Chapter 1

General Introduction

Samta Jain, Arnold J.M. Driessen and Chris van der Does
Introduction

General Introduction

1. Neisseria gonorrhoeae

Neisseria gonorrhoeae is a Gram-negative diplococcus. It belongs to the class β-proteobacteria and the genus Neisseria. This genus has two pathogenic members, N. gonorrhoeae and N. meningitides both of which are obligate human pathogens. The former infects the genital organs causing the sexually transmitted disease “gonorrhea” and the latter infects the nasopharynx causing “meningitis”. The genus Neisseria is named after Albert Ludwig Sigesmund Neisser who discovered the causative agent for gonorrhea and described it in 1879 [1]. Several N. gonorrhoeae strains have developed a high resistance against many different antibiotics which hampers the control of the spread of gonorrhea and not only has adverse effects on vaccine development but also on the transmission of HIV [2, 3].

N. gonorrhoeae is a sensitive bacterium that is prone to undergo autolysis. For growth it requires 3-10% CO₂ [4] and in the laboratory it is sensitive to buffer conditions, the pH, and the temperature. Adverse conditions often induce autolysis, but different members of the Neisseria species show a variation in their ability to autolysse. Peptidoglycan hydrolysis and outer membrane stability seem to be the most important factors affecting this. N-acetyl muramyl-L-alanine amidase is the main peptidoglycan hydrolase and has a very high activity during cell division. Cell division is indeed a prolonged process in N. gonorrhoeae probably leading to the diplococcal morphology [5-10].

A number of lytic transglycosylases producing peptidoglycan fragments have been identified in gonorrhoeae and a "hyperautolytic" effect is seen in gonorrhoeae strain over-expressing the N-Acetylmuramyl-L-Alanine Amidase enzyme AmiC [11-13]. Unlike in many other bacteria, the toxic peptidoglycan fragments are released by N. gonorrhoeae during growth. These fragments along with other cellular components like porins and lipoooligosaccharide contribute to infection of the host cells [14, 15]. Resistance against many different antibiotics has significantly increased in clinical isolates of N. gonorrhoeae over the last decades. The rapid spread of antibiotic resistance in N. gonorrhoeae is caused by its ability to rapidly take up DNA from the
environment. This DNA, among others, released by autolysis is taken up and recombined into the chromosome by homologous recombination [16, 17].

A recently discovered genetic island was shown to contain a Type IV secretion system, which was involved in the secretion of ssDNA. DNA secreted via this system is also taken up by *N. gonorrhoeae*. This transfer of DNA aids the organism to acquire additional genetic information like antibiotic markers and genetic variability resulting in enhanced survival [18]. Neisseria specifically takes up DNA that has a 12 base-pair DUS uptake sequence (5'-ATGCCGTCTGAA-3'). This DUS sequence is present at a frequency of about 1 in 1200 base-pairs [19, 20]. DNA uptake in Neisseria is brought about by the Type IV pili and the competence system which are involved in transport across the outer and the inner membrane, respectively [21].

In order to evade the host immune response, *N. gonorrhoeae* has the capability to undergo phase and antigenic variation of surface exposed components. This includes changes in the number and expression of opacity proteins, the sialylation state of the lipooligosaccharide and recombination dependent pilin variation [22, 23]. Recent genome sequencing data suggests that many more genes might undergo phase variation [24].

These and other remarkable characteristics like twitching motility, IgA protease secretion, iron acquisition and biofilm formation have provided *Neisseria* with an adaptation profile to become a human pathogen making it an interesting and a challenging organism to study.

The aim of this thesis was to study the mechanism of DNA release via the Type IV secretion system encoded within the gonococcal genetic island and to study the outer membrane structural components of the Type IV pili system. Both these systems, involved in the transfer of DNA are described below in detail.

2. **Type IV secretion systems**

Type IV secretion systems (T4SSs) are large multi-protein complexes, consisting of 12-24 proteins, spanning both the inner and outer membrane of Gram-negative bacteria (see Fig. 1 for an overview). They are highly versatile and functionally diverse
secretion systems. Based on their function, the T4SSs can be divided into three subfamilies:

**I) Conjugative systems**

Conjugative systems are the largest family of the T4SS. They are required for the transfer of DNA within and between species. The prototype of this family is the well studied F-plasmid conjugative system [25]. The F-factor was first described in 1953 as the fertility factor determining bacterial sexuality [26]. Conjugation is the mechanism by which single stranded DNA is transferred unidirectionally from the 5’ to the 3’ end via cell to cell contact from a donor to a recipient cell. Plasmids as well as the integrative and conjugative genomic elements (ICEs) harbor the genes required for conjugation [27]. Genes flanking the T4SSs can also be transferred during this process increasing the genomic plasticity [28]. The conjugative systems can be grouped into families, like the F-type (e.g. F-plasmid), I-type (e.g. R64 plasmid) and P-type (e.g. RP4 plasmid) plasmids. Depending upon their ability to coexist in one cell, they can also be classified into several incompatibility groups, for example, the F- plasmid belongs to the IncF plasmid incompatibility group; the RP4 plasmid to the IncP incompatibility group; the R388 plasmid to the IncW and the pKM101 plasmid belongs to the IncN incompatibility group [29].

**II) Substrate secretion systems**

A second large group of T4SSs is formed by the substrates secretion systems that can deliver virulence molecules to the host cells. The adaptive utilization of the T4SS by several microbes has often resulted in the secretion of substrates like toxins or other effector molecules that modify the behavior of the target cell. Thus many of the proteins secreted via T4SS are involved in pathogenicity. Both, intra and extra cellular delivery of substrates to the host cells have been observed. The T4SS of *Agrobacterium tumefaciens* which causes crown gall disease is the best studied example of this family; it however also belongs to the family of conjugative secretions systems as it delivers T-DNA to the plant cell [30]. Substrates of T4SS are diverse and many different ones have been identified in the last years.
There are several systems which have been studied in detail: the T4SS of *A. tumefaciens* that transports proteins like VirE2, VirE3, VirF and VirD5 to the acceptor cell independently from the DNA [31, 32]; the T4SS encoded by the *cag* pathogenicity island of *Helicobacter pylori* that delivers the CagA protein that compromises host cell structure [33]; the Dot/Icm T4SS of *Legionella pneumophila* that secretes, among others, DotF and RalF affecting the host macrophages [34], and the Ptl T4SS of *Bordetella pertussis* that secretes the Ptl toxin into the medium. In contrary to other systems, the Ptl T4SS of *Bordetella pertussis* uses a two-step secretion mechanism. Here, the Ptl toxin is secreted first to the periplasm via the Sec machinery and then across the outer membrane via the T4SS machinery [35]. Examples of pathogens that use T4SS to deliver substrates to their eukaryotic host are given in Table 1.

**Table 1: T4SS and pathogenesis**

<table>
<thead>
<tr>
<th>Pathogens</th>
<th>T4SS</th>
<th>disease caused</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A. tumefaciens</em></td>
<td>VirB</td>
<td>Plant crown gall</td>
<td>[36]</td>
</tr>
<tr>
<td><em>Legionella pneumophila</em></td>
<td>Dot/Icm</td>
<td>Legionnaires' Disease</td>
<td>[37]</td>
</tr>
<tr>
<td><em>Brucella spp.</em></td>
<td>VirB</td>
<td>Various</td>
<td>[38]</td>
</tr>
<tr>
<td><em>Bordetella pertussis</em></td>
<td>Ptl</td>
<td>whooping cough in human</td>
<td>[39]</td>
</tr>
<tr>
<td><em>Bartonella spp.</em></td>
<td>VirB</td>
<td>Various</td>
<td>[40]</td>
</tr>
<tr>
<td><em>H. pylori</em></td>
<td>Cag</td>
<td>human gastric ulcer</td>
<td>[41]</td>
</tr>
<tr>
<td><em>Campylobacter</em></td>
<td>pVir</td>
<td>human bacterial gastroenteritis</td>
<td>[42]</td>
</tr>
</tbody>
</table>

**III) DNA uptake and secretion systems**

The versatility of T4SS is substantiated by this third and smallest subfamily, which involves contact independent uptake or release of DNA. The only example of a Type IV secretion system involved in the uptake of DNA is the ComB system of *H. pylori* [43]. Two operons of this system named *comB2-B4* and *comB6-B10* encode proteins
Introduction

homologous to the VirB T4SS of *A. tumefaciens*. DNA uptake in *H. pylori* is a two step mechanism where double stranded DNA is imported into the periplasm in a ComB T4SS dependent manner and the ComEC channel probably transports the single stranded DNA into the cytoplasm [44]. The only known T4SS involved in mediating DNA release into the milieu is encoded in the gonococcal genetic island (GGI) of *N. gonorrhoeae* [45]. The proteins encoded by the GGI shares similarity to the T4SS encoded on the F-plasmid. The DNA released from the organism is eventually taken up by the neighboring gonococci by natural transformation thereby promoting exchange of genetic material for better survival. GGI of *N. gonorrhoeae* is described in more detail below.

The T4SS of the Ti plasmid of *A. tumefaciens* is currently the best characterized T4SS. It consists of the targeting components VirD1 through VirD4 and 11 other VirB proteins called VirB1 to VirB11. Most of these proteins are conserved among the known T4SSs of other bacteria. T4SSs of other bacteria can contain additional components or have homologs for only a subset of these proteins. Unfortunately naming of the different proteins involved in type IV secretion systems is often confusing. Homologous proteins with a similar function have often been given different names in different systems, for example the homolog of the conjugal coupling protein VirD4 of *A. tumefaciens*, has been named TraD in the F plasmid T4SS, TraG in the RP4 plasmid T4SS, TrwB in the R388 plasmid T4SS and TraJ in the pKM101 T4SS. Also, proteins with similar names can perform very different functions in different systems, for example TraA of F plasmid is the major pilus subunit but in pKM101 TraA is an inner membrane pilus assembly protein and is homologous to TraL of F plasmid. Table 2 includes an overview of the names of the VirD1 through VirD4 and VirB1 through VirB11 proteins in conjugative plasmids of different incompatibility groups, F (IncF), RP4 (IncP), R388 (IncW) and pKM101 (IncN). In this introduction, names of proteins involved in Type IV secretion will be followed by the names of their homologs in the *A. tumefaciens* Ti plasmid (Vir system). This thesis is focused on the Type IV secretion system encoded within the gonococcal genetic island of *N. gonorrhoeae*; many proteins within this system resemble those from well studied F plasmid conjugation system (see below). When applicable, the names will be followed with the system they belong to as subscript (e.g. TraI of the F plasmid as TraI_F).
Table 2: Nomenclature of the different T4SSs

<table>
<thead>
<tr>
<th>pTi</th>
<th>F plasmid</th>
<th>GGI</th>
<th>RP4</th>
<th>R388</th>
<th>pKM101</th>
<th>function</th>
</tr>
</thead>
<tbody>
<tr>
<td>virD2</td>
<td>tral</td>
<td>tral</td>
<td>tral</td>
<td>trwC</td>
<td>tral</td>
<td>DNA relaxase</td>
</tr>
<tr>
<td>virD4</td>
<td>traD</td>
<td>traD</td>
<td>traG</td>
<td>trwB</td>
<td>tral</td>
<td>conjugal coupling protein</td>
</tr>
<tr>
<td>virB1</td>
<td>orf169</td>
<td>trbN</td>
<td>trwN</td>
<td>trl</td>
<td></td>
<td>lytic transglycosylase</td>
</tr>
<tr>
<td>virB2</td>
<td>traA</td>
<td>traA</td>
<td>trbC</td>
<td>trwL</td>
<td>tral</td>
<td>major pilus component</td>
</tr>
<tr>
<td>virB3</td>
<td>traL</td>
<td>traL</td>
<td>trbD</td>
<td>trwM</td>
<td>tral</td>
<td>pilus assembly</td>
</tr>
<tr>
<td>virB4</td>
<td>traC</td>
<td>traC</td>
<td>trbE</td>
<td>trwK</td>
<td>tral</td>
<td>conjugal transfer ATPase</td>
</tr>
<tr>
<td>virB5</td>
<td>traE</td>
<td>traE</td>
<td>trbJ</td>
<td>trwJ</td>
<td>tral</td>
<td>minor pilus subunit</td>
</tr>
<tr>
<td>virB6</td>
<td>traG</td>
<td>traG</td>
<td>trbL</td>
<td>trwL</td>
<td>tral</td>
<td>inner membrane protein</td>
</tr>
<tr>
<td>virB7</td>
<td>traV</td>
<td>traV</td>
<td>trbH</td>
<td>trwH</td>
<td>tral</td>
<td>core complex component</td>
</tr>
<tr>
<td>virB8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>tral</td>
<td>inner membrane protein</td>
</tr>
<tr>
<td>virB9</td>
<td>traK</td>
<td>traK</td>
<td>trbG</td>
<td>trwF</td>
<td>traO</td>
<td>core complex component</td>
</tr>
<tr>
<td>virB10</td>
<td>traB</td>
<td>traB</td>
<td>trbl</td>
<td>trwE</td>
<td>traF</td>
<td>core complex component</td>
</tr>
<tr>
<td>virB11</td>
<td>-</td>
<td>-</td>
<td>trbB</td>
<td>trwD</td>
<td>tral</td>
<td>conjugal transfer ATPase</td>
</tr>
</tbody>
</table>

**Relaxases**

Relaxases have been given their name based on their ability to relax supercoiled DNA. Relaxases make a site and strand specific nick in the oriT region and then covalently bind to the 5’ end of one strand of the DNA [46-48]. Different accessory proteins can aid the relaxase in this step. These accessory proteins have a conserved ribbon-helix-
Introduction

helix motif and are found in many T4SS [49]. For example, the TrwA and NikA accessory proteins stimulate the relaxation activity of TrwC (R388 plasmid) and NikB (R64 plasmid) relaxases, respectively [50, 51]. The accessory proteins assemble into a complex called the relaxosome at the oriT region of the DNA for the initiation of strand transfer [52]. Generally accessory proteins encoded in the same operon as the relaxase are involved in relaxosome formation. They have DNA strand specificity and are proposed to bind and process DNA and then present the cleavage site to the relaxase suggesting that the relaxase recognizes a nucleoprotein complex [50, 53-57]. Several relaxases have a C-terminal domain that can dimerize and is required for interaction with the accessory proteins [22]. In standard, in vitro relaxase/nicking assays, accessory proteins are often essential to nick the supercoiled DNA. All relaxases are large proteins that contain an N-terminal relaxase domain involved in the unidirectional and site specific DNA cleavage reaction [48, 52, 58, 59]. Many relaxases also contain a C-terminal helicase domain involved in unwinding the DNA. When present, both domains are essential for the translocation process [60-62]. Although T4SSs are highly diverse, all conjugative T4SSs contain a relaxase.

Six different families of relaxases have been identified. They were named MOB_F, MOB_H, MOB_Q, MOB_C, MOB_P and MOB_V [63]. The MOB_F relaxases are characterized by the presence of a helicase domain at the C-terminus. Furthermore, the relaxase domain at the N-terminus harbors three conserved motifs, motif I, II and III represented by two catalytic tyrosine residues (motif I), a conserved aspartate (motif II) and a conserved histidine triad (HUH) motif. Examples of this family are TraI of the F plasmid and TrwC of the R388 plasmid. Relaxases of the MOB_C family do not contain the conserved tyrosine or the histidine motif and have a different signature motif represented as D-x6–17-E-x-E-(RL)-x2-K-x3-R-(YF). Examples of this family are MobC of the CloDF13 plasmid and TraX of the pAD1 plasmid. The MOB_Q family resembles the MOB_F family, and also contains three conserved motifs at the N-terminus. However, the aspartate of motif II is located closer to the tyrosine of motif I and this may have implication on the activity. The MobA relaxase of RSF1010 is an example of this family. Relaxases of the MOB_P family also contain three different conserved motifs, these motifs however contain a variation in motif II and have only one catalytic tyrosine residue. This family can be subdivided into at least 7 subfamilies.
Examples of this family are TraI of the RP4 plasmid and VirD2 of pTi plasmid. Relaxases of the MOB_v family harbor the conserved motif I and the extended motif III. They are exemplified by the MobM relaxase of the pMV158 plasmid. MOB_h is a novel family that includes relaxases from conjugative plasmids belonging to incompatibility group IncH, IncJ, IncT, IncP7 and IncA/C as well as those from ICEs and genomic islands. The two conserved sequence motifs of this family are a HD hydrolase motif and an alternate histidine motif. The TraI relaxase of the R27 plasmid and the TraI plasmid encoded within the GGI belong to this family.

It has been shown that relaxases bind DNA at the oriT and that they cleave the DNA at the nic site. The oriT sequence often contains an inverted repeat sequence that forms a hairpin loop close to the nic site [64-66]. This was confirmed when the relaxase domains of the F [67, 68] and R388 [65, 66] plasmids of the MOB_f family were purified and crystallized. These domains bind specific to oriT sequences in vitro with nM affinity [61, 66, 69, 70]. The crystal structure of the relaxase domain bound to DNA showed that the inverted repeat bound in a ‘knob-into-hole’ manner to the relaxase, and also showed that the inverted repeats had complex interactions with the relaxase [67]. The specificity can be altered by a few amino acid substitutions in the relaxase domain [71]. Remarkably, DNA binding, cleavage and conjugation also takes place efficiently when the hairpin structure is disrupted [72]. The crystal structures further showed that the conserved tyrosine residue that forms the covalent bond with the DNA is in close proximity to the cleavage site, and the HUH motif is involved in coordinating a metal ion, the presence of which is essential for the relaxase activity [73, 74]. Both, in vitro and in vivo [75] results indicate that relaxases cleave DNA in a sequence specific manner [75] and it has been suggested that binding and cleavage occur independent of each other [58, 76, 77]. After the relaxase domain nicks the supercoiled DNA, the helicase domain unwinds the nicked DNA. Helicase activity is an energy dependent process [78]. The relaxase of the F plasmid was shown to unwind double stranded DNA at a rate of ~1100 base-pairs/sec [78]. The second conserved tyrosine residue of the relaxase is proposed to play a role in the termination of the reaction by re-ligating the strands together [79]. It has been proposed that when the relaxase reaches the oriT the second tyrosine independently forms a covalent bond with the DNA as a reaction intermediate [79]. Depending on the relaxase family, the re-circularization reaction is
Introduction

either proposed to take place in the donor or the acceptor cell [77, 79, 80]. Since several relaxases can ‘pilot’ the DNA to another cell, relaxases of T4SS have also been called “pilot proteins”.

Coupling proteins

The coupling protein is a membrane bound hexameric ATPase, which is involved in the transfer of the relaxase or the relaxosome complex to the mpf. Coupling proteins are often encoded within the same operon as the relaxase, and phylogenetic analysis also suggests that they have co-evolved [63]. Interactions between either a relaxase or a component of the relaxosome with the coupling protein have been shown previously [81, 82], For example, TraI F is known to interact with TraD F via its C-terminal domain. [83]. Many coupling proteins have an N-terminal transmembrane domain and a C-terminal cytoplasmic region. Certain periplasmic residues were shown to be required for interaction with the T4SS membrane complex [84]. The crystal structure of the cytoplasmic region of the coupling protein TrwB R388 has been solved [85]. It is an hexameric ATPase that binds DNA with a minimum length of 40-45 nucleotides. Remarkably, it is DNA rather than ATP binding or hydrolysis that leads to oligomerization [86]. The ATPase activity was stimulated by the presence of DNA, but no DNA unwinding activity was found associated with the protein. The structure of a TrwB R388 monomer can be divided into two domains, the “nucleotide binding domain” and the “all-alpha domain”. The crystal structure of the TrwB R388 hexamer shows a 20 Å wide channel between the TrwB R388 monomers. Remarkably, six tryptophan residues line the central channel and a mutation of one of these tryptophan residues rendered the protein inactive in conjugation and ATPase activity [86]. Since there is structural similarity between TrwB R388 and the F1-ATPase, three out of six catalytic residues are proposed to be active in TrwB R388 ATP hydrolysis, like the F1-ATPase. Furthermore, the DNA strand is speculated to be analogous to the γ-subunit of F1-ATPase. TrwB would thus act as a motor pumping DNA out of the central channel while cycling between ATP-bound, unbound and empty conformational states [85]. Further details about the energetics of coupling proteins are given below.
**The Mating pair formation (Mpf) complex**

In *A. tumefaciens*, there are several proteins that form the Mpf complex that spans both the inner and the outer membrane. The substrates are putatively transported through this mfp complex. The VirB1 to VirB11 proteins are involved in the assembly of the mfp complex or they form the structural components of this mfp complex. Here, the VirB1 to VirB11 components will be discussed shortly.

VirB1 is a periplasmic protein with muramidase activity that is thought to break peptidoglycan strands in order to facilitate the formation of the Mpf complex [87, 88]. Deletion of VirB1 reduces but does not eliminate translocation [89, 90]. VirB2 is the major pilus subunit that undergoes several processing steps and once mature, the subunits likely form a pool in the inner membrane for polymerization into a pilus that extends out of the cell [91]. The pilus brings about cell to cell contact and is essential for substrate transfer [92, 93]. Assembly and processing of pilin subunits is described in more detail below. VirB3 is an inner membrane protein with two transmembrane domains. It is stabilized by VirB4 [94, 95], VirB7 and VirB8 [96]. There are examples known where VirB3 and VirB4 exist as a VirB3/VirB4 fusion protein [97-99]; these and an engineered fusion protein of VirB3 and VirB4 of *A. tumefaciens* can complement an *A. tumefaciens ΔVirB3/B4* mutant [96]. VirB3 is essential for substrate transfer probably by interacting with other Vir proteins for T-pilus assembly [95]. VirB4 is highly conserved inner membrane localized protein of the T4SS [100]. It contains Walker A and B NTP binding motifs and like other ATPases, it likely assembles into a higher order oligomer (hexamer) [101]. By energizing the Mpf complex, it most likely functions as a motor for pilus assembly and substrate transfer [102, 103].

VirB5 is a minor pilin subunit localized at the tip of the pilus [104]. It is thought to behave as an adhesin [105]. VirB6 is a polytopic inner membrane protein with 5 transmembrane domains and an extended periplasmic loop which is essential for substrate interaction [106]. VirB6 also interacts with other Vir proteins like VirB8 and the VirB7/VirB9/VirB10 complex [107-110]. VirB7 is a small lipoprotein that stabilizes VirB9 by forming a disulphide crosslink with VirB9 [111, 112] and is essential for outer membrane insertion of the VirB7-VirB9-VirB10 complex [113]. However, VirB7 is not well conserved and various forms are found among the different T4SSs [114-117].
VirB8 is a bitopic inner membrane protein with a single transmembrane domain and a periplasmic domain that folds into a dimer [118]. VirB8 forms one of the largest interaction networks with several T4SS proteins including VirB1 [119], VirB4-5 [95], VirB6 [106], Vir8-10 [120, 121] and VirB11 [119] and is also suggested to influence the cell pole localization of the Mpf [122-124]. VirB9 is an outer membrane protein with two conserved domains separated by a non-conserved region [125]. The C-terminus of VirB9 forms a disulphide bond with VirB7 [126] and shares some homology with the secretin superfamily [25]. Both, the conserved C-terminus and N-terminal periplasmic domain of VirB9 are essential for channel assembly and pilus production. VirB10 is a bitopic membrane protein with an essential proline rich region in the periplasm [127]. VirB10 plays an important role during substrate translocation and acts as a bridge for the T4SS’s inner and outer membrane complexes [128]. Together with VirB7 and VirB9, VirB10 forms a core complex. This core complex forms the channel through which the proteins are putatively transported. Remarkably, the core complex of the pKM101 plasmid can be assembled independently of other proteins. VirB11 is a peripheral inner membrane ATPase and oligomerizes into a hexamer. The crystal structure of VirB11 homolog HPO525 of H. pylori [129] indeed showed a hexameric structure with a central pore of 50Å. The structure showed conformational changes between a relaxed and a compact form upon nucleotide binding [130]. Although VirB11 belongs to the family of “traffic ATPases” associated with type II, type III and type IV pili system [131], it is remarkably absent in F plasmid type conjugation systems.

To determine the structure and mechanism of Type IV secretion system, many structural and interaction studies including yeast two hybrid, co-fractionation analysis, localization, detergent extraction and purification, crystallization and mutational studies have been performed. These studies have been very insightful for the understanding of the sub-complexes that work together in T4SS. In recent years, two major breakthroughs have been made. First, the group of Christie determined the pathway of the substrate through the T4SS and secondly, the group of Waksman determined the structure of the T4SS core complex. The determination of the structure of the T4SS core complex is described here, and the TrIP assay that was used to determine the different
proteins and the order in which these proteins are contacted by the substrate during transport is described further below.

Figure 1: Schematic representation of the components of the type IV secretion system (T4SS). Currently known structures of the following components are represented as they are: F-pili/VirB2 [132], VirB5 [133], TraN<sub>pKM101</sub>/VirB7, TraO<sub>pKM101</sub>/VirB9 and TraF<sub>pKM101</sub>/VirB10 core complex [113, 134], periplasmic domain of VirB8 [124], hexameric HP0525<sub>Helicobacter pylori</sub> /VirB11[129], hexameric TrwB<sub>R388</sub>/VirD4 [85] and TraI<sub>IV</sub>/VirD2 bound to ssDNA [67]. IM and OM are abbreviations for inner and outer membrane, respectively.

Structure of the core complex

The answer to the long standing question in the research field of T4SS, “what constitutes the core of the translocation complex that traverses the membrane?” came from a recent cryo-electron microscopy structure of the core complex of the T4SS [113]. TraN<sub>pKM101</sub>, TraO<sub>pKM101</sub> and TraF<sub>pKM101</sub>, the VirB7, VirB9 and VirB10 homologues of the T4SSs encoded on the pKM101 plasmid were over-expressed, purified and isolated from E. coli as a 1 MDa complex. Electron microscopy and single particle analysis revealed that the complex consisted of 14 subunits each of TraN<sub>pKM101</sub>/VirB7, TraO<sub>pKM101</sub>/VirB9 and TraF<sub>pKM101</sub>/VirB10 enclosing a large chamber of 185Å spanning from inner to outer membrane. The structure consists of an O-layer,
facing towards the outer membrane with a 10 Å opening and an I-layer, facing towards the inner membrane with a 55 Å opening. Overall the structure looks like a double walled two cup shaped structure, the mouths of which come together in the periplasm. The entire structure encloses a large cavity.

After the cryo-EM structure, the crystal structure of the O-layer of the complex was solved at a 2.6 Å resolution [134]. The crystal structure revealed several interesting features. The 540 kDa structure showed a hetero-tetradecameric complex consisting of the C-terminal domains of TraO_{pKM101}/VirB9 and TraF_{pKM101}/VirB10 and the entire TraN_{pKM101}/VirB7 protein. Remarkably, TraF_{pKM101}/VirB10 crosses both membranes and α-helices from TraF_{pKM101}/VirB10 span the outer membrane. This was remarkable since it had always been speculated that the TraO_{pKM101}/VirB9 protein would act as the “secretin” of the T4SS. The crystal structure shows a 32 Å opening when viewed from top; this differed from the cryo-EM structure and was suggested to be the “relaxed” state of the confirmation changing structure.

**Energy Components**

T4SSs contain two to three cytoplasmic ATPases which are called VirD4, VirB4 and VirB11 in *A. tumefaciens*. Of these, VirB4 is most conserved and VirB11 is not found in all the T4SSs. Mutations in the Walker A domain in any of the three ATPase leads to a non-transfer phenotype [135-138] demonstrating that they are essential for the translocation process.

As described above, VirD4 is called the coupling protein for its ability to act as a substrate receptor and to “couple” the substrates to the Mpfb complex [82, 139-141]. The coupling protein forms a hexameric structure with an N-terminal transmembrane domain and a C-terminal cytoplasmic domain important for nucleotide binding [137]. The crystal structure of the cytoplasmic region of TrwBR388, the VirD4 homologue from R388 plasmid showed a hexameric structure with a 20 Å large central channel and 3 catalytic sites subunits. TrwBR388 has a DNA dependent ATPase activity suggestive of a “pumping” mechanism [86]. The C-terminus seems important for the interaction with substrate. It interacts directly with the relaxase or with a component of the relaxosome [142]. An arginine rich C-terminus signal sequence found in many substrates defines the interaction specificity with the coupling protein [143]. The N-terminus is involved in
the interaction with the Mpf complex. A VirB10-VirD4 association is believed to be an essential link between the substrate with the core complex [127]. Interestingly, localization studies on the VirD4 homologue from plasmid R27 show accumulation at cell borders [84]. Remarkably, there are currently two examples, e.g. Bartonella spp. [144] and Bordetella pertussis [35] that do not require a coupling protein for T4SS dependent translocation of their respective substrates.

VirB4 is a large (approximately 90 kDa) conserved protein that is essential for pilus assembly and substrate translocation. However, studies on its localization, oligomeric state and ATPase activity show ambiguous results. Initial studies on VirB4 from A. tumefaciens using the PhoA (alkaline phosphatase) and λ cI repressor fusion systems suggested that VirB4 is a dimeric inner membrane protein with two periplasmic domains [100, 145]. For the VirB4 homologues from plasmid RP4 and R388 no ATPase activity could be detected [146]. However, in a later study the VirB4 from plasmid R388 was purified as a monomer in solution but it also formed hexamers and the hexamers showed an increase in ATPase activity in presence of acetate but not chloride ions. Since no effect of phospholipids was observed on its ATPase activity and no transmembrane domains were predicted, it was suggested to be not an integral membrane protein [101]. Interestingly, the VirB4 homologue from plasmid pKM101 could be purified both from the membrane and soluble fraction as dimers and hexamers, respectively. Only the hexameric form was active in hydrolyzing ATP. Remarkably, the VirB4 homologue from plasmid pKM101 could also bind to DNA [103].

VirB11 has homology to the family of “traffic” ATPases found in type II secretion system and in type IV pili systems. Crystal structures from two homologs of VirB11 show a hexameric ring structure with a central cavity [129, 130, 147]. The structure from the H. pylori VirB11 homolog bound to ADP consists of two domains per monomer with the ATP binding site between the two domains of a single monomer [129]. However, in the structure of the Brucella suis VirB11 homolog, the ATP binding site resides between the domains of the consecutive monomers. Most of the VirB11 homologues have a sequence similar to that of B. suis. The structure of the B. suis VirB11 homolog seems to be the result of a “domain swap” made by an elongated linker region not present in proteins that resemble the H. pylori structure [147].
Introduction

Biochemical studies on VirB11 are suggestive of a chaperone resembling activity. Mutants of VirB11 in *A. tumefaciens* have shown uncoupling of pilus biogenesis and substrate selection [148] and mutants in TrbBrp4, the VirB11 homolog encoded of the RP4 plasmid have only a reduced substrate transfer frequency, emphasizing its possible chaperone like activity on Mpf complex formation [149].

The pilus and pilus assembly proteins

T4SSs generally encode a hydrophobic major pilin subunit which is assembled into a pilus structure. They also sometimes encode a minor pilin subunit which is incorporated in the pilus but at lower numbers. The pilin subunits generally undergo several processing and maturation steps before they are assembled into the pilus. Assembly into the pilus requires all of the Mpf proteins [150, 151]. T4SSs are highly versatile and have been studied in different systems but many details on the process of assembly and on the nature of the assembled pilus are still lacking. This is caused by the low sequence homology between the different pilin proteins, and the fact that they are difficult to visualize due to a low abundance and non-constitutive expression. Pilus formation has been studied best for the pilus of the F plasmid, called the F-pili [152] and the pili encoded on the RP4 and Ti plasmid, called P-pili [153, 154]. F-pili are long, flexible filaments used in solid and liquid type mating and P-pili are short, rigid filaments used only during mating on solid surfaces [154]. Other examples also exist, like in *H. pylori*, where a sheathed structure is produced [155] or in *Legionella pneumophila*, where a meshwork of pili are formed [156]. Bacteriophages (e.g. R17) that specifically bind to the F-pilus filament have been used to visualize the pilus structure. Making use of fluorescent bacteriophages, live cell imaging demonstrates F-pili undergoing extension and retraction with mean rates of 39.5 nm/sec and 15.8 nm/sec, respectively. The extension was shown to be proximal with a length of 5 µm made in 148 seconds and the retraction force was shown to draw the donor and recipient cells together [157].

The two best studied classes, the F- and the P type pili, differ in their morphology and in their maturation process. In F pili, the pilin subunit is N-terminally acetylated after signal sequence cleavage. This acetylation is performed by TraXf, an F-plasmid encoded membrane protein [158]. The P-type pilin subunits from the RP4 plasmid and from the *A. tumefaciens* Ti plasmid are circularized after signal sequence cleavage.
Chapter 1

Here, N- and the C-terminal amino acids are joined by a covalent bond. This bond formation is catalyzed by the TraF$_{RP4}$ encoded membrane proteases of the RP4 plasmid and by an unknown chromosomally encoded protein of *A. tumefaciens*, respectively [92, 154].

The topology of the VirB2 pilin subunit of *A. tumefaciens* [103] and the topology of the circularization protease shows that the circularization reaction takes place within the periplasm after pilin insertion into the membrane. A pool of matured pilin subunits is found in the inner membrane ready to be assembled into the pilus. The VirB4 ATPase can release the mature pilin subunit in an ATP dependent manner from the membrane [103]. Based on the structure of the core complex [134], it seems unlikely that the pilus subunits enter the core complex through the periplasm. Thus the VirB4 ATPase might assemble on the cytoplasmic side of the core complex in order to transfer the pilin subunits inside the core complex for the subsequent assembly of the pilus. The exact mechanism of pilus assembly however still needs to be determined [159]. Interestingly, the requirement for assembled pilus for substrate transport by T4SS is not absolute since several mutations have been identified that disrupt pilus assembly but have no effect on secretion by T4SS. These mutations are found in several different T4SS proteins as listed in Table 3.
### Table 3: Examples of T4SS mutants uncoupling pilus assembly and substrate translocation

<table>
<thead>
<tr>
<th>Donor organism</th>
<th>T4SS</th>
<th>protein with the mutation</th>
<th>description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A. tumefaciens</em></td>
<td>Vir</td>
<td>VirB2</td>
<td>G119C mutation of VirB2 selectively blocks substrate transfer (virulence and IncQ transfer) without any effect on pilus production</td>
<td>[160]</td>
</tr>
<tr>
<td><em>A. tumefaciens</em></td>
<td>Vir</td>
<td>VirB6</td>
<td>Insertion at position 140 and 191 led to Pilin- Transfer+ phenotype</td>
<td>[106, 109]</td>
</tr>
<tr>
<td><em>A. tumefaciens</em></td>
<td>Vir</td>
<td>VirB9</td>
<td>Insertion at position 116,236 and 286 led to Pilin-Transfer+ phenotype</td>
<td>[125]</td>
</tr>
<tr>
<td><em>A. tumefaciens</em></td>
<td>Vir</td>
<td>VirB10</td>
<td>Insertion at position 35,40 and 45 of the transmembrane domain led to Pilin-Transfer+ phenotype</td>
<td>[128]</td>
</tr>
<tr>
<td><em>A. tumefaciens</em></td>
<td>Vir</td>
<td>VirB11</td>
<td>several mutations leading to either Pilin+ Transfer- or Pilin- Transfer+ phenotype</td>
<td>[148]</td>
</tr>
<tr>
<td><em>A. tumefaciens</em></td>
<td>Vir</td>
<td>VirB1</td>
<td>Deletion of <em>virB1</em> led to Pilin-Transfer+ phenotype</td>
<td>[89, 151]</td>
</tr>
<tr>
<td><em>E. coli</em> (IncP)</td>
<td>IncP</td>
<td>TrbC</td>
<td>Mutation trbCG114A led to a DNA transfer-deficient phenotype but still allowed production of pilin</td>
<td>[161]</td>
</tr>
<tr>
<td><em>N. gonorrhoeae</em></td>
<td>IncF</td>
<td>TraA</td>
<td>pilus production abolishes DNA secretion</td>
<td>(Thesis Emilia Pachulec-Chapte 2)</td>
</tr>
</tbody>
</table>
Chapter 1

The transport mechanism

The exact mechanism of relaxase transport via the secretion system is not clear. It is not known whether relaxases are transported in a folded or unfolded form or how chaperones or partitioning proteins could play a role in directing/transporting the relaxase to and through the Mpf complex/coupling protein [162, 163]. Nonetheless, it has been shown that some relaxases are transported into the recipient cell [164, 165], more specifically, to the nuclei of the host cell in case of A. tumefaciens [166-169]. The transfer DNA immunoprecipitation (TrIP) assay which is a modification of the chromatin immunoprecipitation assay was used elegantly to determine the translocation pathway of the T4SS substrate. Using this assay, a step-wise and direct interaction of the substrate with VirD4, VirB11, VirB6 and VirB8 and finally with VirB2 and VirB9 was shown [110]. The TrIP assays also revealed that in the initial phase of transport, the three ATPases interact with each other and with VirB10 [102]. The T-DNA first interacts with the VirD4 coupling protein and then with VirB11. VirB4 does not come in contact with T-DNA but is necessary for further transfer of the substrate to VirB6 and VirB8 [110]. The VirD4 coupling protein can function in the initial targeting steps without any energy requirements [102] but probably requires energy to pump out [170] or transfer the DNA to VirB11 [102, 110]. VirB4 is essential for energy dependent pilus assembly [103]. VirD4 and VirB11 mediate conformational changes in VirB10 [110, 127]. VirB10 might then regulate the opening and closing of the outer channel depending on the energy status of the ATPases of the system [171].

It is still under debate whether the substrate passes through the core complex with the pilus protein present or that pilus extension and substrate transfer are two different steps. It has been proposed that extension of the pilus might function as a “piston” to push the substrate out through the core complex. Based on the structure of F pili determined by electron microscopy, it has been proposed that single stranded DNA could pass through the 30 Å cavity of the F-pilus [132] but it is unlikely that a nucleo-protein complex could fit there. Currently there is increasing evidence for the model where pilus biogenesis/assembly and substrate translocation are independent events (see Table 3) [162].
Introduction

The type IV secretion system of Neisseria gonorrhoeae

*N. gonorrhoeae* encodes a 57 kb horizontally acquired island called the gonococcal genetic island (GGI). The GGI consists of three genetic regions transcribed in different directions. The first two regions encode proteins that shares similarity to the T4SS of F plasmid and the third region consists mainly of hypothetical proteins or proteins required for DNA processing. The GGI is present in 78% of gonorrhoeae clinical isolates and at a lower frequency, the island is also found in *N. meningitides* strains [172]. The GGI is integrated into the chromosome at the *dif* site and is flanked by one partial and one complete *dif* site. The *dif* site is a 28 base-pair (aattgcataatgtatatattaat) conserved sequence found at the chromosomal replication terminus among the proteobacteria. It is the binding site for the XerC and XerD sequence specific tyrosine recombinase [173]. When the partial *dif* site flanking the GGI is repaired, the island can be excised from of the chromosome [45]. This excision is mediated by XerD recombinase of the gonococci and the excised GGI can be seen as an extra chromosomal entity [174].

Interestingly, the MS11 strain in which the GGI was first identified was shown to secrete DNA into the medium [45]. Mutations in several genes in the GGI that have homology to T4SS genes strongly reduced the release of DNA. The released DNA is then taken up by natural transformation by other gonococci [45]. The secreted DNA was shown to be single stranded and protected at the 5’ end. Remarkably, DNA secretion assays revealed that an intact T4SS system is required for DNA secretion (Fig. 2) and surprisingly, mutagenesis of the pilus subunit TraA and pilus processing protein TrbI did not affect DNA secretion.

Of the three predicted genetic regions of GGI, the yaf-yaa region encodes the coupling protein TraD<sub>GGI</sub> and the relaxase TraI<sub>GGI</sub> along with two hypothetical proteins Yaa and Yaf. The expression levels of both TraI<sub>GGI</sub> and TraD<sub>GGI</sub> are increased in piliated *N. gonorrhoeae* variants [175]. The TraI<sub>GGI</sub> relaxase belongs to a novel family of relaxases. It is characterized by an N-terminal hydrophobic region and a TraI_2 domain which overlaps with a metal dependent phosphohydrolase domain called the HD domain. The N-terminus catalytic tyrosine is predicted to be at position 93. The C-terminus harbors a conserved domain with unknown function called the DUF1528 domain. Initial
phylogenetic analysis classifies it to the MOB$_H$ family [63] and a further detailed analysis groups the TraI$_{GGI}$ into two. The first group associates TraI$_{GGI}$ with PAGI-2 like genetic islands and the second one with integrases and recombinases (Thesis Emilia Pachulec- Chapter 4).

**Figure 2:** Schematic representation of the genetic map of the gonococcal genetic island (GGI) of *Neisseria gonorrhoeae*. In red are the genes known to be essential for DNA secretion and in blue are the genes not required for the function of DNA secretion [45] (Thesis Emilia Pachulec, Chapter 2).

The coupling protein TraD$_{GGI}$ has a transmembrane domain and conserved Walker motifs. Complete deletion of the $traD_{GGI}$ gene abolished DNA secretion, surprisingly a mutation in the Walker A motif of $traD_{GGI}$ yielded wild type levels of DNA secretion and a C-terminus truncation of $traD_{GGI}$ led to intermediate levels of DNA secretion [175]. Phylogenetic analysis showed that both TraI$_{GGI}$ and TraD$_{GGI}$ are not related to TraI$_{F}$ and TraD$_{F}$, demonstrating that contrary to the proteins forming the mpf complex, the genes involved in substrate processing in GGI are distant from those from the F plasmid.

The $ltgx-ych$ region of the GGI encodes 22 proteins, most of them homologous to T4SS Mpf proteins. The genes required for DNA secretion are $atlA$, $traH$, $traN$, $traF$, $traL$, $traE$, $traK$, $traB$, $dsbC$, $traV$, $traC$, $traW$, $traU$, $trbC$ and $traG$ (Fig. 2). Phylogenetic analysis of the core complex protein TraB$_{GGI}$, the inner membrane protein TraG$_{GGI}$ and the ATPase TraC$_{GGI}$ revealed that they are closely related to the F plasmid T4SS (Thesis Emilia Pachulec, Chapter 4). There are five hypothetical proteins encoded in this region called Yag, Ybe, Ybi, Ycb and Ych. Of these five, only mutation in $yag$ reduced DNA secretion.
Remarkably, the pilus subunit TraA\textsubscript{GGI} and pilus processing protein TrbI\textsubscript{GGI} are dispensable for DNA secretion. TraA\textsubscript{GGI} shares low sequence similarity to the other pilus subunit proteins and TrbI\textsubscript{GGI} is homologous to the pilus processing protein TraF\textsubscript{RP4}. The exp1-parA region encodes for 35 mostly hypothetical proteins. This region contains several DNA processing proteins like the ParA and ParB partitioning proteins, the single stranded DNA binding protein SSB and the topoisomerase TopB. Interestingly, while ParA and ParB are essential for DNA secretion, the remainder of the third operon can be deleted without any effect on DNA secretion.

### 3. Type IV pili systems

Type IV pili have many different functions; they are involved in host cell adhesion [176, 177], twitching motility [178], biofilm formation [179, 180], DNA uptake [17] and other infection/pathogenesis related processes [181]. In \textit{N. gonorrhoeae}, type IV pili are used for all the functions described above [176, 182, 183] and they are essential for pathogenicity [184]. Type IV pili (Tfp) systems can extend and retract Type IV pili. The force generated by the retraction of pilus bundles as measured \textit{in vivo} for \textit{N. gonorrhoeae} has demonstrated that the retraction motor is one of the strongest molecular machines characterized to date [185-187]. Tfp systems consist of large multiprotein complexes of 12-20 proteins with components located in the cytosol, the periplasm and in both the inner and outer membranes. Many proteins of Tfp systems show similarities to Type II secretion systems. Most Tfp machinery genes are named with the \textit{pil} suffix (see \textbf{Fig. 3} for an overview). Many \textit{pil} genes have different names in different systems, and here the nomenclature of the Tfp system of \textit{N. gonorrhoeae} will be used.

#### The mechanism of Type IV pili (Tfp) system

As a first step in assembly of the pilus, the pilin protein encoded by the \textit{pilE} gene [188] is synthesized and inserted in the inner membrane. PilE has an N-terminal class III signal sequence, and a larger C-terminal domain with two conserved cysteines that form a disulphide bond [189]. In \textit{N. gonorrhoeae}, the pilin undergoes, after membrane insertion, several post translational modifications. These include glycosylation, modifications with phosphorylcholine [190, 191], and cleavage of the class III signal sequence at the cytosolic side by PilD, the prepilin peptidase [192]. PilE further
undergoes phase and antigenic variation, a phenomenon used to evade the human immune system response [193]. After cleavage of the signal sequence, PilD also methylates the free N-terminus. The processed pilus subunit is then extruded from the inner membrane by the PilF ATPase and assembled on PilG, a polytopic inner membrane protein [194-197]. A second ATPase, PilT behaves antagonistically and drives pilus retraction [198]. Both PilF and PilT are inner membrane proteins and belong to AAA+ family of hexameric ATPases [199]. PilT has several conserved regions that include the Walker A and B motifs, Asp and His boxes and an AIRNLIRE sequence [200], that are essential for function [201, 202]. The crystal structures of the nucleotide bound and free forms of the PilT ATPase of *P. aeruginosa* showed a hexamer with an inner pore of size 40Å. The monomers contain N- and C-terminal domains, and the nucleotide binding site is located between the two domains.

The PilM, PilN, PilO and PilP proteins, encoded in a conserved gene cluster, form a complex and function in pili biogenesis and assembly [204, 207]. At the outer membrane, PilQ forms a homooligomeric ring through which the pilus can extend into the medium [208-210]. The adhesin PilC [211, 212], bound at the tip of the pilus is involved in adhesion to human cells. Several other pil proteins have specialized functions like PilP [213, 214] and PilW [206], which are involved in proper expression, assembly and localization of the secretin PilQ. If the PilF ATPase is replaced by PilT, the extension of the pilus is reversed and the pilus is retracted [215]. Extension and retraction of the Tfp can be used for movement, and this phenomenon is called twitching motility [178]. In *N. gonorrhoeae*, retraction of the pilus is also necessary for colonization of the host cells at various levels [216]. The pilus fiber extension, attachment and retraction cycles drives the formation of a gonococcal microcolony. The Tfp retraction by these microcolonies helps the formation of large gonococcal motile colony structures on to the human epithelial cells [217]. A structure of the pilus was obtained by fitting the 2.3 Å resolution crystal structure of a PilE monomer into the 12.5 Å resolution Cryo-EM structure of the assembled pilus. The pilus is formed by a 3-start twisted helix with approximately 3.6 PilE subunits per turn. The outer diameter is approximately 60 Å, and the pilus contains a narrow central channel with a diameter between 6–11 Å. The N-terminal α-helix forms the core of the filament and the sites of post-translational modifications along with hyper variable region extend out from the
surface. The surface is covered with positive charges that might explain the DNA binding properties of the pilus [203]. The mechanism of pilus assembly and disassembly is currently unclear. Next to PilF, several other proteins have been proposed to be involved in the retraction mechanism of pili. PilC is an antagonist of PilT, and thus promotes retraction. The increase in transcription level of PilC during adhesion to human cells provides direct evidence for its regulatory role in retraction [218]. Furthermore, in vivo assays performed in a pilT mutant of N. meningitidis suggested that PilG, H, I, J, K and W could work towards counter-retraction [219].

Figure 3: Schematic representation of the components of the type IV pili (Tfp) system. IM and OM are abbreviations for inner and outer membrane, respectively. Structures of the following components are represented as they are: Pilus fibre/PilE [203], PilG [197], periplasmic domain of PilO dimer [204], hexameric PilT [205] and PilW [206].

Competence and Tfp

Tfp systems are also often involved in the uptake of DNA. In N. gonorrhoeae, expression of PilE, but not necessarily pilus fibre expression is required for efficient DNA uptake. Mutagenesis of most proteins involved in type IV pili assembly or retraction abolishes the uptake of DNA. Tfp are thought to be involved in the transport of DNA across the outer membrane. Transport across the inner membrane is driven by competence or Com proteins. These competence related proteins are ComP; the
Since the structure of the pilus has demonstrated that there is no significantly large central channel; it is very likely that the transport across the outer membrane occurs after binding of DNA to the pilus. The following step of pilus retraction driven by the PilT ATPase would then pull the DNA across the outer membrane. Competence pseudopilins, which use the Tfp machinery for assembly, have also been proposed to be involved in this initial step [224]. In the second step, the competence proteins are shown to transport DNA across the inner membrane [21, 225]. N. gonorrhoeae selectively takes up DNA that has a DUS sequence which is highly prevalent in the genome sequence of Neisseria. The 12-mer DUS sequence (5’-ATGCCGTCTGAA-3’) was shown to increase DNA binding and the transformation efficiency [20, 226]. The molecular basis of DUS dependent DNA binding is however poorly understood. Although it is apparent from biochemical studies that DNA binding and uptake are separate events [220]; the search for the DUS receptor for DNA recognition/ binding is still ongoing.

Since the main focus of Chapter 5 of this thesis is the structure of the outer membrane components of the Tfp system, this component of Tfp systems will be discussed below in more detail.

**The secretin complex**

Secretins belong to a large super family of proteins present in many Gram negative bacteria and are found associated with type II, type III, type IV pili systems and with filamentous phages. Secretins are integral outer membrane proteins that multimerize and form gated ion channel [227-229]. Secretins are unique in that they form a very stable heat and SDS resistant multimers. The channel diameter varies among different members probably in accordance with the substrate dimensions (Table 4).

Secretins contain N and C-terminal domains, where the C-terminal domain is highly conserved and involved in the multimer formation. The N-terminus could play a role in defining specificity to the protein [210]. This was shown for the OutD secretin of E.
Introduction

*chrysanthemi* [230] and the phage secretin pIV [231], but this was not the case for PulD secretin of *Klebsiella oxytoca* [232].

Different secretins have been studied in detail. The *K. oxytoca* secretin PulD purified in a 1:1 ratio in complex with the lipoprotein PulS. The purified structure had a 12-fold symmetry. The lipoprotein binds to the c-terminal end and formed radial spokes around the secretin giving outer diameter of 25.8 nm [227, 233]. The purified PulD-PulS complex demonstrated channel activity, however no difference was observed in conductance in the presence of the substrate pullulanase. Unlike most outer membrane proteins, the outer membrane insertion property of PulD was Omp-85 independent [197] and PulD was shown to assemble properly in an *E.coli in vitro* environment which was not the case for PilQ, the secretin from *N. meningitidis* [234]. PilQ exhibited 12-mer C4 quasi symmetry. The 3D reconstructed structure showed a funnel shape with cap, arms, ring and plug regions [209]. The cryo-electron microscopy of phage pIV secretin showed 14-mer symmetry. The structure represents a tail-to-tail dimer formation of the homo-multimer which could be divided into an N-ring, an M-ring and a C-ring with the latter facing the outer side of the membrane. Large conformational changes probably accommodate the release of the phage [235]. The recent dodecameric cryo-EM structure of the purified GspD secretin of *Vibrio cholera* was interestingly obtained in its “closed” state with an extracellular cap region having a diameter of 10Å. The structure has a prominent periplasmic vestibule with a conserved constricted region [236].

Lipoproteins are commonly associated with secretins and it was proposed that their function is to aid the localization and assembly of the secretins. The crystal structure of the MxiM lipoprotein in association with a peptide from its cognate MxiD secretin demonstrated a "cracked β-barrel" structure [237]. The PilP lipoprotein from *N. meningitidis* adopts a similar structure and was proposed to interact with PilQ. PilP however does not affect the assembly of PilQ [213, 214]. PilW from the Tfp system of *N. meningitidis* is the only protein shown to affect the multimerization of a Neisserial PilQ [206]. Interestingly, in the absence of PulS, PulD assembles in the inner membrane suggesting the lipoprotein influencing the localization but not assembly of the secretin in this case [238].
Table 4: Secretins from different families with their substrate

<table>
<thead>
<tr>
<th>Secretin</th>
<th>symmetry</th>
<th>Outer diameter/ Channel size</th>
<th>Substrate</th>
<th>substrate size</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>N. meningitidis</em> PilQ</td>
<td>12-fold</td>
<td>~16.5nm/ 6.5 nm</td>
<td>Tfp fibre</td>
<td>60 Å</td>
<td>[203, 208]</td>
</tr>
<tr>
<td>fl phage pIV</td>
<td>14-fold</td>
<td>~14nm/ 7-8 nm</td>
<td>Phage</td>
<td>60-70 Å</td>
<td>[239]</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> PilQ</td>
<td>unknown</td>
<td>183nm/ 5.3 nm</td>
<td>Tfp fibre</td>
<td>~52 Å</td>
<td>[240, 241]</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> XcpQ</td>
<td>unknown</td>
<td>198nm/ 9.5 nm</td>
<td>Elastase</td>
<td>60 Å</td>
<td>[240, 242]</td>
</tr>
<tr>
<td><em>K. oxytoca</em> PulD</td>
<td>12-fold</td>
<td>11.6nm/ 7.6 nm</td>
<td>Pullulanase</td>
<td>~70 Å</td>
<td>[227, 243]</td>
</tr>
<tr>
<td><em>Y. enterocolitica</em> YscC</td>
<td>13-fold</td>
<td>14nm/ 4.5 nm</td>
<td>Type III needle</td>
<td>60-70 Å</td>
<td>[244]</td>
</tr>
<tr>
<td><em>V. cholera</em> GspD</td>
<td>12-fold</td>
<td>15.5nm/ 10 Å (exterior opening), 5.5 nm (periplasmic constriction), 7.5 nm (periplasmic opening)</td>
<td>cholera toxin</td>
<td>~65 Å</td>
<td>[236]</td>
</tr>
</tbody>
</table>

Extension and retraction of the pilin occurs with a high rate (350 ± 50 nm/s). Since the extending and retracting pilin most likely passes through the secretin, large forces are exerted on the secretin complex. This might explain the observed stability of the secretin complexes. Whether and how the secretin can withstand these forces is currently unknown. Possibly the extending spikes observed in e.g PulD might aid in
Introduction

anchoring the secretin complex to the membrane [206, 209]. Remarkably, ExeA of Aeromonas hydrophila [245] and FimV of Pseudomonas aeruginosa [246] have recently been implicated to function in secretin assembly. Both proteins have peptidoglycan binding domains that might be involved in anchoring the complex.

4. Comparison: Tfp and T4SS

Both T4SSs and the Tfp systems are functionally diverse and can play important roles in pathogenesis. T4SS can extend and sometimes retract pili, and utilize a one step secretion mechanism to deliver/release macromolecules [247]. Tfp system also extend pili [189] and have homology to Type II secretion systems which are involved in substrate transport. Although the systems are not homologous they show many similarities. Pilus retraction occurs in Tfp system [178, 186] and in some T4SS [157]. Remarkably, there are examples where pilus formation is dispensable for substrate translocation in T4SS [160] and for DNA uptake in Tfp systems [248]. The ATPases from both systems function as hexamers [247, 249, 250], and the homologs of N. gonorrhoeae PilT (Tfp) and the homologs of A. tumefaciens VirB11 (T4SS) belong to the same family of secretion ATPases.

At the outer membrane the two systems harbor a multi-protein complex. This complex consists of homologs of TraN\textsubscript{pKM101}/VirB7, TraO\textsubscript{pKM101}/VirB9 and TraF\textsubscript{pKM101}/VirB10 for the T4SSs and homologs of the PilQ secretin for the Tfp system. The PilQ secretin is surrounded by a second ring and extending spikes, formed by currently unknown proteins. Viewed from the top, the crystal structure of the purified T4SS sub-complex [134] showed an outer diameter of 76 Å and a central channel of 32 Å in the crystal structure and of 10 Å in the cryo_EM structure. EM analysis of the PilQ secretin showed a diameter of PilQ of 150Å and a central channel of 60 Å. A difference between T4SSs and the Tfp systems is that the core complex of T4SSs spans both the inner and outer membrane, while there is no evidence that the complex formed for Tfp also spans both membranes. A 14-fold symmetry is observed for the core complex of the T4SS. The PilQ secretin was previously proposed to have a 12 fold symmetry but the ring structure surrounding PilQ has a 14 fold symmetry. In both systems correct assembly of the complex depends on a lipoprotein [113, 206]. It is however not known whether similar to T4SS, the lipoprotein involved in secretin assembly also becomes part of the
outer membrane complex. Both systems require a gated aperture for opening and closing the membrane channel (Fig. 4).

**Figure 4:** Schematic representation of the model of the type IV secretion system (T4SS) and the type IV pili (Tfp) system. Arrow heads indicate the direction of the substrate transfer for the T4SS and the extension and retraction of the pilus for the Tfp system. IM and OM are abbreviations for inner membrane and outer membrane, respectively.

It is proposed for T4SS that opening and closing of the channel depends on an interaction between VirB10 and the ATPases of the system [251]. For type II secretion systems, it was proposed that the growing pseudopilus of the system acts as a piston to push the substrate, thus inducing a conformational change and opening of the secretin complex. This would allow for the opening of the gate for substrate secretion [236]. For Tfp system, direct biochemical and structural evidence is lacking but it has been hypothesized that PilT indirectly interacts with the pilus [250] via the inner membrane complex of PilM, PilN, PilO and PilP, and that the outer and inner membrane complexes could work together synchronously via a PilQ-PilP interaction [207, 214]. PilC could also act as the mediator since it is known to interact with PilQ and its regulatory role in pilus retraction is well established [218]. Thus, although they show little homology, there are many similarities between the Type IV secretion systems and the Type IV pili systems.
Scope of thesis

Scope of the thesis

In this thesis, a study of components of the type IV secretion and the type IV pili systems of the human pathogen *N. gonorrhoeae* is described.

Chapter 1 provides a general introduction to the thesis. Type IV secretion systems are described in general and our current knowledge of the different proteins constituting the system and forming complexes is described in detail. Our insight in the mechanism of transport and secretion is also discussed, including the recent progress in studies on the type IV secretion system encoded within the gonococcal genetic island (GGI). Finally an introduction to the type IV pili system is given with emphasis on the secretin complex of the system with a description of the other components of the system. A comparative analysis of the two systems is shortly presented.

In Chapter 2, the mechanism of processing and membrane insertion of the pilin subunit TraA<sub>GGI</sub> was studied. The role of two inner membrane proteases, the leader peptidase LepB and GGI encoded protease TrbI<sub>GGI</sub>, in the circularization mechanism of the pilin subunit were investigated. Mutational analysis revealed several key residues involved in the circularization reaction of pilus processing. Formation of a stable pilus-protease intermediate was observed during *in vitro* studies on the circularization reaction.

Chapter 3 describes the study of the novel relaxase protein TraI<sub>GGI</sub>. The relaxase TraI<sub>GGI</sub> was found to be associated with the membrane fraction. Possible functionally conserved residues including catalytic tyrosine, phosphohydrolase histidine and aspartate and metal coordinating histidine mutations were analyzed *in vivo* for DNA secretion assay. The secreted DNA was determined to be single stranded and an oriT region was proposed for the GGI. The TraI<sub>GGI</sub> possess N-terminus hydrophobic region with putative signal sequence which was studied in an *in vitro* transport assay.

In the appendix to Chapter 3, the localization of TraI<sub>GGI</sub> was studied further in the membrane and the secreted fraction of *N. gonorrhoeae*. The outer membrane derived vesicles were especially analyzed in the different mutant strains with abolished DNA secretion.
Chapter 4 describes the study on the single stranded DNA binding protein (SSB) encoded by the GGI of *N. gonorrhoeae*. In order to determine the function of SSB pertaining to the DNA secretion mechanism of GGI, various *in vivo* assays were performed in *N. gonorrhoeae*. The changes in DNA release and uptake efficiencies were determined with over expressed levels of SSB. The oligomeric state, DNA binding behavior, binding mode of the purified protein was characterized. The purified SSB was also analyzed for its effect on topoisomerase DNA relaxation activity.

In Chapter 5, the membranes encompassing the secretin complex of the type IV pili system of Neisseria were analyzed using electron microscopy. The membrane embedded complex in its native state revealed novel structure at the outer membrane previously not identified. The differences between the complex from *N. gonorrhoeae* and *N. meningitidis* membranes were studied. Concurrently, several mutants of the type IV pili system were generated in *N. gonorrhoeae* in order to study and identify the extra components of the secretin complex.

In Chapter 6, a brief summary of the work described in this thesis is presented.