

Evaluation of N-terminal labeling mass spectrometry for characterization of partially hydrolyzed gluten proteins

Wanying Cao, Joseph L. Baumert, Melanie L. Downs*

Food Allergy Research and Resource Program, Department of Food Science and Technology, Food Innovation Center, 1901 North 21st Street, University of Nebraska-Lincoln, Lincoln, NE 68588-6205, USA

ARTICLE INFO

Keywords:

LC-MS/MS
Hydrolyzed gluten
Gluten
N-terminal labeling
Mass spectrometry
TAILS

ABSTRACT

Gluten, a group of proteins found in wheat, barley, and rye, is the trigger of celiac disease, an immune disorder that affects about 1% of people worldwide. The toxicity of partially hydrolyzed gluten (PHG) in fermented products is less well understood due to the significant analytical challenges in PHG characterization. In this project, an N-terminal labeling mass spectrometry method, terminal amine isotopic labeling of substrates (TAILS), was optimized for the in-depth analysis of PHG and validated using a test protease (trypsin) with known cleavage specificity. Gluten N-termini in test and control groups were labeled with heavy and light formaldehyde, respectively. Trypsin-generated neo N-termini were identified by exhibiting an MS1 Log2 H:L peak area ratio with a significant difference ($p < .01$) from zero and native N-termini with no significant difference from zero ($p > .01$). Using this strategy, all abundant, theoretical, test protease-generated peptides in exemplar alpha/beta gliadins and gamma gliadins were identified. **Significance:** This study is the first study that modified and evaluated TAILS analysis for the analysis of partially hydrolyzed gluten proteins. The evaluation indicated that the TAILS analysis could be modified and expanded to the identification of multiple protease cleavage sites in gluten proteins and is worth further evaluation as a novel strategy for the analysis of natural hydrolysis of gluten in food processes. This strategy also may be further applied to characterize a broader range of partially hydrolyzed allergens in foods and provide reference for their safety assessment to both industry and regulatory authorities.

1. Introduction

Gluten, which is a complex mixture of storage proteins found in wheat, barley, and rye, is responsible for the unique visco-elasticity of wheat dough [1]. Gluten is composed of alcohol-soluble monomeric prolamins (gliadins in wheat, hordeins in barley, and secalins in rye) and alcohol-insoluble polymeric glutelins [1,2]. Each gluten protein contains various types of subunits, such as α/β , γ , and ω subunits in gliadins, and low molecular weight (LMW) and high molecular weight (HMW) subunits in glutenins [2]. These subunits are highly homologous and each subunit contains numerous isoforms.

Celiac disease is a T-cell mediated immune disorder in genetically predisposed populations that affects about 1% of people worldwide [3]. Gluten can trigger symptoms of celiac disease such as gastrointestinal disorders, chronic nutrition malabsorption, or neurological diseases [3–5]. Gluten can be hydrolyzed or partially hydrolyzed to various unknown degrees during different food processes, such as fermentation. While fermented products containing detectable, residual gluten are almost certainly not suitable for individuals with celiac disease, the relative health risks and

physiological interactions with hydrolyzed or partially hydrolyzed gluten (PHG) is less well understood. Solid conclusions should be made by obtaining a thorough understanding of the PHG in fermented food products.

The detection of gluten is challenging due to various factors including, but not limited to, insolubility in aqueous extraction buffer, lack of reference materials, cultivar heterogeneity, and highly homologous protein isoforms. Enzyme-linked immunosorbent assays (ELISAs) based on the R5 and G12 antibodies have been approved by the AOAC as official methods of analysis for gliadin/gluten in certain food matrices [6,7]. However, results of different ELISA analyses can be inconsistent due to different ELISA formats (i.e. sandwich or competitive), antibody specificity, extraction procedures, and calibration standards [8–11]. In addition, current ELISA methods are not designed to differentiate gluten from different grain sources. Therefore, alternative methodologies based on immunosensors [12], mass spectrometry (MS) [13–17], genomics [18,19], or multiplex profiling [20], have been developed and investigated for the detection and quantification of gluten in food matrices. Although these methods have complemented each other and can quantitatively analyze intact gluten in complex food matrices, the detection

* Corresponding author.

E-mail address: mldowns2@unl.edu (M.L. Downs).

<https://doi.org/10.1016/j.jprot.2019.103538>

Received 9 May 2019; Received in revised form 17 September 2019; Accepted 27 September 2019

Available online 17 October 2019

1874-3919/ © 2019 Elsevier B.V. All rights reserved.

and quantification of PHG in fermented food matrices faces additional challenges. For ELISA assays, the epitope(s) for analysis might be hydrolyzed resulting in false negative results [9,21,22].

For MS analysis, the remaining gluten peptides may be too large for a direct bottom-up MS identification. Bottom-up MS analysis of PHG in processed foods has been investigated by digesting PHG with chymotrypsin or a combination of trypsin and chymotrypsin [16,21,23–27]. These studies have successfully proven the presence of detectable peptides derived from gluten in some fermented products, and some studies have also explored the absolute and relative quantification of gluten proteins. However, the actual size and sequences of relatively large PHG, which elicit a celiac reaction, are difficult to determine by typical bottom-up MS analysis and therefore accurate description and quantification of these large PHG has remained challenging. Traditional methods to characterize PHG such as two-dimensional gel electrophoresis followed by MS analysis [28] are time-consuming and lacking sensitivity for low abundance proteins. Top-down MS analysis could serve as a potential alternative strategy, but the complex natures of gluten and food matrices would make thorough analysis challenging. A novel method is critical in order to characterize these large polypeptides in PHG systems.

Differential stable isotopic labeling coupled with liquid chromatography-mass spectrometry (LC-MS) has been employed for the simultaneous discrimination of proteins from different samples in a single run and has been thoroughly reviewed elsewhere [29]. Relative quantification of proteins in different samples can be obtained by comparing the intensity of the isotopically labeled precursor ions (MS1). Among all isotopic labeling strategies, dimethyl labeling of the primary amines (the N-termini of proteins or peptides and lysine side chains) has served as a reliable, cost-effective, and multiplexed approach to discriminate proteins or peptides from various biological states and identify N-terminal post-translationally modified peptides [30,31]. Specifically, the terminal amine isotopic labeling of substrates (TAILS) strategy developed by Kleinfeld et al. [32] allowed for the high confidence discrimination of the native N-termini of proteins from neo N-termini generated by protease cleavage and therefore the identification of cleavage sites and substrates of a protease with broad or unknown specificity. Detailed step-by-step procedures for performing TAILS have been described elsewhere [33,34]. TAILS was later modified and applied for the global analysis of the N-terminome and degradomics in various cell cultures and controlled biological matrices.

To adapt TAILS for analysis of PHG, additional challenges are faced due to the unique characteristics and complex nature of gluten proteins. Firstly, gluten is only soluble in aqueous alcohol [35,36], which may affect TAILS performance due to the specific pH requirements described in the original protocol [33]. The original HEPES buffer system may not solubilize gluten proteins even in harsh denaturing and reducing conditions. Secondly, the unique amino acid sequence composition of gluten may necessitate the use of non-tryptic enzymes for the MS digestion and alternative strategies to remove labeling reagents. Last but not least, the bioinformatics challenges of highly heterogeneous cultivars, highly homologous isoforms, and incomplete annotation of protein sequence databases would make the data analysis more complex. Therefore, a proof-of-concept study is needed to adapt TAILS for the analysis of gluten proteins and the principle should be validated using a test protease with canonical cleavage specificity. In this study, TAILS was modified specifically for gluten analysis and validated as an N-terminal labeling MS strategy for the in-depth characterization of large polypeptides present in PHG.

2. Experimental methods

2.1. Gluten digestibility and enzyme selection

Due to the complex natures of gluten proteins, monomeric gliadins were chosen as the exemplar gluten source for this study. Digestion of commercial gliadins by trypsin, chymotrypsin, elastase, and pepsin (Promega, WI, USA) was conducted for the selection of the enzyme

combination to use in the validation of N-terminal labeling MS analysis of PHG. Time-related digestibility with the various enzymes was evaluated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and discovery mass spectrometry (MS) analysis. Each enzyme was added to the substrate at 1:100 (w/w) ratio and incubated for 18 h at manufacturer-recommended conditions (trypsin and elastase: 37 °C, pH = 7.8, chymotrypsin: 25 °C, pH = 7.8, pepsin: 37 °C, pH = 1.2). Pepsin digestion was conducted in simulated gastric fluid (0.084 N HCl and 35 mM NaCl in H₂O) and others in 50 mM HEPES.

2.2. Gliadin isolation and discovery MS analysis of isolated gliadin

Commercially available gliadin was investigated for MS analysis, and large, poorly-resolved peaks were found at the end of chromatography and resulted in substantial amounts of chromatographic carry-over as detailed in Fig. S-1. Therefore, an in-house isolated gliadin, which was isolated from a hard white whole wheat flour (Hodgson Mill, IL, USA) purchased from a local market (Lincoln, NE, USA) according to a modified Osborne procedure [18] (Fig. S-2), was employed as the starting material in this study since the same issues were not observed with this isolated gliadin material. In short, to remove albumins and globulins, 0.5 g of wheat flour was washed twice with 10 mL of 0.5 M NaCl at room temperature (RT) with agitation for 1 h, followed by another two washes with 10 mL of deionized water. Extraction of aqueous alcohol-soluble gliadins was then conducted using 10 mL of 70% ethanol. A 2D Quant Kit (GE Healthcare Bio-Sciences, PA, USA) was employed to quantify the extracted gliadin in samples. Aliquots of 0.5 mg crude gliadin were dried in a centrifugal evaporator (Jouan, NJ, USA) without heating and stored in a –20 °C freezer for future experiments.

Discovery MS analysis of the isolated gliadin was conducted in four digestion replicates and duplicate instrument injections to identify the most abundant gliadin isoforms presented in the starting material. An aliquot (0.5 mg) of isolated gliadin was reconstituted with 50 mM HEPES buffer (pH = 7.8) to a concentration of 1 µg total protein/µL HEPES, reduced with 10 mM dithiothreitol (DTT) at 60 °C for 30 min, alkylated with 10 mM iodoacetamide (IAA) at RT in the dark for 30 min, and denatured with 1 M guanidine hydrochloride (GuHCl) at 65 °C for 15 min. Subsequently, 30 µg of the processed gliadin was digested with chymotrypsin (1:100 w/w) at 25 °C overnight. Sample desalting was conducted using Pierce™ C18 Spin Columns (Thermo Scientific, MA, USA) following manufacturer's instructions. The desalted gliadin peptide mixtures were dried in a centrifugal evaporator and reconstituted with 0.1% formic acid (FA) in H₂O with 3% ACN for LC-MS/MS analysis.

2.3. Dimethylation labeling verification

A dimethylation labeling verification of heavy formaldehyde (¹³CD₂O, 20% in D₂O (w/w)) (Cambridge Isotope Laboratories, MA, USA) and light formaldehyde (CH₂O, 37% in H₂O (w/w)) (Sigma-Aldrich, MO, USA) with gliadins was conducted on fully chymotrypsin-digested isolated gliadins. A digested gliadin sample was divided into two aliquots and labeled with 40 mM light formaldehyde [+H(4)C(2)] or 40 mM heavy formaldehyde [+D(4)¹³C(2)], in the presence of 20 mM of sodium cyanoborohydride (NaBH₃CN) (Sigma-Aldrich, MO, USA) at pH 6–7 at 37 °C overnight. Dimethyl labeled samples were desalted using C18 spin columns, dried in a centrifugal evaporator, reconstituted with 0.1% FA in H₂O with 3% acetonitrile (ACN), and combined in equal volumes before LC-MS/MS analysis.

2.4. Initial N-terminal labeling MS analysis

The sample preparation flow diagram for gluten-specific TAILS is shown in Fig. 1. In short, two 0.5 mg aliquots of isolated gliadin were each reconstituted with 125 µL of 50 mM HEPES buffer (pH = 7.8), reduced with 10 mM DTT at 60 °C for 30 min, and alkylated with 15 mM IAA at RT in dark for 30 min. To provide predictable hydrolysis which

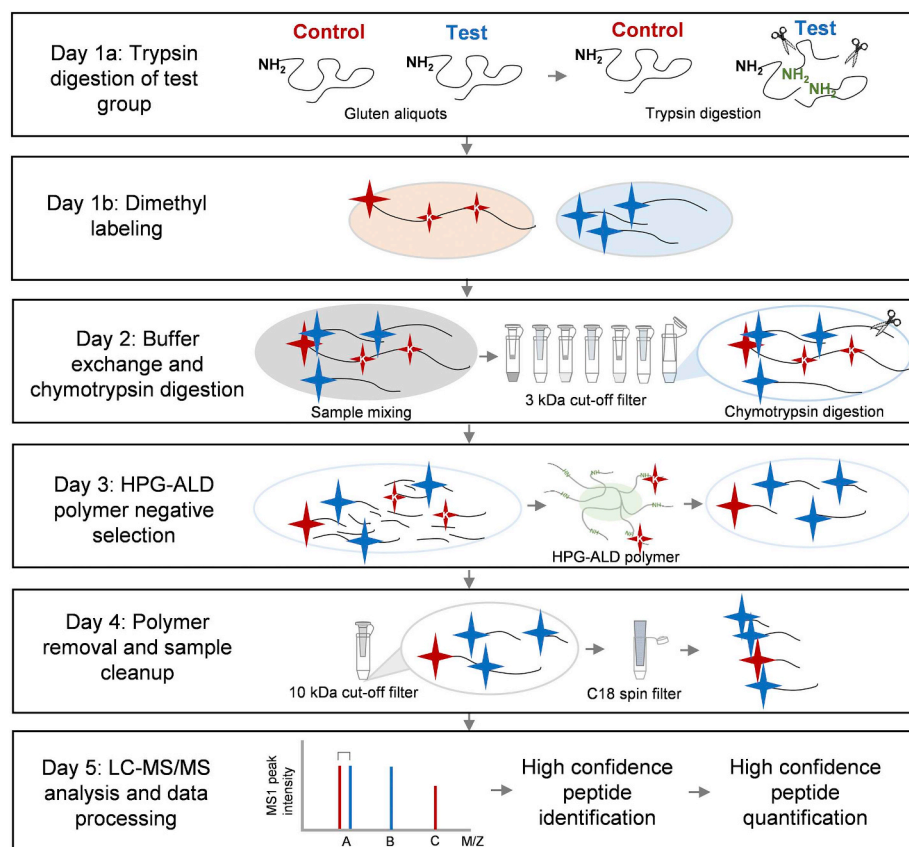


Fig. 1. Flow diagram for gluten N-terminal labeling MS analysis. The primary amines of the native N-termini (black NH₂), neo N-termini (green NH₂) and lysine (K) of gluten proteins from control (red) and test (blue) groups were labeled with light (red star) and heavy (blue star) formaldehyde, respectively. Peptides with equally distributed light and heavy formaldehyde were native N-termini (peak A). Heavy singleton indicated a peptide with trypsin generated neo N-termini (peak B) and light singleton yield a native N-termini with an internal trypsin cleavage (peak C).

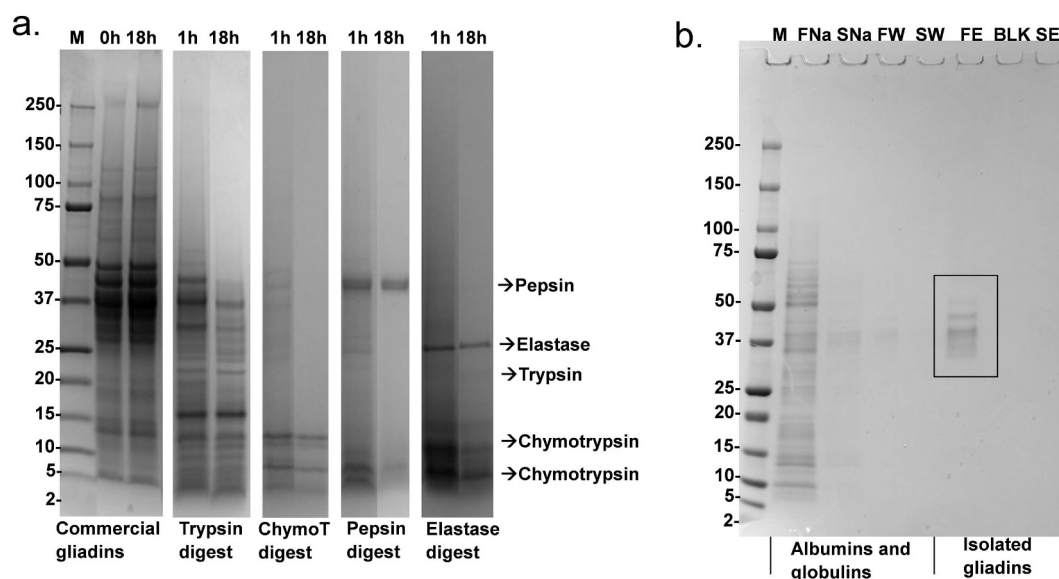
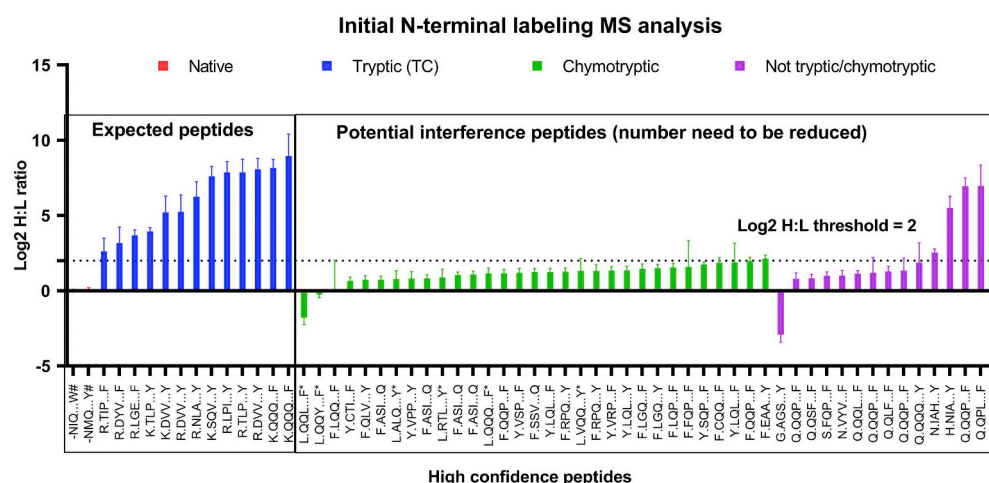
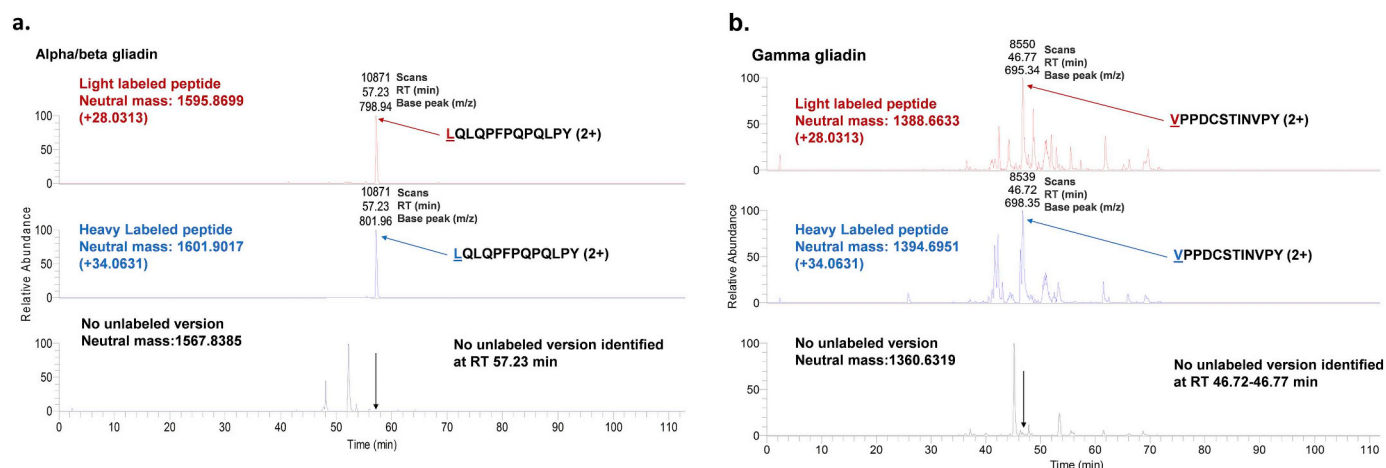


Fig. 2. (a) SDS-PAGE gel of commercial gliadin mixture digested with various enzymes, from left to right: marker, intact gliadin at 0 h and 18 h, trypsin, chymotrypsin, pepsin, and elastase digested gliadin at 1 h and 18 h, and (b) SDS-PAGE gel of different fractions of wheat protein as isolated from different steps of the modified Osborne procedure, from left to right: marker, first/second NaCl extract (FNa/SNa), first/second water extract (FW/SW), first 70% ethanol extract (FE), blank (BLK), second 70% ethanol extract (SE).

could easily be adapted for TAILS validation, one aliquot, the test sample, was digested with trypsin (1:100 w/w) at 37 °C for 4 h, and the other aliquot, the control sample, received no trypsin digestion. After digestion, IAA in both samples was quenched with 15 mM DTT at RT for 25 min to prevent the free amine group on IAA from competing with the amine groups on proteins during dimethylation labeling. Samples were then denatured with 4 M GuHCl at 65 °C for 15 min. Proteins in test and control samples were then labeled with heavy and light formaldehyde,

respectively, following the conditions described above. To ensure complete labeling, 20 mM formaldehyde and 10 mM NaBH₃CN were added at the end of labeling and incubated at 37 °C for another 2 h. Dimethylation reagents were quenched with 100 mM Tris (pH = 6.8) at 37 °C for 1 h. The test and control samples were then combined in equal volumes, and the labeling reagent was removed by a buffer exchange with 50 mM HEPES using 3 kDa centrifugal filters (Millipore Sigma, MA, USA) according to manufacturer's instructions. In the original TAILS protocol



#Log2 H:L of peptide "-NIQ...W" (p=0.575) and "-NMQ...Y" (p=0.026) was not significant from 0.
*Additional chymotryptic peptides resulted by the lower affinity of the chymotrypsin cleavage after L.
All other peptides showed Log2 H:L significant from 0 (p<0.01).

[33], an acetone/methanol precipitation was used to remove labeling reagents but was not employed in gluten analysis since substantial PHG sample loss was observed during preliminary precipitation trial (Fig. S-3). The use of spin filters could lead to the loss of peptides with sizes < 3 kDa, with filtration efficiency also affected by the shape of peptide in buffer [37]. Buffer exchange, however, was found to be more suitable than precipitation for gluten analysis.

After the buffer exchange, chymotrypsin was added at 1:100 (w/w) and incubated at 25 °C overnight to generate peptides of a size more suitable for LC-MS/MS. A free amine-binding high-molecular-weight dendritic polyglycerol aldehyde polymer (HPG-ALD polymer) (The University of British Columbia - UBC, Canada) was subsequently added at 1:5 ratio (w/w, peptide:polymer) for the negative selection of dimethyl-labeled peptides, following manufacturer's instructions. The polymer was removed using 10 kDa spin filters (Millipore Sigma, MA, USA) and samples were desalted with C18 spin columns. The desalted peptides were dried and reconstituted with 0.1% FA in H₂O with 3% ACN for LC-MS/MS analysis. The experiment was conducted in eight digestion replicates, each with three injection replicates (from a single preparation of isolated gliadin, $n = 24$ in total).

2.5. Protocol optimization of N-terminal labeling MS analysis

To enhance the removal of labeling reagents using 3 kDa spin filters,

the efficiency of five reagents (100 mM Tris (original protocol), 200 mM Tris, 100 mM ammonium bicarbonate (AB), 100 mM Tris with the addition of NaBH₃CN (Tris-NaBH₃CN), and 100 mM free lysine) for quenching formaldehyde was further investigated, and conducting the buffer exchange process three times instead of only once was also evaluated to minimize the remaining labeling reagent in the filter. Further, to aid the digestion of PHG with chymotrypsin, the effects of adding 1 M GuHCl or a commercial MS-compatible surfactant (RapiGest) (Waters, MA, USA) to triple buffer exchange treated samples were also probed. Single digestion replicates and duplicate injection replicates were used in all protocol optimization experiments.

2.6. LC-MS/MS analysis

For each sample, 4 μL of reconstituted peptide mixture (300 ng/ μL) was chromatographically separated on a Thermo Scientific™ Ultimate 3000 RSLC ultrahigh performance liquid chromatography (UHPLC) system using a Thermo Scientific™ Hypersil GOLD™ C18 Selectivity LC column (100 mm $\times$ 1 mm i.d., 1.9 μm particle size), at a flow rate of 0.06 mL/min using a linear gradient of 2–40% solvent B over 70 min and retaining at 40% solvent B for another 6 min (solvent A: 0.1% FA in H_2O ; solvent B: 0.1% FA in ACN). The data-dependent MS/MS analysis was performed on a Thermo Scientific™ Q Exactive™ Plus Hybrid Quadrupole-Orbitrap™ mass spectrometer. The top ten most intense

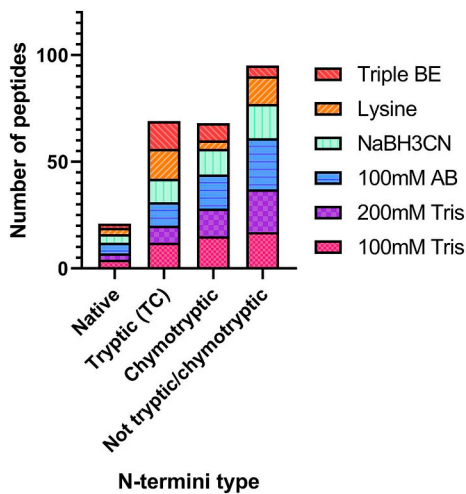


Fig. 5. Number of high confidence peptides (FDR < 0.01) found in different test groups in evaluation of quenching reagents. Each column indicates the number of high confidence peptides with different types of N-termini. Data from different test groups with the same type of N-termini are stacked in one column and illustrated in different color. TC indicates trypsin-chymotrypsin generated target peptides. Experiments were conducted in duplicate injections.

parent ions were fragmented following each full scan (70,000 resolution, 400–1400 m/z scan range, 3e6 automatic gain control (AGC) target, 100 ms maximum injection times (IT), and 1+ to 5+ charge state) for the MS2 analysis (70,000 resolution, 200–2000 m/z scan range, 1e5 AGC target, and 60 ms maximum IT). The minimum AGC target was set at 1.50e3 for an intensity threshold of 2.5e4. A 2.0 m/z isolation window and normalized collision energy of 27 were used.

2.7. LC-MS/MS data processing

A SEQUEST HT search against the *Triticum aestivum* protein sequence database (UniProt, 136,866 sequences, Aug11th, 2017) containing the common Repository of Adventitious Proteins (cRAP) (The Global Proteome Machine, 115 sequences, version 2012.01.01) was performed with Proteome Discoverer 2.1 (PD) (Thermo Scientific, MA, USA). A precursor mass range from 350 Da to 5000 Da (1+ to 5+) and a maximum of two missed cleavages were allowed, and the precursor and product ion mass tolerances were ± 10 ppm and ± 0.02 Da, respectively. The enzymatic cleavage site was set as chymotrypsin (FWY/not P) and chymotrypsin semi-C (only C termini resulted by the cleavage of FWY/not P) for discovery and N-terminal labeling MS data processing, respectively. Fixed modifications included carbamidomethylation (C), and variable modifications included oxidation (M) and deamidation (N/Q). Additional fixed modifications for the N-terminal labeling MS data processing, light (+28.031 Da) and heavy (+34.063 Da) dimethylation (peptide/protein N-termini and Lys), were searched separately and reported in the same output file. Peptides identified with a target-decoy-based False Discovery Rate (FDR) < 0.01 across all replicates ($n = 24$ in initial N-terminal labeling MS) were used as high confidence peptides and imported to the Skyline 4.1.0 (MacCoss Lab, WA, USA) [38] for calculation of the Log2 heavy:light ratio (Log2 H:L) of the peptide MS1 peak area. Peak areas of the same peptide from different charge states and variable modifications were summed.

2.8. Statistical analysis

For each high confidence peptide, a two-tailed Student's *t*-test to compare the Log2 H:L with 0 was performed via R 3.4.0 (R Core Team). Native N-termini were identified by showing a Log2 H:L with no significant difference from 0 ($p > .01$). Tryptic cleavage-generated neo N-

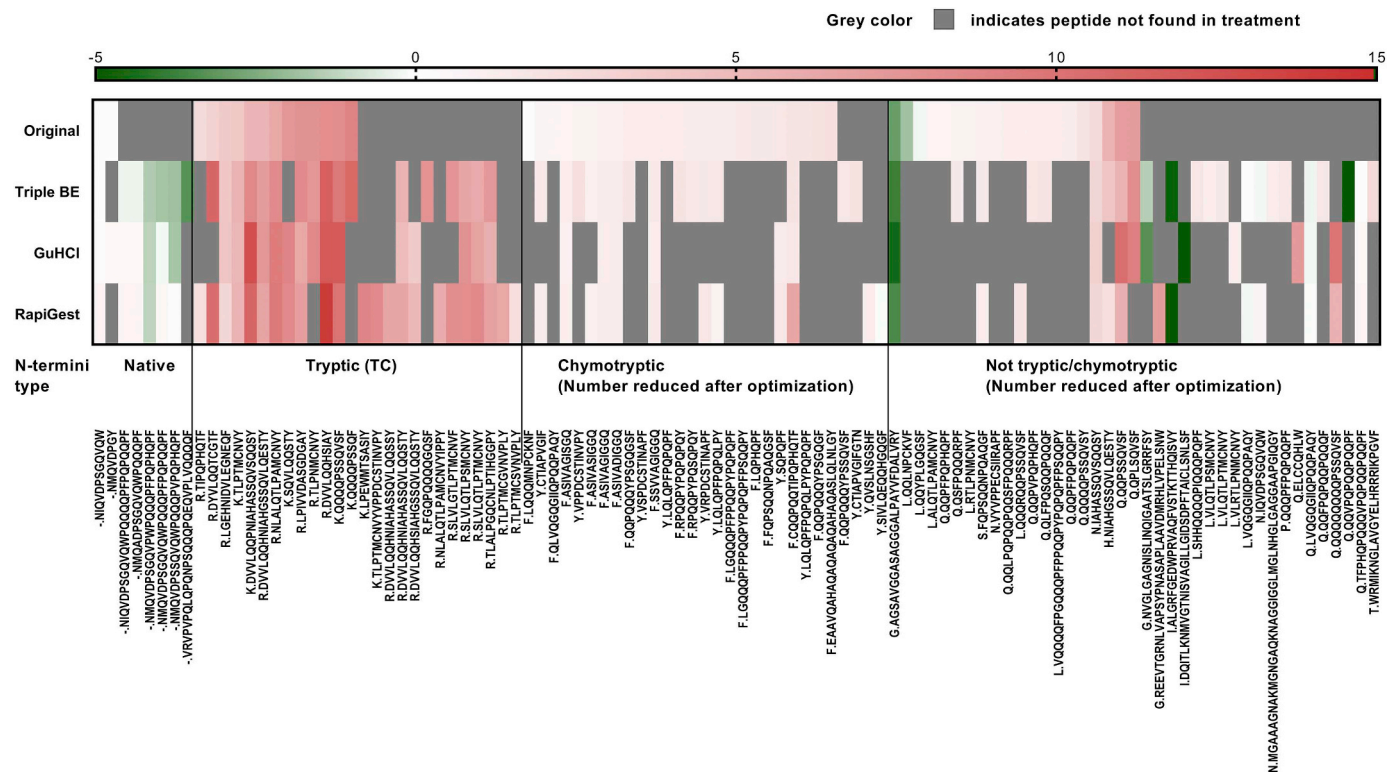


Fig. 6. Heat map for Log2 H:L of high confidence peptides with different types of N-termini as identified in the digestion optimization experiments. TC indicates trypsin-chymotrypsin generated target peptides. Different rows display experiment groups, from top: the original eight digestion replicates trial as control, triple buffer exchange (BE), and chymotrypsin digestion conducted in the presence of 1 M GuHCl or surfactant. Colors indicate the Log2 H:L, as shown in the legend at the top. Log2 H:L values less than -5 are indicated as -5 in the color scale. Grey color indicates peptides not identified in treatment. Digestion optimization experiments were conducted with duplicate injections.

Table 1
Gluten proteins and peptides found in bottom-up discovery MS analysis.

Protein group	Uniprot ID	No. of peptide	Peptide sequence	Additional associated protein ID
Alpha/beta-gliadin	^a X2KVH9	5	F.LGQQQPFPPQYPQPFPSPQPY.L ^b	P04721
			Y.LQLQFPFPQQLPY.S ^b	P18573, P02863
			F.RPQQYYPQPQPY.S ^b	P18573, P02863, P04721, K7XRA1, J7HT09
			F.QQPQQYPLGQGSF.R	
			Y.IPPYCTIAPF.G ^b	P02863, K7XRA1
	^a P18573	5	Y.LQLQFPFPQQLPYQPQLPYQPQPF.R	
			F.QQPQQYPSGQGSF.Q ^b	I0IT51, J7HT09
	P04721	4	Y.IPPYCTIAPVGIF.G ^b	P04724
			Y.LQLQFPQLPY.S	
	P02863	4	F.LQPQLPY.S	
	K7XRA1	4	F.LGQQQPFPPQYPQPFPSPQLPY.L	
			Y.LQLQFPFPQQLPYQPQPF.R	
	I0IT51	3	F.QQPQQYPSGQGF.F	
			Y.QQPQQYPSGQGSFQPSQNPQAQGF.V	
	J7HT09	3	Y.IPPYCTIAPF.G	
	P04724	2	Y.LQLQFPFPQQLPYQPQLPYQPQPF.R	
	P04725	2	Y.MQLQFPFPQQLPYQPQLPYQPQPF.R	
Gamma-gliadin	^a A1EHE7	4	F.QQPQQYPSQVSF.Q	P04727
			Y.IPPHCSTIAPF.G ^b	
			Y.LQLQFPFPQPF.L	
			F.SQQPQQTTFPQQTTFPHQPQQQFPQPQQQF.L	
			F.SQPQQQFPQPQQSFPQQQPPFIQPSLQQQVNPC KNF.L	
	^a P06659	4	Y.VPPECSIHKAPF.S	
			F.SSVVAGIGGQ.-	
			F.SQQPQQIFPQPQQTTFPHQPQQQFPQPQQQF.L	
			Y.VPPYCTIRAPF.A	
			Y.CSTIRAPF.A	
	^a P21292	4	F.ASIVASIGGQ.-	
			F.CQQPQRTIPQPHQTF.H	
			F.LQQQMNPCKNF.L	
			Y.VPPDCSTINVPY.A	
			Y.ANIDAGIGGQ.-	
	M9TG60	2	Y.LQQQMNPCKNY.L	
			F.ASIVADIGGQ.-	
			F.CQQPQQTIPQPHQTF.H	
			Y.VPPECSIIRAPF.A	
			Y.EAIRAIY.S	
LMW-GS ^c	P04729	1	Y.QAIRAIY.S	
	P04730	1	Y.DAIRAIY.S	
	B2BZD1	2	Y.SIVLQEQHQHGGF.N	

^a Protein isoforms used for gluten N-terminal labeling MS coverage analysis.

^b Shared peptide with multiple associated protein IDs.

^c LMW-GS abbreviated for low molecular weight – glutenin subunit.

termini with unambiguously assigned Lys or Arg in the P1 position were identified by showing a significant difference ($p < .01$) from 0.

3. Results and discussion

3.1. Selection of enzymes and starting materials

Selection of the enzyme pairs to be used for validating N-terminal labeling MS analysis was done by evaluating the enzymatic digestibility of commercially available gliadin, which is a crude gliadin mixture (Fig. 2a). The digestibility was evaluated with SDS-PAGE and summarized in Fig. 2a (detailed in Fig. S-4). Resulting peptides were investigated by discovery MS analysis as indicated in Fig. S-1. Trypsin started to partially hydrolyze gluten into polypeptides as early as 0.5 h (Fig. S-4a) but could not completely digest gluten into peptides even after 18 h digestion. Therefore, to easily adapt this first digestion into N-terminal labeling MS analysis protocol, a 4 h trypsin digestion was employed as the test enzymatic hydrolysis to generate PHG polypeptides for the method validation. Chymotrypsin and pepsin can digest gluten to peptides with molecular weights up to 15 kDa and 10 kDa. However, pepsin was not suitable in this study because its theoretical resulting peptides were too short, and the low pH required for pepsin activity made it difficult to be adapted to the sample preparation protocol. Elastase could also digest gluten into peptides < 15 kDa, except

for a less intense band at 50 kDa. However, elastase has too many cleavage sites, which would introduce a considerable challenge to the N-terminal labeling MS data analysis. Consequently, chymotrypsin was incorporated into this study as the second enzyme to generate the peptides of a size more suitable for LC-MS/MS.

As detailed in Fig. S-1b, substantial amounts of chromatographic carry-over were observed with the MS analysis of commercial gliadins. Using the modified Osborne procedure (Fig. S-2), a gliadin mixture was isolated from a washed, hard white whole wheat flour (Fig. 2b) without any notable chromatographic issues. Therefore, the isolated gliadin was used as an MS-compatible exemplar gliadin source for the remaining N-terminal labeling MS analysis in this study.

3.2. Dimethyl labeling verification

A verification of the dimethyl labeling efficiency of gliadin peptides was conducted using the fully chymotrypsin-digested gliadin. Samples were labeled with light or heavy formaldehyde individually and desalted with C18 spin columns without the removal of non-dimethylated peptides, and analyzed by high-resolution MS. As illustrated in Fig. 3, both alpha/beta gliadin and gamma gliadin N-termini can be labeled with light formaldehyde (+28.0313 Da) or heavy formaldehyde (+34.0631 Da) successfully. The mass difference between the heavy and light labeled peptide (6.0318 Da) was large enough to ensure an

Table 2

List of target TC peptides and peptides with native N-termini found from gluten N-terminal labeling MS analysis using the optimized digestion protocol.

ID	P-1	Peptide sequence	Missed tryptic cleavage	Average Log2 H:L ^b	SD ^d
N1	–	NIQVDPGQVQW	0	0.38	0.30
N2	–	NMQVDPGY ^a	0	0.39	NA ^c
N3	–	NIQVDPGQVQWPQQQPFPPQPPF	0	0.28	0.45
N4	–	NMQADPSGQVQWPQQPF	0	0.24	0.43
N5	–	NMQVDPGQVQWPQQPFPPQPPF	0	–1.37	0.11
N6	–	NMQVDPGQVQWPQQPFPPQPPF	0	–0.50	0.92
N7	–	NMQVDPSSQVQWPQQPFPPQPPF	0	–1.14	0.89
N8	–	VRVPVPLQFPQNPQQPQEQVPLVQQQF ^a	1	–3.26	NA
TC1	R	DYVLQQTCTGF	0	10.72	0.20
TC2	R	LGEHNIDVLEGNEQF	0	3.64	0.59
TC3	K	TLPTMCNVY	0	5.29	0.10
TC4	K	DVVLQQPNIAHASSQVSQSY	0	10.75	1.79
TC5	R	DVVLQQHNIHAGSSQVLQESTY	0	7.05	0.39
TC6	R	NLALQTLPMACNVY	0	9.09	0.59
TC7	K	SQVLQSTY	0	8.58	0.22
TC8	R	LPIVVDASGDGAY	0	7.18	0.88
TC9	R	TLPNMCNVY	0	8.02	0.52
TC10	R	DVVLQQHSIAY	0	12.58	1.18
TC11	K	QQQQPSSQVSF	0	10.53	0.93
TC12	K	QQQQPSSQF ^a	0	10.82	NA
TC13	K	LPEWMTSASIY ^a	0	8.74	NA
TC14	K	TLPTMCNVYPPDCSTINVPY ^a	0	8.35	NA
TC15	R	DVVLQQHNIHASSQVLQSSSY ^a	0	6.36	NA
TC16	R	DVVLQQHNIHASSQVLQSTY	0	5.52	0.65
TC17	R	DVVLQQHSIAGSSQVLQSTY	0	4.36	0.71
TC18	R	FGQPQQQGGQSF ^a	0	7.72	NA
TC19	R	NLALQTLPMACNVYIPPY ^a	0	5.97	NA
TC20	R	SLVLGTLPTMCNVF	0	7.89	0.23
TC21	R	SLVLQTLPSMCNVY	0	7.44	0.79
TC22	R	SLVLQTLPTMCNVY	0	7.08	1.01
TC23	R	TLALPGQCNLPTIHGGPY	0	6.59	0.68
TC24	R	TLPTMCGVNVPLY ^a	0	6.23	NA
TC25	R	TLPTMCSNVNVPLY ^a	0	2.30	NA
TC26	R	TIPQPHQTF ^a	0	2.61	NA

The ID of peptide with a native and tryptic N-termini was shown as N and TC, respectively.

^a Indicates peptides found only in one digestion optimization experiment.

^b Data from digestion optimization trial (triple buffer exchange, 1 M GuHCl, and surfactant) were averaged. All experiments were conducted using the same gluten extract.

^c Data not available.

^d SD: standard deviation.

accurate discrimination of the two types of peptides at the same retention time. A dynamic search for the presence of light or heavy dimethylation was conducted to check the labeling efficiency, and no unlabeled peptides were identified in all samples (Table S-1). All related raw data and search result files were uploaded to the PRIDE Archive (project PXD012296).

3.3. Initial gluten N-terminal labeling analysis

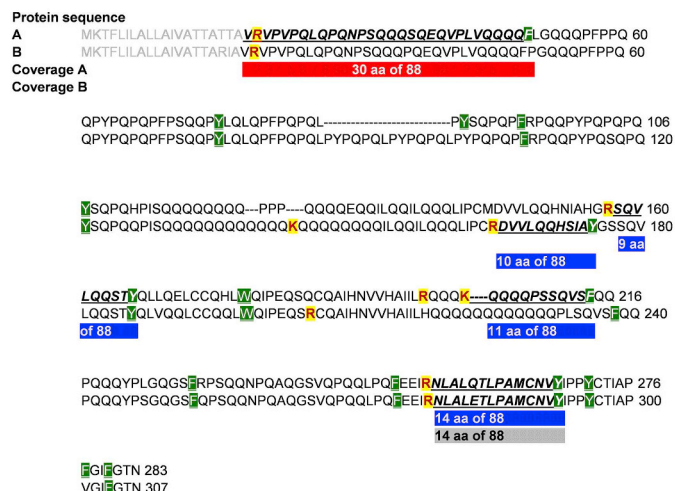
In the initial gluten N-terminal labeling MS analysis, 60 high confidence dimethylated peptides were identified across all replicates ($n = 24$) and peptide information is detailed in Table S-2. As indicated in Fig. 4, two of the 60 peptides were confirmed as peptides with native N-termini by demonstrating a Log2 H:L with no significant difference from zero ($p > .01$). The other 58 peptides all showed Log2 H:L with significant difference ($p < .01$) from zero. Thirteen peptides were confirmed as trypsin-chymotrypsin generated peptides (TC), with unambiguously assigned Lys or Arg in the P1 position. However, the other 45 peptides displayed non-tryptic N-termini, where 31 of those had chymotryptic N-termini and the other 14 showed neither tryptic nor chymotryptic N-termini.

For the 14 peptides with neither tryptic nor chymotryptic N-termini (Log2 H:L ranging from -2.92 to 6.97), various reasons are associated with the presence of those peptides. One peptide (G.AGSAVGGASAGGG-ALPAYVFDALVRY) showed a negative Log2 H:L ratio. Negative values could be observed due to the presence of the internal trypsin cleavage site in the peptide, which was only identified in the light-labeled control sample. As the N-terminus of this peptide in the control sample was

neither tryptic nor chymotryptic, its presence is indicative of proteolysis at this site prior to sample preparation (e.g. proteolysis occurring in seed, during flour production or storage, etc.). Nine of those 14 peptides showed a varied amino acid at P1 (Q (7) / N (1) / S (1)) but a unique independent consistent heavy and light ratio pattern (Log2 H:L close to 1), which could have arisen following unspecific chymotrypsin cleavage, as described below. The last four peptides displayed Log2 H:L values similar to TC peptides, which could due to some cleavage or degradation occurring during sample preparation (with at least one of these peptides, Q.QQPSSQVSF, appearing to be a product of a peptide with a tryptic N-terminus, K.QQQQPSSQVSF).

In addition to the preferential cleavage sites (FWY/not P), chymotrypsin used in this study also could cleave after L and M at a lower rate. While L/M cleavage specificity was not included in the search parameters, it is likely that some of the peptides observed with L/M in the P1 position are the result of chymotrypsin cleavage and therefore were grouped with the other chymotryptic peptides. For peptides with chymotryptic N-termini (31), 26 of them were internal fully chymotrypsin cleaved peptides (both termini were chymotryptic) and the other 5 were chymotryptic C-termini peptides from gamma gliadins. Fully chymotryptic peptides were not expected to be found in this experiment. The objective of chymotrypsin digestion was to generate peptides with a size more suitable for LC-MS/MS after labeling, and the newly generated peptides with chymotryptic N-termini were to be removed from the sample using an HPG-ALD polymer. The consistent unique Log2 H:L (Log2 H:L close to 1) of these peptides led to a hypothesis regarding their presence: some amount of unquenched dimethyl

a. CLUSTAL O(1.2.4) multiple sequence alignment of alpha/beta gliadins (total 74 aa of 88 ~ 84%)



b. CLUSTAL O(1.2.4) multiple sequence alignment of gamma gliadins (total 114 aa of 114 ~ 100%)

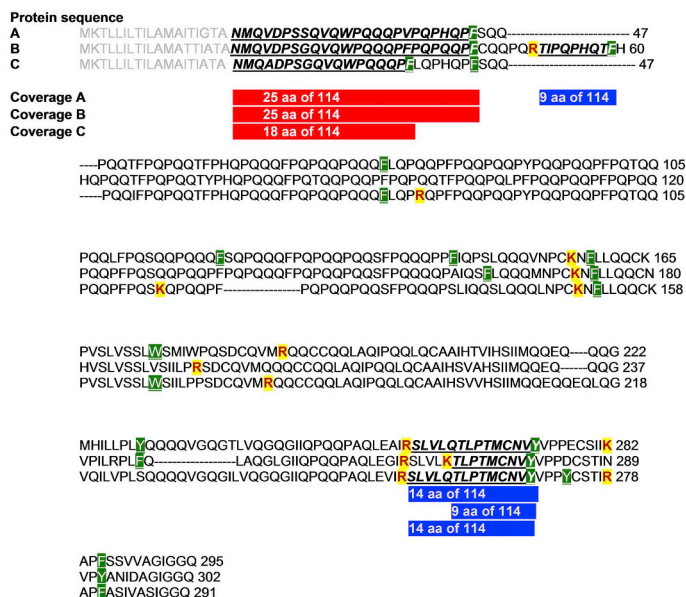


Fig. 7. Efficiency of optimized gluten N-terminal labeling MS of (a) alpha/beta gliadins (UniProt IDs: A: [X2KVH9](#), B: [P18573](#)) and (b) gamma gliadins (UniProt IDs: A: [A1EHE7](#), B: [P21292](#), C: [P06659](#)). All trypsin and chymotrypsin cleavage sites are highlighted in yellow and green, respectively. All in-silico predicted TC peptides are underlined (a total of 88 and 114 amino acids (aa) for alpha/beta gliadins and gamma gliadins, respectively). Native and TC peptides identified by N-terminal analysis are highlighted in red and blue, respectively. Peptides highlighted in grey indicate in-silico predicted peptides that were not found. Efficiency was evaluated by comparing the number of experimental identified amino acids to the in-silico predicted number.

labeling reagents were retained in the 3 kDa spin filter used for buffer exchange and subsequently reacted with free amines on the newly generated chymotryptic N-termini.

To validate the hypothesis, a fully chymotrypsin digested gliadin mixture was labeled with equally mixed heavy and light formaldehyde, and the Log2 H:L were compared to the results from labeling verification (i.e. where heavy- and light-labeled digest samples were mixed after labeling). As illustrated in Fig. S-5, conducting dimethylation with equally mixed heavy and light formaldehyde can result in the unique Log2 H:L pattern (close to 1, average Log2 H:L = 0.8). Further investigation of the flow-through from the buffer exchange indicated the presence of the flow-through was likely related to the unique log2 H:L pattern (Fig. S-6, average Log2 H:L ratio is 1.77 and 1.66 in 20% and 10% spike, respectively). Therefore, peptides with the unique Log2 H:L pattern in Fig. 4, mostly chymotryptic peptides, were likely interference peptides and should be reduced in future experiments. Consequently, a series of methodology optimization trials were conducted to reduce the number of potential interference peptides, and additional data analysis strategies were investigated.

3.4. Protocol optimization of N-terminal labeling MS analysis

To reduce the interference from the remaining labeling reagents in the spin filters, the efficiency of five types of reagents for quenching formaldehyde were evaluated, and conducting buffer exchange three times instead of only once was also investigated to minimize the remaining labeling reagent in the filter. As indicated in Fig. 5, there were no obvious differences in the number of high confidence peptides (FDR < 0.01) identified between the original quenching method (100 mM Tris), 200 mM Tris, 100 mM AB, and Tris-NaBH₃CN. However, both the use of free lysine and triple buffer exchange did reduce the number of potential interference peptides (the last two columns in Fig. 5) without negative effects of the target TC and native N-termini peptides. A detailed peptide list with Log2 H:L ratio is shown in Table S-3 and all related files were uploaded to PXD012296. Although quenching with lysine seems a bit more efficient comparing to the original quenching reagent (100 mM Tris), an extra step of adjusting of

pH from pH 12 to pH 6–7 was needed. Therefore, 100 mM Tris quenching together with a triple buffer exchange was used for all subsequent gluten N-terminal labeling MS analysis.

Effects of 1 M GuHCl or a commercial MS-compatible surfactant (RapiGest) for enhancing the secondary chymotrypsin digestion of PHG for LC-MS/MS were investigated. All digestion evaluation experiments were conducted using triple buffer exchange and results were compared to triple buffer exchange without digestion enhancer as well as the initial N-terminal labeling MS analysis results. As indicated in Fig. 6 (detailed in Table S-4), the use of 1 M GuHCl and surfactant both increased the number of TC peptides from 13 to 15 and 23, respectively. The number of identified peptides with chymotryptic N-termini was reduced from 31 to 7 and 14 when conducting the secondary digestion with the presence of 1 M GuHCl (77% decrease) and surfactant (55% decrease), respectively. A decrease from 14 to 10 and 11 of peptides in the non-trypsin/chymotryptic group was also observed when the sample was treated with GuHCl and surfactant, respectively.

Although interference peptides were still present, it is evident that the combination of triple buffer exchange and a digestion enhancer leads to improved results, where the number of interference peptides, especially fully chymotryptic peptides, was reduced without a negative impact of target TC or native peptides. Since approximately 50 µL of solution always remains presents in the filter unit after centrifugation, as stated by the manufacturer, application of additional data interpretation measures may be needed to aid in distinguishing true test-protease-generated peptides from those acquiring a label from the remaining unquenched reagents. In the mixed-labeling experiments, a labeling bias towards the heavy formaldehyde was observed, resulting in a consistent Log2 H:L of approximately 1 (Fig. S-5). Based on these data, a proposed Log2 H:L threshold of > 2 is proposed to eliminate the interference peptides during data interpretation. Setting a Log2 H:L threshold of 2 reduced the number of chymotryptic and non-specific peptide from 31 to 3 and 14 to 5, respectively, as detailed in Table S-5. As the precise experimental conditions and objectives may vary in future applications of this type of method, individual assessments and quality control checks of appropriate Log2 H:L thresholds will likely be necessary [39]. Other quenching practices, such as quenching with

lysine, could also be further explored for improved results, if needed in a specific application. Other quenching methods may also be necessary if small peptides (< 3 kDa) are of interest, as they may be removed during buffer exchange. In these cases a combination of direct analysis of small peptides and TAILS analysis of large polypeptides may be appropriate.

3.5. Efficiency of gluten N-terminal labeling MS analysis

The efficiency of this method was evaluated by comparing an in-silico prediction of TC peptides and native N-termini peptides with those identified in the experimental N-terminal analysis. To predict and visualize TC peptides, proteins identified in discovery analysis were aligned and used as a protein “map”. In detail, for the isolated gliadin samples digested with chymotrypsin, a total of 40 peptides from different subunits of gluten proteins (gliadins and glutenins) were identified across four digestion replicates (Table 1). Due to the numerous homologous isoforms of gliadins, the top-ranked associated protein ID for each peptide was selected in Proteome Discoverer via ranking all associated proteins according to the number of peptide sequences, the number of PSMs, protein scores, and the sequence coverage. Proteins were then manually grouped by gluten subunits. The protein isoforms with the highest number of identified peptides in alpha/beta gliadins (X2KVH9 and P18573) and gamma gliadins (A1EHE7, P21292, and P06659) were aligned and used to investigate N-terminal analysis coverage.

Table 2 indicates the list of native and TC peptides identified in digestion optimization samples. The TC peptides showed a Log2 H:L > 2, and native peptides showed a Log2 H:L close to 0. Native peptides with an internal tryptic cleavage site had a negative Log2 H:L, such as peptide N8 (VRVPVQLQPQNPSQQPQEQVPLVQQQF).

In theory, peptides with a tryptic N-termini and a chymotryptic C-termini (TC) or the native N-termini with a chymotryptic C-termini (native) that are 6 to 35 amino acid in length should be identified using this method. As illustrated in Fig. 7, all in-silico predicted TC and native peptides (indicated by underlining) in gamma gliadins were identified and only one TC peptide in alpha/beta gliadin was not found using the N-terminal labeling method. The number of total amino acids of all in-silico predicted peptides were counted for each type of gliadin (88 for alpha/beta gliadin and 114 for gamma gliadin), as well as the number for all peptides identified in the experimental N-terminal analysis (74 for alpha/beta gliadin and 114 for gamma gliadin). The coverage was evaluated by comparing the number of N-terminal identified amino acids to the in-silico predicted. In total, 84% (74 of 88) and 100% (114 of 114) of amino acid coverage was obtained for alpha/beta gliadins and gamma gliadins, respectively. In-silico predicted TC peptide NLALETLPAMCNVY in isoform P18573 was not found, likely because deamidation (Q) was included as a dynamic modification in the search, so peptide NLALETLPAMCNVY and NLALQTLPMAMCNVY would essentially be the same. Peptide NLALQTLPMAMCNVY was indeed identified using the method (TC 6 in Table 2) but was assigned to the other isoforms. Several tryptic cleavage sites could not be identified using the currently employed protease combination since the resulted peptides will be too long (> 35) or too short (< 6) for MS identification. In addition, fully tryptic peptides < 3 kDa would have been removed by the 3 kDa spin filter during buffer exchange. The ones with suitable size for MS analysis were identified by an additional fully tryptic search (Table S-6).

Although challenges still exist with detection and interpretation of isoform-level information, the matching of the in-silico predicted TC and native peptides with the experimentally identified peptides in the exemplar gliadins indicated the TAILS protocol could be modified and expanded for the analysis of gluten proteins. Multiple cleavage sites of a test protease in gluten protein mixtures and other co-existing hydrolysis events could be identified using the gluten-specific TAILS analysis protocol. This technique is worth further evaluation for analyzing the hydrolysis of gluten in the real food processes, such as gluten degradation patterns throughout brewing processes.

4. Conclusions

This study investigated various sample preparation practices for gluten N-terminal labeling MS analysis. The N-terminal labeling methodology, TAILS, was modified and validated for analysis of partially hydrolyzed gluten using trypsin as a test protease. Native N-termini and neo N-termini were unambiguously discriminated by their Log2 H:L ratios. Interference peptides, mainly fully chymotryptic peptides, with a specific Log2 H:L pattern (close to 1) were observed and can be decreased by using triple buffer exchange and a digestion enhancer in sample preparation, and setting an appropriate Log2 H:L threshold value. The validation indicated that the TAILS analysis could be modified and expanded to the identification of multiple protease cleavage sites in gluten proteins successfully. An estimated 80% of in-silico predicted peptides with a native or tryptic N-termini in exemplar alpha/beta gliadin and gamma gliadin protein sequences were identified. Although challenges remain when applying this method to more complex food matrices, this study indicated the N-terminal labeling MS analysis is worth further evaluation as a novel strategy for the analysis of natural hydrolysis of gluten in real food processes, such as fermentation. This strategy also could be further optimized, evaluated, and applied to the characterization of partially hydrolyzed proteins in allergenic foods more broadly and may provide reference for their safety assessment to both industry and regulatory authorities.

Associated content

The mass spectrometry proteomics data and search results from PD have been deposited to the PRIDE Archive in project PXD012296.

All results from Skyline have been copied to Panorama Public with the access URL (https://panoramaweb.org/N_terminal_gluten.url) and deposited to the ProteomeXchange in project PXD013679.

Funding sources

Research funding was provided by the Food Allergy Research & Resource Program (FARRP) at the University of Nebraska, a food industry-sponsored consortium of over 100 food processing companies and their suppliers. This research is also part of a collaboration between FARRP and Thermo Scientific.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jprot.2019.103538>.

References

- [1] J.A. Delcour, R.C. Hoseney, *Principles of Cereal Science and Technology*, Third ed., AACC International, Inc., St. Paul, Minnesota 55121, U.S.A., 2010.
- [2] H. Wieser, Chemistry of gluten proteins, *Food Microbiol.* 24 (2) (2007) 115–119.
- [3] A. Fasano, C. Catassi, Clinical practice. Celiac disease, *N. Engl. J. Med.* 367 (25) (2012) 2419–2426.
- [4] L. Shan, Ø. Molberg, I. Parrot, F. Hausch, F. Filiz, G.M. Gray, L.M. Sollid, C. Khosla, Structural basis for gluten intolerance in celiac sprue, *Science* 297 (5590) (2002) 2275–2279.
- [5] K.O. Bushara, Neurologic presentation of celiac disease, *Gastroenterology* 128 (2005) s92–s97.
- [6] AOAC, Gliadin as a Measure of Gluten in Rice- and Corn-Based Foods, <http://www.eoma.aoc.org/methods/info.asp?ID=49873>, (2012) (Accessed July 12th, 2018 2018).
- [7] AOAC, Gluten in Rice Flour and Rice Based Food Products, <http://www.eoma.aoc.org/methods/info.asp?ID=50740>, (2014) (Accessed July 12th, 2018 2018).
- [8] I.D.B. Slot, M. Bremer, I. van der Fels-Klerx, R.J. Hamer, Evaluating the performance of gluten ELISA test kits: the numbers do not tell the tale, *Cereal Chem.* 92 (5) (2015) 513–521.
- [9] R. Panda, H.F. Zoerb, C.Y. Cho, L.S. Jackson, E.A.E. Garber, Detection and quantification of gluten during the brewing and fermentation of beer using antibody-based technologies, *J. Food Prot.* 78 (6) (2015) 1167–1177.
- [10] B. Lexhaller, C. Tompos, K.A. Scherf, Comparative analysis of prolamins and glutelin fractions from wheat, rye, and barley with five sandwich ELISA test kits, *Anal.*

- Bioanal. Chem. 408 (2016) 6093–6104.
- [11] B. Lexhaller, C. Tompos, K.A. Scherf, Fundamental study on reactivities of gluten protein types from wheat, rye and barley with five sandwich ELISA test kits, *Food Chem.* 237 (2017) 320–330.
 - [12] S.P. White, C.D. Frisbie, K.D. Dorfman, Detection and sourcing of gluten in grain with multiple floating-gate transistor biosensors, *ACS Sens.* 3 (2018) 395–402.
 - [13] J.A. Sealey-Voyksner, C. Khosla, R.D. Voyksner, J.W. Jorgenson, Novel aspects of quantitation of immunogenic wheat gluten peptides by liquid chromatography-mass spectrometry/mass spectrometry, *J. Chromatogr. A* 1217 (25) (2010) 4167–4183.
 - [14] L. Uvackova, L. Skultety, S. Bekesova, S. McClain, M. Hajduch, The MS-proteomic analysis of gliadins and glutenins in wheat grain identifies and quantifies proteins associated with celiac disease and baker's asthma, *J. Proteome* 93 (2013) 65–73.
 - [15] K.L. Fiedler, S.C. McGrath, J.H. Callahan, M.M. Ross, Characterization of grain-specific peptide markers for the detection of gluten by mass spectrometry, *J. Agric. Food Chem.* 62 (25) (2014) 5835–5844.
 - [16] A. Manfredi, M. Mattarozzi, M. Giannetto, M. Careri, Multiplex liquid chromatography-tandem mass spectrometry for the detection of wheat, oat, barley and rye prolamins towards the assessment of gluten-free product safety, *Anal. Chim. Acta* 895 (2015) 62–70.
 - [17] M.J. Martinez-Esteso, J. Norgaard, M. Brohee, R. Haraszi, A. Maquet, G. O'Connor, Defining the wheat gluten peptide fingerprint via a discovery and targeted proteomics approach, *J. Proteome* 147 (16) (2016) 156–168.
 - [18] D. Zeltner, M.A. Glomb, D. Maede, Real-time PCR systems for the detection of the gluten-containing cereals wheat, spelt, kamut, rye, barley and oat, *Eur. Food Res. Technol.* 228 (2009) 321–330.
 - [19] J.R. Mujico, M. Lombardía, M.C. Mena, E. Méndez, J.P. Albar, A highly sensitive real-time PCR system for quantification of wheat contamination in gluten-free food for celiac patients, *Food Chem.* 128 (3) (2011) 795–801.
 - [20] C.Y. Cho, W. Nowatzke, K. Oliver, E.A. Garber, Multiplex detection of food allergens and gluten, *Anal. Bioanal. Chem.* 407 (14) (2015) 4195–4206.
 - [21] M.L. Colgrave, H. Goswami, M. Blundell, C.A. Howitt, G.J. Tanner, Using mass spectrometry to detect hydrolysed gluten in beer that is responsible for false negatives by ELISA, *J. Chromatogr. A* 1370 (2014) 105–114.
 - [22] W. Cao, D. Watson, M. Bakke, R. Panda, B. Bedford, P. Kande, L. Jackson, E.A.E. Garber, Detection of gluten during the fermentation process to produce soy sauce, *J. Food Prot.* 80 (5) (2017) 799–808.
 - [23] M.L. Colgrave, H. Goswami, C.A. Howitt, G.J. Tanner, What is in a beer? Proteomic characterization and relative quantification of hordein (gluten) in beer, *J. Proteome Res.* 11 (2012) 386–396.
 - [24] G. Picariello, G. Mamone, A. Cutignano, A. Fontana, L. Zurlo, F. Addeo, P. Ferranti, Proteomics, peptidomics, and immunogenic potential of wheat beer (weissbier), *J. Agric. Food Chem.* 63 (13) (2015) 3579–3586.
 - [25] H. Li, K. Byrne, R. Galiyev, O. Mendoza-Porras, U. Bose, C.A. Howitt, M.L. Colgrave, Using LC-MS to examine the fermented food products vinegar and soy sauce for the presence of gluten, *Food Chem.* 254 (2018) 302–308.
 - [26] K.L. Fiedler, R. Panda, T.R. Croley, Analysis of gluten in a wheat-gluten-incurred sorghum beer brewed in the presence of proline endopeptidase by LC/MS/MS, *Anal. Chem.* 90 (2018) 2111–2118.
 - [27] M.L. Colgrave, K. Byrne, M. Blundell, C.A. Howitt, Identification of barley-specific peptide markers that persist in processed foods and are capable of detecting barley contamination by LC-MS/MS, *J. Proteome* 147 (16) (2016) 169–176.
 - [28] G. Picariello, F. Bonomi, S. Lametti, P. Rasmussen, C. Pepe, S. Lilla, P. Ferranti, Proteomic and peptidomic characterisation of beer: immunological and technological implications, *Food Chem.* 124 (2011) 1718–1726.
 - [29] J.L. Hsu, S.Y. Huang, N.H. Chow, S.H. Chen, Stable-isotope dimethyl labeling for quantitative proteomics, *Anal. Chem.* 75 (24) (2003) 6843–6852.
 - [30] P.L. Ross, Y.N. Huang, J.N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlett-Jones, F. He, A. Jacobson, D.J. Pappin, Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents, *Mol. Cell. Proteomics* 3 (2004) 1154–1169.
 - [31] P.J. Boersema, R. Raaijmakers, S. Lemeer, S. Mohammed, A.J.R. Heck, Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics, *Nat. Protoc.* 4 (4) (2009) 484–494.
 - [32] O. Kleifeld, A. Doucet, U.a.d. Keller, A. Prudova, O. Schilling, R.K. Kainthan, A.E. Starr, L.J. Foster, J.N. Kizhakkedathu, C.M. Overall, Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products, *Nat. Biotechnol.* 21 (4) (2011) 228–237.
 - [33] O. Kleifeld, A. Doucet, A. Prudova, U.a.d. Keller, M. Gioia, J.N. Kizhakkedathu, C.M. Overall, Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates, *Nat. Protoc.* 6 (10) (2011) 1578.
 - [34] The Overall Lab, Protocols and SOPs, <http://clip.ubc.ca/resources/protocols-and-sops/>, (2018) (Accessed July 16th, 2018).
 - [35] E. Garcia, M. Llorente, A. Hernando, R. Kieffer, H. Wieser, E. Mendez, Development of a general procedure for complete extraction of gliadins for heat processed and unheated foods, *Eur. J. Gastroenterol. Hepatol.* 17 (2005) 529–539.
 - [36] A. Fallahbaghery, W. Zou, K. Byrne, C.A. Howitt, M.L. Colgrave, Comparison of gluten extraction protocols assessed by LC-MS/MS, *J. Agric. Food Chem.* 65 (2017) 2857–2866.
 - [37] J.R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis, *Nat. Methods* 6 (5) (2009) 359–362.
 - [38] B. MacLean, D.M. Tomazela, N. Shulman, M. Chambers, G.L. Finney, B. Frewen, R. Kem, D.L. Tabb, D.C. Liebler, M.J. MacCoss, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments, *Bioinform.* 26 (7) (2010) 966–968.
 - [39] D.J. McCarthy, G.M. Smyth, Testing significance relative to a fold-change threshold is a TREAT, *Bioinformatics* 25 (6) (2009) 765–771.

Supporting Information I

Validation of N-terminal Labeling Mass Spectrometry for Characterization of Partially Hydrolyzed Gluten Proteins

Table of Content

Figure S-1. Discovery analysis of commercial gliadin. (p. S-2)

Figure S-2. Gliadin isolation process from a white whole wheat flour. (p. S-3)

Figure S-3. Number of high confidence peptides before and after precipitation. (p. S-4)

Figure S-4. SDS-PAGE gel of commercial gliadins digested with various enzymes. (p. S-5)

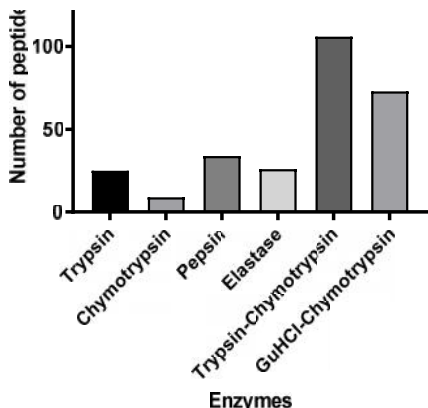
Figure S-5. The Log₂ H:L of peptides identified in labeling verification experiments and exemplar Skyline MS1 peak area of heavy/light labeled peptide pair comparison. (p. S-6)

Figure S-6. The Log₂ H:L and MS1 peak area of peptides from the flow-through investigation experiment. (p. S-7)

Table S-5. Reduction of false discovery rate when Log₂ H:L threshold value was included. (p. S-8)

Table S-6. In-silico predicted fully tryptic peptides in pre-selected gliadin isoforms. (p. S-9)

a.



b.

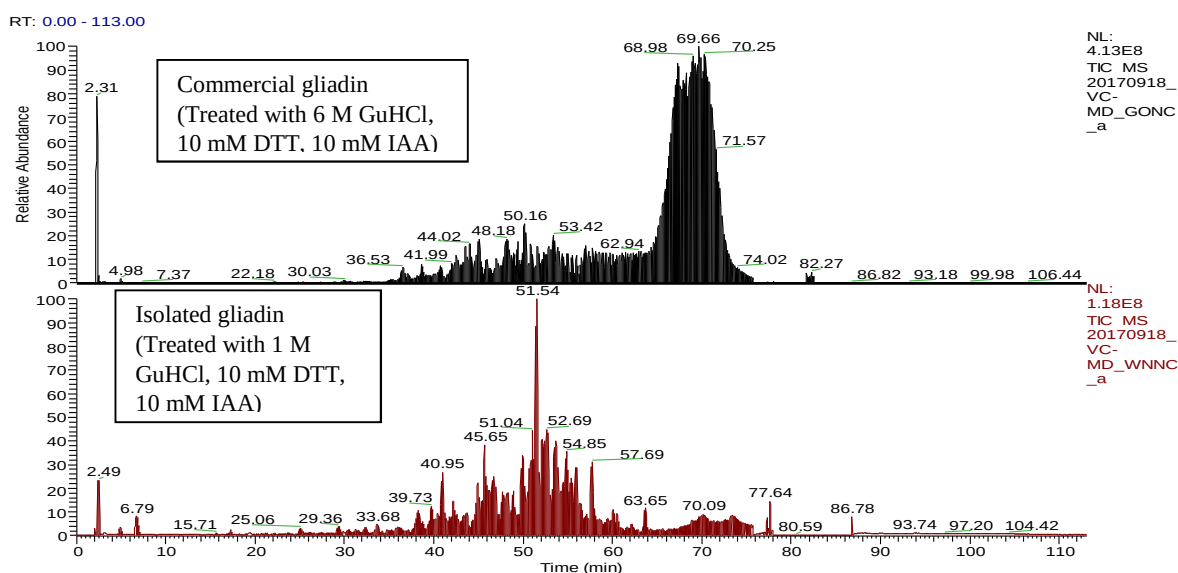


Figure S-1. (a) Number of high confidence peptides as identified in bottom-up discovery MS analysis of commercial gliadin digested with various enzymes (enzyme:substrate at 1:100 w/w and digestion conducted in manufacture recommended conditions). All SEQUEST HT search parameters are the same as stated in manuscript section 2.7 for discovery analysis except for the enzymatic cleavage sites (trypsin: KR/not P, chymotrypsin: FWY/not P, pepsin (pH = 1.2): FL, elastase: AVSGLI). Note that only high affinity cleavage sites for chymotrypsin and pepsin were included in the search. A small number of gliadin peptides were identified regardless of the protease used. The combination of trypsin and chymotrypsin or the incorporation of high concentration of strong chaotropic reagent (6 M GuHCl) could increase the amount of identified peptides. (b) Comparison of total ion chromatograms (TICs) of digested commercial gliadin and digested isolated gliadin. Both samples were denatured, reduced and alkylated before a chymotrypsin digestion (1:100 w/w) and the 6 M GuHCl for denaturing commercial gliadin was diluted to 1 M before digestion. The unusual TICs observed with commercial gliadin digests indicate that enzymatic digestion and/or LC-MS/MS performance were negatively affected by some property of the commercial gliadin.

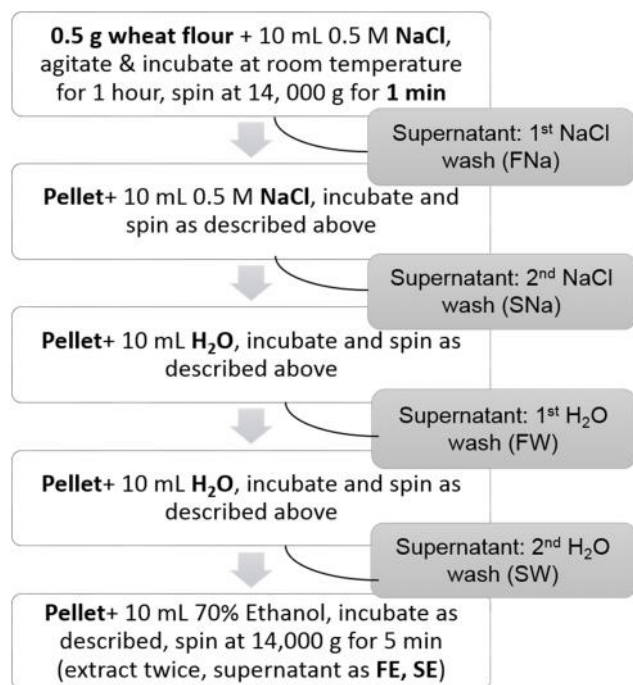


Figure S-2. Gliadin isolation process from a white whole wheat flour.

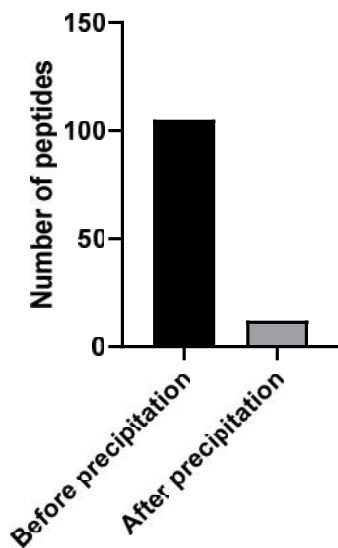


Figure S-3. Number of high confidence peptides as identified in fully (trypsin-chymotrypsin) digested commercial gliadin proteins before and after acetone/methanol precipitation. All samples were labeled with 40 mM light formaldehyde in the presence of 20 mM of (NaBH₃CN) and desalted with C18 spin columns before LC-MS/MS analysis (no negative selection with HPG-ALD polymers).

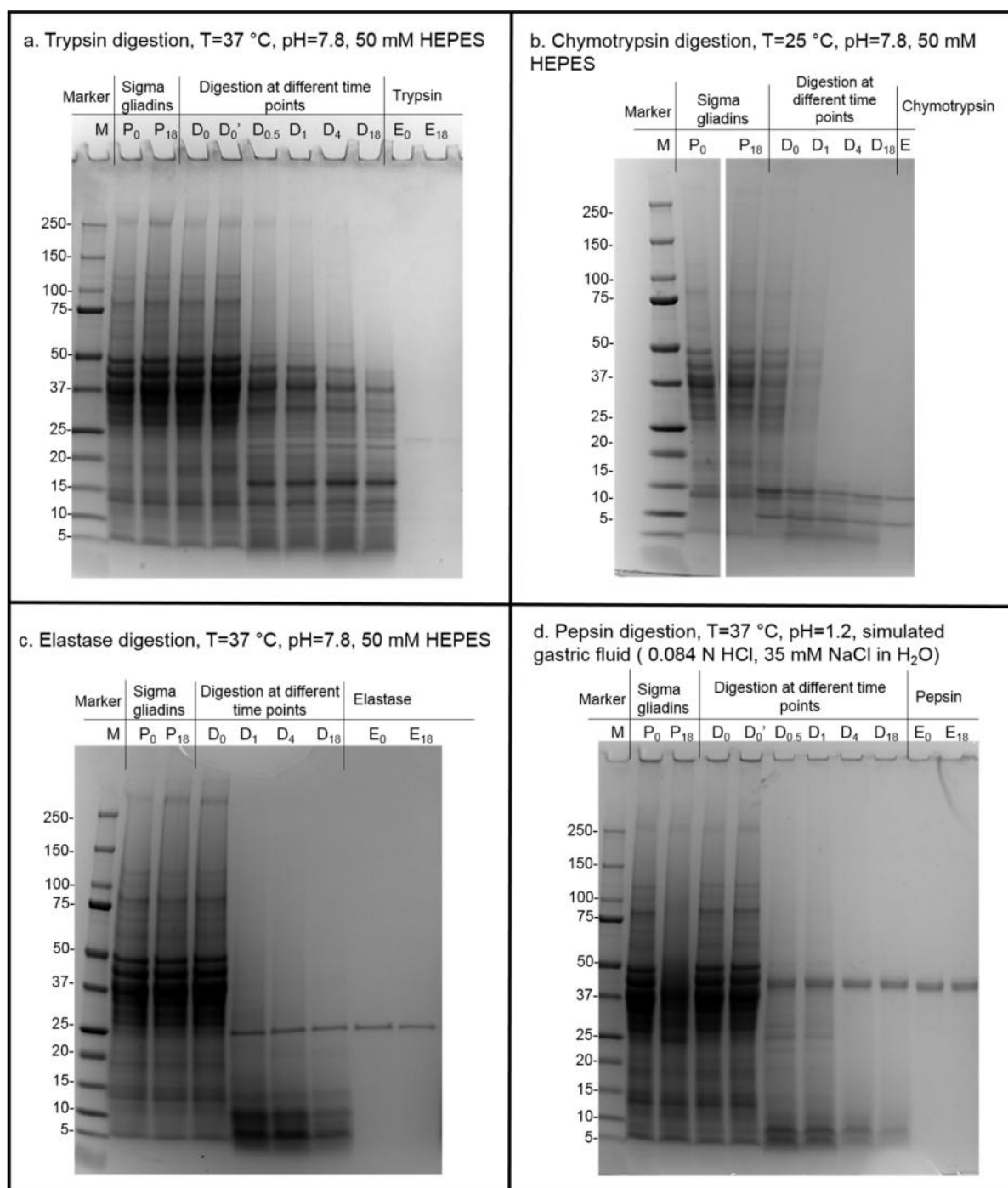


Figure S-4. SDS-PAGE gel of commercial gliadins digested with various enzymes (enzyme:substrate at 1:100 w/w): a. trypsin, b. chymotrypsin, c. elastase, and d. pepsin. M: Marker; P₀ and P₁₈: undigested gliadin at 0 and 18 hour; D₀, D_{0.5}, D₁, D₄, and D₁₈: gliadin incubated with enzymes at 0, 0.5, 1, 4, and 18 hour; E₀ and E₁₈: enzymes at 0 and 18 hour; D₀': undigested gliadin with heat denatured enzyme. All gliadin samples were prepared at 5 µg/µL and loaded 10 µL per lane.

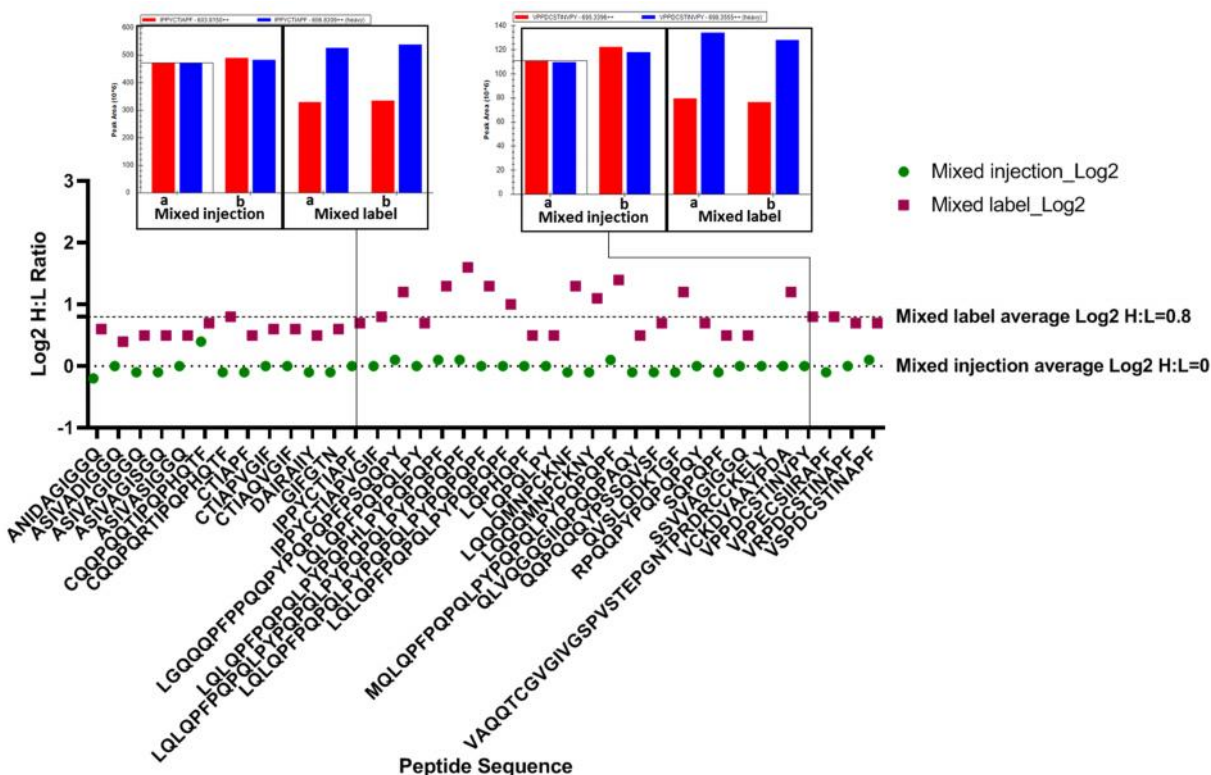


Figure S-5. The Log2 H:L ratio of peptides found in the labeling verification experiments and exemplar Skyline MS1 peak area of heavy labeled (blue) and light labeled (red) peptide pair comparison. Mixed injection indicates peptides labeled with equal amount of light and heavy formaldehyde separately but mixed before injection, which was the data from labeling verification, and no difference between the heavy and light labeled peptide pair were observed. Mixed label indicates peptides labeled with equally mixed light and heavy formaldehyde and the unique pattern (Log2 H:L close to 1) was found. The experiment was conducted in duplicated injection.

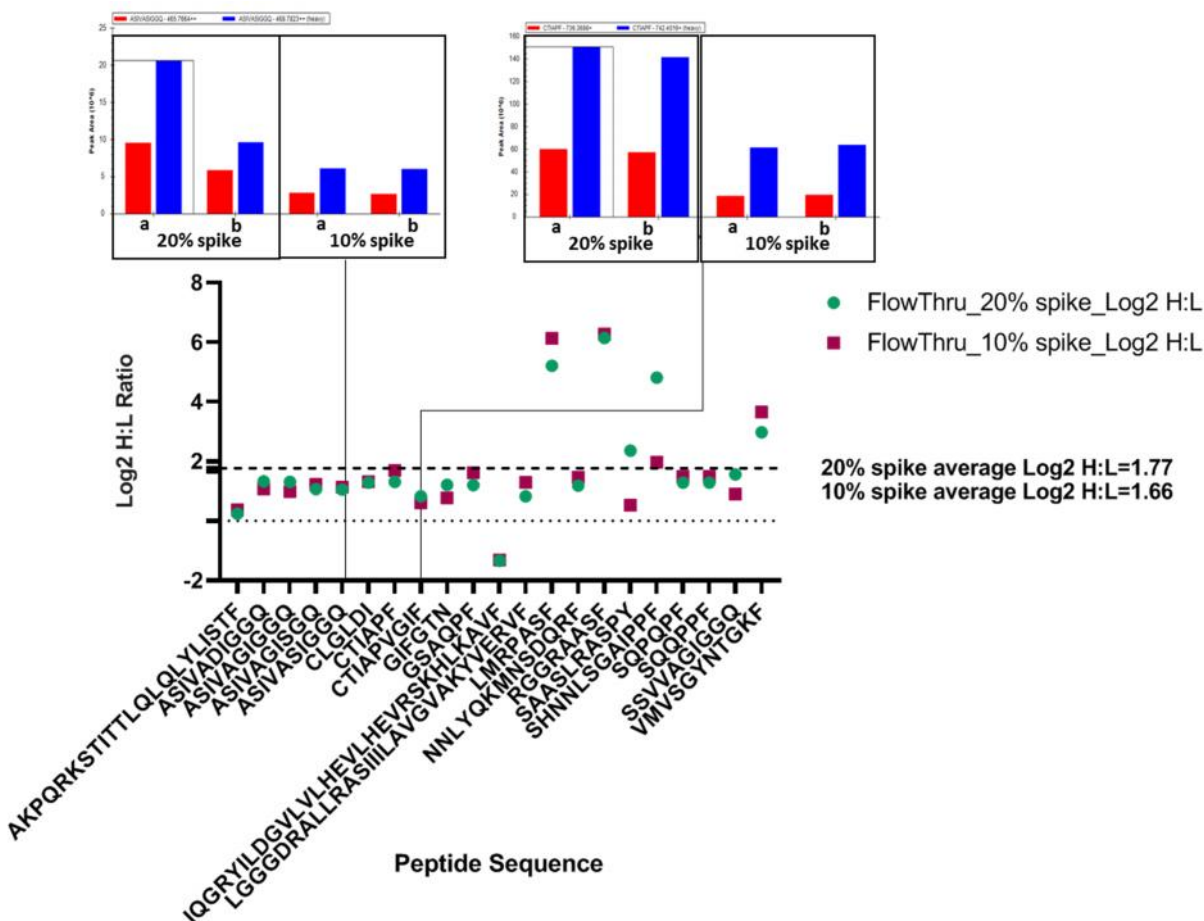


Figure S-6. The Log2 H:L ratio and exemplar MS1 peak area of peptides from the flow-through investigation experiment. Flow-through from buffer exchange was collected and spiked into a gluten discovery analysis sample (20% v/v, 10% v/v) before chymotrypsin digestion. Samples were then digested with chymotrypsin, underwent negative HPG-ALD polymer selection, and desalted as per the gluten N-terminal analysis protocol. Chymotrypsin cleaved peptides can be labeled with left-over light (red) and heavy (blue) formaldehyde present in the flow-through at both 20% and 10% spike. The MS1 peak area of heavy labeled peptides were mostly higher than the light labeled counterpart.

Table S-5. Reduction of false discovery rate when Log2 H:L threshold value was included in the initial gluten N-terminal labeling analysis.

	Peptide type	Original	Possible Log2 H:L threshold value		
			1	2	3
Number of peptide	TC	13	13	13	12
	Chymotryptic	31	23	3	2
	Non-specific	14	12	5	4

ID	Protein	Peptide sequence	AA length	Mw (average) (Da)	Initial	Quench	Digestion Optimization
1	Alpha/beta gliadin	R.QQQK	6	530.58	N ^{&}	N	N
2	Alpha/beta gliadin	K.QQQQQQQQILQQILQQQLI PCR*	24	2719.11	N	N	N
3	Gamma gliadin	R.QPFPQQPQQPYQPQQPQQP FPQTQQPQQPFPQSK*	35	3929.32	H [#]	H	H
4	Gamma gliadin	K.QPQQPFPQPQQPQQSFPQ QQPQLIQQSLQQQLNPCK	38	4198.68	N	N	N

Table S-6. In-silico predicted fully tryptic peptides in pre-selected gliadin isoforms

[&]Peptide is not identified in the experiment.

[#]Heavy dimethyl labeled peptide is identified in the experiment.

*Peptide length is suitable for MS analysis in this study.