

---

[All ETDs from UAB](#)

[UAB Theses & Dissertations](#)

---

2011

## Elucidation of the Elongated Fibrillar Structure of Streptococcus mutans Antigen I/II

Matt Rodney Larson  
*University of Alabama at Birmingham*

Follow this and additional works at: <https://digitalcommons.library.uab.edu/etd-collection>

 Part of the [Medical Sciences Commons](#)

---

### Recommended Citation

Larson, Matt Rodney, "Elucidation of the Elongated Fibrillar Structure of Streptococcus mutans Antigen I/II" (2011). *All ETDs from UAB*. 2221.  
<https://digitalcommons.library.uab.edu/etd-collection/2221>

This content has been accepted for inclusion by an authorized administrator of the UAB Digital Commons, and is provided as a free open access item. All inquiries regarding this item or the UAB Digital Commons should be directed to the [UAB Libraries Office of Scholarly Communication](#).

ELUCIDATION OF THE ELONGATED  
FIBRILLAR STRUCTURE OF  
STREPTOCOCCUS MUTANS ANTIGEN I/II

by

MATTHEW RODNEY LARSON

CHAMPION DEIVANAYAGAM, COMMITTEE CHAIR

LAWRENCE DELUCAS

CATHERINE FULLER

SUZANNE MICHALEK

PETER SMITH

A DISSERTATION

Submitted to the graduate faculty of The University of Alabama at Birmingham,  
in partial fulfillment of the requirements for the degree of  
requirements for the degree of  
Doctor of Philosophy

BIRMINGHAM, ALABAMA

2011

© Copyright by  
Matthew Rodney Larson  
2011

# ELUCIDATION OF THE ELONGATED FIBRILLAR STRUCTURE OF STREPTOCOCCUS MUTANS ANTIGEN I/II

MATTHEW RODNEY LARSON

PHYSIOLOGY AND BIOPHYSICS

## ABSTRACT

*Streptococcus mutans* (*S. mutans*) is the causative agent behind dental caries, an infectious disease also known as tooth decay or dental cavities. *S. mutans* has a cell wall-attached protein known as Antigen I/II (AgI/II) utilized for bacterial adhesion to the tooth surface. Here we have solved the structures of both amino- and carboxy-terminal regions of the AgI/II molecule using X-ray crystallography. Using this structural information we have now built a tertiary model for AgI/II as a fibrillar protein. Further, we have functionally characterized AgI/II and determined minimal regions of AgI/II that are implicated in its adherence to the salivary agglutinin found on the tooth surface.

### **The following results are presented in this thesis work:**

1. We report the high resolution structure (1.8 Å) of an amino-terminal fragment of AgI/II. This fragment reveals that AgI/II is an elongated molecule with a stalk comprised of  $\alpha$ - and polyproline type-II (PPII)-helices. The hybrid structure formed from  $\alpha$ - and PPII helices represents the first example of a new structural class of fibrillar proteins.
2. We report the (2.5 Å) crystal structure of the complete carboxy-terminus of AgI/II. This structure now shows that the AgI/II carboxy-terminus has three domains that each each adopt the DE-variant immunoglobulin-like fold.
3. Using the amino- and carboxy- terminal structures of AgI/II, biophysical characterizations, and electron microscopic imaging, we have now created an overall tertiary

model for AgI/II family of proteins.

4. Adherence studies for AgI/II revealed that AgI/II contains two distinct regions that adhered to human salivary agglutinin (SAG). These regions of AgI/II also adhered non-competitively to SAG. These results now suggest possible models for the *S. mutans* bacterial adherence to SAG.

AgI/II can now be described as an elongated fibrillar protein with globular domains at its termini. The V-region is distal to the cell surface, while three carboxy-terminal domains are close to the cell surface. These globular regions are separated by the alanine-rich and proline-rich repeats of AgI/II that form a unique  $\alpha$ - and PPII helical hybrid fibril which is revealed for the first time by these structures. This research now provides the first overall structural model for AgI/II as a fibrillar protein.

Keywords: *Streptococcus mutans*, Antigen I/II, fibrillar, structure

This work is dedicated to my wife Eppie; who has stood with me all these years in graduate school, both in Birmingham and when we were across the country. Her love has carried me through the difficulties, joys, and wonders of research. Without her it would not have been possible for me to complete this path.

My brother, Brad, led the way in science and provided a path for me to follow. To my parents Jim and Diane, I will always give my greatest thanks to you for making learning an important part of my life and showing me how to live thoughtfully with others in mind.

## ACKNOWLEDGEMENTS

All of our collaborators on this project have been of tremendous help. Sue Michalek deserves considerable thanks for helping initiate the Antigen I/II project. Jeannine Brady and Paula Crowley of UFL have contributed so much in studying Antigen I/II with regard to the science, discussions, and thoughts about the nature of this protein. I likewise thank our colleagues within the United Kingdom, Charles Kelly and Tim Mitchell for their work on this project. X-ray crystallography is a difficult science and this work would not be possible without the Center for Biophysical Sciences and Engineering at the University of Alabama at Birmingham or the beamline staff of NE-CAT and SERCAT. I especially thank Kanagalhatta Rajashankar, Irina Protassevich, and Christie Brouillette who have shared much time, equipment and lab space.

The members of the Deivanayagam Lab have made it more fun to work in a lab day in and out. Manisha Patel has always made it happier in the lab, Joo Hyoung Lee was like a Korean brother, Sangeetha (and Aadithya) Purushotham, Shaan Khaled, and Brian Sharon have all been friends these years.

All the members of my committee have also shared their time and thoughts with me over these last five years. I truly thank you - Cathy Fuller, Sue Michalek, Larry DeLucas, and Peter Smith. Dale Benos, who was not only a member of my committee, but also a driving force behind the Department of Physiology and Biophysics, is deeply missed at the conclusion of my studies.

Finally, I cannot thank Champion Deivanayagam enough for his time, patience, and knowledge - all of which he shared in abundance with me in these past years. Both Champion and his wife Lydia have treated all of us in his lab as if we were part of their family.

Thank you for your kindness and effort during all these years.

## LIST OF ABBREVIATIONS

|                      |   |
|----------------------|---|
| $\alpha$ -helix..... | a common right-handed helical structure found in proteins |
| A-region .....       | Alanine-rich region                                       |
| AEP .....            | acquired enamel pellicle                                  |
| AgI/II.....          | Antigen I/II  |
| DEv-IgG fold.....    | DE-variant Immunoglobulin-like folds                      |
| DexA .....           | Dextranase A  |
| DMBT1.....           | Deleted in Malignant Brain Tumors<br>1                    |
| ECM .....            | Extracellular Matrix                                      |
| FruA.....            | Fructanase A  |
| GbpC.....            | Glucan-binding protein C                                  |
| gp-340.....          | glycoprotein 340, the main component of SAG               |
| kDa.....             | kilodalton, a unit of mass                                |

|                            |   |
|----------------------------|---|
| LPxTG.....                 | Anchoring motif                         |
| MIR.....                   | multiple isomorphous replacement        |
| PCR.....                   | Polymerase chain reaction               |
| PPII.....                  | polyproline type-II                     |
| SAG.....                   | Salivary Agglutinin                     |
| SIDs.....                  | SRCR interspersed domains               |
| sIgA.....                  | Secretory Immunoglobulin A              |
| SRCR domain.....           | scavenger receptor cysteine-rich domain |
| V-region.....              | Variable region                         |
| WapA.....                  | Wall Associated Protein A               |
| WapE.....                  | Wall Associated Protein E               |
| <i>P. gingivalis</i> ..... | <i>Porphyromonas gingivalis</i>         |
| <i>S. mutans</i> .....     | <i>Streptococcus mutans</i>             |

## TABLE OF CONTENTS

|  |             |
|--|-------------|
| <b>ABSTRACT</b>  | <b>iii</b>  |
| <b>DEDICATION</b>  | <b>v</b>    |
| <b>ACKNOWLEDGEMENTS</b>                                  | <b>vi</b>   |
| <b>LIST OF TABLES</b>                                    | <b>xii</b>  |
| <b>LIST OF FIGURES</b>                                   | <b>xiii</b> |
| <b>1 INTRODUCTION</b>                                    | <b>1</b>    |
| 1.1 Oral Bacteria and Tooth Decay . . . . .              | 1           |
| 1.1.1 Resisting Salivary Flow . . . . .                  | 3           |
| 1.1.2 Mechanisms of Adherence . . . . .                  | 4           |
| 1.2 <i>Streptococcus mutans</i> . . . . .                | 6           |
| 1.3 Antigen I/II . . . . .                               | 8           |
| 1.3.1 AgI/II as an adherence protein . . . . .           | 11          |
| 1.3.2 Interaction with Salivary agglutinin . . . . .     | 11          |
| 1.3.3 Mechanism for the SAG-AgI/II interaction . . . . . | 13          |
| 1.3.4 Other roles for the AgI/II family . . . . .        | 14          |
| 1.3.5 Glimpses of the AgI/II structure . . . . .         | 15          |
| 1.4 Aims of this Study . . . . .                         | 16          |

|          |  |            |
|----------|--|------------|
| 1.5      | Organization of this dissertation . . . . .  | 18         |
| <b>2</b> | <b>ELONGATED FIBRILLAR STRUCTURE OF A STREPTOCOCCAL ADHESIN ASSEMBLED BY THE HIGH AFFINITY ASSOCIATION OF ALPHA- AND PPII- HELICES</b>               | <b>19</b>  |
| <b>3</b> | <b>CRYSTAL STRUCTURE OF THE C-TERMINUS OF <i>STREPTOCOCCUS MUTANS</i> ANTIGEN I/II AND CHARACTERIZATION OF SALIVARY AGGLUTININ ADHERENCE DOMAINS</b> | <b>59</b>  |
| <b>4</b> | <b>CONCLUSION</b>  | <b>87</b>  |
| 4.1      | Determination of the AgI/II Structure . . . . .  | 87         |
| 4.1.1    | N-terminal Region Crystallization . . . . .  | 88         |
| 4.1.2    | C-terminal Region Crystallization . . . . .  | 89         |
| 4.1.3    | N-terminal AgI/II structure . . . . .  | 90         |
| 4.1.4    | C-terminal AgI/II structure . . . . .  | 93         |
| 4.1.5    | A fibrillar model of AgI/II . . . . .  | 94         |
| 4.2      | AgI/II Adherence Function . . . . .  | 94         |
| 4.2.1    | Ligand Binding Sites of AgI/II . . . . .   | 98         |
| 4.3      | Other AgI/II Family Members . . . . .  | 100        |
| 4.4      | Future of Adherence Inhibition . . . . .   | 102        |
|          | <b>LIST OF REFERENCES</b>  | <b>104</b> |
|          | <b>APPENDIX: ALIGNMENTDRAW: SEQUENCE ALIGNMENT TOOL</b>  | <b>115</b> |

## LIST OF TABLES

|     |   |     |
|-----|---|-----|
| 2.1 | Isothermal titration calorimetry measurements of the A-P interaction . . . .                    | 30  |
| 2.2 | Sedimentation velocity measurements . . . . .   | 32  |
| 2.3 | Predicted AgI/II secondary structures from circular dichroism (CONTIN/LL)                       | 37  |
| 2.4 | PCR primers used in N-terminal structural study . . . . .                                       | 43  |
| 2.5 | AgI/II A <sub>3</sub> VP <sub>1</sub> X-ray data collection and refinement statistics . . . . . | 47  |
| 3.1 | PCR Primers used in the C-terminal structural study . . . . .                                   | 64  |
| 3.2 | AgI/II C-terminus X-ray data collection and refinement statistics . . . . .                     | 66  |
| 3.3 | Competitive inhibition to SAG-coated hydroxyapatite by AgI/II fragments .                       | 76  |
| 4.1 | Residues with sidechains involved in ligand and ion interactions of AgI/II .                    | 101 |

## LIST OF FIGURES

|      |   |    |
|------|---|----|
| 1.1  | The oral cavity and tooth structure . . . . .                                       | 2  |
| 1.2  | Adherence proteins of bacteria . . . . .  | 5  |
| 1.3  | <i>Streptococcus mutans</i> . . . . .   | 7  |
| 1.4  | <i>Streptococcus mutans</i> Antigen I/II . . . . .                                  | 8  |
| 1.5  | Sequence alignments of the AgI/II family . . . . .                                  | 9  |
| 1.6  | AgI/II A- and P-repeat sequences . . . . .  | 10 |
| 1.7  | SAG/gp-340/DMBT1 . . . . .  | 12 |
| 1.8  | A sequence map of AgI/II function . . . . .   | 14 |
| 2.1  | AgI/II protein and constructs . . . . .   | 22 |
| 2.2  | Antigen I/II layout and A <sub>3</sub> VP <sub>1</sub> structure overview . . . . . | 25 |
| 2.3  | Interactions between the A <sub>3</sub> and P <sub>1</sub> helices . . . . .        | 26 |
| 2.4  | Stereo images of the A <sub>3</sub> and P <sub>1</sub> helices . . . . .            | 27 |
| 2.5  | Avian pancreatic peptide . . . . .  | 29 |
| 2.6  | Calorimetric of the A/P interaction . . . . .                                       | 30 |
| 2.7  | Inhibition of binding of anti-AgI/II MAbs to <i>S. mutans</i> whole cells . . . . . | 33 |
| 2.8  | SPR studies of AgI/II fragments adherence to immobilized SAG . . . . .              | 35 |
| 2.9  | Crystal packing of A <sub>3</sub> VP <sub>1</sub> . . . . .                         | 38 |
| 2.10 | Model of the AgI/II structure and predicted binding with SAG . . . . .              | 40 |
| 2.11 | Crystals of A <sub>1-3</sub> . . . . .  | 41 |
| 2.12 | Purification of A <sub>3</sub> VP <sub>1</sub> . . . . .                            | 45 |

|     |  |    |
|-----|--|----|
| 3.1 | C-terminal structure and primary sequence of AgI/II of <i>S. mutans</i> . . . . .              | 68 |
| 3.2 | C-terminal domains of AgI/II . . . . .   | 70 |
| 3.3 | Comparison of AgI/II C <sub>123</sub> and SspB C23 structures . . . . .                        | 71 |
| 3.4 | Interaction of carbohydrates within the C <sub>12</sub> cleft . . . . .                        | 72 |
| 3.5 | Evaluation of AgI/II polypeptide-SAG interactions using surface plasmon<br>resonance . . . . . | 74 |
| 3.6 | Electron micrographs and composite model of AgI/II of <i>S. mutans</i> . . . . .               | 78 |
| 4.1 | Crystallization of fragments <i>S. mutans</i> AgI/II . . . . .                                 | 88 |
| 4.2 | Structures of A <sub>3</sub> VP <sub>1</sub> and C <sub>123</sub> . . . . .                    | 90 |
| 4.3 | A comparison of fibrillar structures . . . . .   | 91 |
| 4.4 | The fibrillar model of AgI/II . . . . .  | 95 |
| 4.5 | Adherence of an elongated adhesin . . . . .  | 98 |
| 4.6 | Ligands of the A <sub>3</sub> VP <sub>1</sub> and C <sub>123</sub> structures . . . . .        | 99 |

# CHAPTER 1

## INTRODUCTION

Dental caries, also known as tooth decay or cavities, is a common and chronic ailment that amounts to a “silent epidemic” disproportionately affecting poor children, the elderly, and many racial and ethnic minorities [1]. 59% of children ages 5-17 and 86% of adults within the United States have been negatively affected by dental caries or periodontal disease [1]. We commonly consider cavity prevention to be a matter of oral hygiene, but caries is the most common infectious disease affecting humans [2]. The first step of caries is the adhesion of microbial species to the tooth, and this is considered to be a possible target for preventive therapies. The primary etiological agent implicated in dental caries is the gram-positive bacteria *Streptococcus mutans* (*S. mutans*). *S. mutans* adhesion to teeth occurs through proteins of the bacterial cell surface, which mediate specific interactions with the host. In this dissertation, we report the structural determination of the *S. mutans* adherence protein Antigen I/II (AgI/II). All together, this describes the first tertiary structural model for the AgI/II protein family, which we hope can contribute to the understanding of the pathogenic adherence of *S. mutans* and enable the design of compounds to disrupt adherence of the caries bacteria to human teeth.

### 1.1 Oral Bacteria and Tooth Decay

Humans produce their first teeth in infancy, building the initial set of 20 deciduous teeth by age of 3. These are replaced with 32 permanent teeth that fill in the mouth starting at age 6 until completion by age 14. Each tooth has a crowning outermost layer called the

tooth enamel that is made of calcium hydroxyapatite (Fig 1.1). The enamel is the first layer of protection for the human tooth, and surrounds increasingly vulnerable tissues. The next layer inside is the dentin, a bone-like mixture of mineral and collagen. At the core of the tooth is a vascularized tissue full of nerve endings called the pulp. Carious lesions or pulp exposure due to trauma allow oral bacteria to colonize dentinal tubules and create a root canal infection [3]. The acute symptoms of tooth decay are felt when this infection reaches the pulp and irritates the nerves, resulting in intense sensations of pain that are associated with the end stages of tooth decay.

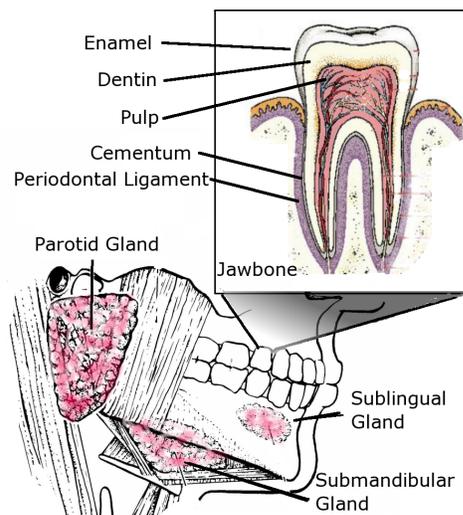


Figure 1.1: *The oral cavity and tooth structure.* Saliva is produced by the parotid, sublingual, and submandibular major glands and also by hundreds of minor glands. The inset details the structure of a human tooth which has several layers including the enamel, dentin, and pulp.

Infections of the tooth arise from microbial species present with the oral cavity. With more than 500 unique bacterial species recorded from samples, there is a vibrant diversity of microbial species within the human oral cavity [4]. As modern techniques such as the polymerase chain reaction (PCR) are employed to identify oral species, the total number of microbes will surely be revised upwards. Among these bacterial species, the

streptococci and lactobacilli are the most commonly associated with dental disease. Gram-positive oral streptococci adhere directly with the tooth surface via human protein receptors within the salivary pellicle. Upon attachment to the tooth, streptococcal colonies multiply and form multilayer biofilms. These bacterial cells digest common dietary sugars through a fermentative process. An unfortunate side product of carbohydrate fermentation is acid production, which lowers the pH of the surrounding fluid at the tooth surface. This is the root cause in the pathogenicity of tooth decay. Under acidic conditions, pits (cavities) are formed on the surface of the tooth as the enamel is eroded and the calcium hydroxyapatite dissolves into solution. Oral bacterial species which both easily adhere to human teeth and metabolize dietary carbohydrates are therefore associated with tooth decay. Of all the species present in the mouth, *S. mutans* is the major causative pathogen that is responsible for dental caries [5]. What makes this species more virulent than the rest? One of its most important traits provides *S. mutans* with an advantage: its ability to adhere to teeth.

## **1.1 Resisting Salivary Flow**

The oral cavity, unlike a surface such as skin, has the unique challenge of shear forces in fluid flow during eating, drinking, and with the constant production of saliva. To colonize the mouth, a microbe must find a way to resist these forces and keep firmly attached to a surface within the oral cavity.

In a healthy oral environment, saliva will be produced continuously at a rate of one liter per day [6]. The flow rate of saliva increases during chewing and eating, while flow decreases during sleep [6]. Salivary fluid is produced mainly by three major glands and contains a secreted mix of proteins, mucus, salts, and buffering agents. The parotid gland produces watery secretions, the sublingual produces largely mucus secretions, and the submandibular produces a mix between the two extremes [7] (Fig 1.1). Under greatly reduced

or absent salivary flow (xerostomia) rampant tooth decay occurs [5], indicating the great protective value of saliva against the oral bacteria.

The tooth surface is the primary adherence site for caries-causing bacteria. Saliva washing over the tooth surface deposits a proteinaceous film called the acquired enamel pellicle (AEP). The AEP is an amorphous layer, 0.1 to 0.3  $\mu\text{m}$  thick, containing a high number of both carboxyl and sulfate groups from the deposited salivary proteins, increasing the net negative surface charge of the teeth [5]. Bacterial cell surfaces are also negatively charged, and so the AEP produces a natural repulsion against bacterial cells which may be considered part of the innate immune defense. Caries-causing bacteria have developed specific adherence proteins that overcome these challenges.

## **1.1 Mechanisms of Adherence**

A single microbe will typically utilize multiple methods for attaching to a host simultaneously. Adherence proteins of gram-positive bacteria may reside within the outer cell membrane or be attached covalently to external cell walls made of peptidoglycan. These surface proteins are the tools that the oral bacteria employ to gain adherence with their host. The most common adherence proteins include the pili, fibrillar proteins, extracellular-matrix binding proteins, lectins, and glucan-producing/binding proteins (Fig. 1.2).

Pili are elongated stalks which are built of many subunits (fibrillin). The stalks are anchored at one end to the cell wall by the action of sortase enzymes. Sortases are also involved in linking the fibrillin subunits to each other [8]. Adhesion sites are often present in the distal domains of the pili far from the cell surface.

Fibrillar structures are formed from long helical coils, which are often intertwined as dimers, with the most famous example being the M-protein that is common to all streptococci [9]. Sortase enzymes attach fibrillar proteins to the cell wall, as occurs with pili.

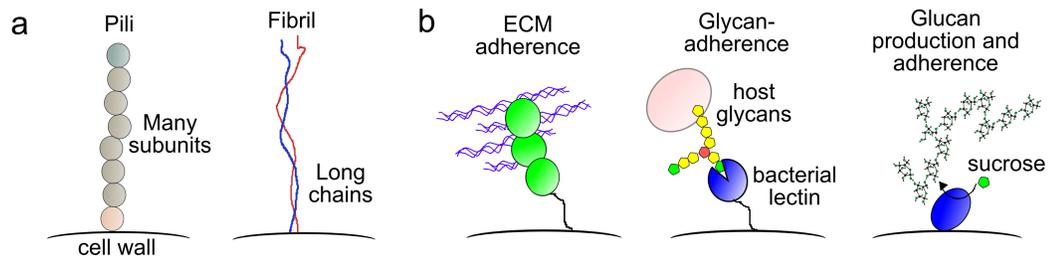


Figure 1.2: *Adherence receptors of bacteria*- Many bacterial adherence proteins are elongated either through repeating subunits (such as in pili) or from fibrillar forms (a). Bacterial proteins adhere to common host receptors such as extracellular matrix molecules (ECM) or glycan signatures on proteins. Oral bacteria also create their own sticky sugar polymers, called glucans, which enhance adhesion and biofilm formation (b).

Differing from the pili, the fibrillar proteins are granted their elongation by helical polypeptides rather than the assembly of multiple globular subunits.

Some bacterial proteins adhere to tissues by attaching to molecules of the supportive frameworks that join the host cells, i.e. the extracellular matrix (ECM). ECM is composed of fibrous molecules such as collagen, fibronectin, elastin, and laminin which are attractive targets for bacterial adherence. Bacterial ECM-binding proteins have received the unfortunately long name “MSCRAMMs”, for microbial surface components recognizing adhesive matrix molecules. MSCRAMMS are commonly employed by bacteria in colonization and virulence owing to their utility in binding the ubiquitous ECM found in tissue surfaces.

Complex structures made of sugar molecules, glycans, are one of the major components of cellular life. Bacterial proteins called lectins recognize these host sugars and are commonly utilized to adhere to the human host [10], as mammalian cells have evolved distinct glycan signatures that identify them apart from lower invertebrates, lower eukaryotic, and prokaryotic cells [11]. Many salivary and mucosal proteins are highly glycosylated and serve as potential targets for host colonization.

In addition to host sugars, oral bacteria have developed their own sticky sugars called

glucans. We are often warned by our dentists to cut down on our sugar intake or risk a return trip- this is reasonable advice as oral streptococci utilize sucrose as the building block for the formation of large sticky polymers. Virulence of both *S. sobrinus* and *S. mutans* has been associated with enzymatic glucosyltransferases, which transform sucrose (white sugar) into sticky glucan that is useful for building biofilms on the human tooth. Glucosyltransferases produce soluble glucan with  $\alpha$ -(1-6) linkages and insoluble glucan with  $\alpha$ -(1-3) linkages. The insoluble form has been shown to play a major role in smooth-surface tooth decay [5]. The development of glucan on the tooth surface is crucial for building surface biofilms, however the streptococcal adherence to the tooth also occurs through direct adherence to the salivary pellicle, without the necessity of dietary sucrose and glucan.

Finally, it is worth noting that all these classes of adherence are indistinct. The large adherence proteins of the gram-positive bacterial cell surface often contain many different forms of adherence domains within a single protein, turning them into “Swiss-army knives” of adherence function.

## **1.2** *Streptococcus mutans*

The mutans group of streptococci, and in particular *S. mutans*, have emerged as the primary causative agents of dental caries [5]. The major virulence of *S. mutans* occurs with dental caries [5], however infections of *S. mutans* also contribute to bacteremia and infective endocarditis [12, 13]. *S. mutans* is believed to have co-evolved with humans and is mainly transmitted from mothers to offspring [14]. Until the rise of human agriculture, and subsequent flood of carbohydrates into the human diet, the interactions between the host and bacterium may have been nondestructive. The dietary grains suited the pathogenic nature

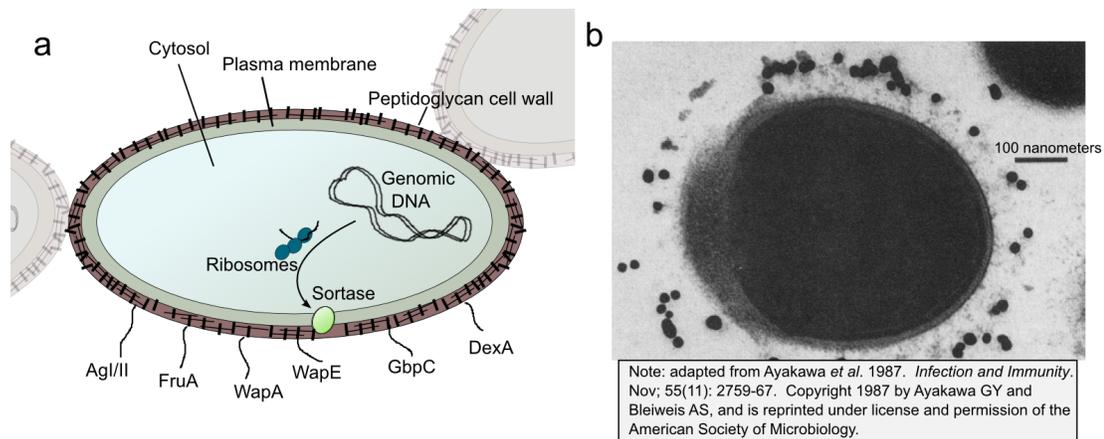


Figure 1.3: *Streptococcus mutans* is a gram positive bacterium with an internal cytoplasmic membrane and a thicker external peptidoglycan cell wall (a). Proteins are produced in the cytosol. Secretion signals direct adherence proteins to the cell surface, where sortase enzymes attach them to the cell wall. An electron micrograph from Ayakawa *et al.* [16] of the *S. mutans* cell shows the fuzzy coat on the cell surface (b). Gold-labeled antibodies against Antigen I/II appear as dark circles on the fuzzy cell surface. The bar represents a 100 nanometer distance.

of *S. mutans*, which genomic sequencing has revealed to have more genes for the breakdown of carbohydrates than any other gram-positive bacterial species sequenced thus far [15]. These highly evolved metabolic abilities provide *S. mutans* with energy sources in the oral environment and high acidogenicity.

*S. mutans* adheres to both smooth surfaces and fissures of human teeth. The bacterium utilizes a set of adherence proteins covalently attached to its cell wall for this task. Amino-terminal signal sequences of these proteins provide surface localization, while their carboxy-terminus have another amino-acid motif (LPxTG) marking their anchorage site for the cell wall [17]. Genomic sequencing of *S. mutans* identified six genes which contain a carboxy-terminal LPxTG anchoring motif required in cell-wall attached surface proteins [15] (Fig. 1.3). These are Antigen I/II (AgI/II), fructanase (FruA), wall-associated protein A (WapA or AgIII), wall-associated protein E (WapE), glucan-binding protein C (GbpC), and dextranase (DexA) [15]. Several of these proteins appear to have roles in bacterial

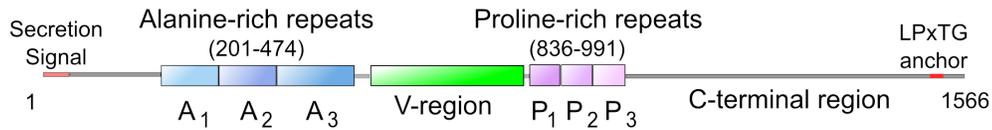


Figure 1.4: *Streptococcus mutans* Antigen I/II is a large 170-kDa protein with several regions: sets of alanine-rich and proline-rich repeats, a variable domain, and a large conserved carboxy-terminus. Secretion to the cell surface is directed from an N-terminal signal, and the protein is anchored via a peptide motif near the end of the carboxy-terminus.

adherence while the others (DexA and FruA) have enzymatic functions. This thesis work focuses on deciphering the structural and functional characteristics of AgI/II, a protein first discovered over 30 years ago [18]. AgI/II is widely considered to initiate the adhesion of *S. mutans* to the human tooth through specific adherence to the salivary pellicle.

### 1.3 Antigen I/II

AgI/II was originally detected as two individual antigenic components (Antigens I and II) in lysates of *S. mutans* [18], only later to be understood that Antigen I and II were sub-components of a single protein [19]. Afterward, the protein was renamed Antigen I/II [18]. AgI/II has also been variously referred in literature as SpaP, PAc, AgB, and P1. Variants of the AgI/II gene were cloned from *S. mutans* [20, 21, 22]. AgI/II variants have been identified in many other oral streptococci and the Group A and B streptococcal strains of *S. pyogenes* and *S. agalactiae* [23].

AgI/II family proteins range from 1310 to 1654 amino acids in length [26], and sequences are conserved across the family (Figs 1.4 and 1.5). At the amino terminus of AgI/II, a signal sequence directs its secretion from the cell. The signal is next followed by the alanine-rich region (A-region) of residues 202-448. The A-region has 22-30% alanine

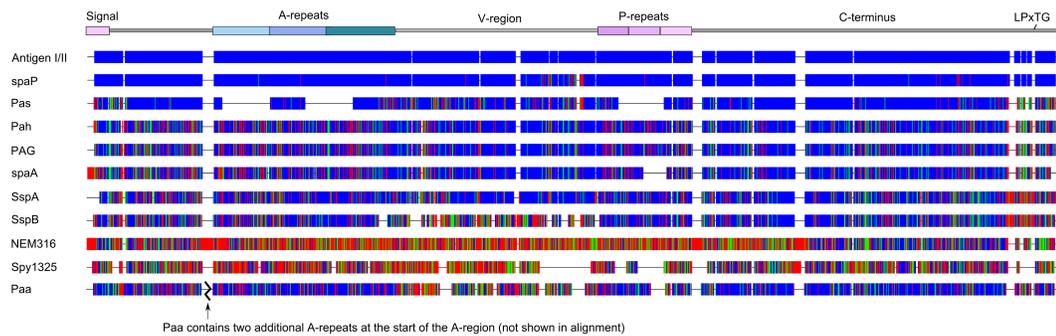


Figure 1.5: *Sequence alignment of the AgI/II family*- A Clustalx2 [24] alignment of the AgI/II family is displayed graphically using AlignmentDraw (See Appendix A). *S. mutans* NG8 AgI/II (GenBank id # ACV69919.1) was aligned with *S. mutans* SpaP (AAN58348.1), *S. intermedius* Pas (BAA96878.2), *S. downei* Pah (AB207813.1), *S. sobrinus* PAG (D90354.1), *S. sobrinus* spaA (X57841.1), *S. gordonii* SspA (AAC44099.1), *S. gordonii* SspB (AAC44100.1), *S. agalactiae* NEM316 (CAD47015.1), *S. pyogenes* M28 Spy1325 (AAX72435.1), and *S. criceti* Paa (AB042239.3). Positions with identical sequence are colored in blue, similar residues are colored in green, and dissimilar residues are shown red. Gaps are illustrated in white. A high level of variation in the V-region in both Paa and SspB. Pas has deletions resulting in fewer A-repeats and P-repeats. Paa has a total of 5 A-repeats, with two extra A-repeats entailing a total of 164 residues which were omitted from this diagram for clarity of the alignment. The *S. agalactiae* and *S. pyogenes* AgI/II family members are the most dissimilar to *S. mutans* AgI/II. [25]

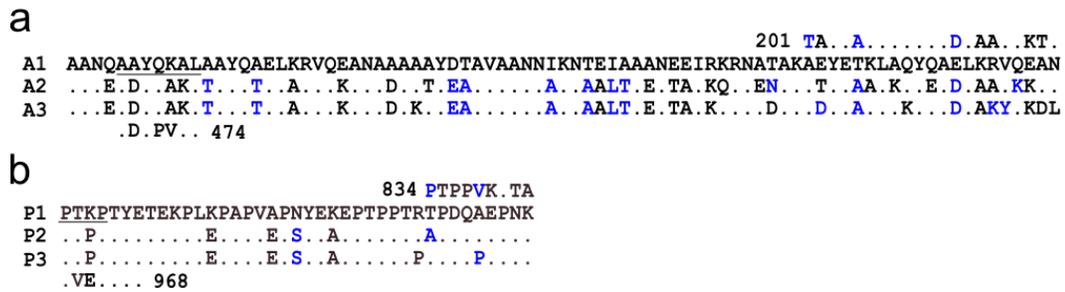


Figure 1.6: *AgI/II* A- and P- repeats- There are three A-repeats of 82 amino acids in length within the N-terminus of *AgI/II*, with partial repeat sequences at their start and end (a). Within the A-repeats are heptad motifs of [TA]xYxxxL. Following the V-region are three P-repeats of 38 amino acids in length (b). Examples of A-repeat heptad and P-repeat PxxP motifs are underlined. Identical (.) and similar residues (blue) are listed in the alignments.

content in the *AgI/II* family, and most *AgI/II* A-regions contain three repeated peptide sequences of 82 amino acids in length, called the A<sub>1</sub> (202-284), A<sub>2</sub> (285-366), A<sub>3</sub> (367-448) (See Fig 1.6).

Following the A-region between residues 495-832 is a variable sequence (V-region) that diverges in the *AgI/II* family. The V-region was the first segment of *AgI/II* to be structurally characterized and is a distorted  $\beta$ -sandwich consisting of two sheets that each contain eight antiparallel strands [27]. The largest divergences in the sequence of the V-region among the oral streptococci are in *S. gordonii* SspB, which has both sequence alterations and deletions (Fig 1.5). Another set of sequence repeats (834-964) immediately follows the V-region and are called the P-region due to repeats with  $\sim 35\%$  proline content. Each repeat is 38 amino acids in length, with the P<sub>1</sub> (844-882), P<sub>2</sub> (883-921), P<sub>3</sub> (922-960) shown in Fig 1.6. One major difference in some *AgI/II* family proteins is varying numbers of the A- and P- repeats, as the *S. intermedius* Pas has only a single A/P repeat and *S. criceti* Paa has an additional two A-repeats.

The A, V, and P regions form the N-terminal *AgI* portion of *AgI/II* [28]. The remaining *AgII* consists of the 550 amino acids at the carboxy-terminus following the P-repeats of

AgI/II. The carboxy-terminus (C-terminus) of AgI/II is the most conserved region of the AgI/II family, with ~62% identity among family members (Fig 1.5). Defining the end of the C-terminus of AgI/II is an LPxTG anchoring motif sequence that is recognized by a sortase enzyme. The sortase enzyme covalently attaches the threonine residue of the LPxTG motifs to the bacterial cell wall and cleaves the remaining C-terminal residues [29]. This motif is critically important for AgI/II to be anchored to the cell surface as disrupting the C-terminus results in unattached proteins that are secreted into the supernatant [30].

Electron micrographs (See Fig 1.3, b) using anti-AgI/II antibodies find that the AgI/II family proteins form an evenly distributed, fuzzy coat on the surface of oral streptococcal species [16, 31].

### **1.3 AgI/II as an adherence protein**

The first hints of AgI/II function were seen in mutational studies that deleted the AgI/II gene from *S. mutans*. The AgI/II-less mutants had decreased surface hydrophobicity and lost their ability to adhere to saliva-coated hydroxyapatite [32]. Additionally, the bacteria lost a phenotype of aggregation by salivary proteins [32]. AgI/II itself adhered to saliva-coated hydroxyapatite beads, which were used as an *in vitro* model for the AEP surface [28]. Together, it became clear that AgI/II was providing *S. mutans* with adherence to the salivary pellicle on the tooth surface.

### **1.3 Interaction with Salivary agglutinin**

At nearly the same time as the discovery of AgI/II, salivary proteins were identified that could cause the aggregation of *S. mutans* [33]. Salivary proteins that interacted with bacteria were thought to promote the aggregation and clearance of the oral bacteria through the salivary flow, or conversely, to become incorporated into the acquired pellicle of the

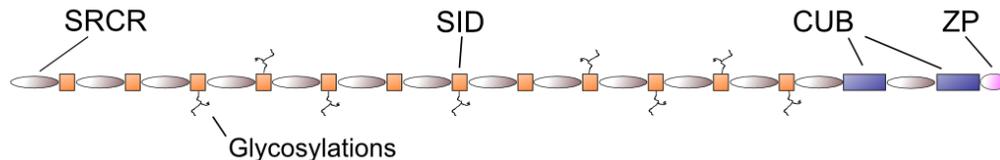


Figure 1.7: *SAG/gp-340/DMBT1*- SAG is a protein complex whose main component is gp-340 (also known as DMBT1), a polypeptide that consists of 14 SRCR domains. Near the carboxy-terminus, the 14<sup>th</sup> SRCR domain is interspersed between two CUB domains and a single ZP domain. Polypeptide segments between the SRCR domains are highly glycosylated.

tooth surface and aid the attachment of bacteria to the tooth [33]. Many proteins in saliva caused bacterial aggregation, such as lysozyme [34], secretory immunoglobulins (sIgA) [35], and high molecular weight secretory glycoproteins [33]. The binding of the high molecular weight glycoprotein salivary agglutinin (SAG) to streptococci and resulting bacterial clumping has been known for almost three decades [36]. Of the salivary proteome, SAG is one of the strongest binding components to *S. mutans* [37].

SAG purified from the human salivary fluid was biophysically characterized to be a protein complex composed of monomeric units of 440 kilodalton (kDa) size that associate as complexes as large as 5000 kDa. [33] The 440 kDa monomers were revealed by mass spectroscopy to be identical to the lung protein gp-340, encoded by the Deleted in Malignant Brain Tumors 1 (DMBT1) gene [38]. SAG can also be found in tear and gastrointestinal fluids secreted by epithelial cells at mucosal surfaces [26].

Gp-340 consists of 14 scavenger receptor cysteine-rich domains (SRCR domains), with two CUB domains and a single zona pellucida domain surrounding the 14<sup>th</sup> SRCR domain (Fig. 1.7). Between adjacent SRCR domains are highly glycosylated regions also called SRCR interspersed domains (SIDs). Variations in these glycosylation patterns of SAG occur among individuals and at various tissue locations [38].

The large body of evidence confirms that *S. mutans* interaction with SAG occurs primarily through the AgI/II protein [39, 40, 41, 42, 43, 44]. Bacterial mutants lacking AgI/II lost their ability to adhere to artificial SAG-coated surfaces [44]. When SAG is immobilized on the tooth surface as part of the salivary pellicle, SAG acts as a receptor for streptococcal adherence [43]. In this role, the interaction between SAG and *S. mutans* AgI/II has become a target in order to disrupt streptococcal adherence to human teeth. A necessary step along this path is the determination of the mechanism driving the interaction of SAG with AgI/II.

### **1.3 Mechanism for the SAG-AgI/II interaction**

The interaction of SAG with AgI/II has proven to be complex and multiple regions of AgI/II are now implicated (Fig. 1.8). The mutans streptococci are specifically aggregated by SAG in a manner dependent on the presence of calcium [33]. Monoclonal antibodies against AgI/II that map to widely different regions of primary sequence will inhibit bacterial SAG adherence and aggregation [39]. Moreover, some AgI/II antibodies that inhibit adherence to SAG coated hydroxyapatite fail to similarly inhibit aggregation by fluid-phase SAG [26]. The Scatchard analysis of the AgI/II binding to SAG found both higher and lower affinity binding [28]. Sequences within the AgI/II A-repeats were identified that might serve as salivary protein binding motifs [45], and the A-repeat fragments, in isolation, bound SAG [40]. Carboxy-terminal fragments of AgI/II also inhibited the binding of SAG to *S. mutans* [41]. AgI/II family proteins bind calcium ions within the AgI/II carboxy-terminus, which are possibly involved in the calcium-dependent binding with SAG [46]. A short peptide (amino acids 1025-1044) whose sequence was taken from the carboxy-terminus of AgI/II could prevent adhesion of *S. mutans* to teeth and the re-colonization of the oral cavity by *S. mutans* in human subjects [47].

A peptide with sequence “QGRVEVLYRGSWGTV C”, taken from the SRCR domains

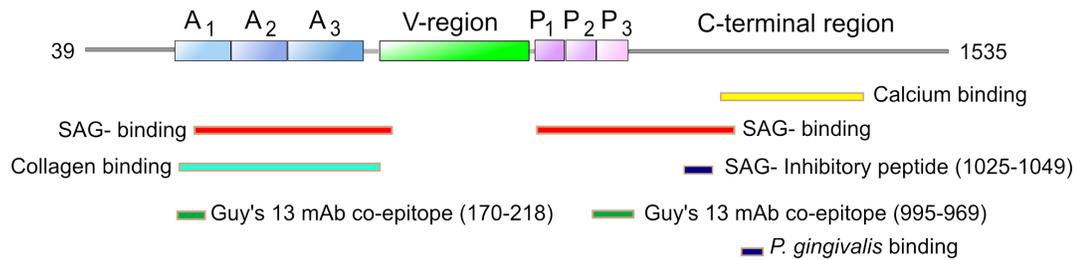


Figure 1.8: *A sequence map of AgI/II function*- Many regions of AgI/II have been implicated in functionality. Fragments of the A-repeats bind SAG [40] and collagen [52]. An AgI/II fragment containing the P-repeats and carboxy-terminus inhibits binding of SAG to *S. mutans* [41], and a peptide from this region acts as an inhibitor for *S. mutans* colonization of teeth [47]. Calcium binding sites exist within the carboxy-termini of some AgI/II family proteins [46]. AgI/II family protein SspB adheres to *P. gingivalis* through a minimal sequence identified in the AgI/II carboxy-terminus [53]. Portions of the A- and P- repeats form a co-epitope recognized by monoclonal antibodies [54].

of gp-340, adheres to oral bacterial strains [48]. However, this short peptide also binds to many streptococcal and non-streptococcal species, including those without the AgI/II protein [49].

Various sugars and amine-containing compounds inhibit the binding of SAG with mutans streptococci [50, 28, 51]. From these results it has been suggested that AgI/II may interact with SAG through its extensive glycosylations, however the identification of a specific sugar(s) to which AgI/II directly adheres and their requirement in AgI/II-mediated adhesion is still to be substantiated.

### 1.3 Other roles for the AgI/II family

AgI/II has functions in addition to host adherence. Within the oral cavity, co-aggregation of multiple species is a widespread phenomenon observed in biofilms [4]. In the complicated oral environment, the AgI/II homologue of *S. gordonii* also adheres to the gram-negative bacterium *Porphyromonas gingivalis* (*P. gingivalis*), the causative agent for periodontal

disease. This interaction is confined to a site in the carboxy-terminus of AgI/II (SspB) mapped by two peptides “KKVQDLLKK” and “NITVK” [53]. Mutans streptococci also invade dentinal tubules after damage to the tooth enamel, in a manner dependent on AgI/II [55]. AgI/II adheres to a growing set of receptors in saliva and extracellular matrix, now including salivary agglutinin (SAG) [40, 41, 28, 47], collagen [56], fibronectin, and fibrinogen [57], and platelets [58].

### **1.3 Glimpses of the AgI/II structure**

The first structural view of AgI/II began with the crystallization of the variable region of AgI/II [27]. This crystal structure revealed a globular structure composed of a  $\beta$ -supersandwich fold made of 16  $\beta$ -strands [27]. This structure was suggested to be reminiscent of a lectin-like fold, able to accommodate a small molecule such as a carbohydrate in a large trench formed within this domain [27].

However, beyond the V-region, the majority of the remaining structure of AgI/II was still unknown. Studies focused on the alanine- and proline- rich repeats have hinted at a tertiary organization of AgI/II. Both of these repeats are essential for surface localization of AgI/II, and when these regions were removed from AgI/II it resulted in an unstable protein that was no longer surface expressed [59]. The “Guy’s 13” monoclonal antibody that provides long-term protection against colonization by mutans streptococci [60] also recognizes a co-epitope formed by the A- and P-repeats, and this has suggested that in a properly folded AgI/II protein, these two regions are closely associated in the structure of AgI/II [54].

A more complete knowledge of the overall tertiary structure of AgI/II, as would be necessary for a structure to function analysis, would now require the further structural determination of the remaining unknown portions of AgI/II. These include the alanine-rich

repeats, proline-rich repeats, and the carboxy-terminus of AgI/II.

## 1.4 Aims of this Study

Many regions of the AgI/II sequence are now described to have a putative interaction with SAG and other proteins, while the AgI/II structure has remained unsolved. Structural models have the potential to clarify the mechanisms driving AgI/II adherence, and define minimal binding regions that could be target-able for the inhibition of microbial adherence. We have hypothesized that *S. mutans* AgI/II binds to SAG through specific sites and that the structural determination of AgI/II can help determine the mechanisms of adherence. This study is directed at the determination of the overall structure of AgI/II and the extents of the minimal regions most involved in the adherence to SAG.

In order to solve the structure of AgI/II, we have produced both N- and C-terminal fragments of AgI/II that were crystallized and used for X-ray crystallographic structure determination. The determined structures were then used to build a composite model of the whole adhesin. The goals of this study are broadly defined below in these three specific aims:

- **Aim 1: Characterize the adherence of the domains of AgI/II to SAG using quantitative techniques.** A set of recombinant AgI/II fragments were cloned, expressed, and purified, while human SAG was also purified from the saliva of healthy volunteers. These proteins were used in studies directed at the measurement of the AgI/II regions which are functionally involved in adherence to SAG. Direct interactions of AgI/II and SAG were measured in real time through surface plasmon resonance techniques. Similarities of conformation were measured for AgI/II fragments by testing

their ability to mimic monoclonal antibody epitopes found on the *S. mutans* cell surface. The ability of the AgI/II fragments to inhibit the adhesion of *S. mutans* to SAG-coated hydroxyapatite was measured to confirm the interactions of these fragments *in vitro*.

- **Aim 2: Determine the structure of the amino-terminal regions of AgI/II.** Protein fragments containing the N-terminal regions were expressed, purified, and crystallized. The structure of the N-terminal regions was determined using X-ray crystallography. The N-terminal structure (A<sub>3</sub>VP<sub>1</sub>), which contained a single alanine- and proline-rich repeat, was solved using molecular replacement. The structure containing a single A- or P-repeat was then used to build a model for all three repeats of the A- and P- regions that encompasses the majority of the amino-terminus of AgI/II. Circular dichroism and analytical ultracentrifugation compared the characteristics of AgI/II found in solution against the determined crystal structures.
- **Aim 3: Determine the structure of the carboxy-terminal regions of Ag I/II.** The structure of the highly conserved AgI/II carboxy-terminus was determined using X-ray crystallography. This region was expressed, purified, and crystallized, and its structure was determined by multiple isomorphous replacement (MIR). The three carboxy-terminal domains identified by the structure were used for the further mapping of SAG adherence within the AgI/II carboxy-terminus. Carbohydrates of the cryo-protectant solution were identified interacting with the structure of the C<sub>12</sub> domains, revealing a putative carbohydrate binding region. Electron microscopy of native AgI/II was used to create a composite tertiary model for the elongated adhesin.

## 1.5 Organization of this dissertation

The following chapters contain two manuscripts describing the structural determination of AgI/II.

- The crystallographic determination of structures of the AgI/II amino-terminus and carboxy-terminus are described in detail separately within Chapters 2 and 3.
- Functional studies are partially described within both of these two Chapters.
- Following the manuscripts is a discussion on the structural and functional characterizations reported in these manuscripts and their implications.

These combined works now describe the first tertiary model of the AgI/II family of proteins as elongated fibrillar structures [25, 26].

## CHAPTER 2

# ELONGATED FIBRILLAR STRUCTURE OF A STREPTOCOCCAL ADHESIN ASSEMBLED BY THE HIGH AFFINITY ASSOCIATION OF ALPHA- AND PPII- HELICES

by

LARSON MR, RAJASHANKAR KR, PATEL MH, ROBINETTE RA, CROWLEY PJ,  
MICHALEK S, BRADY LJ, DEIVANAYAGAM C

*Proceedings of the National Academy of Sciences.* 2010 Mar 30; 107(30):5983-8

Copyright

2010

by

National Academy of Sciences of the United States of America

Used by permission

Format adapted (figure panels enlarged) for dissertation

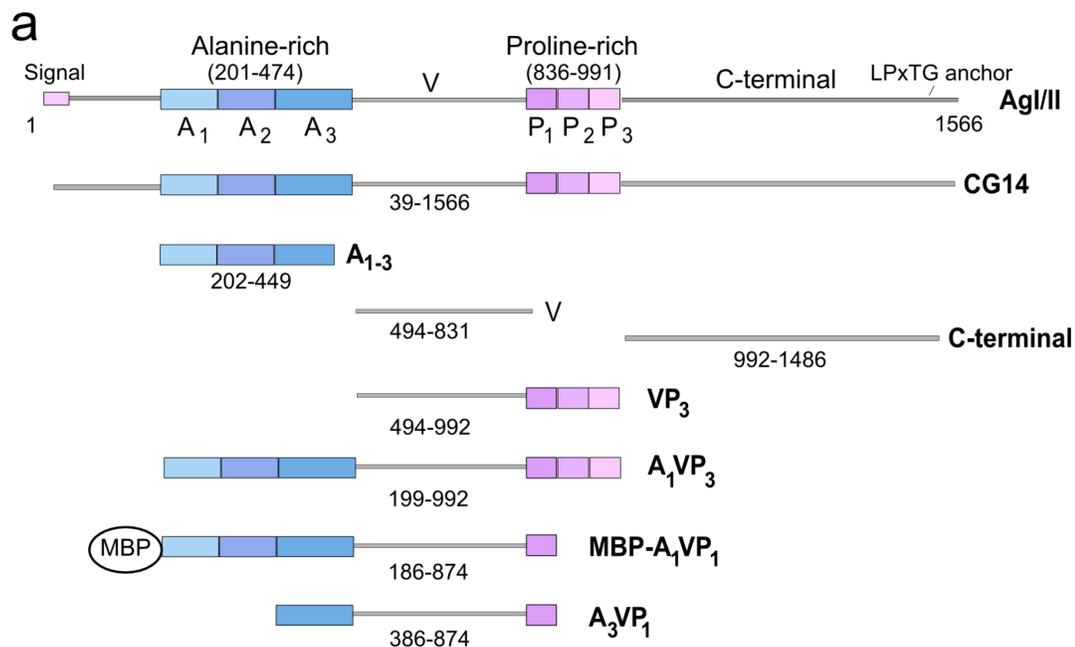
## Abstract

*Streptococcus mutans* Antigen I/II (AgI/II) is a cell surface-localized protein adhesin that interacts with salivary components within the salivary pellicle. AgI/II contributes to virulence and has been studied as an immunological and structural target, but a fundamental understanding of its underlying architecture has been lacking. Here we report a high-resolution (1.8 Å) crystal structure of the A<sub>3</sub>VP<sub>1</sub> fragment of *S. mutans* AgI/II that demonstrates a novel fibrillar form (155 Å) through the interaction of two non-contiguous regions in the primary sequence. The A<sub>3</sub>-repeat of the alanine-rich domain adopts an extended  $\alpha$ -helix that intertwines with the P<sub>1</sub>-repeat polyproline type II (PPII) helix to form a highly extended stalk-like structure heretofore unseen in prokaryotic or eukaryotic protein structures. Velocity sedimentation studies indicate that full-length AgI/II that contains three A/P repeats extends over 50 nanometers in length. Isothermal titration calorimetry revealed that the high affinity association between the A<sub>3</sub> and P<sub>1</sub> helices is enthalpically driven. Two distinct binding sites on AgI/II to the host receptor salivary agglutinin (SAG) were identified by surface plasmon resonance. The current crystal structure reveals that AgI/II family proteins are extended fibrillar structures with the number of alanine- and proline-rich repeats determining their length.

## Introduction

*Streptococcus mutans* is the causative agent of human dental caries (1) and its protein adhesin Antigen I/II (AgI/II) is a known target of protective immunity (2). AgI/II family molecules are expressed by numerous oral streptococci (3) and homologs have also been identified in the invasive pathogens *Streptococcus pyogenes* and *Streptococcus agalactiae* (4). In addition to mediating adhesion to the tooth surface (5), AgI/II influences biofilm formation (6), promotes collagen-dependent bacterial invasion of dentin (7), and mediates adherence to human epithelial cells (8). Elimination of AgI/II results in decreased virulence (9), but despite three decades of study a mechanistic understanding of the functional properties of the molecule has been stymied by a lack of understanding of its structure.

Originally identified as AgI/II (10) (also called P1, PAc, or SpaP), members of this protein family contain between 1310 to 1653 amino acids (aa) beginning with an amino-terminal signal motif that directs secretion, followed by the A-, V-, and P-regions (Fig. 2.1). The A-region typically consists of 3-4 alanine-rich repeats (82 residues each) with 23-30% alanine content. The P-region has 3-4 proline-rich repeats (39 residues each) with ~35% proline content. Nested between the A- and P-repeats is a segment commonly referred to as the V- or variable-region which contains within it a stretch of ~100 amino acids where most of the sequence variation among *S. mutans* AgI/II molecules is clustered (11). The crystal structure of the V-region adopts a globular  $\beta$ -stranded “super-sandwich” fold (12). Finally, the carboxy-terminus contains the LPxTG sortase motif for covalent anchorage to the cell wall (13). *S. mutans* AgI/II possesses both low and high affinity binding sites for salivary agglutinin (SAG) (14), a 600 kDa oligomeric protein complex containing glycoprotein 340 (440 kDa), sIgA (25, 59, 88 kDa), and an unknown 80 kDa polypeptide (15). Studies aimed at developing active and passive immune therapies have analyzed segments of Ag I/II that contribute to bacterial adherence and cariogenicity (2, 16). Several anti-AgI/II



**b Alanine-rich repeats**

201 TAYEAKL AQYQADL AAVQKTN

|    |   |
|----|---|
| A1 | AANQ AAYQKAL AAYQAE L KRVQEANAAA AAYDTAV AANNIKNTEIAAANEIIRKRNATAK AEYETKL AQYQAE L KRVQEAN |
| A2 | AANE ADYQAKL TAYQTEL ARVQKANADA ATYEAAV AANNAKNAALTAENTAIKQRNENAK ATYEAL KQYEADL AAVKKAN    |
| A3 | AANE ADYQAKL TAYQTEL ARVQKANADA KAYEAAV AANNAANAALTAENTAIKKRNADAK ADYEAKL AKYQADL AKYQKDL   |

ADYFVKL 474

**c Proline-rich repeats**

834 PTPPVKPTA

|    |  |
|----|--|
| P1 | PTKPTYETEKPLKPAFVAPNYEKEPTPPTRTPDQAEPNK  |
| P2 | PTPPTYETEKPLEPAFVPEPSYEAEPTPPTRAPDQAEPNK |
| P3 | PTPPTYETEKPLEPAFVPEPSYEAEPTPPTPTPDQPEPNK |

PVEPTYE 968

Figure 2.1: *AgI/II* protein and constructs- (a) The primary sequence layout of *AgI/II* is shown at the top. The alanine-rich repeats are highlighted in shades of blue, as the proline-rich repeats are in shades of purple. Shown below are the fragments of *AgI/II* used in this study, including the crystallized fragment  $A_3VP_1$ . Residue numbers correspond to the primary sequence of *AgI/II* from *S. mutans* strain NG8 (Genbank Accession #GQ456171). (b) The sequence alignments of  $A_1$ ,  $A_2$ ,  $A_3$  repeats, where the 19 heptad motifs are highlighted. (c) The alignments of  $P_1$ ,  $P_2$ , and  $P_3$  repeats, where the proline residues are shown in red.

monoclonal antibodies (MAbs) recognize complex conformational epitopes encompassing non-contiguous sequences within the A- and the P-repeats (16, 17) indicating that these regions are in close proximity, but the exact nature of the intra-molecular interaction required to achieve a functional adhesin was undetermined.

Here we report a high-resolution (1.8 Å) crystal structure of the A<sub>3</sub>VP<sub>1</sub> fragment of *S. mutans* AgI/II that demonstrates a novel fibrillar structure (155 Å) formed by the intimate association of two widely separated segments within the primary sequence. The A<sub>3</sub>-repeat of the alanine-rich domain was found to adopt a long  $\alpha$ -helical structure that intertwines with the P<sub>1</sub>-repeat polyproline type II (PPII) helix to form a highly extended stalk. Competition ELISA experiments employing multiple adherence-inhibiting monoclonal antibodies (MAbs) confirmed that the crystallized fragment represented a functional structure. Thermodynamic quantitation demonstrated a high affinity interaction between the  $\alpha$ - and PPII helices. Velocity sedimentation studies indicated that together the three A/P repeat units comprised almost the entire length of the intact molecule. Adherence studies identified two distinct binding sites on AgI/II to the host receptor salivary agglutinin (SAG). Finally, we propose a model for AgI/II's interaction with SAG.

## Results

### Crystallization and Overall Structure of A<sub>3</sub>VP<sub>1</sub>.

A<sub>3</sub>VP<sub>1</sub> was crystallized in the P2<sub>1</sub> space group. A similar crystallization condition augmented with 50 mM fructose crystallized in the P2<sub>1</sub>2<sub>1</sub>2 space group. The A<sub>3</sub>VP<sub>1</sub> structures (Fig. 2.2, panel c) from each space group superimpose with an average RMSD of 0.875 Å, indicating high structural similarity. A<sub>3</sub> is an extended  $\alpha$ -helix (110 residues) and adopts torsion angles akin to the canonical alpha-helix geometry, followed by the globular V-region that is situated at the apex of the molecule. P<sub>1</sub> (residues 834-874) folds over to

interact with  $A_3$ , and conforms to the PPII helix geometry, with backbone torsion angles that are characteristic of PPII helices ( $\phi$  of  $-75^\circ$  and the  $\psi$  of  $145^\circ$ ) with 3 residues per turn (18). Viewed through an imaginary axis along this PPII helix, the  $\alpha$ -helical segment wraps  $180^\circ$  around the axis to form a left-handed supercoiled structure. The resultant interaction forms an extended ( $\sim 155$  Å) fibrillar structure.

### **Differences between Crystal Forms.**

When the unique molecules of the asymmetric unit from the  $P2_1$ , and the  $P2_12_12$  datasets were superimposed by strictly anchoring the residues over the V-region, a latitudinal shift along the helical axis was observed (Fig. 2.2, panel c). The residues of  $A_3$  (385-474) and  $P_1$  (834-874) show temperature factors increasing with distance from the V-region, indicative of a global flexibility of the extended stalk. Angular deviation in the orientation of the helices of approximately  $7.5^\circ$  between the 2 unique molecules of the  $P2_1$  crystal begins in the vicinity of the kink at proline residue 471, resulting in a displacement of  $17.0$  Å between the termini of the extended stalks. These differences in orientations of the helices resulted in alternative packing with a larger b-axis cell dimension in the  $P2_1$  as compared to the  $P2_12_12$  crystals.

### **Analysis of $A_3$ - $P_1$ interactions.**

Notable within  $A_3$  were seven heptad repeats consisting of an  $AxYxAx[LV]$  motif (Fig. 2.3, top), where every other residue is hydrophobic. The 1<sup>st</sup>, 3<sup>rd</sup>, 5<sup>th</sup>, and 7<sup>th</sup> residues are highly conserved with the remaining positions filled with charged residues. The heptads are a defining signature of the A-region with 19 present across all three repeats (Fig. 2.1, b). The  $A_3$  heptads interact with  $P_1$  in a “knobs-into-holes” fashion (Fig. 2.4, panel a) with the third tyrosine residue on each heptad facing inwards into a hydrophobic pocket formed

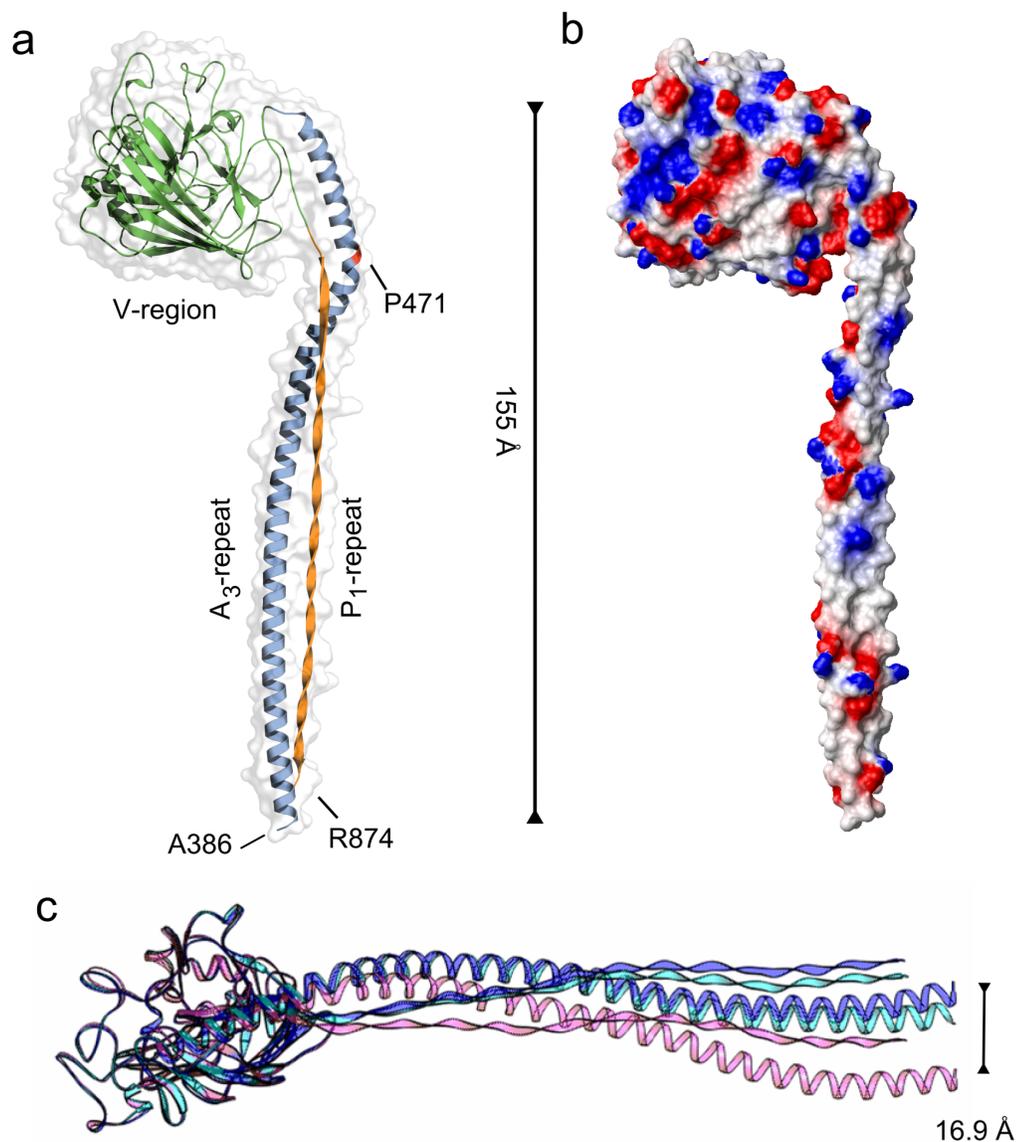


Figure 2.2: *AgIII* layout and  $A_3VP_1$  structure overview- (a) Ribbon diagram of the crystal structure of  $A_3VP_1$  generated by Pymol (48). The alpha helix from  $A_3$  and the PPII helix from the  $P_1$  regions interact to form the 155 Å stalk. (b) Electrostatic map of  $A_3VP_1$  generated using MolMol (49). The stalk formed by the association of the  $A_3$  and  $P_1$  helices displays an extensive hydrophobic surface. (c) Superposition of the unique molecules from the  $P2_1$  and  $P2_12_12_1$  crystal forms by strictly anchoring over the V-region shows  $\sim 17$  Å displacement at the termini of the helices.

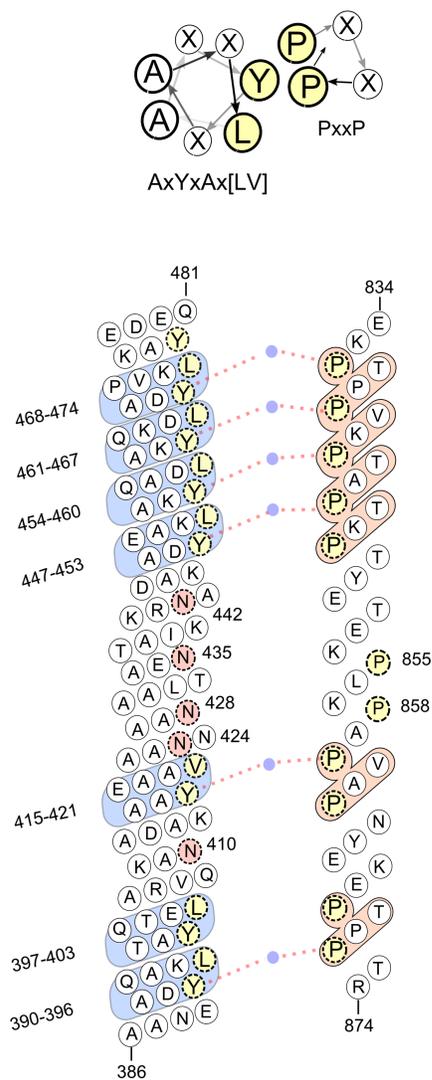


Figure 2.3: *Interactions between the  $A_3$  and  $P_1$  helices*- The helical wheel diagram (top) shows the interaction between the  $A_3$ 's heptad motif 'AxYxAx[LV]' and  $P_1$ 's 'PxxP' motif. The helical net of the interactions between the  $\alpha$ - and PPII helices is also shown, where the  $A_3$ -repeat heptad motifs are highlighted in blue and the  $P_1$  repeat PxxP motifs are highlighted in orange. Asparagine residues within the  $A_3$ -region, intervening the heptads, that are involved in hydrogen bonding with the  $P_1$  PPII helix are highlighted in red. The conserved tyrosines and leucines (highlighted in yellow) of the heptad sequences are nestled between prolines of the P-repeat. Additionally, the phenolic oxygen atom of the tyrosine residues participate in water-mediated hydrogen bonding (red-dashed lines).

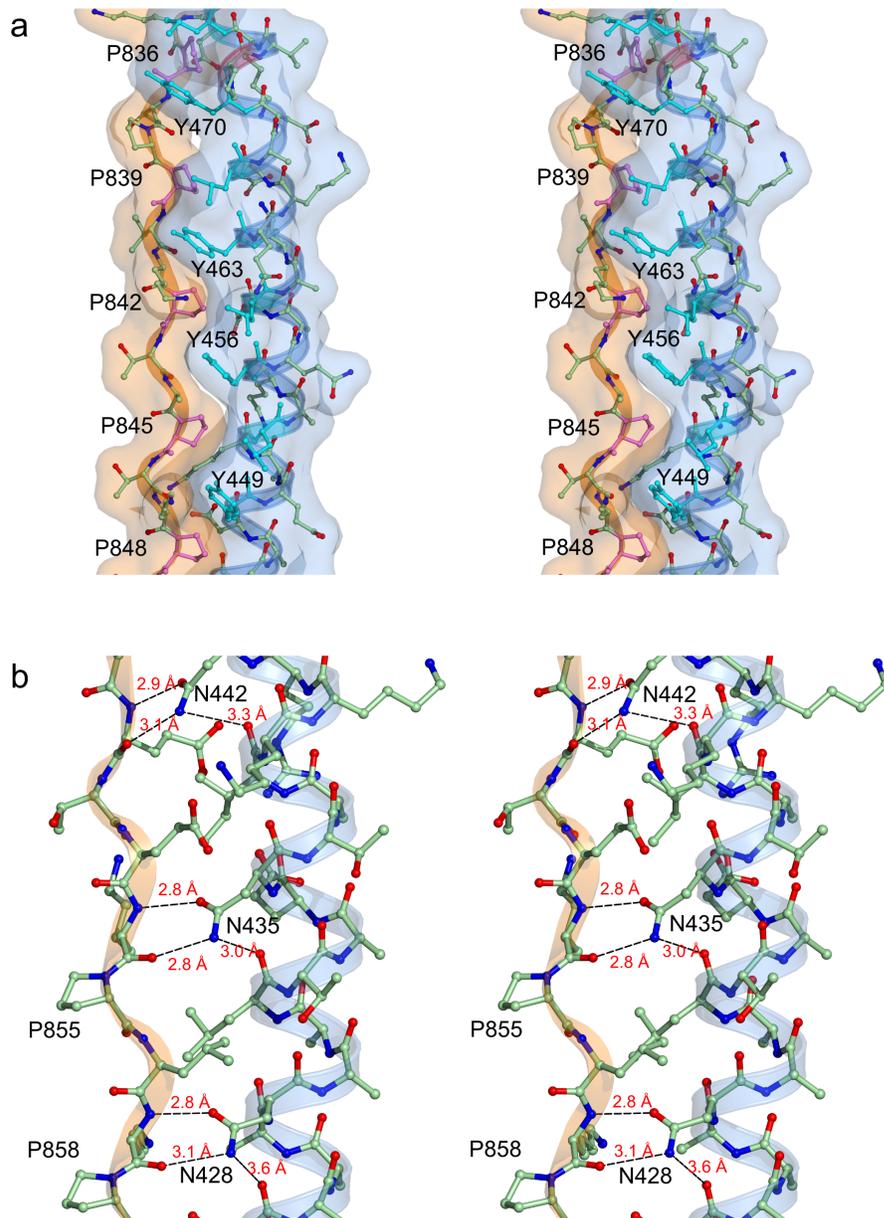


Figure 2.4: *Stereo images of the  $A_3$  and  $P_1$  helices-* (a) The stereo diagram of the heptad interactions shows that tyrosine sidechains nestle between the prolines in a knobs-in-holes interaction, which is highlighted by surface plots for the A- and P- repeats. (b) The stereo diagram of the region intervening the heptads shows a dominant direct hydrogen bonding between the asparagine side chains of  $A_3$  and the main chain oxygen and nitrogen of the PPII helix. Two prolines (Pro855 and Pro858) break this pattern and face outward.

by the P<sub>1</sub> proline sidechains at 'PxxP' motifs. A second hydrophobic leucine or valine residue at the seventh position also faces into this hydrophobic pocket. These interactions are reminiscent of those within the 36-amino acid avian pancreatic polypeptide (Fig. 2.12), in which an 8-residue PPII helix interacts with a 17-residue  $\alpha$ -helix (19). Similar to classical coiled-coil motifs, residues within A<sub>3</sub> display quasi-periodicity, where hydrophobic residues are alternately three and four positions apart. Many of the A<sub>3</sub>-P<sub>1</sub> interactions are hydrophobic as indicated from surface accessibility measurements of the A<sub>3</sub> and P<sub>1</sub> helices (2515.2 Å:920.0 Å, nonpolar:polar contact-surface). Regions that intervene between the heptads are dominated by direct hydrogen bonding between asparagine residues on A<sub>3</sub> and the main chain oxygens and nitrogens of P<sub>1</sub> (Fig. 2.4, panel b). The heptad- and intervening sequences, and PxxP-motifs, are highly conserved in A<sub>1</sub>A<sub>2</sub> and P<sub>2</sub>P<sub>3</sub>, respectively, suggesting similar interactions are maintained in all repeats.

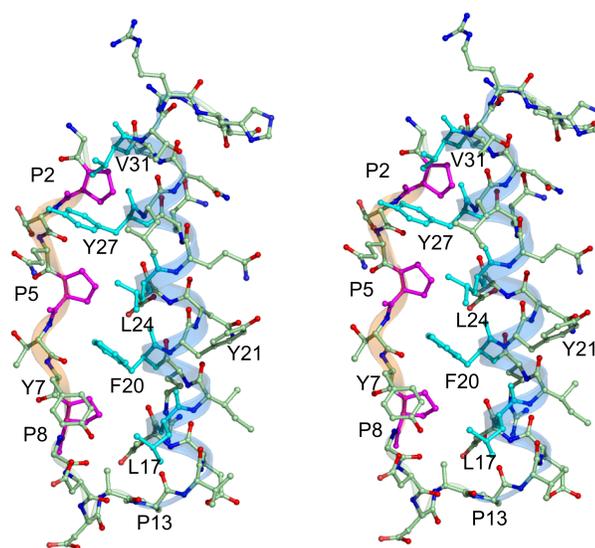


Figure 2.5: *Avian pancreatic peptide*- Stereo representation of the crystal structure of the avian pancreatic peptide (APP) resolved at 0.99 Å resolution (PDB 2BF9). Figure shows the short  $\alpha$ -helix (highlighted in blue) and the short PPII-helix (highlighted in orange). The color schemes adopted are similar to Fig 2.4 in the main text to allow for comparisons. The residues colored in cyan namely, F20, Y26, L17, L24 and V31 in this structure, are homologous to the heptad repeats observed in  $A_3VP_1$  crystal structure. Similarly the PxxP motifs (colored in magenta) are homologous to the  $A_3VP_1$  crystal structure.

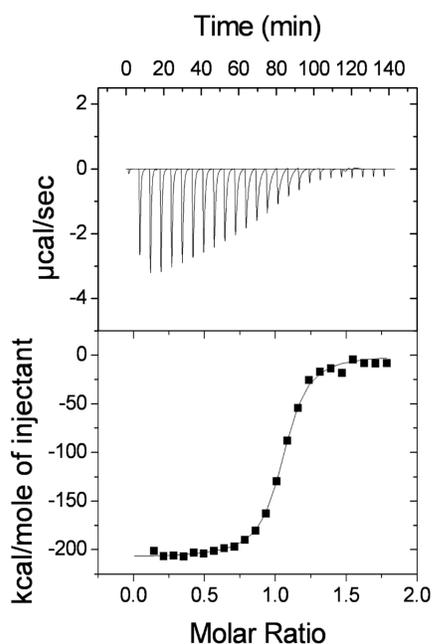


Figure 2.6: *Calorimetry of the A/P interaction*- Top: Calorimetric measurements at 20 °C, where 25 injections of 10  $\mu\text{L}$   $\text{A}_{1-3}$  were made into the cell containing  $\text{VP}_3$ . Bottom: The energy (kcal/mol) released during each injection.

Table 2.1: Isothermal titration calorimetry measurements of the A-P interaction

| Temp<br>°C | $\Delta\text{G}$<br>(kJ /<br>mol) | $\text{K}_a$<br>( $\text{M}^{-1} \times 10^6$ ) | $\text{K}_d$ (nM) | $\Delta\text{H}$ (kJ / mol) | $\Delta\text{S}$ ( $\text{J K}^{-1}$<br>$\text{mol}^{-1}$ ) | $n$               | $\Delta\text{C}_p$ ( $\text{kJ K}^{-1}$<br>$\text{mol}^{-1}$ ) |
|------------|-----------------------------------|---|-------------------|-----------------------------|---|-------------------|--|
| 15         | -36.6                             | $4.26 \pm 0.48$                                 | $234 \pm 26$      | $-808.3 \pm 10.2$           | -2677.8   | $0.858 \pm 0.008$ | $-7.95 \pm 0.58$   |
| 20         | -40.4                             | $15.9 \pm 1.3$                                  | $62.9 \pm 5.1$    | $-870.3 \pm 5.3$            | -2828.4   | $1.030 \pm 0.004$ |  |
| 25         | -44.4                             | $61 \pm 4$                                      | $16.5 \pm 1.1$    | $-892.4 \pm 2.7$            | -2845.1   | $1.050 \pm 0.002$ |  |
| 30         | -49.2                             | $310 \pm 60$                                    | $3.27 \pm 0.64$   | $-933.5 \pm 5.5$            | -2916.2   | $0.822 \pm 0.002$ |  |

## **Energetics of the A- P interaction.**

Isothermal titration calorimetry was used to measure the strength of association between the A and P repeats. The titration of VP<sub>3</sub> by A<sub>1-3</sub> (Fig. 2.6) shows that enthalpically driven binding occurs between A<sub>1-3</sub> and VP<sub>3</sub> with a dissociation constant of  $62.9 \pm 5.1$  nM at 20 °C (Table 2.1). The A<sub>1-3</sub> and VP<sub>3</sub> binding had a large enthalpy change ( $\Delta H = -870.3 \pm 5.3$  kJ/mol) accompanied by a large change in entropy ( $T\Delta S = -827.5$  kJ/mol), with the estimated  $\Delta G_{\text{binding}} = -40.4$  kJ/mol. Measurements of the VP<sub>3</sub> and A<sub>1-3</sub> interaction at multiple temperatures determined that the heat capacity ( $\Delta C_p$ ) of binding as  $-7.95 \pm 0.58$  kJ K<sup>-1</sup> mol<sup>-1</sup>.

## **AgI/II is a Highly Extended Molecule.**

The linear dimensions of AgI/II and its fragments in solution were estimated in velocity ultracentrifugation experiments. The sedimentation coefficients and frictional ratios of A<sub>1-3</sub>, A<sub>1</sub>VP<sub>3</sub>, A<sub>3</sub>VP<sub>1</sub>, VP<sub>3</sub> and full length AgI/II (CG14) exhibited high axial ratios for both prolate and oblate ellipsoids (Table 2.2). The ratios of the A<sub>3</sub>VP<sub>1</sub> prolate ellipsoid axes predict an elongated shape similar to that of the A<sub>3</sub>VP<sub>1</sub> crystal structure. Additionally, the ratio of axes of the A<sub>1</sub>VP<sub>3</sub> prolate ellipsoid is nearly three times of A<sub>3</sub>VP<sub>1</sub>, indicating that AgI/II forms an extended stalk of over 50 nanometers, consistent with the dimensions of the AgI/II “fuzzy coat” visualized on the surface of *S. mutans* by immunoelectron microscopy twenty years ago (20). A<sub>1-3</sub> has elongated dimensions similar to the AP stalk, and VP<sub>3</sub> is characterized by a shorter length. All the polypeptides characterized are monomeric, except A<sub>1-3</sub> that exists as a dimer.

Table 2.2: Sedimentation velocity measurements

| Protein                        | Theoretical MW (kDa) | Fit MW (kDa) | Fit rmsd (OD) | s <sub>20</sub> (S) | f/f <sub>0</sub> | Stokes radius (nm) | Prolate ellipsoid a/b ratio | Prolate ellipsoid dimensions (nm x nm) |
|--------------------------------|----------------------|--------------|---------------|---------------------|------------------|--------------------|-----------------------------|--|
| A <sub>3</sub> VP <sub>1</sub> | 56.54                | 55.83        | 0.004         | 3.197               | 1.642            | 4.16               | 7.47                        | 22.70 x 3.04                           |
| A <sub>1</sub> VP <sub>3</sub> | 88.25                | 92.93        | 0.005         | 2.913               | 2.548            | 7.61               | 24.56                       | 59.33 x 2.42                           |
| CG14                           | 167.52               | 138.48       | 0.004         | 3.796               | 2.536            | 8.70               | 24.52                       | 67.82 x 2.77                           |
| VP <sub>3</sub>                | 56.53                | 56.61        | 0.005         | 2.840               | 1.872            | 4.76               | 12.45                       | 30.65 x 2.46                           |
| A <sub>1-3</sub>               | 27.63                | 55.93        | 0.007         | 2.071               | 2.545            | 6.44               | 27.33                       | 51.56 x 1.89                           |

### Achievement of functional structure by A<sub>3</sub>VP<sub>1</sub>.

To confirm that the folding of A<sub>3</sub>VP<sub>1</sub> is representative of biologically active AgI/II, the ability of A<sub>3</sub>VP<sub>1</sub> and A<sub>1</sub>VP<sub>3</sub> to compete for binding of previously characterized anti-AgI/II MAbs against *S. mutans* cells was evaluated by ELISA (Fig. 2.7). MAbs 1-6F, 4-9D and 4-10A have all been shown to react with *S. mutans* cells (21) and inhibit *S. mutans* adherence to immobilized human salivary agglutinin (SAG) (22). MAb 1-6F binds within the V-region. The 4-9D epitope also maps within the V-region in close proximity to the A-repeats. The discontinuous 4-10A epitope is achieved by the interaction of A- and P-repeat containing fragments (16, 23). In contrast, MAb 3-10E reacts with *S. mutans* cells, but does not inhibit bacterial adherence to SAG (21, 22). As expected, A<sub>3</sub>VP<sub>1</sub> and A<sub>1</sub>VP<sub>3</sub> were similar to CG14 in their ability to compete for binding of the adherence-inhibiting MAbs 1-6F, 4-9D and 4-10A, whereas the C-terminal fragment did not demonstrate substantial competition. On the other hand, A<sub>3</sub>VP<sub>1</sub> and A<sub>1</sub>VP<sub>3</sub> did not competitively inhibit binding of the non-adherence-inhibiting MAb 3-10E, while the full-length CG14 molecule did. Taken together, these experiments demonstrate that A<sub>3</sub>VP<sub>1</sub> and A<sub>1</sub>VP<sub>3</sub> reflect a functional structure in that the epitopes recognized by three different anti-AgI/II MAbs known to inhibit bacterial adherence were replicated in these polypeptides to a degree comparable to that of the full length adhesin.

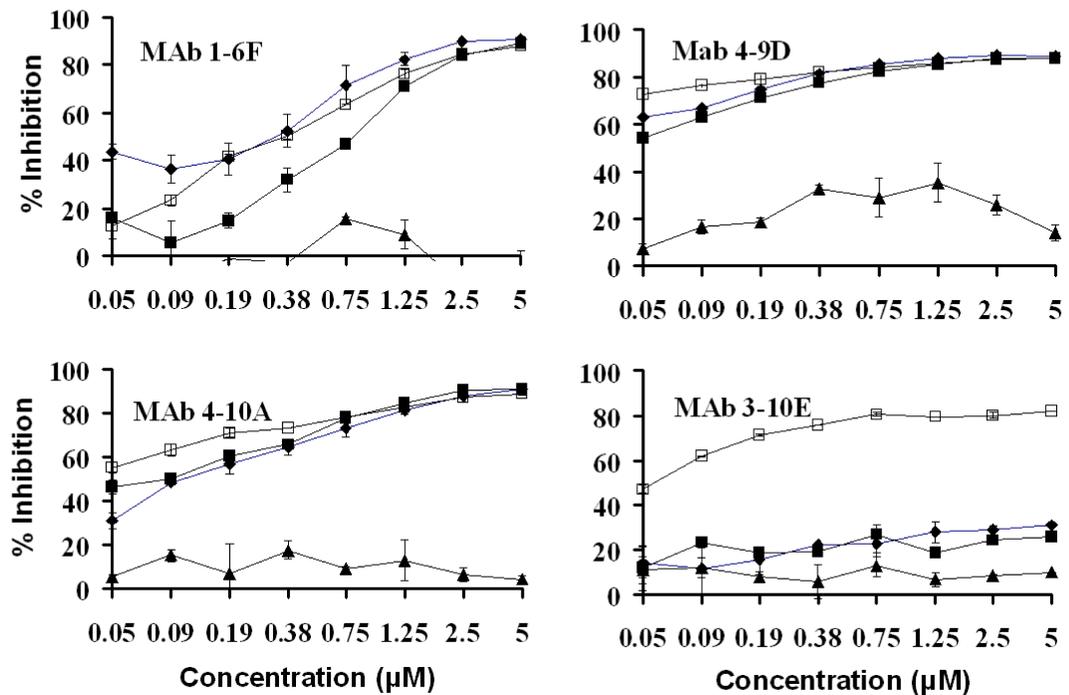


Figure 2.7: *Inhibition of binding of anti-AgI/II MAbs to S. mutans whole cells*- Two-fold serial dilutions of AgI/II polypeptides, beginning at 5  $\mu$ M, were incubated with each indicated MAb. Competition ELISA results are shown as the percent inhibition of binding of the indicated MAb to *S. mutans* whole cells in the presence of each polypeptide. Error bars indicate standard deviations. CG14 (open squares), A<sub>3</sub>VP<sub>1</sub> (closed diamonds), A<sub>1</sub>VP<sub>3</sub> (closed squares), C-terminal construct (closed triangles).

To further confirm adherence properties of AgI/II fragments, binding to immobilized SAG was evaluated using surface plasmon resonance (SPR) (Fig. 2.8). The binding experiments were conducted using a buffer augmented with calcium that is necessary for AgI/II interaction with SAG (14), while under a laminar fluid flow as found in the oral environment. CG14 and A<sub>3</sub>VP<sub>1</sub> both interacted strongly with SAG, while the isolated V-region demonstrated diminished binding activity. As was previously reported (24), the C-terminal fragment also bound to SAG. Modeling the association and dissociation of the AgI/II proteins with SAG using the 1:1 Langmuir binding model led to estimated K<sub>A</sub> values of 3x10<sup>7</sup> for A<sub>3</sub>VP<sub>1</sub>, 3x10<sup>7</sup> for CG14, 5x10<sup>6</sup> for the C-terminal construct, and 3x10<sup>5</sup> for the V-region. The independent adherence of two non-overlapping AgI/II fragments is consistent with previous experimental evidence indicative of multiple binding sites (14). Now viewed in the context of the A<sub>3</sub>VP<sub>1</sub> crystal structure, two distinct regions with independent affinity for SAG appear to be widely separated by the long AgI/II stalk.

## Discussion

Early work on fibrous alpha-keratin predicted formation of extended structures with coiled-coiled  $\alpha$ -helices that was later witnessed in a variety of proteins (25). Similarly, a model for fibrous collagen with three left-handed polyproline type-II (PPII) helices forming a right handed supercoil was proposed (26) and then supported experimentally (27). However, the formation of a single unified hybrid structure comprised of both  $\alpha$ - and PPII helices was unknown. The crystal structure of A<sub>3</sub>VP<sub>1</sub> now provides evidence for a novel interaction between  $\alpha$ - and PPII-helices that results in an elongated fibrillar form. The A-P interactions are well maintained in the molecules from two different space groups. The entire stalk flexes as a unit from the hinge-like area near proline residue 471 in the two crystal

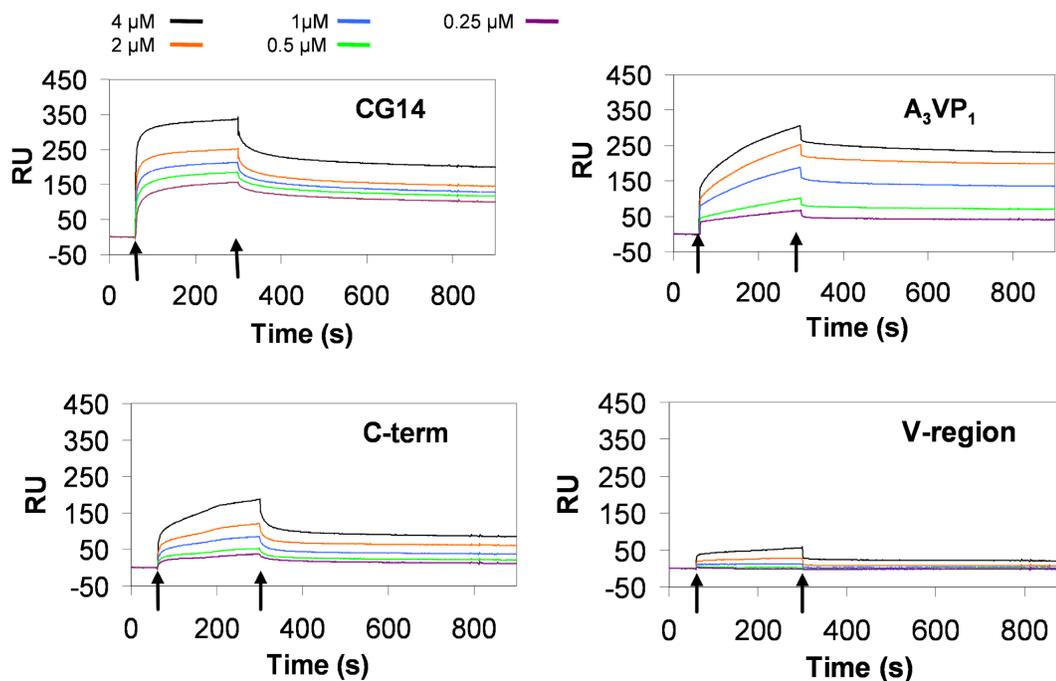


Figure 2.8: *SPR studies of AgI/II fragments with immobilized SAG*- Full-length CG14, A<sub>3</sub>VP<sub>1</sub>, C-terminal, or V-Region fragments (analytes) were flowed over a CM5 chip coated with covalently attached SAG. Analyte proteins ranged from 0.25 μM - 4 μM concentrations. Arrows indicate injection starting and stopping points. During the 4-minute injection of each polypeptide, an increase in the RU is observed as AgI/II polypeptides bind to SAG. CG14 and A<sub>3</sub>VP<sub>1</sub> showed the maximal binding responses (one RU is equivalent to one picogram per square millimeter of sensor surface).

forms of  $A_3VP_1$ . The association of the  $\alpha$ - and PPII-helices in solution is well supported by the high affinity interaction measured between the  $A_{1-3}$  and  $VP_3$  fragments. A negative change in the heat capacity ( $\Delta C_P$ ) supports the burial of hydrophobic surfaces in the association of the helices. The thermodynamic properties of the enthalpically driven association of the alanine-rich and proline-rich repeats are similar in sign and magnitude as observed in the formation of triple-helical collagen-like peptides (28), and in the binding of SH3 to short PPII-conforming peptides (29). In the latter case, the enthalpy and entropy changes were of opposite sign from that expected of a primarily hydrophobic interaction. To explain differences between expected and measured changes of entropy and enthalpy, binding of SH3 was proposed to occur through the folding of the short PPII peptide resulting in a large enthalpic energy release and decrease in entropy (29). It is possible that the  $A_{1-3}$  and  $VP_3$  interaction may follow a similar mechanism of association that is dependent on folding of the PPII helix. Differences in disordered content between  $VP_3$  and  $A_1VP_3$  observed in circular dichroism studies (Table 2.3) support the hypothesis that the PPII helix undergoes a conformational change during its association with the A-repeats. In addition, the shorter estimated length of  $VP_3$  (30 nm) compared to  $A_1VP_3$  (59 nm) derived from velocity sedimentation experiments is consistent with such a change in its conformational state. In contrast, the CD spectrum of the unpaired A-repeats ( $A_{1-3}$ ), which have nearly 100%  $\alpha$ -helical content (Table 2.3), suggest that the A-repeats likely adopt their helical fold independently and prior to association with the P-repeats. The formation of dimers of the  $A_{1-3}$  fragment (Table 2.2) likely result in stabilization of the A-repeats'  $\alpha$ -helix.

While AgI/II has been implicated in *S. mutans* aggregation (30), a multimeric state was not prevalent in constructs with paired A/P repeats. The  $A_3VP_1$ ,  $A_1VP_3$ , and CG14 fragments studied by analytical ultracentrifugation each had a single large peak with their sedimentation coefficient distributions corresponding to the mass of a monomeric molecule. However, packing of the  $A_3VP_1$  fragment within the crystal (Fig. 2.9, panel a) resulted in

numerous hydrophobic contacts between symmetry related molecules, particularly along the A- and P-repeats, suggesting a possible mechanism of aggregation of AgI/II molecules. A notable symmetry interaction involved two prolines (Pro855 and Pro858, Fig 2.9, panel b) of the PPII helix that broke the dominant pattern to face outward and abut with the surface of another symmetry related A<sub>3</sub> (2.5, panel b). The A-repeats of AgI/II are known to bind collagen (8). This symmetry-related interaction may mirror a reported biological interaction in which the collagen proline and hydroxyproline residues of the classical GXY motif face outward to interact with the A-repeats.

Extrapolation based upon the crystal structure of the A<sub>3</sub> (386-474)-P<sub>1</sub> (834-847) interaction predicts that the continued association of A<sub>2</sub> (304-385) with P<sub>2</sub> (875-913), and subsequently A<sub>1</sub> (222-303) with P<sub>3</sub> (914-952), would result in a molecule that exceeds 50 nm in length. In support of this prediction, the estimated dimensions of full-length AgI/II deduced from velocity sedimentation experiments indicate that it is over 50 nm long. In addition, crystals of A<sub>1-3</sub> (Fig. 2.9) diffracted poorly and displayed fiber-like diffraction patterns, but revealed the presence of one very long axis (> 600 Å), agreeing well with the dimensions predicted from ultracentrifugation. It is now apparent that AgI/II is a highly extended fibrillar structure formed by a long and continuous A-P association. This architecture would position the globular V-region at the tip of the stalk away from the cell surface, with the C-terminal region positioned near the cell surface and the pre-A-region in

Table 2.3: Predicted AgI/II secondary structures from circular dichroism (CONTIN/LL)

|   | <b>A<sub>1-3</sub></b> | <b>V</b>   | <b>VP<sub>3</sub></b> | <b>A<sub>3</sub>VP<sub>1</sub></b> | <b>A<sub>1</sub>VP<sub>3</sub></b> | <b>C-terminus</b> |
|---|------------------------|------------|-----------------------|------------------------------------|------------------------------------|-------------------|
| Total residues                                  | 248                    | 356        | 516                   | 517                                | 812                                | 516               |
| <b>α-Helical content % total (no. residues)</b> | 97.9 (243)             | 8.8 (31)   | 9.2 (47)              | 38.5 (199)                         | 61.7 (501)                         | 5.8 (30)          |
| <b>β-Sheet content % total (no. residues)</b>   | 1.9 (5)                | 34.2 (122) | 27.1 (140)            | 18.6 (96)                          | 4.2 (34)                           | 53.8 (278)        |
| <b>β-Turn content % total (no. residues)</b>    | 0.2 (0)                | 22.1 (79)  | 19.0 (98)             | 18.7 (96)                          | 16.1 (131)                         | 17.6 (91)         |
| <b>Other % total (no. residues)</b>             | 0.0 (0)                | 34.9 (124) | 44.6 (230)            | 24.3 (126)                         | 18.1 (147)                         | 22.9 (118)        |

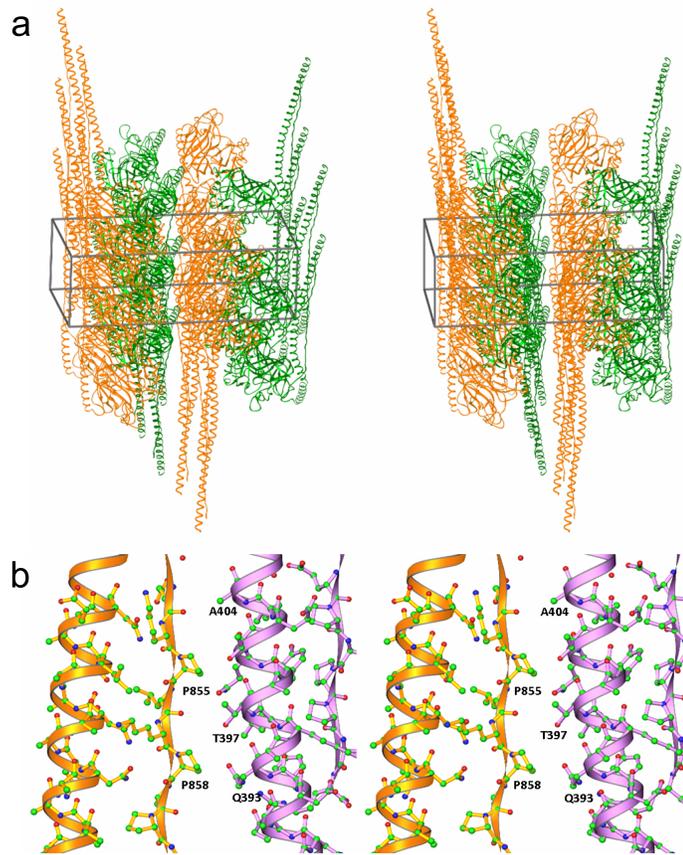


Figure 2.9: *Crystal Packing Diagram of  $A_3VP_1$* - (a) The unit-cell for the  $P2_12_12$  space-group crystals is delineated by the box. Molecules in the crystal pack with the V-regions stacking upon each other, while the long helical  $A_3/P_1$  units lie parallel and in a staggered alignment between the symmetry related molecules. (b) Two prolines point outwards from the  $P_1$ 's PPII helix, and interact with a neighboring symmetry related molecule  $A_3$  helix.

close proximity (Fig. 2.10, panel a). This model is consistent with the known C-terminal anchorage of LPxTG motif containing sortase substrates to the bacterial cell wall (13).

Corresponding well with the current  $A_3VP_1$  crystal structure, a single A-repeat adopting an alpha-helix was proposed to match in length a single P-repeat adopting a PPII helix (17). Alanine- and proline-rich repeats are highly conserved among AgI/II family members, although the total number of repeats varies between one and four (2.11) (31). From the current structure, it is likely that all AgI/II family members have elongated structures, with the overall length of the fibrillar proteins determined by the number of A- and P-repeats. For example, *S. intermedius* Pas possesses only one A- and one P- repeat and would consequently be predicted to be shorter (15-17 nm).

AgI/IIs relevant domains have been predicted based on primary sequence and studied in isolation (14, 24, 32, 33); however, the interactions between the A- and P-regions as a single extended functional unit now indicates that both these segments together would share roles previously ascribed to one or the other in isolation. In addition to the anti-AgI/II MAbs described in this study, antibodies that inhibit *S. mutans* adherence to SAG have also been mapped to the A-region (17, 34-37). Notably, two thirds of the surface area of the A/P stalk of  $A_3VP_1$  is formed by the  $A_3$ -sequence. Due to the highly repetitive nature of  $A_{1-3}$ , it is plausible that adherence-inhibiting antibodies recognize a repeated epitope and disrupt the binding via steric hindrance. SAGs gp340 also contains multiple tandem repeats of SRCR domains (38). While a structure of tandem repeats of these SRCR domains has not been determined, gp340 may be reminiscent of fibronectin, whose repeated domains display an elongated structure (39). There would then be a potential for a longitudinal interaction of SAG with the highly extended AgI/II.

Full-length AgI/II (CG14),  $A_3VP_1$  and the C-terminal fragment all demonstrated measurable adherence to immobilized SAG by surface plasmon resonance. Independent adherence to SAG by two non-overlapping AgI/II fragments is consistent with the multivalent

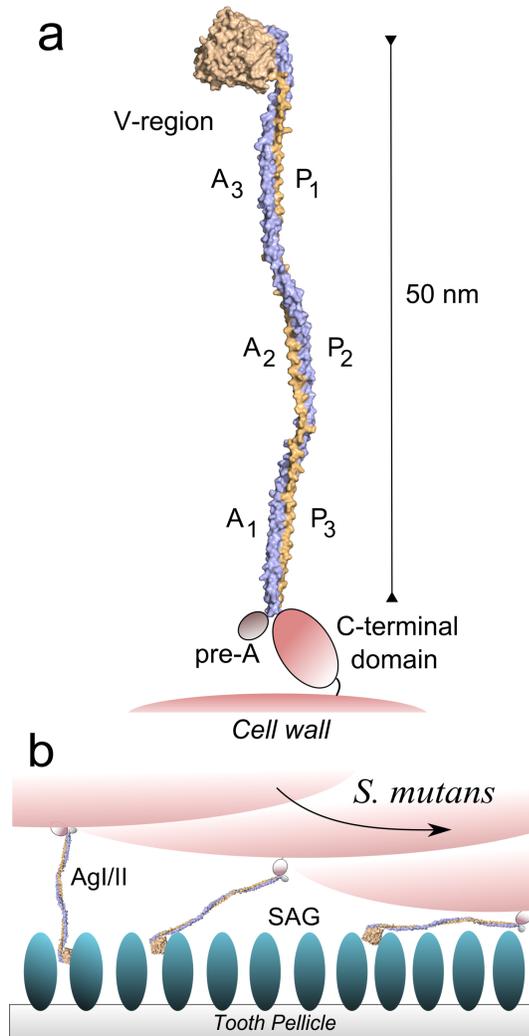


Figure 2.10: *Model of the AgI/II structure and predicted binding with SAG-* (a) The crystal structure of the AgI/II A<sub>3</sub>VP<sub>1</sub> region revealed an extended stalk formed by the A<sub>3</sub> and P<sub>1</sub> repeats. Ultracentrifugation studies predict that the A<sub>2</sub>-P<sub>2</sub> and A<sub>1</sub>-P<sub>3</sub> repeats extend the stalk to a length over 50 nm. (b) Adherence studies indicated the presence of two sites, one within A<sub>3</sub>VP<sub>1</sub> and another within the C-terminal region, which are widely separated in the AgI/II structure. The cartoon represents a possible model for binding to SAG, where interactions occur at both the distal end through the A<sub>3</sub>VP<sub>1</sub> region, and at a secondary adherence site mediated by the C-terminal domain.

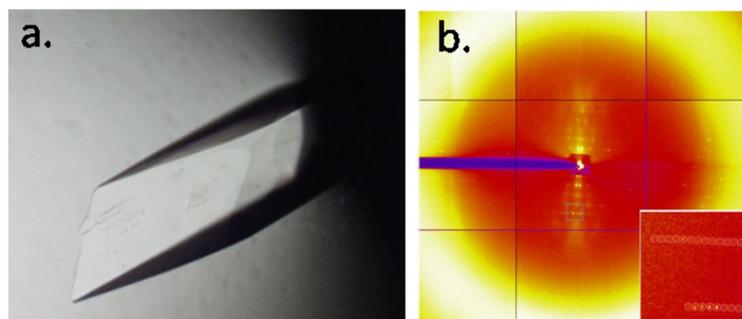


Figure 2.11: *Crystals of A<sub>1-3</sub>*- A<sub>1-3</sub> was purified as described in the Methods Section. (a) Large crystals of the A<sub>1-3</sub> were grown using the hanging drop vapor diffusion method. A droplet containing 2  $\mu\text{L}$  of protein (7.2 mg/ml) was mixed with an equal volume of the reservoir solution that contained, 9% PEG 8000, 100mM Tris pH 8.0, 50 mM L-Arginine and 10 mM EDTA. (b) However, when exposed to X-rays, we observed asymmetric diffraction, and the crystal parameters were estimated to be  $a = 69.96$ ,  $b = 671.43$  and  $c = 74.69$  with  $b = 91.8$ . One of the cell dimensions was  $> 600 \text{ \AA}$ , and most frames contained, a fiber like diffraction as seen in panel b. The asymmetric diffraction pattern, and the quick radiation damage to the crystals, made it impossible to collect a full data set, and they only diffracted at best to  $6 \text{ \AA}$ . Solvent content estimations using the statistical Matthews coefficient, indicate that there could be at least 12 monomers in the asymmetric unit and that the A<sub>1-3</sub> repeats form multimers within the crystal. From the structure of A<sub>3</sub>VP<sub>1</sub>, and ultracentrifugation studies the estimated of the length of  $> 50 \text{ nm}$  compares well with the estimated elongated cell dimensions of  $> 60 \text{ nm}$  observed in the A<sub>1-3</sub> crystals, indicating that they could pack in an elongated fashion along the b-axis.

binding detected by Scatchard analysis (14). In a simple model of binding based on its highly extended structure, one possibility is that high-affinity binding of AgI/II with SAG occurs via the apical fish-hook like structure observed within A<sub>3</sub>VP<sub>1</sub>, and an additional interaction occurring within the C-terminal region (Fig. 2.10 b). The similar affinities observed for the full length AgI/II and A<sub>3</sub>VP<sub>1</sub> suggest that AgI/II may not simultaneously interact with a single molecule of SAG at both the A<sub>3</sub>VP<sub>1</sub> and C-terminal sites, possibly due to competition or to large distance existing between the binding sites on A<sub>3</sub>VP<sub>1</sub> and C-terminal region.

*Streptococcus pyogenes* M Protein is another bacterial surface protein, whose elongated shape is formed by an  $\alpha$ -helical coiled-coil dimer structure (40). The hyper-variable region at the N-terminus forms a conformational binding surface that recognizes the human plasma protein C4BP (41). Similarly, the *Staphylococcus aureus* surface protein CNA has the N-terminal collagen binding region extended away from the cell surface by the B-repeats Inverted IgG domains. (42). It appears, therefore, that gram-positive bacteria have developed several variations of structural units that enable positioning of binding regions of adhesin proteins away from the cell surface. The crystal structure of A<sub>3</sub>VP<sub>1</sub> of *S. mutans* AgI/II displays a novel fibrillar architecture that provides insights into this widespread family of streptococcal surface proteins.

## Methods

### Construction of expression plasmids.

The primary sequence features of AgI/II and extents of the recombinant fragments used in this study are shown in Figure 2.1. Plasmids pDC20 and pRR2 that encode full-length and aa 84-1223 of AgI/II (2, 3), respectively, pCK1 that encodes A<sub>1</sub>VP<sub>3</sub> linked to an amino-terminal MBP tag (3) and pSBR (4) that encodes the alanine-rich repeats A<sub>123</sub>, were

used as templates for PCR. Construction and expression of the recombinant his-tagged full length AgI/II (CG14) was described previously (2). DNA primer sequences for PCR, with appropriate restriction sites for cloning underlined, are listed in Table 2.4. PCR reactions were carried out using the PTC-200 thermocycler (MJ Research) and the DeepVent proof-reading polymerase (New England Biolabs). DNA ligations were carried out with T4 DNA ligase (New England Biolabs). MBP-A<sub>1</sub>VP<sub>1</sub>, A<sub>1</sub>VP<sub>3</sub>, V, VP<sub>3</sub>, and C-term fragments were subcloned into pET23d (Novagen) using endonuclease sites BamHI and XhoI, while A<sub>1-3</sub> was subcloned into pET20b (Novagen) using endonuclease sites NcoI and XhoI. The resulting recombinant plasmids were used to transform *E. coli* XL1-Blue cells (Stratagene, Inc). Colonies were screened for plasmids with the correct size inserts, and these were confirmed by sequencing (UAB Sequencing Center). These plasmids were used subsequently to transform the *E. coli* BL21 ( $\lambda$ DE3) expression strain.

Table 2.4: PCR primers used in N-terminal structural study

| <b>Construct</b>                       | <b>Forward Primer</b>                                  | <b>Reverse Primer</b>                            | <b>Template Vector</b> |
|--|--|--|------------------------|
| <b>MBP-A<sub>1</sub>VP<sub>1</sub></b> | GCAAATGGATCCCC<br>TCGGCTCTCGCC                         | GATAACTCGAGCCT<br>TGTCGGCGGTGTTG<br>GCTCC        | pCK1                   |
| <b>A<sub>1</sub>VP<sub>3</sub></b>     | CGAGTGGATCCCCA<br>GTAAAACAGCTTAT<br>GAAGC              | GCAATACTCGAGAT<br>GAACAGTTGGTACA<br>GATGG        | pRR2                   |
| <b>VP<sub>3</sub></b>                  | CGAGTGGATCCAT<br>AAAAATGAAGACG<br>GAAACTTAACAGA<br>ACC | GCAATACTCGAGAT<br>GAACAGTTGGTACA<br>GATGG        | pRR2                   |
| <b>A<sub>1-3</sub></b>                 | ATATCCATGGGGGC<br>TTATGAAGCTAAAT<br>TGG                | ATATCTCGAGGTAA<br>TCAGCTTTAGCATC<br>CGC          | pSBR                   |
| <b>V</b>                               | CGAGTGGATCCAT<br>AAAAATGAAGACG<br>GAAACTTAACAGA<br>ACC | GCAATACTCGAGGG<br>TAACTTTAGGAACA<br>TTAACCGCACGG | pRR2                   |
| <b>C-term</b>                          | TATAAGGATCCATT<br>TCCATTACTTTAAA<br>CTAGC              | TTATACTCGAGTGA<br>ACTGTAAGTTACCC<br>CATTGACAG    | pDC20                  |

## **Expression and purification of A<sub>3</sub>VP<sub>1</sub>.**

MBP-A<sub>1</sub>VP<sub>1</sub> protein was expressed in the *E. coli* BL21 (λDE3) according to the manufacturers instruction. The cells were lysed using a Sonic Dismembrator Model 500 (Fisher Scientific) on ice, and the lysate was purified with a HisPrep Nickel column (Amersham Biosciences), followed by purification over an Amylose resin column (New England Biolabs). MBP-A<sub>1</sub>VP<sub>1</sub>-containing fractions were pooled and dialyzed overnight in 20 mM HEPES pH 7.5, 100 mM sodium chloride, and 2 mM CaCl<sub>2</sub>. The next day, Factor Xa protease (Sigma-Aldrich) was added (50 ng Factor Xa per mg MBP-A<sub>1</sub>VP<sub>1</sub>) and digestion was continued over 4 days at 22 °C. Factor Xa digestion resulted in a 54.5 kDa protein fragment (size determined by mass spectroscopy). Western blot analysis with mouse monoclonal anti-penta-histidine (Novagen) or rabbit anti-MBP (New England Biolabs) antibodies, followed by alkaline phosphatase-labeled goat anti-mouse or anti-rabbit IgG (Novagen), revealed that over-digestion had occurred at the amino-terminus, such that the resultant product comprises the third alanine-rich region (A<sub>3</sub>), the variable region (V), and the first proline-rich region (P<sub>1</sub>) (Fig. 2.12). Edman degradation analysis at the University of Florida Protein Core Facility confirmed Ala386 of AgI/II as the amino-terminal residue. This stable fragment was further purified with ion exchange over a MonoQ HR10/10 (Amersham Biosciences) and size exclusion over a Superdex 200pg 26/60 (Amersham Biosciences) column (Fig. 2.12). The A<sub>3</sub>VP<sub>1</sub> fragment was concentrated to 15 mg/mL by an Amicon concentrator (Amicon) under nitrogen gas.

## **Expression and purification of A<sub>1-3</sub>, V, VP<sub>3</sub>, A<sub>1</sub>VP<sub>3</sub>, and C-term polypeptides.**

A<sub>1-3</sub>, V, VP<sub>3</sub>, A<sub>1</sub>VP<sub>3</sub> and C-term fragments of AgI/II were expressed in *E. coli* BL21 (λDE3) similar to A<sub>3</sub>VP<sub>1</sub>. The recombinant *E. coli* cells were collected and lysed and

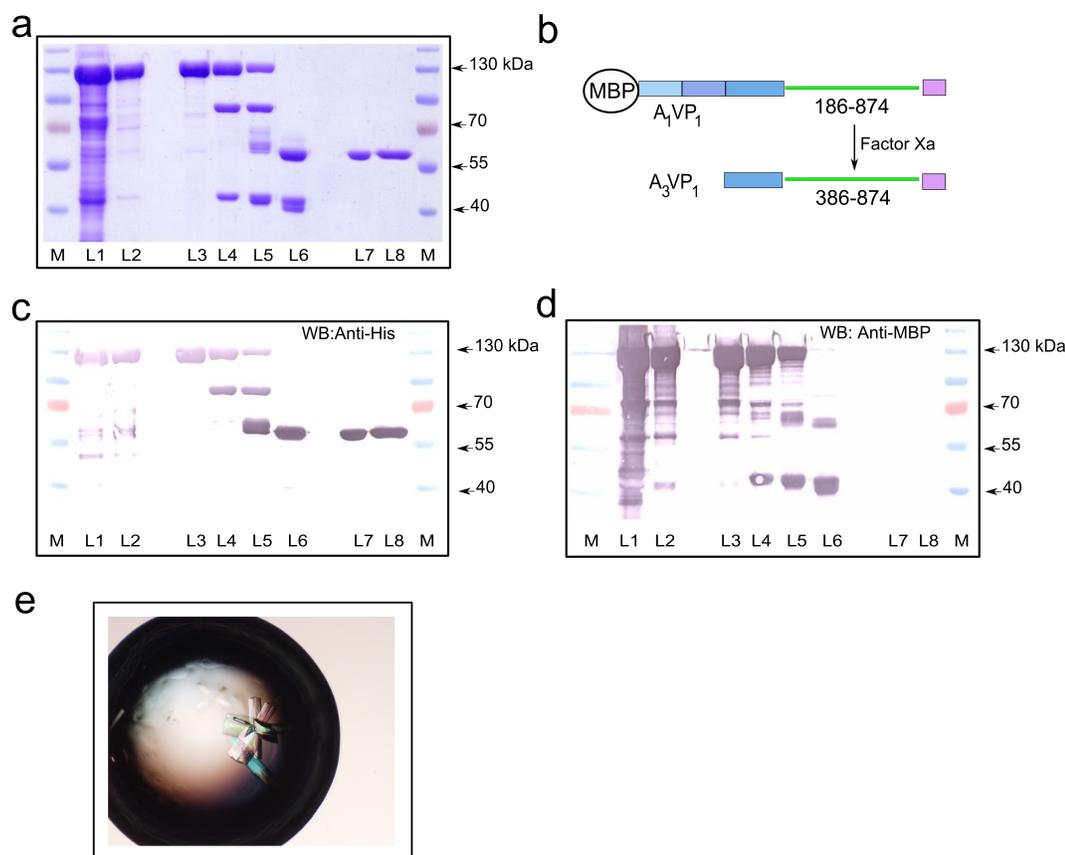


Figure 2.12: *Purification of  $A_3VP_1$* - (a) MBP- $A_1VP_1$  was purified from the initial lysate (L1) by nickel affinity- (L2) followed by MBP affinity-chromatography (L3), and the MBP-moiety sequentially cleaved for 24h (L4), 48h (L5), and 72h (L6) with Factor Xa. After cleavage, the remaining  $A_3VP_1$  was further purified over a MonoS cationic affinity column (L7) and polished by size exclusion chromatography (L8). The final product of Factor Xa cleavage of MBP- $A_1VP_1$  is diagrammed in (b). Western blots (WB) probed with an anti-histidine antibody (c) verified the intact carboxy-terminus, while reactivity with anti-MBP (d) indicated that cleavage occurred at the amino-terminus. The extent of proteolysis was evaluated by mass spectroscopy and Edman degradation confirmed the N-terminal sequence to correspond to  $A_3VP_1$ .  $A_3VP_1$  migrates with an apparent molecular mass of  $\sim 60$  kDa by SDS-PAGE with the molecular mass determined to be 54.5 kDa by mass spectroscopy. (e)  $A_3VP_1$  crystals formed in a 1:1 L hanging drop by vapor diffusion with the reservoir solution containing 30% (v/v) polyethylene-glycol monomethyl ether 2000, 200 mM ammonium sulfate, and 50 mM sodium cacodylate at pH 4.6.

AgI/II polypeptides were purified using a three-step protocol; lysates were sequentially subjected to a HisPrep affinity column (Amersham) followed by ion-exchange chromatography over a MonoQ/S column (Amersham) and then polished over a size exclusion column (Superdex 75/200pg 26/60-Amersham). The A<sub>1-3</sub> purification utilized the MonoS and Superdex 75 columns while all other proteins were purified using the MonoQ and Superdex 200 columns. Proteins were concentrated to ~5-15 mg/ml.

### **Crystallization and data collection.**

Crystals of A<sub>3</sub>VP<sub>1</sub> were optimized from condition 13 of Crystal Screen II (Hampton Research) using the hanging drop vapor diffusion methodology at 22 °C. Crystals were routinely grown from a reservoir solution containing 30% (v/v) polyethylene-glycol monomethyl ether 2000, 200 mM ammonium sulfate, and 50 mM sodium cacodylate at pH 4.6. The addition of 50 mM fructose in the reservoir solution further improved the size and appearance of the protein crystals. Crystals were flash frozen with 15% ethylene glycol and data were collected under a gaseous N<sub>2</sub> stream maintained at 100 °K in the NE-CAT beamline at APS/ANL, Chicago. Data were recorded on a Quantum CCD detector, using a wavelength ( $\lambda$ ) of 1.0 Å, exposure time of 2.0 sec, oscillation angle (*phi*) of 1.0°, and a distance (D) of 150 mm from the crystal to the detector. Diffraction data were integrated and scaled using the HKL2000 (43). A<sub>3</sub>VP<sub>1</sub> crystals grown in the presence or absence of fructose, respectively, adopted the orthorhombic P2<sub>1</sub>2<sub>1</sub>2 and monoclinic P2<sub>1</sub> space groups. Structure Determination and Model Refinement. Molecular replacement was carried out with the V-region (495-828) of AgI/II (PDB 1JMM) (12) as the input model using PHASER (44). Molecular replacement yielded a strong single solution for each dataset and were further refined using CNS (45). The resulting electron density map had clear interpretable

densities for the helical repeat regions, and were then built with Coot (46). Residues 495-828 of the A<sub>1</sub>VP<sub>3</sub> structures were highly similar to the solved AgI/II V-region structure (12) (average RMSD of 0.431 Å), The final Rfactor/Rfree of the models were 18.1/22.2 (native) and 19.0/22.8 (fructose co-crystal). The quality of the final models were assessed using PROCHECK (47), where 99.5% are within the most favored and allowed regions of the Ramachandran plot with an additional 0.5% within the generously allowed region. Complete data collection and refinement statistics are shown in Table 2.5.

Table 2.5: AgI/II A<sub>3</sub>VP<sub>1</sub> X-ray data collection and refinement statistics

|  | <b>A<sub>3</sub>VP<sub>1</sub></b> | <b>A<sub>3</sub>VP<sub>1</sub>+fructose</b> |
|--|------------------------------------|---|
| <b>Data collection</b>                               |                                    |   |
| Space group  | P2 <sub>1</sub>                    | P2 <sub>1</sub> 2 <sub>1</sub> 2            |
| Cell dimension                                       |                                    |   |
| a, b, c (Å)  | 50.03, 164.15, 67.73               | 70.73, 153.06, 49.68                        |
| $\alpha, \beta, \gamma$ (°)                          | 90, 91.03, 90                      | 90, 90, 90                                  |
| Wavelength (Å)                                       | 0.97918                            | 0.97918                                     |
| Resolution (Å)                                       | 50-2.0                             | 35-1.8                                      |
| R <sub>sym</sub> * (%)                               | 6.2 (18.2)                         | 7.2 (30.5)                                  |
| Mean I / $\sigma$ I                                  | 29.3 (8.9)                         | 16.4 (6.3)                                  |
| Completeness (%)                                     | 95.9 (91.0)                        | 90.7 (87.7)                                 |
| Redundancy   | 7.6 (6.4)                          | 5.1 (4.8)                                   |
| Solvent content (%)                                  | 51.33                              | 49.64                                       |
| <b>Refinement</b>                                    |                                    |   |
| R <sub>work</sub> /R <sub>free</sub> (%)             | 18.1/22.2                          | 19.0/22.8                                   |
| <b>Geometry analysis</b> , number residues (% total) |                                    |   |
| Most favored regions                                 | 786 (91.0)                         | 397 (91.9)                                  |
| Additional allowed regions                           | 75 (8.7)                           | 33 (7.6)                                    |
| Generously allowed regions                           | 3 (0.3)                            | 2 (0.5)                                     |
| Disallowed regions                                   | 0 (0.0)                            | 0 (0.0)                                     |
| PDB accession number                                 | 3IPK                               | 3IOX  |

## **Analytical Ultracentrifugation.**

A<sub>3</sub>VP<sub>1</sub>, A<sub>1</sub>VP<sub>3</sub>, and CG14 polypeptides were dialyzed into 20 mM TRIS pH 8.0, 150 mM sodium chloride, 1 mM EDTA to a final protein concentration of 0.6 mg/mL. Sedimentation velocity experiments were carried out using a Beckman Optima XL-A. During velocity runs at 40,000 rpm and 20 °C, frames of the absorbance at 280 nm across the cell were collected every 5 minutes. Sedimentation velocity data were fit using the program Sedfit (5) with the first seven frames of data removed and then every third frame of data used in the analysis. The partial specific volume ( $v_{\text{bar}}$ ) of 0.731 ml/g and buffer density of 1.00517 as well as hydration values for A<sub>3</sub>VP<sub>1</sub>, A<sub>1</sub>VP<sub>3</sub>, and CG14 of 0.45 g/g, 0.44 g/g, and 0.44 g/g, respectively, were calculated in Sednterp and then used in Sedfit to fit with a resolution of 100 (s-grid increments of 0.1 S) using a continuous  $c(s)$  distribution to model the velocity sedimentation and obtain the sedimentation and frictional coefficients. The coefficients of the largest peak (which contributed to 94% or more of the total protein concentration) were used to determine oblate and prolate ratios in Sednterp.

## **Circular Dichroism (CD) Spectroscopy.**

CD Spectra were recorded at 22 °C on the Olis DSM 100 circular dichroism spectrophotometer using a quartz cell (Starna Cells, Inc) with a path length of 0.2 mm. Proteins were diluted to 0.3 - 0.6 mg/mL in a buffer of 20 mM TRIS at pH 8.0, 150 mM sodium chloride, and 1 mM EDTA. Spectra were recorded over 200 - 260 nm in 1-nm steps and averaged over 10 complete scans. The millidegree data were converted to  $\Delta\epsilon$  (liter  $\text{m}^{-1} \text{cm}^{-1}$ ) and then analyzed using the CONTIN/LL algorithm implemented in CDPRO (6) with the SDP48 reference set of proteins.

## **Preparation of Salivary Agglutinin.**

Salivary agglutinin was prepared as previously described (7, 8). Unstimulated saliva was collected over a 1 hour period from healthy volunteers into 50 mL conical tubes on ice. Saliva was clarified by centrifugation (8000 x g) for 20 min at 4 °C. Clarified saliva was mixed with potassium-phosphate buffered saline (KPBS; 2.7 mM potassium chloride, 1.5 mM potassium phosphate, 137 mM sodium chloride, 6.5 mM sodium phosphate, pH 7.2) and *S. mutans* cells (at OD600 = 1.0) at a 1:1:1 (v:v:v) ratio and incubated end-over-end at 37 °C for 30 min. Cells were pelleted at 2000 x g for 15 minutes, washed once with KPBS, and the adsorbed agglutinin eluted from the *S. mutans* cells in one volume of KPBS containing 1 mM EDTA. The cells were removed by centrifugation at 4000 x g for 20 min and the supernatant-containing agglutinin was filter-sterilized with a 0.2 µm filter, dialyzed overnight against KPBS, and stored at -20 °C.

## **Evaluation of AgI/II epitopes by competition ELISA.**

ELISA plate wells were coated with *S. mutans* whole cells (10<sup>5</sup> cfu/well) in carbonate-bicarbonate buffer (pH. 9.6), washed, and blocked with phosphate buffered saline (pH 7.2) containing 0.03% Tween-20 (PBS-Tween) as described (9). One hundred microliters of serial 2-fold dilutions beginning at 5 µM of each polypeptide diluted in PBS-Tween were incubated for 2 h at 37 °C with 100 µl of anti-AgI/II monoclonal antibodies (MAbs) 1-6F, 4-9D, 4-10A, or 3-10E (10) diluted in the same buffer and titrated to result in OD450 ~ 1.0 in the absence of inhibitor. The source of MAbs was murine ascites fluids. After washing, MAb binding to the plate wells was traced with 100 l of peroxidase-labeled goat anti-mouse IgG (MP Biomedicals) diluted 1:1000 in PBS-Tween and incubated for 2 hours at 37 °C. Plates were washed and developed with o-phenylenediamine dihydrochloride substrate solution and the absorbance at 450 nm was recorded using an MPM Titertek model

550 ELISA plate reader (Bio-Rad, Hercules, CA). All assays were performed in triplicate. Percent inhibition was calculated as  $100 \left[ \frac{\text{mean OD450 of MAb} + \text{test polypeptide}}{\text{mean OD450 of MAb alone}} \times 100 \right]$ .

### **Surface Plasmon Resonance.**

Adherence of AgI/II polypeptides to human salivary agglutinin was assessed by surface plasmon resonance using the BIAcore 2000 (BIAcore AB, Uppsala, Sweden). Purified SAG was immobilized to an RU value of 3800 on a CM5 sensor chip via amine coupling, and the experiment or control surfaces were blocked with 1.0 M ethanolamine. 80  $\mu\text{L}$  of each AgI/II polypeptide, in concentrations ranging from 0.25 - 4  $\mu\text{M}$  in 10mM HEPES pH 7.4 and 150 mM sodium chloride (HBS) with 2.5 mM  $\text{CaCl}_2$ , were flowed at a rate of 20 L/min over the sensor chip. An 8 minute dissociation time was used to measure the slow dissociation rates observed. Between experiments, 10  $\mu\text{L}$  of HBS augmented with 3 mM EDTA, 0.0005% Tween-20 (HBSEP) and 10 mM sodium hydroxide was flowed over the chip surface. This was followed by flowing 5  $\mu\text{L}$  of 10 mM HCl over the surfaces to regenerate the cells. Each binding experiment was repeated 3 times in different series of analyte concentrations on the same chip surface as well as on a separately prepared CM5 chip surface that showed similar results. Signal from the non-coated control cell FC1 was subtracted from that of the SAG-coated FC2 surface to obtain the sensograms ( $\Delta\text{RU}$ ). The collected data was fitted using a 1:1 Langmuir kinetic model and the BIAEvaluation software version 4.2 (Biacore AB, Sweden).

### **Isothermal titration calorimetry.**

Calorimetric titrations were carried out with a MicroCal ITC calorimeter (MicroCal, Inc., Northhampton, MA). The  $\text{VP}_3$  protein was loaded into the 1.342 mL stirred cell (at 400

rpm) with a concentration ranging from 7-11  $\mu\text{M}$  and the  $\text{A}_{1-3}$  protein at an 8-fold higher concentration was loaded into a 250  $\mu\text{L}$  syringe. Experiments were carried out at different temperatures (15, 20, 25, and 30  $^{\circ}\text{C}$ ) with the proteins in 20 mM HEPES (pH 7.0), 150 mM sodium chloride, and 1% (v/v) glycerol. After the initial injection of 1  $\mu\text{L}$  (which was excluded in data fitting), 24 injections of 10  $\mu\text{L}$  each were delivered at 300 s intervals. Data were fit to a single binding site model using Origin software (version 7.0383 with MicroCal ITC-analysis module) from which the  $\Delta\text{G}$ ,  $\text{K}_{\text{A}}$ ,  $\text{K}_{\text{D}}$ ,  $\Delta\text{H}$ ,  $\Delta\text{S}$ , and stoichiometry ( $n$ ) of the binding were obtained. The enthalpy of dilution of the  $\text{A}_{1-3}$  protein was measured by titration of  $\text{A}_{1-3}$  into buffer alone to be  $\sim 4$  kcal/mol, small relative to the large enthalpy changes obtained in titrations of  $\text{A}_{1-3}$  into  $\text{VP}_3$ . Within these experimental temperatures,  $\Delta\text{H}$  is well-approximated as a linear function of  $T$ , where  $\Delta\text{C}_p = (\Delta\text{H}T_2 - \Delta\text{H}T_1) / (T_1 - T_2)$  and  $\Delta\text{C}_p$  is the change in heat capacity at constant pressure upon binding. This relationship was used to calculate the  $\Delta\text{C}_p$  as  $-7.95 \pm 0.58 \text{ kJ K}^{-1} \text{ mol}^{-1}$ .

### **Modeling PMS into the structure of $\text{A}_3\text{VP}_1$**

Residues 495-828 of the  $\text{A}_1\text{VP}_3$  structures were highly similar to the solved AgI/II V-region structure (11) (average RMSD of 0.431  $\text{\AA}$ ), however the residue Trp816, located inside a cleft of the V-region, has a large side chain movement in the  $\text{A}_3\text{VP}_1$  molecules compared to the AgI/II V-region structure. This tryptophan residue neighbors a large continuous difference density that suggested the presence of a single covalently linked molecule. As all crystal forms of  $\text{A}_3\text{VP}_1$  were supplemented with PMSF (phenyl-methylsulfonyl fluoride) during purification, we modeled this site with a hydrolyzed form (PMS) of the inhibitor PMSF. The sulfate group of PMS is positioned near the positively charged sidechain of Arg824, while the PMS benzyl group stacks to form a hydrophobic interaction with Trp816.

Structure Figures. All images of the structures were generated using PyMol (12).

### **Solvent-Accessible Surface Area (ASA) of A<sub>3</sub>VP<sub>1</sub> X-ray Structure.**

ASA values were calculated from atomic positions of the Ag I/II structures using the software NAccess (ver. 2.1.1) with a probe size of 1.4 Å and Z-slices of 0.05 Å. To estimate the change in ASA ( $\Delta$ ASA) upon A-P association, the ASA changes were calculated using the structure of A<sub>3</sub>VP<sub>1</sub> from the P2<sub>1</sub>2<sub>1</sub>2 crystal. Polar (ASA<sub>polar</sub>) and non-polar (ASA<sub>apolar</sub>) surface areas were calculated for the associated A<sub>3</sub> and P<sub>1</sub> present in the structure, and then separately for only the A<sub>3</sub> residues (386-475) or the P<sub>1</sub> residues (836-874). The  $\Delta$ ASA for polar and non-polar surfaces were then calculated as:  $\Delta$ ASA<sub>A<sub>3</sub>P<sub>1</sub></sub> = ASA<sub>A<sub>3</sub></sub> + ASA<sub>P<sub>1</sub></sub> - ASA<sub>A<sub>3</sub>P<sub>1</sub></sub>, resulting in 2515.2 Å<sup>2</sup> to 920.0 Å<sup>2</sup> non-polar:polar contact-surface.

### **ACKNOWLEDGMENTS**

The authors thank Dr. Irina Protassevitch and Dr. Christie Brouillette for ITC, CD and BIAcore studies, and Dr. Peter Prevelige for the Analytical Ultracentrifugation studies. ML and CD thank the UAB-CCC X-ray core and NE-CAT facility at APS/ANL. This work was supported by grants NIH-NIDCR RO1 DE017737 (CD), NIH-NIDCR R01 DE013882 (LJB), and by the DART training grant NIDCR T-32 DE0176707 (ML).

### **References**

1. Loesche WJ (1986) Role of *Streptococcus mutans* in human dental decay. *Microbiol Rev* **50**(4):353-380.
2. Russell MW, Childers NK, Michalek SM, Smith DJ, & Taubman MA (2004) A Caries Vaccine? The state of the science of immunization against dental caries.

*Caries Res* **38**(3):230-235.

3. Ma JK, Kelly CG, Munro G, Whiley RA, & Lehner T (1991) Conservation of the gene encoding streptococcal antigen I/II in oral streptococci. *Infect Immun* **59**(8):2686-2694.
4. Zhang S, Green NM, Sitkiewicz I, Lefebvre RB, & Musser JM (2006) Identification and characterization of an antigen I/II family protein produced by group A Streptococcus. *Infect Immun* **74**(7):4200-4213.
5. Jenkinson HF & Demuth DR (1997) Structure, function and immunogenicity of streptococcal antigen I/II polypeptides. *Mol Microbiol* **23**(2):183-190.
6. Pecharki D, Petersen FC, Assev S, & Scheie AA (2005) Involvement of antigen I/II surface proteins in *Streptococcus mutans* and *Streptococcus intermedius* biofilm formation. *Oral Microbiol Immunol* **20**(6):366-371.
7. Love RM, McMillan MD, Park Y, & Jenkinson HF (2000) Coinvasion of dentinal tubules by *Porphyromonas gingivalis* and *Streptococcus gordonii* depends upon binding specificity of streptococcal antigen I/II adhesin. *Infect Immun* **68**(3):1359-1365.
8. Sciotti MA, Yamodo I, Klein JP, & Ogier JA (1997) The N-terminal half part of the oral streptococcal antigen I/II contains two distinct binding domains. *FEMS Microbiol Lett* **153**(2):439-445.
9. Crowley PJ, Brady LJ, Michalek SM, & Bleiweis AS (1999) Virulence of a spaP mutant of *Streptococcus mutans* in a gnotobiotic rat model. *Infect Immun* **67**(3):1201-1206.
10. Russell MW, Bergmeier LA, Zanders ED, & Lehner T (1980) Protein antigens of

*Streptococcus mutans*: purification and properties of a double antigen and its protease-resistant component. *Infect Immun* **28**(2):486-493.

11. Brady LJ, et al. (1991) Restriction fragment length polymorphisms and sequence variation within the spaP gene of *Streptococcus mutans* serotype c isolates. *Infect Immun* **59**(5):1803-1810.
12. Troffer-Charlier N, Ogier J, Moras D, & Cavarelli J (2002) Crystal structure of the V-region of *Streptococcus mutans* antigen I/II at 2.4 Å resolution suggests a sugar preformed binding site. *J Mol Biol* **318**(1):179-188.
13. Navarre WW & Schneewind O (1994) Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in gram-positive bacteria. *Mol Microbiol* **14**(1):115-121.
14. Hajishengallis G, Koga T, & Russell MW (1994) Affinity and specificity of the interactions between *Streptococcus mutans* antigen I/II and salivary components. *J Dent Res* **73**(9):1493-1502.
15. Oho T, Yu H, Yamashita Y, & Koga T (1998) Binding of salivary glycoprotein-secretory immunoglobulin A complex to the surface protein antigen of *Streptococcus mutans*. *Infect Immun* **66**(1):115-121.
16. McArthur WP, et al. (2007) Characterization of epitopes recognized by anti-*Streptococcus mutans* P1 monoclonal antibodies. *FEMS Immunol Med Microbiol* **50**(3):342-353.
17. van Dolleweerd CJ, Kelly CG, Chargelegue D, & Ma JK (2004) Peptide mapping of a novel discontinuous epitope of the major surface adhesin from *Streptococcus mutans*. *J Biol Chem* **279**(21):22198-22203.

18. Creamer TP & Campbell MN (2002) Determinants of the polyproline II helix from modeling studies. *Adv Protein Chem* **62**:263-282.
19. Blundell TL, Pitts, J.E., Tickle, I.J., Wood, S.P. and Wu, C.-W (1981) X-ray analysis (1.4 Å resolution) of avian pancreatic polypeptide: Small globular protein hormone. *Proc Natl Acad Sci U S A* **78**(7):4175-4179.
20. Ayakawa GY, et al. (1987) Isolation and characterization of monoclonal antibodies specific for antigen P1, a major surface protein of mutans streptococci. *Infect Immun* **55**(11):2759-2767.
21. Brady LJ, Piacentini DA, Crowley PJ, & Bleiweis AS (1991) Identification of monoclonal antibody-binding domains within antigen P1 of *Streptococcus mutans* and cross-reactivity with related surface antigens of oral streptococci. *Infect Immun* **59**(12):4425-4435.
22. Oli MW, McArthur WP, & Brady LJ (2006) A whole cell BIAcore assay to evaluate P1-mediated adherence of *Streptococcus mutans* to human salivary agglutinin and inhibition by specific antibodies. *J Microbiol Methods* **65**(3):503-511.
23. Crowley PJ, et al. (2008) Requirements for surface expression and function of adhesin P1 from *Streptococcus mutans*. *Infect Immun* **76**(6):2456-2468.
24. Kelly CG, et al. (1999) A synthetic peptide adhesion epitope as a novel antimicrobial agent. *Nat Biotechnol* **17**(1):42-47.
25. Parry DA, Fraser RD, & Squire JM (2008) Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. *J Struct Biol* **163**(3):258-269.

26. Ramachandran GN & Kartha G (1954) Structure of collagen. *Nature* **174**(4423):269-270.
27. Brodsky B, Thiagarajan G, Madhan B, & Kar K (2008) Triple-helical peptides: an approach to collagen conformation, stability, and self-association. *Biopolymers* **89**(5):345-353.
28. Venugopal MG, Ramshaw JA, Braswell E, Zhu D, & Brodsky B (1994) Electrostatic interactions in collagen-like triple-helical peptides. *Biochemistry* **33**(25):7948-7956.
29. Ferreon JC & Hilser VJ (2004) Thermodynamics of binding to SH3 domains: the energetic impact of polyproline II (PPII) helix formation. *Biochemistry* **43**(24):7787-7797.
30. Lee SF, et al. (1989) Construction and characterization of isogenic mutants of *Streptococcus mutans* deficient in major surface protein antigen P1 (I/II). *Infect Immun* **57**(11):3306-3313.
31. Jakubovics NS, Stromberg N, van Dolleweerd CJ, Kelly CG, & Jenkinson HF (2005) Differential binding specificities of oral streptococcal antigen I/II family adhesins for human or bacterial ligands. *Mol Microbiol* **55**(5):1591-1605.
32. Crowley PJ, Brady LJ, Piacentini DA, & Bleiweis AS (1993) Identification of a salivary agglutinin-binding domain within cell surface adhesin P1 of *Streptococcus mutans*. *Infect Immun* **61**(4):1547-1552.
33. Munro GH, et al. (1993) A protein fragment of streptococcal cell surface antigen I/II which prevents adhesion of *Streptococcus mutans*. *Infect Immun* **61**(11):4590-4598.

34. Brady LJ, Piacentini DA, Crowley PJ, Oyston PC, & Bleiweis AS (1992) Differentiation of salivary agglutinin-mediated adherence and aggregation of mutans streptococci by use of monoclonal antibodies against the major surface adhesin P1. *Infect Immun* **60**(3):1008-1017.
35. Senpuku H, Miyauchi T, Hanada N, & Nisizawa T (1995) An antigenic peptide inducing cross-reacting antibodies inhibiting the interaction of *Streptococcus mutans* PAc with human salivary components. *Infect Immun* **63**(12):4695-4703.
36. Senpuku H, et al. (2001) Inhibitory effects of MoAbs against a surface protein antigen in real-time adherence in vitro and recolonization in vivo of *Streptococcus mutans*. *Scand J Immunol* **54**(1-2):109-116.
37. van Dolleweerd CJ, Chargelegue D, & Ma JK (2003) Characterization of the conformational epitope of Guy's 13, a monoclonal antibody that prevents *Streptococcus mutans* colonization in humans. *Infect Immun* **71**(2):754-765.
38. Holmskov U, et al. (1999) Cloning of gp-340, a putative opsonin receptor for lung surfactant protein D. *Proc Natl Acad Sci U S A* **96**(19):10794-10799.
39. Pankov R & Yamada KM (2002) Fibronectin at a glance. *J Cell Sci* **115**(Pt 20):3861-3863.
40. Phillips GN, Jr., Flicker PF, Cohen C, Manjula BN, & Fischetti VA (1981) Streptococcal M protein: alpha-helical coiled-coil structure and arrangement on the cell surface. *Proc Natl Acad Sci U S A* **78**(8):4689-4693.
41. Andre I, et al. (2006) Streptococcal M protein: structural studies of the hypervariable region, free and bound to human C4BP. *Biochemistry* **45**(14):4559-4568.

42. Deivanayagam CC, et al. (2000) Novel fold and assembly of the repetitive B region of the *Staphylococcus aureus* collagen-binding surface protein. *Structure Fold Des* **8**(1):67-78.
43. Z. Otwinowski WM (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* **276**(A):307-326.
44. McCoy AJ, et al. (2007) Phaser crystallographic software. *Journal of Applied Crystallography* **40**:658-674.
45. Brunger AT (2007) Version 1.2 of the Crystallography and NMR system. *Nat Protoc* **2**(11):2728-2733.
46. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1):2126-2132.
47. Laskowski RA MM, Moss DS and Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**:283-291.
48. DeLano WL &. (The PyMOL Molecular Graphics System (2009). (DeLano Scientific LLC, Palo Alto, California, USA).
49. Koradi R, Billeter M, & Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14**(1):51-55, 29-32.

**CHAPTER 3**

**CRYSTAL STRUCTURE OF THE C-TERMINUS OF  
*STREPTOCOCCUS MUTANS* ANTIGEN I/II AND  
CHARACTERIZATION OF SALIVARY AGGLUTININ  
ADHERENCE DOMAINS**

by

LARSON MR, RAJASHANKAR KR, CROWLEY PJ, KELLY C, MITCHELL TJ,  
BRADY LJ, DEIVANAYAGAM C

Submitted to *The Journal of Biological Chemistry*

Format adapted for dissertation

## Abstract

The *Streptococcus mutans* Antigen I/II (AgI/II) is a cell surface localized protein that adheres to salivary components and extracellular matrix molecules. Here we report the 2.5 Å resolution crystal structure of the complete carboxy-terminal (C-terminal) region of AgI/II. The C-terminal region is comprised of three major domains: C<sub>1</sub>, C<sub>2</sub>, and C<sub>3</sub>. Each domain adopts a DE-variant IgG fold, with two  $\beta$ -sheets whose A and F strands are linked through an intramolecular isopeptide bond. The adherence of the C-terminal AgI/II fragments to the putative tooth surface receptor salivary agglutinin (SAG), as monitored by surface plasmon resonance, indicated that the minimal region of binding was contained within the first and second DE-variant-IgG domains (C<sub>1</sub> and C<sub>2</sub>) of the C-terminus. Inhibition studies further confirmed that C<sub>12</sub> could inhibit *Streptococcus mutans* adherence to SAG. Competition experiments demonstrated that the C- and N-terminal regions of AgI/II adhere to distinct sites on SAG. A cleft formed at the intersection between these C<sub>1</sub> and C<sub>2</sub> domains bound glucose molecules from the cryo-protectant solution, revealing a putative binding site for its highly glycosylated receptor SAG. Finally, electron microscopy images confirmed the elongated structure of AgI/II and enabled building a composite tertiary model that encompasses its two distinct binding regions.

## Introduction

Dental caries (also called tooth decay or dental cavities) is a ubiquitous worldwide disease that affects humans of all age groups. *Streptococcus mutans*, a primary etiological agent of human dental caries (1) and an increasingly recognized cause of bacterial endocarditis (2), adheres to proteins contained within the salivary pellicle on the tooth surface, the extracellular matrix, and other microbial species (3). Antigen I/II (AgI/II, also known as P1, B, SpaP, or PAc) of *S. mutans* has been implicated in bacterial adherence to constituents of the salivary pellicle (4, 5), and has been studied for the past three decades as a target for protective immunity against dental caries. Apart from adherence, AgI/II influences biofilm formation (6) and promotes platelet aggregation (7), collagen-dependent bacterial invasion of dentin (8), and cariogenicity (9). While AgI/II was initially discovered on oral streptococci, it has also been identified in members of the Group A and Group B streptococci (10), suggesting a role for this adhesin in a variety of species.

The AgI/II family proteins range from 140-180 kDa in predicted size and have a primary sequence composed of multiple conserved regions (Fig 3.1 b). Toward the N-terminus, repeated sequences of high alanine content constitute the alanine-rich region, followed by a segment commonly referred to as the variable (V) region. Further C-terminal in the sequence is a region of high proline content that forms a repetitive proline-rich region. Following the AgI/II proline-rich region is a C-terminal region (60 kDa or 550 amino acids) which is the most conserved region of AgI/II, with 62% identity among strains (11). This region also contains an LPxTG consensus motif recognized by the sortase enzyme responsible for covalently attaching AgI/II to the cell wall peptidoglycan (3).

The human receptor for AgI/II is the glycoprotein complex known as salivary agglutinin (SAG), which is secreted by the salivary glands as part of the salivary fluid (12). The main component of SAG is the highly-glycosylated 340 kDa glycoprotein scavenger receptor

called gp340 (13). The interaction of AgI/II with SAG is complex and multivalent, and requires the presence of calcium (14). Segments within both the N- and C- termini of AgI/II have been implicated in the binding and evidence from several studies suggest that the C-terminus contributes substantially to *S. mutans* adherence (5, 15, 16). In addition, monoclonal antibodies recognizing epitopes within the C-terminus of AgI/II effectively inhibit adherence of bacterial cells to SAG immobilized on hydroxyapatite beads (17). Further, the C-terminus of AgI/II itself can competitively inhibit the adherence of *S. mutans* to SAG (5). The peptide “QLKTADLPAGRDETTSFVLV” within the C-terminus both inhibits binding of AgI/II to SAG as well as the recolonization of *S. mutans* to human teeth in vivo (15). Alanine substitutions within this peptide indicated that residues Q, E, and FVLV were required for inhibition by this peptide (15).

In our previous study, we had shown that the discontinuous alanine- (A) and proline- (P) rich regions associate in a hybrid of  $\alpha$ - and polyproline type II helices to form an extended A/P stalk which positions the globular intervening region (V), at the apex and away from the cell surface (16). In addition, we had identified the presence of two distinct AgI/II regions, A<sub>3</sub>VP<sub>1</sub> and the C-terminus, that adhere to SAG. While the functional structure comprising the N-terminal and central regions of *S. mutans* AgI/II has been well characterized (16), the complete and functional C-terminal region has not yet been described. A partial structure of a homologous fragment from the C-terminus of *Streptococcus gordonii* (*S. gordonii*) SspB was recently reported, and contains two  $\beta$ -stranded domains (18).

In this paper, we report the structure of the entire C-terminus of AgI/II from *Streptococcus mutans* at 2.5 Å resolution. The structure reveals three  $\beta$ -sheeted domains, with each domain adopting the DE-variant immunoglobulin-like fold. We have biophysically characterized and delineated the minimal binding region within C-terminus through both surface plasmon resonance studies as well as adherence inhibition studies using *S. mutans* cells in vitro. In addition, several glucose molecules were identified in the C-terminus structure,

suggestive of a binding site which could facilitate adherence to the heavily glycosylated SAG. We have further deciphered AgI/II complex multivalent binding characteristics using competition experiments with AgI/II fragments, and provide evidence for the presence of distinct binding sites on SAG. Finally, transmission electron microscopy images of AgI/II enabled the building of a composite structural model for this elongated fibrillar protein.

## Experimental Procedures

*Expression and Purification of C-terminal Fragments-* DNA encoding the C-terminal region (C<sub>123</sub>) was subcloned from the AgI/II gene of the *S. mutans* NG8 strain into a pET23d vector as described previously (16). Additional constructs of C<sub>1</sub>, C<sub>12</sub>, C<sub>2</sub>, C<sub>23</sub>, and C<sub>3</sub> were likewise subcloned into the pET23d vector using the primers listed in Table 3.1, with extents of the fragments shown in Figure 3.1b. The resulting C-terminal plasmids were transformed into BL21 ( $\lambda$ DE3) *E. coli* cells. 20 mL Terrific Broth cultures were grown overnight at 37 °C and transferred into 1 liter cultures on the following day with protein expression and cell lysis as described previously (16). Lysates were purified over a HisPrep Nickel affinity column (Amersham Biosciences), followed by ion exchange chromatography on a MonoQ HR10/10 column (Amersham Biosciences), and lastly by size exclusion chromatography over a Superdex 75pg 26/60 column (Amersham Biosciences).

\*Plasmid pDC20 containing the AgI/II gene from *S. mutans* NG8 (36) was used as the PCR template

*Crystallization and Data Collection-* C<sub>123</sub> was concentrated to 12.1 mg/mL using a 10-kDa m.w. cutoff membrane in an Amicon concentrator under nitrogen gas pressure and screened by vapor diffusion at 22 °C using commercial crystallization screens in 1:1  $\mu$ l protein to well solution hanging drops. C<sub>123</sub> crystals were optimized from condition 1 of Crystal Screen II (Hampton Research), to a well solution (WS) containing 11%

|                  | <b>Forward Primer</b>                    | <b>Reverse Primer</b>                     |
|------------------|--|---|
| C <sub>123</sub> | TATAAGGATCCATTTCCATT<br>ACTTTAAACTAGC    | TTATACTCGAGTGAAGTGTAA<br>GTTACCCCATTGACAG |
| C <sub>1</sub>   | TATAAGGATCCATTTCCATT<br>ACTTTAAACTAGC    | TTATACTCGAGTGTCACCCGA<br>ACAACATTGG       |
| C <sub>12</sub>  | C <sub>1</sub> forward                   | C <sub>2</sub> reverse                    |
| C <sub>2</sub>   | TATAAGGATCCCAACTCCTG<br>GTAAACCAAATGATCC | TTATACTCGAGCTTAGGAACA<br>TTATTGATAACG     |
| C <sub>23</sub>  | C <sub>2</sub> forward                   | C <sub>3</sub> reverse                    |
| C <sub>3</sub>   | TATAAGGATCCAGATTAACC<br>CTAAGAAAGATGTG   | TTATACTCGAGTGAAGTGTAA<br>GTTACCCCATTGACAG |

Table 3.1: PCR Primers used in the C-terminal structural study

polyethylene-glycol (PEG) 6000, 2 M sodium chloride and 55 mM sodium succinate at pH 4.5. Crystals were flash frozen in WS augmented with cryoprotectants of either 20% PEG 400 for data collection at a home source (Raxis-IV mounted on a RU-H3R X-ray generator operated at 50 mA and 100 KV) or 30% D-glucose at the NE-CAT beamline at 93 °K under a gaseous N<sub>2</sub> stream. High resolution data recorded at the NE-CAT beamline was collected on a Quantum4 CCD detector, using a wavelength ( $\lambda$ ) of 1.0 Å, an exposure time of 2.0 seconds, an oscillation angle ( $\phi$ ) of 1.0°, and a distance (D) of 150 mm from crystal to detector. The diffraction data were integrated and scaled with HKL2000 (19).

*Multiple Isomorphous Replacement (MIR)*- Three heavy atom derivatives of the C<sub>123</sub> region were obtained by soaking freshly grown crystals. Crystals placed in WS augmented with 1M sodium iodide darkened to a brown/yellow tinge after 1 week indicating incorporation of iodide. Derivatives containing thallium (III) acetate were stable under a brief 30 second quick-soaks containing 1 mM thallium (III) acetate followed by flash freezing, while derivatives of thimerosal were soaked in WS augmented with 10 mM thimerosal for 48 hours. For phasing, low resolution derivative and native datasets were collected at the home source. These data were processed with HKL2000 (19) and MIR solutions were

identified by SHARP (version 2.6). 12 sites were identified in the sodium iodide derivative, and an additional 4 thallium and 6 thimerosal sites were identified. The figure of merit (FOM) of the initial solution was 0.35 and 0.40, respectively, for the acentric and centric reflections. After solvent flattening, the FOM<sub>total</sub> improved to 0.914, with a resulting highly interpretable electron density map having clear boundaries between the protein and solvent within the map. The models were manually built using Coot (20), with refinements under CNS (21) over several cycles. At later stages of refinement, a higher resolution (2.5 Å) native dataset collected at NE-CAT was used to complete the model building. The first 8 residues (992-1000) could not be identified on the electron density maps. A complete summary of crystallographic parameters and refinement statistics is included in Table 3.2.

*Surface Plasmon Resonance-Adherence of AgI/II fragments-* SAG was prepared from pooled unstimulated saliva from healthy human volunteers as described previously (17). The adherence of AgI/II polypeptides to human SAG was assessed by surface plasmon resonance (SPR) on the BIAcore 2000 (BIAcore AB, Uppsala, Sweden). SAG was immobilized on a CM5 chip, with buffer and flow conditions as previously described (16) Eighty  $\mu\text{L}$  of each AgI/II polypeptide, in concentrations ranging from 0.25 to 4  $\mu\text{M}$ , were injected over the chip surfaces, and dissociations were measured over 8 minute time periods. Between experiments, 10  $\mu\text{L}$  of 10 mM HCl was used to regenerate the surface. Signals from the uncoated control cell FC1 were subtracted from those of the SAG-coated cell FC2 to produce sensorgrams ( $\Delta$  RU). Binding data were fitted with a 1:1 Langmuir kinetic model using the BIAEvaluation software (v4.2, BIAcore AB, Sweden).

*Competition binding:* To determine whether AgI/II domains bound to the same site on SAG, competition SPR experiments were conducted as follows: One fragment of AgI/II (CG14, A<sub>3</sub>VP<sub>1</sub>, AVP, or C<sub>123</sub>) was flowed over the chip surface for 60 seconds to saturate available SAG binding sites, and immediately afterwards the second fragment was flowed over the chip for 60 seconds to test for adherence, with each experiment performed in

|  | <b>C<sub>123</sub> Native</b>  | <b>C<sub>9</sub>H<sub>9</sub>HgNaO<sub>2</sub>S<br/>(Thimerosal)</b> | <b>Sodium Iodide</b>      | <b>Thallium (III)<br/>Acetate Hydrate</b> |
|--|--|--|---------------------------|---|
| <b>Data collection</b>                   |  |  |                           |   |
| Resolution (Å)                           | 2.5  | 3.2  | 3.0                       | 4.0                                       |
| Space group                              | I222   | I222   | I222                      | I222                                      |
| Cell dimensions                          |  |  |                           |   |
| a, b, c (Å)                              | 108.50, 156.36,<br>213.65  | 108.00, 156.90,<br>213.99  | 108.03, 156.37,<br>213.84 | 108.62, 157.04,<br>213.71                 |
| $\alpha, \beta, \gamma$ (°)              | 90, 90, 90   | 90, 90, 90   | 90, 90, 90                | 90, 90, 90                                |
| Unique Reflections                       | 52164  | 56587  | 68419                     | 25993                                     |
| Completeness (%) <sup>*</sup>            | 97.0 (81.2)  | 98.1 (99.9)  | 100 (100)                 | 86.8 (89.5)                               |
| Redundancy <sup>*</sup>                  | 6.9 (5.3)  | 2.4 (2.3)  | 3.9 (3.8)                 | 2.8 (2.7)                                 |
| I / $\sigma$ I <sup>*</sup>              | 24.5 (3.59)  | 13.4 (3.4)   | 9.4 (1.4)                 | 12.0 (5.7)                                |
| R <sub>sym</sub> (%) <sup>*</sup>        | 9.2 (31.4)   | 8.1 (30.8)   | 14.6 (86.4)               | 13.0 (25.3)                               |
| R <sub>cullis</sub> (acentric/centric)   |  | .789/.807  | .815/.855                 | .932/.957                                 |
| Phasing Power (acentric/centric)         |  | .790/.756  | 1.006/.834                | .333/.299                                 |
| <b>Refinement</b>                        |  |  |                           |   |
| Resolution (Å)                           | 2.5  |  |                           |   |
| No. reflections (work/free)              | 46836 / 5328   |  |                           |   |
| R <sub>work</sub> / R <sub>free</sub>    | 20.8 / 24.4  |  |                           |   |
| <b>No. atoms in asymmetric unit</b>      |  |  |                           |   |
| Protein                                  | 2 molecules, with<br>7610 atoms total                                    |  |                           |   |
| Ligand/ion                               | 6 Ca <sup>2+</sup> , 1 Mg <sup>2+</sup> ,<br>and 8 Glucose<br>(96 atoms) |  |                           |   |
| Water                                    | 324  |  |                           |   |
| <b>Average B-factors (Å<sup>2</sup>)</b> |  |  |                           |   |
| Protein                                  | 44.6   |  |                           |   |
| Ligand/ion                               | 44.6 for ions and<br>67.4 for glucoses                                   |  |                           |   |
| Water                                    | 43.6   |  |                           |   |
| <b>R.m.s. deviations</b>                 |  |  |                           |   |
| Bond lengths (Å)                         | 0.0082   |  |                           |   |
| Bond angles (°C)                         | 1.411  |  |                           |   |
| <b>Ramachandran plot<sup>**</sup></b>    |  |  |                           |   |
| % favored                                | 87.3   |  |                           |   |
| % allowed                                | 12.0   |  |                           |   |
| % generously allowed                     | 0.7  |  |                           |   |
| % disallowed                             | 0.0  |  |                           |   |

Table 3.2: AgI/II C-terminus X-ray data collection and refinement statistics  
<sup>\*</sup>Values in parentheses are for highest-resolution shell. <sup>\*\*</sup> as defined by Procheck (37).

triplicate. The RU values (of the second fragment) were determined as the difference in the maximal RU subtracted by the baseline prior to second injection.

*Inhibition of adherence of S. mutans to SAG-coated Hydroxyapatite-* The ability of fluid-phase recombinant AgI/II polypeptides to compete with *S. mutans* cell-associated AgI/II for binding to SAG-coated hydroxyapatite beads was determined as described previously (17). Assays were performed using *S. mutans* strain NG8 and all assays were performed in triplicate.

*Purification of AgI/II from S. mutans and Electron Microscopy-* *S. mutans* (Guys strain) was grown in supplemented basal medium and full-length AgI/II was prepared from the culture supernatant by ion exchange and gel filtration chromatography as described previously (22). Purified AgI/II was dissolved at a range of protein concentrations (0.2 - 1  $\mu$ M) in 50% (w/w) glycerol, sprayed onto freshly cleaved mica, and after drying for 15 minutes at 10-5 mbar were unidirectionally shadowed with platinum at low angle ( $7.5\sigma$ ). Replicas were floated off onto distilled water. All specimens were examined and photographed in a Siemens 102 transmission electron microscope, with the magnification calibrated using a diffraction grating replica. Electron micrographs were measured on an Apple digitizer. At least 100 measurements of length were recorded for each specimen.

## RESULTS

*Overall Structure.* The C-terminus of AgI/II contains three domains C<sub>1</sub>, C<sub>2</sub> and C<sub>3</sub> (Fig 3.1 a) that are predominantly  $\beta$ -stranded, and possess 11 to 12 strands that are interspersed with short helical stretches. This highly  $\beta$ -sheeted character of the C-terminal region confirms the results of circular dichroism studies on this protein fragment (16). Each of these domains adheres to the DE-variant immunoglobulin like (DEv-IgG) fold (23). The DEv-IgG like domains possess a set of antiparallel strands A-G similar to classical IgG folds, whose

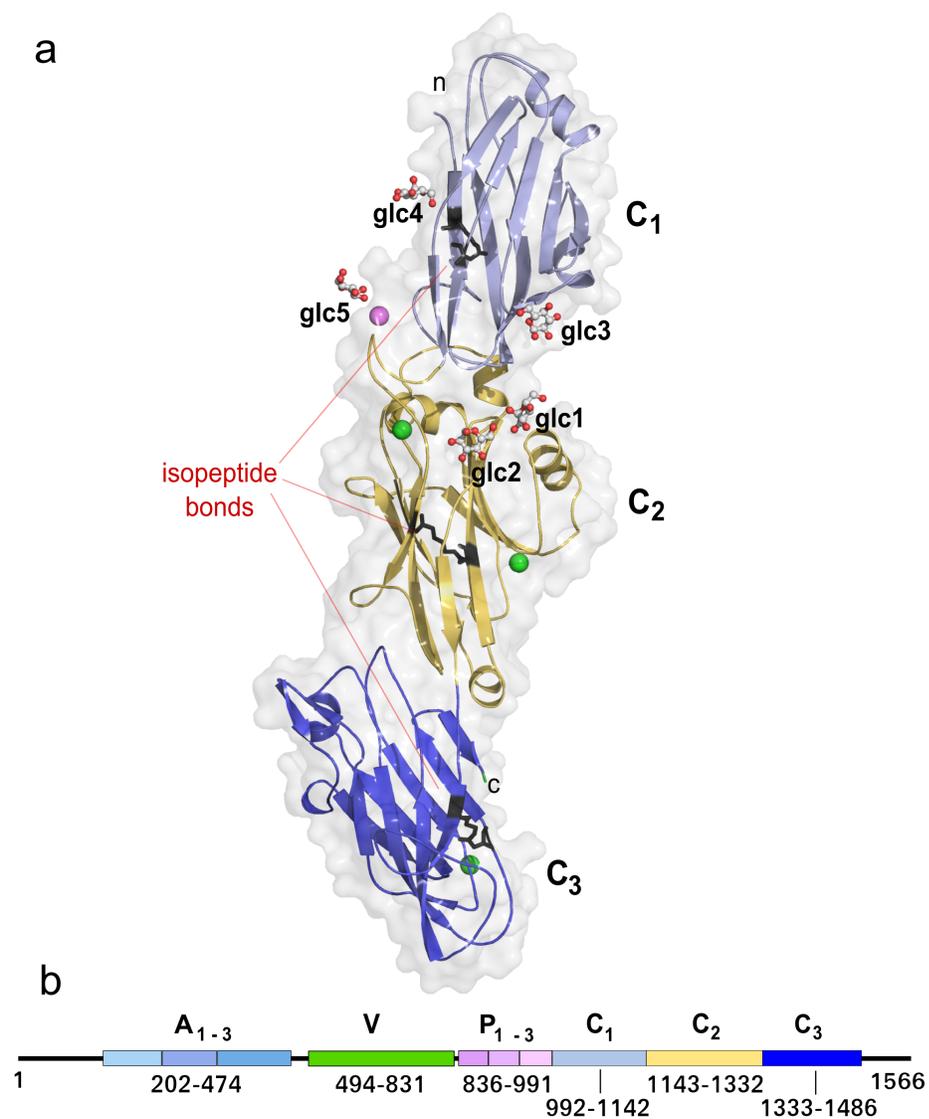


Figure 3.1: a) The AgI/II C-terminal region is comprised of three domains, labelled as C<sub>1</sub>, C<sub>2</sub>, and C<sub>3</sub>. The AgI/II C-terminal domains are shown in a ribbon model overlaid with a surface plot to illustrate the overall contiguous structure of the three domains. Each domain has an isopeptide bond linking two  $\beta$ -sheets (shown in red). The C<sub>1</sub> and C<sub>2</sub> domains have five bound glucoses (labeled glc1,2,3,4,5). Additionally, three calcium ions exist in the C<sub>2</sub> and C<sub>3</sub> domains (green), and one magnesium ion on the C<sub>1</sub> domain (purple). b) Cartoon represents the primary sequence of AgI/II, where the alanine-rich- and proline-rich repeats, as well as the variable-region, and C-terminus (C<sub>1</sub>, C<sub>2</sub>, C<sub>3</sub>) are indicated with distinct colors.

essential variations (additional strands and helices) occur between the D and E strands (Fig 3.2 a). The A, B, E, and D strands form one major  $\beta$ -sheet, while the C, F, and G strands form the second major  $\beta$ -sheet. The variations between the D and E strands are most prominent in the C<sub>2</sub> and C<sub>3</sub> domains, whereas the C<sub>1</sub> domain has the least variation. The C<sub>1</sub> domain connects to the C<sub>2</sub> domain through a short proline-rich segment, while the C<sub>2</sub> domain connects to the C<sub>3</sub> through one very long  $\beta$ -strand. The surface of the C-terminal domains are comprised of an even mix of polar (44%) and non-polar (56%) residues (as calculated by naccess 2.1.1). Each of these domains also carries an isopeptide bond, a characteristic signature present in gram-positive pili proteins (28) that are typically coordinated through three principal residues, lysine (K), asparagine (N), and aspartate (D). Within C<sub>1</sub>, C<sub>2</sub>, and C<sub>3</sub> domains the isopeptide bonds link the residues Lys1006-Asn1121, Lys1161-Asn1311, and Lys1338-Asn1473, respectively. In each of the three C-terminal domains of AgI/II this isopeptide bond covalently links the A and F strands, locking these strands together into a stable conformation (Fig 3.2 b). The isopeptide bonds observed in the C<sub>23</sub> domains of the *S. mutans* AgI/II's C-terminal region occur at equivalent positions in the structure as those identified in the homologous C<sub>23</sub> domains of *S. gordonii* SspB (18). The C<sub>23</sub> of *S. mutans* AgI/II superposes well with C<sub>23</sub> of SspB of *S. gordonii* with an RMSD (on C $\alpha$ ) of 0.906 Å (Fig 3.3).

*Metal ions.* There are three calcium ions present in the structure: two within the C<sub>2</sub> domain and one within the C<sub>3</sub> domain. Each ion is well coordinated by multiple interactions with backbone and sidechain oxygens (Fig 3.4a). In the C<sub>2</sub> domain, the first calcium ion interacts with Asn1155, Asp1189, Asp1191, and Gln1192 side-chains, as well as the main-chain oxygen of Tyr1156 and an apex water molecule, resulting in the pentagonal bi-pyramidal coordination geometry. A second calcium ion is also observed within the C<sub>2</sub> domain and is well coordinated through the sidechain oxygens of Asp1212 and Glu1215, and the backbone oxygens of Tyr1213, Lys1265, and Ala1267 that are located at loops

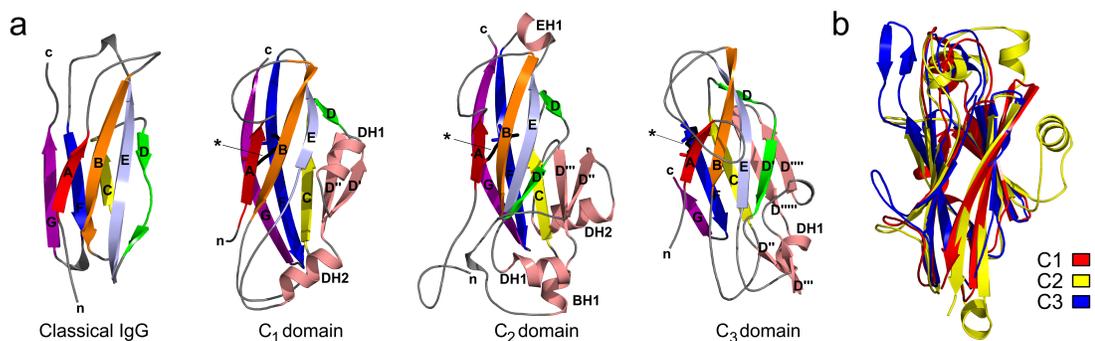


Figure 3.2: a) Each of the AgI/II C-terminal domains is shown using a rainbow colored ribbon model, where each strand is highlighted from A-G strands as seen in a typical IgG like-fold. Variations in the form of additional strands and helices occur between the D-E strands of this fold are highlighted in light pink. Isopeptide bonds are indicated with a (\*). b) Superposition of the  $C_1$ ,  $C_2$ ,  $C_3$  domains shows the similarities in these domains.

between the C and D strands. This second calcium ion is present near residues 1248-1330, which corresponds to a minimal calcium binding region identified in SspB (residues 1168-1250) (24). The  $C_3$  domain has the third calcium ion that is coordinated by the sidechain oxygens of residues Asp1388 and Gln1391 as well as by backbone oxygens of Tyr1389, Lys1434, Gly1435 and a water molecule once again located at loops between C and D strands similar to the second  $Ca^{2+}$  of the  $C_2$  domain. While most of the coordinating residues are also conserved in the loops between the C and D strands of the  $C_1$  domain, the presence of a metal ion site is precluded by the substitution of a Pro1051 for the Asp1212 or Asp1388 which are present in the  $C_2$  and  $C_3$  sites. Additionally, a magnesium ion (Fig 3.1a, purple) is observed, coordinated between the two molecules within the asymmetric unit through residues Asp1150 and Arg1018 of molecule A, and Thr1397 of molecule B.

*Carbohydrates.* In the final stages of refinements, eight positions near the surface of the AgI/II C-terminal domains resolved as large ring-like structures in difference electron density map. The native data had been obtained from a crystal soaked in a cryoprotectant solution that contained 30% w/v glucose and these densities were well fitted with glucose

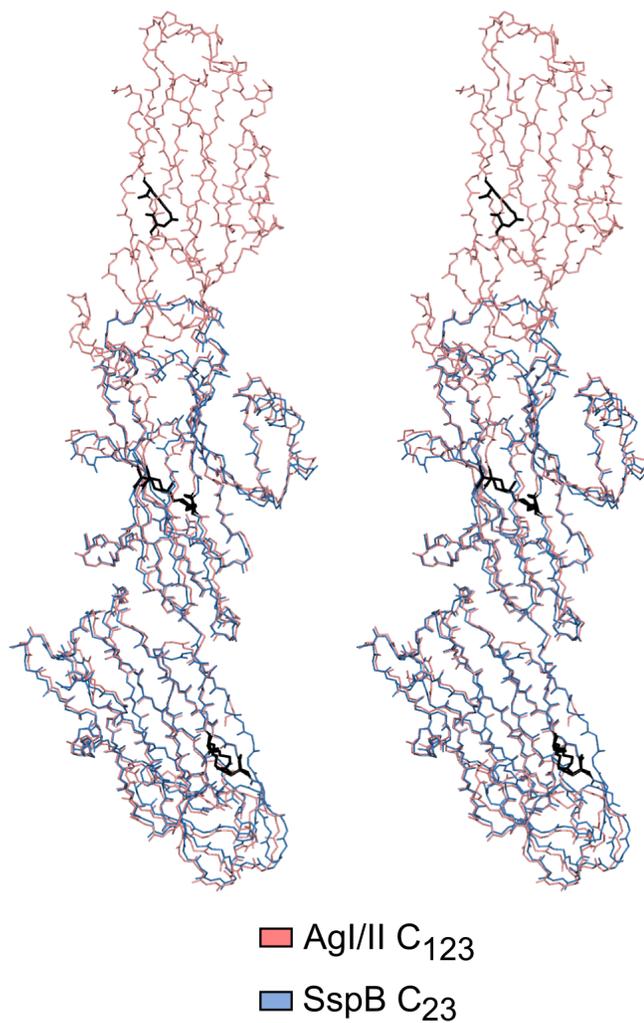


Figure 3.3: A superposition of the *S. mutans* AgI/II (red) and *S. gordonii* SspB (PDB:2WOY) (blue) structures is shown as traces of the backbone atoms. The two different AgI/II family members showed a high degree of similarity, with an average RMSD of 0.906 . The isopeptide bonds (black) of the C<sub>2</sub> and C<sub>3</sub> domains are located in equivalent positions on the two structures.

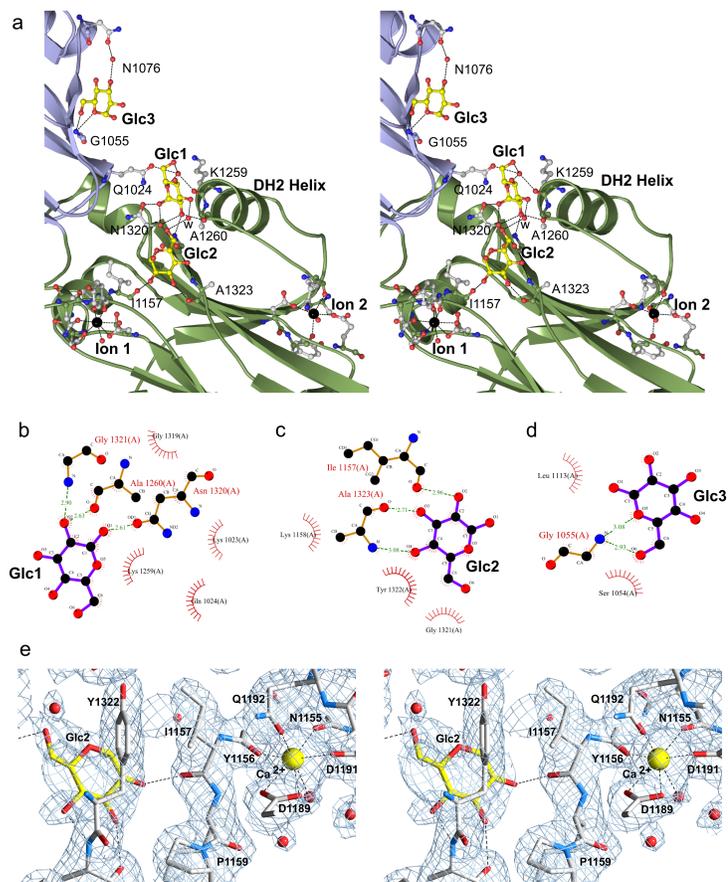


Figure 3.4: a) Stereo image shows the three glucose molecules which are bound within the cleft formed between the  $C_1$  and  $C_2$  domains of AgI/II. Sites Glc1 and Glc3 are present in both the molecules A and B, while Glc2 is only present in molecule A of the asymmetric unit. The DH2-helix of  $C_2$  domain is in close proximity to the cleft, and residues from DH2 interact with Glc1. Each glucose forms bonds with surrounding residues and waters. Glc1 and Glc2 are bridged through a shared water molecule (w). The interactions between Glc1 (b), Glc2 (c), and Glc3 (d) and AgI/II residues are shown below in a Ligplot diagram (35). The electron density map shown in stereo for the Glc2 sugar (e), which is stabilized in the cleft by hydrogen bonding with nearby residues including the backbone oxygen of isoleucine 1157. The backbone oxygen of the neighboring tyrosine 1156 forms part of a coordination site for the nearby calcium ion.

molecules (Glc1-5 shown in Fig 3.1a). Glc1, Glc3, and Glc4 are conserved in the both molecules of asymmetric unit while Glc2 (mol1 only) and Glc5 (interface of mol1 and mol2) each exist only at one position within the asymmetric unit (Fig 3.1 a). Glc1, Glc2, and Glc3 were in close proximity to each other (12 Å) within the cleft formed by the C<sub>1</sub> and C<sub>2</sub>, and may represent a binding site for a complex carbohydrate structure such as a glycosylation or sugar polymer (Fig 3.4a). Glc4 and Glc5, which lie outside the cleft (Fig 3.1 a), are observed at intermolecular contact sites between different C<sub>123</sub> molecules within the crystal. The oxygens of Glc1, Glc2, and Glc3 make direct hydrogen bonds with AgI/II residues (Fig 3.4 b,c,d), as well as water mediated interactions with nearby amino acids that appear to stabilize the position of these glucoses. Glc1 interacts with sidechain nitrogens of Asn1320, Gln1024, backbone nitrogen of Gly1321, and nitrogen and oxygens of backbone of the C<sub>2</sub> DH2 helical residues Lys1259, and Ala1260 (Fig 3.4b). Neighboring Glc1 and Glc2 are linked through a single water molecule (Fig 3.4 a), mimicking a linked sugar chain. Glc2 is additionally hydrogen bonded with Ile1157 and Ala1323. Glc3 is stabilized through interactions with Asn1076, and Gly1055 (Fig 3.4 d).

*Binding studies - Surface plasmon resonance.* AgI/II has two separate segments capable of adherence to immobilized salivary agglutinin (16), first, the N-terminal and central regions comprised of the third A-repeat through the first P-repeat (A<sub>3</sub>VP<sub>1</sub>) and second, the C-terminus. With the identification of three domains within the C-terminus, we attempted to further delineate the binding within the C-terminus. The full-length recombinant AgI/II and C-terminal polypeptides comprising C<sub>123</sub>, C<sub>12</sub>, C<sub>23</sub>, and the individual C<sub>1</sub>, C<sub>2</sub>, and C<sub>3</sub> domains were investigated using surface plasmon resonance (SPR) for their ability to interact with immobilized SAG. Among the C-terminal fragments, C<sub>123</sub> and C<sub>12</sub> exhibited measurable binding (Fig 3.5a). Kinetics experiments conducted at multiple concentrations (Fig 3.5 b), exhibited a relatively higher affinity for C<sub>12</sub> with an estimated dissociation constant (KD) of 57 nM, compared to C<sub>123</sub> and full length (CG14) which had KD values

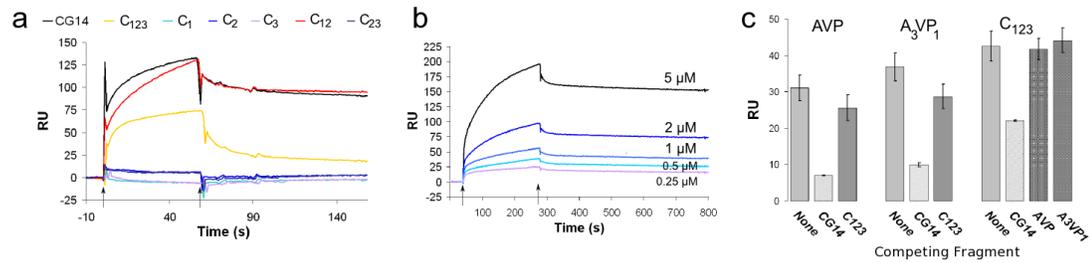


Figure 3.5: a) SAG was immobilized and AgI/II polypeptides ( $2 \mu\text{M}$  each) were injected over the CM5 chip surface. CG14 (black), C<sub>123</sub> (orange), and C<sub>12</sub> (red) each displayed a strong binding response (RU, resonance units,  $1 \text{ RU} = 1 \text{ picogram/mm}^2$ ) accompanied by slower dissociations that are indicative of adherence to the SAG coated surface. b) The interaction between C<sub>12</sub> with immobilized SAG at multiple concentrations is shown. Kinetics fitting with multiple C<sub>12</sub> concentrations estimated the KD to be 57 nM. c) Competition experiments were performed for the binding of AVP, A<sub>3</sub>VP<sub>1</sub>, or C<sub>123</sub> to immobilized SAG. The adherence of AgI/II fragments to SAG without competition (labeled as none) is shown first, followed by measured adherence when competed against other AgI/II fragments (labeled CG14, C<sub>123</sub>, AVP, or A<sub>3</sub>VP<sub>1</sub>). All experiments were carried out in triplicate.

of 410 nM and 69 nM, respectively.

Next we tested whether the N-terminal (AVP or subset A<sub>3</sub>VP<sub>1</sub>) and C-terminal (C<sub>123</sub>) polypeptides of AgI/II could compete with each other for binding to immobilized SAG (Fig 3.5 c) to determine if these regions adhere to the same or different sites on SAG. Binding of full-length CG14 construct substantially reduced the subsequent binding of C<sub>123</sub>, A<sub>3</sub>VP<sub>1</sub>, and AVP, by 48, 73, and 77%, respectively. However, C<sub>123</sub> did not greatly inhibit the binding of the amino-terminal constructs, nor did the amino-terminal fragments inhibit the binding of C<sub>123</sub>. Control SPR experiments, with immobilized N- and C-terminal AgI/II fragments did not show measurable self- or cross-interactions with other AgI/II fragments. Lack of competition between the AgI/II N-terminal and C-terminal regions for SAG suggests that these regions interact with independent or distinct sites on the immobilized SAG.

*S. mutans adherence inhibition.* To further establish the physiological relevance of C<sub>12</sub> binding, competition experiments with *S. mutans* were performed. The binding of

3H-labelled *S. mutans* to SAG-coated hydroxyapatite beads was evaluated in the presence and absence of fluid-phase competitors. In the first set of experiments (Table 3.3, top), the abilities of recombinant full-length AgI/II (CG14) and the N-terminal (A<sub>3</sub>VP<sub>1</sub>) and C-terminal (C<sub>123</sub>) fragments to inhibit the interaction of *S. mutans* with immobilized SAG were measured. Both A<sub>3</sub>VP<sub>1</sub> and the C-terminal fragments demonstrated partial inhibition of bacterial adherence at 5 M concentration, (28% and 43%, respectively), whereas the full-length molecule demonstrated substantial inhibition (70%) at the lower 1 M concentration. Adding the A<sub>3</sub>VP<sub>1</sub> and the C-terminal polypeptides together did not improve the ability to inhibit adherence suggesting that the orientation or spacing of two separate binding sites within the full-length adhesin may contribute to the strength of the binding interaction. Next, we tested the abilities of the C<sub>12</sub> and C<sub>23</sub> fragments to competitively inhibit bacterial adherence compared to the full-length molecule (CG14) and entire C-terminal fragment (C<sub>123</sub>) at 1 and 5 μM concentrations (Table 3.3, bottom). The C<sub>12</sub>, but not the C<sub>23</sub> fragment was able to serve as a competitive inhibitor. Consistent with the results of the SPR binding studies, the C<sub>12</sub> fragment again demonstrated a greater ability to interact with SAG than the entire C-terminal fragment (C<sub>123</sub>).

*Transmission electron microscopy studies.* AgI/II purified from *S. mutans* was visualized by transmission electron microscopy as an elongated molecule that contains larger bulges at its termini (Fig 3.6 a). The average length of AgI/II molecules in these electron microscopic images was estimated to be 65.8 nm and conforms well with the 68 nm length of full-length molecule estimated from our previous structural study (16). The length of the C-terminus was measured to be 13 nm based on the crystal structure described in this report. This is twice the measured diameter (6 nm) of the variable-region domain in crystal structures containing the V-region (16, 25). These longer carboxy termini can be clearly seen on one end of the molecule, with the smaller V-region at the other end of the molecule,

| <b>Inhibitor (<math>\mu\text{M}</math>)</b>   | <b>Inhibition <math>\pm</math> sd*</b> |
|---|--|
| No Inhibitor                                  | 0                                      |
| CG14 (1)                                      | 70.0 $\pm$ 10.6                        |
| AVP (1)                                       | 18.5 $\pm$ 17.0                        |
| AVP (5)                                       | 25.7 $\pm$ 5.84                        |
| C <sub>123</sub> (1)                          | 3.95 $\pm$ 1.17                        |
| C <sub>123</sub> (5)                          | 42.9 $\pm$ 12.9                        |
| AVP + C <sub>123</sub> (1 $\mu\text{l}$ each) | 23.2 $\pm$ 6.70                        |
| AVP + C <sub>123</sub> (5 $\mu\text{l}$ each) | 38.9 $\pm$ 4.94                        |

| <b>Inhibitor (<math>\mu\text{M}</math>)</b> | <b>Inhibition <math>\pm</math> sd*</b> |
|---|--|
| No Inhibitor                                | 0                                      |
| CG14 (1)                                    | 62.8 $\pm$ 3.11                        |
| C <sub>123</sub> (1)                        | 6.33 $\pm$ 4.06                        |
| C <sub>123</sub> (5)                        | 25.1 $\pm$ 6.08                        |
| C <sub>12</sub> (1)                         | 27.7 $\pm$ 7.09                        |
| C <sub>12</sub> (5)                         | 45.2 $\pm$ 5.43                        |
| C <sub>23</sub> (1)                         | 5.53 $\pm$ 2.87                        |
| C <sub>23</sub> (5)                         | 4.87 $\pm$ 6.95                        |

Table 3.3: Competitive inhibition to SAG-coated hydroxyapatite by AgI/II fragments

Top: independent vs cooperative inhibition by N- and C-terminus of AgI/II.

Bottom: inhibition by sub-fragments of C-terminus of AgI/II

\*Values are the means  $\pm$  standard deviation(sd) of triplicate samples.

which are connected through the alanine- and proline-rich stalk. A composite comprehensive tertiary structure that contains two distinct and independent binding sites for SAG was built (Fig 3.6 b) based upon the current and previous X-ray crystallographic studies of AgI/II regions (16). This model now well describes the majority of the AgI/II structure.

## DISCUSSION

The C-terminal region consists of three ( $C_1$ ,  $C_2$ , and  $C_3$ ) DEv-IgG domains (23), each of which displays distinct variations in the structural elements between the “D” and “E” strands on the common underlying architecture of the IgG fold. This fold is mainly observed in proteins connected with protein:protein interactions, such as *Staphylococcus aureus* clfA (23) and CNA (26). The domains of the C-terminus of AgI/II are aligned to give an extended appearance similar to the fibronectin-like domains (27). Each of the  $C_1$ ,  $C_2$  and  $C_3$  domains of *S. mutans* AgI/II contains an isopeptide bond, located at an identical position of the DEv-IgG fold. The isopeptide bonds present in AgI/II’s C-terminus follow an architecture similar to the domains of the *Corynebacterium diphtheriae* pilin SpaA, with the isopeptide bond joining the “A” and “F” strands, and linking the 2 opposing  $\beta$ -sheets (28). These isopeptide bonds likely provides stability to the fold, for protection from proteolytic enzymes and shear forces brought by salivary flow (1 liter/day) that pose particular challenges to design and function of bacterial surface proteins within the oral cavity.

SPR experiments on the C-terminal fragments  $C_1$ ,  $C_2$ ,  $C_3$ ,  $C_{12}$  and  $C_{23}$  indicated that a cooperative surface formed by  $C_{12}$  is required for adherence to SAG. This result is in concurrence with two other important studies on AgI/II. First it is in immediate proximity to the previously identified inhibitory peptide (mapping to residues 1030-1049 on NG8 AgI/II) (15) on  $C_1$ , which is capable of blocking *S. mutans* adhesion to SAG and recolonization in human subjects. Secondly, the DH2 helix present within the  $C_2$  AgI/II family

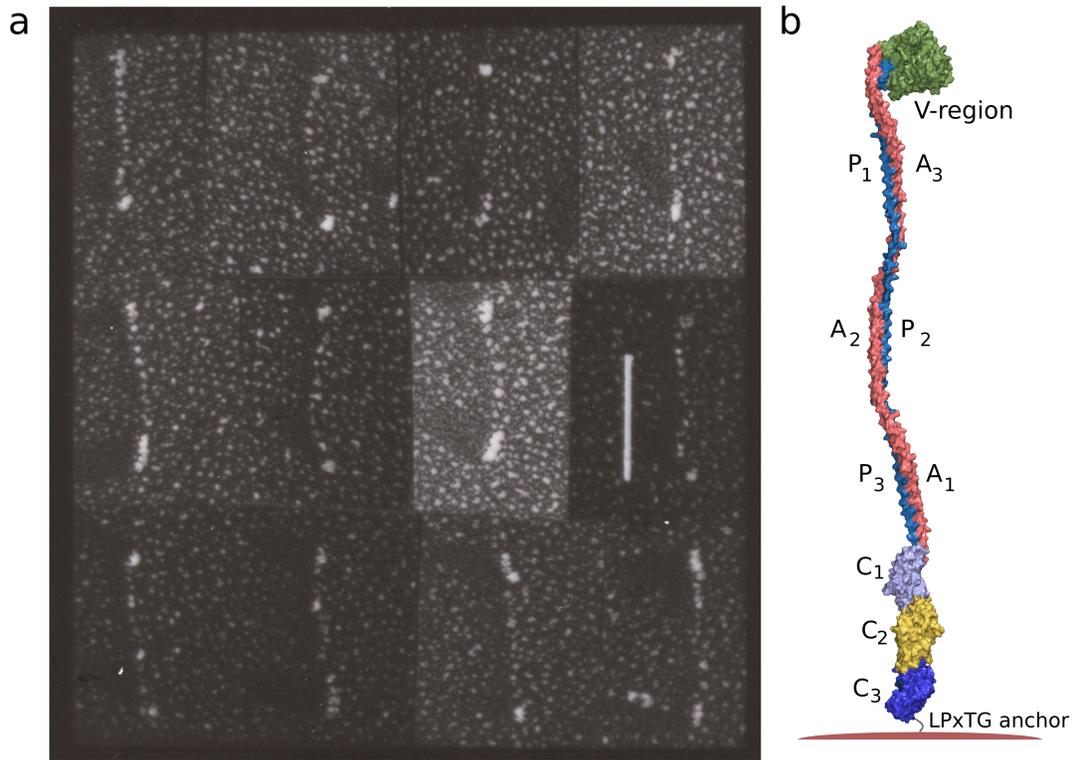


Figure 3.6: a) Figure shows a composite of electron micrographs of unidirectionally metal-shadowed streptococcal AgI/II molecules at magnification  $\times 235,000$  (Bar = 50 nm). The micrographs of AgI/II illustrate an elongated model with globular domains present at the termini. The longer globular C-terminus ( $\sim 13$  nm) present at an opposing end from the smaller V-region ( $\sim 6$  nm) can be clearly visualized in these micrographs. The end-to-end length of the molecules is 65.8 (S.D. + 3.6) nm, similar to predicted dimensions from analytical ultracentrifugation of AgI/II (16). b) Illustrated is a composite model of tertiary structure for AgI/II including the extended hybrid of the  $\alpha$ -helical A-region (red) and the polyproline type II helical P-region (blue), that separates the globular V- (green) and C-terminal regions (light blue-C<sub>1</sub>, yellow-C<sub>2</sub>, blue-C<sub>3</sub>). The two distinct and non-competing binding sites for SAG are contained within the A<sub>3</sub>VP<sub>1</sub> and C<sub>1</sub>C<sub>2</sub> regions, which are located at opposite ends of the A-P stalk. Following the C<sub>3</sub> domain is the LPxTG motif that anchors the protein to the cell wall peptidoglycan layer.

member of *S. gordonii* SspB was shown to specifically recognize the minor fimbrial protein Mfa1 of *Porphyromonas gingivalis* (29,30). In addition, this kind of specific adherence utilizing multiple DEv-IgG domains to latch on to receptors has also been observed for other gram-positive surface adhesins, such as the *Staphylococcus aureus* CNA and *Staphylococcus epidermidis* SdrG (31). Therefore it appears likely from these results that the C-terminal SAG adherence site is contained within the C<sub>12</sub> domains.

Earlier we identified the presence of two distinct SAG adherence regions, A<sub>3</sub>VP<sub>1</sub> and the C-terminus (16). The present SPR studies now further elucidates that the cell surface proximal (C-terminal) and distal (A<sub>3</sub>VP<sub>1</sub>) ends of AgI/II bind non-competitively to SAG (Fig. 3.5 c). The ability of the two different AgI/II fragments to simultaneously adhere to immobilized SAG indicates that the AVP and C-terminus regions bind to distinct sites on SAG. Given that the AVP region (Antigen I) is prone to proteolysis, while the Antigen II (C-terminal region) is more resistant to proteolysis (22, 32), the presence of an independent C-terminal adherence site for SAG suggests that *S. mutans* has evolved secondary adherence mechanisms for bacterial colonization of the tooth surface.

SAG's significant glycosylations have long been studied as a possible mode for the binding of AgI/II with SAG (14, 33, 34). Interactions of *S. mutans* and SAG can be inhibited by various amino-sugars as well as other compounds with primary amine groups and less efficiently by lactose and melibiose (14). In addition, binding of AgI/II homologue SspB from *S. gordonii* to SAG is sialic acid dependent (33). The structural report on the V-region implicated a hydrated pocket as a potential sugar binding site (25), and this site was occupied by a non-carbohydrate ligand in the structure of A<sub>3</sub>VP<sub>1</sub> (16). However, the identification and measurement of an interaction between a simple or complex sugar molecule with any region of AgI/II is yet to be substantiated. In the current structure, the serendipitous identification of multiple glucose molecules within a large trench formed between the C<sub>12</sub> domains suggests a possible mechanism for recognition of the highly glycosylated

SAG.

It is unlikely though that a simple glucose would represent an ideal carbohydrate motif to be recognized by the AgI/II C-terminus. The presence of multiple bound carbohydrates (Glc1, Glc2, and Glc3) in near proximity (12 Å) also suggests a more complex glyco-structure would fit the cleft, with increased affinity and better recognition than a simple monosaccharide. Branched glycosylations or polymers like dextran might be candidates for such recognition. Glucose and other simple monosaccharides have little inhibitory effect on the interaction of *S. mutans* with SAG (14), whereas the interaction of oral streptococci with SAG are impaired upon enzymatic removal of the complex sugar structures on SAG (34). SAG's multiple globular scavenger-receptor (SRCR) domains are interspersed with segments (SIDs) predicted to have O-glycosylation sites that separate the SRCR domains, and these glycosylations are predicted to produce the elongated structure for SAG (12). The binding sites of Glc1, Glc2, and Glc3 could interact with the above mentioned SIDs of from SRCRs, acting alone or in addition to other direct protein:protein interactions between AgI/II and SAG.

Structural changes in the presence or absence of calcium have previously been suggested to be important for the binding of AgI/II and SAG (24). Within the cleft region, structural elements link the two calcium ion sites within the C<sub>2</sub> domain to the Glc1 and Glc2 sites. The main chain oxygen of Tyr1156 interacts with a calcium ion and its adjacent residue (Ile1157) interacts with Glc2 (Fig 3.4, e). Separately, the leading loop to the C<sub>2</sub> DH2 helix hosts a second calcium ion, and Glc1 interacts with Lys1259 and Ala1260 of the C<sub>2</sub> DH2 helix. These ions could stabilize the orientations of residues to preferably interact with carbohydrate chains. Such an arrangement may explain the observed requirement for calcium in SAG binding (14) and it is tempting to speculate that either of these calcium ions could provide a mechanism for conformational changes that modulate an interaction of the AgI/IIs C-terminus to carbohydrate chains.

With the resolution of the C-terminal structure of AgI/II, almost the entire AgI/II protein structure is now known with only a minimal N-terminal portion of the AgI/II preceding the alanine-rich region remaining uncharacterized (Fig. 3.6 b). The C<sub>123</sub> structure, which is 13 nm in length, in addition to the 50 nm extended AVP region (16), produces an estimated overall AgI/II length of 63 nm (Fig 3.6 b) which concurs well with our earlier results from analytical ultracentrifugation (16). The electron microscopy studies (Fig. 3.6 a) confirm the extended nature and dumbbell shape of AgI/II, particularly the visualization of the larger extended C-terminal domain on one end of the molecule and the smaller V-region that is positioned at the opposite apex of the molecule. The C<sub>123</sub> appears as a continuous Van der Waals surface in the electron micrographs. Unlike the A/P repeat motifs, the C-terminus has utilized structural motifs consisting of repetitions of the DEv-IgG fold. The use of repeated sequences and structural units to extend gram-positive surface proteins from the cell-surface is now being seen as a common and widespread architecture for bacterial adhesion proteins.

In summary, the structure of the C-terminus of AgI/II has revealed three domains of the DEv-IgG fold (C<sub>1</sub>, C<sub>2</sub>, and C<sub>3</sub>) and has enabled the construction of a comprehensive structural model of the adhesin as visualized from electron microscopy images. (Fig 3.6). SAG binding function was localized within the first two domains (C<sub>12</sub>) of the C-terminus, which also showed multiple interactions with carbohydrates within the crystal structure. The presence of two distinct N- and C-terminal SAG binding sites within AgI/II, as well as the localization within the C<sub>12</sub> fragment were confirmed by these regions ability to inhibit *S. mutans* adherence to SAG, further indicative of its physiological relevance in the context of the cell-surface localized adhesin as it is displayed on *S. mutans* cells. The distinct types of interaction between AgI/II and SAG warrants further studies that would elaborate on the mechanisms of the two independent forms of adherence. In this context, structural and functional studies on the interaction between gp340's SRCR domains and

AgII fragments would provide intimate details on these bacterial-host interactions, and investigations aimed at disrupting these interactions could produce novel therapeutics to impede the adherence of oral streptococci to tooth surface.

## REFERENCES

1. Loesche, W. J. (1986) *Microbiol Rev* **50**(4), 353-380
2. Nomura, R., Nakano, K., Nemoto, H., Fujita, K., Inagaki, S., Takahashi, T., Taniguchi, K., Takeda, M., Yoshioka, H., Amano, A., and Ooshima, T. (2006) *J Med Microbiol* **55**(Pt 8), 1135-1140
3. Brady, L. J., Maddocks, S. E., Larson, M. R., Forsgren, N., Persson, K., Deivanayagam, C. C., and Jenkinson, H. F. (2010) *Mol Microbiol* **77**(2), 276-286
4. Crowley, P. J., Brady, L. J., Piacentini, D. A., and Bleiweis, A. S. (1993) *Infect Immun* **61**(4), 1547-1552
5. Munro, G. H., Evans, P., Todryk, S., Buckett, P., Kelly, C. G., and Lehner, T. (1993) *Infect Immun* **61**(11), 4590-4598
6. Pecharki, D., Petersen, F. C., Assev, S., and Scheie, A. A. (2005) *Oral Microbiol Immunol* **20**(6), 366-371
7. Matsumoto-Nakano, M., Tsuji, M., Inagaki, S., Fujita, K., Nagayama, K., Nomura, R., and Ooshima, T. (2009) *Oral Microbiol Immunol* **24**(5), 427-430
8. Love, R. M., McMillan, M. D., Park, Y., and Jenkinson, H. F. (2000) *Infect Immun* **68**(3), 1359-1365
9. Crowley, P. J., Brady, L. J., Michalek, S. M., and Bleiweis, A. S. (1999) *Infect Immun* **67**(3), 1201-1206

10. Zhang, S., Green, N. M., Sitkiewicz, I., Lefebvre, R. B., and Musser, J. M. (2006) *Infect Immun* **74**(7), 4200-4213
11. Jenkinson, H. F., and Demuth, D. R. (1997) *Mol Microbiol* **23**(2), 183-190
12. Ligtenberg, A. J., Veerman, E. C., Nieuw Amerongen, A. V., and Mollenhauer, J. (2007) *Biol Chem* **388**(12), 1275-1289
13. Madsen, J., Mollenhauer, J., and Holmskov, U. (2010) *Innate Immun* **16**(3), 160-167
14. Hajishengallis, G., Koga, T., and Russell, M. W. (1994) *J Dent Res* **73**(9), 1493-1502
15. Kelly, C. G., Younson, J. S., Hikmat, B. Y., Todryk, S. M., Czisch, M., Haris, P. I., Flindall, I. R., Newby, C., Mallet, A. I., Ma, J. K., and Lehner, T. (1999) *Nat Biotechnol* **17**(1), 42-47
16. Larson, M. R., Rajashankar, K. R., Patel, M. H., Robinette, R. A., Crowley, P. J., Michalek, S., Brady, L. J., and Deivanayagam, C. (2010) *Proc Natl Acad Sci U S A* **107**(13), 5983-5988
17. Brady, L. J., Piacentini, D. A., Crowley, P. J., Oyston, P. C., and Bleiweis, A. S. (1992) *Infect Immun* **60**(3), 1008-1017
18. Forsgren, N., Lamont, R. J., and Persson, K. (2010) *J Mol Biol* **397**(3), 740-751
19. Z. Otwinowski, W. M. (1997) *Methods in Enzymology* **276**(A), 307-326
20. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1), 2126-2132
21. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Gross-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R.

- J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* **54**(Pt 5), 905-921
22. Russell, M. W., Bergmeier, L. A., Zanders, E. D., and Lehner, T. (1980) *Infect Immun* **28**(2), 486-493
23. Deivanayagam, C. C., Wann, E. R., Chen, W., Carson, M., Rajashankar, K. R., Hook, M., and Narayana, S. V. (2002) *Embo J* **21**(24), 6660-6672
24. Duan, Y., Fisher, E., Malamud, D., Golub, E., and Demuth, D. R. (1994) *Infect Immun* **62**(12), 5220-5226
25. Troffer-Charlier, N., Ogier, J., Moras, D., and Cavarelli, J. (2002) *J Mol Biol* **318**(1), 179-188
26. Symersky, J., Patti, J. M., Carson, M., House-Pompeo, K., Teale, M., Moore, D., Jin, L., Schneider, A., DeLucas, L. J., Hook, M., and Narayana, S. V. (1997) *Nat Struct Biol* **4**(10), 833-838
27. Leahy, D. J. (1997) *Annu Rev Cell Dev Biol* **13**, 363-393
28. Kang, H. J., Paterson, N. G., Gaspar, A. H., Ton-That, H., and Baker, E. N. (2009) *Proc Natl Acad Sci U S A* **106**(40), 16967-16971
29. Brooks, W., Demuth, D. R., Gil, S., and Lamont, R. J. (1997) *Infect Immun* **65**(9), 3753-3758
30. Chung, W. O., Demuth, D. R., and Lamont, R. J. (2000) *Infect Immun* **68**(12), 6758-6762
31. Zong, Y., Xu, Y., Liang, X., Keene, D. R., Hook, A., Gurusiddappa, S., Hook, M., and Narayana, S. V. (2005) *Embo J* **24**(24), 4224-4236

32. Russell, M.W., and Lehner, T. (1978) *Arch Oral Biol* **23**(1), 7-15
33. Demuth, D. R., Golub, E. E., and Malamud, D. (1990) *J Biol Chem* **265**(13), 7120-7126
34. Loimaranta, V., Jakubovics, N. S., Hytonen, J., Finne, J., Jenkinson, H. F., and Stromberg, N. (2005) *Infect Immun* **73**(4), 2245-2252
35. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) *Protein Eng* **8**(2), 127-134
36. Brady, L. J., Cvitkovitch, D. G., Geric, C. M., Addison, M. N., Joyce, J. C., Crowley, P. J., and Bleiweis, A. S. (1998) *Infect Immun* **66**(9), 4274-4282
37. Laskowski RA, M. M., Moss DS and Thornton JM. (1993) *Journal of Applied Crystallography* **26**, 283-291

## Footnotes

The authors thank Shaan Khaled and Manisha Patel for help with the purification AgI/II fragments. MRL and CD also thank the University of Alabama-Comprehensive Cancer Center X-Ray core and Northeastern Collaborative Access Team (NE-CAT) facility at the Advanced Photon Source/Argonne National Laboratory. This work was supported by National Institutes of Health (NIH)-National Institute of Dental and Craniofacial Research (NIDCR) grants RO1 DE017737 (to CD) and by the DART Training Grant NIDCR T-32 DE0176707 (to MRL)

## **Accession codes**

The atomic coordinates of the AgI/II Carboxy-terminus have been deposited in the Protein Data Bank under the accession code 3QE5.

## **Author Contributions**

MRL, LJB, CK, and CD designed research; MRL, KRR, CD conducted crystallographic experiments and structure solution, MRL, PJC, LJB performed functional experiments. CK and TJM performed microscopy studies. MRL, LJB, TJM, CK, and CD analyzed data. MRL, LJB, CK, and CD wrote the manuscript.

## CHAPTER 4

### CONCLUSION

#### 4.1 Determination of the AgI/II Structure

Determining the structure for the full length *Streptococcus mutans* AgI/II protein was challenging due to its large size (170 kDa). The first hurdle was obtaining well-diffracting protein crystals, as large macromolecules often prove to be more difficult targets to crystallize [61]. AgI/II is easily degraded by proteolytic enzymes, a characteristic observed with its discovery [18, 19]. This was a complicating factor for crystallization of AgI/II, as its recombinantly produced protein had proteolytic degradation fragments that were difficult to remove by conventional chromatographic methods. After unsuccessful crystallization attempts for the largest fragments of the AgI/II, we continued to divide AgI/II into smaller pieces using features identified within the AgI/II primary sequence that delineate the extents of the amino acid regions. The success in obtaining crystals of AgI/II occurred after we successfully employed a divide-and-conquer strategy to yield crystallizable fragments (Fig 4.1). Well-defined boundaries for crystallizable fragments were determined by the extents of the globular V-region structure [27], the alanine-rich repeat sequences, proline-rich repeat sequences, and division of the proline-rich repeat sequences with the carboxy-terminus. The next sections describe in detail our efforts to crystallize AgI/II.

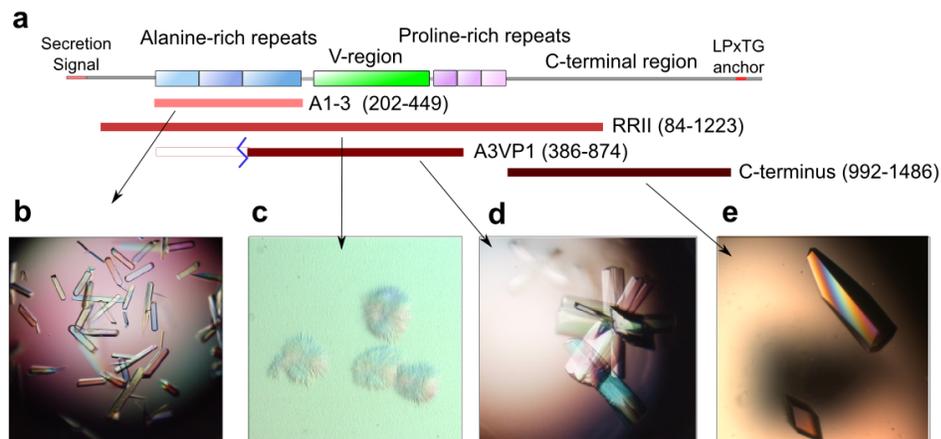


Figure 4.1: *Crystallization of AgI/II*- Several fragments of AgI/II were crystallized to determine the AgI/II structure (a). The A<sub>1-3</sub> crystals (b) led to fiber-like diffraction and the needle-like crystals of RRII (c) that failed grow to large sizes were found first. A<sub>3</sub>VP<sub>1</sub> was obtained by removing the A<sub>1-2</sub> repeats with limited cleavage using the factor Xa enzyme. A<sub>3</sub>VP<sub>1</sub> crystals (d) had high resolution diffraction (1.8 Å). The C-terminal region crystals (e) diffracted to 2.5 Å.

#### 4.1 N-terminal Region Crystallization

After screening for crystallization conditions, we obtained crystals of the A-region (A<sub>1-3</sub>). These crystals displayed fiber-like diffraction patterns (Fig 2.11, pg. 41), where one of the crystal cell dimensions (b) was unusually large at 671.43 Å. The unusual cell dimensions, combined with poor and asymmetric diffraction prevented further structural determination using these data. However, the large cell dimension provided the first hint that the AgI/II A-repeats may adopt an elongated structure. Subsequently, we obtained crystals of needle clusters for the fragment encompassing amino acids 84-1223 (RRII). These crystals could not be grown to larger sizes, making them unsuitable for structure solution. During protein purifications we observed that all the AgI/II fragments that contained both A- and P- repeats, such as RRII, had degradation as did native AgI/II. Protein degradations are known to be highly detrimental, as the degraded fragments can fit sites of a growing crystalline

lattice and prevent further crystal growth.

Seeking to remove protein degradations from the large N-terminal AgI/II fragments, we developed a plasmid expression system with two affinity tags on the N- and C- termini. This allowed sequential affinity purifications to yield intact fragments. The A<sub>1</sub>VP<sub>1</sub> fragment produced with this system was highly pure. The removal of the large N-terminal MBP affinity tag was required prior to crystallization, but the factor Xa enzyme cleaved at sites past A<sub>1</sub>VP<sub>1</sub>'s engineered cleavage site. Once the cleavage reaction was complete, only a single AgI/II fragment existed (Fig 2.12, pg. 45). Mass spectroscopy determined the fragment size and N-terminal amino acid sequencing revealed that the first two A-repeats of A<sub>1</sub>VP<sub>1</sub> had been cleaved, leaving only the A<sub>3</sub>-repeat to match the single P<sub>1</sub>-repeat (A<sub>3</sub>VP<sub>1</sub>). Screening with this fragment, we obtained large crystals that finally yielded high-resolution diffraction. The phases for A<sub>3</sub>VP<sub>1</sub> were resolved by molecular replacement using the V-region structure [27], and we began to build a structure for the AgI/II N-terminus.

#### **4.1 C-terminal Region Crystallization**

Crystallization trials for an AgI/II fragment comprising the complete carboxy-terminus (C<sub>123</sub>, a.a. 992-1486) also succeeded, yielding large crystals from which data were collected at 2.5 Å resolution. No solved structures for any homologous proteins were available at the time of this work, so we could not employ molecular replacement as we had for A<sub>3</sub>VP<sub>1</sub>. In addition, the low percentage of methionine residues within the large C-terminus precluded solving this structure through multi-wavelength anomalous dispersion (MAD), and therefore we employed multiple isomorphous replacement (MIR) for the structure solution. We screened crystals of the C-terminus for possible heavy atom incorporation and collected X-ray data for possible derivatives. Three heavy atom derivatives incorporating

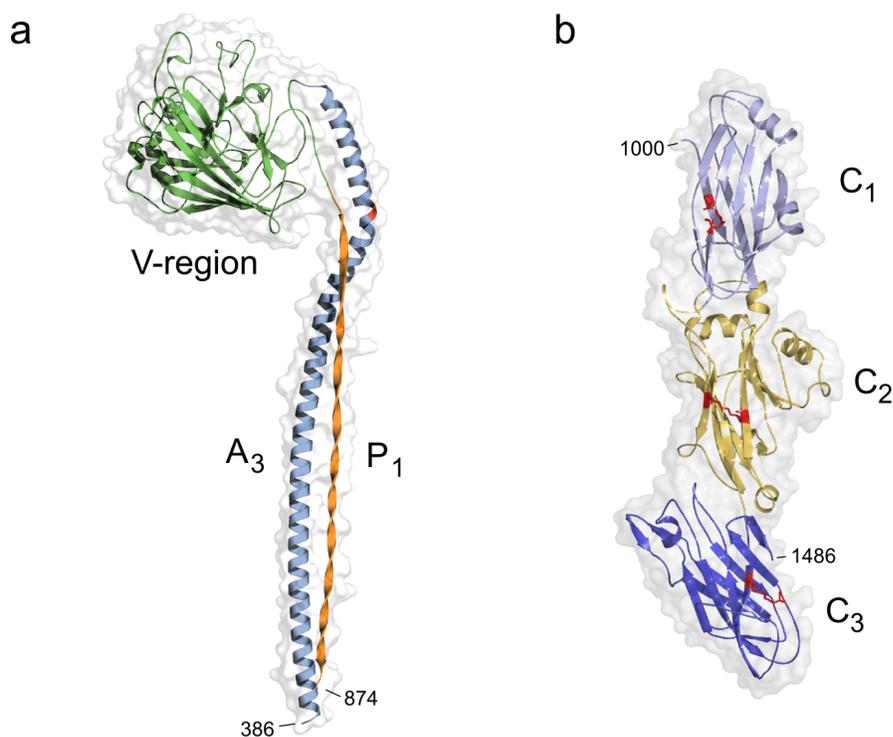


Figure 4.2: *The structures of the N- and C-terminal fragments of AgI/II- A<sub>3</sub>VP<sub>1</sub> has a long stalk composed of both  $\alpha$ -helices (A-repeats) and a polyproline type II-helix (P-repeats) (a). The V-region sits at the apex of the stalk and bridges the two helical segments. The carboxy-terminus of AgI/II consists of 3 domains of DE-variant IgG-like fold (b). These structures are deposited in the Protein Data bank under entries 3IOX, 3IPK (A<sub>3</sub>VP<sub>1</sub>) and 3QE5 (C<sub>123</sub>).*

thimerosal, sodium iodide, and thallium (III) acetate were able to provide phases that resulted in clear electron densities (Table 3.2, pg. 66), at which point we were able to build the C<sub>123</sub> structure manually.

#### 4.1 N-terminal AgI/II structure

The A<sub>3</sub>VP<sub>1</sub> structure revealed that the A- and P- repeats were associated as a single helical stalk. In the stalk, the single A-repeat (A<sub>3</sub>) exists as a long  $\alpha$ -helix, while the single P-repeat (P<sub>1</sub>) is a polyproline-type II (PPII) helix. Earlier, Van Dolleweerd *et al* had identified

a monoclonal antibody that recognized an unusual co-epitope formed from discontinuous AgI/II A- and P- repeat sequences [54]. The A<sub>3</sub>VP<sub>1</sub> structure now clearly explains how two sequence repeats which are discontinuous within AgI/II's primary sequence form a co-epitope, as the A- and P- repeats join within the AP stalk as a contiguous surface. The AP stalk is stable in solution and the dissociation constant for the A-P interaction is 62.9 nM (Table 2.1, pg. 30). This tight association makes it unlikely that the two regions would separate after forming the AgI/II stalk. The globular V-region, interposed between the sequences of the A<sub>3</sub> and P<sub>1</sub> repeats, exists at the apex of the stalk.

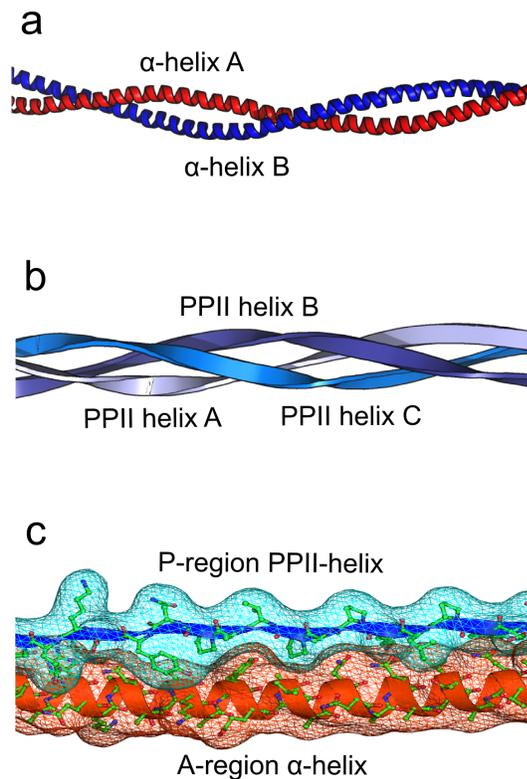


Figure 4.3: A comparison of fibrillar structures, old and new- The most notable feature of the AgI/II structure is the highly extended AP stalk that contributes the majority of AgI/II's length. Shown are the classical  $\alpha$ -helical coiled-coils, *ex. tropomyosin* (a). Fibrils made of three PPII-helices, *ex. collagen* (b). The *S. mutans* AgI/II alanine-rich and proline-rich repeats are a new form of hybrid structure that integrates both  $\alpha$ - and PPII-helices (c).

The AP stalk of AgI/II now represents a new fibrillar form composed of elements found separately in the structures of classical fibrillar proteins. Over half a century ago, Linus Pauling and Robert Corey first proposed that  $\alpha$ -helices could pair to form  $\alpha$ -helical coiled-coil structures able to explain the periodicities present in X-ray diffraction patterns of  $\alpha$ -keratin [62]. When  $\alpha$ -helices have repeating heptad sequence motifs and hydrophobic residues appearing every 3<sup>rd</sup> or 4<sup>th</sup> amino-acid position, they will have a resulting hydrophobic face on one side of the helix. Two such  $\alpha$ -helices can coil about each other, associated by the hydrophobic surfaces. This classical  $\alpha$ -helical coiled-coil fibrillar form exists in a multitude of elongated proteins including keratin, myosin, tropomyosin, and the streptococcal M protein [63]. AgI/II shares similarities with these fibrillar proteins formed by  $\alpha$ -helices, as seen in the A<sub>1-3</sub> sequence where we have identified 19 heptad motifs that provide hydrophobic interactions within the AgI/II fibrillar structure.

A much different fibrillar form is seen in one of the most well known proteins, collagen. Collagen has three polypeptide chains, with each chain having a repeated Gly-X-Y triplet where the X and Y positions are substituted by proline and 4-hydroxyproline residues [64]. These chains display an extended left-handed polyproline type-II (PPII) conformation, and together the three chains form a right-handed supercoil [64]. Short proline-rich sequences also occur on many eukaryotic and prokaryotic proteins and these sequences also have the propensity to adopt the PPII form [65]. PPII helices have an extended form and a perfect threefold rotational symmetry that is seen when viewed down their central helical axis [65]. The AgI/II P-repeats have adopted the left-handed PPII form but within the AgI/II structure one finds only a single PPII chain rather than a set of three as occur in collagen.

What makes AgI/II an exciting new addition to these classical fibrillar forms is that it has paired the  $\alpha$ - or PPII fibrillar structures to build a novel  $\alpha$ - and PPII hybrid form. The AgI/II A<sub>3</sub>VP<sub>1</sub> is the first known fibrillar hybrid of  $\alpha$ - and PPII- helices to be structurally determined and reveals a novel solution for host adherence evolved within the gram positive

bacterial species. It is clear that the gram positive bacteria still possess many secrets yet to be shared for protein structure.

#### **4.1 C-terminal AgI/II structure**

The determination of the AgI/II carboxy-terminal structure ( $C_{1-3}$ ) has revealed that the carboxy-terminus is composed of three domains,  $C_1$ ,  $C_2$ , and  $C_3$ , of an immunoglobulin-like fold. Immunoglobulin folds are globular domains that have seven strands, “A” - “G”, which form two  $\beta$ -sheets. One  $\beta$ -sheet contains the “ABED” strands, while the other  $\beta$ -sheet has the “CFG” strands. In particular, AgI/II’s C-terminal domains have adopted the DE-variant Immunoglobulin-like fold (DEv-IgG fold). The DEv IgG-like fold was so named as it matches the architecture of classical immunoglobulin folds, but has additional structural variations present between the “D” and “E” strands [66]. These variations exist in each of the three  $C_{123}$  domains, with the  $C_2$  domain having the most additional secondary structure elements. DEv IgG-like folds have previously been found among many surface proteins of gram positive bacteria and are also often involved in protein:protein interactions [66, 67].

Within each of the domains, an isopeptide bond links lysine and asparagine residues of the “A” and “F” strands from the two opposing  $\beta$ -sheets. These isopeptide bonds join Lys1006-Asn1121 ( $C_1$ ), Lys1161-Asn1311 ( $C_2$ ), and Lys1338-Asn1473 ( $C_3$ ). Additionally, the C-terminus of AgI/II has three calcium ion sites located within the  $C_2$  and  $C_3$  domains. Altogether, the alignment of the  $C_{1-3}$  domains gives them an extended shape similar to the fibronectin-like domains [68]. This provides the AgI/II C-terminus with an overall length of  $\sim 13$  nanometers. However, as we see next, the length of the C-terminus is only a fraction of the extension that makes AgI/II such a highly elongated protein.

## 4.1 A fibrillar model of AgI/II

Full-length AgI/II has two additional repeats that are absent from A<sub>3</sub>VP<sub>1</sub>, namely the A<sub>1</sub>, A<sub>2</sub> and P<sub>2</sub>, P<sub>3</sub>. After resolving the structure of A<sub>3</sub>VP<sub>1</sub>, we hypothesized that the additional repeats would extend AgI/II to three times the length of the single AP repeat present in A<sub>3</sub>VP<sub>1</sub>. We confirmed this when we characterized the shapes of the AgI/II fragments while they were in solution by using analytical ultracentrifugation (Table 2.2, pg. 32). The velocity sedimentation showed highly elongated shapes for A<sub>3</sub>VP<sub>1</sub> (22.7 nm), A<sub>1</sub>VP<sub>3</sub> (59.3 nm), and AgI/II (67.8 nm).

Electron microscopic images of individual molecules of AgI/II (Fig 3.6, pg. 78) provided direct evidence for a fibrillar structure of AgI/II. Dimensions measured from these images, at ~65 nanometers length, were also very close to those determined above in analytical ultracentrifugation of the full length AgI/II. Together, these results now confirm that AgI/II is indeed an elongated fibrillar protein.

Having determined the structures of the N- and C-terminal fragments of AgI/II, the dimensions, and the overall shape of the AgI/II molecule we are now able to present a complete tertiary structure for AgI/II. This model positions the V-region distal to the streptococcal cell surface as distanced by the long A-P stalk and the three carboxy-terminal domains. This structural model explains the visual presence of the “fuzzy coat” (Fig 1.3, pg. 7) which extends over 50 nanometers from the cell surface of *S. mutans*, as recognized in electron microscopic imaging [16] almost a quarter of a century earlier.

## 4.2 AgI/II Adherence Function

As described in the introduction, AgI/II plays a major role in oral streptococcal adherence to the salivary pellicle. AgI/II specifically adheres to the large glycoprotein complex SAG, whose major component is the gp-340 polypeptide. We have utilized AgI/II fragments

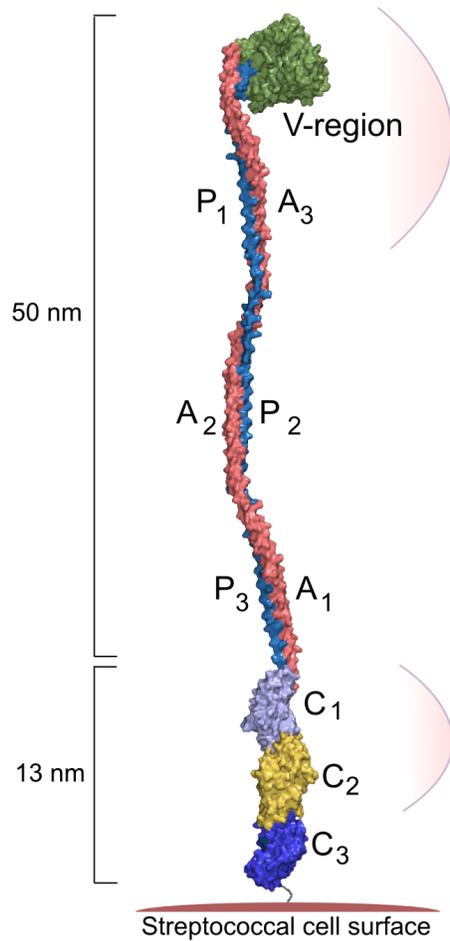


Figure 4.4: *The fibrillar model of AgI/II.* An overall tertiary model of AgI/II combines the information of the A<sub>3</sub>VP<sub>1</sub> and C<sub>13</sub> crystal structures with the biophysical characterizations and microscopy. The biophysical characterizations derive the length of AgI/II to be ~65 nm, which in the above model has the AP stalk contributing the majority of AgI/II's length (50 nm) and the 13 nm long C-terminus near the cell surface providing the remaining extension. AgI/II adherence to SAG minimal N- and C- terminal binding regions of A<sub>3</sub>VP<sub>1</sub> and C<sub>1-2</sub>, positioning adherence regions both proximal and distal to the *S. mutans* cell surface.

that were developed for structural studies to map AgI/II adherence properties with SAG. In contrast with earlier studies, the fragments of AgI/II used here were designed using the AgI/II structure. This has allowed us to test adherence using fragments containing distinct domain regions, without premature termination of sequences inside domain termini that might negatively affect the structure of the fragments. To confirm that this was the case, secondary structures were further characterized using circular dichroism and found to agree well with our determined crystal structures (Table 2.3, pg. 37).

Measuring adherence of AgI/II to SAG was primarily conducted using surface plasmon resonance (SPR). One advantage of the SPR technique is low requirements of protein, which is useful for studies involving SAG, as it is currently purified only in low amounts from human saliva. Through SPR, we have measured the kinetics of both association and dissociation rates for the adherence of regions of AgI/II to SAG. The adherence was studied under flow conditions that mimic the salivary flow of the oral environment. The SPR experiments were qualitatively supported by inhibition studies that measured that ability of the fragments to block adherence by *S. mutans* to SAG coated calcium hydroxyapatite, a widely used *in vitro* model (Table 3.3, pg. 76).

Many fragments of AgI/II showed adherence to SAG. These include the full-length AgI/II, A<sub>1</sub>VP<sub>3</sub>, A<sub>3</sub>VP<sub>1</sub>, V-region, C<sub>123</sub>, and C<sub>12</sub> (Fig 2.8, pg. 35 and Fig 3.5, pg. 74). Affinities to SAG were similar among the full-length AgI/II, A<sub>1</sub>VP<sub>3</sub>, and A<sub>3</sub>VP<sub>1</sub> which had affinity constants ( $K_A$ ) of  $\sim 3 * 10^7$ . The V-region had a lower affinity ( $K_A$ ) at  $3 * 10^5$ . After we determined the structure of the AgI/II C-terminus and found it contained three domains, we then could further delineate SAG adherence where we found that the C<sub>123</sub> had an affinity ( $K_A$ ) of  $2.4 * 10^6$  while the C<sub>12</sub>'s affinity was relatively higher at  $1.7 * 10^7$ . Negligible binding was seen for C<sub>23</sub> and individual C<sub>1</sub>, C<sub>2</sub>, and C<sub>3</sub> domains (Fig 3.5, pg. 74). In sum, the two minimal regions of A<sub>3</sub>VP<sub>1</sub> and C<sub>12</sub> are implicated in adhering to SAG and these are outlined in Fig. 4.4.

Having seen similar affinities between all the N-terminal regions ( $A_1VP_3$  or  $A_3VP_1$ ) versus full-length SAG, we tested the hypothesis that there may be only a single adherence site present on SAG for which regions of AgI/II could compete. We found that the AgI/II fragments consisting of the two major N- and C-terminal SAG-interacting regions adhered non-competitively (Fig 3.5, 74) with SAG. These results indicate that the AgI/II N- and C-terminal regions are interacting with unique sites on SAG. Inhibition studies using either of these fragments have also confirmed their activity. Either fragment of AgI/II partially blocks *S. mutans* adherence, but inhibition occurs to a lesser degree than seen with full-length AgI/II (Table 3.3, pg. 76).

Several possibilities for AgI/II adherence to SAG are now suggested (Fig 4.5). A single site of AgI/II may alone engage with SAG during adherence. In this case, the presence of multiple adherence sites on AgI/II may serve as a bacterial fail-safe, with the secondary site present to continue adherence in the possibility of the loss of the first. This may be necessary for AgI/II in proteolytically active environments, as the N-terminus of AgI/II (the AgI fragment) is highly proteolytically sensitive [18, 19].

It is also possible that AgI/II may make multiple contacts at both the N- and C-terminal sites to a single gp-340 polypeptide or to several gp-340 polypeptides of SAG. Little is still known about the overall structure of SAG as it is found on the tooth surface. The SAG complex has been predicted to have a very high molecular weight ( $5 \times 10^3$  kDa) [33], suggesting it may contain many polypeptide chains. The available electron microscopic imaging of SAG/gp340 have shown aggregates of elongated polypeptides that extend like tentacles [69]. As such, it is quite possible that the SAG complex may form a very large structure that is well matched to the elongation of AgI/II.

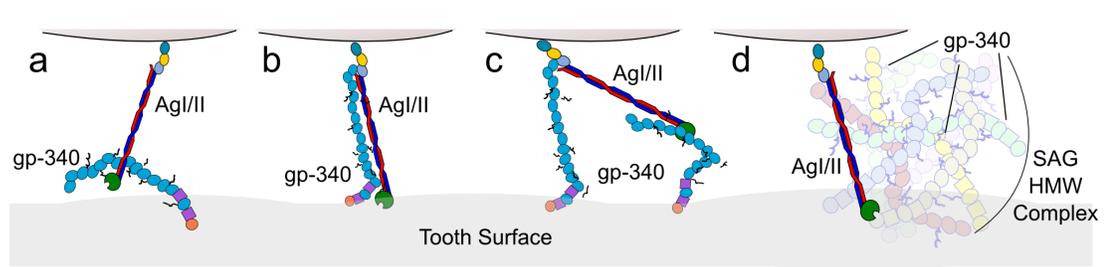


Figure 4.5: *Possible models of AgI/II adherence*- AgI/II adheres to the tooth surface through the SAG complex, whose main component is gp-340. AgI/II adherence to gp-340 may occur through a single contact with AgI/II (a), through multiple contacts of AgI/II to a single gp-340 (b), contacts of one AgI/II with multiple gp-340 proteins (c), or with AgI/II adhering to a very large SAG complex structure that contains multiple gp-340 proteins (d).

## 4.2 Ligand Binding Sites of AgI/II

Further reduction of the binding regions to specific sites of interaction may require another difficult endeavor, the determination of complex structures containing *S. mutans* AgI/II bound to fragments of SAG. However, while we await the solution of a complex structure, the N- and C-terminal structures contain specific interactions that could now be functionally examined.

It has long been suggested that the AgI/II family adherence with SAG might occur through SAG's extensive glycosylations [50]. Several carbohydrates can partially inhibit the binding of *S. mutans* with SAG [50, 28] but it is yet to be determined whether a carbohydrate is specifically recognized by AgI/II and if these interactions are required for AgI/II adherence with SAG. Small ligands are seen interacting with the AgI/II structure within the extents of both the N- and C- terminal regions that were highlighted by the adherence studies above, and the AgI/II residues involved might be tested to determine whether they represent possible adherence sites for SAG's glycosylations.

The first ligand site exists within the variable region. In *S. mutans* AgI/II this globular domain has a deep trench and was previously postulated to be a possible lectin-like fold

[27]. Inside the trench near the residue Trp816 A<sub>3</sub>VP<sub>1</sub>, we identified large electron densities that were well fitted by the protease inhibitor phenyl-methyl-sulfonate (PMS) that was present as an additive in crystallization [25]. No other additives explained the densities which contained both a ring-like segment and a tetrahedral shape that matches the molecule PMS (containing a phenyl group linked to a sulfate group). This molecule had hydrogen bonding between its sulfate group and the positively charged side-chain of Arg824, while its hydrophobic phenyl group stacked with a nearby Trp816 in a manner similar to lectin-sugar binding sites. The side-chain orientation of Trp816 differs from the original V-region structure [27] showing a displacement occurred with PMS binding (Fig 4.6).

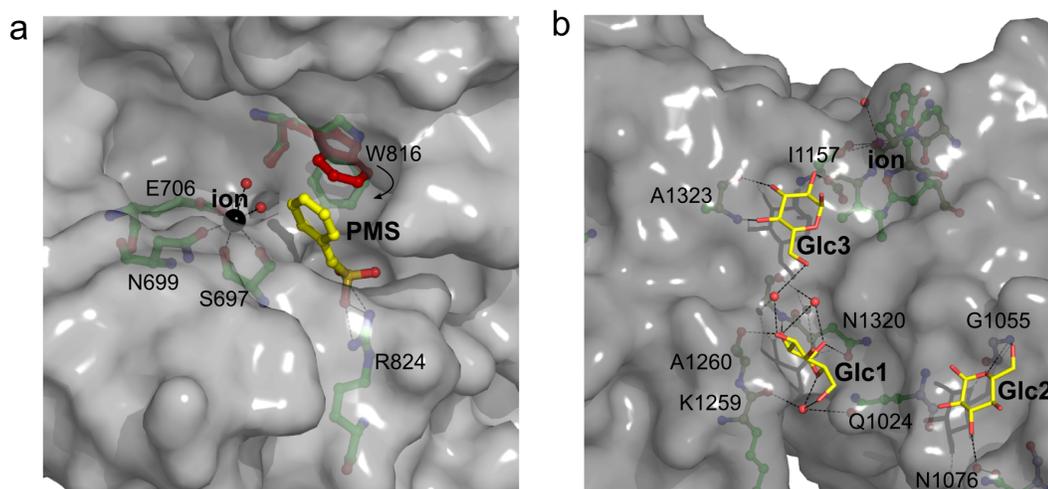


Figure 4.6: a) A cleft within the V-region of A<sub>3</sub>VP<sub>1</sub> contained a phenyl-methyl-sulfonate (PMS) at a putative sugar binding site. The Trp816 sidechain has moved from the original V-region structure's orientation (red) [27] to stack with the phenyl group, while the sulfate group of PMS is hydrogen bonding with Arg824. A calcium ion site nearby may stabilize this interaction.

b) At a small cleft between the C<sub>12</sub> domains of the C-terminal crystal structure are three molecules of glucose interacting with the AgI/II. These glucose molecules may represent a binding site for a complex sugar group. Glc1 forms several hydrogen bonds with the residues of the cleft and the C<sub>2</sub> DH2 helix. Another calcium ion site exists near Glc3.

The second set of ligands were found in the C-terminal structure near the C<sub>1</sub> and C<sub>2</sub>

domains. Five glucose molecules are present which originated from the high concentration glucose cryo-protectant solution (30% weight/volume), making C<sub>123</sub> the first AgI/II structure to show a direct interaction with carbohydrates. Three are located within a cleft between the C<sub>1</sub> and C<sub>2</sub> domains. The other two are located at crystal contacts between C<sub>123</sub> molecules, which may only exist when C<sub>123</sub> is organized in this crystalline lattice. At a minor cleft formed between C<sub>12</sub>, the three glucose molecules are all within 12 Å distance, suggesting they could represent an AgI/II binding site for a more complex glyco-chain.

The relatively few residues involved with either of these sites (listed in Fig 4.6) make them attractive targets for a mutational study. The unresolved question remains, what, if any specific molecule could these sites be designed to recognize? If they are carbohydrate binding sites, then these two sites may provide a clue towards the adherence property of AgI/II. The functional characterization of the AgI/II ligand binding sites may require the determination of a highly specific carbohydrate candidate which could be employed in functional assays. Recently, the mapping of glycosylations on SAG was completed through the chemical removal and mass spectroscopic analysis of the sugars. These glycosylations are predominantly O-linked and contain fucose, N-acetyl-glucosamine, and sialyllactosamine groups [70]. These complex sugars may make good targets in an initial study of sugar binding with AgI/II.

### **4.3 Other AgI/II Family Members**

The structural information determined for *S. mutans* AgI/II will also have implications for other known members of the AgI/II family. The sequence alignment of the family members (Fig 1.5, pg. 9) shows most AgI/II members share the arrangement of three AP- repeats, a variable region, and carboxy-terminus. These proteins can be predicted to

| Location         | Interaction              | Residue                                |
|------------------|--------------------------|--|
| V-region         | PMS site                 | Trp816 and Arg824                      |
| C <sub>1-2</sub> | Glucose 1 site           | Gln1024 and Asn1320                    |
| C <sub>1-2</sub> | Glucose 2 site           | Tyr1322                                |
| C <sub>1-2</sub> | Glucose 3 site           | Asn1076                                |
| V-region         | Calcium 1 site near PMS  | Ser697, Asn699, and Glu706             |
| C <sub>2</sub>   | Calcium 2 site near Glc3 | Asn1155, Asp1189, Asp1191, and Asp1192 |
| C <sub>2</sub>   | Calcium 3 site near Glc1 | Asp1212 and Glu1215                    |
| C <sub>3</sub>   | Calcium 4 site           | Asp1388 and Gln1391                    |

Table 4.1: Residues with sidechains involved in ligand and ion interactions of AgI/II

exist as elongated structures of similar dimension on their respective streptococcal species. However the *S. intermedius* Pas has only a single A/P repeat and would be predicted to exist as a shorter fibrillar protein (15-17 nm) with only a single full AP- repeat set extending Pas a third of the length of *S. mutans* AgI/II. The length of AgI/II may have implications for its adherence function. It has been noted that the *S. intermedius* strains of oral streptococci do not adhere to the salivary pellicle as strongly as *S. mutans* and the *S. intermedius* species colonize different surfaces of the tooth [30].

Antigen I/II is important for biofilm formation for both of these species [71]. Biofilms provide bacteria with advantages over the planktonic state: protection from competing species, protecting from host defense molecules, and environmental homeostasis within the biofilm. It has been observed that thick biofilms require fluid channels between cells that behave similarly to primitive circulatory systems through which nutrients, waste, and other materials can move [72]. As an elongated surface protein, AgI/II could serve a structural purpose to provide separation between cells within these oral biofilms. Electron micrographs showing multiple *S. mutans* cells [16] have gaps of similar dimension as the length of AgI/II proteins. With predicted differences of length for *S. intermedius* Pas and *S. mutans* AgI/II, the biofilm may offer another phenotypical difference related to AgI/II's structure.

## 4.4 Future of Adherence Inhibition

The model of AgI/II presented in this thesis has described both the overall structure of AgI/II and details at the atomic resolution that may provide leads to deciphering the mechanisms of AgI/II-mediated adherence to the salivary pellicle. As such, these structures may aid in the development of the next generation of anti-caries compounds. With the spreading of multiple resistance in bacteria against modern antibiotics, targeting bacterial adhesion for inhibition is a new and growing research topic, with the specific development of anti-caries therapeutics now being pursued globally.

AgI/II has been widely studied as a potential immunogen for passive or active immunization therapies directed against *S. mutans*. Early in these investigations, it was noted that when rabbits were hyper-immunized with *S. mutans* immunogens, cross-reactivity could be seen in heart tissues [73]. This raised concerns whether *S. mutans* could be used in human therapies without similar risk of immune-directed damage of the human heart. To a large extent, these worries have been resolved, and more recent studies have failed to show cross reactivity with streptococcal proteins and specifically indicated that AgI/II was not the cause of these interactions [74]. Studies are on-going using AgI/II fragments in the development of an anti-caries vaccine [75, 76, 77]. Future immunogenic studies may benefit from these structures determined for AgI/II that now reveal the tertiary structure of AgI/II and suggest regions useful for protective inhibition.

The development of small molecule inhibitors which specifically block bacterial adhesion is an alternative path to anti-caries therapeutics. The second most common bacterial infection after *S. mutans* occurs in stomach ulcers due to *Helicobacterium pylori* [10] and small molecule adhesion inhibitors have been studied to combat this pathogen. Sialic-acid containing oligosaccharides were effective inhibitors of *H. pylori* adhesion [78] and can cure or reduce *H. pylori* colonization in rhesus monkeys [79]. These results led to a recent

human clinical trial [80]. While the failure of this trial to cure *H. pylori* colonization in humans was a high profile setback for small molecule adhesion inhibitors, the attempts to find inhibitors for bacterial adhesion continue.

There have been some successes for *S. mutans*. Small peptides that interfere with mutans streptococcal adherence have been derived from the AgI/II amino acid sequence [47, 81]. One of these peptides, taken from a sequence of the AgI/II C-terminus has been successful in inhibiting mutans group colonization of the oral cavity in human trials [47]. These studies give hope that the further determination of the specificity for the AgI/II-SAG mechanisms could lead to cheap and easy-to-produce small molecule inhibitors for widespread use.

There is still much work to be completed on the path to inhibiting pathogenic bacterial adherence. It is likely that bacterial colonization will be even more complex as our understanding of it grows. We hope that the structures of AgI/II will help play a role in the future studies of bacterial adhesion.

## LIST OF REFERENCES

- [1] US Department of Health and Human Services. Oral health in America: a report of the Surgeon General. *J Calif Dent Assoc* **28**, 685–695 Sep (2000).
- [2] M. Balakrishnan, R. S. Simmonds, and J. R. Tagg. Dental caries is a preventable infectious disease. *Aust Dent J* **45**, 235–245 Dec (2000).
- [3] R. M. Love and H. F. Jenkinson. Invasion of dentinal tubules by oral bacteria. *Crit. Rev. Oral Biol. Med.* **13**, 171–183 (2002).
- [4] C. J. Whittaker, C. M. Klier, and P. E. Kolenbrander. Mechanisms of adhesion by oral bacteria. *Annu Rev Microbiol* **50**, 513–552 (1996).
- [5] W. J. Loesche. Role of *Streptococcus mutans* in human dental decay. *Microbiol. Rev.* **50**, 353–380 Dec (1986).
- [6] I. D. Mandel. Relation of saliva and plaque to caries. *J. Dent. Res.* **53**, 246–266 (1974).
- [7] K. Kasper, E. Braunwald, A. Fauci, S. Hauser, D. Longo, and J. Jameson. *Principles of Internal Medicine, 16th Edition*. McGraw-Hill. (2005).
- [8] H. Ton-That, L. A. Marraffini, and O. Schneewind. Protein sorting to the cell wall envelope of Gram-positive bacteria. *Biochim. Biophys. Acta* **1694**, 269–278 Nov (2004).

- [9] P. R. Smeesters, D. J. McMillan, and K. S. Sriprakash. The streptococcal M protein: a highly versatile molecule. *Trends Microbiol.* **18**, 275–282 Jun (2010).
- [10] C. Bavington and C. Page. Stopping bacterial adhesion: a novel approach to treating infections. *Respiration* **72**, 335–344 (2005).
- [11] K. Ohtsubo and J. D. Marth. Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855–867 Sep (2006).
- [12] J. M. Vose, P. W. Smith, M. Henry, and D. Colan. Recurrent *Streptococcus mutans* endocarditis. *Am. J. Med.* **82**, 630–632 Mar (1987).
- [13] R. Nomura, M. Hamada, K. Nakano, H. Nemoto, K. Fujimoto, and T. Ooshima. Repeated bacteraemia caused by *Streptococcus mutans* in a patient with Sjogren’s syndrome. *J. Med. Microbiol.* **56**, 988–992 Jul (2007).
- [14] P. W. Caufield. Tracking human migration patterns through the oral bacterial flora. *Clin. Microbiol. Infect.* **15 Suppl 1**, 37–39 Jan (2009).
- [15] D. Ajdic, W. M. McShan, R. E. McLaughlin, G. Savic, J. Chang, M. B. Carson, C. Primeaux, R. Tian, S. Kenton, H. Jia, S. Lin, Y. Qian, S. Li, H. Zhu, F. Najjar, H. Lai, J. White, B. A. Roe, and J. J. Ferretti. Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14434–14439 Oct (2002).
- [16] G. Y. Ayakawa, L. W. Boushell, P. J. Crowley, G. W. Erdos, W. P. McArthur, and A. S. Bleiweis. Isolation and characterization of monoclonal antibodies specific for antigen P1, a major surface protein of mutans streptococci. *Infect. Immun.* **55**, 2759–2767 Nov (1987).

- [17] V. A. Fischetti, V. Pancholi, and O. Schneewind. Conservation of a hexapeptide sequence in the anchor region of surface proteins from gram-positive cocci. *Mol. Microbiol.* **4**, 1603–1605 Sep (1990).
- [18] M. W. Russell and T. Lehner. Characterisation of antigens extracted from cells and culture fluids of *Streptococcus mutans* serotype c. *Arch. Oral Biol.* **23**, 7–15 (1978).
- [19] M. W. Russell, L. A. Bergmeier, E. D. Zanders, and T. Lehner. Protein antigens of *Streptococcus mutans*: purification and properties of a double antigen and its protease-resistant component. *Infect. Immun.* **28**, 486–493 May (1980).
- [20] C. Kelly, P. Evans, L. Bergmeier, S. F. Lee, A. Progulske-Fox, A. C. Harris, A. Aitken, A. S. Bleiweis, and T. Lehner. Sequence analysis of the cloned streptococcal cell surface antigen I/II. *FEBS Lett.* **258**, 127–132 Nov (1989).
- [21] C. Kelly, P. Evans, J. K. Ma, L. A. Bergmeier, W. Taylor, L. J. Brady, S. F. Lee, A. S. Bleiweis, and T. Lehner. Sequencing and characterization of the 185 kDa cell surface antigen of *Streptococcus mutans*. *Arch. Oral Biol.* **35 Suppl**, 33S–38S (1990).
- [22] A. S. Bleiweis, S. F. Lee, L. J. Brady, A. Progulske-Fox, and P. J. Crowley. Cloning and inactivation of the gene responsible for a major surface antigen on *Streptococcus mutans*. *Arch. Oral Biol.* **35 Suppl**, 15S–23S (1990).
- [23] S. Zhang, N. M. Green, I. Sitkiewicz, R. B. Lefebvre, and J. M. Musser. Identification and characterization of an antigen I/II family protein produced by group A *Streptococcus*. *Infect. Immun.* **74**, 4200–4213 Jul (2006).
- [24] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J.

- Gibson, and D. G. Higgins. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 Nov (2007).
- [25] M. R. Larson, K. R. Rajashankar, M. H. Patel, R. A. Robinette, P. J. Crowley, S. Michalek, L. J. Brady, and C. Deivanayagam. Elongated fibrillar structure of a streptococcal adhesin assembled by the high-affinity association of alpha- and PPII-helices. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5983–5988 Mar (2010).
- [26] L. J. Brady, S. E. Maddocks, M. R. Larson, N. Forsgren, K. Persson, C. C. Deivanayagam, and H. F. Jenkinson. The changing faces of *Streptococcus* antigen I/II polypeptide family adhesins. *Mol. Microbiol.* **77**, 276–286 Jul (2010).
- [27] N. Troffer-Charlier, J. Ogier, D. Moras, and J. Cavarelli. Crystal structure of the V-region of *Streptococcus mutans* antigen I/II at 2.4 Å resolution suggests a sugar preformed binding site. *J. Mol. Biol.* **318**, 179–188 Apr (2002).
- [28] G. Hajishengallis, T. Koga, and M. W. Russell. Affinity and specificity of the interactions between *Streptococcus mutans* antigen I/II and salivary components. *J. Dent. Res.* **73**, 1493–1502 Sep (1994).
- [29] W. W. Navarre and O. Schneewind. Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in gram-positive bacteria. *Mol. Microbiol.* **14**, 115–121 Oct (1994).
- [30] F. C. Petersen, S. Assev, H. C. van derMei, H. J. Busscher, and A. A. Scheie. Functional variation of the antigen I/II surface protein in *Streptococcus mutans* and *Streptococcus intermedius*. *Infect. Immun.* **70**, 249–256 Jan (2002).
- [31] R. G. Holt and H. J. Wynder. Localization on the cell surface of streptococcal protein antigen A. *FEMS Microbiol. Lett.* **118**, 175–179 May (1994).

- [32] S. F. Lee, A. Progulsk-Fox, G. W. Erdos, D. A. Piacentini, G. Y. Ayakawa, P. J. Crowley, and A. S. Bleiweis. Construction and characterization of isogenic mutants of *Streptococcus mutans* deficient in major surface protein antigen P1 (I/II). *Infect. Immun.* **57**, 3306–3313 Nov (1989).
- [33] T. Ericson and J. Rundegren. Characterization of a salivary agglutinin reacting with a serotype c strain of *Streptococcus mutans*. *Eur. J. Biochem.* **133**, 255–261 Jun (1983).
- [34] R. H. Metcalf and R. H. Deibel. Differential lytic response of enterococci associated with addition order of lysozyme and anions. *J. Bacteriol.* **99**, 674–680 Sep (1969).
- [35] W. F. Liljemark, C. G. Bloomquist, and J. C. Ofstehage. Aggregation and adherence of *Streptococcus sanguis*: role of human salivary immunoglobulin A. *Infect. Immun.* **26**, 1104–1110 Dec (1979).
- [36] J. Rundegren and T. Ericson. Saliva-induced aggregation of micro-organisms from skin, tooth surfaces, oral mucosa, and rectum. *J. Oral Pathol.* **10**, 248–260 Aug (1981).
- [37] K. S. Ambatipudi, F. K. Hagen, C. M. Delahunty, X. Han, R. Shafi, J. Hryhorenko, S. Gregoire, R. E. Marquis, J. E. Melvin, H. Koo, and J. R. Yates. Human Common Salivary Protein 1 (CSP-1) Promotes Binding of *Streptococcus mutans* to Experimental Salivary Pellicle and Glucans Formed on Hydroxyapatite Surface. *J. Proteome Res.* **9**, 6605–6614 Dec (2010).
- [38] A. Prakobphol, F. Xu, V. M. Hoang, T. Larsson, J. Bergstrom, I. Johansson, L. Frangsmyr, U. Holmskov, H. Leffler, C. Nilsson, T. Boren, J. R. Wright, N. Stromberg, and S. J. Fisher. Salivary agglutinin, which binds *Streptococcus mutans* and *Helicobacter pylori*, is the lung scavenger receptor cysteine-rich protein gp-340. *J. Biol. Chem.* **275**, 39860–39866 Dec (2000).

- [39] L. J. Brady, D. A. Piacentini, P. J. Crowley, P. C. Oyston, and A. S. Bleiweis. Differentiation of salivary agglutinin-mediated adherence and aggregation of mutans streptococci by use of monoclonal antibodies against the major surface adhesin P1. *Infect. Immun.* **60**, 1008–1017 Mar (1992).
- [40] P. J. Crowley, L. J. Brady, D. A. Piacentini, and A. S. Bleiweis. Identification of a salivary agglutinin-binding domain within cell surface adhesin P1 of *Streptococcus mutans*. *Infect. Immun.* **61**, 1547–1552 Apr (1993).
- [41] G. H. Munro, P. Evans, S. Todryk, P. Buckett, C. G. Kelly, and T. Lehner. A protein fragment of streptococcal cell surface antigen I/II which prevents adhesion of *Streptococcus mutans*. *Infect. Immun.* **61**, 4590–4598 Nov (1993).
- [42] T. Oho, H. Yu, Y. Yamashita, and T. Koga. Binding of salivary glycoprotein-secretory immunoglobulin A complex to the surface protein antigen of *Streptococcus mutans*. *Infect. Immun.* **66**, 115–121 Jan (1998).
- [43] V. Loimaranta, N. S. Jakubovics, J. Hytonen, J. Finne, H. F. Jenkinson, and N. Stromberg. Fluid- or surface-phase human salivary scavenger protein gp340 exposes different bacterial recognition properties. *Infect. Immun.* **73**, 2245–2252 Apr (2005).
- [44] M. W. Oli, W. P. McArthur, and L. J. Brady. A whole cell BIAcore assay to evaluate P1-mediated adherence of *Streptococcus mutans* to human salivary agglutinin and inhibition by specific antibodies. *J. Microbiol. Methods* **65**, 503–511 Jun (2006).
- [45] A. Moisset, N. Schatz, Y. Lepoivre, S. Amadio, D. Wachsmann, M. Scholler, and J. P. Klein. Conservation of salivary glycoprotein-interacting and human immunoglobulin G-cross-reactive domains of antigen I/II in oral streptococci. *Infect. Immun.* **62**, 184–193 Jan (1994).

- [46] Y. Duan, E. Fisher, D. Malamud, E. Golub, and D. R. Demuth. Calcium-binding properties of SSP-5, the *Streptococcus gordonii* M5 receptor for salivary agglutinin. *Infect. Immun.* **62**, 5220–5226 Dec (1994).
- [47] C. G. Kelly, J. S. Younson, B. Y. Hikmat, S. M. Todryk, M. Czisch, P. I. Haris, I. R. Flindall, C. Newby, A. I. Mallet, J. K. Ma, and T. Lehner. A synthetic peptide adhesion epitope as a novel antimicrobial agent. *Nat. Biotechnol.* **17**, 42–47 Jan (1999).
- [48] F. J. Bikker, A. J. Ligtenberg, K. Nazmi, E. C. Veerman, W. van'tHof, J. G. Bolscher, A. Poustka, A. V. Nieuw Amerongen, and J. Mollenhauer. Identification of the bacteria-binding peptide domain on salivary agglutinin (gp-340/DMBT1), a member of the scavenger receptor cysteine-rich superfamily. *J. Biol. Chem.* **277**, 32109–32115 Aug (2002).
- [49] J. T. Leito, A. J. Ligtenberg, K. Nazmi, J. M. deBlieck-Hogervorst, E. C. Veerman, and A. V. Nieuw Amerongen. A common binding motif for various bacteria of the bacteria-binding peptide SRCRP2 of DMBT1/gp-340/salivary agglutinin. *Biol. Chem.* **389**, 1193–1200 Sep (2008).
- [50] D. R. Demuth, E. E. Golub, and D. Malamud. Streptococcal-host interactions. Structural and functional analysis of a *Streptococcus sanguis* receptor for a human salivary glycoprotein. *J. Biol. Chem.* **265**, 7120–7126 May (1990).
- [51] H. Koba, K. Okuda, H. Watanabe, J. Tagami, and H. Senpuku. Role of lysine in interaction between surface protein peptides of *Streptococcus gordonii* and agglutinin peptide. *Oral Microbiol. Immunol.* **24**, 162–169 Apr (2009).
- [52] M. A. Sciotti, I. Yamodo, J. P. Klein, and J. A. Ogier. The N-terminal half part of the oral streptococcal antigen I/II<sub>f</sub> contains two distinct binding domains. *FEMS Microbiol. Lett.* **153**, 439–445 Aug (1997).

- [53] C. A. Daep, R. J. Lamont, and D. R. Demuth. Interaction of *Porphyromonas gingivalis* with oral streptococci requires a motif that resembles the eukaryotic nuclear receptor box protein-protein interaction domain. *Infect. Immun.* **76**, 3273–3280 Jul (2008).
- [54] C. J. vanDolleweerd, C. G. Kelly, D. Chargelegue, and J. K. Ma. Peptide mapping of a novel discontinuous epitope of the major surface adhesin from *Streptococcus mutans*. *J. Biol. Chem.* **279**, 22198–22203 May (2004).
- [55] R. M. Love, M. D. McMillan, Y. Park, and H. F. Jenkinson. Coinvasion of dentinal tubules by *Porphyromonas gingivalis* and *Streptococcus gordonii* depends upon binding specificity of streptococcal antigen I/II adhesin. *Infect. Immun.* **68**, 1359–1365 Mar (2000).
- [56] A. R. Holmes, C. Gilbert, J. M. Wells, and H. F. Jenkinson. Binding properties of *Streptococcus gordonii* SspA and SspB (antigen I/II family) polypeptides expressed on the cell surface of *Lactococcus lactis* MG1363. *Infect. Immun.* **66**, 4633–4639 Oct (1998).
- [57] L. Kelemen, S. Rizk, M. Debreczeny, J. Ogier, and B. Szalontai. Streptococcal antigen I/II binds to extracellular proteins through intermolecular beta-sheets. *FEBS Lett.* **566**, 190–194 May (2004).
- [58] M. Matsumoto-Nakano, M. Tsuji, S. Inagaki, K. Fujita, K. Nagayama, R. Nomura, and T. Ooshima. Contribution of cell surface protein antigen c of *Streptococcus mutans* to platelet aggregation. *Oral Microbiol. Immunol.* **24**, 427–430 Oct (2009).
- [59] T. B. Seifert, A. S. Bleiweis, and L. J. Brady. Contribution of the alanine-rich region of *Streptococcus mutans* P1 to antigenicity, surface expression, and interaction with the proline-rich repeat domain. *Infect. Immun.* **72**, 4699–4706 Aug (2004).

- [60] J. K. Ma and T. Lehner. Prevention of colonization of *Streptococcus mutans* by topical application of monoclonal antibodies in human subjects. *Arch. Oral Biol.* **35 Suppl**, 115S–122S (1990).
- [61] M. Mueller, S. Jenni, and N. Ban. Strategies for crystallization and structure determination of very large macromolecular assemblies. *Curr. Opin. Struct. Biol.* **17**, 572–579 Oct (2007).
- [62] L. PAULING and R. B. COREY. Compound helical configurations of polypeptide chains: structure of proteins of the alpha-keratin type. *Nature* **171**, 59–61 Jan (1953).
- [63] G. N. Phillips, P. F. Flicker, C. Cohen, B. N. Manjula, and V. A. Fischetti. Streptococcal M protein: alpha-helical coiled-coil structure and arrangement on the cell surface. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4689–4693 Aug (1981).
- [64] K. E. Kadler, C. Baldock, J. Bella, and R. P. Boot-Handford. Collagens at a glance. *J. Cell. Sci.* **120**, 1955–1958 Jun (2007).
- [65] B. J. Stapley and T. P. Creamer. A survey of left-handed polyproline II helices. *Protein Sci.* **8**, 587–595 Mar (1999).
- [66] C. C. Deivanayagam, E. R. Wann, W. Chen, M. Carson, K. R. Rajashankar, M. Hook, and S. V. Narayana. A novel variant of the immunoglobulin fold in surface adhesins of *Staphylococcus aureus*: crystal structure of the fibrinogen-binding MSCRAMM, clumping factor A. *EMBO J.* **21**, 6660–6672 Dec (2002).
- [67] J. Symersky, J. M. Patti, M. Carson, K. House-Pompeo, M. Teale, D. Moore, L. Jin, A. Schneider, L. J. DeLucas, M. Hook, and S. V. Narayana. Structure of the collagen-binding domain from a *Staphylococcus aureus* adhesin. *Nat. Struct. Biol.* **4**, 833–838 Oct (1997).

- [68] D. J. Leahy. Implications of atomic-resolution structures for cell adhesion. *Annu. Rev. Cell Dev. Biol.* **13**, 363–393 (1997).
- [69] J. Madsen, J. Mollenhauer, and U. Holmskov. Review: Gp-340/DMBT1 in mucosal innate immunity. *Innate Immun* **16**, 160–167 (2010).
- [70] S. Issa, A. P. Moran, S. N. Ustinov, J. H. Lin, A. J. Ligtenberg, and N. G. Karlsson. O-linked oligosaccharides from salivary agglutinin: *Helicobacter pylori* binding sialyl-Lewis x and Lewis b are terminating moieties on hyperfucosylated oligo-N-acetyllactosamine. *Glycobiology* **20**, 1046–1057 Aug (2010).
- [71] D. Pecharki, F. C. Petersen, S. Assev, and A. A. Scheie. Involvement of antigen I/II surface proteins in *Streptococcus mutans* and *Streptococcus intermedius* biofilm formation. *Oral Microbiol. Immunol.* **20**, 366–371 Dec (2005).
- [72] S. S. Socransky and A. D. Haffajee. Dental biofilms: difficult therapeutic targets. *Periodontol. 2000* **28**, 12–55 (2002).
- [73] I. Van deRijn, A. S. Bleiweis, and J. B. Zabriskie. Antigens in *Streptococcus mutans* cross reactive with human heart muscle. *J. Dent. Res.* **55 Spec No**, 59–64 Apr (1976).
- [74] F. J. Swartzwelder, P. K. Barua, B. Albin, and M. W. Stinson. Heart-reactive antibodies in rabbit anti-*Streptococcus mutans* sera fail to cross-react with *Streptococcus mutans*. *J. Immunol.* **140**, 954–961 Feb (1988).
- [75] P. Zhang, C. Jespersgaard, L. Lamberty-Mallory, J. Katz, Y. Huang, G. Hajishengalis, and S. M. Michalek. Enhanced immunogenicity of a genetic chimeric protein consisting of two virulence antigens of *Streptococcus mutans* and protection against infection. *Infect. Immun.* **70**, 6779–6787 Dec (2002).

- [76] Y. Tsuha, N. Hanada, T. Asano, T. Abei, S. Yamaguchi, M. A. Salam, R. Nakao, H. Takeuchi, N. Kurosaki, and H. Senpuku. Role of peptide antigen for induction of inhibitory antibodies to *Streptococcus mutans* in human oral cavity. *Clin. Exp. Immunol.* **137**, 393–401 Aug (2004).
- [77] P. M. Drake, T. Barbi, A. Sexton, E. McGowan, J. Stadlmann, C. Navarre, M. J. Paul, and J. K. Ma. Development of rhizosecretion as a production system for recombinant proteins from hydroponic cultivated tobacco. *FASEB J.* **23**, 3581–3589 Oct (2009).
- [78] P. M. Simon, P. L. Goode, A. Mobasseri, and D. Zopf. Inhibition of *Helicobacter pylori* binding to gastrointestinal epithelial cells by sialic acid-containing oligosaccharides. *Infect. Immun.* **65**, 750–757 Feb (1997).
- [79] J. V. Mysore, T. Wigginton, P. M. Simon, D. Zopf, L. M. Heman-Ackah, and A. Dubois. Treatment of *Helicobacter pylori* infection in rhesus monkeys using a novel antiadhesion compound. *Gastroenterology* **117**, 1316–1325 Dec (1999).
- [80] F. Parente, C. Cucino, A. Anderloni, G. Grandinetti, and G. Bianchi Porro. Treatment of *Helicobacter pylori* infection using a novel antiadhesion compound (3'sialyllactose sodium salt). A double blind, placebo-controlled clinical study. *Helicobacter* **8**, 252–256 Aug (2003).
- [81] K. Okuda, N. Hanada, Y. Usui, H. Takeuchi, H. Koba, R. Nakao, H. Watanabe, and H. Senpuku. Inhibition of *Streptococcus mutans* adherence and biofilm formation using analogues of the SspB peptide. *Arch. Oral Biol.* **55**, 754–762 Oct (2010).

## **APPENDIX**

### **ALIGNMENTDRAW: SEQUENCE ALIGNMENT TOOL**

#### **Introduction**

The following is a PERL language script which can process the text output of ClustalX [24] and render this output into a visual graph of sequence similarity. This script addresses the problem of how to visually represent total gene sequence alignments in publications. It is common to find entire gene sequences listed out in text or presentation slides. Providing such an overload of information for the reader or viewer rarely conveys the meanings meant by the authors. Instead, they typically put the reader to sleep! This script was written to produce more exciting output that is kinder on the eyes.

An example of the output of this program is given in (Fig 1.5, shown in the introduction of this thesis). Some formatting is required after the output is produced, namely adding labels, titles, and other relevant markings to the final product.

This script is under GNU public domain, feel free to use and modify as you want. For publications, please reference the original publication that it was used in: Larson *et al* (2010).

The requirements for using this software are: PERL compiler (available from <http://www.perl.org>) and the Perl::GD package (available at <http://search.cpan.org/~lds/GD-2.11/GD.pm>). The ClustalX package needed to generate the alignments input file is downloadable at <http://www.clustal.org/>. (I used version

2.0.12)

## Instructions

1. Open **clustalx** to begin creating your alignment.
2. Use “File-Load sequences” to open your file containing all gene sequences for alignment.
3. Use “Alignment-Do complete alignment” to create an alignment file. This text output should have an extension of “.aln”
4. Now run “*perl AlignmentDraw.pl alignment.aln numberseq master output.png*” on the command line. The command options “**alignment.aln**” is your alignment filename from previous step, “**numberseq**” is the number of genes that you are aligning, “**master**” is the number of the sequence which represents the template gene against which you are comparing similarity (first sequence is 0, second is 1, ..), and “**output.png**” is the filename of the graphics file output to be created.
5. Open the graphics file in your favorite figure software (I recommend Inkscape), and begin adding labels or other information to get ready for publication.

## PERL code

### AlignmentDraw.pl

```
#!/usr/bin/perl

# AlignmentDraw: Matt Larson 2009 for GNU Public License use.
# Requires Perl::GD package installed for use.
# Takes .aln files generated by Clustalx2 as input.
$usage =
"\n\nAlignmentDraw: a program to create PNG output files summarizing the
```

```

_output_of_clustalw_sequence_alignments.

Usage: ./AlignmentDraw_alignment.file_number_of_sequences
      master_sequence_output.png \n
      \n\n";

$numargs = @ARGV;

use GD;

# getclass( character $input )
# returns character $type
#
# Classifies amino acids by group for similarity
sub getclass {
    # Classify the residues by type
    my $c = $_[0];
    my $type = "";
    if ($c eq "A" || $c eq "C" || $c eq "I" || $c eq "M" ||
        $c eq "P" || $c eq "V" || $c eq "L" || $c eq "G") {
        $type = "ShortHydrophobic";
    } elsif ($c eq "D" || $c eq "E") {
        $type = "Negative";
    } elsif ($c eq "F" || $c eq "W" || $c eq "Y" || $c eq "H") {
        $type = "BulkyHydrophobic";
    } elsif ($c eq "N" || $c eq "Q" || $c eq "S" || $c eq "T"
        || $c eq "R" || $c eq "K") {
        $type = "PosOrUncharged";
    }
    return ($type);
}

my $input_file = "";
my $output_file = "default.png";
my $number_of_sequences = 2;

# strict on command line arguments
if ($numargs != 4) {
    print $usage;
    exit 0;
}

$input_file = $ARGV[0];
$number_of_sequences = $ARGV[1];
$master_sequence = $ARGV[2];
$output_file = $ARGV[3];

my @array1;
for ($i = 0; $i < $number_of_sequences; $i++) {
    $array1[$i] = "";
}

if ($input_file eq "") {
    print $usage;
    exit 0;
}

# Read in the files
#print "The input file is $input_file\n";
open (INFILE, "<$input_file");
$line_number = 0;

# Read in the *.aln file, merging separate text lines into individual strings.

```

```

while ($line = <INFILE>) {
  if ($line_number < 3) {
    # Header lines, do nothing
  } else {
    @words = split(/\s+/, $line);
    # now words[1] is my sequence.

    # current sequence is?
    $current_sequence = ($line_number - 2) % ($number_of_sequences+2);

    if ($current_sequence <= $number_of_sequences) {
      $array1[$current_sequence] .= $words[1];
      #print "Test $current_sequence : $array1[$current_sequence]\n";
    }
  }
  $line_number += 1;
}

# Determine the total length of characters of the alignment
$characters = length($array1[1]);

# Create the Image canvas with 10 pixel space on top, bottom, sides
# Each alignment will be a bar about 20 pixels.
$im = new GD::Image($characters + 40, $number_of_sequences * 40);

# Setup the colors
$white = $im->colorAllocate(255,255,255);
$blue5 = $im->colorAllocate(128, 128, 255);
$blue10 = $im->colorAllocate(0, 0, 255);
$green10 = $im->colorAllocate(0, 255, 0);
$red10 = $im->colorAllocate(255,0,0);
$black = $im->colorAllocate(0,0,0);

my $master_str = $array1[$master_sequence];

# Draw markers every 100 alignment positions
for ($j = 0; $j < $characters; $j++) {
  if (($j % 100) == 1) {
    $im->filledRectangle($x+$j,0,$x+$j,5,$black);
  }
}

# Draw the alignments
for ($i = 1; $i <= $number_of_sequences; $i++) {
  # Upper coordinates of the sequences;
  my $x = 20;
  my $y = ($i - 1) * 40 + 10;

  my $str = $array1[$i];

  for ($j = 0; $j < $characters; $j++) {

    $color = $white;
    my $c = substr($str,$j,1);
    my $c_type = 'N';
    my $m = substr($master_str, $j,1);
    my $m_type = 'N';

    # Classify the residues by type
    $c_type = getclass($c);
    $m_type = getclass($m);

```

```

if ($c eq "-") {
    $color = $white;
} elseif ($c eq $m) {
    $color = $blue10;
    #print "$c matches $m\n";
} elseif ($c_type eq $m_type) {
    $color = $green10;
} else {
    $color = $red10;
}

# Draw stripes of colors
$im-> filledRectangle($x+$j, $y, $x+$j, $y+20, $color);

# Draw a line for a gap
if ($color eq $white) {
    $im-> filledRectangle($x+$j, $y+10, $x+$j, $y+10, $black);
}
}

# Write out the PNG file
open (MYFILE, ">>$output_file");
print MYFILE $im->png;
close MYFILE;

```