UNIVERSITÀ DEGLI STUDI DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Sede Consorziata: Novartis Vaccines and Diagnostic srl

SCUOLA DI DOTTORATO DI RICERCA IN: Bioscienze

INDIRIZZO: Biologia Cellulare

CICLO: XXIII

TITOLO TESI

Discovering Sortases:
From pilus biogenesis in *Streptococcus pneumoniae*
to applications in biotechnology

Direttore della Scuola : Ch.mo Prof. Giuseppe Zanotti

Coordinatore d'indirizzo: Ch.mo Prof. Cesare Montecucco

Supervisor : Ch.mo Prof. Cesare Montecucco

Dottorando : Antonia Sinisi

TABLE OF CONTENTS

LIST O ABBREVIATIONS

BSA	Bovine serum albumin
CBD	Cholin binding domains
CBPs	Cholin binding proteins
CbpA	Cholin binding protein A
CC	Clonal Complex
CF	Cystis fibrosis
CLSM	Confocal laser scanning microscopy
Con A	Concavalin A
CCSP	Competence stimulating peptide
DIPEA	N, N’ diisopropylethylamine
DMF	N, N’ dimethylformamide
DSC	N,N ‘disuccinimidyl carbonate
ECM	Extracellular component matrix
EPS	Extracellular polymeric substance
FACS	Fluorescence-activated cell sorter
FMOC	9-fluorenylmethyloxycarbonyl
FRET	Fluorescent Resonance Energy Transfer
GFP	Green fluorescentic protein
HMW	High molecular weight
Hyl	Hyaluronate lyase
HOBT	1-Hydroxybenzotriazole
LTAs	Lipoteichoic acids
MFI	Mean fluorescence intensity
MLST	Multilocus sequence typing
NAG	N-acetyl glucosamine
NAM	N-acetyl muramic acid
NVT	Non vaccine type
OD	Optical density
OM	Otitis media
ST	Sequence type
PAF	Platelet activating factor
PBPs	Penicillin-binding proteins
PBS	Phosphate buffered saline
PFGE	Pulsed field gel electrophoresis
PyBOP	benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
PI-1	Pilus Islet 1
PI-2	Pilus Islet 2
Ply	Pnemuolysin
PsaA	Pneumococcal surface antigen A
PspA	Pneumococcal surface protein A
SBA	Soybean agglutinin

SDS	Sodium dodecyl sulphate
SipA	Signal peptidase A
SrtA	Sortase A
SrtC1	Sortase C1
SrtC2	Sortase C2
SrtC3	Sortase C3
SrtG1	Sortase G1
SrtG2	Sortase G2
SLV	Singl locus variants
Tas	Teichoic acids
TSB	Tryptone Soy Broth
THYE	Todd Hewitt Broth Yeast Extract
WGA	White germ agglutinin

SOMMARIO

I batteri Gram positivi espongono sulla loro parete cellulare proteine di superficie, molte di queste hanno una funzione relativa alla virulenza del batterio stesso, tra gli esempi del loro ruolo ci sono l'evasione dal sistema immunitario o l'iterazione con la superficie dell'ospite.

Diverse proteine di superficie sono legate covalentemente allo strato di peptidoglicano dei Gram positivi attraverso un meccanismo di transpeptidazione, enzimi responsabili di quenza reazione sono stati denominati sortasi. Le sortasi sono in grado di riconoscere un motivo C-terminal LPxTG, effettuando un taglio tra i residui aminoacidici di Thr e di Gly, in maniera tale da avere in un primo step di reazione la formazione di un intermedio che viene poi risolto in un secondo momento da un attacco nucleofilo mediato dall'N-terminal del peptydil side chain del Lipid II precursore del peptidoglicano (Ton-That et al., 1999). Oltre all'ancoraggio di proteine di superficie le sortasi sono responsabili della polimerizzazione di nuovi fattori di virulenza recentemente identificati in batteri Gram positivi e noti come Pili.

I Pili consistono in subunità covalentemente legate e assemblate grazie alle sortasi associate al pilo, le quali sono filogeneticamente e maccanicisticamente vicine alle housekeeping sortasi il cui ruolo è quello di ancorare le proteine alla parete cellulare (Telford et al., 2006). Diversi studi descrivono queste strutture come componenti elegibili per lo sviluppo di un vaccino, sebbene il meccanismo di assemblaggio risulta ancora rudimentale.

Per meglio chiarire il ruolo delle sortasi associate al pilo in questo lavoro di tesi si è preso come modello lo studio delle sortasi in *S. pneumoniae*.

In questo studio viene descritto infatti il ruolo delle sortasi nella biogenesi del pilo di *S. pneumonaie*, codificato da una regione nota come Pilus islet 1 (PI-1). Le sortasi relative a questa regione sono state caratterizzate usando diversi metodi biochimici, tra cui saggi di FRET (Fluorescent Resonance Energy Transfer) e saggi di attività basati su analisi HPLC.

Questi studi hanno permesso di fare analisi cinetiche e di studiare la specificità di substrato delle diverse sortasi relativa alle subunità che compongono il pilo, ossia RrgB, RrgA e RrgC.

Analisi *in vivo* hanno poi permesso di confermare i dati ottenuti dallo studio *in vitro*.

Dopo questa prima sessione, in questo lavoro si sono anche valutate le possibili applicazioni biotecnologiche capaci di sfruttare l'attività biologica delle sortasi.
Recentemente svariate applicazioni biotecnologiche hanno permesso di sfruttare la transpeptidazione per di funzionalizzare proteine, processo noto col nome di Sortagging.
Infatti le sortasi esercitano la loro attività anche *in vitro* in presenza di una proteina substrato ricombinante, contenente un motivo LPxTG e composti derivatizzati con oligoglicine che agendo da nucleofili risolvono la reazione di transpeptidazione in un prodotto finito modificato e funzionalizzato. Attraverso il Sortagging è possibile ottenere proteine fluorescenti, circolarizzarle o covalentemente unirle a supporto solido o banalmente è possibile ricreare un legame peptidico tra proteine con prodotti ricombinanti opportunamente modificati col motivo LPxTG e con oligoglicine (e.Gly_3). Per le suddette applicazioni è stata ampliamente utilizzata l'housekeeping sortase A di *S. aureus* che è anche stata impiegata in questa tesi.
Obiettivo è quello di creare un protein array per high troughput screening capace di rilevare iterazioni proteina-proteina. In questo contesto un protein array del *Virus Influenza* è stato sviluppato per lo screening di siero commerciale H1N1. Lisati batterici che esprimono proteine dell'*Influenza virus* modificate al C-terminal con il mofivo sortagging sono state immobilizate su glass-slide opportunamente derivatizzate con Gly_3.
Grazie alla specificità della sortasi A è possibile utilizzare estratti crudi evitandodi step di purificazione e un'immobilizzazione orientata.
É stato applicato per la prima volta il processo di Sortagging per legare proteine di un genoma intero su supporto solido. L'idea è quella di poter applpicare la tecnologia a un ampio proteoma in grado di sviluppare un protein microarray ad alta densità.

SUMMARY

Gram-positive bacteria display various surface proteins on their cell wall, many of them performing virulence related functions, such as host surface attachment and immune system escape. Several of these exposed proteins are covalently linked to the peptidoglycan layer through the sortase-mediated transpeptidation mechanism. Sortases recognize protein substrates with a C-terminal LPXTG motif, cleaving the peptide bond between the motif's threonine and glycine residues, forming a thioacyl intermediate. Transpeptidation proceeds by the nucleophilic attack by the N-terminus of the peptidyl side chain of a lipid II peptidoglycan precursor resulting in the formation of a protein-lipid II adduct (Ton-That and Schneewind, 1999). Incorporation of the protein-loaded lipid II into the peptidoglycan layer finally leads to covalent attachment of the sortase substrate to the cell wall.

Sortases are also responsible for the polymerization of new virulence factors recently identified in Gram-positive bacteria known as Pili. Pili consist of covalently linked subunits assembled by specialized pilus-related sortases, phylogenetically and mechanistically related to housekeeping sortases that anchor proteins to the cell wall (Telford et al., 2006).

Several studies describe pilus-like surface structures as eligible components for vaccine development; however, understanding the mechanism of pilus assembly in Gram positive bacteria is still rudimentary. To elucidate the role of pilus-related sortases, the human pathogen *Streptococcus pneumoniae* was used as a model system.

This thesis describes the role of pilus-related sortases in biogenesis of pili encoded by a region known as Pilus islet 1 (PI-1). PI-1 related sortases of *S. pneumoniae* were functionally characterized using various biochemical methods. FRET (Fluorescent Resonance Energy Transfer) and HPLC-based sortase activity assays were developed allowing kinetic and mechanistic analysis and studying substrate specificity of PI-1 related sortases. *In vivo* studies were performed to confirm *in vitro* data.

Furthermore, this thesis explores possible applications of sortases in biotechnology. Discovery of sortase-catalyzed transpeptidation lead to development of the Sortase tagging technology (Sortagging), a powerful tool to create functionalized protein-conjugates. Sortases may exert their activity *in vitro* in the presence of an LPXTG tagged protein and derivatized glycine compounds that act as nucleophiles to yield a functionalized transpeptidation product. Depending on the function carried by the glycine derivative, it is possible to obtain fluorescent labeled proteins, circularized proteins, covalently linked proteins or proteins that are covalently attached to a solid support. The housekeeping sortase A from *S. aureus* is extensively applied for these biotechnological purposes and it was also used for the novel biotechnological application described here. This thesis explores the application of sortase technology for the creation of protein arrays to be used in high throughput screening of protein-protein interactions. In this context, an *Influenza virus* protein array was developed to screen sera against the H1N1 virus. Bacterial crude extracts were applied to immobilize expressed LPETG-tagged *Influenza virus* proteins at their carboxyl-termini on pre-activated amine-glass slides derivatized with Gly_3 peptide. The specificity of the Sortase A reaction permits avoiding purification steps in array building and allows for immobilization of proteins in an oriented fashion. This section of the work describes how, for the first time, sortagging was implemented to covalently link proteins of a viral genome to a solid support. This approach allows for scaling-up to proteins of larger genomes in order to develop a high density protein array to be used in high throughput screening.

CHAPTER 1

Pilus assembly in *S. pneumoniae*: role of pilus islet 1 sortases

1. Introduction

1.1 Pathogenesis of pneumococcal disease: colonization and invasion

Streptococcus pneumoniae is a noted human pathogen causing various respiratory infections such as acute otitis media, sinusitis and pneumoniae (Fletcher and Fritzell, 2007;O'Brien et al., 2009;Ryan and Antonelli, 2000) but also part of the commensal flora of the upper respiratory tract. Together with other commensal microorganisms like *Moraxella cattarrhalis*, *Haemophilus influenzae*, *Neisseria meningitidis*, *Staphylococcus aureus*, and various hemolytic streptococci, it colonizes the nasopharyngeal niche. Though colonization with pneumococci is mostly symptomless, it can progress to respiratory or even systemic disease (Bogaert et al., 2004). Colonization is commonly followed by horizontal dissemination of the pathogens to individuals in the direct environment, leading to spread within the community (Givon-Lavi et al., 2002). The reported rates of bacterial acquisition and carriage depend on age, geographical area, genetic background, and socioeconomic conditions, so risk groups for diseases caused by pneumococci include young children, elderly people and patients with immunodeficiencies (1997).

S. pneumoniae colonization is a dynamic process; it competes with the resident flora which has an important role in the inhibition of pneumococcus and other virulent microorganisms.

The local host immune response depends on the condition of the upper respiratory tract; a poor mucosal immune response might lead to persistent and recurrent colonization and consequently infection, whereas a brisk local immune response to the pathogen will eliminate colonization and prevent re-colonization.

Nasopharyngeal colonization is the initial event from which mucosal and invasive pneumococcal infection develops and is also the nidus of pneumococcal transmission.

The polysaccharide capsule is the most important virulence factor of pneumococci and has been proven to be highly immunogenic. Antibodies targeting polysaccharides protect against infection with the homologous serotype by induction of opsonophagocytosis (Nelson et al., 2007b).

The layer underneath the capsule, the cell wall, consists of polysaccharides and teichoic acid and serves as an anchor for cell-wall-associated surface proteins. The cell wall is the cause of the intense inflammatory reaction that accompanies pneumococcal infection, since it stimulates the influx of inflammatory cells and activates the complement cascade and cytokine production.

The surface proteins contribute to the hydrophobic and electrostatic surface characteristics of pneumococci and might facilitate adherence to host cells partly through non-specific, physicochemical interactions.

Pneumococal strains with antigenically related capsule types are grouped together as a *serogroup* (for example 7A, 7B, 7C and 7F). There are 48 known serogroups and a total of 92 different serotypes (Henrichsen, 1995). Epidemiological studies, indicating the spread and occurrence of individual serotypes in certain geographic regions are abundant. More recent tools for epidemiological studies are PFGE (pulsed field gel electrophoresis), based on separation pattern of chromosomal DNA fragments, and MLST (multilocus sequence typing), involving the sequencing of seven housekeeping genes. Genomes of several pneumococcal strains have been sequenced and annotated since 2001 (TIGR4, R6, and 19F) (Tettelin et al., 2001). Compared to each other, individual strains show high diversity in their genetic composition. This diversity can be attributed to point mutations, promoted by endogenous hydrogen peroxide causing oxidative damage (Pericone et al., 2002) or the uptake of foreign DNA and incorporation into the genome (Berge et al., 2002).

1.2 Pneumococcal vaccine and Cell-wall anchored proteins

Although *S. pneumoniae* kills over 1.5 million children each year (http://www.who.int/vaccines/en/pneumococcus.shtml) and despite of being studied for over a century, pathogenesis and epidemiology are still not fully understood. Antibiotic therapy still remains efficacious against this pathogen *but*, due to anincrease of multi-drugs resistance, vaccination represents the most effective strategy for preventing infections. Although different serotype-based vaccines are already present on the market, their efficacy is limited.

The 23-valent polysaccharide based vaccine is not effective in children under 2 years of age, while the 7-valent and the 10-valent and 13-valent conjugate vaccines are effective on the target populations, but their strain coverage is serotype-dependent and can favor serotype replacement. In addition, the prohibitive cost of producing conjugate vaccines is a limit for their use in developing countries with the highest burden of disease and poor economical resources. For these reasons, vaccine research is currently focusing toward the use of one or more universally conserved protein antigens as potential vaccine components, able to elicit serotype-independent protection (Ryan and Antonelli, 2000). In this respect, cell-surface proteins have attracted considerable attention, because of their potential as vaccine antigens, able to stimulate the production of opsonic antibodies.

Gram-positive bacteria like *S. pneumoniae* display proteins on their surface permitting them to interact with host cells and tissues, thus conferring them an important role in virulence of bacteria (Navarre and Schneewind, 1999).

Bacteria use different mechanisms for displaying proteins at their surface. Examples are specific structural features such as choline residues decorating teichoic acid (TA) and lipoteichoic acid, a membrane anchoring portion, or lipoprotein and LPxTG motifs.

A major part of these cell wall anchored proteins (CWAPs) are attached to the bacterial surface through a covalent linkage with the peptidoglycan layer. Formation of this linkage is catalyzed by membrane-associated transpeptidases known as sortases, which recognize the C-terminal LPxTG motif of proteins destined to be attached to the cell wall (;Ton-That et al., 2004a).

S. pneumoniae is the only human-pathogenic bacterium known to express surface proteins specifically binding to choline as a mechanism of surface attachment. Interestingly, all three pneumococcal autolysins (LytA, LytB and LytC) and PspA belong to this category, and several Cbps, like CbpA, CbpD, CbpE and CbpG are bacterial factors implicated in colonization processes.

1.3 Pili

Surface-exposed components of a bacterium permit the microorganism to assess its environment; hence, these constituents often are major virulence factors. Pili and fimbriae are long filamentous structures that are presented by bacteria on their surface. They play a key role in host cell invasion, biofilm formation, cell aggregation, DNA transfer and twitching motility as it has been reported in different studies (Proft and Baker, 2009).

The role of pili as adhesive organelles is of particular importance for pathogenic bacteria; in fact they are involved in roles such as attachment to specific host cells during colonization, a key step in establishing an infection.

In Gram-negative bacteria, surface proteins are embedded into the outer membrane lipid layer and a common feature of Gram-negative pili is their role in adhesion to eukaryotic cells.

It has been proposed that bacteria use these structures to form an initial association with host cells, which can then be followed by a more ‘intimate’ attachment that brings the bacterium into close proximity to the host-cell surface. Pili of Gram negative bacteria are known to adhere to components of the extracellular matrix (ECM) and are variable in thickness (from 2 up to 7nm) and are composed by non-covalent homo-polymerized major subunits.

In contrast, Gram-positive bacteria use the cell wall as a scaffold for displaying a wide variety of surface molecules that may be either non-covalently attached to the cell wall or, as mentioned before, covalently linked to the peptidoglycan. Among the latter are pilus-like structures, long and flexible appendages with a diameter ranging from 3 to 10 nm and composed by covalently linked major and ancillary subunits and assembled by specialized sortase enzymes (Telford et al., 2006).

In Gram-positive bacteria, pilus-like structures have been identified by electron microscopy and characterized genetically. These structures were reported in many diverse species such as *Actinomyces naeslundii, Corynebacterium diphtheriae*, and many oral *Streptococcus spp.* (Ton-That et al., 2004b).

Gram-positive pili are usually encoded by islets inserted into specific genomic regions, varying in length and genetic organization, but sharing characteristic features; all pilus encoding islets contain genes coding for LPxTG or LPxTG-like surface-anchored proteins that are the pilus major and minor components. Other genes encode for sortase enzymes, specialized transpeptidases involved in pilus assembly.

The first insights into the assembly mechanism of Gram-positive pili were provided by Ton-That and Schneewind using *Corynebacterium diphtheriae* as a model organism (Ton-That and Schneewind, 2003a). Cell wall sorting signals (CWSS) are needed for an efficient sorting to the cell wall; the CWSS are N-terminus leader peptides. CWSS in Gram-positive bacteria are generally longer, more hydrophobic and more charged at their N-terminus than their counterparts in Gram-negative bacteria (Dramsi et al., 2008).

Various conserved genetic elements present within the major pilin subunit are required for a correct sortase-mediated pilus assembly: the pilin motif (WxxxVxVYPK), providing the ε-amino group of the lysine (K) residue, is required for subunit polymerization, furthermore the E-box domain (YxLxETxAPxGY) and a C-terminal cell wall sorting signal (LPxTG or a variant thereof), followed by an hydrophobic domain and a positively charged tail (Ton-That & Scheneewind, 2003; Ton-That & Scheneewind ,2004; Ton-That et al.2004) (Figure 1). Following this observation, additional experiments were performed in different pilus-expressing streptococci (S. *pneumoniae, S. pyogenes and S. agalactiae*).

Taken together, mutant studies, biochemical data and the recently solved crystal structure of the *Streptococcus pyogenes* pilus backbone (Kang et al., 2007) provide evidence for the presence of peptidyl linkages between adjacent pilus backbone subunits, and lend support to the "sortase-mediated pilus assembly" mechanism proposed by Ton That and Schneewind as a general model for the pilus assembly of Gram-positive bacteria (Ton-That & Schneewind 2003).

Pilus-like surface structures are reported to be virulence factors and various studies describe them as eligible components for vaccine development (Gianfaldoni et al., 2007;Maione et al., 2005;Margarit et al., 2009).

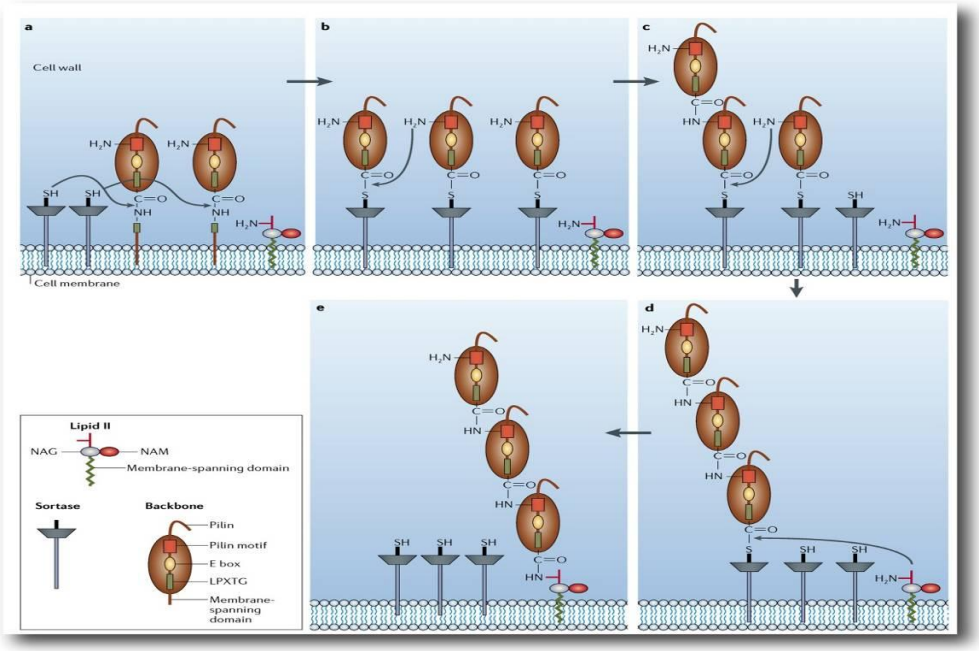


Figure 1. **General model of pilus assembly**. **a** In the first step, proteins with a C-terminal LPxTG motif (where x denotes any amino acid) are targeted to the cell membrane by Sec-dependent secretion (not shown). This is followed by a sortase-mediated reaction (indicated by the arrows) in which the LPXTG motif is cleaved between the threonine (T) and glycine (G) residues. **b** The reaction leads to the formation of an acyl-enzyme intermediate in which a covalent thioester bond is formed between the thiol group of a cysteine residue in the sortase and the carboxyl group of the pilin threonine residue. **c** Oligomerization occurs through the nucleophilic attack by the ε-amino group of the lysine residue in the pilin motif on the thioester bond of the sortase intermediate. **d** Finally, the thioester bond between the pilin subunit and the sortase is targeted by the amino group of the pentapeptide of lipid II, the precursor of peptidoglycan. **e** This leads to the formation of a membrane-associated, covalently linked, elongated pilus. NAG, *N*-acetyl glucosamine; NAM, N-acetyl muramic acid. **Image from J. Telford *et al.;* 2006.**

1.4 Pili in *S. pneumoniae*

Two genomic elements containing genes typical of Gram-positive pili islets have been identified thought analysis of *Streptococcus pneumoniae* genomes: pilus islet 1 (PI-1) and pilus islet 2 (PI-2) (Barocchi et al., 2006);(Bagnoli et al., 2008) (Figure 2).

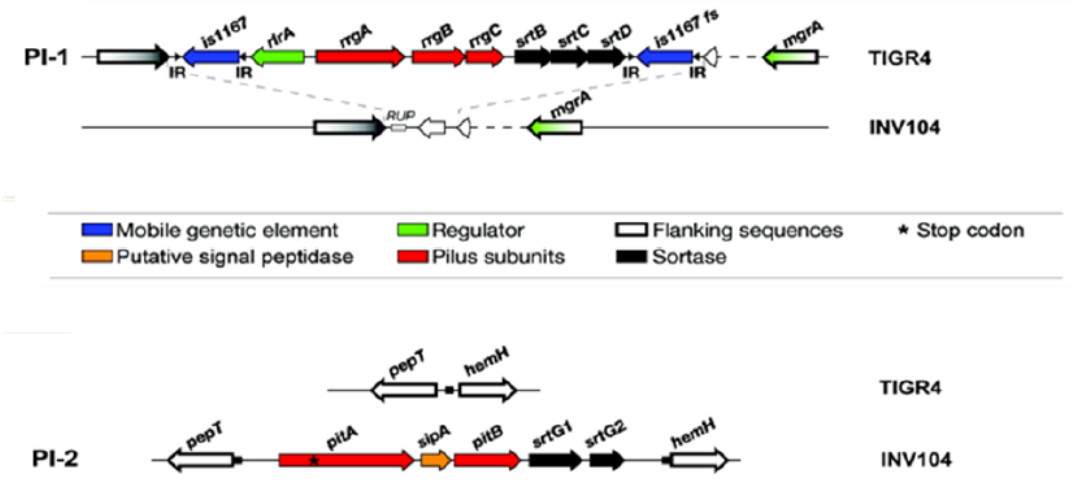


Figure 2. The genomic organization of pilus-encoding islets in *S. pneumoniae*.
Schematic representation of PI-1 andPI-2 (Bagnoli et al; 2008*).*

Both islets were verified to code for pili on the surface of pneumococcal bacteria and to be not widespread in the *S. pneumoniae* population. Interestingly, molecular epidemiology and genomic analyses demonstrated that unlike *S. pneumoniae,* all *S. pyogenes* and *S. agalactiae* isolates encode at least one functional pilus, suggesting different roles for pili in these major streptococcal pathogens (Falugi et al., 2008;Moschioni et al., 2008).

In detail, PI-1, first identified in the genome of the TIGR4 strain, is a chromosomal region of approximately 12 kb, flanked by two insertion sequences (IS1167) and comprising seven genes: *rlrA*, encoding a RolA-like transcriptional regulator, *rrgA*, *rrgB* and *rrgC*, which encode the LPxTG cell-wall anchored pilin subunits, and *srtC*-

1, *srtC-2* and *srtC-3* coding for the three sortases involved in the linkage and assembly of the pilus structure on the bacterial surface.

Recently, it has been demonstrated that the major pilin subunit, forming the backbone of the pilus, is RrgB, while the two ancillary pilins, probably localized at the opposite extremes of the pilus, are RrgA and RrgC, where RrgA is the pilus adhesin and the role of RrgC is still under investigation (Barocchi et al., 2006;Hilleringmann et al., 2008)(Figure 3).

Recent data show that the deletion of all three pilus-associated sortase genes, *srtC-1*, *srtC-2* and *srtC-3*, completely prevented pilus biogenesis, and expression of SrtC-1 alone is sufficient to covalently associate RrgB subunits to one another as well as linking the RrgA adhesin and RrgC to the polymer (Falker et al., 2008;Lemieux et al., 2008).

In particular, Falker's studies suggested that, SrtC-1and SrtC-2 act as redundant pilus subunit polymerases, with SrtC-1 processing all three pilus subunits proteins, while SrtC-2 only RrgA and RrgB. In contrast, SrtC-3 seems to have no pilus polymerase activity, but appears to be required for wild type focal presentation of the pili on the bacterial surface.

PI-2 is similar in its global genetic organization and in its sequence homology to the FCT-3 pilus islet of *S. pyogenes* (Telford et al., 2006). PI-2 islet contains five genes coding for two putative sortases (*srtG1* and *srtG2*), a signal peptidase related protein (*sipA*) essential for pilus assembly (but whose role has not yet been established), and two LPXTG-type surface anchored proteins (*pitB* and *pitA*), where PitB is the backbone subunit, and PitA is the ancillary protein. Interestingly, probably due to the different functions of the two pili (studies on the possible role of PI-2 in pathogenesis are currently ongoing); several copies of PI-1 are present on the surface of the bacteria, whereas only a single copy of PI-2 is detectable for each diplococcus.

Three independent studies analyzing global *S. pneumoniae* collections demonstrated that in both invasive and nasopharyngeal clinical isolates the frequency of the PI-1 islet was ~30% and that the presence of PI-1 was correlated with the genotype of the isolate, as defined by multi locus sequence typing (MLST), but not with the serotype (Moschioni et al., 2008) (Ogunniyi et al., 2007).

Moreover, the incidence of PI-1 was higher among antibiotic resistant clones (~50%) (Moschioni et al., 2008) thus suggesting that the pilus encoded by PI-1 may play a role in the global spread of resistant pneumococci (Sjostrom et al., 2007). A recent study demonstrated that PI-2 is present in ~16% of invasive and nasopharyngeal clinical isolates (Bagnoli et al. 2008), that the presence of PI-2 is also correlated with the genotype and that both islets, PI-1 and PI-1 are concomitantly present in the Taiwan19F-14 (ST236) clone, whose spread is responsible for the increasing incidence of antibiotic resistant isolates in many countries (Maiden, 1998). Extensive sequencing of PI-1 and PI-2 from different PI positive strains, allowed, based on sequence similarities, the identification of three major PI-1 clades while all the PI-2 isolates fell into a single clade. Interestingly, strains with the same genotype were also homogeneous for the PI-1 clade. More detailed studies on the basis of nucleotide homology revealed that the most variable gene within PI-1 is *rrgB*, followed by *rrgA*: for RrgB three major clades (I, II and III) were identified, displaying an overall homology ranging from 49% to 67%, whereas for RrgA two major clades (I and II), with an overall homology of 83% were identified. Interestingly, RrgA clade I isolates harbor RrgB clade I and III whereas RrgA clade II isolates RrgB clade II (Moschioni et al. 2008).

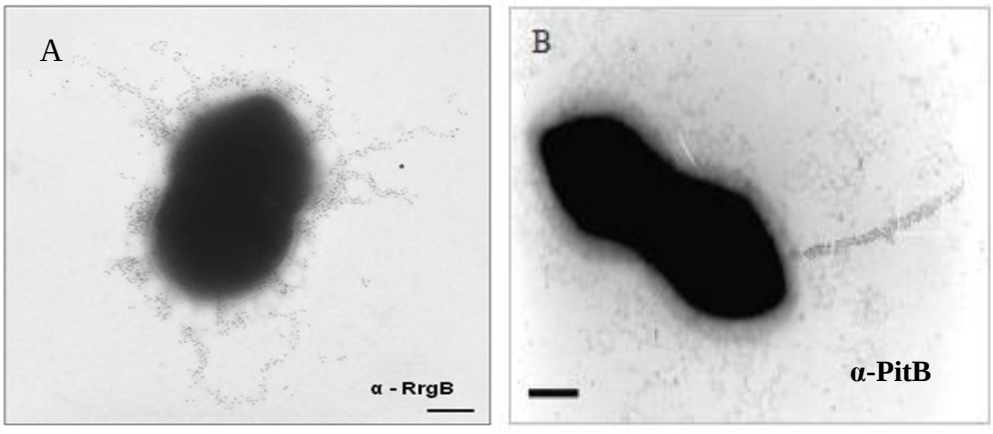


Figure 3: Detection of pneumococcal surface exposed pilus structures. A) Immunogold localization of RrgB in pilus-1 of *S. pneumoniae* TIGR4 using α-RrgB antibodies. **B)** Immunogold localization of PitB in pili of *S. pneumoniae* 19F Taiwan 14.

1.5 Sortases and Sortase-mediated pilus structure in *S. pneumoniae*

Sortases are membrane associated transpeptidases that recognize surface protein precursors with a CWSS containing a penta-peptide followed by a hydrophobic region containing 30-40 amino acids and a charged tail (Scott and Zahner, 2006). The hydrophobic domain and the charged tail anchor the protein to the cell membrane where the sortase is located (Scott and Zahner, 2006) (Figure 4).

Sortase homologues have been identified in all Gram-positive bacterial genomes and can be divided into four general classes, based on sequence alignment and predicted substrate preference (Clancy et al.; 2010). Designations are: sortase A, B, C and D, where the main sortases in class A and C recognize and cleave at the LPXTG domain (Dramsi et al., 2008). Sortase of class A, known as housekeeping sortases, are responsible for anchoring LPXTG-containing proteins to lipid II, many of them are involved in pathogenesis of bacteria. The gene-encoding *srtA* is not clustered with any of its substrates and is constitutively expressed. SrtA is a broad-range enzyme required for anchoring the majority or all the LPXTG-containing proteins.

The sortase B class has more pronounced substrate specificity; it selectively anchors the unique NPQTN-containing protein. The gene encoding *srtB* and its putative substrate are often in the same operon.

Class C forms the largest group of sortases. Class C sortases are present in several copies in the genome and clustered with their substrates and often associated with mobile genetic elements. Only this class of sortase possesses a hydrophobic C-terminal domain, they are associated with pilus or fimbria polymerization, catalyzing the cross linking between two pilin subunits (Dramsi et al; 2008). The role of class C sortases has been demonstrated first by Schneewind et al. in *Corynebacterium diphteriae* (Ton-That and Schneewind, 2003b).

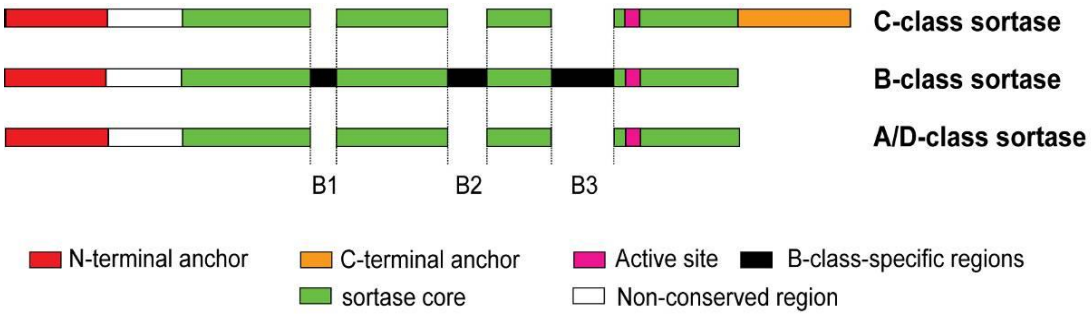


Figure 4: Four classes of sortases in Gram-positive bacteria. All sortases share an N-terminal anchor peptide and a conserved core region. The B-class sortases possess three additional B-class-specific regions. C-class sortases have a typical C-terminal hydrophobic anchor.

Pilus-related sortases are C-class related and recently the SrtC-1 and SrtC-3 structure of *S. pneumoniae* sortases have been resolved. SrtC-1 is composed of eight central, mostly anti-parallel β-strands that associate to form a β-barrel that is surrounded by three major α-helics (Manzano et al., 2008).

The crystal structures of both SrtC-1 and SrtC-3 display active sites whose access is controlled by flexible lids that are not present in non-pilus related sortases. The lid covers the active site region and structural studies observations suggest that it could play a role in substrate specificity providing a potential opening mechanism (Figure 5).

The proposed model suggested that the closed lid potentially prevents recognition of the LPXTG-like sequences, while only the binding with the specific substrate can result in to lid opening and enzymatic activity. In combination with the model proposed by Mandlik and coworkers, predicting that in order to have a proper pilus assembly mechanism, two different sortases must act sequentially (Mandlik et al., 2008). As shown by Manzano et al. (Manzano et al., 2009), the active sites of both crystallized sortases of pneumoniae are very similar, so the lid can only be opened by the specific pilin subunit.

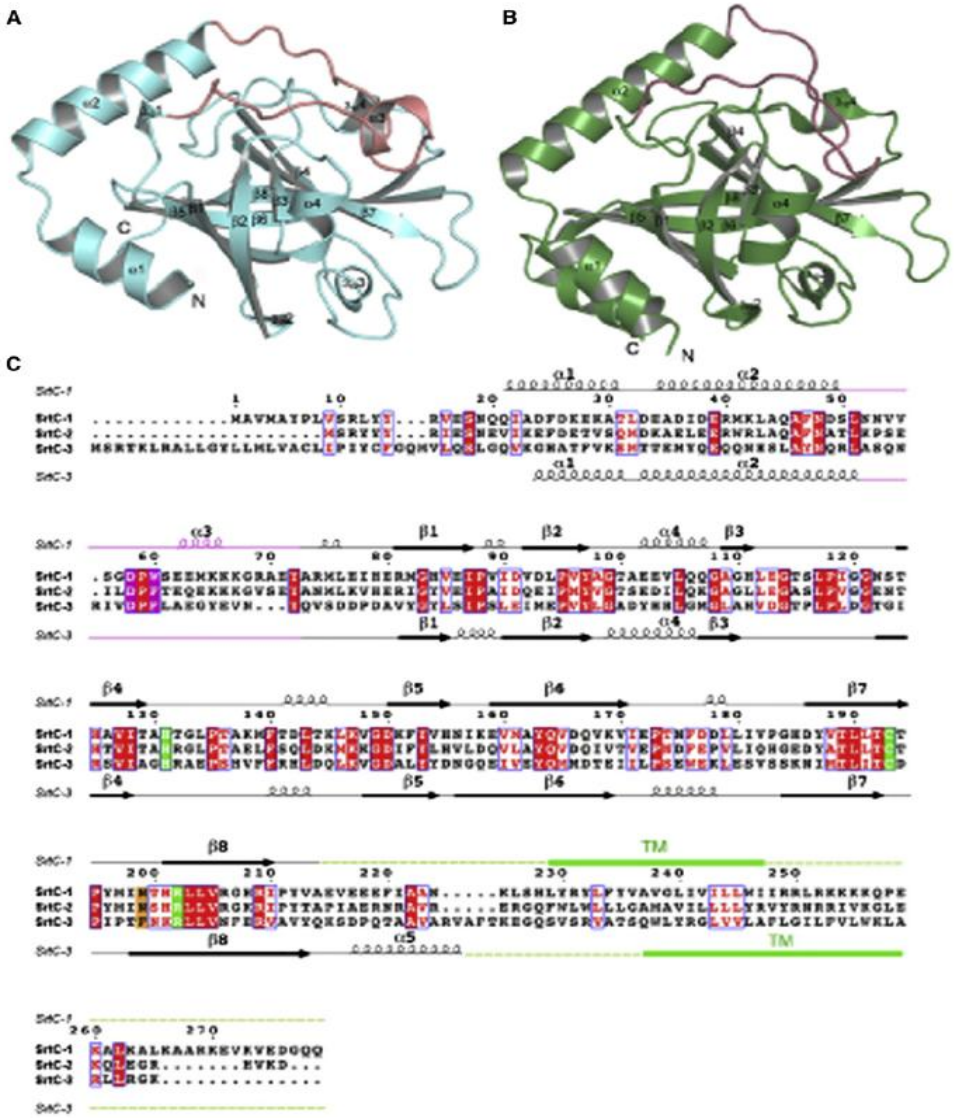


Figure 5: **Sortases of the pilus-formation machinery display similar folds** (A and B). SrtC-1 (A) and SrtC-3 (B) fold into β-barrels surrounded by helices; their active sites, centered on the region surrounding the C terminus of b7 and the N terminus of b8, are covered by a lid (shown in pink). (C) Structure-based sequence alignment of the three pilus-forming sortases of *S. pneumoniae*. Identical residues are shown with a red background, whereas similar residues are shown in red and highlighted with blue boxes. Residues located within the active site cleft of SrtC-1/SrtC-3 are highlighted in green, whereas the Asn residue on b7/b8 that contacts the nucleophilic Cys in SrtC-1 is shown with an orange background. The lid region is shown in magenta. From (Manzano et al., 2008).

2. Materials and methods

Cloning and expression of *S. pneumoniae* sortases. Genes encoding sortases *srtC1, srtC2*, and *srtC3* of *S. pneumoniae* TIGR4 were amplified from genomic DNA by PCR and inserted into the expression vector pET21b (Novagen). Primers used were shown in Table 1. PCR products were purified, digested and transformed into E. coli TOP10. Sortase overexpressing plasmids were transformed into BL21 codon plus (DE3) RIPL (Invitrogen) *E. coli* competent cells. The complete transformation mix was added to Luria broth (LB, 10 g/l bacto-tryptone, 5 g/l bacto-yeast extract, 0.5 g/l NaCl pH 7.4) and supplied with 100 μg/ml of ampicillin and 50 μg/ml of cloramphenicol and incubated overnight (o.n.) at 37°C with shaking. The o.n. cultures were diluted 1:100 in LB medium, shaken at 37°C and grown to an OD_{600} of 0.5-0.8, then expression was induced by addition of IPTG (isopropyl-β-D-thiogalactopyranoside, Sigma) to a final concentration of 1 mM. In the case of *srtC3*, the o.n culture was supplemented with 0.5 M of sorbitol (Sigma) and 1 mM of betaine to induce osmotic stress and increase product solubility. Induced cultures were grown at low temperature (20-25 °C) for 4-18 h. Cells were harvested by centrifugation for 10 min in a Beckmann JA 81000 rotor at 6.000 min^{-1} and 4 °C.

	Primer forw	Primer rev
SrtC-1	GGCGGTACATATGTATCCGCTGGTGTCTCGC	CCAATGGATCCATCGCAACGGCGAATCC
SrtC-2	CGCCCATATGATATATCCATTGGTGTCTCG	CAAACAGGATCCCAACATCAACAAGTATCCC
SrtC-3	TATTGTCATATGCAGATGGTGTTGCAGTCTC	TGACAAGGATCCGGAAAAAGGAACAGCCTTAG

Table 1: Nucleotide sequence of the primers used for *S.pneumoniae* sortases proteins cloning.

Purification of sortases. The cell pellets were resuspended in a lysis mix containing 0.25 mg/ml lysozyme, 50 mM Tris-HCl pH 7.5, 10% glycerol, 0.25% Brij-35, 2.5 mM DTT (1-4 dithiothreitol, Sigma) and 5 mM benzamidine-HCl. The lysate was incubated for 30 min on ice and centrifuged for 30 minutes at 35.000 min^{-1} in a 45 Ti rotor at 5°C.

The pellets were resuspended in buffer containing (500 mM NaCl, 50 mM Tris-HCl pH 7.5, 10% glycerol, 2.5 mM DTT and 5 mM benzamidine-HCl) and centrifuged for 15 minutes at 20.000 min^{-1} in a 45 Ti rotor at 5°C. Finally, pellets were resuspended in buffer containing (1% CHAPS, 50 mM Tris-HCl pH 7.5, 10% glycerol, 2.5 mM DTT, 5 mM benzamidine-HCl) and centrifuged for 15 minutes at 35.000 min^{-1} in a 45 Ti rotor at 5°C .

The supernatant was supplemented with ammonium sulphate (45% saturation) and stirred o.n. at 4°C. Precipitated proteins were collected by centrifugation for 15 minutes 35.000 min^{-1} in a 70 Ti rotor at 5°C. The pellets were resuspended in buffer A containing (50 mM Tris-HCl pH 8.0, 10% glycerol, 0.05% Brij-35 and 2.5 mM DTT) and dialyzed against the same buffer.

The dialyzed sample was applied to a 25 ml column (HR 10/25 column, Q-Sepharose FF, GE Healthcare) and eluted with a 10 CV gradient from 0 to 500 mM NaCl in buffer A. Fractions were concentrated by centrifugation in Centricon Plus-20 (Millipore) cartridges for 30 min at 4000 g at 4°C and dialyzed against 50% glycerol, 50 mM Tris-HCl pH 7.5, 0.05% Brij-35 and 5 mM DTT to be stored at -20°C.

LPXTG Peptides Cleavage. A Fluorescent Resonance Energy Transfer (FRET)-based assay was used to evaluate sortase activity and specificity. Sortases were incubated at 37°C with various substrates derived from the sorting sequence of each pilus component. As control a hyaluronidase-derived peptide known to be processed by the housekeeping sortase *SrtA* was used. The following peptides were obtained from Biomer Technology LLC: dabcyl-KRE**YPRTG**GI-edans (RrgA), dabcyl-KIT**IPQTG**GI-edans (RrgB), dabcyl-RID**VPDTG**EE-edans (RrgC) and dabcyl-KN**LPQTG**EG-edans (HylA). Reactions were carried out in RP buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2.5 mM DTT, 5 mM $CaCl_2$) with 8 µM of enzyme and varying concentration of peptide (1 to 25 µM), and in the presence or absence of H_2N-OH , NH_2 –Gly_3 and Lys in a total volume of 300 µl.

The increase of the fluorescence was measured at 37°C at λ_{em} 490 and λ_{abs} 360 for 70 min using a LS 55 Fluorescence Spectrometer (PerkinElmer).

Transpeptidation of synthetic peptides by SrtC1. Transpeptidation assays were performed in 30 µl reaction mixture containing RP buffer, 8 µM of SrtC1, 300 mM Lys, 500 µM dabcyl-KIT**IPQTG**GI, 250 µM AHVYP**K**NT-edans at 37°C for 2 h. Reactions were blocked with 18 mM EDTA in a total volume of 300 µl. Fluorescence was measured at 37°C at λ_{em} 490 and λ_{abs} 360 over a period of 70 min using a LS 55 Fluorescence Spectrometer.

SDS-PAGE and Western Blot analysis. SDS-PAGE analysis was performed using Nu-PAGE™ 4-12 % Bis-Tris gradient gel (Invitrogen) according to the manufacturer's instructions. Hi-Mark™ pre-stained HMW protein standard (Invitrogen) served as protein standard. Gels were stained with Colloidal Coomassie Blue G-250 (Invitrogen) or processed for Western Blot analysis by using standard protocols. Mouse antibodies raised against recombinant His-Tag-proteins were used at a dilution of 1/3000. Secondary goat anti-mouse IgG alkaline phosphatase conjugated antibodies were used at a dilution of 1/5000 and signals were developed by using Western Blue Stabilized Substrate for Alkaline Phosphatase (Promega).

HPLC analysis. LPXTG peptide cleavage by sortases was carried out in a reaction mixtures containing RP buffer in presence and absence of 200 mM H_2N-OH or 20 mM Lys·HCl in a total volume of 300 µl. Reactions were initiated by the addition of 8 µM of sortase and incubated at 37°C for 4 h. Controls contained either peptide only, sortase only or the complete mix at T_0. 100 µl of each mix were separated on a µRPC C2/C18 ST 4.6/100 column (Pharmacia Biotech) using a linear gradient from 20 to 80% CH_3CN/0,1% TFA over 30 minutes (1 ml/min). To confirm peptide cleavage and the identity of the products, peak fractions were collected, concentrated in a Centrivap vacuum concentrator system (Labconco) and analyzed by MALDI-TOF mass spectroscopy.

Affinity purification antibodies. Affinity-purified antibodies were obtained by blotting several μg of purified protein target protein on nitrocellulose membranes. Poorly bound protein was removed by immersing the membrane in glycin-HCl 100 mM pH 2.5 for 5 minutes and washing twice for 2 minutes in TBS. The part of the membrane containing the protein band was cut out and blocked with TBS/3% BSA for 1 hour, and then washed again in TBS. The serum to be purified was diluted 1:4 in TBS and incubated over night at 4°C with the membrane. The supernatant was removed and the membrane was washed in TBS and PBS. Elution was performed by adding glycin-HCl, 100 mM, pH 2.5 incubating for 10 min, and neutralization with Tris-HCl pH 8.

Immunofluorescent imaging of bacteria. Mid-log phase (OD_{620} = 0.25) pneumococci were grown in THYE medium, washed, struck on glass slides and allowed to air-dry. Paraformaldehyde was added to bacterial pellet to final concentrations of 2% in phosphate buffer pH 7.4. The cells were incubated for 15 min on ice, washed in PBS, and attached to poly-lysine-coated cover slips. The slides were washed in PBS, dipped in ice-cold methanol for 10–60 seconds, and dipped in PBS. The cells were then treated with 1 U/ml Cpl1 in PBS for 10 min at room temperature, washed in PBS, and blocked for 15 min with PBS supplemented with normal goat serum and 1% BSA. Antibodies and dyes were diluted in PBS containing 1% BSA and incubated with the cells in a moist chamber for 1 hour at room temperature. Between each incubation step the cells were washed thoroughly with 1% BSA in PBS. The slides were washed in water and then mounted in ProLong® Gold Antifade Reagent (Molecular Probes/Invitrogen). Images were obtained using a Carl Zeiss LSM 7MP Laser Scanning Microscope. Images were deconvolved using Axiovison software.

Strains and bacterial growth. Unless otherwise noted, pneumococcal strains were stored at -80°C in 12% glycerol and routinely grown at 37°C in 5% CO2 on Tryptic Soy Agar (Becton Dickinson) supplemented with 5% defibrinated sheep blood or in Tryptic Soy Broth (Becton Dickinson).

Flow Cytometry. Bacteria were grown in THYE to exponential phase (OD_{600}= 0.2), fixed with 2% paraformaldehyde, and then stained with mouse antisera raised against RrgB, SrtC-1, or SrtC-2 recombinant proteins (dilution 1:100). After labeling with a FITC-conjugated secondary antibody (Jackson Laboratories), bacterial staining was analyzed and sorted by using a FACSAria (Becton Dickinson).

3. Results

3.1 Efficient expression of *srtC-1*, *srtC-2* and *srtC-3* in *E. coli* requires overproduction of certain tRNAs.

In order to study how pilus-related sortases are related to pilus assembly in *Streptococcus pneumoniae*, *in vitro* assays were performed with purified enzymes. Genes encoding sortases *srtC-1*, *srtC-2*, and *srtC-3* of *S. pneumoniae* TIGR4 were amplified from genomic DNA by PCR and inserted into the expression vector pET21b as described in *Materials and Methods*. Initiation codons were placed right after the end of predicted membrane anchor sequences common to all sortases. Since the natural stop codon was conserved during PCR amplification, the resulting polypeptides do not bear any tag and correspond to the native polypeptides missing only the N-terminal membrane anchor. Plasmid constructs were transformed into BL21 (DE3) and gene expression was induced in logarithmically growing cells by the addition of IPTG. Analysis of crude cell extracts of induced cells on Coomassie blue-stained SDS PAA-gels revealed only very faint bands of over-expressed sortases .Upon closer inspection of the sortase-coding nucleotide sequences the presence of a number of codons was noted, that are rarely used by *E. coli* in highly expressed genes.

In order to obtain an expression level of sortases, the plasmid constructs were transformed into BL21 (DE3) RIPL, an *E. coli* strain carrying a plasmid specifying the usually poorly expressed tRNA genes. Crude extracts of the IPTG-induced cells contained considerable amounts of the corresponding sortases, demonstrating that expression of *srtC-1*, *srtC-2* and *srtC-3* in *E. coli* usually is limited by tRNA supply. SrtC-1 and SrtC-2 were soluble in Bugbuster reagent while SrtC-3 was found mostly in inclusion bodies. A soluble form of SrtC-3 was obtained by growing the cells under osmotic stress conditions in the presence of sorbitol and betain. Enough material was ready to be processed and used for experimental procedures described above.

3.2 Pilus-related sortases were obtained in active form.

Sortases were purified in active form by a four-step procedure. SrtC-1, C-2 and C-3 remained in the insoluble pellet after Brij-lysozyme extraction of induced cells. The bulk of soluble proteins was removed by a washing the insoluble material in high-salt buffer, followed by solubilization of the sortases in 1 % CHAPS buffer. Extracted proteins were precipitated with ammonium sulphate, dialyzed and subjected to anion exchange chromatography on Sepharose Q. The latter step yielded a nearly homogeneous preparation of soluble sortase suitable for the experiments described below (Figure 6).

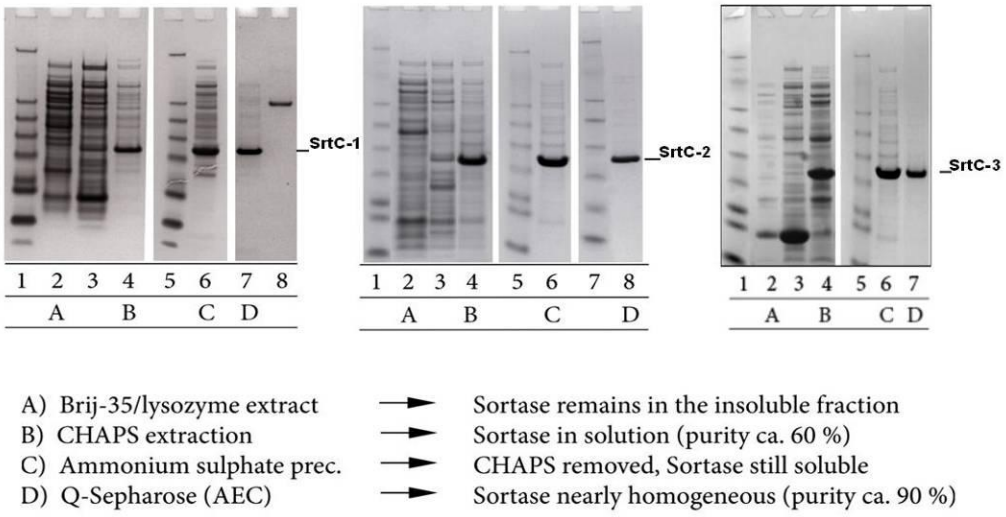


Figure 6: Purification of Sortases and scheme of the four step procedure. The cell pellets were resuspended in a lysis mix containing Brij-35 /lysozyme, then in CHAPS detergent. The supernatant was supplemented with ammonium sulphate and dialyzed in Q-Sepharose buffer. An SDS-PAGE analysis of specific fractions obtained from SrtC-1, SrtC-2 and SrtC-3 purification steps is shown.

3.3 Sortases SrtC-2 and C-3 of Pilus Island 1 are redundant *in vitro*.

Activities of SrtC-1, C-2 and C-3 were tested with synthetic oligopeptides as described in *Materials and Methods*. The peptides were FRET-labeled with Dabcyl/Edans allowing for detection of cleavage by measuring the resulting increase in fluorescence intensity.

In a typical experiment, peptides were incubated with the recombinant sortase enzymes in a temperature-controlled micro-cuvette at 37°C and the development of fluorescence was measured over time.

The three sortases display different cleavage activities on synthetic peptides *in vitro,* but no dramatic differences in the substrate specificities were observed (Figure 7). SrtC-1 shows the highest specific activity with the RrgB-related peptide and a lower activity with RrgA and RrgC-related peptides. In the case of SrtC-2 it has been obtained a very similar profile with a lower specific activity. SrtC-3 has the lowest activity of all three sortases with RrgB and RrgC-related peptide but the highest with the RrgC substrate.

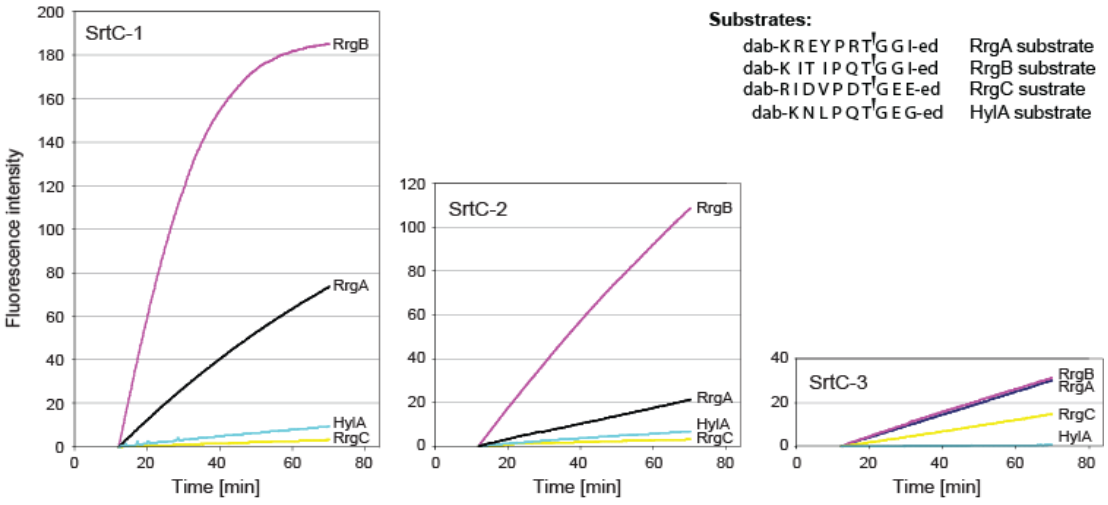


Figure 7: Sortase-catalyzed cleavage reactions. Activities of SrtC-1, C-2 and C-3 were tested with synthetic oligopeptides as described in *Materials and Methods*. The peptides were FRET-labeled with Dabcyl/Edans allowing for detection of cleavage by measuring the resulting increase in fluorescence intensity. In a typical experiment, peptides were incubated with sortase in a temperature-controlled micro-cuvette at 37°C and the development of fluorescence was measured over time.

To see if the surprisingly relaxed substrate specificity of PI-1 sortases *in vitro* is also reflected by the *in vivo*-requirements for pilus assembly, sortase deletion mutants of *S. pneumoniae* strain TIGR4 were constructed and analyzed. In detail, mutanolysine extracts of the mutants were separated on 4-12 % PAA gradient gels and pilus components were detected by western blotting using polyclonal antisera α-RrgA, B or C as indicated above. Presence of assembled pili is indicated by high molecular weight (HMW) multimers of RrgB, the pilus backbone protein, and by incorporation of the ancillary proteins RrgA and RrgC into the HMW material (Figure 8 upper row).

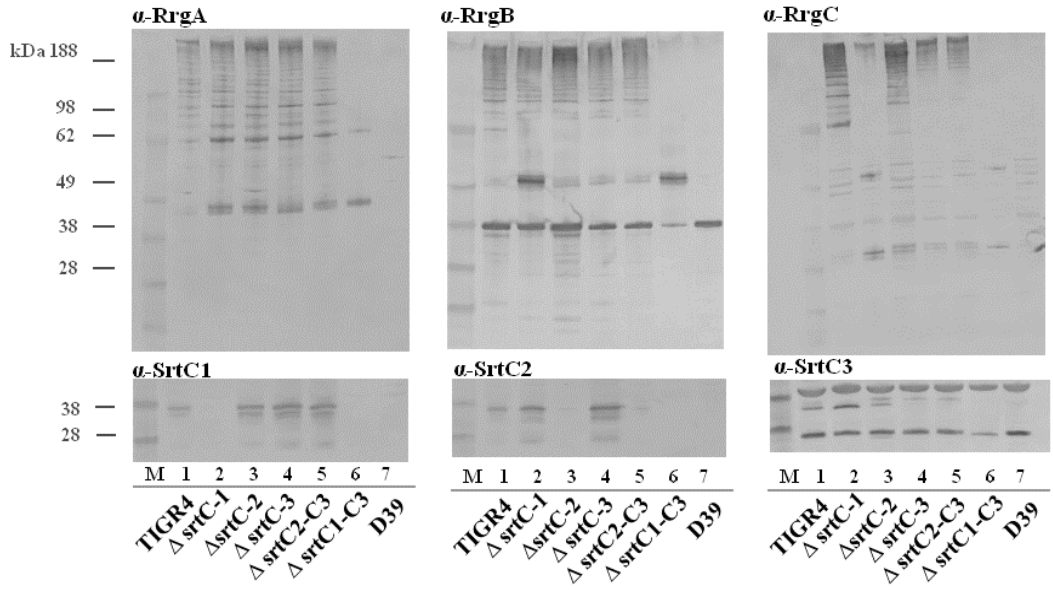


Figure 8: Sortases SrtC-2 and C-3 of Pilus Island 1 are redundant. Production of polymeric high-molecular-weight (HMW) cell wall-associated pili in isogenic T4 and sortase mutant strains was evaluated by immunoblotting detecting pilus components RrgA, RrgB and RrgC (upper gels) and pilus related sortases SrtC-1, SrtC-2 and SrtC-3 (lower gels). In all cases, cell wall proteins were separated by gradient SDS-PAGE, transferred to nitrocellulose membranes and probed against specific antisera. Immunoblots of RrgA, RrgB and RrgC and pilus-related sortases in wild-type T4 (TIGR4), T4 Δ*srtC-1*, T4 Δ*srtC-2*, T4 Δ*srtC-3*, T4 Δ*srtC-2-C-3*, T4 Δ*srtC-1-C3* are shown. Pilus-freeD39 strain was used as control. Lack of polymer formation in T4 Δ*srtC1-C3* shows that at least one of the three pilus-associated sortases, SrtC-1, SrtC-2 or SrtC-3, is required for pilus biogenesis. SrtC-1 and SrtC-2 are sufficient for pilus polymerization, as shown by T4 Δ*srtC3 and* T4 Δ*srtC2-C3*, while SrtC-3 is not capable of pilin polymerization T4 Δ*srtC1-C3*.

Indeed, in the *srtC-2* and the *srtC-3* deletion mutants and even in the double mutant strain lacking *srtC-2* and *srtC-3*, efficient incorporation of RrgB and RrgA into HMW structures was found. RrgC incorporation in the double mutant seems reduced but is clearly detectable. This finding demonstrates that SrtC-2 and SrtC-3 are dispensable for assembly of the complete pilus and suggests that SrtC-1 contributes the principal catalytic activity. The same extract was probed with antisera α-SrtC-1, C-2 or C-3 to confirm the absence or presence of sortases in the respective mutant strains (Figure 8 lower row).

3.4 Nucleophiles influence Sortase-catalyzed cleavage reactions.

Sortase enzyme activity was characterized biochemically by analyzing the influence of H_2N-OH, NH_2–Gly_3 and Lys. The velocity of reactions was influenced by the presence of these nucleophiles. Hydroxylamine enhances the velocity of the reaction in a dose-dependent manner, probably by acting as a nucleophile to resolve transpeptidation intermediate.

Gly_3 is known to stimulate the housekeeping sortase SrtA by mimicking the peptide part of lipid II that, in the course of the transpeptidation reaction, becomes attached to the C-terminus of the protein destined for being anchored to the peptidoglycan. Interestingly, Gly_3 inhibits the in *vitro* reaction of pilus-related sortases, while lysine, which is supposed to provide the ε-amino group as a nucleophile in pilus polymerization, stimulates SrtC-1 and C-2 within a narrow range of concentrations, around 20 mM, suggesting different mechanisms of action for class C and housekeeping sortases. Enhancement of peptide concentration was correlated with sortases activity; however, kinetic studies revealed that sortases do not show simple Michaelis-Menten kinetics. The cleavage reaction is inhibited upon increasing the substrate concentration. In the case of SrtC-1 and SrtC-2 the highest velocity is achieved at a substrate concentration around 5 μM (not shown).

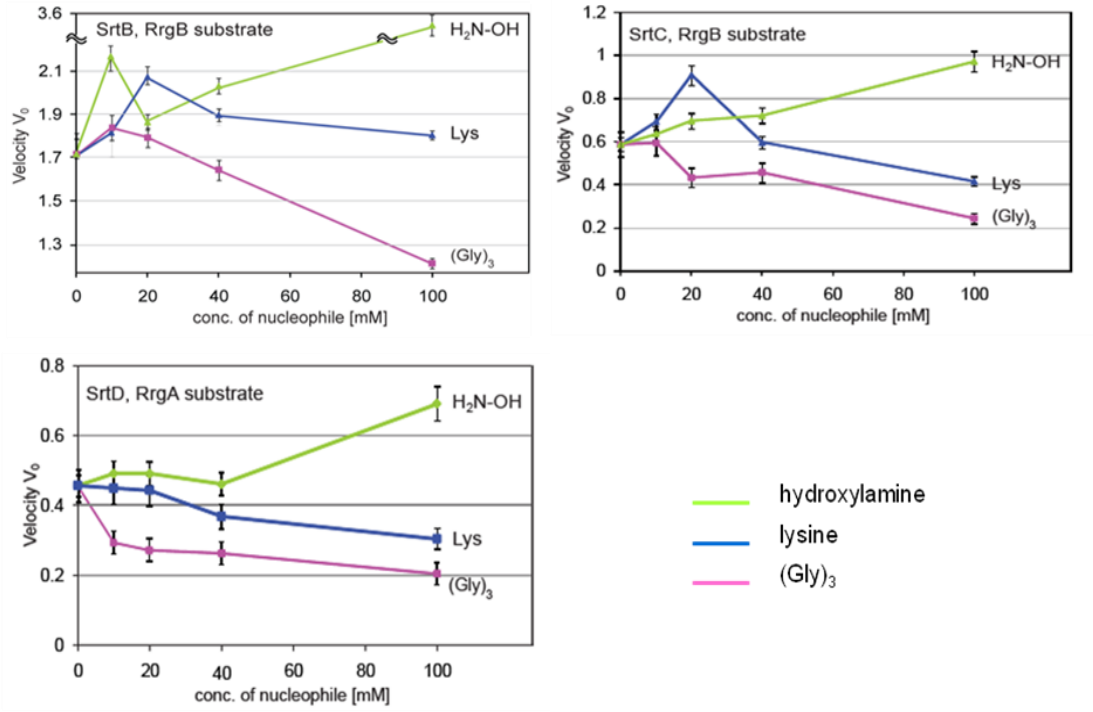


Figure 9: Nucleophiles influence the velocity of sortase-catalyzed cleavage reactions.
A Fluorescent Resonance Energy Transfer (FRET)-based assay was used to evaluate sortase activity and nucleophile specificity. Sortases were incubated at 37°C with the indicated substrates. Reactions were carried out in RP buffer with 8 µM of enzyme, 25 µM peptide substrate and various concentrations of H_2N-OH, Gly_3 or Lys in a total volume of 300 µl. The increase of the fluorescence was measured at 37°C at λ_{em} 490 and λ_{abs} 360 for 70 min using a LS 55 Fluorescence Spectrometer (PerkinElmer) and the resulting reaction velocity V_0 was plotted against the relative nucleophile's concentration.

3.5 Sortases catalyze specific transpeptidation reactions C-terminal processing of RrgB by SrtC-1.

To demonstrate participation of the pilin motif in the transpeptidation reaction and to mimic the reaction that leads to pilus polymerization *in vitro*, two FRET-labeled synthetic peptides were designed: one contained the RrgB-related LPXTG motif, the other one the pilin motif YPKN with the specific lysine residue supposed to form the intermolecular isopeptide bond linking adjacent subunits in the assembled pilus. LPXTG and YPKN peptides were labelled at the N-terminus with dabcyl and at the C-terminus with edands, respectively. While the fluorescence of the separate peptides, due to the presence of the dabcyl fluorophore is high, covalent linkage of the two peptides results in a drop of fluorescence intensity by the FRET effect (Figure 10). Indeed, upon incubating a mixture of the two substrates in the presence of SrtC-1 a drop in fluorescence was observed that exceeds that observed in the absence of sortase. This observation could be a first hint of transpeptidase activity in vitro.

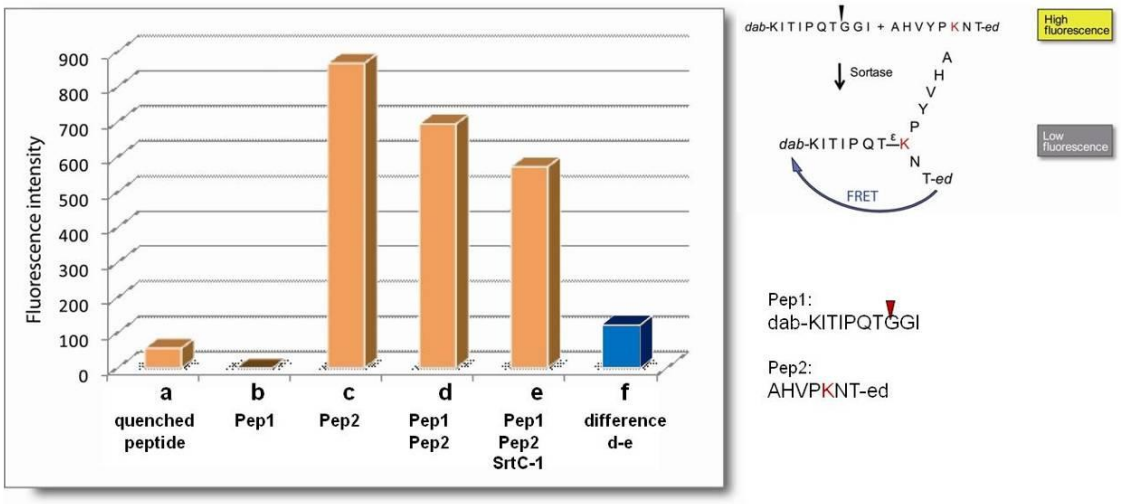


Figure 10: Is SrtC-1 catalyzing tranpeptidation *in vitro*? In a first attempt to demonstrate transpeptidation activity in vitro, synthetic peptides were incubated either alone (a,b,c), as a mixture (d) or in presence of SrtC-1. Due to the inner filtering effect, the mixture of peptides carrying fluorophore and quencher shows lower fluorescence intensity. A further drop in intensity, the extent of which is show in (f) was observed after incubation of the peptides with SrtC-1. This drop might be attributed to the expected FRET effect following ligation of the two peptides. Peptide 1: **dab-KITIPQTGGI**, Peptide2: **AHVPKNT-ed**.

3.6 Characterization of cleaved products by mass spectrometry in sortases/peptide reactions.

Cleavage products generated from LPXTG-like peptides by sortase activity *in vitro* were analyzed by mass spectrometry. As a control a reaction of dab-KNLPQTGEG-ed peptide in the presence of housekeeping sortase from *S. aureus* was carried out either in presence or absence of Gly_3.

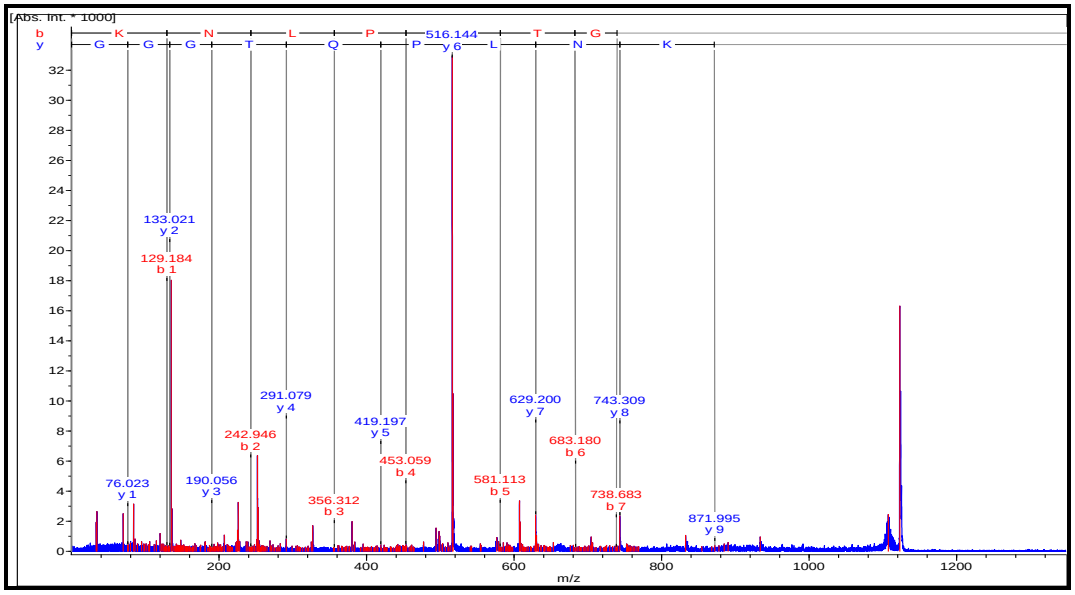


Figure 11: Mass spectroscopy analysis of cleaved product by S.aureus SrtA. LPXTG peptide cleavage by SrtA were carried out in a reaction mixtures containing RP buffer in presence or absence of Gly_3 in a total volume of 300 µl. Reactions were initiated by the addition of 8 µM of sortase (final concentration) and incubated at 37°C for 4 h. 100 µl aliquots from each reaction were separated on a µRPC C2/C18 ST 4.6/100 using a linear gradient from 20 to 80% CH_3CN/0,1% TFA over 30 minutes. To confirm substrate cleavage and the identity of the products, peaks were collected, concentrated and analyzed by MALDI-TOF MS. The mass profile shown above indicates that cleavage occurred between the threonyl and glycyl residues of the LPXTG motif.

To confirm cleavage and to identify the products peaks from reverse phase chromatography were collected, concentrated and analyzed by MALDI-TOF MS (Figure 11).

Table 2 lists predicted molecular weights of HlyA products that could be generated in the presence of housekeeping sortase SrtA and various nucleophiles. Peak fractions obtained after RP chromatography contained a peptide with a molecular weight of 1122.81 (Yellow row) as determined by mass spectrometry. This corresponds to the sum of the masses of dab-KNLPQT HlyA peptide and the Gly_3 nucleophile, confirming that the transpeptidation occurred at the expected site.

HlyA peptide	Predicted MW (KDa)
dab-KNLPQTGEG-ed	1442,59322
dab-KNLPQT	951,07648
dab-KNLPQTGGG	1122,23956
dab-KNLPQT –NHOH h	966,08648
dab-KNLPQTG	1008,127884
dab-KNLPQTGGGG	1179,28192

Table 2: Molecular weights of possible HlyA fragments produced after cleavage by housekeeping sortase SrtA. Various nucleophiles (water, H_2N-OH, and Gly_3) were considered. A yellow background marks the fragment actually detected by mass spectroscopy.

The assay now validated using housekeeping sortase SrtA was then applied for the pilus related sortases (Figure 12-13). While the cleavage site seen with the housekeeping sortase SrtA of *S. aureus* was confirmed to be located between the threonyl and glycyl residues in the LPXTG motif (Figure 11), masses of fragments generated with the pilus related sortases indicate an unusually shifted cleavage site, located after the glycyl residue.

Table 3 shows predicted molecular weights of RrgA and RrgB fragments. Peak analyzed in mass spectrometry correspond to a molecular weight of 1257,37 for RrgA and 1108,25 for RrgB.

RrgA peptide	Predicted MW (Kda)
dab-KREYPRTGGI-ed	1675,87
dab-KREYPRT-CONHOH	1215,33
dab-KREYPRT-COOH	1200,32
dab-KREYPRTG-CONHOH	1272,38
dab-KREYPRTG-COOH	1257,37
H_2N-GGI-ed	493,57
H_2N-GI-ed	436,52
RrgB peptide	
dab-KITIPQTGGI-ed	
dab-KITIPQT-CONHOH	1066,21
dab-KITIPQT-COOH	1051,20
dab-KITIPQTG-CONHOH	1123,26
dab-KITIPQTG-COOH	1108,25
H_2N-GGI-ed	493,57
H_2N-GI-ed	436,52
dab-KITIPQT-iK	1195,37
dab-KITIPQTG-iK	1252,42

Table 3: Molecular weights of predicted RrgA and RrgB fragments produced after cleavage by SrtC-1. A yellow background marks the fragment actually detected by mass spectroscopy.

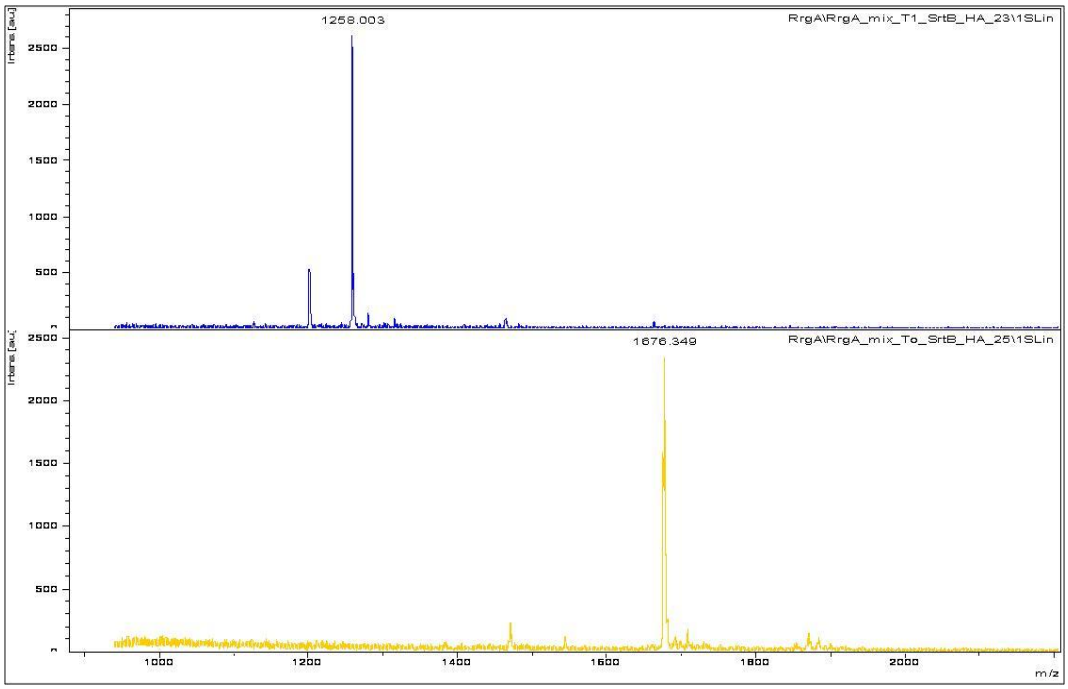


Figure 12: Mass spectroscopy analysis of RrgA peptide cleavage products in the presence of SrtC-1. dab-KREYPRTGGI-ed was incubated in a reaction mixture containing RP buffer in presence 200 mM H_2N-OH s in a total volume of 300 µl. Reactions were initiated by the addition of 8 µM of sortases and incubated at 37°C for 4 h. The mix containing only the pilus-related peptide, only sortase and the complete mix at T_0 were used as controls.100 µl from each reaction were separated on a µRPC C2/C18 ST 4.6/100 using a linear gradient from 20 to 80% CH_3CN/0,1% TFA over 30 minutes. To confirm substrate cleavage and the identity of the products, peaks were collected, concentrated and analyzed by MALDI-TOF MS. The mass profile shown above indicates that cleavage occurred after the glycyl residue of the YPRTG motif.

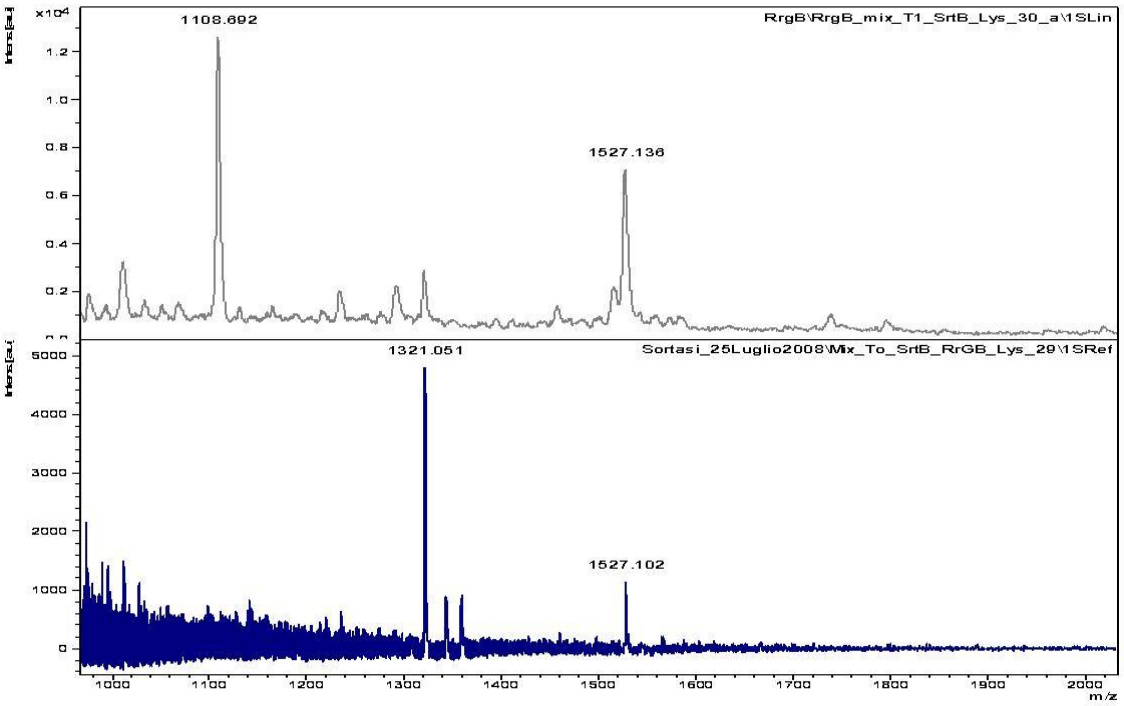


Figure 13: Mass spectroscopy analysis of RrgB peptide cleavage products in the presence of SrtC-1. Peptides were incubated with sortase in a reaction mixture containing RP buffer in presence or absence of 20 mM Lys in a total volume of 300 µl. Reactions were initiated by the addition of 8 µM of sortases and incubated at 37°C for 4 h. The mix containing only the pilus-related peptide, only sortase and the complete mix at T_0 were used as controls.100 µl from each reaction were separated on a µRPC C2/C18 ST 4.6/100 using a linear gradient from 20 to 80% CH_3CN/0,1% TFA over 30 minutes. To confirm substrate cleavage and the identity of the products, peaks were collected, concentrated and analyzed by MALDI-TOF MS. The mass profile shown above indicates that cleavage occurred after the glycyl residue of the IPQTG motif.

Possible explanations for this finding include i) reaction conditions, ii) peptide substrates with an incomplete CWSS missing determinants for cleavage site positioning and iii) absence of the nucleophile native to the transpeptidation reaction.

3.7 PI-1 positive bacteria differentially express the pilus on their surface

To evaluate localization and co-localization of sortases with the major pilus components on whole bacteria, immunofluorescence experiments were performed.

First, to evaluate pilus expression on the bacterial surface, bacteria were processed for immunofluorescence and then stained with anti RrgB antibodies and anti *S. pneumoniae* capsular antibodies (Omniserum).

The RrgB pilus subunit was clearly detected on the bacterial surface. In previous topological studies, Falker and co-workers analyzed the expression and the role of pneumococcal sortases (Falker et al., 2008), and showed that, in the presence of pilus-related sortases, the pilus components are localized in symmetric foci at the cell surface.

Here, the right panel of Figure 14 shows elongated pili extending from the bacterial suface. High density clusters can be detected on the bacteria, probably deriving from collapsed pili, but more analyses are necessary to confirm the presence of foci.

Interestingly, it was observed that, while the capsular staining was homogeneous, pili were differentially expressed in the pneumococcal population. Two distinct sub-populations of bacteria were revealed, one expressing and the other one not expressing (or expressing at undetectable levels) pili on their surface (Figure 14).

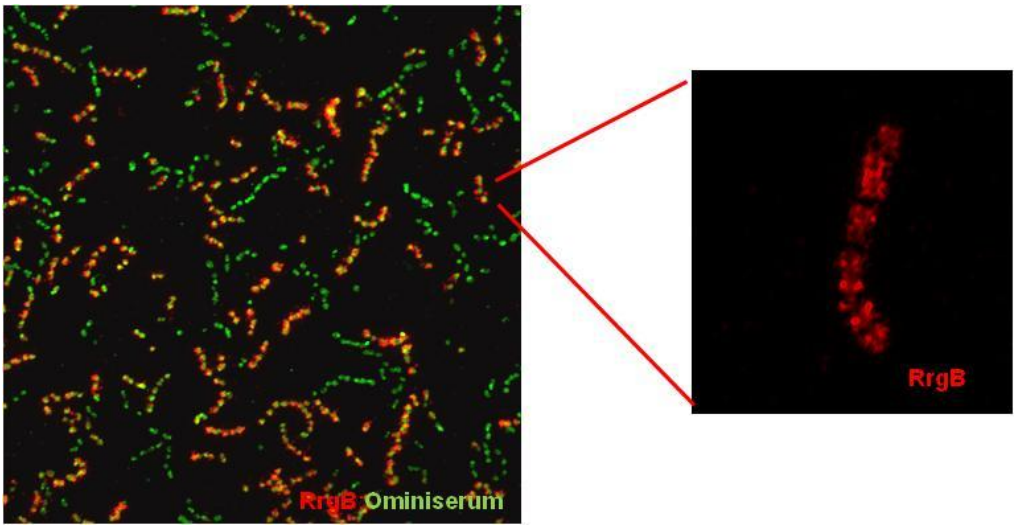


Figure 14: PI-1 positive bacteria differentially express the pilus on their surface by Immunofluorescence. TIGR 4 bacteria were grown to OD_{600}=0.5 and then stained for RrgB (red), and capsule (green) detection. RrgB signal is present only in a portion of the bacterial population suggesting differential expression of pili. The panel on the right shows a zoom of stained bacteria for RrgB. High density clusters are here detected.

The existence of two sub-populations differentially expressing the pilus was also confirmed by flow cytometry experiments, where two clearly distinct peaks can be observed when bacteria were stained with anti RrgB antibodies (Figure 15A). With the attempt to better characterize the two sub-populations, TIGR4 wild type (T4 wt) bacteria were fixed with paraformaldehyde, labeled with primary anti RrgB antibodies and secondary FITC conjugated antibodies, and then sorted by FACSAria into RrgB positive (T4+) and RrgB negative (T4-) populations, on the basis of RrgB signal intensities (Figure 15). The purity of the two distinct populations recovered by FACS sorting is visualized in Figure 15A.

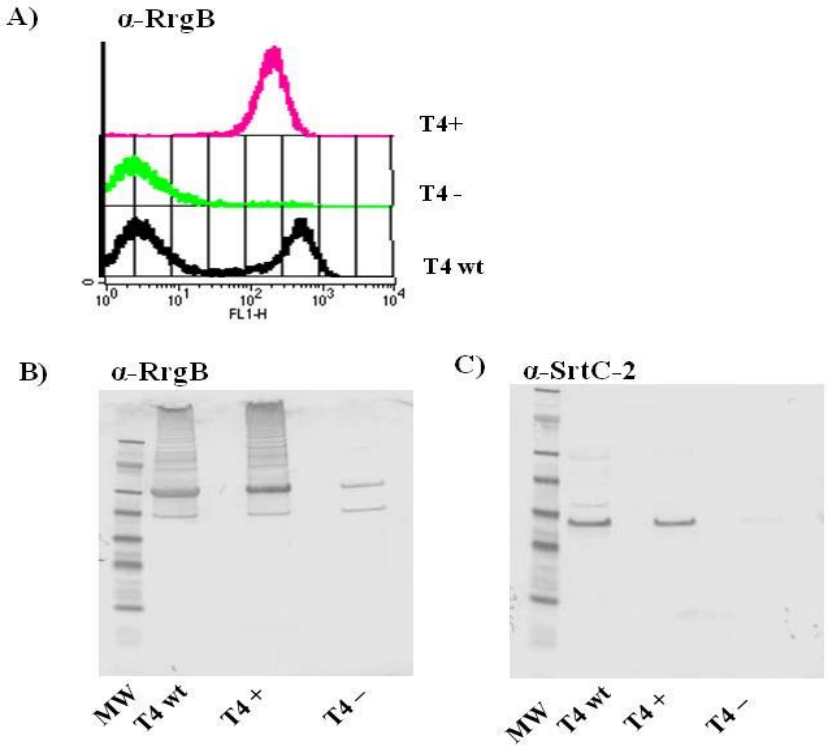


Figure 15: PI-1 positive bacteria differentially express the pilus on their surface. A) FACS analysis and sorting was performed with *S. pneumoniae* TIGR4 (wt) expressing PI-1 islet. Pili were labeled with mouse polyclonal antibodies raised against RrgB and two populations either expressing or not expressing RrgB were observed. Based on this signal, bacteria were sorted into two different populations: TIGR4 pilus positive (T4+) and TIGR4 pilus negative (T4-). B) Western blot analyses of crude cell extracts from either population were performed using antibodies against either RrgB (left panel) or SrtC-2 (right panel). A clear correlation between expression of SrtC-2 and RrgB pilus expression can be observed.

In order to evaluate if there was correlation between pilin and sortases expression, sorted bacteria T4-/T4+ were recovered by centrifugation and then lysed; bacterial lysates were probed by western blot with anti RrgB or SrtC-2 antibodies. (Figure 15 B-C). A clear correlation between expression of SrtC-2 and RrgB expression was detected, as RrgB negative bacteria clearly did not express SrtC-2 (Figure 15C).

In order to evaluate sortase localization in relation to RrgB, bacteria were processed for additional immunofluorescence analysis. Noteworthy, the results previously obtained by western blot analysis of the two sub-populations clearly indicated that, despite its still undetermined localization, the sortases should have displayed a differential expression pattern likewise RrgB.

In previous studies the localization of associated cell wall proteins was facilitated by the use of cell wall hydrolases, able to permeabilize the bacterial cell wall. In the specific case of *S. pyogenes*, a Phage lysine (PlyC) was used to detect the localization of the housekeeping sortase A (Raz and Fischetti, 2008).

In this thesis, the endolysin encoded by pneumococcal Cpl bacteriophage (Clp) was used to partially hydrolyze the pneumococcal cell wall, in the attempt to detect the sortase specific signal. Given the toxicity of the Clp enzyme, to evaluate if there were differences in the signal detected, all the immuno-staining experiments were carried out both in the presence and absence of the enzyme. Although the Clp1 treatment somehow changed the bacterial morphology and pili overall shape (bacteria tended to be more rounded and pili shorter), it did not affect RrgB signal intensity (fig 16 lower panel).

On the other hand, sortase signal obtained with anti SrtC-2 antibodies seems to be localized on the bacterial hull like RrgB does (Figure 16 upper panel), and evenly distributed on the bacterial surface. Interestingly, the signal is detectable independently from Clp treatment also present with similar intensity, in non piliated bacteria.

The latter observation is clearly in contrast with flow cytometry and western blot results, thus indicating that the signal observed with anti SrtC-2 antibodies could be due to a non specific recognition, possibly deriving from the low sortase expression.

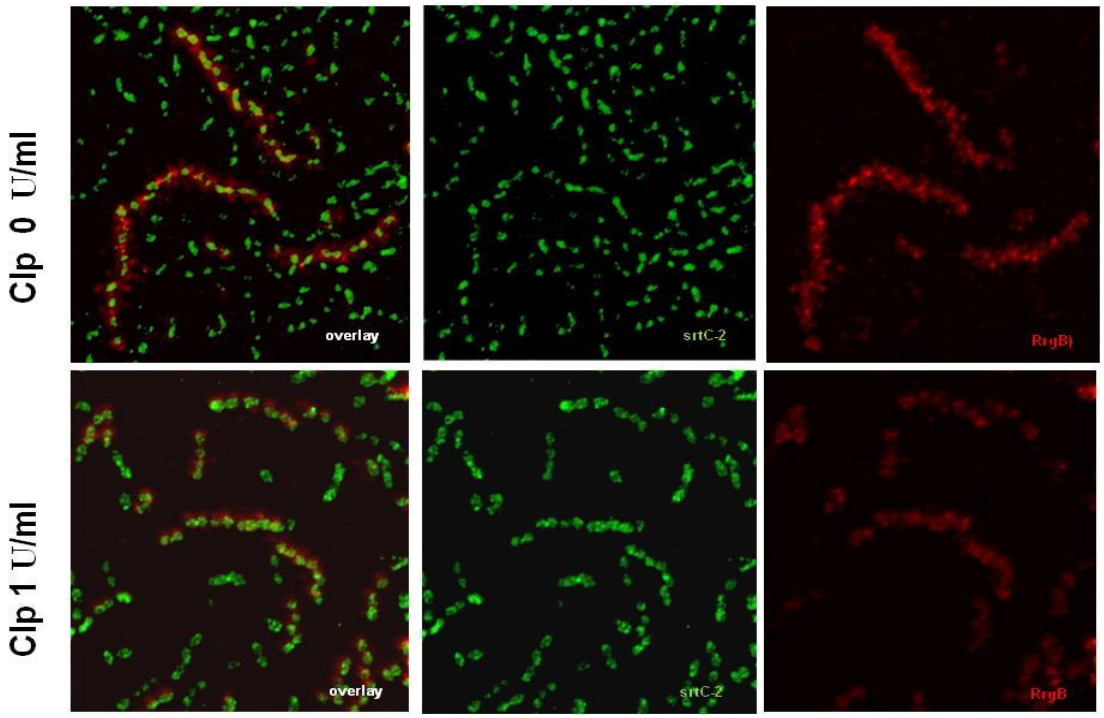


Figure 16: Co-localization of RrgB and SrtC-2 by Immunofluorescence. TIGR4 bacteria were grown to OD_{600}=0.5, stained for RrgB (red), and subsequently labeled for sortase C2 (green). The upper panel shows images of bacteria without Clp1 treatment. The lower panel shows bacteria after treatment with 1U/ml Clp1. No correlation with RrgB expression and sortase localization.

Therefore, to evaluate SrtC-2 antibody specificity, a TIGR4 deletion mutant lacking all the three sortases, but still expressing RrgB (TIGR 4 Δ*srtC1-3*) was analyzed by immunofluorescence with anti RrgB and SrtC-2 antibodies.

The RrgB signal, as expected, evenly surrounds the perimeter of the bacterial cells, but no extracellular structures are detected, while the SrtC-2 signal was undetectable (Figure 17 upper panel). This observation is clearly in contrast to the previously stated hypothesis of antibody non specificity, and does not explain the results obtain in the wt strain with SrtC-2 antibodies.

In conclusion, the results are not sufficient to define a clear localization of pilus-related sortases, and further investigations are still ongoing. In detail, one of the most promising approaches could be the expression of one sortase at a time in a TIGR 4 Δ*srtC1*-3 background by using a plasmid able to replicate in *S. pneumoniae*. The high levels of expression reachable with this expression system in a background where non specific signal is absent should overcome both the low endogenous sortase expression problem and the cross-reactivity observed in the wt strain.

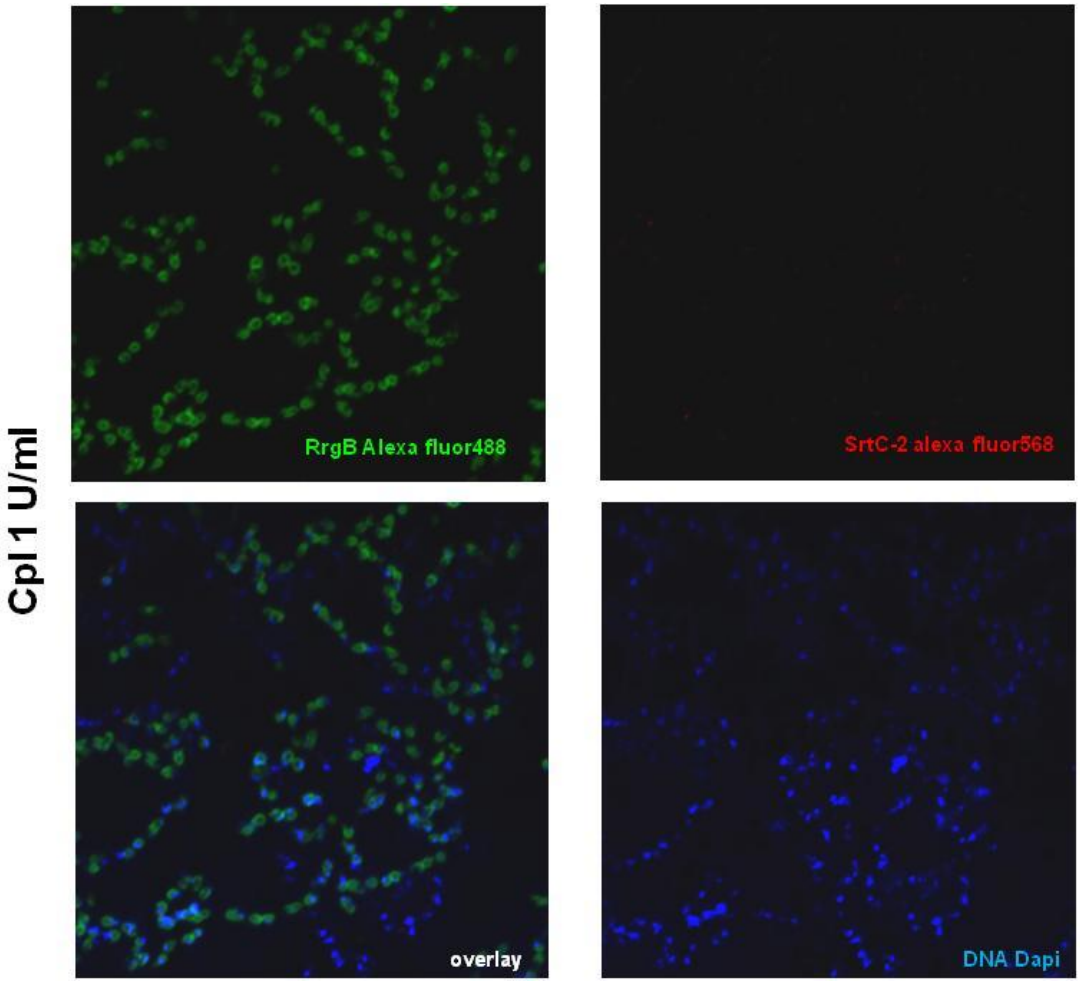


Figure 17: Immunofluorescence of *S. pneumoniae* TIGR 4 ΔsrtC1-3: sortases are required for assembly of extracellular structures. TIGR 4 Δ*srtC1-3* bacteria were grown to OD_{600}=0.5, stained for RrgB (green), subsequently labeled for sortase C1 (red) and labeled with DAPI for DNA (blue).

4. Discussion

Pilus assembly in Gram positive bacteria is an area in microbiology that currently is under development. Since these structures were found, covering the surface of certain Gram positive bacteria, a lot of work has been invested to reveal how pilus components are assembled into an extended pilus.

The *S. pneumoniae* pilus encoded by Pilus islet 1 (PI-1) is composed of a major pilus component, designated RrgB constituting the pilus backbone, while the two minor pilus subunits RrgA and RrgC seem to be located at the tip and the base of the pilus, respectively (Hilleringmann et al., 2009). PI-1 is present in only 30% of pneumococcal strains; it contains three specialized sortases, SrtC-1, SrtC-2 and SrtC-3 that are involved in polymerization and maturation of the pilus. In this thesis, the role of pilus-related sortases in pilus biogenesis of *Streptococcus pneumoniae* has been investigated.

Pilus-1 related sortases in *S. pneumoniae* were characterized using a two-pronged approach, involving biochemical assays and mutant studies. On the basis of the *in vitro* data it was possible to determine the substrate specificities of PI-1 related sortases, data that were also confirmed by *in vivo* studies.

No dramatic differences in the substrate specificities of the three PI-1 sortases were observed. Although SrtC-1 displays the highest cleavage activity with the RrgB-related peptide, all three sortases cleave also RrgA and RrgC-related peptides. SrtC-3 has the lowest specific activity of all sortases with RrgA and B-related substrates but the highest with the RrgC-type substrate.

A clear stimulating effect of both hydroxylamine and lysine on all three PI-1 sortases was observed. While the effect exerted by hydroxylamine, being a small, highly reactive nucleophile, is not surprising, stimulation by lysine might mimic the function of the conserved lysine residue present in virtually all pilus backbone proteins characterized to date.

The ε-amino group of this residue is thought to form, in a sortase-catalyzed reaction, an iso-peptide bond with the C-terminal threonyl residue of the CWSS, linking adjacent RrgB pilus subunits. Iteration of this process leads to formation of the pilus filament, while termination and/or initialization of this polymerization reaction could occur by incorporation of RrgC and RrgA subunits at the proximal or at the distal end of pilus shaft respectively, as demonstrated recently by TEM and scanning transmission EM analysis (Hilleringmann et al., 2009).

RrgA and RrgC seem to be devoid of the pilin motif seen in RrgB, the only pilus component containing both pilin and LPxTG like-motifs and also the only one which has been observed to polymerize.

Absence of polymerization in ancillary pilus subunits supports the proposed model of pilus assembly in *S. pneumoniae* (Fig. 18) (Hilleringmann et al., 2009). Since RrgA apparently does not possess a pilin motif, this model predicts its incorporation only at the tip of the pilus. Location of RrgA at the pilus tip is also consistent with its role as an adhesin as reported in several studies (Bagnoli et al., 2008;Nelson et al., 2007).

In contrast, the ability of RrgC to form heterodimers with RrgA (Hilleringmann et al., 2009;Lemieux et al., 2006a) is consistent with its presence at the base of the pilus, suggesting the existence of a site on RrgC that can be linked to an LPxTG-like motif.

In addition, as shown by *in vitro* experiments reported in this thesis, the RrgC related LPxTG-like motif is cleaved more readily by SrtC-3 which therefore could possess a role in linking the pilus to the cell wall. SrtC-3 seems to have a crucial role in presentation of pilus at the cell surface (Falker et al., 2008).

As demonstrated in this work, absence of SrtC-3 does not influence pilus assembly in the respective mutant strains, although topological studies suggest that SrtC-3 in combination with SrtC-1 is responsible for symmetric focal distribution of pili (Falker et al., 2008).

In *in vitro* polymerization experiments involving SrtC-1, RrgC has been found in association with RrgB (El et al., 2010) suggesting that the IPQTG motif of RrgB might form an isopeptide bond with an unknown residue of RrgC.

Evidence for intermolecular isopeptide bond formation in pili of Gram-positive bacteria is provided by studies conducted on native pili of *S. pyogenes* (Budzik et al., 2009;Kang et al., 2009;Linke et al., 2010). Mass spectroscopy analysis in *S. pyogenes* confirmed the linkage between contiguous backbone-pilus subunits, one containing a Lys residue and the other containing the Thr residue in the LPxTG-like motif, cleaved by sortase (Kang et al., 2007).

A recent study on *S. pneumoniae* pilus organization shows a similar head-to-tail structure (Spraggon et al., 2010). Based on sequence alignments with Spy0128 (*S. pyogenes*) in which the linking Lys residues was identified (Kang et al., 2009), two possible lysine residues were identified in RrgB as candidates potentially involved in intermolecular bond formation.

Mutant studies with *S. pneumoniae* TIGR4, inactivating PI-1 sortases individually or in combinations, confirmed the results of the *in vitro* analysis concerning substrate specificity of PI-1 sortases (see results). Based on mutant strain analysis, it was demonstrated that the presence of a single PI-1 sortase, SrtC-1, is sufficient for pilus backbone polymerization and for incorporation of the ancillary proteins RrgA and RrgC.

Previous studies already had shown the redundancy of SrtC-2 and SrtC-3 in pilus assembly (Falker et al., 2008;Lemieux et al., 2008), data that were confirmed and reproduced also by the *in vitro* model presented here.

It is still not known how the ancillary pilin subunits are attached to the pilus backbone and in which order.

Immunoblot analysis, presented also in this thesis, demonstrates high molecular weight polymers that react with antibodies directed against ancillary pilus subunits, suggesting a covalent attachment to pilus polymers of the latter. On the other hand, sortase specificity appears to play a role in the incorporation of minor pilins, while housekeeping sortase could play a role in pilus anchoring as it does with other LPxTG containing proteins resulting in covalent attachment to the cell surface (Nobbs et al., 2008). While anchoring of pili to the cell wall by housekeeping sortases has been observed in other species, in *S. pneumoniae* this step appears to be independent of SrtA (Lemieux et al., 2006b).

In fact, the role of the housekeeping sortase SrtA in pilus biogenesis was investigated in a parallel analysis demonstrating that that housekeeping sortase appears not to be involved in pilus assembly and anchoring (data not shown). In SrtA mutant strains pili are detectable in crude lysates but not in supernatant fractions (not shown).

In an attempt to elucidate the mechanism of pilus biogenesis further biochemical analyses were performed. To confirm structure and identity of cleaved products produced by pilus related sortases, an *in vitro* assay has been developed using sortases in the presence of substrate peptides, followed by mass spectrometric analysis of cleavage products.

While the well-known cleavage site for the housekeeping sortase SrtA from *S. aureus* was confirmed to be located between the threonyl and glycyl residues in the LPxTG motif, fragments generated with the pilus related sortases from *S. pneumoniae* suggest an unusually shifted cleavage site located after the glycyl residue.

Recently, data have been published on processing of RrgB by SrtC-1 *in vitro* (El et al., 2010;Manzano et al., 2008) showing the formation of a heterodimer between RrgB and RrgC. Interestingly, the mass of this heterodimer, determined by mass spectroscopy, is consistent with an RrgC molecule associated with RrgB cleaved behind the threonyl of the IPQTG motif of RrgB (El et al., 2010), hence at the typical cleavage site described for various types of sortases, but inconsistent with the *in vitro* results described here.

Possible reasons for this discrepancy could consist in i) a substrate peptide that is missing specificity determinants required for correct positioning of the reactive cystein residue with respect to the LPxTG-like motif or ii) the absence of the native incoming nucleophile (the pilin motif). For the pilus-related sortases, it is not known if and how this nucleophile contributes to the specificity of the reaction. In this respect it is noteworthy that structural analysis of the PI-1 sortases (Manzano et al., 2009) revealed a lid region that covers the active site and that is not present in housekeeping sortases.

It has been speculated that opening of this lid is necessary for providing access of substrate molecules to the enzyme's active site. While the roles of the incoming nucleophile or the LPxTG-like motif in lid-opening are unknown, impaired access of synthetic peptides might in fact result in the observed improper cleavage.

Additional studies were conducted to visualize distinct foci for pilus related sortases in association with RrgB by confocal IF microscopy. While another study already reported on pilus topology in *S. pneumoniae* (Falker et al., 2008), the attempt to co-localize pilus-related sortases with the RrgB pilus component failed. The analysis demonstrated that the RrgB signal is homogeneously distributed on the cell surface without displaying distinct foci as reported before (Falker et al., 2008), while the sortase signal appeared unclear and patchy.

Interestingly, fluorescence microscopy analysis revealed that all the strains presented two sub-populations of bacteria, one expressing the pilus at high levels, while the pilus was undetectable in the other. This profile was confirmed by western blot analysis.

A clear correlation between expression of SrtC-2 and RrgB expression was detected, as RrgB negative bacteria clearly did not express SrtC-2. Confocal IF microscopy studies did not reveal the same profile: a constant weak signal was observed with anti SrtC-2 antibodies that probably is due to a non specific reaction, while sortase expression is too low to be detected by this method.

Additional analysis will be necessary to overcome the technical limits encountered in the immunofluorescence studies, in order to overexpress one sortase at a time avoiding background signals and cross-reactivity observed in the wild type. Further studies are underway to explain the bipolar expression behavior and the regulatory mechanism governing pilus expression in pneumococci.

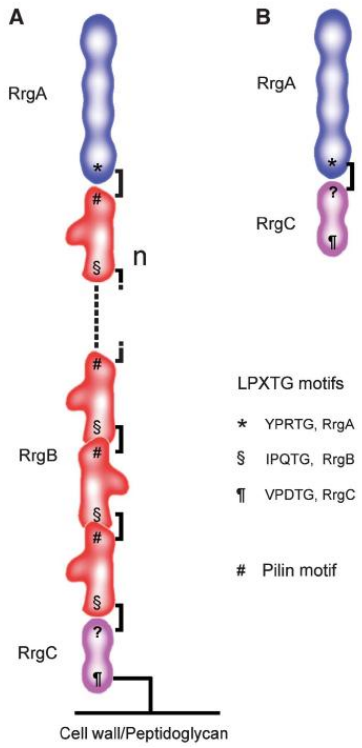


Figure 18: Model of *S. pneumoniae* pilus. (A) The TIGR4 pilus consists of a shaft composed of RrgB, with RrgA at its distal and RrgC at its proximal end. Sortase SrtC-1 mediates the polymerization of RrgB (red) via the LPxTG-like motif (*, IPQTG) and the pilin motif (Manzano et al, 2008) into a single string of monomers covalently linked head-to-tail. Polarity of the nose-like feature of RrgB allowed for localization of RrgA at the distal end of the pilus, while RrgC is at the pilus base, proximal to the bacterium.
(B) The ability of RrgA and RrgC to form a heterodimer (LeMieux et al, 2008) suggests a site in RrgC that can be covalently linked to an LPxTG-like motif from either RrgA or RrgB.
Figure from (Hilleringmann et al., 2009)

In conclusion, this work studies the specificity of pilus-related sortase activity, revealing an interesting redundancy of sortases. The pilus expression profile opens new questions concerning the regulation and role of this virulence factor described as an eligible component for a pneumococcal vaccine.

References Cited

1997, Prevention of pneumococcal disease: recommendations of the Advisory Committee on Immunization Practices (ACIP): MMWR Recomm.Rep., v. 46, no. RR-8, p. 1-24.

Bagnoli, F. et al., 2008, A second pilus type in Streptococcus pneumoniae is prevalent in emerging serotypes and mediates adhesion to host cells: J.Bacteriol., v. 190, no. 15, p. 5480-5492.

Barocchi, M. A. et al., 2006, A pneumococcal pilus influences virulence and host inflammatory responses: Proc.Natl.Acad.Sci.U.S.A, v. 103, no. 8, p. 2857-2862.

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., v. 45, no. 2, p. 411-421.

Bogaert, D., B. A. van, M. Sluijter, A. Luijendijk, G. R. de, H. C. Rumke, H. A. Verbrugh, and P. W. Hermans, 2004, Colonisation by Streptococcus pneumoniae and Staphylococcus aureus in healthy children: Lancet, v. 363, no. 9424, p. 1871-1872.

Budzik, J. M., S.-Y. Oh, and O. Schneewind, 2009, Sortase D forms the covalent bond that links BcpB to the tip of Bacillus cereus pili: Journal of Biological Chemistry, v. 284, no. 19, p. 12989-12997.

Dramsi, S., S. Magnet, S. Davison, and M. Arthur, 2008, Covalent attachment of proteins to peptidoglycan: FEMS Microbiol.Rev., v. 32, no. 2, p. 307-320.

El,ML, R Terrasse, A Dessen, T Vernet, A M Di Guilmi. Stability and Assembly of Pilus Subunits of Streptococcus pneumoniae. Journal of Biological Chemistry 285[16], 12405-12415. 2010.
Ref Type: Magazine Article

Falker, S., A. L. Nelson, E. Morfeldt, K. Jonas, K. Hultenby, J. Ries, Melefors O., S. Normark, and B. Henriques-Normark, 2008, Sortase-mediated assembly and surface topology of adhesive pneumococcal pili: Molecular Microbiology, v. 70, no. 3, p. 595-607.

Falugi, F. et al., 2008, Sequence variation in group A Streptococcus pili and association of pilus backbone types with lancefield T serotypes: J.Infect.Dis., v. 198, no. 12, p. 1834-1841.

Fletcher, M. A., and B. Fritzell, 2007, Brief review of the clinical effectiveness of PREVENAR against otitis media: Vaccine, v. 25, no. 13, p. 2507-2512.

Gianfaldoni, C. et al., 2007, Streptococcus pneumoniae pilus subunits protect mice against lethal challenge: Infect.Immun., v. 75, no. 2, p. 1059-1062.

Givon-Lavi, N., D. Fraser, N. Porat, and R. Dagan, 2002, Spread of Streptococcus pneumoniae and antibiotic-resistant S. pneumoniae from day-care center attendees to their younger siblings: J.Infect.Dis., v. 186, no. 11, p. 1608-1614.

Henrichsen, J., 1995, Six newly recognized types of Streptococcus pneumoniae: J.Clin.Microbiol., v. 33, no. 10, p. 2759-2762.

Hilleringmann, M., F. Giusti, B. C. Baudner, V. Masignani, A. Covacci, R. Rappuoli, M. A. Barocchi, and I. Ferlenghi, 2008, Pneumococcal pili are composed of protofilaments exposing adhesive clusters of Rrg A: PLoS Pathog., v. 4, no. 3, p. e1000026.

Hilleringmann, M., P. Ringler, S. A. Muller, A. G. De, R. Rappuoli, I. Ferlenghi, and A. Engel, 2009, Molecular architecture of Streptococcus pneumoniae TIGR4 pili: EMBO J., v. 28, no. 24, p. 3921-3930.

Kang, H. J., F. Coulibaly, F. Clow, T. Proft, and E. N. Baker, 2007, Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure: Science, v. 318, no. 5856, p. 1625-1628.

Kang, H. J., M. Middleditch, T. Proft, and E. N. Baker, 2009, Isopeptide bonds in bacterial pili and their characterization by X-ray crystallography and mass spectrometry: Biopolymers, v. 91, no. 12, p. 1126-1134.

Lemieux, J., D. L. Hava, A. Basset, and A. Camilli, 2006a, RrgA and RrgB are components of a multisubunit pilus encoded by the Streptococcus pneumoniae rlrA pathogenicity islet: Infect.Immun., v. 74, no. 4, p. 2453-2456.

Lemieux,J, D L Hava, A Basset, A Camilli. RrgA and RrgB are components of a multisubunit pilus encoded by the Streptococcus pneumoniae rlrA pathogenicity islet. Infection & Immunity 74[4], 2453-2456. 2006b.
Ref Type: Magazine Article

Lemieux, J., S. Woody, and A. Camilli, 2008, Roles of the sortases of Streptococcus pneumoniae in assembly of the RlrA pilus: Journal of Bacteriology, v. 190, no. 17, p. 6002-6013.

Linke, C., P. G. Young, H. J. Kang, R. D. Bunker, M. J. Middleditch, T. T. Caradoc-Davies, T. Proft, and E. N. Baker, 2010, Crystal structure of the minor pilin FctB reveals determinants of Group A streptococcal pilus anchoring: J.Biol.Chem., v. 285, no. 26, p. 20381-20389.

Maiden, M. C., 1998, Horizontal genetic exchange, evolution, and spread of antibiotic resistance in bacteria: Clin.Infect.Dis., v. 27 Suppl 1, p. S12-S20.

Maione, D. et al., 2005, Identification of a universal Group B streptococcus vaccine by multiple genome screen: Science, v. 309, no. 5731, p. 148-150.

Mandlik, A., A. Swierczynski, A. Das, and H. Ton-That, 2008, Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development: Trends Microbiol., v. 16, no. 1, p. 33-40.

Manzano, C., C. Contreras-Martel, M. L. El, T. Izore, D. Fenel, T. Vernet, A. M. Di Guilmi, and A. Dessen, 2008, Sortase-Mediated Pilus Fiber Biogenesis in Streptococcus pneumoniae: Structure, v. 16, no. 12, p. 1838-1848.

Manzano, C., T. Izore, V. Job, A. M. Di Guilmi, and A. Dessen, 2009, Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of Gram-positive pathogens: Biochemistry, v. 48, no. 44, p. 10549-10557.

Margarit, I. et al., 2009, Preventing bacterial infections with pilus-based vaccines: the group B streptococcus paradigm: J.Infect.Dis., v. 199, no. 1, p. 108-115.

Moschioni, M. et al., 2008, Streptococcus pneumoniae contains 3 rlrA pilus variants that are clonally related: J.Infect.Dis., v. 197, no. 6, p. 888-896.

Navarre, W. W., and O. Schneewind, 1999, Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope: Microbiol.Mol.Biol.Rev., v. 63, no. 1, p. 174-229.

Nelson, A. L. et al., 2007a, RrgA is a pilus-associated adhesin in Streptococcus pneumoniae: Mol.Microbiol., v. 66, no. 2, p. 329-340.

Nelson, A. L., A. M. Roche, J. M. Gould, K. Chim, A. J. Ratner, and J. N. Weiser, 2007b, Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance: Infect.Immun., v. 75, no. 1, p. 83-90.

Nobbs, A. H., R. Rosini, C. D. Rinaudo, D. Maione, G. Grandi, and J. L. Telford, 2008, Sortase A utilizes an ancillary protein anchor for efficient cell wall anchoring of pili in Streptococcus agalactiae: Infect.Immun., v. 76, no. 8, p. 3550-3560.

O'Brien, K. L. et al., 2009, Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates: Lancet, v. 374, no. 9693, p. 893-902.

Ogunniyi, A. D., M. Grabowicz, D. E. Briles, J. Cook, and J. C. Paton, 2007, Development of a vaccine against invasive pneumococcal disease based on combinations of virulence proteins of Streptococcus pneumoniae: Infect.Immun., v. 75, no. 1, p. 350-357.

Pericone, C. D., D. Bae, M. Shchepetov, T. McCool, and J. N. Weiser, 2002, Short-sequence tandem and nontandem DNA repeats and endogenous hydrogen peroxide production contribute to genetic instability of Streptococcus pneumoniae: J.Bacteriol., v. 184, no. 16, p. 4392-4399.

Proft, T., and E. N. Baker, 2009, Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease: Cell Mol.Life Sci., v. 66, no. 4, p. 613-635.

Raz,A, V A Fischetti. Sortase A localizes to distinct foci on the Streptococcus pyogenes membrane. Proceedings of the National Academy of Sciences of the United States of America 105[47], 18549-18554. 2008.
Ref Type: Magazine Article

Ryan, M. W., and P. J. Antonelli, 2000, Pneumococcal antibiotic resistance and rates of meningitis in children: Laryngoscope, v. 110, no. 6, p. 961-964.

Scott, J. R., and D. Zahner, 2006, Pili with strong attachments: Gram-positive bacteria do it differently: Mol.Microbiol., v. 62, no. 2, p. 320-330.

Sjostrom, K. et al., 2007, Clonal success of piliated penicillin nonsusceptible pneumococci: Proc.Natl.Acad.Sci.U.S.A, v. 104, no. 31, p. 12907-12912.

Spraggon, G. et al., 2010, Supramolecular organization of the repetitive backbone unit of the Streptococcus pneumoniae pilus: PLoS ONE, v. 5, no. 6, p. e10919.

Telford, J. L., M. A. Barocchi, I. Margarit, R. Rappuoli, and G. Grandi, 2006, Pili in gram-positive pathogens: Nat.Rev.Microbiol., v. 4, no. 7, p. 509-519.

Tettelin, H. et al., 2001, Complete genome sequence of a virulent isolate of Streptococcus pneumoniae: Science, v. 293, no. 5529, p. 498-506.

Ton-That,H, G Liu, S K Mazmanian, K F Faull, O Schneewind. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. Proceedings of the National Academy of Sciences of the United States of America 96[22], 12424-12429. 1999.
Ref Type: Magazine Article

Ton-That,H, L A Marraffini, O Schneewind. Protein sorting to the cell wall envelope of Gram-positive bacteria [Review]. Biochimica et Biophysica Acta - Molecular Cell Research 1694[1-3], 269-278. 2004a.
Ref Type: Magazine Article

Ton-That, H., L. A. Marraffini, and O. Schneewind, 2004b, Sortases and pilin elements involved in pilus assemly of Corynebacterium diphtheriae: Molecular Microbiology, v. 53, no. 1, p. 251-261.

Ton-That, H., and O. Schneewind, 1999, Anchor structure of staphylococcal surface proteins. IV. Inhibitors of the cell wall sorting reaction: Journal of Biological Chemistry, v. 274, no. 34, p. 24316-24320.

Ton-That,H, O Schneewind. Assembly of pili on the surface of Corynebacterium diphtheriae. Molecular Microbiology 50[4], 1429-1438. 2003a.
Ref Type: Magazine Article

Ton-That, H., and O. Schneewind, 2003b, Assembly of pili on the surface of Corynebacterium diphtheriae: Molecular Microbiology, v. 50, no. 4, p. 1429-1438.

CHAPTER 2

Sortagging: Design and development of an *Influenza virus* protein array

1. Introduction

1.1 Protein arrays and Sortagging

Protein microarray development represents the challenge of large-scale, qualitative and quantitative revelation of protein-ligand interactions in high-throughput manner (Macbeath, 2002).

In the last decade application of microarray technology has grown exponentially, but a simple and fast technique to immobilize functional proteins on solid supports still represents a central issue (Zhu and Snyder, 2003). The use of specific tags can help to specifically attach proteins to a solid support. Therefore, different strategies have been applied in this direction. Examples are hexahistidine tag fusion proteins or in vivo/in vitro biotinylation procedures, which have the advantage to confer a specific orientation to the conjugated protein. As a drawback, these approaches often give rise to a high level of background, which may compromise the final result of the protein chip itself (Zhu et al., 2001). Chemical procedures or chemioselective reactions between protein α-thioesters and a Cys-containing glass slide are examples of broader but less specific reactions to immobilize proteins to a solid support (Camarero et al., 2004;Kindermann et al., 2003).

The biotechnology field has also an increased demand for modified proteins and protein conjugates and the sortase-catalyzed transpeptidation reaction can be applied for this purpose (Proft, 2010). Sortagging technology represents a powerful tool to create protein conjugates using sortase A from *Staphylococcus aureus.*

Sortase A recognizes substrates with an LPXTG motif, cleaving the peptide bond between the threonine and glycine and resulting in a thioacyl intermediate as already mentioned in the first chapter (Ton-That et al., 2000). In the next step, a modified oligoglycine nucleophile attacks the thioacyl intermediate to yield the transpeptidation product (Figure 1) (Popp et al., 2007a). Depending on the function carried by the glycine derivative, it is possible to obtain fluorescent labeled proteins, circularized proteins or covalently linked proteins. A plethora of publications testifies the wide applicability of the sortagging reaction (Antos et al., 2009a;Antos et al., 2009b;Popp et al., 2009;Yamamoto and Nagamune, 2009). Furthermore, it is possible obtain proteins that are covalently attached to a solid support.

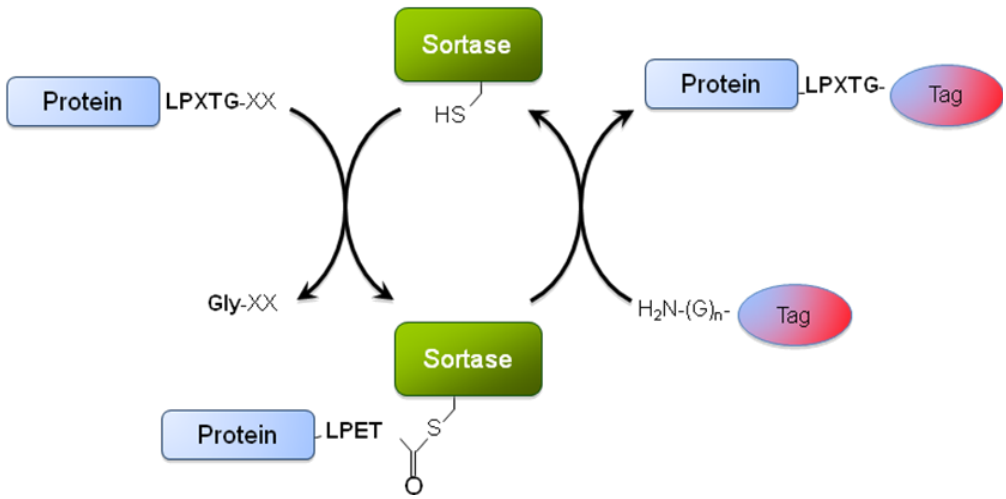


Figure 1: Site-specific labeling of target proteins by sortase-mediated transpeptidation. Sortase A recognizes substrates with an LPXTG motif, cleaving the peptide bond between the threonine and glycine and resulting in a thioacyl intermediate. Subsequently, a modified oligoglycine nucleophile attacks the thioacyl intermediate to yield the transpeptidation product with the probe linked to the target protein by an amide bond (Popp, Antos et al. 2009).

Two soluble forms of Sortase A enzyme have been used in the *in vitro* transpeptidation reaction. Native SrtA has a trans-membrane domain, allowing the enzyme to be anchored in the plasma membrane of the cell.

To favor the *in vitro* expression and purification of the protein the trans-membrane domain has been replaced by a hexahistidine tag. One version (Ton-That et al., 1999) contains an N-terminal deletion of 25 amino acids, replaced by a hexahistidine tag. A second version contains a more extensive deletion (59 amino acids), replaced by a hexahistidine that is followed by a thrombin cleavage site. These two versions have essentially identical activities (Ilangovan et al., 2001), but different mobility in SDS-PAGE gels, a useful attribute when the substrate protein migrates at a similar molecular weight as the sortase.

In this thesis, Sortagging is used as a tool to covalently link proteins of the *Influenza virus* genome onto a solid support; as a proof of principle, an influenza virus protein arrays was generated. The sortase reaction allows immobilizing proteins in an oriented fashion avoiding purification steps in an array building.

1.2 Influenza virus array

Influenza virus (Flu) is the cause of major respiratory illness in the human population, killing millions people in pandemic years. Hemaglutinin (HA) and Neuraminidase (NA) are the two surface proteins identified as the most prominent virulence factors. HA is present on the viral coat. It is synthesized as a precursor (HA0) that trimerizes in the endoplasmatic reticulum and is then transferred to the cell surface via the Golgi system. HA is cleaved by cellular proteases into HA1 and HA2 which convert the protein into a fusion-active form (Skehel and Wiley, 2000).

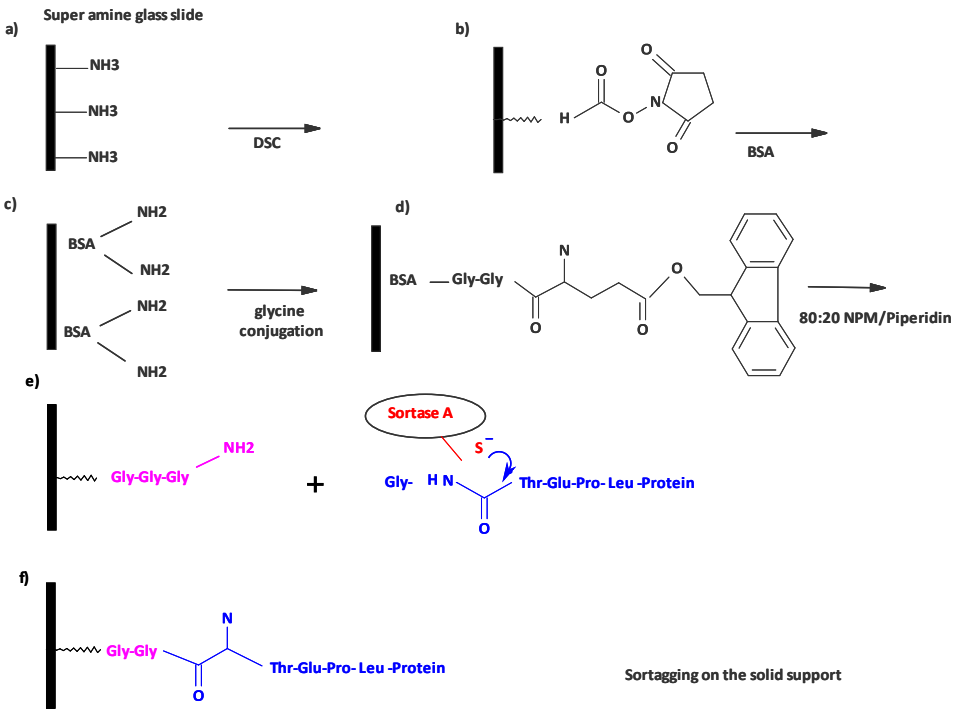
HA2 is the conserved portion of hemagglutinin. Early mapping studies of the immunogenic regions on HA showed that neutralizing antibodies are directed only against the receptor binding HA1 portion, while recent studies showed antibody recognition of a highly conserved influenza virus epitope (Ekiert et al., 2009). These antibodies have been shown to neutralize several subtypes among the virus clades which in principle could allow for a wide coverage. The antibodies that neutralize the stem region of HA (conserved region HA2) presumably have a role preventing the conformational change of protein at low pH, avoiding the membrane fusion of the virus with the host cell.

It has been hypothesized that making use of recombinant epitope antigens could provide the basis for a universal flu vaccine (Bommakanti et al., 2010).

In this perspective it would be desirable to develop tools for generating protein arrays to screen and identify new epitopes on the basis of biological interest. A useful antigen chip with Flu proteins and fragments of HA and NA antigens could be updated seasonally and used for seasonal screening. In this section of the thesis, the use of sortagging for Flu chip construction has been explored.

1.3 Protein array construction for Sortagging

In this thesis a simple process has been developed to pre-activate the solid support, a glass slide chemically derivatized with primary amine groups (Scheme 1). In a few fast steps, the glass slides were readied for the sortagging conjugation reaction. In the first step, BSA-NHS slides were fabricated and activated with the DSC (N,N' di-succimidyl carbonate) reaction in order to display the BSA amino groups (Macbeath and Schreiber, 2000). Conjugation of Gly_3 to the BSA-layer is possible only via its C-terminus since the Gly_3 N-terminus is blocked by the FMOC (9-fluorenylmethyloxycarbonyl) protective group. Activated groups react easily with surface amines on the FMOC-glycine peptide to form amide linkages. A deprotection reaction allows releasing the free nucleophilic glycine N-terminus allowing it to take part as a nucleophile in sortase-mediated reactions.

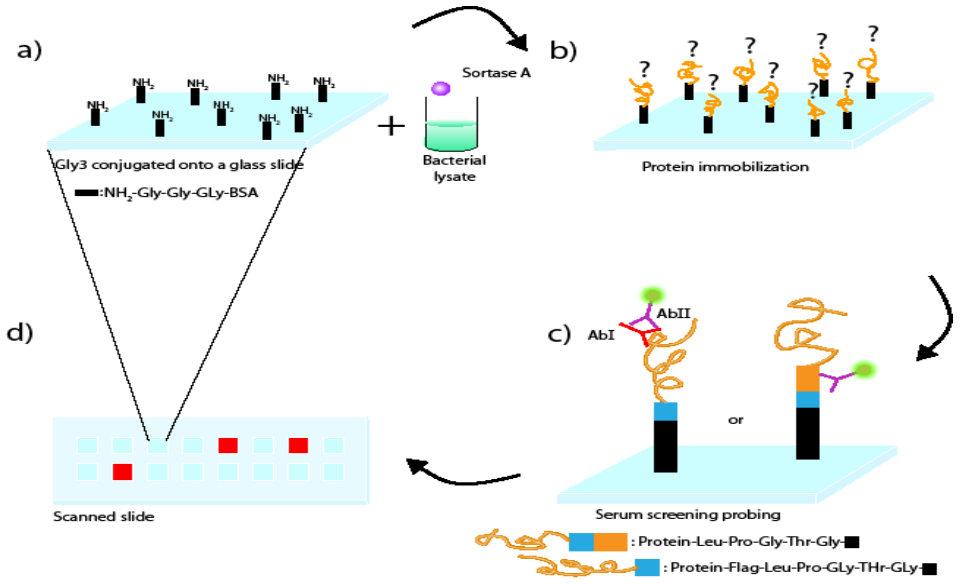


Scheme 1: Creation and use of chemically derivatized glass slides. a-b-c) Super amine slides were derivatized with BSA and conjugated with Fmoc-Gly_3 (d), followed by deprotection using piperidine (20% in NMP). e-f) the Sortase A-mediated conjugation reaction proceeds on the surface of the activated slide in the presence of LPETG-protein.

Scheme 2 illustrates here the critical steps to build the protein array. After the initial derivatization of the slide, Sortagging technology is used to immobilize proteins containing the sortagging motif at their C-terminus.

Briefly, super amine glass slides derivatized with BSA are conjugated with Gly_3 peptide; bacterial lysates in different wells containing protein with a C-terminal LPETG motif are incubated in presence of Sortase A and immobilized.
The unbound material is removed by extensive washing and the immobilized conjugates are ready to be detected using specific anti-sera. Finally, the slide is scanned to visualize proteins specifically recognized by the serum.

The process represents a flexible way to fabricate protein arrays in high throughput fashion, eliminating time-consuming and costly purification steps. Conceivably, a whole genome can be processed and immobilized in order to obtain protein profiles useful for many purposes.



Scheme 2: Sortagging immobilization steps using proteins expressed by a cDNA library: a) Super amine glass slides are derivatized with bovine serum albumine (BSA) using the N-N di-succimidyl carbonate reaction and the slide is conjugated with Gly-Gly-Gly peptide (Gly_3). b) Bacterial lysates expressing proteins containing the LPETG motif are incubated in separate wells in the presence of sortase A. In the process, proteins become sortagged to the slide c) Sortagged conjugates are detected using specific anti sera. d) The slide is scanned to visualize protein interactions.

2. Materials and methods

Protein microarray chip fabrication

Generation of an *Influenza virus* library and expression of proteins. PCR amplifications of each cDNA fragment from *Influenza A virus* (A/WSN/1933/H1N1) were performed; recombinant DNA fragments were inserted into a bacteria expression vector system. Expression of recombinant proteins and array construction completed the building process. PCR primers containing ±20 bp of gene specific regions and adapter sequences were used.

The adapter sequence, which becomes incorporated into the termini flanking the amplified gene, are homologous between linearized vector pET 28 (b+) (Novagen) and inserts, can anneal as desired and become viable plasmids upon transformation.

in chemically competent cells TOP 10 (Invitrogen). Primers used were shown in Table 1.

Sequence-confirmed plasmids were expressed in BL21 Codon Plus competent cells (Stratagene) and grown in auto induction media (15g/L glycerol, 30 g/L Yeast extract, 0,5 g/L $MgOSO_4$, 20 g/L K_2HPO_4, 5 g/L KH_2PO_4) for 12-18 h and harvested after over night (o.n.) growth.

SDS-PAGE and Western Blot analysis. SDS-PAGE analysis was performed using Nu-PAGE™ 4-12% Bis-Tris gradient gels (Invitrogen) according to the manufacturer's instructions. Hi-Mark™ protein standard (Invitrogen) served as protein marker. Gels were stained with Colloidal Coomassie Blue G-250 (Invitrogen) or processed for Western Blot analysis by using standard protocols. Polyclonal antibody specific for Influenza A by IHA, recognizing H1N1 and H3N2 and probably other Flu A strains (Millipore), and polyclonal antibody specific for PB1 and NS1 were used at a dilution of 1/1000. Secondary rabbit anti-goat IgG-HRP conjugated antibodies were used at a dilution of 1/10 000.

PROTEIN NAME	OLIGO SEQUENCE
1AS_For_HAchainA	GAAGGAGATATACATATG aaggcaaaactactggtc
2AS_Rev_HAchainA_Flag	AtcatctttataatcgccgccTCTGTATTGAATGGATGG
3AS_Rev_HAchainA/A3_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCCTCTGTATTGAATGGATGG
4AS_For_HAchainB	GAAGGAGATATACATATGggtctatttggagccattgct
5AS_Rev_HAchainB_flag	atcatctttataatcgccgccGATGCATATTCTGCACTGCAA
6AS_Rev_HAchainB/A5_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCCGATGCATATTCTGCACTGCAA
7AS_For_HA_A1	GAAGGAGATATACATATGaaggcaaaactactggtcctg
8AS_Rev_HA_A1_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCCTTTACATAGTTTCCCGTTGTG
9AS_For_HA_A2	GAAGGAGATATACATATGttaaaaggaatagcccca
10AS_Rev_HA_A2_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCC CCATGAACTTTCCTTGGGAAA
11AS_For_HA_A3	GAAGGAGATATACATATG cccaaccacacattcaacgga
12AS_For_HA_A4	GAAGGAGATATACATATGggtctatttggagccattgct
13AS_Rev_HA_A4_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCC ATTCTTCACATTTAAGTCATG
14AS_For_HA_A5	GAAGGAGATATACATATGctgtacgagaaagtaaaaagc
15AS_For_NA_C1	GAAGGAGATATACATATG aatccaaaccagaaaata
16AS_Rev_NA_C1_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCCCCTATAAGGGCTTCTGTCCTT
17AS_Rev_NA_C1_flag	atcatctttataatcgccgccCCTATAAGGGCTTCTGTCCTT
18AS_For_NA_C2	GAAGGAGATATACATATG gccttaatgagctgccctgtc
19AS_Rev_NA_C2_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCC GCCGGTATCAGGGTAACAGGA
20AS_Rev_NA_C2_flag	atcatctttataatcgccgccGCCGGTATCAGGGTAACAGGA
21AS_For_NA_C3	GAAGGAGATATACATATGaaagtgatgtgtgtgtgcaga
22AS_Rev_NA_C3_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCCACTGTCACTTTTAGTCCTTCC
23AS_For_NA_C4	GAAGGAGATATACATATGtccagacatgggtttgagatg
24AS_Rev_NA_C4_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCC CTTGTCAATGGTGAACGGCAA
25AS_For_M1/2	GAAGGAGATATACATATGagtcttctaaccgaggtcgaa
26AS_REV_M1_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCCCTTGAATCGTTGCATCTGCAC
27AS_REV_M2_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCCCTCCAGCTCTATGTTGAC
28AS_For_NP	GAAGGAGATATACATATGgcgaccaaaggcaccaaacga
29AS_REV_NP_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCCATTGTCGTACTCCTCTGC
30AS_For_PA	GAAGGAGATATACATATGgaagattttgtgcgacaa
31AS_Rev_PA_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCCTCTCAATGCATGTGTGAGGAA
32AS_For_PB1	GAAGGAGATATACATATGgatgtcaatccgactttactt
33AS_Rev_PB1_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCCTTTTTGCCGTCTGAGCTCTTC
34AS_For_PB2	GAAGGAGATATACATATGgaaagaataaaagaactaagg
35AS_Rev_PB2_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCCATTGATGGCCATCCGAATTCT
36AS_For_NS1/2	GAAGGAGATATACATATG gatccaaacactgtgtca
37AS_Rev_NS2_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCCAATAAGCTGAAACGAGAAAGT
38AS_Rev_NS1_linker	CGGTAGGCCGCCGCTGCCGCCGCCGCCAACTTCTGACCTAATTGTTCC

Table 1: Nucleotide sequence of the primers used for cloning of *Influenza virus* genes in the LPETG expression vector.

Chemically derivatized amine glass slides. Super amine slides were purchased from Array it (Arrayit, CA). BSA-NHS slides, displaying activated amino and carboxyl groups on the surface of an immobilized layer of BSA, were fabricated with 100 mM N,N-disuccinimidyl carbonate (DSC), 100 mM N, N' diisopropylethylamine (DIPEA), in a final volume of 400 ml anhydrous N, N' dimethylformamide (DMF). Slides were then immersed under agitation in this solution for 3h at room temperature. The slides were rinsed twice with 95% ethanol and then immersed in 400 ml of PBS pH 7.5 containing 1% BSA for 12 h at room temperature. The slides were rinsed twice with H_2O and twice with 95% ethanol. Slides were then dried by using a Microarray High-Speed Centrifuge (Arrayit, CA). The slides were immersed in Fmoc-Gly-Gly-Gly (Fmoc-Gly_3) peptide conjugation solution, containing 100 mM FMOC-Gly_3, 100 mM PyBOP, 100 mM HOBT, 300 mM DIPEA in NMP.

A multi wells cassette 2x8 (ArrayIt) was used as hybridization system; the slide was inserted in the cassette, a gasket was applied and then tightened. 100 µl of Gly_3 solution was applied on each well of the glass slide. Aldehyde groups were blocked with 100mM ethanolamine, 300 mM DIPEA in NMP for 1 to 24 h. Then the slides were washed 10 min in NMP and then rinsed in 20% piperidine NMP for 30 min. Following washing steps in NMP, ddH2O and sortase buffer (150 mM NaCl, 50 mM Tris pH 8.0, 1 mM $CaCl_2$). Finally, the slide was dried by centrifugation and stored.

Transpeptidation reaction conditions. The bacterial pellet was solubilized in lysis buffer (150 mM NaCl, 20 mM Tris-HCl pH 8.0, 1 mM $CaCl_2$, 0,5% NP40) and/or denaturing buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 20 mM imidazole, 0,5% Triton X-100 ,0,3% SDS, 10% glycerol) and then sonicated. For chemical lysis the Buffer B-PER (Pierce) was used in sortase buffer. SDS-page analysis gel was used for normalizing concentrations of expressed proteins. Total cell lysates of induced bacteria were loaded along with internal controls.

Sortase reaction on glass slides. To perform the sortase reaction on glass slides 100 µl of a solution containing sortase buffer (final concentrations: 150 mM NaCl, 20 mM Tris-HCl pH 8.0, 1 mM $CaCl_2$, 5% milk) were supplemented with 1.5-3 µM of various expressed protein in bacterial lysate and 50 µM sortase A. The mixture was applied to appropriate wells on the glass slides and the sortagging reaction was carried out o.n. at room temperature. Slides were washed three times in sortase buffer in the presence of detergents (150 mM NaCl, 20 mM Tris-HCl pH 8.0, 1 mM $CaCl_2$, 0.5% NP40) and finally the slides were dried by centrifugation.

Serum screning probing and scanner analysis. The slides were blocked with 5% milk , 2% BSA, 1% Tween in PBS at 4°C for 4 h or o.n., and washed in 0,1% tween PBS (PBST) for three times. The slide was incubated with a 1:150 dilution of serum from infected mice or with a commercial *Influenza A virus* H1N1 serum (Millipore) in blocking buffer for 1 h at 4°C. Slides were washed in PBST for three times, then a fluorescent secondary antibody Alexa-647 anti-mouse or goat (Invitrogen) in PBST was applied at a final concentration of 1 ug/ml and incubation was continued for 1 h at 4°C. Washing steps in PBTS and PBS were followed by a final step in ddH_2O. The slides were air dried after brief centrifugation and analyzed using a Genepix 4000B scanner (Molecular Devices). Intensities were quantified using Image J software. All signal intensities were corrected for spot-specific background. Protein concentrations in bacterial lysates in each well were normalized with respect to internal protein concentration standards. Covalent attachment to glass slides was followed for constructs containing a FLAG tag using Alexa-647 antibody and detected by scanning as described above.

Sortase A expression and purification: BL-21 cells were transformed by heat shock with the expression plasmid for sortase A and grown on LB plates containing ampicilin (100 µg/l) overnight. Single colonies were selected and a starter culture was grown overnight. The starter culture was used to inoculate two 1 l cultures of LB containing ampicillin. The bacteria were induced with 1 mM IPTG at OD_{600} ~0.5-0.8 for 4 h at 37 C. The bacteria were then spun down and frozen at -20 °C. The pellet was resuspended in 10% glycerol, 50 mM Tris-HCl pH 8.0, 150 NaCl, 20 mM imidazole and subjected to lysis in a French Press (two times). The lysis mixture was then centrifuged and the supernatant retained. The supernatant was passed over a Ni-NTA column which was subsequently washed with the above buffer (~50 ml). Sortase was eluted with 10% glycerol, 50 mM Tris-HCl pH 8.0, 150 NaCl, 300 mM imidazole. All fractions were analyzed by SDS-PAGE. The purified sortase was then dialyzed against 10% glycerol, 50 mM Tris-HCl pH 8.0, 150 NaCl. The protein concentration was determined by the Bradford assay. The protein solution was aliquoted (200 µl) and stored at -80 C.

3. Results

3.1 Design and construction of an *Influenza A virus* protein array using Sortagging.

A proteome array of *Influenza A virus* (WSN/1933/H1N1) was generated in order to demonstrate a new technology able to immobilize proteins in an oriented fashion to build protein array, furthermore the system was set up in order to study interactions of patient sera with the whole proteome of a cognate strain of *Influenza virus.*

To this end, a cDNA library of influenza virus was cloned in a dedicated expression vector and clones were expressed in frame-fused to a C-terminal LPETG motif followed by a His_6 tag. The small genome of influenza virus which contains eight single (non-paired) RNA strands encoding ten proteins (HA, NA, NP, M1, M2, NS1, NS2, PA, PB1, PB2) was used for protein array building because it seemed easy to manage and interesting for future studies.

The library cloned into the LPETG expression vector was successfully expressed in BL21-Codon Plus cells using auto induction medium (Figure 2). The library represents each of 10 cDNA of viral genome. In the case of Hemagglutinin (HA) two fragments were obtained, here called HA chain A (1 to 342 aa) and HA chain B (343 to 566 aa). For neuraminidase (NA) a fragment called NA/C2, ranging from aa 141 to 273, was cloned and expressed. Other fragments were generated for HA and NA, however, their expression was hardly detectable (Data not shown).

Following the reaction with sortase it seems useful to verify that the sortagging reaction indeed has immobilized the protein of interest before starting the sera screening. For This reason, the latter three constructs contain upstream of the LPETG motif a FLAG tag consisting of eight amino acids (N-DYKDDDDK-C). The small tag has little influence on the solubility of fused proteins and is easily detectable using a fluorescent monoclonal antibody (anti-FLAG Alexa 647) allowing even quantification of the protein on the support.

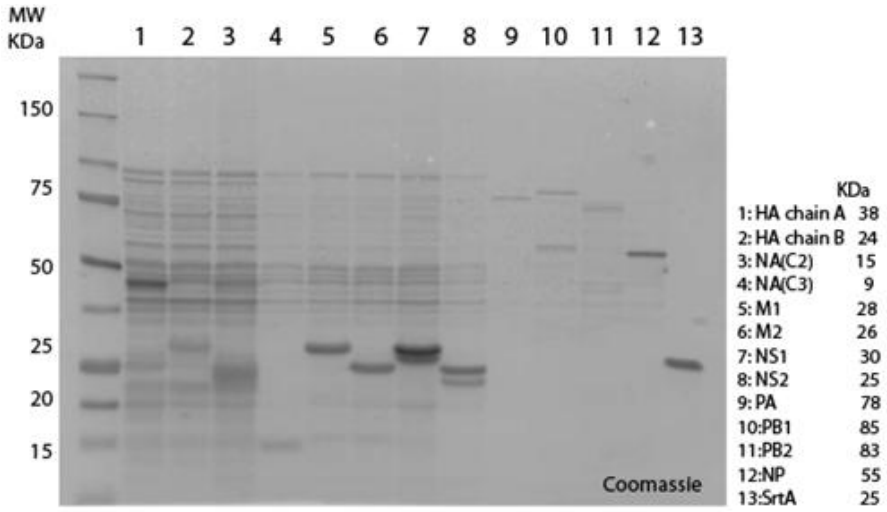


Figure 2: SDS-Page analysis of bacterial lysates containing *Influenza virus* proteins from the library stained with Coomassie Blue.

3.2 Chemical derivatization and visualization of coated Gly_3 glass slides.

Different kinds of chemically modified glass slides were initially tested to evaluate the optimal solid support to use during the Sortagging reaction. For this purpose aldehyde, epoxy and amine glass slides were compared (data not shown) and amine glass slides were chosen for the following analysis.

First, a study of optimal range concentration of nucleophile Gly_3 was conducted. Figure 3 A shows that fluorescence of immobilized mCherry-LPETGG protein enhances in parallel with higher concentration of Gly_3 concentration until saturation is reached. 100 mM of peptide was chosen as the optimal concentration to be used in the following experiments.

Then, in order to visualize random linkage of peptides on the solid support, a fluorescent molecule (TMR-Gly_3) was conjugated to Gly_3. Figure 3 B clearly demonstrates the presence of covalently linked peptides within the wells of the glass slide.

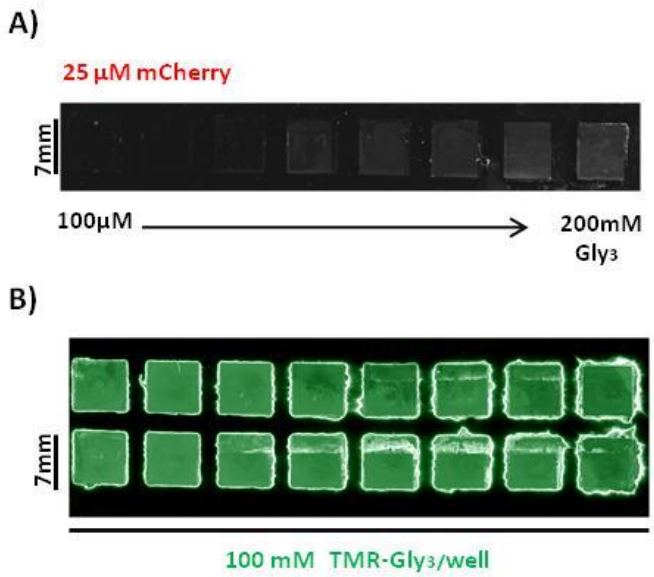


Figure 3: Laser scan analysis of derivatized slides. A) The concentration of tri-glycine peptide was varied between 100 µM and 200 mM during derivatization to find the optimal nucleophile concentration. Subsequently, 25 µM of fluorescent purified protein mCherry-LPETG was immobilized by sortagging using 150 µM of SrtA in sortase reaction buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 10 mM $CaCl_2$) on pre-activated glass slides and the amount of immobilized mCherry was determined by scanner analysis. B) The BSA-layer of a Super amine glass slide was randomly conjugated with the fluorescent peptide TMR- Gly_3. 100mM of peptide were conjugated on the slide and visualized by scanner analysis at 647 nm.

3.3 Sortase labeling in bacterial lysate of Flu protein under various conditions.

In previous studies, the Sortagging reaction was performed with high efficiency using purified starting materials, but evidence for the reaction to have taken place was also obtained by using crude mammalian cell- and bacterial lysates (Popp et al., 2007a).
To set up and verify sortase activity in order to label proteins in BL21 *E. coli* crude cell extracts, a construct from the Flu library was used, allowing verification of the reaction in solution before moving to the solid support. The recombinant enzyme was expressed and purified as previously described (Ton-That et al., 1999).

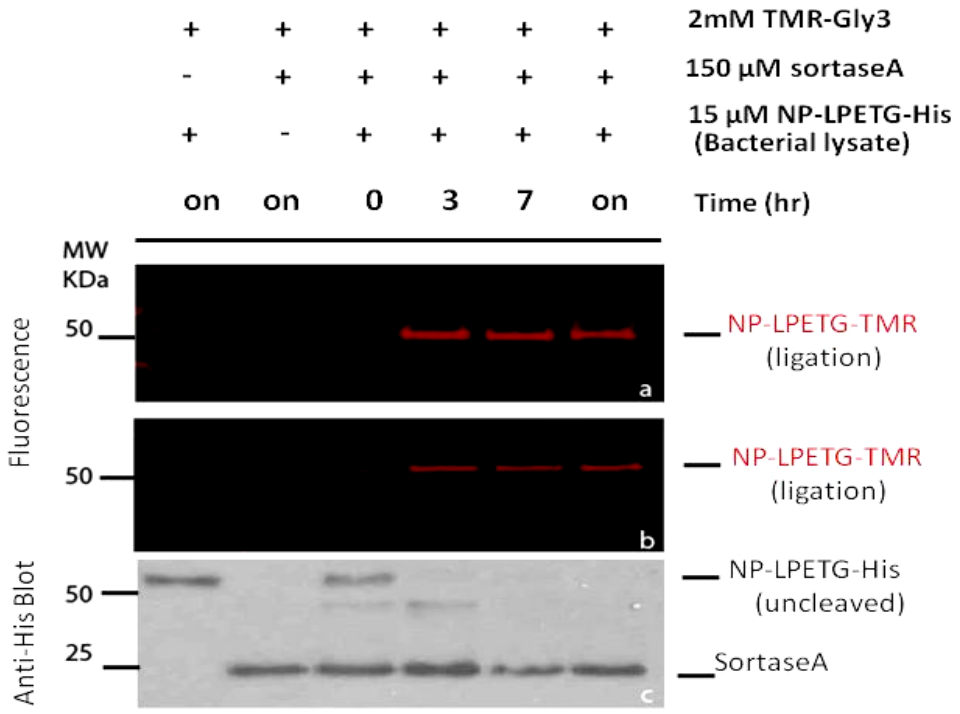


Figure 4: Labeling of NP protein in bacterial lysate with GGG-TMR nucleophile in normal and denaturing conditions. Bacterial lysate of *Influenza virus* protein NP containing a C-terminal LPETG motif was incubated with sortase A for different times in presence and absence of fluorescent TMR- Gly3 peptide. 15 µM of protein in the bacterial lysate was treated with 150 µM of SrtA in sortase reaction buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10 mM $CaCl_2$) in (a) NP-40 lysis buffer and in (b) SDS/Triton buffer (for 0, 4, 7 h and o.n. at 37°C) , (c) His_6/HRP immunoblot showing the presence of unlabeled protein in the absence of sortase A and at T_0, the band corresponding to 25 KDa demonstrates the presence of sortase A.

As shown in Figure 4, 15 µM of expressed LPETG tagged protein from a crude extract was lysed by sonication under mild (a) or harsh (b) conditions and labeled with 2 mM of the fluorescent TMR- Gly_3 peptide. NP-40 agent was used used to solubilize proteins in native condition, while SDS/Triton detergents were used as denaturing agent.

The Sortagging reaction becomes evident after 3 h and is virtually complete after 7 hours of incubation. A mild environment improves enzymatic activity, but the final transpeptidation product is still observed, even when the reaction was carried out in the presence of SDS (Figure 4).

Immuno-blotting using an antibody against His_6 shows the presence of the His_6 tag at the C-terminus of the input material when incubated in the absence of sortase A or before the reaction has taken place (T_0). SrtA contains an N-terminal His tag and is also recognized by the anti-his antibody (Figure 4 c).

In the next step, the optimal SrtA concentration was determined under different conditions. Figure 5 shows that the efficiency of the labeling reaction is not substantially affected by a low concentration of sortase, both under mild or harsh conditions, confirming that this parameter has little influence on the efficacy of the reaction.

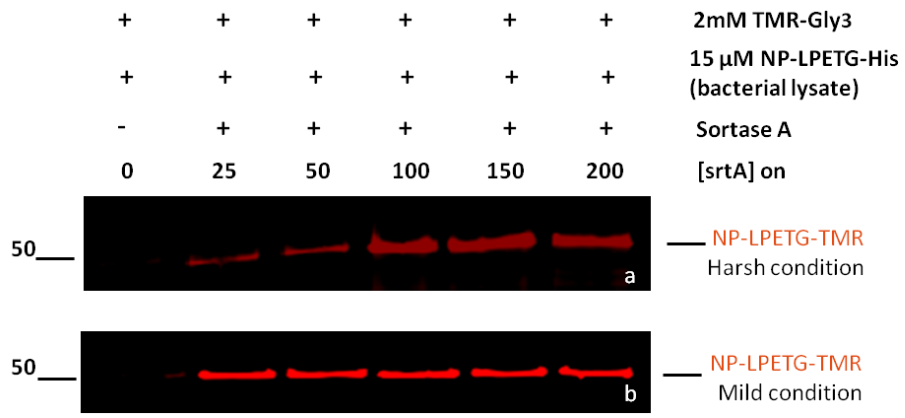


Figure 5: Sortase A working range of concentrations under different reaction conditions. 15 µM of LPETG-tagged protein in bacterial lysate was incubated with various concentrations of SrtA in a range between 0 to 200 µM and with 2 mM of the fluorescent TMR-Gly3 peptide under (a) mild or (b) under harsh conditions. Reactions were analyzed by SDS-PAGE visualizing fluorescence by scanning the gel in a Typhoon 8600 scanner (Amersham Pharmacia Biotech).

3.4 Normalization and validation of system.

As previously shown, the efficiency of the sortase reaction is influenced neither by the presence of unpurified starting material (Figure 4), nor by the presence of SDS (Figure 5). To further explore characteristics of the immobilization reaction the *Influenza virus* NA/C_2 fragment from the Influenza library was used.

Increasing concentrations of NA/C_2 protein in crude extract, ranging from 1.5 to 9 µM, were incubated in the presence and absence of sortase A on a slide, pre-activated with Gly_3.

The fluorescent FLAG antibody was used to reveal the specifically immobilized NA/C_2 fragment and the fluorescence intensity was used as measure for the amount of conjugated protein, to estimate reproducibility of the system, and to normalize the data.

As an internal standard, a range of concentrations of a purified and immobilized auto-fluorescent protein (mCherry) was used (Figure 6).

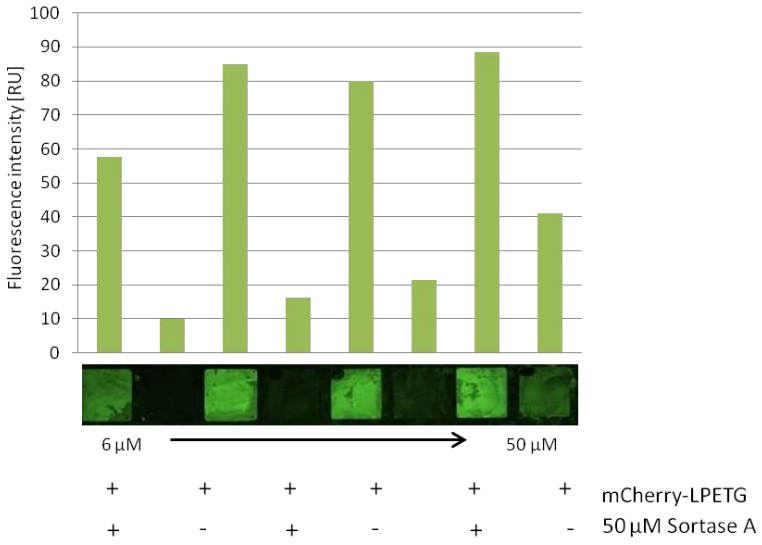


Figure 6: Effect of different concentrations of purified protein on Super amine glass slides. Increasing concentrations of mCherry-LPETG protein ranging from 6 µM to 50 µM were incubated in absence or presence of sortase A on a derivatized glass slide. mCherry is a fluorescent protein and is visualized easily by scanning analysis.

A crude extract of *E. coli* cells incubated on the array in the absence of sortase A was used as a measure of non-specific binding of the control protein to the slide. This resulted in a mean fluorescence intensity not higher than 35 FI (arbitrary units), a value far below the FI detected for the lower concentrations of the protein when incubated with SrtA.

A graphical representation of the fluorescence intensity attributable to immobilized NA/C_2 protein shows a linear range of sortase activity during the reaction on the solid support (Figure 7).

Increasing proteins concentration in the extracts increased the revealed fluorescent intensity until saturation is reached. As mentioned before, the FLAG tag was used to detect linkage of the protein to the glass slide.

A crude cell extract from *E. coli*, incubated on the array in absence of sortase A was used as a measure of non-specific binding of the antibody to the slide.

An anti-FLAG 647 antibody was incubated in the different wells and the effect of different concentrations of cellular lysate expressing LPETG-tagged protein was nicely verified by graphical analysis (Figure 7). Signal intensity increased as protein concentration increased, while the background is very low. After incubation with antibody the slide was scanned using Geneprix scan and analysed using ImageJ software, which allowed to arbitrary quantify the fluorescence in each well.

In previous studies anti-His tag antibodies were used to detect proteins on nitrocellulose glass slides generating a very high background. This shows that the system using the FLAG tag allows for a more efficient detection and a clearer signal.

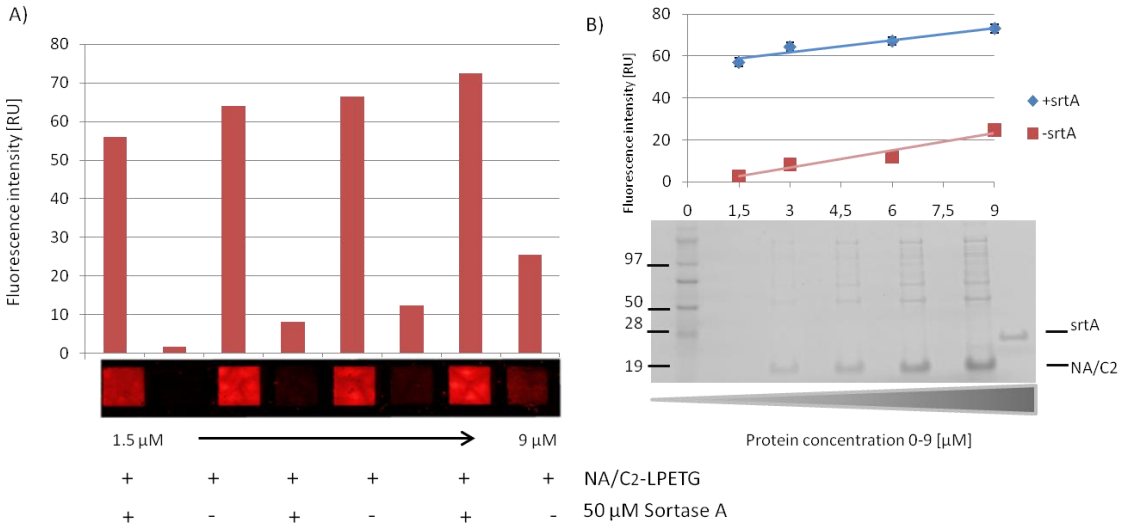


Figure 7: Effect of different concentrations of cellular lysate on Super amine glass slides. A) The lower part represents the detection of the immobilized protein in a concentration dependent manner. Increasing concentrations of Influenza virus protein in crude extract ranging from 1.5 µM to 9 µM were incubated in absence and presence of sortase A and detected by Anti-FLAG 647. Upper part: graphical representation of the fluorescence intensity showing the linear range of activity. Data were analyzed using Image J software. B) Linearity and sensivity of the protein array by experiments with lysate in triplicate (Upper part) with a SDS-Page analysis of the different cellular protein concentration used on the slide (Lower part).

3.5 Screening the *Influenza virus* library

As a first application of the *Influenza virus* protein array the interaction with a commercial anti-H1N1 serum (anti-influenza A virus, polyclonal antibody) has been investigated. Prior to performing binding studies on glass slides a western blot analysis was performed with the crude extracts. Surprisingly, although visible on Coomassie-stained gels, not all the proteins from the crude extracts were detectable (Figure 8, left panel). In fact, western blot analysis shows interactions between some expressed proteins and the α-H1N1 serum. Protein folding and missing glycosylation might influence the interactions with antibodies, so that some of the proteins are not recognized.

Therefore, another blot was performed for proteins not detected by anti-H1N1-serum. Anti-PB1 antibody shows the interaction with PB1 protein in bacterial lysate, interacting also with PA and PB2 (Figure 8, right panel).

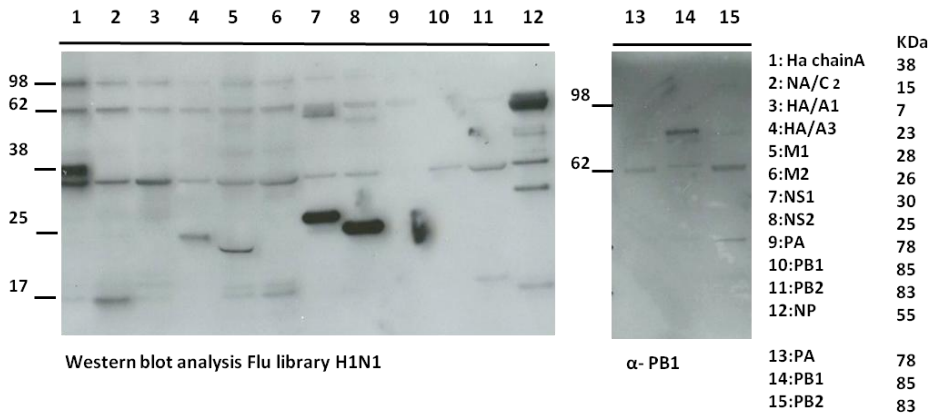


Figure 8: Western blot analysis of Flu library. SDS-Page analysis of bacterial lysates containing *Influenza virus* proteins from the library blotted against H1N1 antibodies (right panel). PA, PB1 and PB2 crude extracts blotted against anti-PB1 antibody (left panel).

To evaluate interactions between proteins of the Flu library, sortagged to the glass slide, and the same serum that was used in western blot analysis, all the crude cell extracts from the Flu library were applied for immobilization of recombinant proteins on the solid support. The total extract from *E. coli* cells expressing NA/C_2 FLAG-tag was printed on the array in presence and absence of sortase A verifying that immobilization of the proteins was sortase-dependent. All other reactions were carried out in presence of sortase A. Proteins were manually spotted at a concentration of 3 μM onto the Gly_3 surface to generate the protein microarray. The array was incubated in over night incubation with sortase A and then washed extensively to remove unbound material. Each well was probed with anti-H1N1/ anti-goat alexa647- or with alexa-647 anti-FLAG (1 μg/ml), washed and scanned with a Geneprix microarray scanner. The background signal generated by incubation with either anti-H1N1 or anti-FLAG 647 was subtracted from the initial fluorescence intensity of each protein on the slide.

As expected, only wells containing NA/C_2 immobilized in the presence of sortase show significant fluorescence intensity, while not in the absence of sortase (Fig 9). The other wells show interactions between proteins and serum however, not all the proteins interact. To demonstrate the presence of these proteins on the slide, specific antibodies were used, analogously to the western blot analysis initially described.

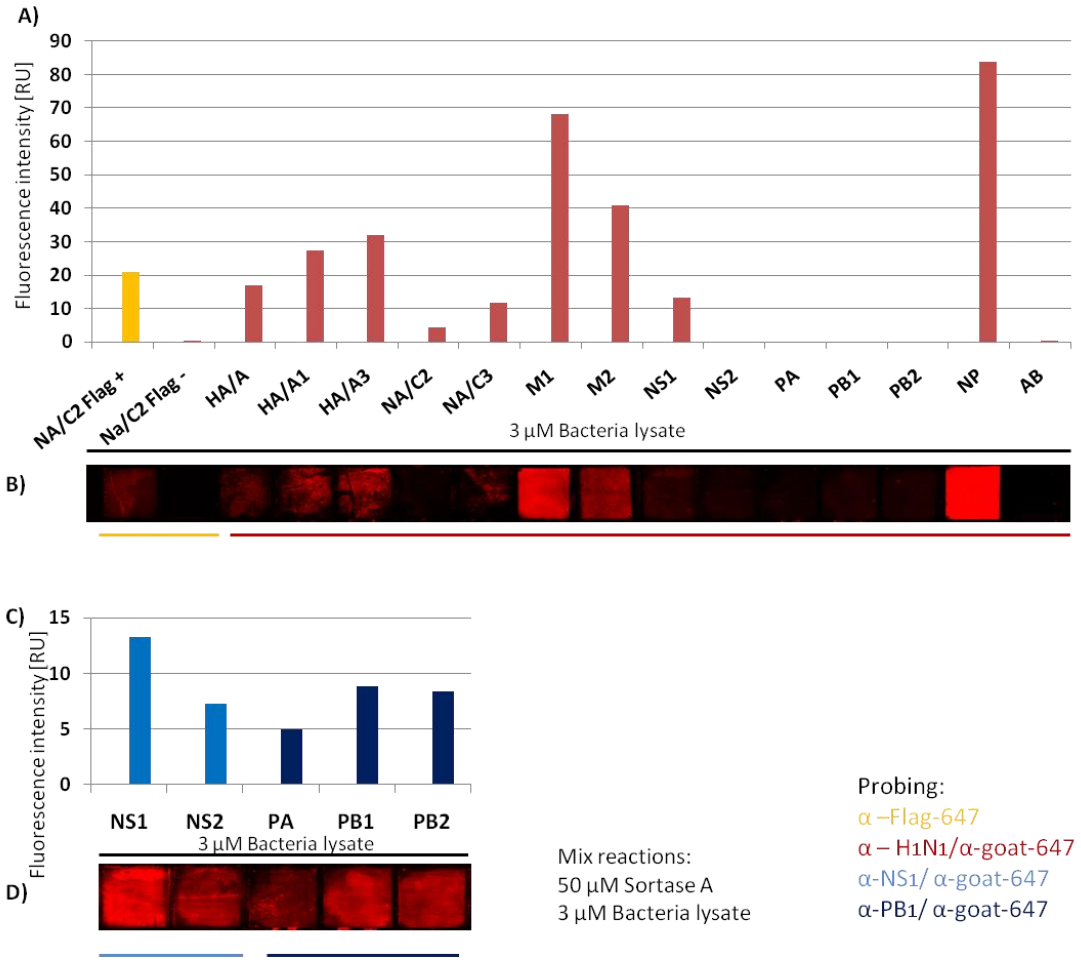


Figure 9: **Probing an immobilized library of *Influenza virus*.** A) 3 µM of Influenza virus proteins in crude extract were incubated in presence of sortase A and detected by H1N1 serum. Graphical representation of fluorescence intensity of constructed protein array shows some interaction with H1N1 serum. The total extract expressing NA/C_2 FLAG-tag was printed on the array in presence and absence of sortase A verifying that immobilization of the proteins was sortase-dependent. B) Visualization of the detected immobilized proteins after scanning. C-D) Graphical representation of fluorescence intensity of constructed protein array shows some interaction with specific antibodies against NS and PB proteins, verifying sortase-mediated ligation.

Figure 9 (C-D) shows these interactions, verifying the presence of recombinant protein on the glass slide. In future, FLAG tagged proteins will be used allowing easier detection of proteins present on the slide, like already done previously with the Na/C_2 fragment.

The protein immobilization method presented here has several advantages over traditional methods. First, proteins can be immobilized from crude extracts without time consuming and costly purification steps, second, high selectivity is reached since *E. coli* or *H. sapiens* do not have endogenous substrates for the sortase reaction.

The course of the immobilization reaction could be verified or even quantified by using the FLAG tag.

4. Discussion

Protein array technology is an emergent tool enabling high throughput screening of protein-protein or protein-lipid interactions as well as a tool to screen immuno-dominant antigens during the course of a bacterial or viral infection (Zhu and Snyder, 2003).

Protein-detecting arrays use different kinds of captured agents that need to be immobilized on a solid support to selectively get rid of complex mixture such as serum or cell lysate samples. Proteins can be immobilized on different kinds of solid supports applying chemical reactions involving aldehyde (Zhu et al., 2001), epoxide (Zhu et al., 2000), photo-crosslinking agents (Kanoh et al., 2005) or amine groups through non specific or covalent manner as discussed in this work.

In a previous study a proteome chip has been constructed covering almost 80% of the yeast proteome, allowing identification of new protein interactions and a common binding motif among the proteins screened (Zhu et al., 2001). This shows that in general a whole bacterial or viral or eukaryotic genome can be cut in a random (or even rational) way in order to generate fragments that can be cloned into a expressed vector and immobilized on solid support (Macbeath and Schreiber, 2000).

In this thesis a new technology to build protein arrays has been explored. As proof of concept, a protein array of the viral genome of *Influenza A virus* H1N1 was set up by the use of sortagging, a technology already used as a versatile method for labeling specific proteins and for protein conjugation (Popp et al., 2007b).

Thanks to a one-step enzymatic reaction mediated by sortase A, the ten proteins from the cDNA library of *Influenza A virus* were immobilized via their C-termini on a super amine glass slide and screened with a commercial serum against the same viral strain.

Chemical derivatization of the solid support is an easy preparation step, permitting site specific immobilization of recombinant proteins via their modified C-termini avoiding alterations of conformation and activity.

The sortase A enzyme has been used in the last years for many biotechnological purposes, its enzymatic activity could be improved by mutation; therefore, the form actually in use is adaptable to many different conditions.

Sortagging of proteins from crude cell extracts was demonstrated to be possible either in solution or to a solid support. Furthermore, the possibility of sortagging proteins in cellular extracts under mild or even under harsh conditions has been demonstrated. On the glass slides, only very small amounts of material (0.03-0.07 μg/μl) were necessary to reach a good signal to background ratio.

The presence of a specific tag, such as the FLAG tag, cloned upstream of the LPETG motif could allow to verify the sortase reaction, and to quantify directly the amount of protein on the chip by using specific fluorescent-labeled antibodies. Here, in one of the cloned proteins the FLAG tag was used to verify and to quantify covalent attachment of the protein to the slide surface.

The sortagging technique can be used to link purified proteins to the solid support in order to identify protein-protein interactions, however, the specificity of the Sortase A reaction avoids purification steps in array building and nevertheless allows immobilization of proteins in an oriented fashion. This technology has been applied already as a versatile method for labeling specific proteins and for protein conjugation. In the present study for the first time the tool is implemented to covalently link proteins of a viral genome onto a solid support. The process readily can be scaled up to larger proteomes in order to develop protein arrays for high throughput screening.

Finally, this methodology can be easily interfaced with cell-free protein expression systems, allowing for rapid access to the high-throughput production of protein chips.

References Cited

Antos,JM, G L Chew, C P Guimaraes, N C Yoder, G M Grotenbreg, M W L Popp, H L Ploegh. Site-Specific N- and C-Terminal Labeling of a Single Polypeptide Using Sortases of Different Specificity. Journal of the American Chemical Society 131[31], 10800-+. 2009a.
Ref Type: Magazine Article

Antos,JM, M W L Popp, R Ernst, G L Chew, E Spooner, H L Ploegh. A Straight Path to Circular Proteins. Journal of Biological Chemistry 284[23], 16028-16036. 2009b.
Ref Type: Magazine Article

Bommakanti, G. et al., 2010, Design of an HA2-based *Escherichia coli* expressed influenza immunogen that protects mice from pathogenic challenge: Proc.Natl.Acad.Sci.U.S.A, v. 107, no. 31, p. 13701-13706.

Camarero, J. A., Y. Kwon, and M. A. Coleman, 2004, Chemoselective attachment of biologically active proteins to surfaces by expressed protein ligation and its application for "protein chip" fabrication: J.Am.Chem.Soc., v. 126, no. 45, p. 14730-14731.

Ekiert, D. C., G. Bhabha, M. A. Elsliger, R. H. Friesen, M. Jongeneelen, M. Throsby, J. Goudsmit, and I. A. Wilson, 2009, Antibody recognition of a highly conserved influenza virus epitope: Science, v. 324, no. 5924, p. 246-251.

Ilangovan, U., H. Ton-That, J. Iwahara, O. Schneewind, and R. T. Clubb, 2001, Structure of sortase, the transpeptidase that anchors proteins to the cell wall of Staphylococcus aureus: Proc.Natl.Acad.Sci.U.S.A, v. 98, no. 11, p. 6056-6061.

Kanoh, N., K. Honda, S. Simizu, M. Muroi, and H. Osada, 2005, Photo-cross-linked small-molecule affinity matrix for facilitating forward and reverse chemical genetics: Angew.Chem.Int.Ed Engl., v. 44, no. 23, p. 3559-3562.

Kindermann, M., N. George, N. Johnsson, and K. Johnsson, 2003, Covalent and selective immobilization of fusion proteins: J.Am.Chem.Soc., v. 125, no. 26, p. 7810-7811.

Macbeath, G., 2002, Protein microarrays and proteomics: Nat.Genet., v. 32 Suppl, p. 526-532.

Macbeath, G., and S. L. Schreiber, 2000, Printing proteins as microarrays for high-throughput function determination: Science, v. 289, no. 5485, p. 1760-1763.

Popp, M. W., J. M. Antos, G. M. Grotenbreg, E. Spooner, and H. L. Ploegh, 2007a, Sortagging: a versatile method for protein labeling: Nat.Chem.Biol., v. 3, no. 11, p. 707-708.

Popp,MW, J M Antos, G M Grotenbreg, E Spooner, H L Ploegh. Sortagging: a versatile method for protein labeling. Nature Chemical Biology 3[11], 707-708. 2007b.
Ref Type: Magazine Article

Popp, M. W., J. M. Antos, and H. L. Ploegh, 2009, Site-specific protein labeling via sortase-mediated transpeptidation: Curr.Protoc.Protein Sci., v. Chapter 15, p. Unit.

Proft,T. Sortase-mediated protein ligation: an emerging biotechnology tool for protein modification and immobilisation [Review]. Biotechnology Letters 32[1], 1-10. 2010.
Ref Type: Magazine Article

Skehel, J. J., and D. C. Wiley, 2000, Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin: Annu.Rev.Biochem., v. 69, p. 531-569.

Ton-That,H, G Liu, S K Mazmanian, K F Faull, O Schneewind. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. Proceedings of the National Academy of Sciences of the United States of America 96[22], 12424-12429. 1999.
Ref Type: Magazine Article

Ton-That, H., S. K. Mazmanian, K. F. Faull, and O. Schneewind, 2000, Anchoring of surface proteins to the cell wall of Staphylococcus aureus. Sortase catalyzed in vitro transpeptidation reaction using LPXTG peptide and NH₂-GLY₃ substrates: Journal of Biological Chemistry, v. 275, no. 13, p. 9876-9881.

Yamamoto, T., and T. Nagamune, 2009, Expansion of the sortase-mediated labeling method for site-specific N-terminal labeling of cell surface proteins on living cells: Chemical Communications, no. 9, p. 1022-1024.

Zhu, H. et al., 2001, Global analysis of protein activities using proteome chips: Science, v. 293, no. 5537, p. 2101-2105.

Zhu, H. et al., 2000, Analysis of yeast protein kinases using protein chips: Nat.Genet., v. 26, no. 3, p. 283-289.

Zhu, H., and M. Snyder, 2003, Protein chip technology: Curr.Opin.Chem.Biol., v. 7, no. 1, p. 55-63.