Structural and functional studies of streptococcal surface adhesins

Åsa Nylander
Till Mor

“42”
- Douglas Adams

“The true delight is in the finding out rather than in the knowing.”
- Isaac Asimov
# Table of Contents

Table of Contents

List of original publications

Abstract

Abbreviations

Enkel sammanfattning på svenska

Introduction

Materials and methods

Aim of thesis

Result and discussion

Structures of the C-terminal domain of the antigen I/II protein family (paper I & IV)

Structural and functional analysis of the N-terminal domain of the *Streptococcus gordonii* adhesin Sgo0707 (paper II)

Protein expression and purification of *Streptococcus pyogenes* adhesin AspA (paper III)

Conclusion

Acknowledgements

References
List of original publications

This thesis is based on the following publications and manuscripts, referred to by their roman numerals throughout the thesis.

Paper I  

Paper II  

Paper III  
Nylander Å., Hall M, Jenkinson H., Persson K. Expression and purification of Streptococcus pyogenes adhesin AspA. Manuscript

Paper IV  
Hall M., Nylander Å., Jenkinson H., Persson K. Structure of the C-terminal domain of the Streptococcus pyogenes antigen I/II-family protein AspA. Manuscript
Abstract

The oral cavity is home to an array of microorganisms that are associated with dental plaque. Some Gram-positive bacteria are common inhabitants of the oral cavity and in order to colonize such a unique environment adhesion becomes essential and is accomplish by adhesins expressed on the bacterial surface. Adhesins can interact with host molecules or with structures on the resident oral microbial flora. Members of the antigen I/II (AgI/II) protein family are commonly found on the surface of oral streptococci and have the unique feature that their putative adhesin domain is located in the centre of the primary sequence. Crystal structures representing parts of the C-terminal domains from two AgI/II members, SpaP from *Streptococcus mutans* and AspA from *Streptococcus pyogenes*, were determined to 2.2 and 1.8 Å resolution respectively. The structures are very similar and consist of two domains with DEv-IgG folds. The proteins are stabilized by intramolecular isopeptide bonds and tightly coordinated metal ions.

Another group of surface proteins is the microbial surface components recognizing adhesive matrix molecules (MSCRAMMs) that have their putative adhesin domain in the N-terminal, presented on a stalk formed by multiples of repeated C-terminal domains. Sgo0707 from *Streptococcus gordonii* is an example of this group of proteins and its N-terminal domain was determined to 2.1 Å resolution. The structure consists of two domains, N1 and N2, both of which adopt β-sandwiches. In the Sgo0707 structure no isopeptide bonds or metal ions were detected. A putative binding cleft is present in the N1 domain. Functional studies revealed collagen type-1 and keratinocytes as possible binding partners.

In order to further characterize the AgI/II protein AspA from *S. pyogenes* a long form of the protein, AspA-AVPC, was expressed and purified. During the purification process it was observed that the protein fragmented into two major parts. This process could be inhibited by the addition of 0.5 mM EDTA during protein purification.

In conclusion, these studies have resulted in adding to the knowledge of protein structures and function of streptococcal surface proteins.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D</td>
<td>Three dimensional</td>
</tr>
<tr>
<td>AgI/II</td>
<td>Antigen I/II</td>
</tr>
<tr>
<td>CLSM</td>
<td>Confocal laser scanning microscopy</td>
</tr>
<tr>
<td>CM</td>
<td>Cytoplasmic membrane</td>
</tr>
<tr>
<td>CW</td>
<td>Cell wall</td>
</tr>
<tr>
<td>CWA</td>
<td>Cell wall anchor</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EPS</td>
<td>Exopolysaccharides</td>
</tr>
<tr>
<td>GCF</td>
<td>Gingival crevicular fluid</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-acetylglucosamine</td>
</tr>
<tr>
<td>IgG-C</td>
<td>Immunoglobulin G-constant</td>
</tr>
<tr>
<td>IMAC</td>
<td>Immobilized metal affinity chromatography</td>
</tr>
<tr>
<td>IPS</td>
<td>Intracellular polysaccharides</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>LTA</td>
<td>Lipoteichoic acids</td>
</tr>
<tr>
<td>MAD</td>
<td>Multiple-wavelength anomalous dispersion</td>
</tr>
<tr>
<td>MALDI-TOF MS</td>
<td>Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry</td>
</tr>
<tr>
<td>MIR</td>
<td>Multiple isomorphous replacement</td>
</tr>
<tr>
<td>MR</td>
<td>Molecular replacement</td>
</tr>
<tr>
<td>MSCRAMM</td>
<td>Microbial surface components recognizing adhesive matrix molecules</td>
</tr>
<tr>
<td>MurNAc</td>
<td>N-acetylmuramic acid</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root-mean-square deviation</td>
</tr>
<tr>
<td>S.</td>
<td><em>Streptococcus</em></td>
</tr>
<tr>
<td>S-layer</td>
<td>Surface layer</td>
</tr>
<tr>
<td>SAD</td>
<td>Single-wavelength anomalous dispersion</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>Sec</td>
<td>Secretary</td>
</tr>
<tr>
<td>SeMet</td>
<td>Selenomethionine</td>
</tr>
<tr>
<td>SRP</td>
<td>Signal recognition particle</td>
</tr>
<tr>
<td>Staph.</td>
<td><em>Staphylococcus</em></td>
</tr>
<tr>
<td>TA</td>
<td>Teichoic acids</td>
</tr>
<tr>
<td>Tat</td>
<td>Twin arginine transport</td>
</tr>
<tr>
<td>vWFA</td>
<td>von Willebrand factor A</td>
</tr>
<tr>
<td>Å</td>
<td>Ångström (1 Å=0.1 nm)</td>
</tr>
</tbody>
</table>
Enkel sammanfattning på svenska

I munhålan finns det bakterier som bildar samhällen på tänderna, s.k. (tand) plack. För att bakterier ska kunna kolonisera munhålan måste de snabbt kunna fästa till molekyler och ytor, annars sväljs de ner i magsäcken där magsyran oskadliggör dem. Bakterier uttrycker därför ytprotein som gör att de kan fästa till molekyler i munhålan och på så sätt stanna där. Mer än 700 olika bakteriearter har påvisats i munhålan, men vanligtvis ligger medelvärdet på ca 100 i en frisk individs mun. Olika streptokockarter är vanliga i munhålan och de uttrycker flera ytproteiner. I denna avhandling beskrivs; pili, antigentyp I/II, och så kallade MSCRAMM.


Denna avhandling baseras i huvudsak på tre strukturer av olika streptokockproteiner; SpaP från *Streptococcus mutans*, Sgo0707 från *Streptococcus gordonii*, och AspA från *Streptococcus pyogenes* som jag här sammanfattar väldigt kort. Strukturrellt är SpaP från *S. mutans* och AspA från *S. pyogenes* från antigen I/II familjen väldigt lika varandra (artikel I och manuskript IV). De är båda uppbyggda av två domäner som består av β-strängar, som i sin tur bildar en s.k. β-sandwich. När modellerna av proteinerna jämförs genom att de placeras på varandra, sammanfaller de flesta av de analyserade kolatomerna i huvudkedjorna, eftersom proteinerna är strukturellt mycket lika. Strukturen av MSCRAMM-proteinet Sgo0707 från *S. gordonii* (artikel II) uppriser även det två domäner, som båda veckar sig i β-sandwichar. En av domänerna har en möjlig bindningsficka så experiment utfördes för att ta reda på vad som kan tänkas binda till denna ficka. Resultatet visade att collagen typ-1 (ett protein som bl.a. finns i senor och ben) och keratinocyter (cellerna som täcker munhålan mjuka delar) binder till Sgo0707 vilket visar att Sgo0707 är inblandad i vidhäftning. Planen var också att kristallisera en längre form av AspA proteinet (AspA-AVPC) när det upptäcktes att proteinet bröts ner till två mindre fragment,
men orsaken till nedbrytningen är ännu inte känd. Vi märkte att om EDTA (en molekyl som binder metalljoner) tillsattes till proteinet så uteblev nedbrytningen. Studier fortgår för att få reda på betydelsen av denna nedbrytning.
Introduction

Microbiology in the oral cavity
The oral cavity is the portal to the respiratory tract as well as to the gastrointestinal tract and is therefore an important gateway into the body. The acidic environment of the stomach is an effective defence mechanism against ingested microorganisms. The human mouth consists of mucosal surfaces, teeth, and saliva, which makes the oral cavity a unique environment. The amounts of microorganisms present on the mucosal surfaces are quite low due to the regular shedding of epithelial cells (desquamation). The papillary structures of the tongue can harbour obligate anaerobes as these surfaces give protection against the salivary flow and have low redox potential. The non-shedding tissues of the teeth can accumulate a larger mass of microorganisms than the mucosal tissue. The local environments of teeth vary; the tooth is divided into a subgingival region (below the gingival border), and a supragingival region (above the gingival border) (Marcotte and Lavoie, 1998). The buccal (cheek) and lingual (tongue) surfaces of the teeth have different salivary flow, and tongue movement on the lingual side effects colonization. The occlusal surfaces (biting surface) and the surfaces between the teeth (approximal surfaces) are more difficult to clean which makes them more susceptible to microbial growth (Marsh and Martin, 2009). Saliva is produced by three major glands (parotid, submandibular and sublingual) and a number of minor glands. Saliva contains bicarbonate to facilitate buffering; calcium, and phosphate for the integrity of minerals in the tooth; and proteins. Salivary proteins range in concentration and comprise proline-rich proteins, amylase, statherin, histatin, mucins, lysozyme, lactoferrin, peroxidase, and secretory IgA (Dodds et al., 2005; Oppenheim et al., 1988; Slomiany et al., 1996). The saliva lubricates the oral cavity, which is necessary for protection, speech, taste, chewing, and swallowing (Mandel, 1987). The oral cavity is covered with saliva, and within the salivary layer, a salivary pellicle is formed. The acquired enamel pellicle, with so far 130 identified proteins, covers the teeth in the oral cavity (Siqueira et al., 2007). The pellicle contains adsorbed salivary glycoproteins, (e.g. gp340 (Loimaranta et al., 2005)) phosphoproteins, lipids (Al-Hashimi and Levine, 1989; Slomiany et al., 1986), and other molecules. These structures can be used for attachment and subsequent colonization by microorganisms of the oral cavity. Hannig showed that the formation of the pellicle is initiated immediately at the contact of a surface to saliva and the thickness of the pellicle increases over time. After 24 hours the teeth’s lingual surfaces is covered by a 100-200 nm thick pellicle, and a 1000-1300 nm pellicle on the buccal surfaces (Hannig, 1999). Gingival crevicular fluid (GCF) is a highly nutrient plasma exudate present in the gingival crevice.
GCF can be used as nutrients for resident bacteria, and contains members of the immune defense (for example immunoglobulins and leucocytes). The flow of GCF is slow in healthy sites, but increases in gingivitis and periodontitis (Marcotte and Lavoie, 1998).

**Biofilms and dental plaque**
The oral cavity is home to a diverse range of microbial life; Gram-positive and Gram-negative bacteria, fungi, mycoplasma, viruses, and protozoa. *Streptococcus*, *Actinomyces*, *Veillonella*, and *Neisseria* are bacteria commonly isolated from early dental plaque (Jenkinson and Lamont, 2005a; Li et al., 2004). The initial contacts with host molecules during colonization are weak (hydrophobic, van der Waals, and electrostatic interactions). If the environment is suitable the interaction becomes stronger and specific through bacterial adhesins (Nikolaev and Plakunov, 2007). The biofilm continues to grow into a mature biofilm with different species and bacterial exopolysaccharides (EPS) being incorporated (Branda et al., 2005). Streptococcal and *Actinomyces* species are Gram-positive commensal, early colonizers of the oral cavity that adhere to proteins, peptides or other molecules adsorbed in the salivary pellicle. The early colonizers can alter the environment by displaying new attachment sites for subsequent microorganisms, changing the pH, metabolizing oxygen and thereby changing the redox potential and by producing metabolites that can be used as nutrients by other microorganisms (Jenkinson and Lamont, 2005b; Marsh, 2006). Late colonizers attaches to the new surfaces available to them through the surfaces of resident bacteria. The majority of the bacteria in the early oral biofilm are commensal, but as the biofilm grows, more pathogenic bacteria accumulate. If dietary conditions are met (high sugar content in the diet), the mutans streptococci group and lactobacilli can thrive resulting in acidic metabolic products lowering the local pH causing enamel destruction i.e. caries (Loesche, 1986; Marsh, 2006). The oral biofilm on teeth is known as dental plaque, and consists of an array of microorganisms in a three dimensional (3D) structure with EPS, channels, and cavities. In a biofilm, the properties of the different species are less important; with the properties as a community instead becoming essential. Since bacteria in a community interact with each other in ways not yet fully understood, and the biofilm itself provides shelter from the environment, the planktonic pure culture format commonly implemented in laboratories results in that approximately 50% of the oral bacteria cannot be cultured. (Wade, 2002). In order to get a better understanding of how microorganisms in dental plaque collaborate, studies of pure cultures may need to be bypassed for the advances of other growth techniques.
**Gram-positive bacteria**

The cell envelope (figure 1) of Gram-positive bacteria contains a **cytoplasmic membrane** (CM) made up of a lipid bilayer, transport proteins, ion pumps, and enzymes. The CM is responsible for electron transport and energy production in the form of ATP (Murray et al., 2002). Outside the CM the **cell wall** (CW) is located, which consists of a thick (30-100 nm) **peptidoglycan** layer. Peptidoglycan consists of polysaccharide chains, made up of repeating units of N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) and the polysaccharide chains are cross-linked (Ghuysen et al., 1965). Teichoic acids (TA) and Lipoteichoic acids (LTA) are present in the peptidoglycan layer and are both polymers of modified ribose or glycerol, linked by phosphate moieties (Armstrong et al., 1958). TA is attached to peptidoglycan in the cell wall, and LTA is attached to the lipid bilayer in the CM (Silhavy et al., 2010). The surfaces of some Gram-positive bacteria are covered by a repeating unit of a (glyco) protein with a molecular mass of 40-200 kDa. This self-assembled crystalline **surface layer** (S-layer) is 5–25 nm thick with pores 2-8 nm in size, and is non-covalently linked to other neighboring molecules and to the cell wall. The S-layer gives the bacterium stability, acts like a sieve, and protects the bacterium from the host immune defense system, however it is usually lost during prolonged cultivation (Sara, 2001). Some Gram-positive bacteria have a **capsule** composed of extracellular polysaccharides as the outermost layer covering also the S-layer when present (Mesnage et al., 1998).

![Figure 1: The Gram-positive envelope consist of; the CM (lipid bilayer) with membrane proteins (pink); the cell wall with repeating units of GlcNAc (blue) and MurNAc (purple) with cross-links (black lines), TA (orange) and LTA (light green); the S-layer (grey pentagons), and the capsular polysaccharide (green ovals). Parentheses around the S-layer and capsule indicate that they are not always present. Adapted from Silhavy et al. (2010).](image)

**Stabilization of Gram-positive surface proteins**

Proteins are stabilized in a number of ways (Kauzmann, 1959). The **hydrophobic effect** is an important driving force when it comes to protein
folding. In short, hydrophobic side chains fold into the interior of the protein while polar and charged side chains form the protein surface. Strong **ionic interactions** (salt bridges) between residues with opposing charges will contribute to the stabilization of proteins when present. **Hydrogen bonds** in the protein interior or exterior between side chains, backbone, or water molecules help to stabilize the protein. Intrinsic **metal ions** (commonly iron, calcium, magnesium, or zinc) bind side chains or main chains and participate both in function and stabilization of the protein. **van der Waals forces** are transient forces that are weak but numerous, adding stability to the proteins. Protein can also be stabilized by covalent bonds; **disulfide bonds** are covalent bonds formed spontaneously between the sulphur atoms of two adjacent cysteine residues during protein folding. Intramolecular **isopeptide bonds** are covalent bonds that also can be spontaneously formed between the nitrogen of a lysine side chain and the gamma carbon of an asparagine/aspartic acid side chain. A catalytically essential glutamic- or aspartic acid must be present in the vicinity for the isopeptide bond to form, contributing with hydrogen bonds to the isopeptide bond (Kang and Baker, 2011; Kang et al., 2007) (figure 2). Isopeptide bonds are quite common in pili or other elongated Gram-positive surface proteins.

![Figure 2: Isopeptide formation. A) A prerequisite for spontaneous intramolecular isopeptide bond formation is close proximity between a lysine, and an asparagine (or aspartic acid). A glutamic acid (or aspartic acid) must also be present to catalyze the bond formation. B) The formation of the covalent isopeptide bonds between the side chain nitrogen from lysine and the gamma carbon from asparagine result in the release of NH$_3$ (ammonia). The catalytically important glutamic acid hydrogen bonds (black dashed lines) to a lysine hydrogen atom, and an asparagine oxygen. Adapted from Kang and Baker (2011).](image-url)
Streptococci

Streptococci are a family of spherical shaped Gram-positive bacteria that grow in pairs or chains. Most species are facultative anaerobes and display partial hemolysis (alpha-hemolysis) on blood agar. Viridans streptococci is a collective term used for oral streptococci based on the green coloration surrounding colonies created by the partial hemolysis on blood agar (the latin word viridis meaning green) (Holman, 1916). To classify streptococcal species Lancefield created a method for grouping cell wall antigens based on surface carbohydrate composition; polysaccharides (group A, B, D, E, F, and G), teichoic acid (D, and N), or lipoteichoic acid (H) (Lancefield, 1933; Nobbs et al., 2009). To further classify streptococci 16s rRNA sequencing can be used, where streptococcal species are divided into six groups; the salivarius, mutans, mitis, and anginosus groups with members commonly isolated from the oral cavity; and the pyogenes, and bovis groups which are not considered oral streptococci (Nobbs et al., 2009). Oral streptococcal species are acidogenic (produces acids) and aciduric (tolerate acids) to some extent. During periods of excess dietary sugar streptococci are able to store sugars as intracellular polysaccharides (IPS) for use as an energy source during times of low dietary sugar supplies (Takahashi et al., 1991; van Houte et al., 1970). Streptococci express a number of surface proteins containing an LPxTG-motif in the C-terminus (described in coming sections); Streptococcus gordonii strain CH1 expresses 20 (Nobbs et al., 2009); Streptococcus mutans strain UA159 expresses 6 (Ajdic et al., 2002), and Streptococcus pyogenes expresses 13 (Ferretti et al., 2001), based on whole genome sequencing.

Streptococcus gordonii

S. gordonii is considered to be a commensal bacteria, meaning that the bacteria benefits by the relationship to the host, while the host is unaffected. This may be a truth with modification, as the colonization by S. gordonii in the oral cavity benefits the host by outcompeting possible pathogenic bacteria, and on the other hand displays surface molecules known to serve as attachment sites for pathogenic bacteria (Brooks et al., 1997). S. gordonii display an array of surface proteins, and can cause infective endocarditis when spread to the blood stream (Douglas et al., 1993). S. gordonii produces EPS from sucrose; which is incorporated into dental plaque (Branda et al., 2005) and are able to catabolize arginine leaving ammonia as a product, which raises local pH and is thus important for pH homeostasis (Burne and Marquis, 2000).

Streptococcus mutans

The acidogenic, and aciduric S. mutans is implicated in dental caries, and converts sugars to acidic products that are detrimental for the enamel
(Loesche, 1986). *S. mutans* are involved in plaque maturation by production of EPS from sucrose (Branda et al., 2005).

**Streptococcus pyogenes**

*S. pyogenes* is a human pathogen responsible for mild (pharyngitis and scarlet fever) to lethal (necrotizing fasciitis and streptococcal toxic shock syndrome) conditions (Cunningham, 2000), and is not specific to the oral cavity. The cells are 0.5-1.0 µm in diameter and are surrounded by a hyaluronic acid capsule; creating a barrier to the host immune system (Murray et al., 2002).

**Gram-positive surface proteins require passage through the cytoplasmic membrane**

The countless proteins produced by cells have different objectives; some fulfill their fate by remaining in the cytoplasm, while others are secreted into the extracellular matrix for destination elsewhere. A group of proteins are displayed at the cell wall, interacting with the environment. A majority of the secreted proteins from Gram-positive bacteria will carry an N-terminal signal peptide sequence destined for the *secretory* (Sec) **pathway** (van Wely et al., 2001). Newly synthesized proteins will be bound by a signal recognition particle (SRP) and transported to the CM (Keenan et al., 2001; Zanen et al., 2006). There the unfolded protein will be transported through the membrane by the aid of specific Sec proteins forming a CM channel (translocon). During secretion through the CM the signal peptide will be cleaved off by signal peptidase I. In Gram-negative bacteria surface proteins are folded in the periplasmic space. Since Gram-positive bacteria lack a periplasmic space, protein folding must take place elsewhere. A single microdomain in the *S. pyogenes* CM, termed ExPortal, has been identified. It corresponds to the location of the Sec translocon and may be of importance for protein biogenesis (Rosch and Caparon, 2004; Rosch and Caparon, 2005). Another way to cross the CM is by the *Twin-arginine transport* (Tat) pathway. Tat can transport folded proteins and is present in some Gram-positive bacteria, but is not yet so well characterized (Dilks et al., 2003).

**Gram-positive surface proteins**

The Gram-positive surface proteins can be divided into four groups based on structure, function, and the method of attachment (Desvaux et al., 2006). 1) **Transmembrane proteins** are amphipathic, with the hydrophobic part of the protein embedded in the lipid bilayer, and the hydrophilic part interacting with the cytoplasm and extracellular milieu. Transmembrane proteins can traverse the membrane once or several times. (Lodish H, 2000). 2) **Lipoproteins** with a lipobox (specific sequence in the signal peptide
containing a cysteine important for lipid attachment) will be attached to the CM. Lipoproteins are often involved in transport (Hutchings et al., 2009). 3) Cell surface proteins with an **LPxTG-motif** (or equivalent) are attached to the cell wall and will be described in more detail in the following sections. 4) **Cell wall binding proteins** are associated to the cell wall by non-covalent attachment to components (e.g. LTA) in the cell wall. They are commonly amidases, involved in cell wall breakdown for cell growth and contain a glycine-tryptophan (GW) repeat modules that associate with LTA (Scott and Barnett, 2006).

The rest of this thesis will focus on Gram-positive bacterial cell surface proteins containing an LPxTG-motif, facilitating cell wall attachment (table 1). This collection of proteins can be divided into three groups: pili, the antigen I/II (AgI/II) protein family, and microbial surface components recognizing adhesive matrix molecules (MSCRAMM).

**Cell wall attachment**

Proteins containing LPxTG-motifs will be translocated unfolded through the CM by the Sec pathway. The signal peptide is cleaved off and the remaining protein is retained at the CM by the C-terminal hydrophobic domain and the positively charged tail. The CM-bound housekeeping enzyme sortase A cleaves the LPxTG-motif between the T and G residues and a covalently linked intermediate is formed (Mazmanian et al., 1999; Navarre and Schneewind, 1994). A nucleophilic attack by the amino group (NH$_2$) from the CW precursor protein lipid II on the carbonyl (C=O) group of the acyl-enzyme intermediate (Ruzin et al., 2002; Ton-That et al., 2000) results in the protein-lipid II complex being incorporated into the CW (figure 3).
Figure 3: Sortase mediated cell wall attachment of surface proteins carrying an LPxTG-motif. 1) The signal peptide is cleaved off during passage through the CM. 2) The protein is retained at the CM by the hydrophobic side chains traversing the CM and the positively charged side chains in the C-terminal. A sortase cleaves the protein between the T and G residue. 3) A protein-sortase intermediate is created, 4) and a nucleophilic attack by Lipid II on the protein-sortase intermediate results in a protein-lipid II complex. 5) The protein is incorporated into the cell wall. Adapted from (Hendrickx et al., 2011; Scott and Barnett, 2006).

Pili (fimbriae)
Pili are elongated structures protruding 1-2 µm from the cell wall that are present on the cell surface of some bacteria (Ton-That and Schneewind, 2003). Pili on Gram-positive bacteria usually consist of a minor base pilin attached to the peptidoglycan in the cell wall, a major pilin polymerized into the backbone (shaft) (figure 4A) and a minor tip pilin, containing an adhesin domain, usually at the tip of the structure (figure 5A). Sortase Class C is pili specific and polymerizes identical pili protein through a lysine residue in the pili motif creating the elongated structure (Kang and Baker, 2012; Ton-That and Schneewind, 2003). The structures of some of the backbone pilins on Gram-positive bacteria have been solved by the use of X-ray crystallography and they show the backbone pilin subunits adopting β-sandwich folds with an immunoglobulin G-constant (IgG-C) fold or variations thereof, e.g. the DEv-IgG and IgG-rev fold (figure 6). S. pyogenes backbone pili Spy0128 consist of two domains which both adopt an IgG-rev fold, with isopeptide
bonds covalently linking adjacent β-strands of the same sheet (Kang and Baker, 2009; Kang et al., 2007). β-sandwiches with stabilizing isopeptide bonds is also seen in the structures of RrgB from *Streptococcus pneumoniae* (El Mortaji et al., 2012; Paterson and Baker, 2011; Spraggon et al., 2010), and BcpA from *Bacillus cereus* (Budzik et al., 2009). SpaA from *Corynebacterium diphtheriae* (Kang et al., 2009) and GBS80 from *Streptococcus agalactiae* (Vengadesan et al., 2011) consist not only of β-sandwiches and isopeptide bonds, but also stabilizing metal ions. Apart from all the above mentioned stabilizing factors, FimA (Mishra et al., 2011) and FimP (Persson et al., 2012) from *Actinomyces oris* also contain disulphide bonds. The tip pilin Spy0125 from *S. pyogenes* (Pointon et al., 2010), RrgA from *S. pneumoniae* (Izore et al., 2010) and GBS104 from *S. agalactiae* (Krishnan et al., 2013) both show stabilizing isopeptide bonds. The top domain of Spy0125 contains a thioester bond between a cysteine and glutamine residue relevant for adhesion to host molecules. RrgA and GBS104 consist of a top domain with a von Willebrand factor A-type domain (vWFA) where six α-helices form a barrel with a β-sheet in the center. The minor base pilin GBS52 from *S. agalactiae*, known to mediate adhesion to lung epithelial cells (Krishnan et al., 2007), and FctB from *S. pyogenes* (Linke et al., 2010) are both believed to be located at the base of the protein and is attached to the cell wall by housekeeping sortase A. The structure of GBS52 and FctB both consist of domains with IgG-rev folds but only GBS52 is stabilized by an isopeptide bond (Kang et al., 2007).

<table>
<thead>
<tr>
<th>A</th>
<th>N</th>
<th>CTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>N</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>N1</td>
<td>N2</td>
</tr>
</tbody>
</table>

**Figure 4:** Domain organization of different groups of surface adhesins with signal peptide in dark grey, and the putative adhesin region in red/orange.  
A) Pilin backbone subunit of Spy0125 from *S. pyogenes*; signal peptide, N-terminal domain, and a C-terminal region (CTR).  
B) Agl/II AspA from *S. pyogenes*; signal peptide, N-terminal domain, Alanine-rich domain, Variable domain, Proline-rich domain, C-terminal domain, and a cell wall anchor segment.  
C) MSCRAMM Sgo0707 from *S. gordonii*; signal peptide, non-repetitive N-terminal domain, two repeat regions, a small non-repetitive C-terminal region, and a cell wall anchor segment.
Figure 5: Models of cell wall anchored adhesins. A) Illustration of the S. pyogenes pili, B) the AgI/II protein, C) MSCRAMM Sgo0707 from S. gordonii. The backbone pilin in A is reduced in number of repeats.

Figure 6: Common β-sandwich folds in Gram-positive adhesins. A) The IgG constant fold. B) The DE-variant IgG (DEv-IgG) fold has additional modules between the D and E β-strands and can contain an isopeptide bond between β-strand A and F. C & D) The two variants of the IgG-rev fold has reversed order of the β-strands and can contain an isopeptide bond between β-strand A and G. Adapted from (Vengadesan and Narayana, 2011).
Antigen I/II

Antigen I/II (AgI/II) is a family of surface proteins expressed by basically all oral streptococci, and also some streptococci not specific to the oral cavity. AgI/II facilitate colonization by mediating adherence to an array of host molecules, such as the salivary agglutinin glycoprotein 340 (gp-340) (Loimaranta et al., 2005; Zhang et al., 2006), fibronectin (Petersen et al., 2002), and collagen (Heddle et al., 2003; Soell et al., 2010), as well as to other microorganisms, Porphyromonas gingivalis (Daep et al., 2006; Park et al., 2005), Candida albicans (Silverman et al., 2010) and Actionomyces (Jakubovics et al., 2005). The protein consist of approximately 1500 amino acids and have the following domain structure (figure 4B); an N-terminal domain (N), an Alanine-rich repeat region (A), a Variable domain (V), a Proline-rich repeat region (P), a C-terminal domain (C), followed by a cell wall anchor (CWA) segment with an LPxTG-motif, a hydrophobic domain, and finally a positively charged tail (Brady et al., 2010; Jenkinson and Demuth, 1997). Despite the central location of the V-domain in the primary sequence, the V-domain is presented at the tip of the protein with the flanking A- and P-domains intertwining, forming a stalk (Larson et al., 2010). The C- and the N-domains form the base of the protein at the cell surface (figure 5B). The elongated protein model has a length of > 50 nm and is stabilized in the C-domain by metal atoms (Ca²⁺) and isopeptide bonds.

The amino acid sequence identity between the full-length AgI/II proteins in different streptococcal species ranges from 80% between S. gordonii SspA and SspB, to only 27% between S. mutans SpaP and S. pyogenes AspA.

The first determined structure of the AgI/II protein family was the V-domain from the caries pathogen S. mutans (Troffer-Charlier et al., 2002). The V-domain of S. mutans is made up of a β-sandwich with eight β-strands in each sheet. The V-domain from the commensal S. gordonii was solved (Forsgren et al., 2009) and showed high structural similarity despite low sequence identity (26%) to the V-domain from S. mutans. A putative binding pocket with a bound metal ion is present in both structures. The C-domain of AgI/II proteins consists of three domains, C1, C2, and C3. The C2-3 domains of SspB from S. gordonii (Forsgren et al., 2010), SpaP from S. mutans (paper I) (Nylander et al., 2011), and AspA from S. pyogenes (paper IV), and C1-3 domains of SpaP from S. mutans (Larson et al., 2011) have been determined revealing a striking structural identity among the proteins. Each domain forms a β-sandwich with DEv-IgG fold, with isopeptide bonds covalently linking the β-sheets in the β-sandwich. All structures contained at least one stabilizing metal ion modeled as Ca²⁺. S. gordonii SspB binds the minor fimbriae Mfa1 on P. gingivalis but despite structural similarity, SpaP cannot bind Mfa1 (Brooks et al., 1997).
Microbial Surface Components Recognizing Adhesive Matrix Molecules (MSCRAMM)

The definitions of MSCRAMMs are microbial molecules located at the cell wall that recognize ligands (for example collagen, laminin, fibronectin, and fibrinogen) in the ECM with high affinity and specificity. The MSCRAMMs share a domain organization with an N-terminal signal peptide, followed by a non-repetitive sequence (putative adhesin domain). Next follows one or two separate repeat regions, a cell wall-spanning domain with an LPxTG-sorting motif, a hydrophobic membrane-spanning domain, and finally a positively charged C-terminal tail (Patti et al., 1994) (figure 4C). The repeat regions of MSCRAMMs are varied both in size and in number of repeats as well as in the number of different repeat regions present. Some fibrinogen-binding proteins (Staphylococcus aureus ClfA (Deivanayagam et al., 2002), Staph. aureus ClfB (Ganesh et al., 2011), and Staphylococcus epidermidis SdrG (Ponnuraj et al., 2003) have 200-300 serine-aspartic acid residue repeats in the C-terminal. The Serine-rich repeat (SRR) proteins GspB and Fap1 from S. gordonii and Streptococcus parasanguinis respectively, have an approximately 2000 residue long segment of SRRs. The structure of the non-repetitive N-terminal part, have been determined, and GspB consist of domains with predominately β-sandwiches. (Pyburn et al., 2011). The non-repetitive N-terminal domain of S. parasanguinis Fap1 can be divided into two domains, the Nα-domain, consisting of α-helices (Garnett et al., 2012) and the Nβ-domain, consisting of a β-sandwich (Ramboarina et al., 2010). S. gordonii Sgo0707 has two repeat regions, one with 84 amino acids repeated eight times, followed by an 88 amino acid sequence repeated five times (paper II) (Nylander et al., 2013). MSCRAMMs are an important virulence factor so the structure of these proteins can give vital information for possible drug design. β-sandwiches are the most common feature of the N-terminal putative adhesin domain in MSCRAMMs protein, especially of the DEv-IgG fold. Staph. aureus display the collagen binding CNA protein on the cell surface and structural studies on the N-terminal half (called A) reveled β-sandwiches of the DEv-IgG fold (Symersky et al., 1997). CNA-A binds collagen in a so called “hug model” by assuming an open conformation where collagen is able to bind in between N1 and N2 (Zong et al., 2005). The C-terminal half of the CNA protein, CNA-B (consist of three repeats, B1-3) displays the IgG-rev fold (Deivanayagam et al., 2000). ACE from Enterococcus faecalis (Liu et al., 2007) and the N-terminal part of RspB from the human pathogen Erysipelothrix rhusiopathiae shows two β-sandwiches with the DEv-IgG fold (Devi et al., 2012). β-sandwiches can be seen in S. pyogenes Epf (Linke et al., 2012), UafA from uropathogen Staphylococcus saprophyticus (Matsuoka et al., 2010), the N-terminal part of ACP (NTACP) from S. agalactiae (Auperin et al., 2005), and Spr1345-MucBD from S. pneumoniae (Du et al., 2011). The authors of the structures of all the
MSCRAMMs above do not mention the presence of any isopeptide bond, or they state that no isopeptide bonds were found. Only two structures seem to contain metal ions, ClfA and UafA.

The M1 protein from *S. pyogenes* and protein G from the Lancefield group G streptococci differ in structure to pili, AgI/II, and MSCRAMM mentioned in this thesis, as they lack the commonly found β-sandwich folds. The N-terminal part of the M1 protein forms an α-helical coiled coil (McNamara et al., 2008). The structures of the three domains (B1-B3) of protein G from the Lancefield group G streptococci (Achari et al., 1992; Derrick and Wigley, 1994; Gallagher et al., 1994) was determined individually and they all display a β-sheet and an α-helix that share high structural similarities between them, with a root-mean-square deviation (RMSD) of 0.4 Å, and sequence identity ranging from 88-98%.
**Table 1:** Structures of Gram-positive surface adhesins determined by X-ray crystallography.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Type of protein</th>
<th>Ligand</th>
<th>Fragment</th>
<th>PDB code</th>
<th>Reference</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Actinomyces oris</em></td>
<td>Major pilin (type-1)</td>
<td></td>
<td>FimP</td>
<td>3UXF</td>
<td>Persson 2012</td>
<td>IgG-like fold. Isopeptide bond, disulphide bond and Ca^{2+}.</td>
</tr>
<tr>
<td><em>S. agalactiae</em></td>
<td>Major pilin(type-1)</td>
<td></td>
<td>GBS80</td>
<td>3PF2</td>
<td>Vengadesan 2011</td>
<td>DEv-IgG and rev-IgG folds. Isopeptide bonds and Ca^{2+}. (α-chymotrypsin treated).</td>
</tr>
<tr>
<td><em>S. agalactiae</em></td>
<td>Minor (tip) pilin</td>
<td></td>
<td>GBS104</td>
<td>3TXA (N2-N3)</td>
<td>Krishnan 2013</td>
<td>β-sandwich and a vWFA -type domain.</td>
</tr>
<tr>
<td><em>S. agalactiae</em></td>
<td>Minor (base) pilin</td>
<td>Lung epithelial cells</td>
<td>GBS52</td>
<td>3PHS</td>
<td>Krishnan 2007</td>
<td>Rev-IgG folds and isopeptide bond.</td>
</tr>
<tr>
<td><em>S. pneumoniae</em></td>
<td>Minor (tip) pilin</td>
<td>Host cells, fibronectin, collagen, and laminin.</td>
<td>RrgA</td>
<td>2WW8</td>
<td>Izore 2010</td>
<td>β-sandwiches and a vWFA -type domain. MIDAS motif with Mg^{2+} and isopeptide bonds.</td>
</tr>
<tr>
<td>Organism</td>
<td>Type of protein</td>
<td>Ligand</td>
<td>Fragment</td>
<td>PDB code</td>
<td>Reference</td>
<td>Comments</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------</td>
<td>--------------</td>
<td>----------</td>
<td>----------</td>
<td>-----------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td><em>S. mutans</em></td>
<td>AgI/II</td>
<td>AgI/II</td>
<td>V</td>
<td>1JMM</td>
<td>Troffer-Charlier 2002</td>
<td>β-sandwich and a metal ion.</td>
</tr>
<tr>
<td><em>S. gordonii</em></td>
<td>AgI/II</td>
<td>SspB-V</td>
<td></td>
<td>2WD6</td>
<td>Forsgren 2009</td>
<td>β-sandwich and Ca²⁺.</td>
</tr>
<tr>
<td><em>S. gordonii</em></td>
<td>AgI/II</td>
<td><em>P. gingivalis</em></td>
<td>SspB-C&lt;sub&gt;2&lt;/sub&gt;-</td>
<td>2WZA</td>
<td>Forsgren 2010</td>
<td>DEv-IgG folds. Isopeptide bonds and Ca²⁺.</td>
</tr>
<tr>
<td><em>S. pyogenes</em></td>
<td>AgI/II</td>
<td>AspA-C&lt;sub&gt;2&lt;/sub&gt;-</td>
<td></td>
<td></td>
<td>Hall TBP</td>
<td>DEv-IgG folds. Isopeptide bonds and Ca²⁺.</td>
</tr>
<tr>
<td><em>S. gordonii</em></td>
<td>AgI/II</td>
<td>P. gingivalis</td>
<td>SspB-C&lt;sub&gt;2&lt;/sub&gt;-</td>
<td>2WZA</td>
<td>Forsgren 2010</td>
<td>DEv-IgG folds. Isopeptide bonds and Ca²⁺.</td>
</tr>
<tr>
<td><em>S. pyogenes</em></td>
<td>AgI/II</td>
<td>AspA-C&lt;sub&gt;2&lt;/sub&gt;-</td>
<td></td>
<td></td>
<td>Hall TBP</td>
<td>DEv-IgG folds. Isopeptide bonds and Ca²⁺.</td>
</tr>
<tr>
<td><em>S. mutans</em></td>
<td>AgI/II</td>
<td>P. gingivalis</td>
<td>SspB-C&lt;sub&gt;2&lt;/sub&gt;-</td>
<td>2WZA</td>
<td>Forsgren 2010</td>
<td>DEv-IgG folds. Isopeptide bonds and Ca²⁺.</td>
</tr>
<tr>
<td>Enterococcus</td>
<td>MSCRAMM</td>
<td>Collagen</td>
<td>ACE</td>
<td></td>
<td>Ponnuraj 2007</td>
<td>DEv-IgG fold on N1 and N2.</td>
</tr>
<tr>
<td>faecalis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Liu 2007</td>
<td>DEv-IgG folds.</td>
</tr>
<tr>
<td>Erysipelothrix</td>
<td>MSCRAMM</td>
<td>Collagen</td>
<td>RspB-N</td>
<td>3V10</td>
<td>Devi 2012</td>
<td>DEv-IgG folds.</td>
</tr>
<tr>
<td>rhusiopathiae</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staph aureus</td>
<td>MSCRAMM</td>
<td>Collagen</td>
<td>CnaA</td>
<td>1AMX (N2)</td>
<td>Symersky 1997</td>
<td>DEv-IgG fold on N1 and N2.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2F6A(collegen)</td>
<td>Zong 2005</td>
<td>Collagen-hug.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1D2P (B1-B2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staph aureus</td>
<td>MSCRAMM</td>
<td>Fibrinogen</td>
<td>ClfA (N2-N3 domains)</td>
<td>1N67</td>
<td>Deivanayagam 2002, Ganesh 2008</td>
<td>DEv-IgG fold on N2 and N3, metal ions.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2VR3 (N2-N3 + fibrinogen)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staph aureus</td>
<td>MSCRAMM</td>
<td>Fibrinogen, Cytokeratin 10, Dermokine.</td>
<td>ClfB (N2-N3 domains)</td>
<td>4F27 (Fibrinogen)</td>
<td>Xiang 2012, Ganesh 2011</td>
<td>DEv-IgG folds.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3AU0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staph epidermidis</td>
<td>MSCRAMM</td>
<td>Fibrinogen</td>
<td>SdrG (N2-N3 domains)</td>
<td>1R19</td>
<td>Ponnuraj 2003, Bowden 2008</td>
<td>DEv-IgG folds.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2RAL (Mutants)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. agalactiae</td>
<td>MSCRAMM</td>
<td>Heparin.</td>
<td>NtACP</td>
<td>1YWM</td>
<td>Auperin 2005</td>
<td>β-sandwich and a three α-helix bundle.</td>
</tr>
</tbody>
</table>

15
<table>
<thead>
<tr>
<th>Organism</th>
<th>Type of protein</th>
<th>Ligand</th>
<th>Fragment</th>
<th>PDB code</th>
<th>Reference</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. gordonii</em></td>
<td>MSCRAMM</td>
<td>Collagen, Keratinocytes.</td>
<td>Sgo0707-N</td>
<td>41GB</td>
<td>Nylander 2013</td>
<td>N1 consist of a β-sandwich, N2 of a DEv-IgG fold.</td>
</tr>
<tr>
<td><em>S. pyogenes</em></td>
<td>MSCRAMM</td>
<td>Human epithelial cells</td>
<td>Epf</td>
<td>4ES8</td>
<td>Linke 2012</td>
<td>β-sandwiches.</td>
</tr>
<tr>
<td><em>S. pyogenes</em></td>
<td>MSCRAMM</td>
<td>Fibrinogen</td>
<td>M1</td>
<td>2OTO</td>
<td>McNamara 2008</td>
<td>α-helices forming a coiled coil.</td>
</tr>
<tr>
<td><em>S. pneumoniae</em></td>
<td>MSCRAMM</td>
<td>Mucin</td>
<td>Spr1345-MucBD</td>
<td>3NZ3</td>
<td>Du 2011</td>
<td>IgG-like fold.</td>
</tr>
<tr>
<td><em>Lancefield group G streptococci</em></td>
<td>MSCRAMM</td>
<td>Serum albumin and Human IgG</td>
<td>Protein G</td>
<td>1PGA/B (B1) 1PGX (B2) 1IGD (B3)</td>
<td>Gallagher 1994 Achari 1992 Derrick and Wigley 1994</td>
<td>β-sheet and an α-helix.</td>
</tr>
<tr>
<td><em>S. gordonii</em></td>
<td>SRR/MSCRAMM</td>
<td>Human platelets</td>
<td>GspB</td>
<td>3QC5</td>
<td>Pyburn 2011</td>
<td>β-sandwiches and a metal ion.</td>
</tr>
<tr>
<td><em>S. parasanguinis</em></td>
<td>SRR/MSCRAMM</td>
<td></td>
<td>Fap1</td>
<td>3RGU (NRα) 2X12 (NRβ)</td>
<td>Garnett 2012 Ramboarina, 2010</td>
<td>NRα consist of α-helices and NRβ of a β-sandwich.</td>
</tr>
</tbody>
</table>
Materials and methods

Cloning, protein expression and purification
The gene of interest is amplified using PCR with gene specific primers. Ligation of the DNA-fragment into the lac operon of the expression vector is done after the digestion of the vector and PCR product with appropriate restriction enzymes. The vector contains the DNA code for six residues of histidine, facilitating protein purification. Antibiotic resistance will be relayed to the correct constructs of gene and vector and can thus be selected on agar plates containing the correct antibiotic. Protein expression is usually done in *Escherichia coli* expression strains and is induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) which activates transcription of the lac operon. The standard protein purification process implemented in this thesis was done in two steps, first by immobilized metal affinity chromatography (IMAC) where histidine residues in the his-tag of the protein adheres to the nickel-matrix, facilitating purification through affinity, secondly by size exclusion chromatography where the proteins are fractioned according to size.

From crystal to structure
The two step purification protocol usually renders the protein of high enough purity (95%) to be set up for crystallization trails; if the purity is not reached additional purification steps can be performed, for example ion exchange chromatography. The protein does not only need to be pure, it also must be homogenous (all molecules must be identical), and monodisperse (exist as monomers, dimers, or larger oligomers, not a mixture) in order to pack into crystals.

Vapour diffusion is a common technique for obtaining crystals, where a drop of protein is mixed with the well solution (mother liquor) and placed either as a hanging drop, or sitting drop in air contact to the well solution. As the well solution is diluted in the drop by the addition of the protein, equilibrium will be reached through vapour diffusion of water from the drop to the well solution. In order to initiate crystallization the protein must be supersaturated and nucleation must occur. The growth of the crystal will continue until the protein in the drop is depleted or if defects, or mechanical stress renders crystal growth impossible. There are commercial crystallization kits available to test the protein against an array of different buffers, pH, precipitants, and additives.
Crystallography is a method to get a 3D image of a protein by subjecting it to X-rays. The content of the protein crystal (the unit cell and the protein electrons) will diffract the X-ray beam and the structure of the protein can be calculated. Acquired crystals are mounted on loops and vitrified in a stream of liquid nitrogen to limit radiation damage during data collection. Data collection can be done at a synchrotron (possibility to vary the wavelength and a sharper beam) or a home source (set wavelength). The mounted crystal is subjected to X-rays, and the crystal is rotated in increments of usually 1.0 degree during which an image of the diffraction pattern is collected. Most of the X-rays will go straight through the crystal, but approximately 5% of the X-rays will be diffracted by the crystal. The diffracted beam will give rise to spots on the detector and the positions and intensity of the spots will be recorded. The amplitude of the diffracted wave can be calculated from the intensities, but its phase is not recorded. This is the so called phase problem of crystallography, and it must be solved in order to obtain an electron density map. To solve the phase problem a number of techniques can be implemented; multiple isomorphous replacement (MIR), multiple-wavelength anomalous dispersion (MAD), single-wavelength anomalous dispersion (SAD) and molecular replacement (MR). To use MIR data sets are collected of the native protein crystal and also of crystals soaked in liquid containing heavy atoms. The native data set is subtracted from the heavy atom soaked data set, resulting in the intensities contributed by the heavy atom, and the location of the heavy atom can be identified. This will enable the phases of all intensities to be determined. MAD and SAD require the introduction of heavy atoms (for example selenium) into the protein, usually by substituting methionine with selenomethionine (SeMet) during protein synthesis. The position of the heavy atoms can be identified and the phases can be determined. If a homologous protein structure is known usually with more than 30% sequence identity, MR can be used to get initial estimates of the phases.

The bottleneck of X-ray crystallography is the requirement of diffracting protein crystals. The ultimate method is to clone, express, purify, and crystallize the full-length protein, but this is often only possible for smaller proteins. For Gram-positive surface proteins this is usually difficult due to the large size of the full-length protein. The large proteins can therefore be divided into domains in order to reduce the size of the protein. The design of constructs can be done by mimicking previously used domain borders in already determined crystal structures and also by sequence analysis. During work with AspA-AVPC from *S. pyogenes* N-terminal sequencing and matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) were used to analyze the fragments. N-terminal sequencing is done by transferring the proteins separated by sodium dodecyl
sulfate polyacrylamide gel electrophoresis (SDS-PAGE) to a PVDF membrane. The protein is cleaved repeatedly at the N-terminal one amino acid at the time and the leaving amino acid is determined. Since only amino acids in the N-terminal part of the proteins are determined, MALDI-TOF MS is preformed in order to get a more complete view of the fragments. In-gel trypsin digestion is carried out and the peptides are subjected to a laser beam that ionizes the fragments and the time it takes for the fragment to reach the detector is dependent on the mass of the fragment.
**Aim of thesis**

The purpose of this thesis was to determine the 3D structure of streptococcal surface adhesins using X-ray crystallography techniques, and also to conduct functional studies in order to elucidate possible binding partners to the structurally determined adhesins.

**Specific aims**

Paper I  
Determine and study the structure of SpaP-C\textsubscript{2-3} from the caries pathogen *S. mutans*, and compare the structure to the commensal SspB from *S. gordonii*.

Paper II  
Determine the structure and perform functional studies of the MSCRAMM Sgo0707 from *S. gordonii*.

Paper III  
Express and purify AspA-AVPC from *S. pyogenes*, and elucidate fragmentation behavior.

Paper IV  
Determine and study the structure of AspA-C\textsubscript{2-3} from *S. pyogenes*, and compare the structure to the homologous proteins SpaP from *S. mutans* and SspB from *S. gordonii*. 
**Result and discussion**

**Structures of the C-terminal domain of the antigen I/II protein family (paper I & IV)**

During the last decade, 3D structures of domains of AgI/II proteins from oral streptococci have been determined using X-ray crystallography. In order to compare the known structure of the commensal *S. gordonii* SspB to the caries implicated *S. mutans* SpaP (paper I) and the human pathogen *S. pyogenes* AspA (paper IV) the structures of the truncated C-terminals from these two proteins were determined using X-ray crystallography.

**The structures of SpaP-C\textsubscript{1136-1489} and AspA-C\textsubscript{971-1306}**

The protein crystals of SpaP-C\textsubscript{1136-1489} belonged to the $P_{21}2_12_1$ space group and contained six copies in the asymmetric unit. The monomeric structure consisted of 354 residues and was solved by MR and refined to a resolution of 2.2 Å. The AspA-C\textsubscript{971-1306} protein crystals also belonged to the $P_{21}2_12_1$ space group and contained two molecules in the asymmetric unit. The structure was solved by MR and refined to a resolution of 1.8 Å, and consisted of 336 residues. The 3D structures of SpaP-C\textsubscript{1136-1489} and AspA-C\textsubscript{971-1306} consist of two domains; C2 and C3 (figure 7A and B). In C2 the β-strands form anti parallel sheets from five and six β-strands resulting in a β-sandwich with a DEv-IgG fold. Two short α-helices are located on top of the β-sandwich forming the apex of the protein. An α-helix is present perpendicular to the β-sandwich facing the surroundings. The C3 domain contains a DEv-IgG folded β-sandwich constructed from two antiparallel sheets, each cosist of five strands.

**Stabilization of SpaP-C\textsubscript{1136-1489} and AspA-C\textsubscript{971-1306}**

SpaP-C\textsubscript{1136-1489} and AspA-C\textsubscript{971-1306} are stabilized in both domains by isopeptide bonds (figure 7A and B). All four isopeptide bonds in the two structures are formed between a lysine side chain in the first β-strand in one sheet, and an asparagine side chain in the last β-strand of the other sheet, resulting in covalently linking the two β-sheets in the β-sandwich, adding stability to the β-sandwich and the protein as a whole. The essential aspartic acid is located in strand C, from the same sheet as the asparagine residue is located. To further stabilize the SpaP protein, two metal ions (modelled as calcium ions) are bound in the vicinity of the isopeptide bonds in both domains (figure 7A). The AspA protein only contains one metal ion, located in the C2 domain close to the isopeptide bond (figure 7B).
Figure 7: A) The crystal structures of S. mutans SpaP-C_{2-3} and B) S. pyogenes AspA-C_{2-3}. Calcium ions and isopeptide bonds are represented by grey sphere and black bonds, respectively. C) Superposition of SpaP (blue), AspA (green), and SspB from S. gordonii (red).

The structures of SpaP-C_{1136-1489} and AspA-C_{971-1306} put in perspective
The structures of SpaP-C_{1136-1489} and AspA-C_{971-1306} are very similar with an RMSD of 1.84 Å on 332 aligned C\(_\alpha\) and a sequence identity of 36%. Also S. gordonii SspB-C are very similar to S. mutans SpaP-C_{1136-1489} with an RMSD of 1.2 of 337 aligned C\(_\alpha\), and a sequence identity of 66% (figure 8). As the RMSD suggest, the structures are very similar as shown by superposition of the three structures (figure 7C). The isopeptide bonds and calcium ions are located in identical positions. The SspB Adherence Region (BAR) is a segment (1167 to 1250) in S. gordonii SspB involved in adherence to P. gingivalis fimbria Mfa1 (Brooks et al., 1997). Studies with a synthetic peptide of 27 residues from BAR in SspB (1167-1193) and the corresponding sequence in SpaP (1242-1268) resulted in adherence between the SspB peptide and Mfa1, but no adherence to SpaP, despite a 59% primary sequence identity in that area. The different residues result in different electrostatic surface charge of the protein in this area. Mutation of two amino acids (1182 N to G, and 1185 V to P) in the SspB peptide to mimic the amino acids in SpaP, showed no adherence. (Demuth et al., 2001). This implies that this part of the BAR is involved in species specific interactions. The two corresponding amino acids in AspA are identical to SpaP, so adherence to Mfa1 is not to be expected.
<table>
<thead>
<tr>
<th></th>
<th>AspA</th>
<th>SpaP</th>
<th>SspB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NLIQPTKIVDSDKQSIDGKSVLNKPTNLTVAGSNYYELTDLQYKNGRSADTIQKGYVYDDYP</td>
<td>NYIKPTKVNENGVVDDKTVLASGTNYYELTDLQYKNDRSADTIQKGYVYDDYP</td>
<td>NYIKPTKVNENGVVDDKTVLASGTNYYELTDLQYKNDRSADTIQKGYVYDDYP</td>
</tr>
<tr>
<td></td>
<td>LYPEGAVVENKNEKSLIDGKSVLDKTSYVEKLGSKYELTDLQYKGDSSKEAIQNGFYVYDDYP</td>
<td>LYPEGAVVENKNEKSLIDGKSVLDKTSYVEKLGSKYELTDLQYKGDSSKEAIQNGFYVYDDYP</td>
<td>LYPEGAVVENKNEKSLIDGKSVLDKTSYVEKLGSKYELTDLQYKGDSSKEAIQNGFYVYDDYP</td>
</tr>
<tr>
<td></td>
<td>DEAIDGHSLVVNSIKAANGDDVTNLLEMRHVLSQDTLDDKLAKSIGSPVGEFYMVW</td>
<td>DEAIDGHSLVVNSIKAANGDDVTNLLEMRHVLSQDTLDDKLAKSIGSPVGEFYMVW</td>
<td>DEAIDGHSLVVNSIKAANGDDVTNLLEMRHVLSQDTLDDKLAKSIGSPVGEFYMVW</td>
</tr>
<tr>
<td></td>
<td>AKDPAAFYKAYVQKGLDITYNLSSFKLQDF</td>
<td>AKDPAAFYKAYVQKGLDITYNLSSFKLQDF</td>
<td>AKDPAAFYKAYVQKGLDITYNLSSFKLQDF</td>
</tr>
<tr>
<td></td>
<td>ELTVHKDVF---DKEGQINSAGTVKGDDEVTRLECGWVPTRYNGYDLYEYKFVDQLHT</td>
<td>ELTVHKDVF---DKEGQINSAGTVKGDDEVTRLECGWVPTRYNGYDLYEYKFVDQLHT</td>
<td>ELTVHKDVF---DKEGQINSAGTVKGDDEVTRLECGWVPTRYNGYDLYEYKFVDQLHT</td>
</tr>
<tr>
<td></td>
<td>HDLYQKDKVLATVDTSDGSVITGTDLAKYTEVNYKETYHELAKQDFLOAKV</td>
<td>HDLYQKDKVLATVDTSDGSVITGTDLAKYTEVNYKETYHELAKQDFLOAKV</td>
<td>HDLYQKDKVLATVDTSDGSVITGTDLAKYTEVNYKETYHELAKQDFLOAKV</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>SAFQAETLYQMRRIAGTFENTYVNTVNVKAYASNTVRTT</td>
<td>S</td>
</tr>
</tbody>
</table>

**Figure 8:** Sequence alignment of the AgI/II C2-3’-domains, with the boxed 27 amino acid sequence involved in *S. gordonii* SspB adherence to *P. gingivalis* fimbriae Mfa1. Stars indicate identical residues in all three sequences.
Structural and functional analysis of the N-terminal domain of the *Streptococcus gordonii* adhesin Sgo0707 (paper II)

*S. gordonii* expresses many different proteins on the cell surface with which it interacts with the environment. Sgo0707 was identified by Davies and co-workers as a putative surface adhesin (Davies et al., 2009) and structural and functional studies were implemented in order to determine the structure and to elucidate possible binding partners.

**Binding studies of Sgo0707**

A *S. gordonii* Sgo0707 deletion mutant (ΔSgo0707) was constructed in order to elucidate structures that interact with Sgo0707. Whole saliva, serum, collagen type-1, and keratinocytes were coated on an ibiTreat µ-slide VI flow-cell. The wild-type and ΔSgo0707 were added to the coated flow-cells and adhered cells were stained and visualized with confocal laser scanning microscopy (CLSM). The ΔSgo0707 showed a 30% reduction in adherence to collagen type-1 and keratinocytes compared to the wild-type. Saliva and serum-coated surfaces showed no differences in adherence between wild-type and the deletion mutant ΔSgo0707. These data suggest that Sgo0707 may be involved in adherence of *S. gordonii* to collagen type-1, and keratinocytes. The N-terminal domain of the Sgo0707 protein was shown to be involved in adherence, by adding different concentrations of recombinant Sgo0707-N protein to flow-cells covered by adhered collagen type-1, adherence of wild-type cells was decreased.

**The structure of Sgo0707-N**

Protein crystals of Sgo0707-N<sub>36-456</sub> belonged to the I222 space group and contained four molecules in the asymmetric unit. The structure was solved with SAD, using SeMet labeled protein, and was refined to a resolution of 2.1 Å. The structure of the N-terminal part of Sgo0707 can be divided into two distinct domains, N1 and N2 (figure 9A). The N<sub>136-311</sub> domain consists mainly of β-strands forming a β-sandwich fold, and contains a putative binding cleft of 333 Å<sup>3</sup> (figure 9B and C), as determined by CASTp server, with mainly negatively surface potential. According to a DALI server search, the variable domain of AgI/II from *S. gordonii* (Forsgren et al., 2009), and *S. mutans* (Troffer-Charlier et al., 2002) revealed to be the closest structural relatives, with an RMSD of 3.7 Å, and approximately 10% sequence identity. The N<sub>2312-456</sub> consists of a β-sandwich with a DEv-IgG-fold with the *B. cereus* backbone pilin BcpA (Budzik et al., 2009) as the closest structural relative with an RMSD of 2.3 Å, and 15% sequence identity. The location of collagen type-1, and keratinocytes adherence to Sgo0707-N is not known, it is however possible that the putative binding pocket is involved. A collagen type-1 fragment was used in web-based docking programs to elucidate
possible locations for adherence on the Sgo0707-N crystal structure. The results indicate the putative binding pocket and the interface between N1 and N2 (figure 9D) as possible collagen type-1 binding sites. The N-terminal part of Sgo0707 is presented at the tip of a putative elongated protein with the shaft constructed of repeat domains, whose structural and functional role is unknown.

**Figure 9:** **A)** The structure of Sgo0707-N as a ribbon presentation, with the N1 domain in pink, N2 in purple, and a putative binding pocket (X). **B)** The N1 domain with negatively charged, polar, and aromatic residues involved in the putative binding pocket, shown as green stick models. **C)** View of the binding pocket tilted 90° into the paper plane. **D)** The second possible binding site of collagen type-1 (X), view rotated 90° from figure A.
Protein expression and purification of *Streptococcus pyogenes* adhesin AspA (paper III)

The human pathogen *S. pyogenes* causes an array of conditions, ranging from mild to severe. Although *S. pyogenes* is not specific for the oral cavity, it does express AspA, a member of the AgI/II protein family.

Cloning, protein expression and purification of AspA

The *M28_spy1325* gene from *S. pyogenes* was amplified almost in full length (AspA-AVPC) using PCR. Protein expression was done in *E. coli* BL21 (DE3) cells and purification was carried out using IMAC followed by size exclusion chromatography. Purification resulted in soluble protein.

Auto fragmentation of AspA-AVPC

During laboratory work with AspA-AVPC, it was noticed that the protein fragmented as observed with *S. mutans* AgI/II proteins in earlier studies (Kelly et al., 1989; Russell et al., 1980). The fragmentation of AspA-AVPC occurred at both 4°C and 22°C, but was accelerated in warmer temperature and resulted in two fragments, approximately 50 and 75 kDa (figure 10). Cysteine and serine protease inhibitors were used during protein purification which makes it unlikely that proteases of those families are responsible for the fragmentation. Metalloproteases are however not inhibited during purification. Studies made by Maddocks and co-workers indicated that AspA-V binds a zinc ion, possibly through histidine residues as observed in certain streptococcal zinc metalloprotease (Hase and Finkelstein, 1993; Maddocks et al., 2011). The biological relevance for the fragmentation is unclear. One possible reason may be to relocate, if the environment is no longer optimal for the bacteria. Or it may be a consequence of overexpression, and may not occur under normal biological conditions. Conformational changes in the cell wall attached protein compared to the soluble protein may also be an explanation.
Figure 10: Auto fragmentation of AspA-AVPC. A) After incubation for four days. B) After incubation for 15 days. C) Chromatogram from purification of fragmented AspA-AVPC showing two peaks. D) Size exclusion chromatography fractions of purified fragmented AspA-AVPC.

N-terminal sequencing and MALDI-TOF MS
The fragments were N-terminally sequenced and subjected to MALDI-TOF MS. The result showed that the 55 kDa fragment contains the C1-3-domain, and the 67 kDa fragment starts in the His-tag in the N-terminal part of the protein and ends in the P-domain.

Inhibition of auto-fragmentation
The protein stability in the presence and absence of EDTA during purification was analyzed (figure 11). The conclusion of these studies was that auto fragmentation was inhibited in the presence of 0.5 mM EDTA. In order to find out if the addition of a metal ion could accelerate the fragmentation, Zn$^{2+}$ was added in a range of concentrations. No increase in fragmentation was seen, however 0.5 mM Zn$^{2+}$ appears to inhibit fragmentation (figure 11A), but how the inhibition is accomplished is not known, possibly by blocking an active site or causing conformational changes to the protein. These results imply that EDTA sequesters some unknown metal ion crucial for the auto fragmentation, or that EDTA itself physically blocks fragmentation.
Figure 11: AspA-AVPC protein purified with or without EDTA and supplemented with ZnCl₂, followed by incubation in 22°C for two days. A) AspA-AVPC purified without EDTA, with ZnCl₂ added at different concentrations. B) AspA-AVPC supplemented with EDTA at different concentrations, and in the presence or absence of ZnCl₂.
Conclusion

The determination of the C$_{2-3}$ domain of SpaP from *S. mutans* (paper I), and AspA from *S. pyogenes* (paper IV) using X-ray crystallography revealed that the two shared highly similar protein structures, despite the low primary sequence identity. They were also very similar to the already known structure of SspB from *S. gordonii*. All three structures consist of two domains, C2 and C3 which consist of DEv-IgG folds in each domain. SpaP and AspA both have stabilizing isopeptide bonds in identical positions to SspB, and calcium ions are present to further stabilize the proteins. Despite the structural similarity, the amino acid sequence differ in such a way that SpaP cannot be bound by *P. gingivalis* fimbria Mfa1 but SspB can. The structure of MSCRAMM *S. gordonii* Sgo0707-N (paper II) revealed two domains, N1 and N2, with a β-sandwich in each domain. No isopeptide bonds or metal ions were present in the model. However the N1 domain consists of a putative binding cleft with mainly negative surface potential. Binding studies was performed on wild-type cells and on a Sgo0707 deletion mutant in order to elucidate possible ligands. Binding studies using flow-cell techniques resulted in collagen type-1, and keratinocytes as possible binding partners. Keratinocytes line the oral cavity and are important ligands in the oral cavity, whereas collagen type-1 may be of interest when *S. gordonii* is spread to non-oral sites, as with infective endocarditis. Expression and protein purification of AspA-AVPC from *S. pyogenes* (paper III) resulted in a soluble protein that easily fragmented into two pieces of 55 and 67 kDa in the absence of EDTA. The biological relevance for the fragmentation is not clear, but one reason may be to relocate when the environment is no longer suitable. Continued work with AspA-AVPC to further elucidate the fragmentation is of outmost importance. It would be of interest to clone, express, and purify individual domains, and different combinations thereof to see whether or not fragmentation occurs, and the putative metalloprotease could in that way possibly be identified. If AspA turns out to be a metalloprotease, the V-domain is the most likely candidate because of the bound metal ion and its location at the tip of the protein. The three determined structures will add to the knowledge of Gram-positive surface proteins generally, and streptococcal adhesins specifically. But much more work need to be done in order to clarify the composition of molecules involved in colonization within the oral cavity.
Acknowledgements

Det finns många människor som förtjänar min tacksamhet! Här nedan nämner jag de jag inte glömt bort! 😊

Först och främst vill jag tacka min handledare Karina Persson för din hjälp med allt inom forskningen, men även för att du vid väl valda tillfällen slängde ut orden: Slappna av!

Min bihandledare Elisabeth Sauer-Eriksson för hennes expertis!


Tack till Malin LP för all hjälp och stöd med allt mellan himmel och jord!

Tack till vår lilla grupp som vuxit sig större! Nina som la grunden, Michael och Patrik (a.k.a. bildexperten) som fortsätter arbetet! Ett särskilt tack till Michael som hjälpt mig med en massa saker, däribland korrekturläsning av denna avhandling!

Jag hade privilegiet att tillhöra den Svenska Nationella Forskarskolan i Odontologisk Vetenskap och har därigenom träffat en hel hög trevliga människor! Våra kurser har varit berikande både akademiskt och socialt. Särskilt tack till Anna Oldin för att du gjort mötena extra roliga!


Livet består ju inte bara av jobb, så tack även till min familj och utökade släkt! Tack till Siv, Hans, Björn och Niklas!

Tack till vänner som finns där för både toppar och dalar! Särskilt tack till Åsa, Daniell och Hobbe!

Till min älskade Mattias, tack för att du alltid stöttat mig! Tack också för att du plockat ner mig från väggarna när stressen blivit för påträffande!

(Som en lite parentes vill jag också tacka mig själv! BRA JOBBAT!!! 😊)
References


Ghuysen, J.M., Tipper, D.J., and Strominger, J.L. (1965). Structure of the cell wall of Staphylococcus aureus, strain Copenhagen. IV. The teichoic acid-


differential ligand-binding properties and mediates biofilm formation. Mol Microbiol 81, 1034-1049.


