Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Summer Bushman April 12, 2022

Role of Streptococcal Pyrogenic Exotoxins A and B (SpeA and SpeB) Mutation in *Streptococcus pyogenes* Virulence

By

Summer Bushman

Christopher LaRock
Adviser

Biology

Christopher LaRock

Adviser

Chris Beck

Committee Member

Sarah Fankhauser

Committee Member

2022

Role of Streptococcal Pyrogenic Exotoxins A and B (SpeA and SpeB) Mutation in *Streptococcus pyogenes* Virulence

By

Summer Bushman

Christopher LaRock

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Biology

2022

Abstract

Role of Streptococcal Pyrogenic Exotoxins A and B (SpeA and SpeB) Mutation in *Streptococcus pyogenes* Virulence

By Summer Bushman

Streptococcus pyogenes (also referred to as group A *Streptococcus*, or GAS) causes about 18.1 million cases of severe infection annually, resulting in 500,000 deaths worldwide. It is important to study the interactions of GAS with its human host because there is no vaccine, and the only treatment is antibiotics, to which resistance is rapidly increasing. This project looked at two specific streptococcal pyrogenic exotoxin (Spe) proteins, SpeA and SpeB. SpeA is a superantigen associated with cases of streptococcal toxic shock-like syndrome (STSS). SpeB is a protease that degrades immunoglobulins and cytokines. Using sequences that encode for SpeA and SpeB to identify naturally occurring synonymous and nonsynonymous polymorphisms, we found that SpeA is undergoing diversifying selection and SpeB is undergoing purifying selection. Nonsynonymous mutations have the potential to alter protein function, so these mutations within SpeB were further examined. Site-directed mutagenesis was used to change the sites of three natural nonsynonymous polymorphisms (G17, E130, and D154) within *speB* to alanine via a pET expression plasmid. After sequence confirmation, the DNA was transformed into *E. coli* where the mutated proteins were expressed. The proteins were subsequently extracted and purified. The activity of each mutant protein was examined *in vitro* using fluorescent activity assays to quantify the degradation of known SpeB substrate and fluorophore, substrate 73. Non-mutant reference-type SpeB from strain M1T1 5448 served as positive controls; Strains with empty pET vectors served as negative controls. SpeB with mutation G17A was found to have similar activity to wildtype SpeB, indicating no alternation in conformation or functional ability. SpeB with mutations E130A or D154A was found to have a slight reduction in activity, indicating potential interference with binding sites. This data can be used to determine the functional consequences of heterogeneity within *speB*. In future studies, this data can be applied more broadly to help develop effective public health policies and identify important and conserved proteins to target for the development of a vaccine, which is critically lacking for GAS.

Role of Streptococcal Pyrogenic Exotoxins A and B (SpeA and SpeB) Mutation in *Streptococcus pyogenes* Virulence

By

Summer Bushman

Christopher LaRock

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Biology

2022

Acknowledgements

I would like to thank Dr. LaRock for mentoring me throughout the past two years and helping me develop my skills as a scientist, a researcher, an academic, and as a person overall. I would also like to thank Dr. Beck and Dr. Fankhauser for the taking the time to serve on my thesis committee, being engaging professors who have influenced my decision to pursue academia, and supporting me throughout my undergraduate career. I would like to thank all members of the LaRock Lab for their patience as they guided me, especially Doris LaRock, Anders Johnson, and Shyra Wilde. Finally, I would like to thank my friends and family for their endless support, especially my parents Sharon and John Bushman. Thank you to all the aforementioned individuals. I am forever inspired to be a better person, and I am grateful for everything you all have done for me.

Table of Contents

Chapter 1: Introduction

Many bacteria, fungi, viruses, and parasites are capable of causing infectious disease (Inhorn and Brown, 1990). Most common infectious diseases have low mortality rates, such as the common cold or chlamydia (Cleveland Clinic, 2018). However, others have high mortality rates, and altogether account for a quarter of all deaths (Drexler, 2010). Risk can vary across groups; in children under five, infectious diseases account for about two-thirds of all deaths (Drexler, 2010). Vaccines to prevent human infection have been a major innovation in medicine for the prevention of infectious diseases including measles, smallpox, tuberculosis, and COVID-19. For infectious diseases caused by bacteria, the primary mode of treatment are antibiotics (Cleveland Clinic, 2018).

One important pathogen for which we lack a vaccine, and which relies on antibiotic treatment is *Streptococcus pyogenes*, also called Group A Streptococcus or GAS (Shulman et al., 2012). GAS often causes pharyngitis, colloquially known as strep throat, which accounts for 37% of sore throats in children under 18 years of age (Shaikh et al., 2010). However, infections by GAS can also lead to fatal complications, such as scarlet fever and streptococcal toxic shock-like syndrome (Kazmi et al., 2001). It is estimated that there are about 18.1 million annual severe cases of GAS infections, resulting in 500,000 annual deaths worldwide (Carapetis et al., 2005).

A major cause of complications is the massive amount of inflammation that is induced by streptococcal pyrogenic exotoxins (Spe) that are released by GAS. Most Spe proteins are superantigens, including the most characterized one, SpeA (Kotb, 1995). Superantigens act by bridging the MHC receptor of an antigen presenting cell with the T cell receptor of T cells in an antigen-independent manner. This leads to high levels of cell activation and an increased release of inflammatory cytokines relative to a standard antigen (Kotb, 1995). The increased levels of inflammation can lead to more severe GAS infections with higher mortality rates and is directly

responsible for streptococcal toxic shock-like syndrome (STSS). In fact, there is a strong association between STSS and strains encoding for SpeA over other superantigens (Musser et al., 1991), with one study finding 85% of its STSS patients testing positive for the presence of SpeA (Hauser et al., 1991).

There are four primary gene alleles encoding for SpeA, and three of the alleles, *speA1, speA2,* and *speA3* differ by a single amino acid substitution (Nelson et al., 1991). Interestingly, both *speA2,* and *speA3* are more strongly correlated with STSS. Both have mutations in the same highly conserved gene segment, which could indicate that they result from selective pressures (Marrack and Kappler, 1990). Furthermore, it was found that the single amino acid change within allele *speA3* forms a more active toxin of SpeA with significantly increased mitogenic activity (Kline and Collins, 1996). Since these studies, many more new strains from clinical isolates have been sequenced and identified.

In contrast to the superantigenic activity of SpeA and all other Spes, SpeB is a prolific cysteine protease that cleaves or degrades immunoglobulins, cytokines, GAS surface proteins, and other immune effectors (Chuan and Wu, 2008). Because it is highly conserved and cleaves or degrades both host and bacterial proteins, it is considered a key virulence factor of GAS pathogenesis (Chuan and Wu, 2008). Variation in expression of SpeB is associated with clinical disease and increased inflammation. For example, the level of protease activity from SpeB was significantly associated with soft tissue necrosis in STSS (Talkington, et al., 1993). The *speB* gene is encoded by all strains of GAS and is considered highly conserved with allelic variation resulting from accumulating point mutations (Kapur et al., 1993). However, the exact role of these molecular variations is not well established, and it is not clear as to what extent or how SpeB expression levels impacts disease severity (Chuan and Wu, 2008).

While some common natural polymorphisms are known within *speA* and *speB,* even more can be observed in the deep sequencing of clinical isolates, and little is known about their functional consequences. Therefore, in chapter 2, I examined the natural polymorphisms present in the sequences encoding SpeA and SpeB. I then classified them by whether they were synonymous or not, to better understand whether they could have an effect on function, and whether there was evolutionary selection on these proteins. Since *speB* is known to be highly conserved, I used those sequences as a baseline to compare variation among *speA* sequences. Based on this analysis, my preliminary work suggests that SpeA is undergoing diversifying selection, while SpeB is undergoing purifying selection. Since the functional significance resulting from the sequence variations of *speB* is not well-characterized, and there were fewer mutations, I chose to continue analysis of those polymorphisms in *in vitro* biochemical experiments. Therefore, in chapter 3, I investigated the role of these natural mutations on the functionality of SpeB and subsequently the virulence GAS. Long-term, I expect these studies to provide insights into how GAS has evolved as a human-specific pathogen and potential targets for its treatment.

Chapter 2: Examining the Heterogeneity of SpeA and SpeB

The GAS superantigen SpeA has four well-characterized alleles; two of these, *speA2* and *speA3*, are associated with greater STSS infection severity. In recent years, a large number of GAS genomes have been sequenced, and upon analysis of this data it is clear that SpeA has acquired many more natural polymorphisms within different strains of GAS beyond what is captured with this allelic assignment. In contrast, the protease SpeB that is encoded by all GAS is highly conserved, with more limited heterogeneity arising from accumulating point mutations. Therefore, heterogeneity within *speB* will serve as a baseline to examine the prevalence of allelic variation within *speA*. I hypothesize that if SpeA has a significantly greater number of mutations than SpeB, then its greater genetic variation is due to diversifying selection. I expect that SpeB has more limited genetic variation, due to purifying selection. The ultimate goal is to test whether polymorphisms arising as part of the natural genetic heterogeneity of GAS have functional consequences that impact virulence. Further biochemical characterization of these mutations is the focus of Chapter 3.

Methods

Collecting Sequences

Using Basic Local Alignment Search Tool (BLAST) and Ensembl Genome Browser, the coding sequence of *speA* was collected from the full genome sequences of 169 unique clinical isolates representing the global diversity of the species, and without reference to time of collection. MGAS5005, a clone of the hypervirulent M1T1 serotype responsible for the resurgence in invasive disease since the 1980s (Walker et al., 2014) was used as the reference strain to which all other strains were compared. The same process was repeated with SpeB using strain M86905.1 as the reference strain to find 100 additional coding sequences.

Sequence Alignment

Using version 1.2.4. of Clustal Omega, a multiple sequence alignment was performed among the 170 total *speA* sequences and 101 total *speB* sequences. Settings were all default with the input being DNA sequences and the output was formatted in Pearson/FASTA (Madeira et al., 2019).

Classification of Polymorphisms

After alignment, all the sequences were analyzed via the Synonymous Nonsynonymous Analysis Program (SNAP) (Korber, 2000). This analysis graphs the positional ratio of synonymous mutations to the ratio of nonsynonymous mutations (dS/dN) across the gene sequence. It also is able to define whether there is an insertional or deletional mutation (indel) present at a particular genetic location. Synonymous mutations were compiled and used to map evolutionary distance of alleles, creating a dendrogram (Korber, 2000). To examine whether nonsynonymous mutations

where clonal, their positional frequency was manually examined using this dendrogram, where occurrence in neighboring clades was assumed to be inheritance from a common source.

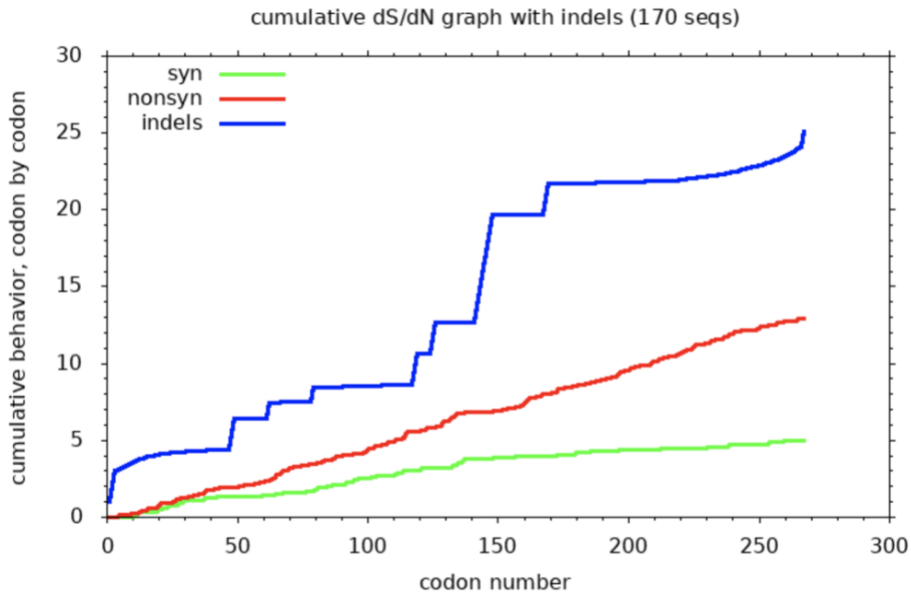
3-D Modeling

The structural model of SpeB (4D8I) was obtained from on the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB). These coordinates were examined and visualized in PyMOL (Schrödinger). All natural polymorphisms were mapped onto this 3-D ribbon structure for SpeB and manually positionally mutated, using stick figure diagrams to examine changes in amino acid side chains.

Results

Synonymous and Nonsynonymous Polymorphisms

A.



B.

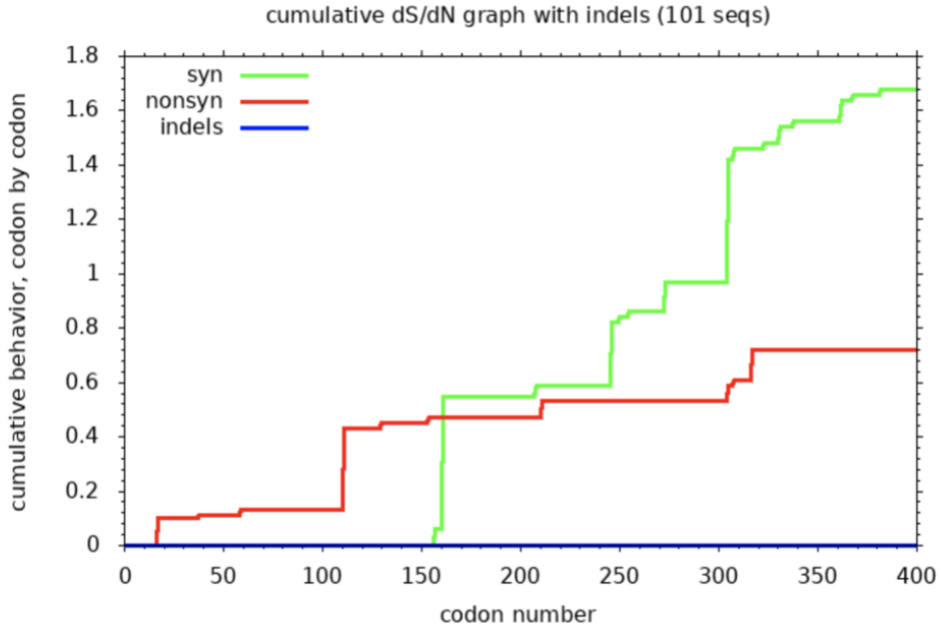
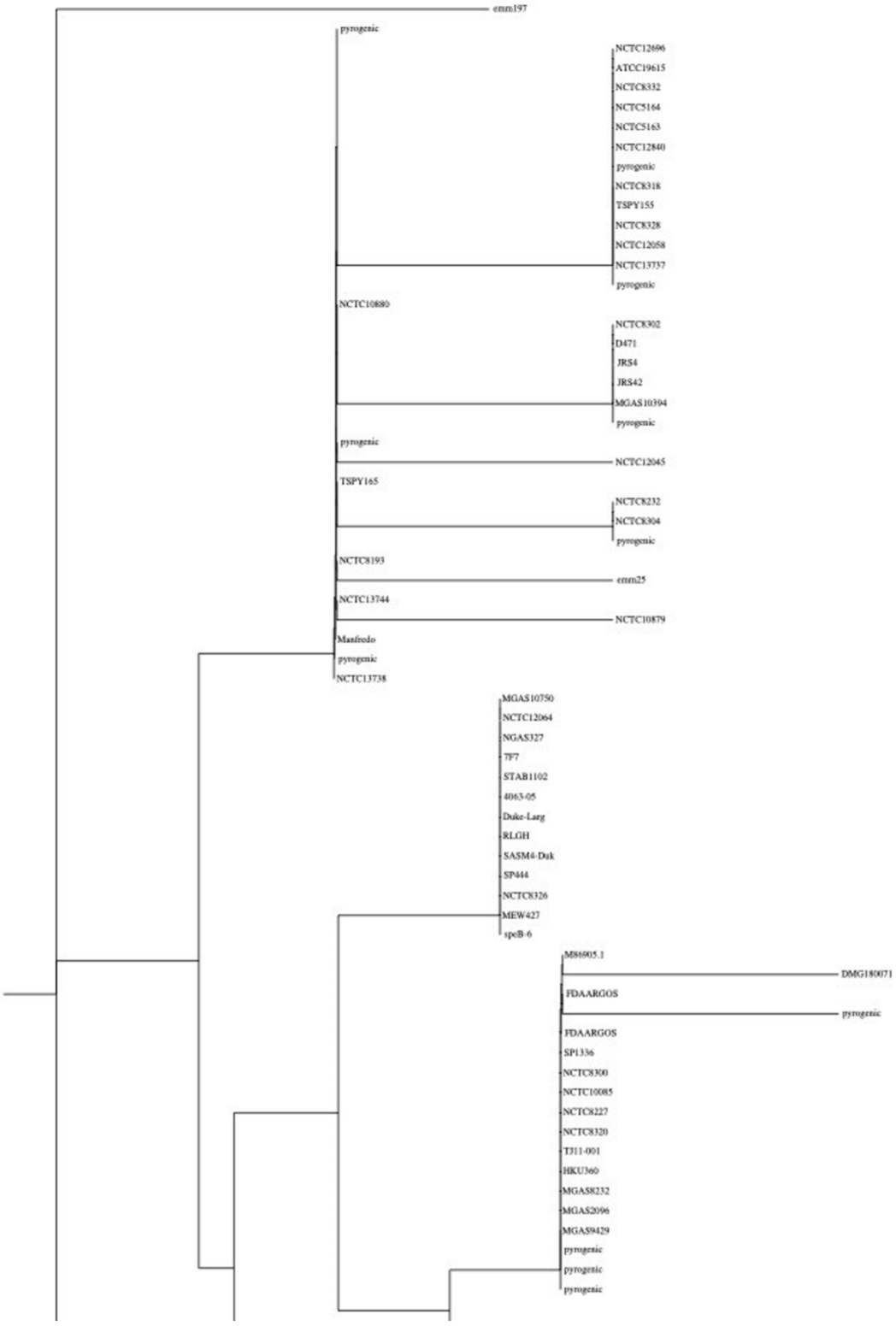


Figure 1.
These graphs generated by SNAP show the location of each mutation. Nonsynonymous mutations are indicated by the red line, synonymous mutations are indicated by the green line, and insertions or deletions are indicated by the blue line. The top graph (**Figure 1A**) classifies the polymorphisms from SpeA, and bottom graph (**Figure 1B**) classifies the polymorphisms from SpeB.

In a dS/dN graph, a positive slope indicates that a polymorphism has been detected at that particular codon site, whereas a flat slope indicates no polymorphisms found at those locations. Thus, an increase in slope indicates a region of change, and a flat slope indicates a region of conservation. The top graph (**Figure 1A**) classifies the polymorphisms from SpeA. Since the slope of both the synonymous and nonsynonymous mutations in **Figure 1A** is relatively constant, we can determine that the polymorphisms are relatively evenly distributed throughout the entire genetic sequence coding for SpeA. The polymorphisms of *speA* are likely to alter the amino acid sequence because SpeA has a larger slope in the red line relative to the green line. The bottom graph (**Figure 1B**) classifies the polymorphisms from SpeB. SpeB has fewer natural polymorphisms as indicated by the long stretches with a slope of zero punctuated by a few regions of change in both the red and green lines. In addition, if SpeB did have a natural polymorphism, it was more likely to be synonymous, as indicated by the larger slope of the green line relative to the red line (**Figure 1B**). By comparing the range of the y-axis scales from **Figure 1**, it is apparent that SpeA has a significantly greater number of natural polymorphisms overall. Another contrasting feature is that SpeB could not tolerate any natural insertions or deletions within its genetic sequence, while SpeA is able to tolerate several insertions and deletions throughout its sequence. This is indicated by no insertions or deletions being detected in the sequence coding for SpeB (**Figure 1B**).

Evolutionary Relationships



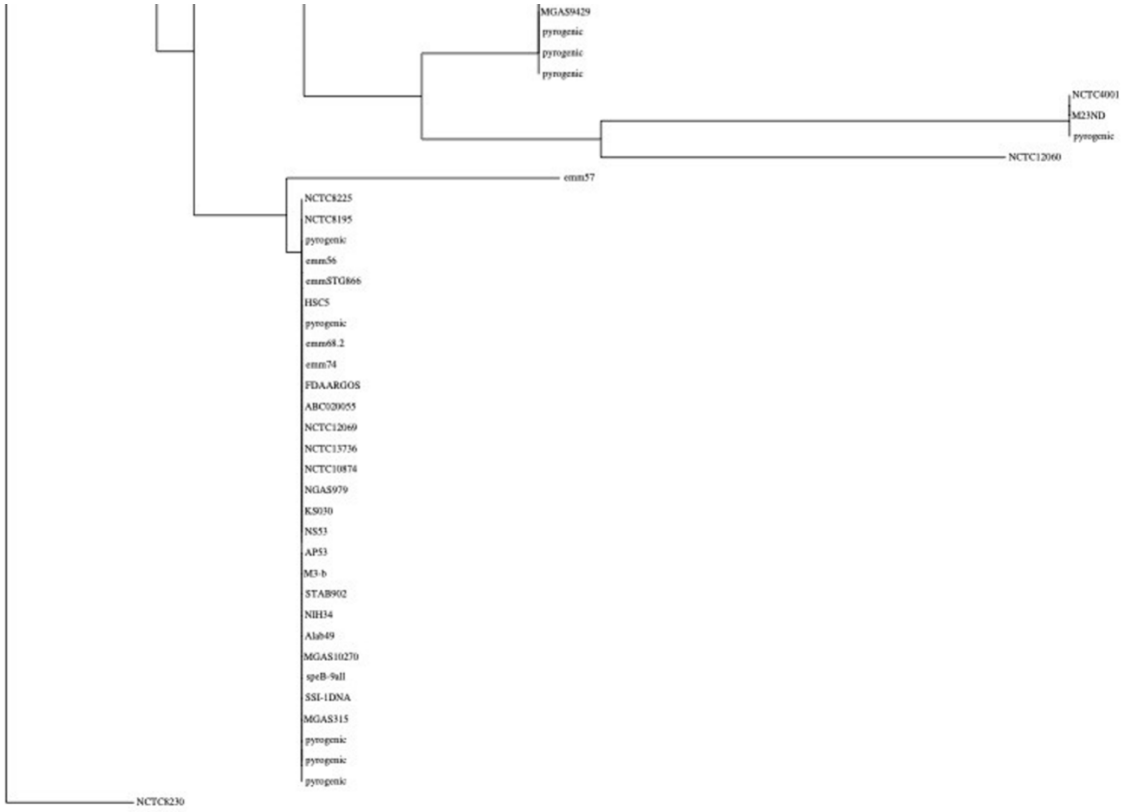


Figure 2.
Dendrogram exhibiting the evolutionary relationships among 101 strains of GAS based on natural polymorphisms found within *speB* sequences.

The genetic relatedness of SpeB shows divergence of multiple lineages by synonymous mutations, with nonsynonymous mutations appearing randomly (**Figure 2**). Occurrence of nonsynonymous mutations in neighboring clades was assumed to be inheritance from a common source. All nonsynonymous mutations examined were unique, without evidence of ancestral inheritance. Overall, these plots help quantify the evolutionary pressures on SpeB.

3-D Mapping of Natural Polymorphisms

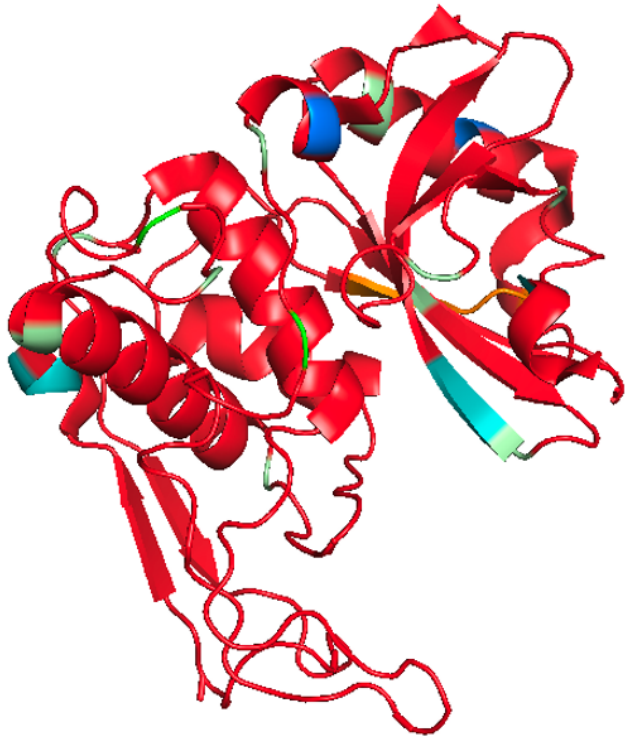


Figure 3. 3-D ribbon diagram of the mature form of SpeB (red). The 3-D diagram was obtained from structure 4D8I on RCSB PDB. The synonymous mutations are highlighted in green and nonsynonymous mutations are highlighted in blue. A pale shade of green or blue indicates a mutation of low frequency, whereas a dark, saturated green or blue indicates a mutation of high frequency (> 0.5).

This 3-D visualization from PyMOL helped determine polymorphisms of high interest. Polymorphisms were considered to be of highest interest if they were near the active site (C192), or dramatically altered an amino acid's polarity or charge. The visualization also corroborates that SpeB is highly conserved because all the natural polymorphisms were able to be simultaneously mapped with majority of the structure retaining the original red color, indicating that no mutations were found at that amino acid site (**Figure 3**).

Discussion

The hypothesis that SpeA will have more genetic variation relative to SpeB was supported. This can be seen most directly in **Figure 1** where the ratio of synonymous to nonsynonymous mutations for SpeA (**Figure 1A**) and SpeB (**Figure 1B**) were plotted. Cumulatively, SpeA had more synonymous mutations, nonsynonymous mutations, and insertions or deletions relative to SpeB, despite presence in a lesser number of isolates. Additionally, SpeA had more nonsynonymous mutations rather than synonymous mutations, as indicated by the red line having a steeper slope than the green line. All these markers suggest that SpeA is undergoing diversifying selection. Diversifying selection rapidly introduces variation to gene sequences, which leads to increased genetic diversity, a prerequisite for adaptive evolution (Zhang and Rosenberg, 2002). To contrast, SpeB had significantly fewer natural polymorphisms of each type, as indicated by the long stretches of with a slope of zero, signifying that there are no natural polymorphisms at those locations. Furthermore, SpeB had more synonymous mutations rather than nonsynonymous mutations, as indicated by the green line crossing the red line and ending at a higher value. This data suggests that SpeB is undergoing purifying selection, which is thought to reduce genetic diversity and preserve biological function (Cvijovic et al., 2018)

One limitation of this analysis is that there were more sequences of *speA* analyzed (174 sequences) when compared with *speB* (101 sequences). This could partially account for why more natural polymorphisms were found within the *speA* sequences, but the disparity is so large that the conclusions drawn above are still relevant. In part, *speB* was undersampled because identical sequences were automatically consolidated in databases. In future works, if more precise frequencies are required, each sequenced genome will be independently examined.

Chapter 3: Examining the Functional Consequences of Heterogeneity in SpeB

Since there was a large number of polymorphisms of SpeA to analyze, and the existing assays for evaluating superantigen function require the use of human donor blood that was difficult to regularly source under university COVID-19 precautions at the time of this research, I focused on the natural polymorphisms of SpeB. We additionally reasoned that as an enzyme, SpeB might be more sensitive to nonsynonymous mutations, and we might have greater likelihood that one could introduce a significant change in activity (**Figure 4**). Based on the comparison of the amino acid side chains before and after the mutation, I developed several hypotheses about how the polymorphisms could impact SpeB's function:

1) If nonsynonymous mutations impact the ability of the protein to fold correctly, altering the overall conformation, then there will be a reduction in or a complete lack of functionality.
2) If nonsynonymous mutations are near the sites used to bind target receptors, then function will be impacted (either positively or negatively).
    a) If the binding interface is disrupted, the mutation can negatively impact activity.
    b) If there is increased binding affinity, the mutation can positively impact activity.
3) If the nonsynonymous mutation is neither at a binding site nor impacting the protein's conformation, then the protein's function will not change.

I investigated these hypotheses via site-directed mutagenesis to change the sites with natural polymorphisms to alanine, a biochemically neutral amino acid. I then purified each recombinant mutant form of SpeB and quantified the activity of the enzymes with a fluorescent activity assay. I found that there is still significant function resulting from the proteins with the natural polymorphisms, indicating that there is either no significant difference in the function of SpeB, or a slight reduction in function.

Methods

3-D Modeling of Site-Directed Mutagenesis

Based on the methodology shown in **Figure 4**, I decided to further investigate the nonsynonymous mutations found within SpeB. Using the Alphafold model of SpeB from RCSB PDB, I introduced each mutation onto the 3-D ribbon structure using PyMOL (AF-P0C0J0-F1-model_v2, Jumper et al., 2021; Varadi et al., 2021) to visualize how the change in amino acid side chain could impact the protein.

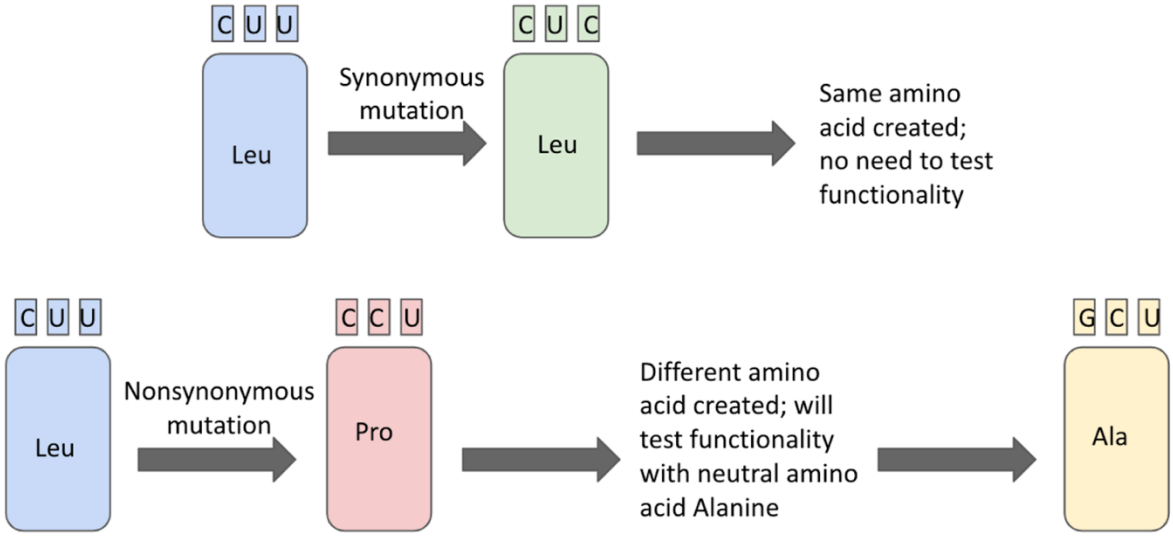


Figure 4.
Methodology for pursuing site-directed mutagenesis with the nonsynonymous mutations.

Designing Primers for Mutagenesis

Primers were designed to introduce non-synonymous mutations in a variation of Polymerase Incomplete Primer Extension (PIPE) cloning previously described by Klock and Lesley (Klock and Lesley, 2009). The forward and reverse primers were at least 20 nucleotides

long and ordered from a commercial vendor (IDT). Using a codon chart, the mutation was designed to change the amino acid to be alanine using the minimum number of base pair changes. The viability of the primer was confirmed using the Melting Point Calculator from New England Biolabs (https://tmcalculator.neb.com/#!/main) to find a melting point between 63-65°C. For the forward primer (5' to 3'), the nucleotide sequence could be copied. For the reverse primer (3' to 5'), the reverse complement of the nucleotide sequence was used. Using this process, the following primers were designed (the codon with the mutation is underlined):

G17A Forward:	5' GCATTAGCAGGATTTGTTCTTGCT 3'
G17A Reverse:	5' AAATCCTGCTAATGCTAAAAGACTTAATAATC 3'
E130A Forward:	5' ATCAAAGCAAACAAAAAATTAGACACTACT 3'
E130A Reverse:	5' TTTGTTTGCTTTGATTTGTTCGACATA 3'
D154A Forward:	5' CTCCTTGCTTCAAAAGGCATTCATTAC 3'
D154A Reverse:	5' TTTTGAAGCAAGGAGAGATTTAACAACTG 3'

Insertion of Primers into Vector

The primer sequence was inserted into the vector pET-SpeB via polymerase chain reaction (LaRock et al., 2016). The master mix of each 25 µL PCR tube included 15.8 µL nuclease free water, 5 µL High-fidelity buffer, 0.5 µL dNTP, 1 µL PET SpeB, 0.3 µL Phusion polymerase (NEB), 1.2 µL of the desired forward primer, and 1.2 µL of the desired reverse primer. This reaction mix was run on a thermocycler with the samples held at 94°C for 5 minutes so that the DNA could denature. Then, the samples were cycled 30 times with each cycle being held at 94°C for 30 seconds, ramped from 55°C to 65°C for 15 seconds, and then held at 72°C for 30 seconds. Finally, the samples were held at 72°C for 5 minutes. The product was then visualized on a 1% agarose gel to verify the insertion.

Transforming Vector into *E. coli*

The mutated vectors were transformed into chemically competent *E. coli* DH5a via heat shock. 10 μL of each sample was incubated on ice for 30 minutes and then submerged in a 42°C water bath for 45 seconds. 200 μL of SOC was then added allow the *E. coli* cells to recover. After incubating at room temperature for an hour, the samples were spread on plates and allowed to grow at 37°C overnight. Colonies were then restreaked on LB + 50 ug/ml kanamycin plates to verify the successful transformation of pET-SpeB, which carries a kanamycin resistance marker. Once colonies grew, plasmids were isolated from single colonies (Qiagen mini-prep) and submitted to Genewiz for analysis by Sanger sequencing. Clones confirmed for the correct sequence were transformed into *E. coli* BL21 cells. This process was repeated for all three mutants: G17A, E130A, and D154A.

Culturing *E. Coli* with Transformed Vector

100 mL of LB broth was inoculated with *E. coli* BL21 carrying pET plasmids encoding each SpeB clone, including wild-type SpeB from hypervirulent M1T1 5448 (positive control), a strain with an empty vector (negative control), and the three experimental strains encoding for the following mutations within SpeB: G17A, E130A, and D154A. After incubating at 37°C for 2 hours, 0.1 mM of IPTG was added to induce the expression of SpeB. The cultures were incubated at 37°C with shaking overnight.

Protein Extraction and Purification

The following procedure was repeated with all five cultures. The cultures were centrifuged and then washed with 20 mL PBS twice. To lyse the cells, the cultures were sonicated and then

centrifuged. A column with 1.0 mL of Ni-NTA affinity resin (Pierce) was equilibrated, then supernatant run through to purify the proteins. The column was washed with 10 mL PBS, followed by 10 mL of 10 mM imidazole. The target protein, SpeB, was eluted using 3 mL 300 mM imidazole. The elution was run on a SDS Page gel to confirm the presence of SpeB. The protein concentration of SpeB variants were normalized using a BSA (Bovine serum albumin; Sigma) standard. The samples were then diluted to have the same concentration as the most diluted protein, which was about 0.9 mg/mL.

Fluorescent Activity Assays

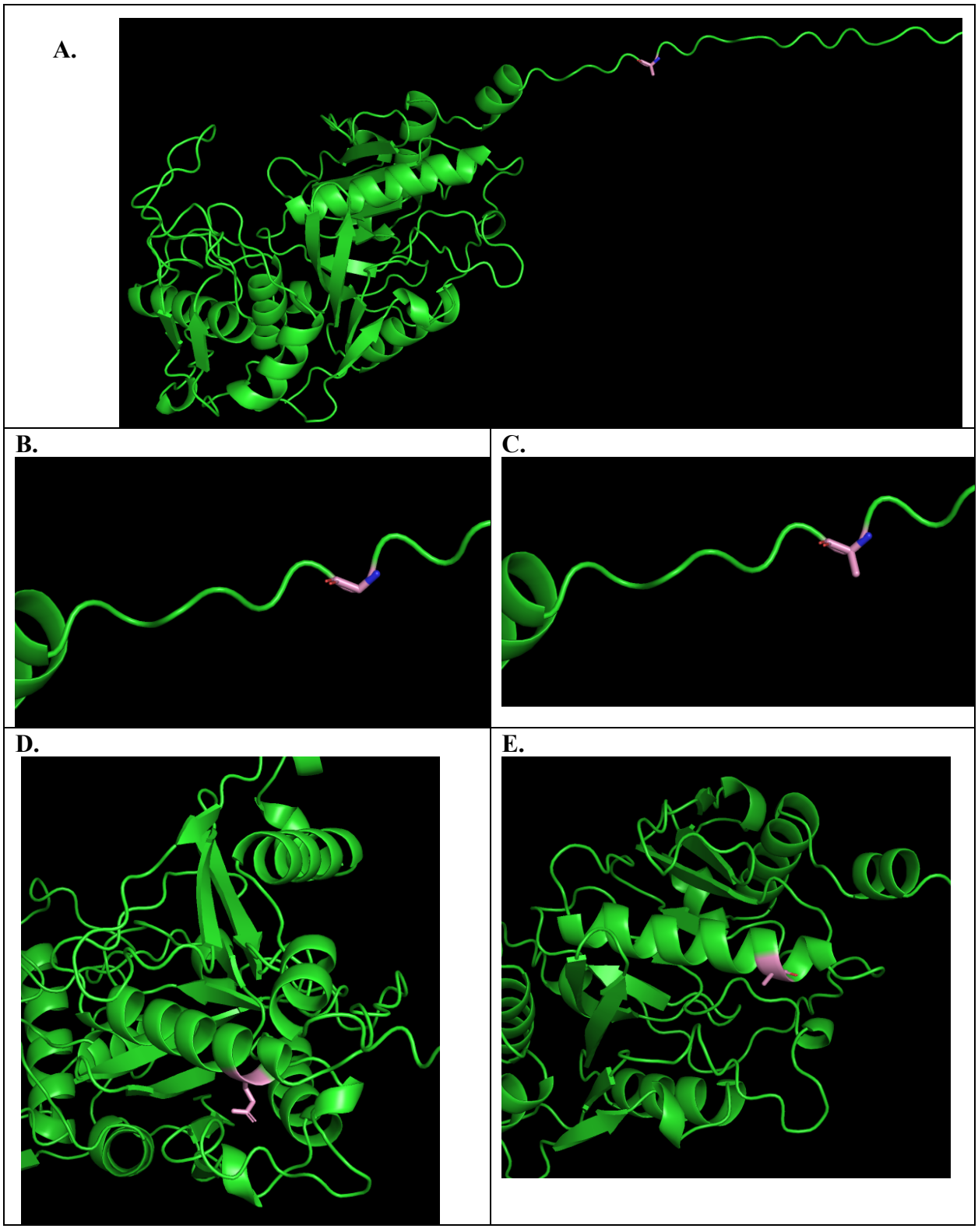
The F355-450 AMC, MCA, DAPI protocol on the VICTOR Nivo plate reader was used to quantify the activity of the SpeB variants. The assay was conducted at 37°C and lasted 30 minutes, with a plate reading fluorescence every 60 seconds, and shaking in between each reading. Substrate 73, a previously characterized substrate of SpeB that only fluoresces after being cleaved, was the substrate used for all five proteins (LaRock et al., 2016). To dilute substrate 73 to 4 μM, the aliquot had to be diluted 1:500 into SpeB buffer. SpeB buffer consists of 10 mL PBS, 1 μL tween-20, and 10 μL DTT. 10 μL of each enzyme in 300 mM imidazole was added to 20 μL of 4 μM substrate. Each variant of SpeB was tested in triplicates.

Data Analysis

To create the line graphs quantifying the activity of the SpeB variants, data at the following time points were taken: 0, 10, 20, and 30 minutes. The three replicate values were averaged and then plotted on an XY graph with time in minutes on the x-axis and relative fluorescent units as measured by the plate reader on the y-axis.

Results

3-D Modeling of Site-Directed Mutagenesis



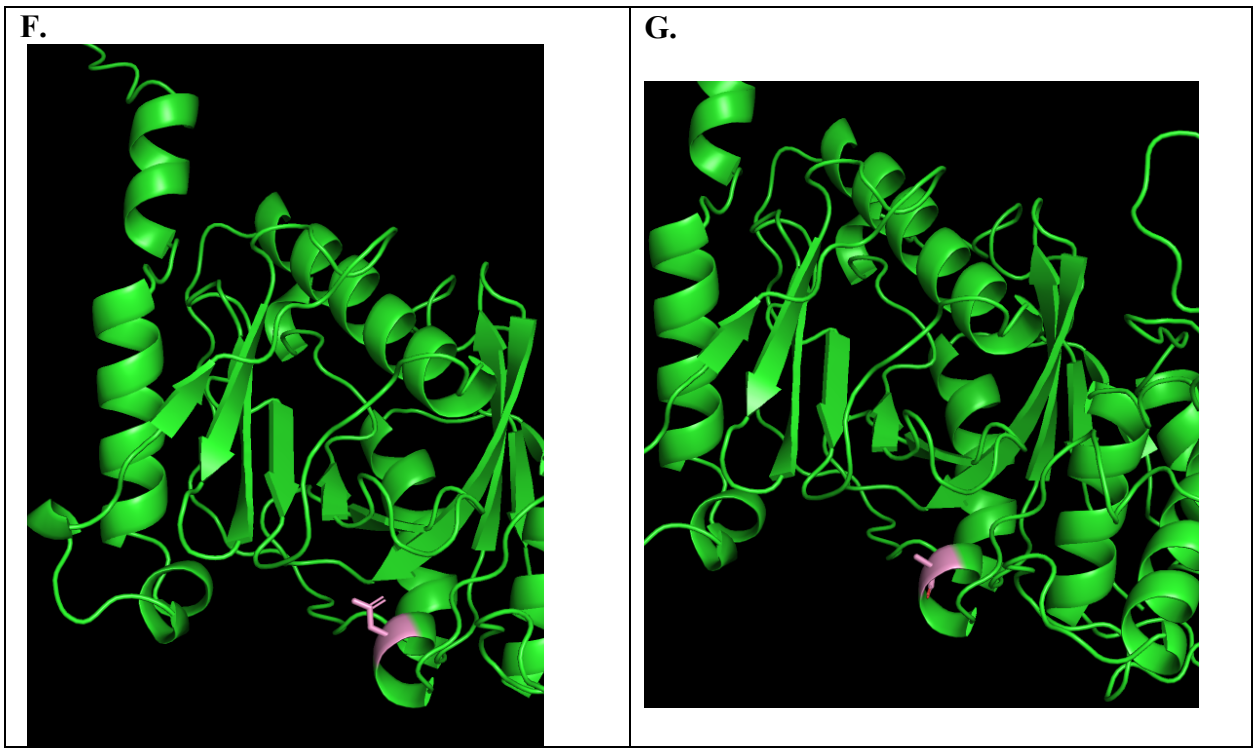


Figure 5.
Mapping of the native amino acid compared with alanine on the 3-D ribbon structure of the Alphafold form of SpeB. The targeted amino acid is shown in pink, with the side chains being represented as sticks. The top image shows a zoomed-out visualization of the entire protein, with the 17th amino acid highlighted in pink to show where the G17A mutagenesis will occur relative to the rest of the protein (**Figure 5A**). All images in the left column represent the side chains of the native residues (**Figure 5B, 5D, 5F**), and all images in the right column represent the site with alanine (**Figure 5C, 5E, 5G**). The next two images represent a zoomed-in perspective of the mutation G17A (**Figure 5B, 5C**). The middle two images represent the mutation E130A (**Figure 5D, 5E**). The bottom two images represent the mutation D154A (**Figure 5F, 5G**). Each pair of images is in a different 3-D orientation to maximize the visibility of the side chains.

The site of mutagenesis is shown in pink with the side chain of the native residue being shown on the left side and the side chain of alanine shown on the right side (**Figure 5**). For G17A, the side chain becomes slightly bulkier, but for both E130A and D154A the side chain from alanine is much less bulky compared with the native residues.

Protein Extraction and Purification

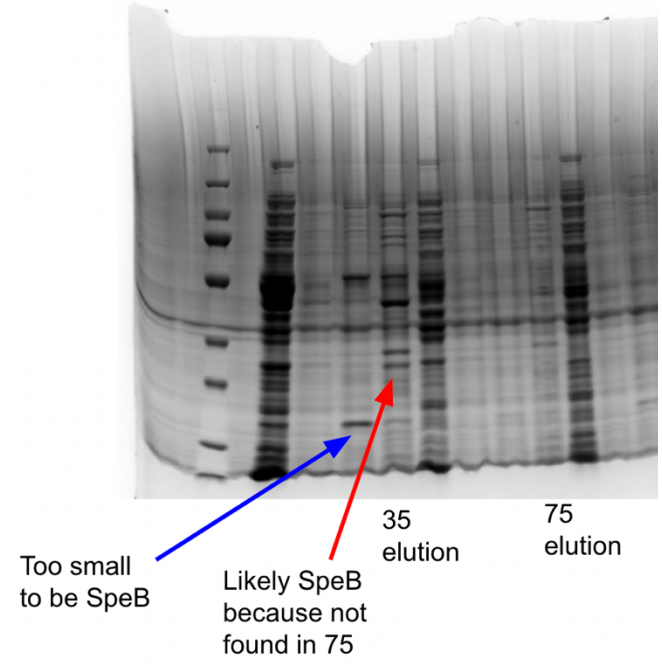


Figure 6.
An SDS Page protein gel verifying the presence of SpeB. The three proteins shown in this image were extracted from the wildtype control strain M1T1 5448 (35), the strain with an empty vector (75), and the experimental strain with mutation G17A (472). The blue arrow shows a protein that was removed via the Nickel column and the red arrow shows the band that represents SpeB.

This is an SDS Page protein gel verifying the presence and purity of SpeB after extraction from *E. coli* and affinity purification on a Nickel column. The band indicated by the blue arrow is an example of a protein that was removed by a PBS wash of the Nickel column because it is not present in the lane with the elution buffer. The band indicated by the red arrow is likely to be SpeB because it is present in the wildtype control strain M1T1 5448 (35), which is serving as a positive control, and absent from the strain with an empty vector (75), which is serving as a negative control. This is in contrast to the band directly above it which is present in the elution of both strains (35 and 75) (**Figure 6**).

Fluorescent Activity Assays

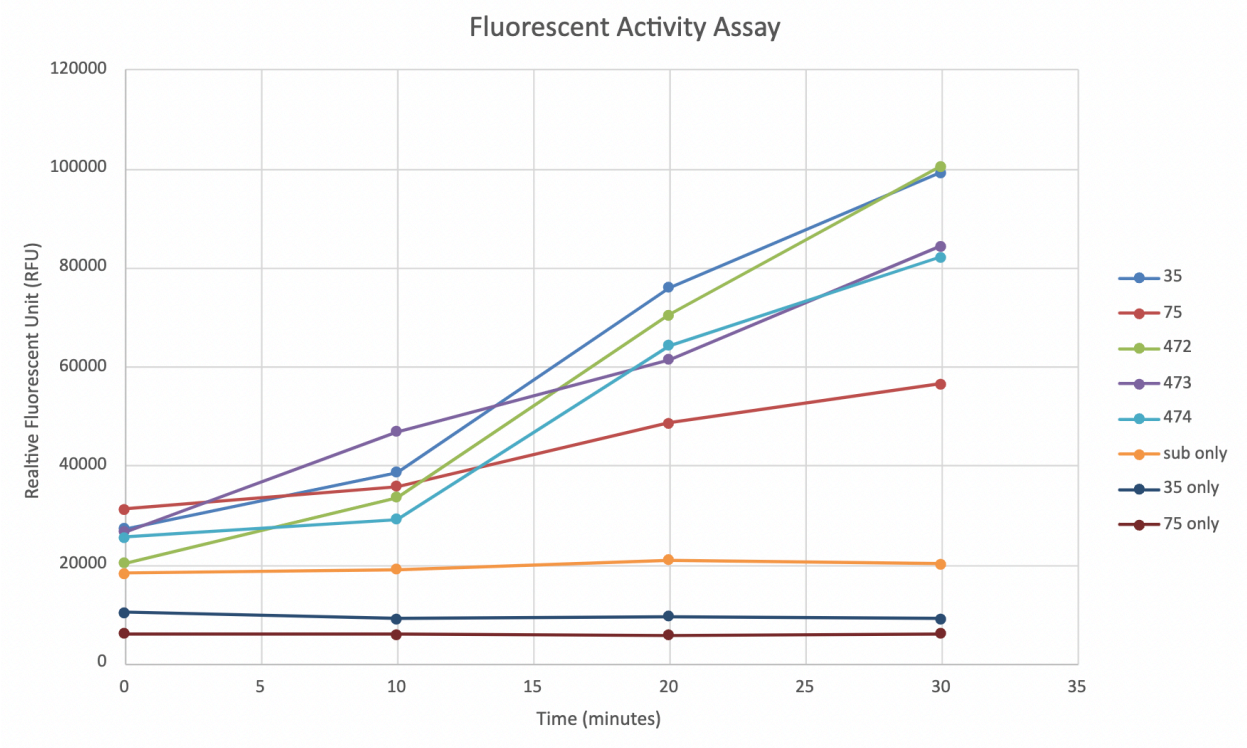


Figure 7.
These are the results are measured by the F355-450 AMC, MCA, DAPI protocol on the VICTOR Nivo plate reader. The substrate being used to measure the activity of the SpeB variants is substrate 73 at a concentration of 4 µM. The five proteins shown in this image were extracted from the wildtype control strain M1T1 5448 (35), the strain with an empty vector (75), and the three experimental strains with mutations G17A, E130A, D154A (472, 472, and 474 respectively). "Sub only" indicates only substrate was added to those wells, and "35 only" or "75 only" indicates that only enzyme was added to those wells.

Over the course of thirty minutes, there is negligible change in fluorescence detected among samples with only substrate or only enzyme added. There is a baseline increase in fluorescence that was detected from a mock purification from the empty vector (75), which demonstrates there is some background or spontaneous activation of the fluorophore substrate 73. However, the four remaining proteins have significantly greater activity. SpeB with the mutation G17A (472) has almost identical activity to the wildtype protein (35). SpeB with mutations E130A and D154A (473 and 474) have slightly worse activity when compared with the wildtype SpeB, but still exhibit functionality (**Figure 7**).

Discussion

One of my hypotheses is that if the nonsynonymous mutation is neither at a binding site nor impacting the protein's conformation, then the protein's function will not change. Overall, the results support my hypotheses, which can be seen with mutation G17A. As seen in the Alphafold 3-D model, this mutation is a part of a long chain of amino acids that gets cleaved when SpeB matures into its active form (**Figure 5A**). Therefore, it makes sense that there is no significant difference between wildtype SpeB and SpeB with the G17A mutation (**Figure 7**). The other two mutations that were examined (E130A and D154A) have a less bulky side chain when mutated to alanine (**Figure 5C, 5D, 5E, 5G**). Recalling my hypothesis that if the nonsynonymous mutations are near the sites used to bind target receptors, then function will be impacted, it seems like the binding interface was somewhat disrupted. This is supported by the reduced activity seen in SpeB with the E130A and D154A relative to the wildtype SpeB (**Figure 7**). However, there is still some function of SpeB with these mutations because there is more activity when compared with the negative control, indicating that the conformation of SpeB was not completely disrupted by these mutations (**Figure 7**).

One limitation of this analysis is that the aliquots of SpeB were unable to be completely purified. This can be seen from the several bands still present in the elution portions of my samples (**Figure 6**). Additionally, this would explain why although strain 75 is an empty vector, there is still a baseline level of cleavage of substrate 73 being detected.

Chapter 4: Future Directions and Conclusions

In chapter 2, we saw that SpeA is undergoing diversifying selection, with several insertions or deletions being tolerated and a higher prevalence of nonsynonymous natural polymorphisms compared with synonymous ones (**Figure 1A**). In contrast, SpeB is undergoing purifying selection, since in comparison of distantly related strains, SpeB appears to be unable to tolerate any insertions or deletions. SpeB also had more synonymous mutations rather than nonsynonymous mutations (**Figure 1B**). Since SpeA had a significantly greater quantity of mutations relative to SpeB, this suggests it is not an intrinsic property of GAS or just divergence of strains, but that these two proteins are under different selective pressures (**Figure 1**). This could be explained by the fact that SpeB is a protease found in all strains of GAS and has many essential functions, such as cleaving other proteins to activate them and interacting with surface proteins (Chuan and Wu, 2008). In contrast, SpeA is a superantigen that is not expressed in all strains of GAS. Some strains have other superantigens that may have redundant function. Superantigens can help GAS colonize its host and cause more severe infections, as is the case with STSS, but it is not essential for GAS in some animal infection models (Chatellier et al., 2000). Together these observations would explain why more natural polymorphisms are tolerated in *speA*, where losing the superantigenic properties does not pose a lethal risk to GAS.

In chapter 2, I also show the dendrograms for SpeB that were created by using synonymous polymorphisms to map the evolutionary distance of alleles (**Figure 2**). Given the extreme conservation of SpeB, this could be useful for the phylogenetic characterization of GAS as a core gene, since the highly dynamic and pleotropic nature of its genome can make analysis of non-clonal strains a challenge (Olsen et al., 2015). It is important to investigate the evolutionary history of GAS to understand why specific strains caused epidemics in the past and to predict emergent

strains more accurately in the future. This data can also be used to develop effective public health policies and identify broadly important and conserved proteins to target for the development of a vaccine, which is critically lacking for GAS (Nasser et al., 2014).

In chapter 3, I showed that the natural nonsynonymous mutation occurring at site 17 likely has no impact on the function of SpeB. This is probably because that portion is removed during the maturation of SpeB and does not appear to impact that process (Doran et al., 1999). My data supports this hypothesis because mutating the native glycine residue to an alanine (G17A) caused no reduction in functionality when compared with wildtype SpeB from the reference strain M1T1 5448 (**Figure 7**). This suggests that there is no change in the mature conformation of SpeB and that SpeB's proteolytic capabilities are negligibly impacted, if at all.

Data from chapter 3 also indicated that SpeB proteins with mutations E130A and D154A show some reduction in function when compared with wildtype SpeB (**Figure 7**). Neither amino acid is known to be directly involved in the processing of substrates. SpeB is a cysteine protease, meaning that it uses the thiol side chain of reactive cysteine (C192) to carry out the hydrolysis of a target protein (Nelson et al., 2011). Previous studies by the lab have shown that mutations at this site (C192S and C192A) fully destroy the activity of SpeB (LaRock et al., 2016). Like in other proteases, the amino acids immediately surrounding the active site (C192) in the structure are expected to determine the substrate specificity of SpeB, where their combinations of charge and hydrophobicity promote the interaction with substrates and exclude non-substrates. Neither E130 nor D154 are located near this site in the structure (**Figure 5**). Therefore, we expect these mutations may more subtly alter the structure of SpeB. For example, D154 is part of a loop that contains Q162, a highly hydrophobic glutamine amino acid that in the structure of SpeB located less than 7 Angstroms from the reactive C192 when measured in PyMOL (Schrödinger).

To more precisely analyze the functional consequences of heterogeneity within SpeB, I hope to repeat the activity assays with several concentrations of substrate 73. This will generate the date needed to create Michaelis-Menten Models of the enzyme activity, which will allow me to calculate and compare the enzymatic velocity of the variants of SpeB more accurately. Furthermore, examining the natural polymorphisms of these sites and comparing their activities to the alanine mutants will allow me to uncover more about the importance of that protein position and which amino acid it holds.

Overall, many alleles of *speA* and *speB* have been identified, but the functional implications of these differences have not been adequately explored. Future studies should continue to pursue the functional consequences of these natural polymorphisms and examine other Spe proteins to determine exactly how they impact the overall virulence of GAS.

References

Carapetis, Jonathan R, et al. "The Global Burden of Group A Streptococcal Diseases." *The Lancet Infectious Diseases*, vol. 5, no. 11, 2005, pp. 685–694., https://doi.org/10.1016/s1473-3099(05)70267-x.

Chatellier, Sonia, et al. "Genetic Relatedness and Superantigen Expression in Group A Streptococcus Serotype M1 Isolates from Patients with Severe and Nonsevere Invasive Diseases." *Infection and Immunity*, vol. 68, no. 6, 2000, pp. 3523–3534., https://doi.org/10.1128/iai.68.6.3523-3534.2000.

Cvijović, Ivana, et al. "The Effect of Strong Purifying Selection on Genetic Diversity." *Genetics*, vol. 209, no. 4, 2018, pp. 1235–1278., https://doi.org/10.1534/genetics.118.301058.

Chuan, Chiang-Ni, and Jiunn-Jong Wu. "Effects of Streptococcal Pyrogenic Exotoxin B on Pathogenesis of Streptococcus Pyogenes." *Journal of the Formosan Medical Association*, vol. 107, no. 9, 2008, pp. 677–685., https://doi.org/10.1016/s0929-6646(08)60112-6.

Doran, John D., et al. "Autocatalytic Processing of the Streptococcal Cysteine Protease Zymogen. Processing Mechanism and Characterization of the Autoproteolytic Cleavage Sites." *European Journal of Biochemistry*, vol. 263, no. 1, 1999, pp. 145–151., https://doi.org/10.1046/j.1432-1327.1999.00473.x.

Drexler M; Institute of Medicine (US). What You Need to Know About Infectious Disease. Washington (DC): National Academies Press (US); 2010. II, Disease Threats. Available from: https://www.ncbi.nlm.nih.gov/books/NBK209711/

Hauser, A R, et al. "Molecular Analysis of Pyrogenic Exotoxins from Streptococcus Pyogenes Isolates Associated with Toxic Shock-like Syndrome." *Journal of Clinical Microbiology*, vol. 29, no. 8, 1991, pp. 1562–1567., https://doi.org/10.1128/jcm.29.8.1562-1567.1991.

"Infectious Diseases." *Diseases and Conditions*, Cleveland Clinic, 2018, https://my.clevelandclinic.org/health/diseases/17724-infectious-diseases

Inhorn, Marcia C., and Peter J. Brown. "The Anthropology of Infectious Disease." *Annual Review of Anthropology*, vol. 19, no. 1, 1990, pp. 89–117., https://doi.org/10.1146/annurev.an.19.100190.000513.

Jumper, John, et al. "Highly Accurate Protein Structure Prediction with Alphafold." *Nature*, vol. 596, no. 7873, 2021, pp. 583–589., https://doi.org/10.1038/s41586-021-03819-2.

Kapur, Vivek, et al. "A Conserved Streptococcus Pyogenes Extracellular Cysteine Protease Cleaves Human Fibronectin and Degrades Vitronectin." *Microbial Pathogenesis*, vol. 15, no. 5, 1993, pp. 327–346., https://doi.org/10.1006/mpat.1993.1083.

Kazmi, Shahana U., et al. “Reciprocal, Temporal Expression of Spea and SpeB by Invasive M1T1 Group A Streptococcal Isolates in Vivo.” *Infection and Immunity*, vol. 69, no. 8, 2001, pp. 4988–4995., https://doi.org/10.1128/iai.69.8.4988-4995.2001.

Kline, J B, and C M Collins. “Analysis of the Superantigenic Activity of Mutant and Allelic Forms of Streptococcal Pyrogenic Exotoxin a.” *Infection and Immunity*, vol. 64, no. 3, 1996, pp. 861–869., https://doi.org/10.1128/iai.64.3.861-869.1996.

Klock, Heath E., and Scott A. Lesley. “The Polymerase Incomplete Primer Extension (PIPE) Method Applied to High-Throughput Cloning and Site-Directed Mutagenesis.” *Methods in Molecular Biology*, 2009, pp. 91–103., https://doi.org/10.1007/978-1-59745-196-3_6.

Korber B. (2000). HIV Signature and Sequence Variation Analysis. Computational Analysis of HIV Molecular Sequences, Chapter 4, pages 55-72. Allen G. Rodrigo and Gerald H. Learn, eds. Dordrecht, Netherlands: Kluwer Academic Publishers.

Kotb, M. “Bacterial Pyrogenic Exotoxins as Superantigens.” *Clinical Microbiology Reviews*, vol. 8, no. 3, 1995, pp. 411–426., https://doi.org/10.1128/cmr.8.3.411.

LaRock, Christopher N., et al. “IL-1β Is an Innate Immune Sensor of Microbial Proteolysis.” *Science Immunology*, vol. 1, no. 2, 2016, https://doi.org/10.1126/sciimmunol.aah3539.

Madeira, Fábio, et al. “The EMBL-Ebi Search and Sequence Analysis Tools Apis in 2019.” *Nucleic Acids Research*, vol. 47, no. W1, 2019, https://doi.org/10.1093/nar/gkz268.

Marrack, P, and J Kappler. “The Staphylococcal Enterotoxins and Their Relatives.” *Science*, vol. 248, no. 4956, 1990, pp. 705–711., https://doi.org/10.1126/science.2185544.

Musser, J M, et al. “Streptococcus Pyogenes Causing Toxic-Shock-like Syndrome and Other Invasive Diseases: Clonal Diversity and Pyrogenic Exotoxin Expression.” *Proceedings of the National Academy of Sciences*, vol. 88, no. 7, 1991, pp. 2668–2672., https://doi.org/10.1073/pnas.88.7.2668.

Nasser, Waleed, et al. “Evolutionary Pathway to Increased Virulence and Epidemic Group A Streptococcus Disease Derived from 3,615 Genome Sequences.” *Proceedings of the National Academy of Sciences*, vol. 111, no. 17, 2014, https://doi.org/10.1073/pnas.1403138111.

Nelson, Daniel C., et al. “Cysteine Proteinase SpeB from Streptococcus Pyogenes – a Potent Modifier of Immunologically Important Host and Bacterial Proteins.” *Biological Chemistry*, vol. 392, no. 12, 2011, pp. 1077–1088., https://doi.org/10.1515/bc.2011.208.

Nelson, K, et al. “Characterization and Clonal Distribution of Four Alleles of the Spea Gene Encoding Pyrogenic Exotoxin a (Scarlet Fever Toxin) in Streptococcus Pyogenes.” *Journal of Experimental Medicine*, vol. 174, no. 5, 1991, pp. 1271–1274., https://doi.org/10.1084/jem.174.5.1271.

Olsen, Randall J., et al. “The Majority of 9,729 Group A Streptococcus Strains Causing Disease Secrete SPEB Cysteine Protease: Pathogenesis Implications.” *Infection and Immunity*, vol. 83, no. 12, 2015, pp. 4750–4758., https://doi.org/10.1128/iai.00989-15.

Shaikh, Nader, et al. “Prevalence of Streptococcal Pharyngitis and Streptococcal Carriage in Children: A Meta-Analysis.” *Pediatrics*, vol. 126, no. 3, 2010, https://doi.org/10.1542/peds.2009-2648.

Shulman, Stanford T., et al. “Clinical Practice Guideline for the Diagnosis and Management of Group A Streptococcal Pharyngitis: 2012 Update by the Infectious Diseases Society of America.” *Clinical Infectious Diseases*, vol. 55, no. 10, 2012, https://doi.org/10.1093/cid/cis629.

Talkington, D F, et al. “Association of Phenotypic and Genotypic Characteristics of Invasive Streptococcus Pyogenes Isolates with Clinical Components of Streptococcal Toxic Shock Syndrome.” *Infection and Immunity*, vol. 61, no. 8, 1993, pp. 3369–3374., https://doi.org/10.1128/iai.61.8.3369-3374.1993.

The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

Varadi, Mihaly, et al. “Alphafold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models.” *Nucleic Acids Research*, vol. 50, no. D1, 2021, https://doi.org/10.1093/nar/gkab1061.

Walker, Mark J., et al. “Disease Manifestations and Pathogenic Mechanisms of Group A Streptococcus.” *Clinical Microbiology Reviews*, vol. 27, no. 2, 2014, pp. 264–301., https://doi.org/10.1128/cmr.00101-13.

Zhang, Jianzhi, and Helene F. Rosenberg. “Diversifying Selection of the Tumor-Growth Promoter Angiogenin in Primate Evolution.” *Molecular Biology and Evolution*, vol. 19, no. 4, 2002, pp. 438–445., https://doi.org/10.1093/oxfordjournals.molbev.a004099.