Understanding the Role of Proteolytic Digestion on Discovery and Targeted Proteomic Measurements Using Liquid Chromatography Tandem Mass Spectrometry and Design of Experiments

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · OCTOBER 2013

Impact Factor: 4.25 · DOI: 10.1021/pr4008442 · Source: PubMed

6 AUTHORS, INCLUDING:

Jack Peng-Yu Wang
North Carolina State University
15 PUBLICATIONS **121** CITATIONS
SEE PROFILE

Quanzi Li
Chinese Academy of Forestry
36 PUBLICATIONS **650** CITATIONS
SEE PROFILE

Ronald Sederoff
North Carolina State University
157 PUBLICATIONS **9,958** CITATIONS
SEE PROFILE

David Muddiman
North Carolina State University
256 PUBLICATIONS **6,923** CITATIONS
SEE PROFILE

Journal of proteome research

Article

pubs.acs.org/jpr

Understanding the Role of Proteolytic Digestion on Discovery and Targeted Proteomic Measurements Using Liquid Chromatography Tandem Mass Spectrometry and Design of Experiments

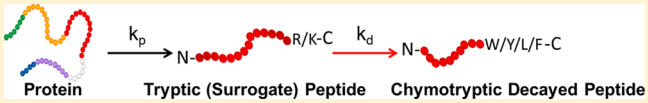
Philip L. Loziuk,[†] Jack Wang,[‡] Quanzi Li,[‡] Ronald R. Sederoff,[‡] Vincent L. Chiang,[‡] and David C. Muddiman*,[†]

[†]W.M. Keck Fourier Transform Mass Spectrometry Laboratory, Department of Chemistry, North Carolina State University, Raleigh, North Carolina, 27695, United States

[‡]Forest Biotechnology Group, Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, North Carolina, 27695, United States

Supporting Information

ABSTRACT: Workflows in bottom-up proteomics have traditionally implemented the use of proteolysis during sample preparation; enzymatic digestion is most commonly performed using trypsin. This results in the hydrolysis of peptide bonds forming tryptic peptides, which can then be subjected to LC−MS/MS analysis. While the structure, specificity, and kinetics of trypsin are well characterized, a lack of consensus and understanding has remained regarding fundamental parameters critical to obtaining optimal data from a proteomics experiment. These include the type of trypsin used, pH during digestion, incubation temperature as well as enzyme-to-substrate ratio. Through the use of design of experiments (DOE), we optimized these parameters, resulting in deeper proteome coverage and a greater dynamic range of measurement. The knowledge gained from optimization of a discovery-based proteomics experiment was applied to targeted LC−MS/MS experiments using protein cleavage-isotope dilution mass spectrometry for absolute quantification. We demonstrated the importance of these digest parameters with respect to our limit of detection as well as our ability to acquire more accurate quantitative measurements. Additionally, we were able to quantitatively account for peptide decay observed in previous studies, caused by nonspecific activity of trypsin. The tryptic digest optimization described here has eliminated this previously observed peptide decay as well as provided a greater understanding and standardization for a common but critical sample treatment used across the field of proteomics.



KEYWORDS: *design of experiments, proteolytic digestion, LC−MS/MS, Populus trichocarpa*

INTRODUCTION

Investigation of sample preparation methods for workflows using LC−MS/MS has continued to be a central focus in the field of proteomics. Advances in sample preparation have made bottom-up proteomics less cumbersome with the introduction of methods such as filter-aided sample preparation (FASP).[1,2] FASP implements the use of a molecular size cutoff filter, allowing the experimenter to remove impurities and unwanted detergents while retaining the protein sample which can then be digested in solution. Subsequently, the resulting peptides can be eluted and collected for analysis. Typical sample treatment for bottom-up proteomics involves the use of enzymatic digestion to generate a mixture of peptides from a protein sample. The peptides' sequences can then be obtained by bioinformatics software using their intact mass and subsequent product-ion spectra; experimental spectra are matched to theoretical spectra obtained from an in silico digestion of a target (or target-reverse) database.[3−6] The most commonly used enzyme for protein digestion in bottom-up proteomics is trypsin. Because of the frequency of arginine (R) and lysine (K) residues within the proteome, trypsin generates peptides with a mass range ideal for obtaining sequences.[7] Because it cleaves at basic residues R and K, the use of trypsin favors the +2 charge state of peptides when conducting experiments under acidic conditions using positive electrospray ionization mass spectrometry (ESI), generating fragments that are easily detectable.

The structure, specificity, and kinetics of trypsin have been widely studied in great detail. Many sources and types of trypsin are currently manufactured and used by researchers in proteomics.[8−11] While nonspecific cleavage activity of unmodified trypsin due to autolysis during enzymatic digestion has previously been reported,[12,13] the cleavage specificity of various trypsin types and the extent to which it influences a proteomic data set have been understudied. With the introduction of modified trypsin through dimethylation of lysine residues, kinetic studies have suggested that the enzyme is capable of greater cleavage specificity of tryptic sites as well as having the ability to maintain optimal activity under more basic conditions and at higher temperatures.[14] Other forms of modified trypsin include trypsin with acetylated lysine residues and immobilized trypsin, which is chemically linked to a

Received: August 18, 2013
Published: October 21, 2013

hydrophilic polymer. Additionally, the source of trypsin, typically, either bovine or porcine, varies as well. Other digestion parameters such as pH and the enzyme-to-substrate ratio used in processing protein samples vary widely depending on the lab. Enzyme-to-protein ratios typically used in discovery-based and absolute quantitative proteomics experiments range from 1:100 to as high as 1:2.5.[15−19] While previous research has shown there to be no significant difference between a 1:20 or a 1:100 enzyme-to-substrate ratio on global proteomic data, investigation of higher enzyme-to-substrate ratios has not been reported.[20] Examination of the effect of the enzyme-to-substrate ratio on measurements made for absolute quantification has been lacking.

While a digestion pH of 8 is most typically reported, manufacturers recommend a pH of anywhere from 7 to 9. This pH range is substantial in a biological context where small fluctuations in pH can drastically alter the activity of enzymes as well as the protonation state of all biomolecules in solution. Previous research has suggested that unmodified bovine trypsin functions optimally at a pH of 7, while more recent research suggests that modified trypsin has an optimal pH closer to 8.[14,21] A comprehensive understanding of the effect of tryptic digestion on bottom-up proteomics remains elusive.[20] With conflicting literature on optimal digest conditions for global and targeted proteomics, this study sought to reduce the ambiguity surrounding this critical aspect of sample preparation for bottom-up proteomic workflows using LC−MS/MS.

Protein cleavage-isotope dilution mass spectrometry (PC-IDMS) is widely used for absolute quantification of peptides and proteins.[22−25] The development of such an assay typically begins with a global proteomic approach in which a digested sample is subjected to LC−MS/MS, and the most abundant peptides are sequenced. Using the list of peptides identified for each target protein, a list of potential surrogate peptides can be generated. The list of potential surrogates can be further filtered based on factors that predict the most appropriate surrogate for development of a PC-IDMS assay.[11] These factors range from sequence length (7−25 amino acids), avoiding or selecting sites of nonsynonymous SNPs, potential sites of oxidation (M,C) as well as sites containing post-translational modifications. Most importantly the peptide must be unique to the protein, and those detected with higher ion abundance are favored.

Once surrogate peptides have been selected, a PC-IDMS assay can be developed using selected reaction monitoring on a triple quadrupole instrument. This method involves two mass filters or quadrupoles. The first mass filter Q1 selects only a single intact precursor ion which is then fragmented in q2, the collision cell. A selected fragment ion can then be allowed to pass through Q3 to the detector. A list of up to several hundred transitions can be monitored in any given experiment.[20,26] This allows for the development of a targeted multiplexed assay to specifically quantify target peptides/proteins in complex biological samples.

We have developed a targeted assay for quantification of 25 proteins involved in monolignol biosynthesis in the model plant *Populus trichocarpa*. In tracking surrogate peptides of these proteins throughout a digestion, it was observed that a peptide underwent a process of decay after it was produced.[11] Different peptides were produced and degraded at different rates. This observation emphasized the importance of when stable isotope-labeled (SIL) standards used for absolute quantification should be added to a sample. Addition of the SIL-standard concurrent with digestion was the optimal introduction time for quantitatively accurate measurements. In this way, any sample loss during digestion can be corrected using the internal standard reference under the same conditions. While quantitative accuracy is largely maintained, native peptides and SIL standards do not experience the exact same production/decay kinetics because the SIL peptides are not produced via proteolysis. This results in an overestimation of the quantity, and in some cases, this overestimation can be quite significant.[11] Additionally, the loss of low abundant peptides during digestion can affect the dynamic range of measurement by hindering the ability to accurately quantify or even detect a peptide. Consequentially, minimizing peptide decay creates another potential factor to consider when selecting surrogate peptides in the development of targeted assays for absolute quantification.

This study addresses this influence of proteolysis in both global and targeted proteomic workflows. We have optimized digest parameters for global proteomic responses as well as minimized this observed peptide decay, thereby providing a lower limit of detection and increasing the ability to obtain more accurate measurements for absolute quantification. Through the optimization of trypsin proteolysis, we emphasize the importance of critical aspects of digestion, which must be considered to obtain accurate, reliable, and optimum data from both discovery proteomics and absolute quantification experiments.

EXPERIMENTAL SECTION

Materials

Unless otherwise stated, all reagents were purchased from Sigma-Aldrich (St. Louis, MO). All solvents were HPLC-grade from Honeywell Burdick & Jackson (Muskegon, MI).

Stem Differentiating Xylem Tissue (SDX) Protein and Filter-Aided Sample Preparation

Stem differentiating xylem was collected from three six-month-old *Populus trichocarpa* (genotype Nisqually-1), grown in a greenhouse as previously described.[27] Protein extracts were prepared from each tree, by grinding 3 g of SDX in liquid nitrogen and then homogenizing the cells (2 min, on ice) in 15 mL of extraction buffer containing 50 mM Bis-Tris (pH 8.0), 20 mM sodium ascorbate, 0.4 M sucrose, 100 mM NaCl, 5 mM DTT, and 10% (w/w) polyvinylpolypyrrolidone. After the cell debris was removed by centrifugation (3000*g*, 4 °C, 15 min − twice), the three protein extracts were pooled and the protein concentration was measured using a Coomassie Plus Bradford assay (Thermo Scientific, Rockford, IL) prior to storage at −80 °C. Filter-aided sample preparation was performed as described previously[20] for evaluating digest pH, digest temperature, enzyme/substrate ratio, and source of trypsin. All solutions were compromised of 50 mM Tris buffer. Protein extracts were incubated for 30 min at 56 °C in a 2-fold dilution of 8 M urea and 100 mM dithiothreitol in order to denature and reduce. The sample was then alkylated using 200 mM iodoacetamide for 1 h at 37 °C. Samples were then washed 3× for 15 min at 14000*g* with digestion buffer containing 2 M urea and 10 mM $CaCl_2$. It should be noted that the Tris digestion buffer was measured prior to incubating at 37 or 48 °C. This is the typical approach to reporting digestion pH in the field of proteomics. Experimentally, it was verified that the Tris digestion buffer will undergo a pH shift of about −0.3 pH units at 37 °C and about −0.5 pH units at 48 °C. Thus, the actual digestion pH becomes more acidic than it was prior to incubation.

dx.doi.org/10.1021/pr4008442 | J. Proteome Res. 2013, 12, 5820−5829

Full Factorial Design of Experiments

A full factorial design of experiments was employed for digest optimization using a DOE custom design within JMP Pro v10.0.0 (SAS Institute Inc., Cary, NC). Four parameters were specifically investigated: pH (7 or 8), incubation temperature (37 or 48 °C), enzyme/substrate ratio (1:50 or 1:5), as well as the type of trypsin (unmodified bovine sequencing-grade trypsin or modified dimethylated porcine trypsin both from Sigma-Aldrich). Within the full factorial design of experiments, 16 different digest conditions were tested in duplicate. Specific experimental conditions tested in each sample can be found in Supporting Information (Table S1).

Following the full factorial study, the enzyme/substrate ratio using modified porcine trypsin was more thoroughly examined in order to determine the amount of enzyme required for complete digestion. In these latter experiments, the optimized digestion conditions determined from the design of experiments (modified porcine trypsin, pH 7 and 37 °C) were used for all subsequent digests, varying only the ratio of enzyme/substrate. Digest volume was kept constant, using 50 μL of trypsin solution which was added to a filter containing 100 μg of total protein, while 2 nmol of SIL cocktail was added postdigestion. In experiments assessing quantitative accuracy of digestion methods, 5 μL (2 nm) of SIL peptides was added concurrently with 45 μL of trypsin solution, keeping the total digest volume constant. When different enzyme substrate ratios were tested for "completeness of digestion" the following enzyme/substrate ratios were used: 1:50, 1:20, 1:10, and 1:5. These digestion reactions were quenched using 500 μL of 0.1% formic acid after 16 h. The resulting data were fit by nonlinear regression as previously described for unmodified bovine trypsin to determine the ratio and concentration of enzyme required for complete digestion.[11]

Time-course experiments compared differences in quantitative discrepancies in peptide decay and chymotryptic peptide production using previous digest conditions for absolute quantification (unmodified bovine trypsin, pH 8, 37 °C, and 1:5 enzyme/substrate ratio) and newly optimized conditions for absolute quantification (modified porcine trypsin, pH 7, 37 °C, and 1:5 enzyme/substrate ratio). When different digestion times were tested, all reactions contained 20 μg of trypsin in a total of 50 μL (400 μg/mL final concentration). Duplicate samples were prepared from 3.2 mg of crude protein from SDX tissue, and 100 μg of denatured, reduced and alkylated protein was added to a 10 kDa FASP filter. Digestion was quenched at 0.5, 1, 2, 4, 6, 8, 12, and 16 h. Following digestion, 2 nmol of SIL standards were added, and the sample was dried down and stored at −20 °C prior to analysis. Quantification of both tryptic peptides and chymotryptic peptides was carried out for each time point.

The final, optimized digestion protocol for protein quantification was then implemented to assess the quantitative accuracy of previous method compared to the new method. Two samples were digested for 8 h with previously used digest conditions for absolute quantification (unmodified bovine trypsin, pH 8, 37 °C, and a 1:5 enzyme/substrate ratio). Two other samples were also digested for 8 h using newly optimized conditions for absolute quantification (modified porcine trypsin, pH 7, 37 °C, and a 1:5 enzyme/substrate ratio). In all cases, FASP samples were dried down under a vacuum and stored at −20 °C until peptides were reconstituted in 0.001% zwittergent 3−16 (Calbiochem, La Jolla, CA) for LC−MS/MS analysis.

Stable Isotope-Labeled Peptide Standards and Transition Characterization

Natural (NAT) and stable isotope-labeled (SIL) peptides were synthesized by the Mayo Clinic Proteomics Research Center (Rochester, MN) with the exception of NGSVVAPPDAMR and LSTSEIASHLPTK NAT and SIL peptides, synthesized by New England Peptide (Gardner, MA). All peptides were used as received with the exception of C3H3.125-134 and CCoAOMT3.217-232. These two cysteine containing peptides were carbamidomethylated as described.[20] SIL peptides stock solutions from the Mayo Clinic were produced by dissolving ∼2 mg in water. Absolute concentrations had been previously confirmed by spectrophotometry using the Scope's method.[20,28] Purity of peptides from New England Peptide was verified by mass spectrometry, and peptides were quantified using amino acid analysis (AAA). The previous selection of surrogate peptides was performed based on shotgun proteomic data obtained from LC−MS/MS analysis of native xylem protein as well as recombinant xylem proteins.[20] Each peptide and transition were then optimized for specific and sensitive detection utilizing collision energy optimization (Table S2, Supporting Information).[20] An SIL cocktail containing isotopically labeled surrogate peptides obtained from the Mayo Clinic were used for all experiments quantifying lignin protein surrogate peptides. This cocktail utilized an equimolar mixture of surrogate peptides which were added to xylem protein samples in the amount of 10 fmol/μg of total protein.[20] Another SIL cocktail was used for quantification of chymotryptic peptide products. This cocktail contained two unique peptides (New England Peptide, Gardner, MA) which were products of a chymotryptic cleavage (WYFL) of the tryptic surrogate peptide for proteins CCoAOMT1 (NGSVVAPPDAMR) and COMT2 (LSTSEIASHLPTK). These two peptides were identified in a preliminary discovery proteomics experiment in which surrogate peptide standards were incubated with unmodified trypsin and were found to be unique in the *P. trichocarpa* database. The peptide NGSVVAPPDAMR was added in 15 fmol/μg of total protein, while LSTSEIASHLPTK was added in 40 fmol/μg of total protein. All SIL peptides were added either concurrently or postdigestion to all samples for quantification such that the SIL signal was within 1 order of magnitude of the NAT signal.

Newly obtained standards from New England Peptide were subjected to direct infusion into a TSQ Vantage triple quadrupole mass spectrometer (Thermo Fisher Scientific, San Jose, CA) at 50 μM to identify the most abundant precursor ion for each peptide. Standards were then fragmented, and the six most abundant product ions were chosen for further characterization. Collision energy optimization experiments were performed as described with the assistance of Skyline (v.1.4.0.4222) in which methods were developed, testing a range of collision energies and choosing the correct energy for maximum signal intensity of each peptide.[20,29,30] Replicate injection LC−MS/MS analysis of standard peptides was performed to characterize the relative abundance (RA) for each particular transition. This allowed us to confirm the purity of each transition in any sample through ion ratio assessment or relative abundance value. The relative abundance value of a particular transition is defined by the ratio of the total area due to all transitions subtracting the area of the transition of interest, finally dividing by the area of the transition of interest.[20] This results in abundant transitions having small RA values, while less abundant transitions have large RA values. A

dx.doi.org/10.1021/pr4008442 | J. Proteome Res. 2013, 12, 5820−5829

significance threshold was then implemented in order to determine whether a transition was contaminated based on deviation from the expected RA value. For highly abundant transitions (RA $\leq$ 1), the tolerance was set at $\pm$15%. For moderate abundance (9 < RA < 1) and low abundance transitions (RA $\geq$ 9), the tolerances were set at $\pm$25% and $\pm$50%, respectively. Finally, transitions were monitored in a digest of crude SDX protein in order to verify the presence and adequacy of the signal/noise ratio (>10) of each transition prior to quantification.

LC–MS/MS Analysis

Samples prepared for the full factorial experiments were analyzed in duplicate using a nanoLC-1D (Eksigent Tecnologies, Dublin, CA, USA) coupled with a hybrid LTQ-Orbitrap XL mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) via a trap and elute configuration in which 5 μL was injected onto a trap column, 200 μm × 0.5 mm in-line with an analytical column, 75 μm × 15 cm (Eksigent Tecnologies, Dublin, CA, USA) both packed with Chrom XP C18-CL (3 μm particle size, 120 Å pore size). A 90-min elution was performed at a flow rate of 350 nL/min utilizing a 5–40% B gradient. Mobile phases A and B were composed of water/acetonitrile/formic acid (98/2/0.2% and 2/98/0.2%, respectively). LTQ-Orbitrap instrument parameters, previously optimized for global proteomics experiments, were implemented for MS analysis.[31]

For targeted experiments, LC-SRM was performed using a nanoLC-2D system equipped with an AS1 autosampler and a cHiPLC-nanoflex system (Eksigent, Dublin, CA) which was coupled to a TSQ Vantage Triple stage quadrupole mass spectrometer (Thermo Scientific, San Jose, CA) using a 10 μm i.d. SilicaTip ESI emitter (New Objective, Woburn, MA). LC-SRM analysis was performed as previously described in the development of a multiplexed assay for quantification of monolignol-pathway enzymes.[20] Each sample was loaded and desalted using a 9 μL metered injection of a 5 μL sample using 100% mobile phase A at 1.5 μL/min. Subsequently, peptides were eluted at 400 nL/min using a 22 min ramp from 5 to 38.5% mobile phase B, after which the column was washed with 95% mobile phase B for 5 min prior to re-equilibrating both columns to initial conditions. The column eluent was ionized using a 1400 V ESI potential and a capillary temperature of 200 °C. Scheduled SRM was performed using the EZ Method with a defined cycle time of 1.5 s, chrom filter of 30 s, Q1 and Q3 peak width (fwhm) of 0.7 Da, a collision pressure of 1.5 mTorr, and a 4.0 min retention time window for each peptide. The entire list of peptide transitions is provided (Table S2, Supporting Information) along with their optimized collision energies. Quantification of surrogate peptides and their decayed products was performed on n = 2 samples for each condition along with replicate injection analysis using two scheduled SRM methods, one for the tryptic surrogates and the other for their decayed product peptides.

Bioinformatics, Global and Targeted Data Analysis

Raw LC–MS/MS data files from SDX protein extracts were converted to ".mgf" files using the Mass Matrix file conversion tool freely available online . Resulting ".mgf" files were searched using Mascot Daemon against a concatenated target-reverse *P. trichocarpa* database JGI *P. trichocarpa* v2.2 (Joint Genome Institute, U.S. Department of Energy) containing 45 134 target sequences including modified sequences recently cloned from *P. trichocarpa*.[27] The database was searched using the following parameters: trypchymo as the enzyme which performs in silico digest of *P. trichocarpa* database at (FWYLRK) tryptic and well-known chymotryptic cleavage site residues, fixed carbamidomethyl modification of cysteine residues, variable oxidation of methionine, variable deamidation of asparagine and glutamine, maximum of nine missed cleavages, 5 ppm precursor tolerance, and 0.6 Da MS/MS tolerance. Finally, the search results (.dat files) obtained from Mascot were imported into ProteoIQ v2.3.08 (Premier Biosoft, Palo Alto, CA) to obtain a final protein identification list at a protein FDR of 1%. Cleavage site information was obtained by exporting peptide sequences and preceding residue information in .csv format from Mascot and selecting only those peptides that were identified by ProteoIQ after filtering at 1% FDR.

Data from targeted LC–MS/MS experiments were imported into Skyline[29] where peak integration was performed after adjusting retention time windows. After consistent integration of all samples was performed, data were exported into Excel. Peak areas were further analyzed using an Excel template in which ion ratio assessments were performed by implementing tolerance thresholds for the relative abundance (RA) value of transitions in order to confirm the presence and purity of individual transitions. In turn, the transitions determined to be pure were then used to quantify peptides as described previously.[20]

RESULTS AND DISCUSSION

Proteolysis and Global Proteomic Measurements

A full factorial design of experiments (DOE) was employed to investigate the influence of particular digestion parameters on the data from a global proteomics experiment utilizing wild-type stem differentiating xylem tissue from the model woody plant *P. trichocarpa*. The selectively manipulated parameters were pH (7 or 8), temperature (37 or 48 °C), enzyme-to-substrate ratio (1:50 or 1:5), and trypsin type (unmodified TPCK treated bovine trypsin or modified dimethylated porcine trypsin). The responses of interest to be maximized were number of protein groups identified, number of unique peptides identified, average protein sequence coverage as well as the ratio of tryptic-to-chymotryptic cleavage sites. Through data analysis in JMP10 pro, contrast values, and p-values were obtained based on each parameter tested. P-values denote statistical significance, whereas contrast values reveal the magnitude with which a particular parameter influences each response.

To maximize the number of protein groups identified, all parameters were significant, while temperature had the greatest influence on the data, with 37 °C being the better condition (Figure 1a). Trypsin type had the second greatest influence on protein identification with modified trypsin being more favorable. Following trypsin type, a digestion pH of 7 and an enzyme-to-substrate ratio of 1:50 were best in maximizing protein identification. In maximizing the number of unique peptides, all the parameters tested also had a significant influence on the data with temperature having the greatest effect, yielding more unique peptides identified at 37 °C versus 48 °C (Figure 1b). The other parameters had a relatively comparable influence on the data, favoring an enzyme-to-substrate ratio of 1:50, a pH of 7, and modified trypsin.

In maximizing protein sequence coverage, the enzyme-to-substrate ratio was the most influential parameter, favoring a ratio of 1:50 (Figure 1c). This result was likely due to

dx.doi.org/10.1021/pr4008442 | J. Proteome Res. 2013, 12, 5820–5829

a)

Parameters Maximizing Protein ID's

Factor	Contrast Value	p-value
Trypsin Type	27.34	<0.0001
Enzyme:Substrate	-14.59	<0.0001
pH	-22.03	<0.0001
Temperature	-113.59	<0.0001

b)

Parameters Maximizing Unique Peptides Sequenced

Factor	Contrast Value	p-value
Trypsin Type	138.53	0.0013
Enzyme:Substrate	-140.59	0.0012
pH	-140.91	0.0012
Temperature	-475.41	<0.0001

c)

Parameters Maximizing Sequence Coverage

Factor	Contrast Value	p-value
Trypsin Type	0.90	0.0007
Enzyme:Substrate	-1.76	<0.0001
pH	-1.03	0.0005
Temperature	-1.00	0.0006

d)

Parameters Maximizing Ratio of Tryptic:Chymotryptic Sites

Factor	Contrast Value	p-value
Trypsin Type	85.87	<0.0001
Enzyme:Substrate	-20.49	0.1183
pH	-18.10	0.1642
Temperature	0.24	0.984

Figure 1. Shows corresponding contrast values resulting from JMP analysis for each of the responses investigated. Contrast values indicate the magnitude with which each factor influences the data in question while the *p*-value determines the statistical significance of the parameter. Contrast values, along with their corresponding *p*-values revealed that for the number of protein groups identified (a) all parameters were observed to be significant, with temperature being the most influential factor followed by trypsin type, pH, and enzyme:substrate ratio. All parameters were shown to be significant in maximizing the number of unique peptides sequenced (b) with temperature being the most influential followed by pH, enzyme/substrate ratio and trypsin type. The conditions maximizing protein identification, unique peptides identified, sequence coverage as well as tryptic/chymotryptic cleavage sites were 37 °C, pH 7, porcine modified trypsin, and a 1:50 ratio of enzyme/substrate. All parameters were also shown to be significant in maximizing the average sequence coverage obtained (c) during an experiment with the enzyme/substrate ratio being the most influential followed by pH, temperature, and trypsin type. Only trypsin type was shown to be significant in maximizing the ratio of tryptic to chymotryptic sites (d).

"overdigestion" of peptides caused by increased tryptic autolysis at higher enzyme-to-substrate ratios and subsequently altering the digestion activity of trypsin. The other parameters were also significant favoring modified trypsin, a pH of 7, and a temperature of 37 °C, all with similar contrast values. Finally, in maximizing the ratio of tryptic to chymotryptic cleavage sites (R|K to F|L|W|Y), the only parameter that significantly influenced the response was trypsin type, with the response favoring modified trypsin (Figure 1d). In comparing digestion conditions previously used by our lab to the newly optimized conditions (Figure 2a–d), we observed a significant increase in the number of protein groups identified, number of unique peptides identified, sequence coverage, and ratio of tryptic-to-chymotryptic cleavage sites observed. While the nonspecific activity of trypsin is not absent when using modified trypsin, it is significantly reduced when compared to unmodified trypsin. These observations also agree with recently published data regarding the cleavage specificity of unmodified and modified trypsin as well as their influence on the number of peptides identified and sequence coverage obtained from global proteome experiments.[32]

The optimized digest conditions appeared to improve all responses. The impact of temperature on the digest was somewhat expected, because 37 °C is the most commonly used incubation temperature for trypsin. However, higher temperatures have been used, particularly when performing shortened digests. The impact of temperature on the data suggests that using higher temperatures to speed up tryptic digestion may be counterproductive. Additionally, while the absolute value of the pH differed depending on the digestion temperature used, it is unlikely the −0.2 pH unit difference observed between 37 and 48 °C could account for the large discrepancies in proteins and peptides identified between samples incubated at different temperatures. It is most likely that the significant decrease in peptide/protein identifications at higher digestion temperature was a result of thermal denaturation of trypsin and rapid loss of tryptic activity during the incubation period.

In the enzyme-to-substrate ratio, a lower enzyme-to-substrate ratio is more effective for global proteomic workflow. Excess trypsin results in greater autolysis, even with modified trypsin. This is likely due to the higher enzyme concentration, resulting in enzyme being more likely to encounter enzyme instead of substrate. The higher amount of enzyme may result in excess digestion and nonspecific cleavage of peptides into fragments too small to be sequenced with good sensitivity by the mass spectrometer. The increase in chymotryptic cleavage (even for the case of modified trypsin) upon the increase in enzyme/substrate ratio supports this hypothesis.

Finally, the favorability of using a digest pH of 7 was interesting, because nearly all reported tryptic digestion is performed at pH of 8. Trypsin may be stable for a longer duration when incubated at pH 7. Most kinetic studies of trypsin have only analyzed the activity during short digestion periods.[14,21,33] These short digestion times could account for the discrepancy. It is also possible that the optimal pH is substrate dependent; however, because peptides have a wide range of pI values it is more likely that pH would more significantly affect the activity of trypsin itself. It is interesting to note that the absolute value of the optimized pH may even be slightly acidic as was noted in the Experimental Section. The pH shift of −0.3 pH units observed at 37 °C and −0.5 pH units at 48 °C with our digestion buffer implies that the absolute value of the pH in the optimized digest buffer is closer to 6.7 rather than 7. This makes the optimized pH finding even more interesting since it has largely been the consensus among the proteomics community that a more basic pH is most suitable for tryptic digestion.

Proteolysis and Targeted Proteomic Measurements

We further investigated the effect of proteolysis on the ability to detect and accurately quantify peptides/proteins in a targeted experiment using protein cleavage isotope-dilution mass

dx.doi.org/10.1021/pr4008442 | J. Proteome Res. 2013, 12, 5820−5829

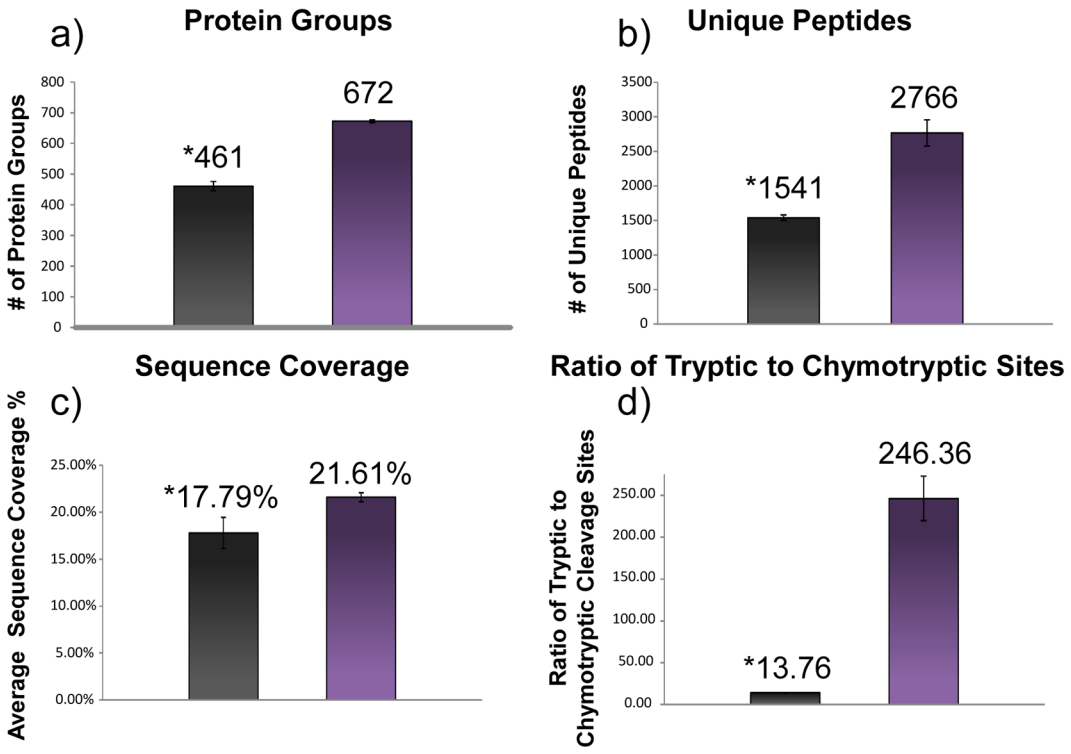


Figure 2. Responses revealed a significantly greater number of unique peptides, proteins identified, greater sequence coverage, and a greater ratio of tryptic-to-chymotryptic cleavage sites using the optimized parameters from the design of experiments (37 °C, pH 7, porcine modified trypsin, and a 1:50 enzyme/substrate ratio) when compared to digestion conditions previously used for global proteomics by our laboratory (37 °C, pH 8, bovine unmodified trypsin, and a 1:50 enzyme/substrate ratio). The responses shown here have greatly affected our ability to identify proteins and discover potential surrogate peptides that can be used for absolute quantification as well as site-specific identification of post-translational modifications.

spectrometry (PC-IDMS). Previously used digestion methods for absolute quantification performed in our laboratory (unmodified bovine trypsin, 37 °C, pH 8, and enzyme:substrate ratio of 1:5) were assessed quantitatively with our newly optimized digestion conditions (modified porcine trypsin, 37 °C, pH 7, and an enzyme/substrate ratio of 1:50) determined from the design of experiments. Surrogate peptides for 22 monolignol proteins were quantified using both concurrent and postdigest addition of stable-isotope labeled (SIL) peptides. Production and decay constants for these 22 peptides had been determined previously.[11] For peptides with rapid production rates and slow decay rates, such as the surrogate peptide IGSFEEELK for the enzymes PAL4|5, the peptide is easily detected using both methods of digestion, and quantitative accuracy is maintained whether the standard is added, concurrent with digestion or postdigestion (Figure 3a).

For surrogate peptides with rapid production and rapid decay rates, such as the surrogate peptide AAGIDSGFFELQPK for protein PAL1 (Figure 3b), the newly optimized digestion method was able to detect and accurately quantify this peptide with little discrepancy between the quantitation of samples with SIL peptides added concurrently and postdigestion, indicating attenuation of peptide decay using newly optimized conditions. In contrast, previous conditions resulted in high variability of quantitation, likely because the target peptide was near the limit of detection due to decay. The absolute quantity observed previously was much greater than that using optimized digest conditions, indicating an overestimation in the quantification. While the peptide was detected in the sample in which the standard was added postdigestion, agreement in the quantitation was not obtained, suggesting significant peptide decay had occurred. Both of these observations demonstrate that the time point at which SIL labeled standards are added to a digest affects the quantitation. Postdigestion of SIL peptides results in a significant underestimation of the target peptide, while addition of SIL peptides concurrently can result in an overestimation of the true value, particularly with peptides having rapid decay rates and slow production rates.[11] In addition to making it more difficult to obtain accurate quantitative data with little variation, peptide decay also raises the limit of detection. In some instances, it can prevent detection altogether, such as the case with the surrogate peptide of CCoAOMT3 (Figure S1, Supporting Information).

In examining peptides with slow production and slow decay, the surrogate peptide AFEIIEDLR of protein PO8 also experienced a difference between the two digestion methods when adding SIL peptides concurrent with digestion (Figure 4). In the newly optimized digestion, we observed about 50% of the absolute abundance obtained from the previous methods. While there was no peptide decay, the peptide was not completely cleaved from the protein due to an insufficient amount of enzyme. Peptides were then quantified while varying the enzyme-to-substrate ratio (1:50, 1:20, 1:10, and 1:5) using modified trypsin and fit using nonlinear regression for each peptide (Figure S4, Supporting Information) as described previously.[20] The peptide requiring the highest enzyme:sub-

dx.doi.org/10.1021/pr4008442 | J. Proteome Res. 2013, 12, 5820−5829

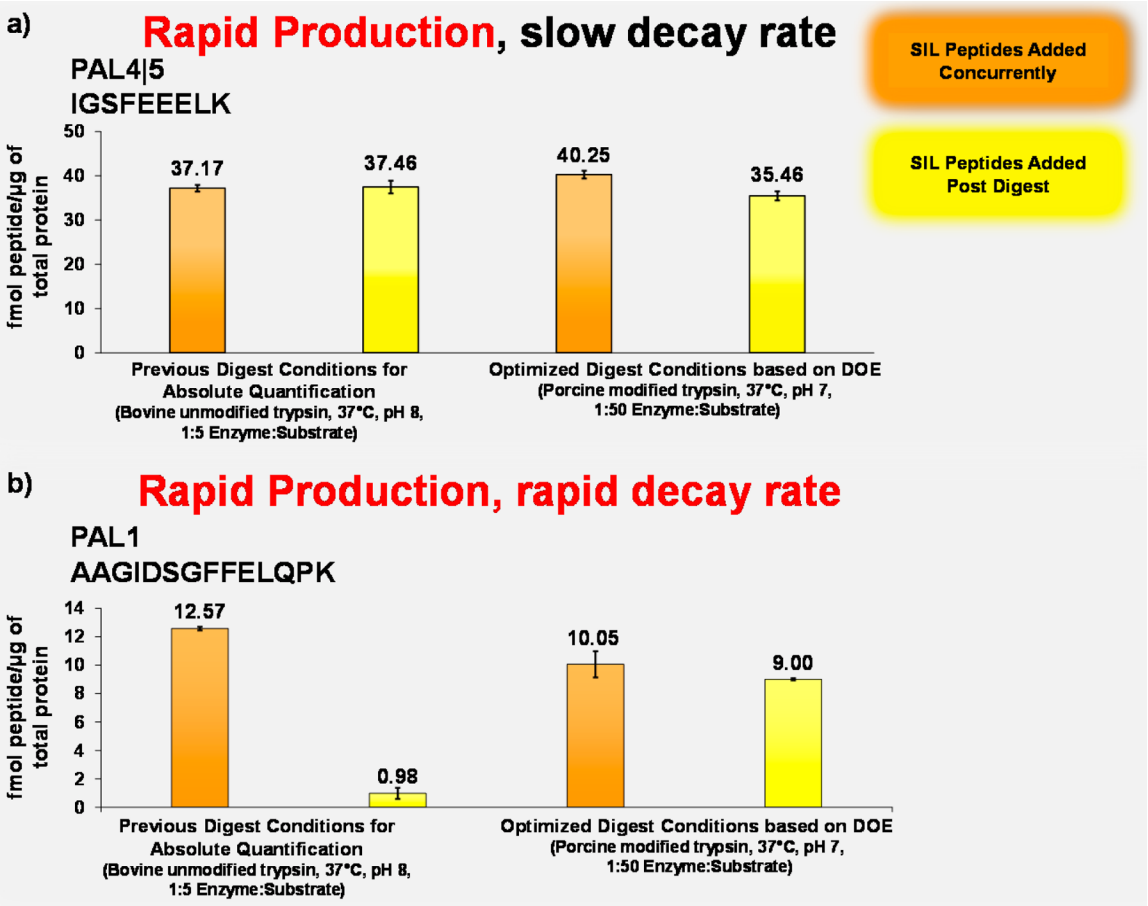


Figure 3. With peptides having low decay rates (a), the same quantitative accuracy can be achieved with either set of digestion parameters and regardless of when SIL peptide is introduced to the sample, as shown here from the surrogate peptide used to quantify the monolignol biosynthetic proteins PAL4 and PAL5. Peptides with high decay rates (b) are more difficult to detect and quantify accurately when close to or at the limit of detection, as shown by the high variability and overestimation of the measurement made using previous digestion conditions for the surrogate peptide of PAL1. A major discrepancy in the quantitation occurs using previous digest conditions and comparing concurrent and postdigest addition of SIL peptides, indicative of peptide decay. Only a slight underestimation in quantification was observed when the peptide was added postdigestion with optimized conditions, suggesting little decay occurs when using optimized digestion parameters.

strate ratio for sufficient digestion was AGPGAFLSTSEIASHLPTK from the most abundant monolignol biosynthesis protein, COMT2. It required a 1:7 enzyme/substrate ratio to achieve 95% peptide production. To ensure reproducible complete peptide production, we chose to use an enzyme/substrate ratio of 1:5 for optimized digestion in targeted proteomics experiments. Many of the surrogate peptides, including that of COMT2, were produced much more rapidly by unmodified trypsin (Figure S3, Supporting Information) than modified trypsin (Figure S2, Supporting Information). This finding suggests modified trypsin may have a reduced activity, possibly due to the modification of lysine residues.

In assessing the quantitative accuracy of the two digestion methods using concurrent addition of SIL peptides, no significant difference in the quantitation was observed for the majority of the peptides (Figure S1, Supporting Information). However, there were significant differences in the quantification of peptides with rapid decay rates such as 4CL5, CCoAOMT1, CCoAOMT2, and PAL1. In these cases, the difference in rates of peptide decay for the native and SIL peptides during digestion (because the natural peptide must be produced first before undergoing decay) results in a significant overestimation of the true quantity using previous digest conditions.

Time Course Studies, Quantifying Chymotryptic Activity, and Assessing Quantitative Accuracy

To determine the extent to which chymotryptic activity accounts for observed peptide decay, we selected two unique peptides for quantification which are the products of a chymotryptic cleavage of two tryptic surrogate peptides previously observed to have a rapid decay rate (Figure 5). A time-course study was implemented in which varying digestion times were used on a single sample of crude xylem protein. SIL standards were then added postdigestion to observe tryptic peptide decay as well as chymotryptic peptide production over time.

The time course of the tryptic surrogate peptide for CCoAOMT1 revealed significant decay over a 16 h digestion using previous digestion conditions with unmodified trypsin (Figure 6a). As the tryptic surrogate was produced, we observed the production of the chymotryptic byproduct of our surrogate, implying that unmodified trypsin over time undergoes autolysis and gains nonspecific activity. Additionally, the amount of chymotryptic peptide accounts for nearly all of the tryptic surrogate decay. This implies that it is the nonspecific activity of trypsin is largely responsible for the peptide decay. There was no observed peptide decay for the

dx.doi.org/10.1021/pr4008442 | J. Proteome Res. 2013, 12, 5820−5829

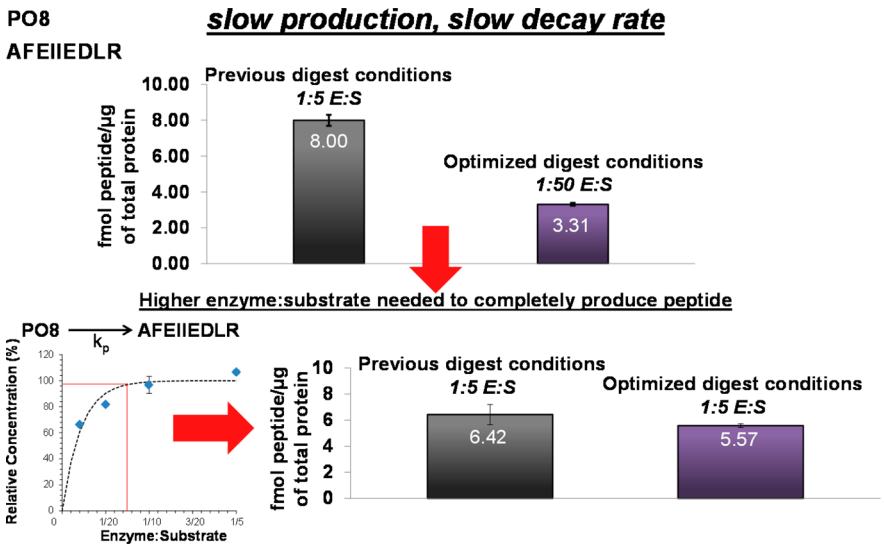


Figure 4. Peptides previously determined to have slow production rates (small K_p) displayed a discrepancy in quantification with the newly optimized conditions (SIL peptides added concurrently), yielding a lower value than previously used conditions suggesting that higher enzyme/substrate ratios are required for "complete digestion" of target proteins into their surrogate peptides. The enzyme to substrate ratio used during digestion was manipulated to confirm previous observations and to verify that a sufficient amount of enzyme was used to completely cleave the surrogate peptide from its respective protein. This ensures reproducible data and that the experimental quantification is an accurate reflection of the true concentration of the protein of interest. After manipulating the enzyme-to-substrate ratio and fitting the data using nonlinear regression, we determined the ratio required for complete digestion. An enzyme/substrate ratio of 1:5 is sufficient for complete production of all surrogate peptides. The agreement in quantitation (concurrent addition of SIL peptides) between both methods for the slowly produced surrogate of PO8 supports our hypothesis.

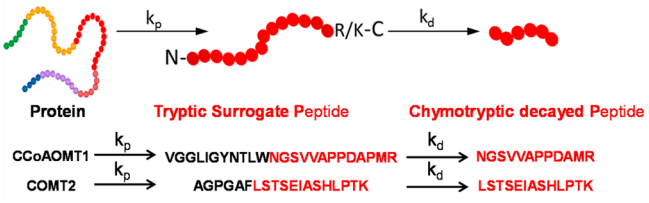


Figure 5. Tryptic surrogate peptides are produced via enzymatic digestion of the target protein at a rate k_p. Tryptic surrogate peptides are proposed to then be further digested through nonspecific chymotryptic activity of trypsin, resulting in overdigested byproducts of the tryptic surrogate. We selected two peptides and chymotryptic byproducts previously identified to quantify their production and decay over a digestion period of 16 h.

surrogate of CCoAOMT1 using optimized conditions with modified trypsin. No chymotryptic product of CCoAOMT1 was observed at any point during the digestion using optimized conditions.

The tryptic surrogate peptide for COMT2 also displayed significant decay over 16 h of digestion using previous digest conditions with unmodified trypsin (Figure 6b). Production of the chymotryptic byproduct was also observed, increasing in amount over the course of the digestion. While the absolute amount of the chymotryptic peptide did not account for the majority of the tryptic peptide decay, it still accounted for a substantial proportion (~30%). Therefore, there are likely to be other nontryptic cleavage sites that could be responsible for decay of this and other peptides. This is further evidenced by the lack of correlation in the frequency of chymotryptic sites (FLYW) and the decay rate of the peptides, suggesting that nonspecific trypsin has unique cleavage site specificity different from that of trypsin or chymotrypsin. While little decay of our tryptic surrogate of COMT2 was observed under optimized digestion conditions, we did observe a detectable amount of a chymotryptic peptide (~5 fmol/µg of protein). This suggests that while the nonspecific activity of trypsin is greatly reduced under optimized conditions, it is not fully absent. However, the 5 fmol of COMT2 chymotryptic peptide per µg of protein is negligible compared to the 350 fmol of COMT2 tryptic peptide per µg of protein. It should also be noted that time-course studies under optimized conditions showed no significant decay for any of the 20 peptides quantified (Figure S2, Supporting Information). Conversely, under previously used conditions, many of the peptides appeared to decay over the course of the 16 h digestion (Figure S3, Supporting Information).

CONCLUSION

Through the implementation of design of experiments, we were able to optimize digestion parameters for both discovery-based and targeted, absolute quantification-based proteomic workflows. We measured the influence of pH, enzyme-to-substrate ratio, temperature, and trypsin type on the quantitation of monolignol biosynthetic proteins obtained from these workflows. Each parameter had a significant influence on the number of proteins affected, unique peptides identified, as well as the sequence coverage obtained in a global proteomics experiment, while the ratio of tryptic-to-chymotryptic cleavage was largely determined by the type of trypsin used. We have demonstrated the ability to attenuate peptide decay through the use of modified trypsin, allowing us not only to detect peptides which were previously undetectable but also to quantify them with lower variability in our measurements. Further examination of the effect of the enzyme-to-substrate ratio revealed that relatively higher concentrations of enzyme were necessary in targeted proteomic workflows for sufficient peptide production to obtain the most accurate and reproducible quantification.

dx.doi.org/10.1021/pr4008442 | J. Proteome Res. 2013, 12, 5820−5829

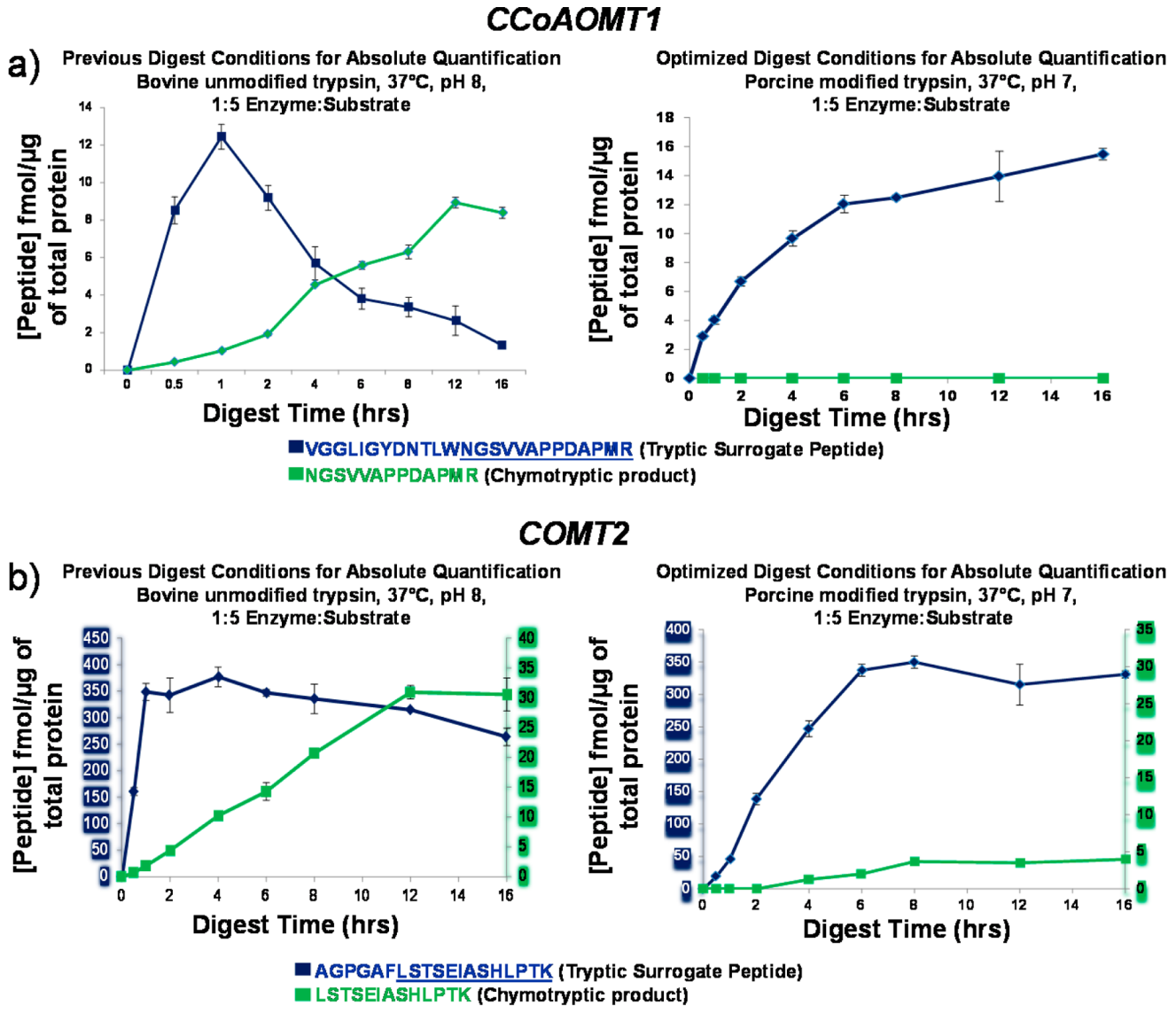


Figure 6. (a, b) Time course experiments revealed the presence of tryptic peptide decay using previous digestion conditions for both surrogate peptides. Chymotryptic activity was observed to start prior to complete production of the tryptic surrogate peptide and increased over the duration of enzymatic digestion. Meanwhile, no significant tryptic peptide decay was observed using optimized digestion conditions for absolute quantification. Additionally, chymotryptic activity was only quantifiable for one peptide under the new optimized digestion conditions, implying that while chymotryptic activity is not fully absent; it is significantly reduced using these digestion parameters.

These findings highlight the importance of digestion parameters in both qualitative and quantitative proteomics.

■ ASSOCIATED CONTENT

Supporting Information

This article contains two supplementary tables (S1–S2) and four supplementary figures (S1–S4). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 919-513-0084. E-mail: dcmuddim@ncsu.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Shan Randall for helpful discussions. The authors gratefully acknowledge The National Science Foundation Plant Genome Research Program (Grant DBI-0922391) for funding this work.

■ REFERENCES

(1) Wisniewski, J. R.; Zougman, A.; Mann, M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J. Proteome Res.* **2009**, *8* (12), 5674–5678.

(2) Wisniewski, J. R.; et al. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6* (5), 359–362.

(3) Chang, K. Y.; Muddiman, D. C. Identification of alternative splice variants in Aspergillus flavus through comparison of multiple tandem MS search algorithms. *BMC Genomics* **2011**, *12*, 358.

(4) Gokce, E.; et al. In-depth analysis of the Magnaporthe oryzae conidial proteome. *J. Proteome Res.* **2012**, *11* (12), 5827–5835.

(5) Cox, J.; Mann, M. Computational principles of determining and improving mass precision and accuracy for proteome measurements in an Orbitrap. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (8), 1477−1485.
(6) Karpievitch, Y. V.; et al. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. *Ann. Appl. Stat* **2010**, *4* (4), 1797−1823.
(7) Vandermarliere, E.; Mueller, M.; Martens, L. Getting intimate with trypsin, the leading protease in proteomics. *Mass Spectrom. Rev.* **2013**, 0.
(8) Freije, J. R.; et al. Chemically modified, immobilized trypsin reactor with improved digestion efficiency. *J. Proteome Res.* **2005**, *4* (5), 1805−1813.
(9) Song, Z.; et al. Trypsin immobilization on silica beads modified by squamous polymer for ultra fast and highly efficient proteome digestion. *Se Pu* **2012**, *30* (6), 549−554.
(10) Olsen, J. V.; Ong, S. E.; Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* **2004**, *3* (6), 608−14.
(11) Shuford, C. M.; et al. Peptide production and decay rates affect the quantitative accuracy of protein cleavage isotope dilution mass spectrometry (PC-IDMS). *Mol. Cell. Proteomics: MCP* **2012**, *11* (9), 814−823.
(12) Frenz, J.; Bourell, J.; Hancock, W. S. High-performance displacement chromatography-mass spectrometry of tryptic peptides of recombinant human growth hormone. *J. Chromatogr.* **1990**, *512*, 299−314.
(13) Vestling, M. M.; Murphy, C. M.; Fenselau, C. Recognition of trypsin autolysis products by high-performance liquid chromatography and mass spectrometry. *Anal. Chem.* **1990**, *62* (21), 2391−2394.
(14) Finehout, E. J.; Cantor, J. R.; Lee, K. H. Kinetic characterization of sequencing grade modified trypsin. *Proteomics* **2005**, *5* (9), 2319−2321.
(15) Murphy, J. P.; Pinto, D. M. Targeted proteomic analysis of glycolysis in cancer cells. *J. Proteome Res.* **2011**, *10* (2), 604−613.
(16) Kuhn, E.; et al. Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and ^{13}C-labeled peptide standards. *Proteomics* **2004**, *4* (4), 1175−1186.
(17) Wisniewski, J. R.; Mann, M. Consecutive Proteolytic Digestion in an Enzyme Reactor Increases Depth of Proteomic and Phosphoproteomic Analysis. *Anal. Chem.* **2012**, *84* (6), 2631−2637.
(18) Norrgran, J.; et al. Optimization of digestion parameters for protein quantification. *Anal. Biochem.* **2009**, *393* (1), 48−55.
(19) Wang, J. P.; et al. Functional redundancy of the two 5-hydroxylases in monolignol biosynthesis of Populus trichocarpa: LC-MS/MS based protein quantification and metabolic flux analysis. *Planta* **2012**, *236* (3), 795−808.
(20) Shuford, C. M. Comprehensive Quantification of Monolignol-Pathway Enzymes in Populus trichocarpa by Protein Cleavage Isotope Dilution Mass Spectrometry. *J. Proteome Res.* **2012**, *11*, 3390−3404.
(21) Farmer, W. H.; Yuan, Z. Y. A continuous fluorescent assay for measuring protease activity using natural protein substrate. *Anal. Biochem.* **1991**, *197* (2), 347−352.
(22) Williams, D. K.; Muddiman, D. C. Absolute quantification of C-reactive protein in human plasma derived from patients with epithelial ovarian cancer utilizing protein cleavage isotope dilution mass spectrometry. *J. Proteome Res.* **2009**, *8* (2), 1085−1090.
(23) Andrews Kingon, G. L.; et al. Multi-peptide nLC-PC-IDMS-SRM-based Assay for the quantification of biomarkers in the chicken ovarian cancer model. *Methods* **2013**, *61* (3), 323−330.
(24) Barnidge, D. R.; et al. Absolute quantification of the model biomarker prostate-specific antigen in serum by LC-Ms/MS using protein cleavage and isotope dilution mass spectrometry. *J. Proteome Res.* **2004**, *3* (3), 644−652.
(25) Shuford, C. M.; et al. Absolute quantification of free glutathione and cysteine in aquatic insects using isotope dilution and selected reaction monitoring. *Anal. Bioanal. Chem.* **2012**, *402* (1), 357−366.
(26) Stahl-Zeng, J.; et al. High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol. Cell. Proteomics: MCP* **2007**, *6* (10), 1809−1817.
(27) Shi, R.; et al. Towards a systems approach for lignin biosynthesis in Populus trichocarpa: transcript abundance and specificity of the monolignol biosynthetic genes. *Plant Cell Physiol.* **2010**, *51* (1), 144−163.
(28) Scopes, R. K. Measurement of protein by spectrophotometry at 205 nm. *Anal. Biochem.* **1974**, *59* (1), 277−282.
(29) MacLean, B.; et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26* (7), 966−968.
(30) Maclean, B.; et al. Effect of collision energy optimization on the measurement of peptides by selected reaction *monitoring (SRM) mass spectrometry*. *Analytical Chem.* **2010**, *82* (24), 10116−10124.
(31) Andrews, G. L.; et al. Improving proteome coverage on a LTQ-Orbitrap using design of experiments. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (4), 773−783.
(32) Bunkenborg, J.; Espadas, G.; Molina, H. Cutting Edge Proteomics: Benchmarking of Six Commercial Trypsins. *J. Proteome Res.* **2013**, *12*, 3631−3641.
(33) Sipos, T.; Merkel, J. R. An effect of calcium ions on the activity, heat stability, and structure of trypsin. *Biochemistry* **1970**, *9* (14), 2766−2775.