

#### pubs.acs.org/acschemicalbiology

# Engineered ChymotrypsiN for Mass Spectrometry-Based Detection of Protein Glycosylation

Balakrishnan Ramesh,<sup>†,§</sup> Shaza Abnouf,<sup>†,§</sup> Sujina Mali,<sup>‡</sup> Wilna J. Moree,<sup>‡</sup> Ujwal Patil,<sup>‡</sup> Steven J. Bark,<sup>‡</sup> and Navin Varadarajan\*<sup>,†</sup>

<sup>†</sup>Department of Chemical and Biomolecular Engineering, University of Houston, Houston, Texas 77204-4004, United States <sup>‡</sup>Department of Biology and Biochemistry, University of Houston, Houston, Texas 77004, United States

S Supporting Information

**ABSTRACT:** We have engineered the substrate specificity of chymotrypsin to cleave after Asn by high-throughput screening of large libraries created by comprehensive remodeling of the substrate binding pocket. The engineered variant (chymotrypsiN, ChyB-Asn) demonstrated an altered substrate specificity with an expanded preference for Asncontaining substrates. We confirmed that protein engineering did not compromise the stability of the enzyme by biophysical characterization. Comparison of wild-type ChyB and ChyB-Asn in profiling lysates of HEK293 cells demonstrated both qualitative and quantitative differences in the nature of the peptides and proteins identified by liquid chromatography and tandem mass spectrometry. ChyB-Asn enabled the identification of partially glycosylated Asn sites within a model glycoprotein and in the extracellular proteome of Jurkat T cells. ChymotrypsiN is a valuable addition to the toolkit of



proteases to aid the mapping of N-linked glycosylation sites within proteins and proteomes.

espite advances in high-throughput sequencing of DNA and RNA, detailed maps of the human proteome with direct measurements of proteins and their modifications are just starting to emerge.<sup>1,2</sup> Comprehensive characterization of biological systems enabled by quantitative differences in protein expression and abundance, cell-type specific and temporal expression patterns, protein-protein interactions, and post-translational modifications (PTMs) is best accomplished using mass spectrometry (MS) proteomics.<sup>1,3</sup> The complexity of the proteome is a formidable barrier when accounting for all modifications; it is estimated that there are >100 000 proteins encoded by the  $\sim$ 20 000 genes and that the relative abundances of these proteins can vary across 10 orders of magnitude.<sup>3</sup>

MS is the most comprehensive and versatile tool for proteomics, and two major approaches have been developed: shotgun proteomics, suited for discovery, including PTMs; and selected reaction monitoring, suited for targeted quantifications and comparisons.<sup>4,5</sup> Both of these methods utilize proteolytic digestion to generate peptide fragments that are then detected using MS. In a typical shotgun proteomics experiment, the proteomic sample is digested with trypsins, fractionated, and subjected to liquid chromatography and tandem mass spectrometry (LC-MS/MS). Trypsin and the trypsin family of enzymes, including chymotrypsin, are widely used as the workhorse proteases of mass spectrometry-based

proteomics.<sup>6,7</sup> This is primarily due to the ability of these enzymes to cleave protein mixtures yielding peptides of mass in the preferred mass range for MS and with defined substrate specificity, thus providing readily interpretable and reproducible fragmentation spectra. Furthermore, these enzymes are fairly stable and can function in the presence of denaturants like urea that assist in the unfolding the target proteins that need to be proteolyzed prior to detection. There are, however, limitations to the nearly exclusive use of these enzymes in proteomics. (a) Any single protease covers only certain regions of the proteome through the library of peptides it generates, and (b) proteolytic efficiency is variable across different proteins within complex mixtures.<sup>7,8</sup> It is clear that the availability of proteases with orthogonal cleavage specificities will dramatically alter the coverage of the human proteome, including the annotation of the PTMs.<sup>7,8</sup>

Protein glycosylation is among the most abundant PTMs found in all domains of life. Despite their considerable compositional and structural diversity, glycosylation plays essential roles from protein folding to cellular homeostasis, and not surprisingly aberrant glycosylation is well-recognized in a number of diseases, including cancer.9,10 N-Linked

Received: June 23, 2019 Accepted: November 11, 2019 Published: November 11, 2019

Downloaded via UNIV OF HOUSTON MAIN on October 21, 2020 at 21:53:11 (UTC). See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.



**Figure 1.** Flow cytometric assay for the screening of recombinant chymotrypsin with expanded specificity for Asn. (A) The mature form of chymotrypsin (mChyB) is displayed on the surface of *E. coli* cells via fusion to the autotransporter, Ag43. A FLAG epitope tag is inserted C-terminal to the *mchyB* gene to enable monitoring of protein expression. (B) High-throughput, flow cytometric, single-cell screening assay. *E. coli* cells expressing mChyB variants were incubated with Asn-sub (selection) and Tyr-sub (counterselection) peptide sequences. *E. coli* cells expressing proteases with the desired specificity label with the Asn-sub show minimal labeling with Tyr-sub and are isolated using flow cytometry. To eliminate issues associated with protease-induced toxicity, plasmids were directly isolated from cells after FACS and re-transformed into fresh *E. coli* competent cells, which were subjected to further rounds of sorting. The FLAG epitope tag was used to normalize protease expression. (C) Sequences of the Asn-sub and Tyr-sub peptide substrates used for flow cytometry. The intact substrates have a +2 charge at pH 7. Proteolysis generates a free N-terminus [denoted by the (+)], and this causes the cleaved fragment to have a +3 charge. The arrow denotes the putative site of proteolysis.

glycosylation represents one of the best-characterized forms of glycosylation in eukaryotes with the glycans being attached to the amide group of asparagine (Asn) residues within proteins. Comprehensive annotation of the N-glycans of proteins requires quantitative identification of (a) the Asn residues that are modified, (b) N-glycan microheterogeneity, the diversity of the glycans that are attached to these Asn residues, and (c) N-glycan macroheterogeneity, the substoichiometric glycosylation of Asn residues.<sup>11</sup> For any given protein, the annotation of all of the Asn residues that are modified by glycans represents a fundamental first step to N-glycan analysis.

Engineering proteolytic specificity to enable detection of modified amino acids provides a direct route to the annotation of PTMs in proteomic experiments. Aided by advances in combinatorial screening, impressive successes have been reported in the systematic engineering of the substrate specificity of a wide range of microbial proteases.<sup>12–21</sup> Despite these advances in engineering, including proteases that target PTMs, none of the engineered proteases have been demonstrated to work in proteomic experiments. Recently, the substrate specificity of trypsin was expanded via engineering to also include citrulline, and proof-of-concept mapping experiments with a model protein were performed.<sup>22</sup> We report here the engineering and comprehensive characterization of the expanded substrate specificity of an engineered variant, chymotrypsiN, to recognize and cleave after unmodified Asn residues in proteins. We demonstrate that the engineered chymotrypsiN can be used to identify glycosylation sites and partially glycosylated Asn residues in

proteins. We envision that this protease not only can be used for mapping N-linked glycosylation sites but also can be added to the toolkit of proteases for proteomics to enable wider coverage of the proteome.

# RESULTS AND DISCUSSION

High-Throughput Assay for Tracking the Activity of Recombinant Chymotrypsin Libraries in *Escherichia coli*. We identified three challenges to the high-throughput engineering of the substrate specificity of proteases in *E. coli*: (i) display of the protease on the surface, (ii) an appropriate screening method for reporting proteolytic activity, and (iii) toxicity. Because proteases like chymotrypsin are toxic to the *E. coli* cells that express them, the gene encoding the variant had to be recovered from cells after sorting but without culturing.

Surface display of proteases ensures that the proteolytic activity can be assayed against any exogenous substrate, including synthetic peptide substrates containing PTMs. To facilitate surface display, we first constructed mature chymotrypsin B (mChyB) that contained only chains B and C. Within this construct, we introduced two additional mutations, Tyr164Ala and Cys140Ser. The Tyr164Ala mutation reduced (but did not abolish) the need for autoproteolytic processing, and the Cys140Ser mutation eliminated an unpaired Cys to minimize chances of misfolding (Figure S1). This engineered mChyB, containing a C-terminal FLAG epitope tag, was displayed on the surface of *E. coli* by fusion of the C-terminus to the *E. coli* autotransporter, Antigen 43 (Figure 1A).<sup>23</sup> This design ensured a free N-terminus of the



**Figure 2.** Isolation of chymotrypsin variants with altered specificity for Asn substrates. (A–D) Substrate preferences of populations of *E. coli* MC1061 cells displaying either WT-mChyB or mChyB variants isolated after sequential sorting. The variant at the end of the active site library was used as a template for the shuffled library, and the variant at the end of the shuffled library was used as a template for the shuffled library, and the variant at the end of the shuffled library was used as a template for the shuffled library, and the variant at the end of the shuffled library was used as a template for the error-prone library to yield the final variant, designated mChyB-Asn. (E) Amino acid changes in the mChyB-Asn variants in comparison to WT-mChyB. The average hydropathy index (HI), reflective of the hydrophobic character of the protein, is also reported for each of these proteins. The net HI of the amino acids that are being compared is also listed in parentheses for comparison.

mChyB, essential for proteolytic activity. The expression of full-length chymotrypsin on the cell surface was confirmed using an antibody specific for the FLAG tag (Figure S2).

To report the proteolytic activity of the individual cells expressing mChyB, we utilized a Förster resonance energy transfer (FRET)-based multiplexed assay (Figure 1B,C).<sup>24</sup> The peptide substrates contain a positively charged Arg tail and the amino acid sequence (AAPXGS, where X = Asn or Tyr) in the linker region sandwiched between the FRET pair (Figure 1C). Proteolysis of the substrate by the surface-displayed enzyme generates two fragments. The C-terminal fragment contains the Arg tail, the fluorescent tag, and the newly generated free N-terminus. This fragment now has an overall +3 charge and is captured locally on the bacterial cell surface due to electrostatic interactions. This provides a link between proteolytic activity and cellular fluorescence and thus enables the identification of cellular populations by fluorescence-activated cell sorting (FACS). By utilizing substrates tagged with different fluorescent tags, individual E. coli cells displaying mChyB that have high Asn and low Tyr activities are isolated using FACS while wild-type (WT)-like and nonspecific variants are rejected (Figure 1B). We specifically chose to counterselect against Tyr peptide substrates because the WT enzyme has high activity toward Tyr peptides. On the basis of other proteases that can cleave after Asn residues, we hypothesized that the steric hindrance provided by the sugar modification at Asn would block proteolysis by chymotrypsin variants.<sup>25,26</sup> Accordingly, we deemed counterselection against the sugarmodified Asn unnecessary.

Lastly, because the expression of mChyB significantly impacts the viability of *E. coli*, we directly isolated plasmids from sorted cells and re-transformed them into electro-competent cells, thus circumventing the toxicity issue.<sup>27</sup>

Collectively, these established a method for the high-throughput screening of chymotrypsin variants in *E. coli* cells.

Chymotrypsin Variant with Improved Activity toward P1 Asn Identified by Iterative Mutagenesis and FACS. Having established the screening method, we next focused on the construction of a suitable library. Within the chymotrypsinfold serine proteases, loops 185-192 and 215-226 have been identified as key determinants of substrate specificity based on their significance in evolutionary divergence.<sup>28</sup> These same loops have also been targeted in all of the previous protein engineering efforts to modify the P1 specificity of trypsin and chymotrypsin.<sup>29-32</sup> Because structural analyses suggested that residues 185-188 were farthest from the substrate, these were immediately excluded. Next, we posited that amino acid residues that are critical for the catalytic activity and structural stability are likely conserved across the protease family and hence should not be targeted for randomization. Accordingly, we utilized MUSCLE<sup>33</sup> to perform multiple-sequence alignment of amino acid sequences of 17 different proteases with comparable tertiary structures but having divergent substrate specificities. This analysis identified that residues 190, 191, 215, 220, and 225 were highly conserved (Figure S3). Excluding these amino acids gave us our list of 11 residues targeted for randomization: 189, 192, 216-219, 221-224, and 226. Independently, we used the crystal structure [Protein Data Bank (PDB) entry 1DLK] of  $\delta$ -chymotrypsin bound to a peptidyl chloromethyl ketone (substrate analogue)<sup>34</sup> as input to the web-based server, Hotspot wizard,<sup>35</sup> to identify key residues for randomization (Figure S4). Because the same 11 amino acids listed above were identified, we targeted these for diversification.

We constructed a partial saturation library by utilizing degenerate NNS (N = T, A, G, or C; S = G or C)



Figure 3. Kinetic and biophysical characterization of rChyB-Asn. (A) Kinetic characterization of purified, soluble WT-rChyB and rChyB-Asn with aminomethyl coumarin (AMC) fluorogenic substrates. Comparative assessments of the stability of proteases (B) to undergo autoproteolysis or (C) to function in the presence of urea. The relative activities are measured using the cognate AMC substrates. Error bars in all panels represent the standard deviation from at least triplicate measurements.

oligonucleotides targeting these 11 amino acids. Cloning and subsequent transformation into E. coli MC1061 cells yielded  $10^7$  transformants. These  $10^7$  transformants represent only a very small subfraction of the entire sequence space comprising randomization at these 11 amino acids (theoretically  $20^{11} \sim 2$  $\times$  10<sup>14</sup>), but it served as a starting point for exploration of enzyme variants with altered specificities. The expression of mChyB was induced by the addition of arabinose, and the library of cells was screened using 20 nM Asn-sub and 20 nM Tyr-sub peptides (Figure 1C). Cells displaying high Atto633 fluorescence (Asn proteolysis) and low BODIPY fluorescence (Tyr proteolysis) were isolated after six rounds of sorting. Ten randomly picked clones from this population were found to contain identical mutations in the mChyB gene. This variant labeled mChyB-Asn-v1 yielded 1.5-fold increased fluorescence with Asn-sub and 3.9-fold decreased fluorescence with Tyrsub, in comparison to that of WT mChyB (Figure 2A,B). Of the 11 positions targeted in the library, mChyB-Asn-v1 had mutations at 10 of them, Gly216 being the exception.

To remove any undesired mutations necessary for the observed change in substrate specificity, a second library was constructed by backcrossing the DNA encoding mChyB-Asnv1 with WT-mChyB. The resulting library of 10<sup>6</sup> transformants was screened using 20 nM Asn-sub and 20 nM Tyr-sub. During sorting of this library, we decreased the stringency of counterselection, i.e., allowed sorting of clones that displayed slightly higher activity toward Tyr-sub in comparison to the sort gate used for the initial library. We reasoned that some residual WT-like activity is essential for functional activation of mChyB (and mChyB variants) leading to active enzyme formation. After four rounds of FACS sorting, DNA sequencing of 10 random clones demonstrated that five amino acids (Ser189, Arg192, Arg218, Trp224, and Gly226) were conserved across all of the different clones tested, suggesting that they were essential for the altered specificity. mChyB-Asn-v2 that showed the highest fluorescence with Asnsub had two mutations (Met189 and Val217) that reverted back to Ser, as in WT mChyB. *E. coli* cells expressing mChyB-Asn-v2 displayed a 2.4-fold increased fluorescence with the Asn-sub in comparison to WT-mChyB (Figure 2C). Consistent with the relaxed counterselection, the activity of mChyB-Asn-v2 toward Tyr-sub also increased, although it was still 1.8-fold lower in comparison to that of WT-mChyB.

To identify beneficial mutations outside the 11 positions targeted, we constructed an error-prone library using the DNA sequence of mChyB-Asn-v2 as the template. The library of  $\sim 10^7$  transformants was again screened using 20 nM Asn-sub and 20 nM Tyr-sub. After four rounds of sorting, mChyB-Asn with an additional mutation, Val99Met, showed the highest fluorescence with Asn-sub [3.4-fold in comparison to that of WT-mChyB (Figure 2D)]. After three complete cycles of diversification, screening, and rescreening, we had established mChyB-Asn as the lead candidate enzyme that had the desired Asn proteolytic activity but diminished Tyr proteolytic activity.

At the molecular level, we observed a positive correlation between the polarity of the active site, quantified using the GRAVY<sup>36</sup> (grand average of hydropathy) score, and engineered activity toward the P1 asparagine residue (Figure 2E). Inspection of the mutations introduced into mChyB-Asn identified Arg192 (positive charge) and ion pair residues (Arg218 and Glu219) at key substrate binding pocket residues (Figure 2E). This change to a more positively charged substrate binding pocket (Figure S5) is consistent with the recognition of Asn as a substrate within other proteolytic enzymes.<sup>37</sup>

Characterization of Purified Chymotrypsin with Peptide Substrates. To characterize the enzymes in a purified, soluble form, we recloned the WT-mChyB and mChyB-Asn constructs into a mammalian pcDNA3.4 expression vector, as the zymogen form. These constructs included (1) the native leader peptide of chymotrypsinogen to facilitate secretion of the protease into the supernatants, (2)



Figure 4. Substrate specificity of WT-rChyB and rChyB-Asn profiled using HEK293F cell lysates, represented as a heat map.

chain A from WT-mChyB that contains the trypsin activation site, and (3) the C-terminal  $\text{His}_6$  tag for protein purification (Figure S6). This strategy yielded the recombinant zymogen of these proteins, WT-rChyB, and rChyB-Asn that could be activated by trypsin.

We transfected these plasmids into HEK293F cells. The supernatant containing the secreted protein was harvested and purified by His-tag affinity and cation exchange chromatography to >95% purity (as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis). Immediately prior to use, the proteases were activated by brief incubation with trypsin, and the trypsin was inactivated using a TLCK inhibitor. Reverse phase high-performance liquid chromatography (HPLC) analysis of the Asn-sub substrate  $(1 \ \mu M)$ digested with trypsin-activated rChyB-Asn (50 nM) showed one extra peak in addition to the substrate. LC-MS analysis confirmed that it was the C-terminal peptide fragment upon proteolysis after Asn (Figure S7). By contrast, we detected the intact Asn-sub substrate only when Asn-sub was co-incubated with trypsin-activated WT-rChyB (up to 250 nM) even after 24 h (Figure S7). These results provided direct confirmation that mChyB-Asn was capable of proteolysis after Asn residues.

Next, we measured the kinetics of WT-rChyB and rChyB-Asn using fluorogenic peptide substrates containing Tyr (LLVY-AMC), Asn (AAPN-AMC), or GlcNAc-linked Asn (AAPN[GlcNAc]-AMC) at the P1 position. WT-rChyB showed very fast kinetics toward LLVY-AMC even at low enzyme concentrations (1 nM). By comparison, at an identical Tyr substrate concentration of 50  $\mu$ M, rChyB-Asn even at a 50-fold higher enzyme concentration demonstrated a decrease in the rate of product generation by 20-fold (Figure 3A), confirming that Tyr activity was efficiently decreased with the aid of counterselection during library screening. The catalytic efficiency  $(k_{cat}/K_M)$  of LLVY-AMC proteolysis by WT-rChyB was  $(3 \pm 2) \times 10^7 \text{ M}^{-1} \text{ min}^{-1}$ , whereas the cleavage of AAPN-AMC by WT-rChyB was not detectable (Figure S8). The catalytic efficiency of AAPN-AMC cleavage by rChyB-Asn was  $(3 \pm 2) \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$  whereas the catalytic efficiency of LLVY-AMC proteolysis was determined to be  $(8 \pm 7) \times 10^3$ 

 $M^{-1}$  min<sup>-1</sup>. Proteolysis of the sugar-modified Asn substrate, AAPN(GlcNAc)-AMC, by rChyB-Asn was undetectable (Figure S8). Collectively, these kinetic data established that the engineered variant demonstrated an increase in Asn specific proteolysis and a >3000-fold reduction in Tyr specific proteolysis in comparison to that of WT-rChyB. These results also validated our hypothesis that the engineered variant was specific for unmodified Asn and could thus be utilized for proteomic experiments.

Because proteomic experiments routinely require digestion for extended periods of time, 12–48 h, and the protease has to function in the presence of protein denaturants like urea, we next evaluated these biophysical properties of the engineered variant in comparison to those of the WT enzyme. rChyB-Asn displayed enhanced resistance to autoproteolysis with a half-life of 67 h, in comparison to that of WT-rChyB that had a half-life of 8 h (Figure 3B). In the presence of 0.5 M urea, both enzymes demonstrated only modest decreases in activity, although at urea concentrations of >1 M both enzymes demonstrated <50% activity and the reduction in activity was larger for rChyB-Asn (Figure 3C). In aggregate, engineered chymotrypsiN demonstrated altered selectivity while demonstrating excellent stability.

Expanded Substrate Specificity of Engineered Chymotrypsin. To derive a more comprehensive map of the altered specificity of rChyB-Asn and to test its utility in a proteomic context, we applied a mass spectrometry-based approach to profile the selectivity of rChyB-Asn. We harvested and concentrated the secreted protein (secretome) fraction from HEK293F cells. This secretome was digested separately with activated WT-rChyB or rChyB-Asn or trypsin (as a control) and subjected to LC-MS/MS. To analyze the resulting proteomic data, we took advantage of the known substrate specificity of trypsin ([RK]|Not Pro) and chymotrypsin ([YWFLMN]|Not Pro). All peptides identified at a false discovery rate of  $\leq 1\%$  were mapped to the human proteome to yield 176 unique proteins within the secretome fraction (Table ST1). These 176 proteins comprised the custom database of proteins identified within the HEK293F

## **ACS Chemical Biology**



**Figure 5.** Mass spectrometric analysis of *S. cerevisiae* invertase (Suc2) digested using rChyB-Asn. Purified invertase was denatured, treated with the glycosidase EndoH, digested with mChyB-Asn, and characterized using LC–MS/MS. The combined mass spectrometry results from three independent digests with the protease were used to map the detected glycosylation patterns. The consensus NXS/T motif is underlined, and all Asn residues are shown in bold. Red denotes parts of the protein sequence detected by mass spectrometry, and green indicates Asn residues that were identified as being partially glycosylated. Asterisks identify sites of HexNAc modification. A representative example of the MS/MS spectra obtained is shown, which illustrates that the AEPILNISN peptide contains two Asn residues, of which only one is glycosylated. Note that rChyB-Asn did not cut after the glycosylated Asn but was able to mediate proteolysis after the unmodified Asn.

secretome in our experiments. Next, to identify the substrate specificity of WT-rChyB and rChyB-Asn in an unbiased manner, we performed nonspecific searches to match spectra to peptides within this custom database. We identified the P2-P2' positions in protein substrates cleaved by WT-rChyB (n =405) or rChyB-Asn (n = 338) (Tables ST2 and ST3) and generated a heat map representation of the substrate specificity of the proteases using IceLogo<sup>38</sup> (Figure 4A). Not surprisingly, WT-rChyB preferred hydrophobic amino acids at the P1 position. Unlike WT-rChyB, rChyB-Asn preferred Asn at position P1 in addition to the hydrophobic amino acids. We next inspected the specificity at the other subsites (P1', P2, and P2') specifically focusing on whether any specific amino acids were strongly disfavored as a result of the engineering. rChyB-Asn did not proteolyze substrates containing Cys at P1', but this is not surprising because the Cys is expected to be alkylated during sample preparation. The rest of the subsite preferences were largely similar to those of WT-chyB (Figure 4A). Thus, our enzyme engineering efforts resulted in expanding the natural substrate preference of chymotrypsin to include Asn at P1 and did not alter subsite preferences at other sites.

One of the advantages of using proteases with orthogonal specificity for mass spectrometry-based proteomics is the generation of unique peptides in a proteomic sample that either increases the coverage of a single protein or increases the number of unique proteins identified.<sup>39</sup> Consistent with the expanded specificity of rChyB-Asn, when we annotated the proteins identified in the HEK293F secretome digested with either WT-rChyB or rChyB-Asn, we observed qualitative differences in the nature of pathways (Tables ST1–ST3).

These results suggested that the coverage of protein mixtures was increased by the use of engineered rChyB-Asn.

Partially Glycosylated Asn Identified in a Model Protein Using Chymotrypsin. Glycan macroheterogeneity, defined as the heterogeneity due to substoichiometric glycosylation, is significant in physiological phenomena such as glycoprotein hormonal regulation and is relevant to quality control of recombinant glycobiologics produced.<sup>40</sup> We hypothesized that rChyB-Asn can be used to identify partially occupied N-glycosites as it would generate two signature peptides: (i) a nonglycosylated peptide with Asn residue in the C-terminus and (ii) a glycopeptide. We used the Saccharomyces cerevisiae invertase (Suc2) as the model protein to test this hypothesis because it contains 14 potential glycosylation sites (based on the NXS/T sequons, where X is any amino acid) that have been mapped previously.<sup>41</sup> This would allow us to test the utility of rChyB-Asn to detect incomplete or partial glycosylation, before using it for studying glycan macroheterogeneity in a proteomic sample.

The heterogeneity in glycan composition at any amino acid (microheterogeneity) presents a challenge for mapping *in silico*-generated peptides to spectra. Typically, a peptide-spectrum match (PSM) scoring function analyzes peptide and spectrum pairs (P, S) to assign a confidence value that the fragmentation of peptide P is observed in the experimental spectrum S. In matching these (P, S) pairs, the diversity of glycans results in low confidence values.

To address this issue, we digested Suc2 with the endoglycosidase H (Endo H) enzyme that trims glycans to leave only the N-acetylglucosamine (GlcNAc) moiety attached to the Asn residue. This eliminates microheterogeneity and ensures that the resulting peptide can harbor only two kinds of

## **ACS Chemical Biology**



Figure 6. Detection of proteins and N-glycan sites within the Jurkat secretome. (A) Pathway analyses of the 2676 unique proteins identified using mass spectrometric LC–MS/MS analyses of the Jurkat cell secretome digested with rChyB-Asn. Pathway and protein function classification was performed using PantherDB (www.pantherdb.org). MS/MS spectra of proteins (B) CALM-1, (C) FAM3B, and (D) aGPCR F5, which illustrate putative N-glycan macroheterogeneity.

Asn: unmodified Asn (no glycosylation) or Asn-GlcNAc (glycosylated). Subsequent to EndoH treatment, Suc2 was digested with rChyB-Asn for 72 h and the digested peptides were subjected to LC-MS/MS. We utilized a nonspecific PSM search against the invertase sequence with the criterion that addition of GlcNAc to Asn could be present as a variable modification, and all of the peptides (nascent and glycosylated) were identified (Table ST4).

To specifically focus on glycan macroheterogeneity, we analyzed the subset of peptides containing the NXS/T sequon (Table 1). Of the 14 sequons, two of them (2 and 11) were in a region of the protein not mapped by MS/MS, and hence, no assignment of glycosylation could be made (Figure 5A and Table 1). For five sequons (1, 3, 4, 6, and 7), we identified only the glycosylated peptide, suggesting complete glycosylation; for sequon 5, we detected only the nonmodified proteolyzed peptide, suggesting no glycosylation. Closely spaced, even overlapping, NXS/T sequons frequently occur, and often, at least one of them remains unglycosylated due to site skipping by the oligosaccharyltransferase complex.<sup>42</sup> In the invertase

sequence, sequons 4 and 5 overlap and only sequon 4 is glycosylated. This feature was correctly resolved using rChyB-Asn as proteolysis was observed at the C-terminus of Asn112 in sequon 5. Sequons 10, 12, and 14 displayed the dual signature characteristic wherein both the unproteolyzed but modified and proteolyzed but unmodified peptides were detected and were thus correctly identified as partially occupied glycosites. These mapping results were consistent with the previously reported glycosylation at these sites.

Mapping with rChyB-Asn also demonstrated some differences. For both sequons 8 and 9 (Asn266 and Asn275), we observed only the glycopeptide but not the nonglycosylated peptide, even though their occupancy frequency was measured previously to be only 10%. Second, although sequon 13 is predicted to be almost fully glycosylated, we detected only the unglycosylated peptide. This detection of the unglycosylated peptide suggests that this site cannot be completely glycosylated. In aggregate, these results collectively show that the engineered substrate specificity of rChyB-Asn can be exploited for glycoproteomic analyses of candidate proteins.

Large-Scale Analysis of N-Linked Glycosylation in the Jurkat Secretome. We next examined the utility of ChyB-Asn for analyzing the extracellular proteome of the human Jurkat T cell line and to identify glycan macroheterogeneity within this compartment. The cell-free supernatant fraction from cells grown in serum-free medium was first digested with rChyB-Asn, subsequently with peptide-N-glycosidase (PNGase) F, and then subjected to LC-MS/MS. A semispecific database search with a false discovery rate (FDR) of 1% identified 4379 peptides mapped to 2676 extracellular proteins (Figure 6A and Figure S9). The number of proteins identified by our engineered enzyme compares favorably with other independent studies of Jurkat extracellular proteins profiled using trypsin (Figure S10).<sup>43,44</sup> Within these peptides, 19% (808) were identified on the basis of the proteolysis at Asn, further confirming the modified specificity of the engineered enzyme (Table ST5).

To identify glycan macroheterogeneity, we analyzed the data set in two ways. First, within the peptides that were identified to arise from Asn proteolysis, we identified a set of 78 peptides that were (1) part of the NXS/T sequon and (2) annotated to be a glycosite by machine learning algorithms implemented within UniProt.45 These 78 peptides represent annotated glycosites that have been detected in our experiments not to have sugar modifications and to have been proteolyzed. This, in turn, implies that these sites within these proteins are completely unmodified or have glycan macroheterogenity. As an example, RTPENFPCKN<sub>318</sub>, which resulted from cleavage at Asn308 of human plasminogen (Figure S11), is known to display glycan macroheterogeneity. The presence or absence of sugar at Asn308 of human plasminogen can affect the rate of activation by tissue plasminogen activator and the binding affinity for cell surface receptors, ultimately affecting the rate of fibrinolysis of blood clots.<sup>2</sup>

Second, within all of the peptides, we identified the subset of 48 peptides with the Asn [N + 1] modification signature associated with PNGase F deglycosylation (Figure S9). Comparison of the 48 peptides, to the 808 Asn-cleaved peptides, led to the identification of three unique protein sites wherein we detected both the modified and unmodified peptides: Asn120 of FAM3B (UniProt ID P58499), Asn61 of calmodulin-1 (UniProt ID P0DP23), and Asn315 of adhesion G-protein-coupled receptor F5 (UniProt ID Q8IZF2) (Figure 6B–D). In summary, our proof-of-concept experiments with the engineered chymotrypsiN enabled the identification of glycosites that are either unglycosylated or substoichiometrically glycosylated within proteomic mixtures.

# DISCUSSION

Protease engineering has fascinated biochemists for decades. Most proteases have defined substrate binding pockets that help them differentiate the many different chemical properties of the amino acid side chains.<sup>47</sup> Trypsin and chymotrypsin, for example, have very similar tertiary structures and almost superimposable substrate binding pocket architecture (layout of the backbone).<sup>47,48</sup> Despite this similarity, site-directed mutational swapping of selected residues within the binding pocket of trypsin did not endow chymotrypsin-like reactivity<sup>49</sup> and required the grafting of extended loops outside the substrate binding pocket to bring about the change in substrate specificity.<sup>29</sup> Aided by advances in combinatorial screening, impressive successes have been reported in the systematic engineering of the substrate specificity of a wide range of

microbial proteases.<sup>12-21</sup> By contrast, engineering the specificity of mammalian proteases remains a formidable barrier with few reports of successful engineering. Comprehensive engineering of mammalian protease substrate specificity is a challenging task due to (1) structural complexity requiring disulfide bonds, (2) the need for zymogen activation, and (3) the capacity of proteases to cleave host proteins, resulting in cell death, or to cleave themselves leading to autoproteolysis. We have utilized a two-step strategy to engineer the substrate specificity of chyB to cleave after Asn and demonstrate its utility for proteomic and Asn-N-linked glycan mapping experiments.

We screened libraries of mChyB via fusion to the *E. coli* autotransporter Ag43 and utilized a high-throughput selection/ counterselection system. After three complete cycles of diversification and screening assisted by flow cytometry, we isolated a variant, chymotrypsiN, that contained nine mutations and that demonstrated both increased activity toward Asn-containing substrates and decreased activity toward Tyr-containing substrates. At the molecular level, chymotrypsiN displayed an increased positive charge within the substrate binding pocket, and this is consistent with the recognition of Asn substrates in other proteases.<sup>37</sup> We profiled the proteolytic activity of purified rChyB-Asn variant using standard AMC substrates, and mass spectrometry confirmed that the variant affected site specific cleavage after Asn.

One of the challenges in engineering proteases for proteomics applications is that the activity of proteases toward small peptide sequences, either natural or synthetically derived, does not always translate to cleavage of the same sequences in full-length proteins.<sup>19,50</sup> We specifically chose chyB as a starting scaffold for our engineering efforts because it is welldocumented that this family of proteases has the ability to cleave full-length proteins at their cognate recognition sites. To ensure that the engineered variants could be utilized for proteomics digestions, we performed biophysical experiments to confirm that rChyB-Asn had the ability to function in the presence of the commonly used denaturant, urea, and had enhanced autostability in comparison to WT-rChyB. Next, we evaluated the secretome of HEK293F cells subsequent to proteolytic digestion with either WT-rChyB or rChyB-Asn. These experiments allowed us to map the specificity of rChyB-Asn by surveying cleavage across a vast sequence landscape and demonstrated that rChyB-Asn had expanded substrate specificity to cleave after Asn.

The availability of proteases with distinct specificities makes it likely that different peptides are generated for mass spectrometry, and therefore, subjecting the same proteome or proteins to proteolysis by these orthogonal proteases increases the likelihood that complementary parts are sequenced leading to enhanced coverage of sequence space.<sup>7,8</sup> This parallel digestion approach using multiple proteases in parallel reactions has been utilized for the quantification of proteomes obtained from mammalian cells and viruses and leads to the increased frequency of identification of PTMs.<sup>7,8</sup> In this context, comparative profiling of the proteins identified by either WT-rChyB or rChyB-Asn showed qualitative differences in the peptides and proteins identified. We thus envision that rChyB-Asn can be added to the toolbox of proteases available for parallel digestion reactions.

To test the utility of mChyB-Asn in mapping N-linked glycosylation sites, we chose the well-studied glycoprotein

invertase because it harbors 14 sequons and up to 50% of the proteins' mass is contributed by complex sugars.<sup>51,52</sup> Our results mapping the glycosylation of invertase using chymotrypsiN were largely consistent with known mapping experiments, but there were also differences. According to our results, sequon 14 (Asn512) is partially glycosylated because we detected both peptides, but prior results have claimed both partial and complete glycosylation.<sup>52,53</sup>

Next, we validated the application of chymotrypsiN to proteomic studies of glycosylation by undertaking the mapping of the extracellular proteome of Jurkat T cells. The number of proteins identified by the aid of chymotrypsin is similar to the proteins identified using trypsin. More importantly, we identified 78 proteins that are either substoichiometrically glycosylated or nonglycosylated at putative N-linked glycosylation sites. For three of these proteins, we identified both the glycopeptide and the unmodified peptide confirming Nglycan macroheterogeneity. Although we have demonstrated the detection of candidate glycopeptides, these likely have some false positives because it has been previously demonstrated that spontaneous deamidation of Asn can happen independently of PNGase activity in MS/MS experiments.<sup>54</sup> To overcome this limitation, a more thorough annotation of N-glycan sites can be undertaken by the direct detection of glycopeptides as illustrated in a number of recent studies.55,56

There are a wide range of PTMs that are currently suboptimally annotated, and thus, engineering proteases specific for PTM offers a route to continuously expanding the set of proteases available for proteome-wide mapping. Although here we have demonstrated specificity for the unmodified parent amino acid as a mechanism for detecting PTM, PTMs like phosphorylation, due to the substoichiometric modification and labile nature, will likely require that the engineered proteases directly detect and cleave after the PTM. We have established proof-of-concept studies to demonstrate the utility of the engineered variant. More comprehensive studies will have to be undertaken to reliably map glycan macroheterogeneity within complex proteomes. Second, our kinetic data demonstrate that while the engineered rChyB-Asn demonstrated the requisite specificity in distinguishing modified and unmodified Asn (Figure S8), the overall catalytic efficiency of the enzyme can be substantially improved, in comparison to that of WT-rChyB. While we have employed chyB as the scaffold for the current studies because trypsin is the most commonly used protease for proteomics experiments and typically leads to detection of larger numbers of peptides and/or proteins, we are also exploring the engineering of more commonly utilized trypsins for detecting PTMs. Because the engineering of trypsin has also been recently reported, we are optimistic that this is a scalable problem.<sup>2</sup>

In summary, as demonstrated by our studies, we envision that chymotrypsiN can serve as a proteomic tool for mapping glycosylation. In broader terms, with the aid of our screening methodology, and surface display, other proteases like trypsin can be engineered to directly detect PTMs, including phosphorylation. This, in turn, will ensure more comprehensive coverage of the proteomes of living organisms leading to more fundamental discoveries.

## METHODS

**Construction of Chymotrypsin Libraries.** We constructed a target library randomizing residues 189, 192, 216–219, 221–224, and

226 of chymotrypsin B (Rattus norvegicus, UniProt ID P07388). We performed polymerase chain reaction (PCR) using primers 1 and 2 (Table ST5) and pBAD\_AChy\_ $700^{23}$  as a template to obtain fragment 1 encoding amino acids 206-245 of chymotrypsin B with a C-terminal FLAG tag and a *Kpn*I site for efficient restriction digestion. Fragment 1 was gel-purified and used as template in a PCR with primers 2 and 3 such that the region encoding amino acids 193-206 is added to its 5' end. This step was repeated using primers 2 and 4 to obtain fragment 2 encoding amino acids 183-245 with all of the desired positions randomized with the NNS codon. Fragment 3 containing the SacI restriction site with an additional overhang, Shine-Dalgarno sequence, and gene encoding the Ag43 signal peptide and amino acids 16-188 of chymotrypsin B was obtained using pBAD AChy 700 as a template and primers 5 and 6. One hundred femtomoles of gel-purified fragments 2 and 3 were first assembled together by 10 cycles of PCR and then amplified by primers 2 and 5 for the next 20 cycles. All primers were purchased from Integrated DNA Technologies (IDT) and are listed in Table ST6. After assembly, the gene library was gel-purified, digested with SacI-HF and KpnI-HF at 37 °C for 1 h, and ligated into digested pBAD\_700. The ligated plasmid was dialyzed against water for 1 h and transformed into E. coli MC1061 cells by electroporation. The resulting transformants were recovered with SOC medium ( $\sim 4 \times 10^{7}$ ) and directly grown in 100 mL of 2xYT medium supplemented with 0.5% glucose and 25  $\mu$ g mL<sup>-1</sup> chloramphenicol at 37 °C for 10 h. The cells were lysed, and their plasmid DNA isolated using a QIAGEN miniprep kit was stored at -20 °C; 300  $\mu$ L aliquots of cells (OD<sub>600</sub> ~ 2) in 20% (v/v) glycerol were frozen using liquid nitrogen and stored at -80 °C.

To create the backcross library, primers 7 and 8 with mixed bases were used. The procedure described above was repeated with primers 1 and 4 replaced by 8 and 7, respectively. Because more than one mixed base was used for randomization, the resulting library members also contained amino acids other than those that correspond to mChyB or mChyB-Asn-v1 in positions (Table ST6). For random mutagenesis of mChyB-Asn-v2, error-prone PCR was performed using the GeneMorph II kit (Agilent Technologies) with the mChyB-Asn-v2 gene as a template and primers 2 and 9. Assembly with the gene fragment encoding the Ag43 signal peptide obtained using primers 5 and 10, cloning into digested pBAD\_700, transformation, and cell recovery was performed as described previously.<sup>6</sup> For all of the libraries, plasmid DNA isolated from 10 randomly picked clones was sequenced (SeqWright Genomic Services) to assess the genetic diversity and the mutation rate for the library created using errorprone PCR was estimated to be 0.65%.

Synthesis of FRET Peptide Substrates. We resuspended the lyophilized peptide (KAAPNGSCGRGR, N-terminally acetylated and C-terminally amidated, >70% pure, Genscript) and dyes, Atto-633 maleimide (Atto-TEC GmbH, Singen, Germany) and QSY21 carboxylic acid succinimidyl ester (Life Technologies), in anhydrous *N*,*N*-dimethylformamide (DMF) to concentrations of 10  $\mu$ g  $\mu$ L<sup>-1</sup> (peptide) and 10 mM (dyes). Twenty microliters of a 1 M NaHCO<sub>3</sub> solution was mixed with 50  $\mu$ L of a peptide solution. Excess salt was pelleted to transfer the supernatant to a fresh tube, and 50  $\mu$ L of Atto-633 maleimide was added to the reaction mixture. After incubation at 25 °C for 1 h, an aliquot was analyzed on a C18 column (catalog no. 00G-4041-E0, Phenomenex) with water/acetonitrile mobile phases containing 1% acetic acid at a flow rate of 1 mL min<sup>-1</sup> by measuring absorbance at wavelengths of 630 and 670 nm via HPLC (Shimadzu scientific). For the second reaction, the peptide/Atto-633 reaction mixture was added to a tube containing 100  $\mu$ L of 1 M 4dimethylamino pyridine (DMAP) (Sigma-Aldrich) and 50  $\mu$ L of a QSY 21 solution. The reaction mixture was diluted with 4.5 mL of water containing 10% (v/v) acetic acid and loaded onto the column using a 5 mL sample loop. The chemical identity of the synthesized substrate was verified using ESI-MS (Rice University, Houston, TX). The desired fractions were pooled together and lyophilized overnight. After reconstitution in 100  $\mu$ L of water, the substrate (Asn-AtQ21) concentration was estimated by measuring the absorbance of dilutions in an Infinite 200 PRO plate reader (Tecan, Männedorf, Switzerland).

Screening of Chymotrypsin Libraries Using FACS. A 3 mL LB culture was inoculated using a glycerol stock of the library to a starting OD<sub>600</sub> of 0.02 and grown to an OD<sub>600</sub> of 0.5. We induced protein expression by the addition of 100  $\mu$ M arabinose at 37 °C for 2 h. The induced cells were washed with a 1% (w/v) sucrose solution and resuspended to an OD<sub>600</sub> of 1. The washed cells were incubated with 20 nM Tyr-BQ7 and 20 nM Asn-AtQ21 at an  $OD_{600}$  of 0.01 in a 1% (w/v) sucrose solution containing 2 mM Tris (pH 7.5) for 10 min at 25 °C. The labeled cells were analyzed using a BD Biosciences FACSJazz cell sorter at an event rate of 7000 s<sup>-1</sup>. Sorting was performed in the labeling time window of 10-25 min because with a longer incubation, cells could become labeled nonspecifically. For all of the libraries, the number of cells screened was at least 3-fold higher than their genetic diversity estimated on the basis of transformation efficiency. We recovered the plasmid DNA from the sorted cells using a Zymo miniprep kit (Zymo Research) and transformed into electrocompetent E. coli cells as described previously.<sup>7</sup> Ten colonies were randomly picked from the sorted population for clonal characterization using flow cytometry with Tyr-BQ7 and Asn-AtQ21 substrates. Mutations in chymotrypsin B corresponding to the clones that showed the desired phenotype on the cell sorter were identified by standard Sanger sequencing (SeqWright Inc.).

Enzyme Expression and Purification. We use the notations mChyB and rChyB to refer to the surface display form and the soluble form of the enzymes, respectively. The pcDNA 3.4-based plasmid vector, used for the expression of chymotrypsinogens in human embryonic kidney (HEK293F) cells, was a kind gift from the Georgiou lab (The University of Texas at Austin, Austin, TX). A gene fragment encoding rChyB-Asn with a His<sub>6</sub> tag and containing the Kozak sequence (GCCACC) and sequences overlapping with the 5' and 3' ends of the AgeI-HF/Bsu36I-digested vector for subsequent cloning by Gibson assembly was commercially purchased (IDT). The wild-type chymotrypsinogen gene (rChyB) was generated by overlapextension PCR of DNA fragments obtained using rChyB-Asn and mChyB as templates. After verifying the cloned plasmids by DNA sequencing, we inoculated a 100 mL 2xYT culture supplemented with  $200 \ \mu g \ mL^{-1}$  ampicillin with cells harboring *rChyB* or *rChyB-Asn*. The culture was grown overnight at 37 °C, and plasmid DNA was isolated using the QIAGEN plasmid maxi kit and QIAvac 24 plus vacuum manifold (Qiagen Inc.); 100 mL of HEK293F cells grown to a density of  $2.5 \times 10^6$  cells mL<sup>-1</sup> after three passages in a suspension medium culture at 37 °C and 5% CO<sub>2</sub> was transiently transfected with 100  $\mu$ g of plasmid DNA using 80 µL of Expifectamine transfection reagent (Life Technologies). Transfection enhancers provided in the kit were then added 16 h following transfection. The culture was grown for a period of 4-5 days to allow for protein secretion in the supernatant. Subsequently, we centrifuged the cells at 4000g for 20 min at 4 °C to collect the supernatant. We confirmed the presence of chymotrypsinogen in the supernatant by Western blotting using the rabbit anti-His antibody (300 ng mL<sup>-1</sup>, GenScript), the goat anti-rabbit antibody conjugated to HRP (40 ng mL<sup>-1</sup>, Jackson ImmunoResearch), and the chemiluminescent HRP substrate (SuperSignal West Pico, Thermo Scientific); 100  $\mu$ L of the supernatant was activated with 1 unit of proteomics-grade trypsin (Sigma-Aldrich) at 25 °C for 1 h. One microliter of a 5  $\mu$ g  $\mu$ L<sup>-1</sup> casein-FITC (50–100  $\mu$ g of FITC/mg of solid, Sigma-Aldrich) solution was added to monitor proteolytic activation by measuring fluorescence (excitation at 488 nm, emission at 530 nm, and cutoff at 515 nm) at 25 °C using a SpectraMAX Gemini EM plate reader (Molecular Devices).

After ascertaining the presence of chymotrypsinogen, we performed purification using an AKTA FPLC system (GE Healthcare Bio-Sciences) at 4 °C. The filtered (0.22  $\mu$ m) supernatant was loaded onto the column, and 12 mL of phosphate buffer was passed to elute any nonspecifically bound proteins. Elution was performed by increasing the imidazole concentration from 0 to 150 mM in a linear gradient in 16 min to collect 5 mL fractions.

An aliquot of the fractions expected to contain the desired protein was exchanged with Tris buffer [50 mM Tris and 10 mM CaCl<sub>2</sub> (pH 7.5)] using Zeba spin columns in a 96-well plate format (Thermofisher Scientific), and protease activity was confirmed using a caseinFITC substrate after activation. The pooled samples were further purified by cation exchange chromatography. One milliliter of concentrated protein was loaded onto a column packed with 0.8 mL of resin (GE Healthcare Bio-Sciences) by injection. The purity of fractions (1 mL volume) expected to contain a majority of the protein was estimated by running samples denatured using sodium dodecyl sulfate (SDS) and  $\beta$ -mercaptoethanol, in a 4 to 20% polyacrylamide gel (Lonza Houston Inc.) and subsequent Coomassie blue staining. Prior to any functional characterization, zymogens were activated using trypsin (1:100 enzyme:chymotrypsinogen ratio) at 25 °C for 1 h and subsequently treated with 500  $\mu$ M tosyl-L-lysine chloromethyl ketone hydrochloride (TLCK, Sigma-Aldrich) to inhibit trypsin.

**HPLC Characterization.** Asn-sub, at a concentration of 1  $\mu$ M, was incubated with activated chymotrypsin variants in Tris buffer (pH 7.5) at 37 °C for ~1 h. To estimate the degree of proteolysis, an aliquot of the reaction mixture was analyzed on a C18 column by measuring the absorbance at 630 and 670 nm corresponding to the Atto633 fluorophore and the QSY21 quencher, respectively, using the same gradient method that was previously used during substrate synthesis. To validate the identity of the peaks observed in the chromatogram, aliquots were also characterized by LC–MS (Mass Spectrometry Laboratory, Department of Chemistry, University of Houston).

Measurement of Kinetic Parameters. The concentration of total proteins was estimated by a bicinchoninic acid (BCA) colorimetric assay (Thermo Scientific) where the calibration curve was generated by measuring dilutions of the albumin standard (1 mg mL<sup>-1</sup>, Sigma-Aldrich). The concentration of functional chymotrypsin after zymogen activation was determined by active site titration against 4-methylumbelliferyl p-trimethylammoniocinnamate chloride (MUTMAC, Sigma-Aldrich). Conversion of relative fluorescence units (RFU) measured at 380 and 445 nm (excitation and emission wavelengths, respectively) to molarity was achieved by calibrating with different concentrations  $(0-5 \ \mu M)$  of the fluorophore, 4methylumbelliferone, in the presence of 50  $\mu$ M MUTMAC, and the rChyB zymogen incubated with TLCK-treated trypsin was used as a negative control. Kinetics of rChyB and rChyB-Asn toward suc-Ala-Ala-Pro-Asn-AMC (custom synthesized, Anaspec), suc-Ala-Ala-Pro-Asn(GlcNAc)-AMC (custom synthesized, Anaspec), suc-Ala-Ala-Pro-Phe-AMC (EMD Millipore), and suc-Leu-Leu-Val-Tyr-AMC (Sigma-Aldrich) were measured. AMC fluorescence, measured at 380 and 460 nm, was calibrated with dilutions of an unconjugated AMC stock solution. Depending on the kinetics, the enzyme concentration was in the range of 2.5-50 nM and substrate concentrations were varied from 20 to 500  $\mu$ M. For the suc-LLVY-AMC substrate, the highest concentration was limited to 50  $\mu$ M due to poor solubility. The enzyme concentration for AAPN(GlcNAc)-AMC ranged from 0 to 560  $\mu$ M, and no proteolysis was observed at any of the enzyme substrate concentrations tested.  $k_{cat}$  and  $K_{M}$  were determined by fitting the kinetic data to standard Michaelis-Menten or Lineweaver-Burk plots.

**Protease Stability Experiments.** Protease activity was assessed by incubating the suc-Ala-Ala-Pro-Phe-AMC substrate in activity buffer (50 mM Tris and 10 mM CaCl<sub>2</sub>) with either active WT rChyB or rChyB-Asn and measuring the increase in fluorescence with time using excitation at 380 nm and emission at 460 nm. To assess the stability in urea, 80 nM WT rChyB or 600 nM rChyB-Asn was incubated at the indicated concentrations at 25 °C for 20 min. For studying autoproteolysis, WT rChyB and rChyB-Asn were incubated at 37 °C at various intervals and aliquots were removed and stored at 4 °C. The cleavage activity of 80 nM WT rChyB and 400 nM rChyB-Asn at time points of 0, 2, 8, 25, and 50 h was measured.

**Digestion of the Secretome of HEK293F Cells.** HEK293F cells were cultured in Freestyle 293 serum-free medium (Thermo Scientific) and passaged three times. Cells were then seeded at a density of  $0.9 \times 10^6$  cells mL<sup>-1</sup> in 200 mL of fresh medium; the supernatant was harvested after 24 h and lyophilized, and the protein sample was dissolved in 0.5% (w/v) SDS. N-Linked glycosylation was trimmed by EndoH (New England Biolabs) treatment following the manufacturer's instructions. Then, the sample was reduced with 4

mM tris(2-carboxyethyl)phosphine (TCEP) at 37 °C for 15 min and alkylated with 4 mM iodoacetamide (IAA) for 30 min in the dark at room temperature (RT). The sample was purified by chloroform/ methanol precipitation and resuspended in 25  $\mu$ L of 6 M urea. The protein concentration was measured with a Nanodrop A280 instrument. Five microliters (65  $\mu$ g) of the sample was digested with engineered chymotrypsin, wild-type chymotrypsin, and trypsin in 100 mM ammonium bicarbonate buffer at 37 °C for 72 h. The reaction was stopped by adding formic acid to a final concentration of 1%.

**Digestion of the Jurkat Secretome.** Jurkat cells were cultured in RPMI 1640 containing 10% fetal bovine serum and supplemented with 50  $\mu$ g mL<sup>-1</sup> gentamicin and an insulin-transferrin-selenium supplement (Thermofisher Scientific) until a cell number of 6 × 10<sup>7</sup> was reached. Cells were resuspended in 100 mL of serum-starved RPMI1640, and the supernatant was harvested after 24 h. Secreted proteins were concentrated using an Amicon 10 kDa column, and protein content was measured using a BCA assay.

Eight milligrams of Jurkat secretome was reduced with 10 mM dithiothreitol (DTT) at 56 °C for 1 h and alkylated with 20 mM IAA at RT in the dark for 1 h. The proteins were subsequently washed thrice with 50 mM ammonium bicarbonate. Trypsin-activated rChyB-Asn was added to the Jurkat secretome at a 1:1 ratio at 37 °C for 3 days. Again, the peptides were washed with 100  $\mu$ L of 50 mM ammonium bicarbonate twice and lyophilized.

Glycopeptide enrichment was performed using the Protein Extract Glycopeptide Enrichment kit (Sigma-Aldrich) following the manufacturer's instructions. Enriched and depleted (peptides that did not bind to the glycocapture resin) samples were treated with PNGase F at a 1:50 ratio overnight at 37 °C. PNGase F-treated proteins were washed with ammonium bicarbonate lyophilized and resuspended in 0.1% formic acid prior to LC-MS/MS analysis.

Mass Spectrometry of the HEK293F Cell Secretome. Liquid chromatography and tandem mass spectrometry (LC-MS/MS) was performed on a ThermoFinnigan LTQ instrument equipped with an Agilent 1290 Infinity UPLC system using solvent A [water and 0.1% (v/v) formic acid] and solvent B (methanol and 0.1% formic acid). Five microliters (8  $\mu$ g) of the sample was injected for each mass spectrometry analysis. The peptides were separated in a home-packed 500  $\mu$ m  $\times$  6 cm C18 reversed phase column by a 30 min linear gradient of 20% (v/v) to 100% (v/v) solvent B with a flow rate of 30  $\mu$ L min<sup>-1</sup>. The electrospray voltage was set to 3.78 kV with a sweep, auxiliary, and sheath gas set to 0 on a standard IonMax ESI Source. The capillary temperature was set to 250 °C, and the mass spectrometer was set for dynamic exclusion data-dependent MS/MS with the three highest-intensity masses observed in the MS scan targeted for MS/MS fragmentation. The RAW data files were converted to MGF format using the MSConvert utility program from the ProteoWizard program suite (http://proteowizard.sourceforge. net). The data were first analyzed using an X! tandem advanced search against the human protein database using the following criteria: specific cleavage site [YWFLMN]|Not Pro for engineered and wildtype chymotrypsin, [RK]|Not Pro for trypsin, one missed cleavage allowed, precursor mass tolerances of +3 and -1 Da, product mass tolerance of 0.4 Da, carbamidomethyl cysteine as fixed modification, and other parameters at the default settings.

For the nonspecific search, peak lists obtained from MS/MS spectra were identified using OMSSA version 2.1.9 and X!Tandem version X! Tandem Sledgehammer (2013.09.01.1).<sup>57,58</sup> The search was conducted using SearchGUI<sup>59</sup> against the custom database of proteins identified (176 sequences) from the previous specific search. Protein identification was conducted against a concatenated target/ decoy version of the custom database. The decoy sequences were created by reversing the target sequences in SearchGUI. The identification settings were as follows: no cleavage specificity; 3.0 Da as MS1 and 0.5 Da as MS2 tolerances; fixed modifications, carbamidomethylation of C (+57.021464 Da); variable modifications, oxidation of M (+15.994915 Da); fixed modifications during the refinement procedure, carbamidomethylation of C (+57.021464 Da); variable modifications during the refinement procedure, acetylation of

the protein N-terminus (+42.010565 Da), Pyrolidone from E (-18.010565 Da), Pyrolidone from Q (-17.026549 Da), Pyrolidone from carbamidomethylated C (-17.026549 Da). Peptides and proteins were inferred from the spectrum identification results using PeptideShaker version 1.16.12.<sup>60</sup> Peptide Spectrum Matches (PSMs), peptides, and proteins were validated at a 1.0% false discovery rate (FDR) estimated using the decoy hit distribution.

Mass Spectrometry of the Jurkat Secretome. Mass spectrometry analysis of the Jurkat secretome was performed commercially. Digested peptides from the Jurkat secretome were analyzed using Nanoflow UPLC: Easy nLC1000 (Thermofisher Scientific) coupled to Orbitrap Q Exactive mass spectrometry (Thermofisher Scientific) using mobile phases A (0.1% formic acid in water) and B [0.1% (v/v) formic acid in acetonitrile]. The peptide mixture was separated in a home-packed 100  $\mu$ m  $\times$  10 cm column with a reverse phase ReproSil-Pur C18-AQ resin (3  $\mu$ m, 120 Å) with a flow rate of 600 nL min<sup>-1</sup> by applying a linear gradient from 6 to 30% B for 38 min, from 30 to 42% B for 10 min, 42 to 90% B for 6 min, and constant 90% to complete the 60 min program. The eluted peptides were electrosprayed at a voltage of 2.2 kV with a capillary temperature of 270 °C. Full MS scans were acquired in the Orbitrap with a resolution of 70000 at m/z400. In each Orbitrap survey scan, a full scan spectrum was acquired in the mass range of m/z 300–1650, followed by fragmentation on the 15 most intense peptide ions from the preview scan in the Orbitrap

The MS raw data of rChyB-Asn were analyzed and searched against a database of human-related proteins using Byonic version 2.15.7. Similarly to HEK293F analysis, protein identification was conducted against a target/decoy database with an estimated FDR of  $\leq$ 1%. A semispecific enzymatic search against YWFLMN was set with a maximum number of missed cleavages of two and a peptide molecular weight tolerance of 10 ppm. The MS/MS tolerances of 0.5 Da were allowed. The sole fixed modification parameter was carbamidomethylation (C), and the variable modification parameters were oxidation (O) and deamination (D) and the glycan modifications HexNAc(1), HexNAc(2), HexNAc(1)Fuc(1), and HexNAc(2)-Fuc(1). The peptide list from enriched and depleted fractions was combined and narrowed to proteins with a taxonomy of Homo sapiens. A total of 4380 peptides belonging to 2678 proteins are annotated as secreted by either (1) the Gene Ontology (GO) cellular component as either extracellular region (GO: 0005576), extracellular matrix (GO: 0031012), extracellular space (GO: 0005615), extracellular exosome (GO: 0070062), cell surface (GO: 0009986), external side of plasma membrane (GO: 0009897), or protein secretion (GO: 0009306), (2) proteins predicted to contain a signal peptide according to SecretomeP 2.0 or SignalP 4.1 servers, or (3) proteins classified as secreted by a nonclassical pathway with a SecretomeP 2.0 score exceeding 0.6.

**Invertase Digestion.** The invertase (Sigma-Aldrich) sample was treated with EndoH to remove all glycosylation except for the core GlcNAc. Then, invertase (0.6  $\mu$ g) was digested with engineered chymotrypsin at 1:1 ratio in 100 mM ammonium bicarbonate buffer and incubated at 37 °C for  $\leq$ 72 h. The reaction was stopped by adding 10% (v/v) formic acid to a final concentration of 1%. Four microliters of digested peptides (70 or 130 ng) was analyzed by Bruker MicrOTOF-Q mass spectrometry as described previously.<sup>61</sup> The data were searched against the invertase protein sequence database created in OMSSA using the following search criteria: no enzyme, two missed cleavages allowed, precursor mass tolerance of 1 Da, product mass tolerance of 0.4 Da, Asn HexNAc, Asn dHexHexNAc, deamidation of N and Q as variable modifications, and *E* value threshold set to 1.000e+000.

#### ASSOCIATED CONTENT

#### **S** Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acschembio.9b00506.

Supplementary figures (PDF) Mass spectrometry data (XLSX) Mass spectrometry data (XLSX)

#### AUTHOR INFORMATION

#### **Corresponding Author**

\*Department of Chemical and Biomolecular Engineering, University of Houston, Houston, TX 77204. Telephone: +1-713-743-1691. E-mail: nvaradar@central.uh.edu.

## ORCID <sup>©</sup>

Navin Varadarajan: 0000-0001-7524-8228

#### Author Contributions

<sup>§</sup>B.R. and S.A. contributed equally to this study.

#### Notes

The authors declare the following competing financial interest(s): N.V., S.A., and B.R. are inventors on a patent disclosure filed by the University of Houston.

## ACKNOWLEDGMENTS

This work was supported by CPRIT (RP180466, RP180674, and RP190160), an MRA Established Investigator Award to N.V. (509800), the Welch Foundation (E1774), the National Science Foundation (1705464), CDMRP (CA160591), and the Owens Foundation. This project was supported by the Cytometry and Cell Sorting Core at the Baylor College of Medicine with funding from the National Institutes of Health (CA125123 and RR024574) and the expert assistance of J. M. Sederstrom.

#### REFERENCES

(1) Kim, M. S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., Madugundu, A. K., Kelkar, D. S., Isserlin, R., Jain, S., Thomas, J. K., Muthusamy, B., Leal-Rojas, P., Kumar, P., Sahasrabuddhe, N. A., Balakrishnan, L., Advani, J., George, B., Renuse, S., Selvan, L. D., Patil, A. H., Nanjappa, V., Radhakrishnan, A., Prasad, S., Subbannayya, T., Raju, R., Kumar, M., Sreenivasamurthy, S. K., Marimuthu, A., Sathe, G. J., Chavan, S., Datta, K. K., Subbannayya, Y., Sahu, A., Yelamanchi, S. D., Jayaram, S., Rajagopalan, P., Sharma, J., Murthy, K. R., Syed, N., Goel, R., Khan, A. A., Ahmad, S., Dey, G., Mudgal, K., Chatterjee, A., Huang, T. C., Zhong, J., Wu, X., Shaw, P. G., Freed, D., Zahari, M. S., Mukherjee, K. K., Shankar, S., Mahadevan, A., Lam, H., Mitchell, C. J., Shankar, S. K., Satishchandra, P., Schroeder, J. T., Sirdeshmukh, R., Maitra, A., Leach, S. D., Drake, C. G., Halushka, M. K., Prasad, T. S., Hruban, R. H., Kerr, C. L., Bader, G. D., Iacobuzio-Donahue, C. A., Gowda, H., and Pandey, A. (2014) A draft map of the human proteome. Nature 509, 575-581.

(2) Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., Alm, T., Edqvist, P. H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., and Ponten, F. (2015) Tissue-based map of the human proteome. *Science* 347, 1260419.

(3) Wilhelm, M., Schlegl, J., Hahne, H., Moghaddas Gholami, A., Lieberenz, M., Savitski, M. M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J. H., Bantscheff, M., Gerstmair, A., Faerber, F., and Kuster, B. (2014) Mass-spectrometry-based draft of the human proteome. *Nature 509*, 582–587.

(4) Yates, J. R., 3rd. (2013) The revolution and evolution of shotgun proteomics for large-scale proteome analysis. *J. Am. Chem. Soc.* 135, 1629–1640.

(5) Kusebauch, U., Campbell, D. S., Deutsch, E. W., Chu, C. S., Spicer, D. A., Brusniak, M. Y., Slagel, J., Sun, Z., Stevens, J., Grimes, B., Shteynberg, D., Hoopmann, M. R., Blattmann, P., Ratushny, A. V., Rinner, O., Picotti, P., Carapito, C., Huang, C. Y., Kapousouz, M., Lam, H., Tran, T., Demir, E., Aitchison, J. D., Sander, C., Hood, L., Aebersold, R., and Moritz, R. L. (2016) Human SRMAtlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. *Cell* 166, 766–778.

(6) Vandermarliere, E., Mueller, M., and Martens, L. (2013) Getting intimate with trypsin, the leading protease in proteomics. *Mass Spectrom. Rev.* 32, 453–465.

(7) Tsiatsiani, L., and Heck, A. J. (2015) Proteomics beyond trypsin. *FEBS J.* 282, 2612–2626.

(8) Giansanti, P., Tsiatsiani, L., Low, T. Y., and Heck, A. J. (2016) Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat. Protoc.* 11, 993–1006.

(9) Christiansen, M. N., Chik, J., Lee, L., Anugraham, M., Abrahams, J. L., and Packer, N. H. (2014) Cell surface protein glycosylation in cancer. *Proteomics* 14, 525–546.

(10) Pinho, S. S., and Reis, C. A. (2015) Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* 15, 540–555. (11) Gil, G. C., Velander, W. H., and Van Cott, K. E. (2009) Nglycosylation microheterogeneity and site occupancy of an Asn-X-Cys

sequon in plasma-derived and recombinant protein C. Proteomics 9, 2555–2567.

(12) Varadarajan, N., Rodriguez, S., Hwang, B. Y., Georgiou, G., and Iverson, B. L. (2008) Highly active and selective endopeptidases with programmed substrate specificities. *Nat. Chem. Biol.* 4, 290–294.

(13) Varadarajan, N., Pogson, M., Georgiou, G., and Iverson, B. L. (2009) Proteases that can distinguish among different post-translational forms of tyrosine engineered using multicolor flow cytometry. *J. Am. Chem. Soc.* 131, 18186–18190.

(14) Varadarajan, N., Georgiou, G., and Iverson, B. L. (2008) An engineered protease that cleaves specifically after sulfated tyrosine. *Angew. Chem., Int. Ed.* 47, 7861–7863.

(15) O'Loughlin, T. L., Greene, D. N., and Matsumura, I. (2006) Diversification and specialization of HIV protease function during in vitro evolution. *Mol. Biol. Evol.* 23, 764–772.

(16) Verhoeven, K. D., Altstadt, O. C., and Savinov, S. N. (2012) Intracellular detection and evolution of site-specific proteases using a genetic selection system. *Appl. Biochem. Biotechnol.* 166, 1340–1354. (17) Sellamuthu, S., Shin, B. H., Han, H. E., Park, S. M., Oh, H. J., Rho, S. H., Lee, Y. J., and Park, W. J. (2011) An engineered viral protease exhibiting substrate specificity for a polyglutamine stretch prevents polyglutamine-induced neuronal cell death. *PLoS One 6*, No. e22554.

(18) Yi, L., Gebhard, M. C., Li, Q., Taft, J. M., Georgiou, G., and Iverson, B. L. (2013) Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. *Proc. Natl. Acad. Sci. U. S. A. 110*, 7229–7234.

(19) Yoo, T. H., Pogson, M., Iverson, B. L., and Georgiou, G. (2012) Directed evolution of highly selective proteases by using a novel FACS-based screen that capitalizes on the p53 regulator MDM2. *ChemBioChem* 13, 649–653.

(20) Callahan, B. P., Stanger, M. J., and Belfort, M. (2010) Protease activation of split green fluorescent protein. *ChemBioChem* 11, 2259–2263.

(21) Kostallas, G., and Samuelson, P. (2010) Novel fluorescenceassisted whole-cell assay for engineering and characterization of proteases and their substrates. *Appl. Environ. Microbiol.* 76, 7500– 7508.

(22) Tran, D. T., Cavett, V. J., Dang, V. Q., Torres, H. L., and Paegel, B. M. (2016) Evolution of a mass spectrometry-grade protease with PTM-directed specificity. *Proc. Natl. Acad. Sci. U. S. A. 113*, 14686–14691.

(23) Ramesh, B., Sendra, V. G., Cirino, P. C., and Varadarajan, N. (2012) Single-cell characterization of autotransporter-mediated Escherichia coli surface display of disulfide bond-containing proteins. *J. Biol. Chem.* 287, 38580–38589.

(24) Varadarajan, N., Rodriguez, S., Hwang, B.-Y., Georgiou, G., and Iverson, B. L. (2008) Highly active and selective endopeptidases with programmed substrate specificities. *Nat. Chem. Biol.* 4, 290–294.

(25) Manoury, B., Hewitt, E. W., Morrice, N., Dando, P. M., Barrett, A. J., and Watts, C. (1998) An asparaginyl endopeptidase processes a microbial antigen for class II MHC presentation. *Nature 396*, 695–699.

(26) An, H. J., Peavy, T. R., Hedrick, J. L., and Lebrilla, C. B. (2003) Determination of N-glycosylation sites and site heterogeneity in glycoproteins. *Anal. Chem.* 75, 5628–5637.

(27) Ramesh, B., Frei, C. S., Cirino, P. C., and Varadarajan, N. (2015) Functional enrichment by direct plasmid recovery after fluorescence activated cell sorting. *BioTechniques* 59, 157–161.

(28) Perona, J. J., and Craik, C. S. (1997) Evolutionary Divergence of Substrate Specificity within the Chymotrypsin-like Serine Protease Fold. *J. Biol. Chem.* 272, 29987–29990.

(29) Hedstrom, L., Szilagyi, L., and Rutter, W. J. (1992) Converting trypsin to chymotrypsin: the role of surface loops. *Science* 255, 1249–1253.

(30) Hung, S. H., and Hedstrom, L. (1998) Converting trypsin to elastase: substitution of the S1 site and adjacent loops reconstitutes esterase specificity but not amidase activity. *Protein Eng., Des. Sel.* 11, 669–673.

(31) Jelinek, B., Katona, G., Fodor, K., Venekei, I., and Gráf, L. (2008) The crystal structure of a trypsin-like mutant chymotrypsin: the role of position 226 in the activity and specificity of S189D chymotrypsin. *Protein J.* 27, 79–87.

(32) Reyda, S., Sohn, C., Klebe, G., Rall, K., Ullmann, D., Jakubke, H.-D., and Stubbs, M. T. (2003) Reconstructing the Binding Site of Factor Xa in Trypsin Reveals Ligand-induced Structural Plasticity. *J. Mol. Biol.* 325, 963–977.

(33) Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.

(34) Mac Sweeney, A., Birrane, G., Walsh, M. a., O'Connell, T., Malthouse, J. P. G., and Higgins, T. M. (2000) Crystal structure of  $\delta$ -chymotrypsin bound to a peptidyl chloromethyl ketone inhibitor. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 56*, 280–286.

(35) Pavelka, A., Chovancova, E., and Damborsky, J. (2009) HotSpot Wizard: a web server for identification of hot spots in protein engineering. *Nucleic Acids Res.* 37, W376–383.

(36) Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132.

(37) Dall, E., and Brandstetter, H. (2013) Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation. *Proc. Natl. Acad. Sci. U. S. A. 110*, 10940–10945.

(38) Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods* 6, 786–787.

(39) Giansanti, P., Tsiatsiani, L., Low, T. Y., and Heck, A. J. R. (2016) Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat. Protoc.* 11, 993–1006.

(40) Zacchi, L. F., and Schulz, B. L. (2016) N-glycoprotein macroheterogeneity: biological implications and proteomic characterization. *Glycoconjugate J.* 33, 359–376.

(41) Zeng, C., and Biemann, K. (1999) Determination of N-linked glycosylation of yeast external invertase by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mass Spectrom.* 34, 311–329.

(42) Shrimal, S., and Gilmore, R. (2013) Glycosylation of closely spaced acceptor sites in human glycoproteins. *J. Cell Sci.* 126, 5513–5523.

(43) Bonzon-Kulichenko, E., Martinez-Martinez, S., Trevisan-Herraz, M., Navarro, P., Redondo, J. M., and Vazquez, J. (2011) Quantitative in-depth analysis of the dynamic secretome of activated Jurkat T-cells. J. Proteomics 75, 561–571. (44) Wu, C. C., Hsu, C. W., Chen, C. D., Yu, C. J., Chang, K. P., Tai, D. I., Liu, H. P., Su, W. H., Chang, Y. S., and Yu, J. S. (2010) Candidate serological biomarkers for cancer identified from the secretomes of 23 cancer cell lines and the human protein atlas. *Mol. Cell. Proteomics 9*, 1100–1117.

(45) The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45, D158–D169.

(46) Mori, K., Dwek, R. A., Downing, A. K., Opdenakker, G., and Rudd, P. M. (1995) The activation of type 1 and type 2 plasminogen by type I and type II tissue plasminogen activator. *J. Biol. Chem.* 270, 3261–3267.

(47) Szabo, E., Venekei, I., Bocskei, Z., Naray-Szabo, G., and Graf, L. (2003) Three dimensional structures of S189D chymotrypsin and D189S trypsin mutants: the effect of polarity at site 189 on a protease-specific stabilization of the substrate-binding site. *J. Mol. Biol.* 331, 1121–1130.

(48) Hedstrom, L., Perona, J. J., and Rutter, W. J. (1994) Converting trypsin to chymotrypsin: residue 172 is a substrate specificity determinant. *Biochemistry* 33, 8757–8763.

(49) Venekei, I., Szilagyi, L., Graf, L., and Rutter, W. J. (1996) Attempts to convert chymotrypsin to trypsin. *FEBS Lett.* 379, 143–147.

(50) Timmer, J. C., Zhu, W., Pop, C., Regan, T., Snipas, S. J., Eroshkin, A. M., Riedl, S. J., and Salvesen, G. S. (2009) Structural and kinetic determinants of protease substrates. *Nat. Struct. Mol. Biol.* 16, 1101–1108.

(51) Reddy, V. A., Johnson, R. S., Biemann, K., Williams, R. S., Ziegler, F. D., Trimble, R. B., and Maley, F. (1988) Characterization of the glycosylation sites in yeast external invertase. I. N-linked oligosaccharide content of the individual sequons. *J. Biol. Chem.* 263, 6978–6985.

(52) Zeng, C., and Biemann, K. (1999) Determination of N-linked glycosylation of yeast external invertase by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mass Spectrom.* 34, 311–329.

(53) Ziegler, F. D., Maley, F., and Trimble, R. B. (1988) Characterization of the glycosylation sites in yeast external invertase. II. Location of the endo-beta-N-acetylglucosaminidase H-resistant sequons. J. Biol. Chem. 263, 6986–6992.

(54) Palmisano, G., Melo-Braga, M. N., Engholm-Keller, K., Parker, B. L., and Larsen, M. R. (2012) Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *J. Proteome Res.* 11, 1949–1957.

(55) Yang, W., Shah, P., Hu, Y., Toghi Eshghi, S., Sun, S., Liu, Y., and Zhang, H. (2017) Comparison of Enrichment Methods for Intact N- and O-Linked Glycopeptides Using Strong Anion Exchange and Hydrophilic Interaction Liquid Chromatography. *Anal. Chem.* 89, 11193–11197.

(56) Zhang, C., Ye, Z., Xue, P., Shu, Q., Zhou, Y., Ji, Y., Fu, Y., Wang, J., and Yang, F. (2016) Evaluation of Different N-Glycopeptide Enrichment Methods for N-Glycosylation Sites Mapping in Mouse Brain. *J. Proteome Res.* 15, 2960–2968.

(57) Geer, L. Y., Markey, S. P., Kowalak, J. A., Wagner, L., Xu, M., Maynard, D. M., Yang, X., Shi, W., and Bryant, S. H. (2004) Open mass spectrometry search algorithm. *J. Proteome Res.* 3, 958–964.

(58) Craig, R., and Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 20, 1466–1467.

(59) Vaudel, M., Barsnes, H., Berven, F. S., Sickmann, A., and Martens, L. (2011) SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics* 11, 996–999.

(60) Vaudel, M., Burkhart, J. M., Zahedi, R. P., Oveland, E., Berven, F. S., Sickmann, A., Martens, L., and Barsnes, H. (2015) PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat. Biotechnol.* 33, 22–24.

(61) Mali, S., Moree, W. J., Mitchell, M., Widger, W., and Bark, S. J. (2016) Observations on different resin strategies for affinity purification mass spectrometry of a tagged protein. *Anal. Biochem.* 515, 26–32.