Karolinska Institutet

Institutionen för medicin, Solna

Structural and biophysical studies of pneumococcal capsular surface proteins

AKADEMISK AVHANDLING
som för avläggande av medicine licentiatexamen vid Karolinska Institutet offentligen försvaras i seminarierummet/lunchrummet, plan 2, SciLifeLab byggnad alfa, Tomtebodavägen 23A, Karolinska Institutet, Solna

Onsdagen den 12 november, 2014, kl 09.00

av

Cecilia Mikaelsson

Huvudhandledare:
Docent Adnane Achour
Karolinska Institutet
Institutionen för medicin, Solna
Science for Life Laboratory

Bihandledare:
Professor Per-Åke Nygren
Kungliga Tekniska Högskolan
Skolan för Bioteknologi

Professor Birgitta Henriques-Normark
Karolinska Institutet
Institutionen för mikrobiologi, tumör- och cellbiologi

Betygsnämnd:
Docent Mikael Karlsson
Karolinska Institutet
Institutionen för mikrobiologi, tumör- och cellbiologi

Professor Daniel Daley
Stockholms Universitet
Institutionen för biokemi och biofysik

Docent Robert Schnell
Karolinska Institutet
Institutionen för medicinsk biokemi och biofysik

Stockholm 2014

From the Department of Medicine, Solna
Karolinska Institutet, Stockholm, Sweden

Structural and biophysical studies of pneumococcal capsular surface proteins

Cecilia Mikaelsson

Karolinska Institutet

Stockholm 2014

The cover figure shows the crystal structure of the $BR_{187\text{-}378}$ dimer (PDB ID: 4V13)

Published by Karolinska Institutet.
Printed by Åtta45

ISBN 978-91-7549-694-8

"No one can make you feel inferior without your consent".

–Eleanor Roosevelt

POPULÄRVETENSKAPLIG SAMMANFATTNING

Streptococcus pneumoniae (pneumokocker) är en av de bakterier som orsakar lung-inflammation och kan även leda till bakteriell blodförgiftning och hjärnhinneinflammation. Tack vare den tilltagande användningen av antibiotika samt vaccinernas intåg har sjukdomarna föranledda av denna bakterie kunnat stävjas. Men på senare tid har dödsfallen ökat i antal och tidigare relativt lättbehandlade infektioner orsakade av bakterien har blivit en större utmaning. En av anledningarna till att pneumokocker återigen börjat bli en allvarlig hälsorisk är att antibiotikaresistensen ökat avsevärt. För att kunna utveckla nya vacciner och läkemedel behövs måltavlor för dessa, i detta fall något som kallas för *virulensfaktorer*. Bakterien använder sig av virulensfaktorer för att kunna infektera oss människor eller för att skydda sig mot vårt immunförsvar. De bidrar till bakteriens *virulens*, vilket är dess förmåga att orsaka sjukdom. Dessa virulensfaktorer är i många fall proteiner som sitter på ytan av bakterien.

En av grundstenarna i att kunna utveckla vaccin och läkemedel mot helt nya måltavlor, är att ta reda på så mycket som vi kan om dem, hur de fungerar och hur de ser ut. För att kunna ta reda på detta använder vi oss av en process som kallas strukturbestämning. I detta använder vi olika tekniker för att skapa en väldigt detaljerad 3D-modell av t.ex. ett protein. Detta måste göras eftersom proteiner är för små att se med ett mikroskåp.

Målet med forskningen som ingår i den här avhandlingen är att strukturbestämma och karakterisera två proteiner som är viktiga för pneumokockers virulens, PsrP och RrgC. PsrP sitter på ytan av baketerien och fungerar som ett lim, i det att den hjälper bakterien att fästa till ytor, som t. ex. vår lungvävnad. Endast en liten del av detta protein är funktionellt, och denna del har vi bestämt strukturen på. Det visar sig att denna del binder till ett protein som heter Keratin-10. Många känner säkert till keratin, hornämne, som en viktig komponent i naglar och framförallt hår. Det som var intressant med att finna att PsrP binder till Keratin-10, är att det i vissa fall finns på ytan av lungceller. Det är ännu oklart om dessa fall inkluderas av infektion av pneumokocker, men om detta visar sig vara fallet betyder det ett stort genombrott för bekämpning av bakterien. RrgC är ett protein som ingår som en del av den struktur på bakterien som kallas pilus, som är ett vidhäftningsorgan. Inte mycket har varit känt om detta protein, men det har nu strukturbestämts och dessa studier kommer förhoppningsvis att leda till att nya läkemedel samt vacciner kan utvecklas för att återfå kontrollen över denna bakterie.

ABSTRACT

Streptococcus pneumoniae (pneumococcus) is a major human pathogen and a leading cause of morbidity and mortality worldwide, especially in children and the elderly. The casualties due to respiratory infections are estimated to be over 4 million per year, where pneumococcus is the predominant species. Moreover, the increasing number of antibiotic-resistant strains and the suboptimal clinical efficacy of available vaccines confines control of this pathogen. In view of this situation, substantial attention has focused on novel virulence-related pneumococcal proteins as potential targets for future drug targets. One such target is the pneumococcal serine-rich repeat protein (PsrP), an important virulence factor present in a majority of the strains capable of causing invasive pneumococcal disease (IPD). The functional binding region (BR) of this protein binds to keratin-10 (KRT10) and also promotes biofilm formation through self-oligomerization. The crystal structure of the KRT10-binding region of PsrP ($BR_{187-385}$) reveals an extended β -sheet on one side of a compressed two-sided β-barrel presents a basic groove that could accommodate the acidic helical rod domain of KRT10. Well-ordered loop regions distort the other side of the barrel and form a paperclip-like sub-structure for more specific interaction with KRT10. *In vitro* alanine substitution of residues localized within this paperclip structure efficiently disrupted $BR_{187-385}$/KRT10 complex formation.

Within the work of this thesis we also found that $BR_{187-385}$, which lacks the putative oligomerization region, forms stable oligomers *in vitro*. Small angle X-ray scattering and circular dichroism experiments revealed a non-globular and possibly disordered structure of the N-terminal region. A comparative analysis of the long ($BR_{120-395}$) and short ($BR_{187-385}$) domain constructs even suggested an inhibitory role for the N-terminal $BR_{122-166}$ domain in oligomerization. Indeed, we could show that the N-terminal region is released by cleavage through the human furin protease that specifically recognizes a sequence localized between the globular $BR_{187-385}$ domain and the disordered N-terminal part. The crystal structure of the dimer of the KRT10-binding domain of PsrP reveals a domain swap mechanism for dimerization, this process, although energetically costly, is probable when PsrP is involved in biofilm formation.

The minor ancillary pilus protein RrgC is believed to anchor the pneumococcal pilus to the cell wall, and until recently very little was known about its structure. In this work we have evaluated the structure of the protein in solution with SAXS, and shown that the protein is a multidomain protein with flexible linkers and adopts extended conformations in solution.

Through this work we have proposed a more specific interaction between PsrP and KRT10 than previously reported, as well as shown that oligomerization is possible despite the removal of the N-terminal region. These findings are very important for a deeper understanding of the details in PsrP's role in pneumococcal invasion. The structural studies of RrgC provide a platform for future studies. The results presented within this thesis will hopefully aid in the future development of novel drugs and vaccines.

LIST OF SCIENTIFIC PAPERS

I. **The basic keratin 10-binding domain of the virulence-associated pneumococcal serine-rich protein PsrP adopts a novel MSCRAMM fold**
Schulte T, Löfling J, **Mikaelsson C**, Kikhney A, Hentrich K, Diamante A, Ebel C, Normark S, Svergun D, Henriques-Normark B, Achour A.
Open Biol. 2014 Jan; **4:**130090. Doi: 10. 1098/rsob. 130090

II. **The BR domain of PsrP forms a domain-swapped dimer and is proteolytically cleaved by a furin protease; implications for pneumococcal biofilm formation?**
Schulte T[1], **Mikaelsson C**[1], Kikhney A, Löfling J, Ebel C, Svergun D, Normark S, Henriques-Normark B, Nygren P-Å , Achour A
Manuscript
[1]These authors contributed equally to this work.

CONTENTS

Reprints of papers and manuscripts

LIST OF ABBREVIATIONS

ASU	Asymmetric unit
AUC	Analytical ultracentrifugation
BR	Basic region
CBP	Choline binding protein
CD	Circular dichroism
CPS	Capsular polysaccharide
CWSS	Cell wall sorting signal
ELISA	Enzyme-linked immunosorbent assay
EM	Electron microscopy
EOM	Ensemble optimization method
EPS	Extracellular polymeric substances
IF	Intermediary filament
IDP	Intrinsically disordered protein
IPD	Invasive pneumococcal disease
LPM	Linear peptide motif
MR	Molecular replacement
MSCRAMM	Microbial surface components recognizing adhesive matrix molecule
MW	Molecular weight
NMR	Nuclear magnetic resonance
NPP	Non-bacteremic pneumococcal disease
KRTs	Keratins
PAFr	Platelet-activating factor receptor
pI	Isoelectric point
PsrP	Pneumococcal serine-rich repeat protein
r.m.s.d.	Root mean squared deviation
SAXS	Small angle X-ray scattering
SEC	Size exclusion chromatography
SRRP	Serine-rich repeat protein
WLC	Worm-like chain

1 INTRODUCTION

1.1 *Streptococcus pneumoniae*: "THE CAPTAIN OF ALL THE MEN OF DEATH"

The statement in the title above was Sir William Osler's description of *Streptococcus pneumoniae* (pneumococcus), in 1918, and it is as fitting today as it once was (1, 2). The human pathogen *S. pneumoniae* is an inhabitant of the upper respiratory tract in about 10-20% of healthy adults and 40-60% of children (2, 3). Pneumococcus, although normally commensal, can cause sinusitis and otitis media as well as more sinister conditions such as pneumonia, bacteremia and meningitis and is a major cause of morbidity and mortality worldwide (4, 5). Unlike other commensals, such as *Streptococcus pyogenes* and *Staphylococcus aureus,* pneumococcus produces a rather limited repertoire of tissue-destructive proteins and toxins, making it an ideal candidate for the study of factors contributing to pathogenicity (3).

Historically, the study of *S. pneumoniae* has contributed greatly to the fields of molecular biology and epidemiology. Louis Pasteur and George M. Sternberg, independently of one another, first described this organism in 1881, leading to it being established as a major cause of human lobar pneumonia. Being a pervasive pathogen, pneumococcus was one of the first bacteria observed during the development of the Gram staining technique, which was key in being able to distinguish pneumococcal pneumonia from other causes. In the beginning of the 20th century, Fred Neufeld demonstrated agglutination as well as capsular swelling when adding specific antisera to pneumococcal suspensions, thus providing a method for serotyping of pneumococcal isolates. The discovery by Neufeld, through Fred Griffith, led Oswald Avery and colleagues to demonstrate that the natural transformation of pneumococcus occurs with the DNA as genetic material. This was shown by transformation of a noncapsulated, avirulent strain of *S. pneumoniae* with the DNA from a capsulated, virulent strain (1, 3, 6). Thence, the studies of this bacterium has led to the sprouting of a whole body of knowledge leading to important discoveries and state-of-the-art techniques used by laboratories across the world.

S. pneumoniae is a Gram-positive bacterium, most commonly seen as diplococcus, and is able to divide without separation of the cell walls, resulting in a continuous growth with the formation of a string of cells (7). The *S. pneumoniae* strains are categorized into subgroups, serotype strains, depending on the composition of their polysaccharide capsule. The surface capsular polysaccharide, a key pathogenicity factor, is composed of repeating units of sugars (8). In the early 1920s Michael Heidelberger and colleagues demonstrated that this complex carbohydrate, capsular polysaccharide (CPS), was antigenic by vaccination of mice. This novel finding changed the field as it up until that point was generally believed that only proteins were able to elicit an immune response (1, 6, 8). There are 93 serotypes of *S. pneumonia* that have been described so far and whose genomes can have up to 10% sequence variation (1, 5, 9). In order to investigate the genetic basis of pneumococcal virulence, two

serotype strains of *S. pneumoniae* were chosen to have their genomes fully sequenced, virulent serotype 4 (TIGR4) and the avirulent R6 (6, 10). The genomes of these strains were compared in order to identify genes associated with virulence. The main difference between these strains is the fact that TIGR4 is capsulated whilst R6 is not, and the capsule has since long been acknowledged as a major virulence factor of pneumococcus (1, 3, 6, 11). Analysis of the genome of TIGR4 is widely used to identify components involved in virulence, such as surface proteins. Examples are pneumococcal surface protein (pspA), the autolysin (lytA) as well as choline binding proteins (CBPs), adhesins, such as pneumococcal serine-rich repeat protein (PsrP) and the pneumococcal pilus. The importance in characterizing novel virulence factors is the potential use they may have as novel drug and vaccine targets (11).

1.2 PNEUMOCOCCAL EPIDEMIOLOGY

Pneumococcus colonizes the human nasopharynx during the first months of life, and is generally believed to spread via aerosols. The nasopharynx acts as a reservoir for pneumococci and is the source of pneumococcal spread between human hosts. After the bacteria enter the nasal cavity and adhere to nasopharyngeal epithelial cells, they can colonize the nasal cavity, rendering the host a carrier. This colonization is usually cleared quite rapidly resulting in serotype-specific immunity, however when this clearance fails it can cause upper respiratory tract infections ore even more severe disease by spreading to other organs resulting in pneumonia, bacteremia, or even meningitis (9, 12). Occurrence of invasive pneumococcal disease varies significantly based on age, immune status, socioeconomic status and geography, among other factors. The serotypes associated with invasive disease vary between geographical areas. This is very important to consider, since the vaccines available are serotype-specific (2, 9).

From studies of animal models it is evident that the nasopharynx carriage is the first step in pneumococcal pathogenesis. These studies demonstrates the development of otitis media or invasive disease following nasal inoculation of pneumococci (4). Carriage is associated with colonization of one serotype at a time, although it is also common with multiple serotype carriage. The natural competence displayed by *S. pneumoniae* lead to the possibility of a switch of capsular serotype through the uptake of extracellular DNA. This means that other techniques than serotyping are required to distinguish clinical isolates, referred to as clonal types, from one another. These molecular techniques are used to analyze the genetic relationships between clonal types. The most common colonization occurs with an “either/or” host-pathogen interaction, in that it is either a clonal type with a invasive pneumococcal disease phenotype, or one with a non-invasive phenotype that allows for long-term carriage, although it must be noted that it is very rare with serotypes causing carriage alone (2, 9).

Current means of controlling this pathogen are vaccines directed towards the polysaccharide capsule and β-lactam based antibiotics. The resistance rates of *S. pneumoniae* to conventional antimicrobiotics have increased, which means that failures in treatment of patients with pneumococcal infections occur more frequently (13). This emphasizes the need for the

development of novel drugs and vaccines against pneumococcus. As previously mentioned, the pneumococcal CPS is highly immunogenic and induces type-specific protective immunity, and the vaccines are based on the polysaccharidic capsule, which differs between serotypes (14, 15). The first polyvalent CPS vaccines were licensed in the 1970s and the current 23-valent one is used in the United States and Europe. This vaccine covers the most common disease causing serotype strains in these parts of the world, however CPS vaccines fail to elicit a proper immune response in young children (< 2 years of age), since CPSs are T-cell-independent. Many bacterial pathogens carry surface antigens that are T-cell-independent, which means that they are able to stimulate antibody production without help from T-cells, and polysaccharide antigens is one of them (16, 17). Pneumococcal CPS-protein conjugate (PCV) vaccines were developed to overcome this shortcoming. The coupling of CPS to protein transporters renders them T-cell-dependent antigens (5, 14, 18-20). The serotypes included in the conjugate vaccines were chosen because their dominance in invasive diseases in the United States, thus they are not adapted to geographic areas displaying different serotype distributions. At the moment there are 7-, 10- and 13-valent vaccines in use.

The reason for the lower number of serotypes included so far in the conjugate vaccines are high costs due to technical reasons. Since the year 2000, when the 7-valent vaccine was included in the childhood vaccination program in the United States, there was an enormous decrease in pneumococcal invasive disease (vaccine-serotype-related) in children. Herd immunity in the adult population has also reduced carriage of vaccine-serotypes. However, it has been recently reported that serotypes reduced by the vaccines available are being replaced by non-vaccine pneumococcal serotypes (serotype replacement). The increase in these replacement strains has brought new concern combined with the observed clonal expansion of non-vaccine clones, as capsular switching may eradicate the efficacy in existing vaccines. Discussions are ongoing as to whether or not it is cost-effective to vaccinate adults with the conjugate vaccine because the 23-valent polysaccharide vaccines effectiveness towards non-bacteremic pneumococcal pneumonia (NPP) appears to be poor. Nonetheless, it is suggested conjugate vaccines might be effective against NPP and invasive pneumococcal disease (IPD) in adults. This is still a contentious issue because, as of yet, the potential effectiveness is unknown, and routine vaccination of children will most likely result in herd immunity, reducing the potential benefits from adult vaccination (9, 14, 19-22).

The standard treatment of infections caused by pneumococci has been the β-lactam penicillin for a very long time. The general use of antibiotics to combat these infections is in the long-term unrealistic due to the raise in antibiotic resistances through genetic plasticity including natural competence the sole use of antibiotics will become less and less effective. Some pneumococcal clonal types carrying resistance are more inclined to spread globally than others. One example is the clonal type ST156 (resistance towards penicillin, trimethoprim sulfamethoxazole) that originated in Spain and has now spread to other countries (9, 23). The widespread resistance towards most antibiotics observed continues to be a cause for concern. The spread of antibiotic-resistant strains, together with the observed serotype-replacement

and capsular type switching emphasizes the need to develop improved vaccines and drugs.

A biofilm is a bacterial formation where a community of microorganisms adhered to a surface are embedded in a matrix consisting of extracellular material. The microorganisms themselves mostly produce the components of this matrix, which can constitute up to and over 90% of the biofilm. The biopolymers that make up the scaffold of the matrix are referred to as extracellular polymeric substances (EPS), which allow for surface adhesion and biofilm cohesion (24).

Up until quite recently, no information about pneumococcal biofilms was available, however, it is now evident that pneumococci can form biofilms, particularly during colonization, and in the middle ear (otitis media). More than 60 % of human infections by *S. pneumoniae* are related to the formation of pneumococcal biofilms, and these bacterial communities have been proven to have an increased tolerance towards the host immune system as well as antibiotics. It has been shown that changes in gene-expression and production of specific virulence factors are more abundant during biofilm-formation. These include pneumolysins, CBPs and PsrP (18, 25).

1.3 PNEUMOCOCCAL VIRULENCE FACTORS: SURFACE MOLECULES

The ever-growing knowledge of the genomes of both invasive and non-invasive strains of *S. pneuomoniae* has enabled detailed comparative analyses. The differences in genomes between strains are located within sites of diversity scattered around the bacterial chromosome, which mainly distinguishes invasive from non-invasive strains. Many of these regions carry the genetic signature of pathogenic islands (2, 3). It's been established that *S. pneumoniae*, as a species, show a high plasticity and variability in its genome, most likely due to recombinational events, both within and between strains, as well as transformation events (26).

In order for tissue invasion by pneumococcus to be successful, a battery of virulence factors needs to be expressed in a coordinated fashion. The most important virulence determinants of pneumococcus includes the capsular polysaccharides (adherence and antiphagocytic properties), adhesins, invasion genes, heavy-metal transporters, oxidative stress protectors, host-defence evasion molecules and biofilm formatting compounds, among others (2). The initial adherence of pneumococci to nasopharyngeal cells is crucial, however, the mechanisms are not fully understood. Although not exclusively involved in adhesion, the surface molecules of pneumococci are undoubtedly required in adhesion and colonization (27, 28).

Considering the increased prevalence of strains resistant to currently available treatments the situation with antibiotic resistance and serotype switching progresses, which means it is absolutely necessary for the development of new therapies. In order to make this challenging task more feasible we need a deeper understanding of the interaction between this pathogen and the host tissues. These interactions mainly involve extracellular virulence factors

expressed by the bacterium, to interact with host tissues. Such virulence factors include the sugar-based (capsule, teichoic and lipoteichoic acids) and the protein based (29).

1.3.1 Pneumococcal serine-rich repeat protein (PsrP)

PsrP is a recently identified host surface and intraspecies bacterial adhesion, a pneumococcal virulence factor that is encoded within the RD10 pathogenicity island and present in 60 % of strains capable of causing pneumonia in children (30, 31). PsrP is an extremely large surface protein, which belongs to the serine rich repeat protein (SRRP) family, and as the name suggests consists largely of repeats of the amino acid serine. SRRPs are glycosylated and thus runs larger on SDS-PAGE than their expected molecular weight (MW), and this is also the case with PsrP (32). The domain structure of SRRPs typically comprises an N-terminal signal peptide, a shorter serine-rich repeat region (SSR2), a non-repeating region (NR) that is unique for different SRRPs followed by a very long serine-rich repeat region (SSR1) and a cell wall anchor (CWAD). The extremely long SSR2 domain is believed to extend the functional NR domain away from the bacterium to mediate adhesion (33, 34). The domain organization of PsrP is outlined in figure 1.

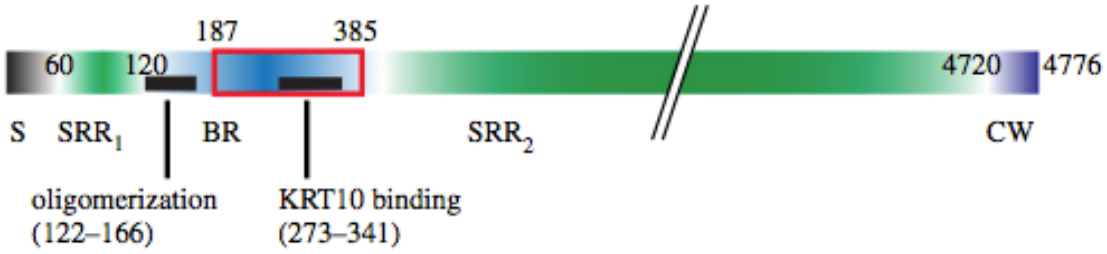

Figure 1. *Domain organization of PsrP, the NR domain of PsrP is called the basic region (BR) (35)*

With its 4776 amino acid residues, PsrP is the largest known bacterial protein. The functional, unique domain, NR, of PsrP is named the basic region (BR), simply because of its high content of basic amino acids, and a predicted isoelectric point of 9.9 (32). PsrP is able to mediate the adherence to lung (bronchial and alveoliar) cells but not to nasopharyngeal or vascular endothelial cells. Studies have shown that *S. pneumoniae* mutants that are deficient in *psrP* have greatly reduced virulence (30, 36). Pneumococci that do not express PsrP are incapable of establishing a lung infection in mouse models (32, 37).

Quite recently, the host ligand of PsrP has been identified as KRT10, which has been shown to be overexpressed on lung cells in certain cases. Immunization with the BR domain of PsrP protects mice against pneumococcal infection, which means that the BR domain, or derivatives there of, might be suitable as a component in a protein vaccine to protect against pneumococcal disease. The location of the KRT10 binding region is indicated in figure 1 (32). PsrP has also been proposed to be of importance in pneumococcal aggregation and biofilm formation. *In vivo* studies have shown that *psrP* deficient mutants were unable to

form pneumococcal aggregates, and the oligomerization region was suggested to be situated in the N-terminal region of BR (figure 1) (25).

1.3.2 The pneumococcal pilus

The bacterial pilus is an extended polymer structure coupled to the cell wall. Pili are long hair-like extensions that commonly have adhesins attached to the tip, to maximize contacts with the surroundings. The Gram-positive pilus differs from the Gram-negative pilus in that the subunits making up the structure are covalently attached to one another, and require sortase enzymes for proper assembly. The Gram-negative pili, on the other hand, require specific chaperons for proper assembly (38-40).

The subunits of the Gram-positive pilus are secreted through the cell membrane, after which, polymerization takes place. The polymerization is catalyzed by sortases (Srt), a family of membrane-associated transpeptidases, which covalently couple proteins containing a C-terminal cell-wall sorting signal (CWSS), to the peptidoglycan cell wall, or to each other. The housekeeping sortase, SrtA, anchors the surface molecules, containing the CWSS followed by a hydrophobic segment and a positively charged tail, and is present in nearly all Gram-positive bacteria. The sortases cleaves the precursor-protein through a recognition peptide, the LP*X*TG (or LP*X*TG-like, where *X* is any amino acid) motif between T and G, after which the protein is coupled to its target. The pili expressed in different bacteria are always co-expressed with the family of sortases that assemble them, (class C sortases), and are components of the pilus islet in the genome (38, 41-43).

1.3.2.1 The pneumococcal pilus is composed of RrgA, RrgB and RrgC

S. pneumoniae strain T4 (TIGR4) expresses a pilus, encoded by the *rlrA* pilus islet 1. The pneumococcal pilus consists of three different component proteins. RrgA, RrgB and RrgC. These are covalently coupled by the pilus-specific sortases in pneumococcus (SrtC-1, SrtC-2 and SrtC-3) and anchored to the peptidoglycan cell wall by the housekeeping SrtA. The recognition peptides (LP*X*TG-like motif) for RrgA, RrgB and RrgC are YPRTG, IPQTG and VPDTG, respectively (43, 44). Covalently linked repeating units of the major component, RrgB, make up the stalk of the pilus. The structure of the first three domains was determined by X-ray crystallography in 2010 (45), and, subsequently fully determined in 2011 (46-48). RrgB consists of four domains (D1-D4) (Figure 2); each domain is stabilized by one intramolecular isopeptide bond, present in all subunits of the pilus. An isopeptide bond is an amide bond that can be formed autocatalytically between the amino acids lysine and asparagine, which is typically stabilized by aspartic acid or glutamic acid. This bond can form within proteins and prior to its discovery in Gram-positive pilin subunits this had never been observed (49). RrgA acts as an adhesin that can bind to lung epithelial cells *in vitro*, which is aided by its location at the tip of the pilus, though it was shown to also assemble along the shaft (41, 50-53). However, more recently converse reports show that RrgA is only localized at the tip of the pilus (54). The crystal structure of RrgA reveals it to be composed of four domains (D1-D4, Figure 2), where D2 and D4 are stabilized by one intra-domain isopeptide

bond each. The crystal structure also unveiled an integrin collagen-recognition domain, confirming its ability to recognize and bind extracellular matrix elements (55).

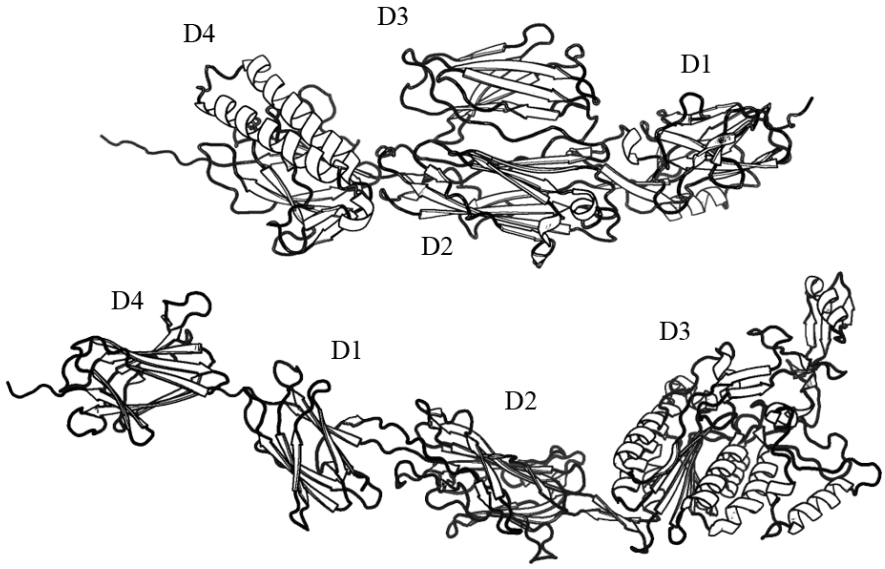

Figure 2. *Crystal structure of RrgB (upper panel), with domains D1-D4, PDB ID: 2Y1V (47). The bottom panel show the crystal structure of RrgA, with domains D1-D4, PDB ID: 2WW8 (55). Figures generated using PyMOL (56)*

The last, and also the smallest, building block of the pneumococcal pilus is the minor ancillary protein RrgC. The information about the role of RrgC in the virulence of the pilus is limited, however, it is now generally accepted that the protein is anchoring the pilus to the peptidoglycan cell wall of *S. pneumoniae*. There is continuing interest into which of the three pilus-related sortases are coupling RrgC to the rest of the pilus structure, as none of the sortases seem to identify the recognition peptide (VPDTG) specifically (43, 44). Therefor, there is a great incentive to structurally characterize this protein to shed insights into its functions, as well as the mechanism of its coupling to the cell wall. Structural information was lacking for RrgC when this work was initiated. However, recently the structure of a modified RrgC has been determined (35) revealing a three-domain structure (D1-D3) (Figure 3) with intra-domain stabilizing isopeptide bonds in D2 and D3. The study suggests that RrgC is mostly dependent on the housekeeping sortase, SrtA, for covalent linking to the peptidoglycan cell wall.

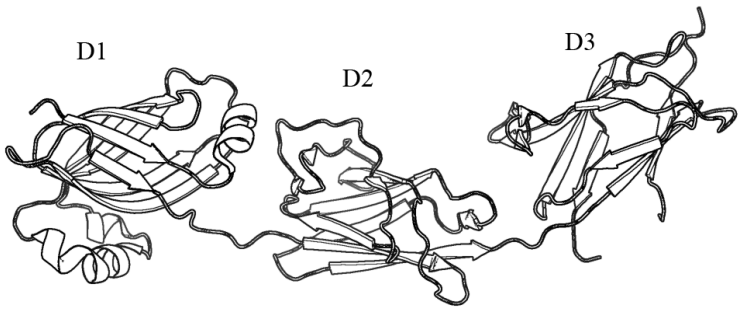

Figure 3. *Crystal structure of RrgC, PDB ID: 4OQ1 (35). Figure generated using PyMOL (56)*

1.4 HOST RECOGNITION AND IMMUNOLOGY

Adherence of *S. pneumoniae* to the alveolar epithelium following proliferation and initiation of host damage responses leads to the development of lobar pneumonia. During the early stages of host response, which is an interaction between secreted products and cell wall anchored surface molecules of the bacterium and the alveolar epithelium together with the innate immune system; the bacterium adheres firmly to the tissues through several cell surface anchored adhesins, including PsrP and pili. If the bacterium subsequently succeeds in "innate inversion" which is the process whereby a pathogenic infection advances from mucosal disease to bacteremia the colonization results in a life-threatening condition for the host. *S. pneumoniae* express phosphorylcholine, which can bind to the human platelet-activating factor receptor (PAFr), as a mean to cross from lung into blood. Mice studies show reduced pneumococcal growth and alleviated survival rates when this receptor is deleted (2, 9).

One of the key components of the innate immune system is the pattern-recognition receptors, which have evolved to recognize conserved motifs expressed by pathogens, referred to as pathogen-associated molecular patterns (PAMPs). One of these pattern-recognition receptors is the C-reactive protein, acute-phase protein, is essential during pneumococcal bacteremia as it blocks PAFr (2, 9, 57). Indeed, many components of the immune system are active in pneumococcal infections, and toll-like receptors (TLRs) are key components as pattern-recognition receptors in initiation of the cellular innate immune response upon invasion of pathogens. This is mainly because they are able to detect a broad range of microbial pathogens at e.g. the cell surface. TLRs mediate the recognition of PAMPs and structural studies indicate that TLRs recognizes several PAMPs as their ligands (57). TLR2 is most commonly thought of as very important in the response to Gram-positive pathogens, and the PAMPs recognized by TLR2 are cell-wall components (peptidoglycan, lipoteichoic acids and lipoproteins). Pneumolysin is one of the primary toxins in the etiology of inflammatory responses and tissue damage in pneumococcal pneumonia, and the receptor involved in the proinflammatory effects of pneumolysin is TLR4 (2, 58).

1.4.1 Keratins: Keratin 10

Keratins (KRTs) are cytoskeletal intermediary filament (IF) proteins expressed in epithelial cells. There are, in the mammalian genome, over 50 keratin genes identified. Biochemical characteristics, such as molecular weight and isoelectric point (pI), determine the classification of the KRTs as type I (acidic) or type II (neutral-basic). The type I and type II KRTs are almost exclusively co-expressed in the tissues for which the KRTs are characteristic and form coiled coil α-helical heteropolymers (59-62). KRTs are important in several aspects of the cell, such as size, migration, differentiation and proliferation. More effort is being put into typing of KRTs and the role they play throughout the development of carcinomas, in this they can be utilized as tools in histopathological diagnosis. There is continuing research into understanding the impact of KRTs in the development of tumor formation and behavior (63).

All KRTs exhibit a tripartite structure consisting of a long α-helical rod domain and non- α-helical terminal domains, one N-terminal and one C-terminal. The highly conserved secondary structure of the rod domain takes the form of a two-stranded coiled-coil, in KRTs a heterodimer of type I and type II. This coiled-coil structure is divided into four segments called 1A, 1B, 2A, 2B (Figure 4) and are segmented by short linking segments (L1, L12, L2), that in turn forms tetramers, octamers (protofibril) and finally keratin filaments (64).

Previous studies have shown that KRTs are readily available on the surface of epithelial cells, rendering them potential docking sites for bacterial adhesins. Specifically, it has been shown that KRT10 is expressed on the surface of desquamated human nasal epithelial cells and epidermal keratinocytes. The adhesin clumping factor B (ClfB) of *Staphylococcus aureus* has been shown to interact with KRT10, and possibly KRT8, extracted from these types of cells (65, 66), while the SRR-1 protein from *Streptococcus agalactie* interacts with KRT4 on the surface of Hep2-cells, derived from human laryngeal carcinoma (67)

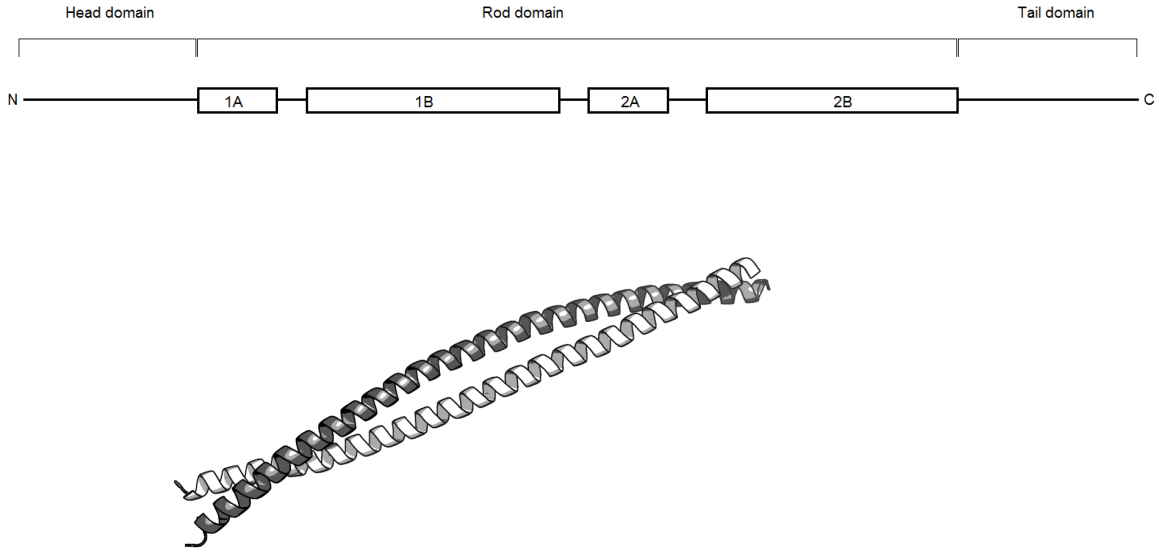

Figure 4. *Upper panel: Schematic domain organization of keratins. Bottom panel: Example structure of a keratin heterodimer, shown here is the crystal structure (PDB ID: 3TNU) of the 2B helical domain of coiled coil heterodimer of KRT5 (type II) and KRT14 (type I) (68)*

1.5 SMALL ANGLE X-RAY SCATTERING

With the structural information regarding macromolecules becoming more extensive, there is an increasing demand on structural determination methods to be able to decipher large complexes as well as highly disordered proteins. The challenges we are facing will most likely not be readily solved by one single biophysical technique. These challenges can be faced by combining high-resolution techniques, such as X-ray crystallography and nucleic magnetic resonance (NMR) with methods providing lower-resolution information e.g. electron microscopy (EM) and small angle scattering (SAS), particularly small angle X-ray scattering (SAXS).

SAXS provides the possibility of sampling huge complexes, multiple conformation samples,

flexibility, assembly, disassembly and aggregation etc. The solution technique is a rapid method to gain three-dimensional structural information, as the samples mostly are easily prepared. On the contrary X-ray crystallography needs well-diffracting crystals, which can be difficult to obtain, and has to prepared in such a way that the proteins are subjected to non-physiological conditions (i.e. precipitants and denaturants). This may alter the conformation and other characteristics of the protein. NMR can be unsuitable when proteins and protein complexes reach a certain size, though there are of course exceptions. Therefore, it is evident that a large fraction of proteins cannot be analyzed using these two methods alone. The advantage SAXS allows for is sampling of proteins in near physiological conditions, and the capture of real-time structural conformations resulting from slight changes in buffer conditions, addition of co-factors etc. The drawback is the loss of detailed information on the atomic level, due to the spherical averaging in the scattering profile (69, 70).

X-ray crystallography and SAXS are principally similar techniques, and most of the hardware can be shared between them. The incident X-ray beam is scattered (SAXS) or diffracted (X-ray crystallography) by the sample and the intensities are measured by an X-ray detector. The fundamental difference is the organization of the target molecules during the collection of data. Solution scattering is isotropic where the signal from all of the orientations of the molecule is relative to one another and to the beam is sampled and averaged together. In X-ray crystallography, the signal is obtained from highly ordered molecules stabilized within the crystal lattice. The information obtained from this method is significantly more detailed than of that acquired with SAXS, as it allows for atomic resolution of the resulting model. Due to the need of the molecules to be in a specific conformation that aids crystallization of the protein, it is believed that for certain multi-domain or inherently flexible proteins, the non-physiological conditions used in the crystallization buffer and the restraints put on proteins when forced into a rigid state during crystal packing can lead to biologically irrelevant conformations (69, 70).

1.5.1 SAXS data collection and processing

The scattering signal in SAXS measurements stem from the difference in average electron density between the molecules of interest and the bulk solvent ($\Delta\rho(\boldsymbol{r}) = \rho_{protein} - \rho_{solvent}$, where ρ is the scattering density). The scattering profile of the buffer must be subtracted from that of the protein, since the difference is quite small. The measurements are made by exposing a protein sample solution to X-rays, and the scattered intensity (I) is recorded. The resulting scattering curve is the solvent subtracted intensity as a function of the momentum transfer (s, also denoted q) and the scattering angle ($s = 4\pi \sin\theta/\lambda$, 2θ is the scattering angle, λ the wavelength of the incident beam) (Eq. 1), where $\rho(\boldsymbol{r})$ is the electron density, $\mathbf{r}$ is the vector describing the spatial position and D_{max} is the maximum dimension of the protein (71). The scattering profile is dependent on the shape and size of the protein, which means that different proteins will have a unique scattering signature (72, 73). The intensity is isotropic because of the randomly orientation and distribution of particles in solution. For a comprehensive review of solution scattering see (69, 74).

$$I(s) = 4\pi \int_0^{D_{max}} \langle \rho(\vec{\boldsymbol{r}}) * \rho(-\vec{\boldsymbol{r}}) \rangle \frac{\sin(sr)}{sr} r^2 dr \quad (1)$$

$$p(r) = r^2 \langle \rho(\boldsymbol{r}) * \rho(-\vec{\boldsymbol{r}}) \rangle \quad (2)$$

The Patterson function is the spatially averaged auto-correlation function of the electron density. Multiplied by r^2 gives the pair distance distribution function, $p(r)$ (Eq. 2). From the $p(r)$ we get direct information about the distances between electrons in the scattering molecules that are a distance r from each other. The resulting function is displayed as a histogram over these distances. The maximum dimension (maximum intraparticle distance), D_{max}, as well as the radius of gyration (R_g) can be determined from this plot. It can also be used to deduce overall shape (globular, elongated) of a protein (75). The Patterson function in X-ray crystallography is related to the *P(r)*, but unlike the *P(r)* the crystallographic Patterson function is derived from molecules that are confined in their rotations. This means that the three-dimensional information about the inter-atomic vectors is maintained. The Patterson function can be used in crystallography to locate the heavy-atoms in the unit cell, which aids in the solving of the phase problem (69, 76).

The radius of gyration (R_g) of a protein is defined as the square root of the average squared distance of the collection of electrons from their common particle center of mass (69, 77). R_g can be calculated from the scattering curves generated from the radial scattering intensities in different ways. The most common is using the Guinier approximation (78) which is measured by linear approximation of the slope of the Guinier plot ($\ln(I(s))$ vs. s) at low values for s. The range of the data used to calculate this is $sR_g<1.3$ for globular proteins and $sR_g<0.8$ for elongated proteins. If the protein is elongated it might be more suitable to calculate the R_g via the Debye approximation, as the Guinier approximation is valid over a narrower range for these type proteins. Another important parameter that can be extracted from the lowest s values is the zero intensity ($I(0)$), which is the intensity at zero angle, s=0. As this is coincident with the direct beam, this value needs to be extrapolated to be determined. $I(0)$ is the square of the number of electrons in the particle that is scattering, on an absolute scale, and can therefore be used to determine the molecular weight of the protein (69).

The Kratky plot can be directly calculated from the scattering curve and plotted as: $s^2I(s)$ vs. s. This can be used in evaluation of the folding in protein samples. Folded domain proteins give rise to the shape of a parabola (roughly), and the peak could provide some information regarding the size of the protein. However, the Kratky, as is R_g, is shape-dependent and can consequently not give absolute information considering molecular weight. When analyzing Kratky plots calculated for e.g. random coil peptides, they follow the Porod-Kratky worm-like chain (WLC) model (79), it is evident that they are lacking the peak characteristic for folded proteins. In the large s region they are linear with respect to s. It is important to note that for more extended proteins and proteins that comprises smaller folded regions with long flexible linkers, the Kratky analysis needs to be more extensive, and in some cases with respect to both the folded regions and the regions that can be described by the WLC model. For these type proteins it is sometimes necessary to calculate the radius of gyration of the cross section (R_c), as well as the length (L) and treat the entire protein according to the WLC

model. This is common when analyzing unfolded or intrinsically disordered proteins (IDPs) using SAXS (69, 80).

1.5.2 Data analysis software

There are many different analysis software and programs to evaluate, analyze and model SAXS data. The availability and user compatibility varies, and here only software from the ATSAS (© ATSAS team 1991-2014) suite, developed by Svergun and colleagues, will be described, as it is readily available and is predominately what has been used in the work presented in this thesis.

Initial data evaluation is usually done through automated analysis pipelines at beam lines, to assess the quality of the data, and if re-measurements are required. This includes calculation of R_g and $I(0)$ from the Guinier plot, and is mainly to determine whether or not the protein sample shows aggregation or interparticle interference. It should be noted that the software used is developed for globular proteins mainly, so it is unwise to blindly use the analysis performed by the automated pipeline, as the protein might not be completely globular. It is always advisable to perform quality checks of the analyzed data.

PRIMUS (81) is mainly used to handle the experimental 1-D SAS data files. Features include averaging, subtraction, scaling, merging, Guinier approximation, calculation of $I(0)$ and R_g (globular, flat and rod-type particles), Porod's volume and the Kratky plot.

GNOM (71) is used to process SAS data through the regularization technique. It is an indirect transform program that requires one-dimensional scattering curves as input files and generates the $p(r)$ (monodisperse systems), which in turn can be used to evaluate D_{max}, R_g and $I(0)$.

CRYSOL (82) can be used to evaluate the SAS scattering data when there is a known atomic structure available. This can also be used reversibly, to evaluate whether the crystal structure has the same conformation in solution as in the crystal packing, or if the crystallization process has forced the protein into a biologically irrelevant conformation. This program fits the atomic structure to the experimental scattering curves.

Solution structure modeling can be performed with several different programs, depending on what protein or system one wants to analyze. When the individual subunits of a protein are known, but not how they are positioned relative one another, rigid body modeling can be used to get a model of the overall assembly of the complex. SASREF (83) and CORAL (84) are two rigid body modeling programs, where SASREF models a complex formed by subunits, where the part of the atomic structure is known, against experimental scattering data. CORAL, based on SASREF, performs modeling of multidomain protein complexes against multiple experimental scattering data sets.

Ensemble optimization method (EOM) is a program suite used to describe the SAXS data of flexible proteins in solution. The program utilizes an ensemble representation of atomic

models that allows for flexibility by allowing different conformations of the protein to coexist (84, 85).

2 AIMS OF THIS THESIS

The casualties due to pneumococcus, although difficult to estimate, are believed to be in the millions every year worldwide. The efficacy of current vaccines is tapering owing to serotype switching, and antibiotic resistant strains are increasing in numbers, which creates a dire need for novel drug targets such as virulence-related pneumococcal proteins. A primary step towards the identification of new drug targets is structural and biochemical characterization of key virulence-factors. In this thesis the structural and partly biochemical basis of adhesion to KRT10 of the functional basic region (BR) domain of PsrP has been addressed. Additionally, the structural and biochemical study of the self-oligomerization of the BR domain has been evaluated, and this has been put in context with the very real possibility that PsrP is indeed important in pneuomococcal invasiveness.

Another protein that is believed to be of great importance in the virulence of pneumococcus is the Gram-positive pilus component RrgC. At the start of this research, this protein had not yet been structurally determined, as the last of the pilus components in the pilus of pneumococcus. The aim was to determine the three-dimensional structure of RrgC, preferably with X-ray crystallography, and to assess the function through small-angle X-ray scattering. This protein proved difficult to crystallize, which can be explained by the SAXS data collected and analyzed, as well as from recently published data regarding the structure (35). I present in this thesis the structural basis of this protein in solution.

The aims of the thesis was to investigate important virulence factors by structural studies, using X-ray crystallography and small angle X-ray scattering, as well as biochemical studies.

3 RESULTS AND DISCUSSION

3.1 STRUCTURAL STUDIES OF THE BASIC REGION (BR) OF PSRP AND ITS BINDING TO KERATIN 10 (PAPER I)

The pneumococcal serine-rich repeat protein (PsrP) is one of the key virulence factors naturally occurring in a majority of strains capable of causing IDP. The functional BR binds to KRT10 and is suggested to play a pivotal role in biofilm formation through self-oligomerization. In this study we present the crystal structure of $BR_{187-385}$, the KRT10 binding portion of BR (**paper I**; figure 1a), determined to 2.0 Å resolution. Structural homology analysis reveals a novel variant of the DEv-IgG fold. This fold is typical in MSCRAMMs (Microbial surface components recognizing adhesive matrix molecule), the adhesive molecules commonly found in Gram-positive bacteria. The structure, a compressed two-sided β-barrel, presents a basic groove on the one side of the barrel that is likely to interact with the acidic rod domain of KRT10. The other side of the barrel, formed by a loop region interspersed with short, stabilizing β-strands, contains the specific binding site for interaction with the tail-rod domain of KRT10.

3.1.1 The KRT10 binding region domain of PsrP adopts a MSCRAMM-type fold

The overall three-dimensional structure of $BR_{187-385}$ can be described as a compressed β-barrel, with two very different faces (**paper I**; figure 1b). The β-barrel consists of one extended twisted β-sheet oppose a distorted β-sheet interspersed with loops and β-turns. Structural homology search reveals that the fold $BR_{187-385}$ adopts is related to the DEv-IgG fold, an MSCRAMM fold-variant (**paper I**; figure 3). The typical topology of this fold can be crudely described as a compressed β-barrel comprising two opposing β-sheets. The fold of $BR_{187-385}$ is a novel variant of this type of fold. A Dali search (86) identified nine structural homologues belonging to the MSCRAMM or SRRP family (**paper I**; table S2).

Analytical ultracentrifugation (AUC) confirms a monomer in solution, with a sedimentation coefficient of 1.85 S that corresponds to a monomer with a hydrodynamic radius of 25 Å (**paper I**; figure 2a). Using size exclusion chromatography (SEC, data not shown), a similar value of the hydrodynamic radius was derived from the retention volume of the $BR_{187-385}$ monomer. Finally, SAXS data collection was performed in order to further establish the monomeric form of $BR_{187-385}$ in solution. Analysis of the data collected clearly shows a monomer in solution (**paper I**; figure 2b; table S1). The concave touchdown observed in the $p(r)$ suggests some elongation and/or several different conformations in solution (80), probably due to some disorder in the N-terminal part of the protein. The experimental SAXS data was subjected to EOM, and the results indicate that the protein construct is globular with N- and C-terminal flexible extensions (**paper I**; figure S1).

3.1.2 The KRT10 binding pocket of $BR_{187-385}$ bares resemblance to a paperclip

The residues 273-341 of PsrP comprise the KRT10 binding region. Within this region there is a substructure that resembles the form of a paperclip, with a front and back (**paper I**; figure 4a). Previously reported experimental data states that only BR constructs containing the entire sequence for KRT10 binding were able to inhibit binding of TIGR4 to $KRT10^{+}$A549 cells (32), which these structural findings might explain. Examination of the distribution of the B-factor values suggests a high flexibility in the loops of $L_{C1/C2}$, $L_{D1/D2}$ and $L_{D3/D4}$ compared to the rest of the structure, which is more rigid (**paper I**; figure 4a; figure S4a). Conformational rearrangements of these loops might be required in order to accommodate KRT10.

The glycine loops of KRT10 are intrinsically disordered protein regions, and interactions with these type regions require hydrophobic interactions (87). When analyzing the surface regions of $BR_{187-385}$ two distinct hydrophobic patches can be found (**paper I**; figure 4b). These pockets might act as the initial anchor points for interaction. Furthermore, analysis of the electrostatic surface of $BR_{187-385}$ unveiled a highly basic groove (**paper I**; figure 4c). It is possible for the acidic helical rod domain of KRT10 to fit in this groove (**paper I**; figure S5). The initial complex formation between $BR_{187-385}$ and the acidic heterodimer of KRT10/KRT1 may be due to the electro-static forces owing to the complementary charges on each molecule (67).

3.1.3 Alanine substitution of several residues within the paperclip-like region of $BR_{187-385}$ disrupts binding to KRT10

A pull-down experiment using Strep-Tag-II $BR_{187-385}$ (STII-$BR_{187-385}$) bound to Ni–NTA bead-immobilized full- length KRT10 confirmed that there is interaction between $BR_{187-385}$ and KRT10 (**paper I**; figure 5a). In order to investigate the specific site of interaction in KRT10, three truncated constructs of KRT10 were designed and produced (**paper II**; figure 5b); head rod domain (HRD), tail rod domain (TRD) and rod domain (ROD). These constructs were tested for interaction with STII-$BR_{187-385}$ in ELISA assays and while the binding capacity was similar for KRT10 full-length (FL) and KRT10-TRD, KRT10-ROD bound significantly less, and binding to KRT-HRD was even lower still (**paper I**; figure 5b).

The head and tail domains of KRT10 both contain glycine-rich tandem repeats, that can be described as loops held together by aromatic residues and/or large nonpolar residues. The tail domain of KRT10 has been suggested to contain more flexible glycine loops and this could be important in its interaction with other proteins (62, 88, 89). Using the peptide-binding prediction site, PepSite (90), two sites, associated with the ‘paperclip’, were identified in $BR_{187-385}$, where several possible LPMs from KRT10 could bind (**paper I**; figure S6). These two predicted sites were used in the design of a mutation study, where ten residues, within or close to these sites, were substituted to alanine (**paper I**; figure 6a). The mutants were then subjected to ELISA, to test the ability to bind KRT10-TRD.

Circular dichroism (CD) performed on all the mutants, and the $BR_{187-385}$ for comparison, showed no significant change in the overall secondary structure, and in the production nor purification by SEC, were any changes devisable. Although it should be noted that the introduced substitutions might have induced small conformational changes, not detectible by CD, that may have had implications of the binding to KRT10-TRD.

3.2 THE BASIC REGION (BR) OF PSRP HOLDS A FURIN PROTEASE CLEAVAGE SITE AND FORMS A DIMER THAT MAY BE IMPORTANT IN BIOFILM FORMATION (PAPER II, MANUSCRIPT)

The proposed self-oligomerization region of PsrP is localized in the N-terminal region of the BR domain. We have, however, found that the KRT10-binding $BR_{187\text{-}385}$, lacking this N-terminal part, forms stable higher oligomers *in vitro*. We show that this N-terminal region is non-globular and probably disordered through SAXS and circular dichroism (CD). In this study we perform a comparative analysis of the long and short constructs of the BR domain of PsrP ($BR_{120\text{-}395}$ and $BR_{187\text{-}385}$, respectively). The results suggest a possibility that the N-terminal region might even play an inhibitory role in the self-oligomerization of BR. We discovered a sequence localized between the globular domain of BR ($BR_{187\text{-}385}$) and the disordered N-terminal region which is recognized by the human furin protease. Cleavage by this protease releases the N-terminal part, which could be a mechanism of activation of oligomerization of PsrP.

3.2.1 $BR_{120\text{-}395}$ carries the previously reported self-oligomerizing portion, but forms a lesser amount of oligomers than the shorter $BR_{187\text{-}385}$ domain

When produced in *Escherichia coli* both constructs of the BR domain forms higher oligomers in addition to the monomeric form. This was unexpected considering that the putative oligomerization domain was supposed to be situated in the N-terminal part of $BR_{120\text{-}395}$ (32), and not present in $BR_{187\text{-}385}$.

The monomeric form of $BR_{187\,-\,385}$ has recently been structurally determined (91). However, during production of the monomeric form we observed higher oligomers during isolation of the monomer using size exclusion chromatography (SEC), (**paper II**, figure 1B). Investigation of these oligomers showed that they are stable and static, i.e. did not convert back to monomeric form when re-applied on the SEC-column (figure 5). The longer construct, $BR_{120\text{-}395}$, forms a lower proportion of oligomers than the construct lacking the oligomerization region (figure 5), (**paper II**, figure 1C).

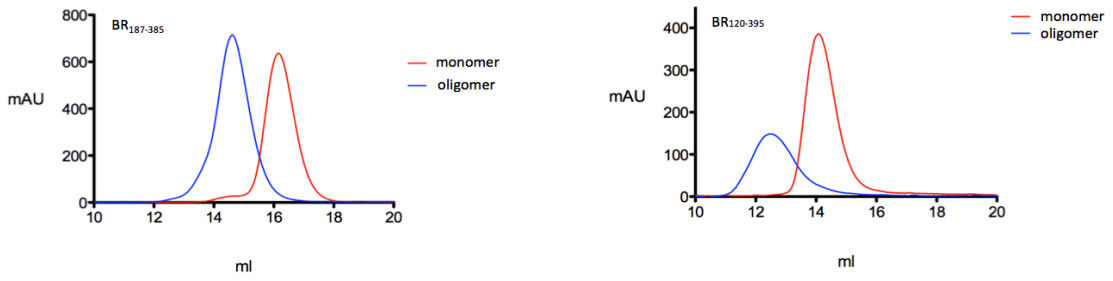

Figure 5. *$BR_{187\text{-}385}$ forms a higher portion of higher oligomeric states than the longer $BR_{120\text{-}395}$, which carries the putative oligomerization domain. The figure shows the chromatograms (SEC) from re-application of isolated monomer and oligomer from (left panel) short construct $BR_{187\text{-}385}$ and (right panel) long construct $BR_{120\text{-}395}$*

3.2.2 N-terminal part of $BR_{120-395}$ comprising the putative self-oligomerizing region resembles the structure of an IDP

When subjected to bioinformatic analysis, the N-terminal part of $BR_{120-395}$, the sequence comprising residues 120-170, was predicted to be disordered (**paper II**; supplementary material; figure S1A).

In order to further investigate the disordered N-terminal region of $BR_{120-395}$, a comparative SAXS study was performed between the longer and shorter constructs (**paper II**; figure 2A,B; table 1). When analyzing the scattering curves for each construct, it is immediately evident that the shape of the longer construct is less globular than the shorter construct. This was expected since we believe that the N-terminal region of $BR_{120-395}$ is disordered, and though the globular domain of $BR_{187-385}$ remains, the overall structure will display more flexible and elongated properties (figure 6). The particle distance distribution function, $p(r)$, (**paper II**; figure 2C) gives D_{max} values of 12.5 ± 1.2 nm and 7.8 ± 0.8 nm for $BR_{120-395}$ and $BR_{187-385}$, respectively. Although it should be noted that when performing the analysis manually (analysis of the scattering data by the indirect Fourier transform method using GNOM (71)), we found that D_{max} values in the range 12 - 18 nm were probable within reasonable degrees of agreement ($\chi^2 \leq 1$) for $BR_{120-395}$. This would suggest that the protein construct takes many different conformations in solution, which is not surprising considering the predicted disorder of the long N-terminal region. Another indication that the protein is elongated and/or takes many different conformations in solution is the concave touchdown of the $p(r)$, for completely globular proteins this would be smooth and convex (80). The peak of $BR_{120-395}$ appears at a slightly longer distance than $BR_{187-385}$, further supporting the theory that the longer construct appears predominately in a more elongated state in solution.

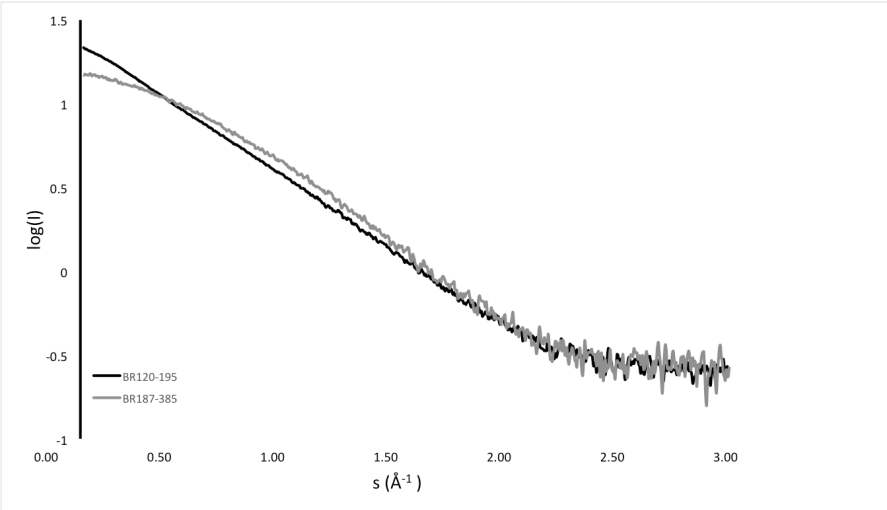

Figure 5 *Experimental scattering curves of $BR_{187-385}$ (grey) and of $BR_{120-395}$ (black). Already here we can see the difference between the constructs in that the curve of the short construct (grey) adopts a more backwards s-shaped curve than does the longer construct. This indicates a more globular shape for $BR_{187-385}$ than for $BR_{120-395}$.*

3.2.3 The N-terminal region of $BR_{120-395}$ is released by cleavage of Furin protease

As previously mentioned, the shorter construct, $BR_{187-385}$, forms a higher portion of oligomeric forms than does the longer construct, $BR_{120-395}$. This might be an indication that the N-terminal part in some way hinders effective oligomerization of $BR_{120-395}$. Interestingly enough, five positive residues in a stretch ($K_{164}RRKR_{168}$) were found in between the KRT10 binding region and the non-globular N-terminal region (**paper II**; figure 1A). The sequence comprises sequence motif R-X-K/R-R that is specifically recognized by ubiquitously expressed furin-like proteases. It is well known that furin proteases activate several pro-protein substrates as well as bacterial and viral pathogenic agents (92-94). Protease cleavage assays confirmed that the N-terminal part could be readily cleaved off $BR_{120-395}$ at the predicted position by human furin protease. The proteolytic activity could be inhibited when a furin-specific inhibitor was present. Moreover, a mutated construct of $BR_{120-395}$, where two arginines where serine substituted ($BR^*_{120-395}$, KRRKR → KSRKS) was produced. The furin protease was unable to cleave off the N-terminal region on the mutant.

A previous study has recently shown that PsrP is capable of self-oligomerization through the BR-domain for pneumococcal biofilm formation (95). Based on these results, we hypothesized that the human, or possibly a putative pneumococcal, furin protease cleaves the N-terminal disordered region off PsrP in order to initiate biofilm formation through self-oligomerization. Microbial biofilms consists mainly of microorganisms adhering to a surface and embedded in exopolysaccharides. Sometimes these are bound together by nucleic acids, lipids and proteins (24, 96), to which the N-terminal region of PsrP might contribute when cleaved from the rest of the protein.

3.2.4 The crystal structure of the $BR_{187-385}$ dimer unveils 3D domain swapping

In order to investigate the stable dimer formation of the $BR_{187-385}$, crystallization trials were set up. However, the crystals obtained proved difficult to reproduce and diffracted poorly. To combat this, a shorter construct was designed and produced. The residues removed were the seven C-terminal residues not built in the model of monomeric $BR_{187-385}$, due to missing electron density (91). The final model shows a dimer structure with two subdomains, each resembling the structure of the $BR_{187-385}$ monomer (**paper II**; figure 5A and 5B).

The crystal structure reveals that the dimer is formed through a 3D domain swap, a well-established mechanism for proteins to form oligomers from their monomers (97). In the $BR_{187-385}$ dimer, the domain swap comprises β-strands from one chain forming a new β-sheet with β-strands from the neighboring chain (**paper II**; figure 6A). The release of proteases and chaperones during forming of the biofilm matrix might also aid in the domain swapping process. The hypothesis we are working with is that the domain-swapped $BR_{187-385}$ dimer could be relevant in establishing stable cell-cell contacts during biofilm formation, and it is possible that the BR domain could form even higher complexes through this mechanism as well.

SAXS data collected for the $BR_{187-385}$ dimer correspond reasonably well to the structure of the $BR_{187-378}$ dimer. Analysis of the data indicated an overall butterfly-like shape, with two distinct domains. D_{max} values of 9.2 ± 0.9 nm were derived from the $p(r)$, which displays a maximum at 2.5 nm with a shoulder at 4.2 nm (**paper II**; figure S2). It should be noted that as for the longer construct $BR_{120-395}$, when performing the analysis manually (analysis of the scattering data by the indirect Fourier transform method using GNOM (71)), we found that D_{max} values in the range 9 - 12 nm were probable within reasonable degrees of agreement ($\chi^2 \leq 1$). Although this would suggest a lesser degree of conformational diversity in solution than for $BR_{120-395}$, it is considerably more than in the shorter $BR_{187-385}$ construct.

3.3 STRUCTURAL ANALYSIS OF THE MINOR ANCILLARY PILUS PROTEIN RrgC IN SOLUTION

There have been many questions regarding the minor ancillary pilus protein RrgC, and therefor much effort has gone into researching this final component of the pneumococcal pilus. As results previous studies have been somewhat contradictory, the need to structurally determine this protein has been great. Secondary structure predictions suggested that RrgC consists of three domains with disordered loop regions between the domains. In this study, we designed a protein construct where the N-terminal signaling peptide and the hydrophobic tail, the CWSS, were removed. The resulting protein could be expressed and purified as a soluble protein. RrgC was studied with SAXS, and the analysis showed a protein comprising short structural elements with flexible linkers, consistent with the structure predictions. Recently, the structure of a mutated full-length RrgC has been solved with X-ray crystallography, and analysis with CRYSOL revealed a nice fit of the determined crystal structure to the experimental SAXS data collected in this study.

3.3.1 SAXS analysis reveals that RrgC adopts an elongated shape in solution.

The protein construct used in this study was designed based on the full-length RrgC, which will be called RrgC. The N-terminal signal peptide and the C-terminal hydrophobic CWSS were omitted in order to acquire a protein construct that could be readily expressed and purified as a soluble protein, and to facilitate crystallization trials as well as SAXS analysis. SAXS experiments and subsequent analysis reveals a protein with an elongated rod shape in solution, most probably containing several structural elements with flexible regions. Structure prediction using Phyre^2 (98) and PsiPred (99, 100) show that RrgC most probably consists of three domains with long flexible, possibly disordered, linkers between them.

Analysis of the scattering curve together with the *p(r)* and the Kratky plot (figure 7) suggests that RrgC is a multidomain protein with characteristics associated to that of a protein comprising smaller domains with flexible linkers, possibly disordered, in solution. The fact that the protein is multi-domain is perhaps most visible in the *p(r)*, with a maximum peak at 2.3 nm, and two shoulders at 4.7 and 7.4 nm (80). The D_{max} values of 12-13.2 nm were probable, suggesting the inherent flexibility and possible conformational diversity of the protein in solution. The Kratky for RrgC is plotted together with the Kratkys for a completely globular protein, as well as a completely extended protein chain, for emphasis. As is clearly visible, the RrgC adopts a conformation in solution more closely related to that of a completely extended protein. The Kratky curves for a globular and extended protein were calculated (80).

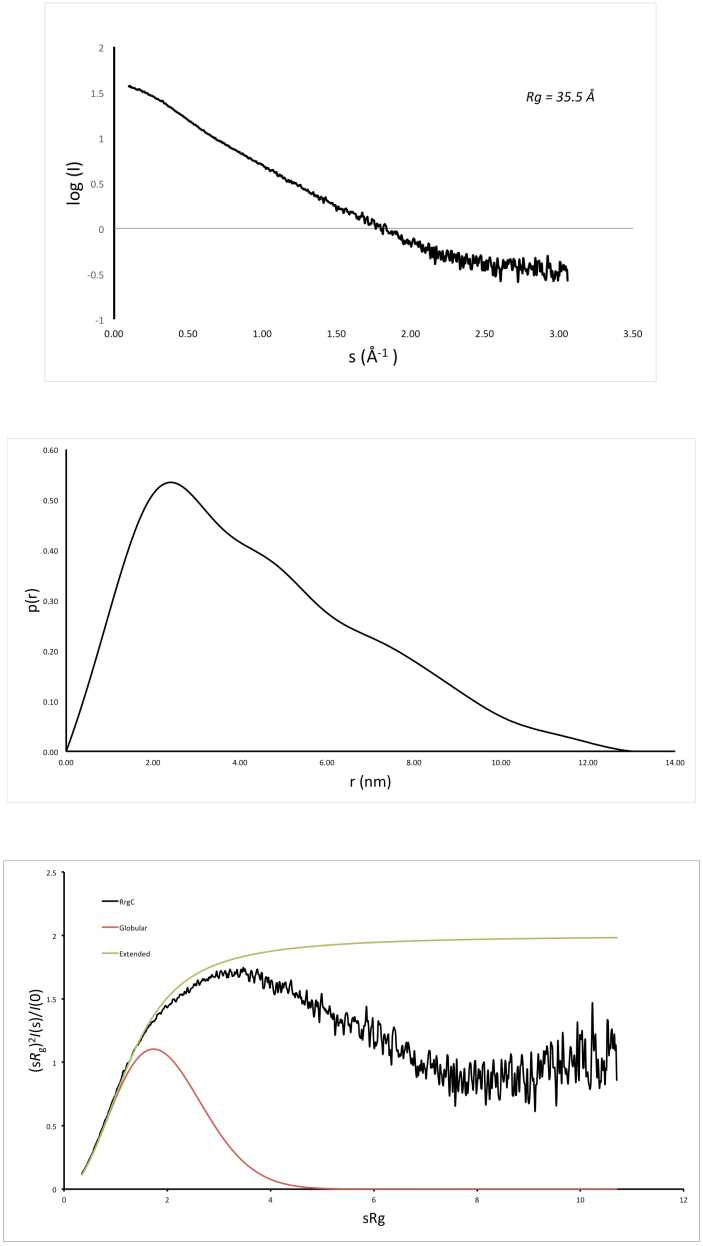

Figure 6. *Top panel shows the buffer-subtracted scattering curve for RrgC, and the middle panel shows the p(r), here for D_{max}=13. The bottom panel shows the Kratky plot with reduced coordinates*

Crystallization trials of RrgC failed, and a mutated version of the full-length RrgC has now been structurally solved (35). In this study, the efforts of trying to crystallize the basic same construct as was designed in our study were unsuccessful. A surface entropy reduction analysis suggested three mutations; Glu-179, Lys-180, Glu-181 into alanines, which is the construct that successfully crystallized. The crystal structure fits nicely with the SAXS data presented here (figure 8), suggesting that the structure contains three domains and takes an elongated rod-like shape in solution.

3.3.2 Experimental procedures

3.3.2.1 Protein production and purification

The coding sequence of the wild type full-length RrgC was obtained from *S. pneumoniae*

TIGR4 (T4) strain chromosomal DNA as previously described (39). The DNA encoding residues 29-268 of the protein RrgC was cloned into the pET21d expression vector (Novagen) using the Fastcloning method (101). The resulting construct containing a C-terminal poly-histidine (HHHHHH) tag was transformed into *E. coli* (T7 express, New England Biolabs) for over-expression at 37°C. Protein expression was induced at OD 0.4-07 with isopropyl-D-1-thiogalacto-pyranoside and cultures were maintained at 25°C overnight.

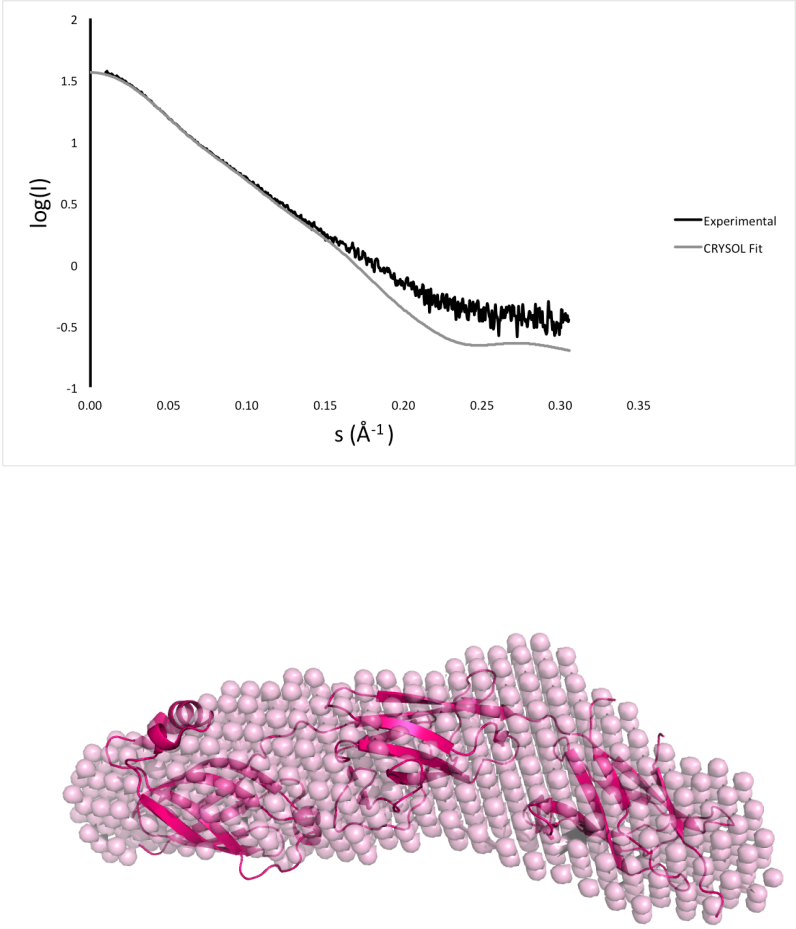

Figure 7. *Top panel shows the CRYSOL fit of the crystal structure of RrgC (PDB-ID 4OQ1), X^2=1.00. Bottom panel depicts the structural envelope for RrgC as modeled ab initio by DAMMIF, superimposed on the crystal structure of RrgC (PDB-ID 4OQ1). Figure created using PyMOL version 1.3 (56)*

The poly-histidine-tagged $RrgC_{29\text{-}268}$ was purified from cell culture lysate using immobilized metal affinity (IMAC) chromatography (HisTrap FF; GE Healthcare) and in a second step SEC, where the protein was loaded on Superdex 75 or 200 columns (GE Healthcare). The purity of <98% was assessed using SDS-PAGE (data not shown) and later confirmed as pure and monodisperse sample with dynamic light scattering (DLS) and SAXS experiments.

3.3.2.2 Small angle X-ray scattering

Sample preparation and data collection. Data sets were collected at beamline 29 (BioSAXS) at the European Synchrotron Radiation Facility (ESRF, Grenoble) (102). Prior to data collection, protein samples were buffer exchanged into phosphate buffered saline (PBS) with 5% glycerol, pH 7.4. Samples were measured at concentrations between 0.3 and 3 mg/ml using a continuous flow cell capillary. Each experimental run was carried out in duplicates in order to evaluate systematic error in the data.

Data analysis. Scattering curves of the buffer was subtracted from each individual protein scattering curves to decimate scattering contributions from the buffer solution. The R_g value (3.55 ± 0.03) were obtained from the Guinier plot (sR_g limit: 0.358-1.101), and the theoretical curve was generated using CRYSOL (82) from the crystal structure of RrgC (PDB-ID: 4OQ1 (35)). The *p(r)* was calculated using GNOM (71). The D_{max} values were obtained from the *p(r)* by running GNOM with varying D_{max} values and D_{max} values in the range 12 – 13.2 nm were probable within reasonable degrees of agreement ($\chi^2 \leq 1$).

The *ab initio* shape reconstruction was performed using DAMMIF (103), was not constrained to any symmetry (P1) and was run twenty times in order to assess the reproducibility of the model reconstruction. The twenty similar, separate 3D structure models obtained were spatially aligned and analyzed for spatial discrepancy using SUPCOMB (104). Models that are dissimilar were rejected and a model representing the volume most populated among the twenty models is presented (figure 8).

4 CONCLUDING REMARKS

The focus of this thesis is the structural elucidation of the function of pneumococcal virulence factors; more specifically the pneumococcal serine-rich repeat protein, PsrP, and the minor ancillary pilus component, RrgC. In order to achieve a greater understanding of the part these proteins play in pneumococcal invasion, a structural approach, involving X-ray crystallography and SAXS, was combined with biochemical methods.

We have shown that the basic region, BR, of Psrp, in its function as an adhesin, is able to interact specifically with the tail rod domain region of keratin 10 and that the most probable mechanism of action is a two-step binding process. This interaction with a human protein is one of several, perhaps many, possible interaction partners for PsrP throughout an *S. pneumonia* infection. This latching to host tissue may be the first step in pneumococcal invasive disease. The next step could be the formation of biofilm in order to consolidate the hold of infection. Incidentally, the putative oligomerization region of the BR domain has been scrutinized as part of this work. The crystal structure of a dimer not containing this region show that BR is able to oligomerize through the mechanism of domain swapping. This together with the fact that the BR domain contains a Furin protease cleavage recognition sequence betwixt the oligomerization and keratin 10 binding region suggests the possibility of a trigger of oligomerization through proteolytic cleavage.

The three-dimensional structure of RrgC has been unknown until recently, and previous reports of this protein have been contradictory. In this thesis we report the structural envelope of this protein, as determined by SAXS, which reveals an elongated, flexible protein in solution. The flexible linkers between the domains provide the possibility of many different conformations in solution. The crystal structure recently published, confirms the SAXS data presented here.

These findings will aid in the future work towards development of novel vaccines and drugs in the quest for abolishment of this microorganisms hold on humanity.

5 ACKNOWLEDGEMENTS

Dear reader,

There are many people who I would like to thank, for seeing me through this challenging time of my life. There are many moments I wished undone, but I will never, ever regret that I took this opportunity, for it has truly educated me in so many ways. There are people who I would never have met, had it not been for this experience, and for that I am grateful. I am tempted to dust off the old saying, “none mentioned, none forgotten”, however there are a few people I will single out.

My main supervisor, **Adnane Achour**, for giving me the opportunity to study at Karolinska Institutet.

My co-supervisor **Per-Åke Nygren**, for the support and fruitful discussions, and for always taking the time to listen.

My co-supervisor **Birgitta Henriques-Normark**, thank you for being such a kind person and for the support during the difficult times.

A special thanks to **John Steen**, doktorandombudsman, for the support and always taking the time to listen and to give good advice.

Kristina Broliden, thank you so much for all the help and support during these last couple of months.

Michael Fored and **Lillemor Melander**, thank you so very much for helping me with this situation, and for the friendly treatment when I joined the department.

Ranjana Sarma, what can I say? There were ups and downs, but we always managed to stay sane together. Thank you for being an amazing mentor, teacher and friend

Eva Allerbring, there were difficult times but we pulled through, like always, thank you for being an amazing friend, **Lara Mentlein** and **Fermín González Bergas**, thank you for the support, always, and for all the laughs. Thank you **Adil** and **Anna-Maria** for all the time you spent listening, **Tim Schulte**, thank you for the support in this process, and thanks to all my other co-workers.

Marco, Anna, Ivan, Tomasz and all other friendly faces sharing the office and labs with the Achour group at SciLife, it would never have been as much fun without you!

Alexis Braun, thank you for always being an amazing friend, I hope we will always be able to see each other through the difficulties life brings.

Peter Järver, thank you for always being there; ready to help with whatever it might be.

To all my friends and former colleagues at KTH, thank you!

Roger, jag är så glad över att ha dig i mitt liv, och skulle aldrig någonsin ha klarat detta utan dig. Jag älskar dig.

Mamma och **pappa**, ni är verkligen de bästa föräldrar man kan tänka sig, jag älskar er.

Lilly, min finaste vän, tack för att du alltid är ett stort stöd och finns där i vått och torrt. Jag älskar dig.

Many thanks to all my friends, family and colleagues not mentioned here, not because of lack of importance but merely due to the lack of space.

6 REFERENCES

1. Watson DA, Musher DM, Jacobson JW, Verhoef J. A brief history of the pneumococcus in biomedical research: a panoply of scientific discovery. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 1993 Nov;17(5):913-24. PubMed PMID: 8286641. Epub 1993/11/01. eng.

2. Poll Tvd, Opal SM. Pathogenesis, treatment, and prevention of pneumococcal pneumonia. The Lancet. 2009 Oct;374.

3. Wilson BA, Salyers AA, Whitt DD, Winkler ME. Bacterial pathogenesis: a molecular approach: American Society for Microbiology (ASM); 2011.

4. Simell B, Auranen K, Kayhty H, Goldblatt D, Dagan R, O'Brien KL. The fundamental link between pneumococcal carriage and disease. Expert review of vaccines. 2012 Jul;11(7):841-55. PubMed PMID: 22913260. Epub 2012/08/24. eng.

5. Kadioglu A, Weiser JN, Paton JC, Andrew PW. The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. Nature reviews Microbiology. 2008 Apr;6(4):288-301. PubMed PMID: 18340341. Epub 2008/03/15. eng.

6. Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Peterson S, et al. Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. Science. 2001 Jul 20;293(5529):498-506. PubMed PMID: 11463916. Epub 2001/07/21. eng.

7. Navarre WW, Schneewind O. Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. Microbiology and molecular biology reviews : MMBR. 1999 Mar;63(1):174-229. PubMed PMID: 10066836. Pubmed Central PMCID: 98962. Epub 1999/03/06. eng.

8. Moxon ER, Kroll JS. The Role of Bacterial Polysaccharide Capsules as Virulence Factors. In: Jann K, Jann B, editors. Bacterial Capsules. Current Topics in Microbiology and Immunology. 150: Springer Berlin Heidelberg; 1990. p. 65-85.

9. Henriques-Normark B, Tuomanen EI. The Pneumococcus: Epidemiology, Microbiology, and Pathogenesis. Cold Spring Harbor Perspectives in Medicine. 2013 July 1, 2013;3(7).

10. Hoskins J, Alborn WE, Jr., Arnold J, Blaszczak LC, Burgett S, DeHoff BS, et al. Genome of the bacterium Streptococcus pneumoniae strain R6. J Bacteriol. 2001 Oct;183(19):5709-17. PubMed PMID: 11544234. Pubmed Central PMCID: 95463. Epub 2001/09/07. eng.

11. Jothi R, Manikandakumar K, Ganesan K, Parthasarathy S. On the analysis of the virulence nature of TIGR4 and R6 strains of Streptococcus pneumoniae using genome comparison tools. Journal of Chemical Sciences. 2007;119(5):559-63.

12. Steel HC, Cockeran R, Anderson R, Feldman C. Overview of Community-Acquired Pneumonia and the Role of Inflammatory Mechanisms in the Immunopathogenesis of Severe Pneumococcal Disease. Mediators of Inflammation. 2013;2013:18.

13. Maestro B, Sanz JM. Novel approaches to fight Streptococcus pneumoniae. Recent patents on anti-infective drug discovery. 2007 Nov;2(3):188-96. PubMed PMID: 18221176. Epub 2008/01/29. eng.

14. Richter SS, Heilmann KP, Dohrn CL, Riahi F, Beekmann SE, Doern GV. Changing epidemiology of antimicrobial-resistant Streptococcus pneumoniae in the United States, 2004-2005. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2009 Feb 1;48(3):e23-33. PubMed PMID: 19115971. Epub 2009/01/01. eng.

15. Romero P, Lopez R, Garcia E. Key role of amino acid residues in the dimerization and catalytic activation of the autolysin LytA, an important virulence factor in Streptococcus pneumoniae. The Journal of biological chemistry. 2007 Jun 15;282(24):17729-37. PubMed PMID: 17439951. Epub 2007/04/19. eng.

16. Mond JJ, Vos Q, Lees A, Snapper CM. T cell independent antigens. Current opinion in immunology. 1995 Jun;7(3):349-54. PubMed PMID: 7546399. Epub 1995/06/01. eng.

17. Lesinski GB, Westerink MA. Novel vaccine strategies to T-independent antigens. Journal of microbiological methods. 2001 Nov;47(2):135-49. PubMed PMID: 11576678. Epub 2001/09/29. eng.

18. García E, Yuste J, Rodríguez-Cerrato V, Moscoso M, García P. Streptococcus pneumoniae: from molecular biology to host-pathogen interactions. Journal of Applied Biomedicine. 2010;8(3):131-40.

19. Byington CL, Samore MH, Stoddard GJ, Barlow S, Daly J, Korgenski K, et al. Temporal Trends of Invasive Disease Due to Streptococcus pneumoniae among Children in the Intermountain West: Emergence of Nonvaccine Serogroups. Clinical Infectious Diseases. 2005 July 1, 2005;41(1):21-9.

20. Steenhoff AP, Shah SS, Ratner AJ, Patil SM, McGowan KL. Emergence of Vaccine-Related Pneumococcal Serotypes as a Cause of Bacteremia. Clinical Infectious Diseases. 2006 April 1, 2006;42(7):907-14.

21. Smith KJ, Wateska AR, Nowalk MP, Raymund M, Nuorti JP, Zimmerman RK. Cost-effectiveness of adult vaccination strategies using pneumococcal conjugate vaccine compared with pneumococcal polysaccharide vaccine. JAMA. 2012 Feb 22;307(8):804-12. PubMed PMID: 22357831. Epub 2012/02/24. eng.

22. Shapiro ED. Prevention of Pneumococcal Infection With Vaccines. JAMA. 2012;307(8):847-49.

23 Lopez R. Pneumococcus: the sugar-coated bacteria. International microbiology : the official journal of the Spanish Society for Microbiology. 2006 Sep;9(3):179-90. PubMed PMID: 17061208. Epub 2006/10/25. eng.

24. Flemming H-C, Wingender J. The biofilm matrix. Nat Rev Micro. 2010;8(9):623-33.

25. Sanchez CJ, Hurtgen BJ, Lizcano A, Shivshankar P, Cole GT, Orihuela CJ. Biofilm and planktonic pneumococci demonstrate disparate immunoreactivity to human convalescent sera. BMC microbiology. 2011;11:245. PubMed PMID: 22047041. Pubmed Central PMCID: 3216281. Epub 2011/11/04. eng.

26. Pericone CD, Bae D, Shchepetov M, McCool T, Weiser JN. Short-sequence tandem and nontandem DNA repeats and endogenous hydrogen peroxide production contribute to genetic instability of Streptococcus pneumoniae. Journal of Bacteriology. 2002;184(16):4392-9.

27. Bergmann S, Hammerschmidt S. Versatility of pneumococcal surface proteins. Microbiology (Reading, England). 2006 Feb;152(Pt 2):295-303. PubMed PMID: 16436417. Epub 2006/01/27. eng.

28. Hammerschmidt S. Adherence molecules of pathogenic pneumococci. Current opinion in microbiology. 2006 Feb;9(1):12-20. PubMed PMID: 16338163. Epub 2005/12/13. eng.

29. Jedrzejas MJ. Extracellular virulence factors of Streptococcus pneumoniae. Frontiers in bioscience : a journal and virtual library. 2004 Jan 1;9:891-914. PubMed PMID: 14766417. Epub 2004/02/10. eng.

30. Obert C, Sublett J, Kaushal D, Hinojosa E, Barton T, Tuomanen EI, et al. Identification of a Candidate Streptococcus pneumoniae core genome and regions of diversity correlated with invasive pneumococcal disease. Infection and immunity. 2006 Aug;74(8):4766-77. PubMed PMID: 16861665. Pubmed Central PMCID: 1539573. Epub 2006/07/25. eng.

31. Munoz-Almagro C, Selva L, Sanchez CJ, Esteva C, de Sevilla MF, Pallares R, et al. PsrP, a protective pneumococcal antigen, is highly prevalent in children with pneumonia and is strongly associated with clonal type. Clinical and vaccine immunology : CVI. 2010 Nov;17(11):1672-8. PubMed PMID: 20861332. Pubmed Central PMCID: 2976083. Epub 2010/09/24. eng.

32. Shivshankar P, Sanchez C, Rose LF, Orihuela CJ. The Streptococcus pneumoniae adhesin PsrP binds to Keratin 10 on lung cells. Molecular microbiology. 2009 Aug;73(4):663-79. PubMed PMID: 19627498. Pubmed Central PMCID: 2753542. Epub 2009/07/25. eng.

33. Takahashi Y, Konishi K, Cisar JO, Yoshikawa M. Identification and characterization of hsa, the gene encoding the sialic acid-binding adhesin of Streptococcus gordonii DL1. Infection and immunity. 2002 Mar;70(3):1209-18. PubMed PMID: 11854202. Pubmed Central PMCID: 127787. Epub 2002/02/21. eng.

34. Takahashi Y, Yajima A, Cisar JO, Konishi K. Functional analysis of the Streptococcus gordonii DL1 sialic acid-binding adhesin and its essential role in bacterial binding to platelets. Infection and immunity. 2004 Jul;72(7):3876-82. PubMed PMID: 15213130. Pubmed Central PMCID: 427394. Epub 2004/06/24. eng.

35. Shaik MM, Maccagni A, Tourcier G, Di Guilmi AM, Dessen A. Structural Basis of Pilus Anchoring by the Ancillary Pilin RrgC of Streptococcus pneumoniae. The Journal of biological chemistry. 2014 Jun 13;289(24):16988-97. PubMed PMID: 24755220. Pubmed Central PMCID: 4059141. Epub 2014/04/24. Eng.

36. Hava DL, Camilli A. Large-scale identification of serotype 4 Streptococcus pneumoniae virulence factors. Molecular microbiology. 2002 Sep;45(5):1389-406. PubMed PMID: 12207705. Pubmed Central PMCID: 2788772. Epub 2002/09/05. eng.

37. Rose L, Shivshankar P, Hinojosa E, Rodriguez A, Sanchez CJ, Orihuela CJ. Antibodies against PsrP, a novel Streptococcus pneumoniae adhesin, block adhesion and protect mice against pneumococcal challenge. The Journal of infectious diseases. 2008 Aug 1;198(3):375-83. PubMed PMID: 18507531. Epub 2008/05/30. eng.

38. Scott JR, Zahner D. Pili with strong attachments: Gram-positive bacteria do it differently. Molecular microbiology. 2006 Oct;62(2):320-30. PubMed PMID: 16978260. Epub 2006/09/19. eng.

39. Barocchi MA, Ries J, Zogaj X, Hemsley C, Albiger B, Kanth A, et al. A pneumococcal pilus influences virulence and host inflammatory responses. Proceedings of the National Academy of Sciences of the United States of America. 2006 Feb 21;103(8):2857-62. PubMed PMID: 16481624. Pubmed Central PMCID: 1368962. Epub 2006/02/17. eng.

40. Kline KA, Dodson KW, Caparon MG, Hultgren SJ. A tale of two pili: assembly and function of pili in bacteria. Trends in microbiology. 2010 May;18(5):224-32. PubMed PMID: 20378353. Pubmed Central PMCID: 3674877. Epub 2010/04/10. eng.

41. Hendrickx AP, Budzik JM, Oh SY, Schneewind O. Architects at the bacterial surface - sortases and the assembly of pili with isopeptide bonds. Nature reviews Microbiology. 2011 Mar;9(3):166-76. PubMed PMID: 21326273. Epub 2011/02/18. eng.

42. Basset A, Trzcinski K, Hermos C, O'Brien KL, Reid R, Santosham M, et al. Association of the pneumococcal pilus with certain capsular serotypes but not with increased virulence. Journal of clinical microbiology. 2007 Jun;45(6):1684-9. PubMed PMID: 17392439. Pubmed Central PMCID: 1933072. Epub 2007/03/30. eng.

43. El Mortaji L, Fenel D, Vernet T, Di Guilmi AM. Association of RrgA and RrgC into the Streptococcus pneumoniae pilus by sortases C-2 and C-3. Biochemistry. 2012 Jan 10;51(1):342-52. PubMed PMID: 22122269. Epub 2011/11/30. eng.

44. El Mortaji L, Terrasse R, Dessen A, Vernet T, Di Guilmi AM. Stability and assembly of pilus subunits of Streptococcus pneumoniae. The Journal of biological chemistry. 2010 Apr 16;285(16):12405-15. PubMed PMID: 20147289. Pubmed Central PMCID: 2852978. Epub 2010/02/12. eng.

45. Spraggon G, Koesema E, Scarselli M, Malito E, Biagini M, Norais N, et al. Supramolecular organization of the repetitive backbone unit of the Streptococcus pneumoniae pilus. PloS one. 2010;5(6):e10919. PubMed PMID: 20559564. Pubmed Central PMCID: 2886109. Epub 2010/06/19. eng.

46. Gentile MA, Melchiorre S, Emolo C, Moschioni M, Gianfaldoni C, Pancotto L, et al. Structural and functional characterization of the Streptococcus pneumoniae RrgB pilus backbone D1 domain. The Journal of biological chemistry. 2011 Apr 22;286(16):14588-97. PubMed PMID: 21367860. Pubmed Central PMCID: 3077656. Epub 2011/03/04. eng.

47. El Mortaji L, Contreras-Martel C, Moschioni M, Ferlenghi I, Manzano C, Vernet T, et al. The full-length Streptococcus pneumoniae major pilin RrgB crystallizes in a fibre-like structure, which presents the D1 isopeptide bond and provides details on the mechanism of pilus polymerization. The Biochemical journal. 2012 Feb 1;441(3):833-41. PubMed PMID: 22013894. Epub 2011/10/22. eng.

48. Paterson NG, Baker EN. Structure of the full-length major pilin from Streptococcus pneumoniae: implications for isopeptide bond formation in gram-positive bacterial pili. PloS one. 2011;6(7):e22095. PubMed PMID: 21760959. Pubmed Central PMCID: 3132780. Epub 2011/07/16. eng.

49. Kang HJ, Baker EN. Structure and assembly of Gram-positive bacterial pili: unique covalent polymers. Current opinion in structural biology. 2012 Apr;22(2):200-7. PubMed PMID: 22342454. Epub 2012/02/22. eng.

50. Orrskog S, Rounioja S, Spadafina T, Gallotta M, Norman M, Hentrich K, et al. Pilus adhesin RrgA interacts with complement receptor 3, thereby affecting macrophage function and systemic pneumococcal disease. mBio. 2012;4(1):e00535-12. PubMed PMID: 23269830. Pubmed Central PMCID: 3531807. Epub 2012/12/28. eng.

51. LeMieux J, Hava DL, Basset A, Camilli A. RrgA and RrgB are components of a multisubunit pilus encoded by the Streptococcus pneumoniae rlrA pathogenicity islet. Infection and immunity. 2006 Apr;74(4):2453-6. PubMed PMID: 16552078. Pubmed Central PMCID: 1418942. Epub 2006/03/23. eng.

52. Fälker S, Nelson AL, Morfeldt E, Jonas K, Hultenby K, Ries J, et al. Sortase-mediated assembly and surface topology of adhesive pneumococcal pili. Molecular microbiology. 2008;70(3):595-607.

53. Hilleringmann M, Giusti F, Baudner BC, Masignani V, Covacci A, Rappuoli R, et al. Pneumococcal pili are composed of protofilaments exposing adhesive clusters of Rrg A. PLoS pathogens. 2008 Mar;4(3):e1000026. PubMed PMID: 18369475. Pubmed Central PMCID: 2265430. Epub 2008/03/29. eng.

54. Hilleringmann M, Ringler P, Muller SA, De Angelis G, Rappuoli R, Ferlenghi I, et al. Molecular architecture of Streptococcus pneumoniae TIGR4 pili. The EMBO journal. 2009 Dec 16;28(24):3921-30. PubMed PMID: 19942854. Pubmed Central PMCID: 2797065. Epub 2009/11/28. eng.

55. Izore T, Contreras-Martel C, El Mortaji L, Manzano C, Terrasse R, Vernet T, et al. Structural basis of host cell recognition by the pilus adhesin from Streptococcus pneumoniae. Structure. 2010 Jan 13;18(1):106-15. PubMed PMID: 20152157. Epub 2010/02/16. eng.

56 Schrödinger L. *PyMOL Molecular Graphics System*. 2010.

57. Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nat Immunol. 2010;11(5):373-84.

58. Cockeran R, Theron AJ, Steel HC, Matlola NM, Mitchell TJ, Feldman C, et al. Proinflammatory Interactions of Pneumolysin with Human Neutrophils. Journal of Infectious Diseases. 2001 February 15, 2001;183(4):604-11.

59. Steven AC, Hainfeld JF, Trus BL, Wall JS, Steinert PM. Epidermal keratin filaments assembled in vitro have masses-per-unit-length that scale according to average subunit mass: structural basis for homologous packing of subunits in intermediate filaments. The Journal of cell biology. 1983 Dec;97(6):1939-44. PubMed PMID: 6196371. Pubmed Central PMCID: 2112722. Epub 1983/12/01. eng.

60. Zhou XM, Idler WW, Steven AC, Roop DR, Steinert PM. The complete sequence of the human intermediate filament chain keratin 10. Subdomainal divisions and model for folding of end domain sequences. The Journal of biological chemistry. 1988 Oct 25;263(30):15584-9. PubMed PMID: 2459124. Epub 1988/10/25. eng.

61. Steinert PM, Marekov LN, Fraser RDB, Parry DAD. Keratin Intermediate Filament Structure: Crosslinking Studies Yield Quantitative Information on Molecular Dimensions and Mechanism of Assembly. Journal of Molecular Biology. 1993;230(2):436-52.

62. Steinert PM. Structure, Function, and Dynamics of Keratin Intermediate Filaments. J Investig Dermatol. 1993;100(6):729-34.

63. Magin TM, Vijayaraj P, Leube RE. Structural and regulatory functions of keratins. Experimental Cell Research. 2007;313(10):2021-32.

64. Parry DA, Steinert PM. Intermediate filaments: molecular architecture, assembly, dynamics and polymorphism. Quarterly reviews of biophysics. 1999 May;32(2):99-187. PubMed PMID: 10845237. Epub 2000/06/09. eng.

65. O'Brien LM, Walsh EJ, Massey RC, Peacock SJ, Foster TJ. Staphylococcus aureus clumping factor B (ClfB) promotes adherence to human type I cytokeratin 10: implications for nasal colonization. Cellular microbiology. 2002;4(11):759-70.

66. Haim M, Trost A, Maier CJ, Achatz G, Feichtner S, Hintner H, et al. Cytokeratin 8 interacts with clumping factor B: a new possible virulence factor target. Microbiology (Reading, England). 2010;156(12):3710-21.

67. Samen U, Eikmanns BJ, Reinscheid DJ, Borges F. The surface protein Srr-1 of Streptococcus agalactiae binds human keratin 4 and promotes adherence to epithelial HEp-2 cells. Infection and immunity. 2007 Nov;75(11):5405-14. PubMed PMID: 17709412. Pubmed Central PMCID: 2168289. Epub 2007/08/22. eng.

68. Lee CH, Kim MS, Chung BM, Leahy DJ, Coulombe PA. Structural basis for heteromeric assembly and perinuclear organization of keratin filaments. Nature structural & molecular biology. 2012 Jul;19(7):707-15. PubMed PMID: 22705788. Pubmed Central PMCID: 3864793. Epub 2012/06/19. eng.

69. Putnam CD, Hammel M, Hura GL, Tainer JA. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Quarterly reviews of biophysics. 2007 Aug;40(3):191-285. PubMed PMID: 18078545. Epub 2007/12/15. eng.

70. Svergun DI, Petoukhov MV, Koch MH. Determination of domain structure of proteins from X-ray solution scattering. Biophysical journal. 2001 Jun;80(6):2946-53. PubMed PMID: 11371467. Pubmed Central PMCID: 1301478. Epub 2001/05/24. eng.

71. Svergun D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. Journal of Applied Crystallography. 1992;25(4):495-503.

72. Sokolova AV, Volkov VV, Svergun DI. Prototype of a database for rapid protein classification based on solution scattering data. Journal of Applied Crystallography. 2003;36(3-1):865-8.

73. Sokolova AV, Volkov VV, Svergun DI. Database for rapid protein classification based on small-angle X-ray scattering data. Crystallogr Rep. 2003 2003/11/01;48(6):959-65. English.

74. Koch MH, Vachette P, Svergun DI. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Quarterly reviews of biophysics. 2003 May;36(2):147-227. PubMed PMID: 14686102. Epub 2003/12/23. eng.

75. Tsutakawa SE, Hura GL, Frankel KA, Cooper PK, Tainer JA. Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography. Journal of structural biology. 2007 May;158(2):214-23. PubMed PMID: 17182256. Epub 2006/12/22. eng.

76. Rhodes G. Crystallography made crystal clear: a guide for users of macromolecular models: Academic press; 2010.

77. Sarma R, Mulder DW, Brecht E, Szilagyi RK, Seefeldt LC, Tsuruta H, et al. Probing the MgATP-bound conformation of the nitrogenase Fe protein by solution small-angle X-ray scattering. Biochemistry. 2007 Dec 11;46(49):14058-66. PubMed PMID: 18001132. Pubmed Central PMCID: 3289971. Epub 2007/11/16. eng.

78. Guinier A, Fournet G, Walker CB, Yudowitch KL. Small-angle scattering of X-rays. 1955.

79. Kratky O, Porod G. Röntgenuntersuchung gelöster fadenmoleküle. Recueil des Travaux Chimiques des Pays-Bas. 1949;68(12):1106-22.

80. Receveur-Brechot V, Durand D. How random are intrinsically disordered proteins? A small angle scattering perspective. Current protein & peptide science. 2012 Feb;13(1):55-75. PubMed PMID: 22044150. Pubmed Central PMCID: 3394175. Epub 2011/11/03. eng.

81. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography. 2003;36(5):1277-82.

82. Svergun D, Barberato C, Koch M. CRYSOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. Journal of Applied Crystallography. 1995;28(6):768-73.

83. Petoukhov MV, Svergun DI. Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophysical journal. 2005;89(2):1237-50.

84. Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, et al. New developments in the ATSAS program package for small-angle scattering data analysis. Journal of Applied Crystallography. 2012;45(2):342-50.

85. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. Structural characterization of flexible proteins using small-angle X-ray scattering. Journal of the American Chemical Society. 2007 May 2;129(17):5656-64. PubMed PMID: 17411046. Epub 2007/04/07. eng.

86. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. Searching protein structure databases with DaliLite v.3. Bioinformatics (Oxford, England). 2008 Dec 1;24(23):2780-1. PubMed PMID: 18818215. Pubmed Central PMCID: 2639270. Epub 2008/09/27. eng.

87. Mészáros B, Tompa P, Simon I, Dosztányi Z. Molecular Principles of the Interactions of Disordered Proteins. Journal of Molecular Biology. 2007;372(2):549-61.

88. Steinert PM, Mack JW, Korge BP, Gan SQ, Haynes SR, Steven AC. Glycine loops in proteins: their occurrence in certain intermediate filament chains, loricrins and single-stranded RNA binding proteins.

International journal of biological macromolecules. 1991 Jun;13(3):130-9. PubMed PMID: 1716976. Epub 1991/06/01. eng.

89. Strnad P, Usachov V, Debes C, Grater F, Parry DA, Omary MB. Unique amino acid signatures that are evolutionarily conserved distinguish simple-type, epidermal and hair keratins. Journal of cell science. 2011 Dec 15;124(Pt 24):4221-32. PubMed PMID: 22215855. Pubmed Central PMCID: 3258107. Epub 2012/01/05. eng.

90. Petsalaki E, Stark A, García-Urdiales E, Russell RB. Accurate Prediction of Peptide Binding Sites on Protein Surfaces. PLoS Comput Biol. 2009;5(3):e1000335.

91. Schulte T, Lofling J, Mikaelsson C, Kikhney A, Hentrich K, Diamante A, et al. The basic keratin 10-binding domain of the virulence-associated pneumococcal serine-rich protein PsrP adopts a novel MSCRAMM fold. Open biology. 2014 Jan;4:130090. PubMed PMID: 24430336. Pubmed Central PMCID: 3909270. Epub 2014/01/17. eng.

92. Klimpel KR, Molloy SS, Thomas G, Leppla SH. Anthrax toxin protective antigen is activated by a cell surface protease with the sequence specificity and catalytic properties of furin. Proceedings of the National Academy of Sciences of the United States of America. 1992 Nov 1;89(21):10277-81. PubMed PMID: 1438214. Pubmed Central PMCID: 50321. Epub 1992/11/01. eng.

93. Thomas G. Furin at the cutting edge: from protein traffic to embryogenesis and disease. Nature reviews Molecular cell biology. 2002 Oct;3(10):753-66. PubMed PMID: 12360192. Pubmed Central PMCID: 1964754. Epub 2002/10/03. eng.

94. Molloy SS, Bresnahan PA, Leppla SH, Klimpel KR, Thomas G. Human furin is a calcium-dependent serine endoprotease that recognizes the sequence Arg-X-X-Arg and efficiently cleaves anthrax toxin protective antigen. The Journal of biological chemistry. 1992 Aug 15;267(23):16396-402. PubMed PMID: 1644824. Epub 1992/08/15. eng.

95. Sanchez CJ, Shivshankar P, Stol K, Trakhtenbroit S, Sullam PM, Sauer K, et al. The pneumococcal serine-rich repeat protein is an intra-species bacterial adhesin that promotes bacterial aggregation in vivo and in biofilms. PLoS pathogens. 2010;6(8):e1001044. PubMed PMID: 20714350. Pubmed Central PMCID: 2920850. Epub 2010/08/18. eng.

96. Moscoso M, Garcia E, Lopez R. Pneumococcal biofilms. International microbiology : the official journal of the Spanish Society for Microbiology. 2009 Jun;12(2):77-85. PubMed PMID: 19784927. Epub 2009/09/29. eng.

97. Bennett MJ, Schlunegger MP, Eisenberg D. 3D domain swapping: a mechanism for oligomer assembly. Protein science : a publication of the Protein Society. 1995 Dec;4(12):2455-68. PubMed PMID: 8580836. Pubmed Central PMCID: 2143041. Epub 1995/12/01. eng.

98. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols. 2009;4(3):363-71. PubMed PMID: 19247286. Epub 2009/02/28. eng.

99. Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Research. 2013 June 8, 2013.

100. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 1999 Sep 17;292(2):195-202. PubMed PMID: 10493868. Epub 1999/09/24. eng.

101. Li C, Wen A, Shen B, Lu J, Huang Y, Chang Y. FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. BMC biotechnology. 2011;11:92. PubMed PMID: 21992524. Pubmed Central PMCID: 3207894. Epub 2011/10/14. eng.

102. Pernot P, Round A, Barrett R, De Maria Antolinos A, Gobbo A, Gordon E, et al. Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. Journal of synchrotron radiation. 2013 Jul;20(Pt 4):660-4. PubMed PMID: 23765312. Pubmed Central PMCID: 3943554. Epub 2013/06/15. eng.

103. Franke D, Svergun DI. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. Journal of Applied Crystallography. 2009;42(2):342-6.

104. Kozin MB, Svergun DI. Automated matching of high- and low-resolution structural models. Journal of Applied Crystallography. 2001;34(1):33-41.