

UNIVERSITY OF CALIFORNIA
Santa Barbara

Methods for redesigning the specificity of secreted proteases

A dissertation submitted in partial satisfaction of the
requirements for the degree

Doctor of Philosophy

in

Chemical Engineering

by

Jennifer Lauren Guerrero

Committee in charge:

Professor Patrick S. Daugherty, Co-Chair

Professor Michelle A. O'Malley, Co-Chair

Professor M. Scott Shell

Professor Stuart C. Feinstein

June 2016

The dissertation of Jennifer Lauren Guerrero is approved.

Stuart C. Feinstein

M. Scott Shell

Michelle A. O'Malley, Co-Chair

Patrick S. Daugherty, Co-Chair

May 2016

Methods for redesigning the specificity of secreted proteases

Copyright © 2016

by

Jennifer Lauren Guerrero

ACKNOWLEDGEMENTS

“Saber es Poder” -Abuelito

I am incredibly grateful for the past six years of graduate school and the amazing support of family and friends which propelled me forward during my PhD. The experiences and discussions shared with colleagues and friends in Santa Barbara have not only fueled my growth as a scientist but also allowed me to become healthier, happier, and more present.

First, I would like to thank my mentors at UCLA, Professor Weiss and Gloria, who first discussed with me what graduate school was and encouraged me to pursue a PhD. Thank you Professor Weiss for giving me my first research experience in your lab and Gloria for being an exceptional research mentor and female role model in science. Professor Tang, Professor Da Silva, Professor Wang, and Professor Sherlock, thank you for welcoming me into your labs and helping me develop my research skills through your advice and mentorship. To the members of the SACNAS organization, thank you for supporting me throughout my undergraduate and graduate schooling by providing an outlet to connect with researchers from underrepresented backgrounds and discuss diversity related issues.

To my advisors, Patrick Daugherty and Michelle O’Malley, thank you for your support and guidance throughout my doctoral studies. There were many times I doubted that my thesis project would ever turn a corner, but you were both incredibly optimistic and your positivity pushed me to continue to troubleshoot any challenges I encountered. Thank you for giving me the freedom to explore different courses of action in tackling our research objectives. Your trust in me was invaluable for my scientific growth to become more creative, independent, and willing to take risks. To my committee members, Scott Shell

and Stu Feinstein, thank you for your insightful feedback and interesting scientific discussions. Thank you Stu for welcoming me to carry out experiments in your lab and thank you to your student Sarah Benbow for teaching me how to culture mammalian cells.

To the members of the Daugherty and O'Malley labs, thank you for your feedback throughout the years and the wonderful memories we shared both inside and outside the lab. To Serra Elliott, Tyler Shropshire and Jen Getz, thank you for mentoring me when I first entered the lab and continuing to be great friends I can go to for career advice. To Silvia Lanati, Amol Shivange, Jack Reifert, Tobias Schoep, and Kelly Ibsen thank you for our entertaining discussions about family and science. To Kat Camacho and Kevin Solomon, thank you for our wonderful lunch breaks outside and dancing our worries away at Zumba class. To Amy Andriano, Luke Andriano, Nikki Schonenbach, and Neil Eschmann, thank you for our hiking and camping trips together that were inspiring and rejuvenating.

To my family, this thesis would not have been possible without your love and support. Mom and Dad, thinking about the challenges you overcame to pursue a college education has always kept things in perspective for me and I am so thankful for the opportunities you were able to provide me and Chris. Mom, your hard work and dedication to teaching has always inspired me to mentor others. Dad, your appreciation and pure joy for living reminds me that life is what you make of it. Thank you Mom, Dad, Chris, and my future in-laws, Kathy and Mark, for your calls, visits to Santa Barbara, and unwavering support.

To my fiancé and best friend Dan, I am so thankful that we met on the first day of graduate school and for all the memories we have shared since that day. Together we made it through our PhDs and had a lot of fun along this journey! Thank you for always

believing in me and introducing me to so many wonderful activities, such as camping and scuba diving, which kept me motivated and happy during graduate school. I am looking forward to many amazing experiences as husband and wife and for all life has in store as we move to Ventura to start our new careers!

Finally, this thesis is dedicated to my *abuelito*, *abuelita*, nana, and grandma from El Paso who have always inspired me to be brave and take risks.

VITA OF JENNIFER L. GUERRERO

May 2016

EDUCATION

- 2010-2016 Doctor of Philosophy in Chemical Engineering
University of California, Santa Barbara
- 2005-2010 Bachelor of Science in Chemical and Biomolecular Engineering
University of California, Los Angeles
Graduated Cum Laude with Distinction

PROFESSIONAL EXPERIENCE

- 2010-2016 Graduate Research Fellow and Teaching Assistant
Department of Chemical Engineering, UC Santa Barbara
Academic Advisors: Patrick Daugherty and Michelle O'Malley
- 2007-2010 Undergraduate Research Assistant
Department of Chemical Engineering, UC Los Angeles
Academic Advisor: Yi Tang
- Summer 2009 Amgen Scholars Program Fellow
Department of Genetics, Stanford University
Academic Advisor: Gavin Sherlock
- Summer 2008 UC LEADS Program Fellow
Department of Chemical Engineering, UC Irvine
Academic Advisors: Nancy Da Silva and Szu-Wen Wang

PUBLICATIONS AND PATENTS

J. L. Guerrero, P. S. Daugherty, M. A. O'Malley, Emerging technologies for protease engineering: new tools to clear out disease. *In Review*.

J. L. Guerrero, M. A. O'Malley, P. S. Daugherty, Intracellular FRET-based screen for redesigning the specificity of secreted proteases, *ACS Chem. Biol.* 2016, 11 (4), 961-970.

J. L. Guerrero, M. A. O'Malley, P. S. Daugherty. "Platform for the Evolution of Secreted Proteases" U.S. Provisional Patent Application. Application No: 62/208,572

X. Xie, I. Pashkov, X. Gao, **J. L. Guerrero**, T. O. Yeates, Y. Tang, Rational improvement of simvastatin synthase solubility in *Escherichia coli* leads to higher whole-cell biocatalytic activity, *Biotechnol. Bioeng.* 2009, 102 (1), 20-28.

PRESENTATIONS

Oct 2015	International Proteolysis Society General Meeting; Penang, Malaysia (<i>talk</i>)
Jan 2015	International Conference on Biomolecular Engineering; Austin, TX (<i>poster</i>)
Nov 2014	AIChE Annual Meeting; Atlanta, GA (<i>talk</i>)
Oct 2014	SACNAS National Conference; Los Angeles, CA (<i>talk</i>)
Oct 2014	Clorox-Amgen Graduate Student Symposium; UC Santa Barbara (<i>talk</i>)
Nov 2013	AIChE Annual Meeting; San Francisco, CA (<i>talk</i>)
Oct 2013	Amgen-Clorox Graduate Student Symposium; UC Santa Barbara (<i>poster</i>)
Oct 2012	SACNAS National Conference; Seattle, WA (<i>poster</i>)
Oct 2012	Clorox-Amgen Graduate Student Symposium; UC Santa Barbara (<i>poster</i>)

AWARDS AND HONORS

2015	International Proteolysis Society Travel Award for IPS Meeting in Malaysia
2015	UCSB Doctoral Student Travel Grant for IPS Meeting in Malaysia
2015	SACNAS Travel Scholarship for 2015 Conference in Washington D.C.
2015	Ford Foundation Dissertation Fellowship Honorable Mention
2015	1 st Place Poster Award at 5 th ICBE Meeting
2014	Outstanding Graduate Oral Presentation at SACNAS National Conference
2013	Best Poster Award at Amgen-Clorox Graduate Student Symposium
2013	UCSB Chemical Engineering Distinguished Service Award
2012	Best Poster Award at Clorox-Amgen Graduate Student Symposium
2010-2013	National Science Foundation Graduate Research Fellowship Program
2010	Heslin Fellowship awarded by UCSB Department of Chemical Engineering

SERVICE AND TEACHING EXPERIENCE

Apr 2016	Guest Lecturer for ChE 171: Introduction to Biochemical Engineering
Feb 2016	Panelist for UCSB Chemical Engineering graduate recruitment
Oct 2015	Guest Lecturer for ChE 170: Molecular and Cell Biology for Engineers
Sept 2015	Panelist for UCSB orientation for new graduate students
Mar 2015	Alumni Panelist and Graduate Mentor for UC LEADS Symposium - Mentored UCSB STEM undergraduate students and judged posters
Oct 2014	Guest Lecturer for ChE 170: Molecular and Cell Biology for Engineers
Aug 2014	Graduate Mentor for the Summer Institute in Mathematics and Science - Mentored four UCSB freshmen on a two-week biotechnology project
July 2014	Graduate Mentor for the Jack Kent Cooke Bridges Program - Mentored community college students on a one-week research project
May 2014	Graduate Representative for UCSB Hispanic Serving Institution designation
Apr 2013	Guest Lecturer for ChE 171: Introduction to Biochemical Engineering
Feb 2013	Co-organizer for UCSB Chemical Engineering graduate recruitment
Oct 2012	Recruiter for UCSB graduate programs at the SACNAS Conference
2010-2015	Graduate research mentor for 3 undergraduates and 2 graduate students
2010-2014	Volunteer at UCSB's Annual Engineering Preview Day - Spoke with engineering undergraduates about graduate school at UCSB

ABSTRACT

Methods for redesigning the specificity of secreted proteases

by

Jennifer Lauren Guerrero

Proteases regulate many biological processes through their ability to activate or inactivate their target substrates and therefore present unique opportunities for therapeutic application. Because proteases catalytically turnover proteins, they could potentially be used at lower doses in therapy, reducing the cost of treatment. However, many proteases are capable of cleaving multiple physiological substrates. Therefore their activity, expression, and localization are tightly controlled to prevent unwanted proteolysis that could lead to side effects. Currently approved protease therapeutics rely on naturally evolved specificities and are often used for protease replacement therapy in genetically deficient patients. The clinical use of proteases in replacement therapy has been successful due to the narrow substrate specificity of these enzymes, which limits their toxicity. However, the application of proteases in therapy could be extended beyond their native biological functions. The emergence of methods for engineering proteases with new activities and narrow specificities toward substrates relevant in disease could greatly expand their therapeutic potential.

Here we have developed a novel intracellular screen in yeast for redesigning the specificity of human secreted proteases which we have termed protease evolution via cleavage of an intracellular substrate (PrECISE). Using PrECISE, a protease library and a

target substrate flanked by fluorescent proteins CyPet and YPet, capable of Förster resonance energy transfer (FRET), are co-expressed in the endoplasmic reticulum (ER) of yeast. The ER provides an oxidizing environment for the formation of disulfide bonds common to human secreted proteases and the co-localization of the protease and substrate promotes cleavage. Protease activity on the selection substrate leads to loss of intracellular FRET and increases cyan fluorescence enabling screening of large protease libraries using fluorescence activated cell sorting (FACS). As a model system, we screened randomly mutated and rationally designed libraries of the secreted protease human kallikrein 7 (hK7) using PrECISE to isolate variants with improved selectivity toward the hydrophobic core of the amyloid beta peptide (A β 8: KLVF↓F↓AED).

Sequential rounds of low error rate random mutagenesis were found to be most effective in altering protease selectivity by incrementally introducing and screening for beneficial substitutions. Findings from our work emphasize the importance of screening large libraries during protease evolution since multiple substitutions were required to alter hK7 selectivity for A β . The substitutions found to improve hK7 selectivity would be impossible to predict since the majority were located far from the hK7 active site. Interestingly, improvements in selectivity were accompanied by a reduction in toxicity of the protease variant toward mammalian cells and improved resistance to wild-type inhibitors. Analysis of the crystal structures of improved variants provided insights to the potential mechanisms that affected hK7 activity and selectivity. The PrECISE method and techniques developed here can be broadly applied to evolve human proteases for specific degradation of toxic proteins involved in disease to enable their greater use in therapy.

TABLE OF CONTENTS

1. Introduction.....	1
1.1. Motivation.....	1
1.2. Organization of the dissertation.....	2
1.3. Proteases and their role in therapy.....	3
1.3.1. Protease substrate recognition.....	4
1.3.2. Regulation of protease activity.....	8
1.3.3. Proteases as therapeutics.....	10
1.4. Methods for engineering protease specificity.....	12
1.4.1. Structure-guided studies.....	13
1.4.2. Directed evolution.....	15
1.5. Alzheimer’s disease and A β proteolysis.....	22
1.5.1. Amyloid hypothesis.....	22
1.5.2. A β degrading proteases.....	24
1.6. Human tissue kallikreins.....	26
1.6.1. Human kallikrein 7 (hK7).....	26
1.7. Techniques.....	29
1.7.1. Fluorescent protein FRET.....	29
1.7.2. Fluorescence activated cell sorting (FACS).....	32
1.7.3. Strategies for library diversification.....	34
2. Development of a cell-based screen for protease engineering.....	38
2.1. Introduction.....	39
2.2. Results.....	43
2.2.1. Bacteria as an expression host for secreted proteases.....	43
2.2.2. Detection of hK7 activity in yeast using the GAL1 promoter.....	47
2.2.3. Development of the PrECISE cell-based screen for protease engineering.....	52
2.3. Discussion.....	58
2.4. Materials and methods.....	61
2.4.1. hK7 bacterial vector construction and expression.....	61
2.4.2. hK7 yeast vector construction and expression.....	62
2.4.3. Construction of FRET reporter yeast strains.....	64
2.4.4. Western blots for hK7 expression.....	66
2.4.5. Detecting hK7 activity in vivo.....	67
3. Design of a human protease targeting amyloid beta peptides.....	68
3.1. Introduction.....	69
3.2. Results.....	70
3.2.1. Screening randomly mutated hK7 libraries for A β 8 specificity.....	70

3.2.2. Characterization of variant hK7-2.7 selectivity switch	78
3.3. Discussion	87
3.4. Materials and methods	90
3.4.1. hK7 random mutagenesis library construction.....	90
3.4.2. Library screening and FACS analysis	91
3.4.3. Construction of non-target FRET reporter substrates	92
3.4.4. Characterizing activity and specificity of hK7 variants	93
3.4.5. PC12 toxicity assay	96
4. Determinants of S1 pocket specificity within hK7	98
4.1. Introduction	99
4.2. Results	100
4.2.1. Improving the dynamic range of the PrECISE method.....	100
4.2.2. Screening a rationally designed S1 pocket library of hK7	102
4.3. Discussion	112
4.4. Materials and methods	115
4.4.1. Addition of A206K substitutions to the FRET reporter substrates	115
4.4.2. Construction of an S1 pocket library of hK7	116
4.4.3. Library screening and FACS analysis	117
4.4.4. Characterization of variant activity in cell lysate	118
5. Incorporation of a non-target substrate within the cell-based screen	119
5.1. Introduction	120
5.2. Results	122
5.2.1. Introducing a counterselection substrate within a YPet exposed loop.....	122
5.2.2. Protease profiling using short-lived fluorescent substrates.....	124
5.3. Discussion	128
5.4. Materials and methods	130
5.4.1. Construction, expression and purification of YPet substrate insertions.....	130
5.4.2. Addition of a degradation tag to yeast expressed YPet.....	131
6. Conclusions.....	133
6.1. Perspectives	133
6.1.1. Development of a cell-based screen for evolving human proteases	133
6.1.2. Application of PrECISE to randomly and rationally mutated hK7 libraries... ..	135
6.1.3. Incorporation of a non-target substrate within the PrECISE screen	137
6.2. Future directions.....	138
6.3. Overall conclusions	140
7. Appendix.....	142
7.1. Appendix A: Expression of hK7 from the CUP1 yeast promoter.....	142

7.1.1. Introduction	142
7.1.2. Results and Discussion	142
7.1.3. Methods	146
8. References	148

LIST OF FIGURES

Figure 1.1. Protease specificity nomenclature.	5
Figure 1.2. Schematic overview of protease engineering.	13
Figure 1.3. Schematic for protease engineering using bacterial display.	18
Figure 1.4. Amyloid metabolism within the human brain.	23
Figure 1.5. Docking model of human kallikrein 7 with A β 8.	28
Figure 1.6. FRET probes link proteolytic activity with fluorescence.	31
Figure 1.7. Cell based libraries screened using fluorescence activated cell sorting.	33
Figure 2.1. hK7 ribbon model with disulfide connectivities.	40
Figure 2.2. Lowering the IPTG inducer concentration increases the ratio of processed to unprocessed hK7.	45
Figure 2.3. hK7 expressed in bacteria is insoluble.	46
Figure 2.4. hK7 is present in the soluble cell lysate after expression.	49
Figure 2.5. hK7 expressed in yeast is correctly folded and active.	50
Figure 2.6. hK7 expression is toxic to yeast.	51
Figure 2.7. Protease evolution via cleavage of an intracellular substrate (PrECISE).	53
Figure 2.8. Detection of hK7 activity <i>in vivo</i> using PrECISE methodology.	55
Figure 2.9. PrECISE methodology allows effective discrimination between different hK7 substrates <i>in vivo</i>	56
Figure 2.10. Workflow for enrichment of active hK7 from a background of cells expressing inactive variants to determine optimal co-expression conditions for library screening.	58
Figure 2.11. Plasmid map of pYC2/NT for hK7 expression and secretion in yeast.	63
Figure 2.12. Plasmid map of pYES3/CT for FRET reporter expression in yeast.	65
Figure 3.1. Sorting a first generation hK7 random mutagenesis library for A β 8 cleavage. ...	71
Figure 3.2. Intracellular cyan fluorescence correlates linearly to A β 8 activity in cell lysates.	73
Figure 3.3. First generation hK7 variants display improved intracellular activity for A β 8.	74
Figure 3.4. hK7 variants with improved <i>in vitro</i> activity show improved expression levels.	75
Figure 3.5. Sorting a second generation hK7 random mutagenesis library for A β 8 cleavage.	76
Figure 3.6. Second generation hK7 variant displays improved A β 8 selectivity.	77
Figure 3.7. Analysis of wt-hK7 and variant hK7-2.7 activity on a panel of substrates.	80

Figure 3.8. hK7-2.7 exhibits enhanced selectivity for A β 8 due to decreased activity on tyrosine.	81
Figure 3.9. hK7-2.7 displays reduced toxicity to PC12 neuronal-like cells.....	82
Figure 3.10. Variant hK7-2.7 confers improved yeast growth during expression.	82
Figure 3.11. hK7-2.7 exhibits reduced sensitivity to wild-type inhibitors.....	84
Figure 3.12. Improved A β 8 selectivity is mediated by a S1 pocket substitution that excludes tyrosine (Y).	86
Figure 3.13. Substitution G233V is partially responsible for A β 8 selectivity improvement in hK7-2.7.....	87
Figure 3.14. Western blot analysis of hK7 concentration before and after purification.	95
Figure 4.1. A206K substitutions to CyPet and YPet decrease FRET efficiency.	101
Figure 4.2. PRVMYYT-A206K FRET reporter displays improved dynamic range.	102
Figure 4.3. hK7 prefers cleavage after tyrosine (Y) at the P1 position.....	103
Figure 4.4. Structural analysis of the S1 subsite of hK7.	104
Figure 4.5. Analysis of hK7 activity <i>in vivo</i> against the PRVMFFT substrate.....	105
Figure 4.6. Sorting an hK7 S1 pocket library for cleavage of PRVMFFT.	106
Figure 4.7. Amino acids C and N are prevalent at position 199 after post sort 5.	107
Figure 4.8. Position 200 tolerates small amino acid substitutions.	108
Figure 4.9. The wild-type residue glycine is strongly preferred at position 222.	108
Figure 4.10. Position 227 displays preference for the wild-type glycine.....	109
Figure 4.11. Multiple residues are tolerated at position 233.....	109
Figure 4.12. S1 pocket variants display reduced activity on PRVMFFT.....	112
Figure 5.1. Substrates inserted into YPet are not susceptible to cleavage <i>in vitro</i> by hK7...	123
Figure 5.2. Cell-based screen for protease engineering using degradation of fluorescently tagged substrates.	125
Figure 5.3. Addition of DegK to the C-terminus of YPet leads to reduced yellow fluorescence after expression.	126
Figure 5.4. Substrate insertion between YPet and DegK does not affect degradation.....	127
Figure 5.5. Active hK7 co-expression does not rescue yellow fluorescence of YPet substrates fused to DegK.....	128
Figure 7.1. hK7 expression from the CUP1 promoter is measurable in yeast lysate.....	143

Figure 7.2. hK7 expression *in vivo* is not detectable with the CUP1 promoter. 144

Figure 7.3. Comparison of hK7 expression levels under the CUP1 and GAL1 promoters. . 146

LIST OF TABLES

Table 3.1. Substitutions identified for first generation hK7 variants.	72
Table 3.2. Substitutions identified for second generation hK7 variants.	78
Table 4.1. Substitutions identified for hK7 variants with highest PRVMFFT activity.	111
Table 4.2. Primers used in SOEing PCR for hK7 S1 pocket library construction.	117

1. Introduction

1.1. Motivation

Proteases represent a class of enzymatic proteins that regulate many biological processes through their ability to irreversibly activate or inactivate their target substrates through peptide bond hydrolysis. Since most proteases have on the order of 10 to 100 physiological substrates, they often participate in multiple and sometimes opposing pathways in the body. For example, thrombin dually functions as a procoagulant by cleaving fibrinogen to fibrin to form blood clots and also elicits anticoagulant effects by activating protein C¹. A generally applicable, high-throughput strategy to engineer proteases to cleave a target substrate with high specificity and high catalytic efficiency, while minimizing off-target activity, would greatly expand the use of proteases for analytical, biotechnological, and therapeutic applications.

Deficiencies or abnormalities in protease activity are common to many diseases, therefore proteases present attractive targets for use as therapeutics. Due to their catalytic properties, proteases provide a significant advantage over antibodies, since they need not be present at stoichiometric ratios to a given target, which could reduce the dosage and cost of treatment. Currently approved protease therapeutics rely on naturally evolved narrow specificities, which limits their toxicity. Anticoagulants such as t-PA and u-PA break down clots for the treatment of heart attacks and stroke, while procoagulants, such as Factor IX and Factor VIIa, are approved to treat haemophilia^{2,3}. Currently, the use of proteases as therapeutics is limited in part by their specificity for a given target substrate, since off-target recognition would lead to unwanted side-effects⁴. And methods for engineering human secreted proteases, candidates for therapy, are lacking. Therefore a high-throughput

screen to engineer human proteases with novel and narrow specificities could greatly expand their use in therapy.

1.2. Organization of the dissertation

This dissertation describes the development of a screening system in yeast for engineering the specificity of human secreted proteases by assaying the activity of large protease variant libraries against a fluorogenic target substrate. This general cell-based assay can be applied toward multiple protease targets that are actively expressed in yeast, to increase their activity and selectivity toward an arbitrary peptide target. Additionally, through application of this screen, a deeper understanding of the molecular mechanisms by which proteases achieve a high level of substrate selectivity can be gained. As a model system for developing highly specific proteases that could be applied in therapy, we chose to engineer the protease human kallikrein 7 (hK7) to selectively cleave the hydrophobic core of the amyloid beta (A β) peptide, which aggregates in the brains of patients with Alzheimer's disease (AD) and is widely considered to be the key pathological event in AD⁵.

The dissertation is composed of six chapters where Chapter 1 describes the mechanisms by which proteases recognize their targets and past work in engineering proteases specificity. In Chapter 2, we describe the development of a novel cell-based method for engineering proteases which belong to the family of human secreted proteases that contain multiple disulfide bonds required for proper folding and activity. In Chapter 3, this cell-based assay is applied to hK7 to screen large randomly mutated libraries for selectivity toward the amyloid beta (A β) peptide. Chapter 4 describes the extension of this approach

to a site-saturation mutagenesis library of the hK7 S1 pocket to determine the amino acids in the S1 subsite that contribute to activity and selectivity. In Chapter 5, opportunities to improve the screen by the addition of a counterselection substrate are explored. Chapter 6 discusses current challenges in the field of protease engineering and potential for future advancements. Collectively, our studies demonstrate the benefit of developing high-throughput screens for protease engineering, since multiple substitutions were required to alter the selectivity of hK7 for A β . Additionally, our work uses key information obtained from analysis of the crystal structures of improved variants to postulate mechanisms responsible for substrate specificity within hK7.

1.3. Proteases and their role in therapy

Proteases constitute 2% of the human genome and are key players in diverse biological pathways including digestion⁶, inflammation⁷, coagulation⁸, and apoptosis⁹ through proteolytic processing of their substrates. Their mechanism of action allows organization of proteases into classes based on the residues responsible for catalysis. These residues and surrounding amino acids that contact the substrate during catalysis make up the active site of the protease. The most abundant human protease class are the metalloproteases followed by serine, cysteine, threonine and aspartyl family members¹⁰. Additionally, the position of the cleaved peptide bond within a substrate, either internal or terminal, is further used to classify a protease as an endopeptidase or exopeptidase, respectively.

The protease class dictates whether covalent or non-covalent hydrolysis is used to cleave the peptide bond¹¹. Serine, threonine, and cysteine proteases participate in covalent catalysis through the utilization of amino acids within their active sites which serve as

nucleophiles. The hydroxyl and thiol groups of serine, threonine, and cysteine effectively attack the substrate peptide backbone resulting in hydrolysis. In non-covalent catalysis, the nucleophile is an activated water molecule and this mechanism is used by the metalloproteases and aspartic proteases. During catalysis, histidines function as bases in covalent catalysis, whereas either aspartic acid, glutamic acid or zinc serve as acids and bases in non-covalent catalysis¹¹. The rate at which a substrate is cleaved depends largely on the interactions it forms when docked into the protease active site. These interactions form the basis for protease substrate recognition, also known as the specificity of a protease.

1.3.1. Protease substrate recognition

In order for hydrolysis of a substrate to occur, the substrate must first dock into the active site of the protease in the correct orientation that permits catalysis. Protease substrate recognition was first described by Schechter and Berger in 1967¹². Using the Schechter and Berger nomenclature, the substrate residues are labeled $H_2N-PX\dots P4-P3-P2-P1 \downarrow P1'-P2'-P3'-P4'\dots PX'-COO^-$, where P1 is the new C-terminal residue and P1' is the new N-terminal residue following catalysis (Figure 1.1). The substrate residues dock into the corresponding subsites that run along the surface of the protease, that are labeled SX-S1 and S1'-SX', respectively. Therefore, the P1 residue of the substrate interacts with the S1 subsite of the protease and so forth. Crystal structure analysis of a substrate docked within a protease can elucidate which amino acids comprise the various subsites. However, the importance of specific residues within a given subsite is often difficult to predict from structural information alone.

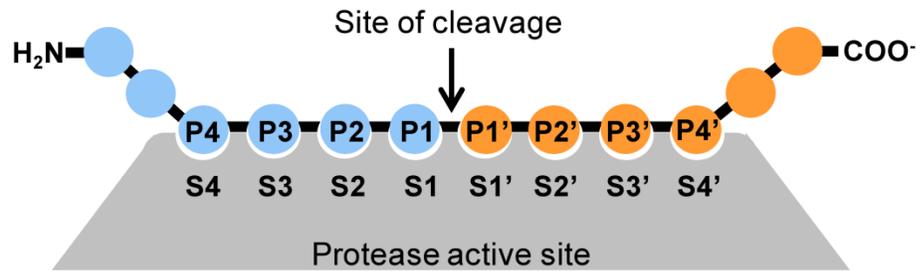


Figure 1.1. Protease specificity nomenclature.

The protease active site (grey) is composed of subsites (SX...S1-S1'...SX') that each interact with the corresponding residue in the substrate. Proteolysis occurs between substrate residues P1 and P1'.

A protease's increased activity and preference for certain residues at a subsite is known as the specificity of a protease. Some subsites within a protease may be restricted to one or two amino acids and are therefore significant in defining the protease's specificity. While other subsites display no discrimination between different amino acids and contribute less to the overall binding energy of the protease-substrate interaction. Because cleavage occurs after the P1 residue, the S1 subsite is considered the primary binding site and critical determinant of protease specificity since it correctly orients the substrate for catalysis. Within the serine protease class, trypsin-like proteases favor cleavage after basic residues arginine and lysine. The preference for arginine and lysine has been attributed in part to the aspartic acid (Asp189) at the bottom of the S1 binding pocket in trypsin¹³.

The subsites which affect protease specificity can be highly variable and can both positively and negatively affect binding of the P1 residue into the S1 pocket. The favorable or deleterious interactions that occur between substrate residues at distinct subsites is known as subsite cooperativity. Studies on the protease subtilisin discovered that favorable amino acids at the P1 or, more significantly, the P4 position in the substrate could

overcome the adverse effects of less favorable residues at other positions¹⁴. Many proteases of viral origin, such as the tobacco etch virus protease (TEVp), have evolved extended subsite recognition to facilitate efficient viral protein processing and assembly within host cells¹⁵. TEVp displays very narrow selectivity for the sequence ENLYFQ↓S/G within the TEV polyprotein¹⁶. Additionally, mammalian proteases involved in cellular pathways have evolved to recognize a limited set of peptide sequences. For example, proteases responsible for the formation and breakdown of blood clots have evolved exquisite substrate specificity¹⁷. While proteases involved in processes such as digestion and in the recycling of proteins display relaxed specificity.

To determine a protease's substrate specificity, typically kinetic assays are run with synthetic substrates that lead to a fluorescent output upon cleavage. Measuring the change in fluorescence over time allows calculation of the specificity constant or k_{cat}/K_M . To increase the throughput of mapping protease specificity, combinatorial fluorogenic peptide libraries can be assayed in microtiter plates with the protease of interest. However, manually synthesized libraries are limited in their size, generally containing 10^5 to 10^6 members, and can be expensive to make¹⁸. Substrate phage display and bacterial peptide display profile protease specificity on much larger libraries containing 10^7 to 10^8 members and are cost effective, since phage and bacteria produce the peptides. In these methods, phage or bacteria encoding protease-sensitive peptides are collected and the optimal cleavage motif is determined by DNA sequencing¹⁹⁻²¹. The preferred substrate motif can then be searched against the human proteome to determine candidate proteins that may serve as physiological substrates. One drawback to phage and bacterial display of substrates is that the P1-P1' cleavage site cannot be directly determined. Further studies

using mass spectrometry to analyze cleaved fragments of synthetic substrates are needed to determine the scissile bond. Once the preferred substrate residues are identified, analysis of the crystal structure can help explain what mechanisms, such as hydrogen bonding, electrostatics, steric hindrance, and hydrophobicity, govern the preference of an amino acid at a subsite.

Protease specificity characterization has mainly focused on residues in the substrate that preferentially dock into the active site. However, proteases can additionally use surfaces distant from the active site called exosites to influence substrate selectivity and increase or decrease the rate of catalysis. The influence of exosites is evident in studies describing the enhanced cleavage of large physiological substrates compared to short peptide sequences that bind solely to the active site^{22,23}. The most widely recognized example of exosite utilization is in the serine protease thrombin. Thrombin has a strong preference for arginine at P1, which is found in its three primary physiological targets, fibrinogen, activated protein C, and protease-activated receptor 1 (PAR1)²⁴. In addition to its active site, the presence of two positively charged surface patches termed exosite I and exosite II on thrombin interact with extended substrate sequences and kinetically control docking of the substrates²³. An extended exosite binding motif in PAR1 (LRNPNDKYEPFWEDEEK) aids in forming the initial protease substrate complex and increases thrombin catalysis by 325-fold over a substrate containing only the residues which interact with the active site alone²⁵.

1.3.2. Regulation of protease activity

Most proteases are predicted to cut ~10-100 physiological substrates through their involvement in various pathways. Therefore protease activity must be tightly regulated to promote activity when needed and also prevent improper processing of substrates. Misregulated proteolysis is lethal and has been linked to a diverse set of pathologies such as cardiovascular²⁶ and inflammatory disease²⁷, cancer^{28,29}, and neurodegenerative disorders³⁰. Therefore, protease activity is tightly controlled in the body through a protease's substrate specificity, as well as through temporal and spatial regulation of protease activity. To minimize unregulated proteolysis, nearly all proteases are produced as zymogens which have undetectable activity until activated through proteolytic processing. Proteases are further regulated through the action of inhibitors, which bind the active enzyme effectively reducing its concentration³¹.

To reduce off-target proteolysis, proteases are localized to specific tissues and compartments of the cell through protein-protein interactions. Proteases may also be secreted and circulate in the blood stream or lymphatic system. Additionally, nearly all proteases are inactive after translation due to the presence of a short N-terminal propeptide that blocks or alters activity of the active site³². These inactive versions termed zymogens become activated after proteolytic processing of the propeptide, which often leads to a conformational change and correct rearrangement of the active site for catalysis. For example, cleavage of the propeptide in many serine proteases produces an N-terminal isoleucine, which inserts itself into the protease and forms an internal salt bridge with Asp194³³. Formation of the internal salt bridge is critical for stabilizing the active site and producing a functional S1 pocket¹³. Interestingly, synthetic peptides which mimic the

activated N-terminus of the serine proteases trypsin and hepatocyte growth factor have been shown to bypass proteolytic processing and reversibly stimulate the functional forms of the proteases^{34,35}. Protease zymogens which participate in catalytic cascades, such as apoptosis, are activated in a hierarchical fashion with the first proteases in the pathway activating proteases below them³⁶. Often the first proteases in the pathway have been shown to self-activate, cleaving and releasing their own propeptides, initiating the catalytic cascade³⁷.

Protease activity is further regulated by the action of inhibitors which bind to the active site or allosteric sites preventing substrate docking and reducing the concentration of active protease³¹. Inhibitors are generally non-specific and can effectively bind to multiple proteases with similar folds³⁸. The mechanism of inhibition may be reversible or irreversible and the strength of inhibition is characterized by the binding constant K_i and association rate constant k_{ass} to describe the binding speed. The largest class of inhibitors are the serine protease inhibitors or serpins, which employ an irreversible mechanism of inhibition by trapping the active site of the protease in a covalent complex with the reactive site of the inhibitor³⁹. In humans, serpins are localized to the bloodstream and regulate the activity of serine proteases responsible for inflammation and blood clotting/fibrinolysis⁴⁰. Because protease activity is necessary for viral replication in diseases such as acquired immunodeficiency syndrome (AIDS), synthetic inhibitors that bind to the active site of the human immunodeficiency virus (HIV) protease have been developed⁴¹. These inhibitors have dramatically increased the life expectancy of those living with AIDS⁴². However in many cases, HIV protease has evolved resistance to inhibitor drugs rendering treatment ineffective⁴³.

1.3.3. Proteases as therapeutics

Because proteases have the unique ability to catalytically turnover their target substrates, they have been commercially used as reagents for research, detergent additives, and increasingly looked to as therapeutics. Currently, around 17 proteases have been clinically approved with the majority of these proteases being used for enzyme replacement therapy in genetically deficient patients. For example, the proteases Factor VIII, Factor IX, and Factor FVII are essential for the formation of blood clots and have been approved to treat patients with hemophilia A, B, and C, respectively, who lack production of these proteins⁴⁴⁻⁴⁶. In patients with cystic fibrosis, a deficiency in pancreatic proteases causes severe intestinal malabsorption. Therefore, replacement therapy with digestive proteases has been used to improve fat and nitrogen uptake⁴⁷.

Proteases used in replacement therapy generally have narrow substrate specificities which has enabled their success in the clinic while limiting their toxicity. However, the application of proteases in therapy does not need to solely arise from their biological functions. The emergence of new tools for protease engineering will allow the design of proteases with novel activities and specificities for therapeutic targets. Proteases can be engineered to neutralize toxic proteins such as the amyloid beta protein or mutant huntingtin protein, which aggregate in neurological disorders. Additionally, the proteolytic inactivation of acute phase response proteins can help treat inflammation observed in asthma and rheumatoid arthritis. Due to their catalytic properties, proteases provide a significant advantage over antibodies, since their therapeutic application does not depend on stoichiometric addition to a given target, which could reduce the dosage and cost of treatment.

The lack of substrate specificity has limited the use of proteases in therapy, since off-target activity would lead to unwanted side-effects. For example, the approved protease therapy activated protein C (APC) participates in both inflammation and blood coagulation and its anti-inflammatory properties provide cytoprotective effects against serious infection in patients with sepsis^{48,49}. However, the excessive anticoagulant properties of APC often lead to serious bleeding in patients, which has complicated its use^{50,51}. Protein engineering studies to reduce APC's inactivation of targets involved in coagulation while maintaining its anti-inflammatory properties would greatly minimize these side effects. Therefore high-throughput methods for engineering target-specific proteolytic enzymes of human origin with high activity and stability are needed.

Another hurdle in using proteases for therapy is their short half-lives in the body. Most protease display half-lives on the order of minutes to hours *in vivo* due to rapid inhibition compared to antibodies, which are active in the bloodstream for weeks. The therapeutic protease tissue plasminogen activator (tPA), approved for the breakdown of blood clots in patients who have recently suffered from heart attack or stroke, displays a biological half-life of 6 minutes due to its rapid inhibition by plasminogen activator inhibitor-1 (PAI-1)⁵². Site-directed mutagenesis of residues in tPA responsible for its fast clearance rate, such as those involved in the interaction with PAI-1, extended the half-life of tPA 3-fold, reducing the dosing frequency⁵³⁻⁵⁵. However, levels of tPA that are too high can lead to severe bleeding and haemorrhage⁵⁶. Therefore caution must be taken in extending the therapeutic window since the multiple biological roles of proteases may produce unintended side effects. General methods for improving protease specificity are thus critical for reducing off-target activity before protease half-lives can be extended.

1.4. Methods for engineering protease specificity

The ability to engineer proteases with novel substrate specificities and activities could greatly expand the use of proteases in therapy. However, the evolution of proteases poses unique complications compared to other protein classes. Because most proteases are capable of cleaving multiple substrates, expression of a protease within a cell often causes toxicity due to cleavage of essential host cell proteins^{57,58}. Variants with improved intrinsic activity or expression display even greater levels of toxicity to the cells. This creates a major hurdle since protease expression is necessary for characterizing the properties of protease variants. Additionally, proteases have the capability of cleaving and inactivating themselves, which could lead to low stability during characterization. Despite these challenges, researchers have made significant progress in engineering proteases of bacterial, viral, and human origin with improved activity, specificity, stability, and expression.

Protease engineering is a relatively new field with the majority of the high-throughput screens for protease evolution being developed in the past decade. Although these methods vary substantially, the main steps for protease engineering remain the same. These steps include selecting a parent protease, generating diversity through random or targeted mutagenesis, screening or selecting for the desired activity, and characterizing the properties of isolated variants (Figure 1.2). Often this process is repeated several times as mutations are incrementally introduced and screened for enhanced properties of the protease. In addition, choosing an optimal protease mutagenesis and screening strategy depends heavily on the engineering objective, information available and critical features of the parent protease. For example, if the protease contains post-translational modifications

that are required for protease activity, library screening would need to be accomplished in eukaryotic cells, which are capable of performing such modifications.

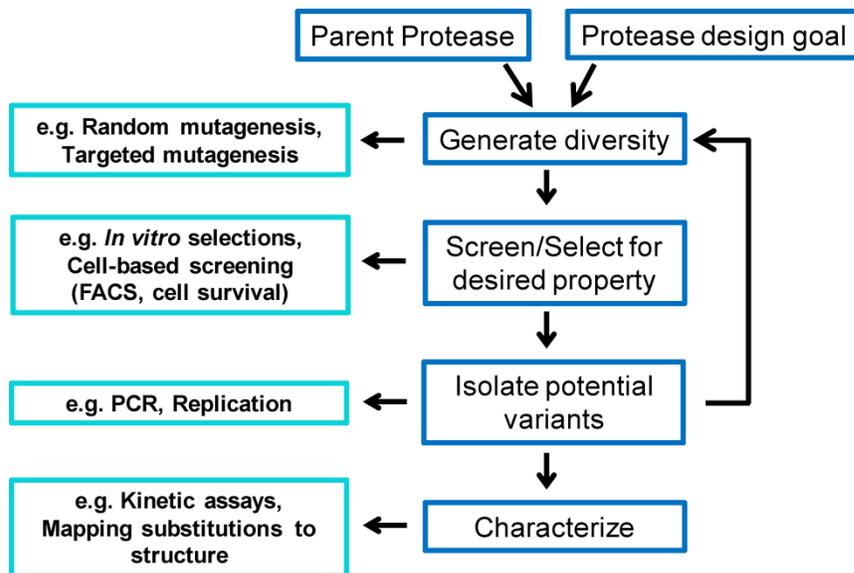


Figure 1.2. Schematic overview of protease engineering.

After selecting a protease design goal and parent protease, the protease is mutated either through targeted mutagenesis or the creation of large libraries. Variants are then screened for a desired property and positive clones are isolated for further characterization. This process can be repeated multiple times before obtaining a variant that satisfies the design criteria.

1.4.1. Structure-guided studies

Early efforts in engineering proteases have generally centered on structure-guided mutagenesis of surface-exposed amino acids lining the active site using low throughput assays. Computational analysis of substrate-protease docking models and comparing homology between closely related proteases with known structure and function can additionally aid in determining sites of interest and predicting favorable substitutions⁵⁹. Using this approach, the secondary structural elements or specific point substitutions can

be incorporated into a protease based on the structure of another protease. For example, trypsin and chymotrypsin have similar tertiary structures, but differ in their specificity. Trypsin cleaves after basic residues, arginine and lysine, whereas chymotrypsin prefers cleavage after large hydrophobic amino acids. By analyzing the structure and active site of both proteases, researchers converted trypsin to have chymotrypsin-like specificity by incorporating four amino acid substitutions and two surface loops found in chymotrypsin for the analogous sites in trypsin⁶⁰.

Studies evolving the specificity of thrombin highlight the clinical utility of redesigning protease specificity. Thrombin has opposing functions in hemostasis, since it can convert fibrinogen to fibrin monomers that form a stable blood clot and can also elicit anticoagulant effects when in complex with thrombomodulin by activating protein C¹. The wild-type protease displays >100 fold higher k_{cat}/K_M for cleavage of fibrinogen over protein C in the presence of thrombomodulin. Alanine scanning mutagenesis of 97 residues covering the solvent accessible surface of thrombin, identified the tryptophan (W) at position 215 as an important determinant of specificity⁶¹. Site-saturation mutagenesis revealed that the substitution W215E led to a 1000-fold switch in selectivity favoring cleavage of protein C over fibrinogen. The enhancement in anticoagulant properties of variant W215E could be efficacious in the breakdown of blood clots after heart attack or stroke.

Another recent study used structure-guided mutagenesis of residues within the active site of neprilysin to engineer its specificity for the amyloid beta peptide. A panel of variants was generated and screened in microtiter plates for activity toward A β ₄₀ and multiple non-target physiological substrates using fluorescence polarization to detect

activity⁶². A neprilysin variant with two substitutions (G399V/G714K) displayed 20-fold improved activity and up to 3200-fold improved selectivity on A β ₄₀ through reductions in cleavage of wild-type substrates. While structure-guided mutagenesis has shown some success in redesigning protease specificity, the ability to predict which residues are responsible for specificity based on the crystal structure is still a very challenging task. Additionally, often multiple substitutions are needed to generate large specificity switches and these substitutions may be located far from the active site⁶³. Nevertheless, site-directed mutagenesis of amino acids in and around the active site of the protease can be used to identify residues that are important for activity and specificity providing guidance toward directed evolution approaches.

1.4.2. Directed evolution

The ability to screen large protease libraries can dramatically improve the probability of isolating rare variants with novel activities and selectivities, since typically a combination of amino acid substitutions is required to alter specificity while maintaining high levels of catalytic activity⁶⁴. Current screening and selection methods are equipped to handle 10⁴ to 10¹³ variants. Protease libraries can be constructed using a variety of mutagenesis methods, such as error-prone polymerase chain reaction (epPCR), site-specific saturation mutagenesis, or genetic recombination. Recent advances in the development of high throughput assays for protease engineering have utilized bacteria and yeast cells for the redesign of bacterial and viral proteases. The advantages and drawbacks of recent screens and selection methods are discussed as well as their potential to be applied toward additional protease targets.

A cell-based assay was developed in bacteria to exploit the toxicity associated with non-specific protease activity. A protease library is expressed in bacteria and non-specific variants leading to toxicity are removed, while surviving variants are screened for cleavage of a co-expressed reporter substrate⁵⁷. Within the screen, a target peptide from the TNF α protein was inserted into a surface accessible loop of the reporter protein β -galactosidase (β -gal), such that cleavage of the substrate leads to a decrease in β -gal activity. Human immune-deficiency virus type I protease (HIV-Pr) was randomly mutated using epPCR and expressed in bacteria to remove toxic, non-specific variants. Surviving variants were then assayed for a decrease in β -gal activity. An HIV-Pr variant with two substitutions (P9S, I50L) displayed 2-fold increased selectivity for the TNF α peptide over a wild-type preferred sequence coupled with non-toxic expression in bacteria. This screen is amenable to proteases of bacterial and viral origin, which can be expressed in an active form in the cytoplasm of *E. coli*.

Another potential strategy in protease engineering is localizing a transcriptional activator at a site separate from the nucleus by fusing it to an anchoring membrane protein with a target substrate linking the two proteins. Cleavage of the target substrate causes translocation of the transcriptional activator to the nucleus where it turns on expression of a reporter protein. This screening design was incorporated into yeast through localization of the transcriptional activator LexA-b42 to the cytoplasmic face of the plasma membrane by attaching it to an integral membrane protein via a cleavable linker⁶⁵. Four residues within the S2 pocket of the hepatitis A virus 3C protease (HAV 3CP) were subjected to saturation mutagenesis. Variants were screened for activity toward a substrate containing a non-preferred glutamine (Q) at the P2 position and a resulting variant displayed a 160-fold

switch in selectivity for Q over T at P2. This method was further used to identify variants of HAV 3CP that can cleave glutamine stretches found in the mutant huntingtin protein that aggregates in Huntington's disease⁶⁶.

To increase the throughput of analyzing large libraries, a screen was developed for the display of protease variants on the surface of *E. coli* followed by fluorescence activated cell sorting (FACS). Cells displaying the protease library are incubated with substrates that upon proteolytic cleavage produce a positively charged fluorescent product that is captured on the negatively charged cell surface. The fluorescent signal measured by flow cytometry is directly proportional to the catalytic activity of the protease and variants with the desired activity can be isolated. Using this screen with a single selection substrate, the *E. coli* outer membrane protein T (OmpT) was engineered to exhibit a 60-fold increase in activity against a non-preferred substrate Arg-Val⁶⁷. However, the engineered enzyme displayed relaxed overall specificity and also cleaved peptide substrates at Ala-Arg sequences with higher efficiency than the wild-type OmpT. This increase in substrate promiscuity has been commonly observed when engineering proteases toward unnatural substrates^{68,69}.

To improve on this method, a counter-selection substrate was added to the cell-based assay and OmpT variants were screened for cleavage of the non-preferred Ala-Arg peptide bond while counter-selecting for variants that did not cleave the Arg-Arg bond preferred by wild-type OmpT (Figure 1.3)⁷⁰. This counter-selection strategy allowed isolation of an OmpT variant that cleaved the Ala-Arg bond with three million-fold selectivity over the wild-type Arg-Arg preferred cleavage. The isolated variant also had a catalytic efficiency for the Ala-Arg sequence similar to wild-type for the preferred substrate. By screening libraries with this selection-counter-selection substrate method, an OmpT variant with

activity toward Glu-Arg bonds was discovered, a specificity that has not been observed among natural proteases⁷¹. Additionally, OmpT variants were isolated that could selectively recognize sulfotyrosine and 3-nitrotyrosine, while discriminating against peptides containing unmodified tyrosine^{72,73}. Although this method has been successful for engineering OmpT to selectively recognize new substrates, it is limited by the display of protease libraries on the surface of bacteria, which may not be feasible for more complex secreted eukaryotic proteases.

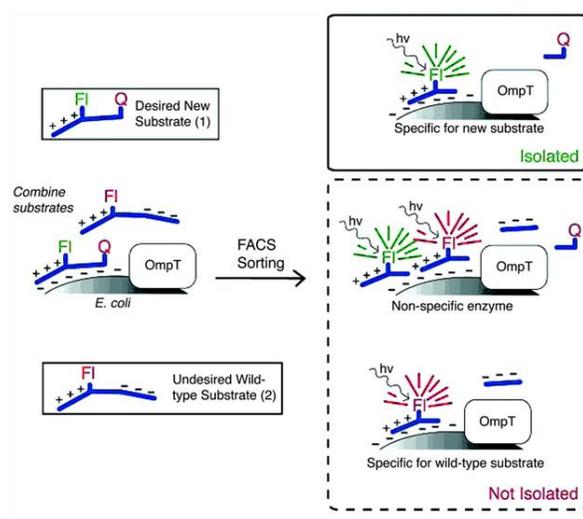


Figure 1.3. Schematic for protease engineering using bacterial display.

As a model for protease specificity engineering, OmpT libraries are displayed on the outer membrane of *E. coli* and cells are incubated with a target substrate capable of FRET and a fluorescent non-target substrate. Cleavage of the target or non-target substrates by an OmpT variant leads to capture of the corresponding fluorescent moiety on the cell surface. Cells which express variants displaying activity solely for the target substrate are isolated by FACS⁷⁰.

A potential cell-based selection for protease engineering is the coupling of proteolytic cleavage of a target substrate to antibiotic resistance permitting cell growth. This concept was utilized in a bacterial-based screen where cleavage of a positive selection substrate allows β -lactamase translocation from the cytoplasm to the periplasm where it is active and can confer antibiotic resistance to carbenicillin⁷⁴. However, proteolytic cleavage of a counter-selection substrate causes retention of β -lactamase in the cytoplasm reducing cell growth on carbenicillin. Utilizing this selection and counter-selection strategy, a focused three member saturation mutagenesis library of the S6 pocket (N171, N176, Y178) of Tobacco Etch Virus (TEV) protease was analyzed for increased selectivity for T and P at P6 over the wild-type preferred residue E. Growth selection on carbenicillin yielded the importance of a single substitution N171D which conferred nearly 12-fold greater *in vitro* selectivity for P over E at P6. While this method employs a counter-selection substrate to increase the probability of isolating selective protease variants, it is restricted to proteases which can fold in the cytoplasm of *E. coli* and are non-toxic to the host. In addition, selection based methods offer less control over the dynamic range between cells containing active or inactive protease variants compared to screens which give a fluorescence or colorimetric output.

The compartmentalization and secretion pathway of yeast were utilized in a general and robust screen for protease engineering by sequestering the protease and substrates within the endoplasmic reticulum (ER) after which cleaved substrates are displayed on the outer surface of the cell⁷⁵. Fusion of the yeast ER retention signal (FEHDEL) to the substrate or protease increases their residence time in the ER improving the opportunity for proteolysis to occur. Cleavage of the selection or counter-selection substrates can be quantified by the

addition of fluorescent tags specific for epitopes downstream of the cleavage site. Loss of fluorescence indicates cleavage of the substrate or counter-selection substrate directly upstream, and cells with desired fluorescence profiles are isolated using FACS. The FACS-based approach offers the advantages of screening protease libraries with greater than 10^7 members and the ability to fine tune the conditions for sorting to isolate rare clones with altered selectivities.

The YESS method was used to screen a four member saturation mutagenesis library of the S1 pocket (T146, D148, H167, and S170) of TEV against a library of substrates with the sequence ENLYFX↓S, where X is any amino acid. The wild-type preferred sequence ENLYFQ↓S was used as a counter-selection substrate to increase selectivity of TEV for new P1 residues. Variants characterized from this screen were subjected to random mutagenesis and additional screening to improve catalytic activity on the desired substrates. Two highly active variants isolated from screening displayed 5,000-fold and 1,100-fold switches in selectivity for E or H at P1, respectively, over the wild-type preferred residue Q. In addition, by removing the ER retention signals from the protease and substrate, the YESS system recovered a TEV variant with 4-fold improved activity on its native substrate ENLYFQS. This screen also displayed sufficient dynamic range for the screening of other viral proteases, as well as the human protease granzyme K, which contains multiple disulfide bonds.

To enable the rapid identification of substitutions that lead to inhibitor resistance in proteases, the phage assisted continuous evolution (PACE) system⁷⁶ was modified to link protease activity on a target substrate to phage survival and propagation⁷⁷. Within protease PACE, phage carry an evolving protease of interest. Protease variants with activity on the

selection substrate trigger activation of T7 RNA polymerase, which is responsible for the production of an essential protein in phage propagation. Phage that produce inactive protease variants are unable to infect *E. coli* and are diluted out of the pool. Throughout the experiment, mutagenesis of the protease is triggered by an inducible, low fidelity DNA polymerase, which increases the speed of continuous evolution with minimal researcher intervention. This system was used to evolve the hepatitis C virus (HCV) protease in the presence of two inhibitors, danoprevir and asunaprevir, for retention of wild-type activity. Growth and selection revealed that substitutions D168E and D168Y to HCV protease resulted in up to 30-fold resistance to inhibition *in vitro* by danoprevir and asunaprevir. This method requires that the protease of interest be produced in a folded and functional form within bacteria and therefore protease PACE would likely not be suitable for human proteases.

Although, the high-throughput screens developed have enabled the rapid identification of protease variants with improved activity, selectivity, and stability, they have only been applied to bacterial and viral proteases, which lack therapeutic potential. Bacterial and viral proteases are easy to express, since they lack post-translational modifications and often display narrow specificities, which limits their toxicity. However, these proteases would elicit an immune response if introduced into humans and would not be candidates for therapy. Many of the current and potential protease therapeutics are human secreted proteases, which contain multiple disulfide bonds. Therefore, this dissertation outlines the development of a screen to evolve the specificity of human secreted proteases toward target substrates which aggregate in disease, such as the amyloid beta peptide in Alzheimer's disease. By minimizing off-target proteolytic cleavage, protease variants

should display lower levels of side effects and increased efficacy for neutralization of the target.

1.5. Alzheimer's disease and A β proteolysis

1.5.1. Amyloid hypothesis

Though there is still much debate on the causes of Alzheimer's disease (AD), the amyloid hypothesis proposes that accumulation and aggregation of amyloid beta (A β) in the brain causes a series of events that leads to cognitive decline in patients with AD⁷⁸. The A β peptide is produced after sequential cleavage of the amyloid precursor protein (APP), an integral membrane protein, by β -secretase and γ -secretase (Figure 1.4)^{79,80}. The A β monomer is 40 or 42 amino acids in length, with A β 42 displaying a higher hydrophobicity, toxicity, and propensity to form fibrils⁸¹. An important feature of the A β peptide is a central hydrophobic core of amino acids (KLVFFAED) that is both necessary and sufficient for A β fibril formation^{82,83}. When A β 42 accumulates, these monomeric species begin to aggregate due to their hydrophobic nature into toxic oligomeric species and fibrils, which are the main components of amyloid plaques⁸⁴. A β plaques interfere with communication between neurons and promote neuronal death⁵.

Clearance of amyloid beta is achieved through proteolytic processing by enzymes within the brain and also removed by crossing the blood brain barrier into the peripheral vasculature^{85,86}. However, most AD patients display increased accumulation of A β in the central nervous systems due to impaired clearance mechanisms⁸⁷. Therefore, a therapeutic agent that decreases the amount of A β species could be efficacious in the treatment of AD. Clinical trials of therapeutics for AD have focused on inhibitors of β -secretase and γ -secretase to reduce production of the A β monomer as well as monoclonal antibodies

toward A β to reduce its aggregation⁸⁸⁻⁹⁰. However, β - and γ -secretase inhibitors have failed to show cognitive improvement in patients with AD and have also led to unwanted side-effects due to shutting down the other biological functions of β - and γ -secretase^{91,92}. Additionally, the clinical efficacy of therapeutic antibodies has only been observed at high dosage concentrations, which have led to side-effects such as vasogenic edema^{93,94}. Alternatively, another therapeutic strategy is the enhanced proteolytic clearance of A β , which could irreversibly inactivate the toxic peptide reducing its aggregation in AD^{85,95}.

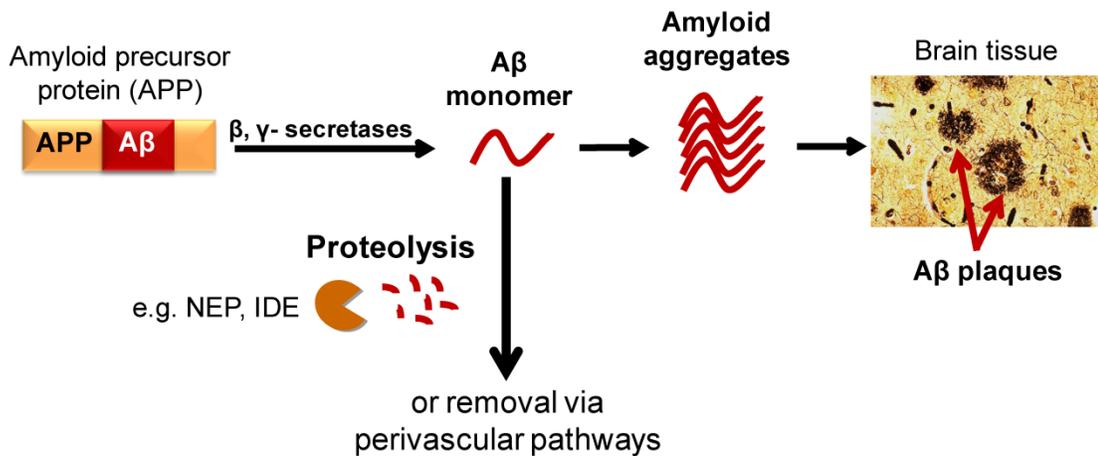


Figure 1.4. Amyloid metabolism within the human brain.

The amyloid precursor protein (APP) is sequentially cleaved by the proteases β -secretase and γ -secretase, respectively, to produce the amyloid beta monomer, which can be 40 or 42 amino acids in length. The physiological role of A β is still uncertain and the monomer is cleared from the brain via proteolytic processing or removal via perivascular pathways to prevent its accumulation. Deficiencies in the clearance pathway of A β lead to build up of monomers in the brain which aggregate due to their hydrophobicity. The presence of A β plaques in the brain has been shown to impair neuronal communication and function resulting in cognitive decline.

1.5.2. *A β* degrading proteases

Multiple proteases have been identified with *in vitro* activity toward the A β peptide and many of these proteases have been confirmed to remove A β *in vivo*. Knockout mice for the proteases neprilysin (NEP)⁹⁶, endothelin-converting enzyme (ECE)⁹⁷, and insulin degrading enzyme (IDE)⁹⁸ display an accumulation of endogenous A β confirming A β regulation by these proteases. The cellular or extracellular localization of A β -degrading proteases is strongly correlated with their ability to cleave monomeric, oligomeric, and fibrillar A β . For example, IDE is localized to the cytosol of neurons and microglia and cleaves only soluble monomeric A β that becomes internalized⁹⁹. The membrane-bound protease NEP cleaves soluble and oligomeric A β ^{100,101}, whereas secreted proteases, such as plasmin and human kallikrein 7 (hK7), have been shown to degrade A β monomers and fibrils^{102,103}. It is likely that the fibrillar forms of A β cause certain cleavage sites of the monomer to become inaccessible leading to the accumulation of A β plaques¹⁰⁴. A β -degrading proteases cleave at different sites within the monomer producing a wide array of fragments that each display differing levels of reduced aggregation. While much work still remains in elucidating the aggregation potential of A β degradation products, overall the majority of A β fragments are less likely to aggregate, less neurotoxic, and more easily cleared than full-length A β ^{105,106}.

Enhancement of A β clearance through proteolytic cleavage into non-aggregating peptides represents a distinct potential alternative to antibody-mediated clearance for the treatment of amyloid diseases including AD. Systemically administered monoclonal antibodies specific for A β have exhibited efficacy in mild AD⁹⁴. However, antibodies toward A β monomers and aggregates have resulted in side effects including brain

inflammation and vasogenic edema or amyloid-related imaging abnormalities (ARIA)^{93,107}. These safety concerns have resulted in lowering the antibody dosage during clinical trials which may reduce the desired clinical effect¹⁰⁸. Therapeutic proteases present an alternative approach and could be used at much lower doses, providing sustained clearance at lower rates. Toward this goal, catalytic antibodies with α -secretase activity toward A β have been explored¹⁰⁹. However, known catalytic antibodies suffer from low specificity and ~100-fold reduced activity toward A β compared to proteases such as neprilysin and insulin degrading enzyme (IDE). Upregulation of neprilysin or IDE activity has exhibited efficacy in mouse models of AD^{110,111}. However due to their large structures and complex substrate recognition mechanisms¹¹², specificity engineering to reduce neprilysin and IDE activity on other physiological substrates has remained challenging⁶². Therefore, the identification of A β -degrading protease candidates that have smaller structures and open active sites suitable for protein engineering could expand the potential of using protease replacement therapy in AD.

Recent work discovered that the protease human kallikrein 7 (hK7) is capable of cleaving A β *in vitro* within the central hydrophobic core (KLVF↓F↓AED) reducing its propensity to aggregate¹⁰³. hK7 cleaves the peptide A β 8 (KLVFFAED) with high activity ($k_{\text{cat}}/K_M = 0.24 \pm 0.05 \mu\text{M}^{-1} \cdot \text{min}^{-1}$), is 4-fold smaller than IDE and NEP, and has an open active site which makes it amenable to use in high-throughput protein engineering screens. Additionally, hK7 is capable of degrading preformed A β fibrils, which has not been demonstrated for IDE and NEP¹⁰³. Because multiple studies have characterized both hK7's structural features, mechanism of action, and specificity, we chose this protease as a model

enzyme to develop a cell-based method for engineering secreted human proteases toward therapeutic targets such as the A β peptide^{103,113,114}.

1.6. Human tissue kallikreins

The human tissue kallikrein family comprises fifteen serine proteases forming the largest contiguous cluster of enzyme-encoding genes within the human genome¹¹⁵. The physiological roles of many of the kallikrein proteases remain unknown, however they are all localized to one or more human tissues, often being co-expressed with other kallikreins, suggesting important biological functions¹¹⁶. A few kallikrein cascades have been described including kallikreins 2, 3, and 5 in seminal plasma¹¹⁷ and kallikreins 5, 7, and 14 in skin^{118,119}. Of the fifteen proteases, human kallikrein 3 (hK3) and human kallikrein 7 (hK7) display chymotrypsin-like specificity and favor cleavage after large hydrophobic residues. All other thirteen proteases show trypsin-like specificity for hydrolysis of basic amino acids at P1¹¹³. While the precise physiological targets are difficult to determine, upregulation and overexpression of the majority of the tissue kallikreins has been linked to multiple cancers¹²⁰. Elevated levels of hK3, also known as prostate specific antigen (PSA), have been linked to the presence of prostate cancer and is used as a biomarker for screening and diagnosis¹²¹.

1.6.1. Human kallikrein 7 (hK7)

hK7 is a small 27 kDa secreted serine protease initially named stratum corneum chymolytic enzyme after being discovered in the skin. hK7 is translated as a pre-proenzyme with a 22 residue signal sequence that becomes processed after secretion of the protease from the cell. Post-translational modifications to hK7 include six disulfide bonds

required for proper folding and one N-linked glycosylation site at N246 (UniProt numbering used). After secretion, the protease remains inactive due to a 7 amino acid propeptide that prevents correct conformation of the active site. This propeptide is processed by cleavage within the sequence EEAQGDK↓I by a protease with trypsin-like specificity. hK5 is the only kallikrein capable of processing the hK7 zymogen *in vitro* and also co-localizes with hK7 in the skin suggesting its role in physiological hK7 activation¹²².

After activation, the new N-terminal isoleucine inserts itself into the hK7 structure forming an internal salt bridge with D204, which produces a functional active site¹¹⁴. The S1 pocket of hK7 differs from the other trypsin-like kallikreins in that at the bottom of the pocket the negatively charged D199 is replaced with N199 in hK7, which reduces interactions with positively charged P1 residues (UniProt numbering used). In addition, the pocket is more hydrophobic, accommodating bulky aromatic amino acids such as tyrosine, phenylalanine, and leucine. Multiple studies have demonstrated hK7's strong preference for cleavage after tyrosine, likely due to the favorable hydrogen-bonding between N199 and the hydroxyl group of tyrosine^{103,113,123}. In addition, the extended substrate specificity of hK7 displays a strong preference for arginine at the P4 or P5 position, and small amino acids such as alanine, glycine and serine at the P1' position¹⁰³. Two positively charged surface patches on hK7 may increase specificity for negatively charged substrates or aid in attachment to the cell surface (Figure 1.5)¹¹⁴.

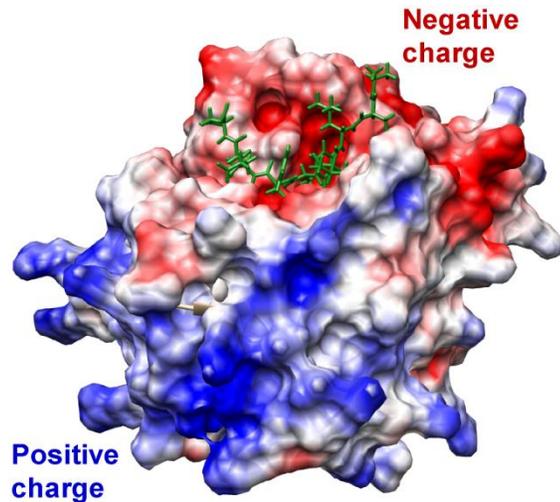


Figure 1.5. Docking model of human kallikrein 7 with A β 8.

hK7 has an open active site for the docking of substrates such as the A β 8 peptide (shown in green). Two positive charged surface patches are found on hK7 which may play a role in the protease's specificity by binding negatively charged regions in substrates. Additionally, the positively charged surface of the protease may bind to the negatively charged cell membrane after secretion. The image above was created with the Chimera software package using PDB:2QXI.

hK7 is most highly expressed in the skin, esophagus, and kidney and at lower levels in the brain and other tissues¹¹⁶. The physiologically preferred substrates of hK7 include intercellular cell adhesion proteins in the cadherin family including desmocollin-1, desmoglein-1, and E-cadherin^{122,124}. hK7 expressed in the skin plays an important role in the shedding of the stratum corneum for continuous regeneration of the skin¹¹⁹. However, elevated levels of hK7 activity in the epidermis have been linked to inflammation in severe skin diseases such as psoriasis, atopic dermatitis, and Netherton syndrome^{125,126}. Netherton syndrome has been genetically linked to a defect in the *SPINK5* gene, which encodes the multidomain protein LEKTI, an inhibitor of hK7^{127,128}. Furthermore, upregulation of hK7 activity has been implicated in metastatic tumor growth, especially in ovarian cancer. Increasing hK7 expression has been shown to correlate with ovarian cancer stage and a

high level of hK7 mRNA is associated with poor prognosis^{129,130}. It has been proposed that hK7 enhances metastasis by degrading the extracellular matrix which allows tumor cells to migrate from the primary tumor.

For hK7 to serve as a potential therapeutic in AD, it would be necessary to reduce its activity toward its physiological substrates and increase specificity for A β . Though hK7 can cleave after the phenylalanines in A β , hK7 has higher selectivity and activity for tyrosine at the P1 position. Because many candidate substrates exhibit tyrosine adjacent to the scissile bond, we hypothesized that a protease variant that does not recognize substrates with tyrosine will be less toxic and more potent. Further, we hypothesized that amino acid substitutions in and around the active site of hK7 can generate a structure that favors phenylalanine and excludes tyrosine.

1.7. Techniques

1.7.1. Fluorescent protein FRET

Förster resonance energy transfer (FRET) is the transfer of energy from an excited donor fluorophore to an acceptor fluorophore when there is significant overlap between the emission spectrum of the donor and absorption spectrum of the acceptor. FRET is strongly dependent on distance, orientation, and spectrum properties of the donor and acceptor fluorophores. The efficiency of energy transfer is proportional to the inverse sixth power of the distance between the centers of the two dipoles. Therefore, the donor and acceptor molecules must be within a distance of 1-10 nm of each other for efficient energy transfer.

Another requirement for FRET is that there must be sufficient separation between the excitation spectra of the donor and acceptor to selectively excite the donor. Also the

emission spectra of the donor and acceptor should be separated so that each fluorophore can be measured independently¹³¹. When these requirements are met, transfer of energy from the excited donor to the acceptor fluorophore results in emission of the acceptor molecule leading to the FRET-On state. When the fluorophores are separated, excitation of the donor fluorophore results in normal donor emission, termed the FRET-Off state.

FRET-based approaches have allowed the detection of protease activity and protein-protein interactions but are limited by the dynamic range between the FRET-On and FRET-Off states and sensitivity of the donor-acceptor pair. To address this problem, a CFP-YFP donor-acceptor pair was evolved for enhanced brightness and an improved dynamic range. Multiple rounds of mutagenesis and screening with fluorescence activated cell sorting (FACS) resulted in a CyPet-YPet pair with a 20-fold ratiometric FRET signal change¹³². Using this optimized FRET pair, events such as proteolytic cleavage within a substrate flanked by the CyPet and YPet molecules and protein-protein interactions can be easily monitored by flow cytometry or fluorescence microscopy (Figure 1.6)¹³³.

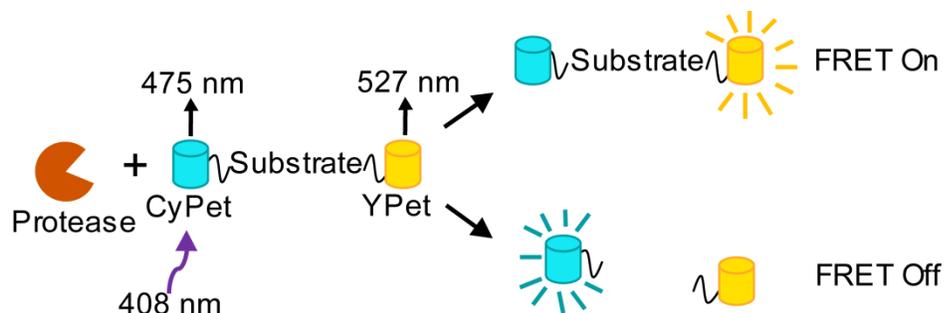


Figure 1.6. FRET probes link proteolytic activity with fluorescence.

A protease susceptible substrate is flanked by the cyan fluorescent protein CyPet and yellow fluorescent protein YPet. Excitation with violet light at 408 nm excites CyPet which can transfer energy to YPet if it is within a distance of 1-10 nm leading to yellow emission (FRET On). Cleavage of the substrate by a protease separates the two fluorophores so that violet light excitation leads to cyan emission (FRET Off).

Fluorescent protein FRET provides a unique tool for flow cytometry-based protease library screens for redesigning protease specificity or turnover rate. Aside from enabling the detection of protease activity, fluorescent protein sensors are relatively non-toxic to host cells and therefore could be co-expressed with a protease library for protease directed evolution. In order to use this screening strategy, the protease of interest must have an open active site compatible with a dumbbell-like substrate that is flanked by donor and acceptor FRET proteins. Also, it is necessary for the protease to be expressed in active form inside the host cell without killing the cell. Non-specific proteases, such as HIV protease, recognize essential proteins in *E. coli*, yeast, and mammalian cells and are cytotoxic^{57,134}.

1.7.2. *Fluorescence activated cell sorting (FACS)*

The use of FRET reporters is appealing for the discovery of proteases with novel activities and specificities. However, screening variants by fluorescence spectroscopy or microscopy is time and labor intensive and would limit the size of variant libraries. Fluorescence activated cell sorting (FACS) is unique among protein library screening methods in that millions of single cells can be analyzed rapidly in a high-throughput manner while measuring subtle changes in cellular fluorescence intensity¹³⁵.

Through flow cytometry, a suspension of cells is hydrodynamically focused into a single file line, where each cell can be individually excited by one or more laser beams. Most cytometers are equipped with lasers that excite at wavelengths of 407 nm, 488 nm, and 633 nm. Photomultiplier tube (PMT) detectors allow simultaneous detection of fluorescence from various dyes. The scattered light and fluorescence emissions from each laser are collected by detectors and optical signals are converted into digital signals. Using a computer attached to the cytometer, size and fluorescence of cell populations can be monitored in real time. Light scattered in the forward (FSC) direction gives a measurement of particle size, where larger particles will scatter more light. Side-scattered (SSC) light is collected at ninety degrees to the laser beam and is proportional to cell granularity or roughness. Gates are drawn on plots of FSC versus SSC and SSC height versus SSC width to focus the analysis on only individual healthy cells and avoid cell aggregates.

If the cytometer is equipped with a cell sorting device, individual cells with desired properties, such as size, shape, or fluorescence can be isolated. A gate is drawn around cells having the fluorescence profile of interest and eliminating low fluorescent cells based on the background fluorescence of a negative control. The stream containing the cell

sample is broken into droplets, with each droplet enclosing a single cell. Target cells can be separated from negative cells by first charging the droplet that encapsulates the target cell. Charged droplets are deflected into a collection chamber after passing between two high voltage deflection plates (Figure 1.7). The uncharged droplets are kept in the waste stream. With this technology, a subpopulation of interest can be separated from negative cells with high purity at speeds greater than ten thousand cells per second¹³⁶.

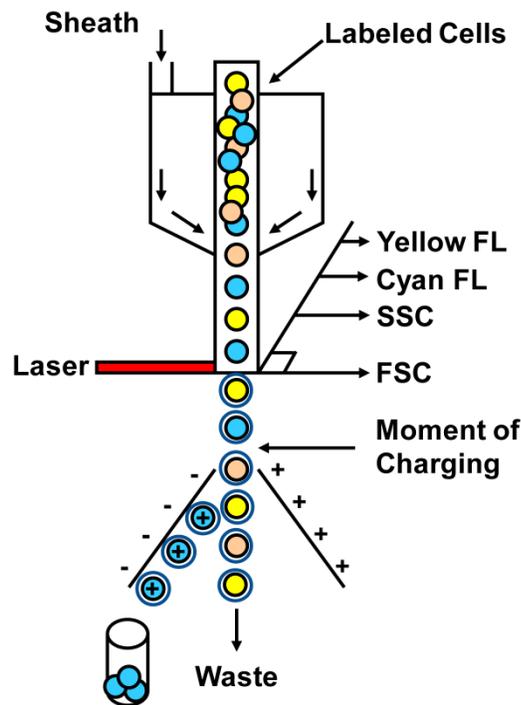


Figure 1.7. Cell based libraries screened using fluorescence activated cell sorting.

Cells are interrogated by a laser and the scattered light is separated into forward scatter (FSC), side scatter (SSC), and fluorescence (FL). Single cells are encapsulated by a droplet and droplets containing cells of interest are charged and collected after passing between two high voltage deflection plates.

1.7.3. *Strategies for library diversification*

The ability to screen large protein libraries can dramatically improve the probability of isolating rare variants which display a novel function of interest. In nature, proteins display slow rates of evolution due to the low rate of DNA mutagenesis found in most organisms. This slow rate of mutagenesis is important for long-term survival, since high mutation rates have been linked to conditions such as cancer¹³⁷. However, through the use of directed evolution, an individual protein of interest can be isolated and randomly or selectively mutated to generate a library of diverse variants. Although it is impossible to cover the entire mutational space of a protein, gene diversification methods perform an optimal sampling of the vast sequence space¹³⁸. This library can then be screened or selected for beneficial properties within the course of weeks. Directed evolution has been extremely effective in evolving hundreds of proteins for functions such as improved catalytic activity, specificity, fluorescence profiles, and stability¹³⁹⁻¹⁴².

The success of a directed evolution experiment is directly linked to the quality of the library generated and the development of a robust screening methodology for the activity of interest. Mutations can be introduced to a gene of interest using a variety of methods including chemicals¹⁴³, mutator strains⁷⁶, error-prone polymerase chain reaction (epPCR)¹⁴⁴, site-directed saturation mutagenesis¹⁴⁵, and genetic recombination^{146,147}. Chemical agents such as ethyl methanesulfonate (EMS) and ultraviolet irradiation are non-specific and can randomly mutate DNA. Mutator strains which contain low-fidelity polymerases and deactivated proofreading and repair enzymes cause the introduction of random mutations at low frequencies within the evolving gene and host genome¹⁴⁸.

Therefore, the ability of the host cell to tolerate random mutations to its genome places an upper bound on the *in vivo* mutagenesis rate.

Error-prone PCR is a conservative method for randomly introducing substitutions of single bases along the length of the gene of interest. This can be achieved through the use of a low-fidelity polymerase such as *Taq* and the addition of manganese ions to the PCR reaction¹⁴⁹. Magnesium ions are important for polymerase activity and the introduction of the larger manganese ions into the reaction causes deformation of the polymerase structure resulting in higher rates of base pair mismatches. In addition, unequal concentrations of dNTP monomers are used to reduce the natural bias of the polymerase for certain mutations promoting a more even representation of substitutions to each of the 20 amino acids. With epPCR, 0.15-3% of nucleotides within a gene can be targeted for mutagenesis by adjusting the concentrations of manganese and magnesium ions within the reaction¹⁴⁹.

Because mutations are randomly scattered using epPCR, different sites of the protein can be sampled for their importance to a given function. Additionally, beneficial substitutions can be introduced at sites that would be impossible to predict. This method is highly favorable when computational modeling is unable to determine beneficial substitutions as in the case of engineering proteins with unsolved structures. A disadvantage of epPCR with *Taq* polymerase is that transitions of A to G and C to T are more commonly seen during mutagenesis than transversions, which causes bias in the amino acid substitutions¹⁴⁹. Additionally, because the genetic code contains 64 codons encoding 20 amino acids, mutations at the wobble position, or the third position within a codon, often result in silent mutations where the amino acid is unchanged. A final consideration with epPCR is that because mutations are random, it is very rare for adjacent

bases to be mutated. This biases substitution of amino acids to residues which can be encoded by a single base pair substitution. These substitutions tend to be similar in property to the original amino acid, therefore dramatic changes in side chain structure or charge are less common.

For proteins with solved structures, computational methods can be used to predict residues in a protein that are important for functions such as binding to a target or substrate recognition. For example, molecular modeling of proteases often involves docking a protease with a substrate and determining which residues in the active site make intermolecular contacts with the substrate. Site-directed mutagenesis and kinetic characterization can aid in further identifying the function of different residues¹⁵⁰. Important residues can then be targeted for site-directed saturation mutagenesis by using primers that enable mutagenesis to all 20 amino acids with little bias. To encode for all 20 amino acids, the codon NNS is introduced into the primer at a specific position, where N is A/C/G/T and S is C/G. Common overlap PCR techniques such as splice overlap extension¹⁵¹ and gene assembly mutagenesis¹⁵² can be used to target multiple sites for saturation mutagenesis.

Site-directed saturation mutagenesis allows for the discovery of the optimal substitution at a given position of interest. By focusing saturation mutagenesis on a few residues, libraries are made at reasonable sizes for screening and retain some of the benefits of random mutagenesis since all residues are explored at a given position. However, determining which residues to mutate is still a very challenging task since often substitutions to one site can disrupt the structure of neighboring residues that are important for binding and catalysis¹⁵³. Because residues involved in the protein's function are

targeted, site-saturation mutagenesis often results in a high portion of the library being non-functional. Additionally, this method would miss residues distant from the active site that may operate synergistically with active site residues, but would be impossible to predict from the crystal structure.

Mutations from distinct clones with improved properties discovered from epPCR and site-saturation mutagenesis experiments can be combined in later rounds using genetic recombination. Recombination methods are most effective on a diverse population of functional variants. Additionally, the wild-type DNA sequence can be introduced during recombination to eliminate deleterious mutations. To access beneficial combinations of mutations, variants are fragmented with DNase and the fragments are reassembled into full-length genes by repeated cycles of overlap extension reaction¹⁴⁶. Another homologous recombination technique which does not require treatment with DNase, is staggered extension process (StEP)¹⁵⁴. StEP is a modified PCR method that uses a shortened annealing and extension time to promote crossover events along the full length of the template sequences. During both these methods, additional point mutations may be introduced if the DNA polymerase used has low-fidelity.

2. Development of a cell-based screen for protease engineering

Proteases have a unique ability to catalytically activate or inactivate proteins or peptides and present potential applications for use in therapy. However, most proteases are capable of cleaving multiple substrates and therefore would cause unwanted and toxic side-effects if introduced into the body. A handful of proteases have been approved for replacement therapy with naturally narrow substrate specificities, which limits their toxicity. However, the application of therapeutic proteases need not be limited to their native biological functions. The emergence of tools to engineer proteases with narrow specificities toward substrates involved in disease pathology could greatly expand their potential in therapy. As a model system, we have chosen to develop a cell-based method for engineering the human secreted protease kallikrein 7 (hK7) toward peptide targets such as the amyloid beta peptide that aggregate in disease. To develop a screening method for protease engineering, suitable expression hosts for hK7 were evaluated. hK7 formed insoluble aggregates when expressed in bacteria likely due to the mispairing of cysteines that form six disulfide bonds within hK7. Therefore yeast, which is equipped with enzymes that can aid in the formation of multiple disulfide bonds within the oxidizing compartment of the endoplasmic reticulum (ER), was assessed for its ability to fold mature, active hK7. hK7 was active and properly folded after expression in the ER of yeast. Therefore, a high-throughput screen for protease engineering was developed in yeast involving co-expression of a protease library and a peptide substrate exhibiting Förster resonance energy transfer (FRET) within the ER. After optimizing the co-expression conditions of the protease and FRET reporter substrate, significant *in vivo* activity of hK7 was measured by flow cytometry above background.

The dynamic range of the assay could enable screening large protease libraries using fluorescence activated cell sorting for the activity of interest.

2.1. Introduction

Human proteases represent a potential source of therapeutics capable of clearing toxic proteins from the body in a catalytic fashion. Because human proteases are naturally made in the body, they would likely cause little immune response when introduced in therapy providing a significant advantage over using proteases from bacterial or viral sources. Some of the most attractive proteases for engineering into therapeutics are the secreted human proteases, which often contain multiple disulfide bonds that can stabilize the tertiary structure of these molecules in the extracellular environment. For example, the approved protease therapeutics tissue plasminogen activator, coagulation factors VII, VIII and IX, and thrombin belong to the family of human secreted proteases and require multiple disulfide bonds for proper folding and activity¹⁵⁵. While engineering screens have been developed for membrane bound and intracellular proteases^{65,70,74,75}, general methods for engineering human secreted proteases are lacking.

Human kallikrein 7 (hK7) is a secreted serine protease that is highly expressed in skin tissue and plays an important role in cleaving cell adhesion proteins for the shedding of the skin^{116,122}. hK7 has a total of 12 cysteines within its mature protein sequence that form six disulfide bonds within the folded protease (Figure 2.1). Because hK7 has the unique ability to cleave both A β monomers and fibrils, but prefers cleavage after substrates containing tyrosine¹⁰³, we chose this protease as a model enzyme for developing a cell-based screen to engineer human secreted protease specificity. A major hurdle in engineering secreted proteases, such as hK7, is finding a suitable expression host capable of forming the correct

disulfide bonds required for conformational stability and biological activity. hK7 has previously been expressed as inclusion bodies in the cytoplasm of bacteria which required refolding¹¹³. Because protein refolding is labor and time intensive¹⁵⁶, this method of hK7 production would be inefficient for screening variants. To analyze the activity of multiple variants in a high-throughput fashion, hK7 would need to be expressed in an active form intracellularly. Ideally, the expression host would be microbial, since mammalian cells suffer from slow grow rates and low transformation efficiency which would limit the ability to screen large protease libraries for the activity of interest.

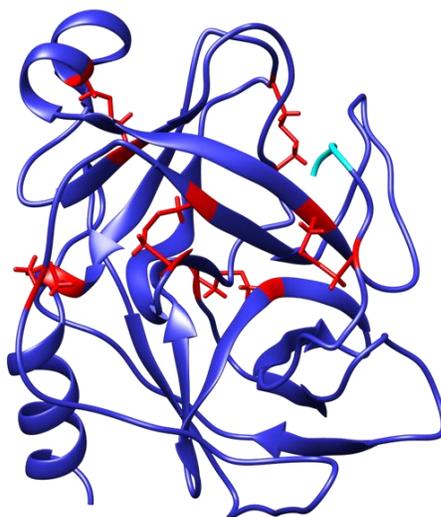


Figure 2.1. hK7 ribbon model with disulfide connectivities.

hK7 has six disulfide bonds that are required to form for proper folding and activity (shown in red). Additionally, after the propeptide sequence of hK7 is processed the N-terminus (teal) inserts itself into the hK7 structure causing a conformational change that produces a functional active site. The image above was created with the Chimera software package using PDB:2QXI.

While the cytoplasm of bacteria is naturally kept in a reduced state that disfavors disulfide bond formation, the periplasmic compartment is oxidizing and equipped with enzymes that efficiently form disulfide bonds within secreted proteins. Proteins become secreted to the periplasm through the secretory pathway which recognizes N-terminal signal sequences on proteins that direct their transport^{157,158}. Within the periplasm of *E. coli*, the enzyme DsbA catalyzes disulfide bond formation by transferring its one disulfide bond to substrate proteins causing DsbA to become reduced in the process^{159,160}. The integral membrane protein DsbB then reoxidizes the resulting reduced cysteines of DsbA to regenerate active DsbA¹⁶¹. Because incorrect disulfide bonds formed by DsbA may trap the protein in a non-native conformation, enzymes DsbC and DsbG act as proofreading chaperones and can break incorrect disulfide bond pairs to form the native disulfide bonds^{162,163}. Overexpression of DsbA and DsbC has been used to improve the formation of native disulfide bonds in eukaryotic proteins produced in bacteria¹⁶⁴⁻¹⁶⁷. However, often eukaryotic proteins contain multiple disulfide bonds with complex bonding patterns that overwhelm the *E. coli* machinery causing improper disulfide linkages. Therefore, although the proteins become oxidized, they are often locked in non-native states that may generate inclusion bodies¹⁶⁸.

An alternative microbial system for the production of recombinant proteins with multiple disulfide bonds is yeast. Yeast, like bacteria, have faster growth rates and higher transformation efficiencies than mammalian cells making them suitable for the development of protein engineering screens. However, yeast are better equipped than bacteria in forming the complex disulfide bonding patterns found in eukaryotic proteins. Within the yeast secretory pathway, disulfide bond formation occurs in the endoplasmic

reticulum (ER). The enzyme protein disulfide isomerase (PDI) resides in the ER and catalyzes both the formation and rearrangement of disulfide bonds until the protein folds into its native conformation¹⁶⁹. PDI achieves both these functions depending on if it is in an oxidized or reduced form. When PDI is in the oxidized state, it is capable of forming a disulfide bridge in a secreted protein by transferring its own disulfide bond. However, reduced PDI acts as an isomerase to break mispaired cysteines in proteins and aid in the formation of the correct disulfide bond¹⁷⁰. Additionally, the chaperone activity of PDI allows it to stabilize the misfolded structures of proteins and fold them into their native state. Another important protein in disulfide bond formation in yeast is the ER resident protein ER oxidoreductin (Ero1). Ero1 introduces oxidizing equivalents into the ER that promote reoxidation of reduced PDI¹⁷¹.

Within yeast, control mechanisms such as endoplasmic reticulum associated degradation (ERAD) are in place to prevent the buildup of misfolded proteins in the cell which can aggregate and cause cell stress. Misfolded proteins can be recognized by chaperones in the ER that bind the proteins and aid in their translocation to the cytosol¹⁷². Once in the cytosol, misfolded proteins are tagged with multiple ubiquitin molecules which target them for degradation by the proteasome¹⁷³. However, if misfolded proteins accumulate in the ER and overwhelm ERAD clearance mechanisms, the unfolded protein response (UPR) is activated. During UPR, translation in the ER is greatly reduced and the expression of molecular chaperones that assist in protein folding is increased¹⁷⁴. If normal cellular function is unable to be restored through these mechanisms, the UPR triggers cell death¹⁷⁵. Here, we have evaluated bacteria and yeast for their ability to correct fold the

secreted human protease hK7 for the development of a cell-based screen for protease engineering.

2.2. Results

2.2.1. Bacteria as an expression host for secreted proteases

Bacteria was first chosen as an expression host for hK7 due to its increased growth rate and transformation efficiency compared to eukaryotic cells and because a previous study had reported successful production of active hK7 in the periplasmic space of *E. coli*¹⁷⁶. Since bacteria lack many of the tRNAs necessary for efficient translation of human genes, the wild-type hK7 gene was codon-optimized for expression in bacteria. An OmpA signal sequence was fused to the N-terminus of hK7 to direct protein secretion to the oxidizing environment of the periplasm for formation of the six disulfide bonds within hK7. hK7 expression was placed under the lactose promoter and different concentrations of the synthetic lactose analogue isopropyl β -D-1 thiogalactopyranoside (IPTG) were tested for their ability to produce soluble, active hK7 in bacteria. Since IPTG is a membrane permeable, small molecule and does not need a transporter, a lower concentration of IPTG will correlate to a lower protein induction level. Using a lower concentration of IPTG for induction of another disulfide-containing protease was shown to increase cell viability after protein expression and increase the formation of correctly folded, active protease^{177,178}. Therefore, slowing the rate of protein synthesis to provide more time for proper folding and formation of disulfide bonds should favor the production of soluble protein.

To characterize the effect of inducer concentration on hK7 production, hK7 was expressed at a range of IPTG concentrations. Cell pellets with hK7 production and

negative control pellets containing a vector without the hK7 gene were ran on a Western blot after denaturing and reducing the samples. All samples expressing hK7 displayed two bands at ~27 and 29 kDa representing hK7 (Figure 2.2). The higher band likely represents unprocessed hK7, which has not been secreted and therefore retains the signal sequence, while the lower band may represent processed hK7. The OmpA signal sequence directing hK7 to the periplasm is ~2 kDa which would account for the molecular weight difference between these two bands. Cells induced with a final concentration of 5 μ M IPTG displayed a higher ratio of processed hK7 to unprocessed material compared to cells induced with 1 mM IPTG (Figure 2.2). A higher concentration of inducer may increase protein synthesis to a level where the secretion system becomes jammed causing a higher portion of the expressed hK7 to stay in the cytoplasm unfolded.

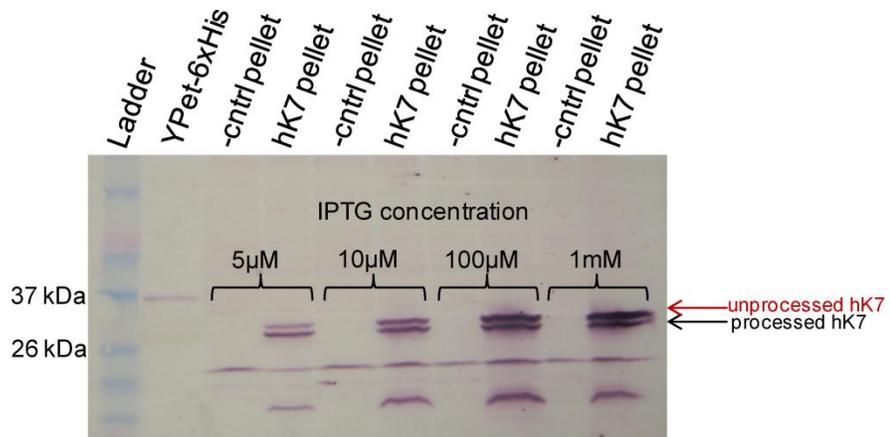


Figure 2.2. Lowering the IPTG inducer concentration increases the ratio of processed to unprocessed hK7.

hK7 expression in bacteria was placed under the lactose promoter which is inducible by the addition of IPTG, a synthetic lactose analogue, to the media. The presence of hK7 expression was visualized using a Western blot with an anti-C-terminal histidine tag antibody. Cells with hK7 expression showed two bands at the correct molecular weights of processed (27 kDa) and unprocessed hK7 (29 kDa), which represent hK7 without and with the periplasmic signal sequence, respectively. Reducing the IPTG inducer concentration increased the ratio of processed to unprocessed hK7 likely due to slowing the secretion rate through the periplasm.

To determine if hK7 produced in bacteria with a low concentration of IPTG was correctly folded and active, lysis buffer was added to cells expressing hK7 to isolate the soluble protein fraction. The presence of hK7 in both the cell pellet and soluble fraction was determined by Western blot after the samples were reduced. hK7 was present in the cell pellet after 1 hour of expression as a single band at ~27 kDa (Figure 2.3). After 4 hours of expression, a band above the original hK7 band was seen representing the accumulation of unprocessed hK7 over time (Figure 2.3). Soluble cell lysate samples showed an absence of the hK7 bands indicating that hK7 expressed in bacteria was not soluble and may be aggregating after expression. Aggregation could likely be attributed to the formation of non-specific intramolecular and intermolecular disulfide bonds due to cysteines within

hK7 mispairing. Additionally, misfolded protein intermediates often expose hydrophobic patches of amino acids, that would otherwise be buried in the core of the protein, that can increase self-association¹⁷⁹. Because the samples are reduced before loading the Western blot, higher order aggregates are broken apart and could not be visualized.

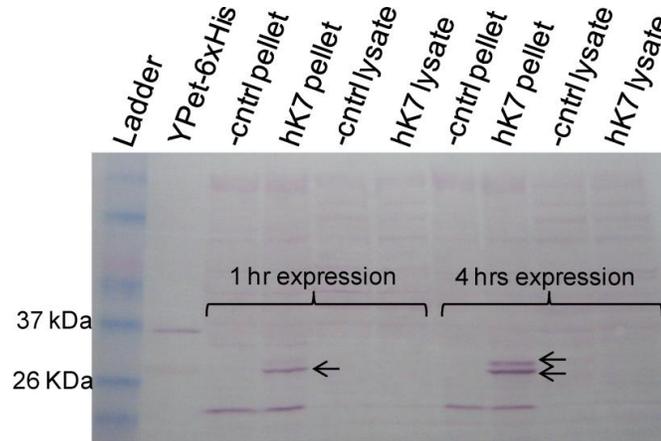


Figure 2.3. hK7 expressed in bacteria is insoluble.

Bacterial samples expressing hK7 for 1 or 4 hours were collected and a portion of the sample was lysed to isolate the soluble lysate fraction. Cell pellets and lysates from negative controls and cell expressing hK7 were run on a Western blot with an anti-C-terminal histidine tag antibody to visualize the presence of hK7. After 1 hour of expression, hK7 was present in the cell pellet as a single band at 27 kDa. A second band representing unprocessed hK7 was present after 4 hours expression representing the accumulation of unprocessed hK7 with its signal sequence. hK7 was not present in the soluble cell lysates indicating that hK7 is misfolded and likely aggregating after it is produced.

To improve the formation of the correct disulfide bonds within hK7, 5 mM reduced glutathione was added to the media during hK7 induction. Reduced glutathione can help to reshuffle disulfide bonds by reducing mispaired disulfide bonds in hK7 and allowing the bacterial disulfide isomerase DsbC to form the correct linkages. Additionally, a low expression temperature of 15°C was used to slow the rate of protein synthesis and secretion

which may give more time for proper folding. Low temperature expression of secreted proteases was shown to increase solubility and production of the correctly folded protease leading to increased activity^{176,177}. However, reduced glutathione and lowering the expression temperature had no effect on hK7 solubility for bacterial expression. To engineer hK7 activity and specificity, it is essential that hK7 be produced in a soluble and active form inside the cell, which would allow screening large libraries in a high-throughput fashion. Therefore, hK7 expression was tested in yeast, which is eukaryotic and better equipped with the machinery to fold complicated human proteins.

2.2.2. *Detection of hK7 activity in yeast using the GAL1 promoter*

Yeast represents an attractive alternative to *E. coli* for the production of proteins with complicated folds and multiple disulfide bonds due to the presence of the endoplasmic reticulum (ER) compartment. Within the oxidizing environment of the yeast ER, the enzyme protein disulfide isomerase (PDI) catalyzes the formation and shuffling of disulfide bonds¹⁶⁹. Additionally, the ER contains quality control mechanisms that dispose of misfolded proteins, reducing the possibility of aggregates building up within the cell¹⁷⁴. The wild-type hK7 gene was isolated from a human brain cDNA library and cloned into a yeast centromeric plasmid with a GAL1 promoter for galactose inducible expression. To ensure proper translation and folding within the ER, an invertase secretion signal sequence was fused to the N-terminus of hK7¹⁸⁰.

Yeast cells were transformed with the hK7 plasmid and protein expression was induced for 6 hours at 22°C by addition of galactose to the media. To determine if hK7 expressed in yeast was soluble, cells producing hK7 were lysed and the cell lysate was ran on a Western

blot to probe for the presence of hK7. As a negative control for activity, an inactive variant of hK7 was created with the catalytic serine at position 205 substituted with alanine (UniProt numbering used). The hK7 S205A variant should not retain any catalytic activity due to the absence of the serine nucleophile within the S1 pocket. A Western blot using a polyclonal antibody specific to hK7 showed a single band for both the inactive hK7 and wild-type hK7 at the correct molecular weight in the lysate of induced cells (Figure 2.4). Interestingly, a positive control sample of active hK7 produced from a mouse myeloma cell line showed a band at a slightly higher molecular weight (Figure 2.4). This may be due to the addition of an N-linked glycan at the known glycosylation site (N246) in the mammalian produced version, which remains unglycosylated in yeast. This initial data suggested that hK7 was efficiently produced in a soluble form in yeast and that activity should be detectable inside the cell lysate.

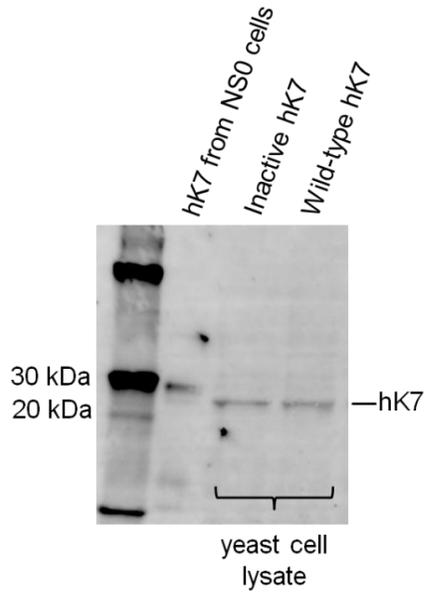


Figure 2.4. hK7 is present in the soluble cell lysate after expression.

Soluble lysate from cells expressing active hK7 and inactive hK7 was run on a Western blot and probed with a polyclonal anti-hK7 antibody. A single band at the correct molecular weight of 27 kDa was present in both samples. As a positive control, hK7 produced from a mouse myeloma cell line was also detected by Western blot. This band appeared slightly higher than the yeast derived hK7, indicating an increase in molecular weight. This increase is likely due to glycosylation of N246, which is the known N-linked glycan site in hK7.

To verify that hK7 was properly folded and active in yeast, hK7 activity in cell lysate was measured by adding peptide substrates flanked by fluorescent proteins capable of FRET and analyzing the change in FRET over time. Lysate of yeast producing active hK7 showed high conversion of the peptide substrates PRVMFFT and PRVMYYT that was comparable to a 10 nM preparation of commercial hK7 expressed in mammalian cells (Figure 2.5). Additionally, the selectivity for cleavage after tyrosine (Y) over phenylalanine (F) was maintained for hK7 produced in yeast cells further confirming formation of the correctly folded structure and S1 pocket for catalysis. Yeast cells expressing the inactive hK7 variant S205A displayed no background activity in crude cell

lysate toward the two probes, verifying that no endogenous yeast proteases are responsible for the activity seen in active hK7 samples (Figure 2.5). Because yeast can fold hK7 in the same manner as mammalian cells, but grow much faster and are easier to transform with DNA, they provide an exceptional host to build our cell-based screen for protease engineering.

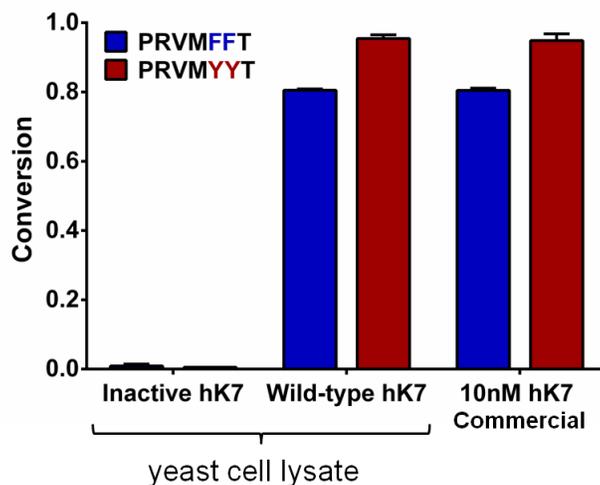


Figure 2.5. hK7 expressed in yeast is correctly folded and active.

Cells expressing active hK7 or inactive hK7 were lysed and FRET reporter substrates were added to the lysate to determine if hK7 produced in yeast was active. Lysate of cells producing active hK7 showed conversion of the two substrates PRVMFFT and PRVMYYT that was comparable to a 10 nM commercial preparation of hK7 produced in mammalian cells. To confirm that activity was specific to hK7, lysate of cells producing the inactive hK7 variant displayed no background activity on the two probes (N=3). Data represented as mean \pm SD.

To determine if protease expression was toxic to the yeast cells, cell growth was measured over time for cells producing active hK7 or inactive hK7 at 30°C and 22°C. Yeast grow best at 30°C, however at this expression temperature, cells producing active hK7 showed a dramatic reduction in growth rate compared to cells expressing inactive hK7

(Figure 2.6). The reduction in cell growth can likely be attributed to the non-specific activity of hK7 which may cut essential yeast proteins compromising the cell's ability to grow and divide. Because protease activity is directly correlated with temperature, the expression temperature was lowered to 22°C to reduce toxicity associated with protease expression. At this lower expression temperature, cell producing active hK7 or inactive hK7 displayed smaller differences in their growth rates over time. Reducing host toxicity is critical in the development of protein engineering methods since cell death during expression would prevent the isolation of variants during screening.

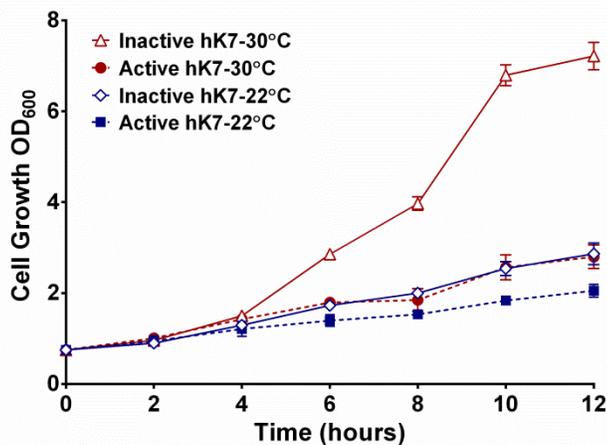


Figure 2.6. hK7 expression is toxic to yeast.

Growth curves of cells expressing active hK7 at 30°C (red circles) and 22°C (blue squares) and cells expressing the inactive hK7 variant S205A at 30°C (red open triangles) and 22°C (blue open diamonds) (N = 3). Expression of active hK7 was toxic to yeast and reduced cell growth rate compared to cells expressing inactive hK7. Reducing the expression temperature from 30°C to 22°C reduced protease activity and toxicity to the cell. Data represented as mean \pm SD.

“Adapted with permission from Guerrero, J. L., O’Malley, M. A., and Daugherty, P. S. (2016) Intracellular FRET-based screen for redesigning the specificity of secreted proteases. *ACS Chem. Biol.* 11, 961–970. Copyright 2016 American Chemical Society.”

2.2.3. *Development of the PrECISE cell-based screen for protease engineering*

To identify A β selective hK7 variants, a screening method termed protease evolution via cleavage of an intracellular substrate (PrECISE) was developed using co-expression of a peptide substrate reporter and candidate protease variant within the endoplasmic reticulum (ER) of yeast (Figure 2.7). Because hK7 is secreted with six disulfide bonds, it is not amenable to expression in or display on *E. coli*, nor to cytosolic expression in yeast or animal cells. To increase substrate accumulation within the ER and probability of cleavage by hK7, a C-terminal yeast ER retention sequence (HDEL) was added to the substrate¹⁸¹. To reduce toxicity of the co-expressed protease (hK7), the protease was targeted for secretion without ER retention. Both protease and reporter substrate are simultaneously expressed from the GAL1 promoter (Figure 2.7). The target substrate was flanked by fluorescent proteins, CyPet and YPet, optimized for intracellular FRET with violet light excitation¹³². Thus substrate cleavage abolishes FRET and increases cyan fluorescence, enabling separation of cells with cleaved substrates using fluorescence activated cell sorting (FACS) (Figure 2.7).

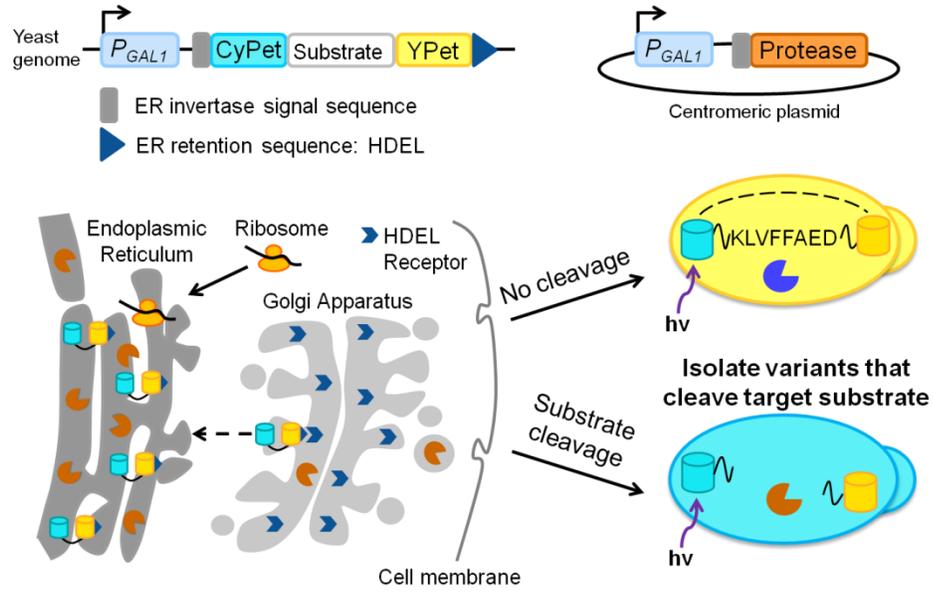


Figure 2.7. Protease evolution via cleavage of an intracellular substrate (PrECISE).

A gene encoding the A β 8 substrate flanked by fluorescent proteins CyPet and YPet was integrated into the yeast genome and hK7 was expressed from a galactose inducible plasmid. Both proteins are translated into the ER while only the FRET reporter substrate contains an ER retention signal (HDEL). Cleavage of the target substrate by a protease variant disrupts FRET and leads to an increase in cyan fluorescence enabling isolation of cells with cleaved substrates using FACS.

Substrate specificity was selected for by utilizing the toxicity associated with intracellular expression of non-specific, heterologous proteases⁵⁷. Thus, we reasoned that variants with narrowed substrate specificity but wild-type activity would effectively outcompete variants with broad specificity. Indeed, hK7 expression with ER retention was toxic, evidenced by rapid arrest of cell growth. For comparison, when expressed at 22°C and secreted without ER retention, hK7 exhibited mild toxicity with a slight reduction in growth rate relative to cells expressing inactive hK7 (S205A) (Figure 2.6). Additionally, to avoid false positives resulting from YPet cleavage, YPet expression was measured by

direct excitation at 488 nm and only cells with high blue-excited yellow fluorescence and increased violet-excited cyan fluorescence were recovered.

Library screening conditions were optimized by analyzing the separation between cells expressing active or inactive hK7 on flow cytometry. The A β 8 substrate encoding gene was integrated into the yeast genome and hK7 was expressed from a galactose inducible plasmid to simplify library generation and clone analysis. Cells co-expressing hK7, or inactive hK7 S205A, and the A β 8 substrate were analyzed using flow cytometry after expression at 22°C, 30°C, and 37°C at various time points. Substrate cleavage was interrogated by measuring the ratio of cyan to yellow fluorescence. Cells expressing active hK7 and the A β 8 FRET reporter at 30°C and 37°C, showed a significant reduction in cell density compared to cells expressing the inactive hK7 S205A variant. The expression temperature was therefore reduced to 22°C to reduce non-specific protease activity and improve cell viability. After 6 hours of expression, significant differences in conversion of the FRET reporter were observed by flow cytometry between the samples expressing active or inactive hK7 (Figures 2.8a,b). The maximum separation between positive and negative cells was observed after 16 hours of expression (Figure 2.8c,d). However at this longer co-expression time, cells expressing active hK7 displayed significant growth rate reduction compared to cell expressing the inactive hK7.

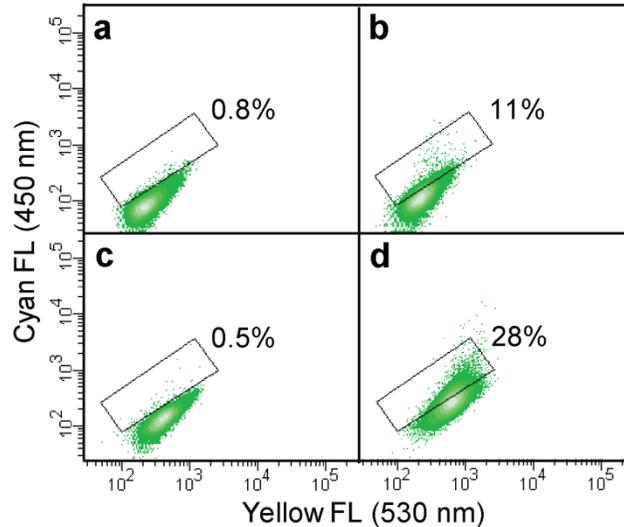


Figure 2.8. Detection of hK7 activity *in vivo* using PrECISE methodology.

Flow cytometry analysis of cell populations co-expressing inactive hK7 variant S205A (a,c) or wild-type hK7 (b,d) with the A β 8 FRET reporter substrate after 6 hours (a,b) and 16 hours (c,d) at 22°C. The dynamic range between the positive and negative cell populations increased with co-expression time.

To verify that the increase in cyan/yellow fluorescence observed for active hK7 co-expression was due to hK7 activity within the substrate linking region of the FRET reporter, intracellular conversion of a negative hK7 substrate (GGSGSGGS) was measured. Co-expression of active hK7 with the GGSGSGGS FRET reporter substrate at 22°C for 16 hours caused no change in the ratio of cyan fluorescence to yellow fluorescence as measured by flow cytometry (Figure 2.9). This result further confirmed that hK7 does not cut outside of the substrate linking region to produce a false positive signal. Furthermore, intracellular conversion of an optimal hK7 substrate (PRVMYYT) was much higher compared to the A β 8 (KLVFFAED) FRET substrate indicating that subtle differences in hK7 cleavage kinetics for various substrates can be effectively measured using the PrECISE screen.

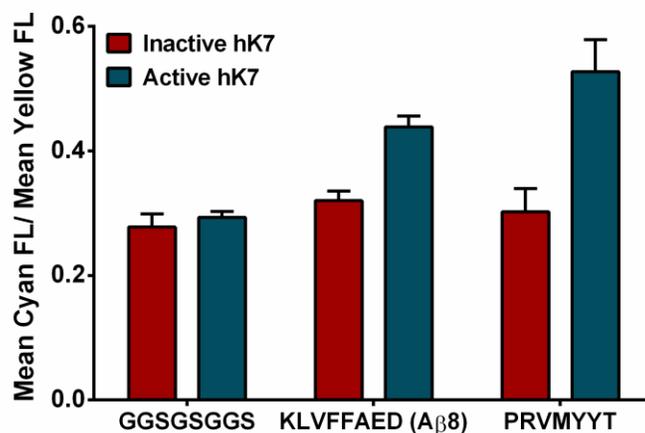


Figure 2.9. PrECISE methodology allows effective discrimination between different hK7 substrates *in vivo*.

Active hK7 or inactive hK7 were co-expressed at 22°C for 16 hours with three different FRET reporters containing either a negative hK7 substrate (GGSGSGGS), the A β 8 peptide (KLVFFAED), or an optimal hK7 substrate (PRVMYYT). Intracellular conversion of the FRET reporters was analyzed by flow cytometry by measuring the ratio of cyan to yellow fluorescence. No significant conversion of the GGSGSGGS reporter by active hK7 was measured. Active hK7 co-expression reduced FRET and caused an increase in cyan fluorescence for the A β 8 and PRVMYYT probes, indicating cleavage within the substrates (N = 3). Data represented as mean \pm SD.

To determine the optimal expression conditions for screening an hK7 library using the PrECISE methodology, enrichment of active hK7 cells from a background of inactive hK7 cells was determined for various co-expression parameters such as time and temperature. We hypothesized that at shorter co-expression times with an expression temperature of 22°C there would be less cellular toxicity associated with wild-type hK7 expression and cells expressing the inactive variant would be less likely to outgrow active hK7 cells during screening. Though the separation between positive and negative cells is much smaller at shorter co-expression times, oversampling the number of cells sorted with high cyan/yellow fluorescence ratio should lead to active hK7 enrichment. Cells expressing

active hK7/A β 8 and cells expressing inactive hK7 S205A/A β 8 were mixed at a ratio of 1:10 based on optical density measurements after expression at 22°C for 6 hours. Yeast cells with high cyan/yellow fluorescence were sorted over two rounds and plated after each round of sorting (Figure 2.10). We sequenced ten random clones after each round of sorting and determined that 7/10 sequences were active hK7 after the second sort. An enrichment factor of ~2.7 for wild-type hK7 was calculated for each round of sorting. In contrast, when the same enrichment experiment was performed after co-expression at 22°C for 16 hours, no enrichment of active hK7 was measured. This indicated that cells expressing the inactive variant had outcompeted cells with active protease likely due to their faster growth rate during long co-expression times. To improve enrichment of active hK7 variants during library screening, cells co-expressing the hK7 library and CyPet-A β 8-YPet were induced at 22°C for 6 hours.

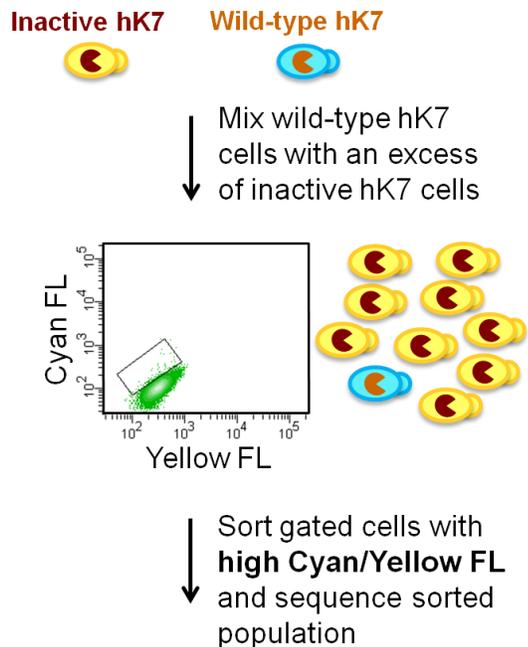


Figure 2.10. Workflow for enrichment of active hK7 from a background of cells expressing inactive variants to determine optimal co-expression conditions for library screening.

Cells expressing active hK7 with the A β 8 FRET reporter were mixed with an excess of cells expressing inactive hK7 and the A β 8 FRET reporter at various ratios (1:10, 1:100). Enrichment of active hK7 cells from the background of inactive variants was determined for various expression parameters by sorting cells with high cyan/yellow fluorescence using FACS. Random colonies from the enriched population were sequenced to identify the percentage of cells containing active protease after one and two rounds of sorting.

2.3. Discussion

To develop a cell-based method for engineering the specificity of secreted proteases, both bacteria and yeast were evaluated for their ability to express active hK7 which requires six disulfide bonds for proper folding. Human secreted proteases are attractive targets for use in therapy due to their enhanced stability imparted by their intermolecular disulfide bonds and because they would cause less immunogenicity if introduced into the body compared to a bacterial or viral protease. However, the expression of correctly folded eukaryotic proteins with multiple disulfide bonds remains a challenging feat to accomplish

in microbial systems. Because of this, no general methods have been developed and applied to engineering secreted proteases in a high-throughput fashion. The lack of methods for engineering secreted proteases greatly limits their potential to be redesigned to cleave new targets which are relevant in disease with high specificity.

Bacteria was first evaluated for its ability to produce active hK7 because of its fast growth rate and high transformation efficiency. hK7 expression was directed to the periplasm of bacteria to promote disulfide bond formation. However, hK7 accumulated within *E. coli* as misfolded aggregates. For short expression times and low inducer concentrations, the majority of misfolded protein was localized to the periplasm. Whereas, longer expression times resulted in the accumulation of aggregated protein in the cytoplasm. This change in localization is likely due to the secretion system becoming overcrowded causing the majority of the expressed hK7 to stay in the cytoplasm unfolded. While multiple parameters were tested to enhance the formation of disulfide bonds in hK7, no improvements were seen in hK7 solubility or activity upon expression in bacteria. Because many of the native disulfide bonded proteins produced in the periplasm of bacteria contain two or fewer disulfides, mispairings are less likely to occur¹⁶³. However, in a eukaryotic protein such as hK7, which contains six disulfide bonds, mismatches have a higher probability of occurring and likely overwhelmed the *E. coli* machinery, which is less equipped to recognize eukaryotic protein folding patterns.

While yeast do not glycosylate proteins in the same manner as mammalian cells, they have been shown to form the correct disulfide bonds in eukaryotic proteins with complicated folds^{182,183}. hK7 secretion to the endoplasmic reticulum of yeast resulted in correctly-folded, active protease indicating formation of the six disulfide bonds within

hK7. hK7 produced in yeast had the same substrate selectivity as mammalian cell produced hK7 confirming that the active site and S1 pocket were equivalent. However, active hK7 expression in yeast mediated excessive toxicity to the cell, which was not observed for expression of an inactive hK7 variant. Therefore, non-specific protease activity rather than merely hK7 expression led to reduced cell growth. It is very hard to determine which yeast proteins become cleaved non-specifically due to hK7 activity since multiple yeast proteins contain tyrosine and phenylalanine residues that could be favored as sites of cleavage by hK7. However, by reducing the expression temperature, we were able to reduce non-specific activity of hK7 and also toxicity to the cell comparable to the inactive hK7 variant.

hK7 intracellular activity in yeast was coupled to the fluorescence output of a co-expressed FRET reporter substrate to enable the screening of large protease libraries by fluorescence activated cell sorting (FACS). The protease and FRET reporter are co-expressed in the endoplasmic reticulum to promote disulfide bond formation in hK7 and also increase the probability of substrate cleavage occurring. Using the PrECISE methodology, protease libraries created through random mutagenesis or site-saturation mutagenesis can be screened for activity toward a peptide sequence of interest. Because the FACS-based approach is a quantitative method, changes to the co-expression parameters such as time and temperature can be easily assessed and optimized in order to favor variants with improved activity. Here, we have optimized screening conditions for hK7 intracellular activity toward the hydrophobic core of the A β peptide. Reducing the co-expression time and temperature were both found to be critical in the enrichment of cells with the active protease from a background of cells expressing an inactive variant.

2.4. Materials and methods

2.4.1. hK7 bacterial vector construction and expression

The active wild-type hK7 gene consisting of 226 amino acids was codon-optimized for expression in *E. coli* (DNA 2.0) and cloned between the EcoRI and HindIII sites in the vector pLAC22¹⁸⁴, which contains a lactose promoter. Expression of genes under the lactose promoter can be turned on by the addition of lactose or a lactose derivative such as the small molecule isopropyl β -D-1 thiogalactopyranoside (IPTG) to the media. An OmpA signal sequence (MKKTAIAIAVALAGFATVAQA) was placed upstream of the hK7 gene to direct its export to the periplasm for the formation of disulfide bonds. Additionally, a histidine tag (GHHHHHH) was fused to the C-terminus of hK7 to simplify downstream purification of the protease. The *E. coli* strain MC1061 was transformed with the pLAC22 plasmid containing the hK7 gene and transformed colonies were selected for on plates containing ampicillin (100 μ g/mL). As a negative control, MC1061 cells were transformed with an empty pLAC22 vector.

To test for hK7 expression in bacteria, overnight cultures containing either the empty vector or the hK7 vector were grown in 5 mL Luria broth (LB) media with ampicillin (100 μ g/mL). In the morning, cells were subcultured 1:50 into 40 mL LB with ampicillin and grown for 2 hours until cell growth reached an optical density of 0.6. Cells were then induced with IPTG for hK7 expression. Protease expression was tested under various induction conditions such as IPTG concentration (5 μ M - 1 mM), temperature (15°C and 25°C), time for expression (1 – 16 hours), and addition of reduced glutathione (5 mM) to the media. After expression, a small sample of 8×10^8 cells was spun down at 3000g for 5 minutes for analysis of the cell pellet fraction by Western blot. The rest of the 40 mL

culture was spun down at 3000g for 10 minutes at 4°C to remove the media. Cell samples were then frozen at -80°C for 2 hours to improve cell lysis and soluble lysate extraction. Cell pellets from 40 mL cultures were lysed by adding 3 mL B-PER II reagent (Thermo Scientific) and pipetting until the mixture was homogenous. Cell samples were then gently shook at room temperature for 30 minutes. Soluble cell lysate was isolated by centrifugation at 20,000g for 25 minutes.

2.4.2. *hK7 yeast vector construction and expression*

The gene encoding the proform of human kallikrein 7 (hK7) was amplified from a human brain cDNA library (BD Biosciences) with primers designed on the basis of the known hK7 coding sequence. The mature form of hK7 was then amplified to contain the invertase secretion signal (MLLQAFLFLLAGFAAKISA) at the N-terminus and a 6x histidine (GHHHHHH) purification tag at the C-terminus. Inactive hK7 variant S205A was created using PCR-driven overlap extension¹⁵¹ with forward primer 5' GCCTGCAATGG TGACGCAGGGGGACCGTTGG and reverse primer 5' CCAACGGTCCCCCTGC GTCACCA TTGCAGGC. The resulting fragments were cloned between the HindIII and PmeI sites in the pYC2/NT plasmid (Invitrogen), which contains a GAL1 promoter for galactose-inducible expression (Figure 2.11). Electrocompetent DH5α *E. coli* were transformed by electroporation with the ligated vector and cells were plated on media supplemented with ampicillin (100 µg/mL). Plasmid DNA was isolated using a Zyppy plasmid miniprep kit (Zymo Research) and the yeast strain JYL69 (MATa *ura3-1 ADE2+ his3-11,15 leu2-3,112 trp1-1 can1-100*; isogenic to W303) was transformed with 1 µg of

the hK7 plasmid using the high efficiency PEG/lithium acetate method¹⁸⁵ and plated on minimal SD-Ura media.

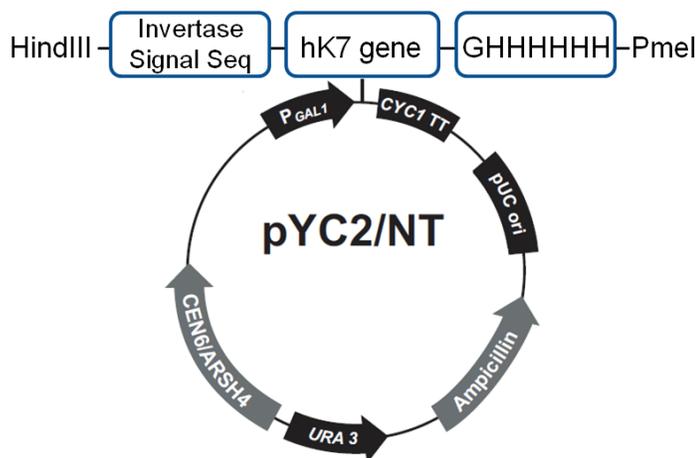


Figure 2.11. Plasmid map of pYC2/NT for hK7 expression and secretion in yeast.

The pYC2 vector has a GAL1 promoter for protein expression in yeast. The vector is centromeric and contains a CEN6/ARSH4 sequence that maintains the plasmid at 1-2 copies per cell. Additionally, ampicillin and *URA3* selection markers allow for selection and maintenance of the plasmid in bacteria and yeast, respectively. The hK7 gene insert contains an N-terminal secretion signal to direct protein translation to the ER and a C-terminal histidine tag for downstream purification and characterization.

To determine if hK7 was correctly folded and active in yeast, cells were cultured overnight in 10 mL SD-Ura media with 2% raffinose to an OD₆₀₀ of 4. Cells were then subcultured to an OD₆₀₀ of 0.75 in 40 mL fresh SD-Ura media with 2% raffinose and 2% galactose for induction at a specific temperature and length of time. Expression parameters were varied to assess the effect of temperature (22°C, 30°C, and 37°C) and expression time (0 – 16 hours) on hK7 activity. After expression, cell pellets were isolated from the media by centrifugation at 5000g for 10 minutes. Cells were lysed by adding 3 mL Y-PER reagent (Thermo Scientific) and gently shaking the homogenous mixture at room

temperature for 30 minutes. Soluble lysate fractions were recovered by spinning the solution at 15,000g for 5 minutes.

2.4.3. Construction of FRET reporter yeast strains

The yeast strain JYL69 (MATa *ura3-1 ADE2+ his3-11,15 leu2-3,112 trp1-1 can1-100*; isogenic to W303) was constructed to express FRET reporter substrates flanked by optimized cyan and yellow fluorescent proteins, CyPet and YPet¹³². Individual yeast strains were constructed for expression of FRET reporter substrates containing the sequences KLVFFAED (A β 8), GGS GSGGS, or PRVMYYT. FRET substrates were amplified to contain the N-terminal invertase secretion signal (MLLQAFLFLLAGFAAKISA) to direct protein translation to the ER and a C-terminal yeast ER retention signal sequence, HDEL¹⁸¹. The resulting fragments were cloned between the HindIII and PmeI sites in the pYES3/CT plasmid (Invitrogen), which contains a GAL1 promoter for galactose-inducible expression (Figure 2.12). Electrocompetent DH5a *E. coli* were transformed by electroporation with the ligated vector with the 2- μ m origin removed and cells were plated on media supplemented with ampicillin (100 μ g/mL). Plasmid DNA was isolated using a Zippy plasmid miniprep kit (Zymo Research) and the resulting plasmid containing the FRET reporter substrate was digested with Bsu361, which cuts once within the TRP1 locus. The linearized plasmid was integrated into the yeast genome using the high efficiency PEG/lithium acetate method¹⁸⁵ and plated on minimal SD-Trp media.

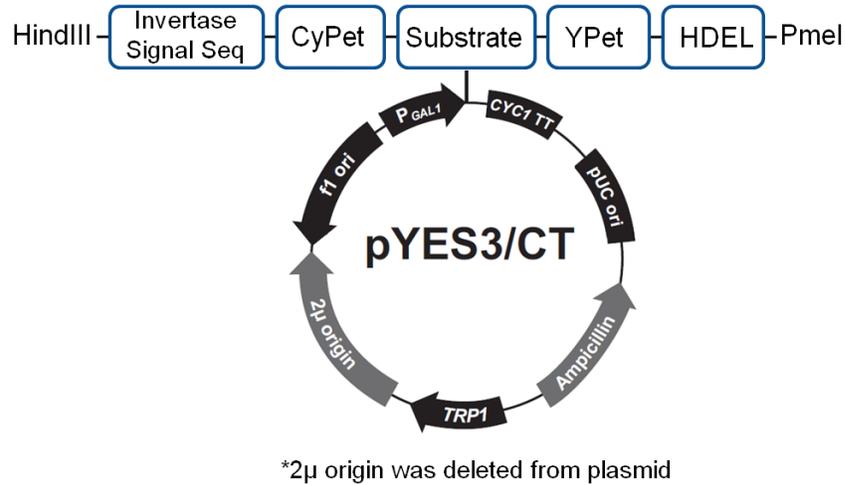


Figure 2.12. Plasmid map of pYES3/CT for FRET reporter expression in yeast.

The pYES3 vector has a GAL1 promoter for protein expression in yeast. The 2μ origin was deleted from the vector and genomic integration into yeast is achieved by cutting once within the TRP1 locus. Integration occurs at 1-2 copies per cell. Additionally, ampicillin and *TRP1* selection markers allow for selection and maintenance of the plasmid in bacteria and yeast, respectively. The FRET reporter gene contains an N-terminal secretion signal to direct protein translation to the ER and a C-terminal sequence HDEL for ER retention in yeast.

To identify a highly fluorescent Aβ8 substrate expressing cell line for subsequent library transformations, individual clones expressing the Aβ8 FRET reporter were assessed via flow cytometry for yellow fluorescence due to FRET. Cells were cultured overnight in SD-Trp media with 2% raffinose to an OD₆₀₀ of 4. Cells were then subcultured to an OD₆₀₀ of 0.75 in fresh SD-Trp media with 2% raffinose and 2% galactose for induction at a specific temperature and length of time. After expression, the cells were pelleted by centrifugation at 5,000g for 5 minutes and resuspended in 1x PBS for analysis on a FACSAria flow cytometer (BD Biosciences) with violet light excitation. Cells were pre-gated for high blue-excited yellow fluorescence to ensure that all cells analyzed had produced full-length FRET reporters. After analysis of 15 individual clones, a clone with

high violet-excited yellow fluorescence was chosen for co-transformations with the protease vector.

2.4.4. *Western blots for hK7 expression*

Cell lysate samples (32 μ L) or cell pellet fraction diluted with 32 μ L ddH₂O were incubated with 8 μ L of 5x SDS-reducing loading buffer (Thermo) for 5 minutes at 95°C. Reduced samples (30 μ L) were then loaded onto a 4-20% Tris-Glycine gel (NuSep) and 5 μ L of Novex Sharp Prestained Protein Standard (Life Technologies) was loaded for a molecular weight ladder. As a positive control, 20 ng of recombinant hK7 (R&D Systems) produced from a mouse myeloma cell line was used. Protein samples were separated in 1x Tris-Glycine running buffer using an XCell SureLock Mini-Cell electrophoresis system (Thermo Fisher) and applying a voltage of 150 V for 1 hour.

Protein samples separated by SDS-PAGE were transferred to a PVDF membrane using a Mini-PROTEAN tetra cell and mini trans blot module (Bio-Rad) for 1 hour at 100 V. PVDF membranes were activated with 100% methanol for 30 seconds then rinsed with ddH₂O prior to transfer. After protein transfer, the membrane was rinsed briefly in 1x Tris-buffered saline (TBS) and then incubated with 8 mL Blocking Buffer for Fluorescent Western blots (Rockland) for 1 hour at room temperature in a closed container. The membrane was then incubated with polyclonal goat anti-hK7 (R&D Systems) diluted 1:400 in 8 mL blocking buffer for 16 hours at 4°C in the dark. To reduce non-specific binding of the primary antibody, membranes were washed three times with 0.05% Tween in TBS. Membranes were then incubated with the Alexa Fluor680 donkey anti-goat secondary antibody (Thermo Fisher) diluted 1:2500 in blocking buffer for 1 hour at room

temperature in the dark. Membranes were washed briefly in 0.05% Tween in TBS before imaging using a ChemiDoc MP Imager (Bio-Rad).

For colorimetric detection of histidine-tagged proteins by Western blot, membranes were incubated with a mouse anti-C-terminal histidine-tag primary antibody (Thermo Fisher) and a goat anti-mouse secondary antibody conjugated to alkaline phosphatase (Thermo Fisher). Blots were developed by the addition of 1-Step NBT/BCIP (Thermo Fisher) for 5-15 minutes until the desired color developed. Developed membranes were then washed with distilled water.

2.4.5. *Detecting hK7 activity in vivo*

Using flow cytometry, expression conditions were optimized to detect hK7 activity in yeast with the co-expressed intracellular A β 8 FRET reporter. hK7 and FRET reporter expression were both under control of the GAL1 promoter, which is inducible by the addition of galactose to the media. Triplicate samples of cells expressing wild-type hK7 and the A β 8 FRET reporter were analyzed after expression at 22°C, 30°C, and 37°C for various time points (0 – 16 hours) and compared to cells co-expressing the inactive hK7 variant S205A and the A β 8 FRET substrate to determine the optimal dynamic range of the assay. Cells were cultured overnight in SD-Trp-Ura media with 2% raffinose to an OD₆₀₀ of 4. Cells were then subcultured to an OD₆₀₀ of 0.75 in fresh SD-Trp-Ura media with 2% raffinose and 2% galactose for induction. After expression, the cells were pelleted by centrifugation at 5,000g for 5 minutes and resuspended in 1x PBS for analysis on a FACSAria flow cytometer (BD Biosciences) with violet excitation.

3. Design of a human protease targeting amyloid beta peptides

Proteases are attractive as therapeutics given their ability to catalytically activate or inactivate their targets. However, therapeutic use of proteases is limited by insufficient substrate specificity, since off-target activity can induce undesired side-effects. In addition, few methods exist to enhance the activity and specificity of human proteases, analogous to methods for antibody engineering. Given this need, a general methodology termed protease evolution via cleavage of an intracellular substrate (PrECISE) was developed to enable engineering of human protease activity and specificity toward an arbitrary peptide target (Chapter 2). PrECISE relies on co-expression of a protease and a peptide substrate exhibiting Förster resonance energy transfer (FRET) within the endoplasmic reticulum of yeast (Chapter 2). Use of the FRET reporter substrate enabled screening large protease libraries using fluorescence activated cell sorting for the activity of interest. To evolve a human protease that selectively cleaves within the central hydrophobic core (KLVF↓F↓AED) of the amyloid beta (A β) peptide, PrECISE was applied to human kallikrein 7, a protease with A β cleavage activity but broad selectivity, with a strong preference for tyrosine (Y) at P1. Screening randomly mutated hK7 libraries for activity on A β 8 yielded a protease variant which displayed up to 30-fold improvements in A β selectivity mediated by a reduction in activity toward substrates containing tyrosine. Additionally, the increased selectivity of the variant led to reduced toxicity toward PC12 neuronal-like cells and 16-1000-fold improved resistance to wild-type inhibitors. PrECISE thus provides a powerful high-throughput capability to redesign human proteases for therapeutic use.

“Adapted with permission from Guerrero, J. L., O’Malley, M. A., and Daugherty, P. S. (2016) Intracellular FRET-based screen for redesigning the specificity of secreted proteases. *ACS Chem. Biol.* 11, 961–970. Copyright 2016 American Chemical Society.”

3.1. Introduction

Deficiencies or abnormalities in protease activity are common to many diseases, therefore proteases present attractive targets for use in therapy⁴. Due to their catalytic nature, protease therapeutics could potentially be used at lower doses reducing the cost of treatment. However, most proteases are thought to have as many as 10-100 physiological substrates and therefore their expression, localization and activity are tightly regulated to prevent unwanted proteolysis. Insufficient target substrate specificity has limited therapeutic uses of proteases since cleavage of non-target substrates can lead to side-effects. Among FDA approved therapeutic proteases, most recognize only a few physiological substrates thereby limiting their toxicity.

Prior efforts to redesign protease specificity have focused upon the use of structure-guided mutagenesis within the active site and surface loops^{60,61,186}. However, rational design of specificity has proven difficult since substitutions far from the active site can alter active site geometry and specificity¹⁴⁰. Furthermore, substitutions within the active site can disrupt the structure of neighboring residues that are important for binding and catalysis. The identification of rare protease variants with increased target specificity has typically required screening large variant libraries⁶⁴. Library screening has been successfully applied to a handful of model proteases including *E. coli* outer membrane protease OmpT, human immunodeficiency virus type 1 (HIV) protease, hepatitis A virus 3C protease, and tobacco etch virus protease (Chapter 1)^{57,65,70,71,75,77}. While these methods

have enabled engineering of membrane bound and intracellular proteases; methods for specificity redesign of human secreted proteases - candidates for therapeutic use - are needed.

We have developed a cell-based assay to enable design and engineering of human proteases by screening protease libraries for variants that cleave a peptide substrate exhibiting FRET (Chapter 2). This method was applied to human kallikrein 7 (hK7), a member of the trypsin-like serine protease superfamily, to identify variants that selectively cleave the central hydrophobic core (KLVFFAED) of the A β peptide involved in the pathology of Alzheimer's disease (AD)¹⁸⁷. Alzheimer's disease is characterized by increased accumulation of A β within the central nervous system due in part to impaired clearance mechanisms⁸⁷. hK7 cleaves A β *in vitro* within the central hydrophobic core KLVF↓F↓AED (A β 8) thereby preventing aggregation¹⁰³. Given that hK7 exhibits modest activity toward A β , but prefers tyrosine (Y) over phenylalanine (F) at P1, we sought to improve activity and narrow hK7 specificity toward A β . Our results demonstrate that screening randomly mutated hK7 libraries using PrECISE yielded a protease variant with improved selectivity for the A β hydrophobic core, reduced *in vitro* toxicity to mammalian cells, and increased resistance to inhibitors. Further, PrECISE provides a general method to evolve human proteases for specific degradation of proteins implicated in a disease state.

3.2. Results

3.2.1. Screening randomly mutated hK7 libraries for A β 8 specificity

To identify variants with increased A β 8 specificity, yeast cells harboring the A β 8 FRET substrate were transformed with a library of 3×10^7 hK7 variants (~2 base pair

substitutions per gene). Cells expressing hK7 variants mediating A β 8 cleavage after co-expression at 22°C for 6 hours were enriched in 6 sorting cycles from 0.8% of gated cells to 10% - a value representing wild-type like activity (Figure 3.1). Enrichment was due to sorting, since repeated regrowth and induction of the library did not yield detectable enrichment. Cells from the enriched library were plated to isolate individual clones for flow cytometry analysis.

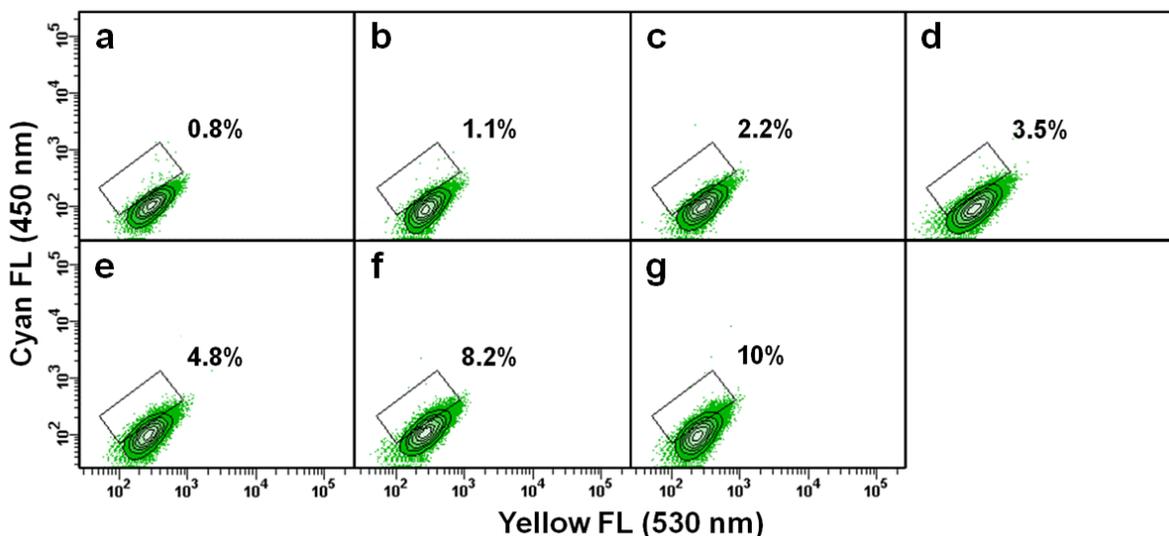


Figure 3.1. Sorting a first generation hK7 random mutagenesis library for A β 8 cleavage.

A first generation hK7 error-prone library was created using the gene encoding wild-type hK7 as a template and transformed into cells harboring the A β 8 FRET reporter. The hK7 library was co-expressed with the A β 8 substrate and cells were sorted as stated in Methods. FACS plots of the initial first generation library (a) and after the first (b), second (c), third (d), fourth (e), fifth (f), and sixth (g) rounds of enrichment. After the second round of enrichment, gates were adjusted to collect the top 1% of cells in the cyan fluorescence gate.

The A β 8 cleavage activity of isolated variants was ranked using the cell-based assay and ten unique protease variants were identified by DNA sequencing (Table 3.1). To validate the clonal ranking obtained using the cell-based assay, protease activity in cell

lysates was measured using freshly added A β 8 FRET reporter. To simplify initial evaluation of library variants, relative second order rate constants ($k_{cat} \cdot [E]/K_M$) were determined which do not take into account differences in expression. Confirming the effectiveness of the cell-based screening assay, *in vitro* measured A β 8 cleavage activities linearly correlated with FRET signals from the whole cell assay (Spearman correlation = 0.88) (Figure 3.2). Three variants displayed lower lysate activity than would be predicted from intracellular cyan fluorescence. But importantly, none of the clones had both low cyan fluorescence and high *in vitro* activity indicating that rare variants with improved activity are unlikely to be excluded during sorting.

Table 3.1. Substitutions identified for first generation hK7 variants.

hK7 variant	substitutions	frequency
1.1	R132M, K198I	1/30
1.2	N230I	1/30
1.3	L53P	4/30
1.4	N119D, K129R, F241S	1/30
1.5	D83G	1/30
1.6	D156G, N246S	2/30
1.7	K251E	4/30
1.8	N186T, K250E	2/30
1.9	R100C, R123M	3/30
1.10	K181R, K240Q	2/30

Nine of the sequenced clones were the wild-type hK7 gene

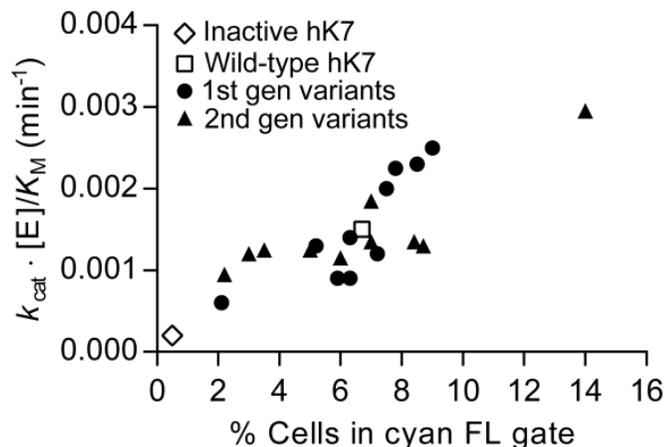


Figure 3.2. Intracellular cyan fluorescence correlates linearly to A β 8 activity in cell lysates.

To validate the cell-based assay, individual variants from the final round of sorting the first and second generation libraries were analyzed by flow cytometry for cyan fluorescence after co-expression with the A β 8 FRET substrate. Relative second order rate constants ($k_{cat} \cdot [E]/K_M$) toward A β 8 *in vitro* linearly correlated with the percentage of cells having cyan fluorescence above background. Spearman correlation coefficients were 0.88 (first generation variants) and 0.82 (second generation variants) giving a combined correlation of 0.77 (all variants).

Four variants obtained from screening the initial library exhibited increased activity toward A β 8 and KLVYYAED (Figure 3.3). However, none of these variants exhibited improved selectivity for A β 8 over KLVYYAED using the cell lysate assay. Variant hK7-1.10 showed a 1.7-fold improvement in activity toward A β 8 compared to wt-hK7. Variants with improved activity *in vitro* possessed substitutions that reduced positive surface charge. hK7 has been reported to have an isoelectric point of ~ 9 and contains two positively charged surface patches that have been proposed to bind to negatively charged substrates or aid in attachment to the cell surface^{114,188}. To determine whether variants with improved A β 8 lysate activity were produced at higher levels per cell compared to wild-type, differences in expression were evaluated using Western blot. Each of the variants

with reduced positive surface charge were expressed at higher levels on a per cell basis (Figure 3.4).

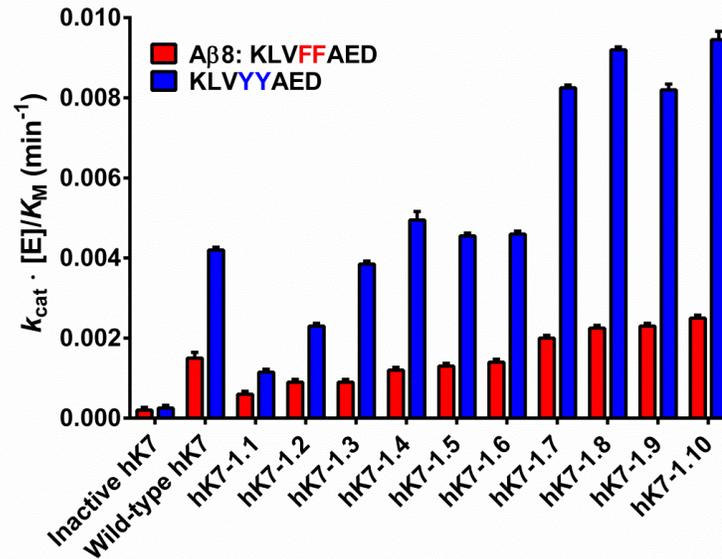


Figure 3.3. First generation hK7 variants display improved intracellular activity for Aβ8.

In vitro activity of unique variants acquired from sorting the first generation library was assessed toward the target Aβ8 and non-target KLVYYAED substrates to determine changes in P1 selectivity. Equal numbers of cells expressing each variant were lysed to determine a relative second order rate constant $k_{cat} \cdot [E]/K_M$ which does not take into account differences in protease expression levels per cell (N = 3). Inactive hK7 variant S205A was used as a control to determine background activity by yeast proteases toward these two substrates. Four variants hK7-1.7, hK7-1.8, hK7-1.9, and hK7-1.10 exhibited increased activity toward both Aβ8 and KLVYYAED. Variant hK7-1.10 showed a 1.7-fold improvement in lysate activity toward Aβ8 compared to wild-type hK7. Data represented as mean \pm SD.

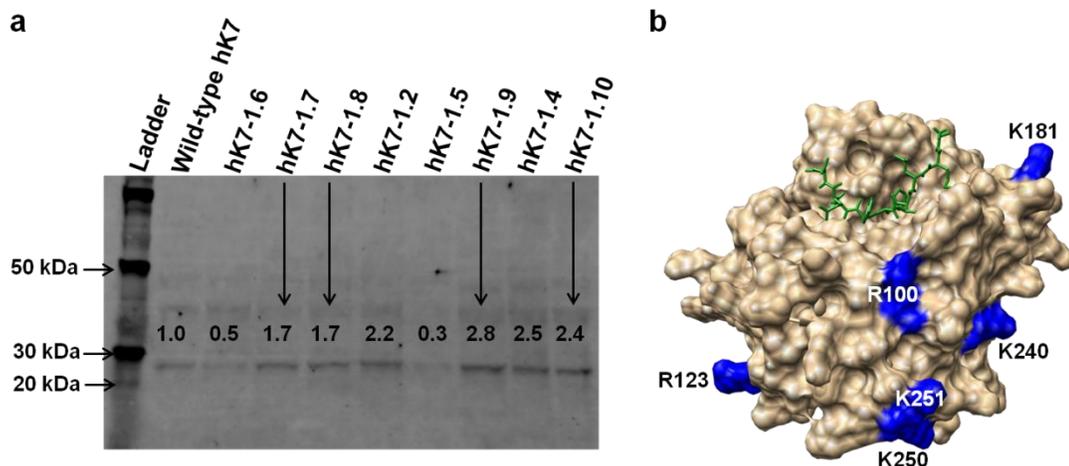


Figure 3.4. hK7 variants with improved *in vitro* activity show improved expression levels.

(a) Equal numbers of cells expressing wt-hK7 or the hK7 variants were lysed and hK7 expression levels were assessed using a Western blot with a poly-clonal anti-hK7 antibody. hK7 expected molecular weight is ~27 kDa. Arrows indicate increased expression of variants hK7-1.7, hK7-1.8, hK7-1.9, and hK7-1.10 per cell normalized to wt-hK7 expression. (b) Model of hK7 docked with Aβ8 (KLVFFAED) substrate (green) indicating the location of positively charged surface residues that are substituted in variants with improved expression.

Collectively, our results indicated that the population of variants enriched by sorting the first generation library was diverse and spanned a range of apparent activities influenced by intrinsic activity and expression level. Improvements in expression had the benefit of improving the assay dynamic range. Given this, we further diversified this pool using error prone PCR, to create a second generation library of 1×10^7 variants with ~5 base pair substitutions per gene. Library members with high cyan fluorescence were enriched using 6 cycles of FACS after co-expression at 22°C for 6 hours, from 0.8% to 15% gated cells in the final round (Figure 3.5). Individual isolated clones ($n = 10$) exhibiting high Aβ8 substrate conversion were then assayed for activity toward Aβ8 and KLVYYAED *in vitro* (Figure 3.6). Variant hK7-2.10 exhibited 2-fold increased activity toward Aβ8 compared to

wt-hK7 and possessed three substitutions of solvent exposed arginines (Table 3.2). Importantly, variant hK7-2.7 exhibited improved A β 8 selectivity, conferred by reduced activity toward KLVYYAED. This A β 8 selective variant possessed seven substitutions, including Q52R, N119S, K129G, V157G, L189Q, Q228R, and G233V (UniProt numbering used).

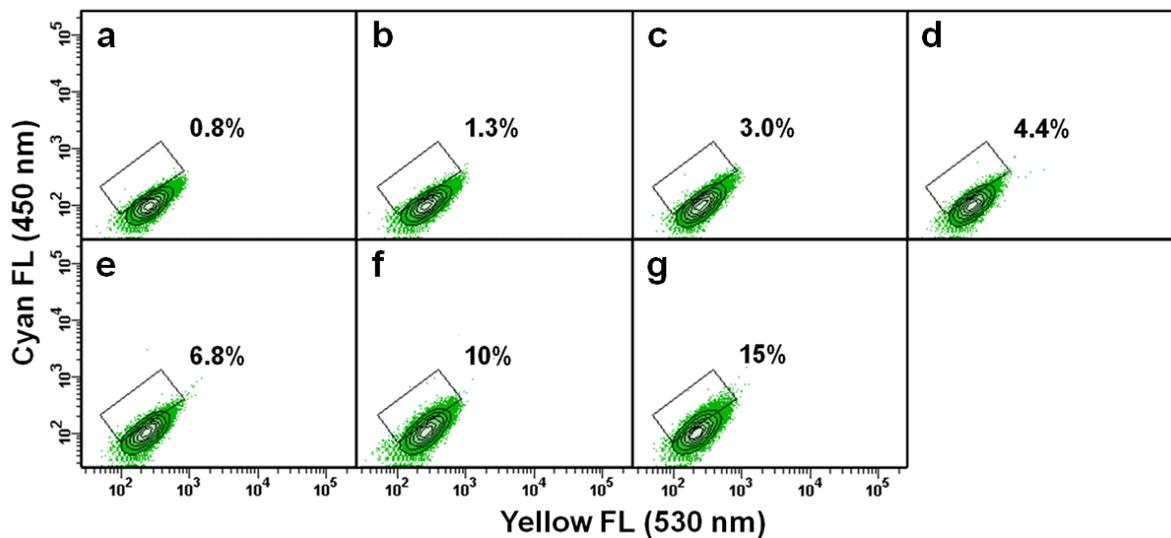


Figure 3.5. Sorting a second generation hK7 random mutagenesis library for A β 8 cleavage.

hK7 second generation error-prone library was created using the final pool of first generation variants as a template and transformed into cells harboring the A β 8 FRET reporter. The hK7 library was co-expressed with the A β 8 substrate and cells were sorted as stated in Methods. FACS plots of the initial second generation library (a) and after the first (b), second (c), third (d), fourth (e), fifth (f), and sixth (g) rounds of enrichment. After the second round of enrichment, gates were adjusted to collect the top 1% of cells in the cyan fluorescence gate.

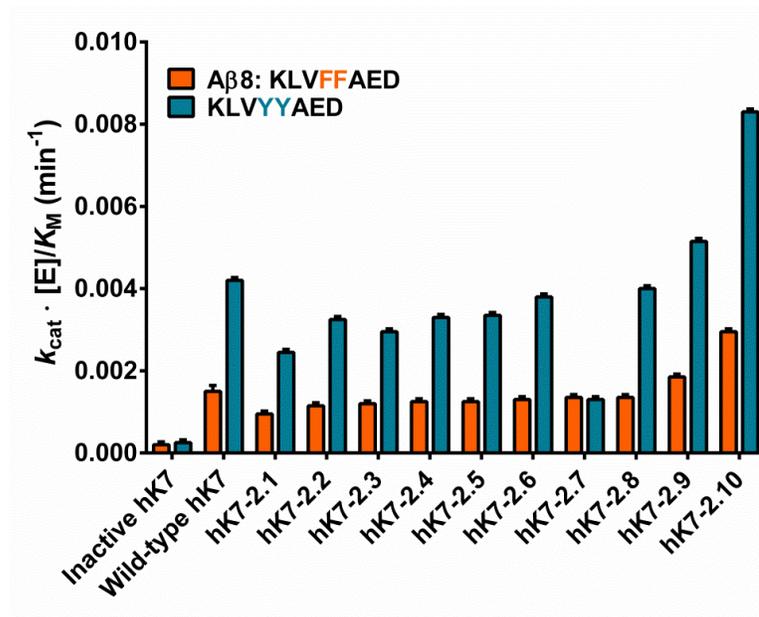


Figure 3.6. Second generation hK7 variant displays improved Aβ8 selectivity.

In vitro activity of unique variants acquired from sorting the second generation library was assessed toward the target Aβ8 and non-target KLVYYAED substrates to determine changes in P1 selectivity. Equal numbers of cells expressing each variant were lysed to determine a relative second order rate constant $k_{cat} \cdot [E]/K_M$ ($N = 3$). Variant hK7-2.10 displayed 2-fold increased activity on Aβ8 compared to wt-hK7 while variant hK7-2.7 displayed a substantial reduction in activity toward the non-target KLVYYAED and maintained activity on Aβ8 comparable to wild-type. Data represented as mean \pm SD.

Table 3.2. Substitutions identified for second generation hK7 variants.

hK7 variant	substitutions	Frequency
2.1	V166G, K181E, K251E	1/20
2.2	R100C, R123M, D162V	1/20
2.3	K198E, K250T	2/20
2.4	R63H, M73I, T106A, K117M, V157A, I171V	3/20
2.5	K169R, T223A, K250E	2/20
2.6	K130N, R136C, M164T	2/20
2.7	Q52R, N119S, K129G, V157G, L189Q, Q228R, G233V	2/20
2.8	N119D, K129R, F241S, K250E	3/20
2.9	Q174L, N186T, R212S, V238A, K250E	1/20
2.10	R100C, R123M, R212S	3/20

3.2.2. Characterization of variant hK7-2.7 selectivity switch

To probe the apparent selectivity improvement of hK7-2.7 toward A β 8, activities toward A β 8 and a panel of non-target substrates were measured after purification. Substrates were designed to probe specificity at multiple sub-sites. This panel consisted of reported hK7 substrates, including a matrix metalloproteinase-9 activation site (RHLY↓GPRP) and an activation sequence in prochemerin (GQFAF↓SKA)^{189,190}. Given hK7's role in skin desquamation¹¹⁶, candidate sites within cell adhesion proteins fibronectin (RYRVTYSS), corneodesmosin (SQVSYSSG), desmoglein-1 (RITKYSTV), desmocollin-1 (RVTIFTVP), and E-cadherin (RTAYFSLD, KVFYSITG) were included^{122,124,191}. hK7 cleaved A β 8 ($k_{cat}/K_M = 0.61 \pm 0.02 \mu\text{M}^{-1} \cdot \text{min}^{-1}$) and exhibited the highest activity toward fibronectin sequence RYRVTYSS ($k_{cat}/K_M = 92.5 \pm 3.5 \mu\text{M}^{-1} \cdot \text{min}^{-1}$)

(Figure 3.7a). The substrate RYRVTYSS is to our knowledge the fastest cleaved substrate reported for hK7 and contains two potential cleavage sites in agreement with hK7's reported preference for Y at the P1 position and R or S at the P1' position¹²³. Analysis of the rates of cleavage for panel substrates supported previous findings that hK7 prefers tyrosine (Y) at the P1 position over phenylalanine (F)^{113,123} and prefers arginine (R) at P4/P5¹⁰³ (Schechter and Berger nomenclature¹²).

In contrast, the A β 8 selective variant hK7-2.7 retained activity on A β 8 ($k_{\text{cat}}/K_{\text{M}} = 0.47 \pm 0.01 \mu\text{M}^{-1} \cdot \text{min}^{-1}$) and other phenylalanine containing substrates, while exhibiting decreased activity toward non-target substrates containing tyrosine (Figure 3.7b). For example, hK7-2.7 retained only 3% of wild-type hK7 activity on substrate RHLYGPRP, 4.5% on RYRVTYSS, 5% on SQVSYSSG, and 8.7% wild-type activity on RITKYSTV resulting in 10- to 30-fold selectivity increases (Figure 3.8). Furthermore, hK7-2.7 exhibited 4-fold improved selectivity toward A β 8 (KLVFFAED) over KLVYYAED, and selectivity toward PRVMFFT over PRVMYYT was increased by 4-fold. Modest selectivity enhancements were observed on substrates containing both phenylalanine and tyrosine (e.g. RTAYFSLD, KVFYSITG). hK7-2.7 displayed activity comparable to wt-hK7 on GQFAFSKA and RVTIFTVP, which both lack tyrosine. Selectivity was also improved outside of the P1 pocket since hK7-2.7 exhibited increased selectivity for KLVFFAED over the A β 8 variant KLVFFAAD. Thus, *in vitro* studies confirmed that library screening yielded an hK7 variant with improved selectivity for A β 8, mediated by exclusion of tyrosine containing substrates.

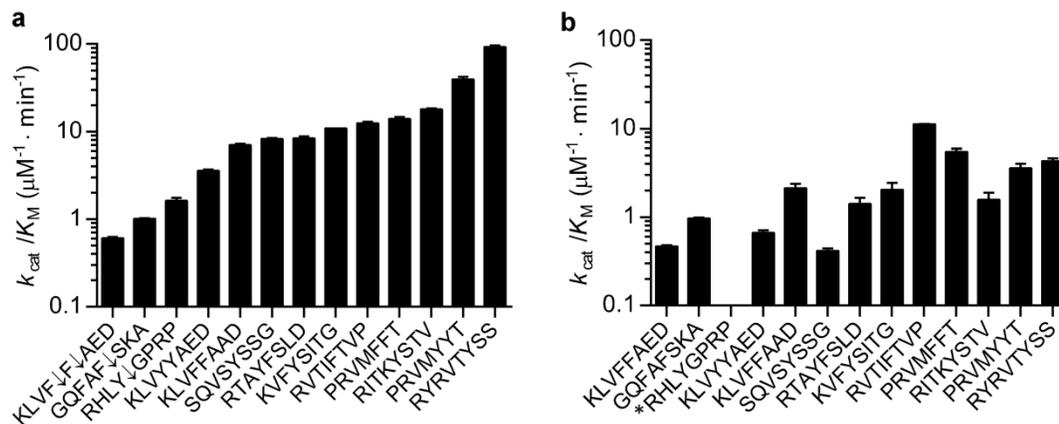


Figure 3.7. Analysis of wt-hK7 and variant hK7-2.7 activity on a panel of substrates.

(a) Wild-type hK7 was purified from yeast cell lysate and activity toward the A β 8 substrate and a panel of non-target substrates was measured *in vitro*. (b) Similarly, k_{cat}/K_M values were determined for variant hK7-2.7 on the same panel of substrates ($k_{cat}/K_M = 0.05 \pm 0.01 \mu\text{M}^{-1} \cdot \text{min}^{-1}$ for *RHLYGPRP).

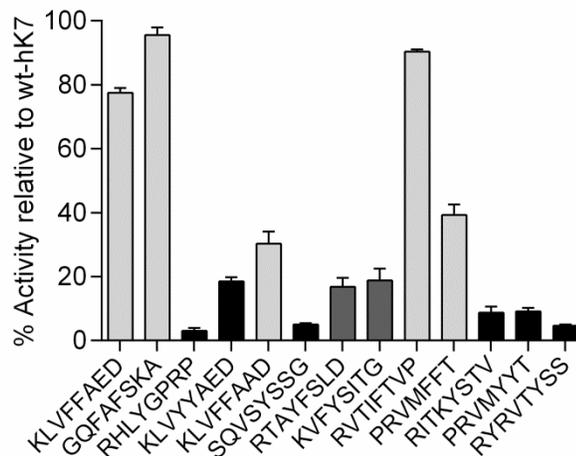


Figure 3.8. hK7-2.7 exhibits enhanced selectivity for A β 8 due to decreased activity on tyrosine.

Variant hK7-2.7 was more selective for A β 8 (KLVFFAED) due to a reduction in activity on substrates containing tyrosine (black bars) and substrates containing tyrosine and phenylalanine (dark gray bars). Importantly, hK7-2.7 retained wild-type activity on A β 8 and other substrates containing phenylalanine (light gray bars). Values are the average activity measurements of quadruplicate experiments each performed in duplicate wells and the error is their standard deviation.

To assess whether the improved selectivity of hK7-2.7 could reduce cytotoxicity due to decreased off-target substrate cleavage, toxicity toward differentiated PC12 neuronal-like cells was compared to that of wt-hK7. The A β 8 selective variant hK7-2.7 exhibited significantly reduced toxicity ($p = 0.03$) when compared to wt-hK7 (Figure 3.9). Furthermore, it was anticipated that improved selectivity would reduce yeast cytotoxicity during screening and improve growth. Therefore, growth rates of hK7-2.7 and wt-hK7 were determined for the expression conditions used during library screening. Cells expressing hK7-2.7 displayed a 30% increased growth rate during expression ($0.13 \pm .01 \text{ hr}^{-1}$) compared to cells expressing wt-hK7 ($0.10 \pm .01 \text{ hr}^{-1}$) (Figure 3.10).

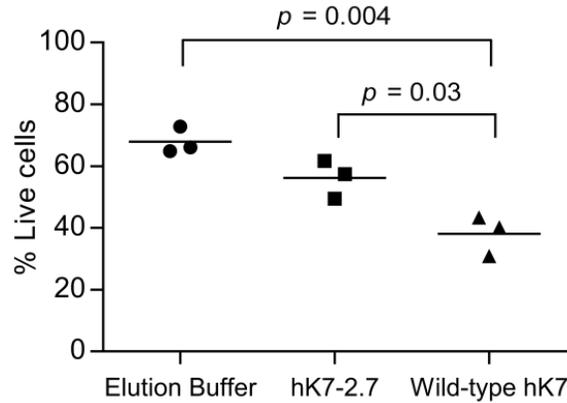


Figure 3.9. hK7-2.7 displays reduced toxicity to PC12 neuronal-like cells.

Cell viability was measured 24 hours after addition of purified wild-type hK7, hK7-2.7, or elution buffer to differentiated PC12 cells. Treatment of PC12 cells with wild-type hK7 (1 nM) reduced cell viability to 38% ($p = 0.004$). Cell viability increased by 47% when cells were incubated with hK7-2.7 ($p = 0.03$). The mean is the result of three experiments each performed in N=5 wells. p -values represent an unpaired t -test.

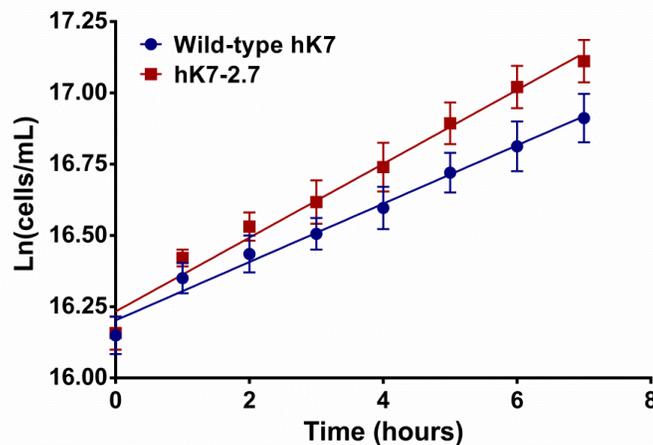


Figure 3.10. Variant hK7-2.7 confers improved yeast growth during expression.

Growth rates of cells expressing wild-type hK7 or variant hK7-2.7 were measured for the expression conditions used during library sorting ($T = 22^{\circ}\text{C}$). Variant hK7-2.7 displayed an improved growth rate of $0.13 \pm .01 \text{ hr}^{-1}$ during expression compared to the wild-type growth rate of $0.10 \pm .01 \text{ hr}^{-1}$ and the divergence of the slopes was clearly seen for the 6 hour expression time used for library sorting (N = 3). Data represented as mean \pm SD.

We next investigated whether selectivity enhancing substitutions within hK7-2.7 had altered its susceptibility to wild-type inhibitors. Inhibitor IC_{50} values for bovine pancreatic trypsin inhibitor (BPTI), soybean trypsin inhibitor (SBTI), α -antichymotrypsin, and lympho-epithelial Kazal-type-related inhibitor (LEKTI) fragment consisting of domains 10-15 were measured for wt-hK7 and hK7-2.7 in assays containing 5 nM protease and 200 nM substrate. Although BPTI and SBTI are not endogenous to humans, they have been reported to be effective *in vitro* inhibitors of wild-type hK7¹⁹². BPTI inhibited wt-hK7 with an IC_{50} value of 47 nM, but did not inhibit hK7-2.7 at the highest concentration tested (30 μ M), giving a >640-fold increased resistance to BPTI (Figure 3.11a). Similarly, SBTI inhibited wt-hK7 with an IC_{50} value of 17 nM but weakly inhibited hK7-2.7 at 50 μ M SBTI (Figure 3.11b). Human α -antichymotrypsin, a serpin inhibitor, potently inhibited wt-hK7 (IC_{50} = 2.5 nM), but was 16-fold less potent toward hK7-2.7 (IC_{50} = 40 nM) (Figure 3.11c). Additionally, we tested inhibition by the physiologically relevant inhibitor LEKTI which has been shown to be a tight binding inhibitor of hK7¹⁹³ and co-localizes with hK7 in the skin¹⁹⁴. LEKTI (domains 10-15) inhibited wt-hK7 with an IC_{50} of 55 nM and displayed no inhibition toward hK7-2.7 at the highest concentration tested (130 nM) (Figure 3.11d). The reduced sensitivity of hK7-2.7 to inhibitors further confirms changes in the active site binding selectivity.

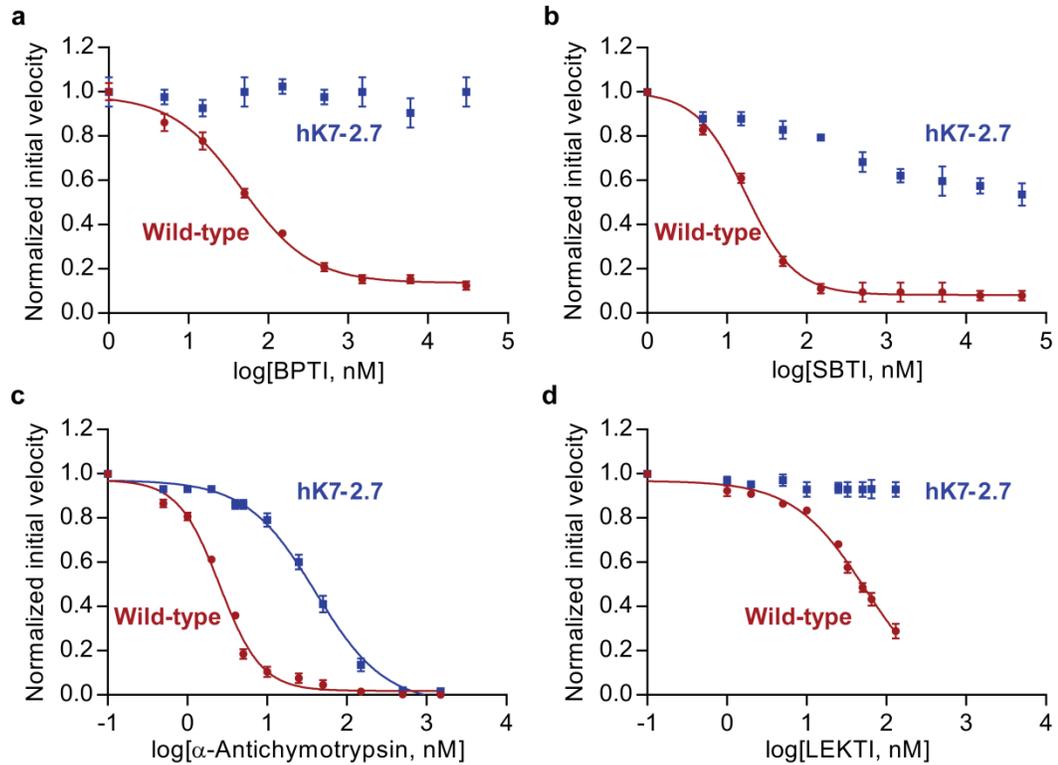


Figure 3.11. hK7-2.7 exhibits reduced sensitivity to wild-type inhibitors.

A β 8 cleavage activity was measured in assays using 5 nM protease and 200 nM substrate in the presence of a range of concentrations of four wild-type inhibitors. (a) hK7 tight-binding inhibitor bovine pancreatic trypsin inhibitor (BPTI) inhibited wild-type hK7 with an IC_{50} value of 47 nM but did not inhibit hK7-2.7. (b) Wild-type hK7 was inhibited by soybean trypsin inhibitor (SBTI) with an IC_{50} value of 17 nM, while hK7-2.7 was partially inhibited in the micromolar range. (c) α -antichymotrypsin potently inhibited hK7 with an IC_{50} of 2.5 nM but showed reduced inhibition on hK7-2.7 with an IC_{50} of 40 nM. (d) The physiologically relevant inhibitor LEKTI (domains 10-15) inhibited wt-hK7 with an IC_{50} of 55 nM while hK7-2.7 activity was not affected. Values are the average activity measurements of three replicate experiments each performed in duplicate wells and the error is their standard deviation.

To investigate the mechanism by which substitutions in hK7-2.7 improved selectivity, a model of hK7 docked with A β 8 was analyzed¹⁰³. Strikingly, hK7-2.7 substitution G233V was located deep within the S1 pocket and contacted the P1 phenylalanine in A β 8 (Figure 3.12a). Thus, G233V partially fills the S1 pocket to exclude the hydroxyl group of tyrosine while retaining favorable hydrophobic interactions with phenylalanine (Figure 3.12b). Indeed, substitution G233V was responsible for roughly 50% of the selectivity increase, with 2-fold reduced activity on KLVYYAED (Figure 3.13a). Nevertheless, G233V exhibited only partial selectivity improvement for A β 8 over the panel of non-targets (Figure 3.13b). Thus, substitutions aside from G233V contribute to selectivity improvements. For example, substitution Q52R could enhance selectivity by forming salt bridges with the substrate prime side residues glutamic acid and aspartic acid.

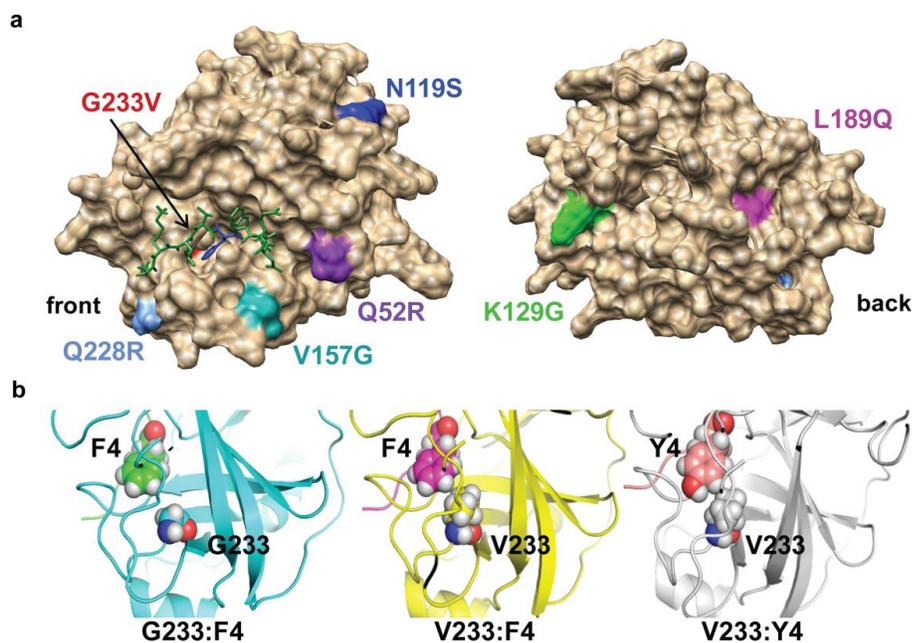


Figure 3.12. Improved A β 8 selectivity is mediated by a S1 pocket substitution that excludes tyrosine (Y).

(a) Seven substitutions in hK7-2.7 shown on the hK7 structure docked with A β 8 substrate KLVFFAED. Substitution G233V (red) occurs within the S1 pocket of hK7 which contacts the P1 phenylalanine (F) in A β 8 (blue residue). Additionally, Q52R (purple) can interact with negatively charged prime side residues within A β 8. (b) The single substitution G233V conferred a 2-fold switch in selectivity for phenylalanine (F4) over tyrosine (Y4). Valine may increase steric hindrance with the hydroxyl group of tyrosine. The images shown were created using PDB:2QXI with the UCSF Chimera Software package.

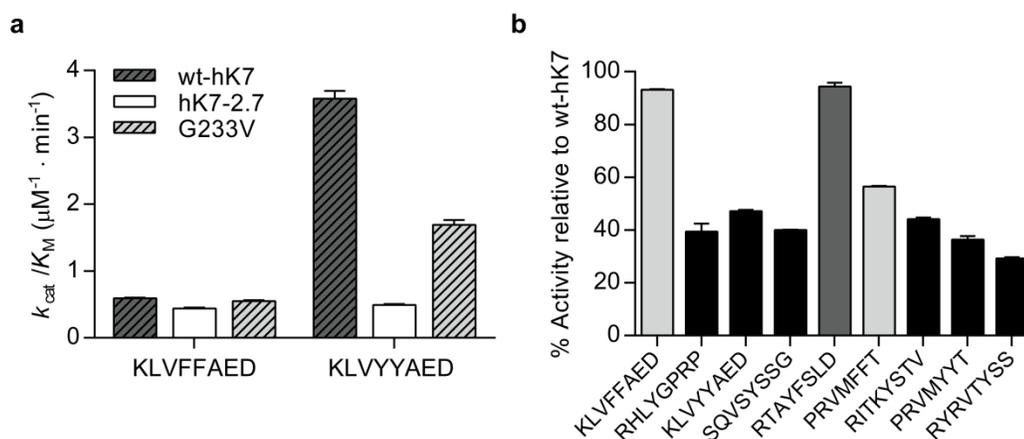


Figure 3.13. Substitution G233V is partially responsible for A β 8 selectivity improvement in hK7-2.7.

(a) The single substitution variant G233V displayed reduced activity on the non-target KLVYYAED but showed only a 2-fold selectivity improvement for A β 8 over KLVYYAED. (b) Activity of variant G233V was assessed with a larger panel of non-target wild-type preferred substrates. Though the single substitution variant G233V exhibited reductions in activity toward a majority of the non-targets, it is not solely responsible for the switch in selectivity of variant hK7-2.7. Values are the average activity measurements of quadruplicate experiments each performed in duplicate wells and the error is their standard deviation.

3.3. Discussion

Wider application of proteases as therapeutics for human disease is hindered by the lack of effective methods to optimize their activity, selectivity, stability, and expression. Although several high-throughput methods have been developed, their use has been limited to model proteases unlikely to find use in therapy^{57,65,70,71,75,77}. Furthermore, the majority of these methods are limited to proteases that can be expressed in the reducing environment of the cytosol or on the outer membrane of *E. coli*. Given these limitations, we developed a general strategy to redesign the specificity of the largest class of human proteases toward therapeutic targets. As a target protease, we selected a member of the

largest family (kallikreins) in the largest superfamily (trypsin-like serine proteases) of human proteases.

Discovery of the A β selective variant hK7-2.7 would have been unlikely without PrECISE to screen large protease libraries, since multiple substitutions contributed to increased selectivity. PrECISE relies on co-expression of the protease and substrate within the ER of yeast which provides an oxidizing environment for the formation of multiple disulfide bonds common to human secreted proteases. Using this method, randomly mutated libraries or rationally designed libraries for proteases with solved structures can be screened for substrate cleavage without exogenous reagents or labeling steps. Additionally, the FACS-based approach provides a quantitative assessment of library enrichment after each round of sorting and the ability to optimize screening conditions to favor desired variants. Here we used sequential rounds of low error rate random mutagenesis to incrementally introduce and screen for beneficial substitutions in hK7 leading to improved A β 8 activity and selectivity due to reduced host cell toxicity. After only two rounds of library generation and screening, variant hK7-2.7 displayed up to 30-fold improvements in A β 8 selectivity over substrates containing tyrosine and exhibited reduced toxicity toward yeast cells and PC12 neuronal-like cells compared to wt-hK7. Among the seven substitutions in hK7-2.7, G233V directly contacted the P1 phenylalanine of A β but was only partially responsible for the selectivity switch of hK7-2.7.

Endogenous protease inhibitors limit the serum half-lives of therapeutic proteases, therefore bypassing inhibition is of significant importance when translating the *in vitro* properties of proteases to therapeutic settings^{195,196}. Prior efforts to reduce protease inhibition have relied on structural analysis to identify residues recognized by inhibitors

but that are not responsible for substrate recognition, a strategy which has proven challenging^{53,197}. Current high-throughput methods for evolving proteases with reduced sensitivity to inhibitors have generated variants with up to 30-fold resistance to inhibition⁷⁷. Here, the redesigned active site of hK7-2.7 was significantly less sensitive to inhibitors. hK7-2.7 displayed a 16-fold reduction in inhibition by the serpin α -antichymotrypsin and greater than a 2-fold reduction in inhibition by LEKTI (domains 10-15). Additionally, hK7-2.7 exhibited greater than 600-fold reduced susceptibility to the known tight binding, but non-endogenous inhibitors SBTI and BPTI. Though we did not directly screen for inhibitor resistance, this trait was acquired along with the increase in substrate specificity.

Although improvements in selectivity can be achieved by counter-selecting for a non-target substrate while simultaneously selecting for target cleavage^{70,71,75}, it can be challenging to identify the most relevant non-target substrate(s). Here, we circumvented this problem by screening for retention of target activity, while reducing host toxicity. Even so, adaptation of PrECISE to enable simultaneous consideration of multiple non-target substrates would likely increase the range of applications. Currently, proteases approved for therapy have been engineered primarily for serum half-life and stability using rational design⁵⁴. The ability to engineer human proteases with narrowed substrate specificities could greatly expand the use of proteases for therapy by reducing toxic side-effects while maintaining efficacy.

3.4. Materials and methods

3.4.1. *hK7* random mutagenesis library construction

The wild-type *hK7* gene was randomized using error-prone PCR to create a library with an average of 2 mutations per gene. For low error-rate mutagenesis, 0.5 mM MnCl₂ was added to the PCR mixture with Taq polymerase, 1.5 mM MgCl₂ and unbalanced dNTPs (dATP = dGTP = 0.2 mM and dCTP = dTTP = 1mM) as described¹⁴⁴. Error-prone PCR was performed using an Eppendorf Mastercycler thermocycler as follows: 1 cycle, 3 min at 95°C; 30 cycles, 30 sec at 95°C, 30 sec at 58°C, 5 min at 68°C; 1 cycle, 8 min at 68°C. The amplified PCR product was concentrated using a Thermo GeneJet PCR purification kit (Thermo Scientific). The insert was then digested with HindIII and PmeI and gel purified using a Zymo gel extraction kit (Zymo Research). The purified DNA (5.5 µg) was ligated with 10 µg of HindIII/PmeI digested pYC2/NT (Invitrogen). Ligation products were concentrated using a Zymo Clean and Concentrator kit (Zymo Research) and desalted on 0.025 µm VSWP membrane filters (Millipore) using ddH₂O for 2 hours. DH5α *E. coli* were transformed by electroporation with library plasmid DNA. Transformed aliquots were each incubated in SOC media for 1 hour at 37°C with shaking. The transformations were pooled and serial dilutions were plated onto LB plates supplemented with 100 µg/mL ampicillin to determine the library size. The transformed cells were added to 500 mL LB supplemented with ampicillin (100 µg/mL) and 0.2% glucose and cultured for 9 hours. Library plasmid DNA was extracted from the transformed cells using a Zyppy Plasmid Maxiprep kit (Zymo Research). Twenty randomly picked colonies were sequenced to estimate the error rate (Genewiz).

Yeast cells containing the integrated A β 8 FRET reporter gene were transformed with library plasmid DNA using the high efficiency PEG/lithium acetate method¹⁸⁵ and plated on SD media lacking tryptophan and uracil to determine library size. The transformed cells were added to 500 mL SD-Trp-Ura media supplemented with 2% glucose and penicillin-streptomycin (pen-strep) at a final concentration of 100 units/mL and 100 μ g/mL (Invitrogen) and allowed to recover for 24 hours. To create the second generation hK7 library, first generation variant plasmid from the sixth round of sorting was recovered from yeast using a Zymoprep yeast plasmid miniprep II kit (Zymo Research) and electrocompetent DH5 α cells were transformed with the recovered plasmid. The transformed sample was recovered from bacteria and used as a template for a second round of random mutagenesis using error-prone PCR as described above. Yeast cells containing the integrated A β 8 FRET reporter gene were transformed with the second generation library for screening.

The single substitution variant G233V was created using PCR-driven overlap extension with forward primer 5' CAACCCAATGACCCAGTAGTCTACACTCAAGTG and reverse primer 5' CACTTGAGTGTAGACTACTGGGTCATTGGGTTG¹⁵¹. This insert was cloned into the pYC2 vector as previously described (Chapter 2).

3.4.2. Library screening and FACS analysis

Cells transfected with the protease-encoding plasmid library and A β 8 FRET reporter were grown overnight in 50 mL SD-Trp-Ura media with 2% raffinose and penicillin-streptomycin (pen-strep) at a final concentration of 100 units/mL and 100 μ g/mL (Invitrogen). Overnight cells were subcultured to a final OD₆₀₀ of 0.75 and approximately 5×10^8 cells were induced with 2% galactose in fresh SD-Trp-Ura media containing 2%

raffinose. After induction, cells were grown at 22°C for 6 hours with shaking. A total of 5×10^7 cells, around 2-fold larger than library sizes, were spun down and resuspended in 1x PBS buffer for screening using a FACSAria flow cytometer (BD Biosciences) to isolate clones exhibiting an increase in cyan fluorescence. Cells not producing a full-length FRET reporter were excluded from the FRET analysis by pregating cells with high yellow emission (530/30 nm) with blue excitation (488 nm). The cyan (450/40 nm) and yellow (530/30 nm) fluorescence intensities with violet excitation (407 nm) were collected. Cells displaying a high cyan to yellow fluorescence ratio were sorted into SD-Trp-Ura media with pen-strep and 2% glucose to repress further expression. After 24 hours, cells were subcultured into fresh SD-Trp-Ura media containing 2% raffinose and pen-strep and allowed to grow for 16 hours before the next round of sorting. After each round of sorting, the population of cells was plated on selective SD-Trp-Ura media. Sixty individual clones from the sixth round of sorting were analyzed on flow cytometry after 6 hours co-expression to compare intracellular A β 8 FRET reporter conversion to wild-type hK7. The top thirty clones with the highest intracellular A β 8 activity were amplified using yeast colony PCR and sent for sequencing (Genewiz) to determine which mutations had occurred.

3.4.3. Construction of non-target FRET reporter substrates

To determine changes in specificity for the hK7 variants isolated from the cell-based screen, a panel of non-target FRET reporters was constructed. The non-target panel included a cleavage site within matrix metalloproteinase-9 that leads to activation (RHLY↓GPRP) and a C-terminal sequence that activates prochemerin (GQFAF↓SKA)

which are confirmed *in vitro* substrates of wild-type hK7, as well as potential cleavage sites within confirmed *in vitro* hK7 substrates such as fibronectin (RYRVTYSS), corneodesmosin (SQVSYSSG), desmoglein-1 (RITKYSTV), desmocollin-1 (RVTIFTVP), and E-cadherin (RTAYFSLD, KVFYSITG)^{122,124,189-191}. To determine the potential cleavage sites, the substrate specificity profile previously described¹⁰³ was aligned with the full protein sequence of confirmed *in vitro* substrates and 8-amino acid regions with high similarity were cloned into FRET reporters to assess wild-type hK7 activity. Non-target FRET reporters were constructed, expressed in *E. coli*, and purified as previously described¹⁹⁸. Confirmed *in vitro* substrates PRVMYYT, PRVMFFT, KLVYYAED, KLVFFAAD, and KLVFFAED were previously constructed as FRET reporters¹⁰³.

3.4.4. Characterizing activity and specificity of hK7 variants

Activity of isolated variants toward the target substrate A β 8 and non-target substrate KLVYYAED was initially measured in yeast lysate. Cultures of variants and wild-type hK7 were grown overnight and subcultured to an OD₆₀₀ of 1 in 30 mL SD-Trp-Ura media with 2% raffinose and 2% galactose and expressed for 6 hours at 22°C. Equal numbers of cells expressing each variant were harvested by centrifugation at 5,000g for 10 minutes and cell pellets were lysed with 500 μ L Y-PER (Thermo Scientific). Protease activity was measured in lysates using the FRET reporter assay described below to determine a relative second order rate constant $k_{\text{cat}} \cdot [\text{E}]/K_{\text{M}}$. For large scale expression and purification of hK7 variants and wild-type hK7, 150 mL starter cultures were grown overnight in SD-Trp-Ura media with 2% raffinose to an OD₆₀₀ of 3. Cells were then subcultured to an OD₆₀₀ of 1 in 360 mL fresh SD-Trp-Ura media with 2% raffinose and 2% galactose and expressed for 6

hours at 22°C. Cells were harvested by centrifugation at 5,000g for 10 minutes and froze overnight. Cell pellets were lysed with 6 mL of Y-PER for 30 minutes at room temperature and then centrifuged for 10 minutes at 5,000g to clarify the lysate. Lysate was incubated with 1 mL Ni-NTA resin (Qiagen) for 1 hour at 4°C. Supernatant was flowed through a gravity flow column and resin was washed twice with 6 mL wash buffer (50 mM NaH₂PO₄, 300 mM NaCl, 20 mM imidazole, pH = 8.0). Histidine-tagged hK7 protease was eluted into 1 mL of elution buffer (50 mM NaH₂PO₄, 300mM NaCl, 250 mM imidazole, pH = 8.0) and determined to be >90% pure as determined by SDS-PAGE with Western blot staining using a polyclonal anti-hK7 antibody (R&D Systems) (Figure 3.14). The concentration of active wild-type hK7 was determined by active site titration with the hK7 tight inhibitor α -antichymotrypsin (Sigma-Aldrich). Due to the reduced sensitivity of variant hK7-2.7 to wild-type inhibitors, no suitable inhibitor for active site titration was found. The fraction of active variant hK7-2.7 after purification was assumed to be the same as wild-type.

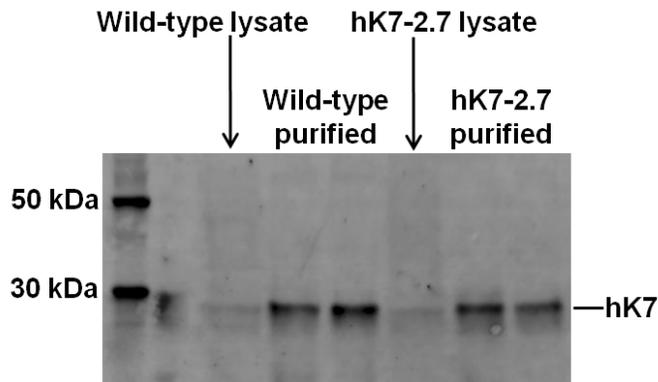


Figure 3.14. Western blot analysis of hK7 concentration before and after purification.

Wild-type hK7 and variant hK7-2.7 were purified from yeast cell lysate using affinity chromatography with nickel beads. Cell lysate samples and purified samples were run on a Western blot with an anti-hK7 polyclonal antibody. The expected molecular weight of hK7 is 27 kDa. The concentration of protease dramatically increased after one purification step from the lysate to the purified fraction. The concentration of active protease within the purified fraction was determined by active site titration using a tight-binding hK7 inhibitor.

Purified protease (5 nM) was incubated with FRET reporter substrate (200 nM) and conversion of the substrate was monitored by measuring the fluorescence emissions at 475 nm and 527 nm every 5 min upon excitation at 433 nm for 4 hours at 37°C using a Tecan Infinite 200 PRO spectrophotometer (Tecan). The *in vitro* conversion of the FRET reporter was calculated by measuring the change in the FRET ratio (yellow fluorescence/ cyan fluorescence) over time and is given by the equation:

$$Conversion = \frac{FR_{max} - FR_{time}}{FR_{max} - FR_{min}}$$

where FR_{max} is the maximum FRET ratio (527/475 nm), FR_{time} is the ratio at a given time within the assay, and FR_{min} is the minimum FRET ratio achievable with 100% cleavage. Second order rate constants (k_{cat}/K_M) were determined for hK7 variants and wild-type hK7

toward the A β 8 FRET reporter and non-target FRET substrates using the simplified Michaelis-Menten equation²⁰:

$$Conversion = 1 - \exp\left(\frac{-k_{cat}}{K_M} [E]t\right)$$

where [E] is the protease concentration and t is time. Reported values are the average k_{cat}/K_M and standard deviation of quadruplicate experiments each performed in duplicate wells.

Four different wild-type hK7 inhibitors were tested for their ability to inhibit purified wild-type hK7 and purified variant hK7-2.7. The above FRET assay was used to determine the rate of A β 8 FRET reporter cleavage (200 nM) with a protease concentration of 5 nM and inhibitor concentrations in the range of 0.5 nM to 50 μ M. The hK7 tight-inhibitors bovine pancreatic trypsin inhibitor (Sigma-Aldrich), soybean trypsin inhibitor (Sigma-Aldrich), α -antichymotrypsin (Sigma-Aldrich) and, LEKTI fragment domains 10-15 (R&D Systems) were dissolved to 300 μ M, 500 μ M, 15 μ M, and 1.3 μ M stock solutions in water, respectively. Inhibitors were diluted to the appropriate concentration and were incubated with purified protease samples for 10 minutes at 37°C prior to the addition of the A β 8 FRET reporter. IC₅₀ values were determined for the assay conditions above and generated by plotting the activity as a fraction of the uninhibited control against log₁₀[inhibitor] and GraphPad Prism (GraphPad Software, San Diego, CA) was used to fit the data.

3.4.5. PC12 toxicity assay

Rat pheochromocytoma PC12 cells were plated at high density on poly-L-lysine coated white, clear bottom 96 well culture dishes (1 x 10⁴ cells/well) in Dulbecco modified Eagle medium (DMEM) with 44 mM NaHCO₃, 5% horse serum, 5% calf serum, and 1x

antibiotic-antimycotic (Life Technologies) and grown at 37°C in a CO₂ incubator. 16 hours after plating, media was replaced with fresh media supplemented with dibutyryl-cAMP (dbcAMP-Sigma) at a final concentration of 0.5 mM to initiate differentiation. 24 hours post dbcAMP addition, media was replaced with fresh media supplemented with nerve growth factor (NGF-Sigma) at a final concentration of 20 ng/mL. Media changes supplemented with NGF were performed every 2 days for a total of 6 days. To assess the ability of variant hK7-2.7 to reduce protease-mediated toxicity to cells compared to wild-type hK7, variant hK7-2.7 was diluted 1/4 into media (50 µL) in the presence of PC12 cells. Similarly wild-type hK7 (5 nM) was diluted 1/5 into media (50 µL) so that the rate of Aβ₈ cleavage was equal for the two samples. Cells were incubated with protease for a total of 24 hours at 37°C in a CO₂ incubator. As a negative control, cells were also incubated with elution buffer diluted 1/4 into media. PC12 viability was measured using the CellTiter Glo assay (Promega) which quantifies the ATP content of live cells. Luminescent values were normalized between untreated control cells (100% viable) and background luminescence measured in media without the presence of cells (0% viable). The mean is the result of three experiments each performed in N=5 wells. Statistical analysis was performed using an unpaired *t*-test using GraphPad Prism.

4. Determinants of S1 pocket specificity within hK7

The PrECISE methodology was developed to enable engineering the activity and specificity of human secreted proteases toward peptide targets relevant in disease. This screening system allows detection of intracellular protease activity by flow cytometry through use of a co-expressed peptide substrate exhibiting FRET (Chapter 2). We previously described the application of this cell-based assay for FACS screening of large randomly mutated libraries of the secreted protease hK7 for improved activity and selectivity toward the A β 8 peptide (Chapter 3). Here we applied PrECISE to screen a focused site-saturation library of the hK7 S1 pocket (N199, A200, G222, G227, G233) for activity on the substrate PRVMFFT to gain a deeper understanding of the molecular mechanisms by which proteases achieve a high level of substrate selectivity. Monomeric variants of CyPet and YPet were constructed to eliminate the weak heterodimer interaction and improve enrichment of active clones during sorting. hK7 favors cleavage of tyrosine (Y) over phenylalanine (F) at P1 and we hypothesized that substitutions within the S1 pocket could lead to a variant with greater selectivity for F over Y. However, screening the S1 pocket focused library of hK7 revealed that substitutions within this subsite were not well tolerated and led to reduced catalytic efficiency of the protease. Sequencing analysis of isolated clones from the post sort 5 population suggested that the original wild-type residues at the five mutated positions within the S1 subsite were optimal for hK7 activity on PRVMFFT.

4.1. Introduction

The development of high-throughput methods for redesigning protease specificity is not only useful for therapeutic applications but also in elucidating the molecular mechanism proteases employ to recognize their substrate^{199,200}. If the crystal structure of a protease has been solved, focused libraries of residues within the active site that make side chain contacts with the substrate can be targeted for mutagenesis^{201,202}. This approach often leads to the discovery of epistatic combinations of substitutions within the active site that are required for dramatic changes in activity or selectivity^{69,71}. The identification of multiple synergistic substitutions within a protease active site using random mutagenesis would be rare given the low probability of introduced mutations to be grouped to a particular location on the protease²⁰³. Because mutations to the active site can often lead to protease unfolding and inactivity, amino acids targeted for mutagenesis can be partially saturated with specific sets of residues that may reduce the fraction of non-functional variants in the library²⁰⁴. Site-saturation library screening has been successfully applied in numerous cases to generate large switches in protease selectivity for non-optimal substrates^{70,71,75}.

Analysis of the hK7 crystal structure has elucidated potential mechanisms for its strong preference for cleavage after bulky aromatic amino acids at the P1 position. The S1 subsite of hK7 is large and hydrophobic which promotes favorable interactions with the hydrophobic side chains of P1 residues such as tyrosine (Y), phenylalanine (F) and leucine (L)¹¹⁴. Further, the presence of asparagine N199 at the bottom of the S1 pocket can create a hydrogen bond with the polar hydroxyl group of tyrosine leading to the high favorability of hK7 cleavage after tyrosine compared to all other amino acids^{103,113,123}. We hypothesized that partial saturation mutagenesis of residues within the S1 pocket could generate a

structure with a higher preference for F over Y at P1. Here, we have generated a focused library of the hK7 S1 subsite (N199, A200, G222, G227, G233) which was screened using the PrECISE cell-based assay for improved activity toward the PRVMFFT substrate.

4.2. Results

4.2.1. Improving the dynamic range of the PrECISE method

CFP and YFP are weak heterodimers with an estimated K_D of $100 \mu\text{M}^{205}$. Within the PrECISE method, the high concentration of FRET reporter in the yeast ER could lead to a false negative signal after proteolysis of the substrate due to non-specific interactions between CyPet and YPet. Therefore we tested if reducing the interaction between CyPet and YPet could improve the dynamic range of the assay allowing enhanced isolation of cells with active proteases toward the target substrate. Previous work discovered that hydrophobic residues at the crystallographic interface of CFP and YFP were responsible for the dimerization of these two proteins²⁰⁵. Substitution of the hydrophobic residue A206 with the positively charged lysine (K) in both CFP and YFP led to monomeric proteins incapable of dimerizing while maintaining sufficient FRET^{141,205}. Therefore this substitution was introduced into the CyPet and YPet of the PRVMYYT FRET reporter and the construct was integrated into the yeast genome. Transformants were analyzed by flow cytometry after expression at 22°C for 6 hours to select a clone expressing high levels of the substrate. Interestingly, addition of the A206K substitutions to the FRET reporter slightly reduced yellow fluorescence due to FRET of the intact reporter likely due to the decrease in CyPet and YPet interaction (Figure 4.1). However, CyPet and YPet brightness were unchanged.

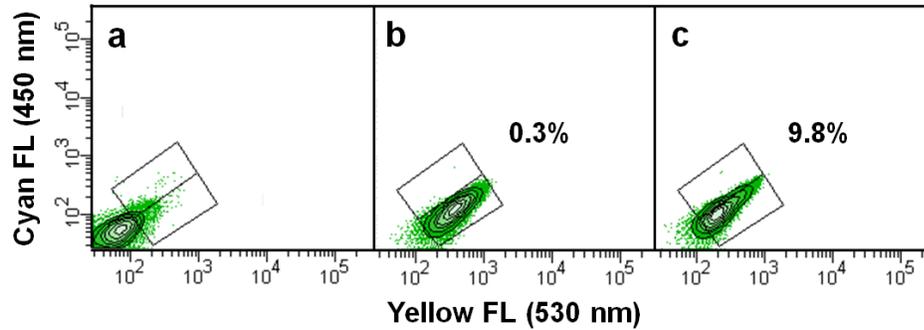


Figure 4.1. A206K substitutions to CyPet and YPet decrease FRET efficiency.

(a) Flow cytometry analysis of cells carrying the PRVMYYT FRET reporter gene that have not been induced with galactose. (b) Fluorescent profile of cells expressing the PRVMYYT FRET reporter for 6 hours at 22°C. (c) Fluorescent profile of cells expressing the engineered PRVMYYT FRET reporter, which contains the substitution A206K to CyPet and YPet, for 6 hours at 22°C. The engineered FRET reporter displayed a decrease in yellow fluorescence due to FRET causing a slight shift in the population to higher cyan fluorescence.

To determine if the variant FRET reporter could improve the dynamic range of the PrECISE screen, cells carrying the engineered PRVMYYT FRET reporter were co-expressed with active or inactive hK7 and compared to co-expressions in the original FRET strain. Co-transformants were analyzed by flow cytometry after co-expression at 22°C for 6 hours. Cells co-expressing the inactive hK7 protease with the engineered FRET reporter displayed reduced yellow fluorescence due to FRET compared to inactive hK7 co-expression in the original strain (Figure 4.2). Co-expression of the engineered PRVMYYT FRET with the wild-type protease led to a higher increase in cyan fluorescence/yellow fluorescence for three unique co-transformants compared to the change in fluorescence with the original FRET reporter (Figure 4.2). This modest improvement in the dynamic range of the assay could enhance the enrichment of cells with active protease and reduce the chance of false negatives occurring due to interactions between free CyPet and YPet in

the ER of yeast. A PRVMFFT FRET reporter with the A206K substitutions was constructed to screen a site-saturation mutagenesis library of the S1 pocket of hK7.

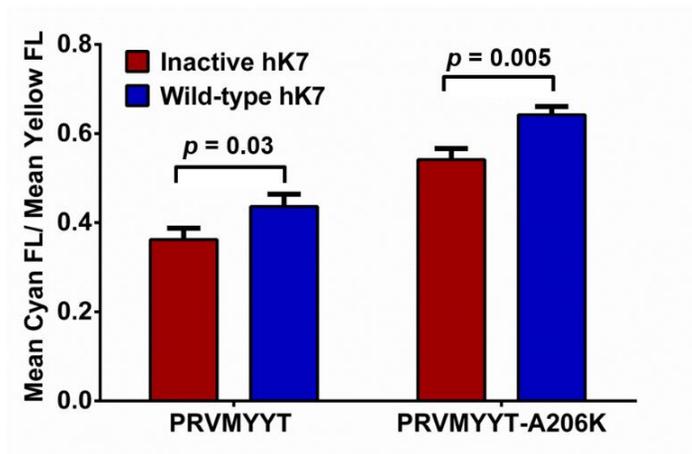


Figure 4.2. PRVMYYT-A206K FRET reporter displays improved dynamic range.

The engineered PRVMYYT FRET reporter, which contains the addition of the A206K substitutions to CyPet and YPet, showed a larger shift in the cyan/yellow fluorescence ratio between inactive hK7 and active hK7 co-expression compared to co-expressions in the original PRVMYYT FRET strain (N = 3). Data represented as mean \pm SD. *p*-values represent an unpaired *t*-test.

4.2.2. Screening a rationally designed S1 pocket library of hK7

Although past studies have determined that hK7 prefers tyrosine (Y) over phenylalanine (F) at the P1 position^{103,113}, little information is known concerning which residues can be cleaved at P1 besides Y and F. To elucidate which amino acids can be tolerated at the P1 and P1' positions by hK7, eight FRET reporters were constructed, expressed and purified from bacteria containing the substrates: PRVMYYT, PRVMFFT, PRVMLLT, PRVM~~MM~~MT, PRVM~~WW~~T, PRVM~~HH~~T, PRVM~~VV~~T, and PRVM~~PP~~T. Cleavage of the 8 substrates was analyzed by fluorimetry in assays using 10 nM wild-type hK7 and 200 nM substrate. hK7 displayed a catalytic efficiency (k_{cat}/K_M) that was 2-fold higher for PRVMYYT compared to PRVMFFT and 10-fold higher for PRVMYYT over PRVMLLT

and PRVMMMT (Figure 4.3). Interestingly, hK7 was also able to accommodate the large hydrophobic tryptophan (W) residue in the S1 pocket although activity was reduced 35-fold compared to PRVMYYT. No activity was detected on PRVMHHT, PRVMVVT, and PRVMPPT.

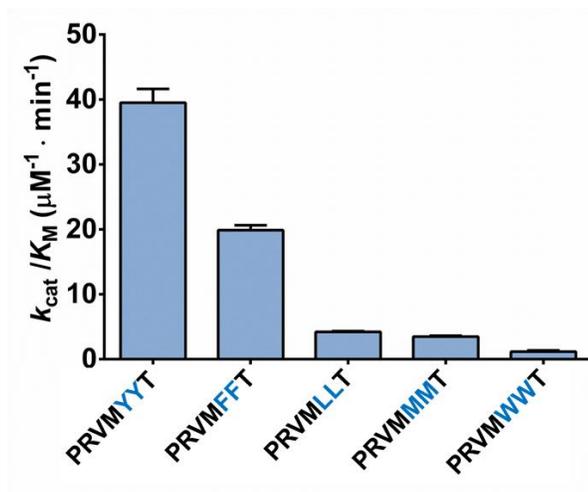


Figure 4.3. hK7 prefers cleavage after tyrosine (Y) at the P1 position.

A panel of FRET reporter substrates was constructed in bacteria to determine which residues can be accommodated within the S1 pocket of hK7. Assays with 10 nM wild-type hK7 and 200 nM substrate revealed a strong preference for tyrosine (Y) at the P1 position. hK7 can also cleave after phenylalanine (F), leucine (L), methionine (M), and tryptophan (W) with reduced activity compared to tyrosine ($N = 3$). Data represented as mean \pm SD.

To target selectivity switches in hK7 toward different P1 residues, we created a rationally designed library of hK7. Five residues in the hK7 S1 pocket (N199, A200, G222, G227, G233) were targeted for saturation mutagenesis with degenerate codons. Residues were chosen based on crystal structural analysis to determine amino acids with close proximity to the P1 phenylalanine in the docked substrate (Figure 4.4). Residues N199, A200, G222, G227 and G233 are each respectively 5.4Å, 4.3Å, 3.8Å, 7.7Å, and 5.5Å away from the benzyl ring of phenylalanine and therefore likely play a role in

determining the activity and specificity of hK7 toward different P1 residues. Within the hK7 S1 pocket library, position N199 was replaced with all possible amino acids while position A200 was substituted with residues A, V, I, L, M, T, or P. Positions G222 and G227 were replaced with G, A, or V and position G233 was replaced with G, A, V, I, L, M, T or P. We chose to limit the substitutions at these positions to reduce the fraction of inactive variants due to more dramatic changes in the amino acid side chains²⁰⁴. The S1 pocket library was created using the hK7-2.10 variant as a template, which has higher intracellular expression levels (Chapter 3).

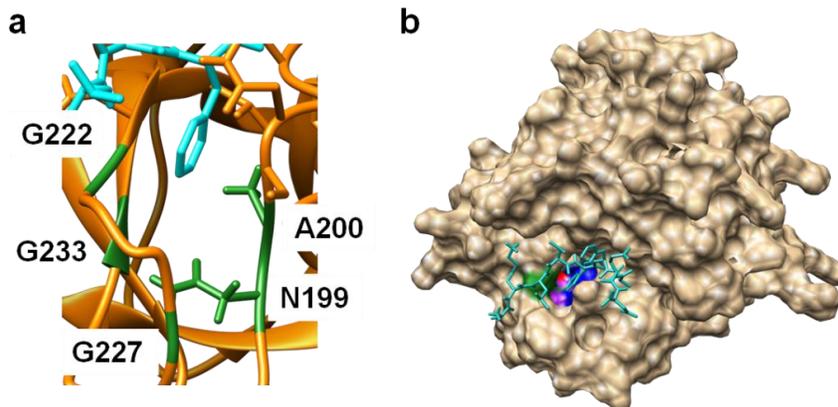


Figure 4.4. Structural analysis of the S1 subsite of hK7.

(a) Five residues within the S1 pocket of hK7 (N199, A200, G222, G227, G233) were selected for site-saturation mutagenesis to create a targeted hK7 library. (b) hK7 docking model with the bound A β 8 peptide (light blue) indicating the localization of position N199 (purple), A200 (blue), G222 (green), and G233 (red). Position G227 is located deeper within the pocket and therefore not shown. The images shown were created using PDB:2QXI with the UCSF Chimera Software package.

To isolate hK7 variants with improved activity toward P1 phenylalanine (F), yeast harboring the engineered PRVMFFT FRET substrate with A206K substitutions were transformed with the hK7 S1 pocket library yielding a library size of 3×10^7 members. Control populations of cells co-expressing the inactive hK7 variant S205A and the

engineered PRVMFFT substrate or the hK7-2.10 variant and the engineered PRVMFFT substrate for 6 hours at 22°C were analyzed by flow cytometry to determine the appropriate gates for library screening (Figure 4.5). Cells carrying the hK7 S1 pocket library were co-expressed with the PRVMFFT substrate at 22°C for 6 hours and cells with high cyan/yellow fluorescence ratio were sorted over six rounds from 0.6% to 10% gated cells in the final round (Figure 4.6). Cells collected after the fifth and sixth rounds of sorting were plated to determine the substitutions within isolated variants.

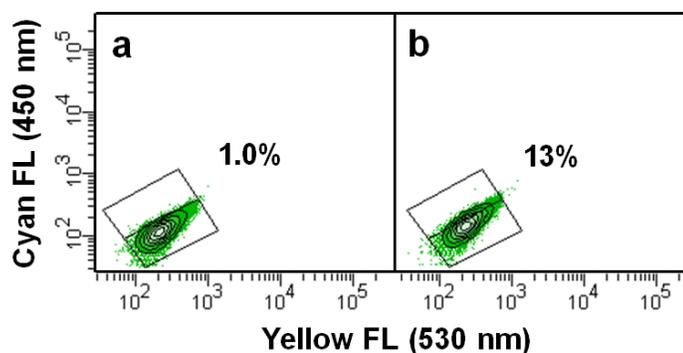


Figure 4.5. Analysis of hK7 activity *in vivo* against the PRVMFFT substrate.

To determine the proper gates for sorting the hK7 S1 pocket library, control populations co-expressing inactive hK7 and the engineered PRVMFFT FRET reporter (a) or variant hK7-2.10 and the engineered PRVMFFT FRET reporter (b) for 6 hours at 22°C were analyzed by flow cytometry. The sorting gate was drawn based on the cyan/yellow fluorescence readout of the negative population allowing for only 1% of negative cells to be included in the sorted fraction.

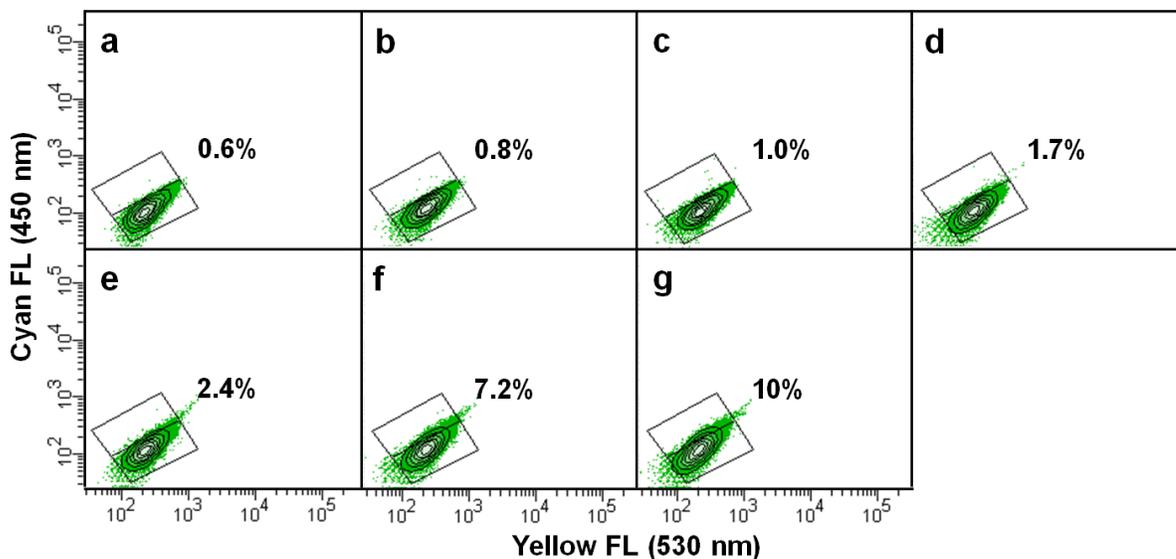


Figure 4.6. Sorting an hK7 S1 pocket library for cleavage of PRVMFFT.

An hK7 S1 pocket library consisting of partial mutagenesis to residues N199, A200, G222, G227, and G233 was transformed into cells harboring the engineered PRVMFFT FRET reporter. The library was co-expressed with the PRVMFFT substrate and cells were sorted as stated in Methods. FACS plots of the initial library population (a) and after the first (b), second (c), third (d), fourth (e), fifth (f), and sixth (g) rounds of enrichment.

Forty individual clones from the fifth and sixth rounds of sorting were sequenced using yeast colony PCR to determine the diversity of unique variants. Analysis of the substitutions occurring at the five mutated positions within the naïve library and after the fifth round of sorting revealed certain preferences for substitutions that may be important for hK7 activity toward F at P1. For example, within the naïve library, position N199 was mutated to all amino acids but by the fifth round of sorting there was a strong preference for the wild-type residue asparagine (N) and also a tolerance for cysteine (C) at this position (Figure 4.7). Position A200 displayed a strong preference for the small amino acid proline (P) and also a tolerance for threonine (T) and valine (V) in variants isolated from the fifth round of sorting (Figure 4.8). Interestingly, position G222 did not seem to

accommodate any other amino acids and over 80% of the variants sequenced had the wild-type glycine at this position by the fifth sort (Figure 4.9). The wild-type glycine was also highly favored at position G227 (Figure 4.10), whereas sequenced variants tolerated multiple substitutions at position G233 (Figure 4.11). From sequencing analysis, we hypothesized that positions N199, A200, and G233 may be more tolerant to substitutions whereas mutations to the glycines at positions 222 and 227 may have deleterious effects to hK7 folding and activity on PRVMFFT.

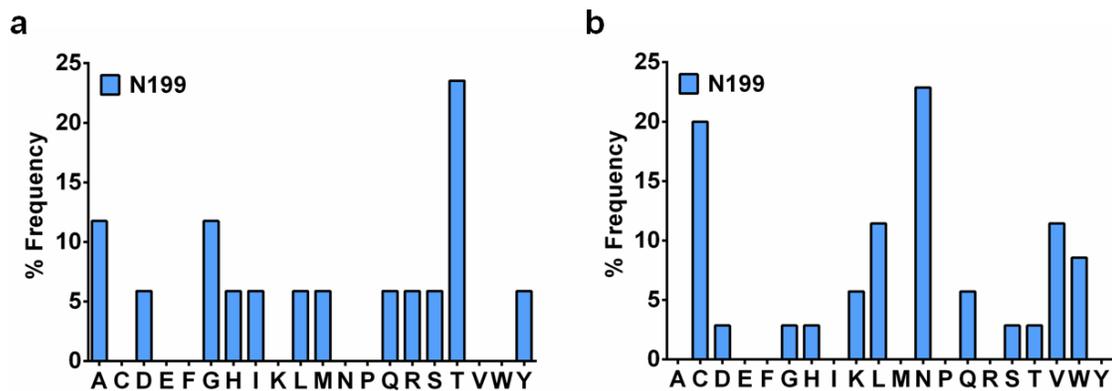


Figure 4.7. Amino acids C and N are prevalent at position 199 after post sort 5.

Substitutions within twenty sequenced clones of the naïve library and forty sequenced clones in the post sort 5 population were analyzed to determine amino acid prevalence at position 199 in the naïve library (a) and after post sort 5 (b). Within the naïve library, N199 was mutated to all amino acids. However, after five rounds of sorting for cleavage of PRVMFFT, the majority of variants analyzed showed a strong preference for either the wild-type residue asparagine (N) and also a preference toward mutations to cysteine (C).

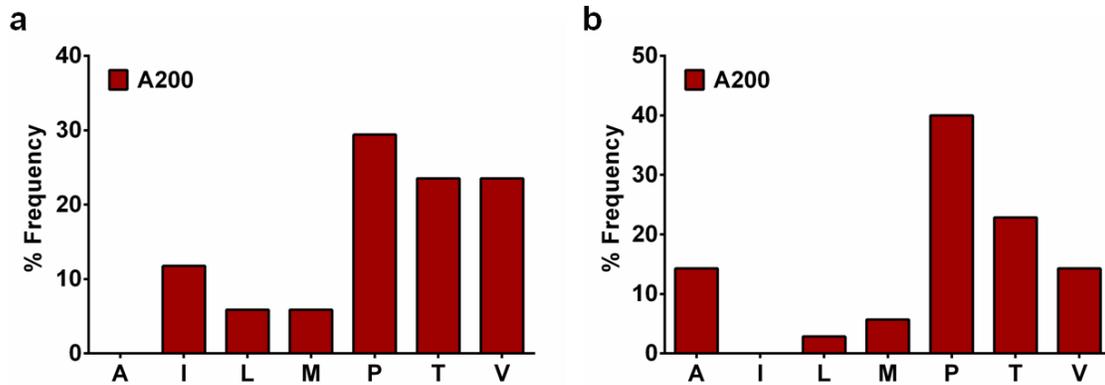


Figure 4.8. Position 200 tolerates small amino acid substitutions.

The frequency of substitutions at position A200 was evaluated in twenty sequenced clones of the naïve library (a) and forty sequenced clones in the post sort 5 population (b). A200 was mutated to residues A, I, L, M, P, T, or V in the focused library. Substitutions to small amino acids such as proline (P), threonine (T) and valine (V) were well tolerated at A200. However, the larger amino acids isoleucine (I), leucine (L), and methionine (M) displayed low substitution frequency at A200 in the post sort 5 clones.

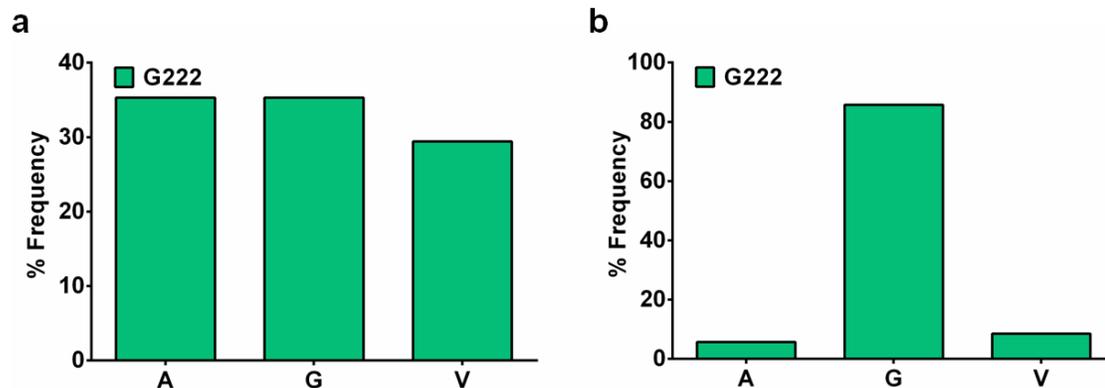


Figure 4.9. The wild-type residue glycine is strongly preferred at position 222.

Analysis of substitution frequency at position G222 in the naïve library (a) and in the post sort 5 population (b) revealed that the wild-type residue glycine is highly preferred at this position. G222 was mutated to G, A, or V in the naïve library but by post sort 5, over 80% of clones had reverted back to the wild-type residue glycine (G).

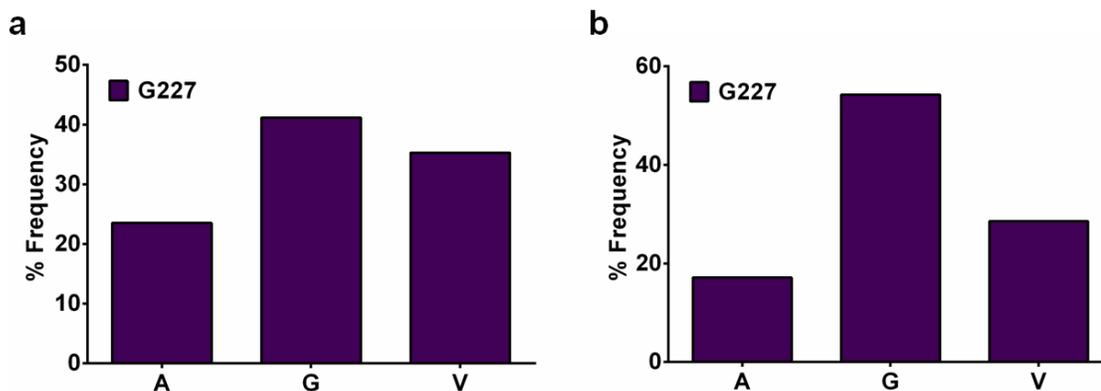


Figure 4.10. Position 227 displays preference for the wild-type glycine.

The amino acid prevalence at position G227 was evaluated in twenty sequenced clones of the naïve library (a) and forty sequenced clones in the post sort 5 population (b). While position G227 was mutated to residues G, A, or V in the S1 pocket library, post sort 5 clones displayed a strong preference for the wild-type residue glycine (G) at this position.

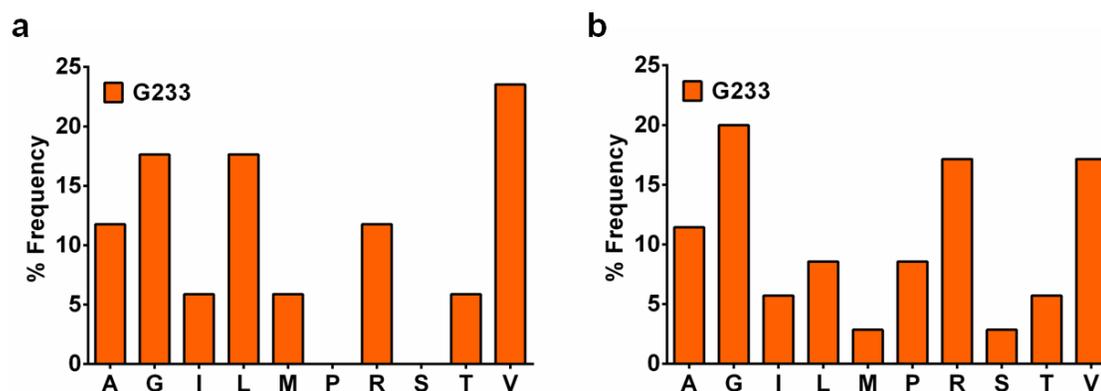


Figure 4.11. Multiple residues are tolerated at position 233.

The substitution frequency at position G233 was evaluated for twenty sequenced clones in the naïve library (a) and forty sequenced clones in the post sort 5 population (b). Although, position 233 was targeted for mutagenesis to G, A, I, L, M, P, T or V, the degenerate codon used for these substitutions also included potential mutations to serine (S) and arginine (R). After five rounds of sorting for PRVMFFT activity, no clear preference was seen at position 233 although the wild-type glycine (G) was seen with highest frequency in sequenced clones.

Sequencing analysis of the post sort 6 population revealed that a single variant hK7-S1.9 which contained three substitutions N199C, A200P, G233A had enriched to over 65% of the population by the sixth sort (Table 4.1). The wild-type gene was present in 5% of the sequenced clones of the post sort 6 population while it was not present in post sort 5 sequenced clones indicating successful enrichment of active clones during the sixth round of sorting (Table 4.1). Interestingly, no variants with substitutions to G222 were sequenced from the post sort 6 population confirming the importance of the wild-type glycine at this position for activity. Unique variants discovered from sequencing analysis of the post sort 5 and post sort 6 populations were evaluated in cell lysates for activity toward the target PRVMFFT substrate and a non-target PRVMYYT substrate.

All variants analyzed displayed decreased activity in cell lysates on the target substrate PRVMFFT compared to the parent protease hK7-2.10, which was used as a template to construct the library. Ten variants with the highest *in vitro* activities toward PRVMFFT were compared to the activity of variant hK7-2.10 (Figure 4.12). Variants with the highest PRVMFFT activity contained substitutions to positions N199, A200, and G233 which were predicted to be more tolerant to amino acid changes from sequencing analysis of enriched clones. However, the activity of these variants was reduced by 2-fold or greater on PRVMFFT when compared to wild-type. Variants with substitutions to G227 all displayed more dramatic reductions in activity for the target substrate PRVMFFT. Overall substitutions within the S1 pocket of hK7 seemed to be deleterious for hK7 activity on PRVMFFT. Variant hK7-S1.10 showed a 1.4-fold improved selectivity on PRVMFFT over PRVMYYT however this modest change in specificity came at a cost to the catalytic activity (Figure 4.12). Variants with improved activity on PRVMFFT may not have been

present in the naïve library because the residues chosen for mutagenesis were critical for activity as seen by the majority of sequenced clones containing the wild-type residues by post sort 5. Or variants with improved activity in the naïve library were present at too low of a frequency for the screen to isolate them from the background of inactive variants.

Table 4.1. Substitutions identified for hK7 variants with highest PRVMFFT activity.

hK7 variant	substitutions	post sort 5 frequency	post sort 6 frequency
S1.1	A200V, G227A, G233V	0/40	1/40
S1.2	N199S, A200T, G227A	1/40	0/40
S1.3	N199S, G227A	0/40	2/40
S1.4	N199Q, A200P, G233P	0/40	1/40
S1.5	N199Q	1/40	0/40
S1.6	G227A	1/40	0/40
S1.7	N199Q, A200P	0/40	1/40
S1.8	N199T, G233V	0/40	1/40
S1.9	N199C, A200P, G233A	3/40	27/40
S1.10	N199A, A200V, G233A	0/40	2/40
S1.wt	none	0/40	2/40

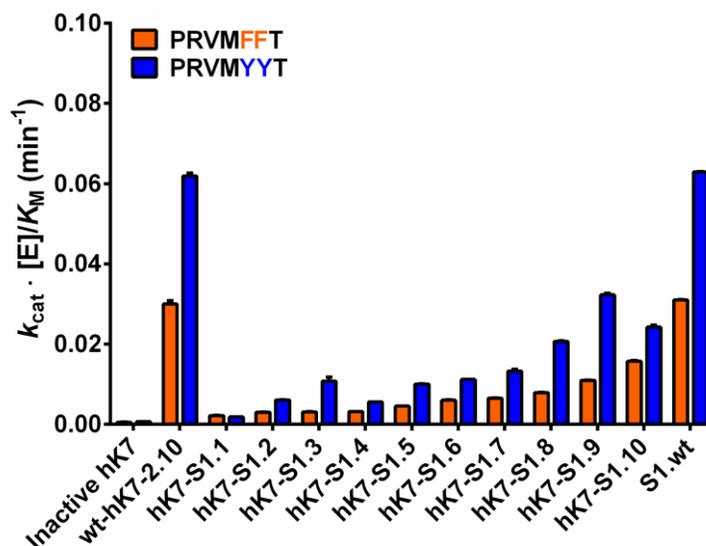


Figure 4.12. S1 pocket variants display reduced activity on PRVMFFT.

Unique variants sequenced from the post sort 5 and post sort 6 populations screened for intracellular PRVMFFT cleavage were assessed for *in vitro* activity. Activity in cell lysates was measured toward the target PRVMFFT and non-target PRVMYYT substrates to determine changes in P1 selectivity. Equal numbers of cells expressing each variant were lysed to determine a relative second order rate constant $k_{cat} \cdot [E]/K_M$ ($N = 3$). Here the top ten variants with activity toward PRVMFFT are shown. All variants displayed reduced activity on the target PRVMFFT, while one variant hK7-S1.10 showed a modest 1.4-fold switch in selectivity for PRVMFFT over PRVMYYT. Data represented as mean \pm SD.

4.3. Discussion

For proteases with solved crystal structures, site-directed mutagenesis of residues within the active site engaging in close contacts with the substrate has been successful in generating changes in protease selectivity⁶⁰⁻⁶². However, at each site chosen for mutagenesis, there are 19 other potential amino acids that could be substituted and the number of variants to be analyzed grows exponentially if more than one site is chosen for mutagenesis. Because it is still difficult to predict which substitutions will be beneficial at a particular position, the development of screens for protease activity, such as PrECISE,

enables screening multiple combinations of substitutions within the active site simultaneously to determine optimal residues at each position. Additionally, by distinguishing which substitutions within the active site increase or decrease the activity of the protease toward a particular substrate, a deeper understanding of the mechanism by which proteases recognize their substrates can be gained. Therefore we applied the PrECISE methodology toward a focused library of the hK7 S1 pocket to elucidate which residue substitutions in this subsite may improve activity toward phenylalanine (F) over tyrosine (Y).

The dynamic range of the PrECISE method was increased by reducing the interaction between YPet and CyPet after proteolytic cleavage of the FRET reporter substrate to enhance isolation of active variants. Addition of the substitution A206K to CyPet and YPet led to an increase in the separation between inactive hK7 and active hK7 clones on flow cytometry while maintaining sufficient FRET for screening²⁰⁵. The engineered FRET reporter was used for screening an hK7 S1 pocket library consisting of partial saturation mutagenesis of five residues (N199, A200, G222, G227, G233) for improved activity on the target substrate PRVMFFT. The five positions in the hK7 S1 subsite were chosen for mutagenesis based their proximity to a P1 phenylalanine in a structural docking model. To reduce the number of inactive variants in the initial library population, four of the five sites (A200, G222, G227, G233) were partially saturated to prevent large changes in amino acid side chain chemistry that could cause the protease to unfold. Screening the hK7 S1 pocket library for PRVMFFT cleavage led to slow enrichment of clones with improved cyan/yellow fluorescence in early rounds indicating that a low fraction of the library was

active. However, by the sixth round of sorting, the cell population displayed improved cyan/yellow fluorescence that was comparable to the positive control.

Sequencing analysis of 40 random clones from the post sort 5 and post sort 6 populations revealed that substitutions to G222 and G227 were not well tolerated, whereas positions N199, A200, and G233 displayed a higher prevalence of amino acid switches in enriched clones. However, at all positions, except A200, the wild-type residue showed the highest frequency of enrichment. Additionally, the wild-type sequence was present in 5% of clones in the post sort 6 population confirming isolation of wild-type-like variants. Interestingly, when activity of the unique hK7 S1 pocket variants was evaluated in cell lysates, all variants displayed reduced activity on PRVMFFT compared to the parent protease hK7-2.10 used to make the library. This result further confirmed that substitutions to the hK7 S1 pocket were not well tolerated and led to deleterious effects in activity of hK7.

There are many potential reasons why enriched clones were not more active for the target substrate compared to wild-type. First, the five positions within the S1 pocket chosen for mutagenesis may be critical for folding and activity. Therefore the initial library contained weaker activity variants and only sub-optimal variants could be enriched. On the other hand, we may not have enriched beneficial variants because substitutions to more than five positions in the active site were necessary to improve hK7 activity and selectivity. Expanding the zone of mutagenesis has previously been shown to improve selectivity when small groups of mutations had no effect^{71,202}. Additionally, the PrECISE screen may not be sensitive enough for the isolation of low frequency variants that have enhanced activity due to the overlap between positive and negative populations and the

inherent toxicity of hK7 variants with increased activity. Further, the screen does not contain a counter-selection substrate which could dramatically improve the probability of isolating a variant with improved selectivity and reduce the appearance of wild-type like variants in the sorted populations.

4.4. Materials and methods

4.4.1. Addition of A206K substitutions to the FRET reporter substrates

To increase the dynamic range of the PrECISE screen, the substitution A206K was added to both CyPet and YPet to reduce dimerization that may occur between the monomers after proteolytic cleavage of the FRET reporter substrate. To introduce the A206K substitution to CyPet, PCR driven overlap extension¹⁵¹ was used with internal forward primer 5' TTGTCTACCCAGTCAAAAATTGTCCAAGGACCCA and internal reverse primer 5' TGGGTCCTTGGACAATTTTTGACTGGGTAGACAA. Similarly, the A206K substitution was added to YPet using internal forward primer 5' TTATCCTATCAATCTAAAAATTATTCAAAGATCCA and internal reverse primer 5' TGGATCTTTGAATAATTTTTAGATTG ATAGGATAA. The FRET reporters carrying the substrates PRVMYYT and PRVMFFT were used as templates for the addition of the A206K substitutions. The resulting full-length FRET reporter fragments were cloned between the HindIII and PmeI sites in the pYES3/CT plasmid (Invitrogen) and transformed into DH5 α *E. coli* as described in Chapter 2. Plasmid DNA was recovered from bacteria and digested with Bsu361 to linearize the plasmid for integration into the genome of the yeast strain JYL69 as described in Chapter 2.

The fluorescence profiles of individual clones expressing the engineered FRET reporters with substrates PRVMFFT or PRVMYYT were assessed via flow cytometry.

Cells were cultured overnight in SD-Trp media with 2% raffinose to an OD₆₀₀ of 4. Cells were then subcultured to an OD₆₀₀ of 0.75 in fresh SD-Trp media with 2% raffinose and 2% galactose for expression at 22°C for 6 hours. After expression, the cells were pelleted by centrifugation at 5,000g for 5 minutes and resuspended in 1x PBS for analysis on a FACSAria flow cytometer (BD Biosciences). Cells were analyzed with violet light excitation (407 nm) to determine the cyan (450/40 nm) and yellow (530/30 nm) fluorescence intensities. Highly fluorescent clones expressing the engineered PRVMFFT FRET reporter or the engineered PRVMYYT FRET reporter were selected for co-transformation with the wild-type hK7 plasmid and inactive hK7 plasmid.

4.4.2. *Construction of an S1 pocket library of hK7*

Five residues within the S1 pocket of hK7 (N199, A2000, G222, G227, G233) were partially saturated to create a targeted library of hK7. Splice-overlap extension PCR was used to introduce the degenerate codon NNS at position 199 which encodes for all the amino acids, codon VYS was introduced at position 200 which encodes for A, V, I, L, M, T, and P, codon GBY was introduced at positions 222 and 227 which encodes for G, A, and V and codon VBS was introduced at position 233 to encode for G, A, V, I, L, M, T, and P (where B, G, N, S, V, Y are standard nucleotide codes). Primers were designed based on coding sequence of the hK7-2.10 variant as a template, which contains three substitutions R100C, R123M and R212S shown to improve intracellular expression levels (Chapter 3) (Table 4.2). Splice overlap extension was performed as described¹⁵¹ with the high fidelity KOD polymerase using an Eppendorf Mastercycler thermocycler with a program as follows: 1 cycle, 3 min at 94°C; 30 cycles, 30 sec at 94°C, 30 sec at 60°C, 1

min at 72°C; 1 cycle, 5 min at 72°C. The amplified library PCR product was digested, ligated into vector pYC2/CT (Invitrogen), and DH5α *E. coli* were transformed by electroporation with library plasmid DNA as described in Chapter 3. Library transformants were picked from LB + ampicillin plates to send for sequencing to determine the library quality.

Table 4.2. Primers used in SOEing PCR for hK7 S1 pocket library construction.

Forward external	5'TAGAAAGCTTGGCCACATAGGCCACCATGTTGTTGCAAGC ATTTTGTTTTTGTTGGCTGGTTTTGCTGCTAAAATTTCTGCT
Reverse internal	5'ACACCAGACCTTGCAGGGTACCACTGCACACCAACGGTCC CCCTGAGTCACCATTGCAS SRBSNN TTTCTTGGAGTCGGG
Forward internal	5'TACCCTGCAAGGTCTGGTGTCTGG GBY ACTTTCCTTGC GBY CAACCCAATGACCCA VBS GTCTACTCAAGTG
Reverse external	5'ATCTGTTTAAACGGCCGAATTGGCCTTATCAATGATGATG ATGATGATGACC

Yeast cells containing the engineered PRVMFFT FRET reporter gene were transformed with library plasmid DNA using the high efficiency PEG/lithium acetate method¹⁸⁵ and plated on SD media lacking tryptophan and uracil to determine library size. The transformed cells were added to 500 mL SD-Trp-Ura media supplemented with 2% glucose and penicillin-streptomycin (pen-strep) at a final concentration of 100 units/mL and 100 µg/mL (Invitrogen) and allowed to recover for 24 hours.

4.4.3. Library screening and FACS analysis

Cells carrying the engineered PRVMFFT FRET reporter and S1 pocket library of hK7 were grown overnight in 50 mL SD-Trp-Ura media with 2% raffinose and penicillin-

streptomycin (pen-strep) at a final concentration of 100 units/mL and 100 μ g/mL (Invitrogen). Overnight cells were subcultured to a final OD₆₀₀ of 0.75 in fresh SD-Trp-Ura media containing 2% raffinose and induced with 2% galactose. The hK7 S1 pocket library and the engineered PRVMFFT FRET reporter were co-expressed at 22°C for 6 hours with shaking. A total of 2×10^7 cells were spun down and resuspended in 1x PBS buffer for screening using a FACSAria flow cytometer (BD Biosciences) to isolate clones displaying an improved cyan/yellow fluorescence ratio. Cells were sorted as described in Chapter 3 and plated on selective SD-Trp-Ura media after each round of sorting. Forty individual clones from the fifth and sixth rounds of sorting were amplified using yeast colony PCR and sent for sequencing (Genewiz) to determine which mutations had occurred.

4.4.4. *Characterization of variant activity in cell lysate*

To determine the activity of individual variants, cells were cultured overnight in 10 mL SD-Trp-Ura media with 2% raffinose to an OD₆₀₀ of 3. Cells were then subcultured to an OD₆₀₀ of 1 in 30 mL fresh SD-Trp-Ura media with 2% raffinose and 2% galactose for expression at 22°C for 6 hours. After expression, cells were centrifuged at 5000g for 5 minutes and the cell pellets were frozen at -80°C overnight. Cells were lysed by adding 450 μ L Y-PER reagent (Thermo Scientific) and gently shaking the homogenous mixture at room temperature for 30 minutes. Cell lysate fractions were recovered by spinning the solution at 10,000g for 5 minutes. Protease activity was measured in lysates using the FRET reporter assay described in Chapter 3 to determine relative second order rate constants $k_{\text{cat}} \cdot [\text{E}] / K_{\text{M}}$ of the variants.

5. Incorporation of a non-target substrate within the cell-based screen

Most human proteases participate in multiple physiological pathways and have been estimated to cut anywhere from 10 to 100 substrates to fulfill their biological functions. Therefore their activity, expression, and localization are tightly controlled to prevent unwanted proteolysis. This lack of substrate specificity has severely limited the use of proteases in therapy since unregulated proteolysis would lead to adverse side-effects. The redesign of proteases with narrow selectivities could be greatly enhanced by the development of protease engineering screens that not only profile protease activity toward a target substrate but also incorporate counterselection substrates to reduce off-target activity. To improve upon the PrECISE methodology for the isolation of selective variants, we sought to incorporate a non-target counterselection substrate within the cell-based assay for simultaneous assessment of protease activity toward a target and non-target substrate. Insertion of a non-target substrate within a surface exposed loop of YPet did not lead to reduced yellow fluorescence in the presence of hK7. Therefore, we developed a new cell-based assay using degradation signals attached to fluorescent substrates to determine if hK7 co-expression could rescue fluorescent proteins from degradation by the yeast proteasome. Although, hK7 co-expression did not lead to higher fluorescence output of fluorescent protein substrates with degradation tags, the studies described here provide a starting point for future incorporation of a counterselection substrate within the PrECISE method.

5.1. Introduction

The ability to engineer proteases with high activity and selectivity for a target substrate could greatly expand their use in therapeutic and biotechnological applications. However, laboratory evolution of enzymes to accept novel substrates has generally led to the discovery of “promiscuous” variants which retain activity on wild-type substrates in addition to cleaving a newly selected target substrate^{67–69,206}. In nature, enzymes evolve new functions through positive selection pressure for the target substrate and negative selection pressure to reduce off-target activity that would lead to adverse side effects. Therefore to emulate natural evolution, protease engineering screens have been developed with positive and negative selection pressure to isolate variants with activity on a target substrate, while removing variants which cut a wild-type preferred sequence. Screens which simultaneously profile protease activity on a target substrate and a counterselection substrate have generated selectivity switches on the order of 1,000-3,000,000 fold^{70,75}. However, the incorporation of a counterselection substrate has not yet been applied for engineering human protease selectivity. Here, we have tested two routes for the incorporation of a non-target substrate within the PrECISE screen for the development of highly selective human secreted proteases for their expanded use in therapy.

A potential strategy for the incorporation of a counterselection substrate within the PrECISE screen would be the insertion of a non-target substrate within an exposed loop of YPet such that cleavage of the non-target substrate would lead to a reduction in blue-excited yellow fluorescence. Insertion of random peptide sequences within exposed loops of GFP has in most cases led to deleterious effects for GFP folding and fluorescence²⁰⁷. However, tyrosine at position 145 of YFP was found to tolerate whole protein insertions in

addition to peptide insertions without affecting yellow fluorescence²⁰⁸. It is hypothesized that the two portions of YFP on either side of Y145 have the ability to independently fold in the presence of an inserted sequence. Interestingly, the insertion of the calcium binding protein calmodulin at position Y145 in YFP led to increased yellow fluorescence in the presence of calcium ions indicating that conformational changes in the insert could affect fluorescence²⁰⁸. However, protease hydrolysis of a peptide substrate inserted at Y145 in YFP has not been tested and may provide a route for counterselection if cleavage of the peptide destabilizes fluorescence.

An additional strategy for the introduction of a counterselection substrate for protease engineering would be splitting the FRET reporter into the individual fluorescent proteins YPet and CyPet which would carry a target or non-target substrate at the C-terminus, respectively, followed by a degradation signal that targets the complex for degradation by the yeast proteasome. In this screen, cleavage of the target or non-target substrate would release the degron tag and increase fluorescence of the corresponding reporter. A similar strategy was employed in bacteria to profile TEV specificity using short-lived fluorescent substrates in the cytoplasm of *E. coli* where protease co-expression caused an increase in fluorescence proportional to protease activity on the substrate^{209,210}. To apply this screening strategy in yeast, C-terminal hydrophobic peptides have been discovered that promote degradation of fusion proteins by the yeast proteasome^{211,212}. These degron tags additionally contain a positively charged residue where attachment of ubiquitin occurs to target degradation by the ubiquitin system^{213,214}. Here we have tested two routes for incorporation of a counterselection substrate within the PrECISE screen to improve isolation of selective variants.

5.2. Results

5.2.1. Introducing a counterselection substrate within a YPet exposed loop

Addition of a non-target counter-selection substrate within the cell-based assay would allow for enrichment of protease variants with a specific selectivity profile and also serve to decrease enrichment of wild-type-like variants. We tested whether the yellow fluorescent protein (YPet) within the FRET reporter would tolerate insertion of a non-target substrate such that cleavage of the non-target would unfold YPet and abolish blue-excited yellow fluorescence as measured by fluorimetry and flow cytometry. Substrates were cloned between Y145 and N146 within YPet which has previously been shown to tolerate peptide insertion without reducing fluorescence²⁰⁸. The four substrates inserted into YPet included the well-cleaved hK7 substrate PRVMYYT, a negative control GGSGSGGS, and extended substrates GGSGPRVMYYTGGSG, GGSGGGSGSGGS GGSG to determine if longer substrates would be more accessible to hK7 proteolysis. These constructs were expressed in bacteria with a His-tag and purified using affinity chromatography for *in vitro* analysis of susceptibility to cleavage by wild-type hK7.

After purification of the YPet substrates, sample concentrations were measured by fluorimetry and YPet substrates were diluted in water to 6 μM . We tested *in vitro* cleavage of YPet with either the PRVMYYT or GGSGSGGS insertion compared to a negative control YPet probe without substrate insertion in assays using 240 nM of the substrate and 5 nM, 16 nM and 80 nM of commercial wild-type hK7. Yellow fluorescence of the substrates was measured over 2.5 hours at 37°C to determine if hK7 cleavage could significantly reduce yellow fluorescence of YPet with the PRVMYYT substrate insertion. As a negative control, the hK7 active site inhibitor SBTI was added at a final concentration

of 10 μM to completely shut down hK7 activity and determine the background yellow fluorescence. No significant decrease in yellow fluorescence was seen for addition of hK7 to YPet with the PRVMYYT insertion at the highest protease concentration tested compared to the background with inhibitor indicating no cleavage of the inserted peptide (Figure 5.1). Cleavage may not have occurred due to the accessibility of the insertion site. Additionally, if proteolysis of the substrate did occur, YPet may have been stable after cleavage.

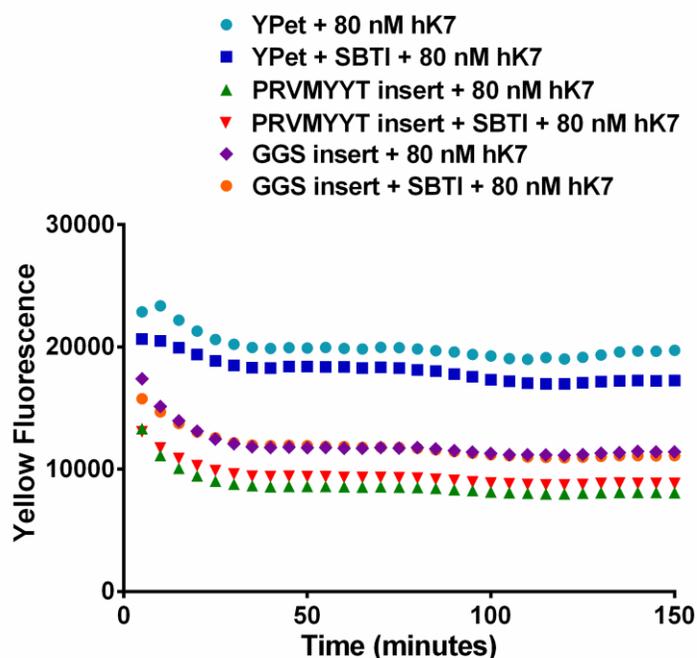


Figure 5.1. Substrates inserted into YPet are not susceptible to cleavage *in vitro* by hK7.

Substrates PRVMYYT and GGSGSGGS were inserted between residues Y145 and N146 in YPet to determine if addition of wild-type hK7 could reduce yellow fluorescence by cleaving the optimized PRVMYYT substrate. As a negative control, YPet without substrate insertion was analyzed for non-specific cleavage. The tight binding active site inhibitor SBTI (10 μM) was additionally used to determine the background fluorescence of each sample. Addition of 80 nM hK7 did not lead to a reduction in yellow fluorescence for YPet with the PRVMYYT insert compared to background.

5.2.2. *Protease profiling using short-lived fluorescent substrates*

To develop a novel cell-based screen with the incorporation of a counterselection substrate, the FRET reporter was split into the individual proteins YPet and CyPet and a C-terminal yeast degradation signal was added to both fluorescent proteins. The hydrophobic amino acid sequence ACKNWFSSLSHFVIHL, which we have termed DegK, has been shown to target proteins for degradation by the ubiquitin system in yeast when fused at the C-terminus²¹¹. Within this novel screen, a target substrate can be inserted between YPet and DegK and multiple non-targets can be inserted between CyPet and DegK. Cleavage of the target substrate will release the degradation tag causing the cell to fluoresce yellow (Figure 5.2). Off-target activity on a wild-type preferred non-target substrate would lead to an increase in cyan fluorescence. The absence of protease activity on either substrate permits attachment of ubiquitin molecules to the degron tag and degradation of the fluorescent proteins by the proteasome causing cells to be non-fluorescent. Using this screen, cells with high yellow fluorescence and minimal cyan fluorescence are isolated via FACS indicating cleavage of only the target substrate (Figure 5.2).

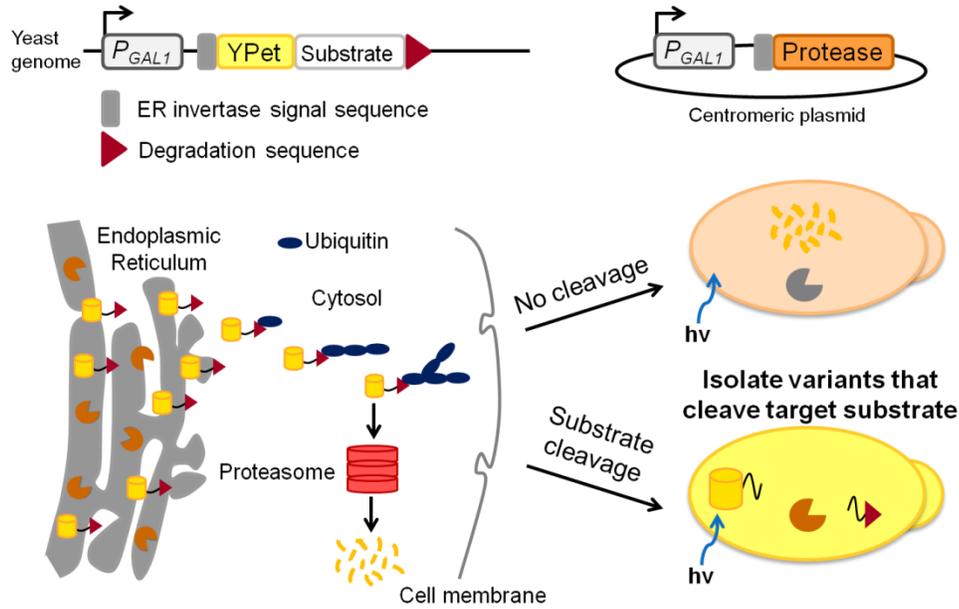


Figure 5.2. Cell-based screen for protease engineering using degradation of fluorescently tagged substrates.

A protease susceptible substrate is flanked by an N-terminal fluorescent protein and a C-terminal degradation signal that targets the complex for proteolysis by the proteasome in the cytosol. Protease co-expression with the substrate in the ER releases the degradation tag allowing the cell to fluoresce the corresponding color of the upstream fluorescent protein. Using this assay, a target substrate can be inserted between YPet and the degradation signal and multiple non-targets can be inserted between CyPet and the degradation signal. Cleavage of the target substrate by a protease variant leads to an increase in yellow fluorescence enabling isolation of cells with cleaved target substrates using FACS.

To test the effectiveness of yeast degradation signals to reduce YPet fluorescence, the degron tag ACKNWFSSLSHFVIHL (DegK) was fused to C-terminus of YPet and yellow fluorescence was monitored after expression in yeast. The yellow fluorescence of cells expressing YPet-DegK was compared to cells expressing YPet using flow cytometry to determine the maximum dynamic range of the system. Cells expressing YPet-DegK at 22°C or 30°C for expression times between 0 to 6 hours displayed reduced yellow

fluorescence compared to cells expressing YPet (Figure 5.3). After 6 hours of expression, the separation between positive and negative cells was greater than 2-fold. Four substrates were inserted between YPet and DegK to determine if substrate insertion affected degradation of the yellow fluorophore. Cells expressing the YPet-substrate-DegK constructs at 22°C or 30°C for 6 hours showed a similar 2-fold or greater reduced yellow fluorescence compared to YPet expressing cells (Figure 5.4). Insertion of the PRVMYYT substrate increased degradation and cells expressing YPet-PRVMYYT-DegK displayed greater than 3.3-fold reduced yellow fluorescence compared to YPet expressing cells.

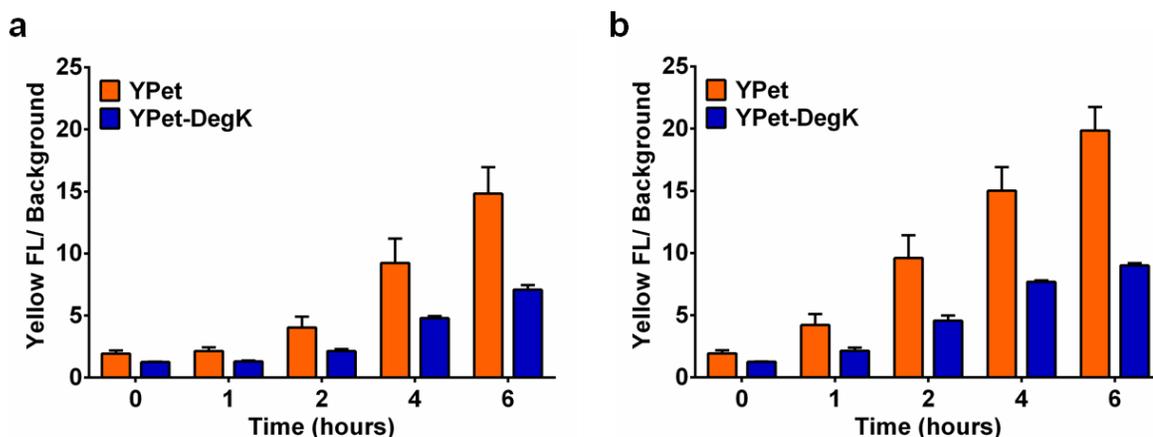


Figure 5.3. Addition of DegK to the C-terminus of YPet leads to reduced yellow fluorescence after expression.

Yellow fluorescence intensity was measured for cells expressing YPet-DegK and compared to cells expressing YPet using flow cytometry after expression at 22°C (a) or 30°C (b) for various lengths of time from 0-6 hours. Expression of YPet-DegK led to a reduction in YPet fluorescence for all expression times analyzed. After 6 hours of expression at 22°C or 30°C, YPet-DegK intracellular yellow fluorescence was reduced greater than 2-fold compared to cells expressing YPet (N = 3). Data represented as mean \pm SD.

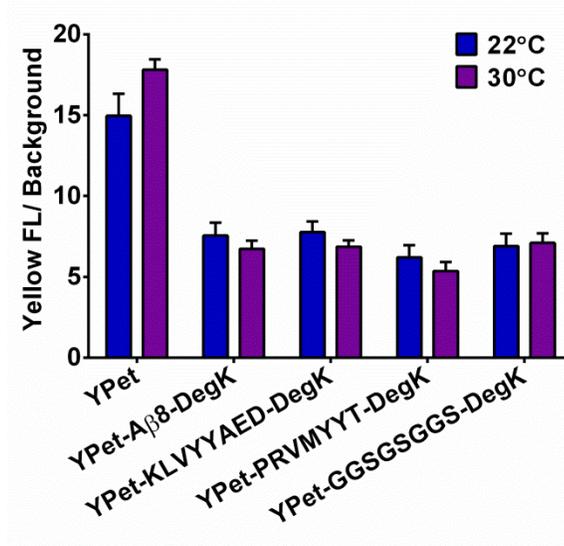


Figure 5.4. Substrate insertion between YPet and DegK does not affect degradation.

The four hK7 substrates KLVFFAED (A β 8), KLVYYAED, PRVMYYT, and GGSGSGGS were inserted between YPet at the N-terminus and DegK at the C-terminus. Expression of the YPet-substrate-DegK constructs at 22°C or 30°C for 6 hours, led to greater than 2-fold reduction in intracellular yellow fluorescence compared to cells expressing YPet (N = 3). Data represented as mean \pm SD.

To determine if substrate proteolysis could rescue yellow fluorescence of the YPet-substrate-DegK constructs, active hK7 and inactive hK7 were co-transformed into yeast strains carrying the four substrates for co-expression analysis using flow cytometry. Because hK7 can potentially cleave the DegK sequence which would give a false positive signal, we first tested hK7 activity on the DegK sequence *in vitro*. hK7 did not effectively cleave the DegK signal, likely due to the lack of tyrosine residues, and therefore off-target cleavage of the degradation signal would not contribute to any increase in intracellular fluorescence. Co-expression of active hK7 with substrates PRVMYYT, KLVYYAED or KLVFFAED flanked between YPet and DegK at 30°C for 6 hours did not lead to an increase in yellow fluorescence compared to cells co-expressing an inactive hK7 variant

with the substrates (Figure 5.5). In contrast, wild-type hK7 co-expression caused a further reduction in yellow fluorescence. This result was also seen for active hK7 co-expression at 22°C for 6 hours.

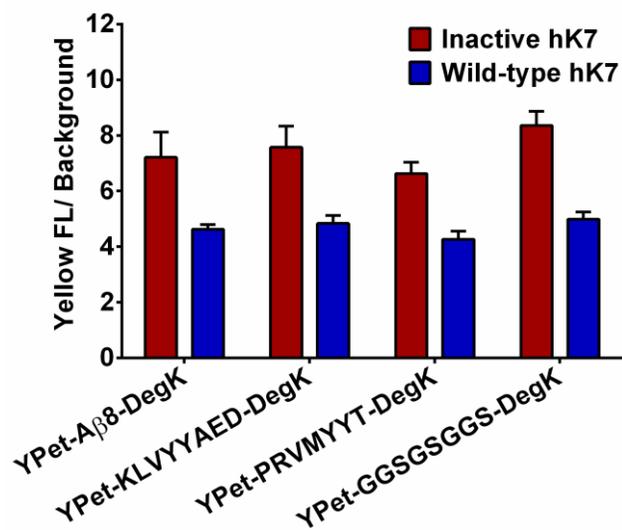


Figure 5.5. Active hK7 co-expression does not rescue yellow fluorescence of YPet substrates fused to DegK.

Cells carrying the YPet-substrate-DegK constructs were co-expressed with either active hK7 or inactive hK7 to determine if active hK7 co-expression could cleave the substrates releasing the degradation tag causing increased intracellular yellow fluorescence. However, active hK7 co-expression led to a reduction in yellow fluorescence of all the substrates tested relative to inactive hK7 co-expression (N = 3). Data represented as mean \pm SD.

5.3. Discussion

Although the majority of protease engineering screens developed only profile protease reactivity with a target selection substrate, the incorporation of a non-target substrate has been shown to improve isolation of selective variants by removing variants which retain wild-type activity^{70,75}. The development of human protease therapeutics that display a high

level of specificity for residues flanking the cleavage site is critical since any off target activity could lead to adverse side effects. Here we tested two different methods for incorporation of a counterselection substrate within the PrECISE screen for engineering human protease selectivity toward peptide targets relevant in disease.

Non-target peptide substrates were inserted within a surface exposed loop of YPet to determine if proteolysis of the substrates could destabilize the protein and lead to a reduction in yellow fluorescence. However, the addition of high concentrations of wild-type hK7 did not cause a decrease in fluorescence *in vitro* of a YPet construct with an optimal hK7 substrate insertion. Potentially the substrate may have not been accessible to cleavage or more likely cleavage of the substrate did not reduce stability of the fluorescent protein. In support of this hypothesis, non-fluorescent GFP fragments have been found to associate strongly resulting in green fluorescence emission²¹⁵. Therefore, further testing would be necessary to determine if the YPet substrate insertion was cleaved and the fragments generated still led to fluorescence output.

An additional strategy for incorporation of a non-target substrate for protease engineering was tested using fluorescent substrates attached to degron signals mediating degradation by the yeast proteasome. In this screen, co-expression of an active protease should release the degron tag and increase fluorescence of the upstream reporter protein proportional to the proteolytic activity on the target substrate. Although, cells expressing YPet-DegK displayed reduced yellow fluorescence compared to cells expressing YPet, co-expression of active hK7 with an optimal hK7 substrate inserted between YPet and DegK did not rescue yellow fluorescence. In contrast, active hK7 co-expression with the YPet-substrate-DegK constructs led to reduced yellow fluorescence compared to co-expression

with an inactive hK7 variant. One explanation for this result is that active hK7 expression triggers the unfolded protein response in yeast which may lead to enhanced ubiquitination of the degron tag and enhanced proteasome activity¹⁷⁴. This screening system may therefore be more amenable to proteases which are easier to express and do not require post-translational modifications such as the tobacco etch virus (TEV) protease.

5.4. Materials and methods

5.4.1. Construction, expression and purification of YPet substrate insertions

Splice overlap extension PCR¹⁵¹ was used to clone four substrates (PRVMYYT, GGSGPRVMYYTGGSG, GGSGSGGS, and GGSGGGSGSGGGSG) between residues Y145 and N146 of YPet¹³². PCR products were cloned between the SfiI sites of the vector pBAD33²¹⁶, which carries an arabinose promoter for bacterial expression. YPet contains a C-terminal histidine tag (HHHHHH) for downstream purification. The *E. coli* strain MC1061 was transformed with the pBAD33 plasmid containing the YPet constructs and transformed colonies were selected on plates containing chloramphenicol (30 µg/mL). Plasmid sequencing (Genewiz) of transformed colonies verified substrate insertions into YPet.

To determine if substrates inserted into YPet were susceptible to proteolysis, each of the four constructs was expressed in bacteria and purified for fluorimetry analysis after protease addition. Cells carrying the sequenced plasmids were grown for 20 hours at 37°C in 5 mL Luria broth (LB) media with chloramphenicol (30 µg/mL). Cells were subcultured 1:100 into 80 mL LB with chloramphenicol and grown for 2 hours at 37°C until cell growth reached an optical density of 0.6. Cells were then induced with a final

concentration of 0.02% arabinose and expression was carried out for 16 hours at room temperature. After expression, cells were spun down at 10,000g for 10 minutes and media was decanted. Samples were then frozen at -80°C for 2 hours to improve cell lysis and soluble lysate extraction. Cell pellets were lysed by adding 5 mL B-PER II reagent (Thermo Scientific) and pipetting until the mixture was homogenous. Lysis mixtures were gently shook at room temperature for 30 minutes. Soluble lysate was isolated by centrifugation at 20,000g for 25 minutes.

To isolate purified YPet substrates, lysate was incubated with 500 µL Ni-NTA resin (Qiagen) for 1 hour at 4°C. Supernatant was flowed through a gravity flow column and resin was washed twice with 6 mL wash buffer (50 mM NaH₂PO₄, 300 mM NaCl, 20 mM imidazole, pH = 8.0). YPet substrates were eluted into 1 mL of elution buffer (50 mM NaH₂PO₄, 300mM NaCl, 250 mM imidazole, pH = 8.0). Recombinant hK7 (R&D Systems) was incubated with the YPet substrates (240 nM) and conversion of the substrate was monitored by measuring the yellow fluorescence emission at 527 nm every 5 min upon excitation at 488 nm for 2.5 hours at 37°C using a Tecan Infinite 200 PRO spectrophotometer (Tecan).

5.4.2. *Addition of a degradation tag to yeast expressed YPet*

To develop a cell based assay with a non-target counterselection step, we tested if a yeast degradation signal fused to the C-terminus of YPet would lead to a loss in yellow fluorescence. We also placed a protease susceptible substrate between YPet and the degradation signal, to determine if co-expression with active hK7 could rescue yellow fluorescence visualized by flow cytometry. The degradation signal

ACKNWFSSLSHFVIHL (DegK) was fused to the C-terminus of YPet containing an N-terminal invertase signal sequence (MLLQAFLFLLAGFAAKISA) for expression in the ER of yeast. Additionally four constructs were made containing the substrate sequences KLVFFAED, KLVYYAED, PRVMYYT, or GGSGSGGS flanked by YPet at the N-terminus and the degradation signal at the C-terminus. All inserts were created using sequential PCR steps using the YPet gene as a template. The resulting full-length YPet reporter fragments were cloned between the HindIII and PmeI sites in the pYES3/CT plasmid (Invitrogen) and transformed into DH5 α *E. coli* as described in Chapter 2. Plasmid DNA was recovered from bacteria and digested with Bsu361 to linearize the plasmid for integration into the genome of the yeast strain JYL69 as described in Chapter 2.

The fluorescence profiles of individual clones expressing YPet-DegK or YPet-substrate-DegK were assessed via flow cytometry. Cells were cultured overnight in SD-Trp media with 2% raffinose to an OD₆₀₀ of 4. Cells were then subcultured to an OD₆₀₀ of 0.75 in fresh SD-Trp media with 2% raffinose and 2% galactose for expression at various temperatures (22°C or 30°C) and times (0-6 hours). After expression, the cells were pelleted by centrifugation at 5,000g for 5 minutes and resuspended in 1x PBS for analysis on a FACSaria flow cytometer (BD Biosciences). Cells were analyzed with blue light excitation (488 nm) to determine the yellow (530/30 nm) fluorescence intensity. Clones with loss of YPet fluorescence due to degradation by the proteasome were selected for co-transformation with the wild-type hK7 plasmid and inactive hK7 plasmid.

6. Conclusions

6.1. Perspectives

6.1.1. *Development of a cell-based screen for evolving human proteases*

The catalytic turnover of proteins by proteases presents unique opportunities to exploit in therapy, which could reduce the dosage and cost of treatment compared to antibodies which bind stoichiometrically to their targets⁴. However, the lack of protease substrate specificity has hindered their therapeutic use since unregulated proteolysis would lead to adverse side effects. Additionally, neutralization by inhibitors found in tissues and serum limits the activity and stability of proteases *in vivo* and can reduce their half-life to on the order of minutes⁵². Currently no general methods have been developed and applied to evolving the specificity and stability of human proteases for their expanded use in the clinic. In an effort to fill this need, we developed a high-throughput, general method for screening large libraries of human secreted proteases for activity on a fluorogenic substrate to redesign their specificity for therapeutic targets.

The reasoning for protease template selection is often omitted from reports, however it is important to select an optimal parent protease based on its downstream application²¹⁷. Additionally, information on the protease crystal structure and the protease substrate specificity profile is critical in guiding directed evolution experiments. We chose to focus our efforts on engineering human secreted proteases since these proteases would be less immunogenic in downstream therapeutic applications compared to bacterial or viral proteases. In addition, common to human secreted proteases are post-translational modifications such as glycosylation and multiple disulfide bonds which stabilize the active form of the molecule in the extracellular environment. As a model system, we chose to

engineer the secreted protease human kallikrein 7 (hK7). hK7 was selected because of its open active site which made it amenable to screening with larger substrates such as the FRET reporter substrate used in these studies¹¹⁴. Additionally, hK7 displays basal levels of activity toward a therapeutic target of interest- the amyloid beta (A β) peptide, which provided a significant starting point in improving its activity and stability rather than designing a protease *de novo*¹⁰³.

Engineering human secreted proteases poses additional obstacles compared to bacterial or viral proteases since multiple post-translational modifications such as disulfide bonds, glycosylation, and correct processing of the N-terminal signal peptide are required for proper folding and activity. In developing a high-throughput screen, it is crucial that the protease be produced in an active form to expedite analysis and characterization. hK7 expression in bacteria resulted in insoluble aggregates due to the mispairing of cysteines that form six disulfide bonds in active hK7. However yeast, which has the machinery to form complex disulfide bonds, was found to be a suitable host for active hK7 production. In addition, yeast offered the advantages of faster growth rate and higher transformation efficiency compared to mammalian cells, making it an ideal cell type to build a screen for protease engineering.

To screen through large libraries of hK7 variants using FACS, protease activity was linked to fluorescence output through the use of FRET reporter substrates^{58,132}. Co-expression of the fluorogenic substrate and protease variant in the ER of yeast was essential for the formation of hK7's six disulfide bonds and to increase the local concentration of protease and substrate for catalysis to occur. However, overexpression of hK7 activity in yeast led to degradation of essential host cell proteins resulting in toxicity

to the yeast. The toxicity associated with active protease expression presented a unique challenge in engineering hK7's specificity and activity compared to other protease targets such as the tobacco etch virus (TEV) protease which is relatively non-toxic. It was necessary to reduce hK7 toxicity to the cell without eliminating the fluorescence signal correlating to intracellular activity. Reducing the time and temperature of co-expression alleviated most of the cellular toxicity associated with active protease expression and allowed the successful isolation of active variants during screening.

6.1.2. Application of PrECISE to randomly and rationally mutated hK7 libraries

The success of any directed evolution experiment not only requires the development of a robust screening methodology for the activity of interest, but is also influenced by the quality of the protease library generated. Therefore we compared two methods of diversification for generating hK7 libraries to determine the advantages and drawbacks of random mutagenesis and site-saturation mutagenesis in evolving human proteases using PrECISE. Low error-rate mutagenesis was employed to reduce the fraction of inactive variants and to incrementally screen for beneficial substitutions within hK7 in an iterative process as has been suggested¹⁴⁰. The first round of random mutagenesis allowed enrichment of hK7 variants with improved cell lysate activity, whereas a second round of random mutagenesis enabled isolation of an hK7 variant with improved selectivity toward A β 8. This result suggested that multiple rounds of mutagenesis and screening were required to isolate variants with changes in selectivity since random single substitutions on their own likely did not drastically change protease specificity. The hK7 variant with improved selectivity for A β 8 was found to have seven substitutions, with the majority of

these substitutions being distant from the active site. Therefore random mutagenesis enabled the acquisition of beneficial substitutions within hK7 at unusual sites. Further, our results indicated that these distant substitutions work cooperatively with the active site substitution G233V to enable large changes in selectivity of the hK7-2.7 variant for phenylalanine (F) at P1 over tyrosine (Y). Interestingly, the switch in selectivity of variant hK7-2.7 was accompanied by reduced toxicity toward mammalian cells and improved resistance to wild-type inhibitors, which further confirmed rearrangements to the active site⁵⁸.

While random mutagenesis has the advantage of not requiring crystal structure information, it often misses substitutions localized to the active site that operate synergistically since the probability that two random substitutions will occur near each other is rare. Therefore, we additionally screened a site-saturation library of five residues within the S1 pocket of hK7 for improved activity toward F over Y at P1 to determine if multiple substitutions within this subsite could cooperate favorably to switch selectivity. To reduce the fraction of inactive variants within the naïve library, four of the five sites were partially saturated by predicting which residues would be tolerated due to similar side chain chemistry to the wild-type residue. However, a high portion of the naïve library was found to be inactive and multiple rounds of screening for hK7 activity on PRVMFFT were required to see enrichment by FACS. Isolated variants from the final rounds of screening were found to have reduced activity on the selection substrate compared to wild-type, without large changes in selectivity indicating that the majority of substitutions to the hK7 S1 pocket were deleterious. Overall, low-error rate random mutagenesis and screening was more effective in introducing beneficial substitutions since biasing mutations to the active

site of hK7 dramatically reduced activity. Additionally, variants with improved activity in the rationally designed library were likely at too low of a frequency to become enriched with the PrECISE screen.

6.1.3. Incorporation of a non-target substrate within the PrECISE screen

Although the PrECISE screen only profiles protease cleavage on a selection substrate, the toxicity associated with non-specific intracellular hK7 activity was exploited to enrich variants with improved growth rates as a result of narrowed specificity. This strategy enabled isolation of an hK7 variant with improved selectivity and growth. However, the addition of a counter-selection substrate to the PrECISE method could greatly improve the probability of isolating selective variants by removing proteases which retain activity on a preferred wild-type substrate. Screening randomly mutated and rationally designed libraries of hK7 often gave back the wild-type parent protease, which could have been eliminated using negative selection on a wild-type substrate.

However, addition of a counterselection substrate to PrECISE was a non-trivial task since the availability of fluorescent reporters which do not overlap with the FRET reporter used in these studies is limited. Therefore, we chose to test incorporation of a non-target substrate within a surface exposed loop of YPet. While position Y145 within YPet tolerated peptide insertions, addition of hK7 to the YPet substrate *in vitro* did not reduce yellow fluorescence indicating that the inserted substrate was not accessible to proteolysis or cleavage did not destabilize the protein. Another tested strategy for non-target substrate incorporation was attachment of degradation signals to the fluorescent protein substrates. While, the degradation signal chosen for this assay was effective in degrading an N-

terminal YPet-substrate fusion partner after expression, hK7 co-expression with the short-lived fluorescent substrate did not rescue fluorescence. Future work in adding a non-target substrate to the screen could test tethering a yeast transcription factor to a ER integral membrane protein where cleavage of a non-target substrate linker releases the transcription factor to turn on expression of a reporter gene.

6.2. Future directions

While proteases present a unique opportunity to exploit their catalytic properties in therapeutic and biotechnological applications, the toxicity commonly associated with protease expression is a major barrier to evolving proteases with novel specificities and activities^{57,58}. Cell death cause by non-specific intracellular protease activity prohibits proper isolation of active variants during screening and additionally hinders downstream characterization of the properties of protease variants. For this reason, tobacco etch virus protease (TEVp) has been used as a model for the vast majority of protease engineering screens due to its high level of characterization, ease of expression in different host platforms, and naturally narrow substrate specificity, which greatly minimizes toxic expression effects^{74,75,218}. However, TEVp lacks potential for use in therapy since it is of non-human origin. Therefore, future efforts in developing screens for protease engineering should focus on the more challenging problem of evolving proteases of human origin.

Most proteases of human origin have complicated folds and multiple post-translational modifications such as glycosylation and disulfide bonds required for proper activity. Therefore, production of active protease would require expression in the secretory pathway of eukaryotic hosts. Additionally, many human proteases are non-selective; often they

cleave essential host cell proteins and are toxic when expressed. Therefore it would be critical to determine assay conditions that allow sufficient protease activity for screening, while minimizing host cell toxicity. Current methods developed for TEVp would need to be optimized to become extendable for human proteases. Alternatively, recent developments in the screening of free enzymes using emulsions could circumvent toxicity issues associated with intracellular protease expression and may provide an advantageous route for protease evolution²¹⁹.

Protease engineering methods that solely incorporate positive selection on a target substrate have shown success in narrowing protease selectivity by constantly challenging the protease to form stronger interactions with the selection substrate^{57,58,65}. However, incorporation of a counterselection substrate within the PrECISE methodology would be highly beneficial in isolating variants with selectivity changes directed at specific residues within the substrate. This precise control over selectivity switches is critical in the development of human protease therapeutics that display a high level of specificity for residues flanking the cleavage site. The development of screens where protease activity can be interrogated on multiple non-target substrates simultaneously will further increase the rate at which variants with narrow specificities can be identified. Additionally, among current high-throughput methods for protease engineering, only linear peptide targets have been used as selection substrates rather than fully folded target proteins. Because substrate conformation and accessibility plays an important role in determining the level of protease activity, the incorporation of fully folded target proteins within screens will aid in the development of effective protease therapies.

6.3. Overall conclusions

Most naturally occurring proteases lack the specificity and stability to be directly translated for therapeutic use. A key contribution of this work was the development of a novel cell-based screen for evolving human protease specificity toward therapeutic targets. To the best of our knowledge, this is the first application of using random mutagenesis and high-throughput screening to redesign human protease specificity. Screening methodologies for the directed evolution of protease activity and specificity hold strong promise over traditional methods of structure-guided rational design, since variants with novel activities may require multiple substitutions that would be impossible to predict *a priori*. Random mutagenesis provided an unbiased diversification approach to incrementally introduce a small number of substitutions throughout the hK7 structure. Multiple round of screening using PrECISE allowed isolation of beneficial substitutions in hK7, leading to improved activity and selectivity for A β 8. Within our studies, we found that multiple substitutions were required to alter hK7's specificity for A β . These substitutions were scattered throughout the protease, with the majority being distant from the active site and therefore impossible to predict. Analysis of the crystal structure of an improved variant assisted in hypothesizing a potential mechanism within hK7 that was responsible for changes in selectivity, which was confirmed using site-directed mutagenesis. The PrECISE screening methodology and techniques for characterizing isolated variants developed here can be generalized for the directed evolution of other human secreted proteases that require post-translational modifications and are toxic to the host during expression. In summary, harnessing the catalytic properties of human proteases provides an attractive alternative to antibodies for the irreversible clearance of toxic

proteins and peptides implicated in disease. Future advances in the development and optimization of screens for evolving the activity, selectivity, and stability of human proteases will undoubtedly bolster their application as therapeutics.

7. Appendix

7.1. Appendix A: Expression of hK7 from the CUP1 yeast promoter

7.1.1. Introduction

To decouple expression of the protease and substrate from the GAL1 promoter within the PrECISE screen, the gene encoding hK7 was cloned into a yeast centromeric vector containing the weaker CUP1 promoter. Placing expression of the protease under a separate regulated promoter could facilitate optimizing the concentration of protease and substrate in the ER of yeast for proteolysis by individually tuning their expression. The CUP1 promoter is one of the few regulated promoters in yeast and is inducible by addition of copper ions in the form of CuSO_4 to the media. Since low micromolar amounts of Cu^{2+} can inhibit the enzymatic activity of hK7, a copper resistant variant, hK7 H99F, was cloned into the CUP1 vector¹¹⁴. From studies of the hK7 crystal structure, the histidine at position 99 was determined to be the structural basis for Cu^{2+} inhibition of hK7¹¹⁴.

7.1.2. Results and Discussion

To determine if the H99F substitution impacted activity of hK7, the hK7 H99F variant was first cloned under the GAL1 promoter and activity was measured in cell lysates after expression. Activity of the hK7 H99F variant was unchanged from wild-type hK7 indicating that the mutation did not affect hK7 enzymatic activity. We next tested expression of the hK7 H99F variant from the CUP1 promoter for various amounts of time (1-16 hours) and with different copper induction concentrations (1 μM , 10 μM , 100 μM , and 1 mM) at 22°C and 30°C in minimal media supplemented with either raffinose or glucose. Protease activity was not detectable for expression of hK7 H99F in minimal

media containing raffinose. When the expression media was replaced with minimal media supplemented with 2% glucose and hK7 was expressed from the CUP1 promoter for 4 hours, hK7 activity was detected in cell lysates with the FRET PRVMYYT reporter (Figure 7.1). Yeast grow faster in glucose containing media compared to raffinose and therefore likely diverted more energy toward hK7 production in glucose media.

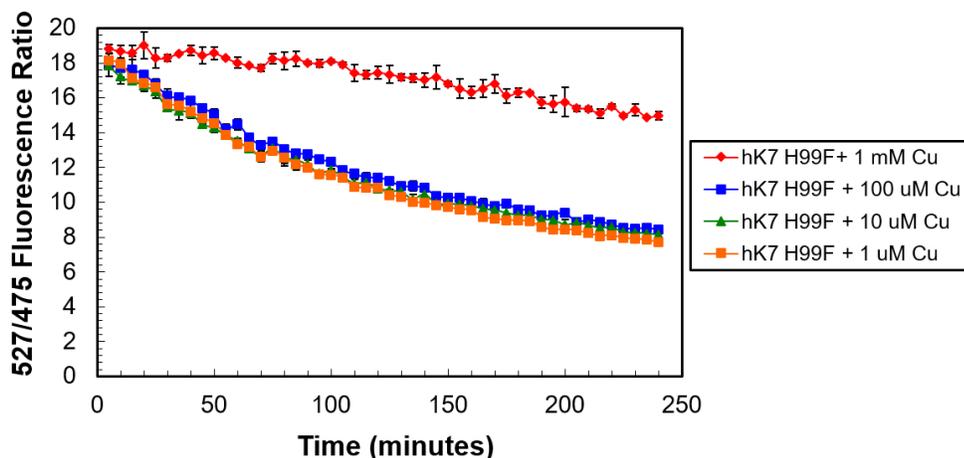


Figure 7.1. hK7 expression from the CUP1 promoter is measurable in yeast lysate.

The copper resistant variant hK7 H99F was expressed under the CUP1 promoter for 4 hours at 22°C in glucose containing media with various concentrations of copper for induction. Activity of hK7 H99F was then measured in cell lysate using the PRVMYYT FRET reporter by monitoring the change in yellow fluorescence/cyan fluorescence over time with fluorimetry. Induction with the highest concentration of copper tested (1 mM) gave a weak signal for activity, whereas lower concentrations of copper (1-100 μ M) were effective in producing higher quantities of active hK7.

Expression of the protease in glucose media would inhibit simultaneous production of the FRET reporter substrate from the GAL1 promoter in the PrECISE screen. Therefore, we tested if expression of the FRET reporter substrate prior to hK7 H99F expression could allow for intracellular detection of hK7 activity from the CUP1 promoter. Producing the FRET reporter prior to induction of hK7 allows more time for the probe to fold prior to

protease expression. Also, by halting FRET probe production, the cell can divert more of its energy to producing folded hK7. The FRET reporter containing the optimized hK7 substrate PRVMYYT under the GAL1 promoter was expressed for 6 hours at 22°C in minimal media with 2% galactose. After expression of the FRET probe, cells displayed high yellow fluorescence with minimal cyan fluorescence (Figure 7.2). Cells were then spun down and the media was removed and replaced with minimal media containing 2% glucose and 100 μ M CuSO₄ for hK7 H99F expression. Cells were analyzed using flow cytometry after 2 and 4 hours to measure the change in FRET. After 4 hours of protease expression, no shift in FRET was detected and the amount of FRET probe was dramatically reduced since the addition of glucose inhibited further production of the fluorescent substrate (Figure 7.2).

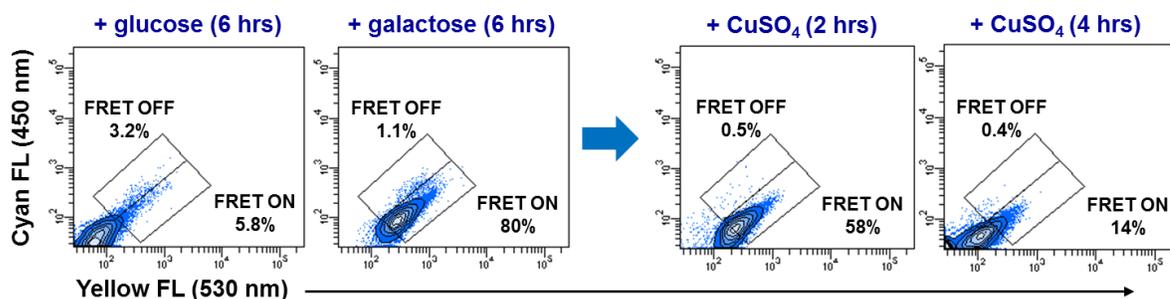


Figure 7.2. hK7 expression *in vivo* is not detectable with the CUP1 promoter.

Flow cytometry analysis of sequential expression of the PRVMYYT FRET substrate from the GAL1 promoter at 22°C for 6 hours in minimal media with galactose prior to induction of hK7 H99F from the CUP1 promoter for 2 or 4 hours at 22°C in minimal media with glucose and 100 μ M CuSO₄. Expression of hK7 H99F under the CUP1 promoter did not cause a shift in the cyan/yellow fluorescence visualized by FACS indicating minimal intracellular cleavage of the FRET reporter.

Because no hK7 H99F activity was detectable *in vivo* under the CUP1 promoter, but hK7 activity had been previously visualized *in vivo* under the GAL1 promoter, we decided to characterize hK7 expression levels from both promoters using a Western blot. hK7 was expressed under the GAL1 promoter for 4 hours at 22°C with 2% galactose in minimal media. Induction from the CUP1 promoter has been reported to be quite rapid^{220,221}. Therefore, hK7 H99F was similarly expressed under the CUP1 promoter for 4 hours at 22°C in minimal media with 2% glucose. Cells were collected and lysed after expression and lysates were run on an SDS-PAGE gel and transferred to a membrane. As a negative control, the cell lysate of the parent yeast strain JYL69 was also ran on the gel. Samples were incubated with a polyclonal antibody against human kallikrein 7 and a secondary antibody conjugated to an Alexa Fluor 680 dye. After imaging for fluorescence, hK7 expression was detected under the GAL1 promoter in the yeast cell lysate, but no expression was seen under the CUP1 promoter (Figure 7.3). The cell lysate of the parent strain, JYL69, also did not show a band for hK7 expression as expected (Figure 7.3). Since no hK7 expression could be detected in the cell lysate after expression from the CUP1 promoter, all samples were loaded onto a nickel column to purify and concentrate any His-tagged hK7 in the lysate. The purified samples were then run on a Western blot to probe for hK7 expression. The band for hK7 expressed under the GAL1 promoter became stronger and more defined after purification indicating that the protein had been successfully concentrated (Figure 7.3). A faint band for hK7 appeared for the purified sample expressed under the CUP1 promoter which suggested that hK7 H99F had been expressed, but the expression level was much less than that obtained under the GAL1 promoter (Figure 7.3).

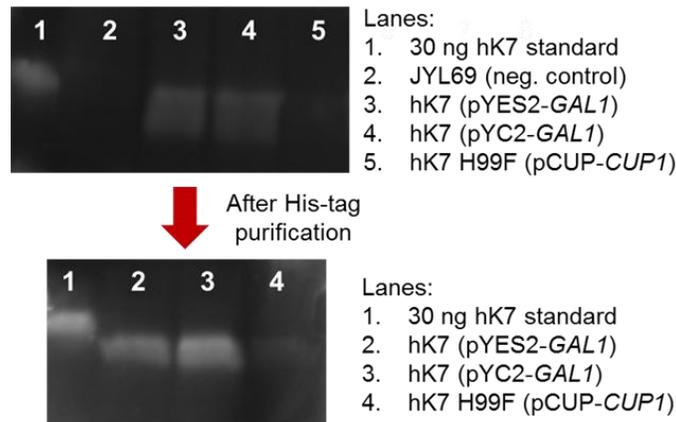


Figure 7.3. Comparison of hK7 expression levels under the CUP1 and GAL1 promoters.

Western blot analysis for hK7 expression in yeast cell lysate and after purification after expression from the GAL1 promoter in an integrated vector (pYES2) or a centromeric vector (pYC2) compared to expression under the CUP1 promoter from a centromeric vector (pCUP). As a negative control, untransformed yeast cells (strain JYL69) were probed for the presence of hK7 with a polyclonal anti-hK7 antibody. Expression of hK7 was detectable in cell lysate for expression under the GAL1 promoter but not under the CUP1 promoter (top blot). Purification caused an increase in intensity of the hK7 band for GAL1 promoter expression and the appearance of a faint band for CUP1 promoter expression (bottom blot).

7.1.3. Methods

The gene encoding mature human kallikrein 7 (hK7) with an N-terminal invertase secretion signal (MLLQAFLFLLAGFAAKISA) and a C-terminal 6x histidine (GHHHHHH) tag was amplified for insertion into pCu416CUP1 (pCUP). Copper resistant hK7 variant H99F was created using PCR-driven overlap extension¹⁵¹ with forward primer 5' TACTCC ACACAGACCTTTGTTAATGACCTCATG and reverse primer 5' CATGAGGTCATT AACAAAGGTCTGTGTGGAGTA. The resulting fragments were cloned between the SmaI and ClaI sites in the pCUP plasmid (Addgene), which contains a CUP1 promoter for copper-inducible expression. DH5 α *E. coli* were transformed by electroporation with the ligated vector and cells were plated on media supplemented with

ampicillin (100 µg/mL). Plasmid DNA was isolated using a Zyppy plasmid miniprep kit (Zymo Research) and the yeast strain JYL69 was transformed with 1 µg of the hK7 H99F plasmid using the high efficiency PEG/lithium acetate method¹⁸⁵ as described in Chapter 2.

Expression of hK7 H99F from pCUP was analyzed by culturing cells overnight in 10 mL SD-Ura media with 2% glucose to an OD₆₀₀ of 4. Cells were then subcultured to an OD₆₀₀ of 0.75 in 40 mL fresh SD-Ura media with 2% glucose with 1 µM – 1 mM CuSO₄ for induction. The expression temperature (22°C, 30°C) and time (1-16 hours) were varied to optimize production of active hK7. After expression, cell pellets were isolated by centrifugation at 5000g for 10 minutes. Cells were lysed by adding 3 mL Y-PER reagent (Thermo Scientific) and gently shaking the homogenous mixture at room temperature for 30 minutes. Soluble lysate fractions were recovered by spinning the solution at 15,000g for 5 minutes. Western blot analysis and activity in cell lysate was assessed as described in Chapter 2. Co-expression with the PRVMYYT FRET reporter was determined via FACS as stated in Chapter 2.

8. References

- (1) Di Cera, E. (2007) Thrombin as procoagulant and anticoagulant. *J. Thromb. Haemost.* 5, 196–202.
- (2) Rijken, D. C., and Lijnen, H. R. (2009) New insights into the molecular mechanisms of the fibrinolytic system. *J. Thromb. Haemost.* 7, 4–13.
- (3) Mannucci, P. M. (2008) Back to the future: A recent history of haemophilia treatment. *Haemophilia* 14, 10–18.
- (4) Craik, C. S., Page, M. J., and Madison, E. L. (2011) Proteases as therapeutics. *Biochem. J.* 435, 1–16.
- (5) Tanzi, R. E., and Bertram, L. (2005) Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective. *Cell* 120, 545–555.
- (6) Kageyama, T. (2002) Pepsinogens, progastricsins, and prochymosins: structure, function, evolution, and development. *Cell. Mol. Life Sci.* 59, 288–306.
- (7) Pham, C. T. N. (2006) Neutrophil serine proteases: specific regulators of inflammation. *Nat. Rev. Immunol.* 6, 541–550.
- (8) Chambers, R. C., and Laurent, G. J. (2002) Coagulation cascade proteases and tissue fibrosis. *Biochem. Soc. Trans.* 30, 194–200.
- (9) Mcilwain, D. R., Berger, T., and Mak, T. W. (2013) Caspase functions in cell death and disease. *Cold Spring Harb. Perspect. Biol.* 5:a008656, 1–28.
- (10) Puente, X. S., Sánchez, L. M., Gutiérrez-Fernández, A., Velasco, G., and López-Otín, C. (2005) A genomic view of the complexity of mammalian proteolytic systems. *Biochem. Soc. Trans.* 33, 331–334.
- (11) Turk, B. (2006) Targeting proteases: successes, failures and future prospects. *Nat. Rev. Drug Discov.* 5, 785–799.
- (12) Schechter, I., and Berger, a. (1967) On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 27, 157–162.
- (13) Huber, R., and Bode, W. (1978) Structural basis of the activation and action of trypsin. *Acc. Chem. Res.* 11, 114–122.
- (14) Gron, H., and Breddam, K. (1992) Interdependency of the binding subsites in subtilisin. *Biochemistry* 31, 8967–8971.
- (15) Anindya, R., and Savithri, H. S. (2004) Potyviral NIa proteinase, a proteinase with novel deoxyribonuclease activity. *J. Biol. Chem.* 279, 32159–32169.

- (16) Phan, J., Zdanov, A., Evdokimov, A. G., Tropea, J. E., Peters, H. K., Kapust, R. B., Li, M., Wlodawer, A., and Waugh, D. S. (2002) Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem.* 277, 50564–50572.
- (17) Krem, M. M., and Cera, E. Di. (2002) Evolution of enzyme cascades from embryonic development to blood coagulation. *Trends Biochem. Sci.* 27, 67–74.
- (18) Deperthes, D. (2002) Phage display substrate: A blind method for determining protease specificity. *Biol. Chem.* 383, 1107–1112.
- (19) Diamond, S. L. (2007) Methods for mapping protease specificity. *Curr. Opin. Chem. Biol.* 11, 46–51.
- (20) Boulware, K., and Daugherty, P. (2006) Protease specificity determination by using cellular libraries of peptide substrates (CLiPS). *Proc. Natl. Acad. Sci. U. S. A.* 103, 7583–7588.
- (21) Boulware, K. T., Jabaiah, A., and Daugherty, P. S. (2010) Evolutionary optimization of peptide substrates for proteases that exhibit rapid hydrolysis kinetics. *Biotechnol. Bioeng.* 106, 339–346.
- (22) Cornille, F., Martin, L., Lenoir, C., Cussac, D., Roques, B. P., and Fournié-Zaluski, M.-C. (1997) Cooperative exosite-dependent cleavage of synaptobrevin by tetanus toxin light chain. *J. Biol. Chem.* 272, 3459–3464.
- (23) Bock, P. E., Panizzi, P., and Verhamme, I. M. A. (2007) Exosites in the substrate specificity of blood coagulation reactions. *J. Thromb. Haemost.* 5, 81–94.
- (24) Di Cera, E. (2008) Thrombin. *Mol. Aspects Med.* 29, 203–254.
- (25) Vu, T.-K. H., Hung, D. T., Wheaton, V. I., and Coughlin, S. R. (1991) Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation. *Cell* 64, 1057–1068.
- (26) Lutgens, S. P. M., Cleutjens, K. B. J. M., Daemen, M. J. A. P., and Heeneman, S. (2007) Cathepsin cysteine proteases in cardiovascular disease. *FASEB J.* 21, 3029–3041.
- (27) Yasuda, Y., Kaleta, J., and Brömme, D. (2005) The role of cathepsins in osteoporosis and arthritis: rationale for the design of new therapeutics. *Adv. Drug Deliv. Rev.* 57, 973–993.
- (28) Koblinski, J. E., Ahram, M., and Sloane, B. F. (2000) Unraveling the role of proteases in cancer. *Clin. Chim. Acta* 291, 113–135.
- (29) Coussens, L. M., Fingleton, B., and Matrisian, L. M. (2002) Matrix metalloproteinase inhibitors and cancer: trials and tribulations. *Science* 295, 2387–2392.
- (30) Vardy, E. R. L. C., Catto, A. J., and Hooper, N. M. (2005) Proteolytic mechanisms in amyloid- β metabolism: therapeutic implications for Alzheimer's disease. *Trends Mol. Med.* 11, 464–472.

- (31) Rawlings, N. D., Tolle, D. P., and Barrett, A. J. (2004) Evolutionary families of peptidase inhibitors. *Biochem. J.* 378, 705–716.
- (32) Khan, A. R., and James, M. N. G. (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.* 7, 815–836.
- (33) Hedstrom, L. (2002) Serine protease mechanism and specificity. *Chem. Rev.* 102, 4501–4523.
- (34) Bode, W., and Huber, R. (1976) Induction of the bovine trypsinogen-trypsin transition by peptides sequentially similar to the N-terminus of trypsin. *FEBS Lett.* 68, 231–236.
- (35) Landgraf, K. E., Steffek, M., Quan, C., Tom, J., Yu, C., Santell, L., Maun, H. R., Eigenbrot, C., and Lazarus, R. A. (2014) An allosteric switch for pro-HGF/Met signaling using zymogen activator peptides. *Nat. Chem. Biol.* 10, 567–573.
- (36) Slee, E. A., Harte, M. T., Kluck, R. M., Wolf, B. B., Casiano, C. A., Newmeyer, D. D., Wang, H.-G., Reed, J. C., Nicholson, D. W., Alnemri, E. S., Green, D. R., and Martin, S. J. (1999) Ordering the cytochrome c-initiated caspase cascade: hierarchical activation of caspases-2,-3,-6,-7,-8, and -10 in a caspase-9-dependent manner. *J. Cell Biol.* 144, 281–292.
- (37) Salvesen, G. S., and Dixit, V. M. (1999) Caspase activation: the induced-proximity model. *Proc. Natl. Acad. Sci. U. S. A.* 96, 10964–10967.
- (38) Luo, L.-Y., and Jiang, W. (2006) Inhibition profiles of human tissue kallikreins by serine protease inhibitors. *Biol. Chem.* 387, 813–816.
- (39) Schechter, N. M., and Plotnick, M. I. (2004) Measurement of the kinetic parameters mediating protease-serpin inhibition. *Methods* 32, 159–168.
- (40) Travis, J., and Salvesen, G. S. (1983) Human plasma proteinase inhibitors. *Annu. Rev. Biochem.* 52, 655–709.
- (41) De Clercq, E. (2009) The history of antiretrovirals: key discoveries over the past 25 years. *Rev. Med. Virol.* 19, 287–299.
- (42) Tan, J. J., Cong, X. J., Hu, L. M., Wang, C. X., Jia, L., and Liang, X.-J. (2010) Therapeutic strategies underpinning the development of novel techniques for the treatment of HIV infection. *Drug Discov. Today* 15, 186–197.
- (43) Grabar, S., Pradier, C., Le Corfec, E., Lancar, R., Allavena, C., Bentata, M., Berlureau, P., Dupont, C., Fabbro-Peray, P., Poizot-Martin, I., and Costagliola, D. (2000) Factors associated with clinical and virological failure in patients receiving a triple therapy including a protease inhibitor. *AIDS* 14, 141–149.
- (44) Nilsson, I. M., Berntorp, E., Löfqvist, T., and Pettersson, H. (1992) Twenty-five years' experience of prophylactic treatment in severe haemophilia A and B. *J. Intern. Med.* 232, 25–32.

- (45) Gitschier, J., Drayna, D., Tuddenham, E. G., White, R. L., and Lawn, R. M. (1985) Genetic mapping and diagnosis of haemophilia A achieved through a BclI polymorphism in the factor VIII gene. *Nature* 314, 738–740.
- (46) Hedner, U. (1990) Factor VIIa in the treatment of haemophilia. *Blood Coagul. Fibrinolysis* 1, 307–317.
- (47) Littlewood, J. M., Wolfe, S. P., and Conway, S. P. (2006) Diagnosis and treatment of intestinal malabsorption in cystic fibrosis. *Pediatr. Pulmonol.* 41, 35–49.
- (48) Esmon, C. T., Fukudome, K., Mather, T., Bode, W., Regan, L. M., Stearns-Kurosawa, D. J., and Kurosawa, S. (1999) Inflammation, sepsis, and coagulation. *Haematologica* 84, 254–259.
- (49) Riewald, M., Petrovan, R. J., Donner, A., Mueller, B. M., and Ruf, W. (2002) Activation of endothelial cell protease activated receptor 1 by the protein C pathway. *Science* 296, 1880–1882.
- (50) Bernard, G. R., Ely, E. W., Wright, T. J., Fraiz, J., Stasek, J. E. J., Russell, J. A., Mayers, I., Rosenfeld, B. A., Morris, P. E., Yan, S. B., and Helterbrand, J. D. (2001) Safety and dose relationship of recombinant human activated protein C for coagulopathy in severe sepsis. *Crit. Care Med.* 29, 2051–2059.
- (51) Gentry, C. A., Gross, K. B., Sud, B., and Drevets, D. A. (2009) Adverse outcomes associated with the use of drotrecogin alfa (activated) in patients with severe sepsis and baseline bleeding precautions. *Crit. Care Med.* 37, 19–25.
- (52) Andreasen, P. A., Egelund, R., and Petersen, H. H. (2000) The plasminogen activation system in tumor growth, invasion, and metastasis. *Cell. Mol. Life Sci.* 57, 25–40.
- (53) Madison, E. L., Goldsmith, E. J., Gerard, R. D., Gething, M. J. H., and Sambrook, J. F. (1989) Serpin-resistant mutants of human tissue-type plasminogen activator. *Nature* 339, 721–724.
- (54) Van De Werf, F. J. (1999) The ideal fibrinolytic: can drug design improve clinical results? *Eur. Heart J.* 20, 1452–1458.
- (55) Semba, C. P., Sugimoto, K., and Razavi, M. K. (2001) Alteplase and tenecteplase: applications in the peripheral circulation. *Tech. Vasc. Interv. Radiol.* 4, 99–106.
- (56) Bhana, N., and Spencer, C. M. (2000) Lanoteplase. *BioDrugs* 13, 217–224.
- (57) O’Loughlin, T. L., Greene, D. N., and Matsumura, I. (2006) Diversification and specialization of HIV protease function during in vitro evolution. *Mol. Biol. Evol.* 23, 764–772.
- (58) Guerrero, J. L., O’Malley, M. A., and Daugherty, P. S. (2016) Intracellular FRET-based screen for redesigning the specificity of secreted proteases. *ACS Chem. Biol.* 11, 961–970.

- (59) 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., and Smith, J. W. (2014) Basis for substrate recognition and distinction by matrix metalloproteinases. *Proc. Natl. Acad. Sci. U. S. A.* *111*, E4148–E4155.
- (60) Hedstrom, L., Szilagyi, L., and Rutter, W. J. (1992) Converting trypsin to chymotrypsin: the role of surface loops. *Science* *255*, 1249–1253.
- (61) Marino, F., Pelc, L. A., Vogt, A., Gandhi, P. S., and Di Cera, E. (2010) Engineering thrombin for selective specificity toward protein C and PAR1. *J. Biol. Chem.* *285*, 19145–19152.
- (62) 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., and Andersson, C. (2014) Engineering neprilysin activity and specificity to create a novel therapeutic for Alzheimer’s disease. *PLoS One* *9*, e104001.
- (63) Spiller, B., Gershenson, A., Arnold, F. H., and Stevens, R. C. (1999) A structural view of evolutionary divergence. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 12305–12310.
- (64) Pogson, M., Georgiou, G., and Iverson, B. L. (2009) Engineering next generation proteases. *Curr. Opin. Biotechnol.* *20*, 390–397.
- (65) Sellamuthu, S., Shin, B. H., Lee, E. S., Rho, S.-H., Hwang, W., Lee, Y. J., Han, H.-E., Kim, J. Il, and Park, W. J. (2008) Engineering of protease variants exhibiting altered substrate specificity. *Biochem. Biophys. Res. Commun.* *371*, 122–126.
- (66) Sellamuthu, S., Shin, B. H., Han, H.-E., Park, S. M., Oh, H. J., Rho, S.-H., Lee, Y. J., and Park, W. J. (2011) An engineered viral protease exhibiting substrate specificity for a polyglutamine stretch prevents polyglutamine-induced neuronal cell death. *PLoS One* *6*, e22554.
- (67) Olsen, M. J., Stephens, D., Griffiths, D., Daugherty, P., Georgiou, G., and Iverson, B. L. (2000) Function-based isolation of novel enzymes from a large library. *Nat. Biotechnol.* *18*, 1071–1074.
- (68) Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S. M., Roodveldt, C., and Tawfik, D. S. (2005) The “evolvability” of promiscuous protein functions. *Nat. Genet.* *37*, 73–76.
- (69) Khersonsky, O., Roodveldt, C., and Tawfik, D. S. (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* *10*, 498–508.
- (70) Varadarajan, N., Gam, J., Olsen, M. J., Georgiou, G., and Iverson, B. L. (2005) Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 6855–6860.
- (71) Varadarajan, N., Rodriguez, S., Hwang, B.-Y., Georgiou, G., and Iverson, B. L. (2008) Highly active and selective endopeptidases with programmed substrate specificities. *Nat. Chem. Biol.* *4*, 290–294.

- (72) Varadarajan, N., Georgiou, G., and Iverson, B. L. (2008) An engineered protease that cleaves specifically after sulfated tyrosine. *Angew. Chemie* 120, 7979–7981.
- (73) Varadarajan, N., Pogson, M., Georgiou, G., and Iverson, B. L. (2009) Proteases that can distinguish among different post-translational forms of tyrosine engineered using multicolor flow cytometry. *J. Am. Chem. Soc.* 131, 18186–18190.
- (74) Carrico, Z. M., Strobel, K. L., Atreya, M. E., Clark, D. S., and Francis, M. B. (2016) Simultaneous selection and counter-selection for the directed evolution of proteases in *E. coli* using a cytoplasmic anchoring strategy. *Biotechnol. Bioeng.* 113, 1187–1193.
- (75) Yi, L., Gebhard, M. C., Li, Q., Taft, J. M., Georgiou, G., and 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, 7229–7234.
- (76) Esvelt, K. M., Carlson, J. C., and Liu, D. R. (2011) A system for the continuous directed evolution of biomolecules. *Nature* 472, 499–503.
- (77) Dickinson, B. C., Packer, M. S., Badran, A. H., and Liu, D. R. (2014) A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nat. Commun.* 5, 5352.
- (78) Hardy, J., and Selkoe, D. J. (2002) The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* 297, 353–356.
- (79) Wolfe, M. S., Xia, W., Ostaszewski, B. L., Diehl, T. S., Kimberly, W. T., and Selkoe, D. J. (1999) Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and gamma-secretase activity. *Nature* 398, 513–517.
- (80) Hussain, I., Powell, D., Howlett, D. R., Tew, D. G., Meek, T. D., Chapman, C., Gloger, I. S., Murphy, K. E., Southan, C. D., Ryan, D. M., Smith, T. S., Simmons, D. L., Walsh, F. S., Dingwall, C., and Christie, G. (1999) Identification of a novel aspartic protease (Asp 2) as beta-secretase. *Mol. Cell. Neurosci.* 14, 419–427.
- (81) Jarrett, J. T., Berger, E. P., and Lansbury, P. T. (1993) The carboxy terminus of the β amyloid protein is critical for the seeding of amyloid formation: implications for the pathogenesis of Alzheimer's disease. *Biochemistry* 32, 4693–4697.
- (82) Tjernberg, L. O., Callaway, D. J. E., Tjernberg, A., Hahne, S., Lilliehöök, C., Terenius, L., Thyberg, J., and Nordstedt, C. (1999) A molecular model of Alzheimer amyloid β -peptide fibril formation. *J. Biol. Chem.* 274, 12619–12625.
- (83) Balbach, J. J., Ishii, Y., Antzutkin, O. N., Leapman, R. D., Rizzo, N. W., Dyda, F., Reed, J., and Tycko, R. (2000) Amyloid fibril formation by A β 16-22, a seven-residue fragment of the Alzheimer's β -amyloid peptide, and structural characterization by solid state NMR. *Biochemistry* 39, 13748–13759.
- (84) Iwatsubo, T., Odaka, A., Suzuki, N., Mizusawa, H., Nukina, N., and Ihara, Y. (1994) Visualization of A β 42(43) and A β 40 in senile plaques with end-specific A β monoclonals: Evidence that an initially deposited species is A β 42(43). *Neuron* 13, 45–53.

- (85) Miners, J. S., Barua, N., Kehoe, P. G., Gill, S., and Love, S. (2011) A β -degrading enzymes: potential for treatment of Alzheimer disease. *J. Neuropathol. Exp. Neurol.* 70, 944–959.
- (86) Preston, S. D., Steart, P. V., Wilkinson, A., Nicoll, J. A. R., and Weller, R. O. (2003) Capillary and arterial cerebral amyloid angiopathy in Alzheimer's disease: defining the perivascular route for the elimination of amyloid β from the human brain. *Neuropathol. Appl. Neurobiol.* 29, 106–117.
- (87) Mawuenyega, K. G., Sigurdson, W., Ovod, V., Munsell, L., Kasten, T., Morris, J. C., Yarasheski, K. E., and Bateman, R. J. (2010) Decreased clearance of CNS β -amyloid in Alzheimer's disease. *Science* 330, 1774.
- (88) Ghosh, A. K., Brindisi, M., and Tang, J. (2012) Developing β -secretase inhibitors for treatment of Alzheimer's disease. *J. Neurochem.* 120, 71–83.
- (89) Sheridan, C. (2015) Pivotal trials for β -secretase inhibitors in Alzheimer's. *Nat. Biotechnol.* 33, 115–116.
- (90) Schenk, D. (2002) Amyloid- β immunotherapy for Alzheimer's disease: the end of the beginning. *Nat Rev Neurosci* 3, 824–828.
- (91) Wolfe, M. S. (2008) Gamma-secretase inhibition and modulation for Alzheimer's disease. *Curr. Alzheimer Res.* 5, 158–164.
- (92) Extance, A. (2010) Alzheimer's failure raises questions about disease-modifying strategies. *Nat. Rev. Drug Discov.* 9, 749–751.
- (93) Sperling, R., Salloway, S., Brooks, D. J., Tampieri, D., Barakos, J., Fox, N. C., Raskind, M., Sabbagh, M., Honig, L. S., Porsteinsson, A. P., Lieberburg, I., Arrighi, H. M., Morris, K. A., Lu, Y., Liu, E., Gregg, K. M., Brashear, H. R., Kinney, G. G., Black, R., and Grundman, M. (2012) Amyloid-related imaging abnormalities in patients with Alzheimer's disease treated with bapineuzumab: a retrospective analysis. *Lancet Neurol.* 11, 241–249.
- (94) Ratner, M. (2015) Biogen's early Alzheimer's data raise hopes, some eyebrows. *Nat. Biotechnol.* 33, 438.
- (95) Nalivaeva, N. N., Beckett, C., Belyaev, N. D., and Turner, A. J. (2012) Are amyloid-degrading enzymes viable therapeutic targets in Alzheimer's disease? *J. Neurochem.* 120, 167–185.
- (96) Iwata, N., Tsubuki, S., Takaki, Y., Shirotani, K., Lu, B., Gerard, N. P., Gerard, C., Hama, E., Lee, H. J., and Saido, T. C. (2001) Metabolic regulation of brain A β by neprilysin. *Science* 292, 1550–1552.
- (97) Eckman, E. A., Watson, M., Marlow, L., Sambamurti, K., and Eckman, C. B. (2003) Alzheimer's disease β -amyloid peptide is increased in mice deficient in endothelin-converting enzyme. *J. Biol. Chem.* 278, 2081–2084.

- (98) Farris, W., Mansourian, S., Chang, Y., Lindsley, L., Eckman, E. A., Frosch, M. P., Eckman, C. B., Tanzi, R. E., Selkoe, D. J., and Guenette, S. (2003) Insulin-degrading enzyme regulates the levels of insulin, amyloid β -protein, and the β -amyloid precursor protein intracellular domain in vivo. *Proc. Natl. Acad. Sci. U. S. A.* 100, 4162–4167.
- (99) Morelli, L., Llovera, R., Gonzalez, S. A., Affranchino, J. L., Prelli, F., Frangione, B., Ghiso, J., and Castaño, E. M. (2003) Differential degradation of amyloid β genetic variants associated with hereditary dementia or stroke by insulin-degrading enzyme. *J. Biol. Chem.* 278, 23221–23226.
- (100) Numata, K., and Kaplan, D. L. (2010) Mechanisms of enzymatic degradation of amyloid β microfibrils generating nanofilaments and nanospheres related to cytotoxicity. *Biochemistry* 49, 3254–3260.
- (101) Kanemitsu, H., Tomiyama, T., and Mori, H. (2003) Human neprilysin is capable of degrading amyloid β peptide not only in the monomeric form but also the pathological oligomeric form. *Neurosci. Lett.* 350, 113–116.
- (102) Tucker, H. M., Kihiko, M., Caldwell, J. N., Wright, S., Kawarabayashi, T., Price, D., Walker, D., Scheff, S., McGillis, J. P., Rydel, R. E., and Estus, S. (2000) The plasmin system is induced by and degrades amyloid- β aggregates. *J. Neurosci.* 20, 3937–3946.
- (103) Shropshire, T. D., Reifert, J., Rajagopalan, S., Baker, D., Feinstein, S. C., and Daugherty, P. S. (2014) Amyloid β peptide cleavage by kallikrein 7 attenuates fibril growth and rescues neurons from A β -mediated toxicity in vitro. *Biol. Chem.* 395, 109–118.
- (104) Crouch, P. J., Tew, D. J., Du, T., Nguyen, D. N., Caragounis, A., Filiz, G., Blake, R. E., Trounce, I. A., Soon, C. P. W., Laughton, K., Perez, K. A., Li, Q. X., Cherny, R. A., Masters, C. L., Barnham, K. J., and White, A. R. (2009) Restored degradation of the Alzheimer's amyloid- β peptide by targeting amyloid formation. *J. Neurochem.* 108, 1198–1207.
- (105) Mukherjee, A., Song, E., Kihiko-Ehmann, M., Goodman, J. P., Pyrek, J. S., Estus, S., and Hersh, L. B. (2000) Insulysin hydrolyzes amyloid β peptides to products that are neither neurotoxic nor deposit on amyloid plaques. *J. Neurosci.* 20, 8745–8749.
- (106) Hu, J., Igarashi, A., Kamata, M., and Nakagawa, H. (2001) Angiotensin-converting enzyme degrades Alzheimer amyloid β -peptide (A β); retards A β aggregation, deposition, fibril formation; and inhibits cytotoxicity. *J. Biol. Chem.* 276, 47863–47868.
- (107) Carlson, C., Estergard, W., Oh, J., Suhy, J., Jack, C. R., Siemers, E., and Barakos, J. (2011) Prevalence of asymptomatic vasogenic edema in pretreatment Alzheimer's disease study cohorts from phase 3 trials of semagacestat and solanezumab. *Alzheimer's Dement.* 7, 396–401.

- (108) Salloway, S., Sperling, R., Fox, N. C., Blennow, K., Klunk, W., Raskind, M., Sabbagh, M., Honig, L. S., Porsteinsson, A. P., Ferris, S., Reichert, M., Ketter, N., Nejadnik, B., Guenzler, V., Miloslavsky, M., Wang, D., Lu, Y., Lull, J., Tudor, I. C., Liu, E., Grundman, M., Yuen, E., Black, R., and Brashear, H. R. (2014) Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. *N. Engl. J. Med.* 370, 322–333.
- (109) Rangan, S. K., Liu, R., Brune, D., Planque, S., Paul, S., and Sierks, M. R. (2003) Degradation of β -amyloid by proteolytic antibody light chains. *Biochemistry* 42, 14328–14334.
- (110) Marr, R. A., Rockenstein, E., Mukherjee, A., Kindy, M. S., Hersh, L. B., Gage, F. H., Verma, I. M., and Masliah, E. (2003) Neprilysin gene transfer reduces human amyloid pathology in transgenic mice. *J. Neurosci.* 23, 1992–1996.
- (111) Leissring, M. A., Farris, W., Chang, A. Y., Walsh, D. M., Wu, X., Sun, X., Frosch, M. P., and Selkoe, D. J. (2003) Enhanced proteolysis of β -amyloid in APP transgenic mice prevents plaque formation, secondary pathology, and premature death. *Neuron* 40, 1087–1093.
- (112) Malito, E., Hulse, R. E., and Tang, W.-J. (2008) Amyloid β -degrading cryptidases: insulin degrading enzyme, presequence peptidase, and neprilysin. *Cell. Mol. Life Sci.* 65, 2574–2585.
- (113) Debela, M., Magdolen, V., Schechter, N., Valachova, M., Lottspeich, F., Craik, C. S., Choe, Y., Bode, W., and Goettig, P. (2006) Specificity profiling of seven human tissue kallikreins reveals individual subsite preferences. *J. Biol. Chem.* 281, 25678–25688.
- (114) Debela, M., Hess, P., Magdolen, V., Schechter, N. M., Steiner, T., Huber, R., Bode, W., and Goettig, P. (2007) Chymotryptic specificity determinants in the 1.0 Å structure of the zinc-inhibited human tissue kallikrein 7. *Proc. Natl. Acad. Sci. U. S. A.* 104, 16086–16091.
- (115) Yousef, G. M., and Diamandis, E. P. (2001) The new human tissue kallikrein gene family: structure, function, and association to disease. *Endocr. Rev.* 22, 184–204.
- (116) Shaw, J. L. V., and Diamandis, E. P. (2007) Distribution of 15 human kallikreins in tissues and biological fluids. *Clin. Chem.* 53, 1423–1432.
- (117) Michael, I. P., Pampalakis, G., Mikolajczyk, S. D., Malm, J., Sotiropoulou, G., and Diamandis, E. P. (2006) Human tissue kallikrein 5 is a member of a proteolytic cascade pathway involved in seminal clot liquefaction and potentially in prostate cancer progression. *J. Biol. Chem.* 281, 12743–12750.
- (118) Brattsand, M., Stefansson, K., Lundh, C., Haasum, Y., and Egelrud, T. (2005) A proteolytic cascade of kallikreins in the stratum corneum. *J. Invest. Dermatol.* 124, 198–203.

- (119) Borgono, C. A., Michael, I. P., Komatsu, N., Jayakumar, A., Kapadia, R., Clayman, G. L., Sotiropoulou, G., and Diamandis, E. P. (2007) A potential role for multiple tissue kallikrein serine proteases in epidermal desquamation. *J. Biol. Chem.* 282, 3640–3652.
- (120) Borgono, C. A., and Diamandis, E. P. (2004) The emerging roles of human tissue kallikreins in cancer. *Nat Rev Cancer* 4, 876–890.
- (121) Stamey, T. A., Yang, N., Hay, A. R., McNeal, J. E., Freiha, F. S., and Redwine, E. (1987) Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. *N. Engl. J. Med.* 317, 909–916.
- (122) Caubet, C., Jonca, N., Brattsand, M., Guerrin, M., Bernard, D., Schmidt, R., Egelrud, T., Simon, M., and Serre, G. (2004) Degradation of corneodesmosome proteins by two serine proteases of the kallikrein family, SCTE/KLK5/hK5 and SCCE/KLK7/hK7. *J. Invest. Dermatol.* 122, 1235–1244.
- (123) Oliveira, J. R., Bertolin, T. C., Andrade, D., Oliveira, L. C. G., Kondo, M. Y., Santos, J. A. N., Blaber, M., Juliano, L., Severino, B., Caliendo, G., Santagada, V., and Juliano, M. A. (2015) Specificity studies on Kallikrein-related peptidase 7 (KLK7) and effects of osmolytes and glycosaminoglycans on its peptidase activity. *Biochim. Biophys. Acta - Proteins Proteomics* 1854, 73–83.
- (124) Johnson, S. K., Ramani, V. C., Hennings, L., and Haun, R. S. (2007) Kallikrein 7 enhances pancreatic cancer cell invasion by shedding E-cadherin. *Cancer* 109, 1811–1820.
- (125) Ekholm, E., and Egelrud, T. (1999) Stratum corneum chymotryptic enzyme in psoriasis. *Arch. Dermatol. Res.* 291, 195–200.
- (126) Hansson, L., Bäckman, A., Ny, A., Edlund, M., Ekholm, E., Hammarström, B. E., Törnell, J., Wallbrandt, P., Wennbo, H., and Egelrud, T. (2002) Epidermal overexpression of stratum corneum chymotryptic enzyme in mice: a model for chronic itchy dermatitis. *J. Invest. Dermatol.* 118, 444–449.
- (127) Chavanas, S., Bodemer, C., Rochat, A., Hamel-Teillac, D., Ali, M., Irvine, A. D., Bonafé, J. L., Wilkinson, J., Taïeb, A., Barrandon, Y., Harper, J. I., de Prost, Y., and Hovnanian, A. (2000) Mutations in SPINK5, encoding a serine protease inhibitor, cause Netherton syndrome. *Nat. Genet.* 25, 141–142.
- (128) Egelrud, T., Brattsand, M., Kreutzmann, P., Walden, M., Vitzithum, K., Marx, U. C., Forssmann, W. G., and Mägert, H. J. (2005) hK5 and hK7, two serine proteinases abundant in human skin, are inhibited by LEKTI domain 6. *Br. J. Dermatol.* 153, 1200–1203.
- (129) Yousef, G. M., Polymeris, M. E., Yacoub, G. M., Scorilas, A., Soosaipillai, A., Popalis, C., Fracchioli, S., Katsaros, D., and Diamandis, E. P. (2003) Parallel overexpression of seven kallikrein genes in ovarian cancer. *Cancer Res.* 63, 2223–2227.

- (130) Zheng, Y., Katsaros, D., Shan, S. J. C., de la Longrais, I. R., Porpiglia, M., Scorilas, A., Kim, N. W., Wolfert, R. L., Simon, I., Li, L., Feng, Z., and Diamandis, E. P. (2007) A multiparametric panel for ovarian cancer diagnosis, prognosis, and response to chemotherapy. *Clin. Cancer Res.* 13, 6984–6992.
- (131) Pollok, B. A., and Heim, R. (1999) Using GFP in FRET-based applications. *Trends Cell Biol.* 9, 57–60.
- (132) Nguyen, A. W., and Daugherty, P. S. (2005) Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat. Biotechnol.* 23, 355–360.
- (133) You, X., Nguyen, A. W., Jabaiah, A., Sheff, M. A., Thorn, K. S., and Daugherty, P. S. (2006) Intracellular protein interaction mapping with FRET hybrids. *Proc. Natl. Acad. Sci. U. S. A.* 103, 18458–18463.
- (134) Baum, E. Z., Beberitz, G. A., and Gluzman, Y. (1990) Isolation of mutants of human immunodeficiency virus protease based on the toxicity of the enzyme in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 87, 5573–5577.
- (135) Daugherty, P. S., Iverson, B. L., and Georgiou, G. (2000) Flow cytometric screening of cell-based libraries. *J. Immunol. Methods* 243, 211–227.
- (136) Boder, E. T., and Wittrup, K. D. (1998) Optimal screening of surface-displayed polypeptide libraries. *Biotechnol. Prog.* 14, 55–62.
- (137) Loeb, L. A., Loeb, K. R., and Anderson, J. P. (2003) Multiple mutations and cancer. *Proc. Natl. Acad. Sci. U. S. A.* 100, 776–781.
- (138) Packer, M. S., and Liu, D. R. (2015) Methods for the directed evolution of proteins. *Nat. Rev. Genet.* 16, 379–394.
- (139) Bershtein, S., and Tawfik, D. S. (2008) Advances in laboratory evolution of enzymes. *Curr. Opin. Chem. Biol.* 12, 151–158.
- (140) Tracewell, C. A., and Arnold, F. H. (2009) Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr. Opin. Chem. Biol.* 13, 3–9.
- (141) Shaner, N. C., Steinbach, P. A., and Tsien, R. Y. (2005) A guide to choosing fluorescent proteins. *Nat. Methods* 2, 905–909.
- (142) Roodveldt, C., Aharoni, A., and Tawfik, D. S. (2005) Directed evolution of proteins for heterologous expression and stability. *Curr. Opin. Struct. Biol.* 15, 50–56.
- (143) Lai, Y.-P., Huang, J., Wang, L.-F., Li, J., and Wu, Z.-R. (2004) A new approach to random mutagenesis in vitro. *Biotechnol. Bioeng.* 86, 622–627.
- (144) Cadwell, R. C., and Joyce, G. F. (1992) Randomization of genes by PCR mutagenesis. *Genome Res.* 2, 28–33.
- (145) Reidhaar-Olson, J. F., and Sauer, R. T. (1988) Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science* 241, 53–57.

- (146) Stemmer, W. P. C. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature* 370, 389–391.
- (147) Ostermeier, M., Shim, J. H., and Benkovic, S. J. (1999) A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* 17, 1205–1209.
- (148) Scheuermann, R., Tam, S., Burgers, P. M. J., Lu, C., and Echols, H. (1983) Identification of the epsilon-subunit of Escherichia coli DNA polymerase III holoenzyme as the dnaQ gene product: a fidelity subunit for DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* 80, 7085–7089.
- (149) Fromant, M., Blanquet, S., and Plateau, P. (1995) Direct random mutagenesis of gene-sized DNA fragments using polymerase chain reaction. *Anal. Biochem.* 224, 347–353.
- (150) Cozzi, R., Malito, E., Nuccitelli, A., D’Onofrio, M., Martinelli, M., Ferlenghi, I., Grandi, G., Telford, J. L., Maione, D., and Rinaudo, C. D. (2011) Structure analysis and site-directed mutagenesis of defined key residues and motives for pilus-related sortase C1 in group B Streptococcus. *FASEB J.* 25, 1874–1886.
- (151) Heckman, K. L., and Pease, L. R. (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protoc.* 2, 924–932.
- (152) Stemmer, W. P. C., Cramer, A., Ha, K. D., Brennan, T. M., and Heyneker, H. L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* 164, 49–53.
- (153) Bloom, J. D., Meyer, M. M., Meinhold, P., Otey, C. R., MacMillan, D., and Arnold, F. H. (2005) Evolving strategies for enzyme engineering. *Curr. Opin. Struct. Biol.* 15, 447–452.
- (154) Zhao, H., Giver, L., Shao, Z., Affholter, J. A., and Arnold, F. H. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.* 16, 258–261.
- (155) Davie, E. W., Fujikawa, K., and Kisiel, W. (1991) The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry* 30, 10363–10370.
- (156) Singh, S. M., and Panda, A. K. (2005) Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng.* 99, 303–310.
- (157) Pugsley, A. P. (1993) The complete general secretory pathway in gram-negative bacteria. *Microbiol. Rev.* 57, 50–108.
- (158) Natale, P., Brüser, T., and Driessen, A. J. M. (2008) Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-Distinct translocases and mechanisms. *Biochim. Biophys. Acta - Biomembr.* 1778, 1735–1756.
- (159) Bardwell, J. C. A., McGovern, K., and Beckwith, J. (1991) Identification of a protein required for disulfide bond formation in vivo. *Cell* 67, 581–589.

- (160) Kishigami, S., Akiyama, Y., and Ito, K. (1995) Redox states of DsbA in the periplasm of *Escherichia coli*. *FEBS Lett.* *364*, 55–58.
- (161) Bardwell, J. C. A., Lee, J.-O., Jander, G., Martin, N., Belin, D., and Beckwith, J. (1993) A pathway for disulfide bond formation in vivo. *Proc. Natl. Acad. Sci. USA* *90*, 1038–1042.
- (162) Rietsch, A., Belin, D., Martin, N., and Beckwith, J. (1996) An in vivo pathway for disulfide bond isomerization in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* *93*, 13048–13053.
- (163) Nakamoto, H., and Bardwell, J. C. A. (2004) Catalysis of disulfide bond formation and isomerization in the *Escherichia coli* periplasm. *Biochim. Biophys. Acta - Mol. Cell Res.* *1694*, 111–119.
- (164) Qiu, J., Swartz, J. R., and Georgiou, G. (1998) Expression of active human tissue-type plasminogen activator in *Escherichia coli*. *Appl. Environ. Microbiol.* *64*, 4891–4896.
- (165) Joly, J. C., Leung, W. S., and Swartz, J. R. (1998) Overexpression of *Escherichia coli* oxidoreductases increases recombinant insulin-like growth factor-I accumulation. *Proc. Natl. Acad. Sci. U. S. A.* *95*, 2773–2777.
- (166) Kurokawa, Y., Yanagi, H., and Yura, T. (2000) Overexpression of protein disulfide isomerase DsbC stabilizes multiple-disulfide-bonded recombinant protein produced and transported to the periplasm in *Escherichia coli*. *Appl. Environ. Microbiol.* *66*, 3960–3965.
- (167) Zhang, Z., Li, Z.-H., Wang, F., Fang, M., Yin, C.-C., Zhou, Z.-Y., Lin, Q., and Huang, H.-L. (2002) Overexpression of DsbC and DsbG markedly improves soluble and functional expression of single-chain Fv antibodies in *Escherichia coli*. *Protein Expr. Purif.* *26*, 218–228.
- (168) Sriubolmas, N., Panbangred, W., Sriurairatana, S., and Meevootisom, V. (1997) Localization and characterization of inclusion bodies in recombinant *Escherichia coli* cells overproducing penicillin G acylase. *Appl. Microbiol. Biotechnol.* *47*, 373–378.
- (169) Wilkinson, B., and Gilbert, H. F. (2004) Protein disulfide isomerase. *Biochim. Biophys. Acta* *1699*, 35–44.
- (170) Hatahet, F., and Ruddock, L. W. (2007) Substrate recognition by the protein disulfide isomerases. *FEBS J.* *274*, 5223–5234.
- (171) Gross, E., Kastner, D. B., Kaiser, C. A., and Fass, D. (2004) Structure of Ero1p, source of disulfide bonds for oxidative protein folding in the cell. *Cell* *117*, 601–610.
- (172) Xie, W., and Ng, D. T. W. (2010) ERAD substrate recognition in budding yeast. *Semin. Cell Dev. Biol.* *21*, 533–539.
- (173) Meusser, B., Hirsch, C., Jarosch, E., and Sommer, T. (2005) ERAD: the long road to destruction. *Nat. Cell Biol.* *7*, 766–772.

- (174) Patil, C., and Walter, P. (2001) Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals. *Curr. Opin. Cell Biol.* 13, 349–356.
- (175) Szegezdi, E., Logue, S. E., Gorman, A. M., and Samali, A. (2006) Mediators of endoplasmic reticulum stress-induced apoptosis. *EMBO Rep.* 7, 880–885.
- (176) Fernández, I. S., Ständker, L., Forssmann, W.-G., Giménez-Gallego, G., and Romero, A. (2007) Crystallization and preliminary crystallographic studies of human kallikrein 7, a serine protease of the multigene kallikrein family. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 63, 669–672.
- (177) Takagi, H., Morinaga, Y., Tsuchiya, M., Ikemura, H., and Inouye, M. (1988) Control of folding of proteins secreted by a high expression secretion vector, pIN-III-ompA: 16-fold increase in production of active subtilisin E in *Escherichia coli*. *Nat. Biotechnol.* 6, 948 – 950.
- (178) Donovan, R. S., Robinson, C. W., and Glick, B. R. (1996) Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter. *J. Ind. Microbiol.* 16, 145–154.
- (179) Georgiou, G., and Valax, P. (1996) Expression of correctly folded proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* 7, 190–197.
- (180) Carlson, M., and Botstein, D. (1982) Two differentially regulated mRNAs with different 5' ends encode secreted and intracellular forms of yeast invertase. *Cell* 28, 145–154.
- (181) Pelham, H. R. B., Hardwick, K. G., and Lewis, M. J. (1988) Sorting of soluble ER proteins in yeast. *EMBO J.* 7, 1757–1762.
- (182) Shusta, E. V., Raines, R. T., Pluckthun, A., and Wittrup, K. D. (1998) Increasing the secretory capacity of *Saccharomyces cerevisiae* for production of single-chain antibody fragments. *Nat. Biotechnol.* 16, 773–777.
- (183) O'Malley, M. A., Lazarova, T., Britton, Z. T., and Robinson, A. S. (2007) High-level expression in *Saccharomyces cerevisiae* enables isolation and spectroscopic characterization of functional human adenosine A2a receptor. *J. Struct. Biol.* 159, 166–178.
- (184) Warren, J. W., Walker, J. R., Roth, J. R., and Altman, E. (2000) Construction and characterization of a highly regulable expression vector, pLAC11, and its multipurpose derivatives, pLAC22 and pLAC33. *Plasmid* 44, 138–151.
- (185) Gietz, R. D., and Woods, R. A. (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* 350, 87–96.

- (186) Lim, E. J., Sampath, S., Coll-Rodriguez, J., Schmidt, J., Ray, K., and Rodgers, D. W. (2007) Swapping the substrate specificities of the neuropeptidases neurolysin and thimet oligopeptidase. *J. Biol. Chem.* 282, 9722–9732.
- (187) Karran, E., Mercken, M., and De Strooper, B. (2011) The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics. *Nat. Rev. Drug Discov.* 10, 698–712.
- (188) Hansson, L., Strömquist, M., Bäckman, A., Wallbrandt, P., Carlsteint, A., and Egelrud, T. (1994) Cloning, expression, and characterization of stratum corneum chymotryptic enzyme: a skin-specific human serine proteinase. *J. Biol. Chem.* 269, 19420–19426.
- (189) Ramani, V. C., Kaushal, G. P., and Haun, R. S. (2011) Proteolytic action of kallikrein-related peptidase 7 produces unique active matrix metalloproteinase-9 lacking the C-terminal hemopexin domains. *Biochim. Biophys. Acta* 1813, 1525–1531.
- (190) Schultz, S., Saalbach, A., Heiker, J. T., Meier, R., Zellmann, T., Simon, J. C., and Beck-Sickinger, A. G. (2013) Proteolytic activation of prochemerin by kallikrein 7 breaks an ionic linkage and results in C-terminal rearrangement. *Biochem. J.* 452, 271–280.
- (191) Ramani, V. C., and Haun, R. S. (2008) The extracellular matrix protein fibronectin is a substrate for kallikrein 7. *Biochem. Biophys. Res. Commun.* 369, 1169–1173.
- (192) Franzke, C. W., Baici, A., Bartels, J., Christophers, E., and Wiedow, O. (1996) Antileukoprotease inhibits stratum corneum chymotryptic enzyme: evidence for a regulative function in desquamation. *J. Biol. Chem.* 271, 21886–21890.
- (193) Miyai, M., Matsumoto, Y., Yamanishi, H., Yamamoto-Tanaka, M., Tsuboi, R., and Hibino, T. (2014) Keratinocyte-specific mesotrypsin contributes to the desquamation process via kallikrein activation and LEKTI degradation. *J. Invest. Dermatol.* 134, 1665–1674.
- (194) Deraison, C., Bonnart, C., Lopez, F., Besson, C., Robinson, R., Jayakumar, A., Wagberg, F., Brattsand, M., Hachem, J. P., Leonardsson, G., and Hovnanian, A. (2007) LEKTI fragments specifically inhibit KLK5, KLK7, and KLK14 and control desquamation through a pH-dependent interaction. *Mol. Biol. Cell* 18, 3607–3619.
- (195) Andreasen, P. A., Georg, B., Lund, L. R., Riccio, A., and Stacey, S. N. (1990) Plasminogen activator inhibitors: hormonally regulated serpins. *Mol. Cell. Endocrinol.* 68, 1–19.
- (196) Oldenburg, J., and Pavlova, A. (2006) Genetic risk factors for inhibitors to factors VIII and IX. *Haemophilia* 12, 15–22.
- (197) Berg, D. T., Gerlitz, B., Shang, J., Smith, T., Santa, P., Richardson, M. A., Kurz, K. D., Grinnell, B. W., Mace, K., and Jones, B. E. (2003) Engineering the proteolytic specificity of activated protein C improves its pharmacological properties. *Proc. Natl. Acad. Sci. U. S. A.* 100, 4423–4428.

- (198) Jabaiah, A., and Daugherty, P. S. (2011) Directed evolution of protease beacons that enable sensitive detection of endogenous MT1-MMP activity in tumor cell lines. *Chem. Biol.* 18, 392–401.
- (199) Kaur, J., and Sharma, R. (2006) Directed evolution: an approach to engineer enzymes. *Crit. Rev. Biotechnol.* 26, 165–199.
- (200) Yep, A., Kenyon, G. L., and McLeish, M. J. (2008) Saturation mutagenesis of putative catalytic residues of benzoylformate decarboxylase provides a challenge to the accepted mechanism. *Proc. Natl. Acad. Sci.* 105, 5733–5738.
- (201) Cobaugh, C. W., Almagro, J. C., Pogson, M., Iverson, B., and Georgiou, G. (2008) Synthetic antibody libraries focused towards peptide ligands. *J. Mol. Biol.* 378, 622–633.
- (202) Varadarajan, N., Cantor, J. R., Georgiou, G., and Iverson, B. L. (2009) Construction and flow cytometric screening of targeted enzyme libraries. *Nat. Protoc.* 4, 893–901.
- (203) Morley, K. L., and Kazlauskas, R. J. (2005) Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* 23, 231–237.
- (204) Balint, R. F., and Larrick, J. W. (1993) Antibody engineering by parsimonious mutagenesis. *Gene* 137, 109–118.
- (205) Zacharias, D. A., Violin, J. D., Newton, A. C., and Tsien, R. Y. (2002) Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* 296, 913–916.
- (206) Rothman, S. C., and Kirsch, J. F. (2003) How does an enzyme evolved in vitro compare to naturally occurring homologs possessing the targeted function? Tyrosine aminotransferase from aspartate aminotransferase. *J. Mol. Biol.* 327, 593–608.
- (207) Abedi, M. R., Caponigro, G., and Kamb, A. (1998) Green fluorescent protein as a scaffold for intracellular presentation of peptides. *Nucleic Acids Res.* 26, 623–630.
- (208) Baird, G. S., Zacharias, D. A., and Tsien, R. Y. (1999) Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl. Acad. Sci. U. S. A.* 96, 11241–11246.
- (209) Kostallas, G., and Samuelson, P. (2010) Novel fluorescence-assisted whole-cell assay for engineering and characterization of proteases and their substrates. *Appl. Environ. Microbiol.* 76, 7500–7508.
- (210) Kostallas, G., Löfdahl, P.-Å., and Samuelson, P. (2011) Substrate profiling of tobacco etch virus protease using a novel fluorescence-assisted whole-cell assay. *PLoS One* 6, e16136.
- (211) Gilon, T., Chomsky, O., and Kulka, R. G. (1998) Degradation signals for ubiquitin system proteolysis in *Saccharomyces cerevisiae*. *EMBO J.* 17, 2759–2766.

- (212) Gilon, T., Chomsky, O., and Kulka, R. G. (2000) Degradation signals recognized by the enzyme pair degradation signals recognized by the Ubc6p-Ubc7p ubiquitin-conjugating enzyme pair. *Mol. Cell. Biol.* 20, 7214–7219.
- (213) Jentsch, S. (1992) The ubiquitin-conjugation system. *Annu. Rev. Genet.* 26, 179–207.
- (214) Johnson, E. S., Bartel, B., Seufert, W., and Varshavsky, A. (1992) Ubiquitin as a degradation signal. *EMBO J.* 11, 497–505.
- (215) Cabantous, S., Terwilliger, T. C., and Waldo, G. S. (2005) Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol.* 23, 102–107.
- (216) Guzman, L.-M., Belin, D., Carson, M. J., and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* 177, 4121–4130.
- (217) Bommarius, A. S., Blum, J. K., and Abrahamson, M. J. (2011) Status of protein engineering for biocatalysts: how to design an industrially useful biocatalyst. *Curr. Opin. Chem. Biol.* 15, 194–200.
- (218) Renicke, C., Spadaccini, R., and Taxis, C. (2013) A tobacco etch virus protease with increased substrate tolerance at the P1' position. *PLoS One* 8, e67915.
- (219) Romero, P. A., Tran, T. M., and Abate, A. R. (2015) Dissecting enzyme function with microfluidic-based deep mutational scanning. *Proc. Natl. Acad. Sci.* 112, 7159–7164.
- (220) Labbe, S., and Thiele, D. J. (1999) Copper ion inducible and repressible promoter systems in yeast. *Methods Enzymol.* 306, 145–153.
- (221) Koller, A., Valesco, J., and Subramani, S. (2000) The CUP1 promoter of *Saccharomyces cerevisiae* is inducible by copper in *Pichia pastoris*. *Yeast* 16, 651–656.