Study of the natural transformation pilus in streptococcus pneumoniae

Raphael Laurenceau

Université Pierre et Marie Curie

Complexité Du Vivant

Institut Pasteur, G5 Biologie Structurale de la Sécrétion Bactérienne

Study of the natural transformation pilus in *Streptococcus pneumoniae*

Par Raphaël Laurenceau

Thèse de doctorat de Microbiologie

Dirigée par Rémi Fronzes

Présentée et soutenue publiquement le 18 septembre 2014

Devant un jury composé de :
Guennadi Sezonov, professeur des universités, UMPC, Paris, président du jury
Eric Cascales, directeur de recherche, CNRS, Marseille, rapporteur
Vladimir Pelicic, Principal investigator, Imperial College, Londres, rapporteur
Olivera Francetic, chargée de recherche, Institut Pasteur, Paris, examinatrice
Xavier Charpentier, chargé de recherche, Chargé de recherche, INSERM, Lyon, examinateur

As for every other achievements in my life,

To Norman,

Whom I love,

And who makes it all possible.

Acknowledgments

First among all the great people I had the chance to meet during my PhD, I want to give warm thanks to Rémi Fronzes. Not only for being a tremendous PhD director, but also for being the nicest person I have worked with. He is enthusiastic, dedicated, passionate, and deeply human in his contact with others. He showed me that research is not necessarily linked to stress, pressure, competition and yield, but much more to collaboration, mutual help and understanding.

I want to thank Annick Dujeancourt, who started the same day as I this great adventure in the bacterial secretion lab. She has been from the first day a great colleague, a great help, and a great friend.

Many thanks to Natalia de Val Alda, Francesca Gubellini, Chiara Rapisarda, Petya Violinova Krasteva and Amy Diallo, all brilliant researchers that have been very generous in sharing their experience with me. All dear friends as well.

Special thanks to Sahra Ouarti, a devoted, talented and intelligent student I had the chance to supervise.

I thank my friend and collaborator Gerard Pehau-Arnaudet, for guiding my first steps and mentoring me in the art of Electron microscopy.

I thank the whole Jean-Pierre Claverys and Patrice Polard lab, especially Nathalie Campo. They have been invaluable collaborators throughout the whole of my PhD by teaching me and sharing their expertise on the genetics of *S. pneumoniae*.

I thank all the other collaborators to this work: Lena Dewenter and Berenike Maier; Nadia Izadi and Idir Malki; Joseph Gault, Christian Malosse and Julia Chamot-Rooke; Sonia Baconnais and Eric Lecam; Michel Debarbouille and Tarek Msadek; Jean-Yves Tivenez; Adeline Mallet; and many others. They have been, whether the outcome of each project was a success or not, people that I really enjoyed to work with.

I thank Bruno Dupuy, my tutor, for his kind and insightful guidance.

I thank Luc Talini for hiring me as a consultant in the company Quattrocento. Thank you for giving me this terrific opportunity, and thank you for your valuable advice.

Chiara, Nora, Amelie, Amy, Clémence, Quentin, thank you for being faithful beer buddies, so vital to get through the ups and downs of a research project.

Finally I want to thank my family and friends that I love dearly, and who pushed me forward all along these 3 years.

Contents

Figure index

Abbreviations

°C: celsius degree

3D: three dimension

AUC: analytical ultracentrifugation

BSA: bovine serum albumine

CSP: compétence stimulating peptide

Da: Dalton

DNA: deoxyribonucleic acid

dsDNA: double-stranded DNA

EMSA: electrophoretic mobility shift assay

FRET: fluorescence resonance energy transfer

GFP: green fluorescent protein

HEPES: N-(2-hydrocyethyl)piperazine-N'-2-ethanesulfonic acid

HGT: horizontal gene transfer

HSQC: heteronuclear single quantum coherence spectroscopy

LB: luria broth

M: molar (mol/L)

M: molarity

NHS: N-hydroxysuccinimide

NMR: nuclear magnetic resonance

O.D.: optical density

PBS: phosphate buffered saline

PDB: Protein Data Bank

RPM: rotation per minute

SDS-PAGE: sodium-dodecyl-sulfate polyacrylamide gel electrophoresis

ssDNA: single-stranded DNA

T2SS: type II secretion

T4P: type IV pilus

TB: terrific broth

TBS: tris buffered saline

TEM: transmission electron microscopy

TIRF: total internal reflection fluorescence

WT: wild type

Introduction

I. A historical perspective: the discovery of DNA

Natural transformation in *Streptococcus pneumoniae* is a particularly exciting topic. Its discovery is considered as the starting point of one of the most important step forward in life sciences: the fact that a deoxyribonucleic acid polymer (DNA) is the molecule of heredity, carrying genetic information in all living organisms.

In 1928, Frederic Griffith showed that *S. pneumoniae* was able to « transform » itself from one type of colony, rough and avirulent, to another, smooth and virulent, by mixing a lysed mixture of the second type to cultures of the first type (Griffith 1928). This discovery led Oswald Avery, Colin MacLeod and Maclyn McCarty to isolate and identify the « transforming principle » behind this phenomenon: DNA (Avery et al., 1944). Their results were published in 1944 after a long debate over the nature of the biomolecule carrying the genetic information, which was widely believed to be proteins at that time. Indeed, how could such a simple and almost inert molecule as DNA contain information?

The answer came in 1953, when James Watson and Francis Crick discovered the double-helix structure of DNA. They showed how the specific base-pairing enables the perfect copying, while the order of bases along the polymer forms the blueprint for the sequence of amino acids in proteins (Watson et al., 1953).

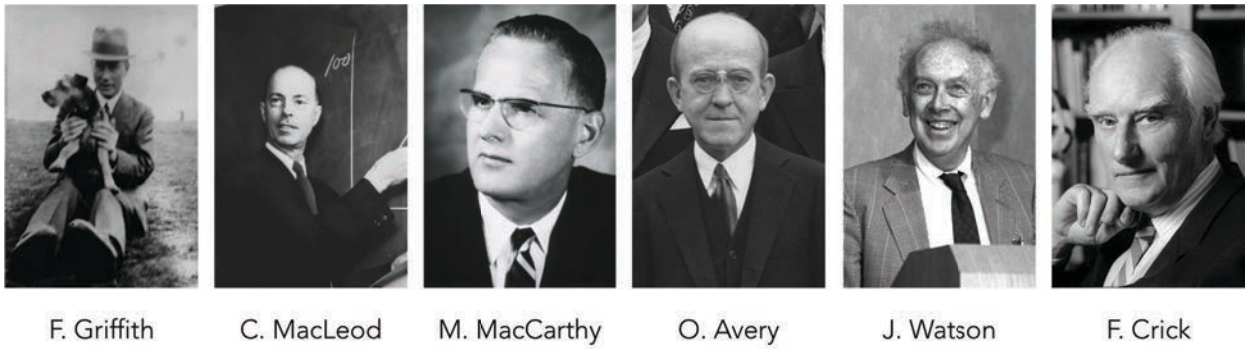


Figure 1 - The scientists who discovered and unveiled the mechanism of the molecule of heredity

Frederic Griffith showed how to transform one pneumococcal type into another in 1928; Colin MacLeod, Maclyn McCarthy and Oswald Avery showed that DNA is the support of heredity information in the Griffith experiment in 1944; James Watson and Francis Crick discovered the structure and mechanism of the DNA molecule in 1953;

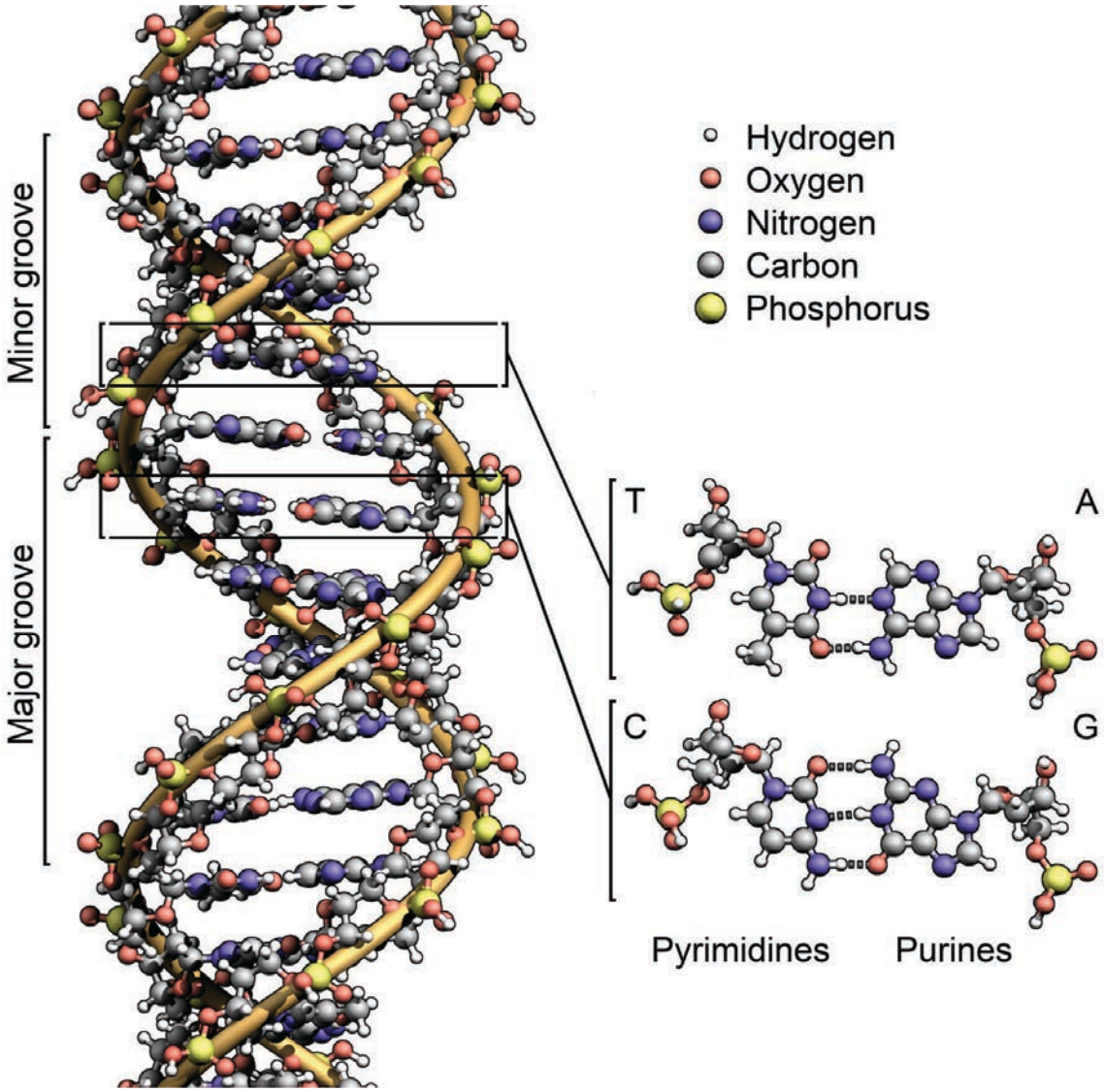


Figure 2 - Double-helix structure of the DNA molecule

Atoms in the structure are color coded. The interaction of nucleotides A-T and C-G is detailed on the bottom right.

After this milestone discovery, almost two decades were necessary to decode entirely the information encrypted in a DNA sequence. In 1961 Francis Crick et al. demonstrated that a triplet of nucleotide correspond to one amino (Crick et al., 1961). The same year Marshall Nirenberg showed that a poly-uracil RNA molecule can be translated using a cell-free system into a poly-phenylalanine peptide, thus demonstrating that the triplet TTT codes for the amino acid Phenylalanine. Using different combination of RNA bases the whole genetic code could be solved during the following years. Meanwhile, the understanding of the DNA language at another level came from the description of the *lac* operon by Jacques Monod and François Jacob. How could *Escherchia coli* regulate the production of a lactose digestion enzyme only when lactose is present? They showed the existence of a protein binding to the genomic DNA at the site where the enzyme gene is located. They called it a 'repressor', since it prevented the gene transcription, and therefore the protein production. They showed that lactose could also bind competitively to this repressor, consequently lifting the repression when it was present. This feedback loop constitutes the first description of a genetic regulation mechanism, which revealed to be ubiquitous in bacteria (Jacob et al., 1961).

The last chapter to date of this story occurred more recently, through a work published in 2007 by scientists from the J. Craig Venter Institute: by inserting whole genomic DNA, naked, of *M. mycoides* species into a *M. capricolum* cell, they changed one species into another (Lartigue et al., 2007). The newly created *M. mycoides* cells were phenotypically identical to their donor strain. More than « transforming » one type of pneumococcal cell into another, DNA was able to « transform » a species into a different one. Going even further, they synthesized an entirely new chromosome based on *M. mycoides* genome, that they transplanted in *M. capricolum* cells, to create a new species (Gibson et al., 2010). After a few generations, all the proteins of the new cells were originating from the synthetic chromosome, definitely proving that DNA contains the software of life, while cells are the hardware.

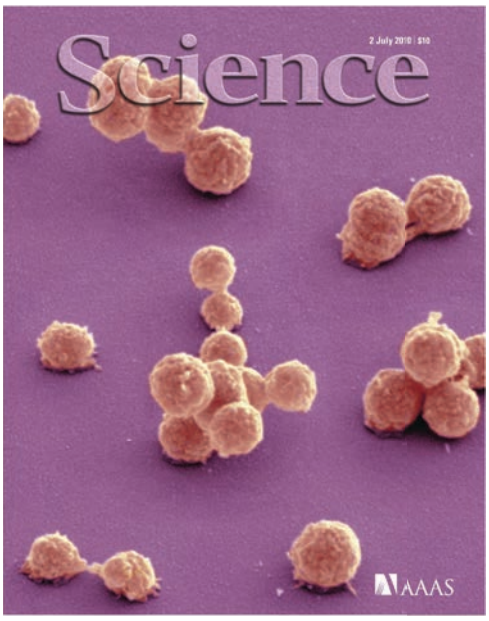


Figure 3 - Cover of the Science journal showing M. mycoides JCVI-syn1.0 cells

Electron micrograph (magnification ~25,000×) showing the first cells that are controlled by a chemically synthesized genome.

II. Horizontal gene transfer in bacteria

Creation of genetic variability is commonly considered as it happens in eukaryotes, by combining of two distinct parental genomes into a completely new offspring genome, a process named vertical gene transfer. Appart from the accumulation of punctual mutations, prokaryotes are only able generate clonal populations of themselves and cannot create this sort of genetic variability. Instead they use a process called horizontal gene transfer (HGT) in which pieces of genetic information are exchanged directly from cell to cell. There are several ways for these transfers to occur, all completely independent with distinct origins and distinct mechanisms.

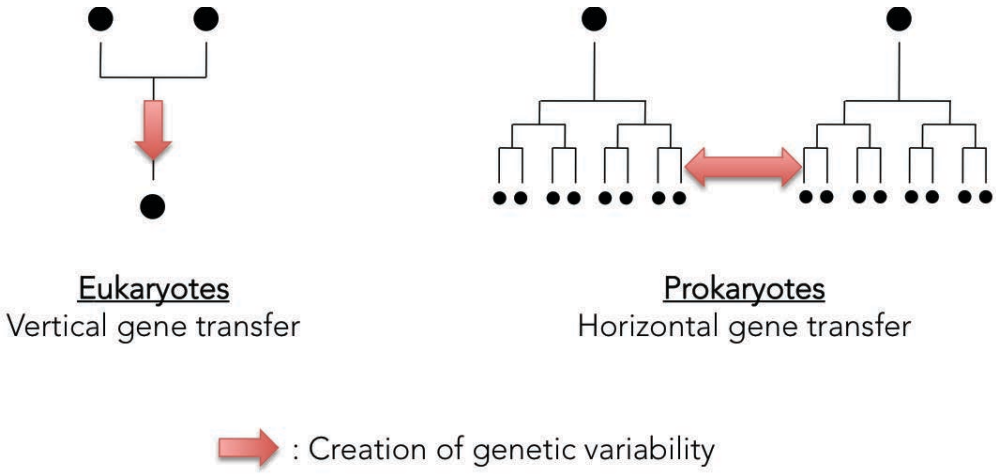


Figure 4 - Vertical versus horizontal gene transfer
The flux of creation of genetic variability (red arrow) goes along with the offsprings in vertical gene transfer, while it goes from one « branch » of a phylogenetic tree to the other in horizontal gene transfer.

A. Transduction

Transduction is the exchange of DNA mediated by a bacterial virus. A virion particle can sometimes be accidentally loaded with a piece of the host genome while its own genome is being packaged, and transfer this piece of bacterial DNA to the cell it will infect (Figure 5). As viruses have generally a narrow host range, the resulting HGT is mostly intra-species. It is likely the most ancient of all form of gene transfer, since viruses already existed in the RNA world, prior the emergence of DNA (Forterre et al., 2005).

B. Natural transformation

Natural transformation is the capture, incorporation, and chromosomal integration of exogenous DNA by a bacterial cell, thanks to a dedicated protein machinery (Figure 5). Among the different HGT mechanisms, it is the only one to be mediated by the recipient bacterium, thanks to proteins all present in its the core genome (i.e. found in every individual cell of the species).

So far, it has been described in approximately 80 species, but new ones are being discovered continuously (Morikawa et al., 2012; Johnston et al., 2014). Even distantly related species share similar DNA uptake and processing systems. It becomes clear that the main roles of such systems are both creation of genetic variability (Johnston et al., 2013), and chromosome repair (Stevens et al., 2011).

C. Conjugation

Conjugation is the transfer of a plasmid or genomic DNA from a donor cell to a receptor cell through a direct cell to cell contact mediated by a Type IV secretion system (Figure 5). Long continuous DNA fragment can be transferred. Some conjugative plasmids have a broad host range, spanning different bacterial clades. Even if they can provide increased fitness to their host in some cases (i.e. antibiotic resistance), conjugative elements can be seen as selfish genetic material, spreading and parasitizing their hosts (Smillie et al., 2012; Cascales et al., 2003).

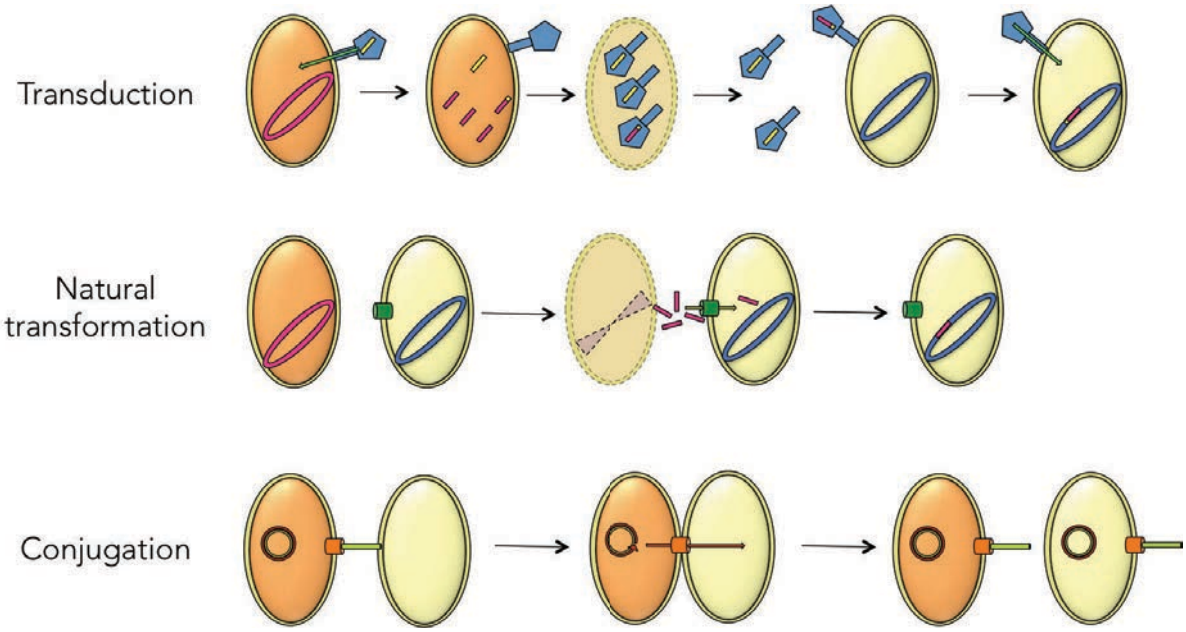


Figure 5 - The three canonical HGT mechanisms

In each kind of genetic transfer, the donor cell is orange and the recipient cell is yellow. Viruses are shown in blue transduction. Chromosomal DNA is shown by an oval in transduction and transformation. Plasmid DNA is shown as a circle in conjugation.

D. Other means

Other gene transfer means have been described, although they are not well documented yet. Among them are gene transfer agents, which are phage-like particles, with the major difference that they package random pieces of DNA, instead of their own coding sequence (Lang et al., 2012). HGT has also been shown to occur by the exchange of membrane vesicles containing a DNA fragment (Mashburn-Warren et al., 2006), or through membrane nanotubes connecting two cells together (Dubey et al., 2011).

E. Public health repercussions, or the rise of superbugs

The intensive traffic of genetic material provides a great genomic plasticity to bacteria, and makes them difficult target for any treatment. Of particular gravity, the rapid spread of antibiotic resistance genes is considered today as a major threat for public health worldwide (Cars et al.,

2008; Rask-Andersen et al., 2010; Laxminarayan et al., 2011). The massive use of antibiotics in healthcare settings or in livestock, the lack of new antibacterial agents, and the fact that sublethal dose of antibiotics are sufficient to select and maintain resistances (Gullberg et al., 2011) are quickly generating « superbugs », bacteria resistant to every available treatments. The 2014 report on antimicrobial resistance from the World Health Organization warns that « a post-antibiotic era—in which common infections and minor injuries can kill—far from being an apocalyptic fantasy, is instead a very real possibility for the 21st century ». Drug resistance has often been compared to global warming: a worldwide multi-causes problem, impossible to tackle without a concerted action.

F. Limits of Horizontal Gene Transfer

It is important to highlight that, although the consequences of HGT are huge in terms of public health, HGT events are not as simple as a free sharing of genetic material among bacteria. Many barriers exist, greatly limiting this kind of transfer (Popa et al., 2011). For example, the difference in genomic GC content from donor to recipient is most of the time inferior to 5 % for a successful foreign gene acquisition to be maintained. Genes travel more easily between similar bacteria, in which they have more chances to be functional, and avoid being discarded. Lastly, bacteria have to share a same ecological niche for HGT to occur.

III. Natural transformation in *S. pneumoniae*

A. A human pathogen

S. pneumoniae, also called the pneumococcus, is a Gram-positive diplococcus bacterium living exclusively in the human nasopharinx. Considered as a commensal bacterium, it can nonetheless become virulent in susceptible individuals such as young children, elderly or immunocompromised people. It can cause severe invasive diseases such as meningitis and bacteremia, as well as many mucosal diseases including pneumonia, sinusitis, and otitis.

S. pneumoniae is closely related to other commensal streptococci, *Streptococcus infantis*, *Streptococcus oralis*, and especially *Streptococcus mitis*. It was proposed that *S. pneumoniae* and *S. mitis* share a common ancestor that was a human pathogen some thousand years ago. While *S. mitis* evolved into a truly commensal bacterium by loss of virulence genes and masking from the immune system, *S. pneumoniae* has kept numerous virulence genes, still induces immunity, and is more similar to this pathogenic ancestor (Kilian et al., 2008; Donati et al., 2010). The recent expansion of *S. pneumoniae* lineage could be accounted for the recent skyrocketing human population on earth. Being a deadly pathogen for its unique host in a small and sparse population is conceivably a strong bottleneck for expansion (around 1000 human individuals estimated to live 100,000 years ago) (Cavalli-Sforza et al., 2003), which is not the case in a dense population, constantly mixing at the global scale.

Despite medical advances and vaccination campaigns *S. pneumoniae* remains a leading mortality cause worldwide (Bogaert et al., 2004; Walker et al., 2013). According to the 2008 Bulletin of the World Health Organization, the childhood pneumonia, caused mainly (but not exclusively) by *S. pneumoniae*, is responsible for about 19 % of all deaths in children aged less than 5 years, of which more than 70 % take place in sub-Saharan Africa and south-east Asia (Rudan et al., 2013).

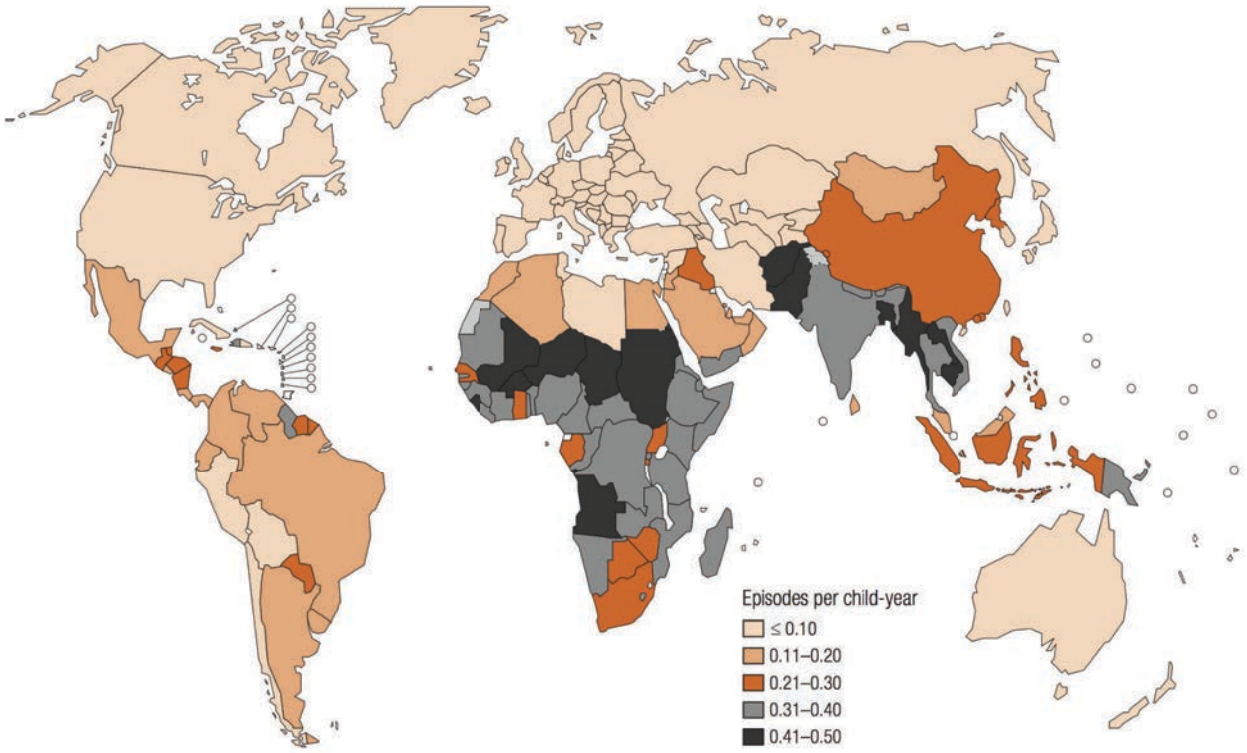


Figure 6 - Incidence of childhood clinical pneumonia at the country level

Taken from Rudan et al., 2013.

S. pneumoniae infection mechanism relies mostly on its capsule, a polysaccharide shell exposed at its surface. Vaccines have to cover the whole range of virulent capsule types, called serotypes. However, the bacteria constantly shuffle their capsule gene locus by HGT, causing vaccine escape, and rise of new virulent serotypes that were not covered in the vaccine (Croucher et al., 2011).

B. A model organism

S. pneumoniae has always been used as a model to study natural transformation. Surely because of its early discovery by F. Griffith, but also because *S. pneumoniae* has pushed this HGT mechanism to its maximum efficiency (Johnston et al., 2014). Devoid of a SOS response pathway, this bacterium seems to have developed natural transformation as way to react from

stressful conditions (Prudhomme et al., 2006). Recent sequencing studies of clinical isolates have shown high level of recombination events, including transfer of pathogenic islands and capsule switching. (Croucher et al., 2012; Hu et al., 2012). As an example, a child clinical isolate was shown to have had 23 recombination events and substituted 7.8 % of its genome in a seven months period (Hiller et al., 2010).

Overall these studies make clear that *S. pneumoniae* species genome can be seen as a large gene pool, from which individual pneumococcal cells regularly pick up and recombine pieces of their chromosome by DNA uptake. This gene pool exceeds by twice the size of any pneumococcal chromosome (Donati et al., 2010).

C. The competence state

Natural transformation in *S. pneumoniae* is a strictly regulated event. It occurs during a transitory state of the bacterium's life cycle – competence – and requires the timed expression of a dedicated set of genes (Claverys et al., 2009). Here is a comprehensive figure of the whole process:

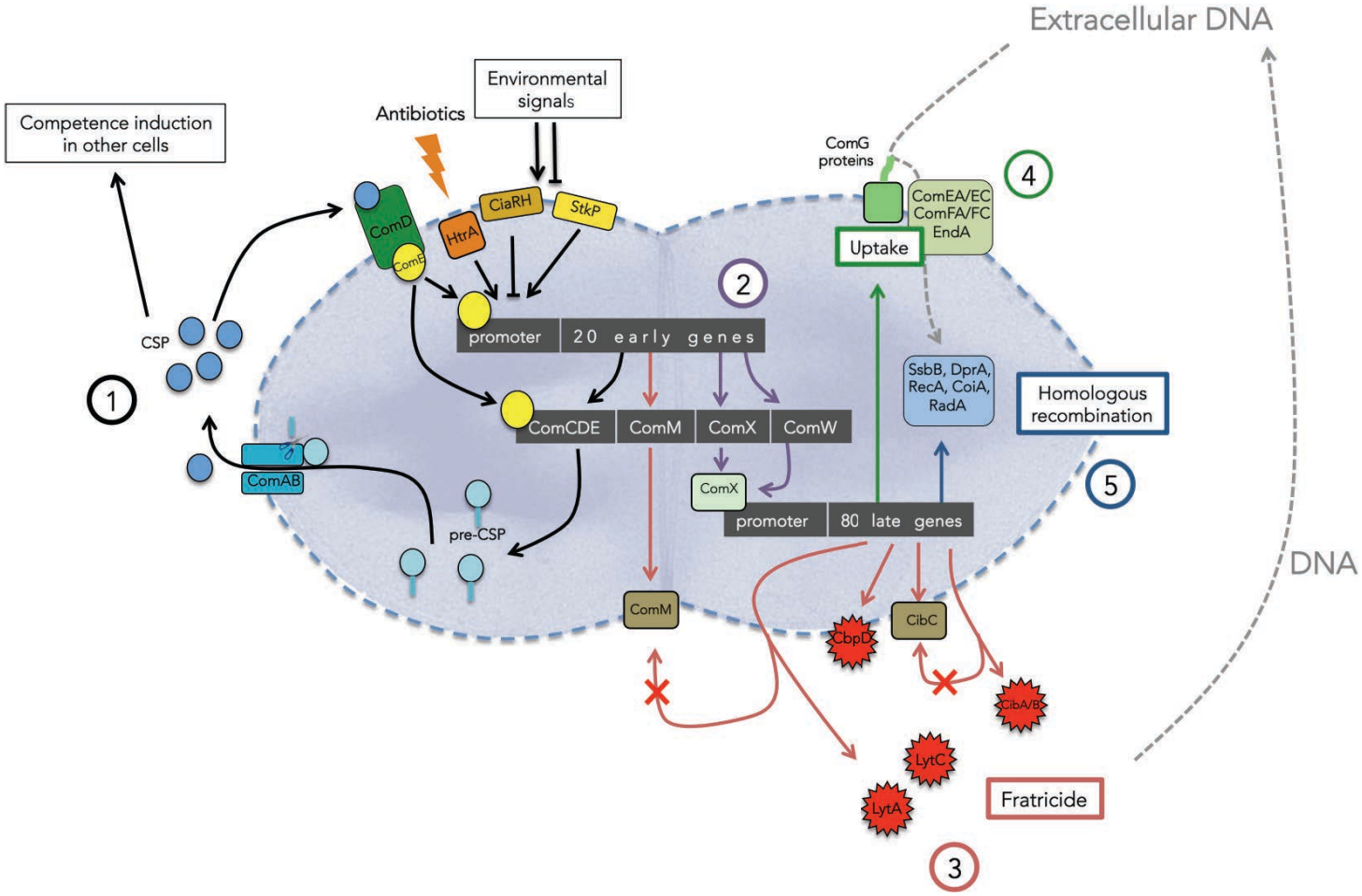


Figure 7 - The competence state

Adapted from Johnsborg et al., 2009. The competence state is divided in 5 steps: 1) the autocatalytic loop mediating the production of the CSP peptide; 2) the activation of comX and the late competence genes; 3) the fratricide behaviour; 4) the DNA uptake apparatus assembly; 5) the recombination of imported DNA.

The competence state can be divided in 5 steps (Figure 7):

1) Environmental signals are sensed by the two-components system CiaRH or the stress-regulated global regulator StkP, that regulate the basal transcription of the « early genes » promoters (Johnsborg et al., 2007; Martin et al., 2000; Prudhomme et al., 2006; Claverys et al., 2006). The protease HtrA is also involved by measuring the level of misfolded proteins in the cell (Stevens et al., 2011). In case of stress, like the presence of sublethal doses of antibiotics in the environment, the early genes promoters are activated. Among these early genes, *comC* encodes a precursor peptide, pre-CSP, that will be cleaved and secreted through the dedicated ComAB transporter into its mature form, the competence stimulating

peptide (CSP). This pheromone is sensed by ComD, a membrane-embedded histidine kinase, subsequently phosphorylating ComE, which will further activates the early genes promoter, resulting in an autocatalytic loop (Martin et al., 2000; Martin et al., 2010).

This activation generates an increasing concentration of CSP in the extracellular medium that will trigger competence in neighboring cells. The whole activation cascade ends in a competence peak within the cell population, occurring spontaneously during exponential phase in a cell culture (figure 8). Interestingly, a *comC* mutant strain, unable to produce CSP, is particularly helpful for laboratory experiments, by allowing to precisely control the triggering of competence in 100 % of the cells by the addition of CSP to the extracellular milieu.

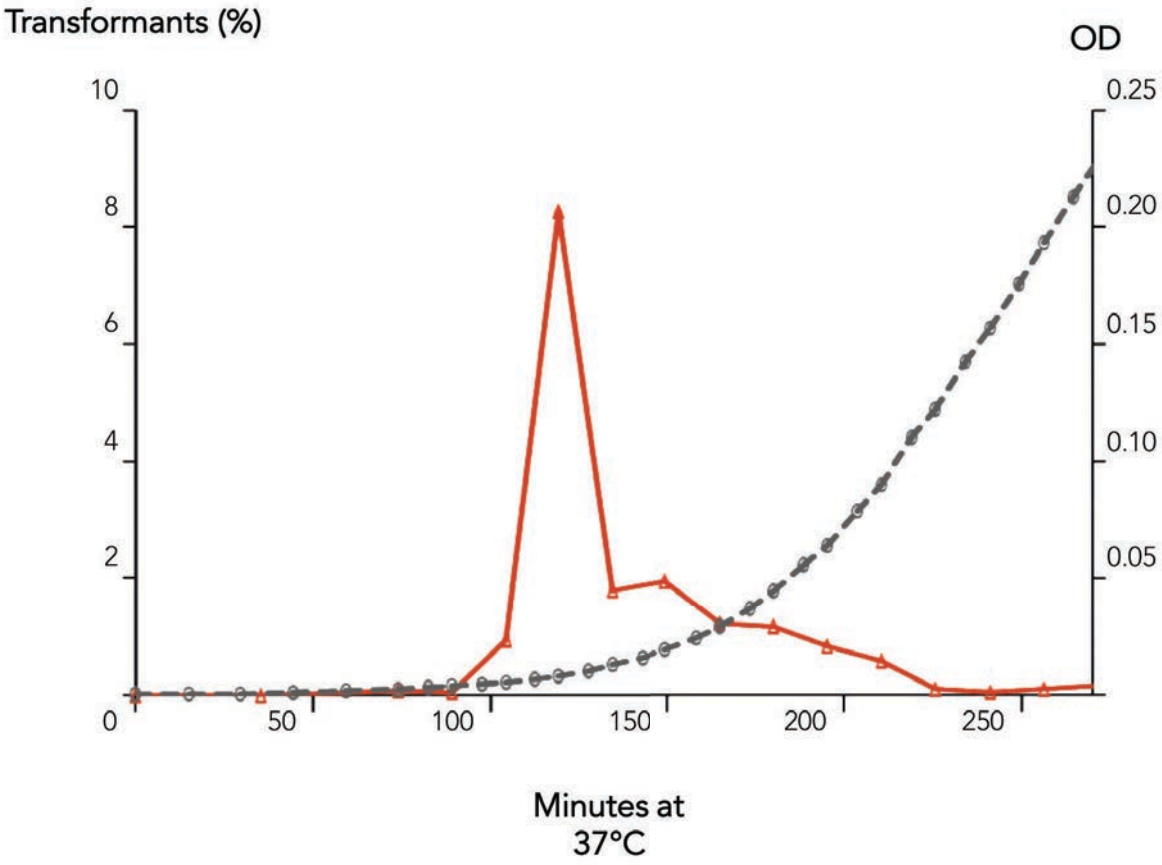


Figure 8 - Competence peak in a growing pneumococcal culture
Transformation efficiency assay with DNA conferring an antibiotic resistance was realized at different time points during the cell culture (red curve). Cell density is measured by the optical density (OD) at 492 nm (grey curve).

2) The sigma factor ComX, protected by the protein ComW, is responsible for the activation of a promoter controlling the transcription of 80 « late competence genes ».
3) Two hydrolases, CbpD and LytA, as well as two bacteriocins (narrow spectrum antibacterial molecules) CibA and CibB are part of these late genes, and are involved in fratricide, a unique behavior restricted to *S. pneumoniae* and closely related streptococci (Guiral et al., 2005; Claverys et al., 2007). These proteins combined together have the peculiar property to lyse *S. pneumoniae* cells. LytC, another hydrolase expressed outside of competence, is also involved: it is secreted and found in significative amount in the supernatant of a growing pneumococcal liquid culture, whether cells are competent or not. However it has been shown to necessitate the action of CbpD to actively digest the pneumococcal cell wall (Perez-Dorado et al., 2010). LytA is cytoplasmic, and requires cell death to actively engage in the pneumococcal population autolysis. CbpD is secreted, but it stays attached to the cell wall. Consequently, competent cells can only kill other cells in a contact-dependent manner (Eldholm et al., 2009). ComM is an immunity protein, protecting the competent cells from their own hydrolases. Altogether, LytA, LytC, CbpD and ComM are promoting the lysis of non-competent cells by competent cells (Johnsborg et al., 2008). Similarly to ComM, the competence induced protein CibC gives immunity from CibAB (Guiral et al., 2005).
4) All the proteins forming the DNA uptake apparatus, responsible for binding and internalization of DNA are expressed specifically during competence. The only known exception is the membrane-associated nuclease EndA, which is constitutively expressed and recruited at the site of DNA incorporation by ComEA (Berge et al., 2013). The whole apparatus is assembled and functional within 15 minutes after CSP induction. Double-stranded DNA (dsDNA) is preferentially bound at the cell surface. One strand is digested by EndA, while the other is incorporated through the channel ComEC, at a speed of ~ 90 nucleotides per second (5,4 kb / min) with a 3' to 5' polarity (Mejean et al., 1993).
5) Several intracellular DNA binding proteins are expressed specifically during competence, like DprA, SsbB and possibly RadA and CoiA. RecA is constitutively expressed but its

expression is increased during competence. SsbB is expressed in very high quantity, enough to cover entirely the size of *S. pneumoniae* chromosome. It is thought to make a reservoir of uptaken DNA, available for recombination (Attaiech et al., 2011). DprA, a competence specific RecA loader, displaces SsbB to initiate RecA polymerization on the foreign ssDNA. Follows RecA mediated homologous recombination with the chromosome. DprA is then involved the shutting down of competence state, by interacting with phosphorylated ComE (Mirouze et al., 2013).

The secretion of antimicrobial agents by bacteria is a widespread phenomenon, but is usually directed at other species, competing for a given niche. Why would bacteria secrete lysins specialized in the killing of its sister cells? Although no quantitative answer exist to this question, a possible way to rationalize this fratricide behavior is to take into account the bacterial natural habitat: biofilms inside the human oral cavity and nasopharynx. These biofilms are a collaborative bacterial lifestyle in which several bacterial species live together, densely packed inside an extracellular matrix, protected from the immune system and from antimicrobial agents. Pneumococcal biofilm formation has been linked to colonization of the host (Munoz-Elias et al., 2008; Domenech et al., 2011), and it was recently shown that the competence state is necessary for biofilm growth on human respiratory epithelial cells (Vidal et al., 2013). Interestingly, *S. pneumoniae* biofilms contain high amounts of extracellular DNA, and the secreted peptidoglycane hydrolase LytC, essential for biofilm formation, was shown to bind DNA (Domenech et al., 2013). In addition, CbpD/LytC mediated fratricide is active within a biofilms, and efficiently promoting genetic exchange by natural transformation in this particular environment (Wei et al., 2012).

Biofilm formation and fratricide are consequently linked together, and both favor uptake by the pneumococcus of DNA from its own species (or from closely related species, since fratricide lysins are only active on similar bacteria), rather than random exogenous DNA. This is further

supported by the fact that CbpD, absolutely necessary for fratricide to occur, stays bound to the bacterial surface, presumably making fratricide a very localized event within the biofilm. Moreover, it has been described that *S. pneumoniae* is able to produce genetic variability even by incorporating a piece of its own genome from a sister cell (self-transformation) (Johnston et al., 2013). Taken together, these data reinforce the hypothesis of natural transformation as a way to repair damaged DNA (preferentially taking up intra-species DNA for this purpose), and increasing genetic plasticity of the population.

D. Natural transformation in other bacteria

The ever-growing number of sequenced genomes has revealed that a vast number of bacteria carry a complete set of proteins necessary for natural transformation. However, it has not been proved that they are actually functional, and stays the possibility that these proteins are only remnants of the past ability to transform. As shown recently in *Staphylococcus aureus*, in which duplication of a normally repressed sigma factor gene is necessary to activate competence (Morikawa et al., 2012), or in *Vibrio cholerae* where chitin is necessary for induction (Meibom et al., 2005), it is likely that these transformation systems have yet unknown regulation mechanisms which makes them hard to demonstrate as functional in laboratory conditions.

Despite these difficulties, natural transformation has been well-studied in many distantly related organisms, providing a great wealth of information by comparing the different systems (Johnston et al., 2014).

Generally speaking, the main differences are:

- Various signals can trigger a transient competence state in many bacterial species such as *B. subtilis*, *V. cholerae* or *S. pneumoniae*. On the other hand, *Neisseria* species, *Thermus thermophilus* or *Helicobacter pylori* (Averhoff et al., 2009) are described as constitutively competent.

- The regulation mechanism can occur at the transcriptional, translational and/or post-translational level (Johnston et al., 2014).
- Specificity for the incoming DNA can vary. A DNA motif is recognized in Pasteurellaceae and Neisseriacae, able to select preferentially intra-species DNA for uptake. This DNA motif, called DUS (DNA-uptake sequence), is found in high quantity throughout their genome (Frye et al., 2013). All other species show no preference for any DNA, not even for closely related GC content DNA.

Main similarities are based on the DNA uptake machinery as a whole, staying strikingly similar throughout all these species:

- All rely on a Type IV pilus-like (T4P) operon, always required for the binding and uptake of DNA. This ubiquitous family of pili will be described in detail in the last part of the introduction.
- All possess a translocation apparatus, containing the proteins ComEA and ComEC homologs (nomenclature from *B. subtilis*), necessary for DNA entry inside the cell.
- All rely on single-stranded DNA (ssDNA) binding proteins for protection of the incoming DNA, the recombinase RecA and the highly conserved competence-specific protein DprA .

Of note, *Helicobacter Pylori* is an exception to the first point, as it uses a Type IV secretion system-like rather than T4P-like. It still relies on ComEC and DprA/RecA proteins for DNA translocation throughout the inner-membrane and recombination. Being almost the only bacterium described with such a system so far, it has been named the « black sheep » of natural transformation (Johnston et al., 2014).

IV. The DNA uptake apparatus in *S. pneumoniae*

S. pneumoniae natural transformation regulation is well understood, and its DNA uptake apparatus can serve as an archetypal system.

The process of DNA uptake can be divided in 3 steps: binding of exogenous DNA, degradation of one strand, and internalization of the other through the cell membrane. Data on the nature of the subunits involved and the interaction between them provides a partial view of how these sequential steps are orchestrated in the pneumococcus.

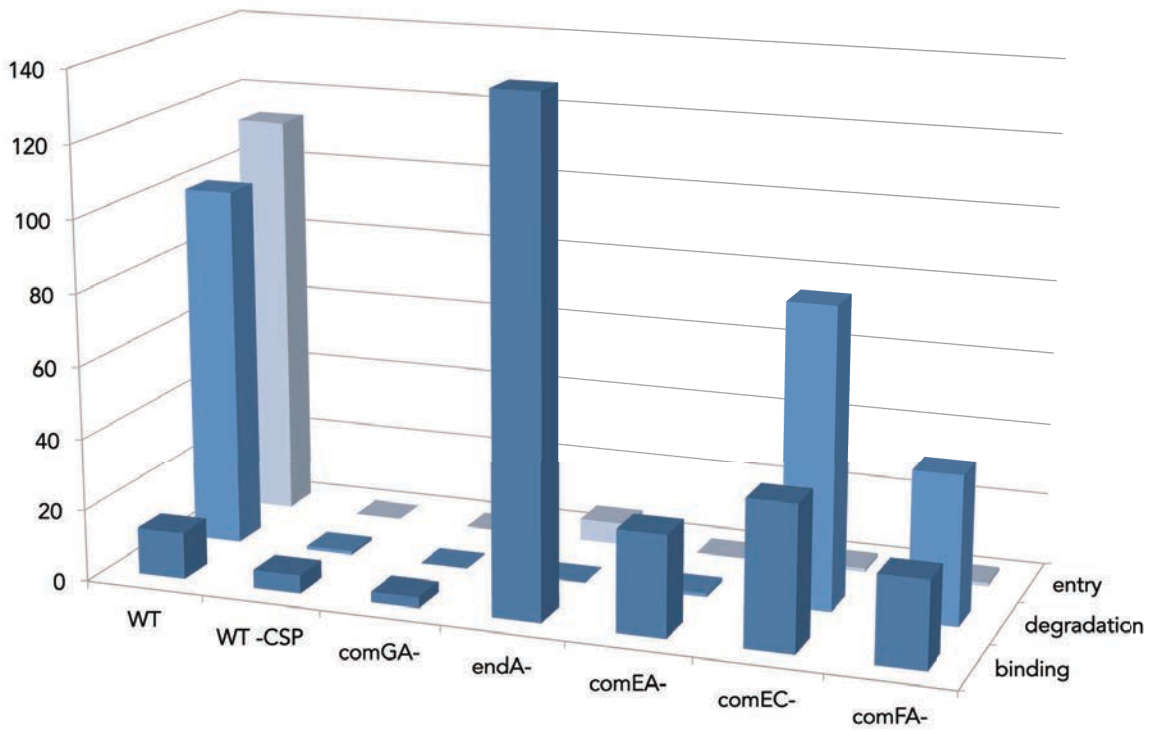


Figure 9 - Characterization of binding, degradation and internalization of donor DNA in wild-type and entry mutant strains

Data from Bergé et al., 2002. mutants were generated by mariner transposon mutagenesis. All strains used were unable to spontaneously induce competence state, which was induced by the addition of CSP peptide to the medium. Radioactively labelled donor DNA was added to the medium of competent cells. Degradation was measured from the supernatant radioactivity after precipitation of linear DNA. For DNA binding, cells were pelleted, resuspended in a buffer containing DNAse, and pelleted again. The radioactivity was measured in this second supernatant. For DNA internalization, the pellet from the previous step was resuspended, so that radioactivity measured in this last step corresponds to DNAse-protected DNA.

A. Binding

The *comG* operon is homologous to operons encoding T4P in Gram-negative bacteria. ComGA is an intracellular membrane-associated ATPase and ComGB is an integral membrane protein. Both could be responsible for the assembly of a transformation pilus using some or all the pilin

homologs (ComGC-GG). These pilins would only polymerize after cleavage of a signal peptide at their N-terminus by a dedicated pre-pilin peptidase PilD. Details on the possible mechanism of action of this operon will be developed below in the light of other well documented T4P systems.

By looking at the amount of radioactively-labeled DNA associated with competent cells, it was shown that deleting ComEA reduces the amount of DNA bound to the cell compared to a ComEC mutant for example. However the binding could only be completely abolished in a ComGA mutant, suggesting a two-steps binding of extracellular DNA, first depending on ComGA, and then on ComEA (Bergé et al., 2002). Of note, a similar two-steps mechanism for DNA binding has been proposed in *B. subtilis.* However, it was shown that in a deletion mutant of the *comGC-G* (encoding the pilin homologs), DNA binding is not entirely abolished. As in *S. pneumoniae,* DNA binding is only abolished in a *comGA* mutant, indicating that ComGA seems to be responsible for initial DNA binding to the cell, without the help of the pilin proteins. This led the authors to hypothesize that ComGA, which is intracellular, might be contacting an unidentified membrane protein responsible for the initial DNA binding (Briley Jr et al., 2011).

B. Generation of the incoming ssDNA

EndA is a nuclease inserted in the membrane, localized all over the cell surface. It is the only known protein so far to be involved in the DNA uptake apparatus without being specifically expressed during competence in *S. pneumoniae* (Claverys et al., 2009). Outside of competence, it is involved in human innate immunity escape by degrading neutrophil extracellular traps, consisting of antimicrobial agents bound to a DNA scaffold (Beiter et al., 2006; Moon et al., 2011). Recruited by ComEA, EndA is responsible for digesting one strand of the incoming DNA. Interestingly, although degradation of one strand of the incoming DNA is a

conserved feature of all DNA uptake systems, *S. pneumoniae* is the only bacterium in which the nuclease responsible has been identified.

C. Internalization

Incorporation inside the cell is both abolished when mutating the integral membrane protein ComEC, or the helicase-like membrane-associated ATPase ComFA. Since DNA binding and degradation is maintained in these mutants, it suggests that the ComG proteins, ComEA and EndA are functional independently from ComEC and ComFA.

By catching a µm-sized DNA-coated bead inside a laser beam trap, a setup called optical tweezers, and by approaching the bead from a competent *B. subtilis*, it was possible to evaluate the DNA binding and rate of incorporation of a unique DNA molecule. These experiments confirmed the two-step binding mechanism, with initial binding dependent on the comG operon, and second more tight binding and internalization dependent on ComEC. The speed of incorporation measured was 80 nucleotides per second, consistent with the ~ 90 nucleotides per second estimation in *S. pneumoniae* (Mejean et al., 1993). These experiments additionally revealed that the proton electrochemical gradient across the membrane, termed proton motive force, is necessary (but might not be sufficient) for internalization (Maier et al., 2004).

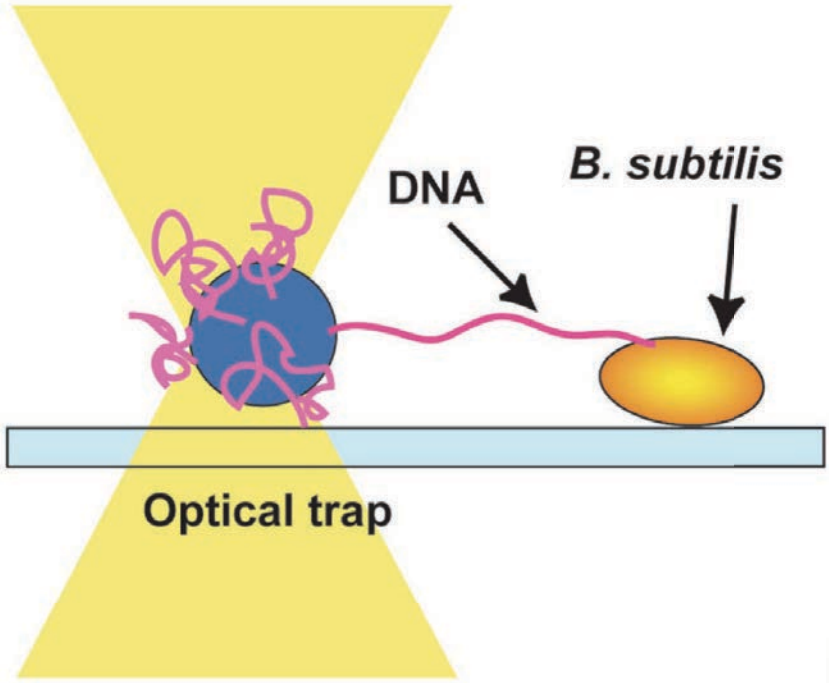


Figure 10 - Optical tweezers setup for single molecule DNA internalization measurements

Taken from Maier et al., 2004. Pulling on the DNA tether by the bacterium creates a measurable deflection of the bead from the center of the optical trap. Alternatively, at a higher trap stiffness, the bacterium can slide on the surface towards the bead while incorporating the DNA tether, providing a direct information on the rate of incorporation.

Interestingly, using the same setup, it was shown that the Gram-negative bacterium *H. pylori*, with its hybrid Type IV secretion system/ComEC DNA uptake machinery, is incorporating DNA more than 10 times faster than *B. subtilis*, but with a weaker force. Indeed DNA internalization could be reversed by applying an external force of 40 pN on the DNA tether, while it was not even decreasing the speed of incorporation in *B. subtilis* (Stingl et al., 2010).

D. The transformasome

ComEA and EndA co-localize together at midcell, at the spot where fluorescently labeled DNA is seen to accumulate during transformation (Bergé et al., 2014). ComGA, ComFA, and several of the ssDNA binding proteins (RecA, DprA) have been shown to co-localize by FRET studies in *B. subtilis*. Moreover, ComEA, ComEC and ComFA show localization dependency in *B.*

Subtilis (Kramer et al., 2007), and pulling specifically on the ComGG protein co-purifies ComFA, ComEC, RecA, SsbB, DprA (Mann et al., 2013). Strikingly, Kaufenstein and colleagues showed through immunofluorescence and FRAP experiments (Fluorescence Recovery After Photobleaching) that this complex seems highly stable, and does not consists of dynamically exchanging sub-complexes (Kaufenstein et al., 2011).

Altogether these data suggest the formation of a "transformasome", a protein complex embedded in the bacterial membrane responsible for binding exogenous DNA, processing it into ssDNA, incorporating it inside the cell, and protecting it from endogenous nucleases.

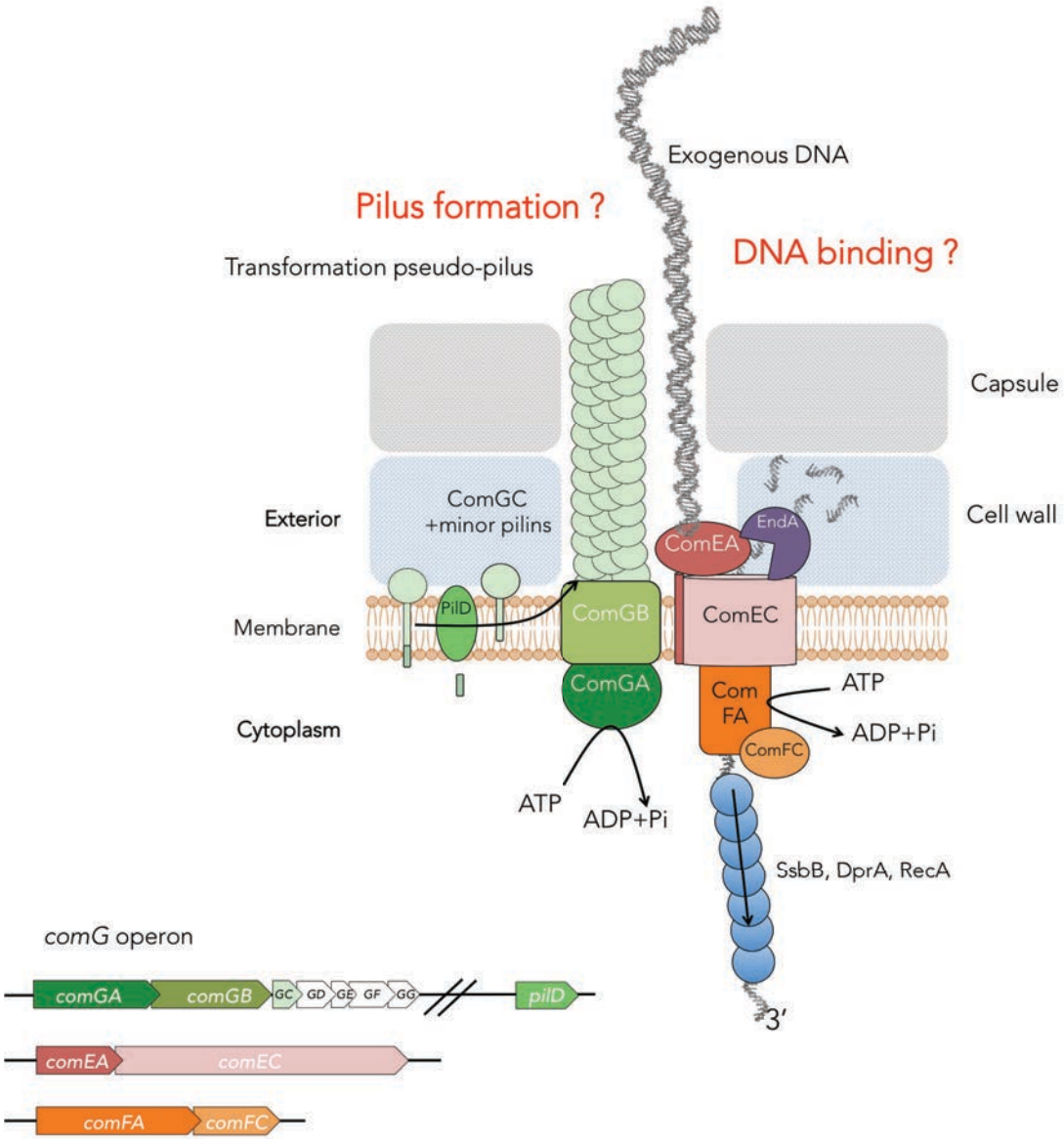


Figure 11 - Model of S. pneumoniae transformasome from the literature

Proteins are color-coded, and the operon structure of the genes involved is shown on the bottom left. Genes are drown to scale (ComGA is ~1 kb).

The aim of my PhD is to study the role of the *comG* operon in natural transformation. Being conserved in almost all DNA uptake systems, and necessary for DNA binding to the cell, it is the first logical target to study in order to understand the mechanism of DNA uptake. Moreover, operons from the *com* family have never been clearly described in any Gram-positive bacteria, which raises the question of their similarity to the Gram-negative counterparts. A comprehensive view of T4P systems is necessary to picture the range of possible structures and possible roles of this particular operon.

V. Type IV pili: a case study for evolution at work

T4P systems form a family of incredibly diverse and versatile appendages, widespread among prokaryotes. They can be seen as a concentrate of bacterial ingenuity. Well studied because of their high potential as vaccine targets, they keep surprising researchers by the extent of their functions (Giltner et al., 2012). Even though some of them could have been transferred through HGT, their presence all over the eubacterial and archeal branches suggests that they appeared particularly early during evolution.

A. Overview of Type IV pili systems

i) Mechanism of assembly

The central mechanism behind T4P systems is very simple: a membrane-embedded motor, using ATP hydrolysis energy to polymerize a single subunit, "the pilin", into a helicoidal fiber "the pilus". Upon polymerization, the pilins that were initially inserted in the membrane have their hydrophobic residues gathered in the core of the pilus, providing a strong cohesion to the architecture through hydrophobic interactions.

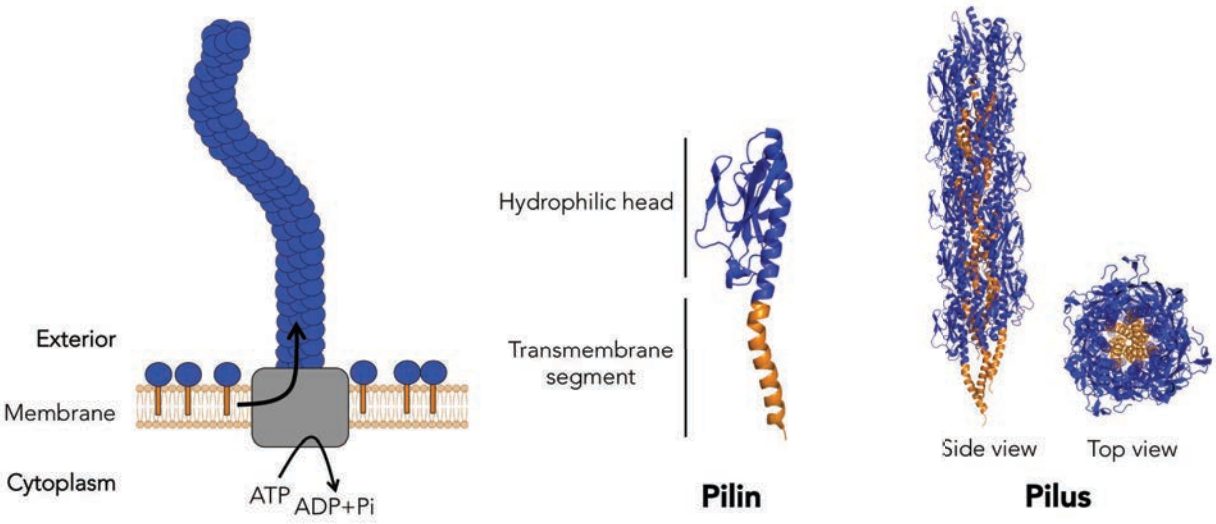


Figure 12 - Schematic view of the T4P biogenesis mechanism
The pilin displayed is PilE1 from Nesseria gonorrhoeae (uniprot accession number P02974; PDB code 1AY2). The pilus model (PDB code 2HI2) was obtained by fitting this pilin structure in an electron microscopy map of the native pili (Craig et al., 2006). The N-terminal transmembrane helix is colored in orange, the C-terminal soluble globular domain is colored in blue (realized with the PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC). Pilus and pilin structure are not at the same scale.

Added to this mechanism, an upstream post-translational control operates on the pilin, which needs a maturation step before being assembled by the platform: a positively-charged signal sequence on its N-terminus has to be cleaved by a dedicated aspartyl protease, the ‘prepilin peptidase’. Alternative version of the pilin, the ‘minor pilins’, are part of the system, and usually in minority within the fiber.

ii) Conserved features

There is usually very little similarity between the solvant-exposed domain of the pilins from different T4P systems. However, the hydrophobic N-terminal helix has some very conserved features, mainly the prepilin peptidase cleavage motif, and a glutamic acid residue in position 5 (Glu5) after the cleavage site. The positively charged cleaved peptide is important for the orientation of the pilin in the membrane, and might prevent extraction of the pilin to form the pilus. The negatively-charged Glu5 is thought to form a salt bridge within the pilus hydrophobic core with the positively-charged N-terminal amine group of the next pilin (Craig et al., 2006).

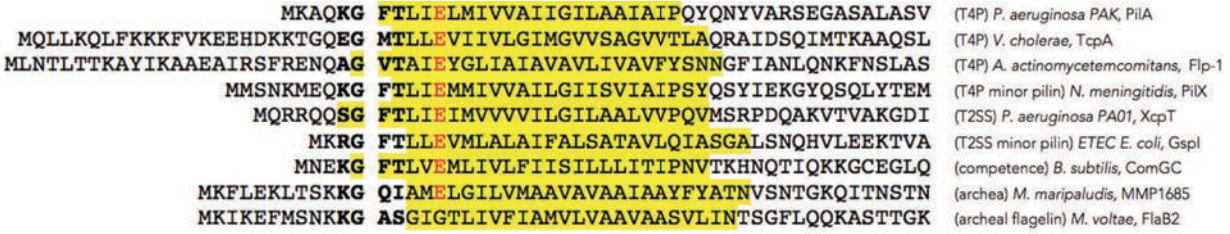


Figure 13 - Alignment of the N-terminal part of a wide range of Type IV pilin

Adapted from Giltner et al., 2012. Transmembrane regions highlighted in yellow were determined using the TMHMM software on the full length pilin. The recognition motif of the prepilin peptidase is shown in bold letters, with a space indicating the cleavage site. The highly conserved Glu5 residue on the mature pilin is in red.

This conservation enabled the implementation of an efficient Type IV pilin finder software, by scanning genomes for the signature motif of the N-terminus portion (Imam et al., 2012).

In addition to cleavage, it has been shown in many Gram-negative T4P systems that the prepilin peptidase adds a methyl moiety to the N-terminus of the pilin (Aly et al., 2013). Moreover, a dedicated oxydoreductase forms an intra-subunit disulfide bridge in most pilins, although some Gram-positive pilins, like *S. pneumoniae comG* pilins, lack this post-translational modification system. Calcium binding is, as well, a conserved feature of many pilins, most of them having a calcium binding loop. Interestingly, the first crystal structure of a Gram-positive pilin was obtained recently, and revealed the incorporation of a structural zinc ion, maintained in an original binding pocket, involving 3 cystein residues (Piepenbrinck et al., 2013).

iii) Families of Type IV pili systems

In Gram-negative bacteria T4P are usually separated in 2 categories, T4a and T4b, based on the size of the signal peptide (shorter for T4a, longer for T4b), the size of the mature pilin (between 150 and 160 amino acids for T4a, either longer or much shorter for T4b). While the T4a category represents a homogeneous family among distantly related bacteria, sharing a

similar architecture of operons, the T4b family is not as well delineated (Pelicic et al., 2008, Giltner et al., 2012).

When comparing to Gram-positive bacteria, especially in *Clostridia* species, the distinct features of all categories are mixed, suggesting that pili in these species are closely related to ancient T4P systems before they diverged into the families observed today (Melville et al., 2013).

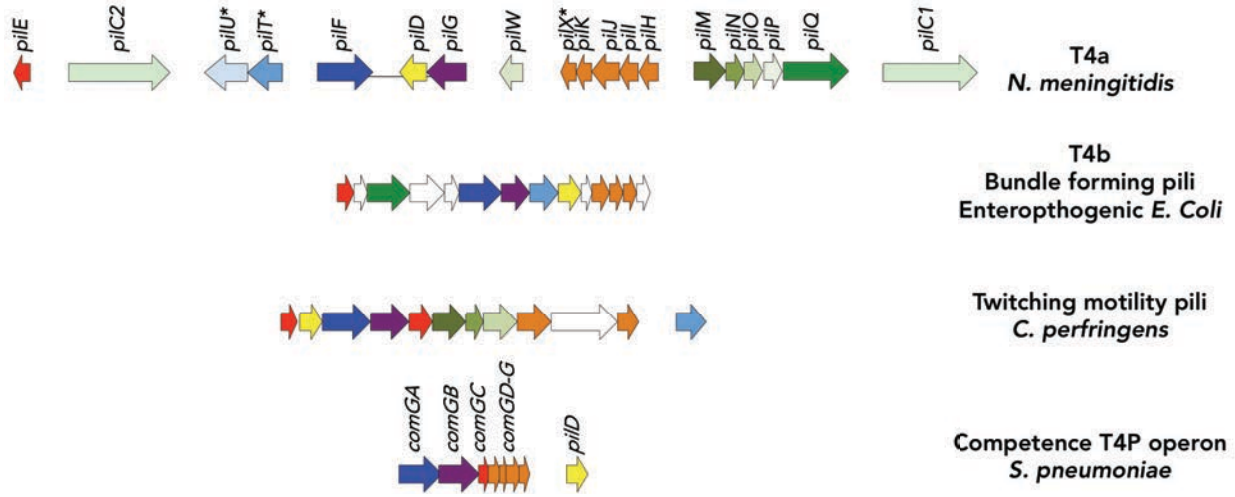


Figure 14 - Comparison of T4P gene clusters

Adapted from Pelicic et al., 2008. Homologous proteins are shown in the same color. Genes are drown to scale (ComGA from S. pneumoniae is ~1 kb). Genes displayed are essential for pilus biogenesis in the different systems, except for those indicated by star (the retraction ATPase PilT, the secondary assembly ATPase PilU, and the additional minor pilin PilX). Major pilins are in red; minor pilins in orange; inner-outer channeling proteins in shades of green; ATPases in shades of blue; integral membrane platform in purple; prepilin peptidase in yellow.

Importantly, while a minimum of around 12 proteins are necessary for most T4P systems, only 8 proteins have been identified as forming part of the competence T4P system in *S. pneumoniae* or *B. subtilis.* If a T4P is truly assembled by these systems, they would represent one of the simplest bacterial T4P known so far.

B. Functional diversity of Type IV pili systems

Starting from this core apparatus, a wide range of molecular machineries have been generated throughout billion years of evolution, covering an impressively diverse range of functions.

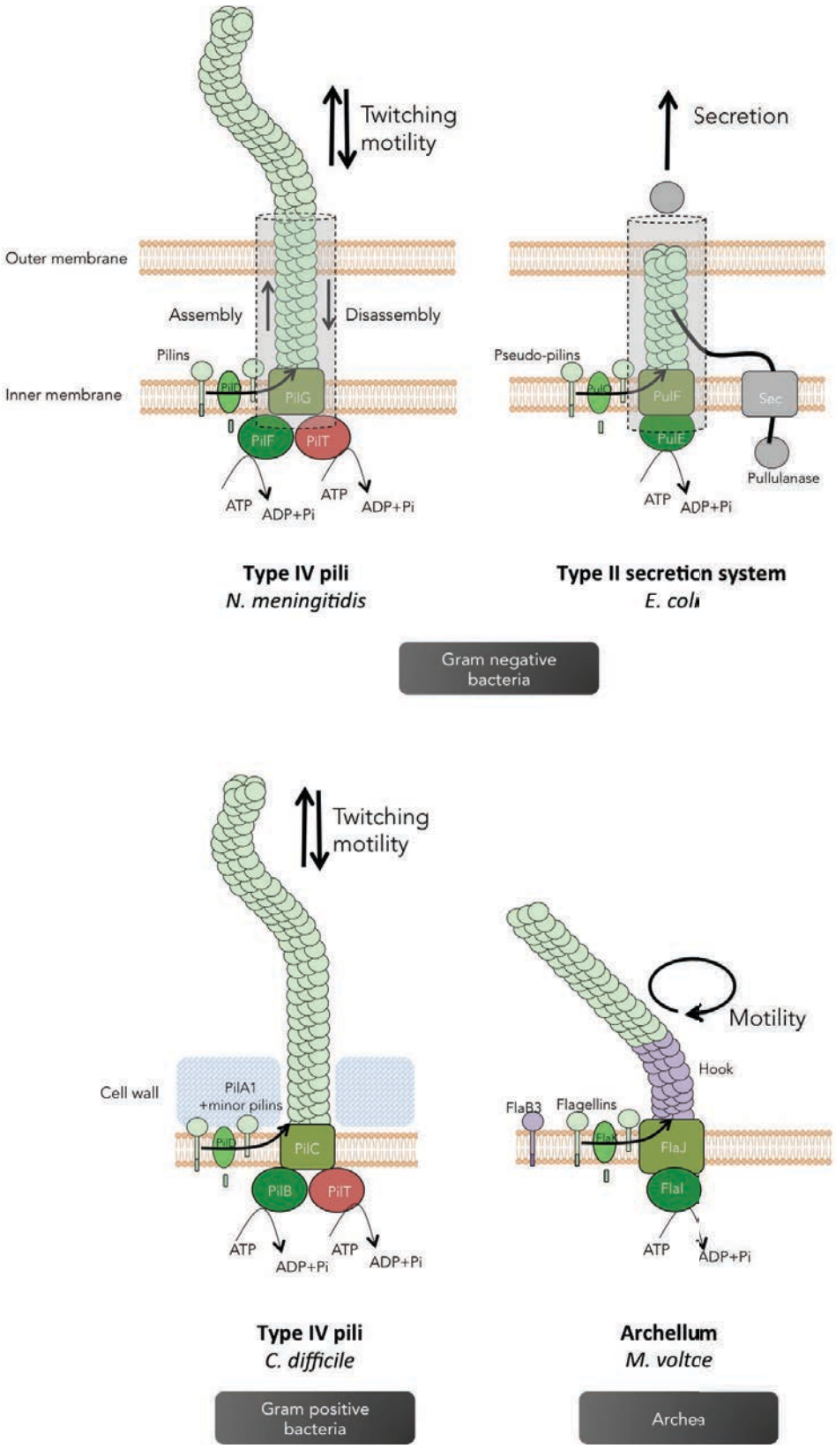


Figure 15 - Comparison of varied T4P systems and their functions

Homologous proteins are shown in the same color. In N. meningitidis: PilD is the prepilin peptidase; PilF is the assembly ATPase; PilT is the retraction ATPase; PilG is the integral membrane platform. Inner-outer channeling proteins for Gram-negative bacteria are shown as a transparent grey cylinder.

i) Biofilms

Fibers were adapted into a sticking device used to form biofilms. By binding to each other and to other cells, T4P form huge bundles and strongly attach cells to each other. Four major human pathogens are using T4P mediated biofilm formation as the keystone of their infection process: *Pseudomonas aeruginosa*, leading cause of nosocomial infections, *Nesseria gonorrhoeae* colonizing the urogenital tract and causing a widespread sexually transmitted disease, enteropathogenic *Echerischia coli*, an important pediatric diarrhea pathogen, and *Nesseria meningitidis* able to cross the blood brain barrier to cause central nervous system deadly infections.

For example, during the well described colonization process of *N. meningitidis*, T4P are required for their initial attachment onto human endothelial cells, thanks to the major pilin PilE binding specifically with the beta2-adrenergic receptor (Coureuil et al., 2010; Coureuil et al., 2012).

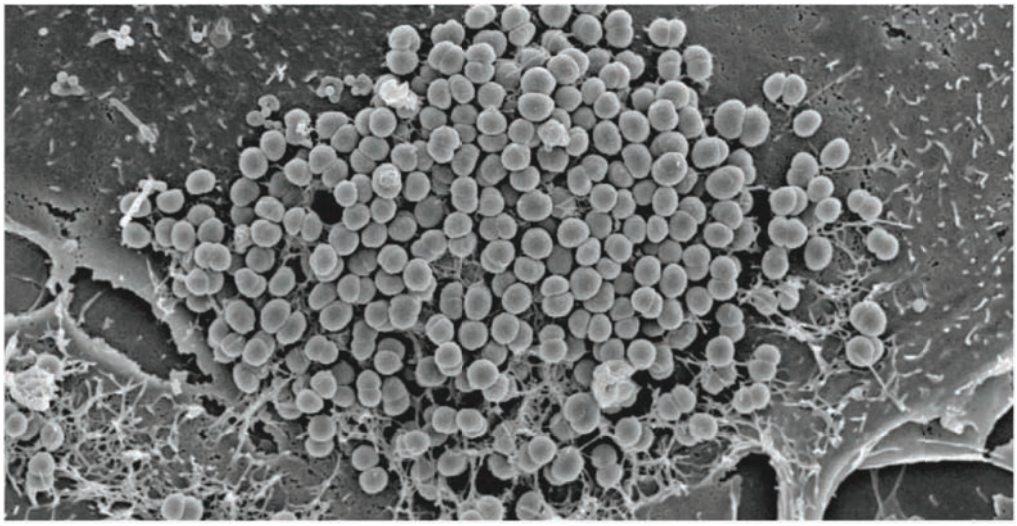


Figure 16 - Scanning electron micrograph of a N. meningitidis biofilm on a human cell
Taken from Chamot-Rooke et al., 2011. Web-forming filaments observed are T4P.

This major pilin sequence undergoes antigenic variability, a process of gene shuffling consisting of intragenic recombination of part of the *pilE* gene locus with silent pilin gene cassettes (Hill et al., 2009). This high variability hides it from the immune system, and although well exposed on the surface PilE was abandoned as a potential vaccine target for this reason. Post-translational modifications like the addition of sugars or phosphates on the pilin is used to expand even more the functions of the pili. By changing the sugars exposed on its surface, the bacteria have another way to mask themselves efficiently from the immune system. Moreover it enables them to "tune" their adherence according to their environment: by regulating the number of added phosphates on the major pilin, *N. meningitidis* is able to reverse the biofilm formation process after a few hours, and return to a planctonic state to allow better propagation (Chamot-Rooke et al., 2011). A controversy exists on the precise role of *N. meningitidis* minor pilins. Like the major pilin, PilV binds specifically to the beta2-adrenergic receptor, and was proposed to be necessary for triggering the opening of the blood brain barrier (Mikaty et al., 2009). PilX was proposed to be essential for pilus mediated adhesion and aggregation, presumably by forming protruding hooks on the pilus surface, preventing pili to slide on one another (Helaine et al., 2007). These properties would occur upon a conformational change in the pilus when it is elongated by a pulling force (Brissac et al., 2012). This force is provided by retraction of the pili, a process that will be described in the next part. On the other hand, Imhaus et al. suggested recently that PilV and PilX would exert their function from the periplasm rather than by being incorporated in the pilus, and that their mere involvement in pilus biogenesis could account for the defective phenotype described in the work cited above (Imhaus et al., 2014).

Far from infection, T4P biofilms are involved in other fascinating processes: *G. sulfurreducens* and *G. metallireducens* use their T4P as electron transfer cables, consequently named nanowires. Their T4P are conductive under physiological conditions, giving a high reductive power to biofilms of these bacteria. They are notably able to precipitate potentially toxic metals,

making them important players of microbial ecology (Malvankar et al., 2012, Reguera et al., 2005).

ii) Twitching motility

Thanks to their strong adherence properties, T4P provide a way for the cell to « crawl » on a surface, a process called twitching motility. Used as a grappling hook, the pilus by its random movement gets stuck to the surface, and by retraction, the bacterium is able to pull itself towards the point of attachment. Retraction is the exact opposite of pilus growth: a dedicated ATPase is coupled to the assembly platform in the membrane, but use ATP-hydrolysis energy to depolymerize the pilins instead (Satyshur et al., 2007; Misic et al., 2010), therefore generating a force of retraction. The available structure of this retraction ATPase, PilT, from the hyperthermophile *Aquifex aeolicus* and the human pathogen *P. aeruginosa* provided some clues on a possible retraction mechanism. The hexametric ring ATPase shows large domain motions upon interaction and hydrolysis of ATP, that could drag a pilin back inside the membrane.

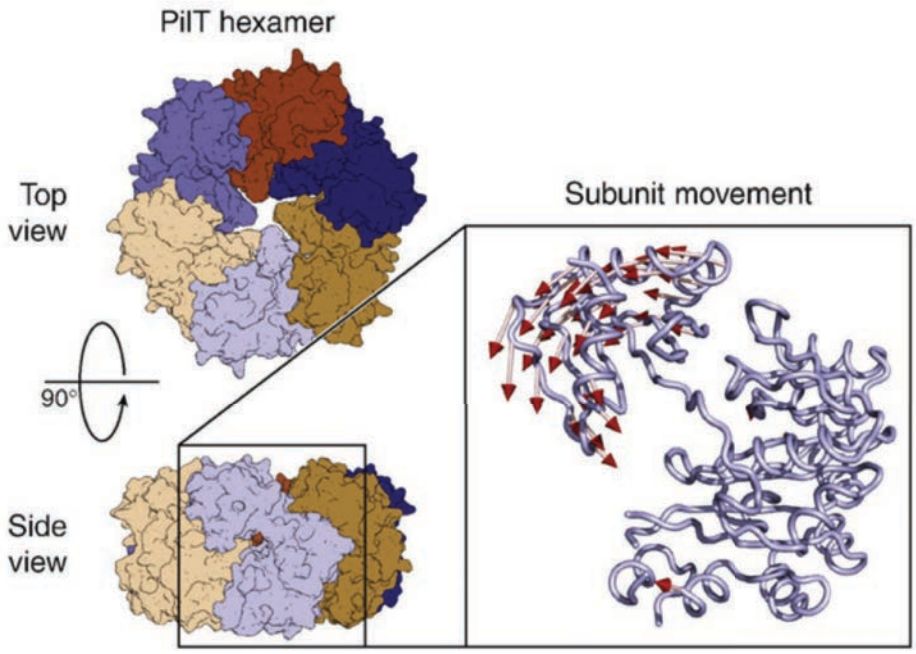


Figure 17 - P. aeruginosa PilT structure, and domain motion visualization

Taken from Misic et al., 2010. ATP binding leads to large domain movements (red arrows) within the subunit (light-blue cartoon). Left panels show two separate hexamer views, with each subunit individually colored. The zoom-in is of a single subunit (light blue) illustrating the motions of the moving domain during ATP binding, with the RecA domains fixed.

Whether this motion transfer occurs entirely through the conserved integral membrane platform, or the ATPase is interacting directly with the pilin transmembrane helix is still unclear.

The pulling force generated by pilus retraction can be accurately measured using the optical tweezers setup. When the pili from a bacterium stuck on the coverslide surface are bound to the bead in the optical trap, the displacement of the bead from the center of the trap is directly related to the force generated by the PilT motor. Doing so, it was evaluated that individual motors (i.e. the force generated by a single pilus retraction ATPase) exceeds 100 pN (Maier et al., 2002). Not single pili but rather bundles of pili working together are thought to mediate twitching motility, through forces reaching the nanonewton, making of pilus retraction the strongest micro-scale biological motor known so far (Burrows et al., 2005; Biais et al., 2008). Twitching motility can be guided by a chemotaxis mechanism (Miller et al., 2008). Concerning

Gram-positive bacteria, twitching motility has only been described in a few Clostridia species (Varga et al., 2006).

Minor pilins are again involved. In *P. aeruginosa* PAO1 strain, minor pilins have been shown by immunospecific labeling to be incorporated in the pilus, although it was not clear if they were only at the tip or incorporated inside the pilus. Mutating these minor pilins led to pilus deficient strains (pili were greatly reduced, but not completely abolished), and complementation by the low expression of the pilin gene on a plasmid restored piliation and twitching motility. However, complementing by overexpressing the first three minor pilins intriguingly led to twitching motility deficient phenotypes, while the last two minor pilins led to an increased twitching motility proportional to the expression level (Giltner et al., 2010). These data show the involvement of the minor pilins in modulating the pilus function, and suggest that the entire set of minor pilins, extremely conserved in all T4P systems, evolved as one functional unit that can be perturbed by changing its stochiometric ratio.

iii) Secretion

Far from the previous examples, this fiber assembly mechanism was adapted into a secretion apparatus, called the Type II secretion (T2S) system. A very short filament is formed within the periplasm (consequently named pseudopilus), and mediates the secretion of substrates through the bacterial outer-membrane (Vignon et al., 2003).

The architecture of the T2S system is the best documented from the T4P superfamily. Many crystal and electron microscopy structures are shaping a more and more detailed overview of the whole apparatus (McLaughlin et al., 2012).

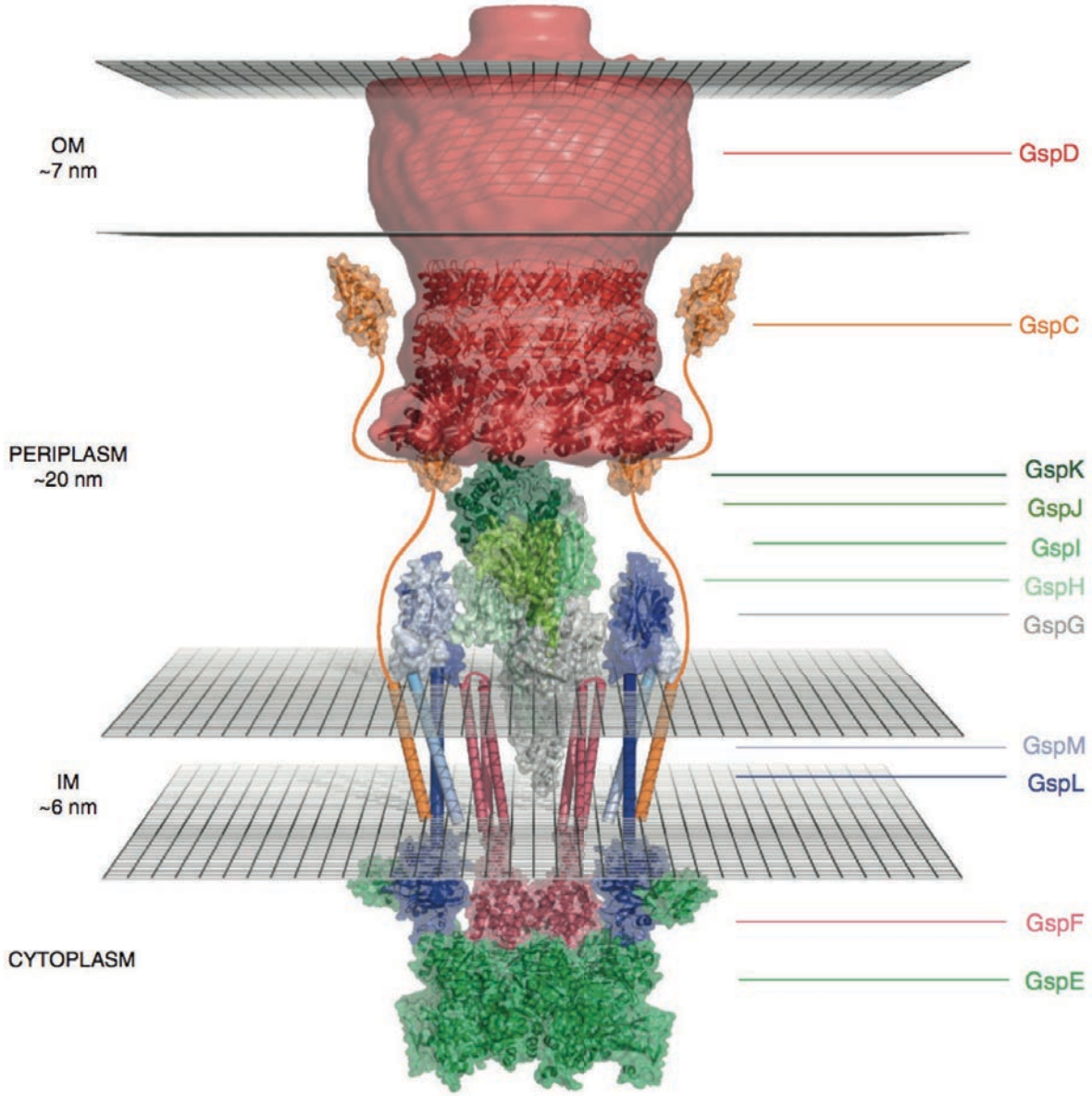


Figure 18 - Gathering all T2S subunit structures

Taken from McLaughlin et al., 2012.

- *GspE: pseudopilus assembly ATPase, cytoplasmic. The putative hexametric conformation has been modeled using the P. aeruginosa retraction ATPase PilT structure.*
- *GspF: inner membrane assembly platform, supposedly positioned at the base of the pilus. Only a small cytoplasmic part of its structure is available.*
- *GspL-M: integral membrane proteins, forming a channel around the pseudo pilus. Their transmembrane domain is missing.*
- *GspG: major pilin, GspH-I-J-K are minor pilins.*
- *GspC: Couples this inner membrane complex to the the pore protein on the outer membrane, the secretin GspD, for which a low resolution electron microscopy structure is available.*

Of particular interest, an insightful crystal structure of a complex of the three minor pilins, GspI, J and K, from enterotoxigenic *E. coli* T2S system was obtained (Korotkov n et al., 2008). GspK pilin, which is significantly bigger than the other pilins, is placed on top of GspI and J. It was hypothesized that this complex could be placed at the tip of the pseudopilus. This idea was strongly supported by a study showing that this complex was also formed by the homologous proteins from the pul system of *K. oxytoca*, and was necessary for the initiation of pseudopilus polymerization, but not for its elongation (Cisneros et al., 2011). By molecular dynamics simulation of the GspI-J-K complex within a lipid bilayer, it was also shown that GspK protein was moved upward by its interaction with the GspI-J complex, creating a deformation in the membrane. This conformation could potentially be the thermodynamic trigger for activation of the pseudopilus assembly platform.

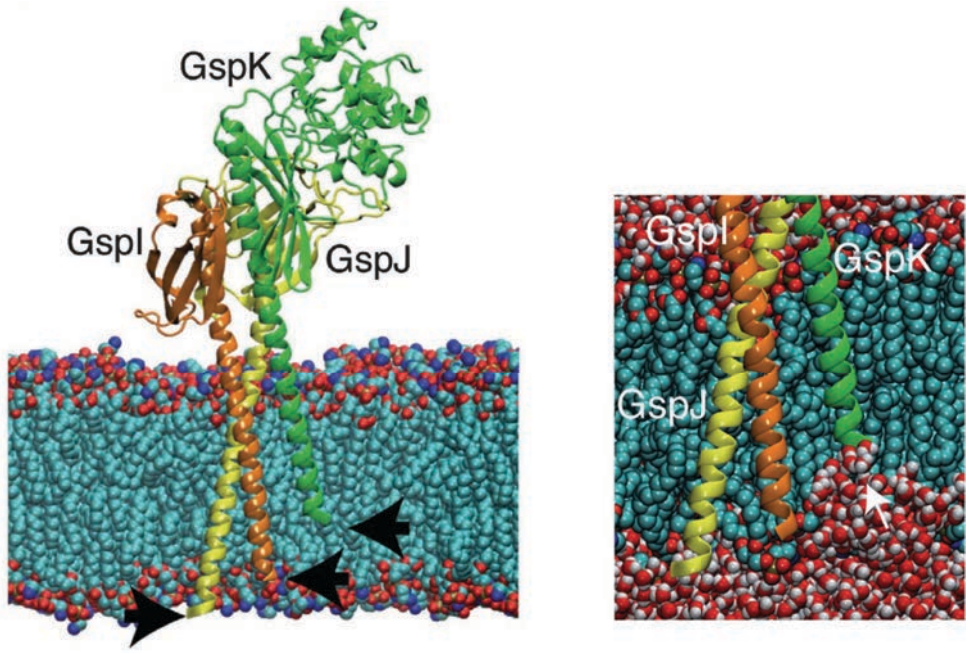


Figure 19 - Molecular dynamics simulation of the GsPI-J-K ternary complex inside a membrane

Taken from Cisneros et al., 2012. Left: GspI-J-K complex placed inside a bilipidic membrane, after addition of the transmembrane helices on the crystal structure. N-termini are shown by black arrows. Right: result of the molecular dynamics experiment showing the membrane perturbation by GspK upward position, white arrow.

Given the fact that minor pilins are almost constantly required for pilus biogenesis, and sharing similar features, their initiation role is likely conserved in many T4P systems.

Despite the wealth of data on this apparatus, the precise mechanism of secretion remains unclear. Two different models are currently proposed: the piston model, in which the pseudopilus would undergo cycles of assembly and disassembly to push substrates through the outer-membrane channel (called secretin); and the screw model, in which the pseudopilus is constantly assembled at its base and disassembled at its tip, and powers through its assembly the passage of substrates from the periplasm to the exterior through the secretin (Nivaskumar et al., 2014).

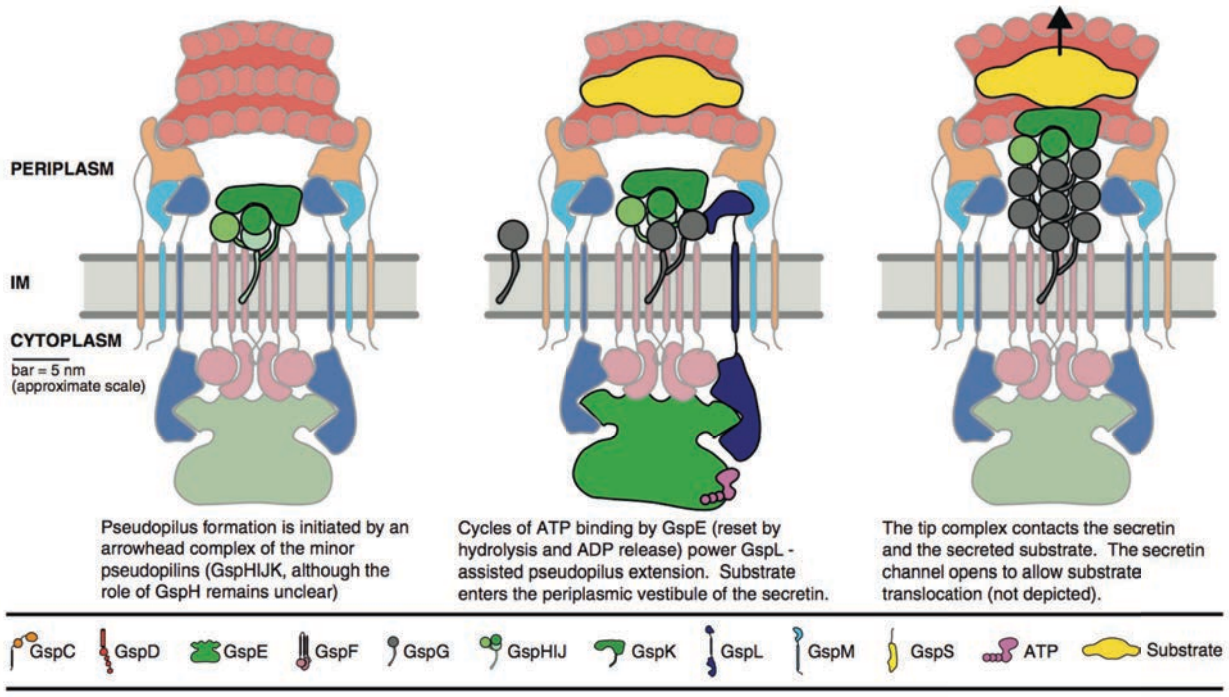


Figure 20 - T2S mechanism from current data

Taken from McLaughlin et al., 2012. Assembly and function of the pseudopilus. For clarity, the panel focuses on the IM complex, and the C-terminal domains of the secretin and channel clamp are not shown.

iv) Swimming

Perhaps the most structurally distinct adaptation is the so-called archellum, a T4P derived Flagellum in archea. With a hook shape at its base formed by a minor pilin, this long appendage is thought to motorize the cell by rotation in a similar fashion as Flagella in Gram-negative bacteria (Pohlschroder et al., 2012; Van Wolferen et al., 2013; Albers et al., 2006; Ng et al., 2008)

v) Type IV pili structures

T4a and T4b pilus structures were obtained by fitting a crystal structure of the pilin in a high resolution cryo-electron microscopy map (Craig et al., 2006; Li et al., 2012). The T4a pilus from *N. gonorrhoeae* is densely packed, with numerous electrostatic interactions between the globular C-terminal domains. It was shown to resist high temperatures, proteolysis, and up to 8M urea! T4b pili globular domains show a more relaxed conformation, with less inter-pilin interactions, forming deep grooves on its surface. This could account for their weaker resistance compared to T4a pili.

By overexpressing the major pilin, a long T2S pseudo-pilus could be generated and its helical parameters could be determined by electron microscopy (Kohler et al., 2004). Combining these data with the crystal structure of the pilin, several pseudopilus models were generated by defining the minimum energy conformations by molecular modeling, and the accurate model was validated using biochemical restraints on the structure (double cystein substitution crosslinking and charge inversion mutants) (Campos et al., 2010).

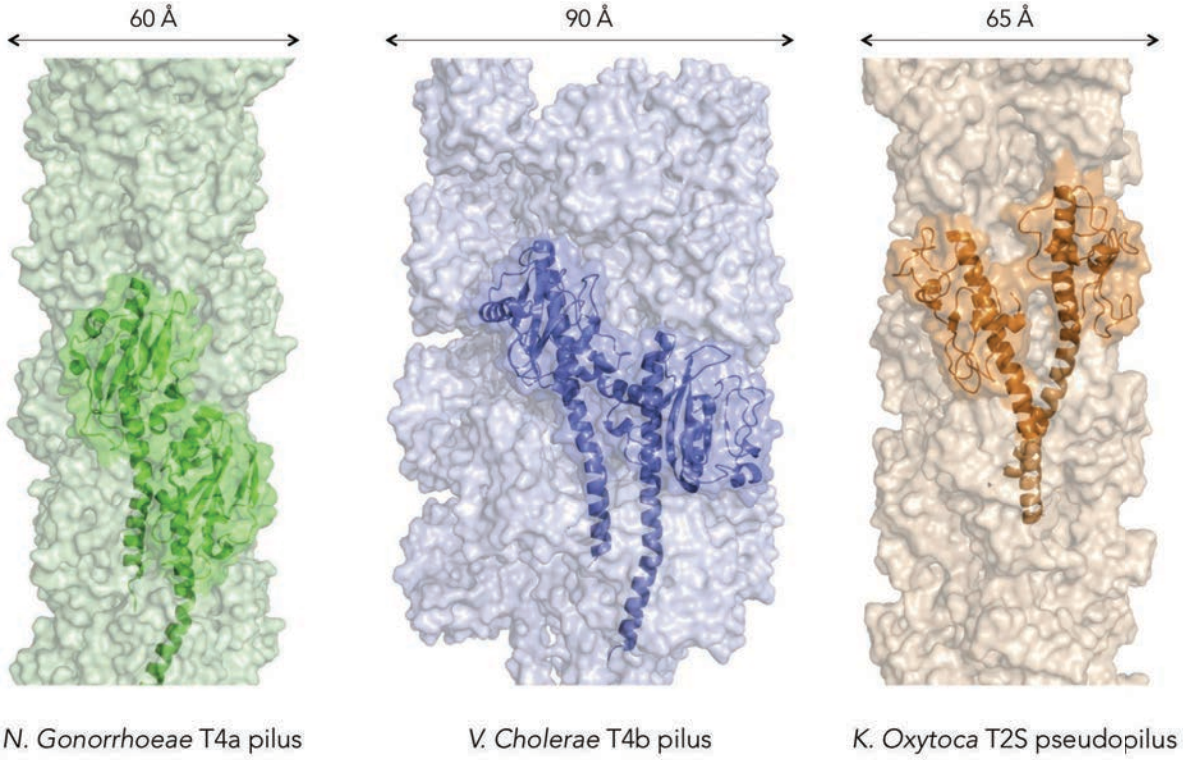


Figure 21 - Models of T4P structure

Figure taken from Giltner et al., 2012. Two pilins are highlighted within the fiber. The T4a pilus has 3.6 pilins per turn and a helical pitch of 37.8 Å, T4b has 3.7 and 129 Å, while T2S pseudopilus has 4.25 and 43.8 Å. Despite these differences in helicity, the axial rise of a pilin is similar, 10.5 Å, 8.4 Å and 10.4 Å respectively. In the three cases, hydrophobic interactions are the major source of cohesion in the fiber.

C. *Type IV pili and DNA uptake*

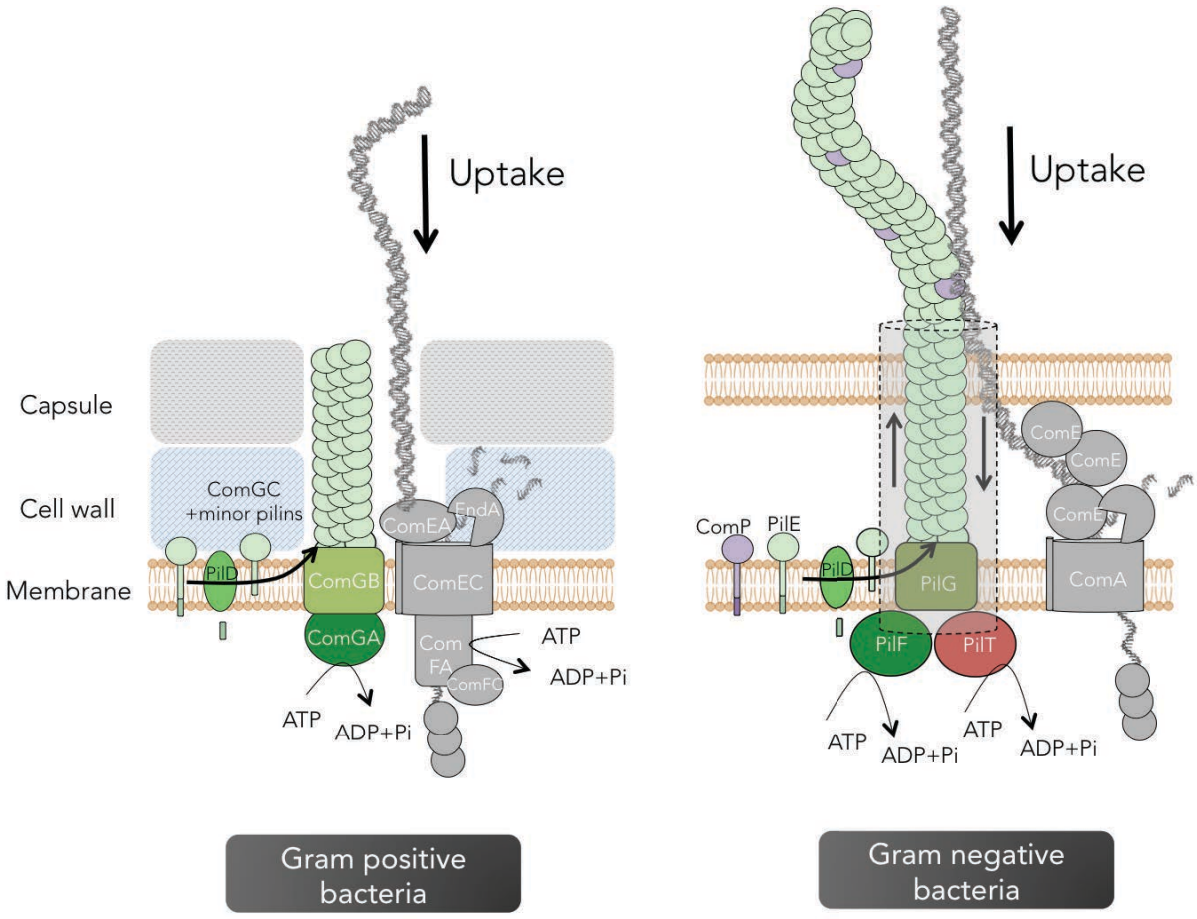


Figure 22 - T4P in DNA uptake

Schematic DNA uptake systems based on the Gram-positive bacterium B. subtilis and the Gram-negative bacterium N. meningitidis. Homologous proteins are shown in the same color. In N. meningitidis: PilE is the major pilin; ComP is one of the minor pilins; PilD is the prepilin peptidase; PilF is the assembly ATPase; PilT is the retraction ATPase; PilG is the integral membrane platform. Inner-outer channeling proteins are shown as a transparent grey cylinder. Proteins involved DNA translocation across the inner membrane are shown in grey.

As already said, T4P mediate DNA uptake in Gram-negative bacteria, as shown in *V. cholerae* (Seitz et al., 2013), *T. thermophilus* (Friedrich et al., 2002), and different Neisseiria species (Chen et al., 2004). In *N. meningitidis,* whole pili have been shown to bind DNA (Lang et al., 2009), although weakly, and it remains unclear if this binding is physiologically relevant. In *N. gonorrhoeae*, it was proposed from the high resolution structure obtained that a positive groove on the pilus surface could mediate this binding (Craig et al., 2006). It is only recently that a DNA binding receptor has been unambiguously identified in *N. meningitidis*: the minor pilin ComP (Cehovin et al., 2013). Indeed, the soluble domain of this minor pilin alone is able to bind DNA,

and preferentially DNA containing the DNA-uptake sequence, which all other pilins are unable to do under the same experimental conditions.

An electron microscopy structure of the assembly ATPase PilF in *T. thermophilus*, involved in pilus biogenesis and DNA uptake, was recently obtained. This protein has an unusually large N-terminal domain consisting of a 3 GspII domains repeat, while only one is present in most T2S assembly ATPases. An insertion mutant showed that this large N-terminal domain is not necessary for piliation, but essential for twitching motility (Friedrich et al., 2002), while deletion of the whole gene resulted in a piliation deficient strain, indicating that N-terminal and C-terminal domains can function independently (Salzer et al., 2014). Intriguingly, the large N-terminal domain was shown to bind DNA, and preferentially RNA, but it is not known yet if this interaction has a physiological role.

Whether an actual pilus is formed by *comG* operons in naturally competent Gram-positive bacteria remains an open question. As mentioned above, all distinctive proteins for T4P biogenesis are present, and seem essential for uptake (Chen et al., 2004; Bergé et al., 2002). It is an important question, considering that a wide variety of T4P operon have been found in Gram-positive bacterial genomes, without only a few of them clearly characterized (Imam et al., 2012; Rakotoarivonina et al., 2002; O'Connell Motherway et al., 2011).

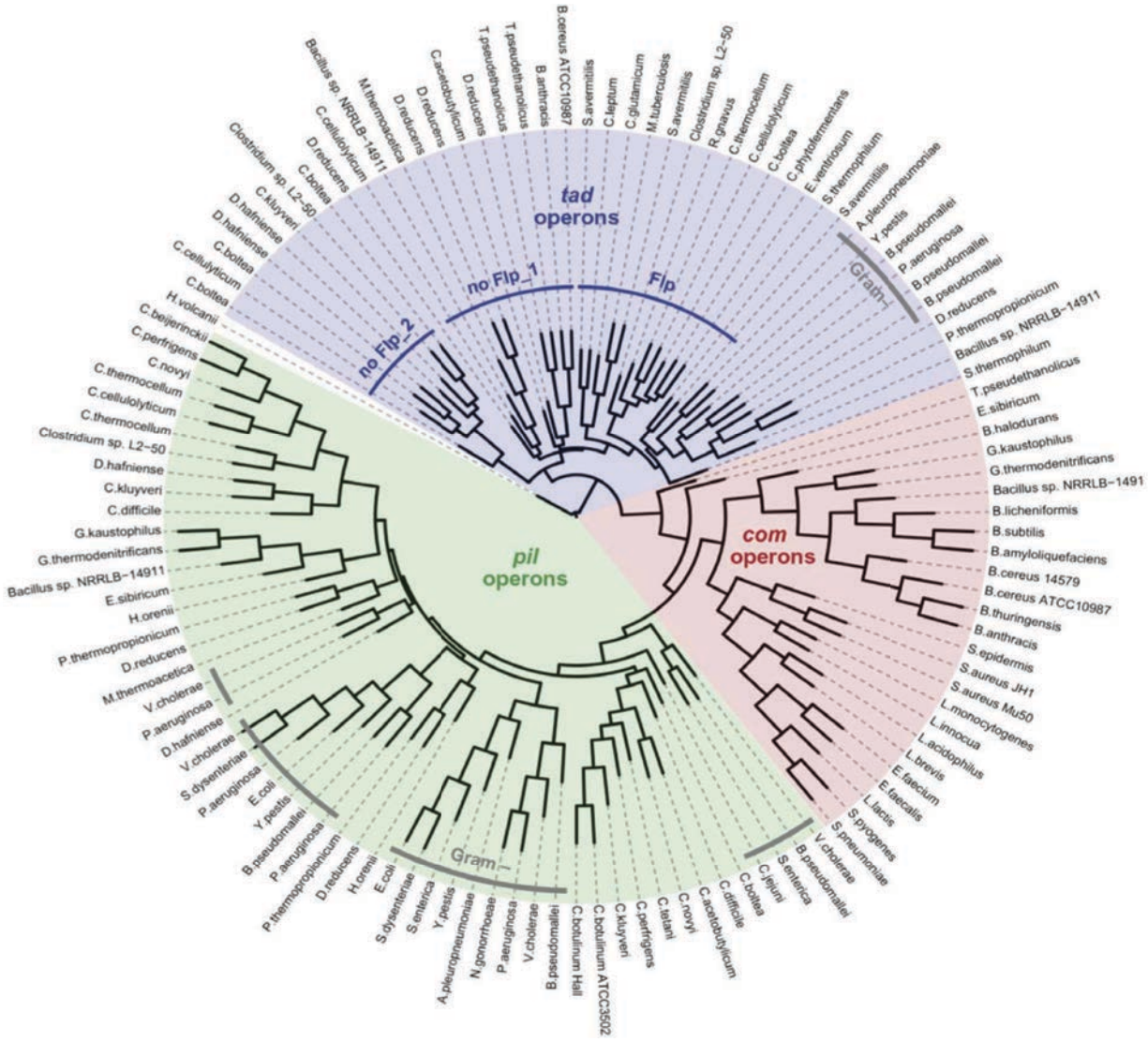


Figure 23 - Diversity of T4P operon found in Gram-positive bacteria
Figure taken from Imam et al., 2012. Phylogenetic tree based on the ComGA/TadA/PilF protein (assembly ATPase) using H. volcanii FlaI as an outgroup. Note that pil, com and tad operons form distinct clades, with Gram-negative bacteria grouped into clusters within each clade (displayed by a grey line).

In *B. Subtilis*, the ComGC pilin has been shown to multimerize, possibly in a filament structure. From the molecular weight of the released ComGC multimer, it was assessed they would measure 40 to 100 nm in a filamentous form. These results gave rise to the "transformation pseudopilus" hypothesis, stating that the *comG* operon is forming a very short filament, comparable to the T2SS pseudopilus, responsible for helping the DNA to cross the peptidoglycane layer (Chen et al., 2006). Minor pilins are necessary for ComGC multimerization in this bacterium, and they have been shown to interact directly with one another, and not with

ComGC (Mann et al., 2013). Additionally, the ComGG minor pilin was shown to have a special role: when affinity-purified, most of the « transformasome » proteins are co-purified with it; unlike other pilins it is not accessible to protease treatment in its non-processed form, suggesting that it faces the cytoplasm before processing; it is unstable in the absence of other *comG* proteins, and needs ComGA and ComGD-GF presence for its processing to occur (Mann et al., 2013).

D. Focus on the comG operon of S. pneumoniae

Pilins of the *S. pneumoniae comG* operon are partially annotated with the *cgl* nomenclature in R6 strain genome, but we will rather use the *com* nomenclature of *B. subtilis* for clarity.

Strain	comGA	comGB	comGC	comGD	comGE	comGF	comGG	pilD
R6	spr1864 /cglA	spr1863 /cglB	spr_1862 /cglC	spr_1861 /cglD	Not annotated	spr_1859	spr_1858	pilD/spr1628
D39	SPD_1863 /cglA	SPD_1862 /cglB	SPD_1861 /cglC	SPD_1860 /cglD	SPD_1859	SPD_1858	SPD_1857	SPD_1593

Table 1 - Clarification on genes nomenclature
The gene number and names in the R6 and D39 pneumococcal strains are shown under the name they will be given in this manuscript.

Except for ComGE, pilins are significantly positively-charged, with a high number of lysin residues, a feature that might have a relevant role in the uptake of DNA.

	comGC	comGD	comGE	comGF	comGG
Theoretical pI	8.15	9.64	5.78	8.83	9.30
aa content	Lys 15.5% Ala 12.7% Glu 11.3%	Ser 10.8% Lys 9.8% Gln 9.8%	Gln 13.6% Glu 13.6% Leu 10.6%	Gln 10.5% Glu 10.5% Lys 9.6%	Lys 18.5% Glu 13.9% Val 8.3%

Table 2 - pKa and amino acid (aa) content of pilins C-terminal domain

C-terminal domain is defined as starting right after the transmembrane helix predicted by the TMHMM software. The theoretical pI is obtained by the Protparam software. The 3 most abundant amino acids (aa content) for each pilin is displayed.

They all have the recognizable Type IV pilin characteristics, and when compared to the best studied set of pilins, *B. subtilis* competence and *E. coli* T2S, they show a similar organization. Based on this comparison, ComGC is supposedly the major pilin, followed by 4 minor pilins. ComGG has a degenerated cleavage motif, and could have like GspK and ComGG$^{B.\ subtilis}$ a special role.

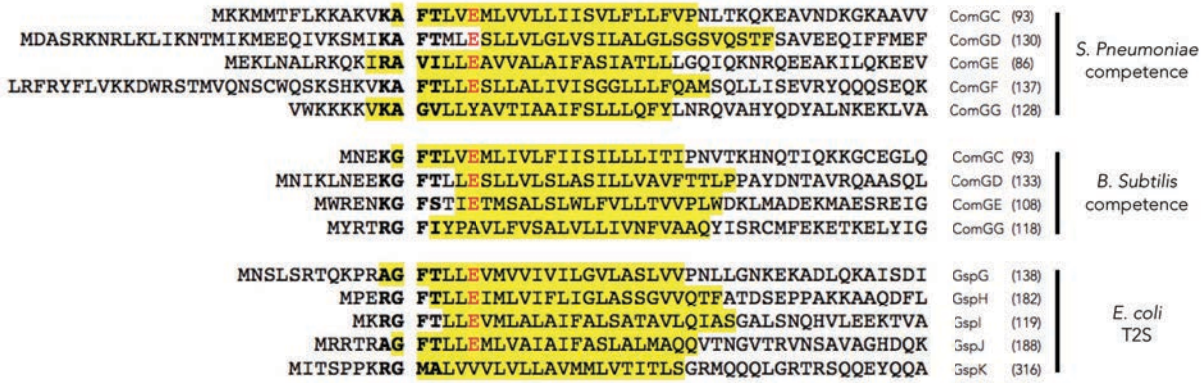


Figure 24 - Comparison of N-terminal sequences of S. pneumoniae pilins with B. subtilis competence pilins and enterotoxigenic E. coli T2S pilins

Transmembrane regions highlighted in yellow were determined using the TMHMM software on the full length pilin. The recognition motif of the prepilin peptidase is shown in bold letters, with a space indicating the cleavage site. The highly conserved Glu5 residue on the mature pilin is in red

ComGA is the only identified ATPase, supposedly powering assembly of the pilus. If retraction occurs during DNA uptake, like in Gram-negative bacteria, it will be by an unknown mechanism or an unidentified ATPase, since retraction has so far always been described only in the presence of a dedicated disassembly ATPase. The same question remains on T2S: a piston-like mechanism is currently the favored model, in which the pseudopilus would cyclically disassemble and assemble to push substrates through the secretin channel. However there is no retraction ATPase, and retraction has never been shown to occur.

In Gram-negative bacteria, a default in pilus retraction was shown to abolish DNA uptake in *N. gonorrhoeae* (Wolfgang et al., 1998); *P. sutzeri* (Graupner et al., 2001) or *V. cholerae* (Seitz et al., 2013). However, a recent study showed that *T. thermophilus* possesses 2 retraction ATPases, both of them essential for twitching motility, but not for natural transformation (Salzer et al., 2014). This result suggest that either retraction is not necessary for DNA uptake in this bacterium, or retraction can occur by a different mechanism, like a switching mode of the assembly ATPase.

VI. Aim and research strategy

First of all, I embarked on this PhD by addressing the outstanding question of whether or not a pilus is formed by the *comG* operon in *S. pneumoniae*. T4P operons being widespread in Gram-positive bacteria, barely described in this family, and the *comG* operon being a minimal T4P system, this question was of central importance. By mutagenesis, biochemical characterization, optical and electron microscopy techniques we were able to identify long, micrometer-sized appendages protruding from the surface of competent cells. We confirmed the T4P nature of these appendages, we showed that they bind DNA, and are absolutely

required for DNA uptake. We overthrew the pseudopilus hypothesis at least in *S. pneumoniae*, and provided crucial information concerning the initial step of DNA uptake.

Then, I studied the biochemical properties of the different *comG* subunits alone, and tried crystallize them in order to get some structural insights into the pilus architecture. Although no structure could be obtained, important interaction data were found.

Finally, I investigated in more detail the mechanism of DNA uptake by looking into a possible retraction of the pili by a wide range of techniques, and by studying the role of the minor pilins in pilus biogenesis and function.

Existence and characteristics of the transformation pilus in S. pneumoniae

I. Identification, biochemical characterization and visualization of the transformation pilus

Since all the work in this section has been published, it will be presented as the final edited publication.

Note: The strain *R1501, but comGC E20A* (point mutation of ComGC pilin) was not consistently annotated in the publication (different name in the figure and in the supplementary table). The glutamic acid residue in position 20 on full length ComGC was changed for an alanine and not a valine. It will be referred as the RL003 strain in this manuscript.

OPEN ACCESS Freely available online

PLOS | PATHOGENS

A Type IV Pilus Mediates DNA Binding during Natural Transformation in *Streptococcus pneumoniae*

Raphaël Laurenceau[1,2], Gérard Péhau-Arnaudet[2], Sonia Baconnais[3], Joseph Gault[2,4], Christian Malosse[2,4], Annick Dujeancourt[1,2], Nathalie Campo[5,6], Julia Chamot-Rooke[2,4], Eric Le Cam[3], Jean-Pierre Claverys[5,6], Rémi Fronzes[1,2]*

1 Institut Pasteur, Groupe Biologie Structurale de la Sécrétion Bactérienne, Paris, France, **2** CNRS, UMR3528, Paris, France, **3** Maintenance des Génomes et Microscopies Moléculaire, UMR 8126 CNRS-Université Paris Sud, Institut Gustave Roussy, Villejuif, France, **4** Institut Pasteur, Unité de Spectrométrie de Masse Structurale et Protéomique, Paris, France, **5** CNRS, UMR5100, Toulouse, France, **6** Université de Toulouse, UPS, Laboratoire de Microbiologie et Génétique Moléculaires, Toulouse, France

Abstract

Natural genetic transformation is widely distributed in bacteria and generally occurs during a genetically programmed differentiated state called competence. This process promotes genome plasticity and adaptability in Gram-negative and Gram-positive bacteria. Transformation requires the binding and internalization of exogenous DNA, the mechanisms of which are unclear. Here, we report the discovery of a transformation pilus at the surface of competent *Streptococcus pneumoniae* cells. This Type IV-like pilus, which is primarily composed of the ComGC pilin, is required for transformation. We provide evidence that it directly binds DNA and propose that the transformation pilus is the primary DNA receptor on the bacterial cell during transformation in *S. pneumoniae*. Being a central component of the transformation apparatus, the transformation pilus enables *S. pneumoniae*, a major Gram-positive human pathogen, to acquire resistance to antibiotics and to escape vaccines through the binding and incorporation of new genetic material.

Citation: Laurenceau R, Péhau-Arnaudet G, Baconnais S, Gault J, Malosse C, et al. (2013) A Type IV Pilus Mediates DNA Binding during Natural Transformation in *Streptococcus pneumoniae*. PLoS Pathog 9(6): e1003473. doi:10.1371/journal.ppat.1003473

Editor: Carlos Javier Orihuela, The University of Texas Health Science Center at San Antonio, United States of America

Received April 23, 2013; **Accepted** May 17, 2013; **Published** June 27, 2013

Funding: The Agence Nationale pour la Recherche, Institut Pasteur and the Centre National de la Recherche Scientifique have supported this work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: remi.fronzes@pasteur.fr

Introduction

Natural transformation, first discovered in *Streptococcus pneumoniae* [1], is observed in many Gram-negative and Gram-positive bacteria [2]. It increases bacterial adaptability by promoting genome plasticity through intra- and inter-species genetic exchange [3]. In *S. pneumoniae*, a major human pathogen responsible for severe diseases such as pneumonia, meningitis and septicemia, transformation is presumably responsible for capsular serotype switching and could therefore reduce the efficiency of capsule-based vaccines after a short period [4]. In this species, it occurs during a genetically programmed and differentiated state called competence that is briefly induced at the beginning of exponential growth. During this competent state, pneumococci secrete a peptide pheromone called Competence-Stimulating-Peptide (CSP) [5], which spreads competence in the pneumococcal population. Interestingly, in *S. pneumoniae*, some antibiotics and DNA-damaging agents induce competence, which would act as an alternative SOS response and ultimately increases bacterial resistance to external stresses [6].

During transformation, environmental DNA is bound at the surface of competent cells and transported through the cell envelope to the cytosolic compartment. This process has been mostly studied in the Gram-positive bacterium *Bacillus subtilis* with additional information coming from studies in *S. pneumoniae* [7,8]. In both species, a DNA translocation apparatus mediates the transfer of DNA through the cellular membrane. In *S. pneumoniae*, it is composed of ComEA, EndA, ComEC and ComFA. Incoming double-stranded DNA would bind the membrane receptor ComEA. One DNA strand crosses the membrane through ComEC while the endonuclease EndA degrades the other strand. On the cytoplasmic side, ComFA, an ATPase that contains a helicase-like domain, would facilitate DNA internalization through ComEC. Once inside the bacterium, single-stranded DNA is either integrated into the chromosome by RecA-mediated homologous recombination or entirely degraded.

Strikingly, all transformable Gram-positive bacteria also carry a *comG* operon that resembles operons encoding Type IV pili and Type II secretion pseudopili in Gram-negative bacteria, as well as a gene encoding a prepilin peptidase homolog, *pilD* [7]. In *B. subtilis* and *S. pneumoniae*, *comG* and *pilD* genes are exclusively expressed in competent cells and are essential for transformation [9,10,11]. In *S. pneumoniae*, the *comG* operon encodes a putative ATPase (ComGA), a polytopic membrane protein (ComGB) and five prepilin candidates named ComGC, ComGD, ComGE, ComGF and ComGG (Figure 1A and B and table S1). By homology with Type IV pili, it is generally proposed that these proteins could be involved in the assembly of a transformation pseudo-pilus at the surface of competent cells [7,8,12]. So far, two studies show that a large macromolecular complex containing ComGC can be found at the surface of competent *B. subtilis* cells [9,12]. In this complex, ComGC subunits appear to be linked together by disulfide bridges [9]. All the other ComG proteins and the PilD homolog, ComC, are necessary for the formation of this

Author Summary

Natural genetic transformation, first discovered in *Streptococcus pneumoniae* by Griffith in 1928, is observed in many Gram-negative and Gram-positive bacteria. This process promotes genome plasticity and adaptability. In particular, it enables many human pathogens such as *Streptococcus pneumoniae, Staphylococcus aureus* or *Neisseria gonorrhoeae* to acquire resistance to antibiotics and/or to escape vaccines through the binding and incorporation of new genetic material. While it is well established that this process requires the binding and internalization of external DNA, the molecular details of these steps are unknown. In this study, we discovered a new appendage at the surface of *S. pneumoniae* cells. We show that this appendage is similar in morphology and composition to appendages called Type IV pili commonly found in Gram-negative bacteria. We demonstrate that this new pneumococcal pilus is essential for transformation and that it directly binds DNA. We propose that the transformation pilus is an essential piece of the transformation apparatus by capturing exogenous DNA at the bacterial cell surface.

complex [9]. It was proposed that this complex could correspond to the transformation pseudo-pilus.

Despite these first clues, no transformation appendage could be directly visualized at the surface of any competent Gram-positive bacterium. In addition, the function of the ComG proteins during transformation remains unclear. Mutations in the cytosolic ComGA protein abolish DNA binding at the surface of both *B. subtilis* and *S. pneumoniae* [13,14,15]. This strongly suggests that the ComGC-containing macromolecular complex detected at the surface of competent *B. subtilis* cells could bind DNA. However, it was recently shown that ComGA is the only ComG protein essential to the initial DNA binding at the surface of competent *B. subtilis* cells [14]. This protein would interact with an unknown DNA receptor at the surface of competent cells while the other ComG proteins would only be required at a later stage during transformation.

In this study, we provide the first direct evidence for the existence of a transformation pilus in a Gram-positive bacterium. We discovered a new appendage at the surface of competent pneumococci that we could visualize using immuno-fluorescence and electron microscopy. Competent cells harbor one or a few appendages that are morphologically similar to Type IV pili found in Gram-negative bacteria. We were able to purify this pilus and showed that it is essentially composed of the ComGC pilin. We also demonstrate that pilus assembly is required for transformation. As we provide direct evidence that the transformation pilus binds extracellular DNA, we propose it is the primary DNA receptor at the surface of competent pneumococci.

Results

An extracellular appendage containing ComGC is detected at the surface of competent *S. pneumoniae*

Mechanical shearing is frequently used to release bacterial surface appendages and to study their protein composition [16]. To see if ComGC was part of a macromolecular complex at the surface of competent *S. pneumoniae*, we adapted the method to this bacterium and raised antibodies against the purified soluble domain of ComGC. Using this antibody, we showed that ComGC could be detected by immunoblotting in the sheared fraction of competent bacteria (Figure 2A). While ComGC level in the cell fraction was not affected, no ComGC could be found in the sheared fraction in a *comGA* knockout mutant (Figure 2A). These data strongly suggest that ComGC is part of an extra-cellular appendage and that ComGA is necessary to its assembly.

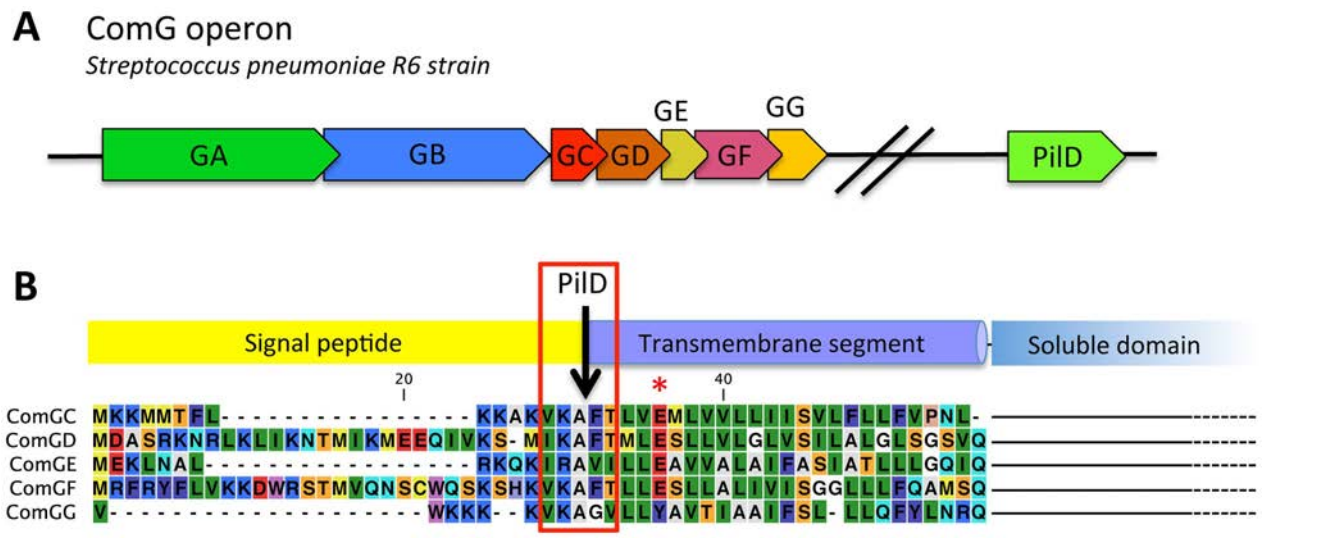


Figure 1. Genes potentially essential for transformation pilus assembly in *S. pneumonia and* prepilin sequences. (**A**) Genes that are potentially essential in transformation pilus assembly. The *comG* operon contains 7 genes named *comGA* to *comGG*. A gene that encodes a pre-pilin peptidase homolog, *pilD*, is found elsewhere on the chromosome. The name used to designate the *comG* and *pilD* genes varies in different pneumoccocal strains. For clarity, we refer to the comG nomenclature used in B. subtilis. Names of the corresponding genes in different S. pneumoniae strains are found in the Table S1. (**B**) Alignment of the N-terminal extremity of the five possible prepilins show that ComGC, ComGD, ComGE and comGF contain a canonical prepilin cleavage motif (red rectangle) [20]. All four proteins also contain a conserved glutamic acid residue in position 5 after the cleavage site (red asterisk), a conserved feature of all Type IV pilins [43]. ComGG has a degenerated peptidase motif and does not contain a conserved glutamic acid in position 5. Secondary structure predictions for the five proteins propose a N-terminal hydrophobic helix and a C-terminal soluble domain of variable size (not shown in the figure), the typical topology found in Type IV pilins.
doi:10.1371/journal.ppat.1003473.g001

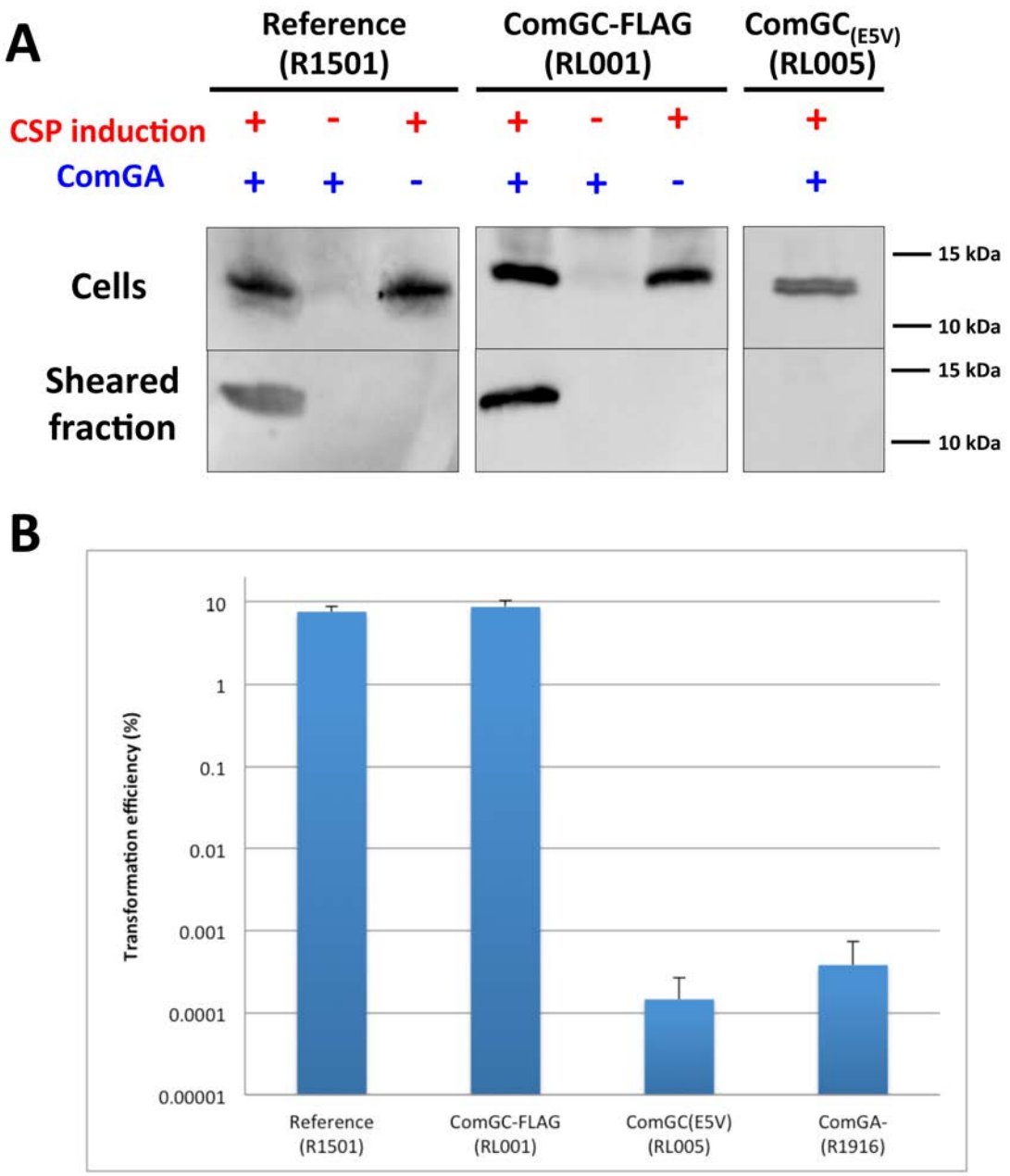


Figure 2. Competence-induced appendages assembly and transformation efficiency. (**A**) Detection of ComGC in the sheared and cellular fractions by immunoblot. ComGC was detected in the sheared fraction of reference (R1501) and ComGC-FLAG (RL001) competent strains. ComGC was not detected in the sheared fraction of cultures that were not competence-induced. Deletion of *comGA* completely abolished detection of the ComGC pilin in the sheared fraction in both reference and ComGC-FLAG strains. Substitution of the conserved glutamic acid in position 5 of the mature ComGC pilin (Figure 1) by alanine ($ComGC_{(E5V)}$) also completely abolished detection of ComGC in the sheared fraction. For the reference and $ComGC_{(E5V)}$ strains, detection of ComGC was performed with a polyclonal rabbit antibody raised against the soluble domain of ComGC. With the ComGC-FLAG strain, the FLAG-tagged pilin was detected by an anti-FLAG monoclonal antibody. (**B**) Transformation assay using the reference strain (R1501), comGC-FLAG strain (RL001), $comGC_{(E5V)}$ mutant (RL005) and *comGA* deletion mutant (R1916). Both $comGC_{(E5V)}$ and *comGA* mutants were defective for transformation.
doi:10.1371/journal.ppat.1003473.g002

A ComGC-containing appendage is visualized at the surface of competent pneumococci

We inserted a FLAG tag at the C-terminus of ComGC to directly visualize the competence-induced appendages by immuno-fluorescence. It was not possible to insert the sequence encoding the tag at the *comGC* locus on the chromosome because *comGC* and *comGD* genes overlap in the *comG* operon. Therefore, a copy of *comGC* encoding a C-terminally FLAG-tagged ComGC (ComGC-FLAG) was integrated ectopically into the chromosome of *S. pneumoniae* under the control of a competence-induced promoter [17]. The transformation efficiency was not affected in this strain (Figure 2B). Using anti-FLAG antibodies, we could show by immuno-fluorescence that almost all the cells appeared to harbour one or a few ComGC foci or distinct fluorescent appendages (Figure 3A and B; Figure S1A). Due to sample

preparation, many broken appendages were also found in the background. No preferential location of the foci/appendages at the cell surface was observed. They are absent in *comGA* knockout cells (Figure 3A). Note that anti-ComGC antibodies were not able to label the competent cells. They probably recognize epitopes that are masked when ComGC is included in the appendages.

Competence-induced appendage morphology and composition

Using electron microscopy, we observed filaments attached to the cell surface of negatively stained competent pneumococci (Figure 4A and B). These flexible filaments are 5–6 nm in diameter. Their length could reach up to 2–3 micrometers (Figure 4A). A maximum of 2–3 filaments per cell could be observed. Their average length was difficult to assess because they break easily into smaller fragments during sample preparation. Using the ComGC-FLAG expressing strain, we confirmed by immunogold-labelling that they contain ComGC (Figure 4C).

Appendages were then purified using anti-FLAG affinity chromatography after mechanical shearing. Appendage fragments of between 50 and 500 nm in length were observed by electron microscopy (Figure 5A), showing that these filamentous structures do not disassemble during purification. SDS-PAGE analysis of the purified fraction showed that ComGC is the major component of the appendages (Figure 5B). Using whole protein mass profiling by high-resolution mass spectrometry [18], we could only detect ComGC and ComGC-FLAG in the purified material (Figure 5C), confirming that ComGC is the major constituent of these appendages. Monoisotopic mass measurements of intact proteins and top-down fragmentation using a variety of activation techniques confirmed that the ComGC prepilin is cleaved after the alanine residue in position 15 and that the first amino acid of the mature protein is methylated, presumably by PilD (Figure S2). Indeed, PilD homologs in Gram-negative bacteria catalyze this post-translational modification of the Type IV pilins [19]. No other post-translational modification was detected in ComGC. Other proteins, including other ComG proteins, were not detected in the purified material by the methods used in this study. This suggests that these proteins are either absent, present in very low amount within the appendage or weakly bound to it and lost during sample preparation. These morphological and biochemical features are typical of Gram-negative Type IV pili. Therefore, we propose that the competence-induced appendage observed in *S. pneumoniae* belongs to the Type IV pilus family.

The competence-induced pili are required for transformation

It was important to determine whether these competence-induced pili were involved in transformation. Indeed, it was previously shown that *S. pneumoniae* and *B. subtilis comGA* knockout could not be transformed (Figure 2B) [9] [13]. In this study, we were able to show in *S. pneumoniae* that *comGA* mutant cells lack pili (Figure 2A and 3A). It was enticing to conclude that competence-induced pili assembly is essential for transformation. However, it was recently shown that a *comGA* mutation could have a pleiotropic effect on transformation in *B. subtilis* [14]. Therefore, we generated a *comGC* mutant in *S. pneumoniae* in which the conserved glutamic acid in position 5 was substituted by an alanine (Figure 1B). Such a substitution was shown to impair Type IV pilus assembly in Gram-negative bacteria [20]. ComGC cellular level was not affected by this point mutation (Figure 2A). Our results show that this mutant strain could not assemble any pilus and that it was defective for transformation (Figure 2A and B). Therefore we conclude from the analysis of both *comGA* and $comGC_{(E5A)}$ mutants that the assembly of the competence-induced pilus is required for transformation.

Transformation pili bind extracellular DNA

The nature of the primary DNA receptor at the surface of transformable Gram-positive bacteria is not known. It is generally proposed that the transformation pseudopilus would bind extracellular DNA at the surface of competent Gram-positive bacteria [8,21]. However, this hypothesis has never been confirmed experimentally. Using affinity purification, we show that DNA naturally released in the culture medium co-fractionates with the purified pili. No DNA could be found in the purified fraction in absence of the pilus (Figure 6A). These data were a first hint suggesting that DNA present in the environment could bind to the transformation pilus. However, it was not clear if this binding was related to the transformation process or fortuitous. By using specific electron microscopy methods [22], we visualized DNA directly bound to the transformation appendage after adding linear double stranded DNA (dsDNA) to competent bacteria. Long stretches of dsDNA interacting with the transformation pilus were observed with clearly visible multiple contact points (Figure 6 B–E). Interestingly, it was extremely difficult to see DNA bound on the pilus in the reference bacteria (R1501 strain), which are known to internalize exogenous DNA quickly [23]. On the other hand, in *ComEC* and *comFA* mutants, we could easily observe bound DNA on transformation pili. These strains are defective for DNA uptake and accumulate bound DNA at their surface [13]. Given that the dsDNA was added in large excess, no difference between the reference and mutant strains should be observed if DNA binding on the pili was a coincidental event. The fact that the uncoupling of DNA binding and uptake processes facilitates the observation of the DNA/pilus interaction is a strong indication that DNA binding on the transformation pilus is related to the transformation process.

Discussion

The *comG* operon is conserved in all transformable Gram-positive bacteria. This operon encodes proteins that are homologous to proteins involved in Type IV pilus assembly in Gram-negative bacteria. Therefore it has been proposed that a pilus (or pseudopilus) could be assembled at the surface of competent Gram-positive bacteria. Since all *comG* genes are essential for transformation, this pilus could be directly involved in transformation. The first biochemical clues for the existence of a transformation pilus were found in *B. subtilis* although decisive observational support was lacking. In addition, it was not clear if the ComGC-containing macromolecular complex found in *B. subtilis* was a common feature of competent Gram-positive bacteria or specific to this species. Finally, the function of this putative transformation pilus, and in general of the ComG proteins, was unclear.

Discovery of a new pneumococcal appendage

The pneumococcal transformation pilus represents a newly discovered pneumococcal surface structure. For a long time, no external appendage could be found at the surface of *S. pneumoniae* cells while many electron microscopy images were published in the literature. Recently, sortase-mediated pili have been discovered in some pathogenic *S. pneumoniae* strains [24]. To our knowledge, no specific ultrastructural study of competent *S. pneumoniae* has ever been described. Here, we analysed a laboratory strain that is

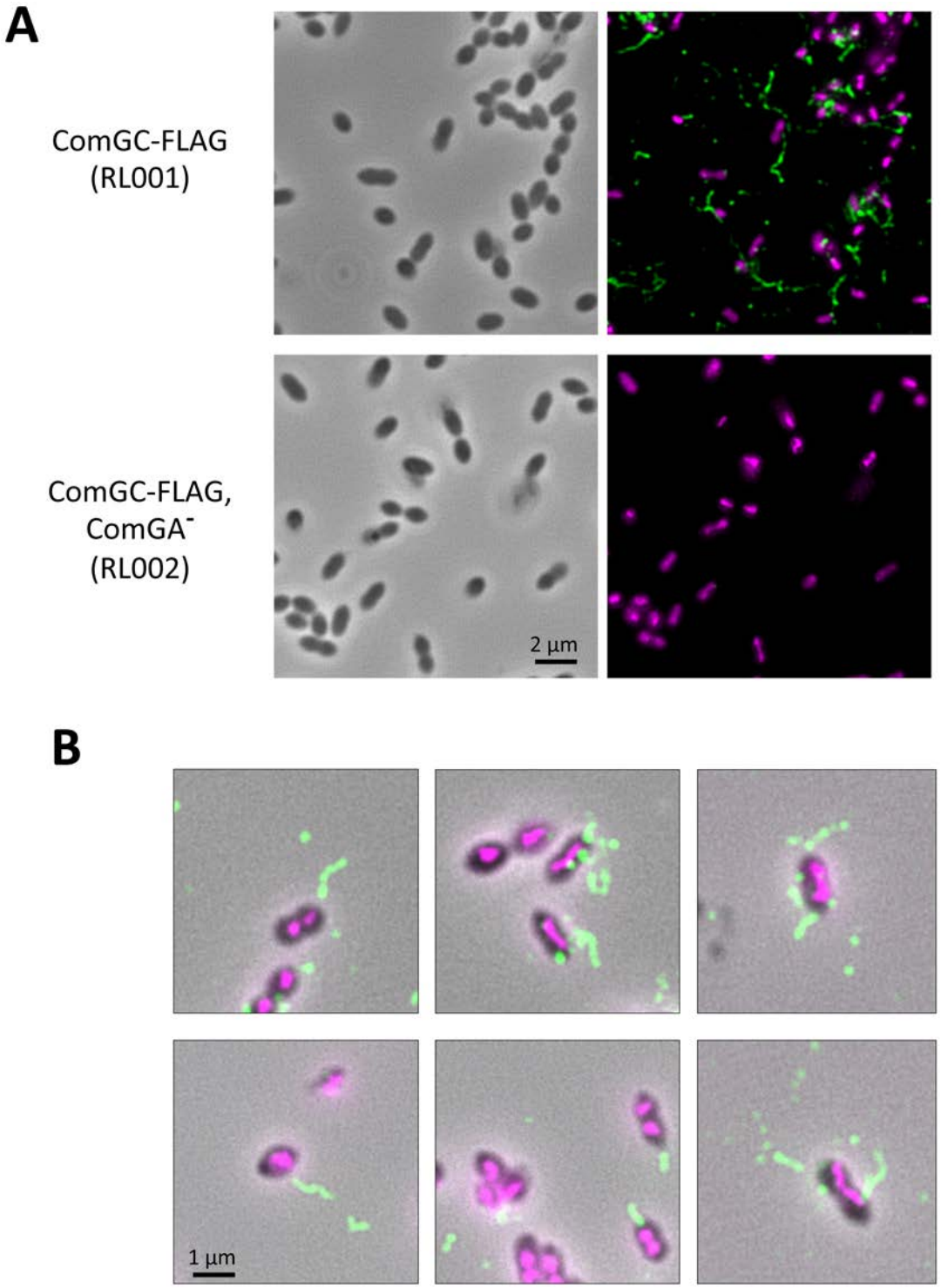


Figure 3. Direct visualization of competence-induced appendages by Immuno-fluorescence. (**A**) Immuno-fluorescence microscopy showing intact competent cells expressing a FLAG-tagged ComGC pilin in presence (top row) or absence of ComGA (bottom row). Left column correspond to bright field image, right column to overlay between anti-FLAG antibody fluorescence (green) and DAPI fluorescence (magenta). (**B**) Zoom on several bacterial cells visualized by immuno-fluorescence. Overlay of bright filed image, anti-FLAG antibody fluorescence (green) and DAPI fluorescence (magenta). Distinct appendages are visible on competent cells.
doi:10.1371/journal.ppat.1003473.g003

commonly used to study the transformation process in *S. pneumoniae* [10] [13]. In this strain, competence can be induced in a rapid and synchronous manner upon addition of synthetic CSP in the medium of an exponentially growing culture [5,25]. To make sure that the appearance of the transformation pilus is a common feature of competent pneumococci and not a mere one-

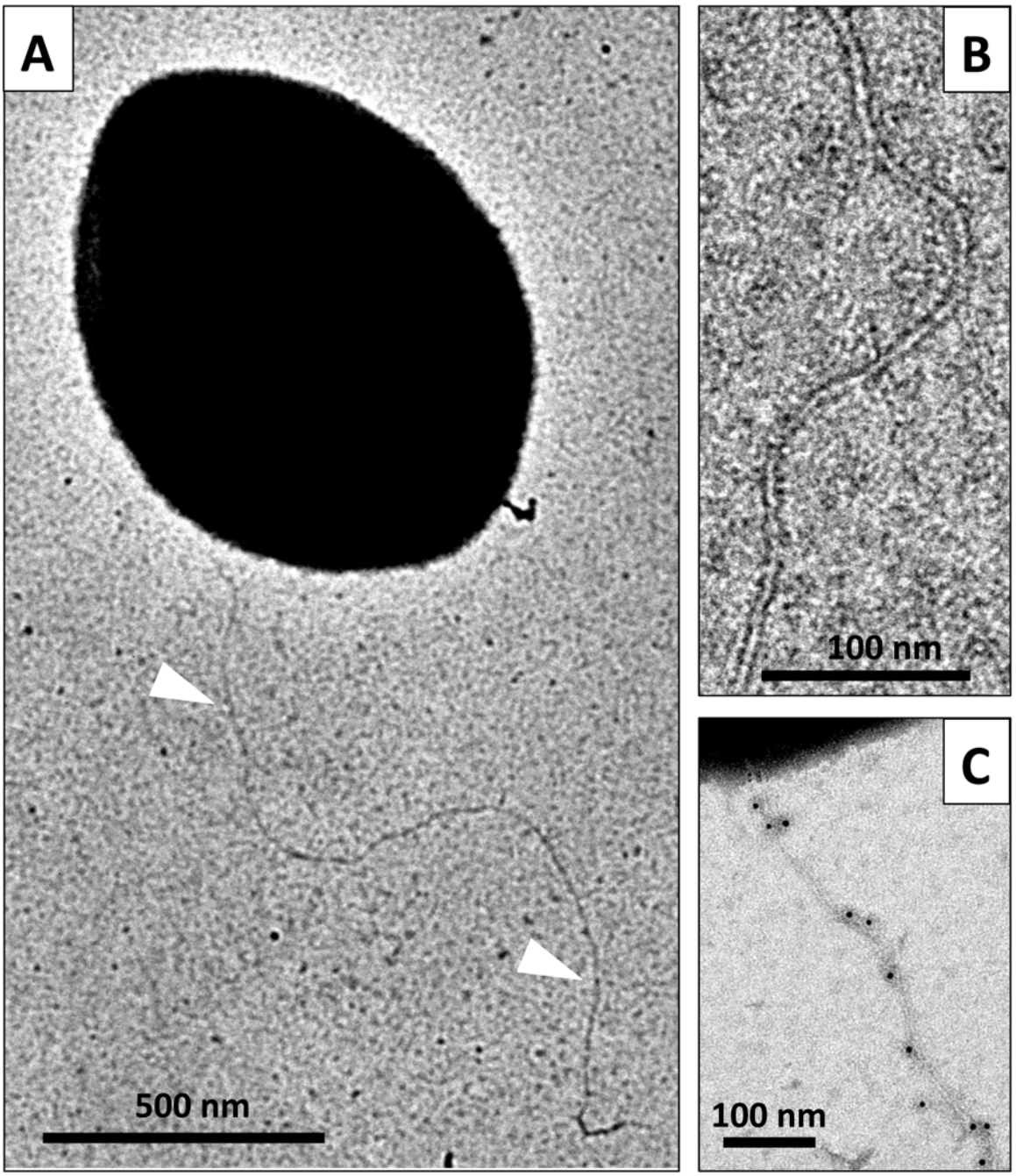


Figure 4. Direct visualization of the competence-induced appendage. Competent reference bacteria observed by transmission electron microscopy. Single appendages, 5–6 nm wide, were observed at the surface most of the competent cells in the culture. Many long filaments were observed, reaching up to several micrometers in length. (**A**) a competent *S. pneumoniae* cell with a long pilus (white triangle). (**B**) closer view of a transformation pilus. (**C**) A pilus observed by transmission electron microscopy after immunogold labeling with anti-FLAG antibody (5 nm gold beads) using the ComGC-FLAG strain. ComGC-FLAG proteins are detected within the appendages.
doi:10.1371/journal.ppat.1003473.g004

off property of our reference strain, we observed negatively stained G54 and CP strains by electron microscopy. The G54 strain is a wild-type clinical strain. The CP strain is a laboratory strain that has a different genetic background than our reference strain [26]. In both cases, transformation pili were observed at the surface of competent cells (Figure S3). Therefore, we think that transformation pili are found at the surface of most, if not all, pneumoccocal strains, including clinical strains.

The transformation pilus is a Type IV pilus

The pneumococcal transformation pilus is morphologically very similar to Type IV pili found in many Gram-negative bacteria. Its major component, the ComGC pilin, is cleaved and probably methylated by a PilD homolog. We therefore propose that the transformation pilus is a *bona fide* Type IV pilus. Since its length can reach up to 2–3 μm, we think that the "pseudo-pilus" appellation does not apply to the pneumococcal transformation

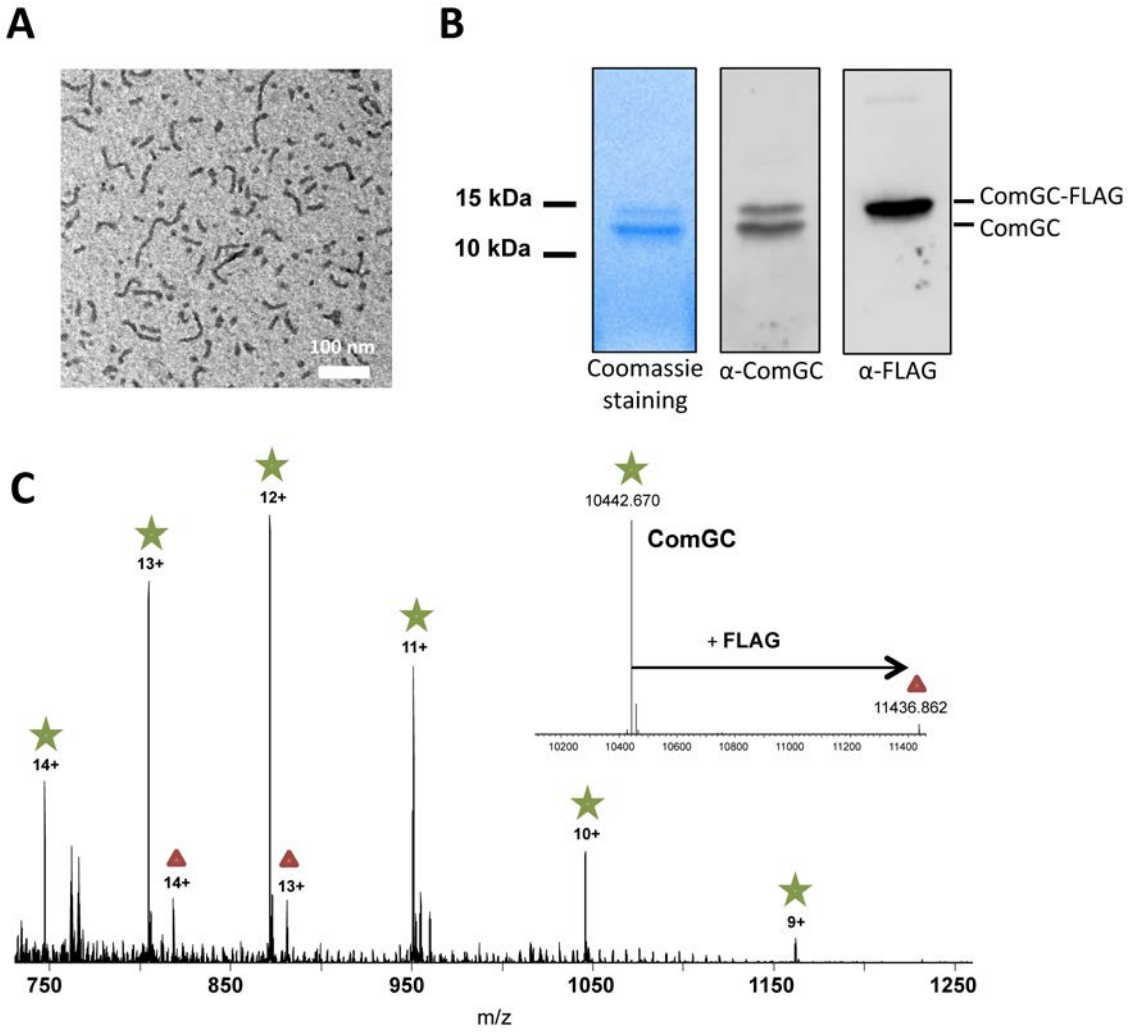


Figure 5. Nature of the transformation pilus. (**A**) Purified pili visualized by negative stain electron microscopy. Short pilus fragments ranging from 50 to 500 nm were observed. (**B**) SDS-PAGE analysis of the purified fraction. Left lane, Coommassie blue staining. Middle lane, Western blot using anti-ComGC antibody. Right lane, Western blot analysis using anti-FLAG analysis. (**C**) Nano-ESI FT-MS spectrum of purified pili. Peaks corresponding to ComGC are labelled with a green star those corresponding to ComGC-FLAG with a red triangle. ComGC-FLAG is present at approximately 6+/−2% abundance of the native form. The deconvoluted spectrum showing monoisotopic masses of the neutral protein forms is presented in the inset. The measured mass of methylated comGC (10,442.670 Da) compares very well to the calculated theoretical mass (10,442.636 Da) with an error of +3 ppm.
doi:10.1371/journal.ppat.1003473.g005

appendage. By comparison, the type II secretion pseudo-pilus is just 50–100 nm long [27]. The transformation pilus is the first Type IV pilus clearly observed in a Gram-positive bacterium. So far, Type IV pilus-dependent gliding motility had been described in *Clostridium* species [28]. However, no clear picture of this pilus was provided. A recent genomic study show the existence of numerous and diverse Type IV pilus-like operons in a wide range of Gram-positive bacteria [29]. This suggests that many other Type IV-like pili remain to be discovered in these bacteria. The conservation of *comG* operons argues in favor of the presence of a transformation pilus in all naturally transformable Gram-positive bacteria. However, species-specific variations in pilus length can be anticipated because of variations in thickness of the capsule and/or the cell wall.

Pilus function. We envision the transformation pilus to act as a ''DNA-trap'' to capture DNA in the environment. In Gram-negative bacteria, Type IV pilus assembly is also essential for natural transformation [30]. It is proposed that this pilus interacts directly with DNA but the molecular details of this interaction remains enigmatic [31,32,33]. No DNA binding protein could be identified in the pilus [34]. Recently a minor pilin, ComP, has been identified as a DNA receptor specific of genus-specific DNA uptake sequence (DUS) motifs in *Neisseria meningitidis* [35]. Our data explicitly show that the pneumoccocal transformation pilus binds DNA. At this stage, it is not clear if this DNA binding ability is due to the physicochemical properties of the pilus and/or due to a yet undetected minor pilin. Interestingly, only ComGA homolog was found indispensable for initial DNA binding at the surface of *B. subtilis* [14]. These bacteria could assemble only short transformation pili that are not sufficient to bind DNA at the surface of the competent cells. An unknown DNA receptor that interacts with ComGA ATPase could be required for efficient DNA binding in this species. It is possible that the mechanism of initial DNA binding at the surface of competent Gram-positive bacteria vary.

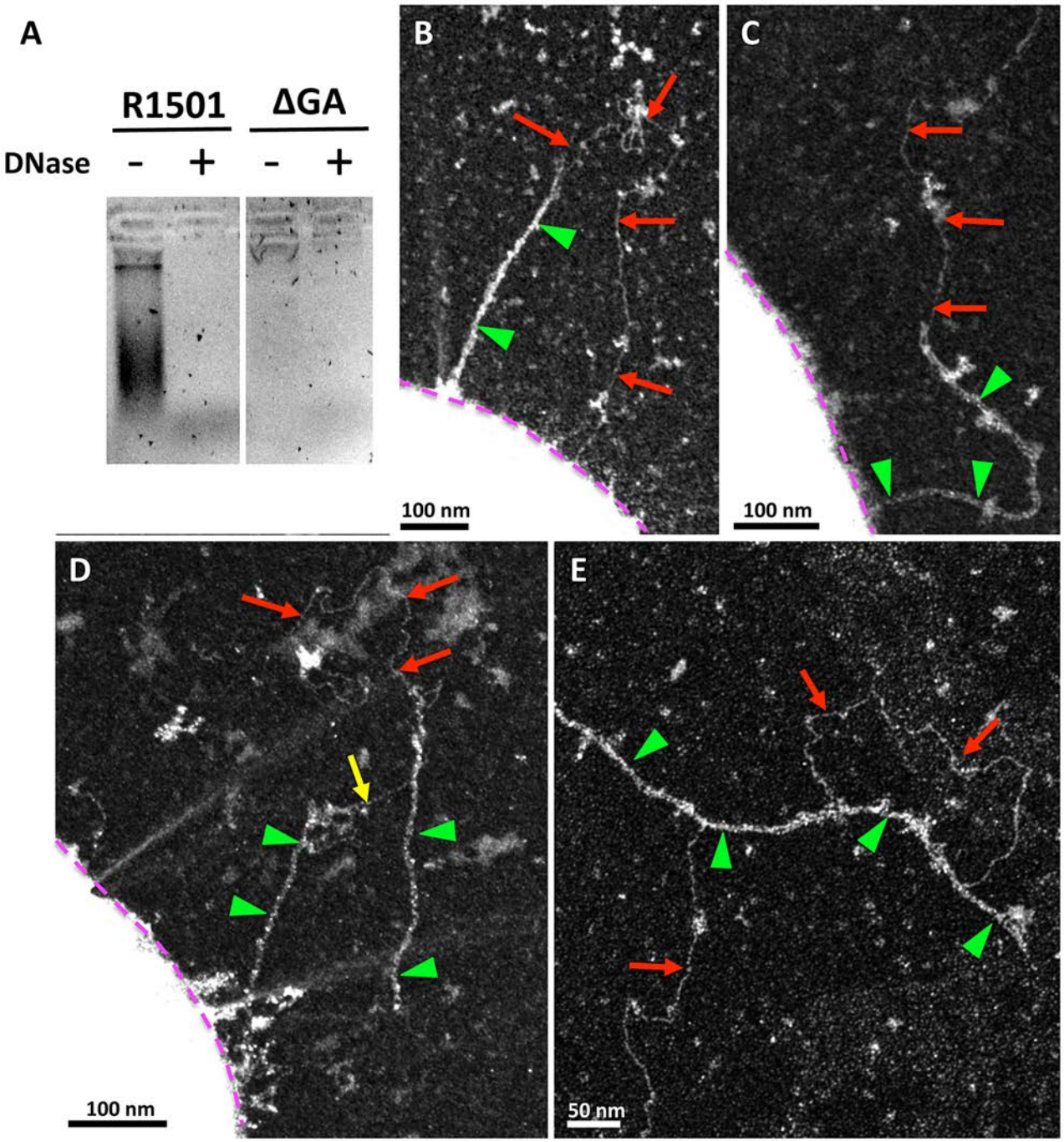


Figure 6. Interactions between transformation pili and DNA. (**A**) DNA co-purified with the pili by affinity chromatography was revealed after migration on agarose gel and Gel Green staining. DNA was present in the ComGC-FLAG strain. In the absence of ComGA, DNA was not detected. DNase treatment unambiguously showed that the bands detected on the gel are DNA. Linear DNA (red arrow) bound to the pilus (green triangle) in *ΔcomFA* (**B**) and *ΔcomEC* (**C**), (**D**) and (**E**) strains. (**B**) and (**C**), linear DNA interacting with a single pilus. The DNA molecule contacts the pilus at several points. **D**, a DNA molecule maintains a broken pilus close to the bacterium (yellow arrow), demonstrating the existence of several DNA binding sites on the pilus. (**E**) Detached pilus found in the medium with bound DNA. In (**B**), (**C**) and (**D**), bacterial cell envelope is highlighted by a purple dotted line.
doi:10.1371/journal.ppat.1003473.g006

Pilus assembly and retraction. In Gram-negative bacteria, 12 to 15 proteins are necessary for Type IV pilus biogenesis [36,37]. The assembly of the pneumococcal transformation pilus could require the simplest Type IV pilus assembly apparatus discovered so far in bacteria. Indeed, only 7 ComG proteins might be sufficient to assemble this pilus. Strikingly, Type IV pili are retractile appendages in Gram-negative bacteria [38]. It is therefore appealing to propose that the transformation pilus could retract to guide bound DNA through the cell wall and the polysaccharide capsule that is often present in bacteria (as in *S. pneumoniae*), and, ultimately deliver it to the ComEA/EndA complex located in the membrane [7]. It is unlikely that this

DNA would find its way along several hundreds of nanometers through the cell wall and capsule without pilus retraction. However, a dedicated ATPase called PilT powers pilus retraction in Gram-negative bacteria and *S. pneumoniae* genome does not encode any PilT homolog. All the *comG* operons that have been identified so far in Gram-positive bacteria encode only one ATPase, ComGA, that is required for the pilus assembly. Therefore, the possible retraction of the transformation pilus and role of ComGA in this process should be assessed.

Our data clearly establish the existence and function of a transformation pilus at the surface of competent pneumococci. As an essential component of the transformation apparatus, it enables this major human Gram-positive pathogen to acquire resistance to antibiotics and to escape vaccines through the binding and incorporation of new genetic material. Future work should establish whether transformation pili exist and play similar roles in other transformable pathogens. Intriguingly, ComG proteins also appear to play an important and direct role in phagosomal escape and virulence in *Listeria monocytogenes* [39]. It would be interesting to determine whether they also assemble into a pilus to play this role.

Materials and Methods

Strain construction and growth

Cells were grown at 37°C under anaerobic condition, without agitation, in a Casamino Acid Tryptone medium (CAT) up to $OD_{600} = 0.3$ for stock cultures [40]. After addition of 15% glycerol, stocks were kept frozen at −80°C. For competence induction, cells were grown in CAT supplemented with BSA (2 g/L), calcium chloride (1 mM) and adjusted to pH = 7.8. Competence was triggered by incubating cells with the Competence Stimulating Peptide (CSP) at $OD_{600} = 0.1$ for 12 min as described previously [40]. For transformation, DNA was then added and transformants were selected on CAT agar plates [17]. Competence was induced following the same protocol in G54 and TCP1251 strains.

For transformation efficiency assays, 100 µL of competent bacteria were transformed by the addition of 100 ng of *S. pneumoniae* R304 genomic DNA (contains the streptomycin resistance gene *str41*). Bacteria were plated in presence and absence of streptomycin (100 µg/mL final concentration) and incubated at 37°C overnight before colony counting.

The annotated names of the *comG* genes in different strains of *S. pneumoniae* are listed in Table S1. The *S. pneumoniae* strains used derived from the non-capsulated R6 strain and are listed in Table S2. The *comGC-FLAG* gene was cloned by PCR using genomic DNA of pneumococcal R6 strain (ATCC BAA-255) as template. The resulting fragment was digested with *NcoI* and *BamHI* and inserted into the same sites of the pCEPx vector [17]. RL001 strain was constructed by transformation of R1501 cells with the pCEPx plasmid containing *comGC-FLAG*, followed by selection with kanamycin (Kan). RL002 was obtained by transformation of RL001 with R1062 chromosomal DNA and selection with spectinomycin (Spc). For RL003, a 2 kb genomic fragment of R6 genome containing *comGC* in the middle was amplified, and the codon 20 was changed from GAG to GTG by cross-over PCR. R1501 was transformed with this modified genomic fragment, and clones were screened by sequencing the *comGC* gene.

Chemically competent *Escherichia coli* BL21 Star (Life Technologies) were used for heterologous production of ComGC soluble domain. The corresponding DNA sequence was amplified from genomic DNA of strain R800 and cloned into pET15b expression vector (Novagen), using *NdeI/XhoI*. The protein was purified from the soluble fraction using IMAC affinity and gel filtration in 50 mM Tris/HCl pH = 8, 200 mM NaCl. The anti-ComGC were raised against the purified protein (Eurogentec).

Detection and purification of cell surface appendages

Shearing experiments were adapted from Sauvonnet et al. [41]. Competence was induced exactly as described above in a 50 mL culture. Cells were harvested by centrifugation 15 min at 4,500 g, 4°C. The pellet was suspended in 1 mL LB and immediately vortexed for 1 min to apply mechanical pressure. The suspension was then centrifuged twice at 13,000 g for 5 min to separate the bacteria (pellet fraction) from the pilus-enriched supernatant (sheared fraction). The supernatant was then precipitated with 10% trichloroacetic acid for 30 min on ice. Both fractions were loaded on SDS 15% polyacrylamide gels and subjected to electrophoresis and immunoblot with rabbit polyclonal antibodies raised against ComGC soluble domain (38–108) or anti-FLAG M2 antibody (Sigma-Aldrich F1804).

The pili containing ComGC-FLAG were purified from the sheared fraction of a 1 L culture. Shearing was performed in 2 mL Tris Buffered Saline (TBS, Tris pH 7.6 0.05 M, NaCl 0.15 M, protease inhibitor cocktail Roche 11873580001) and incubated overnight on a rotating wheel at 4°C with ANTI-FLAG M2 affinity resin (Sigma-Aldrich A2220). After washing with TBS, the pili were eluted by adding 3×FLAG-peptide at 100 µg/mL (Sigma Aldrich F4799) 30 min at room temperature under agitation.

To prevent DNA aspecific binding on the ANTI-FLAG M2 affinity resin, the resin was saturated 2 h at 4°C with a 1.5 kb PCR fragment (20 ng/µL). For DNA detection, 20 µL of the eluted pili were run on a 1% agarose gel and stained with SYBR safe (Life technologies S33102).

Visualization of pili and immunogold labelling

Competence was induced exactly as described above in a 10 mL culture. Cells were harvested by centrifugation 15 min at 4,500 g, 4°C. The pellet was suspended in 60 µL phosphate-buffered saline (PBS) (Sigma-Aldrich P4417). A drop of this suspension was placed on a glow discharged carbon coated grid (EMS, USA) for 1 min. The grid was then placed on a drop of PBS-3% formaldehyde, 0.2% glutaraldehyde for 10 min, and washed on drops of distilled water. The grids were then treated with 2% uranyl acetate in water. Specimens were examined using a Philips CM12 transmission electron microscope operated at 120 kV. Pictures were recorded using a camera KeenView (SIS, Germany) and ITEM software. For immunogold labelling, additional steps were applied after fixation: 3 washes with PBS, PBS–50 mM NH_4Cl (10 min), 3 washes with PBS, PBS with 1% BSA (5 min), 1 hour incubation with ANTI-FLAG M2 antibody (Sigma-Aldrich F1804) diluted 1/100 in PBS with 1% BSA, 3 washes with PBS-BSA 1% (5 min), 1 hour incubation with goat anti-mouse antibody (5 nm gold particles, BritishBioCell, UK) diluted 1/25 in PBS containing 1% BSA.

Fluorescence microscopy

S. pneumoniae cells were grown in the same conditions as above for visualization by electron microscopy. Cells were harvested by centrifugation for 15 min at 4,500 g, 4°C. The pellet was suspended in 500 µL PBS and directly immobilized on poly-L-lysine-coated coverslips. Samples were fixed for 30 min with 3.7% formaldehyde, washed 3 times with PBS containing 1% BSA and incubated on a 100 µL drop of anti-FLAG antibodies (1:300) and secondary Alexa Fluor 488- coupled anti-mouse IgG (Invitrogen). Samples were examined with an Axio Imager.A2 microscope

(Zeiss). Images were taken with AxioVision (Zeiss) and processed in ImageJ [42].

Mass spectrometry

Protein samples were desalted and eluted directly into a 10 µL spray solution of methanol:water:formic acid (75:25:3). Approximately 4 µL was loaded into a coated, medium sized, nano-ESI capillary (Proxeon) and introduced into an Orbitrap Velos mass spectrometer, equipped with ETD module (Thermo Fisher Scientific, Bremen, Germany) using the off-line nanospray source in positive ion mode. A full set of automated positive ion calibrations was performed immediately prior to mass measurement. The transfer capillary temperature was lowered to 100°C, sheath and axillary gasses switched off and source transfer parameters optimised using the auto tune feature. Helium was used as the collision gas in the linear ion trap. For MSn experiments, ions were selected with a 3 Da window and both CID and HCD were performed at normalised collision energies of 15–25%, with the appropriate HCD charge state set and other activation parameters left as default. For ETD the reagent gas was fluoranthene and the interaction time 10 ms. Supplemental activation was used as noted. The FT automatic gain control (AGC) was set at 1×10^6 for MS and 2×10^5 for MSn experiments. Spectra were acquired in the FTMS over several minutes with one microscan and a resolution of 60,000 @ m/z 400 before being summed using Qualbrowser in Thermo Xcalibur 2.1. Summed spectra were then deconvoluted using Xtract and a, b, c−1, y, z, z+1 ions assigned using in house software at a tolerance of 5 ppm. N-terminal ions were verified manually.

Positive staining electron microscopy

Five microliters of bacterial culture (wild-type, *ΔcomFA* or *ΔcomEC*) were diluted in 45 µL of Tris 10 mM, pH 8, NaCl 150 mM. Bacteriophage lambda DNA (0,1 mg/ml final) was then added to bacteria. Five µL of mix were immediately adsorbed onto a 600 mesh copper grid coated with a thin carbon film, activated by glow-discharge. After 1 min, grids were washed with 0,02% (w/vol) uranyl acetate solution (Merck, France) and then dried with filter paper. TEM observations were carried out with a Zeiss 912AB transmission electron microscope in filtered crystallographic dark field mode. Electron micrographs were obtained using a ProScan 1024 HSC digital camera and Soft Imaging Software system.

Supporting Information

Figure S1 Visualization of competence-induced appendages by Immuno-fluorescence. Same picture as in figure 3A, in high resolution. Left column correspond to bright field image, right column to overlay between anti-FLAG antibody fluorescence (green) and DAPI fluorescence (magenta).
(PDF)

Figure S2 Mass spectrometry analysis of the major pilus component. Fragment map of GomGC generated from several top-down mass spectrometry experiments. Sequence coverage is 74%. MS/MS spectra formed through different fragmentation techniques were deconvoluted and de-isotoped in Xtract and the resulting peak lists combined. Fragment peaks were picked and assigned from this combined list using in house software at a tolerance of 5 ppm. Individual experimental conditions were as follows; ETD 14+ charge state 7 ms activation time; 13+ charge state 10 ms activation time, 5 ms activation time with and without supplementary activation; HCD 30 eV collision energy, 13 eV collision energy; CAD 20 eV collision energy.
(PDF)

Figure S3 Transformation pili are observed in other pneumococcal strains. Competent G54 and TCP1251 *S. pneumoniae* cells were observed by transmission electron microscopy. The same appendages were detected in these strains.
(PDF)

Table S1 *ComG* and *pilD* genes in different pneumococcal strains. The name used to designate the comG and pilD genes varies in different pneumoccocal strains. For clarity, we refer to the comG nomenclature used in B. subtilis. Names of the corresponding genes in different S. pneumoniae strains are found in the table.
(DOCX)

Table S2 Strains and plasmids. The strains and plasmids used in this study are listed in this table.
(DOCX)

Acknowledgments

We thank Olivera Francetic and David Cisneros for their help to set up the shearing assay. We thank all the members of the BSSB group for stimulating discussions.

Author Contributions

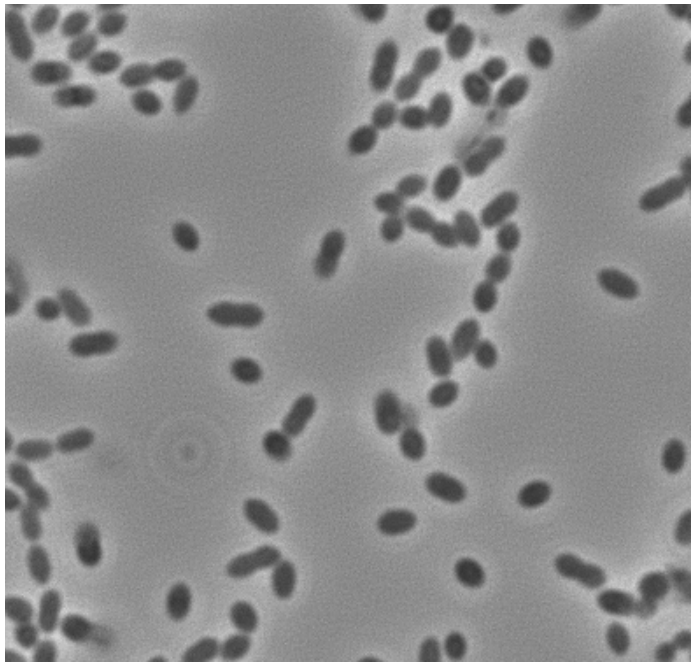
Conceived and designed the experiments: JCR ELC JPC RF. Performed the experiments: RL GPA SB JG CM AD NC. Analyzed the data: JCR ELC JPC RF. Wrote the paper: RL JG NC JCR JPC RF.

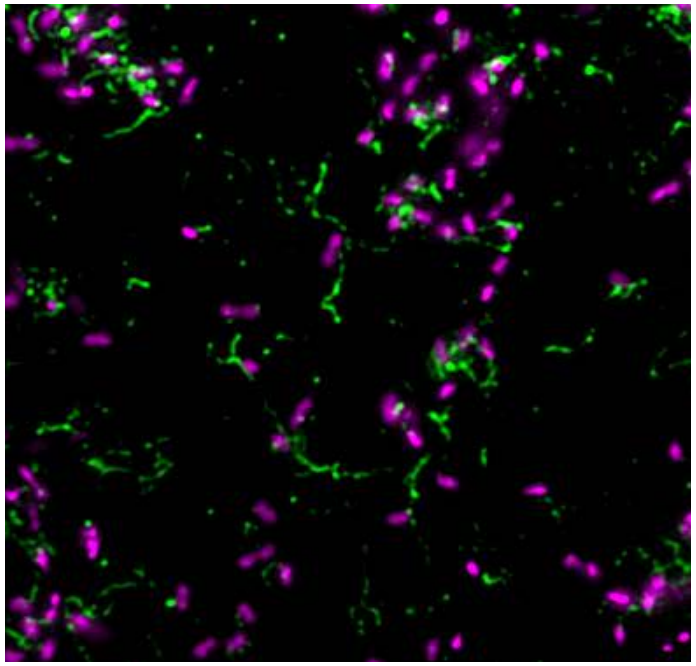
References

1. Griffith F (1928) The Significance of Pneumococcal Types. The Journal of hygiene 27: 113–159.
2. Johnsborg O, Eldholm V, Havarstein LS (2007) Natural genetic transformation: prevalence, mechanisms and function. Research in microbiology 158: 767–778.
3. Popa O, Dagan T (2011) Trends and barriers to lateral gene transfer in prokaryotes. Current opinion in microbiology 14: 615–623.
4. Hiller NL, Ahmed A, Powell E, Martin DP, Eutsey R, et al. (2010) Generation of genic diversity among Streptococcus pneumoniae strains via horizontal gene transfer during a chronic polyclonal pediatric infection. PLoS pathogens 6: e1001108.
5. Havarstein LS, Coomaraswamy G, Morrison DA (1995) An unmodified heptadecapeptide pheromone induces competence for genetic transformation in Streptococcus pneumoniae. Proceedings of the National Academy of Sciences of the United States of America 92: 11140–11144.
6. Prudhomme M, Attaiech L, Sanchez G, Martin B, Claverys JP (2006) Antibiotic stress induces genetic transformability in the human pathogen Streptococcus pneumoniae. Science 313: 89–92.
7. Chen I, Dubnau D (2004) DNA uptake during bacterial transformation. Nature reviews Microbiology 2: 241–249.
8. Claverys JP, Martin B, Polard P (2009) The genetic transformation machinery: composition, localization, and mechanism. FEMS microbiology reviews 33: 643–656.
9. Chen I, Provvedi R, Dubnau D (2006) A macromolecular complex formed by a pilin-like protein in competent Bacillus subtilis. The Journal of biological chemistry 281: 21720–21727.
10. Dagkessamanskaia A, Moscoso M, Henard V, Guiral S, Overweg K, et al. (2004) Interconnection of competence, stress and CiaR regulons in Streptococcus pneumoniae: competence triggers stationary phase autolysis of ciaR mutant cells. Molecular microbiology 51: 1071–1086.
11. Peterson SN, Sung CK, Cline R, Desai BV, Snesrud EC, et al. (2004) Identification of competence pheromone responsive genes in Streptococcus pneumoniae by use of DNA microarrays. Molecular microbiology 51: 1051–1070.
12. Kaufenstein M, van der Laan M, Graumann PL (2011) The three-layered DNA uptake machinery at the cell pole in competent Bacillus subtilis cells is a stable complex. Journal of bacteriology 193: 1633–1642.
13. Berge M, Moscoso M, Prudhomme M, Martin B, Claverys JP (2002) Uptake of transforming DNA in Gram-positive bacteria: a view from Streptococcus pneumoniae. Molecular microbiology 45: 411–421.

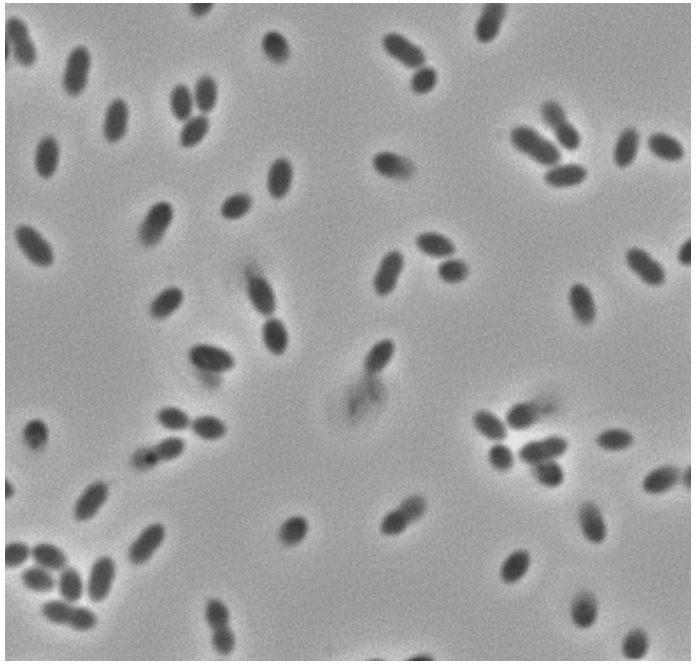
14. Briley K, Jr., Dorsey-Oresto A, Prepiak P, Dias MJ, Mann JM, et al. (2011) The secretion ATPase ComGA is required for the binding and transport of transforming DNA. Molecular microbiology 81: 818–830.
15. Chung YS, Dubnau D (1998) All seven comG open reading frames are required for DNA binding during transformation of competent Bacillus subtilis. Journal of bacteriology 180: 41–45.
16. Nunn D, Bergman S, Lory S (1990) Products of three accessory genes, pilB, pilC, and pilD, are required for biogenesis of Pseudomonas aeruginosa pili. Journal of bacteriology 172: 2911–2919.
17. Martin B, Granadel C, Campo N, Henard V, Prudhomme M, et al. (2010) Expression and maintenance of ComD-ComE, the two-component signal-transduction system that controls competence of Streptococcus pneumoniae. Molecular microbiology 75: 1513–1528.
18. Chamot-Rooke J, Mikaty G, Malosse C, Soyer M, Dumont A, et al. (2011) Posttranslational modification of pili upon cell contact triggers N. meningitidis dissemination. Science 331: 778–782.
19. Strom MS, Nunn DN, Lory S (1993) A single bifunctional enzyme, PilD, catalyzes cleavage and N-methylation of proteins belonging to the type IV pilin family. Proceedings of the National Academy of Sciences of the United States of America 90: 2404–2408.
20. Strom MS, Lory S (1991) Amino acid substitutions in pilin of Pseudomonas aeruginosa. Effect on leader peptide cleavage, amino-terminal methylation, and pilus assembly. The Journal of biological chemistry 266: 1656–1664.
21. Burton B, Dubnau D (2010) Membrane-associated DNA transport machines. Cold Spring Harbor perspectives in biology 2: a000406.
22. Dupaigne P, Le Breton C, Fabre F, Gangloff S, Le Cam E, et al. (2008) The Srs2 helicase activity is stimulated by Rad51 filaments on dsDNA: implications for crossover incidence during mitotic recombination. Molecular cell 29: 243–254.
23. Mejean V, Claverys JP (1993) DNA processing during entry in transformation of Streptococcus pneumoniae. The Journal of biological chemistry 268: 5594–5599.
24. Barocchi MA, Ries J, Zogaj X, Hemsley C, Albiger B, et al. (2006) A pneumococcal pilus influences virulence and host inflammatory responses. Proceedings of the National Academy of Sciences of the United States of America 103: 2857–2862.
25. Alloing G, Martin B, Granadel C, Claverys JP (1998) Development of competence in Streptococcus pneumonaie: pheromone autoinduction and control of quorum sensing by the oligopeptide permease. Molecular microbiology 29: 75–83.
26. Pestova EV, Havarstein LS, Morrison DA (1996) Regulation of competence for genetic transformation in Streptococcus pneumoniae by an auto-induced peptide pheromone and a two-component regulatory system. Molecular microbiology 21: 853–862.
27. Campos M, Cisneros DA, Nivaskumar M, Francetic O (2013) The type II secretion system - a dynamic fiber assembly nanomachine. Research in microbiology [Epub ahead of print] doi: 10.1016/j.resmic.2013.03.013.
28. Varga JJ, Nguyen V, O'Brien DK, Rodgers K, Walker RA, et al. (2006) Type IV pili-dependent gliding motility in the Gram-positive pathogen Clostridium perfringens and other Clostridia. Molecular microbiology 62: 680–694.
29. Imam S, Chen Z, Roos DS, Pohlschroder M (2011) Identification of surprisingly diverse type IV pili, across a broad range of gram-positive bacteria. PloS one 6: e28919.
30. Craig L, Pique ME, Tainer JA (2004) Type IV pilus structure and bacterial pathogenicity. Nature reviews Microbiology 2: 363–378.
31. Biswas GD, Sox T, Blackman E, Sparling PF (1977) Factors affecting genetic transformation of Neisseria gonorrhoeae. Journal of bacteriology 129: 983–992.
32. Dougherty TJ, Asmus A, Tomasz A (1979) Specificity of DNA uptake in genetic transformation of gonococci. Biochemical and biophysical research communications 86: 97–104.
33. van Schaik EJ, Giltner CL, Audette GF, Keizer DW, Bautista DL, et al. (2005) DNA binding: a novel function of Pseudomonas aeruginosa type IV pili. Journal of bacteriology 187: 1455–1464.
34. Lang E, Haugen K, Fleckenstein B, Homberset H, Frye SA, et al. (2009) Identification of neisserial DNA binding components. Microbiology 155: 852–862.
35. Cehovin A, Simpson PJ, McDowell MA, Brown DR, Noschese R, et al. (2013) Specific DNA recognition mediated by a type IV pilin. Proceedings of the National Academy of Sciences of the United States of America 110: 3065–3070.
36. Carbonnelle E, Helaine S, Nassif X, Pelicic V (2006) A systematic genetic analysis in Neisseria meningitidis defines the Pil proteins required for assembly, functionality, stabilization and export of type IV pili. Molecular microbiology 61: 1510–1522.
37. Georgiadou M, Castagnini M, Karimova G, Ladant D, Pelicic V (2012) Large-scale study of the interactions between proteins involved in type IV pilus biology in Neisseria meningitidis: characterization of a subcomplex involved in pilus assembly. Molecular microbiology 84: 857–873.
38. Merz AJ, So M, Sheetz MP (2000) Pilus retraction powers bacterial twitching motility. Nature 407: 98–102.
39. Rabinovich L, Sigal N, Borovok I, Nir-Paz R, Herskovits AA (2012) Prophage excision activates Listeria competence genes that promote phagosomal escape and virulence. Cell 150: 792–802.
40. Martin B, Prudhomme M, Alloing G, Granadel C, Claverys JP (2000) Cross-regulation of competence pheromone production and export in the early control of transformation in Streptococcus pneumoniae. Molecular microbiology 38: 867–878.
41. Sauvonnet N, Vignon G, Pugsley AP, Gounon P (2000) Pilus formation and protein secretion by the same machinery in Escherichia coli. The EMBO journal 19: 2221–2228.
42. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nature methods 9: 671–675.
43. Hansen JK, Forest KT (2006) Type IV pilin structures: insights on shared architecture, fiber assembly, receptor binding and type II secretion. Journal of molecular microbiology and biotechnology 11: 192–207.

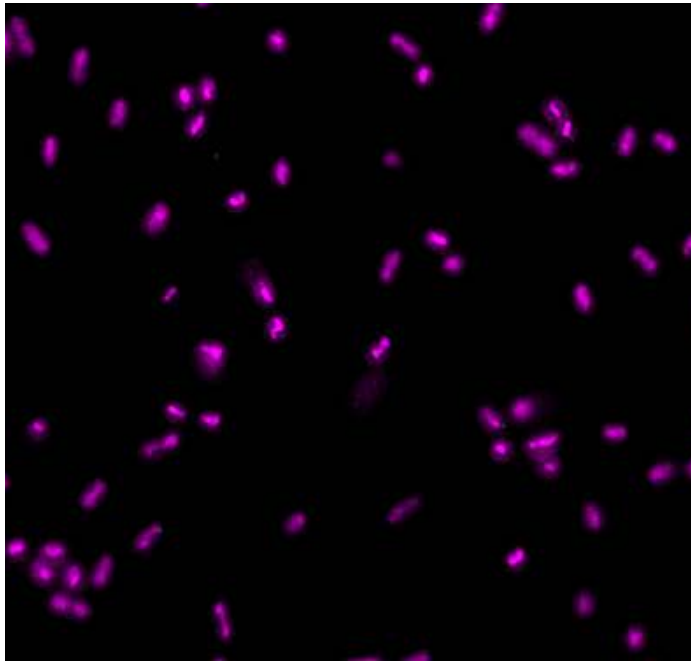
ComGC-FLAG
(RL001)

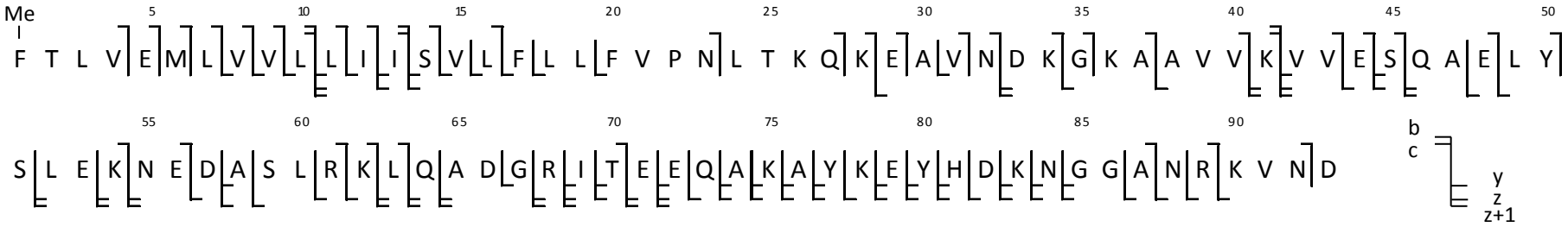


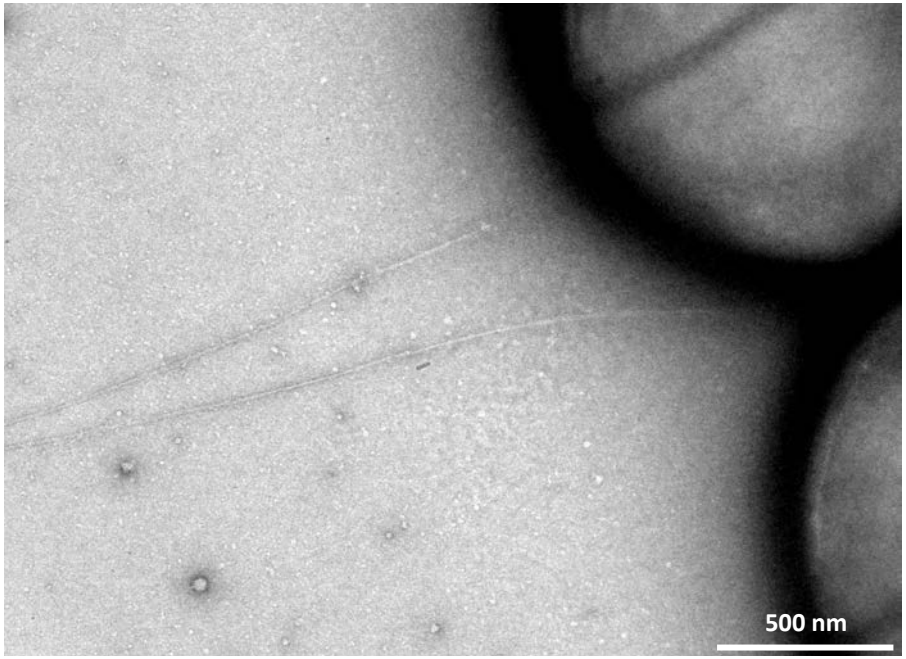


ComGC-FLAG, ComGA^{-}
(RL002)

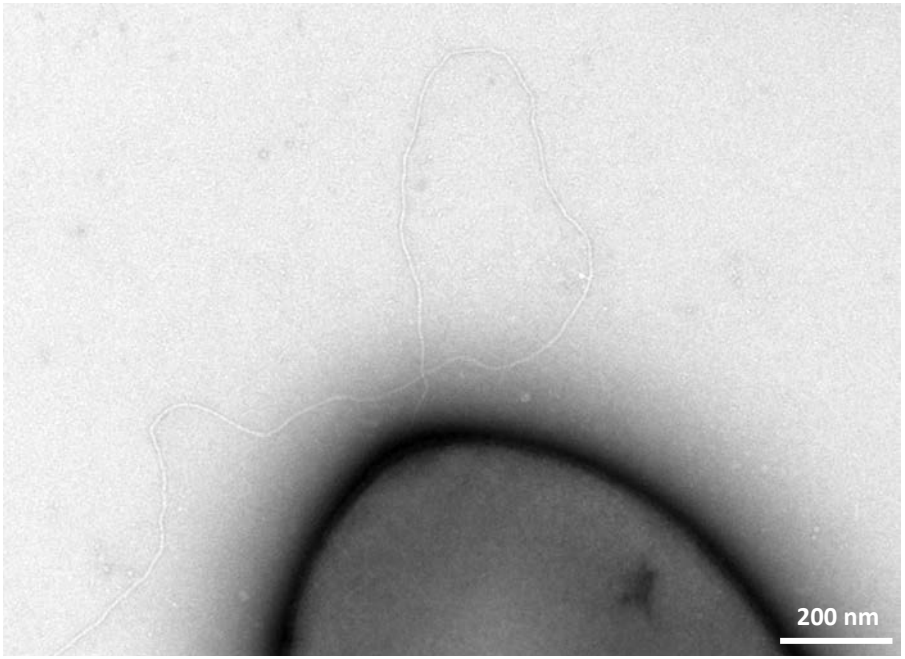





Me
5
10
15
20
25
30
35
40
45
50
F T L V E M L V V L L I I S V L F L L F V P N L T K Q K E A V N D K G K A A V V K V V E S Q A E L Y
55
60
65
70
75
80
85
90
S L E K N E D A S L R K L Q A D G R I T E E Q A K A Y K E Y H D K N G G A N R K V N D
b
c
y
z
z+1



S. Pneumoniae G54 strain
(capsulated)



S. Pneumoniae TCP1251 strain

Table S1

Strain	*comGA*	*comGB*	*comGC*	*comGD*	*comGE*	*comGF*	*comGG*	*pilD*
R6	*spr1864 /cglA*	*spr1863 /cglB*	*spr_1862 /cglC*	*spr_1861 /cglD*	*Not annotated*	*spr_1859*	*spr_1858*	*pilD/spr1628*
D39	*SPD_1863 /cglA*	*SPD_1862 /cglB*	*SPD_1861 /cglC*	*SPD_1860 /cglD*	*SPD_1859*	*SPD_1858*	*SPD_1857*	*SPD_1593*
G54	*SPG_1968*	*SPG_1967*	*SPG_1966*	*SPG_1965*	*SPG_1964*	*SPG_1963*	*SPG_1962*	*SPG_1704*
CP	*Not Sequenced*	*Not Sequenced*	*Not Sequenced*	*Not Sequenced*	*Not Sequenced*	*Not Sequenced*	*Not Sequenced*	*Not Sequenced*
TIGR4	*SP_2053*	*SP_2052*	*SP_2051*	*SP_2050*	*SP_2049*	*SP_2048*	*SP_2047*	*SP_1808*

Table S2

Strain number	Genotype/relevant feature[a]	Reference
R800	R6 derivative	[1]
G54	Clinical isolate of serotype 19F	[2]
TCP1251	Rx derivative but *malM511, rpsL1, bgl1;* Sm^R	[3]
R1501	R800 but Δ*comC*	[4]
R304	R800 derivative, *nov1, rif23, str41;* Nov^R, Rif^R, Sm^R	[5]
R1916	R1501 but *ssbB::luc (ssbB*$^+$*), comGA::kan;* Cm^R, Kan^R	Claverys' strain collection
R998	R1501 but *comEC::spc;* Cm^R, Spc^R	Claverys' strain collection
R1063	R1501 but *comFA::spc;* Cm^R, Spc^R	Claverys' strain collection
RL001	R1501, but CEPx-*comGC-FLAG* (from plasmid pCEPx-*comGC-FLAG*); Kan^R	This study
RL002	RL001 , but *comGA::spc3*^{C} (from strain R1062); Kan^R, Spc^R	This study
RL003	R1501 but *comGC* E20A (point mutation of ComGC pilin)	This study

Plasmids		
pCEPx	ColE1 (pBR322) derivative containing the ComX-dependent promoter, P_X, and the RBS of ssbB; Kan^R	[6]

[a] ^{R}, resistance.

1. Lefevre JC, Claverys JP, Sicard AM (1979) Donor deoxyribonucleic acid length and marker effect in pneumococcal transformation. Journal of bacteriology 138: 80-86.
2. Dopazo J, Mendoza A, Herrero J, Caldara F, Humbert Y, et al. (2001) Annotated draft genomic sequence from a Streptococcus pneumoniae type 19F clinical isolate. Microbial drug resistance 7: 99-125.
3. Pestova EV, Havarstein LS, Morrison DA (1996) Regulation of competence for genetic transformation in Streptococcus pneumoniae by an auto-induced peptide pheromone and a two-component regulatory system. Molecular microbiology 21: 853-862.
4. Dagkessamanskaia A, Moscoso M, Henard V, Guiral S, Overweg K, et al. (2004) Interconnection of competence, stress and CiaR regulons in Streptococcus pneumoniae: competence triggers stationary phase autolysis of ciaR mutant cells. Molecular microbiology 51: 1071-1086.
5. Mortier-Barriere I, de Saizieu A, Claverys JP, Martin B (1998) Competence-specific induction of recA is required for full recombination proficiency during transformation in Streptococcus pneumoniae. Molecular microbiology 27: 159-170.
6. Martin B, Granadel C, Campo N, Henard V, Prudhomme M, et al. (2010) Expression and maintenance of ComD-ComE, the two-component signal-transduction system that

controls competence of Streptococcus pneumoniae. Molecular microbiology 75: 1513-1528.

II. Controversy on *S. pneumoniae* transformation pilus morphology and mechanism of action

Petya Violinova Krasteva, postdoctoral fellow in the lab, as well as Sahra Ouarti, a student under my supervision, have greatly contributed to the work presented in this chapter.

A. A Type II secretion pseudopilus-like device rather than a Type IV pilus?

Controversy was ignited by a recent study by Balaban and colleagues (Balaban et al., 2014). Studying competence-induced *S. pneumoniae* cells, the authors visualized a completely different structure: a short, plaited polymer that is released in the medium.

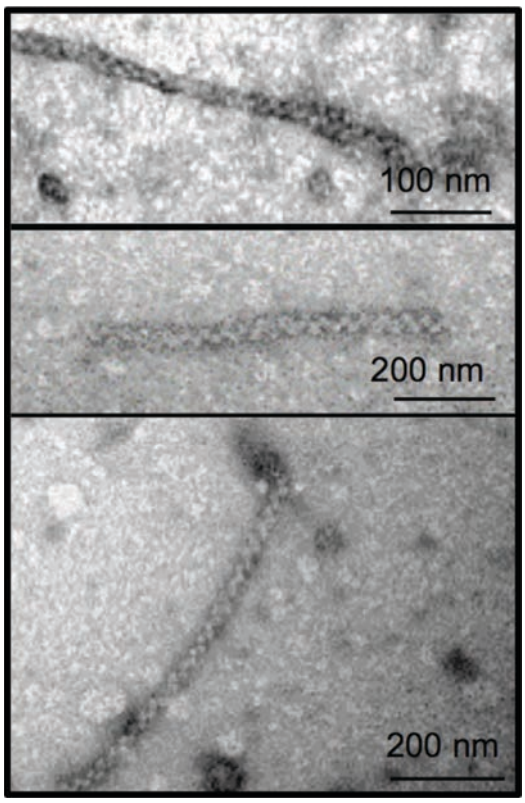


Figure 25 - Plaited filaments found in the supernatant of S. pneumoniae competent cells

Taken from Balaban et al., 2014. Plaited pilus structures that were found in the supernatants of induced pneumococcal cultures after 30 min of CSP induction.

Biochemical observation of significant ComGC release in the medium during competence convinced the authors that the plaited structures correspond to transformation pili. After failing to immunolabel these structures, they expressed heterologously the whole *comG* operon in *E. coli* and detected release of similar plaited polymers. Finally, they proposed a model in which, rather than acting as a DNA receptor, the pilus acts as a peptidoglycane-drilling device, which upon release leaves a gateway for transforming DNA to find the uptake machinery.

We observed the structures visualized by Balaban and colleagues along with the transformation pili (Figure 26). Approximately 8-12 nm-wide and a hundred nanometers in average length, these structures were indeed morphologically distinct from the ~5-6 nm wide, several micrometer-long transformation pili.

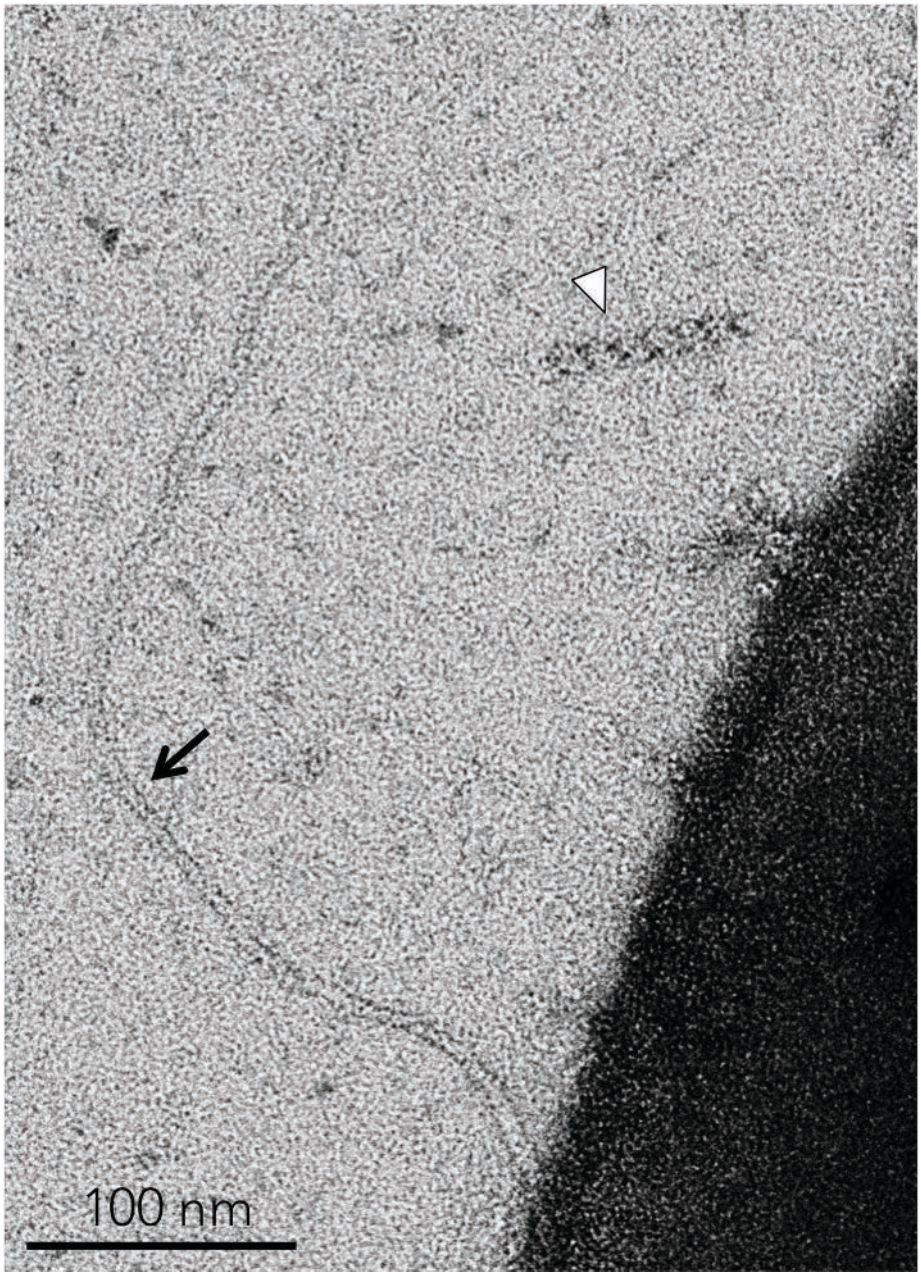


Figure 26 - Plaited filament next to a transformation pilus
The transformation pilus is indicated by a black arrow, the plaited filament by a white arrowhead.

To investigate whether these structures could be enlaced ComGC pili, we started by checking the supernatant of pilus-deficient strains. We visualized the plaited filaments in the ΔcomGB and ΔcomGA strains, seemingly in the same amount as in a strain able to assemble the transformation pilus. To our surprise, we further checked in the supernatant of cells that cannot trigger themselves the competence state (deleted for *comC*), and were equally able to detect

the plaited filaments in absence of CSP. These structures are therefore not related to the transformation pilus, and not even related to competence.

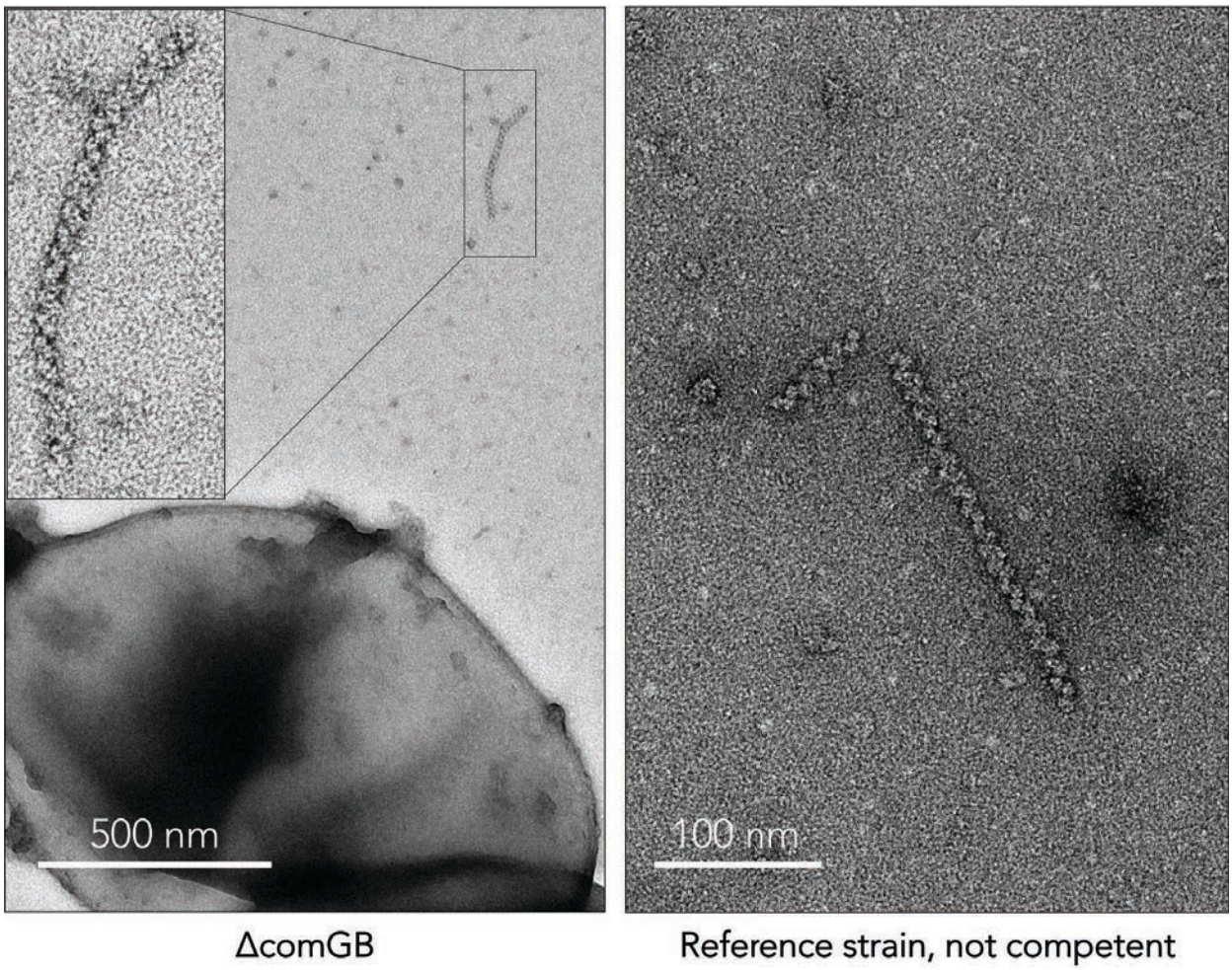


Figure 27 - Plaited filaments are a contaminant unrelated to transformation pili

Plaited filaments are found in the supernatant of competence induced ΔComGB cells, and in the supernatant of non induced R1501 cells.

Balaban and colleagues dismissed the existence of a long transformation pilus in *S. pneumoniae* by implying that our results were overexpression artifacts. However this is an erroneous report on our study, since we observed micrometer-long transformation pili on different pneumococcal strains with a single *comGC* gene copy at its genomic locus. Among these are two highly transformable laboratory strains of different genetic background – R6 and TCP1251 – as well as a capsulated clinical isolate – the G54 strain. Only in a subset of experiments, we expressed an additional Flag-tagged ectopic copy of the *comGC* gene, which

enabled us to immunolabel the pili and unambiguously identify them as transformation-associated cellular appendages.

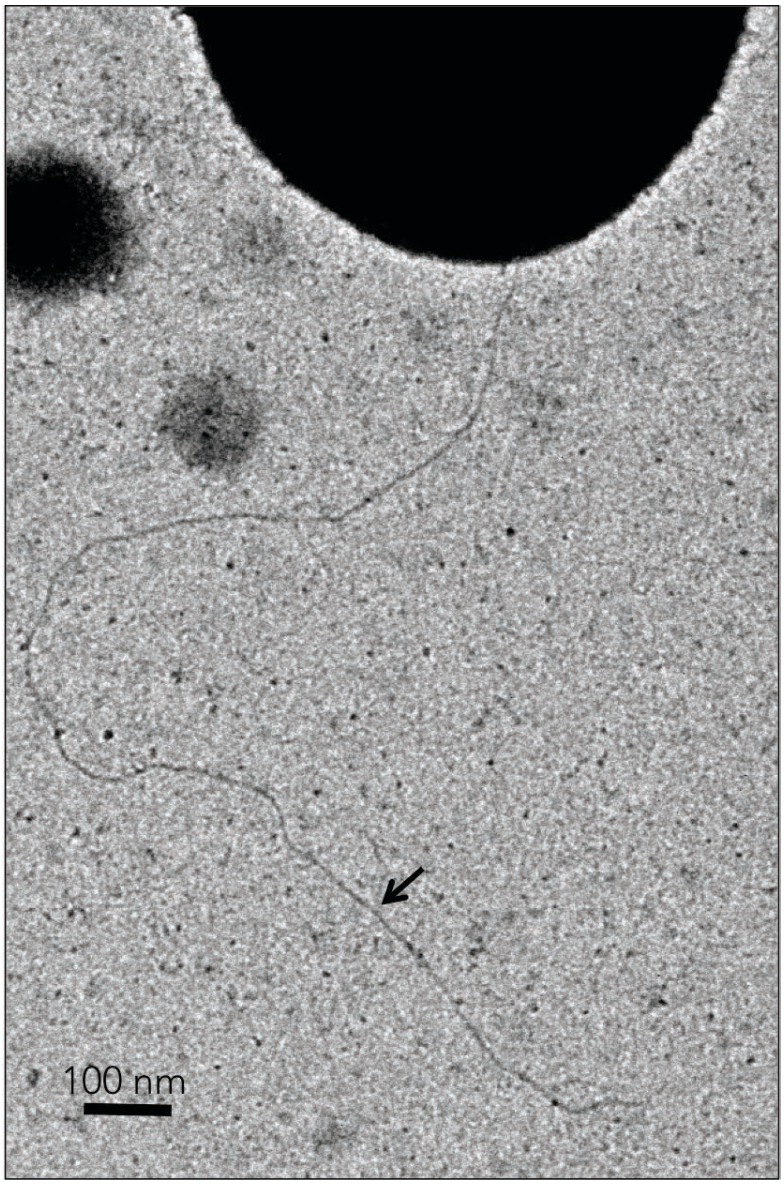


Figure 28 - Micrometer-sized transformation pilus on the surface of R1501 cells without overexpression of ComGC
The transformation pilus is indicated by a black arrow.

Balaban and colleagues analyzed the quantity of ComGC released in the supernatant of different *S. pneumoniae* strains after centrifugation and showed that it correlates with the peak of transformation efficiency. Since we previously showed that the transformation pilus is absolutely required for DNA uptake, it is not surprising to observe correlation between

extracellular ComGC and transformation efficiency. However the idea that transformation pili are actively released in the medium to serve a physiological role in competence remains dubious. We showed previously that the transformation pili are fairly sensitive to mechanical stress, as short vortexing and/or fast centrifugation are routinely used for pili shearing and isolation. Such mechanical forces, however, are exclusive to laboratory practice and are unlikely to be exerted in nature. Moreover, competence in *S. pneumoniae* has been linked to biofilm formation (Vidal et al., 2013) – a collaborative bacterial life-style that offers protection from noxious stimuli and mechanical stress.

B. DNA binding

We showed that native pili bind and co-purify with DNA already present in the cell culture. While no mechanistic or quantitative data on this binding are available, electron microscopy showed extensive contact interfaces between the pili and DNA filaments. It is plausible that multiple weak interactions in the context of the helical pilus lattice stabilize this interaction and allow its reversal upon DNA uptake. Such a scenario would also explain why no DNA binding to a monomeric ComGC truncation was ever reported (Provvedi et al., 1999). As exposed in the next chapter we did not observe any DNA binding to the soluble domain of ComGC either. Balaban and colleagues also failed to observe DNA binding in the context of a gel shift assay, using as a probe cellular supernatant, supposedly enriched in polymeric ComGC through ultracentrifugation pelleting. Nevertheless, the fact that the authors could not visualize any transformation pili strongly suggest that pilus structure and function were severely compromised during sample handling, leading to ComGC filament aggregation or degradation, rather than enrichment. While pili are broken in short pieces as we showed during our affinity chromatography purification on the anti-Flag resin, they are still filamentous, disperse, very pure and concentrated. Of note, after treating the pili with DNase on the anti-FLAG resin to remove the bound DNA, we did not detect any binding activity of the purified pili fraction by a

gel shift assay. We can think of several explanations for this result: the DNase enzyme does not digest DNA efficiently when it is bound to a protein receptor. Some residual DNA might consequently remain on the pili and prevent additional binding to occur. Alternatively, pili might loose their ability to bind once they are eluted from the column, because broken in pieces too small to bind efficiently, or because a DNA binding partner has been removed in the purification.

We are confident that our data on DNA binding by the pilus remain unchallenged and deserve further experimental investigation. Moreover, they are strongly supported by the fact that DNA binding at the surface of competent *S. pneumoniae* is abolished in a pilus-deficient strain (Bergé et al., 2002; Bergé et al., 2013).

C. On the nature of the plaited filaments

Even if these plaited filaments are not related to transformation, we were curious to know what they are. They are constantly found in pneumococcal cultures and could be of importance. Also, we want to submit a manuscript to show that the plaited filaments are not transformation pili and that the conclusions of the study published by Balaban et al. are wrong. To do so, we have to identify the nature of these filaments.

i) RecA filaments?

As the appearance of the short polymers were reminiscent of polymerized RecA protein from other bacterial species (Williams et al., 1986; Yu et al., 1990), we hypothesized that the structures observed by Balaban and colleagues are RecA filaments, randomly released in the medium after cell lysis. A cytosolic protein, RecA is massively expressed during competence and its polymerization on the incoming single-stranded DNA is essential for DNA integration in

the genome by homologous recombination. RecA could be detected by western blot on the supernatant of competent cells, without shearing. Nevertheless, RecA null cells display normal DNA uptake during competence induction (Bergé et al., 2003), so extracellular release of RecA is likely the result of random cell lysis and bears no physiological relevance to natural transformation.

The structure of polymerized RecA homologs has been well documented in the literature: forming characteristic helical coils, RecA filaments can be more or less extended depending on the presence and type of DNA and small-molecule ligands (Yu et al., 1990; Okorokov et al., 2010; Vanloock et al., 2003; Yu et al., 2004). Moreover, in vitro reconstituted filaments using purified *S. pneumoniae* RecA were indeed similar to the plaited filaments, and class averages of the ‘plaited’ structures are strikingly similar to two-dimensional reprojections generated from known structures of RecA-homologous proteins.

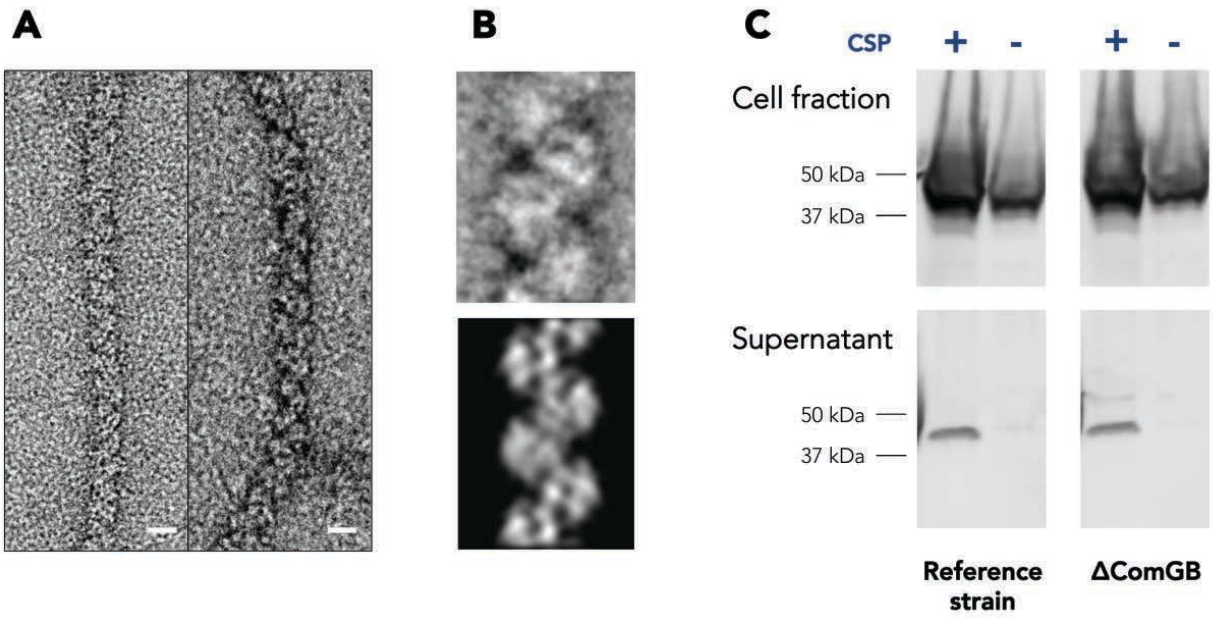


Figure 29 - Plaited filaments compared to RecA filaments
Data courtesy of Petya Violinova Krasteva. (A) In vitro reconstituted S. pneumoniae RecA – ssDNA nucleofilaments (left) compared to short 'plaited' filaments found in culture supernatant (right). Scale bars are 10 nm. (B) Comparison of class averages of the 'plaited' filaments from culture supernatant to two-dimensional reprojections of hDmc1 nucleofilament structure (Okorokov et al., 2010). (C) anti-RecA(E. coli) western blot of R1501 (reference strain) and R1918 (ΔcomGB). anti-RecA(E.coli) detects RecA(S. pneumoniae) due to the high percentage of identity (63% calculated with Blastp).

Despite these convincing elements, we showed that the plaited filaments are not polymerized RecA, by observing the same structures in the supernatant of ΔrecA mutant cells. This result, combined with the work of Balaban and colleagues, underlines how careful any identification by TEM should be.

ii) PBP3 filaments?

To identify the component(s) of the filaments, we performed the ultracentrifugation of *S. pneumoniae* cell supernatant (material and methods II.G.). The filaments were significantly enriched by this protocol as seen by TEM (figure 30). The different components were separated by SDS-PAGE, and the bands were analyzed by MALDI -TOF mass spectrometry.

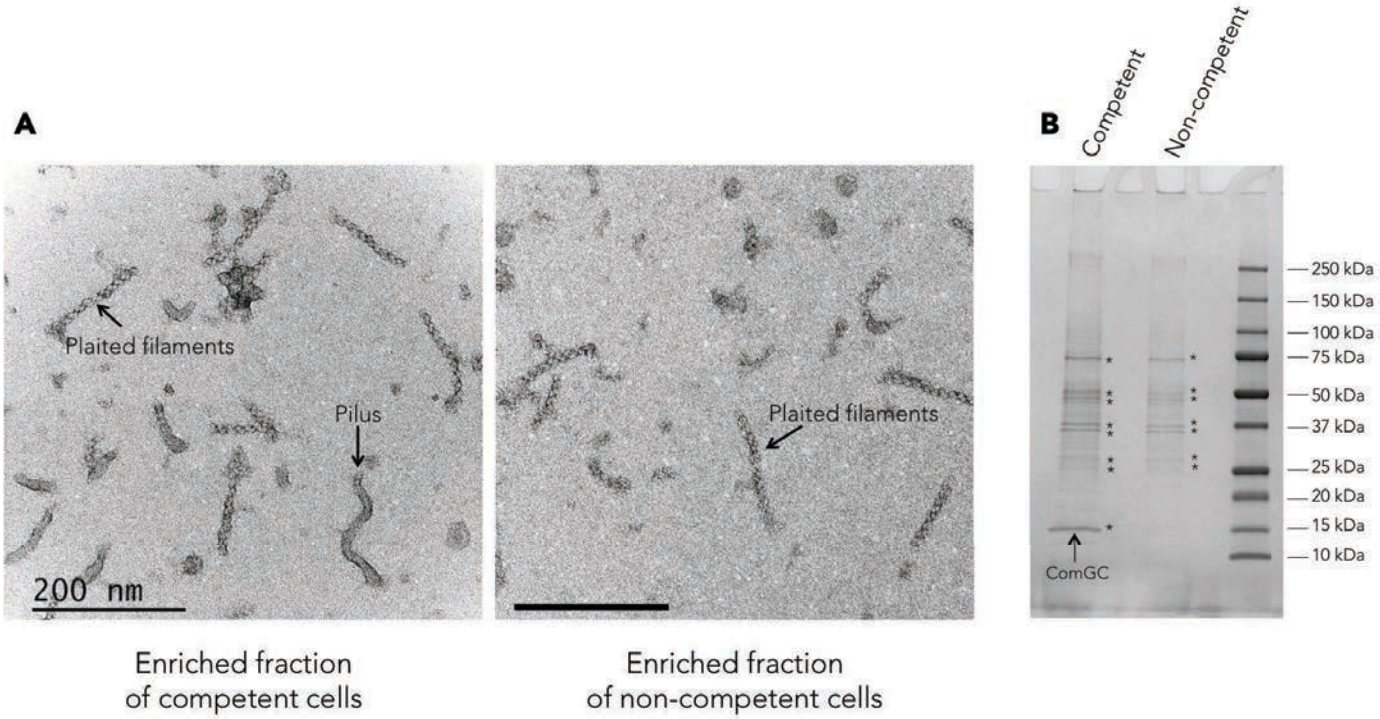


Figure 30 - Plaited filament identification by mass spectrometry

(A) UAc 2% negative stain picture of samples obtained after ultracentrifugation enrichment of R1501 cell supernatant. Cells were induced with CSP (left) or not induced (right). Small broken pili are visible in the supernatant of competent cells along with plaited filaments. (B) Coomassie stained SDS-PAGE gel of both samples. A ComGC band appears only in the competent cell supernatant. Bands sent for MALDI-TOF analysis are shown with a star.

No obvious candidate (protein already known to form filaments) was identified. After careful examination of all proteins, the carboxypeptidase PBP3 (Uniprot Q8DQ99) turned out to be interesting, due to a distant mitochondrial homolog that was described to form strikingly similar filaments (Polianskyte et al., 2009).

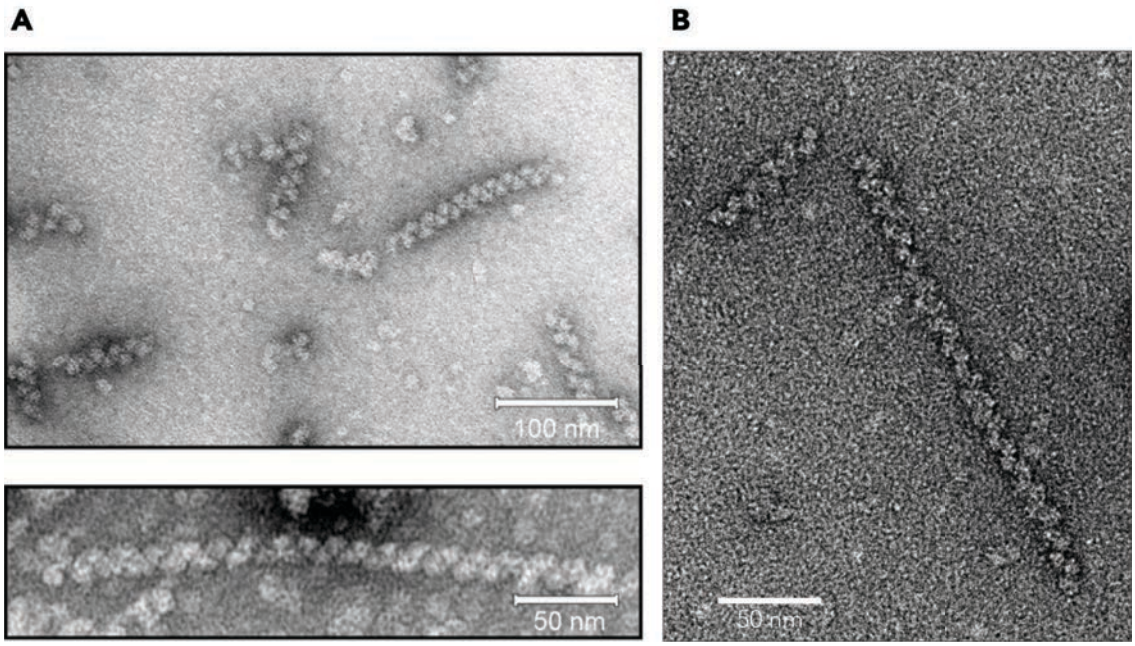


Figure 31 - Plaited filaments compared to LACTB filaments
(A) Taken from Polianskyte et al., 2009: negative stain picture of LACTB filaments. (B) Negative stain picture of the pneumococcal plaited filaments.

PBP3 is involved in peptidoglycane synthesis and well conserved among bacteria (25% identity between *S. pneumoniae* PBP3 and its homolog PBP5 in *E. coli*). Although the protein family has been extensively studied (Morlot et al., 2004; Morlot et al., 2005), none of the bacterial homologs have been described to polymerize. We performed immunogold labeling (see protocol in Laurenceau et al., 2013) using antibodies raised against a C-terminal truncated form of PBP3 (Morlot et al., 2005). No labeling was observed. In the meantime, we obtained a ΔPBP3 strain to check for the presence of the plaited filaments in the supernatant (gift from C. Grangeasse). We did not observe any plaited filaments in this mutant. However the ΔPBP3 strain is extremely sick, showing abnormal cell shape due to severe cell division deficiency. The absence of plaited filaments could therefore be due to the particularly severe phenotype.

We are still investigating PBP3 as a potential candidate, and increasing the purity of plaited filaments by size exclusion chromatography in order to reduce the list of protein candidates detected by mass spectrometry. Notably, we found entirely by chance similar filaments in the supernatant of *C. difficile* cultures. The sample preparation was not done in parallel to an

experiment with *S. pneumoniae* cells, so that a contamination is impossible. This finding suggest that the protein contained in these filaments might be well conserved in bacteria, and would explain why Balaban and colleagues observed similar structures in *E. coli* after heterologous expression of the *comG* operon in this bacteria (Balaban et al., 2014). We are currently checking if these filaments are found in *E. coli* cultures as well.

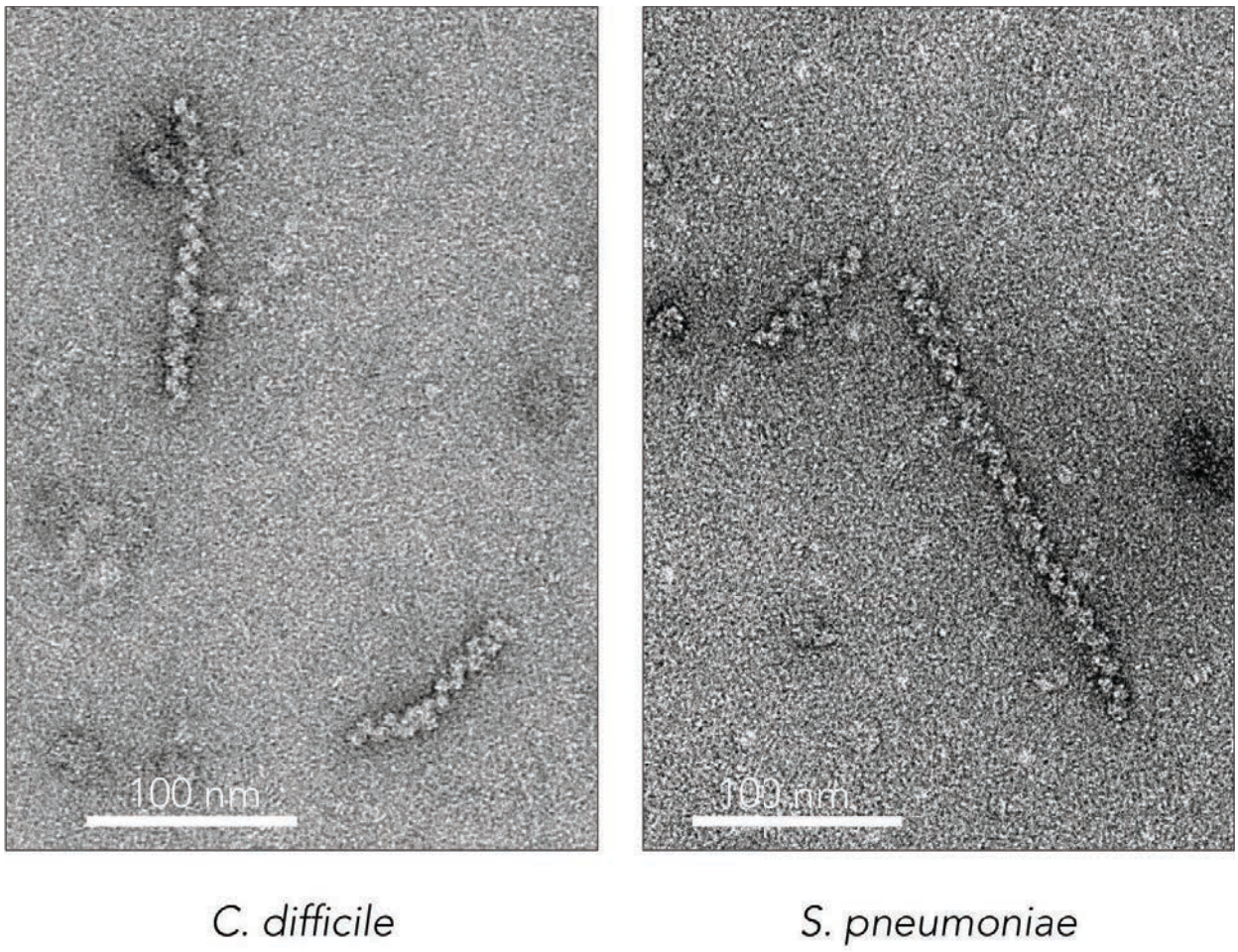


Figure 32 - Plaited filaments found in C. difficile culture supernatant
Negative stain UAc 2%.

Structural study of the transformation pilus subunits

I. The structure-function approach

Visualizing the transformation pilus raised new questions related to its morphology: what is the precise mechanism of DNA binding on the pilus, and how is this actual T4P assembled by such simple assembly platform? A particularly powerful way to answer is by going for a structural biology approach.

Three techniques are widely used in structural biology to obtain molecular scale information: X-ray crystallography, Transmission Electron Microscopy (TEM), and Nuclear Magnetic Resonance (NMR).

A. X-ray crystallography

X-ray crystallography is the most powerful structural biology technique, which has provided since its beginnings in the 50s a vast number of near atomic resolution structures. This method has triggered a new era in biology: the understanding of life at its most intimate scale, the atomic scale. As an outstanding example, the almost complete structure of the ATP synthase could be solved by assembling crystal structures of its different parts from different organisms, combined with structures obtained by the other techniques described below. This complex motor is responsible for synthesizing ATP in all living organisms, and subsequently power all other molecular machineries in the cell.

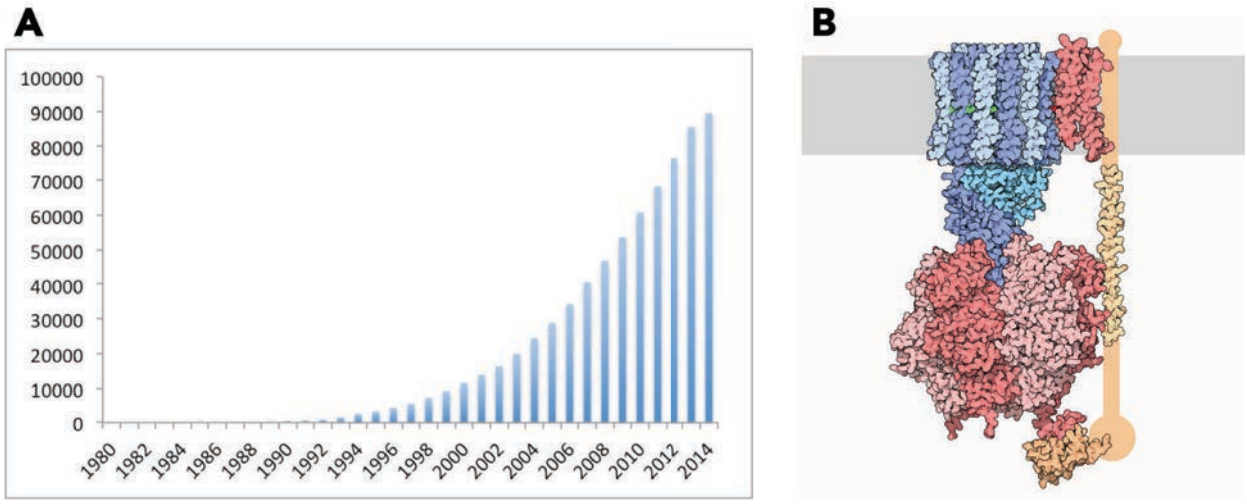


Figure 33 - Total number of X-ray structures in the PDB per year

(A) Total number of X-ray structures in the PDB per year. (B) Model of the ATP synthase embedded in a lipid bilayer (grey), generated by combining 3 different crystal structures and a NMR structure. Data from the Protein Data Base (http://www.rcsb.org/pdb).

For this technique to be used, the biological molecule, in our case a protein, has to be purified and concentrated, a first significant hurdle for success. Then, it needs to be crystallized. A crystal is a periodic arrangement of a lattice containing one or more copies of the molecule of interest. The lattice is repeated identically throughout the crystal. Crystallization is induced by changing the state of the protein from soluble to solid phase. This phase transition can be divided in two stages, nucleation and crystal growth. To induce nucleation, the protein solubility needs to be slowly decreased to reach super saturation.

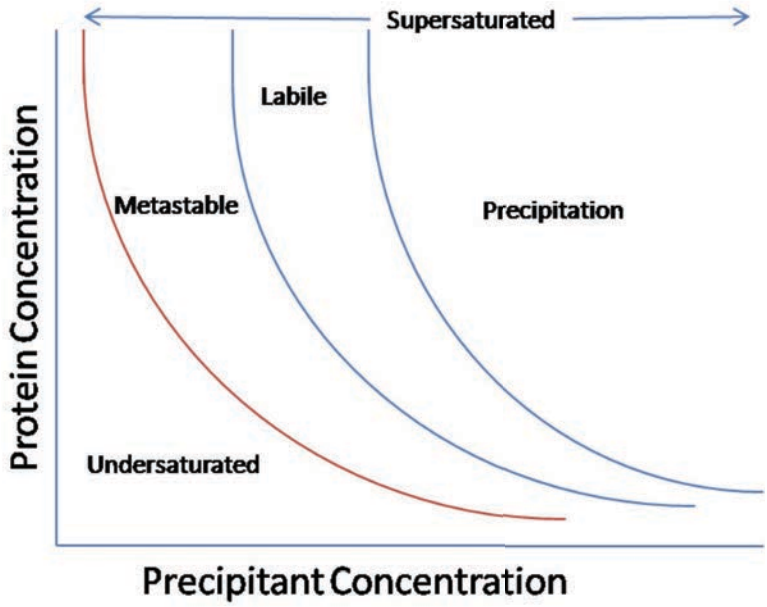


Figure 34 - Phase diagram for protein crystallization
The red curve corresponds to the protein solubility. Crystallization can only occur above this curve in the supersaturated area. The labile zone in the targeted area, in which protein crystals can nucleate and grow. In the metastable area, crystals can only grow from a preexisting crystal seed, and cannot nucleate.

Unfortunately, triggering crystal nucleation (ordered arrangement of the proteins) rather than precipitation (random aggregation of the proteins) depends on so many parameters that it is still today impossible to control nor predict, and a wide variety of conditions (temperature, pH, precipitants, salts, additives, etc.) has to be screened for this purpose. All these elements are going to affect protein stability and the distribution of charges on its surface, so that a nucleation favorable environment can be found. It represents the second major hurdle to X-ray crystallography. The system then evolves towards its thermodynamic equilibrium by entering growth phase of the crystal.

When exposed to an X-ray beam, the crystal arrangement will cause scattering and interference of the beam, resulting in a diffraction pattern. The angles and intensities of the diffracted rays are directly related to the electronic density within the crystal lattice. If the resolution is good enough (typically below 4 Å), meaning that the crystal periodicity is precisely

maintained throughout the crystal, this electronic density can be calculated, and the 3D structure of the protein can be solved.

B. Transmission Electron Microscopy

TEM uses the interaction of an electron beam with an object to form an image, like optical microscopy uses the interaction with a beam of visible light photons. The resolution that can be obtained from a given beam cannot exceed a theoretical limit imposed by the diffraction phenomenon. The theoretical maximum resolution reachable is given by the equation $d=\lambda/(2.NA)$, where d is the minimum distance at which two points are seen separately, lambda is the beam wavelength, and NA (numerical aperture) a number characterizing the optical setup. The equation indicates that the smaller the wavelength, the better the resolution, which explains the huge potential of electron microscopy: electrons wavelength is around 0,01 nm in TEM, while visible light is only above 400 nm. Current transmission electron microscopes typically reach a resolution of about 0,5 nm, while optical microscopy is limited to 200 nm.

Moreover a 3D structure of an object like a protein can be obtained by combining a great number projection images of this object in random orientations, a process called single particle analysis. After aligning, classifying, averaging, and assigning angles of thousands of particle images, a 3D reconstruction can be generated. TEM is better suited for big biological molecules, like protein complexes, and is consequently a complementary technique to X-ray crystallography, which is better adapted to small molecules.

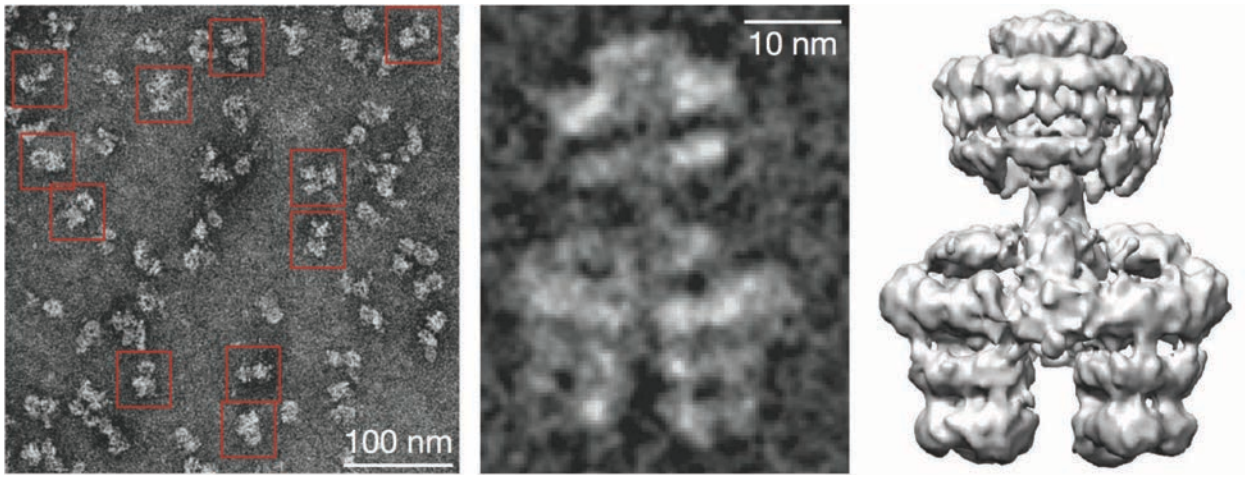


Figure 35 - 3D reconstruction of a Type IV secretion system
Taken from Low et al., 2014. From left to right: Negative stain electron microscopy image with selected single particles; class average of a front view; electron density map of the R388 conjugative plasmid type IV secretion system.

C. Nuclear Magnetic Resonance

The NMR phenomenon is based on the fact that nuclei of atoms have magnetic properties that can be used to yield chemical information when the sample is in a strong magnetic field. The chemical information (interpreted as chemical shift parameter in a spectra) depends on the nature of the atom and its environment in the molecule. Signals from 13C, 1H and 15N nuclei being among the isotopes that are observable by NMR, this technique is particularly efficient to study biological molecules. The advantage of NMR is to provide structural and dynamics information on molecules in solution. The NMR allows also to study proteins with flexible and disordered regions, usually difficult targets for X-ray crystallization, to determine their structures at atomic resolution and to investigate their dynamics and molecular interactions.

Tridimensional structures of proteins (usually below 50 kDa) can be solved by NMR spectroscopy, in a two-step process: chemical shifts of all 13C, 1H and 15N atoms in the protein are assigned using 2 and 3D heteronuclear experiments, and distance restraints between these atoms are used for structure calculations. The main limitation however is the

requirement of a stable sample during the time of NMR experiment and containing 13C and 15N isotopically labeled proteins in relatively large amount.

II. ComGA, the pilus assembly motor

Chiara Rapisarda, a postdoctoral fellow in the lab, has greatly contributed to this work.

A. Heterologous expression in E. coli and crystallization of different ComGA homologs

The protein ComGA is essential for pilus growth, presumably by providing the energy necessary to polymerize the pilins. Although some structures of ComGA homologs have been solved, they all belong to Gram-negative twitching motility T4P or T2S systems. No reliable homology model could be obtained for ComGA using these structures as template. As protein crystallization has a low probability of success, we selected several ComGA homologs from different Gram-positive bacteria in addition to the *S. pneumoniae* homolog: ComGA from *B. subtilis* (uniprot P25953) and *Geobacillus thermodenitrificans* (uniprot A4IQW5), sharing respectively 34% and 38% similarity with ComGA$^{S.\ pneumoniae}$.

We used *E. coli* BL21 strain as a protein production platform. A great number of plasmids allowing protein expression with different tags are commercially available for expression in *E. coli.* (see materials and methods I.A. for cloning procedure, and II.A. for protein expression procedure).

ComGA expression was evaluated by looking both in the soluble and membrane fraction. Indeed, the protein is predicted to be associated to the membrane. ComGA$^{B.\ subtilis}$ was proteolysed in *E. coli*, whether a Strep-tag was placed at the N-terminal or C-terminal end of

the protein sequence. ComGA$^{G.\ thermodenitrificans}$ *expressed with a C-terminal Strep-tag* seemed to associate with the membranes, even after washing the membrane fraction with 1M NaCl buffer. It was extracted from the membrane fraction by using the detergent n-Dodecyl-β-D-maltoside (DDM), but could not be retrieved after affinity chromatography, possibly because the tag was not accessible.

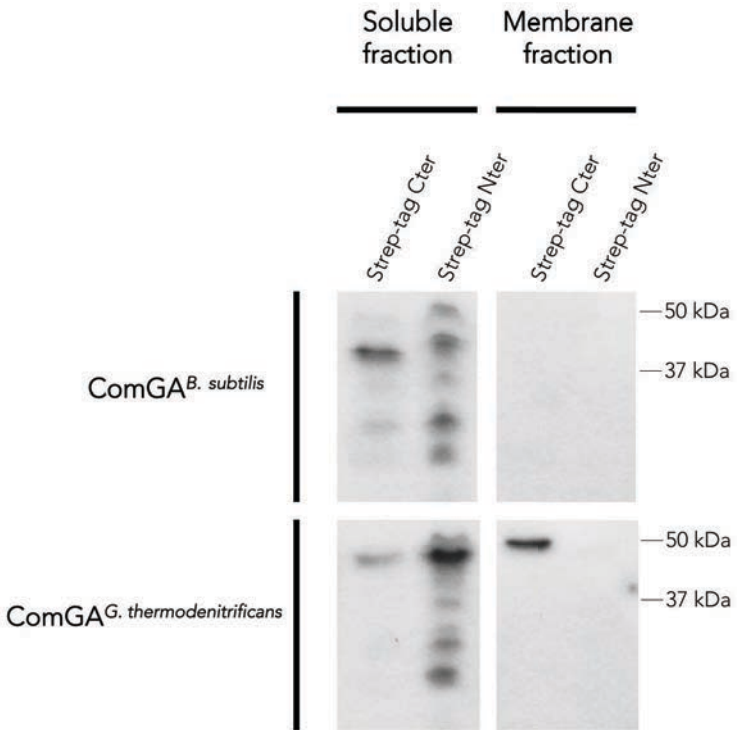


Figure 36 - ComGA homologs expression test
Western blot anti-Strep comparing the amount of ComGA-Strep in the soluble and membrane fractions. Each fraction has the same volume, so that the comparison between each fraction is quantitative. Multiple bands in the lane shows that the protein is degraded in several pieces. The absence of bands shows either non-expression or complete degradation of the protein.

ComGA$^{S.\ pneumoniae}$ with a N-terminal 6xHistidine-tag was found in small amount in the membrane fraction, however it aggregated rapidly after solubilization by detergents. With a C-terminal Strep-tag, ComGA was expressed in high amounts, and found in vast majority within the soluble fraction. Since a fraction of the protein always remained associated with the membrane fraction, membranes were always separated from the sample (materials and methods II.B.) by ultracentrifugation in order to avoid having a mix of 2 different ComGA

populations in the sample. After affinity purification and size-exclusion chromatography, the purity of the protein was satisfactory, although a 25 kDa contaminant could not be avoided. N-terminal sequencing (Edman degradation of the protein after transfer of the proteins from the gel to a PVDF (polyvinylidene fluoride) membrane and extracting the protein from the membrane) indicated the amino-acid sequence LHFWFQ, corresponding to a degradation product of the ComGA protein. A ComGA 3D structure was modeled with the Swiss Model server, using the crystal structure of *V. cholerae* T2S ATPase GspsE (PDB 4KSS) as template. GspE shares 26,56% identity with ComGA. Keeping in mind that this model can only provide a rough idea about the actual ComGA structure, the cleavage site was positioned in a linker region separating the 23 kDa N-terminal domain from the 13 kDa C-terminal domain. It could explain why this region is particularly susceptible to proteolysis. These degradation products could never be completely avoided, but were removed before crystallization by concentrating the sample on a membrane on which they were not retained (100 kDa cutoff).

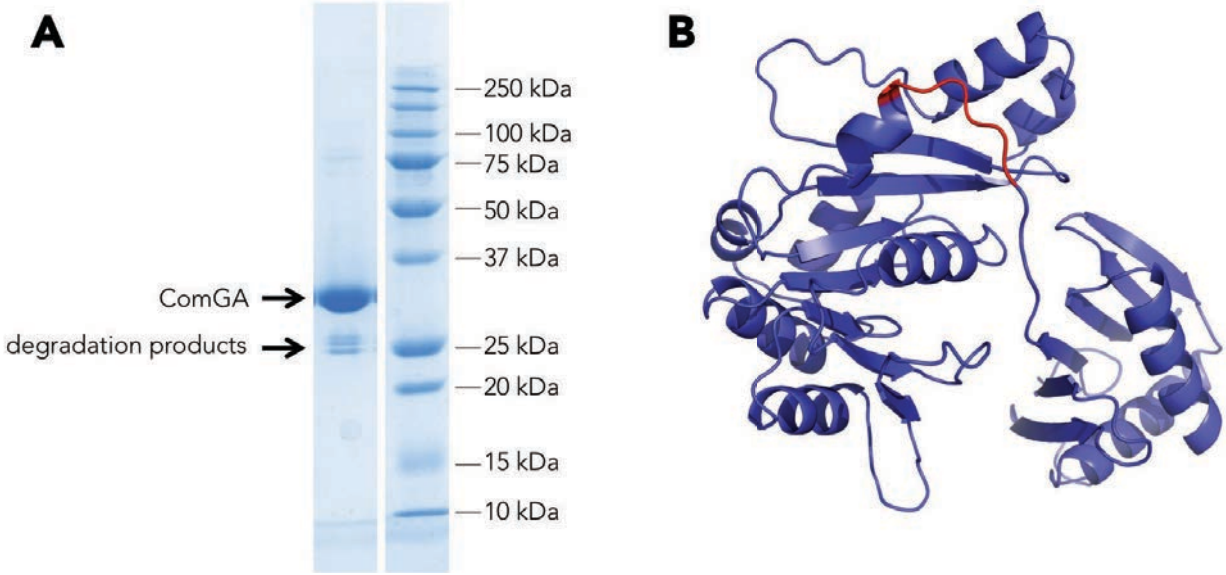


Figure 37 - Purified ComGA sample

(A) Coomassie stained SDS-PAGE of purified ComGA; (B) ComGA model generated by Swiss Model, with the sequence LHFWFQ colored in red (performed with the PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

In the end, ComGA from *S. pneumoniae* was the only ComGA homolog that we were able purify, despite its instability during the purification process. Crystallization trials were performed as described in the materials and methods section (see materials and methods II.C.i.). By optimizing the buffer conditions and the concentration procedure, the protein could be concentrated up to ~ 9 mg/mL. Above this threshold, the protein was massively precipitating. Two different crystal forms were obtained:

- diamond-shaped crystals (Tris 100 mM pH=7; NaCl 2,5 M)
- parallelepiped rectangle crystals (Bis-tris propane 100 mM pH=7,9; NaCl 3,2 M)

Crystal size could be increased by varying the buffer conditions (pH, salt, precipitant concentration), and tens of these crystals were tested at the ESRF synchrotron X-ray source.

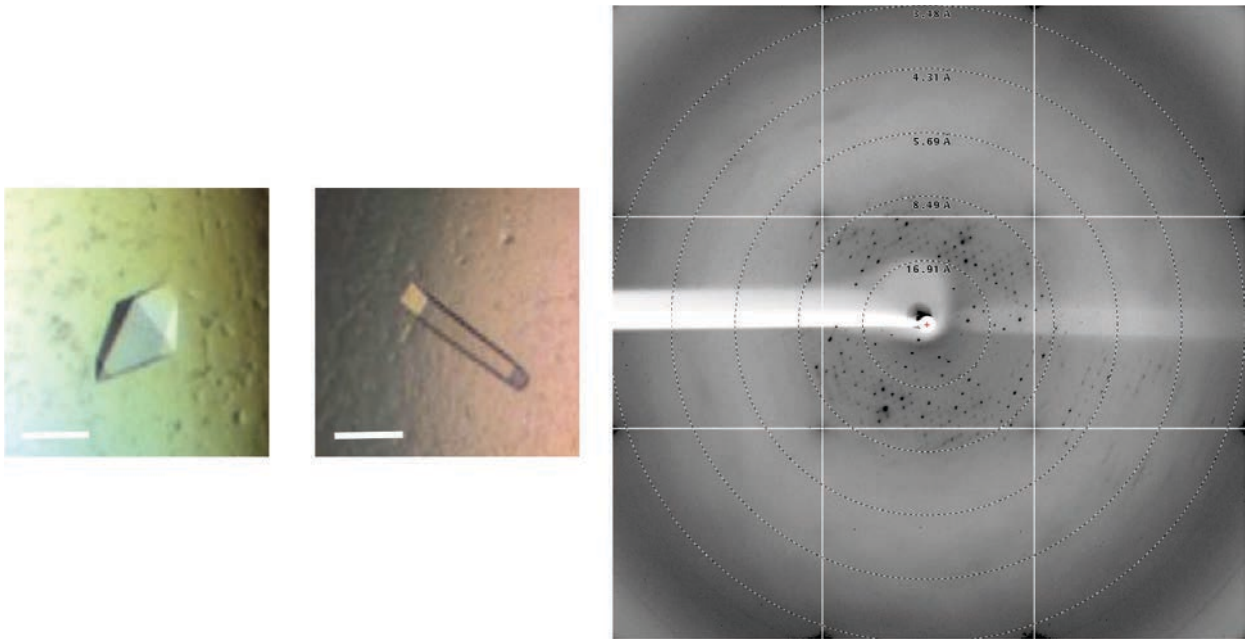


Figure 38 - ComGA crystals
(A) ComGA crystals obtained. Scale bar is 1 µm. (B) Diffraction pattern of a ComGA crystal. The pattern is strongly anisotropic, only in one direction some spots at 5 Å are detected.

Despite extensive trials to optimize crystal quality, the best X-ray diffraction data obtained from these crystals did not reach a sufficient resolution to solve the ComGA structure.

B. *Oligomerisation state*

The purified ComGA oligomerisation state was tested by analytical ultracentrifugation (AUC) with the help of Bertrand Raynal from the biophysics facility headed by Patrick England at Institut Pasteur. Macromolecules when subjected to the ultracentrifugation force migrate differently according to their size and shape. By monitoring the sedimentation velocity of the most abundant macromolecule populations in a given sample, their molecular weight can be calculated by considering the protein as a globular object. The different oligomerisation states of a protein from the same sample can consequently be separated and identified by AUC (materials and methods II.D.).

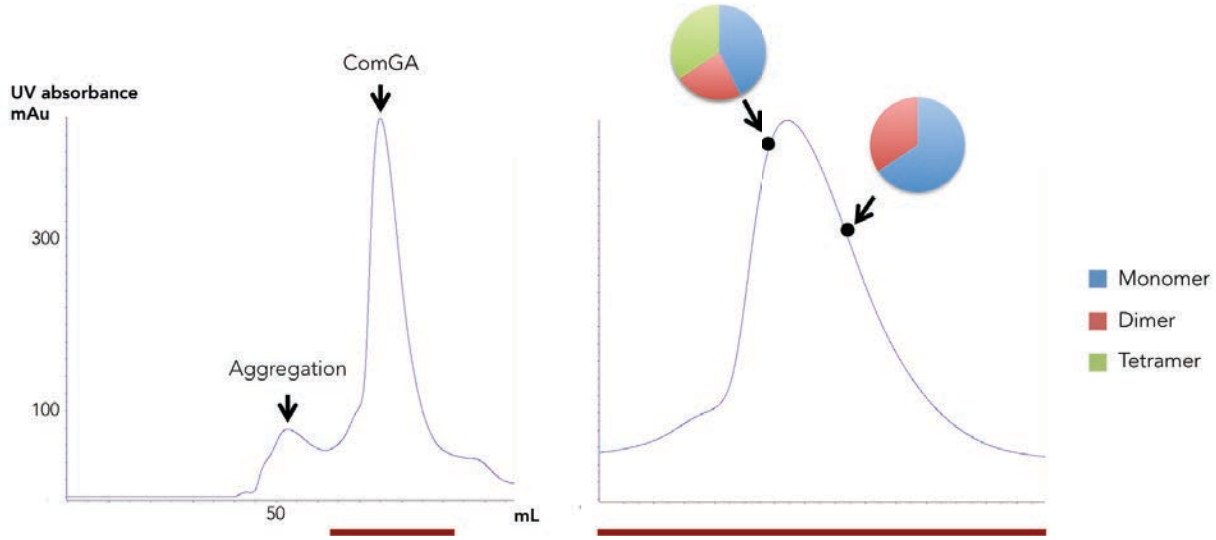


Figure 39 - Purified ComGA oligomerisation state.
Gel filtration profile of affinity-purified ComGA. The different ComGA oligomerisation states were not separated on this column: AUC results are shown for 2 samples taken on different sides of the ComGA peak.

We obtained from these experiments that ComGA is in majority a monomer in solution, with a significant subpopulation of dimer and tetramer. Interestingly, the condition that gave rise to crystals was consistently obtained by concentrating on a 100 kDa cutoff membrane. 35 kDa ComGA monomers and dimers will go through this membrane, while the tetramers will be retained. Moreover, if an equilibrium exists between monomeric and tetrameric states, the

tetrameric state will be favored at higher ComGA concentrations. These results consequently suggest that the tetramer might be the ComGA oligomerisation state that is crystallizing.

C. ATPase activity

Purified ComGA was clearly active for ATP hydrolysis when tested for ATPase activity (materials and methods II.E.). The ATP hydrolysis rate was low (~ 9.5 nmol ATP min^{-1} mg^{-1}) which is common for recombinant ATPases: in comparison the archeal GspE (T2S) had an hydrolysis rate of only to 2.5 nmol ATP min^{-1} mg^{-1} when purified with a C-terminal His-tag (Yamagata et al., 2012), and TrwD (type IV secretion system) had an hydrolysis rate of 4.5 nmol ATP min^{-1} mg^{-1} when purified with a C-terminal GST-tag (Ripoll-Rozada et al., 2012). Although activity is a good sign indicating that proper folding of the enzyme has been maintained throughout the purification process, it can potentially hinder crystallization, that requires the protein in a stable conformation. Two strategies were used to address this aspect: a point mutation in the catalytic site of ComGA to prevent ATP hydrolysis, and the use of non-hydrolizable ATP analogs.

The lysine in the consensus $G(X)_4GKT$ sequence (walker A motif) was replaced by an alanine, to inactivate ATP binding. Unfortunately, this mutant was not as stable as the native protein, and could not be concentrated at levels required for crystallization trials.

The non-hydrolizable ATP analogs ATPɣS, ATPαS, AMP-PNP, AppNHp, AppCp, ApCpp were tested by addition to the concentrated protein before setting up crystallization screens (see materials and methods II.C.ii). The diamond-shaped crystals were obtained in presence of ATPɣS, but always diffracted at even lower resolution than crystals of ComGA alone.

D. Perspectives

In parallel to ComGA, crystallization trials were performed on ComGB in our lab by Natalia de Val and Chiara Rapisarda. Unfortunately, this integral membrane protein was reluctant to crystallization, despite our ability to purify it in high quantity and high purity. Electron microscopy could provide an alternative approach to get structural insights about this pilus assembly platform. By successive pulldown on comGA and ComGB, we obtained preliminary data showing the *in vitro* interaction of ComGA and ComGB. Each protein alone, even in an oligomeric form, would be a too small object for 3D reconstruction using electron microscopy. However a stable complex containing multiple copies of ComGA (most probably a ComGA hexamer, like all ATPases from this family) and ComGB would be big enough to obtain an informative structure.

Therefore, I constructed a *S. pneumoniae* strain carrying a *comGA* gene at the *comG* locus expressing a N-terminally Strep-tagged protein (see Laurenceau et al., 2013, same cloning procedure as for RL003). Purifying ComGA from its *in vivo* environment under mild conditions is the best strategy to obtain the protein in a functional complex with its partners. However the ComGA-Strep mutant was completely deficient for transformation (see Laurenceau et al., 2013 for transformation efficiency assay procedure). The Strep-tag might impair ComGA function, maybe preventing its interaction with a partner, or its coding sequence might disrupt the *comG* operon structure. Another tag should be tried, and alternatively a complementation of a ΔComGA strain with a tagged ectopic copy of *comGA* could solve this problem.

III. ComGC, the major pilin

We have shown in chapter I that the *comG* operon encodes a *bona fide* T4P made of the major pilin ComGC, which is the first competence pilus directly observed in a Gram-positive bacterium. Therefore, ComGC is a particularly interesting target for structural studies. Its structure could be used to generate a model of the *S. pneumoniae* transformation pilus (see introduction section V.B.v). After biochemical validation this model could provide insights on the surface chemistry of the transformation pilus. In particular we showed that transformation pili are able to bind DNA. However no information on the DNA binding mechanism were obtained from these experiments. ComGC is a strong candidate for mediating DNA binding since it is the major pilin and it is exposed on the pilus surface.

A. Biochemical characterization of the major pilin

The full-length mature ComGC (ComGC^{mature}) was heterologously expressed in *E. coli* (plasmid pBSSB36), extracted from the membrane using detergents, and successfully purified by Natalia de Val, a former postdoctoral member of the lab. Crystallization trials were performed, but did not yield any crystallization hits. The fact that the protein could only be produced at low yields, and that detergents must be kept in the solution to prevent aggregation of the hydrophobic transmembrane segment, certainly add a great difficulty to the crystallization process.

AUC experiments could be performed, and showed that ComGC^{mature} is found in solution in a monomeric state (69%), trimeric state (27%) and hexameric state (4%). Although it needs further experimental confirmation, these oligomerisation states suggest that purified ComGC^{mature} could initiate its own helical polymerization in solution. Indeed, the T2S pseudopilus model from *K. oxytoca* (Campos et al., 2010; Nivaskumar et al., 2014) and T4P model from *N. gonorrhoeae* (Craig et al., 2006) show interaction of each pilin protomer (P) with

protomers above (P+1, P+3 and even P+4) and below. No higher molecular structures were detected, suggesting that spontaneous polymerization cannot exceed the hexameric state.

I consequently engaged in the resolution of the ComGC soluble domain structure (ComGC^{soluble}). In this construct the N-terminal transmembrane segment from ComGC was removed. Due to their small size and the absence of detergents in the purification process, type IV pilin soluble domains have been much easier to crystallize than their full-length counterparts (Giltner et al., 2012). However, the loss of the N-terminal helix can certainly affect the global structure, which imposes an extra-carefulness in the use of the structure obtained.

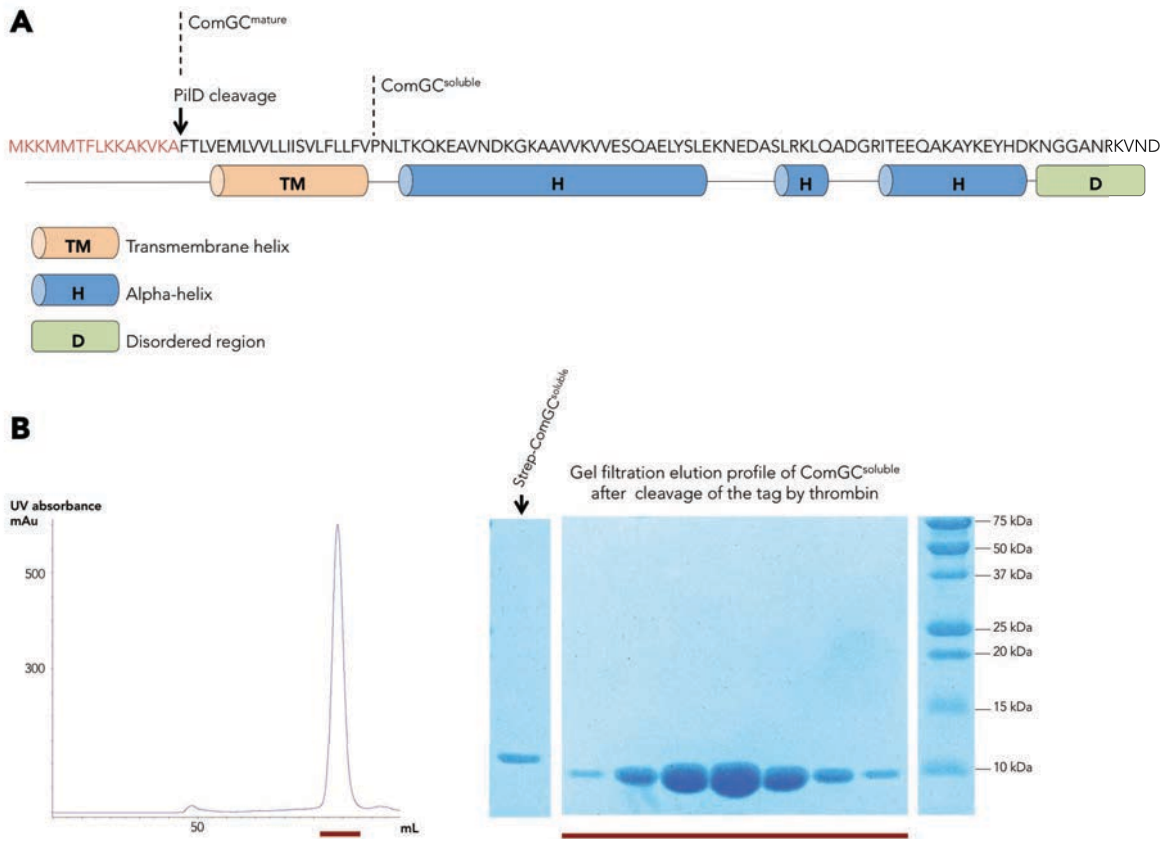


Figure 40 - ComGC purification

(A) ComGC secondary structure prediction generated by the Quick2D software; (B) Gel filtration profile and corresponding (red line section) SDS-PAGE of purified ComGC^{soluble} after thrombin digestion.

ComGC^{soluble} was expressed in high quantity, using either a N-terminal His-tag or Strep-tag (plasmids pBSSB39 and pBSSB31). The highest purity was obtained by Strep-Tactin (IBA)

resin purification of the Strep-tagged protein, and removal of the tag thanks to a thrombin cleavage site inserted between the Strep-tag and the first ComGC^{soluble} amino acid (see materials and methods II.B.iii.). Extensive crystallization trials were carried out with the different constructs, using a protein concentration as high as 100 mg/mL. No crystal was obtained. Interestingly, little precipitation was observed in the crystallization drops. Indeed, the vapor-diffusion method creates a slow concentration of the protein, pushing it towards its insoluble point. If not crystallization, precipitation is usually observed. ComGC^{soluble} turned out to be extremely soluble and recalcitrant to crystallization.

Alternatively, crystallization trials using *in situ* proteolysis were performed: dscifferent proteases are added directly to the protein sample before a usual crystallization screening (see materials and methods II.C.iii.). This technique can significantly increase the chances of crystallization of proteins with flexible parts. But no crystals were obtained.

B. ComGC involved in the DNA binding process?

ComGC^{soluble} forms the part of the pilus exposed to the solvent. We showed by whole protein mass profiling by high-resolution mass spectrometry in chapter I that this domain is present in the pilus with no post-translational modifications. DNA binding experiments were therefore conducted on ComGC^{soluble} in parallel to the structural studies.

i) Electrophoretic mobility shift assay

After incubating a protein with a labelled DNA fragment, electrophoretic mobility shift assay (EMSA) is the separation by electrophoresis of DNA alone from DNA attached to proteins, which have different electrophoretic properties. Generally, the DNA/protein complex is shifted up in the gel compared to DNA alone. Since *S. pneumoniae* has no preference for a particular

DNA sequence during transformation, a 40 nucleotide long random sequence was used (long enough to bind several ComGC^{soluble}) coupled to the DY-680 fluorophore in 5' (see materials and methods III.A.). The soluble domain of ComEA (ComEA^{soluble}) was expressed and purified from *E. coli* as already described (plasmid pBSSB30, materials and methods II.A. and II.B.) and used as a positive control. Indeed this protein from the DNA-uptake machinery has been shown to bind DNA (see introduction IV.A.).

Although ComEA^{soluble} consistently showed one or several shifted bands, no shift was observed for ComGC^{soluble} in the same experimental conditions, even at high DNA/protein molar ratio (up to 1000:1).

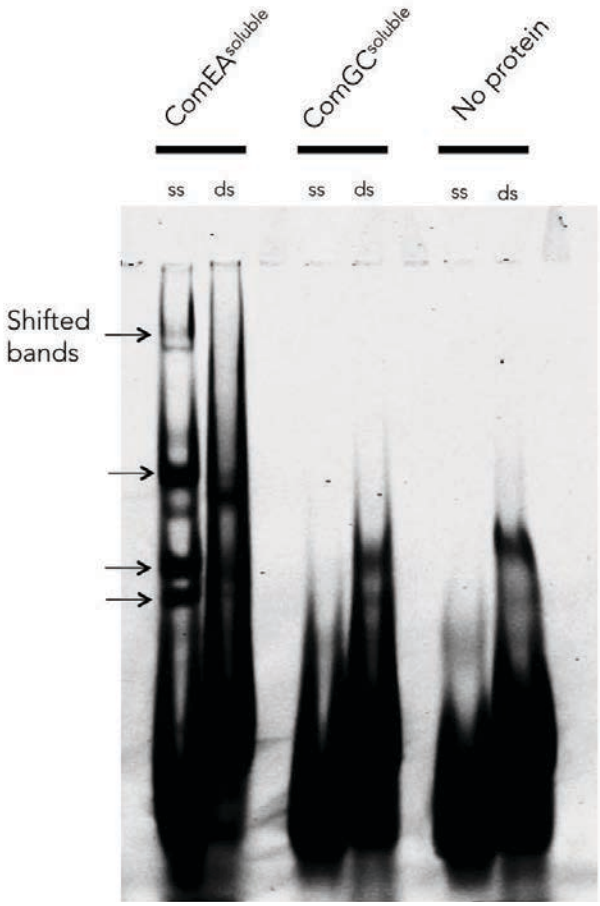


Figure 41 - EMSA on ComGC^{soluble}

EMSA gel revealed with a LICOR Odyssey scanner. The DNA probe emission is detected at 700 nm. ComGC^{soluble}, ComEA^{soluble}, and buffer alone were incubated with the DNA probe (ss for single stranded, ds for double stranded) before running on the gel. Shifted bands revealing interaction of ComEA^{soluble} with ssDNA are shown with black arrows.

Before concluding that ComGC^{soluble} cannot bind DNA *in vitro*, a more sensitive technique was used to probe this interaction.

ii) Fluorescence anisotropy

Fluorescence anisotropy experiments were performed with the help of Bruno Baron from the biophysics platform headed by Patrick England at Institut Pasteur.

Fluorescence anisotropy of a given fluorophore refers to the difference of intensity of its emission that can be observed in different directions. It can be measured by a fluorimeter by

recording emission at 2 different angles (usually separated by 180°) and plotting the ratio of the 2 intensities vs time. This anisotropy is particularly sensitive to the environment of the fluorophore: it is close to 0 when the fluorophore is free to move by Brownian motion in every direction. It increases when the fluorophore loses degrees of freedom. Typically, a DNA primer coupled to a fluorescent probe will see its fluorescence anisotropy increase when it binds to another molecule.

We used the same 40 bp primer that was used during the EMSA experiments. For anisotropy measurements, it was coupled to a fluorescein probe at its 5' end. ComEA^{soluble} was again used as a positive control. BSA was used as a negative control to check if crowding effect could trigger a change in fluorescence anisotropy. Protein concentration was gradually incremented every 5 minutes before measuring fluorescence anisotropy of the probe (see materials and methods III.B.). ComEA^{soluble} clearly had concentration dependent effect on the DNA fluorescence anisotropy, while no change could be detected with ComGC^{soluble} even at high concentration. Results are too preliminary to determine the DNA binding affinity of ComEA, however they clearly show no interaction of ComGC^{soluble} with DNA.

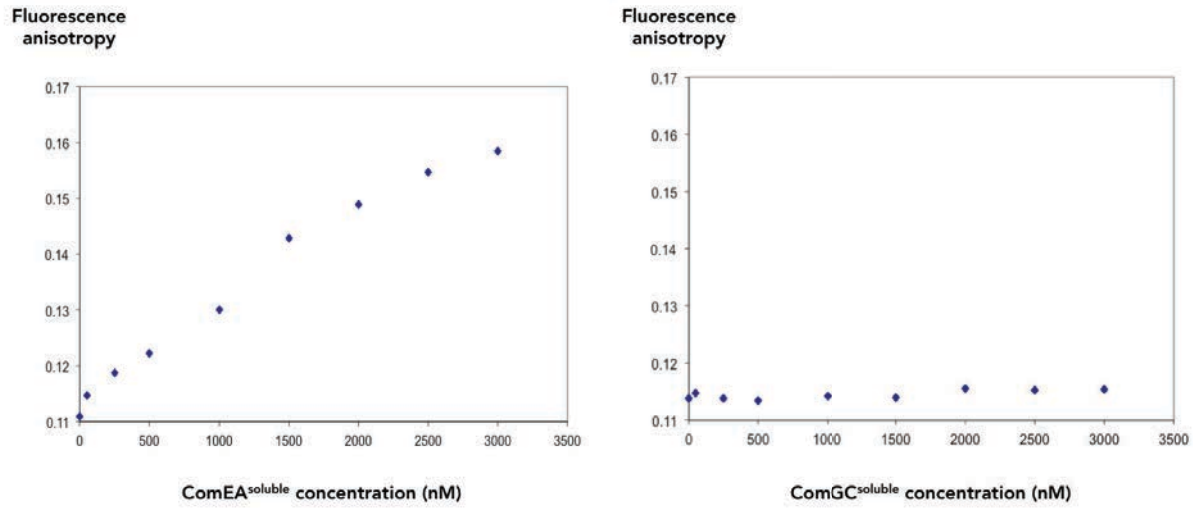


Figure 42 - Fluorescence anisotropy measurement of a DNA probe in presence of ComEA^{soluble} and ComGC^{soluble} increasing concentration
In Y axis the measure of fluorescence anisotropy is a dimensionless value, ratio of the fluorescence monitored at 2 different angles.

We concluded from these experiments that ComGC^{soluble} does not bind to DNA on its own. Several explanations can be raised for this result:

- ComGC^{soluble} does not bind DNA *in vivo*.
- ComGC^{soluble} is subjected to a conformational change necessary for DNA binding when polymerized in the pilus.
- ComGC^{soluble} purified from *E. coli* is not properly folded.
- ComGC^{soluble} interaction with DNA is too weak to be detected by the two techniques we used. The pilin alone could be unable to bind DNA significantly, even though the pilus could. As postulated for *N. gonorrhoeae* T4P, DNA binding could be mediated through a positively-charged groove on the pilus surface formed by the addition of multiple pilin monomers (see introduction part V.C.). Our results cannot exclude that the pilin monomer could be involved in the formation of a DNA binding pattern on the pilus surface.

C. *Structural study by Nuclear Magnetic Resonance*

Since ComGC^{soluble} could be produced in high quantity, very pure and very stable, we undertook its structural study by NMR, in collaboration with Nadia Izadi and Idir Malki from the NMR of biomolecules unit headed by Muriel Delepierre at Institut Pasteur.

As a preliminary study we determined by AUC the oligomerisation state of ComGC^{soluble} and the homogeneity of the sample in the range of concentration used for NMR. Fortunately, ComGC^{soluble} stays strictly monomeric in solution, showing that the trimer and hexamer states of ComGC^{soluble} were mediated by the N-terminal hydrophobic helix.

For its structure determination by NMR, ComGC^{soluble} needed to be ^{13}C and ^{15}N labeled. Double labeled ComGC^{soluble} was produced by growing *E.coli* in minimal medium in the presence of ^{13}C glucose and ^{15}N ammonium chloride as the sole carbon and nitrogen source (see materials and methods II.A.ii.).

The first structural information of ComGC^{soluble} was obtained by analysis of its ^{15}N-^{1}H HSQC (Heteronuclear Single Quantum Coherence Spectroscopy) spectrum. In this spectrum, each residue of the protein (except proline) is represented by a signal resulting of the correlation between a nitrogen and its attached proton. In ComGC^{soluble} HSQC spectrum, the peaks are large with a weak resolution, indicating the presence of multiple conformations. Moreover, the ^{1}H resonances are not well dispersed ("crowded" between 7.5-8.5 ppm) suggesting that a significant part of the protein is exposed to the solvent and therefore disordered. These spectral properties were modified after 6 days. The modification of the shape of the peaks reflects structural and/or chemical changes. Mass spectrometry analysis of ComGC^{soluble} at different time points revealed that the C-terminal end was gradually degraded. This C-terminal end, predicted to be a disordered region represents a good target for proteases which potentially

remain present in trace amount in the purified sample. In order to avoid this degradation, I cloned and expressed a shorter version of ComGC^{soluble} lacking the 9 C-terminal amino acids: this shorter version was unfortunately very poorly expressed. These 9 amino acids, although disordered, could be required for proper initial folding of the protein.

We consequently tried to remove these nine C-terminal amino acids by limited proteolysis on the entire ComGC^{soluble}. This technique consists in the addition of proteases at very low concentration in a protein sample, so that they digest primarily the exposed flexible parts of the protein of interest, leaving its compactly folded core intact (materials and methods II.B.iv.). Trypsin seemed to remove a little piece of ComGC^{soluble} after 30 min incubation. However, mass spectrometry analysis revealed that trypsin was creating a heterogeneous population, with proteins digested from both the C-terminal and N-terminal ends. Carboxypeptidase Y was consequently tried, because this protease can only digest from the C-terminal end: even at high concentration of the protease, no significant digestion of ComGC^{soluble} was obtained.

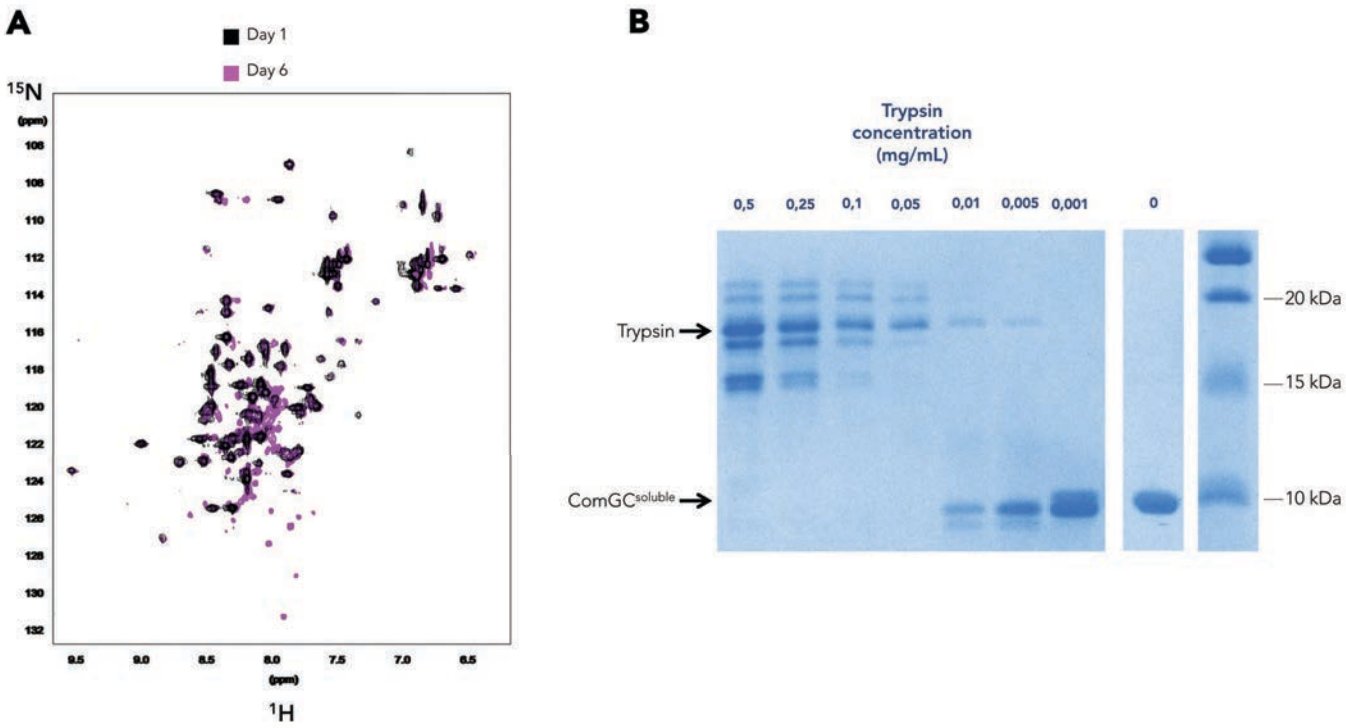


Figure 43 - HSQC spectra of soluble ComGC^{soluble}, and limited proteolysis

(A) Overlay of ComGC^{soluble} HSQC spectra after 1 day (black) and 6 days (purple), revealing a drastic change in the protein overall structure; (B) SDS-PAGE of ComGC^{soluble} after limited proteolysis with trypsin. Above 0,01 mg/mL, the trypsin entirely digest ComGC^{soluble}. At 0,005 mg/mL trypsin, ComGC^{soluble} is cut to a lower band.

Alternatively we observed by NMR that decreasing the temperature (from 30 to 20°C) and the addition of a mixture of protease inhibitors helped to avoid the degradation and to stabilize the sample for few days. Finally, several samples prepared at the same conditions were used for NMR data collecting and almost all backbone atoms (^{1}H, ^{15}N, ^{13}Cα, ^{13}Cβ, ^{13}CO) were assigned. However, the analysis of heteronuclear 3D experiments confirmed that the protein was highly flexible adopting several conformations (presence of multiple sets of peaks belonging to the same residue).

This structural dynamics of the protein was further confirmed by hydrogen-deuterium exchange experiment. This experiment allows to determine the solvent accessibility of amide protons of

protein, by monitoring how fast the protein hydrogens are exchanged with the deuterium of the solvent. It normally takes several weeks or months to exchange all ^{1}H to deuterium of a well structured protein. After one hour, all backbone proton amide of ComGC were exchanged by deuterium. These backbone atoms seem consequently not involved in stable hydrogen bonds and leave the whole protein accessible to the bulk solvent, suggesting again a very flexible state.

These structural features prevented us to go further in the structural study of the ComGC protein by NMR. The extreme flexibility might also explain the failure to detect DNA binding by this soluble domain.

The atoms assignment however will certainly prove useful for testing the interaction of $ComGC^{soluble}$ with any identified partner. Upon interaction with $ComGC^{soluble}$ atoms, a partner will change their chemical shift and the precise amino acids involved in the interaction can be detected. Despite the high protein flexibility, an interacting partner could stabilize $ComGC^{soluble}$ in one particular conformation. The addition of 4mM of a 40mer oligonucleotide on 0.5 mM of $ComGC^{soluble}$ was tested, without any change in the protein signature, further confirming the EMSA and fluorescence anisotropy results. Binding to calcium, as many type IV pilins do (Giltner et al., 2012) was also tested, and $ComGC^{soluble}$ was not chelating calcium ions.

DNA-uptake mechanism

I. Role of the minor pilins

From the current knowledge of Gram-negative T4P and T2S systems, two hypothesis can be put forward about the role of the minor pilins in the *comG* operon: 1) minor pilins are essential for initiation of the pilus polymerization; and/or 2) minor pilins are incorporated in the pilus and directly involved in pilus function, for example in DNA binding (see introduction part V.B.iii for 1 and part V.C. for 2).

An important question is whether or not the minor pilins are essential for pilus growth. If not, are they essential for DNA uptake? The study by Balaban and colleagues already mentioned in chapter I provides an answer (Balaban et al., 2014). Although they failed to visualize the transformation pilus, the authors generated knockout mutants of all the minor pilins and assessed the phenotype of each mutant. In particular, they detected the release of ComGC in the extracellular medium thanks to antibodies generated against the *in vitro* purified ComGC^{soluble} protein. ComGC release in the medium was abolished for each minor pilin mutant, showing that minor pilins are essential for pilus growth. Consistently with our results showing that the pilus is required for DNA uptake, they showed that transformability was abolished as well in all the mutants.

Most of the experiments in this section were done by Sahra Ouarti, a talented student preparing a professional bachelor degree, whom I had the opportunity to supervise all along her internship in the lab (6 months).

A. PilD processing and binary interactions

Annick Dujeancourt, technician in the lab, has largely contributed to the work presented in this section.

The soluble domain of each minor pilins was expressed recombinantly in *E. coli*, exactly as described for ComGC^{soluble} (plasmids pBSSB63, pBSSB64, pBSSB65; materials and methods II.A.). In contrast with ComGC^{soluble}, each of them was extremely unstable (either using a Strep-tag or a His-tag, expressed in the cytoplasm or the periplasm), and crystallization trials could not be attempted. Co-expression were tried (pBSSB66/51, pBSSB67/42, pBSSB67/51), and although expression was weak, interactions between ComGD^{soluble}/ComGG^{soluble} and ComGF^{soluble}/ComGG^{soluble} could be shown by two sequential affinity chromatography steps pulling successively on the tag of each pilin. We discovered only later the existence of ComGE, that was wrongly annotated in *S. pneumoniae* R6 strain genome. Testing the interaction with this fourth minor pilin is in process and is required to conclude whether all minor pilins interact together to form a quaternary complex.

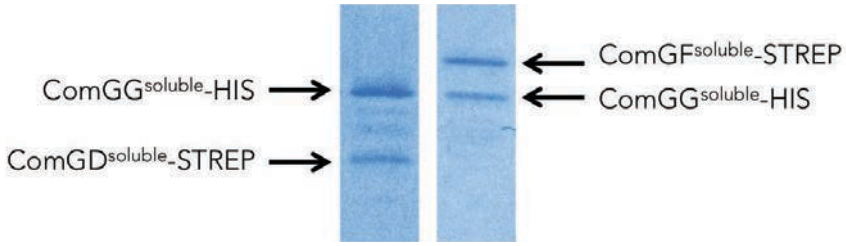


Figure 44 - Minor pilins binary interactions
Data courtesy of Annick Dujeancourt. Two minor pilins soluble domain having either a Strep- of His- N-terminal tag were co-expressed in E. coli. SDS-PAGE of the eluted fraction after serial Strep-tag and His-tag affinity chromatography steps is shown.

The cleavage motif by the prepilin peptidase PilD is well conserved in all minor pilins except ComGG (see introduction part V.D.). We therefore tested the ComGG cleavage by co-

expressing PilD with the full length pilin in *E. coli.* PilD was indeed cleaving the ComGG degenerated signal peptide, as obtained in *B. subtilis*.

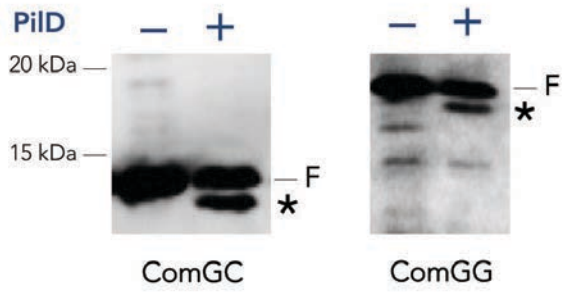
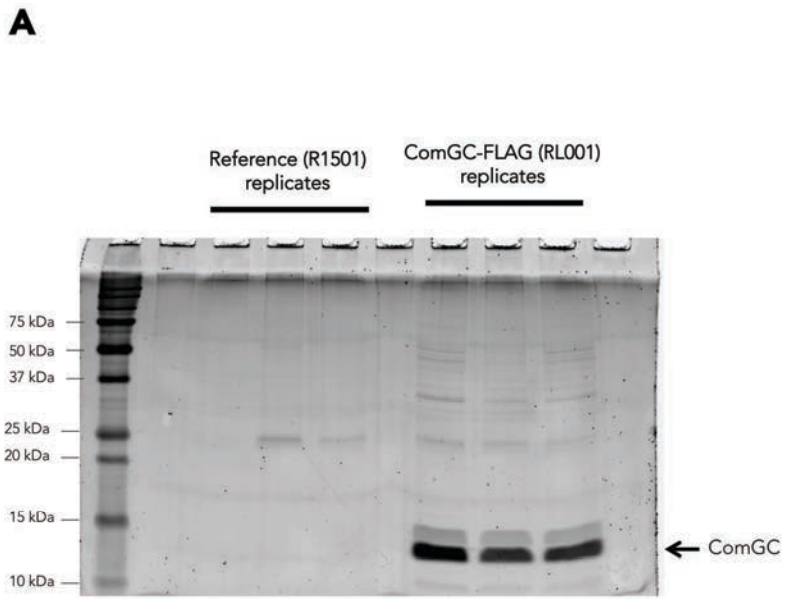


Figure 45 - In vitro processing by PilD
*Data courtesy of Annick Dujeancourt. PilD cleavage of ComGC and ComGG shown by western blot analysis. Each pilin (Full length band indicated by F) is partially processed to a lower band (indicated by *) when co-expressed with PilD in E. coli.*

B. Mass spectrometry analysis of purified pili

Taking advantage from the fact that we were able to purify the transformation pili directly from *S. pneumoniae*, broken in pieces but still polymerized, the presence of the minor pilins inside the pilus was tested by mass spectrometry in collaboration with Joseph Gault, Christian Malosse and Julia Chamot-Rooke from the Mass spectrometry of biomolecules unit at Institut Pasteur. The difficulty in these experiments comes from the extreme sensitivity of the mass spectrometry analysis, able to detect trace amounts of minor pilins coming from a contaminating cell debris on the resin. The purification protocol on the anti-Flag resin was consequently optimized to minimize cell lysis (see materials and method II.F.), and triplicates experiments were done using the reference strain R1501 as a negative control. These cells make pili without the Flag-pilin which consequently go through the resin. Any protein retained on the resin for these samples was considered as a contaminant. To avoid any co-purification mediated by the DNA binding of the pili, DNAse was incubated overnight with the sheared

fraction, and the resin was extensively washed with the buffer supplemented with 1M NaCl before eluting.



B

Accession	Description
P0A2W1	Acyl carrier protein
P0A472	50S ribosomal protein L7/L12
P64031	Elongation factor Tu
P66582	30S ribosomal protein S5
Q8CWM8	30S ribosomal protein S2
Q8CWP9	Cell division protein DivIVA
Q8CWR9	30S Ribosomal protein S1
Q8CWS4	tRNA (guanine-N(1)-)-methyltransferase
Q8CWU8	50S ribosomal protein
Q8DPP2	Tagatose 1,6-diphosphate aldolase
Q8CWN6	Glyceraldehyde 3-phosphate dehydrogenase
Q8DMY4	General stress protein GSP-781
Q8DNN3	Putative uncharacterized protein
Q8DN86	ComGA
Q8DN88	ComGC
Q8DN89	ComGD
	ComGE
Q8DN90	ComGF

Figure 46 - MALDI analysis of proteins co-purified with the transformation pili

(A) Lavapurple-stained SDS-PAGE of the replicates used for the analysis.
(B) Final list of proteins found only in the pili fraction. ComGC is included although it was found in minor amount in the negative control.

A list of only 17 proteins detected in the pilus fraction was obtained. Most of them were still clearly contaminants, corresponding to proteins in high amount within the cells. ComGD, ComGE, and ComGF were detected in the pili fractions in this final triplicate experiment, as well in all the previous optimization experiments. These experiments gave us confidence in the fact that these minor pilins are indeed incorporated in the pilus, while ComGG is not. However since membrane proteins and ComGA are still found within the reduced list, these results need to be complemented with other experiments to clearly localize the minor pilins within the pilus, and rule out the possibility that they could just be stuck to its base. Indeed the PilV and PilX minor pilins of *N. meningitidis* have recently been suggested to exert their function from the periplasm rather than from the pilus (Imhaus et al., 2014).

C. Localization of the minor pilins

In parallel to the MS analysis experiment, genes corresponding to each each mature minor pilin with a C-terminal Flag-tag were placed in the CEP cassette for *in vivo* studies, (same procedure as for RL001 in Laurenceau et al, 2013). Additionally, a *comGG-Flag* construct at locus was made, the only *comG* gene for which it was feasible since all genes are overlapping in the operon.

i) ComGG

The ComGG-Flag strain (RL005) had a normal piliation phenotype seen by TEM (Figure 47), but had a ~ 10 times reduced transformability. Pulling on ComGG-Flag on the anti-Flag resin showed no pilus purification when the sample was checked by TEM (Figure 47), suggesting consistently with the MS results that the pilin is not incorporated to the transformation pilus. Immunogold labeling experiments were performed using the same protocol used for the labeling of comGC-Flag in the pilus (see protocol in Laurenceau et al, 2013), and absolutely no labeling of the pili was observed despite extensive trials. By western blot analysis (see shearing experiment protocol in Laurenceau et al, 2013) ComGG-Flag was detected inside the cell fraction in low amounts, and in residual amount in the sheared fraction (Figure 47). Since different techniques correlate to show that ComGG is not inserted in the pilus, we think that this residual amount could be due to release of ComGG by cell lysis in the medium.

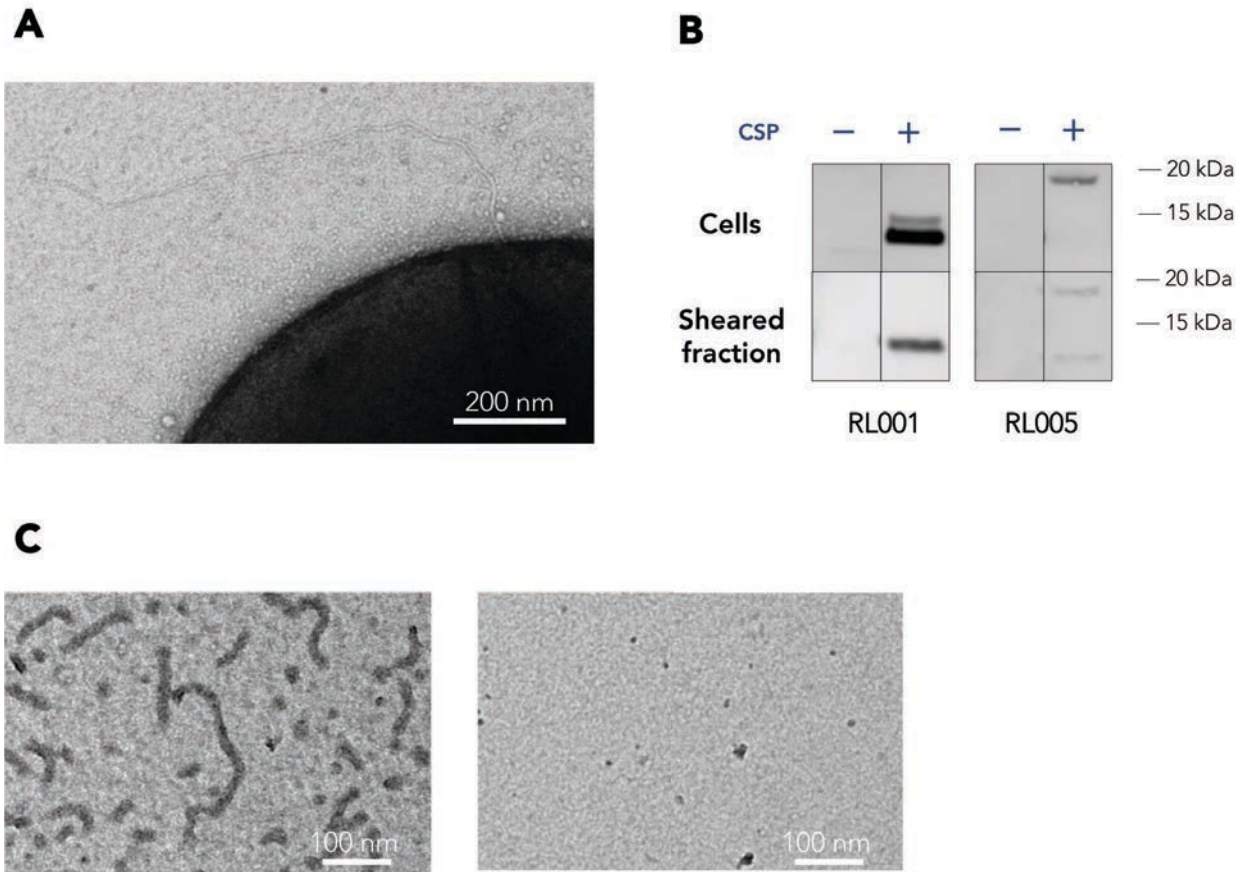


Figure 47 - RL005 strain analysis

(A) Negative stain picture of competent RL005 cells. (B) anti-Flag Western blot of RL005 (ComGG-Flag) and RL001 (CEPx-comGC-Flag). The 12 kDa band in RL001 is ComGC-Flag, the 19kDa band in RL005 is ComGG-Flag. The lower band detected in RL005 sheared fraction is most probably an aspecific band, as it was not detected in other western blots. (C) Negative stain picture of purified RL001 sheared fraction on anti-Flag resin (left), compared to RL005 sheared fraction purified in parallel (right).

Altogether these data suggest a special role in transformation for the ComGG minor pilin, as it was concluded for *B. subtilis* ComGG by Mann and colleagues (Mann et al., 2013). Knowing that ComGG is not inserted in the pilus, if piliation in RL005 is truly unaffected as it seems by TEM (same number of pili), but transformability is, then this functional defect would suggest a role for ComGG other than purely in piliation. There is too little information to speculate on this role for now. However the similar observations in *B. subtilis* and *S. pneumoniae* concerning this pilin are intriguing, and suggest a conserved role in DNA uptake that will require a proper study to be elucidated.

ii) On the importance of the minor pilins mutual interactions.

A disappointing result came from the expression of each minor pilin inside the CEP cassette: the cellular levels of pilin detected by western blot were either low or null, except for ComGE (Figure 48). The CEP cassette is integrated ectopically in the chromosome, and consists of a strong promoter induced specifically during competence (the *ssbB* gene promoter, see introduction part III.C.) followed by the protein coding sequence, and a kanamycin resistance gene. Expression of ComGC-Flag in exactly the same conditions provided high cellular levels, and competing of the tagged pilin with the native ComGC for integration in the pilus. The ratio ComGC-Flag:ComGC inside the purified pili was estimated to approximately 1:10 by mass spectrometry and SDS-PAGE (see Laurenceau et al., figure 5.B and 5.C). Different explanations can account for the fact that the minor pilins are not expressed in similar quantities as ComGC-Flag:

- Genes are regulated post-transcriptionally.
- An inhibition occurs in the transcription of genes under the CEP cassette promoter (secondary structure of the mRNA for example)
- Proteins are degraded post-translationally.

Comparing the cellular levels of RL005 (ComGG-Flag at locus) and SO001 (CEPx-ComGG-Flag) revealed that the cellular level of ComGG-Flag were higher when expressed in the *comG* operon rather than in the CEP cassette. Therefore minor pilins on their own could be unstable, and/or unable to fold properly when expressed in the cassette. We showed that they interact with one another most probably to form a complex as it was shown for *K. oxytoca* T2S minor pilins. The overlapping operon structure might be required for instantaneous interactions of the minor pilins after translation. This hypothesis is consistent with the fact that minor pilins when expressed recombinantly in *E. coli* were all extremely unstable, in contrast with ComGC. Immunogold experiments were still performed to see if some pilins were incorporated in the

pilus at levels too low to be detected by shearing experiments, however no labeling was observed. The *CEPx-comGE-Flag* strain was obtained only recently. It is very promising since it shows cellular levels comparable to ComGC-Flag. Further experiment are currently under way for this construct.

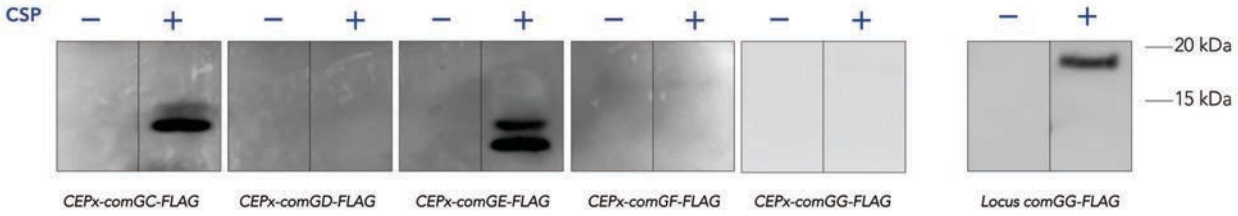


Figure 48 - Ectopic expression of the minor pilins

Expression test performed by western blot anti-Flag on cell lysate after 15 min incubation at 37°C in presence or absence of the CSP peptide. The 2 bands detected for ComGC-Flag and ComGE-Flag correspond to PilD processing, the upper band being the unprocessed pilin.

Interestingly, Giltner and colleagues provided similar observations in an in-depth study of the minor pilins of the twitching motility T4P in *P. aeruginosa* (Giltner et al., 2010): when expressing the minor pilins ectopically on their own, they are able to complement a pilin deletion mutant at locus for pilus growth and twitching motility. However this is only the case with a low expression. If in the same complementation experiment the ectopic copy is overexpressed, function is not restored. Interestingly, this ectopic overexpression is not detrimental anymore to function when done in a functional T4P genetic background (with no minor pilin deleted). It is not mentioned however if like in our case they observe degradation of the surplus pilin.

They conclude from these experiments that only the formation of a complete minor pilin complex matters for function: once formed, it does not matter that one of the minor pilin is in surplus copies. However in a deletion mutant background, overexpression leads to drastic perturbation of the operon stoichiometry, and impedes proper assembly of the complex.

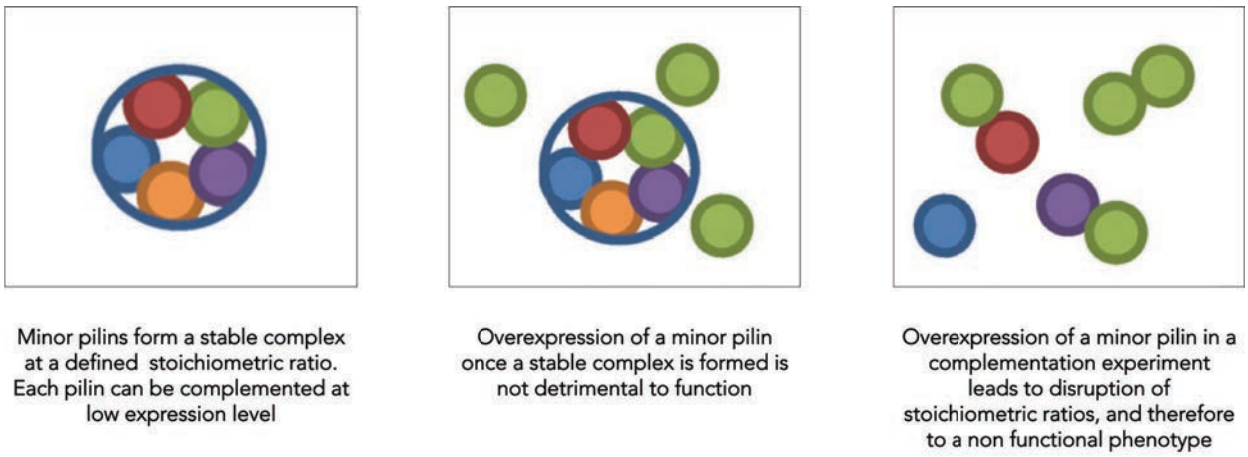


Figure 49 - Model for the minor pilins complex formation in P. aeruginosa

Adapted from Giltner et al., 2010. The different minor pilins are shown as colored circles. The stable complex formation is shown as a blue circle around the 5 minor pilins.

The general conclusion that can be drawn from all studies performed on clustered minor pilins (leaving aside cases of minor pilin expressed separately in the chromosome, like *N. gonorrhoeae* ComP) in different T4P systems (see introduction part V.B.iii and V.D) is that they are functioning together as a single functional unit, even if one of them always seems to play a special role (ComGG in competence, PulK in T2S, PilX in twitching motility).

The next experiments we devised take this aspect into account. Instead of expressing minor pilins on their own, we are now placing comGC-D-E-F-G-Flag, comGC-D-E-F-Flag, ComGD-E-F-G-Flag, and ComGD-E-F-Flag inside the CEP cassette. We hope that keeping the operon structure will form a stable complex and that tagged pilin will be incorporated in the pilus and detectable by immunogold labeling. Will overexpression of all pilins lead to increased number of pili? Will overexpression of all minor pilins only be detrimental to pilus growth? Will overexpression of all minor pilins except ComGG lead to a different phenotype? These experiments will provide crucial answers to understand the role of minor pilins in competence.

II. Is DNA uptake mediated by pilus retraction?

A. Background

T4P retraction was shown to be essential for DNA uptake of several Gram-negative bacteria (see introduction part V.D.). In Gram-positive bacteria the absence of an outer membrane suggests that the uptake mechanism could be significantly different. However, the discovery of a long T4P at the surface of *S. pneumoniae* suggests otherwise. Could retraction be involved to carry the bound DNA down to the translocation channel? Many *S. pneumoniae* strains have a thick capsule in addition to their peptidoglycane layer. A retractile µm sized pilus seems like an efficient option for importing DNA across this complex envelop.

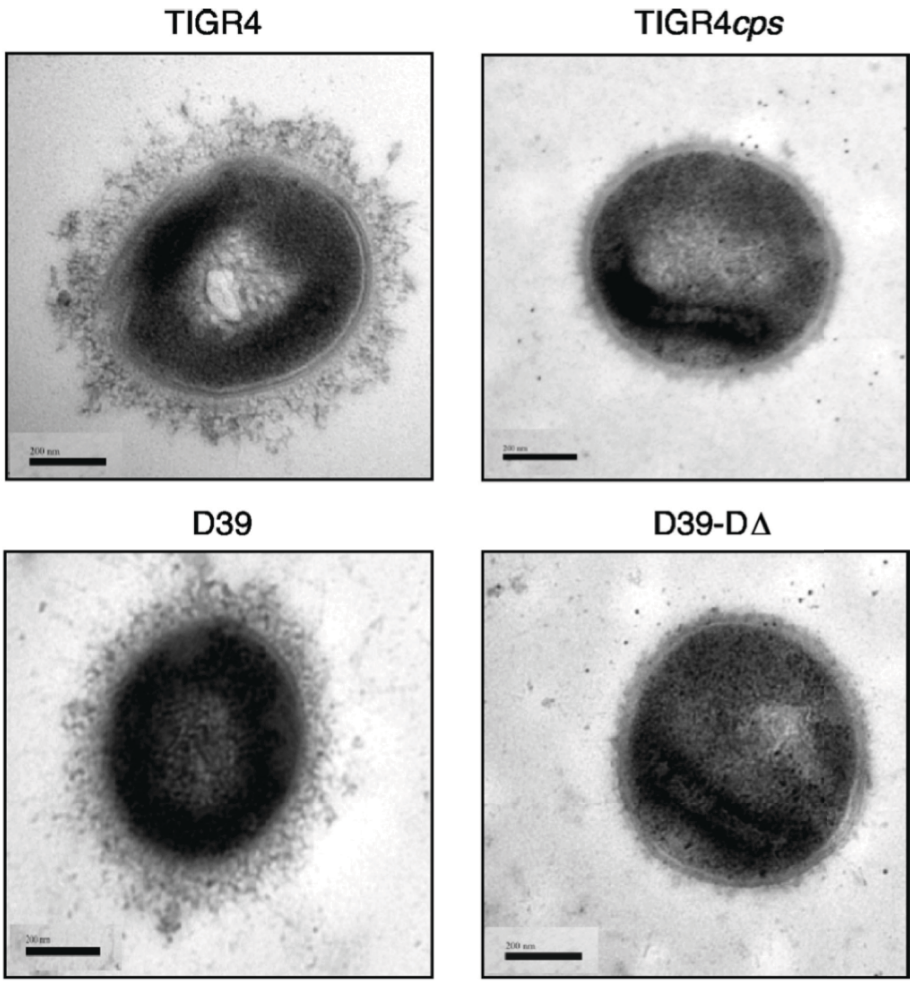


Figure 50 - S. pneumoniae capsule

Taken from Hyams 2010. Scale bars are 200 nm. Cells were cryofixed, and capsule contrast was enhanced using ruthenium red. TIGR4 and D39 strains are shown on the left, with their uncapsulated counterparts on the right.

No retraction ATPase is present in the genome of *S. pneumoniae*, as well as in other naturally transformable Gram-positive bacteria. However alternative ways to power retraction are plausible, like the proton motive force that has been shown to be involved in the DNA uptake through the cell envelop in *B. subtilis* (Maier et al., 2004). Moreover pulling on a DNA strand should not require strong forces, contrary to twitching motility during which the whole cell body has to be displaced by the retraction force.

B. Live immunofluorescence

Live immunofluorescence experiments were performed with the help of Jean-Yves Tivenez from the Imagopole platform headed by Spencer Shorte at Institut Pasteur (see materials and methods IV.A.).

i) Pilus fluorescent labeling

We started to look for possible retraction by live immunofluorescence because this technique was successfully used to visualize twitching T4P in action in P. aeruginosa (Skerker et al., 2001). Skerker and colleagues used a monofunctional succinimidyl ester coupled to the Cy3 fluorophore, a chemical that reacts with amino acids, and consequently label non specifically all proteins in the sample.

Figure 51 - Cy3 monofunctional succinimidyl ester labeling of P. aeruginosa T4P
Taken from Skerker et al., 2010. Each picture is separated by 6 s. Some pili are under tension (straight), while the one above is retracting.

We used NHS (N-HydroxySuccinimide) coupled to Cy3 (Amersham) to label competent pneumococci. The conditions were perfectly optimized for competence development, close to the conditions in which we were able to detect pili by immunofluorescence in chapter I. Unfortunately the transformation pili could not be detected using either confocal or TIRF microscopy. Only the cell bodies were labeled. Since twitching motility T4P are in high number

on the cell surface of P. aeruginosa, and have a tendency for bundling, we think the filaments observed in the figure above are actually bundles of T4P.

ii) Anti-Flag antibody labeling

We consequently tested a solution closer to the immunofluorescence protocol used previously, with the RL001 strain expressing an extra-copy of ComGC-Flag in the CEP cassette. To avoid a two-step antibody incubation we labeled the Flag tags directly with an anti-Flag A647 antibody (Life technologies). We checked first that we were able to detect the pili with this antibody by fixating the cells (see immunofluorescence protocol in Laurenceau et al., 2013). Indeed the filaments could be observed, although fluorescence was weak and a 5 second exposure was necessary to detect them.

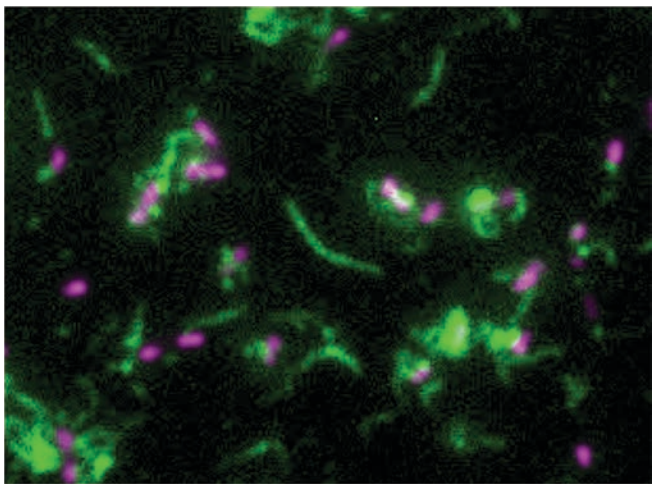


Figure 52 - Competent RL001 cells labeled with an anti-Flag A647 antibody

A467 fluorescence is displayed in green, overlaid with DAPI fluorescence in purple. Fluorescence is brighter in cell aggregates. Many pili are found detached in the medium.

The fluorescence could be detected in live imaging on the cell surface by confocal microscopy. Although filaments could not be clearly distinguished, a fluorescent focus was present on the surface of most bacteria, and absent in a non-induced sample. Fluorescence was significantly brighter on cell aggregates, suggesting that aggregates possibly contain tangled pili.

Even by performing 5 recording cycles per second, no dynamic of the foci could be observed. The same result was obtained by observing the samples on a spinning disk confocal microscope, in order to be more sensitive at lower time scales. It is however not yet possible to conclude from this single experiment that the pili cannot retract: the anti-Flag A647 antibody might be blocking retraction once it is attached, as it was observed for *P. aeruginosa* T4P (Bradley et al., 1972; Bradley et al., 1972).

iii) ComGC fused to the mCherry

We performed another attempt to visualize the pilus dynamics by live immunofluorescence by fusing the major pilin ComGC to a fluorescent protein. We chose the mCherry protein because it is small, monomeric, and highly photostable (Shaner et al., 2005). The cloning of an extra-copy of the *comGC* gene fused to the mCherry coding sequence inside the CEPx cassette was done by Rémi Fronzes. We fused the mCherry coding sequence behind the *comGC* gene using the Flag-tag coding sequence as a linker. Since the Flag-tag was accessible to antibodies at the surface of the transformation pilus, we rationalized that the mCherry in that position should be oriented towards the exterior of the fiber. After successful cloning, we had the disappointment to observe that absolutely no fluorescence could be detected after competence induction. Even by using a sensitive fluorimeter and waiting for up to an hour post induction. The plasmid *pCEP-comGC-Flag-mCherry* when introduced as a shuttle in *E. coli* was apparently leaking, and the mCherry fluorescence could be detected easily in these cells. We therefore think that the absence of fluorescence in *S. pneumoniae* is due to a maturation problem of the fluorophore. Indeed, the maturation process depends on multiple factors, and can vary from a few minutes to a couple hours even between different *E. coli* strains (Hebisch et al., 2013).

C. *Optical tweezers*

We then opted for a powerful technique, able to provide a conclusive result whether or not pili are retracting: optical tweezers. These experiments were performed in collaboration with Lena Dewenter and Berenike Maier from the biophysics group at the University of Cologne. Worldwide expert on this technique, Berenike Maier has obtained crucial information on the nanomechanics of T4P.

i) Setup description

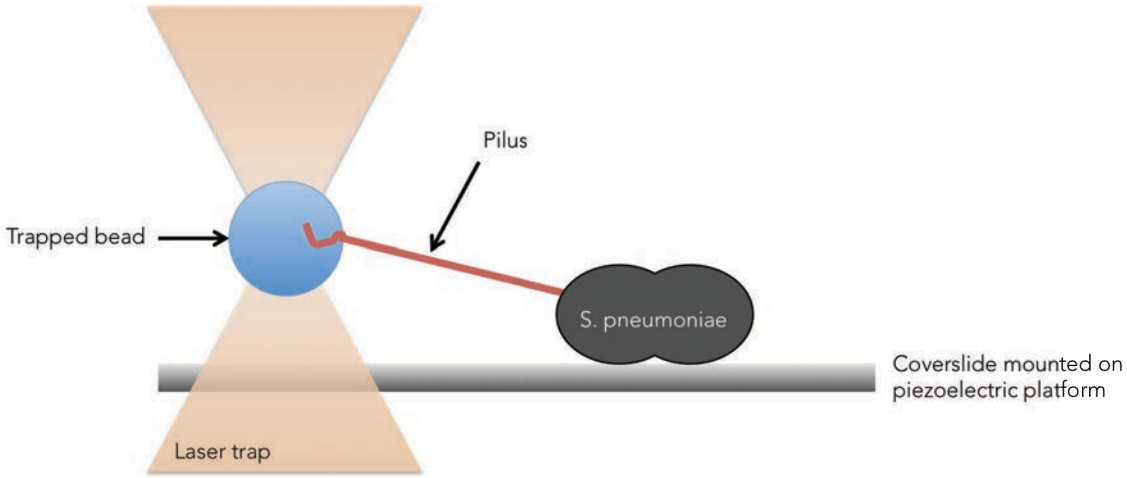


Figure 53 - Optical tweezers setup

The bacterium is stuck on the cover slide. If retraction occurs in this setup, the force of retraction will be measured by the displacement of the bead from the center of the optical trap. If retraction does not occur, the bacteria can be moved away from the bead, and the presence of the pilus will be detected if pulling of the bead in the direction of the movement is sensed.

The optical tweezers setup is thermoregulated at 37°C, so that *S. pneumoniae* cells could be probed by approaching the bead at different time point after CSP induction of competence, thereby ensuring that the competence window was not missed (see materials and methods IV.B.).

We used the strain R2546 (*CEPx-GFP*) expressing GFP during competence to check that competence induction was indeed occurring with our protocol. The strain was kindly provided by our collaborators from the LMGM at the University Paul Sabatier in Toulouse.

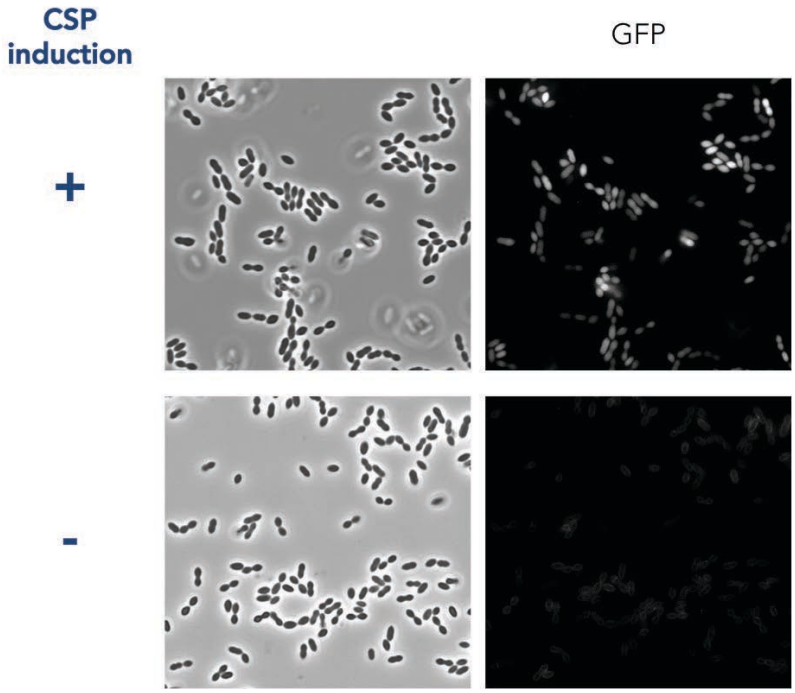


Figure 54 - R2546 strain
The R2546 strains expresses GFP in the competence state. Left panel correspond to bright field image, right panel to GFP fluorescence.

ii) Binding of the pili to the bead

First, we tried to see if transformation pili were naturally sticking to polystyrene beads, like the *N. gonorrhoeae* T4P do. No binding event could be detected despite extensive trials. We consequently used the RL001 strain, which has ~10 % of ComGC-Flag incorporated in its transformation pili, and is perfectly functional for DNA uptake (see chapter I), to bind pili on beads coated with anti-Flag antibodies (see materials and methods IV.B.). Less than 1% of bead approach event led to a binding event. A variety of conditions were tested to optimize binding without success:

- DNAse was added during competence induction to digest DNA released from cell lysis and try to avoid covering of the pili with DNA.

- The medium of incubation was changed to check for any binding improvement.
- Polylysine coated beads were tried, since pili are nicely deposited on polylysine-coated electron microscopy grids.
- Plastic covered coverslides were used, in case pili were sticking irreversibly to the common coverslides.
- An alternative strategy was used for bead coating: the mouse anti-Flag antibodies were bound to protein G coated beads (materials and methods VI.), and proper coating was checked by fluorescence using an anti-mouse fluorescent antibody.

The fact that *S. pneumoniae* competent cells harbor only one (or a few) pilus per cell should not be a major issue, since it was shown using an engineered *N. gonorrhoeae* strain that retraction from a single pilus could be detected and measured (Maier et al., 2002).

From the few binding events monitored, we determined that a force of 0,08 pN/nm was breaking the interaction, and most probably the pilus. This suggests that transformation pili could be ~10 times weaker than *N. gonorrhoeae* T4P, however more binding events are required to confirm this first approximation. Since we used a low trap stiffness (more sensitive) when scanning around the bacteria, more fragile pili should have been detected equally well.

Given the little number of binding event collected due to experimental limitations that we could not identify, we are still unable to conclude whether *S. pneumoniae* transformation pili are able to retract or not.

III. Universality of the long transformation pilus?

Morikawa et al. showed that the Gram-positive bacterium *S. aureus* is naturally transformable when expressing the sigma factor σ^H, controlling the expression of competence genes. The *sigH* gene is repressed *in vivo*, but is naturally duplicated in a fraction of *S. aureus* population and triggers the competence state (Morikawa et al., 2012). Van der Kooi-Pol et al. further showed by immunofluorescence experiments that the prepilin peptidase ComC (PilD homolog) is required for optimal surface display of ComGC. Filaments of ComGC were not detected with the polyclonal antibodies raised against ComGC (Van der Kooi-Pol et al., 2012), as we obtained in *S. pneumoniae*. In our case only the C-terminal Flag-tag enabled the visualization of *S. pneumoniae* transformation pili by immunofluorescence (Laurenceau et al., 2013).

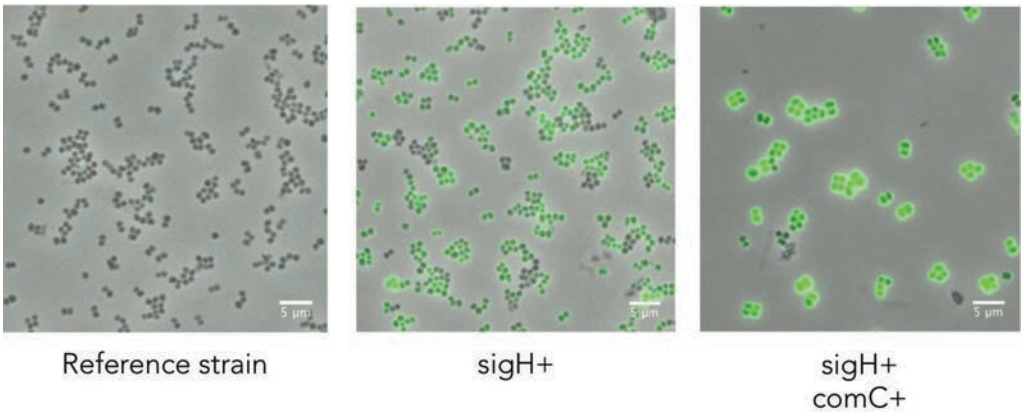


Figure 55 - Surface display of ComGC in S. aureus
Taken from Van der Kooi-Pol et al., 2012. Immunofluorescence microscopy on S. aureus Newman strain, using primary rabbit polyclonal antibodies specific for ComGC, and anti-rabbit A488 secondary antibodies (Like technologies). Strains carrying the pRIT:sigH plasmid are indicated by sigH+, and strains carrying the pCN51:comC plasmid are indicated by comC+.

We started a collaboration to study the transformation pilus morphology in *S. aureus* with Michel Debarbouille and Tarek Msadek from the Biology of Gram-positive pathogens unit headed by Patrick Trieu-Cuot at Institut Pasteur. Using the same plasmids as in the Van der Kooi Pol et al. study, we visualized filaments on *S. aureus* cells strikingly similar to the *S. pneumoniae* transformation pilus by TEM.

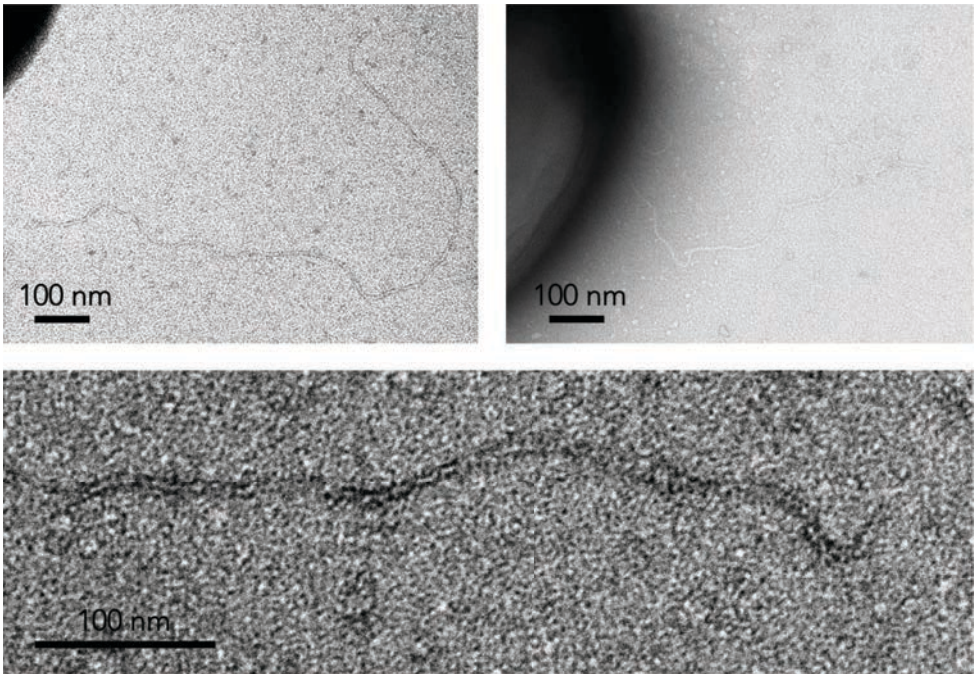


Figure 56 - Filaments observed by TEM on S. aureus expressing σ^H and ComC

Some filaments like in the top right panel were attached to the cell surface.

However we failed to reproduce this observation in many trials, playing on the induction conditions of ComC expression. Filaments were observed only one more time after reinsertion of both plasmids pRIT:sigH and pCN51:comC in the reference strain. Since pRIT:sigH constitutively expresses σ^H, we hypothesized that continuous induction of competence genes might lead overtime to a defect in pilus biogenesis. Therefore, we will reiterate the experiments after construction of a plasmid carrying only one inducible plasmid, expressing both *sigH* and c*omC*.

These preliminary results are encouraging. If confirmed they will show that the transformation pilus is a surprisingly conserved appendage among Gram-positive bacteria. However a ComGA mutant as well as immunogold labeling of the filament are necessary before drawing any conclusions.

Discussion and perspectives

The goal of my PhD project was to study the mechanism of action of the DNA uptake machinery in the major human pathogen *S. pneumoniae*. In particular, I focused on the *comG* operon which is highly conserved in Gram-positive bacteria.

I started by investigating which kind of molecular device was built by this operon, since it belongs to the ubiquitous and diverse family of T4P systems. Through the course of highly effective collaborations, I could show that the pilin ComGC is assembled in a micrometer-sized, DNA-binding pilus on the surface of competent *S. pneumoniae*. I confirmed the actual T4P nature of these appendages, and showed that they are absolutely required for DNA uptake.

A unified mechanism for DNA uptake in bacteria?

These findings have many implications, primarily for the general understanding of the DNA uptake mechanism. While T4P were known to be responsible for initializing DNA uptake in Gram-negative bacteria, it has remained unclear how transforming DNA is captured by Gram-positive bacteria. The prevailing hypothesis has been that the transformation pilus is a short pseudopilus that destabilizes the cell wall peptidoglycane for DNA entry (Chen et al., 2004). Given that *S. pneumoniae* is the first Gram-positive species in which the transformation pilus was directly observed, not much credit is left to support this hypothesis. The group of David Dubnau, who was the first to provide data to support the existence of a pseudopilus in *B. subtilis* (Chen et al., 2006) has been trying for years to visualize the transformation pilus/ pseudopilus without success. After our publication, they started again the search for a transformation pilus on competent *B. subtilis*, with no more results to date. An additional difficulty with *B. subtilis* is that the switching-on of the competence state is not as sharp and

intense as for *S. pneumoniae*. Competence rises slowly and lasts longer, with less than 10% of the cells becoming simultaneously competent in the population (Claverys et al., 2006). Moreover *B. subtilis* has many flagella and prophages, which make the detection of a pilus structure by TEM more difficult. Characterization of the transformation pilus in other Gram-positive bacteria is required to know whether *S. pneumoniae*, an organism particularly optimized for natural transformation, could be an exception regarding its transformation pilus. We started a collaboration to look into this aspect on the human pathogen *S. aureus*.

Importantly, the *S. pneumoniae* transformation is more similar than previously thought to the one in Gram-negative DNA uptake systems. Following our publication, a published study showed the existence of a T4P on competent *V. cholerae*, which shares many features with the pneumococcal transformation pilus: its expression is competence-induced, it is a prerequisite for DNA uptake, and roughly a single copy is found per cell (Seitz et al., 2013). These striking similarities suggests that the DNA uptake process could be more conserved than expected in the whole bacterial kingdom. Interestingly, Imhaus et al. recently showed that reducing the number of T4P to roughly one copy per diplococcus in *N. meningitidis* leaves competence to its optimal level. In contrast, this reduction leads to an impaired ability for adhesion and twitching motility. The high number of T4P in this bacterium seems to be related to twitching motility and adhesion only, but not competence (Imhaus et al., 2014), implicating that these functions are completely independent. From these observations comes a very interesting question: could there be simultaneously two kinds of T4P pili in *N. meningitidis*, some for adherence and twitching, and one (or a few) for DNA uptake? Could the minor pilin ComP be the main element differentiating the two types of pili? Indeed ComP is not involved in T4P biogenesis, but absolutely required for DNA uptake (Wolfgang et al., 1999; Cehovin et al., 2013). Overexpression of ComP in *N. gonorrhoeae* increases the amount of DNA bound to the cell surface and it increases DNA uptake in general (Gangel et al., 2014). The hypothesis could be tested by immunogold labeling of ComP at the surface of *N. meningitidis*.

Besides, the presence of one or a few pili per cell for transformation raises questions considering that the ComGA-ComGB apparatus is known to co-localize with the translocation channel ComEC-ComFA in *S. pneumoniae* and *B. subtilis* (see introduction IV.D.). Does the pilus base corresponds to a unique active entry pore? How is the formation of only one or a few pili regulated? These questions need to be investigated, and might reveal important features of the DNA uptake mechanism. Bergé and colleagues recently showed that CY3-labeled DNA fragments form foci on the surface of competent e*ndA*- *S. pneumoniae* cells (Bergé et al., 2013). They showed that this DNA signal co-localizes with the site where the nuclease EndA is recruited by the DNA receptor ComEA. Interestingly, in the absence of the transformation pilus (*ΔcomGA* strain), the recruitment of DNA does not occur, consistently with the model of the pilus as a 'DNA trap'. To investigate the hypothesis that 'one pilus corresponds to one active entry pore', it would be interesting to observe by immunofluorescence whether the pilus co-localizes with the DNA signal in a *ΔendA* strain. Moreover, it is possible to go even further by observing this phenomenon at nanoscale resolution. By providing DNA labelled with a gold bead to competent cells deficient for DNA incorporation, the entry pore(s) on the cell surface can be precisely located at the site where the gold bead(s) is(are) observed. By cryo-electron tomography, a technique in which a 3D reconstruction of a cryofixed object is reconstituted from 2D tilted images, the surface of cells could really be scanned for entry pores. The pili should be observable by this technique, which will unambiguously answer the question of co-localization pilus/pore. Francesca Gubellini, a researcher in the lab, has initiated this project in collaboration with Gérard Péhau-Arnaudet from the Ultrastructural Microscopy platform at the Institut Pasteur.

A powerful way to look into the aspect of subunit regulation would be to use a targeted proteomics approach. Instead of identifying all proteins inside a complex sample, only a selected group is analyzed quantitatively (Picotti et al., 2011; Brun et al., 2009). Therefore, all

the proteins from the transformasome in *S. pneumoniae* could be precisely quantified by this technique. Knowing the copy number of each protein (even low copy number proteins) at different time points after competence induction might potentially unveil the mechanism of regulation of this machinery. How does each protein amount vary with the transformation efficiency? Are some of the proteins in limiting amounts? Does the quantity of some proteins vary from one pneumococcal strain to the other? Answers to these questions would provide rich information for all DNA uptake systems. However the technique is long and costly to develop: each targeted protein needs to go through the single reaction monitoring (SRM) development process, in order to be quantifiable by a quadrupole mass spectrometer. This perspective could be included in a long-term project.

DNA binding is another striking feature of the transformation pilus. We were able to observe that pili are ‘stuck’ to DNA by TEM, and that DNA from lysed cells is co-purified with pili. However the precise mechanism of this binding has not been resolved yet. The major pilin soluble domain does not bind DNA on its own, but could be involved in forming a binding pattern on the surface of the pilus. Alternatively, DNA binding on the pili could be indirect, involving an additional DNA binding partner. The only clearly characterized T4P DNA binding to date is the gonococcal T4P, through the ComP minor pilin (Cehovin et al., 2013). Consequently the *comG* minor pilins are attractive candidates to mediate this binding. Unfortunately they are essential for pilus biogenesis (Balaban et al., 2014), and a knockout mutant cannot be used to investigate this possibility. All minor pilins except ComGG were found to co-purify withe the pili by proteomics experiments, strongly suggesting that they are incorporated in the pilus. Immunogold labeling experiments on Flag-tagged minor pilins will allow us to confirm this point. Subsequently, obtaining their atomic structure could help to understand their precise function by designing point mutation that alter their surface. The discovery that *comG* minor pilins interact together, and that they are extremely unstable when expressed alone in *E. coli* and in *S. pneumoniae*, strongly suggests that they need to interact with each other to be stabilized.

Instead of co-expressing them in pairs as we tried, with little success, the expression of all minor pilins together preserving their operon structure in *E. coli* should be tried.

A platform to dissect the T4P biogenesis mechanism

Secondly, T4P have been extensively studied in Gram-negative bacteria notably for their potential as vaccine targets, but were described in only a few Gram-positive bacteria. *S. pneumoniae* transformation pilus revealed to be an extremely simple T4P of this widespread family in bacteria, in terms of essential components for pilus biogenesis. Moreover, the number of T4P-like operons present in Gram-positive bacteria (Imam et al., 2012) suggests that many other T4P remain to be discovered in these bacteria. Due to its simplicity, the *comG* operon represents the ideal platform for studying the mechanism of T4P assembly. Is the ComGA-ComGB complex a rotary motor? Then what is the stator? Is it able to perform pilus retraction by depolymerizing the pilins? Is the proton motive force involved in powering pilus assembly? I tried to obtain the atomic structure of ComGA, the assembly ATPase, and ComGC, the major pilin. Both proteins could be purified and biochemically studied, but their structure has not been obtained yet. Additionally, we showed through a proteomics approach that minor pilins are co-purified the pilus, and we constructed strains to study their localization.

To take advantage even further of the *comG* operon simplicity, I started the ambitious project of reconstituting the transformation pilus *in vitro*. This approach, if successful, will truly provide a platform for dissecting the T4P biogenesis mechanism. By easily inserting point mutations and screening for functional piliation, the precise role of each subunit could be investigated. Moreover, it would provide an easy source of functional pilus assembly complex for purification and structural study by cryo-electron microscopy. I am constructing in that perspective a synthetic operon containing all the known essential component for pilus biogenesis: the *comG* genes followed by *pilD*. This synthetic operon will be inserted in a vector for *in vitro* expression,

and liposomes will be added to the *in vitro* synthesis mix to provide a recipient for the pilus assembly machinery. *In vitro* reconstitution of pili will then be screened by TEM.

Another question concerning T4P biogenesis in Gram-positive bacteria could be investigated in the case of the *comG* platform: how can the pilus cross the peptidoglycane layer? And more related to competence, how can the incoming DNA cross this layer? It was evaluated that the peptidoglycane in *B. subtilis* is permeable to ~25 kDa proteins (Demchick et al., 1996). In this condition, it is hard to imagine how a long and straight structure such as the pilus could pass through this mesh without local restructuring. However, this cutoff size fluctuates a lot, even on a single cell, according to the reticulation degree, the osmotic pressure, the pH etc. (Forster et al., 2012). The peptidoglycane dynamics is still far from understood. Consequently, given the little information on the *S. pneumoniae* peptidoglycane, the need for local restructuring of the cell wall might simply not be required if the cell wall is relaxed enough during competence. Moreover several experiments indicate that DNA uptake takes place at midcell (Bergé et al., 2013), where the peptidoglycane is synthesized in the pneumococcus. This particular site might be more easily crossed. Even so, this topic remains important to investigate. In *S. pneumoniae* especially, several peptidoglycane hydrolases are involved in competence, through fratricide behavior (see introduction part III.C.). CbpD in particular is secreted by competent bacteria, but remains attached to the cell surface. The enzyme breaks the murein stem-peptide with its CHAP domain. Alone, it only performs a partial digestion of the peptidoglycane and does not lead to cell lysis (Eldholm et al., 2009). LytC, inactive on its own, becomes active after CbpD ‘predigestion’, and leads to cell lysis of non-competent cells. Competent cells are protected by the immunity protein ComM, but the mechanism behind this protection remains unknown. Does ComM protect the cell from the activity of CbpD or from the activity of LytC? This question is important, because the second case would mean that competent cells peptidoglycane does get remodeled, by their own CbpD enzymes. This partial digestion could help the pilus biogenesis and/or the entry of DNA. CbpD deletion in the laboratory strain R6 was shown to have no effect

on transformation efficiency (Johnsborg et al., 2008). However a different result was obtained in the clinical isolate G54, where CbpD deletion had a 50 fold reduced transformation efficiency (Rimini et al., 2000). Interestingly, a comparison between the R6 laboratory strain and its progenitor strain (the one used by Avery et al. in 1944; see introduction part I) showed that CbpD is among the genes that have mutated during laboratory selection (Lanie et al., 2007). Moreover, a similar importance in transformation efficiency was found for the CbpD homolog in *S. mutans* (Eaton et al., 2010). Of note, the R6 strain is uncapsulated, which might account for the fact that CbpD is dispensable for DNA uptake in this strain. Consequently, investigating the role of CbpD, apart from its involvement in the fratricide process, might shed light on some missing parts of the DNA uptake mechanism. A structural approach, coupled to functional assays and immunolocalisation experiments on a clinical strain could provide the information that might have been missed by studying the laboratory strain R6. The mechanism of action of the ComM immunity protein is also one of the major remaining puzzle. Again, a structure/ function approach seems to be a reasonable strategy to address such an intricate question, since the LytC protein structure is already solved (Perez-Dorado et al., 2010).

Is competence related to adhesion?

A conceptual difficulty to understand the maintenance of competence in *S. pneumoniae* is the absence of discrimination in the DNA taken up: foreign distantly-related DNA has more chances to be harmful than beneficial for the bacteria. Moreover, since protein misfolding is able to trigger competence (Stevens et al., 2012), DNA repair seems a particularly good explanation for the maintenance of competence in the pneumococcus, by providing a short-term advantage to this costly behavior. However, DNA repair necessitates the uptake of clonal DNA in priority. Fratricide, the killing of non-competent sister cells, is also a behavior difficult to explain. Since stressful conditions can also trigger competence (Prudhomme et al., 2006), how can the killing of the fittest bacteria (the ones that do not trigger competence) by the less fit bacteria (the ones

that are stressed) provide any advantage? Here again, the strategy can be explained provided that intra-species DNA is taken-up: if the fittest bacteria is lysed by a competent bacteria, and its DNA gets recombined, then the 'winners' are the genes of the fittest bacteria that got recombined. They will overwrite the genes of the competent bacteria, and continue to replicate.

Since no discrimination exists on the incoming DNA, biofilm formation would certainly be the simplest strategy to ensure that intra-species DNA is taken-up. Moreover, competence in *S. pneumoniae* has been linked to biofilm formation (Vidal et al., 2013). Interestingly, a recent study evaluated the expression levels of different genes that were shown to be involved in the pneumococcal colonization. Among them, the *lytC* gene showed the highest expression levels, followed by *lytA* and *comD* (Sakai et al., 2013), all three involved in the competence state (see introduction part III.C.). It is likely that a high LytC protein production results from this high transcription rate, since it is detected in high amounts in the supernatant of a *S. pneumoniae* culture (Eldholm et al., 2009). Importantly, LytC is highly produced constitutively and not only during competence, contrary to LytA and ComD. We saw that LytC is inactive in the absence of CbpD. Why would *S. pneumoniae* invest so much energy in producing and secreting an inactive protein? The involvement of LytC in biofilm formation seems to be the explanation for such an investment. LytC was shown to bind DNA, and the LytC/DNA interaction was shown to be a major 'glue' of the *S. pneumoniae* biofilm extracellular matrix (Domenech et al., 2013). Therefore, the primary function of LytC seems to be the structuring of the biofilm extracellular matrix. Like a ticking bomb, it accumulates rapidly in the supernatant of a pneumococcus culture. Then as soon as competence is triggered, most cells in the population become rapidly competent, CbpD is expressed, and non-competent cells are killed. Their released genomic DNA will be used by LytC to build the matrix trapping competent bacteria. The concerted action of the pneumococcal culture will therefore produce a rapid change from the planctonic state to the biofilm state, ensuring that most of the DNA ready to be recombined in the matrix is pneumococcal DNA rather than foreign DNA. Could the transformation pilus be involved in this

process? The discovery of a long DNA-binding appendage, induced specifically during competence naturally raises the question. Indeed, UV-inducible T4P in the hyperthermophile *Sulfolobales* archea mediate species-specific aggregation (discriminating even between different *Sulfolobales* species) and subsequent genetic exchange for chromosomal DNA repair by homologous recombination (Ajon et al., 2011).

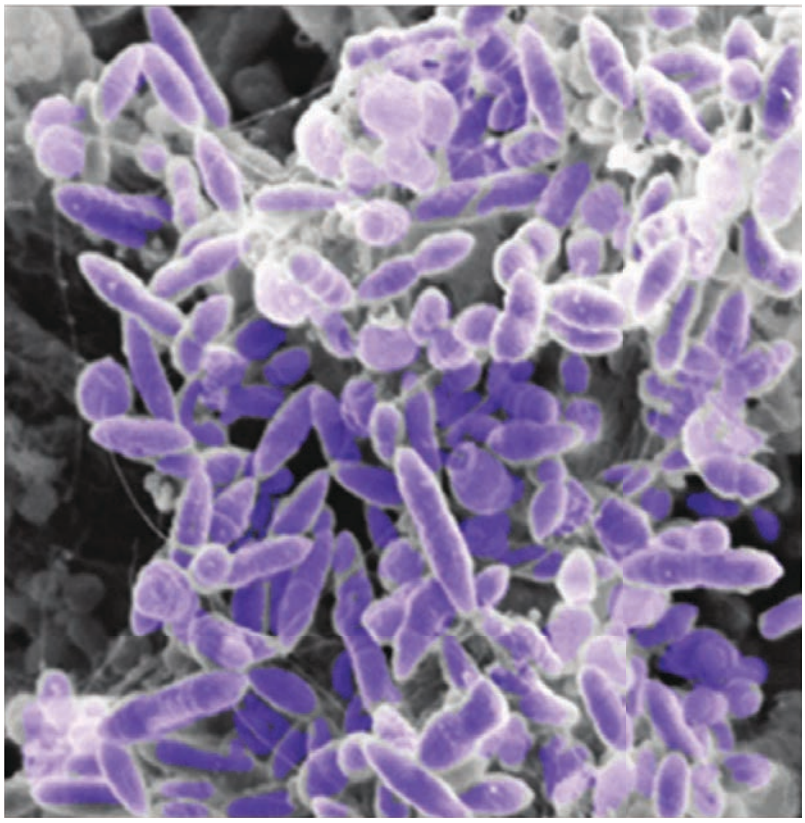


Figure 57 - Scanning electron micrograph of a pneumococcal biofilm.
Cover photograph of the Infection and Immunity journal, April 2013, volume 81, issue 4. Cells were colored in purple using Photoshop.

The regulatory pathway triggering competence in laboratory conditions is extremely well understood, however experimentation in more physiological conditions is now needed to address this kind of questions. Vidal and colleagues recently developed a smart reactor for the culture of human respiratory epithelial cells (HREC), with a continuous flow of nutrients, mimicking the respiratory epithelium microenvironment. By flowing *S. pneumoniae* cells in the setup the biofilm formation process can be monitored. Doing so, they showed that the

competence state is involved in early biofilm formation on HREC (Vidal et al., 2013). The reactor is consequently the perfect setup to test biofilm formation in quasi *in vivo* conditions. We have just initiated a collaboration with Vidal lab, to investigate in a first step whether a pilus deficient strain has any defect in biofilm formation.

Finally, a long-term perspective could arise from this PhD project: could the transformation pilus serve as a vaccine candidate? Current vaccines to treat pneumococcal disease are targeting the *S. pneumoniae* polysaccharide capsule (see introduction part III.A.). Although this strategy has shown to be effective, the gradually reduced efficiency of these vaccines makes it clear that better coverage will be needed in the future, by including protein antigens (non capsular) (Rodgers et al., 2011). Rodgers et al. define the criteria for finding such antigens:

- common to all serotypes,
- highly conserved within all serotypes,
- present on the cell surface,
- expressed in all stages of pathogenesis of the disease.

ComGC, the major pilin forming the transformation pilus, fulfills the first three requirements. Moreover T4P have successfully served as vaccine candidates for some Gram-negative bacteria (Giltner et al., 2012). Difficulties have hindered vaccine development against *Nesseria* T4P, due to the pilin high antigenic variability and post-translational modifications. ComGC does not undergo the antigenic variability described for *N. gonorrhoeae* PilE, and we showed by whole protein mass spectrometry that the protein has no post-translational modification other than a simple methylation (Laurenceau et al., 2013). Concerning the last requirement, in vivo study of the transformation pilus, as mentioned above, is necessary to conclude if the transformation pilus would be detectable in all stages of the disease. The transformation pilus is only expressed during the transitory competence state, for a short window of 15 to 30 min (see introduction III.C.) (Balaban et al., 2014). Moreover, if only a part of the cell population trigger competence, some cells will not be recognized by the immune system. Nevertheless the

competence state, central to *S. pneumoniae* metabolism (Johnston et al., 2014), is likely an obligatory step at every stage of the infection process. The potential involvement of these pili in biofilm formation could increase their attractiveness as a vaccine target. If such an application was to be achieved, it would nicely illustrate how fundamental and applied research are two intertwined and self-complementary disciplines.

Materials and methods

I. Molecular biology

A. Gene cloning in E. coli

Chemically competent *E. coli* BL21 Star (Life Technologies) were used for heterologous production of proteins, and chemically competent *E. coli* Top10 (Life Technologies) for the cloning procedure. The corresponding DNA sequence was amplified from genomic DNA of the pneumococcal strain R800, and subsequently inserted in a commercial plasmid using restriction enzymes sites. Plasmids were transformed and plaited on selective antibiotic media. Top10 clones were screened by PCR for the insert. Positive clones were then cultivated overnight in 3 mL of selective Luria Broth (LB) and plasmids were extracted using a commercial kit (Qiagen). Cloning was validated by sequencing (Eurofins MWG) using primers covering the whole gene sequence from upstream and downstream on the plasmid sequence. The plasmid was finally transformed in BL21 cells, and freshly transformed cells were used for protein production.

B. Point mutation insertion

A point mutation in the Walker A motif of ComGA was performed by inserting the mutation (K144A, codon AAG>GCG) inside a primer hybridizing around the mutation site, that was used to amplify the whole pBSSB20 plasmid by PCR. The PCR product was purified and digested by the DPN1 enzyme (Thermo Scientific) to eliminate all methylated DNA fragments (the template plasmid without the mutation). Both primers were phosphorylated, to allow recircularization of

the plasmid by a T4 DNA ligase (Thermo Scientific). The plasmid was transformed in Top10 *E. coli* cells, and positive clones were screened by sequencing the *comGA* gene.

C. Strains and plasmids

All pneumococcal strains and all *E. coli* plasmids used in this study are listed in the following tables. Most of the minor pilins clonings in *E. coli* were performed by Annick Dujeancourt, technician in the lab, and most of the minor pilins clonings in *S. pneumoniae* were performed by Sarah Ouarti, a student under my supervision.

Organism	Strain number	Genotype/relevant feature	Reference
S. pneumoniae	R800	*R6 derivative*	*Lefevre et al., Journal of bacteriology 1979*
	G54	*Clinical isolate of serotype 19F*	*Dopazo et al., Microbial drug resistance 2001*
	TCP1251	*Rx derivative but malM511, rpsL1, bgl1; SmR*	*Pestova et al., Molecular microbiology 1996*
	R1501	*R800 but ΔcomC*	*Dagkessamanskaia et al., Molecular microbiology 2004*
	R304	*R800 derivative, nov1, rif23, str41; NovR, RifR, SmR*	*Mortier-Barriere et al., Molecular microbiology 1998*
	R2548	*R1501 but CEPx-GFP, recA::ermAM; kanR*	*Claverys' strain collection*
	R1916	*R1501 but ssbB::luc (ssbB+), comGA::kan; CmR, KanR*	*Claverys' strain collection*
	R998	*R1501 but comEC::spc; CmR, SpcR*	*Claverys' strain collection*
	R1063	*R1501 but comFA::spc; CmR, SpcR*	*Claverys' strain collection*
	R1918	*R1501 but comGB::kan; KanR*	*Havarstein et al., Molecular microbiology 2006*
	R2546	*R1501, but CEPx-GFP(optimized); KanR*	*Claverys' strain collection*
	RL001	*R1501, but CEPx-comGC-FLAG (from plasmid pCEPx-comGC-FLAG); KanR*	*Laurenceau et al., PloS pathogens 2013*
	RL002	*RL001 , but comGA::spc (from strain R1062); KanR, SpcR*	*Laurenceau et al., PloS pathogens 2013*
	RL003	*R1501 but comGC E20A (point mutation of ComGC pilin)*	*Laurenceau et al., PloS pathogens 2013*
	RL005	*R1501 but comGG-FLAG (tag added at locus, Cterminal)*	*This study*
	RF001	*R1501, but CEPx-comGC-FLAG-mcherry; KanR*	*This study*
	SO001	*R1501, but CEPx-comGG-FLAG; KanR*	*This study*
	SO002	*R1501, but CEPx-comGD-FLAG; KanR*	*This study*
	SO005	*R1501, but CEPx-comGE-FLAG; KanR*	*This study*
	SO006	*R1501, but CEPx-comGF-FLAG; KanR*	*This study*
E. coli	BL21	*lon-, F-, ompT, hsdS(rB- mB-), gal, [malB +]K-12(λS)*	

Table 3 - Strains

Organism	Plasmid	Cloned genes/relevant features	Origin, resistance	Reference
S. pneumoniae, shuttle in *E. coli*	pCEPx	(pBR322) derivative containing the ComX-dependent promoter, PX, and the RBS of ssbB; KanR in S. pneumoniae, AmpR in E. coli	ColE1, KanR in S. pneumoniae, AmpR in E. coli	Martin et al., Molecular microbiology 2010.
E. coli	pASK-IBA3	C-terminal Strep-tag	ColE1, AmpR	commercial plasmid
	pASK-IBA12	Periplasmic expression, Thrombin-cleavable N-terminal Strep-tag	ColE1, AmpR	commercial plasmid
	pCDFDuet-1	2 cloning sites	CDF, AmpR	commercial plasmid
	pRSFDuet-1	2 cloning sites	RSF, KanR	commercial plasmid
	pET-20b	C-terminal 6xHis-tag	ColE1, AmpR	commercial plasmid
	pET-15b	N-terminal 6xHis-tag	ColE1, AmpR	This study
	pBSSB17	comGA in pCDF-Duet	CDF, AmpR	This study
	pBSSB18	comGA in pET15b	ColE1, AmpR	This study
	pBSSB20	comGA in pASK-IBA3plus	ColE1, AmpR	This study
	pBSSB21	comGA in pET20b	ColE1, AmpR	This study
	pBSSB59	comGA(G. thermodenitrificans) in pASK-IBA3plus	ColE1, AmpR	This study
	pBSSB60	comGA(G. thermodenitrificans) in pASK-IBA13	ColE1, AmpR	This study
	pBSSB61	comGA(B. subtilis) in pASK-IBA3plus	ColE1, AmpR	This study
	pBSSB62	comGA(B. subtilis) in pASK-IBA13	ColE1, AmpR	This study
	pBSSB36	comGC in pET-20b	ColE1, AmpR	This study
	pBSSB31	comGC(38-108) in pASK-IBA12	ColE1, AmpR	This study
	pBSSB32	comGC(16-108) in pASK-IBA3plus	ColE1, AmpR	This study
	pBSSB33	comGC(38-99)in pASK-IBA13	ColE1, AmpR	This study
	pBSSB34	comGC(38-108) in pASK-IBA13	ColE1, AmpR	This study
	pBSSB36	comGC(16-108) in pET-20b	ColE1, AmpR	This study
	pBSSB39	comGC(38-108) in pCDF	ColE1, AmpR	This study
	pBSSB63	comGD(51-160) in pASK-IBA12	ColE1, AmpR	This study
	pBSSB66	comGD(51-160) in pASK-IBA13	ColE1, AmpR	This study
	pBSSB40	comGD(31-160) in pASK-IBA3plus	ColE1, AmpR	This study
	pBSSB41	comGD(51-160) in pET-15b	ColE1, AmpR	This study
	pBSSB42	comGD(51-160, N-terminal His-tag) in pCDF MCS1	CDF, AmpR	This study
	pBSSB43	comGD(51-160) in pRSF	ColE1, AmpR	This study
	pBSSB64	comGF(55-160) in pASK-IBA12	ColE1, AmpR	This study
	pBSSB67	comGF(55-160) in pASK-IBA13	ColE1, AmpR	This study
	pBSSB45	comGF(32-160) in pASK-IBA3plus	ColE1, AmpR	This study
	pBSSB46	comGF(55-160) in pET-15b	ColE1, AmpR	This study
	pBSSB47	comGF(55-160, N-terminal His-tag) in pCDF MCS1	CDF, AmpR	This study
	pBSSB65	comGG(31-137) in pASK-IBA12	ColE1, AmpR	This study
	pBSSB68	comGG(31-137) in pASK-IBA13	ColE1, AmpR	This study
	pBSSB49	comGG(10-137) in pASK-IBA3plus	ColE1, AmpR	This study
	pBSSB50	comGG(31-137) in pET-15b	ColE1, AmpR	This study
	pBSSB51	comGG(31-137, N-terminal His-tag) in pCDF MCS1	CDF, AmpR	This study
	pBSSB52	comGG(31-137) in pRSF	RSF, KanR	This study
	pBSSB30	comEA(32-216) in pET-20b	ColE1, AmpR	This study

Table 4 - Plasmids

II. Protein biochemistry

A. Protein expression in E. coli BL21

i) General procedure

A single colony of freshly transformed *E. coli* BL21 cells was inoculated and cultivated overnight at 37°C, 200 RPM in 100 mL of TB supplemented with the selective antibiotic for a given plasmid. Selective antibiotics were maintained in the culture media all along the procedure at the following concentration :

- Ampicillin (Sigma-Aldrich): 1 µg/mL.
- Kanamycin (Sigma-Aldrich): 1 µg/mL.

4 to 8L of LB were inoculated by the overnight preculture at OD = 0.1, and cultivated at 37°C, 200 RPM. At OD = 0.8 protein expression was induced and cultures were switched to 20°C, 200 RPM overnight.

- For *tet*-promoters (IBA plasmids) cultures were induced by the addition of anhydrotetracycline (IBA) at 200 µg/L.
- For T7 promoters (pET, pRSF and pCDF plasmids) cultures were induced by the addition of IPTG (Sigma-Aldrich) at 1 µg/mL.

ii) Minimum medium expression and isotopic labeling

The minimum medium used for ComGC expression (pBSSB31) is the following :

- Na_2HPO_4 (6 g/L)
- KH_2PO_4 (3 g/L)
- NaCl (0.5 g/L)
- NH_4Cl (1 g/L, containing 15N for labeling)
- Glucose (4 g/L, containing 13C for labeling)

- $MgSO_4$ (0.2 mM)
- $CaCl_2$ (0.1 mM)
- Thiamine (10 mg/L)

The 100 mL preculture (see above) was performed in minimum medium, to avoid contamination by nutriments from the rich medium.

B. Protein purification

i) General procedure

Cells were harvested and resuspended in 50 mL of lysis buffer (Tris 50 mM pH=8; lisozyme (Anatrace); DNAse 1.25 µg/mL, protease inhibitor cocktail (Roche)) before being lysed by 3 passage through a cell disruptor (Emulsiflex) at 15,000 PSI, 4°C. Lysate was cleared by centrifuging 20,000 g for 45 min at 4°C, and passed on an affinity chromatography resin (buffer tris 50 mM pH=8, NaCl 200 mM, protease inhibitor cocktail (Roche)) following the instruction manual:

- HisTrap HP column (Amersham) for His-tagged proteins,
- a column packed with 10 mL of Strep-Tactin Superflow high capacity (IBA) for Strep-tagged proteins.

Eluted fractions were checked by SDS-PAGE. Fractions containing the protein of interest were pooled, concentrated on a concentrator (Millipore), 100 kDa cutoff for ComGA, 3 kDa cutoff for pilins. The concentrated protein was further purified by size exclusion chromatography on Superdex 200 (GE Healthcare) for ComGA, Superdex 75 (GE Healthcare) for pilins. Eluted fractions were checked by SDS-PAGE and concentrated again. For the first purification of each protein, the protein band was sent for mass spectrometry analysis to confirm its identity.

ii) Membrane protein purification

For membrane proteins, the cell lysate was cleared by centrifuging 20,000 g for 15 min at 4°C, and the membranes were pelleted by ultracentrifugation at 100,000 g for 45 min at 4°C. The membrane pellet was resuspended in a maximum of 50 mL of solubilization buffer (tris 50 mM pH = 8, NaCl 50 mM, protease inhibitor cocktail (Roche), DDM (Anatrace) 1%, LDAO (Anatrace) 10 mM) for protein solubilization, or in wash buffer (tris 50 mM pH = 8, NaCl 1 M, protease inhibitor cocktail (Roche)) for removing any protein partially attached with the membranes. For protein solubilization the sample was incubated for 20 min at room temperature on a rocker. An ultracentrifugation step at 100,000 g for 20 min at 4°C was performed to remove unsolubilized material. The clear solubilized sample was used for subsequent purification steps always keeping 5 mM LDAO (Anatrace) in the buffer to avoid precipitation of the membrane protein.

iii) Cleavage of the Strep-tag

For the pBSSB31 ComGC construct (thrombin-cleavable C-terminal Strep-tag), the tag was cleaved by the addition of 30 units of thrombin enzyme (Novagen) and 20 mM $CaCl_2$ to the pooled fractions after the Strep-Tactin column. The solution was incubated overnight at 4°C while dialyzing in a 3 kDa membrane to a bulk volume of buffer B (tris 50 mM pH=8, NaCl 200 mM, $CaCl_2$ 20 mM) in order to dilute the desthiobiotin (IBA). The cleaved ComGC was run again on the Strep-Tactin column, this time flowing through, while the remaining uncut ComGC-Strep was retained. The buffer B was exchanged to buffer C (Phosphate buffered saline 100 mM pH = 6.8, protease inhibitor cocktail (Roche)) on the Superdex 75 for NMR study. Protease inhibitors were not used prior the thrombin digestion, to avoid inhibiting the cleavage by this enzyme.

iv) Limited proteolysis

For limited proteolysis, the protease trypsin (Sigma-Aldrich) was mixed with ComGC^{soluble} to a molar ratio of trypsin:ComGC of 1:1500. The sample was incubated for 30 min on ice. Proteolysis was stopped by the addition of protease inhibitor cocktail (Roche). ComGC was then separated from the trypsin by purifying on a Superdex 75.

C. Protein crystallization

i) General procedure

Crystallization condition screening and crystal optimization were performed with the help of Ahmed Haouz and Patrick Weber from the Crystallography platform at Institut Pasteur.

Crystals were generated using the vapor diffusion by hanging drop technique. In brief, a drop of the concentrated protein was diluted 1:1 (volume:volume) in a crystallization solution containing precipitant(s). The mixed drop was placed in a hermetically closed well with a large volume of the precipitant solution, and incubated at 19°C. The protein and precipitant concentration will gradually increase as the water in the drop evaporates, until the protein reaches super saturation.

Commercial kits (Jena Bioscience, Hampton research) were used to screen for crystallization conditions. The protein was dispatched using the Mosquito robot. every hit was manually reproduced and optimized.

Crystals need to be frozen to avoid instantaneous degradation by the X-ray beam. They are frozen by plunging in liquid nitrogen (-196°C). A cryoprotectant is used to avoid the formation of water crystals that create parasite diffraction pattern. Glycerol, xylithol, polyethylene glycol

(PEG), were used for ComGA, and crystals cryoprotected in 15% glycerol were the ones giving the best diffraction.

Data were collected by Ahmed Haouz using the X-ray source from the Institut Pasteur crystallography platform, and the X-ray source from the Soleil synchrotron.

ii) ATP analogs

ATP analogs (Jena Bioscience) were added to the concentrated protein before manually setting the hanging drops, at a concentration of 2 mM, supplemented with 10 mM $MgCl_2$. $MgCl_2$ allows the ATPase hydrolysis cycle to take place, and allows the analog to take place in the ATP binding pocket.

List of the ATP analogs :

- ATPγS: Adenosine-5'-(γ-thio)-triphosphate, Lithium salt
- ATPαS: Adenosine-5'-(α-thio)-triphosphate, Sodium salt
- AppNHp: Adenosine-5'-[(β,γ)-imido]triphosphate, Sodium salt
- ApCpp: Adenosine-5'-[(α,β)-methyleno]triphosphate, Sodium salt
- AppCp: Adenosine-5'-[(β,γ)-methyleno]triphosphate, Sodium salt

iii) *In situ* proteolysis

In situ proteolysis of ComGC was performed using the Floppy-choppy kit (Jena Bioscience), following the instruction manual.

List of proteases used :

- Trypsin from bovine pancreas
- α-Chymotrypsin from bovine pancreas
- Subtilisin from *Bacillus licheniformis*
- Papain from papaya latex

- Carboxypeptidase Y (Thermo scientific)

D. Analytical ultracentrifugation

AUC analysis was performed by Bertrand Raynal from the biophysics platform headed by Patrick England at Institut Pasteur. Samples were used quickly directly after elution from the gel filtration column, to avoid degradation and aggregation. Sedimentation velocity experiments were done at 25°C using a ProteomeLab XL-I analytical ultracentrifuge (Beckman Coulter) equipped with an AN60-Ti rotor, spinning at 42,000 rpm. Measurements were performed at a concentration of 1.4 mg/mL in cells assembled with double sector epoxy centerpiece. The protein concentration was monitored by optical density measurements at a wavelength of 290 nm. Data were analyzed with the sedfit 12.0 software.

E. ATPase activity assay

The ATPase assay kit from Innova Bioscience was used to test ComGA activity, following the instruction manual. The presence of free phosphate in solution, resulting from ATP hydrolysis, is titrated through a colorimetric assay. readout was done on a plate reader (Tecan) at 600 nm. Buffer composition for measuring activity is: tris 50 mM pH = 7.4, NaCl 100 mM, $MgCL_2$ 2.5 mM, ATP 0.5 mM.

F. Proteomics analysis of purified pili

Proteomics analysis has been done in collaboration with Joseph Gault, Christian Malosse and Julia Chamot-Rooke from the Mass spectrometry unit at the Institut Pasteur.

The pili were purified as described in Laurenceau et al., 2013, with few differences:

- Cells were not sheared by vortexing 1 min, but merely resuspended in an iso-osmotic buffer (tris 100 mM pH = 7.5, $MgCl_2$ 1 mM, sucrose 1 M, DNAse 12.5 µg/mL) to minimize cell lysis.
- Resin was washed extensively using 5 incubations in wash buffer (tris 100 mM pH = 7.5, NaCl 1 M) for 5 min on a wheel at 4°C.

For mass spectrometry analysis, samples were resuspended in solvent A (98% H2O/ 2% ACN/ 0.1% CHOOH), and were separated prior to mass spectrometry using a nano-HPLC system (UltiMate 3000, Thermo Scientific) equipped with a pre-column (5 mm x 300 µm I.D., C18 PepMap100, 5 µm, 100 Å) and a homemade analytical column (15 cm, 75 µm I.D., 3 µm beads, 100 Å pore diameter) packed with C18 reverse phase (3 µm ReproSil-Pur Basic-C18-HD® -Maisch). After injection peptides were concentrated on the precolumn for several minutes then by means of a switching valve, eluted from the extraction column onto the analytical column in a standard back flush configuration. Peptides were separated, using a linear gradient of solvents A and B (20% H2O/ 80% ACN/ 0.08% CHOOH) increasing in hydrophobicity from 2% to 55% B at a flow rate of 300 nl/min. The eluent was directly electrosprayed (1.7 kV) to the coupled Q Exactive hybrid quadrupole-Orbitrap mass spectrometer (Thermo Scientific). Mass spectra were acquired in the positive-ion mode applying a data dependent switch between the survey scan and MS/MS acquisition. A single survey scan was acquired between m/z 300 to 2000 with an AGC target value of 1e6 at a resolution of 70000 at m/z 400. MS/MS was then performed on the 15 most intense ions above a threshold of 1e5 and with charge >1+. Ions were selected using an isolation window of 2Da and fragmented by high energy C-trap

dissociation (HCD) at a normalized collision energy of 27.5. Isotope clustering was enabled and a dynamic exclusion was set at 20 seconds. Data produced by mass spectrometry were re-processed with Mascot Distiller version 2.4.3 and searched using the Mascot search engine version 2.4.1 against a database of *S. pneumoniae*.

G. Enrichment of the plaited filaments

R1501 *S. pneumoniae* cells cultivated in 1 L LB at 37°C in a closed bottle without agitation, to avoid aeration. At O.D = 0.35 cells were pelleted and resuspended in 1 mL TBS (tris 100 mM pH = 7.5, NaCl 150 mM). Cells were vortexed for 1 min, and the suspension was then centrifuged twice at 10,000 g for 5 min to separate the bacteria from the supernatant containing plaited filaments. The supernatant was further filtered on 0.45 µm filter, and filaments were enriched by ultracentrifugation for 1 h at 50,000 g, 4°C, in a table-top ultracentrifuge (Beckman) and subsequent resuspension of the pellet in 50 µL TBS.

III. DNA binding experiments

A. EMSA

Random sequence 40mer oligonucleotide labelled with DY682 and the unlabeled reverse oligonucleotide were obtained from Eurofins MWG operon. For generating a labelled double stranded oligonucleotide, single stranded oligonucleotides were mixed together at 500 mM each in TE buffer (tris 10 mM pH = 8, EDTA 1 mM), and annealed using a thermocycler (Biorad) with the following program:

- 94°C for 3 min

- 94°C for 1 min, -0.7°C per cycle, 99 cycles
- 4°C forever

4 µL of protein sample were mixed with 4 µL of probe sample (either single stranded DY682 oligonucleotide of annealed double stranded DY682 oligonucleotide) at 2.5 nM, and incubated for 20 min at room temperature. Various molar ratio protein:oligonucleotide were tried. 1 µL of loading buffer was added (sucrose 20 %, bromophenol blue 0.12%), and the whole sample was loaded on an EMSA gel. EMSA gels were homemade: 8% acrylamid in TBE buffer (tris 89 mM, boric acid 89 mM, EDTA 2 mM). The gel was run for 1 h at 100 V, 4°C, in the dark, and revealed by a Licor Odyssey scanner.

B. Fluorescence anisotropy experiments

Fluorescence anisotropy experiments were performed with the help of Bruno Baron from the Biophysics platform headed by Patrick England at Institut Pasteur. Fluorescein-labeled and unlabeled oligonucleotides were obtained from Sigma-Aldrich. Annealing was performed as described above. The binding was performed in PBS containing labelled ssDNA or dsDNA at a concentration of 50 nM, increasing the protein concentration by 0.5 µmol increments. Fluorescence anisotropy was monitored at 25 °C using a PTI Quanta-Master QM4CW spectrofluorometer (PTI, Lawrenceville, NJ, USA) equipped with polarizers for the excitation and emission beams. All experiments were carried out in a 1 cm path-length cuvette at 25 °C. Band-pass of excitation (488 nm) and emission (525 nm) monochromators was set at 5 and 20 nm, respectively.

IV. Retraction experiments

A. Live immunofluorescence experiments

Live immunofluorescence experiments were performed by Jean-Yves Tivenez from the Imagopole platform headed by Spencer Shorte at Institut Pasteur.

The CY3-NHS (Amersham) and anti-FLAG A647 (Life technologies) labeling were performed using aliquots of pre-competent cells, so that experiments could be reproduced quickly, using the same batch of cells. The competence state being transitory (around 15 min), this experimental procedure was key to be able to screen efficiently many samples in the right time window.

Preparation and use of pre-competent cells:

S. pneumoniae R1501 cells were grown at 37°C under anaerobic condition, without agitation, in 50 mL of Casamino Acid Tryptone medium (CAT) up to OD600 = 0.15. Cells were pelleted by centrifuging 20 min at 4,500 g, 4°C, and resuspended in 5 mL of CAT supplemented with 15% glycerol. The suspension was aliquoted in 100 µL in 1.5 mL tubes, and aliquots were frozen at -80°C. For competence induction, an aliquot was thawed in ice, and 0.9 mL of CAT supplemented with $CaCL_2$ (1 mM) and adjusted to pH = 7.8, was added. Competence was triggered by incubating cells with CSP for 15 min at 37°C on a heat block.

Labeling:

For CY3-NHS labeling, 1 mL of competent cells were then centrifuged at 3,000 g for 2 min at room temperature, and resuspended in 1 mL of buffer Hepes 100 mM pH = 8, EDTA 50 µM, CY3-NHS 20 µg/mL for 2 min. Cells were pelleted again and the pellet was washed (without resuspending the pellet) with 1 mL of buffer without CY3-NHS to remove the excess. Cells were

resuspended in 100 µL of CAT medium, and the suspension was sealed without air between a glass slide and a coverslip. The preparation was observed immediately under the microscope for 15 min, before using another sample.

For the anti-FLAG A647 labeling, 1 mL of competent cells were then centrifuged at 3,000 g for 2 min at room temperature, and resuspended in 120 µL of CAT medium to concentrate the cells. 2 µL of anti-FLAG A647 were added, and the suspension was sealed without air between a glass slide and a coverslip. The preparation was observed immediately under the microscope for 15 min, before using another sample.

We used a LSM700 confocal microscope (Zeiss) with an objective PlanApochromat 64x, oil NA=1.4. TIRF microscopy was performed using a TIRF microscope Olympus PlanApo 100x, oil NA=1.45. For faster readout of pili dynamics, we used a spinning-disk UltraView VOX (Perkin-Elmer), mounted on a Zeiss platform, with a PlanApochromat 63x objective, oil NA=1.4, and a camera EMCCD Hamamatsu ImagEM C9100 back illuminated.

B. Optical tweezers

Optical tweezers experiments were performed in collaboration with Lena Dewenter and Berenike Maier from the biophysics group at the University of Cologne.

Coupling of the beads to monoclonal anti-Flag M2 antibodies (Sigma-Aldrich):

Antibody coupling to polystyrene beads with COOH groups beads (Kisker) were done following the instruction manual using a NHS esterification. For the coupling to protein G polystyrene beads (Gentaur), 50 µL of bead resin was incubated overnight at 4°C with anti-Flag antibodies (10 µg antibodies / mg of beads) in PBS. Beads were washed 3 times in PBS and were ready to be used.

Retraction was tested using aliquots of pre-competent cells, so that experiments could be reproduced quickly, using the same batch of cells. As mentioned above, this procedure was key to be able to screen efficiently many samples in the competence window.

Preparation and use of pre-competent cells:

S. pneumoniae R1501 cells were grown at 37°C under anaerobic condition, without agitation, in 50 mL of Casamino Acid Tryptone medium (CAT) up to OD600 = 0.15. Cells were pelleted by centrifuging 20 min at 4,500 g, 4°C, and resuspended in 5 mL of CAT supplemented with 15% glycerol. The suspension was aliquoted in 100 µL in 1.5 mL tubes, and aliquots were frozen at -80°C. For competence induction, an aliquot was thawed in ice, and 0.9 mL of CAT supplemented with $CaCL_2$ (1 mM) and adjusted to pH = 7.8, was added. Competence was triggered by incubating cells with CSP for 15 min at 37°C on a heat block. A 100 µL of this suspension supplemented with the polystyrene beads was sealed without air between a glass slide and a coverslip. The preparation was observed immediately in the optical tweezers setup, thermostated at 37°C.

A bead was trapped and approached to the bacterium that was fixed to a polystyrene coated glass slide while the entire sample was mounted on a piezo stage. The trapped bead was imaged directly onto a four-quadrant photo diode, allowing to monitor the displacements with high spatial and temporal resolution. The piezo stage was actuated by a computer-controlled force feedback loop to record single pilus dynamics at constant force.

Hereafter, the displacement by moving the piezo stage with the bacterium relative to centre of the optical tweezers allowed to screen for presence of a pilus bound to bead.

References

Ajon, M., S. Frols, M. van Wolferen, K. Stoecker, D. Teichmann, A. J. Driessen, D. W. Grogan, S. V. Albers and C. Schleper (2011). "UV-inducible DNA exchange in hyperthermophilic archaea mediated by type IV pili." Mol Microbiol **82**(4): 807-817.

Akrigg, A. and S. R. Ayad (1970). "Studies on the competence-inducing factor of Bacillus subtilis." Biochem J **117**(2): 397-403.

Albers, S. V., Z. Szabo and A. J. Driessen (2006). "Protein secretion in the Archaea: multiple paths towards a unique cell surface." Nat Rev Microbiol **4**(7): 537-547.

Avery, O. T., C. M. Macleod and M. McCarty (1944). "Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III." J Exp Med **79**(2): 137-158.

Balaban, M., P. Battig, S. Muschiol, S. M. Tirier, F. Wartha, S. Normark and B. Henriques-Normark (2014). "Secretion of a pneumococcal type II secretion system pilus correlates with DNA uptake during transformation." Proc Natl Acad Sci U S A **111**(7): E758-765.

Beiter, K., F. Wartha, B. Albiger, S. Normark, A. Zychlinsky and B. Henriques-Normark (2006). "An endonuclease allows Streptococcus pneumoniae to escape from neutrophil extracellular traps." Curr Biol **16**(4): 401-407.

Berge, M., I. Mortier-Barriere, B. Martin and J. P. Claverys (2003). "Transformation of Streptococcus pneumoniae relies on DprA- and RecA-dependent protection of incoming DNA single strands." Mol Microbiol **50**(2): 527-536.

Berge, M., M. Moscoso, M. Prudhomme, B. Martin and J. P. Claverys (2002). "Uptake of transforming DNA in Gram-positive bacteria: a view from Streptococcus pneumoniae." Mol Microbiol **45**(2): 411-421.

Berge, M. J., A. Kamgoue, B. Martin, P. Polard, N. Campo and J. P. Claverys (2013). "Midcell recruitment of the DNA uptake and virulence nuclease, EndA, for pneumococcal transformation." PLoS Pathog **9**(9): e1003596.

Biais, N., B. Ladoux, D. Higashi, M. So and M. Sheetz (2008). "Cooperative retraction of bundled type IV pili enables nanonewton force generation." PLoS Biol **6**(4): e87.

Bogaert, D., A. van Belkum, M. Sluijter, A. Luijendijk, R. de Groot, H. C. Rumke, H. A. Verbrugh and P. W. Hermans (2004). "Colonisation by Streptococcus pneumoniae and Staphylococcus aureus in healthy children." Lancet **363**(9424): 1871-1872.

Bradley, D. E. (1972). "Evidence for the retraction of Pseudomonas aeruginosa RNA phage pili." Biochem Biophys Res Commun **47**(1): 142-149.

Briley, K., Jr., P. Prepiak, M. J. Dias, J. Hahn and D. Dubnau (2011). "Maf acts downstream of ComGA to arrest cell division in competent cells of B. subtilis." Mol Microbiol **81**(1): 23-39.

Brissac, T., G. Mikaty, G. Dumenil, M. Coureuil and X. Nassif (2012). "The meningococcal minor pilin PilX is responsible for type IV pilus conformational changes associated with signaling to endothelial cells." Infect Immun **80**(9): 3297-3306.

Brun, V., C. Masselon, J. Garin and A. Dupuis (2009). "Isotope dilution strategies for absolute quantitative proteomics." J Proteomics **72**(5): 740-749.

Burrows, L. L. (2005). "Weapons of mass retraction." Mol Microbiol **57**(4): 878-888.

Campos, M., M. Nilges, D. A. Cisneros and O. Francetic (2010). "Detailed structural and assembly model of the type II secretion pilus from sparse data." Proc Natl Acad Sci U S A **107**(29): 13081-13086.

Cars, O., L. D. Hogberg, M. Murray, O. Nordberg, S. Sivaraman, C. S. Lundborg, A. D. So and G. Tomson (2008). "Meeting the challenge of antibiotic resistance." BMJ **337**: a1438.

Cascales, E. and P. J. Christie (2003). "The versatile bacterial type IV secretion systems." Nat Rev Microbiol **1**(2): 137-149.

Cavalli-Sforza, L. L. and M. W. Feldman (2003). "The application of molecular genetic approaches to the study of human evolution." Nat Genet **33 Suppl**(3s): 266-275.

Cehovin, A., P. J. Simpson, M. A. McDowell, D. R. Brown, R. Noschese, M. Pallett, J. Brady, G. S. Baldwin, S. M. Lea, S. J. Matthews and V. Pelicic (2013). "Specific DNA recognition mediated by a type IV pilin." Proc Natl Acad Sci U S A **110**(8): 3065-3070.

Chen, I. and D. Dubnau (2004). "DNA uptake during bacterial transformation." Nat Rev Microbiol **2**(3): 241-249.

Cisneros, D. A., P. J. Bond, A. P. Pugsley, M. Campos and O. Francetic (2012). "Minor pseudopilin self-assembly primes type II secretion pseudopilus elongation." EMBO J **31**(4): 1041-1053.

Claverys, J. P. and L. S. Havarstein (2007). "Cannibalism and fratricide: mechanisms and raisons d'etre." Nat Rev Microbiol **5**(3): 219-229.

Claverys, J. P., B. Martin and P. Polard (2009). "The genetic transformation machinery: composition, localization, and mechanism." FEMS Microbiol Rev **33**(3): 643-656.

Claverys, J. P., M. Prudhomme and B. Martin (2006). "Induction of competence regulons as a general response to stress in gram-positive bacteria." Annu Rev Microbiol **60**(1): 451-475.

Craig, L., N. Volkmann, A. S. Arvai, M. E. Pique, M. Yeager, E. H. Egelman and J. A. Tainer (2006). "Type IV pilus structure by cryo-electron microscopy and crystallography: implications for pilus assembly and functions." Mol Cell **23**(5): 651-662.

Crick, F. H., L. Barnett, S. Brenner and R. J. Watts-Tobin (1961). "General nature of the genetic code for proteins." Nature **192**: 1227-1232.

Croucher, N. J., S. R. Harris, L. Barquist, J. Parkhill and S. D. Bentley (2012). "A high-resolution view of genome-wide pneumococcal transformation." PLoS Pathog **8**(6): e1002745.

Demchick, P. and A. L. Koch (1996). "The permeability of the wall fabric of Escherichia coli and Bacillus subtilis." J Bacteriol **178**(3): 768-773.

Donati, C., N. L. Hiller, H. Tettelin, A. Muzzi, N. J. Croucher, S. V. Angiuoli, M. Oggioni, J. C. Dunning Hotopp, F. Z. Hu, D. R. Riley, A. Covacci, T. J. Mitchell, S. D. Bentley, M. Kilian, G. D. Ehrlich, R. Rappuoli, E. R. Moxon and V. Masignani (2010). "Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species." Genome Biol **11**(10): R107.

Dubey, G. P. and S. Ben-Yehuda (2011). "Intercellular nanotubes mediate bacterial communication." Cell **144**(4): 590-600.

Eaton, R. E. and N. A. Jacques (2010). "Deletion of competence-induced genes over-expressed in biofilms caused transformation deficiencies in Streptococcus mutans." Mol Oral Microbiol **25**(6): 406-417.

Eldholm, V., O. Johnsborg, K. Haugen, H. S. Ohnstad and L. S. Havarstein (2009). "Fratricide in Streptococcus pneumoniae: contributions and role of the cell wall hydrolases CbpD, LytA and LytC." Microbiology **155**(Pt 7): 2223-2234.

Forster, B. M. and H. Marquis (2012). "Protein transport across the cell wall of monoderm Gram-positive bacteria." Mol Microbiol **84**(3): 405-413.

Forterre, P. (2005). "The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells." Biochimie **87**(9-10): 793-803.

Friedrich, A., C. Prust, T. Hartsch, A. Henne and B. Averhoff (2002). "Molecular analyses of the natural transformation machinery and identification of pilus structures in the extremely thermophilic bacterium Thermus thermophilus strain HB27." Appl Environ Microbiol **68**(2): 745-755.

Gangel, H., C. Hepp, S. Muller, E. R. Oldewurtel, F. E. Aas, M. Koomey and B. Maier (2014). "Concerted spatio-temporal dynamics of imported DNA and ComE DNA uptake protein during gonococcal transformation." PLoS Pathog **10**(4): e1004043.

Gibson, D. G., J. I. Glass, C. Lartigue, V. N. Noskov, R.-Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z.-Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison, H. O. Smith and J. C. Venter (2010). "Creation of a bacterial cell controlled by a chemically synthesized genome." Science **329**(5987): 52-56.

Giltner, C. L., M. Habash and L. L. Burrows (2010). "Pseudomonas aeruginosa minor pilins are incorporated into type IV pili." J Mol Biol **398**(3): 444-461.

Giltner, C. L., Y. Nguyen and L. L. Burrows (2012). "Type IV pilin proteins: versatile molecular modules." Microbiol Mol Biol Rev **76**(4): 740-772.

Graupner, S., N. Weger, M. Sohni and W. Wackernagel (2001). "Requirement of novel competence genes pilT and pilU of Pseudomonas stutzeri for natural transformation and suppression of pilT deficiency by a hexahistidine tag on the type IV pilus protein PilAI." J Bacteriol **183**(16): 4694-4701.

Griffith, F. (1928). "The Significance of Pneumococcal Types." J Hyg **27**(2): 113-159.

Guiral, S., T. J. Mitchell, B. Martin and J. P. Claverys (2005). "Competence-programmed predation of noncompetent cells in the human pathogen Streptococcus pneumoniae: genetic requirements." Proc Natl Acad Sci U S A **102**(24): 8710-8715.

Hebisch, E., J. Knebel, J. Landsberg, E. Frey and M. Leisner (2013). "High variation of fluorescence protein maturation times in closely related Escherichia coli strains." PLoS One **8**(10): e75991.

Helaine, S., D. H. Dyer, X. Nassif, V. Pelicic and K. T. Forest (2007). "3D structure/function analysis of PilX reveals how minor pilins can modulate the virulence properties of type IV pili." Proc Natl Acad Sci U S A **104**(40): 15888-15893.

Hiller, N. L., A. Ahmed, E. Powell, D. P. Martin, R. Eutsey, J. Earl, B. Janto, R. J. Boissy, J. Hogg, K. Barbadora, R. Sampath, S. Lonergan, J. C. Post, F. Z. Hu and G. D. Ehrlich (2010). "Generation of genic diversity among Streptococcus pneumoniae strains via horizontal gene transfer during a chronic polyclonal pediatric infection." PLoS Pathog **6**(9): e1001108.

Hu, F. Z., R. Eutsey, A. Ahmed, N. Frazao, E. Powell, N. L. Hiller, T. Hillman, F. J. Buchinsky, R. Boissy, B. Janto, J. Kress-Bennett, M. Longwell, S. Ezzo, J. C. Post, M. Nesin, A. Tomasz and G. D. Ehrlich (2012). "In vivo capsular switch in Streptococcus pneumoniae--analysis by whole genome sequencing." PLoS One **7**(11): e47983.

Imhaus, A. F. and G. Dumenil (2014). "The number of Neisseria meningitidis type IV pili determines host cell interaction." EMBO J **33**(16): 1767-1783.

Jacob, F. and J. Monod (1961). "Genetic regulatory mechanisms in the synthesis of proteins." J Mol Biol **3**(3): 318-356.

Johnsborg, O., V. Eldholm, M. L. Bjornstad and L. S. Havarstein (2008). "A predatory mechanism dramatically increases the efficiency of lateral gene transfer in Streptococcus pneumoniae and related commensal species." Mol Microbiol **69**(1): 245-253.

Johnsborg, O., V. Eldholm and L. S. Havarstein (2007). "Natural genetic transformation: prevalence, mechanisms and function." Res Microbiol **158**(10): 767-778.

Johnston, C., N. Campo, M. J. Berge, P. Polard and J. P. Claverys (2014). "Streptococcus pneumoniae, le transformiste." Trends Microbiol **22**(3): 113-119.

Johnston, C., S. Caymaris, A. Zomer, H. J. Bootsma, M. Prudhomme, C. Granadel, P. W. Hermans, P. Polard, B. Martin and J. P. Claverys (2013). "Natural genetic transformation generates a population of merodiploids in Streptococcus pneumoniae." PLoS Genet **9**(9): e1003819.

Johnston, C., B. Martin, G. Fichant, P. Polard and J. P. Claverys (2014). "Bacterial transformation: distribution, shared mechanisms and divergent control." Nat Rev Microbiol **12**(3): 181-196.

Kaufenstein, M., M. van der Laan and P. L. Graumann (2011). "The Three-Layered DNA Uptake Machinery at the Cell Pole in Competent Bacillus subtilis Cells Is a Stable Complex." J Bacteriol **193**(7): 1633-1642.

Kilian, M., K. Poulsen, T. Blomqvist, L. S. Havarstein, M. Bek-Thomsen, H. Tettelin and U. B. Sorensen (2008). "Evolution of Streptococcus pneumoniae and its close commensal relatives." PLoS One **3**(7): e2683.

Kohler, R., K. Schafer, S. Muller, G. Vignon, K. Diederichs, A. Philippsen, P. Ringler, A. P. Pugsley, A. Engel and W. Welte (2004). "Structure and assembly of the pseudopilin PulG." Mol Microbiol **54**(3): 647-664.

Korotkov, K. V. and W. G. Hol (2008). "Structure of the GspK-GspI-GspJ complex from the enterotoxigenic Escherichia coli type 2 secretion system." Nat Struct Mol Biol **15**(5): 462-468.

Kramer, N., J. Hahn and D. Dubnau (2007). "Multiple interactions among the competence proteins of Bacillus subtilis." Mol Microbiol **65**(2): 454-464.

Lang, A. S., O. Zhaxybayeva and J. T. Beatty (2012). "Gene transfer agents: phage-like elements of genetic exchange." Nat Rev Microbiol **10**(7): 472-482.

Lang, E., K. Haugen, B. Fleckenstein, H. Homberset, S. A. Frye, O. H. Ambur and T. Tonjum (2009). "Identification of neisserial DNA binding components." Microbiology **155**(Pt 3): 852-862.

Lanie, J. A., W. L. Ng, K. M. Kazmierczak, T. M. Andrzejewski, T. M. Davidsen, K. J. Wayne, H. Tettelin, J. I. Glass and M. E. Winkler (2007). "Genome sequence of Avery's virulent serotype 2 strain D39 of Streptococcus pneumoniae and comparison with that of unencapsulated laboratory strain R6." J Bacteriol **189**(1): 38-51.

Lartigue, C., J. I. Glass, N. Alperovich, R. Pieper, P. P. Parmar, C. A. Hutchison, H. O. Smith and J. C. Venter (2007). "Genome transplantation in bacteria: Changing one species to another." Science **317**(5838): 632-638.

Laxminarayan, R., A. Duse, C. Wattal, A. K. Zaidi, H. F. Wertheim, N. Sumpradit, E. Vlieghe, G. L. Hara, I. M. Gould, H. Goossens, C. Greko, A. D. So, M. Bigdeli, G. Tomson, W. Woodhouse, E. Ombaka, A. Q. Peralta, F. N. Qamar, F. Mir, S. Kariuki, Z. A. Bhutta, A. Coates, R. Bergstrom, G. D. Wright, E. D. Brown and O. Cars (2013). "Antibiotic resistance-the need for global solutions." Lancet Infect Dis **13**(12): 1057-1098.

Li, J., E. H. Egelman and L. Craig (2012). "Structure of the Vibrio cholerae Type IVb Pilus and stability comparison with the Neisseria gonorrhoeae type IVa pilus." J Mol Biol **418**(1-2): 47-64.

Maier, B., I. Chen, D. Dubnau and M. P. Sheetz (2004). "DNA transport into Bacillus subtilis requires proton motive force to generate large molecular forces." Nat Struct Mol Biol **11**(7): 643-649.

Maier, B., L. Potter, M. So, C. D. Long, H. S. Seifert and M. P. Sheetz (2002). "Single pilus motor forces exceed 100 pN." Proc Natl Acad Sci U S A **99**(25): 16012-16017.

Malvankar, N. S. and D. R. Lovley (2012). "Microbial nanowires: a new paradigm for biological electron transfer and bioelectronics." ChemSusChem **5**(6): 1039-1046.

Mann, J. M., V. J. Carabetta, I. M. Cristea and D. Dubnau (2013). "Complex formation and processing of the minor transformation pilins of Bacillus subtilis." Mol Microbiol **90**(6): 1201-1215.

Martin, B., C. Granadel, N. Campo, V. Henard, M. Prudhomme and J. P. Claverys (2010). "Expression and maintenance of ComD-ComE, the two-component signal-transduction system that controls competence of Streptococcus pneumoniae." Mol Microbiol **75**(6): 1513-1528.

Martin, B., M. Prudhomme, G. Alloing, C. Granadel and J. P. Claverys (2000). "Cross-regulation of competence pheromone production and export in the early control of transformation in Streptococcus pneumoniae." Mol Microbiol **38**(4): 867-878.

Mashburn-Warren, L. M. and M. Whiteley (2006). "Special delivery: vesicle trafficking in prokaryotes." Mol Microbiol **61**(4): 839-846.

McLaughlin, L. S., R. J. F. Haft and K. T. Forest (2012). "Structural insights into the Type II secretion nanomachine." Curr Opin Struc Biol **22**(2): 208-216.

Mejean, V. and J. P. Claverys (1993). "DNA processing during entry in transformation of Streptococcus pneumoniae." J Biol Chem **268**(8): 5594-5599.

Mikaty, G., M. Soyer, E. Mairey, N. Henry, D. Dyer, K. T. Forest, P. Morand, S. Guadagnini, M. C. Prevost, X. Nassif and G. Dumenil (2009). "Extracellular bacterial pathogen induces host cell surface reorganization to resist shear stress." PLoS Pathog **5**(2): e1000314.

Mirouze, N., M. A. Berge, A. L. Soulet, I. Mortier-Barriere, Y. Quentin, G. Fichant, C. Granadel, M. F. Noirot-Gros, P. Noirot, P. Polard, B. Martin and J. P. Claverys (2013). "Direct involvement of DprA, the transformation-dedicated RecA loader, in the shut-off of pneumococcal competence." Proc Natl Acad Sci U S A **110**(11): E1035-1044.

Misic, A. M., K. A. Satyshur and K. T. Forest (2010). "P. aeruginosa PilT Structures with and without Nucleotide Reveal a Dynamic Type IV Pilus Retraction Motor." J Mol Biol **400**(5): 1011-1021.

Morikawa, K., A. J. Takemura, Y. Inose, M. Tsai, T. Nguyen Thi le, T. Ohta and T. Msadek (2012). "Expression of a cryptic secondary sigma factor gene unveils natural competence for DNA transformation in Staphylococcus aureus." PLoS Pathog **8**(11): e1003003.

Morlot, C., M. Noirclerc-Savoye, A. Zapun, O. Dideberg and T. Vernet (2004). "The D,D-carboxypeptidase PBP3 organizes the division process of Streptococcus pneumoniae." Mol Microbiol **51**(6): 1641-1648.

Morlot, C., L. Pernot, A. Le Gouellec, A. M. Di Guilmi, T. Vernet, O. Dideberg and A. Dessen (2005). "Crystal structure of a peptidoglycan synthesis regulatory factor (PBP3) from Streptococcus pneumoniae." J Biol Chem **280**(16): 15984-15991.

Ng, S. Y., B. Zolghadr, A. J. Driessen, S. V. Albers and K. F. Jarrell (2008). "Cell surface structures of archaea." J Bacteriol **190**(18): 6039-6047.

Nivaskumar, M., G. Bouvier, M. Campos, N. Nadeau, X. Yu, E. H. Egelman, M. Nilges and O. Francetic (2014). "Distinct docking and stabilization steps of the Pseudopilus conformational transition path suggest rotational assembly of type IV pilus-like fibers." Structure **22**(5): 685-696.

Nivaskumar, M. and O. Francetic (2014). "Type II secretion system: a magic beanstalk or a protein escalator." Biochim Biophys Acta **1843**(8): 1568-1577.

Okorokov, A. L., Y. L. Chaban, D. V. Bugreev, J. Hodgkinson, A. V. Mazin and E. V. Orlova (2010). "Structure of the hDmc1-ssDNA filament reveals the principles of its architecture." PLoS One **5**(1): e8586.

Perez-Dorado, I., A. Gonzalez, M. Morales, R. Sanles, W. Striker, W. Vollmer, S. Mobashery, J. L. Garcia, M. Martinez-Ripoll, P. Garcia and J. A. Hermoso (2010). "Insights into pneumococcal fratricide from the crystal structures of the modular killing factor LytC." Nat Struct Mol Biol **17**(5): 576-U572.

Picotti, P. and R. Aebersold (2012). "Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions." Nat Methods **9**(6): 555-566.

Pohlschroder, M., A. Ghosh, M. Tripepi and S. V. Albers (2011). "Archaeal type IV pilus-like structures-evolutionarily conserved prokaryotic surface organelles." Curr Opin Microbiol **14**(3): 357-363.

Polianskyte, Z., N. Peitsaro, A. Dapkunas, J. Liobikas, R. Soliymani, M. Lalowski, O. Speer, J. Seitsonen, S. Butcher, G. M. Cereghetti, M. D. Linder, M. Merckel, J. Thompson and O. Eriksson (2009). "LACTB is a filament-forming protein localized in mitochondria." Proc Natl Acad Sci U S A **106**(45): 18960-18965.

Popa, O. and T. Dagan (2011). "Trends and barriers to lateral gene transfer in prokaryotes." Curr Opin Microbiol **14**(5): 615-623.

Prudhomme, M., L. Attaiech, G. Sanchez, B. Martin and J. P. Claverys (2006). "Antibiotic stress induces genetic transformability in the human pathogen Streptococcus pneumoniae." Science **313**(5783): 89-92.

Rask-Andersen, M., M. S. Almen and H. B. Schioth (2011). "Trends in the exploitation of novel drug targets." Nat Rev Drug Discov **10**(8): 579-590.

Reguera, G., K. D. McCarthy, T. Mehta, J. S. Nicoll, M. T. Tuominen and D. R. Lovley (2005). "Extracellular electron transfer via microbial nanowires." Nature **435**(7045): 1098-1101.

Rimini, R., B. Jansson, G. Feger, T. C. Roberts, M. de Francesco, A. Gozzi, F. Faggioni, E. Domenici, D. M. Wallace, N. Frandsen and A. Polissi (2000). "Global analysis of transcription kinetics during competence development in Streptococcus pneumoniae using high density DNA arrays." Mol Microbiol **36**(6): 1279-1292.

Ripoll-Rozada, J., A. Pena, S. Rivas, F. Moro, F. de la Cruz, E. Cabezon and I. Arechaga (2012). "Regulation of the type IV secretion ATPase TrwD by magnesium: implications for catalytic mechanism of the secretion ATPase superfamily." J Biol Chem **287**(21): 17408-17414.

Rodgers, G. L. and K. P. Klugman (2011). "The future of pneumococcal disease prevention." Vaccine **29 Suppl 3**: C43-48.

Rudan, I., K. L. O'Brien, H. Nair, L. Liu, E. Theodoratou, S. Qazi, I. Luksic, C. L. Fischer Walker, R. E. Black, H. Campbell and G. Child Health Epidemiology Reference (2013). "Epidemiology and etiology of childhood pneumonia in 2010: estimates of incidence, severe morbidity, mortality, underlying risk factors and causative pathogens for 192 countries." J Glob Health **3**(1): 010401.

Sakai, F., S. J. Talekar, K. P. Klugman and J. E. a. Vidal (2013). "Expression of Streptococcus pneumoniae Virulence-Related Genes in the Nasopharynx of Healthy Children." PloS one **8**(6): e67147.

Salzer, R., F. Joos and B. Averhoff (2014). "Type IV pilus biogenesis, twitching motility, and DNA uptake in Thermus thermophilus: discrete roles of antagonistic ATPases PilF, PilT1, and PilT2." Appl Environ Microbiol **80**(2): 644-652.

Satyshur, K. A., G. A. Worzalla, L. S. Meyer, E. K. Heiniger, K. G. Aukema, A. M. Misic and K. T. Forest (2007). "Crystal structures of the pilus retraction motor PilT suggest large domain movements and subunit cooperation drive motility." Structure **15**(3): 363-376.

Seitz, P. and M. Blokesch (2013). "DNA-uptake machinery of naturally competent Vibrio cholerae." Proc Natl Acad Sci U S A **110**(44): 17987-17992.

Shaner, N. C., P. A. Steinbach and R. Y. Tsien (2005). "A guide to choosing fluorescent proteins." Nat Methods **2**(12): 905-909.

Smillie, C., M. P. Garcillan-Barcia, M. V. Francia, E. P. Rocha and F. de la Cruz (2010). "Mobility of plasmids." Microbiol Mol Biol Rev **74**(3): 434-452.

Stevens, K. E., D. Chang, E. E. Zwack and M. E. Sebert (2011). "Competence in Streptococcus pneumoniae Is Regulated by the Rate of Ribosomal Decoding Errors." Mbio **2**(5).

Stingl, K., S. Muller, G. Scheidgen-Kleyboldt, M. Clausen and B. Maier (2010). "Composite system mediates two-step DNA uptake into Helicobacter pylori." Proc Natl Acad Sci U S A **107**(3): 1184-1189.

van der Kooi-Pol, M. M., E. Reilman, M. J. Sibbald, Y. K. Veenstra-Kyuchukova, T. R. Kouwen, G. Buist and J. M. van Dijl (2012). "Requirement of signal peptidase ComC and thiol-disulfide oxidoreductase DsbA for optimal cell surface display of pseudopilin ComGC in Staphylococcus aureus." Appl Environ Microbiol **78**(19): 7124-7127.

van Wolferen, M., M. Ajon, A. J. M. Driessen and S.-V. Albers (2013). "How hyperthermophiles adapt to change their lives: DNA exchange in extreme conditions." Extremophiles **17**(4): 545--563.

VanLoock, M. S., X. Yu, S. Yang, A. L. Lai, C. Low, M. J. Campbell and E. H. Egelman (2003). "ATP-mediated conformational changes in the RecA filament." Structure **11**(2): 187-196.

Varga, J. J., V. Nguyen, D. K. O'Brien, K. Rodgers, R. A. Walker and S. B. Melville (2006). "Type IV pili-dependent gliding motility in the Gram-positive pathogen Clostridium perfringens and other Clostridia." Mol Microbiol **62**(3): 680-694.

Vignon, G., R. Kohler, E. Larquet, S. Giroux, M. C. Prevost, P. Roux and A. P. Pugsley (2003). "Type IV-like pili formed by the type II secreton: specificity, composition, bundling, polar localization, and surface presentation of peptides." J Bacteriol **185**(11): 3416-3428.

Walker, C. L., I. Rudan, L. Liu, H. Nair, E. Theodoratou, Z. A. Bhutta, K. L. O'Brien, H. Campbell and R. E. Black (2013). "Global burden of childhood pneumonia and diarrhoea." Lancet **381**(9875): 1405-1416.

Watson, J. D. and F. H. C. Crick (1953). "Molecular structure of nucleic acids." Nature.

Williams, R. C. and S. J. Spengler (1986). "Fibers of Reca Protein and Complexes of Reca Protein and Single-Stranded Phi-X174 DNA as Visualized by Negative-Stain Electron-Microscopy." J Mol Biol **187**(1): 109-118.

Wolfgang, M., P. Lauer, H. S. Park, L. Brossay, J. Hebert and M. Koomey (1998). "PilT mutations lead to simultaneous defects in competence for natural transformation and twitching motility in piliated Neisseria gonorrhoeae." Mol Microbiol **29**(1): 321-330.

Wolfgang, M., J. P. van Putten, S. F. Hayes and M. Koomey (1999). "The comP locus of Neisseria gonorrhoeae encodes a type IV prepilin that is dispensable for pilus biogenesis but essential for natural transformation." Mol Microbiol **31**(5): 1345-1357.

Yamagata, A., E. Milgotina, K. Scanlon, L. Craig, J. A. Tainer and M. S. Donnenberg (2012). "Structure of an essential type IV pilus biogenesis protein provides insights into pilus and type II secretion systems." J Mol Biol **419**(1-2): 110-124.

Yu, X. and E. H. Egelman (1990). "Image analysis reveals that Escherichia coli RecA protein consists of two domains." Biophys J **57**(3): 555-566.

Yu, X., M. S. VanLoock, S. Yang, J. T. Reese and E. H. Egelman (2004). "What is the structure of the RecA-DNA filament?" Curr Protein Pept Sc **5**(2): 73-79.

Étude du pilus de transformation de *Streptococcus pneumoniae*

Résumé

La transformation naturelle est la capacité de certaines bactéries à incorporer et à recombiner activement de l'ADN extra-cellulaire. Ce procédé majeur augmente la plasticité et l'adaptabilité des bactéries à Gram positif et négatif en réalisant des échanges génétiques intra- et inter-espèces. *S. pneumoniae* est un pathogène majeur de l'Homme. Cette bactérie est responsable d'infections sévères telles que des pneumonies, des méningites et des septicémies. Dans cette espèce, la transformation naturelle est corrélée au phénomène de changement de capsule et à la baisse d'efficacité des vaccins.

La plupart des bactéries à Gram positif naturellement transformables possèdent un opéron *comG*, semblable aux opérons codant pour la famille des pili de type IV, extrêmement répandus chez les bactéries à Gram négatif. Il a été proposé que l'opéron *comG* est responsable de la formation d'un petit filament, nommé pseudo-pilus. Cependant, un tel filament n'a jamais été observé. Par des techniques de mutagenèse, de caractérisation biochimique, de microscopie optique et électronique, nous sommes parvenus à identifier des filaments de plusieurs micromètres de long à la surface de bactéries *S. pneumoniae* compétentes. Nous avons confirmé l'appartenance de ces filaments à la famille des pili de type IV. Par conséquent, nous avons infirmé l'hypothèse de la formation d'un pseudo-pilus par l'opérons *comG* chez *S. pneumoniae.* De plus, nous avons montré que les pili se lient à l'ADN et qu'ils sont requis pour la capture de l'ADN extra-cellulaire. Ces résultats apportent des informations cruciales concernant les premières étapes de capture de l'ADN durant la transformation naturelle. Nous proposons un nouveau modèle dans lequel le pilus agirait comme un « piège à ADN », capturant l'ADN à la surface des bactéries compétentes pour le guider jusqu'au pore d'entrée dans la cellule.

Study of the natural transformation pilus of *Streptococcus pneumoniae*

Summary

Natural transformation is the ability of bacteria to actively take up and recombine extracellular DNA. This crucial process increases genome plasticity and adaptability of Gram-negative and Gram-positive bacteria through intra- and inter-species genetic exchange.
S. pneumoniae is a major human pathogen responsible for severe diseases such as pneumonia, meningitis and septicemia. In this species, transformation has been linked to capsular serotype switching and reduced vaccine efficiency.

Most transformable Gram-positive bacteria carry a *comG* operon that resembles operons encoding a widespread family of pili in Gram-negative bacteria, the type IV pili. It has been commonly proposed that the *comG* operon is responsible for the formation of a short pseudo-pilus filament. However, such an appendage had never been visualized in any bacterium. By mutagenesis, biochemical characterization, optical and electron microscopy techniques we were able to identify long, micrometer-sized appendages protruding from the surface of competent *S. pneumoniae*. We confirmed the Type IV pili nature of these appendages, we showed that they bind DNA, and are absolutely required for DNA uptake. We consequently overthrew the pseudopilus hypothesis at least in *S. pneumoniae*, and provided crucial information concerning the initial step of DNA uptake. We propose a revised model in which the transformation pilus acts as a “DNA trap” capturing DNA at the surface of competent cells, guiding it to the translocation channel.