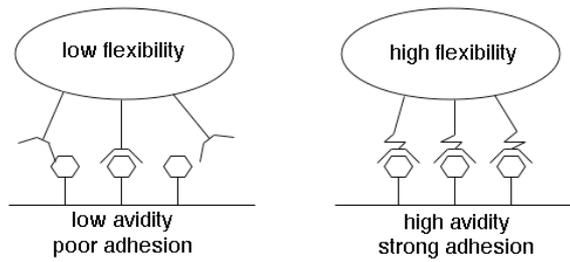


Figure 3.8. Low flexibility in the linker region might lead to low avidity and poor adhesion. High flexibility mediates high avidity and strong adhesion.

3.2 F1 antigen (Paper I and II)

The capsular F1 antigen from *Yersinia pestis* is the prototype system for atypical adhesins assembled by the FGL chaperone/usher pathway. No structure was previously available from any of the components of an FGL system.

3.2.1 The Caf1M:Caf1 binary complex

Caf1 subunits have been shown to assemble into short Caf1M:(Caf1)_n fibres in the periplasm in the absence of the Caf1A usher (Zavialov et al., 2002). In order to prevent this pre-assembly and to facilitate purification, the amino-terminal donor strand was exchanged for a 6-his-tag, which completely abolished fibre formation (Paper I). The Caf1M:Caf1 chaperone:subunit complex could thus be over-expressed, purified and finally crystallised in 25% PEG 4000. The structure was solved by a combination of Selenium MAD and platinum SIRAS, see Chapter 2. ARP/wARP (Morris et al., 2002, CCP4, 1994) traced a large part of the molecule and improved the phases further, and the missing part of the molecule could be built with the graphics program O (Jones et al., 1991).

The structure of the Caf1M chaperone is very similar to the structures of the FGS chaperones. It consists of two Ig-like domains, joined at 90° angle. Two cysteines, Cys98 and Cys137, have been shown to be important for Caf1M folding and are conserved in all FGL chaperones (Zavialov et al., 2002, Hung et al., 1996). In the structure they can be seen to stabilise the long F₁-G₁ loop by forming a disulfide bridge between the two strands, a disulfide bridge likely to be conserved throughout the FGL chaperone family. The F₁-G₁ loop itself was only partly visible in the electron density, indicating that the loop is flexible in the crystal (Paper II).

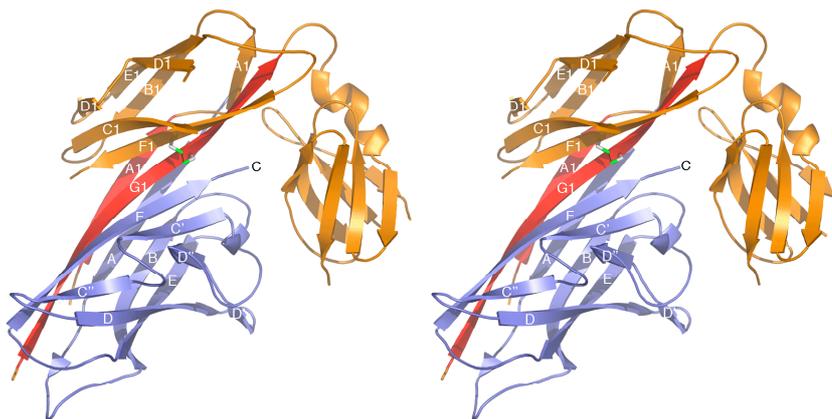


Figure 3.9. Stereo-figure of the binary complex, with the Caf1 subunit in blue and Caf1M chaperone in orange with the G₁ and the A₁ strand in red. The disulfide bridge between the G₁ and the F₁ strand shown in ball-and-stick.

The Caf1 subunit, despite virtually no sequence similarity, has a surprisingly similar fold to the pilin subunits in the FGS-systems. It shows the typical Ig-like 6 strand beta barrel with the 7th strand missing, thus exposing part of its hydrophobic core. The G₁ strand of the Caf1M chaperone is inserted parallel to the F strand of the Caf1 subunit (Figure 3.9), donating large hydrophobic residues to the subunit core. Similarly the A₁ strand of the chaperone binds in an anti-parallel mode to the A strand of Caf1. The two proteins thereby form a 'super barrel' from two beta sheets with a fused hydrophobic core. One sheet of the super barrel is made up from strand C and F from Caf1 and G₁, F₁, C₁ and D₁ from the chaperone. The other sheet consists of strands E, B and A from Caf1, and strand A₁, B₁ and E₁ from the chaperone. The same type of super barrel can be seen in the FGS-system complexes, although less apparent (Figure 3.10).

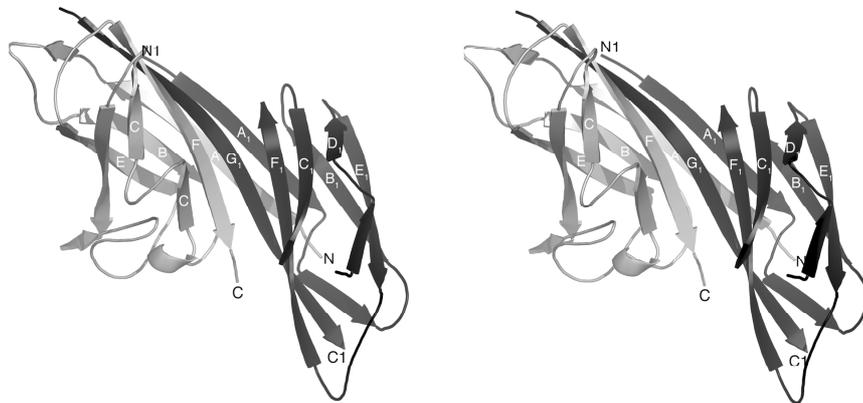


Figure 3.10. Stereo-figure of the Caf1M:Caf1 super barrel. Strands belonging to Caf1 in grey and Caf1M strands in black.

The 4 first hydrophobic side chains of the chaperone, Val126:128:130 and Phe132, are inserted into the Caf1 core but the 5th hydrophobic residue Ile134 instead packs sideways onto the cleft. The A₁-strand side chains are not inserted into the core of the subunit, but pack onto the side chains of Caf1 A-strand to form a second layer of the fused hydrophobic core. The carboxy-terminus of Caf1 is anchored in the cleft formed by the two chaperone domains, hydrogen bonding to the conserved residues Arg20 and Lys139. The amino-terminal histidine-tag is not visible in the electron density.

The long G₁-donor strand of Caf1M, typical of a FGL chaperone, is matched by a long F-strand in Caf1, which gives a correspondingly long acceptor cleft. The FGS chaperones donate three hydrophobic residues in the donor strand complementation, while Caf1M donates five. The longer acceptor cleft and

chaperone donor strand might reflect the lower complexity of the FGL systems. The FGL chaperones have one, or maximally two different subunits to recognise, and would thereby be able to bind with a higher specificity given by the longer G₁-strand with more donated residues. The FGS chaperones need to recognise and bind to several different subunits in order to assemble the more complex pilus. A shorter G₁-strand might give more flexibility to this binding, thus making the FGS chaperones more promiscuous.

3.2.2 The *Caf1M:Caf1:Caf1* ternary complex

In order to get direct evidence for donor strand exchange and to visualise the structure of the smallest possible F1 fibre, the structure of the Caf1M:Caf1:Caf1 chaperone:subunit:subunit complex was determined. A mutation of Ala9Arg in the N-terminal strand had previously been shown to hinder larger assemblies from forming (Zavialov et al., 2002). This mutant was expressed and the Caf1M:Caf1:Caf1 ternary complex purified and crystallised. The structure was solved by molecular replacement using the Caf1M:Caf1 binary complex as a search model, and refined to 2Å.

The Caf1M chaperone and the chaperone bound Caf1 subunit are very similar to the Caf1M:Caf1 binary complex, as expected. They have an overall r.m.s.d. of 0.49Å over 324 equivalent C α -positions. The major difference is observed in the amino-terminal extension, which in the binary complex is exchanged for a his-tag and disordered in the structure. In the ternary complex the amino-terminal extension has replaced the G₁ strand of the chaperone, and hydrogen bonds anti-parallel to the F strand of the Caf1 fibre subunit (Figure 3.11) providing the first direct evidence for donor strand exchange (Paper II).

The donated strand from the Caf1 subunit is termed G_{donor}, since it replaces the missing G strand in the subunit Ig-fold. The Arg9 mutation lies within the G_{donor} strand, and hinders polymerisation of Caf1 into larger assemblies. In the structure, Arg9 has a conformation that seems to be a mimic of an alanine. The first atom of the side chain, CB, points into the core of the subunit while the following atoms of the arginine side chain make a sharp turn and point out towards the solvent. Although the arginine is bent in a mimic of an alanine, it is still inevitably larger. The arginine side chain packs against Tyr23 on the Caf' A-strand, and the area around the arginine shows elevated B-factors, indicating that the region might be destabilised.

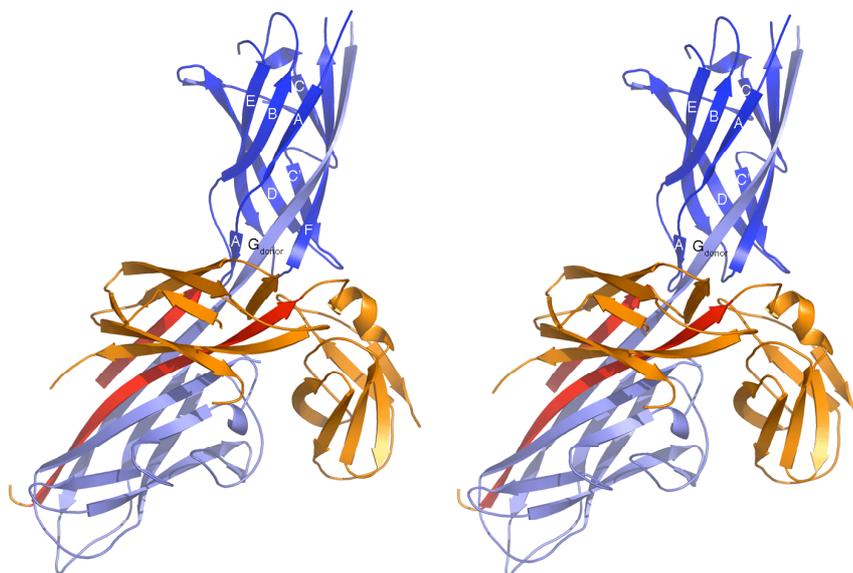


Figure 3.11. Stereo figure of the Caf1M:Caf1:Caf1 ternary complex. Chaperone in black, chaperone bound subunit in white, and the Caf1 fibre subunit in grey.

To simplify the following discussion, the chaperone bound Caf1 subunit will hereafter be termed Caf1' and the fibre subunit Caf1". Caf1" is structurally relatively different from Caf1', with an r.m.s.d. of 1.2Å over 130 equivalent C α atoms. The N-terminal strand of Caf1' replaces the G₁ plus the A₁ strand of the chaperone, hence two strands are replaced by one. There is also a size difference between the side chains of the chaperone G₁ strand and the subunit G_{donor}, the bulkier chaperone side chains Val, Val, Val, Phe and Ile are replaced by Leu, Ala, Ala, Thr and Ala (Figure 3.12).

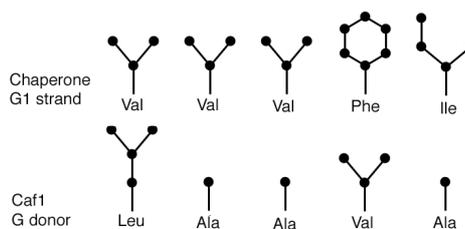


Figure 3.12. Comparison of Caf1M G₁ side chains to Caf1 G_{donor} side chains

In order to match these new conditions, the entire core of Caf1" has been rearranged to become more condensed. Aligning Caf1' with Caf1" by matching their F and C strands clearly visualises this. The entire beta sheet 1 of Caf1", made up of strands A, B, E and D, is rotated inwards in order to narrow the acceptor cleft and to fill the gaps in the core (Figure 3.13). This is particularly obvious in the upper part of the Caf1" molecule, where previously the A₁ strand of

the chaperone was bound. The Caf1" A strand has moved closer to the donor strand by around 2Å in this part of the molecule.

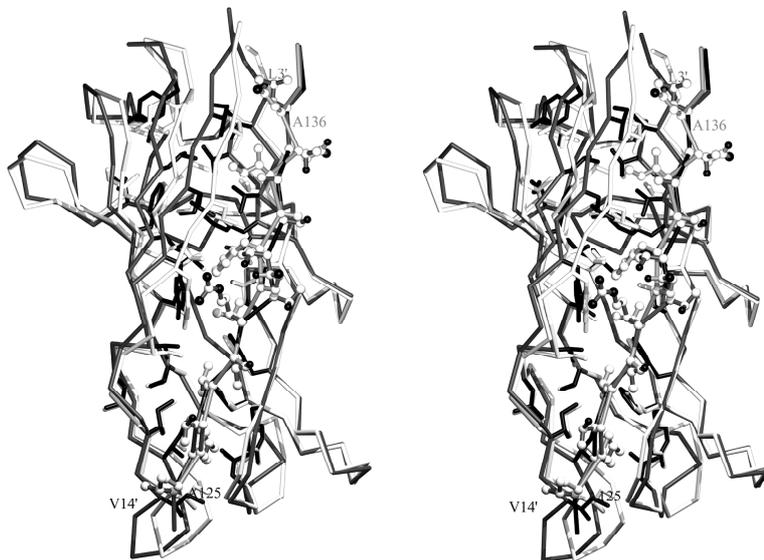


Figure 3.13. Stereo figure of the collapse of the Caf1 hydrophobic core. Caf1' is shown in black and Caf1'' in white, with the residues of the Caf1' G_{donor} in ball-and-stick to be compared with the Caf1M G₁ strand in grey.

3.2.3 Folding energy is preserved by the chaperone

Assembly of organelles via the chaperone:usher pathway does not require input of external energy. In donor strand exchange, the subunits switch from a chaperone:subunit complex to a subunit:subunit complex, which means two chaperone strands (G₁ and A₁) are replaced with only one subunit strand, the G_{donor} strand. Owing to the transition from a two-strand to a one-strand interaction, the contact area in the chaperone:subunit complex is larger than the corresponding area in the subunit:subunit complex. While the chaperone:subunit contact buries an area of 3600Å² total and 2250Å² hydrophobic surface area, the subunit:subunit interface buries only 2250Å² total and 1400Å² hydrophobic surface area. This implies a larger enthalpy of binding between the subunit and the chaperone than the subunit:subunit, unless the fit of the chaperone:subunit complex is poor.

In order to compare how well two proteins fit together in a complex, the shape correlation statistics (Sc) can be calculated (Lawrence and Colman, 1993). Sc = 0 corresponds to no geometrical fit, while Sc = 1 would be a perfect fit. Sc for the Caf1M:Caf1 complex gives a value of 0.74 for the whole interface and 0.76 for the G₁ strand. The Caf1:Caf1 interface gives a Sc of 0.78. These values are all in the range for well-fitting protein complexes, and suggest an equally good fit between

the chaperone:subunit complex and the subunit:subunit complex. This raises the question as to how donor strand exchange will ever occur without the input of external energy.

The answer is suggested by the rearranged core of the Caf1" subunit. An Sc correlation of the fit between the two beta sheets in the barrel of the Caf1' and the Caf1" subunit can be calculated. The internal fit of the Caf1' (chaperone bound) subunit gives a value of 0.58 which is indicative of a poorly packed hydrophobic core. This value should be compared to an Sc of 0.71 for the packing of the beta sheets in the Caf1" subunit, indicative of a well-packed core. The chaperone thus traps the Caf1' subunit in a poorly packed conformation, whereas the Caf1" subunit has collapsed to a near optimal packing. The chaperone can be imagined to jam the final step of the folding process, thereby preserving some of the folding energy. Release of the chaperone would allow folding of the subunit to be completed, and the drop in energy between the chaperone-bound subunit-conformation and the final, collapsed conformation, is suggested to drive fibre assembly (Figure 3.14).

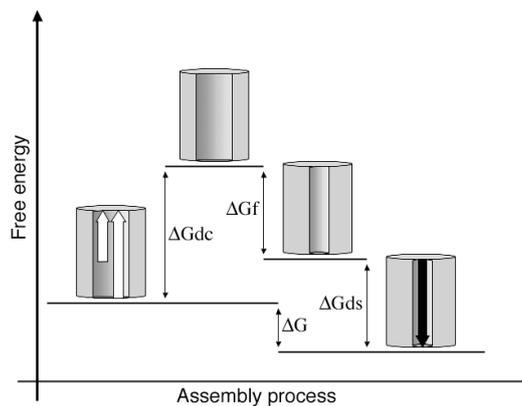


Figure 3.14. Free energy of a subunit during the assembly process. To the left in the diagram a stable complex between a subunit and a chaperone, with the G_1 and the A_1 strands of the chaperone shown as white arrows. When removing the chaperone, the open high-energy conformation of the subunit would be very unstable, with a free energy increase of ΔG_{dc} . If packing of the chaperone bound subunit proceeds to the compact form the energy would decrease with ΔG_f . The donated strand from a subunit (black arrow) can complete the fold that would drop the energy with ΔG_{ds} to the final, low-energy conformation of a subunit in a fibre. The drop in energy between the first Caf1' and the Caf1" subunit, ΔG , drives fibre assembly.

The conformation of Caf1' is expected to be very unstable in solution, whereas the Caf1" fibre subunit is folded much more like a native protein, and could be imagined to have some stability (Figure 3.14). A study by Vetsch et al (Vetsch et al., 2002) demonstrates that the pilin-domain of FimH can fold spontaneously in

the absence of the FimC chaperone, although in such conditions the subunit is only marginally stable. Self-folded FimH does not bind to the chaperone FimC to a very high extent, whereas binding of unfolded FimH is much more efficient (Vetsch et al., 2002). This implies that once the subunit has collapsed into its final packing, it is energetically unfavourable to go back to the open form needed to bind the chaperone.

The chaperone preferentially binds to unfolded subunits, and folding of subunits is promoted by the chaperone (Vetsch et al., 2002). Based on our structures we proposed a model for how the chaperone promotes subunit folding (Paper II). The chaperone can be seen as a scaffold for folding, where it provides a nucleus of large hydrophobic side chains around which the subunit can efficiently fold. The hydrophobic side chains of the chaperone would consequently be inserted deep into the hydrophobic core of the subunit, thereby effectively jam the folding process and trapping the subunit in the high-energy state.

3.3 The immunoglobulin fold in organelle subunits

3.3.1 Introduction

Members of the immunoglobulin fold family (IgFF) are evolutionary distantly related proteins, or possibly analogous proteins that have evolved towards a common, stable fold. No sequence signature can be defined for the family, and in some cases the protein 3D structure is needed for identification of a new family member. The Ig fold can be found in a large variety of organisms, and has been encountered in vertebrates, invertebrates, bacteria, viruses, fungi and plants. Proteins with this fold show highly heterogenic tissue distribution and diverse biological roles (Williams and Barclay, 1988, Halaby et al., 1999, Halaby and Mornon, 1998).

Typical for the immunoglobulin fold is the existence of 7-10 beta strands, forming a beta barrel from two beta sheets of conserved topology and connectivity. The two sheets typically consist of strands ABED/CFG, where the D strand is not obligatory and the C strand in some cases is divided into C, C' and C". B, E and F, G are the most conserved strands, while the A strand varies and is difficult to superimpose in distantly related proteins. The Ig domains often, but not always, have a disulfide bridge connecting strand B and F (Halaby et al., 1999).

3.3.2 The immunoglobulin fold in organelle subunits

The 3D structure of seven different domains from subunits of organelles assembled by the chaperone/usher pathway has been determined so far, four pilin domains and three lectin domains. They include PapK, PapE and PapG_{lectin} from P pili, FimH_{lectin} and FimH_{pilin} from type-1 pili, F17-G_{lectin} from F17 fimbriae and CafI from the F1 capsule. All domains share a common Ig-fold, as was pointed out in a recent article (Buts et al., 2003), and there is reason to believe that the Ig fold is conserved in most subunits included in the chaperone/usher pathway organelles. For the pilin domains the Ig-fold is easily identified but the lectins have evolved almost beyond recognition, which is why their membership in the Ig fold family was not initially recognised.

The beta strands belonging to the Ig-fold from all the structures can be superimposed (Figure 3.15), but extra strands and loops are quite different, in particular for the lectin domains.

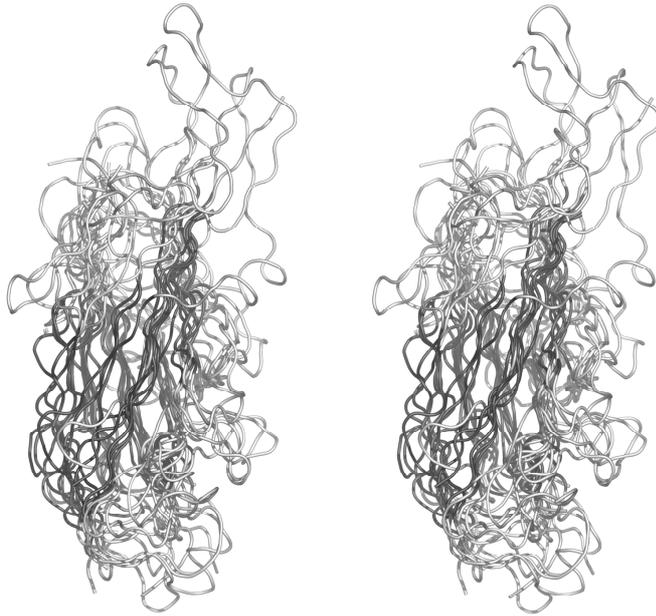


Figure 3.15. Stereo figure with all domains superimposed. The beta strands belonging to the Ig-fold are coloured black and the variable loops are in grey.

From the structural superimposition it is possible to obtain a structure based sequence alignment. To align the sequences without using 3D information turns out to be impossible, since even the most closely related of the structures determined (PapE and PapK) share only 16% sequence identity. In the sequence alignment there is not one single amino acid that is 100% conserved in all seven domains.

Conserved features

All structures determined share a pattern of strands: A, B, C, (C'), (C''), D, D', D'', E, F and (G) (Figure 3.16). All pilin subunits lack strand G of the Ig fold since they participate in DSE, and instead have an N-terminal G_{donor} strand. The structures seem to have a more or less pronounced division into three beta sheets, two front sheets and one back sheet. The upper front sheet is made up from strand A, B, E and the lower sheet from B, E, D, with some minor variations. Strands B and E consequently participate in both sheets running the whole length of the front side. The back sheet consists of strand A, G/G_{donor} , F, C, D'. The A strand is thus part of both the upper front sheet and the back sheet. It starts hydrogen bonding to the upper part of the front sheet, makes a switch in the middle and continues as part of the back sheet. In CafI this switch comes late, and it is only the very last part of strand A that forms the back sheet. In PapG lectin domain the situation is the opposite, and the A strand is part of the back sheet only and no switch occurs.

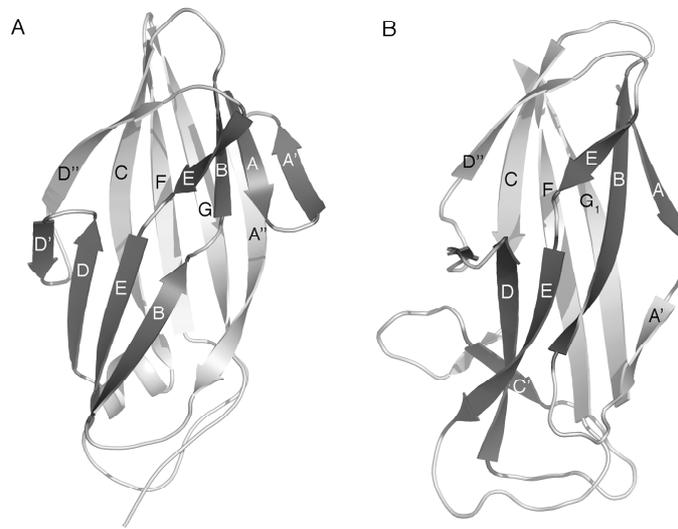


Figure 3.16. Two examples of the Ig-fold shared by the subunits. (A) FimH lectin domain and (B) FimH pilin domain.

3.3.3 Aligning the pilin subunits

The pilin domains are more similar to each other than the lectin domains, possibly because of the requirement that they must be able to participate in DSE. The four pilin domain sequences were aligned based on the structural alignment. A profile hidden Markov model (profile hmm) was built from the alignment with the online tool NPS@ (Network Protein Sequence @analysis) (Guermeur et al., 1999), and the Swiss Prot database was searched with this profile in an hmm search, again using NPS@. In the hmm search every sequence found matching the profile is added to the previous alignment, and the whole process can be used iteratively. This proved an efficient way of finding organelle subunits, and 96 sequences were discovered and aligned, all of them from FGS chaperone systems.

Investigation of the sequence alignment together with the structural superimposition revealed several features. First, the hydrophobic core is quite conserved, and a hydrophobic pattern can be seen in the alignment of strands A, B, C, D, D', E and F. The conserved hydrophobic core seem to be a general feature in the IgFF, and is thought to be the reason why the Ig fold can be maintained despite the low sequence similarity (Halaby et al., 1999). A large majority of the FGS pilin subunits posses a conserved disulfide bridge linking strand A to strand B, an unusual position for a disulfide bridge in the IgFF. This disulfide bridge is not present in the FGL subunits, which is probably one of the reasons as to why no FGL sequences were found in the hmm search.

Handle-like features on the β -barrel

Typical for pilin subunits seems to be the topology of strands C', C'' and D', D'' (Figure). The C strand starts out being part of the back sheet, then loops out into a beta hairpin consisting of C' and C'', the latter is also part of the lower front sheet. This beta hairpin is matched by a similar loop in strand D. Strand D hydrogen bonds to strand E in the lower front sheet, then loops out in two short beta strands D' and D'', where D'' in some cases continues to take part of the back sheet. These two loop structures create a handle-like feature on the surface of the Ig-beta barrel (Figure 3.17).

Based on the structure of the FimC:FimH complex, a model for the pilus rod has been constructed (Choudhury et al., 1999). The two loops are on the outside of the pilus model, where the C'-C'' loop folds up towards the subunit above, coming very close to the tip of the D'-D'' loop of that subunit. This suggests a possible stabilising function of these two loops in the final helical structure of the pilus.

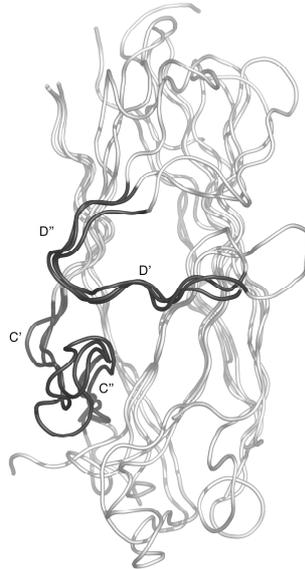


Figure 3.17. Comparison of the C'-C'' loop and the D'-D'' loop in the three FGS subunits.

The subunit acceptor cleft

The F strand is by far the most conserved strand, as earlier recognised (Hung et al., 1996). The -14 Gly of the F strand is nearly 100% conserved, only two out of 96 sequences do not have a glycine at this position. Almost equally well conserved is the penultimate Tyr. A distinct pattern of every second residue being hydrophobic can be seen, and particularly position -6 is often a phenylalanine or an equally bulky side chain while position -8 and -10 are often small, like an

alanine or a valine. The deepest part of the acceptor cleft starts just before the F-strand -6 residue, with one wall of the deep pocket defined by residue -8 and -10 (Figure 3.18).

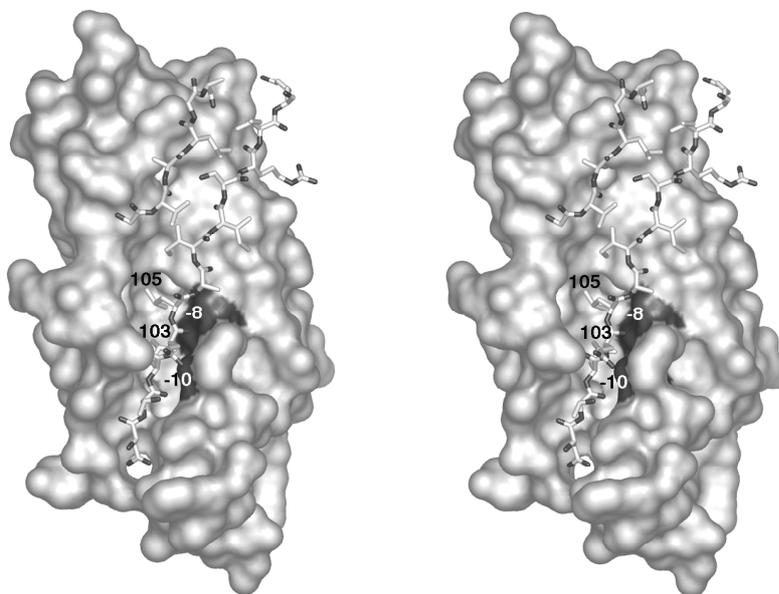


Figure 3.18. Stereo figure of FimH donor strand cavity, with position -8 and -10 coloured black. Chaperone G_1 - and A_1 -strand shown as sticks.

This deep pocket is conserved among the subunits, and aligning the structures of the three FGS pilin domains shows three structurally very conserved side chains defining the bottom of the deep cleft (Val223, Leu225 and Ala254 in FimH), in addition to the -8 and -10 positions (Figure 3.19). The FGS chaperone FimC has two leucines (103 and 105) inserted into this part of the cleft and PapD has a leucine and an isoleucine (103 and 105), both pairs superimposing very well in the two chaperone structures. Leucine 103 is quite conserved in the FGS chaperone-family (Hung et al., 1996), and is also the side chain making interactions with all three conserved amino acids in the cleft (Figure 3.19). The two bulky chaperone side chains also packs against the F-strand residues -8 and -10 , which may be why these two residues are required to be small.

One might speculate that the area comprised of the three conserved subunit residues at the bottom of the deep cleft constitutes a nucleus for chaperone-subunit interactions. The chaperone leucine 103 together with position 105 could be imagined to interact with this subunit nucleus, which could catalyse folding of the subunit around the chaperone scaffold.

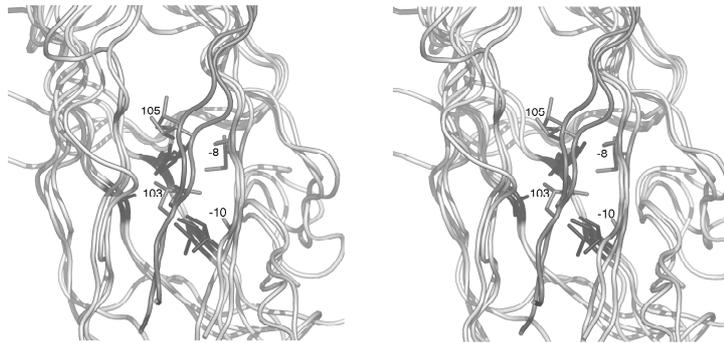


Figure 3.19. Three conserved residues at the bottom of the acceptor cleft coloured black. The chaperone FimC G₁ strand is coloured black, and FimD G₁ strand grey. The bulky side chains in position 103 and 105 are shown in sticks, as is position -8 and -10 of the subunit F-strand.

3.3.4 The lectin domains

The Ig fold can still be recognised in the lectin domains, but the variation of the fold is much greater. The lectin domains do not perform DSE, neither do they have to take part in the packing into a pilus. The only structural demand on the lectin domains, apart from the actual carbohydrate binding, is probably a size limit for passage through the outer membrane channel created by the usher. The lectin domains are accordingly very similar in width, just under 30Å, while the length varies between 53Å for FimH, 58Å for F17-G and 64Å for PapG.

Less functional pressure naturally allows for a greater structural diversity and evolution, which can be seen in the great variation of the lectin domain sequences and structures. The domains can be superimposed (Figure 3.20) and a structural based sequence alignment was made from the 3 lectin domain structures, but the hmm profile built from this alignment picked up very few, closely related sequences when used for searching the Swissprot database.

The lectin domains all contain a disulfide bridge, but the position is highly diverse. FimH_{lectin} has kept the disulfide bond between strand A and B seen in the FGS pilin domains. F17-G_{lectin} has a disulfide bridge linking strand C and D", and the disulfide bridge of PapG_{lectin} links the BC-loop to the D"E-loop.

The carbohydrate binding sites differ among the three lectins, both in shape and position (Figure 3.17). The binding sites are all created by the domain top part loops. The loop regions are the most variable part of a protein fold, and the easiest part to adapt to a binding site for a particular receptor, which might be the reason for the position of the binding sites.

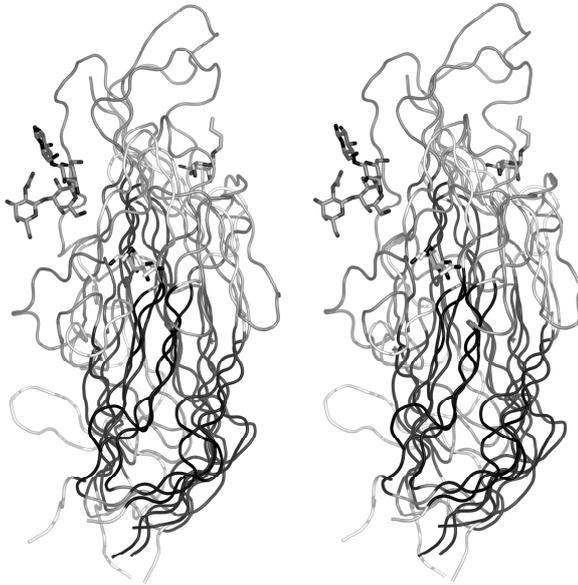


Figure 3.17. Stereo figure of the three lectin domains superimposed, with their ligands shown in sticks. The Ig-fold beta strands are coloured black.

4. Future perspectives

As is often the case as one project goes towards an end, many new issues are raised. There are always loose ends to pick up and new questions to ask. This is a selection of questions that I find of particular importance in this field of research, and to which the future hopefully will provide an answer.

4.1 Structure and sequence comparison of the organelle subunits

A very large number of organelles assembled by the chaperone:usher pathway can be found in nature. Only in the search done in this thesis 96 sequences were discovered, and that is only of FGS organelle subunits. We present the first structure of a FGL chaperone and subunit, and as more structures become known, better sequence alignments can be made and additional organelle systems can be identified. Even though all structures determined so far resemble each other, every new structure will provide more information to the puzzle. Alignments of the pilin domains give clues of DSE and the importance of the conserved hydrophobic core, and comparison of the lectin domains answer questions about receptor specificity and binding mechanism. More structures determined would therefore be of great value in understanding the fine details of the various systems.

4.2 The usher

In this thesis a theory of the driving force of donor strand exchange is presented, which is a significant step towards understanding assembly of the chaperone:usher organelles. An important question remaining to be answered is the role of the usher, and thereby the exact nature of the mechanisms behind DSE. The usher is catalysing DSE, and a structure of the usher, possibly in complex with a chaperone:subunit complex, would greatly help understanding the catalysing mechanism.

4.3 Binding phenotypes of FimH

The variation in binding phenotypes of type-1 pili to different receptors and surfaces is still not brought to a complete understanding. Binding studies to a FimH mutant where the tyrosine gate, Tyr48 and Tyr137, is changed would be of interest, both binding to FimH incorporated in a pilus and to the purified protein. Also, 3D structures of FimH from the two strains of different phenotypes, CI#4 and F18, would confirm the theory that there is no structural difference in the identified mannose-binding site, and that the underlying causes for the altered binding pattern is to be found elsewhere.

Valuable information would be whether FimH has one or more binding sites. It would be relatively simple to study by NMR, and the answer would lead one step further in the attempts to fully understand the mechanism of the binding.

5. References

- Abraham, S. N., Sun, D., Dale, J. B. and Beachey, E. H. (1988) *Nature*, **336**, 682-4.
- Backhed, F., Alsen, B., Roche, N., Angstrom, J., von Euler, A., Breimer, M. E., Westerlund-Wikstrom, B., Teneberg, S. and Richter-Dahlfors, A. (2002) *J Biol Chem*, **277**, 18198-205.
- Bahrani-Mougeot, F. K., Buckles, E. L., Lockatell, C. V., Hebel, J. R., Johnson, D. E., Tang, C. M. and Donnenberg, M. S. (2002) *Mol Microbiol*, **45**, 1079-93.
- Berglund, J. and Knight, S. D. (2003) *Adv Exp Med Biol*, **535**, 33-52.
- Binet, R., Letoffe, S., Ghigo, J. M., Delepelaire, P. and Wandersman, C. (1997) *Gene*, **192**, 7-11.
- Boisier, P., Rahalison, L., Rasolomaharo, M., Ratsitorahina, M., Mahafaly, M., Razafimahefa, M., Duplantier, J. M., Ratsifasoamanana, L. and Chanteau, S. (2002) *Emerg Infect Dis*, **8**, 311-6.
- Buts, L., Bouckaert, J., De Genst, E., Loris, R., Oscarson, S., Lahmann, M., Messens, J., Brosens, E., Wyns, L. and De Greve, H. (2003) *Mol Microbiol*, **49**, 705-15.
- CCP4, C. C. P., number 4 (1994) *Acta Crystallogr D*, **50**, 760-763.
- Chanteau, S., Ratsitorahina, M., Rahalison, L., Rasoamanana, B., Chan, F., Boisier, P., Rabeson, D. and Roux, J. (2000) *Microbes Infect*, **2**, 25-31.
- Cheng, Y. and Prusoff, W. H. (1973) *Biochem Pharmacol*, **22**, 3099-108.
- Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S. J. and Knight, S. D. (1999) *Science*, **285**, 1061-6.
- Connell, I., Agace, W., Klemm, P., Schembri, M., Marild, S. and Svanborg, C. (1996) *Proc Natl Acad Sci U S A*, **93**, 9827-32.
- Cowan, C., Jones, H. A., Kaya, Y. H., Perry, R. D. and Straley, S. C. (2000) *Infect Immun*, **68**, 4523-30.
- Drancourt, M. and Raoult, D. (2002) *Microbes Infect*, **4**, 105-9.
- Drenth, J. (1994) *Principles of Protein X-ray Crystallography*, Springer-Verlag, New York.
- Du, Y., Rosqvist, R. and Forsberg, A. (2002) *Infect Immun*, **70**, 1453-60.
- Firon, N., Ashkenazi, S., Mirelman, D., Ofek, I. and Sharon, N. (1987) *Infect Immun*, **55**, 472-6.
- Firon, N., Ofek, I. and Sharon, N. (1984) *Infect Immun*, **43**, 1088-90.
- Foxman, B. (2002) *Am J Med*, **113 Suppl 1A**, 5S-13S.
- Giacovazzo, C., Monaco, H. L., Artiolo, G., Viterbo, D., Ferraris, G., Gilli, G., Zanotti, G. and Catti, M. (2002) *Fundamentals of Crystallography*, Oxford University Press.
- Glish, G. L. and Vachet, R. W. (2003) *Nat Rev Drug Discov*, **2**, 140-50.
- Guermeur, Y., Geourjon, C., Gallinari, P. and Deleage, G. (1999) *Bioinformatics*, **15**, 413-21.
- Halaby, D. M. and Mornon, J. P. (1998) *J Mol Evol*, **46**, 389-400.
- Halaby, D. M., Poupon, A. and Mornon, J. (1999) *Protein Eng*, **12**, 563-71.
- Hanisch, F. G., Hacker, J. and Schrotten, H. (1993) *Infect Immun*, **61**, 2108-15.

- Henderson, I. R., Navarro-Garcia, F. and Nataro, J. P. (1998) *Trends Microbiol*, **6**, 370-8.
- Hetenyi, C. and van der Spoel, D. (2002) *Protein Sci*, **11**, 1729-37.
- Hinnebusch, B. J., Rudolph, A. E., Cherepanov, P., Dixon, J. E., Schwan, T. G. and Forsberg, A. (2002) *Science*, **296**, 733-5.
- Holmgren, A. and Branden, C. I. (1989) *Nature*, **342**, 248-51.
- Horovitz, A. and Levitzki, A. (1987) *Proc Natl Acad Sci U S A*, **84**, 6654-8.
- Hueck, C. J. (1998) *Microbiol Mol Biol Rev*, **62**, 379-433.
- 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) *Mol Microbiol*, **44**, 903-15.
- Hung, D. L., Knight, S. D., Woods, R. M., Pinkner, J. S. and Hultgren, S. J. (1996) *Embo J*, **15**, 3792-805.
- Jacob-Dubuisson, F., Striker, R. and Hultgren, S. J. (1994) *J Biol Chem*, **269**, 12447-55.
- Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard (1991) *Acta Crystallogr A*, **47 (Pt 2)**, 110-9.
- Khan, A. S., Kniep, B., Oelschlaeger, T. A., Van Die, I., Korhonen, T. and Hacker, J. (2000) *Infect Immun*, **68**, 3541-7.
- Knight, S. D. (2000) *Acta Crystallogr D Biol Crystallogr*, **56 (Pt 1)**, 42-7.
- Knight, S. D., Choudhury, D., Hultgren, S., Pinkner, J., Stojanoff, V. and Thompson, A. (2002) *Acta Crystallogr D Biol Crystallogr*, **58**, 1016-22.
- Korhonen, T. K., Vaisanen-Rhen, V., Rhen, M., Pere, A., Parkkinen, J. and Finne, J. (1984) *J Bacteriol*, **159**, 762-6.
- Kraulis, P. J. (1991) *Journal of applied crystallography*, **24**, 946-950.
- La Fortelle, E. and Bricogne, G. (1997) In *Macromolecular Crystallography*, Vol. 276 (Eds, Carter, C. W. and Sweet, R. M.) Academic Press, pp. 472-494.
- Lawrence, M. C. and Colman, P. M. (1993) *J Mol Biol*, **234**, 946-50.
- Mansotte, F. (1997) *Sante Publique*, **9**, 135-44.
- Martinez, J. J., Mulvey, M. A., Schilling, J. D., Pinkner, J. S. and Hultgren, S. J. (2000) *Embo J*, **19**, 2803-12.
- McRee, D. E. (1999) *Practical Protein Crystallography*, Academic Press.
- Min, G., Stolz, M., Zhou, G., Liang, F., Sebbel, P., Stoffler, D., Glockshuber, R., Sun, T. T., Aebi, U. and Kong, X. P. (2002) *J Mol Biol*, **317**, 697-706.
- Morris, G. M., Goodsell, D. S., Huey, R. and Olson, A. J. (1996) *J Comput Aided Mol Des*, **10**, 293-304.
- Morris, R. J., Perrakis, A. and Lamzin, V. S. (2002) *Acta Crystallogr D Biol Crystallogr*, **58**, 968-75.
- Mulvey, M. A. (2002) *Cell Microbiol*, **4**, 257-71.
- Mulvey, M. A., Schilling, J. D. and Hultgren, S. J. (2001) *Infect Immun*, **69**, 4572-9.
- Nowicki, B., Selvarangan, R. and Nowicki, S. (2001) *J Infect Dis*, **183 Suppl 1**, S24-7.
- Nunn, D. (1999) *Trends Cell Biol*, **9**, 402-8.
- Ofek, I., Hasty, D. L. and Doyle, R. J. (2003) *Bacterial Adhesion to Animal Cells and Tissues*, ASM Press, Washington.

- Pellecchia, M., Guntert, P., Glockshuber, R. and Wuthrich, K. (1998) *Nat Struct Biol*, **5**, 885-90.
- Perry, R. D. and Fetherston, J. D. (1997) *Clin Microbiol Rev*, **10**, 35-66.
- Petterson, J., Holmstrom, A., Hill, J., Leary, S., Frithz-Lindsten, E., von Euler-Matell, A., Carlsson, E., Titball, R., Forsberg, A. and Wolf-Watz, H. (1999) *Mol Microbiol*, **32**, 961-76.
- Price, S. B., Cowan, C., Perry, R. D. and Straley, S. C. (1991) *J Bacteriol*, **173**, 2649-57.
- Ramalingaswami, V. (1995) *Nat Med*, **1**, 1237-9.
- Rini, J. M. (1995) *Annu Rev Biophys Biomol Struct*, **24**, 551-77.
- Ronald, A. (2002) *Am J Med*, **113 Suppl 1A**, 14S-19S.
- Sauer, F. G., Barnhart, M., Choudhury, D., Knight, S. D., Waksman, G. and Hultgren, S. J. (2000) *Curr Opin Struct Biol*, **10**, 548-56.
- Sauer, F. G., Futterer, K., Pinkner, J. S., Dodson, K. W., Hultgren, S. J. and Waksman, G. (1999) *Science*, **285**, 1058-61.
- Schembri, M. A., Hasman, H. and Klemm, P. (2000) *FEMS Microbiol Lett*, **188**, 147-51.
- Sokurenko, E. V., Chesnokova, V., Doyle, R. J. and Hasty, D. L. (1997) *J Biol Chem*, **272**, 17880-6.
- Sokurenko, E. V., Chesnokova, V., Dykhuizen, D. E., Ofek, I., Wu, X. R., Krogfelt, K. A., Struve, C., Schembri, M. A. and Hasty, D. L. (1998) *Proc Natl Acad Sci U S A*, **95**, 8922-6.
- Sokurenko, E. V., Courtney, H. S., Abraham, S. N., Klemm, P. and Hasty, D. L. (1992) *Infect Immun*, **60**, 4709-19.
- Sokurenko, E. V., Courtney, H. S., Maslow, J., Siitonen, A. and Hasty, D. L. (1995) *J Bacteriol*, **177**, 3680-6.
- Stathopoulos, C., Hendrixson, D. R., Thanassi, D. G., Hultgren, S. J., St Geme, J. W., 3rd and Curtiss, R., 3rd (2000) *Microbes Infect*, **2**, 1061-72.
- Thanassi, D. G. and Hultgren, S. J. (2000a) *Methods*, **20**, 111-26.
- Thanassi, D. G. and Hultgren, S. J. (2000b) *Curr Opin Cell Biol*, **12**, 420-30.
- Thanassi, D. G., Saulino, E. T. and Hultgren, S. J. (1998a) *Curr Opin Microbiol*, **1**, 223-31.
- Thanassi, D. G., Saulino, E. T., Lombardo, M. J., Roth, R., Heuser, J. and Hultgren, S. J. (1998b) *Proc Natl Acad Sci U S A*, **95**, 3146-51.
- Thomas, W. E., Trintchina, E., Forero, M., Vogel, V. and Sokurenko, E. V. (2002) *Cell*, **109**, 913-23.
- Titball, R. W., Hill, J., Lawton, D. G. and Brown, K. A. (2003) *Biochem Soc Trans*, **31**, 104-7.
- Titball, R. W. and Williamson, E. D. (2001) *Vaccine*, **19**, 4175-84.
- Vetsch, M., Sebbel, P. and Glockshuber, R. (2002) *J Mol Biol*, **322**, 827-40.
- Williams, A. F. and Barclay, A. N. (1988) *Annu Rev Immunol*, **6**, 381-405.
- Wu, X. R., Sun, T. T. and Medina, J. J. (1996) *Proc Natl Acad Sci U S A*, **93**, 9630-5.
- Zavialov, A., Berglund, J. and Knight, S. D. (2003a) *Acta Crystallogr D Biol Crystallogr*, **59**, 359-62.
- Zavialov, A. V., Batchikova, N. V., Korpela, T., Petrovskaya, L. E., Korobko, V. G., Kersley, J., MacIntyre, S. and Zav'yalov, V. P. (2001) *Appl Environ Microbiol*, **67**, 1805-14.

- Zavialov, A. V., Berglund, J., Pudney, A. F., Fooks, L. J., Ibrahim, T. M., MacIntyre, S. and Knight, S. D. (2003b) *Cell*, **113**, 587-96.
- Zavialov, A. V., Kersley, J., Korpela, T., Zav'yalov, V. P., MacIntyre, S. and Knight, S. D. (2002) *Mol Microbiol*, **45**, 983-95.
- Zhou, G., Mo, W. J., Sebbel, P., Min, G., Neubert, T. A., Glockshuber, R., Wu, X. R., Sun, T. T. and Kong, X. P. (2001) *J Cell Sci*, **114**, 4095-103.

6. Acknowledgments

Stefan, you are a genuinely cool person, and I want to be just like you when I grow up. Thanks for these past years, you have taught me more than I can remember ;-) Science is like rock'n'roll, right? A big thanks also goes to the rest of the Knight family, **Mia**, **Adrian** and **Edgar** for great dinners and late discussions at stugan! **Anton**, it was after you came to the lab things started rolling, and the projects suddenly worked. You are an amazing scientist, and I have watched and learned (I wish...). **Deva**, thanks for all the help to set me off in the beginning. **Mark Schembri** for providing me with all the FimH_r constructs, and **Thomas Norberg** for helping me with MS measurements without hesitation, and **Pär Säfsten** for Biacore collaboration.

Emma, we have come a long way together. Thanks for all these years, both in and outside the lab. You are the best! I don't want to believe we end here, Jakobsson & Berglund AB lives on in my dreams! We just have to find that idea...

Johanna Björkman, I still can't believe you are not here to share this day...

All of you who left the lab: **Andreas**, you are a fantastic combination of being really cool, really nice, and extremt gnällig! Winning combination and Traktorflickan misses you... **Adam**, your enthusiasm for science is what got me started in the first place. **Anke**, we really should spend more time outside the lab! **Anke & Remco**, thanks for skating, Talisman, dinners and home made wine. **Jeff** and **Karl**, without you the beer-club is dying and the corridors silent. Thanks for many great, late nights! **Kenth**, thanks for beer, dart and shopping in Grenoble. **Susanna**, **Tove**, **Elles**, **Karin K** and **Isabella** you are the best of girls! **Richard**, I miss your laugh and all your white teeth, **Rams** thanks for your energy and **Gisela** for your enthusiasm, and always being up for some fun.

alla@xray: **Mark**, where do I begin? Thanks for kayaking, sailing, climbing, skating and endless discussions about houses, house renovations and cordless powertools. **Gösta**, I am sorry we bought such a small car, but very grateful that you despite this always are up for a ride in it. **Talal**, thanks for all the great dinners at your place, and for always noticing new hairstyle and clothes! **Linda**, you are the absolute best roommate, I just wish you would be slightly less efficient and organised ;-) **Malin and Rosie**, thanks for shaping up the group-meetings and for having the energy to care. **Fredrik** (slem), I miss the competition on the climbing wall, **Andrea**, thanks for always arranging a party or a ski trip. **Martin W**, **Urszula** and **Ola** for bringing new spirit in when most needed, **Louise** for keeping track of everyones age, **Yonathan** for making me try to understand Danish and **Nisse** for at least trying to shape up my computer. Thank you **Calle**, **Magnus**, **Sara**, **Alexandra** and **Nic** for bringing new life into the Janos-corridor, **Gunilla** for Trocadero, folkdancing and office-sharing and

Llano for talking so much. **Karin V** for always arranging the Friday lunches, and for letting me win the car-race sometimes. **Erling, Christer** and **David** for invaluable support! **Elleanor** for all the paper-work help, and for your fantastic clothes. You have brightened up many of my days! Thank you **Ulla** for true care, **Hasse** for all the semlas and **Margareta** for keeping things in order. **Janos** and **Inger** for generosity and scientific enthusiasm. **Saeid, Martin S., Tom** you are great guys, thanks for all the help! **Ulrika**, thanks for the hint about Carlsberg, and **Eva-Lena** for synchrotron trips and thesis-support. **Jimmy** for being my sambo for so long, **Patrik** for all the great outdoors sessions and **Annette & Klätter-Nisse** for fun at the climbing wall. **Alina** for true girl-power and **Lena** for great dancing at the defense parties. **Tex**, thanks for your great sense of humor, **Gunnar** for being Gunnar, **Mats** for always helping, **Anna** for climbing and help with SHARP. Good luck with your baby! **Jerry** for appreciating crazy clothes, **Lars** for great lunches, we are almost related you know! **Seved** for nice Friday visits and **Totte** for wings and gloria. **Nina** for going house-spotting with me, **Adrian** for your smile, **Wimal** for being so nice and **Daniel** for being the only one who knows more stupid facts than Al. **Jonas** for crazy discussions at lunch, **Gerard** for grrreat coupons, **Alwyn** for help with O. **Johan Åquist** and all the mother-in-law-dream-guys in your groups, thanks for filling the coffee-breaks with discussions about poker nights. You really need a girl in your group ;-) **Sherry, Wojtec, Fariborz, Henke, Marian, Viktor, Kaspars** THANK YOU!

A special thanks to **Stallet climbing wall** and **www.hemnet.se** for being great time-wasters...

Familjen: **mamma** och **pappa**, tack för att ni tror att jag kan klara vad som helst! **Mormor** och **Farmor**, ni är världens tuffaste tanter! Alla syskonen: **Poie, Oskar** och **Kalle**, tack för att ni finns. **Nicole**, du är ett pucko, men en himla bra tjej.

Tack alla vänner och ett speciellt tack till **Falu-gänget**, Louise, Johan, Pippin den lille Nicolin, Sara E, Anna-Eva, Jesper, Hanseman, Sofia och Lova, plus heders faluborna Emma & Iker, det känns himla skönt att ni alla fortfarande finns kvar.

Al, my crazy, red-haired, pierced and dread-locked Tasmanian. What are the chances for a girl from Falun to ever meet someone like you? (Actually 1 in 2000000000000 according to Mark...) You are truly unique and you are my hero...

Jenny