PROGRAMA DE PÓS-GRADUAÇÃO EM ALIMENTOS E NUTRIÇÃO

CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE

UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO

Thais de Oliveira Alves

DETERMINAÇÃO DE PEPTÍDEOS E PROTEÍNAS ASSOCIADOS À DOENÇA CELÍACA E ALERGIAS EM FARINHAS DE TRIGO POR UPLC-MS^{E}

Rio de Janeiro

2018

Thais de Oliveira Alves

DETERMINAÇÃO DE PEPTÍDEOS E PROTEÍNAS ASSOCIADOS À DOENÇA CELÍACA E ALERGIAS EM FARINHAS DE TRIGO POR UPLC-MS^{E}

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Alimentos e Nutrição, da Universidade Federal do Estado do Rio de Janeiro como requisito parcial para obtenção do título de Mestre em Ciência de Alimentos.

Orientadora: Profa. Dra. Mariana S. L. Ferreira

Rio de Janeiro

2018

Thais de Oliveira Alves

DETERMINAÇÃO DE PEPTÍDEOS E PROTEÍNAS ASSOCIADOS À DOENÇA CELÍACA E ALERGIAS EM FARINHAS DE TRIGO POR UPLC-MSE

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Alimentos e Nutrição da Universidade Federal do Estado do Rio de Janeiro

Aprovado em: _____/_____/_____

BANCA EXAMINADORA

_____

Dra. Mariana Simões Larraz Ferreira
Universidade Federal do Estado do Rio de Janeiro – UNIRIO

_____

Dra. Maria Gabriela Bello Koblitz
Universidade Federal do Estado do Rio de Janeiro – UNIRIO

_____

Dr. Fábio César Souza Nogueira
Universidade Federal do Rio de Janeiro – UFRJ

Dedico este trabalho a Deus, aos meus pais e a todos que, de alguma forma, contribuíram para a minha caminhada de sucesso até esse momento.

AGRADECIMENTOS

Agradeço à CAPES pela bolsa concedida para execução do Mestrado.

À FAPERJ e ao CNPq por financiarem a pesquisa.

Ao Programa de Pós-Graduação em Alimentos e Nutrição por tudo que me foi oferecido durante esses dois anos, pela oportunidade de participar de eventos científicos apresentando meu trabalho.

Ao prof. Dr. Luiz Carlos Gutkoski e à empresa OR Melhoramento de Sementes Ltda. pelo fornecimento das amostras.

À minha orientadora mais linda e inteligente, profª. Dra. Mariana Simões Larraz Ferreira, por toda ajuda, paciência e conselhos durante essa trajetória. Muito obrigada por acreditar em mim, por me estimular e mostrar que eu era capaz, nos momentos em que eu pensei em desistir.

À profª. Dra. Édira Castello Branco de Andrade Gonçalves, coordenadora do PPGAN, por toda a confiança depositada em mim, desde a graduação, e por ser meu exemplo e uma das maiores incentivadoras a seguir o caminho da Docência.

À profª. Dra. Maria Gabriela Bello Koblitz pelos sorrisos encantadores pelos corredores da Universidade, por toda contribuição ao trabalho e por todo conhecimento compartilhado.

Ao prof. Dr. Fábio César Souza Nogueira, por aceitar fazer parte da banca avaliadora dessa dissertação.

Ao prof. Dr. Paulo Sergio Marcellini, meu primeiro (e melhor) orientador de Iniciação Científica, por me ensinar a admirar a ciência e por me direcionar, ao final da graduação, ao caminho trilhado nesses dois anos. Essa foi a melhor escolha que eu jamais poderia ter feito, se não fosse pela sua ajuda!

Ao Dr. Gustavo Henrique Martins Ferreira de Souza por todo auxílio prestado na execução deste trabalho.

Ao Dr. Andrés Rodríguez Vega pelas conversas, piadas e por estar sempre pronto a ajudar a resolver as minhas dificuldades com o Excel. *Muchas Gracias*!

A todos os membros do *Team Gluten*. Os "raiz", Carolina Thomaz, Diego Kalili, Millena Santos, Sónia Rodrigues e Verônica Mayrinck, por todas as horas que passamos pesquisando, sofrendo juntos, repetindo análises, perdendo feriados de Sol, pelas festinhas surpresa, pelos momentos em que vocês foram minha fortaleza, pois era aqui que eu conseguia me distrair um pouco e dar algumas risadas, e, principalmente, pela amizade que nós criamos. Espero não perdê-los nunca! E aos membros "Nutella", Amanda Moutella, Fabiana Ramos, Luciana

Ribeiro, Oscar Vidal, Talita Braga, Talita Pimenta e Tatiana Medina, pela amizade e por se integrarem tão bem ao nosso time.

Ao projeto Unidança e todas as pessoas maravilhosas que conheci ao fazer parte do mesmo, por me darem uma nova família linda chamada Ohana, pela oportunidade de aprender a dançar forró e arriscar alguns outros ritmos, e por ser um refúgio nos meus maiores momentos de stress.

Às minhas mais novas, mas não menos queridas, melhores amigas, Camilla Dale e Mariana Novaes, por todas as aventuras, festas, muretas, conversas, conselhos e puxões de orelha.

Aos meus afilhados de casamento, Andressa e Thiago Santos, por me mostrarem o quanto eu era amada e que eu jamais estaria só. Eu realmente não sei o que seria de mim sem a amizade e o amor de vocês no momento mais crítico da minha vida, e por me darem um dos meus presentes mais lindos, nossa princesa Ana Luiza.

Às amigas Maria Clara Nunes e Mayara Pereira, por serem meu "núcleo cômico", sempre me colocando pra cima e me fazendo rir.

Ao meu querido amigo Vinicius Gonçalves, por me mostrar que em tão pouco tempo é possível construir uma relação de tamanha confiança, e por ser meu psicólogo nas horas vagas, sempre disposto a ouvir meus choros e reclamações.

A todos os meus amigos, que tanto torceram por mim neste período.

À minha amada família por todo amor, carinho e confiança.

Aos meus pais, Francisco José Alves e Jane de Oliveira Alves, por serem meu porto seguro e por terem dedicado uma vida inteira a mim. Todo meu esforço até aqui é baseado na luta de vocês, e espero um dia poder retribuir um pouco do que fizeram e ainda fazem pelo meu bem-estar. Obrigada por me amarem de forma incondicional e por serem os melhores pais que alguém poderia ter.

Ao meu padrinho José Roberto Alves por todo auxílio durante a minha vida acadêmica. Se não fosse por você, eu jamais teria chegado aqui.

À memória da minha madrinha Cidineia Suzano Alves, por tanto me amar e ser a verdadeira "segunda mãe" que eu tive.

À memória da minha avó Josefa Porcina de Oliveira, pelos deliciosos brigadeiros, por todas as bênçãos que me dava ao sair de casa diariamente, e pelo alívio que sentia ao me ver chegar em segurança.

Aos meus afilhados João Eduardo e Pedro Jaime, por me ensinarem o que é o amor de verdade e por encherem a "Didi" de orgulho a cada nova etapa das suas vidas.

À minha irmã de leite, Pamela Freitas, pelas orações, por interceder junto a Jesus tantas vezes em meu favor, e por ser minha companheira desde sempre.
Às minhas cadelinhas Alice e Petit, por todo calor e afago nos momentos mais difíceis.
Acima de tudo, agradeço a Deus. Se não fosse pela Sua vontade, nada do que está sendo concluído hoje teria acontecido.

Muito obrigada.
E aqui vamos nós, rumo à próxima etapa. Que venha o Doutorado.
T.

“Aprender é a única coisa que a mente nunca se cansa, nunca tem medo e nunca se arrepende”

(Leonardo da Vinci)

RESUMO

O trigo é um cereal de grande importância econômica, nutricional e tecnológica, devido à presença das proteínas do glúten. O glúten é composto por prolaminas álcool-solúveis e gluteninas álcool-insolúveis, que conferem viscoelasticidade à massa. A qualidade tecnológica do trigo é determinada pelas propriedades reológicas da farinha e define sua aplicação na indústria. Diversas doenças são relacionadas ao consumo de glúten, podendo ser autoimunes, como a doença celíaca (DC), que se manifesta no intestino delgado; alergias mediadas pela IgE e ainda sensibilidades. Outras doenças, como alergias respiratórias, estão associadas às proteínas da fração solúvel (não-glúten), albuminas e globulinas (AG). A detecção e quantificação das proteínas do trigo são de extrema importância, pois o tratamento para a maioria dessas doenças é a exclusão da dieta de alimentos que contenham essas proteínas. As abordagens proteômicas têm sido apontadas como as técnicas não imunológicas mais promissoras para a detecção das proteínas do glúten (GLU). O objetivo desse trabalho foi caracterizar o perfil imunogênico e alergênico de farinhas de trigo (*Triticum aestivum*) brasileiro de diferentes qualidades tecnológicas em ambas as frações proteicas (GLU e AG), utilizando a cromatografia líquida de ultraperformance acoplada à espectrometria de massas de alta definição (UPLC-MS^E). As proteínas foram extraídas de nove farinhas, agrupadas em pool pela qualidade tecnológica em fraca (LW), média (MD) e forte (SP). Os peptídeos trípticos foram analisados pelos métodos MS^E ou $UDMS^E$ em um sistema NanoUPLC Acquity acoplado ao Synapt G2-HDMS (Waters). Os dados foram processados no software PLGS, usando o banco de dados *T. aestivum* (UNIPROT) e Propepper, específico para prolaminas. Na fração GLU foram identificados 29 peptídeos imunogênicos, sendo 10 comuns às 3 amostras, 4 presentes apenas na amostra SP, 3 na amostra MD e 5 na amostra LW. Foram identificadas 19 proteínas, sendo as subunidades de glutenina as principais proteínas associadas à DC e outras alergias. A amostra LW apresentou abundância total de peptídeos imunogênicos 27% maior em relação às outras amostras, bem como um maior número de epítopos associados à DC. Para a identificação da presença de proteínas alergênicas na fração AG, foi construída uma lista com 72 proteínas previamente reconhecidas como alergênicas. Foram identificadas 13 proteínas alergênicas nas frações AG, sendo os inibidores de α-amilase e as serpinas as principais proteínas identificadas responsáveis por desencadear alergias respiratórias. Embora não tenha sido observada diferença na expressão total das proteínas alergênicas entre as diferentes farinhas analisadas, as farinhas LW apresentaram maior expressão de duas isoformas de serpinas. Conclui-se que aplicação de métodos UPLC-MS^E permitiu a identificação e quantificação relativa de peptídeos imunogênicos e proteínas alergênicas em diferentes genótipos de farinha de trigo. Esse trabalho sugere uma expressão diferenciada de alérgenos entre farinhas de trigo de diferentes qualidades tecnológicas, além de demonstrar que as farinhas de qualidade tecnológica mais fraca podem apresentar maior potencial de causar reações imunológicas que as outras qualidades.

Palavras-chave: alergia, cromatografia líquida, doença celíaca, espectrometria de massas, peptídeos.

ABSTRACT

DETERMINATION OF PEPTIDES AND PROTEINS ASSOCIATED TO CELIAC DISEASE AND ALLERGIES IN WHEAT FLOURS BY UPLC-MSE

Wheat is a cereal of great economic, technologic and nutritional importance, due to the presence of proteins of gluten. Gluten is composed of alcohol-soluble prolamins and alcohol-insoluble glutenins, which confer viscoelasticity to dough. The technological quality of wheat is determined by the rheological properties of flour and defines its application in the industry. Various diseases are related to the consumption of gluten, which may be autoimmune, such as celiac disease (CD), that manifests in the small intestine; IgE mediated allergies and also sensitivities. Other diseases, such as respiratory allergies, are associated with the proteins of the soluble fraction (non-gluten), albumins and globulins (AG). The detection and quantification of wheat proteins are extremely important, because the treatment for most of these diseases is the exclusion of foods containing these proteins from diet. Proteomics approaches have been identified as the most promising immunological techniques to detect gluten proteins (GLU). The aim of this work was to characterize the immunogenic and allergenic profile of wheat (*Triticum aestivum*) flours of different technological qualities in both protein fractions (GLU and AG), using ultra performance liquid chromatography coupled to the high definition mass spectrometry (UPLC-MS^E). Proteins were extracted from nine flours and grouped in pools by technological quality in weak (LW), medium (MD) and strong (SP). The tryptic peptides were analyzed by methods MS^E or $UDMS^E$ in a NanoUPLC system coupled to the Synapt G2-Acquity HDMS (Waters). Data were processed in PLGS software, using the database *T. aestivum* (UNIPROT) and Propepper, specific to prolamins. In the GLU fraction 29 immunogenic peptides were identified, being 10 common to the 3 samples, 4 only in the sample SP, 3 in the sample MD and 5 in LW. 19 proteins have been identified, and glutenin subunits were the major proteins associated with CD and other allergies. The LW sample presented total abundance of the immunogenic peptides 27% higher than the other samples, as well as a larger number of epitopes associated with CD. To identify the presence of allergenic proteins in the AG fraction, a list of 72 previously recognised allergenic proteins was built. 13 allergenic proteins have been identified in the AG fractions, and α-amylase inhibitors and serpins the major identified proteins responsible for trigger respiratory allergies. Although it has not been observed difference in the expression of allergenic proteins between the different flours, LW flours presented a higher expression of two isoforms of serpins. It is concluded that application of UPLC-MS^E allowed the identification and relative quantification of immunogenic peptides and allergenic proteins in different genotypes of wheat flours. This work suggests a different expression of allergens between wheat flours of different technological qualities, as well as it shows that the flours of weaker technological quality may have greater potential to cause immunological reactions the other qualities.

Key words: allergy, celiac disease, liquid chromatography, mass spectrometry, peptides.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AACCI	*American Association of Cereal Chemists International*
ABITRIGO	Associação Brasileira das Indústrias de Trigo
ANOVA	Análise de variância
AOAC	*Association of Official Analytical Chemists*
ASM	Bomba auxiliar
BPI	*Base peak intention*
BSA	Albumina de soro bovino
CV	Coeficiente de variância
DDA	Aquisição dependente de dados
DIA	Aquisição independente de dados
DTT	Ditiotreitol
FAOSTAT	*Food and Agriculture Organization of the United Nations*
FDR	Taxa de falsos positivos (*false discovery rate*)
GFP	[Glu]^1-fibrinopeptídeo B
HDMS^E	Espectrometria de massas de alta definição (*High Definition MS^E*)
HMW-GS	Subunidades de glutenina de alto peso molecular (*High-molecular-weight glutenin subunits*)
IAM	Iodoacetamida
IMS	Espectrometria por mobilidade iônica
Leu-Enc	Leucina-Encefalina
LMW-GS	Subunidades de glutenina de baixo peso molecular (*Low molecular weight - glutenin subunits*)
m/z	Relação massa/carga
LW	Fraca (*Low*)
MALDI-TOF	Ionização e dessorção a laser assistida por matriz- tempo de voo
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
MD	Média (*Medium*)
MS	Espectrometria de massas (*Mass Spectrometry*)
NH_4HCO_3	Bicarbonato de amônio
ppm	partes por milhão
Sc	Corridas de prospecção (*Scouting Runs*)
SDS	Dodecil sulfato de sódio

SP	Forte (*Superior*)
TIC	Contagem total de íons (*total ion count*)
TOF	Tempo de voo (*time-of-flight*)
UDMS^{E}	Espectrometria de massas de ultradefinição (*Ultra Definition MS^{E}*)
UPLC	Cromatografia líquida de ultraperformance (*Ultra Perfomance Liquid Chromatography*)

SUMÁRIO

INTRODUÇÃO

O trigo (*Triticum aestivum*) é um dos cereais mais consumidos no mundo, por isso apresenta grande importância econômica e nutricional. A farinha de trigo é considerada a matéria-prima mais adequada para panificação, devido às suas propriedades viscoelásticas exclusivas. Sua produção tem se mantido constante ao longo dos anos, atualmente fica atrás apenas do milho e é seguido pelo arroz (FAO, 2017). Nos dados mais recentes informados pela USDA (2018), a produção mundial de trigo foi de aproximadamente 752 milhões de toneladas, suprindo o consumo de 740 milhões de toneladas. No Brasil, esse consumo foi de cerca de 12 milhões de toneladas. A estimativa de produção do trigo no país no ano de 2017 foi de em torno de 6,7 milhões de toneladas, tendo sido o restante do trigo consumido importado de outros países, como Argentina e Paraguai (ABITRIGO, 2017).

O trigo pode ser classificado de acordo com a qualidade tecnológica da farinha que servirá como base para definir a orientação industrial de uso do trigo. As análises de caracterização da farinha de trigo envolvem testes de análises químicas, reológicas e teste de panificação, realizados de acordo com metodologias aprovadas pelo órgão não governamental *American Association of Cereal Chemists International* (AACCI). Dentre as análises reológicas, destacam-se a farinografia (AACC International Method 54-21.02), a extensografia e a alveografia (AACC International Method 54-30.02).

No Brasil, o padrão de qualidade da farinha de trigo destinada ao segmento da panificação é definido pela Instrução Normativa nº 38 (MAPA, 2010) que define uma classificação dos tipos de farinha pela força do glúten (W), determinada por alveografia, estabilidade da massa, determinada por farinografia e número de queda pelo método de tempo de queda (Falling Number) (AACC International Method 56-81.03), usado para avaliar a atividade da enzima α-amilase e detectar possíveis danos causados pela germinação na espiga. Assim, a partir dessas análises, a farinha de trigo brasileiro é classificada em cinco classes: “Melhorador”, “Pão”, “Doméstico”, “Básico” e “Outros Usos”, e destinada à produção de diferentes alimentos, na indústria, ou para uso doméstico, conforme apresentado na Tabela 1.

A qualidade tecnológica da farinha está relacionada, portanto, com a textura e com as proteínas do glúten, gliadinas e gluteninas. Estas proteínas abrangem 80-85% do total de proteínas do grão de trigo, sendo as gliadinas subunidades monoméricas e as gluteninas subunidades poliméricas, formadas por polipeptídeos de alta e baixa massa molecular, estabilizados por pontes dissulfeto intermoleculares (Wieser, 2007).

Tabela 1. Classificação da qualidade tecnológica do trigo destinado à moagem e outras finalidades.

Classes	**Força do glúten (W)** (valor mínimo, expresso em 10^{-4}J)	**Estabilidade** (valor mínimo, tempo expresso em min)	**Número de queda** (valor mínimo, tempo expresso em s)
Melhorador	300	14	250
Pão	220	10	220
Doméstico	160	6	220
Básico	100	3	200
Outros Usos	Qualquer	Qualquer	Qualquer

Fonte: Adaptado Instrução Normativa nº 38 de 2010 (MAPA).

Devido à complexidade da estrutura e tamanho molecular que as proteínas do glúten podem alcançar, suas subunidades têm sido amplamente estudadas e a qualidade de panificação tem sido fortemente atribuída à composição de subunidades de gluteninas, principalmente às de alta massa molecular (Dhaka & Khatkar, 2015) e ao perfil de distribuição de massa molecular dos polímeros (Chaudhary et al., 2016).

Além da importância tecnológica e nutricional, uma vez que os produtos à base de cereais representam a base da pirâmide alimentar, o trigo tem um grande impacto em relação à saúde, já que diversas doenças são relacionadas ao consumo de cereais, principalmente associadas ao contato com as proteínas do glúten. As desordens relacionadas ao glúten (DRG) classificam-se em três tipos de acordo com a resposta desencadeada no organismo (Sapone et al., 2012). Elas podem ser de origem autoimune, como a dermatite herpetiforme, a ataxia induzida pelo glúten e a mais conhecida, a doença celíaca (DC); alergias imunomediadas pelos anticorpos IgE (Imunoglobulina E), como a WDEIA (do inglês, *Wheat-dependent exercise-induced anaphylaxis*), urticária de contato, alergia alimentar e alergias respiratórias; e por fim as sensibilidades, conhecidas como NCGS (do inglês, Non-celiac gluten sensitivity), que são desordens de origem não-alérgica e não-imunes (Sapone et al., 2012).

Os estudos sobre NCGS ainda são controversos e além do glúten, outras proteínas como inibidores da tripsina e aglutinina, e carboidratos fermentáveis como oligo-, di-e os monossacáridos e polióis (FODMAP) podem contribuir para os sintomas (Burkhardt et al., 2017; Sapone et al., 2012). Já as alergias respiratórias são majoritariamente relacionadas com as proteínas da fração solúvel, albuminas e globulinas e são conhecidas como “asma de padeiro” (Larre et al., 2011).

O tratamento para as DRG é a exclusão permanente do glúten da dieta. Assim, a detecção e a quantificação das proteínas de glúten são extremamente importantes, não só devido a seu efeito direto na qualidade dos alimentos, mas também por razões de segurança alimentar e saúde pública. Considerando os aspectos abordados, este trabalho tem como objetivo a caracterização do perfil imunogênico de farinhas de trigo brasileiro de diferentes genótipos em ambas as frações proteicas (glúten e albuminas/globulinas), por meio da utilização de abordagens proteômicas como cromatografia líquida de ultraperformance e espectrometria de massas de alta definição (UPLC-MS^E). Os resultados encontrados no estudo estão apresentados nessa dissertação na forma de artigo, conforme descrição dos capítulos a seguir.

No primeiro capítulo deste trabalho será apresentado um artigo de revisão bibliográfica, publicado na forma de capítulo no livro *Celiac Disease and Non-Celiac Gluten Sensitivity*. O capítulo intitula-se *Determination of Gluten Peptides Associated with Celiac Disease by Mass Spectrometry*. Além de introduzir conceitos sobre o trigo e as definições de cereais, glúten e doença celíaca, esse artigo aborda o uso de ferramentas proteômicas, como LC-MS, como as técnicas não-imunológicas mais promissoras na detecção de glúten, visto que esta é uma proteína de difícil identificação e quantificação devido à alta complexidade e alta taxa de homologia entre seus polipeptídeos. Ainda, essa revisão apresenta os avanços na identificação de epítopos responsáveis pelo desencadeamento de doenças relacionadas ao glúten, bem como na detecção de peptídeos marcadores de glúten em diferentes cereais por ferramentas proteômicas.

No capítulo 2, será apresentado um artigo original submetido para publicação à revista *Journal of Food Composition and Analysis*. O artigo é intitulado *Immunogenic and allergenic profile of wheat flours from different technological qualities revealed by UPLC-MS^E*. Nele, são apresentados os resultados obtidos do estudo de nove variedades de trigo brasileiro de diferentes genótipos, agrupadas em três classes por qualidade tecnológica. As frações proteicas, solúveis e insolúveis (albuminas/globulinas e glúten) foram extraídas sequencialmente e digeridas. Avaliou-se, por meio de UPLC-MS^E, a presença de peptídeos capazes de desencadear diferentes tipos de reações imunes. A partir desse estudo foi possível traçar um perfil imunogênico dessas farinhas, indicando quais variedades possuem um maior potencial para causar doenças relacionadas ao seu consumo.

CAPÍTULO 1 – DETERMINATION OF GLUTEN PEPTIDES ASSOCIATED WITH CELIAC DISEASE BY MASS SPECTROMETRY

Thais O. Alves[1,2]; Carolina T. S. D'Almeida [2,3]; Mariana S. L. Ferreira[1-3]

[1] *Food and Nutrition Graduate Program (PPGAN), Laboratory of Bioactives, Federal University of State of Rio de Janeiro (UNIRIO), Rio de Janeiro, Brazil.*

[2] *Laboratory of Protein Biochemistry - Center of Innovation in Mass Spectrometry (LBP-IMasS), UNIRIO, Rio de Janeiro, Brazil.*

[3] *Department of Food Science, School of Nutrition, UNIRIO, Rio de Janeiro, Brazil.*

Capítulo publicado no livro "Celiac Disease and Non-Celiac Gluten Sensitivity". Prof. Luis Rodrigo (Ed.), InTech, DOI: 10.5772/67547. 1ed., p. 43-58, 2017.

Abstract

Gluten is a big protein network composed of monomeric fraction (prolamins) and polymeric fraction (glutelins), occurring in many cereal-based products, especially in those containing wheat. Gluten peptides can trigger food allergies and intolerances, including inflammatory reactions as the celiac disease, an autoimmune disorder of the small intestine characterized by mucosal degeneration and villous atrophy. The treatment is the permanent exclusion of gluten from diet. However, gluten analysis is a very difficult task, due to the high complexity of polypeptides and the lack of consensus on the most appropriate analytical method. Proteomics approaches, combining liquid chromatography and mass spectrometry *in tandem* (LC-MS/MS), have been pointed as the most promising non-immunological techniques for gluten detection. LC-MS analyses associated with bioinformatics and specific-prolamin database can solve methodological limitations since it is based on the accurate molecular mass of peptide biomarkers. One of the major contributions of proteomics has been the identification of epitopes of gluten peptides responsible for wheat-related diseases. Recent works have defined grain-specific gluten peptides and also the lowest concentration at which peptides could be confidently detected. Proteomic application for gluten quantification should support not only regulatory limits in processed foods, but also the safety of consumers about food labeled as gluten-free.

Keywords: gluten peptides, LC-MS/MS, prolamins, proteomics, wheat.

1. Introduction

Gluten is defined as a complex protein network present in the cereals endosperm, responsible to confer viscoelasticity to pasta. It is composed by the cereal storage proteins, divided in two protein fractions: monomers, formed by alcohol-soluble prolamins; and polymers, formed by alcohol-insoluble glutelins (Wieser, 2007). This insoluble complex occurs when the gluten proteins are hydrated and submitted to mechanical force. Dry gluten is composed about 75-85% proteins and 5-10% lipids, the rest being residual starch and non-amylaceous carbohydrates (Wieser, 2007).

The wheat gluten network presents exclusive rheological properties as viscosity, extensibility and elasticity conferred by the storage proteins: gliadins and glutenins (Shewry & Halford, 2002). An appropriate proportion of both protein fractions in dough is essential to guarantee the viscoelastic properties and end-product quality (Wieser, 2007). Due to these properties, wheat is recognized as the most suitable raw material for bread and pasta making. Vital wheat gluten is a raw material widely added in gluten-based food products to improve quality and sensory properties and can be obtained from washing the viscoelastic dough, removing the water-soluble components (Day et al., 2006; Giannou & Tzia, 2016).

Besides the technological aspect, the gluten proteins can trigger food allergies and intolerances, including inflammatory reactions in patients with celiac disease (CD). CD is a gluten-sensitive enteropathy defined as an immune-mediated disorder triggered by gluten in genetically predisposed individuals.

The family of storage proteins of gluten occurs in wheat grains (*Triticum* spp.; gliadins and glutenins), barley (*Hordeum vulgare*; hordeins), rye (*Secale cereale*; secalins) and oats (*Avena sativa*; avenins). In the context of gluten intolerance, one of the most common definitions of gluten is provided by the European Commission Regulations: “protein fraction from wheat, rye, barley, oats or their crossbred varieties and derivatives thereof, to which some persons are intolerant and which is insoluble in water and 0.5 M sodium chloride solution” (Vassiliou, 2009).

The gluten proteins are present in various types of cereal-based food products, mainly in wheat-based products. However, due to the incorporation of gluten as an ingredient in foods that traditionally not contain wheat proteins, there is also a growing concern about gluten alergenicity in hidden sources of gluten, incorrect labeling or cross-contamination in manufacturing, transportation and storage (Day et al., 2006). Hence, because of its nutritional

and economic importance, there is a big effort to characterize these proteins. Since the treatment for gluten-sensitivity is the exclusion of gluten from diet, the detection and quantification of these proteins are extremely important, not only due to its direct effect on the food quality, but also for food safety reasons.

Nevertheless, the gluten analysis in food products is a very difficult task, due to the need to properly extract the proteins before analysis and to the high complexity and homology of polypeptides. Hence, the first point to be addressed is the appropriate protein extraction, whose steps involve sequential buffers to perform prolamins extraction and the reduction of disulphide bonds of glutenins, formerly insoluble, releasing their polypeptides (Wieser et al., 1998).

The second point is about the lack of consensus on the most appropriate analytical method to identify and quantify gluten in food. The most commonly used methods are based on ELISA, PCR and also electrophoresis, but these methods differ in terms of sensitivity and present several drawbacks. The main faced problem is related to the lack of certified reference material (Martinez-Esteso et al., 2016). In fact, the immunological methods are based on the use of developed antibodies for the detection of gliadins and, therefore, are not suitable for all classes of gluten proteins. In addition, current methods are unable to distinguish the source of cereals.

The protein composition of the grain varies among different species and varieties, and it leads to methodological difficulties in the allergenic food analysis. In this context, modern proteomic approaches based on sensitive and reliable techniques combining liquid chromatography (LC) coupled with mass spectrometry *in tandem* (MS/MS) have been pointed as the most promising non-immunological techniques for identification and quantification of gluten proteins, even in traces level (Colgrave et al., 2016; Fiedler et al., 2014; Martinez-Esteso et al., 2016; Uvackova et al., 2013a).

2. Cereals

Cereal grains are essentially composed by endosperm (~ 83%, on weight basis), germ (or embryo, ~ 3%) and bran (or external layers, ~ 14%) (Colgrave et al., 2015; Feillet, 2000). The endosperm contains about 80-90% of starch and can contain 8-20% protein (on dry

basis), that correspond mainly to gluten proteins (Feillet, 2000). These proteins are important due to its impact on technological processing of cereals.

The most representative species of this class are rice, wheat, rye, barley and corn. Wheat is one of the most important and most consumed cereals in the world, and is considered the most suitable raw material for baking and pasta-making. Its production and consumption have remained constant over the years, being the second most produced and consumed cereal (the first one is corn and rice the third one) (FAOSTAT, 2016).

Rye, barley and oats also have significant production and consumption; they are mainly used for baking, especially in the case of rye; barley malt is an important ingredient for beer production but can also be found in the form of meal, flakes or flour, whereas bran and other oat-based products are largely available for immediate consumption (Owusu-Apenten, 2002).

The cereals proteins are classically divided according to Osborne (1907), in four groups consistent with its solubility, being albumins soluble in water, globulins in diluted saline solutions, prolamins in alcoholic solutions and glutelins in diluted acids or bases. Albumins and globulins are metabolic proteins, which represent 20% of total protein content and participate in important functions in plant development and responses to environment (Gao et al., 2009), while prolamins and glutelins, cumulatively referred to as gluten, represent the major class of storage proteins (i.e. 80% of total protein), which function is to store nutrients, providing nitrogen during seeds germination (Feillet, 2000).

3. Gluten proteins

Gluten proteins are represented by the storage proteins, that are divided in two groups: prolamins (e.g. gliadin, hordein, secalin, avenin), which are monomerics; and glutelins (e.g. glutenin, D-hordein, secalinin or simply HMW-secalin), which are polymerics. The last ones, as a result of the numerous covalent associations between polypeptides, may remain insoluble even in strongly denaturant buffers such as sodium dodecyl sulfate (SDS) (Shewry & Halford, 2002; Shewry et al., 1995).

The gluten proteins present common structural characteristics. The primary structure of these proteins is subdivided into distinct domains that may have repeated sequences of some specific amino acids (Shewry & Halford, 2002). These proteins are unique in terms of

amino acid composition, characterized by high levels of proline (P) and glutamine (Q)[1] and low levels of amino acids with charged side groups. Glutamin generally predominates (15–31%), followed by proline in the case of wheat, rye, and barley (12–14%) (Koehler & Wieser, 2013). The cysteines represent only 2% of the amino acids of the gluten proteins but are extremely important to the structure and functionality of gluten (Wieser, 2007). The nutritionally essential amino acids tryptophan (0.2–1.0%), methionine (1.3–2.9%), histidine (1.8–2.2%), and lysine (1.4–3.3%) are also present only at very low levels (Koehler & Wieser, 2013).

Breeding and genetic engineering have been successful applied to improve the content of essential amino acids, such as the case of high-lysine barley and corn. However, these approaches may be used to develop celiac-safe wheat, this remains a formidable challenge due to the complex multigenic control of gluten protein composition, besides the requirement of acceptable technological properties for bread and pasta making (Kucek et al., 2015; Shewry & Tatham, 2016).

The cereals present variable levels of Osborne's fractions (albumins, globulins, prolamins and glutelins). The amino acid composition of prolamins can be correlated to the botanical genealogy of cereals, where wheat, rye and barley belong to the subtribe *Triticeae* and oat to *Aveneaea* (Belitz et al., 2009). The amino acid composition is similar in wheat, rye and barley, whereas in oats, the prolamins composition is intermediate between the *Triticeae* and other cereals. The amount of glutamine in oat prolamins is similar of the *Triticeae*, while the amounts of proline and leucine in oats prolamins are smaller and larger, respectively, to those found in *Triticeae* (Belitz et al., 2009).

Gliadins are the group of monomeric proteins present in wheat gluten, whose molecular weight (MW) range from 30 to 75 kDa. Gliadins are regrouped based on its electrophoretic mobility and structural similarity: α/β-gliadins, γ-gliadins e ω-gliadins. As the other cereals prolamins, they are all soluble in alcohol, a characteristic of this group (Veraverbeke & Delcour, 2002). The α/β- and γ-gliadins are smaller (30-60 kDa) than the ω-gliadins (<75 kDa) (Shewry & Halford, 2002). The first ones have very similar primary sequences and present N-terminal domain with repetitive sequences with 7-11 amino acids (P/Q) and C-terminal homologue domains, with 6-8 cysteines able to form intrachain

[1] Typical of all cereal flours is the fact that glutamic acid almost entirely occurs in its amidated form as glutamine.

disulphide bonds (Shewry et al., 1995). The ω-gliadins show the highest levels of proline and glutamine, with repetitive sequences of 8-10 of these amino acids.

The wheat glutenins are formed by a heterogeneous mixture of polypeptides with high molecular weight, which can reach until 1 million Da. They are considered one of the biggest proteins found in nature (Wrigley, 1996). Depending on the polymerization degree, these polymers remain insoluble even in denaturating buffers such as SDS, leading to a difficult solubilization. Glutenin polymers are formed by monomeric glutenin subunits (GS), subdivided according to the MW and stabilized by interchain disulphide bonds. The HMW-GS (high molecular weight glutenin subunits) present MW ranging between 65-90 kDa and can be subdivided into x-type and y-type; while the LMW-GS (low molecular weight glutenin subunits) present 30-60 kDa and are subdivided into B-, C-, and D- groups according to electrophoretic mobility (D'Ovidio & Masci, 2004; Veraverbeke & Delcour, 2002).

In other cereals, HMW group contains HMW-secalins and D-hordeins, respectively in rye and barley. They comprise polymers (glutelins) possessing around 600–800 amino acid residues, MW of 70 and 90 kDa and a high content of glutamine, glycine and proline which represent around 60% of residues (Koehler & Wieser, 2013). HMW and MMW proteins are missing in oats. The MMW group consists of monomeric ω-secalins and C-hordeins, including 300 and 400 amino acid residues and MW around 40 kDa. They are characterized by high contents of glutamine, proline, and phenylalanine, which together account for 80% of residues.

The LMW group includes monomers such as γ-40 k-secalins, γ-hordeins and avenins of oats, but also polymers including γ-75 k-secalins, and B-hordeins. They have between 200 and 430 amino acid residues, with MW ranging from 23-50 kDa and its amino acid composition is dominated by glutamine and proline, and by relatively high levels of hydrophobic amino acids, leucine and valine (Scherf et al., 2016).

Wheat gluten is of great importance in the food industry because it promotes the dough ability to retain carbon dioxide produced during fermentation, resulting in the rising of dough that presents good gas holding properties. Barley and rye flours are also able to form gluten because of its chemical composition, whose proteins are similar to gliadins and glutenins. However, the gluten network formed by them is more fragile since these proteins are present in a smaller amount than in wheat flour (Belitz et al., 2009). Due to the unique

viscoelastic characteristics conferred by the wheat gluten proteins, wheat flour becomes an essential ingredient for the food production (Day et al., 2006).

4. Celiac disease (CD)

CD is an autoimmune disorder of the small intestine characterized by mucosal degeneration and villi loss, mainly affecting the capacity of nutrient absorption. Its origin is related with the presence of genes HLA-DQ2 or HLA-DQ8 (HLA: human leukocyte antigen) and both genotypes cause the predisposition for the disease (Presutti et al., 2007), but 95% of CD patients exhibit the DQ2 serotype class (Scherf et al., 2016). In predisposed individuals, it can manifest in any stage of life, since that the contact with the protein fraction of wheat, barley or rye was established (Bongiovanni et al., 2010).

Diagnosed patients cannot consume foods containing gluten or its traces, because even a minimal amount of this protein can trigger the reaction, causing the most varied symptoms, ranging from abdominal pain, bloating and diarrhea, to osteoporosis and infertility in long term. The severity of the reaction can be due to the degree of intolerance of each individual (Banerjee, 2010; Pietzak & Fasano, 2005).

Current knowledge about the pathogeneses of CD has been associated with the long chain and amino acid composition of the peptides generated during gastrointestinal digestion of the gluten proteins (Kucek et al., 2015). Due to the lack of lysine and arginine residues in gluten proteins, the action of the proteases, such as trypsin, but also chymotrypsin and pepsin, is very difficult, making the proteolysis practically ineffective. Because of its hard cleavage, those proline and glutamine-rich polypeptides act as mediators of immune reactions in the intestinal epithelium cells of the predisposed subjects (Scherf et al., 2016).

The most celiac-active T-cell epitopes are present on the α-gliadins, but T-cell epitopes derived from either γ- and ω-gliadins as well as from HMW and LMW-GS have been reported (Shewry & Tatham, 2016; Sollid et al., 2012). However, T-cell epitopes from hordeins and secalins have been also described, it can be explained by their high homology to those found in wheat (Sollid et al., 2012). While the consumption of wheat, rye and barley has been proved to cause harm to CD patients, there is still a discussion about the safety consumption of oats by CD patients.

In this context, there are controversies about the reactivity of oat gluten, since only a few number of celiac patients have demonstrated to be affected by oat consumption (Food & Drug, 2007; Pietzak & Fasano, 2005). Recent reports suggest a tolerated oat consumption for a great part of celiac patients, showing a safe long-term feeding (Farage et al., 2015; Kaukinen et al., 2013). Although some authors consider oats a gluten-free cereal, the main problem is the risk of cross contamination by gluten-based cereals during harvest, milling or industry processing (Hernando et al., 2008; Koerner et al., 2011; Vassiliou, 2009). For this reason, this cereal cannot be completely discarded as CD trigger and its consumption by celiacs is still considered unsafe (Hallert et al., 1999; Sealey-Voyksner et al., 2010).

The *Codex Alimentarius* proposed in 2008 a standard international labeling, where products labeled as "gluten-free" must not exceed the limit of 20 ppm of wheat, barley or rye gluten, which corresponds to approximately 1 mg of gluten in 50 g of food (Codex Alimentarius, 2008). The maximum amount of gluten tolerated by celiac patients is not completely known, because of the variable reactivity of gluten among different species and also the unpredictable sensitivity among individuals. However, several studies have indicated that 10 mg of gluten daily are well tolerated, while intestinal mucosa damage has been observed with doses around 50 mg (as reviewed by Hischenhuber et al., 2006).

The difference in the amino acid composition of prolamins and glutelins from each cereal has been pointed as responsible by the different reactivity associated with the CD (Belitz et al., 2009; Colgrave et al., 2015). Compared to other cereals, grains belonging to subtribe *Triticeae* (wheat, barley and rye) contain significantly higher levels of glutamine and proline than others, being these amino acids the principal responsible for triggering the immune response in celiac disease (Scherf et al., 2016). A direct correlation between the immunogenicity of the different oat varieties and the presence of specific peptides with differential reactivity has been proposed as the origin of the wide range of variation of potential immunotoxicity of oat cultivars (Real et al., 2012).

Triticum species exhibits an important genetic variability, resulting in different toxicities, what can be a promising alternative for obtaining suitable varieties for consumption by celiac patients (Gell et al., 2015; Shewry & Tatham, 2016; Spaenij–Dekking et al., 2005). Higher levels of immunogenic peptides related to CD were attributed to a modern Canadian wheat when compared to old varieties of common wheat and tetraploid wheat (van den Broeck et al., 2015). Despite the importance of genotypic variation within species and

cultivars, specific knowledge about CD, especially regarding the structure of the allergens and the immunoreactive epitopes is not fully known and requires new information.

5. Gluten detection techniques

Several methods have been developed to guarantee the safety of foods labeled or expected to be gluten-free for celiac patients. However, there is no consensus about the analytical method considered more appropriate to identify and quantify gluten in foods (Sealey-Voyksner et al., 2010). The main used methods are based on different techniques for the detection of DNA sequences, related proteins, such as the Enzyme-Linked Immunosorbent Assay (ELISA) and the Polyacrylamide Gel Electrophoresis (PAGE) methods or more recently, the detection of digested peptides by means of LC-MS.

These methods differ widely from each other, especially in terms of sensitivity, specificity and cost. Other reasons for this divergence can be related with food processing (heat or hydrolysis steps), matrix type, polymorphic variants of wheat, rye and barley, type of extraction and possible cross-reaction with other prolamins.

5.1. ELISA

Currently, the ELISA method is the most common and recognized approach for detection of gluten, because it presents low cost, it is easy to perform and promotes results quickly. It is the technique recommended by the Codex Alimentarius for the detection of gluten in industrialized foods (Codex Alimentarius, 2006). This technique is based on the immunological reaction between known toxic peptides from gluten proteins and mono- or poly-clonal antibodies.

There are two variations of the method, the R5 ELISA sandwich and competitive R5 ELISA. In ELISA-sandwich, samples containing the antigens are incubated to form an antibody-antigen complex, and then a labeled antibody is incubated, conjugated to another antigen epitope, forming two layers of antibodies. This method requires at least two binding sites (epitopes) for the antibody and is only suitable for large peptides or intact proteins quantitation, being unfeasible to detect partially hydrolyzed gluten (e.g. fermented foods).

The competitive ELISA only requires one epitope and is indicated for detecting minor antigens, present in partially degraded gluten. In this method occurs a competitive binding process performed by original antigen (sample antigen) and the added antigen, leading to the competition of the antigens by the limited number of epitopes. When available, quantification can be done through calibration curves with reference proteins (Arendt & Dal Bello, 2008).

Some ELISA-based studies were successfully applied in the detection of wheat, barley and rye contamination, with confirmation of the results by MS and PCR (Hernando et al., 2008; Koerner et al., 2011). However, measurements by commercial ELISA kits are inconsistent and require standardization of results due to the lack of certified reference material and the diversity of kits using different test conditions (Manfredi et al., 2015; Martinez-Esteso et al., 2016).

Current methods are based on the use of antibodies that are not accurate and may have false negative results. These antibodies were especially developed for the detection of gliadins and therefore are not suitable for all classes of gluten, especially in matrices that are difficult to analyze (Colgrave et al., 2014; Martinez-Esteso et al., 2016; Tanner et al., 2013a; Tanner et al., 2013b). The accuracy of ELISA method is also compromised since the result is converted into gluten by multiplication by two, assuming that the gliadin/glutenin ratio is constant. Moreover, the current methods are not able to distinguish the cereal source (wheat, barley, rye) or cultivar (Diaz-Amigo & Popping, 2013; Wieser & Koehler, 2009).

The development of standardized gluten material represents significant progress towards the accurate analysis of gluten in low levels. However, this is a challenging task due to polymorphism of gluten proteins, which vary from sample to sample (Manfredi et al., 2015; Martinez-Esteso et al., 2016; Tanner et al., 2013b). When comparing the use of modern techniques such as LC-MS and ELISA, previous studies show no correlation between ELISA results and the relative content of peptides determined by MS (Tanner et al., 2013b). The authors concluded that ELISA methods are no longer sufficient for gluten quantification and should eventually be replaced by MS based methods.

In this context, methods based on MS have been alternatively proposed for gluten quantification, since it can detect specific and comprehensive peptides with good sensitivity and precision, due to the high throughput data analysis capacity (Manfredi et al., 2015; Uvackova et al., 2013a). A progressive number of approaches using MS have been developed,

offering great potential in this area (Fiedler et al., 2014; Koeberl et al., 2014; Manfredi et al., 2015; Sealey-Voyksner et al., 2010).

5.2. Proteomic tools for gluten detection

Proteomics is the large-scale analysis of the set of proteins encoded by the genome responsible for controlling almost all biological processes in a particular biological system at a certain time. Proteomics includes not only the structural and functional knowledge of proteins but also the study of their modifications, interactions, localization and quantification. The proteome of an organism is dynamic; it will reflect the momentaneous response of those cells to a determinate stimulus. It means that a single genome can give origin to infinite different proteomes (Graves & Haystead, 2002).

The most practical application of proteomics refers to the analysis of target proteins as opposed to entire proteomes (Graves & Haystead, 2002). The use of proteomics in food analysis has become a key technological tool for characterization and quantification of proteins and peptides, especially when it comes to the evaluation of biological markers (Herrero et al., 2012).

The protein composition of cereals is variable between different species and varieties, leading to methodological difficulties for food allergen analysis and also for selection of genotypes. The high similarity of amino acid sequences of the different prolamins, together with limitations on the available methodologies, make difficult the exact identification of the allergens and immunoreactive epitopes related to CD, as well as its genotypic frequency, variability and stability (Juhász et al., 2015a).

In this context, proteomic approaches based on reliable and sensitive techniques such as high resolution LC-MS, reveal themselves as important tools for the identification, quantification and also discrimination of gluten proteins, since it is based on accurate molecular mass of peptide biomarkers.

In the last years, MS techniques have overcome some limitations associated to antibody-based methods, such as cross-reactivity and discriminating capacity of gluten protein sources in a single run (Manfredi et al., 2015). Recently, label-free MS experiments have been improved in order to quantify specifically CD-epitopes (van den Broeck et al., 2015).

This type of research is very important, since accurate quantification and identification of the cereal source and protein type of contamination is critical to the health and well-being of celiac patients (Colgrave et al., 2016). Furthermore, labeled "gluten-free" food products have showed contamination with gluten-containing protein fractions above the acceptable (20 ppm) (Sharma et al., 2015).

One of the major contributions of proteomics related with gluten sensitivity diseases, especially CD, has been the identification of epitopes sequences of gluten peptides of known immunogenic action. A number of gluten T-cell epitopes restricted by CD associated HLA-DQ molecules have been characterized over the last few years and a compiled a list of epitopes from gluten peptides able to activate the immune system was proposed (Table 2) (as reviewed by Sollid et al., 2012). It is interesting to note that the identified sequences were not only from prolamins but also from glutelins. A website dedicated to these epitopes was created to update the list, but until now presented no recent inputs (Sollid et al., 2012).

More recently, a database (ProPepper™) built from *in silico* results was proposed to assist the identification of epitopes, peptides and prolamins associated with DC and other types of wheat and cereal disorders (Juhász et al., 2015a). This database contains sequences of specific peptides, *in silico* digested, from prolamins available in public databases (UniprotKB, NCBI GenBank), and currently presents 37,914 peptides and 833 epitopes.

Table 2. List of gluten peptides epitopes recognized by immune system. Adapted from Sollid et al. (2012).

Epitope	Sequence of peptides recognized
DQ2.5-glia-α1a	PFPQPELPY
DQ2.5-glia-α1b	PYPQPELPY
DQ2.5-glia-α2	PQPELPYPQ
DQ2.5-glia-α3	FRPEQPYPQ
DQ2.5-glia-γ1	PQQSFPEQQ
DQ2.5-glia-γ2	IQPEQPAQL
DQ2.5-glia-γ3	QQPEQPYPQ
DQ2.5-glia-γ4a	SQPEQEFPQ
DQ2.5-glia-γ4b	PQPEQEFPQ
DQ2.5-glia-γ4c	QQPEQPFPQ
DQ2.5-glia-γ4d	PQPEQPFCQ
DQ2.5-glia-γ5	QQPFPEQPQ
DQ2.5-glia-ω1	PFPQPEQPF
DQ2.5-glia-ω2	PQPEQPFPW
DQ2.5-glut-L1	PFSEQEQPV
DQ2.5-glut-L2	FSQQQESPF
DQ2.5-hor-1	PFPQPEQPF
DQ2.5-hor-2	PQPEQPFPQ
DQ2.5-hor-3	PIPEQPQPY
DQ2.5-sec-1	PFPQPEQPF
DQ2.5-sec-2	PQPEQPFPQ
DQ2.5-ave-1a	PYPEQEEPF
DQ2.5-ave-1b	PYPEQEQPF
DQ2.2-glut-L1	PFSEQEQPV
DQ8-glia-α1	EGSFQPSQE
DQ8-glia-γ1a	EQPQQPFPQ
DQ8-glia-γ1b	EQPQQPYPE
DQ8-glut-H1	QGYYPTSPQ
DQ8.5-glia-α1	EGSFQPSQE
DQ8.5-glia-γ1	PQQSFPEQE
DQ8.5-glut-H1	QGYYPTSPQ

5.2.1. Liquid Chromatography coupled to Mass Spectrometry (LC-MS)

LC-MS is an analytical technique that consists in the separation process based on differential interaction of sample components of a mixture, combining a powerful technology of the generation of molecular ions (ionization), which are separated and detect based on their mass/charge ratio (*m/z*) (Gross, 2004).

In nowadays, *tandem* designs (also referred to as MS/MS) makes up most of the instruments in research laboratories. In this configuration, high energy is applied to produce fragments from precursors ions; hence the selected peptides are then submitted to fragmentation in order to elucidate the amino acids sequence, allowing the confirmation and identification of sequences differing from one single amino acid (Ferreira et al., 2014; Graves & Haystead, 2002). LC-MS/MS is considered a gold standard for the analysis of biomolecules in complex samples, due to high levels of sensitivity and specificity, and has been used in food analysis and forensic science (Castiglioni et al., 2006; Fenn et al., 1989; Lehotay et al., 2005).

The main current strategies to identify gluten markers use both discovery (known as shotgun analysis) and targeted-based proteomic approaches. Basically, combined strategies can be applied based on primary fractionation of gluten proteins using RP-HPLC or SE-HPLC followed by a multi-enzymatic based digestion of the protein resulting fractions and high resolution MS or MS/MS measurements (Colgrave et al., 2016; Fiedler et al., 2014; Martinez-Esteso et al., 2016). The investigated gluten markers peptides can be identified by comparison via theoretical (*in silico*) and experimental results (e.g. *de novo* peptide sequencing), using current protein databank (NCBI, UniprotKB) or specific cereal prolamins epitopes involved in CD pathogenesis (Juhász et al., 2015b).

For the selection of gluten markers, the main used MS technique is the SRM or MRM (Selected or Multiple Reaction Monitoring) that allows targeted analysis, especially for quantification even in trace levels. The MRM method uses a mass spectrometer of triple quadrupole type (QqQ), where the precursor ions will be selected and focused on the first quadrupole (Q1). The second quadrupole (q2) is actually a collision cell, where occurs the injection of a collision gas (usually argon) leading to ion fragmentation. The third quadrupole (Q3) is the mass analyzer, responsible for defining which are the fragments generated in the collision cell according to their *m/z* (Kitteringham et al., 2009).

In recent studies, some authors evaluated the presence of gluten peptides markers in beers by using MRM techniques (Tanner et al., 2013b). These authors revealed the superiority of LC-MS in relation to the ELISA method when comparing analytical methods to quantify low levels of gluten peptides, since MS quantification is undertaken using peptides that are specific and unique, enabling the quantification of individual hordein isoforms.

Looking for more reliable results for celiac patients, other studies have sought to define gluten specific peptides in an attempt to validate the MS as high sensitivity analytical method for gluten detection. Fiedler et al. (2014) applied MS to identify grain-specific peptides marker for wheat, barley, rye and oats, to assess gluten contamination in various types of commercial flours. Martinez-Esteso et al. (2016) identified a set of unique wheat gluten peptides and proposed its use as markers for the presence of gluten related to CD manifestation. The same authors reinforce the idea that this strategy can be applied to other food allergens, and may be considered the first step for developing certified reference materials and defining a new methodology, more sensitive than ELISA, to detect gluten in foods.

For complex samples, such as gluten proteins, multiplex methods of acquisition, called DIA (data independent acquisition) or MS^E allow to recover sample all the ions and minimize data loss (e.g. non-fragmented precursors) (Uvackova et al., 2013a, 2013b). In MS^E methods, all the ions generated at ionization source are transmitted to the collision cell, which alternates between high and low energy (c.a. from 15 to 55 eV), sending to the TOF analyzer, simultaneous, the precursors and fragments of the peptides (Egertson et al., 2015).

Modern technologies can be applied to surmount cross-reactivity problems associated to antibody recognition, that are particularly challenging in gluten analysis due to high level of homology between different prolamins. For a consistent analysis of primary structures, showing a high degree of homology, it is also possible to separate peptides applying the ion-mobility system (IMS) that consists of an orthogonal separation technique, where for each value of *m/z*, a spectrum of *drift time* (dt) is added. The dt corresponds to the time taken by the ion to cross the ion-mobility cell, full of an inert gas, allowing the determination of cross-shock sections (Michaelevski et al., 2010).

The integration of IMS into MS^E workflows, provides an additional dimension of separation, improving system peak capacity while concomitantly reducing chimeric and composite interferences; ions can be distinguished by size, shape and charge, besides to the

m/z (Distler et al., 2016). MS^E is also able to provides absolute quantitative analysis by examining the signal response of a known internal standard spiked into the sample (Uvackova et al., 2013a). Developing MS^E methods to quantitatively measure gluten peptides could support advancement in understanding the natural variability in protein expression of clinically relevant wheat grain allergens. Proteomic application for gluten quantification should support not only regulatory limits in processed foods, but also the safety of consumers about food labeled as gluten-free.

CAPÍTULO 2 – IMMUNOGENIC AND ALLERGENIC PROFILE OF WHEAT FLOURS FROM DIFFERENT TECHNOLOGICAL QUALITIES REVEALED BY UPLC-MS^E

Thais O. Alves[a]; Carolina T. S. D'Almeida[a]; Verônica C. M. Victorio[a]; Gustavo H. M. F. Souza[b]; L. C. Cameron[c]; Mariana S. L. Ferreira[a,c]

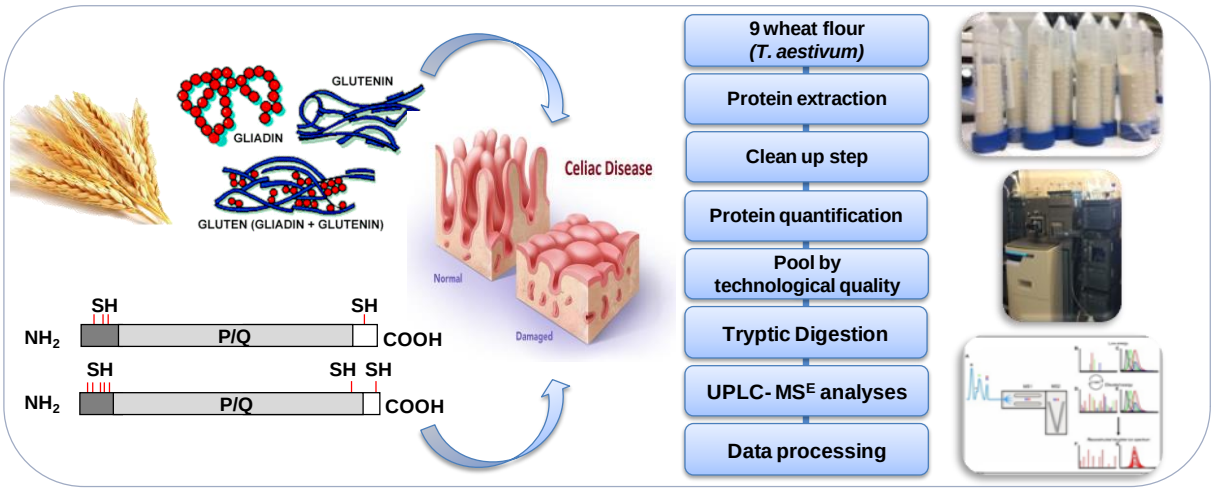
[a] *Laboratory of Bioactive Compounds, Food and Nutrition Graduate Program (PPGAN); Federal University of the State of Rio de Janeiro, UNIRIO. Av. Pasteur, 296, 22290-240 Rio de Janeiro, Brazil.*

[b] *MS Applications Research and Development Laboratory, Health Sciences Research, Waters Corporation, São Paulo, SP, Brazil.*

[c] *Laboratory of Protein Biochemistry, Center of Innovation in Mass Spectrometry, UNIRIO. Av. Pasteur, 296, 22290-240 Rio de Janeiro, Brazil.*

Artigo submetido à publicação no periódico "Journal of Food Composition and Analysis". Elsevier. ISSN: 0889-1575.

Graphical abstract



Abstract

Wheat proteins, especially gluten, are known to be potentially allergenic and able to trigger various kinds of immune reactions in predisposed individuals, including celiac disease (CD). A two-step protein extraction (soluble, AG and gluten, GLU) was performed from nine common wheat flours of different technological qualities (low, LW, media, MD and SP, superior). The immunogenic amd allergenic peptides were evaluated by UPLC-MS^E and $UDMS^E$ analysis, confronting the results with *T. aestivum* database from UniProt and ProPepper™. Collectively, 3,603 and 91,609 peptides were identified and after filtering around 450 and 1,708 were identified in all samples, respectively from GLU and AG fraction. About 19% of the identified peptides shown immunogenic action associated with prolamins, where only 10 peptides were common to all samples. The most part of peptides (60%) belong to HMW-glutenin subunits and were related to CD and WDEIA (wheat-dependent exercise-induced anaphylaxis), although the immunogenicity is mainly associated with gliadins. The immunogenic peptides were more expressed in LW than SP and MD flours. LW flours presented also the highest number of epitopes (28) related to wheat allergies, suggesting a larger ability to cause immunological reactions than the other flours qualities. The major proteins associated to respiratory allergy from AG were α-amylase inhibitors and serpins. Two serpins were found more expressed in LW flours. This work presented the most comprehensive and detailed wheat proteins immunogenic profile based on proteomics analysis to date. The UPLC-MS^E approach allowed the identification and relative quantification of allergenic peptides of AG and GLU proteins. LW quality flours showed greater potential to cause immunological reactions than the other qualities.

Keywords: celiac disease; gluten; LC-MS; mass spectrometry; prolamins; proteomics; wheat allergy; wheat quality**.**

1. Introduction

Wheat is one of the most important and most consumed cereals in the world, and it is considered the most suitable raw material for bread-making. The wheat proteins are classically divided into metabolic and storage proteins. The metabolic proteins involve the group of albumins and globulins, which are respectively water and salt-soluble proteins and represent 20% of the total protein content of the grain. They are mainly represented by enzymes involved in important functions in the development of the plant and in the responses to the environment (Gao et al., 2009).

The storage proteins are represented by gluten proteins, which are divided in monomeric, alcohol-soluble prolamins and polymeric, alcohol-insoluble glutelins (Shewry & Halford, 2002; Shewry et al., 1995; Wieser, 2007). In wheat, these proteins are called gliadins and glutenins, or, generically, only prolamins (Shewry et al., 1986), due to the high content in proline (P) and glutamine (Q) residues. Gliadins are divided in subgroups α, β, γ, and ω based on electrophoretic mobility, while glutenin subunits are grouped into high (HMW-GS, type-x or type-y) or low (LMW-GS) based on their apparent molecular weight (D'Ovidio & Masci, 2004).

The wheat proteins, specially prolamins are the main responsible for triggering some inflammatory diseases, such as celiac disease (CD), atopic dermatitis (AD), urticarial and wheat allergies (WA) such as baker's asthma and wheat dependent exercise induced anaphylaxis (WDEIA) (Uvackova et al., 2013b).

CD is the most known autoimmune enteropathy related to gluten consumption in genetically predisposed individuals. It is characterized by mucosal degeneration and villi loss of the small intestine, leading to a malabsorption of nutrients. It is related with the presence of genes HLA-DQ2 or HLA-DQ8 (HLA: human leukocyte antigen) (Presutti et al., 2007) and can manifest in any stage of life, since contact with the protein fraction of wheat, barley or rye has been established (Bongiovanni et al., 2010). While the consumption of wheat, rye and barley has been proved to be harmful to CD patients, oat is generally well tolerated but the safe consumption of oats by CD patients remains a topic of discussion (Comino et al., 2015). Indeed, oat prolamins represent only 10-20% of the total grain protein amount, differing from wheat, barley, and rye where glutelins and prolamins accounting for 60-85% of total protein, besides to present an amount of proline 40-50% lower than those cereals (Belitz et al., 2009).

In CD, the epitopes associated to the reactions are frequently related to the α-gliadins, but some derived from γ-, ω-gliadins, HMW and LMW-GS have been also reported (Shewry & Tatham, 2016; Sollid et al., 2012). In WDEIA, the main allergens are the ω-gliadins and HMW-GS, however LMW-GS can also trigger WDEIA because they share epitopes with the major wheat allergen known as ω5-gliadin (Bouchez-Mahiout et al., 2010; Hofmann et al., 2012; Matsuo et al., 2005). In addition, metabolic proteins from the soluble fraction, such as α-amylase/trypsin inhibitors, β-amylase, peroxidases, serpins and α-purothionin have been also reported by their ability to trigger IgE-mediated allergies related to wheat consumption (Burkhardt et al., 2017; Larre et al., 2011; Tatham & Shewry, 2008). Although soluble proteins are the main allergens related to WA, gliadins are also reported to cause respiratory allergy such as baker's asthma (Bittner et al., 2008).

As members of the *Poaceae* family and belonging to the same tribe *Triticeae*, some proteins are shared among different species such as wheat, barley, and rye, and their prolamins present a high homologous sequences and repetitive motifs, making the investigation of allergenic gluten peptides a very difficult task (Fallahbaghery et al., 2017). The development of modern methods for determination of clinically relevant wheat allergens is essential to comprehend the exposure thresholds, to support clinical allergy study designs (Uvackova et al., 2013a) and also to direct the plant breeding programs (Rogniaux et al., 2015).

Proteomic techniques such as liquid chromatography coupled to mass spectrometry *in tandem* (LC-MS/MS) have been considered the most promising non-immunological method to identify, quantify and discriminate gluten proteins in food matrices, since it is based on accurate molecular mass of peptide biomarkers (Alves et al., 2017). Proteomic application for wheat allergens quantification should support regulatory limits in processed foods and the safety of consumers about food labeled as gluten-free, since this protein can be identified even in small concentrations (Manfredi et al., 2015; Martinez-Esteso et al., 2016; Rogniaux et al., 2015).

One of the major contributions of proteomics in this field has been the identification of epitopes sequences of gluten peptides responsible for wheat-related diseases (Sollid et al., 2012; Uvackova et al., 2013a, 2013b). In the same way, open-source curated databases built from *in silico* digestions assist the identification of epitopes and peptides of prolamins associated with CD and other cereal-related disorders (Juhász et al., 2015a). Moreover, the

use of specific database can improve the targeted proteomic workflows and the development of methods to detect and quantify gluten proteins (Bromilow et al., 2017).

In this context, proteomic approaches based on reliable and reproducible methods such as the label-free quantitative method (MS^E) based on data-independent acquisition (DIA) have been lately applied to characterize different wheat proteins and peptides (Victorio et al., 2018; Bromilow et al., 2016; Uvackova et al., 2013a, 2013b). In this method, the *quasi*-simultaneous fragmentation of multiple precursor ions regardless of intensity, lead to a more comprehensive sequence coverage and peptides identification, even in less abundant protein fractions of wheat endosperm (Victorio et al., 2018). Modern multiplex techniques such as Ultra Definition MS^E ($UDMS^E$) promotes the addition of another dimension of peptide separation through the ion mobility system (IMS), where homologous peptides can be separated through their different spatial conformations, reducing interferences and increasing the selectivity of precursor (Victorio et al., 2018). In this work, allergenic/immunogenic peptides were monitored for the first time by using coupled proteomic techniques, nanoUPLC-MS^E and nanoUPLC-$UDMS^E$, in both soluble and gluten protein fractions of nine wheat flours from different technological qualities.

2. Material and methods

2.1. Samples

Samples from nine Brazilian wheat (*Triticum aestivum*) genotypes were kindly provided by OR Melhoramento de Sementes (ORS) Ltda (Passo Fundo, RS, Brazil). Wheat flours were previously characterized by physical tests based on gluten force (AACC International Method 54-30.02), dough stability and water absorption (AACC International Method 54-21.02). The flours were then classified into different technological qualities as Campeiro, ORS25 and ORS27 were classified as low (LW) (gluten force: < $219x10^{-4}$J); Marfim, ORS1401, ORS1402 as medium (MD) (gluten force: 220-$299x10^{-4}$J); and Ametista, Guabiju and Jadeíte 11 varieties as superior (SP) (gluten force: > $400x10^{-4}$J).

2.2. Chemicals

Ammonium bicarbonate (NH_4HCO_3) (PubChem CID: 14013), β-mercaptoethanol (PubChem CID: 1567), bovine serum albumin (BSA) (PubChem CID: 16132389), Bradford reagent (p/n B6916 Sigma-Aldrich), 1,4-dithiothreitol (DTT) (PubChem CID: 19001),

Iodoacetamide (IAM) (PubChem CID: 3727), sodium dodecil sulfate (SDS) (PubChem CID: 3423265), Tris-HCl (PubChem CID: 93573), trifluoroacetic acid (TFA) (PubChem CID: 6422) as well as acetonitrile LC–MS grade (PubChem CID: 6342) and formic acid (PubChem CID: 284) for LC–MS solvents were purchased from Sigma Aldrich (St. Louis, USA). RapiGest Surfactant (p/n 186002118 Waters) and [Glu1]-Fibrinopeptide B human (GFP) (PubChem CID: 16172403) were purchased from Waters Corp. (Milford, MA), while trypsin Gold (p/n V5280 Promega) and trypsin/Lys-C (p/n V5071 Promega) of mass spectrometry grade for enzymatic digestion were bought from Promega Corporation (Madison, USA).

2.3. Sequential extraction of proteins: albumin/globulin and gluten fractions

The extraction of the soluble proteins, albumins and globulins (AG), was performed according Victorio et al. (2018). Briefly, 1 mL of Tris-HCl buffer (80 mM, pH 8.0) and Iodoacetamide (IAM) (40mM) was added to 100 mg of samples. Samples were incubated in a shaker (60 min, 200 rpm, 25 °C) (TE-420, Tecnal, Brazil) and then centrifuged (10 min, 10.600 x g, 20 °C) (Heraeus Megafuge 16R Centrifuge, Thermo Fisher Scientific, Germany). The extraction was performed twice, the respective supernatants were merged together and gluten proteins (GLU) extraction proceeded with the pellets.

Pellets were ressuspended in 1 mL of PBS buffer (pH 7.4), 0.5% SDS and 2% β-mercaptoethanol (Martinez-Esteso et al., 2016). The samples were incubated (14h, 200 rpm, 25 °C), centrifuged (15 min, 10,600 x g at 4 °C) and then submitted to concentration and buffer exchange steps by using Amicon 3kDa filters (Amicon, Millipore, Ireland). Three washes with NH_4HCO_3 (50 mM, pH 8.5) were carried out with centrifugation step between them (14,000 x g, 8 °C, 60 min).

2.4. Quantification of the protein extracts and pooling samples

After extraction, the protein content of each extract was determined by Bradford method (1976) using BSA as standard with concentration between 1.5 and 0.125 $mg.mL^{-1}$. Samples were diluted and 20 µL of sample was added in triplicate to 1 mL of Bradford reagent. After 5 min of reaction, the absorbance was measured at 595 nm in a spectrophotometer (UV-2700, Shimadzu, Japan). Then, the extracts were diluted in 50 mM NH_4HCO_3 (pH 8.5) to reach the final concentration of protein of 1 µg/µL. Pools of protein extracts were performed according

to the technological qualities (LW, MD and SP). Hence, from 20 µL of each sample, three pools of AG and GLU were built with final volume of 60 µL each one.

2.5. Samples digestion and preparation for NanoUPLC-MS^E analysis

To an aliquot containing 60 µg of protein within each extract were added 25 µL of 0.2% v/v RapiGest SF solution, heated (80 °C, 15 min) and centrifuged (14.000 x g; 10 min; 4 °C). Afterward 2.5 µL of 100 mM DTT (prepared in 50 mM NH_4HCO_3 pH 8.5) were added to sample, incubated (30 min; 60 °C) and centrifuged (14.000 x g; 10 min; 4 °C). Then, 2.5 µL of 300 mM IAM (prepared in 50 mM NH_4HCO_3 pH 8.5) was added to the samples and remained at 20 °C in the dark for 30 min. Samples were centrifuged in microcentrifuge spin for 30s and 20 µL of trypsin (in the case of AG extracts) or trypsin/LysC (in the case of gluten extracts), both prepared in NH_4HCO_3 (50 mM; pH 8.5) were added (1:60 enzyme:protein ratio). The samples were homogenized and incubated in a heater block (12 h, 37 °C). After digestion, 10 µL of 5% TFA were added to stop proteolysis, incubated (90 min; 37 °C) and centrifuged (14.000 x g; 90 min; 4 °C). All the supernatants were transferred, centrifuged (14,000 x g; 60 min; 4 °C) and placed into total recovery vials (Waters, USA) for proteomic analysis. Vials were stored in ultrafreezer (-80 °C) (IULT 335 D, Indrel, Brasil) until data acquisition.

2.6. Peptides analysis by NanoUPLC-MS^E

The nanoUPLC analysis of tryptic peptides was performed using nanoACQUITY UPLC system (Waters Corp., Milford, MA) coupled to Synapt G2-S High Definition Mass Spectrometer (HDMS) (Waters Corp., Manchester, UK) equipped with a travelling-wave ion mobility (TWIM) separation device and hybrid analyzers such as a quadrupole/ion mobility mass spectrometry/orthogonal acceleration time-of-flight (Qq-IMMS-oaTOF) geometry.

All analyses were carried out using nano-electrospray ionization in the positive ion mode nanoESI(+) and a NanoLockSpray ionization source (Waters Corp., Manchester, UK). The lock mass was derived from an auxiliary pump using a flow rate of 250 nL.min^{-1} at a concentration of 200 fmol.uL^{-1} of GFP of the NanoLockSpray source. The lock mass channel was sampled at a frequency of 30 s. The time-of-flight analyzer of the mass spectrometer was calibrated with a MS/MS spectrum of GFP. The final instrument calibration was obtained by the specific y+ and b+ fragments obtained by the double charged precursor ion $[M + 2H]^{2+}$ =

785.8426 signal. For all measurements, the mass spectrometer was operated in the resolution mode with a typical *m/*z resolving power of at least 35.000 FWHM and an ion mobility cell filled with helium gas. The ion source block temperature was set to 70 ºC and capillary voltage was set to 2.8 kV.

The chromatographic system was equipped with a nanoEase BEH130 C18 (5 µm, 180 µm x 20 mm) TRAP column (Waters Corp., USA) and a nanoAcquity HSS T3 C18 (1.8 µm, 100 µm x 100 mm) reversed-phase column (Waters Corp., USA). Mobile phase A consisted of ultra-pure water (Milli-Q®, Millipore) and mobile phase B consisted of acetonitrile, both added of 0.1% formic acid (FA). The analytical column temperature was maintained at 55 °C and the sample manager temperature was 8 °C. Prior to analysis, stoichiometric measurements based on scouting runs were performed with injection of 1 µL of each extract to ensure standardized molar values across all conditions based on the integrated total ion account (TIC) area. For the scouting runs, chromatographic conditions were applied as follow: 0 min – 7% phase B; 0-7.20 min – 40% phase B; 7.20-8.51 min – 85% phase B; 8.51-11.13 min – 85% phase B; 11.13- 12.44 min – 7% phase B, with run time of 20 min and flow rate of 600 $nL.min^{-1}$. From these results, the volume of injection of each sample was adjusted.

2.6.1. AG peptides analysis by NanoUPLC-UDMS^{E}

For the AG extracts, multiplexed DIA scanning with added specificity and selectivity of a non-linear TWIM (HDMS^{E}/UDMS^{E}) experiments were performed according to previous described by Victorio et al. (2018) (see chromatograms in Supplementary Fig. 1). The chromatographic methods with the following gradient were carried out: 7-40% of mobile phase B in 91.31 min, followed by a cleaning column gradient of 40 to 85% of B for 1 column volume (cV); maintained in 85% of B for 2 cV; then from 85 to 7% of B in 1 cV. The flow rate was 600 $nL.min^{-1}$. The exact mass retention time (EMRT) signals from multiplexed ion-mobility DIA scanning (UDMS^{E}) were collected in an alternating low energy and elevated energy acquisition mode. In the low energy mode, data were collected at 6 eV. In the elevated collision energy, quasi *m/*z-specific collision energies were applied to the different drift time bins (collected in the HDMS^{E} acquistion) and were used to fragment precursor ions prior to orthogonal acceleration time-of-flight (oa-TOF) analysis, applied to the transfer travelling-wave (T-wave), collision-induced dissociation (CID) cell filled with argon gas. A miliseconds scan time were previously adjusted based on the linear velocity of the

chromatography peak delivered through nanoACQUITY UPLC to get a minimum of 20 scan points for each single peak, both in low energy and at high-energy transmission at an oa-TOF from *m/z* 50 to 2000.

2.6.2. GLU peptide analysis by NanoUPLC-MS^E

The samples were acquired only by MS^E multiplex mode, applying simultaneously low and high energy fragmentation (15-55V) (see chromatograms in Supplementary Fig. 2). The flow of 600 nL/min of mobile phases in the following gradient: 7-60% of mobile phase B in 91.31 min, followed by a cleaning column gradient of 60 to 85% of B for 1 column volume (cV); maintained in 85% of B for 2 cV; then from 85 to 7% of B in 1 cV. The flow rate was 600 nL.min^{-1}. Ions were acquired with *m*/z between 50-2000, scanning time of 0.5 s, cone voltage of 30 V, capilar voltage of 2.7 kV and source offset voltage of 30 V.

2.7. Data and Statistical analysis

The software Protein Lynx Global Server v3.0.1 (PLGS) (Waters Corp., Manchester, UK) was used to identify and relatively quantify the peptides from a database of Triticum aestivum from UniProt KB 2016_07v (http://www.uniprot.org). To the original database, a reversed protein database was created and appended to monitor the false discovery rate (FDR) during the queries. The parameters for database searching were tryptic peptides with only one missed cleavage allowed, carbamidomethylation (Cys-CAM) as fixed modification and oxidation of methionine (MetO) as variable modification. For peptides and proteins assignments the following filters were used: 3 of the 3 technical replicates (3/3 replicates), pass 1 only, score more than 8.15/raw score, error < 20 ppm, and maximum FDR of 4%. Peptide lists were used to final sequence identifications. Relative quantification was determined from the absolute intensities with the use of ion accounting Hi3 (Top3) based quantitation method, as described(Silva et al., 2006). All the informations about peptides and proteins identified in each replicate is available as Supplementary material (excel sheet), i.e., the complete protein and peptides list containing accession code, protein description, ion counting, score, error (ppm).

To identify the immunogenic/allergenic GLU peptides, the remaining peptides (after filtering) were searched into a prolamin-specific database (ProPepper®) (Juhász et al., 2015a). In the case of the AG fraction, a list of major allergens related to the wheat allergies

was created from different studies (Larre et al., 2011; Lupi et al., 2013; Lupi et al., 2014; Rogniaux et al., 2015). Then, this list (Supplementary Table 1) material containing 72 proteins was confronted to our results to identify peptides and proteins in our samples.

The results were expressed as the mean of ion abundance intensities measured of each peptide ± SD within the replicates available, and the *p* value ($p < 0.05$) calculated using GraphPad software to refer to the differences between the flour samples.

3. Results and discussion

3.1. Identification and quantification of immunogenic and allergenic gluten peptides

For the identification and analysis of peptide and immune-responsive epitopes from the GLU extracts, a curated database currently containing 2,484 unique and complete protein sequences and 37,914 unique peptide sequences obtained with single and multi-enzyme *in silico* digestion, as well as a list of linear epitopes responsible for wheat-related food disorders were used in this study (Juhasz et al., 2015). The optimized enzymatic system (trypsin-LysC) used for the GLU protein extracts is appointed to improve prolamin digestion and was also taken account in this database (Juhasz et al., 2015). Moreover, in the present work only 3% of identified GLU peptides presented missed cleavage, which indicates efficiency of digestion (shown in supplementary Figure 3). The low percentage of basic residues in the gluten proteins, associated with high levels of proline, can lead to an inefficient digestion, resulting in large peptides, and consequently more difficult to be identified by mass spectrometry analysis (Ferreira et al., 2014; Vensel et al., 2011).

Globally, in all replicates 3,603 peptides were identified from the GLU extracts in all samples. After filtering, basically taking account minimum score, only peptides found in all 3 replicates, error less than 10 ppm and considering absence of reverse (0% FDR), around 150 peptides were identified for each sample. From these, about 19% of the identified peptides shown immunogenic action associated with prolamins, representing 29 peptides. Amongst them, 10 peptides were common to the 3 samples (LW, MD and SP flours) and about 12 peptides were shared by two of them as shown in the Venn diagram (Fig. 1). Only 4 unique peptides were found in SP, 3 in MD and 5 in LW flour samples.

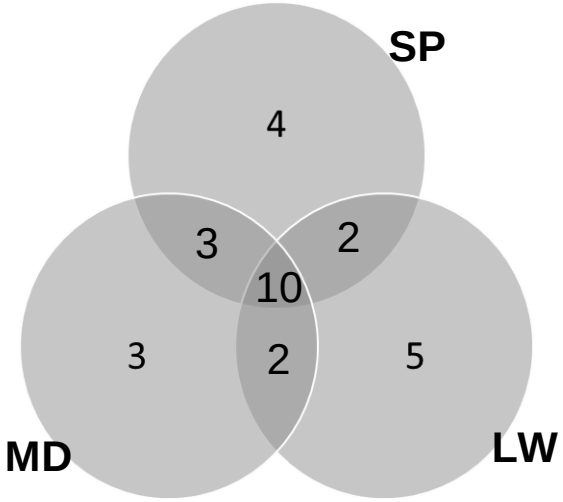


Figure 1. Venn diagram representing the number of gluten peptides identified in each sample.

The allergenic and CD immunogenic peptides from GLU extracts identified in this study and their 19 proteins of origin are presented in the Table 1. Despite the literature associates the biggest toxicity potential to peptides from gliadins, especially to the D-genome encoding α2-gliadins that contains the most immunodominant gluten peptide called 33-mer peptide (LQLQPFPQPQLPYPQPQLPYPQPQLPYPQPQPF) (Camarca et al., 2009; Shan et al., 2005), in this work glutenin subunits were the major proteins associated with immunogenic peptides in GLU fraction. These peptides are mainly related to trigger CD and other allergies, such as WDEIA allergy and atopic dermatitis (Table 1). Similarly, Uvackova et al. (2013a) identified 15 wheat allergen protein isoforms in nine samples of milled wheat grains, where the most abundant proteins extracted to the isopropanol-Gliadin fraction were HMW-GS and γ-gliadin and HMW and LMW-GS were the most abundant proteins in the isopropanol-DTT-glutenins fraction. The 33-mer is considered a model peptide to study CD immune response, since it contains three overlapping T-cell epitopes (PFPQPQLPY, PYPQPQLPY, PQPQLPYPQ), which result in a strong reactivity (Shan et al., 2005; Sollid et al., 2012). According to a BLAST search within 897 entries for α-gliadins from *Triticinae* in the UniProtKB database, the amino acid sequence of the 33-mer was found in 16 protein sequences from *T. aestivum* and in three from *T. spelta* with an identity of 100% (Schalk et al., 2017). In this work, the complete 33-mer was present in the (K7X1I9) and its epitopes were identified in a peptide from α-gliadin (X2KWL1) (Supplementary Fig. 4). The α-gliadin presenting the 33-mer was identified only in LW samples, while the α-gliadin carrying those epitopes was identified in both LW and SP samples.

Table 3. Peptides sequence and proteins of origin from the gluten peptides identified.

Nº	**Peptide sequence**	**Protein type (acession number)**	**Disease**
1	QQDQQSGQGQQPGQR	HMW-GS (A0A060MZP1)	WDEIA
	QQPGQGQQLR (PEP1)	HMW-GS (A0A060MZP1)	WDEIA
	GQQGQQSGQGQQLGQGQQGQQPGQK (PEP2)	HMW-GS (A0A060MZP1)	WDEIA
	YYPTSSQQPGQLQQLAQGQQGQQPER (PEP3)	HMW-GS (A0A060MZP1)	WDEIA
2	QGQQLEQGQQPGQGQQTR	HMW-GS 1By9 (Q03871)	WDEIA
	QQPGQGQQIGQGQQLGQGR (SP3)	HMW-GS 1By9 (Q03871)	WDEIA
	QQPGQGQQPEQGQQPGQGQQGYYPTSPQQPGQGK (PEP4)	HMW-GS 1By9 (Q03871)	CD, WDEIA
3	QQGYYPTSPQQPGQGQQLGQGQPGYYPTSQQPGQK	HMW-GS x-type Bx7 (Q1KL95)	CD, WDEIA
	QQSGQGQQPGQGQQSGQGQQGQQPGQGQQAYYPTSSQQSR (PEP8)	HMW-GS x-type Bx7 (Q1KL95)	WDEIA, ALLERGY
	YYPSVTSSQQGSYYPGQASPQQSGQGQQPGQEQQPGQGQQDQQPGQR (PEP9)	HMW-GS x-type Bx7 (Q1KL95)	WDEIA
4	**QGQQGQQSGQGQPR** (MD1)	HMW-GS Ax2 (Q41553)	WDEIA
	YYPTSPQQPGQEQQPR	HMW-GS Ax2 (Q41553)	WDEIA
5	**QGQQPGQGQQGYYPTSSQQPGQGK** (LW1)	HMW-GS 1By16 (A5HMG2)	WDEIA
6	**QQPGQGQHPEQGK** (LW4)	HMW-GS Y-type (Q0Q5D3)	WDEIA
7	GQQGYYPTSLQQPGQGQQGYYPTSLQHTGQR (PEP5)	HMW-GS Y-type (Q0Q5D3)	WDEIA
	QGSYYPGQASPQQPGQGQQPGK (PEP6)	HMW-GS Y-type (Q0Q5D3)	WDEIA
	QQPVQGQQPEQGQQPGQWQQGYYPTSPQQLGQGQQPR (PEP7)	HMW-GS Y-type (Q0Q5D3)	CD, WDEIA
8	VFLQQQCSPVAMPQSLAR	LMW-GS (Q6J160)	ALLERGY
9	ETFPQQPPSSQQQQPFPQQPPFLQQQPSFSQQPLFSQK	LMW-GS 0154A5-M (B2BZC7)	CD, AD
10	**QLPQIPEQSR** (SP4)	LMW-GS (Q1ZZT4)	CD
11	**NLALQTLPAMCNVYIPPYCTIAPFGIFGTN** (LW2)	Alpha-gliadin (K7X1I9)	CD, ALLERGY
12	**NLALQTLPAMCNVYIPPHCSTTIAPFGIFGTN** (LW5)	Alpha-gliadin (K7X1L1)	CD, ALLERGY
13	QQPSSQVSFQQPQQQYPSSQVSFQPSQLNPQAQGSVQPQQLPQFAEIR	Alpha-gliadin (X2KWL1)	CD, ALLERGY
14	**NLALQTLPAMCNVYIPPYCSTTIAPFGIMSTN** (SP2)	Alpha/beta-gliadin (I0IT53)	CD, ALLERGY
15	**QPQQPFPQPQQPQQSFPQQQPSLIQQSLQQQLNPCK** (LW3)	Gamma-gliadin B (P06659)	CD, WDEIA, ALLERGY, AD
16	**SDQPQQSFPQQPQQK** (SP1)	Gliadin/avenin-like seed protein (D2KFH0)	CD
17	QLVQIPEQAR (PEP10)	Putative uncharacterized protein (Q7X9M4)	CD
18	**QLSQIPEQFR** (MD3)	Avenin-like protein (V5M127)	CD
19	**QGFGQSQQQQQPGQR** (MD2)	Avenin-like protein (G9I0T6)	WDEIA

Bold: peptides present only in one sample, underlined: common peptides to the three samples. Celiac disease (CD), wheat-induced exercise-induced anaphylaxis (WDEIA), allergies and atopic dermatitis (AD).

From the spectra ion counting, the average of total ion abundance of all allergenic peptides was calculated for each replicate of samples. The LW flour samples showed the highest immunogenic potential, being around 27% higher than MD and SP flour samples (Fig. 2). This result reinforces the observation about the presence of the 33-mer only in this sample.

The highest abundance of allergenic peptides found in the pool sample of LW quality flours can be associated to greater amounts of gliadin. Song and Zheng (2008) showed that greater amounts of gliadin in wheat flour lead to a decrease of mixing time and dough stability, as a consequence of a weakening of the protein network, resulting in weak doughs, suitable for biscuit production, for example. Indeed, one of the current strategies to develop food products directed to celiac patients is to reduce the gliadin content. It has been supported the use of wheat lines lacking in α-gliadins encoded by chromosome 6A and 6D could contribute to the production of wheat with reduced coeliac toxicity (Shewry & Tatham, 2016). However, the challenge remains is not loose dough functionality. Gil-Humanes et al. (2014) proposed a near gliadin-free bread with acceptable baking, nutritional and organoleptic quality.

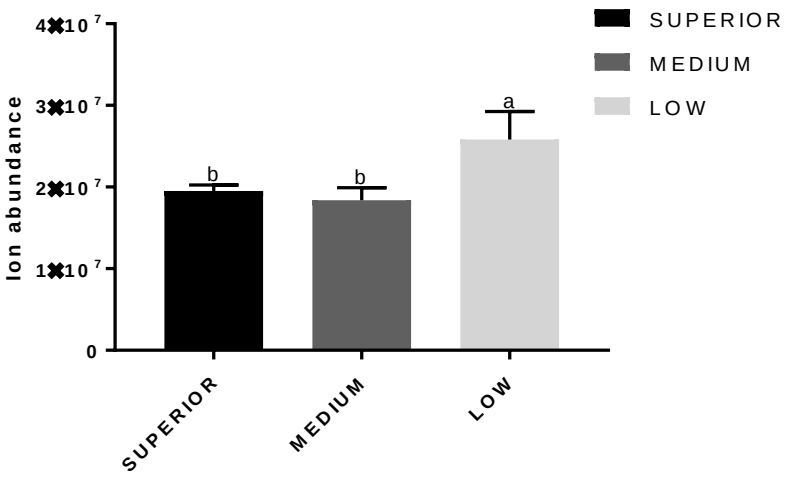


Figure 2. Total ion abundance of all 29 allergenic peptides in each pool sample of flours.

Ten peptide sequences that exhibit immunogenicity were commonly found in the three samples (SP, MD, LW) (Fig 3). Nine of them originated from glutenin subunits and one from an uncharacterized protein (Table 3). The same profile of allergenic peptides abundance was found with these results, LW samples presented a higher abundance of allergenic peptides than MD and SP samples that presented similar profiles (Fig. 3B).

The ProPepper® database provides complete information about the peptides, allowing the recognition of immunogenic or non-immunogenic peptides and therefore the number of epitopes associated. Epitopes are defined as the smallest portion of the antigen that are able to generate the immune response and, therefore, in this context, responsible for wheat-related eating disorders (Juhász et al., 2015a). From the ten common peptide sequences, 22 epitopes were identified.

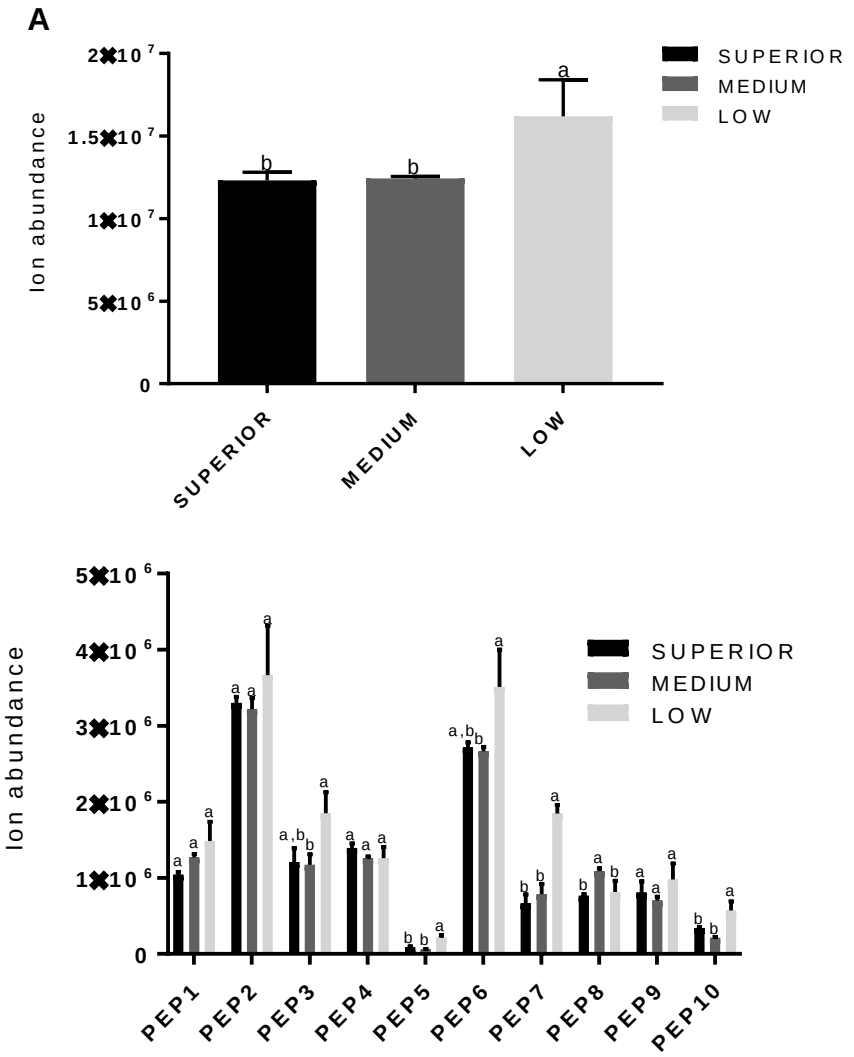


Figure 3. Abundance of the common peptides found in all samples. A. Sum of the abundances and B. Abundance of each one of the 10 common peptides.

Some peptides were exclusively identified only in one of the samples (Fig. 4). From these results, it is observed that the LW flours showed 5 unique peptide sequences and 28

epitopes in total; while the MD flours presented 3 unique peptide sequences and 3 epitopes in total and the SP flours had 4 unique peptide sequences and 6 epitopes in total. It is possible to note that even when the unique peptides are analyzed, the weak flours (LW) seem to have the highest allergenic potential due to the highest number of epitopes. When analyzing the expression of the peptides, SP flour samples although presenting fewer unique peptides and epitopes (compared to the LW samples) showed a higher expression due to two peptides sequences with greater abundance, being approximately 45% higher than the third most abundant peptide sequence of LW sample (LW3). The unique peptides found in each sample could be also useful to help the selection of possible peptides biomarkers related to celiac toxicity (e.g. LW2, LW5, SP1, SP2, SP4) and WDEIA (LW1, LW4, MD1 and SP3).

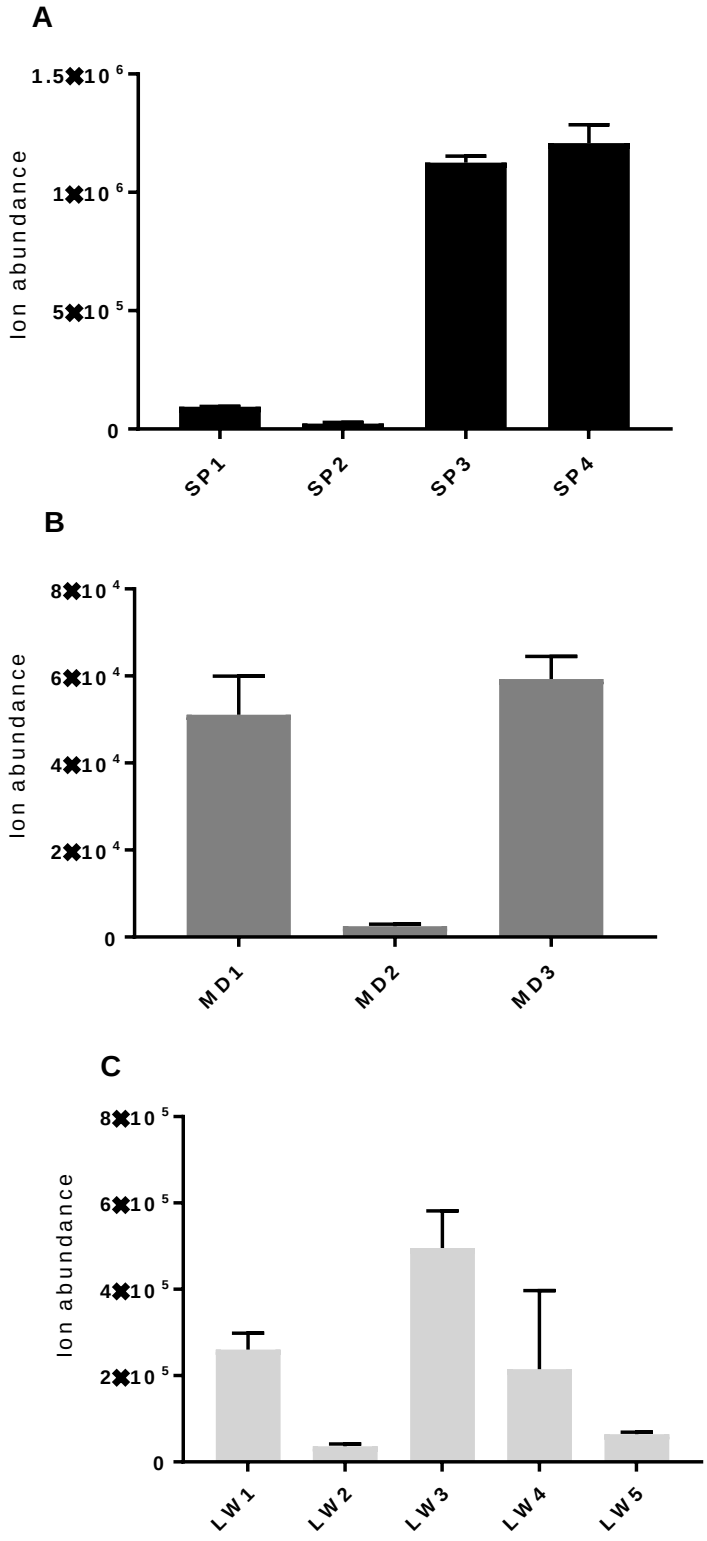


Figure 4. Relative abundance of the unique immunogenic gluten peptides found in each sample (SP, superior; MD, medium and LW, low).

From all the identified allergenic peptides (29 peptides), it was possible to indicate the type of disease triggered by each specific peptide (Table 3). Hence, it can be observed that the most prevalent disease is celiac disease (CD), followed by wheat-induced exercise-induced anaphylaxis (WDEIA), allergies and atopic dermatitis (AAD) (Fig. 5). There is also the triggering of two diseases by the same epitope, such as celiac disease + atopic dermatitis and WDEIA + allergy, but in a smaller percentage, together representing 6%.

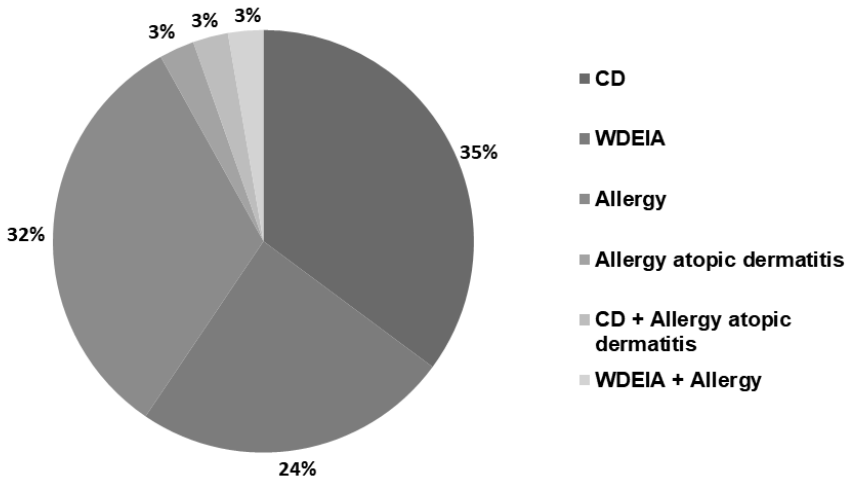


Figure 5. Percentage distribution of diseases triggered by the epitopes found in the samples.

3.2. Identification and quantification of allergenic proteins from the AG fraction

Collectivelly, in all samples were identified 91,609 peptides from the AG extracts, resulting in 1,708 peptides found in 3/3 technical replicates after filtering. There is a lack of information about peptides sequences and epitopes responsible for the allergies triggering by AG proteins in the literature. Hence, a list of 72 recognized wheat allergens from metabolic soluble proteins were prepared from previous published works, as described before, based on two databases, the Structural Database of Allergenic Proteins (SDAP, http://fermi.utmb.edu) and the WHO/IUIS allergen databases (http://www.allergen.org/index.php). Therefore, in the present work, 13 allergenic proteins were identified in AG fraction, as presented in Table 4.

Table 4. Abundance of allergenic proteins from the albumin and globulin fractions.

Protein (Accession Number)	**Average Ion Abundance**					
	SP	**CV**	**MD**	**CV**	**LW**	**CV**
α-amylase inhibitor 0.28 (P01083)	$2.55 \times 10^{+6\,a}$	43.3	$2.67 \times 10^{+6\,a}$	2.1	$2.67 \times 10^{+6\,a}$	10.4
α-amylase inhibitor 0.19 (P01085)	$3.67 \times 10^{+6\,a}$	13.4	$3.05 \times 10^{+6\,a}$	14.7	$4.11 \times 10^{+6\,a}$	11.2
α-amylase/trypsin inhibitor CM1 (P16850)	$9.31 \times 10^{+5\,a}$	27.5	$9.97 \times 10^{+5\,a}$	4.4	$8.94 \times 10^{+5\,a}$	5.0
α-amylase/trypsin inhibitor CM2 (P16851)	$1.55 \times 10^{+6\,a}$	15.5	$1.50 \times 10^{+6\,a}$	28.5	$1.60 \times 10^{+6\,a}$	14.6
α-amylase/trypsin inhibitor CM3 (P17314)	$5.78 \times 10^{+6\,a}$	7.6	$4.83 \times 10^{+6\,a}$	20.7	$5.48 \times 10^{+6\,a}$	11.7
Serpin-Z1A (Q41593)	$8.86 \times 10^{+5\,a,b}$	14.7	$7.58 \times 10^{+5\,b}$	3.9	$9.46 \times 10^{+5\,a}$	1.9
Serpin-Z2A (Q9ST57)	$5.85 \times 10^{+5\,a}$	7.0	$4.30 \times 10^{+5\,b}$	5.1	$5.19 \times 10^{+5\,a}$	1.6
Serpin-Z1B (P93693)	$9.34 \times 10^{+5\,a,b}$	16.1	$8.42 \times 10^{+5\,b}$	6.2	$1.10 \times 10^{+6\,a}$	2.3
Serpin-Z2B (P93692)	$5.59 \times 10^{+5\,b}$	11.9	$5.32 \times 10^{+5\,b}$	9.6	$7.56 \times 10^{+5\,a}$	5.1
Serpin-Z1C (Q9ST58)	$6.25 \times 10^{+5\,b}$	21.4	$6.16 \times 10^{+5\,b}$	2.4	$8.07 \times 10^{+5\,a}$	0.6
Superoxide dismutase (P93606)	$2.79 \times 10^{+5\,a}$	9.1	$2.75 \times 10^{+5\,a}$	11	$2.18 \times 10^{+5\,a}$	5.0
1-Cys peroxiredoxin PER1 (Q6W8Q2)	$1.26 \times 10^{+6\,a,b}$	2.1	$1.34 \times 10^{+6\,a}$	6.6	$1.16 \times 10^{+6\,b}$	2.4
Glyceraldehyde-3-phosphate dehydrogenase (C7C4X1)	$1.02 \times 10^{+6\,a}$	1.9	$9.44 \times 10^{+5\,a,b}$	7.4	$8.02 \times 10^{+5\,b}$	7.5
Total ion abundance	**$2.06 \times 10^{+7}$**	**13.7**	**$1.88 \times 10^{+7}$**	**7.8**	**$2.11 \times 10^{+7}$**	**3.2**

The major allergenic proteins were associated to enzymatic inhibitors, such as α-amylase/trypsin inhibitors and serpins, representing, respectively, 70% and 18% of total ion abundance in the samples. These results corroborate the findings of Pastorello et al. (2007) and Šotkovský et al. (2008), that showed different types of serpins and α-amylase/trypsin inhibitor family as the most allergenic proteins associated to the AG fraction. Moreover, Larre et al. (2011) characterized allergen polypeptides related to wheat food allergy in the salt-soluble protein fraction and found similar results to the previously published in the literature, identifying α-amylase inhibitors, β- amylase, serpins and other proteins considered to be potent allergens. Rogniaux et al. (2015) used targeted mass spectrometry approach to compare the relative abundance of allergens in the AG fraction of seven wheat varieties, two hexaploids, three tetraploids and two diploids. Twelve allergens have been quantified in these varieties, between them peptides from the AAI (alpha-amylase inhibitors) family, being them

expressed at lower levels in the diploid species. This kind of result can explain a smaller allergenicity associated with diploid wheat varieties (Spaenij–Dekking et al., 2005).

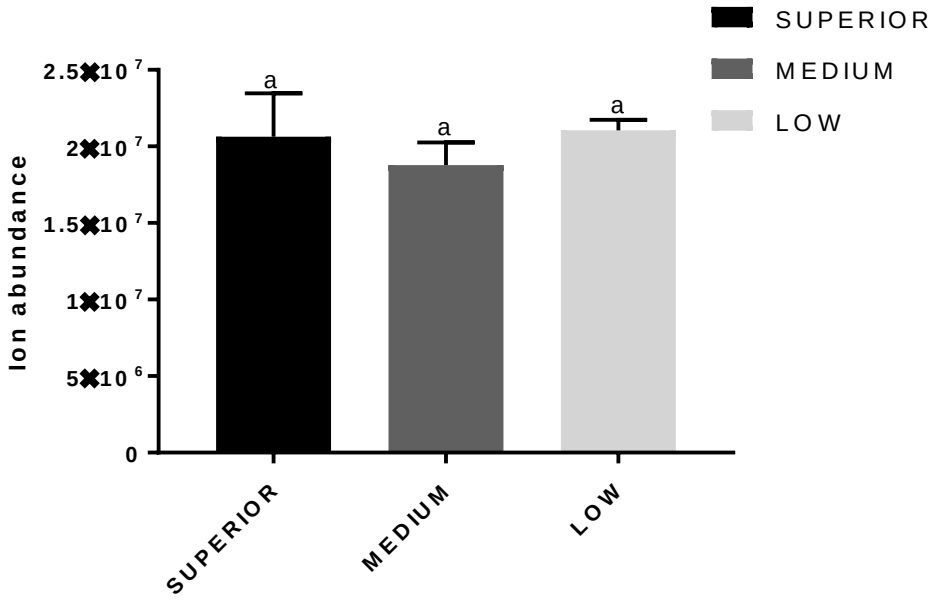


Figure 6. Total abundance of immunogenic peptides from the albumin fraction in each pool.

No significantly difference was found between the samples concerning the abundance of allergenic proteins identified in the flour samples (Fig. 6). This result could be expected since the flours were previous grouped based on technological quality. Although, recently some proteins of the soluble fraction, albumins and globulins, were pointed as possible biomarkers of technological quality (Victorio et al., 2018), these classification is essentially based on gluten force. Indeed, in this recent proteomic study, the authors found two of these AG proteins (Serpin-Z1A and Serpin-Z1C), differentially expressed between LW, MD and SP flours. In the present work, three proteins Serpin-Z1A, Serpin-Z1C and Serpin-Z2B were respectively 15%, 30% and 28% higher expressed in LW samples when compared to MD and SP samples (Table 4), corroborating the results found for GLU extracts.

4. Conclusions

To the best of our knowledge this is the first paper reporting on the identification and relative quantification of allergenic and immunogenic peptides from both fraction of wheat proteins (prolamins and non-prolamins) applying modern MS based methods in wheat flours of different technological qualities.

In this approach, 29 allergenic peptides carrying a high number of epitopes were identified from 19 different protein subunits in gluten extracts. The most part of immunogenic peptides identified belonged to glutenin subunits, especially HMW-GS. Overall, 12 unique peptides were identified belonging only to one sample. This result can be useful to help the selection of peptides biomarkers related to celiac toxicity and also WDEIA.

From a list of 72 recognized allergens of AG proteins, 13 proteins were identified in this study, mainly α-amylase inhibitors and serpins. The total ion abundance of these proteins was similar between the different samples. However, LW flours showed the highest expression of two isoforms of serpins.

Taken together, the results of this work showed that there is a variability of allergen expression between the wheat flours of different technological qualities. LW samples presented a distinguished profile, showing the highest expression of peptides and proteins related to CD and other wheat allergies and the highest number of epitopes associated to CD.

Since strongly dependent of protein curated database, the combined use of ELISA techniques with modern LC-MS techniques can be used to unravel new allergens from wheat proteins. Moreover, as future works, the influence of cereal-based food processing and hydrolysis over the allergenic peptides can also be evaluated, to improve the knowledge about this topic.

5. Acknowledgements

The authors would like to thank Prof. Luiz Carlos Gutkoski (UPF. Brazil) for the rheological characterization of wheat flours and OR Melhoramento de Sementes Ltda for providing samples. This work was supported by the Foundation for Research Support of State of Rio de Janeiro (FAPERJ) (26/111.484/2014; 26/203.312/2015). National Council for the Scientific and Technological Development (CNPq) (448171/2014-7). Coordination for Improvement of Personnel with Higher Education (CAPES) and Federal University of the State of Rio de Janeiro (UNIRIO).

REFERÊNCIAS

AACC International. Approved Methods of Analysis. 11th ed. Method 54-30.02. – Alveograph Method for Soft and Hard Wheat Flour. St Paul, MN: AACC International.

AACC International. Approved Methods of Analysis. 11th ed. Method 54-21.02. – Rheological Behavior of Flour by Farinograph: Constant Flour Weight Procedure. St Paul, MN: AACC International.

AACC International. Approved Methods of Analysis. 11th ed. Method 56-81.03. – Determination of Falling Number. St Paul, MN: AACC International.

ABITRIGO. (2017). Evolução mensal e anual do preço do trigo, consumo, produção e estoque mundial de trigo. Retrieved from http://www.abitrigo.com.br/estatisticas-trigo.php

Alves, T. O., D'Almeida, C. T. S., & Ferreira, M. S. L. (2017). Determination of Gluten Peptides Associated with Celiac Disease by Mass Spectrometry *Celiac Disease and Non-Celiac Gluten Sensitivity*: InTech.

Arendt, E., & Dal Bello, F. (2008). *Gluten-free cereal products and beverages.*: Elsevier Science & Technology.

Banerjee, B. (2010). *Nutritional management of digestive disorders*: CRC Press.

Belitz, H. D., Grosch, W., & Schieberle, P. (2009). *Food chemistry, 4th revised and extended edn* (4 ed.).

Bittner, C., Grassau, B., Frenzel, K., & Baur, X. (2008). Identification of wheat gliadins as an allergen family related to baker's asthma. *Journal of Allergy and Clinical Immunology, 121*(3), 744-749.

Bongiovanni, T. R. S., Clark, A. L., Garnett, E. A., Wojcicki, J. M., & Heyman, M. B. (2010). Impact of gluten-free camp on quality of life of children and adolescents with celiac disease. *Pediatrics, 125*(3), e525-e529.

Bouchez-Mahiout, I., Snégaroff, J., Tylichova, M., Pecquet, C., Branlard, G., & Laurière, M. (2010). Low molecular weight glutenins in wheat-dependant, exercise-induced anaphylaxis: allergenicity and antigenic relationships with omega 5-gliadins. *International archives of allergy and immunology, 153*(1), 35-45.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry, 72*(1-2), 248-254.

Bromilow, S., Gethings, L. A., Buckley, M., Bromley, M., Shewry, P. R., Langridge, J. I., & Mills, E. N. C. (2017). A curated gluten protein sequence database to support development of proteomics methods for determination of gluten in gluten-free foods. *Journal of proteomics*.

Burkhardt, J. G., Chapa-Rodriguez, A., & Bahna, S. L. (2017). Gluten sensitivities and the allergist: Threshing the grain from the husks. *Allergy*.

Camarca, A., Anderson, R. P., Mamone, G., Fierro, O., Facchiano, A., Costantini, S., . . . Sette, A. (2009). Intestinal T cell responses to gluten peptides are largely heterogeneous: implications for a peptide-based therapy in celiac disease. *The journal of Immunology, 182*(7), 4158-4166.

Castiglioni, S., Zuccato, E., Crisci, E., Chiabrando, C., Fanelli, R., & Bagnati, R. (2006). Identification and measurement of illicit drugs and their metabolites in urban wastewater by liquid chromatography-tandem mass spectrometry. *Analytical Chemistry, 78*(24), 8421-8429.

Chaudhary, N., Dangi, P., & Khatkar, B. S. (2016). Relationship of molecular weight distribution profile of unreduced gluten protein extracts with quality characteristics of bread. *Food chemistry, 210*, 325-331.

Codex Alimentarius, C. (2006). *Alinorm 06/29/23*. Retrieved from Geneva, Switzerland:

Codex Alimentarius, C. (2008). Codex standard for foods for special dietary use for persons intolerant to gluten. *Codex stan*, 118-1979.

Colgrave, M. L., Byrne, K., Blundell, M., & Howitt, C. A. (2016). Identification of barley-specific peptide markers that persist in processed foods and are capable of detecting barley contamination by LC-MS/MS. *Journal of proteomics, 147*, 169-176.

Colgrave, M. L., Goswami, H., Blundell, M., Howitt, C. A., & Tanner, G. J. (2014). Using mass spectrometry to detect hydrolysed gluten in beer that is responsible for false negatives by ELISA. *Journal of Chromatography A, 1370*, 105-114.

Colgrave, M. L., Goswami, H., Byrne, K., Blundell, M., Howitt, C. A., & Tanner, G. J. (2015). Proteomic profiling of 16 cereal grains and the application of targeted proteomics to detect wheat contamination. *Journal of proteome research, 14*(6), 2659-2668.

Comino, I., de Lourdes Moreno, M., & Sousa, C. (2015). Role of oats in celiac disease. *World journal of gastroenterology, 21*(41), 11825.

D'Ovidio, R., & Masci, S. (2004). The low-molecular-weight glutenin subunits of wheat gluten. *Journal of Cereal Science, 39*(3), 321-339.

Day, L., Augustin, M. A., Batey, I. L., & Wrigley, C. W. (2006). Wheat-gluten uses and industry needs. *Trends in food science & technology, 17*(2), 82-90.

Dhaka, V., & Khatkar, B. S. (2015). Effects of Gliadin/Glutenin and HMW-GS/LMW-GS Ratio on Dough Rheological Properties and Bread-Making Potential of Wheat Varieties. *Journal of Food Quality, 38*(2), 71-82.

Diaz-Amigo, C., & Popping, B. (2013). Accuracy of ELISA detection methods for gluten and reference materials: a realistic assessment. *Journal of agricultural and food chemistry, 61*(24), 5681-5688.

Distler, U., Kuharev, J., Navarro, P., & Tenzer, S. (2016). Label-free quantification in ion mobility-enhanced data-independent acquisition proteomics. *Nat. Protocols, 11*(4), 795-812. doi:10.1038/nprot.2016.042

Egertson, J. D., MacLean, B., Johnson, R., Xuan, Y., & MacCoss, M. J. (2015). Multiplexed peptide analysis using data-independent acquisition and Skyline. *Nature protocols, 10*(6), 887-903.

Fallahbaghery, A., Zou, W., Byrne, K., Howitt, C. A., & Colgrave, M. L. (2017). Comparison of Gluten Extraction Protocols Assessed by LC-MS/MS Analysis. *Journal of agricultural and food chemistry, 65*(13), 2857-2866.

FAO. (2017). World Food Situation - FAO Cereal Supply and Demand Brief. Retrieved from http://www.fao.org/worldfoodsituation/csdb/en/

FAOSTAT. (2016). *Food and agricultural organization of the United Nations*. Retrieved from http://www.fao.org/worldfoodsituation/csdb/en/

Farage, P., Boas, G. V., Gandolfi, L., Pratesi, R., & Zandonadi, R. P. (2015). Is the consumption of oats safe for celiac disease patients? A review of literature. *Food and Nutritional Disorders, 2014*.

Feillet, P. (2000). *Le grain de blé: composition et utilisation*. Paris: Editions Quae.

Fenn, J. B., Whitehouse, C. M., Dreyer, R. N., & Yamashita, M. (1989). Electrospray ionization for mass-spectrometry of large biomolecules. *Science, 246*(4926), 64-71.

Ferreira, M. S. L., Mangavel, C., Rogniaux, H., Bonicel, J., Samson, M.-F., & Morel, M.-H. (2014). A MALDI-TOF based study of the in-vivo assembly of glutenin polymers of durum wheat. *Food research international, 63*, 89-99.

Fiedler, K. L., McGrath, S. C., Callahan, J. H., & Ross, M. M. (2014). Characterization of grain-specific peptide markers for the detection of gluten by mass spectrometry. *Journal of agricultural and food chemistry, 62*(25), 5835-5844.

Food, U. S., & Drug, A. (2007). Food labeling: gluten-free labeling of foods. *Federal Register, 21*.

Gao, L., Wang, A., Li, X., Dong, K., Wang, K., Appels, R., . . . Yan, Y. (2009). Wheat quality related differential expressions of albumins and globulins revealed by two-dimensional difference gel electrophoresis (2-D DIGE). *Journal of proteomics, 73*(2), 279-296.

Gell, G., Kovács, K., Molnár, I., Bugyi, Z., Tömösközi, S., & Juhász, A. (2015). Celiac disease-specific prolamin peptide content of wheat relatives and wild species determined by ELISA assays and bioinformatics analyses. *Cereal Research Communications, 43*(1), 133-143.

Giannou, V., & Tzia, C. (2016). Addition of vital wheat gluten to enhance the quality characteristics of frozen dough products. *Foods, 5*(6).

Gil-Humanes, J., Pistón, F., Altamirano-Fortoul, R., Real, A., Comino, I., Sousa, C., . . . Barro, F. (2014). Reduced-gliadin wheat bread: an alternative to the gluten-free diet for consumers suffering gluten-related pathologies. *PLoS One, 9*(3), e90898.

Graves, P. R., & Haystead, T. A. J. (2002). Molecular biologist's guide to proteomics. *Microbiology and Molecular Biology Reviews, 66*(1), 39-63.

Gross, B. A. (2004). *BioMEMs.* Paper presented at the IEMBS'04. 26th Annual International Conference of the IEEE.

Hallert, C., Olsson, M., Störsrud, S., Lenner, R. A., Kilander, A., & Stenhammar, L. (1999). Oats can be included in gluten-free diet. *Lakartidningen, 96*(30-31), 3339-3340.

Hernando, A., Mujico, J. R., Mena, M. C., Lombardía, M., & Mendez, E. (2008). Measurement of wheat gluten and barley hordeins in contaminated oats from Europe, the United States and Canada by sandwich R5 ELISA. *European journal of gastroenterology & hepatology, 20*(6), 545-554.

Herrero, M., Simó, C., García-Cañas, V., Ibáñez, E., & Cifuentes, A. (2012). Foodomics: MS-based strategies in modern food science and nutrition. *Mass spectrometry reviews, 31*(1), 49-69.

Hischenhuber, C., Crevel, R., Jarry, B., Mäki, M., Moneret-Vautrin, D. A., Romano, A., . . . Ward, R. (2006). Review article: safe amounts of gluten for patients with wheat allergy or coeliac disease. *Alimentary pharmacology & therapeutics, 23*(5), 559-575.

Hofmann, S. C., Fischer, J., Eriksson, C., Bengtsson Gref, O., Biedermann, T., & Jakob, T. (2012). IgE detection to α/β/γ-gliadin and its clinical relevance in wheat-dependent exercise-induced anaphylaxis. *Allergy, 67*(11), 1457-1460.

Juhász, A., Haraszi, R., & Maulis, C. (2015a). ProPepper: a curated database for identification and analysis of peptide and immune-responsive epitope composition of cereal grain protein families. *Database, 2015,* bav100.

Juhász, A., Haraszi, R., & Maulis, C. (2015b). ProPepper: a curated database for identification and analysis of peptide and immune-responsive epitope composition of cereal grain protein families. *Database: The Journal of Biological Databases and Curation, 2015,* bav100. doi:10.1093/database/bav100

Kaukinen, K., Collin, P., Huhtala, H., & Mäki, M. (2013). Long-term consumption of oats in adult celiac disease patients. *Nutrients, 5*(11), 4380-4389.

Kitteringham, N. R., Jenkins, R. E., Lane, C. S., Elliott, V. L., & Park, B. K. (2009). Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. *Journal of Chromatography B, 877*(13), 1229-1239.

Koeberl, M., Clarke, D., & Lopata, A. L. (2014). Next generation of food allergen quantification using mass spectrometric systems. *Journal of proteome research, 13*(8), 3499-3509.

Koehler, P., & Wieser, H. (2013). Chemistry of cereal grains. In M. Gobbetti & M. Gänzle (Eds.), *Handbook on Sourdough Biotechnology* (pp. 11-45). Boston, MA: Springer US.

Koerner, T. B., Cleroux, C., Poirier, C., Cantin, I., Alimkulov, A., & Elamparo, H. (2011). Gluten contamination in the Canadian commercial oat supply. *Food Additives and Contaminants, 28*(6), 705-710.

Kucek, L. K., Veenstra, L. D., Amnuaycheewa, P., & Sorrells, M. E. (2015). A grounded guide to gluten: how modern genotypes and processing impact wheat sensitivity. *Comprehensive Reviews in Food Science and Food Safety, 14*(3), 285-302.

Larre, C., Lupi, R., Gombaud, G., Brossard, C., Branlard, G., Moneret-Vautrin, D. A., . . . Denery-Papini, S. (2011). Assessment of allergenicity of diploid and hexaploid wheat genotypes: identification of allergens in the albumin/globulin fraction. *Journal of proteomics, 74*(8), 1279-1289.

Lehotay, S. J., Kok, A. d., Hiemstra, M., & Bodegraven, P. v. (2005). Validation of a fast and easy method for the determination of residues from 229 pesticides in fruits and vegetables using gas and liquid chromatography and mass spectrometric detection. *Journal of AOAC International, 88*(2), 595-614.

Lupi, R., Denery-Papini, S., Rogniaux, H., Lafiandra, D., Rizzi, C., De Carli, M., . . . Larré, C. (2013). How much does transgenesis affect wheat allergenicity?: Assessment in two GM lines over-expressing endogenous genes. *Journal of proteomics, 80*, 281-291.

Lupi, R., Masci, S., Rogniaux, H., Tranquet, O., Brossard, C., Lafiandra, D., . . . Larré, C. (2014). Assessment of the allergenicity of soluble fractions from GM and commercial genotypes of wheats. *Journal of Cereal Science, 60*(1), 179-186.

Manfredi, A., Mattarozzi, M., Giannetto, M., & Careri, M. (2015). Multiplex liquid chromatography-tandem mass spectrometry for the detection of wheat, oat, barley and rye prolamins towards the assessment of gluten-free product safety. *Analytica chimica acta, 895*, 62-70.

INSTRUÇÃO NORMATIVA Nº 38, DE 30 DE NOVEMBRO DE 2010, (2010).

Martinez-Esteso, M. J., Nørgaard, J., Brohée, M., Haraszi, R., Maquet, A., & O'Connor, G. (2016). Defining the wheat gluten peptide fingerprint via a discovery and targeted proteomics approach. *Journal of Proteomics*.

Matsuo, H., Kohno, K., Niihara, H., & Morita, E. (2005). Specific IgE determination to epitope peptides of ω-5 gliadin and high molecular weight glutenin subunit is a useful tool for diagnosis of wheat-dependent exercise-induced anaphylaxis. *The journal of Immunology, 175*(12), 8116-8122.

Michaelevski, I., Kirshenbaum, N., & Sharon, M. (2010). T-wave ion mobility-mass spectrometry: basic experimental procedures for protein complex analysis. *Journal of visualized experiments*(41), e1985-e1985.

Osborne, T. B. (1907). *The proteins of the wheat kernel*: Carnegie Institution of Washington.

Owusu-Apenten, R. (2002). *Food protein analysis: quantitative effects on processing* (Vol. 118): CRC Press.

Pastorello, E. A., Farioli, L., Conti, A., Pravettoni, V., Bonomi, S., Iametti, S., . . . Ballmer-Weber, B. (2007). Wheat IgE-mediated food allergy in European patients: α-amylase inhibitors, lipid transfer proteins and low-molecular-weight glutenins. *International archives of allergy and immunology, 144*(1), 10-22.

Pietzak, M. M., & Fasano, A. (2005). Celiac disease: a new paradigm of an immune-mediated disorder due to dietary gluten *Reviews in food and nutrition toxicity* (Vol. 3, pp. 243): Citeseer.

Presutti, R. J., Cangemi, J. R., Cassidy, H. D., & Hill, D. A. (2007). Celiac disease. *Small, 100*(2), 5-17.

Real, A., Comino, I., de Lorenzo, L., Merchán, F., Gil-Humanes, J., Giménez, M. J., . . . Sousa, C. (2012). Molecular and immunological characterization of gluten proteins isolated from oat cultivars that differ in toxicity for celiac disease. *PLoS One, 7*(12), e48365.

Rogniaux, H., Pavlovic, M., Lupi, R., Lollier, V., Joint, M., Mameri, H., . . . Larré, C. (2015). Allergen relative abundance in several wheat varieties as revealed via a targeted quantitative approach using MS. *Proteomics, 15*(10), 1736-1745.

Sapone, A., Bai, J. C., Ciacci, C., Dolinsek, J., Green, P. H. R., Hadjivassiliou, M., . . . Schumann, M. (2012). Spectrum of gluten-related disorders: consensus on new nomenclature and classification. *BMC medicine, 10*(1), 13.

Schalk, K., Lang, C., Wieser, H., Koehler, P., & Scherf, K. A. (2017). Quantitation of the immunodominant 33-mer peptide from α-gliadin in wheat flours by liquid chromatography tandem mass spectrometry. *Scientific reports, 7*, 45092.

Scherf, K. A., Koehler, P., & Wieser, H. (2016). Gluten and wheat sensitivities – an overview. *Journal of Cereal Science, 67*, 2-11.

Sealey-Voyksner, J. A., Khosla, C., Voyksner, R. D., & Jorgenson, J. W. (2010). Novel aspects of quantitation of immunogenic wheat gluten peptides by liquid chromatography–mass spectrometry/mass spectrometry. *Journal of Chromatography A, 1217*(25), 4167-4183.

Shan, L., Qiao, S.-W., Arentz-Hansen, H., Molberg, Ø., Gray, G. M., Sollid, L. M., & Khosla, C. (2005). Identification and analysis of multivalent proteolytically resistant peptides from gluten: implications for celiac sprue. *Journal of proteome research, 4*(5), 1732-1741.

Sharma, G. M., Pereira, M., & Williams, K. M. (2015). Gluten detection in foods available in the United States – a market survey. *Food chemistry, 169*, 120-126.

Shewry, P. R., & Halford, N. G. (2002). Cereal seed storage proteins: structures, properties and role in grain utilization. *Journal of experimental botany, 53*(370), 947-958.

Shewry, P. R., Napier, J. A., & Tatham, A. S. (1995). Seed storage proteins: structures and biosynthesis. *The plant cell, 7*(7), 945.

Shewry, P. R., & Tatham, A. S. (2016). Improving wheat to remove celiac epitopes but retain functionality. *Journal of Cereal Science, 67*, 12-21.

Shewry, P. R., Tatham, A. S., Forde, J., Kreis, M., & Miflin, B. J. (1986). The classification and nomenclature of wheat gluten proteins: a reassessment. *Journal of Cereal Science, 4*(2), 97-106.

Silva, J. C., Gorenstein, M. V., Li, G.-Z., Vissers, J. P. C., & Geromanos, S. J. (2006). Absolute quantification of proteins by LCMSE a virtue of parallel MS acquisition. *Molecular & Cellular Proteomics, 5*(1), 144-156.

Sollid, L. M., Qiao, S.-W., Anderson, R. P., Gianfrani, C., & Koning, F. (2012). Nomenclature and listing of celiac disease relevant gluten T-cell epitopes restricted by HLA-DQ molecules. *Immunogenetics, 64*(6), 455-460.

Song, Y., & Zheng, Q. (2008). Influence of gliadin removal on strain hardening of hydrated wheat gluten during equibiaxial extensional deformation. *Journal of Cereal Science, 48*(1), 58-67.

Šotkovský, P., Hubálek, M., Hernychová, L., Novák, P., Havranová, M., Šetinová, I., . . . Tučková, L. (2008). Proteomic analysis of wheat proteins recognized by IgE antibodies of allergic patients. *Proteomics, 8*(8), 1677-1691.

Spaenij–Dekking, L., Kooy–Winkelaar, Y., van Veelen, P., Drijfhout, J. W., Jonker, H., van Soest, L., . . . Koning, F. (2005). Natural variation in toxicity of wheat: potential for selection of nontoxic varieties for celiac disease patients. *Gastroenterology, 129*(3), 797-806.

Tanner, G. J., Blundell, M. J., Colgrave, M. L., & Howitt, C. A. (2013a). Quantification of hordeins by ELISA: the correct standard makes a magnitude of difference. *PLoS One, 8*(2), e56456.

Tanner, G. J., Colgrave, M. L., Blundell, M. J., Goswami, H. P., & Howitt, C. A. (2013b). Measuring hordein (gluten) in beer – a comparison of ELISA and mass spectrometry. *PLoS One, 8*(2), e56452.

Tatham, A. S., & Shewry, P. R. (2008). Allergens to wheat and related cereals. *Clinical & Experimental Allergy, 38*(11), 1712-1726.

Uvackova, L., Skultety, L., Bekesova, S., McClain, S., & Hajduch, M. (2013a). The MSE-proteomic analysis of gliadins and glutenins in wheat grain identifies and quantifies proteins associated with celiac disease and baker's asthma. *Journal of proteomics, 93*, 65-73.

Uvackova, L., Skultety, L., Bekesova, S., McClain, S., & Hajduch, M. (2013b). MSE based multiplex protein analysis quantified important allergenic proteins and detected relevant peptides carrying known epitopes in wheat grain extracts. *Journal of proteome research, 12*(11), 4862-4869.

van den Broeck, H. C., Cordewener, J. H. G., Nessen, M. A., America, A. H. P., & van der Meer, I. M. (2015). Label free targeted detection and quantification of celiac disease immunogenic epitopes by mass spectrometry. *Journal of Chromatography A, 1391*, 60-71.

Vassiliou, A. (2009). Commission regulation (EC) no. 41/2009 of 20 January 2009 concerning the composition and labelling of foodstuffs suitable for people intolerant to gluten. *Off J Eur Union, 52*, 3-5.

Vensel, W. H., Dupont, F. M., Sloane, S., & Altenbach, S. B. (2011). Effect of cleavage enzyme, search algorithm and decoy database on mass spectrometric identification of wheat gluten proteins. *Phytochemistry, 72*(10), 1154-1161.

Veraverbeke, W. S., & Delcour, J. A. (2002). Wheat protein composition and properties of wheat glutenin in relation to breadmaking functionality. *Critical Reviews in Food Science and Nutrition, 42*(3), 179-208.

Victorio, V. C. M., Souza, G. H. M. F., Santos, M. C. B., Vega, A. R., Cameron, L. C., & Ferreira, M. S. L. (2018). Differential expression of albumins and globulins of wheat flours of different technological qualities revealed by nanoUPLC-UDMSE. *Food chemistry, 239*, 1027-1036.

Wieser, H. (2007). Chemistry of gluten proteins. *Food microbiology*, *24*(2), 115-119.

Wieser, H., Antes, S., & Seilmeier, W. (1998). Quantitative determination of gluten protein types in wheat flour by reversed-phase high-performance liquid chromatography. *Cereal Chemistry, 75*(5), 644-650.

Wieser, H., & Koehler, P. (2009). Is the calculation of the gluten content by multiplying the prolamin content by a factor of 2 valid? *European Food Research and Technology, 229*(1), 9-13.

Wrigley, C. W. (1996). Biopolymers - giant proteins with flour power. *Nature, 381*(6585), 738-739.

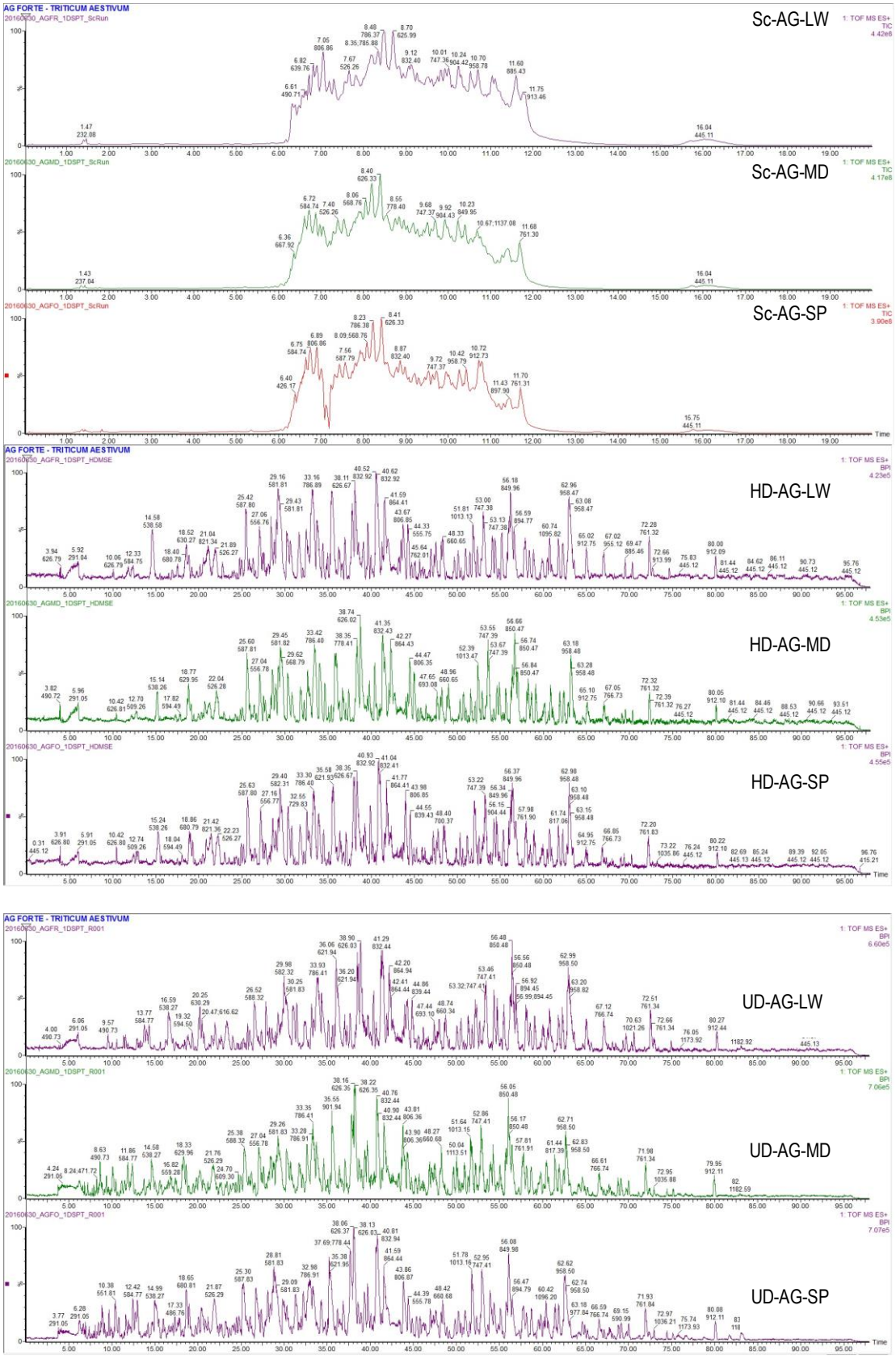
APÊNDICES

Supplementary Table 1. List of the recognized wheat allergens evaluated in this work.

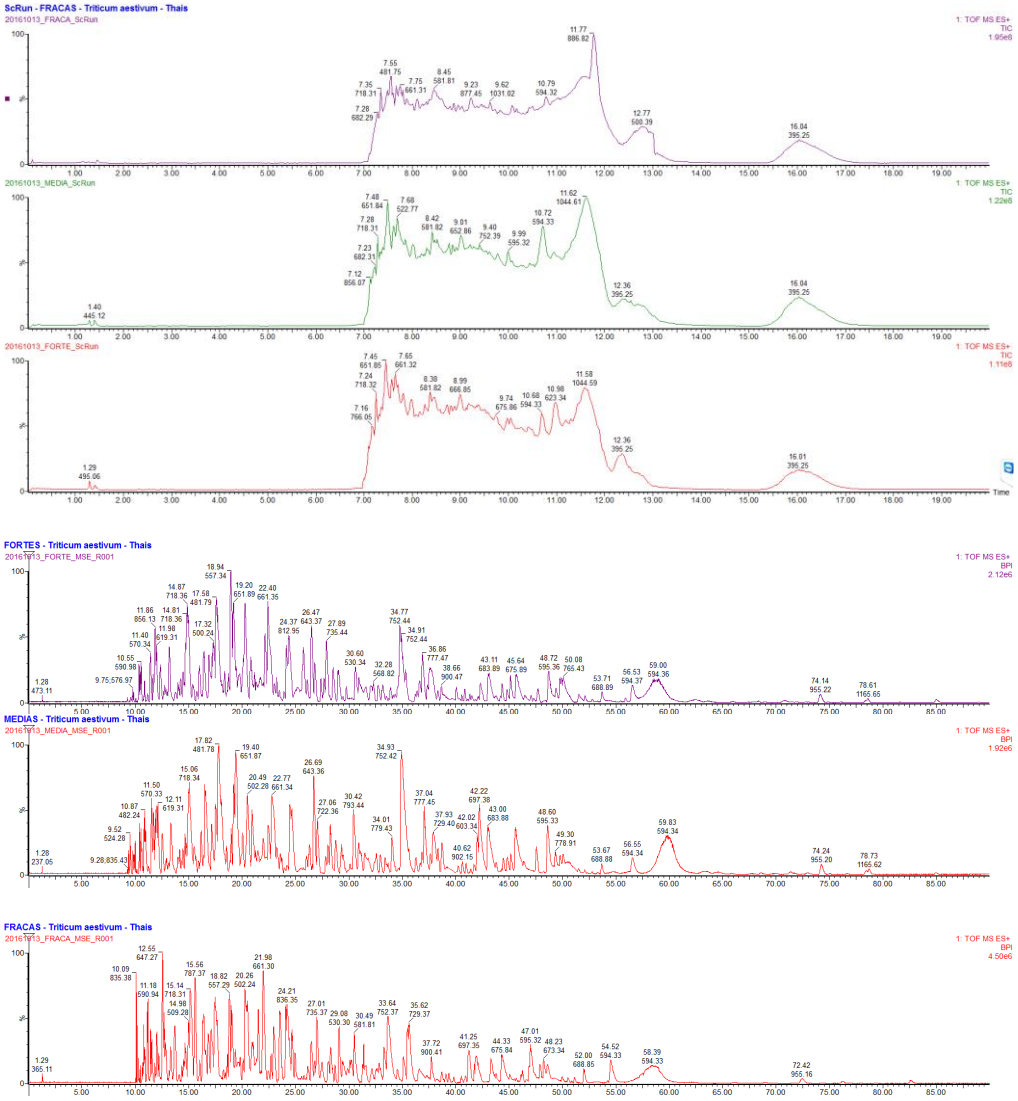
Uniprot Acession	Uniprot best homologue protein name
TC383884	Enolase; Oryza sativa [1]
TC425761	Enolase; Oryza sativa [1]
TC374294	Serpin; Triticum aestivum [1]
TC380640	Serpin; Triticum aestivum [1]
B7U6L4	Globulin 3 — Triticum aestivum [1]
B7U6L3	Globulin 3C (Fragment) — Triticum aestivum [1]
Q8LK23	Peroxidase — Triticum aestivum [1]
B7U6L5	Globulin 3B (Fragment) — Triticum aestivum [1]
TC424895	Beta-glucosidase — Hordeum vulgare [1]
C0LF30	Serpin 1 — Triticum aestivum [1]
Q9ST57	Serpin-Z2A — Triticum aestivum [1]
C0LF31	Serpin 2 — Triticum aestivum [1]
Q401N7	Aspartic proteinase — Triticum aestivum [1]
Q6S5B1	Alpha amylase inhibitor CM3 — Triticum turgidum [1]
C3VWA4	Dimeric alpha amylase inhibitor — Triticum turgidum [1]
B5B0D5	Major allergen CM16 — Triticum aestivum [1]
C3VW80	Dimeric alpha amylase inhibitor — Triticum turgidum [1]
A4ZIU3	Monomeric alpha amylase inhibitor, Triticum monococcum [1]
TC435176	Alpha amylase inhibitor 0.28 precursor; Triticum aestivum [1]
C4P5B9	Monomeric alpha amylase inhibitor — Triticum turgidum [1]
TC431321	Dimeric alpha amylase inhibitor; Triticum aestivum [1]
A4ZIW9	Monomeric alpha-amylase inhibitor; Triticum aestivum [2]
TC402211	WHEAT Alpha-amylase/trypsin inhibitor CM3; Triticum aestivum [2]
A4GFN8	Dimeric α-amylase inhibitor;Triticum turgidum subsp. dicoccoides [2]
TC377918	Cupin family protein. expressed [Oryza sativa Japonica Group] [2]
CA613733	Triosephosphate isomerase; Triticum aestivum [2]
TC450362	Cupin family protein. expressed; Oryza sativa Japonica [2]
TC388221	Beta amylase; Triticum aestivum [2]
TC425413	Globulin-2 precursor; Zea mays [2]
A7UME2	Xylanase inhibitor 725ACCN; Triticum aestivum [2]
P93693	Serpin-Z1B; Triticum aestivum [2]
D2T2K2	Non-specific lipid transfer protein 1 [3]
P01083	Monomeric alpha-amylase inhibitor [3]
Q9LDX4	Thioredoxin [3]
Q4W0V7	Dimeric alpha-amylase inhibitor [3]
C7C4X0	Tetrameric alpha-amylase inhibitor CM1/CM2 [3]
D2TGC2	Tetrameric alpha-amylase inhibitor CM1/CM2 [3]
P17314	Tetrameric alpha-amylase inhibitor CM3 [3]
Q9FS79	Triosephosphate-isomerase [3]

Q6W8Q2	1-cys-peroxiredoxin [3]
Q41593	Serpins [3]
P93692	Serpins [3]
C7C4X1	Glyceraldehyde-3-phosphate-dehydrogenase [3]
TC373663	Aspartate aminotransferase (Fragment); Triticum aestivum [4]
TC382510	Phytepsin precursor (Aspartic proteinase); Hordeum vulgare [4]
TC377861	Beta-amylase; *Hordeum vulgare* [4]
P93606	Superoxide dismutase; *Triticum aestivum* [4]
TC403071	Cupin family protein. expressed; *Oryza sativa* Japonica Group (Rice) [4]
TC428590	Triosephosphate isomerase; *Triticum aestivum* (Wheat) [4]
TC369624	Dehydroascorbate reductase; *Triticum aestivum* [4]
TC392505	Cupin family protein. expressed; *Oryza sativa Japonica* (Rice) [4]
T389920	Triosephosphate isomerase; *Triticum aestivum* (Wheat) [4]
TC389327	Embryo globulin; Triticeae [4]
TC399106	Embryo globulin; *Triticeae* (Barley) [4]
TC421662	Hydroxyproline-rich glycoprotein DZ-HRGP precursor; Volvox carteri f. Nagariensis [4]
CD914053	Seed maturation protein [*Oryza sativa* Japonica Group] [4]
TC368606	Protein disulfide isomerase 2 precursor; Triticum [4]
TC394284	Dehydroascorbate reductase; *Triticum aestivum* [4]
TC417836	Superoxide dismutase; *Triticum aestivum* [4]
	Embryo globulin; Triticeae [4]
TC368657	Xylanase inhibitor; *Triticum aestivum* [4]
TC383936	Embryo globulin; *Triticeae Hordeum vulgare* (Barley) [4]
TC404101	Endo-1.4-beta-glucanase*Malus x domestica* [4]
TC369421	Xylanase inhibitor XIP-III; *Triticum aestivum* (Wheat) [4]
TC408407	Glucose and ribitol dehydrogenase homolog Barley [4]
TC374136	Glutathione transferase; *Triticum aestivum* [4]
CK215488	Lactoylglutathione lyase; Oryza [4]
CJ541649	Embryo globulin; Triticeae [4]
TC374296	Lactoylglutathione lyase; *Oryza sativa* (Rice) [4]
Q9ST58	Serpin-Z1C; *Triticum aestivum* [4]
TC380730	Beta-amylase precursor; *Hordeum vulgare* subsp. spontaneum (Barley) [4]
Q9ZR33	Glycosyltransferase 75; *Triticum aestivum* [4]

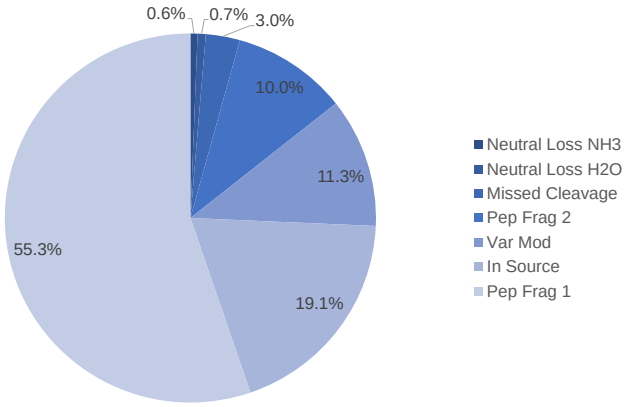
[1] Larré. 2011. [2] Lupi. 2014. [3] Rogniaux. 2015. [4] Lupi. 2013



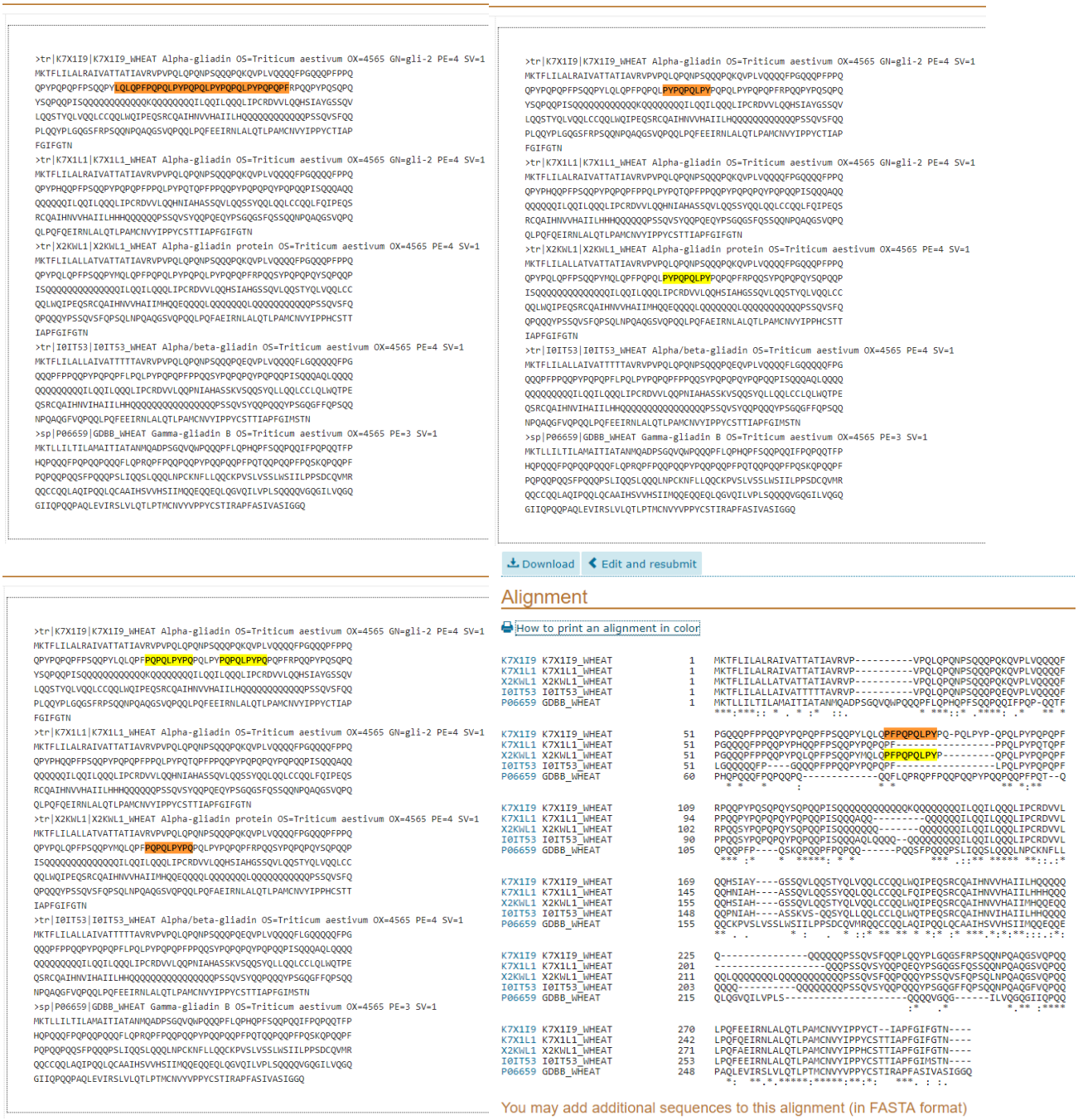
Supplementary Figure 1. Scouting runs (Sc). HDMS^{E} and UDMS^{E} chromatograms of AG samples.



Supplementary Figure 2. Scouting runs (Sc) and MS^E chromatograms of gluten samples.



Supplementary Figure 3. Gluten peptides match type.



Supplementary Figure 4. Alignment of α-gliadins identified in the work and search for 33-mer epitopes.