

ENGINEERING SUBSTRATE SPECIFICITY OF MAMMALIAN PROTEASES

A Dissertation

Presented to

the Faculty of the Department of Chemical and Biomolecular Engineering

University of Houston

In partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

in Chemical Engineering

by

Balakrishnan Ramesh

August 2015

ENGINEERING SUBSTRATE SPECIFICITY OF MAMMALIAN PROTEASES

Balakrishnan Ramesh

Approved:

Chair of the Committee
Navin Varadarajan, PhD
Assistant Professor
Chemical and Biomolecular Engineering

Committee Members:

Patrick C. Cirino, PhD
Associate Professor
Chemical and Biomolecular Engineering
Biology and Biochemistry

Jacinta C. Conrad, PhD
Assistant Professor
Chemical and Biomolecular Engineering

Timothy Cooper, PhD
Associate Professor
Biology and Biochemistry

Laura Segatori, PhD
Associate Professor
Chemical and Biomolecular Engineering
Bioengineering & Biochemistry
Cell Biology

Suresh K. Khator, PhD
Associate Dean
Cullen College of Engineering

Michael P. Harold, PhD
Chair and M.D. Anderson Professor
Chemical and Biomolecular Engineering

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Navin Varadarajan, for his constant guidance to navigate through the difficult maze of research. He was an inspiration for strong work ethics.

I enjoyed a great work environment with the past and present members in the laboratories of Drs. Varadarajan, Cirino, Willson and Conrad. I appreciate the contribution of our collaborator, Dr. Bark's lab, in mass spectrometry experiments and analysis. I also thank my external committee members, Drs. Cooper and Segatori.

This PhD would have been less enjoyable, if not impossible, without my family and my friends in Houston and around the world. I dedicate the successful completion of the degree to my parents. Sorry Amma, I didn't know it was the other doctor!

ENGINEERING SUBSTRATE SPECIFICITY OF MAMMALIAN PROTEASES

An Abstract
of a
Dissertation
Presented to
the Faculty of the Department of Chemical and Biomolecular Engineering
University of Houston

In partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
in Chemical Engineering

by
Balakrishnan Ramesh

August 2015

ABSTRACT

Proteases, enzymes that catalyze the hydrolysis of amide bonds in peptides and proteins, are ubiquitous across all forms of life and affect every protein in its natural life cycle. Engineering programmed specificity onto stable protease scaffolds holds promise as a route to a new generation of useful enzymes with requisite clinical and biotechnological properties. Combinatorial approach of screening targeted protein libraries (size $> 10^7$) has been successfully applied for engineering the substrate specificity of disulfide-free microbial proteases. To date however, the largest size of the library screened for any mammalian protease system is 150. This is primarily due to paucity of high-throughput screening platforms that are compatible with post-translational modifications such as zymogen activation and disulfide bond formation typically required for functional mammalian proteases. In the present study, we have established a comprehensive methodology that allows efficient FACS screening of large libraries of recombinant proteases in *E.coli* by developing and integrating three different components: (i) surface display system compatible with mammalian proteases such as chymotrypsin and caspase-3, (ii) protease activity assay with single cell resolution, and (iii) genotype recovery method to facilitate iterative enrichment of desired phenotype.

Characterizing post-translational modifications (PTMs) has been of increasing interest due to their biological functions and this knowledge is used for novel target discovery and drug development. Since PTMs can be labile, proteomic techniques relying on tandem MS have difficulty in achieving comprehensive coverage of PTM sites within proteins. Proteases which selectively recognize PTMs may improve their identification in complex protein mixtures and are easily adaptable to current proteomic methods. Using

our FACS-based methodology, we have screened large ($> 10^7$) libraries of chymotrypsin, and engineered its substrate binding pocket to cleave after Asn residues for mapping N-linked glycosylation sites. In comparison to the wild-type chymotrypsin, the engineered variant displayed a 2×10^4 -fold reversal in catalytic selectivity for asparagine at P1 position over tyrosine. Unlike engineered microbial proteases, we demonstrate that the ability of the engineered chymotrypsin variant to cleave after Asn in peptide substrates translates to full length proteins, an essential attribute for application to proteomics. Our results show that comprehensive engineering of the substrate specificity of mammalian proteases is feasible and when executed in the context of chymotrypsin, can enable the detection of various kinds of PTMs.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
ABSTRACT	vi
TABLE OF CONTENTS.....	viii
LIST OF FIGURES	xi
LIST OF TABLES	xiii
CHAPTER 1 ENGINEERING SUBSTRATE SPECIFICITY OF PROTEASES	1
1.1 Introduction.....	1
1.2 Significance of engineering the specificity of proteases.....	4
1.2.1 Engineering therapeutic proteases to overcome endogenous inhibitors.....	4
1.2.2 Engineering extended specificity to decrease off-target proteolysis	8
1.2.3 Engineering catalytic activity towards novel substrates.....	12
1.3 Specificity determinants of proteases are tunable by evolution.....	13
1.3.1 Evolution of substrate specificity in natural proteases	13
1.3.2 Engineering proteases by grafting specificity determinants.....	16
1.3.3 Success of directed evolution approach in engineering specificity	18
1.4 Tools of the trade: Advances in technologies for engineering specificity.....	20
1.4.1 Computational tools for constructing targeted libraries	21
1.4.1.1 Homology guided mutagenesis.....	21
1.4.1.2 Structure and sequence guided mutagenesis.....	22
1.4.1.3 <i>In silico</i> mutagenesis based on energy minimization	24
1.4.2 Platforms for high-throughput characterization	25
1.4.2.1 Platforms based on fluorescence activated cell sorting	26
1.4.2.2 Microfluidic devices for cell sorting.....	29
1.4.2.3 <i>In vivo</i> continuous evolution.....	31
1.4.3 Techniques for profiling specificity of engineered proteases.....	32
1.4.3.1 Selection/ screening of genetically encoded substrate libraries.....	32
1.4.3.2 Mass spectrometry based approaches	34
1.4.3.3 Quantitative characterization of specificity	35
1.5 An outlook for the future of engineering protease specificity	37
1.6 Conclusions.....	41

CHAPTER 2 BACTERIAL PLATFORM FOR ANCHORED DISPLAY OR SECRETION OF RECOMBINANT PROTEINS WITH DISULFIDE BONDS	43
2.1 Introduction.....	43
2.2 Experimental Procedures	47
2.2.1 Plasmid construction.....	47
2.2.2 Expression and labeling of M18 scFv	48
2.2.3 Expression and labeling of rChyB.....	50
2.2.4 Flow-cytometry	51
2.2.5 Immunofluorescence microscopy	51
2.2.6 Western blotting	52
2.3 Results.....	53
2.3.1 Surface display of functional single-chain antibody via fusion to Ag43 revealed by flow-cytometry	53
2.3.2 C-terminal domains comprising the β -barrel and the α -helix are sufficient to achieve efficient surface display.....	55
2.3.3 Binding of bulky antigens is not sterically hindered	56
2.3.4 The extended signal peptide region is not indispensable for functional surface display.....	57
2.3.5 Surface display of functional M18 scFv is not improved by the addition of reducing agents or the use of protease-deficient strains	58
2.3.6 Surface display of chymotrypsin	59
2.3.7 rChyB displayed on the surface is enzymatically active	62
2.4 Discussion	64
2.5 Conclusions.....	67
CHAPTER 3 HIGH-THROUGHPUT SCREENING OF RECOMBINANT PROTEASE LIBRARIES IN <i>E.COLI</i>	68
3.1 Introduction.....	68
3.2 Materials and Methods.....	72
3.2.1 Molecular cloning.....	72
3.2.2 Crude membrane preparation	72
3.2.3 Enzyme characterization using HPLC.....	73
3.2.4 Substrate synthesis.....	73
3.2.5 Plasmid recovery from sorted cells	74
3.2.6 Phenotype enrichment by FACS sorting	75

3.3 Results.....	76
3.3.1 Functional characterization of proteases displayed on the bacterial surface....	76
3.3.2 Protease activity assay improved by optimizing pH, cell density and charge of the FRET peptide substrate.....	78
3.3.3 Phenotype enrichment by direct plasmid recovery after FACS	81
3.4 Discussion	83
3.5 Conclusions.....	86
CHAPTER 4 ENGINEERING SUBSTRATE SPECIFICITY OF CHYMOTRYPSIN FOR MAPPING N-GLYCOSYLATION SITES	87
4.1 Introduction.....	87
4.2 Material and Methods	90
4.2.1 Construction of chymotrypsin libraries	90
4.2.2 Synthesis of FRET peptide substrates	92
4.2.3 Screening of chymotrypsin libraries using FACS	94
4.2.4 Enzyme expression and purification.....	95
4.2.5 HPLC characterization	100
4.2.6 MS/MS characterization	101
4.2.7 Kinetic parameter measurements.....	102
4.3 Results.....	103
4.3.1 Surface loops critical for substrate specificity randomized in a chymotrypsin B library.....	103
4.3.2 Isolation of rChyB-N-v1 variant with activity towards Asn using FACS.....	105
4.3.3 Minimal number of mutations required for activity towards P1 Asn identified by backcross shuffling.....	109
4.3.4 Purified rChyB-N-v2 is functional towards the peptide substrate containing Asn.....	111
4.3.5 Improved proteolytic activity towards P1 Asn by random mutagenesis	113
4.3.6 Variants with high Asn activity displayed increase in polarity of active site.	116
4.3.7 Kinetic characterization of engineered chymotrypsin with fluorogenic peptide substrates	118
4.3.8 Engineered protease cleaves after Asn in protein substrate	120
4.4 Discussion	122
4.5 Conclusions.....	124
REFERENCES	125

LIST OF FIGURES

Figure 1.1 Serpin regulation of coagulation, protein C and fibrinolytic pathways.....	10
Figure 1.2 Similarity of active site architecture in serine proteases.	15
Figure 1.3 Sequence alignment of human matrix metalloproteinases.	17
Figure 1.4 Engineering of specificity towards different residues at P1 and P1` positions from a single parental enzyme, outer membrane protease T (OmpT).	19
Figure 1.5 Graphical user interface of HotSpot Wizard results.....	23
Figure 1.6 Schematic directed evolution cycles by lysate screening in droplets.....	31
Figure 1.7 Proteomic identification of cleavage sites (PICS) library generation and specificity assay	34
Figure 2.1 Schematic describing the transport of AT to the outer membrane.	54
Figure 2.2 Surface expression of functional M18 scFv as a function of native passenger length.....	55
Figure 2.3 Surface expression of functional M18 as a function of N-terminal leader sequence.....	57
Figure 2.4 Effect of reducing agent and host outer membrane protease on the surface display of functional M18 scFv... ..	60
Figure 2.5 Flow-cytometry analysis of cells expressing rChyB labeled with biotin conjugated anti-ChyB antibody (chain C specific) and streptavidin-PE	61
Figure 2.6 Phase contrast and immunofluorescence microscopy showing surface display of chymotrypsin.. ..	62
Figure 2.7 Chymotrypsin displayed on the surface of <i>E. coli</i> is enzymatically active... ..	63
Figure 2.8 Inhibition of the catalytic activity of surface displayed chymotrypsin... ..	64
Figure 3.1 HPLC analysis of surface displayed caspase-3 and chymotrypsin using peptide substrates.....	77
Figure 3.2 Effect of pH and cell density on flow-cytometric protease assay	79
Figure 3.3 Concentration of the peptide bound to the surface of <i>E.coli</i> MC1061 cells decreases with the charge of the peptide-fluorophore conjugate.....	80
Figure 3.4 Comparison of phenotype enrichment by plasmid recovery and culturing methods in chymotrypsin expression system.....	83
Figure 4.1 Chemical synthesis of N-AtQ21, Asn FRET peptide substrate.... ..	93
Figure 4.2 Schematic of chymotrypsinogen forms used for protein expression.....	97
Figure 4.3 Sequence alignment of chymotrypsin-fold proteases with differing substrate specificities and species origin.....	105

Figure 4.4 Schematic for high-throughput screening of chymotrypsin libraries using FRET peptide substrates for iterative enrichment of active variants showing specificity towards Asn substrate. .	107
Figure 4.5 Evolution of naïve chymotrypsin B library to a population enriched with variants displaying high Asn activity through multiple rounds of sorting.....	108
Figure 4.6 Western blot showing chymotrysinogens expressed in mammalian cells....	112
Figure 4.7 HPLC analysis of purified wild-type rChyB or engineered rChyB-N-v2 incubated with N-AtQ21, FRET peptide substrate containing Asn.....	112
Figure 4.8 HPLC analysis of rChyB-N-v3 with unconjugated peptide substrate.	114
Figure 4.9 LC-MS analysis of N-AtQ21 peptide substrate digested with (A) wild-type chA-rChyB and (B) engineered chA-rChyB-N-v3..	115
Figure 4.10 Electrostatic surface potentials of wild-type and engineered chymotrypsins. .	116
Figure 4.11 Kinetic comparison of wild-type (chA-rChyB) and engineered chymotrypsins (chA-rChyB-N-v3) with fluorogenic peptide substrates.....	119
Figure 4.12 Sequence coverage of bovine serum albumin (BSA) based on MS/MS analysis of peptides generated by digestion with engineered chymotrypsin, chA-rChyB-N-v3..	121

LIST OF TABLES

Table 1.1 Proteases with therapeutic potential improved by engineering native substrate specificity.....	5
Table 2. 1 List of primers used in this study.	49
Table 2. 2 List of plasmids used in this study.	50
Table 4.1 List of primers used in this study.....	91
Table 4.2 Library construction for shuffling engineered variant, rChyB-N-v1, with wild-type rChyB.	92
Table 4.3 Analysis of individual clones randomly picked from the site-saturated chymotrypsin library enriched by six rounds of sorting.	107
Table 4.4 Mass spectrometry analysis of N-AtQ21 substrate digested with cells expressing inactive rChyB or rChyB-N-v1 to identify peptide fragments generated by proteolysis.	109
Table 4.5 Analysis of individual clones randomly picked from the population enriched by four rounds of sorting the library obtained by backcrossing rChyB-N-v1 with rChyB. .	110
Table 4.6 Analysis of individual clones randomly picked from the population enriched by four rounds of sorting the library obtained by error-prone PCR with rChyB-N-v2 as template.....	113
Table 4.7 Hydropathic analysis of chymotrypsin variants.....	117

Chapter 1

Engineering substrate specificity of proteases

1.1 Introduction

Proteases, enzymes that catalyze the hydrolysis of amide bonds in peptides and proteins, are ubiquitous across all forms of life and affect every protein in its natural life cycle. By virtue of site-specific processing, proteases have a profound effect on biological processes ranging from simple molecular switches altering activities of protein substrates (Davie et al., 1991; Kitamoto et al., 1994) to enabling migration and differentiation of cells (Lamkanfi et al., 2007; Seeds et al., 1997) and finally to orchestrating tissue remodeling and immune responses (Heutinck et al., 2010; Saito, 2005). Consequently, dysregulated proteases have been shown to be pathological in several disorders including hypertension (Mullins et al., 1990), diabetes (Marguet et al., 2000), rheumatoid arthritis (Jørgensen et al., 2011) and tumor progression (Gialeli et al., 2011). The increased protease activity in diseased tissues has been utilized for targeted delivery of therapeutics and toxins (Choi et al., 2012; Desnoyers et al., 2013). However, in most cases, the therapeutic strategy involves inhibition of the pathological proteolytic activity. Protease inhibitors account for 5-10% of all drug targets and make a ~\$12 billion industry (Drag and Salvesen, 2010). Viral protease inhibitors are one of the key components of highly active antiretroviral therapy (HAART) which substantially decreases mortality in the HIV infected patients (Hernández-Ávila et al., 2015; Wensing et al., 2010).

Proteases when directly used as therapeutics can inactivate multiple target molecules as opposed to the simple stoichiometric drugs such as antibodies and small molecules. Urokinase was the first protease approved as a therapeutic for myocardial infarction and pulmonary embolism to dissolve clots that block arteries (Brochier, 1978). Since then, several proteases involved in the coagulation cascade have been developed as anticoagulants to treat ailments such as stroke, heart attack and muscle atrophy depending on the location of the clot formation. Factors IX and VIIa which promote clot formation have also been developed as therapeutics for platelet dysfunction and hemophilia (Craik et al., 2011). Thrombin, a procoagulant protease, is used for control of surface bleeding (in patches/bandages) and surgical hemostasis (Bowman et al., 2010). Significant improvements in the pharmacokinetic properties have been achieved by engineering the native form of the protease. Extension of half-life using several strategies such as Fc fusion (Mahlangu et al., 2014), albumin fusion (Mannucci, 2015), domain truncation (Bode et al., 1996) and glycosylation (Bolt et al., 2012) and decrease in immunogenicity by PEGylation (Liu et al., 2014) and site-directed mutagenesis of epitopes corresponding to anti-drug antibodies (Nguyen et al., 2014) have increased the bioavailability of therapeutic agents. Protease variants with faster kinetics have also been engineered for better efficacy as in the case of factor VIIa for treatment of hemophilia (Holmberg et al., 2009).

Proteases collectively account for > 40 % of total enzyme sales worldwide with the major consumers being laundry detergent and food industries for processes such as milk clotting, meat tenderizing and gluten removal (Feijoo-Siota and Villa, 2010; Li et al., 2013). Several applications are emerging in the field of sustainable biotechnology. Use of

serine proteases in enzymatic hydrolysis of wastes from fishing industry to produce bioactive peptides and as a nitrogenous source for microbial media has been recently demonstrated (Najafian and Babji, 2012; Safari et al., 2009). Similarly, keratinases and collagenases are used in processing recalcitrant wastes such as feathers and hides from poultry and meat industries (Gupta et al., 2013). With rise in the awareness about environmental impact of conventional pesticides, the possibility of using proteases identified from plant defense systems, from microbes which infect insects and from venoms of arthropod predators is currently being explored (Harrison and Bonning, 2010). Initial engineering efforts modifying the natural enzymes mainly focused on improving thermal and chemical stability for better performance in non-native environments using different techniques such as random mutagenesis, introduction of disulfide bonds and surface charge optimization (Jakob et al., 2013; Vojcic et al., 2015). Recently variants with active site mutations are also characterized for their ability to catalyze non-amidolytic reactions such as ligation and perhydrolysis (Despotovic et al., 2013; Franke et al., 2013). Such enzymes could lead to manufacturing processes with green chemistry and less energy-intensive operating conditions apart from higher yield of fractions with desired regioselectivity. PT121 protease isolated from *P. aeruginosa* has been used in the enzymatic synthesis of Z-aspartame (low calorie sweetener) by ligation of N-protected aspartic acid and L-phenylalanine methyl ester (Liu et al., 2015). It has also been used in enzymatic synthesis of the analgesic protein endomorphin-1 from dipeptides Tyr-Pro and Trp-Phe (Sun et al., 2011).

Thus, proteases are an extremely useful class of proteins, and protease engineering has substantially increased their usability potential. In the present review, we will focus

on a subset of protease engineering studies which aim to modify the native substrate specificity for various clinical and biotechnological applications (**Table 1.1**). Protease specificity is described using the nomenclature outlined by Berger (Schechter and Berger, 1967) where amino acids N-terminal to the scissile bond are designated P1, P2, P3 etc. starting at the scissile bond and moving outwards. Similarly, amino acids C-terminal to the scissile bond are designated P1', P2', P3' etc. The corresponding substrate binding pockets on the protease are designated S1-Sn, and S1'-Sn' respectively.

1.2 Significance of engineering the specificity of proteases

1.2.1 Engineering therapeutic proteases to overcome endogenous inhibitors

Serpins refer to a superfamily of proteins classified based on phylogeny and the majority of them are large (330 - 500 amino acids), irreversible serine protease inhibitors though some of them affect cysteine proteases as well. *In vivo*, they perform a variety of functions such as regulation of inflammatory response, prevention of unregulated apoptotic activation, tumour immunity, inactivation of pathogen-derived proteases and even non-inhibitory roles such as hormone transport (Law et al., 2006; Mangan et al., 2008). Serpins, plasminogen activator inhibitor-1 (PAI-1) and α_2 -antiplasmin, regulate the anticoagulant proteases, Tissue-type plasminogen activator (tPA) and plasmin, to control the process of fibrinolysis i.e dissolving blood clots (**Figure 1.1**) (Rau et al., 2007). Life-threatening conditions such as heart attack, stroke and pulmonary embolism could ensue if blood clots blocking fine arteries in essential organs are not immediately removed by fibrinolysis. The standard of care treatment for such emergencies involves administering tPA in an initial bolus dose within 60 min of onset of symptoms followed by a constant infusion for 90 min (Jauch et al., 2010).

Table 1.1 Proteases with therapeutic potential improved by engineering native substrate specificity.

Protease	Enzyme classification no.	Engineering	Target disease	Publication	Stage of development
Tissue-type plasminogen activator (PA)	3.4.21.68	Decreased activity towards PA inhibitor type 1 without loss in activity towards plasminogen	Acute myocardial infarction	(Tachias, 1997)	Clinical use
Granzyme B	3.4.21.79	Decreased activity towards serpin B9 without loss in activity towards pro-apoptotic substrates	Tumor	(Losasso et al., 2012)	Pre-clinical
Activated protein C	3.4.21.69	Decreased activity towards factor Va (coagulant) without loss in activity towards protease-activated receptor 1 (pro-inflammatory)	Septic shock	(Mosnier et al., 2007a)	Withdrawn
Thrombin	3.4.21.5	Decreased activity towards fibrinogen and protease-activated receptor 1 (pro-coagulant) without loss in activity towards protein C (anti-coagulant)	Ischemic stroke	(Marino et al., 2010)	Pre-clinical
Neprilysin	3.4.24.11	Decreased activity towards 8 off-target substrates such as neurotensin and angiotensin while improving activity towards amyloid beta peptide by 20-fold	Alzheimer's disease	(Webster et al., 2014)	In vitro
Kumamolisin - As	3.4.21.B48	Switched P2 specificity from R to Q to cleave gluten-derived immunogenic peptides	Celiac disease	(Gordon et al., 2012)	In vitro

tPA binds to the fibrin network in clots and cleaves plasminogen in circulation to generate active plasmin which subsequently hydrolyzes fibrin and also activates inhibitors corresponding to pro-coagulant proteases. As recombinant tPA used in the

clinic had a half-life of only 3.5 min, it was hypothesized that engineering resistance towards PAI-1 while preserving or improving the catalytic efficiency of tPA could improve the efficacy of the treatment (Huber, 2001). Inactivation of tPA is initiated by the interaction between the positively charged surface loop (positions 296-299) of tPA and the negatively charged reactive center region (positions 350-355) of PAI-1 (Madison et al., 1989). Typically, when a protease cleaves the reactive center loop region of a serpin molecule, its active site is distorted by the conformational change induced by serpin. Therefore the subsequent hydrolysis reaction does not occur and the protease is trapped in the acyl intermediate state (Rau et al., 2007). When charge reversal mutations (Lys296Glu, Arg298Glu and Arg299Glu) were introduced to disrupt electrostatic interactions between tPA and PAI-1, the resulting tPA variant showed substantially lower affinity ($K_i = 500 \text{ M}^{-1}\text{s}^{-1}$) towards PAI-1 in comparison to the wild-type ($K_i = 1.4 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$) while retaining the plasminogen activation kinetics (Madison et al., 1990). Indeed, Tenecteplase, an engineered tPA variant containing these mutations, showed 6-fold higher half-life *in vivo* and lower incidence of intracranial hemorrhage due to 15-fold increase in fibrin specificity, ultimately allowing administration in 5-10 seconds as a single bolus dose without the previous requirement of a follow-up infusion over 90 min (Tanswell et al., 2002). It is to be noted that Tenecteplase contains two other additional modifications, Thr103Asn mutation to create a new N-linked glycosylation site and Asn117Gln to remove a high mannose glycosylation site possibly involved in faster clearance. Therefore exact contribution of serpin resistance mutations towards improvement in efficacy is not known.

Granzyme B is a pro-apoptotic serine protease involved in clearance of cells infected with pathogens, tumour cells and other harmful cells. The possibility of using granzyme B as an immunotoxin to target tumours by fusion to an antibody molecule with high affinity towards a specific tumour antigen is a promising therapeutic strategy. As opposed to immunotoxins of bacterial or chemical origin which lead to generation of anti-drug antibodies by the patient's immune system, granzyme B being a human protein is expected to be less immunogenic (Berges et al., 2014). When mice grafted with a breast cancer model known to express fibroblast growth factor antigen (Fn14) were treated with granzyme B fused to a single chain antibody against Fn14, significant tumour growth inhibition was observed in comparison to saline or granzyme B only control (Zhou et al., 2014). Physiologically, activity of granzyme B is regulated by Serpin B9 for cytoprotection of lymphocytes. However, overexpression of Serpin B9 in several tumours is used as an escape mechanism to avoid immune surveillance (Ray et al., 2012; van Houdt et al., 2005). With the objective of an immunotoxin that would be effective even in the presence of endogenous protease inhibitors in serum and tumour microenvironment, computational alanine scanning method was used to identify the residues that contribute towards the binding free energy of granzyme B: Serpin B9 complex (Losasso et al., 2012). The choice of mutations was made based on the degree of instability introduced in the protease:serpin complex assessed using molecular dynamics simulations. Surprisingly, a single mutation with amino acid of similar identity (Arg201Lys), allowed 94% retention of activity even after incubation with 3-fold molar excess of inhibitor at 37 °C for 1 h. When primary mononuclear cells isolated from blood of acute myelomonocytic leukemia patients were treated with engineered granzyme B fused to an

antibody fragment targeted against CD64 antigen, the enzyme was able to induce apoptosis selectively in CD64⁺ leukemic cells and the reduction in cell viability was superior in comparison to the wild-type granzyme B fusion protein (Schiffer et al., 2014). Efficacy of engineered granzyme B fused to a single chain variable fragment (scFv) directed against CD30 antigen in classical Hodgkin lymphoma cells known to overexpress Serpin B9 (Bladergroen et al., 2002) was demonstrated in a mouse model (Schiffer et al., 2013).

1.2.2 Engineering extended specificity to decrease off-target proteolysis

Physiologically, most of the proteases cleave multiple substrates and are involved in disparate pathways (Overall and Dean, 2006; Spronk et al., 2014). When such a protease is used as a therapeutic, toxicity due to off-target cleavages undermines the efficacy of the treatment (**Table 1.1**). Severe sepsis is a life-threatening emergency condition with 30% mortality rate (Stevenson et al., 2014). Inflammatory response triggered by an infection is amplified by simultaneous cytokine signaling and activation of coagulation and complement cascades. Symptoms are multiple organ failure, thrombosis and hypotension, ultimately leading to septic shock when death can occur by tissue hypoxia (Nguyen et al., 2006). The protease activated protein C (APC) was previously shown (Bernard et al., 2001) to be effective in reducing mortality in patients with severe sepsis due to its cytoprotective effects to inhibit cytokine release and induced cell apoptosis (Mosnier et al., 2007b). However, a higher risk of bleeding was observed as side-effect during APC treatment possibly due to its anticoagulant activity by inactivation of factor Va (**Figure 1.1**) (Rice and Bernard, 2004). Based on long term survival of APC-treated patients, FDA approval was limited to use only when there was high risk of death (Angus

et al., 2004) and ultimately the product was withdrawn voluntarily by the manufacturer after assessing no additional survival benefit to septic shock patients in an independent clinical trial (Ranieri et al., 2012). It is possible that unrelated functions performed by APC in addition to its desired anti-inflammatory protease-activated receptor 1 (PAR-1) signaling could have decreased the efficacy of the treatment. As a first step towards decoupling different pleiotropic functions of APC by engineering specificity, five mutations (Arg229Ala/Arg230Ala/Lys191Ala/Lys192Ala/ Lys193Ala) to decrease the positive charge of the exosite and thereby the activity towards procoagulant factor Va while retaining its anti-inflammatory properties were identified (Mosnier et al., 2007a). An engineered APC variant was able to effectively protect mice subjected to sepsis trigger without increasing the risk of bleeding and thus demonstrated the possibility of selectively abrogating one dimension of a pleiotropic protease (Kerschen et al., 2007).

Thrombin is the effector enzyme of the coagulation cascade and cleaves twelve different protein substrates to auto-regulate its activity and provide positive/negative feedback for controlling blood coagulation (**Figure 1.1**) (Huntington, 2012). In the previous examples highlighted here, specificity was engineered by mutagenesis of exosites which are substrate interacting regions remote from the active site where catalysis happens. In contrast, the specificity of thrombin was engineered by modifying Trp215 and Glu217 residues in the active site. The key determinants of the specificity were identified using alanine scanning of 97 residues constituting the active site and the region around it (Marino et al., 2010). Wild-type thrombin cleaves fibrinogen and PAR-1 substrates (procoagulant) with kinetics 100-fold faster than the rate it displays towards protein C substrate (anticoagulant). While the mutation Trp215Gly is 10-fold more

specific for protein C, a variant containing the mutation Trp215Glu discriminated against protein C for PAR-1 substrate by 1000-fold, thus demonstrating the possibility of tuning the specificity towards protein substrates by modifying the side chain at a single position. The utility of the thrombin double variant, Trp215Ala/Glu217Ala, to treat several disorders including autoimmune encephalomyelitis (Verbout et al., 2015), collagen-induced arthritis (Flick et al., 2011) and ischemic stroke (Berny-Lang et al., 2011) by selective activation of endogenous anticoagulant protein C has been successfully shown in mice models.

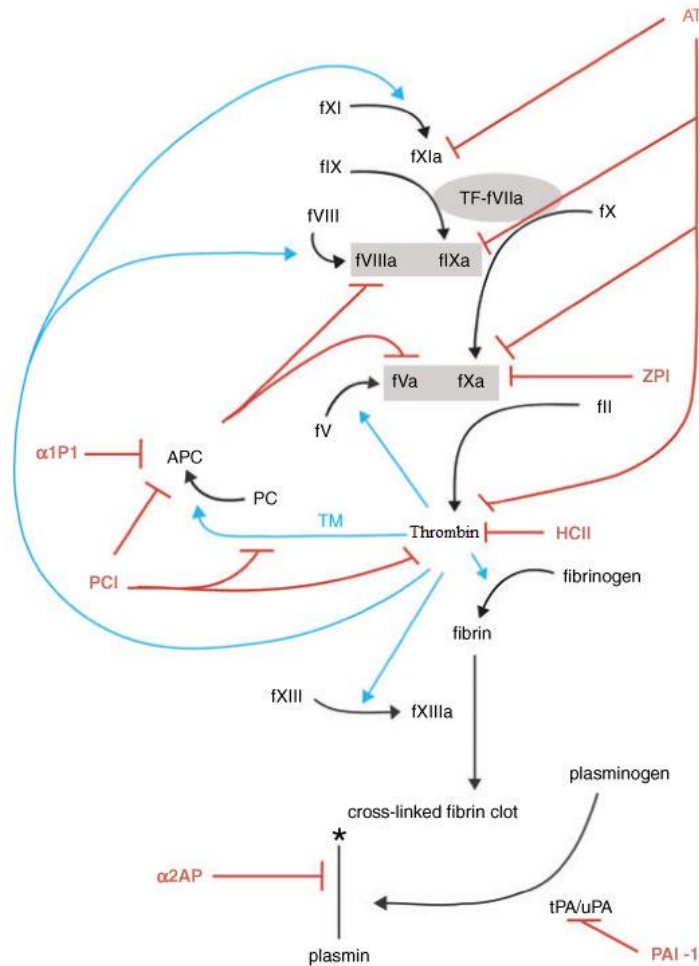


Figure 1.1 Serpin regulation of coagulation, protein C and fibrinolytic pathways. Figure copied from (Huntington, 2012). Reprinted with permission from John Wiley and Sons.

Neprilysin (NEP) is a membrane bound zinc-dependent protease known to be involved in regulating concentrations of signaling peptides including hormones such as glucagon, neurotensin and oxytocin. Several inhibitors for the treatment of hypertension, heart failure (Gu et al., 2010), chronic pain, depression (Roques et al., 2012) and sexual disorders (Wayman et al., 2010) are currently in clinical trials. Also known as CD10 in oncology studies, NEP serves as a known biomarker in several lymphomas and plays a key role in tumorigenesis (Maguer-Satta et al., 2011). The downregulation of NEP expression or decrease in its activity by oxidative damage in the brain has been implicated in the progression of Alzheimer's disease due to decrease in clearance of β -amyloid ($A\beta$) peptides (Madani et al., 2006; Wang et al., 2003). Considering the variety of roles played by the enzyme, it is obvious that a NEP-based therapeutic candidate for Alzheimer's disease should selectively cleave β -amyloid peptides without any residual activity towards other cognate substrates. Towards this objective, a library of NEP single variants with mutations across 134 different positions lining the active site was screened to identify a preliminary list of mutations that reduced the proteolytic activity towards eight different peptide substrates while improving the activity towards $A\beta$ peptide (Webster et al., 2014). For the second round of screening, a library of double mutants with all of them containing Gly714Lys, the most optimal mutation identified from the previous library, and one of the other mutations from the preliminary list was constructed. The lead variant containing the mutations, Gly714Lys and Gly399Val, identified from the second round of screening was further characterized with sixteen different peptide substrates to compare its specificity with wild-type enzyme. Except for three substrates, the variants had activity lower than NEP by at least 10-fold and the

highest decrease of 3200-fold was observed for bradykinin peptide while the kinetics towards the desired A β substrate improved by 20-fold. It remains to be seen if the engineered ‘specialist’ NEP could be used as a therapeutic for the Alzheimer’s disease without any deleterious side-effects due to residual activity towards any essential peptide substrates.

1.2.3 Engineering catalytic activity towards novel substrates

Apart from narrowing down the broad specificity of a native enzyme, protease engineering has also been successfully applied to generate novel specificity which did not exist in the parental scaffold. Celiac disease is a gastrointestinal disorder caused by allergic reaction to dietary gluten and its incidence rate is on the rise affecting 0.3-1% of the global population currently. It involves damage to the digestive tract by the inflammatory response to peptides derived from partial digestion of protein components of gluten and symptoms manifested include diarrhea, malnutrition and edema (Bai et al., 2013; Kaukinen et al., 2014). Oral supplements of different enzymes such as lipases, proteases and glucosidases from plant, animal and microbial origins have been successfully used to treat a variety of digestive disorders such as lactose intolerance and exocrine pancreatic insufficiency (Roxas, 2008). Therefore efficacy of peptidases identified from gluten-degrading microbes and plant seed kernels were tested to treat celiac disease and in fact, a mixture of cysteine protease from barley and prolyl endopeptidase from *Sphingomonas capsulate* is currently studied in a phase 2b clinical trial (Lähdeaho et al., 2014). An ideal therapeutic candidate should adhere to several constraints such as optimal kinetics at acidic intestinal pH of 2-4, resistance to endogenous digestive proteases and specificity towards immunogenic peptides to avoid

competitive inhibition by non-immunogenic peptides and proteolysis of essential gut proteins. With this objective, Kumamolisin-As, a stable bacterial protease from *alicyclobacillus sendaiensis* which prefers PRX motif at P2-P1' positions, was chosen as a starting point to engineer specificity to cleave after PQX motif commonly found in α -gliadin, a monomeric protein component of gluten (Gordon et al., 2012). Based on computational analysis by Rosetta modeling, a focused library of 261 variants with the number of mutations in the active site region varying from one to seven was constructed to decrease the negative charge of the substrate binding pocket, fill up the vacant space that would be generated by the replacement of bulky Arg side chain of the native substrate and provide H-bonds to stabilize the polar amide group of Gln residue in the modified substrate. Based on plate screening with FRET peptide substrate, a set of seven mutations that improves the kinetics towards PQX motif by 100-fold while abolishing the activity towards native PRX motif was identified. The engineered variant retained its kinetics to hydrolyze α 9-gliadin peptide in a digestive-like environment (pH 4, 0.1 mg/mL pepsin) *in vitro*. Thus, the ability to engineer specificity of a protease allows use of a scaffold that already displays desired chemical and thermal stability as a starting point and obviates the need for conventional protein engineering approach of identifying mutations that confer stability. The possibility of clinical use will depend on the ability of the engineered enzyme to cleave immunogenic peptides *in vivo* and its toxicity profile.

1.3 Specificity determinants of proteases are tunable by evolution

1.3.1 Evolution of substrate specificity in natural proteases

In nature, the specificity of proteases evolves according to the needs dictated by the selection pressure. Proteinase inhibitors in plants have driven the mutagenesis of

chymotrypsins in herbivorous larvae and selection of inhibitor-resistant variants (Dunse et al., 2010). The P2 specificity of the egg hatching enzyme, choriolysin L, in killifish and medaka fish correlates with the amino acid sequence of substrate protein in the egg envelope and is not cross-reactive, thus demonstrating the possible protease-substrate co-evolution over 100 million years (Kawaguchi et al., 2013). Therefore, even proteases with high sequence similarity display catalytic activity towards diverse peptide or protein substrates in a highly selective or non-specific fashion depending on their native function. A textbook example used to demonstrate the preference towards chemically distinct substrates in spite of close structural similarity is trypsin – positively charged amino acids, chymotrypsin – hydrophobic amino acids and elastase – small aliphatic amino acids (Fersht, 1999). Enzymes with novel functions are postulated to evolve through multifunctional ‘generalist’ intermediates (Khersonsky and Tawfik, 2010) and subsequent specialization after gene duplication (Hittinger and Carroll, 2007). Granule associated serine proteases of immune defense (GASPIDs) refers to a family of serine proteases involved in various immune-related processes such as phagocytosis of pathogens and apoptosis of infected cells. Based on P1 specificity, GASPIDs could be classified into subfamilies such as tryptases, chymases and Met-ases. Evolutionary analysis of GASPID sequences has suggested divergence of proteases with varying specificity from a broadly specific tryptase, ancestral granzyme K in cartilaginous fish (Ahmad et al., 2014). The hypothesis is also supported by the functional roles of proteases at different hierarchical levels in the phylogenetic tree. For example, granzymes A and K which share high sequence identity with ancestral sequences mediate inflammation against a broad range of infectious microbes unlike evolutionarily young

granzymes B and H directed towards specific viral components for mammalian immunity. Previously (Wouters et al., 2003), an ancestral sequence reconstructed based on the phylogenetic analysis of 56 GASPID sequences was shown to be active towards a broad range of P1 residues (Asp/Arg/Phe/Met) and amenable to mutagenesis in specificity determinant positions (SDP) without loss in catalytic efficiency unlike ‘specialist’ enzymes that evolved later (Caputo et al., 1999).

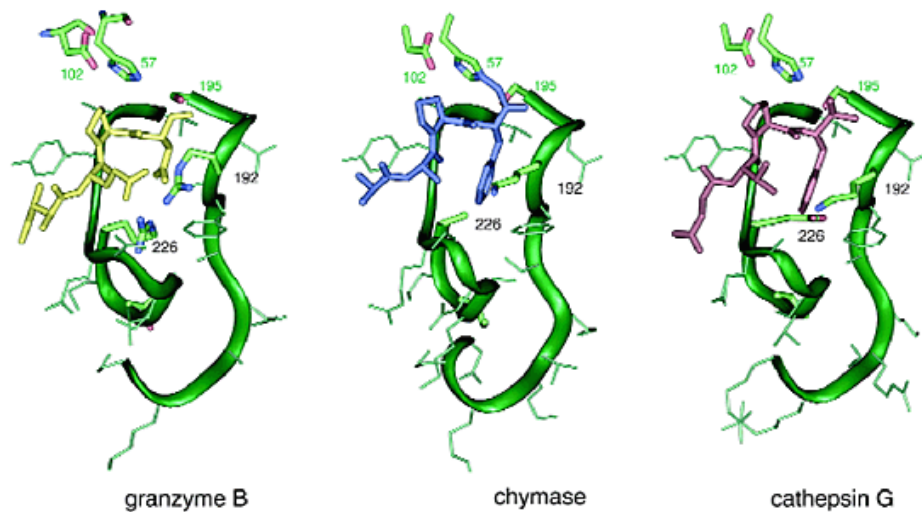


Figure 1.2 Similarity of active site architecture in granzyme B/Asp (yellow), chymase/Phe (blue) and cathepsin G/Lys (pink)). Figure copied from (Waugh et al., 2000). Reprinted with permission from Nature Publishing Group.

Selective preference towards drastically different P1 residues has been acquired even with minimal number of side chain changes in SDP and an imperceptible change in the overall architecture of the active site as exemplified by evolution of granzyme B which prefers Asp/Glu, chymase which prefers Phe/Tyr and Cathepsin G which can cleave after Lys in addition to Leu/His (**Figure 1.2**) in mammals (Waugh et al., 2000). The degree of specialization made possible by evolution allows fine discrimination between chemically similar substrates as shown by guinea pig chymase which cleaves after Leu but not Tyr/Phe or Val/Ala unlike other mammalian chymases (Caughey et al., 2008). Hence *in*

vitro evolution under a selective pressure to generate variants with desired specificity appears feasible.

1.3.2 Engineering proteases by grafting specificity determinants

While nature can take millions of generations to identify the mutations necessary for modifying the substrate specificity of a protease by hit and trial, protein engineers are expected to achieve the same feat in a much accelerated pace. One of the strategies employed was transferring specificity determinant positions (SDPs) from one enzyme to another as demonstrated by the successful engineering of chymotrypsin-like specificity in trypsin (Hedstrom et al., 1992). This approach is however challenging as all the residues that contribute towards substrate recognition needs to be identified. The prediction of SDPs based on analysis of large volume of phylogenetic and structural information using computational tools has been reviewed recently (Chakraborty and Chakrabarti, 2015). It was inferred that their success in identifying catalytic residues of unknown proteases could not be replicated in SDP prediction due to lack of experimental mutagenesis data for the design of machine learning algorithms and validation of predicted SDPs. The amount of experimental data required for SDP prediction can be explained using the following example. In a previous study, apparent kinetics of eight matrix metalloproteinases (MMPs), representatives from three different phylogenetic subfamilies, towards 1369 different peptides selected from a random hexamer library displayed on phage was measured (Ratnikov et al., 2014). Using this large set of experimental data, pairwise correlation between substrate specificity of any two MMPs was determined. Functional phylogeny determined based on substrate specificity was shown to correlate with sequence identity of fifty-seven positions and were identified as

SDPs (**Figure 1.3**). The correlation between sequence and functional phylogeny decreased when any of the SDPs were omitted from the analysis showing that all of the residues identified participate in defining the substrate specificity. By interchanging only 4 residues in the SDPs spatially close to the substrate, preference of MMP-16 was transmutated to that of MMP-17. It is possible that remaining residues assigned as SDPs contribute to fine-tuning of the specificity as hybrid MMP-16/17 did not cleave 12 out of 82 substrates preferred by MMP-17.

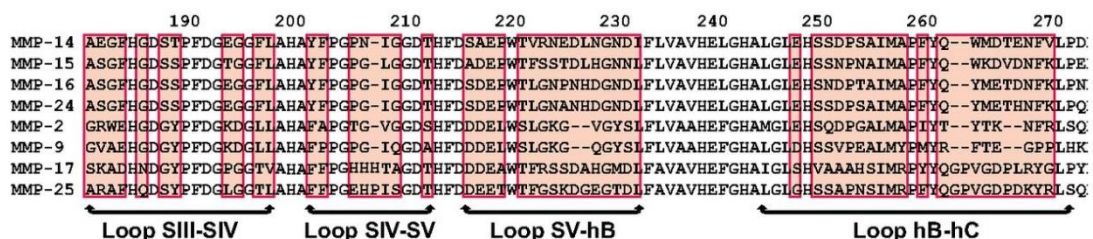


Figure 1.3 Sequence alignment of catalytic domains of human matrix metalloproteinases. Specificity determinants located in the variable surface loops are highlighted in pink. Figure copied from (Ratnikov et al., 2014).

It is to be noted that all the residues predicted as specificity determinants in MMPs were present in surface loops. Role of surface loops in evolutionary divergence of substrate specificity in chymotrypsin-fold protease has previously been described (Perona and Craik, 1997). It is tempting to generalize and claim that SDPs are typically present in motifs such as surface loops that can withstand high mutational load so that specificity can be tuned as per the requirements of evolutionary pressure without affecting the general scaffold that allows functional proteolytic activity. There are several intrinsic constraints for a functional protease scaffold such as proper stabilization of the transition state over ground state for fast turnover, right orientation of the base depending on the stereochemistry of catalysis and tighter packing with low helical content and high C_{α} densities to avoid autolysis (Buller and Townsend, 2013; Stawiski et al., 2000). Though

23 different catalytic triads/dyads have been observed in nature (Gherardini et al., 2007), it would be more efficient to allow divergence of substrate specificity on the same scaffold without redesigning the wheel. Supporting the hypothesis, proteases were shown to have higher loop content than non-proteases and it was used as one of the structural features to predict putative proteases from genomic databases (Stawiski et al., 2000).

1.3.3 Success of directed evolution approach in engineering specificity

Though substrate binding pocket appears to be plastic, it is not straightforward or easy to engineer novel specificity as any incompatible modification in SDPs will affect the amidase activity of the enzyme (Hung and Hedstrom, 1998). In several cases, gain of function mutations in the active site destabilize the protein fold and require introduction of functionally neutral mutations that improve protein stability (Bommarius et al., 2011). It is to be noted that the effect of these mutations need not be additive but can be epistatic. SDPs determine the substrate binding affinity but it has been suggested that flexible motion by distal loops drives the chemical catalysis and mutations in these regions remote from the active-site influence the turnover number (Currin et al., 2014). Thus, identifying all the subsites interacting with the substrate and grafting them onto corresponding positions in a different protease may not necessarily yield a functional enzyme with engineered specificity.

High-throughput screening of a diverse library spanning large evolutionary breadth is a powerful approach to identify potential ‘hits’ as it increases the chances of identifying one of the right combinations of epistatic mutations. Outer membrane protease T (OmpT) which prefers to cleave between dibasic residues was engineered to cleave after peptides containing acidic Glu residue in P1 position (Varadarajan et al.,

2008). The necessary mutations were identified by screening a library of million variants which was constructed to target three residues in the active site by saturation mutagenesis. While the variant (ER-OmpT) containing triple mutations (Glu27Leu, Asp208Arg and Ser223Gly) showed high catalytic efficiency ($k_{\text{cat}}/K_M = 10^5 \text{ M}^{-1}\text{s}^{-1}$) towards the new substrate and 1000-fold discrimination against the wild-type substrate, variants containing single mutations did any show any activity towards peptides containing Glu residue thus showing the need for epistatic mutations to reverse the polarity of the active site.

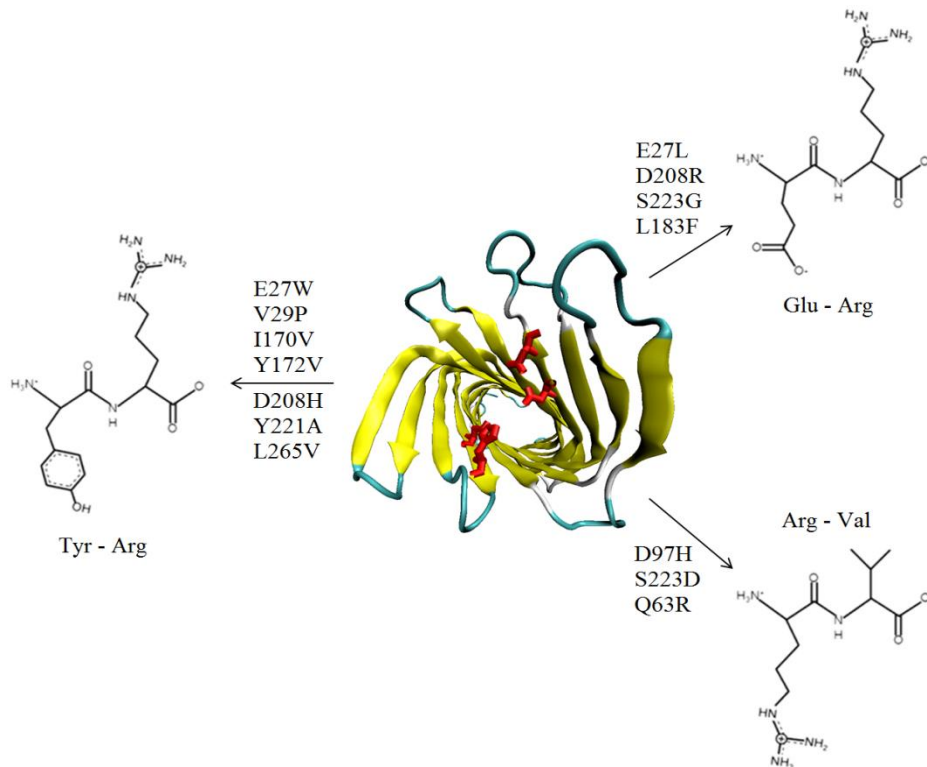


Figure 1.4 Engineering of specificity towards different residues at P1 and P1' positions from a single parental enzyme, outer membrane protease T (OmpT).

The engineering effort was further extended to alter P1' specificity from Arg to Ala by screening 100 million variants of ER-OmpT with random mutations introduced by error-prone PCR. A more comprehensive screening of 200 million variants with

mutations distributed across 21 residues lining the active site was required to identify a variant which can cleave after hydrophobic Tyr at P1 position. Starting from a single parental enzyme with preference towards basic residues, directed evolution approach had allowed successful engineering of selectivity towards acidic Glu, hydrophobic Tyr and tyrosine containing post-translational modifications such as sulfation and nitration without loss in catalytic efficiency (**Figure 1.4**). One of the key reasons behind the success achieved in a relatively short time scale was the high degree of genetic diversity screened, as facilitated by the screening throughput, with a resolution to simultaneously cull broadly specific variants and enrich kinetics of desired activity.

1.4 Tools of the trade: Advances in technologies for engineering specificity

Initial success in protein engineering by laboratory evolution approach has been consolidated by the advances in technologies allowing the scientific community to apply the same principles more efficiently. Newer methods for library construction and understanding of the fitness landscape observed in diverse protein libraries serve as a guide to navigate the sequence space efficiently to identify useful enzymes more often (Currin et al., 2014). The costs and time associated with custom DNA synthesis and DNA sequencing continue to decrease drastically. While gene synthesis was priced at \$12 per base pair 15 years ago, today we can purchase a 500 bp gene fragment for less than \$100 in a turnaround time of 3-5 days (Mueller et al., 2009). Therefore in the context of technologies available, protease engineering is riding the wave of next-generation biocatalysts pursued by both academic and industrial interests (Bornscheuer et al., 2012).

1.4.1 Computational tools for constructing targeted libraries

1.4.1.1 Homology guided mutagenesis

The number of protein sequences classified as peptidases in MEROPS database is 413,834 and it continues to grow with accelerating inputs from genome sequencing and metagenomics studies (Rawlings et al., 2014). Even to non-bioinformatics scientists, tools to mine the sequence database are becoming increasingly available for making an informed choice of residues to be targeted instead of relying on random mutagenesis techniques for creating the genetic diversity. The basic tenets in using sequence homology information for modifying specificity of enzymes are: (i) residues conserved in a superfamily of enzymes with varying specificity should not be mutated as they are potentially required for catalysis (ii) residues that are not conserved in the superfamily but conserved across the subfamily with similar specificity should be targeted as they could contribute towards substrate recognition. By restricting the codon choice at selected sites based on multiple sequence alignment of 1750 α/β hydrolases, the size of the library was decreased by 100-fold and a variant with 240-fold enhancement in specific activity towards the new substrate was identified (Jochens and Bornscheuer, 2010). Peptidase databases can be searched to identify an optimal starting point for the protein engineering studies based on the desired characteristics expected of the end product. For example, to generate an oral enzyme therapeutic candidate for celiac disease, a preliminary list of enzymes known to be active in acidic pH 2 – 4 were identified from MEROPS database (Gordon et al., 2012). In addition to desired pH stability, kumamolisin-As also had preference for Pro residue in P2 position and was chosen as the starting point to engineer specificity towards Pro-Gln motif commonly present in immunogenic gluten-derived

peptides. It is even possible to identify an enzyme with the desired specificity only by mining databases without any library screening as demonstrated by *in-silico* identification of transaminases with desired (R)-selectivity (Höhne et al., 2010). Based on the crystal structures available for (S)-selective transaminases and L-branched chain transaminases, key motifs that are expected to be present and mutations that should not be present for the desired (R)-selectivity were identified to be used as the database search criteria. Of the 24 enzymes selected as putative (R)-selective transaminases for experimental verification, 17 of them showed desired enantioselectivity towards a variety of amine substrates tested. It remains to be seen whether the same strategy could be extended to searching protease sequence database using motifs predicted for any desired substrate specificity.

1.4.1.2 Structure and sequence guided mutagenesis

Though not as populated as the sequence database, database of enzyme structures contains a wealth of information for targeted library construction. For example, analysis using the CASTp server can identify the residues constituting the putative binding pockets and cavities present in the enzyme by taking only the corresponding PDB id as input (Dundas et al., 2006). In the absence of structural information for the desired enzyme, most of the analysis tools can also use homology-based structure models as inputs which can be easily constructed using online servers such as PHYRE2 provided an experimental structure of a related protein which shares at least 30% sequence identity is available (Kelley et al., 2015). Another useful tool that needs to be mentioned in this context is the Hotspot wizard (Pavelka et al., 2009) which integrates several bioinformatics tools developed to collect both evolutionary and structural information corresponding to a given input and generates a list of candidate residues for library

construction identified based on spatial proximity to the substrate and the degree of mutability depending on how conserved the residue is (**Figure 1.5**). This computational tool has been successfully used in redesigning the specificity of a variety of enzymes including deaminases and dehalogenases (Kohila et al., 2012; Pavlova et al., 2009). With regard to peptidases, it was used to identify residues in methionine aminopeptidases (MetAPs) that are responsible for fine differences in specificities between variants from three different species, *Mycobacterium tuberculosis*, *Enterococcus faecalis* and *Homo sapiens* (Kishor et al., 2013). Three hotspots identified were targeted to design drug candidates that can selectively inhibit bacterial proteases without affecting the human analogue, thereby decreasing toxicity associated with the inhibitor treatment. A comprehensive list of online tools and web servers available for protein engineering to harness sequence and structural information are available elsewhere (Sebestova et al., 2014; Verma et al., 2012).



Figure 1.5 Graphical user interface of HotSpot Wizard results. Figure copied from reference (Sebestova et al., 2014) with permission from Springer.

1.4.1.3 *In silico* mutagenesis based on energy minimization

One of the approaches that have been used in engineering specificity is to estimate the change in substrate binding free energy for a given mutation in the active site. The specificity of HIV-1 protease which cleaves after three different domains in HIV-1 Gag polyprotein was altered to selectively cleave reverse transcriptase domain. Three mutations beneficial for substrate discrimination were identified after sampling the effect of amino acid changes in eight different subsites (S4-S4') on binding free energy (Alvizo et al., 2012). Similarly α -lytic protease was engineered to identify a variant with 1000-fold improvement in P1 activity towards Leu (Wilson et al., 1991). In experimental studies, obtaining sufficient quantity of enzyme variants with high mutational load for characterization could be challenging due to various reasons such as aggregation induced by misfolding. *In silico* screening approaches can sample mutations at several positions without any such limitations and generate a small set of potentially beneficial mutations which can be experimentally validated. Typically during sampling, mutations are introduced in a sequential fashion so that energy minimization and repacking can be performed using tools such as FoldX (Schymkowitz et al., 2005) and FASTER (Desmet et al., 2002) in the intermediate steps.

It is somewhat surprising that active site mutations chosen based only on substrate binding energy without regard to stabilization of reaction intermediate can lead to successful engineering of protease specificity as strong binders can also act as inhibitors (Baggio et al., 1996). HIV-1 protease, though specifically cleaves at particular sites in its endogenous substrates, displays no consensus in the primary sequence of its cleavage sites. Variants of α -lytic protease though highly active towards the novel substrate were

shown to be broadly specific (Bone et al., 1991). It is possible that the overlap between substrate recognition and catalysis is minimal in the scaffolds which do not recognize specific consensus sequence and mutations identified based on binding free energy are sufficient to introduce the activity towards novel substrates. General efficacy of this approach in engineering specificity will be known only when applied to other classes of proteases.

Most of the computational tools which study the effect of point mutations are limited in their ability to include interactions due to water molecules or other cofactors and dynamic interactions by the flexible loop regions of the enzyme, both known to play a significant role in enzyme catalysis (Ma et al., 2005). Therefore assessing whether a given mutation would introduce any steric clashes or affect the stability of the protein fold has been more straightforward than to estimate its effect on substrate binding and catalysis. However it is possible to overcome these limitations by using them in junction with more computationally rigorous tools such as Rosetta molecular modeling suite to include backbone flexibility or impose geometric constraints corresponding to a given catalytic mechanism. For example, after mutations that affect the substrate binding energy were identified using computational alanine scanning (Lise et al., 2009), molecular dynamics simulations were performed to assess the stability of granzyme B variant containing the mutation Arg201Lys bound to Serpin B9 in presence of water molecules (Losasso et al., 2012).

1.4.2 Platforms for high-throughput characterization

More often than not, computational enzyme design efforts are followed by directed evolution experiments since the kinetics and stability of the endpoint variants do not meet

the final requirements of the application. However, dearth of assays for determining protease specificity in high-throughput limited the span of genetic diversity explored. Even with the advent of robotics for liquid handling or colony picking, the maximum size of the libraries screened using agar/ microwell plates was restricted to 10^4 . Here, we will focus on newer platforms and strategies developed for high-throughput screening (HTS) of large libraries ($> 10^7$ variants).

1.4.2.1 Platforms based on fluorescence activated cell sorting

Before we proceed further, a brief introduction to the workflow of fluorescence activated cell sorting (FACS) would be helpful. Dilute suspension of cells is introduced from the sample line into sheath fluid in laminar flow such that cells are aligned one after another in a single line to pass through a nozzle of size less than 100 μm . Laser and detection optics are aligned below the nozzle to measure fluorescence and scatter properties at single cell resolution. The nozzle is oscillated to generate droplets at a desired piezo frequency which is at least 3 or 4-fold higher than the number of cells passing through the nozzle per second. Number of cells per droplet can be determined by Poisson distribution and 30% of the droplets are expected to contain only one cell. All droplets are directed towards the waste reservoir but when a droplet is known to contain a desired cell, it can be deflected by electrostatic or acoustic methods for collection into a separate tube. Further information on working mechanism of FACS can be found elsewhere (Davies, 2007). Though FACS has been around since 1969, its use was restricted to mammalian cells. With the advances in optics and fluorescent dyes suitable for bacterial analysis, cell sorters capable of screening 10^4 variants per second were made commercially available. A whole-cell sensor was designed where a GFP variant with its

C-terminal fused to a quenching peptide through a protease sensitive linker was co-expressed with the protease in the cytoplasm of *E.coli* (Lindenburg and Merkx, 2014). If a protease variant is active towards the linker sequence, then it would lead to release of quenching peptide and an increase in GFP fluorescence. When screening a library of protease variants using FACS, only the cells showing high GFP fluorescence can be isolated to identify the mutations corresponding to active protease variants. The strategy of quenching the fluorescence by C-terminal fusion to a hydrophobic quencher peptide was shown effective in yellow, cyan and red fluorescent proteins as well to allow a multiplexed assay (Wu et al., 2013). It is to be noted that translation of this platform to a new protease is possible only if the amino acids that remain in the C-terminal of the fluorescent protein after proteolysis do not interfere with the subsequent chromophore formation and optimization of expression levels of the protease and the substrate might be required to improve the signal/noise.

As opposed to substrates that are genetically encoded, exogenously added substrates allow a higher range of concentration to screen the protease library. Also, substrates synthesized in vitro can be modified to contain non-natural amino acids, post-translational modifications absent in the host and non-natural dyes which exhibit Forster resonance energy transfer (FRET). FRET is the phenomenon of radiationless transfer of energy from an excited fluorophore to a quencher molecule in close proximity (< 10 nm) leading to a decrease in emission of the excited fluorophore. When a pair of dyes known to exhibit FRET is conjugated to a peptide such that the amino acid sequence sandwiched between the dyes corresponds to the specificity of the desired protease, then any proteolysis event will disrupt the spatial proximity of the quencher to the fluorophore and

can be measured by the corresponding increase in fluorescence. In a previous study (Olsen et al., 2000), a peptide conjugated to a FRET pair was designed to contain an additional positively charged tail (amino acids Gly-Arg-Gly-Arg) such that after proteolysis, fragment containing the fluorophore is locally captured on the surface of the bacterial cell which is typically negatively charged (zeta potential ~ -30 mV). Therefore, when a population of cells expressing a protease library was incubated with the peptide substrate, only those cells which contained variants catalytically active towards the peptide substrate became fluorescent while rest of the population did not. Moreover, the strategy was extended to measuring proteolytic activity towards three different substrates simultaneously using spectrally different fluorophores allowing isolation of variants that are highly specific towards only one of the substrates FACS (Varadarajan et al., 2009b). Permeability of outer membrane of bacteria is typically restricted to hydrophilic molecules of size than 600 Da (Nikaido and Vaara, 1985). As peptide-dye conjugates are larger (2.5 -3 kDa) than this size limit, the platform has been used only for native membrane proteases or recombinant proteases displayed on the surface (Ramesh et al., 2012).

Invention of α -agglutinin based platform for display of recombinant protein on the surface of yeast spurred the use of eukaryotic host in protein engineering (Boder and Wittrup, 1997). Unlike prokaryotes, presence of organelles such as endoplasmic reticulum (ER) and Golgi apparatus allow proper expression of recombinant proteins containing multiple disulfide bonds or glycosylation modifications. Recombinant proteins can be targeted towards ER in yeast using an N-terminal signal peptide and can be retained there using a C-terminal peptide. In a previous study, a substrate was designed to

contain ER signaling and retention peptide, a protease sensitive linker, a-agglutinin domains for display of the substrate on the surface and epitope tags for detection using fluorophore-conjugated antibodies (Yi et al., 2013). In the absence of any proteolysis, the substrate is sequestered in ER and no fluorescence is detected upon antibody labeling. Proteolysis towards multiple sequences can be monitored by flanking additional epitope tags. The platform was shown to be amenable to detect activities of hepatitis C virus protease, tobacco etch virus (TEV) protease and human granzyme K. In another study using yeast, a transcription factor was anchored by fusion to the intracellular domain of a membrane protein through a protease-sensitive linker thereby restricting transcription of the reporter gene in nucleus in the absence of any proteolytic activity cleaving the linker (Kim et al., 2000). Using this platform, P2 specificity of hepatitis A virus 3C protease was altered from Thr to Gln by screening a library containing one million variants (Sellamuthu et al., 2008). Depending on the nature of the reporter gene, selection or screening can be performed to isolate variants that exhibit proteolytic activity towards the linker sequence. In spite of the ease in culturing and the increasing knowledge for genetic manipulation, yeast is not preferred for screening very large libraries due to its lower transformation efficiency ($10^6/\mu\text{g}$ of plasmid) even with the improved LiAc/PEG method in comparison to typical transformation efficiency of $10^{10}/\mu\text{g}$ of plasmid achieved with *E.coli* (Gietz and Schiestl, 2007).

1.4.2.2 Microfluidic devices for cell sorting

Recently, advances in microfluidics and *in vitro* compartmentalization (IVC) methods have fueled the efforts to design next generation sorters with features that are lacking in the commercially available FACS machines (van Vliet et al., 2015). In a

previous study, enzyme libraries were sorted at single cell resolution similar to FACS without the need for periplasmic or surface display techniques. This was achieved by encapsulation of cells expressing variants of arylsulfatase in droplets containing a mixture of lysis reagent and fluorogenic substrate such that each droplet contains a variant that is released from the cytoplasm by cell lysis (**Figure 1.6**) (Kintses et al., 2012). This technology is especially relevant to protease engineering as it allows expression of the enzyme in its zymogen form in the cytoplasm which can be subsequently activated for labeling with substrates. Another key advantage of the microfluidics-based sorting methods is the exquisite control on the fate of the drops containing enzyme variants. As each variant is labeled separately and the mass exchange between droplets is negligible due to separation by the oil phase, droplets could be incubated with the substrate off the flow path for a longer time period when the kinetics of library members is expected to be slow. Though fluorescence activated droplet sorters have lot of potential, they are not used prevalently as most of them are custom built and require considerable expertise in microfluidics. In fact, there has been only one study that used IVC technique to engineer proteases where the activity of subtilisin in presence of the inhibitor antipain was improved (Tu et al., 2011a). Apart from flow sorting, microfluidic chips which are miniaturized versions of microtiter plates but containing 1000-fold more number of wells have been used for kinetic measurement of proteolytic activity observed in mammalian cells at single cell resolution by fluorescence microscopy (Wu et al., 2015). As the dimensions of nanowells can be easily tuned, we can anticipate adaptation of these imaging based methods for screening protease libraries. However, the

advantage of high-content screening and easy retrieval of desired variants is offset by two orders of magnitude lower throughput in comparison to flow sorting.

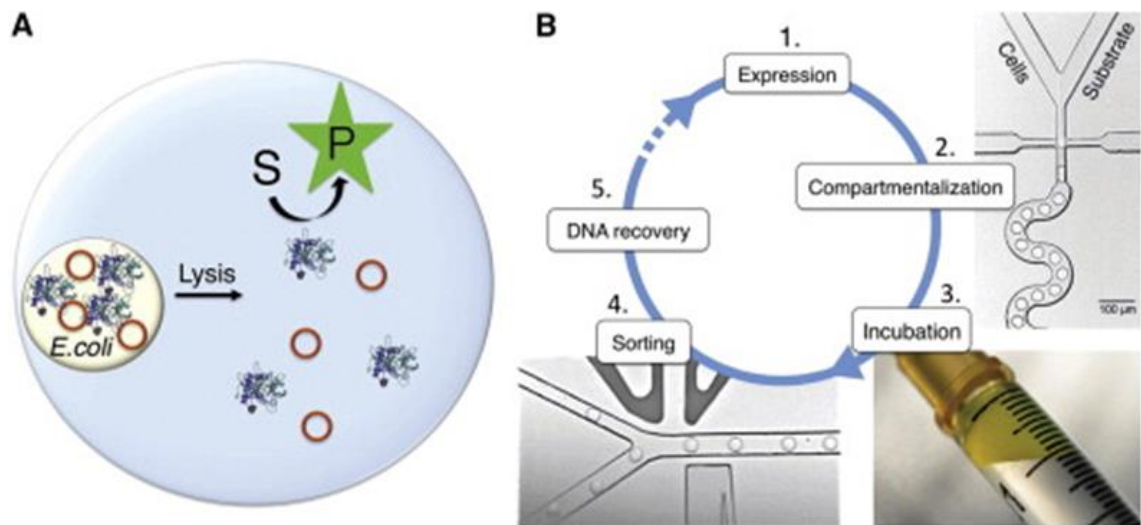


Figure 1.6 Schematic directed evolution cycles by lysate screening in droplets. Figure copied from (Kintses et al., 2012).

1.4.2.3 *In vivo* continuous evolution

In all the previous platforms described so far, protease libraries were screened in discrete steps for iterative enrichment of desired variants. Alternatively, continuous enrichment of desired mutations was achieved in a system where enzyme activity of the protease variant displayed by a phage determines its propagation by infection of *E.coli* cells. With an objective of identifying mutations in hepatitis C virus protease gene that lead to resistance towards inhibitor drugs, a reactor of *E.coli* cells was infected with phage displaying the viral protease and maintained at a constant cell density by influx of fresh cells and outflux of phage-infected cells. Genetic diversity of the viral protease was created by inducing overexpression of mutator proteins such as DNA polymerase without the proofreading subunit in *E.coli*. In addition, *E.coli* cells express an inactive form of T7 RNA polymerase which can be converted to an active form by the viral protease upon

phage infection to upregulate the expression of protein III under T7 promoter. As protein III is critical for viral replication, only phages that display protease variants active in presence of the inhibitor propagate while others cannot due to inability to activate T7 RNA polymerase. Therefore, phages displaying active protease continue to survive and enrich in the reactor by infecting fresh cells while phages that do not get diluted by the constant outflux. High-throughput sequencing of phages isolated from the reactor after 28 h of continuous evolution, identified mutations that have been observed in inhibitor drug-resistant hepatitis C patients in the clinic. Fast life cycle of phage (progeny release in 10 min) allows enrichment of mutations through several generations in a short duration of time. Considering the ease of setting up a phage-assisted continuous evolution experiment, this system can also be used in fine-tuning the properties of potential ‘winners’ identified using other methods. It should be interesting to see how PACE can be translated for engineering specificity of other proteases as phage display has already been demonstrated for enzymes such as enterokinase and trypsin previously (Gasparian et al., 2013; R. Corey et al., 1993). Unlike the hepatitis C virus protease studied previously, when PACE is applied for proteases such as chymotrypsin which does not display any extended specificity, enrichment of phage containing active variants can be affected by the effect of protease toxicity on host viability.

1.4.3 Techniques for profiling specificity of engineered proteases

1.4.3.1 Selection/ screening of genetically encoded substrate libraries

Profiling the activity of potential ‘winners’ isolated from protease libraries towards substrates with different P_n - P_n' residues is essential for validating the successful engineering of specificity. Unlike chemically synthesized peptide-dye conjugates,

genetically encoded substrates provide a facile way to increase the diversity of the amino acid sequences tested. Randomized hexameric peptide libraries displayed on phage have been successfully used to determine substrate preferences of several poorly characterized proteases (Clara et al., 2011; Shi, 1995). In a typical experiment, the protease of interest is incubated with phages displaying randomized peptide – epitope tag fusion such that phage particles displaying substrates that did not undergo proteolysis can be depleted using antibody against the epitope tag. Phage variants which have lost the epitope tag due to proteolysis are propagated and the cycle of negative selection is repeated multiple times. Amino acid sequence of optimal substrates can be identified by DNA sequencing of enriched variants. Apart from phage display, another platform that has been used for profiling specificity of proteases is cellular libraries of peptide substrates (CLiPS) (Boulware and Daugherty, 2006). Peptide library fused to a streptavidin-binding motif is displayed on the surface of *E.coli* by fusion to a membrane protein and labeled with streptavidin-fluorophore conjugate. Homogenously fluorescent population obtained by FACS is incubated with the protease of interest and sorted again to isolate non-fluorescent cells that have lost the streptavidin-binding motif by proteolysis. Unlike selection by phage panning, CLiPS screening allows a higher degree of control on members that are carried forward for subsequent rounds of enrichment albeit with at least 10-fold decrease in peptide diversity screened as limited by bacterial transformation efficiency. Some of the platforms described earlier which use genetically encoded substrates to detect protease activity can be modified such that positions in the protease-sensitive linker region of the substrate are randomized and substrates optimal for a given protease can be enriched in a way identical to screening of protease libraries.

1.4.3.2 Mass spectrometry based approaches

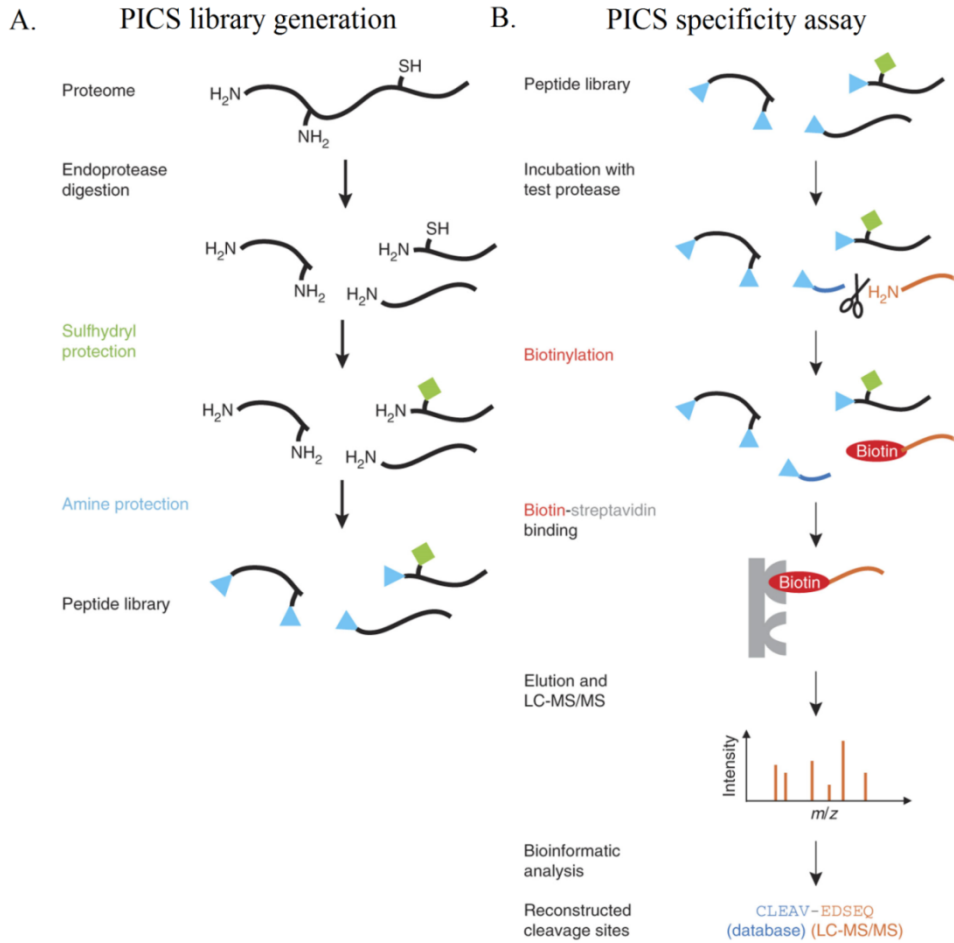


Figure 1.7 Proteomic identification of cleavage sites (PICS) library generation and specificity assay. Figure copied from reference (Schilling et al., 2011b) with the permission from Nature Publishing group

Though large sequence space can be comprehensively screened to identify the amino acid sequence of the optimal substrate, determining the exact cleavage site in the peptide requires subsequent characterization by N-terminal sequencing or mass spectrometry. Typical workflow of a proteomics based platform that utilizes tandem mass spectrometry (MS/MS) and bioinformatics to determine all the sites cleaved by a given protease is shown in **Figure 1.7** (Schilling et al., 2011b). Peptide libraries are obtained from lysates of mammalian or bacterial cells and are expected to contain a variety of post-translational

modifications, thus increasing the chemical repertoire of constituent amino acids for profiling specificity of the protease. For example, a metalloprotease isolated from *Methanosarcina acetivorans* was characterized to cleave at the N-terminal of methylated or dimethylated lysines in addition to free lysines (Huesgen et al., 2015). All free amine groups are chemically protected during library preparation so that any new N-terminus generated subsequently can be attributed to proteolysis catalyzed by the protease of interest. Peptides with a free N-terminus isolated by selective conjugation to succinimidyl ester of biotin are analyzed by MS/MS coupled to liquid chromatography and the spectra obtained are assigned to amino acid sequences using database search software such as MASCOT. P1` sequence corresponding to each peptide identified can be inferred from the protein database corresponding to the cell line used for the peptide library generation using the freely available online tool, WebPICS (Schilling et al., 2011a). Overall, the system has been shown applicable to all classes of proteases and subsite cooperativity analysis can be performed as hundreds of unique sequences are identified for each protease. In the case of HIV-1 protease, 767 unique sequences were identified to determine P6-P6` specificity and presence of Ala in P1 position was shown to correlate with P3 Leu residue (Schilling and Overall, 2008). Though utility of the original proteomic identification of cleavage sites (PICS) method is restricted to endoproteases, a modified version was developed to profile the specificity of human metallocarboxypeptidases A1, A2 and A4 (Tanco et al., 2013).

1.4.3.3 Quantitative characterization of specificity

The standard method of kinetic characterization to obtain Michaelis-Menten parameters uses colorimetric or fluorogenic peptide substrates where the dye is quenched

either by conjugation to C-terminal of an amino acid or by a separate quencher dye. Combinatorial peptide library substrates can be purchased commercially and the rate of fluorescence generated under varying protease/substrate concentrations can be conveniently measured in a plate reader. Sometimes, higher substrate concentrations are to be avoided due to poor solubility of hydrophobic dyes in aqueous buffer solutions and the calibration procedure to convert absorbance or fluorescence to molarity requires inner filter effect correction. In a previous study, to avoid the hassle of chemical synthesis for each peptide substrate, phage display coupled with enzyme-linked immunosorbent assay (ELISA) was used to measure the apparent proteolysis kinetics of eight matrix metalloproteinases towards 1370 different substrates (Ratnikov et al., 2014). Though the kinetic parameter obtained from phage ELISA can be used for a quantitative comparison of substrate preferences of proteases, they are not identical to Michaelis-Menten parameters due to several reasons. Substrate concentration corresponding to the typically used value of 10^{11} phage particles/well is in picomolar whereas the Michaelis constant K_m for most of the proteases is in micromolar. In fact, nanomolar protease concentration is required to generate sufficient chemiluminescence for subsequent detection and also substrate molecules are surface-bound, not in free solution. Quantitation of product generated can be performed without relying on any chromophore or fluorophore using mass spectrometry by using tandem mass tags or internal peptide standards with identical amino acid sequence labeled with heavy isotope (Bantscheff et al., 2007). Majority of the protease characterization efforts utilize peptide substrates and in some cases, they do not represent the ability of the protease to cleave corresponding protein substrates possibly due to the lack of ability to recognize a different secondary structure associated with the

same cleavage site motif. It is possible that some of the techniques used in terminomics (studies focused on proteins with C- or N- termini generated post-translationally) which involves identifying physiological substrates of proteases can be adapted for profiling the specificity of engineered proteases with protein substrates instead of peptides.

1.5 An outlook for the future of engineering protease specificity

Recombinant protein production allowed industrial use of natural proteases other than digestive enzymes isolated from slaughtered cattle. Protease engineering confers desired properties absent in natural proteases to meet the existing industrial needs and also enables emergence of new applications such as enzymatic synthesis of small molecule drugs. The key thrust area for proteases with modified specificity is in development of new therapeutic candidates for clearance of pathological target proteins. One of the major requirements of any therapeutic protease is localized activity towards specific targets as unintended proteolysis can lead to fatal side effects such as intracranial hemorrhage in the case of thrombolytic therapy for heart attack. In nature, apart from regulation by inhibitors and cofactors, localization of proteolytic activity is also achieved by non-catalytic domains present in the proteases. For example, kringle domains present in anticoagulant proteases bind selectively to fibrin, thus localizing the fibrinolytic activity to the clot site. Presence of fibrin binding domains in tissue-type plasminogen activator has allowed its administration in a systemic infusion method as opposed to urokinase which has to be locally administered at the clot site (Craik et al., 2011). It has been shown that regulatory proteases have evolved from an ancestral trypsin-like protease that contained only the signal peptide, zymogen activation peptide and catalytic domain by accretion of binding modules through various mechanisms such as exon

shuffling and transposition (Patthy, 1985). As the binding domains are independent units of the protease, it is plausible that they can be engineered for higher affinity towards a target without affecting the catalytic activity. Variants of kringle domains have been identified to bind target proteins such as tumour necrosis factor (TNF) and TNF-related apoptosis-inducing ligand receptors with a dissociation constant of 10 nM (Lee et al., 2010). Granzyme B, a pro-apoptotic protease that does not contain any binding domain, was fused to an antibody fragment targeted against a tumour antigen and shown to be effective in different lymphoma mouse model studies (Schiffer et al., 2014, 2013). Though this strategy can decrease the required dosing by increasing the local concentration of the protease in the region containing target molecules, it is to be noted that the protease is active all along even before the target binding event. Off-target cleavages can be further minimized if extended substrate recognition motifs can be grafted within the catalytic domain. IgG degrading enzyme from *Streptococcus pyogenes* (IdeS) specifically cleaves the lower hinge region of IgGs and does not react towards analogous peptide substrates or denatured proteins containing the putative recognition site, Glu - Leu - Leu - Gly (Vincent et al., 2004). This exquisite specificity is attributed to an exosite that interacts with CH₂ domain of the Fc region. We could envision next-generation proteases with grafted exosites to carry out proteolysis of specific target substrates present in complex protein mixtures with precision. As a first step in this direction, backbone grafting of binding motifs, even those discontinuous in the primary sequence, onto different scaffolds was successfully performed using a computation-guided directed evolution approach (Azoitei et al., 2011). Though extension of this strategy to grafting in protease scaffolds without loss in catalytic efficiency will be

formidable, the real challenge in engineering exosites for selectivity towards a target will be identification of all the relevant specificity determining positions (SDPs) in the scaffold.

The contribution of protease engineering towards industrial biotechnology is evident from the \$2.7 billion sales market in 2014 which is pegged to grow further in the future (86). Efforts are focused on improving the chemical and thermal stability of proteases such that they are active in the non-native working environment. Most of the industrially useful proteases have broad specificity and engineering their specificity is not relevant to improving their utility. To the best of our knowledge, currently no protease with modified specificity is used in any biotechnological application. However it is very possible that in the near future engineered proteases along with innovations from other technologies can revolutionize multiple fields such as protein sequencing. One of the factors behind renaissance in the field of DNA sequencing last decade was DNA polymerases engineered for high fidelity and ability to incorporate chemically modified nucleotides (Murakami and Trakselis, 2013). Though amino acid sequence of a protein can be inferred from its cDNA, direct sequencing is required to show evidence at the protein level and for identification of post-translational modifications such as proteolytic processing. Peptide sequencing using the classic Edman degradation method which releases amino acids one at a time from the N-terminus using phenylisothiocyanate chemistry can be modified using engineered proteases. A cysteine protease was redesigned to release the derivatized N-terminal amino acid without the need for a strong acid or heat which typically causes high background by additional hydrolysis elsewhere in the peptide (Borgo and Havranek, 2015). Currently, samples with modified N-termini

are treated with aminopeptidases for subsequent Edman sequencing. It is possible to imagine a microfluidics-based platform that uses different aminopeptidases with stringent P1 selectivity to identify the entire sequence of a protein without any restriction on the size that could be analyzed. One of the major limitations of Edman sequencing that led to the rise of mass spectrometry (MS) as the gold standard for protein sequencing is the inability to handle protein mixtures. Substantial advances in both hardware and software for tandem MS have allowed two independent attempts, Human Proteome Project and ProteomicsDB, to characterize the entire human proteome (Kim et al., 2014; Wilhelm et al., 2014). Trypsin, due to its strict P1 specificity towards Lys/Arg, has been the workhorse enzyme to digest proteomic samples into easily ionizable peptides of desired size for downstream analysis. Current proteomic studies focusing on post-translational modifications (PTMs) such as phosphorylation face several challenges including locating the site of modification and avoiding losses during sample enrichment and MS fragmentation. As other proteases in addition to trypsin can easily be used during protein digestion, attempts were made to engineer specificity of proteases such as subtilisin and OmpT for improved identification of PTMs (88, 89). Though previously engineered variants suffered from limitations such as poor digestion efficiency or poor discrimination between peptides with/without PTM, efforts to create protein analogues of DNA restriction enzymes continue and could make the characterization simpler in the future. Apart from protein analytics, proteases are used by synthetic biologists for chemical or even light modulated degradation of proteins to design next generation sensors and memory devices (Bonger et al., 2011; Prindle et al., 2012; Renicke et al., 2013). For example, a recent report demonstrated using a mouse model that engineered

bacteria can be orally administered to colonize liver tumors and secrete reporter enzymes for cancer detection by *in vivo* imaging or simple fluorogenic substrate assay using urine sample (Danino et al., 2015). Multiple proteases with extended, orthogonal specificities can open up a whole new array of applications by allowing synthesis of integrated genetic circuits with minimal cross talk. Progress in both exosite and active site engineering of proteases is expected to come from a multipronged approach where the initial design from computational methods which take cues from natural diversity by mining sequence and structural databases is subsequently optimized using directed evolution techniques. With an increased understanding of the rules that govern substrate recognition and catalysis at an atomistic level, we can aspire for *de novo* arrangement of chemical groups in a 3-D scaffold which is not necessarily amino acid based.

1.6 Conclusions

Proteases constitute the largest group of industrially useful enzymes. With the increase in understanding of their physiological roles, they are emerging as a new class of biologics for clinical treatment. Unlike antibodies, proteases can achieve faster clearance of pathological targets due to high turnover. However, their therapeutic potential is undermined by a variety of shortcomings such as short half-life due to endogenous inhibitors and toxicity due to off-target proteolysis. Engineering the specificity of those proteases to overcome these limitations seems to be a viable solution. Under natural selection, proteases with high sequence similarity have evolved to display catalytic activity towards diverse peptide or protein substrates in a highly selective or non-specific fashion depending on the requirement for fitness advantage. The apparent plasticity of substrate determining motifs in proteases was taken advantage of using *in vitro* evolution

methods to generate variants with exquisite selectivity without compromising on high catalytic efficiency. The success rate and time involved in identifying beneficial mutations have substantially improved with advances in targeted diversity creation methods, high-throughput screening/selection platforms and comprehensive specificity profiling techniques. With the rise of next-generation biocatalysts, only creativity can limit the plethora of suitable applications which includes synthetic biology, protein analytics and protease therapeutics.

Chapter 2

Bacterial platform for anchored display or secretion of recombinant proteins with disulfide bonds

This chapter has been published as a research article titled ‘Single-cell Characterization of Autotransporter-mediated *Escherichia coli* Surface Display of Disulfide Bond-containing Proteins’ in the Journal of Biological Chemistry, 287:38580–38589. I performed all the experiments and the manuscript was written in collaboration with the co-authors listed in the paper.

2.1 Introduction

The display of recombinant proteins on the surface of microorganisms has attracted substantial interest from both an application (biotechnological or clinical) and from a basic microbiological standpoint towards understanding mechanisms of protein translocation. The utility of surface display of functional peptides and proteins for biotechnological applications has been demonstrated in several different contexts including: whole-cell biocatalysis based on esterases (Schultheiss et al., 2008), bioremediation using recombinant display of organophosphorous hydrolases and metallothioneins (Li et al., 2008; Valls et al., 2000), glucose-responsive biosensors (Shibasaki et al., 2001) and protein engineering of displayed libraries (Binder et al., 2010;

Jose and Meyer, 2007). In the context of surface display of antigenic recombinant peptides/proteins for delivery of live vaccines, besides the obvious safety advantage of using surface display on non-pathogenic bacteria, there are several other benefits like the cost and ease of manufacturing, the ability of bacterial components like lipopolysaccharide (LPS) to function as adjuvants to stimulate the immune system via Toll-like receptors leading to sustained immunity (Nhan et al., 2011) and the ability of the mammalian innate immune system to recognize prokaryotic mRNA (present only in live cells) leading to protective immunity (Sander et al., 2011).

In parallel, surface display of both recombinant and engineered native proteins has been used to delineate the mechanism of protein translocation. In gram negative bacteria like *Escherichia coli* in particular, display of proteins on the cell surface requires that the protein that is translated in the cytoplasm traverse across two separate lipid bilayers, the inner and outer membranes (Dautin and Bernstein, 2007). Consequently, gram-negative bacteria have evolved a diverse array of protein transport machinery (designated types I-VIII) dedicated to facilitate the translocation and ultimately secretion or surface display of proteins (Tseng et al., 2009). Autotransporters (AT, type Va) are believed to be the most abundant secretion pathway with >700 members that are ubiquitous in bacterial genomes (Henderson and Navarro-garcia, n.d.; Pallen et al., 2003).

ATs comprise an extended N-terminal leader sequence that is cleaved at the inner membrane (Szabady et al., 2005) followed by an N-terminal passenger (20-400 kDa) typically associated with virulence functions (adhesion, proteolysis, pore formation etc.) and finally a conserved C-terminal ~30 kDa β -barrel (Wells et al., 2007). ATs are attractive candidates for the surface display of recombinant proteins because they are

displayed at high-copy numbers (>100,000 protein copies) with minimal host-toxicity (Rutherford and Mourez, 2006). The nomenclature of ATs was based on the assumption that the primary sequence of the protein encodes all the information necessary for accurate translocation and ultimately surface display or secretion of the passenger domain. Although it was initially hypothesized that the C-terminal β -barrel formed a pore through which the N-terminal passenger was translocated (Pohlner et al.) this model has subsequently been challenged and an alternative model based on the aid of accessory proteins like the Bam complex has been developed (Bernstein, 2007; Sauri et al., 2009). Recent biochemical and functional studies on the C-terminal domain indicated that a conserved α -helix and the β -barrel comprise the minimum functional transport unit (Marín et al., 2010) but the role of the α -helix has since been suggested to be required for cleavage rather than actual secretion (Dautin and Bernstein, 2011).

There exists considerable controversy in the ability of ATs to transport passengers, either native or recombinant, containing folded elements, especially those containing disulfide bonds. Since disulfide bond formation is catalyzed in the periplasm of *E. coli* by the *Dsb* family of oxidoreductases, proteins containing thiols that form disulfide bonds are expected to be oxidized in the periplasm (Oudega, 2003). Using both native (EspP, IcsA) and recombinant (single-chains of antibodies) passengers, several groups have independently reported the ability of the passenger domain to fold in the periplasm in a proteolytically resistant state and subsequently be transported across the outer-membrane (Brandon and Goldberg, n.d.; Ieva et al., 2008; Skillman et al., 2005; Veiga et al., 2004, 1999). In contrast, studies utilizing native passengers engineered to contain cysteine residues (Hbp, Pet) have indicated that to a large extent, disulfide bonds between

cysteines that are not closely spaced stall passenger transport (Jong et al., 2007; Junker et al., 2009; Leyton et al., 2011).

An understanding of the variations in functional protein display at a single-cell level and quantifying the frequencies of cells in a clonal population capable of expressing functional protein can provide insight in to constraints imposed by the folded state of passenger on its transport. At the same time, quantifying the ability of ATs to display recombinant antigenic proteins containing multiple disulfide bonds would facilitate the adaptation of ATs for live-cell vaccine applications. Similarly, biotechnological applications that involve displaying libraries of protein molecules require knowledge of the number of functional recombinant protein molecules displayed on every single-cell. Although flow-cytometry has been previously applied to study of ATs, these have been predominantly restricted to quantifying epitope tags that indicate surface protein display and not functional protein display (Marín et al., 2010; Nhan et al., 2011; Pyo et al., 2009).

Here, we systematically investigate the ability of the Antigen 43 (Ag43) AT to display two different recombinant reporter proteins that are known to fold in the periplasm, a single-chain antibody (scFv) that contains two disulfides and chymotrypsin (rChyB) that contains four disulfides (Harvey et al., 2004; Venekei et al., 1996a). Using flow-cytometry to quantify surface display of functional protein at the single-cell level, we demonstrate that in spite of the known propensity of these passengers to fold in the periplasm, surface display of these proteins is rather efficient and can be achieved using only the C-terminal domain containing the α -helix and the β -barrel. Our flow-cytometric data is consistent with data obtained by immunofluorescence microscopy and western blotting on whole cell lysates. In contrast to previous reports (Jose et al., 1996), no

genetic manipulation like the use of *dsbA*- strains or the use of reducing agents like β -mercaptoethanol (BME) was necessary to accomplish efficient display. Our results indicate that display of recombinant proteins containing multiple disulfides can be achieved by employing the Ag43 system and that the vast majority of native AT including the autochaperone domain are not indispensable for heterologous protein display. These results have important implications for both understanding protein translocation by ATs and for the recombinant display of heterologous proteins for catalysis and engineering.

2.2 Experimental Procedures

2.2.1 Plasmid construction

Gene fragments coding for recombinant passenger (M18 scFv and *Rattus norvegicus* chymotrypsin (ChyB)) and signal peptides of Antigen 43 (Ag43), outer membrane protease T (OmpT), and pectate lyase (PelB) were obtained by PCR using oligonucleotides (oligos) (Integrated DNA Technologies, USA), as listed in Table 1.1. Templates for PCR to obtain gene fragments coding for M18 scFv and rChyB were kind gifts from the lab of G. Georgiou (University of Texas, Austin). Genomic DNA of *E. coli* MC1061 was used as template to obtain other gene fragments. *E. coli* MGB263 as a kind gift from Dr. Marcia Goldberg (Harvard school of Public Health, Boston). The genes coding for fragments of Ag43 (138 – 1039 aa, 552 – 1039, and 700 – 1039 aa) were amplified by PCR with oligos designed to encode a (GGGGS)₂ linker (5') and His₆ tag, in addition to restriction enzyme recognition sites (3') (**Table 2.1**). The PCR product was subsequently digested with KpnI and HindIII at 37 °C for 3h and was ligated using T4

DNA ligase at 25 °C for 4h to pBAD33 cut using the same restriction enzymes. The cells were then transformed into competent *E. coli* MC1061 via electroporation and verified by standard Sanger sequencing. This family of vectors designated pBAD_138, pBAD_552 and pBAD_700 containing the C-terminal fragments from Ag43 were used for the easy cloning of different passenger/leader combinations. The genetic fusion of signal peptide to 5' region of recombinant passenger gene was accomplished via the use of complementary oligos (**Table 2.1**) by overlap extension PCR (OE-PCR), essentially as described previously (Varadarajan et al., 2009a). Following OE-PCR, the product was gel-purified, digested with SacI and KpnI at 37 °C for 3h and ligated into the appropriate plasmid constructs. Ligated plasmids were transformed in to *E. coli* MC1061 cells (Varadarajan et al., 2009a) by electroporation and verified by standard Sanger sequencing. All plasmids use a standardized nomenclature (**Table 2.2**) that indicates the leader, passenger and the residues that originate from AT. For example, plasmid pBAD_AM18_138 contains the gene encoding for a fusion protein that has signal peptide of Ag43, M18 scFv, and 138 – 1039 aa of Ag43 AT.

2.2.2 Expression and labeling of M18 scFv

A standard 3 mL culture of cells harboring plasmids to surface display M18 scFv (e.g. pBAD_M18_138) was grown in LB medium (BD Diagnostics, NJ) in presence of 30 µg/mL of chloramphenicol (Thermo Fisher Scientific, NJ) to an optical density (OD₆₀₀) of 0.6 at 25 °C. Cells were then induced via the addition of 0.2% L-arabinose (Sigma, USA) to express M18 scFv for 12 h at 25 °C. Presence of M18 scFv on the surface of *E. coli* was characterized by the ability of intact cells to bind antigens (PA₆₃ and PAD4) using flow-cytometry.

Table 2.1 List of primers used in this study.

S.No	Primer name	Sequence	Comment
1	Ag43 leader-f	AGCCCGAGCTCCTAAGGAAAAGCTGATGAAAC GACATCTGAATACCT	1 and 2 were used to obtain PCR fragment coding for Ag43 leader. SacI site and Shine-Dalgarno (SD) sequence are present in 1.
2	Ag43 leader-r	AGCCAGCACCGGGAAGTGAC	
3	M18-f	ATGGCGGACTACAAAGATAT	3 and 4 were used to obtain PCR fragment coding for M18 ScFv. KpnI site is present in 4.
4	M18-r	CTACAGGTACCCGAGGAGACGGTGACTGAGGTT CCTTG	
5	Ag43 leader- M18-f	CTCCCGGTGCTGGCTGCTATGGCGGACTACAA GATATTC	1, 4, and 5 were used to genetically fuse Ag43 leader and M18 ScFv gene fragments by overlap extension (OE) PCR.
6	pelB leader-M18-f	AGCCCGAGCTCCTAAGGAAAAGCTGATGAAAT ACCTATTGCCTACGGCAGCCGCTGGATTGTTAT TACTCGCGGCCAGCCGGCCAATGGCGGACTACA AAGATATTCAG	4 and 6 were used in PCR to obtain gene fragment coding for PelB leader-M18 ScFv fusion protein. SacI site, SD sequence, pelB leader and region that can bind to M18 ScFv are present in 6.
7	ompT leader-M18-f	AGCCCGAGCTCCTAAGGAAAAGCTGATGCGGGC GAAACTGGGAATAGTCCTGACAACCCCTATTG CGATCAGCTCTTTTATGGCGGACTACAAAGAT ATTCAG	4 and 7 were used in PCR to obtain gene fragment coding for ompT leader-M18 ScFv fusion protein. SacI site, SD sequence, ompT leader and region that can bind to M18 ScFv are present in 7.
8	M18_stop_r	ATCCCAAGCTTTTACGAGGAGACGGTGACTAGG	6 and 8 were used in PCR to obtain gene fragment coding for PelB leader-M18 ScFv. Region that can bind to M18 ScFv, stop codon and HindIII site are present in 8.
9	Ag43 leader- rChyB-f	CTCCCGGTGCTGGCTGCTATCGTCAACGGAGAGG ATGC	1, 8 and 9 were used in PCR to obtain gene fragment coding for Ag43 leader – rChyB fusion protein. KpnI site is present in 9.
10	rChyB-r	CGGGGTACCGTTGGCTTCCAAGATCTGCTG	
11	Ag43 138 –f	GGGGTACCGgtggaggtggttctggtggaggtggttctGGAC AGAGCCTTCAGGGACG	10 and 13 were used to obtain PCR fragment coding for 138-1039 aa of Ag43. 11 and 13 were used to obtain PCR fragment coding for 552-1039 aa of Ag43. 12 and 13 were used to obtain PCR fragment coding for 700-1039 aa of Ag43. 10, 11, and 12 contain KpnI site and introduces (GGGS) ₂ linker N-terminal to Ag43 fragment. 13 contains HindIII site and introduces His ₆ tag at the C-terminal of Ag43 fragment
12	Ag43 552 –f	GGGGTACCGgtggaggtggttctggtggaggttctccaca aatgtcactctgc	
13	Ag43 700-f	GGGGTACCGgtggaggtggttctggtggaggttctCGCA GTGAAAATGCTTATCGTG	
14	Ag43-r	CCCAAGCTTTCAAGTATGGTGATGATGGTGGCT GCCGAAAGTCAATTCACTGTGGC	
15	M18_C28A_r	TAATTCCTAATGTCCTGACTTGCCCTGGCACTG	1 and 15 were used in PCR to obtain Ag43 leader-M18 fragment containing C28A mutation. 15 contains region of M18 that can bind to 16 to perform OE-PCR subsequently. Similarly other M18 fragments containing mutations (C28A, C93A, C155A, and C229A) were obtained using primer pairs 16/17, 18/19, 20/21, and 22/4.
16	M18_C28A_f	GCAAGTCAGGACATTAGGAATTA	
17	M18_C93A_r	ATTGGCACTTACTTTGCTCAACAGGGCAATAC	
18	M18_C93A_f	GTATTGCCCTGTTGAGCAAAGTAAGTGCCAAT ATC	
19	M18_C155A_r	CAGTGAAGATTTCCGCCAAAGATTCTGGCTAC	
20	M18_C155A_f	GTAGCCAGAAATCTTTGGCGGAAATCTTCACTG	
21	M18_C229A_r	CGTAGCAACCCCGATCTTGGCGGCAAT	
22	M18_C229A_f	GCCGCAAGATCGGGGTGCTACG	

Table 2.2 List of plasmids used in this study. pBAD33 was used as the cloning vector and contains chloramphenicol resistance marker.

Plasmid	Leader	Recombinant passenger	aa from Ag43 AT
pBAD_AM18_138	Ag43	M18 scFv	138 – 1039 aa
pBAD_AM18_552	Ag43	M18 scFv	552 – 1039 aa
pBAD_AM18_700	Ag43	M18 scFv	700 – 1039 aa
pBAD_OM18_700	OmpT	M18 scFv	700 – 1039 aa
pBAD_PM18_700	PelB	M18 scFv	700 – 1039 aa
pBAD_AChy_700	Ag43	34-263 aa of rat chymotrypsin	700 – 1039 aa
pBAD_AM18_C/A_700	Ag43	M18 scFv with C28A, C93A, C155A and C229A mutations	700 – 1039 aa
pBAD_PM18	PelB	M18 scFv	Soluble expression, no fusion to Ag43.

PAD4 containing FLAG epitope tag (rPAD4) was recombinantly expressed in *E. coli* and purified as described previously (Jeong et al., 2007) and PA₆₃-FITC was obtained from List biological laboratories, Inc., CA. 100 µl of cells expressing M18 scFv from the appropriate construct normalized to OD₆₀₀ of 2 units were washed twice in PBS and incubated with 200 nM of rPAD4 for 1 h at 25 °C. Cells were then washed, resuspended in 20 µl of PBS, and incubated with 40nM of anti-FLAG-Phycoerythrin (PE) (ProZyme, Inc., CA) at 25 °C for 45 min in the dark. For labeling using PA₆₃, 25 µl of cells at OD₆₀₀ 2 were washed twice in PBS and incubated with 300 nM of PA₆₃-FITC at 25 °C for 1 h in the dark. A 5 µL aliquot of cells labeled with fluorophore was diluted with 0.5 mL of PBS and analyzed by flow-cytometry.

2.2.3 Expression and labeling of rChyB

Cells harboring pBAD_AChy_700 were grown in 3 mL of LB medium supplemented with 30 µg/mL of chloramphenicol to OD₆₀₀ 0.6 at 37 °C. Protein expression was induced at 37 °C for 2h using 0.2% L-arabinose. To characterize the

proteolytic activity of chymotrypsin, a peptide substrate (Chy-BQ7) was obtained by conjugating the synthetic peptide (Ac-CAAPYGSKGRGR-CONH₂) (Genscript, NJ) to a fluorophore (BODIPY) (Invitrogen, NJ) and a non-fluorescent quencher (QSY7) (Invitrogen, NJ) as previously described (Varadarajan et al., 2005). 1 mL of cells at OD₆₀₀ 0.1 were washed twice in 1% sucrose (Sigma, USA) solution and incubated for 1 h with 20 nM of Chy-BQ7 at 25 °C in the dark. Surface display of chymotrypsin was independently verified by incubating whole cells with an antibody that can bind to chain C of rChyB (Santa Cruz Biotechnology, Inc, CA). 20 µl of cells expressing chymotrypsin at OD₆₀₀ 2 were washed twice in PBS and incubated with 50 µg/mL of biotin conjugated anti-chyB antibody for 1 h. Cells were subsequently washed with PBS and incubated with 5 µg/mL of streptavidin conjugated PE (Jackson Immuno Research, PA) for 30 min at 25 °C in the dark. A 5 µL aliquot of cells labeled with fluorophore was diluted with 0.5 mL of PBS or 1% sucrose and analyzed by flow-cytometry.

2.2.4 Flow-cytometry

Cells labeled with fluorophore were analyzed using BD Jazz cell sorter (BD Biosciences, CA) at ~ 8000 events/second and offset 1. For all samples, a minimum of 10,000 events were recorded and the events were triggered using side scatter. Fluorophores were excited using 488 nm laser and emission was measured using 530/40 (BODIPY and FITC) and 580/30 (PE) filters.

2.2.5 Immunofluorescence microscopy

Cells expressing chymotrypsin were labeled using biotinylated antibody against chain C of chymotrypsin (Santa Cruz Biotechnology, Inc, CA) and PE conjugated to

streptavidin as described above. After washing with PBS to remove excess fluorophore, cells were resuspended to an OD₆₀₀ 1. Slide preparation was performed as described previously (Zeng and Golding, 2011). Briefly, 2 μ L of cell suspension was pipetted on a glass slide and a small piece of agarose pad was placed on top of it to prevent cell movement. Microscopy was performed using an inverted epifluorescence microscope (Eclipse Ti, Nikon) with a 100x objective (Plan Fluo, NA 1.4, oil immersion) and Cy3 filter cube (Nikon). Images were captured in a cooled 1024x1024 EMCCD camera (Cascade II 1024, Photometrics, USA) at a gain of 2000 and an exposure time of 100 ms using Nikon Elements software (Nikon, USA).

2.2.6 Western blotting

A standard 3 mL culture of cells displaying recombinant protein (M18 scFv or rChyB) was obtained as described above. A whole-cell pellet obtained by centrifugation at 16,000xg for 2 min was resuspended in 2X sample buffer (Bio-rad, USA) diluted with equal volume of 50 mM Tris (Sigma, USA) and boiled at 100 °C for 5 min. SDS-PAGE of samples (10⁷ cells/lane) was performed at 120V for 90 min in a 4-12% polyacrylamide gel (Lonza, USA) using a Mini-PROTEAN gel electrophoresis system (Bio-rad, USA). After electrophoresis, protein molecules were transferred to a PVDF membrane (Bio-rad, USA) at 100V for 60 min using a wet electroblotting system (Bio-rad, USA). Procedures for SDS-PAGE and blotting were essentially performed as described previously (Ausubel, 1999). To reduce non-specific binding, the membrane was blocked by overnight treatment with Tris-buffered saline containing 5% milk and 0.1% Tween-20 (Bio-rad, USA). Membrane was labeled using rabbit anti-His antibody (300 ng/mL, Genscript, USA), goat anti-rabbit antibody conjugated to HRP (40 ng/mL, Jackson

Immunoresearch, USA) and chemiluminescent HRP substrate (SuperSignal west pico, Thermo Scientific, USA). Protein marker (15-225 kDa, EMD Millipore, USA) was used as reference in estimating the size of bands. Chemiluminescent imaging of developed blot was performed using a CCD camera in an imaging cabinet (Alpha Innotech Fluorchem SP, USA).

2.3 Results

2.3.1 Surface display of functional single-chain antibody via fusion to Ag43 revealed by flow-cytometry

Ag43, a native *E. coli* AT, mediates bacterial autoaggregation via self-recognition of the passenger domains (van der Woude and Henderson, 2008). It is synthesized as a 1039 amino acid (aa) polypeptide containing the following domains: (i) leader peptide (aa 1-52) (ii) N-terminal unstructured domain of unknown function (aa 53-138) (iii) passenger with a β -helical core (aa 138-552) containing a proteolytic site (551-552) and a putative autochaperone domain (aa 600-707) (iv) α -helix (aa 710-730) that is threaded via the (v) C-terminal β -barrel pore (aa 731-1039) (**Figure 2.1A**). The domains were identified by homology modeling of Ag43 using the PHYRE server (Kelley and Sternberg, 2009). In order to investigate the ability of the AT to mediate display of recombinant proteins containing disulfides, we synthesized a genetic construct coding for the native Ag43 leader, single-chain antibody (scFv) M18 and aa 138-1039 from Ag43 via standard overlap extension PCR in a two-step cloning strategy.

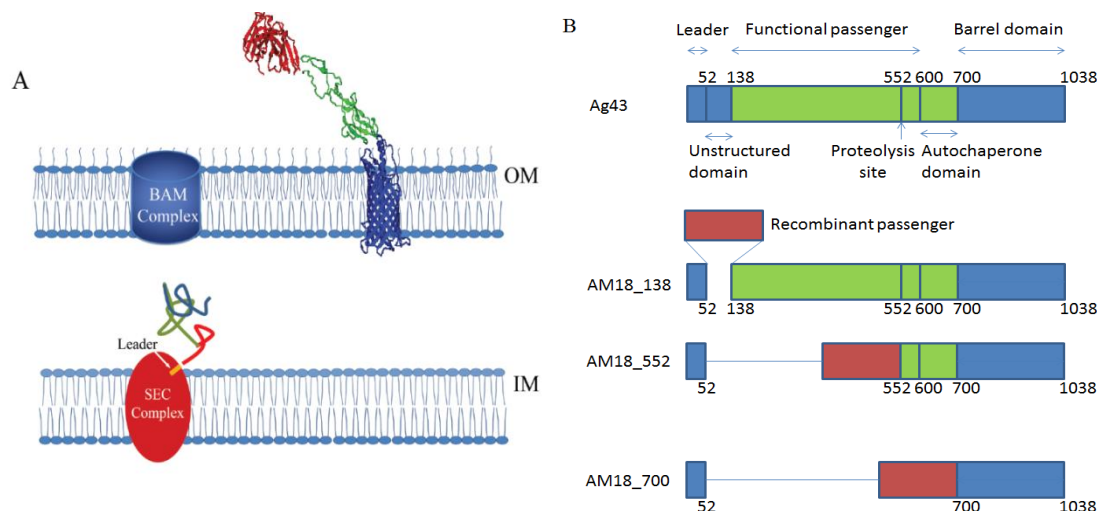


Figure 2.1 (A) Schematic describing the transport of AT to the outer membrane. (B) Schematic of Ag43 and fusion proteins containing recombinant passenger (red), Ag43 passenger (green) and Ag43 barrel domain (blue) used in this study.

The M18 scFv contains two disulfides, one in each of the variable regions, and is ~25 kDa globular protein. It was previously isolated using bacterial display as a high-affinity ($K_d = 35$ pM) binder to anthrax toxin protective antigen (PA) (Harvey et al., 2004). The functional status of M18 scFv when displayed as a fusion to Ag43 was investigated using flow-cytometry. Protein expression was induced from pBAD_AM18_138 using L-arabinose for 12h and the cells were incubated with FITC labeled heptameric PA₆₃. Flow-cytometry revealed that cells expressing M18 scFv were bimodal (**Figure 2.2A**, 30% positive, Mean = 83) whereas cells surface displaying an irrelevant protein or uninduced cultures were uniformly negative (Mean = 3). In order to ensure that our single-cell functional assay only detects folded disulfide-bond containing proteins, a cysteine-free M18 scFv variant (AM18_C/A_700) was synthesized by standard overlap PCR. As expected, cells expressing AM18_C/A_700 showed poor labeling with PA₆₃-FITC (2% positive, Mean = 28). Similarly, in order to establish the

requirement of the AT for surface display, cells expressing the soluble form of M18 (PM18) were labeled and shown to be negative.

In parallel, western blotting experiments were performed with whole-cell lysates to confirm the presence of the fusion construct. Recombinant fusion protein expressed from pBAD_AM18_138 was immunoblotted using a rabbit anti-His antibody (**Figure 2.2B**). In addition to the full length protein (135 kDa) additional bands corresponding to degraded protein were observed (35 kDa), consistent with the prolonged induction.

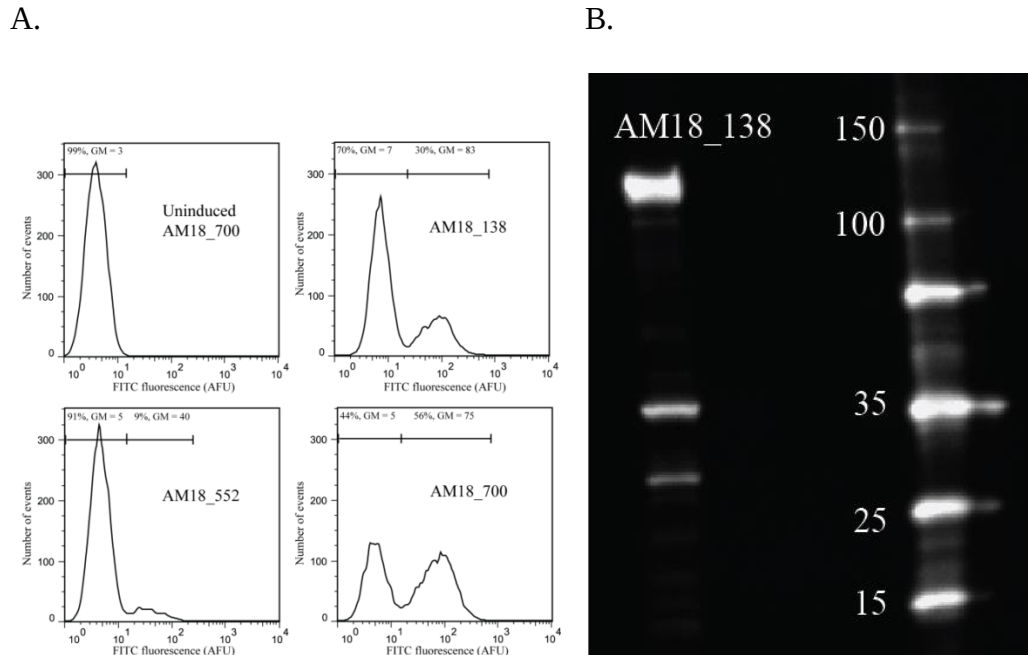


Figure 2.2 Surface expression of functional M18 scFv as a function of native passenger length. (A) Flow-cytometric quantification of AT mediated display by fusion of M18 scFv (B) Western blot showing presence of AM18_138 (expected size = 135 kDa).

2.3.2 C-terminal domains comprising the β -barrel and the α -helix are sufficient to achieve efficient surface display

Since initial studies utilizing cells harboring pBAD_AM18_138 indicated that functional surface display of M18 scFv could be achieved, we next sought to

systematically investigate the contribution of the different domains of Ag43 to the surface display of M18 scFv. Accordingly, two separate plasmids designated pBAD_AM18_552 and pBAD_AM18_700 (**Figure 2.1B**) were constructed using site-overlap extension. pBAD_AM18_552 encodes a truncated N-terminal passenger but an intact autochaperone domain, while pBAD_AM18_700 does not encode the autochaperone domain (**Figure 2.1B** and **Table 2.2**).

Protein expression was induced by the addition of L-arabinose for 12h and the cells were labeled using PA₆₃-FITC and analyzed on the flow cytometer. Cells expressing pBAD_AM18_552 were reproducibly less efficient (**Figure 2.2A**, 9% positive, Mean = 40, n = 3) at functional M18 scFv display compared to pBAD_AM18_138. In contrast, greater than half of all cells harboring pBAD_AM18_700 showed expression of M18 scFv on the surface (56% positive, Mean = 75), demonstrating that at least for recombinant expression of M18 scFv the C-terminal domains containing the α -helix and β -barrel are sufficient.

2.3.3 Binding of bulky antigens is not sterically hindered

The ability of the carbohydrate chains of lipopolysaccharide (LPS) to interfere with protein labeling on the cell-surface is well-documented (Voorhout et al., 1986). Since our labeling strategy utilized the large heptameric antigen PA₆₃ to probe the functional status of M18 scFv, we tested if steric hindrance from LPS could explain the heterogeneity of the population in the efficiency of surface display. The epitope of M18 scFv has been mapped to domain 4 within PA (PAD4) and the construction, expression and purification of recombinant PAD4 (rPAD4) with an N-terminal FLAG epitope tag (DYKDDDDK) has been reported previously (Leysath et al., 2009). Cells expressing pBAD_AM18_700

were incubated first with rPAD4 (200nM) and then with anti-FLAG-PE (40 nM) and interrogated on the flow-cytometer. Based on comparison of frequency of cell population that can bind to rPAD4 (46% positive) and PA₆₃-FITC (51% positive), we can infer that binding of M18 scFv to its antigen does not appear to be sterically hindered with this expression system (**Figure 2.3**).

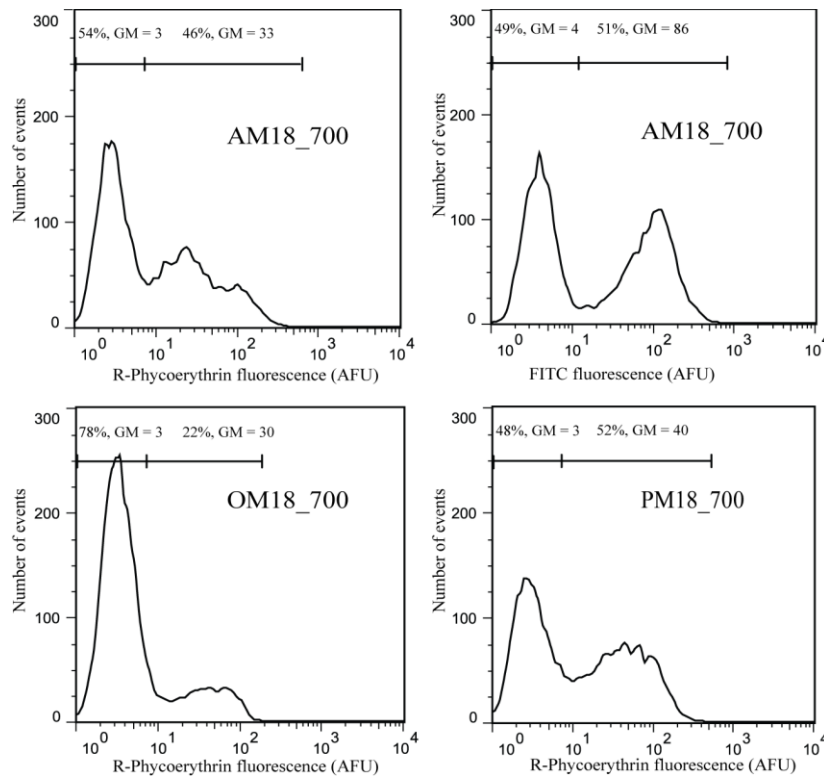


Figure 2.3 Surface expression of functional M18 as a function of N-terminal leader sequence and size of the antigen.

2.3.4 The extended signal peptide region is not indispensable for functional surface display

Signal peptides are responsible for translocation of AT across the inner-membrane. In contrast to other periplasmic proteins or even other outer-membrane proteins, the signal peptides of AT have an extended N-terminal region and it has been previously

hypothesized that the extended N-terminal region prevents the passenger domain from acquiring a conformation incompetent for translocation (Szabady et al., 2005). In order to quantitatively determine the significance of the extended native signal peptide at the single-cell level, two additional fusion constructs were synthesized by PCR and cloned into pBAD33. Since our data indicated that efficient surface display is accomplished via fusion to the C-terminal domains of Ag43 (700-1039 aa) (**Figure 2.2A**) these were used as optimal fusion partners. pBAD_PM18_700 encodes for a tripartite fusion between the leader sequence of pectate lyase B (pelB), a periplasmic protein, M18 scFv and Ag43 (residues 700-1039) and pBAD_OM18_700 for an identical tripartite fusion with the leader sequence of OmpT, an outer membrane protein (**Table 2.2**). Flow-cytometric analysis of M18 scFv using rPAD4 immunofluorescent sandwiches (**Figure 2.3**) demonstrated that the frequency of cells harboring pBAD_AM18_700 (native leader, 46 % positive) was not largely different from those harboring pBAD_PM18_700 (pelB leader, 52 % positive). Populations of cells expressing pBAD_OM18_700 (ompT leader, 22% positive) were however not as efficient in the functional display of M18 scFv. These results indicate that while the native Ag43 leader is not a requirement for AT mediated surface display, substituting the leader sequence from other outer membrane proteins like ompT leads to suboptimal display of functional protein molecules on the surface (**Figure 2.3**).

2.3.5 Surface display of functional M18 scFv is not improved by the addition of reducing agents or the use of protease-deficient strains

Prior work has suggested that the surface display of disulfide bond-containing passengers using the AT system are hampered by the formation of disulfide bonds in the

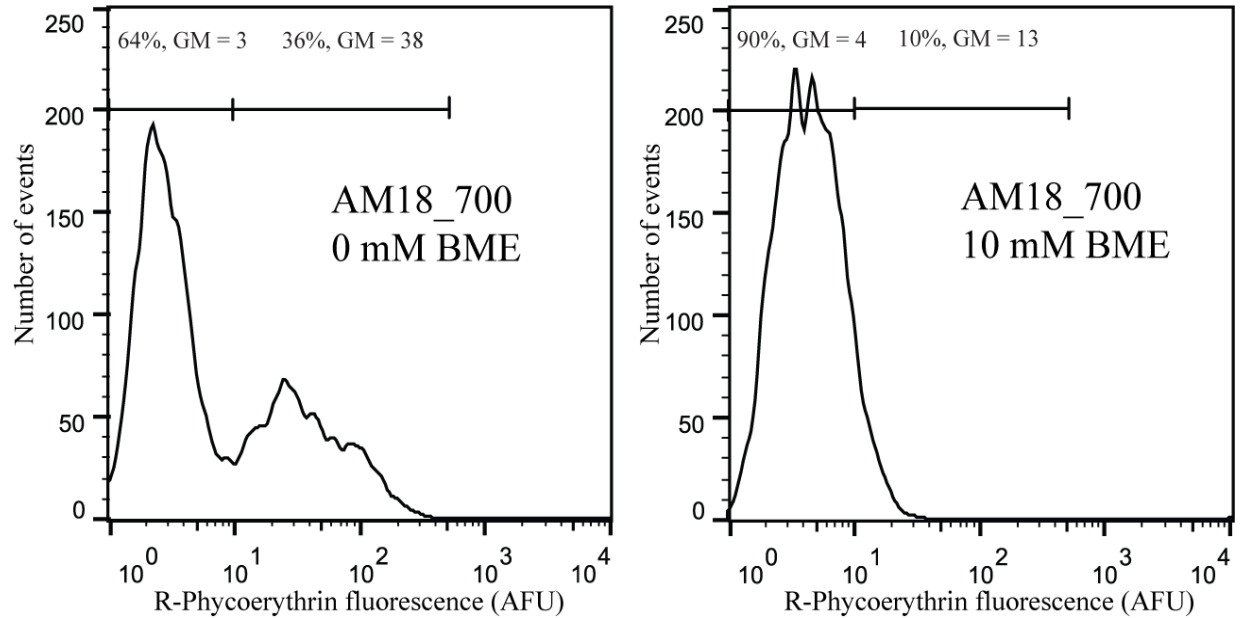
periplasm and that the yield of surface displayed protein can be improved either by the use of genetically engineered *E. coli* strains (*dsbA*-) or by the addition of reducing agents like β -mercaptoethanol (BME) during cell growth and protein expression (Jose and Zangen, 2005; Rutherford and Mourez, 2006; Veiga et al., 1999). In order to test this possibility in the current context, cells expressing pBAD_AM18_700 were grown to mid-log phase and expression of M18 scFv was induced for 12h in LB media containing 10 mM BME. Subsequently, the cells were transferred to media devoid of BME for 1h to facilitate refolding prior to analysis. An aliquot of cells was incubated with rPAD4 and anti-FLAG-PE and the populations were analyzed on the flow cytometer. Growth and induction in the presence of BME had a negative effect on functional M18 scFv display (10% positive, Mean = 13) compared to an identical culture grown without BME (36 % positive, Mean = 38) (**Figure 2.4A**). Secondly, in order to test if the use of OmpT knockout strains would improve the frequency of cells expressing functional M18 scFv, we used the *E. coli* MGB263 strain as the host for surface display. Flow-cytometric analysis of cells expressing pBAD_AM18_700 revealed that both the frequency and mean of cells expressing functional M18 scFv was identical in both strains (**Figure 2.4B**) indicating that OmpT mediated proteolysis had no discernible effect on functional protein display in our current system.

2.3.6 Surface display of chymotrypsin

Rattus norvegicus chymotrypsin B (*chyB*) is a prototypical serine protease that cleaves immediately after amino acids with aromatic side chains like tyrosine. ChyB has three chains (A, B and C) held together by disulfide bonds and mature recombinant ChyB (rChyB) containing only chains B and C (aa 34-263) contains the peptidase unit and is

catalytically active. rChyB contains four disulfide linkages (aa 60-76, 154-219, 186-200 and 209-238, chymotrypsinogen numbering) and folds into a globular structure that is very different from the standard β -solenoid-like folds of AT passengers.

A.



B.

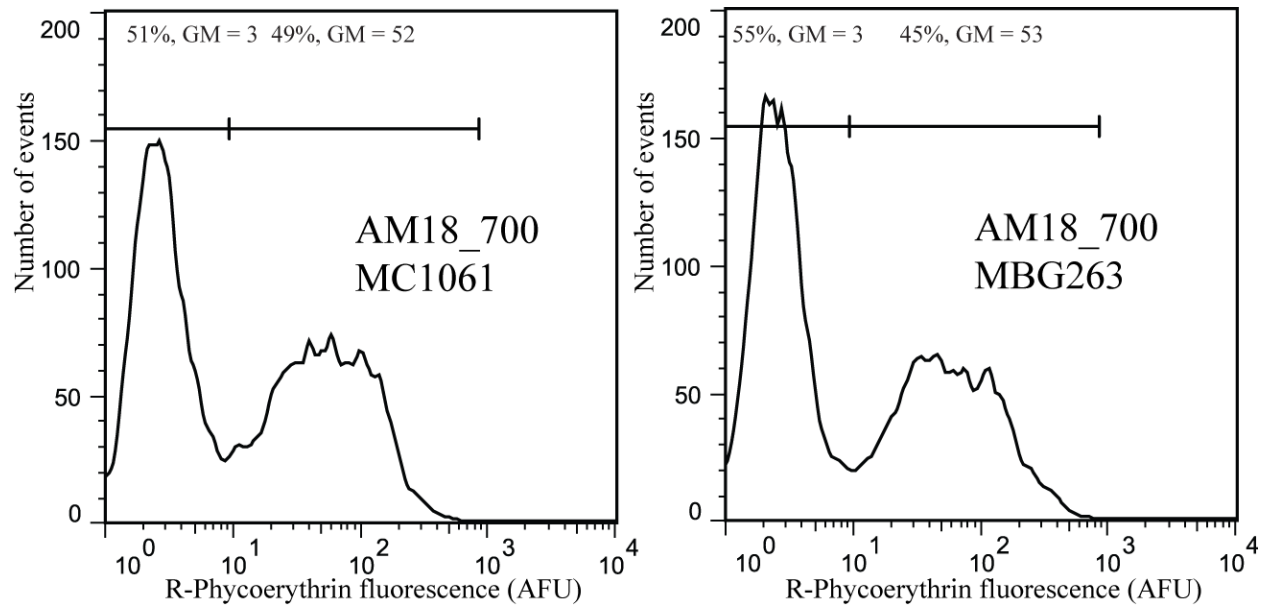


Figure 2.4 (A) Effect of reducing agent and (B) host membrane protease OmpT on the surface display of functional M18 scFv.

In order to investigate the ability of Ag43 to display rChyB on the cell surface, a genetic fusion with both *ag43* leader and the gene coding for C-terminal domains (aa 700-1039) of Ag43 was constructed using PCR and cloned into pBAD33 to yield pBAD_AChy_700 (**Table 2.2**). Since chymotrypsin non-specifically degrades cellular proteins we rationalized that lower protein expression would preserve cell viability. Secondly, since chymotrypsin is an enzyme, we reasoned that we could detect lower levels of expressed protein using a catalytic assay in comparison to the antibody-based stoichiometric assay. Cells containing pBAD_AChy_700 were therefore induced at 37 °C for only 2h. Cells were subsequently labeled with biotin-conjugated anti-ChyB antibody (chain C specific) and streptavidin-PE and analyzed on the flow-cytometer. Cells expressing pBAD_AChy_700 (Mean = 11) were uniformly labeled and could be reproducibly detected (**Figure 2.5**) compared to either un-induced or M18 scFv-expressing cells (Mean = 4) confirming the surface display of rChyB using the AT system. The presence of full-length fusion protein was also independently confirmed by immunoblotting employing an anti-His antibody.

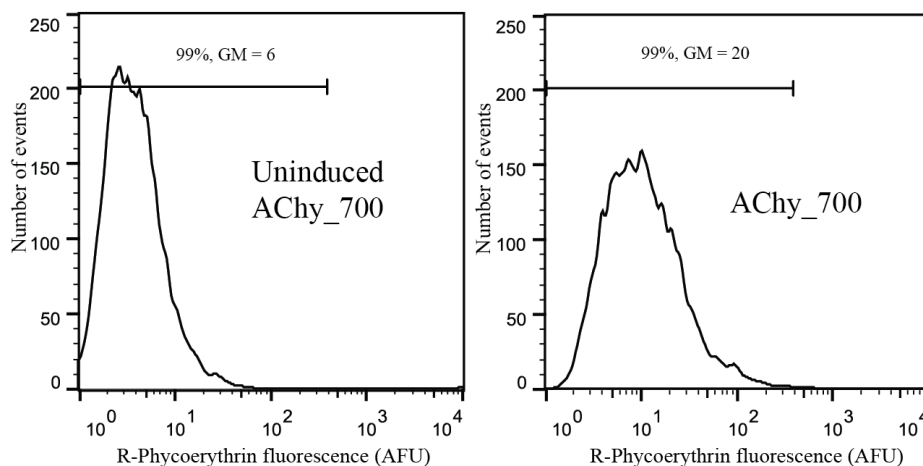


Figure 2.5 Flow-cytometry analysis of cells expressing rChyB labeled with biotin conjugated anti-ChyB antibody (chain C specific) and streptavidin-PE.

Next, we evaluated the surface localization of the recombinant chymotrypsin molecules using immunofluorescent microscopy. *E. coli* cells expressing rChyB were detected using a biotinylated antibody specific for chain C. Consistent with our flow-cytometry data, pBAD_AChy_700 displayed uniform expression on both the polar and lateral surfaces (**Figure 2.6**). No fluorescence was observed when either uninduced cells or cells expressing pBAD_AM18_138 were labeled under identical conditions. Examination of the morphology of cells expressing AChy_700 indicated that the cells were coccoid. We confirmed that the coccoid nature was an experimental artifact due to the extended time required for labeling and subsequent imaging (~3h) of live cells and not due to protein expression by imaging control samples.

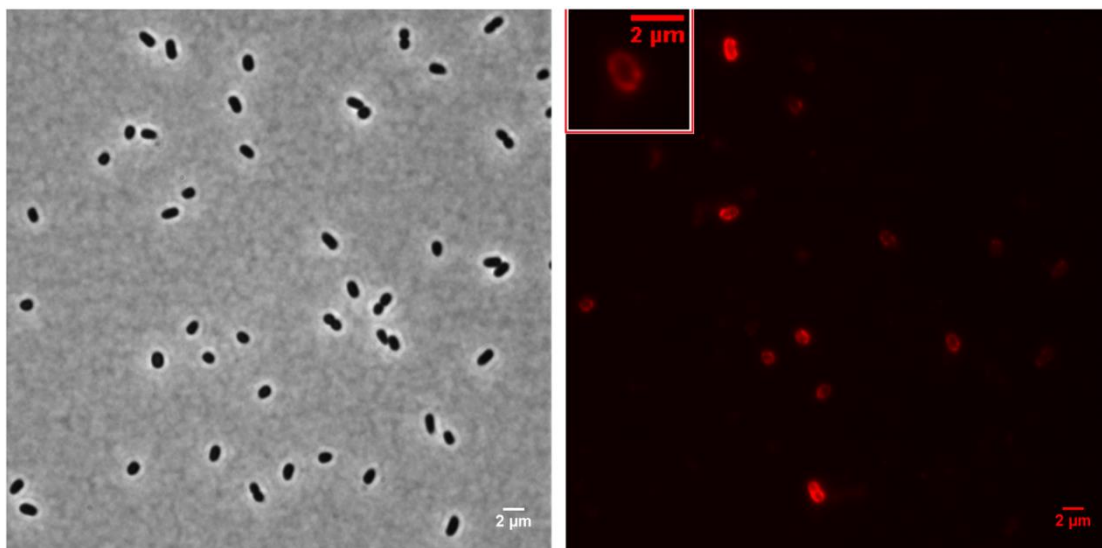


Figure 2.6 Phase contrast and immunofluorescence microscopy showing surface display of chymotrypsin. Enlarged image of a cell is shown in the inset.

2.3.7 rChyB displayed on the surface is enzymatically active

The enzymatic activity of surface-displayed rChyB was investigated using a Forster Resonance Energy Transfer (FRET)-based peptide assay on the flow-cytometer, essentially as described previously (Varadarajan et al., 2005). Briefly, cells displaying the

protease are incubated with a positively charged peptide substrate that contains a protease-sensitive linker sandwiched between a FRET pair. Cleavage of the peptide linker leads to loss of FRET and the product molecules are captured locally on the same cells using electrostatic interactions. Cells expressing pBAD_AChy_700 were incubated with 20 nM of the Chy-BQ7 FRET substrate containing a chymotrypsin-sensitive linker for 1h in 1% sucrose. Subsequent flow-cytometric analysis (**Figure 2.7**) revealed that the cells expressing pBAD_AChy_700 (79% positive, Mean = 251) were uniformly labeled compared to pBAD_AM18_700 that expressed M18 scFv which was uniformly negative (Mean = 7). Furthermore, proteolytic activity of rChyB expressed from pBAD_AChy_700 could be inhibited by pre-incubation with 0.1 mg/mL soybean trypsin/chymotrypsin inhibitor (**Figure 2.8**).

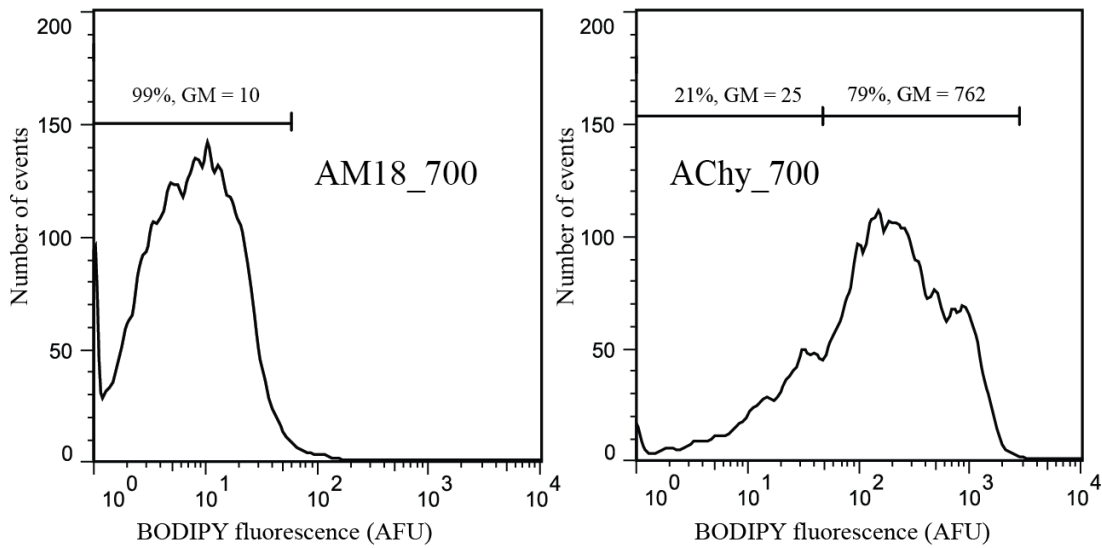


Figure 2.7 Chymotrypsin displayed on the surface of *E. coli* is enzymatically active. Flow-cytometry analysis of cells expressing M18 antibody or rChyB using Chy-BQ7 FRET substrate.

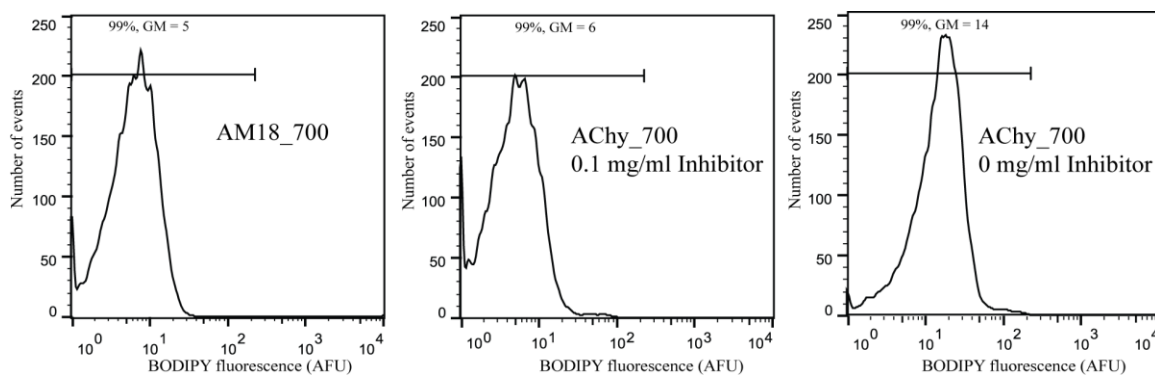


Figure 2.8 Inhibition of the catalytic activity of surface displayed chymotrypsin.

2.4 Discussion

ATs comprise a large superfamily of extracellular virulence proteins secreted by Gram-negative bacteria. In spite of the fact that the naturally occurring passengers of ATs exhibit diversity in sequence, function and size, most of them do not contain widely spaced cysteines. This conserved feature has been reported to have resulted from structural constraints required for translocation across the outer-membrane mediated by the C-terminal β -barrel. Since disulfide bonds are oxidized in the periplasm prior to translocation across the outer-membrane, the export of native passenger variants containing widely spaced disulfides is severely restricted due to the aforementioned structural constraint (Leyton et al., 2011). The inconsistencies in literature regarding the ability of ATs to export disulfide bond-containing passengers is further complicated by the fact that different reports use different ATs with different replacement points for the native or heterologous passengers. In this report, with the aid of quantitative flow-cytometry we have studied at the single-cell level the ability of an *E. coli* AT Ag43 to export functional recombinant passenger molecules containing multiple disulfides by varying (a) length of the native passenger by including/deleting the different domains comprising the (i) β -helical core (ii) auto-proteolysis site and (iii) autochaperone domain

containing the C-terminal β -hairpin and (b) the leader sequence that facilitates inner-membrane translocation. As reported here, the efficient display of globular proteins like the single-chain antibody fragment (2 disulfides) against anthrax toxin or the serine protease chymotrypsin (4 disulfides) can be accomplished in the absence of both the β -helical core (predicted aa 138- 600 of Ag43) and the autochaperone domain (predicted aa 600-707 of Ag43). Based on our results, an intact β -barrel with the invariant α -helix is sufficient for the extracellular display of globular recombinant protein in a functional state. This is somewhat surprising given that slow folding of C-terminal region of passenger which folds into a β -helix has been implicated to be a crucial element for translocation/export (Junker et al., 2006; Leyton et al., 2012). While it is likely that the globular structure of chymotrypsin is compatible with extracellular export via ATs given that the IgAP/Hbp ATs contain a chymotrypsin/trypsin-like protease at their N-terminus, two important distinctions need to be made (Johnson et al., 2009; Otto et al., 2005). First, the native IgAP/Hbp proteolytic components are located at the N-terminus of the conserved β -helical core that is believed to facilitate their translocation whereas our recombinant construct, pBAD_AChy_700 displaying chymotrypsin is devoid of the β -helical passenger or the auto-chaperone domain. Second, the native passengers of IgAP/Hbp do not contain disulfide bonds whereas rChyB is expected to contain four disulfide bonds, with one of the disulfide bonds bridging >50 aa (154-219, chymotrypsinogen numbering). These data suggest that our recombinant constructs might be similar in structure to *Pseudomonas aeruginosa* AT, EstA. The recently solved crystal structure of EstA indicated that unlike classical ATs, the esterase passenger of EstA adopts a globular structure dominated by α -helices and loops (van den Berg, 2010). Our

results are consistent with the hypothesis that the requirement of the autochaperone domain is tied to the nature of the N-terminal passenger but is not a universal requirement. Similarly, the ability of Ag43 to translocate disulfide bond containing proteins suggests that regardless of the spacing of the cysteines within the primary sequence, the overall globular structure of the passenger might dictate translocation efficiency, at least when heterologous proteins are being displayed. It remains to be tested if other recombinant passengers that adopt a similar fold to single chain antibodies or chymotrypsin can be efficiently displayed using Ag43. The use of protease-deficient strains or reducing agents during protein expression did not have a positive effect on the surface display of recombinant proteins using the Ag43 system. In parallel, our data investigating the effect of the Sec-dependent extended N-terminal leader sequence indicated that genetic replacement with leader sequences of either outer-membrane proteins or periplasmic proteins yielded display of functional passenger. In this regard our data is consistent with a recent report indicating that the native signal peptide was not essential for either secretion or function of the plasmid encoded toxin (Pet) AT (Leyton et al., 2010).

It remains to be seen if the modular architecture of Ag43 for heterologous display as demonstrated here is also applicable to other ATs. Our work also opens avenues for the engineering of recombinantly displayed proteins mediated by the Ag43 system comprising only the α -helix and the β -barrel by employing high-throughput flow-cytometry (Becker et al., 2007; Varadarajan et al., 2008). A special feature of the AT extracellular transport system is that by including the auto-proteolysis domain, switching

between surface-display and secretion of recombinant protein is rather straightforward (Wargacki et al., 2012).

2.5 Conclusions

Autotransporters (ATs) are a family of bacterial proteins containing a C-terminal β -barrel forming domain that facilitates the translocation of N-terminal passenger domain whose functions range from adhesion to proteolysis. Genetic replacement of the native passenger domain with heterologous proteins is an attractive strategy not only for applications like biocatalysis, live-cell vaccines and protein engineering but also for gaining mechanistic insights towards understanding AT translocation. The ability of ATs to efficiently display functional recombinant proteins containing multiple disulfides has remained largely controversial. By employing high-throughput single-cell flow-cytometry, we have systematically investigated the ability of the *E. coli* AT Antigen 43 (Ag43) to display two different recombinant reporter proteins, a single-chain antibody (M18 scFv) that contains two disulfides and chymotrypsin (rChyB) that contains four disulfides by varying the signal peptide and deleting the different domains of the native protein. Our results indicate that only C-terminal β -barrel and threaded α -helix are essential for efficient surface display of functional recombinant proteins containing multiple disulfides. These results imply that there are no inherent constraints for functional translocation and display of disulfide bond containing proteins mediated by the AT system and should open new avenues for protein display and engineering.

Chapter 3

High-throughput screening of recombinant protease libraries in *E.coli*

A part of this chapter has been accepted for publication as an article titled ‘Functional enrichment by direct plasmid recovery after Fluorescence Activated Cell Sorting’ in the BioTechniques journal. I performed all the experiments for the chymotrypsin system and the manuscript was written in collaboration with the listed authors.

3.1 Introduction

Protease engineering has been successfully employed to confer desired properties artificially for clinical and biotechnological applications. An example of early success was improving the half-life of tissue-type plasminogen activator (t-PA), a thrombolytic agent, by decreasing its propensity to cleave plasminogen activator inhibitor-1 without compromising its plasminogen activation kinetics (Madison et al., 1989). This was accomplished by deleting a loop predicted to interact with the inhibitor based on the structure of trypsin co-crystallized with bovine pancreatic trypsin inhibitor, homology between catalytic domain of t-PA and trypsin and structural similarity in serine protease-serpin complexes. As engineering proteases by site-directed mutagenesis (SDM) involves identification of essential amino acids based on either intuition or experimental data accumulated over a period of time, the knowledge obtained by engineering one

enzyme does not readily translate to other enzymes even if they are structurally similar. Not surprisingly, despite extensive engineering efforts to modify trypsin or chymotrypsin, they have primarily resulted in enzymes with poor kinetics or specificity or both (Hung and Hedstrom, 1998; Rauh et al., 2002; Venekei et al., 1996b). The significance of directed evolution in protein engineering is evident from the therapeutic success of antibodies and their fragments engineered using phage display technology (Thie et al., 2008). Unfortunately, unlike antibody or protein engineering that only aims to optimize binding efficiency, enzyme engineering is much more challenging since binding is only the first step in the catalysis, and catalytic efficiency does not have an obvious relationship to the strength of substrate binding. Furthermore, high affinity binders are likely to function as inhibitors that impede catalysis (Baggio et al., 1996). Second, assays for binding can modify one of the interacting partners to append tags and thus binding can be directly monitored by the use of the tag. In enzymatic screening however, the substrate gets turned over into product and hence the ability to continuously monitor reaction progress is not straight forward. These constraints are exemplified by the observation that there are over twenty FDA-approved engineered protease inhibitors but no proteases with modified substrate binding pocket are currently used in the clinic.

Advances in technologies to generate DNA libraries with diversity focused on hot spots, to screen protein libraries for presence of desired phenotype in high-throughput and to repeat both the steps iteratively in quick succession until a variant with desired characteristics is identified, have provided a robust ‘shotgun’ approach to achieve the moving target of engineering enzymes as opposed to ‘sniper’ SDM approach. For example, based on 84 serine proteases of the chymotrypsin family, the number of critical

residues making contact with P1-P4 residues of the substrate is predicted to be eleven (Rose and Di Cera, 2002). If all these residues are chosen for randomization with 12 chemically distinct codons (Reetz et al., 2008) simultaneously, the theoretical size of the library is $11^{12} = 3 \times 10^{12}$. It is clear that screening every variant of this library is not feasible, and for any new desired functionality it is desirable to maximize the number of variants screened. This is a particularly significant consideration in the context of engineering new activities and specificities that are not inherently detectable in the parent enzyme, i.e. the evolving something out of nothing paradigm (Dalby, 2011; Khersonsky and Tawfik, 2010) since the desired variants are expected to be sparsely populated in the sequence space of the parent enzyme (Weinreich et al., 2006). Low throughput of agar- or well plate-based assays will strongly limit the diversity screened and thus the chances of identifying winners.

Fluorescence-activated cell sorting (FACS) is a powerful tool that can screen thousands of clones per second, provided a protein's function of interest can be coupled to and correlated with cell fluorescence. Several previous studies have successfully screened diverse protease libraries using innovative experimental designs to fit active protease, corresponding genotype and fluorescent sensor of specificity in a micron-sized compartment that can be sorted using FACS (Smith et al., 2015; van Vliet et al., 2015). One of the pioneering studies demonstrated the ability to systematically modify the substrate specificity of *E. coli* outer membrane protease T (OmpT), without loss in catalytic efficiency, by high-throughput screening (HTS) of a library with 10^8 unique members, containing mutations randomly distributed across the entire active site comprised of 21 residues. Successful engineering of OmpT was facilitated by several

factors: natural localization on the outer membrane providing access to bulky FRET peptide substrates and minimal host toxicity upon overexpression, stable β -barrel structure allowing variants with high mutational load to fold properly and expression in a constitutively active form without any requirement for zymogen activation. Despite this being a significant milestone in protease engineering, it is not clear whether these results are generalizable, particularly in the context of human proteases. To date, no substrate binding pocket of mammalian protease has been comprehensively engineered, primarily due to the paucity of high-throughput screening platforms for screening targeted libraries.

In the present study, we have established a comprehensive methodology that allows efficient FACS screening of large libraries of recombinant mammalian proteases in *E.coli* by integrating three different components: (i) surface display system compatible with proteases that need zymogen activation, (ii) protease activity assay with single cell resolution, and (iii) genotype recovery method to facilitate iterative enrichment of desired phenotype. We have displayed two well-studied proteases, chymotrypsin and caspase-3, by fusion to a native membrane protein, Antigen 43, and characterized their functional activity and specificity. A multiplexed assay previously developed to measure protease specificity using FRET peptide substrates was adapted to the chymotrypsin system. We optimized in a quantitative manner, the labeling parameters such as cell density and net charge of the peptide substrate to improve the dynamic range of the assay. Despite displaying chymotrypsin on the surface, *E. coli* cells showed poor viability due to the proteolytic nature of the enzyme. In order to overcome this limitation, we developed a method to directly recover plasmid DNA from clones sorted by FACS and quantitatively compared it with the more standard method of re-culturing sorted cells. We evaluated

both methods by measuring the degree of retention of the genetic diversity and by quantifying the yield of the sub-populations expressing the desired phenotype.

3.2 Materials and Methods

3.2.1 Molecular cloning

The DNA construct coding for a fusion protein with the following domains: signal peptide of Antigen 43, small and large subunits of human caspase-3 and FLAG tag was obtained by overlap extension PCR (OE-PCR). Template to obtain caspase-3 subunit gene fragments was kindly provided by Georgiou lab (University of Texas, Austin). *Rattus norvegicus* chymotrypsin B used in this study is constitutively active and does not contain the chain A. Using pBAD_Achy_700 as template, Cys122Ser and Tyr164Ala mutations were introduced by OE-PCR to obtain a modified version of wild-type chymotrypsin (rChyB). Catalytic serine (position 195) was mutated to alanine by OE-PCR to create an inactive chymotrypsin variant. All these genes were cloned into pBAD_700 vector as described previously (Ramesh et al., 2012).

3.2.2 Crude membrane preparation

Three mL LB cultures were inoculated to a starting OD₆₀₀ of 0.02 using overnight cultures and grown up to an OD₆₀₀ of 0.5 for induction with 0.2% (w/v) arabinose at 37 °C for 2 h. Cells expressing protein (3 X 3 mL cultures) were spun down, resuspended in 750 µL of lysis buffer (50 mM Tris, pH 7.5), sonicated (3 cycles of 30s on/off/on, XL-2000, QSonica, New Town, CT) and centrifuged (170,000 x g, 4 °C, 1 h) to pellet the membrane fraction. The pellet was incubated with solubilization buffer (50 mM Tris, pH 7.5, 10 mM CaCl₂, 100 mM NaCl, and 1% n-octylglucoside) overnight at 4 °C and 500

rpm. The mixture was centrifuged (13,000 x g, 1 min) to remove insoluble fraction and the supernatant was filtered using 0.45 µm filter plate (EMD Millipore, MA).

3.2.3 Enzyme characterization using HPLC

For chymotrypsin experiments, membrane protein fraction was incubated with 10 µM Y-BQ7 substrate at 37 °C for 1 h. For caspase-3 experiments, cells expressing caspase-3 were incubated with 20 µM DEVD-AFC substrate (Santa cruz Biotechnology Inc., TX) at 37 °C for 1 h. Purified α-chymotrypsin (Sigma-Aldrich, MO) and caspase-3 (BD Biosciences, CA) were used as positive controls. Digested samples were analyzed on a C18 column (Cat. No. 00G-4041-E0, Phenomenex, Torrance, CA) by measuring absorbance in HPLC (Shimadzu scientific, Kyoto, Japan) using the following gradient method with water/1% acetic acid (mobile phase A) and acetonitrile/1% acetic acid (mobile phase B): 95% A/5% B for 1 min, linear increase to 40% A/60% B for 30 min, ramp up to 5% A/95% B in 2 min, 5% A/95% B for 5 min, ramp down to 95% A/5% B in 2 min and 95% A/5% B for 5 min.

3.2.4 Substrate synthesis

Lyophilized peptides (>70% purity, N-terminal acetylated and C-terminal amidated) were purchased from Genscript (Piscataway, NJ). Atto-633 maleimide (Atto-TEC GmbH, Sigen, Germany) was conjugated to the peptide and subsequently purified using HPLC. Electrospray ionization mass spectrometry (Shared equipment authority, Rice University, Houston, TX) was used to verify chemical identity of synthesized substrates. Chymotrypsin substrate Y-BQ7, CAAPYGSKGRGR peptide conjugated to BODIPY-

FL-SE and QSY7 C5 maleimide dyes, was a generous gift from Georgiou lab, University of Texas, Austin.

3.2.5 Plasmid recovery from sorted cells

Three different lysis conditions were evaluated for the cells expressing chymotrypsin B: i) 1X Bugbuster (EMD Millipore, MA) with 1 mg/mL lysozyme in 50 mM Tris-HCl pH 7.5 (buffer 1); ii) 100 mM EDTA and 1% Triton-X in 50 mM Tris-HCl pH 7.5 (buffer 2); and iii) 1X lysis buffer (buffer 3) for plasmid miniprep (Zymo research, CA). Typical composition of buffer 3 is 0.1 M NaOH, 0.7% SDS, 0.7 mM EDTA, and 1% isopropanol (Jia, 2011). For evaluating buffers 1 and 2, *E. coli* MC1061 cells ($N_0 = 50,000$ in ~ 200 μ L) harboring plasmid pBAD_AChy_700 (p15A origin, 15-20 copies/cell) were sorted based on a gate in FSC-SSC plot and incubated with equal volume of 2X lysis at 25 °C for 15 min (Chang and Cohen, 1978; Ramesh et al., 2012). Lysates were purified using Zymo DNA clean and concentrator kit (Zymo research, CA). For buffer 3, an equal number of cells were sorted and the protocol prescribed for Zyppy plasmid miniprep kit (Zymo research) was followed using the DNA binding columns (Cat. No. D4004, Zymo research, CA). Each sample was eluted with 10 μ L of Zyppy elution buffer, and 2 μ L was used to transform electrocompetent *E. coli* MC1061 cells (Varadarajan et al., 2009a) (transformation efficiency $> 10^9$ cfu/ μ g), and the total number of transformants was estimated by plating transformation dilutions onto LB-agar plates supplemented with chloramphenicol (LB-Cm plates).

3.2.6 Phenotype enrichment by FACS sorting

Cells expressing inactive chymotrypsin were mixed with 1 % cells expressing wild-type (WT) chymotrypsin (ratio determined by OD₆₀₀) and incubated with 20 nM Y-BQ7 for 15 min at 25 °C. After the optics alignment and drop delay were calibrated using 3 µm UltraRainbow fluorescent particles (Spherotech Inc., IL) and Accudrop beads (BD Biosciences, CA), the labeled bacterial population was sorted at an event rate of 7000/s using 'pure' as the sort mode to isolate cells showing highest (top 1%) fluorescence in the 530/40 nm channel. Known number of cells in the sort gate were collected in 100 µl of 2xYT media and plated onto LB-Cm plates to estimate viability. For the plasmid recovery method, 50,000 cells in the sort gate were collected into an empty microcentrifuge tube and DNA isolated using the buffer 3 protocol. Subsequent to transformation into *E. coli* MC1061 cells, the total number of transformants was estimated by plating transformation dilutions onto LB-Cm plates. Transformants recovered with 1 mL of SOC media (plasmid recovery method) or 50,000 sorted cells collected in 250 µl of 2xYT (culturing method) were directly grown in 100 mL of 2xYT media supplemented with 0.5% glucose and 25 µg/mL chloramphenicol at 37 °C for 10 h, and used to seed a subculture. After 2 hours of induction with 0.2% arabinose, cells were labeled with Y-BQ7 and analyzed by flow cytometry. The presence of full-length chymotrypsin variants on the bacterial surface was characterized by incubating induced cells at with 40 nM anti-FLAG - Phycoerythrin (ProZyme Inc., CA) for 30 min at 25 °C.

3.3 Results

3.3.1 Functional characterization of proteases displayed on the bacterial surface

Caspase-3 is a cysteine protease belonging to the caspase cascade involved in cellular apoptosis (Lamkanfi et al., 2007). It is typically expressed as a zymogen to be activated by caspase-8/9 by generation of a small subunit, derived from the C-terminal region, and a large subunit, derived from the N-terminal region. An engineered recombinant human version of caspase-3 with reversed ordering of the subunits was previously designed to produce a zymogen that can auto-activate to form the active enzyme (Srinivasula et al., 1998). We genetically fused the N-terminal of auto-activable caspase-3 to Ag43 signal peptide and C-terminal to β -barrel domain of Ag43 to display the enzyme on the surface of *E.coli* MC1061 cells. To test the functional activity of caspase-3, cells displaying the enzyme were incubated with its preferred peptide substrate DEVD conjugated to the amine group of 7-amino-4-trifluoromethylcoumarin (AFC). When the digested peptide substrate was analyzed in a C18 column, free hydrophobic dye released by proteolysis eluted at higher retention time (RT = 31 min, 78% AcN) unlike the unreacted peptide substrate containing polar Asp residues which eluted at RT = 26 min, 54% AcN (**Figure 3.1A**). No proteolysis was observed when the substrate was incubated with cells that did not contain any plasmid showing that host proteases do not cleave the substrate. Chromatogram peak profile of substrate incubated with cells expressing caspase-3 was identical to that of commercially purchased caspase-3 sample, thereby confirming its functional activity and substrate specificity.

To ensure that the wild-type chymotrypsin displayed on the cell surface is active, a peptide substrate (Y-BQ7) which contains the sequence AAPYGS sandwiched between

the fluorophore Bodipy-FL and the quencher QSY7 was used. The membrane fraction of cells displaying wild-type chymotrypsin was purified by ultracentrifugation and incubated with 10 μ M of Y-BQ7. The membrane purification step was necessary because we observed proteolysis of Y-BQ7 even with cells containing no plasmid, which was possibly due to proteases on the outer membrane.

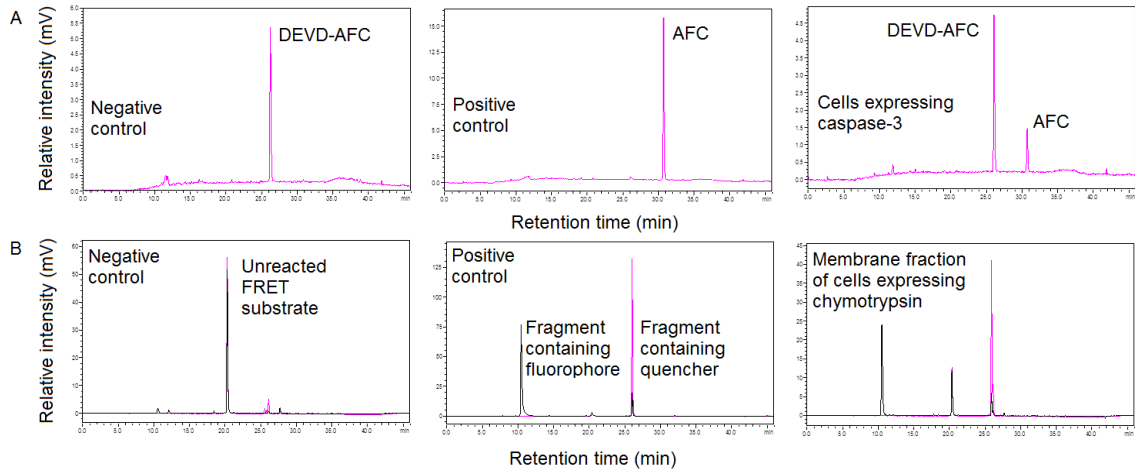


Figure 3.1 HPLC analysis of caspase-3 (A) and chymotrypsin (B) incubated with peptide substrates.

The reaction mixture was analyzed using HPLC by measuring absorbance at 503 nm (Bodipy-FL) and 560 nm (QSY7). No proteolysis in the linker between the FRET pair was observed when the substrate was incubated with the membrane fraction of cells that did not express chymotrypsin, since 503 nm and 560 nm peaks eluted at the same retention time (20 min, 42% AcN) (Figure 3.1B). The chromatogram peak profile of the membrane fraction of cells expressing chymotrypsin was identical to that of bovine chymotrypsin sample with fragment 1 containing Bodipy-FL and arginine tail (RT = 12 min, 28% AcN) and fragment 2 (RT = 26 min, 54% AcN) containing the quencher. Identities of these fragments were also verified by LC-MS. Expected masses of fragments 1 (1387.5 Da) and 2 (989.5 Da) match the abundant m/z peaks observed.

3.3.2 Protease activity assay improved by optimizing pH, cell density and charge of the FRET peptide substrate

For HTS of protein libraries, it is essential to maintain the genotype-phenotype connection to identify the mutations responsible for any desired phenotype observed. In the case of proteases, when substrates are exogenously added, proteolysis products can diffuse away after catalysis leading to loss of information about which member of the population was responsible for the reaction. To address this issue, in a previous study (Olsen et al., 2000), a peptide substrate was designed to contain a protease-sensitive linker sandwiched between a FRET pair and a positively charged tail such that after proteolysis, the fragment containing the fluorophore is locally captured on the surface of the bacterial cell which is typically negatively charged. Each bacterial cell encodes for an individual member of the protease library and only those cells which contain variants catalytically active towards the peptide substrate will become fluorescent while rest of the population does not.

As localized capture of the fluorescent fragment on the surface of the cell is driven by electrostatic interaction between positively charged tail of the peptide and negative charge of the bacterial surface, labeling can be done only in buffers with low ionic strength. In the previous study, libraries of ompT, an asparaginyl protease ($pK_a = 5.6$) (Kramer et al., 2000), was screened by labeling in water (18 M Ω .cm resistivity) containing 30 mM sucrose. To engineer serine proteases, it is essential to keep the His in the catalytic triad in a deprotonated state ($pK_a = 6.8$) for conserving the catalytic activity. After trying different bases including guanidine and NaOH to increase the pH of the labeling buffer without increasing the ionic strength, 2 mM Tris was found to be

satisfactory as it raised the pH to 7.5 and the labeling with Y-BQ7 electrostatic capture of peptide substrate increased 10-fold (**Figure 3.2A**). pH measurements were performed using pH paper (colorpHast pH 5 -10, EMD Millipore, MA) as electrodes cannot be used with buffers of low ionic strength. In the previous study (Varadarajan et al., 2005), cells at OD₆₀₀ 0.1 were labeled with 20 nM of peptide substrate which is 2 orders of magnitude lower than K_m of OmpT or any typical enzyme. Fluorescence generated by proteolysis was still detected in FACS possibly due to the partitioning of positively charged substrate molecules leading to a higher substrate concentration near the negatively charged cell surfaces (zeta potential $\sim$ -30 mV) (Razatos et al., 1998) than the bulk concentration and sensitive detection by photomultiplier tubes (PMTs).

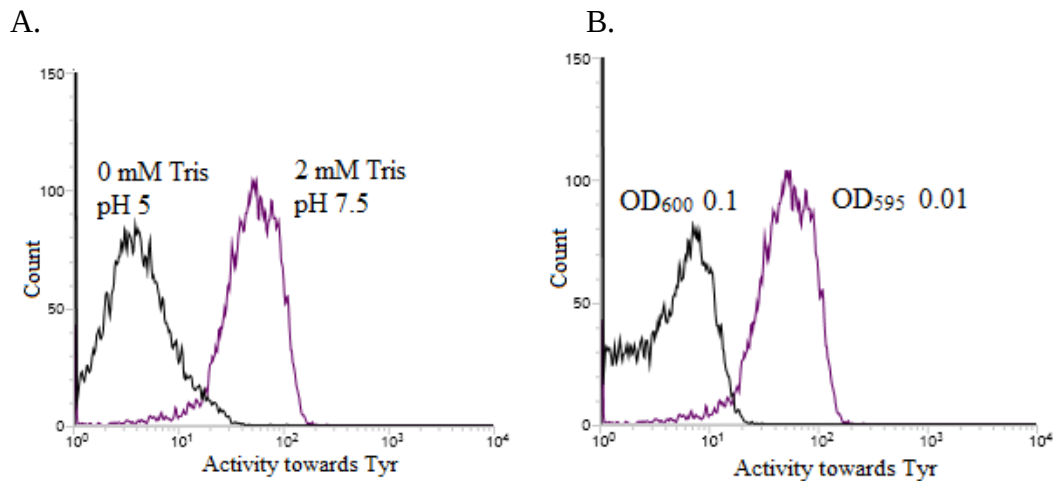


Figure 3.2 Effect of pH (**A**) and cell density (**B**) on labeling of *E. coli* MC1061 cells expressing rChyB with Y-BQ7 substrate in 1% sucrose solution.

We optimized the cell density during labeling to improve the detection of positive fluorescent signal. If each cell contains 10^5 enzyme molecules, then substrate: enzyme concentrations ratio starts at $\sim$ 1:1 and decreases further as the reaction progresses. When the density of cells expressing chymotrypsin was decreased during the labeling with Y-BQ7 by a factor of 10, there was a corresponding increase in geometric mean

fluorescence by a factor of 10 as number of substrate molecules available per cell had increased (**Figure 3.2B**).

We then sought to estimate the effect of peptide charge on the degree of substrate labeling. It is to be noted that the peptide substrate used did not contain the quencher and hence, the fluorescence observed is only dependent on the concentration of peptide substrate bound to the bacterial cell surface. We observed that, when incubated with peptides conjugated to fluorophore Atto-633, *E.coli* MC1061 cells expressing no recombinant enzyme showed increased labeling when the peptide had increased charge (**Figure 3.3**). Of note, the net charge of the peptide substrate was determined by the pKa of the side chains of constituent amino acids and accounting for the fluorophore which has a +1 charge at pH 7.5. Applying aforementioned FRET peptide strategy to screen libraries of caspase-3 could be challenging as multiple carboxylic acid groups in the preferred substrate sequence hinders favorable electrostatic interactions with negatively charged bacterial surface. It remains to be seen whether addition of arginines to the N-terminal of the peptide could improve surface concentration of caspase-3 substrate.

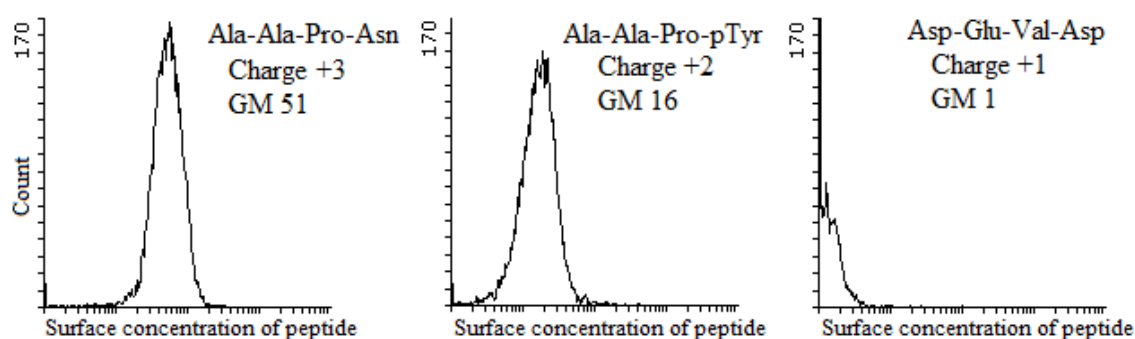


Figure 3.3 Concentration of the peptide bound to the surface of *E.coli* MC1061 cells decreases with the charge of the peptide-fluorophore conjugate. pTyr refers to phosphorylated Tyr.

3.3.3 Phenotype enrichment by direct plasmid recovery after FACS

Screening of libraries for proteins with improved functional traits is typically accomplished by expression in a suitable microbial host and by iterative sorting, pooling and culturing of the sorted clones (19, 20). Unfortunately, enhancements in the protein's function of interest (*e.g.* enzymatic activity) and/or cell fluorescence (*e.g.* fluorescent protein expression) compromise cell viability or growth (21, 22). Recently, a methodology to directly recover high-copy plasmid (1000 copies/cell) from small numbers of cells (10^3 - 10^5) has been reported (23), but its utility after FACS especially for low- and medium-copy plasmids is unknown. To develop a direct plasmid recovery method that can be used for screening protease libraries iteratively, we compared three different lysis conditions to release plasmid DNA from sorted cells by estimating the recovery efficiency, defined as the ratio of total number of transformants obtained to total number of sorted cells used for plasmid recovery. Alkaline lysis yielded highest recovery efficiency (buffer 3, 44 ± 24 %) in comparison to BugbusterTM (16 ± 13 %) or Tris/EDTA/Triton-X (4 ± 2 %). Recovery efficiency can also be improved by decreasing the volume of eluted plasmid prep to 1 μ L per transformation and using freshly prepared electrocompetent cells with $OD_{600} > 200$. While custom-built microfluidic devices allow harvesting 10^5 sorted droplets into a tiny volume of 1-2 μ L, a typical FACS instrument with a 100 μ m nozzle can lead to greater than 300 μ L volumes, presenting the requirement to concentrate sorted cells prior to genotype recovery. Centrifugation followed by removal of supernatant was used for volume reduction when N_{sort} was more than 10^5 , as this approach yielded 4-fold higher recovery efficiency as compared to using lyophilization for the same purpose.

A population of cells expressing inactive chymotrypsin variant was doped with 1% wild-type chymotrypsin (as determined by OD₆₀₀) to mimic a protease library with low frequency of desired variants. Thousand cells displaying highest activity (top 1%) towards Y-BQ7 substrate were collected in 100 µL of 2xYT media and plated onto LB-Cm plates to count the number of colony forming units to be less than 50. Poor viability of sorted cells ($4 \pm 1\%$) is possibly due to both proteolytic activity of chymotrypsin and labeling with peptide substrate in 1% sucrose.

To assess the utility of the plasmid recovery method in recovering the genotype of sorted cells with poor viability, plasmid DNA from 50,000 cells in the sort gate was isolated using the buffer 3 protocol. Total number of transformants obtained was estimated to be 31,500 by plating dilutions on LB-Cm plates. To compare the yield of active chymotrypsin when performing direct plasmid recovery from sorted cells versus post-sort culturing, recovered transformants or sorted cells, both directly grown in 2xYT media supplemented with 0.5% glucose and 25 µg/mL chloramphenicol at 37 °C for 10 h were used to inoculate subcultures. After 2 h of induction with 0.2% arabinose, cells were labeled with Chy-BQ7 and analyzed by flow cytometry. The frequency of the subpopulation displaying chymotrypsin activity increased by 80-fold in comparison to the parent population when the plasmid recovery method was employed (**Figure 3.4**). In contrast, the post-sort culturing method induced the loss of a majority of the cells expressing full-length chymotrypsin variants after just one round of sorting (**Figure 3.4B**).

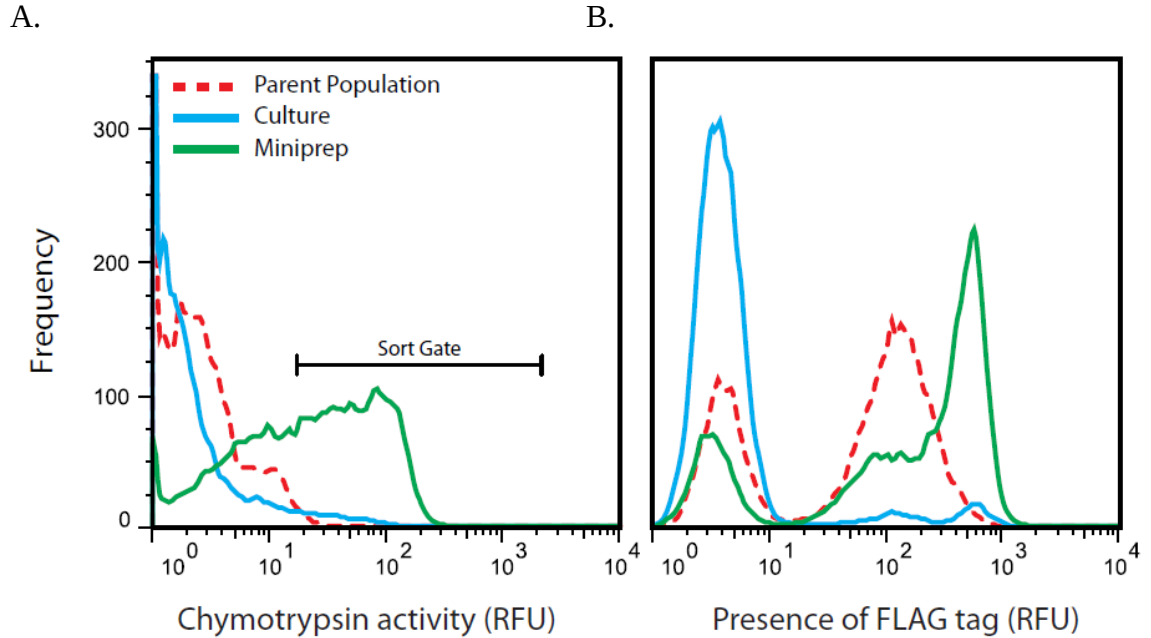


Figure 3.4 Comparison of phenotype enrichment by plasmid recovery and culturing methods in chymotrypsin expression system using (A) peptide substrate and (B) anti-FLAG tag antibody.

3.4 Discussion

A platform to screen libraries of protease variants at high-throughput is highly desirable as the prediction of a phenotype *a priori* based on its primary sequence can be challenging. Though proteases are typically expressed in their stable, inactive zymogen form, it is possible to engineer them to be auto-activable or be constitutively active (Pozzi et al., 2013; Srinivasula et al., 1998). Conversion of zymogen into active form, particularly in serine proteases such as factor C (Kobayashi et al., 2014) and chymotrypsin (Berg et al., 2002), involves cleavage of propeptide in the N-terminal region of the enzyme such that newly generated N-terminus leads to structural changes such as formation of substrate-binding site. Our autotransporter-mediated platform enables display of recombinant protein through C-terminal fusion such that N-terminal of

the protease is available to participate in the necessary interactions, ultimately leading to a functional enzyme that can catalyze proteolysis.

Several assays were described to screen protease libraries using FACS to isolate variants with increased kinetics (Kostallas and Samuelson, 2010; Wu et al., 2013; Yoo et al., 2012). However in most cases, only the high-throughput feature of FACS is taken advantage of to funnel down the size of libraries to numbers that can be handled by plate based methods without regard to the dynamic range (ratio of largest and smallest possible fluorescence outputs when protease variants with differing k_{cat}/K_m values are present). When the dynamic range is infinitesimally small, the assay output would be on/off and the signal associated with all the protease variants displaying an activity below the chosen threshold value would be indistinguishable from noise. Protein engineering typically progresses through iterative mutagenesis and identifying hits with subtle improvement in desired phenotype during HTS is crucial to avoid stalling in the directed evolution process. For example, it would be advantageous to tune the dynamic range of the assay such that variants showing even weak amplitude of desirable activity would generate high signal during initial rounds of sorting. We showed that by decreasing the cell density and increasing the pH when labeling with the peptide substrate, fluorescence generated by wild-type protease can be increased 10-fold. Analogous to adjusting voltage gain of PMT detectors in FACS, it appears that we can tune labeling conditions depending on characteristics of the population being sorted and the variants desired.

Methods alternate to direct culturing of sorted cells are required for iterative enrichment of phenotype when sorted cells are unviable due to toxicity associated with protease expression or assay reagents. Furthermore, some whole-cell protein assays

require that the cells are first permeabilized to provide access to exogenously added substrates and thus rendered unviable. Apart from the subsequent re-cloning being time-consuming and laborious, “rescue” PCR when used to amplify the corresponding genes of the sorted cells (29), has potential to alter the frequency distribution of DNA sequences by amplification bias and homologous recombination (30). As our genotype recovery method involves direct isolation of plasmids from sorted cells, it obviates the need for re-cloning and transformation into fresh cells after each round of sorting helps overcome plasmid instability observed in aging cells and chromosomal mutations that might affect the phenotype observed. When the plasmid recovery method was applied in a disparate AraC biosensor system (Tang et al., 2013), similar phenotype enrichment was observed. Surprisingly, poor recovery efficiency ($< 5\%$) was achieved when plasmid DNA isolated from cells displaying human IL7 was transformed although alkaline lysis and DNA binding/elution from resin is not expected to be independent on nature of the DNA sequence. When iterative sorting of a protein library is performed, total number of transformants carried forward to the next round of sorting should be at least three fold the theoretical degeneracy of the sorted population for retaining any given unique variant with 95% confidence level (Bosley and Ostermeier, 2005). Hence, when the top 1% of a library with 10 million variants is sorted, at least 300,000 transformants should be carried forward to the next round of sorting. As the recovery efficiency is $\sim 60\%$, 500,000 sorted cells should be collected for plasmid DNA isolation and at an event rate of 7000/s, a library with 10 million members can be screened in 2 hours.

3.5 Conclusions

We have established a pipeline for FACS screening of recombinant protease libraries expressed in a bacterial host that offers high transformation efficiency. Autotransporter-mediated surface display system previously used for scFv screening was extended to chymotrypsin B and caspase-3 and the ability to cleave peptide substrates when anchored on the cell surface was characterized. We showed that the dynamic range of the protease activity assay can be tuned using labeling condition such as pH, cell density and net charge of the peptide. A technique for the direct isolation of plasmid DNA with medium-to-high copy-number from whole-cells sorted using FACS and subsequent transformation into competent cells with recovery efficiency sufficiently high for complete screening of large, diverse libraries iteratively was developed. This robust method for phenotype enrichment is well suited as an alternative to culturing sorted cells or genotype recovery by PCR especially when protein expression or screening protocol induces host cell toxicity.

Chapter 4

Engineering substrate specificity of chymotrypsin for mapping N-glycosylation sites

4.1 Introduction

The diversity of proteins expands beyond the limits of the genetic code due to a wide repertoire of post-translation modifications (PTMs). The biological significance of PTMs has been demonstrated using classic examples such as gene regulation by lysine acetylation/deacetylation of histones (Strahl and Allis, 2000). Therefore large scale PTM profiling efforts have been undertaken with the objective of improving understanding of cell physiology and discovering novel targets for drug development (Horiya et al., 2014; Lundby et al., 2013). Also, analytical characterization of PTMs is required for the quality control of industrial production of therapeutic biologics (Berkowitz et al., 2012). Mass-spectrometry (MS) based methods have emerged as the preferred approach for identifying new PTM sites and for quantitative comparison of different samples. Typically such an MS experiment involves digestion of protein samples into peptides using trypsin and subsequent depletion of peptides that do not contain any PTM as modified peptides are present in low abundance. Enriched peptides are analyzed by LC-MS/MS (liquid chromatography coupled to tandem mass spectrometry) where a precursor ion selected from the first MS is further fragmented into product ions for analysis by the second MS. Peptide sequences corresponding to MS/MS spectra are identified using

bioinformatics tools and the mass increment due to PTM in fragment ions is used to identify the modified residue in the peptide. Presence of multiple PTMs in a single peptide and loss during fragmentation due to labile nature of PTMs such as Ser/Thr phosphorylation and O-linked glycosylation makes precise mapping of sites of modification difficult (Olsen and Mann, 2013). One of the strategies that have been implemented to overcome this limitation was chemical modification of phosphorylated Ser and Thr into Lys analogs for phosphospecific proteolysis using Lys-C protease (Knight et al., 2003). As sample loss in chemical derivatization methods was high due to multiple washing/desalting steps (Zhao and Jensen, 2009), alternate methods that use enzymes which can discriminate between modified and unmodified amino acids were sought after. For example, the inability of trypsin to cleave after Arg which is modified to citrulline has been used to identify citrullinated proteins in the synovial fluid of rheumatoid arthritis patients (Kasper B. Lauridsen, 2013). There are 400 different PTMs identified so far and enzymes for detecting most of them are not available in nature. Therefore protease engineering has been employed to incorporate selectivity towards specific PTM such that position of an amino acid in the primary sequence of the peptide can yield information on the presence of PTM independent of its detection in MS. Engineered proteases can be readily integrated into the existing bottom-up proteomics workflow by performing tandem digestion with trypsin. Detection of a variety of PTMs can be multiplexed by using a cocktail of engineered enzymes. Though this strategy seems attractive, proteases engineered so far were limited either by poor selectivity towards desired PTM or absence of robust digestion of different protein substrates. For example, an engineered subtilisin variant distinguished phosphorylated Tyr from naked

Tyr by 5-fold and from sulfated Tyr by only 2-fold (Knight et al., 2007). Poor discrimination between modified and unmodified residues could lead to false annotation of PTM. It is to be noted that protease specificity can only be used as an additional criteria during bioinformatics analysis of MS/MS spectra to assign peptide identity but cannot be relied on as the sole determinant to annotate PTMs as even highly specific trypsin is known to exhibit at least 20% missed cleavages (Burkhart et al., 2012). In another system, *E.coli* outer membrane protease T was engineered using directed evolution approach to display 3600-fold higher activity towards peptide substrates containing 3-nitrotyrosine residue in comparison to sulfated Tyr (Varadarajan et al., 2009b). However the proteolytic activity of these highly selective variants was restricted to peptide substrates and they did not digest proteins containing putative cleavage sites (Yoo et al., 2012).

In this study, we have engineered the specificity of chymotrypsin to cleave after asparagine by high-throughput screening of large libraries ($> 10^7$ variants) targeting selected residues in the substrate binding pocket. The engineered variant containing eight different mutations identified by iterative mutagenesis showed 10-fold higher activity towards P1 Asn than the wild-type chymotrypsin. We demonstrated its ability to generate peptides of desired size by digestion of full length protein substrates for LC-MS/MS analysis using the model protein bovine serum albumin. Our method can be used to directly detect asparagines that are modified with carbohydrates without loss of information on non-glycosylated peptides. The high-throughput screening platform used in this study can be readily extended to engineer variants for detection of other PTMs.

4.2 Material and Methods

4.2.1 Construction of chymotrypsin libraries

For construction of the site-saturated library of chymotrypsin B (rChyB), residues 189, 192, 216-219, 221-224 and 226 were targeted for randomization. PCR was performed using primers **1** and **2** (**Table 4.1**) and pBAD_AChy_700 (Ramesh et al., 2012) as template to obtain fragment **1** coding for 206 – 245 aa of chymotrypsin B with a FLAG tag and a KpnI site with overhang for efficient restriction digestion. After gel purification, fragment **1** was used as template in a PCR reaction with primers **2** and **3** such that region coding for 193-206 aa is added to its 5'. The above step was repeated using primers **2** and **4** to obtain fragment **2** coding for 183 -245 aa with all the desired positions randomized with NNS codon. Fragment **3** containing SacI restriction site with an additional overhang, Shine-Dalgarno sequence and gene coding for Ag43 signal peptide and 16-188 aa of chymotrypsin B was obtained using pBAD_AChy_700 as template and primers **5** and **6**. 100 fmols of gel-purified fragments **2** and **3** were first assembled together by 10 cycles of PCR and then amplified by primers **2** and **5** for the next 20 cycles. All primers were purchased from Integrated DNA Technologies (IDT), IA and are listed in **Table 4.1**. Even though primers **1** and **3** were longer than 80 bases, they were not HPLC or PAGE purified anticipating loss in genetic diversity. Assembled gene library was gel-purified, digested with SacI-HF and KpnI-HF at 37 °C for 1 h and ligated into digested pBAD_700. Enzymes for molecular cloning were obtained from New England Biolabs (Ipswich, MA). Ligated plasmid was dialyzed with water for 1 h and transformed into *E. coli* MC1061 cells by electroporation. Transformants recovered with SOC media ($\sim 4 \times 10^7$) were directly grown in 100 mL of 2xYT media supplemented

with 0.5% glucose and 25 µg/mL chloramphenicol at 37 °C for 10 h. 10 mL of cells were lysed and their plasmid DNA isolated using QIAGEN miniprep kit was stored at -20 °C. 300 µl aliquots of cells (OD₆₀₀ ~2) in 20% glycerol were frozen using liquid nitrogen and stored at -80 °C.

Table 4.1 List of primers used in this study. Codons containing mixed bases for randomization are highlighted in red. Notation used for mixed bases are: n - a/c/g/t, s - c/g, w - a/t, y - c/t, k - g/t, m - a/c, and r - a/g.

S. No.	Primer	Sequence
1	chy_216-219_221-224_226_NNS_f	gtctggaccctggcaggcattgtgtctctgg nnsnnsnnsnns tg tnnnsnnsnns cct nns gtgtattcccg agtcacagcc
2	chy_FLAG_KpnI_r	atcgggggtacctttgtcatcgtcatcttataatcgttggtccaagatctgtgatcgggtacctttgtcatcgtcatc
3	chy_193-219_f	ggtgactccgggtggcccccctcgtctgccagaagaatggagtgctggaccctggcaggcattgtgtctctgg
4	chy_M192_S189_NNS_f	gcaggcgctagcgggtgtc nns tcctgc nns ggtgactccgggtggccc
5	SacI_Ag43 leader-f	agccc gagctc tctaaggaaaagctgatgaacgacatctgaatacct
6	chy_182-188_r	gacaccgctagcgctgcgca
7	chy_M192_S189_wt/v1_f	gcaggcgctagcggcgtc wykt ctctgc mkg gggtgactccgggtggccc
8	chy_216-219_221-224_226_wt/v1_f	gtctggaccctggcaggcattgtgtctctgggg crktsgggwg tgt wccrscwccws gc ctgsc gtgtattccga gtcacagcc
9	Ag43 leader- rChyB-f	ctcccgggtgctggctctatcgtcaacggagaggatgc
10	Ag43 leader-r	agccagcaccgggagtgac

To create a library where positions 189, 192, 217-219, 221, 222, 224 and 226 were to contain the codon corresponding to rChyB or rChyB-N-v1 sequence, primers **7** and **8** with mixed bases were used. Procedure described earlier was repeated with primers **1** and **4** replaced by **8** and **7** respectively. It is to be noted that library members contained amino acids other than those which correspond to rChyB or rChyB-N-v1 in positions where more than one mixed base was used for randomization (**Table 4.2**). For random mutagenesis of rChyB-N-v2, error-prone PCR was performed using GeneMorph II kit (Agilent Technologies, CA) with 2 fmols of pBAD_700 containing *rChyB-N-v2* gene as template and primers **2** and **9**. Assembly with the gene fragment coding for Ag43 signal peptide obtained using primers **5** and **10**, cloning into digested pBAD_700, transformation and cell recovery were performed as described earlier. For all the libraries,

plasmid DNA isolated from 10 randomly picked clones were sequenced (SeqWright Genomic Services, TX) to assess the genetic diversity and mutation rate for the library created using error-prone PCR was estimated to be 0.65%.

Table 4.2 Library construction for shuffling engineered variant, rChyB-N-v1, with wild-type rChyB.

Position	189		192		217		218		219		221		222		223		224		226	
rChyB	tct	Ser	atg	Met	agt	Ser	ggc	Gly	gtc	Val	tcc	Ser	act	Thr	tcc	Ser	act	Thr	gct	Ala
rChyB-N-v1	atg	Met	cgg	Arg	gtg	Val	cgg	Arg	gag	Glu	acc	Thr	ggc	Gly	acc	Thr	tgg	Trp	ggc	Gly
Mixed codon	wyk		mkg		rkt		sgg		gwg		wcc		rsc		wcc		wsg		gsc	
Possible codons/ amino acids	tcg	Ser	agg	Arg	agt	Ser	ggg	Gly	gtg	Val	tcc	Ser	acc	Thr	tcc	Ser	acg	Thr	gct	Ala
	tct	Ser	cgg	Arg	att	Val	cgg	Arg	gag	Glu	acc	Thr	ggc	Gly	acc	Thr	tgg	Trp	ggg	Gly
	atg	Met	atg	Met	ggg	Gly							agc	Ser			agg	Arg		
	ttg	Leu	ctg	Leu	ggt	Val							gcc	Ala			tcg	Ser		
	ttt	Phe																		
	acg	Thr																		
	act	Thr																		
	att	Ile																		

4.2.2 Synthesis of FRET peptide substrates

Lyophilized peptide (KAAPNGSCGRGR, N-terminal acetylated and C-terminal amidated, >70% purity, Genscript, NJ) and dyes, Atto-633 maleimide (Atto-TEC GmbH, Sigen, Germany) and QSY21 carboxylic acid succinimidyl ester (Life technologies, NY), were resuspended in anhydrous *N,N*-dimethylformamide (DMF) to a concentration of 10 mM and 10 µg/µL, respectively. Pilot reactions were carried out to optimize the HPLC gradient method for proper separation of peptide species conjugated to both the dyes (desired) from the peptide molecules conjugated to either one of the dyes (undesired). A schematic describing the chemistry of conjugation of FRET dyes to the peptide is shown in Figure 4.1. 20 µL of 1 M NaHCO₃ solution was mixed with 50 µL of peptide solution. Excess salt was pelleted to transfer the supernatant to a fresh tube and 50 µL of Atto-633 maleimide was added to it. After incubation at 25 °C for 1 h, an aliquot was analyzed on

a C18 column (Cat. No. 00G-4041-E0, Phenomenex, CA) with water/acetonitrile mobile phases containing 1% acetic acid at a flow rate of 1 mL/min by measuring absorbance at wavelengths 630 and 670 nm in HPLC (Shimadzu scientific, Kyoto, Japan). For the second reaction, peptide/Atto-633 reaction mixture was added to a tube containing 100 μ L of 1M 4-dimethyl amino pyridine (DMAP, Sigma Aldrich, MO) and 50 μ L of QSY 21 solution.

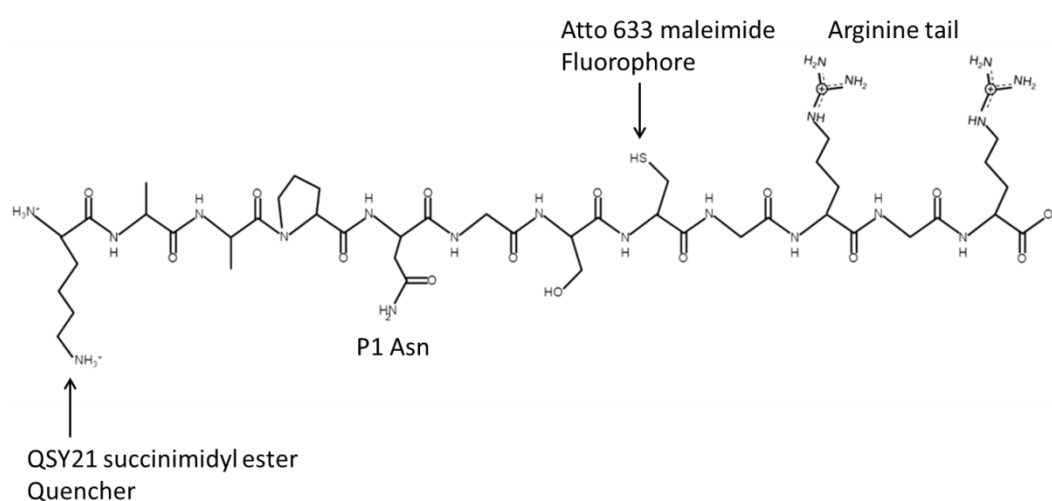


Figure 4.1 Chemical synthesis of N-AtQ21, Asn FRET peptide substrate.

To estimate the yield of desired species after 1 h incubation at 25 °C, an aliquot was analyzed using the following method: 5% acetonitrile for 5 min, increasing to 28% in 2 min and maintained for 5 min (peptide conjugated to only one of the dyes eluted in this step), increased to 35% in 2 min and maintained for 4 min (peptide conjugated to both the dyes, N-AtQ21, eluted in this step), increased to 95% in 2min (unconjugated dyes eluted in this step) and maintained for 5 min, decreased to 5% and maintained for 5 min. The reaction mixture was diluted with 4.5 mL of water containing 10% v/v acetic acid and loaded on to the column using a 5 mL sample loop. 500 μ l fractions were collected manually based on absorbance measurements observed. Chemical identity of the

synthesized substrate was verified using ESI-MS (Rice core mass spectrometry facility, TX). Desired fractions were pooled together and lyophilized overnight. After reconstitution in 100 μ L of water, substrate (N-AtQ21) concentration was estimated by measuring absorbance of dilutions in Infinite 200 PRO plate reader (Tecan group Ltd, Mannedorf, Switzerland). Y-BQ7 substrate, CAAPYGSKGRGR peptide conjugated to BODIPY-FL-SE and QSY7 C₅ maleimide dyes, to measure activity towards tyrosine was a generous gift from Georgiou lab, University of Texas, Austin.

4.2.3 Screening of chymotrypsin libraries using FACS

For protein expression, 3 mL LB culture was inoculated using a glycerol stock of the library to a starting OD₆₀₀ of 0.02 and grown up to an OD of 0.5 for induction with 100 μ M arabinose at 37 °C for 2 h. Induced cells were washed with 1% sucrose solution and resuspended to an OD of 1. Washed cells were incubated with 20 nM of Y-BQ7 and N-AtQ21 at an OD 0.01 in 1% sucrose solution containing 2 mM Tris (pH 7.5) for 10 min at 25 C. Labeled cells were analyzed using BD Biosciences FACSJazz cell sorter at an event rate of 7000/s. Optimal PMT settings were determined using cells displaying inactive rChyB whose catalytic Ser195 was mutated to Ala. Sort gate, typically containing 0.5-1% of the event rate, was an intersection of a gate in 670/30 nm channel containing cells with high Asn activity and a gate in 530/40 channel to exclude cells which display Tyr activity higher than that of the top 5% of inactive rChyB population. As the fluorescence profile of the population changed gradually during sorting, parent gates were adjusted to maintain their initial frequencies and restrict the frequency of the sort gate to less than 1%. Sorting was performed in the labeling time window of 10-25 min as with longer incubation, cells could get labeled non-specifically. For all the

libraries, number of cells screened was at least 3-fold higher than their genetic diversity estimated based on transformation efficiency. When sorting was performed for long periods, it was ensured that induced cultures were used within 1 h of bench time as a decrease in proteolytic activity with the age of the culture was observed. Sorting was performed using the 'pure' mode and top 1% events corresponding to the sort gate were collected in an empty microcentrifuge tube. Plasmid DNA was recovered from the sorted cells using Zymo miniprep kit (Zymo Research, CA) as described in chapter 3. Subsequent to transformation into *E. coli* MC1061 cells, total number of transformants was estimated by plating dilutions onto LB-Cm plates. Transformants recovered with SOC media were directly grown in 100 mL of 2xYT media supplemented with 0.5% glucose and 25 µg/mL chloramphenicol at 37 °C for 10 h, and used to seed a subculture for the next round of sorting. Screening of site-saturated mutagenesis library and the other two libraries (backcross shuffling and error-prone PCR) involved 6 and 4 rounds of sorting respectively. 10 colonies were randomly picked from the sorted population for clonal characterization using flow cytometry with Y-BQ7 and N-AtQ21 substrates. Mutations in chymotrypsin B corresponding to the clones that showed desired phenotype on the cell sorter were identified by standard Sanger sequencing (SeqWright Inc., Houston, TX).

4.2.4 Enzyme expression and purification

For bacterial expression of chymotrypsin in its zymogen form, the gene fragment coding for chain A of trypsin (Phe-Pro-Val-Asp-Asp-Asp-Asp-Lys) was introduced in the 5' region of the chymotrypsin gene using forward primer overhang in a PCR. Using KpnI and SacI restriction sites, gene fragment coding for fusion of chain A of

trypsin and chains B and C of rChyB or other variants was introduced in-frame into a pET45b based vector containing N-terminal His₆ tag under the regulation of T7 promoter. For protein expression, 100 mL LB culture was inoculated to a starting OD₆₀₀ of 0.02 and grown up to an OD of 0.5 for induction with 500 μ M isopropyl β -D-1-thiogalactopyranoside (IPTG) at 37 °C for 6 h. Purification of inclusion bodies and refolding was performed as described previously (Wang et al., 2011). Briefly, cell pellet was resuspended in phosphate buffer (pH 8.0) containing 4 M guanidine hydrochloride (GuHCl) and 5 mM dithiothreitol and lysed by sonication (3 cycles of 30s on/off/on, XL-2000, QSonica, CT). Inclusion body fraction was isolated by centrifugation (10,000x g, 15 min) and washed with lysis buffer containing 10% Triton-X detergent twice to remove other insoluble contaminants. Small scale histidine tag affinity purification was performed using HisPur Ni-NTA spin plates (Thermo Scientific, MA). Refolding of denatured protein was achieved by flash dilution to a concentration of 10 μ g/mL in Tris buffer (pH 7.5) containing 1 M GuHCl and 3 mM glutathione (reduced/oxidized = 2:1) and incubation at 4 °C overnight. Prior to activation of zymogens, protein was exchanged into Tris buffer (pH 7.5) containing 10 mM CaCl₂ using Zeba spin desalting plates (7 kDa molecular weight cut-off (MWCO), Life technologies, NY) or Amicon centrifugal filters (10 kDa MWCO, Fisher Scientific, IL) depending on the volume as the activity of enterokinase in the refolding buffer is low.

pcDNA 3.4 based plasmid vector used for expression of chymotrypsinogens in human embryonic kidney (HEK293F) cells was a kind gift from Georgiou lab (University of Texas, Austin). A gene fragment coding for signal peptide and chain A of trypsin-1 (Homo Sapiens), rChyB-N-v2 and His₆ tag and containing Kozak sequence (GCCACC)

and sequences overlapping with 5' and 3' ends of AgeI-HF/Bsu36I digested vector for subsequent cloning by Gibson assembly, was commercially purchased (IDT, IA). For rChyB-N-v3, mutations, V99M and T221S, were introduced by overlap extension PCR using the plasmid for expression of rChyB-N-v2 as template. Cloning was also performed to obtain chymotrypsinogens with native chain A (1-15 aa, chA) instead of chain A corresponding to trypsin (**Figure 4.2**). Position 122 was reverted back to Cys from Ser as this residue is involved in a disulfide bond between chain A and chain B. *chA-rChyB-N-v3* gene fragment was commercially purchased (IDT, IA) and wild-type chymotrypsinogen gene, *chA-rChyB*, was generated by overlap-extension PCR of DNA fragments obtained using *chA-rChyB-N-v3* and *rChyB* as templates.

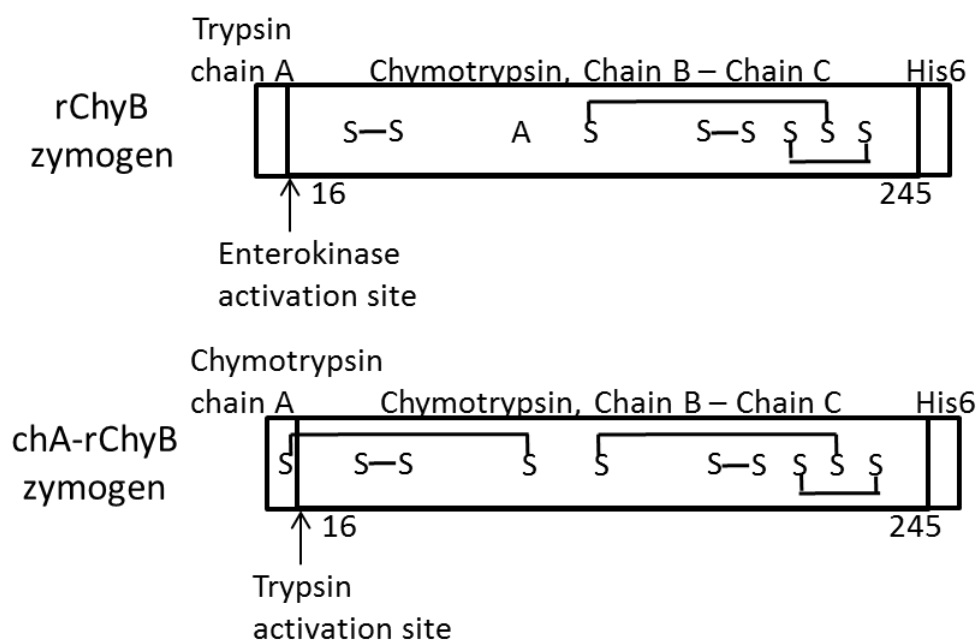


Figure 4.2 Schematic of chymotrypsinogen forms used for protein expression.

After verification of success of cloning by DNA sequencing, 100 mL 2xYT culture supplemented with 200 ug/ml ampicillin was inoculated with a colony of cells harboring pcDNA3.4 based plasmid containing *rChyB* or other engineered mutants, grown

overnight at 37 °C and plasmid DNA was isolated using QIAGEN plasmid maxi kit and QIAvac 24 plus vacuum manifold (Qiagen Inc, CA) to draw the solution quickly through the DNA binding column. With the high copy number of the vector used, typical yield from a 100 mL culture is at least 150 µg of plasmid DNA. 100 mL HEK293F cells grown to a density of 2.5×10^6 cells/mL after three passages in a suspension media culture at 37 °C and 5% CO₂ was transiently transfected with 100 µg of plasmid DNA using 80 µl of Expifectamine transfection reagent (Life technologies, NY). Transfection enhancers provided in the kit were then added 16 h following transfection. Protein secretion was allowed for 4-5 days, usually when viability (as measured by trypan blue) was 30-50%, and cells were spun down at 4000 xg for 20 min at 4 °C to collect the supernatant. Presence of chymotrypsinogen in the supernatant was verified by Western blotting as described previously (Ramesh et al., 2012) using rabbit anti-His antibody (300 ng/mL, GenScript, NJ), goat anti-rabbit antibody conjugated to HRP (40 ng/mL, Jackson ImmunoResearch, PA), and chemiluminescent HRP substrate (SuperSignal West Pico, Thermo Scientific, MA). After activating 100 µL of supernatant with 1 U of recombinant enterokinase (EMD Millipore, MA) at 25 °C for 1h, 1 µL of 5 µg/µL casein-FITC (50-100 µg FITC/mg solid, Sigma Aldrich, MO) solution was added to monitor proteolytic activity by measuring fluorescence (excitation: 488 nm, emission: 530 nm and cut-off: 515 nm) at 25 °C using SpectraMAX Gemini EM plate reader (Molecular Devices, CA). In the case of zymogens containing chain A of chymotrypsin, activation was performed using proteomics-grade trypsin (Sigma Aldrich, MO).

After ascertaining the presence of chymotrypsinogen, purification was performed using AKTA FPLC system (GE Healthcare Bio-Sciences, PA) at 4 °C. 12 mL of 100 mM

NiSO₄ solution was passed through a column packed with 4 mL of chelation-sepharose resin (Cat. No. 17-057501, GE Healthcare Bio-Sciences, PA) to immobilize nickel and purify chymotrypsinogen by histidine tag affinity chromatography. At a flow rate of 5 mL/min, the column was washed with water (8 mL) and 100 mM sodium acetate buffer (pH 4, 20 mL) to remove weakly bound nickel ions and equilibrated with phosphate buffer (50 mM Na₂HPO₄, 500 mM NaCl, pH 8). Filtered (0.22 µm) supernatant was loaded on to the column and 12 mL of phosphate buffer was passed to elute any non-specifically bound proteins. Elution was performed by increasing imidazole concentration from 0 to 150 mM in a linear gradient in 16 min to collect 5 mL fractions. The resin was regenerated by passing 20 mL of 1 M NaOH and 12 mL of water, nickel ions were stripped using 10 mL of phosphate buffer containing 50 mM EDTA and the column was recharged with nickel sulfate before using it for purification of a different protein.

An aliquot of the fractions expected to contain the desired protein was exchanged to Tris buffer (50 mM Tris, 10 mM CaCl₂, pH 7.5) using Zeba spin columns in a 96 well plate format (Thermo Scientific, MA) and protease activity was confirmed using casein-FITC substrate after activation. Fractions containing majority of the protein were pooled together and exchanged to acetate buffer (pH 4) using 10KDa MWCO Amicon columns (15 mL capacity, EMD Millipore, MA) for further purification by cation exchange chromatography as the isoelectric point of chymotrypsinogen is 5.3. 1 mL of concentrated protein was loaded on to a column packed with 0.8 mL of resin (GE Healthcare Bio-Sciences, PA) by injection. After washing the column with 2.4 mL of acetate buffer, the desired protein eluted when the conductivity was 44 mS/cm during the increasing NaCl concentration gradient from 20 mM to 1 M in 80 min. Purity of fractions

(1 mL volume) expected to contain majority of the protein was estimated by running samples denatured using SDS and β -mercaptoethanol, in a 4-20% polyacrylamide gel (Lonza Houston Inc, TX) and subsequent coomassie blue staining. Prior to any functional characterization, zymogens were activated using enterokinase (1 U per 25 μ g of chymotrypsin variants containing chain A of trypsin) or trypsin (1:100 ratio of enzyme to zymogen containing chain A of chymotrypsin) at 25 °C for 1 h and subsequently treated with 500 μ M of tosyl-L-lysine chloromethyl ketone hydrochloride (TLCK, Sigma Aldrich, MO) to inhibit activity of the activating enzymes. Activated zymogens were incubated with equimolar concentration of ActivX fluorophosphonate probe conjugated to 5-carboxytetramethylrhodamine (TAMRA) dye (Thermo Scientific, MA) at 25 °C for 45 min and analyzed by electrophoresis to identify the size of the polypeptide chain containing catalytic Ser195.

4.2.5 HPLC characterization

FRET peptide substrate, N-AtQ21, at concentration 1 μ M, was incubated with activated chymotrypsin variants (1 μ M of rChyB-N-v2 or 50 nM of chA-rChyB-N-v3) in Tris buffer (pH 7.5) at 37 °C for about 1 h. To estimate the degree of proteolysis, an aliquot of the reaction mixture was analyzed on a C18 column by measuring absorbance at 630 and 670 nm corresponding to Atto633 fluorophore and QSY21 quencher respectively using the same gradient method that was previously used during substrate synthesis. To validate the identity of the peaks observed in the chromatogram, aliquots were also characterized by LC-MS (Mass spectrometry laboratory, Department of Chemistry, University of Houston).

Lyophilized peptide (WKAAPNGSSGRGR, N-terminal acetylated and C-terminal amidated, >70% purity, Genscript, NJ) was resuspended in phosphate buffer (pH 8) to a concentration of 10 mM and stored at -20 °C. 100 μ M peptide substrate was digested with 5 μ M of purified rChyB or rChyB-N-v3 in Tris buffer at 25 °C for 30 min and analyzed on a C18 column by measuring tryptophan absorbance at 280 nm. Following gradient method was used with water/acetonitrile mobile phases containing 1% acetic acid at a flow rate of 1 mL/min: 5% acetonitrile for 5 min, increasing to 30% in 20 min, further increased to 95% in 5 min and maintained for 5 min, decreased to 5% and maintained for 5 min.

4.2.6 MS/MS characterization

Activated chymotrypsins were cleaned using Zeba spin desalting columns to remove excess TLCK. MS/MS analysis of bovine serum albumin (BSA) digested with chymotrypsin variants, was performed in collaboration with Bark lab (Department of Biochemistry, University of Houston, TX) as described previously (Bark et al., 2001). Briefly, 1 mg of BSA dissolved in 100 μ L of 100 mM ammonium bicarbonate buffer, pH 7.5 was reduced by adding 5 μ L of 100 mM tris(2-carboxyethyl)phosphine (TCEP, diluted from 500 mM stock with ammonium bicarbonate buffer). After incubation at 37 °C for 15 min, 5 μ L of 100 mM iodoacetamide (IAA, diluted from 500 mM stock with ammonium bicarbonate buffer) was added to the reaction mixture for incubation at 25 °C for 30 min in dark as IAA is both heat and light sensitive. Excess reagents were removed by chloroform:methanol:water precipitation and the protein was reconstituted in ammonium carbonate buffer to a concentration of 100 μ g/ μ L. 50 μ g of treated BSA was digested 1 μ g of purified wild-type or engineered chymotrypsin overnight at 37 °C in 50

μL of ammonium bicarbonate buffer. Digested samples were first chromatographically separated (Agilent 1200 Nanoflow, Agilent Technologies, CA) and then analyzed in a quadrupole time-of-flight mass spectrometer (Bruker MircoTOF-Q, Bruker Daltonics, MA) equipped with an interface for electrospray ionization. MS spectra was matched to a custom protein database containing BSA, rChyB, rChyB-N-v3 and enterokinase using open mass spectrometry search algorithm (OMSSA) to identify the sequences of peptides generated.

4.2.7 Kinetic parameter measurements

All measurements were made using Infinite 200 PRO plate reader (Tecan group Ltd, Mannedorf, Switzerland) at 25 °C. Total protein concentration was estimated by bicinchoninic acid (BCA) colorimetric assay (Thermo Scientific, MA) where the calibration curve was generated by measuring dilutions of albumin standard (1 mg/mL, Sigma Aldrich, MO). Concentration of functional chymotrypsin after zymogen activation was determined by active-site titration against 4-Methylumbelliferyl p-trimethylammoniocinnamate chloride (MUTMAC, Sigma Aldrich, MO). Conversion of relative fluorescence units (RFU) measured at 380/445 nm (excitation/emission wavelengths) to molarity was achieved by calibrating with different concentrations (0-5 μM) of the fluorophore, 4-Methylumbelliferone, in presence of 50 μM MUTMAC and chA-rChyB zymogen incubated with TLCK-treated trypsin was used as negative control. Kinetics of chA-rChyB and chA-rChyB-N-v3 towards three different substrates, suc-Ala-Ala-Pro-Phe-AMC (EMD Millipore, MA), suc-Ala-Ala-Pro-Asn-AMC (custom synthesized, Anaspec, CA), and suc-Leu-Leu-Val-Tyr-AMC (RnD Systems Inc, MN) were measured where suc refers to succinylation at the N-terminal and AMC refers to 7-

amino-4-methylcoumarin, the fluorophore quenched by conjugation to the C-terminal of the peptide. AMC fluorescence measured at 380/460 nm was calibrated with dilutions of unconjugated AMC stock solution whose concentration was calculated using the values of pathlength (0.3 cm), molar absorption coefficient of $16,000 \text{ M}^{-1}\text{cm}^{-1}$ and background subtracted absorbance measured at 354 nm. Calibration was also performed at varying gain settings and concentrations of peptide substrates. Optimal enzyme/substrate concentrations and gain settings that allowed a linear increase in fluorescence in the first 5 min of incubation were identified. Depending on the kinetics, enzyme concentration was in the range of 2.5-50 nM and substrate concentrations were varied from 20-200 μM except for suc-LLVY-AMC substrate where the highest concentration was limited to 50 μM due to poor solubility.

4.3 Results

4.3.1 Surface loops critical for substrate specificity randomized in a chymotrypsin B library

Chymotrypsin B is a typical trypsin-fold serine protease containing the catalytic triad (His57, Asp102, and Ser195) at the cleft between its two β -barrel domains. Previously, based on structural and sequence alignment of trypsin, chymotrypsin and elastase, triplet positions 189, 216 and 226 which are in close proximity to the P1 substrate residue were considered to be responsible for substrate specificity (Wouters et al., 2003). Site-directed mutagenesis studies targeting these positions to modify enzyme specificity did not yield functional variants. In this study, we have used a combinatorial approach for modifying substrate specificity of chymotrypsin B (rChyB) by screening protein libraries containing

randomized variations at targeted positions. HotSpot Wizard (Pavelka et al., 2009) was used to generate a preliminary list of candidate residues for library construction based on their proximity to active site and how conserved they were evolutionarily. Crystal structure of δ -chymotrypsin bound to a peptidyl chloromethyl ketone inhibitor (Mac Sweeney et al., 2000) and multiple sequence alignment of amino acid sequences of proteases with chymotrypsin-like tertiary structure and differing substrate specificities from different organisms performed using MUSCLE (Edgar, 2004), were used as inputs (**Figure 4.3**). The underlying assumption for using sequence alignment was that residues critical for catalytic activity should be conserved and hence should be not randomized to retain the catalytic efficiency. Though structure of rat chymotrypsin B with mutations Ser189Asp and Ala226Gly was available, structure of δ -chymotrypsin was used as it was bound to a substrate analogue. Loops 185-192 and 215-226 were identified as key determinants of substrate specificity and was supported by their significance in evolutionary divergence of substrate specificity within chymotrypsin-fold serine proteases (Perona and Craik, 1997) and in all the protein engineering efforts to modify P1 specificity of trypsin and chymotrypsin (Hedstrom et al., 1992; Hung and Hedstrom, 1998; Jelinek et al., 2008; Reyda et al., 2003). Ideally all the variable positions identified in the loops should have been randomized while keeping the number of mutations per library member to be low (<10). However, techniques like parsimonious mutagenesis (Balint and Larrick, 1993) or DNA shuffling (Stemmer, 1994) with wild-type typically used for such a purpose could not be applied as most of the residues are in the primary sequence contiguously. As a trade-off to decrease the mutational load and increase the fraction of library diversity screened, residues 185-188 which were the farthest from the

substrate analogue among all the candidate residues were not randomized. Positions 189, 192, 216-219, 221-224 and 226 were targeted using NNS codons.

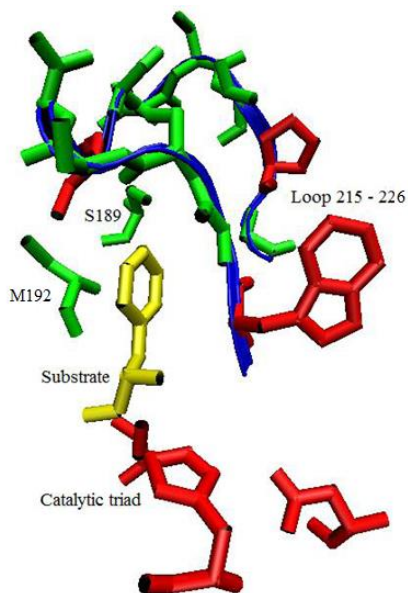
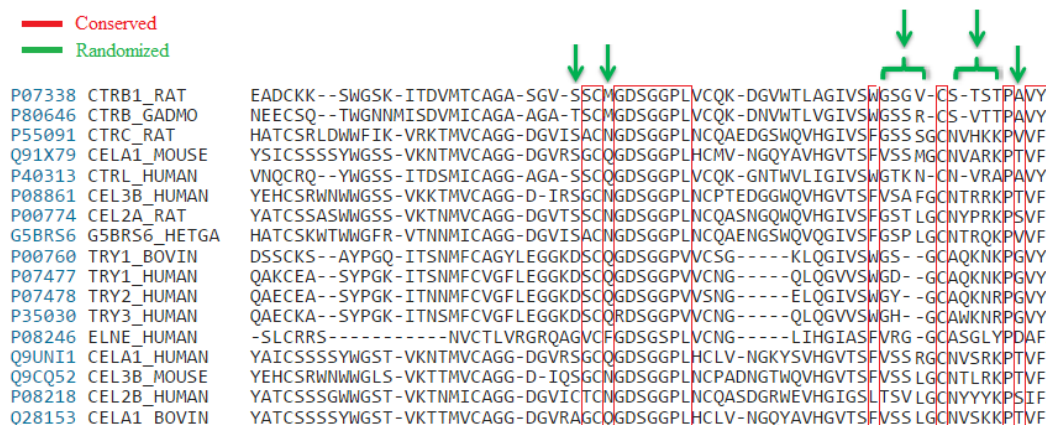


Figure 4.3 Sequence alignment of chymotrypsin-fold proteases with differing substrate specificities and species origin. Residues Ser189 and Met192, and loop 215 -226 were randomized (marked in green) barring conserved residues Trp215, Cys220 and Pro225 (marked in red).

4.3.2 Isolation of rChyB-N-v1 variant with activity towards Asn using FACS

The site-saturated chymotrypsin B library (size $\sim 10^7$ variants) was screened by fluorescence activated cell sorting (FACS) using a dual-substrate selection-counterselection strategy described previously (Varadarajan et al., 2008). Activities of

library members towards Asn and Tyr residues were monitored simultaneously using FRET peptide substrates, N-AtQ21 and Y-BQ7, containing positively charged tail for localized electrostatic capture of fluorescent peptide fragment on the surface of bacteria, thus allowing isolation of variants with high Asn activity while culling out inactive or non-specific variants using FACS (**Figure 4.4**). The protease assay was validated by high Tyr activity observed for cells expressing wild-type chymotrypsin (rChyB) in comparison to inactive rChyB (catalytic Ser195Ala). When cells expressing rChyB were labeled with N-AtQ21 substrate, measured activity towards Asn was similar to that of the negative control. In addition to functional activity, presence of chymotrypsin on the cell surface was verified using an antibody specific for FLAG epitope tag. As NNS codons were used for randomization during library construction, several library members were expected to contain premature stop codons as confirmed by the FLAG negative sub-population in the naïve library. The plasmid recovery method described in the previous chapter was employed to retrieve the genotype of sorted cells for further rounds of sorting as viability of the sorted cells was poor. Starting from the naïve site-saturated chymotrypsin library, stepwise enrichment of variants displaying high Asn activity through six rounds of sorting is shown in **Figure 4.5**. From the round six population, five clones were randomly picked for identifying mutations in the target positions and their Tyr/Asn activities were also assessed by flow cytometry using FRET peptide substrates (**Table 4.3**). Three of them were identical in their sequence and showed 2-fold higher Asn activity than rChyB. The round 5 population was sorted again and five random clones picked from the newly obtained round 6 population were also identical to rChyB-N-v1 sequence. As rChyB-N-

v1 showed the highest Asn activity in comparison to the other two clones (v1c3 and v1c5), only this clone was carried forward for further analysis.

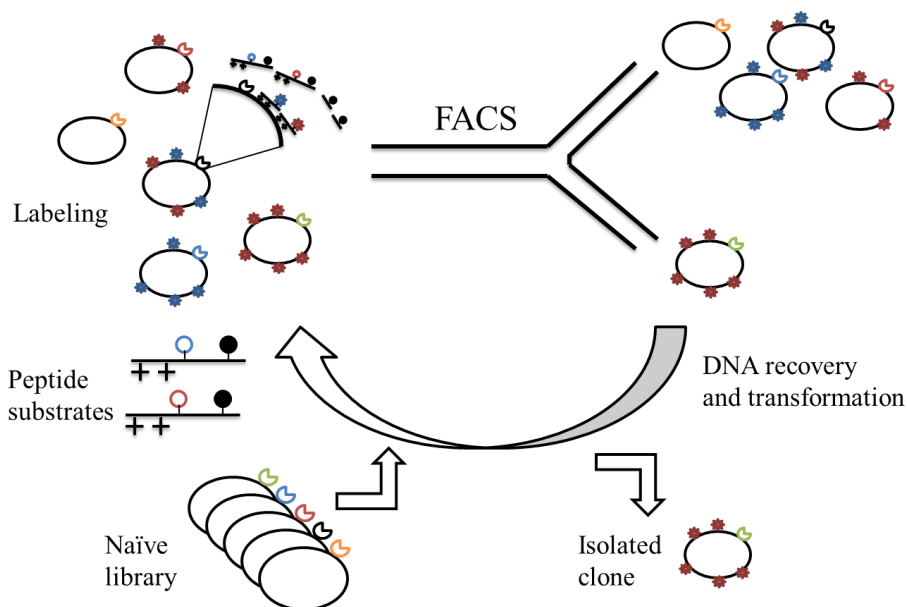


Figure 4.4 Schematic for high-throughput screening of chymotrypsin libraries using FRET peptide substrates for iterative enrichment of active variants showing specificity towards Asn substrate.

Table 4.3 Analysis of individual clones randomly picked from the site-saturated chymotrypsin library enriched by six rounds of sorting. Best variant of the three which was carried forward for further engineering is highlighted in bold.

Clone	Activity towards Asn	Activity towards Tyr	Amino acids at target positions (chymotrypsin numbering)										
	(ratio to background)	(ratio to background)	189	192	216	217	218	219	221	222	223	224	226
rChyB	1	2.5	Ser	Met	Gly	Ser	Gly	Val	Ser	Thr	Ser	Thr	Ala
v1c3	1	1.5	Leu	Val	Arg	Leu	Gly	Asp	Pro	Ala	Ala	Gly	Gly
v1c5	1.8	1.5	Ser	Trp	Gly	Ser	Lys	Phe	Gly	Gly	Gly	Arg	Gly
rChyB-N-v1	2	1	Met	Arg	Gly	Val	Arg	Glu	Thr	Gly	Thr	Trp	Gly

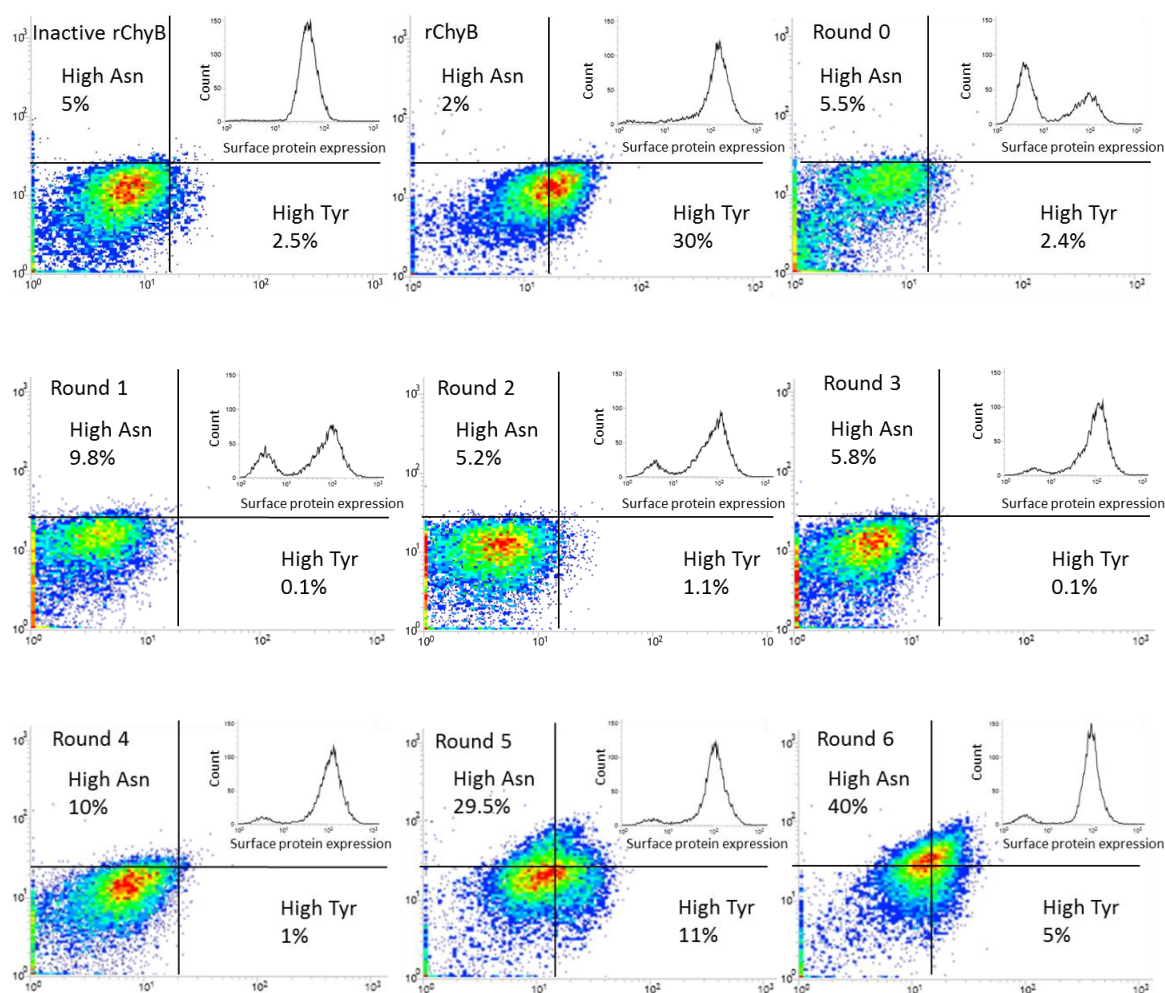


Figure 4.5 Evolution of naïve chymotrypsin B library to enrich variants displaying high Asn activity through multiple rounds of sorting. Corresponding histograms representing the presence of full-length chymotrypsin displayed is shown as inset.

rChyB-N-v1 yielded 2-fold higher fluorescence with N-AtQ21 and not Y-BQ7 suggesting its preference for proteolysis at Asn as the amino acid sequence (AAPXGS) in the linker region between the FRET pair was identical in both peptide substrates except for X which was either Asn or Tyr. To ascertain that the fluorescence observed was not an artifact but due to FRET disruption by proteolysis, cells displaying rChyB-N-v1 or inactive rChyB were incubated with N-AtQ21 substrate and the reaction mixture after 0.22 μ m filtration was analyzed by tandem mass spectrometry coupled to liquid chromatography (LC-MS/MS). All the background peaks present in both inactive rChyB

and rChyB-N-v1 samples were neglected. The peak corresponding to intact N-AtQ21 substrate was observed only in inactive rChyB sample. In the case of rChyB-N-v1, peaks corresponding to the masses of the peptide fragments expected to be generated by proteolysis at Asn residue in N-AtQ21 substrate were identified (**Table 4.4**). Other peaks whose identity could not be matched to proteolysis at any residue in N-AtQ21 were also present. Starting from wild-type chymotrypsin with negligible activity towards Asn, evidence for proteolysis at Asn residue in N-AtQ21 substrate provided sufficient rationale to proceed with further engineering of rChyB-N-v1 variant and characterization with purified enzyme as opposed to enzyme displayed on the cell surface.

Table 4.4 Mass spectrometry analysis of N-AtQ21 substrate digested with cells expressing inactive rChyB or rChyB-N-v1 to identify peptide fragments generated by proteolysis.

m/z	Intensity	z	Neutral mass (Da)	Matched identity
Peaks unique to inactive rChyB				
643.09	2.5	4	2568.3	Unreacted N-AtQ21 substrate
				Ac-Lys(quencher)-Ala-Ala-Pro-Asn- Gly-Ser-Cys(fluorophore)-
				Gly-Arg-Gly-Arg-NH ₂
Peaks unique to rChyB-N-v1				
507.74	2.5	2	1014.5	
585.79	1.2	2	1170.6	
409.88	0.6	3	1227.6	Ac-Lys(quencher)-Ala-Ala-Pro-Asn
648.86	4.73	2	1296.7	
455.92	1.7	3	1365.7	Gly-Ser-Cys(fluorophore)-Gly-Arg-Gly-Arg-NH ₂
484.91	0.6	3	1452.7	

4.3.3 Minimal number of mutations required for activity towards P1 Asn identified by backcross shuffling

Of the 11 positions targeted in the site-saturated library, rChyB-N-v1 had mutations at 10 of them except G216. To verify if all the mutations were necessary for the observed change in substrate specificity and to eliminate deleterious mutations if any, a library was

constructed where each of these ten positions could contain the amino acid corresponding to rChyB or rChyB-N-v1. As amino acid changes were introduced by using mixed bases with a valency of 2 at 16 nucleotide positions, the theoretical size of the library is $2^{16} = 65536$.

Table 4.5 Analysis of individual clones randomly picked from the population enriched by four rounds of sorting the library obtained by backcrossing rChyB-N-v1 with rChyB. Best variant of different clones tested is highlighted in bold.

Clone	Activity towards Tyr	Activity towards Asn	Amino acids at target positions (chymotrypsin numbering)									
	(ratio to background)	(ratio to background)	189	192	217	218	219	221	222	223	224	226
rChyB	1.0	2.5	Ser	Met	Ser	Gly	Val	Ser	Thr	Ser	Thr	Ala
rChyB N-v1	1.0	2.0	Met	Arg	Val	Arg	Glu	Thr	Gly	Thr	Trp	Gly
rChyB N-v2	1.4	4.0	Ser	Arg	Ser	Arg	Glu	Thr	Gly	Thr	Trp	Gly
v2c1	1.3	3.3	Ser	Arg	Ile	Arg	Val	Thr	Gly	Thr	Trp	Gly
v2c2	1.2	3.0	Ser	Arg	Val	Arg	Glu	Ser	Ala	Ser	Trp	Gly
v2c3	1.3	3.3	Ser	Arg	Ser	Arg	Val	Ser	Gly	Ser	Trp	Gly
v2c5	1.3	2.2	Ser	Arg	Ile	Arg	Glu	Ser	Ser	Ser	Ser	Gly
v2c7	2.0	3.3	Ser	Arg	Ser	Arg	Glu	Ser	Gly	Ser	Trp	Gly
v2c8	1.5	3.8	Ser	Arg	Ser	Arg	Glu	Ser	Gly	Ser	Trp	Gly
v2c9	1.2	2.0	Ser	Arg	Ser	Arg	Glu	Ser	Gly	Ser	Trp	Gly
v2c10	1.3	3.8	Ser	Arg	Ile	Arg	Glu	Thr	Gly	Ser	Trp	Gly

Ten clones randomly picked from the population obtained after 4 rounds of sorting were sequenced to identify the mutations and were also analyzed by flow cytometry to measure activity towards Tyr and Asn peptide substrates. Several positions (S189, Arg192, Arg218, Trp224 and G226) were conserved across all the different clones tested showing their significance towards Asn activity. rChyB-N-v2 which showed the highest fluorescence, had two mutations (M189 and V217) reverted back to Ser, as in wild-type

rChyB, leading to an increase in activity towards N-AtQ21 by 2-fold in comparison to rChyB-N-v1.

4.3.4 Purified rChyB-N-v2 is functional towards the peptide substrate containing Asn

Unlike native chymotrypsinogen which requires activation by trypsin, rChyB-N-v2 was expressed as a zymogen where chains B and C of chymotrypsin B were N-terminally fused to signal peptide and chain A of trypsin (Jelinek et al., 2004) such that activation can be performed using enterokinase with extended specificity. Recombinant enzymes were overexpressed in the cytoplasm of *E.coli* BL21(DE3) strain to form inclusion bodies which were solubilized and purified by histidine tag affinity chromatography. As the conditions for refolding of denatured chymotrypsin C was available in literature, rChyB, inactive rChyB and rChyB-N-v2 zymogens were quickly obtained without further optimization. Though enterokinase activated rChyB-N-v2 was found to be functional using casein-FITC assay, it did not cleave N-AtQ21 peptide even when incubated with equimolar concentrations of enzyme and substrate. As an alternative to bacterial expression, mammalian host (HEK293F cells) which allows secretion of folded zymogens was chosen to eliminate the need for refolding. After expression in HEK293F cells, the supernatant containing the secreted protein was characterized using Western blot. Upon activation with enterokinase, formation of chains B and C with lower sizes could be observed in the samples rChyB and rChyB-N-v2 whereas the brightest band for the inactive rChyB sample was still the intact 25 kDa zymogen (**Figure 4.6**). The presence of catalytic Ser195 in the 10 kDa band corresponding to chain C was independently verified by PAGE analysis of activated chymotrypsins incubated with a

fluorophosphonate probe conjugated to rhodamine dye. Chymotrypsinogens from 100 mL HEK293F culture were then purified by His-tag affinity and cation exchange chromatography to obtain 2-3 mgs of protein with greater than 95% purity (as determined by SDS-PAGE). When activated chymotrypsins were incubated with N-AtQ21 and analyzed on HPLC, we observed proteolysis only with rChyB-N-v2 but not rChyB, thus confirming that the signal associated with N-AtQ21 substrate observed previously in FACS was not a FRET related artifact or due to host proteases.

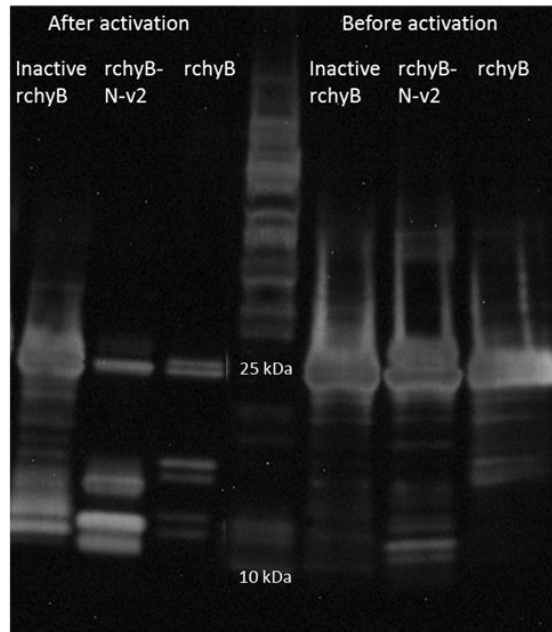


Figure 4.6 Western blot showing chymotrypsinogens expressed in mammalian cells.

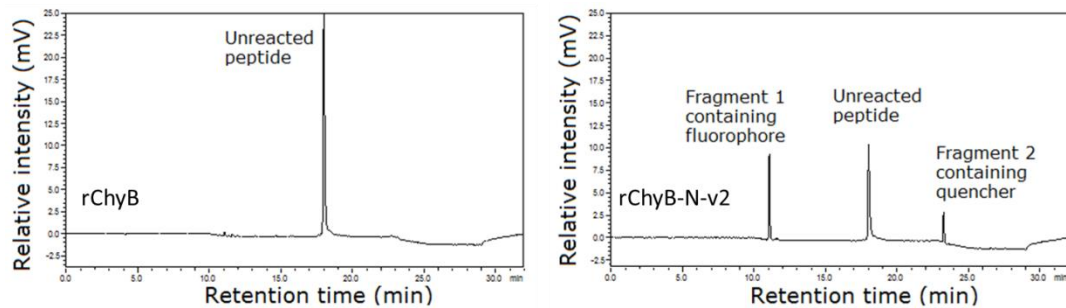


Figure 4.7 HPLC analysis of purified wild-type rChyB or engineered rChyB-N-v2 incubated with N-AtQ21, FRET peptide substrate containing Asn.

4.3.5 Improved proteolytic activity towards P1 Asn by random mutagenesis

To identify beneficial mutations outside the 11 positions targeted, we screened an error-prone PCR library constructed using rChyB-N-v2 as template. After 4 rounds of sorting, ten clones randomly picked were sequenced to identify the enriched mutations and were also analyzed by flow cytometry to measure activity towards Tyr and Asn peptide substrates. No consensus mutation was observed across all the clones sequenced and only marginal improvement in Asn activity over rChyB-N-v2 was displayed. rChyB-N-v3 variant with an additional mutation, Val99Met, in chain B which showed the highest fluorescence was selected for protein expression.

Table 4.6 Analysis of individual clones randomly picked from the population enriched by four rounds of sorting the library obtained by error-prone PCR with rChyB-N-v2 as template. Best variant of different clones tested is highlighted in bold.

Clone	Activity towards Tyr (ratio to background)	Activity towards Asn (ratio to background)	Mutations in rChyB-N-v2
rChyB-N-v2	1.4	4.0	-
rChyB-N-v3	1.4	4.3	Val99Met
v3c2	1.3	2.9	Glu49Lys, Gln87Arg, Ala158Val
v3c3	1.2	2.3	Val99Leu, Lu116Val
v3c4	1.0	2.0	Val99Met, Thr110Ser, Leu149Pro
v3c5	1.4	3.7	Gly59Arg
v3c7	0.9	1.1	Ala148Val
v3c9	1.7	3.0	Gly74Ala, Asn125Ser, Pro161His

In order to ensure that the engineered enzyme can cleave unconjugated peptides, we used a substrate with amino acid sequence identical to that of N-AtQ21 FRET substrate but with a Trp residue added to the N-terminal for characterization by monitoring absorbance at 280 nm, instead of any dyes. Also Cys residue which was used for conjugation to Atto-633 dye in N-AtQ21 substrate was mutated to Ser to avoid any disulfide bond formation between two peptide molecules. The chromatograms of peptide

samples digested with wild-type rChyB and engineered variants, rChyB-N-v1 and rChyB-N-v2 were identical to that of undigested peptide. Multiple peaks in the negative control sample could be due to impurities as the peptide was purchased from the manufacturer at crude purity (>70%). When incubated with rChyB-N-v3, an additional fragment eluting at higher retention time was observed possibly due to proteolysis after Asn removing the positively charged arginine tail (**Figure 4.8**).

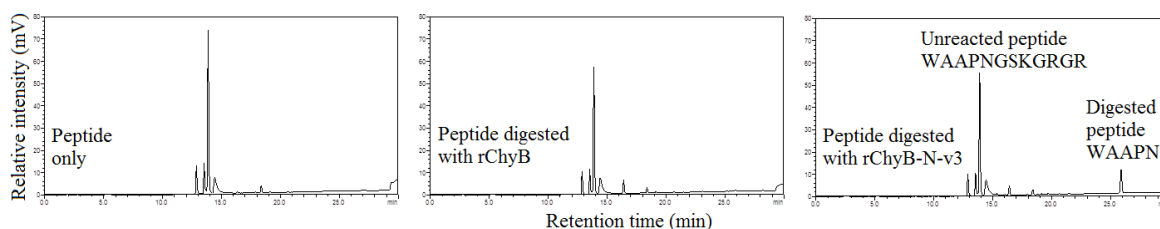


Figure 4.8 HPLC analysis of rChyB-N-v3 with unconjugated peptide substrate.

In a previous study, replacing the native chain A (1-15 aa) of chymotrypsin B with chain A of trypsin was shown to substantially decrease the spontaneous activation of chymotrypsinogen in the absence of enterokinase without affecting the kinetics towards peptide substrates (Venekei et al., 1996a). However additional disulfide bond between Cys1 residue in chain A and Cys122 in chain B is known to increase the thermal and chemical stability of the enzyme (Kardos et al., 1999). Considering the need for stability in downstream proteomic applications, we expressed and purified rChyB-N-v3 with native chain A (chA-rChyB-N-v3). LC-MS analysis of N-AtQ21 substrate digested with chA-rChyB-N-v3 located the exact scissile bond at C-terminal of Asn residue as the Gly-Ser-Cys-Gly-Arg-Gly-Arg fragment containing the fluorophore was reliably detected (**Figure 4.9**). However the corresponding Lys-Ala-Ala-Pro-Asn fragment containing the quencher was not detected possibly due to poor ionization.

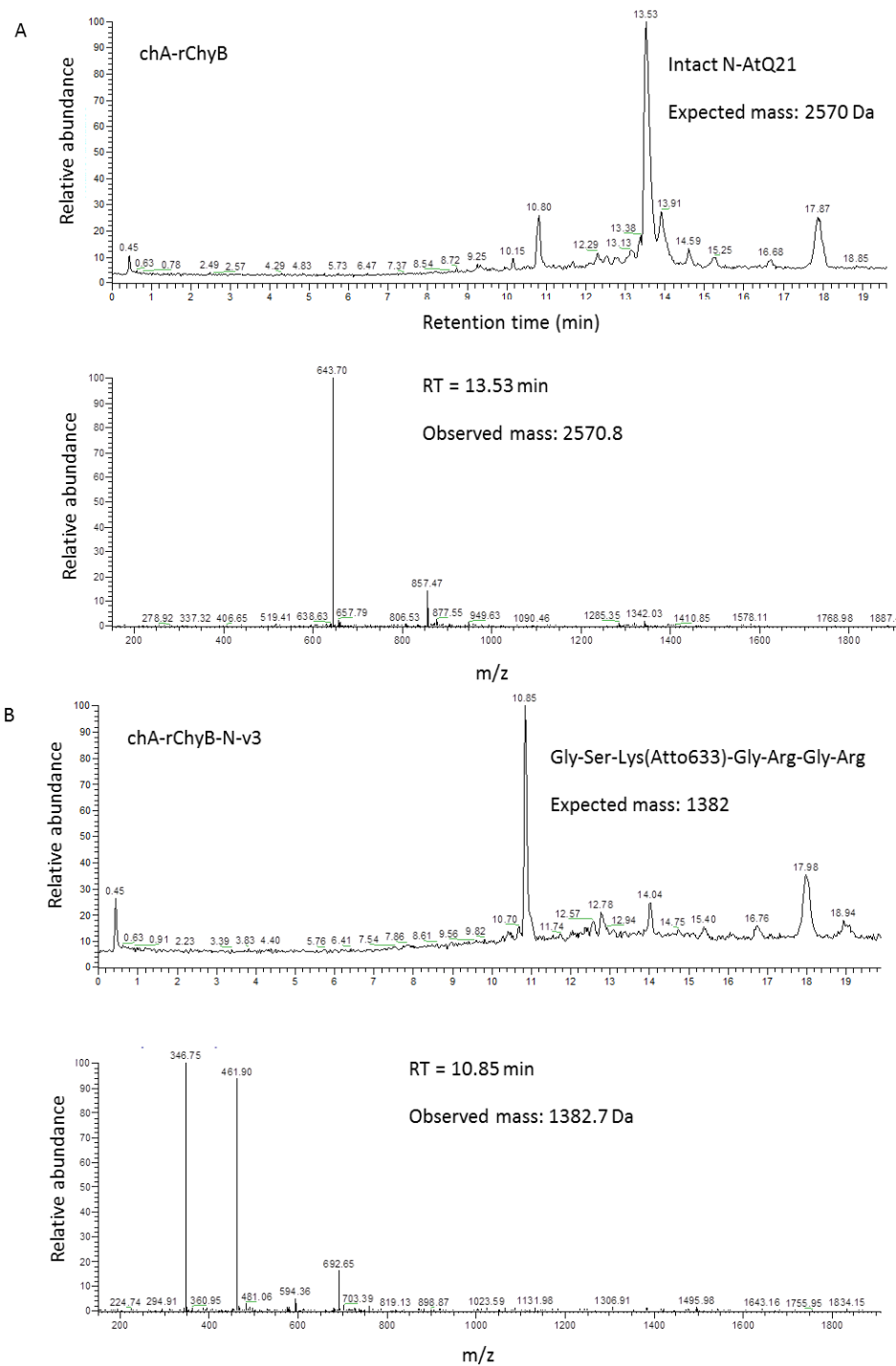


Figure 4.9 LC-MS analysis of N-AtQ21 peptide substrate digested with (A) wild-type chA-rChyB and (B) engineered chA-rChyB-N-v3.

Only intact N-AtQ21 substrate was detected when analogous wild-type rChyB sample was characterized. In contrast to the previous study, it appears that chain A of chymotrypsin could improve the kinetics of proteolysis as the concentration of rChyB-N-v3 (without chain A) required to digest 1 μ M of N-AtQ21 for detection by HPLC was at least 15-fold higher than chA- rChyB-N-v3.

4.3.6 Variants with high Asn activity displayed increase in polarity of active site

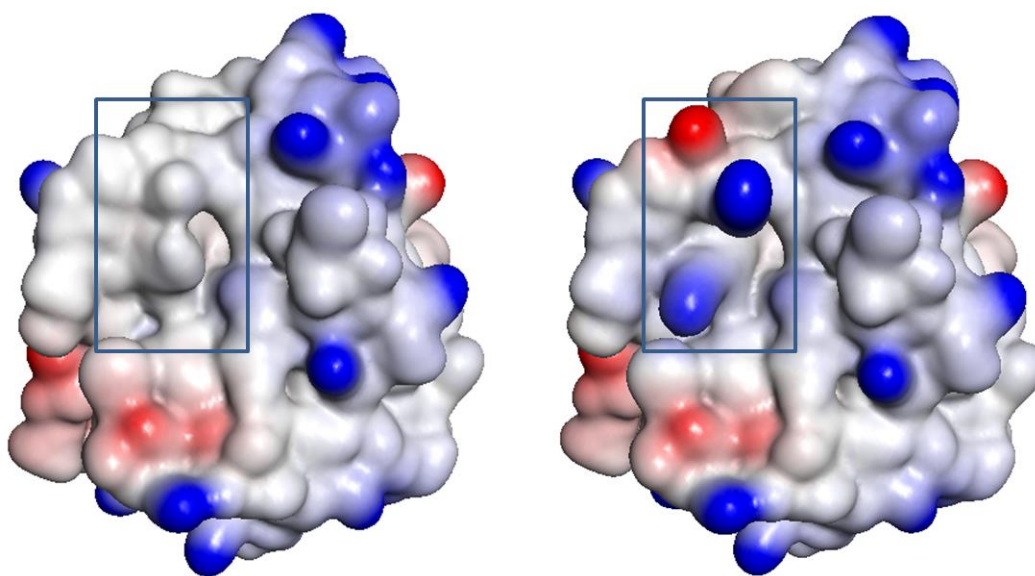


Figure 4.10 Electrostatic surface potentials of wild-type and engineered chymotrypsins.

We sought to understand how mutations in rChyB-N-v3 led to high Asn activity. To this end, corresponding mutations were introduced into the experimentally determined structure of wild-type chymotrypsin (PDB id 1DLK) using PyMOL. Electrostatic nature of substrate binding pocket (highlighted by the blue rectangle) in wild-type and engineered chymotrypsin were compared using Discovery Studio 4.1 (**Figure 4.10**). In the case of wild-type enzyme, only apolar (white) amino acids were present. In contrast, two positively charged (blue) residues and a negatively charged residue (red) provided a zwitterionic environment in the substrate binding pocket of engineered chymotrypsin. It

is worth mentioning that legumain (asparaginyl endopeptidase) also displays a similar feature to interact with electrophilic carbonyl group and nucleophilic nitrogen in the amide side chain (Dall and Brandstetter, 2013).

Clone	Asn/Tyr activity ratio	Amino acids corresponding to targeted positions												Net Hydropathy index (HI)
		99	189	192	216	217	218	219	221	222	223	224	226	
rChyB	0.4	Val	Ser	Met	Gly	Ser	Gly	Val	Ser	Thr	Ser	Thr	Ala	6.8
rChyB-N-v1	2	Val	Met	Arg	Gly	Val	Arg	Glu	Thr	Gly	Thr	Trp	Gly	-5.8
rChyB-N-v2	2.8	Val	Ser	Arg	Gly	Ser	Arg	Glu	Thr	Gly	Thr	Trp	Gly	-13.5
rChyB-N-v3	3.1	Met	Ser	Arg	Gly	Ser	Arg	Glu	Thr	Gly	Thr	Trp	Gly	-15.7

Table 4.7. Hydropathic analysis of chymotrypsin variants.

To quantify the change introduced in the active site of engineered chymotrypsin, hydropathy index (HI), a measure of free energy of hydration, was used as the metric (Kyte and Doolittle, 1982). Positive values represent hydrophobic nature and their magnitude correlate with degree of hydrophobicity. Amino acid changes at mutated positions in chymotrypsin variants identified during iterative mutagenesis are shown in **Table 4.7**. In addition to the color code previously used, black and green colors represent small/apolar and hydrophobic amino acids respectively. Net hydropathy index for each enzyme was calculated by summation of individual HI values assigned to amino acids. Starting from a hydrophobic pocket (HI = 6.8) in wild-type chymotrypsin, polarity of the active-site was progressively increased by introduction of charged residues and reversal of hydrophobic residues. Not surprisingly, increase in polarity of substrate binding pocket

correlates with increase in selectivity towards polar Asn over hydrophobic Tyr as measured by flow cytometry using FRET peptide substrates.

4.3.6 Kinetic characterization of engineered chymotrypsin with fluorogenic peptide substrates

To quantify the effect of mutations introduced on the substrate specificity, kinetics of wild-type and engineered chymotrypsins were measured using fluorogenic peptide substrates containing Tyr or Asn residues at P1 position. We could not estimate individual Michaelis-Menten parameters as saturation of initial velocity rates were not observed at any of the concentrations tested (determined by solubility limits). In the case of LLVY substrate, highest concentration tested was 50 μ M due to its poor solubility. Wild-type chymotrypsin showed very fast kinetics towards LLVY peptide that the enzyme concentration had to be decreased to 1 nM to avoid saturation of the detector. In comparison, engineered chymotrypsin demonstrated at least three orders of magnitude decrease in the propensity to cleave after Tyr or Phe residues (**Figure 4.11A**) suggesting that variants with high wild-type like activity were efficiently eliminated by counter-selection with Y-BQ7 substrate during library screening .

In comparing the kinetics of the Asn substrate, engineered chymotrypsin showed 10-fold increase in the rate of product generation in comparison to wild-type chymotrypsin (**Figure 4.11B**). To explore the possibility of further increasing P1 Asn activity, we are currently working on screening a library based on rChyB-N-v3 with randomized codons in the loop 185-188 as substrate affinity could be improved by providing additional subsites for favorable interactions.

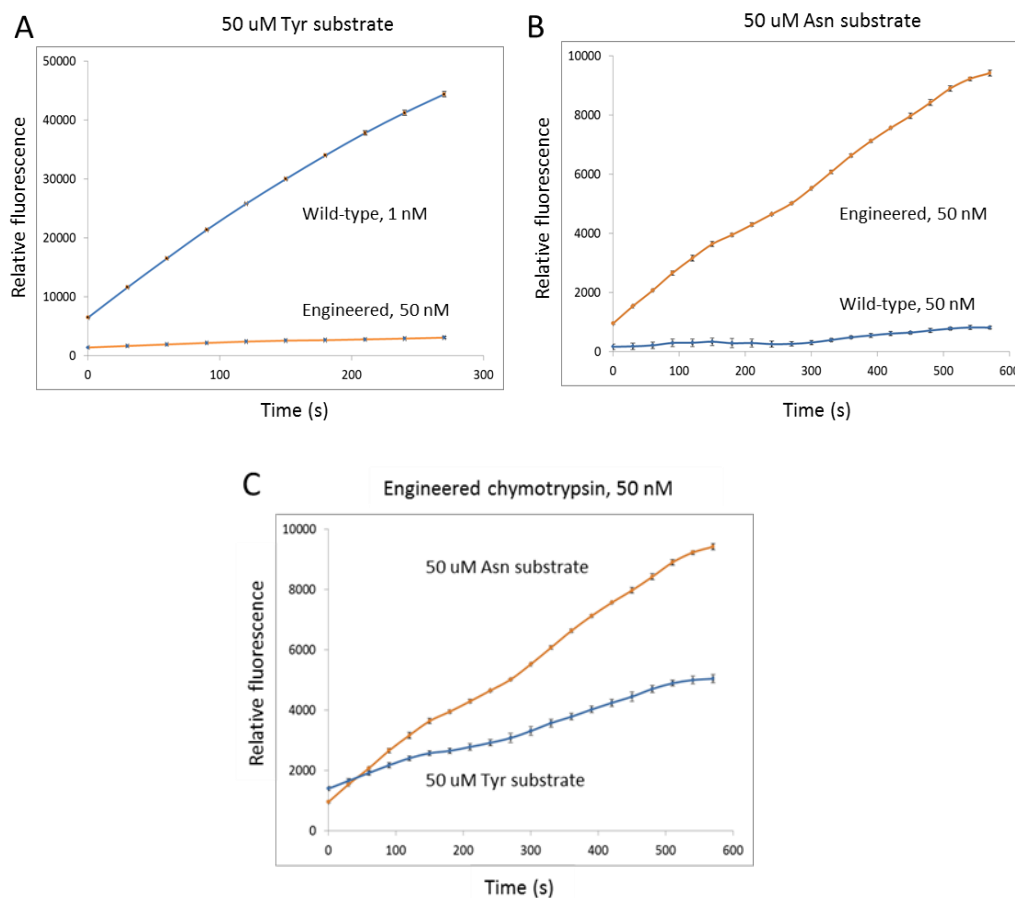


Figure 4.11. Substrate preference of wild-type (chA-rChyB) and engineered chymotrypsins (chA-rChyB-N-v3) with fluorogenic peptide substrates, suc-Leu-Leu-Val-Tyr-AMC and suc-Ala-Ala-Pro-Asn-AMC.

Though counter-selection decreased the P1 Tyr activity substantially, specificity of the engineered enzyme is not exclusive towards P1 Asn as shown by its preference towards Asn over Tyr substrate by only 2-fold (**Figure 4.11C**). During activation of zymogen, the release of N-terminus of Ile16 by trypsin is followed by autoproteolysis of chymotrypsin at Y146 to form chains B and C. Elimination of Tyr activity by engineering might hinder the ability of chymotrypsin to process itself into an active mature form and hence we believe that at least some residual wild-type activity is essential for function. It should be noted that the retention of activity towards hydrophobic amino acids in the

engineered variant is also beneficial for the application of protein sequencing by tandem mass spectrometry with high sequence coverage.

4.3.7 Engineered protease cleaves after Asn in protein substrate

Proteases which efficiently process peptide substrates often do not cleave the same sequence motif present in a protein scaffold possibly due to lack of the preferred β -strand conformation (Tyndall et al., 2005). Microbial proteases that were previously engineered to cleave after novel P1 residues in peptides did not display any catalytic activity towards protein substrates (Varadarajan, 2006). To assess the ability of engineered chymotrypsin to digest proteins, bovine serum albumin (BSA) was used as a model substrate. When reduced/alkylated BSA was digested chA-rChyB-N-v3, 132 different peptides with unique amino acid sequences were identified by MS/MS analysis (**Figure 4.12**). Of all the 14 Asn residues in BSA, proteolysis was reliably detected after 5 of them (positions 123, 144, 185, 341 and 414). Near all the missed Asn residues except Asn290, proteolysis at the neighboring amino acids (< 2 -3 aa distance) was detected and this could have decreased the size of peptides with C-terminal Asn to less than 500 Da which cannot be detected using MS analysis. Typically proteases miss a putative cleavage site when specific residues are present near the P1 site. For example, trypsin is known to cleave after Lys/Arg less efficiently when an Asp/Glu residue is present in the P2, P1' or P2' positions (Gershon, 2014). It is possible that Asn290 cleavage site was missed by the engineered chymotrypsin due to such non-P1 position constraints.

10	20	30	40	50
MKWVTFISLL	LLFSSAYSRG	VFRDTHKSE	IAHRFKDLGE	EHFKGLVLIA
60	70	80	90	100
FSQYLQQCPF	DEHVKLVEL	TEFAKTCVAD	ESHAGCEKSL	HTLFGDELCK
110	120	130	140	150
VASLRETYGD	MADCCEKQEP	ERNECFLSHK	DDSPDLPLK	PDPNTLCDEF
160	170	180	190	200
KADEKKFWGK	YLYEIARRHP	YFYAPELLYY	ANKYNGVFQE	CCQAEDKGAC
210	220	230	240	250
LLPKIETMRE	KVLASSARQR	LRCASIQKFG	ERALKAWSVA	RLSQKFPKAE
260	270	280	290	300
FVEVTKLVTD	LTKVHKECCH	GDLLCADDR	ADLAKYICDN	QDTISSKLKE
310	320	330	340	350
CCDKPLLEKS	HCIAEVEKDA	IPENLPPLTA	DFAEDKDVCK	NYQEAKDAFL
360	370	380	390	400
GSFLYEYSRR	HPEYAVSVLL	RLAKEYEATL	EECCAADDPH	ACYSTVFDFKL
410	420	430	440	450
KHLVDEPQNL	IKQNCDFEK	LGEYGFQNAL	IVRYTRKVPQ	VSTPTLVEVS
460	470	480	490	500
RSLGKVGTRC	CTKPESERMP	CTEDYLSLIL	NRLCVLHEKT	PVSEKVTGCC
510	520	530	540	550
TESLVNRRPC	FSAITPDETY	VPKAFDEKLF	TFHADICTLP	DTEKQIKKQT
560	570	580	590	600
ALVELLKHKP	KATEEQIKTV	MENFVAFVDK	CCAADDKEAC	FAVEGPKLVV
STQTALA				

Figure 4.12 Sequence coverage of bovine serum albumin (BSA) based on MS/MS analysis of peptides generated by digestion with engineered chymotrypsin, chA-rChyB-N-v3.

In addition to Asn, several peptides were generated by proteolysis at C-terminal of hydrophobic amino acids. This is in accordance with the kinetic data measured using peptide substrates and demonstrated that in addition to cleaving after Asn the engineered variant also had activity against the hydrophobic amino acids. Secondly, the engineering of the S1 pocket did not introduce any unintended but restricting preferences for other extended positions of the substrate. A few cuts after Lys/Arg were also observed possibly due to incomplete TLCK inactivation of trypsin that was used for activation of zymogens. Generation of peptides of desired size (500-2000 Da) allowed high coverage (58%) of the entire BSA sequence (608 aa long) for rChyB-N-v3 in comparison to 46% for wild-type rChyB and addition of chain A further improved the coverage to 73% in the case of chA-rChyB-N-v3 variant. Currently we are testing several glycosylated proteins

including RNase B and cytokine IL-7 for assessing the ability of chA-rChyB-N-v3 to map their glycosylation sites.

4.4 Discussion

Current proteomic technologies continue to develop at a rapid pace allowing large scale mapping of post-translation modifications in complex clinical and environmental samples. One of the key components in the workflow that has not been modified over time is the protease used for sample digestion as the proteomics community has faithfully relied only on trypsin. Proteases with engineered specificity to discriminate between post-translationally modified and unmodified residues can easily be added for tandem digestion with trypsin to provide additional information that can facilitate database search algorithms to locate PTM sites. Previously disulfide free microbial proteases have been engineered for detection of PTMs. However they were either inactive towards protein substrates due to extended specificity/conformation requirements or they were highly non-specific leading to generation of peptides that are too small (<500 Da) for MS analysis.

We developed a platform that is compatible with features such as multiple disulfide bonds and zymogen activation for display of mammalian proteases in a functional form on the surface of bacteria. Anchored surface display allowed us to assay for activity using peptide substrates containing FRET dyes which can neither diffuse through outer membrane nor be genetically encoded. Chymotrypsin which is already used in proteomic studies, especially for membrane protein samples known to contain hydrophobic regions, due to its strong P1 specificity with no extended preference otherwise, was engineered

using a combinatorial approach to cleave after Asn and the ability of engineered variants to digest protein substrates for MS/MS analysis was demonstrated. To our knowledge, this is the first time substrate binding pocket of a mammalian protease has been comprehensively engineered by screening large ($>10^7$) libraries. Prior to this study, the size of the largest library for a mammalian protease system was 150 and was screened using a plate reader (Webster et al., 2014). Our FACS-based screening method can readily be extended to human proteases of therapeutic interest or to other proteases for biotechnological applications, thereby allowing the transition of engineering efforts from site-directed mutagenesis to combinatorial approach of screening targeted libraries.

The final engineered version of chymotrypsin contained seven mutations in the substrate binding pocket. Though these mutations were verified by backcross shuffling to contribute towards Asn activity, precise role of these mutations in substrate recognition is unknown. Qualitative comparison of substrate binding pocket of the isolated variant with that of legumain (Dall and Brandstetter, 2013) (asparaginyl endopeptidase of cysteine protease family) yields similarities in zwitterionic character (Met192Arg, Gly218Arg and Val219Glu) to interact with the amide side chain of Asn residue and presence of bulky side chain (Thr224Trp) to provide steric constraints for substrate binding. We are currently pursuing experimental determination of structure of the engineered enzyme co-crystallized with a substrate analogue (Ala-Ala-Pro-Asn-CHO). Additional information obtained by structural characterization can help understand the mechanism of substrate discrimination and the degree of plasticity exhibited by the substrate binding pocket. In chymotrypsin-fold proteases, Trp215, Cys220 participating in a disulfide bond and Pro225 residues act like hinges conserved for a stable horse-shoe shaped loop

surrounding the substrate. While side chains of the amino acids constituting the loop are typically modified to engineer specificity, it appears that size of the loop can also be a tunable parameter for accommodating bulky PTMs such as phosphorylated serine.

4.5 Conclusions

Characterizing post-translational modifications (PTMs) has been of increasing interest due to their biological functions and this knowledge is used for novel target discovery and drug development. Since PTMs can be labile, proteomic techniques relying on tandem MS have difficulty in achieving comprehensive coverage of PTM sites within proteins. Proteases which selectively recognize PTMs may improve their identification in complex protein mixtures and are easily adaptable to current proteomic methods. In this study, we have utilized a combinatorial approach to engineer the substrate binding pocket of chymotrypsin, and isolate variants that can cleave after Asn residues to map N-linked glycosylation sites. The ability of engineered variants to generate peptides of a model protein (BSA) for MS/MS analysis leading to 73% sequence coverage was demonstrated. We anticipate that that our results can be the foundation for engineering novel substrate specificities onto the chymotrypsin scaffold for detecting various kinds of PTMs.

References

1. Ahmad, J., Bird, P.I., Kaiserman, D., 2014. "Analysis of the evolution of granule associated serine proteases of immune defence (GASPIDs) suggests a revised nomenclature." *Biol. Chem.* 395, 1253–62. doi:10.1515/hsz-2014-0174
2. Alvizo, O., Mittal, S., Mayo, S.L., Schiffer, C.A., 2012. "Structural, kinetic, and thermodynamic studies of specificity designed HIV-1 protease." *Protein Sci.* 21, 1029–41. doi:10.1002/pro.2086
3. Angus, D.C., Laterre, P.-F., Helterbrand, J., Ely, E.W., Ball, D.E., Garg, R., Weissfeld, L.A., Bernard, G.R., 2004. "The effect of drotrecogin alfa (activated) on long-term survival after severe sepsis." *Crit. Care Med.* 32, 2199–2206. doi:10.1097/01.CCM.0000145228.62451.F6
4. Ausubel, F.M., 1999. Short protocols in molecular biology: a compendium of methods from Current protocols in molecular biology. Wiley.
5. Azoitei, M.L., Correia, B.E., Ban, Y.-E.A., Carrico, C., Kalyuzhniy, O., Chen, L., Schroeter, A., Huang, P.-S., McLellan, J.S., Kwong, P.D., Baker, D., Strong, R.K., Schief, W.R., 2011. "Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold." *Science* 334, 373–6. doi:10.1126/science.1209368
6. Baggio, R., Shi, Y.Q., Wu, Y.Q., Abeles, R.H., 1996. "From poor substrates to good inhibitors: Design of inhibitors for serine and thiol proteases." *Biochemistry* 35, 3351–3353. doi:10.1021/bi952879e
7. Bai, J.C., Fried, M., Corazza, G.R., Schuppan, D., Farthing, M., Catassi, C., Greco, L., Cohen, H., Ciacci, C., Eliakim, R., Fasano, A., González, A.,

- Krabshuis, J.H., LeMair, A., 2013. "World Gastroenterology Organisation global guidelines on celiac disease." *J. Clin. Gastroenterol.* 47, 121–6.
doi:10.1097/MCG.0b013e31827a6f83
8. Balint, R.F., Larrick, J.W., 1993. "Antibody engineering by parsimonious mutagenesis." *Gene* 137, 109–18.
 9. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B., 2007.
"Quantitative mass spectrometry in proteomics: a critical review." *Anal. Bioanal. Chem.* 389, 1017–31. doi:10.1007/s00216-007-1486-6
 10. Bark, S.J., Muster, N., Yates, J.R., Siuzdak, G., 2001. "High-Temperature Protein Mass Mapping Using a Thermophilic Protease." *J. Am. Chem. Soc.* 123, 1774–1775. doi:10.1021/ja002909n
 11. Becker, S., Michalczyk, A., Wilhelm, S., Jaeger, K.-E., Kolmar, H., 2007.
"Ultrahigh-throughput screening to identify E. coli cells expressing functionally active enzymes on their surface." *Chembiochem* 8, 943–9.
doi:10.1002/cbic.200700020
 12. Berg, J.M., Tymoczko, J.L., Stryer, L., 2002. Many Enzymes Are Activated by Specific Proteolytic Cleavage. W H Freeman.
 13. Berges, N., Hehmann-Titt, G., Hristodorov, D., Melmer, G., Thepen, T., Barth, S., 2014. "Human Cytolytic Fusion Proteins: Modified Versions of Human Granzyme B and Angiogenin Have the Potential to Replace Bacterial Toxins in Targeted Therapies against CD64+ Diseases." *Antibodies* 3, 92–115.
doi:10.3390/antib3010092

14. Berkowitz, S.A., Engen, J.R., Mazzeo, J.R., Jones, G.B., 2012. “Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars.” *Nat. Rev. Drug Discov.* 11, 527–40. doi:10.1038/nrd3746
15. Bernard, G.R., Vincent, J.L., Laterre, P.F., LaRosa, S.P., Dhainaut, J.F., Lopez-Rodriguez, A., Steingrub, J.S., Garber, G.E., Helterbrand, J.D., Ely, E.W., Fisher, C.J., 2001. “Efficacy and safety of recombinant human activated protein C for severe sepsis.” *N. Engl. J. Med.* 344, 699–709.
doi:10.1056/NEJM200103083441001
16. Bernstein, H.D., 2007. “Are bacterial ‘autotransporters’ really transporters?” *Trends Microbiol.* 15, 441–7. doi:10.1016/j.tim.2007.09.007
17. Berny-Lang, M.A., Hurst, S., Tucker, E.I., Pelc, L.A., Wang, R.K., Hurn, P.D., Di Cera, E., McCarty, O.J.T., Gruber, A., 2011. “Thrombin mutant W215A/E217A treatment improves neurological outcome and reduces cerebral infarct size in a mouse model of ischemic stroke.” *Stroke.* 42, 1736–41.
doi:10.1161/STROKEAHA.110.603811
18. Binder, U., Matschiner, G., Theobald, I., Skerra, A., 2010. “High-throughput sorting of an Anticalin library via EspP-mediated functional display on the Escherichia coli cell surface.” *J. Mol. Biol.* 400, 783–802.
doi:10.1016/j.jmb.2010.05.049
19. Bladergroen, B.A., Meijer, C.J.L.M., Ten Berge, R.L., Hack, C.E., Muris, J.J.F., Dukers, D.F., Chott, A., Kazama, Y., Oudejans, J.J., Van Berkum, O., Kummer, J.A., 2002. “Expression of the granzyme B inhibitor, protease inhibitor 9, by tumor cells in patients with non-Hodgkin and Hodgkin lymphoma: A novel

- protective mechanism for tumor cells to circumvent the immune system?” *Blood* 99, 232–237. doi:10.1182/blood.V99.1.232
20. Bode, C., Smalling, R.W., Berg, G., Burnett, C., Lorch, G., Kalbfleisch, J.M., Chernoff, R., Christie, L.G., Feldman, R.L., Seals, A.A., Weaver, W.D., 1996. “Randomized comparison of coronary thrombolysis achieved with double-bolus reteplase (recombinant plasminogen activator) and front-loaded, accelerated alteplase (recombinant tissue plasminogen activator) in patients with acute myocardial infarction. The RA.” *Circulation* 94, 891–8.
 21. Boder, E.T., Wittrup, K.D., 1997. “Yeast surface display for screening combinatorial polypeptide libraries.” *Nat. Biotechnol.* 15, 553–7. doi:10.1038/nbt0697-553
 22. Bolt, G., Bjelke, J.R., Hermit, M.B., Hansen, L., Karpf, D.M., Kristensen, C., 2012. “Hyperglycosylation prolongs the circulation of coagulation factor IX.” *J. Thromb. Haemost.* 10, 2397–8. doi:10.1111/j.1538-7836.2012.04911.x
 23. Bommarius, A.S., Blum, J.K., Abrahamson, M.J., 2011. “Status of protein engineering for biocatalysts: how to design an industrially useful biocatalyst.” *Curr. Opin. Chem. Biol.* 15, 194–200. doi:10.1016/j.cbpa.2010.11.011
 24. Bone, R., Fujishige, A., Kettner, C.A., Agard, D.A., 1991. “Structural basis for broad specificity in alpha-lytic protease mutants.” *Biochemistry* 30, 10388–10398. doi:10.1021/bi00107a005
 25. Bonger, K.M., Chen, L., Liu, C.W., Wandless, T.J., 2011. “Small-molecule displacement of a cryptic degron causes conditional protein degradation.” *Nat. Chem. Biol.* 7, 531–7. doi:10.1038/nchembio.598

26. Borgo, B., Havranek, J.J., 2015. "Computer-aided design of a catalyst for Edman degradation utilizing substrate-assisted catalysis." *Protein Sci.* 24, 571–579.
doi:10.1002/pro.2633
27. Bornscheuer, U.T., Huisman, G.W., Kazlauskas, R.J., Lutz, S., Moore, J.C., Robins, K., 2012. "Engineering the third wave of biocatalysis." *Nature* 485, 185–194. doi:10.1038/nature11117
28. Bosley, A.D., Ostermeier, M., 2005. "Mathematical expressions useful in the construction, description and evaluation of protein libraries." *Biomol. Eng.* 22, 57–61. doi:10.1016/j.bioeng.2004.11.002
29. Boulware, K.T., Daugherty, P.S., 2006. "Protease specificity determination by using cellular libraries of peptide substrates (CLiPS)." *Proc. Natl. Acad. Sci. U. S. A.* 103, 7583–8. doi:10.1073/pnas.0511108103
30. Bowman, L.J., Anderson, C.D., Chapman, W.C., 2010. "Topical recombinant human thrombin in surgical hemostasis." *Semin. Thromb. Hemost.* 36, 477–84. doi:10.1055/s-0030-1255441
31. Brandon, L.D., Goldberg, M.B., n.d. "Periplasmic Transit and Disulfide Bond Formation of the Autotransported Shigella Protein IcsA Periplasmic Transit and Disulfide Bond Formation of the Autotransported Shigella Protein IcsA" 183. doi:10.1128/JB.183.3.951
32. Brochier, M., 1978. "Moderate doses of urokinase (UK) in the treatment of myocardial infarct and pulmonary embolism." *Ann. Anesthesiol. Fr.* 19, 735–8.

33. Buller, A.R., Townsend, C.A., 2013. “Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad.” *Proc. Natl. Acad. Sci. U. S. A.* 110, E653–61. doi:10.1073/pnas.1221050110
34. Burkhardt, J.M., Schumbrutzki, C., Wortelkamp, S., Sickmann, A., Zahedi, R.P., 2012. “Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics.” *J. Proteomics* 75, 1454–62. doi:10.1016/j.jprot.2011.11.016
35. Caputo, A., Parrish, J.C., James, M.N.G., Powers, J.C., Bleackley, R.C., 1999. “Electrostatic reversal of serine proteinase substrate specificity.” *Proteins Struct. Funct. Genet.* 35, 415–424. doi:10.1002/(SICI)1097-0134(19990601)35:4<415::AID-PROT5>3.0.CO;2-7
36. Caughey, G.H., Beauchamp, J., Schlatter, D., Raymond, W.W., Trivedi, N.N., Banner, D., Mauser, H., Fingerle, J., 2008. “Guinea pig chymase is leucine-specific: a novel example of functional plasticity in the chymase/granzyme family of serine peptidases.” *J. Biol. Chem.* 283, 13943–51. doi:10.1074/jbc.M710502200
37. Chakraborty, A., Chakrabarti, S., 2015. “A survey on prediction of specificity-determining sites in proteins.” *Brief. Bioinform.* 16, 71–88. doi:10.1093/bib/bbt092
38. Chang, A.C., Cohen, S.N., 1978. “Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid.” *J. Bacteriol.* 134, 1141–1156.

39. Choi, K.Y., Swierczewska, M., Lee, S., Chen, X., 2012. "Protease-activated drug development." *Theranostics* 2, 156–78. doi:10.7150/thno.4068
40. Clara, R.O., Soares, T.S., Torquato, R.J.S., Lima, C.A., Watanabe, R.O.M., Barros, N.M.T., Carmona, A.K., Masuda, A., Vaz, I.S., Tanaka, A.S., 2011. "Boophilus microplus cathepsin L-like (BmCL1) cysteine protease: specificity study using a peptide phage display library." *Vet. Parasitol.* 181, 291–300. doi:10.1016/j.vetpar.2011.04.003
41. Craik, C.S., Page, M.J., Madison, E.L., 2011. "Proteases as therapeutics." *Biochem. J.* 435, 1–16. doi:10.1042/BJ20100965
42. Currin, A., Swainston, N., Day, P.J., Kell, D.B., 2014. "Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently." *Chem. Soc. Rev.* 44, 1172–1239. doi:10.1039/c4cs00351a
43. Dalby, P.A., 2011. "Strategy and success for the directed evolution of enzymes." *Curr. Opin. Struct. Biol.* 21, 473–80. doi:10.1016/j.sbi.2011.05.003
44. Dall, E., Brandstetter, H., 2013. "Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation." *Proc. Natl. Acad. Sci. U. S. A.* 110, 10940–5. doi:10.1073/pnas.1300686110
45. Danino, T., Prindle, A., Kwong, G.A., Skalak, M., Li, H., Allen, K., Hasty, J., Bhatia, S.N., 2015. "Programmable probiotics for detection of cancer in urine." *Sci. Transl. Med.* 7, 289ra84. doi:10.1126/scitranslmed.aaa3519
46. Dautin, N., Bernstein, H.D., 2011. "Residues in a conserved α -helical segment are required for cleavage but not secretion of an Escherichia coli serine protease

- autotransporter passenger domain.” *J. Bacteriol.* 193, 3748–56.
doi:10.1128/JB.05070-11
47. Dautin, N., Bernstein, H.D., 2007. “Protein secretion in gram-negative bacteria via the autotransporter pathway.” *Annu. Rev. Microbiol.* 61, 89–112.
doi:10.1146/annurev.micro.61.080706.093233
 48. Davie, E.W., Fujikawa, K., Kisiel, W., 1991. “The coagulation cascade: initiation, maintenance, and regulation.” *Biochemistry* 30, 10363–70.
 49. Davies, D., 2007. Flow Cytometry. Humana Press.
 50. Desmet, J., Spriet, J., Lasters, I., 2002. “Fast and accurate side-chain topology and energy refinement (FASTER) as a new method for protein structure optimization.” *Proteins* 48, 31–43. doi:10.1002/prot.10131
 51. Desnoyers, L.R., Vasiljeva, O., Richardson, J.H., Yang, A., Menendez, E.E.M., Liang, T.W., Wong, C., Bessette, P.H., Kamath, K., Moore, S.J., Sagert, J.G., Hostetter, D.R., Han, F., Gee, J., Flandez, J., Markham, K., Nguyen, M., Krimm, M., Wong, K.R., Liu, S., Daugherty, P.S., West, J.W., Lowman, H.B., 2013. “Tumor-specific activation of an EGFR-targeting probody enhances therapeutic index.” *Sci. Transl. Med.* 5, 207ra144. doi:10.1126/scitranslmed.3006682
 52. Despotovic, D., Vojcic, L., Blanusa, M., Maurer, K.-H., Zacharias, M., Bocola, M., Martinez, R., Schwaneberg, U., 2013. “Redirecting catalysis from proteolysis to perhydrolysis in subtilisin Carlsberg.” *J. Biotechnol.* 167, 279–86.
doi:10.1016/j.jbiotec.2013.06.017
 53. Drag, M., Salvesen, G.S., 2010. “Emerging principles in protease-based drug discovery.” *Nat. Rev. Drug Discov.* 9, 690–701. doi:10.1038/nrd3053

54. Dumon-Seignovert, L., Cariot, G., Vuillard, L., 2004. "The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3)." *Protein Expr. Purif.* 37, 203–6. doi:10.1016/j.pep.2004.04.025
55. Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., Liang, J., 2006. "CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues." *Nucleic Acids Res.* 34, W116–8. doi:10.1093/nar/gkl282
56. Dunse, K.M., Kaas, Q., Guarino, R.F., Barton, P.A., Craik, D.J., Anderson, M.A., 2010. "Molecular basis for the resistance of an insect chymotrypsin to a potato type II proteinase inhibitor." *Proc. Natl. Acad. Sci. U. S. A.* 107, 15016–21. doi:10.1073/pnas.1009327107
57. Edgar, R.C., 2004. "MUSCLE: multiple sequence alignment with high accuracy and high throughput." *Nucleic Acids Res.* 32, 1792–7. doi:10.1093/nar/gkh340
58. Feijoo-Siota, L., Villa, T.G., 2010. "Native and Biotechnologically Engineered Plant Proteases with Industrial Applications." *Food Bioprocess Technol.* 4, 1066–1088. doi:10.1007/s11947-010-0431-4
59. Fersht, A., 1999. Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. W. H. Freeman.
60. Flick, M.J., Chauhan, A.K., Frederick, M., Talmage, K.E., Kombrinck, K.W., Miller, W., Mullins, E.S., Palumbo, J.S., Zheng, X., Esmon, N.L., Esmon, C.T., Thornton, S., Becker, A., Pelc, L.A., Di Cera, E., Wagner, D.D., Degen, J.L., 2011. "The development of inflammatory joint disease is attenuated in mice

- expressing the anticoagulant prothrombin mutant W215A/E217A.” *Blood* 117, 6326–37. doi:10.1182/blood-2010-08-304915
61. Franke, L., Liebscher, S., Bordusa, F., 2013. “Engineering the oxyanion hole of trypsin for promoting the reverse of proteolysis.” *J. Pept. Sci.* 128–136. doi:10.1002/psc.2597
 62. Gasparian, M.E., Bobik, T. V, Kim, Y. V, Ponomarenko, N.A., Dolgikh, D.A., Gabibov, A.G., Kirpichnikov, M.P., 2013. “Heterogeneous catalysis on the phage surface: Display of active human enteropeptidase.” *Biochimie* 95, 2076–81. doi:10.1016/j.biochi.2013.07.020
 63. Gershon, P.D., 2014. “Cleaved and missed sites for trypsin, lys-C, and lys-N can be predicted with high confidence on the basis of sequence context.” *J. Proteome Res.* 13, 702–9. doi:10.1021/pr400802z
 64. Gherardini, P.F., Wass, M.N., Helmer-Citterich, M., Sternberg, M.J.E., 2007. “Convergent evolution of enzyme active sites is not a rare phenomenon.” *J. Mol. Biol.* 372, 817–45. doi:10.1016/j.jmb.2007.06.017
 65. Gialeli, C., Theocharis, A.D., Karamanos, N.K., 2011. “Roles of matrix metalloproteinases in cancer progression and their pharmacological targeting.” *FEBS J.* 278, 16–27. doi:10.1111/j.1742-4658.2010.07919.x
 66. Gietz, R.D., Schiestl, R.H., 2007. “High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method.” *Nat. Protoc.* 2, 31–4. doi:10.1038/nprot.2007.13

67. Gordon, S.R., Stanley, E.J., Wolf, S., Toland, A., Wu, S.J., Hadidi, D., Mills, J.H., Baker, D., Pultz, I.S., Siegel, J.B., 2012. "Computational design of an α -gliadin peptidase." *J. Am. Chem. Soc.* 134, 20513–20. doi:10.1021/ja3094795
68. Gu, J., Noe, A., Chandra, P., Al-Fayoumi, S., Ligueros-Saylan, M., Sarangapani, R., Maahs, S., Ksander, G., Rigel, D.F., Jeng, A.Y., Lin, T.-H., Zheng, W., Dole, W.P., 2010. "Pharmacokinetics and pharmacodynamics of LCZ696, a novel dual-acting angiotensin receptor-neprilysin inhibitor (ARNi)." *J. Clin. Pharmacol.* 50, 401–14. doi:10.1177/0091270009343932
69. Gupta, R., Sharma, R., Beg, Q.K., 2013. "Revisiting microbial keratinases: next generation proteases for sustainable biotechnology." *Crit. Rev. Biotechnol.* 33, 216–28. doi:10.3109/07388551.2012.685051
70. Harrison, R.L., Bonning, B.C., 2010. "Proteases as insecticidal agents." *Toxins (Basel)*. 2, 935–53. doi:10.3390/toxins2050935
71. Harvey, B.R., Georgiou, G., Hayhurst, A., Jeong, K.J., Iverson, B.L., Rogers, G.K., 2004. "Anchored periplasmic expression, a versatile technology for the isolation of high-affinity antibodies from Escherichia coli-expressed libraries." *Proc. Natl. Acad. Sci. U. S. A.* 101, 9193–8. doi:10.1073/pnas.0400187101
72. Hedstrom, L., Szilagyi, L., Rutter, W.J., 1992. "Converting trypsin to chymotrypsin: the role of surface loops." *Science* 255, 1249–53.
73. Henderson, I.R., Navarro-garcia, F., n.d. "Type V Protein Secretion Pathway : the Autotransporter Story" 692–744. doi:10.1128/MMBR.68.4.692
74. Hernández-Ávila, J.E., Palacio-Mejía, L.S., Hernández-Romieu, A., Bautista-Arredondo, S., Sepúlveda Amor, J., Hernández-Ávila, M., 2015. "Effect of

Universal Access to Antiretroviral Therapy on HIV/AIDS Mortality in Mexico
1990 - 2011.” *J. Acquir. Immune Defic. Syndr.*

doi:10.1097/QAI.0000000000000645

75. Heutinck, K.M., ten Berge, I.J.M., Hack, C.E., Hamann, J., Rowshani, A.T., 2010. “Serine proteases of the human immune system in health and disease.” *Mol. Immunol.* 47, 1943–55. doi:10.1016/j.molimm.2010.04.020
76. Hittinger, C.T., Carroll, S.B., 2007. “Gene duplication and the adaptive evolution of a classic genetic switch.” *Nature* 449, 677–81. doi:10.1038/nature06151
77. Höhne, M., Schätzle, S., Jochens, H., Robins, K., Bornscheuer, U.T., 2010. “Rational assignment of key motifs for function guides in silico enzyme identification.” *Nat. Chem. Biol.* 6, 807–13. doi:10.1038/nchembio.447
78. Holmberg, H.L., Lauritzen, B., Tranholm, M., Ezban, M., 2009. “Faster onset of effect and greater efficacy of NN1731 compared with rFVIIa, aPCC and FVIII in tail bleeding in hemophilic mice.” *J. Thromb. Haemost.* 7, 1517–22. doi:10.1111/j.1538-7836.2009.03532.x
79. Horiya, S., MacPherson, I.S., Krauss, I.J., 2014. “Recent strategies targeting HIV glycans in vaccine design.” *Nat. Chem. Biol.* 10, 990–999. doi:10.1038/nchembio.1685
80. Huber, K., 2001. “Plasminogen activator inhibitor type-1 (part two): role for failure of thrombolytic therapy. PAI-1 resistance as a potential benefit for new fibrinolytic agents.” *J. Thromb. Thrombolysis* 11, 195–202.
81. Huesgen, P.F., Lange, P.F., Rogers, L.D., Solis, N., Eckhard, U., Kleifeld, O., Goulas, T., Gomis-Rüth, F.X., Overall, C.M., 2015. “LysargiNase mirrors trypsin

- for protein C-terminal and methylation-site identification.” *Nat. Methods* 12, 55–8. doi:10.1038/nmeth.3177
82. Hung, S.H., Hedstrom, L., 1998. “Converting trypsin to elastase: substitution of the S1 site and adjacent loops reconstitutes esterase specificity but not amidase activity.” *Protein Eng. Des. Sel.* 11, 669–673. doi:10.1093/protein/11.8.669
 83. Huntington, J.A., 2012. “Thrombin plasticity.” *Biochim. Biophys. Acta* 1824, 246–52. doi:10.1016/j.bbapap.2011.07.005
 84. Ieva, R., Skillman, K.M., Bernstein, H.D., 2008. “Incorporation of a polypeptide segment into the beta-domain pore during the assembly of a bacterial autotransporter.” *Mol. Microbiol.* 67, 188–201. doi:10.1111/j.1365-2958.2007.06048.x
 85. Jakob, F., Martinez, R., Mandawe, J., Hellmuth, H., Siegert, P., Maurer, K.-H., Schwaneberg, U., 2013. “Surface charge engineering of a *Bacillus gibsonii* subtilisin protease.” *Appl. Microbiol. Biotechnol.* 97, 6793–802. doi:10.1007/s00253-012-4560-8
 86. Jauch, E.C., Cucchiara, B., Adeoye, O., Meurer, W., Brice, J., Chan, Y.Y.-F., Gentile, N., Hazinski, M.F., 2010. “Part 11: adult stroke: 2010 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care.” *Circulation* 122, S818–28. doi:10.1161/CIRCULATIONAHA.110.971044
 87. Jelinek, B., Antal, J., Venekei, I., Gráf, L., 2004. “Ala226 to Gly and Ser189 to Asp mutations convert rat chymotrypsin B to a trypsin-like protease.” *Protein Eng. Des. Sel.* 17, 127–31. doi:10.1093/protein/gzh014

88. Jelinek, B., Katona, G., Fodor, K., Venekei, I., Gráf, L., 2008. "The crystal structure of a trypsin-like mutant chymotrypsin: the role of position 226 in the activity and specificity of S189D chymotrypsin." *Protein J.* 27, 79–87.
doi:10.1007/s10930-007-9110-3
89. Jeong, K.J., Seo, M.J., Iverson, B.L., Georgiou, G., 2007. "APEX 2-hybrid, a quantitative protein-protein interaction assay for antibody discovery and engineering." *Proc. Natl. Acad. Sci. U. S. A.* 104, 8247–52.
doi:10.1073/pnas.0702650104
90. Jia, X.Y., 2011. "Plasmid DNA Isolation." US 20110117641A1.
91. Jochens, H., Bornscheuer, U.T., 2010. "Natural diversity to guide focused directed evolution." *Chembiochem* 11, 1861–6. doi:10.1002/cbic.201000284
92. Johnson, T.A., Qiu, J., Plaut, A.G., Holyoak, T., 2009. "Active-site gating regulates substrate selectivity in a chymotrypsin-like serine protease the structure of haemophilus influenzae immunoglobulin A1 protease." *J. Mol. Biol.* 389, 559–74. doi:10.1016/j.jmb.2009.04.041
93. Jong, W.S.P., ten Hagen-Jongman, C.M., den Blaauwen, T., Slotboom, D.J., Tame, J.R.H., Wickström, D., de Gier, J.-W., Otto, B.R., Lührink, J., 2007. "Limited tolerance towards folded elements during secretion of the autotransporter Hbp." *Mol. Microbiol.* 63, 1524–36. doi:10.1111/j.1365-2958.2007.05605.x
94. Jørgensen, I., Kos, J., Krašovec, M., Troelsen, L., Klarlund, M., Jensen, T.W., Hansen, M.S., Jacobsen, S., 2011. "Serum cysteine proteases and their inhibitors

- in rheumatoid arthritis: relation to disease activity and radiographic progression.”
Clin. Rheumatol. 30, 633–8. doi:10.1007/s10067-010-1585-1
95. Jose, J., Krämer, J., Klauser, T., Pohlner, J., Meyer, T.F., 1996. “Absence of periplasmic DsbA oxidoreductase facilitates export of cysteine-containing passenger proteins to the *Escherichia coli* cell surface via the Iga(β) autotransporter pathway.” *Gene* 178, 107–110. doi:10.1016/0378-1119(96)00343-5
 96. Jose, J., Meyer, T.F., 2007. “The autodisplay story, from discovery to biotechnical and biomedical applications.” *Microbiol. Mol. Biol. Rev.* 71, 600–19. doi:10.1128/MMBR.00011-07
 97. Jose, J., Zangen, D., 2005. “Autodisplay of the protease inhibitor aprotinin in *Escherichia coli*.” *Biochem. Biophys. Res. Commun.* 333, 1218–26. doi:10.1016/j.bbrc.2005.06.028
 98. Junker, M., Besingi, R.N., Clark, P.L., 2009. “Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion.” *Mol. Microbiol.* 71, 1323–32. doi:10.1111/j.1365-2958.2009.06607.x
 99. Junker, M., Schuster, C.C., McDonnell, A. V, Sorg, K.A., Finn, M.C., Berger, B., Clark, P.L., 2006. “Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins.” *Proc. Natl. Acad. Sci. U. S. A.* 103, 4918–23. doi:10.1073/pnas.0507923103
 100. Kalle, E., Kubista, M., Rensing, C., 2014. “Multi-template polymerase chain reaction.” *Biomol. Detect. Quantif.* 2, 11–29. doi:10.1016/j.bdq.2014.11.002

101. Kardos, J., Bódi, Á., Závodszky, P., Venekei, I., Gráf, L., 1999. "Disulfide-Linked Propeptides Stabilize the Structure of Zymogen and Mature Pancreatic Serine Proteases †." *Biochemistry* 38, 12248–12257. doi:10.1021/bi990764v
102. Kasper B. Lauridsen, T.B., 2013. "Optimizing the Identification of Citrullinated Peptides by Mass Spectrometry: Utilizing the Inability of Trypsin to Cleave after Citrullinated Amino Acids." *J. Proteomics Bioinform.* 06, 288–295. doi:10.4172/jpb.1000293
103. Kaukinen, K., Lindfors, K., Mäki, M., 2014. "Advances in the treatment of coeliac disease: an immunopathogenic perspective." *Nat. Rev. Gastroenterol. Hepatol.* 11, 36–44. doi:10.1038/nrgastro.2013.141
104. Kawaguchi, M., Inoue, K., Iuchi, I., Nishida, M., Yasumasu, S., 2013. "Molecular co-evolution of a protease and its substrate elucidated by analysis of the activity of predicted ancestral hatching enzyme." *BMC Evol. Biol.* 13, 231. doi:10.1186/1471-2148-13-231
105. Kelley, L. a, Sternberg, M.J.E., 2009. "Protein structure prediction on the Web: a case study using the Phyre server." *Nat. Protoc.* 4, 363–71. doi:10.1038/nprot.2009.2
106. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. "The Phyre2 web portal for protein modeling, prediction and analysis." *Nat. Protoc.* 10, 845–858. doi:10.1038/nprot.2015.053
107. Kerschen, E.J., Fernandez, J.A., Cooley, B.C., Yang, X. V, Sood, R., Mosnier, L.O., Castellino, F.J., Mackman, N., Griffin, J.H., Weiler, H., 2007.

- “Endotoxemia and sepsis mortality reduction by non-anticoagulant activated protein C.” *J. Exp. Med.* 204, 2439–48. doi:10.1084/jem.20070404
108. Khersonsky, O., Tawfik, D.S., 2010. “Enzyme promiscuity: a mechanistic and evolutionary perspective.” *Annu. Rev. Biochem.* 79, 471–505.
doi:10.1146/annurev-biochem-030409-143718
 109. Kim, M.-S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S., Thomas, J.K., Muthusamy, B., Leal-Rojas, P., Kumar, P., Sahasrabudhe, N.A., Balakrishnan, L., Advani, J., George, B., Renuse, S., Selvan, L.D.N., Patil, A.H., Nanjappa, V., Radhakrishnan, A., Prasad, S., Subbannayya, T., Raju, R., Kumar, M., Sreenivasamurthy, S.K., Marimuthu, A., Sathe, G.J., Chavan, S., Datta, K.K., Subbannayya, Y., Sahu, A., Yelamanchi, S.D., Jayaram, S., Rajagopalan, P., Sharma, J., Murthy, K.R., Syed, N., Goel, R., Khan, A.A., Ahmad, S., Dey, G., Mudgal, K., Chatterjee, A., Huang, T.-C., Zhong, J., Wu, X., Shaw, P.G., Freed, D., Zahari, M.S., Mukherjee, K.K., Shankar, S., Mahadevan, A., Lam, H., Mitchell, C.J., Shankar, S.K., Satishchandra, P., Schroeder, J.T., Sirdeshmukh, R., Maitra, A., Leach, S.D., Drake, C.G., Halushka, M.K., Prasad, T.S.K., Hruban, R.H., Kerr, C.L., Bader, G.D., Iacobuzio-Donahue, C.A., Gowda, H., Pandey, A., 2014. “A draft map of the human proteome.” *Nature* 509, 575–81. doi:10.1038/nature13302
 110. Kim, S.Y., Park, K.W., Lee, Y.J., Back, S.H., Goo, J.H., Park, O.K., Jang, S.K., Park, W.J., 2000. “In vivo determination of substrate specificity of hepatitis C virus NS3 protease: genetic assay for site-specific proteolysis.” *Anal. Biochem.* 284, 42–8. doi:10.1006/abio.2000.4662

111. Kintsès, B., Hein, C., Mohamed, M.F., Fischlechner, M., Courtois, F., Lainé, C., Hollfelder, F., 2012. "Picoliter cell lysate assays in microfluidic droplet compartments for directed enzyme evolution." *Chem. Biol.* 19, 1001–9. doi:10.1016/j.chembiol.2012.06.009
112. Kishor, C., Arya, T., Reddi, R., Chen, X., Saddanapu, V., Marapaka, A.K., Gumpena, R., Ma, D., Liu, J.O., Addlagatta, A., 2013. "Identification, biochemical and structural evaluation of species-specific inhibitors against type I methionine aminopeptidases." *J. Med. Chem.* 56, 5295–305. doi:10.1021/jm400395p
113. Kitamoto, Y., Yuan, X., Wu, Q., McCourt, D.W., Sadler, J.E., 1994. "Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of a distinctive assortment of domains." *Proc. Natl. Acad. Sci.* 91, 7588–7592. doi:10.1073/pnas.91.16.7588
114. Knight, Z.A., Garrison, J.L., Chan, K., King, D.S., Shokat, K.M., 2007. "A remodelled protease that cleaves phosphotyrosine substrates." *J. Am. Chem. Soc.* 129, 11672–3. doi:10.1021/ja073875n
115. Knight, Z.A., Schilling, B., Row, R.H., Kenski, D.M., Gibson, B.W., Shokat, K.M., 2003. "Phosphospecific proteolysis for mapping sites of protein phosphorylation." *Nat. Biotechnol.* 21, 1047–54. doi:10.1038/nbt863
116. Kobayashi, Y., Shiga, T., Shibata, T., Sako, M., Maenaka, K., Koshiba, T., Mizumura, H., Oda, T., Kawabata, S., 2014. "The N-terminal Arg residue is essential for autocatalytic activation of a lipopolysaccharide-responsive protease zymogen." *J. Biol. Chem.* 289, 25987–95. doi:10.1074/jbc.M114.586933

117. Kohila, V., Jaiswal, A., Ghosh, S.S., 2012. "Rationally designed *Escherichia coli* cytosine deaminase mutants with improved specificity towards the prodrug 5-fluorocytosine for potential gene therapy applications." *Medchemcomm* 3, 1316. doi:10.1039/c2md20209c
118. Kostallas, G., Samuelson, P., 2010. "Novel fluorescence-assisted whole-cell assay for engineering and characterization of proteases and their substrates." *Appl. Environ. Microbiol.* 76, 7500–8. doi:10.1128/AEM.01558-10
119. Kramer, R.A., Zandwijken, D., Egmond, M.R., Dekker, N., 2000. "In vitro folding, purification and characterization of *Escherichia coli* outer membrane protease OmpT." *Eur. J. Biochem.* 267, 885–893. doi:10.1046/j.1432-1327.2000.01073.x
120. Kyte, J., Doolittle, R.F., 1982. "A simple method for displaying the hydropathic character of a protein." *J. Mol. Biol.* 157, 105–32.
121. Lähdeaho, M.-L., Kaukinen, K., Laurila, K., Vuotikka, P., Koivurova, O.-P., Kärjä-Lahdensuu, T., Marcantonio, A., Adelman, D.C., Mäki, M., 2014. "Glutenase ALV003 attenuates gluten-induced mucosal injury in patients with celiac disease." *Gastroenterology* 146, 1649–58. doi:10.1053/j.gastro.2014.02.031
122. Lamkanfi, M., Festjens, N., Declercq, W., Vanden Berghe, T., Vandenabeele, P., 2007. "Caspases in cell survival, proliferation and differentiation." *Cell Death Differ.* 14, 44–55. doi:10.1038/sj.cdd.4402047
123. Law, R.H.P., Zhang, Q., McGowan, S., Buckle, A.M., Silverman, G.A., Wong, W., Rosado, C.J., Langendorf, C.G., Pike, R.N., Bird, P.I., Whisstock, J.C., 2006.

- “An overview of the serpin superfamily.” *Genome Biol.* 7, 216. doi:10.1186/gb-2006-7-5-216
124. Lee, C.-H., Park, K.-J., Sung, E.-S., Kim, A., Choi, J.-D., Kim, J.-S., Kim, S.-H., Kwon, M.-H., Kim, Y.-S., 2010. “Engineering of a human kringle domain into agonistic and antagonistic binding proteins functioning in vitro and in vivo.” *Proc. Natl. Acad. Sci. U. S. A.* 107, 9567–71. doi:10.1073/pnas.1001541107
 125. Leysath, C.E., Monzingo, A.F., Maynard, J. a., Barnett, J., Georgiou, G., Iverson, B.L., Robertus, J.D., 2009. “Crystal Structure of the Engineered Neutralizing Antibody M18 Complexed to Domain 4 of the Anthrax Protective Antigen.” *J. Mol. Biol.* 387, 680–693. doi:10.1016/j.jmb.2009.02.003
 126. Leyton, D.L., de Luna, M.D.G., Sevastsyanovich, Y.R., Tveen Jensen, K., Browning, D.F., Scott-Tucker, A., Henderson, I.R., 2010. “The unusual extended signal peptide region is not required for secretion and function of an *Escherichia coli* autotransporter.” *FEMS Microbiol. Lett.* 311, 133–9. doi:10.1111/j.1574-6968.2010.02081.x
 127. Leyton, D.L., Rossiter, A.E., Henderson, I.R., 2012. “From self sufficiency to dependence: mechanisms and factors important for autotransporter biogenesis.” *Nat. Rev. Microbiol.* 10, 213–25. doi:10.1038/nrmicro2733
 128. Leyton, D.L., Sevastsyanovich, Y.R., Browning, D.F., Rossiter, A.E., Wells, T.J., Fitzpatrick, R.E., Overduin, M., Cunningham, A.F., Henderson, I.R., 2011. “Size and conformation limits to secretion of disulfide-bonded loops in autotransporter proteins.” *J. Biol. Chem.* 286, 42283–91. doi:10.1074/jbc.M111.306118

129. Li, C., Zhu, Y., Benz, I., Schmidt, M.A., Chen, W., Mulchandani, A., Qiao, C., 2008. "Presentation of functional organophosphorus hydrolase fusions on the surface of *Escherichia coli* by the AIDA-I autotransporter pathway." *Biotechnol. Bioeng.* 99, 485–490. doi:10.1002/bit.21548
130. Li, Q., Yi, L., Marek, P., Iverson, B.L., 2013. "Commercial proteases: present and future." *FEBS Lett.* 587, 1155–63. doi:10.1016/j.febslet.2012.12.019
131. Lindenburg, L., Merckx, M., 2014. "Engineering genetically encoded FRET sensors." *Sensors (Basel).* 14, 11691–713. doi:10.3390/s140711691
132. Lise, S., Archambeau, C., Pontil, M., Jones, D.T., 2009. "Prediction of hot spot residues at protein-protein interfaces by combining machine learning and energy-based methods." *BMC Bioinformatics* 10, 365. doi:10.1186/1471-2105-10-365
133. Liu, J., Wang, Z., He, J., Wang, G., Zhang, R., Zhao, B., 2014. "Effect of site-specific PEGylation on the fibrinolytic activity, immunogenicity, and pharmacokinetics of staphylokinase." *Acta Biochim. Biophys. Sin. (Shanghai).* 46, 782–91. doi:10.1093/abbs/gmu068
134. Liu, K., Zhu, F., Zhu, L., Chen, G., He, B., 2015. "Highly efficient enzymatic synthesis of Z-aspartame in aqueous medium via in situ product removal." *Biochem. Eng. J.* 98, 63–67. doi:10.1016/j.bej.2015.01.012
135. Losasso, V., Schiffer, S., Barth, S., Carloni, P., 2012. "Design of human granzyme B variants resistant to serpin B9." *Proteins* 80, 2514–22. doi:10.1002/prot.24133
136. Lundby, A., Andersen, M.N., Steffensen, A.B., Horn, H., Kelstrup, C.D., Francavilla, C., Jensen, L.J., Schmitt, N., Thomsen, M.B., Olsen, J. V, 2013. "In

- vivo phosphoproteomics analysis reveals the cardiac targets of β -adrenergic receptor signaling.” *Sci. Signal.* 6, rs11. doi:10.1126/scisignal.2003506
137. Ma, W., Tang, C., Lai, L., 2005. “Specificity of Trypsin and Chymotrypsin: Loop-Motion-Controlled Dynamic Correlation as a Determinant.” *Biophys. J.* 89, 1183–1193. doi:10.1529/biophysj.104.057158
 138. Mac Sweeney, A., Birrane, G., Walsh, M. a., O’Connell, T., Malthouse, J.P.G., Higgins, T.M., 2000. “Crystal structure of δ -chymotrypsin bound to a peptidyl chloromethyl ketone inhibitor.” *Acta Crystallogr. Sect. D Biol. Crystallogr.* 56, 280–286. doi:10.1107/S0907444999016583
 139. Madani, R., Poirier, R., Wolfer, D.P., Welzl, H., Groscurth, P., Lipp, H.-P., Lu, B., El Mouedden, M., Mercken, M., Nitsch, R.M., Mohajeri, M.H., 2006. “Lack of neprilysin suffices to generate murine amyloid-like deposits in the brain and behavioral deficit in vivo.” *J. Neurosci. Res.* 84, 1871–8. doi:10.1002/jnr.21074
 140. Madison, E.L., Goldsmith, E.J., Gerard, R.D., Gething, M.J., Sambrook, J.F., 1989. “Serpine-resistant mutants of human tissue-type plasminogen activator.” *Nature* 339, 721–4. doi:10.1038/339721a0
 141. Madison, E.L., Goldsmith, E.J., Gerard, R.D., Gething, M.J., Sambrook, J.F., Bassel-Duby, R.S., 1990. “Amino acid residues that affect interaction of tissue-type plasminogen activator with plasminogen activator inhibitor 1.” *Proc. Natl. Acad. Sci.* 87, 3530–3533. doi:10.1073/pnas.87.9.3530
 142. Maguer-Satta, V., Besançon, R., Bachelard-Cascales, E., 2011. “Concise review: neutral endopeptidase (CD10): a multifaceted environment actor in stem cells,

physiological mechanisms, and cancer.” *Stem Cells* 29, 389–96.

doi:10.1002/stem.592

143. Mahlangu, J., Powell, J.S., Ragni, M. V, Chowdary, P., Josephson, N.C., Pabinger, I., Hanabusa, H., Gupta, N., Kulkarni, R., Fogarty, P., Perry, D., Shapiro, A., Pasi, K.J., Apte, S., Nestorov, I., Jiang, H., Li, S., Neelakantan, S., Cristiano, L.M., Goyal, J., Sommer, J.M., Dumont, J.A., Dodd, N., Nugent, K., Vigliani, G., Luk, A., Brennan, A., Pierce, G.F., 2014. “Phase 3 study of recombinant factor VIII Fc fusion protein in severe hemophilia A.” *Blood* 123, 317–25. doi:10.1182/blood-2013-10-529974
144. Mangan, M.S.J., Kaiserman, D., Bird, P.I., 2008. “The role of serpins in vertebrate immunity.” *Tissue Antigens* 72, 1–10. doi:10.1111/j.1399-0039.2008.01059.x
145. Mannucci, P.M., 2015. “Half-life extension technologies for haemostatic agents.” *Thromb. Haemost.* 113, 165–76. doi:10.1160/TH14-04-0332
146. Marguet, D., Baggio, L., Kobayashi, T., Bernard, A.M., Pierres, M., Nielsen, P.F., Ribet, U., Watanabe, T., Drucker, D.J., Wagtman, N., 2000. “Enhanced insulin secretion and improved glucose tolerance in mice lacking CD26.” *Proc. Natl. Acad. Sci. U. S. A.* 97, 6874–9. doi:10.1073/pnas.120069197
147. Marín, E., Bodelón, G., Fernández, L.Á., 2010. “Comparative analysis of the biochemical and functional properties of C-terminal domains of autotransporters.” *J. Bacteriol.* 192, 5588–602. doi:10.1128/JB.00432-10

148. Marino, F., Pelc, L.A., Vogt, A., Gandhi, P.S., Di Cera, E., 2010. "Engineering thrombin for selective specificity toward protein C and PAR1." *J. Biol. Chem.* 285, 19145–52. doi:10.1074/jbc.M110.119875
149. Mosnier, L.O., Yang, X. V, Griffin, J.H., 2007a. "Activated protein C mutant with minimal anticoagulant activity, normal cytoprotective activity, and preservation of thrombin activable fibrinolysis inhibitor-dependent cytoprotective functions." *J. Biol. Chem.* 282, 33022–33. doi:10.1074/jbc.M705824200
150. Mosnier, L.O., Zlokovic, B. V, Griffin, J.H., 2007b. "The cytoprotective protein C pathway." *Blood* 109, 3161–72. doi:10.1182/blood-2006-09-003004
151. Mueller, S., Coleman, J.R., Wimmer, E., 2009. "Putting synthesis into biology: a viral view of genetic engineering through de novo gene and genome synthesis." *Chem. Biol.* 16, 337–47. doi:10.1016/j.chembiol.2009.03.002
152. Mullins, J.J., Peters, J., Ganten, D., 1990. "Fulminant hypertension in transgenic rats harbouring the mouse Ren-2 gene." *Nature* 344, 541–4. doi:10.1038/344541a0
153. Murakami, K., Trakselis, M.A., 2013. Nucleic Acid Polymerases. Springer Science & Business Media.
154. Najafian, L., Babji, A.S., 2012. "A review of fish-derived antioxidant and antimicrobial peptides: their production, assessment, and applications." *Peptides* 33, 178–85. doi:10.1016/j.peptides.2011.11.013
155. Nguyen, H.B., Rivers, E.P., Abrahamian, F.M., Moran, G.J., Abraham, E., Trzeciak, S., Huang, D.T., Osborn, T., Stevens, D., Talan, D.A., 2006. "Severe sepsis and septic shock: review of the literature and emergency department

- management guidelines.” *Ann. Emerg. Med.* 48, 28–54.
doi:10.1016/j.annemergmed.2006.02.015
156. Nguyen, P.-C.T., Lewis, K.B., Ettinger, R.A., Schuman, J.T., Lin, J.C., Healey, J.F., Meeks, S.L., Lollar, P., Pratt, K.P., 2014. “High-resolution mapping of epitopes on the C2 domain of factor VIII by analysis of point mutants using surface plasmon resonance.” *Blood* 123, 2732–9. doi:10.1182/blood-2013-09-527275
 157. Nhan, N.T., Gonzalez de Valdivia, E., Gustavsson, M., Hai, T.N., Larsson, G., 2011. “Surface display of Salmonella epitopes in Escherichia coli and Staphylococcus carnosus.” *Microb. Cell Fact.* 10, 22. doi:10.1186/1475-2859-10-22
 158. Nikaido, H., Vaara, M., 1985. “Molecular basis of bacterial outer membrane permeability.” *Microbiol. Rev.* 49, 1–32.
 159. Olsen, M.J., Stephens, D., Griffiths, D., Daugherty, P., Georgiou, G., Iverson, B.L., 2000. “Function-based isolation of novel enzymes from a large library.” *Nat. Biotechnol.* 18, 1071–4. doi:10.1038/80267
 160. Olsen, J. V, Mann, M., 2013. “Status of large-scale analysis of post-translational modifications by mass spectrometry.” *Mol. Cell. Proteomics* 12, 3444–52.
doi:10.1074/mcp.O113.034181
 161. Otto, B.R., Sijbrandi, R., Luirink, J., Oudega, B., Heddle, J.G., Mizutani, K., Park, S.-Y., Tame, J.R.H., 2005. “Crystal structure of hemoglobin protease, a heme binding autotransporter protein from pathogenic Escherichia coli.” *J. Biol. Chem.* 280, 17339–45. doi:10.1074/jbc.M412885200

162. Oudega, B., 2003. Protein Secretion Pathways in Bacteria. Springer Netherlands.
163. Overall, C.M., Dean, R.A., 2006. "Degradomics: systems biology of the protease web. Pleiotropic roles of MMPs in cancer." *Cancer Metastasis Rev.* 25, 69–75.
doi:10.1007/s10555-006-7890-0
164. Pallen, M.J., Chaudhuri, R.R., Henderson, I.R., 2003. "Genomic analysis of secretion systems." *Curr. Opin. Microbiol.* 6, 519–27.
165. Patthy, L., 1985. "Evolution of the proteases of blood coagulation and fibrinolysis by assembly from modules." *Cell* 41, 657–663. doi:10.1016/S0092-8674(85)80046-5
166. Pavelka, A., Chovancova, E., Damborsky, J., 2009. "HotSpot Wizard: a web server for identification of hot spots in protein engineering." *Nucleic Acids Res.* 37, W376–83. doi:10.1093/nar/gkp410
167. Pavlova, M., Klvana, M., Prokop, Z., Chaloupkova, R., Banas, P., Otyepka, M., Wade, R.C., Tsuda, M., Nagata, Y., Damborsky, J., 2009. "Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate." *Nat. Chem. Biol.* 5, 727–33. doi:10.1038/nchembio.205
168. Perona, J.J., Craik, C.S., 1997. "Evolutionary Divergence of Substrate Specificity within the Chymotrypsin-like Serine Protease Fold." *J. Biol. Chem.* 272, 29987–29990. doi:10.1074/jbc.272.48.29987
169. Pohlner, J., Halter, R., Beyreuther, K., Meyer, T.F.,. "Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease." *Nature* 325, 458–62. doi:10.1038/325458a0

170. Pozzi, N., Chen, Z., Zapata, F., Niu, W., Barranco-Medina, S., Pelc, L.A., Di Cera, E., 2013. "Autoactivation of thrombin precursors." *J. Biol. Chem.* 288, 11601–10. doi:10.1074/jbc.M113.451542
171. Prindle, A., Samayoa, P., Razinkov, I., Danino, T., Tsimring, L.S., Hasty, J., 2012. "A sensing array of radically coupled genetic 'biopixels'." *Nature* 481, 39–44. doi:10.1038/nature10722
172. Pyo, H.-M., Kim, I.-J., Kim, S.-H., Kim, H.-S., Cho, S.-D., Cho, I.-S., Hyun, B.-H., 2009. "Escherichia coli expressing single-chain Fv on the cell surface as a potential prophylactic of porcine epidemic diarrhea virus." *Vaccine* 27, 2030–6. doi:10.1016/j.vaccine.2009.01.130
173. R. Corey, D., Shiau, A.K., Qing, Y., Janowski, B.A., Craik, C.S., 1993. "Trypsin display on the surface of bacteriophage." *Gene* 128, 129–134. doi:10.1016/0378-1119(93)90163-W
174. Ramesh, B., Sendra, V.G., Cirino, P.C., Varadarajan, N., 2012. "Single-cell characterization of autotransporter-mediated Escherichia coli surface display of disulfide bond-containing proteins." *J. Biol. Chem.* 287, 38580–9. doi:10.1074/jbc.M112.388199
175. Ranieri, V.M., Thompson, B.T., Barie, P.S., Dhainaut, J.-F., Douglas, I.S., Finfer, S., Gårdlund, B., Marshall, J.C., Rhodes, A., Artigas, A., Payen, D., Tenhunen, J., Al-Khalidi, H.R., Thompson, V., Janes, J., Macias, W.L., Vangerow, B., Williams, M.D., 2012. "Drotrecogin alfa (activated) in adults with septic shock." *N. Engl. J. Med.* 366, 2055–64. doi:10.1056/NEJMoa1202290

176. Ratnikov, B.I., Cieplak, P., Gramatikoff, K., Pierce, J., Eroshkin, A., Igarashi, Y., Kazanov, M., Sun, Q., Godzik, A., Osterman, A., Stec, B., Strongin, A., Smith, J.W., 2014. "Basis for substrate recognition and distinction by matrix metalloproteinases." *Proc. Natl. Acad. Sci.* doi:10.1073/pnas.1406134111
177. Rau, J.C., Beaulieu, L.M., Huntington, J.A., Church, F.C., 2007. "Serpins in thrombosis, hemostasis and fibrinolysis." *J. Thromb. Haemost.* doi:10.1111/j.1538-7836.2006.02283.x
178. Rauh, D., Reyda, S., Klebe, G., Stubbs, M.T., 2002. "Trypsin Mutants for Structure-Based Drug Design: Expression, Refolding and Crystallisation." *Biol. Chem.* 383, 1309–1314.
179. Rawlings, N.D., Waller, M., Barrett, A.J., Bateman, A., 2014. "MEROPS: the database of proteolytic enzymes, their substrates and inhibitors." *Nucleic Acids Res.* 42, D503–9. doi:10.1093/nar/gkt953
180. Ray, M., Hostetter, D.R., Loeb, C.R.K., Simko, J., Craik, C.S., 2012. "Inhibition of Granzyme B by PI-9 protects prostate cancer cells from apoptosis." *Prostate* 72, 846–55. doi:10.1002/pros.21486
181. Razatos, A., Ong, Y.-L., Sharma, M.M., Georgiou, G., 1998. "Molecular determinants of bacterial adhesion monitored by atomic force microscopy." *Proc. Natl. Acad. Sci.* 95, 11059–11064. doi:10.1073/pnas.95.19.11059
182. Reetz, M.T., Kahakeaw, D., Lohmer, R., 2008. "Addressing the numbers problem in directed evolution." *Chembiochem* 9, 1797–804. doi:10.1002/cbic.200800298

183. Renicke, C., Schuster, D., Usherenko, S., Essen, L.-O., Taxis, C., 2013. "A LOV2 Domain-Based Optogenetic Tool to Control Protein Degradation and Cellular Function." *Chem. Biol.* 20, 619–626. doi:10.1016/j.chembiol.2013.03.005
184. Reyda, S., Sohn, C., Klebe, G., Rall, K., Ullmann, D., Jakubke, H.-D., Stubbs, M.T., 2003. "Reconstructing the Binding Site of Factor Xa in Trypsin Reveals Ligand-induced Structural Plasticity." *J. Mol. Biol.* 325, 963–977. doi:10.1016/S0022-2836(02)01337-2
185. Rice, T.W., Bernard, G.R., 2004. "Drotrecogin alfa (activated) for the treatment of severe sepsis and septic shock." *Am. J. Med. Sci.* 328, 205–14.
186. Roques, B.P., Fournié-Zaluski, M.-C., Wurm, M., 2012. "Inhibiting the breakdown of endogenous opioids and cannabinoids to alleviate pain." *Nat. Rev. Drug Discov.* 11, 292–310. doi:10.1038/nrd3673
187. Rose, T., Di Cera, E., 2002. "Substrate recognition drives the evolution of serine proteases." *J. Biol. Chem.* 277, 19243–6. doi:10.1074/jbc.C200132200
188. Roxas, M., 2008. "The role of enzyme supplementation in digestive disorders." *Altern. Med. Rev.* 13, 307–14.
189. Rutherford, N., Mourez, M., 2006. "Surface display of proteins by gram-negative bacterial autotransporters." *Microb. Cell Fact.* 5, 22. doi:10.1186/1475-2859-5-22
190. Safari, R., Motamedzadegan, A., Ovissipour, M., Regenstein, J.M., Gildberg, A., Rasco, B., 2009. "Use of Hydrolysates from Yellowfin Tuna (*Thunnus albacares*) Heads as a Complex Nitrogen Source for Lactic Acid Bacteria." *Food Bioprocess Technol.* 5, 73–79. doi:10.1007/s11947-009-0225-8

191. Saito, H., 2005. "Role of mast cell proteases in tissue remodeling." *Chem. Immunol. Allergy* 87, 80–4. doi:10.1159/000087572
192. Sander, L.E., Davis, M.J., Boekschoten, M. V, Amsen, D., Dascher, C.C., Ryffel, B., Swanson, J.A., Müller, M., Blander, J.M., 2011. "Detection of prokaryotic mRNA signifies microbial viability and promotes immunity." *Nature* 474, 385–9. doi:10.1038/nature10072
193. Sauri, A., Soprova, Z., Wickström, D., de Gier, J.-W., Van der Schors, R.C., Smit, A.B., Jong, W.S.P., Luirink, J., 2009. "The Bam (Omp85) complex is involved in secretion of the autotransporter haemoglobin protease." *Microbiology* 155, 3982–91. doi:10.1099/mic.0.034991-0
194. Schechter, I., Berger, A., 1967. "On the size of the active site in proteases. I. Papain." *Biochem. Biophys. Res. Commun.* 27, 157–62.
195. Schiffer, S., Hansen, H.P., Hehmann-Titt, G., Huhn, M., Fischer, R., Barth, S., Thepen, T., 2013. "Efficacy of an adapted granzyme B-based anti-CD30 cytolytic fusion protein against PI-9-positive classical Hodgkin lymphoma cells in a murine model." *Blood Cancer J.* 3, e106. doi:10.1038/bcj.2013.4
196. Schiffer, S., Rosinke, R., Jost, E., Hehmann-Titt, G., Huhn, M., Melmer, G., Barth, S., Thepen, T., 2014. "Targeted ex vivo reduction of CD64-positive monocytes in chronic myelomonocytic leukemia and acute myelomonocytic leukemia using human granzyme B-based cytolytic fusion proteins." *Int. J. Cancer* 135, 1497–508. doi:10.1002/ijc.28786
197. Schilling, O., auf dem Keller, U., Overall, C.M., 2011a. "Factor Xa subsite mapping by proteome-derived peptide libraries improved using WebPICS, a

- resource for proteomic identification of cleavage sites.” *Biol. Chem.* 392, 1031–7.
doi:10.1515/BC.2011.158
198. Schilling, O., Huesgen, P.F., Barré, O., Auf dem Keller, U., Overall, C.M., 2011b. “Characterization of the prime and non-prime active site specificities of proteases by proteome-derived peptide libraries and tandem mass spectrometry.” *Nat. Protoc.* 6, 111–20. doi:10.1038/nprot.2010.178
 199. Schilling, O., Overall, C.M., 2008. “Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites.” *Nat. Biotechnol.* 26, 685–94. doi:10.1038/nbt1408
 200. Schultheiss, E., Weiss, S., Winterer, E., Maas, R., Heinzle, E., Jose, J., 2008. “Esterase autodisplay: enzyme engineering and whole-cell activity determination in microplates with pH sensors.” *Appl. Environ. Microbiol.* 74, 4782–91.
doi:10.1128/AEM.01575-07
 201. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., Serrano, L., 2005. “The FoldX web server: an online force field.” *Nucleic Acids Res.* 33, W382–8.
doi:10.1093/nar/gki387
 202. Sebestova, E., Bendl, J., Brezovsky, J., Damborsky, J., 2014. “Computational tools for designing smart libraries.” *Methods Mol. Biol.* 1179, 291–314.
doi:10.1007/978-1-4939-1053-3_20
 203. Seeds, N.W., Siconolfi, L.B., Haffke, S.P., 1997. “Neuronal extracellular proteases facilitate cell migration, axonal growth, and pathfinding.” *Cell Tissue Res.* 290, 367–70.

204. Sellamuthu, S., Shin, B.H., Lee, E.S., Rho, S.-H., Hwang, W., Lee, Y.J., Han, H.-E., Kim, J. Il, Park, W.J., 2008. "Engineering of protease variants exhibiting altered substrate specificity." *Biochem. Biophys. Res. Commun.* 371, 122–6. doi:10.1016/j.bbrc.2008.04.026
205. Shi, L., 1995. "Rapid Identification of Highly Active and Selective Substrates for Stromelysin and Matrilysin Using Bacteriophage Peptide Display Libraries." *J. Biol. Chem.* 270, 6440–6449. doi:10.1074/jbc.270.12.6440
206. Shibasaki, S., Ueda, M., Ye, K., Shimizu, K., Kamasawa, N., Osumi, M., Tanaka, A., 2001. "Creation of cell surface-engineered yeast that display different fluorescent proteins in response to the glucose concentration." *Appl. Microbiol. Biotechnol.* 57, 528–33. doi:10.1007/s002530100767
207. Sjostrom, S.L., Bai, Y., Huang, M., Liu, Z., Nielsen, J., Joensson, H.N., Andersson Svahn, H., 2014. "High-throughput screening for industrial enzyme production hosts by droplet microfluidics." *Lab Chip* 14, 806–13. doi:10.1039/c3lc51202a
208. Skillman, K.M., Barnard, T.J., Peterson, J.H., Ghirlando, R., Bernstein, H.D., 2005. "Efficient secretion of a folded protein domain by a monomeric bacterial autotransporter." *Mol. Microbiol.* 58, 945–58. doi:10.1111/j.1365-2958.2005.04885.x
209. Smith, M.R., Khera, E., Wen, F., 2015. "Engineering Novel and Improved Biocatalysts by Cell Surface Display." *Ind. Eng. Chem. Res.* 54, 4021–4032. doi:10.1021/ie504071f

210. Spronk, H.M.H., de Jong, A.M., Crijns, H.J., Schotten, U., Van Gelder, I.C., Ten Cate, H., 2014. "Pleiotropic effects of factor Xa and thrombin: what to expect from novel anticoagulants." *Cardiovasc. Res.* 101, 344–51.
doi:10.1093/cvr/cvt343
211. Srinivasula, S.M., Ahmad, M., MacFarlane, M., Luo, Z., Huang, Z., Fernandes-Alnemri, T., Alnemri, E.S., 1998. "Generation of Constitutively Active Recombinant Caspases-3 and -6 by Rearrangement of Their Subunits." *J. Biol. Chem.* 273, 10107–10111. doi:10.1074/jbc.273.17.10107
212. Stawiski, E.W., Baucom, A.E., Lohr, S.C., Gregoret, L.M., 2000. "Predicting protein function from structure: unique structural features of proteases." *Proc. Natl. Acad. Sci. U. S. A.* 97, 3954–8. doi:10.1073/pnas.070548997
213. Stemmer, W.P., 1994. "DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution." *Proc. Natl. Acad. Sci. U. S. A.* 91, 10747–51.
214. Stevenson, E.K., Rubenstein, A.R., Radin, G.T., Wiener, R.S., Walkey, A.J., 2014. "Two decades of mortality trends among patients with severe sepsis: a comparative meta-analysis." *Crit. Care Med.* 42, 625–31.
doi:10.1097/CCM.0000000000000026
215. Strahl, B.D., Allis, C.D., 2000. "The language of covalent histone modifications." *Nature* 403, 41–5. doi:10.1038/47412
216. Sun, H., He, B., Xu, J., Wu, B., Ouyang, P., 2011. "Efficient chemo-enzymatic synthesis of endomorphin-1 using organic solvent stable proteases to green the synthesis of the peptide." *Green Chem.* 13, 1680. doi:10.1039/c1gc15042a

217. Szabady, R.L., Peterson, J.H., Skillman, K.M., Bernstein, H.D., 2005. "An unusual signal peptide facilitates late steps in the biogenesis of a bacterial autotransporter." *Proc. Natl. Acad. Sci. U. S. A.* 102, 221–226.
doi:10.1073/pnas.0406055102
218. Tachias, K., 1997. "Variants of Tissue-type Plasminogen Activator That Display Extraordinary Resistance to Inhibition by the Serpin Plasminogen Activator Inhibitor Type 1." *J. Biol. Chem.* 272, 14580–14585.
doi:10.1074/jbc.272.23.14580
219. Tanco, S., Lorenzo, J., Garcia-Pardo, J., Degroeve, S., Martens, L., Aviles, F.X., Gevaert, K., Van Damme, P., 2013. "Proteome-derived peptide libraries to study the substrate specificity profiles of carboxypeptidases." *Mol. Cell. Proteomics* 12, 2096–110. doi:10.1074/mcp.M112.023234
220. Tang, S.-Y., Qian, S., Akinterinwa, O., Frei, C.S., Gredell, J.A., Cirino, P.C., 2013. "Screening for enhanced triacetic acid lactone production by recombinant *Escherichia coli* expressing a designed triacetic acid lactone reporter." *J. Am. Chem. Soc.* 135, 10099–103. doi:10.1021/ja402654z
221. Tanswell, P., Modi, N., Combs, D., Danays, T., 2002. "Pharmacokinetics and pharmacodynamics of tenecteplase in fibrinolytic therapy of acute myocardial infarction." *Clin. Pharmacokinet.* 41, 1229–45. doi:10.2165/00003088-200241150-00001
222. Thie, H., Meyer, T., Schirrmann, T., Hust, M., Dübel, S., 2008. "Phage display derived therapeutic antibodies." *Curr. Pharm. Biotechnol.* 9, 439–46.

223. Tseng, T.-T., Tyler, B.M., Setubal, J.C., 2009. "Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology." *BMC Microbiol.* 9 Suppl 1, S2. doi:10.1186/1471-2180-9-S1-S2
224. Tu, R., Martinez, R., Prodanovic, R., Klein, M., Schwaneberg, U., 2011a. "A flow cytometry-based screening system for directed evolution of proteases." *J. Biomol. Screen.* 16, 285–94. doi:10.1177/10870571110396361
225. Tu, R., Martinez, R., Prodanovic, R., Klein, M., Schwaneberg, U., 2011b. "A flow cytometry-based screening system for directed evolution of proteases." *J. Biomol. Screen.* 16, 285–94. doi:10.1177/10870571110396361
226. Tyndall, J.D.A., Nall, T., Fairlie, D.P., 2005. "Proteases universally recognize beta strands in their active sites." *Chem. Rev.* 105, 973–99. doi:10.1021/cr040669e
227. Valls, M., Atrian, S., de Lorenzo, V., Fernández, L.A., 2000. "Engineering a mouse metallothionein on the cell surface of *Ralstonia eutropha* CH34 for immobilization of heavy metals in soil." *Nat. Biotechnol.* 18, 661–5. doi:10.1038/76516
228. Van den Berg, B., 2010. "Crystal structure of a full-length autotransporter." *J. Mol. Biol.* 396, 627–33. doi:10.1016/j.jmb.2009.12.061
229. Van der Woude, M.W., Henderson, I.R., 2008. "Regulation and function of Ag43 (flu)." *Annu. Rev. Microbiol.* 62, 153–69. doi:10.1146/annurev.micro.62.081307.162938
230. Van Houdt, I.S., Oudejans, J.J., van den Eertwegh, A.J.M., Baars, A., Vos, W., Bladergroen, B.A., Rimoldi, D., Muris, J.J.F., Hooijberg, E., Gundy, C.M.,

- Meijer, C.J.L.M., Kummer, J.A., 2005. "Expression of the apoptosis inhibitor protease inhibitor 9 predicts clinical outcome in vaccinated patients with stage III and IV melanoma." *Clin. Cancer Res.* 11, 6400–7. doi:10.1158/1078-0432.CCR-05-0306
231. Van Vliet, L.D., Colin, P.-Y., Hollfelder, F., 2015. "Bioinspired genotype-phenotype linkages: mimicking cellular compartmentalization for the engineering of functional proteins." *Interface Focus* 5, 20150035–20150035. doi:10.1098/rsfs.2015.0035
232. Varadarajan, N., 2006. Engineering highly active and specific protease variants. University of Texas, Austin.
233. Varadarajan, N., Cantor, J.R., Georgiou, G., Iverson, B.L., 2009a. "Construction and flow cytometric screening of targeted enzyme libraries." *Nat. Protoc.* doi:10.1038/nprot.2009.60
234. Varadarajan, N., Gam, J., Olsen, M.J., Georgiou, G., Iverson, B.L., 2005. "Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity." *Proc. Natl. Acad. Sci. U. S. A.* 102, 6855–60. doi:10.1073/pnas.0500063102
235. Varadarajan, N., Pogson, M., Georgiou, G., Iverson, B.L., 2009b. "Proteases that can distinguish among different post-translational forms of tyrosine engineered using multicolor flow cytometry." *J. Am. Chem. Soc.* 131, 18186–90. doi:10.1021/ja907803k

236. Varadarajan, N., Rodriguez, S., Hwang, B.-Y., Georgiou, G., Iverson, B.L., 2008. “Highly active and selective endopeptidases with programmed substrate specificities.” *Nat. Chem. Biol.* 4, 290–4. doi:10.1038/nchembio.80
237. Veiga, E., de Lorenzo, V., Fernández, L.A., 2004. “Structural tolerance of bacterial autotransporters for folded passenger protein domains.” *Mol. Microbiol.* 52, 1069–80. doi:10.1111/j.1365-2958.2004.04014.x
238. Veiga, E., de Lorenzo, V., Fernández, L.A., 1999. “Probing secretion and translocation of a beta-autotransporter using a reporter single-chain Fv as a cognate passenger domain.” *Mol. Microbiol.* 33, 1232–43.
239. Venekei, I., Gráf, L., Rutter, W.J., 1996a. “Expression of rat chymotrypsinogen in yeast: a study on the structural and functional significance of the chymotrypsinogen propeptide.” *FEBS Lett.* 379, 139–142. doi:10.1016/0014-5793(95)01483-7
240. Venekei, I., Szilágyi, L., Gráf, L., Rutter, W.J., 1996b. “Attempts to convert chymotrypsin to trypsin.” *FEBS Lett.* 379, 143–147. doi:10.1016/0014-5793(95)01484-5
241. Verbout, N.G., Yu, X., Healy, L.D., Phillips, K.G., Tucker, E.I., Gruber, A., McCarty, O.J.T., Offner, H., 2015. “Thrombin mutant W215A/E217A treatment improves neurological outcome and attenuates central nervous system damage in experimental autoimmune encephalomyelitis.” *Metab. Brain Dis.* 30, 57–65. doi:10.1007/s11011-014-9558-8
242. Verma, R., Schwaneberg, U., Roccatano, D., 2012. “Computer-Aided Protein Directed Evolution: a Review of Web Servers, Databases and other

- Computational Tools for Protein Engineering.” *Comput. Struct. Biotechnol. J.* 2, e201209008. doi:10.5936/csbj.201209008
243. Vincents, B., von Pawel-Rammingen, U., Björck, L., Abrahamson, M., 2004. “Enzymatic characterization of the streptococcal endopeptidase, IdeS, reveals that it is a cysteine protease with strict specificity for IgG cleavage due to exosite binding.” *Biochemistry* 43, 15540–9. doi:10.1021/bi048284d
 244. Vojcic, L., Pitzler, C., Wirtz, G., Jakob, F., Ronny Martinez, Maurer, K.-H., Schwaneberg, U., 2015. “Advances in protease engineering for laundry detergents.” *N. Biotechnol.* doi:10.1016/j.nbt.2014.12.010
 245. Voorhout, W.F., Leunissen-Bijvelt, J.J., Leunissen, J.L., Verkleij, A.J., 1986. “Steric hindrance in immunolabelling.” *J. Microsc.* 141, 303–10.
 246. Wang, D.-S., Iwata, N., Hama, E., Saido, T.C., Dickson, D.W., 2003. “Oxidized neprilysin in aging and Alzheimer’s disease brains.” *Biochem. Biophys. Res. Commun.* 310, 236–41.
 247. Wang, H., Yuan, D., Xu, R., Chi, C.W., 2011. “Purification, cDNA cloning, and recombinant expression of chymotrypsin C from porcine pancreas.” *Acta Biochim. Biophys. Sin. (Shanghai)*. 43, 568–575. doi:10.1093/abbs/gmr047
 248. Wargacki, A.J., Leonard, E., Win, M.N., Regitsky, D.D., Santos, C.N.S., Kim, P.B., Cooper, S.R., Raisner, R.M., Herman, A., Sivitz, A.B., Lakshmanaswamy, A., Kashiya, Y., Baker, D., Yoshikuni, Y., 2012. “An engineered microbial platform for direct biofuel production from brown macroalgae.” *Science* 335, 308–13. doi:10.1126/science.1214547

249. Waugh, S.M., Harris, J.L., Fletterick, R., Craik, C.S., 2000. "The structure of the pro-apoptotic protease granzyme B reveals the molecular determinants of its specificity." *Nat. Struct. Biol.* 7, 762–5. doi:10.1038/78992
250. Wayman, C.P., Baxter, D., Turner, L., Van Der Graaf, P.H., Naylor, A.M., 2010. "UK-414,495, a selective inhibitor of neutral endopeptidase, potentiates pelvic nerve-stimulated increases in female genital blood flow in the anaesthetized rabbit." *Br. J. Pharmacol.* 160, 51–9. doi:10.1111/j.1476-5381.2010.00691.x
251. Webster, C.I., Burrell, M., Olsson, L.-L., Fowler, S.B., Digby, S., Sandercock, A., Snijder, A., Tebbe, J., Haupts, U., Grudzinska, J., Jermutus, L., Andersson, C., 2014. "Engineering neprilysin activity and specificity to create a novel therapeutic for Alzheimer's disease." *PLoS One* 9, e104001. doi:10.1371/journal.pone.0104001
252. Weinreich, D.M., Delaney, N.F., Depristo, M.A., Hartl, D.L., 2006. "Darwinian evolution can follow only very few mutational paths to fitter proteins." *Science* 312, 111–4. doi:10.1126/science.1123539
253. Wells, T.J., Tree, J.J., Ulett, G.C., Schembri, M. a, 2007. "Autotransporter proteins: novel targets at the bacterial cell surface." *FEMS Microbiol. Lett.* 274, 163–72. doi:10.1111/j.1574-6968.2007.00833.x
254. Wensing, A.M.J., van Maarseveen, N.M., Nijhuis, M., 2010. "Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance." *Antiviral Res.* 85, 59–74. doi:10.1016/j.antiviral.2009.10.003
255. Wilhelm, M., Schlegl, J., Hahne, H., Moghaddas Gholami, A., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T.,

- Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J.-H., Bantscheff, M., Gerstmair, A., Faerber, F., Kuster, B., 2014. "Mass-spectrometry-based draft of the human proteome." *Nature* 509, 582–7. doi:10.1038/nature13319
256. Wilson, C., Mace, J.E., Agard, D.A., 1991. "Computational method for the design of enzymes with altered substrate specificity." *J. Mol. Biol.* 220, 495–506.
257. Wouters, M.A., Liu, K., Riek, P., Husain, A., 2003. "A Despecialization Step Underlying Evolution of a Family of Serine Proteases." *Mol. Cell* 12, 343–354. doi:10.1016/S1097-2765(03)00308-3
258. Wu, L., Claas, A.M., Sarkar, A., Lauffenburger, D.A., Han, J., 2015. "High-throughput protease activity cytometry reveals dose-dependent heterogeneity in PMA-mediated ADAM17 activation." *Integr. Biol. (Camb)*. 7, 513–24. doi:10.1039/c5ib00019j
259. Wu, P., Nicholls, S.B., Hardy, J.A., 2013. "A tunable, modular approach to fluorescent protease-activated reporters." *Biophys. J.* 104, 1605–14. doi:10.1016/j.bpj.2013.01.058
260. Yi, L., Gebhard, M.C., Li, Q., Taft, J.M., Georgiou, G., Iverson, B.L., 2013. "Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries." *Proc. Natl. Acad. Sci. U. S. A.* 110, 7229–34. doi:10.1073/pnas.1215994110
261. Yoo, T.H., Pogson, M., Iverson, B.L., Georgiou, G., 2012. "Directed evolution of highly selective proteases by using a novel FACS-based screen that capitalizes on

the p53 regulator MDM2.” *Chembiochem* 13, 649–53.

doi:10.1002/cbic.201100718

262. Zeng, L., Golding, I., 2011. “Following cell-fate in E. coli after infection by phage lambda.” *J. Vis. Exp.* e3363. doi:10.3791/3363
263. Zhao, Y., Jensen, O.N., 2009. “Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques.” *Proteomics* 9, 4632–41. doi:10.1002/pmic.200900398
264. Zhou, H., Mohamedali, K.A., Gonzalez-Angulo, A.M., Cao, Y., Migliorini, M., Cheung, L.H., LoBello, J., Lei, X., Qi, Y., Hittelman, W.N., Winkles, J.A., Tran, N.L., Rosenblum, M.G., 2014. “Development of human serine protease-based therapeutics targeting Fn14 and identification of Fn14 as a new target overexpressed in TNBC.” *Mol. Cancer Ther.* 13, 2688–705. doi:10.1158/1535-7163.MCT-14-0346

