Functional Characterization of Fibrinolysis-Modulating Proteins Using High Throughput Phage Display

by

Zachary M. Huttinger

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cellular and Molecular Biology)
in The University of Michigan
2018

Doctoral Committee:

Professor David Ginsburg, Chair
Professor Daniel Lawrence
Associate Professor Maria Sandkvist
Associate Professor Jordan Shavit

Zachary M. Huttinger

zmhutt@med.umich.edu

ORCID iD: 0000-0001-8879-8616

DEDICATION

This thesis is dedicated to my grandmothers, Nancy Sharp and Doris Huttinger, both of whom inspired and challenged me to pursue a career in science and medicine. And to my wife, Katie Brown, who inspires and challenges me daily to be a better individual.

ACKNOWLEDGEMENTS

Thank you to my Dissertation Committee (David Ginsburg, Dan Lawrence, Maria Sandkvist, and Jordan Shavit) for your time, insight, support, and guidance. This thesis would not have been possible without your support. I would like to offer a special thanks to my mentor, David Ginsburg who throughout the past four years has constantly pushed me to be a better scientist and student and has served as a role model for me as an aspiring physician-scientist.

Thank you to the University of Michigan Medical Scientist Training Program. Ron, Justine, Ellen, Hillka, and Laurie have been invaluable resources throughout my educational training thus far and have constantly provided guidance and support. Similarly, I would like to thank the Department of Cellular and Molecular Biology, especially Bob Fuller, who helped work with me during my first year as a CMB student to navigate the process of advancing to candidacy.

Thank you to the Cardiovascular Translational Research and Entrepreneurship training grant (Jose Jalife, Dan Michele, Kerri Briesmiester, and Marilyn Cramer) for providing me with funding and giving me the opportunity to explore the entrepreneurial side of biomedical research.

A heartfelt thank you to my friends and family, especially my parents, Mike and Sally, for always trying to understand what it is I study and at least faking interest. To my amazing MSTP cohort: thank you all, you have made the last 6 years more enjoyable than I thought possible and I look forward to the coming two. I would like to especially thank Jason Keil and Danny Triner for their camaraderie during graduate school and their (sometimes) helpful comments on my research. I hope after this you both know what a phage is. To friends (and clinical mentors) far away: Tom and Tiffany, your advice and guidance has been very much appreciated and I look forward to learning more from you.

Finally, thank you to my beautiful wife Katie. I truly would not be here today without your love and support throughout graduate school. Your selfless dedication to us and our home while completing your Juris Doctor has been truly inspiration. I hope one day to be able to achieve the same level of dedication and hard work that you do on a daily basis.

TABLE OF CONTENTS

LISTS OF TABLES

LIST OF SUPPLEMENTARY TABLES

LIST OF FIGURES

LIST OF SUPPLEMENTARY FIGURES

ABSTRACT

The advent of modern high throughput sequencing (HTS) technology has resulted in the resurgence of phage display as a powerful technique in screening peptide and protein libraries for mutations and/or protein fragments with desired properties. As the protein or peptide is covalently tethered to (or "displayed" on) the surface of the bacteriophage, its corresponding genetic material may be rapidly identified and analyzed using DNA sequencing technologies. As the earliest applications of phage display were limited by the available methods (low-throughput Sanger DNA sequencing), repetitive rounds of selection and amplification were required, and only a small portion of the selected phage could be analyzed. Coupling phage display with HTS creates a new *high throughput platform*, termed high throughput phage display, with which to quickly and without bias screen vast libraries of peptides and proteins for novel binding interactions, stability, and kinetics. High throughput phage display allows characterization of the population of a library of $>10^6$ phage prior to and after selection to determine the level of enrichment or depletion of individual clones within the library. In this dissertation, I will highlight the potential of high throughput phage display with two different yet complimentary applications:

Streptococcus pyogenes (GAS) is a pathogenic bacteria that activates the fibrinolytic system via a secreted virulence factor, streptokinase, facilitating GAS dissemination throughout the human host and evasion of the host immune system. A small, bioactive

compound that inhibits streptokinase expression was previously developed and validated in a humanized mouse model of GAS septicemia. The mechanism of action and target of the compound within GAS remain unknown. A candidate gene was ruled out using genetic studies so an unbiased biochemical approach to identify GAS proteins involved in its mechanism of action was undertaken. Randomly fragmented GAS genomic DNA was cloned into a phage display vector in order to create a library of phage displaying GAS peptides on their surface. Although the utility of this library was hindered by technical difficulties during preparation, it demonstrated feasibility of constructing large and diverse phage display libraries that could be useful in other applications.

As a serine protease inhibitor (serpin), plasminogen activator inhibitor-1 (PAI-1) regulates the fibrinolytic system by inhibiting the plasminogen activators, tissue-type and urokinase-type plasminogen activator PAI-1 undergoes spontaneous conformational conversion to an inactive, latent state with a half-life of 2 hours. Two previous whole gene mutagenesis studies identified mutations that stabilize the active conformation of PAI-1 but ultimately only characterized the effects of 1.2% of all possible amino acid substitutions on PAI-1 stability. Here we report the construction and screening of a phage display library of PAI-1 variants to examine the mutational landscape with respect to functional stability. Using high throughput phage display, we surveyed 74% of all possible amino acid substitutions within PAI-1 resulting in identification of 1509 substitutions that decrease the functional half-life of PAI-1 and 492 mutations that stabilize the active conformation, including the majority of previously reported mutations. Our approach provides a massively parallel platform to comprehensively

analyze the largely unknown mutational landscape of PAI-1 and the effect of single amino acid substitutions on functional stability of PAI-1. This approach could be applied in the future to characterize the clinical significant of all possible human mutations and to generate PAI-1 variants with altered target specificity and the potential to treat human diseases due to serpinopathies.

CHAPTER 1:
INTRODUCTION

FILAMENTOUS BACTERIOPHAGE

F-specific filamentous bacteriophage ("phage") are DNA viruses that infect bacterial hosts via interaction with the F-pilus on the surface of some bacteria.[1-4] F-specific filamentous phage include fd, f1, and M13.[5] Phage particles (virions) resemble long cylindrical filaments where one end is capped by coat proteins pVII and pIX, the central portion is covered with pVIII (the major coat protein) and the opposite end is capped by pIII and pVI (Figure 1-1).[3, 6, 7] Contained within the viron body is the single-stranded phage DNA genome. Phage infection is dependent upon the N-terminal domain of the pIII coat protein, which first binds to the F-pilus on the bacterial outer membrane, inducing a conformational change in pIII. This change allows binding of a C-terminal domain of pIII to the TolA receptor in the periplasm.[1, 6, 8-10] Following binding of pIII to the TolA receptor, the phage is inserted into the inner membrane of the bacteria and its DNA genome is trafficked to the cytoplasm by an unknown mechanism.[6]

Within the cytoplasm, the single-stranded phage genome replicates episomally by a rolling circle mechanism to generate a double stranded genome that serves as a template for transcription and translation of viral proteins.[2, 6] After translation, viral coat proteins

span the inner membrane of the bacteria in preparation for virion assembly.[11] Once sufficient viral proteins have been produced, pV binds to the single stranded DNA viral genome once the double stranded replicative form has been unwound for replication, preventing further replication and sequestering for assembly into viral particles.[12]

The single stranded viral DNA contains a packaging signal that forms a hairpin and interacts with coat proteins pVII and pIX, which comprise one end of the virion.[6, 13, 14] This binding initiates virion assembly, which occurs as the phage genome is secreted into the periplasm and then the extracellular space. The single stranded DNA genome acts as a scaffold upon which the major coat protein pVIII, located along the inner membrane, is assembled as the DNA is translocated across the membrane.[15] Assembly is complete once pIII and pVI bind the terminal end of the virion and it is released from the cell. Because filamentous phage do not require host cell lysis for release of newly synthesized phage, very high amounts of virus can be continually released into cell culture media (on the order of 10^{13} phage / mL).[16]

The wild-type phage genome contains approximately 6400 nucleotides, but the virion size is capable of expanding proportionally to the genome contained within, allowing foreign DNA of up to twice that size to be cloned into the phage genome, making phage a valuable tool for cloning and other biochemical applications, especially with large inserts.[3, 7, 17, 18]

PHAGE DISPLAY

In 1985, George Smith described cloning a fragment of foreign DNA between two domains of the M13 pIII gene.[19] When these fragments were cloned in such a way that the open reading frame was kept intact, the function of pIII was not disrupted, and the peptide coded by the foreign DNA was displayed on the surface of the phage such that it could be recognized by specific antibodies.[19, 20] This discovery gave rise to the field of phage display as large libraries of these "fusion phage" could be constructed such that each phage displayed a different peptide, peptide fragment, or mutant protein on its surface.[20-22] Scott and Smith went on to develop an epitope library where each phage displayed a unique random hexamer peptide on it's surface[23] (discussed further in "Applications of Phage Display"). This study demonstrated the possibility of constructing large (2×10^8 unique clones) phage display libraries that could be screened to identify clones that interact with a specific target.[23] Because these fusion phage retained the ability to infect bacteria, a large pool of phage could be generated, screened against a target to select clones that bind, and then bound phage could transduce bacteria. These phage-transduced bacteria now carried the phage DNA sequence, which could be isolated and sequenced in order to determined DNA inserted into pIII and thus the peptide displayed on the surface that bound to the target; or the DNA could be used to generate additional phage to repeat the screen with an already purified pool of phage in order to isolate phage displaying the highest affinity to the target molecule.[18, 20, 23, 24]

The phage vector used by Scott and Smith, fUSE5, contained a tetracycline resistance gene and only a single copy of the pIII gene.[23] This allowed for quantifying phage by

infecting bacterial cells sensitive to tetracycline and counting the number of transduced units (cells conferred resistance by transduction with phage carrying tetracycline resistance).[23] As only a single copy of the pIII gene was present, upon cloning of the foreign DNA into the pIII gene, all copies displayed on the surface of the phage (5 in total) would contain the foreign peptide fusion (Figure 2A).[23] Any foreign DNA insert that altered the reading frame of pIII, or introduced a stop codon, would result in lack of production of pIII and thus non-infectious phage that could not be propagated.[23] An alternative vector for phage display, termed a phagemid, was developed by Carlos Barbas and Richard Lerner, shortly thereafter.[25, 26] Phagemids contain an antibiotic resistance cassette and are capable of replicating within bacteria because they contain both a phage and bacterial origin of replication, allowing phage-independent propagation.[27]. Inclusion of a phage packaging signal allows phagemids to be packaged within phage particles like a typical phage genome.[18, 25, 27] Unlike the fUSE vector or a typical phage genome, a phagemid only carries the pIII gene and not the rest of the major coat proteins required for phage assembly and secretion.[25-27] "Super-infection" of bacteria harboring the phagemid by a filamentous phage whose genome codes for the additional coat proteins (a "helper phage") is thus required for assembly and secretion of fusion phage particles.[25-27] In the phagemid constructed by Barbas and Lerner, the foreign peptide was cloned as a fusion to the C-terminal domain of pIII, replacing the N-terminal domain required for infection.[25, 26, 28]. The WT pIII provided by the helper phage had an intact, full-length pIII, enabling virions produced as a result of super-infection to transduce naïve bacteria and propagate the phagemid vector.[18, 25-27] Likewise, the super-infected cells producing phage particles will contain a mixture of WT and foreign-peptide fused pIII coat proteins

that will be incorporated into the secreted phage (Figure 1-2B).[18, 27] In contrast to the fUSE5 vector, a phagemid containing a foreign DNA-pIII fusion where insertion of foreign DNA that contains stop codons or alters the reading frame of pIII can still be propagated by phage particles because the WT pIII provided by the helper phage can bind the F-pilus of bacteria and initiate canonical filamentous phage transduction.[6, 18, 23, 27, 28] In such an instance, however, no fusion peptide would be displayed on the surface of the phage. Felici and colleagues demonstrated that foreign peptides could also be displayed as fusions to the major coat protein, pVIII, as well.[18, 29] Similar to the fUSE system, this fusion resulted in a multivalent display, with many copies of the foreign peptide displayed on the surface of the phage (Figure 1-2C).[29] A considerable limitation using pVIII to display a foreign peptide was the restriction in size; only very small peptides were capable of being displayed as a fusion to pVIII.[3, 6, 7, 18]

APPLICATIONS OF PHAGE DISPLAY

Phage display offered a novel approach to screening large libraries of peptides to assay binding, function, stability, and much more.[18, 27]. A significant advantage that phage display offered over other techniques for surveying proteins was that the displayed peptide was physically linked to the coding DNA, all within a particle capable of replicating within bacteria.[19] Using phage display, vast libraries of peptides can be screened against a target and then the DNA contained within the phage (or phagemid) of selected clones can be sequenced to determine the identity of the displayed peptide.

Parmley and Smith demonstrated that it was possible to utilize this replicative ability of phage in order to purify clones present at very low frequency within a starting library by introducing an amplification step during the affinity purification process.[20] Using this approach, a library was screening against a target and after non-specifically bound phage were washed away and remaining phage were eluted, the elution fraction was used to infect fresh host bacteria in order to amplify the eluted phage and construct a "post-selection" library with which the screen could be repeated (Figure 1-3A).[20] Presumably, after a single round of amplification, the background phage that did not bind the target specifically were decreased while the frequency of the desired phage was increased and amplification of this pool would result in a greater number of the desired phage to be screened in the second process. Parmley and Smith observed that such an approach resulted in enrichment of 10^4 after a single round of selection, 10^5 after two rounds of selection and re-amplification, and 10^7 after three rounds.[20] Using this iterative selection and re-amplification procedure, phage display became a powerful technique with which to study proteins.[18, 24, 27]

By inserting a string of 18 random nucleotides [$(NNK)_6$, where N = A, C, G, or T and K= G or T, sufficient to generate codons for all twenty amino acids and only one stop codon] between the N and C domains of pIII, Smith and Scott constructed a very large (2×10^8 unique clones) library of phage displaying unique, random hexapeptide on their surface.[23] These "epitope libraries" as they were called were quickly adopted and used to map the epitopes of proteins recognized by antibodies, including a monoclonal antibody raised against human plasminogen activator inhibitor-1 (PAI-1)[30] (an extensive list can be found

in Ref. [18]). Epitope libraries also displayed diagnostic potential as they were screened against sera from patients suffering from rheumatoid arthritis to identify antigenic epitopes targeted in autoimmune disease.[31, 32] This epitope mapping strategy later evolved to include cysteine residues flanking the random region to form a disulfide bond and conformationally constrain the displayed random peptide.[33] Random peptide libraries were also utilized to identify novel peptides that bind to streptavidin[22] and biotin[34]. The therapeutic potential of phage display was demonstrated in 1996, when a random peptide library was screened against the erythropoietin receptor.[35] By sequencing phage screened against the receptor and identifying the displayed peptide, Wrighton and colleagues were able to decipher a specific motif that was involved in receptor binding. Using this as a guide, they were able to rationally design novel agonists of the erythropoietin receptor.[35] Random peptide libraries have also been used for *in vivo* screens, such as when Pasqualini and Ruoslahti injected phage displaying a random sequence of seven amino acids into mice and demonstrated specific homing of certain peptides to different vascular beds.[36] They identified a novel cerebrovascular targeting peptide that when tagged to red blood cells, led to an accumulation of those blood cells within the brains of mice.[36] Finally, as identification of the displayed peptide is confirmed by the foreign DNA cloned into the phage DNA or phagemid and not the protein displayed on the phage directly, phage display has been used to characterize the substrate specificity of proteases using an approach termed "substrate display" where phage are immobilized with a random peptide between the immobilization target and the phage body, containing the DNA, and then subjected to cleavage by a protease of interest. Phage which express a peptide recognized and cleaved by the protease will be released from the immobilization

agent and can be sequenced in order to determine the substrate specificity of the screened protease. Such screens have characterized the substrate specificity of enzymes involved in coagulation, including Factor Xa[37], tissue plasminogen activator (tPA)[38], urokinase-type plasminogen activator (uPA)[39], thrombin[40], and ADAMTS13[41] (Reviewed in 42).

Phage display also offers a powerful tool for studying structure-function relationship within specific proteins. Libraries can be constructed from a fragmented gene, a randomly mutated gene, or even from a whole genome. Using a mutant phage library generated from a thrombin receptor ligand, Doorbar *et al* isolated a novel peptide that acted as a thrombin receptor antagonist and inhibited platelet aggregation.[43] Similar work has been done using random fragments of von Willebrand Factor to confirm that its platelet binding ability is exclusive to the A1 domain and that the cystine loop is required for optimal binding.[44] Jacobsson and colleagues constructed a library from randomly fragmented DNA from *Staphylococcus aureus*, demonstrating that bacterial chromosomal genetic material could be used to generate a phage display library.[45, 46] This library resulted in the identification of novel IgG and fibronectin binding proteins not previously characterized in *S. aureus.*[45]

It is worth noting that all phage studies reported before 2010 (and all but one of the studies discussed up to this point) were performed using the original procedure described by Parmley and Smith[20] and Scott and Smith[23], in which repetitive rounds of selection and re-amplification are required to enrich for clones present in the starting library at very small levels. [20] This procedure involved screening the library against a target of interest,

washing away unbound or loosely associating phage, eluting specifically bound phage, using eluted phage to infect bacteria and re-amplify the library, and repeating the screening and re-amplification process for several cycles.[19, 23, 27]. After the final selection, the eluted phage is used to transduce bacteria, allowing infected cells to grow in the presence of the selection antibiotic. Single colonies of bacteria are harvested, and the phage vector or phagemid is sequenced by Sanger sequencing to determine the identity of the displayed peptide (Figure 1-3A).[27] Sanger sequencing allows for accurate sequencing of long (~900 basepairs for a single read) DNA fragments but is low throughput.[47] Because many phage libraries are very diverse, Sanger sequencing is an impractical approach for assessing the complexity of an entire phage display libraries of up to 10^6 unique clones; in 2009, Dias-Neto and colleagues estimated that to assay 1×10^6 phage clones by Sanger sequencing, it would take well over 1000 days and $1 million.[48] This limits the power of these studies significantly, because without knowing the complexity of the starting library, it is impossible to determine to what degree a specific clone has been enriched (or depleted).

Another important limitation of conventional phage display is the requirement for repetitive rounds of selection and re-amplification required to enrich for, and thus detect by sequencing, clones that are present at very low levels in the starting library. In the early days of phage display, Jamie Scott and George Smith postulated that specific high-affinity interactions between targets and phage in the library might not be detected if those phage are lost during the library amplification steps between selections and that differences in growth rates of clones might introduce bias into the selected library.[23] The

latter was later demonstrated to be a legitimate concern by Brammer *et al* who described a mutation present in the Shine Delgarno sequence of a particular clone within a commercially available phage display library. This mutation resulted in faster amplification and false identification of the displayed peptide, HAIYPRH, as a novel zinc binding protein.[49] The same peptide had previously been identified as a novel peptide that binds to human umbilical vein endothelial cells[50], human bronchial epithelial cells[51], and polystyrene plates[52].

HIGH THROUGHPUT SEQUENCING AND HIGH THROUGHPUT PHAGE DISPLAY

High throughput sequencing (HTS; also called Next Generation or "Next Gen" sequencing) was developed in the early and mid-2000s. There are several different platforms that can be used to perform HTS, but this discussion will focus mainly on the Illumina sequencing platform (https://www.illumina.com/). In contrast to Sanger Sequencing, HTS relies on reversible chain termination in combination with fluorescent nucleotides.[47, 53] DNA molecules are annealed to a chip and fluorescent chain terminating bases are added.[53] The parent strand of DNA to be sequenced is extended one nucleotide at a time, with a camera capturing the fluorescent signal of the nucleotide incorporated at each position. After each nucleotide is incorporated, the chain termination is reversed and the next fluorescent nucleotide is incorporated. [53] Because DNA is amplified and sequenced in clusters, many thousands of sequences can be simultaneously read on a single chip.[47, 53-55] In 2008 it was estimated that several hundred machines performing

Sanger DNA sequencing would be required to match the amount of data generated by a single HTS machine in one day.[54] As a result of such high throughput sequencing, when using HTS to analyze phage libraries, iterative rounds of screening and re-amplification are not required and the entire eluted pool of phage can be analyzed after a single selection event (Figure 1-3B).

In 2010, phage display was adopted for use with HTS, termed high throughput phage display. Fowler *et al* screened a library of over 600,000 variants of a human WW domain β-sheet displayed on phage and characterized the mutational landscape with respect to binding a natural ligand. Their high throughput mutational scan validated the results of previous targeted studies on the WW domain and laid the groundwork for future high throughput phage display experiments. [56] Since then, synthetic human peptidome libraries have been displayed on phage and used to identify the targets of autoantibodies found in the serum of patients with various autoimmune diseases.[57, 58] By displaying the antibody variable region on phage, antibody discovery and maturation was expedited. [59] Phage libraries have even made headlines in the lay press, such as when a library constructed from the genomes of all viruses known to specifically infect humans was generated (https://www.nytimes.com/2015/06/05/health/single-blood-test-for-all-virus-exposures.html). Screening this library against human patient serum offers the potential to identify all previous viral infections that have resulted in an immunologic response in that patient.[60]

High throughput phage display also offers the potential to assess the complexity of the library before and after selection, making it possible to measure the precise enrichment of specific clones within the library. Not only can enrichment be measured in order to identify candidate peptides that bind the target, but depletion can also be measured to determine the sequences of peptides that do not interact with the target. Kretz and colleagues took advantage of this strategy in using HTS to refine earlier phage display approaches designed to characterize the substrate specificity and preferences of thrombin.[61] Whereas a previous study by Vindigni[40] probed the effect of eight targeted mutations on substrate specificity for thrombin, Kretz examined over 6,700 unique sequences and was able to characterize not only which residues were preferred, but which residues interfered with cleavage and were selected against by thrombin substrate phage display.[61]

The ability of HTS to examine the complexity of the phage library at every point in the screening process has also opened the door for large scale kinetic assays. Kretz *et al* constructed a library of randomly mutagenized fragment of von Willebrand factor (VWF) to examine the substrate recognition landscape of ADAMTS13. By determining the number of specific VWF displaying phage clones left uncleaved at multiple specific time points, they could estimate the apparent kinetics of ADAMTS13 for specific VWF mutants, something not previously done using phage display.[41] This "massively parallel enzyme kinetics" as it was dubbed, offers a new application of high throughput phage display to study enzymes and their reactions.

THE SERPIN FAMILY OF PROTEINS

As the name suggests, Serine protease inhibitors (serpins) are inhibitors of serine proteases, acting as suicide inhibitors, where a single serpin molecule can only inhibit one protease molecule.[62] Serpins are found in all domains of life: animals, plants, prokaryotes, and viruses.[63] In humans the serpin superfamily of proteins consists of 36 members, most of which possess an inhibitory function.[63] Some serpins are capable of inhibiting other families of proteases, including cysteine proteases.[64] Non-inhibitory serpins serve other functions, such as transporting hormones in the blood [65, 66], assisting in protein folding following a stress response as chaperone proteins[67, 68], and non-protease-related inhibition of tumor growth[69] (Reviewed in 62 and 63).

All inhibitory serpins share a common mechanism of action wherein a flexible reactive center loop (RCL) extends above the central structure of the molecule and acts as "bait" for the target protease by mimicking the natural substrate (Figures 1-4A,B).[70] The serpin and protease form a stable acyl intermediate and the active-site serine of the protease hydrolyses the P1-P1' bond within the RCL (Figure 1-4B). Following this cleavage event, the P1 residue forms a covalent bond with the active site serine, tethering the protease to the serpin.[71, 72] This is followed by the energetically favorable insertion of the N-terminal residues of the RCL into the central portion of the molecule as β-strand 4 in β-sheet A (s4A), carrying the tethered protease to the opposite side of the molecule in a pole-to-pole transition.[73] This structural transformation renders both the serpin and the serine protease inactive and the entire serpin-protease complex is targeted for degradation.[74-76] Serpins can also undergo cleavage of the P1-P1' carbonyl bond without

subsequent inhibition of the protease, an event termed substrate behavior (Figure 1-4C).[63, 77]

Serpin dysfunction has been implicated in a number of human diseases (Reviewed in 70 and 63). The first described and most well characterized is emphysema secondary to α1-antitrypsin (A1AT, *SERPINA1*) deficiency.[78, 79] It is estimated that 10% of individuals of European descent are heterozygous for one of the two most common mutations that result in A1AT deficiency.[78, 80] Polymerization of serpins, specifically neuroserpin (*SERPINI1*), has been implicated in neurological diseases secondary to inclusion body formation.[81, 82] Alteration of serpin specificity can lead to disease, as seen in α1-antitrypsin Pittsburgh, which contains a point mutation at the P1 site of the RCL.[83] This single substitution from a methionine to an arginine results in altered specificity of the serpin from inhibiting neutrophil elastase to inhibiting the pro-coagulant enzyme thrombin leading to fatal bleeding in one patient.[83]

PLASMINOGEN ACTIVATOR INHIBITOR-1

Plasminogen activator inhibitor-1 (PAI-1, *SERPINE1*) is a serpin that contributes to the regulation of thrombosis and hemostasis as the primary inhibitory of tissue-type and urokinase-type plasminogen activators (tPA and uPA, respectively).[84, 85] tPA and uPA are serine proteases that activate the zymogen plasminogen to plasmin.[86] Activated plasmin dissolves blood clots by degrading fibrin, an insoluble protein that forms the structural backbone of most blood clots.[87] Plasma PAI-1 down regulates the fibrinolytic response by inhibiting plasminogen activation, with loss of PAI-1 resulting in unrestricted

degradation of fibrin clots and a bleeding diathesis.[88-90] Individuals homozygous for a null mutation in *SERPINE1* display bleeding phenotypes that can range from mild to life threatening, while heterozygous individuals are asymptomatic.[91, 92]

PAI-1 has been implicated in a number of other diseases, including cardiovascular disease with increased levels being associated with increased risk of atherosclerosis and subsequent myocardial infarction[93-95] Elevated levels of PAI-1 have also been associated with development of diabetes mellitus type 2 as well as the metabolic syndrome.[96, 97] In malignant disease, PAI-1 has been implicated as a factor that aids in tumor growth and development[98, 99] Decreased levels of PAI-1 on the other hand, have been associated with improved wound healing in mice[100]. In 2017, it was reported that Amish individuals heterozygous for a null mutations in *SERPINE1* showed diminished signs of cellular aging and had longer life spans compared with individuals from the same community who were wildtype for *SERPINE1*.[101]

Among serpins, PAI-1 exhibits two unique properties. First, the mature, active form lacks any cysteine residues facilitating large scale recombinant production in bacterial expression systems.[102, 103] Second, PAI-1 undergoes a spontaneous transition from an active conformation to a more thermodynamically stable, but inactive, latent conformation (Figure 1-4D).[104] At physiologic temperature and pH, this latency transition occurs rapidly, with a half-life of approximately two hours.[105, 106] During the latency transition, the N-terminal portion of the RCL spontaneously inserts into the central β-sheet A as a new β-strand, (strand 4, s4A) via a mechanism that mirrors that which occurs

during serine protease inhibition. However, insertion of the RCL as s4A renders the P1-P1' residues inaccessible and the latent PAI-1 incapable of inhibiting its target protease (Figure 1-4D).[107] The latent molecule can only be reactivated by denaturation followed by refolding.[108] Interestingly, this PAI-1 latency conversion is conserved in many vertebrates, including cartilaginous fish, despite less than 50% sequence identity with humans, suggesting an important function of this conformational switch.[109]

PAI-1 was first displayed on filamentous phage in 1993[110] and until recently was the only serpin to have been displayed successfully on the surface of a bacteriophage (Scott *et al* reported display of A1AT on T7 lytic phage in 2014[111]). In 2001, Stroop *et al* screened a library of approximately 10 million PAI-1 variants that had been randomly mutagenized in order to screen for mutations that increased functional stability.[112, 113] This resulted in the identification of several clones that increased the functional half-life of PAI-1 more than 50-fold.[112] Combining these data with those of an earlier study that used randomly mutated PAI-1 expressed in λ phage to identify mutations with a prolonged half-life for the latency transition[114], only 84 of the over 7,500 possible amino acid substitutions have been characterized with respect to the functional half-life of PAI-1. Refining these screens with the power of high throughput phage display has the potential to offer insight into the remaining possible substitutions that have not yet been characterized.

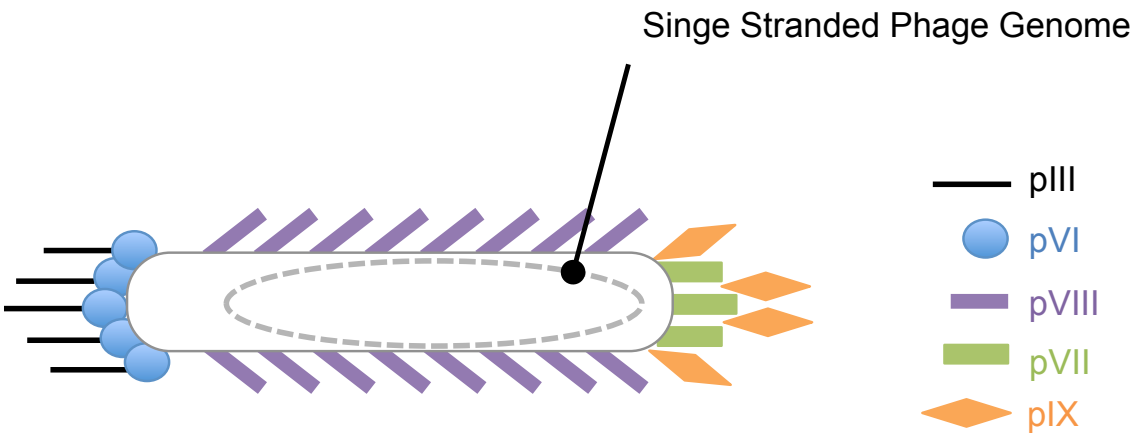


Figure 1-1: Structure of filamentous phage

Filamentous bacteriophage, such as M13, are composed of five coat proteins that surround the single stranded phage genome. pVIII, the major constituent of the coat protein forms a cylinder which is capped at either end by a pVI/pIII or pVII/pIX complex. pIII binds to the F-pilus of bacteria to initiate infection.

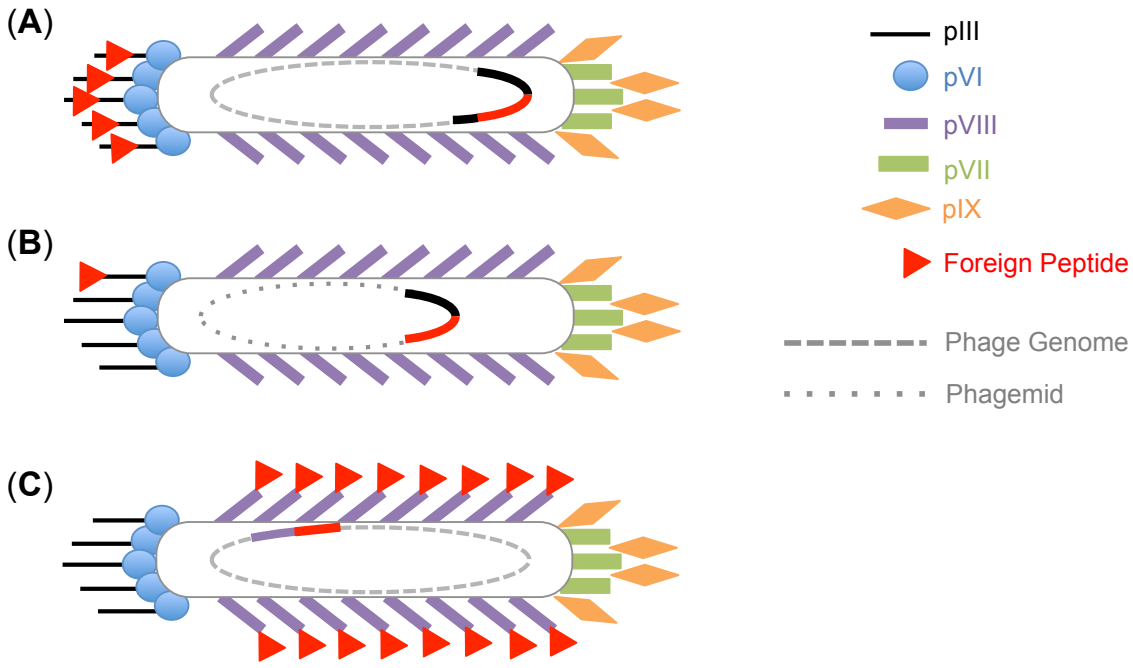


Figure 1-2: Options for display of fusion peptides on surface of filamentous phage

Fusion peptides can be displayed on different coat proteins of the filamentous phage. (**A**) depicts foreign peptide displayed using a fUSE5 vector approach, where the foreign DNA is cloned between the N-terminal and C-terminal domains of pIII. This results in display of the fusion peptide on all copies of pIII on the phage surface. As the N-terminal domain of pIII is intact, these phage retain their infective ability. In (**B**), an alternate form of pIII display is depicted, where the N-terminal domain of pIII has been replaced by the foreign DNA coding for the peptide to be displayed. This fusion gene is carried on a phagemid that does not contain all genes necessary for production of the other major coat proteins, requiring the addition of a WT helper phage for production of phage particles. As the helper phage provides the WT pIII gene, phage will display a heterogeneous mix of WT and fusion-pIII on their surface with only WT pIII maintaining their infective ability. Foreign peptides can also be displayed on different coat proteins, such as pVIII, as depicted in (**C**). Display on pVIII is limited to small peptides. A pVIII phagemid:helpher phage system can be used to produce phage with a heterogeneous mix of WT and fusion-pVIII proteins on their surface (not shown). In (**A**) and (**C**) the foreign DNA is cloned directly into the phage genome vector, whereas in (**B**), the foreign DNA is cloned into a phagemid system.

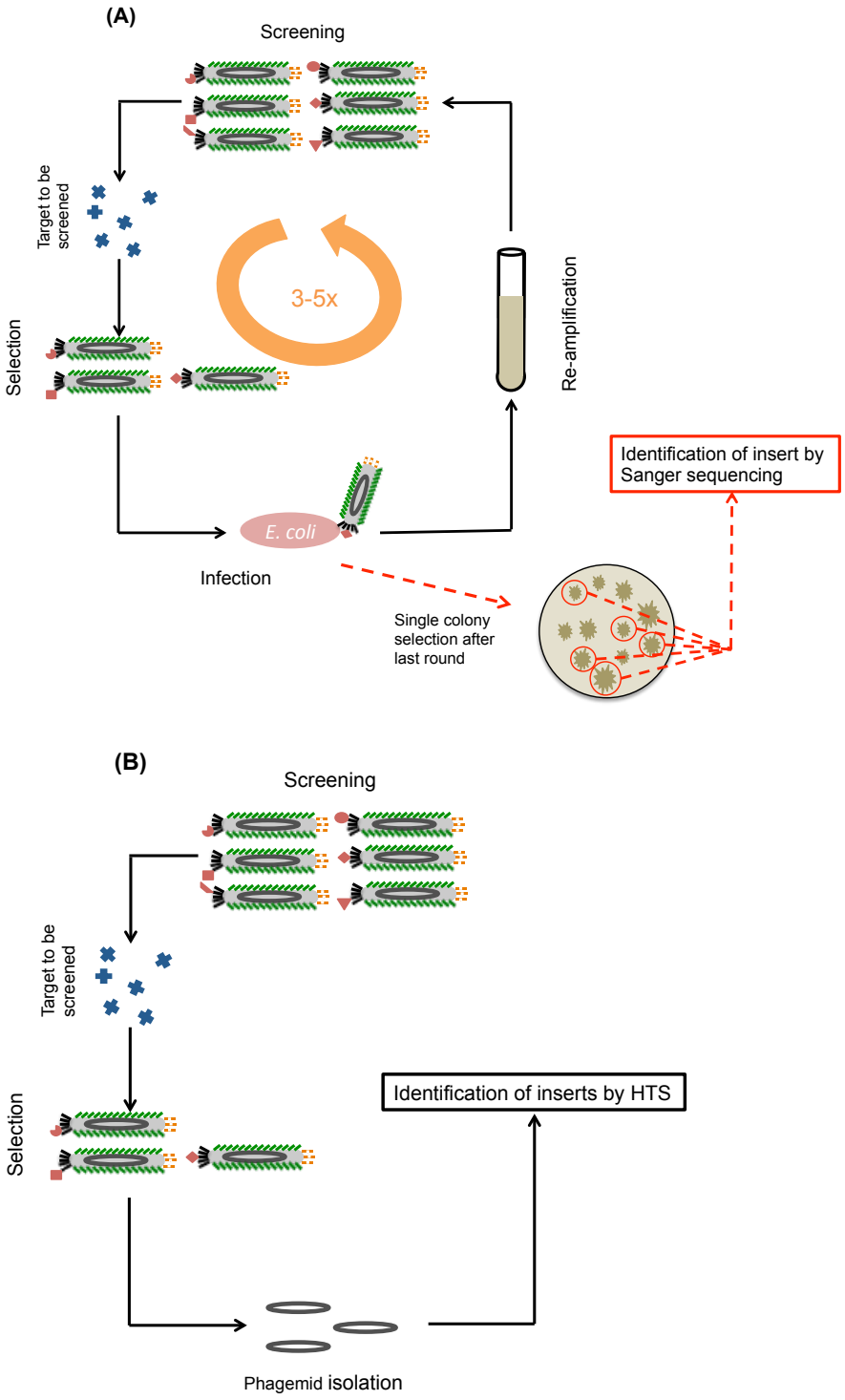


Figure 1-3: Methods of screening for enrichment from phage libraries

(**A**) The process of screening a phage display library as pioneered by George Smith and colleagues, involving several repetitive rounds of selection and library re-amplification to enrich for specific binding interactions within the library. (**B**) A high throughput phage display approach where a single round of selection is coupled to HTS and thus no re-amplification of the selected library is needed.

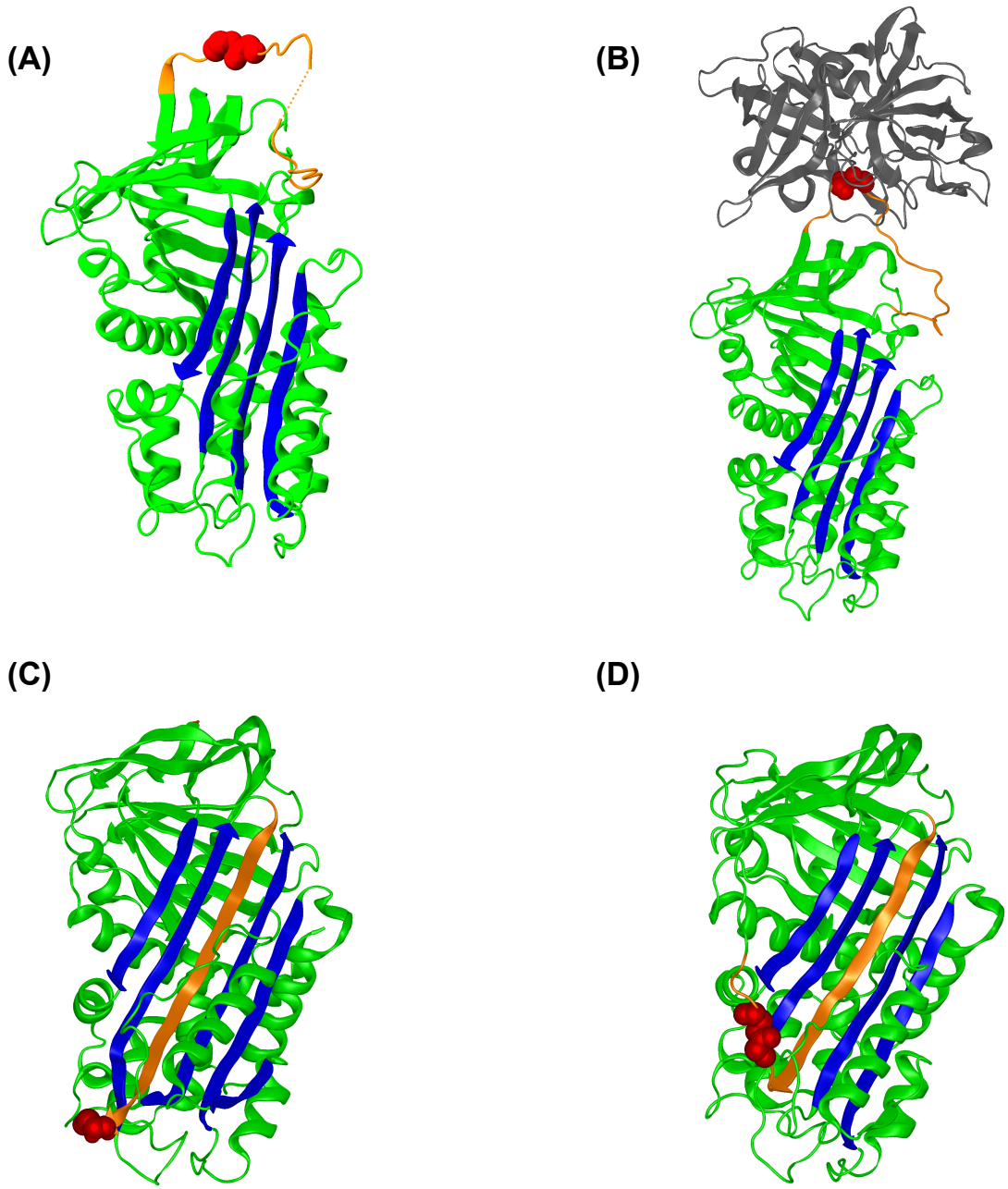


Figure 1-4: Plasminogen activator inhibitor-1 (PAI-1, *SERPINE1*) structural conformations

Structures of the (**A**) active form of PAI-1, demonstrating the flexible RCL (orange) extended above central portion of serpin body, including β-sheet A (blue) (PDB: 1DVM) (Orange dotted line indicates residues not resolved in original crystal structure). (**B**) Michaelis complex of PAI-1 with uPA (gray). P1-P1' residues are depicted as red spheres. P1-P1' scissile bond is cleaved by the protease active site serine (PDB: 3PB1). (**C**) Cleaved PAI-1 that is no longer capable of inhibition of uPA (PDB: 3CVM). (**D**) Latent PAI-1 in which RCL has inserted into β-sheet A as strand 4 (s4A). Latent PAI-1 is incapable of inhibition of uPA or tPA despite an intact P1-P1' bond (PDB: 1DVN).

CHAPTER 2:
CODY IS NOT REQUIRED FOR CCG-2979-MEDIATED INHIBITION OF STREPTOKINASE EXPRESSION IN GROUP A *STREPTOCOCCUS*

INTRODUCTION

Infectious disease is the second leading cause of death worldwide.[115] Unfortunately, several challenges face modern medicine in the fight against infectious disease, including emerging resistance to existing drugs and the lack of new antibiotic compounds. The field of antibiotic discovery grew rapidly following the identification and characterization of penicillin but has slowed considerably, with only seven new antibiotics gaining FDA approval from 2000 to 2010.[116] Occurring in tandem with this drop-off in discovery of drugs targeting pathogenic bacteria is the rise of antibiotic resistance, especially in hospital settings. Contributing to the evolution of antibiotic resistance is the selective pressure placed on the bacteria by the antibiotics themselves and the limited scope of bacterial pathways targeted by available antibiotics.[117, 118] Conventional antibiotics target bacterial processes essential for viability and replication, thus promoting the selection of bacterial cells harboring mutations that limit the effectiveness of the drug without attenuating growth and replication.[119] A novel approach to antibiotic therapy involves targeting the genes and proteins responsible for bacterial virulence factors within a given host. By targeting and inhibiting virulence rather than growth and viability, the bioactive

compound would theoretically exert less selective pressure while at the same time rendering the bacteria non-infectious.[120]

Streptococcus pyogenes, also known as group A *streptococcus* (GAS), is a gram-positive bacterium responsible for a wide range of infections and a major cause of infectious disease-related morbidity and mortality. The World Health Organization estimates GAS is responsible for over two million new cases of severe infections and invasive disease every year leading to at least 780,000 deaths worldwide.[121] Among the virulence factors produced by GAS that aid in its colonization, invasion, and dissemination is streptokinase (SK).[122, 123] SK activates the fibrinolytic system by allosterically binding to the circulating zymogen plasminogen and inducing a conformational change, thereby enabling the streptokinase-plasminogen complex to activate plasminogen to active plasmin, which can in turn degrade fibrin or generate additional units of plasmin from plasminogen, leading to an amplification of fibrinolysis and ultimate degradation of a thrombus. [124-129] Previous work has demonstrated that SK from human pathogenic GAS exhibits species specificity toward human plasminogen and the interaction between SK and human-plasminogen is crucial for virulence and infectivity in a humanized-mouse model of invasive streptococcal disease.[130]

As a known virulence factor in invasive streptococcal disease, SK was proposed as a potential candidate target for anti-virulence drug development.[123, 130, 131] With this in mind, previous work from our laboratory identified bioactive compounds that inhibit streptokinase gene (*ska*) expression without attenuating bacterial growth. A lead

compound (CCG-2979) was identified from a library of 55,000 small molecules and was shown to inhibit streptokinase activity and reduce mortality in a humanized mouse model of GAS septicemia.[132] Transcriptome profiling of GAS treated with a commercially available analog of CCG-2979 showed significantly altered expression profiles for ~29% of GAS transcripts, including a nearly 4-fold reduction in *ska* mRNA levels. [132] However, the direct target of this small molecule remained unknown.

CodY was originally identified as a transcriptional repressor in *Bacillus subtilis,* but is now thought to function as a global transcriptional regulatory protein found in many low-GC content gram-positive pathogenic organisms, including multiple *Bacillus* and *Clostridium* species, *Listeria monocytogenes*, *Staphylococus aureus*, and GAS.[133, 134] Within GAS, CodY has been shown to regulate numerous genes (17% of the genome), including *ska*.[135] The major function of CodY in bacteria is to serve as a sensor of available nutrients, specifically guanosine triphosphate (GTP) and branched chain amino acids (isoleucine, leucine, and valine).[136] When concentrations of these nutrients are sufficiently high, CodY binds to DNA to modulate the expression of various genes.[137, 138] Inactivation of GAS *codY* resulted in 2.5-fold reduction in *ska* expression and an overall gene expression profile similar to that of GAS treated with CCG-2979 (Supplementary Figure 2-1).[139] Based on these observations, we hypothesized that CodY could be the target of CCG-2979, leading to inhibition of SK expression.

MATERIALS AND METHODS

Bacterial strains

Studies were performed using virulent GAS strains UMAA2616 (serotype M1) and MEW123 (serotype M28). [140-142] All GAS strains were cultured in Todd Hewitt-broth (Becton, Dickinson and Company, Sparks, MD) supplemented with 0.2% yeast extract (US Biological, Swampscott, MA) (THY) and 100 μg/mL streptomycin.

A CodY knockout strain was generated by insertional mutagenesis using pSPC18 (a generous gift from Michael Watson Jr., University of Michigan), which harbors a spectinomycin resistance cassette (*aad9*).[143] Using NEBuilder High Fidelity DNA Assembly (New England BioLabs, Ipswich, MA) a 617 bp fragment of *codY* (+83 to +699) was cloned into the BamHI and HindIII sites of the pSPC18 plasmid (pSPC18::*codY*) using primers (5'-TAAAACGACGGCCAGTGCCAATAACACAATGGCATCTCG-3' and 5'-ATTCGAGCTCGGTACCCGGGAATCCCTGCACTTTCTAG-3') (Figure 2-1) and transformed into either competent UMAA2616 or MEW123 GAS. *codY*-null GAS was cultured in THY supplemented with 100 μg/mL streptomycin and 100 μg/mL spectinomycin. All cultures were grown under anaerobic conditions.

Streptokinase activity assay

Cultures were grown in 2xYT (Novagen, Darmstadt, Germany) as described above in the presence of either 50 μM CCG-2979[132] (treated) or 1% DMSO (untreated) for 16 hours at 37°C. Cells were separated from supernatant by centrifugation at 6800g for 10 minutes at

4°C. The supernatant was then concentrated 10-fold using size exclusion concentrator columns (10 kDa cutoff, Merck Millipore Ltd, Cork, Ireland). SK activity in the supernatant was assayed as previously described.[132, 144] Briefly, 20 µL of supernatant was added to 100 uL Tris-buffered saline (50 mM Tris, 150 mM NaCl, pH 7.4; TBS), 10 µL citrated human plasma, and 10 µL S-2403 (1 mg/mL; Chromogenix, Bedford, MA) in a 96-well plate. Absorbance at 405 nm was measured following 2 h incubation at 37°C, and SK activity was calculated relative to SK activity in wildtype ($CodY^+$; WT), untreated GAS.

Phage display library construction

UMAA2616 GAS (100 mL) was grown for 16 hours at 37°C in THY-streptomycin supplemented with 20 mM glycine. Cells were pelleted by centrifugation, washed with TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), and lysed for 1 hour at 37°C with lysozyme and mutanolysin (Sigma-Aldrich, St. Louis, MO). Lysed pellets were further disrupted by three freeze-thaw cycles before phenol/chloroform extraction of nucleic acids, which were treated with RNase and purified by phenol/chloroform extraction. The resulting genomic DNA was sheared using sonication (5 min cycles of 30 s of sonication, 30 s of rest) yielding DNA fragments between 100-1200 bp which was verified on a 1% agarose gel. Adapters pre-cut for annealing into pAYE phagemid[44] digested with NotI and AscI were ligated to sheared and end-repaired DNA, and subsequently purified by gel extraction (300-1200 bp) (Table 2-2). Adapted GAS genomic DNA fragments were cloned into the AscI and NotI digested-pAYE (pAYE-GAS) vector and electroporated into XL-1 Blue MRF' *E. coli* (Agilent Technologies, Santa Clara, CA). The depth and

insert efficiency was determined by titering transformed cells and performing PCR on single colonies using primers annealing outside the insertion site (5'-TTCATGCTGCCGGCTTTCTCGG -3' and 5'-TCGTCATCGTCCTTGTAGTCACCACCA-3'; Table 2-1). GAS DNA fragments in frame with the N-terminal signal peptide and C-terminal FLAG tag should be translated and displayed on the phage surface as a C-terminal FLAG-tag, E-tag, pIII phage coat protein fusion. [44]

Phage production and purification

Phage were prepared as previously reported.[44] Briefly, *E. coli* harboring pAYE-GAS were grown in LB Broth (Invitrogen, Carlsbad, CA) supplemented with 100 μg/mL ampicillin and 2% glucose at 37°C to log phage (OD_{600} 0.3-0.4) before transfection with M13KO7 helper phage (GE Healthcare, Chicago, IL) at a multiplicity of infection of approximately 100 for 1 hour at 37°C. Infected cells were pelleted by centrifugation at 4250g for 10 minutes at 4°C and then resuspended in 2xYT supplemented with 100 μg/mL ampicillin and 30 μg/mL kanamycin and cultured at 30°C for 16 hours. Cells were pelleted (4250g, 10 min, 4°C) and the phage-containing supernatant was cleared of remaining insoluble material by centrifugation (4750g, 10 min, 4°C). Supernatant was decanted and phage were precipitated by addition of 0.16 volume of 20% polyethylene glycol-8000/2.5 M NaCl and incubation on ice overnight before centrifugation at 20,000g for 20 min at 4°C. The phage pellet was resuspended in sterile TBS containing 0.005% Tween-20 (TBS-T). Phage titer was determined by transducing naïve XL-1 Blue MRF'

E. coli grown to log phase for 1 hour at 37°C and plating on LB-agar supplemented with 100 μg/mL ampicillin and 2% glucose.

Phage screen

Magnetic protein G beads (New England BioLabs) were washed twice with TBS before incubation with 5 μg of either a polyclonal anti-SK (PA1-85891, Pierce, Rockport, IL) or anti-streptococal pyogenic exotoxin C (speC) (ab53404, Abcam, Cambridge, UK) antibody for 1 hour at room temperature. Beads were then blocked with TBS containing 5% bovine serum albumin (TBS-BSA). pAYE-GAS phage were incubated with antibody conjugated beads for 2 to 4 hours at 4°C before being washed with TBS-BSA five times. Specifically bound phage were eluted by enterokinase (EK) cleavage (New England BioLabs) in EK Elution Buffer (20 mM Tris-HCl, 50 mM NaCl, 2 nM $CaCl_2$, 5% bovine serum albumin, pH 8.0). EK cleavage occurs at peptide sequence DDDDK, in the FLAG tag, which is cloned between the pIII coat protein and displayed GAS peptide. Cleavage reactions were performed for 16 hours at 4°C. Eluted phage were subjected to sequencing as described below.

Phage sequencing

The GAS insert of eluted pAYE-GAS phage was amplified using primers external to the insertion site (5'-TTCATGCTGCCGGCTTTCTCGG-3' and 5'-TCGTCATCGTCCTTGTAGTCACCACCA-3'; Table 2-1). PCR products were amplified and gel purified before being sheared by sonication to an average size of 300 bp. Illumina adapters containing barcodes to allow for multiplexing were ligated,

followed by High-Throughput Sequencing (HTS) on an Illumina MiSeq Platform using paired end, 150 cycle reads (v3, Illumina, San Diego, CA). DNA sequences that contained a junction between the insert and pAYE backbone aligned as peptides to the GAS genome to determine the gene encoding for the displayed peptide product. Sequences that did not contain a vector between the insert and phagemid backbone were discarded because reading frame and orientation of insert could not be determined.

Statistics

Statistics for activity assays were done using GraphPad Prism 7 (GraphPad Software, La Jolla, CA). All groups were compared using 1-way analysis of variance (ANOVA) adjusted for multiple comparisons. *P* value of <0.05 was considered significant.

RESULTS

ska promoter contains potential CodY binding sites

Although the CodY-DNA binding motif in GAS has not been established, motifs in *B. anthracis* and *L. lactis* have been identified.[145, 146] Using these sequence motifs as a guide, analysis of the 200-base pair promoter sequence directly upstream of *ska* resulted in identification of two potential binding sites for CodY, located at 194 to 184 (5'-TCGGAAgATT-3') and 173 to 163 (5'-TTGGAAgATT-3') bp upstream of the ATG start codon of *ska* (Figure 2-2). Additionally, electrophoretic mobility shift assays suggest that GAS CodY binds to the *ska* promoter sequence in a dose-dependent manner in the

presence of branched chain amino acids (valine) *in vitro* (Dr. Hongmin Sun, University of Missouri, personal communication). Taken together, these data support the hypothesis that CodY directly regulates SK expression via transcriptional control.

Construction of *codY*-null GAS

To test if CodY is required for CCG-2979-mediated inhibition of SK expression, a *codY*-null strain of GAS was generated using homologous recombination as described in "Materials and Methods". Although no colonies were obtained from the transformed UMAA2616 GAS, suggesting the spectinomycin resistance gene (*aad9*) was not successfully inserted into the bacterial chromosome, several colonies were obtained from the MEW123 GAS strain. Of the latter colonies, a single colony was selected, grown in the presence of spectinomycin, and disruption of *codY* confirmed by PCR analysis (Supplementary Figure 2-2).

CodY is not required for CCG-2979-mediated inhibition of *ska* expression

SK activity from both WT and *codY*-null strains of MEW123 GAS, was measured either in the presence or absence of CCG-2979. Treatment of *codY*-null GAS resulted in a 4-fold decrease in SK activity, demonstrating that CodY is not the only target of CCG-2979 (Figure 2-3). This was greater than the decrease seen in WT GAS when treated with compound (32.1% decrease in SK activity with treatment). There was a significant difference in SK activity of the *codY*-null GAS grown in the absence of compound compared with the WT control GAS, supporting the hypothesis that CodY regulates SK expression.[139]

GAS phage display library construction

In an effort to identify the target GAS protein(s) interacting with CCG-2979, a phage display library expressing randomly sheared GAS DNA fragments covering the entire genome of UMAA2616 strain was constructed. The library contained 3.9×10^7 independent clones, with 60.4% containing a GAS genomic DNA insert (2.34×10^7). After correcting for appropriate orientation of insert (50%), insert in frame at the 5' (33.3%), and 3' (33.3%) ends of the insert, the library was estimated to appropriately display a total of ~1.3×10^6 peptides encoded by independent GAS genomic fragments, with an average size of 525 bp, thus providing over 370-fold coverage of the GAS genome.

To test the capacity of this library to display peptides encoded by the GAS genome, the library was panned against two commercial polyclonal antibodies (anti-SK and anti-speC). Following selection, phage were sequenced and compared to the input (naïve) library in order to determine those clones that had been enriched during the screening process; however, peptides mapping to SK or speC were not the most highly enriched in either screen—suggesting that the pAYE-GAS phage display library screening approach was specific enough to detect the target peptide displayed on phage, but was not sensitive enough to detect those peptides above background within the library. (Supplementary Figure 2-3, Supplementary Tables 2-1 and 2-2). Because of the failure to enrich for specific interactions over background, the pAYE-GAS phage display library was not screened against CCG-2979 in order to identify GAS peptides that specifically interact with the bioactive compound.

DISCUSSION

Our data agrees with previous reports suggesting that the transcription regulator CodY positively regulates SK expression in GAS.[135, 139] Compared to an untreated control strain of GAS, inactivation of CodY and treatment of WT GAS with CCG-2979 led to a significant reduction in SK activity. Treatment of CodY-null GAS with CCG-2979 led to a greater reduction of SK activity than either deletion of CodY or treatment with CCG-2979 alone, suggesting that CodY is not directly required for CCG-2979-mediated inhibition of SK expression in GAS. Given the similar activity assays for *codY*-null untreated and WT-treated GAS, it is possible that CodY is indirectly involved in the mechanism of action. CodY regulates a significant portion of the GAS genome, including additional transcriptional regulators.[139, 147] Additionally, SK expression is regulated by several factors other than CodY, including the control of virulence two component regulator system (CovRS), and the FasX regulatory RNA.[139, 148-150] It is possible CCG-2979 targets a GAS protein downstream of CodY signaling, accounting for the similar expression profiles of GAS treated with the small molecule and *codY*-null GAS.

Of note, UMAA2616, the original strain of GAS used in the high-throughput screen to identify CCG-2979 as a small molecule inhibitor of SK expression, harbors mutations in both components of the Cov system, *covR* and *covS* (also known as *csrRS*), resulting in higher basal expression of SK. CodY and the CovRS system overlap significantly in

regulatory targets but are thought to mainly act antagonistically[135, 139] To date, only one study has examined the effect of *covRS* and *codY* double mutants, constructed in serotype M49 GAS strain NZ131, and found that knockout of both regulators resulted in significantly increased levels of *ska* expression, above the increase seen in a *covRS* mutant alone.[135] We were unable to generate a *codY* mutant in strain UMAA2616, possibly due to the existing *covRS* deletion in the M1 GAS strain. As a result, *codY* deletion and all SK activity assays were performed in the M28 MEW123 stain, which has an intact CovRS. It is possible that CCG-2979 behaves differently in strains lacking CovRS and further work could be undertaken with the existing double knockout strain to examine the effect of the compound in *covRS* and *codY* single and double knockout strains. Additionally, *covRS* could be deleted in our existing *codY*¯ MEW123 strain to determine if the compound behaves differently in different serotypes of GAS as serotypes M1 (UMAA2616) and M28 (MEW123) have been show to cause more invasive disease than M49 (NZ131).[151]

Given the over 1800 genes in the GAS genome, construction of single mutant strains and individual testing for SK activity in response to treatment with CCG-2979 is not a viable strategy to determine GAS proteins involved in the mechanism of action. A higher throughput screening method, either genetic or biochemical, is required to more rapidly identify peptides that interact with or mediate the activity of the bioactive small molecule. A GAS phage display library could provide such a high throughput biochemical approach; however we were unable to validate our library with antibody immunoprecipitations. Upon examination of the sequencing obtained of the unselected

library, it was noted that peptides mapping to the β-galactosidase gene were overrepresented, with almost 4% of all reads mapping to this gene (Figure 2-4). Further analysis of these peptides revealed over 99% consisted of the peptide sequence PSSRW. This peptide sequence was also found to be the protein product produced when the two adapters used to ligate the blunted GAS genomic DNA into the phagemid formed heterodimers (Figure 2-5). Given this contamination with heterodimers, the actual depth of the library was likely much lower than estimated.

Phage display libraries have been successfully generated from the genetic material of *Streptococcus agalactiae*, *Streptococcus dysgalactiae, Streptococcus equi*, *Staphylococcus epidermidis*, and *Staphylococus aureus* but to date a phage display library constructed from GAS genomic DNA has not been reported. [45, 152-156] Additionally, these previously reported bacterial DNA phage libraries were screened using the iterative panning technique where the whole library was screened, selected phage were eluted and amplified in order to create a new library. After several rounds of this screening, selection, and re-amplification, only a handful of clones were selected and analyzed using a limited Sanger sequencing approach. This method does not allow for analysis of the diversity present in the naïve, unselected library and thus makes it difficult to determine if selected clones are true positive binders or simply the product of over-representation in the starting library. "High throughput" phage display, by comparison, uses HTS to analyze the diversity of inserts in the selected and unselected libraries to measure enrichment and depletions of specific peptides after a single round of selection with no re-amplification required. This not only saves time, but more importantly avoids

introducing baises into the library that might result from repeated re-amplification within *E. coli*. Although next-generation phage display is gaining popularity, to date it has mostly been used to examine libraries of synthetic peptides or mutagenic libraries constructed from a single gene, rather than whole genome libraries.[41, 48, 56, 61, 157] Analysis of highly diverse phage libraries with different insert sizes and reading frames made from whole genomes may prove too complicated to interpret after a single round of selection, even with the increased power of HTS. Nevertheless, we hope continued advances in HTS and techniques related to phage library construction will allow us to construct improved GAS phage display libraries in the future with the ultimate goal of screening against CCG-2979 to identify its interactions with the GAS proteome.

Table 2-1: Primer sequences

Primer	5' ➔ 3' Sequence
pSPC18::*codY* Cloning	
CodY-pSPC18 For	TAAAACGACGGCCAGTGCCAATAACACAATGGCATCTCG
CodY-pSPC18 Rev	ATTCGAGCTCGGTACCCGGGAATCCCTGCACTTTCTAG
codY Genotyping	
codY For	AAAAAATGCCTAACTTATTAGAAAAAACTCG
codY Rev	TTCTATTTAAAATTCTTTTAATTTAGCAAAAATACC
aad9 For	GAATGGACTAATGAAAATGTAAATTTA
pAYE Phagemid Sequencing	
pAYE-Sequencing For	TTCATGCTGCCGGCTTTCTCGG
pAYE-Sequencing Rev	TCGTCATCGTCCTTGTAGTCACCACCA

Table 2-2: pAYE adapter sequences

Adapter	5' ➔ 3' Sequence
Individual Adapters	
pAYE-AscI For	CGCGCCATCGGGAGGAGGG
pAYE-AscI Rev	CCCTCCTCCCGATGG
pAYE-NotI For	TGGAGGAGGTGC
pAYE-NotI Rev	GGCCGCACCGCCGCCA

Supplementary Table 2-1: Most enriched proteins identified in anti-SK pAYE-GAS screen

Rank	Enrichment over Unselected	Protein
1	16.59	AAK33437, SPy_0405, Conserved hypothetical protein
2	13.37	AAK33705, SPy_0762, Hypothetical protein
3	13.17	AAK34236, SPy_1425, Hypothetical protein
4	**13.04**	**AAK34665, SPy_1979, Streptokinase (SK)**
5	11.63	AAK33205, SPy_0076, 50S Ribosomal protein B

Supplementary Table 2-2: Most enriched proteins identified in anti-speC pAYE-GAS screen

Rank	Enrichment over Unselected	Protein
1	8.15	AAK33826, SPy_0910, Putative DNA topoisomerase IV (subunit C)
2	7.88	AAK34771, SPy_2128, Hypothetical protein- phage associated
3	7.25	AAK33570, SPy_0589, Hypothetical protein
4	6.92	AAK33960, SPy_1073, 50S Ribosomal protein L7/L12
5	6.80	AAK33761, SPy_0830, Putative pyrimidine regulatory protein
8	**5.26**	**AAK33664, SPy_0741, streptococcal pyogenic exotoxin C (speC)**

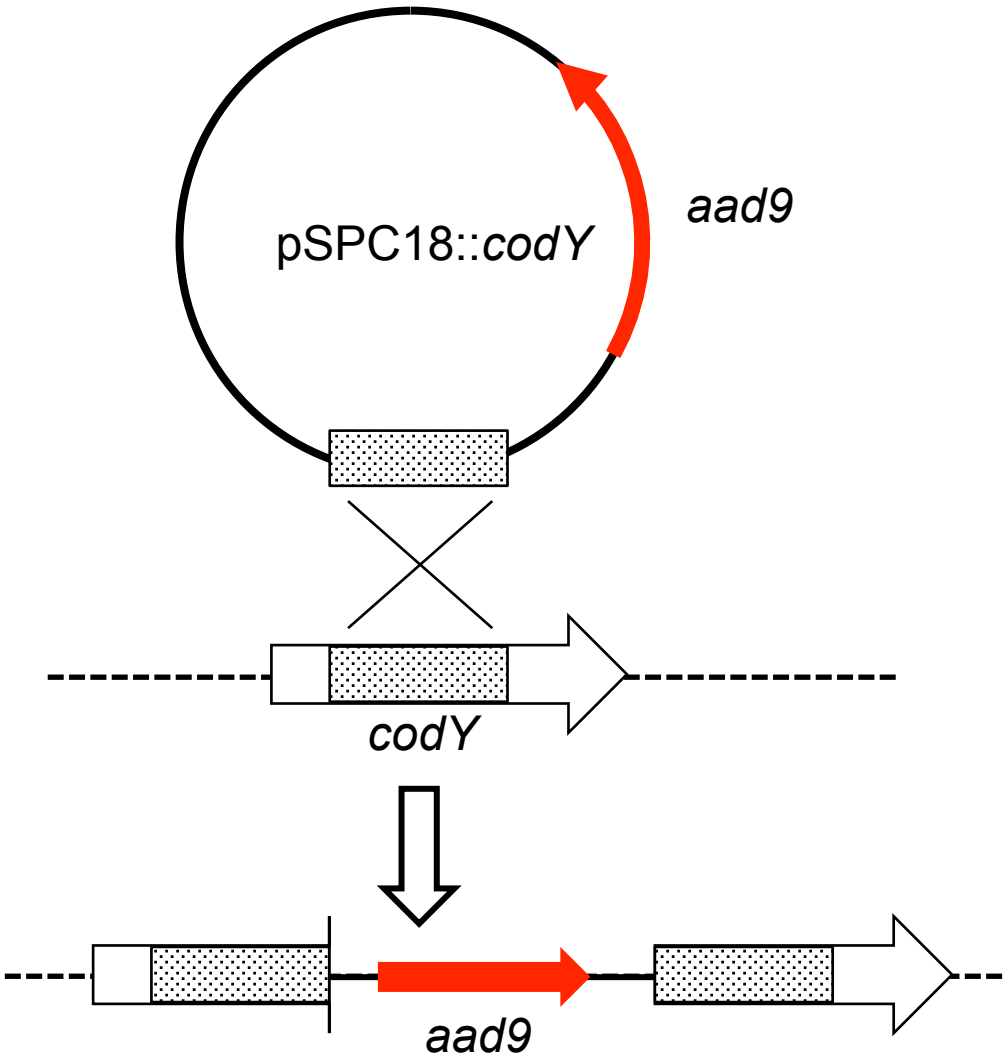


Figure 2-1: CodY insertion mutant construction using pSPC18::*codY*

codY was knocked out in GAS via insertional mutagenesis using pSPC18::*codY*. A 617 bp region of *codY* (+83 +699) was cloned into pSPC18, propagated in *E. coli*, and transformed into competent GAS. pSCP18 contains no origin of replication recognized by GAS. Thus resistance to spectinomycin via *aad9* is only conferred if the resistance cassette integrates chromosomally via homologous recombination at *codY*.

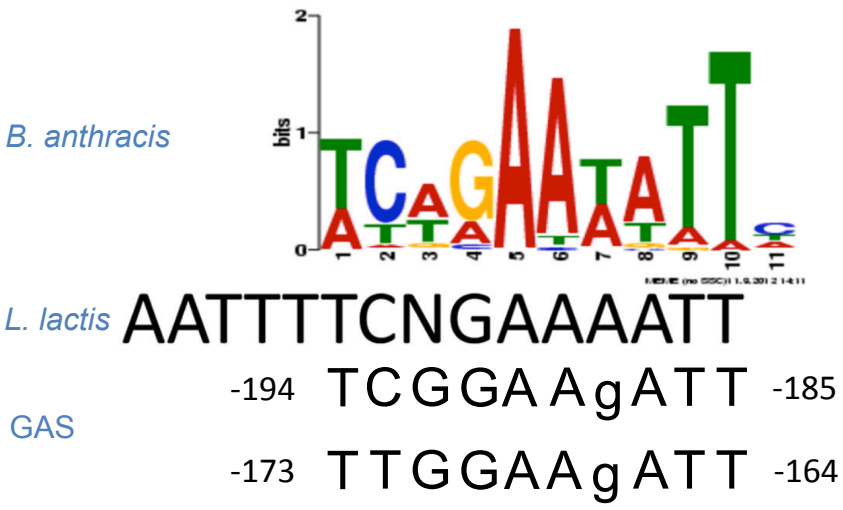


Figure 2-2: Potential CodY binding sites in *ska* promoter

Using the published CodY binding motifs from *B. anthracis* (top)[145] and *L. lactis* (middle)[146] as a guide, two potential CodY binding sites (-194 to -185 and -173 to -164, relative to the ATG start codon) were identified in the *ska* promoter directly upstream of *ska* (bottom).

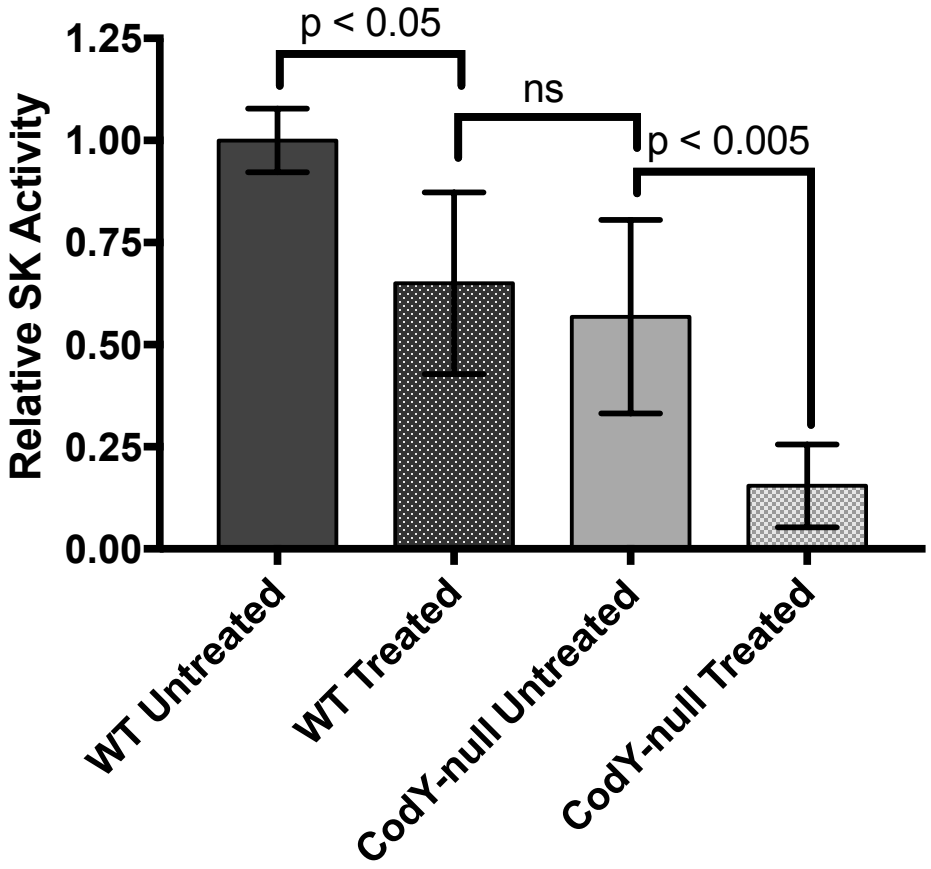


Figure 2-3: Relative SK activity in WT and *codY*-null GAS strains grown in the presence of CCG-2979 or DMSO

Relative to WT untreated GAS (control), a significant decrease in SK activity was observed in both WT-treated and *codY*-null untreated. A significant decrease in SK activity was observed when CodY-null GAS was treated with CCG-2979, suggesting CodY is not directly required for CCG-2979-mediated inhibition of SK activity. Activity is reported as the mean relative activity ± SD (n=9). Significance was determined using ANOVA adjusted for multiple comparisons.

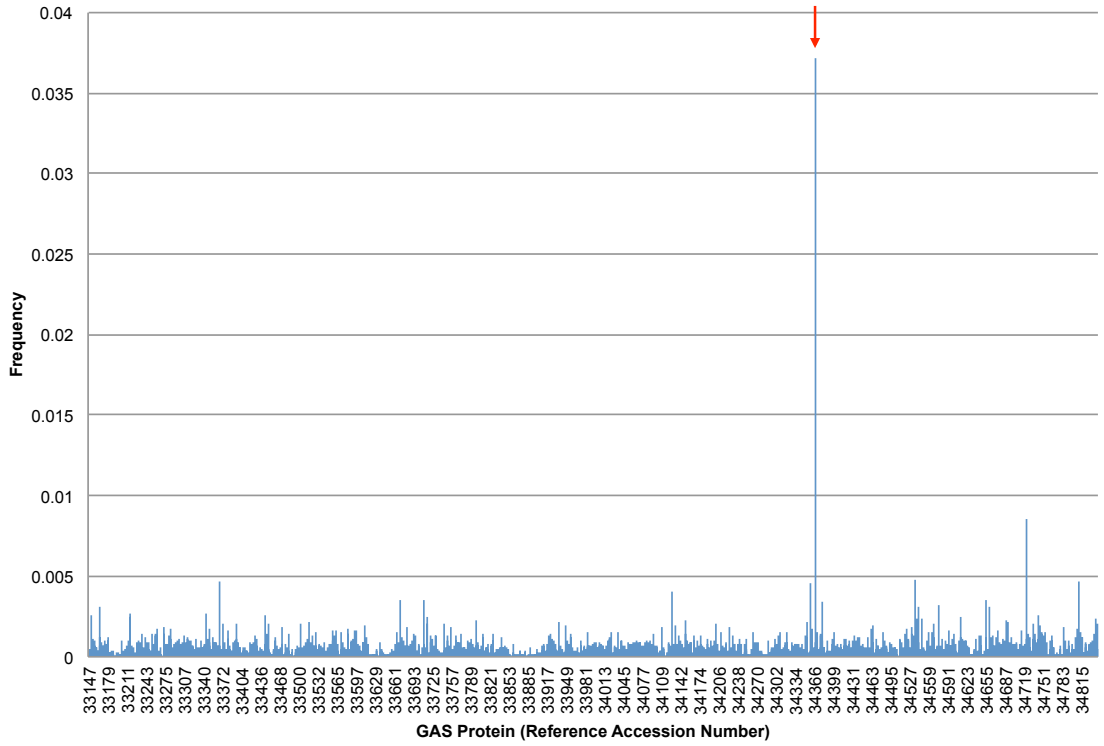


Figure 2-4: pAYE-GAS inserts mapping to GAS genes within unselected (input) library

Input GAS phage display library was subjected to HTS and inserts in pAYE-GAS phagemid were mapped against the GAS genome to identify the gene to which the displayed peptide belonged. Protein products of GAS genes are displayed on the x-axis, organized by protein reference number. Peak height reflects frequency of identification relative to the total number of mapped reads. Red arrow denotes β-galactosidase (Reference Accession Number AAK34368).

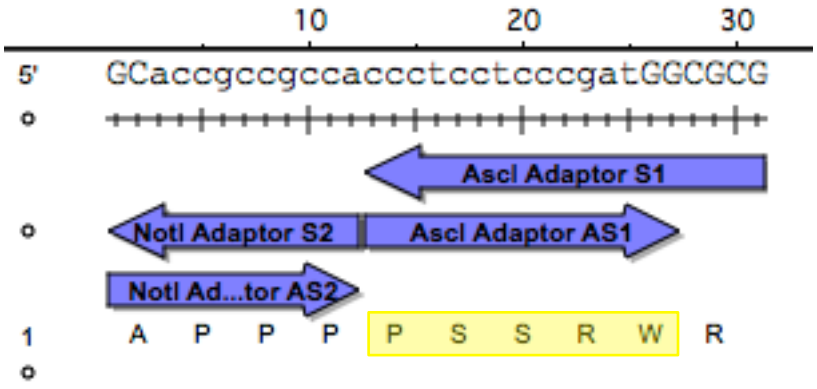
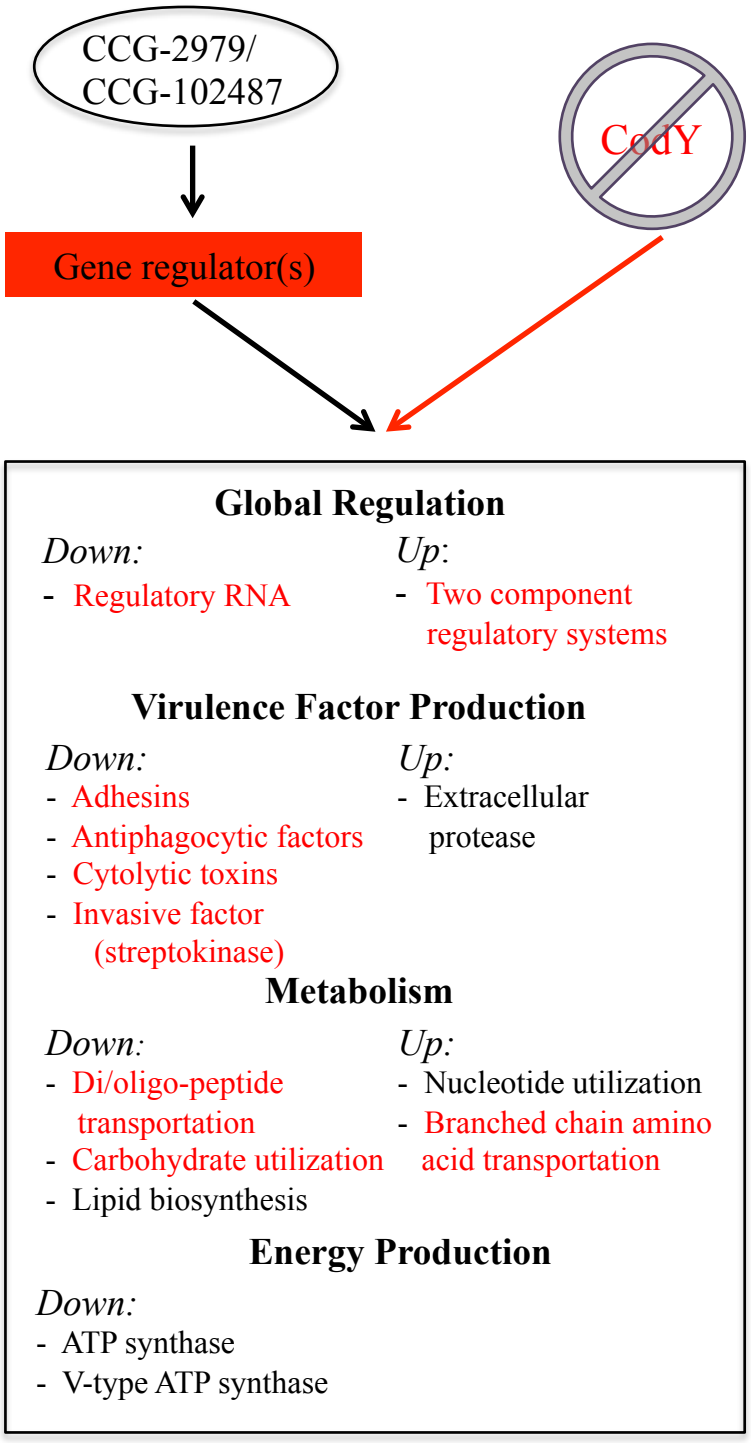


Figure 2-5: Adapter dimer mapping to β-galactosidase

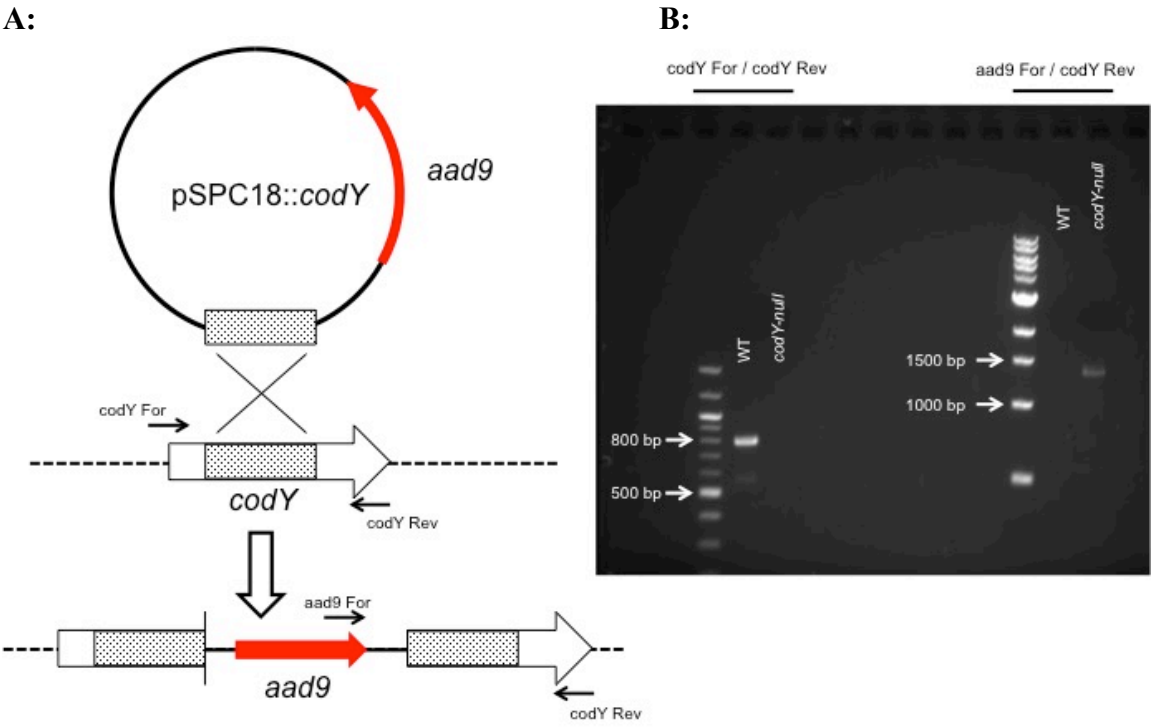
Translation of adapter heterodimer formed when adapters used to ligate blunted GAS genomic DNA fragments into pAYE form heterodimers. 99.5% of all reads (11801/11864) mapping to β-galactosidase consisted of peptide sequence PSSRW, which maps to P184-W188 of β-galactosidase, suggesting contamination of input library with adapter-heterodimers.



Supplementary Figure 2-1: Microarray data from CodY KO GAS and CCG-102487 shows overlap

Broad categories in which genes were found to be significantly (Log2Fold-Change >2) up or down regulated by microarray analysis in either GAS treated with CCG-102487 (commercially available analog to CCG-2979). Gene categories highlighted in red were common between GAS treated with CCG-102487 and a *codY*-null strain of GAS (red categories only). Relatively high degree of overlap led to hypothesis

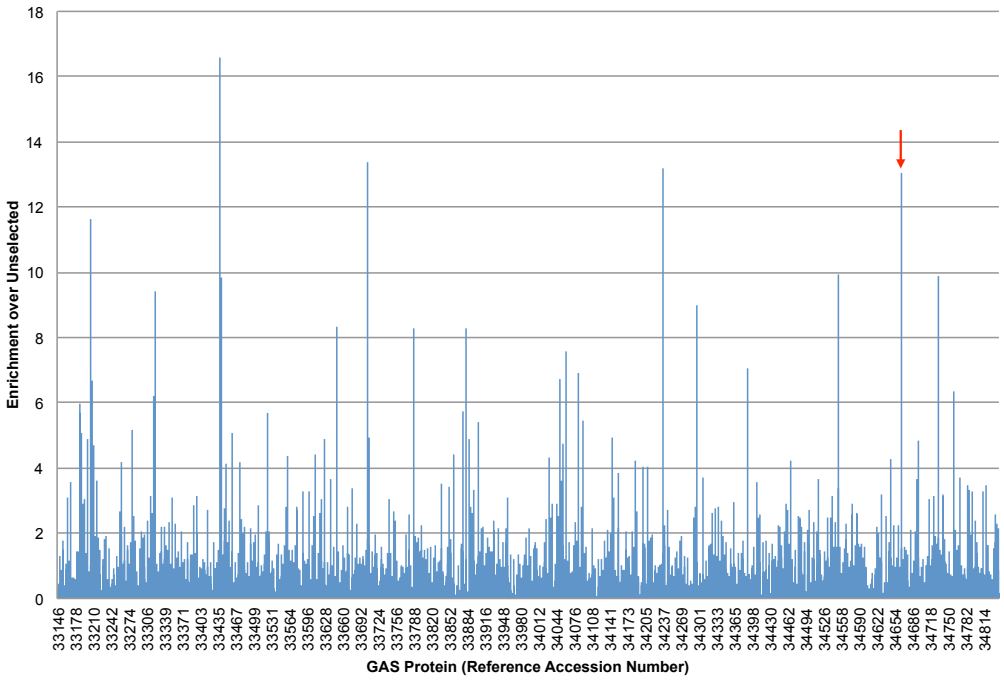
that CCG-2979 inhibited SK expression via a CodY-dependent mechanism. Figure adapted from Xu *et al.* Virulence (2012) and Sun *et al.* PNAS (2012). [158] [132]



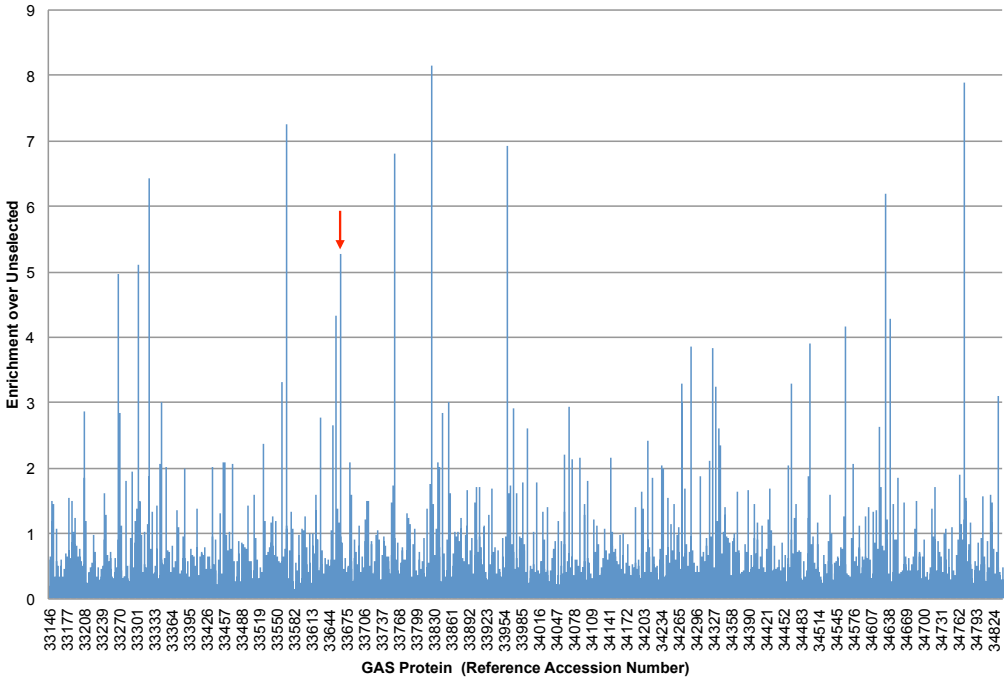
Supplementary Figure 2-2: PCR confirmation of *codY* disruption

Positive disruption of *codY* by *aad9* was confirmed using PCR. **(A)** Cartoon depiction of *codY* knockout via insertional mutagenesis and location of *codY* genotyping primers. (**B**) Agarose gel demonstrating positive disruption of *codY*. Expected product sizes for WT GAS was 794 bp using codY For/CodY Rev primer pair. Expected product size for *codY*⁻ GAS was 1364 bp using aad9 For / codY Rev.

A:



B:



Supplementary Figure 2-3: Enrichment in pAYE-GAS following selection by immunoprecipitation

Sequencing of GAS phage display library following selection using polyclonal antibodies raised against SK or speC. Eluted phage were sequenced using HTS and displayed peptide was identified by mapping GAS

insert to GAS peptidome. (**A**) Enrichment of eluted phage following selection with anti-SK polyclonal antibody. SK, denoted by red arrow, was identified as the fourth most enriched peptide, suggesting selection was not specific. Higher specificity was expected following library screening against a specific antibody. Top 5 most enriched proteins are listed in Supplementary Table 2-1. (**B**) Enrichment of eluted phage following selection with anti-speC polyclonal antibody. Red arrow denotes speC, which was identified as the eighth most enriched GAS peptide, again suggesting the screening approached lacked the specificity to enrich for specific binding interactions. Top 5 most enriched peptides from anti-speC screen are listed in Supplementary Table 2-2.

CHAPTER 3:
COMPREHENSIVE ANALYSIS OF SINGLE AMINO ACID SUBSTITUTIONS AND PAI-1 FUNCTIONAL STABILITY USING PHAGE DISPLAY

INTRODUCTION

Plasminogen activator inhibitor-1 (PAI-1, *SERPINE1*) plays a critical role in the control of thrombosis and hemostasis as the principle inhibitor of tissue-type and urokinase-type plasminogen activators (tPA and uPA, respectively).[84, 85] tPA and uPA are serine proteases that activate the zymogen plasminogen to plasmin.[86] Activated plasmin dissolves blood clots by degrading fibrin, an insoluble protein-complex that forms the structural backbone of most blood clots.[87] Plasma PAI-1 down regulates the fibrinolytic response by inhibiting plasminogen activation, with deficiency of PAI-1 resulting in accelerated degradation of fibrin clots and a bleeding diathesis.[88-90] Individuals homozygous for a null mutation in *SERPINE1* display bleeding phenotypes that can range from mild to life threatening while heterozygous individuals are asymptomatic.[91, 92]

PAI-1 belongs to the serine protease inhibitor (serpin) class of proteins.[159] Serpins are suicide inhibitors that covalently bind to their target proteases.[75] All inhibitory serpins share a common mechanism of action wherein a flexible reactive center loop (RCL) extends above the central structure of the molecule and acts as "bait" for the target protease by mimicking the natural substrate.[70]. During hydrolysis of the P1-P1' bond

within the RCL by the protease, a stable acyl intermediate covalent bond is formed between the newly generated C-terminal residue for the RCL and the active site serine, tethering the protease to the serpin.[71, 72] This is followed by the energetically favorable insertion of the N-terminal residues of the RCL into the central portion of the molecule as β-strand 4 in β-sheet A (s4A), carrying the tethered protease to the opposite side of the molecule in a pole-to-pole transition.[73] This structural transformation renders both the serpin and the serine protease inactive.[74-76] Serpins can also undergo cleavage of the P1-P1' carbonyl bond without subsequent inhibition of the protease, an event termed substrate behavior.[77]

Among human serpins, PAI-1 exhibits two unique properties. First, the mature protein lacks any cysteine residues facilitating large scale recombinant production in bacterial expression systems.[102, 103] Second, PAI-1 undergoes a spontaneous transition from an active conformation to a more thermodynamically stable, but inactive, latent conformation.[104] At physiologic temperature and pH, this latency transition occurs rapidly, with a half-life of approximately two hours.[105, 106] During the latency transition, the N-terminal portion of the RCL spontaneously inserts into the central β-sheet as s4A via a mechanism that mirrors that which occurs during serine protease inhibition but renders the P1-P1' residues of the RCL inaccessible and the serpin incapable of inhibiting its target protease.[107]

Multiple studies have used targeted mutagenesis to examine the role of individual PAI-1 residues and secondary structures in the latency transition (Reviewed in 160). Attempts to

rationally design a more stable variant of PAI-1, however, have resulted in mixed success.[160]. Mutations introduced in order to slow the insertion of the RCL into β-sheet A mostly resulted in decreased inhibitory activity due to substrate behavior of the mutant PAI-1, as the resulting rate of insertion and protease inhibition was slower than deacylation of the serpin:protease intermediate.[161] Some of these mutations, however, did result in slight prolongation of the latency transition half-life.[162-164]

Two previous studies used randomly mutagenized PAI-1 to perform an unbiased screen aimed at characterizing mutations that stabilize the active conformation and prolong the half-life of latency conversion.[112, 114]. Berkenpas *et al* designed a relatively low mutation frequency library (~3 mutations per molecule) and expressed the library in λ phage.[114] Screening this library for mutants capable of binding to tPA after prolonged incubation at 37°C resulted in identification of 14 unique clones, including PAI-1 14-1B, which harbors four amino acid substitutions, has a significantly prolonged half-life of 145 hours, and was sufficiently stable for crystallization of the active form.[114, 165] Stroop *et al* constructed a library of PAI-1 mutants displayed on the surface of M13 filamentous phage and performed a similar screen to identify stabilizing mutations.[112] This library had a much higher mutation frequency (~10 mutations per molecule) and a total of 6 clones were identified with increased stability.[112] Together, these two studies analyzed a combined total of $1.3x10^7$ mutant clones, providing more than enough depth to survey the 7,580 possible single amino acid substitutions within PAI-1 (379 amino acids × 20 possible substitutions). However, only a combined 20 clones were sequenced from the selected material, ultimately identifying 84 substitutions that stabilize the active

conformation of PAI-1. As this dataset represents approximately 1.1% of all possible substitutions, the majority of the PAI-1 mutational landscape remains uncharacterized. Here we report the application of high throughput DNA sequencing (HTS) to screen a phage display library of randomly mutagenized PAI-1 in order to characterize the majority of the remaining mutational landscape of PAI-1.

MATERIALS AND METHODS

PAI-1 cloning

For display of PAI-1 on the M13 filamentous bacteriophage, human *SERPINE1* cDNA was cloned into the pAYE phagemid using AscI and NotI restriction enzyme sites (New England BioLabs, Ipswitch MA).[44, 102] The mature PAI-1 protein is N-terminally fused to a myc tag and C-terminally fused with a FLAG tag, E, tag, and a truncated M13 pIII coat protein, in tandem.

PAI-1 mutant phage display library construction

Wildtype (WT) *SERPINE1* in pAYE was randomly mutagenized using the GeneMorph II Random Mutagenesis Kit (Agilent Technologies, Santa Clara, CA) following manufacturer's instructions. Briefly, primers used for PCR mutagenesis (5'-TTCTCGGCCTAGCCGGCCGG -3' and 5'-GTCCTTGTAGTCACCACCACCTGCGGCC -3', Table 3-1) covered the AscI and NotI restriction sites so as to conserve these sites for ligation of the restriction digested

amplicons into pAYE. After PCR, products were seperated on a 1% agarose gel and the band at ~1200 bp was gel extracted, digested with AscI and NotI, ligated into pAYE, and transformed into XL-1 Blue MRF' *E. coli* (Agilent Technologies)[166]. Library depth was determined by titering transformed cells on LB agar (Invitrogen, Carlsbad, CA) supplemented with 100 μg/mL ampicillin and 2% glucose.

Phage production and purification

Phage were prepared as previously reported with slight modification to minimize PAI-1 conversion to the latent conformation.[44] Briefly, *E. coli* harboring pAYE- mutant PAI-1 were grown in LB Broth (Invitrogen) glucose at 37°C to log phase (OD_{600} 0.3-0.4) and infected with M13KO7 helper phage (GE Healthcare, Chicago, IL) at a multiplicity of infectivity of approximately 100, with incubation for 1 hour at 37°C. Cells were pelleted by centrifugation and then resuspended in 2xYT with ampicillin (100 μg/mL) and kanamycin (30 μg/mL; Sigma-Aldrich, St. Louis, MO). To induce PAI-1 production from the phagemid, IPTG was added to 2xYT (0.6 mM). To minimize the time for PAI-1 latency transition though maintaining sufficient phage production, growth time was shortened from 16 hours at 30°C to 2 hours at 37°C. All subsequent phage preparation steps were carried out on ice. The PEG/NaCl phage precipitation step was also shortened from overnight to 30 minutes. After precipitation, the phage pellet was resuspended in sterile TBS containing 0.005% Tween-20 (TBS-T). Phage titer was determined by transducing naïve XL-1 Blue MRF' *E. coli* grown to log phase for 1 hour at 37°C and plating on LB-agar supplemented with 100 μg/mL ampicillin and 2% glucose.

uPA:phage displayed PAI-1 capture assay

Phage displaying the A3 domain of von Willebrand Factor (VWF-A3; cloned into pAYE)[44] were used as a negative control for binding to uPA. Phage displaying PAI-1 were diluted in phage displaying VWF-A3 and then incubated with high molecular weight urokinase (HMW-uPA, Molecular Innovations, Novi, MI; 1.5 nM) for 30 minutes at 37°C. EDTA-free protease inhibitor cocktail (Roche, Basel, Switzerland) was added and incubated for 10 min before immunoprecipitation. Phage:uPA complexes were immunoprecipitated using magnetic protein G beads (New England BioLabs, Ipswitch, MA) charged with polyclonal anti-uPA antibody (Abcam Inc, Cambridge, UK; ab2412). Beads were washed 5x with TBS-BSA and eluted by enterokinase (New England BioLabs; 16U) digestion for 16 hours at 4°C in enterokinase elution buffer (20 mM Tris-HCl, 50 mM NaCl, 2 mM $CaCl_2$, and 5% BSA, pH 8.0). The eluted phage pool was used to infect naïve XL-1 Blue MRF' *E. coli*. Eluted phage were quantified by counting the number of transduced colonies that grew in the presence of ampicillin.

Colony screening

To determine the composition of the phage pools before and after selection, single colonies of resistant bacteria were selected and DNA amplified by PCR using primers annealing outside the insertion site, to a region common to both pAYE-PAI-1 and pAYE-VWF-A3 (5'-TTCATGCTGCCGGCTTTCTCGG -3' and 5'-TCGTCATCGTCCTTGTAGTCACCACCA -3'). Phagemid inserts were distinguished by PCR product size (1271 bp for PAI-1 and 677 for VWF-A3). Mutation frequency of

PAI-1 mutant libraries was calculated by Sanger sequencing of the *SERPINE1* inserts from selected individual colonies.

PAI-1 latency screen

100 μL resuspended phage were added to 800 μL TBS-BSA (5%) and incubated at 37°C for the desired time before adding HMW-uPA (1.5 nM; Molecular Innovations) and incubating for an additional 30 minutes at 37°C. EDTA-free protease inhibitor cocktail (Roche, Basel, Switzerland) was added and incubated for 10 min before immunoprecipitation. PAI-1:uPA complexes were immunoprecipitated, washed, and eluted as described above.

PAI-1 phage display sequencing

For Illumina high throughput sequencing (HTS), 150 bp overlapping amplicons were PCR amplified from pAYE-mutant PAI-1(Supplementary Tables 3-1 and 3-2). All primer pairs used to generate amplicons included a 5'-6N wobble (where N is A, C, G, or T) to provide sequence diversity and aid in base calling on the Illumina platform, resulting in a final PCR product size for each amplicon of 162 bp. PCR amplicon products were gel purified and prepared using NEBNext DNA Library Prep Master Mix Kit, following manufacturer's instructions (New England BioLabs). Illumina sequencing adapters containing barcodes were purchased from Bioo Scientific (Austin, TX). 2x150 bp paired end HTS was performed using the Illumina HiSeq2500 platform (Illumina, San Diego, CA) at the University of Michigan DNA Sequencing Core (Ann Arbor, MI) or MedGenome, Inc (Foster City, CA).

The Illumina HiSeq Platform was chosen for its relatively high accuracy among HTS technologies.[167] Specifically, paired end sequencing was utilized to minimize the possibility of falsely identifying a sequencing error as a mutation in a PAI-1 insert. The error rate of Illumina HTS has been reported at 0.0026 errors per base for read 1 (R1, "forward read") and 0.0040 errors per base for read 2 (R2, "reverse read").[168] Given these data, the probability of a sequencing error occurring at the same position in the forward and reverse reads (and thus being incorrectly classified as a true mutation) are ~1 in 100,000 ($0.0026 \times 0.0040 = 1.04 \times 10^{-5}$). Given current limits for read lengths on this platform, the mutagenized PAI-1 cDNA insert of 1.2 kbp was sequenced in 12 unique amplicons (Supplementary Table 3-2).

Recombinant PAI-1 expression and purification

Human *SERPINE1* cDNA was subcloned from pAYE-PAI-1 into the pET Flag TEV LIC cloning vector (2L-T) using primers pET PAI-1 LIC For and pET PAI-1 LIC Rev (Supplementary Table 3-1), transformed into BL21(DE3) chemically competent *E. coli* (Invitrogen) and grown on LB agar supplemented with 100 μg/mL ampicillin. Resistant colonies were scraped into LB media supplemented with 100 μg/mL ampicillin and 20% glycerol and stored as glycerol stocks at -80°C. Point mutations were introduced using the QuickChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies) following manufacturer's instructions. The full length of WT and all mutant PAI-1 constructs was confirmed by Sanger sequencing. The pET Flag TEV LIC cloning vector (2L-T) was a gift from Scott Gradia (Addgene plasmid #29715).

For recombinant PAI-1 expression and purification, *E. coli* containing pET-PAI-1 constructs were grown at 37°C in LB media (100 mL) supplemented with 100 μg/mL ampicillin to mid-log phase (OD_{600} ~0.40-0.45) at which point IPTG was added (1 mM) and cells were grown for an additional 2 hours at 37°C. Cells were pelleted by centrifugation (3500g, 20 min, 4°C) and washed twice with 0.85M sodium chloride. The pellet was resuspended in suspension buffer (0.05M sodium phosphate, pH 7.0, 0.1M sodium chloride, 4 mM EDTA), T4 lysozyme (1 mg/mL) was added and the reaction incubated at room temperature for 30 min. Cells were subjected to three freeze-thaw cycles (1 min in liquid nitrogen, 5 min in 37°C water bath) and DNA was sheared by passing the crude extract through a 25 gauge needle 5-6 times. Lysates were cleared by centrifugation at 16000g for 20 min at 4°C. PAI-1 was purified from crude cell extracts using anti-FLAG agarose beads (Sigma-Aldrich) and eluted with 3X FLAG peptide (Sigma-Aldrich) following manufacturer's instructions.

PAI-1 activity assay

To determine half-life for the latency transition, recombinant PAI-1 was incubated for varying amounts of time at 37°C and then its residual inhibitory activity against HMW uPA was measured as follows. Purified PAI-1 (7.5 nM) was incubated with HMW uPA (2.5 nM) in PAI-1 activity assay buffer (40 mM Hepes, 100 mM NaCl, 0.005% Tween-20, pH 7.4) at room temperature for 30 min. Chromogenic substrate (zGly-Gly-Arg-AMC, 50μM; Bachem, Bubendorf, Switzerland) was then added and residual uPA

activity was determined by measuring substrate hydrolysis for 10 minutes (λ_{ex} 370 nM, λ_{em} 440 nM).

Half-life calculation

In order to determine half-life for PAI-1 displayed on phage or recombinant PAI-1, the number of phage recovered or relative activity of recombinant protein at a given time were plotted using GraphPad Prism 7 (GraphPad Software, La Jolla, CA). Within the program, nonlinear regression analysis was used to fit a one-phase decay curve to data following the equation:

Equation 1: $$y = (y_o - B)e^{(-Kt)} + B$$

Where y_o is equal to the y value at time (t) zero. As all values are relative to this point, y_o is set to 1. B is the plateau, or the value of y at infinite time. For all half-life calculations, B was set to 0, resulting in the equation:

Equation 2: $$y = e^{(-Kt)}$$

Where K is the rate constant expressed as $^1/_h$. Half-life was determined using the equation:

Equation 3: $$t_{1/2} = \frac{\ln(2)}{K}$$

Massively parallel half-life calculation

To calculate a half-life of a particular clone based on the sequencing data, the frequency of reads mapping to an individual clone relative to all reads at that time point were multiplied by the estimated number of phage selected at that time point to estimate the number of phage displaying a particular clone. This was repeated for all time points and the number of a particular clone relative to the quantity at time 0 was determined, plotted in GraphPad, and a half-life was estimated as described above. As mutations exist on different amplicons, the half-lives were compared to the estimated half-life for phage with a WT sequence in the same amplicon as the mutation.

Statistics

Statistics for HTS were performed using DeSeq2 with adjusted *p*-value (p_{adj}) of <0.01 considered significant.[169] Statistics for activity assays were calculated using GraphPad Prism 7 (GraphPad Software).

RESULTS AND DISCUSSION

PAI-1 displayed on phage maintains its inhibitory function

To determine if PAI-1 displayed on phage maintained its ability to covalently bind to and inhibit uPA, phage displaying WT-PAI-1 were diluted into phage displaying VWF-A3. This pool was screened against uPA as described in "Materials and Methods." The selected phage pool showed marked enrichment for PAI-1 phage over VWF,

demonstrating that phage displayed PAI-1 maintains its inhibitory function with respect to uPA (Figure 3-1).

PAI-1 displayed on phage exhibits latency kinetics similar to native PAI-1.

In order for a phage display library to accurately identify mutations that increase the functional stability of PAI-1, the phage-displayed PAI-1 must undergo a latency transition similar to native PAI-1. To confirm this, the half-life for the latency transition of WT PAI-1 displayed on phage was determined over the course of a 72 hour incubation at 37°C by quantifying the number of functional phage immunoprecipitated in complex with uPA at each time point, relative to time 0 (Figure 3-2). The half-life of phage displayed PAI-1 was calculated to be 2.2 h (95% CI: 1.7 – 2.9h), similar to that of native PAI-1 ($t_{1/2} \sim 2$ h).[105, 106] Importantly, in the absence of strong detergents, the latency transition half-lives of glycosylated PAI-1 produced in human cell lines and non-glycosylated PAI-1 produced in *E. coli* are equivalent.[170]

To exclude the possibility that loss of PAI-1 function was due to non-specific degradation or proteolysis of PAI-1 phage rather than transition to latency, a myc immunoprecipitation with anti-myc was also performed after prolonged incubation. PAI-1 displayed on phage is fused to an N-terminal myc-tag. As the display of the myc tag is not dependent on the conformation of PAI-1, anti-myc antibody should bind equally well to active and latent PAI-1. Proteolysis of PAI-1 on pIII, however, should result in loss of the myc tag. The number phage recovered by myc-immunoprecipitation remained unchanged over the 72 h time course, confirming that the decrease over time in phage

recovered by uPA capture was due to loss of PAI-1 function secondary to latency transition (Figure 3-2).

PAI-1 mutant phage libraries encompass broad coverage of potential single amino acid substitutions

Two independent libraries (Library A and Library B) were constructed for use as biological replicates in HTS screening experiments. The libraries were similar in terms of mutation frequency and depth (Table 3-1). Deep sequencing of the libraries revealed that all possible single nucleotide substitutions (3411/3411) were represented within the libraries and 74% (5613/7580) of all possible amino acid substitutions were represented (Supplementary Figure 3-1). In examining the nucleotide substitutions identified by HTS, there was a bias for transitions (purine (A and G) to purine or pyrimidine (C and T) to pyrimidine mutations) over transversions (purine to pyrimidine, or vice versa) (Table 3-2, Supplementary Table 3-3). Though $^1/_3$ of all possible nucleotide substitutions are transitions, they constituted 65.2% of substitutions within the libraries, consistent with the known bias of the Mutazyme II error prone polymerase (Agilent Technologies, GeneMorph II Random Mutagenesis Kit, Catalog #200550, Instruction Manual). A slight preference of the polymerase to introduce mutations at A-T over G-C base pairs was also observed (Supplementary Table 3-3). A challenge in generating all possible amino acid substitutions using random mutagenesis is the degeneracy of the genetic code with respect to amino acids; two amino acids (methionine and tryptophan) are coded by a single codon, while the remaining eighteen amino acids can all be made using at least two degenerate codons, with stop codons being coded for by three codons. Given this, it was

not surprising the most common substitution not represented in the library were those to tryptophan (10.5% of all missing substitutions). Somewhat surprisingly, substitutions to cysteine (coded by two codons) were more commonly absent (8.3%) than methionine substitutions (8.0%). (Supplementary Table 3-4). A recent bioinformatics approach demonstrated that when using single nucleotide substitution, which is what we expect in our library given our low mutation frequency, only 43.8% of all amino acid substitution events are possible and those that do occur tend to favor substitutions of amino acid with similar properties (polar, non-polar, charged, etc).[171] Given these data, our library depth covers more mutations than would be expected.

HTS confirms previously identified stabilizing and destabilizing mutations in PAI-1

Mutant libraries were incubated for 0, 24, 48, and 72 hours at 37°C and then screened to select for PAI-1 mutants that retained their functional ability to bind uPA. Mutant libraries A and B were kept separate throughout screening and treated as biological replicates for DeSeq analysis. DeSeq was used to detect mutations that were significantly depleted (p_{adj}<0.01, Log2-Fold Change <0) or enriched (p_{adj}<0.01, Log2-Fold Change >0) relative to the unselected library at all time points individually and to all incubation time points (24, 48, 72h) combined (Table 3-3).

Mutations that were significantly depleted were predicted to cause a destabilization of the active conformation of PAI-1. As expected, within the depleted populations, mutations resulting in substitution of a proline or stop codon were most represented (9.4 and 8.0%

of all depleted mutations, respectively, expected proportion assuming equal distribution: 5%) (Supplementary Figure 3-3).

Two previous studies have used libraries of randomly mutagenized PAI-1 to identify mutations that stabilize the active conformation and increase the half-life for the latency transition; combined these data identified a total of 84 amino acid substitutions that enhance PAI-1 stability.[112, 114] The substitutions classified as stabilizing in our analysis included a majority of the previously identified mutations (Table 3-3). When the samples incubated for 24, 48, and 72 hours were compared to the unselected pool independent of one another, the most significantly enriched mutation identified at each time point was I91L, which has been previously reported as the most stabilizing single amino acid substitution (Supplementary Tables 3-5 – 3-7).[114]

A limitation of the present study is the restriction of maximum amplicon size imposed by sequencing read lengths available on the Illumina platform. As a result, the entire length of PAI-1 cDNA insert (1.2 kb) could not be covered in a single sequencing read, instead requiring analysis of 12 independent amplicons (Supplementary Table 3-2). As a consequence, we are unable to determine whether multiple mutations identified in different amplicons occur in cis (together in the same PAI-1 phage clone) or trans (in different PAI-1 phage). Thus, potential nonadditive interactions between substitutions at two different positions (epistasis) can only be determined when both amino acid substitutions are located in the same amplicon. The most stable PAI-1 variant described by Berkenpas *et al*, 14-1B, harbored 4 mutations: N150H, K154T, Q319L, and M354I.[114]

Using recombinant expression, it was demonstrated that N150H alone was destabilizing, causing a reduction in the functional half-life of PAI-1 (1.3 hours) but when expressed in cis with K154T, the double mutant showed a greater increase in functional stability ($t_{1/2}$ = 6.7 hours) than for K154T alone (3.4 hours). [114] The same epistatic interaction could be assessed from our data since positions 150 and 154 are covered by the same sequencing amplicon (Table 3-4). when comparing the sequencing data from the unselected library to the pooled sequencing data from the prolonged incubation samples, no enriched mutations were detected at position 150, but significantly depleted mutations were (Table 3-4). Three double mutations, with one of the mutations occurring at position 150 and the other at position 154 were found to be significantly enriched after prolonged incubation, suggesting an epistatic interaction between the two residues. A lone 150/154 double mutation was found to be significantly depleted, but the mutation at position 154 was a nonsense mutation, introducing a stop codon (Table 3-4).

Because of its role in inhibition of tPA and uPA as well as its involvement in the latency transition, we chose to examine more closely the mutational landscape of the RCL (Figure 3-3). It is worth noting that because PAI-1:uPA complexes are immunoprecipitated with an anti-uPA antibody in our screen, any mutation that prevents the formation of a covalent bond between the serpin and protease (a dead mutant or mutant that exhibits substrate behavior) would be depleted. The RCL residues are thought to drive the specificity of serpins, especially at P1.[83, 172, 173] Consistent with this model, our data found no enrichment, but substantial depletion, for alternative amino acids at P1 (Figure 3-3). These data agree with previous saturation mutagenesis of the P1-P1'

residues, which demonstrated the requirement for an arginine or lysine at the P1 position with little preference at P1' for uPA inhibitory activity.[174] Consistent with Sherman *et al*, our data demonstrated the P1' position to be highly tolerant for substitution mutations, as seen with the lack of either significantly enriched or depleted mutations at the P1' (Figure 3-3).[174] In the latent conformation of PAI-1, residues P14-P4 of the RCL insert into β-sheet A as strand 4 (s4A), and it was theorized that introducing large, inflexible, or charged residues in this region would inhibit the insertion and stabilize the active conformation.[162, 164] As this insertion is critical for the inhibition of target proteases, however, this hypothesis was proven incorrect by several studies which demonstrated that mutation of these residues to charged or less flexible amino acids resulted in cleavage of the RCL without protease inhibition, termed substrate behavior.[162, 163, 175] In 1994 Audenaert *et al* performed a proline-scanning mutagenesis, introducing proline mutations at P12 (A335P), P10 (S337P), P8 (T339P), and P6 (V341P).[162] In all positions except P6, the substitution of a proline converted PAI-1 into a substrate for tPA; proline substitution at P6 resulted in a mutant that behaved similar to WT PAI-1.[162] This result is mirrored exactly in our data: prolines were observed to be depleted at P12, P10 and P8, but not at P6 (Figure 3-3). Shortly thereafter, Lawrence *et al* demonstrated that substitution of a charged arginine for threonine at position 333 (P14) destroyed all inhibitory activity against tPA and uPA by converting the mutant PAI-1 to a substrate for the proteases.[163]In our mutation catalog, arginine was observed to be the most significantly depleted mutation at that position (Figure 3-3).

Finally, residues C-terminal to the scissile bond (P2'-P4', amino acid positions 348-350) are more tolerant of mutations than the N-terminal segment, consistent with the requirement for insertion of this segment into β-sheet A during the transition to latency (Figure 3-3).[107]

Characterization of stabilizing mutations not previously identified

Three mutations not previously identified as stabilizing, I91F, K176E, and Y221C, were found in our enriched data set and selected for further analysis. All three mutations were among the top 25 most significantly enriched mutations at each time point (Supplementary Tables 3-5 – 3-7). I91F is similar to the previously identified I91L. K176E substitutes a negatively charged glutamic acid for a positively charged lysine. Y221C was the most significantly enriched substitution involving a cysteine substitution. FLAG-labeled rPAI-1 mutants were expressed, purified, and their half-lives measured over a 24 h time course, as described in "Materials and Methods" (Figure 3-4). rPAI-1 I91F and K176E had significantly prolonged half-lives ($p<0.05$) compared to WT rPAI-1 at the 95% confidence interval. The half-life of Y221C rPAI-1 was also prolonged but the difference was not significant (Table 3-5, Figure 3-4). It is possible that PAI-1 Y221C did not show a significantly prolonged half-life for the latency conversion because it required epistatic interactions with a mutation present on another amplicon. As discussed above, we are unable to determine if multiple mutations occurring on different amplicons occur in cis or trans and it is possible that Y221C in conjunction with another, distant, substitution is required for significant increase in structural stability. In such an instance, Y221C would be enriched due to stabilization of a molecule harboring the mutation along

with others, but we would be unable to determine which mutations must occur together if they are found on a different amplicon.

Catalog of mutations that significantly affect PAI-1 stability

By comparing the pooled sequencing data from the prolonged incubation time points with the data from the unselected library, we were able to generate a catalog of all mutations predicted to significantly stabilize or destabilize the active conformation of PAI-1. (Figure 3-5). Two-thousand and one of the 7,580 possible single amino acid substitutions (379 residues × 20 possible substitutions, including stop codons) exhibited statistically significant changes in stability including 492 enriched and 1,509 depleted. A computer algorithm developed in 2012, Sorting Intolerant From Tolerant (SIFT), predicts whether an amino acid substitution will affect protein function.[176] SIFT analysis of PAI-1 predicted that 4,194 of 7,201 (premature stop codons are not included in SIFT analysis) single amino acids substitutions would be deleterious to PAI-1 protein function (Supplementary Figure 3-4). Given this, the relatively high number of depleted mutations, predicted to destabilize the active conformation of PAI-1, identified in the present screen is not unexpected. The large discrepancy in the number of predicted deleterious mutations between our data and SIFT suggest high-throughput functional protein screens are of higher value than computer algorithms with respect to determined the effect of single amino acid substitutions on protein function.

Massively parallel half-life determination

Our screen identified over 400 novel mutations that we predict to stabilize the active conformation of PAI-1 and slow the transition to latency. To determine the half-life for the full set of substitutions, we wanted to develop a high-throughput approach using sequencing read numbers and phage quantification at different time points. To test this, we calculated a predicted functional half-life for the recombinantly generated clone that showed significant prolongation of half-life, I91F and K176E, and the previously described most stabile point mutant, I91L, using a massively parallel approach as proof of principle (Figure 3-6). Half-lives estimated using this approached trended as expected, with I91L showing the most prolonged half-life, followed by I91F, and K176E. The calculated half-lives did not, however, agree with half-lives determined for the mutants by recombinant protein expression and activity assay (Table 3-6). We believe earlier time points are required in order to more accurately estimate the half-life of an individual mutant using this high throughput approach. Another limitation to this approach is that massively parallel half-life determination is likely complicated by epistasis within the library, as discussed above.

Conclusions

Our data show a high degree of agreement with previously published studies using random library of PAI-1 mutants to identify mutations that increase the functional half-life. We also arrived at the same conclusions of several studies using targeted mutagenesis of the RCL. Taken together, these data demonstrate the power and validity of our high throughput phage display approach, especially to perform comprehensive analysis on over 5,600 mutations and their effect on the functional stability of PAI-1.

Using phage display coupled to HTS, we were able to perform these analyses in a massively parallel fashion and ultimately characterized 2,001 mutations as significantly affecting the functional half-life of PAI-1. This was not possible in previous studies using phage display to examine random mutations of PAI-1 and their effect on functional stability.[112]

Currently, our ability to fully characterize mutations that stabilize the active conformation of PAI-1 is limited by the available sequencing technology. Due to DNA read size restrictions, we are unable to examine the role of multiple mutations occurring in tandem, hindering our ability to characterize epistatic interactions between two or more mutations acting synergistically to increase the half-life of PAI-1. This limits our ability to harness the full potential of this high throughput phage display approach and perform massively parallel latency transition half-life estimations and other possible assays such as massively parallel kinetics of uPA inhibition by PAI-1 mutants within the library. Other HTS platforms, such as Pacific Biosciences Single Molecule Real Time (SMRT) sequencing (https://www.pacb.com/products-and-services/pacbio-systems/rsii/) offer sequencing reads of far greater length that would allow sequencing of the entirety of a PAI-1 insert in the phagemid, but the error rate of this technology is in excess of 10%.[167] Such a high error rate would prove a substantial hurdle in examining a mutant library as it would be difficult to differentiate between a true mutation and a sequencing error..

Another limitation with our libraries described herein is the lack of complete coverage of all possible substitution mutations. Although all possible single nucleotide substitutions

were present within the library, 26% (2,167/7,580) of all possible amino acid mutations were not observed. Synthetically constructed DNA oligonucleotides have been described as a viable alterative approach to generating mutant libraries and removes the biases associated with error prone DNA-polymerases.[177] This approach, however, has not been applied to large DNA fragments (such as PAI-1) and may prove too expensive to be seen as a viable approach for mutating whole genes.

Finally, we believe our methodology and results presented here have potential translational applications. In 2010 a "very long half-life" PAI-1 variant was shown to reduce total blood loss and bleeding time in mouse tail clip assays.[178] This particular PAI-1 variant had a half-life approaching 30 days, but required to formation of a disulfide bond between two cysteine substitution mutations and could not be produced in bacteria.[179] Identification of a highly stable PAI-1 variant from our library that does not require the formation of a disulfide bond and could be recombinantly produced in bacteria might hold similar therapeutic applications. Additionally, our comprehensive profiling of the role of single amino acids substitutions on PAI-1 stability could be applied to help better understand the clinical significance of so-called "variants of unknown significance" (VUS) that are routinely identified in patients after genetic testing. By applying our approach to other proteins, this comprehensive profiling approach could help better understand if VUSs are in fact significant, and what role they might play in protein function. Additionally, our mutant library can be screened against other serine proteases to identify potential novel inhibitors or serpins with altered specificity. An obvious target is neutrophil elastase, which is normally inhibited by the

serpin α1-antitrypsin (A1AT).[63] Deficiency of A1AT results in early-onset emphysema due to unregulated elastin breakdown in pulmonary tissue.[78] Currently, treatment for A1AT replacement therapy involves weekly infusions of purified protein from plasma (https://www.nhlbi.nih.gov/health-topics/alpha-1-antitrypsin-deficiency). As PAI-1 can be made in large quantities recombinantly in bacteria, a novel variant of PAI-1 with altered specificity to inhibit neutrophil elastase could prove a valuable therapeutic in treating A1AT-deficieny. This treatment could be hindered, however, by the latency conversion of PAI-1. Alternatively, our screen could be used to describe a "hyper-stable" backbone of PAI-1, which could then be further mutated and tested for altered specificity.

By developing a high throughput system with which to examine the role of individual amino acid substitutions on functional stability of PAI-1 we have been able to profile the largely-unknown mutational landscape. Additionally, we have laid the groundwork for future studies to perform massively parallel functional or enzyme kinetic assays on PAI-1, as well as other studies which could have great therapeutic impact, such as rapid characterization of VUSs or novel therapeutics evolved from mutagenized PAI-1.

Table 3-1: PAI-1 phage display mutation library characteristics

Library	Mutation Frequency (mutations / molecule)	Library Depth (clones)
PAI-1 Random Mutation A	4.2 ± 1.8	8.04×10^{6}
PAI-1 Random Mutation B	2.8 ± 1.8	4.20×10^{6}

Table 3-2: Nucleotide mutation rates in mutant libraries

Raw counts of mutations within pooled libraries					
		WT Nucleotide			
		A	C	G	T
Mutant Nucleotide	A	0	2121915	30542305	14167510
	C	6666641	0	1849651	18764140
	G	16202283	1564727	0	7517170
	T	10147583	19535798	1309239	0
Frequency of mutations within pooled libraries					
		WT Nucleotide			
		A	C	G	T
Mutant Nucleotide	A		0.016	0.234	0.109
	C	0.051		0.014	0.144
	G	0.124	0.012		0.058
	T	0.078	0.150	0.010	

Table 3-3: Number of significantly depleted and enriched mutations identified and agreement with previous reports

	Significantly Depleted Mutations	Significantly enriched mutations		
		Total	Number of Previously Identified (exact)	Number of Previously Identified (similar)
Input vs 0 h	1103	429	57	69
Input vs 24 h	1351	394	54	66
Input vs 48 h	1556	517	57	69
Input vs 72 h	1421	388	53	65
Input vs Prolonged Incubation	1509	492	49	65

Table 3-4: Epistatic interactions between N150 and K154	
Mutation	Log2 fold change relative to input
Single Mutations	
N150I	-1.47
N150Y	-1.45
N150K	-0.84
N150T	-0.62
N150S	-0.48
K154L	2.77
K154I	2.73
K154V	2.37
K154T	1.40
K154X	-4.13
Double Mutations	
N150D / K154I	2.77
N150S / K154I	2.74
N150D / K154E	1.27
N150D / K154X	-3.07

Table 3-5: Half-lives of rPAI-1 mutants

rPAI-1 Mutant	Mean $t_{1/2}$ (h)	95% CI $t_{1/2}$ (h)	Significant to $p < 0.05$
WT	2.2	1.9 – 2.5	-
I91F	5.4	4.4 – 6.7	Yes
K176E	4.8	2.9 – 7.9	Yes
Y221C	3.0	2.0 – 4.5	No

Table 3-6: Latency transition half-lives estimated by a massively parallel approach

Mutant	Half-life estimated using massively parallel approach Mean (95% CI)	Reported half-life Mean (95% CI)	Ref.
WT PAI-1 (phage displayed) (0, 2, 4, 8, 12, 24, 48, 72 h time points)	N/A	2.2 h (1.9 – 2.5 h)	This Report (Figure 3-2)
WT PAI-1 (phage displayed) (0, 24, 48, 72 h time points)	N/A	5.7 h (3.8 – 7.2 h)	This Report (Figure 3-3)
Amplicon 4			
WT	5.7 h (5.5 – 6.0 h)	N/A	N/A
I91L	11.8 h (10.1 – 13.6 h)	18.4 ± 1.07 h	Berkenpas *et al*[114]
I91F	10.1 h (9.0 – 11.3 h)	5.4 h (4.4 – 6.7 h)	This Report
Amplicon 6			
WT	5.7 h (5.5 – 6.0 h)	N/A	N/A
K176E	9.2 (8.3 – 10.1 h)	4.8 h (2.9 – 6.7 h)	This Report
Amplicon 7			
WT	5.8 h (5.5 – 6.0 h)	N/A	N/A
K176E	9.4 h (8.5 – 10.3 h)	4.8 h (2.9 – 6.7 h)	This Report
Amplicon 8			
WT	5.7 h (5.5 – 5.9 h)	N/A	N/A
Y210X	6.9 h (6.3 – 7.5 h)	N/A	N/A
Y221C	12.8 (10.7 – 15.0 h)	3.0 (2.0 – 4.5 h)	This Report

Supplementary Table 3-1: Primer sequences

Primer	5' ➔ 3' Sequence
pAYE-PAI-1 Cloning Primers	
pAYE-PAI-1 For	GCCGGCGCGCCAGAACAAAAACTCATCTCAGAAGAGGATC TGGGTGGAGGTTCAGTGCACCATCCCCCATCCTAC
pAYE PAI-1 Rev	ACCTGCGGCCGCTGAACCACCTCCGGGTTCTATCACTTGGCCCAT
pAYE-PAI-1 Mutagenesis Primers	
PAI-1 Mutagenesis For	TTCTCGGCCTAGCCGGCCGG
PAI-1 Mutagenesis Rev	GTCCTTGTAGTCACCACCACCTGCGGCC
Sanger Sequencing/Colony PCR Primers	
pAYE Sequencing For	TTCATGCTGCCGGCTTTCTCGG
pAYE Sequencing Rev	TCGTCATCGTCCTTGTAGTCACCACCA
HTS Sequencing Primers	
PAI-1 Amplicon 1 For	nnnnnnGAACAAAAACTCATCTCAGAAGAGGATCTG
PAI-1 Amplicon 1 Rev	nnnnnnGGGTGAGAAAACCACGTTGCG
PAI-1 Amplicon 2 For	nnnnnnTTTCAGCAGGTGGCGCAG
PAI-1 Amplicon 2 Rev	nnnnnnCTTGTCATCAATCTTGAATCCCATAGCTGC
PAI-1 Amplicon 3 For	nnnnnnATGCTCCAGCTGACAACAGGA
PAI-1 Amplicon 3 Rev	nnnnnnTGTGGTGCTGATCTCATCCTTGTT
PAI-1 Amplicon 4 For	nnnnnnGCCCTCCGGCATCTGTACAAG
PAI-1 Amplicon 4 Rev	nnnnnnTTGCTTGACCGTGCTCCGG
PAI-1 Amplicon 5 For	nnnnnnAAGCTGGTCCAGGGCTTCATG
PAI-1 Amplicon 5 Rev	nnnnnnCCCAAGCAAGTTGCTGATCATACC
PAI-1 Amplicon 6	nnnnnnATTCATCATCAATGACTGGGTGAAGAC

For	
PAI-1 Amplicon 6 Rev	nnnnnnCTGGAGTCGGGGAAGGGAG
PAI-1 Amplicon 7 For	nnnnnnAATGCCCTCTACTTCAACGGC
PAI-1 Amplicon 7 Rev	nnnnnnGGGCGTGGTGAACTCAGTATAG
PAI-1 Amplicon 8 For	nnnnnnCCCATGATGGCTCAGACCAAC
PAI-1 Amplicon 8 Rev	nnnnnnGGCAGAGAGAGGCACCTCT
PAI-1 Amplicon 9 For	nnnnnnAGCATGTTCATTGCTGCCCC
PAI-1 Amplicon 9 Rev	nnnnnnTTCAGTCTCCAGGGAGAACTTGG
PAI-1 Amplicon 10 For	nnnnnnCAGGCTGCCCCGCCT
PAI-1 Amplicon 10 Rev	nnnnnnACGTGGAGAGGCTCTTGGTC
PAI-1 Amplicon 11 For	nnnnnnAGACAGTTTCAGGCTGACTTCAC
PAI-1 Amplicon 11 Rev	nnnnnnGGGGGCCATGCGGGC
PAI-1 Amplicon 12 For	nnnnnnTCATCCACAGCTGTCATA
PAI-1 Amplicon 12 Rev	nnnnnnTGCGGCCGCTGAACCACCTCC
pET-PAI-1 LIC Primers	
pET PAI-1 LIC For	TACTTCCAATCCAATGCAGTGCACCATCCCCCATCCTAC
pET PAI-1 LIC Rev	TTATCCACTTCCAATGTTATTATCAGGGTTCCATCACTTGGCCCA
pET PAI-1 Site Directed Mutagenesis Primers	
PAI-1 I91F SDM For	GCTCATGGGGCCATtGAACAAGGATGAGATCAGCACCACAGACGC
PAI-1 I91F SDM Rev	ATCGCGTCTGTGGTGCTGATCTCATCCTTGTTCaATGGCCCCATG
PAI-1 K176E SDM For	GCCCTCTACTTCAACGGCCAGTGGgAGACTCCCTTCCCCGACTCC
PAI-1 K176E SDM Rev	GGAGTCGGGGAAGGGAGTCTcCCACTGGCCGTTGAAGTAGAGGGC
PAI-1 Y221C	CCCGATGGCCATTACTgCGACATCCTGGAACTGCCCTACCACGGG

SDM For	
PAI-1 Y221C SDM Rev	GGTAGGGCAGTTCCAGGATGTCGcAGTAATGGCCATCGGGCGTGG

Supplementary Table 3-2: PAI-1 residues covered by amplicons

PAI-1 Amplicon	Residues Covered
1	V1 – D29
2	A26 – A60
3	G52 – W86
4	E81 – L116
5	P111 – K145
6	H143 – K176
7	Q174 – F208
8	K207 – E242
9	Y241 – L275
10	L273 – S310
11	S308 – S344
12	V343 – P379

Supplementary Table 3-3: Mutational bias of error prone polymerase

	Number	Proportion of Total
Transitions	85044526	0.652
Transversions	45344436	0.348
A➔N, T➔N	73465327	0.563
G➔N, C➔N	56923635	0.437

Supplementary Table 3-4: Substitutions absent from PAI-1 mutant library

Substitution Mutation	Total Number Missing From Library	Percentage of all missing substitutions
D	115	5.8
E	137	7.0
K	118	6.0
R	46	2.3
H	134	6.8
S	12	0.6
T	23	1.2
C	164	8.3
Y	110	5.6
N	83	4.2
Q	149	7.6
G	77	3.9
A	46	2.3
V	22	1.1
L	24	1.2
I	48	2.4
M	158	8.0
F	110	5.6
W	207	10.5
P	83	4.2
X (Stop)	101	5.1

Supplementary Table 3-5: Enrichment after 24 hour incubation

Amplicon	Mutation	Log2 Fold Change
4	**I91L**	2.96
4	*I91F*	2.95
12	*F372V*	2.84
2	**M45K**	2.66
4	*I91M*	2.46
5	*K145N*	2.38
7	K176E	2.35
9	*R271L*	2.27
8	*D222Y*	2.25
6	*K154I*	2.20
12	**F372I**	2.20
9	L272R	2.18
6	*K145N*	2.18
12	E378D	2.15
6	K176E	2.11
10	D311V	2.08
8	Y221C	2.07
11	**S331G**	2.06
3	W86L	2.06
12	**M354I**	2.04
6	L151M	2.03
9	Y241H	2.01
9	**R271C**	1.96
6	*M147K*	1.91
12	*F372L*	1.90

25 most enriched mutations after 24 hour incubation at 37°C compared to unselected library. Previously identified mutations are in bold (exact) or italics (similar mutation at position). Novel mutations are listed in red.

Supplementary Table 3-6: Enrichment after 48 hour incubation		
Amplicon	Mutation	Log2 Fold Change
4	**I91L**	3.86
4	*I91F*	3.73
12	*F372V*	3.64
4	*I91M*	3.54
2	**M45K**	3.46
6	*K154I*	3.16
5	*K145N*	3.15
7	K176E	3.12
9	*R271L*	2.97
12	*F372I*	2.96
12	*F372C*	2.90
8	*D222Y*	2.88
11	**S331G**	2.86
6	K176E	2.84
11	G332S	2.78
12	E378D	2.75
6	*L169H*	2.75
11	G332D	2.69
3	W86L	2.66
6	*K145N*	2.65
8	Y221C	2.63
11	*T333A*	2.62
9	Y241H	2.60
9	**R271C**	2.57
4	W86L	2.53

25 most enriched mutations after 48 hour incubation at 37°C compared to unselected library. Previously identified mutations are in bold (exact) or italics (similar mutation at position). Novel mutations are listed in red.

Supplementary Table 3-7: Enrichment after 72 hour incubation

Amplicon	Mutation	Log2 Fold Change
4	**I91L**	4.19
12	*F372V*	3.97
4	*I91M*	3.87
4	*I91F*	3.86
2	**M45K**	3.78
12	*F372C*	3.26
12	*F372I*	3.19
9	*R271L*	3.19
6	*K154I*	3.07
5	*K145N*	3.03
8	*D222H*	2.97
8	*D222Y*	2.93
8	Y221C	2.93
7	K176E	2.81
2	S41W	2.80
6	*L169H*	2.73
9	**R271C**	2.73
9	L272R	2.72
12	*F372L*	2.72
12	E378D	2.68
11	G332D	2.65
6	K176E	2.62
11	G332S	2.59
8	*D222V*	2.55
9	*Y241H*	2.52

25 most enriched mutations after 72 hour incubation at 37°C compared to unselected library. Previously identified mutations are in bold (exact) or italics (similar mutation at position). Novel mutations are listed in red.

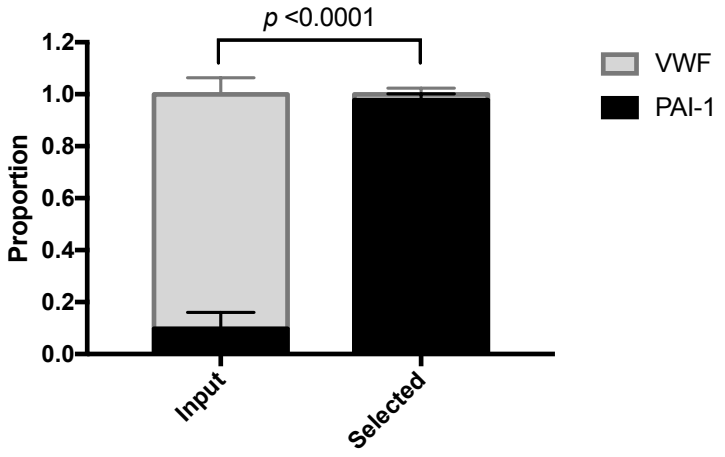


Figure 3-1: Phage-displayed PAI-1 binds HMW-uPA

Phage displaying WT PAI-1 were diluted into phage displaying the A3 domain of VWF at a ratio of 1:9. By analyzing single colonies with primers common to both the PAI-1 and VWF phagemid, the mean proportion of PAI-1 within the unselected (Input) library was 0.097. After selection, the mean proportion of PAI-1 was increased to 0.972. Mean proportion of phage ± SD is displayed, N=3, 96 colonies selected for each replicate (48 for Input pool, 48 for selected pool). $p<0.0001$, compared using two-way ANOVA.

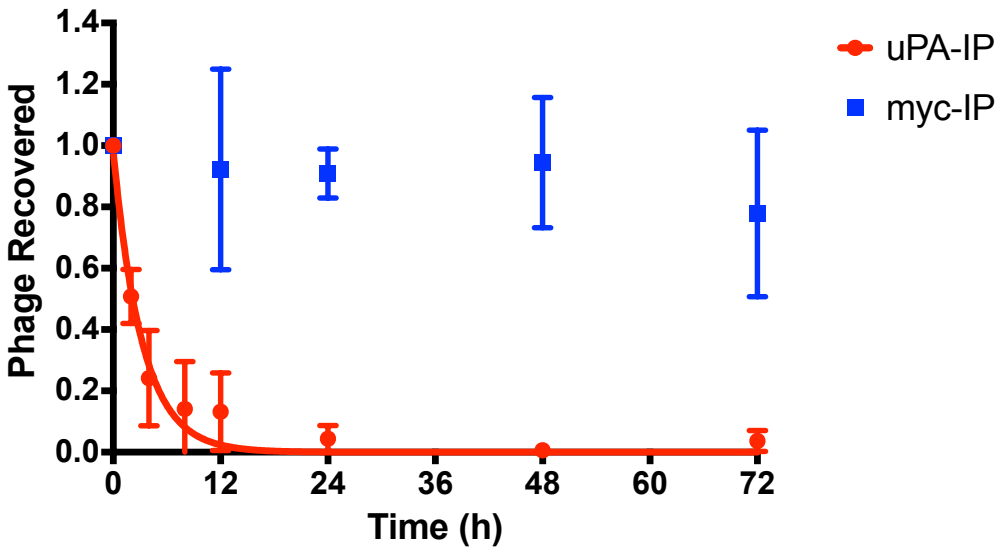


Figure 3-2: Latency transition of PAI-1 displayed on phage

Relative quantity of phage recovered by either uPA complex formation and anti-uPA immunoprecipitation (red) or myc immunoprecipitation (blue) over 72 hours. PAI-1 displayed on phage experiences a latency transition with a half-life of 2.2 hours (95% CI: 1.7 – 2.9h), similar to that of native PAI-1 ($t_{1/2}$ ~ 2h). Myc-immunoprecipitation (blue) shows no significant decrease in number of phage recovered after 72 hour incubation. Mean phage ± SD relative to amount immunoprecipitated at time 0 is plotted. N=3 per group for all time points.

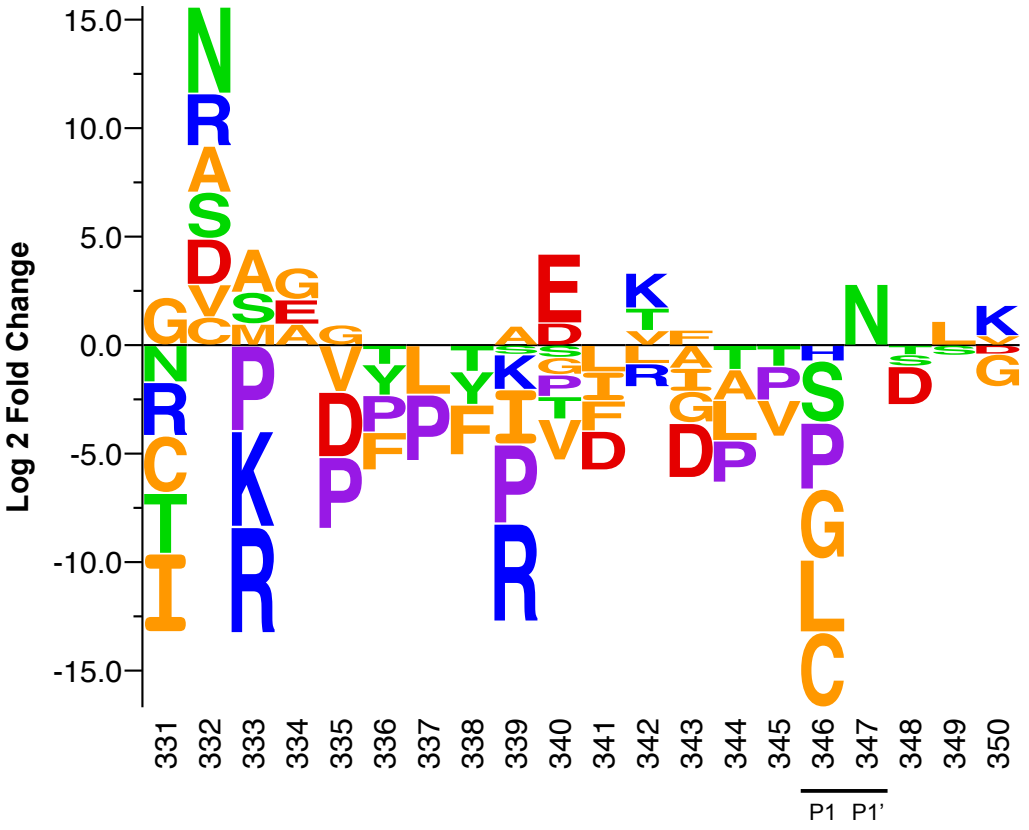


Figure 3-3: Mutational landscape of PAI-1 RCL

Log2 fold change of mutations within the RCL of PAI-1 found to be significantly (p_{adj}<0.01) enriched or depleted in the prolonged incubation subset (24, 48, 72 hours) compared to unselected library. Amino acids above the x-axis are enriched (Log2 fold change >0) while amino acids below the x-axis are depleted (Log2 fold change < 0). Height of letter corresponds to Log2 fold change. Residues are color coded by properties: acidic (negatively charged) residues in red, basic (positively charged) residues in blue, polar amino acids in green, non-polar amino acids in orange, and proline in purple. Logo was generated using SeqLogo2.0: (http://www.cbs.dtu.dk/biotools/Seq2Logo/).[180]

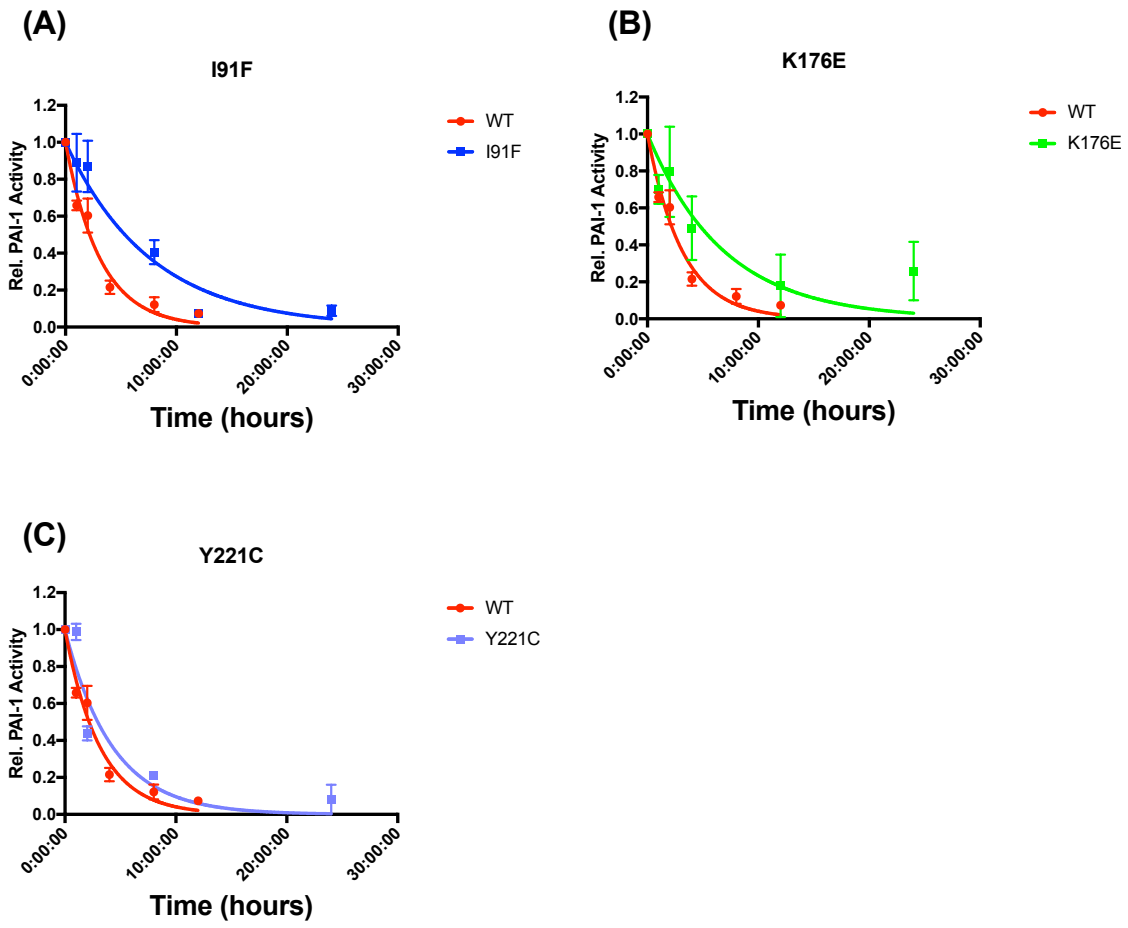


Figure 3-4: Half-life of recombinant PAI-1 mutants

Recombinant PAI-1 mutants were expressed, purified and half-lives were determined as described in "Materials and Methods." I91F and K176E demonstrated a significantly prolonged half-life compared to WT PAI-1. Y221C demonstrated a prolonged half-life but the difference was not significant (Table 3-5). Mean relative PAI-1 activity ± SD is plotted. Significance was determined using the 95% confidence interval for half-life.

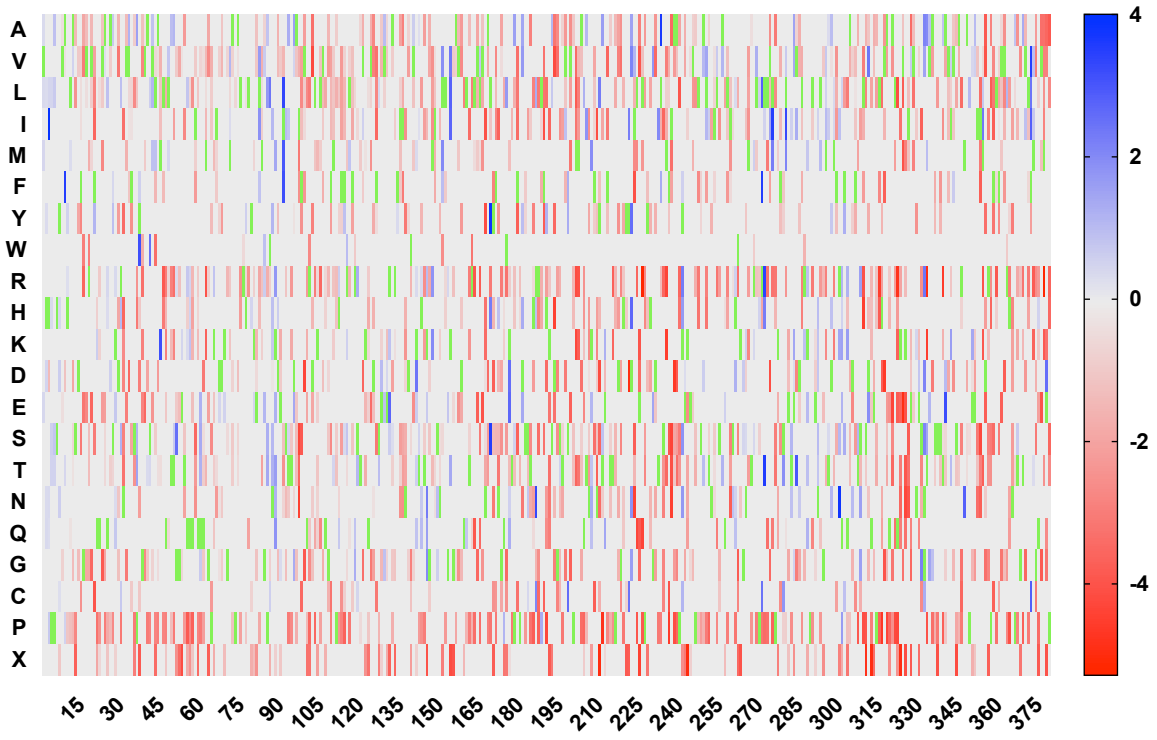


Figure 3-5: Catalog of the effect of all single amino acid substitutions on the functional stability of PAI-1

Log2 fold change of all mutations found to be either significantly depleted or enriched when comparing the prolonged incubation group to the unselected pool. Map in displays the level of enrichment (blue) or depletion (red) of the individual mutations that were altered significantly from the unselected library. Green represents the WT amino acid at each position. Mutations in light gray were either not present in the starting library or were not significantly enriched nor depleted after selection.

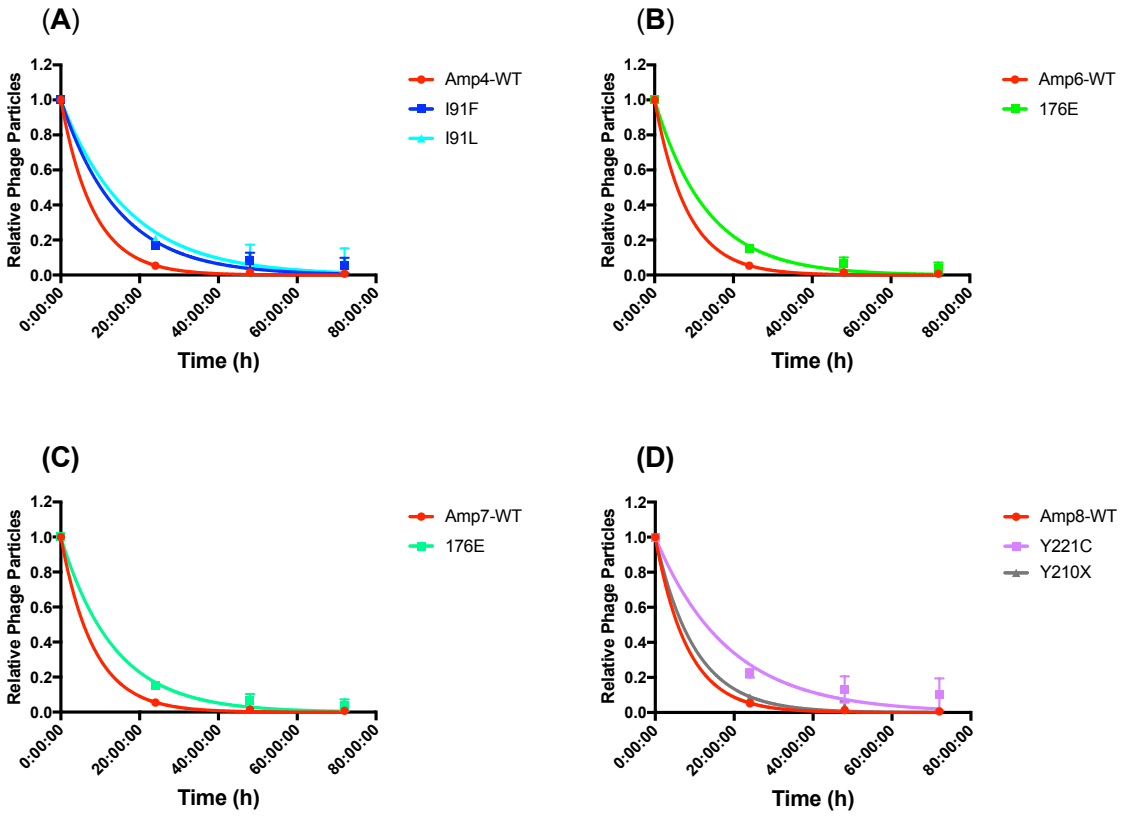
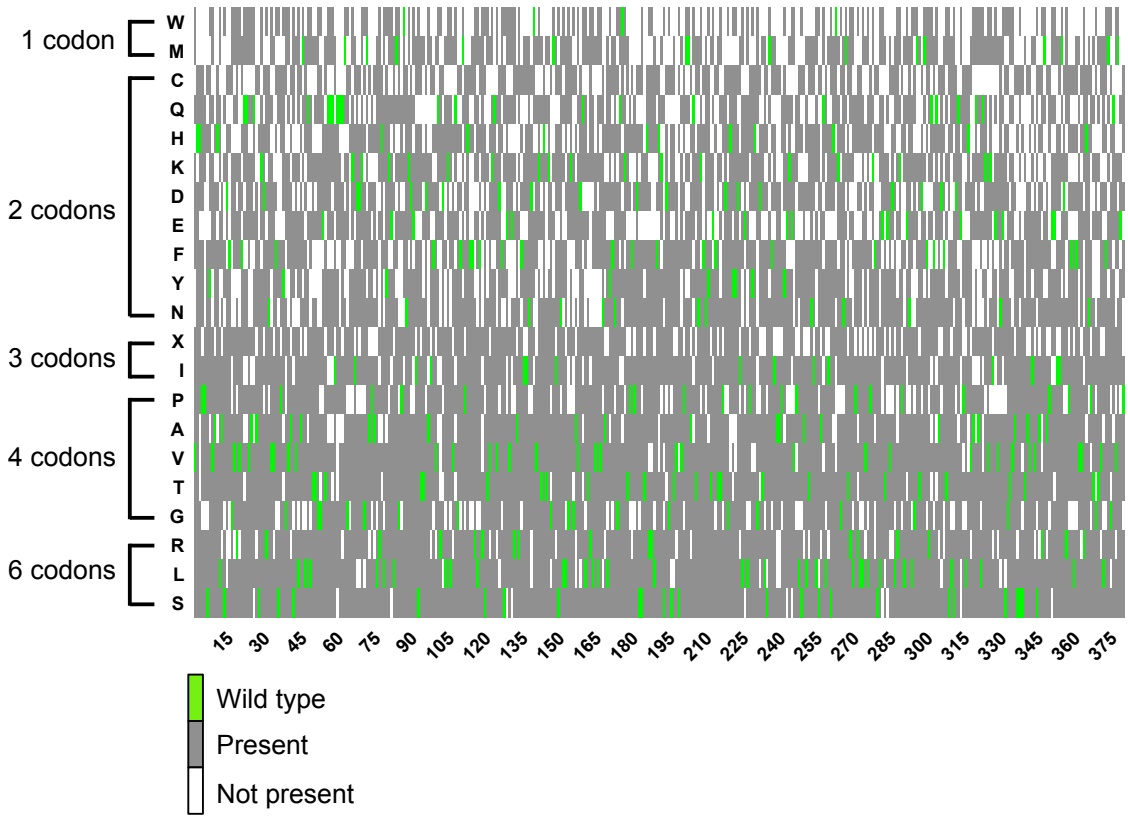


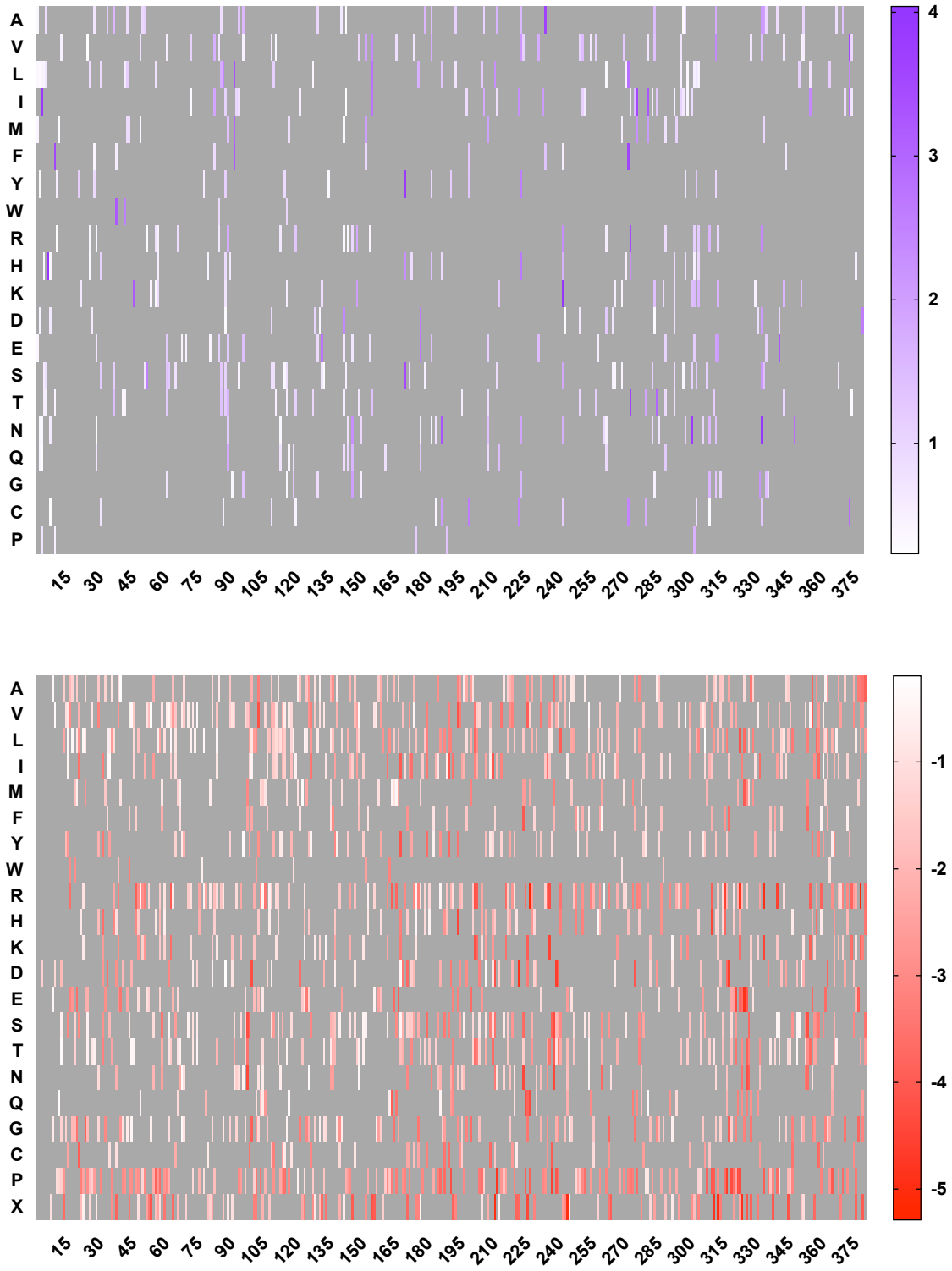
Figure 3-6: Massively parallel latency transition half-life estimation

Estimation of the latency transition half-life using a massively parallel approach for mutations occurring in amplicon 4 (**A**), amplicon 6 (**B**), amplicon 7 (**C**), and amplicon 8 (**D**). All mutations with the exception of Y210X were estimated to have a significantly prolonged half-life relative to the estimate of half-life for the corresponding WT amplicon (Table 3-6). Mean relative phage ± SD relative to amount at time 0 is plotted. N=5 per group for all time points.



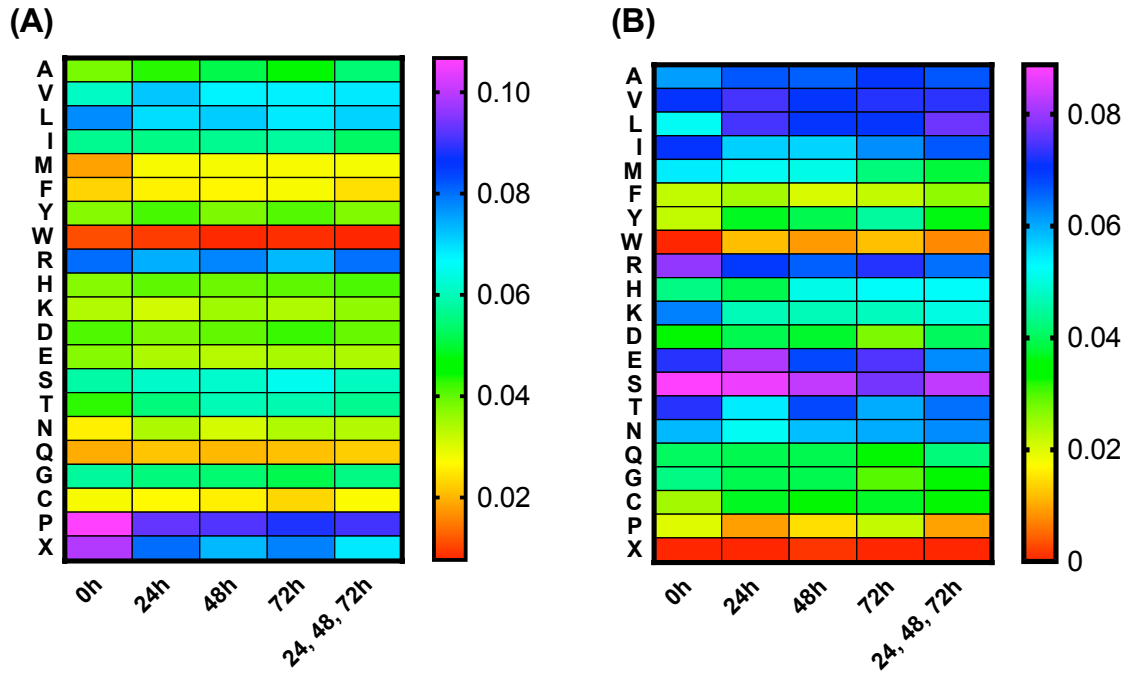
Supplementary Figure 3-1: Mutations present within randomly mutagenized libraries

HTS of Libraries A and B was used to determine the coverage of all possible amino acid substitutions within the unselected library. Mutations observed after sequencing are shown in gray, the WT residue is shown in green, and mutations not found in either library are shown in white. 74% (5613/7530) of all possible amino acid substitutions were present within the libraries.



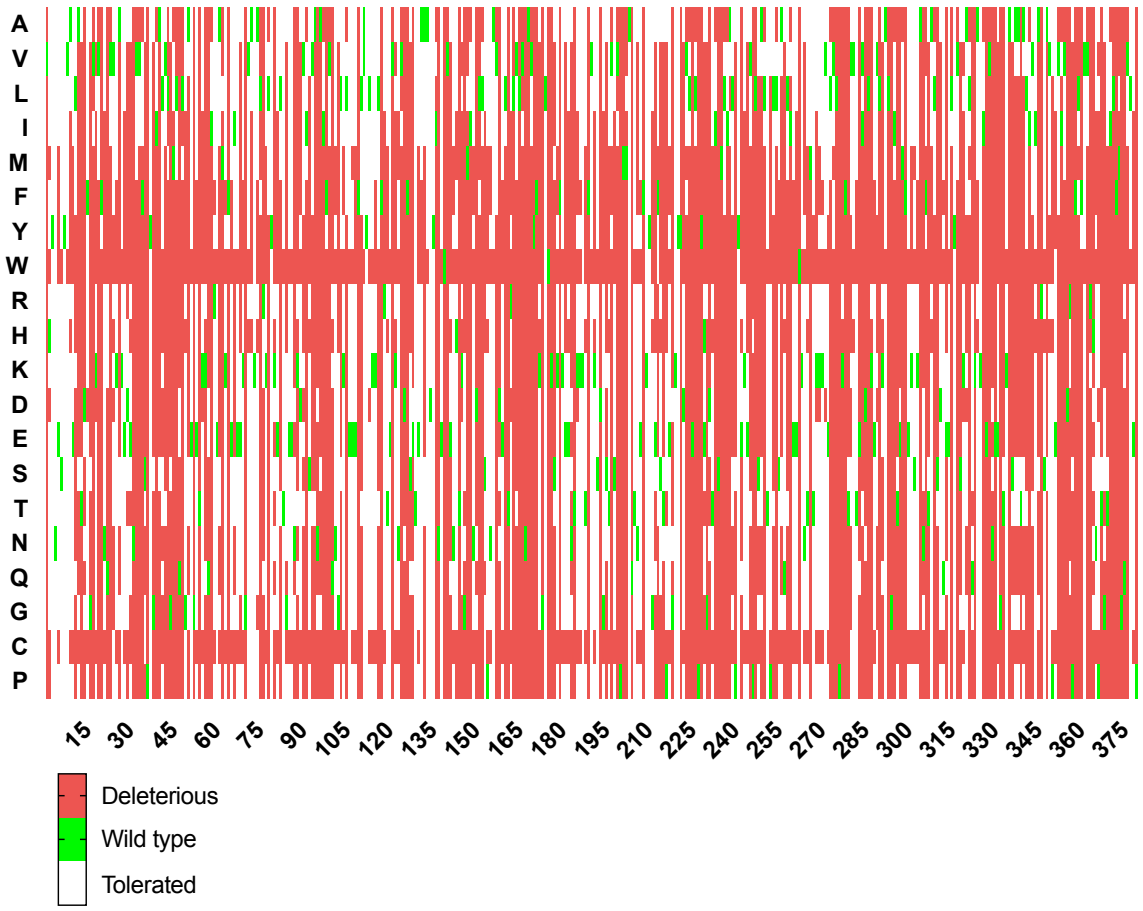
Supplementary Figure 3-2: Significantly enriched and depleted mutations within PAI-1 after prolonged incubation

(**A**) All significantly enriched mutations (p_{adj}<0.01, Log2 Fold Change > 0) found within the PAI-1 mutant library when comparing samples subjected to prolonged incubation with the unselected pool. White indicates lower average Log2 fold change, purple indicates higher average Log2 fold change, and gray indicates mutation was not significantly enriched. Mutations to stop codons were purposefully omitted from graph as no stop codons were found to be enriched, as was to be expected. (**B**) All significantly depleted mutations (p_{adj}<0.01, Log2 Fold Change < 0) found within the PAI-1 mutant library when comparing samples subjected to prolonged incubation with the unselected pool. White indicates higher (closer to 0, less negative) average Log2 fold change, red indicates lower average Log2 fold change, and gray indicates mutation was not significantly depleted. Data was combined at categorized broadly into stabilizing (Log2 fold change > 0) or destabilizing (Log2 fold change < 0) and displayed in Figure 3-6.



Supplementary Figure 3-3: Frequencies of mutations found within enriched and depleted pools

Frequencies of substitution for each amino acid and stop codon found within the significantly depleted (**A**) or enriched (**B**) pool of mutations relative to unselected material after incubation for 0, 24, 48, 72 hours or the pooled prolonged incubation pool (24, 48, 72 hours). Mutations to stop codons and prolines are the most frequent mutations found within the depleted populations. In contrast to the significantly depleted mutations in **(A)**, no stop codons are enriched for after prolonged incubation of PAI-1.



Supplementary Figure 3-4: Mutations predicted to be deleterious by SIFT

Sorting Intolerant From Tolerant (SIFT)[176] analysis predicted 4,194 (not including stop codon substitutions) possible amino acid substitutions to be deleterious to PAI-1 function.

CHAPTER 4:
CONCLUSIONS AND FUTURE DIRECTIONS

CONCLUSIONS

Chapter 2 describes hypothesis-driven experiments aimed at deciphering the mechanism of action by which CCG-2979, a small bioactive compound, inhibits the expression of streptokinase (SK) in Group A *Streptococcus* (GAS). Based on previously published microarray data and identification of two possible binding sites in the SK promoter, the transcriptional regulator CodY seemed like a potential target of the compound. When CodY was knocked out in GAS, however, treatment with CCG-2979 still resulted in decreased SK activity, suggesting that CodY was not directly involved in CCG-2979-mediated inhibition of SK expression or that it was not the sole target. Given this result, an unbiased screen using a GAS phage display library was designed and undertaken in order to identify GAS proteins that interact with the compound, described in the second portion of Chapter 2. This phage display library was unable to be validated, suggesting it would not prove a useful tool in characterizing the GAS-interactome with CCG-2979. It was subsequently discovered that the GAS phage display library contained a high degree of contamination resulting from heterodimeric end adapters. The high frequency at which this specific insert was observed within the pool of phagemids from the GAS phage display library suggests the depth of our library was far lower than expected. Despite our failure to validate the GAS phage display library, we believe phage display, especially

when coupled with high throughput sequencing (HTS), is a powerful technique for studying protein structure-function, as shown by our results in Chapter 3.

The experiments and data in Chapter 3 more accurately represent the power of high throughput phage display. By screening two randomly mutagenized PAI-1 phage display libraries, we were able to comprehensively profile the largely-unknown mutational landscape of PAI-1 with respect to the latency transition half-life in a massively parallel fashion. When selected against uPA after prolonged incubation to promote that latency transition, 2,001 single amino acid substitutions (26.4% of all possible substitutions) were found to be significantly altered relative to the unselected library. 1,509 of these mutations were significantly depleted, including stop codons, mutations within the RCL known to convert PAI-1 from an inhibitor to a substrate of uPA[162, 163, 175], and mutations at the P1 site that would be predicted to diminish or destroy the ability of PAI-1 to inhibit uPA.[174] Contained within the pool of 492 mutations that were significantly enriched, and predicted to stabilize the active conformation, were the majority of mutations previously identified by two research groups using randomly mutagenized libraries of PAI-1 to identify stable mutants.[112, 114] Taken together, these data suggest our library and approach to characterizing single amino acid substitutions is valid. We believe this approach to massively parallel comprehensive analysis of the mutational landscape has potential to offer insight into the true significance of "variants of unknown significance."

Given that over 400 previously uncharacterized single amino acid substitutions were identified as enriched after prolonged incubation, we attempted to design a high

throughput method to estimate half-lives for all mutants. The results obtained from this method did not agree with previously published half-lives or with those we calculated for recombinant PAI-1 mutants, likely due to the lack of early time points included in our calculations, a limitation that can easily be addressed with additional experiments which are ongoing.

FUTURE DIRECTIONS

Group A *Streptococcus* phage display

As no phage display libraries from GAS have been reported, the construction and characterization of such a library could be a valuable resource in understanding how GAS interacts with the human host. Our laboratory has a special interest in the coagulation cascade and endothelial cell heterogeneity. A GAS phage display library would prove useful in further dissecting the differences between vascular beds by screening the library for GAS proteins that specifically localize to the vasculature of certain organs.

A potential limitation in the construction of the GAS phage display library was the choice of a phagemid vector designed for use with a helper phage system. Although phagemids allow for display of larger proteins on the surface of phage, mutations that alter the reading frame of the phagemid encoded pIII are tolerated. A phagemid containing a DNA insert that is not in frame with pIII will not be displayed on the surface of the phage, but because of the packaging signal, will still be packaged into a phage particle displaying

five copies of WT pIII on the surface. Only 1 out of every 18 clones will have an insert that is in the correct reading frame at the 5' and 3' ends and be in the correct orientation.[19] This estimate is likely too conservative because it does not take into account inserts that might contain non-coding DNA or stop codons.[181] Given this, in our current phagemid approach, close to 95% of phage produced will not display a GAS peptide on their surface, but will still contain a GAS insert in the phagemid. An alternative phage vector such as the fUSE[23] or the hyperphage[182] system might prove more useful in constructing a library from randomly fragmented GAS genomic DNA. These vectors contain only a single copy of the pIII gene, so only inserts that are in-frame with pIII will lead to successful phage production.

Applications of PAI-1 mutant libraries

We are interested in expanding the scope of our experiments to screen our mutant libraries against other serine proteases with the goal of engineering new protease inhibitors with therapeutic potential. As discussed earlier, PAI-1 can be made recombinantly in bacteria in large quantities[102, 103], does not require glycosylation, and does not have any native cysteines, making it a versatile backbone on which to develop novel therapeutics. There is evidence of a naturally occurring point mutation in a serpin resulting in altered specificities of inhibition, in A1AT Pittsburgh. In this specific mutant, a substitution of arginine for methionine at the P1 residue results in the mutant A1AT acting as an antithrombin and causing lethal bleeding.[83] Mutations leading to altered specificity have also been engineered. In 2014, Scott and colleagues generated a phage display library of A1AT Pittsburgh randomly mutated at the P7-P3 sites within the

RCL.[111] They screened this library against thrombin to select for mutants capable of inhibiting the protease and identified two novel mutants that had a increased rate of thrombin inhibition relative to A1AT Pittsburgh.[111] In 2017, Polderdijk *et al* introduced targeted single or combinatorial mutations into P2-P1' of the A1AT Pittsburgh RCL with the goal of switching the specificity of the serpin from inhibition of thrombin to inhibition of activated protein C (APC).[183] Based on protein modeling, this group predicted that large, positively charged residues at P2 and P1' could preferentially inhibit APC over thrombin and subsequently demonstrated that mutation of the WT sequence from PRS to KRK at P2-P1' resulted in a mutant A1AT Pittsburgh with altered specificity for APC.[183] These studies demonstrate the ability to construct serpins with altered specificity, either through screening random mutant libraries or rational design, but both examined a very limited portion of a single serpin. Interestingly, these studies focused on affects of mutations within the RCL on specificity, but in 1990 Lawrence *et al* cloned the RCL of PAI-2 into PAI-1, replacing the native PAI-1 RCL and demonstrated that regions outside the RCL of PAI-1 are important for specificity of inhibition.[184] Given this, an unbiased, high-throughput screen examining mutations across the entire molecule of PAI-1, such as we have outlined here, may be needed in order to identify mutants with markedly altered specificity.

Our mutant library can be screened against other serine proteases to 1) identify positions outside of the PAI-1 RCL that might be important for specificity and 2) engineer novel protease inhibitors with translational applications, such as a novel inhibitor of neutrophil elastase that could be produced recombinantly and used to treat A1AT deficiency. As

discussed in Chapter 1, A1AT deficiency results in uninhibited degradation of the pulmonary tissue by neutrophil elastase. Treatment involves weekly infusions with A1AT purified from human plasma. A stable PAI-1 variant that can be made recombinantly in bacteria and inhibits neutrophil elastase would have great therapeutic potential. Screening the library against neutrophil elastase is a logical starting point, as a PAI-1 variant that inhibits neutrophil elastase at rates comparable to its natural inhibitors has been previously described, providing a positive control we would expect to find in our selected population.[185] Mutation of the P1 arginine in PAI-1 to alanine was shown to alter the specificity from tPA and uPA to neutrophil elastase. [185] Despite this, it was later shown that neutrophil elastase could cleave the mutant PAI-1 RCL at a different, non-inhibitory site within the RCL.[185] This provides an argument for an unbiased screen using our randomly mutagenized library to identify novel inhibitor rather than targeted rational mutagenesis as other residues outside of the targeted region might affect the specificity or viability of a mutant PAI-1.

The latency conversion of PAI-1 could complicate the ability to screen other proteases or develop stable therapeutics. To minimize this, a new randomly mutagenized library can be constructed from a stable PAI-1 variant. Options for such a backbone include, I91L, which has been described as the most stabilizing point mutation ($t_{1/2}$: 18.4 hours), mutant 14-1B, a mutant that harbors 4 different substitutions and has a half-life of over 140 hours[114], or even a previously unknown stabilizing mutation that is present within our library.

Finally, although there is ample knowledge that could be gained from screening our current PAI-1 mutant library, future directions should not be limited to just PAI-1. High throughput phage display is a powerful technique that can and hopefully will be applied to other proteins in order to characterize novel binding interactions, novel functions, or even just to analyze the mutational landscape of a protein with respect to its function.

REFERENCES

1. Armstrong, J., R.N. Perham, and J.E. Walker, *Domain structure of bacteriophage fd adsorption protein.* FEBS Lett, 1981. **135**(1): p. 167-72.
2. Model, P. and M. Russel, *Filamentous Bacteriophage*, in *The Bacteriophages*, R. Calendar, Editor. 1988, Plenum Publishing: New York. p. 375-456.
3. Marvin, D.A., *Filamentous phage structure, infection and assembly.* Curr Opin Struct Biol, 1998. **8**(2): p. 150-8.
4. Frost, L.S., *Conjugative Pili and Pilus-Specific Phages*, in *Bacterial Conjugation*, D.B. Clewell, Editor. 1993, Springer US: Boston, MA. p. 189-221.
5. Day, L., *Family Inovidiae*, in *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Ciruses*, A. King, et al., Editors. 2011, Elsevier Academic Press: San Diego. p. 375-384.
6. Rakonjac, J., et al., *Filamentous Phage: Structure and Biology*, in *Recombinant Antibodies for Infectious Diseases*, T.S. Lim, Editor. 2017, Springer International Publishing: Cham. p. 1-20.
7. Rakonjac, J., et al., *Filamentous bacteriophage: biology, phage display and nanotechnology applications.* Curr Issues Mol Biol, 2011. **13**(2): p. 51-76.
8. Riechmann, L. and P. Holliger, *The C-terminal domain of TolA is the coreceptor for filamentous phage infection of E. coli.* Cell, 1997. **90**(2): p. 351-60.
9. Karlsson, F., et al., *The mechanism of bacterial infection by filamentous phages involves molecular interactions between TolA and phage protein 3 domains.* J Bacteriol, 2003. **185**(8): p. 2628-34.
10. Deng, L.W. and R.N. Perham, *Delineating the site of interaction on the pIII protein of filamentous bacteriophage fd with the F-pilus of Escherichia coli.* J Mol Biol, 2002. **319**(3): p. 603-14.
11. Endemann, H. and P. Model, *Location of filamentous phage minor coat proteins in phage and in infected cells.* J Mol Biol, 1995. **250**(4): p. 496-506.
12. Guan, Y., H. Zhang, and A.H. Wang, *Electrostatic potential distribution of the gene V protein from Ff phage facilitates cooperative DNA binding: a model of the GVP-ssDNA complex.* Protein Sci, 1995. **4**(2): p. 187-97.
13. Bauer, M. and G.P. Smith, *Filamentous phage morphogenetic signal sequence and orientation of DNA in the virion and gene-V protein complex.* Virology, 1988. **167**(1): p. 166-75.
14. Russel, M. and P. Model, *Genetic analysis of the filamentous bacteriophage packaging signal and of the proteins that interact with it.* J Virol, 1989. **63**(8): p. 3284-95.
15. Russel, M., N.A. Linderoth, and A. Sali, *Filamentous phage assembly: variation on a protein export theme.* Gene, 1997. **192**(1): p. 23-32.
16. Waldor, M.K. and J.J. Mekalanos, *Lysogenic conversion by a filamentous phage encoding cholera toxin.* Science, 1996. **272**(5270): p. 1910-4.

17. Wang, Y.A., et al., *The structure of a filamentous bacteriophage.* J Mol Biol, 2006. **361**(2): p. 209-15.
18. Smith, G.P. and V.A. Petrenko, *Phage Display.* Chem Rev, 1997. **97**(2): p. 391-410.
19. Smith, G.P., *Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface.* Science, 1985. **228**(4705): p. 1315-7.
20. Parmley, S.F. and G.P. Smith, *Antibody-selectable filamentous fd phage vectors: affinity purification of target genes.* Gene, 1988. **73**(2): p. 305-18.
21. Cortese, R., et al., *Selection of biologically active peptides by phage display of random peptide libraries.* Curr Opin Biotechnol, 1996. **7**(6): p. 616-21.
22. Devlin, J.J., L.C. Panganiban, and P.E. Devlin, *Random peptide libraries: a source of specific protein binding molecules.* Science, 1990. **249**(4967): p. 404-6.
23. Scott, J.K. and G.P. Smith, *Searching for peptide ligands with an epitope library.* Science, 1990. **249**(4967): p. 386-90.
24. Kehoe, J.W. and B.K. Kay, *Filamentous phage display in the new millennium.* Chem Rev, 2005. **105**(11): p. 4056-72.
25. Barbas, C.F., 3rd, et al., *Assembly of combinatorial antibody libraries on phage surfaces: the gene III site.* Proc Natl Acad Sci U S A, 1991. **88**(18): p. 7978-82.
26. Barbas, C.F. and R.A. Lerner, *Combinatorial immunoglobulin libraries on the surface of phage (Phabs): Rapid selection of antigen-specific fabs.* Methods, 1991. **2**(2): p. 119-124.
27. Barbas III, C.F., et al., *Phage Display: A Laboratory Manual*. 2001, Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
28. Pratt, D., H. Tzagoloff, and J. Beaudoin, *Conditional lethal mutants of the small filamentous coliphage M13. II. Two genes for coat proteins.* Virology, 1969. **39**(1): p. 42-53.
29. Felici, F., et al., *Selection of antibody ligands from a large library of oligopeptides expressed on a multivalent exposition vector.* Journal of Molecular Biology, 1991. **222**(2): p. 301-310.
30. Hoess, R.H., et al., *Identification of a structural epitope by using a peptide library displayed on filamentous bacteriophage.* J Immunol, 1994. **153**(2): p. 724-9.
31. Dybwad, A., et al., *Identification of new B cell epitopes in the sera of rheumatoid arthritis patients using a random nanopeptide phage library.* Eur J Immunol, 1993. **23**(12): p. 3189-93.
32. Sioud, M., O. Forre, and A. Dybwad, *Selection of ligands for polyclonal antibodies from random peptide libraries: potential identification of (auto)antigens that may trigger B and T cell responses in autoimmune diseases.* Clin Immunol Immunopathol, 1996. **79**(2): p. 105-14.
33. O'Neil, K.T., et al., *Identification of novel peptide antagonists for GPIIb/IIIa from a conformationally constrained phage peptide library.* Proteins, 1992. **14**(4): p. 509-15.
34. Saggio, I. and R. Laufer, *Biotin binders selected from a random peptide library expressed on phage.* Biochem J, 1993. **293 (Pt 3)**: p. 613-6.
35. Wrighton, N.C., et al., *Small peptides as potent mimetics of the protein hormone erythropoietin.* Science, 1996. **273**(5274): p. 458-64.

36. Pasqualini, R. and E. Ruoslahti, *Organ targeting in vivo using phage display peptide libraries.* Nature, 1996. **380**(6572): p. 364-6.
37. Matthews, D.J. and J.A. Wells, *Substrate phage: selection of protease substrates by monovalent phage display.* Science, 1993. **260**(5111): p. 1113-7.
38. Ding, L., et al., *Origins of the specificity of tissue-type plasminogen activator.* Proc Natl Acad Sci U S A, 1995. **92**(17): p. 7627-31.
39. Ke, S.H., et al., *Optimal subsite occupancy and design of a selective inhibitor of urokinase.* J Biol Chem, 1997. **272**(33): p. 20456-62.
40. Vindigni, A., Q.D. Dang, and E. Di Cera, *Site-specific dissection of substrate recognition by thrombin.* Nat Biotechnol, 1997. **15**(9): p. 891-5.
41. Kretz, C.A., et al., *Massively parallel enzyme kinetics reveals the substrate recognition landscape of the metalloprotease ADAMTS13.* Proc Natl Acad Sci U S A, 2015. **112**(30): p. 9328-33.
42. Deperthes, D., *Phage display substrate: a blind method for determining protease specificity.* Biol Chem, 2002. **383**(7-8): p. 1107-12.
43. Doorbar, J. and G. Winter, *Isolation of a peptide antagonist to the thrombin receptor using phage display.* J Mol Biol, 1994. **244**(4): p. 361-9.
44. Yee, A., F.L. Tan, and D. Ginsburg, *Functional display of platelet-binding VWF fragments on filamentous bacteriophage.* PLoS One, 2013. **8**(9): p. e73518.
45. Jacobsson, K. and L. Frykberg, *Cloning of ligand-binding domains of bacterial receptors by phage display.* Biotechniques, 1995. **18**(5): p. 878-85.
46. Jacobsson, K. and L. Frykberg, *Phage display shot-gun cloning of ligand-binding domains of prokaryotic receptors approaches 100% correct clones.* Biotechniques, 1996. **20**(6): p. 1070-6, 1078, 1080-1.
47. Morozova, O. and M.A. Marra, *Applications of next-generation sequencing technologies in functional genomics.* Genomics, 2008. **92**(5): p. 255-64.
48. Dias-Neto, E., et al., *Next-generation phage display: integrating and comparing available molecular tools to enable cost-effective high-throughput analysis.* PLoS One, 2009. **4**(12): p. e8338.
49. Brammer, L.A., et al., *A target-unrelated peptide in an M13 phage display library traced to an advantageous mutation in the gene II ribosome-binding site.* Anal Biochem, 2008. **373**(1): p. 88-98.
50. Maruta, F., et al., *Use of a phage display library to identify oligopeptides binding to the lumenal surface of polarized endothelium by ex vivo perfusion of human umbilical veins.* J Drug Target, 2003. **11**(1): p. 53-9.
51. Rahim, A., C. Coutelle, and R. Harbottle, *High-throughput Pyrosequencing of a phage display library for the identification of enriched target-specific peptides.* Biotechniques, 2003. **35**(2): p. 317-20, 322, 324.
52. He, X., S. Liu, and K.L. Perry, *Identification of epitopes in cucumber mosaic virus using a phage-displayed random peptide library.* J Gen Virol, 1998. **79 (Pt 12)**: p. 3145-53.
53. Bentley, D.R., et al., *Accurate whole human genome sequencing using reversible terminator chemistry.* Nature, 2008. **456**(7218): p. 53-9.
54. Schuster, S.C., *Next-generation sequencing transforms today's biology.* Nat Methods, 2008. **5**(1): p. 16-8.

55. Tucker, T., M. Marra, and J.M. Friedman, *Massively parallel sequencing: the next big thing in genetic medicine.* Am J Hum Genet, 2009. **85**(2): p. 142-54.
56. Fowler, D.M., et al., *High-resolution mapping of protein sequence-function relationships.* Nat Methods, 2010. **7**(9): p. 741-6.
57. Larman, H.B., et al., *PhIP-Seq characterization of autoantibodies from patients with multiple sclerosis, type 1 diabetes and rheumatoid arthritis.* J Autoimmun, 2013. **43**: p. 1-9.
58. Larman, H.B., et al., *Autoantigen discovery with a synthetic human peptidome.* Nat Biotechnol, 2011. **29**(6): p. 535-41.
59. Ravn, U., et al., *Deep sequencing of phage display libraries to support antibody discovery.* Methods, 2013. **60**(1): p. 99-110.
60. Xu, G.J., et al., *Viral immunology. Comprehensive serological profiling of human populations using a synthetic human virome.* Science, 2015. **348**(6239): p. aaa0698.
61. Kretz, C.A., et al., *High throughput protease profiling comprehensively defines active site specificity for thrombin and ADAMTS13.* Sci Rep, 2018. **8**(1): p. 2788.
62. Carrell, R. and J. Travis, *α1-Antitrypsin and the serpins: variation and countervariation.* Trends in Biochemical Sciences, 1985. **10**(1): p. 20-24.
63. Law, R.H., et al., *An overview of the serpin superfamily.* Genome Biol, 2006. **7**(5): p. 216.
64. Schick, C., et al., *Cross-class inhibition of the cysteine proteinases cathepsins K, L, and S by the serpin squamous cell carcinoma antigen 1: a kinetic analysis.* Biochemistry, 1998. **37**(15): p. 5258-66.
65. Hammond, G.L., et al., *Primary structure of human corticosteroid binding globulin, deduced from hepatic and pulmonary cDNAs, exhibits homology with serine protease inhibitors.* Proc Natl Acad Sci U S A, 1987. **84**(15): p. 5153-7.
66. Pemberton, P.A., et al., *Hormone binding globulins undergo serpin conformational change in inflammation.* Nature, 1988. **336**(6196): p. 257-8.
67. Clarke, E.P., et al., *A collagen-binding protein in the endoplasmic reticulum of myoblasts exhibits relationship with serine protease inhibitors.* J Biol Chem, 1991. **266**(26): p. 17230-5.
68. Nagata, K., *Hsp47: a collagen-specific molecular chaperone.* Trends Biochem Sci, 1996. **21**(1): p. 22-6.
69. Zou, Z., et al., *Maspin, a serpin with tumor-suppressing activity in human mammary epithelial cells.* Science, 1994. **263**(5146): p. 526-9.
70. Stein, P.E. and R.W. Carrell, *What do dysfunctional serpins tell us about molecular mobility and disease?* Nat Struct Biol, 1995. **2**(2): p. 96-113.
71. Lawrence, D.A., et al., *Serpin-protease complexes are trapped as stable acyl-enzyme intermediates.* J Biol Chem, 1995. **270**(43): p. 25309-12.
72. Silverman, G.A., et al., *The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature.* J Biol Chem, 2001. **276**(36): p. 33293-6.
73. Huntington, J.A., R.J. Read, and R.W. Carrell, *Structure of a serpin-protease complex shows inhibition by deformation.* Nature, 2000. **407**(6806): p. 923-6.

74. Fredenburgh, J.C., A.R. Stafford, and J.I. Weitz, *Conformational changes in thrombin when complexed by serpins.* J Biol Chem, 2001. **276**(48): p. 44828-34.
75. Huntington, J.A., *Serpin structure, function and dysfunction.* J Thromb Haemost, 2011. **9 Suppl 1**: p. 26-34.
76. Peterson, F.C. and P.G. Gettins, *Insight into the mechanism of serpin-proteinase inhibition from 2D [1H-15N] NMR studies of the 69 kDa alpha 1-proteinase inhibitor Pittsburgh-trypsin covalent complex.* Biochemistry, 2001. **40**(21): p. 6284-92.
77. Declerck, P.J., et al., *Identification of a conformationally distinct form of plasminogen activator inhibitor-1, acting as a noninhibitory substrate for tissue-type plasminogen activator.* J Biol Chem, 1992. **267**(17): p. 11693-6.
78. Carrell, R.W. and D.A. Lomas, *Alpha1-antitrypsin deficiency--a model for conformational diseases.* N Engl J Med, 2002. **346**(1): p. 45-53.
79. Laurell, C.B. and S. Eriksson, *The Electrophoretic α;1-Globulin Pattern of Serum in α;1-Antitrypsin Deficiency.* Scandinavian Journal of Clinical and Laboratory Investigation, 1963. **15**(2): p. 132-140.
80. K., F.M. and T.O. W., *SERUM Pi TYPES IN SOME EUROPEAN, AMERICAN, ASIAN AND AFRICAN POPULATIONS.* Acta Pathologica Microbiologica Scandinavica, 1968. **72**(4): p. 601-608.
81. Davis, R.L., et al., *Familial encephalopathy with neuroserpin inclusion bodies.* Am J Pathol, 1999. **155**(6): p. 1901-13.
82. Lomas, D.A. and R.W. Carrell, *Serpinopathies and the conformational dementias.* Nat Rev Genet, 2002. **3**(10): p. 759-68.
83. Owen, M.C., et al., *Mutation of antitrypsin to antithrombin. alpha 1-antitrypsin Pittsburgh (358 Met leads to Arg), a fatal bleeding disorder.* N Engl J Med, 1983. **309**(12): p. 694-8.
84. Sprengers, E.D. and C. Kluft, *Plasminogen activator inhibitors.* Blood, 1987. **69**(2): p. 381-7.
85. Alessi, M.C., et al., *Purification and characterization of natural and recombinant human plasminogen activator inhibitor-1 (PAI-1).* Eur J Biochem, 1988. **175**(3): p. 531-40.
86. Vassalli, J.D., A.P. Sappino, and D. Belin, *The plasminogen activator/plasmin system.* J Clin Invest, 1991. **88**(4): p. 1067-72.
87. Cesarman-Maus, G. and K.A. Hajjar, *Molecular mechanisms of fibrinolysis.* Br J Haematol, 2005. **129**(3): p. 307-21.
88. Dieval, J., et al., *A lifelong bleeding disorder associated with a deficiency of plasminogen activator inhibitor type 1.* Blood, 1991. **77**(3): p. 528-32.
89. Lee, M.H., et al., *Deficiency of plasma plasminogen activator inhibitor 1 results in hyperfibrinolytic bleeding.* Blood, 1993. **81**(9): p. 2357-62.
90. Schleef, R.R., et al., *Bleeding diathesis due to decreased functional activity of type 1 plasminogen activator inhibitor.* J Clin Invest, 1989. **83**(5): p. 1747-52.
91. Fay, W.P., et al., *Human plasminogen activator inhibitor-1 (PAI-1) deficiency: characterization of a large kindred with a null mutation in the PAI-1 gene.* Blood, 1997. **90**(1): p. 204-8.

92. Fay, W.P., et al., *Brief report: complete deficiency of plasminogen-activator inhibitor type 1 due to a frame-shift mutation.* N Engl J Med, 1992. **327**(24): p. 1729-33.
93. Sobel, B.E., *Increased plasminogen activator inhibitor-1 and vasculopathy. A reconcilable paradox.* Circulation, 1999. **99**(19): p. 2496-8.
94. Juhan-Vague, I., et al., *Fibrinolytic factors and the risk of myocardial infarction or sudden death in patients with angina pectoris. ECAT Study Group. European Concerted Action on Thrombosis and Disabilities.* Circulation, 1996. **94**(9): p. 2057-63.
95. Ploplis, V.A., *Effects of Altered Plasminogen Activator Inhibitor-1 Expression on Cardiovascular Disease.* Current drug targets, 2011. **12**(12): p. 1782-1789.
96. Alessi, M.-C. and I. Juhan-Vague, *PAI-1 and the Metabolic Syndrome.* Links, Causes, and Consequences, 2006. **26**(10): p. 2200-2207.
97. Festa, A., et al., *Elevated Levels of Acute-Phase Proteins and Plasminogen Activator Inhibitor-1 Predict the Development of Type 2 Diabetes.* The Insulin Resistance Atherosclerosis Study, 2002. **51**(4): p. 1131-1137.
98. McMahon, B. and H.C. Kwaan, *The plasminogen activator system and cancer.* Pathophysiol Haemost Thromb, 2008. **36**(3-4): p. 184-94.
99. Placencio, V.R. and Y.A. DeClerck, *Plasminogen Activator Inhibitor-1 in Cancer: Rationale and Insight for Future Therapeutic Testing.* Cancer research, 2015. **75**(15): p. 2969-2974.
100. Chan, J.C., et al., *Accelerated skin wound healing in plasminogen activator inhibitor-1-deficient mice.* Am J Pathol, 2001. **159**(5): p. 1681-8.
101. Khan, S.S., et al., *A null mutation in SERPINE1 protects against biological aging in humans.* Science Advances, 2017. **3**(11).
102. Ginsburg, D., et al., *cDNA cloning of human plasminogen activator-inhibitor from endothelial cells.* J Clin Invest, 1986. **78**(6): p. 1673-80.
103. Lawrence, D., et al., *Purification of active human plasminogen activator inhibitor 1 from Escherichia coli. Comparison with natural and recombinant forms purified from eucaryotic cells.* Eur J Biochem, 1989. **186**(3): p. 523-33.
104. Declerck, P.J., et al., *Purification and characterization of a plasminogen activator inhibitor 1 binding protein from human plasma. Identification as a multimeric form of S protein (vitronectin).* J Biol Chem, 1988. **263**(30): p. 15454-61.
105. Kooistra, T., E.D. Sprengers, and V.W. van Hinsbergh, *Rapid inactivation of the plasminogen-activator inhibitor upon secretion from cultured human endothelial cells.* Biochem J, 1986. **239**(3): p. 497-503.
106. Levin, E.G. and L. Santell, *Conversion of the active to latent plasminogen activator inhibitor from human endothelial cells.* Blood, 1987. **70**(4): p. 1090-8.
107. Mottonen, J., et al., *Structural basis of latency in plasminogen activator inhibitor-1.* Nature, 1992. **355**(6357): p. 270-3.
108. Hekman, C.M. and D.J. Loskutoff, *Endothelial cells produce a latent inhibitor of plasminogen activators that can be activated by denaturants.* J Biol Chem, 1985. **260**(21): p. 11581-7.
109. Jendroszek, A., et al., *Latency transition of plasminogen activator inhibitor type 1 is evolutionarily conserved.* Thromb Haemost, 2017. **117**(9): p. 1688-1699.

110. Pannekoek, H., et al., *Functional display of human plasminogen-activator inhibitor 1 (PAI-1) on phages: novel perspectives for structure-function analysis by error-prone DNA synthesis.* Gene, 1993. **128**(1): p. 135-40.
111. Scott, B.M., et al., *Phage display of the serpin alpha-1 proteinase inhibitor randomized at consecutive residues in the reactive centre loop and biopanned with or without thrombin.* PLoS One, 2014. **9**(1): p. e84491.
112. Stoop, A.A., et al., *Different structural requirements for plasminogen activator inhibitor 1 (PAI-1) during latency transition and proteinase inhibition as evidenced by phage-displayed hypermutated PAI-1 libraries.* J Mol Biol, 2001. **305**(4): p. 773-83.
113. Stoop, A.A., et al., *High-density mutagenesis by combined DNA shuffling and phage display to assign essential amino acid residues in protein-protein interactions: application to study structure-function of plasminogen activation inhibitor 1 (PAI-I).* J Mol Biol, 2000. **301**(5): p. 1135-47.
114. Berkenpas, M.B., D.A. Lawrence, and D. Ginsburg, *Molecular evolution of plasminogen activator inhibitor-1 functional stability.* EMBO J, 1995. **14**(13): p. 2969-77.
115. Butler, M.S. and A.D. Buss, *Natural products--the future scaffolds for novel antibiotics?* Biochem Pharmacol, 2006. **71**(7): p. 919-29.
116. Demain, A.L., *Antibiotic discovery: a step in the right direction.* Chem Biol, 2011. **18**(8): p. 939.
117. Pathania, R., et al., *Chemical genomics in Escherichia coli identifies an inhibitor of bacterial lipoprotein targeting.* Nat Chem Biol, 2009. **5**(11): p. 849-56.
118. Rasko, D.A. and V. Sperandio, *Anti-virulence strategies to combat bacteria-mediated disease.* Nat Rev Drug Discov, 2010. **9**(2): p. 117-28.
119. Hogberg, L.D., A. Heddini, and O. Cars, *The global need for effective antibiotics: challenges and recent advances.* Trends Pharmacol Sci, 2010. **31**(11): p. 509-15.
120. Alekshun, M.N. and S.B. Levy, *Targeting virulence to prevent infection: to kill or not to kill?* Drug Discovery Today: Therapeutic Strategies, 2004. **1**(4): p. 483-489.
121. Carapetis, J.R., et al., *The global burden of group A streptococcal diseases.* Lancet Infect Dis, 2005. **5**(11): p. 685-94.
122. Bisno, A.L., M.O. Brito, and C.M. Collins, *Molecular basis of group A streptococcal virulence.* Lancet Infect Dis, 2003. **3**(4): p. 191-200.
123. Hynes, W., *Virulence factors of the group A streptococci and genes that regulate their expression.* Front Biosci, 2004. **9**: p. 3399-433.
124. Tillett, W.S. and R.L. Garner, *THE FIBRINOLYTIC ACTIVITY OF HEMOLYTIC STREPTOCOCCI.* J Exp Med, 1933. **58**(4): p. 485-502.
125. Milstone, H., *Factor in Normal Human Blood which participates in Streptocoeeal Fibrinolysis.* Journal of Immunology, 1941. **42**(2): p. 109-16.
126. Mullertz, S. and M. Lassen, *An activator system in blood indispensable for formation of plasmin by streptokinase.* Proc Soc Exp Biol Med, 1953. **82**(2): p. 264-8.
127. Werkheiser, W.C. and G. Markus, *THE INTERACTION OF STREPTOKINASE WITH PLASMINOGEN. II. THE KINETICS OF ACTIVATION.* J Biol Chem, 1964. **239**: p. 2644-50.

128. Gonzalez-Gronow, M., G.E. Siefring, Jr., and F.J. Castellino, *Mechanism of activation of human plasminogen by the activator complex, streptokinase-plasmin.* J Biol Chem, 1978. **253**(4): p. 1090-4.
129. Reddy, K.N. and G. Markus, *Mechanism of activation of human plasminogen by streptokinase. Presence of active center in streptokinase-plasminogen complex.* J Biol Chem, 1972. **247**(6): p. 1683-91.
130. Sun, H., et al., *Plasminogen is a critical host pathogenicity factor for group A streptococcal infection.* Science, 2004. **305**(5688): p. 1283-6.
131. McArthur, J.D., et al., *The role of streptokinase as a virulence determinant of Streptococcus pyogenes--potential for therapeutic targeting.* Curr Drug Targets, 2012. **13**(3): p. 297-307.
132. Sun, H., et al., *Inhibitor of streptokinase gene expression improves survival after group A streptococcus infection in mice.* Proc Natl Acad Sci U S A, 2012. **109**(9): p. 3469-74.
133. Slack, F.J., et al., *A gene required for nutritional repression of the Bacillus subtilis dipeptide permease operon.* Mol Microbiol, 1995. **15**(4): p. 689-702.
134. Sonenshein, A.L., *CodY, a global regulator of stationary phase and virulence in Gram-positive bacteria.* Curr Opin Microbiol, 2005. **8**(2): p. 203-7.
135. Kreth, J., et al., *Counteractive balancing of transcriptome expression involving CodY and CovRS in Streptococcus pyogenes.* J Bacteriol, 2011. **193**(16): p. 4153-65.
136. Ratnayake-Lecamwasam, M., et al., *Bacillus subtilis CodY represses early-stationary-phase genes by sensing GTP levels.* Genes Dev, 2001. **15**(9): p. 1093-103.
137. Levdikov, V.M., et al., *The structure of CodY, a GTP- and isoleucine-responsive regulator of stationary phase and virulence in gram-positive bacteria.* J Biol Chem, 2006. **281**(16): p. 11366-73.
138. Shivers, R.P. and A.L. Sonenshein, *Activation of the Bacillus subtilis global regulator CodY by direct interaction with branched-chain amino acids.* Mol Microbiol, 2004. **53**(2): p. 599-611.
139. Malke, H. and J.J. Ferretti, *CodY-affected transcriptional gene expression of Streptococcus pyogenes during growth in human blood.* J Med Microbiol, 2007. **56**(Pt 6): p. 707-14.
140. Engleberg, N.C., et al., *Spontaneous mutations in the CsrRS two-component regulatory system of Streptococcus pyogenes result in enhanced virulence in a murine model of skin and soft tissue infection.* J Infect Dis, 2001. **183**(7): p. 1043-54.
141. Heath, A., et al., *A two-component regulatory system, CsrR-CsrS, represses expression of three Streptococcus pyogenes virulence factors, hyaluronic acid capsule, streptolysin S, and pyrogenic exotoxin B.* Infect Immun, 1999. **67**(10): p. 5298-305.
142. Watson, M.E., Jr., et al., *Murine vaginal colonization model for investigating asymptomatic mucosal carriage of Streptococcus pyogenes.* Infect Immun, 2013. **81**(5): p. 1606-17.
143. Kang, S.O., M.G. Caparon, and K.H. Cho, *Virulence gene regulation by CvfA, a putative RNase: the CvfA-enolase complex in Streptococcus pyogenes links*

nutritional stress, growth-phase control, and virulence gene expression. Infect Immun, 2010. **78**(6): p. 2754-67.
144. Mahboubi, A., et al., *Biological activity analysis of native and recombinant streptokinase using clot lysis and chromogenic substrate assay.* Iran J Pharm Res, 2012. **11**(4): p. 1087-93.
145. Chateau, A., et al., *Identification of CodY targets in Bacillus anthracis by genome-wide in vitro binding analysis.* J Bacteriol, 2013. **195**(6): p. 1204-13.
146. Guedon, E., et al., *Overall control of nitrogen metabolism in Lactococcus lactis by CodY, and possible models for CodY regulation in Firmicutes.* Microbiology, 2005. **151**(Pt 12): p. 3895-909.
147. Malke, H., et al., *Linking the nutritional status of Streptococcus pyogenes to alteration of transcriptional gene expression: the action of CodY and RelA.* Int J Med Microbiol, 2006. **296**(4-5): p. 259-75.
148. Vega, L.A., H. Malke, and K.S. McIver, *Virulence-Related Transcriptional Regulators of Streptococcus pyogenes*, in *Streptococcus pyogenes : Basic Biology to Clinical Manifestations*, J.J. Ferretti, D.L. Stevens, and V.A. Fischetti, Editors. 2016, University of Oklahoma Health Sciences Center
(c) The University of Oklahoma Health Sciences Center.: Oklahoma City (OK).
149. Ramirez-Pena, E., et al., *The group A Streptococcus small regulatory RNA FasX enhances streptokinase activity by increasing the stability of the ska mRNA transcript.* Mol Microbiol, 2010. **78**(6): p. 1332-47.
150. Churchward, G., et al., *Regulation of streptokinase expression by CovR/S in Streptococcus pyogenes: CovR acts through a single high-affinity binding site.* Microbiology, 2009. **155**(Pt 2): p. 566-75.
151. O'Loughlin, R.E., et al., *The epidemiology of invasive group A streptococcal infection and potential vaccine implications: United States, 2000-2004.* Clin Infect Dis, 2007. **45**(7): p. 853-62.
152. Zhang, L., et al., *A second IgG-binding protein in Staphylococcus aureus.* Microbiology, 1998. **144 (Pt 4)**: p. 985-91.
153. Vasi, J., et al., *M-like proteins of Streptococcus dysgalactiae.* Infect Immun, 2000. **68**(1): p. 294-302.
154. Nilsson, M., et al., *A fibrinogen-binding protein of Staphylococcus epidermidis.* Infect Immun, 1998. **66**(6): p. 2666-73.
155. Lindmark, H. and B. Guss, *SFS, a novel fibronectin-binding protein from Streptococcus equi, inhibits the binding between fibronectin and collagen.* Infect Immun, 1999. **67**(5): p. 2383-8.
156. Beckmann, C., et al., *Identification of novel adhesins from Group B streptococci by use of phage display reveals that C5a peptidase mediates fibronectin binding.* Infect Immun, 2002. **70**(6): p. 2869-76.
157. Ge, X., et al., *Rapid construction and characterization of synthetic antibody libraries without DNA amplification.* Biotechnol Bioeng, 2010. **106**(3): p. 347-57.
158. Xu, Y., Y. Ma, and H. Sun, *A novel approach to develop anti-virulence agents against group A streptococcus.* Virulence, 2012. **3**(5): p. 452-3.
159. Irving, J.A., et al., *Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function.* Genome Res, 2000. **10**(12): p. 1845-64.

160. De Taeye, B., A. Gils, and P.J. Declerck, *The story of the serpin plasminogen activator inhibitor 1: is there any need for another mutant?* Thromb Haemost, 2004. **92**(5): p. 898-924.
161. Lawrence, D.A., et al., *Partitioning of serpin-proteinase reactions between stable inhibition and substrate cleavage is regulated by the rate of serpin reactive center loop insertion into beta-sheet A.* J Biol Chem, 2000. **275**(8): p. 5839-44.
162. Audenaert, A.M., et al., *Conversion of plasminogen activator inhibitor-1 from inhibitor to substrate by point mutations in the reactive-site loop.* J Biol Chem, 1994. **269**(30): p. 19559-64.
163. Lawrence, D.A., et al., *Serpin reactive center loop mobility is required for inhibitor function but not for enzyme recognition.* J Biol Chem, 1994. **269**(44): p. 27657-62.
164. Tucker, H.M., et al., *Engineering of plasminogen activator inhibitor-1 to reduce the rate of latency transition.* Nat Struct Biol, 1995. **2**(6): p. 442-5.
165. Stout, T.J., et al., *Structures of active and latent PAI-1: a possible stabilizing role for chloride ions.* Biochemistry, 2000. **39**(29): p. 8460-9.
166. Soltes, G., et al., *On the influence of vector design on antibody phage display.* J Biotechnol, 2007. **127**(4): p. 626-37.
167. Quail, M.A., et al., *A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.* BMC Genomics, 2012. **13**: p. 341.
168. Schirmer, M., et al., *Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data.* BMC Bioinformatics, 2016. **17**: p. 125.
169. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
170. Gils, A., et al., *Biochemical importance of glycosylation of plasminogen activator inhibitor-1.* Thrombosis and Haemostasis, 2003. **90**(2): p. 206-217.
171. Abdullah, T., et al., *An Analysis of Single Nucleotide Substitution in Genetic Codons - Probabilities and Outcomes.* Bioinformation, 2016. **12**(3): p. 98-104.
172. Chuang, Y.J., et al., *The antithrombin P1 residue is important for target proteinase specificity but not for heparin activation of the serpin. Characterization of P1 antithrombin variants with altered proteinase specificity but normal heparin activation.* Biochemistry, 2001. **40**(22): p. 6670-9.
173. Rashid, Q., et al., *Understanding the specificity of serpin-protease complexes through interface analysis.* J Biomol Struct Dyn, 2015. **33**(6): p. 1352-62.
174. Sherman, P.M., et al., *Saturation mutagenesis of the plasminogen activator inhibitor-1 reactive center.* J Biol Chem, 1992. **267**(11): p. 7588-95.
175. Gils, A., I. Knockaert, and P.J. Declerck, *Substrate behavior of plasminogen activator inhibitor-1 is not associated with a lack of insertion of the reactive site loop.* Biochemistry, 1996. **35**(23): p. 7474-81.
176. Sim, N.-L., et al., *SIFT web server: predicting effects of amino acid substitutions on proteins.* Nucleic Acids Research, 2012. **40**(W1): p. W452-W457.
177. Aitao, L., et al., *Beating Bias in the Directed Evolution of Proteins: Combining High‐Fidelity on‐Chip Solid‐Phase Gene Synthesis with Efficient Gene Assembly for Combinatorial Library Construction.* ChemBioChem, 2018. **19**(3): p. 221-228.

178. Jankun, J., et al., *Very long half-life plasminogen activator inhibitor type 1 reduces bleeding in a mouse model.* BJU Int, 2010. **105**(10): p. 1469-76.
179. Jankun, J., et al., *Highly stable plasminogen activator inhibitor type one (VLHL PAI-1) protects fibrin clots from tissue plasminogen activator-mediated fibrinolysis.* Int J Mol Med, 2007. **20**(5): p. 683-7.
180. Thomsen, M.C. and M. Nielsen, *Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion.* Nucleic Acids Res, 2012. **40**(Web Server issue): p. W281-7.
181. Hust, M., et al., *Enrichment of open reading frames presented on bacteriophage M13 using hyperphage.* Biotechniques, 2006. **41**(3): p. 335-42.
182. Rondot, S., et al., *A helper phage to improve single-chain antibody presentation in phage display.* Nat Biotechnol, 2001. **19**(1): p. 75-8.
183. Polderdijk, S.G., et al., *Design and characterization of an APC-specific serpin for the treatment of hemophilia.* Blood, 2017. **129**(1): p. 105-113.
184. Lawrence, D.A., et al., *Structure-function studies of the SERPIN plasminogen activator inhibitor type 1. Analysis of chimeric strained loop mutants.* J Biol Chem, 1990. **265**(33): p. 20293-301.
185. Stefansson, S., et al., *Mutants of plasminogen activator inhibitor-1 designed to inhibit neutrophil elastase and cathepsin G are more effective in vivo than their endogenous inhibitors.* J Biol Chem, 2004. **279**(29): p. 29981-7.