Waters™

Application Note

A Label-free, Data-independent Approach to Defining the Gluten Proteome

Sophie Bromilow, Lee A. Gethings, Peter Shewry, E.N. Clare Mills

Respiratory and Allergy Centre, Institute of Inflammation and Repair, Manchester Academic Health Science Centre, Manchester Institute of Biotechnology, University of Manchester, Waters Corporation, Rothamsted Research

Abstract

This study highlights improved gluten extraction efficiency using the acid labile detergent *Rapi*Gest followed by identification of the celiac-toxic motif regions using a label-free LC-HDMS^E (LC-DIA-IM-MS) approach.

The extraction efficiency of gluten proteins has shown to improve with buffers containing *Rapi*Gest when compared with other extraction buffers. Subsequent analysis with SYNAPT G2-*Si* provided 25% more protein identifications when using HDMS^E compared with a competitor platform, which utilized a DDA strategy. The number of unique peptides identified was also 10-fold higher for HDMS^E data, providing a greater number of glutenin and associated celiac-toxic motifs. Qualitative and quantitative information is obtained within a single acquisition. Label-free quantitation varies depending on the sub-group of gluten proteins.

Benefits

Enhanced protein extraction of gluten from wheat grain is achieved through the use of extraction buffers that contain *Rapi*Gest, which reproducibly increases the contents of gliadins and glutenins. Chymotrypsin digestion followed by data-independent acquisition with ion mobility mass spectrometry (DIA-IM-MS) provides a higher number of protein identifications with greater sequence coverage when compared to more traditional data-

dependent approaches. In particular, higher sequence coverage of the glutenins allows a greater frequency of celiac-toxic motifs to be identified, which are quantified using a label-free approach with Progenesis QI for Proteomics.

Introduction

Wheat is the most widely grown food crop worldwide and forms a staple part of the modern diet.[1] However, it poses a health risk to a small proportion of the population because it contains gluten – a protein fraction able to trigger celiac disease (CD).[2] Celiac disease is a non-IgE immune mediated adverse reaction to gluten that affects about 70 million people globally.[3] Gluten is formed from the major storage proteins of wheat grain and comprises two groups of proteins, which are classified based on their solubility in aqueous alcohol mixtures. The gliadins are alcohol-soluble while the glutenins are alcohol-insoluble, requiring reduction for solubilization due to their presence as high molecular mass aggregates stabilized by inter-chain disulphide bonds.[4,5] The gliadin proteins can be further sub-grouped into α-, γ-, and ω-gliadins, which are differentiated by both the repetitive peptide motifs present in their sequences and in the number of cysteine residues. The glutenin proteins can be further classified into two groups: high molecular weight (HMW) and low molecular weight (LMW) subunits of glutenin, based on molecular weight. The different solubility properties of the gliadin and glutenin proteins make their simultaneous extraction, detection, and quantification problematic. Gluten proteins contain few, if any, lysine and arginine residues, resulting in poor digestion by trypsin. Therefore, an alternative protease – such as chymotrypsin – must be used to undertake the digestion step in proteomic workflows. The resulting digests comprise longer peptides, and their identification by means of mass spectrometry can be challenging. This study highlights improved gluten extraction efficiency using the acid labile detergent *Rapi*Gest followed by identification of the celiac toxic motif regions using a label-free LC-HDMS^E (LC-DIA-IM-MS) approach.

Experimental

Sample preparation

Gluten proteins were extracted individually from six wheat seeds (cv Hereward) by sonication for 15 minutes at

60 °C using five different extraction solutions (Table 1). These solutions ranged from simple reducing buffers, to solutions containing *Rapi*Gest, and to a complex chaotropic solution – previously used for extraction of grain proteins prior to SDS-PAGE. All five solutions included the buffer recommended by the Codex Alimentarius Commission extraction.[5] After centrifugation (10,000 xg for 10 minutes), the supernatant was removed and stored at -20 °C. The protein concentrations of extracts were determined using the 2-D Quant Kit (GE Healthcare, UK). Samples prepared using extraction solution 3 were reduced, alkylated, and digested overnight using chymotrypsin prior to LC-MS analysis.

Extraction protocol	Extraction solution
1	7 M urea, 2 M Thiourea, 2% CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate, w/v)
2	50 mM Tris-HCl (pH 8.8), 50 mM Dithiotreitol (DTT)
3	50 mM Tris-HCl (pH 8.8), 50 mM DTT, *Rapi*Gest (0.2% w/v)
4	50 mM Tris-HCl (pH 8.8), 50 mM DTT, *Rapi*Gest (0.2% w/v), 75% Ethanol (v/v)
5	Two-step extraction, 50 mM Tris-HCl (pH 8.8), 50 mM DTT, *Rapi*Gest (0.2% w/v), with the pellet resuspended using 50 mM Tris-HCl (pH 8.8), 50 mM DTT, *Rapi*Gest (0.2% w/v), 75% Ethanol (v/v)
6	Ethanol (40% v/v)

Table 1. Summary of the extraction methods compared.

LC conditions

LC system:	M-Class ACQUITY UPLC
Columns:	ACQUITY UPLC M-Class Symmetry C_{18} Trap, 5 µm, 180 µm x 20 mm (P/N 186007497) ACQUITY UPLC M-Class HSS T3, 1.8 µm, 75 µm x 150 mm (P/N 186007473)
Column temp.:	35 °C
Flow rate:	300 nL/min
Mobile phase A:	water (0.1% formic acid)
Mobile phase B:	acetonitrile (0.1% formic acid)
Gradient:	3–40% B in 90 min
Injection volume:	1 µL

MS conditions

MS system:	SYNAPT G2-*Si*
Ionization mode:	ESI+ at 3.2 kV
Cone voltage:	30 V
Acquisition mode:	HDMS^E 50–2000 *m/z* both functions (low and elevated energy)

Acquisition rate:	Low and elevated energy functions at 0.5 s
Collision energy:	5 eV (low energy function) and from 19–45 eV (elevated energy function)
Resolution:	25,000 full width at half maximum (FWHM)
IMS T-wave velocity:	700 m/s
IMS T-wave pulse height:	40 V

Data management

Progenesis QI for Proteomics

GraphPad Prism

Bioinformatics

The LC-MS peptide data was processed and searched with Progenesis QI using a curated gluten database. Progenesis enabled normalized label-free quantification to be achieved along with peptide/protein identifications.

Results and Discussion

Previously reported single-step extraction methods favor the extraction of either water-soluble or water-insoluble proteins.[6] The glutenins are often underrepresented using these methods, with many studies focusing only on the gliadins.[7] As a result, extraction methods were assessed initially to identify buffers capable of extracting both gliadin and glutenin polypeptides in a representative fashion. Analysis of five different gluten extracts of grain (Figure 1) showed that buffers containing *Rapi*Gest (extractions 3, 4, and 5 versus 2) significantly increased protein extraction compared to buffers without *Rapi*Gest. The addition of *Rapi*Gest with heating disrupts protein structure, thereby increasing the solubility. The addition of ethanol in extraction buffer 4 – a classical solvent

used for solubilizing the gliadin fraction of gluten – slightly reduces protein recovery when compared to extraction 3. This is likely due to the ethanol affecting the critical micelle concentration of the *Rapi*Gest. Sequential extraction using buffer 3 followed by buffer 4 (buffer 5) improved protein recovery. However, the two-step protocol compromised the reproducibility of the extraction with a high level of variance.

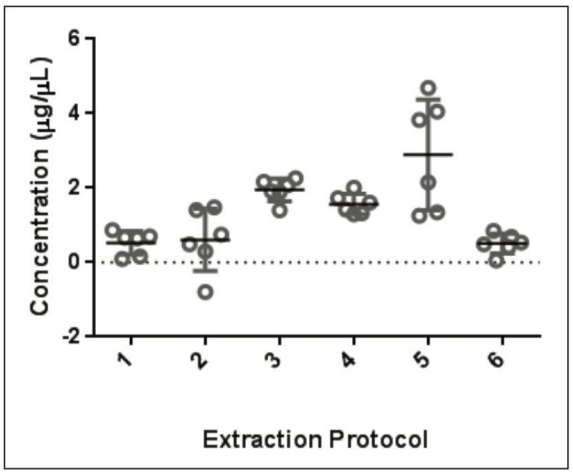

Figure 1. A comparison of protein recovery across a number of extractions, including a typical ELISA-kit extraction method that is based on the Codex Alimentarius Commission recommendations. Investigating three extractions containing RapiGest compared to the codex recommended extraction. Protein recovery was calculated (protein extracted (mg)/total protein present (mg; determined by Kjeldahl analysis)) x 100. See Table 1 for extraction protocol details.

Mass spectrometry was conducted using two different platforms for comparative purposes, focusing on grain extracted with buffer 3 due to its good protein recovery and reproducibility. SYNAPT was used in conjunction with the DIA-IM-MS workflow, while data-dependent acquisitions (DDA) were performed using a competitive platform. In both cases, identical inlets and column chemistries were used. Comparisons were made on the basis of equal sample loadings. Reviewing both datasets showed that while many peptides and proteins were

identified on both platforms (Figure 2), SYNAPT was able to detect 10-fold more unique peptides than the competitor instrument. Additional proteins unique to each platform were also identified, with 50 assigned to SYNAPT and 9 to the competitor instrument. From a protein group perspective, comparable protein identifications can be assigned to the gliadins, while the majority of unique SYNAPT identifications are attributed to the glutenins (Figure 3).

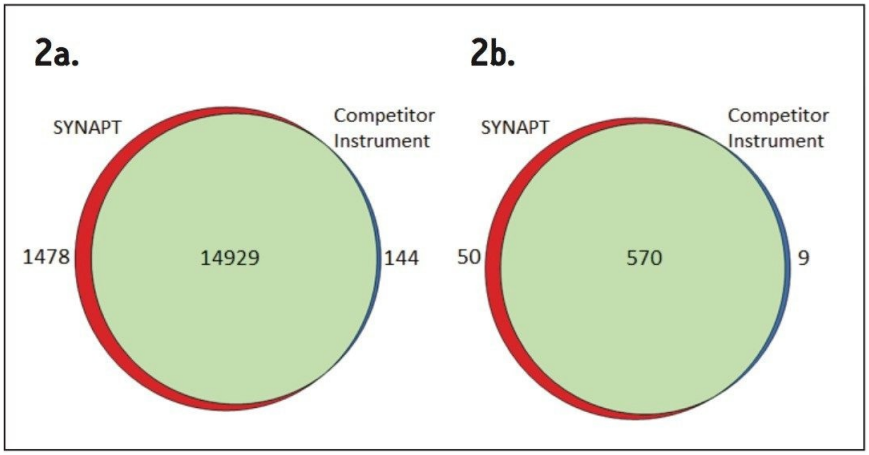

Figure 2. (a) Venn diagram representing the total number of peptides identified across the two platforms; (b) Venn diagram representing the unique and shared proteins across both platforms colored as before. Red: Peptides/proteins uniquely identified by the SYNAPT G2-Si System. Blue: Peptides/proteins uniquely identified by the competitor system. Green: Peptides/proteins identified by both platforms.

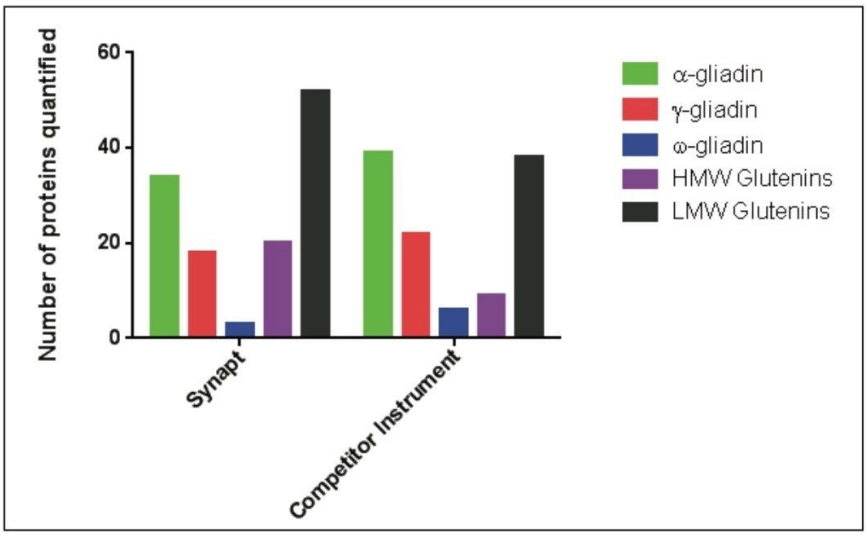

Figure 3. Distribution of gliadin and glutenin proteins identified and quantified
from a RapiGest-based extract. Data was collected using HDMS^E (SYNAPT G2 Si) and on a data-dependent acquisition competitive instrument.

While chymotrypsin is more effective at digesting gluten proteins than trypsin, it was predicted to result in the formation of longer peptides. These longer peptides were readily identified using SYNAPT. 79% of the longer peptides consisted of between 10–24 residues (Figure 4b). SYNAPT identified 8% of the peptides with greater than 24 residues, compared to only 5% identified on the competitor instrument. The majority (87%) of the peptides identified using the competitor instrument contained <19 residues. On average, peptides identified with SYNAPT consisted of 16 residues (Figure 4a). The LMW-glutenin subunit encoded by the UniProt sequence A9YSH4 is a representative example of a gluten protein identified in this way. Figure 5 shows 20% greater sequence coverage across the repetitive domain where the physiologically-relevant celiac-toxic motifs are located. The longest peptide present in the A9YSH4 protein – identified using SYNAPT – contained 35 residues, while the competitor instrument identified a corresponding peptide that contained only 22 residues. Often, the smaller peptides identified on the competitor instrument form part of the longer peptide identified with SYNAPT. This suggests SYNAPT is capable of detecting large peptides that contain a number of miscleavages.

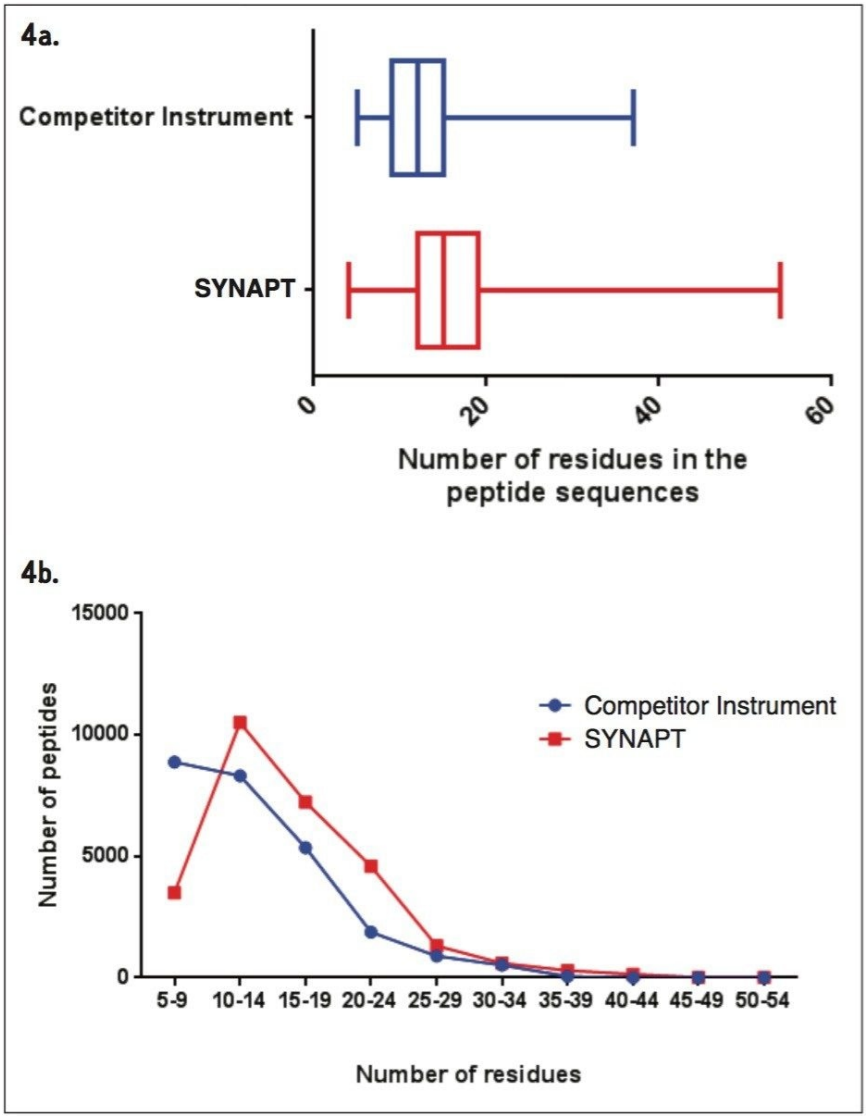

Figure 4. (a) Box and whisker plot distribution of peptide length. (b) Line graph highlighting variation distribution of the number of peptides containing a set number of residues for SYNAPT (red) and competitor platform (blue).

A detailed review has defined 23 celiac-toxic motifs in wheat gliadins and glutenins.[8] A selection of the celiac-toxic motifs was searched using data from both platforms to determine the number of times that individual motifs occurred in all observed peptides. Higher occurrences of five out of six celiac-toxic motifs in the SYNAPT data were observed (Figure 6). This higher rate of identification for celiac-toxic motifs with SYNAPT can be

attributed to the identification of high molecular weight peptides as shown in Figure 5.

Figure 5. LMW-glutenin protein (A9YSH4) highlighting the sequence coverage for both MS platforms. Common peptides (red) observed on both platforms and unique peptides (blue) are represented. Underlined are the clinically relevant celiac-toxic motifs as defined by the Goodman database.[9]

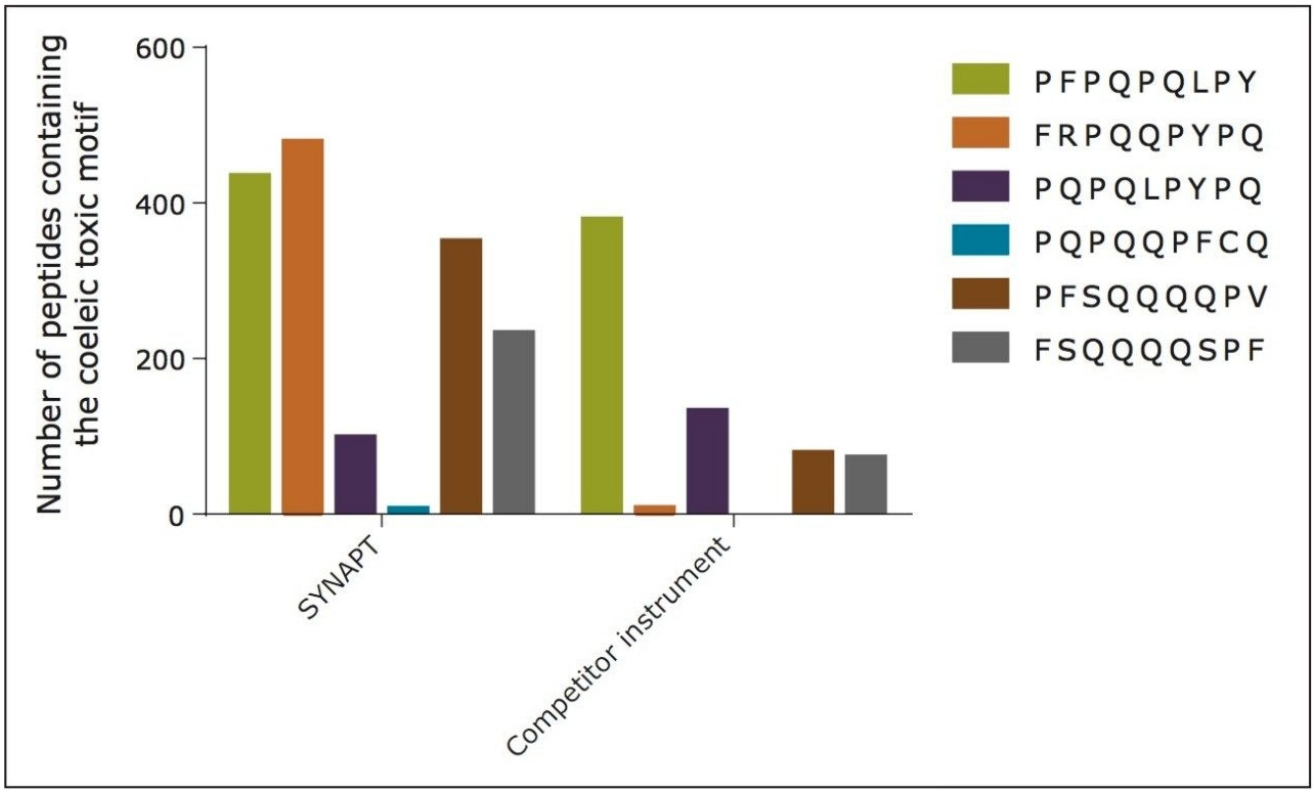

Figure 6. Distribution of peptides identified by both platforms as containing recognized celiac-toxic motifs.[1]

An analysis of the quantitative distribution – using top3 quantification on the various protein groups based on the *Rapi*Gest extraction protocol – shows γ-gliadins to be the most abundant, while LMW glutenins and ω-gliadins are of lowest abundance.

The α-gliadins appear to be present across the entire dynamic range (Figure 7). The quantitation curve is representative of the single extraction system investigated, and the distribution and abundance of the gluten protein groups would be expected to differ depending on the extraction conditions.

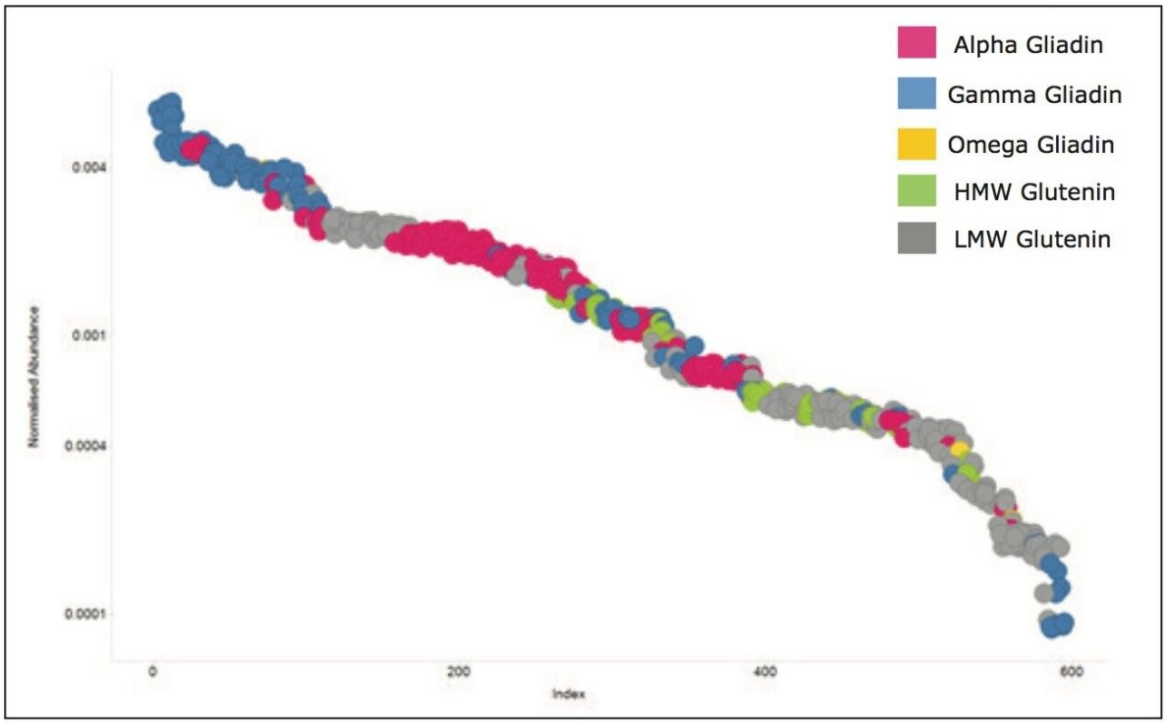

Figure 7. Quantitation curve constructed from normalized relative abundances of gluten proteins, identified using HDMS^E for an extract containing RapiGest. Individual protein identifications are colored on the basis of the five protein sub-groups.

Conclusion

The extraction efficiency of gluten proteins has shown to improve with buffers containing *Rapi*Gest when compared with other extraction buffers – includingthe industry standard, based on the Codex Alimentarius Commission recomme ndation. Subsequent analysis with SYNAPT G2-S*i* provided 25% more protein identifications when using HDMS^E compared with a competitor platform, which utilized a DDA strategy. The number of unique peptides identified was also 10-fold higher for HDMS^E data, providing a greater number of glutenin and associated celiac-toxic motifs. Qualitative andquantitative information is obt ained within a single acquisition. Label-free quantitation varies depending on the sub-group of gluten proteins. For example, peptides derived from γ-gliadins and LMW-glutenins are the most and least abundant respectively.

References

1. Shewry, P.R. and Tatham, A.S. Improving Wheat to Remove Celiac Epitopes But Retain Functionality. *Journal of Cereal Science*. 2016. 67:12–21.
2. Colgrave, M.L., et al. Proteomic Profiling of 16 Cereal Grains and the Application of Targeted Proteomics to Detect Wheat Contamination. *Journal of Proteome Research.* 2015. 14(6):2659–2668.
3. Lebwohl B., L.J.F., and Green P.H. Celiac Disease and Nonceliac Gluten Sensitivity. *British Medical Journal*. 2015.
4. Shewry, P.R., et al. CHAPTER 8: Wheat Grain Proteins, in WHEAT: Chemistry and Technology. *AACC International*, Inc. 2009. 223–298.
5. Wieser, H. Chemistry of Gluten Proteins. *Food Microbiology.* 2007. 24(2):115–119.
6. Haraszi, R., et al. Analytical Methods for Detection of Gluten in Food-Method Developments in Support of Food Labeling Legislation. *Journal of AOAC International.* 2011. 94(4):1006–1025.
7. Fiedler, K.L., et al. Characterization of Grain-Specific Peptide Markers for the Detection of Gluten by Mass Spectrometry. *Journal of Agricultural and Food Chemistry*. 2014. 62(25):5835–5844.
8. Sollid, L., et al. Nomenclature and Listing of Celiac Disease Relevant Gluten T-cell Epitopes Restricted by HLA-DQ Molecules. *Immunogenetics*. 2012. 64(6):455–460.
9. Goodman, R.E. AllergenOnline. Home of the FARRP allergen protein database. Available from: http://www.allergenonline.org/.

Featured Products

SYNAPT G2-Si Mass Spectrometry <https://www.waters.com/134740653>

ACQUITY UPLC M-Class System <https://www.waters.com/134776759>

Progenesis QI for Proteomics <https://www.waters.com/134790665>

Available for Purchase Online

ACQUITY UPLC M-Class Symmetry C18 Trap Column, 100Å, 5 µm, 180 µm x 20 mm, 2D, V/M, 1/pkg <
https://www.waters.com/waters/partDetail.htm?partNumber=186007497>

ACQUITY UPLC M-Class HSS T3 Column, 1.8 µm, 75 µm X 150 mm, 1/pkg <
https://www.waters.com/waters/partDetail.htm?partNumber=186007473>

720005672, April 2016