

APPLYING PHAGE DISPLAY TO ENGINEER  $\alpha_1$ -PROTEINASE INHIBITOR

APPLYING PHAGE DISPLAY TO SCREEN A LIBRARY OF  $\alpha_1$ -PROTEINASE  
INHIBITOR MUTANTS FOR IMPROVED THROMBIN BINDING ACTIVITY

By

BENJAMIN M. SCOTT, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

© Copyright by Benjamin M. Scott, 2013

MASTER OF SCIENCE (2013)

McMaster University

(Medical Sciences)

Hamilton, Ontario

TITLE: Applying Phage Display to Screen a Library of  $\alpha_1$ -Proteinase Inhibitor Mutants  
for Improved Thrombin Binding Activity

AUTHOR: Benjamin M. Scott, B.Sc. (McMaster University)

SUPERVISOR: Dr. William P. Sheffield

COMMITTEE MEMBERS: Dr. Jonathan Bramson, Dr. Murray Junop

NUMBER OF PAGES: xiv, 153

## ABSTRACT

$\alpha_1$ -proteinase inhibitor ( $\alpha_1$ -PI) is the most abundant serine protease inhibitor (serpin) in plasma. The  $\alpha_1$ -PI M358R mutant exhibits greatly increased rates of thrombin inhibition compared to wild type  $\alpha_1$ -PI, which predominantly inhibits neutrophil elastase. M358R (P1) lies at the reactive centre (P1-P1') bond of the reactive centre loop (RCL) of  $\alpha_1$ -PI, cleaved by cognate proteases as they become trapped in the serpin-type inhibitory complex. The relationship between RCL structure and serpin inhibitor function is incompletely understood and has not been subjected to saturation mutagenesis.  $\alpha_1$ -PI M358R is a less potent inhibitor of thrombin than natural thrombin-inhibitory serpins, suggesting room for engineered improvement into an antithrombotic protein drug.

Phage display is a powerful tool for screening mutant protein libraries, but only one serpin (PAI-1) has previously been mutated and expressed in this manner. In this study the T7Select10-3b (Novagen) phage display system was used to express  $\alpha_1$ -PI variants and PAI-1, fused to the first 348 residues of the T7 10B coat protein. Following confirmation that  $\alpha_1$ -PI M358R retained inhibitory activity when fused to T7Select10-3b phage, this system was used to express a library of  $\alpha_1$ -PI mutant proteins with all possible codon combinations at positions P2 (P357) and P1 (M358) (441 mutants). The library was biopanned using a novel technique in order to amplify only the  $\alpha_1$ -PI P2P1 mutants capable of forming stable complexes with thrombin. The P357/M358R mutant was the only P2P1 mutant enriched, indicating that the  $\alpha_1$ -PI M358R protein has the optimal P2P1 sequence for thrombin inhibition.

A second T7Select10-3b library of  $\alpha_1$ -PI mutant proteins was generated to identify the optimal sequence at positions P7 through to P3 (amino acids 352-356) for thrombin inhibition. The P2 and P1 positions were maintained at P357/M358R, while all possible codon combinations at positions P7 through to P3 were represented (>4.08 million mutants). The library was biopanned using the protocol developed for the P2P1 library, before sequences were inserted into an *E. coli* expression vector and  $\alpha_1$ -PI M358R P7-P3 mutants were screened for thrombin inhibitory activity. 80 individual colonies were screened, yielding 22 unique P7-P3 mutants with thrombin inhibitory activity greater than the M358R RCL sequence. The consensus observed in sequences with improved activity matched thrombin's known substrate specificity and also general RCL trends: P7-Not Aromatic/P6-Hydrophobic/P5-T or S/P4-Hydrophobic/P3-Not Aromatic.

Kinetic characterization of selected mutants with improved thrombin inhibitory activity yielded two mutants, P7-P3 sequence DITMA and AAFVS, with a second order rate constant of  $1.0 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ . This represents a >2-fold increase in the rate of thrombin inhibition versus  $\alpha_1$ -PI M358R. Both the DITMA and AAFVS mutants were found to have a lower stoichiometry of inhibition compared to  $\alpha_1$ -PI M358R, indicating that an improved thrombin inhibitory mechanism was also enriched during biopanning.

These findings suggest that based on the scaffold of the  $\alpha_1$ -PI protein, improved thrombin inhibitory activity can be engineered and selected via phage display. Additionally, this work represents a proof-of-principle for the application of this system to screen libraries of up to 10 million mutants in order to better engineer serpins towards a desired activity.

## **ACKNOWLEDGEMENTS**

First, I would like to thank my supervisor Dr. Bill Sheffield for his constant support, always helpful advice, and needed humor. His guidance and “firing solutions” were always spot-on and encouraged every direction I wished to pursue during my project. I enjoyed my work and research immensely thanks to him, and I couldn’t have imagined a better supervisor. I would also like to thank my committee members Dr. Jonathan Bramson and Dr. Murray Junop for their excellent suggestions on how to improve my project and collegial attitude during every meeting.

In the lab, I would like to especially thank Varsha Bhakta for teaching me everything I now know about lab work and techniques, and selflessly taking time out of her busy schedule whenever I had a question. My lab mates Louise Eltringham-Smith, Sharon Gataiance, Rick Gierczak, Melissa Lambourne, Leigh Ann Roddick each helped me out innumerable times and my three years in the Sheffield lab was always enjoyable because of them. Thanks to Dr. Derek Lobb for giving me a sounding board for my ideas, and a picture of where research can go. I would also like to thank Mac Mok and Seiji Sugiman-Marangos for their assistance with the ZMM software.

Finally, I would like to thank my family and friends for keeping me sane throughout my Master’s. Even if you didn’t completely understand what I was doing all those hours in the lab, your support was always there. Mom, Dad, Linda, Ian and Eden and everyone else, thanks! This time has gone by incredibly fast, and I’m looking forward to the next steps.

## **PREFACE**

All of the work presented in this thesis was done independently by me, with the exception of producing purified  $\alpha_1$ -PI M358R HATVS mutant and native  $\alpha_1$ -PI M358R which was performed by Varsha Bhakta and Sharon Gataiance respectively. This work was presented in preliminary form at the 9<sup>th</sup> Annual Protein Engineering Summit (PEGS) in Boston, MA. and at the XXIV International Society of Thrombosis and Haemostasis (ISTH) Congress in Amsterdam, Netherlands.

## TABLE OF CONTENTS

Title Page .....	i
Descriptive Note .....	ii
Abstract .....	iii
Acknowledgements .....	v
Preface.. .....	vi
Table of Contents .....	vii
List of Figures .....	x
List of Tables .....	xi
List of Symbols and Abbreviations.....	xii
<b>1. INTRODUCTION.....</b>	<b>1</b>
1.1 Coagulation .....	1
1.1.1 The Coagulation Cascade .....	1
1.2.1 Haemostasis .....	5
1.2 Thrombin.....	6
1.2.1 Structural Features .....	6
1.2.2 Specificity and Substrate Binding Pocket.....	10
1.2.3 Pharmaceutical Inhibitors of Thrombin.....	11
1.3 Serpins.....	13
1.3.1 Mechanism of Function .....	13
1.3.2 Importance within the Reactive Center Loop.....	17
1.3.3 Natural Thrombin Inhibitory Serpins .....	22
1.3.4 $\alpha_1$ -Proteinase Inhibitor.....	24
1.4 Engineering $\alpha_1$ -Proteinase Inhibitor.....	27
1.4.1 Primary Engineering Goals.....	27
1.4.2 Improving $\alpha_1$ -PI's Thrombin Inhibitory Function .....	28
1.4.3 Serpin Engineering Challenges.....	30
1.5 Applying Phage Display to Engineering Serpins.....	32
1.5.1 Overview of Phage Display Techniques.....	32
1.5.2 PAI-1.....	35
1.5.3 Modeling Protease Specificity .....	40
1.6 Rationale for Study .....	42
1.7 Hypotheses .....	44
1.8 Project Outline .....	45
<b>2. MATERIALS .....</b>	<b>47</b>
2.1 Source of Chemicals and Reagents .....	47
2.2 Oligonucleotides .....	48
2.3 Computer Software .....	48
<b>3. METHODS .....</b>	<b>51</b>
3.1 Cloning H <sub>6</sub> PAI-1 into pET-11d Vector.....	51
3.2 Expression of H <sub>6</sub> PAI-1 and Purification from <i>E. coli</i> Inclusion Bodies..	51
3.2.1 Culture Conditions for Expression of H <sub>6</sub> PAI-1 .....	51

3.2.2 Confirmation of H <sub>6</sub> PAI-1 Expression by Immunoblot.....	52
3.2.3 Purification of H <sub>6</sub> PAI-1 from <i>E. coli</i> Inclusion Bodies.....	52
3.2.4 Bradford Assay to Quantify Purified H <sub>6</sub> PAI-1 Protein.....	53
3.3 Gel Based Serpin-Enzyme Complex Assays (PAI-1 and tPA).....	54
3.4 Expressing PAI-1 and $\alpha_1$ -PI M358R on T7Select10-3b Phage.....	54
3.4.1 Preparing PAI-1 and $\alpha_1$ -PI M358R Inserts.....	54
3.4.2 Ligating Inserts and Packaging into T7Select Phage.....	55
3.5 Plaque Assay to Determine Titer of Phage.....	55
3.6 Amplification of T7Select10-3b Phage.....	56
3.6.1 Plate Lysate Method.....	56
3.6.2 Liquid Lysate Method.....	56
3.7 Purifying Phage.....	57
3.8 Detection of Expressed Serpins using Immunoblotted Plaque Lifts.....	57
3.9 Unsuccessful Biopanning Methods.....	58
3.9.1 Protease-Coated Well.....	58
3.9.2 Antibody-Coated Well.....	59
3.9.3 T25-Direct Infection.....	59
3.10 Detection of Phage-Expressed Serpin-Enzyme Complexes.....	60
3.11 Generation of $\alpha_1$ -PI P2P1 and $\alpha_1$ -PI M358R P7-P3 T7Select10-3b Phage Libraries.....	61
3.11.1 Preparing pUC19 $\alpha_1$ -PI M358R Plasmid for Sub-cloning.....	61
3.11.2 Generation of Plasmid Libraries using PCR Mutagenesis.....	62
3.11.3 Propagation of Plasmid Libraries.....	64
3.12 Inserting DNA Libraries into T7Select10-3b Phage.....	65
3.13 Confirming Degeneracy of Libraries.....	65
3.13.1 $\alpha_1$ -PI P2P1 Phage Library.....	65
3.13.2 $\alpha_1$ -PI M358R P7-P3 DNA Library.....	66
3.14 Biopanning using Magnetic Dynabeads.....	66
3.14.1 Mock-Biopanning T7Select10-3b $\alpha_1$ -PI M358R.....	66
3.14.2 Biopanning $\alpha_1$ -PI P2P1 and $\alpha_1$ -PI M358R P7-P3 Libraries.....	69
3.15 Determining Optimal P2 and P1 Residues for Thrombin Inhibition.....	70
3.15.1 Sequencing Plaques from Biopanned $\alpha_1$ -PI P2P1 Library.....	70
3.15.2 Inserting Sequences into pBAD Expression Plasmid.....	70
3.15.3 ELISA-Based Lysate Screen.....	70
3.15.4 Purification and Quantification of Recombinant Proteins.....	71
3.15.5 Gel Based Serpin-Enzyme Complex Assays.....	72
3.16 Determining Optimal P7 to P3 Residues for Thrombin Inhibition.....	72
3.16.1 Inserting Sequences into pBAD Expression Plasmid.....	72
3.16.2 ELISA-Based Lysate Screen.....	72
3.16.3 Sequencing Biopanned $\alpha_1$ -PI M358R P7-P3 Library.....	73
3.16.3.1 Randomly Selected Plaque Method.....	73
3.16.3.2 Sequencing based on ELISA-Based Lysate Screen Results.....	73
3.16.4 Purification and Quantification of Recombinant Proteins.....	73
3.16.5 Gel Based Serpin-Enzyme Complex Assays.....	73

3.16.6 Kinetic Analysis of $\alpha_1$ -PI M358R P7-P3 Mutants of Interest ...	74
3.17 Molecular Modeling of $\alpha_1$ -PI Interactions with Thrombin .....	75
<b>4. RESULTS</b> .....	76
4.1 Purification of H <sub>6</sub> PAI-1 from <i>E. coli</i> Inclusion Bodies .....	76
4.2 Purified H <sub>6</sub> PAI-1 Forms SDS Stable Complexes with tPA .....	76
4.3 Expression of PAI-1 and $\alpha_1$ -PI M358R on T7Select10-3b Phage .....	76
4.4 Phage-Expressed $\alpha_1$ -PI M358R Forms a Stable Serpin-Enzyme Complex with Thrombin .....	79
4.5 Generation of $\alpha_1$ -PI P2P1 and $\alpha_1$ -PI M358R P7-P3 T7Select10-3b Libraries ...	84
4.6 Mock-Biopanning T7Select10-3b $\alpha_1$ -PI M358R .....	87
4.7 Determining Optimal P2 and P1 Residues for Thrombin Inhibition .....	90
4.7.1 Sequencing Plaques from Biopanned $\alpha_1$ -PI P2P1 Library .....	90
4.7.2 ELISA-Based Lysate Screen of Unique $\alpha_1$ -PI P2P1 Sequences... ..	90
4.7.3 Gel Based Serpin-Enzyme Complex Assays .....	93
4.8 Determining Optimal P7 to P3 Residues for Thrombin Inhibition .....	98
4.8.1 Sequencing Plaques from Biopanned $\alpha_1$ -PI M358R P7-P3 Library.....	98
4.8.2 ELISA-Based Lysate Screen as a “6 <sup>th</sup> Round” of Biopanning ..	102
4.8.3 Analysis of Sequences with Improved Thrombin Binding.....	107
4.9 Functional Characterisation of $\alpha_1$ -PI M358R P7-P3 Mutants of Interest .....	107
4.9.1 Selecting Mutants of Interest .....	107
4.9.2 Gel Based Serpin-Enzyme Complex Assays .....	110
4.9.3 Kinetic Analysis of $\alpha_1$ -PI M358R P7-P3 Mutants of Interest ...	110
4.10 Molecular Modeling of $\alpha_1$ -PI Interactions with Thrombin .....	113
<b>5. DISCUSSION</b> .....	125
5.1 Recombinant PAI-1 must be Refolded and can be Expressed using T7Select10-3b .....	126
5.2 $\alpha_1$ -PI M358R can be Expressed using T7Select10-3b and Forms Stable Serpin- Enzyme Complexes with Thrombin.....	127
5.3 Mock-Biopanning Successfully Enriches T7Select10-3b $\alpha_1$ -PI M358R phage ....	128
5.4 The Optimal P2 and P1 Residues for Thrombin Inhibition .....	129
5.5 The P7 to P3 Region Influences Thrombin Inhibition .....	130
5.6 Phage Displayed Serpins account for Protease Specificity and RCL Insertion .....	132
5.7 Characterisation of Selected Mutants of Interest.....	135
<b>6. CONCLUSIONS AND FUTURE DIRECTIONS</b> .....	138
<b>7. REFERENCES</b> .....	141
<b>APPENDIX A</b> .....	150
<b>APPENDIX B</b> .....	153

## LIST OF FIGURES (abbreviated titles)

Figure 1	Coagulation Cascade .....	3
Figure 2	Tertiary Structure of Thrombin .....	9
Figure 3	Tertiary Structure of a Serpin and the Serpin Mechanism .....	16
Figure 4	Human Serpins Capable of Inhibiting Thrombin .....	21
Figure 5	Structure of T7Select Phage .....	37
Figure 6	Construction of $\alpha_1$ -PI Phage Libraries .....	63
Figure 7	Biopanning Procedure .....	68
Figure 8	H <sub>6</sub> PAI-1 Forms SDS Stable Complexes with tPA .....	78
Figure 9	Immunoblotted Plaque Lifts .....	81
Figure 10	Detection of Phage-Expressed Serpin-Enzyme Complexes.....	83
Figure 11	Confirmatory Sequencing of $\alpha_1$ -PI Libraries .....	86
Figure 12	Enrichment of $\alpha_1$ -PI M358R during Mock-Biopanning.....	89
Figure 13	Frequencies of $\alpha_1$ -PI P2P1 Sequences after Biopanning.....	92
Figure 14	Amount of $\alpha_1$ -PI P2P1 Mutant Protein Bound to Iia .....	95
Figure 15	$\alpha_1$ -PI P2P1 P357/M358R and Iia Serpin-Enzyme Complexes.....	97
Figure 16	P7-P3 Sequences from Plaques following Biopanning.....	100
Figure 17	Amount of $\alpha_1$ -PI M358R P7-P3 Mutant Protein Bound to Iia.....	101
Figure 18	Frequencies of Each Amino Acid for $\alpha_1$ -PI M358R P7-P3 Sequences found to have Greater Thrombin Binding Activity than $\alpha_1$ -PI M358R.....	109
Figure 19	$\alpha_1$ -PI M358R P7-P3 Mutant and Iia Serpin-Enzyme Complexes....	112
Figure 20	Encounter Complexes between $\alpha_1$ -PI Variants and Iia .....	119

## LIST OF TABLES

Table 1	List of Oligonucleotide Primers Used During this Project.....	50
Table 2	$\alpha_1$ -PI M358R P7-P3 Sequences Found to Have Greater Thrombin Binding Activity than $\alpha_1$ -PI M358R .....	106
Table 3	Second Order Rate Constant of Thrombin Inhibition .....	115
Table 4	Stoichiometry of Inhibition with Thrombin .....	117
Table 5	Specific Energies of Serpin-Thrombin Encounter Complexes .....	123

## LIST OF SYMBOLS AND ABBREVIATIONS

$\alpha_1$ -PI	$\alpha_1$ -proteinase inhibitor
$\beta$	beta
$\mu$ L	microliter
$\mu$ M	micromolar
$^{\circ}$ C	degrees Celsius
AAFVS	alanine, alanine, phenylalanine, valine, serine
Ala	alanine
AP	alkaline phosphatase
APC	activated protein C
Arg	arginine
Asp	aspartic acid
AT	antithrombin
BCIP	5-bromo-4-chloro-3-indolyl phosphate
bp	base pair
BSA	bovine serum albumin
CS	cleaved serpin
C-terminal	carboxy-terminal
DAB	diaminobenzidine tetrahydrochloride hydrate
ddH <sub>2</sub> O	double distilled water
DEAE	diethylaminoethyl
DITMA	aspartic acid, isoleucine, threonine, methionine, alanine
DNA	deoxyribonucleic acid
E	enzyme
EDTA	ethylenediaminetetraacetic acid
EATVS	glutamic acid, alanine, threonine, valine, serine
<i>E. coli</i>	<i>Escherichia coli</i>
EISLQ	glutamic acid, isoleucine, serine, leucine, glutamine
F	Factor
Fab	antibody fragment
GAG	glycosaminoglycan
Glu	glutamic acid

Gly	glycine
HATVS	histidine, alanine, threonine, valine, serine
HCl	hydrochloric acid
HCII	heparin cofactor II
His	histidine
HRP	horseradish peroxidase
IgG	immunoglobulin G
Ila	thrombin
Ile	isoleucine
IPTG	isopropyl $\beta$ -D-1-thiogalactopyranoside
$k_2$	second order rate constant
kb	kilobase
kDa	kiloDalton
L	litre
LB	Luria-Bertani broth
Leu	leucine
M	molar
mAB	monoclonal antibody
mg	milligram
ml	milliliter
mM	millimolar
Met	methionine
MOBIX	Molecular Biology and Biotechnology Institute, McMaster University
N-terminal	amino-terminal
Na <sup>+</sup>	sodium ion
NaCl	sodium chloride
NaHCO <sub>3</sub>	sodium bicarbonate
NaOAc	sodium acetate
NBT	nitro blue tetrazolium
Ni-NTA	nickel-nitrilotriacetic acid
ng	nanograms
nm	nanometer

nM	nanomolar
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PAI-1	plasminogen activator inhibitor-1
PBS	phosphate-buffered saline
PBST	phosphate-buffered saline with tween
PCR	polymerase chain reaction
pfu	plaque forming unit
PN1	protease nexin 1
PPACK	D-phenylalanyl-L-prolyl-arginine chloromethyl ketone
Pro	proline
R <sup>2</sup>	regression coefficient
RCL	reactive center loop
Rd5 +IIa	library biopanned for five rounds with thrombin
Rd5 -IIa	library biopanned for five rounds without thrombin
S	serpin
S-2238	H-D-Phenylalanyl-L-pipecolyl-L-arginine-p-nitroaniline
SD	standard deviation
SDS	sodium dodecyl sulfate
SEC	serpin-enzyme complex
Ser	serine
SI	stoichiometry of inhibition
TAFI	thrombin activable fibrinolysis inhibitor
TBST	Tris-buffered saline with tween
TEMED	N,N,N',N'-tetramethylethylenediamine hydrochloride
TF	tissue factor
TFPI	tissue factor pathway inhibitor
Thr	threonine
TMB	3,3',5,5'-Tetramethylbenzidine
tPA	tissue plasminogen activator
Trp	tryptophan
Tyr	tyrosine

# 1. INTRODUCTION

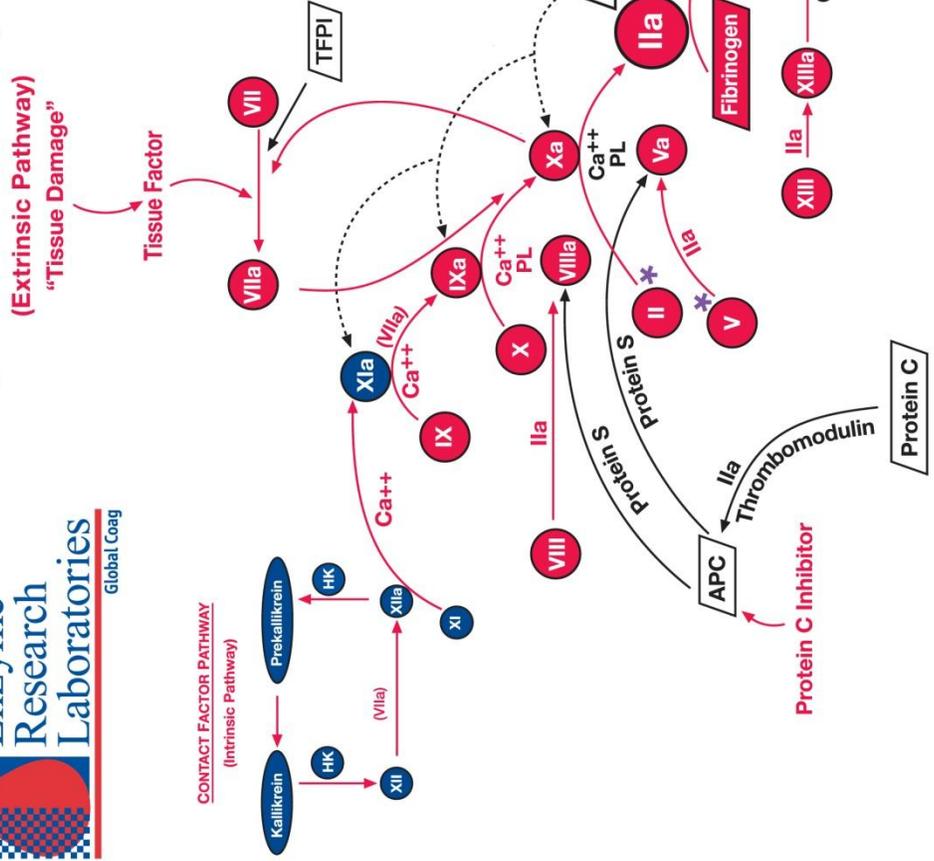
## 1.1 Coagulation

### 1.1.1 The Coagulation Cascade

The coagulation cascade is critical to responding to tissue damage, acting to halt blood loss rapidly through the coordinated action of several serine proteases. This vital biological mechanism is an example of a serine protease cascade, wherein a protease initially exists as a zymogen until it is specifically activated by an active protease upstream (1). This topology serves to rapidly amplify a local signal, such as tissue damage, providing an acute response in the required area. The coagulation cascade can be described as two connected pathways, ultimately leading to the formation of a fibrin clot: the intrinsic and extrinsic pathways (**Figure 1**). The extrinsic pathway begins with tissue damage which exposes circulation to the transmembrane receptor tissue factor (TF), which is normally hidden in the subendothelial space (2). TF binds and activates Factor VII (FVII), which also exists in the activated form (FVIIa) in low levels in plasma in order to prime coagulation (3). The membrane associated TF-FVIIa complex then activates FX via proteolytic cleavage and FXa binds to negatively charged phospholipids (i.e. an activated platelet) (2). The proteolytic activity from one molecule of FXa, in complex with its protein cofactor FVa (prothrombinase), is able to generate thousands of molecules of thrombin (IIa) via the cleavage of its zymogen prothrombin (4). Thrombin then completes the end goal of coagulation by converting fibrinogen into fibrin, allowing the fibrin monomers to spontaneously aggregate (2). FXIII is also activated by thrombin, serving to stabilise fibrin by covalent cross-linking (2).

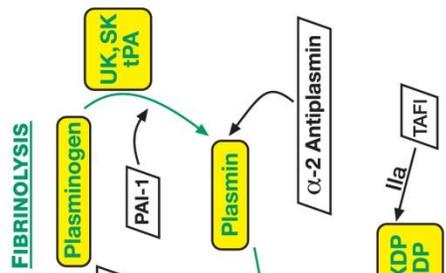
**Figure 1: Coagulation Cascade.** A schematic illustrating the intrinsic, extrinsic, and fibrinolytic pathways including physiological inhibitors. Courtesy of Enzyme Research Laboratories, Inc., 2004. Reprinted with written permission of Enzyme Research Laboratories Inc., South Bend, IN, provided on July 15, 2013.

**TISSUE FACTOR PATHWAY**  
(Extrinsic Pathway)  
"Tissue Damage"



Component	Molecular Weight	Plasma Concentration µg/ml	Plasma Concentration µM
Fibrinogen (I)	330,000	3000	9.09
Prothrombin (II)	72,000	90	1.388
Factor V	330,000	10	0.03
Factor VII	50,000	0.5	0.01
Factor VIII	330,000	0.1	0.0003
Factor IX	56,000	5	0.08928
Factor X	58,800	8	0.13605
Factor XI	160,000	5	0.031
Factor XII	80,000	30	0.375
Factor XIII	320,000	10	0.03125
Protein C	62,000	4	0.0645
Protein S	69,000	10 (free)	0.1449
Protein Z	62,000	2.2	0.0355
Prekallikrein	86,000	50	0.5814
HK	110,000	70	0.6363
Fibrinectin	450,000	300	0.6667
Antithrombin III	58,000	140**	2.4**
Plasminogen	90,000	216	2.4
Urokinase	53,000	0.1	0.001887
Heparin Cofactor II	66,000	90	1.3636
Alpha <sub>2</sub> -Antiplasmin	63,000	60	0.9524
Protein C Inhibitor	57,000	4	0.0702
Alpha <sub>2</sub> -Macroglobulin	725,000	2100	2.8966

Inhibitors in black  
\*Thrombophilia gene mutation  
\*\*Corrected value



The intrinsic pathway is a similar serine protease cascade, converging with the extrinsic pathway with the activation of FX (5). Contact with negatively charged surfaces *in vitro* activates FXII, initiating the intrinsic pathway (2). Kallikrein is activated by FXIIa which acts to amplify the intrinsic pathway response by activating both additional FXIIa and kallikrein (6). A sequential cascade then follows, with FXIIa activating FXI by proteolysis, and FXIa activating FIX (5). FIXa and its cofactor FVIIIa form a complex akin to the TF-VIIa complex in that it also functions to activate FXa and thus generate more thrombin (2). The specific function of the intrinsic pathway is poorly defined, as while FVIII or FIX deficiency causes haemophilia, FXI or FXII deficiency results in only a mild increased bleeding risk (5). This is due to the fact that both FVIII and FIX can be activated outside of the intrinsic pathway by thrombin and the TF-FVIIa complex respectively, negating a crucial need for FXI and FXII (5). Besides its end goal of activating fibrinogen, thrombin also activates FV and the intrinsic pathway proteases FVIII and FXI (2). Thus, the actions of the intrinsic pathway play a large role in amplifying thrombin generation, and combined with the positive feedback in the extrinsic pathway, the rate of thrombin generation increases over 300,000 fold (7).

Neither the intrinsic nor extrinsic pathway is sufficient to form a stable clot, highlighting the interconnectedness within the coagulation cascade. Platelets also play a crucial role, as once activated by von Willebrand factor bound to damaged collagen, they create the physical basis of the clot on which fibrin is crosslinked (6). Activated platelets also serve as a surface for prothrombinase; they release additional coagulation proteases, and activate additional platelets to rapidly form a clot (2, 6, 8). Ensuring that the

formation of thrombin and the activation of the upstream proteases remain localised is crucial to maintaining haemostasis, regulation of which occurs via a variety of proteins to counterbalance this rapid response to bleeding.

### **1.1.2 Haemostasis**

Like other serine protease cascades, coagulation is tightly regulated via control of the proteolytic process that activates the zymogens and by direct inactivation of the proteases. Besides activating procoagulant proteins, thrombin also has anticoagulant functions. When bound to the cell surface receptor thrombomodulin, thrombin preferentially cleaves protein C, yielding activated protein C (APC) (9). The coagulation cascade then begins to slow as APC directly degrades FVa and FVIIIa (9). Tissue factor pathway inhibitor works further upstream in the coagulation cascade, inhibiting the TF-VIIa complex in a Xa dependant manor, preventing additional formation of Xa (8).

Simultaneously activated with the coagulation cascade is fibrinolysis, which breaks down the fibrin clot, and is a process governed by another serine protease cascade (**Figure 1**). Following injury, endothelial cells release tissue plasminogen activator (tPA), which binds fibrin as the clot is formed (2). The zymogen plasminogen also binds fibrin and is converted to its active form plasmin by tPA, which goes on to degrade both fibrin and fibrinogen (2). Fibrinolysis is in turn regulated by the serine protease inhibitors  $\alpha$ -2 antiplasmin and plasminogen activator inhibitor-1 (PAI-1), which inhibit plasmin and tPA respectively (Section 1.3) (10). Thrombin also controls fibrinolysis through the thrombin-activatable fibrinolysis inhibitor (TAFI), which removes the terminal lysines from fibrin to which both plasmin and tPA bind (2). The serine protease inhibitors antithrombin (AT)

and heparin cofactor II (HCII) are the two primary inhibitors of thrombin (Section 1.3.3). At the site of injury, glycosaminoglycans (GAGs) expressed in the sub-endothelium activate these inhibitors, resulting in a  $10^2$ – $10^5$  fold increase in the rate of thrombin inhibition (10). AT also has significant inhibitory activity towards FIXa and FXa, reducing additional thrombin generation (10). Thus, much like thrombin activation, thrombin inhibition is also localised.

Haemostasis functions as a fine balance due to the competing actions of clot formation and degradation, and the variety of regulatory mechanisms therein. This balance can be offset in a variety of disease states including: atherosclerosis, cancer, sepsis, and atrial fibrillation. The regulation of haemostasis is a primary focus of pharmaceutical development, particularly in the creation of anticoagulants which work to impede coagulation and therefore the formation of aberrant clots, thrombosis. Due to the critical importance of thrombin in coagulation it is one of the primary drug targets for both preventing and treating thrombosis.

## **1.2 Thrombin**

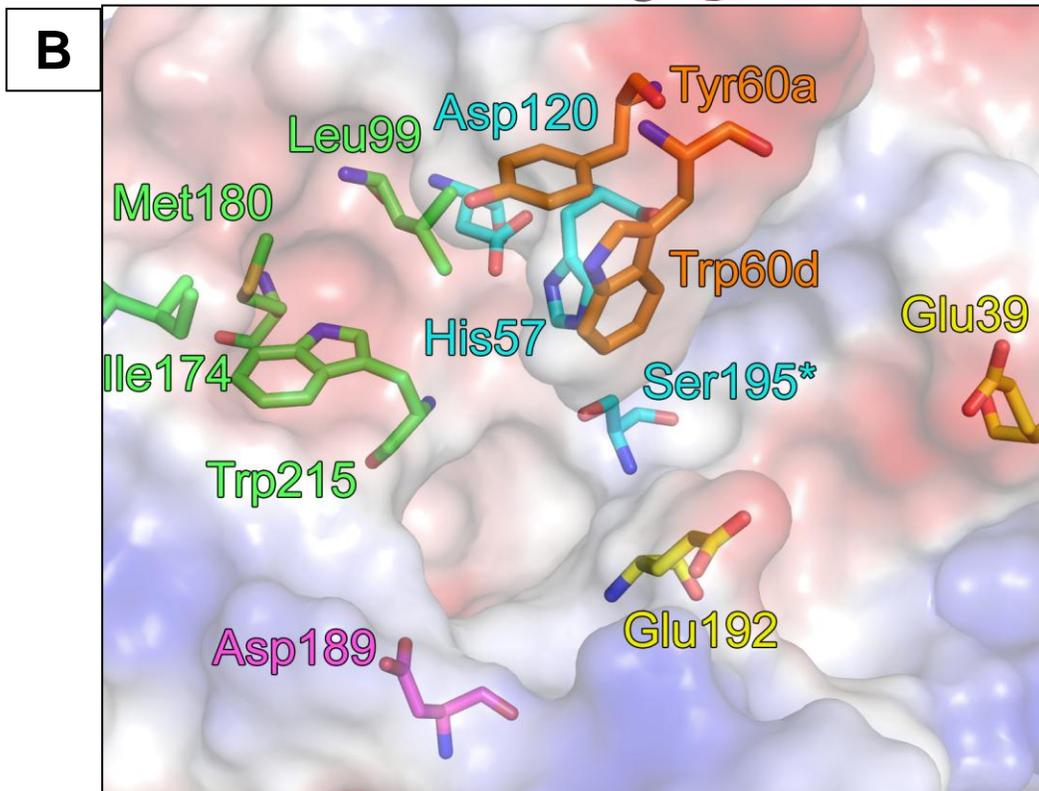
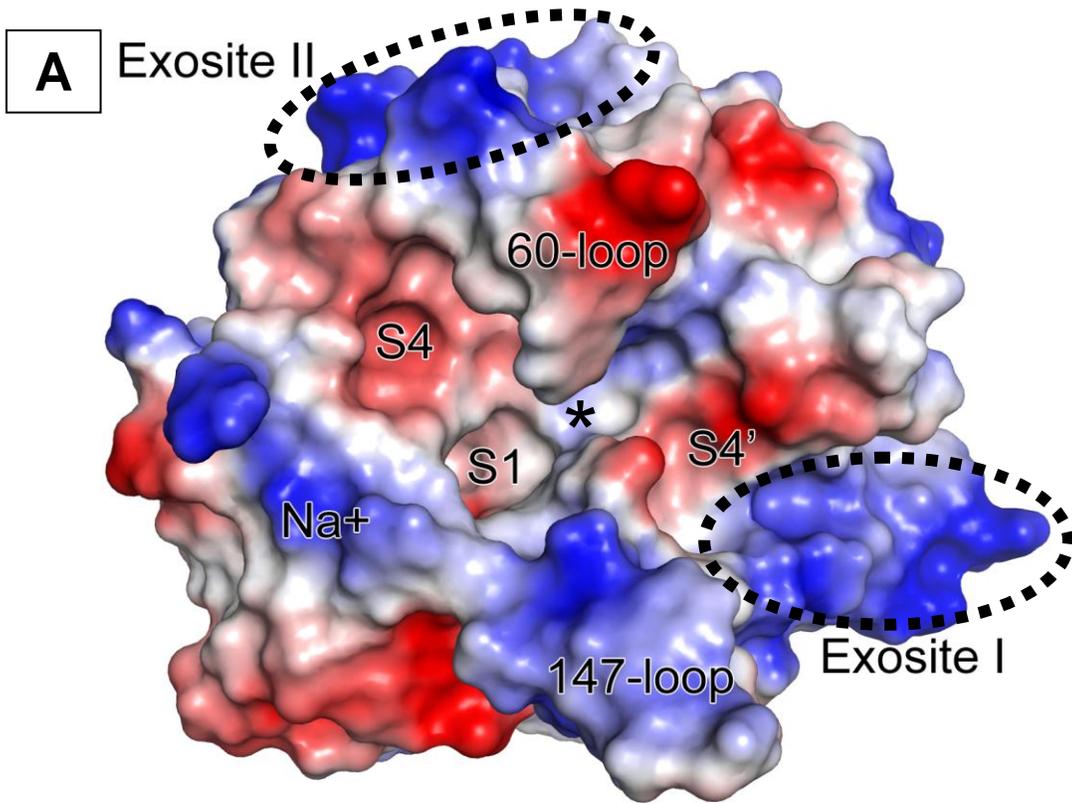
### **1.2.1 Structural Features**

The unique tertiary structure of thrombin affords it a wide variety of interactions, helping to explain its critical role within haemostasis. Like many of the coagulation factors, thrombin is a serine protease in the chymotrypsin family and circulates as an inactive zymogen, prothrombin. Following proteolysis by FXa a 37 kDa product is produced, separated into a light (49 amino acid) and heavy (249 amino acid) chain connected by a single disulfide bond (11). This fully processed thrombin ( $\alpha$ -thrombin, IIa) contains

several distinct regions, and the protease's function varies greatly depending on which regions are engaged (**Figure 2A**) (12). Outside of the active site, two large exosites comprised of basic residues host a variety of haemostatic proteins and molecules. Exosite I is proximal to where the C terminal residues of a substrate sit in the active site, and via exosite I thrombin binds thrombomodulin and it helps to position fibrinogen within the active site (13). Exosite II, located further away from the active site, serves as the GAG binding domain and also interacts with platelets and several coagulation factors, including FVa and FVIIIa (12, 14). In order to prevent prothrombin from competing with active thrombin via the exosites, both exosites are blocked prior to processing by FXa, prompting Huntington (2012) to observe that prothrombin acts as a “super-zymogen” (14). A small Na<sup>+</sup> binding pocket nearby the active site is also important for modulating thrombin's function. Substrate hydrolysis increases when Na<sup>+</sup> is bound, but when the pocket is unoccupied thrombin binds thrombomodulin with a higher affinity (14). Thus, this relatively small change has a central effect on determining whether thrombin acts in a procoagulant or anticoagulant manor.

Approaching the active site, thrombin contains two unique “insertion loops” on either side of where the substrate binds, creating a canyon-like cleft (14). This structure serves to physically limit access to the active site, and presents a large surface area for interactions with the substrate (14). The 147-loop protrudes outwards into the solvent and assists in fibrinogen, AT and Na<sup>+</sup> binding (15, 16). The 60-loop forms a rigid pocket above the catalytic site, comprised of residues Tyr60a and Trp60d (**Figure 2B**) (17). Based on the proximity of the 60-loop to the catalytic site, it is an important factor for

**Figure 2: A) Tertiary Structure of Thrombin.** The active site (\*) is flanked by two unique insertion loops, 60-loop and 147-loop. Substrate binding pockets S4 through to S4' govern thrombin's substrate specificity, which can be influenced by ligands binding either exosite and when the Na<sup>+</sup> pocket is occupied. Surface is coloured based on electrostatic potential (red: negative, blue: positive). **B) Structural Features Governing Substrate Specificity.** Several hydrophobic residues (green) within the S4 binding pocket govern thrombin's substrate preference for a hydrophobic residue at substrate position P4. Asp189 deep within the S1 pocket preferentially binds arginine residues at P1. Catalytic triad is noted in blue, with access to this site limited by two bulky residues within the 60-loop (orange). Glu39 and Glu192 influence specificity within the S2'-S4' region, preferentially binding basic residues. Ser195 marked with \* as the original structure contains a mutant alanine at this position. Generated using PyMOL, PDB ID 1JMO (18).



determining thrombin's substrate specificity, as the pocket it forms explains thrombin's strong preference for proline at the P2 position of a substrate (17, 19). When Na<sup>+</sup> is bound by thrombin the 60-loop opens to allow access to the active site, largely governed by the occluding Trp60d residue (20). This provides yet another example of thrombin's "plasticity" and how this enzyme is able to perform a range of tightly regulated functions.

### 1.2.2 Specificity and Substrate Binding Pocket

Thrombin's substrate specificity profile has been determined by a variety of means, which have provided mostly complimentary results. Gallwitz *et al.* (2012) provide a useful summary, with thrombin specificity fitting the general consensus sequence: P4-hydrophobic/P3-broad specificity/P2-Pro/P1-Arg/P1'-Ser/P2'-broad specificity/P3'-Arg/P4'-broad specificity (P side is N-terminal to protease cleavage site, P' side is C-terminal) (21). The preferences for the P2-P1' residues are the most tightly constrained, and specificity has not been determined beyond P4-P4' (19, 21, 22). A structural investigation of thrombin's substrate binding pocket helps to explain these findings. On the P side, determinants of specificity are relatively easy to observe as noted in the previous section regarding the 60-loop. Similar to FXa and trypsin, the conserved Asp189 residue within the deep S1 pocket preferentially binds to arginine which orients the backbone of a substrate towards the catalytic site, explaining the strong P1 preference (**Figure 2B**) (23). The shallow S4 region of thrombin is highly hydrophobic, which accommodates the widely published observation that substrates of thrombin typically contain a hydrophobic P4 residue which is inserted into this pocket (19, 21). Promiscuity for P3 residues can again be explained by thrombin's structure, as the S3 region is

relatively open and an inserted P3 residue faces away from the body of thrombin leading to minimal interactions (17).

Specificity determinants on the P' side of the binding pocket are more difficult to observe. The preference for serine at P1' is a result of the confined S1' region beneath the 60-loop, supported by the finding that only similarly small uncharged residues can be tolerated without significantly reducing the rate of proteolysis (Ala, Gly, Thr) (21, 22). Glu39 and Glu192 direct specificity within the S2'-S4' region, maintaining a preference for positively charged residues at the P3' position and interacting unfavourably with acidic residues (24, 25). In general the P' region can accommodate a wide range of residues, though its specificity can be modulated when the nearby exosite I binds a ligand (14, 25). In addition to this substrate flexibility, there is a level of cooperativity between substrate residues in that they may influence the binding of one another. Gallwitz *et al.* (2012) observed that the absence of a proline at P2 lead to a stronger preference for arginine at P3', which they suggested compensated for the lack of the preferred proline residue (21). Understanding the properties of thrombin's substrate binding pocket is critical for the rational design of novel inhibitors for use as anticoagulants, and for elucidating the mechanisms of naturally occurring thrombin inhibitors.

### **1.2.3 Pharmaceutical Inhibitors of Thrombin**

Clinically approved inhibitors of thrombin either bind it directly to disrupt its function or activate the body's own thrombin inhibitory proteins. Glycosaminoglycans (GAGs) are part of the latter category as they help to bridge interactions between serpins and thrombin by binding both exosite II of thrombin and a serpin exosite (Section 1.3.3)

(26). Heparins, a clinically approved type of GAG isolated from pig intestine, range in size from 2-30 kDa and are highly negatively charged which requires them to be injected (12). Heparin preferentially binds circulating thrombin over fibrin-bound thrombin, although they must contain at least 18 saccharide units in order to form a complex between AT and thrombin (12, 27). Hirudins are another class of thrombin inhibitors, which inhibit thrombin directly with an extremely high specificity (28). Peptides of ~65 amino acids in length, hirudins bind exosite I of thrombin via an acidic tail and also insert via non-canonical hydrophobic interactions into the active site of thrombin, preventing fibrinogen from binding (9, 28). Originally discovered in medicinal leeches, a number of recombinant hirudin peptides have been approved for clinical usage. Unlike heparins, hirudins can bind both circulating and fibrin-bound thrombin, and do not have the associated risk of thrombocytopenia (28). However, because hirudins have a high risk for bleeding events, a short half-life *in vivo*, and high cost of production, the clinical uses have been limited and require supervised administration (28).

In 2010 a new direct thrombin inhibitor, dabigatran, was approved for usage to prevent stroke due to atrial fibrillation and to reduce the risk of venous embolism following hip and knee replacement (29). The prodrug dabigatran etexilate (Pradax™ in Canada, Pradaxa™ elsewhere), is administered orally and can be taken without routine monitoring, offering significant advantages over classical thrombin inhibitors (30). Unlike peptide-based hirudins, dabigatran is a small molecule drug and binds reversibly to only the active site of both fibrin bound and circulating thrombin (30). It was developed over the course of 10 years via high-throughput screening of chemical leads,

beginning with known active site inhibitors and optimising both their efficacy and oral bioavailability (31). No reversal agent or antidote was available at the time of clinical approval, and in 2011 542 patient deaths in the U.S. were linked to dabigatran, more than any other drug that year (32). Following an FDA investigation, no significant increase in bleeding was discovered compared to other anticoagulants, though the need for an antidote remained (33). Boehringer Ingelheim, the original producers of dabigatran, have recently published evidence that dabigatran can be counteracted using a novel agent. A recombinant antibody fragment (Fab) specific to dabigatran was developed by immunizing mice with the drug (34). Bleeding models in rats indicated that the Fab successfully reversed the actions of dabigatran infusion, decreasing anticoagulant activity without an increase in clotting activity (34). This promising reversal agent is now being carried forward to human trials, although future thrombin inhibitors will ideally not need such drastic countermeasures (35). Better understanding the function of natural thrombin inhibitors offers an additional route towards the development of novel anticoagulants, by using what evolution has provided and extending protein engineering techniques to better direct function.

## **1.3 Serpins**

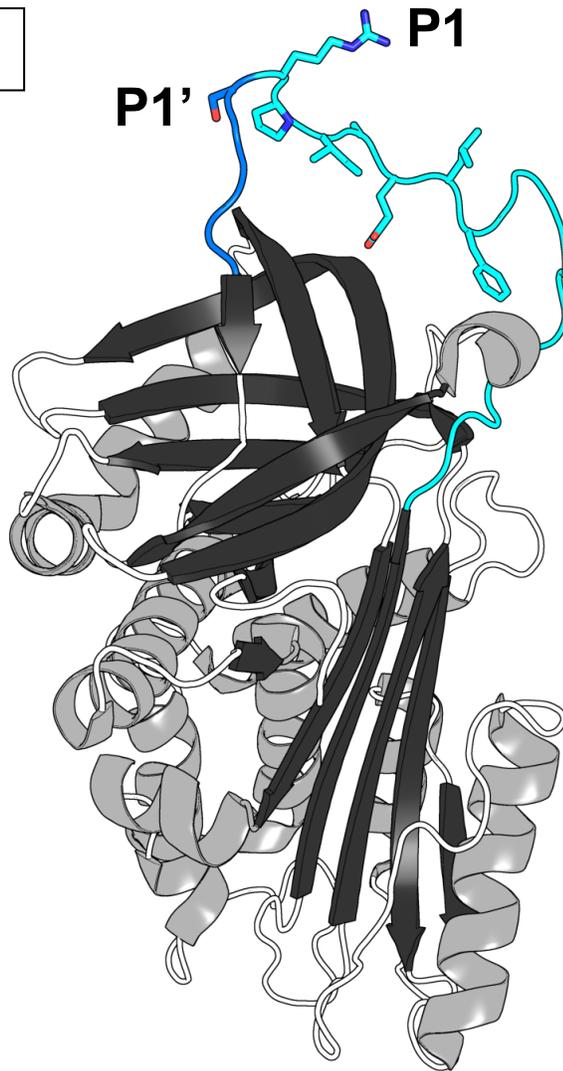
### **1.3.1 Mechanism of Function**

Serine protease inhibitors (serpins) are primarily inhibitors of proteases in the chymotrypsin family and play vital roles in regulating serine protease cascades by limiting the amount of active protease present (36, 37). Over 1500 serpin genes have been identified in genomes from all forms of life, including viruses, with eukaryotic serpins

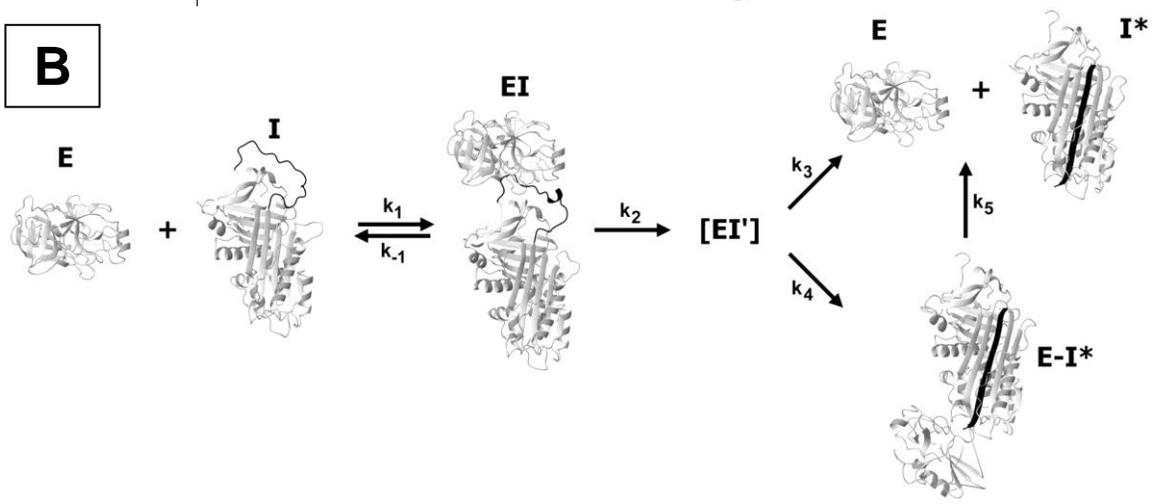
categorized into 16 clades (38). Of the 36 human serpins, 27 function as protease inhibitors while the non-inhibitory serpins' functions include acting as chaperone proteins and hormone transporters (38). The core structure of serpins includes three  $\beta$ -sheets and eight to nine  $\alpha$ -helices, which is highly conserved across phyla due to the requirement for retention of a functional structure (**Figure 3A**) (38). The majority of variability occurs in the reactive center loop (RCL), a ~25 residue unstructured loop which functions as the bait for the protease targeted for inhibition (39, 40). Conventional nomenclature labels residues N-terminal to the protease cleavage site as P residues (P15-P1) and those that are C-terminal P' (P1'-P10') (41, 42). A target protease initially forms a non-covalent Michaelis-Menten "encounter complex" with the RCL, wherein the active site of the protease forms favourable interactions with RCL residues proximal to the P1-P1' site, resembling interactions between the protease and its natural substrate(s) (10). Crystal structures of encounter complexes indicate that the RCL inserts into the active site freely, with neither the protease nor the RCL undergoing significant conformational change (10, 26). Via a nucleophile attack from the protease's active site serine, the peptide bond between RCL residues P1 and P1' is cleaved by the protease, resulting in a covalent acyl-enzyme intermediate between the hydroxyl group of the serine and the P1 residue's carbonyl carbon (10). As a result of the P1-P1' cleavage, a "stressed-to-relaxed" conformational change occurs, allowing the RCL to insert into the central  $\beta$ -sheet of the serpin (39). This process translocates the protease over 70 angstroms to the opposite pole of the serpin, distorting the tertiary structure of the protease due to steric hindrance from

**Figure 3: A) Tertiary Structure of a Serpin ( $\alpha_1$ -PI M358R).** The general structure of serpins is highly conserved, consisting of three  $\beta$ -sheets (black) and eight to nine  $\alpha$ -helices (grey). The sequence of the reactive center loop (RCL) is poorly conserved as it helps to govern specificity towards a target protease, which cleaves between the P1 and P1' residues. The P side of the RCL is shown in light blue, the P' side in dark blue. Residues relevant to this study are shown as sticks (P7-P1, P1' for reference). Generated using PyMOL, PDB ID 1OPH (18). **B) The Serpin Inhibitory Mechanism.** A protease (E) recognizes the RCL (black) of a serpin (I) with a second order rate constant of  $k_1$ , forming an encounter complex (EI). The protease is able to disassociate with rate  $k_{-1}$ , but it is more likely to cleave the RCL to form an acyl-enzyme intermediate (EI') with rate  $k_2$ . Cleavage of the RCL allows it to insert as an additional strand in the central  $\beta$ -sheet, taking the protease along with it to the opposite pole of the serpin. The serpin mechanism then proceeds down one of two opposing paths. If the protease is able to complete the hydrolysis reaction ( $k_3$ ), cleaved serpin ( $E^*$ ) and free protease are formed. If the RCL completely inserts into the serpin body ( $k_4$ ) the protease is deformed against the base of the serpin and trapped, forming the serpin-enzyme complex ( $E-I^*$ ). The hydrolysis reaction can eventually complete ( $k_5$ ), although the complex is typically stable for days and is cleared before this occurs. This research was originally published in *The Journal of Biological Chemistry* (43). Copyright the American Society of Biochemistry and Molecular Biology, academic reuse is licensed.

**A**



**B**



the now proximal lower hemisphere of the serpin (10, 44, 45). Distortion of the protease's active site traps the covalent acyl-enzyme intermediate, resulting in a SDS-stable serpin-enzyme complex which can be visualized using electrophoresis.

The serpin mechanism is described as a branched pathway (**Figure 3B**), as the rate of RCL insertion into the central  $\beta$ -sheet is not instantaneous ( $k_4$ ), providing an opportunity for the protease to complete the hydrolysis reaction and escape inhibition ( $k_3$ ) (10, 46). Using the rates of the two branches, the efficiency of forming a stable serpin-enzyme complex can be defined as the stoichiometry of inhibition (SI), where  $SI = (k_3 + k_4)/k_4$  (10). This value provides the number of moles of serpin required to inhibit one mole of protease, thus a value around 1 is expected *in vivo* while significantly higher values are indicative of a poor rate of RCL insertion resulting in cleaved serpin and active protease (10). Following RCL insertion and the distortion of a protease's active site it remains possible for the protease to escape and regenerate its activity (10). However, the rate of protease release from the complex ( $k_5$ ) is at least 7 orders of magnitude slower than normal protease reactions once the serpin-enzyme complex is formed, and the complex is cleared before any significant release of protease can occur (36, 47). This specific suicide substrate mechanism, where both the inhibitor and target are consumed, makes serpins ideal for the regulation of tightly controlled pathways.

### **1.3.2 Importance within the Reactive Center Loop**

Despite the variable nature of the RCL, several structural and sequence homologies exist within it which provides clues as to the function of the RCL and its parent serpin. In all but two human serpins the RCL region N-terminal to the P1-P1'

cleavage site is 17 residues in length (26). The length of this RCL region, which is inserted into the central  $\beta$ -sheet, is critical for the correct function of serpins as it ensures the protease can be completely translocated and then distorted against the base of the serpin (48). Additional residues allow the protease to escape while the deletion of residues increases the steric strain leading to poor complex stability, particularly if the protease contains large loops proximal to the pole (i.e. thrombin) (48). The length of the RCL region C-terminal to the P1-P1' cleavage site is also significant, and has likely evolved to ensure the RCL is in an optimal position for inhibition of its target protease (49, 50).

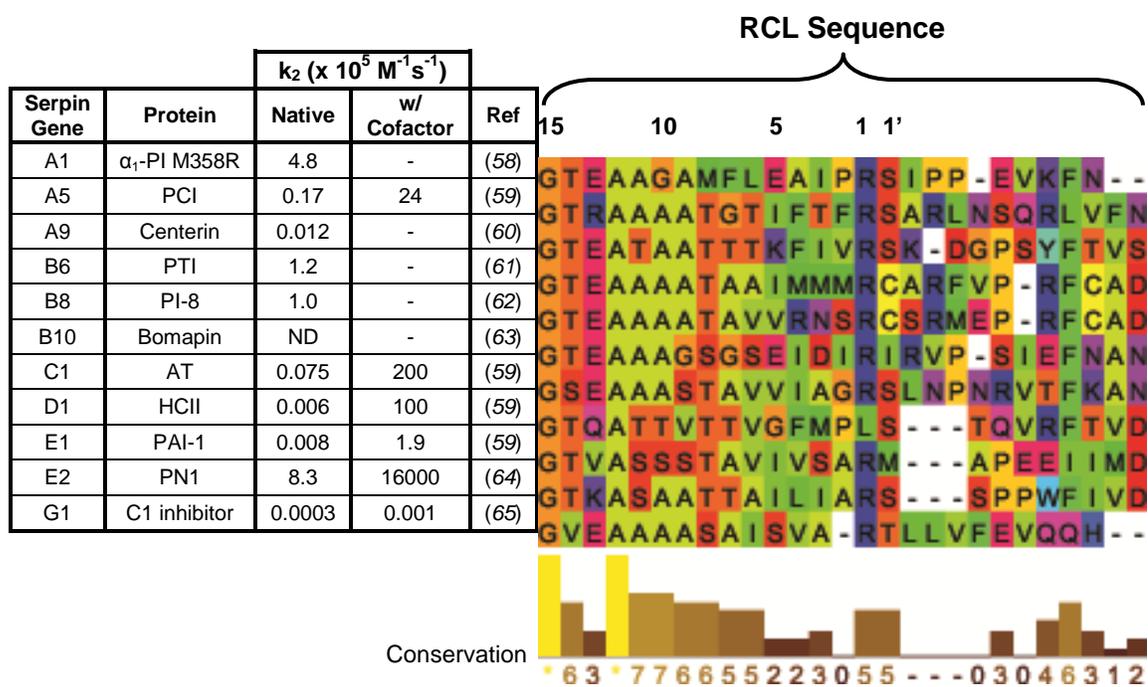
The hinge region, located from positions P15-P8, contains primarily alanine and other small uncharged residues, a pattern which is conserved in all inhibitory serpins but not in non-inhibitory serpins which are not required to undergo loop insertion (26). This homology is due to physical constraints as the hinge region must remain flexible to allow for rapid insertion of the RCL (51). Following RCL cleavage, the highly conserved P14 threonine residue begins the process of RCL insertion into the central serpin  $\beta$ -sheet (26). Gettins (2002) asserts that “the smaller the side chain that needs to be inserted, the smaller the cavity that needs to be created in the interior of the serpin body to accommodate it and so the lower the activation energy for that step is likely to be” (26). As each residue is inserted, the stability of the overall complex increases and assists the “zipper-like” addition of the next RCL residue into the  $\beta$ -sheet (26).

The RCL sequence flanking the P1-P1' site is hypervariable which reflects the evolutionary path of each serpin to have a specific target protease it is able to inhibit. The

identity of the P1 residue is a major factor in determining serpin specificity, as this residue must favourably insert into a protease's S1 pocket, leading to the cleavage of the P1-P1' bond (39). For example, AT has Arg at the P1 site making it an inhibitor of the Arg-specific proteases thrombin and FXa (52). A variety of mutagenesis experiments and natural mutations have revealed that the specificity of a serpin can be significantly influenced by changing only the P1 residue (**Appendix A**). However, 13 of the 27 human inhibitory serpins contain Arg at P1 yet their specificity varies dramatically, clearly indicating that the P1 residue is not the only specificity determinant (53). X-ray crystal structures of serpin-enzyme encounter complexes show that RCL residues in the P4-P3' region are inserted into the active site of the target protease, indicating that this region is also likely to influence specificity (10, 54). Alignments of RCL sequences from divergent serpin genes provide a useful way to visualize this balance between the structural constraints required for hinge flexibility,  $\beta$ -sheet insertion, and the residues required for specificity towards a target protease (**Figure 4**).

Despite the poor sequence conservation within the hypervariable region of the RCL, there are limits to the residues which can be tolerated. A mechanism for insertion of the RCL into the central  $\beta$ -sheet suggests that following P8, the RCL begins to displace helix F (26). As helix F overlaps the central  $\beta$ -sheet in the native conformation, the hypervariable region of the RCL will only continue to be inserted if the net energy change of insertion is sufficient to displace helix F (26). Once the final residue of the RCL is inserted, helix F acts as a "spring loaded ratchet" returning to its original position

**Figure 4: Human Serpins Capable of Inhibiting Thrombin.** Note the amino acid conservation within the hinge region (P15-P8), and hypervariability within the region flanking the P1-P1' cleavage site. Second order rate constants of thrombin inhibition ( $k_2$ ) of the native serpin or with a cofactor were retrieved from the noted references. “ND” not determined, “-” not determined or no increase in rate. Amino acid sequences were aligned using ClustalW 2.0 and displayed using Jalview 2.8 (55, 56). Residues coloured based on their physiochemical properties (57).



resulting in the distortion of the trapped protease (26). Mutations within the RCL support this mechanism, as the introduction of bulky or electrostatically unfavourable residues into the RCL increases the SI dramatically, presumably by interfering with  $\beta$ -sheet insertion (46, 66-68). The observation that evenly numbered hypervariable RCL residues (i.e. P6, P4, P2) are typically hydrophobic also coincides with this mechanism as this pattern mirrors general  $\beta$ -sheet structure, and crystal structures of RCL-inserted serpins indicate these residues are placed in hydrophobic pockets (26, 67, 69-74). Thus, although the hypervariable region is more diverse than the hinge region, both parts of the RCL are constrained by the functional requirement to insert properly into the body of the serpin. The specificity of a serpin towards a protease is a multifaceted interaction and is difficult to predict without an X-ray crystal structure of the inhibitory complex, of which there are now 10 available for human serpins: 2 covalent complexes and 8 non-covalent encounter complexes (10). Notably, the RCL is not the only determinant of serpin function as several encounter complex crystal structures indicate the importance of exosites beyond the RCL (16, 53).

### **1.3.3 Natural Thrombin Inhibitory Serpins**

The serpins which inhibit thrombin provide canonical examples of regions beyond the RCL which influence function, and the importance of cofactors in regulating serpin activity. AT is the primary thrombin inhibitory serpin and also has significant activity towards FIXa, FXa and FXIa, making it the “most important physiological inhibitor of the coagulation pathway” (59). AT primarily circulates in a latent form with its RCL partially buried within its central  $\beta$ -sheet (26). At the site of injury, heparan sulfate and

other GAGs expressed on the subendothelium come in contact with AT, leading to a conformational change which frees the RCL (26). However, the rate of thrombin inhibition is primarily increased due to the “bridging” GAGs create between exosite II of thrombin and a specific GAG exosite within the body of AT, bringing them in close proximity to one another and resulting in a >2400 fold increase in the rate of thrombin inhibition (26). AT also contains a unique 3-residue insertion on the P’ side of the RCL, which accommodates thrombin’s bulky 60 and 147-loops in order to form an optimal encounter complex (16).

HCII is the second most important serpin for regulating thrombin, estimated to contribute 20-30% of thrombin inhibition *in vivo* (59). HCII binds GAGs via an AT-like exosite, preferentially binding the GAG dermatan sulfate (26). Binding of a GAG cofactor drastically increases the rate of thrombin inhibition by as much as 17,000 fold, partially due to bridging interactions but mostly via a mechanism unique to HCII (26). The N-terminal region of HCII is a ~75 amino acid extension not found in any other serpin, which contains two hirudin-like repeats of acidic residues (75). When activated by a GAG, this acidic “tail” is released from stabilising interactions with the serpin body and instead associates with exosite I of thrombin. As HCII’s RCL contains leucine at P1, which is not expected to interact favourably with thrombin, this mechanism is crucial for its thrombin inhibitory activity and helps to explain the drastic increase in the rate following the acidic tail release (75). This interaction also reveals why HCII is extremely specific to thrombin and does not inhibit any other coagulation protease (59).

Protease nexin 1 (PN1) is the final serpin which significantly regulates thrombin, and has both the fastest native thrombin inhibitory rate and the fastest rate when in complex with heparin (26). Fully activated PN1 inhibits thrombin ~100 times faster than AT, and has activity towards FXa, FXIa, and fibrinolytic proteases (76). Paradoxically, PN1 is barely detectable in plasma and expression occurs primarily within soft tissue such as heart muscle (76). Due to PN1's dual coagulation and fibrinolytic regulatory abilities, overexpression of PN1 is suggested to be a regulatory mechanism to rebalance haemostasis in disease states (76).

A number of other human serpins are capable of forming a stable inhibitory complex with thrombin, although likely with little physiological relevance (**Figure 4**). Thrombin's substrate preferences are evident in their RCL sequences, with arginine found exclusively at P1, with the exception of HCII, and most contain the expected P1' serine. Other than the trend towards a hydrophobic P4 residue, any further structure-function relationship of an RCL granting thrombin inhibitory activity is difficult to observe.

#### **1.3.4 $\alpha_1$ -Proteinase Inhibitor**

The serpin  $\alpha_1$ -proteinase inhibitor ( $\alpha_1$ -PI), also called  $\alpha_1$ -antitrypsin, is the most abundant serpin in human plasma, constituting over 70% of serpins by weight (77). As  $\alpha_1$ -PI was the first serpin discovered, it serves as the archetype for the serpin mechanism and the diverse superfamily of serpins (78).  $\alpha_1$ -PI contains three glycosylation sites and appears in several isoforms *in vivo*, each approximately 52 kDa in size (79).

Glycosylation is not critical for the maintenance of  $\alpha_1$ -PI structure nor function, thus recombinant  $\alpha_1$ -PI with comparable activity can be produced using *E. coli*, resulting in a

protein with a size of 45 kDa (80).  $\alpha_1$ -PI primarily inhibits neutrophil elastase at the rapid rate of  $6.5 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$ , helping to balance the action of this protease as it degrades the extracellular matrix and elastin in the lungs (81, 82).  $\alpha_1$ -PI also has appreciable inhibitory activity towards the proteases: chymotrypsin, cathepsin G, trypsin, kallikrein 14, and Factor XIa (81, 83-86). A variety of naturally occurring mutations have been identified, the majority outside of the RCL, some of which lead to polymerisation of  $\alpha_1$ -PI in the liver leading to a deficiency in the circulation (82, 87, 88).  $\alpha_1$ -PI deficiency is believed to affect one in every 2000-5000 individuals, and can lead to liver damage due to the accumulation of polymerised  $\alpha_1$ -PI and emphysema due to unchecked elastase activity in the lungs (82). The pathogenesis of emphysema caused by smoking has also been linked to  $\alpha_1$ -PI, as the P1 methionine residue is susceptible to oxidation, leading to an imbalance of active  $\alpha_1$ -PI to elastase in the lungs (89).

The first indication that the function of  $\alpha_1$ -PI could be modified for a therapeutic use was discovered by chance. A mutation from methionine to arginine at the P1 position, referred to as  $\alpha_1$ -PI M358R or “Pittsburgh” mutant, was found to increase its ability to inhibit thrombin by 6,500 fold (90). This property was discovered as it resulted in a rare but fatal bleeding disorder with haemophilic-like symptoms due to the AT-like activity towards thrombin (91). However, unlike AT and other primary thrombin inhibitory serpins  $\alpha_1$ -PI M358R does not require a cofactor for optimal activity.  $\alpha_1$ -PI M358R inhibits thrombin 17 times faster than AT activated by pentasaccharide, a GAG which releases the AT RCL but does not contribute to bridging (54). Following the discovery of  $\alpha_1$ -PI M358R, an *in vitro* study revealed the mutation also increased the rate of inhibiting

the anticoagulant protease APC by over 4,400 fold (92). This anti-APC activity proved significant in a second patient identified with the M358R mutation, as the concentration of APC in their plasma was found to be significantly reduced (93). Unlike the first patient discovered to have the M358R mutation, the second patient did not suffer from fatal bleeding, believed to be due to the novel balance between the anticoagulant effect of inhibiting thrombin and the procoagulant effect of inhibiting APC (93). Additional *in vitro* studies have indicated that the M358R mutation also grants a significant enhancement in the rate of inhibiting coagulation factors IXa, Xa, XIa, and XIIa, along with kallikrein and the fibrinolytic protease plasmin (90, 94-97). The structure of thrombin's active site and its natural substrate preferences can help to explain the dramatic effect of the M358R mutation. The P2-P1' residues of  $\alpha_1$ -PI M358R (Pro-Arg-Ser), which are the most important for thrombin recognition, perfectly match the substrate preferences of thrombin (Section 1.2.2). The P4 alanine residue likely forms hydrophobic interactions in thrombin's S4 pocket, and alanine appears more frequently than any other amino acid at this position in identified substrates of thrombin (19). The relatively broad specificity of  $\alpha_1$ -PI M358R towards other coagulation proteases is due to the fact that they primarily favour arginine at the P1 position as well (17, 19).

Contemporary to the discovery of  $\alpha_1$ -PI M358R was a desire for a novel anticoagulant for the treatment of sepsis and reducing the associated risk of thrombosis. The fact that recombinant  $\alpha_1$ -PI M358R can be produced in large quantities, that it does not require a cofactor, and that it has broad activity towards coagulation proteases prompted it to be hailed as an "ideal agent" for fitting this demand (94, 98). Although

showing moderate success in a pig model of sepsis, when the use of  $\alpha_1$ -PI M358R was applied to a similar sepsis model in baboons it resulted in a counter intuitive collapse of haemostasis and worsening of septic shock (98, 99). The authors reasoned that the inhibition of cytoprotective APC overwhelmed the benefits of inhibiting thrombin and other coagulation factors, as APC activity remained below 40% following the injection of  $\alpha_1$ -PI M358R (98). The inhibition of  $\alpha_1$ -PI M358R's intended target, thrombin, would have further contributed to reducing APC levels due to thrombin's importance in the processing of the protein C zymogen. Thus, the broad specificity of  $\alpha_1$ -PI M358R originally touted as an asset served to work against its efficacy, resulting in a reduced interest for the use of this protein in the clinic. Attenuating or otherwise improving the specificity of  $\alpha_1$ -PI M358R remains a primary goal before clinical trials can be contemplated. It is important to note that the P2-P1' residues of  $\alpha_1$ -PI M358R fit the substrate specificity of thrombin purely by chance as this serpin did not evolve to inhibit thrombin. Additional mutations within the RCL or elsewhere within the protein may serve to improve its thrombin inhibitory activity while decreasing its activity towards other proteases.

## **1.4 Engineering $\alpha_1$ -Proteinase Inhibitor**

### **1.4.1 Primary Engineering Goals**

Due to the unique serpin mechanism and the abundance of pathologically relevant proteases, serpins have become important scaffolds for protein engineering.  $\alpha_1$ -PI continues to be one of the principal serpins of interest for engineering, primarily due to the relative ease of producing it using *E. coli* and yeast, and the fact that it does not need a

cofactor for optimal activity (80, 100, 101). The discovery of the M358R mutant and the role of  $\alpha_1$ -PI in emphysema coincided with the development of DNA manipulation and recombinant protein production, circumstances combining to spark general interest in engineering  $\alpha_1$ -PI. Prior to the work outlined in this study, 127 variants of human  $\alpha_1$ -PI have been engineered (**Appendix A**); PAI-1 is the only serpin to be mutated more thoroughly (102). The specific goals of engineering  $\alpha_1$ -PI, though diverse, can be grouped into four key categories: oxidation resistance; identifying structural and functional constraints; improving recombinant production; and modifying protease specificity. The latter category has been investigated in the most detail, which has concentrated the focus of mutagenesis within the RCL due to the importance of its sequence in governing serpin function. Knowledge of the serpin mechanism has been greatly enriched through the development of these variants, and in concert with modern mutagenesis and protein modeling techniques it is possible to envision designer serpins with specifically engineered function.

#### **1.4.2 Improving $\alpha_1$ -PI's Thrombin Inhibitory Function**

A total of 35 variant  $\alpha_1$ -PI M358R proteins have been created in an attempt to further the thrombin inhibitory activity of this protein, aiming also to improve its selectivity over other proteases (**Appendix A**). Hopkins *et al.* (1995) recognized that compared to natural thrombin inhibitory serpins,  $\alpha_1$ -PI M358R differed throughout the P7-P3' hypervariable region (**Figure 4**) (58). Although replacing either the P2 Pro or P1' Ser residue was detrimental to the rate of thrombin inhibition, substituting the P7-P3' sequence for that of AT's resulted in a >1000 fold increase in selectivity for thrombin

over APC (58). In addition to this, changing the P2 residue from AT's glycine back to  $\alpha_1$ -PI's proline increased the rate of thrombin inhibition to near that of  $\alpha_1$ -PI M358R, while keeping the rate of APC inhibition >100 fold lower (58). Hopkins *et al.* (2000) investigated additional AT substitutions, progressively replaced each residue in the P7-P3' region to determine the effects on the rate of thrombin, APC, and FXa inhibition (103). Replacing the P2 and P3' residues of  $\alpha_1$ -PI with those of AT resulted in a significant loss of APC inhibition (103). However, the  $\alpha_1$ -PI mutant with the best ratio of thrombin to APC inhibition had AT's P7-P2' sequence with the exception of P2, again indicating that this proline was important and suggesting that AT's P3' asparagine residue was not favoured (103). This mutant, termed "RCL 5" by Sutherland *et al.* (2007), has a comparable rate of thrombin inhibition to M358R but with a 26–168 fold decrease in the rate of APC inhibition, resulting in an impressive thrombin to APC ratio of 2526–3221 to one (103, 104).

Taking additional cues from nature, Sutherland *et al.* (2006) fused HCII's unique 75 residue N-terminal acidic tail to the N-terminus of  $\alpha_1$ -PI M385R (105). This mutant, dubbed HAPI M358R, has the fastest engineered rate of thrombin inhibition yet with a second order rate constant of  $3.83 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ , which is more rapid than any natural thrombin inhibitory serpin not maximally activated with a glycosaminoglycan (26, 59). Seeking to further improve this mutant by engineering greater thrombin specificity, Sutherland *et al.* (2007) combined previous successes to generate the "HAPI RCL5" mutant;  $\alpha_1$ -PI with the N-terminal HCII tail and the "RCL 5" sequence (104). Although the rate of thrombin inhibition was reduced by ~40% compared to HAPI M358R, the rate

of APC inhibition collapsed by over 138 fold providing the best thrombin to APC ratio yet of 10,769 to one (104). HAPI RCL5 was also found to have a lower rate of FXa, FXIa, and FXIIa inhibition, and hence an improved selectivity for thrombin in general (97). Recently, animal model work has solidified the *in vitro* evidence of HAPI RCL5's promising function, as the protein proved effective at reducing clot size, reducing fibrin deposition during sepsis, and lengthening occlusion time after injury (97).

Thus far, engineered  $\alpha_1$ -PI mutants with improved thrombin inhibitory activity have entirely been based on RCLs and exosites from natural thrombin inhibitory serpins. As these borrowed sequences did not evolve to interact favourably with the rest of the  $\alpha_1$ -PI protein it is unlikely that they can provide the optimal inhibitory activity. Cooperative interactions with other portions of  $\alpha_1$ -PI and subtle differences within the serpin body may also contribute to how thrombin interacts with and is inhibited by  $\alpha_1$ -PI, implying the optimal  $\alpha_1$ -PI RCL for inhibiting thrombin remains unknown.

### **1.4.3 Serpin Engineering Challenges**

Few examples of engineering serpins have been met with as much success as improving thrombin inhibitory function. As with engineering other proteins, introducing novel function or tuning existing activity is difficult. Despite the best thought out additions and mutations, proteins rarely behave as expected and engineering serpins is further compounded by the complex nature of the serpin inhibitory mechanism (106, 107). Residues within the RCL must favourably interact with both the target protease and the central serpin  $\beta$ -sheet in order for the serpin to act as a functional inhibitor. Even if the RCL is cleaved at a rapid rate, poor RCL insertion will hinder the function of a serpin.

Modification of a serpin's wild type sequence often leads to a collapse in inhibitory activity or serpin-enzyme complex stability. This is despite the recognition of the RCL sequence as a critical determinant of serpin function and therefore an ideal target for mutagenesis. Several examples highlight that simply replacing the RCL with that of another serpin does not guarantee intended activity. For example, two groups have replaced the RCL of the serpin  $\alpha_1$ -antichymotrypsin with various RCL's from other serpins (108, 109). Both groups observed that despite granting modest improvements in activity towards new proteases, the activity was only a fraction of the original serpins and the inherent specificity was poorly retained. RCL-protease interactions are therefore only part of the picture, and despite the highly conserved structure of the serpin body it too plays a role in determining function.

Attempting to rationally introduce RCL mutations also presents challenges. Seeking to inhibit the protease furin, Dufour and colleagues have published several papers in an effort to direct  $\alpha_1$ -PI or  $\alpha_1$ -antichymotrypsin activity towards this protease (67, 68). Recognizing that furin preferentially cleaves peptides with arginine at P6, P4, P2, and P1 they modified the RCL's of these serpins to contain various combinations of arginine. Unfortunately, the more arginines that were introduced the greater the SI was increased, indicative of a poor rate of  $\beta$ -sheet insertion, negating the introduced anti-furin activity (67, 68). This was particularly evident when replacing  $\alpha_1$ -PI's P6-Leu residue, which they reasoned was due to the hydrophobic pocket this residue inserts in within the  $\beta$ -sheet, providing an ideal example of the structural restraints of the serpin mechanism (67). They also noted that residues interacted cooperatively, as replacing both  $\alpha_1$ -PI's P6

and P2 residues lead to a nonadditive decrease in the inhibitory rate compared to those mutations on their own (67).

Engineering serpins must take into account the balance between RCL-protease interactions, the structural requirements for favourable loop insertion, and interactions unique to each serpin scaffold. Saturation mutagenesis throughout the RCL and testing each possible combination of amino acids would reveal the changes required to meet this balance. As Antalis & Lawrence (2004) note, the challenge facing this approach “lies not with the generation of mutants, but with determining the functional significance of the mutants generated” (107). Introducing codon degeneracy at only 2 RCL positions generates 441 proteins (including stop codons) with unknown function and each additional degenerate position increases the number by 21 fold. A method for screening a large number of serpin mutants is required in order to engineer optimal activity.

## **1.5 Applying Phage Display to Engineering Serpins**

### **1.5.1 Overview of Phage Display Techniques**

Phage display has been in use for over 25 years, beginning with the first genetic fusion of a protein fragment to the coat protein of f1 phage (110). The general technique employs an engineered bacteriophage to express a protein or peptide sequence of interest fused to the coat of a phage particle. Only those phage expressing proteins capable of binding an immobilized target will remain to be amplified in *E. coli* after washing away unbound phage, a process referred to as biopanning. Following amplification, up to  $10^{11}$  plaque forming units per mL (pfu/mL) can be generated, meaning an extraordinarily large number of mutants can be screened rapidly (111). However, libraries are typically

constrained to  $10^8$  unique sequences in order to ensure each sequence in the library is represented multiple times during biopanning, and the fact that bias in the sequences can develop after multiple rounds of amplification (112). Bias can be introduced due to differences in growth rate, nonspecific binding, or for undefined reasons (40, 112-114). Improved sequencing technology has helped to address this issue, by identifying insert sequences with poor growth but high affinity to the target, which would normally be hidden by “parasite sequences” which grow rapidly but do not bind to the target (115).

Several types of bacteriophage have been adapted for use in phage display, with the choice between them based on the design and aim of the study. Filamentous bacteriophage, such as f1, fd and M13, were the first to be developed for use in phage display due to the relative ease of manipulating their genes (116). M13 phage has since been used extensively due to the availability of commercial M13 peptide libraries and the development of high-diversity M13 antibody libraries (113, 117, 118). Inserted genes or peptides are typically fused to the N-terminus of minor coat protein 3 (p3), though several additional coat proteins have been modified to allow inserts in either N-terminal or C-terminal orientation (117). The number of times the inserted sequence is expressed by M13 phage, or the “copy number”, ranges from hundreds of copies of small peptides, to only one copy for proteins >100 amino acids (117). As filamentous phage are secreted from host *E.coli* during replication, any sequence fused to a coat protein must also be able to pass through the membrane or they will not be assembled to the rest of the phage (116).

Lytic species of phage, such as T7, T4 and  $\lambda$ , are not limited by inserted sequences which cannot pass the bacterial membrane, as lytic phage assembly occurs completely

within the infected cell before daughter phage are released by cell lysis (116). Thus, lytic phage display systems are typically used in order to express full sized proteins rather than comparatively small peptides on the surface of phage (116). If samples of phage lysate must remain free of cytoplasmic contents, or if the inserted protein contains many disulphide bonds which require *E. coli* chaperone proteins, filamentous phage display techniques would be more appropriate (119, 120). Contrasting filamentous phage, genes or peptides are typically fused to the C-terminus of lytic phage minor coat proteins, however N-terminal fusions for several lytic phage species have also been developed (116). As with filamentous phage, the copy number of the inserted sequence expressed by lytic phage depends on the size of the insert, though several commercially available kits have been developed to better regulate expression levels (116).

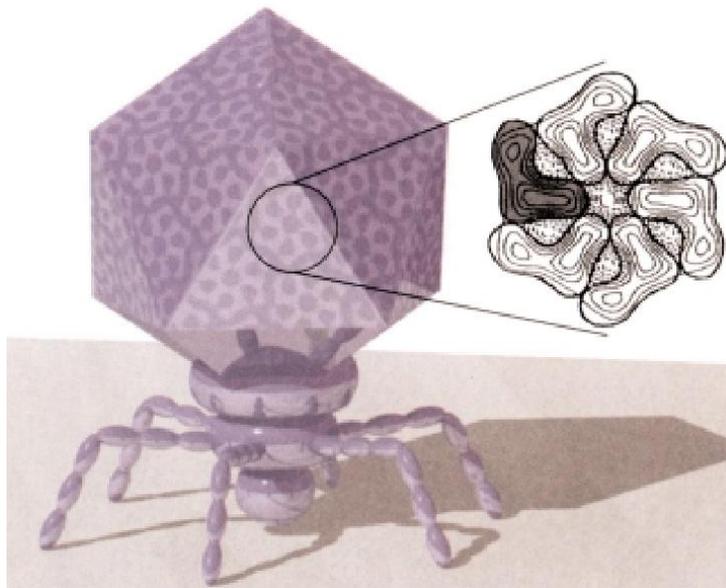
The T7Select phage display system developed by Novagen has three phage display vectors available which differ in the copy number and size limit of the insert. The T7Select10-3b vector provides the best of both options, allowing for inserts as large as 1200 amino acids displayed with a mid range copy number of 5-15 proteins per phage, although in practice copy number can be lower (121, 122). Additional benefits of T7Select, as reported by the manufacturer, include more rapid replication than other phage species and that T7Select phage are more robust, allowing for stringent biopanning conditions (121). An independent study also indicated that T7Select libraries contained a lower sequence bias than comparable M13 libraries, likely due to the differences between lytic and filamentous phage reproduction (123). The T7Select vectors contain a multiple cloning site following amino acid 348 of the 10B minor coat protein, resulting in the

fusion of the N-terminus of an inserted sequence to the C-terminus of the coat protein (**Figure 5**) (121). To assist with T7Select replication, *E. coli* strains provided with the commercial kit have been specifically selected, including the addition of a plasmid into strains BLT5403 and BLT5615 which allows for inducible expression of coat protein 10A to help with virion assembly (121). This mirrors the use of “helper phage” to provide the necessary phage proteins required for reproduction of many filamentous phage species, although T7Select is able to replicate without extra 10A coat protein albeit at a slower rate (116, 121). The ability to accommodate inserts up to 1200 amino acids, and without being limited by having to cross the bacterial membrane, opens the door to express the majority of human proteins, including serpins, using T7Select phage.

### 1.5.2 PAI-1

The benefits of phage display have yet to be fully exploited to screen serpin mutants, as only plasminogen activator inhibitor type 1 (PAI-1) has been produced using phage display. The selection of PAI-1 for phage display was originally based on its successful production using *E. coli*, that X-ray crystal structures were available, and to better understand the unique latent-to-active conformational change PAI-1 undergoes *in vivo* (124, 125). Pannekoek *et al.* (1993) were the first to successfully express PAI-1 using phage display, using the pComb3 phagemid vector and M13 helper phage to assist in viral assembly (124). This method expressed PAI-1 as a fusion protein, with the pComb3 phagemid coding for an immunoglobulin A QVVKL sequence N-terminal to the mature PAI-1 sequence, followed by a GGGGS linker, and finally the C-terminal half of the p3 M13 minor coat protein. With the addition of M13 helper phage, M13 virions with

**Figure 5: Structure of T7Select Phage and DNA Map of T7Select Vector Arms.** Coat protein monomers to which the insert is fused are shown in grey. Primers used in this study were designed to ensure serpin inserts were fused to the C terminus of the 10B minor coat protein in the indicated orientation. Courtesy of Novagen (121, 126). Reprinted with written permission of Novagen Inc., Madison, WI, provided on July 16, 2013.



**gene 10B →**  
 left arm...GATCCG      AATTXXXX(N)XXXX      AGCTT...right arm  
 left arm...CTAGGCTTAA      XXXX(N)XXXXTCGA      A...right arm  
    ...AsnPro      AsnSer...

a single copy of PAI-1 expressed on their surface were secreted from *E. coli* (124). The authors' reason for including the N-terminal immunoglobulin sequence was based on the previous success of expressing antibodies using M13 phage (127). Fusing the C-terminus of PAI-1 to the cpIII protein was an unintuitive choice, as once the RCL is cleaved the C-terminal region is no longer covalently bonded to the rest of the serpin, creating the potential for PAI-1 to disassociate from the phage (128). However, the M13-expressed PAI-1 retained inhibitory activity and remained associated with the phage virions, based on the number of phage recovered after incubating with its target protease tPA, and biopanned using wells coated with anti-tPA mAB (124). Further supportive of retained inhibitory function, was the finding that the activity of tPA was inhibited in a linear relationship to the amount of PAI-1 phage in solution (124). A mutant library was then generated using error-prone PCR which introduced 1-6 mutations at random positions within the full PAI-1 sequence, producing a total library size of  $1.7 \times 10^6$  distinct mutants (124). The authors noted that the use of error-prone PCR to construct the library "restricts the range of amino acid alterations" and that "only a limited set of amino acid substitutions will be encountered" (129). Stoop *et al.* (2000) modestly improved the overall diversity of the M13-expressed PAI-1 phage library by using DNA shuffling, which introduced mutations via both error-prone PCR and homologous recombination (130). This generated a PAI-1 library of  $4 \times 10^7$  individual mutants which was then biopanned to map sites of interaction with anti-PAI-1 mAbs, and to identify mutations which stabilised PAI-1 in its active conformation (130, 131).

Berkenpas *et al.* (1995) used an entirely different phage display technique in order to create a library of PAI-1 mutants (125). Error-prone PCR was again used to generate random mutations throughout the entire PAI-1 sequence, however rather than producing PAI-1 fused to the coat of the bacteriophage,  $\lambda$ -phage was used to introduce the DNA library into *E. coli* for recombinant expression (125). This allowed the PAI-1 mutants to be expressed in their native state, not fused to bacteriophage, which likely provided a more accurate model for the serpin's *in vivo* activity (107). Due to the instability of natural PAI-1, which is found in both an active and latent (inactive) conformation, Berkenpas *et al.* (1995) aimed to find a mutant within the library in a stable active conformation (125). The library was biopanned by incubating lawns of infected *E. coli* overnight at 37°C, then probing for PAI-1 mutants which retained their inhibitory activity using tPA-coated filters (125). This led to the discovery of a hyperstable mutant containing four mutations, which has since been used extensively to better understand PAI-1 structure and function (102, 128). Only the four mutations in combination enhanced stability, indicating that without the use of phage display it is unlikely that this mutant would have been discovered (125).

These two examples of expressing PAI-1 using phage display highlight the fact that even proteins such as serpins which undergo drastic conformational changes remain functional when expressed using bacteriophage. However, the methodology employed in these examples pales in comparison to more modern protein library construction and phage display techniques. The use of error-prone PCR limited the scope of the studies, as mutations were introduced with random frequency and at random positions, leaving

sequence space poorly defined. PAI-1 contains 349 amino acids, meaning of the  $20^{349}$  possible amino acid combinations created using error-prone PCR even the largest PAI-1 phage library ( $4 \times 10^7$ ) contained only a fraction (130). In the case of  $\lambda$ -phage, this technique is only able to screen an even smaller subset of the mutant library, as it is limited to the number of colonies per lawn of bacteria which can be screened at one time (10,000 - 100,000) (107). Zani & Moreau (2010) suggest that more modern techniques, such as the use of T7 phage, could be used to address these issues in order to extend the use of phage display to other serpin proteins (132).

### **1.5.3 Modeling Protease Specificity**

Due to the nature of the serpin inhibitory mechanism phage display can also be used to engineer serpin function without expressing the entire serpin protein. The RCL of a serpin must interact favourably with the active site of a target protease, thus the hypervariable region proximal to the P1-P1' cleavage site typically mimics a protease's natural substrate (26). Phage display is used extensively to elucidate protease specificity as it can rapidly generate a wide range of potential substrates (133). This "substrate phage" technique utilizes a library of peptides expressed on the surface of phage with an affinity domain, such as a His-tag, terminal to the degenerate sequence (40, 134). The phage library is immobilized via the affinity domain, such as to nickel resin, before being exposed to the protease of interest (40). Phage expressing proteolysed peptides are then eluted and sequenced to provide the protease's specificity profile (40). Recently, this technique was used to elucidate the extended substrate specificity of thrombin, indicating a level of specificity as far as the P4 and P4' positions (21).

Several studies have extended the use of substrate phage to predict serpin-protease interactions and to help engineer designer serpins (40, 86, 135, 136). Deperthes and colleagues pioneered the use of substrate phage to modify serpin function, with the goal of inhibiting human kallikrein proteases (86, 135). Peptides susceptible to kallikrein 2 or kallikrein 14 cleavage were first identified by substrate phage before being inserted into the RCL of two scaffold serpins: human  $\alpha_1$ -antichymotrypsin and  $\alpha_1$ -PI (86, 135). Compared to the wild type sequences, the mutations successfully introduced modest kallikrein 2 inhibitory activity and doubled the rate of kallikrein 14 inhibition (86, 135). The rate of inhibiting homologous proteases was also decreased, indicative of improved specificity (86, 135). Notably, the choice of serpin scaffold affected the rate and stability of inhibitory complex formation, with the  $\alpha_1$ -antichymotrypsin mutants being better in both respects than the  $\alpha_1$ -PI mutants (86). This provides further evidence that residues or structural constraints beyond the RCL play a role in determining activity.

Despite these successes, designing serpins based on protease specificity alone is not a reliable method for ensuring optimal activity. Each of the engineered kallikrein 2 inhibitory serpins had a high SI ranging from 9 to 139, indicating poor complex stability leading to the majority of serpin protein being cleaved and inactivated (135). This is mirrored by the previously discussed studies which aimed to engineer inhibitory activity towards the protease furin, using  $\alpha_1$ -PI as a scaffold. Though the rate of inhibition was increased by mutating the RCL to reflect furin's known substrate preference again the SI was increased significantly (67, 68). Song *et al.* (2011) rightly caution that "cleavage within the RCL region might not result in the inhibition of the participating protease"

(40). As summarised in previous sections, other determinants of serpin function such as exosites and the speed of RCL insertion affect the specificity and function of a serpin. Despite having an RCL that is specific to a target protease, designer serpin RCLs may be unable to undergo  $\beta$ -sheet insertion due to bulky or electrostatically unfavourable mutant residues. Although phage display is useful in determining a protease's specificity, it cannot identify the optimal RCL sequence required for inhibition unless the entire serpin protein can be expressed and functionally screened.

## 1.6 Rationale for Study

The thrombin inhibitory activity that  $\alpha_1$ -PI M358R exhibits makes it a promising candidate for further engineering, with the goal of creating a potent and specific anticoagulant protein drug. Improving  $\alpha_1$ -PI M358R's thrombin inhibitory activity is possible, exemplified by previous projects in the Sheffield lab, with two caveats:  $\alpha_1$ -PI M358R did not evolve to inhibit thrombin; and improvements in its thrombin inhibition have solely been based on borrowing sequences from natural thrombin inhibitory serpins. The hypervariable region of the RCL of a serpin plays a large role in determining inhibitory activity, and must interact favourably with both the targeted protease and the unique body of the serpin protein once the RCL is inserted. The optimal engineering method would therefore consist of producing every possible combination of hypervariable RCL sequences in conjunction with the entire  $\alpha_1$ -PI protein, and functionally screening for thrombin inhibition. Out of the millions of such  $\alpha_1$ -PI RCL sequences that theoretically exist, only 35 mutants have been published with the goal of improving thrombin inhibitory activity. Phage display offers a way to express and screen large

numbers of proteins, using an *in vitro* evolution approach where only the sequences with the desired function are amplified. The M13 and  $\lambda$  phage display systems have been used to express and screen the mutants of one serpin, PAI-1. The reason no other serpins have been expressed using phage display remains unknown. For this study, the T7Select phage display system was selected as it can express large inserts, it is commercially available, and has fewer constraints on the type of protein fused to its coat. Additionally, N-terminal fusions of sequences to  $\alpha_1$ -PI do not hinder inhibitory function (104, 105, 137). My rationale for this study was therefore to improve  $\alpha_1$ -PI engineering methodology by applying modern phage display techniques, and to discover the optimal  $\alpha_1$ -PI hypervariable RCL sequence for inhibiting thrombin, which may reveal novel aspects of the serpin's inhibitory mechanism relevant to the successful extension of  $\alpha_1$ -PI to the clinic.

## 1.7 Hypotheses

1. Randomization of the P2 and P1 residue within the RCL of  $\alpha_1$ -PI will allow for an initial study of the optimal residues at these positions for thrombin inhibitory activity;
2. Phage display of  $\alpha_1$ -PI will prove to be an efficient means by which to screen for thrombin inhibitory activity. Applying the concepts of *in vitro* evolution, only the phage expressing  $\alpha_1$ -PI mutants which effectively bind thrombin will be amplified;
3. Due to the known efficacy of the  $\alpha_1$ -PI M358R mutant in binding thrombin, and that its P2P1 sequence already fits thrombin's known specificity,  $\alpha_1$ -PI M358R will be the dominant mutant after screening the P2P1 library;
4. As the entire  $\alpha_1$ -PI protein will be expressed by phage display, a larger library will lead to the discovery of the optimal residues required both for thrombin specificity and rapid RCL insertion. By maintaining the P2 and P1 positions as the optimal residues, but randomizing the P7-P3 region, the majority of the hypervariable RCL region will be screened for the first time to reveal additional aspects of the structure-function relationship of the RCL;
5.  $\alpha_1$ -PI P7-P3 mutants which appear after multiple rounds of screening will have an improved rate of thrombin inhibition and a lower stoichiometry of inhibition (SI) than  $\alpha_1$ -PI M358R, as these mutants will have been evolved *in vitro* for improved thrombin inhibitory activity.

## 1.8 Project Outline

The aim of this project was to apply the versatility of phage display to screen mutants of  $\alpha_1$ -PI which effectively inhibit thrombin. By leveraging previous work in expressing serpins using phage display while applying newer phage display techniques, this project was thought to offer a rapid way to probe the function of  $\alpha_1$ -PI and improve its thrombin inhibitory activity. My objectives were as follows:

1. Produce purified PAI-1 (His-tagged) using *E. coli*, for comparison to phage-produced PAI-1;
2. Express PAI-1 on T7Select10-3b phage based on published results expressing PAI-1 on M13 phage;
3. Express  $\alpha_1$ -PI M358R on T7Select10-3b phage based on the putative success of expressing PAI-1 on T7Select10-3b. Attempt to detect SDS-stable phage-thrombin complexes;
4. Mock-bioplan the T7Select10-3b  $\alpha_1$ -PI M358R phage mixed with a negative control phage in order to assess the effectiveness of a novel biopanning protocol. Success of the protocol was planned to be determined based on enrichment of T7Select10-3b  $\alpha_1$ -PI M358R phage after 5 rounds of biopanning with thrombin;
5. Express a library of  $\alpha_1$ -PI mutants with degenerate codons at the P2 and P1 (amino acids 357 and 358) positions on T7Select10-3b phage, based on the putative success of mock-biopanning T7Select10-3b  $\alpha_1$ -PI M358R. Screen the library for mutants which bind to thrombin effectively by biopanning for 5 rounds with thrombin, and comparing to results of same library biopanned without thrombin;

6. Isolate T7Select10-3b phage which express  $\alpha_1$ -PI P2P1 mutants of interest and perform sequence analysis to determine the identity of the mutations. Insert His-tagged  $\alpha_1$ -PI mutants of interest into individual pBAD plasmids and purify the products. Quantitatively test thrombin binding/inhibitory activity compared to  $\alpha_1$ -PI M358R using kinetic methods;
7. Express a library of  $\alpha_1$ -PI mutants with degenerate codons at the P7 to P3 (amino acids 352 to 356) positions on T7Select10-3b phage, maintaining the optimal P2 and P1 residues putatively determined by the P2P1 library. The P7-P3 library screening was planned to use the same methods as for the P2P1 library;
8. Isolate the T7Select10-3b phage which express  $\alpha_1$ -PI P7-P3 mutants of interest and perform sequence analysis to determine the identity of the mutations. Insert His-tagged  $\alpha_1$ -PI mutants of interest into individual pBAD plasmids and purify the products. Quantitatively test thrombin binding/inhibitory activity compared to  $\alpha_1$ -PI M358R using kinetic methods;
9. Apply computer assisted molecular modeling to help determine the  $\alpha_1$ -PI M358R and novel  $\alpha_1$ -PI mutants' interactions within the active site of thrombin.

## 2. MATERIALS

### 2.1 Source of Chemicals and Reagents

Chemicals and biological reagents were purchased from the following suppliers: New England Biolabs (Pickering, ON) Phusion DNA polymerase, calf intestine alkaline phosphatase; Thermo Fisher Scientific (Burlington, ON) T4 DNA Ligase, DNA restriction enzymes (BspHI, BamHI, NcoI, EcoRI, HindIII, PmlI (Eco72I), SauI (Bsu36I)), guanidine hydrochloride, Coomassie Plus Protein Assay Reagent, TMB developing solution, NanoDrop Spectrophotometer; Origene (Burlington, ON) human PAI-1 cDNA; Qiagen (Carlsbad, CA) QIAquick PCR purification kit, QIAprep Spin Miniprep kit, QIAquick Gel Extraction kit, Plasmid Midi kit, HotStarTaq Plus Master Mix, nickel-nitrilotriacetic acid (Ni-NTA) resin, HotStar HiFidelity DNA polymerase; EMD Millipore (Billerica, MA) pET-11d *E. coli* plasmid, BL21(DE3) *E. coli*, Amicon Ultra MWCO 10000 centrifugation filter; Promega (Madison, WI) HB101 *E. coli*, AP-conjugated goat anti-mouse antibody; Invitrogen (La Jolla, CA) DH5 $\alpha$  *E. coli*, iBlot Gel Transfer Stacks, iBlot machine, rabbit anti-human PAI-1 antibody, DSB-X Biotin Protein Labeling kit, Qdot 625 streptavidin conjugate, pUC19 *E. coli* plasmid, streptavidin-coated FlowComp Dynabeads, Top10 *E. coli*; Bioshop (Burlington, ON) ampicillin, bovine serum albumin (BSA), electrophoresis grade agarose, Tween 20; GE Healthcare (Mississauga, ON) mouse anti-His tag antibody, SP sepharose Fast Flow, DEAE sepharose Fast Flow; Roche (Mississauga, ON) nitro blue tetrazolium/5-bromo-4chloro-3-indolyl phosphate (NBT/BCIP); Gibco (Grand Island, NY) Tween 80; BDH (Toronto, ON)  $\beta$ -mercaptoethanol, Triton X-100; Bio-Rad (Mississauga, ON) nitrocellulose paper;

Genentech (San Francisco, CA) human tPA; Novagen (Madison, WI) T7Select10-3b Vector Arms, T7Select Packaging Extract, T7Select 10-3 Cloning kit, T7Select Control Insert, BLT5615 *E. coli*, provided T7 Elution Buffer; Affinity Biologicals (Ancaster, ON) HRP-conjugated sheep anti-human  $\alpha_1$ -PI antibody; Cedarlane (Burlington, ON) AP-conjugated goat anti-rabbit antibody; Abcam (Toronto, ON) mouse anti-human tPA antibody; Enzyme Research Laboratories (South Bend, IN) human thrombin; Calbiochem (La Jolla, CA) phenylalanyl-L-prolyl-arginine chloromethyl ketone (PPACK); Diapharma (West Chester, OH) chromogenic substrates S2238 (H-D-Phenylalanyl-L-pipecolyl-L-arginine-p-nitroaniline-dihydrochloride). All other chemicals and biological reagents not specifically listed above were of the highest quality and grade available.

## **2.2 Oligonucleotides**

The oligonucleotides listed in **Table 1** were synthesized at the McMaster Institute for Molecular Biology and Biotechnology (MOBIX) at McMaster University (Hamilton, ON). All confirmatory DNA sequencing was also performed at MOBIX.

## **2.3 Computer Software**

DNA sequences were analysed with Clone Manager 7.11 and Align Plus 5.11 (Sci-Ed Software, Cary, NC). Gels and phage plaques were quantified with Quantity One 4.6.7 (Bio-Rad, Mississauga, ON). Amino acid sequences were aligned with ClustalW 2.0 and displayed using Jalview 2.8 (55, 56). Amino acid conservation was quantified with ConSurf (138). Protein crystal structures were manipulated and imaged with PyMOL 1.3r1 (Schrödinger LLC, Cambridge, MA). Monte Carlo minimizations of crystal structures was performed using ZMM (ZMM Software Inc., Hamilton, ON).

**Table 1: List of Oligonucleotide Primers Used During this Project.** Note that primer APIP2P1dgnAS modified the P361 codon from CCC to CCA, and APIP7P3dgnAS modified the same codon to CCG. These silent mutations were introduced in order to differentiate the two libraries from each other, and from the original unmodified DNA.

<b>Primer Name</b>	<b>Description</b>	<b>Primer Sequence (5' - 3')</b>	<b>#</b>
H6PAI1p11S	Forward primer for addition of His-tag and insertion of PAI-1 into pET-11d vector	AGTCATGAGCCATCATCATC ATCATCATGTGCACCATCCAC CATCCTAC	3310
H6PAI1p11AS	Reverse primer for insertion of PAI-1 into pET-11d vector	GCTGGATCCTCAGGGTTCCA TCACTTGGCCCATGAA	3311
PAI1T710bS	Forward primer for insertion of PAI-1 into T7Select10-3b phage	GATCCGAATTCAGTGCACCAT CCACCATCCTAC	4120
PAI1T710bAS	Reverse primer for insertion of PAI-1 into T7Select10-3b phage	GCTAAGCTTCAGGGTTCCATC ACTTGGCCCATGAA	4121
M358RT7S	Forward primer for insertion of $\alpha_1$ -PI M358R into T7Select10-3b phage	GATCCGAATTCAGAGGATCC CCAGGGAGATGCT	5918
M358RT7AS	Reverse primer for insertion of $\alpha_1$ -PI M358R into T7Select10-3b phage	GCTAAGCTTCATTTTTGGGTG GGATTCACCAC	5919
LP-PmII	Forward primer for generation of $\alpha_1$ -PI libraries	CAAGGACACCGAGGAAGAGG ACTT	4980
APIP2P1dgnAS	Reverse primer for generation of $\alpha_1$ -PI P2P1 library	CTTGACCTCAGGTGGGATAG ANNNNNNTATGGCCTCTAAAA AC	5968
APIP7P3dgnAS	Reverse primer for generation of $\alpha_1$ -PI M358R P7-P3 library	TGACCTCAGGCGGGATAGAT CTGGNNNNNNNNNNNNNNNN CATGGCCCCAGCAGCTTC	0201
pBADrev	Reverse primer for sequencing pBAD constructs	AAATTCTGTTTTATCAGACC	6612

### **3. METHODS**

#### **3.1 Cloning H<sub>6</sub>PAI-1 into pET-11d Vector**

PCR was performed using Phusion polymerase and primers H6PAI1p11S and H6PAI1p11AS with cDNA coding for PAI-1 as the template. These primers added a 5' BspHI restriction site followed by a His-tag (6x histidine codons) and a 3' BamHI restriction site. The 1172 bp PCR product was visualized on a 1% agarose gel before being purified using the QIAquick PCR purification kit. The pET-11d *E. coli* plasmid vector was subjected to mini-plasmid DNA extraction and purification using the QIAprep Spin Miniprep kit in order to isolate it from the HB101 host *E. coli* strain. The purified H<sub>6</sub>PAI-1 PCR product was digested with BspHI and BamHI, and the purified pET-11d plasmid was digested with NcoI (site compatible with BspHI) and BamHI. The fragments were visualized on a 1% agarose gel and gel purified using the QIAquick Gel Extraction kit. The digested H<sub>6</sub>PAI-1 gel extraction product was ligated into the pET-11d plasmid using T4 DNA Ligase and transformed into DH5 $\alpha$  competent *E. coli* cells. DH5 $\alpha$  cells containing the pET-11d H<sub>6</sub>PAI-1 plasmid were subjected to mini-plasmid DNA extraction and purification using the QIAprep Spin Miniprep Kit, before being sent to MOBIX for sequencing with primers H6PAI1ps11s and H6PAI1p11AS. The pET-11d H<sub>6</sub>PAI-1 plasmid was then transformed into BL21(DE3) *E. coli* cells for expression.

#### **3.2 Expression of H<sub>6</sub>PAI-1 and Purification from *E. coli* Inclusion Bodies**

##### **3.2.1 Culture Conditions for Expression of H<sub>6</sub>PAI-1**

50 mL LB containing 100  $\mu$ g/mL ampicillin was inoculated with 1/50 volume of BL21(DE3) pET-11d H<sub>6</sub>PAI-1 overnight culture, and incubated on a shaker at 37°C until

OD<sub>600</sub> = 0.5. Expression of H<sub>6</sub>PAI-1 was induced via the addition of IPTG to 1 mM; cultures were incubated for 6 hours on a shaker at 37°C.

### **3.2.2 Confirmation of H<sub>6</sub>PAI-1 Expression by Immunoblot**

SDS-PAGE gel electrophoresis on BL21(DE3) pET-11d H<sub>6</sub>PAI-1 samples was performed. Resulting gels were transferred to nitrocellulose using iBlot Gel Transfer Stacks and an iBlot machine as per the manufacturer's instructions. The membrane was incubated for 1.5 hours in 5% milk/TBST with rocking. Mouse anti-His antibody was diluted 1/2000 in 10 mL 5% milk/TBST and added to the nitrocellulose membrane, and incubated for 1.5 hours with rocking. The membrane was washed 3 times in TBST for 10 minutes each. AP-conjugated goat anti-mouse antibody was diluted 1/10000 in 15 mL 5% milk/TBST and added to the membrane, and incubated for 1.5 hours with rocking. The membrane was washed 3 times in TBST for 10 minutes each. The membrane was developed with NBT/BCIP (one tablet in 10 mL ddH<sub>2</sub>O).

### **3.2.3 Purification of H<sub>6</sub>PAI-1 from *E. coli* Inclusion Bodies**

The following protocol has been adapted from (139). 500 mL LB containing 100 µg/mL ampicillin was inoculated with 1/50 volume of BL21(DE3) pET-11d H<sub>6</sub>PAI-1 overnight culture, and incubated on a shaker at 37°C until OD<sub>600</sub> = 0.5. Expression of H<sub>6</sub>PAI-1 was induced with the addition of IPTG to 1 mM, and incubated for 3 hours on a shaker at 37°C. The culture was centrifuged at 4300 x g and 4°C for 10 minutes using the Sorvall RC 5B Plus centrifuge. The supernatant was decanted and cells suspended in 20 mL equilibrium buffer (0.5 M NaCl, 0.05% Tween 80, 20 mM NaOAc, pH 5.6). Cells were sonicated for 3 minutes (amplitude 30, 6 second pulses), then centrifuged at 9800 x g and

4°C for 30 minutes to retrieve inclusion bodies using the Sorvall RC 5B Plus centrifuge. The pellet was dissolved in 20 mL denaturation buffer (1 M NaCl, 0.01% Tween 80, 20 mM NaOAc, 4 M guanidine hydrochloride, 0.05% β-mercaptoethanol, pH 5.6) and rocked overnight at 4°C. The sample was diluted into 500 mL 2x NaCl equilibrium buffer (1 M NaCl, 0.01% Tween 80, 20 mM NaOAc, pH 5.6) and stirred at room temperature for 1 hour. The solution was further diluted with prechilled 500 mL of the same buffer but without NaCl. The sample was applied over a 2 mL SP-sepharose column which had been equilibrated with equilibrium buffer, and the column was washed with 25 mL equilibrium buffer to remove unbound sample. The sample was eluted with the same buffer adjusted to have a NaCl gradient from 0.75-1.5 M with 0.25 M increments, as per (139). SDS-PAGE gel electrophoresis was performed to determine the purity of eluted fractions. Pure fractions were pooled then concentrated using an Amicon Ultra MWCO 10000 centrifugation filter and centrifuged at 3000 rpm for 20 minutes using the Beckman Coulter Allegra 6R.

#### **3.2.4 Bradford Assay to Quantify Purified H<sub>6</sub>PAI-1 Protein**

A protein standard was created using eight dilutions of 2 mg/mL BSA ranging from 0.1 mg/mL – 1.0 mg/mL. BSA was diluted with the buffer in which the refolded H<sub>6</sub>PAI-1 protein was eluted during SP-Sepharose purification (0.75 M NaCl, 0.05% Tween 80, 20 mM NaOAc, pH 5.6). Undiluted, 1 in 2 and 1 in 4 diluted samples were created for the purified protein. 5 μL of each sample was added in duplicate to a 96 well microtiter plate including 2 blanks composed only of buffer. To each 5 μL sample, 200 μL of Coomassie

Plus Protein Assay Reagent was added. The plate was read using a BioTek EL808 plate reader.

### **3.3 Gel Based Serpin-Enzyme Complex Assays (PAI-1 and tPA)**

5  $\mu$ M of refolded H<sub>6</sub>PAI-1 retrieved from inclusion bodies with 1  $\mu$ M tPA in a total volume of 100  $\mu$ L was incubated at 37°C. 33  $\mu$ L samples were taken at for 5, 10, and 20 minutes. This reaction was repeated for nonfolded H<sub>6</sub>PAI-1 which was purified using a Ni-NTA resin column, described in “The QIAexpressionist” handbook 5<sup>th</sup> edition, and passed over a SP-Sepharose column as described in Section 3.2.3. Reactions were stopped with the addition of 4x SDS-PAGE dye. Reaction products were electrophoresed on 10% SDS-PAGE gels and stained with Coomassie Blue. Equimolar amounts of the purified protein products were used as molecular weight standards.

### **3.4 Expressing PAI-1 and $\alpha_1$ -PI M358R on T7Select10-3b Phage**

#### **3.4.1 Preparing PAI-1 and $\alpha_1$ -PI M358R Inserts**

Two separate PCR reactions using Phusion polymerase were performed using PAI-1 cDNA and pBAD H<sub>6</sub> $\alpha_1$ -PI M358R (CS) plasmid as the templates (See (140) for design of pBAD H<sub>6</sub> $\alpha_1$ -PI M358R (CS) plasmid). Primers PAI1T710bS and PAI1T710bAS were used for the PAI-1 reaction, while primers M358RT7S and M358RT7AS were used for the  $\alpha_1$ -PI M358R reaction. These primers added a 5' EcoRI restriction site, followed by a serine residue to ensure the insert was in-frame when fused to the T7Select10-3b phage, and a 3' HindIII restriction site. The 1160 bp PAI-1 and 1205 bp  $\alpha_1$ -PI M358R inserts were visualized and purified as in Section 3.1. The purified PCR product was digested

with EcoRI and HindIII restriction enzymes and the fragments were visualized on a 1% agarose gel, and gel purified using the QIAquick Gel Extraction kit.

### **3.4.2 Ligating Inserts and Packaging into T7Select Phage**

The purified PAI-1 and  $\alpha_1$ -PI M358R digested DNA products and provided Control Insert (coding for the 15 amino acid S-tag) were each ligated to T7Select10-3b Vector Arms as per the T7Select System Manual, using T4 DNA Ligase. Each 5  $\mu$ L ligation product was incubated with 25  $\mu$ L of the provided T7Select Packaging Extract for 2 hours at room temperature. The reaction was stopped by the addition of 270  $\mu$ L LB and 20  $\mu$ L of chloroform before storing the packaged T7Select10-3b phage at 4°C.

### **3.5 Plaque Assay to Determine Titer of Phage**

5 mL of M9LB was inoculated with BLT5615 *E. coli* from the provided glycerol stock, and was incubated on a shaker at 37°C until OD<sub>600</sub> = 0.5. Expression of additional 10A coat protein was induced with 1 mM IPTG and the culture was incubated on a shaker at 37°C until OD<sub>600</sub> = 1.0. A series of dilutions ( $10^3$ - $10^7$ ) of phage samples was prepared using LB as the diluent. 250  $\mu$ L of BLT5615 cells and 100  $\mu$ L of diluted phage sample were pipetted into a 15 mL sterile tube, followed by 3 mL of melted top agarose (see T7Select System Manual) and the solution was poured over a prewarmed (37°C) LB/ampicillin (100  $\mu$ g/mL) plate. The agarose was allowed to harden before the plates were inverted and incubated overnight at room temperature. The phage titer was calculated based on the number of plaques visible on the plate, times the dilution, times 10, to give a titer value expressed in plaque forming units per mL (pfu/mL). Plaques were counted using a Bio-Rad Gel Doc XR imager and Quantity One software.

### **3.6 Amplification of T7Select10-3b Phage**

#### **3.6.1 Plate Lysate Method**

The packaged reaction (Section 3.4.2) was centrifuged briefly to separate chloroform. The entire packaging reaction was used to infect 1 mL BLT5615 cells ( $OD_{600} = 0.5$ ), before 10 mL of melted top agarose was added. The solution was poured over two LB/ampicillin (100  $\mu\text{g}/\text{mL}$ ) plates, 5 mL each, and the agarose was allowed to harden before the plates were inverted and incubated overnight at room temperature. Each plate was covered with 10 mL Phage Extraction Buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 6 mM  $\text{MgSO}_4$ ), before incubating the plates at 4°C overnight and the eluted phage were combined. To confirm the presence of the introduced PAI-1 or  $\alpha_1$ -PI M358R inserts, a PCR reaction using the HotStarTaq Plus Master Mix was performed with primers PAI1T710bS and PAI1T710bAS for the PAI-1 reaction, M358RT7S and M358RT7AS for the  $\alpha_1$ -PI M358R reaction. 1/10 of the final volume of the PCR reaction was a sample of unpurified phage lysate. The product was visualized on a 1% agarose gel and gel purified using the QIAquick Gel Extraction kit before being sent to MOBIX for sequencing with the above primers. The titer of amplified lysates was determined by plaque assay (Section 3.5).

#### **3.6.2 Liquid Lysate Method**

A 50 mL culture of BLT5615 in LB/ampicillin (100  $\mu\text{g}/\text{mL}$ ) (induced with 1 mM IPTG when  $OD_{600} = 0.5$ , and incubated for an additional 30 minutes) was infected with  $10^7$ - $10^8$  plaque forming units of desired phage, as per T7Select System Manual. The infected culture was incubated at 37°C on a shaker until lysis was observed (1 - 4 hours). The lysed culture was then centrifuged at 8000 x g for 10 minutes using the Sorvall RC 5B

Plus centrifuge and the supernatant (phage lysate) stored at 4°C. The titer of amplified lysates was determined by plaque assay (Section 3.5).

### **3.7 Purifying Phage**

To precipitate phage 1/6 volume of 50% polyethylene glycol 8000 was added to phage lysate and the solution was vortexed before incubating on ice at 4°C for 2 hours. The solution was then centrifuged for 30 minutes and 4°C at 10000 x g using the Sorvall RC 5B Plus centrifuge. The supernatant was decanted and excess liquid was allowed to drain before resuspending pellet in 1.2 mL Phage Concentrating Buffer (1 M NaCl, 10 mM Tris-HCl pH 8, 1 mM EDTA). The solution was transferred to a 1.5 mL tube and centrifuged at 12000 x g for 10 min in Eppendorf Microcentrifuge and the supernatant (purified phage) was transferred to a fresh 1.5 mL tube.

### **3.8 Detection of Expressed Serpins using Immunoblotted Plaque Lifts**

A plaque assay plate (Section 3.5) were first chilled for 1 hour at 4°C before a circular nitrocellulose membrane was applied to the plate and removed after 5 minutes. The membrane was air dried for 15 minutes at room temperature then blocked with 10 mL 5% milk/TBST for 30 minutes with rocking. The blocking solution was replaced with fresh 10 mL 5% milk/TBST and 1:5000 dilution of either rabbit anti-human PAI-1 antibody or HRP-conjugated sheep anti-human  $\alpha_1$ -PI antibody before the membrane was incubated for 2 hours at room temperature with rocking. The membrane was then washed 3 times for 5 minutes each with 10 mL TBST. If probing for  $\alpha_1$ -PI, HRP-conjugated sheep anti-human  $\alpha_1$ -PI antibody was developed with 5 mL HRP detection solution per membrane.

If probing for PAI-1, the rabbit anti-human PAI-1 antibody solution was replaced with fresh 10 mL 5% milk/TBST and 1:5000 dilution of AP-conjugated goat anti-rabbit antibody. The nitrocellulose membrane was then incubated for 1.5 hours at room temperature with rocking, before washed as described above and developed by the addition of NBT/BCIP, (one tablet in 10 mL ddH<sub>2</sub>O).

### **3.9 Unsuccessful Biopanning Methods**

#### **3.9.1 Protease-Coated Well**

The following protocol was adapted from (121). Wells of a microtiter plate were prepared by applying 100  $\mu$ L of 10  $\mu$ g/mL tPA diluted in TBS per well before incubating for 3 hours at room temperature, followed by washing the wells three times with 300  $\mu$ L TBS to remove unbound protein. Wells were blocked with 200  $\mu$ L of 5% milk/TBS per well and incubated for 1 hour at room temperature before being washed 5 times with ddH<sub>2</sub>O.  $1 \times 10^7$  pfu of T7Select10-3b PAI-1 phage from lysate and 90  $\mu$ L TBST was then added to each well. The plate was incubated at room temperature for 1 hour, and then washed 5 times with 300  $\mu$ L TBST. Bound phage were eluted by the addition of 200  $\mu$ L of T7 Elution Buffer (1% SDS) and incubated for 20 minutes at room temperature. The 200  $\mu$ L solution of eluted phage was used to infect a 50 mL BLT5615 culture, as per Section 3.6.2. This biopanning process was repeated for a total of 5 rounds. To judge successful biopanning a PCR reaction using the HotStarTaq Plus Master Mix was performed with primers PAI1T710bS and PAI1T710bAS, where 1/10 of the final volume was a sample of phage lysate. The product was visualized on a 1% agarose gel.

### **3.9.2 Antibody-Coated Well**

The following protocol was adapted from (124). Wells of a microtiter plate were prepared by applying 100  $\mu$ L of 10  $\mu$ g/mL mouse anti-human tPA antibody diluted in 0.1 M  $\text{NaHCO}_3$  pH 8.3 per well before incubating overnight at 4°C, followed by washing the wells three times with 300  $\mu$ L TBS to remove unbound protein. Wells were blocked with 200  $\mu$ L of 5% milk/TBS per well and incubated for 1 hour at room temperature before being washed 5 times with ddH<sub>2</sub>O.  $6 \times 10^5$  pfu of T7Select10-3b PAI-1 phage and T7Select10-3b Control Phage from lysate were each incubated with 0.3 nM tPA in 300  $\mu$ L LB for 1 hour at 37°C. Each reaction was applied to a prepared well and incubated for 2 hours at 37°C, before being washed 5 times with TBST. Bound phage were eluted with a pH pulse using 50  $\mu$ L 0.1 M HCl (pH 2.2) per well, incubated for 1 minute then neutralized using 5  $\mu$ L 1 M Tris-HCl pH 8.0, before used to infect a 50 mL culture of BLT5615 as per Section 3.6.2. This biopanning process was repeated for a total of 5 rounds. To judge success of biopanning, PCR was performed as in Section 3.9.1.

### **3.9.3 T25-Direct Infection**

The following protocol was adapted from (141). T25 tissue culture flasks were prepared by applying 5 mL of 10  $\mu$ g/mL thrombin diluted in PBS to each before incubating overnight at 4°C. The flasks were blocked with 5 mL of 0.4% milk/PBS and incubated for 2 hours at room temperature before being washed 3 times with ddH<sub>2</sub>O.  $2 \times 10^{11}$  pfu of purified T7Select10-3b  $\alpha_1$ -PI M358R or T7Select10-3b Control Phage were added to their respective flasks before incubating at room temperature for 2 hours with rocking. The flasks were then washed 10 times with 5 mL PBST and 5 times with 5 mL ddH<sub>2</sub>O. 5 mL

of BLT5615 cells were added (induced with 1 mM IPTG when  $OD_{600} = 0.5$ , incubated for an additional 30 minutes) and incubated at 37°C with rocking until lysis observed (1-3 hours). To judge success of biopanning, plaque assays performed as in Section 3.5.

### **3.10 Detection of Phage-Expressed Serpin-Enzyme Complexes**

$1 \times 10^{10}$  pfu of purified T7Select10-3b  $\alpha_1$ -PI M358R and T7Select10-3b Control Phage were each mixed with 20 nM thrombin in PBS to a final volume of 50  $\mu$ L. Separate tubes containing only phage and PBS without thrombin were also prepared. The four reactions were incubated at 37°C for 30 minutes before being loaded into wells of a 1.5 mm 8% SDS-PAGE gel. Following electrophoresis the gel was transferred to a nitrocellulose membrane using iBlot Gel Transfer Stacks and an iBlot machine as per the manufacturer's instructions, and the membrane was incubated overnight in 3% BSA/TBST with rocking. Sheep anti-prothrombin affinity-purified IgG antibody (137), which had been biotinylated using a DSB-X Biotin Protein Labeling kit, was diluted 1/2000 in 10 mL of 3% BSA/TBST and added to the membrane before incubating for 1.5 hours with rocking. The membrane was then washed 3 times with TBST for 10 minutes each. Qdot 625 streptavidin conjugate was diluted 1/2000 in 10 mL of 3% BSA/TBST and added to the nitrocellulose membrane and incubated for 1 hour with rocking. The membrane was again washed 3 times in TBST for 10 minutes each. The membrane was visualized using a Bio-Rad Gel Doc XR under ultraviolet light (302 nm). To compare to serpins not fused to T7Select phage, detection was similarly completed using 20 ng purified recombinant  $\alpha_1$ -PI M358R, with or without 50 ng of thrombin, on a separate gel and blot.

### 3.11 Generation of $\alpha_1$ -PI P2P1 and $\alpha_1$ -PI M358R P7-P3 T7Select10-3b Phage

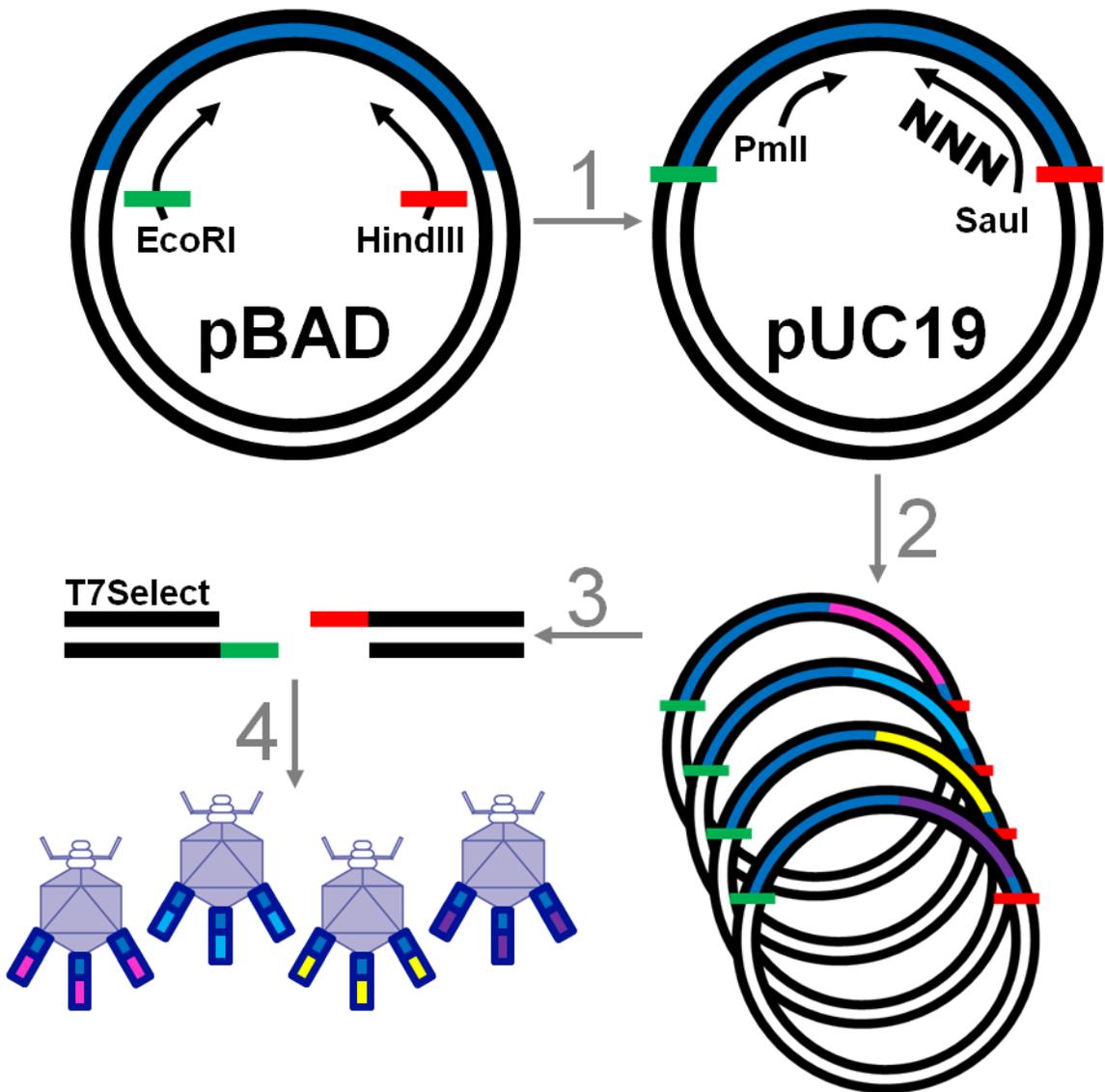
#### Libraries

See **Figure 6** for a schematic of how the libraries were generated.

#### 3.11.1 Preparing pUC19 $\alpha_1$ -PI M358R Plasmid for Sub-cloning

The pUC19 *E. coli* plasmid was subjected to mini-plasmid DNA extraction and purification using the QIAprep Spin Miniprep Kit to isolate it from the DH5 $\alpha$  host *E. coli* strain. To isolate DNA coding for  $\alpha_1$ -PI M358R, a PCR reaction was performed using HotStarTaq Plus Master Mix with the plasmid pBAD H<sub>6</sub> $\alpha_1$ -PI M358R (CS) as the template. Primers M358RT7S and M358RT7AS were used to add a 5' EcoRI restriction site, followed by a serine residue to ensure the insert was in-frame when inserted into phage, and a 3' HindIII restriction site. These primers did not include the His-tag. The "CS" annotation denotes previous modifications to add ClaI and SauI restriction sites within the RCL of  $\alpha_1$ -PI M358R to assist additional modification (140). This annotation is not mentioned in future constructs although they also contain the modification. The 1205 bp PCR product was visualized on a 1% agarose gel before being purified using the QIAquick PCR purification kit. The purified  $\alpha_1$ -PI M358R PCR product and the purified pUC19 plasmid were each digested with EcoRI and HindIII restriction enzymes, with cow intestine phosphatase added to the pUC19 digestion reaction. The fragments were visualized on a 1% agarose gel, and gel purified using the QIAquick Gel Extraction kit. The digested  $\alpha_1$ -PI M358R gel extraction product was ligated into the pUC19 plasmid using T4 DNA Ligase, and transformed into DH5 $\alpha$  competent *E. coli* cells. Transformed DH5 $\alpha$  cells containing the pUC19  $\alpha_1$ -PI M358R

**Figure 6: Construction of  $\alpha_1$ -PI Phage Libraries.** **1)** The entire coding region of  $\alpha_1$ -PI M358R was PCR amplified from the pBAD host plasmid, with the addition of a 5' EcoRI restriction site followed by a serine residue to ensure the insert was in-frame when inserted into phage, and a 3' HindIII restriction site. The resulting PCR product was inserted into the pUC19 plasmid via EcoRI and HindIII restriction. **2)** Degeneracy was introduced at either the P2 and P1, or P7-P3  $\alpha_1$ -PI codons via PCR mutagenesis with the appropriate degenerate primer and a primer which included the upstream PmlI restriction site. The resulting PCR product was inserted into the pUC19  $\alpha_1$ -PI M358R plasmid via PmlI and SauI restriction, and the plasmid library amplified in *E. coli*. **3)** The plasmid library was isolated from *E. coli* and inserted into the T7Select10-3b phage vector arms via EcoRI and HindIII restriction. **4)** The T7Select10-3b phage vector containing the  $\alpha_1$ -PI library was packaged into phage as per manufacturer's instructions, producing the final T7Select10-3b  $\alpha_1$ -PI phage library.



plasmid were subjected to mini-plasmid DNA extraction and purification using the QIAprep Spin Miniprep Kit, before being sent to MOBIX for sequencing with primers M358RT7S and M358RT7AS.

### **3.11.2 Generation of Plasmid Libraries using PCR Mutagenesis**

A PCR reaction for each desired library was performed using purified pUC19  $\alpha_1$ -PI M358R as the template. HotStarTaq Plus Master Mix was used with primers LP-PmlI and either APIP2P1dgnAS or APIP7P3dnAS, to generate the  $\alpha_1$ -PI P2P1 and  $\alpha_1$ -PI M358R P7-P3 libraries respectively. These primers introduced silent mutations at amino acid P361 (for rationale see **Table 1**). The products were visualized on a 1% agarose gel before being purified using the QIAquick PCR purification kit. PCR products and purified pUC19  $\alpha_1$ -PI M358R plasmid were each digested with PmlI and SauI restriction enzymes, with calf intestine alkaline phosphatase added to the pUC19  $\alpha_1$ -PI M358R digestion reaction. The fragments were visualized on a 1% agarose gel, and gel purified using the QIAquick Gel Extraction kit. The digested  $\alpha_1$ -PI P2P1 and  $\alpha_1$ -PI M358R P7-P3 gel extraction products were each ligated into the pUC19  $\alpha_1$ -PI M358R fragment using T4 DNA Ligase, and each transformed into 50  $\mu$ L DH5 $\alpha$  competent *E. coli* cells. In order to observe the degeneracy of the  $\alpha_1$ -PI M358R P7-P3 library, 1% of the cells transformed with the pUC19  $\alpha_1$ -PI M358R P7-P3 plasmid were plated on a LB/ampicillin (100  $\mu$ g/mL) agar plate.

### **3.11.3 Propagation of Plasmid Libraries**

Rather than the standard protocol of plating transformed cells, the entire volume of transformed cells was added to 200 mL LB/ampicillin (100  $\mu$ g/mL) and incubated

overnight at 37°C. The pUC19  $\alpha_1$ -PI P2P1 and pUC19  $\alpha_1$ -PI M358R P7-P3 plasmid libraries were each isolated using Plasmid Midi Kit. Each plasmid library was digested with EcoRI and HindIII and the fragments were visualized on a 1% agarose gel, and gel purified using the QIAquick Gel Extraction kit.

### **3.12 Inserting DNA Libraries into T7Select10-3b Phage**

The  $\alpha_1$ -PI P2P1 and  $\alpha_1$ -PI M358R P7-P3 EcoRI/HindIII fragments were each ligated to T7Select10-3b Vector Arms as per the T7Select System Manual, using T4 DNA Ligase. Each 5  $\mu$ L ligation product was incubated with 25  $\mu$ L of the provided T7Select Packaging Extract for 2 hours at room temperature. The reaction was stopped by the addition of 270  $\mu$ L LB and 20  $\mu$ L of chloroform before storing the packaged T7Select10-3b phage at 4°C. Libraries were amplified as in Section 3.6.2.

### **3.13 Confirming Degeneracy of Libraries**

#### **3.13.1 $\alpha_1$ -PI P2P1 Phage Library**

Plaque assays were completed to determine the titer of the amplified libraries, followed by immunoblotted plaque lifts to identify plaques expressing the  $\alpha_1$ -PI insert (Section 3.5 and Section 3.8). In order to observe the degeneracy of the T7Select10-3b  $\alpha_1$ -PI P2P1 library, the top agarose from 5 random plaques expressing the  $\alpha_1$ -PI insert were scraped using a sterile pipette tip, and resuspended in 20  $\mu$ L ddH<sub>2</sub>O. The plaque solution was incubated at 65°C for 15 minutes, cooled to room temperature, and centrifuged at 14000 x g for 3 minutes. PCR was performed using 4  $\mu$ L plaque solution, HotStarTaq Plus Master Mix, and primers M358RT7S and M358RT7AS. PCR products were visualized on a 1%

agarose gel before being purified using the QIAquick PCR purification kit. Purified PCR samples were sent to MOBIX for sequencing with primers M358RT7S and M358RT7AS.

### **3.13.2 $\alpha_1$ -PI M358R P7-P3 DNA Library**

Prior to inserting sequences into phage, 10 colonies of DH5 $\alpha$  containing clones from the pUC19  $\alpha_1$ -PI M358R P7-P3 plasmid library were randomly selected and amplified in 5 mL LB/ampicillin (100  $\mu$ g/mL) incubated overnight at 37°C. The colonies were subjected to mini-plasmid DNA extraction and purification using the QIAprep Spin Miniprep Kit, and sent to MOBIX for sequencing with primers M358RT7S and M358RT7AS.

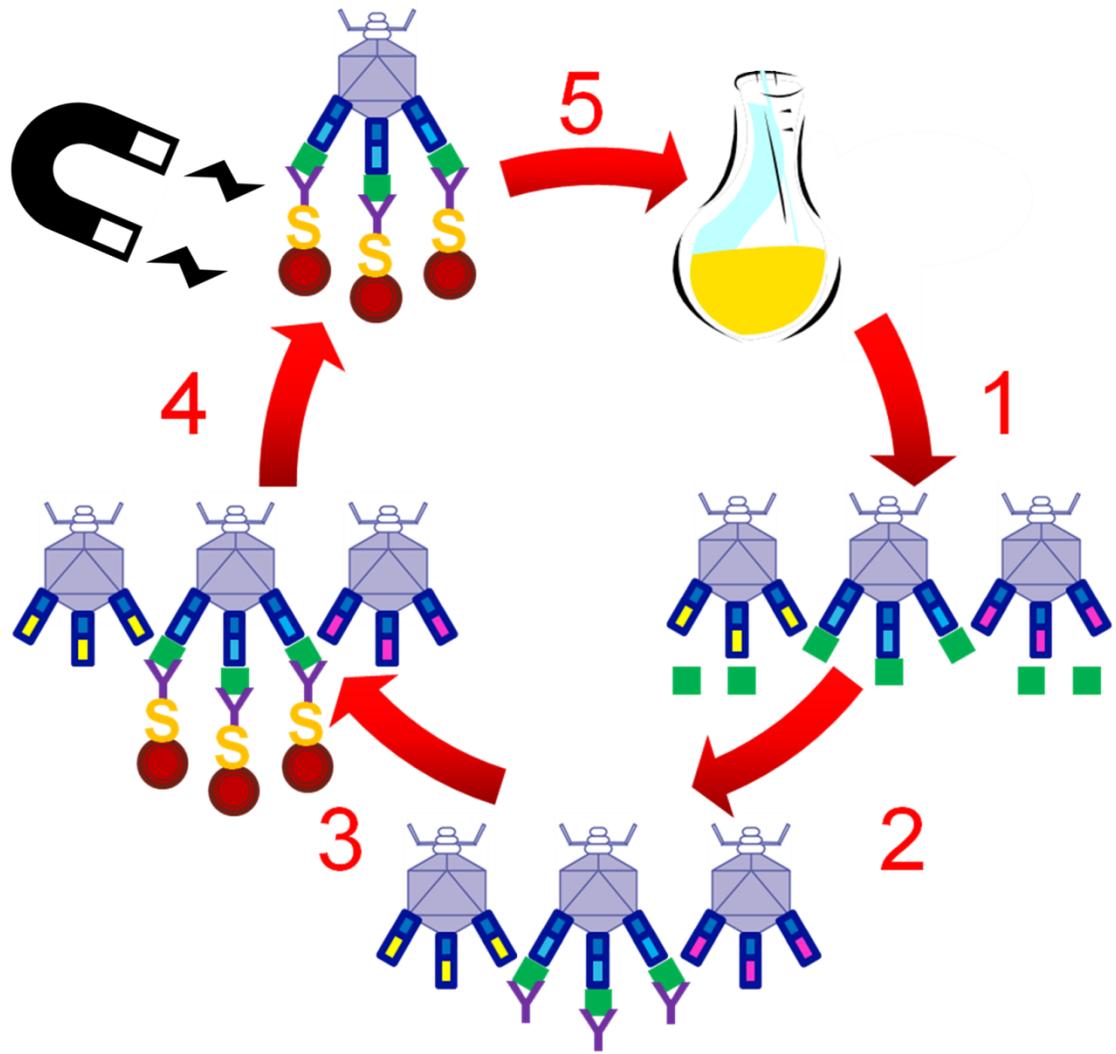
## **3.14 Biopanning Using Magnetic Dynabeads**

See **Figure 7** for a schematic of the biopanning procedure used.

### **3.14.1 Mock-Biopanning T7Select10-3b $\alpha_1$ -PI M358R**

The following protocol was adapted from (122). T7Select10-3b  $\alpha_1$ -PI M358R was mixed 1/300 with T7Select10-3b Control Phage, to a final titer of  $1 \times 10^{10}$  pfu/mL based on immunoblotted plaque lifts. In a 1.5 mL microcentrifuge tube, lysate containing  $1 \times 10^9$  pfu of the mixed phage were combined with 0.5 pmol (0.5 nM) thrombin to 1 mL in 3% BSA/PBS, and the reaction was incubated at 37°C for 30 minutes. A separate tube without the addition of thrombin was used as a control. PPACK was added to 10 nM, along with 2  $\mu$ L of biotinylated sheep anti-prothrombin antibody, and incubated at room temperature with constant rotating for 30 minutes. 75  $\mu$ L of streptavidin-coated magnetic Dynabeads were added, with an additional 30 minute incubation at room temperature with rotating. Tubes were placed against a magnet for 3 minutes, and the supernatant

**Figure 7: Biopanning Procedure.** **1)** T7Select10-3b expressing serpin library was exposed to thrombin (green block). **2)** The solution was exposed to an active site inhibitor of thrombin (PPACK) and a biotinylated anti-prothrombin antibody. **3)** The solution as exposed to streptavidin-coated magnetic beads, and **4)** only those phage expressing thrombin inhibitory serpin were selected by magnet. **5)** The nonspecific phage were removed via washing, while selected phage are amplified in a culture of *E. coli*.



containing unbound phage was removed. The beads were washed 5 times with PBST (0.05% Tween 20), with each wash step lasting 5 minutes with rotating. Beads were resuspended in 500  $\mu$ L PBS, and added directly to a 35 mL culture of BLT5615 *E. coli* (induced with 1 mM IPTG when  $OD_{600} = 0.5$ , incubated for an additional 30 minutes) in LB/ampicillin (100  $\mu$ g/mL). The infected culture was incubated at 37°C on a shaker until lysis was observed (1-4 hours). The lysed culture was centrifuged at 8000 x g for 10 minutes using the Sorvall RC 5B Plus centrifuge. Phage lysate was stored at 4°C for the next round of biopanning, with a total of 5 rounds performed. Round 5 was completed using three different washing protocols. The tubes were replaced after every other wash to minimize non-specific binding.

**Wash A:** 10 times with PBST (0.05% Tween 20), for 5 minutes each wash with rotating.

**Wash B:** 10 times with PBST (1% Tween 20), for 5 minutes each wash with rotating.

**Wash C:** 10 times with PBS made to 1M NaCl and 1% Triton X-100, for 5 minutes each wash with rotating.

A 6<sup>th</sup> round of biopanning was performed using only the “Wash C” method, with or without the addition of thrombin. After each round, plaque lifts were immunoblotted as described in Section 3.8 in order to assess the quality of enrichment.

### **3.14.2 Biopanning $\alpha_1$ -PI P2P1 and $\alpha_1$ -PI M358R P7-P3 Libraries**

Using the "Wash C" method outlined in Section 3.14.1, biopanning was otherwise performed identically, using the T7Select10-3b  $\alpha_1$ -PI P2P1 and  $\alpha_1$ -PI M358R P7-P3 libraries (+/- thrombin).  $1 \times 10^9$  pfu of phage were used for each round, or ~20  $\mu$ L of unpurified phage lysate. Phage lysate was stored at 4°C for the next round of biopanning.

Libraries were annotated as "Rd5 +IIa" if thrombin was added during biopanning or "Rd5 –IIa" if thrombin was not.

### **3.15 Determining Optimal P2 and P1 Residues for Thrombin Inhibition**

#### **3.15.1 Sequencing Plaques from Biopanned $\alpha_1$ -PI P2P1 Library**

Following 5 rounds of biopanning, DNA from individual plaques was obtained as outlined in Section 3.13.1. A total of 20 plaque PCR reactions from the T7Select10-3b  $\alpha_1$ -PI P2P1 library biopanned for 5 rounds with thrombin (P2P1 Rd5 +IIa), and the same number from the library biopanned for 5 rounds without thrombin (P2P1 Rd5 –IIa) were sent to MOBIX for sequencing, with primer M358RT7AS.

#### **3.15.2 Inserting Sequences into pBAD Expression Plasmid**

Unique PCR products from Section 3.15.1 and purified pBAD H<sub>6</sub> $\alpha_1$ -PI M358R (CS) plasmid were each digested with PmlI and SauI restriction enzymes, with calf intestine alkaline phosphatase added to the pBAD H<sub>6</sub> $\alpha_1$ -PI M358R (CS) digestion reaction. The fragments were visualized on a 1% agarose gel, and gel purified using the QIAquick Gel Extraction kit. Each unique sequence was ligated to the pBAD H<sub>6</sub> $\alpha_1$ -PI M358R (CS) vector fragment using T4 DNA Ligase, and transformed into competent Top10 *E. coli* cells.

#### **3.15.3 ELISA-Based Lysate Screen**

Each unique pBAD H<sub>6</sub> $\alpha_1$ -PI P2P1 construct in Top10 *E. coli* (Section 3.15.2) was individually grown from a glycerol stock in 6 mL LB/ampicillin (100  $\mu$ g/mL) overnight at 37°C. Each culture was induced with a final concentration of 0.002% arabinose and returned to the shaker for 3 hours to allow for expression of the mutant serpin. A 1 mL

sample of cells was harvested and resuspended in 0.5 mL PBS. Each sample of cells was lysed using the microtip of a Fisher Scientific Sonic Dismembrator 100, and centrifuged at 13000 x g at 4°C for 20 minutes. 100 µL of each lysate was applied to microtiter wells (in duplicate) that had previously been coated with 5 µg/mL of thrombin, and blocked with 5% milk powder/PBST. Lysates were incubated for 1 hour at room temperature, washed 3 times with PBST, before 100 µL of a 1/5000 dilution of HRP-conjugated sheep anti-human  $\alpha_1$ -PI antibody in 5% milk powder/PBST was added. The antibody solution was incubated for 1 hour at room temperature, washed 3 times, and exposed to TMB developing solution as per manufacturer's instructions.

#### **3.15.4 Purification and Quantification of Recombinant Proteins**

$\alpha_1$ -PI P2P1 mutant P357/M358R (which is identical to  $\alpha_1$ -PI M358R) and P357/M358P were selected for further characterization. 1 L of LB/ampicillin (100 µg/mL) was inoculated with Top10 *E. coli* containing the appropriate pBAD H<sub>6</sub> $\alpha_1$ -PI P2P1 plasmid. Each culture was induced with a final concentration of 0.002% arabinose when the OD<sub>600</sub> reached 0.5 and returned to the shaker for 4 hours to allow for expression of the mutant serpin. Purification was completed using standard nickel affinity column as in (142), although the elution fractions were dialyzed overnight in 20mM sodium phosphate pH 6.8 prior to ion exchange chromatography using DEAE Sepharose. A NaCl gradient ranging from 50 to 200 mM in 20 mM sodium phosphate pH 6.8 was utilized for DEAE Sepharose purification. Fractions containing the protein of interest as determined by SDS-PAGE were pooled and concentrated using an Amicon Ultra MWCO 10000 centrifugation filter and stored at -80°C. Protein concentration was determined by Bradford assay.

### **3.15.5 Gel Based Serpin-Enzyme Complex Assays**

1  $\mu$ M of  $\alpha_1$ -PI P2P1 mutants P357/M358R or P357/M358P with 0.2 or 0.5  $\mu$ M thrombin brought to a total volume of 20  $\mu$ L was incubated at 37°C for 1 minute and 5 minute intervals. Reactions were stopped with the addition of 4x SDS-PAGE dye. Reaction products were electrophoresed on 10% SDS-PAGE gels and stained with Coomassie Blue.

### **3.16 Determining Optimal P7 to P3 Residues for Thrombin Inhibition**

#### **3.16.1 Inserting Sequences into pBAD Expression Plasmid**

Following 5 rounds of biopanning, a PCR reaction was performed using 2  $\mu$ L lysate from the 5<sup>th</sup> round ( $\sim 1.4 \times 10^8$  pfu of Rd5 +IIa biopanned with thrombin, and a separate reaction for Rd5 –IIa biopanned without thrombin), HotStar HiFidelity polymerase, and primers M358RT7S and M358RT7AS. Following purification, the PCR product was digested and inserted into the pBAD H<sub>6</sub> $\alpha_1$ -PI M358R (CS) vector as in Section 3.15.2. Two subsequent pBAD H<sub>6</sub> $\alpha_1$ -PI M358R P7-P3 plasmid libraries resulted: one enriched for thrombin inhibition (P7-P3 Rd5 +IIa), and the other acting as a negative control (P7-P3 Rd5 –IIa). Each was transformed into competent Top10 *E. coli* cells.

#### **3.16.2 ELISA-Based Lysate Screen**

Randomly selected colonies that had been transformed with either the pBAD H<sub>6</sub> $\alpha_1$ -PI M358R P7-P3 Rd5 +IIa or Rd5 –IIa plasmid library were analyzed as per Section 3.15.3. 80 colonies were selected from the Rd5 +IIa library, and 40 colonies were selected from the Rd5 –IIa library. Colonies which indicated a greater level of thrombin complexing than the  $\alpha_1$ -PI M358R protein were screened a second time using this method.

### **3.16.3 Sequencing Biopanned $\alpha_1$ -PI M358R P7-P3 Library**

#### **3.16.3.1 Randomly Selected Plaque Method**

DNA from individual plaques was obtained as outlined in Section 3.13.1. A total of 40 plaque PCR reactions were sent to MOBIX for sequencing with primer M358RT7AS, from each of the T7Select10-3b  $\alpha_1$ -PI M358R P7-P3 Rd5 +IIa and Rd5 –IIa libraries.

#### **3.16.3.2 Sequencing based on ELISA-Based Lysate Screen Results**

Colonies which indicated a greater level of thrombin binding than the  $\alpha_1$ -PI M358R protein were subjected to mini-plasmid DNA extraction and purification using the QIAprep Spin Miniprep Kit. Samples were then sent to MOBIX for sequencing, using the pBAD specific primer pBADrev. Unique sequences were input into ConSurf to quantify amino acid conservation at each position in order to observe trends, using  $\alpha_1$ -PI M358R as the template sequence and PDB ID 1OPH as the template structure (138).

#### **3.16.4 Purification and Quantification of Recombinant Proteins**

$\alpha_1$ -PI M358R P7-P3 mutants of interest were selected based on their apparent thrombin inhibitory activity (Section 3.16.2) and the frequency in which they appeared in sequencing results (Section 3.16.3). Five mutants were ultimately selected and purified as in Section 3.15.4. Final protein concentration was determined using a Thermo Scientific NanoDrop Spectrophotometer, calibrated to the extinction coefficient for  $\alpha_1$ -PI (19940 M<sup>-1</sup> s<sup>-1</sup>) and molecular weights unique to each mutant protein.

#### **3.16.5 Gel Based Serpin-Enzyme Complex Assays**

Assays were completed as in Section 3.15.5 for  $\alpha_1$ -PI M358R P7-P3 mutants of interest.

### 3.16.6 Kinetic Analysis of $\alpha_1$ -PI M358R P7-P3 Mutants of Interest

$\alpha_1$ -PI M358R P7-P3 mutants of interest or native  $\alpha_1$ -PI M358R (200 nM) were incubated with thrombin (10 nM), in PPNE buffer (20 mM  $\text{Na}_2\text{HPO}_4$ , pH 7.4, 100 nM NaCl, 0.1 mM EDTA, 0.1% polyethylene glycol 8000) for up to 70 seconds with 10 second intervals. Chromagenic substrate S2238 was added to 10  $\mu\text{M}$  for native  $\alpha_1$ -PI M358R or 25  $\mu\text{M}$  for  $\alpha_1$ -PI M358R P7-P3 mutants to improve the accuracy of results by slowing the reaction between serpin and thrombin. Reactions were quenched with chromogenic substrate (150  $\mu\text{M}$  S2238) to determine residual protease activity. The change in absorbance over 5 minutes at 405 nm was measured with a BioTek EL808 plate reader and second order rate constants ( $k_2$ ) were determined. This was performed by graphing the time in seconds (x-axis) by  $\ln(P_o/P_t)$  (y-axis):  $P_o$ , initial thrombin activity;  $P_t$ , thrombin activity at time t. Linear regression was performed; data were only included in cases where the regression coefficient ( $R^2$ )  $\geq 0.9$ . The slope was used in the equation  $k_2 = \text{slope} / [\text{serpin}] \times (1 + [\text{S2238 to slow rxn}] / K_m)$ , where the  $K_m$  was determined previously to be 2.9  $\mu\text{M}$  by Jason Sutherland in his PhD thesis (McMaster University, 2007). Stoichiometries of thrombin inhibition (SI) of the recombinant serpins were determined by incubating inhibitor (range of concentration 0-500 nM initially, then focused to 100nM, 75nM, 50nM, and 25nM) with 50 nM thrombin for two hours at room temperature, proteins diluted in PPNE. This reaction was quenched with 150  $\mu\text{M}$  chromogenic substrate S2238. Residual thrombin activity was measured by reading the absorbance at 405 nm over 5 minutes. The number of serpin molecules required to inhibit

one molecule of thrombin was calculated by plotting the percent activity of thrombin vs the ratio of serpin to thrombin.

### 3.17 Molecular Modeling of $\alpha_1$ -PI Interactions with Thrombin

The following files were imported into PyMOL, with waters and other non-amino acids removed: Encounter complex between  $\alpha_1$ -PI M358R and S195A trypsin (PDB ID 1OPH), encounter complex between HCII and S195A thrombin (PDB ID 1JMO). Thrombin's amino acid 195 was mutated back to the wild type serine. HCII was selected due to its similar P2 proline residue, and when aligned it had the smallest root-mean-square deviation (RMSD) compared to AT and PN1 thrombin encounter complexes (PDB ID 2BEH and 4DY7), calculated by the PyMOL `align` command. The RCL residues P7-P3' of  $\alpha_1$ -PI M358R and HCII were aligned, which placed  $\alpha_1$ -PI M358R in a comparable orientation within the active site of thrombin (RMSD = 0.227Å). Due to irresolvable steric clashes between  $\alpha_1$ -PI M358R and thrombin the following mutations were introduced using PyMOL in  $\alpha_1$ -PI M358R: R282G, S283G; and in thrombin: R221a rotated 97°. To generate P7-P3 mutants of interest, mutations were manually introduced using PyMOL with mutant residues rotated to limit the steric clashes. Mutants were selected based on their kinetic data and on consensus sequences in sequenced plaques and mutants with improved thrombin binding. The  $\alpha_1$ -PI-thrombin encounter complexes were exported as PDB files and loaded into the ZMM program. The ZMM program was modified to accept multiple ligand identifiers, and the ligand was defined as  $\alpha_1$ -PI residues P7-P3' (352-361). Full-fledged Monte Carlo minimizations (MCM) were completed using ZMM as outlined in (143), 5000 MCM of a single trajectory.

## **4. RESULTS**

### **4.1 Purification of H<sub>6</sub>PAI-1 from *E. coli* Inclusion Bodies**

The protocol developed by (139) with modifications as described in Section 3.2 was applied. Sonicated bacterial cells were subjected to denaturing to free H<sub>6</sub>PAI-1 believed to be contained within inclusion bodies, and purified via SP-Sepharose column. 5 fractions eluted from the SP-Sepharose column with 0.75 M and 1 M NaCl buffer were pooled, as described in section 3.2.3. A Bradford assay determined the concentration to be 0.5 mg/mL, indicating 0.375 mg of refolded H<sub>6</sub>PAI-1 was retrieved from a 500 mL culture. Comparatively, sonicated bacteria not subjected to denaturing but purified via nickel column and SP-Sepharose column yielded a pooled fraction containing 2.3 mg/mL Soluble H<sub>6</sub>PAI-1, or 3.45 mg from a 500 mL culture.

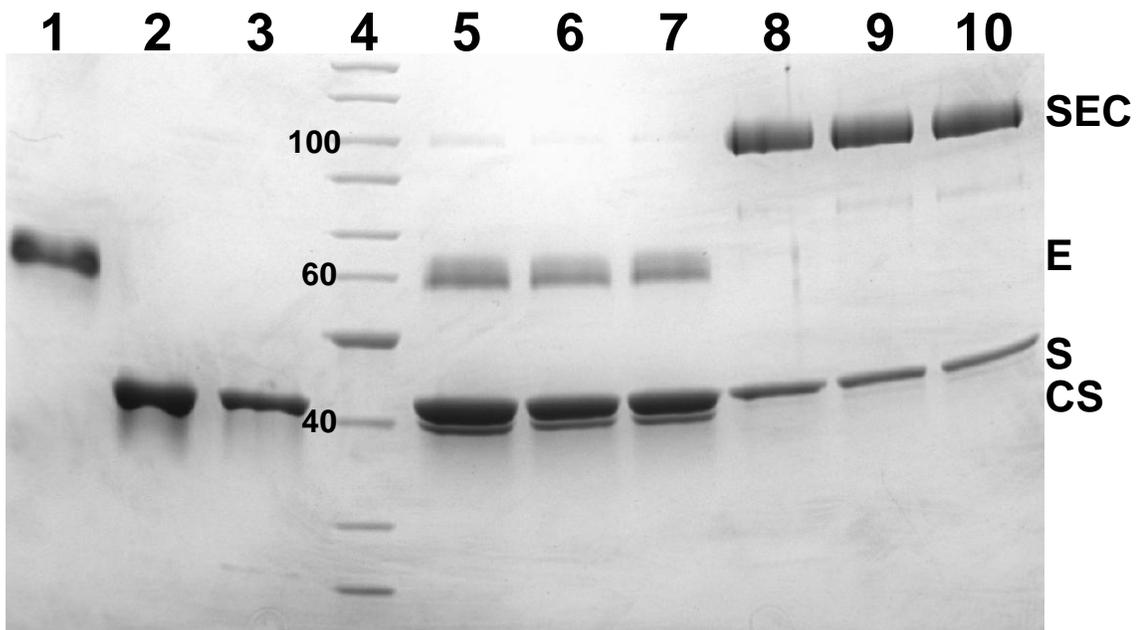
### **4.2 Purified H<sub>6</sub>PAI-1 Forms SDS Stable Complexes with tPA**

H<sub>6</sub>PAI-1 purified from inclusion bodies or soluble H<sub>6</sub>PAI-1 was incubated with tPA and SDS-PAGE gel electrophoresis was performed. The H<sub>6</sub>PAI-1 purified from inclusion bodies effectively formed SDS stable serpin-enzyme complexes, while only a small fraction of soluble H<sub>6</sub>PAI-1 was capable of forming serpin-enzyme complexes (**Figure 8**). Refolding of soluble H<sub>6</sub>PAI-1 was not attempted as for the purposes of this study as the objective to retrieve functional PAI-1 protein was met.

### **4.3 Expression of PAI-1 and $\alpha_1$ -PI M358R on T7Select10-3b Phage**

The PAI-1 and  $\alpha_1$ -PI M358R inserts were successfully separately cloned into T7Select10-3b phage, fusing it to the C-terminus of the 10B coat protein, with the correct

**Figure 8: Purified H<sub>6</sub>PAI-1 Forms SDS Stable Complexes with tPA.** Lane 1, 1 uM tPA; Lane 2, 5 μM purified soluble H<sub>6</sub>PAI-1; Lane 3, 5 μM purified H<sub>6</sub>PAI-1 from inclusion bodies; Lane 4, PageRuler Protein Ladder (Fermentas); Lane 5-7, 5 μM purified soluble H<sub>6</sub>PAI-1 and 1 uM tPA incubated for 5, 10 and 20 min at 37°C; Lane 8-10, 5 μM purified H<sub>6</sub>PAI-1 from inclusion bodies and 1 uM tPA incubated for 5, 10 and 20 min at 37°C. Each well volume is 33 μL. SEC, serpin-enzyme complex; E, enzyme; S, unreacted serpin; CS, cleaved serpin. Molecular weight markers are shown in kDa.

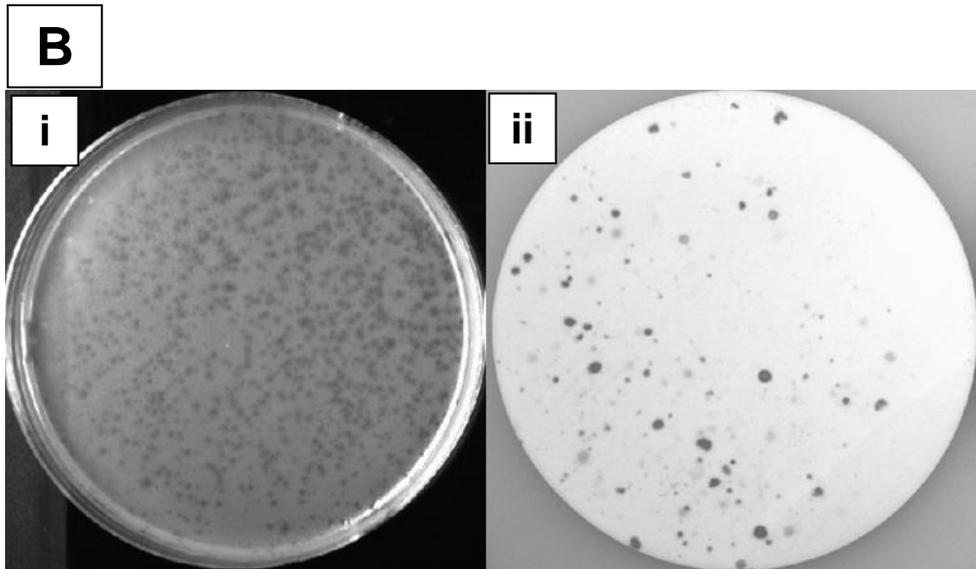
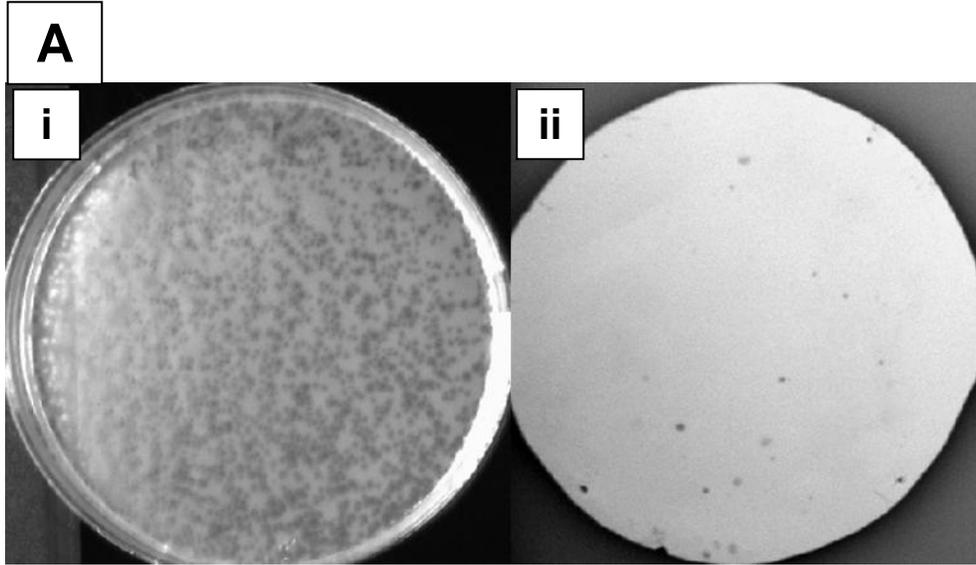


DNA sequence confirmed by PCR amplification from phage lysate. An immunoblotted plaque lift confirmed the presence of PAI-1 and  $\alpha_1$ -PI M358R on the surface of phage particles prior to biopanning, with an average of 2% of phage detectably expressing PAI-1 or 13% of phage detectably expressing  $\alpha_1$ -PI M358R (**Figure 9**). Sequencing of plaques which did not detectably express serpin protein indicated that they also contained the inserted sequence (data not shown). This suggests that either the copy number expressed by those plaques was too low to detect by immunoblot, or these plaques did not express the serpin for an undefined reason.

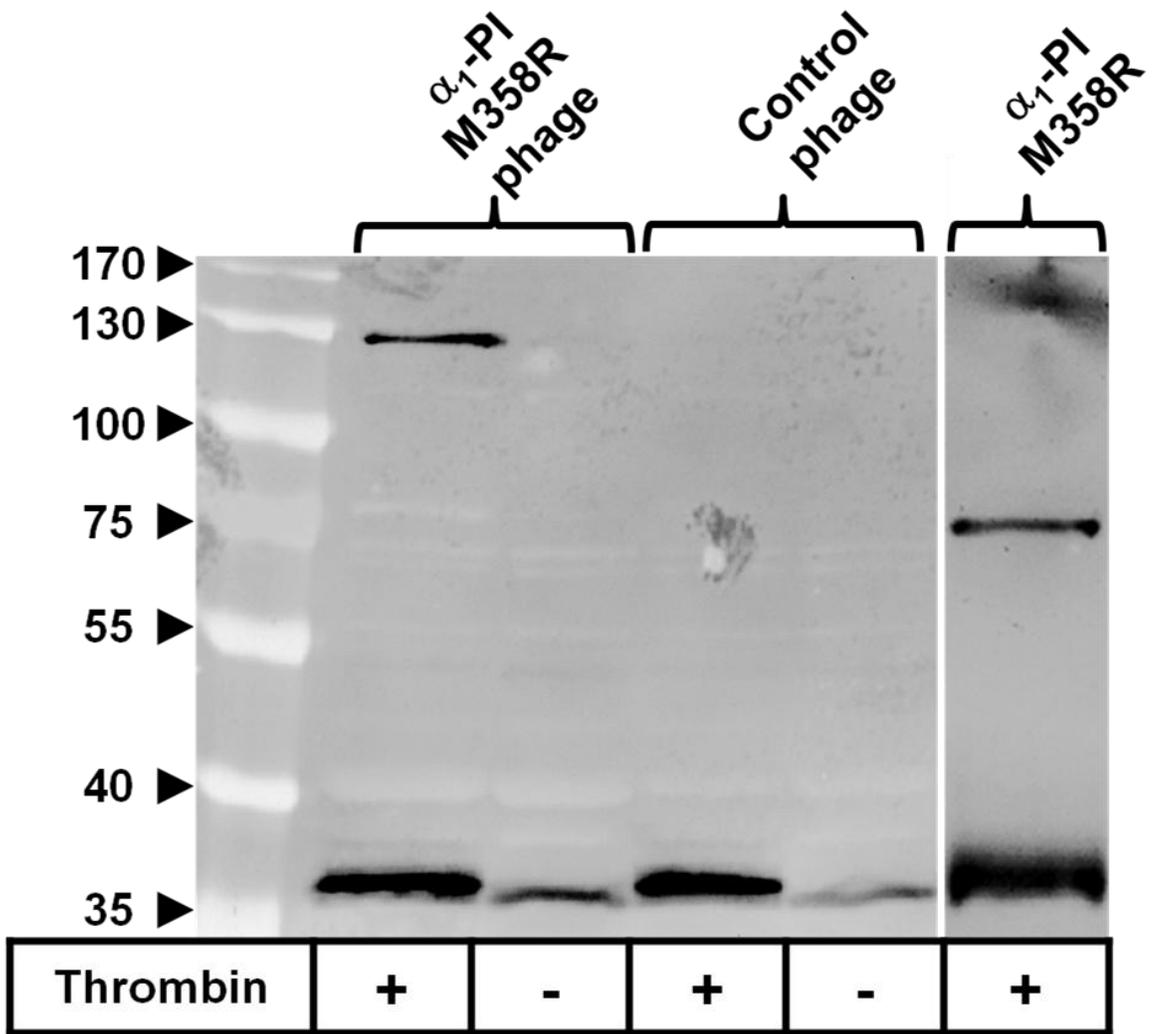
#### **4.4 Phage-Expressed $\alpha_1$ -PI M358R Forms a Stable Serpin-Enzyme Complex with Thrombin**

Thrombin was incubated with samples of purified T7Select10-3b  $\alpha_1$ -PI M358R and T7Select10-3b Control Phage. SDS-PAGE gel electrophoresis and immunoblotting using a biotinylated antibody to thrombin was performed, and the nitrocellulose membrane exposed to streptavidin-conjugated nanocrystals that were illuminated by UV light. This process revealed the presence of phage-expressed  $\alpha_1$ -PI M358R-thrombin complexes at approximately 124 kDa (**Figure 10**). This was the expected size of the full complex as  $\alpha_1$ -PI M358R is fused to the surface of T7Select10-3b phage (10B coat protein 398 amino acids/42 kDa,  $\alpha_1$ -PI M358R 394 amino acids/45 kDa, thrombin 295 amino acids/37 kDa). No such complexes were observed without the addition of thrombin, nor were they observed in incubations between T7Select10-3b Control Phage and thrombin. Recombinant native  $\alpha_1$ -PI M358R was also able to form a complex with thrombin, detected using the same methods on a separate immunoblot.

**Figure 9: Immunoblotted Plaque Lifts A) Detection of T7Select10-3b Expressing PAI-1.** **i)** Plaque assay of amplified T7Select10-3b PAI-1 phage, ~2040 plaques. **ii)** Immunoblotted plaque lift of plate shown in i), 26 visible hits indicating expression of PAI-1. Average detected expression of PAI-1 found to be ~2% of plaques. **B) Detection of T7Select10-3b Expressing  $\alpha_1$ -PI M358R.** **i)** Plaque assay of amplified T7Select10-3b  $\alpha_1$ -PI M358R phage, ~1340 plaques. **ii)** Immunoblotted plaque lift of plate shown in i), 206 visible hits indicating expression of  $\alpha_1$ -PI M358R. Average detected expression of  $\alpha_1$ -PI M358R found to be ~13% of plaques.



**Figure 10: Detection of Phage-Expressed Serpin-Enzyme Complexes.**  $1 \times 10^{10}$  pfu of purified T7Select10-3b  $\alpha_1$ -PI M358R and T7Select10-3b Control Phage were each incubated with or without 20 nM thrombin (50  $\mu$ L reaction), before immunoblotted with a biotinylated anti-prothrombin antibody. Complex formation was detected only between T7Select10-3b  $\alpha_1$ -PI M358R phage and thrombin (~124 kDa). Purified  $\alpha_1$ -PI M358R protein-thrombin complexes are indicated for reference (~82 kDa), detected using the same methods on a separate immunoblot (20 ng  $\alpha_1$ -PI M358R, with 50 ng of thrombin). Molecular weight markers are shown in kDa.



#### 4.5 Generation of $\alpha_1$ -PI P2P1 and $\alpha_1$ -PI M358R P7-P3 T7Select10-3b Libraries

Mutagenesis required to generate both the  $\alpha_1$ -PI P2P1 and  $\alpha_1$ -PI M358R P7-P3 libraries was accomplished using PCR mutagenesis to introduce every possible codon combination at desired sites: amino acids 357 and 358 for the P2P1 library, and amino acids 352 through to 356 for the P7-P3 library. The resulting  $\alpha_1$ -PI P2P1 library therefore contained  $21^2$  or 441 possible mutants, while the  $\alpha_1$ -PI M358R P7-P3 library contained  $21^5$  or 4,084,101 possible mutants. The pUC19 plasmid in *E. coli* was used to amplify the amount of DNA to assist with insertion into T7Select10-3b bacteriophage. The supplier of the degenerate primer noted that the manufacturing process introduced a nucleotide bias within the degenerate region, with a T>G>C>A ratio of approximately 30:26:24:20, instead of the natural 25:25:25:25. As this was the reverse primer, A would be expected to occur 10% more often within codons than T. The potential amino acid frequency is summarised in **Figure 11A**, with arginine expected to represent the most codons with a frequency of 10%, and tryptophan expected to represent the least with a frequency of 1%. The successful creation of each library was determined by DNA sequencing prior to biopanning, which indicated that the correct mutations were introduced (**Figure 11B**, **11C**). Amino acids proline and leucine were observed in higher frequencies than expected in the sampled sequences, suggesting a bias for codons containing C instead of the expected frequencies noted in **Figure 11A**. No contamination with  $\alpha_1$ -PI DNA sequences commonly used in the laboratory was also confirmed, as the silent mutation introduced at amino acid P361 was specific to the phage library construction scheme: CCC to CCA for the P2P1 library, and CCC to CCG for the P7-P3

**Figure 11: Confirmatory Sequencing of  $\alpha_1$ -PI Libraries** A) **Expected Amino Acid Frequency in Degenerate Region.** Based on a T>G>C>A ratio of approximately 30:26:24:20 within the degenerate region of the reverse primers used to construct the libraries. B) **Randomly Selected  $\alpha_1$ -PI P2P1 Sequences before Biopanning.** Five T7Select10-3b plaques containing the  $\alpha_1$ -PI P2P1 insert each had a unique P2P1 sequence. C) **Randomly Selected  $\alpha_1$ -PI M358R P7-P3 Sequences before Biopanning.** Ten pUC19 plasmids containing the  $\alpha_1$ -PI M358R P7-P3 insert each had a unique P7-P3 sequence. Clone 1 and 9 contained a mixed signal within the degenerate region but the introduced C to G silent mutation at P361 was observed, indicative of the correct library construction. Residues coloured based on their physiochemical properties (57).

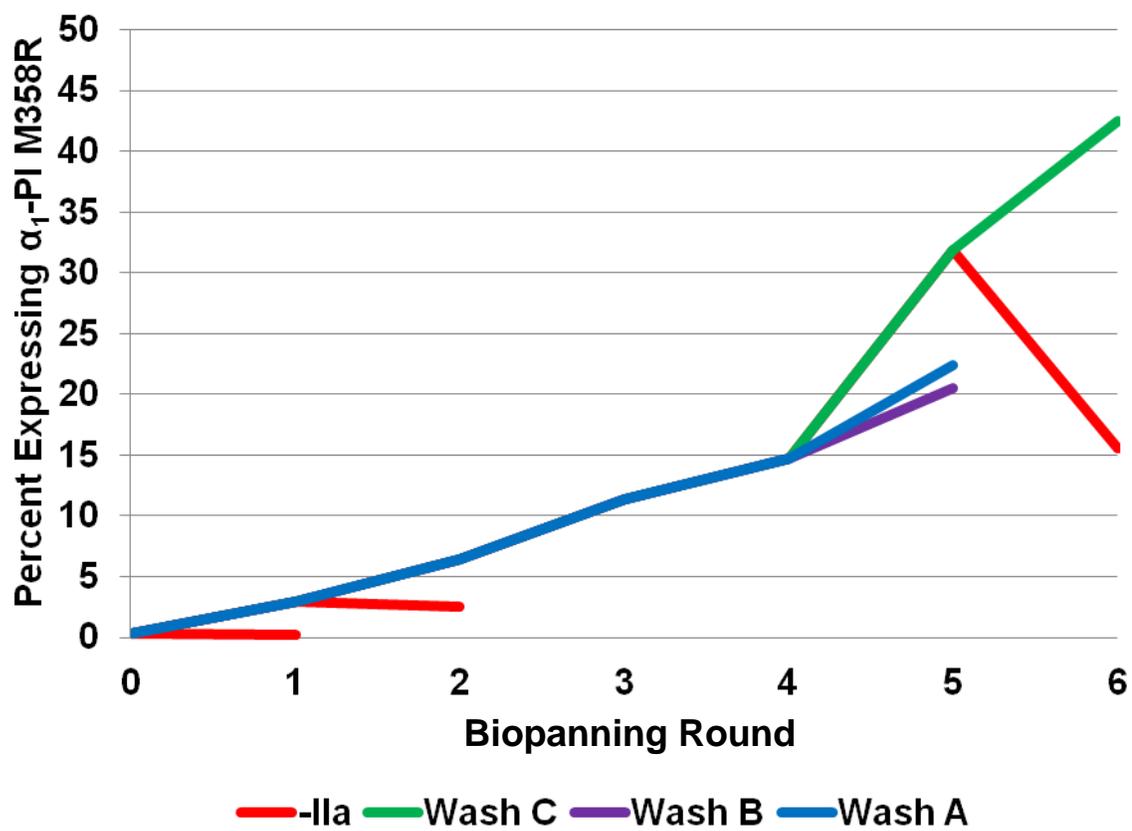
<b>A</b>		<b>B</b>			<b>C</b>							
<b>W</b>	0.012	<b>Clone</b>	<b>P2</b>	<b>P1</b>	<b>Clone</b>	<b>P7</b>	<b>P6</b>	<b>P5</b>	<b>P4</b>	<b>P3</b>	<b>Notes</b>	
<b>Y</b>	0.028	<b>P2P1</b>	<b>1</b>	<b>R</b>	<b>P7-P3</b>	<b>1</b>	<b>?</b>	<b>?</b>	<b>F</b>	<b>?</b>	<b>?</b>	Correct P361 mutation
<b>F</b>	0.018	<b>P2P1</b>	<b>2</b>	<b>D</b>	<b>P7-P3</b>	<b>2</b>	<b>S</b>	<b>P</b>	<b>H</b>	<b>S</b>	<b>F</b>	
<b>L</b>	0.074	<b>P2P1</b>	<b>3</b>	<b>P</b>	<b>P7-P3</b>	<b>3</b>	<b>T</b>	<b>A</b>	<b>L</b>	<b>H</b>	<b>V</b>	
<b>I</b>	0.046	<b>P2P1</b>	<b>4</b>	<b>P</b>	<b>P7-P3</b>	<b>4</b>	<b>N</b>	<b>R</b>	<b>I</b>	<b>L</b>	<b>T</b>	
<b>V</b>	0.048	<b>P2P1</b>	<b>5</b>	<b>N</b>	<b>P7-P3</b>	<b>5</b>	<b>T</b>	<b>P</b>	<b>L</b>	<b>T</b>	<b>Y</b>	
<b>M</b>	0.014				<b>P7-P3</b>	<b>6</b>	<b>V</b>	<b>Q</b>	<b>V</b>	<b>N</b>	<b>G</b>	
<b>A</b>	0.062				<b>P7-P3</b>	<b>7</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>D</b>	
<b>G</b>	0.058				<b>P7-P3</b>	<b>8</b>	<b>Q</b>	<b>C</b>	<b>T</b>	<b>L</b>	<b>L</b>	
<b>C</b>	0.022				<b>P7-P3</b>	<b>9</b>	<b>P</b>	<b>?</b>	<b>?</b>	<b>?</b>	<b>R</b>	Correct P361 mutation
<b>P</b>	0.068				<b>P7-P3</b>	<b>10</b>	<b>L</b>	<b>Y</b>	<b>Y</b>	<b>P</b>	<b>S</b>	
<b>T</b>	0.078											
<b>S</b>	0.085											
<b>Q</b>	0.042											
<b>N</b>	0.041											
<b>E</b>	0.039											
<b>D</b>	0.033											
<b>H</b>	0.036											
<b>K</b>	0.049											
<b>R</b>	0.101											
<b>-</b>	0.047											

library. P7-P3 Clone 1 and 9 contained a mixed signal but the introduced C to G silent mutation at P361 was observed, indicative of correct library construction.

#### **4.6 Mock-Biopanning T7Select10-3b $\alpha_1$ -PI M358R**

A novel protocol for biopanning was developed in order to select only the T7Select10-3b phage which expressed functional  $\alpha_1$ -PI M358R, out of a background of inactive T7Select10-3b Control Phage (1/300 mix). The protocol centered on the use of a biotinylated antibody to prothrombin (that also reacted with thrombin) and streptavidin-coated magnetic beads, so that in theory only the phage which formed a stable serpin-enzyme complex with thrombin would be retrieved and amplified. As a proof of concept, the 1/300 mixture of  $\alpha_1$ -PI M358R phage to inactive control phage was biopanned for 5 rounds, with or without the addition of thrombin for the first two rounds. The biopanning procedure ultimately resulted in a successful enrichment in the percentage of T7Select10-3b  $\alpha_1$ -PI M358R in the mixture, approximately 4% per round as determined by immunoblotted plaque lift (**Figure 12**). No enrichment was observed when the phage mixture was biopanned without the addition of thrombin. In order to improve enrichment and to remove a greater percentage of non-specific phage, the protocol was refined for the 5<sup>th</sup> round of biopanning. The use of a more stringent washing protocol, the “Wash C” method, improved biopanning significantly resulting in an approximate enrichment of 15% in only a single round. A 6<sup>th</sup> round of biopanning using the Wash C method confirmed that it was able to decrease non-specific binding compared to the original washing methods, and therefore was chosen as the biopanning method for the P2P1 and P7-P3 libraries.

**Figure 12: Enrichment of  $\alpha_1$ -PI M358R during Mock-Biopanning.** The amount of phage expressing the  $\alpha_1$ -PI M358R protein was enriched after each round of biopanning with thrombin, as detected by immunoblotted plaque lift. The number of plaques expressing  $\alpha_1$ -PI M358R is represented as a percentage of the total plaques. Biopanning without thrombin (-IIa) resulted in no enrichment. Refinement of the washing methods increased the enrichment in rounds 5 and 6, while decreasing the amount of non-specific binding.



## **4.7 Determining Optimal P2 and P1 Residues for Thrombin Inhibition**

### **4.7.1 Sequencing Plaques from Biopanned $\alpha_1$ -PI P2P1 Library**

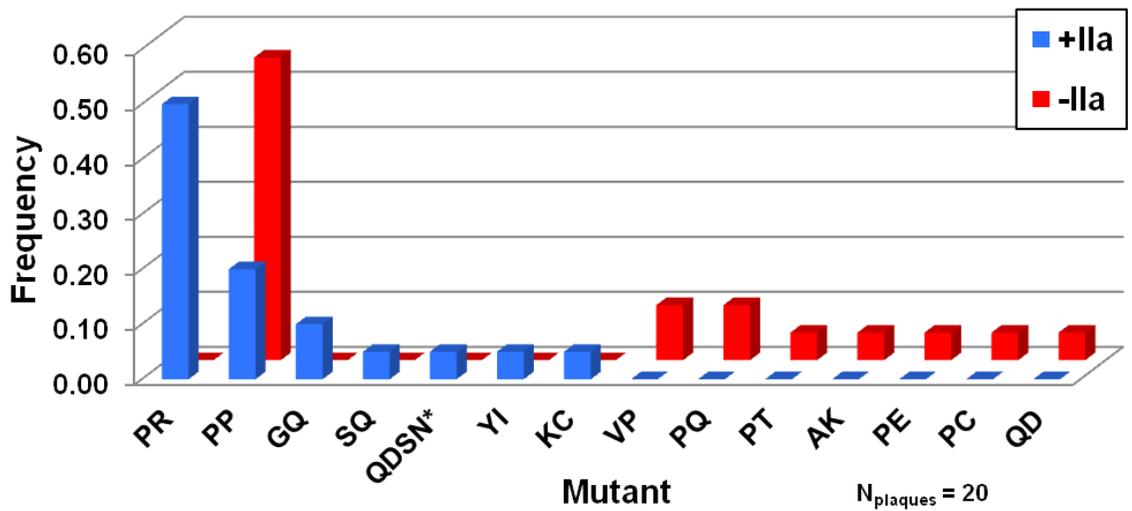
Following 5 rounds of biopanning the T7Select10-3b  $\alpha_1$ -PI P2P1 library either with thrombin (Rd5 +IIa) or without thrombin (Rd5 –IIa), the RCL sequences from 20 randomly selected plaques from each biopanned library were determined (**Figure 13**). The P2P1 mutant P357/M358R, which is identical to  $\alpha_1$ -PI M358R, appeared with a 50% frequency in the Rd5 +IIa plaques and did not appear in any Rd5 –IIa plaques. Codon usage varied between the P357/M358R sequences. The P2P1 mutant P357/M358P appeared with a 20% frequency in the Rd5 +IIa sequenced plaques, and in 55% of Rd5 –IIa plaques. The remaining novel P2P1 sequences did not appear in more than 10% of sequenced plaques and no other sequence was found to overlap between the two biopanned libraries. Mutant “QDSN” appeared due to an unintended point mutation introduced at P2', resulting in the mutation L360N at this position in addition to the novel P2P1 sequence.

### **4.7.2 ELISA-Based Lysate Screen of Unique $\alpha_1$ -PI P2P1 Sequences**

An ELISA-based screen was employed to semi-quantitatively assess the ability of each  $\alpha_1$ -PI P2P1 mutant to bind to immobilized thrombin. Each unique mutant identified by sequencing plaques was inserted into an arabinose inducible *E. coli* expression plasmid. Following induction of  $\alpha_1$ -PI P2P1 production, cultures were lysed and the lysate incubated directly on thrombin-coated microtiter wells. Unbound protein was washed away and remaining  $\alpha_1$ -PI P2P1 protein was detected using an HRP-conjugated antibody.

**Figure 13: Frequencies of  $\alpha_1$ -PI P2P1 Sequences after Biopanning.** After biopanning the T7Select10-3b  $\alpha_1$ -PI P2P1 library for five rounds either with thrombin (+IIa) or without thrombin (-IIa) 20 plaques were sequenced from each biopanned library. The P2P1 mutant “PR”, which is identical to  $\alpha_1$ -PI M358R, appeared in 50% of plaques biopanned with thrombin, and did not appear in the library biopanned without thrombin. The P2P1 mutant “PP” appeared in 55% of plaques biopanned without thrombin, and 20% of plaques biopanned with thrombin. Mutant “QDSN” was the result of a point mutation at the P2’ codon, resulting in a mutant P2P1 and P2’ sequence, while maintaining the wild type P1’ serine residue.

### Sequenced P2P1 Mutants After Biopanning



N<sub>plaques</sub> = 20

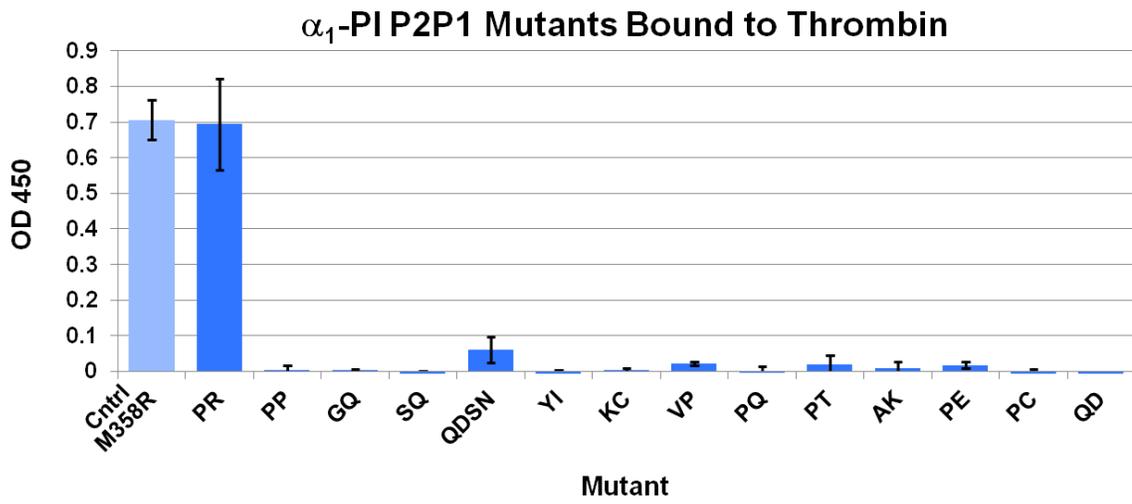
\*Point mutation at P2'

Only the P357/M358R mutant protein was able to bind to the thrombin coated wells, with an identical efficiency as separately produced  $\alpha_1$ -PI M358R lysate (**Figure 14**).

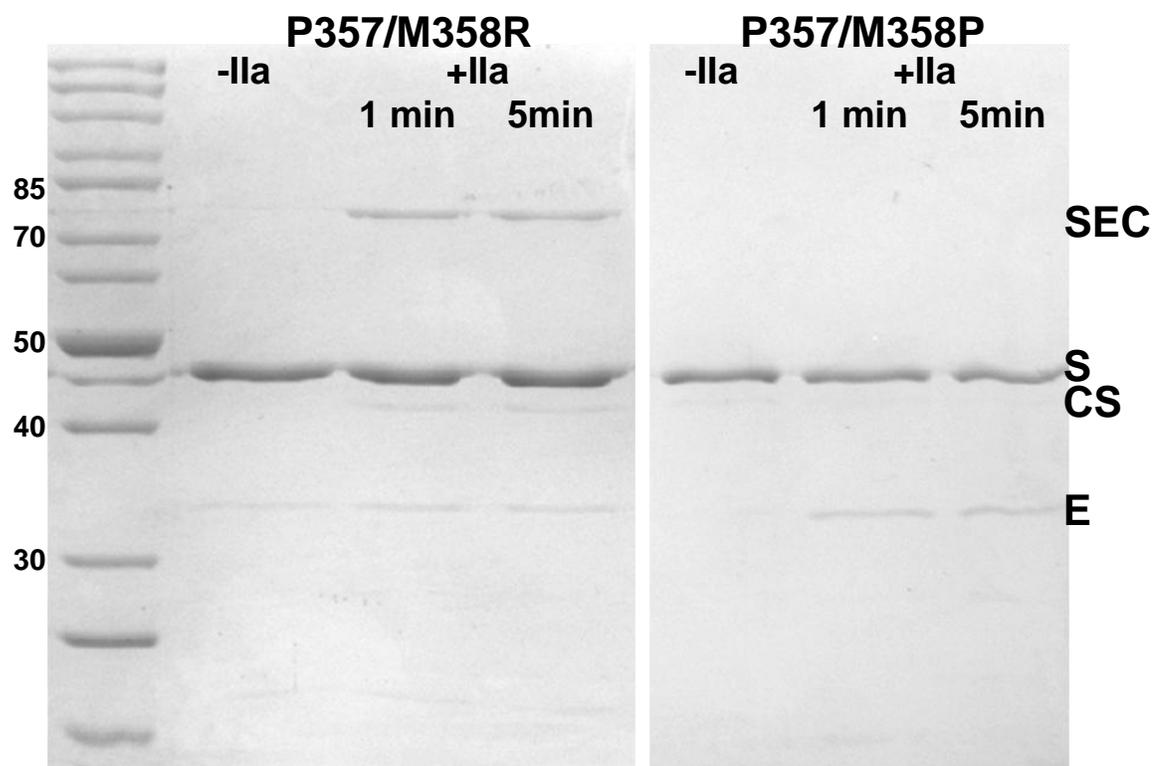
#### 4.7.3 Gel Based Serpin-Enzyme Complex Assays

The P2P1 mutants P357/M358R and P357/M358P were selected for further characterisation based on their frequency in sequenced plaques. To determine if these mutants were capable of forming a SDS-stable serpin-enzyme complex with thrombin, DNA was amplified from phage and inserted into an arabinose inducible *E. coli* expression plasmid. It should be noted that the mutant P357/M358R has an identical protein sequence to  $\alpha_1$ -PI M358R but the codons used at the mutant positions were unique to the phage-derived sample. Recombinant P357/M358R and P357/M358P mutant proteins were then purified using a nickel affinity column and a DEAE sepharose column. Equimolar amounts of each serpin were incubated with thrombin, and complex formation was assessed via SDS-PAGE electrophoresis. Serpin-enzyme complexes were observed between the P357/M358R mutant and thrombin, while no complex formation was observed in reactions with the P357/M358P mutant (**Figure 15**). As such, the optimal  $\alpha_1$ -PI P2P1 sequence for thrombin inhibition was determined as P357/M358R, which had previously been identified as reasonably rapid thrombin inhibitor and was known to match the substrate specificity of thrombin (21, 90).

**Figure 14: Amount of  $\alpha_1$ -PI P2P1 Mutant Protein Bound to Thrombin.** *E. coli* cells containing expression plasmids for each corresponding mutant were lysed, and exposed to thrombin coated wells. The relative amount of each serpin protein bound to thrombin is represented as an OD<sub>450</sub> value on the y-axis (i.e. amount of anti- $\alpha_1$ -PI antibody bound). Only the “PR” mutant, which is identical to  $\alpha_1$ -PI M358R, was found to bind to thrombin effectively. Mutant “QDSN” was the result of a point mutation at the P2’ codon, resulting in a mutant P2P1 and P2’ sequence, while maintaining the wild type P1’ serine residue. Each mutant was tested in duplicate, error bars represent SD.



**Figure 15: Serpin-Enzyme Complexes between  $\alpha_1$ -PI P2P1 mutant P357/M358R and Thrombin.** P2P1 mutants P357/M358R and P357/M358R (at 1  $\mu$ M) were incubated with thrombin (indicated as “IIa” at 0.2  $\mu$ M) for one minute or five minutes at 37°C before the reaction was stopped with 4x SDS dye. These samples were then electrophoresed on a 10% SDS-PAGE gel and stained with Coomassie Blue. SEC, serpin-enzyme complex; S, unreacted serpin; CS, cleaved serpin; E, enzyme. Molecular weight markers are shown and labeled in kDa.



## 4.8 Determining Optimal P7 to P3 Residues for Thrombin Inhibition

### 4.8.1 Sequencing Plaques from Biopanned $\alpha_1$ -PI M358R P7-P3 Library

The optimal P2 and P1 residues for  $\alpha_1$ -PI's ability to inhibit thrombin were found to be proline and arginine respectively, as determined by biopanning the  $\alpha_1$ -PI P2P1 library. These residues were therefore maintained in the P7-P3 library (T7Select10-3b  $\alpha_1$ -PI M358R P7-P3). Following 5 rounds of biopanning the T7Select10-3b  $\alpha_1$ -PI M358R P7-P3 library, with thrombin (Rd5 +IIa) or without thrombin (Rd5 -IIa), 40 plaques were randomly selected from each of the Rd5 +IIa and Rd5 -IIa libraries. The P7-P3 sequences within these plaques were determined in order to look for trends, such as certain sequences appearing frequently and conservation of physiochemical properties between residues at a certain position (**Figure 16**). The sequences from the Rd5 +IIa library were highly diverse with 33 unique sequences out of 40 plaques, fitting the general trend: P7-Not Aromatic/P6-Hydrophobic/P5-T or S/P4-Hydrophobic/P3-Not Aromatic. This matches thrombin's known specificity and also fits the general serpin trend to have hydrophobic residues at the P6 and P4 positions (21, 26). Named after the P7-P3 sequence, the following mutants appeared in more than one sequenced plaque: DITMA, DAFVT, QPPPS, PLFVS, SLELK, NLIPT. A consensus sequence based on the amino acid appearing most frequently at a given position was found to be DLTVS, although it did not appear in any of the sequenced plaques.

The sequences from the Rd5 -IIa library had poor diversity, with only 10 unique sequences out of 40 plaques, and no discernable trends were observed. One sequence, PKSEG, was also found in the Rd5 +IIa sequenced plaque samples.

**Figure 16: P7-P3 Sequences from Plaques following Biopanning** A) **Randomly Selected T7Select10-3b  $\alpha_1$ -PI M358R P7-P3 Plaque Sequences after Biopanning with Thrombin.** 33 unique sequences were found out of 40 plaques. A total of 53 plaques were sequenced, with nine PCR or sequencing failures, three plaques containing the original M358R sequence identified due to the unmodified P361 codon, and one plaque containing a sequence with a stop codon. Consensus sequence: DLTVS B) **Randomly Selected T7Select10-3b  $\alpha_1$ -PI M358R P7-P3 Plaque Sequences after Biopanning without Thrombin.** 10 unique sequences were found out of 40 plaques. A total of 43 plaques were sequenced, with three PCR or sequencing failures. Plaques numbered in order out of the total number of plaques screened. Note that these numbers do not match sequences in Table 2. Amino acid sequences were aligned using ClustalW 2.0 and displayed using Jalview 2.8 (55, 56). Residues coloured based on their physiochemical properties (57).

**A**

## Rd5 +IIa

Sequences 1 - 53

Aligned via ClustalW2

N = 40

9 PCR or sequencing failures

3  $\alpha_1$ -PI M358R contaminants

1 with stop codon

	7	5	3	11'
$\alpha_1$ -PI M358R	GTEAAGAMFLEAIPRSIPP			
+2/1-5	DITMA			
+16/1-5	DITMA			
+5/1-5	DAFVT			
+38/1-5	DAFVT			
+26/1-5	DAFAA			
+31/1-5	TAHVT			
+17/1-5	QATFL			
+44/1-5	EAHFR			
+12/1-5	QPPPS			
+19/1-5	QPPPS			
+3/1-5	HPPSL			
+6/1-5	PKSEG			
+4/1-5	DATVS			
+46/1-5	DVTVS			
+32/1-5	HATIS			
+50/1-5	HATVS			
+25/1-5	YATLS			
+27/1-5	TL SAV			
+8/1-5	PLFVS			
+23/1-5	PLFVS			
+9/1-5	LVFVS			
+43/1-5	PLQLS			
+22/1-5	SLELK			
+24/1-5	SLELK			
+7/1-5	EASLI			
+52/1-5	SLAMT			
+13/1-5	ELLAA			
+33/1-5	ELLAL			
+14/1-5	QLTAT			
+15/1-5	NLIPT			
+34/1-5	NLIPT			
+35/1-5	NLIPT			
+45/1-5	LLPYS			
+18/1-5	PLAPI			
+20/1-5	QRPHQ			
+41/1-5	RYHYI			
+36/1-5	PAMPK			
+42/1-5	ARSST			
+48/1-5	WNPVI			
+53/1-5	NECAI			

**B**

### Rd5 -IIa

Sequences 1 - 43

Aligned via ClustalW2

N = 40

3 PCR or sequencing failures

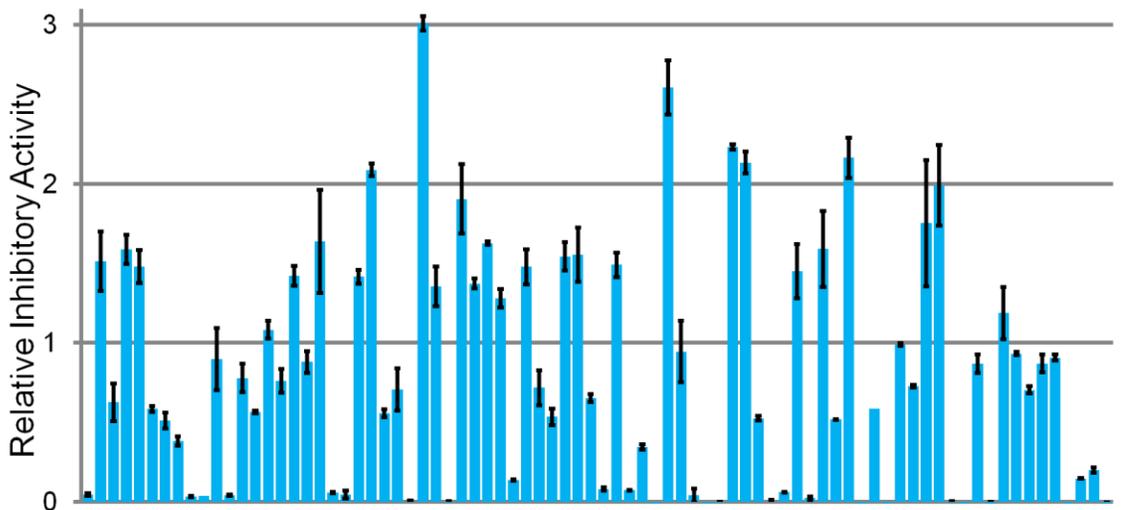
	7	5	3	11'
$\alpha_1$ -PI M358R	GTEAAGAMFLEAIPRSIPP			
-40/1-5	PPAPL			
-43/1-5	PPAPL			
-38/1-5	PPAPL			
-34/1-5	PPAPL			
-26/1-5	PPAPL			
-25/1-5	PPAPL			
-22/1-5	PPAPL			
-13/1-5	PPAPL			
-9/1-5	PPAPL			
-8/1-5	PPAPL			
-5/1-5	PPAPL			
-4/1-5	PPAPL			
-2/1-5	PPAPL			
-17/1-5	PKSEG			
-36/1-5	PKSEG			
-3/1-5	SRHLP			
-31/1-5	RLLVK			
-18/1-5	PFLMH			
-30/1-5	PFLMH			
-1/1-5	FGSNI			
-10/1-5	FGSNI			
-19/1-5	FGSNI			
-20/1-5	FGSNI			
-23/1-5	FGSNI			
-24/1-5	FGSNI			
-27/1-5	FGSNI			
-28/1-5	FGSNI			
-29/1-5	FGSNI			
-32/1-5	FGSNI			
-39/1-5	FGSNI			
-42/1-5	FGSNI			
-6/1-5	RTFIN			
-21/1-5	RTFIN			
-33/1-5	RTFIN			
-35/1-5	RTFIN			
-41/1-5	RTFIN			
-11/1-5	LTMTN			
-14/1-5	LTMTN			
-16/1-5	IGHTA			
-12/1-5	VPVVT			

#### 4.8.2 ELISA-Based Lysate Screen as a “6<sup>th</sup> Round” of Biopanning

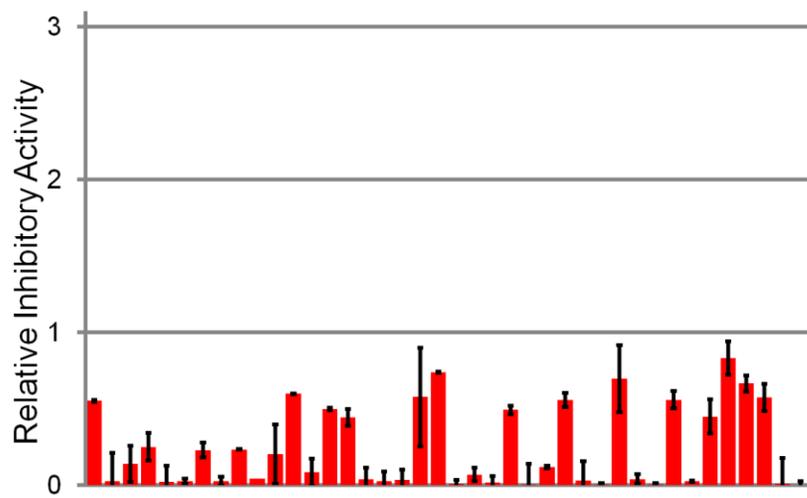
Due to the high sequence diversity remaining after five rounds of biopanning, an ELISA-based screen was employed to semi-quantitatively assess thrombin binding activity of randomly selected  $\alpha_1$ -PI M358R P7-P3 mutant proteins. To better represent the results of biopanning this larger library random sequences from the Rd5 +IIa and Rd5 – IIa were retrieved via PCR of the phage lysate rather than individual plaques, allowing for a greater number of candidate sequences to be screened. Sequences were inserted into an arabinose inducible *E. coli* expression plasmid. After transforming the plasmid libraries into competent *E. coli* cells, 80 colonies containing the Rd5 +IIa sequences and 40 colonies containing the Rd5 –IIa sequences were randomly selected. The lysate from these colonies was applied in duplicate to thrombin coated wells, incubated then washed to remove non-specific binding, before being probed with an anti- $\alpha_1$ -PI antibody to observe the amount of bound  $\alpha_1$ -PI M358R P7-P3 protein. Samples which initially indicated thrombin binding greater than  $\alpha_1$ -PI M358R were analysed again to confirm their activity. Out of the 80 Rd5 +IIa colonies screened, 27 were found to have greater thrombin binding than  $\alpha_1$ -PI M358R, while none of the 40 Rd5 –IIa colonies had thrombin binding activity greater than  $\alpha_1$ -PI M358R (**Figure 17**). Colonies with improved thrombin binding activity were then sequenced, yielding 22 unique P7-P3 sequences with greater activity than  $\alpha_1$ -PI M358R (**Table 2**). Six of these 22 sequences were previously found when sequencing random plaques from Rd5 +IIa, as indicated in Section 4.8.1. Based on this semi-quantitative ELISA-based screen, the top sequence AAFVS indicated a level of thrombin binding 3-fold greater than  $\alpha_1$ -PI M358R.

**Figure 17: Amount of  $\alpha_1$ -PI M358R P7-P3 Mutant Protein Bound to Thrombin Relative to  $\alpha_1$ -PI M358R. A) From Sequences Biopanned with Thrombin (Rd5 +IIa).** Out of 80 colonies 27 indicated an improved level of thrombin binding over  $\alpha_1$ -PI M358R. **B) From Sequences Biopanned without Thrombin (Rd5 –IIa).** None of the 40 selected samples had thrombin binding activity greater than  $\alpha_1$ -PI M358R. Each sample performed in duplicate, with samples that showed an initially greater level of thrombin binding repeated. Error bars represent SD.

**A**



**B**



**Table 2:  $\alpha_1$ -PI M358R P7-P3 Sequences Found to Have Greater Thrombin Binding Activity than  $\alpha_1$ -PI M358R.** OD<sub>450</sub> is the mean value out of two determinations, each performed in duplicate. “vs  $\alpha_1$ -PI M358R” is the relative amount of thrombin binding versus unmodified  $\alpha_1$ -PI M358R, and ranked out of 27. “40seq” notation marks whether the sequence appeared in the 40 randomly sequenced plaques; six out of the 22 unique sequences matched those found in plaques. Consensus sequence: LATVS. Colonies numbered in order out of the 80 Rd5 +IIa colonies screened. Note that these numbers do not match plaque sequences in Figure 16. Displayed using Jalview 2.8 and residues coloured based on their physiochemical properties (56, 57).

Colony	P7-P3	OD <sub>450</sub>	vs $\alpha_1$ -PI M358R	Rank	40seq
$\alpha_1$ -PI M358R	FLEAI	0.71	1.00	-	
+27	AAFVS	2.13	3.01	1	
+46	EISLQ	1.84	2.61	2	
+51	QATFL	1.58	2.23	3	*
+60	LASMR	1.53	2.16	4	
+52	LHTLG	1.51	2.13	5	
+23	LTTLR	1.48	2.09	6	
+67	HATVS	1.41	1.99	7	*
+30	VTTIT	1.35	1.90	8	
+66	EATVS	1.24	1.75	9	
+19	DITMA	1.16	1.64	10	*
+32	LASMR	1.15	1.63	11	
+58	QATFL	1.12	1.59	12	*
+4	DAFAA	1.12	1.59	13	*
+39	EATVS	1.10	1.55	14	
+38	LINPI	1.09	1.54	15	
+2	EAHFR	1.07	1.51	16	
+42	TVSVS	1.05	1.49	17	
+5	EATVS	1.05	1.48	18	
+35	ISSAN	1.04	1.48	19	
+56	LAITS	1.02	1.45	20	
+17	DVTVS	1.00	1.42	21	*
+22	ISSAN	1.00	1.42	22	
+31	QVKPA	0.97	1.37	23	
+28	TL SAV	0.96	1.36	24	*
+33	LSELA	0.91	1.28	25	
+72	LGSFT	0.84	1.19	26	
+15	RVNAK	0.77	1.08	27	

### **4.8.3 Analysis of Sequences with Improved Thrombin Binding**

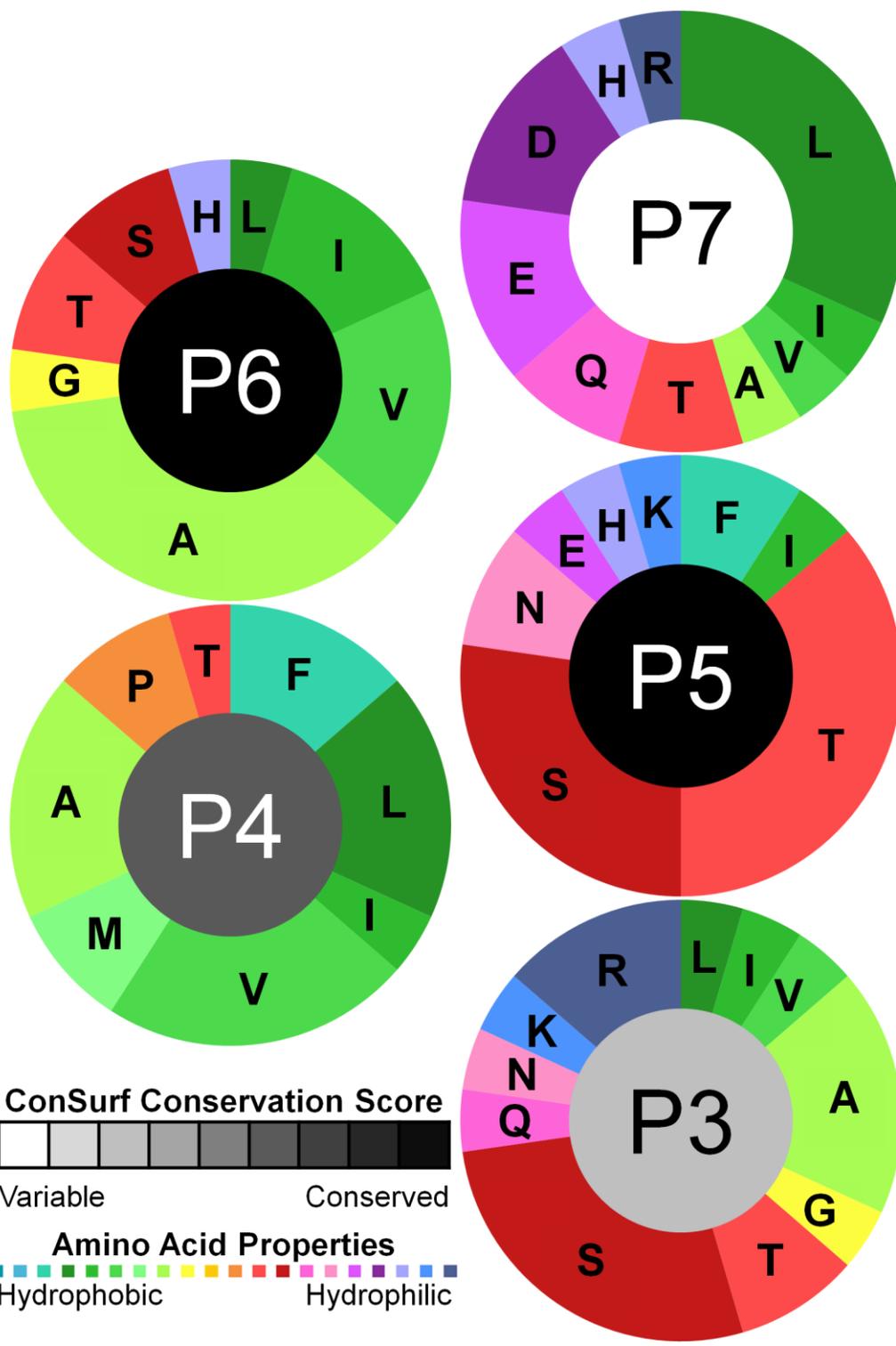
A trend within sequences with improved thrombin binding was observed, matching the trend initially observed in the 40 randomly sequenced Rd5 +IIa plaques; P7-Not Aromatic/P6-Hydrophobic/P5-T or S/P4-Hydrophobic/P3-Not Aromatic. A consensus sequence based on the amino acid appearing most frequently at a given position was found to be LATVS, which varied slightly from the consensus sequence DLTVS determined from randomly sequenced Rd5 +IIa plaques. Again this consensus sequence did not appear in any of the sequenced plasmids. The 22 unique sequences with improved thrombin binding were uploaded to ConSurf (138). This bioinformatics program is typically used to help determine evolutionary conservation within proteins, indicative of a functional significance. For the purposes of this study, ConSurf scored each RCL position P7 through to P3 based on amino acid conservation within the sequences with improved thrombin binding (**Figure 18**). Positions P6 and P5 were found to be highly conserved, position P4 was found to be moderately conserved, position P3 was found to be moderately variable, and position P7 was found to be highly variable.

## **4.9 Functional Characterisation of $\alpha_1$ -PI M358R P7-P3 Mutants of Interest**

### **4.9.1 Selecting Mutants of Interest**

$\alpha_1$ -PI M358R P7-P3 Mutants AAFVS and EISLQ were selected solely based on the results of the ELISA-based screen, as they indicated the best and second best thrombin binding activity respectively. Mutants DITMA, EATVS, and HATVS were selected based on both their increased activity versus  $\alpha_1$ -PI M358R in the ELISA-based screen, and that they appeared within the sequenced Rd5 +IIa plaques several times, or

**Figure 18: Frequencies of Each Amino Acid for  $\alpha_1$ -PI M358R P7-P3 Sequences found to have Greater Thrombin Binding Activity than  $\alpha_1$ -PI M358R.** The outside circles represent the number of times a designated amino acid appeared at the indicated RCL position, while the interior circle represents the level of conservation at each site, as determined by ConSurf (138). The general trend required for optimal thrombin inhibition was determined as: P7-Not Aromatic/P6-Hydrophobic/P5-T or S/P4-Hydrophobic/P3-Not Aromatic. Positions P6 and P5 were found to be highly conserved, position P4 was found to be moderately conserved, position P3 was found to be moderately variable, and position P7 was found to be highly variable. 22 unique sequences, with residues coloured based on their physiochemical properties (57).



were similar to a sequence which was also represented in this group. EATVS did not appear in Rd5 +IIa sequenced plaques, but appeared twice in colonies with improved thrombin binding. Additionally, EATVS only differed from HATVS at the P7 position and therefore offered an opportunity to determine the functional significance of this RCL position.

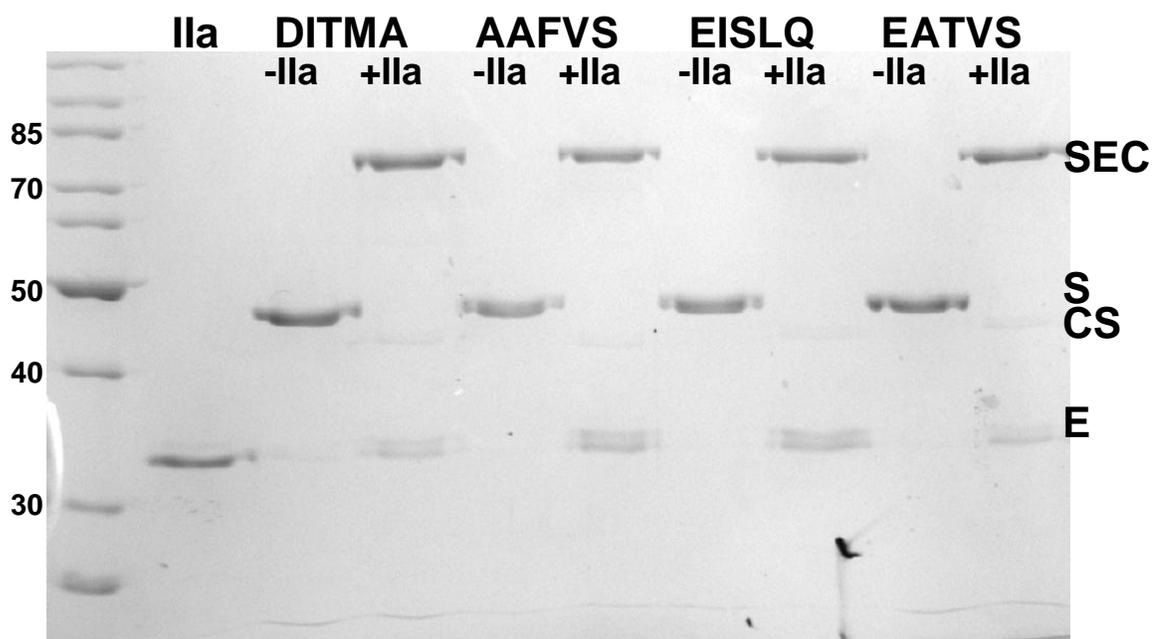
#### **4.9.2 Gel Based Serpin-Enzyme Complex Assays**

Grown using the same colonies analysed in the ELISA-based lysate assay, recombinant  $\alpha_1$ -PI M358R P7-P3 mutants DITMA, AAFVS, EISLQ, EATVS and HATVS were purified using a nickel affinity column and a DEAE sepharose column (from colonies +19, +27, +46, +66 and +67 respectively). The total purified protein retrieved from a 1 L culture of each  $\alpha_1$ -PI M358R P7-P3 mutant, as determined by the extinction coefficient method, were as follows: DITMA 0.53 mg/L of bacterial culture, AAFVS 0.72 mg/L, EISLQ 0.60 mg/L, EATVS 0.70 mg/L, HATVS 0.60 mg/L. Each purified serpin was incubated individually with thrombin for one minute, and complex formation was assessed via SDS-PAGE electrophoresis. Serpin-enzyme complexes were observed between each of the  $\alpha_1$ -PI M358R P7-P3 mutants and thrombin, which provided a qualitative measure of their effectiveness as thrombin inhibitors (**Figure 19**; HATVS data not shown). A faint band which appeared just below the ~45 kDa serpin band suggested cleaved serpin product, although the majority of serpin protein formed a stable serpin-enzyme complex.

#### **4.9.3 Kinetic Analysis of $\alpha_1$ -PI M358R P7-P3 Mutants of Interest**

The second order rate constants ( $k_2$ ) of thrombin inhibition were determined for

**Figure 19: Formation of Serpin-Enzyme Complexes between  $\alpha_1$ -PI M358R P7-P3 Mutants and Thrombin.** Each serpin (at 1  $\mu$ M) was incubated with thrombin (indicated as “IIa” at 0.5  $\mu$ M) for one minute at 37°C before the reaction was stopped with 4x SDS dye. These samples were then electrophoresed on a 10% SDS-PAGE gel and stained with Coomassie Blue. SEC, serpin-enzyme complex; S, unreacted serpin; CS, cleaved serpin; E, enzyme. Molecular weight markers are shown and labeled in kDa.



each of the  $\alpha_1$ -PI M358R P7-P3 mutants of interest and for unmodified  $\alpha_1$ -PI M358R (**Table 3**). The  $\alpha_1$ -PI M358R P7-P3 mutants DITMA and AAFVS were each found to have an approximately 2-fold enhancement in the rate of reaction with thrombin, as compared to  $\alpha_1$ -PI M358R. The mutant EISLQ was found to have a 1.6 fold improvement over  $\alpha_1$ -PI M358R, while mutants EATVS and HATVS did not have significantly increased rates of thrombin inhibition. The second order rate constant determined independently for  $\alpha_1$ -PI M358R in this study matches previous published data (97).

The stoichiometry of inhibition (SI) was determined for each of the  $\alpha_1$ -PI M358R P7-P3 mutants of interest and for unmodified  $\alpha_1$ -PI M358R (**Table 4**). This value provides a measure of how many moles of serpin are required to inhibit one mole of thrombin (26). The SI values for mutants DITMA and AAFVS were found to be significantly lower than the SI of  $\alpha_1$ -PI M358R, approaching the theoretical minimum of 1, while the other selected mutants did not show a significant difference compared to  $\alpha_1$ -PI M358R.

#### **4.10 Molecular Modeling of $\alpha_1$ -PI Interactions with Thrombin**

Based on the crystal structures of serpin-enzyme encounter complexes, computer assisted modeling was used to help identify interactions between  $\alpha_1$ -PI mutants and thrombin. As  $\alpha_1$ -PI has not been crystallised with thrombin, models were built based upon the closest matching encounter complexes:  $\alpha_1$ -PI M358R with S195A trypsin (PDB ID 1OPH); and HCII with S195A thrombin (PDB ID 1JMO). Models were then minimised using ZMM, which revealed the interactions between  $\alpha_1$ -PI's RCL residues P7-P3' with both thrombin and other  $\alpha_1$ -PI residues (**Figure 20** details P7-P3 interactions).

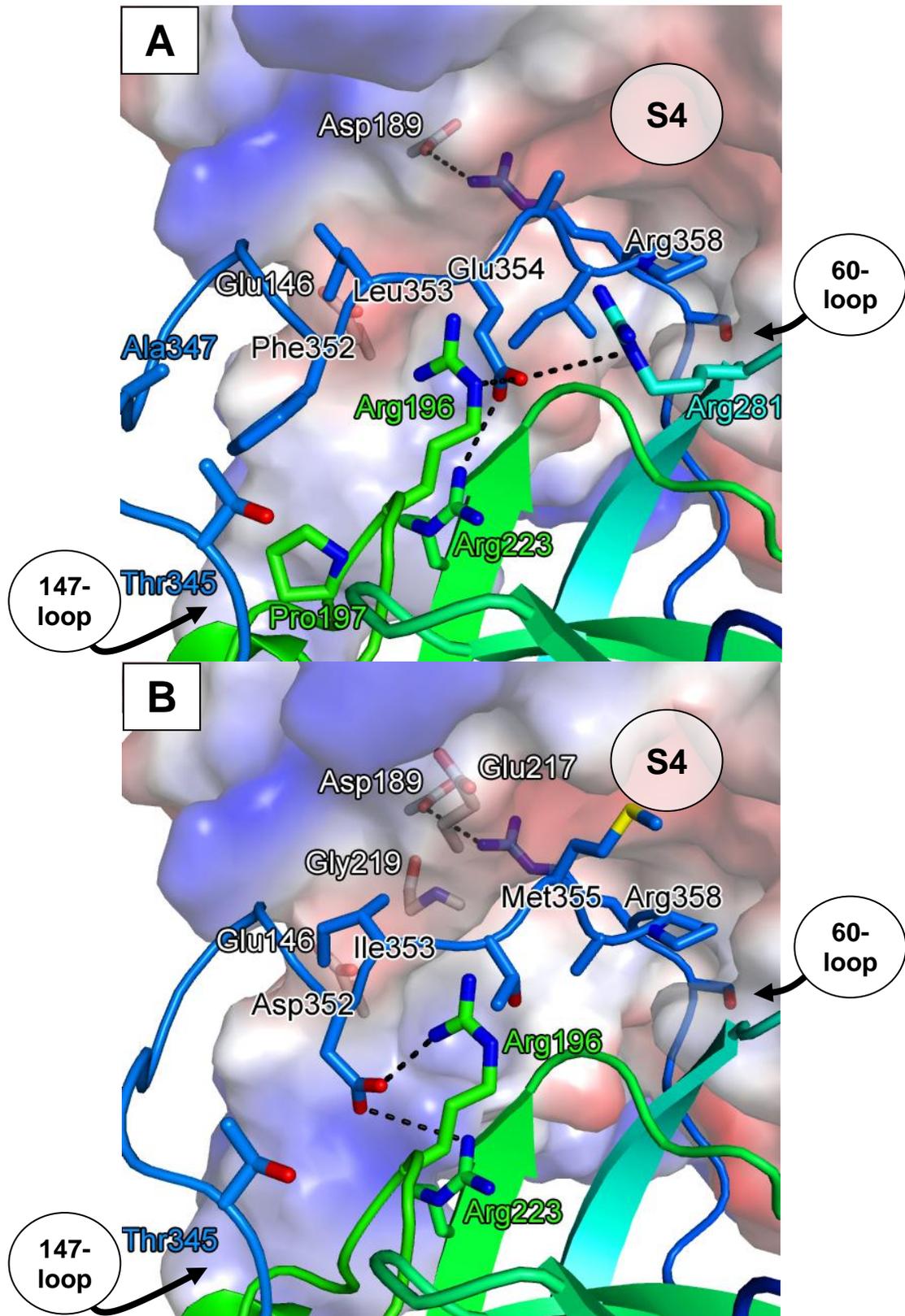
**Table 3: Second Order Rate Constant of Thrombin Inhibition.** The mean  $\pm$  SD of 5 determinations is reported. Statistical significance compared to  $\alpha_1$ -PI M358R was determined by Tukey-Kramer Multiple Comparisons test. “ns” not significant. The rate for wild type  $\alpha_1$ -PI was determined by Sutherland *et al.* (2006) (105).

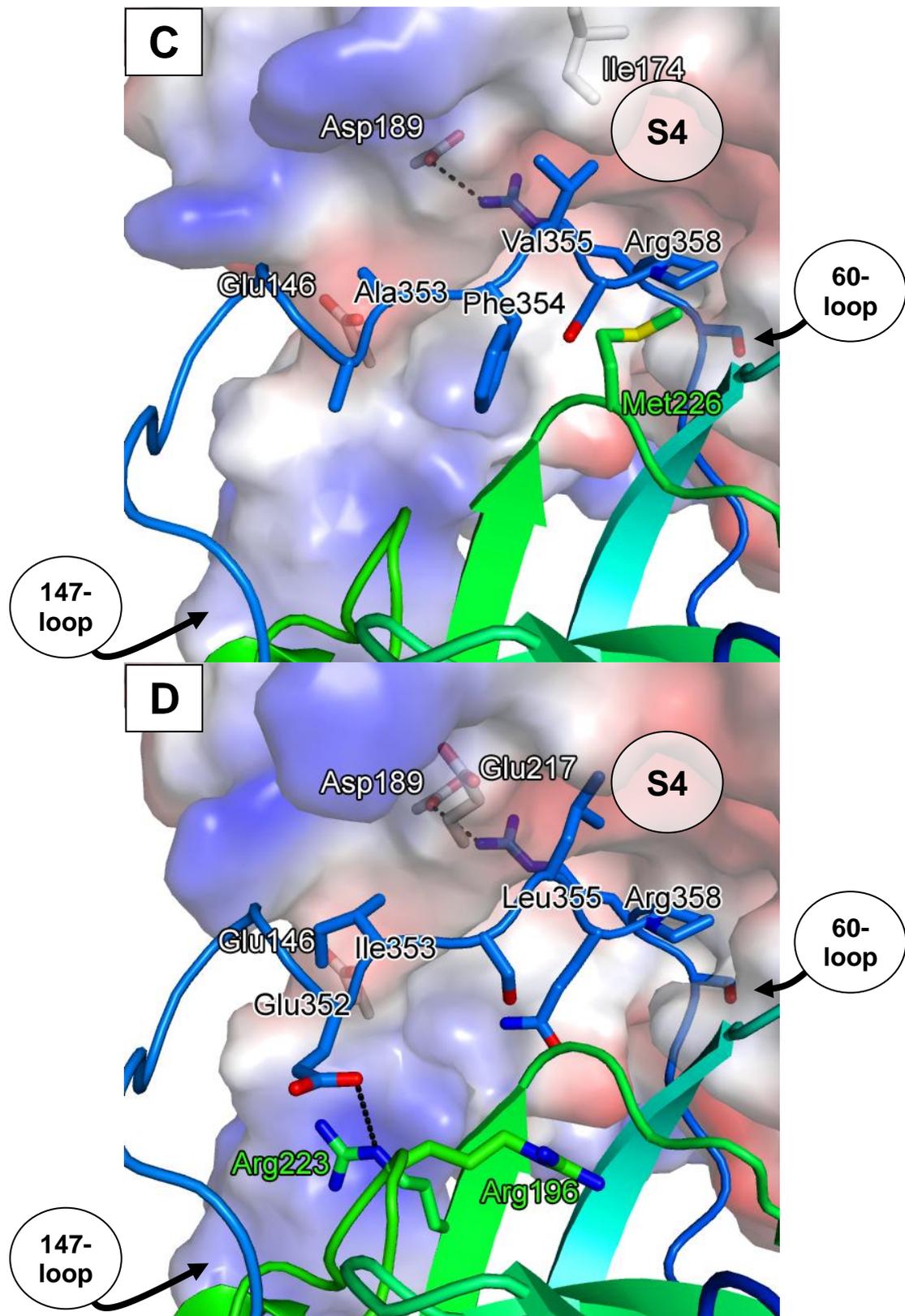
	<b>Protein</b>	<b><math>k_2</math> (<math>\times 10^5 \text{ M}^{-1} \text{ s}^{-1}</math>)</b>	<b>Significance</b>
	$\alpha_1$ -PI	$0.0005 \pm .00003$	-
	$\alpha_1$ -PI M358R	$4.86 \pm 0.43$	-
$\alpha_1$ -PI M358R P7-P3	DITMA	$10.0 \pm 0.68$	*** P<0.001
	AAFVS	$10.1 \pm 0.73$	*** P<0.001
	EISLQ	$7.83 \pm 0.93$	*** P<0.001
	EATVS	$6.16 \pm 0.52$	ns
	HATVS	$5.03 \pm 0.85$	ns

**Table 4: Stoichiometry of Inhibition with Thrombin.** The mean  $\pm$  SD of 5 determinations is reported. Statistical significance compared to  $\alpha_1$ -PI M358R was determined by Tukey-Kramer Multiple Comparisons test. “ND” not determined, “ns” not significant.

	<b>Protein</b>	<b>SI</b>	<b>Significance</b>
	$\alpha_1$ -PI	ND	-
	$\alpha_1$ -PI M358R	2.00 $\pm$ 0.03	-
$\alpha_1$ -PI M358R P7-P3	DITMA	1.57 $\pm$ 0.11	*** P<0.001
	AAFVS	1.74 $\pm$ 0.17	* P<0.05
	EISLQ	1.97 $\pm$ 0.17	ns
	EATVS	1.86 $\pm$ 0.12	ns
	HATVS	ND	-

**Figure 20: Energy Minimised Encounter Complexes between  $\alpha_1$ -PI Variants and Thrombin.**  $\alpha_1$ -PI RCL residues P7-P1' (352-359) and all other residues contributing  $\geq 0.5$  kcal mol<sup>-1</sup> to stabilising RCL residues P7-P3 (352-356) are shown as sticks. P7-P3 residues which contribute to stabilising interactions are labelled in black font. Dashed lines indicate salt bridges. Arg358 salt bridge with Asp189 is shown for reference.  $\alpha_1$ -PI residues coloured based on distance from N terminus. Thrombin's surface is coloured based on electrostatic potential (red: negative, blue: positive). Structural features of thrombin are labeled. Each P6 residue (353) interacts with thrombin's Glu146 via van der Waals interactions. **A)  $\alpha_1$ -PI M358R.** Phe352 is stabilised by Pro197, Thr354 and Ala347 via van der Waals interactions. Glu354 forms salt bridges with Arg196, Arg223, and Arg281. **B) DITMA.** Asp352 forms a salt bridge with Arg196 and Arg223, and van der Waals interactions with Thr345. Met355 is stabilised by Glu217 and Gly219 via van der Waals interactions. **C) AAFVS.** Phe354 and Val355 are stabilised by Met226 and Ile174 respectively via van der Waals interactions. **D) EISLQ.** Glu352 forms a salt bridge with Arg223 and van der Waals interactions with Arg196. Leu355 is stabilised by Glu217 via van der Waals interactions. Generated using PyMOL (18).





The total free energies of the minimised  $\alpha_1$ -PI-thrombin encounter complexes were within 1000 kcal mol<sup>-1</sup> of the minimised HCII-thrombin encounter complex, indicating that although the HCII-thrombin complex was more stable, no significant steric clashes or defects occurred when the encounter complexes were generated (**Table 5**).

Treating RCL residues P7-P3' as the "ligand" and all other residues as the "receptor", of the  $\alpha_1$ -PI variants examined  $\alpha_1$ -PI M358R was found to have the lowest ligand-receptor energy indicating the most favourable interactions. This was primarily due to the Glu354 residue at RCL position P5, which granted at least a 5.9 kcal mol<sup>-1</sup> improvement in free energy compared to the  $\alpha_1$ -PI P7-P3 mutants. Glu354 inserted into a basic pocket facing away from thrombin, forming salt bridges with  $\alpha_1$ -PI residues Arg196, Arg223, and Arg281, matching the results of previous studies (144, 145). The P7-P3 mutants that were investigated had uncharged residues at this position, thus their P5 residues were not stabilised in this manner. However, mutants DITMA and EISLQ were found to interact with the same  $\alpha_1$ -PI basic pocket via their respective acidic residues at the P7 position. In contrast, the  $\alpha_1$ -PI M358R P7 residue Phe352 was stabilised primarily by van der Waals interactions while the AAFVS P7 residue Ala352 did not extend into the same area and consequently had a less favourable free energy. The P6 residues did not contribute significantly to the stability of the encounter complexes, interacting primarily with thrombin's Glu146 residue via van der Waals interactions. The total free energy at the P4 position was most favourable for  $\alpha_1$ -PI M358R, due to fewer steric clashes with residues within thrombin's hydrophobic S4 subsite. The AAFVS mutant's P4 Val355 residue interacted favourably with thrombin's Ile174 within the

**Table 5: Specific Energies of Serpin – Thrombin Encounter Complexes.** “Total” column is the total free energy ( $\Delta G$ ) of the encounter complex. “LR” columns are the ligand receptor energies, with the ligand defined as the indicated RCL residues (P7-P3’; P7-P3) and the receptor defined as all other residues. Individual columns for P7-P3 indicate the contribution of each residue to the ligand receptor energy. All values in kcal mol<sup>-1</sup>, coloured from most favourable (green) to least favourable (red) energies.

		$\Delta G$ (kcal mol <sup>-1</sup> )							
Protein	P7-P3 Seq	Total	LR P7-P3'	LR P7-P3	P7	P6	P5	P4	P3
HCII	TVGFM	-11776	-7.0	-2.7	3.5	-1.5	-0.2	-3.2	-1.3
$\alpha_1$ -PI M358R	FLEAI	-10722	-17	-13	-4.2	1.0	-6.1	-1.7	-2.3
$\alpha_1$ -PI M358R P7-P3	DITMA	-10693	-0.57	0.6	-5.1	4.3	0.8	1.5	-0.9
	AAFVS	-10701	4.1	4.1	0.0	1.8	1.1	2.3	-1.2
	EISLQ	-10767	-4.6	-2.3	-4.8	3.2	-0.2	0.6	-1.1

hydrophobic S4 subsite, although the total free energy at this position was not favourable due to steric clashes with thrombin's Glu217. The DITMA mutant's P4 Met355 residue however interacted favourably with Glu217 and Gly219 via van der Waals interactions. Despite the relative diversity of the P3 residues investigated (Met, Ile, Ser, Ala, Gln) the stability of the complex was not drastically affected, supporting previous findings that thrombin can accommodate a variety of residues at this position (17, 21, 22). RCL residues P2-P3' were also investigated, and primarily interacted with thrombin's 60-loop, Asp189, and Glu192 (refer to **Figure 2**). These interactions were also identified in previous studies, and mirror canonical thrombin-substrate interactions (17, 146, 147).

## 5. DISCUSSION

Previous studies have indicated that the serpin  $\alpha_1$ -PI M358R is a rapid inhibitor of thrombin and that mutations based on known thrombin inhibitory serpins can improve this activity. Due to the unique structural characteristics of the serpin inhibitory mechanism, the RCL sequence has evolved to be specific to a serpin-protease pair and therefore these borrowed sequences are likely sub-optimal. Limited mutagenesis within the functionally significant hypervariable region has left the structure-function relationship poorly defined, thus rationally introduced mutations could not be made. Additionally, no method previously developed was able to screen the large number of  $\alpha_1$ -PI mutants required to test every possible codon combination in the hypervariable region. It was hypothesised that phage display, which had been used to screen mutants of the serpin PAI-1, could be extended to better engineer  $\alpha_1$ -PI mutants. Using T7Select10-3b bacteriophage and an *in vitro* evolution approach, in theory only the  $\alpha_1$ -PI mutants with improved thrombin inhibitory activity would eventually be enriched. These novel  $\alpha_1$ -PI RCL mutants will have been selected to both favourably bind thrombin as well as insert rapidly into the body of the serpin. Any conservation between the RCL sequences of these mutants would also offer insight into the function of the hypervariable region, and serpin inhibitory mechanism in general. The first step was to produce recombinant PAI-1 protein as well as expressing PAI-1 using T7Select10-3b phage in order to demonstrate that this unmodified serpin can be produced in our lab, and that modern phage display techniques can replicate previous serpin engineering work.

## **5.1 Recombinant PAI-1 must be Refolded and can be Expressed using T7Select10-3b**

Mirroring the results published by Lee & Im (2003), from which the protocol for producing PAI-1 was adapted, PAI-1 was required to be released from *E. coli* inclusion bodies and refolded for optimal function (139). The majority of PAI-1 was produced as soluble protein though primarily in an inactive latent conformation. Refolding of this soluble fraction could be attempted if more protein was required, but for the purposes of this study functional PAI-1 was successfully obtained from inclusion bodies.

Using the commercially available T7Select10-3b bacteriophage, the full length PAI-1 sequence was successfully fused to the 10B minor coat protein and expressed, although with only 2% of plaques detectably expressing the PAI-1 insert. The reason for this poor expression level was not clear. The manufacturer stated that 5-15 copies of a protein up to 1200 amino acids in size can be expressed per phage. As the T7Select system had not previously been used to express serpins no comparison was available, although Dai *et al.* (2008) noted that the expression of green fluorescent protein (GFP) variants by T7Select were limited to only 1-3 copy per phage (122). GFP is 238 amino acids in size or approximately 40% smaller than PAI-1. This size difference may account for the further reduction in expression levels, despite the manufacturer's assertions, or the copy number was below the detectable limit of the antibody that was used. The function of phage-expressed PAI-1 was not assessed, as the biopanning procedures employed at this stage of the study (Protease-Coated Well and Antibody-Coated Well methods) were not able to enrich PAI-1 expressing phage. As the focus of the project was on the serpin  $\alpha_1$ -PI, no further experiments were performed using T7Select10-3b expressing PAI-1.

## **5.2 $\alpha_1$ -PI M358R can be Expressed using T7Select10-3b and Forms Stable Serpin-Enzyme Complexes with Thrombin**

Applying the same techniques used to generate T7Select10-3b PAI-1 phage, the serpin  $\alpha_1$ -PI M358R was successfully expressed on the surface of T7Select10-3b. This represented the first time that phage display was extended to  $\alpha_1$ -PI, making it only the second serpin to be produced using this method. T7Select10-3b expression of  $\alpha_1$ -PI M358R was found to be greater than PAI-1, with an average of 13% of plaques detectably expressing the  $\alpha_1$ -PI M358R insert. The difference in expression remains undefined, but may have been influenced by either the ligation or packaging efficiency when constructing the phage. The two proteins are of similar size, indeed  $\alpha_1$ -PI M358R is a slightly larger protein than PAI-1 (394 amino acids versus 379), thus the size of the inserted sequence likely did not play a role.

The definitive test of a serpin's function is its ability to form a SDS-stable serpin-enzyme complex with its target protease. In this study, such complexes were directly observed between  $\alpha_1$ -PI M358R fused to the coat of T7Select10-3b and thrombin, as detected by an immunoblot. Previous studies suggested that phage-expressed PAI-1 could form a stable complex with tPA, but they were not observed directly (124, 129). The observed phage-expressed  $\alpha_1$ -PI M358R-thrombin complexes provided the first evidence that not only can the T7Select system express serpins, but that they also retain their function when fused to coat protein 10B. N-terminal fusion to a large complex such as a phage virion does not disrupt  $\alpha_1$ -PI M358R's activity, mirroring previous work tethering this serpin to mammalian cells (137).

### **5.3 Mock-Biopanning Successfully Enriches T7Select10-3b $\alpha_1$ -PI M358R phage**

Several biopanning protocols based on the T7Select manual and previously published work were initially employed, but none was able to successfully enrich serpin-expressing phage (121, 124, 141). A novel biopanning protocol was therefore developed, based on the method used by Dai *et al.* (2008), centering on the use of streptavidin-coated magnetic beads and a biotinylated anti-prothrombin antibody (122). This antibody was shown to bind phage-expressed  $\alpha_1$ -PI M358R-thrombin complexes, thus these complexes and the phage associated with them could be pulled down via the biotin-streptavidin interaction. After washing the beads several times, *E. coli* was directly infected with the bead-phage solution to ensure the tightest interactions were not lost during an elution step. As the serpin was expressed on the non-infective half of the phage, interaction with the protease, antibody, and beads were not expected to influence infection.

This method proved effective during mock-biopanning of T7Select10-3b  $\alpha_1$ -PI M358R purposely mixed 1/300 with T7Select10-3b expressing a 15 amino acid “Control Insert”. Only modest enrichment of the  $\alpha_1$ -PI M358R phage was observed for the first four rounds, so the protocol was modified for the fifth round of mock biopanning. Due to the resistant nature of the serpin-protease and antibody-bead interactions (covalent and strong non-covalent interactions respectively), a stronger washing solution was employed in an effort to decrease all weaker nonspecific interactions. The success of this modified biopanning protocol laid the foundation for the rest of the study, as additional mutations within  $\alpha_1$ -PI could now be screened for thrombin inhibitory activity.

#### 5.4 The Optimal P2 and P1 Residues for Thrombin Inhibition

To test the effectiveness of biopanning a library of  $\alpha_1$ -PI mutants using this novel phage display approach, the P2 and P1 RCL positions were investigated first. The  $\alpha_1$ -PI M358R mutant had previously been identified as a rapid inhibitor of thrombin, so the appearance of this mutant would be a useful gauge of successful biopanning. Although arginine is highly conserved at P1 in thrombin inhibitory serpins proline is not conserved at P2, and previous mutagenesis experiments had not investigated every P2P1 combination (**Appendix A**). Therefore, the phage library was constructed so that every codon combination was represented at the P2 and P1 positions, to generate a total of 441 possible mutants. The library was biopanned with thrombin, and separately without the addition of thrombin so that overlapping sequences could be ruled out. Results from biopanning the P2P1 library indicated that the only mutant capable of forming a stable inhibitory complex with thrombin was  $\alpha_1$ -PI M358R. The varied codon usage in the sequenced  $\alpha_1$ -PI M358R expressing plaques suggested that this mutant was functionally significant and not preferred due to bias. This result was generally expected as the P2P1 sequence of  $\alpha_1$ -PI M358R already fit thrombin's known substrate specificity, and other amino acids do not interact as favourably in thrombin's S2 and S1 subsites (21, 22). The observation that no other thrombin inhibitory serpins have this P2P1 sequence is likely due to their own evolutionary path. AT, HCII have unique RCL sequences and exosites which specifically evolved to interact with thrombin, while the thrombin inhibitory activity  $\alpha_1$ -PI M358R was the result of a rare mutation, thus the P2P1 residues more are tightly constrained (16, 75).

The P357/M358P mutant persisted through biopanning, representing 20% of sequences when biopanned with thrombin but over 50% of sequences in the negative control biopanning in which thrombin was not added. These frequencies were higher than expected, despite results that this mutant is incapable of forming a stable complex with thrombin. The nucleotide cytosine appeared with a greater frequency than expected in the degenerate region, and this bias lead to an over representation of the proline codon (CCN). Improved library construction methods with more evenly distributed nucleotides may eliminate this issue. Ultimately the codon bias did not influence the finding that the P2 and P1 residues must be proline and arginine respectively in order for  $\alpha_1$ -PI to inhibit thrombin, providing proof of concept that thrombin-specific  $\alpha_1$ -PI mutants could be enriched using this novel phage display method.

### **5.5 The P7 to P3 Region Influences Thrombin Inhibition**

A second T7Select10-3b library was generated using the same methods as the first, maintaining the optimal P2P1 residues while randomising RCL positions P7 through to P3. This region of the RCL was selected based on previous mutagenesis studies which indicated that its sequence affects the function and protease specificity of a serpin (58, 67, 103, 104, 109, 148, 149). With five degenerate codons, the  $\alpha_1$ -PI M358R P7-P3 library contained over 4.08 million ( $21^5$ ) possible mutants. Based on the novel biopanning protocol which was developed for this study, 5 degenerate positions was the maximum number that could be investigated in a single library. This was to ensure at least 100 copies of each mutant were present during biopanning, and  $1 \times 10^9$  pfu screened per round.

Following biopanning with (Rd5 +IIa) or without thrombin (Rd5 –IIa), a collapse of sequence diversity was observed in the Rd5 –IIa samples retrieved from randomly selected plaques due to the lack of a selection pressure. Unlike the results of the P2P1 library, no single mutant was dominant within the Rd5 +IIa sequences though they were found to fit a general trend: P7-Not Aromatic/P6-Hydrophobic/P5-T or S/P4-Hydrophobic/P3-Not Aromatic. No nucleotide bias for cytosine, leading to an abundance of proline residues, was observed in the P7-P3 Rd5 +IIa results. This was likely a result of the P7-P3 region's greater tolerance for diversity; a greater number of functional mutants meant that biased sequences were eventually outgrown and removed. In order to further refine results to uncover the best possible P7-P3 sequence for thrombin inhibition, biopanned sequences were inserted into a bacterial expression vector and semi-quantitatively tested. The results of this screen, acting as a 6<sup>th</sup> round of biopanning, mirrored the randomly sequenced plaques; the general trend at positions P7 through to P3 matched, and the consensus sequences were remarkably similar (DLTVS and LATVS respectively). This suggested that thrombin inhibitory function was enriched via biopanning as intended, supported by the observation that colonies containing Rd5 –IIa sequences did not retain thrombin binding activity while the majority of Rd5 +IIa sequences had thrombin inhibitory activity equal to or greater than  $\alpha_1$ -PI M358R. Notably, neither consensus sequence was found as an individual sequence. This may have been a result of the limited number of plaques and colonies screened, or could suggest cooperative interactions between residues.

The bioinformatics program ConSurf was used to help quantify the conservation of the observed trend, using only sequences found to have a greater level of thrombin binding than the original  $\alpha_1$ -PI M358R. Based on the evolutionary concept that conservation is linked to function, the conserved P6, P5, and P4 positions played a significant role in determining the thrombin inhibitory function of  $\alpha_1$ -PI. Positions P7 and P3 were not well conserved and therefore did not substantially influence function. The relatively high diversity of sequences found after screening the P7-P3 library suggests that this region is an important but not a crucial determinant for thrombin inhibition. Unlike the P2P1 residues, the P7-P3 region acts to fine-tune the  $\alpha_1$ -PI RCL and can accommodate a wider, albeit restricted, range of amino acids.

### **5.6 Phage Displayed Serpins account for Protease Specificity and RCL Insertion**

The trend observed following biopanning of the P7-P3 library is consistent with the known substrate specificity of thrombin. A hydrophobic P4 residue is required to favourably interact with thrombin's S4 subsite, while thrombin can tolerate a variety of residues at the P3 position (17, 21, 22, 26). Anecdotally, serine appeared frequently at P3 while aspartic acid did not, matching some indications of thrombin's substrate specificity (21). The substrate specificity of thrombin N-terminal to P4 is not known, however general serpin structure provides an indication of preferred residues at these RCL positions. As discussed by Gettins (2002), there is a trend within serpins towards a hydrophobic residue at position P6 which also matches general  $\beta$ -sheet structure (26, 69). This trend was clearly evident, as P6 was conserved as a hydrophobic residue both in P7-P3 sequences with improved thrombin binding ability and the majority of sequences

identified in randomly selected plaques. This finding suggests that a hydrophobic P6 residue is important for serpin function in general and not simply in order to inhibit thrombin.

Surprisingly the P5 residue indicated a strong preference for either serine or threonine, with 14 of the 22 improved mutants fitting this criterion. The P5 position is not conserved among natural thrombin inhibitory serpins, so this finding likely reflects a thrombin inhibitory mechanism unique to  $\alpha_1$ -PI. In wild type  $\alpha_1$ -PI, the P5 glutamic acid residue faces away from the protease to form salt bridges with a basic pocket in the body of the serpin, adding a level of rigidity to the RCL structure (144). Chaillan-Huntington & Patston (1998) concluded that these P5 interactions position the RCL into the optimal position for neutrophil elastase and trypsin inhibition (144). The molecular modeling performed in this study also identified these interactions, although they were not observed in the models of the P7-P3 mutants with improved rates of thrombin inhibition.

Therefore, in order for  $\alpha_1$ -PI to inhibit thrombin it may be more favourable if the RCL is flexible in order to manoeuvre into the active site, entrance of which is sterically hindered by thrombin's unique insertion loops (14). It is also possible that an uncharged P5 residue is more favourable during RCL insertion into the central  $\beta$ -sheet (26). Although mutating the P5 residue negatively impacts the SI for neutrophil elastase and trypsin inhibition, thrombin is not a natural target of  $\alpha_1$ -PI and is also known to be sensitive to RCL insertion due to its bulky loops proximal to the active site (48, 144). Cooperative interactions between RCL residues or undefined compounding factors may also affect RCL insertion, which may be strengthened by replacing the P5 glutamic acid residue.

The P7 residue, which is furthest from the active site of thrombin, did not appear to influence activity significantly as the physiochemical properties of amino acids varied considerably. This is exemplified when the mutants EATVS and HATVS are compared. These mutants differ in only the P7 position, and accordingly their rates of thrombin inhibition were not found to be significantly different from each other (**Table 3**). The single constraint appears to be that P7 cannot be an aromatic residue, which are physically larger and would likely impede loop insertion. Although not found to be significant, non-aromatic hydrophobic residues and acidic residues appeared frequently at the P7 position. In the molecular models of DITMA and EISLQ encounter complexes with thrombin, several P7 interactions were observed. The acidic P7 residues of these two mutants formed salt bridges with the same basic residues in  $\alpha_1$ -PI that interact with the wild type P5 glutamic acid residue. These interactions may position the RCL more favourably in thrombin's active site, such as improving the fit of the P4 residue or pulling the RCL away from steric clashes with the 60-loop. Evaluation of additional P7-P3 mutants may reveal more information about this trend.

The trends observed in this study are unique to  $\alpha_1$ -PI, particularly at the highly conserved P5 residue. This supports the hypothesis that borrowing RCL sequences from natural thrombin inhibitory serpins would not create an optimal thrombin inhibitor, due to the unique interactions within the body of  $\alpha_1$ -PI or elsewhere between the serpin and thrombin. As such, rationally introducing these mutations would have been extremely difficult, especially without the use of phage display which exponentially increased the ability to screen engineered mutants.

## 5.7 Characterisation of Selected Mutants of Interest

Five P7-P3 mutants were selected for further characterization with each matching the apparent trend required for optimal thrombin inhibition. The mutant AAFVS was an exception to the trend as it did not contain a P5 threonine or serine residue, and was selected solely based on its thrombin binding activity in the lysate assay. Mutants EATVS and HATVS were also selected based on their similarities to the observed consensus sequences, DLTVS and LATVS, from randomly sequenced plaques and mutants indicating improved thrombin binding respectively.

Mutants DITMA and AAFVS were found to have the most rapid rate of thrombin inhibition, granting an over 2-fold increase in their rates of thrombin inhibition as compared to  $\alpha_1$ -PI M358R, with second order rate constants of  $1.0 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ . Their rapid rate of thrombin inhibition is comparable to the most active  $\alpha_1$ -PI mutant previously engineered, which contains the AT-based “RCL5” mutation (103, 104). Both DITMA and AAFVS were also more rapid thrombin inhibitors than PN1 ( $8.3 \times 10^5$ ), the fastest native thrombin inhibitor (64). However, serpins that have evolved specifically to inhibit thrombin remain an order of magnitude faster than these engineered mutants (10). An important caveat is that these natural serpins require glycosaminoglycan activation to achieve these higher rates, which is not required for the engineered mutants of  $\alpha_1$ -PI (26). This difference can also be attributed to additional exosites and regions beyond the RCL which improve binding, which could be added to  $\alpha_1$ -PI to attribute both improved rates of thrombin inhibition and greater specificity (53, 104).

The failure of the EATVS and HATVS mutants to significantly improve thrombin inhibitory activity is likely a result of the biopanning methods employed. The phage display technique effectively amplified functional sequences but not necessarily the optimal sequences for thrombin inhibition. Thus, many mutants end up being “good enough” rather than optimal thrombin inhibitors, leading to a consensus sequence of middling activity. Normalising the results of the ELISA-based lysate assay to expression levels of serpin may alleviate this bias to reveal only the top inhibitors of thrombin. This would be particularly useful in helping to eliminate the discrepancy in activity between mutants EISLQ and DITMA, in which the former indicated a higher activity than the latter in the lysate assay, which did not follow through when the mutants were purified.

The SIs of both the DITMA and AAFVS mutants were significantly lowered compared to  $\alpha_1$ -PI M358R, indicating a greater propensity to form a stable complex with thrombin as opposed to being cleaved. Despite the fact that the P7-P3 sequences of DITMA and AAFVS vary considerably, they have similar activities. The structural differences between DITMA and AAFVS make it difficult to draw conclusions as to why both the  $k_2$  and SI was improved. The smaller P7 and P6 residues of AAFVS versus DITMA may have been enough to compensate for the energy required to insert the bulkier P5 residue. Indeed, such cooperative interactions between the residues, noted in previous studies, likely play a role (21, 103). Based on the mutants selected for kinetic characterisation, a functional trend for P5 is difficult to determine. Both the DITMA and AAFVS mutants were found to have a significantly lower SI despite widely dissimilar P5 residues, while EATVS did not have a lower SI.

One feature maintained in the characterised P7-P3 mutants was the larger, more hydrophobic P4 residue, compared to wild type  $\alpha_1$ -PI's alanine. Based on the molecular models, the mutant P4 residues occupied more of the hydrophobic S4 pocket. Paradoxically, the ZMM program interpreted  $\alpha_1$ -PI M358R as having the most favourable free energy at the P4 position. Instead of being due to stabilising interactions, this favourable energy was due to the lack of steric clashes, which masked the probable hydrophobic interactions between the larger mutant P4 residues and thrombin. HCII's P4 phenylalanine residue interacted the most favourably, despite being physically larger than any of the  $\alpha_1$ -PI P4 residues investigated. This was likely because the HCII-thrombin encounter complex was a proper crystal structure, while the  $\alpha_1$ -PI-thrombin encounter complexes were manually generated. Modifications of the  $\alpha_1$ -PI and thrombin structures were introduced in order to perform energy minimisation of the encounter complexes, and therefore do not represent optimal interactions. A crystal structure of the  $\alpha_1$ -PI M358R-thrombin encounter complex would help to alleviate these discrepancies, although the developed models were useful in identifying some interactions.

Overall, the rapid rates of thrombin inhibition combined with lower SI values support the original hypothesis: that this phage display technique is able to enrich RCL sequences matching thrombin's substrate preferences and also maintain the constraints of the serpin mechanism. In effect, RCL sequences were evolved *in vitro* to interact favourably with both the target protease, and the unique body of the serpin in order for the RCL to properly insert.

## 6. CONCLUSIONS AND FUTURE DIRECTIONS

This study has successfully shown that two serpins can be expressed using the T7Select phage display system, that  $\alpha_1$ -PI M358R retained its thrombin inhibitory activity when fused to the surface of the phage, and that improved activity can be engineered using an *in vitro* evolution approach. By expressing the full length protein, modifications to the RCL were screened for both specificity and the stability of complex formation. This represents significant advantages over previous serpin engineering techniques, which relied on natural sequences and the specificity profiles of target proteases in order to introduce function. Due to the high-throughput nature of the phage display technique, it successfully enriched desired activity out of over 4 million possible mutants which is over 30,000 times the total number of  $\alpha_1$ -PI mutants screened in all previously published studies (**Appendix A**). Importantly, the mutations were focused within a specific region of the RCL which tightly controlled the sequence space investigated. This generated specific evidence of relationships between RCL structure and function, unlike initial phage display studies with the serpin PAI-1 which introduced mutations randomly through the entire protein.

The sequences enriched for thrombin inhibition helped to both confirm the crucial importance of the P2 and P1 residues for thrombin inhibitory function, and to elucidate previously unknown specificity and functional requirements in the P7-P3 RCL region of  $\alpha_1$ -PI. These findings aid general understanding of serpin biology and may lead to the development of novel thrombin inhibitory serpins with therapeutic relevance. Further

towards that goal, several steps could be taken to extend the use of the developed phage display method, and refine the results.

There are two primary functional limits to the phage display screen which could be addressed in future studies. Due to the size of the P7-P3 library, only a small fraction of the enriched mutants were sequenced and functionally characterised. Next generation sequencing techniques could provide additional information, helping to elucidate remaining P7-P3 mutants with a possible improvement in thrombin inhibition. These “deep sequencing” methods can screen hundreds of thousands of sequences, and may reveal additional trends not observed in this study due to a smaller sampling size (113, 115). Work is currently being performed to analyse the sequences from this study using such methods. Secondly, although the ELISA-based lysate screen was useful as a semi-quantitative screen, it is labour intensive and can not completely screen every sequence which was enriched after five rounds of biopanning. Thus, steps could be taken to limit the sequence diversity by introducing additional selection pressures. The P7-P3 mutants were only screened against thrombin and not against any other coagulation protease.  $\alpha_1$ -PI M358R is known to inhibit several other proteases, particularly APC, which has held it back from clinical uses (98). To further refine the thrombin inhibitory activity of the P7-P3 mutants, a negative selection using the same phage library could be performed. This would involve biopanning against a protease known to also inhibit  $\alpha_1$ -PI M358R, such as APC, and only amplifying the mutants which do not bind to the protease. Alternating between positive selection rounds with thrombin and negative selection rounds would

ensure sequence diversity does not collapse, while removing sequences with broad specificity.

Using the methods outlined in this study, up to 10 million variant serpins could be screened for a desired activity. Additional libraries could be developed to investigate the function of other regions within the RCL. Positions C-terminal to the P1-P1' cleavage site, the P' region, have also been shown to influence specificity (49, 58, 103).

Modification of the P2' and P3' residues can improve the specificity towards thrombin over APC, thus a library investigating these positions and up to three more positions, could be screened (103). Borrowing from previous engineering successes, the unique HCII tail region could be added to the P7-P3 mutants with improved activity to improve their thrombin selectivity over APC. Two mutants with promising activity, HAPI RCL5 and HAPI M358R provide clues as to how such mutants would behave. HAPI RCL5 has the best thrombin to APC ratio yet, while HAPI M358R has the best rate of thrombin inhibition (104, 105). Using one of the mutant P7-P3 RCL sequences which was evolved *in vitro* to inhibit thrombin, it may be possible to engineer a "HAPI DITMA" or "HAPI AAFVS" variant which is both an exceptionally selective and rapid inhibitor of thrombin.

Now that these phage display methods have been developed, the doors have been opened to using  $\alpha_1$ -PI as a scaffold for additional engineering. Other proteases relevant to disease, which do not have readily available inhibitors, could be targeted using these methods. Indeed, this novel phage display system could be extended to engineering other serpins as well.

## 7. REFERENCES

1. Krem, M. M., & Di Cera, E. . (2002) Evolution of enzyme cascades from embryonic development to blood coagulation, *Trends in Biochemical Sciences* 27, 68-74.
2. Gross, P. L., Murray, R. K., and Rand, M. L. (2011) Hemostasis & Thrombosis, In *Harper's Illustrated Biochemistry* (Bender, D. A., Botham, K. M., Weil, P. A., Kennelley, P. J., Murray, R. K., and Rodwell, V. W., Eds.) 29th ed., McGraw-Hill, New York.
3. Morrissey, J. H., Macik, B. G., Neuenschwander, P. F., and Comp, P. C. (1993) Quantitation of activated factor VII levels in plasma using a tissue factor mutant selectively deficient in promoting factor VII activation, *Blood* 81, 734-744.
4. Ansell, J. (2007) Factor Xa or thrombin: is factor Xa a better target?, *Journal of thrombosis and haemostasis : JTH* 5 Suppl 1, 60-64.
5. Gailani, D., and Renne, T. (2007) Intrinsic pathway of coagulation and arterial thrombosis, *Arteriosclerosis, thrombosis, and vascular biology* 27, 2507-2513.
6. Hamad, O. A., Back, J., Nilsson, P. H., Nilsson, B., and Ekdahl, K. N. (2012) Platelets, complement, and contact activation: partners in inflammation and thrombosis, *Advances in experimental medicine and biology* 946, 185-205.
7. Borissoff, J. I., Spronk, H. M., Heeneman, S., and ten Cate, H. (2009) Is thrombin a key player in the 'coagulation-atherogenesis' maze?, *Cardiovascular research* 82, 392-403.
8. Monroe, D. M., Hoffman, M., and Roberts, H. R. (2002) Platelets and thrombin generation, *Arteriosclerosis, thrombosis, and vascular biology* 22, 1381-1389.
9. Bode, W. (2006) Structure and interaction modes of thrombin, *Blood cells, molecules & diseases* 36, 122-130.
10. Olson, S. T., and Gettins, P. G. (2011) Regulation of proteases by protein inhibitors of the serpin superfamily, *Progress in molecular biology and translational science* 99, 185-240.
11. Carter, I. S., Vanden Hoek, A. L., Pryzdial, E. L., and Macgillivray, R. T. (2010) Thrombin a-chain: activation remnant or allosteric effector?, *Thrombosis* 2010, 416167.
12. Huntington, J. A., and Baglin, T. P. (2003) Targeting thrombin--rational drug design from natural mechanisms, *Trends in pharmacological sciences* 24, 589-595.
13. Stubbs, M. T., Oschkinat, H., Mayr, I., Huber, R., Angliker, H., Stone, S. R., and Bode, W. (1992) The interaction of thrombin with fibrinogen. A structural basis for its specificity, *European journal of biochemistry / FEBS* 206, 187-195.
14. Huntington, J. A. (2012) Thrombin plasticity, *Biochimica et biophysica acta* 1824, 246-252.
15. Di Cera, E., and Cantwell, A. M. (2001) Determinants of thrombin specificity, *Annals of the New York Academy of Sciences* 936, 133-146.
16. Huntington, J. A. (2011) Serpin structure, function and dysfunction, *Journal of thrombosis and haemostasis : JTH* 9 Suppl 1, 26-34.
17. Backes, B. J., Harris, J. L., Leonetti, F., Craik, C. S., and Ellman, J. A. (2000) Synthesis of positional-scanning libraries of fluorogenic peptide substrates to define the extended substrate specificity of plasmin and thrombin, *Nat Biotechnol* 18, 187-193.
18. Schrodinger LLC. (2010) The PyMOL Molecular Graphics System, Version 1.3r1.
19. Rawlings, N. D., Barrett, A. J., and Bateman, A. (2012) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors, *Nucleic acids research* 40, D343-350.
20. Huntington, J. A., and Esmon, C. T. (2003) The molecular basis of thrombin allostery revealed by a 1.8 Å structure of the "slow" form, *Structure* 11, 469-479.
21. Gallwitz, M., Enoksson, M., Thorpe, M., and Hellman, L. (2012) The extended cleavage specificity of human thrombin, *PLoS one* 7, e31756.

22. Bianchini, E. P., Louvain, V. B., Marque, P. E., Juliano, M. A., Juliano, L., and Le Bonniec, B. F. (2002) Mapping of the catalytic groove preferences of factor Xa reveals an inadequate selectivity for its macromolecule substrates, *J Biol Chem* 277, 20527-20534.
23. Bhunia, S. S., Roy, K. K., and Saxena, A. K. (2011) Profiling the structural determinants for the selectivity of representative factor-Xa and thrombin inhibitors using combined ligand-based and structure-based approaches, *Journal of chemical information and modeling* 51, 1966-1985.
24. Le Bonniec, B. F., MacGillivray, R. T., and Esmon, C. T. (1991) Thrombin Glu-39 restricts the P'3 specificity to nonacidic residues, *J Biol Chem* 266, 13796-13803.
25. Marque, P. E., Spuntarelli, R., Juliano, L., Aiach, M., and Le Bonniec, B. F. (2000) The role of Glu(192) in the allosteric control of the S(2)' and S(3)' subsites of thrombin, *J Biol Chem* 275, 809-816.
26. Gettins, P. G. (2002) Serpin structure, mechanism, and function, *Chemical reviews* 102, 4751-4804.
27. Petitou, M., Herault, J. P., Bernat, A., Driguez, P. A., Duchaussoy, P., Lormeau, J. C., and Herbert, J. M. (1999) Synthesis of thrombin-inhibiting heparin mimetics without side effects, *Nature* 398, 417-422.
28. Corral-Rodriguez, M. A., Macedo-Ribeiro, S., Pereira, P. J., and Fuentes-Prior, P. (2010) Leech-derived thrombin inhibitors: from structures to mechanisms to clinical applications, *Journal of medicinal chemistry* 53, 3847-3861.
29. Malin, R. (2010). PRADAX™ (dabigatran etexilate) gains approval in Canada for stroke prevention in atrial fibrillation. Retrieved July 9, 2013: [http://www.boehringer-ingenheim.com/news/news\\_releases/press\\_releases/2010/27\\_october\\_2010\\_dabigatran.html](http://www.boehringer-ingenheim.com/news/news_releases/press_releases/2010/27_october_2010_dabigatran.html).
30. Hankey, G. J., and Eikelboom, J. W. (2011) Dabigatran etexilate: a new oral thrombin inhibitor, *Circulation* 123, 1436-1450.
31. Scott, B. M. (2013) High-throughput Screening and the Rational Design of Anticoagulants, *MUMJ* 10, 29-32.
32. Cohen, M. R. (2012). QuarterWatch Monitoring FDA MedWatch Reports Anticoagulants the Leading Reported Drug Risk in 2011. Retrieved July 9, 2013: <http://www.ismp.org/quarterwatch/pdfs/2011Q4.pdf>.
33. U.S. Food and Drug Administration. (2012). Pradaxa (dabigatran etexilate mesylate): Drug Safety Communication - Safety Review of Post-Market Reports of Serious Bleeding Events. Retrieved July 9, 2013: <http://www.fda.gov/safety/medwatch/safetyinformation/safetyalertsforhumanmedicalproducts/ucm282820.htm>.
34. Schiele, F., van Ryn, J., Canada, K., Newsome, C., Sepulveda, E., Park, J., Nar, H., and Litzenburger, T. (2013) A specific antidote for dabigatran: functional and structural characterization, *Blood* 121, 3554-3562.
35. Liew, A., Eikelboom, J. W., O'Donnell, M., and Hart, R. G. (2013) Assessment of anticoagulation intensity and management of bleeding with old and new oral anticoagulants, *The Canadian journal of cardiology* 29, S34-44.
36. Silverman, G. A. (2001) The Serpins Are an Expanding Superfamily of Structurally Similar but Functionally Diverse Proteins, *The Journal of Biological Chemistry* 276, 33293-33296.
37. LeMosy, E. K., Hong, C. C., and Hashimoto, E. (1999) Signal transduction by a protease cascade, *Cell Biology* 9.

38. Law, R. H., Zhang, Q., McGowan, S., Buckle, A. M., Silverman, G. A., Wong, W., Rosado, C. J., Langendorf, C. G., Pike, R. N., Bird, P. I., and Whisstock, J. C. (2006) An overview of the serpin superfamily, *Genome biology* 7, 216.
39. Kaiserman, D., Whisstock, J. C., and Bird, P. I. (2006) Mechanisms of serpin dysfunction in disease, *Expert reviews in molecular medicine* 8, 1-19.
40. Song, J., Matthews, A. Y., Reboul, C. F., Kaiserman, D., Pike, R. N., Bird, P. I., and Whisstock, J. C. (2011) Predicting serpin/protease interactions, *Methods in enzymology* 501, 237-273.
41. Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain, *Biochemical and biophysical research communications* 27, 157-162.
42. UniProtKB/Swiss-Prot. (2013). Alpha-1-antitrypsin precursor - Homo sapiens (Human). Retrieved May 11, 2013: <http://www.uniprot.org/uniprot/P01009>.
43. Dobo, J., and Gettins, P. G. (2004) alpha1-Proteinase inhibitor forms initial non-covalent and final covalent complexes with elastase analogously to other serpin-proteinase pairs, suggesting a common mechanism of inhibition, *J Biol Chem* 279, 9264-9269.
44. Stratikos, E., and Gettins, P. G. (1999) Formation of the covalent serpin-proteinase complex involves translocation of the proteinase by more than 70 Å and full insertion of the reactive center loop into beta-sheet A, *Proceedings of the National Academy of Sciences of the United States of America* 96, 4808-4813.
45. Stratikos, E., and Gettins, P. G. (1998) Mapping the serpin-proteinase complex using single cysteine variants of alpha1-proteinase inhibitor Pittsburgh, *J Biol Chem* 273, 15582-15589.
46. Lawrence, D. A., Olson, S. T., Muhammad, S., Day, D. E., Kvassman, J. O., Ginsburg, D., and Shore, J. D. (2000) 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* 275, 5839-5844.
47. Calugaru, S. V., Swanson, R., and Olson, S. T. (2001) The pH dependence of serpin-proteinase complex dissociation reveals a mechanism of complex stabilization involving inactive and active conformational states of the proteinase which are perturbable by calcium, *J Biol Chem* 276, 32446-32455.
48. Zhou, A., Carrell, R. W., and Huntington, J. A. (2001) The serpin inhibitory mechanism is critically dependent on the length of the reactive center loop, *J Biol Chem* 276, 27541-27547.
49. Bottomley, S. P., and Stone, S. R. (1998) Protein engineering of chimeric Serpins: an investigation into effects of the serpin scaffold and reactive centre loop length, *Protein engineering* 11, 1243-1247.
50. Avron, A., Reeve, F. H., Lickorish, J. M., and Carrell, R. W. (1991) Effect of alanine insertion (P'5) on the reactive centre of alpha 1-antitrypsin, *FEBS letters* 280, 41-43.
51. Hopkins, P. C., Carrell, R. W., and Stone, S. R. (1993) Effects of mutations in the hinge region of serpins, *Biochemistry* 32, 7650-7657.
52. Bjork, I., and Olson, S. T. (1997) Antithrombin. A bloody important serpin, *Advances in experimental medicine and biology* 425, 17-33.
53. Gettins, P. G., and Olson, S. T. (2009) Exosite determinants of serpin specificity, *J Biol Chem* 284, 20441-20445.
54. Dementiev, A., Simonovic, M., Volz, K., and Gettins, P. G. (2003) Canonical inhibitor-like interactions explain reactivity of alpha1-proteinase inhibitor Pittsburgh and antithrombin with proteinases, *J Biol Chem* 278, 37881-37887.

55. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0, *Bioinformatics* 23, 2947-2948.
56. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench, *Bioinformatics* 25, 1189-1191.
57. Taylor, W. R. (1997) Residual colours: a proposal for aminochromography, *Protein engineering* 10, 743-746.
58. Hopkins, P. C., Crowther, D. C., Carrell, R. W., and Stone, S. R. (1995) Development of a novel recombinant serpin with potential antithrombotic properties, *J Biol Chem* 270, 11866-11871.
59. Rau, J. C., Beaulieu, L. M., Huntington, J. A., and Church, F. C. (2007) Serpins in thrombosis, hemostasis and fibrinolysis, *Journal of thrombosis and haemostasis : JTH* 5 Suppl 1, 102-115.
60. Paterson, M. A., Horvath, A. J., Pike, R. N., and Coughlin, P. B. (2007) Molecular characterization of centerin, a germinal centre cell serpin, *The Biochemical journal* 405, 489-494.
61. Riewald, M., and Schleef, R. R. (1996) Human cytoplasmic antiproteinase neutralizes rapidly and efficiently chymotrypsin and trypsin-like proteases utilizing distinct reactive site residues, *J Biol Chem* 271, 14526-14532.
62. Dahlen, J. R., Foster, D. C., and Kisiel, W. (1997) Expression, Purification, and Inhibitory Properties of Human Proteinase Inhibitor 8, *Biochemistry* 36, 14874-14882.
63. Riewald, M., and Schleef, R. R. (1995) Molecular cloning of bomapin (protease inhibitor 10), a novel human serpin that is expressed specifically in the bone marrow, *J Biol Chem* 270, 26754-26757.
64. Evans, D. L., McGrogan, M., Scott, R. W., and Carrell, R. W. (1991) Protease specificity and heparin binding and activation of recombinant protease nexin I, *J Biol Chem* 266, 22307-22312.
65. Caccia, S., Castelli, R., Maiocchi, D., Bergamaschini, L., and Cugno, M. (2011) Interaction of C1 inhibitor with thrombin on the endothelial surface, *Blood coagulation & fibrinolysis : an international journal in haemostasis and thrombosis* 22, 571-575.
66. Hood, D. B., Huntington, J. A., and Gettins, P. G. (1994) Alpha 1-proteinase inhibitor variant T345R. Influence of P14 residue on substrate and inhibitory pathways, *Biochemistry* 33, 8538-8547.
67. Dufour, E. K., Denault, J. B., Bissonnette, L., Hopkins, P. C., Lavigne, P., and Leduc, R. (2001) The contribution of arginine residues within the P6-P1 region of alpha 1-antitrypsin to its reaction with furin, *J Biol Chem* 276, 38971-38979.
68. Dufour, E. K., Desilets, A., Longpre, J. M., and Leduc, R. (2005) Stability of mutant serpin/furin complexes: dependence on pH and regulation at the deacylation step, *Protein science : a publication of the Protein Society* 14, 303-315.
69. Mandel-Gutfreund, Y., and Gregoret, L. M. (2002) On the significance of alternating patterns of polar and non-polar residues in beta-strands, *Journal of molecular biology* 323, 453-461.
70. Baumann, U., Huber, R., Bode, W., Grosse, D., Lesjak, M., and Laurell, C. B. (1991) Crystal structure of cleaved human alpha 1-antichymotrypsin at 2.7 Å resolution and its comparison with other serpins, *Journal of molecular biology* 218, 595-606.

71. Mourey, L., Samama, J. P., Delarue, M., Petitou, M., Choay, J., and Moras, D. (1993) Crystal structure of cleaved bovine antithrombin III at 3.2 Å resolution, *Journal of molecular biology* 232, 223-241.
72. Huntington, J. A., Read, R. J., and Carrell, R. W. (2000) Structure of a serpin-protease complex shows inhibition by deformation, *Nature* 407, 923-926.
73. Yamasaki, M., Arii, Y., Mikami, B., and Hirose, M. (2002) Loop-inserted and thermostabilized structure of P1-P1' cleaved ovalbumin mutant R339T, *Journal of molecular biology* 315, 113-120.
74. Huntington, J. A., Kjellberg, M., and Stenflo, J. (2003) Crystal structure of protein C inhibitor provides insights into hormone binding and heparin activation, *Structure* 11, 205-215.
75. Baglin, T. P., Carrell, R. W., Church, F. C., Esmon, C. T., and Huntington, J. A. (2002) Crystal structures of native and thrombin-complexed heparin cofactor II reveal a multistep allosteric mechanism, *Proceedings of the National Academy of Sciences of the United States of America* 99, 11079-11084.
76. Bouton, M. C., Boulaftali, Y., Richard, B., Arocas, V., Michel, J. B., and Jandrot-Perrus, M. (2012) Emerging role of serpinE2/protease nexin-1 in hemostasis and vascular biology, *Blood* 119, 2452-2457.
77. Gils, A., and Declerck, P. J. (1998) Structure-function relationships in serpins: current concepts and controversies, *Thrombosis and haemostasis* 80, 531-541.
78. Carrell, R. W. (2004) What we owe to alpha(1)-antitrypsin and to Carl-Bertil Laurell, *Copd* 1, 71-84.
79. Kolarich, D., Weber, A., Turecek, P. L., Schwarz, H. P., and Altmann, F. (2006) Comprehensive glyco-proteomic analysis of human alpha1-antitrypsin and its charge isoforms, *Proteomics* 6, 3369-3380.
80. Pearce, M. C., and Cabrita, L. D. (2011) Production of recombinant serpins in *Escherichia coli*, *Methods in enzymology* 501, 13-28.
81. Beatty, K., Bieth, J., and Travis, J. (1980) Kinetics of association of serine proteinases with native and oxidized alpha-1-proteinase inhibitor and alpha-1-antichymotrypsin, *J Biol Chem* 255, 3931-3934.
82. Stoller, J. K., and Aboussouan, L. S. (2005) Alpha1-antitrypsin deficiency, *Lancet* 365, 2225-2236.
83. Futamura, A., Stratikos, E., Olson, S. T., and Gettins, P. G. (1998) Change in environment of the P1 side chain upon progression from the Michaelis complex to the covalent serpin-proteinase complex, *Biochemistry* 37, 13110-13119.
84. Scott, C. F., Schapira, M., James, H. L., Cohen, A. B., and Colman, R. W. (1982) Inactivation of factor XIa by plasma protease inhibitors: predominant role of alpha 1-protease inhibitor and protective effect of high molecular weight kininogen, *The Journal of clinical investigation* 69, 844-852.
85. Patston, P. A., Roodi, N., Schifferli, J. A., Bischoff, R., Courtney, M., and Schapira, M. (1990) Reactivity of alpha 1-antitrypsin mutants against proteolytic enzymes of the kallikrein-kinin, complement, and fibrinolytic systems, *J Biol Chem* 265, 10786-10791.
86. Felber, L. M., Kundig, C., Borgono, C. A., Chagas, J. R., Tasinato, A., Jichlinski, P., Gygi, C. M., Leisinger, H. J., Diamandis, E. P., Deperthes, D., and Cloutier, S. M. (2006) Mutant recombinant serpins as highly specific inhibitors of human kallikrein 14, *The FEBS journal* 273, 2505-2514.
87. Salahuddin, P. (2010) Genetic variants of alpha1-antitrypsin, *Current protein & peptide science* 11, 101-117.

88. Seo, E. J., Im, H., Maeng, J. S., Kim, K. E., and Yu, M. H. (2000) Distribution of the native strain in human alpha 1-antitrypsin and its association with protease inhibitor function, *J Biol Chem* 275, 16904-16909.
89. Johnson, D., and Travis, J. (1979) The oxidative inactivation of human alpha-1-proteinase inhibitor. Further evidence for methionine at the reactive center, *J Biol Chem* 254, 4022-4026.
90. Travis, J., Matheson, N. R., George, P. M., and Carrell, R. W. (1986) Kinetic studies on the interaction of alpha 1-proteinase inhibitor (Pittsburgh) with trypsin-like serine proteinases, *Biological chemistry Hoppe-Seyler* 367, 853-859.
91. Owen, M., Brennan, S., Lewis, J., and Carrell, R. (1983) Mutation of antitrypsin to antithrombin. alpha 1-antitrypsin Pittsburgh (358 Met leads to Arg), a fatal bleeding disorder, *New England Journal of Medicine* 309, 649-698.
92. Heeb, M. J., Bischoff, R., Courtney, M., and Griffin, J. H. (1990) Inhibition of activated protein C by recombinant alpha 1-antitrypsin variants with substitution of arginine or leucine for methionine358, *J Biol Chem* 265, 2365-2369.
93. Vidaud, D., Emmerich, J., Alhenc-Gelas, M., Yvart, J., Fiessinger, J. N., and Aiach, M. (1992) Met 358 to Arg mutation of alpha 1-antitrypsin associated with protein C deficiency in a patient with mild bleeding tendency, *The Journal of clinical investigation* 89, 1537-1543.
94. Scott, C. F., Carrell, R. W., Glaser, C. B., Kueppers, F., Lewis, J. H., and Colman, R. W. (1986) Alpha-1-antitrypsin-Pittsburgh. A potent inhibitor of human plasma factor XIa, kallikrein, and factor XIIIf, *The Journal of clinical investigation* 77, 631-634.
95. Schapira, M., Ramus, M. A., Jallat, S., Carvallo, D., and Courtney, M. (1986) Recombinant alpha 1-antitrypsin Pittsburgh (Met 358---Arg) is a potent inhibitor of plasma kallikrein and activated factor XII fragment, *The Journal of clinical investigation* 77, 635-637.
96. Izaguirre, G., Rezaie, A. R., and Olson, S. T. (2009) Engineering functional antithrombin exosites in alpha1-proteinase inhibitor that specifically promote the inhibition of factor Xa and factor IXa, *J Biol Chem* 284, 1550-1558.
97. Sheffield, W. P., Eltringham-Smith, L. J., Bhakta, V., and Gataiance, S. (2012) Reduction of thrombus size in murine models of thrombosis following administration of recombinant alpha1-proteinase inhibitor mutant proteins, *Thrombosis and haemostasis* 107.
98. Harper, P. L., Taylor, F. B., DeLa Cadena, R. A., Courtney, M., Colman, R. W., and Carrell, R. W. (1998) Recombinant antitrypsin Pittsburgh undergoes proteolytic cleavage during E. coli sepsis and fails to prevent the associated coagulopathy in a primate model, *Thrombosis and haemostasis* 80, 816-821.
99. Colman, R. W., Flores, D. N., De La Cadena, R. A., Scott, C. F., Cousens, L., Barr, P. J., Hoffman, I. B., Kueppers, F., Fisher, D., Idell, S., and et al. (1988) Recombinant alpha 1-antitrypsin Pittsburgh attenuates experimental gram-negative septicemia, *The American journal of pathology* 130, 418-426.
100. Kaiserman, D., Hitchen, C., Levina, V., Bottomley, S. P., and Bird, P. I. (2011) Intracellular production of recombinant serpins in yeast, *Methods in enzymology* 501, 1-12.
101. Karnaukhova, E., Ophir, Y., and Golding, B. (2006) Recombinant human alpha-1 proteinase inhibitor: towards therapeutic use, *Amino acids* 30, 317-332.

102. De Taeye, B., Gils, A., and Declerck, P. J. (2004) The story of the serpin plasminogen activator inhibitor 1: is there any need for another mutant?, *Thrombosis and haemostasis* 92, 898-924.
103. Hopkins, P. C., Pike, R. N., and Stone, S. R. (2000) Evolution of serpin specificity: cooperative interactions in the reactive-site loop sequence of antithrombin specifically restrict the inhibition of activated protein C, *Journal of molecular evolution* 51, 507-515.
104. Sutherland, J. S., Bhakta, V., and Sheffield, W. P. (2007) The appended tail region of heparin cofactor II and additional reactive centre loop mutations combine to increase the reactivity and specificity of alpha1-proteinase inhibitor M358R for thrombin, *Thrombosis and haemostasis* 98, 1014-1023.
105. Sutherland, J. S., Bhakta, V., and Sheffield, W. P. (2006) Investigating serpin-enzyme complex formation and stability via single and multiple residue reactive centre loop substitutions in heparin cofactor II, *Thrombosis research* 117, 447-461.
106. Brannigan, J. A., and Wilkinson, A. J. (2002) Protein engineering 20 years on, *Nature reviews. Molecular cell biology* 3, 964-970.
107. Antalis, T. M., and Lawrence, D. A. (2004) Serpin mutagenesis, *Methods* 32, 130-140.
108. Djie, M. Z., Stone, S. R., and Le Bonniec, B. F. (1997) Intrinsic specificity of the reactive site loop of alpha1-antitrypsin, alpha1-antichymotrypsin, antithrombin III, and protease nexin I, *J Biol Chem* 272, 16268-16273.
109. Plotnick, M. I., Schechter, N. M., Wang, Z. M., Liu, X., and Rubin, H. (1997) Role of the P6-P3' region of the serpin reactive loop in the formation and breakdown of the inhibitory complex, *Biochemistry* 36, 14601-14608.
110. Smith, G. P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface, *Science* 228, 1315-1317.
111. Sidhu, S. S., Lowman, H. B., Cunningham, B. C., and Wells, J. A. (2000) Phage display for selection of novel binding peptides, *Methods in enzymology* 328, 333-363.
112. Szardenings, M. (2003) Phage display of random peptide libraries: applications, limits, and potential, *Journal of receptor and signal transduction research* 23, 307-349.
113. Matochko, W. L., Chu, K., Jin, B., Lee, S. W., Whitesides, G. M., and Derda, R. (2012) Deep sequencing analysis of phage libraries using Illumina platform, *Methods* 58, 47-55.
114. Newton-Northup, J. R., and Deutscher, S. L. (2012) Contending With Target Unrelated Peptides from Phage Display, *J Mol Imaging Dynam* 2, 1-3.
115. Derda, R. (2013) Deep Sequencing Analyses of Phage Libraries, In *Ninth Annual Protein Engineering Summit*, Boston, MA.
116. Bratkovic, T. (2010) Progress in phage display: evolution of the technique and its application, *Cellular and molecular life sciences : CMLS* 67, 749-767.
117. Sidhu, S. S. (2001) Engineering M13 for phage display, *Biomolecular engineering* 18, 57-63.
118. Sidhu, S. S., Li, B., Chen, Y., Fellouse, F. A., Eigenbrot, C., and Fuh, G. (2004) Phage-displayed antibody libraries of synthetic heavy chain complementarity determining regions, *Journal of molecular biology* 338, 299-310.
119. Danner, S., and Belasco, J. G. (2001) T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries, *Proceedings of the National Academy of Sciences of the United States of America* 98, 12954-12959.
120. Missiakas, D., and Raina, S. (1997) Protein folding in the bacterial periplasm, *Journal of bacteriology* 179, 2465-2471.
121. Novagen (2011) *T7Select® System Manual*, Vol. TB178 Rev. D 0311JN.

122. Dai, M., Temirov, J., Pesavento, E., Kiss, C., Velappan, N., Pavlik, P., Werner, J. H., and Bradbury, A. R. (2008) Using T7 phage display to select GFP-based binders, *Protein engineering, design & selection : PEDS* 21, 413-424.
123. Krumpe, L. R., Atkinson, A. J., Smythers, G. W., Kandel, A., Schumacher, K. M., McMahon, J. B., Makowski, L., and Mori, T. (2006) T7 lytic phage-displayed peptide libraries exhibit less sequence bias than M13 filamentous phage-displayed peptide libraries, *Proteomics* 6, 4210-4222.
124. Pannekoek, H., van Meijer, M., Schleef, R. R., Loskutoff, D. J., and Barbas, C. F., 3rd. (1993) Functional display of human plasminogen-activator inhibitor 1 (PAI-1) on phages: novel perspectives for structure-function analysis by error-prone DNA synthesis, *Gene* 128, 135-140.
125. Berkenpas, M. B., Lawrence, D. A., and Ginsburg, D. (1995) Molecular evolution of plasminogen activator inhibitor-1 functional stability, *The EMBO journal* 14, 2969-2977.
126. Rosenberg, A., Griffin, K., Studier, F. W., McCormick, M., Berg, J., Novy, R., and Mierendorf, R. (1996) T7Select Phage Display System: A powerful new protein display system based on bacteriophage T7, *in* *Innovations*, 1-6.
127. Barbas, C. F., 3rd, Kang, A. S., Lerner, R. A., and Benkovic, S. J. (1991) Assembly of combinatorial antibody libraries on phage surfaces: the gene III site, *Proceedings of the National Academy of Sciences of the United States of America* 88, 7978-7982.
128. Jensen, J. K., and Gettins, P. G. (2008) High-resolution structure of the stable plasminogen activator inhibitor type-1 variant 14-1B in its proteinase-cleaved form: a new tool for detailed interaction studies and modeling, *Protein science : a publication of the Protein Society* 17, 1844-1849.
129. van Meijer, M., Roelofs, Y., Neels, J., Horrevoets, A. J., van Zonneveld, A. J., and Pannekoek, H. (1996) Selective screening of a large phage display library of plasminogen activator inhibitor 1 mutants to localize interaction sites with either thrombin or the variable region 1 of tissue-type plasminogen activator, *J Biol Chem* 271, 7423-7428.
130. Stoop, A. A., Jespers, L., Lasters, I., Eldering, E., and Pannekoek, H. (2000) 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), *Journal of molecular biology* 301, 1135-1147.
131. Stoop, A. A., Eldering, E., Dafforn, T. R., Read, R. J., and Pannekoek, H. (2001) 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, *Journal of molecular biology* 305, 773-783.
132. Zani, M. L., and Moreau, T. (2010) Phage display as a powerful tool to engineer protease inhibitors, *Biochimie* 92, 1689-1704.
133. Deperthes, D. (2002) Phage display substrate: a blind method for determining protease specificity, *Biological chemistry* 383, 1107-1112.
134. Matthews, D. J., and Wells, J. A. (1993) Substrate phage: selection of protease substrates by monovalent phage display, *Science* 260, 1113-1117.
135. Cloutier, S. M., Kundig, C., Felber, L. M., Fattah, O. M., Chagas, J. R., Gygi, C. M., Jichlinski, P., Leisinger, H. J., and Deperthes, D. (2004) Development of recombinant inhibitors specific to human kallikrein 2 using phage-display selected substrates, *European journal of biochemistry / FEBS* 271, 607-613.
136. Jegot, G., Derache, C., Castella, S., Lahouassa, H., Pitois, E., Jourdan, M. L., Remold-O'Donnell, E., Kellenberger, C., Gauthier, F., and Korkmaz, B. (2011) A substrate-based approach to convert SerpinB1 into a specific inhibitor of proteinase 3, the Wegener's

- granulomatosis autoantigen, *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 25, 3019-3031.
137. Gierczak, R. F., Sutherland, J. S., Bhakta, V., Toltl, L. J., Liaw, P. C., and Sheffield, W. P. (2011) Retention of thrombin inhibitory activity by recombinant serpins expressed as integral membrane proteins tethered to the surface of mammalian cells, *Journal of thrombosis and haemostasis : JTH* 9, 2424-2435.
  138. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids, *Nucleic acids research* 38, W529-533.
  139. Lee, H. J., and Im, H. (2003) Purification of recombinant plasminogen activator inhibitor-1 in the active conformation by refolding from inclusion bodies, *Protein expression and purification* 31, 99-107.
  140. Filion, M. L., Bhakta, V., Nguyen, L. H., Liaw, P. S., and Sheffield, W. P. (2004) Full or partial substitution of the reactive center loop of alpha-1-proteinase inhibitor by that of heparin cofactor II: P1 Arg is required for maximal thrombin inhibition, *Biochemistry* 43, 14864-14872.
  141. Wirsching, F., Keller, M., Hildmann, C., Riester, D., and Schwienhorst, A. (2003) Directed evolution towards protease-resistant hirudin variants, *Molecular genetics and metabolism* 80, 451-462.
  142. Cunningham, M. A., Bhakta, V., and Sheffield, W. P. (2002) Altering heparin cofactor II at VAL439 (P6) either impairs inhibition of thrombin or confers elastase resistance, *Thrombosis and haemostasis* 88, 89-97.
  143. Tikhonov, D. B., and Zhorov, B. S. (2007) Sodium channels: ionic model of slow inactivation and state-dependent drug binding, *Biophysical journal* 93, 1557-1570.
  144. Chaillan-Huntington, C. E., and Patston, P. A. (1998) Influence of the P5 residue on alpha1-proteinase inhibitor mechanism, *J Biol Chem* 273, 4569-4573.
  145. Elliott, P. R., Lomas, D. A., Carrell, R. W., and Abrahams, J. P. (1996) Inhibitory conformation of the reactive loop of alpha 1-antitrypsin, *Nature structural biology* 3, 676-681.
  146. Le Bonniec, B. F., Guinto, E. R., and Stone, S. R. (1995) Identification of thrombin residues that modulate its interactions with antithrombin III and alpha 1-antitrypsin, *Biochemistry* 34, 12241-12248.
  147. Whisstock, J., Lesk, A. M., and Carrell, R. (1996) Modeling of serpin-protease complexes: antithrombin-thrombin, alpha 1-antitrypsin (358Met-->Arg)-thrombin, alpha 1-antitrypsin (358Met-->Arg)-trypsin, and antitrypsin-elastase, *Proteins* 26, 288-303.
  148. Chaillan-Huntington, C. E., Gettins, P. G., Huntington, J. A., and Patston, P. A. (1997) The P6-P2 region of serpins is critical for proteinase inhibition and complex stability, *Biochemistry* 36, 9562-9570.
  149. Irving, J. A., Pike, R. N., Dai, W., Bromme, D., Worrall, D. M., Silverman, G. A., Coetzer, T. H., Dennison, C., Bottomley, S. P., and Whisstock, J. C. (2002) Evidence that serpin architecture intrinsically supports papain-like cysteine protease inhibition: engineering alpha(1)-antitrypsin to inhibit cathepsin proteases, *Biochemistry* 41, 4998-5004.

## APPENDIX A

This appendix provides a summary of every published  $\alpha_1$ -PI mutant, as of August 2013.

**Table A: List of Mutations made at  $\alpha_1$ -PI RCL Positions P17 – P10'.** “+ w. mut.”

column indicates how many of the  $\alpha_1$ -PI mutants contain a mutation at the corresponding RCL position. “Out of 21” column indicates the diversity of mutations made at the RCL position, scored out of 21 for each possible amino acid or a deletion. The other columns correspond to how many times the indicated amino acid has been introduced at the position, with the wild type residues outlined with a single line, and naturally occurring mutations outlined with a double line. The P1 RCL position has been investigated the most thoroughly, with 89 mutants containing a mutant P1 residue, and with 10 of the possible 21 amino acids or a deletion having been introduced. Coloured based on frequency (white: low, black: high). **Table B: Goals of  $\alpha_1$ -PI Mutants and References.**

A summary of the goals for engineering  $\alpha_1$ -PI mutants is presented. Modifying protease specificity has been the primary goal, representing 73 of the 127 mutants. PubMed IDs of the source papers are listed for reference.

**A**

	# w. mut.	Out of 21	W	Y	F	L	I	V	M	A	G	C	P	T	S	Q	N	E	D	H	K	R	Δ	
P17	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
P16	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0
P15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P14	5	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0
P13	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0
P12	4	3	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
P11	7	3	0	0	0	0	0	2	0	0	0	0	0	4	0	0	0	1	0	0	0	0	0	0
P10	11	4	0	0	0	0	0	0	0	4	0	0	1	5	0	0	0	1	0	0	0	0	0	0
P9	8	3	0	0	0	0	0	5	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	0
P8	17	4	0	0	0	0	0	2	0	2	0	0	0	9	0	0	0	4	0	0	0	0	0	0
P7	22	4	0	0	0	1	0	0	0	13	0	0	0	5	0	0	0	3	0	0	0	0	0	0
P6	23	5	0	0	0	0	2	16	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2	1
P5	28	6	0	0	0	0	0	13	0	0	9	0	0	2	0	1	0	0	0	1	0	0	2	0
P4	37	8	0	0	6	2	12	2	0	0	4	0	0	1	0	0	0	0	0	0	0	0	6	4
P3	32	8	0	0	4	1	0	1	6	13	0	0	0	0	0	0	0	2	2	0	0	0	0	3
P2	27	6	0	0	0	0	0	0	0	4	17	0	0	0	0	1	0	0	1	0	2	2	0	0
<b>P1</b>	<b>89</b>	<b>10</b>	1	1	1	8	1	3	0	1	0	1	0	0	4	0	0	0	0	0	0	0	<b>68</b>	0
<b>P1'</b>	7	3	0	0	0	0	0	0	0	2	0	0	0	3	0	0	2	0	0	0	0	0	0	0
P2'	29	5	0	0	0	12	0	0	0	2	0	1	4	10	0	0	0	0	0	0	0	0	0	0
P3'	24	4	0	0	0	2	0	0	0	4	0	0	0	0	0	9	9	0	0	0	0	0	0	0
P4'	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P5'	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
P6'	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P7'	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P8'	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P9'	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P10'	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

**B**

Goal	Oxidation Resistance	Identifying Functional Constraints	Improving Recombinant Production	Modifying Protease Specificity
# of mutants	9	38	7	73
PubMed ID	6151045 6387509 3006558 2102901 10867014 12009885 3872299 3525544	8144550 11700071 9748317 11325972 11714919 9446551 8347575 2009965 12498804 9468513 8031789 9624149 10220375	7764716 12034861 11690693 16546284 2826966 21972922	19010776 16981704 2191958 15544357 6604220 6333253 3484756 3497179 3278932 3257651 1765073 1644824 8227051 1569192 3509863 7495819 9843177 9636142 11584020 12860985 7707552 9707432 11062061 11479287 18000606 9930674 11939796 12787886 23552726 16704423 2298753 3539143 412531 7547966 3484755 11080374 3541206 22370677 2503514 9598975 7744836 15659365 18061307 9236002

**Total 127**

## APPENDIX B

Chris Duffin <chris@enzymeresearch.com>  
To: "scottbm@mcmaster.ca" <scottbm@mcmaster.ca>

Mon, Jul 15, 2013 at 2:31 PM

Hi Ben  
Thanks for checking with us. Please feel free to use as described.  
Best regards,

-Chris

Begin forwarded message:

**From:** Enzyme Research Laboratories <info@enzymeresearch.com>  
**Date:** July 12, 2013, 4:47:12 PM EDT  
**To:** Chris Duffin <chris@enzymeresearch.com>  
**Subject:** FW: Requesting use of a figure (coagulation cascade)

~Chas

---

**From:** Ben Scott [mailto:scottbm@mcmaster.ca]  
**Sent:** Friday, July 12, 2013 4:41 PM  
**To:** Enzyme Research Laboratories  
**Subject:** Requesting use of a figure (coagulation cascade)

To whom it may concern,

I would like to request usage of the Enzyme Research Laboratory's "Global Coag" (coagulation) figure for my Master's thesis. It provides the most comprehensive view of coagulation that I've been able to find, and it would compliment the introductory section in my thesis.

Written citation beside the figure and within the body of my thesis will be included in order to provide full credit to Enzyme Research Laboratory.

Sincerely,

-Ben Scott

Bioscience San Diego Tech Service <bioscienceshelp@emdchemicals.com>  
To: "scottbm@mcmaster.ca" <scottbm@mcmaster.ca>

Tue, Jul 16, 2013 at 12:13 PM

Dear Mr Scott,

Thank you for contacting EMD Millipore.

It will be of no problem for you to reproduce the figures mentioned in your email message for your thesis or other non-commercial uses provided that you mention us (Novagen / EMD Millipore) as the source of your figures.

Sincerely,

EMD Millipore Technical Service Team