

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Engineering protein tools to understand changes in cell surface proteolysis and viral protein-protein interactions

Permalink

<https://escholarship.org/uc/item/8cg7x91v>

Author

Lui, Irene

Publication Date

2024

Peer reviewed|Thesis/dissertation

Engineering protein tools to understand changes in cell surface proteolysis and viral protein-protein interactions

by
Irene Lui

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

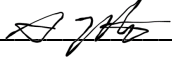
GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:



71181EABCF1940B...


Arun P Wiita

Chair

DocuSigned by:



DocuSigned by:



5F2F4D1A06164C2...

Dr. William Weiss

James A. Wells

Committee Members

Copyright 2024

by

Irene Lui

This work is dedicated to everyone – family, friends, and coworkers- who has supported me through the years. I couldn't have done it without you.

Acknowledgements

Science is not an adventure to be undertaken alone. That is doubly true for any scientist's journey through graduate school. Without the love and support of a whole village of people, I would not be the scientist I am today.

First, I'd like to thank my parents, Jim and Linda, for always pushing me to do my best. Second, my brother, Kevin, who has always been a staunch supporter of me following my passion. Their unwavering belief in my abilities to achieve what seemed impossible.

I would especially like to thank my PhD mentor, Dr. Jim Wells. Without his enthusiasm and love for science, I may have lost my way many times. His patient guidance as I found my own path means more than he can know. Thank you for creating an amazing lab environment that demanded scientific rigor and creativity and his support through the highs and lows have lead me to where I am today. I would also like to thank my thesis committee, Dr. Arun Wiita and Dr. William Weiss, for providing an outside perspective, support, and advice.

The support of many friends and colleagues has been greatly appreciated throughout this journey. First and foremost, Dr. Kaitlin Schaefer, my partner in crime in N-terminomics, from whom I learned so much about developing a project and had many conversations about many challenges. Dr. Jeff Glasgow, Dr. Jamie Byrnes, Dr. Zi Yao, Dr. Johnathan Maza, Dr. Zhi (Lindsey) Lin, Dr. Soumya Remesh, Dr. Shion Lim, Dr. Jie Zhou, Dr. Kevin Leung, and Dr. Xin Zhou were always limitless sources of knowledge and were always willing to help brainstorm new ideas. Dr. Susanna Elledge, Dr. Nicholas Rettko, Dr. Emily Kang, Dr. Lisa Kirkemo, Dr. Paige Solomon, Dr. Katarina Pance, Dr. Cole Bracken, Sophie Kong, and Madison Seto were fellow graduate students who were always willing to help out and I wouldn't have made it through all the late nights and early mornings in lab without them.

To my cohort, Dr. Hannah Toutkoushian, Dr. Chris Waters, and Adam Longhurst, it has been a long journey and graduate school would have been impossible without our hangouts and fun times. To my roommate Chris Mills, thank you for keeping me healthy and fed through these difficult years and making sure I had a life outside of science.

Finally, I'd like to thank the institutions who believed in my work enough to provide funding for it. Specifically, the National Science Foundation Graduate Research Fellowship Program for supporting my training and the work presented herein.

Contributions

Chapter 1: Adapted version of the following published manuscript: The contributors and title of this work are: Schaefer, K., **Lui, I.**, Byrnes, J.R., Kang, E., Zhou, J., Weeks, A.M., and Wells, J.A. Direct identification of proteolytic cleavages on living cells using a glycan tethered peptide ligase. *ACS Central Science*, 2022.

Chapter 2: Adapted version of the following unpublished manuscript: **Lui, I.** Kirkemo, L.L., Zhou, J., Leung, K.K., Schaefer, K. and Wells, J.A. Hypoxia induces extensive proteolytic and protein remodeling of the cell surface in pancreatic adenocarcinoma (PDAC)

Chapter 3: Adapted version of the following unpublished manuscript: **Lui, I.**,* Zhou, X.X.,* Lim, S.A.,* Elledge, S.A., Solomon, P., Rettko, N.J., Zha, B.S., Kirkemo, L.L., Gramespacher, J.A., Liu, J., Muecksh, F., Lorenzi, J.C.C., Schmidt, F., Weisblum, Y., Robbiani, D.F., Nussenzweig, M.C., Hatzioannou, T., Bieniasz, P.D., Roseburg, O.S., Leung, K.K., and Wells, J.A. Trimeric SARS-CoV-2 Spike interacts with dimeric ACE2 with limited intra-Spike avidity. *bioRxiv*, 2021.

*Denotes equal contribution to this work.

**Engineering protein tools to understand changes in cell surface proteolysis and viral
protein-protein interactions**

Irene Lui

Abstract

Cell surface proteolysis plays a crucial role in regulating various biological and cellular processes, including but not limited to cell signaling, protein maturation, and extracellular matrix remodeling. Dysregulation of these proteolytic events is often a hallmark of numerous diseases. Understanding these changes in proteolysis may present opportunities for therapeutic intervention as neo-epitopes are created that may be preferentially displayed in diseased states. Additionally, it is important to understand the mechanism of action for protein-protein interactions in disease, particularly in the case of viral infection, as these interactions may be therapeutic targets for disruption. The following work describes the development of novel protein tools to understand how cell surface proteolysis changes when various oncogenes are overexpressed in breast cancer and how SARS-CoV-2 Spike protein interacts with the human ACE2 receptor. In Chapter 1, I describe the work done to develop a method utilizing subtiligase to study changes to the proteolytic landscape of MCF10A cells under the influence of KRAS G12V and Her2 overexpression. Chapter 2 describes the work done to study changes in cell surface proteolysis and protein secretion in pancreatic ductal adenocarcinoma under hypoxic conditions. In Chapter 3, I describe the development of various protein-Fc fusions to study the protein-protein interactions of Spike protein and ACE2.

Table of Contents

CHAPTER 1	1
DIRECT IDENTIFICATION OF PROTEOLYTIC CLEAVAGES ON LIVING CELLS USING A GLYCAN TETHERED PEPTIDE LIGASE.....	1
ABSTRACT	2
INTRODUCTION	2
RESULTS	4
DISCUSSION	15
METHODS.....	18
REFERENCES	40
CHAPTER 2	46
HYPOXIA INDUCES EXTENSIVE PROTEOLYTIC AND PROTEIN REMODELING OF THE CELL SURFACE IN PANCREATIC ADENOCARCINOMA (PDAC)	46
ABSTRACT	47
INTRODUCTION	47
RESULTS	49
DISCUSSION	53
METHODS.....	56
REFERENCES	71
CHAPTER 3	76
TRIMERIC SARS-COV-2 INTERACTS WITH ACE2 WITH LIMITED INTRA-SPIKE ACTIVITY.....	76
ABSTRACT	77
INTRODUCTION	78

RESULTS	80
DISCUSSION	85
METHODS.....	88
REFERENCES	104

List of Figures

FIGURE 1.1: AN N-TERMINOMICS APPROACH FOR CAPTURING PROTEOLYTIC NEO-N-TERMINI BY CHEMICALLY TETHERING SUBTILIGASE TO GLYCANS ON LIVING CELLS.....	29
FIGURE 1.2: CHEMICAL STRUCTURE OF THE BIOTINYLATED PEPTIDE ESTER SUBSTRATE FOR STABILIGASE.	30
FIGURE 1.3: SCHEMATIC SHOWING THE SYNTHETIC METHOD FOR N-TERMINAL CONJUGATION OF STABILIGASE(A1S) AND CORRESPONDING INTACT PROTEIN MASS SPECTROMETRY TRACES.	31
FIGURE 1.4: STABILIGASE CONJUGATED TO AN N-TERMINAL NUCLEOPHILE FORMS A COVALENT TETHER TO NATIVE GLCYANS ON LIVING CELLS AND EFFICIENTLY LABELS CELL SURFACE N-TERMINI	32
FIGURE 1.5: CLICK-BASED TETHERING OF STABILIGASE ONTO HEK293T CELLS	33
FIGURE 1.6: GT-STABILIGASE ATTACHED TO CELLS BIOTIBYLATES PREDOMINANTLY MEMBRANE PROTEINS AS SHOWN BY IMMUNOBLOT DETECTION WITH STREPTAVIDIN	34
FIGURE 1.7: GT-STABILIGASE TETHERING AND N-TERMINAL LABELING DOES NOT SIGNIFICANTLY AFFECT CYTOTOXICITY	35
FIGURE 1.8: GT-STABILIGASE N-TERMINOMICS BROADLY CAPTURES PRECISE NEO-N- TERMINI ACROSS DIFFERENT CELLULAR CONTEXTS.....	36
FIGURE 1.9: EXTRACELLULAR N-TERMINI DISTRIBUTION ACROSS DOMAINS, DIFFERENT PREDICTED SECONDARY STRUCTURE, AND SOLVENT ACCESSIBILITY.....	37
FIGURE 1.10: COMPARISONS BETWEEN N-TERMINI FOLD CHANGES AND PROTEIN ABUNDANCE CHANGES FOR SINGLE ONCOGENE DRIVEN DATASETS SHOWS MODDEST CORRELATION	38

FIGURE 1.11: INDIVIDUAL ONCOGENES DRIVE COMMON AND UNIQUE PROTEOLYTIC CLEAVAGES ON CELL SURFACES.....	39
FIGURE 2.1: SUBTILIGASE LABELING OF CELL SURFACE N-TERMINI ON PDAC CELL LINES UNDER NORMOXIC AND HYPOXIC CONDITIONS.....	60
FIGURE 2.2: PREDICTED SECONDARY STRUCTURE AT CLEAVAGE SITES.....	62
FIGURE 2.3: N-TERMINI IDENTIFIED BY SUBTILIGASE.....	63
FIGURE 2.4: HYPOXIC CONDITIONS INDUCE EXTENSIVE PROTEOLYTIC REMODELING ON THE CELL SURFACE.....	64
FIGURE 2.5: ANALYSIS OF THE SECRETOME OF PDAC CELL LINES	65
FIGURE 2.6: GO ANALYSIS OF SECRETOMICS CLUSTERS.....	67
FIGURE 2.7: SECRETED PROTEASES AND PROTEINS UNDER HYPOXIC CONDITIONS AND CORRELATION OF N-TERMINI AND SECRETOMICS ANALYSIS.	68
FIGURE 2.8: SHARED PROTEASES IDENTIFIED IN N-TERMINOMICS AND SECRETOMICS.....	69
FIGURE 2.9: CORRELATION PLOTS BETWEEN N-TERMINOMICS AND SECRETOMICS DATA.....	70
FIGURE 3.1: PURIFICATION AND CHARACTERIZATION OF SPIKE AND ACE2 VARIANTS.....	94
FIGURE 3.2: SDS-PAGE ANALYSIS OF SEC FRACTIONS	95
FIGURE 3.3: MELTING TEMPERATURES OF REAGENTS MEASURED BY DIFFERENTIAL SCANNING FLUORIMETRY	96
FIGURE 3.4: BLI CHARACTERIZATION OF BINDING AFFINITY AND KINETICS OF SPIKE AND ACE VARIANTS	97
FIGURE 3.5: SATURATING FL-SPIKE ON BIOSENSOR REVEALS INTERMOLECULAR AVIDITY OF ACE2-Fc.....	98

FIGURE 3.6: ACE2 SPLIT LUCIFERASE EXPERIMENTS DEMONSTRATES MORE THAN ONE RBD
IN FL-SPIKE ARE AVAILABLE TO BIND ACE2..... 99

FIGURE 3.7: C-TERMINALLY FUSED ACE2-Fc SPLIT LUCIFERASE REPORTER IS NOT AS
SENSITIVE AS N-TERMINAL FUSIONS. 100

FIGURE 3.8: ACE2-Fc NEUTRALIZES PSEUDOTYPED VIRUS MORE POTENTLY THAN ACE2
MONOMER 101

FIGURE 3.9: ACE2-Fc NEUTRALIZES SARS-CoV2 LIVE VIRUS MORE POTENTLY THAN ACE2
MONOMER 102

FIGURE 3.10: MODEL OF THE ACE2/SPIKE INTERACTION AND THERAPEUTIC STRATEGIES..... 103

List of Abbreviations

A1S – Alanine 1 to serine

Abu – Aminobutyric acid

AC₄GalNAz – N-Azidoacetylgalactosamine-tetraacylated

ACE2 – Angiotensin converting enzyme 2

ADAM – A disintegrin and metalloproteinase

ADGR6 – Adhesion G protein-coupled receptor G6

AGR2 – Anterior gradient protein 2

BirA – Biotin Ligase

BLI – Biolayer Interferometry

BSA – Bovine serum albumin

CAD13 – T-cadherin

CCR6 – CC chemokine receptor type 6

CD44 – Cluster of differentiation 44

CD99 – Cluster of differentiation 99

CDCP1 – CUB domain containing protein 1

CSC – Cell surface capture

CSPA – Cell surface protein atlas

COFRADIC – Combined fractional Diagonal chromatography

DSF – Differential scanning fluorimetry

DSG2 – Desmoglein 2

DTNB – 5,5'-dithiobis-(2-nitrobenzoic acid); Ellman's reagent

EM – Electronic microscopy

EMT – Endothelial-mesenchymal-transition

EPHA2 – Ephrin receptor A2

ER – Endoplasmic reticulum

Fc – Fragment crystallizable

FL- Full length

GO – Gene ontology

GPCR – G protein-coupled receptor

GPI - glycosylphosphatidylinositol

GT – glycan tethered

HEK293T – Human embryonic kidney 293T cells

HER2 – Human epidermal growth factor receptor 2

IC₅₀ – Half-maximal inhibitory concentration

K_D – Binding affinity

k_{on} – association rate of a protein

KRAS – Kirsten rat sarcoma virus

LC-MS-MS – Liquid chromatography with tandem mass spectrometry

LDLR – Low density lipoprotein receptor

LgBiT – N-terminal fragment of luminescent enzyme

MCF10A – Michigan Cancer Foundation epithelial cell line 10A

MS – Mass spectrometry

N-TAILS – N-Terminal amine isotopic labeling of substrates

NanoLuc – Nanoluciferase

Ni-NTA – Nickel-nitrilotriacetic acid

nM – nanomolar

nm - nanometer

PEG – poly(ethyleneglycol)

PDAC – Pancreatic ductal adenocarcinoma

PDB – Protein Database

PTK7 – Protein tyrosine kinase 7

PTM – Post-translational modification

PTPRK – Protein tyrosine phosphatase receptor type K

RBD – Receptor binding domain

RON – Tyrosine kinase receptor for macrophage stimulating protein

SDC – Syndecan

SEC – Size exclusion chromatography

SmBiT – C-terminal fragment of luminescent enzyme

TEV – Tobacco etch virus

TNB – 2-nitro-5-thiobenzoate

UROK – Urokinase

Chapter 1

Direct identification of proteolytic cleavages on living cells using a glycan-tethered peptide ligase

Abstract

Proteolysis is a critical and irreversible post-translational modification that is prominent in the extracellular milieu. These cleavage events trigger processes such as cell-cell interactions, receptor activation, and shedding of proteins for paracrine and autocrine signaling. Dysregulated cell-surface proteases are common among malignant cell phenotypes including most cancers. Nonetheless, extracellular proteolysis remains poorly understood due in large part to our inability to identify precise proteolytic events in a broad and unbiased manner. To capture proteolytic modifications on the surface of living cells, we have developed a proteomics approach in which an engineered peptide ligase is tethered to native surface glycans, and then used to selectively label N-terminal amines with a biotinylated peptide ester containing a mass tag. We demonstrate the broad utility of this approach by identifying a total of 1532 neo-N-termini across a panel of six different cell types, including primary immune cells. The vast majority of these peptides were not observed in a public knowledgebase of protein N-termini, highlighting the uncharted nature of cell surface proteolysis. We further use this technology to uncover how single oncogenes, *KRAS(G12V)* and *HER2*, induce autologous proteolytic remodeling of the cell surface proteome. Precise cleavage sites in proteins implicated in cancerous cell growth, invasion, and migration are observed, revealing neo-epitopes that can serve as potential new targets for immunotherapeutic approaches. Overall, this work provides a generalizable proteomic platform for understanding how proteases tailor the extracellular proteome.

Introduction

The cell surface proteome comprises approximately 3,000 proteins and is functionally critical for cellular fate and response to environmental stimuli³. Whereas intracellular proteins may be functionally altered by hundreds of different post-translational modifications (PTMs)⁴, protein

modifications in the extracellular space are far more limited. Proteolysis is distinctly prominent among cell surface PTMs, and a repertoire of membrane-embedded and secreted proteases modulate many processes including cell-cell interactions, signal transduction, and cytokine secretion⁵. It is well known that aberrant proteolysis contributes to inflammatory diseases and most cancers⁶. Cleavage events create extracellular new N-termini (neo-N-termini; **Fig. 1.1a**). Characterizing these proteolytic modifications may reveal new proteoforms that can be selectively targeted for immunotherapy⁶.

Advances in mass spectrometry (MS) have greatly improved the global identification of proteolysis, but extracellular proteolytic modifications remain challenging subjects to characterize with current techniques⁷⁻¹⁰. One common approach is to isolate proteins that are proteolytically-shed, otherwise called the secretome, into the supernatant of cell cultures¹¹. Although this method generates substantial information regarding shed proteins, it does not precisely identify cleavage sites and is primarily limited to proteins cleaved close to or within the membrane. Another approach is to enrich and identify C- and N- proteolytic termini peptides from whole cell lysates^{9,12-14}. The high complexity of the proteome and the challenging properties of many membrane proteins – most frequently poor solubility and low abundance relative to intracellular proteins – lead to incomplete coverage of extracellular proteolysis using these approaches.

Here, we develop an N-terminomics approach for characterizing extracellular proteolytic modifications across diverse cell types without genetic manipulation by attaching stabiligase (a subtiligase variant) directly to the surface of living cells. We conjugate stabiligase to an α -nucleophile that forms a covalent linkage to native extracellular glycans and then efficiently labels neo-N-termini. Using this approach, we profiled hundreds of neo-N-termini displayed on the surface of a range of cell types, including both immortalized adherent cells and primary immune

cells. Collectively, we observed 1532 proteolytic modifications across 449 diverse membrane proteins. Lastly, we applied a quantitative N-terminomics approach to reveal how prominent oncogenes, *KRAS(G12V)* and *HER2*, induce extracellular remodeling through proteolysis. We believe this general technology will greatly accelerate the ability to identify and target cell surface proteolytic neoepitopes in healthy and diseased cells.

Results

Stabiligase attachment to native glycans enables efficient neo-N-termini labeling on intact cell surfaces

We previously developed a proteomics technology (N-terminomics) based on subtiligase, a mechanistically-engineered ligase that can specifically label N-terminal α -amines on diverse proteins in a complex milieu (**Fig. 1.1b**)^{15,16}. The basic activity of subtiligase is to catalyze peptide ligation between a donor peptide with a C-terminal ester and the N-terminal amine of an acceptor oligopeptide or protein. The ligase and its further engineered variants are used for diverse biotechnological applications, including peptide cyclization and protein synthesis¹⁷. In N-terminomics, subtiligase typically labels target proteins with a short peptide comprising a biotin handle, a TEV-protease cleavage site, and an aminobutyric acid (Abu) mass tag (**Fig. 1.2**)^{17,18}. In brief, biotinylated proteins are enriched, proteolytically-digested, and the N-terminal peptides are identified after LC-MS-MS analysis by the α -mass-tag (Aminobutyric-acid, Abu; **Fig. 1.1b**). By labeling proteolytic neo-N-termini in whole cell lysates with subtiligase, we have extensively annotated the substrates of soluble proteases like caspases¹⁹. In an attempt to identify extracellular proteolysis, however, we found that using soluble subtiligase to label intact cells or whole cell lysates yielded few cell surface N-termini. By genetically-encoding a transmembrane subtiligase in HEK293T cells, however, we observed substantial improvement in labeling membrane proteins

with the cell surface-displayed ligase; this enabled the identification of hundreds of extracellular neo-N-termini but requires cell engineering²⁰.

Here we sought a cell surface N-terminomics method applicable to different cell types and without genetic manipulation. To achieve this, we developed a chemical strategy to tether the N-terminus of the stable subtiligase variant, stabiligase²¹, to extracellular glycans on intact cells (**Fig. 1**). First, we designed a conjugation strategy for labeling the N-terminus of stabiligase with an α -nucleophile that would readily react with extracellular glycans after cells have been treated with a mild oxidant^{22,23}(**Fig. 1.1b-c**). Auto-prodomain removal generates an N-terminal alanine (A1) on mature stabiligase; to site-selectively modify the ligase, we mutated A1 to serine (A1S) and created a vicinal α -amino-alcohol. This mutation did not alter expression or purification of stabiligase (**Fig. 1.3**). We found that a ten minute sodium periodate oxidation of stabiligase(A1S) is sufficient to completely convert the N-terminal amino-alcohol to a glyoxyl-aldehyde (**Fig. 1.3**). However, this treatment also created a minor product consistent with the oxidation of the active site cysteine, C221. To protect C221, we treated stabiligase(A1S) first with Ellman's reagent (5,5'-dithiobis-(2-nitrobenzoic acid; DTNB) and generated a TNB-C221 adduct²⁴. We then oxidized the ligase with periodate, incubated the N-terminal aldehyde-stabiligase overnight with either bis-aminooxy- or bis-hydrazido- reagent in molar excess to introduce an α -nucleophile, and lastly removed the TNB-protecting group (**Fig. 1.1c**). This strategy produced stabiligase quantitatively functionalized with either an N-terminal α -aminooxy- or α -hydrazido- group.

To pilot stabiligase attachment to cells, we treated HEK293T cells with sodium periodate for ten minutes on ice to form cell surface aldehydes^{22,23}, and then incubated with either of the two conjugated-stabiligases and an amine catalyst (aniline) for fifteen-minutes on ice^{14,15}. Robust tethering of both α -nucleophilic-stabiligases was determined by flow cytometry (**Fig. 1.4a**),

although significantly higher levels of attachment were observed for aminoxy-stabiligase under these conditions, consistent with faster reported kinetic rates of aminoxy-nucleophiles²⁵. Importantly, both the α -nucleophile conjugate and periodate treatment on cells were necessary for stabiligase attachment (**Fig. 1.4a**). Furthermore, we imaged HEK293T cells stained with AlexaFluor647-anti-histidine antibody, which monitors the C-terminal histidine tag on stabiligase, and fluorescent microscopy confirmed that the α -aminoxy-stabiligase was indeed anchored to the membrane (**Fig. 1.4b**). To assess tethering specificity, we pre-treated cells with *V. cholerae* sialidase^{23,26}, a hydrolase that trims the terminal sugars of glycans, and observed dramatically reduced attachment of the aminoxy-stabiligase (**Fig. 1.4c**). We conclude that stabiligase modified with an N-terminal α -nucleophile stably attaches to cell membranes through oxidized cell surface glycans.

Alternate methods for covalent attachment of stabiligase to the cell surface were also considered. We conjugated an N-terminal alkyne onto stabiligase(A1S) to test a click-based approach. Cells were fed Ac₄GalNAz to metabolically incorporate azido-groups into cell surface glycans, and then incubated with alkynyl-stabiligase under copper-based click conditions suitable to living cells^{27,28}. However, only modest attachment of alkynyl-stabiligase was observed by flow cytometry (**Fig. 1.5**). Given this result, we moved forward with an oxidative-coupling approach to tether stabiligase.

To assess the ligase activities of stabiligases tethered to the glycans of HEK293T cells, we incubated cells with a biotinylated peptide ester substrate for 15 minutes at room temperature. Flow cytometry analysis showed that biotinylation was significantly higher for cells tethered with α -nucleophilic stabiligases compared to cells incubated with a soluble stabiligase and the peptide ester (**Fig. 1.4d**). Cytoplasmic and membrane fractions were isolated and immunoblotted

with streptavidin. Biotinylated protein labeling was observed almost exclusively in membrane fractions, and relative biotinylation intensities were congruent with flow cytometry results (**Fig 1.4e and Fig. 1.6**). Likewise, fluorescent microscopy of HEK293T cells stained with AlexaFluor488-streptavidin further showed that N-terminal labeling took place along the cell membrane (**Fig. 1.4e** and see also **Fig. 1.6**). Cell toxicity was evaluated after peptide ligation, and we observed only a modest decrease in cell viability (15%) (**Fig. 1.7**). Collectively, these data show that the glycan-tethering (GT)-stabiligase labels the cell membrane proteome, and that a-aminooxy-functionalized stabiligase is a better conjugate for protein ligation. We were also curious as to whether the proximity of the stabiligase domain to the glycan affected ligation and prepared two additional GT-stabiligases with an N-terminal aminooxy group attached via a 2 or 7 unit poly(ethyleneglycol) (PEG) linker. Although these alternative conjugates add flexibility and theoretical distance between the glycan and ligase domain (up to a three-fold increase in linker length), we observed virtually no difference in biotin labeling (**Fig. 1.4f**).

Mapping neo-N-termini with glycan-tethered (GT) stabiligase N-terminomics

Robust GT-stabiligase tethering and subsequent biotinylation of membrane proteins on HEK293T cells encouraged us to pursue N-terminomics experiments. We treated HEK293T cells with sodium periodate, GT-stabiligase, and the biotinylated peptide ester as described above. Labeled proteins were enriched using neutravidin, digested on-bead with trypsin, and lastly incubated with TEV-protease to release the mass-tagged (Abu) N-terminal peptides for LC-MS-MS analysis (**Fig. 1.1a**). Using features retrieved from UniProt knowledge database²⁹, we identified 507 Abu-tagged peptides (protein neo-N-termini) that mapped to extracellular topology within membrane proteins, extracellular secreted proteins, or GPI-anchored proteins localized in the plasma membrane (**Fig. 1.8a**). Among the proteins observed via N-terminal peptides, most

proteins were type 1 single-pass proteins (53%), which is not surprising since type 1 membrane proteins comprise the majority of cell surface proteins and display an extracellular N-terminus available to both native extracellular proteases and GT-stabiligase³. We also identified neo-N-termini corresponding to multi-pass proteins (20%), secreted proteins (15%), and GPI-anchored (11%) proteins. In contrast, only a few cleavages were observed in type II membrane proteins (2%) which are oriented with a cytoplasmic N-terminus. We repeated the experiment using α -aminoxy-PEG₇-stabiligase and observed similar numbers of cell surface peptides (407 neo-N-termini) which further supports the notion that GT-stabiligase is flexibly incorporated into the cell surface proteome. These data indicate that there is sufficient length, flexibility and mobility in the membrane for GT-stabiligase to access N-termini and hereafter we use the original α -aminoxy-stabiligase for cell surface N-terminomics.

Further analysis showed that identified neo-N-termini were distributed across several types of proteolytic events: the removal of initiator methionine, signal peptide cleavage, propeptide removal, and post-maturation cleavage within the extracellular regions. The majority of neo-N-termini (74%) mapped to the latter group and represent potential cleavage sites of extracellular proteases. Alignment of residues (P4-P4') flanking these inferred cleavage sites did not reveal a significant consensus sequence around the scissile bond (**Fig. 1.8a**), which suggests, not surprisingly, that multiple proteases are responsible for generating these neo-N-termini. We also considered the protein structure at extracellular cleavage sites; the neo-N-termini mapped predominantly to either interdomain, disordered regions or beta-strand regions within domains, consistent with proteolytic substrate preferences (**Fig. 1.9**)³⁰.

To evaluate utility of GT-stabiligase N-terminomics in other cell types, we applied this technology to six different cell types including adherent cells and primary immune cells (**Fig.**

1.8b). Across cell types we observed hundreds of neo-N-termini, ranging from 500-600 for adherent cell lines and 200-400 for immune cells. As seen with HEK293T cells, the majority of extracellular neo-N-termini observed were generated by post-maturation cleavages (mean, 74%) while the remainder were predominantly signal sequence cleavages. In total, 1532 cell surface N-termini from 449 cell surface proteins were captured across the six cell types (**Fig. 1.8b**). An over-representation analysis based on gene ontology (GO) annotations was explored for proteins with proteolytic extracellular N-termini, and specific cellular processes were enriched for adherent cells and primary immune cells (**Fig. 1.9**)³¹. Although this finding is anticipated based on the underlying biological differences between the cell types tested, it highlights the generality of profiling extracellular proteolysis across cellular contexts. Notably, multiple closely spaced cleavage sites were observed within some proteins suggesting sensitive regions of proteolysis. To better characterize how many functionally unique cleavages were observed within proteins, we grouped closely spaced cleavages (less than three residues apart) and observed 936 unique cleavage regions on 449 cell surface proteins.

We also assessed how GT-stabiligase N-terminomics compares to other proteomics methods. Topfind 4.1 is a database containing experimentally-observed N-termini from other proteomic methods (*e.g.*, subtiligase lysate labeling¹⁵, N-TAILS⁹, COFRADIC^{8,12}) and we compared Topfind N-termini to our GT-stabiligase data, grouping N-termini by cleavage type, and subdividing extracellular peptides by the type of membrane protein¹. Strikingly, only 143 N-termini in our data were also found in the Topfind 4.1 database (~9%). Even among the well-annotated protein maturation events detected by GT-stabiligase N-termini labeling (*i.e.*, those identified in Uniprot as signal peptide removal, propeptide cleavage, etc.), only a small percentage were previously characterized by other methods. We also noted that ~50% of the shared N-terminal

peptides originated from extracellular regions of single-pass or secreted proteins, whereas no cleavage sites within multi-pass proteins were found in Topfind 4.1. We then compared our data to the CSPA (Cell Surface Protein Atlas) project, which used cell surface capture (CSC) proteomics to identify 1492 cell surface proteins across 41 human cell types^{22,32}. As expected, we observed significant overlap in proteins between GT-stabiligase N-terminomics and CSPA (67%). Notably, proteins uniquely identified by GT-stabiligase were predicted to be modestly glycosylated (median, 2 glycosites) compared to shared proteins (median, 5 glycosites). We speculate that these proteins were not identified in CSPA because CSC proteomics requires glycosylation for enrichment whereas surface-anchored GT-stabiligase may label neighboring proteins. These comparisons further support the notion that GT-stabiligase yields broad coverage of N-termini on the cell surface with distinct utility relative to other methods.

N-terminomics with GT-stabiligase also gives several lines of evidence as to which proteases are present and active on the cell surface. Proteases are commonly synthesized as inactive precursors that require the removal of an inhibitory N-terminal propeptide for activation³³. Molecular function analysis by gene ontology annotations for proteins identified with cleavages at the pro-mature junction (57 neo-N-termini) showed significant enrichment proteins of endopeptidase activities (**Fig. 1.8d**)³¹. In total, we observed 11 mature, extracellular proteases from several hydrolase families processed at their precise pro-mature junctions, including seven metalloproteases. The latter group includes 4 catalytically-active ADAMs, dedicated sheddases that cleave proteins within their juxtamembrane region¹¹, and we thought that their activity should be reflected in the N-terminomics data. To estimate how many shed proteins were observed, we approximated the membrane proximal cleavage site that is typical for shed proteins. About 140 cleavage sites were located within 30 amino acids of the membrane and are considered candidate

shed proteins (**Fig. 1.8e**). Consistent with this hypothesis, we observed the precise cleavage site for well-studied examples of shed proteins including Notch (*e.g.*, Notch 1,2)³⁴, receptor kinases (*e.g.*, PTK7, PTPRK)³⁵, syndecans (*e.g.*, SDC -1,-4)³⁶, and cell surface receptors (*e.g.*, CD99, CD44, CCR6)¹¹.

Membrane-proximal shedding is a subset of extracellular proteolysis, and more than half of observed extracellular neo-N-termini were located further than 100 amino acids from the membrane (**Fig. 1.8e**). Concurrently, we also observed activated proteases that are not typically considered sheddases. To better characterize extracellular neo-N-termini, we determined structural features surrounding the cleavage sites: relative domain distances, predicted secondary structure, and solvent accessibility (**Fig. 1.9**). Like initial N-terminomics with HEK293T cells, neo-N-termini localized to solvent-exposed regions and primarily unstructured or beta-strand regions. Similar to HEK293T cells, we observed inter- and intra- domain cuts across all extracellular neo-N-termini. Examples of previously characterized cleavages between domains include cleavages between the ephrin-binding domain and fibronectin domains of Eph -A2 and B2, and proteolysis between the light and heavy chains of the urokinase plasminogen activator^{37,38}. Precise intradomain cleavages were also identified, including the known furin-cleavage within the Sema domain of the RON kinase receptor and autoproteolytic GPS domains for two adhesion GPCRs (AGR2 and ADGR6)^{29,39,40}. Across all single-pass membrane proteins, over half of neo-N-termini (65%) were located between the first and last extracellular loop (**Fig. 1.9**). Although these events are not membrane-proximal shedding events, the position of these N-termini suggests they may have significant functional impact. Together, these findings validate that GT-stabiligase N-terminomics is a useful technology for broadly capturing cell surface proteolytic modifications.

Profiling proteolytic changes to the cell surface landscape induced by oncogenes

Cellular disease states are commonly associated with dysregulated proteolytic modifications, but identifying and quantifying the cleavages induced by specific oncogenes remains challenging. We previously quantified oncogene-induced changes in the surface expression of membrane proteins using an immortalized, non-tumorigenic cell line (MCF10A) transformed with individual oncogenes^{41,42}. Two oncogenes, *KRAS(G12V)* and *HER2*, contributed to significant alterations to the cell surface proteome through changes in both protein expression and glycosylation, and we wondered if these transformations might also alter the proteolytic landscape. Importantly, we previously found that CSC proteomics was not biased by glycan alterations⁴¹. Using flow cytometry, we first assessed whether glycan variations may affect the tethering of GT-stabiligase or peptide ligation. Encouragingly, no significant differences were observed among the parent MCF10A transduced with an empty vector (ev) and the two oncogenic cell lines (**Fig. 1.10**).

For quantitative N-terminomics, MCF10A cell lines were cultured in stable isotopic labeling of amino acids (SILAC) media. The oncogene-transformed (*HER2* or *KRAS(G12V)*) cell lines were combined with parental MCF10A cells transformed with an empty vector (ev), labeled with GT-stabiligase, and incubated with the peptide ester as described above (**Fig. 1.11a**). N-terminomics was performed on five biological replicates for both oncogene sets. From these we quantified 303 neo-N-terminal peptides mapped to 151 proteins, and observed 233 N-termini on 89 proteins with differential abundances (1.8-fold threshold). Among these N-termini, 35-40% of extracellular neo-N-termini overlapped between the *HER2*-overexpression and *KRAS(G12V)* datasets, and the fold-change trends were similar for the vast majority of N-termini (**Fig. 1.11b**). In both oncogenic-transformations, as shown in **Fig. 1.11c**, enriched extracellular neo-N-termini

predominantly mapped to cell adhesion proteins and transmembrane signal receptors, two pathways known to undergo proteolytic modifications^{35,43}.

Next, we assessed whether changes in cell surface N-termini coincided with differences in protein abundance in the presence of either oncogene. We plotted the ratios of extracellular neo-N-termini alongside protein abundance values, as previously determined by CSC proteomics (**Fig. 1.11d**, and also **Fig. 1.10**). As shown in Figure 4d, 52 neo-N-termini with greater than 1.8-fold change in abundance mapped to 31 proteins. 80% of these proteins were observed by CSC, and interestingly, the protein abundance values were modestly correlated with N-termini abundance. We note that proteolytic removal of large extracellular domains may contribute to contradictory changes. For instance, syndecan-4 (SDC4) shedding is highly upregulated in both oncogene datasets and the protein was not observed in CSC proteomics. It is likely that cleavage leaves behind a juxtamembrane, neo-N-terminus not suitable for CSC identification. Transcript levels for SDC4, however, were not significantly altered in the presence of *KRAS(G12V)* suggesting that regulation is at the level of the protease. Similar observations were made with individual oncogene datasets (**Fig. 1.10**). Together these data indicate that oncogenes induce changes in expression including proteases and the consequences can be orthogonal, upregulated proteins may or may not correlate with up-regulation of the proteolysis of that protein.

To provide additional validation, we performed immunoblot analysis of selected proteins detected by both CSC and N-terminomics in the parent and transformed cell lines. These experiments used commercially available antibodies that recognize both the full-length and cleaved proteoforms (**Fig. 1.11e**). Notch2 is a receptor and transcription factor in adjacent-cell signaling pathways that is activated by a series of proteolytic cleavages. Mature Notch2 is first cleaved by a furin-like convertase in the Golgi (S1 site), and once on the cell surface, ligand-

binding induces membrane-proximal cleavage by an ADAM metalloprotease (S2) followed by cleavage within the membrane by γ -secretase (S3)³⁴. In the parent and both transformed cell lines we observed N-terminal peptides from both S1 and S2 sites and could observe their cleavage products by immunoblot. In *KRAS(G12V)* cells, cleavage at S1 was increased concurrent with decreased cleavage at S2; in contrast, only an enriched S2 cleavage site was observed in *HER2*-expressing cells. For both cleavage sites, the N-termini ratios were in good agreement with the protein intensities visualized by immunoblot analysis. We analyzed three other proteins of interest, DSG-2, LDLR, and T-cadherin. Proteolysis of cell-adhesion protein DSG-2 plays a role in both cancer and inflammatory cells⁴⁴, and the neo-N-termini characterized here map to domains reportedly cleaved by metalloproteases and ADAM-proteases⁴⁴. Immunoblot analysis showed two intense bands beneath the intact DSG-2 protein in lysates of *KRAS(G12V)* and *HER2* -expressing cells consistent neo-N-termini locations. The LDLR receptor is involved in lipid homeostasis among other functions⁴⁵. The enriched neo-N-terminus of LDLR was observed in both oncogene datasets, and matches a previous report that a metalloprotease cleavage site that results in loss of LDL-class A ligand binding domains 1-4⁴⁵. In agreement with these data, we observe strong protein signal for a species of a molecular weight consistent with the expected product of the cleavage event. Lastly, we observed enriched N-termini mapped to propeptide and extracellular cleavages of a GPI-linked cadherin called T-cadherin (CAD13) which affects cell migration in various cancer types. Similar to our N-terminomics results, we indeed observe increased proteolytic bands consistent with its propeptide-activation and further extracellular cleavage for both *HER2*- and *KRAS(G12V)* - transformed cells. While not exhaustive, these examples and the fact that we find other reported cleavages precisely matching literature reports shows that GT-

stabiligase N-terminomics can accurately capture changes to the cell surface under oncogenic transformations.

Discussion

In summary, we have developed a versatile proteomic technology to precisely profile proteolytic modifications on the surface of intact cells. Using brief and gentle methods, we covalently-tethered stabiligase to extracellular glycans, a prevalent extracellular PTM, and showed that GT-stabiligase efficiently labels N-termini of both non-glycosylated and glycosylated cell surface proteins. We first validated that GT-stabiligase N-terminomics was compatible with varied cell types, including adherent immortalized lines and primary immune cells, and then characterized proteolytic alterations induced by common oncogenes. Collectively, we identified 1637 unique cell surface neo-N-termini across 507 proteins with diverse structures and functions. From these N-termini, we find evidence that proteases impose marked changes to cell surface proteins that include the shedding of entire extracellular portions, removal of discrete protein domains, or the release of inhibitory domains.

Using GT-stabiligase N-terminomics on different cell types revealed proteolytic modifications in distinct molecular pathways that reflect differences in underlying cell function. A prime example is that we observed proteolytic modifications on proteins enriched in immunological pathways among activated T-cell and natural killer cells. Consistently, extracellular proteases, such as ADAMs, are widely expressed and modulate immunity in immune cells⁴⁶. These observations highlight the generality of the method for characterizing proteolysis in different cell-types and environmental contexts. Notably, we also identified activating cleavages within extracellular proteases from mechanistically-diverse families. Although the direct linkage of hydrolase-substrate pairs is challenging due to the complexity of proteolytic networks, GT-

stabiligase N-terminomics in combination with protease knockouts may be a useful adjunct for connecting these relationships.

To uncover how single oncogenes may affect the regulation of cell surface proteolysis, we quantified neo-N-termini on isogenic cell lines expressing the dominant oncogenes *HER2* and *KRAS(G12V)*. We found that these cancerous transformations affect extracellular proteolysis in shared and distinct ways. This was not surprising as the action of multiple proteases within biological pathways suggest that there is not a specific proteolytic profile common to all cancers⁶. Under the influence of either oncogene, for example, we observed increased proteolysis of proteins with important roles in cell growth, proliferation, and metastasis. These effects included increased juxtamembrane shedding of syndecan-4 and CD44, which are cut by ADAM- and metalloproteases to release their soluble domains and thereby modulate ligand binding and intercellular signaling. Concordantly, previous studies suggest that upregulated shedding promotes cancerous proliferation and cell migration^{36,47}. In another example, we observed that *KRAS(G12V)* cells displayed higher levels of proteolytically-modified EphA2, a receptor tyrosine kinase that inhibits Ras-induced growth upon ligand-binding. Metalloproteases remove the ligand-binding domain of EphA2 to promote tumor growth at precisely the junction we observed^{48,49}. In contrast to the examples above, several intriguing proteins underwent differential cleavage between the *HER2* and *KRAS(G12V)* transformed cells, most notably Notch2. Notch signaling requires ADAM-protease cutting close to the membrane^{34,50,51}, and we observed upregulation of the corresponding neo-N-terminus in *KRAS(G12V)*-transformed cells and downregulation in *HER2*-expressing cells. Consistently, Notch2 signaling has been reported to be up- or down-regulated in different cancer cells, and may either promote or suppress tumor growth⁵⁰.

Looking forward, we note that we identified many oncogene-induced proteolytic modifications either poorly characterized or previously not annotated. In future studies, biochemical and cellular experiments are necessary to better understand the functional consequences of these modifications and may provide further insight into how proteins contribute to cancerous phenotypes. We also believe that this method is also amenable to technological improvements. Previous subtiligase engineering efforts identified multiple variants that prefer different N-terminal amino acids, and tethering a cocktail of GT-stabiligase variants may provide additional coverage of cell surface neo-N-termini. We also think it's possible to use GT-stabiligase N-terminomics with alternative quantitative methods, such as isotopically-labeled peptide substrates or TMT-labeling methods,

A major application of GT-stabiligase N-terminomics is that it may identify disease-relevant, neo-epitopes for immunotherapeutic opportunities. For example, we and others have shown that the cleaved form of the cell surface receptor, CDCP1, is highly-specific epitope on different cancer cell types^{42,52,53}. We recently engineered antibodies that selectively bind to cleaved CDCP1 and demonstrated their ability to block tumor growth⁵³. In comparison to those that recognize both full-length and cleaved proteoforms of CDCP1, antibodies against the cleaved proteoform were significantly less toxic. The fact that proteolysis and expression are orthogonal events suggests these neo-epitopes would provide greater selectivity for disease over healthy tissues. Expanding the GT-stabiligase N-terminomics platform across different physiological and pathological contexts will provide insight into how proteases shape the functions of cell surface proteins and may uncover neo-epitopes for precise immunotherapies.

Methods

Cell lines and materials

HEK293T cells were purchased from the UCSF cell culture facility, A549 was gifted by the Rosenberg lab at UCSF, isogenic MCF10A cell lines were constructed previously,^{1,2} A549 cells were gifted from the Rosenberg lab at UCSF, isogenic MCF10A cell lines were constructed previously, PL5 were a gift from the laboratory of E. Scott Seeley (Stanford University, Stanford, California, USA) as described previously³. Primary human T-cells were isolated from leukoreduction chamber residuals following Trima Apheresis (Blood Centers of the Pacific, San Francisco, CA) using established protocols⁴. Cell cultures were maintained in ATCC recommended conditions. All cells were cultured at 37 °C in a 5% CO₂ humidified incubator and passaged no more than 15 times. Cells were tested for mycoplasma contamination yearly.

Biotinylated peptide ester was prepared by solid phase peptide synthesis as previously described⁵. The following reagents were purchased from Sigma Aldrich unless otherwise described: O-[9-(aminooxy)nonyl]hydroxylamine, 5,5-dithio-bis-(2-nitrobenzoic acid) (Thermo Scientific, #22585), Bis-aminooxy-PEG2 (BroadPharm, BP-23592), Bis-aminooxy-PEG7 (Broad Pharm BP-23591), sodium periodate, aniline, iodoacetamide.

N-terminal mutagenesis and B. subtilis transformation with a prodomain-AIS-stabiligase construct

A site-directed mutagenesis reaction mixture was prepared with the forward primer (5'-GATCACGTAGCACATGCGTACCCGTGCCTTACGGCGTATCAC-3') and the reverse primer (5'-GTGATACGCCGTAAGGCACGGGTACGCATGTGCTACGTGATC-3') at 0.5 μM each, the plasmid prep-pro-stabiligase-his6 (100 ng)⁵, dNTPs (0.2 mM), 2.5 mM MgSO₄, 1X KOD Hot start DNA polymerase buffer, and KOD HOT Start DNA polymerase (0.02 U/μl). After

thermocycle amplification, the reaction mixture was digested with Dpn1 (0.8 U/ μ l) at 1 hr, 37 °C at 220 rpm. The PCR product was then added to *E. coli* XL10 for transformation. The A276S-stabiligase plasmid was then transformed into *E. coli* ER1821 and concatemeric DNA was minipreped. *B. subtilis* BG2864 was then transformed using previous methods⁶. Transformed bacteria was then plated on Luria Bertani agar containing 10 μ g/ml chloramphenicol.

Expression and purification of A1S-stabiligase

An overnight culture of *B. subtilis* BG2864 transformed with A1S-stabiligase was grown in Luria Bertani (LB) broth supplemented with 5 μ g/ml chloramphenicol. The culture was then diluted 1:100 into 200 ml of 2XYT broth in a 1 L baffled flask. 1 mM CaCl₂ and 10 μ g/ml chloramphenicol was added to the media, and cells grew 37 °C shaking at 280 rpm for 18-20 hours. Supernatant was clarified (4000xg, 15 min, 4 °C) and a 3:1 v/v of cold EtOH was added. Precipitated material was collected by centrifugation (4000xg, 15 min, 4 °C), and protein was resolubilized in 40 ml of buffer A (50 mM sodium phosphate (pH 8), 300 mM NaCl, and 5 mM imidazole). After removing insoluble debris (4,000xg, 15 minutes, 4 °C), clarified solution was incubated with nickel resin pre-equilibrated in buffer A for 1 hour at 4°C. Unbound material was removed using gravity filtration, and the resin was washed with two consecutive additions of buffer A provided at eight times the resin volume, and then a final addition of buffer A + 20 mM imidazole. Buffer A + 400 mM imidazole was then added to the resin to elute protein. Intact protein mass of stabiligase(A276) was evaluated by intact protein mass spectrometry (MS) using a Xevo G2-XS Mass Spectrometer (Waters) equipped with a LockSpray (ESI) source and a Acquity Protein BEH C4 column (2.1 mm inner diameter, 50 mm length, 300 Å pore size, 1.7 μ m particle size) connected to an Acquity I-class liquid chromatography system (Waters). Deconvolution of mass spectra was performed using the maximum entropy (MaxEnt) algorithm in MassLynx 4.1

(Waters). Using a nanodrop, protein concentration was measured and if required, concentrated using an amicon 10K MWCO tube further to 4-6 mg/ml prior to N-terminal conjugation.

N-terminal conjugation of alpha-nucleophilic handles on AIS-stabiligase

After his-tag affinity purification, solid 5,5-dithio-bis-(2-ntirobenzoic acid) (DNTB, Ellman's reagent; Sigma Aldrich) was added to a final concentration of 2 mM, and the solution was then rocked at room temperature for 30 minutes. TNB-Cys221 addition was monitored by intact protein MS and after quantitative conversion, excess DNTB was removed using a disposable P10-desalt column packed with Sephadex G-25 resin (Cytiva, Catalog # 17085101) and protein was eluted into 50 mM sodium phosphate (pH 8), 300 mM NaCl. Five-fold molar excess of sodium periodate (100 mM in dH₂O; Sigma Aldrich) was added and the N-terminal oxidation proceeded for 10 minutes at 4 °C. After verifying full oxidation with intact protein MS, the excess sodium periodate was removed and the protein was buffer exchanged into buffer B (50 mM sodium phosphate (pH 6.5), 300 mM NaCl). Protein concentration was measured using a nanodrop, and concentrated to a range between 2-6 mg/ml. For hydrazone or propanyl-linked oxime conjugation, 80-100 molar excess of either adipic acid or O-[9-(aminooxy)nonyl]hydroxylamine (Sigma Aldrich) was respectively added with the catalyst aniline (10 mM). N-terminal conjugation proceeded at 4 °C overnight with intact protein MS monitoring after 12-16 hours. Conjugated stabiligase was desalted and buffer exchanged into buffer C (50 mM sodium phosphate (pH 7.4), 300 mM NaCl) using the P-10 desalting column. To remove the Ellman's protecting group, 50 mM TCEP (500 mM stock solution in 1M Tris (pH 8.5); solid from Sigma Aldrich, #75259) was added to the solution and incubated for approximately 10-30 minutes. After MS confirmed complete deprotection, conjugated protein was desalted and buffer exchanged into buffer C using the P-10 desalting column. Concentrated stabiligase was injected into S75 10/300 GL column or HiLoad

16/600 Superdex 75 pg (Cytiva) equilibrated with buffer C and fractions corresponding to monomeric conjugated-stabiligase were collected. Single-use aliquots at 150-200 μM were flash-frozen in liquid nitrogen and then stored at $-80\text{ }^{\circ}\text{C}$.

General protocol for cell surface stabiligase tethering and N-terminal labeling protocol for flow cytometric analysis and western blot detection

HEK293T cells seeded in T75 cm^3 plates was incubated at $37\text{ }^{\circ}\text{C}$ in 5% CO_2 and after reaching confluency, cells were gently washed with DBPS and then incubated with PBS with 0.04% EDTA, free of $\text{Ca}^{2+}/\text{Mg}^{2+}$ for 10 minutes. Dissociated cells were collected and pelleted by centrifugation at 400xg for 5 min at $4\text{ }^{\circ}\text{C}$, washed with cold DPBS (pH 7.4), repelleted, and then transferred to 1.5 mL Eppendorf tubes with acidified DPBS (pH 6.5). Cells were treated with 500 μM sodium periodate (stock solution of 100 mM sodium periodate in dH_2O) and gently rocked for 15 min at $4\text{ }^{\circ}\text{C}$. Cells were re-pelleted (400xg, 5 min, $4\text{ }^{\circ}\text{C}$), washed with cold DPBS (pH 6.5), re-pelleted, and resuspended in stabiligase-tethering mixture containing DPBS (pH 6.5) to which 10 mM aniline and then 5 μM functionalized stabiligase was added. Cells were then gently rotated at $4\text{ }^{\circ}\text{C}$ for 15 min, re-pelleted, washed with DPBS (pH 7.5), re-pelleted, and then resuspended in stabiligase-reaction mixture consisting of 50 mM Tricine (pH 8), 150 mM NaCl, and then 1% DMSO for a final 1 mM concentration of the biotinylated peptide ester (200 mM stock in DMSO). Cells were then gently rotated for 15 min at room temperature before pelleting cells (400xg, 5 min, $4\text{ }^{\circ}\text{C}$), washing cells with PBS (pH 7.4), and then re-pelleting. To assess tethering and ligase activity by flow cytometry, cells were then stained with 1:2000 dilution of Alexa647-anti-histidine for monitoring the his-tagged stabiligase and Alexa-1:5000 dilution of Alexa488-streptavidin for protein biotinylation in DPBS (pH 7.4) with 3% BSA. Cells were pelleted and washed in DPBS

with 3% BSA prior to analysis. Samples were analyzed on a Beckman Coulter CytoFlex flow cytometry and then using FlowJo software.

For immunoblot analysis, pelleted cells were fractionated using a subcellular fractionation kit (Thermo Fisher) according to the manufacturer's instructions for both the cytoplasmic extraction and the membrane extraction. Protein concentration was measured by BCA (Thermo Fisher) and fraction aliquots were frozen at -20°C before use. 15 µg of protein from each fraction was subjected to SDS-PAGE and blotted on PVDF membranes using an iBlot transfer stack (Thermo Fisher, #IB301002). Biotinylation was detected using IRDye 800CW Streptavidin (1:5000 dilution; LI-COR Biosciences, #926-32230) and total protein loaded in each lane was assessed using REVERT protein staining kit (LI-COR Biosciences, #926-11015) according to the manufacturer's instructions. Near-infrared (NIR) immunoblot images were analyzed using an Odyssey Li-COR imaging system and further analyzed using ImageJ.

General protocol for cell surface N-terminomics

For adherent cells (HEK293T, MCF10A, PL5, A549), cells were cultured to approximately 90% confluency in plates that yielded 7-12x10⁶ cells. For SILAC-cultured MCF10A cells were passaged at least eight times prior in the light and heavy-isotopic media. Cells were harvested by washing plates with cold DPBS and incubating with PBS with 0.04% EDTA, free of Ca²⁺/Mg²⁺ incubation. For SILAC-based experiments, cells were combined in equal numbers after cell counting using an Bio-rad TC20 automatic cell counter. Cells were pelleted by centrifugation (400xg, 5 min, 4°C), washed with 1 mL DPBS (pH 6.4) and re-pelleted before proceeding further. For primary immune cells (CD4⁺ T cells, CD8⁺ T cells, NK), cells were counted using a hemocytometer and 20-30x10⁶ cells were collected, pelleted by centrifugation (400xg, 5 min, 4°C), washed with DPBS (pH 7.4) and pelleted before proceeding further.

After harvesting cells, all cell types were treated similarly. Cells were resuspended in 1 mL DPBS (pH 6.4) containing 500 μ M NaIO₄ (100 mM NaIO₄ stock solution in dH₂O) in a low-bind Axygen 1.7 mL tube (Thermo Fisher). The sodium periodate reaction mixture was then covered from light and rotated on ice at 300 rpm for 10 minutes. Cells were re-pelleted by centrifugation (400xg, 5 min, 4 °C), washed with DPBS (pH 7.4), and then resuspended in DPBS (pH 6.4). To a final reaction volume of 1 mL, 10 mM aniline (Sigma Aldrich) and then 5 μ M GT-stabiligase was added. The tethering reaction proceeded at 300 rpm on ice. Cells were re-pelleted by centrifugation (400xg, 5 min, 4 °C), washed with 1 mL DPBS (pH 7.4), and then resuspended in the subtiligase master mix (100 mM Tricine (pH 8), 150 mM NaCl, 1 mM biotinylated peptide ester (1% final DMSO)). Cells were gently rocked at room temperature for 15 minutes before centrifugation at 400xg, 5 min, 4 °C. Cells were washed two times with DPBS (pH 7.4) and then flash-frozen in liquid nitrogen for storage at -80 °C prior to mass spectrometry sample work-up.

Sample preparation for LC-MS-MS analysis

For sample processing, frozen cell pellets were thawed on ice and then resuspended in 1 mL lysis buffer (RIPA, Radioimmunoprecipitation assay buffer) supplemented with a complete protease inhibitor cocktail (Roche) and DNase 1 (100 μ g/ml). After a 30 min incubation at 16 °C, cells were briefly sonicated (3 pulses, 20% amplitude) and cell debris was removed by centrifugation (20,000xg for 10 min, 4 °C). Clarified supernatant was added to 250 μ l NeutrAvidin agarose beads (Pierce, #29200) and incubated for 3 hours at 16 °C. Non-specific bound proteins were removed by collecting beads with spin columns (Pierce, #69725) and washing iteratively with 5 x 1 mL of RIPA buffer, 5 x 1 mL 100 mM Tris (pH 7.5) and 1 M NaCl, 5 x 1 mL 50 mM ammonium bicarbonate and 2 M Urea. Beads were then transferred to 1.5 mL maximum recovery tubes (Axygen, MCT-150-L-C; 1 mL of 50 mM ammonium bicarbonate and 2 M Urea). TCEP (5

mM) was added to the resuspended mixture and then samples were incubated 55 °C for 30 minutes at 800 rpm. After samples were equilibrated to rt, the iodoacetamide (14 mM) was added and incubated for 30 minutes. Beads were centrifuged (800xg, 1 min) and washed two times with 50 mM ammonium bicarbonate with 2 mM Urea. 6 µg of sequencing grade trypsin (Promega, #V5113) was added for an overnight incubation at rt. Beads were collected and washed as described above prior to trypsin digestion. The collected beads were then washed with an additional 3 x 1 mL dH₂O and then resuspended in TEV buffer (700 µl, 50 mM Tris (pH 8.0), 0.5 mM EDTA, 1 mM dithiothreitol). 8 µg of TEV protease was added to each sample and the beads were incubated at rt overnight. Beads were removed with spin columns, and the supernatant was collected into a fresh 1.5 mL maximum recovery tube. Beads were washed with 2 x 100 µl of dH₂O. The collected supernatant was dried to completion using a genovac drying system and then desalted using preomics desalting columns per the manufacturer's instructions.

Liquid chromatography mass spectrometry analysis of N-termini peptides

Desalted peptides (200 ng) were loaded onto a timsTOF pro equipped with a CaptiveSpray source and a nanoElute line (Bruker; Hamburg, Germany). The peptides were separated on a 25 cm, ReproSil c18 1.5 µM 100 A column (PepSep, PN. # PSC-25-150-15-UHP-nc) using a step-wise linear gradient method with H₂O in 0.1% Formic acid and acetonitrile with 0.1% formic acid (solvent B): 5-30% solvent B for 90 min at 0.5 µl/min, 30-35% solvent B for 10 min at 0.6 µl/min, 35-95% solvent B for 4 min at 0.5 µl/min, 95% hold for 4 min at 0.5 µl/min). Acquired data was collected in a data-dependent acquisition mode with ion mobility activated in PASEF mode. MS and MS/MS spectra were collected with m/z ranging from 100 to 1700 in positive mode.

All acquired data was searched using PEAKS online Xpro 1.6 (Bioinformatics Solutions Inc.; Ontario, Canada)⁷. Spectral searches were performed using a custom FASTA-formatted

dataset containing swissprot-reviewed human proteome file with gene ontology localized the plasma membrane (downloaded from Uniprot knowledge database)⁸. A precursor mass error tolerance was set to 20 ppm and a fragment mass error tolerance was set at 0.03 ppm. Peptides, ranging from 6 to 45 amino acids in length, were searched in semi-specific tryptic digest mode with a maximum of two missed cleavages. Carbidomethylation (+57.0214 Da) on cysteines was set as a static modification and methionine oxidation (+15.994) was set as a variable modification. N-terminal specific peptides were identified by the N-terminal amino-butyric mass tag (+85.052764). Lastly, peptides were filtered based on a false discovery rate (FDR) of 1%.

SILAC-labeled datasets were also analyzed in PEAKS online Xpro 1.6 with the following alterations. Two additional variable modifications were set to account for the isotopic-labels: $^{13}\text{C}(6)^{15}\text{N}(2)$ and $^{13}\text{C}(6)^{15}\text{N}(4)$, quantified peptides were matched between experimental replicates and enrichments were normalized based on total ion chromatograph (TIC). Output PEAKS files contained peptide ratios. If peptides overlapped with the same N-terminal site (*i.e.*, tryptic C-termini, different oxidation), the peptides were grouped together and the average ratio was reported for a given experimental dataset. Across replicates, peptides were removed if observed with high variation (coefficient of variation > 0.7) and peptides were quantified if present in two biological replicates. The final N-terminal peptide ratio was then reported as the median \log_2 fold-change value alongside the shortest N-terminal tryptic peptide sequence.

Analysis of N-terminomics datasets

With custom python scripts and the Swiss-prot reviewed human proteome exported from the uniprot knowledge database, MS files processed in PEAKS were further processed to filter neo-N-termini according to the following features: topological domain of peptide, type of membrane protein, subcellular localization, predicted glycosites, distance to signal sequence sites,

distance to propeptide sequence sites. Cell surface displayed proteins were annotated if identified by the following criteria: extracellular topology of membrane proteins (single-pass or multi-pass); proteins with a subcellular localization as extracellular secreted proteins; or GPI-linked proteins localized to the plasma membrane. For the location of the following proteolytic cleavages--initiator methionine removal, signal peptide cleavage, propeptide cleavages—a precision ruler of 10 amino acids from the annotated site was set. Additionally, neo-N-termini were paired to cell surface capture surfaceomics data from the previous reference¹ and the cell surface atlas public datasets (<https://wlab.ethz.ch/cspa/#downloads>)⁹. Additional analyses included amino acid distances from neo-N-termini to the most proximal transmembrane helix or GPI-linked anchor and neo-N-termini relative to annotated domain architectures. In some instances, incomplete annotations (missing topological domain orientations or single-pass membrane proteins lacking topology annotations) were observed and topology of N-termini were manually annotated based on the Protein Protter server (<http://wlab.ethz.ch/protter/start/>)⁹. Gene ontology enrichments were obtained using the webgestalt server using an FDR cut-off at 1%¹⁰, and functional protein assignments were made using the Panther server¹¹. For solvent accessibility and secondary structure analysis, features in Uniprot and predicted structural features in AlphaFold 2.0 were used^{8,12,13}.

For a topfind 4.1 comparison¹⁴, protein accessions and peptide sequences were submitted as lists with default parameters and a precision range within 10 amino acids. For sequence logo generation, IceLogo was used with a precompiled Swiss-Prot *Homo sapiens* composition as the reference set.

For surface accessibility and secondary structure analyses, custom R scripts recalled PDB structures and AlphaFold 2.0 structural predictions^{13,15}.

Immunoblot analysis of proteolytic substrates

For validating quantitative ratios of enriched neo-N-termini, MCF10A cells (empty vector (ev), *her2*, *krasG12V*) were grown in 10 cm³ plates to 90% confluency. Cells were harvested by washing plates with cold DPBS and incubating with PBS with 0.04% EDTA, free of Ca²⁺/Mg²⁺ incubation. Cells were pelleted, washed three times with DPBS, and then resuspended in RIPA buffer supplemented with complete protease inhibitor cocktail (Roche) and DNase 1 (100 µg/ml). After a 30 min incubation at 16 °C, cell debris was removed by centrifugation (20,000xg for 10 min, 4 °C). Total protein concentration was determined using a Pierce BCA Protein Assay kit (#23225) and 20 µg of protein was incubated with 4X-SDS loading buffer containing 5% β-mercaptoethanol on ice for thirty minutes. Samples were separated using a 4-12% Bolt, Bis-Tris, invitrogen gel and then transferred to a PVDF membrane. The membrane was blocked with TBS Intercept blocking buffer (LI-COR, #927-60001) rotating for 1 hour at room temperature. The following primary antibody solutions were added for the respective protein targets: anti-Notch2 (1:1000 dilution; Cell Signaling, #5372), anti-LDLR (1:1000 dilution, R&D, # AF2148); anti-DSG-2 (1:1000 dilution, Abcam, #ab226258), Anti-T-cadherin (1:2000, CDH13; EMD Millipore, #ABT121). All primary antibody solutions also contained anti-actin (anti-mouse or anti-rabbit; 1:2000 dilution; Cell Signaling, #3700 or #4970). All, with the exception of anti-DSG-2, were rocked overnight at 4 °C. For DSG-2, the blot was incubated with anti-DSG-2 solution for 1 hour at room temperature. After iterative washing three times with TBS-T, blots for detecting Notch2, DSG-2, T-cadherin were incubated with either IRDye-680/800-anti-mouse and IRDye-680/800 anti-rabbit (1:10000 dilution) based on the primary antibody used. For blots detecting LDLR, IRDye-800-anti-goat (1:10000 dilution) and IRDye-680-anti-mouse were used. Blots were rocked at room temperature for 1 hour in the dark before iterative washing three times with TBS-T. Near-

infrared (NIR) immunoblot images were analyzed using an Odyssey Li-COR imaging system and further analyzed using ImageJ.

Figures and Tables

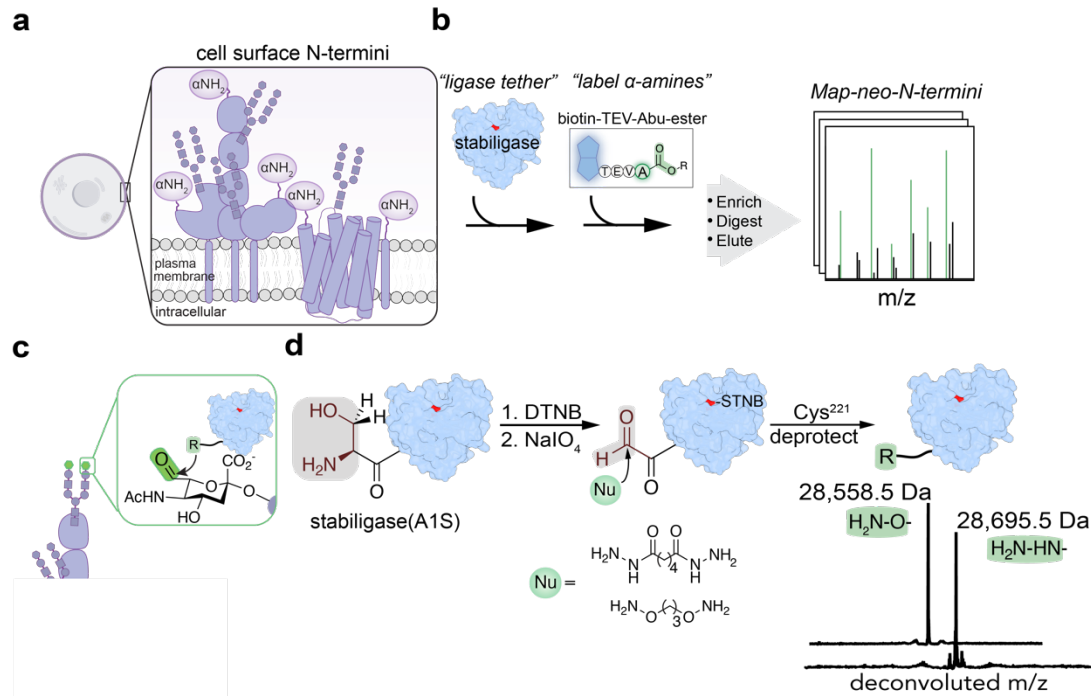


Figure 1.1: An N-terminomics approach for capturing proteolytic neo-N-termini by chemically-tethering stabiligase to glycans on living cells.

a, Membrane-embedded and secreted proteases modify cell surface proteins and create new N-termini (neo-N-termini) often exposed within the extracellular environment. To characterize proteolytic modifications on the cell surface, we considered attaching the engineered ligase, stabiligase, to cell surfaces. **b**, In the presence of accessible N-termini, stabiligase tags α -amines with a peptide ester containing a biotin (blue), a TEV-protease cleavage site, and an amino-butyrinic acid mass tag (A, Abu, green sphere)². After a MS workflow (protein enrichment on neutravidin, proteolytic digestion, release from neutravidin), Abu-N-termini peptides are identified using LC-MS-MS. **c**, Strategy for cell surface tethering of α -nucleophile stabiligase. Treating cells with a mild periodate condition creates aldehydes on extracellular glycans which may react with N-terminal nucleophiles to facilitate a direct stabiligase tether to cell surfaces. **d**, Synthetic conjugation scheme to modify the N-terminus of stabiligase with an α -nucleophile. Purified stabiligase(A1S) was incubated with Ellman's reagent to create a C221-adduct, and then incubated briefly with sodium periodate to generate an N-terminal aldehyde. The aldehyde-stabiligase then reacted completely after an overnight incubation with either excess bis-amino-oxy or bis-hydrazide-based reagents. Following the removal of the Cys-TNB with a reducing agent, fully-functionalized α -nucleophilic-stabiligases were obtained.

Biotin-EEENLYFQ-Abu-glycolate-R

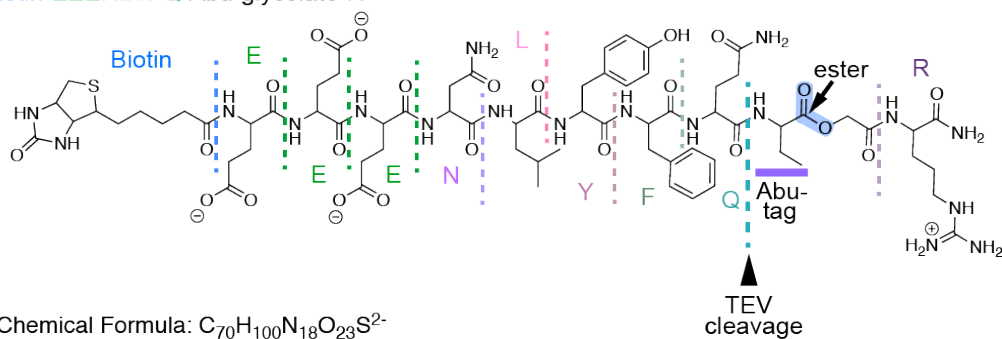


Figure 1.2: Chemical structure of the biotinylated peptide ester substrate for stabiligase.

The peptide ester was prepared using solid-phase peptide synthesis as reported previously⁵.

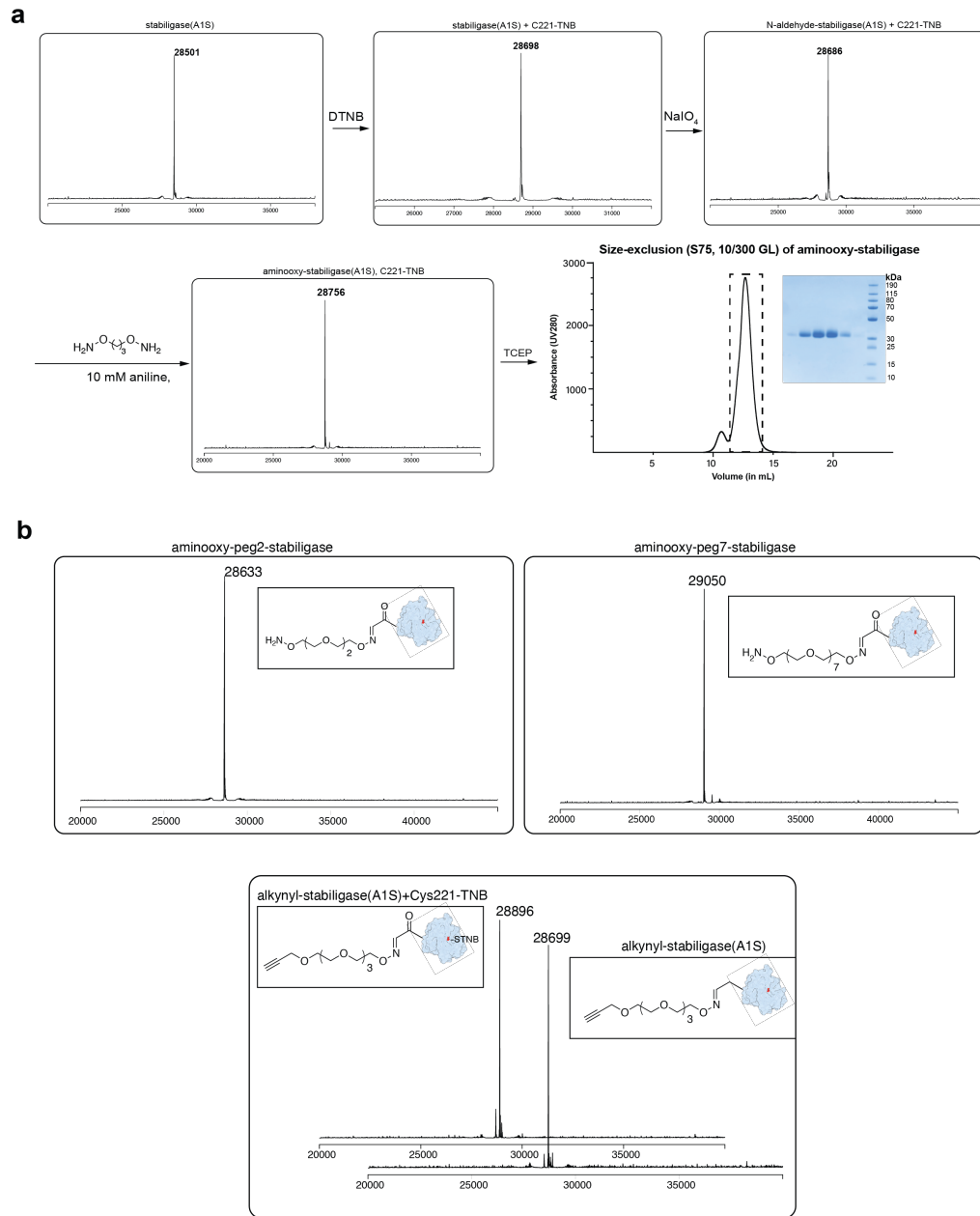


Figure 1.3: Schematic showing the synthetic method for N-terminal conjugation of stabiligase(A1S) and corresponding intact protein mass spectrometry traces.

a, Stabiligase(A1S) is purified using similar conditions as reported⁵. In the presence of Ellman's reagent (5,5-dithio-bis-(2-nitrobenzoic acid), DTNB), the active site cysteine (C221) is protected. Then, TNB-protected stabiligase is treated with sodium periodate briefly on ice and forms a mass consistent with the formation of a N-terminal glyoxyl group. After an overnight incubation with aminoxy-reagents (O-[9(aminoxy)nonyl]-hydroxylamine shown, see SI Fig. 2 for additional conjugate masses), complete conversion to the N-terminal aminoxy-stabiligase is observed. **b**, Deconvoluted intact mass spectrometry traces show complete conversion of N-terminal serine stabiligase into an N-terminal functional group based on respective bis-aminoxy-reagents or an aminoxy-alkynyl-reagent.

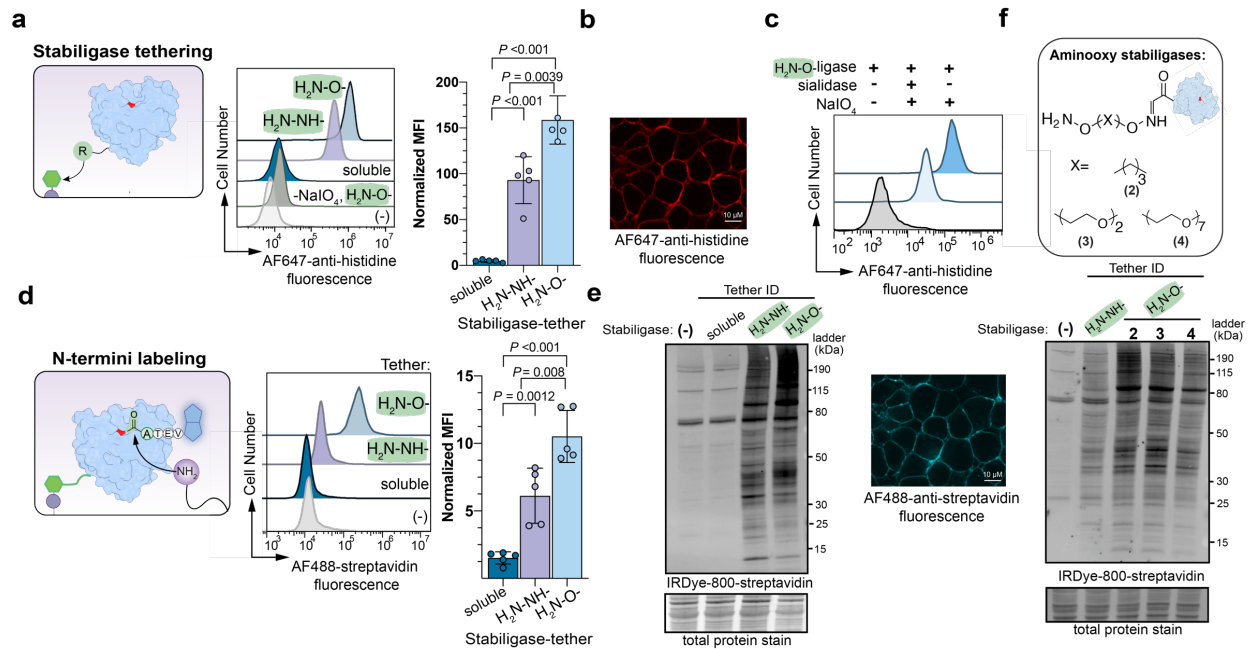


Figure 1.4: Stabiligase conjugated to an N-terminal nucleophile forms a covalently tether to native glycans on living cells and efficiently labels cell surface N-termini

a, Stabiligase tethering to HEK293T cells, initially treated with sodium periodate and then conjugated stabiligase variants, was monitored by flow cytometry with AF647-anti-histidine staining. An N-terminal aminoxy- group mediates higher stabiligase tethering onto cells compared to a hydrazide group. **b**, Fluorescence microscopy of HEK293T cells tethered to aminoxy-stabiligase showed exclusive membrane staining. **c**, Pre-treatment of HEK293T cells with *Vibrio cholerae* sialidase showed greatly reduced attachment of aminoxy-stabiligase. **d**, Compared to unconjugated soluble stabiligase, glycan-tethered (GT)-stabiligases showed dramatically improved ligase activity in the presence of a biotinylated peptide ester substrate. Cells were stained with AF-488-streptavidin to detect labeling by flow cytometry analysis. **e**, Immunoblot detection of biotinylated proteins within the subcellular membrane fraction was consistent with flow cytometry staining of cells in panel **d**. **f**, Labeling efficiencies of GT-stabiligases with varying linker lengths were compared using immunoblot analysis. Increasing the linker distances between the glycan and the stabiligase domain did not affect N-termini biotinylation. Data are representative of at least three independent experiments with similar results. In **a**, and **d**, data are presented as the mean \pm s.e.m., and P values were calculated using one-way ANOVA followed by Tukey's multiple comparisons test.

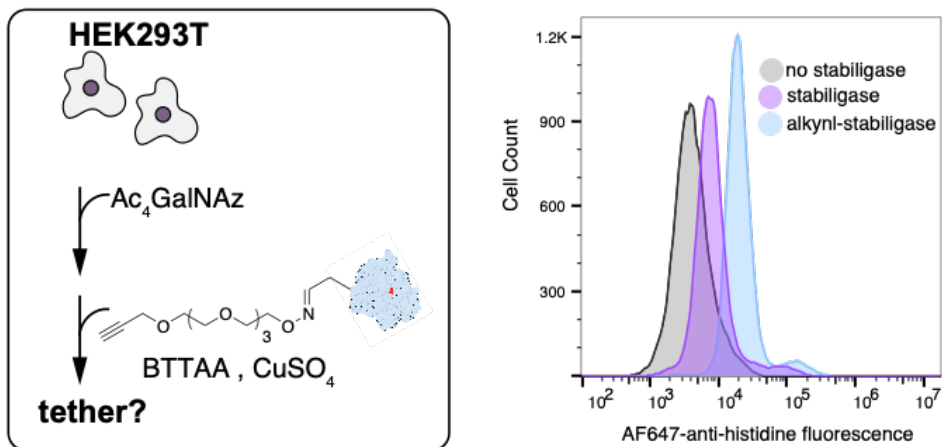


Figure 1.5: Click-based tethering of stabiligase onto HEK293T cells.

Cells were cultured with azido-GalNAz similar to previous literature. After 48 hours, cells were then harvested, washed three times with DPBS, and then incubated with BTTAA, $CuSO_4$, and an N-terminal-alkynyl-stabiligase using previous methods suitable for living cells. HEK293T cells were subsequently washed, incubated with AF647-anti-histidine (1:5000), and analyzed using flow cytometry. As shown by analysis obtained from FlowJo, modest attachment of stabiligase was achieved.

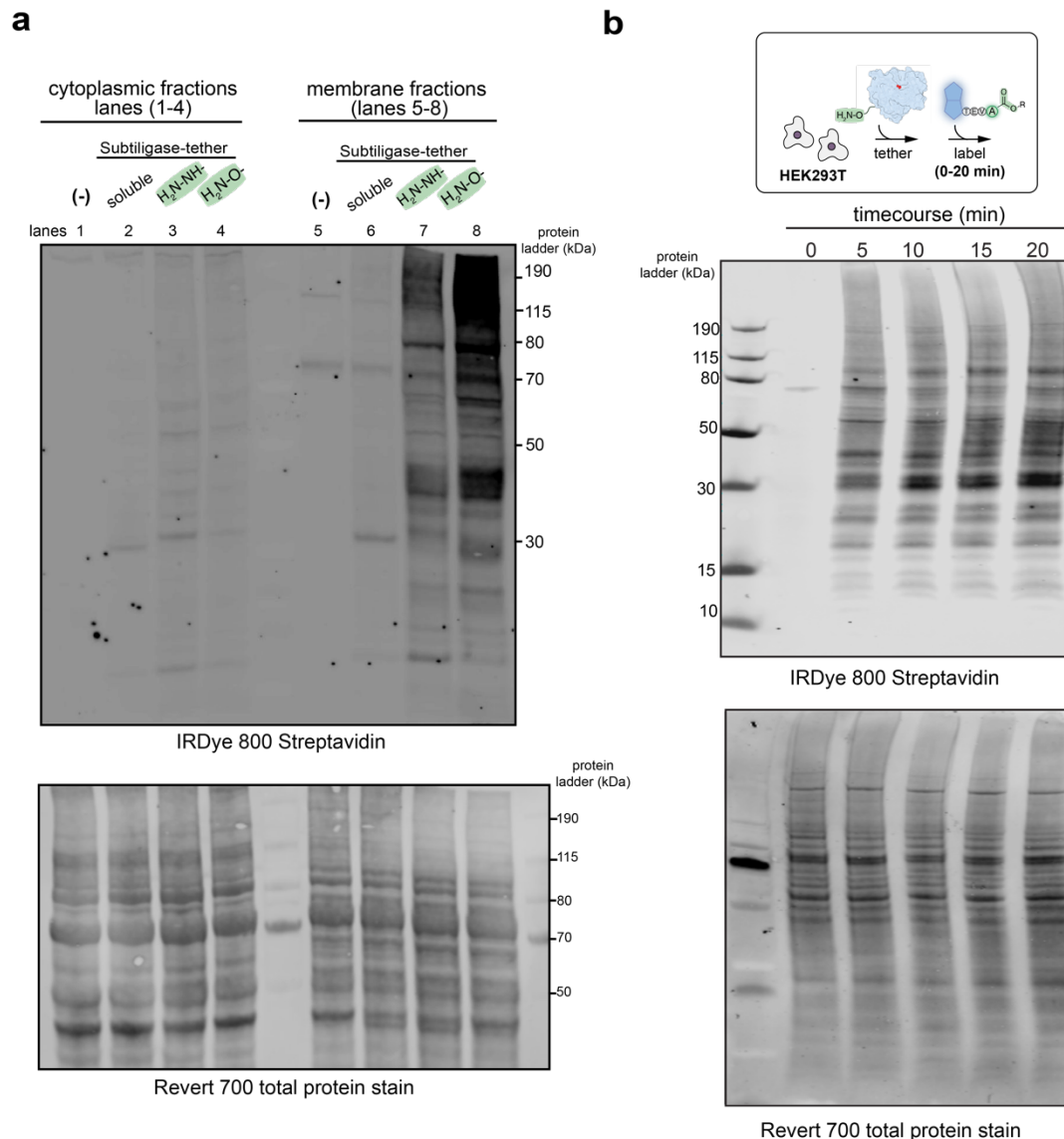


Figure 1.6: GT-stabiligase attached to cells biotinylates predominantly membrane proteins as shown by immunoblot detection with streptavidin.

a, In the following order, HEK293T cells were tethered with GT-stabiligase, labeled with the biotinylated peptide ester substrate, and then lysed cells were fractionated using subcellular fractionation. 20 μ g of cytoplasmic and membrane fractions were loaded onto a SDS-PAGE gel, transferred to PVDF, and then blotted with Revert 700 total protein stain (LI-COR) and IRDye-800-Streptavidin (LI-COR). **b**, HEK293T cells tethered with aminooxy-stabiligase pelleted (400xg, 4 °C for 5 min), washed, and kept on ice before subcellular fractionation (Thermo Scientific subcellular protein fractionation kit, Catalog #78840). Membrane fractions (20 μ g) were separated through PAGE and then transferred onto PVDF for immunoblot analysis as described in panel **a**. These experiments were performed on at least three individual times with similar results.

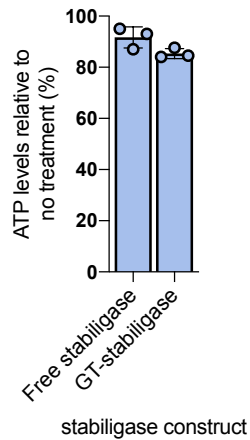


Figure 1.7: GT-stabiligase tethering and N-terminal labeling does not significantly affect cytotoxicity.

HEK293T cells were treated with sodium periodate, and then incubated with either soluble or N-terminal aminoxy-stabiligase as described for other stabiligase tethering experiments. Subsequently, 1 mM biotinylated peptide ester (1% DMSO final) was added to cells. For soluble stabiligase reactions, an additional 2 μ M soluble stabiligase was added alongside the peptide ester. After a 15 min ligation, cells were pelleted and washed with cold DPBS (7.4). To evaluate cell viability, the CellTiter Glo luminescence assay was used according to the manufacturer's instructions to quantify ATP levels and luminescence for stabiligase reactions was normalized relative to untreated HEK293T cells as a control. Independent experiments were performed in triplicate, and the mean \pm S.D. is reported.

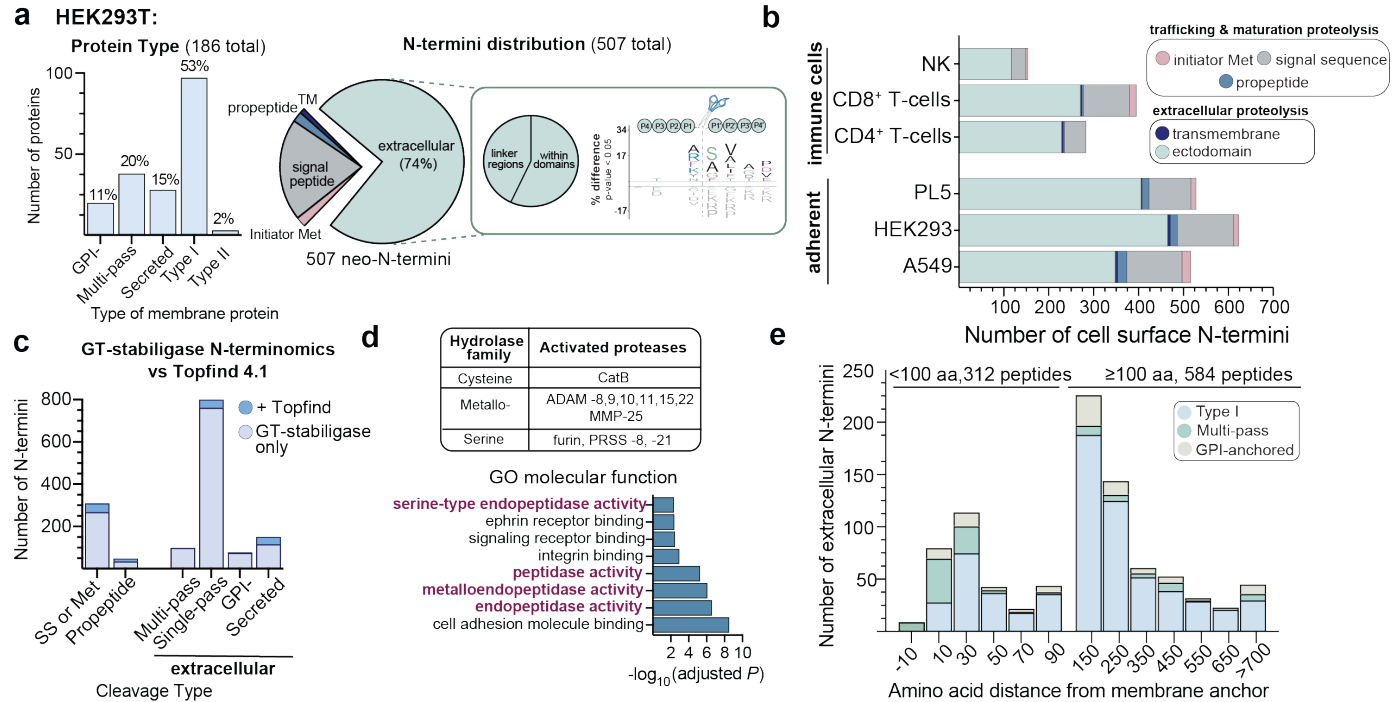


Figure 1.8: GT-stabiligase N-terminomics broadly captures precise neo-N-termini across different cellular contexts

a, Initial GT-stabiligase N-terminomics was performed on HEK293T cells, which yielded 507 N-termini on 186 cell surface proteins. The distribution of membrane proteins mapped with N-termini was similar to the population ratios of different membrane proteins. N-termini peptides were grouped based on the location of cleavage: initiator methionine (Met), signal peptide, propeptide junction, transmembrane, and extracellular regions of proteins. The vast majority of cleavages (74%) mapped to extracellular regions of proteins and were localized either to linker regions or within domains that were predominantly predicted as beta-strands. The icelogo of the P4-P4' residues flanking the cut-site (scissors) shows a range of amino acids at the P1 position. **b**, GT-stabiligase cell surface N-terminomics captures neo-N-termini across adherent cell types and primary immune cells. **c**, N-termini in panel **b**, were compared to N-termini deposited to Topfind 4.1 and only a small percentage of N-termini were observed previously (dark blue)¹. **d**, For proteins identified with pro-mature junction cleavages, endopeptidase activities were over-represented based on molecular functions annotated based on Gene Ontology (GO). These extracellular proteases represent different hydrolase families and substrate profiles. **e**, The number of amino acids between the membrane anchor (either the proximal transmembrane helix or GPI-anchor) and the cleavage site was used to approximate how far the proteolytic cleavage occurred relative to the membrane.

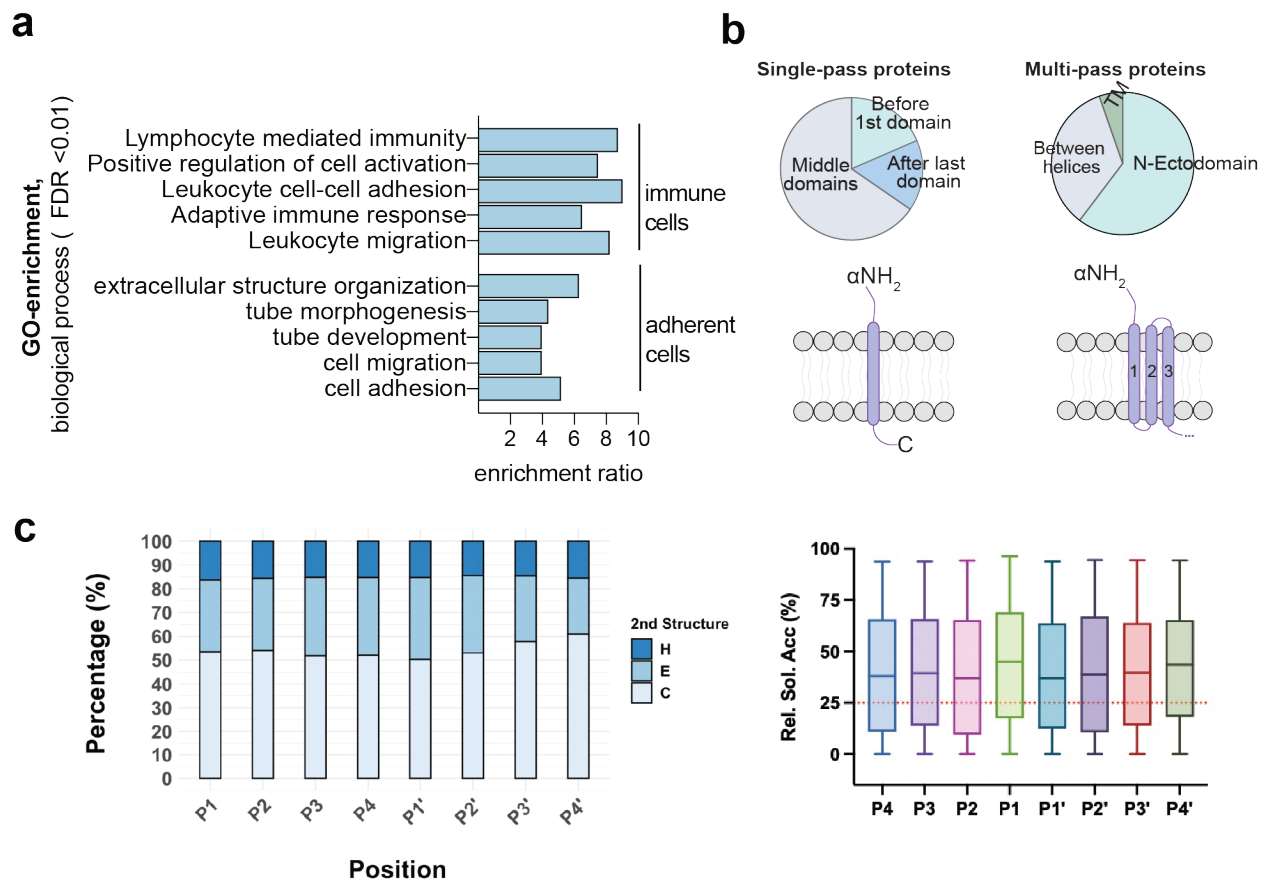


Figure 1.9: Extracellular N-termini distribution across domains, different predicted secondary structure, and solvent accessibility.

a, Gene ontology analysis (GO) of biological processes for extracellular neo-N-termini for adherent cells and immune cell types². **b**, Localization of extracellular cleavage sites relative to domains for single-pass membrane proteins and regions between multi-pass transmembrane helices^{3,4}. **c**, Using PDB deposited structures and structural predictions made by AlphaFold 2.0⁶, predicted secondary structures and surface accessibility of the extracellular cleavage sites were made. For secondary structure assignment, H represents helix, E represents sheets, C represents loops.

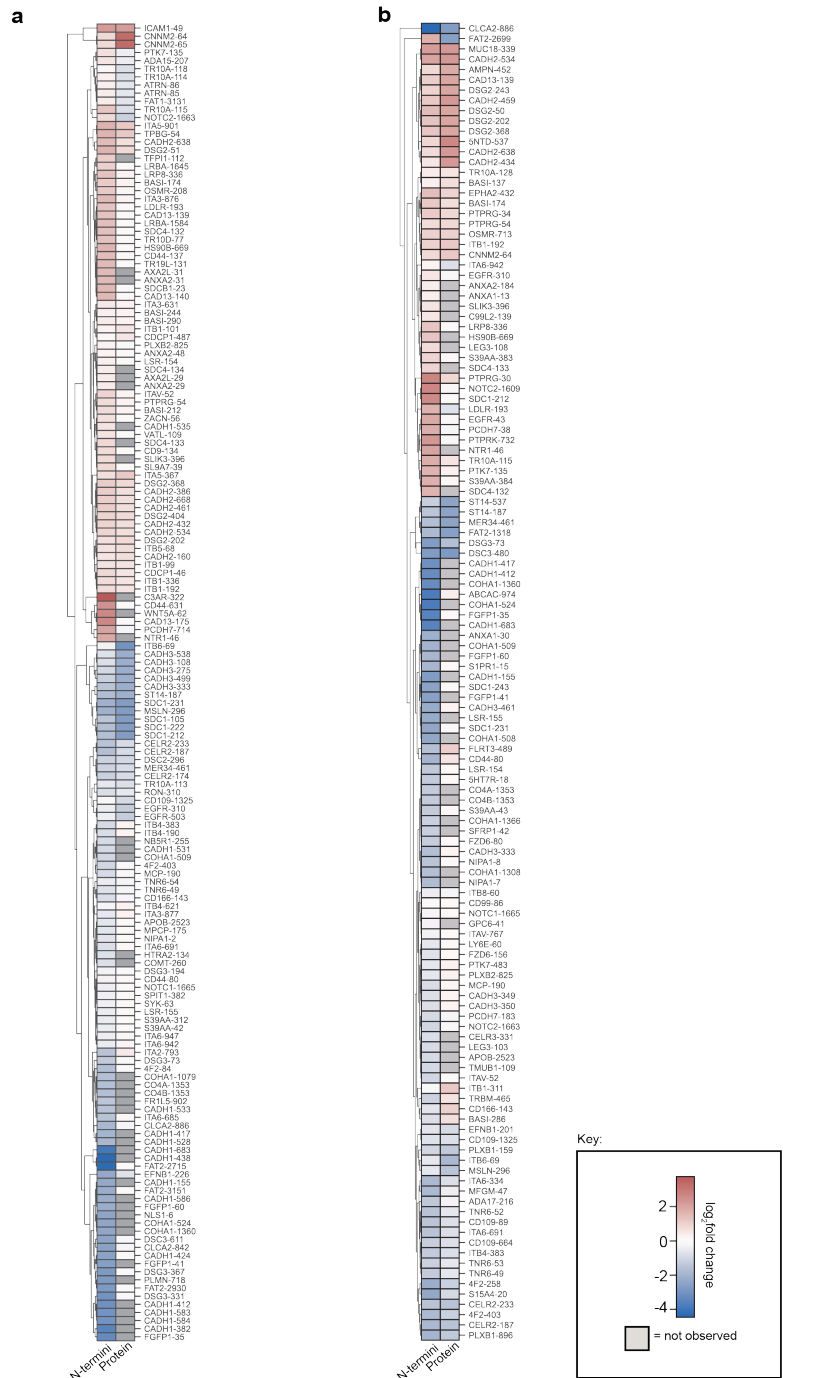


Figure 1.10: Comparisons between N-termini fold changes and protein abundance changes for single oncogene-driven datasets shows modest correlation.

Hierarchical heat maps represent fold-changes (1.8 or higher threshold) in the presence of the oncogenes her2 for panel a, or kras(g12V) for panel b,. The protein and N-termini peptide start position are annotated for each row. For proteins not observed in CSC proteomics¹, the protein abundance row is colored gray to indicate a missing value.

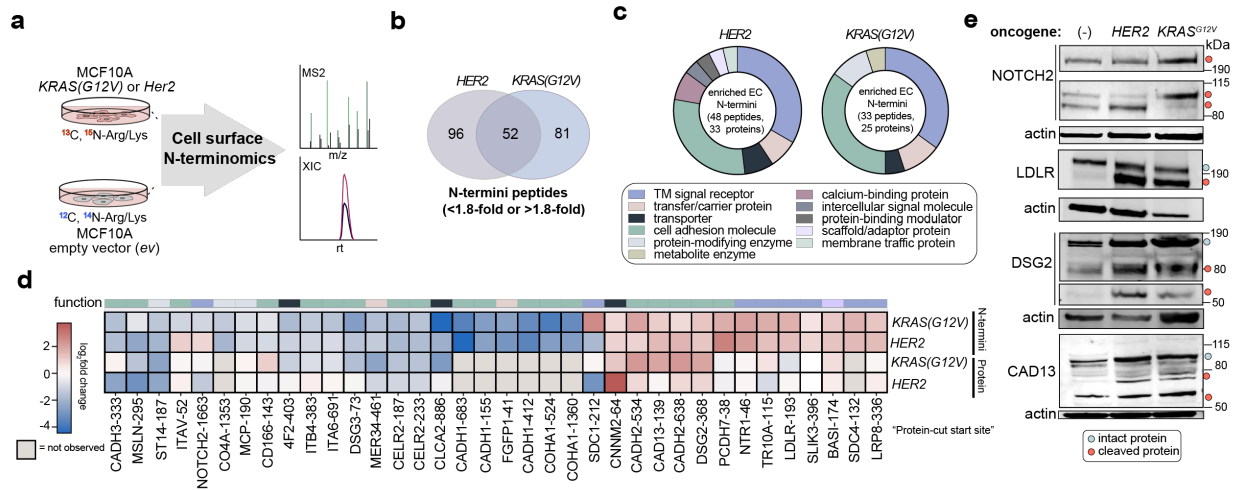


Figure 1.11: Individual oncogenes drive common and unique proteolytic cleavages on cell surfaces.

a, Schematic depicting the application of quantitative GT-stabiligase N-terminomics to identify differences in proteolytic, neo-N-termini in the presence of individual oncogenes, HER2 or KRAS(G12V). After growing cells in SILAC conditions, non-tumorigenic MCF10A cells (empty vector, ev) were combined with cells harboring the single oncogenes for GT-stabiligase N-terminomics as described earlier. **b**, Venn diagram for neo-N-termini for KRAS(G12V) or HER2 transformed cells that were substantially changed (1.8-fold threshold) in comparison to the control MCF10A cells. **c**, Protein classes represented by enriched neo-N-termini observed in the presence of either oncogene shows an enrichment of transmembrane signal receptors and cell adhesion proteins. **d**, A heat map shows comparisons between shared proteolytic neo-N-termini observed in the presence of Her2 or KRas^{G12V} and corresponding CSC-based protein enrichments reported previously⁴¹. **e**, Immunoblot analysis of select proteins (NOTCH2, DSG2, LDLR, CAD13) is consistent with quantitative proteolytic differences observed using GT-stabiligase N-terminomics. Independent experiments were performed in triplicate with similar results. Of note, the S2 and S3 cleavage products of NOTCH2 likely overlap in the blot, and we would not expect to see the S3 cut by cell surface N-terminomics because the neo-N-terminus is intracellular.

References

- 1 Fortelny, N., Yang, S., Pavlidis, P., Lange, P. F. & Overall, C. M. Proteome TopFIND 3.0 with TopFINDER and PathFINDER: database and analysis tools for the association of protein termini to pre- and post-translational events. *Nucleic Acids Res* **43**, D290-297 (2015).
- 2 Weeks, A. M. & Wells, J. A. Subtiligase-Catalyzed Peptide Ligation. *Chem Rev* **120**, 3127-3160 (2020).
- 3 Bausch-Fluck, D. *et al.* The in silico human surfaceome. *Proc Natl Acad Sci U S A* **115** (2018).
- 4 Aebersold, R. *et al.* How many human proteoforms are there? *Nat Chem Biol* **14**, 206-214 (2018).
- 5 Werb, Z. ECM and cell surface proteolysis: regulating cellular ecology. *Cell* **91**, 439-442 (1997).
- 6 Dudani, J. S., Warren, A. D. & Bhatia, S. N. Harnessing Protease Activity to Improve Cancer Care. *Annual Review of Cancer Biology, Vol 2* **2**, 353-376 (2018).
- 7 Griswold, A. R. *et al.* A Chemical Strategy for Protease Substrate Profiling. *Cell Chem Biol* **26**, 901-907 (2019).
- 8 Staes, A. *et al.* Protease Substrate Profiling by N-Terminal COFRADIC. *Methods Mol Biol* **1574**, 51-76 (2017).
- 9 Kleifeld, O. *et al.* Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat Protoc* **6**, 1578-1611 (2011).

- 10 Dix, M. M., Simon, G. M. & Cravatt, B. F. Global identification of caspase substrates using PROTOMAP (protein topography and migration analysis platform). *Methods Mol Biol* **1133**, 61-70 (2014).
- 11 Lichtenthaler, S. F., Lemberg, M. K. & Fluhrer, R. Proteolytic ectodomain shedding of membrane proteins in mammals—hardware, concepts, and recent developments. *EMBO J* **37** (2018).
- 12 Staes, A. *et al.* Selecting protein N-terminal peptides by combined fractional diagonal chromatography. *Nat Protoc* **6**, 1130-1141 (2011).
- 13 Van Damme, P. *et al.* Complementary positional proteomics for screening substrates of endo- and exoproteases. *Nat Methods* **7**, 512-515 (2010).
- 14 Prudova, A. *et al.* TAILS N-Terminomics and Proteomics Show Protein Degradation Dominates over Proteolytic Processing by Cathepsins in Pancreatic Tumors. *Cell Reports* **16**, 1762-1773 (2016).
- 15 Calvo, S. E. *et al.* Comparative Analysis of Mitochondrial N-Termini from Mouse, Human, and Yeast. *Mol Cell Proteomics* **16**, 512-523 (2017).
- 16 Abrahmsen, L. *et al.* Engineering subtilisin and its substrates for efficient ligation of peptide bonds in aqueous solution. *Biochemistry* **30**, 4151-4159 (1991).
- 17 Weeks, A. M. & Wells, J. A. N-Terminal Modification of Proteins with Subtiligase Specificity Variants. *Curr Protoc Chem Biol* **12**, e79 (2020).
- 18 Julien, O. *et al.* Quantitative MS-based enzymology of caspases reveals distinct protein substrate specificities, hierarchies, and cellular roles. *Proc Natl Acad Sci U S A* **113**, E2001-2010 (2016).

- 19 Agard, N. J., Maltby, D. & Wells, J. A. Inflammatory Stimuli Regulate Caspase Substrate Profiles. *Molecular & Cellular Proteomics* **9**, 880-893 (2010).
- 20 Weeks, A. M., Byrnes, J. R., Lui, I. & Wells, J. A. Mapping proteolytic neo-N termini at the surface of living cells. *Proc Natl Acad Sci U S A* **118** (2021).
- 21 Weeks, A. M. & Wells, J. A. Engineering peptide ligase specificity by proteomic identification of ligation sites. *Nat Chem Biol* **14**, 50-57 (2018).
- 22 Wollscheid, B. *et al.* Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat Biotechnol* **27**, 378-386 (2009).
- 23 Zeng, Y., Ramya, T. N., Dirksen, A., Dawson, P. E. & Paulson, J. C. High-efficiency labeling of sialylated glycoproteins on living cells. *Nat Methods* **6**, 207-209 (2009).
- 24 Masamune, S. *et al.* Bio-Claisen Condensation Catalyzed by Thiolase from *Zoogloea Ramigera* - Active-Site Cysteine Residues. *Journal of the American Chemical Society* **111**, 1879-1881 (1989).
- 25 Wang, S. J. *et al.* Saline Accelerates Oxime Reaction with Aldehyde and Keto Substrates at Physiological pH. *Scientific Reports* **8** (2018).
- 26 Debets, M. F. *et al.* Metabolic precision labeling enables selective probing of O-linked N-acetylgalactosamine glycosylation. *Proc Natl Acad Sci U S A* **117** (2020).
- 27 Mockl, L. *et al.* Quantitative Super-Resolution Microscopy of the Mammalian Glycocalyx. *Dev Cell* **50**, 57-72 e56 (2019).
- 28 Hong, V., Steinmetz, N. F., Manchester, M. & Finn, M. G. Labeling live cells by copper-catalyzed alkyne-azide click chemistry. *Bioconjug Chem* **21**, 1912-1916 (2010).
- 29 The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **45**, D158-D169 (2016).

- 30 Madala, P. K., Tyndall, J. D. A., Nall, T. & Fairlie, D. P. Update 1 of: Proteases Universally Recognize Beta Strands In Their Active Sites. *Chemical Reviews* **110**, Pr1-Pr31 (2010).
- 31 Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* **47**, W199-W205 (2019).
- 32 Bausch-Fluck, D. *et al.* A mass spectrometric-derived cell surface protein atlas. *PLoS One* **10**, e0121314 (2015).
- 33 Boon, L., Ugarte-Berzal, E., Vandooren, J. & Opdenakker, G. Protease propeptide structures, mechanisms of activation, and functions. *Crit Rev Biochem Mol Biol* **55**, 111-165 (2020).
- 34 Kopan, R. & Ilagan, M. X. The canonical Notch signaling pathway: unfolding the activation mechanism. *Cell* **137**, 216-233 (2009).
- 35 Huang, H. Proteolytic Cleavage of Receptor Tyrosine Kinases. *Biomolecules* **11** (2021).
- 36 Fitzgerald, M. L., Wang, Z. H., Park, P. W., Murphy, G. & Bernfield, M. Shedding of syndecan-1 and-4 ectodomains is regulated by multiple signaling pathways and mediated by a TIMP-3-sensitive metalloproteinase. *Journal of Cell Biology* **148**, 811-824 (2000).
- 37 Mahmood, N., Mihalcioiu, C. & Rabbani, S. A. Multifaceted Role of the Urokinase-Type Plasminogen Activator (uPA) and Its Receptor (uPAR): Diagnostic, Prognostic, and Therapeutic Applications. *Front Oncol* **8**, 24 (2018).
- 38 Atapattu, L., Lackmann, M. & Janes, P. W. The role of proteases in regulating Eph/ephrin signaling. *Cell Adh Migr* **8**, 294-307 (2014).
- 39 Tseng, C. C. *et al.* Matriptase shedding is closely coupled with matriptase zymogen activation and requires de novo proteolytic cleavage likely involving its own activity. *Plos One* **12** (2017).

- 40 Gherardi, E., Love, C. A., Esnouf, R. M. & Jones, E. Y. The sema domain. *Curr Opin Struct Biol* **14**, 669-678 (2004).
- 41 Leung, K. K. *et al.* Broad and thematic remodeling of the surfaceome and glycoproteome on isogenic cells transformed with driving proliferative oncogenes. *Proc Natl Acad Sci U S A* **117**, 7764-7775 (2020).
- 42 Martinko, A. J. *et al.* Targeting RAS-driven human cancer cells with antibodies to upregulated and essential cell-surface proteins. *Elife* **7** (2018).
- 43 Berx, G. & van Roy, F. Involvement of Members of the Cadherin Superfamily in Cancer. *Cold Spring Harbor Perspectives in Biology* **1** (2009).
- 44 Kamekura, R. *et al.* Inflammation-induced desmoglein-2 ectodomain shedding compromises the mucosal barrier. *Mol Biol Cell* **26**, 3165-3177 (2015).
- 45 Banerjee, S. *et al.* Proteolysis of the low density lipoprotein receptor by bone morphogenetic protein-1 regulates cellular cholesterol uptake. *Sci Rep* **9** (2019).
- 46 Lambrecht, B. N., Vanderkerken, M. & Hammad, H. The emerging role of ADAM metalloproteinases in immunity. *Nat Rev Immunol* **18**, 745-758 (2018).
- 47 Okamoto, I. *et al.* CD44 cleavage induced by a membrane-associated metalloprotease plays a critical role in tumor cell migration. *Oncogene* **18** (1999).
- 48 Sugiyama, N. *et al.* EphA2 cleavage by MT1-MMP triggers single cancer cell invasion via homotypic cell repulsion. *J Cell Biol* **201**, 467-484 (2013).
- 49 Koshikawa, N. *et al.* Proteolysis of EphA2 Converts It from a Tumor Suppressor to an Oncoprotein. *Cancer Res* **75**, 3327-3339 (2015).
- 50 Baker, A. T., Zlobin, A. & Osipo, C. Notch-EGFR/HER2 Bidirectional Crosstalk in Breast Cancer. *Front Oncol* **4**, 360 (2014).

- 51 Meurette, O. & Mehlen, P. Notch Signaling in the Tumor Microenvironment. *Cancer Cell* **34**, 536-548 (2018).
- 52 Kryza, T. *et al.* Substrate-biased activity-based probes identify proteases that cleave receptor CDCP1. *Nature Chemical Biology* **17**, 776-783 (2021).
- 53 Lim, S. A. *et al.* Targeting a proteolytic neopeptide on CUB domain containing protein 1 (CDCP1) for RAS-driven cancers. *J Clin Invest* **132** (2022).

Chapter 2

Hypoxia induces extensive proteolytic and protein remodeling of the cell surface in pancreatic adenocarcinoma (PDAC)

Abstract

The tumor microenvironment (TME) plays a crucial role in cancer progression. Hypoxia, a hallmark of the TME, induces a cascade of molecular events that affect cellular processes involved in metabolism, metastasis, and proteolysis. In pancreatic adenocarcinoma (PDAC), tumor tissues are extremely hypoxic, and it is important to understand how it can affect cellular expression and function. Here, we utilized various mass spectrometry methods to analyze how proteolysis and protein secretion changed under hypoxic conditions. Four cell lines were assayed and we found that changes in proteolysis affected molecules involved in cellular adhesion and motility. Under hypoxic conditions, PDAC cells secreted/shed fewer molecules responsible for the regulating the humoral immune and inflammation response, while there was an upregulation in proteins involved in metabolic processing and tissue development. While proteolysis did not directly contribute to changes in the protein levels in the secretome, the two methods played complementary roles in exploring the breadth of the effect of hypoxia on the cell surface landscape.

Introduction

Pancreatic ductal adenocarcinoma (PDAC) represents more than 90% of pancreatic cancers. Over 50% of patients are diagnosed in the advanced stages of the disease when the tumor cells have metastasized¹. A lack of detectable biomarkers and non-specific symptoms results in a delay in diagnosis and treatment. While treatments for many cancers have improved over the past few years, the 5-year survival rate for pancreatic cancer remains low at 12.5%¹. Understanding the cellular physiology of PDAC and its environment may lead to improved efforts to treat the disease.

The tumor microenvironment (TME) has a major impact on cancer progression, including effects on proliferation, invasion, metabolism, angiogenesis, metastasis, and immunosuppression².³ One characteristic of the TME is hypoxia, caused by the uncontrolled cellular proliferation and

changes to metabolism^{4,5}. The hypoxic environment can then cause additional changes to multiple regulatory processes of the cellular population, some of which may confer resistance to treatments⁴⁻⁶. It can also suppress the immune response by creating a hostile environment for immune cells⁷⁻⁹. Most tumor tissues range from 2-10 times more hypoxic than their respective normal tissues⁴, but pancreatic tumors have been found to be as much as 25 times more hypoxic^{4,10}.

Another hallmark of cancer is the dysregulation of proteolysis¹¹. Proteolysis is a post-translational modification that contributes to many cellular and physiological processes, including cell signaling, protein processing, tissue remodeling, cellular migration, and programmed cell death to name a few¹²⁻¹⁴. However, aberrant proteolysis can contribute to the progression of disease by affecting cellular growth, responses to apoptosis and senescence, angiogenesis, invasion, metastasis, and inflammation^{15,16}. In order to study the role of proteolysis in biological processes, various mass spectrometry (MS) methods have been developed to identify cleavage sites¹⁷⁻²² and shed or secreted proteins (secretome)²³⁻²⁶.

Previous studies have studied the role of hypoxia in PDAC and its effect on various cellular phenotypes. It is known to promote migration via *miR-150* downregulation²⁷ and can contribute to the endothelial mesenchymal transition (EMT) and proliferation^{28,29}. While not directly studying hypoxia, RNA analysis of patient samples^{30,31} and MS analysis of extracellular protein expression in mouse xenografts³² have provided insight on the proteomic landscape in hypoxic PDAC tumors. However, we do not know how a hypoxic environment affects proteolysis of extracellular proteins and the shedding and secretion of proteins in PDAC.

Here we utilized a glycan-tethered subtiligase to identify protein N-terminal peptides, including those created from proteolytic cleavage events²² in 4 PDAC cell lines (KP4, Panc-1, PaTu 8902, and MIA PaCa-2) and address how a hypoxic environment impacts the proteolytic

landscape. Additionally, we studied how hypoxia affects the shedding and secretion of proteins into the extracellular space. In total, we identified 906 unique N-termini that mapped to 384 membrane proteins and 517 membrane proteins as part of secretome. Hypoxia induces extensive remodeling of the proteolytic and extracellular protein landscape of PDAC, especially affecting proteins involved cellular adhesion, motility, and response to the innate immune system.

Results

Mass spectrometry based identification of neo N-termini in extracellular proteins in PDAC under hypoxic conditions.

One method to identify protein N-termini utilizes subtiligase, an engineered ligase that preferentially binds N-termini and catalyzes a reaction between it and the C-terminal ester of a peptide substrate, creating a peptide bond. A previously described method²² was utilized to tether subtiligase to the cell surface to explore how hypoxia affects cell surface proteolysis in PDAC. In short (**Fig. 2.1a**), each cell line was cultured in stable isotopic labeling of amino acid (SILAC) media and grown under hypoxic (1% O₂) or normoxic (5% O₂) conditions for 72 hours. Cells were mixed before tethering subtiligase to cell surface proteoglycans via a bis-aminoxy reagent. This mixture was then incubated with a biotinylated peptide substrate containing an aminobutyric acid (Abu) mass tag, labeling the exposed N-termini of proteins. These proteins were then processed until only Abu-tagged peptides remained and were analyzed via mass spectrometry to identify unique peptides across both conditions. Across the 4 cell lines (KP4, Panc-1, PaTu 8902, and MIA PaCa-2) that were assayed, 906 unique peptides, mapping to 384 membrane proteins were identified (**Fig. 2.1b**). Most proteins identified were type I single-pass transmembrane proteins, making up ~58% of proteins. The next largest group of N-termini identified correlated with multi-pass proteins (22%), followed by secreted proteins (10%), glycosphosphatidylinositol (GPI)-

anchored proteins (7.4%), and finally, type II transmembrane proteins (1.6%). Analysis of the cleavage topology (**Fig. 2.1c**) revealed that over 88% of the identified peptides could be mapped to the extracellular domain of proteins, while just a small number were associated with the removal of signal peptides, initiator methionines, pro-domains, or cleavages in the transmembrane domain. Using PDB deposited structures (**Fig. 2.1d**) and AlphaFold 2.0 predicted structures (**Fig. 2.2**), the secondary structure surrounding these cleavage sites correlated mostly with loops/unstructured domains. We also measured the distance between the identified N-termini and where the proteins were tethered to the membrane (**Fig 2.1e**) and found a distance distribution indicating that while many cleavages correlate with the shedding of whole protein ectodomains (0-40 amino acids), a significant portion occurred away from the membrane.

Hypoxia causes widespread changes in proteolytic activity

Analysis of the identified N-termini did not show a general consensus in proteolytic activity across the 4 cell lines (**Fig. 2.3a**, n=5). Each cell line demonstrated its own proteolytic profile in response to hypoxia. Gene ontology (GO) analysis of biological processes of all peptides (**Fig. 2.3b**) revealed an over-representation of proteins involved in receptor activity: receptor tyrosine kinase (RTK) signaling, transducer and transmembrane signaling receptor activity. We identified 84 neo N-termini that had a greater than or equal to 1.8-fold change in abundance under hypoxic conditions in at least 2 cell lines (**Fig. 2.4a**). In this enriched dataset, the presence of many N-termini were lower under hypoxic conditions in the KP4, Panc-1, and PaTu 8902 lines. MIA PaCa-2 cells demonstrated an increase in proteolysis at the same cut sites. Western blot analysis of the full length (fl) and cleaved (c) forms of MUC18, RTN4R, and CDCP1 were consistent with mass spectrometry analysis (**Fig. 2.4b**). With several proteins, there were some N-termini that differed only by a few amino acids. For example, MUC18 cleavage is increased under hypoxic condition

at residue 338 in MIA PaCa-2 cells, while showing slightly lower levels in the other cell lines. However, in KP4 and Panc-1 cells, a cleavage at residue 339 becomes more prominent and results in the same sized cleaved product identified by western blot. It is the combination of data for these “tandem” cut sites that can provide a clearer picture on the presence of such cleavage products. An over-representation of proteins involved in cell adhesion, migration, and locomotion were found based on GO analysis (**Fig. 2.4c**). Hypoxia is known to affect cellular adhesion and migration³³⁻³⁶ and this data suggests that it may also impact proteolysis of such molecules.

Hypoxia induces changes in the secreted proteome

Studying N-termini and cleavage sites provided some insight on how the cell surface proteome changes in hypoxic PDAC, however, it does not paint a full picture of the changes, particularly with the case of secreted proteins. The residence time of these proteins near the cell surface may not be long enough to enable subtiligase labeling of their N-termini. **Fig. 2.5a** briefly outlines the method²³ used to analyze proteins secreted or shed into the supernatant. Cells are cultured in SILAC media with N-azidoacetymannosamine in normoxic or hypoxic conditions to label glycoproteins with an azide handle. The supernatant from these cultures is then collected and the free-floating glycoproteins enriched via lectin-based pull down and click chemistry. The azido-labeled proteins are then processed for mass spectrometry. Similar to the N-terminomics analysis, a majority of the 517 proteins identified in the supernatant were type I single-pass transmembrane proteins (**Fig. 2.5b**) and as expected, a larger number of secreted proteins was identified via this method. A global look at the identified proteins (**Fig 2.5c**) reveals three clusters of proteins across the four cell lines. Cluster 1 (**Figs. 2.5d and 2.6a**) represents a group of heavily downregulated proteins that with GO analysis, show to be involved in inflammatory and immune responses. Cluster 2 (**Figs. 2.5d and 2.6b**) includes proteins that are slightly upregulated and are over-

represented in molecules that regulate cellular metabolism. Cluster 3 (**Figs. 2.5d and 2.6c**) show an increased number of shed/secreted molecules involved in antigen processing and presentation as well as those involved in endothelium and tissue development. Collectively, this data corroborates existing data that hypoxic conditions can aid in immune evasion^{37,38} and affect cellular metabolism to promote growth of the diseased cells³⁹⁻⁴¹.

Proteolytic activity does not directly correlate with the presence of shed and secreted proteins

Thirty-six membrane proteases were identified in the secretome data set (**Fig. 2.7a**). Serine proteases accounted for almost half of the proteases identified, followed by metalloproteases and cysteine proteases (**Fig. 2.7b**). Many of the upregulated proteases, such as ADAM10, ADAM17, the Cathepsins, MMP2, etc are known to be involved in extracellular matrix degradation to increase tissue invasion. The downregulated proteases are involved in coagulation and angiogenesis. These proteases are primarily known to be secreted or associated with the cellular membrane, but hypoxia induces the secretion of some that normally reside in the cytoplasm, lysosome, endoplasmic reticulum, or golgi apparatus (**Fig. 2.7c**). IceLogo visualization (**Fig. 2.7d**) does not converge on a definitive consensus sequence, as expected with the breadth of proteases identified here, but the distinct presence of arginine and lysine in the P1 position reflects the number of the identified proteases known to cleave with those amino acids in that position. Eleven proteases were identified in both the N-terminomics and secretomics datasets (**Fig. 2.8a**), with activating cut sites for ADAM10 and Urokinase were observed (**Fig. 2.8b**). Side by side comparison of the enriched proteolysis sites with their respective proteins does not directly correlate with the protein levels identified in the secretomics analysis (**Fig. 2.7e**). Hierarchical clustering groups the neo N-termini datasets separately from the secretomics datasets, and a lack

of correlation when the full datasets are plotted against each other (**Fig. 2.9**) suggest that the effect of hypoxia on proteolysis may not contribute greatly to changes in the proteins shed or secreted.

Discussion

Hypoxia is an important feature of the tumor microenvironment and greatly impacts the cellular and molecular functions of cells. In the case of a disease like PDAC, where tumor cores can be nearly 25-times more hypoxic than normal tissue^{4,10}, it is important to understand how it can affect protein expression and processing. Utilizing various mass spectrometry techniques, we discovered that hypoxia strongly impacts proteolysis, protein shedding, and protein secretion, but that the changes in proteolysis did not heavily alter changes in protein secretion. Through N-terminomics, we identified 906 peptides associated with 384 membrane proteins, with type I single-pass transmembrane proteins being the most abundant (**Fig. 2.1b**) and most cleavages occurred in the unstructured loops of protein ectodomains (**Figs. 2.1c, 2.1d**). These results reflected a similar pattern to those previously identified in breast cancer cells under normoxic conditions²², indicating that hypoxia does not change the overall proportions of cell surface proteins.

Cell surface N-terminomics profiles for each cell line varied greatly under hypoxic conditions (**Fig. 2.3a**). Like most cancers, PDAC is a complex disease with variable cell types^{30,42} and the cell lines chosen for this study reflect that. PaTu 8902 cells express more epithelial differentiation genes, while Panc-1, MIA PaCa-2, and KP4 cells express more basal-like molecules, like those involved in stem-cell and epithelial-to-mesenchymal transition (EMT) markers⁴². Additionally, they carry different KRAS mutations which may contribute to different activation levels of the Ras pathway and response to stimuli⁴³. When peptides were narrowed down to those that have a >1.8-fold enrichment, either in normoxic or hypoxic conditions, it was found

that while MIA-PaCa-2 cells displayed increased proteolysis at these sites, the other three cell lines showed a general decrease in proteolysis. GO analysis revealed an over-representation of proteins involved in cellular adhesion and migration (**Fig. 2.3c**). In particular, the enrichment ratio for molecules involved in cell adhesion mediated by integrins was over 45. Included in this category was ITGB1^{44,45}, ITGAV⁴⁵, and ITGB5⁴⁵, proteins known to be upregulated in expression under hypoxic conditions to promote cellular migration. Downregulating the cleavage of these molecules may be integral in maintaining that specific function. It has also been shown that cleavage of cell adhesion molecules can contribute to signal transduction^{46,47}. The proteolytic ectodomain cleavage of CADH2 is involved in signal transduction and the degradation of CREB-binding protein (CBP), a transcriptional coactivator⁴⁷. Exposure to hypoxia may trigger a different response in MIA PaCa-2 cells that results in increased proteolysis for signaling rather than adhesion. The western blot validation of RTN4R in particular demonstrates how hypoxia can push proteolysis in opposite directions in different cell lines (**Fig. 2.3b**). We observed an increase in the cleavage product in PaTu 8902 cells while proteolysis was completely inhibited in the other three cell lines.

We also studied how hypoxia affected what proteins were shed or secreted into the media. 517 membrane proteins were identified in the culture supernatant and as with the N-terminomics data, type I single-pass transmembrane proteins dominated the IDs. As expected with a method tailored to isolating untethered glycoproteins or fragments, an increased portion of the IDs were secreted proteins (**Fig. 2.5b**). This dataset converged across the four cell lines to reveal three clusters of proteins. In cluster 1 (**Fig. 2.5d**), GO analysis revealed an over-representation of proteins involved in regulating inflammation and the humoral immune response. Canonically, proteins in the complement system defend against, however studies have shown that pancreatic cancer increases the expression of complement proteins that can then trigger pro-inflammatory

cytokines^{48,49}. In other contexts, hypoxia further increases expression of complement proteins^{50,51}, but interestingly, we found that these proteins were down regulated, indicating that there are likely additional factors in the tumor microenvironment outside of hypoxia that influence complement expression in PDAC. Cluster 2 shows an upregulation of molecules that impact cellular metabolism. This corroborates existing data that hypoxia upregulates molecules such as TGF- β ⁵² and alpha-enolase⁵³ that are involved in hypoxia tolerance. Cluster 3 shows increased secretion of molecules involve in tissue development and an over-representation of protease inhibitors. Taking a closer look at the secreted proteases (**Fig. 2.7a**) a moderate number that are overexpressed under hypoxic conditions are involved in degrading the extracellular matrix to promote metastasis. Those involved in coagulation are generally downregulated. Interestingly, some of the proteases over expressed in cluster 3 are inhibitors to those found in **Fig. 2.7a**. TFPI1⁵⁴ and APLP2⁵⁵ are inhibitors of coagulation factors X & XI (FA10, FA11); GDN⁵⁶ inhibits thrombin and urokinase; and SPIT1⁵⁷ inhibits matriptase (ST14). APLP2 is also known to be shed by ADAM10 and ADAM17⁵⁸. Although a number of proteases are overexpressed and secreted into the supernatant, it is possible that the presence of the inhibitors is enough to reduce proteolysis seen in **Fig. 2.4a**.

Comparing the N-terminomics and secretomics datasets, there are 11 shared proteases (**Fig. 2.8**). However, only the activating cleavage sites for ADAM10 and urokinase have been identified by subtiligase. While this is not an indication that the other proteases are inactive, this mass spec method can only confirm that ADAM10 and urokinase are active. We also compared the enriched N-terminomics dataset to their respective secreted protein levels (**Fig. 2.7e**) and did not see significant correlation; hierarchical clustering maintained them as separate groups. This was supported by correlation plots of the data sets for each cell line (**Fig. 2.9**). This suggests that the cleavage sites observed do not contribute directly to the protein levels in the supernatant.

Subtiligase only labels N-termini and does not provide comprehensive coverage of all cleavages. Additionally, if entire ectodomains are shed, it is possible the remaining transmembrane domains may be recycled before subtiligase can be used to label the remnants.

Hypoxia greatly impacts the cell surface proteome in PDAC cell lines and is just one component of the tumor microenvironment that continues to be studied. Identification of neo N-termini could lead to the development of new targeted therapies with fewer off-target effects. The study of the secretome allows us to learn about molecules that may interfere with existing drugs, but that could also be potential biomarkers for a disease that currently has few early diagnostic targets.

Methods

Cell Culture

KP4 and MIAPaCa-2 cells were a generous gift from Dr. Rushika Perera (UCSF) PANC-1, and PaTu 8902 were purchased from ATCC (Manassas, Virginia) were cultured in IMDM SILAC media (Thermo Fisher Scientific; Waltham, MA) + 10% dialyzed FBS (GeminiBio, Sacramento, CA) and 1% Penicillin/Streptomycin (Thermo Fisher Scientific) containing L-[¹³C₆, ¹⁵N₂] lysine and L-[¹³C₆, ¹⁵N₄] arginine for heavy labeling (Cambridge Isotope Laboratories, Tewksbury, MA) or L-{¹²C₆, ¹⁴N₂] lysine and L-[¹²C₆, ¹⁴N₄] arginine for light labeling (Sigma-Aldrich, St. Louis, MO). Under normoxic conditions, cells were grown at 37°C, 5% O₂ for 72 hours before harvest. Under hypoxic conditions, cells were grown at 37°C, 1% O₂ for 72 hours before harvest.

Cell Surface Capture

Labeling and capture of cell surface proteins has been described in detail previously ([ref](#)). Cells were harvested with Versene (Thermo Fisher Scientific). Cells were washed with phosphate

buffered saline pH6.5 (Cytiva, Marlborough, MA) and incubated with WGA peroxidase (Vector Labs, Newark, CA) and biotin tyramide (Apex Bio, Houston, TX) at 37°C. Hydrogen peroxide (Sigma-Aldrich) was added and the reaction allowed to proceed for 2 minutes before being quenched and washed with sodium pyruvate (Cytiva). Cells were collected and lysed in RIPA buffer with protease inhibitor cocktail (Sigma-Aldrich). Cell lysate was incubated with neutravidin agarose beads (Thermo Fisher Scientific at 4°C for 30 minutes. Beads were then washed with RIPA buffer, 50 mM PBS + 1M NaCl, and 50 mM ammonium bicarbonate +2M urea buffer. The beads were then processed for tryptic peptide elution and mass spectrometry analysis using the iST desalting kit (Preomics, Martinsried, Germany) per manufacturer instructions.

Secretomics Capture

Secretome analysis has been described in detail previously²³. Cells were incubated under normoxic or hypoxic conditions with 100 μ M N-Azidoacetymannosamine (Thermo Fisher Scientific) for 72 hours. Conditioned media was collected incubated with Concanavalin A Agarose beads (G-Biosciences, Overland, MO) for 2 hours at room temperature. Proteins were eluted from the ConA beads with methyl-alpha-mannopyranoside (Sigma-Aldrich) and incubated with DBCO beads (Vector Laboratories; Newark, CA) overnight at 4°C. Protein disulfide bonds were reduced with DTT (GoldBio, Olivette, MO) and cysteines alkylated with iodoacetamide (Sigma-Aldrich). The protein was digested with sequencing grade trypsin and the tryptic peptides were processed for mass spectrometry analysis using the iST desalting kit per manufacturer instructions.

Cell Surface N-terminomics

N-terminomics was performed as previously described [[ref 22](#)]. Cells were harvested with Versene and treated with sodium periodate (Sigma-Aldrich) to oxidize cell surface glycans. Aminoxy-peg2-stabiligase and analine (Sigma-Aldrich) were then added to cells to tether

stabiligase to the cell surface for 15 minutes at 4°C. After washing, cells were incubated at room temperature with a biotinylated peptide ester substrate for 15 minutes and washed with PBS. Cells were collected and lysed in RIPA buffer with protease inhibitor cocktail. Cell lysate was incubated with neutravidin agarose beads at 4°C for 30 minutes. Beads were then washed with RIPA buffer, 50 mM PBS + 1M NaCl, and 50 mM ammonium bicarbonate +2M urea buffer. The beads were then incubated with sequencing grade trypsin (Promega, Madison, WI) overnight at room temperature to release tryptic peptides. Beads were collected and washed with 50 mM ammonium bicarbonate buffer and incubated with TEV protease and DTT overnight. The TEV protease elution was collected and processed for mass spectrometry analysis using the iST desalting columns per manufacturer instructions.

Liquid chromatography mass spectrometry analysis

200 ng of prepared samples were injected onto a 25 cm, ReproSil c18 1.5 μ M 100 A column (PepSep) on a timsTOF Pro with a Captive Spray source and a nenoElute line (Bruker; Hamburg, Germany). a stepwise linear gradient method with H₂O in 0.1% Formic acid and acetonitrile with 0.1% formic acid (solvent B): 5-30% solvent B for 90 min at 0.5 μ l/min, 30-35% solvent B for 10 min at 0.6 μ l/min, 35-95% solvent B for 4 min at 0.5 μ l/min, 95% hold for 4 min at 0.5 μ l/min) was used. Acquired data was collected in a data-dependent acquisition (DDA) mode with ion mobility activated in PASEF mode. Data was then analyzed using PEAKS online Xpro 1.6 (bioinformatics Solutions Inc.; Ontario, Canada), using the SwissProt GOCC Plasma Membrane Database. Analyzed data was further processed using a python script previously described²².

Western Blots

20 μ g of cellular lysate were loaded onto NuPAGE 4%-12%, Bis Tris Mini Protein Gels (Thermo Fisher Scientific). Gels were then transferred to PVDF membranes using the iBlot2

(Thermo Fisher Scientific). Membranes were blocked with 5% bovine serum albumin (BSA; Gemini Bio) for 1 hour at room temperature. Primary antibodies were added to the blocking buffer and incubate at 4°C overnight. β -actin (1:2000), CDCP1 (1:1000), and MUC-18 (1:500) antibodies were purchased from Cell Signalling Technology (Danver, MA). The RTN4R (1:250) antibody was purchased from Proteintech (Rosemont, IL). Membrane was then washed 3X with PBST and incubated with LI-COR secondary antibodies (Lincoln, NE) for 1 hour at room temperature. Membrane was washed three times and imaged on the Odyssey DLx system (LI-COR).

Figures and Tables

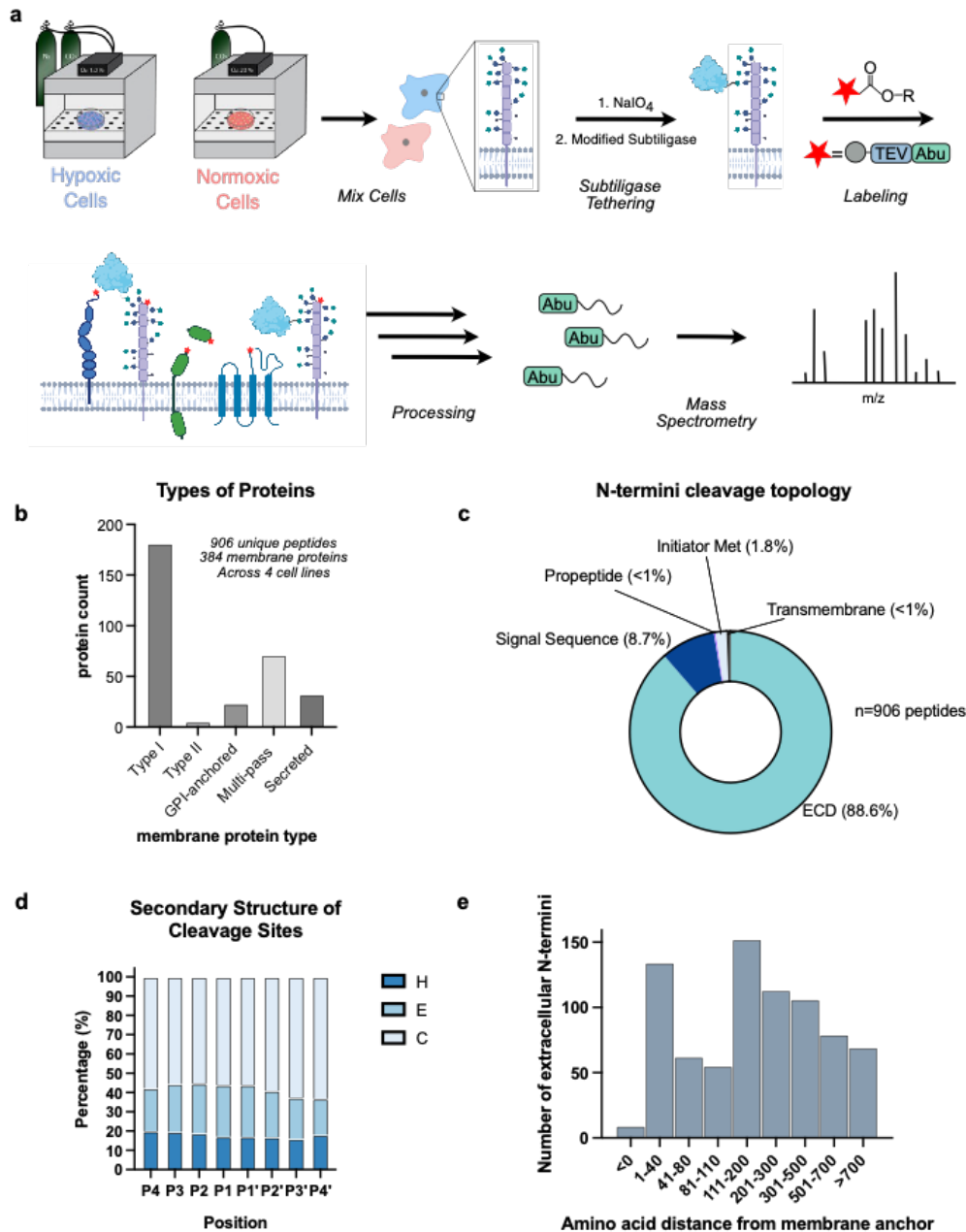


Figure 2.1: Subtiligase labeling of cell surface N-termini on PDAC cell lines under normoxic and hypoxic conditions

a. Schematic of procedure to identify N-termini. PDAC cells are grown in heavy or light SILAC media under normoxic or hypoxic conditions. Cells were then mixed and subtiligase tethered to the cell surface via attachment to proteoglycans. The mixture was incubated with a peptide substrate that label N-termini with an Aminobutyric Acid (Abu) mass tag. Cells were then processed, resulting in peptides with an Abu tag that were then quantified on the timsTOF mass spectrometer. **b.** Across four cell lines, 906 unique N-termini were identified, corresponding to 384 membrane proteins. (Figure caption continued on the next page.)

(Figure caption continued from the previous page.) Of those annotated in UniProt, the majority of proteins identified were Type I single-pass transmembrane proteins, followed by a number of multi-pass and secreted proteins. **c.** Neo N-termini cleavage topology was analyzed and over 88% of peptide mapped to extracellular domains of proteins. Cleavages were also identified at the initiator methionine, propeptide domains, signal sequences, and transmembrane domains. **d.** Using PDB deposited structures, predicted secondary structures for cleavage sites were made. H represents helix, E represents sheets, and C represents loops, with the majority of cleavages occurring in loops or unstructured areas. **e.** Distances between the cleavage site and the cell membrane were approximated based on the number of amino acids between the cleavage site and the membrane anchor.

Secondary Structure of Cleavage Sites (AlphaFold)

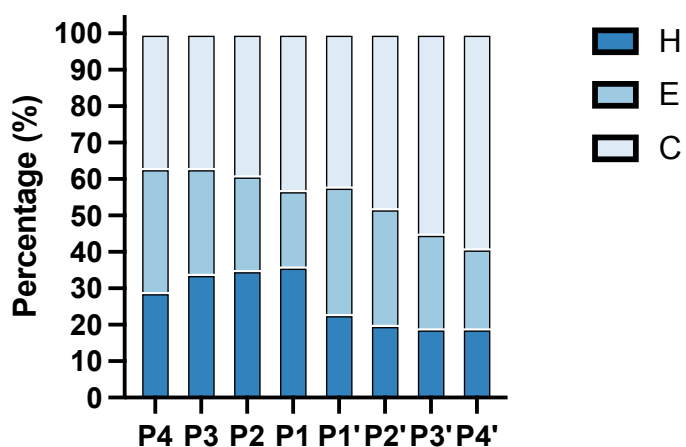


Figure 2.2: Predicted secondary structure at cleavage sites

For proteins that did not have a structure deposited on PDB, AlphaFold 2.0 models were used to predict the secondary structure surrounding the identified cleavage sites. H represents helices, E represents residues in sheets, and C represents residues in loop and unstructured locations.

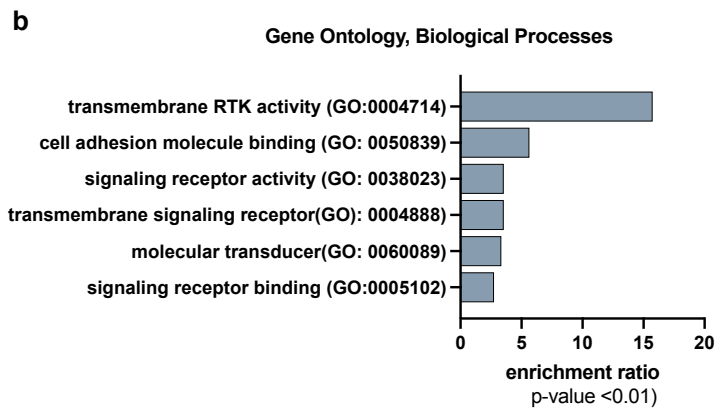
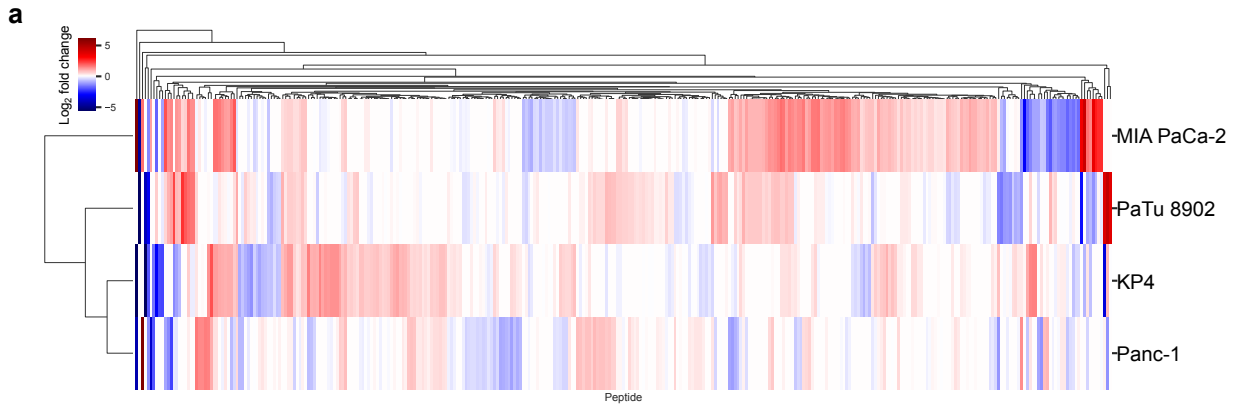


Figure 2.3: N-termini identified by subtiligase

a. Heatmap of identified N-termini depicting the change in peptide abundance under hypoxic conditions. Red indicates an increased presence in hypoxia while blue indicates an increased presence in normoxia. **b.** Go analysis of all N-termini reveals an over-representation of proteins involved in signaling receptor activity.

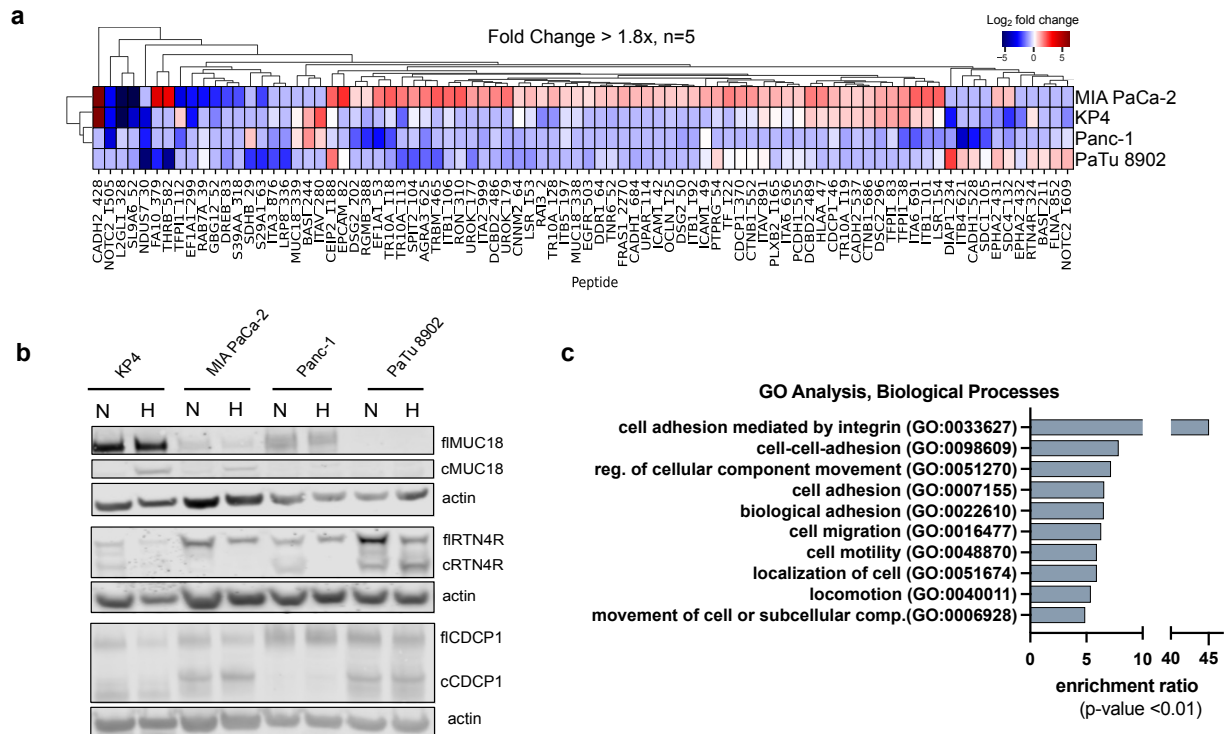


Figure 2.4: Hypoxic conditions induce extensive proteolytic remodeling on the cell surface.
a. Heatmap of significant N-termini with ≥ 1.8 -fold change under hypoxic conditions. Peptides overrepresented under hypoxic conditions are represented in red while those with a higher presence in normoxic conditions are represented in blue. In KP4, Panc-1, and PaTu 8902 cells, proteolysis is decreased under while MIA PaCa-2 cells showed an the opposite effect. **b.** Western blot analysis of full length (fl) and cleaved (cl) isoforms of MUC18, RTN4R, and CDCP1 were consistent with quantitated proteolytic events identified by mass spectrometry. **c.** Gene ontology (GO) analysis of biological processes showed that cleavages affecting proteins involved in cellular adhesion and migration were overrepresented under hypoxic conditions.

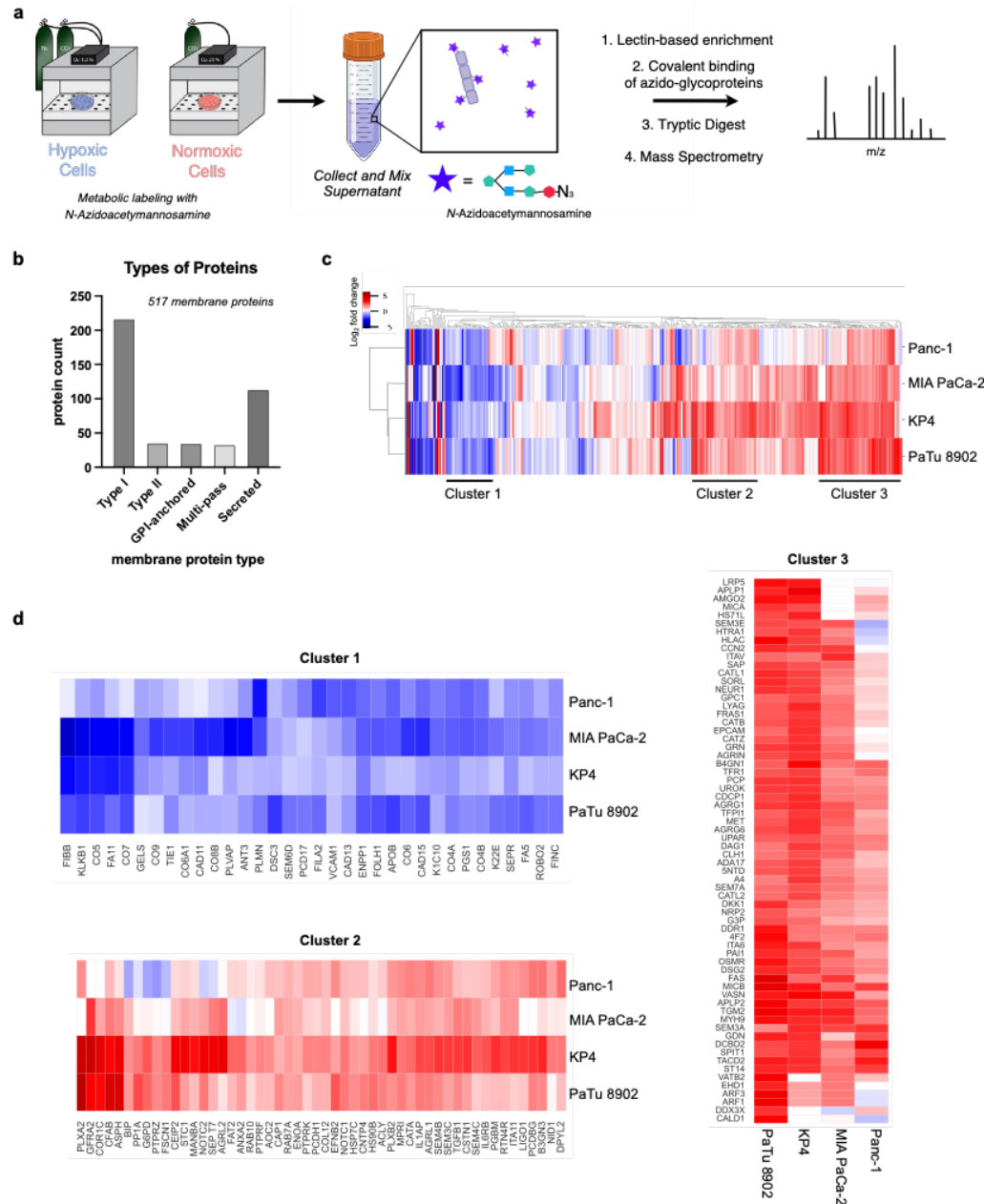


Figure 2.5: Analysis of the secretome of PDAC cell lines

a. Simplified schematic of the method to analyze secreted proteins. Cells are grown in heavy or light SILAC media under normoxic or hypoxic conditions and incubated with *N*-azidoacetymannosamine for 72 hours, resulting in azido-labelled proteins. The supernatant from each condition was collected and mixed before lectin-based enrichment isolated all glycoproteins from the supernatant. Click chemistry was then employed to further enrich for azido-glycoproteins. Samples were digested with trypsin and peptides analyzed by mass spectrometry. **b.** 517 membrane proteins were identified, and of those annotated, type I single-pass transmembrane proteins made up the largest portion of identified proteins, followed by secreted proteins, with a similar number of type II, multi-pass, and GPI-anchored identified. (Figure caption continued on the next page.)

(Figure caption continued from the previous page.) **c.** Global heat map analysis across 4 cell lines revealed three clusters of proteins. **d.** Cluster 1 contained proteins with decreased identification under hypoxic conditions in all four cell lines. GO analysis of biological processes (**Fig 2.5**) revealed a decrease in shed proteins involved in inflammatory and immune responses. Cluster 2 identified an increase in proteins involved in regulation of cellular metabolism. Cluster 3 showed an upregulation in proteins in tissue development and antigen processing and presentation.

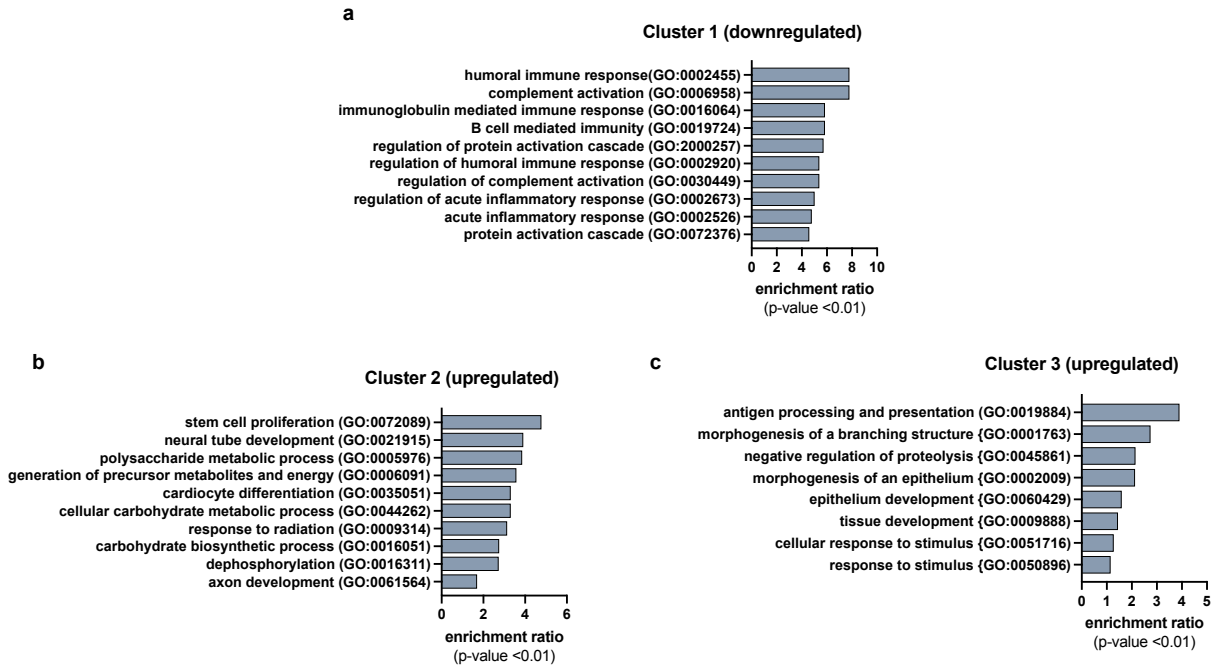


Figure 2.6: GO Analysis of secretomics clusters

a. Cluster 1 identifies downregulated proteins involved in regulating the immune and inflammatory responses. **b.** Cluster 2 identifies moderately upregulated proteins involved in regulating cellular metabolism and development **c.** Cluster 3 identifies upregulated proteins involved in antigen processing/presentation, as well as those involved in tissue development and response to stimuli.

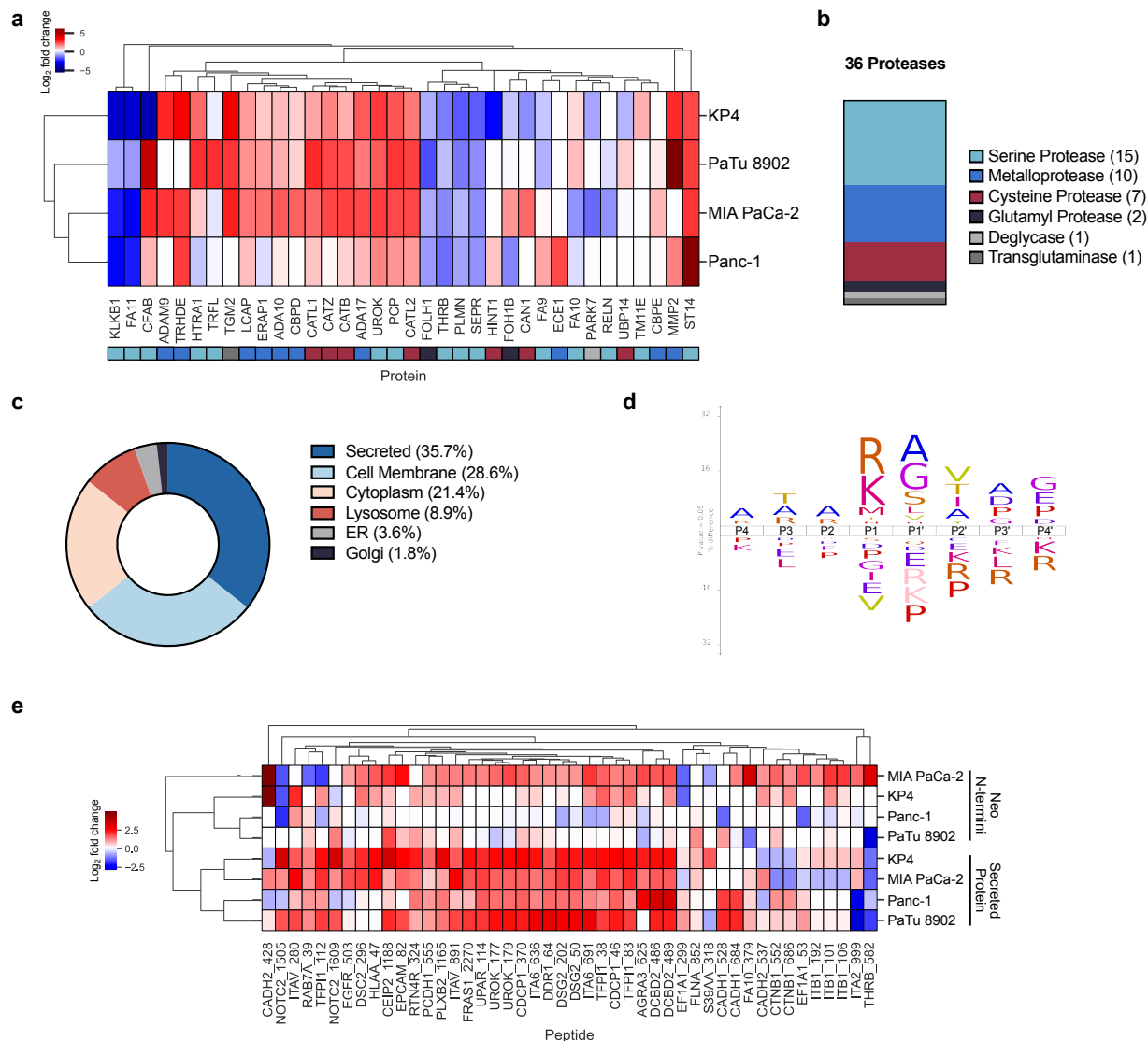


Figure 2.7: Secreted proteases and proteins under hypoxic conditions and correlation of N-termini and secretomics analysis

a. Proteases identified in cellular supernatant under hypoxic conditions across 4 cell lines. Colored tiles beneath the heat map annotate the type of protease, correlating with those in (b). **b.** Classification of identified proteases. **c.** Annotated location(s) for identified proteases according to Uniprot. **d.** iceLogo visualization of P4-P4' residues flanking the cleavage site across four cell lines. **e.** Heatmap comparison between shared N-termini and corresponding secreted protein enrichments.

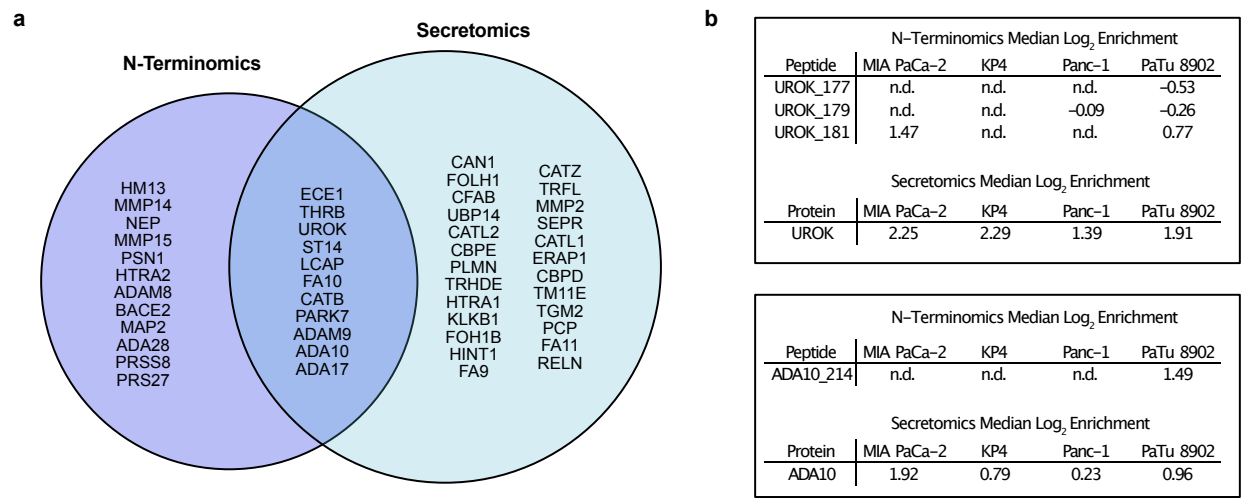


Figure 2.8: Shared proteases identified in N-terminomics and secretomics

a. Venn diagram displaying proteases identified in N-terminomics, secretomics, and shared between the datasets. **b.** Cleavages at activating sites of urokinase and ADAM10 and their associated secreted protein levels.

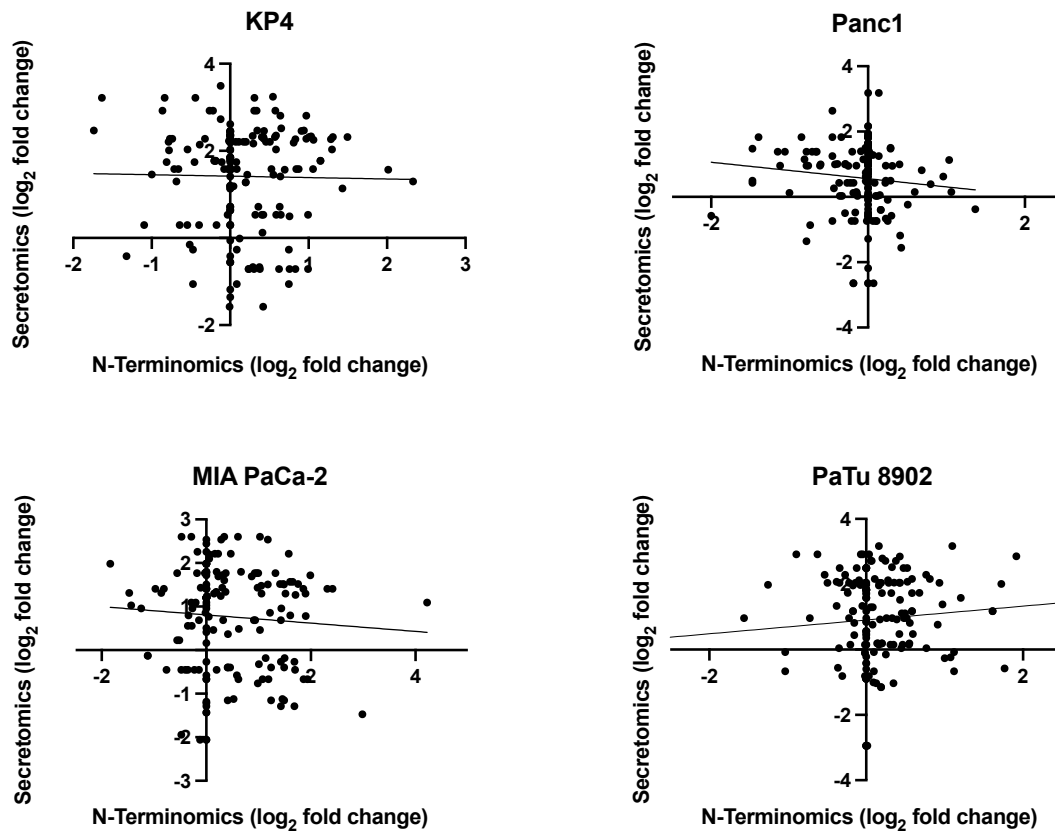


Figure 2.9: Correlation plots between N-terminomics and secretomics data
 Identified N-termini (x-axis) and their changes under hypoxic conditions plotted against their respective change in protein levels in secretomics (y-axis) datasets show little correlation between proteolysis and shed/secreted proteins.

References

1. Siegel, R.L., Gianquinto, A.N., & Jemal, A. Cancer Statistics, 2024. *CA Cancer J. Clin* **74**, 12-49 (2024).
2. Murakami, T., Hiroshima, Y., Matsuyama, R., Homma, Y., Hoffman, R., & Endo, I. *Ann. Gast. Surg.* **2**, 130-137 (2019).
3. Sherman, M.H. & Beatty, G.L. *Annu. Rev. Pathol. Mech. Dis.* **18**,, 123-148 (2022).
4. Muz, B., de la Puente, P., Azab, F., & Azab, A.K. *Hypoxia* **3**, 83-92 (2015).
5. Li, Y., Zhao, L., Li, & X-F. *Technol. Cancer Res. Treat.* **20** (2021).
6. Chen, Z., Han, F.m Du, Y., Shi, H., & Zhou, W. *Sig. Transduct. Target Ther.* **8** (2023).
7. Ruan, K. J. Cellular Biochemistry (2009)
8. Wang, B., Zhao, Q., Zhang, Y., Liu, Z., Zheng, Z., Liu, S., Meng, L., Xin, Y. & Jiang, X. *J. Exp. Clin. Canc. Res.* **40** (2021).
9. Damgaci, S., Ibrahim-Hashim, A., Enriquez-Navas, P.M., Pilon-Thomas, S., Guvenis, A., & Gillies, R.J. *Immunology* **154**, 354-362 (2018).
10. Khouzam, R.A., Goutham, H.V., Zaarour, R.F., Chamseddine, A.N., Francis, A., Buart, S., Terry, S., & Chouaib, S. *Semin. Cancer Biol.* **65**, 140-154 (2020).
11. Koong, A.C., Mehta, V.K., Le, Q.T., Fisher, G.A., Terris, D.J., Brown, J.M., Bastidas, A.J., & Vierra, M. *Int J. Radiat Oncol. Biol. Phys.* **48**, 919-922 (2000).
12. Vizovisek, M., Ristanovic, D., Menghini, S., Christiansen, M.G., & Schuerle, S. *Int J. Mol. Sci.* **22**, 2514-2533 (2021).
13. Turk, B. Turk, D, & Turk, V. *EMBO J.* **4**, 1630-1643 (2012).
14. Werb, Z. *Cell* **91**, 439-442 (1997).
15. Wolf, D.H. & Menssen, R. *FEBS Letters* **592**, 2515-2524 (2018).

16. Dudani, J.S., Warren, A.D., & Bhatia, S.N. *Annu. Rev. Canc. Biol.* **2**, 353-376 (2018).
17. Eatemadi, A., Aiyelabegan, H.T., Negahdari, B., Mazlomi, M.A., Daraee, H., Daraee, N., Eatemadi, R., & Sadroddiny, E. *Biomed. Pharmacother.* **86**, 221-231 (2017).
18. Griswold, A.R., Cifani, P., Rao, S.S., Axelrod, A.J., Miele, M.M., Hendrickson, R.C., Kentsis, A., & Bachovchin, D.A. *Cell Chem. Biol.* **26**, 907-907 (2019).
19. Staes A., Damme, P., Timmerman, E., Ruttens, B., Stes, E., Gevaert, K., & Impens, F. *Methods Mol Biol* **1574**, 51-76 (2017).
20. Prudova, A., Gocheva, B., Auf dem Keller, U., Eckhard, U., Olson, O.C., Akkari, L., Butler, G.S., Fortelny, N., Lange, P.F., Mark, J.C., Joyce, J.A., & Overall, C.M. *Cell Rep.* **16**, 1762-1773 (2016).
21. Dix, M.M., Simon, G.M., & Cravatt, B.F. *Methods Mol Biol.* **1113**, 6-70 (2014).
22. Weeks, A.M. & Wells, J.A. *Curr. Protoc. Chem. Biol.* **12** (2020).
23. Schaefer, K.S., Lui, I., Byrnes, J.R., Kang, E., Zhou, J., Weeks, A.M., & Wells, J.A. *ACS Centr. Sci.* **8**, 1447-1456 (2022).
24. Tüshaus, J., Müller, S.A., Kataka, E.S., Zaucha, J., Monasor, L.S., Su, M., Güner, G., Jocher, G., Tahirovic, S., Frishman, D., Simons, M., & Lichtenthaler, S.F. *EMBO J.* **15** (2020).
25. Meissner, F., Scheltema, R.A., Mollenkopf, H-J., & Mann, M. *Science* **340**, 475-478 (2013).
26. Dieterich, D.D., Link, A.J., Graumann, J., Tirrell, D.A., & Schuman, E.M. *PNAS* **103**, 9482-9487 (2006).
27. Eichelbaum, K., Winter, M., Berriel Diaz, M., Herzig, S., & Krijgsveld, J. *Nat. Biotechnol.* **30**, 984-990 (2012).

28. Sun, J.S., Zhang, X.L., Yang, Y.J., & Nie, Z.G. *Oncol. Lett.* **10**, 835-840 (2015).
29. Jiang, J., Tang, Y.L. & Liang, X.H. *Canc. Biol. Ther.* **11**, 714-723 (2011).
30. Chen, S., Chen, J.Z., Zhang, J.Q., Chen, H.X., Yan, M.L., Huang, L., Tian, Y.F., Chen, Y.L., & Wang, Y.D. *Canc. Lett.* **383**, 73-84 (2016).
31. Collison, E.A., Sadanandam, A., Olson, P., Gibb, W.J., Truitt, M., Gu, S., Cooc, J., Weinkle, J., Kim, G.E., Jakkula, L., Feiler, H.S., Ko, A.H, Olshen, A.B., Danenberg, K.L., Tempero, M.A., Spellman, P.T., Hanahan, D., & Gray, J.W. *Nat. Med.* **17**, 500-503 (2011).
32. Werba, G., Weissinger, D., Kawaler, EA., Zhao, E., Kalfakakou, D., Dhara, S., Wang, L., Lim, H.B., Oh, G., Jing, X., Beri, N.m Khanna, L., Gonda, T., Pberstein, P., Hajdu, C., Loomis, C., Heguy, A., Sherman, M.H., Lund, A.W., Welling, T.H., Dolgalev, I., Tsirigos, A., & Simone, D.M. *Nat. Commun.* **14** (2023).
33. Swietlik, J.J., Bärthel, S., Falcomatà, C., Fink, D., Sinha, A., Cheng, J, Ebner, S., Landgraf, P., Dieterich, D.C., Daub, H., Saur, D., & Meissner, F. *Nat. Commun.* **14** (2023).
34. Lash, G.E., Fitzpatrick, T.E., & Graham, C.H. *Biochem. Biophys. Res. Commun.* **287**, 622-629 (2001).
35. Crossin, K.L. *Cell Adh. Migr.* **6**, 49-58 (2012).
36. Fei, M., Guan, J., Xue, T., Qin, L., Tabg, C., Cui, G., Wang, Y., Gong, H., & Feng, W. *Cell. Mol. Biol. Lett.* **23** (2018).
37. Saxena, K., Jolly, M.K., & Balamurugan, K. *Transl. Oncol.* **13** (2020).
38. Semenza, G.L. *Physiology* **36**, 73-83 (2021).

39. Khouzam, R.A., Zaarour, R.F., Brodaczewska, K., Azakir, B., Venkatesh G.H., Thiery, J, Terry, S., & Chouaib, S. *Front. Immunol.* **13** (2022).
40. Eales, K.L., Hollinshead, K.E.R., & Tennant D.A. *Oncogenesis* **5**, e190 (2016).
41. Kierans, S.J. & Taylor, C.T. *J. Physiol.* **599**, 23-37 (2021).
42. Frezza, C., Zheng, L., Tennant, D.A., Papkovsky, D.B., Hedley, B.A., Kalna, G., Watson, D.G., & Gottlieb, E. *PLoS ONE* **6**, e24411 (2011).
43. Adams, C.R., Htwe, H.H., Mash, T., Wang, A.L., Montoya, M.L., Subbaraj, L., Tward, A.D., Bardeesy, N., & Perera, R.M. *eLife* **8**, e45313 (2019).
44. Johnson, C., Burkhart, D.L., & Halgis, K.M. *Canc. Discov.* **12**, 913-923 (2022).
45. Ju, J.A., Godet, I., Ye, I.C., Byu, J., Jayatilaka, H., Lee, S.J., Xiang, L., Samanta, D., Lee, M.H., Wu, P., Wirtz, D., Semenza, G.L., & Gilkes, D.M. *Mol. Cancer. Res.* **15**, 723-734 (2017).
46. Befani, C. & Liakos, P. *Cell Biol. Intl.* **41**, 769-778 (2017).
47. Nagappan-Chettiar, S., Johnson-Venkatesh, E.M., & Umemori, H. *Neurosci Res.* **116**, 60-69 (2016).
48. Reiss, K., Maretzky, T., Ludwig, A., Tousseyn, T., de Strooper, B., Hartmann, D., & Saftig, P. *EMBO J.* **24**, 742-752 (2005).
49. Afshar-Kharghan, V. *J. Clinic. Invest.* **127**, 780-789 (2017).
50. Hussain, N., Das, D., Pramanik, A., Pandey, M.K., Joshi, V., & Pramanik, K.C. *Cancer Drug Resist.* **5**, 317-327 (2022).
51. Mueller-Buehl, A.M., Buehner, T., Pfarrer, C., Deppe, L., Peters, L., Dick, B.H., & Joachim, S.C. *Cells* **10**, 3575 (2021).

52. Khan, M.A., Shamma, T., Kazmi, S., Altuhami, A., Ahmed, H.A., Assiri, A.M., & Broering, D.C. *J. Transl. Med.* **18**, 147 (2020).
53. Mallikarjuna, P., Zhou, Y., & Landström, M. *Biomolecules* **12**, 635 (2022).
54. Sedoris, K.C., Thomas, S.D., & Miller, D.M. *BMC Cancer* **10**, 157 (2010).
55. Mast, A.E. & Ruf, W. *J. Thromb Haemost.* **20**, 1290-1300 (2022).
56. Xu, F., Previti, M.L., Nieman, M.T., Davis, J., Schmaier, A.H., & Van Nostrand, W.E. *J. Neurosci.* **29**, 5666-5670 (2009).
57. Festoff, B.W., Rap, J.S., & Chen, M. *Neurology* **42**, 1361 (1992).
58. Skovbjerg, S., Holt-Danborg, L., Nonboe, A.W., Hong, Z., Frost, A.K., Schar, C.R., Thomas, C.C, Vitved, L., Jensen, J.K., & Vogel, L.K. *Biochem J.* **477**, 1779-1794 (2020).
59. Endres, K., Postina, R., Schroeder, A., Mueller, U., & Fahrenholz, F. *FEBS J.* **272**, 5808-5820 (2005).

Chapter 3

Trimeric SARS-CoV-2 Spike interacts with dimeric ACE2 with limited intra-Spike avidity

Abstract

A serious public health crisis is currently unfolding due to the SARS-CoV-2 pandemic. SARS-CoV-2 viral entry depends on an interaction between the receptor binding domain of the trimeric viral Spike protein (Spike-RBD) and the dimeric human angiotensin converting enzyme 2 (ACE2) receptor. While it is clear that strategies to block the Spike/ACE2 interaction are promising as anti-SARS-CoV-2 therapeutics, our current understanding is insufficient for the rational design of maximally effective therapeutic molecules. Here, we investigated the mechanism of Spike/ACE2 interaction by characterizing the binding affinity and kinetics of different multimeric forms of recombinant ACE2 and Spike-RBD domain. We also engineered ACE2 into a split Nanoluciferase-based reporter system to probe the conformational landscape of Spike-RBDs in the context of the Spike trimer. Interestingly, a dimeric form of ACE2, but not monomeric ACE2, binds with high affinity to Spike and blocks viral entry in pseudotyped virus and live SARS-CoV-2 virus neutralization assays. We show that dimeric ACE2 interacts with an RBD on Spike with limited intra-Spike avidity, which nonetheless contributes to the affinity of this interaction. Additionally, we demonstrate that a proportion of Spike can simultaneously interact with multiple ACE2 dimers, indicating that more than one RBD domain in a Spike trimer can adopt an ACE2-accessible “up” conformation. Our findings have significant implications on the design strategies of therapeutic molecules that block the Spike/ACE2 interaction. The constructs we describe are freely available to the research community as molecular tools to further our understanding of SARS-CoV-2 biology.

Introduction

In late 2019, a novel, pathogenic coronavirus (SARS-CoV-2) entered the human population and has since spread throughout the world. The number of people suffering from the associated disease (COVID-19) continues to rise, increasing the need for effective therapeutic interventions. SARS-CoV-1 and SARS-CoV-2 Spike proteins are highly homologous (~76% sequence identity). Similar to SARS-CoV-1, the interaction between the SARS-CoV-2 Spike protein and the angiotensin-converting enzyme 2 (ACE2) on human cells is critical for viral entry into host cells¹⁻³. SARS-CoV-2 Spike is an obligate trimer, while ACE2 presents as a dimer on the cell surface⁴. Several high-resolution structures of SARS-CoV-2 Spike receptor binding domain (Spike-RBD) bound to ACE2 have been published⁵⁻⁶. However, as of this writing, structures of SARS-CoV-2 Spike trimer in complex with either the dimeric or monomeric form of ACE2 have not been reported, resulting in an incomplete understanding of the nature of this interaction.

Structural studies of trimeric SARS-CoV-2 and SARS-CoV-1 Spike protein demonstrate that each of the Spike-RBDs, as in other coronaviruses, can undergo hinge-like movements to transition between “up” or “down” conformations. The host ACE2 receptor can only interact with an RBD in the “up” conformation, whereas the “down” conformation is inaccessible to ACE2⁷. The RBDs of SARS-CoV-1 Spike can rotate away from the “down” position by different angles to an “up” position⁸. Several cryo-EM structures report that approximately half of the SARS-CoV-1 and 2 Spike particles are in the “three-down” closed conformation and half in the “one-up” open conformation⁹⁻¹⁰, while another cryo-EM study on SARS-CoV-1 Spike reported 39% of Spike in the “two-up” conformation and 3% in the “three-up” conformation¹¹. The different conformations of Spike observed by these cryo-EM studies may be affected and/or limited by the properties of the grid and sample preparation conditions, and may also reflect differences between

SARS-CoV-1 and 2. Thus, it remains unknown how many RBDs of SARS-CoV-2 are accessible within a trimeric Spike to bind ACE2 under physiological conditions, and thus the degree to which intra-Spike avidity plays in the interaction of ACE2 with SARS-CoV-2 Spike. Recombinant ACE2 and an engineered dimeric ACE2-Fc fusion have been shown in several studies to neutralize SARS-CoV-2 virus¹²⁻¹⁴. However, it remains unknown whether the dimeric form of ACE2 offers any affinity enhancements through avidity compared to a monomer. Understanding the role and mechanism of intra-Spike avidity in binding is important for engineering tight binding antagonists to neutralize virus infection.

To elucidate the nature of the interaction between dimeric ACE2 and trimeric SARS-CoV-2 Spike, we performed a thorough characterization of the binding interactions of the different multimeric forms of Spike-RBD and ACE2 (**Fig. 3.1A**). The results reveal that while both the ACE2 monomer and ACE2-Fc dimer can bind the isolated Spike-RBD, only the ACE2-Fc dimer can bind tightly to the trimeric Spike ectodomain (FL-Spike). Interestingly, the affinity of the ACE2-Fc dimer is much higher to the RBD-Fc dimer than to FL-Spike, suggesting that although intra-molecular avidity plays a role in both interactions, its effect is compromised in the context of FL-Spike. Consistent with this, we find that ACE2 associates more slowly to FL-Spike, which indicates that the RBDs in FL-Spike protein are not readily accessible to ACE2. To further probe the conformational landscape of the RBDs in FL-Spike, we engineered ACE2-Fc as split-Nanoluciferase (NanoLuc) reporters. We found that a proportion of FL-Spike can interact with multiple ACE2-Fc molecules simultaneously, indicating that more than one RBD domain in a Spike trimer can adopt an “up” conformation. Using pseudotyped and SARS-CoV-2 virus neutralization assays, we further show that ACE2-Fc dimer is much more potent at neutralizing virus than ACE2 monomer. These findings extend our biochemical insight into how a Spike trimer

binds to an ACE2 dimer, and have important implications for how multimeric therapeutic molecules, such as dimeric ACE2 or antibody-based biologics, can effectively target SARS-CoV-2.

Results

To study the interaction between ACE2 and SARS-CoV-2 Spike, we constructed a panel of Spike and ACE2 proteins in various multimeric formats (**Fig. 3.1A**). Trimeric SARS-CoV-2 Spike ectodomain (aa 1-1213) (FL-Spike) was expressed and purified as described. The construct generously provided by the F. Krammer Lab has the furin cleavage site removed, a pair of stabilizing mutations and a C-terminal T4 trimerization motif followed by a 6xHis tag added. These modifications have been widely used for structural analysis^{8-11,15}. We designed a dimeric form of the Spike Receptor Binding Domain (aa 328-533) (Spike-RBD-Fc) containing a TEV-cleavable Fc-fusion molecule with a C-terminal Avi tag for biotinylation¹⁶. We also generated a monomeric form of Spike-RBD (aa 328-533) (Spike-RBD-monomer) with a C-terminal TEV-8xHis-Avi. The human ACE2 ectodomain contains a N-terminal peptidase domain (aa 18-614) and a C-terminal dimerization domain (aa 615-740). We designed the monomeric form of ACE2 (aa 18-614) (ACE2-monomer) with a C-terminal TEV-8xHis-Avi tag, and the dimeric form of ACE2 (aa 18-614) (ACE2-Fc) as a TEV-cleavable Fc-fusion molecule with a C-terminal Avi tag¹⁶. All of the ACE2 and Spike proteins were expressed in BirA-ER-expressing Expi293 cells^{17,18}.

The Fc-fusion molecules were purified by Protein A affinity chromatography, and the Spike-RBD-monomer and FL-Spike by Ni-NTA affinity chromatography (**Fig. 3.1B, 3.1C**). However, the ACE2 monomer did not express but was generated instead by TEV release from ACE2-Fc (**Fig. 3.1A, 3.1B**). All these proteins, except the ACE2 monomer were >95% biotinylated during expression (**Fig. 3.1B**), facilitating their use on avidin-functionalized surfaces

and beads. Size exclusion chromatography was performed to confirm the oligomerization state of the different proteins. FL-Spike, Spike-RBD-Fc, ACE2 monomer, and ACE2-Fc all eluted at their expected elution times (**Fig. 3.1D**), indicating successful generation of the different multimeric forms of Spike and ACE2 proteins. The Spike-RBD-monomer eluted later than expected, but further analysis of the associated SEC fractions by SDS-PAGE showed the pure protein at the correct molecular weight (**Fig. 3.2**). Differential scanning fluorimetry (DSF) of ACE2-Fc and ACE2 monomer showed these two proteins had similar T_m values (**Fig. 3.3**).

ACE2 dimerization is important for binding to the trimeric SARS-CoV-2 Spike

To understand how oligomerization affects Spike/ACE2 interaction, we determined the affinity and binding kinetics of the different Spike-RBD and ACE2 proteins by bio-layer interferometry (BLI) (**Fig. 3.4**). Spike-RBD-monomer, Spike-RBD-Fc and FL-Spike were immobilized on Streptavidin or Ni-NTA sensors, and allowed to bind ACE2 monomer or ACE2-Fc in solution. A four-fold avidity effect was observed for the Spike-RBD-monomer/ACE2-Fc interaction ($K_D = 5.5$ nM, **Fig. 3.4D**) compared to the Spike-RBD-monomer/ACE2-monomer interaction ($K_D = 22.4$ nM, **Fig. 3.4A**). In contrast, the Spike-RBD-Fc/ACE2-Fc interaction ($K_D < 10^{-12}$ M, **Fig. 3.4E**) showed >1000-fold increase in binding affinity compared to the Spike-RBD-Fc/ACE2-monomer interaction ($K_D = 13.2$ nM, **Fig. 3.4B**). This dramatic increase in affinity is driven by a massive decrease in the off rate and little change in on rate, which indicates a strong intramolecular two-on-two avidity between Spike-RBD-Fc and ACE2-Fc.

We next tested binding of ACE2 monomer and ACE2-Fc to FL-Spike to determine the affinity and avidity effect. Surprisingly, we found that the binding interaction between ACE2 monomer and FL-Spike is very weak and could not be measured accurately (**Fig. 3.4C**). By contrast, the ACE2-Fc interacts with FL-Spike with a K_D of 5.9 nM, similar to the affinity for the

isolated Spike-RBD-monomer (**Fig. 3.4F**). This suggests that the presence of two ACE2 molecules in close proximity in ACE2-Fc is essential for a productive interaction with FL-Spike, and that a single ACE2 monomer is not sufficient to bind an RBD on FL-Spike. However, the affinity between FL-Spike/ACE2-Fc interaction was substantially less than Spike-RBD-Fc/ACE2-Fc interaction ($K_D < 10^{-12}$ M, **Fig. 3.4E**), suggesting that the high-avidity two-on-two interaction is compromised in the context of FL-Spike. This could be due to geometric/steric constraints or the unique conformations of the RBDs in the FL-Spike context. When FL-Spike is loaded to a much higher density on the BLI sensor (load to 2.0 nm) and probed with ACE2-Fc, we see that avidity can be recovered (**Fig. 3.5A**). This indicates that if Spike is presented at high density, the ACE2-Fc arms can engage two RBDs if neighboring Spike trimers are close enough. In contrast, monomeric ACE2 did not bind FL-Spike strongly even when FL-Spike was loaded until saturation, further demonstrating the importance of ACE2 dimerization for interacting with FL-Spike (**Fig. 3.5B**).

Interestingly, the k_{on} of the FL-Spike/ACE2-Fc interaction (**Fig. 3.4C and 3.4F**) is ~10-fold lower than the interactions between Spike-RBD-monomer or Spike-RBD-Fc with ACE2 or ACE2-Fc (**Fig. 3.4A, 3.4B, 3.4D, and 3.4E**), while the k_{off} is also ~10- to 20-fold lower. The decreased k_{on} suggests that the RBDs in FL-Spike protein may have to undergo a conformational change for binding to ACE2. Previous cryo-EM studies on SARS-CoV-2 FL-Spike show that approximately half of the particles have the three RBD domains in the “down” conformation¹⁰. Additionally, molecular dynamics simulations of SARS-CoV-2 FL-Spike suggest that the RBD exists in a series of conformations between the “down” and “up” states, and any RBD with an angle lower than 52.2° from the body of the trimer are inaccessible to ACE2¹⁹. Moreover, biochemical analysis identified that SARS-CoV-2 Spike-RBD, albeit more potent, is less exposed

than SARS-CoV-1 Spike-RBD²⁰. Our data supports that a significant proportion of the RBDs in SARS-CoV-2 FL-Spike protein are in a “closed” or partially “closed” state inaccessible to ACE2, and the RBD has to open up to allow binding to ACE2. The decreased k_{off} , on the other hand, suggests that the presence of multiple RBDs within the context of a FL-Spike could slow down the dissociation of ACE2-Fc.

Split reporter assays indicate more than one RBD in a Spike trimer can be in the “up” conformation when it binds ACE2

To further investigate the RBD conformational landscape in FL-Spike, we designed a split-NanoLuc system to orthogonally probe the Spike/ACE2 interaction. Split-NanoLuc enzymes, in particular NanoBiT (Promega), have been broadly used to detect protein-protein interactions and to study analyte concentrations²¹. The NanoBiT system is composed of LgBiT and SmBiT. SmBiT is an 11 amino acid peptide which has a low intrinsic affinity to LgBiT ($K_D = 190 \mu\text{M}$), but when SmBiT and LgBiT are in close proximity, the two subunits assemble to form an active luciferase enzyme. To interrogate the interaction between ACE2 and Spike, we engineered ACE2-Fc reporter molecules where SmBiT or LgBiT were fused at the N- or C-termini (**Fig. 3.6A**). All constructs expressed at high yield and purity (**Fig. 3.6B, 3.6C**).

To functionally validate the split reporter system, we immobilized Spike-RBD-Fc on streptavidin magnetic beads at high-density. Incubation with 1 nM of ACE2-Fc-SmBiT and ACE2-Fc-LgBiT, or 1 nM of SmBiT-ACE2-Fc and LgBiT-ACE2-Fc with substrate showed dose-dependent luminescence signal, consistent with an assembled functional split enzyme reporter and intermolecular proximity (**Fig. 3.6D, 3.6E**). We found the N-terminal fusion reporter pair showed higher sensitivity compared to the C-terminal fusion reporter pair (**Fig. 3.7**) suggesting that the

increased entropy from the flexible linker and Fc domain reduces productive luciferase reconstitution.

We next used these split reporters to interrogate the soluble FL-Spike trimer (**Fig. 3.6F, 3.6G**). Increasing concentrations of soluble FL-Spike were incubated with ACE2-Fc-SmBiT/LgBiT or SmBiT/LgBiT-ACE2-Fc, followed by the addition of substrate. The SmBiT/LgBiT-ACE2-Fc reporters showed dose-dependent increase in luminescence signal with 1-10 nM FL-Spike. This result suggests that although FL-Spike cannot form a high-avidity two-on-two interaction with both arms of ACE2-Fc at a time, there is a proportion of FL-Spike proteins with “two-up” or perhaps even “three-up” RBDs that can simultaneously interact with multiple ACE2-Fc molecules. We cannot resolve if these “two-up” or “three-up” conformations are present prior to ACE2 binding, or appear because ACE2 binding induces a conformational change, which enables two or more of the RBD domains to engage in the binding to a second ACE2-Fc domain.

Pseudotyped virus and SARS-CoV-2 virus neutralization assays show that dimeric ACE2-Fc can efficiently block viral entry

To translate these studies to cells, we compared the ability of ACE2 monomer and ACE2-Fc to neutralize SARS-CoV-2 pseudotyped virus to infect cells. Pseudotyped HIV-1 particles carry the wildtype SARS-CoV-2 Spike protein and are capable of delivering a NanoLuc luciferase reporter gene to ACE2-expressing HEK293T cells. Cells and pseudovirus were pre-incubated with serially diluted ACE2-Fc or monomeric ACE2, and luciferase activity was measured in cell lysates at 48 hours post infection. ACE2-Fc neutralized the SARS-CoV-2 pseudotyped particles at an $IC_{50} = 0.96 \mu\text{g/mL}$ (4.75 nM), while ACE2 monomer did not substantially neutralize at the concentrations we tested, up to 48 $\mu\text{g/mL}$ (237.5 nM) (**Fig. 3.8A, 3.8B**). This stark difference in

efficacy between monomeric and dimeric ACE2 confirms that ACE2 monomer binds poorly to Spike.

We further tested the ability of ACE2-Fc and ACE2 monomer to neutralize SARS-CoV-2 virus. SARS-CoV-2 live virus was pre-incubated with 100 nM of ACE2-Fc or ACE2 monomer prior to infecting VeroE6 cells, a monkey kidney epithelial cell line broadly used for studying viral infectivity. 16 hours post infection, cells were lysed and intercellular viral and host RNA was isolated, converted to cDNA, and quantified by qPCR. Consistent with the pseudotyped virus neutralization results, ACE2-Fc neutralized SARS-CoV-2 much more potently than ACE2 monomer (**Fig. 3.9A, 3.9B**).

Discussion

Worldwide efforts are currently underway to develop effective and fast-acting clinical interventions to control the spread and mortality of SARS-CoV-2. While it is clear that the interaction between ACE2 and Spike-RBD plays a crucial role in viral entry, our current understanding is insufficient for the design of maximally effective therapeutic options. In this work, we systematically interrogated how an ACE2 dimer interacts with SARS-CoV-2 Spike trimer to understand the fundamental avidity properties of the Spike/ACE2 interaction.

The results from the BLI experiments, split reporter assays, and virus neutralization assays provide important insight into the conformational landscape of the RBDs in SARS-CoV-2 Spike and are summarized in **Fig. 3.10A**. The decreased k_{on} of ACE2-Fc suggests the majority of RBDs in FL-Spike are in an ACE2-inaccessible, “down” conformation, which requires opening up for binding ACE2 (**Fig. 3.10A**). The poor binding of monomeric ACE2 (**Fig. 3.10B**) and strong binding to dimeric ACE2-Fc (**Fig. 3.10C**) indicate intra-Spike avidity and ACE2 rebinding plays an important role in promoting this interaction. However, while we observed a highly productive

avidity effect for the Spike-RBD-Fc/ACE2-Fc interaction, we did not observe the massive two-on-two avidity for the FL-Spike/ACE2-Fc interaction (**Fig. 3.10D**). This weaker interaction could be due to RBD conformations in Spike, or a sub-optimal geometry of the two proteins for high-avidity binding. Nonetheless, we show that two or more of the RBD domains can be in the “up” conformation allowing binding of two ACE2-Fc on one Spike trimer (**Fig. 3.10E**). This is evidenced in our BLI experiment where once bound to FL-Spike, the presence of the other RBDs slows down the dissociation of ACE2-Fc (**Figure 3.4F** in comparison to **3.4D**). This is further shown in the split-NanoLuc reporter experiments where ACE2 domains from two separate ACE2-Fc split-reporters can simultaneously bind a Spike trimer to generate an active luciferase (**Figure 3.6**). Together these results indicate that trimeric SARS-CoV-2 Spike interacts with ACE2 with limited but important intra-Spike avidity.

In addition to intra-Spike avidity, inter-Spike avidity could contribute to the improved affinity of ACE2-Fc especially in the intact virus. It is possible that ACE2-Fc could bridge two Spike molecules on the viral surface and bind with increased avidity, similar to what we observed in the high-density loading BLI experiment (**Fig. 3.5A**). Previous structural studies of SARS-CoV-1 shows the Spike proteins are densely distributed²², with ~100 Spike molecules displayed on a viral particle with a diameter of ~100 nm^{23,24}. Using values from these studies, we estimate an inter-Spike distance of ~180 Å. The structure of the native ACE2 dimer (PDB: 6M17) shows that the two ACE2 arms (~80 Å apart) would not be able to span this predicted inter-Spike distance. An ACE2 monomer fused Fc with flexible linker (ACE2-Fc) may be sufficient to only bridge two Spikes that are <150 Å apart. Structural studies of IgGs have shown the antigen binding sites in the Fab arms on a flexible-hinge region of an Fc to be roughly ~117-134 Å apart²⁵. This would still fall short of the inter-Spike distance. However, these estimates are based on a rigid membrane

where Spike proteins are uniformly distributed and also do not account for heterogeneity in Spike presentation across viral particles. This may not be the case as coronaviruses are an enveloped virus with a fluid lipid membrane that allows for Spike protein mobility and clustering, and potentially enable inter-Spike binding by ACE2 dimers or IgGs. Indeed, cryo-EM images of SARS-CoV-1 show irregularity and lack of symmetry of Spike distribution on the coronavirus envelope²³. As there are no direct reports of inter-Spike avidity for dimeric ACE2, further experiments such as engineering bivalent binders that can span various distances are needed to thoroughly examine the Spike/ACE2 interaction on the viral surface²⁶⁻²⁸.

While all cryo-EM structural studies to date on SARS-CoV-2 Spike have identified only the “closed” or “one-up” RBD conformation^{10,15,29,30}, our split reporter assay identified a population of Spike in “two-up” or “three-up” conformation. The relative population of these RBD conformers and whether they exist in an ACE2-unbound state or emerge only upon ACE2 binding remains unknown (**Fig. 3.10F**). Previous studies have suggested that ACE2 binding could lead to conformational change in SARS-CoV-1 or 2 Spike^{9,31}. Molecular dynamic simulation of SARS-CoV-2 Spike bound to ACE2 found that there is significant flexibility in the RBD conformation¹⁹, and an EM study of SARS-CoV-1 Spike reported that the distribution of the RBD conformers was very different in the ACE2-bound structure compared to the unbound structure¹¹. In accordance with these studies, our results support the model that ACE2 binding induces a conformational change in Spike and enables two or more RBDs to be in the “up” conformation (**Fig. 3.10F**). This complex and dynamic nature of the Spike/ACE2 interaction is likely to play a key role in the biology of SARS-CoV-2 infection. We anticipate that further characterization of the interaction between SARS-CoV-2 Spike and native ACE2 will elucidate the exact nature of this binding.

Furthermore, our findings have important ramifications in developing a successful protein therapeutic for COVID-19. Recombinant ACE2 proteins or antibodies which can block the host ACE2-viral Spike interaction are promising as anti-SARS-CoV-2 therapeutics. The much higher potency of ACE2-Fc in comparison to ACE2 monomer we observed in the live virus neutralization assays suggest that monovalent therapeutics such as ACE2 monomers or Fab domains will likely be much less effective than multimeric formats such as dimeric ACE2¹⁴, ACE2-Fc^{6,12} or IgG (**Fig. 3.10G**). Finally, our results highlight the importance of understanding the ACE2-Spike interaction in the context of the FL-Spike trimer. We also hope that the Spike and ACE2 constructs generated here are useful tools to the greater research community to enable better understanding of SARS-CoV-2 biology.

Methods

Plasmids construction

Plasmids were constructed by standard molecular biology methods. The FL-Spike plasmid was a generous gift from the Pak lab (Chan Zuckerberg Initiative Biohub) and Krammer lab (Icahn School of Medicine at Mount Sinai). The DNA fragments of Spike-RBD, ACE2 and LgBiT were synthesized by IDT Technologies. The Spike-RBD-TEV-Fc-AviTag, ACE2-TEV-Fc-AviTag, Spike-RBD-8xHis-AviTag, ACE2-8xHis-AviTag plasmids were generated by subcloning the Spike-RBD or ACE2 DNA fragment into a pFUSE-hIgG1-Fc-AviTag vector (adapted from the pFUSE-hIgG1-Fc vector from InvivoGen). The ACE2-Fc-LgBiT fusion plasmids were generated by subcloning the gene fragments of LgBiT to the N- or C-terminus of the ACE2-TEV-Fc-AviTag vector with a 10-amino acid (N-terminal fusion) or 5-amino acid (C-terminal fusion) linker. The SmBiT tag in the ACE2-Fc-SmBiT fusion plasmids was generated by overlap-extension PCR, which also has a 10-amino acid (N-terminal fusion) or 5-amino acid (C-terminal fusion) linker to

the ACE2 or Fc domains. The C-terminal AviTag was removed from all the ACE2-Fc reporter plasmids. Complete plasmid sequences are available upon request.

Expression and purification of ACE2 and Spike constructs

The ACE2 and Spike proteins were expressed and purified from Expi293 BirA cells according to established protocol from the manufacturer (Thermo Fisher Scientific). Briefly, 30 µg of pFUSE (InvivoGen) vector encoding the protein of interest was transiently transfected into 75 million Expi293 BirA cells using the Expifectamine kit (Thermo Fisher Scientific). Enhancer was added 20 h after transfection. Cells were incubated for a total of 3 d at 37 °C in an 8% CO₂ environment before the supernatants were harvested by centrifugation. Fc-fusion proteins were purified by Protein A affinity chromatography and His-tagged proteins were purified by Ni-NTA affinity chromatography. Purity and integrity were assessed by SDS/PAGE. Purified protein was buffer exchanged into PBS and stored at -80 °C in aliquots.

Generation of ACE2 monomer

ACE2 monomer was obtained by TEV treatment of ACE2-Fc and subsequent purification. 50 µl Ni-NTA agarose (Qiagen) and 50 µl Neutravidin resin (Thermo Fisher Scientific) were washed with PBS-25 mM imidazole twice and combined in 100 µl PBS-25 mM imidazole. Next, 20 µg His-Tagged recombinant TEV protease and 1 mg purified ACE2-Fc protein were mixed, and the reaction tube was rotated at 4 °C for 30 minutes. The cleavage reaction was then incubated with the washed beads, rotating, at 4 °C for 30 minutes. While the incubation occurred, an additional 25 µl of magnetic Protein A beads and 25 µl or Ni-NTA beads were prepared as described before. Supernatant from the first bead clearance was transferred to the newly prepared beads and allowed to incubate for an additional 30 minutes at 4 °C. To remove beads from the

protein supernatant, reaction mixture was spin filtered at 1000 g for 2 min and washed with an additional 250 μ L of PBS-25 mM imidazole. The His-tagged TEV, biotinylated Fc, and the uncut ACE2-Fc remained on the beads while the monomeric ACE2 was isolated in the flow-through. The purity of monomeric ACE2 was confirmed by SDS-PAGE electrophoresis. Purified protein was buffer exchanged to PBS and store at -80°C in aliquots.

Differential scanning fluorimetry

To assess the stability of proteins, we measured the melting temperature (T_m) by doing differential scanning fluorimetry (DSF) as the method described previously³². Briefly, purified protein was diluted to 0.5 μM or 0.25 μM in DSF buffer containing Sypro Orange 4x (Invitrogen) and PBS. 10 μL of reaction mixture was transferred to one well of a 384-well PCR plate. Duplicate was prepared as needed. In a Roche LC480 LightCycler, the reaction was heated from 30°C to 95°C with a ramp rate of 0.3°C per 30 sec. The intensities of the fluorescent signal at an ~ 490 nm and ~ 575 nm (excitation and emission wavelengths) were continuously collected. The curve peak corresponds to the melting temperature of the protein. Data was processed and T_m was calculated using the Roche LC480 LightCycler software.

In vitro binding experiments

Biolayer interferometry data were measured using an Octet RED384 (ForteBio). Biotinylated Spike or Spike RBD protein were immobilized on the streptavidin (SA) biosensor (ForteBio). After blocking with biotin, purified ACE2 proteins in solution was used as the analyte. PBS with 0.05% Tween-20 and 0.2% BSA was used for all diluents and buffers. A 1:1 monovalent binding model was used to fit the kinetic parameters (k_{on} and k_{off}).

Magnetic bead and solution based NanoBiT assays

For the Spike-Fc magnetic bead assay, magnetic beads were prepared by taking 100 μ L of Streptavidin Magnesphere Paramagnetic Particles (Promega) and incubated with 5 μ M of Spike-Fc-AviTag for 30 minutes, rotating at room temperature. Following, the beads were blocked with 10 μ M biotin for 10 minutes. The beads were washed three times with PBS + 0.05% Tween + 0.2% BSA. 10 μ L of 10-fold dilutions of the beads were incubated with 10 μ L of premixed 2 nM ACE2-Fc-SmBiT and ACE2-Fc-LgBiT fusions. The sample was incubated shaking at room temperature for 20 minutes. NanoGlo Luciferase substrate (Promega) diluted in NanoGlo Luciferase buffer was added to each well (15 μ L) and luminescence was measured on a Tecan M1000 plate reader after 10 minutes. For the FL-Spike detection, 10 μ L of FL-Spike dilutions were combined with 10 μ L of premixed 2 nM SmBiT-ACE2-Fc and LgBiT-ACE2-Fc. Samples were incubated with substrate and luminescence was detected as described above.

Pseudotyped virus neutralization assay

HEK293T cells overexpressing full-length ACE2 carrying two inactivating mutations in the catalytic domain (H374N & H378N) were generated using standard lentivirus transduction. The cells were maintained in DMEM (Gibco) with 10% heat-inactivated fetal bovine serum, Gentamycin and 5 μ g/ml Blasticidin.

Pseudotyped HIV-1 particles expressing the SARS-CoV-2 Spike glycoprotein and NanoLuc luciferase as a reporter were generated by transfection of HEK293T cells with pNL4-3DEnv-NanoLuc and pSARS-CoV2-S_{trunc}. pNL4-3DEnv-NanoLuc was derived from pNL4-3³³ by inserting a 940 bp deletion after the *vpu* stop-codon, resulting in loss of Env-expression. The NanoLuc Luciferase reporter gene (*Nluc*, Promega) was inserted in place of bp 1-100 in the *nef*-frame. pSARS-CoV2-S_{trunc} was generated by insertion of a human-codon optimized gene encoding for 19 AAs C-terminally truncated SARS-CoV-2 Spike (Genart) into pCR3.1.

Supernatants containing virus were harvested and filtered 48 hr post transfection and used for infection of ACE2-overexpressing 293T cells. Pseudotyped virus was pre-incubated with serially diluted ACE2-Fc or ACE2 monomer at 37 °C for 1 hr before addition to cells., Cells were washed twice with PBS 48 hr post infection and lysed with Luciferase Cell Culture Lysis 5x reagent (Promega). Relative luminescence units were normalized to those derived from cells infected with SARS-CoV-2 pseudotyped virus in the absence of Ace2-Fc/Ace2 monomer. The half maximal inhibitory concentration (IC₅₀) was determined using 4-parameter nonlinear regression (GraphPad Prism). Each concentration of Ace2-Fc/Ace2 monomer was tested in duplicate and reported as an average and standard deviation for each experiment. Two biological replicates of ACE2-Fc treatment were obtained (exp #1, exp #2).

SARS-CoV-2 virus neutralization assay

All handling and experiments using SARS-CoV-2 was conducted under Biosafety Level 3 containment with approved BUA and protocols. SARS-CoV-2 clinical isolate 2019-nCoV/USA-WA1/2020 was obtained through BEI Resources (Harcourt et al., 2020). Prior to experiments, virus was passed in Vero E6 cells to create working stocks and titers were measured by plaque formation assay. For experiments, Vero E6 cells were cultured in Minimal Essential Media (MEM), 10% Fetal Bovine Serum (FBS), 1% Pen-Strep and seeded on 6-well culture plates at 3.8E5 cells/well the day prior. Infection with SARS-CoV-2 was performed using MOI of 0.1. Virus was incubated in infection media (EMEM 0% FBS) containing 100 nM monomeric ACE2, 100 nM dimeric ACE2-Fc, or no blocker for 1 hour at 37 °C. Culture media was removed from Vero E6 cells and 300 µL of the blocker/virus inoculum was added for 1 hour at 37 °C. Subsequently, 1 mL of cell culture media was added and cells were incubated at 37 °C for an additional 16 hours before RNA harvest.

Viral entry into cells and cellular transcription of viral genes was measured by qPCR. Cellular RNA was isolated and converted to cDNA using RNAeasy RNA extraction kit (Qiagen) and Quantitect Reverse-transcriptase kit (Qiagen) according to manufacturer instructions. qPCR reactions were prepared using SYBR Select Master Mix (Thermo) and the following conditions: for N gene, E gene, and hGUSB gene primer (cross reactive with *Cercopithecus aethiops*) concentration was 400 nM and annealing temperature was 58 °C, and for hACTB gene primer (cross reactive with *Cercopithecus aethiops*) concentration was 500 nM and annealing temperature was 60 °C. Primer sequences (IDT) were the following – viral genes: N_F = CACATTGGCACCCGCAATC; N_R = GAGGAACGAGAAGAGGCTTG; E_F = ACAGGTACGTTAATAGTTAATAGCGT; E_R = ATATTGCAGCAGTACGCACACA; and host genes: hGUSB_F = CTCATCTGGAATTTTGCCGATT; hGUSB_R = CCGAGTGAAGACCCCCTTTTTA; hACTB primers were IDT PrimeTime assay reagent. Relative copy number of viral transcript level compared to cellular transcript was determined using the $\Delta\Delta CT$ method.

Figures and Tables

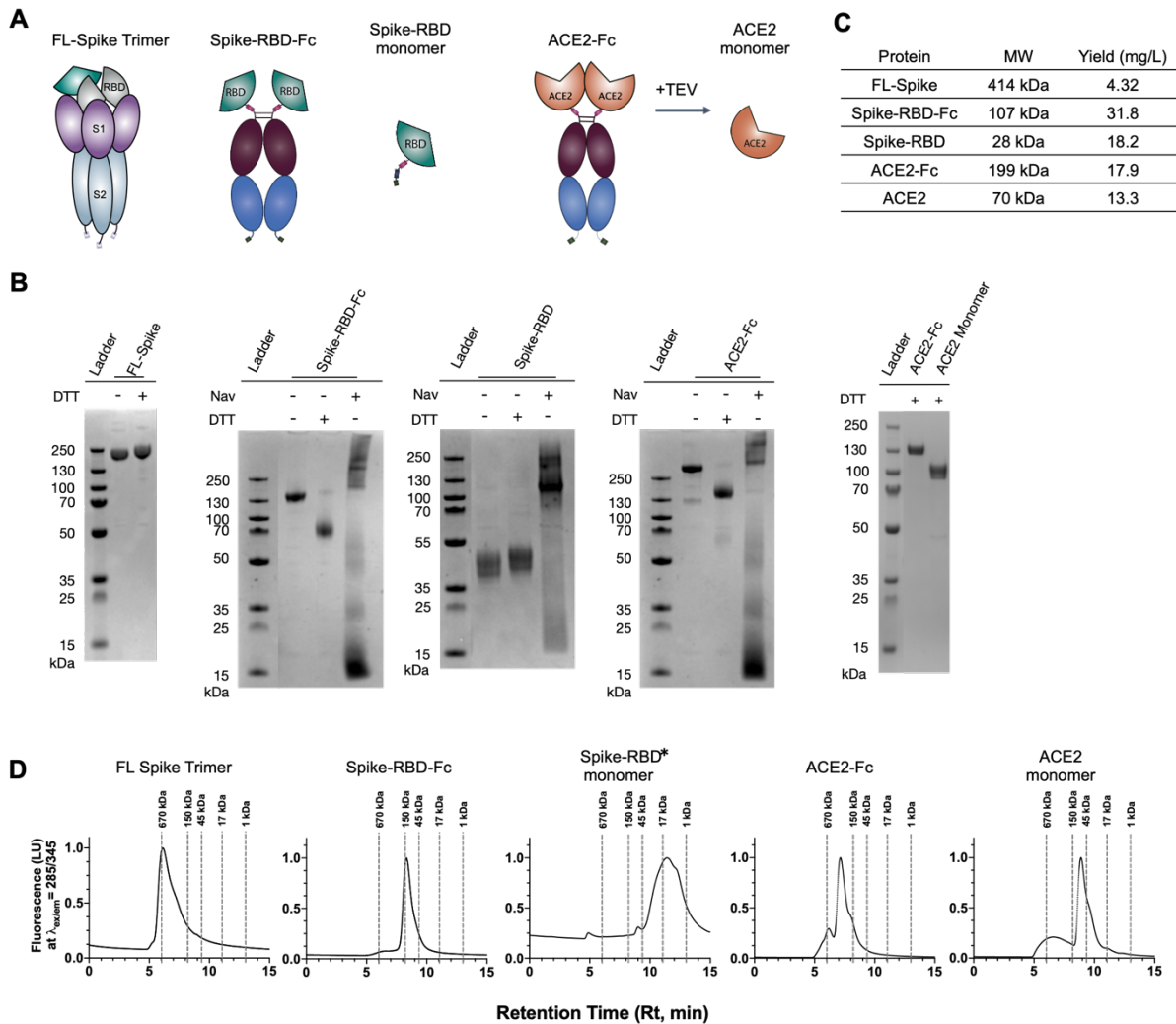


Figure 2.1: Purification and characterization of Spike and ACE2 variants.

(A) Cartoon representation of antigens: FL-Spike, Spike-RBD-Fc, Spike-RBD, ACE2-Fc and ACE2. (B) SDS-PAGE gel (4-12%) analysis of purified proteins stained with Coomassie blue. Protein was incubated with reducing agent or neutravidin to determine biotinylation. (C) Purification method and expression yield per construct. (D) SEC traces of purified protein. Dotted lines mark retention time of a molecular weight standard. *Retention time of Spike-RBD monomer does not precisely match elution profile of standard, but further SDS-Page gel analysis of SEC fractions shows pure protein at the correct molecular weight (Fig. 3.2).

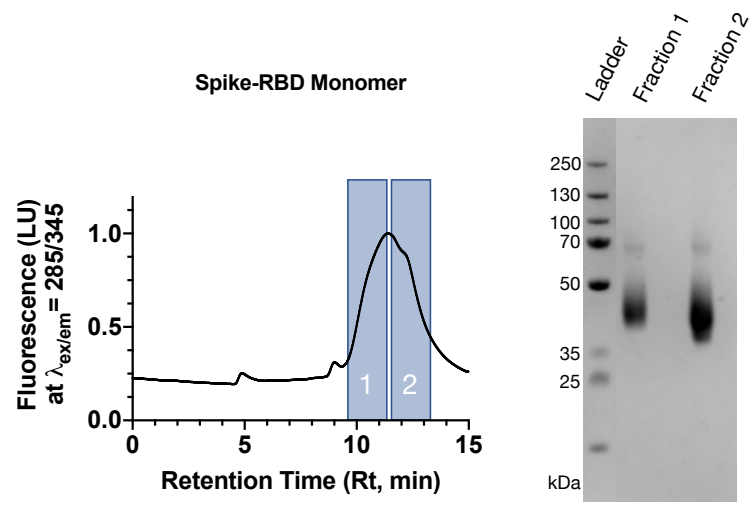


Figure 2.2: SDS-PAGE analysis of SEC fractions shows pure Spike-RBD protein at the correct molecular weight.

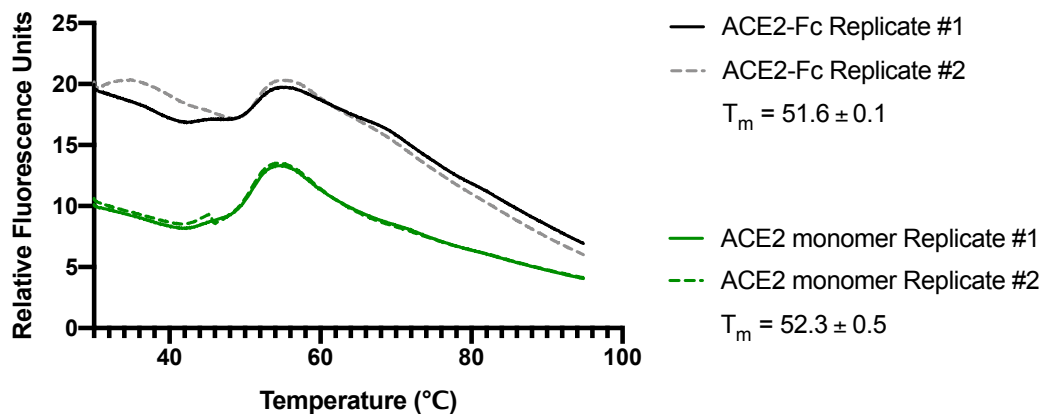


Figure 2.3: Melting temperature of reagents measured by Differential Scanning Fluorimetry (DSF). T_m is reported as an average of two replicates.

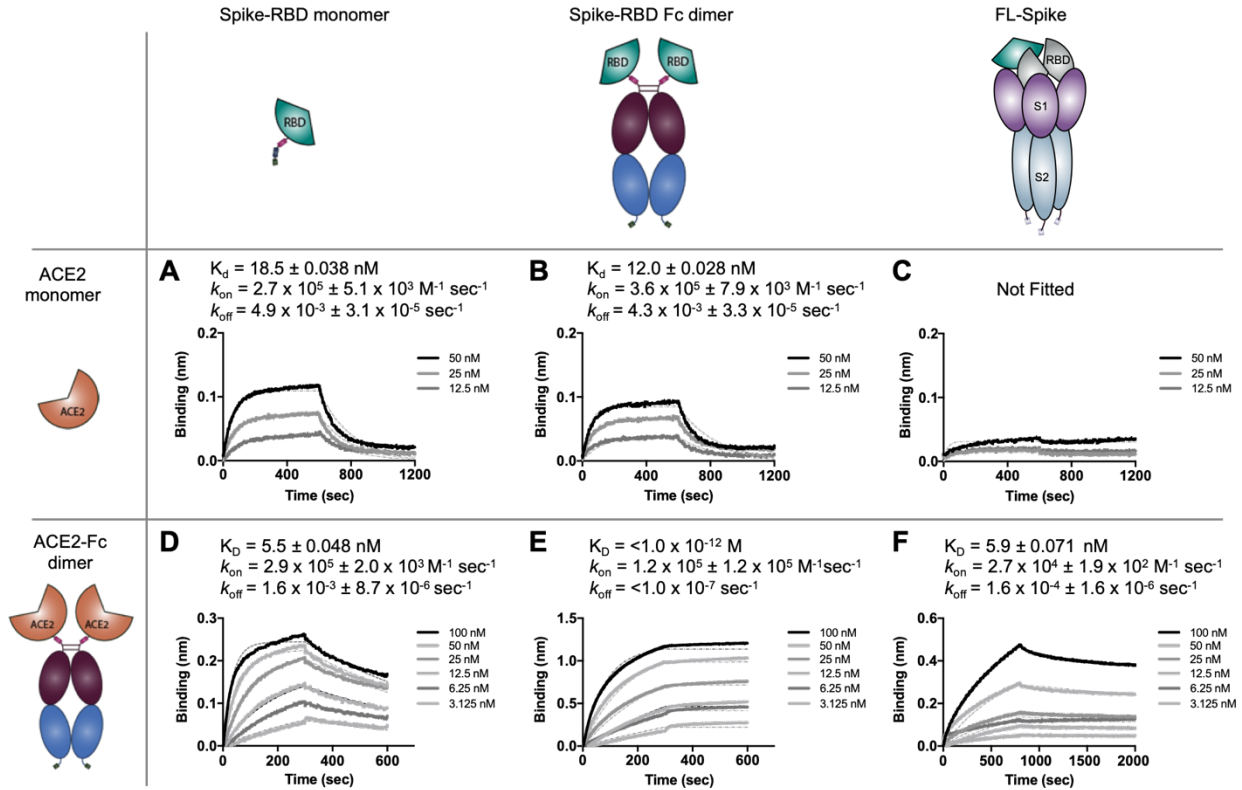


Figure 2.4: BLI characterization of binding affinity and kinetics of Spike and ACE variants
 Bio-layer Interferometry (BLI) data of ACE2 monomer or ACE2-Fc dimer binding to immobilized Spike RBD-monomer, RBD-Fc dimer, or trimeric FL-Spike ectodomain. Spike variants were loaded onto sensors until binding was between 0.4-0.6 nm. ACE2 monomer or ACE2-Fc were used as analytes in solution, with concentrations ranging from 100 nM to 3.125 nM. Although ACE2 monomer bound Spike-RBD-monomer and Spike-RBD-Fc at K_D of 18.5 and 12.0 nM, respectively, it did not bind FL-Spike strongly (A-C). In contrast, the ACE2-Fc dimer bound to these forms with affinities of 5.5 nM, <1 pM, and 5.9 nM (D-F). Although ACE2-Fc can bind Spike-RBD-Fc with strong intramolecular avidity, this 2-on-2 interaction is not present in the context of FL-Spike, indicating that only one arm of ACE2-Fc can engage an RBD in FL-Spike. The decreased k_{on} and k_{off} of ACE2-Fc in the context of FL-Spike (F) compared to Spike-RBD-monomer (D) indicates that conformational changes of the RBD in the context of the Spike trimer are important in the interaction of ACE2-Fc to FL-Spike. Representative data sets are shown. All assays were performed in duplicate with separate batches of protein.

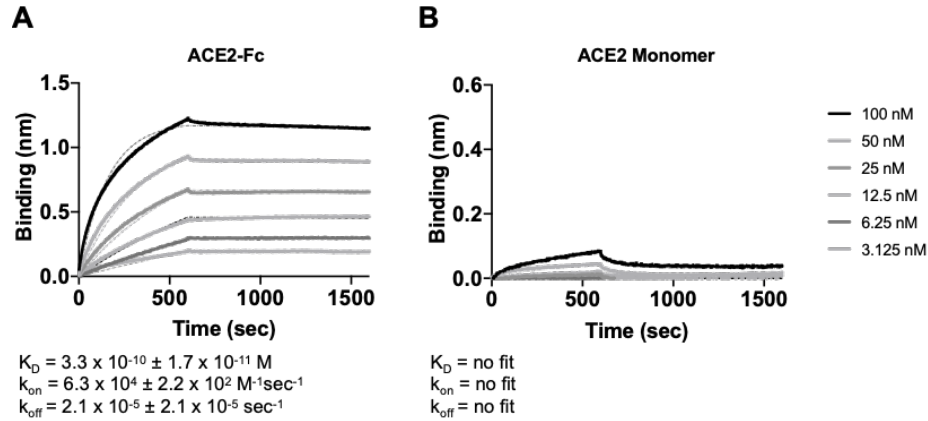


Figure 2.5: Saturating FL-spike on biosensor reveals intermolecular avidity of ACE2-Fc.

FL-Spike was loaded onto sensors until saturation. **(A)** ACE2-Fc was used as analytes in solution, with concentrations beginning at 100 nM to 3.125 nM at 2-fold dilutions. The affinities for the ACE2 dimeric to saturating trimeric FL-Spike was 0.33 nM, or 18-fold higher in affinity compared to non-saturating conditions (5.9 nM, Figure 2F). **(B)** ACE2 monomer did not bind FL-Spike strongly even when FL-Spike was loaded until saturation.

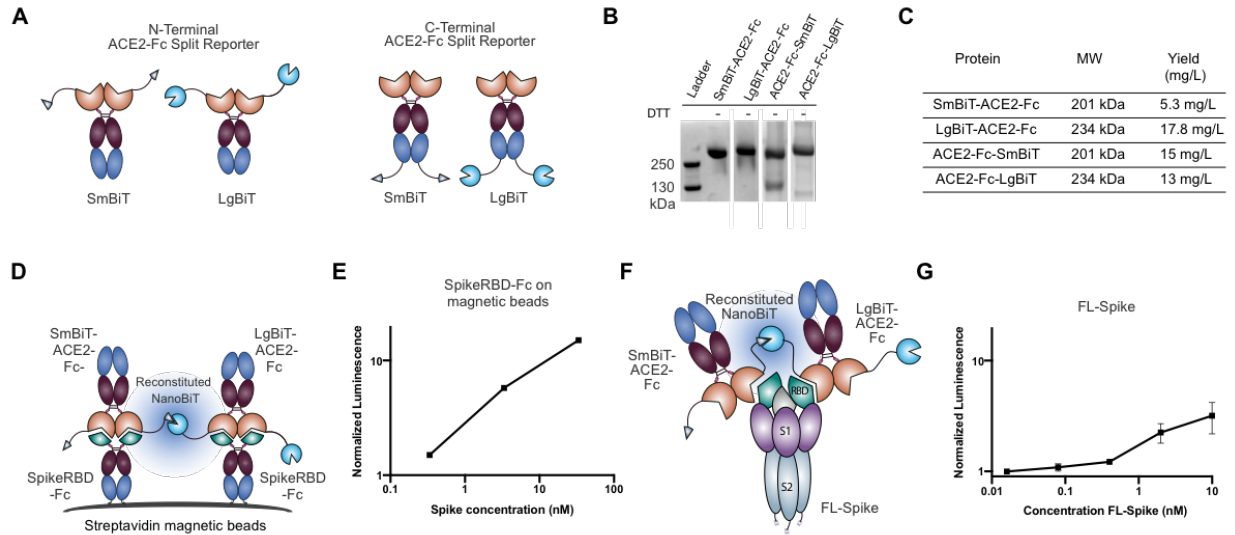


Figure 2.6: ACE2-Fc split luciferase experiments demonstrates more than one RBD in FL-Spike are available to bind ACE2.

(A) Cartoon depiction of the N-terminal and C-terminal ACE2 NanoBiT split luciferase constructs. (B) SDS-PAGE gel showing expression and purification of ACE2 fusions. All show a major band at the expected molecular weight of the dimer. (C) All of the ACE2 split-reporter fusions had good yields (>5 mg/L). (D) Cartoon depiction of the N-terminal ACE2 NanoBiT sensor system to detect Spike-RBD-Fc bound to streptavidin magnetic beads. (E) The N-terminal ACE2 NanoBiT system is able to detect Spike-RBD-Fc bound magnetic beads with better sensitivity than the C-terminal fusions (Fig. 3.7). (F) Cartoon depiction of the N-terminal ACE2 NanoBiT sensor system to detect FL-Spike in solution. (G) The N-terminal ACE2 NanoBiT system is able to detect FL-Spike in solution, indicating that more than one RBD in a FL-Spike trimer can be in the “up” conformation for two ACE2-Fc molecules to bind and reconstitute the split enzyme. For all assays, the luminescence signal is normalized to a no-bead control and the average and standard deviation is plotted (N=3).

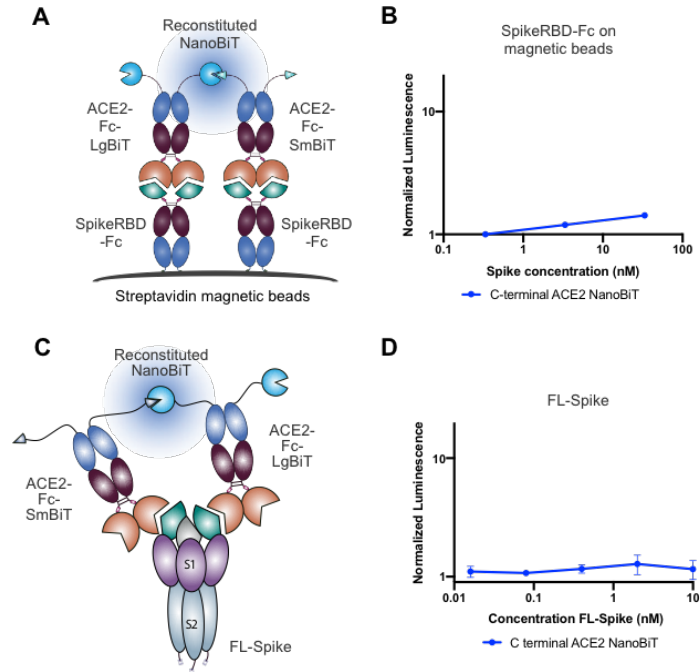


Figure 2.7: C-terminally fused ACE2-Fc Split Luciferase Reporter is not as sensitive as the N-terminal fusions.

(A) Cartoon depiction of the C-terminal ACE2 NanoBiT sensor system to detect Spike-RBD-Fc bound to streptavidin magnetic beads. (B) The C-terminal ACE2 NanoBiT system is able to detect Spike-RBD-Fc bound magnetic beads, but with lower sensitivity than the N-terminal fusions. (C) Cartoon depiction of the C-terminal ACE2 NanoBiT sensor system to detect FL-Spike in solution. (D) The C-terminal ACE2 NanoBiT system is not able to detect Spike-RBD-Fc bound magnetic beads due to lower sensitivity. For all assays, the luminescence signal is normalized to a no-bead control and the average and standard deviation is plotted (N=3).

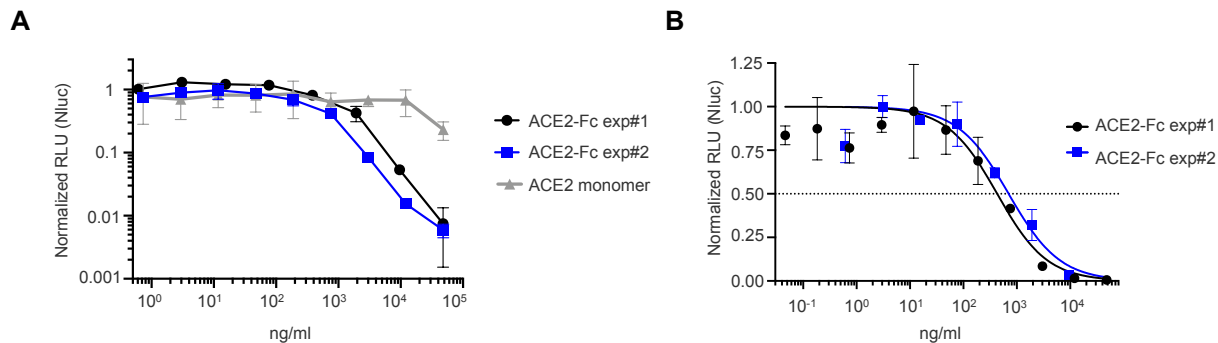


Figure 2.8: ACE2-Fc neutralizes pseudotyped virus more potently than ACE2 monomer.

(A) Neutralization of luciferase-encoding pseudotyped lentivirus expressing SARS-CoV-2 Spike glycoprotein. Pseudotyped virus pre-incubated with ACE2-Fc or ACE2 monomer at indicated concentrations was used to infect HEK293T cells overexpressing ACE2. Luciferase activity in cell lysates were determined at 48 hours post infection and normalized to no treatment control. **(B)** The half-maximal inhibitory concentration (IC_{50}) was obtained for the two ACE-Fc treatment by fitting the normalized data to a 4-parameter nonlinear regression. The average IC_{50} between two biological replicates of ACE2-Fc treatment was 0.98 μ g/ml (4.8 nM).

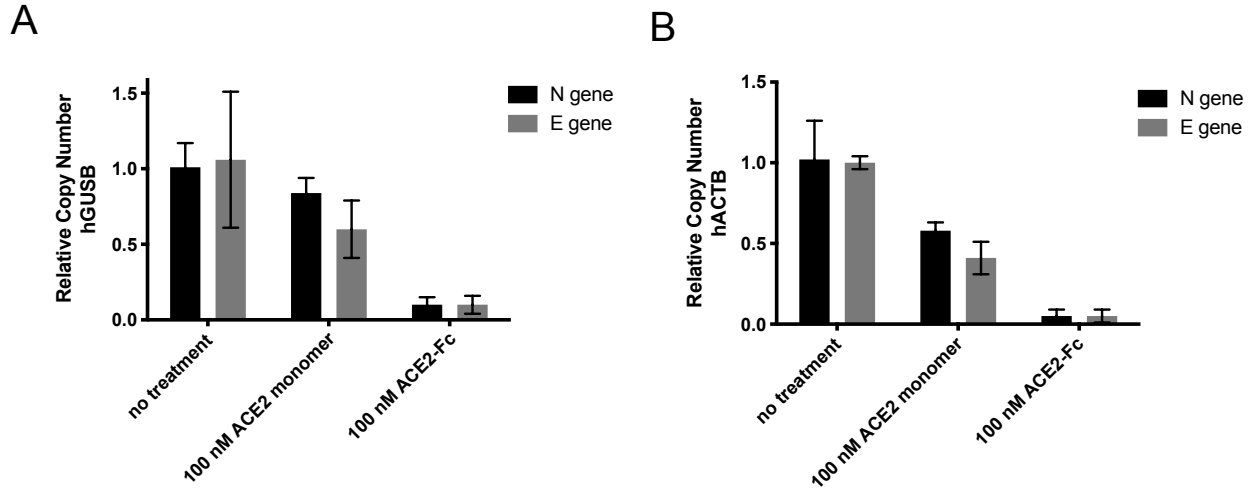


Figure 2.9: ACE2-Fc neutralizes SARS-CoV-2 live virus more potently than ACE2 monomer.

SARS-CoV-2 live virus pre-incubated with 100 nM of ACE2-Fc or ACE2 monomer was used to infect VeroE6 cells. 16 hours post infection, cells were lysed and intercellular RNA was isolated and converted to cDNA. Viral entry and cellular transcription of viral genes was measured by qPCR. Relative copy number of viral N or E gene transcript for each treatment arm was determined using either (A) host gene GUSB (hGUSB) or (B) host gene ACTB (hACTB) as the reference control. Primers are cross reactive between *Homo sapiens* and *Cercopithecus aethiops* (VeroE6 cells). Two biological replicates were conducted, with two technical replicates per assay. For both panels, the average relative copy number from two biological replicates are shown. Error bars show the standard deviation (S.D.).

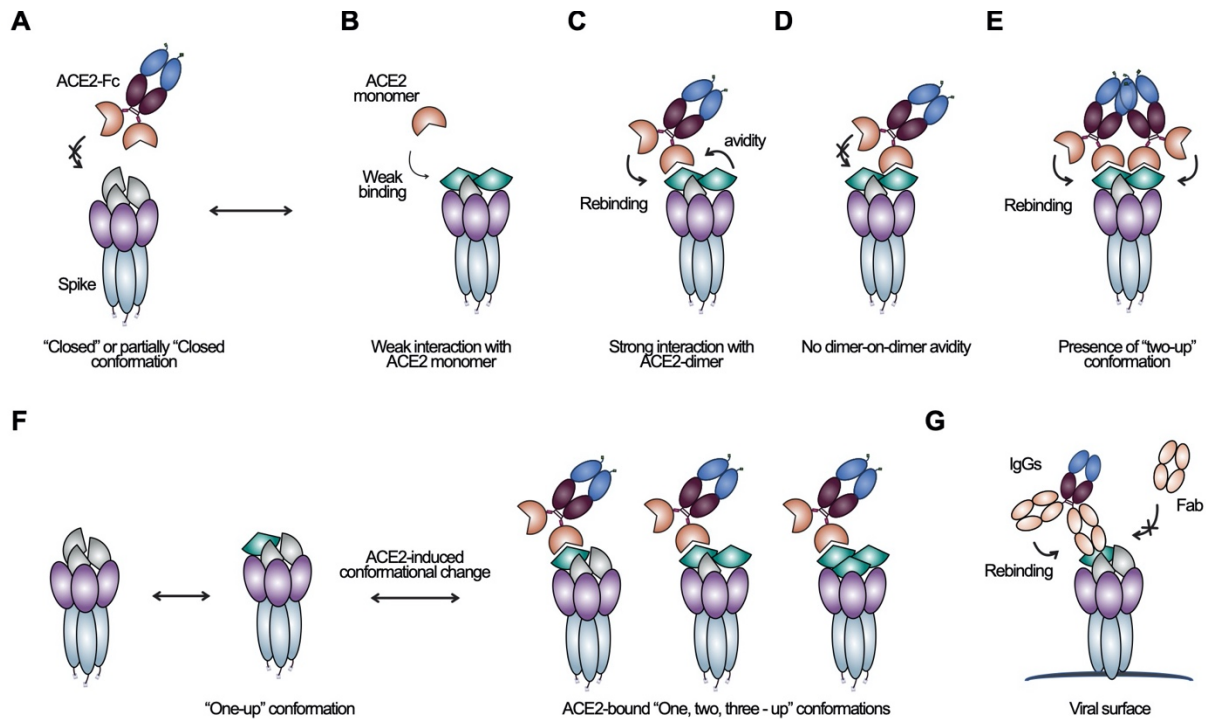


Figure 2.10: Model of the ACE2/Spike interaction and therapeutic strategies.

(A) The majority of FL-Spike proteins are in the “closed” or partially “closed” conformation and “opening up” of RBD on Spike is necessary for ACE2 binding. (B) Monomeric ACE2 binds poorly to trimeric Spike, (C) ACE2-Fc binds much stronger to Spike than monomer, indicating that rebinding and intra-Spike avidity contribute significantly to the high-affinity nature of this interaction. (D) Dimeric ACE2-Fc does not interact with FL-Spike with a full two-on-two intramolecular avidity. (E) More than one RBD can be in the “up” conformation, enabling the engagement of separate ACE2 molecules. For (B-E), only the “two-up” conformation is shown but other RBD conformations are also possible. (F) A proposed model where ACE2 binding may induce a conformational change in Spike, resulting in “two-up” or “three-up” RBD conformations. (G) Dimeric molecules such as ACE2-Fc or IgG will be more potent than a monomeric inhibitor for neutralizing SARS-CoV-2 virus. For all panels, RBD in the “open” conformation is colored in green and RBD in the “closed” formation colored in grey.

References

1. Gralinski, L. E., & Menachery, V. D. Return of the coronavirus: 2019-nCoV. *Viruses* **12**, 135 (2020).
2. Tai, W., He, L., Zhang, X., Pu, J., Voronin, D., Jiang, S., ... Du, L. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol Immun* (2020).
3. Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., ... Zhang, Y. Z. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
4. Chen, Y., Liu, Q., & Guo, D. (2020). Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J Med Virol* **92**, 418-423 (2020).
5. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., ... Wang, X. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).
6. Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448 (2020).
7. Wrapp, D., Vlieger, D. De, Corbett, K. S., Torres, G. M., Breedam, W. Van, Roose, K., ... McLellan, J. S. Structural Basis for Potent Neutralization of Betacoronaviruses by Single-domain Camelid Antibodies. *BioRxiv* (2020).
8. Gui, M., Song, W., Zhou, H., Xu, J., Chen, S., Xiang, Y., & Wang, X. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res* **27**, 119–129 (2017).

9. Song, W., Gui, M., Wang, X., & Xiang, Y. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Path* **14**, e1007236 (2018).
10. Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **180**, 281–292 (2020).
11. Kirchdoerfer, R. N., Wang, N., Pallesen, J., Wrapp, D., Turner, H. L., Cottrell, C. A., ... Ward, A. B. Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Sci Rep* **8** 15701 (2018).
12. Lei, C., Fu, W., Qian, K., Li, T., Zhang, S., Ding, M., & Hu, S. Potent neutralization of 2019 novel coronavirus by recombinant ACE2-Ig. *BioRxiv* (2020).
13. Li, Y., Wang, H., Tang, X., Ma, D., Du, C., Wang, Y., ... Zhong, G. Potential host range of multiple SARS-like coronaviruses and an improved ACE2-Fc variant that is potent against both SARS-CoV-2 and SARS-CoV-1. *BioRxiv* (2020).
14. Monteil, V., Kwon, H., Prado, P., Hagelkruys, A., Wimmer, R., Stahl, M., ... Penninger, J. Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-grade soluble human ACE2. *Cell* (2020).
15. Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., ... McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
16. Czajkowsky, D. M., Hu, J., Shao, Z., & Pleass, R. J. Fc-fusion proteins: New developments and future perspectives. *EMBO Mole Med* **4**, 1015–1028 (2012).
17. Howarth, M., Liu, W., Puthenveetil, S., Zheng, Y., Marshall, L. F., Schmidt, M. M., ...

- Ting, A. Y. Monovalent, reduced-size quantum dots for imaging receptors on living cells. *Nat Meth* **5**, 397–399 (2008).
18. Martinko, A. J., Truillet, C., Julien, O., Diaz, J. E., Horlbeck, M. A., Whiteley, G., ... Wells, J. A. (2018). Targeting RAS-driven human cancer cells with antibodies to upregulated and essential cell-surface proteins. *eLife* **7**, e31098 (2018).
 19. Peng, C., Zhu, Z., Shi, Y., Wang, X., Mu, K., Yang, Y., ... Zhu, W. Exploring the Binding Mechanism and Accessible Angle of SARS-CoV-2 Spike and ACE2 by Molecular Dynamics Simulation and Free Energy Calculation. *Chemrxiv.Org* (2020).
 20. Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., & Li, F. (2020). Cell entry mechanisms of SARS-CoV-2. *PNAS* **117** 11727–11734 (2020).
 21. Dixon, A. S., Schwinn, M. K., Hall, M. P., Zimmerman, K., Otto, P., Lubben, T. H., ... Wood, K. V. NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chem Biol.* **11**, 400–408 (2016).
 22. Lin, Y., Yan, X., Cao, W., Wang, C., Feng, J., Duan, J., & Xie, S. (2004). Probing the structure of the SARS coronavirus using scanning electron microscopy. *Antiviral Ther* **9**, 287–289 (2004).
 23. Beniac, D. R., Andonov, A., Grudeski, E., & Booth, T. F. Architecture of the SARS coronavirus prefusion spike. *Nat Struct Mol Biol* **13**, 751–752 (2006).
 24. Neuman, B. W., Adair, B. D., Yoshioka, C., Quispe, J. D., Orca, G., Kuhn, P., ... Buchmeier, M. J. (2006). Supramolecular Architecture of Severe Acute Respiratory Syndrome Coronavirus Revealed by Electron Cryomicroscopy. *J Virol* **80**, 7918–7928 (2006).
 25. Sosinick, T.R., Mayne, L, Hiller, R., & Englander, S.W. The barriers in protein folding.

- Nat Struct Biol* **1**, 149-156 (1992).
26. Einav, T., Yazdi, S., Coey, A., Bjorkman, P. J., & Phillips, R. Harnessing Avidity: Quantifying the Entropic and Energetic Effects of Linker Length and Rigidity for Multivalent Binding of Antibodies to HIV-1. *Cell Systems* **9**, 466–474 (2019).
 27. Galimidi, R. P., Klein, J. S., Politzer, M. S., Bai, S., Seaman, M. S., Nussenzweig, M. C., ... Bjorkman, P. J. Intra-spike crosslinking overcomes antibody evasion by HIV-1. *Cell* **160**, 433–446 (2015).
 28. Klein, J. S., & Bjorkman, P. J. Few and far between: How HIV may be evading antibody avidity. *PLoS Pathogens* **6**, e1000908 (2010).
 29. Cao, Y., Su, B., Guo, X., Sun, W., Deng, Y., Bao, L., ... Xie, X. S. Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell* (2020).
 30. Pinto, D., Park, Y., Beltramello, M., Walls, A. C., Tortorici, M. A., Bianchi, S., ... Corti, D. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* (2020).
 31. Yuan, M., Wu, N. C., Zhu, X., Lee, C.-C. D., So, R. T. Y., Lv, H., ... Wilson, I. A. A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. *Science* (2020).
 32. Hornsby, M., Paduch, M., Miersch, S., Sääf, A., Matsuguchi, T., Lee, B., ... Wells, J. A High Through-put Platform for Recombinant Antibodies to Folded Proteins. *Mol Cell Proteom* **14**, 2833–2847 (2015).
 33. Adachi, A., Gendelman, H. E., Koenig, S., Folks, T., Willey, R., Rabson, A., & Martin, M. A. (1986). Production of acquired immunodeficiency syndrome-associated retrovirus

- in human and nonhuman cells transfected with an infectious molecular clone. *J Virol* **59**, 284–291 (1986).
34. Harcourt, J., Tamin, A., Lu, X., Kamili, S., Sakthivel, S. K., Murray, J., ... Thornburg, N. J. Severe Acute Respiratory Syndrome Coronavirus 2 from Patient with 2019 Novel Coronavirus Disease, United States. *Emerg Infect Dis* **26** (2020).

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

Irene Lui

6701D5D83E64479...

Author Signature

3/19/2024

Date