I, Stefanie L. Johns, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Molecular Genetics, Biochemistry, & Microbiology.

It is entitled:
Mechanistic insights for protein-dependent biofilm formation in Staphylococcus epidermidis and beyond

Student’s name: Stefanie L. Johns

This work and its defense approved by:

Committee chair: Andrew Herr, PhD
Committee member: Jay Degen, PhD
Committee member: Rhett Kovall, PhD
Committee member: Thomas Thompson, PhD
Committee member: Alison Weiss, PhD
Mechanistic insights for protein-dependent biofilm formation in *Staphylococcus epidermidis* and beyond

A dissertation submitted to the Division of Graduate Studies and Research of the University of Cincinnati

In partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY (Ph.D.)

In the Department of Molecular Genetics, Biochemistry, Microbiology, and Immunology of the College of Medicine

December 2011

**Stefanie L. Johns**

B.S. Wright State University, 2007

Committee Chair: Andrew B. Herr, Ph.D
Abstract

Eighty percent of all microbial infections are associated with the formation of a biofilm, a highly stable aggregated community. Staphylococcal species in particular utilize biofilm formation to establish chronic, antibiotic-resistant infections that are primarily associated with medical devices, artificial implants, and catheters. *Staphylococcus epidermidis* is the most common colonizer of epithelial and mucosal surfaces of the human body and is an important component of the normal flora; however, more recently this species has also emerged as one of the leading causes of opportunistic and hospital-acquired infections world-wide. *S. epidermidis* establishes a biofilm using a zinc-mediated, protein-dependent mechanism; the protein responsible for this mechanism is the Accumulation-associated protein (Aap). Aap and other protein homologues found in staphylococci contain an unusual protein domain organization that consists of five or more almost identical G5 domain repeats. The studies in this thesis were focused on elucidating the function of multiple, tandem G5 domain repeats within Aap. The results have implications for understanding i) the conservation of G5 domain function in diverse proteins, ii) the relationship of G5 domain assembly to staphylococcal biofilm formation, and iii) the development of therapeutic treatments for the prevention and dispersion of staphylococcal biofilms.

Protein binding studies of variable length G5 domain repeats (3, 5, 7, and 9) from Aap in the presence of zinc were used to determine the function of multiple, tandem G5 domains within a single protein. Our studies demonstrated an inverse correlation between the concentration of zinc required for association with respect to the number of repeats in tandem; as the number of G5 domain repeats increased, the concentration of required zinc decreased. These studies also demonstrated that extending the number of tandem G5 domain repeats increases the
potential to form large, elongated protein complexes. Further characterization of these elongated complexes revealed that Aap forms amyloid fibers during the zinc-mediated self-assembly mechanism. Amyloid fiber formation by Aap was determined to be dependent on incubation time and temperature, producing at least two morphologically unique fibers. Additional protein interaction studies using a multi-G5 domain protein from *Streptococcus sanguinis* suggest that G5 domains are universally conserved as metal-binding self-adhesion modules.

Based on these results, we identified critical environmental factors that influence both Aap amyloidogenesis and biofilm formation by *S. epidermidis*. Acidification and metal chelation were shown to inhibit biofilm formation and Aap self-assembly during early phases. However, mature Aap amyloid fibers and *S. epidermidis* biofilms were found to be resistant to metal chelation. These studies will help generate future therapeutic treatments for chronic infections caused by *Staphylococcus epidermidis*. 
Acknowledgements

This work never could have been completed without the help, support, and encouragement of so many people and I would like to express my gratitude for all of their assistance during graduate school. First, I would like to thank my advisor Dr. Andrew Herr. I have never had to “prove myself” to Andy, he has always just trusted my judgment and believed in my research skills without doubt. He has developed the scientist I am today, and I would never have the confidence in myself that I have now. Thank you Andy for all of your support throughout my graduate career.

I would also like to thank my committee members, Dr. Jay Degen, Dr. Rhett Kovall, Dr. Tom Thompson, and Dr. Alison Weiss. All of you have helped mentor me through graduate school and served as outstanding role models as scientists for me, in addition to all of your helpful suggestions during committee meetings. Special thanks to Rhett and Tom who have spent countless hours of their time helping me with my crystallography research project.

Thanks to my fellow Herr lab members Jen, Monica, Deb, and Catie (and surrogate lab member Jeannette!), and former member Jeff Wilson. It has been the best of times and we have all been the crutch to support each other on the worst of days. Special thanks to former lab member Michelle Gomes, who has guided me through the lab from day one; you are a true friend and mentor to me.

I would also like to thank Dr. Peter Schuck at NIH, your brilliance is a true inspiration to any scientist and I am still shocked that I have had the privilege to work with you.
I have had the opportunity to be mentored by so many great scientists during my graduate career, however, there is no chance I would have survived this journey without the love and support of my friends and family.

To Mary Jane – research started our friendship and made us sisters from day one (Missouri loves company). Thanks for being an “ear” to cry on for the past 7 years.

To Chris and Taryn – thanks for keeping my feet on the ground and a safe haven for me to unwind. You both have always kept me “care-free” even in my darkest hour. Also to Patrick, Julia, and Christi – thanks for keeping me laughing and accepting my nerdy-ness.

To my grandparents, Bill and Janet Kay and Paul and JoAnn Ward. You all have been such a great support in my education in so many ways, and have always made me feel like a star.

A big thanks to my parents, Joe and Debbie Ward. I think it would be an understatement if I said, “it has been a ride.” Thanks for always supporting my aspirations, dreams, and goals no matter how big or unrealistic they might be. You both have shaped my creativity and intellect through my life and made me the strong person I am today. Also thanks to my new set of parents, Steve and Maureen Johns, whom I have shared so many great conversations with sitting at your kitchen table. Thanks for all of your love, support, and acceptance in your family.

And last but most definitely not least, to my husband Dennis. You are the highlight of my graduate career, my biggest cheerleader in life, and my best friend. You have been there every time I came home crying and defeated to pick me up and give me the strength to persevere. You are truly my better half and give me the courage to conquer all obstacles in life. And of course thanks to our “furr-kids” Buddy and Xander who have spent everyday of writing this manuscript by my side.
Table of Contents

Abstract .......................................................................................................................... ii

Acknowledgements ....................................................................................................... v

Table of Contents .......................................................................................................... vii

Figures ......................................................................................................................... x

Tables ........................................................................................................................... xiii

Abbreviations ............................................................................................................. xiv

Symbols ....................................................................................................................... xvii

Chapter 1. *Staphylococcus epidermidis*, the opportunistic pathogen: A literature review... 1

Introduction .................................................................................................................. 2

*Staphylococcus epidermidis* Bacterial Profile ............................................................... 3

Stages of Biofilm Formation .......................................................................................... 7

Communication within the Biofilm .................................................................................. 19

Genetic Regulation of Biofilm Formation ........................................................................ 21

Clinical Insights to *S. epidermidis* Infection and Treatment .......................................... 23

Scope and Thesis Outline .............................................................................................. 25

Chapter 2. Functional Analysis of Proteins Containing Multiple, Tandem G5 Domains ... 26

Abstract ....................................................................................................................... 27

Introduction ................................................................................................................... 28

Materials and Methods ................................................................................................. 32
Results.................................................................................................................................35
Discussion............................................................................................................................46

Chapter 3. A Literature Review of Functional Amyloid and Amyloidogenesis..............49
Introduction............................................................................................................................50
Overview of Amyloid Characteristics .................................................................51
Functional Amyloid Proteins.........................................................................................60

Chapter 4. The Biofilm Adhesion Protein Aap from *Staphylococcus epidermidis* Forms Zinc-Dependent Amyloid Fibers.................................................................63
Abstract............................................................................................................................64
Introduction............................................................................................................................65
Materials and Methods.........................................................................................69
Results...............................................................................................................................74
Discussion.........................................................................................................................93

Chapter 5. Connecting the Mechanistic Relationship of Aap Self-Assembly and Biofilm Formation by *Staphylococcus epidermidis*........................................100
Abstract............................................................................................................................101
Introduction............................................................................................................................102
Materials and Methods.........................................................................................104
Results...............................................................................................................................110
Discussion.........................................................................................................................125

Chapter 6. Summary and Future Directions.............................................................129
Summary............................................................................................................................130
Future Directions.................................................................135
Concluding Remarks.............................................................142

Appendix I. Structural Analyses Suggest New Roles for Calcium in the Enzymatic Mechanism and Dimerization of Human Calcium-Activated Nucleotidases........144

Abstract.........................................................................................145
Introduction......................................................................................146
Materials and Methods.................................................................149
Results..........................................................................................154
Discussion.......................................................................................179

Bibliography....................................................................................186
Figures

Chapter 1.

Figure 1. Micrograph of a *S. epidermidis* biofilm……………………………………………….6

Figure 2. Model of *S. epidermidis* biofilm formation……………………………………………….8

Figure 3. Protein domain organization of the Accumulation-associated protein from *S. epidermidis* RP62a……………………………………………………………………………….13

Chapter 2.

Figure 1. Sequence alignment of G5 domains with diverse protein origins……………….28

Figure 2. Species diversity of bacteria expressing G5 domains…………………………………29

Figure 3. Comparison of protein domain architectures of Aap and zmpC………………………35

Figure 4. Solution characterization of recombinantly expressed G5 domain proteins………37

Figure 5. Brpt zinc-mediated self-assembly is dependent on the number of tandem Brpts and zinc concentration…………………………………………………………………………………….38

Figure 6. StsG5 self-assembles in the presence of zinc…………………………………………39

Chapter 3.

Figure 1. Model of amyloid “cross-β” structure……………………………………………….51

Figure 2. Amyloid fiber morphologies………………………………………………………………53

Figure 3. Kinetic diagram of nucleation-dependent amyloid fiber assembly…………………55

Figure 4. Conserved fiber diffraction of amyloid fibers…………………………………………58

Figure 5. Variability of packing in the amyloid fiber core………………………………………59
Chapter 4.

Figure 1. Characterization of the Brpt5.5 construct from *S. epidermidis* Aap…………………66, 75

Figure 2. Sedimentation behavior of Brpt5.5 in the presence of Zn2+…………………………77

Figure 3. Amyloid properties of Brpt5.5 in the presence of Zn2+………………………………79

Figure 4. HPLC quantification of Zn2+-mediated Brpt5.5 oligomer and fibril species………82

Figure 5. Temperature and time-course analysis of Zn-mediated Brpt5.5 fiber assembly………86

Figure 6. AUC c(s, ff0) analysis of early-stage Brpt5.5 amyloidogenic intermediates…………90

Figure 7. Amyloid fibers are early structural components in *S. epidermidis* biofilms………………92

Figure 8. Model for Zn2+-induced amyloidogenesis by the B-repeat region of Aap……………..98

Chapter 5.

Figure 1. The requirements of Zn and Cu to support *S. epidermidis* strain RP62a biofilm formation………………………………………………………………..………………111

Figure 2. The influence of pH on *S. epidermidis* biofilm formation…………………………113

Figure 3. Inhibition of *S. epidermidis* Biofilm formation by metal chelators…………………116

Figure 4. Treatment with DTPA before or after complete *S. epidermidis* biofilm formation…118

Figure 5. Brpt fibers are resistant to acid and metal chelator treatment………………………120

Figure 6. Mature Brpt fibers maintain fiber morphology after acid and metal chelator treatments…………………………………………………………………………………121

Figure 7. Sedimentation velocity analysis of Brpt3.5 in the presence of zinc and DTPA………..124

Chapter 6.

Figure 1. *S. epidermidis* biofilm inhibition by amyloid-specific small molecules……………139
Appendix 1.

Figure 1. Substrate coordination by wild-type and I170K CAN...........................................156

Figure 2. A structural shift in the central calcium coordination by WT-CAN in the presence of substrate..................................................................................................................157

Figure 3. Calcium coordination by Apo-CAN.................................................................161

Figure 4. Calcium binding in the active site is dynamic upon substrate interaction...........163

Figure 5. Arginine 113 regulates dimer enhanced enzymatic activity.........................166-168

Figure 6. The R113A SCAN mutant dimerizes but does not exhibit dimer-enhance nucleotidase activity ..............................................................................................................................172

Figure 7. Differences in the substrate orientation within the binding pocket of monomeric and dimeric CAN forms.................................................................178

Figure 8. A proposed mechanism of enzyme catalysis in dimeric and monomeric CAN forms.................................................................180
Tables

Chapter 2.
Table 1. Comparison of secondary structure characteristics of G5 domain proteins……………38

Chapter 5.
Table 1. Components of the CDM + RPMI medium………………………………………………105
Table 2. List of metal chelators used in this study………………………………………………115

Appendix 1.
Table 1. Amino acid numbering conventions of current CAN structures……………………..148
Table 2. Summary of crystallographic data and refinement statistics …………………………..158
Table 3. Comparison of surface area and volume of wild-type and I170K CAN structures…..174
Table 4. Total known CAN residue mutations and their respective enzymatic activities………184
Abbreviations

1,10 – P, 1, 10 – phenanthroline
Aap, Accumulation-associated protein
ADP, adenosine diphosphate
AMP, adenosine monophosphate
ATP, adenosine triphosphate
AFM, atomic force microscopy
AUC, analytical ultracentrifugation
BHI, Blood-heart infusion medium
Brpt, B-domain repeat
Ca, calcium
CAN, calcium-activated nucleotidase
CBS1, calcium-binding site 1
CBS2, calcium-binding site 2
CD, circular dichroism
CDM, chemically-defined medium
CFM, confocal fluorescence microscopy
CryoEM, cryo-electron microscopy
Cu, copper
DTPA, diethylenetriaminepentacetic acid
EDTA, ethylenediaminetetra-acetic acid
EPS, exopolysaccharide
FA, formic acid
f/f₀, frictional ratio
FTIR, Fourier transfer infrared spectroscopy
GDP, guanosine diphosphate
GMPCP, guanosine-5’-[(α,β) – methyleno] diphosphate
HAI, hospital-acquired infection
HCl, hydrochloric acid
HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His, 6 x histidine tag
HPLC, high-pressure liquid chromatography
I170K, monomer mutant of CAN
IPTG, Isopropyl β-D-1-thiogalactopyranoside
PEG, polyethylene glycol
PGA, poly-γ-DL-glutamic acid
pH, measure of acidity or basicity of a solution
PIA, polysaccharide intercellular adhesion
PNAG, poly-N-acetylglucosamine
PVE, prosthetic valve endocarditis
MBP, mannose-binding protein
MES, morpholino-ethane sulfonic acid
MIC, minimal inhibitory concentration

MRSA, methicillin-resistant *Staphylococcus aureus*

MSCRAMM, microbial surface components recognizing adhesive matrix molecules

NaOH, sodium hydroxide

NMR, nuclear magnetic resonance

RPMI, 50X concentrated amino acid stock

SCAN, soluble calcium-activated nucleotidase

SDS, sodium dodecyl sulfate

StsG5, construct of G5 domains from zmpC

TEM, transmission electron microscopy

ThT, thioflavin T

TLS, Translation/Libration/Screw

TSB, tryptic soy broth

TPEN, N,N,N',N'-Tetrakis(2-pyrdiylmethyl) ethylenediamine

VC, vehicle control

zmpC, zinc metalloprotease

Zn, zinc
Symbols

α - alpha
β - beta
γ - gamma
δ - delta
σ – sigma
θ - degree
ºC - degrees Celsius
Å - Angstroms
μ - micro
m- milli
nm - nanometer
l- liter
M- molar
Chapter 1

*Staphylococcus epidermidis*: The Opportunistic Pathogen
A Literature Review
Introduction

*Staphylococcus epidermidis* has established a long-standing commensal relationship with mammals and is the most common Staphylococcal species in the human body, colonizing the mucous membranes and epithelia. *S. epidermidis* is a critical component of the protective microflora, preventing attachment and infection by non-native pathogens and helping to control population densities of other native microbial species. In the past, there has been little interest to study the molecular physiology of *S. epidermidis* in its natural and beneficial state because the research spotlight has always focused on the life-threatening and disease-causing microorganisms. However, the advancements of medical technology and frequent exposure to antibiotic environmental pressures have required commensal microbes to evolve in order to survive in their native habitat. As a result, *S. epidermidis* has developed an ever-growing resistance to antibiotics and the ability to colonize prosthetic devices, catheters, and biomaterials that has allowed a broad spectrum of tissues and cell types to become susceptible to infection. Therefore, *S. epidermidis* has become one of the leading causes of nosocomial infections, bacteremia, and device-related infections, affecting millions of people world-wide every year. However, the most remarkable component for this species to cause infection is that *S. epidermidis* does not genetically house the stockpile of virulence factors that are relied upon by other bacterial pathogens to penetrate host defenses. Instead, *S. epidermidis* relies solely on its ability to establish a highly chemically-resistant cellular accumulation called a biofilm. Genes for biofilm formation have been shown to be detrimental to its native commensal colonization [14, 25], suggesting that new environmental stress has driven this species to a more defensive, and therefore pathogenic, state.
I. **Staphylococcus epidermidis** Bacterial Profile

*Bacterial classification and physiology*

*S. epidermidis* is a gram-positive cocci that divides in grape-like cluster motifs and forms small, pin-point, white colonies. *S. epidermidis* is further differentiated into the coagulase-negative sub-group from other Staphylococci, such as its close relative *Staphylococcus aureus*, which means that it does not carry the gene for coagulase [7]. Coagulase is a protein that combines with prothrombin in the blood, forming the staphylothrombin complex, in order to proteolytically activate fibrinogen for the formation of blood clots [8, 9]. All gram-positive species contain peptidoglycan and teichoic acid within their cell wall; *S. epidermidis*, however, contains a unique teichoic acid composition containing α-glucosylglycerol teichoic acid [10].

*Native Colonization*

*S. epidermidis* is most associated with its commensal relationship with humans, in which it is a major colonizer of mucosal and epithelial surfaces: primarily the armpits, head, and nostrils at densities ranging from $10^4 – 10^5$ CFU/cm$^2$ [2, 11]. Epidemiological evaluation of *S. epidermidis* colonization suggests that every healthy person contains between 10 and 24 unique strains of this species at any given time [12]. *S. epidermidis* as a species is genetically diverse [13], which is attributed to the presence or absence of prophages and genetic islands [2]. As a result, *S. epidermidis* has generated as many as 74 genetically unique isolates from humans that have been identified [2] [14]. The prevalence of *S. epidermidis* on external surfaces and its ability to survive in harsh environmental conditions is attributed to the expression of a multitude of proteins dedicated to osmoprotection, including sodium ion pumps and metal transport systems [15, 16].
Antibiotic resistance

While *S. epidermidis* is a key component of the human normal flora, it has acquired genes for pathogenesis, presumably by horizontal gene transfer of transposable elements [16]. *S. epidermidis* has developed antibiotic resistance to a broad spectrum of drug therapies as a result of selective pressures within an antibiotic-abundant environment and an availability of resistance genes that can be obtained via conjugation [17]. Methicillin was previously the common method of treatment against staphylococcal infections; however, 80% of coagulase-negative staphylococci have become methicillin-resistant within the last decade [18]. Methicillin resistance is conferred by the *mec* gene within the staphylococcal cassette chromosome (called *SCCmec*), which is a mobile element that can potentially be spread to other species by horizontal gene transfer [19]. A representative study in 2006 found that of all *S. epidermidis* strains clinically isolated from The University of Geneva Hospitals (Switzerland), 65%, 61%, 47%, and 44% demonstrated resistance to erythromycin, ciprofloxacin, clindamycin, and gentamicin, respectively [18]. Vancomycin is currently the last resort treatment for *S. epidermidis* strains that have multi-drug resistance, although even vancomycin resistance is developing among clinical isolates [20].

Virulence factors

As “tough” as this bacterial species may be, it actually possesses very few aggressive virulence factors and is genetically focused towards a long-term commensalism with its host [2]. *S. epidermidis* does produce a poly-y-DL-glutamic acid (PGA) capsule that provides protection from host defenses [21]. While capsule production is likely required to survive in the harsh conditions of its native colonization environment, it has also been shown to be an important
virulence factor during pathogenesis [21]. *S. epidermidis* also produces a mild class of toxin called phenol-soluble modulins (PSMs), which are small, amphipathic peptides capable of mediating pore-forming, cytolytic effects on both bacterial and mammalian cells [22]. However, PSMs are hypothesized to play a role in the commensal survival of *S. epidermidis* on human skin rather than pathogenesis, providing protection against antimicrobial peptides produced by other commensal species [22]. Unlike the more virulent *S. aureus*, coagulase-negative staphylococci do not typically possess prophage-mediated toxin genes. An *S. epidermidis* strain (FRI909) was recently identified to contain an enterotoxin gene within a transposable element derived from *S. aureus*, suggesting the potential for *S. epidermidis* to become more proactively pathogenic in the future [23].

**Biofilm formation: virulence factor or artifact of commensal colonization?**

The majority of *S. epidermidis* strains have acquired or developed genes to support biofilm formation, the primary means of pathogenesis that causes infection associated with surgery, implanted medical devices, and deep wounds [24] (Figure 1). Unlike the previously mentioned pathogenicity traits, biofilm formation provides no function for commensal colonization and has actually been shown to be detrimental [25]. In correlation with the idea that biofilm formation is strictly for pathogenesis, the gene cluster *Ses* (I, G, and C) has only been found in clinical isolates that are primarily biofilm-forming strains [26] and additionally, *Ses* mutants have reduced biofilm-forming and fibrinogen-binding capabilities *in vitro* [27]. This indicates that *S. epidermidis* has two very distinct genetic “personalities,” governed by genes required for commensal colonization and genes involved in pathogenesis. Therefore, to truly
understand *S. epidermidis* as a pathogen, the genetic regulation and molecular mechanisms that promote biofilm formation must be determined.

**Figure 1.** Fluorescence confocal microscopy image of a *Staphylococcus epidermidis* biofilm formed within a plastic catheter. The left image represents a top-down view of a 1-µm section of the biofilm, cells stacking to form dense structures. The right image is a stack of these sectioned images rotated 90° to demonstrate the three-dimensional nature of the biofilm. Green represents living cells, red represents dead cells, yellow is merged. Images were taken using a Zeiss LSM 510 Inverted Microscope and the image was processed using Zen2009 Light edition.
II. Stages of Biofilm Formation

What is a biofilm?

Biofilms are formed by an array of gram-positive (e.g. Bacillus sp., Corynebacterium sp., Enterococcus sp., Finegoldia sp., Staphylococcus sp., Streptococcus sp.) or gram-negative (e.g. Pseudomonas sp., Salmonella sp., Actinobacillus sp.) species for survival within the environment and for pathogenesis [33]. A biofilm is a multi-layered bacterial community that consists of a network of aggregated bacterial cells encased by a polysaccharide matrix [34] (Figure 2). Biofilm formation is favorable for prolonged bacterial survival because it prevents components from the exterior environment (antibiotics, host immune cells, and chemical biocides) from harming individual cells [34]. Biofilm formation alone increases the minimal inhibitory concentration for most antibiotics 1,000-fold [18]. This phenomenon is potentially caused by the lack of antibiotic penetration or because cells within the biofilm have decreased metabolic activity to limited nutrient availability [34]. It was recently shown that S. epidermidis biofilm formation also prevents deposition of complement factor C3b and IgG antibodies, which would prevent opsonization by the host immune system [35]. Biofilm formation occurs through four processes: adherence, accumulation, maturation, and detachment.
Phase I: Adherence

The adherence of *S. epidermidis* to a surface is the first critical step for infection and biofilm formation. Most surfaces in the human body are coated with fibrous proteins, such as collagen, and bathed in plasma proteins, such as fibrinogen and fibronectin, that provide epitopes for *S. epidermidis* attachment [33, 36]. *S. epidermidis* expresses multiple surface proteins, called MSCRAMMs (*microbial surface components recognizing adhesive matrix molecules*), with an LPXTG cell wall binding motif whose function is for biomaterial-bacterial adhesion [37]. For example, *S. epidermidis* produces MSCRAMMs SdrF, lipase GehD, EmbpI, and Fbe/SdrG to attach to type I collagen [38], types I, II, IV collagen [39], fibronectin [40], and fibrinogen [41, 42], respectively.

---

**Figure 2. Model of *S. epidermidis* biofilm formation.** Bacterial cells (blue spheres) first adhere to surface (I) using specific adherence to host proteins or non-specific interactions to artificial surfaces. This phase is followed by an intercellular accumulation event that is critical for biofilm formation (II) and is mediated by both polysaccharide and protein-dependent factors. Biofilm maturation (III) is indicated by formation of three-dimensional structures and production of an exopolysaccharide matrix (orange). Long term persistence of the biofilm will initiate detachment of individual cells (IV) from the biofilm in order to spread the infection and prevent nutrient depletion. This model was adapted from Otto, M. *Nature Microbiology Reviews*, 2009 [2].

**Phase I: Adherence**

The adherence of *S. epidermidis* to a surface is the first critical step for infection and biofilm formation. Most surfaces in the human body are coated with fibrous proteins, such as collagen, and bathed in plasma proteins, such as fibrinogen and fibronectin, that provide epitopes for *S. epidermidis* attachment [33, 36]. *S. epidermidis* expresses multiple surface proteins, called MSCRAMMs (*microbial surface components recognizing adhesive matrix molecules*), with an LPXTG cell wall binding motif whose function is for biomaterial-bacterial adhesion [37]. For example, *S. epidermidis* produces MSCRAMMs SdrF, lipase GehD, EmbpI, and Fbe/SdrG to attach to type I collagen [38], types I, II, IV collagen [39], fibronectin [40], and fibrinogen [41, 42], respectively.
While many of these adhesion targets are quite abundant in the extracellular milieu, adhesion factors must overcome the force of fluids moving within the body as well as degradative proteases to establish a highly stable surface attachment. One mechanism that is proposed as to how MSCRAMMs maintain attachment is the “dock, lock, and latch” mechanism [43]. This mechanism for attachment proposes that the MSCRAMM protein will adopt an open-to-closed conformational change upon ligand binding, securing the cellular contact [43]. This mechanism has specifically been shown by the protein SdrG through structural analysis and differences in the Förster resonance energy transfer of the amino acid contacts in ligand bound and unbound states [43]. In addition to adherence to host proteins, \textit{S. epidermidis} produces surface proteins for attachment to artificial surfaces as well. The surface-associated autolysins AtlE and Aae have been shown to adhere to not only biomaterials (fibrinogen, fibronectin, and vitronectin), they additionally mediate non-specific hydrophobic interactions to polystyrene that enables bacterial adhesion to synthetic and implanted medical devices [44, 45]. Teichoic acid, an important cell wall component that is chemically-unique to \textit{S. epidermidis} from other staphylococci, has also been implicated to mediate adhesion to fibronectin [46]. This provides a physiological rationale for the unique composition of teichoic acid in \textit{S. epidermidis}, particularly since pathogenic traits are secondary in this species. An \textit{S. epidermidis} surface-displayed polysaccharide, called polysaccharide intercellular adhesion (PIA), has also been shown to mediate cellular attachment, although it is primarily utilized for the second accumulation phase of biofilm formation [47]. A collection of additional host extracellular matrix proteins have also been identified to potentially act as ligands for \textit{S. epidermidis} attachment, which displays the redundancy and importance of adhesion for survival and biofilm formation [45].
Phase II: Accumulation

Bacterial accumulation refers specifically to the interaction between bacterial cells and is completely separate from the primary attachment event [48]. Intercellular accumulation for *S. epidermidis* occurs through two primary methods: polysaccharide or protein-mediated intercellular adhesion. Both methods can be utilized independently or synergistically and have demonstrated an equal contribution for the persistence of *S. epidermidis* infection [49-53], although there is currently no correlation to specific types of infection and the utilization of either type of accumulation agent [54].

Polysaccharide intercellular adhesion

Aptly described as “slime,” exopolysaccharide (EPS) is an important component of the biofilm and aids in bacterial survival in both gram-positive and gram-negative bacteria [34, 55], although the molecular composition of EPS differs significantly among these bacterial groups [56]. The EPS of *S. epidermidis* biofilms, polysaccharide intercellular adhesin (PIA, or now referred to as PNAG: poly-N-acetylglucosamine), was first identified to be responsible for accumulation by biofilm negative strains that were produced using random transposon mutagenesis [48]. It was first suspected that the chemical composition of PNAG was considerably different from the other previously characterized gram-negative EPS moieties due to a noticeable difference in charge; other EPS was previously found to be polyanionic or neutral whereas PNAG was found to be polycationic [57]. The composition of PNAG was later determined to be composed of two glycan polymers, termed polysaccharide I and II [57]. Polysaccharide I is a predominately homo-polymerized glycan with at least 130 β-1,6-linked 2-deoxy-2-amino-D-glucopyranosyl residues, 80-85% of which are N-acetylated (a type of N-
acetylglucosamine) [57]. Polysaccharide II has a similar base structure to polysaccharide I although additionally containing phosphate and/or ester-linked succinate and an increase of N-acetylation (91%) [57].

The genes encoding PNAG production are chromosomally encoded within the ica operon, which includes icaADBC [58, 59]. IcaA, in conjunction with its accessory protein IcaD, was determined to be a membrane-associated glycosyltransferase that uses UDP-N-acetylglucosamine as the substrate to produce variable length N-acetylglucosamine oligomers (20 and less) [59]. IcaC functions to increase the polymer length of the glycan chain, as full length polysaccharide I has at least 130 glycan residues [59]. IcaD and IcaC are both functionally dependent on the presence of IcaA, since IcaA is the only protein involved in PNAG synthesis that has minimal activity alone [59]. IcaB is a cell surface-associated protein that is responsible for PNAG deacetylation, and therefore the net positive charge, and has been shown to be essential for biofilm accumulation and S. epidermidis virulence by preventing recognition by host immune cells [60]. In addition, the positive charge generated by deacetylation allows PNAG to restrict the flow of cationic molecules, such as antibiotics, to protect the cells within the biofilm [61]. PNAG has also been shown to mediate erythrocyte agglutination [62] and overall, a critical factor for immune evasion [55].

Protein-dependent intercellular adhesion

Previously, polysaccharide-mediated intercellular adhesion was the only known method for bacterial accumulation until staphylococcal strains that lacked the ica operon were found to still form biofilms via proteinaceous products [48, 63-65]. An S. epidermidis accumulation-negative mutant, M7, was previously characterized to be missing expression of a 115 kD protein
Further analysis of the M7 mutant by Hussain et al revealed expression of variable length protein products from 115 to 180 kD and addition of an antibody specific to this protein inhibited *S. epidermidis* accumulation [67], indicating that this protein was sufficient to function alone for accumulation. This protein was therefore classified as the Accumulation-associated protein, or Aap [67].

The importance of Aap for *S. epidermidis* biofilm formation and pathogenicity was further shown by analysis of the icaA transposon mutant strain 5179 from a cerebral-spinal fluid shunt infection [68]. Biofilms formed by this strain demonstrated sensitivity to only proteinase K and not sodium metaperiodate, a carbohydrate oxidation agent, indicating a protein-dependent and polysaccharide-independent biofilm was possible [68]. The protein responsible for this protein-dependent biofilm was identified as a truncated form of Aap (amino acids 596 – 1507); this was deduced from the mismatch between the predicted open reading frame (4521 nucleotides) and the N-terminal sequencing product of the isolated protein [68]. Further sequence analysis of Aap revealed a multi-domain protein consisting of: an N-terminal export signal followed by a 564 amino acid A domain (11 degenerate, 16 amino acid repeats), a putative globular α/β domain, a B domain of five repeats of 128 amino acids each and a sixth partial repeat of 68 amino acids, a collagen-like, proline/glycine rich repeat domain, and an LPXTG cell wall anchor motif (Figure 3) [68].
It was determined that proteolytic processing of the A and α/β domains is required to expose the B domain for protein accumulation, in that expression of full length Aap in the biofilm-negative *Staphylococcus carnosus* TM300 does not result in biofilm formation [68]. The function of the A domain has not been completely resolved, however, one study does show that this domain binds to surface epithelial cells and therefore functions for native colonization [69]. Another study demonstrates that adding recombinantly expressed A-domain to a biofilm-forming strain of *S. epidermidis* inhibits biofilm formation, suggesting that this domain has a minor role in biofilm regulation [68]. In addition, while proteolytic processing of Aap is required for protein function, the proteases responsible and the time point of cleavage remains unknown [68]. Through sequence analysis, the B domain was found to contain a variable number (3-17, strain dependent) of tandem, nearly identical, B domain repeats of 128 amino acids each [70] followed by a partial 68 amino acid repeat that has been implicated for protein stability [3, 68]. *S. aureus* also expresses a protein homologue of Aap, SasG, which has been implicated for intercellular adhesion and contains seven G5 domains that are 60-67% identical to the G5 domains of Aap [71]. G5 domains can be found in thirty-five other gram-positive species and also within proteins associated with bacterial pathogenesis and antibiotic resistance [72]; although this
number has since grown to 398 species since the sequencing of new genomes [73]. Other G5 domains within this family have been suggested to interact with N-acetylglucosamine, which would enable an interaction between Aap and PNAG [72]; however, this has yet to be shown experimentally. Aap has also been shown to form fibrils (159 +/- 35 nm) on the surface of planktonic S. epidermidis cells, although the exact physiological function of these fibrils has not been determined [75].

Biofilm experiments by Corrigan et al have demonstrated that five G5 domain repeats of SasG are necessary and sufficient for S. aureus biofilm formation, suggesting the importance of multiple, tandem G5 domain repeats for protein function [71]. Recently, Conrady et al have characterized a single B-repeat with the terminal half-repeat cap of Aap and determined that the protein self-associates to form a dimer exclusively upon the addition of zinc (EC$_{50}$ = 5.4 mM) [3]. While this concentration of zinc is much greater than physiological concentrations (~15 µM in the blood), it is hypothesized that binding cooperativity occurs as the number of tandem G5 domains increases, providing a function for the multiple, tandem G5 domains in Aap and SasG. This is directly inferred from additional studies of a B-domain construct containing two B-repeats that required less zinc to achieve a 50% bound state (EC$_{50}$ = 3.7 mM) [3]. In correlation with this data, biofilm assays with S. epidermidis strain RP62a and S. aureus MRSA strain USA300 demonstrated biofilm inhibition upon addition of the metal chelator diethylenetriaminepentacetic acid (DTPA) and biofilm restoration with add back of 15-20 µM zinc chloride, which does match the physiological concentration of zinc [3]. Based on this work, a model for protein-dependent intercellular adhesion has been hypothesized. This model suggests that G5 domains associate in the presence of zinc to form an intercellular network that connects adjacent bacterial cells [3]. Consistent with this hypothesis, a construct containing two
G5 domains from SasG was recently shown to self-associate to form a dimer in the presence of zinc [74]. This model for zinc-mediated intercellular adhesion does provide critical insights for staphylococcal biofilm formation; however, it does not completely address the molecular mechanism of the full-length constructs of Aap or SasG and specifically the function of multiple, tandem G5 domains contained within these proteins.

While Aap is believed to be the primary protein that contributes to protein-dependent intercellular accumulation, other proteins have been shown to be independently responsible for accumulation as well. First identified in *S. aureus* using transposon mutagenesis, the “biofilm-associated protein” (Bap) was shown to support biofilm formation independently [76]. Bap contains similar domain architecture to Aap, consisting of multiple domains that are composed of repetitive sequences with a C-terminal LPXTG cell wall anchor; however, Bap does not share significant sequence identity with Aap [76]. A Bap homolog, Bhp, has been identified within the genome of *S. epidermidis* that is 97% identical to Bap [77]. Bap has been implicated as an important *S. aureus* virulence factor for bovine mastitis infection [78], however, only 50% of human clinical *S. epidermidis* isolates contained Bhp [70]. This suggests that it is not likely to be a critical factor for human pathogenicity compared to Aap and PNAG. Another protein that has recently been shown to mediate intercellular adhesion in *S. epidermidis* is Embp1 [79]. This giant 1 MDa protein complex is capable of adhering to fibronectin as well as being independently sufficient for biofilm formation in a PNAG- and Aap-negative *S. epidermidis* mutant [79]. Biofilms formed in the presence of both Embp1 and PNAG have a synergistic effect, demonstrating resistance to both proteinase K and the N-acetylglucosamine hydrolase dispersin B [79]. *S. epidermidis* strains expressing Embp1 were also shown to have decreased uptake by mouse macrophages, indicating that Embp1 has a protective function from the host.
immune system [79]. This new method of protein-dependent accumulation demonstrates the possibility of other, unknown proteins that can be utilized for biofilm accumulation, maximizing the survival of *S. epidermidis* as a pathogen.

*Accumulation factors with respect to pathogenesis*

PNAG has been characterized as a critical virulence factor for *S. epidermidis* pathogenesis within two animal models, in which a PNAG-negative *S. epidermidis* mutant was shown to not cause an invasive infection [49, 50]. These models, however, used an *S. epidermidis* strain 1457) that has been shown to not express protein-dependent accumulation factors [68] and therefore does not represent an accurate model of *S. epidermidis* infection. A third PNAG infection model in *C. elegans* did demonstrate a decrease in virulence using an icaADBC deletion mutant strain [55]; however, infection was not completely abolished and the contributions of protein-dependent accumulation were not determined. A genetic analysis study in 1997 using southern blots of clinically pathogenic strains vs. commensal isolates of *S. epidermidis* has demonstrated that 85% of pathogenic strains contain the ica gene locus compared to 6% of isolates from healthy individuals, supporting the idea that PNAG is truly a virulence factor only [65]. A second analysis by PCR in 2000, however, reported the presence of ica genes in only 68% and 37% of clinical and saprophytic strains, respectively [80]; neither study measured protein-dependent factors. While these new statistics may represent an increase in the presence of community-acquired infectious strains of *S. epidermidis*, it also suggests that PNAG is not necessarily the sole determinant of pathogenicity. Anti-Aap antibodies were shown to inhibit polysaccharide, PNAG-based biofilms, which suggests that Aap could be required for PNAG cell attachment or accumulation [81]. In addition, a more recent study showed that only
61% of prosthetic joint infection isolates were PNAG-positive despite the presence of Aap in 90% of these strains [70]. This data, in correlation to other supporting studies, suggests that PNAG-independent, proteinaceous biofilm formation is clinically relevant [51-53, 82]. While both PNAG- and protein-dependent factors have been shown to be important for *S. epidermidis* pathogenesis, the relative contributions of each accumulation factor and their potential interactions still remain unclear.

**Phases III and IV. Biofilm Maturation and Detachment**

Biofilm maturation is characterized by the formation of multi-layered, three-dimensional structures and the secretion of an exopolysaccharide matrix [34]. Despite the stationary structure of the biofilm, it is still a dynamic community with actively dividing cells and metabolic requirements [83]. Therefore, clusters of mature biofilm or planktonic cells will eventually be released from the biofilm to control the size of the biofilm and to potentially spread the infection [2, 84]. Based on the dynamics of other related bacterial biofilms, such as *S. aureus*, it is known that enzymatic degradation of biofilm components is required for development of nutrient channels and biofilm detachment; however, such proteases and glycan hydrolases have not been identified in *S. epidermidis* [2, 85]. While the mechanism for biofilm maturation and natural dispersal is still quite mysterious, it is well accepted that complete biofilm maturation is critical for long-term survival of *S. epidermidis* from host defenses and antibiotics. Bacterial cells within immature biofilms have been shown to be more susceptible to antibiotics and host immune defenses than cells within a mature biofilm [86].
PGA: the other glycan

PNAG is one of the primary components of this extracellular matrix and provides resistance to anti-microbial peptides, such as β-defensins, and prevents phagocytosis by neutrophils [87]. However, as previously mentioned, PNAG is not found in all strains of *S. epidermidis*, and a second, ubiquitous glycan component was therefore suspected to be aiding in formation of the exopolysaccharide matrix. Originally discovered in *Bacillus* species, poly-γ-DL-glutamic acid (PGA) was recently shown to be important in *S. epidermidis* biofilms as well [21]. Genes for PGA synthesis are found within the cap locus, which is conserved among all strains of *S. epidermidis*, some species of coagulase-negative staphylococci, and other pathogenic bacterial species [21]. In addition to its contributions to the biofilm extracellular matrix and protection from host defenses, PGA is also considered to play a role in the commensal colonization by protecting the cell surface from extreme conditions of the skin, such as high salt concentrations [21]. The overlapping role of PGA – protection of *S. epidermidis* both in its native environment and in a biofilm - demonstrates the ability of *S. epidermidis* to be an opportunistic pathogen.
III. Communication within the Biofilm

Quorum sensing controls biofilm detachment

An important component of biofilm architecture is the formation of deep channels between thick multi-layers of cells that allow delivery of nutrients to living cells within the biofilm [2, 83]. In order to regulate this dynamic rearrangement process, *S. epidermidis* has developed a method of communication within the biofilm called quorum sensing [88]. Quorum sensing utilizes a bacterially-secreted signal, the pheromone, to sense the cell density within the biofilm community [89]. *S. epidermidis* has two known quorum-sensing systems: *agr* and *luxS* [88, 90].

The accessory gene regulator system (*agr*) is a two-component response regulator system that consists of AgrC and AgrA, for signal transduction, and AgrB, the pheromone to be exported [88]. The *luxS* quorum-sensing system responds to the signaling molecule autoinducer-2 (AI-2) that is also hypothesized to be utilized for inter-species communication [90]. Deletion of either the *agr* or *luxS* locus in *S. epidermidis* induces a phenotype that produces a thicker biofilm with increased attachment and infectivity; this demonstrates that quorum sensing controls biofilm expansion and structural maintenance [89, 90]. The presence of glucose can inhibit *agr* regulation and has been shown to activate *ica* transcription, suggesting a related regulatory pathway [85, 91, 92]. However, *S. epidermidis* biofilms exposed to glucose-limiting conditions over a long period of time demonstrate disintegration of the primary biofilm rather than detachment, which suggests that a secondary regulatory pathway must exist [93]. *agr* suppression of biofilm expansion could also be regulated by the two component regulatory system *saeRS*; *sae* operon deletion strains have been shown to produce thicker biofilms under
anaerobic conditions [94]. While the complete molecular mechanism of *S. epidermidis* biofilm detachment is not well understood, the *agr* quorum sensing system has been shown in *S. aureus* to regulate this process, in which *agr* activation by auto-inducing peptides initiates the activation of proteases that would cause detachment [85]. δ toxin, a phenol-soluble modulin (PSM), is also proposed to aid in detachment by interrupting cellular contacts [95, 96]. PSMs are a protein complex composed of three amphipathic peptides and are highly associated with the ability to cause an inflammatory response [251]. PSMs are only expressed during late stages of infection by activation of *agr* transcription, which is consistent with the idea of PSMs being involved in biofilm detachment [95].
IV. Genetic Regulation of Biofilm Formation

Regulation of PNAG production by the transcriptional repressor IcaR

While biofilm formation is favorable for long-term bacterial survival as a pathogen, it is still a tightly-regulated process. The ica gene locus can be transcriptionally activated by environmental stress factors, such as high salt, temperature, and ethanol [97]. Environmental stress factors, such as ethanol and high osmolarity, activate an unidentified cell surface response regulator that further activates the gene rsbU of the sigma B operon [97]. Transcription of this operon produces alternative σB, which activates PNAG production [97]. σB activates the ica operon by repressing expression of a transcriptional repressor, IcaR [91]. Recent crystallographic studies of IcaR indicate that two IcaR dimers bind to a 28 bp region that is located 17 bases upstream from the ica start codon [250]. Oxygen has also been shown to regulate ica transcription through the sigma B operon. During anaerobic conditions, icaR transcription is suppressed and abnormally high levels of oxygen cause ica inhibition by increasing icaR transcription [98].

Alternative regulatory pathways

Based on these studies, it has been accepted that sigma B controls transcription of IcaR-dependent regulatory pathways; however, there are additional regulatory pathways that exist independently of IcaR [91, 99]. For example, glucose activates ica transcription via an indirect, unknown factor that is independent of sigma B regulation [91, 92]. In addition, deletion of the sigma B locus prevents the production of PNAG without preventing cell accumulation, suggesting an alternative regulation pathway for protein-dependent accumulation factors [91,
Alternative regulation of protein-dependent factors is additionally supported by a PNAG deletion strain that experiences an increase of Aap transcription in the presence of ethanol [101]. This strain also shows a decrease in Aap expression in the presence of high salt and no induction of protein expression by glucose, the latter of which is contrary to PNAG regulation [101]. Later studies determined that SarA, a known transcriptional regulator for virulence factors in *S. aureus*, activates PNAG production directly by binding *icaA* [102]. Although *sarA* deletion mutants also demonstrated biofilm restoration upon addition of ethanol or high osmolarity, suggesting still other, independent pathways exist and that SarA pathways regulate transcription independently of σB [102, 103]. Both the *ica* and *sigma B* operons are conserved among all biofilm-forming strains of *S. epidermidis*; however, there is still a high degree of variability in the amount of *ica* gene products produced in the presence of identical environmental factors within the same strain [97, 104].
V. Clinical Insights to *S. epidermidis* Infection and Treatment

*S. epidermidis: the opportunistic pathogen*

The combination of skin surface exposure, acquired antibiotic resistance, and the ability to form a biofilm have enabled *S. epidermidis* to take advantage of new surfaces for colonization and become one of the most prevalent opportunistic pathogens [7], causing 66% of all hospital-acquired infections [105]. Pathogenesis of *S. epidermidis* is clinically associated with a broad range of diseases, consisting of bacteremia, skin graft infection, endocarditis, prosthetic device cardiac and joint infections, and surgical site infection [24]. Coagulase-negative staphylococci are the leading cause of nosocomial bloodstream infections [24], causing 30-40% of infections, and the second leading cause of prosthetic valve endocarditis (PVE), the majority of which are caused by *S. epidermidis* [106, 107]. While PVE only occurs in roughly 3% of patients with cardiac valve disease, PVE is associated with a mortality rate of 19-50% [108]. Both *S. aureus* and *S. epidermidis* endocarditis infections also have a higher incidence of complications during PVE infections, such as heart failure, cardiac abscess, subclinical brain embolization, and acute brain embolism [106, 109]. An additional obstacle for treatment of these infections is the growing problem of antibiotic resistance, given that 70-90% of all hospital isolates of *S. epidermidis* in the United States are methicillin-resistant [2]. While 50-60% of staphylococcal PVE isolates are antibiotic-resistant [106, 110], there is still a substantial non-resistant population that sustains infection through other methods, such as biofilm formation. The synthetic biomaterials in prosthetic valves particularly provide a surface to support bacterial adherence and biofilm formation [107, 111-113].
Current therapeutic methods

Organizations such as the Center for Disease Control (CDC), the Society for Healthcare Epidemiology of America (SHEA), the Infectious Diseases Society of America (IDSA), and the American Heart Association (AHA) currently provide guidelines for the prevention of coagulase-negative staphylococcal infections, estimating that 10 – 70% of these infections are preventable [24]. However, the guidelines require cooperation of all healthcare personnel to follow proper protocols and only recommend antibiotic prophylaxis in specialized, high-risk cases due to a greater risk to the patient than benefit from treatment [24, 114, 115]. Therefore, novel methods for the prevention of nosocomial infections and infectious endocarditis, particularly caused by staphylococci, are needed in order to increase patient survival and decrease the morbidity and mortality related to disease caused by coagulase-negative staphylococci such as S. epidermidis.
Scope and Thesis Outline

The primary goal of this thesis was to better understand the molecular mechanism underlying Accumulation-associated protein (Aap)-dependent zinc-mediated biofilm formation by *Staphylococcus epidermidis*. More specifically, the aim of this thesis is to i) molecularly characterize the self-assembly of multiple, tandem B domain repeats of Aap in the presence of zinc and ii) determine the functional relationship between the multiple B domain repeats and *S. epidermidis* biofilm formation. Chapter 2 discusses the solution characterization of variable length constructs of the B domain of Aap in the presence of zinc. This work provided a baseline for the behavior of longer B domain repeat constructs and a single, representative longer construct was selected (Brpt5.5) to continue more detailed biochemical studies in Chapter 4. Due to the unexpected and unusual nature of formation of amyloid fibers by the Brpt5.5 construct, an additional literature background on functional amyloid formation was provided in Chapter 3 to provide better insight to the work in Chapter 4. Chapter 5 investigates critical environmental factors that affect the Aap assembly mechanism in relation to biofilm formation. Finally, Chapter 6 summarizes the results and provides future experimental directions to help address the missing elements of the staphylococcal biofilm formation mechanism in order to develop better therapeutic and preventative treatments for chronic staphylococcal infections.
Chapter 2

Functional Analysis of

Proteins Containing Multiple, Tandem G5 Domains
Abstract

G5 domains are functional protein domains that have been characterized by five, non-positionally conserved glycine residues contained within a 76 – 84 amino acid sequence. G5 domains have been identified in proteins from 213 different bacterial species, the majority of which are gram-positive biofilm formers. While two G5 domain-containing proteins, Aap and SasG, have been implicated to function in biofilm formation in Staphylococci, the diversity of protein domain architectures and species suggests there is a more universal function for this protein domain. A double G5 domain from Aap has been previously characterized to self-assemble in the presence of zinc, suggesting that the G5 domain is a metal-binding adhesion module. In addition, G5 domains are often found in multiple, repetitive copies within a single protein. The function of this repetitive protein organization with respect to protein assembly in the presence of zinc has never been evaluated. Therefore, we have generated multiple G5 domain constructs to provide insight on the ability of this domain to universally induce self-assembly in the presence of zinc and influence the higher-order species formed with increasing numbers of G5 domains in tandem. Using biophysical characterization, we demonstrate that protein constructs containing variable numbers of multiple, tandem G5 domains form different zinc-assembled complexes. These constructs demonstrated a trend that increasing the number of G5 domains increases the size of oligomer formed. We also show a relationship between the zinc concentration and the ability to form higher order, complex species. In addition, we demonstrate that a unique G5 domain-containing protein self-assembles in the presence of zinc. Overall, these results suggest that G5 domains have the potential to be universal metal-binding adhesion modules and provide insight into the function of G5 domains with unique protein architectures.
Introduction

The G5 domain was first identified by the Pfam protein sequence database and was characterized by its conserved pattern of glycine residues [73]. While the sequence motif as a whole differs considerably between proteins, there are some highly conserved residues, that are mostly hydrophobic amino acids, and the G5 domain motif is consistently 76 - 84 amino acids long [72] (Figure 1). The G5 domain motif can be found in 118 different protein architectures, 213 species, and 398 genetically unique strains, demonstrating the diversity and high utilization of this protein domain [73] (Figure 2). The arrangement of the G5 domain within these proteins may be diverse, but the function of this protein group is fairly conserved for bacterial pathogenesis and biofilm formation [72]. 132 of these species are known to form biofilms in diverse habitats in the environment (i.e. Bacillus amyloliquefaciens), for pathogenesis (i.e. Staphylococcus epidermidis and S. aureus), and for commensal colonization (i.e. Streptococcus

**Figure 1. Sequence alignment of G5 domains with diverse protein origins.** Single G5 domain sequences were taken from the VanW proteins of Clostridium perfringens (Cp) and Chloroflexus aggregens (Ca, a gram-negative species), a protein used in vancomycin resistance; beta-N-acetylhexosaminidase (BNAHase) from Streptococcus pneumoniae (Sp); a putative metalloendopeptidase (MEdase) from gram-negative Symbiobacterium thermophilum; a hydrolase (Hydrase) from Finegoldia magna (Fm); a putative cell wall anchor protein (CWAnch) from gram-negative Catonella morbi (Cm); the Accumulation-associated protein (Aap) from Staphylococcus epidermidis (Se); and the putative zinc-metalloprotease (zmpC) from Streptococcus sanguinis (Ss). Residues that are conserved by all proteins are highlighted in yellow. Residues highlighted in cyan are conserved in a majority of the proteins. The sequence alignment was performed in the alignment program t-coffee.
oralis) [73]. Sixty-three species are known pathogens, including multiple species from the staphylococcus, streptococcus, mycobacterium, and clostridium families, many of which also express more than one protein containing a G5 domain [73]. A large majority of these species are found in extreme environmental conditions, such as thermal, deep sea vents. Another commonality shared by this group is that they are almost all gram-positive species; only nineteen of the 213 species are gram-negative [73].

**Figure 2. Tree diagram of diverse species containing G5 domain proteins.** The majority of species are divided in the Firmicute phyla between the Clostridia, Lactobacillales, and Bacillli classes. The two primary species families are Clostriales and Streptococccaeae within the Firmicute division. The second largest division is Actinobacteria, containing Corynebacterium, Nocardia, and Mycobacterium species. Newly discovered species within Chloroflexi, Fusobacteria, and Deinococci have yet to be given a primary division and are separated. The numbers represent total sequences in each division. This tree diagram was created using Pfam data.
The majority of these unique protein architectures contain a single G5 domain, such as the vancomycin resistance protein VanW, zinc metalloproteinases, and other various peptidases. However, there is also a propensity for the G5 domain to be found in multiple, tandem copies; there are forty different protein architectures that contain more than one G5 domain. There are as few as two tandem G5 domains in proteins such as the beta-N-acetylhexosaminidase of Streptococcus pneumoniae and a cell wall anchor peptidase from Actinomyces coleocanis, to as many as eleven without any linker sequence between repeats, such as the uncharacterized surface protein from Corynebacterium aurimucosum, or fifteen total with sporadic gaps between a series of tandem domains, which is seen in an uncharacterized protein (C2HIK0) from the pathogenic anaerobe Finegoldia magna. The majority of these proteins that contain multiple, tandem G5 domains are cell wall anchored and surface expressed proteins, although the diversity of the overall protein function of this group questions what the significance of including multiple, tandem G5 domains is in the protein architecture.

The final two G5 domains from the S. epidermidis Accumulation-associated protein (Aap), out of the 12 total G5 domains, has previously been characterized as a zinc-adhesion module [3], in which the G5 domain self-associated to a dimer exclusively in the presence of zinc. While this study provides insight into the function of a double G5 domain, it does not address the purpose of multiple G5 domains within one protein. Another study has demonstrated the physiological relevance of multiple G5 domains using the S. aureus homologue to Aap, SasG [71]. Like Aap, SasG is required for biofilm formation, and this study demonstrates that a particular number of G5 domain repeats (5) are required to support biofilm formation [71]. Therefore, we aim to address two primary questions: i) is a G5 domain conserved as a metal-binding adhesion module in all proteins and ii) what is the function of multiple, tandem G5
domains in a single protein? To answer these questions, we have generated multiple G5 domain constructs from two different proteins and characterized them biophysically. To begin to determine if the G5 domain is a zinc-adhesion module universally, we have generated a construct containing approximately five G5 domains from the zinc metalloprotease (zmpC) from Streptococcus sanguinis. To determine the significance of multiple, tandem G5 domains within a single protein, we have generated four different constructs of the Accumulation-associated protein (Aap) containing three, five, seven, or nine tandem G5 domains. Using these constructs, we have been able to demonstrate that G5 domains are potentially metal-binding adhesion modules. In addition, proteins that contain multiple, tandem G5 domains exhibit a repeat-length dependency during self-association in the presence of metal.
Materials and Methods

Bacterial Strains and Media - *Staphylococcus epidermidis* strain RP62A (ATCC 35984) and *Streptococcus sanguinis* strain SK36 (ATCC BAA-1455) were purchased directly from ATCC as a glycerol stock and cultured in tryptic soy broth (TSB) or as a freeze-dried sample in brain-heart infusion (BHI) broth, respectively.

Construct design and PCR amplification of StsG5 and Brpt – PCR primers were designed to amplify starting at the most N-terminal G5 domain and continue through the most C-terminal G5 domain (amino acid residues 282 – 1,521 or nucleotides 846 – 4,563) of the zinc metalloprotease, zmpC (NCBI ABN44434). The primer sequences used are:

Forward primer: 5’-AAAAGCCAAGAAAGCAGCCAAAC-3’
Reverse primer: 5’-CTACTCTTCTAGACTGCTTTCGGTGCTTC – 3’

Genomic DNA was extracted from a 3 ml overnight culture of *S. sanguinis* SK36 and used as the DNA template for PCR amplification. The Brpt3.5, 5.5, 7.5, and 9.5 constructs (amino acids 1706 – 2223, 1504 – 2223, 1248 – 2223, and 992-2223, respectively) of Aap (NCBI AAW53239.1) was PCR amplified from RP62A genomic DNA. The primer sequences used are:

Forward primer: 5’-CACCACCTCCTAAAGCAGTCGATGGAGATCCAATTAT-3’
Reverse primer: 5’- CTATGTTGGACCATACTCAACAATTTTCGCAACAGG-3’

The forward primer can prime at any G5 domain, therefore, the PCR products were gel purified (Qiagen) to separate each variable length construct and then verified by nucleotide sequencing (Cincinnati Children’s DNA Sequencing Core) and product size for the number of
G5 domains contained within each product. All PCR products were inserted into the expression vector pHisMBP-DEST (kindly provided by Artem Evdokimov) using Gateway technology, which adds an N-terminal hexahistidine-maltose binding protein (His-MBP) fusion to the protein construct, with an intervening tobacco etch virus protease site. The plasmids were each transformed into the *E. coli* expression cell line BLR(DE3) (Novagen).

**Protein Expression and Purification** - 1 L cultures were inoculated with 10 ml of His-MBP-Brpt/BLR(DE3) or His-MBP/StsG5/BLR(DE3) culture and then allowed to incubate shaking at 37 °C until an OD$_{600}$ of 0.6-0.8 was reached. The cells were cold shocked in an ice bath and doped with 2% ethanol. Protein expression was then induced using 200 µM IPTG overnight shaking at 20 °C. The cells were then harvested, re-suspended, frozen and thawed prior to lysis by french press. The cell lysate was centrifuged, and the soluble fraction was decanted onto a nickel-NTA gravity column. All His-MBP/StsG5 and His-MBP/Brpt proteins were eluted by imidazole step gradient. Fractions confirmed by SDS gel to contain Brpt proteins were then further purified by anion exchange using a 5 ml anionQ fast-flow column (GE Healthcare) followed by size-exclusion chromatography using a superdex 200 column; post-affinity chromatography, His-MBP/StsG5 was only purified by size-exclusion chromatography using a superdex 75 column. The His-MBP fusion was cleaved using TEV protease, followed by repurification over a Ni-NTA and final purification using a superdex 200 (for all Brpt proteins except His-MBP/Brpt5.5 for the AUC experiment in the presence of zinc).

**Circular Dichroism** - Protein samples were dialyzed into 10 mM Tris pH 7.4 and 150 mM NaF. Far-UV CD spectra were obtained using an Aviv 215 spectrophotometer. The concentration of
cleaved Brpt (fusion tag removed) was determined using the molar extinction coefficients of 11,920 M$^{-1}$cm$^{-1}$ (Brpt3.5), 19,370 M$^{-1}$cm$^{-1}$ (Brpt5.5), 25,330 M$^{-1}$cm$^{-1}$ (Brpt7.5), and 29,800 M$^{-1}$cm$^{-1}$ (Brpt9.5), as calculated using the online server ProtParam (web.expasy.org/protparam), and the concentration of His-MBP/StsG5 (uncleaved) was determined using the molar extinction coefficient of 100,620 M$^{-1}$cm$^{-1}$ calculated by ProtParam. Data were analyzed using the CDSSTR program on the online Dichroweb server with reference set 4 (http://dichroweb.cryst.bbk.ac.uk) [116].

*Analytical Ultracentrifugation (AUC)* - Experiments were performed with a Beckman XL-I analytical ultracentrifuge using absorbance optics at 280 nm. Sedimentation velocity experiments were performed at 36,000 rpm at 20 °C (except for the Brpt3.5 and His-MBP/StsG5 datasets, which were performed at 48,000 rpm) with and without ZnCl$_2$ as indicated. Data were analyzed using the Sedfit software [117] and the c(s) model.
Results

*Design and Expression of unique G5 domain constructs* – In order to determine the behavior of multiple, tandem G5 domains within the same protein as well as the universal function of G5 domains from diverse species, we designed new protein constructs using the G5 domain containing Accumulation-associated protein (Aap) from *Staphylococcus epidermidis* and the zinc metalloprotease (zmpC) from *Streptococcus sanguinis* (Figure 3).
Aap from *S. epidermidis* strain RP62a contains up to twelve, 80 amino acid, tandem G5 domains that are each separated by a 48 amino acid linker (called a B-repeat) and are 83-100% identical by sequence (Figure 3A). Because of the high sequence identity, even to the nucleotide level, generating unique primers to isolate specific G5 domains becomes extremely challenging, particularly for Aap. Therefore, we utilized this redundancy and generated a forward primer for Aap that is capable of priming any G5 domain and a unique sequence for the reverse primer at the C-terminal half repeat. This produces a diverse mixture of PCR products containing variable numbers of G5 domains within Aap. These PCR products were then isolated by gel extraction and purified prior to inserting them into the Gateway destination expression vector pHis/MBP-DEST. The number of G5 domains within a single construct was determined by nucleotide sequencing and by the molecular weight on DNA gels. These protein constructs are therefore named by the number of G5 domains contained within the construct (all include the conserved C-terminal half-repeat for stability): Brpt3.5, 5.5, 7.5, and 9.5.

zmpC also contains a series of tandem G5 domains, however, the domain architecture and amino acid sequence is considerably different than Aap (Figure 3B). The first G5 domain of zmpC is 73 amino acids and only 33% identical to all other G5 domains, which are separated by a 131 amino acid spacer. The remaining twelve tandem G5 domains each only have a 20 amino acid spacer, although the first G5 domain (G5 2) of the tandem group is somewhat more unique, 73 amino acids and roughly 67-60% identical; all other G5 domains (3-13) are almost completely identical, 76 amino acids and 95-82% identity. Because the first two G5 domains contain more sequence diversity than Aap, we were able to design primers to amplify G5 domains 1-13; however, due to the high identity in G5 domains 3-13 and therefore a propensity for polymerase slippage, we were only able to produce a PCR product through G5 domain five.
This construct was also inserted into the pHis/MBP-DEST expression vector and verified by nucleotide sequencing and molecular weight approximation by gel electrophoresis and will be called StsG5 hereafter.

**Solution characterization of Brpt and StsG5 proteins** - All of the Brpt and StsG5 constructs were solubly expressed in an IPTG-inducible *E. coli* BLR(DE3) system and successfully purified with molecular weights that are consistent with the predicted number of G5 domains by nucleotide analysis ([Figures 4A and B](#)). The N-terminal MBP fusion tag was proteolytically cleaved after purification for all Brpt constructs. The His-MBP/StsG5 construct was not susceptible to proteolytic cleavage and therefore also contains the His-MBP fusion tag. To ensure that the proteins maintained the predicted fold after recombinant expression, far-UV circular dichroism (CD) was performed ([Figures 5A, B, and Table 1](#)). Our lab previously characterized a single G5 domain and the half-repeat cap (called Brpt1.5) and determined the secondary structure content to be primarily β-strand (4% helix, 39% β-strand, 24% turn, and 32% coil) [3]. The high sequence identity of each B-repeat predicts that the secondary structure is maintained and constructs containing multiple B-repeats would amplify this signal. Far-UV CD data of the new Brpt3.5, 5.5, 7.5, and 9.5 constructs revealed a similar,
### Table 1. Secondary structure content of G5 domain proteins.

<table>
<thead>
<tr>
<th>Protein</th>
<th>% Helix</th>
<th>% β-Strand</th>
<th>% Turn</th>
<th>% Coil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brpt3.5</td>
<td>5%</td>
<td>38%</td>
<td>22%</td>
<td>35%</td>
</tr>
<tr>
<td>Brpt5.5</td>
<td>6%</td>
<td>35%</td>
<td>24%</td>
<td>34%</td>
</tr>
<tr>
<td>Brpt7.5</td>
<td>5%</td>
<td>36%</td>
<td>24%</td>
<td>33%</td>
</tr>
<tr>
<td>Brpt9.5</td>
<td>6%</td>
<td>35%</td>
<td>23%</td>
<td>35%</td>
</tr>
<tr>
<td>His-MBP/Brpt5.5</td>
<td>11%</td>
<td>35%</td>
<td>23%</td>
<td>31%</td>
</tr>
<tr>
<td>His-MBP/StsG5</td>
<td>15%</td>
<td>31%</td>
<td>24%</td>
<td>29%</td>
</tr>
</tbody>
</table>

**Figure 5.** Far-UV circular dichroism spectra of recombinantly expressed G5 domain proteins.

A. Aap Brpt constructs 3.5 (blue), 5.5 (red), 7.5 (green), and 9.5 (violet).

B. Comparison of His-MBP/Brpt5.5 (red) with His-MBP/StsG5 (blue).

---

5 domain et al.  B.
high β-strand secondary structure content to the previously characterized Brpt1.5 (Figure 5A and Table 1). The G5 domains of zmpC were also predicted to be similar in secondary structure to Brpt G5 domains due to the similar amino acid residue content contained within this region. The precise secondary structure content of His-MBP/StsG5 is difficult to determine with the MBP tag remaining (which has high helical content), however, the CD data does align well with the His-MBP/Brpt5.5 construct (Figure 5B and Table 1); this suggests that His-MBP/StsG5 does have a similar secondary structure to the Brpt proteins.

To characterize the solution behavior of the Brpt and StsG5 proteins, we performed analytical ultracentrifugation (AUC) sedimentation velocity experiments. Remarkably, we observed similar sedimentation coefficients for all of the larger Brpt constructs (Figure 6A and B), despite the differences in molecular weight from increasing the number of Brpt domains. The primary difference between the constructs is the total length of the protein, which greatly affects the frictional ratio, \( f/f_0 \). A single Brpt has been shown to be an elongated protein in solution [3]. The addition of Brpt domains to a single construct would further elongate the protein, which is supported by the increase in frictional ratio as the number of Brpt domains in a single construct increases (Figure 6B). Brpt9.5 does have a slightly lower frictional ratio than Brpt7.5 that would break this trend; although this could be a result of the protein being so elongated that it begins to double over or curl at the ends during sedimentation. His-MBP/StsG5 has a slightly higher sedimentation coefficient (~3S) than the Brpt proteins, although it is more globular in shape (Figure 6A and B). There is also an additional small peak that sediments around 1.8S that is likely to be self-cleaved MBP based on the molecular weight approximation.
Figure 6. Solution characterization of recombinantly expressed G5 domain proteins by analytical ultracentrifugation. A. Sedimentation velocity analysis of Aap Brpt3.5 (blue), Brpt5.5 (red), Brpt7.5 (green), Brpt9.5 (violet), and His-MBP/StsG5 (cyan). B. Table of respective sedimentation coefficients (s) and frictional ratios (f/f₀).

<table>
<thead>
<tr>
<th></th>
<th>s</th>
<th>f/f₀</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brpt3.5</td>
<td>1.96</td>
<td>2.25</td>
</tr>
<tr>
<td>Brpt5.5</td>
<td>2.05</td>
<td>2.58</td>
</tr>
<tr>
<td>Brpt7.5</td>
<td>2.08</td>
<td>2.91</td>
</tr>
<tr>
<td>Brpt9.5</td>
<td>2.16</td>
<td>2.79</td>
</tr>
<tr>
<td>His-MBP/StsG5</td>
<td>3.12</td>
<td>2.10</td>
</tr>
</tbody>
</table>
The fact that His-MBP/StsG5 is more globular is slightly surprising since the secondary structure content of StsG5 is relatively similar to the Brpt proteins, but this does indicate that StsG5 is a unique G5 domain protein.

**Zinc-mediated Brpt self-assembly is influenced by the number of tandem Brpt domains** – The purpose of the multiple, tandem G5 domain organization of Aap for protein function has been of considerable interest after a single Brpt domain (Brpt1.5) was found to dimerize in the presence of zinc [3]. Why are twelve, tandem G5 domains required for protein function? The primary hypothesis was to facilitate protein association at physiological concentrations of zinc, which is 15 – 20 µM, while the EC$_{50}$ of Brpt1.5 self-association is 5.4 mM zinc [3]. The zinc requirement was further reduced to 3.7 mM by increasing the number of G5 domain-containing B-repeats to two (Brpt2.5), which supports the original hypothesis that Brpt self-association is a cooperative binding event. Based on this hypothesis and the zinc required for the association of shorter Brpt constructs, we can extrapolate the concentration of zinc required for the assembly of longer Brpt constructs. Furthermore, the ability for longer Brpt constructs to form a variety of higher order oligomeric species rather than simply dimerize is likely since each Brpt is almost identical, providing multiple potential binding conformations for association.

To test these hypotheses, we performed AUC sedimentation velocity experiments with the larger Brpt constructs at three different concentrations of zinc (**Figure 7**). 1.5 mM, 3 mM, and 6 mM zinc was added directly to Brpt3.5 (~12 µM) and immediately loaded into AUC cells for sedimentation velocity analysis at 20 °C (**Figure 7A**). The lowest concentration of zinc, 1.5 mM, did not change the sedimentation of Brpt3.5, indicating that the concentration of zinc was
Figure 7. Solution characterization of Aap Brpt extended constructs in the presence of zinc. A. Sedimentation velocity analysis of Brpt3.5 (blue, *normalized) and in the presence of 1.5 mM (red), 3mM (green), or 6 mM zinc (violet). B. Comparison of normalized Brpt3.5 (blue), 5.5/MBP (red), 7.5 (green), and 9.5 (violet) in the presence of 3 mM zinc. The blue shading contains all small oligomer complexes between dimer and trimer for all constructs. Sedimentation of the apo-Brpt constructs is represented by the black dashed line at 2S.
too low to induce assembly. However, there was a clear monomer to oligomer species shift in the sample incubated with 3 mM zinc. There is a small peak remaining that has a sedimentation coefficient consistent with monomer Brpt3.5, although the majority of species are assembled into approximate dimer or trimer species. There is also an obvious shoulder around 5S that suggests there are additional oligomeric species that could not be resolved. The addition of 6 mM zinc induced a complete monomer to oligomeric transition, producing a major dimeric peak around 4S, an additional peak at 10S that is an approximate tetramer, and increasingly large species greater than 20S that cannot be represented in this analysis. While the zinc requirements to induce oligomerization were not well defined in this experiment, it is evident that the formation of higher order species by Brpt is possible.

Similar experiments were performed with Brpt5.5, 7.5, and 9.5 at three different concentrations of zinc; however, it became quickly apparent that these elongated constructs did not behave as typical proteins in the presence of zinc. Therefore, to provide a more simplified assessment of the effect of longer Brpt constructs on oligomerization, a single concentration of zinc (3 mM) was selected for all Brpt proteins for AUC analysis (Figure 7B). In addition, the addition of the MBP fusion tag was found to enhance resolution of oligomeric complexes, and was therefore left on the Brpt5.5 for this experiment only (Brpt5.5/MBP). Using 3mM ZnCl$_2$, there is a clear shift from monomer to oligomeric in all Brpt proteins; the data has been normalized in order to see the variety of oligomeric species in more detail. The longer Brpt constructs (Brpt5.5/MBP, 7.5, 9.5) have predominately larger oligomeric species greater than 6S, whereas Brpt3.5 primarily remains within the 2-5S boundaries that are approximate dimeric or trimeric species. The Brpt7.5 data does not accurately represent all oligomeric species because the majority of species sedimented early because they were large aggregates. This is also true.
for Brpt9.5, although more mid-sized oligomers (approximate tetramer) were represented in the data. Because the majority of protein visually aggregated upon the addition of zinc, particularly the Brpt7.5 and Brpt9.5 constructs, it is difficult to assess the progressive order of Brpt assembly into higher order oligomers. However, there does appear to be a consistent trend that suggests a dimer to tetramer assembly method that would potentially serve as “building blocks” for larger oligomers. This is particularly evident from the Brpt5.5/MBP data because the predominant species are dimer, tetramer, and octamer. Another interesting component is that the overall frictional ratio for all of the longer Brpt constructs decreased to ~2, suggesting that these longer repeats can potentially fold-over on themselves to form an assembly complex. Regardless, this data demonstrates that Brpt assembly is influenced by both the concentration of zinc and the total number of consecutive Brpt domains within a single protein.

*SsG5 self-assembles in the presence of zinc* – The G5 domains of Aap have been shown to self-associate in the presence of zinc and this has shown to be critical for biofilm formation in *S. epidermidis*. Therefore, in order to determine if the G5 domain is a universal metal-binding adhesion module, additional diverse G5 domains must be characterized. The G5 domains of the *S. sanguinis* zmpC were selected because this protein contains a unique G5 domain arrangement; the first G5 domain has very low sequence identity to the other tandem G5 domains that are found further downstream, whereas Aap contains only tandem and almost identical G5 domains. In order to determine if His-MBP/StsG5 can bind and self-assemble in the presence of zinc, 1 mM or 2 mM zinc was added to 4 µM His-MBP/StsG5 for analysis by AUC sedimentation velocity. Upon the addition of 1 mM zinc, there is a clear shift in the sedimentation profile of monomer His-MBP/StsG5 to a higher order species that sediments around 5S (*Figure 8*), only
4% of monomer remaining. The approximate molecular weight calculated for the 5S peak suggests there is a mixture of a highly elongated dimer and trimer species, as determined by an increase in the frictional ratio (3.41) from the zinc-unbound His-MBP/StsG5 (2.1). There was no further oligomerization by increasing the concentration of zinc to 2 mM, suggesting that the trimer is the highest order and stable oligomer formed by His-MBP/StsG5. Regardless, this data strongly supports that the G5 domain is a metal-binding adhesion module.

![Figure 8. Solution characterization of His-MBP/StsG5 in the presence of zinc.](image)

Analytical ultracentrifugation sedimentation velocity performed at 20 °C. Sedimentation coefficient distribution (c(s,*)) plot analysis generated by Sedfit. His-MBP/StsG5 (blue) or in the presence of 1mM (red) or 2mM (green) zinc. Note that there is not change in oligomerization by doubling the concentration of zinc. In the presence of zinc, the small peak around 3S is approximate monomer, whereas the 5S is an approximate trimer.
Discussion

The G5 domain motif is found across diverse bacterial species (Figure 2), although the majority of species share commonalities in their ability to colonize harsh environments and form biofilms. Based on the gene annotation and characterization of the proteins within the Pfam database, it seems that not all proteins with G5 domains are involved in biofilm formation. Together, this suggests a different overall role for the G5 domain, such as a universal metal-binding adhesion module. There are currently only two other G5 domain-containing proteins that have been experimentally characterized: SasG, a homologue of Aap found in S. aureus [71, 74], and RpfB of Mycobacterium tuberculosis [118]. SasG has been shown to have similar zinc-binding properties and function to Aap, which is not surprising since these proteins have high sequence identity [74]. The zinc-binding properties of RpfB have never been explored, and therefore the only known G5 proteins to bind and assemble in the presence of zinc are highly similar in both sequence and function. zmpC was selected as a candidate because it contains a unique and interesting G5 domain architecture and is fairly diverse from Aap in sequence (Figures 1 and 3). In addition, while Streptococcus sanguinis is known to form biofilms, the peptidase domains of zmpC suggest an alternative physiological function.

Despite differences in the primary G5 domain amino acid sequences of Aap and zmpC, all of these protein constructs (Brpt3.5, 5.5, 7.5, 9.5, and His-MBP/StsG5) contain a large amount of β-strand content (Figure 5), which correlates with the crystal structure of RpfB [118] and the biophysical characterization of Brpt1.5 from Aap [3]. The solution characterization of StsG5 construct does, however, suggest some differences in the overall protein structure from the Brpt G5 domains of Aap. The higher helical content by CD and decreased frictional ratio by AUC suggest that His-MBP/StsG5 is a less elongated and more globular protein than Brpt G5.
domains. This could be largely due to the long, 131 amino acid spacer region between G5 domains 1 and 2 of His-MBP/StsG5 that does not occur in the completely tandem G5 domains of the Brpt constructs, which only have a fifty amino acid spacer region.

In the presence of zinc, we have shown that the diverse G5 domains of both Brpt and StsG5 self-assemble (Figures 7 and 8). However, the AUC data clearly suggests that the conformation of these assembled complexes is quite different. The assembly of Brpt constructs demonstrates a correlation between the number of consecutive Brpt domains and the size of the assembled complexes formed, since increasing the number of Brpt domains increases the potential to form extremely large protein complexes in the presence of zinc (Figure 7B). These complexes are also highly elongated and on the order of megaDalton molecular weights in some cases, which suggests that these B-repeat complexes could actually be a type of ordered protein fiber and not just a protein complex. Describing the physical nature of these large protein complexes will be the focus of future studies.

His-MBP/StsG5 does assemble into a protein complex in the presence of zinc (Figure 8), although it appears to be limited to a trimer as its highest order species. This result cannot be attributed to a lack of tandem G5 domains, as seen in the Brpt1.5 construct previously [3]. This is because there are approximately four intact tandem G5 domains within this construct, despite the unusually long spacer after the first G5 domain. Both of the Brpt3.5 and Brpt5.5 constructs, which are comparable by their number of G5 domains, demonstrated the ability to assemble into protein complexes that are larger than a trimer (Figure 7B). One component that could potentially limit the His-MBP/StsG5 complex assembly is the spacer region between tandem G5 domains. Brpt G5 domains have a 50 amino acid spacer, whereas His-MBP/StsG5 has only 20. Shortening of this spacer sequence could provide steric restraints that would limit complex
formation to a single type of oligomer. Future experiments studying the spacer region of G5 domains are required to determine their importance in protein assembly. Another interesting feature of the His-MBP/StsG5 construct is that assembly is not affected by increasing the concentration of zinc (Figure 8). The Brpt constructs demonstrate a distinct dependence on the zinc concentration that determines both the order of species complexity and the percentage of total protein that is assembled (Figure 7). However, because there are additional tandem G5 domains within zmpC, it is possible that extending the His-MBP/StsG5 construct to include these domains could produce additional, higher order species. Another possibility is that 1 mM zinc could be a saturating concentration for protein assembly and would explain why there is no affect upon increasing the zinc to 2 mM. Constructs including more zmpC G5 domains and further studies using lower concentrations of zinc would need to be performed to make these hypotheses conclusive.

While these results provide compelling evidence that G5 domains serve as zinc-binding adhesion models, there are still so many other G5 domains from more diverse bacterial species and protein architectures. Both S. epidermidis and S. sanguinis are pathogenic species where zinc is very bioavailable. Other species that live in deep-sea thermal vents may utilize other metals, such as iron and copper, which would be more bioavailable in this environment [226]. Further biophysical analysis of other unique G5 domain containing proteins is needed to accurately determine the functions of G5 domains for this diverse group of bacteria.

Acknowledgements

I would like to thank Greg Bick, a graduate rotation student, who helped generate the data for the His-MBP/StsG5 construct.
Chapter 3

A Literature Review of
Functional Amyloid and Amyloidogenesis
Introduction

“Amyloid” has long been considered an aberrant and aggregated state of protein folding and therefore only associated with the pathogenesis of human disease. Many amyloid-related disorders are neurodegenerative diseases, such as Alzheimer’s, Parkinson’s, Huntington’s, and Creutzfeld-Jacob; although the pathology is not limited to only affecting brain function as other proteins throughout the body can contribute to amyloid-related disease states, such as type II diabetes. Amyloid disease states typically arise from the generation of extracellular amyloid plaques or intracellular inclusion bodies, and therefore depletion of the properly-folded, functional protein [119]. However, the close association of “amyloid” and “disease” is rapidly changing. The amyloid field has recently exploded with an abundance of new publications detailing the mechanisms of assembly, influential environmental factors, and more importantly, the discovery of a new class of amyloid proteins that provide a physiological function in the fibrous state, called functional amyloid. These “functional amyloid” proteins are required for cellular function in their fibrous or aggregated state and are not considered misfolded or disease causing entities. The majority of these functional amyloid proteins are found in lower order species, such as bacteria, fungi, and insects, and it has been hypothesized that amyloid fiber formation is an ancient protein folding mechanism to enhance structural stability [120]. However, as research progresses, we are learning that not all amyloidogenic formations are the pathogenic and disease-causing state in humans.
I. Overview of Amyloid Characteristics

What is “Amyloid”?

“Amyloid” specifically refers to the structural conformation of the protein fiber resulting from the assembly of many copies of a single (monomeric) protein unit. All amyloid fibers are characterized by the “cross-β” structure that is generated by stacking of the subunit β-strands perpendicular to the long axis of the fiber, with backbone hydrogen bonds stabilizing the fiber in parallel [121-123] (Figure 1). To demonstrate this generalized fold of the amyloid cross-β structure, antibodies generated from one type of amyloid fiber can often be used to recognize any other amyloid fibers, independent of the critical amino acids forming the fibril structure [124].

While the three-dimensional structure of amyloid fibers are conserved, almost every protein has the potential to form amyloid fibers; there is, however, a higher amyloidogenic tendency for

**Figure 1. Diagram of classical amyloid “cross – β” structure.**
A. Model of stacked, parallel β-strands, represented by blue arrows. The channel that runs down the center of the fiber, the core containing the amino acid side chains (R), is shown in B. Figure is used with permission from Biancalana et al, Biochimica et Biophysica Acta, 2010 [1].
proteins with high \(\beta\)-sheet secondary structure content [125]. There have been many attempts to predict amyloidogenic sequences within proteins [126, 127] but there is currently no common consensus sequence or residue preference that universally drives all amyloid fiber formation. Because there is no consensus sequence for this folding mechanism, it is suggested that amyloidogenesis is as primitive as the creation of proteins themselves [120]. The variability of residues within the fibril core produces a library of amyloid fibers that each have unique chemical properties, fiber morphologies, and exhibit different mechanisms for mature fiber formation.

**Chemical Properties of Amyloid Fibers**

The canonical cross-\(\beta\) structure of amyloid fibers is well known for its high degree of stability [128], in that the majority of amyloid fibers are resistant to most chemical denaturants, heat, and sodium dodecyl sulfate (SDS). However, formic acid and sonication are often used to disintegrate amyloid fibers into their oligomeric subunits [124, 129]. Amyloid fibers are specifically identified by their ability to induce thioflavin T fluorescence and congo red birefringence [130]. Thioflavin T is a small molecule that specifically binds within the core of the cross-\(\beta\) structure (Figure 1B) [1, 131]. When bound in this specific confirmation, thioflavin T exhibits fluorescence when excited between 385 – 450 nm and produces an emission maxima between 445 – 482 nm [131]. The exact excitation and emission of thioflavin T fluorescence is unique to each amyloid fiber and is dependent on the orientation of thioflavin T in the fibril core [1]. Congo red dye is also used to identify protein fibers as amyloid; however, congo red binding is not a specific interaction and therefore not completely reliable for identification [132]. Congo red can be used in a microscope assay [130], visualizing a red to green birefringence, or
spectrophotometrically, measuring the absorbance at 403 nm (un-bound congo red) and 543 nm (fiber bound congo red) [132]. Earlier research studies defined amyloid fibers by their ability to bind these small molecules; however, more recent studies have demonstrated protein fibers that display other amyloid properties but do not produce robust thioflavin T fluorescence [133], demonstrating the molecular diversity within amyloid fibers.

**Amyloid Fiber Morphologies**

Another key technique used to identify amyloid fibers is transmission electron microscopy. While there is no limitation to the fibril length, the majority of traditional amyloid fibers are smooth in appearance and approximately 100 nm in width [4] (Figure 2A and B). However, recent studies to further understand environmental factors that influence amyloid fiber formation have demonstrated more non-traditional amyloid fiber morphologies that still maintain amyloid properties [5] (Figure 2C and D). It has also been shown that a single protein can produce multiple fiber types, even under identical solution conditions [134, 135]. This phenomenon demonstrates the molecular complexity of amyloid fiber formation that still remains

![Figure 2. Amyloid fiber morphologies](image)

Classical amyloid fibers (A and B) are long, smooth, intertwined and un-branched compared to the non-traditional, more rough and branched fibers (C and D). Images used with permission from Kodali et al, J. Molec. Bio., 2010 [5].
quite elusive due to multiple environmental factors that can potentially influence fiber assembly. The specific environmental conditions of the resulting fiber morphology can determine the pathogenic phenotype [136]. Therefore, determining the resulting fiber morphology from physiologically relevant environmental factors, such as temperature, metals, pH, osmolarity, and time [5, 137], is of the utmost importance for this field for creating therapeutic treatments for diseases mediated by amyloidosis.

*Amyloid Fiber Formation*

The precise molecular mechanism of fiber formation will differ between specific amyloid morphologies; however, there is still a generalized, multi-stage assembly mechanism that is universally accepted for the production of mature amyloid fibers. Because amyloid fibers are produced from a pool of a single protein or short peptides, the assembly mechanism is highly driven by a protein-protein association equilibrium, collectively called “nucleation-dependent polymerization” ([Figure 3](#)) [138]. This mechanism is defined by an initial lag phase that produces critical, “nucleating” oligomers that are the building blocks for mature fiber assembly [139, 140]. These nucleating complexes are morphologically different from mature fibers by microscopy, with a typically spherical or colloidal shape [138].

Nucleation occurs through a controlled aggregation process, although not all aggregates that are formed are in the correct conformation to produce fibers [141]. These pre-fibril aggregates are typically considered unstable and highly dynamic and are in exchange with the pool of unassociated monomers [141]. Therefore, nucleus formation is typically the rate limiting step in fiber formation because the nucleating species formed do not necessarily proceed to elongation [142]. The rate of oligomeric assembly, and therefore the lag phase, is dependent
on the concentration of soluble, monomeric protein and the environment temperature, since higher, physiological temperatures decrease the lag phase [5]. In addition, it is suspected that there is a critical number or size for the nucleating oligomeric complex required to proceed to elongation [140]. When added to monomeric protein in solution, these oligomeric complexes have the ability to rapidly “seed” new fiber growth and significantly decrease the lag phase [143] (Figure 3).

**Figure 3. Kinetic diagram of nucleation-dependent amyloid fiber assembly.** Lag phase occurs as monomeric protein assembles into oligomers that successfully produce critical nuclei. However, once a critical nucleus is formed, amyloid fiber assembly occurs rapidly (elongation) (black line). Seeding with critical nuclei can reduce lag phase time by providing a template for monomers to form the nuclei (red line). Despite the highly ordered and stable nature of mature amyloid fibers, there is still a dynamic equilibrium with un-assembled monomer (red dashed line). Diagram used with permission from Eichner and Radford, *Molec. Cell*, 2011 [4].
Once the critical nucleus is formed, elongation may occur via rapid assembly of these oligomeric complexes into amyloid fibers, which is marked by a large increase in thioflavin T fluorescence [130]. Elongation can also be measured by the increase of β-stand secondary structure content by circular dichroism, although proteins with high native β-strand content cannot be monitored accurately using this method [124]. There have been several mechanisms of elongation that have been proposed, in which the nucleating oligomers can i) directly stack units vertically, ii) stack protein chains directly in a cross-β spine, iii) associate through domain-swapping, or iv) stack vertically as a cross-β spine with domain swapping (globular protein domains are not included in the fibril core) [123]. Rather than directly producing mature amyloid fibers, some oligomers will instead first form proto-fibril intermediates that are fibrous in nature but do not yet conform to the cross-β structure, eventually undergoing structural rearrangements to generate mature amyloid fibers [144]. There is also new evidence that suggests a non-nucleated pathway of amyloid fiber formation that rapidly produces unconventional “worm-like” fibers; however, these fibers are not as stable [145]. Some amyloid proteins form fibers via refolding of the native secondary fold [146] or are natively unstructured and are therefore susceptible to adopting an amyloid confirmation [147, 148]. Despite the high degree of stability of mature amyloid fibers, recent studies have shown that mature amyloid fibers are still in a dynamic equilibrium with their soluble, monomeric counterparts [149-151] (Figure 3), and that maintaining some unassembled monomer is required to maintain protein fibers [152]. Mature amyloid fibers have also been shown to break and rejoin, which supports the reversible and dynamic nature of amyloid fibers [153]. While amyloid fiber formation does occur spontaneously, environmental factors often act as catalysts to enhance the kinetics of formation and therefore also influence the resulting fiber morphology [5].
Amyloid fiber formation is a controlled aggregation process that is driven by many of the same chemical forces that influence secondary structure folding. This is primarily because the fibril core is stabilized by amino acid side chain contacts along the long fiber axis [6] and therefore is sensitive to changes within the local chemical environment, such as pH and salt content [145]. Low pH has been shown to induce amyloid fiber formation by causing protonation of acidic amino acid residues and allowing charge repulsion to drive assembly [142]. Similarly, the local salt concentration will influence the hydrophobic interactions between amino acid contacts [154].

In addition to these solution properties, the presence of other charged molecules, such as metal ions, can influence fiber assembly as well [155]. Copper and zinc have particularly been shown to influence both amyloidogenesis and fiber pathogenicity in a number of different amyloid proteins such as amyloid-β [156, 157], prions [158], β2-microglobulin [159], α-synuclein [160], and the islet amyloid peptide [161]. It is suggested that metal ions bind amino acid residues within the protein monomer, typically acidic residues (glutamic or aspartic acid) or histidine, initiating a structural confirmation that supports fibrillization [156]. Amyloid fiber formation is often a very slow process that occurs on a time scale of days, months, or even years. Therefore, it is not surprising that external environmental factors, such as metal ions, are utilized as catalysts to increase the rate of fiber assembly [162].

**Amyloid Structure**

The transition between nucleation and elongation is still rather elusive because of the difficulties in studying the structure of both the monomeric and fibrous protein states. However, the differential packing of these critical oligomers during fiber assembly is believed to give rise
to fiber morphologies that cause disease [133-135, 163]. Therefore, it is of great interest to determine the structures of these protein states in order to understand why some amyloid fibers cause disease and other morphologies of the same protein do not [164]. The first method that was used to characterize the “cross-β” amyloid structure was x-ray diffraction [165], the characteristic diffraction pattern showing reflections at 4.7 and 10 Å that represent the fibril core dimensions (Figure 4) [166]. While fiber diffraction determines the structural dimensions of a fiber, it does not provide information on the specific amino acid contacts of the fibril core.

![Fiber diffraction image of an amyloid fiber.](image)

**Figure 4. Fiber diffraction image of an amyloid fiber.** Image used with permission from Eichner and Radford, *Molec. Cell*, 2011 [4].

Therefore, other methods, such as cryo-electron microscopy [146, 167] and more recently solid state NMR [168] and x-ray crystallography [6, 169, 170] have been used in conjunction with the fiber diffraction data to generate atomic models of amino acid residue packing within the fibril core. These structural studies have revealed that while all amyloid fibers share a cross-β structural fold, there is great diversity in how these β-strands are arranged within the fiber (Figure 5). Using structures generated by microcrystallography of amyloidic peptide sequences, Sawaya et al demonstrated eight unique “steric zipper” confirmations within the residues that compose the fibril core [6]. Another model generated by NMR proposes a β-solenoid confirmation for the fibril core of the fungal prion protein HETs(218-289) [168]. In addition, there have been other
atomic models for proteins that form amyloid fibers through an accordion-like fold called “super β-pleated sheets” [171, 172] or assemble from stacking of β-helical protofibers [173]. While all of these studies have provided huge advancements to the amyloid field, most have utilized non-physiological, short peptides and therefore may not completely represent pathological amyloid.

Figure 5. Structural models demonstrating the diversity of residues that compose the fibril core of the amyloid peptides tau, prion, amyloid-β, and insulin. The critical residues that make contacts within the core are labeled. Models were used with permission from Sawaya et al, Nature 2007 [6].
II. Functional Amyloid Proteins

Defining “Functional” Amyloid

Proteins that formed amyloid structures were originally discovered in proteinaceous plaques and were therefore associated as an artifact of protein misfolding. However, amyloid fibers have since been discovered just in the past decade in lower organisms such as bacteria [174], fungi [175], and insects [176]. This changed the idea that amyloid fibers simply caused disease, but instead were used for cellular function and therefore could be a “functional” amyloid fiber [177]. In addition, the oligomeric species, rather than the fibrous form, of amyloid proteins have actually been found to be the pathogenic component in some diseases caused by amyloidosis [178]. Lower organisms originally utilized amyloid fibers for their stability, despite the risk imposed by toxicity from protein misfolding [119, 177]. Functional amyloid proteins have also recently been found in mammals, such as the Pmel17 protein, which forms amyloid fibers to assist in the deposition of melanin in most eukaryotes [179-181], and are also involved in multiple aspects for human hemostasis [182, 183]. As more proteins are being found to form functional amyloid fibers, there is already a large group of bacteria that have been established to utilize amyloid fibers for biofilm formation [184].

Naturally Occurring Amyloid Fibers in Bacterial Biofilms

Bacterial biofilms have long been associated as a persistent state of long term stability and protection from environmental stresses; it is therefore not surprising that the proteins involved in the biofilm structure are capable of forming amyloid fibers. Interestingly, the majority of bacterial species that have been identified to produce amyloid fibers are capable of
forming biofilms; gram-positive and gram-negative as well as environmental and pathogenic species have all been identified as amyloid formers [184]. The diversity of these species reinforces the hypothesis for functional amyloid, in which amyloid is a type of protein quaternary structure utilized for its stability and protection rather than a form of protein toxic aggregation. A study from 2007 tested water samples from four geographically and environmentally different habitats and found that at least 5 – 45% of all bacterial species were amyloid-positive by amyloid-specific antibodies and thioflavin T [184]. While the proteins responsible for the formation of amyloid fibers have not been specifically identified in these environmental species, hydrophobic proteins called chaplins from the soil species *Streptomyces coelicolor* have been more thoroughly characterized [175]. Chaplins form a hydrophobic matrix via amyloid fiber assembly that reduces the surface tension of water and allows aerial hyphae to grow [175].

Curli, the biofilm-associated matrix protein from *Enterobacteriaceae*, form filamentous amyloid fibers called fimbriae or pili [129, 174]. Unlike the canonical folding of disease-causing amyloid proteins, curli have a unique amyloid assembly mechanism that requires other accessory proteins, CsgB and F, to mediate the nucleation and folding events, respectively [185, 186]. However, the requirement of accessory proteins helps control amyloid formation so that it can only occur on the cell surface or extracellularly [185]. In addition to providing the primary biofilm structural component, curli are also important for adhesion to abiotic surfaces and host proteins [187, 188]. *Bacillus subtilis* also utilizes amyloid fibers formed by the protein TasA for its biofilm extracellular matrix [189]. Similar to curli fimbriae, TasA also requires an accessory protein to mediate amyloidogenesis as well as its attachment to the cell wall [190]. FapC, recently discovered in the environmental species *Pseudomonas fluorescens*, also forms
“amyloid-like fimbriae” that are required for biofilm formation and have high protein sequence identity to the fimbriae of *Escherichia coli* and *Salmonella sp.* [191]. The *fap* operon also contains genes that are hypothesized to be nucleator proteins required for amyloid formation, analogous to the Csg proteins required for curli amyloid formation [191]. FapC is conserved in other pseudomonas species, such as the opportunistic pathogen *P. aeruginosa* [191], demonstrating a linkage between environmental and pathogenic colonization.
Chapter 4

The Biofilm Adhesion Protein Aap from *Staphylococcus epidermidis* Forms Zinc-Dependent Amyloid Fibers*


Author Contributions: SLJ performed all the experiments. SLJ, PS, and ABH analyzed the data. SLJ, PS, and DGC contributed materials and analysis tools. SLJ and ABH wrote the paper.
Abstract

The skin-colonizing, commensal bacterium *Staphylococcus epidermidis* has recently emerged as a leading cause of hospital-acquired and device-related infections. The primary determinant for *S. epidermidis* pathogenesis is its ability to form biofilms, which are multi-layered, surface-adherent bacterial accumulations that show remarkable resistance to chemical and physical stresses. Accumulation-associated protein (Aap) from *S. epidermidis* and its *S. aureus* homolog SasG have been shown to be necessary and sufficient for mature biofilm formation. These proteins have a repetitive domain architecture, containing up to 17 tandem B-repeats; the presence of at least five tandem repeats has been shown to be critical for *S. aureus* biofilm formation. We previously demonstrated that a single B-repeat with a C-terminal half-repeat cap dimerizes in the presence of zinc and that zinc-dependent self-association of Aap was necessary for biofilm formation. However, the zinc-dependent behavior of longer tandem B-repeat constructs remains unclear. In this study, we have characterized a biologically relevant construct containing five Aap B-repeats with the C-terminal cap using biophysical techniques and determined that it forms amyloid fibers in the presence of zinc. Thioflavin T fluorescence assays, transmission electron microscopy, and confocal fluorescence microscopy experiments were used to analyze the time- and temperature-dependence of amyloid fiber formation. We have also utilized a recently developed analytical approach to resolve multiple amyloidogenic precursors using sedimentation velocity analytical ultracentrifugation. Furthermore, we have demonstrated the presence of amyloid fibers during early stages of *S. epidermidis* biofilm formation and as a major structural component in mature biofilms by confocal microscopy. This work provides new insights into *S. epidermidis* biofilm formation and architecture that will potentially lead to new therapeutic treatments for persistent staphylococcal infections.
Introduction

*Staphylococcus epidermidis* is a critical component of the normal human flora that prevents the colonization and invasion of potentially dangerous microbial pathogens. However, *S. epidermidis* has recently emerged as a leading opportunistic pathogen due to its high prevalence on epithelial surfaces and ability to colonize prosthetic medical devices [18]. *S. epidermidis* specifically is the leading cause of nosocomial infections and device-related infections [105, 192], and is, along with its other coagulase-negative relatives, the leading cause of bacteremia [24, 193]. While *S. epidermidis* infections are typically non-aggressive, they are extremely resistant to antibiotic therapy [194]. Therefore, staphylococcal infections often require invasive treatment methods and frequently lead to chronic morbidity, mortality, and high healthcare costs [2, 7, 24]. *S. epidermidis* pathogenesis and chronic persistence is primarily associated with its ability to form a biofilm [16, 35], a multi-layered bacterial aggregation encased by an extracellular matrix [34]. Biofilm formation occurs in three stages: primary adherence to a surface followed by an intercellular accumulation event, and finally secretion of extracellular matrix components, which can include the polysaccharide poly-N-acetyl-glucosamine (PNAG) or other macromolecules such as proteins, DNA or teichoic acid, depending on the strain and growth conditions [57, 195, 196]. Biofilm formation can occur through both polysaccharide-dependent and -independent pathways, the latter mediated by protein-protein interactions [70]. The protein Aap (Accumulation-associated protein) is primarily responsible for protein-dependent intercellular accumulation of *S. epidermidis* biofilm formation [67], and is also required for the polysaccharide-dependent mechanism [81]. Rohde et al. and Corrigan et al. described a protein-based mechanism for staphylococcal biofilm formation.
that is independent of PNAG; in several strains of \textit{S. epidermidis} or \textit{S. aureus}, the protein Aap or its homolog SasG can mediate biofilm formation in the absence of polysaccharide secretion [68, 71]. Indeed, 39\% of biofilm positive clinical isolates are PNAG-negative while 90\% are positive for Aap [52, 70, 197].

\textbf{Aap} is a multi-domain protein consisting of an N-terminal export signal followed by the A-repeat region (11 degenerate, 16-amino acid repeats), a putative globular $\alpha/\beta$ domain, the B-repeat region containing five to seventeen conserved repeats of a 128-amino acid sequence, a
collagen-like, proline/glycine-rich domain (P/G-rich), and an LPXTG cell wall anchor motif, as shown in **Figure 1A**. The N-terminal portion of Aap containing the A-repeat region and α/β domain can be proteolytically cleaved to expose the B-repeat region, which then initiates bacterial accumulation into microcolonies [68]. Each B-repeat (Brpt) contains an 78-amino acid G5 domain and a 50-amino acid spacer motif; the B-repeat sequences in Aap are highly conserved, with 83-100% sequence identity (**Figure 1D**). The final repeat in the B-repeat region is comprised of a single G5 domain without the spacer motif (**Figure 1A**); this C-terminal half-repeat “cap” plays a role in stabilizing the protein [3]. We have previously demonstrated that a single Brpt domain with the half-repeat cap (Brpt1.5) will self-associate in the presence of Zn\(^{2+}\) to form a dimer, leading to a model for Zn\(^{2+}\)-mediated protein-dependent intercellular accumulation between staphylococci in a nascent biofilm [3]. Recent work has shown similar Zn\(^{2+}\)-dependent self-association behavior for the B-repeat region of SasG [74]. Full-length Aap contains 5 – 17 of these nearly identical B-repeats; Corrigan et al have previously shown the importance of multiple, tandem B-repeats in the Aap homolog from *S. aureus*, SasG [71]. In that study, they demonstrated that at least five tandem B-repeats were required for *S. aureus* biofilm formation, suggesting that the biological function of SasG and, presumably, Aap relies on tandem B-repeats. Therefore, the focus of this study is to characterize a minimal, biologically relevant construct of Aap consisting of five consecutive B-repeats and the C-terminal cap (called Brpt5.5; see **Figure 1A**), and to determine the role of tandem B-repeats in *S. epidermidis* biofilm formation. In the presence of Zn\(^{2+}\), His-MBP/Brpt5.5 assembled into a range of oligomeric states, including highly elongated fibers. We show that the Zn\(^{2+}\)-induced Brpt5.5 fibers are functional amyloid fibers that assemble in a temperature- and time-dependent fashion. We apply a newly developed analytical approach to deconvolute the early-stage assembly of intermediates.
using analytical ultracentrifugation. Finally, we show that amyloid fibers form during early and late stages of *S. epidermidis* biofilm growth, suggesting the importance of Aap amyloid fiber formation in the pathogenesis of *S. epidermidis*. This is the first report to directly identify an amyloid-forming protein in staphylococcal biofilms; these findings could potentially lead to new therapeutic approaches to relieve the extensive morbidity and mortality caused by *S. epidermidis* infections.

**Figure 1D.** Sequence identity comparison of the tandem Brpt domains of Aap. *S. epidermidis* strain RP62a contains 12 tandem, B-repeat regions. Each Brpt region amino acid sequence was aligned, using the most N-terminal repeat as the reference sequence and the online ISREC-Server program LALIGN (version 2.1.30) [Program by William Pearson using algorithm by Huang and Miller, Adv. Appl. Math. (1991) 12:337-357]. Sequence alignment of each Brpt domain demonstrates a highly conserved sequence identity for each repeat with minor amino acid differences. In addition, some of the other repeat regions that are found internally are up to 98-100% identical (such as repeats 4 and 5 or 9, 10, and 11) despite sequence differences with the N-terminal reference Brpt.
Materials and Methods

Bacterial Strains and Media - *S. epidermidis* strain RP62A (ATCC 35984) was purchased directly from ATCC as a glycerol stock and was cultured in tryptic soy broth (TSB).

Expression Construct Generation - The Brpt5.5 construct (amino acids 1504 – 2223) of Aap (NCBI AAW53239.1) was PCR amplified from RP62A genomic DNA and inserted into the expression vector pHisMBP-DEST (kindly provided by Artem Evdokimov) using Gateway technology, which adds an N-terminal hexahistidine-maltose binding protein (His-MBP) fusion to the Brpt5.5 construct, with an intervening tobacco etch virus protease site. The plasmid was then transformed into the *E. coli* expression cell line BLR(DE3) (Novagen).

Protein Expression and Purification - 1 L cultures were inoculated with 10 ml of His-MBP-Brpt5.5/BLR(DE3) culture at an OD$_{600}$ of 0.6-0.8, and then allowed to incubate overnight with shaking at 37 °C. Protein expression was then induced using 250 µM IPTG for 6 hours at 25 °C. The cells were then harvested, re-suspended, frozen and thawed prior to lysis by french press. The cell lysate was centrifuged, and the soluble fraction was decanted onto a nickel-NTA gravity column. The His-MBP/Brpt5.5 protein was eluted by imidazole step gradient. Fractions confirmed by SDS gel to contain Brpt5.5 were then further purified by anion exchange using a 5 ml anionQ fast-flow column (GE Healthcare) followed by size-exclusion chromatography using a superdex 200 column. The presence of the His-MBP fusion tag was shown not to alter Brpt5.5 assembly when compared to cleaved protein and was therefore left attached to enhance protein
resolution during complex formation studies with zinc, except for initial CD and AUC experiments (Figure 1).

Circular Dichroism - Protein samples were dialyzed into 10 mM Tris pH 7.4 and 150 mM NaF. Far-UV CD spectra were obtained using an Aviv 215 spectrophotometer. The concentration of cleaved Brpt5.5 (fusion tag removed) was determined using the molar extinction coefficient of 19,370 M$^{-1}$cm$^{-1}$, as calculated using the online server ProtParam (web.expasy.org/protparam), and the concentration of His-MBP/Brpt5.5 (uncleaved) was determined using the molar extinction coefficient of 87,210 M$^{-1}$cm$^{-1}$ calculated by ProtParam. An additional sample of His-MBP/Brpt5.5 was treated with 10% formic acid (FA) to depolymerize amyloid fibers, followed by dialysis of the sample into 10 mM Tris pH 7.4 and 150 mM NaF. Data were analyzed using the CDSSTR program on the online Dichroweb server with reference set 4 (http://dichroweb.cryst.bbk.ac.uk) [116].

Analytical Ultracentrifugation - Experiments were performed with a Beckman XL-I analytical ultracentrifuge using absorbance optics at 280 nm. Sedimentation velocity experiments were performed at 36,000 rpm at 20 °C with and without 3 mM ZnCl$_2$ (Figure 2) or at 25 °C and 37 °C with or without 500 µM or 1 mM ZnCl$_2$ as indicated (Figure 7). Data were analyzed using the Sedfit software [117] and the c(s) (Figures 1 and 2) or c(s,ff$_0$) models (Figure 7). The c(s,ff$_0$) model describes the sedimentation behavior of the species in solution as a two-dimensional distribution based on both sedimentation coefficient and frictional ratio, which allows resolution of multiple species of widely differing size and shape sedimenting at the same sedimentation coefficients [198]. Parameters of buffer density and viscosity and the partial
specific volume of His-MBP/Brpt5.5 were calculated using SEDNTERP [199] at all relevant experimental temperatures. The data described in Figure 7B and 7C were analyzed using several different models in Sedfit; the c(s,ff₀) model yielded fits with the lowest value for the summed square of the residuals (SSR). The standard c(s) model fit gave an SSR value of 0.5140; the best-fit worm-like chain model gave an SSR value of 0.5102; the c(s, ff₀) model gave an SSR values of 0.5031 when setting the frictional ratio limits from 1 to 3; and the c(s, ff₀) model gave an SSR values of 0.5029 when setting the frictional ratio limits from 1 to 5.

Transmission Electron Microscopy - 5 µl samples of 10 µM His-MBP/Brpt5.5 incubated with 500 µM ZnCl₂ were applied to 200 mesh formvar carbon/copper grids for 2 minutes, washed with diH₂O, stained by 1% uranyl acetate drop-wise for 30 seconds, and washed a second time with diH₂O. Samples were then dried for 1 hour prior to viewing on a Hitachi 7600 transmission electron microscope at an accelerating voltage of 80 KV. Images were captured using an AMT 2k CCD camera.

Thioflavin T Protein Fluorescence Assay - 10 µM His-MBP/Brpt5.5 samples were treated with 10% formic acid (FA) to depolymerize amyloid fibers [129], followed by dialysis into standard buffer. The FA-treated His-MBP/Brpt5.5 was then incubated with or without 500 µM ZnCl₂ for 24 hours at 20 °C prior to adding thioflavin T (Sigma) to a final concentration of 10 µM. In parallel, one sample of His-MBP/Brpt5.5 was treated with 10% FA and was tested for ThT fluorescence without removing the FA (labeled “FA + MBP/Brpt5.5 + ThT” in Figure 3E). Fluorescence was measured using a Perkin Elmer LS50B Luminescence Spectrophotometer at an
excitation of 434 nm and collecting the complete emission spectrum between 450 and 600 nm. The software package FLWinLab was used to collect the data.

**HPLC Assays** - For the HPLC fiber/oligomer quantification assay, 10 µM His-MBP/Brpt5.5 samples were incubated with or without 500 µM ZnCl₂ for 1, 4, or 30 days at 20 °C or 37 °C. Samples were prepared based on a protocol previously described by O’Nuallain et al [124]. Briefly, samples were centrifuged for 1 hour at 13,500 rpm and then 250 µl supernatant was added to 250 µl of 5% acetonitrile. Samples were loaded on a C4 reverse-phase column (Phenomenex) and run on an Äkta purifier with a linear gradient of 0-95% acetonitrile over 10 column volumes. Peaks were integrated using the Unicorn software, normalizing the peak area to a standardized elution volume for monomer and oligomer peaks.

**Confocal Microscopy** - For analysis of Brpt5.5 protein fibers, 10 µM samples were incubated with 500 µM ZnCl₂ for 2, 6, 12, 18, or 24 hours prior to addition of thioflavin T (10 µM final concentration). Samples were then added to 8-well borosilicate glass slides (Nunc) and viewed using a Zeiss LSM 710 inverted microscope using an Apochromat 63x/1.40 oil DIC M27 objective for brightfield and a laser set to 458 nm with emission filters at 469 – 580 nm to measure thioflavin T fluorescence. For analysis of amyloid formation in biofilms, overnight cultures of *S. epidermidis* RP62A were diluted 1:200 and 400 µl aliquots were added to 8-well borosilicate glass slides. The slides were incubated 18 hours at 37 °C without shaking, followed by addition of thioflavin T to 10 µM as indicated. The media was aspirated and each well was washed 3x with diH₂O. The biofilms were viewed using a Zeiss LSM 510 inverted microscope.
as previously described above. Control biofilm samples were alternatively labeled with the LIVE/DEAD stain kit (Invitrogen), using 3 µl stain diluted in 1 ml of sterile saline.
Results

Solution characterization of five tandem B-repeats from Aap - In order to determine the functional relationship between tandem copies of the Aap B-repeat region and *S. epidermidis* biofilm formation, we have generated a construct containing the C-terminal five intact B-repeats. The conserved C-terminal 78 amino acid half-repeat was additionally included in this construct (called Brpt5.5); we previously demonstrated this C-terminal half-repeat to be important for protein stability [3] (Figure 1A). The design of this construct was based on the minimum number of B-repeats previously shown to support biofilm formation by SasG, a close Aap homolog from *S. aureus* [71]. The final expression construct contained an N-terminal His-tagged MBP fusion protein followed by the Brpt5.5 construct. In order to verify proper folding of the Brpt5.5 region, the fusion tag was removed and the resulting cleaved Brpt5.5 protein was analyzed by far-UV circular dichroism (CD). Each B-repeat has a highly conserved amino acid sequence (Figure 1D) and was therefore predicted to have a similar secondary structure for each repeated domain. Far-UV CD confirmed that the cleaved Brpt5.5 construct has a similar secondary structure profile to the previously characterized Brpt1.5 construct (the intact C-terminal B-repeat with the half-repeat cap) [3] (Figure 1B). The Brpt5.5 fold consists primarily of β-strand and coil (6% helix, 35% β-strand, 24% turn, and 34% coil for Brpt5.5). Analytical ultracentrifugation (AUC) sedimentation velocity experiments revealed that Brpt5.5 sediments as a 2.05 S species with a frictional ratio of 2.58, which corresponds to a highly elongated 78 kDa monomer (Figure 1C). We found that the N-terminal MBP fusion tag provided additional stability to Brpt5.5 in solution without perturbing Zn$^{2+}$-dependent protein assembly, so the fusion tag was left intact for the subsequent experiments. The CD spectrum of uncleaved His-
MBP/Brpt5.5 is shown in Figure 1E, revealing significant helical secondary structure corresponding to the fold of the MBP fusion tag [200]. AUC sedimentation velocity was repeated with uncleaved Brpt5.5; both cleaved and uncleaved Brpt5.5 proteins sediment as well-behaved, highly elongated monomers in solution. The uncleaved His-MBP/Brpt5.5 has a larger sedimentation coefficient, consistent with the additional 43 kDa mass due to the MBP fusion tag (Figure 1C).

**Figure 1E. Secondary structure analysis of Brpt5.5/MBP.** Circular dichroism spectra of Brpt5.5/MBP (black) and Brpt5.5/MBP treated with FA (blue) demonstrates that the secondary structure of Brpt5.5/MBP was not affected by treatment with FA. The secondary structure profile of Brpt5.5/MBP (11% α-helix, 35% β-sheet, 22% turn, 31% coil) is almost identical to Brpt5.5/MBP treated with formic acid (11% α-helix, 33% β-sheet, 23% turn, 32% coil). Secondary structure content was determined using CDSSTAR on Dichroweb.
Brpt5.5 assembles into multiple higher-order species in the presence of Zn\(^{2+}\) - We previously reported that Aap B-repeat constructs containing the C-terminal one or two intact B-repeats followed by the half-repeat (i.e., Brpt1.5 and Brpt2.5) specifically self-associated to form dimers in the presence of Zn\(^{2+}\). These data indicated that tandem B-repeats self-associate in a modular fashion, with each B-repeat region capable of forming an adhesive contact with another B-repeat region in the presence of approximately two Zn\(^{2+}\) ions, and that longer repeats bind Zn\(^{2+}\) in a cooperative manner [3]. For the longer Brpt5.5 construct described here, self-association in the presence of Zn\(^{2+}\) could be significantly more complicated than a simple monomer-dimer equilibrium, due to the potential for multiple adhesive interactions within a stretch of five intact B-repeats. We therefore analyzed the behavior of 5 µM His-MBP/Brpt5.5 in the presence of 3 mM ZnCl\(_2\) by analytical ultracentrifugation (AUC) to characterize the self-assembly process. The raw sedimentation velocity data clearly shows that addition of Zn\(^{2+}\) to His-MBP/Brpt5.5 induces a dramatic change in the sedimentation behavior, with much faster-sedimenting species observed in the presence of Zn\(^{2+}\) (Figure 2A, 2B). Analysis of the sedimentation velocity data using the c(s) model in SEDFIT showed that the data were consistent with self-assembly of Brpt5.5 to form a wide range of oligomeric species (Figure 2C). The predominant peaks in the sedimentation coefficient distribution plot correspond to a putative dimer transitioning to a tetramer or hexamer, with significant populations of higher-order oligomeric species of His-MBP/Brpt5.5. However, there clearly are multiple species that are not well-resolved, and there is potential for overlap of species with the same sedimentation coefficient but different shapes (i.e., co-sedimenting species with different frictional coefficients). Therefore, resolution of the various oligomeric species of His-MBP/Brpt5.5 in the presence of Zn\(^{2+}\) requires a more sophisticated analytical approach, as described below.
**Figure 2. Sedimentation behavior of His-MBP/Brpt5.5 in the presence of Zn\(^{2+}\).**

<table>
<thead>
<tr>
<th>Panel</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A)</td>
<td>Raw sedimentation velocity data for 5 µM His-MBP/ Brpt5.5 and B) His-MBP/ Brpt5.5 in the presence of 3 mM ZnCl(_2), both at 36,000 rpm and 20 °C. In both panels, scans 1-100 were loaded, with every fourth scan plotted. Note the dramatic increase in spacing between scans in panel B compared to panel A; the much faster-moving sedimentation boundaries of His-MBP/Brpt5.5 in the presence of Zn(^{2+}) compared with His-MBP/Brpt5.5 alone indicate the sedimentation of very large species. C) Sedimentation coefficient distribution of Brpt5.5 alone (black) and His-MBP/Brpt5.5 in the presence of 3 mM ZnCl(_2) (red). The broad peak distribution in the presence of Zn(^{2+}) indicates that His-MBP/Brpt5.5 sediments as a mixture of assembled oligomers that appear to range from dimer up to putative 24-mer; however, resolution of individual species is difficult due to overlapping sedimentation profiles. Furthermore, highly elongated or non-globular species can potentially co-sediment with smaller oligomeric complexes, preventing accurate representation of all species in solution.</td>
</tr>
</tbody>
</table>

**Brpt5.5 forms amyloid fibers in presence of Zn\(^{2+}\).** To characterize the nature of these oligomeric and aggregated species, samples of His-MBP/Brpt5.5 incubated with Zn\(^{2+}\) for 2, 4, or 7 days were visualized by transmission electron microscopy (TEM) (**Figure 3, A-D**). Negative-stained TEM images revealed large assemblies of protein fibers with multiple, distinct morphologies. Individual fibers within these fiber complexes measured as small as 15 nm and as large as 100 nm in diameter. One possibility raised by the TEM images is that His-MBP/Brpt5.5 can form amyloid fibers in the presence of Zn\(^{2+}\). Amyloid fibers are highly stable protein fibers that have a characteristic β-strand fibril architecture [121]. Formation of amyloid fibers was initially linked to human diseases; however, a new class of “functional” amyloid proteins required for specific cell processes has recently been established in organisms ranging from bacteria to humans [180]. Functional amyloid proteins from bacterial species, such as *Escherichia coli*, *Bacillus subtilis*, *Enterobacteriaceae*, and *Pseudomonas sp.*, have recently been implicated in
Figure 3. Amyloid properties of Brpt5.5 in the presence of Zn\(^{2+}\). (A-D) Negative-stained TEM images of His-MBP/Brpt5.5 protein fibers generated 2 days (A), 4 days (B and C), and 7 days (D) post-incubation with 500 µM ZnCl\(_2\). His-MBP/Brpt5.5 alone formed very low levels of fibers; when they did form, they resembled the morphology shown in panel (C). Scale bars represent 500 nm (A, B, D) or 1 µm (C). E) Fluorescence emission spectra of 10 µM Thioflavin T (ThT) alone (black) or ThT in the presence of: His-MBP/Brpt5.5 with 500 µM ZnCl\(_2\) (red), His-MBP/Brpt5.5 alone (dark blue), and His-MBP/Brpt5.5 with 10% formic acid (cyan) after excitation at 432 nm. His-MBP/Brpt5.5 without ThT (green) did not produce fluorescence. The His-MBP/Brpt5.5 alone and His-MBP/Brpt5.5 + Zn\(^{2+}\) samples were pre-treated with 10% formic acid to remove aggregates, followed by dialysis back into working buffer and addition of ZnCl\(_2\).
biofilm formation; specifically, these functional amyloid fibers are integral to the overall stability of the biofilm [174, 184, 189, 191].

To confirm that these His-MBP/Brpt5.5 protein fibers were amyloid, samples were incubated with Thioflavin T (ThT) and characterized by fluorescence spectroscopy. Thioflavin T is a small-molecule fluorophore that can intercalate within amyloid fibers, which restricts rotation about an internal bond and results in a dramatic increase in quantum yield and characteristic fluorescence emission at 482 nm [131, 201]. To ensure that the ThT fluorescence produced was truly from Zn$^{2+}$-dependent amyloid fibers, His-MBP/Brpt5.5 was first pre-treated with 10% formic acid (FA) to depolymerize any fibers that may have spontaneously formed [129]. As seen with other amyloid proteins, His-MBP/Brpt5.5 with Zn$^{2+}$ produced minimal ThT fluorescence in the presence of FA (Figure 3E). This effect was not due to FA-induced structural changes in the Brpt5.5 monomer, since the secondary structure of the protein was unchanged upon addition of FA (Figure 1E). Upon removal of FA and incubation for 24 hours, His-MBP/Brpt5.5/Zn$^{2+}$ protein fibers demonstrated strong ThT fluorescence (Figure 3E). However, after FA removal and incubation, His-MBP/Brpt5.5 alone also produced a low level of ThT fluorescence, suggesting that some fibers could potentially form in the absence of Zn$^{2+}$. Based on the AUC data on His-MBP/Brpt5.5 alone, the proportion of protein that forms fibers in the absence of Zn$^{2+}$ is 3% or less (Figures 1 and 2).

*Brpt5.5 fiber assembly is time- and temperature-dependent* - Metal-induced amyloid fiber assembly has been shown to be a critical factor for disease pathogenesis for the amyloid-β peptide in Alzheimer’s disease [155, 156, 202, 203]. However, functional amyloid-forming proteins from bacteria assemble into amyloid fibers without a known requirement for zinc ions or
other triggering molecules. Rather, these amyloid proteins form as a result of the coordinated action of accessory proteins expressed within operons, including nucleator proteins and chaperones to prevent aberrant polymerization in the cytoplasm [174, 185, 190, 191]. Thus, the mechanism of Zn$^{2+}$-dependent amyloid formation by the B-repeat region of Aap is of significant interest for comparison to other systems. We have therefore used a combination of HPLC, TEM, confocal microscopy, and AUC to analyze the assembly of Aap B-repeat amyloid formation as a function of time, temperature, and solution conditions.

To assess the effect of temperature on the rate and morphology of amyloid fiber formation, we used a HPLC quantification assay previously shown to effectively differentiate between monomer and oligomer or soluble fiber species [124]. A relative time course for His-MBP/Brpt5.5 fiber formation was established by incubating 10 µM His-MBP/Brpt5.5 with 500 µM ZnCl$_2$ for 0, 2, and 6 hours and 1, 4, and 30 days at both 20 °C and 37 °C (Figure 4). Samples were first centrifuged for one hour at 13,000 rpm to remove any insoluble aggregates prior to separation by a C4 HPLC column; therefore, the elution profile only reports on monomer, oligomer, and/or soluble amyloid fiber species. Upon comparison of the HPLC elution profile of His-MBP/Brpt5.5 alone (Figure 4) with its sedimentation coefficient distribution (Figure 1B), it is clear that the primary peak eluting at 28.5 ml corresponds to monomeric His-MBP/Brpt5.5, while the low-amplitude peak eluting at 29.5 ml corresponds to an oligomer or fiber species. Upon addition of Zn$^{2+}$ to His-MBP/Brpt5.5, samples gradually showed a shift in the distribution of peaks in the elution profiles. The elution profiles at 2 and 6 hours were unchanged from time point zero (data not shown), indicating slow assembly kinetics. As time progressed, the monomer peak decreased and shifted to the right as the oligomer peak increased. The monomer peak transition was more pronounced after 1, 4, or 30 days for samples
incubated at 37 °C (Figure 4B) compared to the samples incubated at 20 °C (Figure 4A), indicating accelerated His-MBP/Brpt5.5 self-assembly at higher temperature.
**Characterization of Brpt5.5 fibers by microscopy** - The HPLC analysis clearly demonstrated a temperature-dependent difference in Brpt5.5 fiber assembly, but did not provide insight into the assembly of early proto-oligomer or fiber species observed by AUC that lead to mature fiber assembly. We therefore used TEM to characterize both early oligomeric assemblies and mature fibers formed by His-MBP/Brpt5.5 with Zn\(^{2+}\). Samples of 10 µM His-MBP/Brpt5.5 were incubated with 500 µM ZnCl\(_2\) for 2, 6, 12, 18, 24, 48, and 96 hours, either at 20 °C or 37 °C, prior to imaging by TEM (Figure 5). After viewing all of the microscopy images, it is evident that there are multiple morphologies, with at least two distinct forms of mature fibers produced by His-MBP/Brpt5.5. The formation of heterogeneous amyloid fiber morphologies by the same protein, even within a single sample, is common among other amyloid-forming proteins [5, 134, 135]; therefore, both morphologies could potentially be physiologically relevant. Due to the broad heterogeneity of the system, the time course experiment was repeated multiple times in order to determine the best representative morphologies and to better understand the relationship between particular early-oligomer intermediates and mature fibers. In parallel, we used confocal fluorescence microscopy (CFM) at each time point, utilizing ThT as the fluorophore, to help identify: 1) the fiber morphologies that have an amyloid conformation and 2) the time point during assembly at which the intermediate oligomers become amyloid.

By TEM, the 2 hour time point contained similar oligomer assemblies for both 20 °C and 37 °C samples, a mixture of round or oblong complexes that sometimes associate to form small, worm-like species. These compact oligomeric complexes are reminiscent of early proto-fibril species seen in the formation of other amyloid proteins [138, 143, 144], suggesting that they could be the critical nucleating unit for mature fiber assembly. By CFM, we were surprised that these small oblong complexes (about 1 or 2 nm thick) exhibited ThT fluorescence, indicating
that these smaller assemblies have an amyloid conformation (Figure 5A and 5B). The 37 °C CFM samples additionally indicated that the worm-like species have a low level of ThT fluorescence. Between 2-18 hours, both the TEM and CFM images show that the round/oblong complexes continue to associate into longer worm-like fibers (Figure 5, asterisks) that are used as scaffolding for a thread-like fiber (Figure 5, arrows). There is however, a slight modification in the fiber morphology between samples at each temperature. By 24 hours, a new, more ordered fiber morphology is present at both temperatures that appears to be composed of a braided bundle of the thin, thread-like fibers (Figure 5, arrowheads). This fiber morphology is verified as amyloid by ThT fluorescence (Figure 5C and 5D) and will be called the “rope-like” fiber morphology. However, in addition to the rope-like fibers, there was a mature fiber morphology visible as early as 6 – 12 hours in some preparations, which we will term “rod-like” fibers (Figure 5E). These rod-like fibers were also confirmed to have an amyloid confirmation by CFM (Figure 5F). The assembly mechanism for these fibers is not as clear as the progression to the rope-like fiber; however, we propose that fiber assembly may occur by stacking of critical intermediate species to form a mature fiber, as observed for other amyloid proteins [4]. This hypothesis is based on the smooth fiber morphology as well as the punctate proteinaceous material that decorates the sides of many of the fibers, an indication that portions of the monomer subunit are not being assembled in the fiber core [204].
Figure 5. Temperature and time-course analysis of Zn-mediated Brpt5.5 fiber assembly. 10 µM Brpt5.5 was incubated with 500 µM ZnCl$_2$ at A) 20 °C or B) 37 °C (panel B). 5 µl samples were removed at 2, 6, 12, 18, and 24 hours and fixed on copper grids for TEM and stained with 1% uranyl acetate (scale bars = 500 nm). Parallel samples at each time point were stained with 10 µM thioflavin T for CFM at 63x with an excitation wavelength of 458 nm (scale bars = 10 µm). The time course images show a progression of oblong intermediate species that eventually assemble into short, worm-like fibers, thin thread-like fibers (arrows), and then mature, “rope-like” amyloid fibers by 24 hours. C) Mature fibers at 24 hours were additionally imaged with brightfield background to demonstrate ThT specificity at 20 °C or D) at 37 °C. E) An additional ‘rod-like’ mature fiber morphology formed under both temperature conditions that also exhibited ThT fluorescence F) TEM image of the rod-like fiber morphology. Note in panels E and F that the rod-like fiber exhibits discrete punctate regions of fluorescence or proteinaceous material alongside the fiber consistent with globular protein regions not incorporated into the regular amyloid fiber (arrows).
Resolution of early assembly intermediates during amyloidogenesis by AUC - A newly developed AUC analysis method was used to provide additional mechanistic information about early-stage oligomers and soluble amyloid fibers that cannot be resolved by HPLC. As described in Figure 2, the c(s) analysis of sedimentation velocity data showed a sedimentation coefficient distribution with multiple oligomeric states. In order to resolve the large number of species, we utilized the c(s, ff₀) analysis in SEDFIT that characterizes each sedimenting species in terms of both sedimentation coefficient and frictional ratio [198]. Thus, this approach separates distinct species based on both size and shape, and is able to provide detailed resolution of the oligomeric states populated during initial stages of amyloid fiber formation. For these experiments, we used samples of 5 µM His-MBP/Brpt5.5 alone or His-MBP/Brpt5.5 in the presence of 500 µM or 1 mM ZnCl₂. The samples were analyzed at both 25 °C and 37 °C. In order to capture the earliest assembly events, the samples were incubated after addition of Zn²⁺ only for the time required to pull vacuum (approximately 30 minutes). As expected, the Brpt5.5 alone sample yielded a single predominant species corresponding to a highly elongated monomer, as seen in Figure 7A. Both samples incubated with Zn²⁺ formed similar oligomeric states, but the higher oligomeric species were more heavily populated for the sample incubated with 1 mM ZnCl₂, allowing for easier resolution of the species involved (data not shown). The c(s, ff₀) analysis yields a 3-dimensional plot (Figure 6, A and B) that separates the species based on sedimentation coefficient on the x-axis, frictional ratio (f/f₀; i.e., the frictional coefficient of the sedimenting species compared to that of an ideal sphere of identical volume) on the y-axis, and peak amplitude on the z-axis. We also show a top-down view of the distribution (Figure 6C) superimposed on the standard sedimentation coefficient distribution that has been labeled with the putative oligomeric states, implied by the combination of sedimentation coefficient and
frictional ratio. The data show that His-MBP/Brpt5.5 incubated with 1 mM ZnCl$_2$ at 37 °C forms multiple species including: monomeric His-MBP/Brpt5.5, several mostly compact oligomeric species (dimer, trimer, and pentamer), followed by larger oligomers of ever-increasing degrees of elongation. In the presence of Zn$^{2+}$, peaks for both elongated and compact monomer species are observed ($f/f_0$ values of 3.1 and 1.2, respectively), in contrast to the Brpt5.5 alone data that shows only elongated monomer (Figure 6A). The dimer and trimer species are mostly compact, with frictional ratios comparable to those of globular proteins (between 1.2 and 1.4). The pentamer species is moderately elongated, with a frictional ratio of 1.7, but all higher oligomers are highly elongated. The putative 9-mer and 14-mer species show frictional ratios of 2.2 or 2.5, respectively, while the fiber-like 15-mer through 65-mer peaks (circled with magenta ovals in Figure 6C) gave $f/f_0$ values between 3.9 and 4.7. (As a control, the analysis of the same sedimentation data with an upper limit of 2.5 for $f/f_0$ values resulted in a worse fit to the measured sedimentation boundaries.) These $f/f_0$ values of 3.9 or greater indicate extremely elongated fiber-like morphologies, with approximate axial ratios that range from 72 to 108 (assuming prolate ellipsoids), suggesting that these species may be nascent amyloid fibers. The His-MBP/Brpt5.5 sample incubated with 1 mM ZnCl$_2$ analyzed at 25 °C showed a very similar distribution of oligomeric species, although the slower sedimentation rates (due to lower temperature) allowed resolution of a few higher-order oligomeric species, such as putative 75-mer and 115-mer fibers (data not shown). In consideration of both the AUC data and microscopy images, we suggest a plausible mechanistic model for two distinct proto-fibril assembly pathways for amyloidogenesis by His-MBP?Brpt5.5, as described in the Discussion and shown in Figure 8.
Figure 6. AUC c(s, \(f/f_o\)) analysis of early-stage His-MBP/Brpt5.5 amyloidogenic intermediates. Sedimentation velocity (36,000 rpm at 37 °C) data were analyzed using the c(s,\(f/f_o\)) analysis model in Sedfit. A) 3-D shape and size distribution plot for His-MBP/Brpt5.5. Sedimenting species are distinguished based on sedimentation coefficient (plotted uncorrected for buffer conditions) along the x-axis and frictional ratio (\(f/f_o\)) along the y-axis. Increasing values of \(f/f_o\) correspond to more highly elongated or non-globular species. The heat map indicates species concentration, from lowest population density (blue) to highest (red). His-MBP/Brpt5.5 alone sediments as a single dominant species of 4.53 S with an elongated frictional ratio (\(f/f_o = 2.3\)). B) 3-D shape and size distribution plot for His-MBP/Brpt5.5 in the presence of 1 mM Zn\(^{2+}\). In the presence of Zn\(^{2+}\), there is a broad distribution of species that vary both in sedimentation coefficient values as well as frictional ratios. Note in particular the series of extremely elongated species (\(f/f_o\) values of approximately 4 or higher) highlighted by the magenta oval. C) To illustrate the putative species present, the three-dimensional plot from panel B has been simplified to a two dimensional distribution of the sedimentation coefficient (x-axis) and frictional ratio (y-axis), labeled with the putative species present as implied by the given pairs of s- and \(f/f_o\)-values. Elongated species with \(f/f_o\) values of approximately 3 or greater are highlighted by ovals and putative species labels in magenta. Compact species with \(f/f_o\) values between 1 and 2.5 are delineated by dark red lines along with putative species labels under the distribution plot. The solid black line depicts a 2-dimensional representation of c(s,*), showing the relative total amount of material at any sedimentation coefficient.
**Amyloid fibers are established early in S. epidermidis biofilm formation** - While it has been well established that Aap is critical for *S. epidermidis* biofilm formation, the mechanism of how Aap promotes intercellular adhesion is not well understood. Others have shown that bundles of Aap fibrils extend outward from the cell wall on planktonic *S. epidermidis* cells [69, 75], but these were not identified as amyloid fibers. To characterize Aap amyloidogenesis in the context of developing and mature biofilms, we followed biofilm formation by *S. epidermidis* strain RP62A at 2, 6, 12, 18, and 24 hours post-inoculation using CFM, with the addition of ThT to the media during the initial inoculation (**Figure 7A**). In parallel, biofilms grown at the same time intervals were stained with LIVE/DEAD fluorescent dye to ensure that ThT was not affecting cell growth or biofilm formation (**Figure 7B**). As early as 2 hours post-inoculation, punctate ThT fluorescence was visible on non-accumulated cells. This suggests that fibers start to form to a limited degree prior to cellular accumulation or biofilm assembly (**Figure 7A**). At 6 hours, the formation of microcolonies, and therefore intercellular accumulation, begins to occur. At this stage, ThT fluorescence is located at the boundaries between associating cells, which is consistent with the role of Aap as the critical factor for intercellular adhesion. ThT fluorescence increases throughout the biofilm over time, with a predominance of fluorescence in the core of the biofilm with the highest cellular densities. These results support our hypothesis that Aap is important both for intercellular accumulation, by the formation of amyloid fibers between bacteria, and for stabilizing mature biofilms due to the remarkable resistance of amyloid fibers to physical and chemical insults.
Figure 7. Amyloid fibers are early structural components in *S. epidermidis* biofilms. A) Biofilm formation by *S. epidermidis* was characterized 2, 6, 12, 18, and 24 hours post-inoculation by confocal microscopy, staining with 10 µM ThT (cyan) over brightfield. In the 2-hour panel, a zoomed inset is shown with increased contrast and brightness to highlight punctate ThT fluorescence visible around non-accumulated *S. epidermidis* cells. ThT fluorescence is seen primarily at cell-cell junctions at the 6-hour time point. Mature biofilms (24 hours) show ThT throughout the biofilm as a major structural component. B) *S. epidermidis* biofilms at 2, 6, 12, 18, and 24 hours post-inoculation were stained in parallel with LIVE/DEAD stain (green/red) as a control. (scale bar = 10 µm). Images were generated in Zen 2009 Light Edition.
Discussion

Our previous work established that the B-repeat region of Aap can undergo Zn^{2+}-mediated self-association in a modular fashion and that Zn^{2+} chelation was able to inhibit biofilm formation by both *S. epidermidis* and *S. aureus* [3]. This initial work was carried out with the short B-repeat constructs Brpt1.5 and Brpt2.5, which showed reversible self-association upon addition of chelator or exposure to moderate acidification (i.e., lowering the pH from 7.4 to 6.0 completely abrogated Zn^{2+}-induced dimerization of Brpt1.5). Given that biofilms typically undergo acidification over time, it was unclear how this Zn^{2+}-mediated self-assembly mechanism could be maintained within the biofilm *in vivo*. Furthermore, it is known that at least five tandem B-repeats are required to support biofilm formation. Therefore, the characterization of a longer B-repeat construct is of significant interest.

The results presented here using a minimal biologically relevant construct (Brpt5.5) identify for the first time a specific staphylococcal protein capable of forming Zn^{2+}-dependent functional amyloid fibers within both nascent and mature biofilms. While the formation of amyloid fibers in biofilms has been established in other bacterial species, the mechanism of Aap amyloid fiber assembly is unique among this group of bacterial biofilm proteins. Unlike the amyloidogenic biofilm proteins curli (*E. coli* and *Salmonella sp.*) [185, 186] TasA (*B. subtilis*) [189, 190] or FapC (*Pseudomonas sp.*) [191] that require additional proteins for amyloid fiber assembly, our data suggest that Aap utilizes Zn^{2+} as a catalyst to drive amyloid fiber formation. This mechanism for metal-dependent amyloid nucleation is instead reminiscent of several mammalian amyloid-forming proteins including amyloid-β [157, 162, 202], prions [158, 205], and β_{2}-microglobin [159, 206]. It has been shown that Aap requires cleavage of the N-terminal
A-repeat region and $\alpha/\beta$ domain by an unknown protease in order to support biofilm formation [68]. Once the B-repeat region is unmasked and exposed to appropriate concentrations of Zn$^{2+}$, Aap would self-assemble and nucleation of amyloid fibers could begin, leading to the formation of elongated, interlocking assemblies that can further intertwine with fibers on adjacent cells.

The molecular characteristics of diverse types of amyloid fiber architecture have been well described in multiple studies using a combination of techniques, including AFM, fiber diffraction, CD, FTIR, and NMR [133, 163, 169, 207, 208], and more recently cryoEM [209] and x-ray crystallography [6, 170, 210]. AUC approaches have been used to characterize mature amyloid fibers in solution [211, 212]. However, there have been few studies about the assembly of early oligomers and proto-fibril intermediates that produce mature amyloid fibers [153, 213, 214], particularly in terms of resolving discrete intermediate species. For example, pre-amyloid oligomerization of transthyretin has been characterized by AUC but without resolution of individual species [215]. We have applied the $c(s, \mu_0)$ analysis to sedimentation velocity AUC data to resolve the broad array of species present at early stages of Aap amyloidogenesis, including oligomeric complexes and nascent fibers. This is the first use of this new analysis approach to an amyloid-forming protein, and its ability to separate many species with differing sizes and shapes in solution holds promise for resolving assembly intermediates in other amyloid systems as well. Another approach designed to resolve size and shape information from sedimentation velocity data has been applied to mature amyloid fibers [216], but concerns have been raised about the validity of this approach [217]. The $c(s, \mu_0)$ analysis model is capable of distinguishing species of different size and shape that are sedimenting at the same rate, which can occur in amyloid assembly systems due to the complicated nature of the assembly process and the number of species to be resolved. This analysis works particularly well in systems such
as this one with slow kinetics, which allows resolution of discrete assembly intermediates rather than simply showing broad reaction boundaries representing multiple species in rapid exchange. The c(s, ff₀) analysis of Aap presented here clearly discriminates between compact oligomers and fibers that show similar sedimentation coefficients due to the increased drag and resulting slower sedimentation of the highly elongated fibers. We have thus been able to define the early oligomeric complexes leading to mature amyloid fibers and can propose a model for Aap amyloidogenesis (Figure 8).

Based on the c(s, ff₀) AUC analysis, His-MBP/Brpt5.5 alone sediments as an elongated monomer. Shortly after addition of Zn²⁺, both elongated and compact monomers are observed, combined with a large array of higher-order species. It is quite interesting that the AUC analysis reveals two distinct populations of oligomers: relatively compact dimers, trimers, pentamers, 9-mers and 14-mers; and highly elongated fibers starting at approximately 15-mer and increasing to 65-mer or larger. Furthermore, the spike at the far right of the sedimentation coefficient distribution represents large aggregates (Figure 6C). Similarly, our TEM analyses of Zn²⁺-induced Brpt5.5 aggregation showed two very distinct fiber morphologies, along with amorphous aggregate. We propose that the braided rope-like fiber is formed by overlapping self-assembly of individual B-repeat regions in the presence of Zn²⁺, as previously described [3]. This type of extended oligomer is likely represented by the initial fiber-like species observed in the c(s, ff₀) analysis (i.e., the 15mer, 20mer, or 26mer peaks highlighted by magenta ovals in Figure 6C). In contrast, the compact oligomers likely represent a condensed form of such an oligomer in which the flexible Brpt5.5 chains curve to make downstream contacts with other B-repeats (see Figure 8). We propose that these compact oligomers may stack to nucleate the ‘rod-type’ fiber seen by TEM. This is suggested by the observation of additional punctate proteinaceous
material at the boundaries of the rod-like fibers by TEM and CFM, respectively (Figure 5E, 5F). The assembly of the rod-like fibers might then be similar to fiber assembly by superpleated β-sheet stacking, as proposed for Ure2p or amylin [171, 172]. Based on these models, it can be imagined how the rope-like fibers would be able to form between adjacent S. epidermidis cells upon interaction with Zn\(^{2+}\). This would provide a stable interaction between cells during the accumulation phase of the biofilm. The idea that the rod-like fiber is formed by stacking of compact oligomers suggests that this mechanism might stabilize the interaction between cells that have made an initial interaction via Aap. It is possible that the same mode of lateral interaction between chains that nucleates amyloid formation in the rod-shaped fiber could also occur in conjunction with the rope-like fibers. Detailed knowledge of the early-stage amyloidogenic assemblies will be important information to help guide strategies for inhibition of amyloidogenesis, since the mature amyloid fibers are highly resistant to chemical treatment.

We have previously shown that Zn\(^{2+}\) is required for biofilm formation by S. epidermidis and S. aureus [3], which would indicate that the Zn\(^{2+}\)-induced amyloid fibers are a critical component for staphylococcal biofilms. However, we do observe Brpt5.5 amyloid fibers that appear to form in the absence of Zn\(^{2+}\), based on ThT fluorescence and TEM data, although their prevalence is very low as shown by AUC and HPLC. It is possible that the Zn\(^{2+}\)-free amyloid morphology supports functions of Aap other than intercellular adhesion, such as adherence of planktonic S. epidermidis to mammalian host cells. Aap fibrils have been observed extending outward from planktonic S. epidermidis cells [69, 75], and the N-terminal region of uncleaved Aap has been implicated in binding to nasal epithelial cells and corneocytes [69, 218]. It is interesting that factors that promote biofilm formation have been shown to prevent commensal colonization of the skin [24]. Therefore, protease cleavage of Aap followed by Zn\(^{2+}\)-mediated
amyloid fiber formation could be a pathway specifically leading to *S. epidermidis* pathogenesis whereas Zn\(^{2+}\)-independent Aap fibers might mediate commensal colonization. This phenomenon would be consistent with other amyloid proteins, for which environmental conditions such as divalent cations, pH, and salt govern the formation of pathogenic versus non-pathogenic amyloid confirmations [137, 145, 156, 161].

Biofilm formation is the primary characteristic responsible for pathogenicity and it contributes to antibiotic resistance in chronic infections caused by *S. epidermidis*. One of the biggest challenges with recurrent infections caused by biofilms is their highly adhesive and cohesive nature and their resistance to chemical and physical insults. Our data indicate that intercellular amyloid fibers appear early during the accumulation phase of nascent biofilms, and they continually increase until they are ubiquitous throughout the mature biofilm. Furthermore, we predict that homologous surface proteins containing tandem B-repeats in *S. aureus* (i.e., SasG and Pls) and other gram-positive bacteria will form similar Zn\(^{2+}\)-dependent amyloid fibers between cells in biofilms. The data presented here will provide new insights for both potential prevention and treatment of chronic staphylococcal infections, particular with regard to methods that would depolymerize amyloid fibers and thus destabilize the biofilm.
Figure 8. Model for Zn$^{2+}$-induced amyloidogenesis by the B-repeat region of Aap. The c(s, ff) AUC analysis of early-stage oligomeric intermediates of Brpt5.5 in the presence of Zn$^{2+}$ revealed both compact oligomers and long fiber-like species in addition to compact aggregate. TEM and CFM revealed the existence of worm-like and thread-like intermediate fibers, amorphous aggregate, and mature rope-like and rod-like fibers. We propose that the compact oligomers may stack laterally to nucleate the rod-like fibers, which resemble other well-characterized amyloid fibers. The presence of punctate regions of protein density along the rod-like fibers (Figure 5F) is proposed to be due to portions of the Brpt5.5 protein not incorporated into the amyloid fiber, such as the MBP fusion tag. We propose that the braided rope-like fibers observed by TEM after 24 hours incubation are formed by the overlapping self-assembly of individual B-repeat domains, as previously described [3].
Acknowledgments

The authors thank Georgianne Ciralo and Cincinnati Children’s Department of Pathology for assistance and use of the electron microscopy facility, Dr. Dan Hassett and lab members and Birgit Ehmer for assistance with the confocal microscope, Andy Deng and Dr. Tom Thompson’s lab members for help with the HPLC experiments and analysis, and Dr. Nicolas Nassar for use of their fluorescence spectrophotometer. The authors would also like to thank Dr. Rhett Kovall, Dr. Tom Thompson, and Catie Shelton for comments on the manuscript.
Chapter 5

Connecting the Mechanistic Relationship of Aap Self-Assembly and Biofilm Formation by Staphylococcus epidermidis*

*Part of this work has been included in the patent: Herr, AB, Conrady, DG, Brescia, CC, Johns (Ward), SL. Use of Zinc Chelators to Inhibit Biofilm Formation. U.S. patent pending, filed May 29, 2009. International application #PCT/US2009/045623.
Abstract

Biofilm-forming staphylococci, including the highly antibiotic-resistant species \textit{S. epidermidis} and \textit{S. aureus}, are one of the leading causes of hospital-acquired infections worldwide. Despite the high frequency of these infections, there are still no effective treatments to both reduce the number of infections and disperse already formed infectious biofilms. It is well established that staphylococci utilize a zinc-mediated protein-dependent mechanism for biofilm formation, although some of the mechanistic details still remain unresolved. In addition, it has recently been shown that the Accumulation-associated protein (Aap), which is primarily responsible for this mechanism in \textit{S. epidermidis}, forms amyloid fibers in the presence of zinc. The resilience of biofilms to harsh chemical treatments correlates well with the inclusion of amyloid fibers in the biofilm framework, which are also known to be highly stable. Environmental factors, such as pH and zinc concentration, can dually influence both biofilm formation and amyloidogenesis. Therefore, the goal of this chapter is to determine if there is a mechanistic link between \textit{S. epidermidis} biofilm formation and Aap amyloidogenesis. Our results indicate that both early stages of biofilm development and amyloid fiber formation are susceptible to acid degradation and metal chelation. However, both mature biofilms and amyloid fibers were found to be resistant to chemical denaturants. These studies reinforce the mechanistic relationship between \textit{S. epidermidis} biofilm formation and Aap amyloidogenesis.
Introduction

The U.S. National Institute of Health estimates that 80% of all microbial infections are biofilm related. Biofilm-forming, gram-positive cocci, and more specifically staphylococci, are the leading cause of all hospital-acquired infections (HAI), causing roughly 40-60% of the 1.7 million HAIs that occur yearly in the United States alone [18, 24, 219, 220, 222]. Biofilm-associated infections are resistant to antibiotic therapy because bacterial cells within the biofilm are not actively growing; metabolic components are the target of most antibiotics [194]. In addition, the antimicrobial-resistance of HAI clinical isolates has nearly doubled between 1990 and 2004 [220], making these infections additionally difficult to treat. Together, these factors have left prevention as the current best practice to reduce morbidity and mortality caused by staphylococcal biofilm-related and hospital-acquired infections. Therefore, in order to develop better therapeutic treatments for biofilm-associated infections by staphylococci, it is imperative to understand the underlying mechanism of biofilm formation.

It is well accepted that there are two major contributing pathways to biofilm formation in staphylococci: protein-dependent, primarily by the Accumulation-associated protein (Aap or SasG), and polysaccharide-dependent [68], which is mediated instead by the polysaccharide intracellular adhesion poly-N-acetylglucosamine (PNAG) [57]. While both protein and polysaccharide factors have been shown to be clinically relevant [49-53, 55, 82], Aap has been identified in 90% of clinical isolates from a study of prosthetic joint infections compared to only 61% for PNAG-positive isolates [70]. In addition, anti-Aap antibodies were shown to prevent biofilm formation despite the presence of polysaccharide factors, suggesting that Aap could be
required for PNAG cell attachment or accumulation [81] and therefore is a better therapeutic target.

Aap contributes to biofilm formation by providing cell-to-cell adhesion, or accumulation. Aap-driven accumulation has been previously shown to be a zinc-mediated mechanism [3]. Recently, Aap has been shown to form amyloid protein fibrils in vitro, suggesting that Aap amyloidogenesis is the underlying mechanism for intracellular adhesion [Chapter 4]. Amyloidogenesis is influenced by a variety of environmental factors such as osmolarity, pH, metal ions, and temperature [5, 133], factors that could therefore also affect biofilm formation. The environmental pH has previously been shown to influence self-association of short Aap B-repeat constructs in the presence of zinc, inhibiting zinc-mediated dimerization at pH values less than 6.7 [3]. However, over time, biofilms are known to become acidified as a result of the build-up of metabolic products produced by densely packed cells within the biofilm [246]. This questions the ability of Aap to maintain the stable biofilm structure in an acidic environment and must be investigated further. Staphylococcal biofilm formation has been shown to be inhibited by the divalent cation chelator DTPA [3], although this treatment has never been specifically related to the inhibition of Aap-driven accumulation of amyloidogenesis. Therefore, we aim to address the functional correlation between the molecular assembly mechanism of Aap and biofilm formation by Staphylococcus epidermidis. Specifically, we will assess the requirements of zinc and environmental stresses, such as low pH, for biofilm formation and Aap amyloidogenesis.
Materials and Methods

*Preparation of reagents and biofilm inhibitors* - Crystal violet: 1 gram of crystal violet was dissolved in 1 liter of filtered water, followed by filtration by a nitrocellulose 0.45 µm filter. Diethylene-triamine pentacetic acid (DTPA): 100 mM stock solution was prepared by dissolving in 500 mM HCl, incubated at room temperature, rocking for 2 hours until all powder was dissolved. This stock solution was used to prepare 10x stocks to use for the assay, diluting concentrated DTPA into filtered water: 200, 400, 800, 1600, 3200, 6400, 13000 µM. Ethylenediaminetetra-acetic acid (EDTA): A 0.5 M stock was prepared by dissolving EDTA powder in filtered water and titrating in 10M NaOH until the pH reaches 8.8. This stock solution was used to prepare 10x stocks for the biofilm assay (as above), diluting into filtered water. N,N,N’,N’-Tetrakis(2-pyridylmethyl) ethylenediamine (TPEN): A 100 mM stock was prepared by dissolving TPEN powder in 100% ethanol. This stock solution was used to prepare 10x stocks for assay (as above), diluting into 100% ethanol. 1,10 – phenanthroline (1, 10 – P): A 100mM stock was prepared by dissolving 1, 10 – P powder in 100 mM HCl. This stock solution was used to prepare 10x stocks for assay (as above), diluting into filtered water.

*Preparation of Chemically Defined Media (CDM)* – Chemically defined media was made according to a modified version of the protocol by Hussain, Hastings, and White [248, 249]. The chemically defined media was created by adding three primary sterile stock solutions to 749 ml of sterilized water: 10X Solution A, 50X Solution B, and 25X Solution C (components defined in Table 1), each autoclaved separately. A Vitamins stock solution (components in Table 1), adenine sulfate, guanine sulfate, 20X stock of Trace Elements, zinc chloride, cobalt chloride,
and copper chloride (concentrations in Table 1) were added separately. Finally, a 50X RPMI amino acid (20 ml, Sigma) and 100x L-glutamine stock (10 ml) were added (concentrations in Table 1). This media will be called CDM + RPMI.

Table 1. Components of the CDM + RPMI Medium. All concentrations are written as the final concentration at 1 L of prepared media.

<table>
<thead>
<tr>
<th>Solution A</th>
<th>g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na$_2$HPO$_4$</td>
<td>8.95</td>
</tr>
<tr>
<td>KH$_2$PO$_4$</td>
<td>3.0</td>
</tr>
<tr>
<td>(NH$_4$)$_2$SO$_4$</td>
<td>2.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Solution B</th>
<th>g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>MgSO$_4$ x 7H$_2$O</td>
<td>0.05</td>
</tr>
<tr>
<td>MnSO$_4$ x H$_2$O</td>
<td>0.0038</td>
</tr>
<tr>
<td>FeSO$_4$ x 7H$_2$O</td>
<td>0.0028</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Solution C</th>
<th>g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vitamins</th>
<th>µg/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biotin</td>
<td>1</td>
</tr>
<tr>
<td>Nicotinic Acid, Thiamine Hydrochloride, D-pantothenic acid (Ca salt)</td>
<td>2</td>
</tr>
<tr>
<td>Riboflavin</td>
<td></td>
</tr>
<tr>
<td>Pyridoxine hydrochloride</td>
<td>4</td>
</tr>
</tbody>
</table>

| Adenine Sulfate | 2 |
| Guanine Sulfate | 2 |

<table>
<thead>
<tr>
<th>Trace Elements</th>
<th>g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>CaCl$_2$</td>
<td>0.7</td>
</tr>
<tr>
<td>NaMoO$_4$</td>
<td>0.0006</td>
</tr>
<tr>
<td>ZnCl$_2$</td>
<td>12 µM</td>
</tr>
<tr>
<td>CuCl$_2$</td>
<td>2.5 µM</td>
</tr>
<tr>
<td>CoCl$_2$</td>
<td>2.5 µM</td>
</tr>
</tbody>
</table>
Bacterial Cultures and Maintenance - *S. epidermidis* strain RP62a (ATCC 35984) was obtained as a glycerol stock from ATCC. 3 ml overnight cultures were grown in tryptic soy broth (TSB, BD Biomedical) or CDM as indicated, shaking at 37°C, and cultures were diluted 1:200 in TSB for use in the biofilm assay.

Semi-quantitative biofilm assay - Each metal or metal chelator was added directly to a 96 well microtiter plate (Costar) in a series of dilutions, 20 µl of each 10x stock solution for “Pre-Treatment” experiments. 200 µl of diluted bacterial culture (diluted 1:200 for TSB; 1:100 for CDM + RPMI) was then added to each well. The media used for the pH dependence biofilm assay was adjusted to the desired pH value and used to dilute the bacterial stock prior to adding to the assay. Controls with bacterial culture media only and vehicle controls (HCl) only (final concentrations 10mM HCl, 10% ethanol, and 10mM NaOH) were added to wells alone without chelator. The 96 well plates were then incubated for 24 hours statically at 37°C. Media was then aspirated by pipet and the wells were washed twice with filtered water. The plates were dried by inversion for 1 hour and stained with 0.1% crystal violet for 15 minutes. Crystal violet was then aspirated and washed once with filtered water. The plates were then read on a spectrophotometer at a wavelength of 570 nm. An absorbance greater than 0.1 indicates biofilm formation. The “Post-Treatment” biofilm assays allowed biofilms to form over 24 hours as described above, alternatively adding back fresh TSB or additionally containing DTPA after the initial aspiration of media. These assays were then incubated and additional 24 hours, continuing with the washing and staining protocol. Assays were set up in duplicate for each environmental factor tested and repeated at least three separate times.

Calculation of Metal Chelator Minimal Inhibitory Concentration (MIC) – 3 ml TSB cultures were inoculated with a 1:200 dilution of *S. epidermidis* RP62a overnight culture (~3 x 10⁷ cells).
Each chelator was then added at concentrations of 0, 10, 20, 40, 80, 160, 320, 640, and 1300 µM. These cultures were incubated overnight, shaking at 37 °C. Each culture was first visibly assessed; clear cultures indicate growth inhibition, followed by optical density (OD) measurement at 595 nm using a spectrophotometer. The lowest concentration of chelator that produces an absorbance less than or equal to the OD at inoculation indicates the MIC. 50 ul samples were additionally removed from each culture and streaked on a tryptic soy agar plate (incubating overnight at 37 °C) to determine if the chelator was bacteriostatic or bactericidal by counting surviving colonies.

Confocal microscopy – To evaluate the effect of metal chelators on biofilm formation, overnight cultures of S. epidermidis RP62a were diluted as described above and 400 µl aliquots were added to 8-well borosilicate glass slides pre-dosed with either 50 or 100 µM DTPA (final concentration). The slides were incubated for 24 hours at 37 °C without shaking. The media was aspirated and each well was washed 3x with diH2O. The biofilms were stained with the LIVE/DEAD stain kit (Invitrogen), using 3 µl stain diluted in 1 ml of sterile saline, and viewed using a Zeiss LSM 510 inverted microscope.

Protein Expression and Purification - 1 L cultures were inoculated with 10 ml of His-MBP-Brpt3.5/BLR(DE3) culture and then allowed to incubate shaking at 37 °C until an OD600 of 0.6-0.8 was reached. The cells were cold shocked in an ice bath and doped with 2% ethanol. Protein expression was then induced using 200 µM IPTG overnight shaking at 20 °C. The cells were then harvested, re-suspended, frozen and thawed prior to lysis by french press. Alternatively, 1 L cultures were inoculated with 10 ml of His-MBP-Brpt5.5/BLR(DE3) culture at an OD600 of 0.6-0.8, and then allowed to incubate overnight with shaking at 37 °C. Protein expression was
then induced using 250 µM IPTG for 6 hours at 25 °C. The cells were then harvested, re-suspended, frozen and thawed prior to lysis by french press. For both His-MBP/Brpt3.5 and His-MBP/Brpt5.5 preparations post-lysis by french press, the cell lysate was centrifuged, and the soluble fraction was decanted onto a nickel-NTA gravity column. Both Brpt proteins were eluted by imidazole step gradient. Fractions confirmed by SDS gel to contain Brpt proteins were then further purified by anion exchange using a 5 ml anionQ fast-flow column (GE Healthcare) followed by size-exclusion chromatography using a superdex 200 column. The His-MBP fusion of Brpt3.5 was cleaved using TEV protease, followed by re-purification over a Ni-NTA and final purification using a superdex 200 column (His-MBP/Brpt5.5 remained uncleaved for the HPLC experiments).

*HPLC Assays* - For the HPLC fiber/oligomer quantification assay, 10 µM Brpt5.5 samples were incubated with or without 500 µM ZnCl₂ for 1 or 4 days at 20 °C or 37 °C, followed by the addition of an equal volume of: 1) buffer (20mM sodium citrate pH 7.4, 150mM NaCl); 2) dilute hydrochloric acid sufficient to lower the pH to 5.0; or 3) 2 mM Na₅-diethylenetriaminepentacetic acid (DTPA). Samples were further incubated for 2 hours, mixed with equal volume of 5% acetonitrile and then separated on a C4 column based on a protocol previously described by O’Nuallain et al [124]. Samples were loaded on a C4 reverse-phase column (Phenomenex) and run on an Äkta purifier with a linear gradient of 0-95% acetonitrile over 10 column volumes. Peaks were integrated using the Unicorn software, normalizing the peak area to a standardized elution volume for monomer and oligomer peaks.
**Transmission Electron Microscopy** - 5 µl samples of each HPLC sample condition were applied to 200 mesh formvar carbon/copper grids for 2 minutes, washed with diH₂O, stained by 1% uranyl acetate drop-wise for 30 seconds, and washed a second time with diH₂O. Samples were then dried for 1 hour prior to viewing on a Hitachi 7600 transmission electron microscope at an accelerating voltage of 80 KV. Images were captured using an AMT 2k CCD camera.

**Analytical Ultracentrifugation** - Experiments were performed with a Beckman XL-I analytical ultracentrifuge using absorbance optics at 238 nm. Sedimentation velocity experiments of 8 µM Brpt3.5 were performed at 48,000 rpm at 25 °C with and without 3 or 6 mM ZnCl₂ ([Figure 7A](#)) followed by direct treatment of equivalent DTPA ([Figure 7B](#)). Data were analyzed using the Sedfit software [117] and the c(s) distribution model. Parameters of buffer density and viscosity and the partial specific volume of Brpt3.5 were calculated using SEDNTERP [199] at all relevant experimental temperatures.
Results

*Transition Metal Requirements for Staphylococcus epidermidis biofilm formation* – Aap-mediated protein accumulation has been previously hypothesized to be driven by zinc based on *in vitro* biophysical experiments. These analytical ultracentrifugation (AUC) experiments indicated that the B-domain region (Brpt) dimerized only in the presence of zinc but not with the other divalent metals tested (magnesium, nickel, cobalt, strontium, manganese, and calcium) [3]. While these experiments were additionally supported by the inhibition of *S. epidermidis* biofilms by a divalent metal chelator that has a higher specificity for zinc, the biofilm assays were performed using an enriched media with unknown metal concentrations. In addition, it has never been tested if copper could potentially mediate Aap protein accumulation; copper is a transition metal that has similar chemical properties and bioavailability to zinc [224]. In addition, the precise physiological concentration of zinc required for Aap-mediated accumulation by *S. epidermidis* has not been determined.

To evaluate the metal requirements for *S. epidermidis* biofilm formation, we have created a chemically-defined medium (Table 1). This chemically defined media (CDM) has been previously shown to support staphylococcal biofilm formation, particularly selecting protein-dependent pathways, and similar growth as enriched media [248, 249]. The original media recipe was altered slightly, removing zinc, copper and cobalt from the Trace Metals stock so that the transition metals of interest could be evaluated (called CDM + RPMI). The *S. epidermidis* cultures grown in the CDM + RPMI medium did display slightly slower growth kinetics [data not shown], which was compensated in the biofilm assay by decreasing the dilution factor from 1:200 (for enriched media) to 1:100 to produce comparable biofilm formation in the CDM +
RPMI (Figure 1). A second version of the CDM + RPMI was created that contained no copper or zinc (called CDM No Metals). While the growth of *S. epidermidis* in the CDM No Metals medium was comparable to growth in the original, complete CDM + RPMI media (data not shown), biofilm formation was largely reduced or completely inhibited in most cases (Figure 1).

To specifically assess the transition metal requirements for *S. epidermidis* biofilm formation, RP62a cultured in enriched (TSB), CDM + RPMI (Complete), or CDM No Metal (0 uM) media were assayed for biofilm formation. Biofilm formation is measured by crystal violet staining and measuring the absorbance at 570 nm. Zinc (grey) or copper (white) was added to the CDM No Metal cultures at 2, 4, 8, or 16 μM. The black dashed line marks the critical absorbance for biofilm formation, in that absorbances less than this line indicate no biofilm formation.

**Figure 1. The requirements of zinc and copper to support *S. epidermidis* strain RP62a biofilm formation.** RP62a cultured in enriched (TSB), CDM + RPMI (Complete), or CDM No Metal (0 uM) media were assayed for biofilm formation. Biofilm formation is measured by crystal violet staining and measuring the absorbance at 570 nm. Zinc (grey) or copper (white) was added to the CDM No Metal cultures at 2, 4, 8, or 16 μM. The black dashed line marks the critical absorbance for biofilm formation, in that absorbances less than this line indicate no biofilm formation.
formation, a semi-quantitative biofilm assay was used. *S. epidermidis* strain RP62a was grown in the CDM + RPMI (control) or CDM No Metal media overnight at 37 °C. Zinc or copper was titrated into 96 well plates separately at concentrations of 2, 4, 8, or 16 µM prior to the addition of the CDM No Metal RP62a stock diluted 1:100 in fresh CDM No Metal media. The assay was incubated for 24 hours at 37 °C, not shaking. The biofilm assays were then washed with filtered water and stained with 0.1% crystal violet to be measured spectrophotometrically at 570 nm. Absorbance values less than 0.1 indicate that no biofilm as formed and indicated under the black dashed line. Biofilms were capable of forming at all concentrations of zinc with no apparent statistical difference between the different concentrations, although the 16 µM samples generally did not produce a robust biofilm (Figure 1). The samples with varied concentrations of copper, however, did show biofilm inhibition as the concentration of copper increased, particularly at 8 and 16 µM (Figure 1). Overall, all samples prepared in the CDM No Metal media with only zinc or copper did not form as robust biofilms as the CDM + RPMI or TSB enriched media, suggesting that both zinc and copper are required for the formation of a robust biofilm.

*Staphylococcus epidermidis* biofilm formation is effected by pH – Previous biophysical studies of a truncated construct of the Brpt region demonstrated that the zinc-mediated, self-association event is sensitive to pH, initiating complex dissociation at pH values less than 6.0 [3]. However, biofilm acidification is known to develop over time due to a build-up of metabolic by-products produced by the bacterial inhabitants [246]. Biofilm acidification does not however cause destabilization of the biofilm, which contradicts the pH sensitivity of the protein-zinc complex by the truncated Brpt protein. The early stages of *S. epidermidis* biofilm development have not been assessed for pH sensitivity, which is when the initial Aap interaction would occur.
Therefore, *S. epidermidis* biofilms will be studied in different pH environments in order to determine the pH sensitivity of this process. The pH of tryptic soy broth (TSB) enriched media was adjusted to values between 5 – 9.4, varying the pH at increments of 0.4, prior to autoclaving. Overnight cultures, grown in standard TSB (pH 7.3) were diluted 1:200 into each of the TSB pH intervals and incubated in a 96 well plate for 24 hours, statically at 37 °C. The biofilms were stained with crystal violet and measured by an absorbance at 570 nm. The data shows that *S. epidermidis* biofilm formation is quite sensitive to pH, with optimized formation occurring between pH 7.4 to 7.8 (Figure 2). At pH values less than 7.4, biofilm formation steadily drops.

**Figure 2. The influence of pH on *S. epidermidis* biofilm formation.** *S. epidermidis* biofilms were grown in TSB with pH values between 5 – 9.4 at 0.4 pH increments. The biofilms were assayed using a semi-quantitative method by crystal violet staining and measuring the absorbance at 570 nm. The assay was performed in duplicate on three different days.
through pH 6.2, with almost complete biofilm inhibition between pH 5 - 6 (Figure 2). There was however a large amount of variability in the total biofilm formed between pH values 7-6.2, suggesting there is an equivalent inhibition at these values. Basic pH values, pH greater than 7.8, show a sharp decline in biofilm formation, decreasing the total biofilm by almost 50% increasing the pH from 7.8 to 8.2 (Figure 2). The data between pH 8.2 and 8.6 is also highly variable, suggesting that there is a minimal affect by pH on biofilm formation in this range. The high alkaline pH values, 9 and 9.4, show relatively the same level of biofilm inhibition as the highly acidic values. Overall, the data suggests that pH can influence *S. epidermidis* biofilm development.

*Broad Specificity Metal chelators have diverse effects on Staphylococcus epidermidis biofilm formation –* *S. epidermidis* biofilm formation has been previously shown to be inhibited by a divalent cation metal chelator, DTPA, and rescue of inhibition was achieved by the add-back of zinc [3]. Based on this data, the use of metal chelators to inhibit this zinc-mediated mechanism and broadly prevent staphylococcal biofilm formation has been of interest. DTPA has a high affinity for zinc, although it is capable of binding other divalent cations, such as iron, that could also affect both biofilm formation and bacterial growth. Therefore, the goal of these experiments is to determine if metal chelators are effective for inhibiting biofilm formation and if metal specificity is important when selecting a chelator for therapeutic purposes.

Four different metal chelators were selected to determine if zinc metal specificity is required to inhibit biofilm formation by *Staphylococcus epidermidis*. The chelators selected have diverse metal specificity (Table 2), in which some chelators are more specific to zinc, such as DTPA and TPEN, while are others less specific to zinc, such as EDTA and 1, 10 – P. S.
epidermidis strain RP62a was assayed with all four chelators at seven concentrations between 20 - 1300 µM by using a semi-quantitative biofilm assay (Figure 3A). All chelators were added prior to inoculation with the bacterial culture, allowing biofilm formation to occur over a 24 hour period. Biofilm formation was measured by crystal violet staining and measured for absorbance at 570 nm; an absorbance over 0.1 indicates biofilm formation. All assays were performed in duplicate on at least three different days.

Treatment with TPEN had the most drastic effect on biofilm formation, causing an average 70% reduction in biofilm formation using only 20 µM (Figure 3A). TPEN is, however, a cell membrane permeable chelator that was found to be bactericidal (MIC = 40 µM TPEN) and is likely killing the cells rather than specifically inhibiting biofilm formation. DTPA was also fairly effective to inhibit biofilm formation, showing a 90% biofilm reduction using 40 µM DTPA and complete abolishment of biofilm formation increasing the concentration to 80 µM (Figure 3A). The MIC for DTPA is 200 µM, well above the concentrations that were required to inhibit biofilm formation; this confirms that DTPA is acting to inhibit biofilm formation and just not cell growth. There was no significant difference treating with EDTA and 1,10-Phenanthroline, and neither were effective to inhibit biofilm formation (Figure 3A). EDTA and 1,10-phenanthroline required 640 and 1300 µM, respectively, to achieve complete inhibition of

<table>
<thead>
<tr>
<th>Chelator</th>
<th>Chemical name</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetra-acetic acid</td>
<td>Membrane impermeable; highest affinity for Ca and Fe</td>
</tr>
<tr>
<td>TPEN</td>
<td>N,N,N',N'-Tetrakis(2-pyrdiylmethyl) ethylenediamine</td>
<td>Membrane permeable; transition metal chelator, particularly Zn</td>
</tr>
<tr>
<td>DTPA</td>
<td>Diethylenetriaminepentacetic acid</td>
<td>Membrane impermeable, highest affinity for Zn and Fe</td>
</tr>
</tbody>
</table>

Table 2. List of metal chelators used in this study and their chemical properties.
A. Dose-response curve of *S. epidermidis* (blue diamonds), EDTA (red squares), TPEN (green triangles), or 1,10-Phenanthroline (purple diamonds). A vehicle control was generated each for DTPA and TPEN, which had no significant differences from the wild-type biofilm. Total (100%) biofilm formation was based on an average of untreated, RP62a biofilms. The black dashed line indicates biofilm formation. B. Fluorescence confocal microscopy biofilm images of wild-type *S. epidermidis* treated with 50 and 100 µM DTPA (treated prior to biofilm formation). Biofilms were allowed to form over 24 hours prior to staining with Baclight and viewing. Cells that are living are stained green, cells that are dead are red. Images were generated using Zen 2009.
biofilm formation. However, both of these concentrations were still greater than or equivalent to the MIC (640 µM for both chelators), suggesting that these chelators likely do not have an impact on biofilm formation. DTPA, EDTA, and 1,10-phenanthroline were all found to be bacteriostatic inhibitors, allowing normal cell growth once the chelator was removed (data not shown).

DTPA proved to be the most effective chelator to inhibit *S. epidermidis* biofilm using the semi-quantitative biofilm assay; however, this assay does not provide information about how the chelator affects the biofilm structure. Therefore, fluorescence confocal microscopy was used to examine the impact of DTPA on the architecture of the biofilm (Figure 3B). 0, 50 or 100 µM DTPA was added to glass microscope chambers prior to inoculating with *S. epidermidis* strain RP62a and incubating for 24 hours, statically, at 37 °C; DTPA is only soluble in HCl, so a vehicle control (treated with 500 mM HCl) was also included (Figure 3B). Each well was gently aspirated and washed with sterile saline prior to staining with Baclight LIVE/DEAD staining reagents, which stain all living cells green and all dead cells red. Treatment with only 50 µM DTPA prevented robust biofilm development, showing large “bare” patches where no biofilm has formed and a decrease in the overall biofilm thickness. The untreated and vehicle control biofilms had an average thickness of 30 µm, whereas treatment with 50 µM DTPA reduced the biofilm thickness of the tallest structures to 16 µm. The sample treated with 100 µM DTPA was even further reduced to either small clusters of cells or single, unattached cells and an overall thickness around 9 µm. In addition, the majority of cells in both DTPA treated conditions are stained green, indicating that DTPA is not having an effect on cell viability. Overall, these studies suggest that DTPA is an effective inhibitor of the *S. epidermidis* biofilm formation mechanism.
**DTPA is not an effective agent to disperse S. epidermidis biofilms** – The availability of effective preventative agents would greatly reduce the overwhelming number of opportunistic, chronic infections caused by *S. epidermidis*. However, preventative measures are not a “cure all” for already established biofilm-related and antibiotic-resistant infections. Therefore, chemical agents that can disperse already formed biofilms are of great interest. We have already shown that DTPA is effective to prevent *S. epidermidis* biofilm formation, although the ability of this chelator to disperse pre-formed biofilms is unknown.

To determine if DTPA is additionally able to act as a dispersion agent, *S. epidermidis* strain RP62a biofilms were formed in 96 well plates with enriched media over 24 hours at 37 °C. The media was then carefully aspirated and replaced with fresh media containing 0, 20, or 80
µM DTPA, allowing the biofilms to incubate at 37 °C for another 24 hours prior to assaying as previously described for Figure 3A. Post-biofilm treatment with 20 µM demonstrated the same biofilm reduction as pre-treated samples, demonstrating roughly 50% decrease from untreated biofilms (Figure 4). However, there is a dramatic difference between samples treated before or after biofilm formation using 80 µM DTPA. RP62a pre-treated with 80 µM DTPA resulted in complete biofilm inhibition, however, biofilms treated with 80 µM DTPA after complete formation show almost no difference in biofilm formation with the untreated samples (Figure 4 and Figure 3A). This data suggests that DTPA is not capable of altering the biofilm architecture and destabilizing intercellular accumulation once it has already formed.

*Mature amyloid fibers produced by the Accumulation-associated protein (Aap) are resistant to acid and chelator treatment* - Our original hypothesis for DTPA to act as a biofilm inhibitor is to chelate zinc, therefore preventing Aap self-association that mediates accumulation. Previous studies on a truncated version of the B-domain region (Brpt) of Aap revealed that this shorter construct formed a reversible Zn$^{2+}$-dependent dimer that was sensitive to a moderate decrease in pH (from 7.4 to 6.0) [3]. If the interactions of full length Aap were truly reversible, DTPA would be capable of disrupting mature biofilms. However, we demonstrated in Figure 3 that DTPA cannot disperse biofilms, suggesting that full-length Aap forms a more stable interaction. While amyloid fibers are not permanently assembled, they are characterized as a highly stable state of protein structure [149-151]. In order to test the stability of Zn$^{2+}$-induced amyloid fibers formed by the biologically relevant His-MBP/Brpt5.5 (5 tandem B-repeat domains and the half repeat cap of Aap), we allowed Brpt fibers to form in the presence of Zn$^{2+}$, followed by treatment with the chelator DTPA or incubation under acidic conditions. Samples of 10 µM His-
MBP/Brpt5.5 with 500 µM ZnCl₂ were incubated for 24 hours at either 20 °C or 37 °C prior to addition of sufficient dilute HCl to lower the pH to 5, addition of 2 mM DTPA, or addition of buffer as a control. After further incubation for 2 hours, each sample was loaded onto a C4 reverse-phase HPLC column without spinning out the fibrous aggregates (Figure 5, A and B). There were no significant differences in the relative distributions of monomeric versus oligomeric/soluble fiber species in the elution profiles for the 24-hour samples at either

![Figure 5, A and B](image)

samples of 10
120

or A) 20 °C or
ICl to lower the
led that
DTPA (green)
. Additional
days at either
|BP/Brpt5.5
creased fiber
temperature, indicating that incubation with DTPA or HCl did not induce any significant depolymerization of the oligomer/fiber species. Interestingly, however, the samples at both 20 °C and 37 °C did show changes within the oligomer/fiber peaks of the elution profiles, suggesting that while the fibers are stable, incubation with DTPA or HCl could result in limited remodeling of the oligomer/fiber morphology.

To assess the stability of mature amyloid fibers, a second set of His-MBP/Brpt5.5 samples was incubated with Zn$^{2+}$ for 4 days prior to treatment with DTPA or HCl (Figure 5, C).
and D). At 4 days, the 20 °C and 37 °C elution profiles differed drastically. The elution profile for the 20 °C sample resembled the 24 hour samples in terms of the monomer vs oligomer distribution. Likewise, the fibers were generally stable but showed evidence of limited remodeling upon incubation with DTPA or HCl. In contrast, the elution profile for the 37 °C sample showed that the oligomer/fiber peak predominated, and that it was more resistant to the action of DTPA or HCl. TEM was performed on each sample to confirm that fiber structure was maintained (Figure 6A and B). The TEM images of DTPA- or acid-treated samples show a single, rod-like fiber morphology that is consistent with previous characterization of Brpt fibers (Chapter 4). Furthermore, although some conformational rearrangement of the fibers may occur at lower temperatures and early time points [144], mature fibers at 37 °C are highly resistant to these environmental conditions.

These HPLC and TEM data indicate that the Brpt amyloid fibers are stable and not simply reversibly-associated oligomers as observed with the shorter (Bpt1.5) construct of the B-repeat region [3]. Therefore, to more thoroughly explore the stability of the protein-zinc interactions, we used analytical ultracentrifugation (AUC) analysis using a construct containing 3 B-repeats with the conserved half-repeat cap (Brpt3.5) of Aap (Chapter 2 and 4). 8 µM of Brpt3.5 was incubated with 0, 3 or 6 mM Zn\(^{2+}\) for 1 hour at 20 °C prior to beginning the centrifugation experiment at 48, 000 rpm. Data analysis using the software Sedfit revealed assembly of Brpt3.5 into multiple oligomeric forms, an approximate dimer (~4S) and tetramer (~5S) for the 3mM and 6mM samples, respectively (Figure 7A). Equivalent amounts of DTPA were then added to these original AUC samples, approximately 24 hours after the initial addition of zinc. After 1 hour of incubation with DTPA, the samples were re-centrifuged at 48, 000 rpm. Analysis of the data by Sedfit only shows a single peak around 2S, the sedimentation of Brp3.5
as a monomer (**Figure 7B**). The non-treated, apo-Brpt3.5 sample was the same concentration as the zinc and DTPA treated samples, and therefore should be the same height (absorbance) if the complexes were completely reversible. Because the zinc/DTPA treated sample peaks are significantly less than the non-treated Brpt3.5 sample, protein likely remained in large complexes that sedimented prior to data collection and are not present in the data analysis. In addition, the height of the DTPA-treated monomer at both zinc concentrations (**Figure 7B**) is also significantly less than even the of the original zinc-protein oligomer (**Figure 7A**). While we cannot directly conclude that Brpt3.5 remains in ordered complexes, such as amyloid fibers, the discrepancy of protein and HPLC studies support the formation of highly stable amyloidogenic complexes.
Figure 7. Sedimentation velocity analysis of Brpt3.5 in the presence of zinc and DTPA. Analytical ultracentrifugation data was collected at 20 C and 48,000 rpm. A. 8 µM Brpt3.5 was incubated with 0 (black), 3 (grey), or 6 (red) mM ZnCl₂ for 1 hour prior to the start of centrifugation. The graph shows the Sedfit analysis sedimentation coefficient distribution, indicating Brpt oligomer assembly (peaks at 4 and 5 S). B. 24 hours after the initial sedimentation, equivalent amounts of the chelator DTPA were added to the original zinc-treated samples and subjected to centrifugation again. Note that while all of the peaks, treated and untreated samples, all have the same sedimentation coefficient although are not equivalent in absorbance. This indicates that protein has remained out of solution, potentially in large protein assemblies.
Discussion

The primary focus of these studies was to connect the biophysical characterization of the Accumulation-associated protein (Aap) with the factors that influence Staphylococcus epidermidis biofilm formation. While it is well accepted that Aap is an essential component for accumulation [67, 68, 70], the precise mechanism by which Aap supports biofilm formation has remained unclear. Recently, zinc has been suspected to be the primary catalyst that induces Aap-mediated accumulation [3]. Based on this hypothesis, we have focused our analysis of Aap-mediated accumulation specifically on the relevant environmental factors that would influence a zinc-driven protein assembly mechanism, such as pH and the concentration of available transition metals.

The physiological concentration of zinc in human serum is roughly 15 – 20 uM, a drastic difference from the concentrations of zinc required to induce Brpt1.5 dimerization, 3 - 5 mM [3]. However, this initial characterization was performed using a truncated and non-biologically relevant protein construct of the B-repeat region of Aap. Because the B-domain is highly redundant, it was suspected that full-length Aap would exhibit cooperatively upon protein association that would require concentrations of zinc closer to the physiological availability. Regardless, the absolute concentration of zinc required for S. epidermidis biofilm formation has never previously been determined. In addition, copper has never been tested as a potential metal to induce Aap self-association, although it is a transition metal that has similar chemical properties and equivalent serum bioavailability as zinc. Our results suggest that while zinc does play an important role to support biofilm formation, it is not an absolutely required element. Weak and inconsistent biofilm formation by S. epidermidis was still observed under conditions
containing no zinc (Figure 1), although residual amounts of zinc could be leftover from the original colonization on enriched or complete CDM + RPMI media. *S. epidermidis* biofilm formation was maximized at 2 μM zinc, showing an unexpected decrease in biofilm formation as the concentration of zinc from increased 2 μM to 16 μM (Figure 1). In addition, copper was capable of supporting biofilm formation in the absence of zinc, although the biofilms were generally weaker than the biofilms grown under zinc. This data indicates that while zinc can drive biofilm formation, it is also tightly regulated and balanced with other metals to achieve maximum biofilm formation. This is supported by a previous study that shows silver can inhibit *S. epidermidis* biofilm formation [247], presumably by aberrantly replacing zinc for Aap-mediated accumulation.

While we show that zinc is not exclusively required for biofilm formation, the presence of metal chelators can profoundly inhibit *S. epidermidis* biofilm formation. In addition, the metal specificity of the metal chelator determined the efficacy to inhibit biofilm formation; chelators that were more specific to zinc (DTPA and TPEN) showed more efficient inhibition compared to chelators that had a higher specificity to alternative metals (Figure 3A). The chelator DTPA was found to be most effective to inhibit biofilm formation while not influencing cell viability. DTPA prevented *S. epidermidis* cellular aggregation and stacking that is required for a mature biofilm (Figure 3B), although the mode of DTPA inhibition is still unknown. If DTPA was simply binding all of the zinc so it was not available, biofilms should have been completely inhibited when grown in the absence of zinc (Figure 1). This suggests that DTPA could be acting as an amyloid inhibitor, small molecules that bind within the fibril core and destabilize the fiber structure or interrupt amyloidogenesis [221]. Other characterized amyloid small molecule inhibitors are also known metal chelators, supporting this hypothesis [221].
However, DTPA was also not capable of dispersing pre-formed biofilms (Figure 4). The fact that DTPA does not disperse pre-formed biofilms also supports that zinc is a structural component within the Aap amyloid fiber and not simply bridging nearby fibers or monomeric protein. The dense exopolysaccharide matrix has been shown to prevent entry of negatively-charged molecules into the biofilm, such as DTPA, due to charge-repulsion and could also explain this phenomenon.

The B-repeat region amino acid sequence of Aap contains multiple histidine, aspartic and glutamic acid residues that would potentially participate in zinc binding and be sensitive to changes in pH. In addition, zinc-mediated B-repeat dimerization has been shown to be abolished at pH less than 6.0 [3], but this data had never been confirmed biologically by S. epidermidis biofilm formation. Here, our results demonstrate a dependence on pH to form robust biofilms ideally at pH levels around 7.4 (Figure 2). Biofilm formation was inhibited at pH values less than 6.2, which correlates well with the Aap biophysical data and the involvement of histidine for zinc coordination (side chain $pK_a = 6.0$). However, S. epidermidis cell viability was not assessed for each pH level and acidic pH values have been shown to increase but not completely kill cells within a biofilm over time [246]. In addition, biofilm formation by S. epidermidis is not observed on the skin, where pH is typically slightly acidic, compared to the robust biofilms that are formed on blood-exposed medical devices and external catheters, where the pH is tightly buffered.

Aap has been recently been shown to form amyloid fibers in the presence of zinc (Chapter 4), which are highly stable protein structures. To understand the potential implications of DTPA and pH biofilm inhibition, amyloid fibers formed by the B-repeat region of Aap were subjected to similar conditions. Using highly purified Aap protein in solution, we show that
mature amyloid fibers are highly resistant to chemical stresses such as zinc chelation or acidic pH (Figures 5 and 6), suggesting that the amyloid fibers forming the intercellular network are one reason *S. epidermidis* biofilms are so resistant to harsh environmental conditions. This also explains why the addition of DTPA to pre-formed, mature staphylococcal biofilms does not disperse the biofilm (Figure 4). However, early amyloidogenic oligomers and fibril intermediates were found to be susceptible to some acid degradation (Figure 5A) and DTPA chelation (Figures 5A, 6, and 7B). This explains why *S. epidermidis* biofilms are sensitive to both acidic pH and DTPA in early stages of biofilm development (Figures 2 and 3), but not after the biofilm has been established. Amyloidogenesis does however occur rapidly, producing true amyloidogenic species within hours of incubation with zinc despite a majority of oligomeric species (Chapter 4). These same “fast-forming” amyloidogenic species are also found to be resistant to DTPA chelation (Figures 5A and 7B), although it is not known if the fiber morphology formed is biologically relevant for biofilm accumulation (Figure 6, Chapter 4). Amyloid fibers formed by the B-repeat region have been shown to form at least two distinct fiber morphologies, which is common for amyloid proteins [5], and fiber morphology can determine functionality [136]. Further studies of these amyloidogenic intermediates and stable fibers will need to be performed in order to resolve the physical nature of these species and their role in biofilm accumulation.

**Acknowledgements**

I would like to thank Andy Deng and the Tom Thompson lab for assistance with their HPLC equipment and Cincinnati Children’s Hospital Department of Pathology for allowing me to use the transmission electron microscope.
Chapter 6

Summary and Future Directions
Summary

Historically, *Staphylococcus epidermidis* and other coagulase-negative staphylococci were only considered as a commensal species and therefore were not a threat to host defenses [2, 11]. However, staphylococci have recently emerged as one of the leading bacterial pathogens worldwide [2, 24, 222], and have been implicated in a wide range of diseases such as bacteremia, urinary tract infections, infectious endocarditis, medical device-related infections, and skin infections [18, 24, 193]; all of which are primarily seen as chronic hospital-acquired infections mediated by the formation of a highly stable and chemically-resistant biofilm [16, 35]. A biofilm is a multi-layered, adherent cellular community that is encased within an extracellular matrix [34] and formation occurs via a three step process: primary attachment, accumulation, and secretion of the extracellular matrix components [2]. Intercellular accumulation provides the foundation of the biofilm and is mediated by two, distinct polysaccharide and protein-dependent pathways, although the protein-based mechanism has been suspected as a preliminary requirement for polysaccharide-mediated adhesion [81]. In addition, protein-dependent intercellular adhesion has been demonstrated to be solely sufficient to support biofilm formation [68] and protein factors are more prevalent than polysaccharide-dependent factors among clinical staphylococcal isolates [52, 70, 197].

A zinc-mediated mechanism has been established as the predominant mechanism of protein-dependent accumulation [3], demonstrating conservation of this mechanism in *S. epidermidis* and *S. aureus* [74] and it is hypothesized to be universal among all gram-positive cocci, biofilm-forming species. Given the broad utilization of this mechanism among biofilm-forming staphylococci, the primary focus of this thesis is to determine the detailed molecular...
mechanism of zinc-dependent intercellular adhesion and its universal implications for protein function for the future development of therapeutic treatments.

The protein responsible for zinc-mediated intercellular accumulation in *S. epidermidis* is the Accumulation-associated protein (Aap) [67]. Aap contains multiple functional domains, although only the B-domain has been shown to be important for biofilm formation [68]. The B-domain has a unique domain architecture, containing 5 – 17 almost identically repeated amino acid motifs; each repeat contains a G5 domain motif and a conserved 48 amino acid spacer region. A single B-repeat of Aap was shown to dimerize specifically in the presence of zinc, giving rise to the model for zinc-dependent biofilm formation [3]. However, the redundancy of this B-repeat region raises the question of why multiple, tandem B-repeats within Aap are needed for biofilm formation. A homologue of Aap in *S. aureus*, SasG, has demonstrated a minimum requirement of 5 tandem B-repeats to support biofilm formation, as well as a need for zinc-dependent assembly [71, 74]. Our studies indicate that Aap B-repeat self-association in the presence of zinc exhibits cooperative binding that inversely correlates with the number of B-repeats in tandem; increasing the number of tandem B-repeats decreases the concentration of zinc required to induce oligomer assembly. Binding cooperativity of the B-repeat region is likely to compensate for the low micromolar quantities of available zinc in serum. The B-repeat region of Aap also demonstrates a relationship between the number of tandem repeats and the self-assembled oligomeric complexes that can be formed in the presence of zinc; as the number of B-repeats in tandem increases, the formation of higher order species larger than a tetrameric complex is perpetuated. The majority of these large oligomeric complexes are highly elongated, fibril-like species that can exist as visible, insoluble aggregates.
Further characterization of these highly elongated protein complexes revealed that the B-repeat region of Aap forms amyloid fibers. Amyloid fibers have been found in a variety of other biofilm-forming bacterial species, both for pathogenesis and survival in harsh environments [184]. Amyloid fibers are characterized by their highly stable structure that is resistant to heat and chemical denaturants [128] and the use of amyloid fibers in the structure of bacterial biofilms would be advantageous to increase biofilm stability. Amyloid fiber formation by the B-repeat region is facilitated by zinc and is a time and temperature-dependent mechanism, in which the rate of fiber formation is increased at physiological temperature. B-repeat amyloid fibers also formed under in vitro conditions on a time course that is consistent with biofilm formation in *S. epidermidis*. In addition, we identified the presence of amyloid fibers embedded throughout *S. epidermidis* biofilms, although primarily concentrated at cell junctions between 6 and 12 hours that correlates well with intercellular accumulation. Amyloid fibers were also found on non-accumulated *S. epidermidis* cells prior to biofilm formation. Fibrils of Aap have been previously shown on planktonic *S. epidermidis* cells [75], although have not been identified as amyloid fibers. In addition, these fibers are primarily associated with commensal colonization [69]. Biofilm formation is strictly utilized for pathogenesis [25], suggesting that Aap could form different fibril phenotypes depending on the resulting function for native colonization or biofilm formation. Two distinct fibril phenotypes have been identified in our in vitro amyloid characterization studies of the B-repeat region that further supports this hypothesis.

The discovery of amyloid fibers within the *S. epidermidis* biofilm architecture provides clear evidence why biofilms are resistant to harsh environmental conditions and chemical denaturants. Mature *S. epidermidis* biofilms, as well as B-repeat amyloid fibers, were found to be resistant to degradation by the metal chelator DTPA. We did find, however, that both premature
biofilms and B-repeat amyloidogenic-intermediates were susceptible to inhibition or degradation by both acid and metal chelators. This suggests that amyloidogenic structural rearrangements occur over time to form a stable network of amyloid fibers within the biofilm. While biofilm formation was abolished in the presence of the metal chelator DTPA, biofilm growth was only decreased when zinc was removed from the environment. Biofilm formation is a multi-factorial event, relying on potentially several different protein and polysaccharide components; however, there is speculation that DTPA could be affecting biofilm formation more than by merely zinc chelation. In addition, our studies suggest that the environmental or intercellular concentration of metal is tightly regulated for maximal biofilm growth. In chemically-defined media, sub-physiological levels of zinc were found to have more biofilm growth compared to physiological concentrations. In experiments where zinc was the only transition metal present, biofilm formation was significantly decreased compared to biofilm growth in media containing zinc, copper, and cobalt.

We have focused our characterization of the G5 domain motif with respect to its role in zinc-mediated protein-dependent biofilm formation. However, to date this motif has been found in 213 unique bacterial species and 118 different protein domain architectures, suggesting that the G5 domain is not conserved for just biofilm formation. Rather, it is hypothesized that the G5 domain is conserved as a metal-binding adhesion module. Our studies with a unique G5 domain containing protein, a hypothesized zinc metalloprotease from Streptococcus sanguinis, indicate that the G5 domain does have the potential to be a universal metal-binding adhesion module. While these G5 domains also bind zinc, the mechanism of assembly and end product oligomers are different from the G5 domains from Aap, in that there is no cooperativity or dependency on the concentration of zinc that is observed for Aap. This suggests that not only the G5 domain,
but also potentially the spacer regions around this motif, determine the assembly complex. The spacer region has never been viewed as an important feature to the G5 domain motif and therefore has not been investigated previously. In addition, the G5 domains of Aap mediate amyloid fiber formation; therefore, the G5 domain may also be an amyloidogenic domain rather than just a metal-binding motif. The amyloid fold is believed to be an ancient protein fold that was originally utilized by lower organisms for stability [120]. The majority of the bacterial species that contain G5 domain proteins form biofilms and are inhabitants of harsh environments that would require highly stable proteins as well as biofilms for protection. Determining the relationship between G5 domain motifs and amyloidogenesis could provide insight to the function of amyloid proteins in nature.
Future Directions

The studies in this thesis have literally opened a “can of worms,” leading the research in an entirely unexpected and exciting direction. While our results have provided answers to long-standing questions about the mechanism of biofilm formation, even more questions have now been generated. There are three primary categories of questions that have been formulated from this work:

I. *The mechanism of biofilm formation in S. epidermidis* - How is amyloid fiber formation regulated and how is this related to biofilm formation in *S. epidermidis*? How do amyloid fibers participate in intercellular accumulation? Are the fibrils seen on planktonic *S. epidermidis* cells amyloid and what is the function of these fibers during native colonization? Are amyloid-specific inhibitors clinically effective to inhibit and disperse staphylococcal biofilms?

II. *Development of new techniques to study amyloid proteins* - Can a new analytical technique be developed that more accurately identifies amyloidogenic intermediate oligomeric species universally for all amyloid proteins?

III. *Insights into the native function of G5 domains across species* - Are G5 domains universally amyloidogenic domains and/or metal-binding adhesion modules? What other staphylococcal or gram-positive cocci bacterial species utilize amyloid fibers for biofilm formation and are they metal/ G5 domain mediated?

These questions will be the primary focus of our future work.
I. The mechanism of biofilm formation in S. epidermidis

Amyloidogenesis is considered a “high risk with big rewards” process. If achieved successfully, a highly stable protein structure is produced; however, there is a large amount of risk to incidentally produce aberrantly misfolded proteins that are toxic to the cell as an amyloidogenesis by-product [120]. Therefore, nature has provided mechanisms to tightly regulate this process to ensure that amyloid proteins are only produced extracellularly. Other bacterial amyloid proteins require additional factors to create amyloid fibers, such as chaperones or nucleating proteins [185, 186, 191]. However, there is no strong evidence of a regulatory mechanism in S. epidermidis to prevent Aap from spontaneously forming amyloid fibers, particularly since we were successful in producing amyloid fibers under in vitro conditions. One hypothesis is that the A and α/β domains could act as a latch, binding to the B-domain region to prevent early self-association. This is supported by the inhibition of biofilm formation when these domains are not cleaved from Aap [68]. Our MBP/Brpt5.5 construct, however, had no difficulty forming amyloid fibers despite the location of this fusion tag on the N-terminal portion of the construct in the same location of the A and α/β domains. This suggests that it is not simply blocking association by its position and could instead be interacting with the B-domain itself. This hypothesis is additionally supported by the fact that purified A domain also has some inhibitory effects on biofilm formation [68]. We plan on characterizing the A and α/β domains to determine if these domains can interact with the B-domain region of Aap and the effects of zinc to modify this interaction. In addition, we will generate a construct that contains a known cleavage sequence between the α/β domain and B-domain region and characterize this construct by its ability to form amyloid fibers, regulating the cleavage of the A and α/β domains with a known protease. Brpt protein constructs that do not contain the Hi-MBP fusion tag will also be
characterized for their ability to form amyloid fibers. This is to ensure that the tag is not favoring a particular fiber morphology since some forms of this tag are known to favor oligomeric assemblies.

We have demonstrated that amyloid fibers are utilized for intercellular accumulation and are present in a mature biofilm matrix of *S. epidermidis*; however, we have not directly identified that these fibers are composed of Aap. In addition, we do not understand the nature of how these amyloid fibers provide the intercellular interactions required for accumulation. Based on our previous TEM studies, our hypothesis is that Aap fibers from one cell intertwine with fibers from an adjacent cell, forming a strong interaction. Characterizing this event using microscopy could provide insight how accumulation is facilitated during pathogenesis. Structural studies using fiber diffraction of Aap fibers are also of interest in order to characterize the fibril core of Aap as well as provide insights into how Aap folds to form amyloid fibers.

As previously mentioned, fibers have been found on planktonic cells of *S. epidermidis* [75], although it has not been determined whether these are in fact amyloid fibers. During the characterization of His-MBP/Brpt5.5, we found that this construct was capable of producing low levels of amyloid fibers in the absence of zinc. In addition, two distinct fibril morphologies formed under in vitro conditions. Together, these factors suggest that there are potentially two distinct functions for amyloid fibers by *S. epidermidis*, for commensal colonization and pathogenesis. Because *S. epidermidis* is an opportunistic pathogen, and biofilm formation is detrimental to its native colonization, amyloidogenesis in the presence of specific concentrations of zinc could be the driving force for pathogenesis. The concentration of zinc in the epidermis is roughly 6-fold higher than the dermis and approximately 17-times higher than that of the blood [224]. However, His-MBP/Brpt5.5 amyloidogenesis is inhibited or limited at saturating
concentrations of zinc [unpublished data]. In addition, while the amount of zinc in the epidermis is higher than the blood, it is not necessarily bioavailable or at a pH that would allow zinc to bind Aap; We demonstrated that *S. epidermidis* biofilm formation is inhibited at a pH less than 6 and the pH of skin can vary between 4 and 6.5. *S. epidermidis* is both a commensal organism and a harmful pathogen, and determining the role of amyloidogenesis in these diverse roles could be critical in the treatment of infections caused by this species, as well as maintaining *S. epidermidis* as a member of the microflora.

As previously mentioned, Aap of *S. epidermidis* has recently been shown to form amyloid protein fibers that are believed to be the end product of zinc-mediated protein accumulation during biofilm formation. Amyloid fibers are defined by their canonical “cross-beta” structure that produces a similar structural conformation of the fibril core [4]. This defined structural epitope has been shown to bind a variety of small molecules that influence the overall fiber formation process, inhibiting as well as enhancing fibrillogenesis [221]. Therefore, if these small molecules are capable of inhibiting amyloidogenesis, we predict that they would additionally be able to inhibit *S. epidermidis* biofilm formation. Necula et al have characterized 29 small molecules that affect amyloid fiber formation [221]. Using this list, we selected five small molecules (curcumin, neocuproine, o – vanillin, thioflavin T, and eosin Y) to determine their inhibitory properties on *S. epidermidis* biofilm formation (Figure 1).

In order to evaluate the relative inhibitory properties of these amyloid-specific small molecules on *S. epidermidis* biofilm formation, we performed preliminary studies using the same semi-quantitative biofilm assay previously used with the metal chelators. All of the amyloid-specific small molecules tested had robust inhibitory effects, with no effect by the ethanol vehicle control. Curcumin was the most effective small molecule to inhibit biofilm formation,
requiring only 10 µM of curcumin to cause biofilm inhibition. All of the other small molecules, except thioflavin T, demonstrated complete biofilm inhibition at 20 µM; thioflavin T only inhibited at concentrations greater than 20 µM. These results confirm that Aap amyloidogenesis is required for *S. epidermidis* biofilm formation and that these small molecules have the potential

**Figure 1. *S. epidermidis* biofilm inhibition by amyloid-specific small molecules.** A. Dose-response curve of *S. epidermidis* strain RP62a cultures treated with the amyloid inhibitors curcumin (blue diamonds), neocuproine (red squares), o-vanillin (green triangles), thioflavin T (purple x), or eosin Y (cyan stars). Each inhibitor was added to the 96 well plate prior to inoculation with 1:200 diluted (TSB) RP62a. Biofilm formation was measured using a semi-quantitative assay method, measuring crystal violet staining at 570 nm. A 10% ethanol (used to solubilize each chelator) vehicle control was also generated, which had no significant differences from the wild-type biofilm. Total (100%) biofilm formation was based on an average of 10% ethanol treated, RP62a biofilms. Error is based on 3 replicates on 3 different days.
to be utilized to inhibit this process. Future studies will include studying the effect of these amyloid specific inhibitors on cell viability, determining MIC values, and using in vivo animal models of infection. In addition, there is significant potential for these amyloid-specific small molecules to be used against other biofilm-forming, gram-positive cocci. Future studies will also characterize the inhibitory properties of these small molecules against other bacterial species such as *Staphylococcus aureus*, *Streptococcus suis*, and *Streptococcus oralis*; all of which are opportunistic, biofilm-forming pathogens.

II. *Development of new techniques to study amyloid proteins*

During our early analyses of the B-repeat protein constructs using analytical ultracentrifugation (AUC), it became evident that protein fibers and highly elongated proteins exhibit non-ideality during sedimentation and pose much difficulty during data analysis. A new sedimentation model that specifically separates proteins based on their shape (frictional ratio) has recently been developed by Peter Schuck to be used with the analysis software Sedfit. While there have been no previous publications characterizing this mechanism, our studies in this thesis (Chapter 4) utilized this sedimentation model and found a novel method for distinguishing early amyloidogenic species. This method of analysis provided insights to two distinct methods of assembly for the two mature fibril morphologies. Many amyloid proteins have multiple, distinct amyloid morphologies that can potentially determine the disease state [136]; and it is therefore of great interest to understand the early amyloid oligomers that lead to pathogenic vs. non-pathogenic states. In future studies, we will explore the assembly of other amyloid proteins using this method of analysis to determine if AUC can be used as a universal tool to characterize amyloid assembly. In addition, we plan to develop other experiments using the spectral
properties of amyloid-specific small molecules, such as thioflavin T, in order to identify the critical nuclei and the time point at which these proteins become truly amyloid during the assembly process.

III. Insights into the native function of G5 domains across species

G5 domains were originally hypothesized to function in the recognition of N-acetylglucosamine [72], this supposition was modified by the discovery that a G5 domain containing protein self-associates in the presence of zinc [3]. This research formed a new hypothesis that G5 domains were zinc-binding adhesion modules, primarily playing a role in biofilm formation by gram-positive cocci. This assumption, however, was based on the protein annotation in the Pfam database in 2008. Currently, the majority of species expressing G5 domains still fit this profile, although almost 40% do not. This suggests alternative functions for G5 domains outside of biofilm formation. In addition, 19 gram-negative species expressing G5-domain proteins have been identified since this initial analysis, provoking more curiosity about the conservation of the G5 domain as a metal-binding domain. Our recent results also suggest that the G5 domain could also be an amyloidogenic domain, particularly since this domain is found primarily in bacterial inhabitants of extreme environmental conditions. Our future studies will aim to characterize more diverse G5 domain architectures from diverse bacterial species as both metal-binding adhesion modules and amyloidogenic domains. In addition, our recent work characterizing unique G5 domains from Streptococcus sanguinis suggests that the amino acid spacer region between tandem G5 domains regulates self-association in the presence of zinc. Therefore, future studies will also specifically address the ability of the spacer region to modify the activity of these diverse domains.
Concluding Remarks

Humans utilize the microbial normal flora to provide a natural barrier from other evasive and harmful species. However, as medical technology and antibiotic therapies have advanced, there has been a disruption in the natural balance that defines commensal colonization and infection by these native microbial species. The primary colonizer of human epithelial surfaces, *Staphylococcus epidermidis*, is now one of the leading causes of opportunistic infections worldwide. It is therefore critical to understand how opportunistic pathogens cause infection in order to regain control in the natural balance of this delicate, mutualistic relationship. Our recent discovery of amyloid fibers within *S. epidermidis* biofilms demonstrates that there is still much to be learned about this mechanism, and in particular, the environmental and bacterial factors that control amyloidogenesis and biofilm formation. Further studies of Aap amyloid fiber formation will not only aid in the development of new therapeutics to inhibit and disperse *S. epidermidis* biofilms, but also provide novel insights into larger proteins that are capable of forming amyloidogenic species; previous studies have primarily focused on the characterization of amyloidogenic small peptides. In addition, little is known about *S. epidermidis* in its native commensal state and characterizing amyloid fibers on planktonic cells would provide information about its natural function in the normal flora. Furthermore, determining the degree of species conservation of zinc-mediated amyloidogenesis for biofilm formation would provide critical information for future clinical applications, as multiple other gram-positive commensal species are on the rise as opportunistic pathogens. Understanding the mechanisms that promote opportunistic infections by our own commensal inhabitants is imperative to reduce the morbidity and mortality caused by these species world-wide.
Appendix I:

Structural Analyses Suggest New Roles for Calcium in the Enzymatic Mechanism and Dimerization of Human Calcium-Activated Nucleotidases*

*The following text is in preparation for publication:


Author contributions: SLJ performed all protein crystallography and AUC experiments. TLK performed mutational analysis and enzymatic activity studies. JW advised crystallography and analyzed crystallographic data. MY expressed and purified WT and I170K protein. SLJ, TLK, and ABH wrote the paper.
Abstract

Blood-sucking insects secrete nucleotidases that prevent host blood clotting by hydrolyzing ADP, a platelet agonist, to AMP in the blood. Mammals also express a homologous, calcium-activated nucleotidase (CAN or solubly expressed SCAN), however, the mammalian ortholog prefers GDP to ADP as a substrate and therefore is not efficient at preventing thrombosis and its native function is unknown. Calcium is required for CAN catalytic activity in addition to inducing dimerization, which additionally enhances activity, although the role for calcium in these mechanisms has yet to be revealed. Therefore, we have introduced a point mutation at a key residue in the dimer interface, Ile170Lys (I170K), to produce an obligate monomer species for crystallographic analysis. I170K was crystallized in the presence of calcium, resulting in the formation of two distinct, but corresponding, crystal forms; these crystals diffracted to 1.6 and 1.8 Å and solved by molecular replacement, revealing novel calcium binding sites within the active site. I170K and Wild-type SCAN were additionally crystallized with the non-hydrolyzable substrate GMPCP in the presence of calcium and solved by molecular replacement. Structural alignments of the new SCAN structures with substrate, I170K and wild-type, implicate a new residue (Arg113) that regulates dimerization and dimer-enhanced activity. We generated mutation of this residue (R113A) that maintains dimerization with reduced requirements of calcium but does not experience dimer-enhanced activity, confirming our hypothesis. The wild-type SCAN structure also shows that calcium is required for substrate coordination and suggests a “two-metal ion” mechanism of catalysis; this is the first model of catalytic mechanism for CANs.
Introduction

Stroke affects 795,000 Americans each year and cardiovascular related diseases are responsible for approximately one third of all deaths in the United States [245]. Platelets are required to maintain vascular hemostasis, however, they are additionally a major contributor to stroke and cardiovascular disease due to aberrant activation and aggregation that lead to thrombotic occlusions [225]. Mammals regulate platelet activation through a membrane-bound ectonucleotidase (CD39) that enzymatically de-phosphorylates ADP, the primary platelet receptor agonist [225]. Current drug therapies target nucleotide receptors on platelets required for activation, which causes unfavorable blood complications due to non-responsive platelets. Therefore, therapeutic drug therapies that prevent extraneous platelet aggregation while maintaining hemostasis are of considerable interest.

Mammals also express a nucleotidase ortholog, the calcium-activated nucleotidase (CAN). Human CAN functions for nucleotide hydrolysis, although prefers GDP to ADP as a substrate and located intracellularly [226], and is therefore not involved in hemostasis [227]. CAN does exhibit 48% amino acid sequence identity to the soluble apyrases of blood-sucking insects that demonstrate highly efficient ADP hydrolysis [228], and is capable of being released extracellularly in a soluble form (SCAN). Therefore, elucidating the mechanism of nucleotide hydrolysis of human CAN and identifying residues for efficient ADP hydrolysis are both critical in order to use human CAN as a therapeutic agent to prevent blood clotting.

As implied, calcium is required for catalytic activity; however, it is additionally required as a key structural component for tertiary structure folding and stability [227, 229]. Previously, we determined that soluble CAN (SCAN) has substantially higher enzymatic activity as a dimer [230]. Analytical ultracentrifugation analysis determined that both calcium and strontium will
induce SCAN dimerization, although only calcium-induced dimerization supports nucleotidase activity [230, 231]. In order to elucidate key residues for CAN function, Dai et al and our laboratory previously generated crystal structures of apo-SCAN (PDB codes 1S18 and 2H2N, respectively) [231, 232]. Both structures show a conserved, structurally central calcium coordinated by residues aspartate 99, serine 275, glutamate 214, glutamate 326, serine 98, and glutamate 145 [231, 232] (See Table 1 for sequence numbering conventions and conversions); however, neither structure provides any insight into why calcium is required for hydrolysis. Structures 1S18 and 1S1D (wild type SCAN co-crystallized with a non-hydrolyzable substrate) are both high resolution structures, although crystallized in the presence of strontium rather than calcium [232]. Structure 2H2N, which was crystallized in the presence of calcium although density for calcium could not be resolved nor was anomalous scattering for calcium found [231].

In order to elucidate the mechanism of calcium-induced dimerization, and therefore allosteric changes that increase enzyme activity, we have generated a series of SCAN crystal structures in monomer and dimer states in the presence of calcium. To produce an obligate monomer SCAN species, a point mutation was introduced at a key residue within a hydrophobic pocket of the dimer interface, Ile170Lys (I170K) [230]. This mutation was previously characterized as a monomer in solution in the presence of calcium by analytical ultracentrifugation [230]. Electron density maps of our new substrate-bound, wild-type SCAN and I170K SCAN with and without substrate structures in the presence of calcium have identified novel calcium binding sites within the substrate binding pocket. Using our new wild-type SCAN structure with calcium, we have been able to identify a new residue implicated in the regulation of dimer-induced allosteric changes, Arg113. In addition, we show for the first time
that calcium is required for substrate coordination and have proposed a model “two-metal ion” mechanism of catalysis.

### Table 1. Numbering Conventions in Current SCAN Structures

<table>
<thead>
<tr>
<th>Human SCAN</th>
<th>Residue number in crystal structures of current publication and 2H2N</th>
<th>1S1D/1S18</th>
</tr>
</thead>
<tbody>
<tr>
<td>D72</td>
<td>42</td>
<td>44</td>
</tr>
<tr>
<td>D74</td>
<td>44</td>
<td>46</td>
</tr>
<tr>
<td>A129</td>
<td>89</td>
<td>91</td>
</tr>
<tr>
<td>E130</td>
<td>90</td>
<td>92</td>
</tr>
<tr>
<td>E136</td>
<td>96</td>
<td>98</td>
</tr>
<tr>
<td>S138</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>D139</td>
<td>99</td>
<td>101</td>
</tr>
<tr>
<td>D142</td>
<td>112</td>
<td>114</td>
</tr>
<tr>
<td>R143</td>
<td>113</td>
<td>115</td>
</tr>
<tr>
<td>I170</td>
<td>130</td>
<td>132</td>
</tr>
<tr>
<td>K183</td>
<td>143</td>
<td>145</td>
</tr>
<tr>
<td>E185</td>
<td>145</td>
<td>147</td>
</tr>
<tr>
<td>K224</td>
<td>184</td>
<td>186</td>
</tr>
<tr>
<td>D228</td>
<td>188</td>
<td>190</td>
</tr>
<tr>
<td>E254</td>
<td>214</td>
<td>216</td>
</tr>
<tr>
<td>R270</td>
<td>230</td>
<td>232</td>
</tr>
<tr>
<td>S315</td>
<td>275</td>
<td>277</td>
</tr>
<tr>
<td>E335</td>
<td>295</td>
<td>297</td>
</tr>
<tr>
<td>K364</td>
<td>324</td>
<td>326</td>
</tr>
<tr>
<td>E366</td>
<td>326</td>
<td>328</td>
</tr>
</tbody>
</table>
Materials and Methods

Materials. The QuikChange site-directed mutagenesis kit and *Escherichia coli* competent bacteria were purchased from Stratagene. I170K protein was sequenced by the K. Greis laboratory at the University of Cincinnati. The 96-well screens, PACT and index, for crystallography were purchased from Qiagen and Hampton Research, respectively. The non-hydrolyzable substrate GMPCP was purchased from Jena Bioscience.

Site-Directed Mutagenesis and Protein Purification. The I170K SCAN mutant was generated with the Stratagene QuikChange site directed mutagenesis kit [230]. Both the wild type and I170K mutant SCAN were expressed and purified as previously described (Yang, Biochemistry, 2008). Briefly, the I170K construct was expressed in *E. coli* expression cell line BL21(DE3) cells, inducing with IPTG. Inclusion bodies containing the SCAN protein were isolated, denatured, and refolded. The protein was purified by an N-terminal hexa-histidine tag, thrombin-cleaved to remove the tag, and further purified by anion-exchange chromatography. The I170K SCAN was further treated with clostripain digestion in order to remove an N-terminal flexible region at arginine 8 for crystallization. Purified I170K SCAN (0.2mg/ml) was digested with 0.4 µg/ml (1:500) clostripain in 20mM MOPS pH 7.4 buffer for 2 hours at 21°C. The digested products were re-purified over an anion exchange column. The clostripain digested I170K was tested for nucleotidase activity as previously described (Biochemistry 2008) and was unchanged from non-clostripain digested I170K. The cleavage site was confirmed using liquid chromatography-mass spectrometry. Protein purity was assayed by SDS-PAGE.

Crystallization and Data Collection of Wild Type SCAN with Substrate. Crystallization was performed using a hanging drop vapor diffusion method at 21°C in 24 well trays. Crystallization
conditions were similar to the previous wild type SCAN conditions (2H2N [231]). 1 µl of protein (6.16 mg/ml in 20mM MOPS pH 7.4, 100 mM NaCl) was mixed with 1 µl well solution, 0.2M ammonium sulfate, 0.1M sodium acetate pH 4.8, 5mM calcium chloride, and 14% PEG 4000. Crystals formed within 24 hours and were then transferred to a soaking solution with identical mother liquor conditions and additionally containing 3mM guanosine-5’-[(α,β)-methyleno]diphosphate (GMPCP), a non-hydrolyzable substrate analog, 10% PEG 1500, and a reduced calcium chloride concentration (2mM) to prevent substrate precipitation. The crystals soaked for 2 hours at 21°C, then transferred briefly to a cryoprotectant solution with 10% glycerol added to the soaking solution composition (without substrate) and flash frozen in liquid nitrogen. The data was collected at Argonne National Laboratory Advanced Photon Source beamline NE-CAT 24-ID-E. The crystals diffracted to 1.7 Å and belong to space group P1 with two molecules in the asymmetric unit (a = 43.42 Å, b = 53.09 Å, c = 77.19 Å; α = 74.63°, β = 74.75°, γ = 80.44°).

Crystallization and Data Collection of I170K SCAN. Initial crystallization of apo-I170K SCAN was performed using a sitting drop vapor diffusion method in 96 well plates and screened for optimal conditions using the PACT and index reservoir buffer suites. Apo-I170K SCAN crystallization conditions were optimized using a hanging drop vapor diffusion method in 24 well trays at 21°C for two crystal forms. Form 1 crystallized using 1 µl well solution (0.1M MES pH 6.5, 50mM calcium chloride, and 14% PEG) and 1 µl of protein ([I170K] = 6.9 mg/ml). Form 2 crystallized using similar conditions by mixing 4 µl of 40% n-propanol with 4 µl well solution (0.1M MES pH 6.5, 100mM calcium chloride, and 18% PEG 6000), then adding 1 µl of this solution to 1 µl protein ([I170K] = 3.7 mg/ml). Each crystal form was placed into a cryoprotectant solution containing 25% ethylene glycol or 20% glycerol added to the mother
liquor, respectively, and flash frozen in liquid nitrogen. Data was collected at Argonne National Laboratory Advanced Photon Source beamline GM/CA-CAT 23-ID-D. The crystals diffracted to 1.88 Å ($P_2_1$, 4 molecules in the asymmetric unit, unit cell dimensions $a = 79.63$ Å, $b = 79.38$ Å, $c = 107.79$ Å; $\alpha = 90.00^\circ$, $\beta = 103.04^\circ$, $\gamma = 90.00^\circ$) and 1.6 Å ($P1$, 2 molecules in the asymmetric unit, unit cell dimensions $a = 44.33$ Å, $b = 57.51$ Å, $c = 68.43$ Å; $\alpha = 77.76^\circ$, $\beta = 78.18^\circ$, $\gamma = 73.72^\circ$). I170K SCAN substrate bound crystals formed under similar conditions as the apo-I170K crystals (0.1 M HEPES pH 6.8, 0.2 M calcium chloride, and 25% PEG 6000) using a co-crystallization method by mixing I170K [3.8 mg/ml] with GMPCP [2.6 mM]. Crystals were cryoprotected by adding 10% glycerol to the mother liquor and flash frozen in a cold nitrogen stream maintained at 110 °K. Data was collected on the home source beamline using a Rigaku Micromax-007 generator and an R-Axis IV++ image plate detector. The crystals diffracted to 2.6 Å (resolution cut-off to 3.0 Å for completeness) and belong to space group $P1$ with 2 molecules in the asymmetric unit (unit cell dimensions $a = 44.57$ Å, $b = 58.14$ Å, $c = 68.91$ Å; $\alpha = 77.01^\circ$, $\beta = 78.00^\circ$, $\gamma = 73.23^\circ$).

Data Processing, Refinement, and Analysis. The wild type and I170K SCAN with substrate and both apo-I170K SCAN data sets were processed using HKL2000. For all data sets, phases were obtained using molecular replacement with Phaser1.3.3, using the wild-type CAN structure crystallized with calcium as the search model (Protein Data Bank code 2H2N). All refinement was performed using Refmac5 of the CCP4i suite and TLS models generated from the TLS Motion Determination server. Structures were validated using MolProbity. Structure alignments and atomic measurements were performed using coot 0.6.2. Binding pocket analysis was performed using the HotSpot Wizard server. Figures were generated using Pymol.
Generation and Characterization of R113A SCAN mutant - site-directed mutagenesis of WT SCAN in PET28a bacterial expression vector was performed as described previously, using the QuikChange II site-directed mutagenesis kit. The sense oligonucleotide used for R113A mutagenesis was, with the substitution sites underlined and bolded (corresponding required antisense oligonucleotide not shown):

5’-CTACTCCGTGGATGACGCGACGGGGGTCGTCTAC-3’,

encoding the amino acid sequence, LYSVDDA113TGVYQ.

The R113A cDNA was used to transform BL21(DE3) E. Coli, and after induction of expression with IPTG, bacterial inclusion bodies containing the SCAN protein were purified as described previously. 12 mg of the inclusion body protein was denatured and refolded, purified via its N-terminal hexa-histidine tag, thrombin cleaved to remove the N-terminal tag, and further purified by anion exchange chromatography as described previously to yield 6.5 mg of purified R113A SCAN protein. The purified protein was quantified by 280 nm absorbance (using an extinction coefficient of 8.14 X 10⁴), and its purity confirmed by SDS-PAGE.

Nucleotidase activity was determined using a modification of the technique of Fiske and Subbarow as previously described [233]. Prior to assay, SCAN protein was diluted to the appropriate concentration with either Tris-buffered saline (TBS) or TBS containing 0.05% Tween 20 (Tween detergent added when making very dilute protein solutions of SCAN, to minimize protein adsorption to surfaces. Nucleotide hydrolyzing units are expressed in micromoles of Pᵢ liberated per milligram of protein per hour. Generally, assays were conducted in 20 mM MOPS, pH 7.4, containing 5 mM CaCl₂ and 2.5 mM nucleotide substrate at 37°C. Assays to measure the effect of protein concentration on ADPase activity were performed at
22°C to allow comparison with previously published experiments [230] (note that data for wt SCAN ADPase activity at 22°C shown in the current study is reproduced from Figure 3 with permission of publisher).

Protein intrinsic tryptophan fluorescence, a measure of Ca-induced dimerization, was measured as described, using a Hitachi F-2000 fluorescence spectrophotometer. A 0.85 μM (32 μg/mL) concentration of SCAN in 20 mM MOPS, pH 7.4 was excited at 295 nm, and fluorescence emission was recorded at its maximum (340 nm).

Sedimentation velocity experiments to determine dimerization status were carried out at 20°C in a Beckman XL-I analytical ultracentrifuge using absorbance optics at 230 nm. Samples of 0.85 μM of the R113A SCAN mutant (20 mM MOPS (pH 7.4)) were centrifuged at 48,000 rpm in the presence or absence (apo) of 25 μM CaCl₂. Data were analyzed using the program SEDFIT [234] and the c(s) distribution model. Analyzed data for wild-type SCAN (with and without calcium) were borrowed with permission and performed as previously described [231].
Results

**Crystallization of wild-type and I170K CAN in the presence of calcium and the substrate GMPCP** - Previous crystal structures of the soluble, human Calcium-Activated Nucleotidase (SCAN) have provided critical information for enzymatic function, demonstrating key residues for substrate specificity and catalysis [232] and residues important for dimerization [231]. However, while calcium has been shown to be essential for enzymatic activity [227, 228, 235] and dimerization [231], there is no current crystallographic evidence to suggest how calcium participates in either of these processes. In addition, mutant forms of SCAN that cannot dimerize, such as the Ile170Lys (I170K), demonstrate a marked reduction of enzymatic activity for ATP and ADP (40% and 35% of wild-type activity, respectively) [230], indicating there is an enzymatic advantage to calcium-induced dimerization. Therefore, we crystallized both the wild-type SCAN and I170K SCAN monomer mutant in the presence of calcium and the non-hydrolyzable substrate analog GMPCP to determine the contributions of calcium to enzymatic activity and dimerization. Wild-type SCAN crystals were grown under similar conditions to our previous conditions for 2H2N, containing 5mM calcium chloride. These crystals were then soaked with GMPCP because attempts to co-crystallization the protein-substrate complex were not successful. The wild-type CAN with GMPCP crystals diffracted to 1.7 Å and contain two SCAN monomers in dimeric form, demonstrating distinct density for the GMPCP substrate in the active site (**Table 2, Figure 1A**).
### Table 2. Summary of crystallographic data

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>WT + GMPCP</th>
<th>Apo-II70K (a)</th>
<th>Apo-II170K (b)</th>
<th>II170K + GMPCP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
<td>APS NE-24ID-E</td>
<td>APS GM/CA 23ID-D</td>
<td>APS GM/CA 23ID-D</td>
<td>Rigaku RA-Micro007</td>
</tr>
<tr>
<td><strong>Wavelength (Å)</strong></td>
<td>0.97918</td>
<td>1.0332</td>
<td>1.0332</td>
<td>1.5418</td>
</tr>
<tr>
<td><strong>Space group</strong></td>
<td>P1</td>
<td>P1</td>
<td>P1</td>
<td>P1</td>
</tr>
<tr>
<td><strong>Unit cell a, b, c (Å)</strong></td>
<td>43.42, 53.09, 77.19</td>
<td>79.63, 79.38, 107.79</td>
<td>44.33, 57.51, 68.43</td>
<td>44.57, 58.14, 68.91</td>
</tr>
<tr>
<td><strong>α, β, γ (°)</strong></td>
<td>74.63, 74.75, 8.044</td>
<td>90.00, 103.64, 90.00</td>
<td>77.76, 78.18, 73.72</td>
<td>77.01, 78.00, 73.23</td>
</tr>
<tr>
<td><strong>Resolution Range (Å)</strong></td>
<td>72.32 - 1.70</td>
<td>41.63 - 1.88</td>
<td>38.16 - 1.60</td>
<td>19.93 - 2.60</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>95.1(83.5)</td>
<td>95.0(85.0)</td>
<td>98.1(91.3)</td>
<td>95.6(79.8)</td>
</tr>
<tr>
<td><strong>Reflections (observed/unique)</strong></td>
<td>135441/69456</td>
<td>217283/106004</td>
<td>159014/79509</td>
<td>23767/12175</td>
</tr>
<tr>
<td><strong>Average Redundancy</strong></td>
<td>1.7(1.5)</td>
<td>2.6(3.2)</td>
<td>2.2(2.1)</td>
<td>1.9(1.9)</td>
</tr>
<tr>
<td><strong>Average I/σ(I)</strong></td>
<td>11.4(2.9)</td>
<td>9.2(2.3)</td>
<td>21.0(3.8)</td>
<td>5.5(3.2)</td>
</tr>
<tr>
<td><strong>R_{merge} (%)</strong></td>
<td>13.0(35.8)</td>
<td>7.9(41.6)</td>
<td>3.9(20.9)</td>
<td>11.4(23.4)</td>
</tr>
<tr>
<td><strong>B-factor</strong></td>
<td>17.21</td>
<td>19.93</td>
<td>16.82</td>
<td>41.9</td>
</tr>
<tr>
<td><strong>% Solvent</strong></td>
<td>43.50%</td>
<td>43.96%</td>
<td>45.20%</td>
<td>43.67%</td>
</tr>
</tbody>
</table>

#### Model refinement

| **Resolution Range (Å)**     | 72.32 - 1.70 | 41.63 - 1.88 | 38.16 - 1.60 | 19.93 - 3.00 |
| **Reflections (work/test)**  | 62643/3341   | 106663/5311  | 75543/3754   | 11572/603    |
| **R_{work}/R_{free} (%)**    | 18.5/21.2    | 18.1/21.9    | 16.8/19.7    | 21.3/28.0    |
| **Root mean square deviations from ideality** |                           |                           |                           |                           |
| **Bond lengths (Å)**         | 0.006        | 0.008        | 0.011        | 0.006        |
| **Bond angles (°)**          | 1.01         | 1.08         | 1.34         | 1.12         |
| **Number of residues in molecule (A/B/C/D)** | 316/314      | 316/317/316/314 | 316/317     | 317/316      |
| **Total atoms**              | 5562        | 11170        | 5768         | 5178         |
| **Calcium ions**             | 11          | 13           | 7            | 4            |
| **GP2**                      | 2           | 0            | 0            | 2            |
| **Water molecules**          | 472         | 1026         | 580          | 16           |
| **Other**                    | NH₄ (1), ACT (1) | POL(20), GOL(2), | MES(2)        | MES(5)       |

#### Ramachandran Plot

| **Residues**                 | 623 residues, 99.2% | 1241 residues, 98.6% | 617 residues, 97.8% | 612 residues, 97% |
| **Outliers**                 | 5 residues, 0.9%    | 22 residues, 2.1%    | 8 residues, 1.5%    | 21 residues, 3.9% |
| **Outliers**                 | 0 residues, 0%      | 0 residues, 0%       | 0 residues, 0%      | 0 residues, 0%    |
Figure 1. Substrate coordination by WT and I170K SCAN.  A. Wild-type SCAN with GMPCP bound, displaying a 2Fo – Fc map contoured to 1.2 sigma. Calcium is shown as red spheres.  B. Coordination of GMPCP by I170K SCAN, displaying density of 2Fo-Fc map contouring to 1.0 sigma. Calcium is shown as green spheres. The α-phosphate is rotated in the binding pocket, as shown by the distance to the two primary residues that coordinate this position, Glu214 and Glu145.
In order to first determine the structural differences of SCAN as a monomer versus a dimer, we additionally crystallized the obligate monomer I170K SCAN mutant in the presence of calcium without the substrate, producing two unique crystal forms. Both crystal forms produced high resolution data, diffracting to 1.6 Å (P1 form, 100mM calcium) and 1.88 Å (P2₁, 50mM calcium) and the structures were solved by molecular replacement using the coordinates of the apo-wild-type SCAN structure with calcium (2H2N) (Table 2). Both the P1 crystal form (2 protein monomers) and P2₁ crystal form (4 protein monomers) are almost identical to the previously published CAN structures (1S18 and 2H2N), with no major conformational changes to the overall protein structure or fold. One I170K SCAN monomer within the P2₁ form does have a unique confirmation of the flexible loop (residues 162-171) that has previously been shown to interact with the nitrogen base of the nucleotide substrate. Previous structures, both substrate bound or unbound, have only shown this loop in a “closed” position, whereas this structure indicates that the loop is capable of opening via a hinge-like mechanism to allow the substrate to enter the phosphate binding pocket. I170K was also successfully co-crystallized with GMPCP (3.8 mg/ml I170K with 2.6 mM GMPCP) with space group symmetry P1 and two monomers in the asymmetric unit. However, these crystals only diffracted to 2.6 Å and the data resolution was cut off to 3.0 Å for completeness (Table 2). While this is not as high of resolution as the wild-type SCAN/substrate and apo-I170K SCAN structures, there was distinct density for the substrate in the binding pocket that was not present in other I170K substrate complex crystals (Figure 1B).
Central calcium coordination is effected by dimerization - The central calcium of SCAN is an important structural component of the enzyme that supports the native five-blade β-propeller fold [232]; chelation of this calcium results in SCAN thermal instability and increased susceptibility to proteolysis [229]. Therefore, it was predicted that the residues coordinating the central calcium (Asp99, Ser275, Glu214, Glu326, Ser98, Glu145, and water) would be maintained in the apo-I170K SCAN structure. Upon substrate binding, there is a conformational change of central calcium coordination that has previously been shown in the 1S1D SCAN structure [232], in which the side chain of Asp99 shifts away from the central calcium towards Leu147. However, 1S1D was crystallized in the presence of strontium, which induces dimerization but does not support enzymatic activity [231]. Our structure of wild-type SCAN that was crystallized with calcium in the presence of GMPCP also shows a small 0.63 Å shift in the Asp99 side chain away from the calcium; however, this shift is not as exasperated as that seen in the 1S1D and is not close enough to form a hydrogen bond with the carboxyl oxygen of Leu147 (Figure 2A).

The obligate monomer mutant I170K SCAN in the presence of GMPCP displays no shift in the Asp99 side chain and instead loses heptahedral coordination with water to assume hexahedral calcium coordination during substrate binding (Figure 2B). This suggests that the shift in central calcium coordination in the presence of substrate is a result of SCAN dimerization, Asp99 acting as a “communicative switch” to residues of the phosphate binding pocket to coordinate the substrate and alter enzymatic activity with respect to the dimer confirmation. Because 1S1D was crystallized with strontium, and therefore as an enzymatically inactive dimer, this hyper-shift of Asp99 could be responsible for the lack of enzymatic activity in the presence of strontium, particularly since it is in close proximity with residues of the
Figure 2. A shift in central calcium coordination in the presence of substrate. A. Structural alignment of apo-wild-type 2H2N (Blue) and in the presence of substrate (this paper, Green, water = red) or the 1S1D structure (Yellow, water = orange). Calcium is colored by the respective colors for each structure. There is a noticeable rotation in the Asp99 rotamer towards L147, although this is more pronounced by 1S1D. The comparison of I170K (B, violet) shows no rotation of Asp99, but instead a shift in the entire central calcium binding site.
phosphate binding pocket. In addition, Glu145, Glu214, and Glu326, exhibit a rotamer shift upon substrate binding in the wild-type CAN structure that is absent in the residues of the I170K structure (Figure 2B). Instead, there is a minor shift in the central calcium towards the binding pocket, but all residue contacts are maintained. These residues, particularly Glu145, have been previously shown to be important for enzyme activity [232] and changes to the central calcium coordination could be critical to communicate the allostERIC changes that are associated with dimerization.

Apo-I170K CAN coordinates calcium in the phosphate-binding pocket - SCAN enzymatic activity and substrate binding have previously been shown to be strictly dependent on the presence of calcium [229], however, there has been no structural evidence to explain or support this biochemical data. The first apo-SCAN crystal structure, 1S18, was crystallized with strontium, and the second apo-SCAN structure, 2H2N, while crystallized with calcium was too low in resolution to determine additional calcium binding sites. Here, our apo-I170K SCAN crystal structures (two distinct crystal forms) display two additional and unique calcium binding sites within the phosphate binding pocket with significant density for the calcium ion (Figure 3). The additional calcium binding sites within the enzymatic active site are coordinated by the same residues in all six I170K asymmetric unit monomers of the two unique crystal forms with little bond length deviation and considerably low local B-factors, suggesting that these calcium binding sites are important for the non-substrate bound form of the enzyme; these sites will be referred to hereafter as Calcium Binding Site 1 (CBS1) and 2 (CBS2). CBS1 is hexahedrally coordinated by Asp42, Asp44, Glu96, Asp112, and two water molecules; all of these residues have previously been shown to be important for enzymatic activity by mutational analysis [Dai et
CBS2 is close in proximity and shares some residue coordination by Asp42 and Asp112 with CBS1 and additionally coordinated by Ser98, Glu145, Glu326, and two water molecules, one of which is also shared with CBS1 (Figure 3). With the exception of the Glu145 side chain, which is retracted an additional 1.48 Å away from CBS2, all of the other coordinating residues are maintained with the previous apo-SCAN 2H2N structure, which was also crystallized with calcium. This suggests that these two additional calcium binding sites are important for the apo-enzyme, particularly since the residues of the phosphate binding pocket have a relatively high B-factor, which indicates structural flexibility.

Figure 3. Calcium coordination by apo-I170K SCAN. There are three calcium ions (green) bound in the active site of I170K in the absence of substrate (2Fo – Fc map contoured to 1.2 sigma). Water is shown as red spheres.
Calcium coordination in the phosphate-binding pocket is dynamic to accommodate substrate binding and support nucleophilic attack - While both CBS1 and CBS2 are almost identically coordinated in all apo-I170K crystal structures, composite omit map data suggests that CBS2 is somewhat dynamic and not always 100% occupied despite significant 2F_o−F_c map density for calcium and acceptance by the F_o-F_c map. The dynamic nature of CBS2 became evident after solving the SCAN structures in the presence of substrate and calcium, indicating a shift in the calcium binding sites when substrate is bound. Chain A of both the wild-type and I170K SCAN with GMPCP structures are 100% substrate bound, determined by the low B-factor and significant density for the substrate within the phosphate binding pocket; the wild-type structure also exhibits the previously described shift in Asp99. However, chain B of these structures is likely in a partially-bound substrate state, based on the lack of Asp99 shift for central calcium coordination (for wild-type) and less distinct density for placement of the substrate. By aligning both of the 100% and partially-bound substrate states (molecule A and B, respectively of wild-type with GMPCP) (Figure 4A) with the apo-I170K structure (Figures 4B), it is evident that both calcium molecules move out of the phosphate binding pocket, away from the central calcium, as the substrate binding pocket increases in volume to accommodate the substrate. Wild-type molecule B displays two, partially occupied, overlapping calcium binding sites at CBS1; one calcium overlays identically with the apo-I170K CBS1, and the second calcium aligns with the shifted calcium of the 100% substrate-bound state of molecule A (Figure 4). The calcium that originates in the apo-CBS2 site also shifts to dually coordinate the α- and β-phosphates of the substrate (Figure 4B). This shift in CBS2 in the presence of substrate appears to be mediated by Glu145, retracting the side chain out of the phosphate binding pocket space. The calcium at this site in molecule B of the wild-type SCAN/GMPCP structure remains aligned
with the apo-CBS2 site, and Glu145 remains protruding into the binding pocket. A water molecule is shown to replace each calcium in the original, apo-CBS1 and CBS2 positions.

**Figure 4. Calcium binding in the phosphate binding pocket is dynamic upon substrate interaction.** A. Alignment of WT + GMPCP chains A (lt. green, Ca – green, H$_2$O - red) and B (dk. green, Ca – yellow, H$_2$O - orange), chain B is partially substrate bound. WT + GMPCP (chains A and B) is additionally aligned with apo-I170K (B, blue). The arrows indicate the movement of calcium from the substrate unbound CBS1 and CBS2 (circled in red) to other sites further removed from the binding pocket. C and D demonstrate that these calcium rearrangements are necessary to support nucleophilic attack, as the calcium in the I170K + GMPCP structure crowds the binding pocket (purple, C) whereas the calcium position of WT + GMPCP allows for proper substrate coordination (yellow, C) and shown in D. Green spheres in C are calcium in the apo-state and red spheres in D are water.
The I170K SCAN structure with the substrate bound does not exhibit the same positional shifts in calcium as the wild-type structure (Figure 4C). CBS1 is essentially absent in the presence of the substrate and CBS2 only shifts 0.9 Å laterally compared to the 2 Å vertical shift seen in the wild-type structure. The lack of calcium shift in the I170K structure suggests that calcium rearrangements only occur in the dimer form of the enzyme and that this positional change likely enhances enzymatic activity. However, this data also suggests that the I170K dimer mutant is not deficient in binding calcium, but simply unable to initiate allosteric changes that are required for full dimer mediated hydrolytic activity.

As mentioned above, alignment of chain A of the wild-type SCAN/GMPCP structure with the apo-I170K structure reveals rearrangement of calcium in both CBS1 and CBS2 during substrate binding. The calcium at CBS1 moves 2.05 Å out of the phosphate binding pocket, away from Asp112 (called Ca1), and the original CBS1 site is replaced with water. This new calcium binding site also reduces to a tetrahedral coordination state, with dual interactions provided by Asp42, Asp44, and the β-phosphate of GMPCP, and a water molecule that has replaced calcium in the original CBS1 site (Figure 4D). The position of Asp42 mediates the shift of Ca1 and a similar movement of the Asp42 side chain is observed upon substrate binding. This new organization of calcium coordination in the presence of the substrate places Ca1 1.3 Å closer to the β-phosphate, enabling a covalent bond to be formed. In addition to aligning the substrate with nucleophilic residues, the location of this calcium would potentially stabilize the leaving group and enhance nucleophilic attack by polarizing the phosphate P = O bond [236]. The water that replaces calcium in CBS1 is coordinated by Asp112; the position of this water is conserved in the 1S1D structure as well, suggesting its importance in the active site. Both Asp42 and Asp112 have been proposed to be involved in nucleophilic attack [232] Asp44 is positioned
perpendicular to the attacking water and the β-phosphate and has previously shown a significant
decrease in enzymatic activity when mutated [232], suggesting its role in stabilizing a transition
state during nucleophilic attack as well as stabilizing other residues in the active site.

Similar to CBS1, the calcium in CBS2 also shifted 2.08 Å further out of the phosphate
binding pocket to accommodate (called Ca2) substrate binding (Figure 4D). This shift allows
Ca2 to contact both the α- and β-phosphates of the substrate. Ca2 is also highly coordinated in
this position, an octahedral coordination by Glu145, Glu214, Glu326, and three water molecules
in addition to the substrate α- and β-phosphates. The shift in Ca2 appears to be primarily
mediated by Glu145, in which bond lengths between Glu145 and calcium are maintained from
the apo-enzymatic state despite a 1.9 Å shift of the Glu145 side chain. Glu145 and Asp112
dually contact one of the calcium-coordinated waters that would enable the side chain carboxyl
groups to act as a general base for nucleophilic attack. In addition, the close proximity and
location of Glu145 to the substrate suggest that Glu145 could contact the α-phosphate and act as
the nucleophile directly through a dissociative mechanism of attack. This is also supported by
the highly coordinated state of the substrate by multiple acidic residues of the active site (Asp42,
Asp44, Glu214, Glu145, Glu326) and calcium that could produce a stabilized covalent
intermediate, as seen in the FLAP endonucleases [237]. Previous mutational analysis of Glu145
also suggests that it is a critical residue for nucleophilic attack [232].
**Figure 5. Arg113 coordinates dimer-enhanced enzymatic activity.**

A. Arg113 coordinates water in the CBS1 site in the WT-GMPCP structure (green; Ca yellow). However, apo-I170K (blue), I170K + GMPCP (purple; water circled), and 1S1D (yellow structure) does not demonstrate a shift of Arg113 into the binding pocket to coordinate water with the substrate. B. Diagram of Arg113 residue associations that generate communication with the dimer interface when the substrate is bound. The side chain of Arg113 moves to contact Glu96 (originally 4.52 Å apart, red), causing a shift of the entire bridged section between R93 to A89 and allowing dimer contacts at the carbonyl of A89. The side chain of Glu90 also moves 1.87 Å into the binding pocket, breaking a 2.81 Å salt bridge with K143 that exists in the apo-state. C. Arg113 causes a main chain shift ripple through A89 that initiates inter-molecule dimer contacts. Wild-type/GMPCP Green (lime=Molecule A; Dark green = Molecule B); Apo-I170K = purple (molecule B not shown). D. The salt bridge lock between E90 and K143 is broken during the R113 transition. E. Dimer interface residue contacts of the WT-GMPCP structure. Residues 135-140 and 48-55 were hidden to expose residues of the dimer interface.
Key residues responsible for dimer enhanced enzymatic activity - Our structural data indicates that calcium rearrangement upon substrate binding is important for maximal enzymatic activity. In correlation with calcium rearrangement, the positioning of water within the binding pocket is equally important in order to generate a metal hydroxide with calcium for nucleophilic attack. Arg113 coordinates the water molecule associated with Ca1 and is stabilized by forming a salt bridge with Glu96 in the wild-type structure with the substrate (Figure 5A). Glu96 additionally forms a contact with Arg93; Arg93 is linked to Glu90, which is a key residue involved in dimerization (Figure 5B). Evidence from our wild-type substrate bound structure suggests that Arg113 contact with Glu96 causes the allosteric conformational change that induces dimerization. The main chain is rippled as a result of Arg113 contacting Glu96 to push the Glu90 side chain and the Ala89 carbonyl oxygen into the dimer interface, the carbonyl oxygen of Ala89 making contact with Ala89 on the partnering molecule (Figure 5C). The position of the Glu90 side chain is additionally controlled by Lys143, in which they are electrostatically bound in the apo-I170K structure (2.8 Å distance) and are pushed 10.5 Å apart in the presence of substrate in the dimer form (Figure 5D). Ile130, the position of the I170K mutation, is also within close proximity to this region and is likely capable of making contact with the partnering I130 once the first association occurs at Ala89. The third dimer contact, Lys184 and Asp188 are far away from the Glu90 site and likely form a salt bridge as a result of the previous associations at Glu90 and Ile130 (Figure 5E). Glu96 is also near Asp99, the primary residue that alters contact with the central calcium upon substrate binding. It is likely that the Arg113 conformational change enhances enzymatic activity by influencing the central calcium coordination shift, which also affects many of the critical residues for nucleophilic attack.
This conformational lock provided by Arg113 is supported by the partially substrate-bound wild-type CAN structure (molecule B), in which Arg113 is retracted out of the binding pocket. In the partially-substrate bound structure, CBS1 is still 50% occupied at the original, apo-enzyme position as well as the shifted, substrate bound position. The apo-I170K and substrate bound structures, as well as 1S18 and 1S1D, do not show this shift of Arg113 into the binding pocket, also confirming that this is a calcium-mediated and dimer-specific conformational change. The positive charge of arginine would prevent the calcium shift to the substrate bound position, suggesting that Arg113 acts as the key regulatory residue with respect to dimer-enhanced enzymatic activity. In addition to acting as a communicative residue, the retracted Arg113 confirmation of the apo and mutant structures there are two waters that are identically coordinated to the side chain amine groups, suggesting that this is an important pre-substrate bound coordination state. Arg113 also coordinates the only water in contact with CaⅠ in the substrate-bound state; it is therefore possible that Arg113 is involved in generating the proton for the leaving group [238] as well as serving as the key residue to communicate dimer-related conformational changes and enhanced activity.

To test this hypothesis, we generated an alanine mutant, R113A, to determine the role of this residue for dimer-enhanced activity (Figure 6). Initial enzymatic assays indicated that the R113A mutation causes significant decreases in SCAN enzymatic activity; the R113A mutant maintained only 17.6%, 4.1%, and 10.8% of wild-type activity for substrates GDP, ADP, and ATP, respectively (Figure 6A). The nucleotidase activity experiments were performed at low protein concentrations (0.85 µM) that are consistent with our previously published data [231] [230]. However, ADPase activity has previously been shown to correlate with protein concentration [230]. Therefore, we repeated the nucleotidase experiment over increasing protein
concentrations and 5mM CaCl\textsubscript{2} (Figure 6B); however, the R113A mutant did not show an increase in activity despite increasing the protein concentration. To ensure that this mutation was not additionally preventing dimerization, and therefore decreasing nucleotidase activity, we characterized dimer formation of R113A using tryptophan fluorescence assays and analytical ultracentrifugation (AUC) sedimentation velocity in the presence of calcium. The tryptophan fluorescence assays showed enhanced dimerization compared to wild-type, with an EC\textsubscript{50} around 5 \(\mu\text{M}\) compared to 86 uM for wild-type SCAN (Figure 6C). This decreased requirement for calcium-induced dimerization was confirmed by AUC, in which 95\% of species in solution sedimented as a dimer at only 25 \(\mu\text{M}\) calcium whereas wild-type SCAN requires 200 \(\mu\text{M}\) calcium to see an almost complete monomer to dimer shift (Figure 6D). In addition, the R113A mutant without calcium demonstrated a small population of species (22\%) that exist in a partially associated state (shoulder peak at 4S and approximate molecular weight of 53 kD). This supports our hypothesis that Arg113 acts as a regulator to control residues that participate in dimerization.
Figure 6. R113A SCAN mutant dimerizes but does not exhibit dimer-enhanced nucleotidase activity. A. Characterization of R113A mutant enzymatic activity, represented as a fraction of wild-type activity. B. The activity of the R113A mutant is not enhanced by increasing the concentration of protein, despite the enhanced ability to dimerize that is shown by enhanced tryptophan fluorescence (C) and analytical ultracentrifugation compared to wild type SCAN (D).
**CAN dimerization stabilizes the phosphate binding pocket** - The dimer interface is within close proximity of the phosphate binding pocket (mediated by residues I90, I130, K184, and D188), however, does not share any residues that are critical for enzymatic activity. It was therefore surprising that a single mutation to a residue on the dimer interface would drastically decrease in enzymatic activity, particularly for the adenosine bases that are important for regulation of the blood clotting pathway [230]. After initial overview of the entire protein with the substrate bound, it was evident that the phosphate binding pocket was considerably smaller in volume in the monomeric mutant I170K structure compared with wild-type SCAN. We therefore measured the active site volume and surface area of both the wild-type and I170K structures, with and without the substrate using the HotSpot Wizard online server (Table 3). The apo-wild-type SCAN structure with calcium, 2H2N, has a surface area and volume of 725.5 Å² and 1251.3 Å³, respectively. Upon substrate binding, the surface area and volume of the binding pocket increase by 26% and 13% (Table 3). However, the phosphate binding pocket of the apo-I170K SCAN has a surface area and volume of 539.7 Å² and 915.0 Å³ (74% and 73% of wild-type SCAN). While the surface area and volume increase by 17% and 16%, somewhat comparable to the change seen in wild-type SCAN, the overall surface area and volume is still 284 Å² and 350 Å³ smaller than the wild-type binding pocket (Table 3). It appears that the primary reason for narrowing of the I170K SCAN binding pocket is a lack of restraint on the residues lining the channel, many of the side chains are extending far into the pocket. It is likely that the dimer interface acts a rigid lock that restraints the residues of the binding pocket from aberrantly closing the channel.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Surface Area (Å²)</th>
<th>% Surface Area of Apo-WT</th>
<th>Volume (Å³)</th>
<th>% Volume of Apo-WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apo-WT CAN (2H2N)</td>
<td>725.45</td>
<td>-</td>
<td>1251.28</td>
<td>-</td>
</tr>
<tr>
<td>WT CAN + GMPCP (+ Ca²⁺)</td>
<td>916.55</td>
<td>126%</td>
<td>1416.25</td>
<td>113%</td>
</tr>
<tr>
<td>WT CAN + GMPCP (1S1D, +Sr²⁺)</td>
<td>621.43</td>
<td>86%</td>
<td>1087.14</td>
<td>87%</td>
</tr>
<tr>
<td>Apo-I170K</td>
<td>539.71</td>
<td>74%</td>
<td>915.01</td>
<td>73%</td>
</tr>
<tr>
<td>I170K + GMPCP</td>
<td>631.93</td>
<td>87%</td>
<td>1066.50</td>
<td>85%</td>
</tr>
</tbody>
</table>

Table 3. Surface area and volume of SCAN proteins in the presence of GMPCP, Ca, and/or Sr. Values generated by HotSpot Wizard.

Contributions of calcium-dependent CAN dimerization for enzymatic activity. In order to determine the how SCAN function is increased in a dimeric state, we aligned the wild-type and I170K SCAN structures with substrate (Figure 7A). The overall conformation of the nitrogen base and ribose ring of the substrate is relatively similar, with only a 1.87 Å lateral shift of the ribose ring out of the binding pocket while maintaining the bond orientations. Residues that coordinate the guanidine ring are relatively conserved between the two structures, despite a high degree of flexibility in this region. The phosphate groups, however, show a large deviation in both bond orientation and residue coordination. The α-phosphate is rotated 90° counterclockwise and shifted 2.15 Å out of the binding pocket compared to the wild-type...
orientation, preventing any residues of the phosphate binding pocket to coordinate this phosphate position (Figure 7B). This includes Glu145, the residue predicted to primarily mediate nucleophilic attack. The α-phosphate of the wild-type structure is additionally coordinated by Glu214 and Arg230, both of which are well out of reach of the I170K α-phosphate (5.8 and 4.9 Å away, respectively).

The β-phosphate group of the I170K structure is also rotated 90° in the active site, placing the β-phosphate 3 Å deeper into the binding pocket (Figure 7C). Because Ca2 does not move out of CBS2, the β-phosphate alone contacts Ca2 instead of both phosphate groups; Ca1 is absent and instead CBS1 is occupied by the β-phosphate group. As previously discussed, the position of Ca2 is primarily mediated by Glu145; therefore, Glu145 also remains in the original apo-state position and instead coordinates both calcium (Ca2) and the β-phosphate group (Figure 7D). While Glu145 could still participate in catalysis by supplying hydroxide or stabilizing the β-phosphate, it is no longer optimally aligned or within a reasonable distance to be the primary source of nucleophilic attack. The β-phosphate group of the I170K structure is additionally hyper-coordinated by Asp42, Asp44, Glu96, Ser98, Asp112, and Lys324; whereas the β-phosphate group of the wild-type structure is coordinated by Asp44, Arg230, Lys324, and Glu326.

Both Glu145 and Asp112 were implicated to participate in nucleophilic attack in the wild-type SCAN structure, however, the skewed orientation of the substrate in the I170K binding pocket does not suggest Glu145 to be a primary nucleophile. Therefore, in monomeric form, SCAN is likely to utilize Asp112 and a single calcium ion (Ca2) for enzymatic hydrolysis. The proximity and angle of Asp112 to the cleavage site suggests that the side chain carboxyl oxygen would act directly as the nucleophile. Arg113 dually coordinates a water molecule with Asp112
that could be used for nucleophilic attack, although this water is not associated with calcium as in the wild-type structure. The position of the substrate has shifted in the binding pocket towards Glu295, which coordinates a water molecule in both wild-type and I170K structures. This water is located in a reasonable position to act as a nucleophile as well, and could be the alternative mechanism for catalysis in the monomeric form. Regardless, there is only one calcium ion to stabilize and enhance nucleophilic attack, potentially reducing enzymatic activity in monomeric form. In this mechanism, Ca2 would stabilize the leaving group and enhance nucleophilic attack as Ca1 does in the wild-type structure. The coordination of a single calcium ion in the binding pocket could be caused by the retraction of Arg113, and therefore the associated water from the binding pocket. Arg113 is additionally “locked” through a contact with Lys143, which was previously discussed to participate in dimer-related activity enhancement via Glu90. Arg113 also does not contact Glu96 and transition of Glu90 into the dimer interface is not seen, strengthening our hypothesis for Arg113 controlling this allosteric change for dimer activity.
Figure 7. Differences in substrate coordination by the dimeric and monomeric forms of SCAN. A. Substrate coordination by wild-type SCAN (green) compared to I170K (violet), which is rotated 90° at the α-phosphate position. A detailed view is shown in B, demonstrating that important contacts with E214 and E145 are not maintained for catalysis. Rather, the β-phosphate is hyper-coordinated in the active site (C). Therefore, I170K can only bind a single Ca in the active site (purple sphere, D) compared to 2 Ca ions in the wild type active site (green spheres, D).
**Discussion**

The human calcium-activated nucleotidase (SCAN) can be secreted as a soluble and highly stable enzyme that requires calcium for phosphatidic cleavage of nucleotide di- and tri-phosphates. While its physiological function is still relatively unknown, SCAN has previously been characterized biochemically and structurally in multiple studies by our lab and others in order to potentially use SCAN as a therapeutic drug to prevent heart attack and stroke, since ADP is the platelet agonist that initiates the blood clotting pathway. While we have previously determined the importance of calcium for activity and dimerization [227, 231] and Dai et al have demonstrated the critical residues for substrate specificity [232], the molecular contributions of calcium for enzymatic activity and dimerization were still unknown. In this study, we have provided the first structural insights to the role of calcium for SCAN enzymatic activity as well as enhanced enzymatic activity upon dimerization.

Our crystal structures of the apo-I170K SCAN indicate that two calcium ions bind in the phosphate binding pocket regardless of the presence of a nucleotide base. This suggests that calcium is pre-bound to neutralize the acidic, negatively charged residues of the phosphate binding pocket to facilitate binding of the negatively-charged substrate. In addition, our structures of both wild-type and I170K SCAN in the presence of substrate indicate that calcium is important for substrate coordination, and allowing calcium to bind during the apo-state of the enzyme would prevent calcium coordination from being a rate limiting step in catalysis. While the position of these calcium-binding sites seems to be conserved and important for stability of the active site in the apo-state, the transition of these calcium-binding sites in the presence of substrate seems to be equally as important for maintaining substrate coordination and efficient enzymatic activity.
A. “Two-Metal ion” Mechanism

B. “One-Metal ion” Mechanism

Figure 8. Model of the catalytic mechanism by CAN in monomeric and dimeric forms. A. “Two-Metal” mechanism utilized by dimeric CAN, which uses R113 to stabilize the leaving group and activate the coordinated water to aide in nucleophilic attack by D112 and E145. B. In the monomer state, however, only a single metal ion can be used. Therefore, R113 and D112 are likely to be the critical residues for nucleophilic attack.
The increase in SCAN enzymatic activity as a dimer is not well understood, particularly since dimerization itself was not discovered until 2004 despite the multiple, previous biochemical and structural studies of SCAN. Our new wild-type SCAN and I170K monomer mutant structures suggests there is a large deviation in the catalytic mechanism between the monomer and dimer forms (Figure 8). Our wild-type SCAN structure, dimerized in the presence of calcium, suggests a “two-metal” mechanism for catalysis [239] (Figure 8A). This mechanism of catalysis has been previously characterized in bacterial restriction enzymes [240, 241], RNA processing enzymes [242], and Flap endonucleases, which similar to CAN, uniquely contain 8 active site carboxylates [237] compared to only 4 in most metallonucleases. The typical “two-metal” mechanism utilizes two metals that are dually bound to the same non-bridging oxygen of the leaving phosphate; one metal ion (metal “A”) additionally coordinates the nucleophile and the other metal ion (metal “B”) coordinating a water molecule that is a proton donor to the leaving group [239]. In the wild-type SCAN structure, the two calcium ions in the phosphate binding pocket are within close proximity of each other (5.4 Å, 4 Å or less is optimal) and are both within a reasonable distance from the same non-bridging oxygen of the β-phosphate. Ca2 would play the metal “A” role, coordinating the carboxyl oxygens of Glu145 and Glu214, while Ca1 would act as metal “B,” primarily coordinating the β-phosphate and a water molecule. Also in correlation with the two-metal mechanism of catalysis, both calcium ions can be potentially coordinated by the same residue, Asp112. In our structure, Ca1 and Ca2 are 4 and 4.4 Å, respectively from Asp112; however, because GMPCP is a non-hydrolyzable substrate, it is possible that the substrate is not in the exact native coordination of the natural substrate. Arg113 is also likely involved in the nucleophilic attack mechanism, coordinating a water molecule with Ca1 (Figure 8). The R113A mutant demonstrates nucleotidase activity that is lower than I170K;
if R113A was merely involved in dimer enhanced activity, the enzymatic activity would match that of I170K [230]. Outer sphere, positively charged residues have been previously shown to assist with water activation [240]. Therefore, Arg113 likely helps activate the water associated with Ca1 that would assist with the leaving group. This is the first publication of a mutation to Arg113 and demonstrating its importance for SCAN activity (Table 4).

As previously mentioned, the active site of I170K in the presence of substrate only contains one calcium that coordinates the β-phosphate, removing the CBS1/metal “B” position. While the two-metal ion mechanism proposed for the dimerized CAN is likely more efficient for catalysis, a one-metal ion mechanism has also been shown to support catalytic activity in other nucleases [243]. In this one-metal ion mechanism, the position of metal “A” is maintained in order to support nucleophilic attack [243]. In our I170K structure, there is slight deviation in the position of CBS2 from wild-type, although the calcium is still in close proximity to nucleophilic residues. The calcium is utilized primarily for stability of the leaving group and transition state, and nucleophilic attack would occur via water (dually coordinated by Asp112 and Arg113), Asp112, and Glu145 (Figure 8B).

Calcium-induced conformational changes in SCAN have been previously identified as dimerization, although all previous wild-type crystal structures have not provided clues as to how calcium mediates this event. Our high resolution wild-type and apo-I170K CAN structures show clear density for calcium in the phosphate-binding pocket, however, there is no indication of additional calcium binding sites on the dimer interface. Therefore, it is likely that the calcium ions within the binding pocket induce the conformational changes that facilitate dimerization. This is additionally supported by our R113A mutant that lacks enhanced enzymatic activity despite the ability to dimerize; Arg113 regulation of dimerization is controlled by calcium in the
binding pocket. In addition, R113A can partially self-associate without the presence of calcium which has never been seen previously, even in the dimer activity-enhanced mutant E130Y [231]. Another calcium-activated enzyme that exhibits higher activity in dimeric form, PAD4, binds calcium near the enzymatic active site rather than within the dimer interface [244]. While PAD4 does not require calcium specifically to form a dimer, calcium does have an effect on the conformational stability of the dimer form [244].
<table>
<thead>
<tr>
<th>Mutation</th>
<th>Ca&lt;sup&gt;2+&lt;/sup&gt; GDPase Activity (Percentage of Wild-type)</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>D42N</td>
<td>1.10</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>D44N</td>
<td>9.00</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>S69C</td>
<td>59.00</td>
<td>Yang, 2006</td>
</tr>
<tr>
<td>E90Y</td>
<td>202.46</td>
<td>Yang, 2006</td>
</tr>
<tr>
<td>E96Q</td>
<td>4.90</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>S98A</td>
<td>0.81</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>D99N</td>
<td>4.30</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>D112N</td>
<td>0.04</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td><strong>R113A</strong></td>
<td><strong>17.60</strong></td>
<td><strong>This Publication</strong></td>
</tr>
<tr>
<td>I130E</td>
<td>58.00</td>
<td>Yang, 2008</td>
</tr>
<tr>
<td>I130K</td>
<td>83.00</td>
<td>Yang, 2008</td>
</tr>
<tr>
<td>I130C</td>
<td>50.82</td>
<td>Yang, 2006</td>
</tr>
<tr>
<td>S132C</td>
<td>56.56</td>
<td>Yang, 2006</td>
</tr>
<tr>
<td>E145Q</td>
<td>0.67</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>K165A</td>
<td>1.60</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>W167A</td>
<td>0.86</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>T166A</td>
<td>6.00</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>E176M</td>
<td>460.00</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>S186C</td>
<td>90.16</td>
<td>Yang, 2006</td>
</tr>
<tr>
<td>D188T</td>
<td>84.00</td>
<td>Yang, 2008</td>
</tr>
<tr>
<td>E214Q</td>
<td>33.00</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>C217S</td>
<td>74.59</td>
<td>Yang, 2006</td>
</tr>
<tr>
<td>R230A</td>
<td>2.20</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>R232A</td>
<td>0.64</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>S238C</td>
<td>73.77</td>
<td>Yang, 2006</td>
</tr>
<tr>
<td>Y241A</td>
<td>2.00</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>A247C</td>
<td>58.19</td>
<td>Yang, 2006</td>
</tr>
<tr>
<td>S275A</td>
<td>69.00</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>E295Q</td>
<td>14.00</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>K324A</td>
<td>7.40</td>
<td>Dai et al, 2004</td>
</tr>
<tr>
<td>E326Q</td>
<td>6.80</td>
<td>Dai et al, 2004</td>
</tr>
</tbody>
</table>

*Table 4. Total known residue* mutations and respective activities of SCAN (*Uses numbering conventions of this publication*)
Acknowledgements

This work is based upon research conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team beamlines, which are supported by award RR-15301 from the National Center for Research Resources at the National Institutes of Health and the GM/CA CAT has been funded in whole or in part with Federal funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Sciences (Y1-GM-1104). Use of the Advanced Photon Source, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the U.S. DOE under Contract No. DE-AC02-06CH11357.

We would also like to thank Dr. Tom Thompson, Dr. Jennifer Cash, and Dr. Rhett Kovall for their help with crystallization and data processing, as well as insightful discussions.
Bibliography


