

Extending Proteome Coverage with Legumain as a Highly Specific Digestion Protease

Wai Tuck Soh,[#] Fatih Demir,[#] Elfriede Dall, Andreas Perrar, Sven O. Dahms, Maithreyan Kuppusamy, Hans Brandstetter, and Pitter F. Huesgen*

Cite This: *Anal. Chem.* 2020, 92, 2961–2971

Read Online

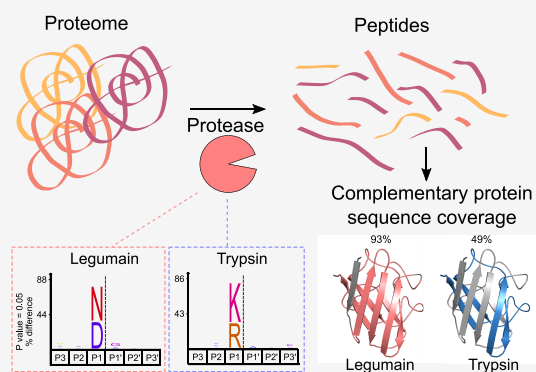
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Bottom-up mass spectrometry-based proteomics utilizes proteolytic enzymes with well characterized specificities to generate peptides amenable for identification by high-throughput tandem mass spectrometry. Trypsin, which cuts specifically after the basic residues lysine and arginine, is the predominant enzyme used for proteome digestion, although proteases with alternative specificities are required to detect sequences that are not accessible after tryptic digest. Here, we show that the human cysteine protease legumain exhibits a strict substrate specificity for cleavage after asparagine and aspartic acid residues during in-solution digestions of proteomes extracted from *Escherichia coli*, mouse embryonic fibroblast cell cultures, and *Arabidopsis thaliana* leaves. Generating peptides highly complementary in sequence, yet similar in their biophysical properties, legumain (as compared to trypsin or GluC) enabled complementary proteome and protein sequence coverage. Importantly, legumain further enabled the identification and enrichment of protein N-termini not accessible in GluC- or trypsin-digested samples. Legumain cannot cleave after glycosylated Asn residues, which enabled the robust identification and orthogonal validation of N-glycosylation sites based on alternating sequential sample treatments with legumain and PNGaseF and vice versa. Taken together, we demonstrate that legumain is a practical, efficient protease for extending the proteome and sequence coverage achieved with trypsin, with unique possibilities for the characterization of post-translational modification sites.



Current “bottom-up” mass spectrometry-based proteomics, also termed shotgun proteomics, can achieve near-complete proteome coverage and allows for extensive mapping of post-translational modification sites.¹ The basis of this approach is the selective protease-mediated digestion of isolated proteomes into peptides, which are then typically separated by reverse-phase liquid chromatography under acidic conditions and analyzed by tandem mass spectrometry (MS/MS). Peptides are subsequently identified by computational matching of the acquired spectra to proteome databases or spectral libraries, and the proteins present in the sample are inferred on the basis of the identified peptides.² The serine protease trypsin has become the dominant workhorse for the proteome digestions due to its high cleavage efficiency, high specificity for cleavage after Arg or Lys, and affordable price, even for high-quality preparations.³ Proteomes digested with trypsin therefore consist of predictable peptides with a C-terminal basic residue favorable for the ionization and generation of a dominant y-ion series, which facilitates database searches and peptide identification. However, about half of the peptides generated by trypsin are less than six residues long and therefore too small for identification and/or unambiguous assignment to specific protein sequences.⁴ Thus, many protein segments, including critical post-translational

modification sites, and even whole proteins remain invisible in proteome analyses relying on trypsin alone.³ This is especially true for proteolytic processing, a site-specific post-translational protein modification that can irreversibly alter protein function, interaction, and localization^{5,6} and thereby exert important signaling functions.⁷ Processed proteoforms are unambiguously identified by their new protease-generated neo-N-, or C-termini.^{8,9} The identification of neo-N-, and C-terminal peptides, which constitute a minor fraction among all peptides in a proteome digest, is facilitated by a variety of methods that have been developed to allow for their selective enrichment.⁵ However, many neo-N-, or C-terminal peptides are too short for mass spectrometry-based identification when only a single protease is used.⁹

Alternative proteases with a high sequence specificity are therefore of great interest and increasingly applied in bottom-up proteomics, including termini profiling approaches.^{3,10}

Received: August 7, 2019

Accepted: January 17, 2020

Published: January 17, 2020

Established proteases include AspN for cleavage before Asp and Glu; chymotrypsin for cleavage after Phe, Tyr, Leu, Trp, and Met; GluC (also known as *Staphylococcus aureus* protease V8) for cleavage after Asp and Glu; LysC for cleavage after Lys; LysN for cleavage before Lys;^{3,11} LysargiNase for cleavage before Arg and Lys;¹² and the prolyl endopeptidase neprosin that selectively cleaves after Pro and Ala.¹³ Also, proteases with broader sequence specificity such as elastase and thermolysin,¹⁴ proteinase K,¹⁵ subtilisin,¹⁶ and thermolysin WaLP and MaLP¹⁷ are occasionally applied but less favored due to the increased sample complexity with overlapping peptides and the less efficient spectrum-to-sequence matching due to the lack of a defined cleavage specificity as a restraint.¹⁸ Notably, digest with a single additional protease increases the number of protein identifications by an average of 7–8%¹¹ and enables the discovery of critical PTMs including phosphorylation sites^{16,19} and N-terminal processing sites^{10,20} that are missed in tryptic digests. Hence, there is a persistent strong demand for new, highly specific proteolytic enzymes with improved, complementary, or unexplored sequence specificity.³

Human legumain, also known as asparaginyl endopeptidase (AEP), is a well characterized caspase-like human cysteine protease known to cleave model substrates selectively after Asn and Asp residues.²¹ Recently, legumain cleavage specificity was further characterized by in-gel digestion of denatured complex proteomes that revealed pH-dependent differences in sequence specificity, with an optimal pH for cleavage after Asn and Asp at pH 6 and 4.5, respectively.²² On the basis of this data, it was further suggested that legumain may be a suitable choice as a precision digestion enzyme in proteomics applications.²² Encouraged by these reports, we reasoned that legumain might also be an attractive enzyme for standard in-solution digestion proteomics workflows. We show that the parallel digestion of proteomes isolated from *Arabidopsis thaliana* (*A. thaliana*) leaves, mouse embryonic fibroblasts (MEF), or *Escherichia coli* (*E. coli*) cell cultures with legumain, trypsin, and GluC results in the identification of distinct peptides that together increase protein sequence and proteome coverage. Legumain retained its remarkable specificity even under unfavorable conditions. N-terminome profiling demonstrated a strong complementarity to trypsin and superior performance compared to that of GluC. Asn is also the site of N-linked glycosylation, a common protein post-translational modification important in protein stability, folding, and protein–protein interaction.²³ By sequential processing with PNGase F and legumain, and vice versa, we demonstrate that N-glycosylation prevents legumain cleavage and propose that this tandem treatment strategy can provide orthogonal validation of N-glycosylation sites. Taken together, our data demonstrate that legumain is an attractive and reliable protease for the specific digestion of proteomes after Asn and Asp, with particular advantages for PTM site identification including processed N-termini and N-glycosylation sites.

■ EXPERIMENTAL SECTION

Expression, Purification, and Activation of Human Legumain. Human legumain was produced using the *Leishmania tarentolae* expression system (LEXSY) following a previously published protocol.²⁴ Briefly, legumain was recombinantly expressed as a secreted protein by a LEXSY suspension culture at 26 °C. The supernatant containing prolegumain protein was harvested by centrifugation and subjected to Ni²⁺-NTA affinity purification, followed by

desalting using PD-10 columns (GE Healthcare). Purified legumain was activated at 20 °C in a buffer containing 100 mM citric acid (pH 4.0), 100 mM NaCl, and 2 mM DTT. The progress of autoactivation was monitored by SDS-PAGE. Activated legumain was further purified using a PD-10 column (GE Healthcare) followed by size exclusion chromatography to have the active protein in a final buffer composed of 20 mM citric acid (pH 4.0), 50 mM NaCl, and 2 mM DTT. Legumain activity was evaluated using the legumain specific fluorescent substrate Z-Ala-Ala-Asn-AMC (AAN-AMC; Bachem) at a concentration of 50 μ M in assay buffer composed of 50 mM citric acid (pH 5.5), 100 mM NaCl, and 2 mM DTT at 37 °C. Fluorescence was detected using an Infinite M200 Plate Reader (Tecan) at 460 nm after excitation at 380 nm.

A. *thaliana* Proteome Preparation. *A. thaliana* Columbia (Col-8) leaves were harvested from 10 week old plants grown on soil under short day conditions (9 h/15 h photoperiod, 22 °C/18 °C, 120 μ mol of photons m⁻² s⁻¹) and snap frozen in liquid nitrogen. Leaves were ground in liquid nitrogen and resuspended in 10 mL/g fresh weight of extraction buffer (6 M Gua-HCl, 0.1 M HEPES (pH 7.4), 5 mM EDTA, 1 mM DTT, and HALT protease inhibitor cocktail; ThermoFisher, Dreieich, Germany). The suspension was homogenized using a Polytron PT-2500 (Kinematica, Luzern, Switzerland) and filtered through Miracloth (Merck, Darmstadt, Germany), and debris and nuclei were removed by centrifugation at 500g, 4 °C for 10 min. Proteins in the supernatant were purified by chloroform–methanol precipitation,²⁵ resuspended in extraction buffer, and reduced with 5 mM DTT at 56 °C, 30 min followed by alkylation with 15 mM iodoacetamide for 30 min at 25 °C. The reaction was quenched by the addition of 15 mM DTT for 15 min. The proteome extract was purified again with chloroform–methanol precipitation, resuspended in 0.2 mL of 0.1 M NaOH, and diluted with water and 1 M Hepes (pH 7.4) to a final concentration of 4 mg/mL in 0.1 M HEPES (pH 7.4). The protein concentration was quantified using the BCA assay (ThermoFisher, Dreieich, Germany). For digestion, aliquots of the concentrated *A. thaliana* proteome extracts were diluted at least four times to reach the required digestion buffer conditions, and the pH was confirmed with pH strips (Merck, Darmstadt, Germany).

Mouse Embryonic Fibroblast Proteome Preparation. Mouse embryonic fibroblast (MEF) cells were cultured in DMEM GlutaMax high glucose (Gibco 61965-026) supplemented with 10% FBS and 1 $\times$ penicillin/streptomycin (Gibco 15140-122) at 37 °C, 5% CO₂. Once the cells reached a confluency of up to 90%, the media were removed, washed with warm PBS, and trypsinized (Gibco 25300-054). The trypsinized cells were pelleted, washed twice with warm PBS to remove excess media, and lysed with 1% SDS 100 mM HEPES (pH 7.5) containing 1:50 (v/v) protease inhibitor cocktail (Sigma P8340). The sample was heated to 95 °C for 5 min, cooled, sonicated for 2 min, and heated again to 95 °C for 5 min to shear DNA. The protein concentration was measured, and 100 μ g of protein was used for each proteome digestion. Proteins were reduced with 10 mM DTT for 30 min at 37 °C and alkylated by the addition of 50 mM chloroacetamide (CAA) and incubation for 30 min at RT in the dark. The reaction was quenched by incubation with 50 mM DTT for 20 min at room temperature (RT) before purification with SP3 beads²⁶ and elution in the required digestion buffer.

E. coli Proteome Preparation. *E. coli* Dh5 α cells were grown in 200 mL of LB media until an optical density of OD₆₀₀ nm of 0.7. Cells were harvested by centrifugation at 400g for 15 min at 4 °C, washed by adding ice-cold PBS, and resuspended in 1 mL of lysis buffer (4% (v/v) SDS, 50 mM HEPES (pH 7.4), 5 mM EDTA, 1 $\times$ HALT protease inhibitor cocktail (ThermoScientific)) per 0.1 g of fresh weight. The cells were lysed by heating to 95 °C two times for 5 min, with 10 min of cooling with ice. Proteins were purified by chloroform–methanol precipitation and resuspended in 6 M Gua-HCl, 100 mM HEPES (pH 7.4), 5 mM EDTA, and the concentration was estimated using the BCA assay (Thermo-Fisher, Dreieich, Germany). One hundred micrograms of proteome was reduced by the addition of 10 mM DTT for 30 min at 37 °C and alkylated by the addition of 50 mM CAA for 30 min at RT in the dark, and the reaction was quenched by incubation with 50 mM DTT for 20 min at RT. The proteome was purified by chloroform–methanol precipitation and resolubilized in the appropriate digestion buffer.

Proteome Digestions. Proteome aliquots of 100 μ g were individually digested by legumain, GluC, or trypsin. The digestion with legumain was carried out in a reaction containing 0.1 M MES (pH 6.0), 0.1 M NaCl, and 2 mM DTT at a protease to proteome ratio of 1:50 (m:m), unless otherwise stated. For GluC (SERVA Electrophoresis, Heidelberg, Germany) digestion, the same amount of proteome was digested in PBS (pH 7.4) with a protease to proteome ratio of 1:50, whereas a 1:100 ratio was used for trypsin (SERVA Electrophoresis, Heidelberg, Germany) digestion in 0.1 M HEPES (pH 7.4) supplemented with 5% acetonitrile and 5 mM CaCl₂. The pH was confirmed using pH strips (Merck, Darmstadt, Germany), and the digestions were carried out at 37 °C overnight. For pH shift assays with legumain, an aliquot of the MEF proteome was digested at pH 6.0 for 5 h at 37 °C, and then the pH was lowered by the stepwise addition of 1 M HCl until pH 4.0 was reached. An additional 2 μ g of legumain and 1 mM DTT were added and incubated for another 5 h at 37 °C.

Mass Spectrometry. All samples were desalted using self-packed C18 Stop and Go Extraction tips as previously described.²⁷ Analysis was performed on a two-column nano-HPLC setup (Ultimate 3000 nano-RSLC system with Acclaim PepMap 100 C18, i.d. 75 μ m, particle size 3 μ m; trap column of 2 cm and analytical column of 50 cm length; ThermoFisher) with a binary gradient from 5 to 32.5% B for 80 min (A, H₂O + 0.1% FA; B, ACN + 0.1% FA) and a total runtime of 2 h per sample coupled to a high-resolution Q-TOF mass spectrometer (Impact II, Bruker) as previously described.²⁸ Data was acquired with the Bruker HyStar Software (v3.2, Bruker Daltonics) in line-mode in a mass range from 200 to 1500 *m/z* at an acquisition rate of 4 Hz. The top 17 most intense ions were selected for fragmentation with a dynamic exclusion of previously selected precursors for the next 30 s, unless an intensity increase of factor 3 compared to the previous precursor spectrum was observed. Intensity-dependent fragmentation spectra were acquired between 5 Hz, for low-intensity precursor ions (>500 cts), and 20 Hz, for high-intensity (>25k cts) spectra. Fragment spectra were acquired with stepped parameters, each with 50% of the acquisition time dedicated for each precursor: 61 μ s transfer time, 7 eV collision energy, and a collision radio frequency (RF) of 1500 Vpp followed by a 100 μ s transfer time, 9 eV collision energy, and a collision RF of 1800 Vpp.

Mass Spectrometry Data Analysis. Database searches were performed with MaxQuant²⁹ v.1.6.0.16 using standard Bruker Q-TOF settings that included peptide mass tolerances of 0.07 Da in the first search and 0.006 Da in the main search. *A. thaliana*, *M. musculus*, and *E. coli* protein databases were downloaded from UniProt (*A. thaliana* release 2018_01, 41350 sequences) with appended common contaminants as embedded in MaxQuant. The “revert” option was enabled for decoy database generation. For shotgun proteome samples, specificity was set to “unspecific” for the characterization of the cleavage specificity, otherwise according to the enzyme used (cleavage at K/RLX for trypsin, D/EIX for GluC, or D/NIX for legumain). Oxidation (M) and acetylation (protein N-term) were set as variable modifications, and the “match between runs” option was disabled. The analysis of the label-free shotgun data was performed with Perseus³⁰ v.1.6.1.1; the validation of the protein identification required at least two unique peptides for each protein and label-free quantification (LFQ) in at least two replicates. Searches for the N-termini were performed as described above, except that the enzyme specificity was set as Arg-C/GluC (DE)/legumain semispecific with a free N-terminus and duplex dimethyl labeling with light ¹²CH₂O formaldehyde or heavy ¹³CD₂O formaldehyde (peptide N-term and K). Oxidation (M), acetyl (N-term), Gln $\rightarrow$ pyro-Glu, and Glu $\rightarrow$ pyro-Glu were set as dynamic modifications, and the requantify option was turned off; the unspecific search window was set to 8–40 amino acids. Data evaluation and positional annotation for N-termini analyses were performed using an in-house Perl script (MANTL.pl; available at <http://MANTL.sourceforge.io>) that combines information provided by MaxQuant and UniProt to annotate and classify identified N-terminal peptides. In short, MaxQuant peptide identifications are consolidated by removing nonvalid identifications (peptides identified with N-terminal pyro-Glu peptides that do not contain Glu or Gln as N-terminal residue, peptides with dimethylation at N-terminal Pro), contaminant, reverse database peptides, and nonquantifiable acetylated peptides in multichannel experiments (no K in peptide sequence to determine labeled channel). For N-terminal peptides mapping to multiple entries in the UniProt protein database, a “preferred” entry was determined in a binary decision tree. Protein entries where the identified peptide matched positions 1 or 2 were preferred over alternative positions, and then manually reviewed UniProt protein entries were favored over alternative models. If multiple entries persisted, the alphabetically first entry was used to retrieve positional annotation information. For the visualization of protein sequence coverage, protein structures were modeled with the Phyre2 server.³¹

Enrichment of N-Terminal Peptides. Protein N-terminal peptides were enriched using the high-efficiency undecanal-based N-termini enrichment (HUNTER) method essentially as previously described.³² Briefly, equal amounts of *A. thaliana* proteome were dimethyl labeled with 20 mM heavy (¹³CD₂O) or light (CH₂O) formaldehyde and 20 mM sodium cyanoborohydride at 37 °C for 16 h to block all primary amines. To ensure a complete reaction, the same concentration of reagents was added again and incubated for another 2 h. Proteins were purified by chloroform–methanol precipitation to remove excess reagents and dissolved in 0.1 M HEPES (pH 7.4), and the protein concentration was estimated using the BCA assay according to manufacturer instructions (Thermo-Fisher, Dreieich, Germany). The samples (400 μ g/sample)

were digested with legumain, GluC, and trypsin at 37 °C for 16 h in the respective digestion buffers and protease–proteome ratios as described above. The protease-generated peptides were hydrophobically tagged with undecanal using an undecanal–proteome ratio of 50:1 and supplemented with 20 mM sodium cyanoborohydride in 40% ethanol at 50 °C for 45 min. The reaction was extended by the addition of 20 mM sodium cyanoborohydride for another 45 min. The reaction was then acidified with a final 1% TFA and centrifuged at 21000g for 5 min to precipitate free undecanal. Supernatant was injected to a preactivated HR-X (M) cartridge (Macherey-Nagel, Düren, Germany). The flow-through containing N-terminal peptides was collected. Remaining N-terminal peptides on the HR-X (M) cartridge were eluted with 40% ethanol containing 0.1% TFA, pooled with the first eluate, and subsequently evaporated in the SpeedVac to a small volume suitable for C18 StageTips purification.

Identification of Glycosylation Sites. Apoplastic fluid proteome enrichment was carried out as described³³ with some modifications. The whole *A. thaliana* rosettes were infiltrated with cold sterile water in a SpeedVac for 3 min at a pressure between 600 and 2500 Pa. The infiltrated rosettes were then centrifuged at 4 °C, 3000g for 10 min into a collection tube containing a Halt protease inhibitor cocktail (ThermoFisher, Dreieich, Germany). Extracted apoplastic fluid proteins were purified by chloroform–methanol precipitation and resuspended in 50 mM HEPES (pH 7.4). The protein concentration was quantified by using the BCA assay. The sample was then reduced with 5 mM DTT at 56 °C for 30 min and alkylated with 15 mM iodoacetamide at 25 °C for 30 min in the dark, and the reaction was quenched with 15 mM DTT at 25 °C for 15 min. The protein extract was then separated into two aliquots. One aliquot of 100 μ g of apoplast proteome was treated with PNGase F (SERVA Electrophoresis, Heidelberg, Germany) for 2 h at 37 °C before legumain digestion with protease at a ratio of 1:50 at 37 °C, pH 6 (pH adjusted with final concentration of 0.1 M MES pH 6.0). In parallel, another 100 μ g of protein extract was predigested with legumain and then treated with PNGase F using the same conditions. The samples were subsequently dimethyl labeled with 20 mM heavy ($^{13}\text{CD}_2\text{O}$) and light (CH_2O) formaldehyde and 20 mM sodium cyanoborohydride at 37 °C for 2 h. The reactions were quenched with 0.1 M Tris pH 7.4 at 37 °C for 1 h and pooled in a 1:1 ratio, and peptides were purified by C18 StageTips.

Data Deposition. MS data have been deposited to the ProteomeXchange Consortium³⁴ (<http://www.proteomexchange.org>) via the PRIDE³⁵ (<https://www.ebi.ac.uk/pride/archive/>) partner repository: PXD014696 for data relating to comparative proteome digestion with legumain, GluC, and trypsin, PXD014699 for *A. thaliana* proteome digested by legumain in the presence of various denaturants, PXD014698 for various pHs, PXD014697 for HUNTER N-termini profiling of *A. thaliana* leaves, and PXD014680 for N-glycosylation site mapping.

RESULTS

Legumain Cleaves Denatured Proteomes Exclusively after Asn and Asp. Previous data obtained by in-gel protein digestion-based specificity profiling²² and by biochemical characterization with test peptides²¹ suggested that legumain cleaves substrates C-terminally to Asn and Asp residues in a pH-dependent manner, with optimal activity and high

selectivity for Asn-containing substrates near pH 6. To test whether this exquisite specificity holds true under in-solution proteome digest conditions, we digested three aliquots of a denatured *A. thaliana* proteome with legumain at pH 6.0 for 18 h. In parallel, we digested three aliquots of the same proteome with trypsin and GluC at pH 7.4. To determine protease cleavage site specificity, peptides were analyzed by nano-LC–MS/MS and the acquired spectra were matched to the UniProt *A. thaliana* proteome database using nonspecific search settings, i.e., without defining an enzyme cleavage specificity. This unbiased search identified 4452, 4078, and 7985 peptide sequences in legumain, GluC, and tryptic digests, respectively, from which we compiled 6300, 5673, and 12107 unique nonredundant cleavage sites based on the sequence surrounding both ends of the identified peptides. For legumain, 93.3% of the observed cleavage sites were Asn and Asp (51.0% after Asn, 42.3% after Asp). A small percentage of unspecific cleavage is expected because of endogenous background proteolysis. The percentage of specific cleavage in a whole proteome is comparable to 96.7% of cleavages after Lys and Arg, as observed for trypsin (58.0% after Lys, 38.7% after Arg), and more stringent than the 85.4% cleavages after Glu and Asp (72.7% after Glu, 12.7% after Asp), as observed for GluC. The visualization of the relative amino acid abundance surrounding the cleavage sites with IceLogos reflected the strict specificity at the P1 position, preceding the hydrolyzed peptide bond in all three enzymes (Figure 1a–c). While GluC (Figure 1b) and trypsin (Figure 1c) do not allow cleavage before proline (P1' position), this is not the case for legumain (Figure 1a). We further analyzed a single replicate of a mouse embryonic

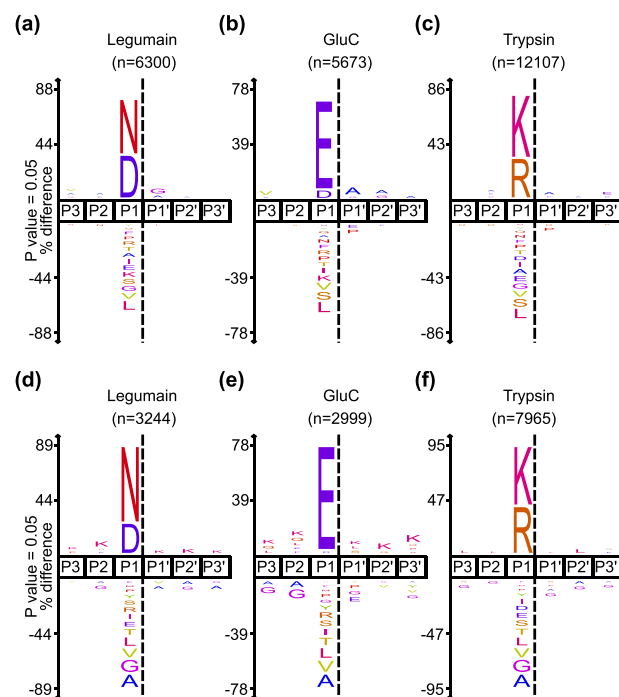


Figure 1. Substrate cleavage specificity of legumain, GluC, and trypsin. IceLogos visualize the amino acid frequencies surrounding the cleavage sites inferred from peptides identified by nonspecific database searches after digestion of (a–c) an *A. thaliana* leaf proteome or (d–f) mouse embryonic fibroblast cell lysate proteome with (a,d) legumain, (b,e) GluC, or (c,f) trypsin. The numbers of nonredundant cleavage sites for each logo are indicated.

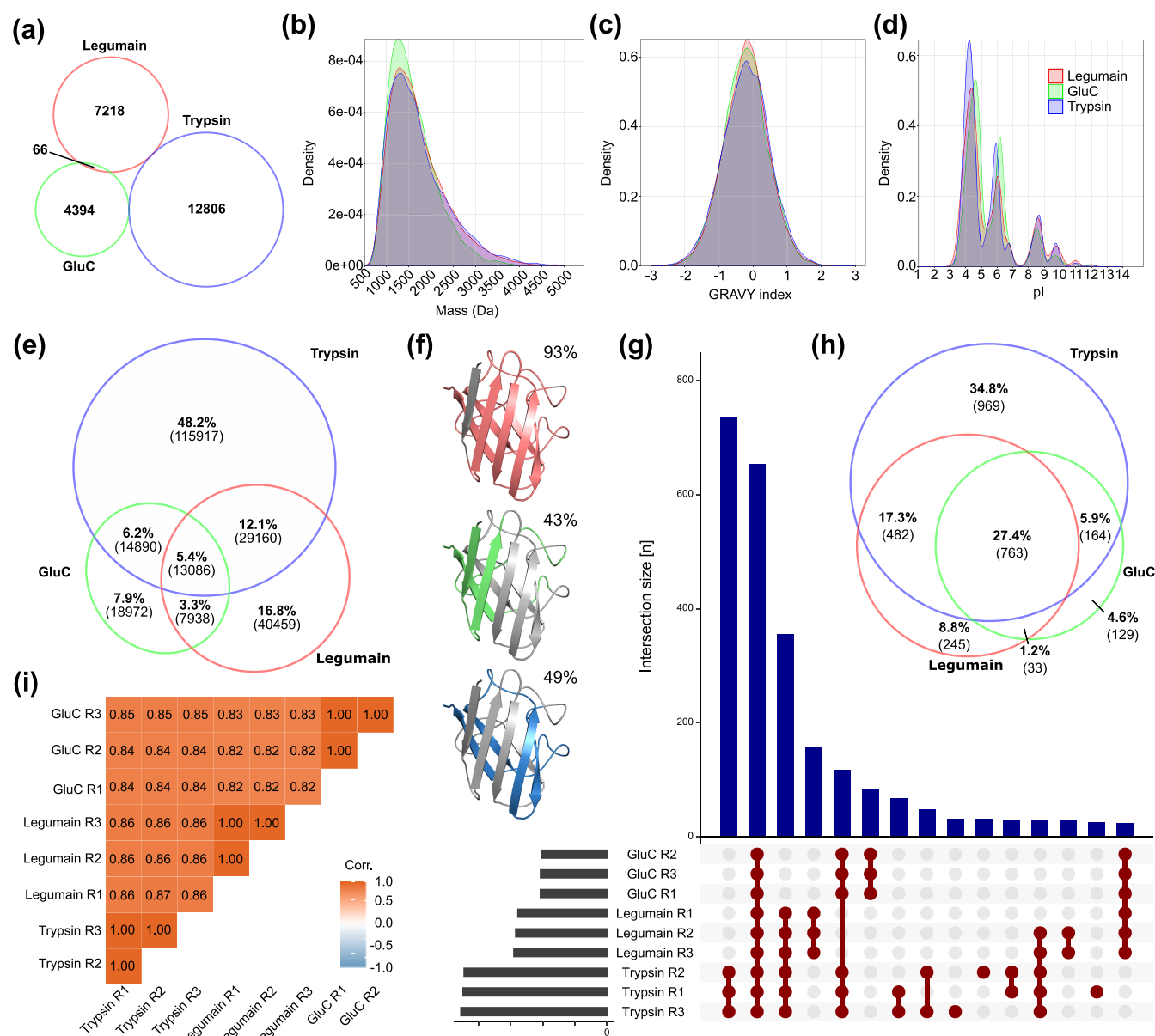


Figure 2. Analysis of an *A. thaliana* leaf proteome digested with legumain, GluC, and trypsin, each performed in three technical repeats. (a) Overlap of unique peptide sequences identified using enzyme-specific database queries. Analysis of the (b) mass, (c) hydrophobicity, and (d) isoelectric point of the identified peptides. (e) Overlap in unique amino acids identified by digestion with the three proteases. (f) Protein sequence coverage observed for superoxide dismutase (At1g08830) in legumain (red, 93%), GluC (green, 43%), and trypsin (blue, 49%) proteome digests. (g) Upset plot showing the overlap in protein groups identified in individual technical digestion replicates. (h) Venn diagram showing the total overlap of protein groups identified by the three enzymes. (i) Reproducibility of proteome quantification (MaxQuant LFQ). Only proteins quantified with two or more peptides were considered. Value indicates the Pearson correlation between the LFQ values obtained for technical replicates.

fibroblast proteome and identified 1893, 1722, and 4377 peptides using nonspecific database searches after digestion with legumain, GluC, and trypsin. Similar specificity profiles were obtained on the basis of the 3244, 2999, and 7965 nonredundant cleavage sites derived from the peptides in legumain (Figure 1d), GluC (Figure 1e), and trypsin (Figure 1f) digests, again showing that legumain tolerates Pro at P1' (Figure 1d). Of the cleavages observed in legumain digest, 94.5% matched the expected specificity (63.6% after Asn, 30.9% after Asp), 97.6% in the tryptic digest (51.9% after Arg, 45.7% after Lys), and 85% in the GluC digest (76.6% after Glu, 8.4% after Asp). These observations were further confirmed by

analyses of an *E. coli* proteome (Supporting Information (SI) Figure S1), where 2681 peptides identified after legumain digestion yielded 4187 cleavage sites with 86.2% cleavage after Asn and Asp (53.1% after Asn, 33.1% after Asp), while 85.3% of the 8597 unique cleavages observed in 5374 peptides identified after tryptic digest matched the expected specificity (44.1% after Arg, 41.2% after Lys).

Complementary Protein Sequence Coverage by Digestion with Legumain Compared to GluC and Trypsin. With the strict cleavage specificity of legumain under proteome digest conditions confirmed by the unbiased database search, we repeated spectra-to-sequence matching

using standard enzyme-specific settings with up to three missed cleavages, using cleavage after Asn and Asp as a specificity rule for legumain. As expected, the smaller search space significantly increased the number of peptide identifications in the *A. thaliana* data set by 64%, 8%, and 66% to 7284, 4394, and 12806 unique peptide sequences for legumain, GluC, and trypsin, respectively (Figure 2a). Specific searches of the MEF proteome data set increased peptide identifications by 129%, 73%, and 61% to 4296, 2983, and 8489 unique peptides for legumain, GluC, and trypsin, respectively, compared to results for nonspecific searches. In *E. coli*, peptide identifications improved by 33% and 7% to 3568 and 5767 unique peptides for legumain and trypsin.

While trypsin showed the expected superior performance, legumain digests resulted in the identification of more peptides than GluC, for example, 66% more in the *A. thaliana* data set. Interestingly, the legumain and GluC data sets showed only a minimal overlap of 66 identical peptides delimited by cleavages after Asp on both sides, which may occur with both enzymes but are not favored by GluC under the applied reaction conditions (Figure 2a). The analysis of the mass (Figure 2b), hydrophobicity (Figure 2c), and isoelectric point (Figure 2d) of the identified *A. thaliana* peptides revealed very similar properties for all three enzymes. In contrast, the biophysical properties of all theoretical peptides in *in silico*-digested *A. thaliana* and *M. musculus* proteomes predicted a higher number of peptides with $pI > 9$ in GluC- and legumain-digested proteomes compared to those with trypsin (SI Figure S2a,b). However, a comparison to our data (Figure 2b–d) suggests that such peptides are rarely identified with the standard experimental setup with reverse-phase chromatography under acidic conditions and ionization and mass spectrometric analysis in positive ion mode. Despite these physical similarities, peptides identified after digestion with the three proteases covered distinct amino acids in the identified *A. thaliana* proteins (Figure 2e).

In total, the parallel application of legumain, GluC, and trypsin in technical triplicates identified 1524, 1090, and 2380 protein groups in the *A. thaliana* proteome, respectively, combining to a total of 2785 protein groups, with legumain contributing 8.8% exclusive identifications (Figure 2g,h, SI Table S1). As expected from the number of peptide identifications, a large majority of 2057 proteins (74.3%) had the highest sequence coverage in the tryptic digest, followed by 507 (18.3%) in legumain digests and 206 (7.4%) in GluC digests (SI Table S1). For example, the sequence coverage of superoxide dismutase (At1g08830) (Figure 2f, SI Figure S3a) was a remarkable 93% in legumain digests compared to 43% and 49% in the GluC and trypsin data sets, and sequence coverage of the germin-like protein 1 (At1g72610) was at 63% with legumain compared to only 23% and 8% with GluC and trypsin (SI Figure S3b). Notably, for each of the three proteases >80% of the proteins were identified in all three replicates, indicating a high degree of reproducibility in the digests (SI Figure S4a). On the single replicate level, the combination of any tryptic digest with any legumain or GluC digest resulted in a slightly higher number of protein identifications than any two tryptic replicates combined (SI Figure S4b). We further compared reproducibility by label free proteome quantification (LFQ) with MaxQuant after filtering for protein groups quantified by two or more peptides (SI Table S2). This demonstrated excellent correlation of the LFQ values between the technical digestion replicates and also a

high correlation between the LFQ values obtained from digests of the three different proteases (Figure 2i).

In the MEF proteome, 1469, 1140, and 2242 protein groups were identified in legumain, GluC, and tryptic digests, combining to 2587 protein groups in total, with 7.7% exclusively identified in the legumain digests (SI Figure S4c). A larger overlap was observed between the *E. coli* proteome digests, where 842 and 1180 protein groups were identified after legumain and tryptic digestion, respectively, but only 37 (3%) of these were exclusive for legumain (SI Figure S4d).

Legumain Cleaves after Asn More Efficiently than after Asp. The digestion efficiency of a protease can be reflected by the number of missed potential cleavage sites within the identified peptides. In the *A. thaliana* data set, legumain generated on average 53% of the peptides without missed cleavage sites, 34% with one missed potential cleavage site, and 13% with more than one missed cleavage site (Figure 3a). GluC performed worse, with only 30% of the peptides

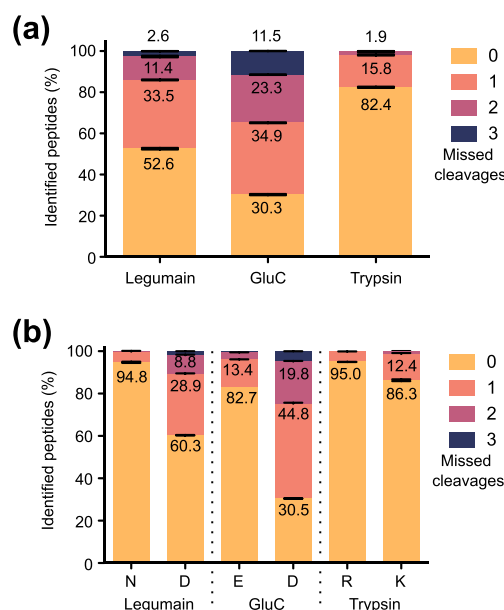


Figure 3. Potential cleavage sites missed by legumain, GluC, and trypsin in *A. thaliana* leaf proteome digests. (a) Percentage of peptides containing up to three missed cleavage sites. (b) Missed cleavage sites sorted by missed amino acid residues.

with no missed cleavage, but almost 12% of the identified peptides containing three missed cleavage sites. Trypsin was the best performing enzyme, with only 18% of the peptides containing one or more missed cleavage sites (Figure 3a). When we further considered the identity of the amino acid residue, we noted that legumain reliably cleaved after internal Asn residues, with only 5% of the peptides containing an internal Asn, but it missed one or more Asp in 40% of the peptides (Figure 3b). Most missed cleavage sites in GluC-digested proteomes were at Asp, and even trypsin showed a higher fidelity at Arg than at Lys (Figure 3b). Remarkably, legumain cleaved after Asn residues as efficiently as trypsin at the favored Arg-containing cleavage sites. Similar trends were observed in digests of MEF and *E. coli* proteomes, where legumain digests consistently showed a high cleavage efficiency at Asn sites with more missed cleavages at Asp (SI Figure S5).

Assessing Legumain Efficiency in Different Reaction Conditions. Previous publications have shown that legumain

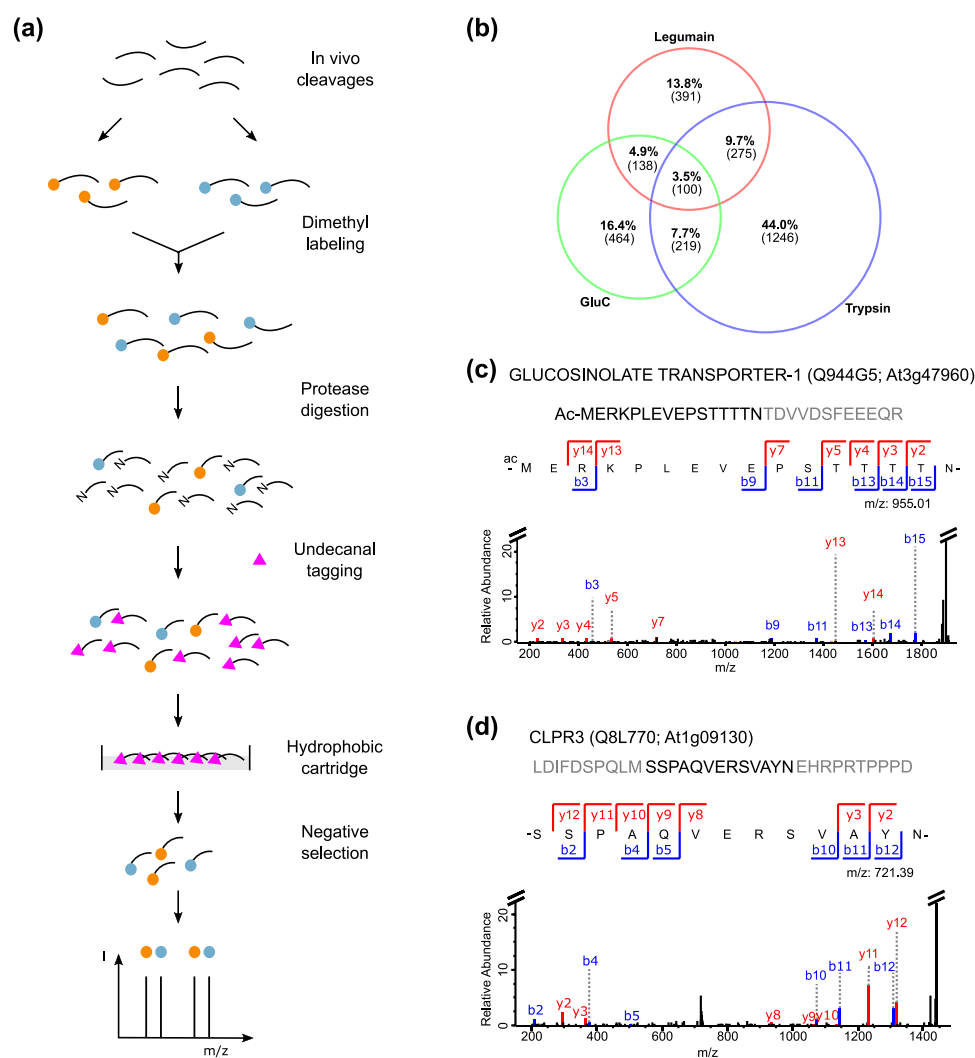


Figure 4. Complementary N-terminome coverage by parallel digestion with legumain, GluC, and trypsin. (a) Experimental workflow for the enrichment of N-terminal peptides using HUNTER. For detailed description, see the main text. Light blue and orange circles indicate differential stable isotope labeling by reductive dimethylation, and magenta triangles indicate undecanal modification. (b) Overlap in N-termini identification based on the first seven amino acids of each N-terminal peptide identified in the experiments with the three proteases. Peptide MS/MS fragmentation spectra of (c) the acetylated mature N-terminus of glucosinolate transporter-1 and (d) a proteolysis-derived dimethylated N-terminus in the CLPR3 subunit of the ATP-dependent Clp protease. Both termini were identified in legumain digests, with sequence context surrounding the identified peptide indicated in gray. UniProt accession code and gene accession numbers are indicated.

is more active at a lower pH²⁴ and that cleavage after Asn is favored at a higher pH.²² To test if this is also the case with the digest conditions applied here, we digested whole-leaf *A. thaliana* proteome at varying pHs between 5.0 and 6.5 for shorter (2 h) and longer (24 h) incubation times (SI Figure S6a). We observed the highest number of peptide identifications at pH 5.5 and pH 6, which may have been caused by the higher propensity for proteome precipitation at a lower pH that we observed in concentrated samples. As expected, legumain showed an increasing preference for Asn with an increasing pH, and this kinetic preference was also reflected at different digestion times. Short proteome digestions (2 h) and/or lower pH (pH 5.0) resulted in a higher proportion of Asn cleavages (SI Figure S6b), whereas longer incubation (24 h) and/or higher pH yielded more complete cleavage after Asp (SI Figure S6b). On the basis of this observation, we tested whether acidification of the MEF proteome digest after an initial incubation at pH 6 would result in more complete cleavage at Asp residues. Indeed, this two-

step incubation maintained efficient cleavage at Asn residues while decreasing the number of peptides containing missed Asp cleavage sites (SI Figure S5).

Denaturants are commonly used for proteome preparations but are problematic during digestion. We tested the tolerance of legumain to urea and guanidinium hydrochloride but observed dramatically decreased digestion efficiency (SI Figure S7a), reflected in decreased peptide identifications (SI Figure S7b) with an increased frequency of missed cleavage (SI Figure S7c). In contrast, legumain tolerated the organic solvent acetonitrile quite well with little decrease in efficiency up to 10% acetonitrile concentration (SI Figure S7).

We also assessed the amount of legumain necessary to achieve optimal digest by varying the protease to proteome ratio. Digestion appeared equally efficient in several dilutions down to a legumain to proteome ratio of 1:100, as judged by the number of identified peptides from an equal starting material (SI Figure S8a). Another important enzyme property for routine use is the shelf-life time, where our recombinant

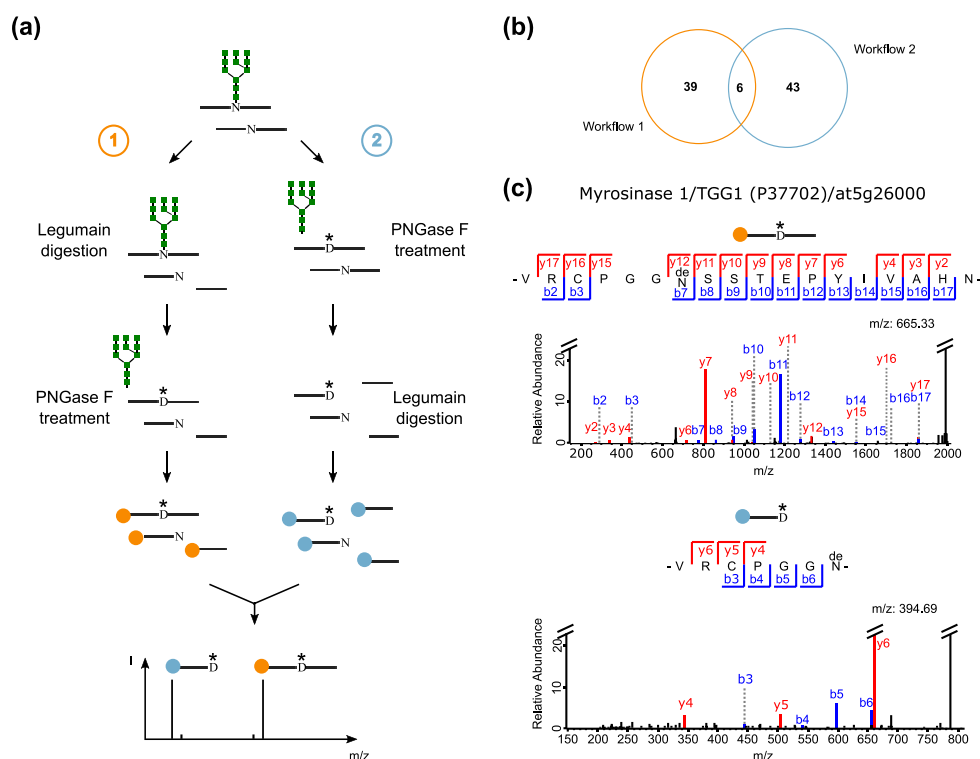


Figure 5. Identification of N-glycosylation sites by sequential processing with legumain and PNGase F. (a) Scheme of the experimental workflow. For details, see the main text. Light blue and orange circles indicate differential stable isotope labeling by dimethylation. Asterisks indicate deamidated asparagine residue arising from PNGase F treatment. (b) Overlap of N-glycosylation identified with internal deamidated Asn in workflow 1 and with C-terminal deamidated Asn in workflow 2. (c) MS/MS fragmentation spectra of an N-glycosylation site in MYROSINASE 1 identified in both workflows. UniProt and *A. thaliana* gene accession codes are indicated.

legumain preparations withstood 10 freeze/thaw cycles without loss of peptidase activity (SI Figure S8b).

Legumain is Highly Complementary for Protein N-Terminal Profiling. The complementarity of different digestion enzymes is particularly helpful for the identification of specific post-translational modification sites such as phosphorylations^{14,16} and protein termini,^{10,20} as these may reside in sequences that are not accessible by trypsin. To demonstrate the value of legumain for this purpose, we profiled N-termini in the *A. thaliana* leaf proteome with our recently established HUNTER protocol (Figure 4a).³⁶ In three replicates per enzyme, two aliquots of *A. thaliana* leaf proteome were differentially dimethyl labeled to block all unmodified primary amines. Thus, all protein N-termini are modified, either by endogenous modifications such as acetylation or by *in vitro* dimethylation. Differentially labeled duplicates are unified and digested in parallel with legumain, GluC, or trypsin. This digestion generates new N-terminal primary amines in all internal and C-terminal peptides, which are then undecanal labeled while the blocked N-terminal peptides remain inert. Undecanal tagging increases the hydrophobicity of the digest-generated peptides, which enables their selective retention on a C18 cartridge, while the dimethyl labeled (or otherwise modified) protein N-terminal peptides are highly enriched in the flow-through for selective analysis (Figure 4a). With this negative selection, we identified a total of 4773 N-terminal peptides (SI Table S3), with 1167, 1209, and 2342 N-terminal peptides identified in legumain, GluC, and tryptic digests, respectively. The differential labeling demonstrated equivalent accuracy in quantification for all three enzymes (SI Figure S9). For comparison of the overlap

of identified protein N-termini, we extracted the first seven residues of each N-terminal peptide (Figure 4b). Only a minority of 100 protein N-termini were identified by all three proteases, and an additional 632 were identified by two proteases, with a majority of 2101 N-termini identified only in digests of a single enzyme (Figure 4b). For example, the acetylated, native N-terminus of the glucosinolate transporter-1 NPF2.10 was only identified in legumain digests, whereas multiple Glu in the N-terminal peptide excluded identification in GluC digests, while the tryptic digest would deliver a very long peptide with unfavorably high content in acidic amino acids (Figure 4c). Similarly, legumain digests uniquely identified an endoproteolytic processing site in CLPR3 (Figure 4d).

Legumain as a Tool for N-Glycosylation Site Mapping. N-Glycosylation is an important and frequent modification of secreted proteins.^{23,37} The removal of the glycan by PNGase F results in deamidation of the Asn to Asp and facilitates mass spectrometry-based identification of occupied N-glycosylation sites.³⁸ We speculated that N-glycosylation would prevent legumain from hydrolyzing adjacent peptide bonds, on the basis of the crystal structure of human legumain that revealed that the zwitterionic character of its S1 subsite provides an ideal binding site for Asn, but no space to accommodate a glycosylated Asn residue.³⁹ In contrast, Asp residues resulting from deglycosylation by a PNGase F treatment would be cleaved. Thus, a sequential treatment with legumain and PNGase F should result in longer peptides containing a missed deamidated Asn (Figure 5a, workflow 1), whereas a PNGase F treatment before legumain digest should result in shorter peptides ending with a

deamidated Asn (Figure 5a, workflow 2). In proof of concept, we isolated *A. thaliana* apoplastic fluid proteome enriched in secreted N-glycosylated proteins and sequentially treated two aliquots with legumain and PNGase F and vice versa in two parallel reactions (Figure 5a). Treated peptides were differentially dimethyl labeled with heavy and light formaldehyde and combined before nano-LC–MS/MS analysis. Indeed, we found several peptides that fulfilled the expectations (Figure 5b, SI Table S4). Peptides from 45 proteins contained a deamidated Asn as missed cleavage in workflow 1, whereas peptides from 49 proteins ended with deamidated Asn. For 6 proteins, including myrosinase 1 (TGG1), an important glycoprotein involved in plant defense,⁴⁰ we observed peptides matching to the same N-glycosylation sites in both workflows, providing intrinsic orthogonal validation (Figure 5c). Notably, this glycosylation site has also been reported previously.⁴¹

DISCUSSION

It is well established that the use of complementary proteases with different specificity in bottom-up proteomic workflows can improve proteome coverage and provide access to sequences that are missed in tryptic digests.^{3,11} This not only allows identification of “missing proteins” that have not been identified by mass spectrometry before,⁴² one of the central goals of the Human Proteome Project,⁴³ but also is important for comprehensive mapping of post-translational modification sites including phosphorylations^{14,16} and global identification of protein termini.^{10,20} Here we characterize human legumain as a new digestion protease in the proteomic toolbox. Legumain exhibited strict sequence specificity for cleavage after Asn and Asp and a high cleavage efficiency that makes it a highly suitable alternative proteolytic enzyme for proteomics.

We have established conditions for reliable in-solution proteome digestion with legumain and show that the alternative cleavage site at Asn yields an entirely different set of peptides than trypsin does, with only minimal overlap in the number of identified peptides delimited by Asp on both sides, in comparison to those with GluC digests. In agreement with the kinetic cleavage preferences determined with peptide substrates,^{22,24} Vidmar et al. reported only minimal cleavage at Asp residues at pH 6 during in-gel digestion of a denatured proteome.²² In contrast, we have observed a much higher cleavage efficiency at Asp residues at pH 6 in our data set, which likely arises from the different digest conditions (in-gel digestion for 2 h with citrate buffer²² compared to in-solution digestion for 16 h in a MES buffer). We noted that the data set of Vidmar et al.²² contains a higher proportion of missed cleavages at Asn and Asp residues than our data set (SI Figure S10), suggesting that the prolonged reaction under more favorable in-solution conditions enables legumain to have a more complete cleavage at Asp residues even at pH 6.0. Notably, a similar effect was observed for Ulilysin/LysargiNase, which has a strong preference for Arg when tested with peptide substrates⁴⁴ but results in a near-complete digestion at Lys residues under proteome digest conditions.¹²

Digestion with legumain consistently identified more peptides than digestion with GluC, but trypsin was far superior. This has been reported for various other digestion proteases, particularly those that do not select for cleavage at basic residues.^{4,11} One explanation is that digestion with enzymes such as legumain and GluC generates peptides with internal basic residues. This can give rise to internal fragment ions during collision-induced dissociation (CID) and result in

unassignable, complex spectra.⁴ In contrast, fragmentation by electron transfer dissociation (ETD) is not affected by the position of the basic residues and has been reported to improve peptide identifications after digestion with proteases that generate long peptides or peptides with internal basic residues.⁴

Parallel digests with all three enzymes increased proteome and protein sequence coverage and were particularly beneficial for protein N-termini identification, where a single digest often generated N-terminal peptides that are too short, too long, or otherwise unfavorable for identification.⁹ By extension, similar benefits may be expected for other post-translational modifications. Furthermore, using a sequential incubation with legumain and PNGase F, we have demonstrated that legumain cannot cleave after glycosylated Asn residues, in contrast to deamidated deglycosylated Asn after PNGase F treatment. On a larger scale, evidence for N-glycosylation can be obtained by PNGase F treatment in ¹⁸O-water, which results in deamidation of Asn to partially ¹⁸O-labeled Asp.³⁸ However, this partial labeling makes relative quantification across samples challenging, while omission of the ¹⁸O-labeling decreases confidence of the site identification as deamidation can also occur spontaneously. On the basis of our proof-of-concept experiment with *A. thaliana* apoplast proteome, we propose tandem sequential PNGase F/legumain treatment as an alternative strategy for experimental validation of N-glycosylation sites.

There are many further potential applications for legumain in peptide-centric proteome workflows. We have previously used legumain to generate high-quality *E. coli* proteome-derived peptide libraries, which enabled detailed cleavage specificity profiling of the vitamin K-dependent coagulation protease sirtlin that would not be possible in trypsin-generated libraries.⁴⁵ Legumain maintains activity at a low pH, down to pH 4.0, and is active in nonreducing conditions;²¹ therefore it is also suitable for protein disulfide bond determination at the low-pH environment required to prevent disulfide reshuffling.⁴⁶ Currently pepsin is used for these experiments due to its high activity under acidic conditions. However, pepsin generates a large number of overlapping peptides due to its broad specificity with a nonexclusive preference for cleavage after Tyr, Phe, Trp, and Leu that complicate the spectra assignment, whereas legumain's high cleavage specificity would alleviate this problem. Taken together, we propose that recombinant human legumain is an attractive protease to complement trypsin in bottom-up mass spectrometry-based proteomics.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.9b03604>.

Figures of substrate cleavage specificity, biophysical properties of peptides, comparisons of protein sequence coverage, overlap of protein group identifications, potential cleavage sites, optimization of legumain digestion conditions, cleavage specificity analysis, legumain digestion efficiency, graphs of numbers of peptides and missed cleavage sites, proteolytic activity of legumain, quantification of protein N-terminal peptides, and potential legumain cleavage sites present in peptides (PDF)

Tables of lists of proteins identified after digestion, N-termini identified, and identification of N-glycosylation sites (XLSX)

AUTHOR INFORMATION

Corresponding Author

Pitter F. Huesgen — Central Institute for Engineering, Electronics and Analytics, ZEA-3, Forschungszentrum Jülich, 52428 Jülich, Germany; Cologne Excellence Cluster on Cellular Stress Responses in Aging Associated Diseases, Medical Faculty and University Hospital and Institute for Biochemistry, Faculty of Mathematics and Natural Sciences, University of Cologne, 50931 Cologne, Germany; orcid.org/0000-0002-0335-2242; Email: p.huesgen@fz-juelich.de

Authors

Wai Tuck Soh — Department of Biosciences, University of Salzburg, 5020 Salzburg, Austria; orcid.org/0000-0003-0082-7983

Fatih Demir — Central Institute for Engineering, Electronics and Analytics, ZEA-3, Forschungszentrum Jülich, 52428 Jülich, Germany; orcid.org/0000-0002-5744-0205

Elfriede Dall — Department of Biosciences, University of Salzburg, 5020 Salzburg, Austria

Andreas Perrar — Central Institute for Engineering, Electronics and Analytics, ZEA-3, Forschungszentrum Jülich, 52428 Jülich, Germany

Sven O. Dahms — Department of Biosciences, University of Salzburg, 5020 Salzburg, Austria

Maithreya Kuppasamy — Central Institute for Engineering, Electronics and Analytics, ZEA-3, Forschungszentrum Jülich, 52428 Jülich, Germany; orcid.org/0000-0001-6866-0417

Hans Brandstetter — Department of Biosciences, University of Salzburg, 5020 Salzburg, Austria; orcid.org/0000-0002-6089-3045

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.analchem.9b03604>

Author Contributions

#W.T.S. and F.D. contributed equally.

Author Contributions

F.D. and P.F.H. conceived the project. W.T.S., F.D., E.D., S.O.D., H.B., and P.F.H. designed the experiments and analyzed the data. E.D. and S.O.D. provided the recombinant human legumain, and W.T.S., F.D., M.K., and A.P. performed the experiments. The manuscript was written by W.T.S., F.D., and P.F.H. and edited by all authors. All authors have approved the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. Ulrich Eckhard for critically reading this manuscript. W.T.S. is a Ph.D. student in the Immunity in Cancer and Allergy Ph.D. program funded by the Austrian Science Fund FWF (project W_01213). This project was in part supported by a starting grant of the European Research Council, with funding from the European Union's Horizon 2020 program (grant 639905, to P.F.H.) and the German Research Foundation DFG (FOR2743, grant HU1756/3-1 to P.F.H.).

REFERENCES

- (1) Aebersold, R.; Mann, M. *Nature* **2016**, 537 (7620), 347–55.
- (2) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., 3rd *Chem. Rev.* **2013**, 113 (4), 2343–94.
- (3) Tsiatsiani, L.; Heck, A. J. *FEBS J.* **2015**, 282 (14), 2612–26.
- (4) Swaney, D. L.; Wenger, C. D.; Coon, J. J. *Proteome Res.* **2010**, 9 (3), 1323–9.
- (5) Perrar, A.; Dissmeyer, N.; Huesgen, P. F. *J. Exp. Bot.* **2019**, 70 (7), 2021–38.
- (6) Lange, P. F.; Overall, C. M. *Curr. Opin. Chem. Biol.* **2013**, 17 (1), 73–82.
- (7) Turk, B.; Turk, D.; Turk, V. *EMBO J.* **2012**, 31 (7), 1630–43.
- (8) Klein, T.; Eckhard, U.; Dufour, A.; Solis, N.; Overall, C. M. *Chem. Rev.* **2018**, 118 (3), 1137–1168.
- (9) Niedermaier, S.; Huesgen, P. F. *Biochim. Biophys. Acta, Proteins Proteomics* **2019**, 1867 (12), 140138.
- (10) Vogtle, F. N.; Wortelkamp, S.; Zahedi, R. P.; Becker, D.; Leidhold, C.; Gevaert, K.; Kellermann, J.; Voos, W.; Sickmann, A.; Pfanner, N.; Meisinger, C. *Cell* **2009**, 139 (2), 428–39.
- (11) Giansanti, P.; Tsiatsiani, L.; Low, T. Y.; Heck, A. J. *Nat. Protoc.* **2016**, 11 (5), 993–1006.
- (12) Huesgen, P. F.; Lange, P. F.; Rogers, L. D.; Solis, N.; Eckhard, U.; Kleifeld, O.; Goulas, T.; Gomis-Ruth, F. X.; Overall, C. M. *Nat. Methods* **2015**, 12 (1), 55–58.
- (13) Schröder, C. U.; Lee, L.; Rey, M.; Sarpe, V.; Man, P.; Sharma, S.; Zabrouskov, V.; Larsen, B.; Schriemer, D. C. *Mol. Cell. Proteomics* **2017**, 16, 1162–1171.
- (14) Schlosser, A.; Vanselow, J. T.; Kramer, A. *Anal. Chem.* **2005**, 77, 5243–5250.
- (15) Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Yates, J. R., 3rd *Nat. Biotechnol.* **2003**, 21 (5), 532–8.
- (16) Gonczarowska-Jorge, H.; Loroch, S.; Dell'Aica, M.; Sickmann, A.; Roos, A.; Zahedi, R. P. *Anal. Chem.* **2017**, 89 (24), 13137–13145.
- (17) Meyer, J. G.; Kim, S.; Maltby, D. A.; Ghassemian, M.; Bandeira, N.; Komives, E. A. *Mol. Cell. Proteomics* **2014**, 13 (3), 823–35.
- (18) Wu, C. C.; Yates, J. R. *Nat. Biotechnol.* **2003**, 21, 262–267.
- (19) Giansanti, P.; Aye, T. T.; van den Toorn, H.; Peng, M.; van Breukelen, B.; Heck, A. J. *Cell Rep.* **2015**, 11 (11), 1834–43.
- (20) Lange, P. F.; Huesgen, P. F.; Nguyen, K.; Overall, C. M. *J. Proteome Res.* **2014**, 13 (4), 2028–44.
- (21) Dall, E.; Brandstetter, H. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, 110 (27), 10940–5.
- (22) Vidmar, R.; Vizovisek, M.; Turk, D.; Turk, B.; Fonovic, M. *EMBO J.* **2017**, 36 (16), 2455–2465.
- (23) Hebert, D. N.; Lamriben, L.; Powers, E. T.; Kelly, J. W. *Nat. Chem. Biol.* **2014**, 10 (11), 902–10.
- (24) Dall, E.; Brandstetter, H. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **2012**, 68 (1), 24–31.
- (25) Wessel, D.; Flugge, U. I. *Anal. Biochem.* **1984**, 138 (1), 141–3.
- (26) Hughes, C. S.; Moggridge, S.; Muller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsvelde, J. *Nat. Protoc.* **2019**, 14 (1), 68–85.
- (27) Rappsilber, J.; Mann, M.; Ishihama, Y. *Nat. Protoc.* **2007**, 2 (8), 1896–906.
- (28) Rinschen, M. M.; Hoppe, A. K.; Grahmmer, F.; Kann, M.; Volker, L. A.; Schurek, E. M.; Binz, J.; Hohne, M.; Demir, F.; Malisic, M.; Huber, T. B.; Kurschat, C.; Kizhakkepathu, J. N.; Schermer, B.; Huesgen, P. F.; Benzing, T. *J. Am. Soc. Nephrol.* **2017**, 28 (10), 2867–2878.
- (29) Tyanova, S.; Temu, T.; Cox, J. *Nat. Protoc.* **2016**, 11 (12), 2301–2319.
- (30) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. *Nat. Methods* **2016**, 13 (9), 731–40.
- (31) Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. *Nat. Protoc.* **2015**, 10 (6), 845–58.
- (32) Weng, S. S. H.; Demir, F.; Ergin, E. K.; Dirnberger, S.; Uozio, A.; Tüscher, D.; Nierves, L.; Tsui, J.; Huesgen, P. F.; Lange, P. F. *Mol. Cell. Proteomics* **2019**, 18, 2335.
- (33) Joosten, M. H. *Methods Mol. Biol.* **2012**, 835, 603–10.

- (34) Deutsch, E. W.; Orchard, S.; Binz, P. A.; Bittremieux, W.; Eisenacher, M.; Hermjakob, H.; Kawano, S.; Lam, H.; Mayer, G.; Menschaert, G.; Perez-Riverol, Y.; Salek, R. M.; Tabb, D. L.; Tenzer, S.; Vizcaino, J. A.; Walzer, M.; Jones, A. R. *J. Proteome Res.* **2017**, *16* (12), 4288–4298.
- (35) Vizcaino, J. A.; Csordas, A.; del-Toro, N.; Dienes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; Xu, Q. W.; Wang, R.; Hermjakob, H. *Nucleic Acids Res.* **2016**, *44* (D1), D447–56.
- (36) Weng, S. S. H.; Demir, F.; Ergin, E. K.; Dirnberger, S.; Uozzie, A.; Tuscher, D.; Nierves, L.; Tsui, J.; Huesgen, P. F.; Lange, P. F. *Mol. Cell. Proteomics* **2019**, *18*, 2335.
- (37) Clerc, F.; Reiding, K. R.; Jansen, B. C.; Kammeijer, G. S. M.; Bondt, A.; Wuhler, M. *Glycoconjugate J.* **2016**, *33*, 309–343.
- (38) Kuster, B.; Mann, M. *Anal. Chem.* **1999**, *71*, 1431–1440.
- (39) Dall, E.; Brandstetter, H. Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation. Proceedings of the National Academy of Sciences of the United States of America, 2013.
- (40) Barth, C.; Jander, G. *Plant J.* **2006**, *46* (4), 549–62.
- (41) Liebminger, E.; Grass, J.; Jez, J.; Neumann, L.; Altmann, F.; Strasser, R. *Phytochemistry* **2012**, *84*, 24–30.
- (42) Wang, Y.; Chen, Y.; Zhang, Y.; Wei, W.; Li, Y.; Zhang, T.; He, F.; Gao, Y.; Xu, P. *J. Proteome Res.* **2017**, *16* (12), 4352–4363.
- (43) Paik, Y. K.; Overall, C. M.; Deutsch, E. W.; Van Eyk, J. E.; Omenn, G. S. *J. Proteome Res.* **2017**, *16* (12), 4253–4258.
- (44) Tallant, C.; Garcia-Castellanos, R.; Marrero, A.; Canals, F.; Yang, Y.; Reymond, J. L.; Sola, M.; Baumann, U.; Gomis-Ruth, F. X. *Biol. Chem.* **2007**, *388* (11), 1243–53.
- (45) Dahms, S. O.; Demir, F.; Huesgen, P. F.; Thorn, K.; Brandstetter, H. *J. Thromb. Haemostasis* **2019**, *17* (3), 470–481.
- (46) Gorman, J. J.; Wallis, T. P.; Pitt, J. J. *Mass Spectrom. Rev.* **2002**, *21* (3), 183–216.