



PROGRAMA DE PÓS-GRADUAÇÃO EM ALIMENTOS E NUTRIÇÃO
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO

TESE DE DOUTORADO

VERÔNICA CRISTINA MAYRINCK VICTORIO

**CARACTERIZAÇÃO DO PERFIL PROTEICO DE TRIGO COMUM EM FUNÇÃO DA
APTIDÃO TECNOLÓGICA DO GRÃO: uma abordagem proteômica**

Rio de Janeiro
2021



Verônica Cristina Mayrinck Victorio

**CARACTERIZAÇÃO DO PERFIL PROTEICO DE TRIGO COMUM EM FUNÇÃO DA
APTIDÃO TECNOLÓGICA DO GRÃO: uma abordagem proteômica**

***CHARACTERIZATION OF COMMON WHEAT PROTEIN PROFILE AS A FUNCTION
OF GRAIN TECHNOLOGICAL APTITUDE: a proteomic approach***

Tese de doutorado 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 Doutora em Alimentos e Nutrição.

Orientadora: Prof^a. Dr^a. Mariana Simões Larraz
Ferreira

Rio de Janeiro

2021

Catálogo informatizada pelo(a) autor(a)

V642 Victorio, Verônica Cristina Mayrinck
Caracterização do perfil proteico de trigo comum em função da aptidão tecnológica do grão: uma abordagem proteômica / Verônica Cristina Mayrinck Victorio. -- Rio de Janeiro, 2022.
219 p.

Orientadora: Mariana Simões Larraz Ferreira.
Tese (Doutorado) - Universidade Federal do Estado do Rio de Janeiro, Programa de Pós-Graduação em Alimentos e Nutrição, 2022.

1. Análise proteômica. 2. Aptidão tecnológica de farinhas de trigo. 3. Biomarcadores proteicos. 4. Espectrometria de massas. 5. Química de cereais. I. Ferreira, Mariana Simões Larraz, orient. II. Título.

Verônica Cristina Mayrinck Victorio

CARACTERIZAÇÃO DO PERFIL PROTEICO DE TRIGO COMUM EM FUNÇÃO DA
APTIDÃO TECNOLÓGICA DO GRÃO: UMA ABORDAGEM PROTEÔMICA

Tese de doutorado apresentada ao Programa de
Pós-Graduação em Alimentos e Nutrição da
Universidade Federal do Estado do Rio de Janeiro.

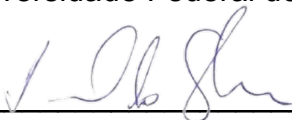
Aprovada em: 19 / 04 / 2022

BANCA EXAMINADORA



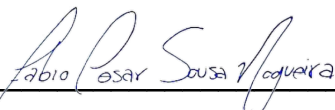
Dra. Mariana Simões Larraz Ferreira

Universidade Federal do Estado do Rio de Janeiro – UNIRIO




Dr. Vanildo Silveira

Universidade Estadual do Norte Fluminense Darcy Ribeiro – UENF



Dr. Fabio César Sousa Nogueira

Universidade Federal do Rio de Janeiro – UFRJ




Dra. Cristina Yoshie Takeiti

Embrapa Agroindústria de Alimentos



Dr. Luiz Carlos Gutkoski

Universidade Federal do Estado do Rio de Janeiro – UNIRIO



Dedico este trabalho a Deus, Pai amantíssimo, que me dirigiu e sustentou até aqui. Gratidão por me mostrar que, contigo, eu posso realizar o que eu acreditava ser impossível. “TUDO POSSO NAQUELE QUE ME FORTALECE” (Filipenses 4:13)

AGRADECIMENTOS

Agradeço à FAPERJ e ao CNPq por financiarem essa pesquisa.

Ao prof. Dr. Luiz Carlos Gutkoski, à empresa OR Melhoria de Sementes Ltda. e à Dr^a. Valérie Lullien-Pellerin pelo fornecimento das amostras.

À Dra. Mariana Simões Larraz Ferreira, minha querida orientadora, por todo apoio, incentivo e carinho de sempre. Gratidão pelo seu esforço tão grandioso para me ajudar.

À Dra. Édira Castello Branco de Andrade Gonçalves pelo incentivo para ingressar no doutorado e por toda ajuda nesta trajetória.

À Dra. Cristina Yoshie Takeiti e aos doutores Fábio César Souza Nogueira, Luiz Carlos Gutkoski e Vanildo Silveira, membros da banca examinadora, e às doutoras Andréa Furtado Macedo e Talita Pimenta do Nascimento, membros suplentes, por terem aceitado o convite e pelas contribuições que permitiram melhorar este trabalho.

Ao Dr. Gustavo Henrique Martins Ferreira Souza pelo direcionamento e por todo auxílio na execução deste trabalho.

Ao Dr. Vanildo Silveira e ao Dr. Felipe Astolpho de Almeida pela eficiência e rapidez com que executaram as análises e me auxiliaram no trabalho.

À toda equipe do Laboratório de Bioativos I. Vocês são o extrato do que há de melhor em termos de cientistas! Agora é só tripsinizar e analisar por LC-MS e teremos a caracterização do melhor time no qual eu poderia trabalhar!

À equipe do Laboratório de Bioquímica de Proteínas.

Aos queridos técnicos Henrique Abreu e Matheus Grilo pela presteza em ajudar e pelas palavras de encorajamento.

À minha querida amiga Mônica Cristine, por me ouvir e incentivar sempre! Você é um grande exemplo para mim.

À Cinthia Couto, amiga e recém Doutora, que trilhou o caminho comigo desde o mestrado. Parabéns minha amiga! Que felicidade! Um brinde a nós!

À Cláudia Farina, Nathália Pereira, Valdemir Matos Souza, Rachel Marques e Renata do Nascimento, amigos de trabalho da Equipe de Nutrição do IEISS, por me apoiarem e torcerem por mim sempre. Vocês foram fundamentais para eu chegar até aqui.

À Isabel Amaral, Juliana Libório, Juliana Lyra, Karine Correa e Quésia Vieira, assistentes sociais e Beatriz Bezerra, fonoaudióloga do IEISS, por serem fechamento sempre.

A todos os meus amigos por me amarem e orarem por mim.

Aos meus amados pais Adilson Francisco Victorio e Selma Mayrinck Victorio por me darem a vida e fazer dela uma existência feliz a cada dia. É por vocês e para vocês que me esforço tanto. Essa vitória é nossa! Agora sim sou sua Doutorinha, mãezinha!

Aos meus irmãos Adilson Victorio Júnior, Anselmo Mayrinck e Vera Lucia Mayrinck por torcerem por mim e se alegrarem com minhas vitórias. Amo muito vocês!

À minha querida Fabíola Felipe por ser essa *sister in law* e *belle soeur* tão amada!

A todos os meus sobrinhos e sobrinhos-netos por alegrarem minha vida.

À minha querida Bárbara Souza por ser minha amiga e continuar sendo minha fã todos esses anos. E, olha, eu também me orgulho muito de você!

Às minhas irmãs de vida Ana Paula Ramos e Marion Santana. Gratidão pela força e incentivo de sempre. Obrigada Mari, por emprestar sua casa para eu estudar e ainda fazer o almoço!

À minha prima Sandra Mendenque, por cuidar da Heleninha enquanto eu realizava este trabalho.

À Marisete Pio da Fonseca, minha sogrinha, por todo auxílio neste período. Gratidão por cuidar tão bem do nosso bem mais precioso.

Ao meu sogro Luiz Carlos da Fonseca e à Rosângela Sodré por entenderem as noites de luzes acesas e por cuidarem da Heleninha para eu trabalhar/estudar.

Ao Warllan Pio da Fonseca, meu marido lindo, por ter caminhado essa jornada comigo. Obrigada por entender as madrugadas em claro e o entra e sai do quarto para fazer a Helena dormir e voltar ao trabalho. Desejo que todo esse esforço seja muito bem recompensado.

À Helena Mayrinck Victorio Pio da Fonseca que me deu o mais grandioso título que tenho, o de MÃE. Você chegou no meio desta jornada e não a facilitou em nada, mas a fez muito mais feliz do que antes da sua chegada. Eu amo você infinita e eternamente!

E a todos que, de algum modo, contribuíram para o sucesso deste trabalho.

Deus abençoe cada um de vocês.

*“A gente demora a se tornar pesquisadora,
mas um dia a chave vira.”*

(Mariana Simões Larraz Ferreira)

VICTORIO, V.C.M.V. Caracterização do perfil proteico de trigo comum em função da aptidão tecnológica do grão: uma abordagem proteômica.

RESUMO

O trigo (*Triticum aestivum*) é um alimento básico de grande importância nutricional e econômica, sendo amplamente consumido e empregado em diversos produtos alimentícios. A aptidão do trigo para produção de bolos, biscoitos e massas está atrelada principalmente à fração proteica deste cereal. As proteínas de reserva conferem propriedades viscoelásticas, enquanto as proteínas metabólicas incluem enzimas que garantem o enovelamento adequado das proteínas de reserva. Sendo assim, o teor e o perfil proteico impactam na aptidão tecnológica do grão que, por sua vez, define a orientação final de uso do trigo. A espectrometria de massas (MS) oferece uma abordagem original para identificar proteínas diferenciais associadas à aptidão tecnológica. O objetivo deste trabalho foi caracterizar os diferentes perfis proteicos de farinhas de trigo com aptidões tecnológicas contrastantes, aplicando proteômica comparativa não dirigida com base na cromatografia líquida de ultra performance acoplada à espectrometria de massas de alta definição com abordagem independente de dados. A identificação das proteínas diferencialmente abundantes revelou subunidades de glutenina e de pequenas proteínas de choque térmico como possíveis marcadores das farinhas de alta força do glúten, bem como puroindolinas e proteína de maciez do grão como possíveis marcadores das farinhas de baixa força do glúten. Por predição de domínios e avaliação da anotação funcional, foram reveladas diferentes funções proteicas e maior diversidade de processos biológicos nas amostras de maior força do glúten. A análise quimiométrica foi consistente com os demais resultados distinguindo e agrupando essas amostras com as de fenótipo *hard*. Conclui-se que os perfis proteicos diferentes permitiram a distinção entre as amostras de alta força do glúten daquelas de baixa força do glúten. Com resultados abrangentes, que podem ser aplicados em programas de melhoramento do trigo, esta tese permitiu avançar na compreensão do perfil proteico atrelado à aptidão tecnológica e ainda apresenta informações inéditas sobre o perfil proteico do trigo brasileiro.

Palavras-chave: ontologia genética, predição de domínios, proteoma do trigo, quimiometria.

VICTORIO, V.C.M.V. Characterization of common wheat protein profile as a function of grain technological aptitude: a proteomic approach.

ABSTRACT

Wheat (*Triticum aestivum*) is a staple food of great nutritional and economic importance, being widely consumed and used in various food products. The aptitude of wheat to produce cakes, cookies and pasta is mainly linked to the protein fraction of this cereal. Storage proteins confer viscoelastic properties, while metabolic proteins include enzymes that ensure the correct folding of storage proteins. Thus, the protein content and profile impact the technological aptitude of the grain, which defines the final orientation of wheat use. Mass spectrometry (MS) offers an original approach to identify differential proteins associated with technological aptitude. The objective of this work was to characterize the different protein profiles of wheat flours with contrasting technological aptitudes, applying non-targeted comparative proteomics based on ultra-performance liquid chromatography coupled to high-definition mass spectrometry with data-independent approach. The identification of differentially abundant proteins revealed glutenin subunits and small heat shock proteins as possible markers of high gluten-force flours, as well as puroindolines and grain softness protein as possible markers of low gluten-force flours. By domain prediction and functional annotation evaluation, different protein functions and greater diversity of biological processes were revealed in samples of higher gluten-force. The chemometric analysis was consistent with the other results, distinguishing and grouping these samples with the hard phenotype ones. The different protein profiles allowed the distinction between the samples of high gluten-force from those of low gluten-force. With comprehensive results, which can be applied in wheat breeding programs, this thesis allowed the better understanding of the protein profile linked to technological aptitude and also presents new information about the protein profile of Brazilian wheat.

Keywords: domain prediction, chemometrics, genetic ontology, wheat proteome.

LISTA DE ILUSTRAÇÕES

CAPÍTULO 1:

Figura 1. Estrutura da tese..... 28

Figura 2. O grão de trigo e suas diferentes camadas. Adaptado de Surget e Barron (2005). 35

Figura 3. Composição das proteínas do trigo: aproximação entre as classificações de Osborne (1907) e Shewry (1986). 37

Figura 4. Estruturas das α - e γ -gliadinas com ligações dissulfeto intramoleculares apresentadas por Wieser (2007). 39

Figura 5. Esquema proposto para as estruturas típicas de LMW-GS por D’ovidio & Masci (2004). As cisteínas são designadas por números ou de acordo com o sistema de letras proposto por Kohler *et al.*(1993). SH ou asterisco representam as cisteínas que formam pontes dissulfeto intermolecular. S, peptídeo sinal, N, N-terminal, REP, domínio repetitivo e C-TER, domínios C-terminal. 40

CAPÍTULO 2:

Fig. 1. Dynamic range of wheat flour samples, low (LW), medium (MD) and superior (SP) quality classification. The rank of identified proteins is plotted in the x axis and y axis is represented by the log of abundance quantified proteins. 64

Fig. 2. Principal component analysis (PCA) of wheat flour samples: low (LW), medium (MD) and superior (SP) quality classification. Identified and annotated proteins are in gray and red. 65

Fig. 3. Volcano plots comparing the wheat flour samples in pairs, low (LW), medium (MD) and superior (SP) ($p < 0.05$). The green dots represent up-regulated proteins (max fold change > 1.5), while red dots represent down-regulated proteins (max fold change < -1.5), black empty dots the considered unchanged proteins and gray empty dots were out of filters (minimum one unique peptide, two peptide count, CV of 0.3, max fold change < 1.5 , $P < 0.05$ and/or proteins in 3/3 replicates). 66

Fig. 4. Heat map illustrating the clustering of functional domains from uncharacterized differentially abundant proteins that were found in low (LW), medium

(MD) and superior (SP) samples. The horizontal axis presents replicates from each sample and the vertical axis presents the domains. Color gradient represents the variation of protein abundance. Highest quantities are represented by dark red and lowest quantities by dark blue. PUF (protein of unknown function), GCK domain is related with intracellular signaling pathways. 70

Fig. 5. Sodium dodecyl sulfate polyacrylamide gel electrophoresis patterns of high molecular weight glutenin subunits (HMW-GS) and low molecular weight glutenin subunits (LMW-GS) from wheat genotypes. The numbers on the top of the figure represent the genotypes. The patterns are Bermude (1 and 12) and Toback (2 and 13). The analyzed genotypes are Campeiro (3), ORS 27 (5) and ORS 25 (7) of low quality, Marfim (4), ORS 1402 (6) and ORS 1401 (10) of medium quality and Jadeíte 11 (8), Ametista (9) and Guabiju (11) of superior quality. HMW-GS are labeled according to the nomenclature of Payne & Lawrence (1983). Unknown bands are marked by an interrogation point. 72

CAPÍTULO 3:

Graphical abstract 1. Summary of NanoUPLC-MS^E analysis of wheat flour. 54

Fig. 1. Dynamic range of wheat flour samples, low (LW), medium (MD) and superior (SP) quality classification. The rank of identified proteins is plotted in the x axis and y axis is represented by the log of abundance quantified proteins. 64

Fig. 2. Principal component analysis (PCA) of wheat flour samples: low (LW), medium (MD) and superior (SP) quality classification. Identified and annotated proteins are in gray and red. 65

Fig. 3. Volcano plots comparing the wheat flour samples in pairs, low (LW), medium (MD) and superior (SP) ($p < 0.05$). The green dots represent up-regulated proteins (max fold change > 1.5), while red dots represent down-regulated proteins (max fold change < -1.5), black empty dots the considered unchanged proteins and gray empty dots were out of filters (minimum one unique peptide, two peptide count, CV of 0.3, max fold change < 1.5 , $P < 0.05$ and/or proteins in 3/3 replicates). 66

Fig. 4. Heat map illustrating the clustering of functional domains from uncharacterized differentially abundant proteins that were found in low (LW), medium

(MD) and superior (SP) samples. The horizontal axis presents replicates from each sample and the vertical axis presents the domains. Color gradient represents the variation of protein abundance. Highest quantities are represented by dark red and lowest quantities by dark blue. PUF (protein of unknown function), GCK domain is related with intracellular signaling pathways. 70

Fig. 5. Sodium dodecyl sulfate polyacrylamide gel electrophoresis patterns of high molecular weight glutenin subunits (HMW-GS) and low molecular weight glutenin subunits (LMW-GS) from wheat genotypes. The numbers on the top of the figure represent the genotypes. The patterns are Bermude (1 and 12) and Toback (2 and 13). The analyzed genotypes are Campeiro (3), ORS 27 (5) and ORS 25 (7) of low quality, Marfim (4), ORS 1402 (6) and ORS 1401 (10) of medium quality and Jadeíte 11 (8), Ametista (9) and Guabiju (11) of superior quality. HMW-GS are labeled according to the nomenclature of Payne & Lawrence (1983). Unknown bands are marked by an interrogation point. 72

CAPÍTULO 4:

Figure 1. Overview of the work and quality of proteomic data. (A) Schematic diagram of wheat flour sampling and untargeted proteomic workflow. (B) Number of identified and quantified proteins that appear in 3/3 replicates in each sample. (C) Dynamic range presenting more than 6 orders of magnitude ($\text{Log}_{10} > 6$). The rank of identified proteins is expressed in the x axis, while log of normalized total ion count (TIC) abundance is in the y axis. NILs: near-isogenic lines, RWF: refined wheat flour and WWF: whole wheat flour. **Erro! Indicador não definido.**

Figure 2. Gene ontology (GO) association data for protein fraction of whole wheat flours (WWF, in orange) and refined wheat flours (RWF, in blue). (A and D) Pie charts represent the percentage of proteins associated or not with GO categories. (B and E) Venn diagrams show the number of proteins associated with biological process, molecular function and cellular component. (C and F) Bar charts show the number of proteins associated with the top 20 biological process.

Figure 3. Volcano plots showing the differentially accumulated proteins between samples with contrasting gluten-force (A and B) and samples differing by PINB allele (C). The blue dots represent up-accumulated proteins (UP), while red dots represent down-

accumulated proteins (DOWN). Gray dots represented unchanged proteins and gray empty dots were out of p-value filter (p_{OUT} , $p > 0.05$). WWF: whole wheat flour; RWF: refined wheat flour; NILs: near-isogenic lines.....

Figure 4. Up accumulated proteins (UP) associated to biological processes in each sample. (A) whole wheat flour (WWF) high gluten-force (GF), (B) WWF low GF, (C) refined wheat flour (RWF) high GF, (D) RWF low GF, (E) RWF hard near-isogenic lines (NILs) and (F) RWF soft NILs.....**Erro! Indicador não definido.**

Figure 5. Chemometric data obtained by Principal Component Analysis - PCA of (A) all samples, separation of whole wheat flours - WWF (triangle) and refined wheat flours – RWF (parallelogram); (B) RWF, separation of near isogenic lines - NILs (rectangle) from gluten-force (GF) based genotypes (trapezium) and by Partial Least Squares-Discriminant Analysis – PLS-DA (C) including all samples, highlighting the separation of high GF and hard-type phenotype (rounded rectangle) from those with low GF and soft-type phenotype (ellipse) along component 2, which was also marked in (D) the PLS-DA of RWF, along component 1; (E) PLS-DA of NILs and (F) PLS-DA of GF based samples.

CAPÍTULO 5

Figura. 1. Exemplo de espectro obtido por HDMS^E processados no programa DriftScope..... 87

Figura. 2. Cromatogramas das corridas de prospecção (*scouting run*) dos extratos sequenciais (AG, Gli, GS e GI) das amostras Campeiro e Ametista em TIC (*total ion count*). 88

Figura. 3. Cromatogramas das aquisições por HDMS^E dos extratos sequenciais (AG, Gli, GS e GI) das amostras Campeiro e Ametista em BPI (*base peak intention*).. 90

Figura. 4. Cromatogramas das aquisições por UDMS^E dos extratos sequenciais (AG, Gli, GS e GI) das amostras Campeiro e Ametista em BPI (*base peak intention*).. 91

Figura. 5. Espectros UDMS^E das aquisições em triplicata dos extratos AG1 (superiores) e Gli1 (inferiores), em roxo os peptídeos associados à identificação de proteínas..... 92

Figura. 6. Espectros UDMS^E das aquisições em triplicata dos extratos GS1 (superiores) e GI1 (inferiores), em roxo os peptídeos associados à identificação de proteínas..... 93

Figura. 7. Comparação do número de identificações de peptídeos e proteínas em HDMS^E e UDMS^E nos extratos Albuminas/Globulinas (AG), Gliadinas (Gli), Gluteninas Solúveis (GS) e Gluteninas Insolúveis (GI). 94

Figura. 8. Comparação do número total de identificações de proteínas e peptídeos identificados com intervalo de confiança (IC) acima de 95%, em HDMS^E e UDMS^E. ... 95

Figura. 9. Distribuição de fragmentação dos íons *b* (vermelho) e *y* (azul) obtidos nas corridas das amostras AG da variedade Ametista em HDMS^E (superior) e UDMS^E (inferior). 96

Figura. 10. Distribuição de fragmentação dos íons *b* (vermelho) e *y* (azul) obtidos nas corridas das amostras Gli (superior), GS (meio), GI (inferior) da variedade Ametista em HDMS^E (esquerda) e UDMS^E (direita). 97

Figura. 11. Total de proteínas identificadas em conjunto em 2 de 3 replicatas e em 3 de 3 replicatas. 98

Figura. 12. Proporção (eixo y) do número de peptídeos identificados por proteína (eixo x, 1 a 18). 98

Figura. 13. Diagrama de Venn com todas as proteínas identificadas nos quatro extratos sequenciais para as duas variedades estudadas. 99

CAPÍTULO 6:

Fig. 1. Percentual de clivagem perdida comparando extratos digeridos com tripsina (Trip) e tripsina/Lys-C (Lys-C) em triplicata. 110

Fig. 2. Faixas dinâmicas obtidas com os métodos (A) MS^E, para proteínas de reserva e (B) HDMS^E, para proteínas totais. 111

LISTA DE TABELAS

CAPÍTULO 1:

Tabela 1. Produção e consumo de trigo em 2020/21. 31

Tabela 2. Classes do trigo do Grupo II destinado à moagem e outras finalidades.

..... 32

Tabela 3. Limites de tolerâncias para farinhas de trigo. 34

CAPÍTULO 3:

Table 1. Technological quality characterization of different Brazilian wheat genotypes according to rheological and physico-chemical parameters, harvest 2014/2015..... 59

Table 2. Differentially abundant and annotated proteins..... 68

CAPÍTULO 4:

Table. 1. List of selected differentially abundant proteins (DAPs) and exclusive proteins (EP) in each sample. Up-accumulated (UP) proteins with cellular response to unfolded protein and protein folding as biological process (BP) term and nutrient reservoir activity as molecular function (MF) term are shown. All exclusive proteins with maximum coefficient of variation of 0.3, irrespective of BP and MF, are presented.

CAPÍTULO 5:

Tabela. 1. Caracterização da qualidade tecnológica de duas cultivares de trigo brasileiro, segundo parâmetros reológicos e físico-químicos. 83

Tabela. 2. Lista das proteínas de reserva identificadas nos diferentes extratos. 101

CAPÍTULO 6:

Tabela: 1. Resumo das abordagens proteômicas utilizadas para caracterizar proteínas do trigo..... 113

Tabela: 2. Principais resultados ligados aos objetivos alcançados. 114

LISTA DE SIGLAS

A/G	Albuminas/globulinas (<i>albumins/globulins</i>)
AACCI	American Association of Cereal Chemists International
ALPs	<i>Avenin-like proteins</i>
ANOVA	Análise de variância
ATP	Adenosina trifosfato (<i>adenosine triphosphate</i>)
AVE	Média (<i>Average</i>)
BP	processo biológico (<i>biological process</i>)
C	Cisteína
CC	Componente celular (<i>celular component</i>)
CD	<i>Celiac disease</i>
CRT	calreticulin
CV	Coeficiente de variação (<i>coefficient of variation</i>)
DAP	Proteínas diferenciamente abundantes/acumuladas (<i>differentially abundant proteins/differentially accumulated proteins</i>)
DIA	Aquisição independente de dados (<i>data-independent acquisition</i>)
DOWN	<i>Down-regulated proteins</i> ou <i>down-accumulated proteins</i>
DTT	Ditiotreitol (<i>dithiothreitol</i>)
EP	proteínas exclusivas (<i>exclusive proteins</i>)
FDR	Taxa de descoberta falsa (<i>false discovery rate</i>)
GF	<i>Gluten-force</i>
GFP	[Glu1]-Fibrinopeptide B human ([Glu1]-Fibrinopeptide B human)
GI	Gluteninas insolúveis
Gli	Gliadinas
GO	<i>Gene ontology</i>
GS	Gluteninas solúveis
GSP	<i>Grain softness protein</i>
HDMS	<i>High-definition mass spectrometry</i>
HDMS ^E	<i>High-definition mass spectrometry multiplex method</i>
HMW-GS	Subunidades de glutenina de alta massa molecular (<i>low-molecular weight glutenin subunit</i>)
HSPs	Proteínas de choque térmico (<i>heat shock proteins</i>)
IAM	Iodoacetamida (<i>iodoacetamide</i>)
IMS	Espectrometria por mobilidade iônica (<i>ion mobility spectrometry</i>)
LC-MS	Cromatografia líquida acoplada à espectrometria de massas (<i>liquid chromatography coupled o mass spectrometry</i>)
LMW-GS	Subunidades de glutenina de baixa massa molecular (<i>low-molecular weight glutenin subunit</i>)
LW	<i>Low</i>
M	Metionina
<i>m/z</i>	relação massa/carga
MD	Média
MF	Função molecular (<i>molecular function</i>)
MS	Espectrometria de massas (<i>mass spectrometry</i>)
MS/MS	Espectrometria de massas sequencial (<i>tandem mass spectrometry</i>)

MS ^E	<i>Mass spectrometry multiplex method</i>
Mt	Milhões de toneladas
NA	Não associado (<i>not associated</i>)
NH ₄ HCO ₃	Bicarbonato de amonio
NILs	Linhagens <i>quase-isogênicas</i> (<i>near-isogenic lines</i>)
P	Prolina
PC	Componente principal (<i>principal component</i>)
PCA	Análise de componentes principais (<i>principal component analysis</i>)
PDI	Proteína dissulfeto isomerase (<i>protein disulfide isomerase</i>)
PH	Peso do hectolitro
PINA	Puroindolina-A (<i>puroindoline-A</i>)
PINB	Puroindolina-B (<i>puroindoline-B</i>)
PINs	Puroindolinas (<i>puroindolines</i>)
PLS-DA	Análise discriminante por mínimos quadrados parciais (<i>partial least squares discriminant analysis</i>)
PU	Peptídeos únicos (<i>unique peptides</i>)
Q	Glutamina
RP	Peptídeos relatados (<i>reported peptides</i>)
RWF	<i>Refined wheat flours</i>
S	Serina
SC	cobertura de sequência (<i>sequence coverage</i>)
SDS	Dodecil sulfato de sódio (<i>sodium dodecyl sulfate</i>)
SP	Superior
T	Treonina
TFA	Ácido trifluoracético (<i>trifluoroacetic acid</i>)
TIC	Contagem total de íons (<i>total ion count</i>)
TOF	Tempo de voo (<i>time-of-flight</i>)
UP	Proteínas superexpressas (<i>up-regulated</i> ou <i>up-accumulated proteins</i>)
UPLC	Cromatografia líquida de ultra performance (<i>ultraperformance liquid chromatography</i>)
VIP	Importância variável na projeção (<i>variable importance in projection</i>)
WWF	<i>Whole Wheat Flours</i>
Y	Tirosina

SUMÁRIO

INTRODUÇÃO	22
CAPÍTULO 1 – O TRIGO	30
1 O TRIGO	30
1.1 PRODUÇÃO E CONSUMO.....	30
1.2 PADRÃO DE QUALIDADE DO TRIGO E DA FARINHA DE TRIGO NO BRASIL	32
1.3 ESTRUTURA E COMPOSIÇÃO QUÍMICA DO GRÃO.....	35
1.4 PROTEÍNAS DO TRIGO E SEU IMPACTO NA APTIDÃO TECNOLÓGICA DO GRÃO	36
1.4.1 Proteínas metabólicas.....	37
1.4.2 Proteínas de Reserva	38
1.4.3 Impacto das proteínas metabólicas e de reserva na aptidão tecnológica do grão de trigo	41
CAPÍTULO 2 – ANÁLISE DAS PROTEÍNAS DE TRIGO E DE OUTROS CEREAIS POR ABORDAGEM PROTEÔMICA	44
2 ANÁLISE DAS PROTEÍNAS DE TRIGO E DE OUTROS CEREAIS POR ABORDAGEM PROTEÔMICA	44
2.1 DESAFIOS PARA ANÁLISE PROTEÔMICA DO TRIGO	44
2.2 CROMATOGRAFIA LÍQUIDA DE ULTRA PERFORMANCE ACOPLADA À ESPECTROMETRIA DE MASSAS	46
2.3 ESTRATÉGIAS E MÉTODOS EM ESPECTROMETRIA DE MASSAS..	48
2.4 ABORDAGEM PROTEÔMICA EM TRIGO E OUTROS CEREAIS	49

CAPÍTULO 3 – NanoUPLC-MS^E REVEALS DIFFERENTIAL ABUNDANCE OF GLUTEN PROTEINS IN WHEAT FLOURS OF DIFFERENT TECHNOLOGICAL QUALITIES	54
HIGHLIGHTS	54
GRAPHICAL ABSTRACT	54
ABSTRACT	55
1 INTRODUCTION	56
2 MATERIALS AND METHODS	58
2.1 PLANT MATERIAL	58
2.2 PROTEINS EXTRACTION	59
2.3 SAMPLE PREPARATION FOR NanoUPLC-MS ^E ANALYSIS	60
2.4 NanoUPLC-MS ^E ANALYSIS	60
2.5 DATA PROCESSING	61
2.6 1D SODIUM DOCEDYL SULFATE POLYACRYLAMIDE GEL ELECTROPHORESIS (1D SDS-PAGE)	62
2.7 STATISTICAL ANALYSIS	63
3 RESULTS AND DISCUSSION	63
3.1 NanoUPLC-MS ^E RESULTS	63
3.1.1 Protein profile cluster of nanoUPLC-MS^E data	65
3.2 PROTEIN ABUNDANCE PROFILES OF SAMPLES COMPARED IN PAIRS	66
3.2.1 Cluster of functional domains of differentially abundant and not described proteins	69
3.3 GLUTENIN SUBUNITS CHARACTERIZATION BY ELECTROFORESIS	71
4 CONCLUSIONS	72

ACKNOWLEDGEMENTS	73
CAPÍTULO 4 – UNTARGETED PROTEOMICS UNRAVELS THE PROTEIN PROFILE OF WHEAT FLOURS OF CONTRASTING TRAITS BASED ON GLUTEN-FORCE AND NEAR-ISOGENIC LINES DIFFERING BY <i>PINB-D1</i> ALLELE	74
ABSTRACT	74
CAPÍTULO 5 – COMPARAÇÃO DE DOIS MÉTODOS DE ESPECTROMETRIA DE MASSAS APLICADOS PARA MELHOR CARACTERIZAÇÃO DAS PROTEÍNAS DO TRIGO: HDMS^E x UDMS^E	75
RESUMO	75
ABSTRACT	76
1 INTRODUÇÃO	78
2 METODOLOGIA	80
2.1 MATERIAL VEGETAL	80
2.2 REAGENTES QUÍMICOS	80
2.3 ALQUILAÇÃO E EXTRAÇÃO SEQUENCIAL DAS PROTEÍNAS DE FARINHA DE TRIGO.....	81
2.4 QUANTIFICAÇÃO DE PROTEÍNAS.....	82
2.5 ETAPA DE LAVAGEM (<i>CLEAN-UP</i>)	84
2.6 DIGESTÃO TRÍPTICA.....	84
2.7 AQUISIÇÃO EM 1D NanoUPLC-MS ^E	85
2.8 ETAPA DE SCOUTING RUN	85
2.9 MÉTODOS HDMS ^e E UDMS ^E	86
2.10 PROCESSAMENTO DE DADOS	87
3 RESULTADOS E DISCUSSÃO	88
4 CONCLUSÃO	106
AGRADECIMENTOS	106

CAPÍTULO 6- DISCUSSÃO	108
6 DISCUSSÃO	108
6.1 PRINCIPAIS DESAFIOS ENCONTRADOS	108
6.2 ESTRATÉGIAS UTILIZADAS PARA CARACTERIZAÇÃO DO PERFIL PORTEICO DAS AMOSTRAS DE DIFERENTES APTIDÕES TECNOLÓGICAS...	111
6.3 PRINCIPAIS RESULTADOS	114
CAPÍTULO 7 – CONSIDERAÇÕES FINAIS	116
REFERÊNCIAS	118
APÊNDICES	130

INTRODUÇÃO

A ampla utilização do trigo na dieta humana está relacionada às suas características viscoelásticas únicas que permitem seu emprego em vários produtos, incluindo massas, pães e outros produtos de panificação (BÉKÉS, 2012; VENSEL *et al.*, 2014). Essa versatilidade faz do trigo uma matriz alimentícia de grande importância econômica e nutricional cuja produção mundial ultrapassa 700 milhões de toneladas (Mt) por ano. No Brasil, a estimativa da produção de trigo para o ano de 2022 é de aproximadamente 7,9 Mt, o que o faz ser o terceiro cereal em produção no país, atrás apenas do milho (112 Mt) e do arroz (10,6 Mt). No entanto, a demanda de trigo, estimada em 12,7 Mt, é muito superior à produção e o país necessita importar o cereal para garantir segurança alimentar à população (CONAB, 2022a).

Produzir trigo de forma competitiva e sustentável tem sido um dos principais desafios da agricultura brasileira, não somente no que tange técnicas de cultivo adequadas, mas também na obtenção de bons rendimentos e aptidão tecnológica (PIRES *et al.*, 2011). A aptidão tecnológica define a orientação industrial de uso do trigo (PIRES *et al.*, 2011) e está relacionada com o teor e a qualidade de suas proteínas. Em grãos maduros, as proteínas metabólicas (albuminas e globulinas, A/G) ou não-prolaminas, compreendem de 15 a 20 % do total de proteína do grão. São compostas por proteínas monoméricas fisiologicamente ativas ou estruturais, representadas principalmente por enzimas metabólicas, reguladoras e protetoras (GOESAERT *et al.*, 2005; HILL *et al.*, 2008; SAULNIER, 2012). Enquanto que as proteínas de reserva ou prolaminas representam 80-85%, sendo compostas por subunidades monoméricas, gliadinas, capazes de formar pontes dissulfeto intramoleculares, e polipeptídeos de gluteninas, formados por subunidades de alta e baixa massa molecular, estabilizados por pontes dissulfeto intermoleculares (WIESER, 2007).

Ambas as frações proteicas, prolaminas e não-prolaminas possuem propriedades que são determinantes para a aptidão tecnológica do trigo. A relação das proteínas de reserva (prolaminas) com a qualidade da farinha há tempos tem sido estudada (AGHAGHOLIZADEH *et al.*, 2017; BARAK *et al.*, 2013; PAYNE, 1987; POPINEAU *et al.*, 1994a). Enquanto as gliadinas fornecem plasticidade/viscosidade, as gluteninas contribuem para a força e elasticidade das massas de farinha de trigo e uma relação

gliadina/glutenina adequada é importante para a qualidade da proteína do glúten na panificação (BARAK *et al.*, 2013; GOESAERT *et al.*, 2005). Ressalta-se que, a presença de subunidades específicas dessas proteínas pode ser mais importante do que a quantidade total delas (LIU *et al.*, 2012). Por exemplo, na panificação, uma maior quantidade de subunidades de glutenina de alta massa molecular (*high molecular weight - glutenin subunits*, HMW-GS) levou a um melhor volume do pão e menor firmeza do miolo (BARAK *et al.*, 2013).

No que tange às proteínas metabólicas, determinadas proteínas podem impactar na aptidão tecnológica do grão pela função que exercem. Foi relatada a participação de chaperonas e isomerases, como a proteína dissulfeto isomerase (PDI), no enovelamento adequado das proteínas de reserva (KIMURA *et al.*, 2015; OSIPOVA *et al.*, 2012). Enquanto PDI participa do enovelamento tridimensional e assegura que as pontes dissulfeto aconteçam corretamente (DONG *et al.*, 2012; OSIPOVA *et al.*, 2012), proteínas de choque térmico (*heat shock proteins*, HSPs) atuam no reenovelamento de proteínas e as proteínas de choque térmico de baixa massa molecular (*small HSPs*) ligam-se e sequestram proteínas com enovelamento incorreto (MOGK *et al.*, 2019). Outras proteínas dessa fração, as puroindolinas (PINs), são conhecidas por influenciar a dureza do grão alterando a textura do endosperma (PAULY *et al.*, 2013), bem como podem afetar o comportamento agregativo de proteínas de reserva (GENEIX *et al.*, 2020; LESAGE *et al.*, 2012).

No Brasil, análises físico-químicas e reológicas são amplamente empregadas para avaliação da aptidão tecnológica do trigo e objetivam caracterizar as propriedades do grão, da farinha e o comportamento da massa por meio de análises como, por exemplo, alveografia e farinografia (AACCI, 2010). No entanto, aspectos como o teor de proteína total, relação gliadina/glutenina, presença de ligações cruzadas por pontes dissulfeto intermoleculares, porcentagem de massa extraível do macropolímero de glutenina subunidades de gliadinas e gluteninas são os principais fatores investigados para compreensão do papel das proteínas na aptidão tecnológica do trigo, principalmente para panificação (LANGENKÄMPER; ZÖRB, 2019).

Sabe-se que a análise de proteínas é uma tarefa complexa devido às inúmeras possibilidades de combinações aminoacídicas e pela complexidade das interações

proteicas e enzimáticas (SOUZA *et al.*, 2017). A análise das proteínas de reserva do trigo apresenta dificuldades adicionais devido às limitações dos bancos de dados disponíveis e à presença de conjuntos de proteínas homólogas (AGHAGHOLIZADEH *et al.*, 2017). Neste sentido, abordagens proteômicas com base em técnicas sensíveis e confiáveis, revelam-se como importantes ferramentas de análise das proteínas do trigo (BROMILOW *et al.*, 2017b; VICTORIO *et al.*, 2018). A espectrometria de massas (*Mass Spectrometry*, MS) é a principal ferramenta para análises proteômicas em larga escala por apresentar alto rendimento de identificação e expressão de proteínas e também por fornecer informações detalhadas sobre interações proteína-proteína e modificações pós-traducionais (CUNSOLO *et al.*, 2014; DISTLER *et al.*, 2014; FERREIRA *et al.*, 2014).

A caracterização de amostras complexas é facilitada pela utilização de métodos conhecidos como aquisição independentes de dados (*Data-Independent Acquisition*, DIA), onde todos os íons gerados na fonte são transmitidos até a célula de colisão, que alterna entre baixa e alta energia, gerando dados de massa exata de precursores e fragmentos simultaneamente (DISTLER *et al.*, 2014; GEROMANOS *et al.*, 2009; SOUZA *et al.*, 2017). Esses métodos podem ainda ser complementados pela espectrometria por mobilidade iônica (*Ion Mobility Spectrometry*, IMS) que permite a separação adicional de íons que apresentam a mesma relação massa/carga (DISTLER *et al.*, 2016; SOUZA *et al.*, 2017).

Recentemente, estudos baseados em abordagens ômicas demonstraram a existência de diferentes perfis proteicos e de compostos bioativos e alergênicos em amostras de trigo brasileiro de diferentes aptidões tecnológicas (ALVES *et al.*, 2018; SANTOS *et al.*, 2022; VICTORIO *et al.*, 2018). Mais especificamente, ferramentas proteômicas foram aplicadas com sucesso no estudo das proteínas metabólicas, mostrando que as farinhas classificadas como fortes em relação à força do glúten apresentaram perfil proteico diferente daquelas de menor força do glúten (VICTORIO *et al.*, 2018). Visando contribuir neste campo de pesquisa, este trabalho teve como ponto de partida o estudo, por meio de abordagem proteômica, do perfil proteico de farinhas de trigo em função da aptidão tecnológica das amostras.

Esta tese de doutorado faz parte do projeto de pesquisa “Metabolômica e proteômica de alimentos”, cujo foco é a aplicação de ferramentas “ômicas” para gerar resultados inovadores na análise de alimentos e está atrelado à linha de pesquisa

“Processamento, qualidade, valorização de alimentos, coprodutos e resíduos” do Programa de Pós-Graduação em Alimentos e Nutrição (PPGAN). Seguindo essa linha, duas hipóteses nortearam este trabalho: i) existem diferenças no perfil proteico que explicam as diferentes aptidões tecnológicas do trigo; ii) a caracterização do perfil proteico de farinhas de trigo pela utilização de abordagem proteômica pode revelar possíveis marcadores proteicos relacionados à aptidão tecnológica do trigo.

Neste contexto, o objetivo central deste trabalho foi caracterizar os diferentes perfis proteicos de farinhas de trigo com aptidões tecnológicas contrastantes, aplicando proteômica comparativa não dirigida com base na cromatografia líquida de ultra performance (*Ultra Performance Liquid Chromatography*, NanoUPLC) acoplada à espectrometria de massas de alta definição (*High Definition Mass Spectrometry*, HDMS) com abordagem independente de dados.

Assim, alguns objetivos específicos puderam ser definidos: i) Revelar os diferentes perfis proteicos de amostras de farinha de trigo comum evidenciando as proteínas que contribuem para a aptidão tecnológica e uso final da farinha; ii) Identificar e quantificar proteínas do glúten em amostras agrupadas por diferentes aptidões tecnológicas distinguindo os perfis proteicos; iii) Revelar e avaliar o perfil proteico de amostras de farinha de trigo com força do glúten contrastantes (farinhas integrais e refinadas) e de farinhas refinadas de linhagens quase-isogênicas com fenótipos diferentes; iv) Apresentar e comparar métodos de aquisição de alta definição em NanoUPLC-MS/MS que permitem caracterização das proteínas do trigo.

Com esta pesquisa, buscou-se aprimorar o conhecimento sobre a aptidão tecnológica do trigo com base na expressão diferencial e na categorização funcional das proteínas das diferentes amostras de farinha de trigo.

Este manuscrito é composto por 7 capítulos divididos em 3 partes principais. A primeira parte foi destinada à revisão bibliográfica direcionada aos temas que embasam a produção científica obtida com esta tese. Os primeiros pontos abordados (capítulo 1) versam sobre a importância e padrão de qualidade do trigo, bem como sobre a estrutura do grão, suas frações proteicas e o impacto destas na aptidão tecnológica do grão. No capítulo 2 estão apresentadas as temáticas ligadas à análise das proteínas de trigo e de outros cereais utilizando abordagem proteômica: principais desafios encontrados para

análise das proteínas do trigo, apresentação da técnica de cromatografia líquida de ultra performance acoplada à espectrometria de massas de alta definição (do inglês, UPLC-HDMS), apresentação das estratégias e métodos em MS, como o uso da abordagem DIA, IMS e os diferentes métodos MS^E que permitem alavancar a análise da complexa fração proteica do trigo e a utilização da abordagem proteômica em cereais.

A segunda parte trata da produção científica gerada a partir das análises experimentais conduzidas durante o desenvolvimento da tese. Neste sentido, estão apresentados três artigos originais. O primeiro (capítulo 3), foi publicado na revista *Journal of Proteomics* (FI:4.044; Qualis único CAPES A1) e intitula-se *NanoUPLC-MS^E reveals differential abundance of gluten proteins in wheat flours of different technological qualities*. Neste artigo foi aplicada a técnica hifenada nanoUPLC-MS^E em busca da melhor compreensão das proteínas de reserva do trigo em amostras de diferentes aptidões tecnológicas, previamente classificadas pela força do glúten em fraca, média e forte. Os resultados apontaram a distinção do perfil proteico das amostras de mais baixa força do glúten em comparação ao das amostras com elevada força do glúten. Além disso, o trabalho mostrou a efetividade da aplicação de nanoUPLC-MS^E para caracterização do glúten, contribuindo para a consolidação das pesquisas no campo da aptidão tecnológica do trigo e trazendo informações inéditas principalmente sobre o trigo brasileiro.

Compõe o capítulo 4, o artigo intitulado *Untargeted proteomics unravels the protein profile of wheat flours in contrasting gluten-force genotypes and near-isogenic lines differing by Pinb-D1 allele*, que será submetido à revista *Food Chemistry* (FI:7.514; Qualis único CAPES A1). Neste trabalho, foram analisadas amostras brasileiras de farinha de trigo integral e refinada com características contrastantes em termos de força do glúten, além de duas amostras de origem francesa de farinha de trigo refinada que diferiam pelo alelo *Pinb-D1*, que codifica puroindolina B, conferindo textura *hard* ou *soft* ao grão. Como principal resultado, as amostras apresentaram diferentes perfis proteicos apontando marcadores específicos de amostras contrastantes. Maior expressão de gliadinas, LMW-GS, HMW-GS e *small HSPs* (do inglês, *small heat shock proteins*) foi verificada em amostras de alta força do glúten, enquanto as amostras de baixa força do glúten expressaram mais PINA, PINB2 e GSP (do inglês, *grain softness protein*). Finalmente, as

amostras que apresentavam alta força do glúten e fenótipo *hard* foram agrupadas e separadas de seus pares com características contrastantes por análise quimiométrica. Assim, o trabalho apontou proteínas que podem ser usadas como possíveis biomarcadores para reconhecimento de aptidão tecnológica.

Por fim, o artigo que compõe o capítulo 5 foi publicado como um capítulo no livro *Nutrição em foco: uma abordagem holística*, volume III, Editora Conhecimento Livre, com o título: *Comparação de dois métodos de espectrometria de massas aplicados para melhor caracterização das proteínas do trigo: HDMS^E x UDMS^E*. Neste trabalho foram analisadas duas amostras de trigo brasileiro com aptidões tecnológicas contrastantes por nanoUPLC acoplada à espectrometria de massas. Foram comparados dois métodos DIA, o de alta definição - *High Definition MS^E* (HDMS^E) e o de ultra definição - *Ultra Definition MS^E* (UDMS^E), que utiliza o sistema de mobilidade iônica onde os íons dos peptídeos são separados em função de sua conformação tridimensional. Os resultados evidenciaram a superioridade do método UDMS^E em termos da cobertura de sequências proteicas, em virtude da melhor eficiência na fragmentação. Ao fornecer novas perspectivas para avaliação da aptidão tecnológica do trigo, este trabalho contribuiu com a pesquisa e a inovação na cadeia produtiva do cereal.

No intuito de articular os dados apresentados nos capítulos de produção científica, foi adicionado na terceira parte deste trabalho, um capítulo de discussão (capítulo 6) que une os principais resultados encontrados para melhor entendimento do perfil proteico de amostras de farinha e, conseqüentemente, da aptidão tecnológica do trigo. Por fim, o capítulo 7 foi dedicado às considerações finais deste trabalho.

A seguir, é apresentada a estrutura da tese (Figura 1).

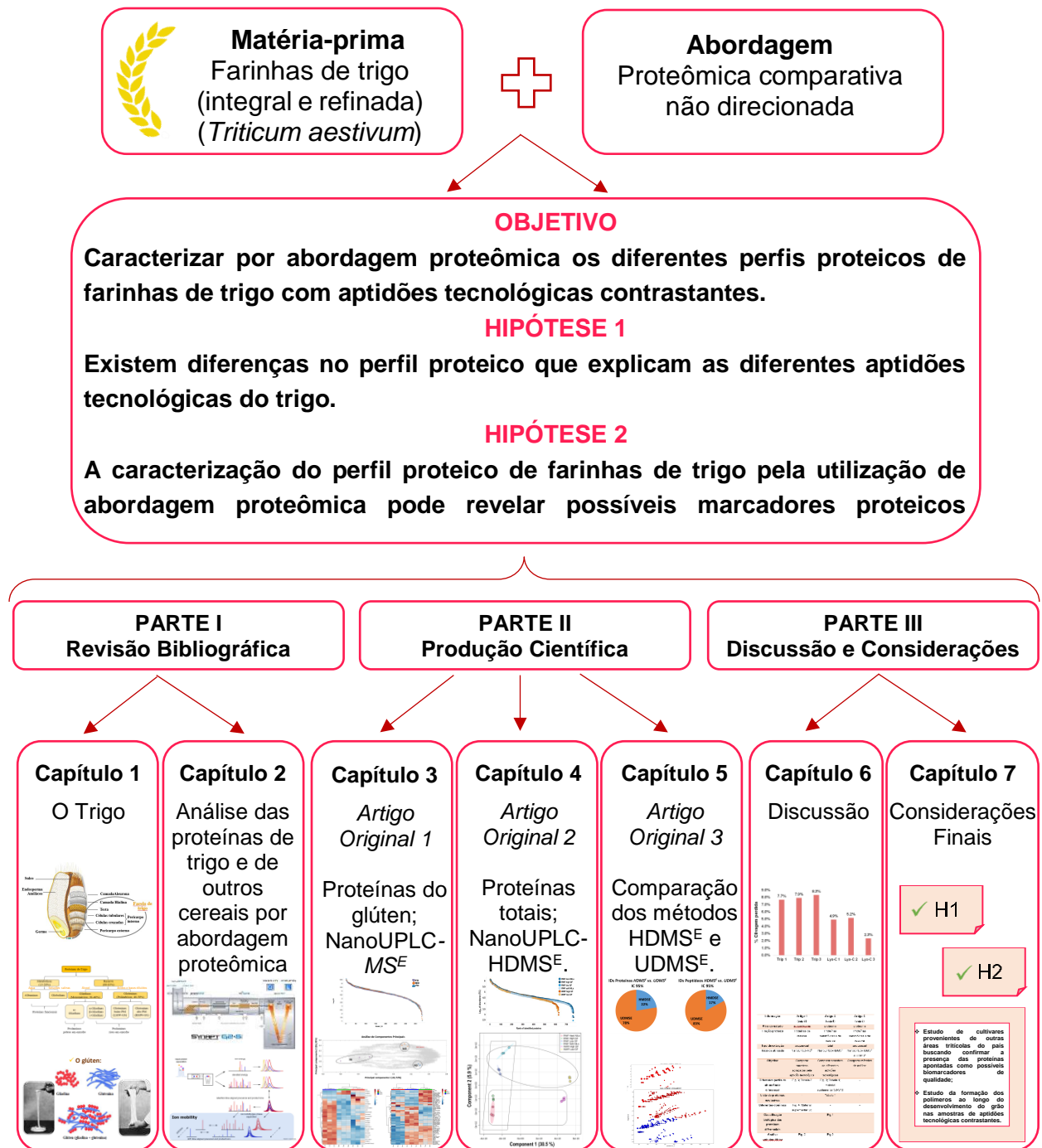


Figura 1. Estrutura da tese.

PARTE I

REVISÃO BIBLIOGRÁFICA

CAPÍTULO 1. O TRIGO

**CAPÍTULO 2. ANÁLISE DAS PROTEÍNAS DE TRIGO E DE OUTROS CEREAIS POR
ABORDAGEM PROTEÔMICA**

CAPÍTULO 1 – O TRIGO

- 1.1 Produção e consumo;
 - 1.2 Padrão de qualidade do trigo e da farinha de trigo no Brasil
 - 1.3 Estrutura e composição química do grão;
 - 1.4 Proteínas do trigo e seu impacto na aptidão tecnológica do grão
-

1 O TRIGO

1.1 PRODUÇÃO E CONSUMO

O trigo é um cereal pertencente à família botânica *Poaceae*, à subfamília *Pooideae* e ao gênero *Triticum*. Os tipos de trigo de maior interesse comercial são o trigo comum (*Triticum aestivum*), espécie hexaploide (AABBDD), mais utilizada na panificação, produção de bolos, biscoitos, massas e produtos de confeitaria, e o trigo duro (*Triticum durum*), espécie tetraploide (AABB), especialmente destinada ao preparo de massas alimentícias (SERNA-SALDIVAR, 2010; SHEWRY, 2009). O trigo é amplamente utilizado na alimentação humana na forma de grão laminado ou na forma de farinha. Além do uso nos produtos citados, pode ser usado como agente espessante em molhos, sopas, pudins e recheios de tortas (BORÉM; SCHEEREN, 2015).

A possibilidade de aplicação em diversos produtos alimentícios e o amplo consumo mundial (de mais de 769 Mt) fazem do trigo um alimento básico de grande importância econômica e nutricional. A produção mundial do cereal, considerando os anos de 2020/21, foi em torno de 773 Mt, com aumento de 4,6% na produção nos últimos 5 anos (ABITRIGO, 2022) e a estimativa de produção para 2022 é de 790 Mt (FAO, 2022). Países como Canadá, Estados Unidos, Índia, Rússia e Ucrânia e a União Europeia possuem a produção maior do que o consumo, enquanto no Brasil a produção representa apenas cerca de 51% da demanda (tabela1) (ABITRIGO, 2022).

Tabela 1. Produção e consumo de trigo em 2020/21.

Local	produção (Mt)	consumo (Mt)
Brasil	6,300	12,200
Canadá	35,183	10,000
China	136,000	140,000
Estados unidos	49,691	21,380
índia	107,592	103,017
rússia	84,000	41,500
Ucrânia	25,500	8,100
união europeia	135,800	118,500
total mundial	773,663	769,320

Fonte: Serviço Agrícola Estrangeiro - Departamento de Agricultura dos Estados Unidos. 16/02/2021 (ABITRIGO, 2022).

No Brasil, o cultivo de trigo é direcionado para a espécie *T. aestivum*, sobretudo por imposições edafoclimáticas (solo, clima, temperatura e foto exposição), sendo classificado como uma cultura de inverno. Apesar de ser uma cultura de baixa necessidade hídrica, a estiagem e a baixa precipitação diminuem o rendimento, bem como geadas na fase reprodutiva e altas temperaturas, também são prejudiciais (CONAB, 2022b). Contudo, os esforços dos programas de melhoramento genético e as melhorias nos sistemas de produção têm proporcionado aumento do rendimento do trigo (BORÉM; SCHEEREN, 2015). Com isso, segundo a Companhia Nacional de Abastecimento (CONAB), o trigo é o quarto grão em termos de produção no país, ficando atrás apenas da soja (leguminosa; 125,47 Mt) e dos cereais milho (112,34 MT) e arroz (10,56). A produção de trigo está estimada em cerca de 7,98 Mt para o ano de 2022 e, diferente do que ocorre para outros grãos, a demanda do trigo supera a sua produção. Dessa forma, para suprir o consumo de cerca de 12,7 Mt, em 2022 está prevista a importação de 6,5 Mt de trigo (CONAB, 2022a).

No Brasil, o melhoramento de trigo é realizado por empresas públicas (ex. Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA e Fundação Estatal de Pesquisa Agropecuária do Rio grande do Sul - FEPAGRO) e privadas (ex. OR Melhoramento de

Sementes e Biotrigo Genética). Os principais objetivos do melhoramento de trigo são a avaliação e a identificação de novos materiais com genes de características importantes de resistência, de adaptação, de produtividade e/ou aptidão tecnológica de uso (BORÉM; SCHEEREN, 2015).

1.2 PADRÃO DE QUALIDADE DO TRIGO E DA FARINHA DE TRIGO NO BRASIL

No Brasil, os parâmetros de qualidade do trigo são definidos pelo Regulamento Técnico do Trigo estabelecido pela Instrução Normativa (IN) nº 38 (BRASIL, 2010) do Ministério da Agricultura, Pecuária e Abastecimento (MAPA) que classifica o trigo em grupos, classes e tipos. Quanto aos grupos, são dois: I – trigo destinado diretamente à alimentação humana e II – trigo destinado à moagem e outras finalidades. O trigo do grupo II é classificado em classes, segundo parâmetros apresentados na tabela 2. Os trigos das classes Melhorador e Pão são indicados para a fabricação de pães industriais, massas alimentícias secas e biscoitos tipo cracker; os trigos das classes Pão e Doméstico são destinados à produção de pães caseiros e uso doméstico; enquanto os trigos das classes Básico e Outros usos são indicados na produção de biscoitos e bolos. Quanto ao tipo, o trigo pode ser tipo 1, 2, 3 e fora de tipo, considerando limites máximos de tolerância de impurezas, matérias estranhas e defeitos.

Tabela 2. Classes do trigo do Grupo II destinado à moagem e outras finalidades.

Classes	Força do Glúten (Valor mínimo expresso em $10^{-4}J$)	Estabilidade (Tempo expresso em minutos)	Número de Queda (Valor mínimo expresso em seg.)
Melhorador	300	14	250
Pão	220	10	220
Doméstico	160	6	220
Básico	100	3	200
Outros usos	Qualquer	Qualquer	Qualquer

Fonte: Instrução Normativa do MAPA nº 38 (BRASIL, 2010)

A IN38 (BRASIL, 2010) estabelece alguns conceitos como: trigo - grãos provenientes das espécies *Triticum aestivum* L e *Triticum durum* L; estabilidade - o tempo, em minutos, que uma massa mantém estável suas características viscoelásticas, quando submetida ao processo de amassamento; força do glúten - o trabalho mecânico necessário para expandir a massa até a sua ruptura, expressa em Joules (J); número de queda (*Falling Number*) - medida indireta da atividade da enzima alfa-amilase, determinada em trigo moído com valor expresso em segundos e peso do hectolitro (ou peso hectolítrico) - a massa de 100 (cem) litros de trigo, expressa em quilogramas (kg). Para avaliar a qualidade do grão são usados parâmetros físicos, como peso de mil grãos e peso do hectolitro (PH) bem como parâmetros físico-químicos como número de queda, umidade, teor de proteínas e cinzas.

Para fins de comercialização o peso do hectolitro (PH) é o parâmetro mais utilizado em vários países, pois representa uma avaliação indireta da qualidade dos grãos (COSTA *et al.*, 2008), onde impurezas e grãos malformados influenciam e reduzem a qualidade. Já o peso de mil grãos, é um parâmetro que classifica o trigo pelo seu tamanho. O tamanho excessivo não é bem visto pelas indústrias por impor dificuldades de regulação de equipamentos de limpeza e moagem e grãos pequenos podem passar pelas peneiras de limpeza, em ambos os casos, gerando perdas (GUTKOSKI *et al.*, 2008). Além de ser influenciada pela qualidade do produto, a definição do preço do trigo sofre influência de diversos outros fatores como a frustação da safra por condições climáticas adversas, os preços dos produtos associados (como o milho), a relação cambial e a disponibilidade de trigo nos países vizinhos (Argentina, Uruguai e Paraguai). Para além desses fatores, a inocuidade e a aptidão tecnológica também contribuem para definição do preço do cereal (BORÉM; SCHEEREN, 2015).

Para nomenclatura do trigo, no Brasil, não são utilizados os termos “*soft*” e “*hard*” como acontece em outros países. Em geral, na América do Norte essas duas terminologias são usadas para *T. aestivum* e o termo “*durum*” para *T. turgidum* ssp. *durum*. No entanto, em alguns países da Europa o termo “*soft*” refere-se ao *T. aestivum* e o termo “*hard*” ao *T. turgidum* ssp. *durum*. Eles apresentam diferentes texturas do endosperma que afetam o uso preferencial da farinha. Em geral, a farinha de trigo mole

(*soft*) é usada para bolo e biscoito, farinha de trigo duro (*hard*) para pão e sêmola de trigo *durum* para massas (PAULY *et al.*, 2013).

Os parâmetros de qualidade da farinha foram estabelecidos pelo Regulamento Técnico de Identidade e Qualidade da Farinha de Trigo de que trata a IN nº 08 (BRASIL, 2005). Segundo a IN08, farinha de trigo é um produto elaborado com grãos de trigo (*Triticum aestivum* L.) ou com grãos de outras espécies de trigo do gênero *Triticum*, ou combinações por meio de trituração ou moagem e outras tecnologias ou processos. A farinha de trigo integral tem as mesmas especificações, mas a partir do processamento completo do grão limpo, contendo ou não o gérmen. As farinhas são, então, classificadas em tipos conforme a tabela 3.

Tabela 3. Limites de tolerâncias para farinhas de trigo.

Tipos	Teor de Cinzas (b.s. máximo)	Granulometria	Teor de Proteínas (b.s.; mín.)	Acidez Graxa (mg de KOH /100g) (máx)	Umidade (máximo)
Tipo 1	0,8	95% do produto deve passar em peneira com abertura de malha de 250 µm.	7,5%	100	15%
Tipo 2	1,4		8,0%	100	15%
Integral	2,5	-	8,0%	100	15%

Fonte: Instrução Normativa 08 (BRASIL, 2005). (b.s.) base seca.

A taxa de extração utilizada para moagem dos grãos afeta a composição química da farinha e corresponde ao grau de separação do endosperma e, conseqüentemente ao rendimento. Assim, uma separação eficiente e limpa do endosperma do farelo durante a moagem é importante para o rendimento e, quanto mais frações da camada aleurona permanece junto ao endosperma, maior é o teor de cinzas da farinha (HEMERY *et al.*, 2007; JI *et al.*, 2020).

1.3 ESTRUTURA E COMPOSIÇÃO QUÍMICA DO GRÃO

O trigo é um cereal com inflorescência do tipo espiga, onde as flores crescem ao longo de um eixo. Botanicamente, trigo é uma monocotiledônea por apresentar um só cotilédono na ocasião da germinação (SERNA-SALDIVAR, 2010). Em termos de composição química, os grãos de trigo são compostos por carboidratos (65-75%, representados por amido e fibras), proteínas (7-12%), lipídios (2-6%), água (12-14%) e micronutrientes (HEMERY *et al.*, 2007). O grão de trigo é chamado de cariopse e é essencialmente composto pelo gérmen, envelopes externos (farelo) e endosperma (SURGET; BARRON, 2005) (Figura 2).

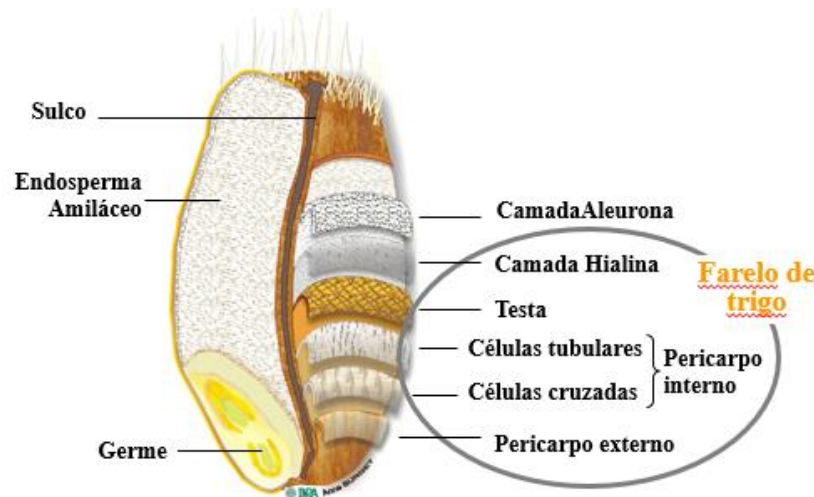


Figura 2. O grão de trigo e suas diferentes camadas. Adaptado de Surget e Barron (2005).

Cada tecido possui uma estrutura e composição diferente: o gérmen (3 % do grão) é rico em lipídeos, vitaminas do complexo B e minerais; os envelopes externos (pericarpo e testa) são ricos em fibras alimentares, e possuem a maior parte das vitaminas, minerais e compostos fenólicos (HEMERY *et al.*, 2007). O endosperma constitui-se da camada aleurona e do endosperma amiláceo. Este último representa cerca de 80-85% do grão, apresentando células repletas de grânulos de amido incorporados em uma matriz proteica, formada essencialmente por proteínas de reserva (SAULNIER, 2012). A camada aleurona compõe cerca de 7 % da massa seca do grão de trigo, contém a maior parte das vitaminas do complexo B e cerca de metade do conteúdo mineral total. Comparada com as camadas mais periféricas, a camada aleurona também possui um alto teor de proteínas e, estas, apresentam melhor equilíbrio de aminoácidos do que o

proteínas do endosperma, principalmente por possuírem níveis mais altos de lisina (HEMERY *et al.*, 2007).

Considerando a composição química, a retirada do gérmen é interessante para a produção de farinha com maior estabilidade no armazenamento, uma vez que a presença de lipídios e de enzimas hidrolíticas e oxidativas do gérmen podem causar rancidez à farinha. Adicionalmente, para moagem dos grãos, pode ser realizado um pré-processamento com retirada das camadas mais externas, de forma a tornar o processo mais eficiente, melhorar a qualidade dos produtos e separar frações de interesse. (HEMERY *et al.*, 2007).

1.4 PROTEÍNAS DO TRIGO E SEU IMPACTO NA APTIDÃO TECNOLÓGICA DO GRÃO

De acordo com a solubilidade, as proteínas do trigo foram classificadas por Osborne (1907) em quatro classes: albuminas solúveis em água, globulinas em soluções salinas, gliadinas em soluções alcoólicas e gluteninas em ácidos ou bases diluídas. Mais tarde, outra classificação foi proposta por Shewry *et al.* (1986), considerando o grau de polimerização (determinado pelas pontes dissulfeto intermoleculares) e o teor de aminoácidos sulfurados das proteínas de reserva, conhecidas genericamente no trigo como prolaminas: as gliadinas são prolaminas monoméricas e as gluteninas são prolaminas poliméricas, constituídas de subunidades de baixa e alta massa molecular, que formam agregados estabilizados por pontes dissulfeto (Figura 3).

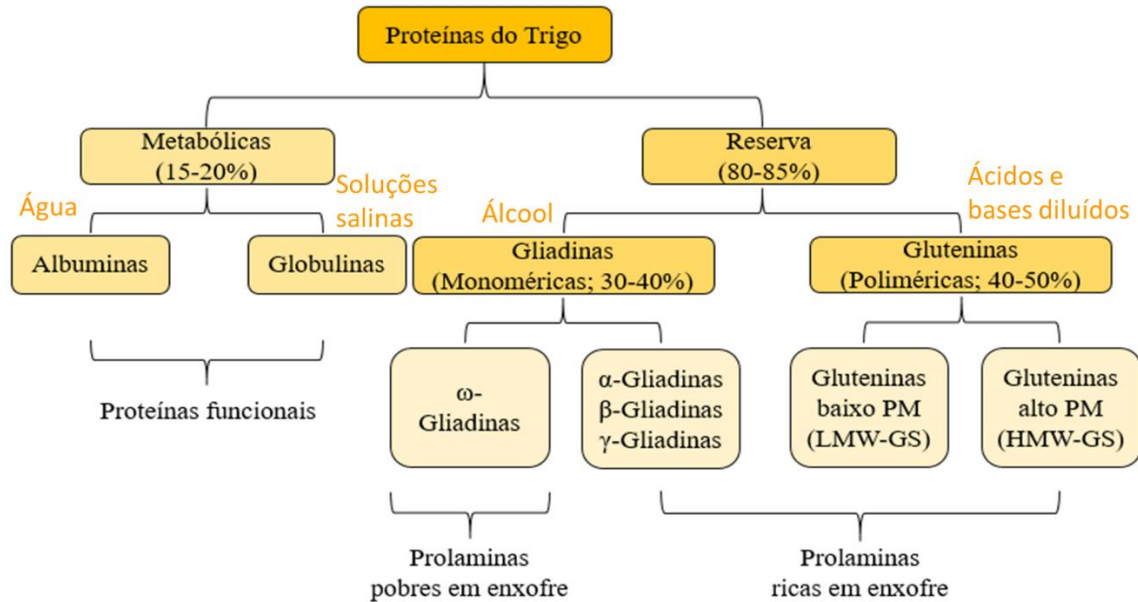


Figura 3. Composição das proteínas do trigo: aproximação entre as classificações de Osborne (1907) e Shewry (1986).

1.4.1 Proteínas metabólicas

As albuminas e globulinas são proteínas metabólicas conhecidas como não-prolaminas e, em cereais, encontram-se no farelo do grão e também nas células da camada aleurona que constituem o endosperma (KHAN *et al.*, 2015). Estas proteínas compreendem de 15 a 20% do total de proteína do grão e são, principalmente enzimas com massa molecular em torno de 25 kDa, podendo variar de 60 a 70 kDa (GOESAERT *et al.*, 2005). Nutricionalmente, são ricas em aminoácidos essenciais como triptofano, metionina e lisina (KHAN *et al.*, 2015).

As albuminas são α -amilases, inibidoras de proteases e outras enzimas com diferentes funções fisiológicas (TOMIĆ *et al.*, 2015). Na avaliação feita por Gao *et al.* (2009), em duas amostras de trigo, a distribuição funcional de A/G mostrou que as proteínas foram mais associadas ao metabolismo de carboidratos (30%), seguido de estresse/defesa (22%), proteínas de reserva (15%) e síntese/montagem/degradação de proteínas (7%). Khan *et al.* (2015) avaliaram a fração albumina de 14 variedades de trigo e verificaram que a massa molecular variou 8 a 72 kDa e suas funções biológicas principais foram relacionadas com a quebra do amido. Victorio *et al.* (2018) avaliaram a fração A/G em amostras de trigo agrupadas por aptidão tecnológica e encontraram a

distribuição nas funções moleculares de ligação (45%), ação enzimática de hidrolases, transferases e oxido redutases (29%), proteínas de ligação iônica (9%), estrutural (4%), transporte (4%) e reguladora (3%).

As puroindolinas, proteínas desta fração, parecem apresentar função biológica de defesa nas plantas. Mas elas são conhecidas por influenciar na textura do endosperma do grão, estando ligadas à base molecular da dureza do trigo (PAULY *et al.*, 2013). PINs são proteínas básicas e ricas em cisteína que aparecem em duas isoformas no trigo, puroindolina A (PINA) e puroindolina B (PINB). (LULLIEN-PELLERIN; MARION, 2002). PINA e PINB possuem 10 resíduos de cisteína. Os genes (*Pina-D1* e *Pinb-D1*) que codificam estas proteínas fazem parte do locus *Ha* do braço curto do cromossomo 5D. Mutações nestes genes resultam em alteração na textura do endosperma. Para o gene *Pina-D1* a mutação mais comum é uma mutação nula (*Pina-D1b*), enquanto para *Pina-D1* existem várias (*Pina-D1b-g*) todas resultando em grãos duros (PAULY *et al.*, 2013).

1.4.2 Proteínas de Reserva

As gliadinas e gluteninas são consideradas proteínas de reserva, pois suas funções são armazenar nutrientes e fornecer nitrogênio durante a germinação das sementes (FEILLET, 2000). Elas são caracterizadas por serem insolúveis em água e por possuírem alto teor de prolina e glutamina (VAN DEN BROECK *et al.*, 2015). Essas proteínas correspondem a 80-85% do conteúdo proteico total do grão e são encontradas principalmente no endosperma amiláceo (GOESAERT *et al.*, 2005). Estão divididas em dois grupos de acordo com sua habilidade em formar polímeros através de pontes dissulfeto intermoleculares entre subunidades proteicas (DAHESH *et al.*, 2014). Estas ligações fazem com que essas proteínas estejam envolvidas com a aptidão tecnológica do trigo (LUTZ *et al.*, 2012). As gliadinas, solúveis em álcool, conferem viscosidade e extensibilidade à massa e as gluteninas, insolúveis em álcool, relacionam-se com força e elasticidade (VENSEL *et al.*, 2014; WIESER, 2007). A insolubilidade em água e a possibilidade de associações covalentes e não-covalentes são propriedades que permitem a essas proteínas serem isoladas como uma massa coesa, o glúten (SHEWRY, 2009), que proporciona viscoelasticidade à massa e boa retenção de gás, o que constitui um importante papel na funcionalidade de panificação (BÉKÉS, 2012).

- GLIADINAS

As gliadinas podem ser divididas em α/β -, γ - e ω -gliadinas (Figura 2), com base na mobilidade em gel de eletroforese. Por possuírem pontes dissulfeto intramoleculares, apresentam-se como monômeros, sendo α/β - e γ -gliadinas encontradas em maiores proporções do que ω 1,2-, ω 5-gliadinas. Este fato está relacionado com o genótipo da variedade e com as condições de crescimento do trigo (WIESER, 2007). Os genes que codificam as gliadinas estão no braço curto dos cromossomos 1 e 6 dos genomas A e B e seus lócus são chamados *Gli-1* e *Gli-2*, respectivamente (JOPPA *et al.*, 1983).

Cada tipo de gliadina tem como característica dois domínios diferentes: o N-terminal com sequências repetitivas ricas nos aminoácidos glutamina, prolina, fenilalanina e tirosina e o C-terminal que não é repetitivo e contém as cisteínas (WIESER, 2007). Em geral, α -gliadinas contêm seis resíduos de cisteínas e γ -gliadinas possuem oito podendo fazer, respectivamente, três e quatro ligações intramoleculares (Figura 4), enquanto ω -gliadinas não possuem cisteínas (LUTZ *et al.*, 2012).

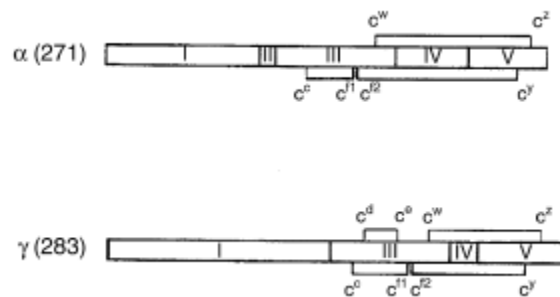


Figura 4. Estruturas das α - e γ -gliadinas com ligações dissulfeto intramoleculares apresentadas por Wieser (2007).

Apesar da separação das gliadinas como monômeros e gluteninas como polímeros, algumas gliadinas contêm número ímpar de cisteínas decorrentes de mutações de aminoácido pontuais (LUTZ *et al.*, 2012). As cisteínas livres fazem com que essas gliadinas também sejam capazes de formar pontes dissulfeto intermoleculares, agindo como terminadores de cadeia. Elas têm estrutura de gliadina, mas funcionam como gluteninas (MUCCILLI *et al.*, 2010; VENSEL *et al.*, 2014).

- GLUTENINAS

As gluteninas são divididas em subunidades de alta massa molecular (*High-molecular-weight glutenin subunits* - HMW-GS) e de baixa massa molecular (*Low-molecular-weight glutenin subunits* - LMW-GS). Apresentam-se como polímeros porque possuem cisteínas livres que possuem grupamentos tióis (SH), podendo fazer, adicionalmente, pontes dissulfeto intermoleculares (MUCCILLI *et al.*, 2010; ROMBOUTS *et al.*, 2013; VENSEL *et al.*, 2014). Devido à natureza polimérica, a massa molecular das gluteninas pode variar de 500.000 a mais de 10 milhões de Daltons (Da) (LUTZ *et al.*, 2012; WIESER, 2007; WRIGLEY, 1996), tornando-as insolúveis mesmo em tampões desnaturantes como o dodecil sulfato de sódio (SDS), necessitando de redução e sonicação para solubilização (VENSEL *et al.*, 2014).

Assim como as gliadinas, as LMW-GS, possuem um domínio N-terminal repetitivo rico em prolina e glutamina e um C-terminal não repetitivo homólogo ao das α/β -, γ -gliadinas (WIESER, 2007). As LMW-GS contêm seis resíduos de cisteína em posições homólogas às de α/β -, γ -gliadinas, com três ligações intramoleculares, e dois resíduos adicionais responsáveis pelas ligações intermoleculares (LUTZ *et al.*, 2012; MUCCILLI *et al.*, 2010). Estas subunidades de glutenina podem ser classificadas como extensoras ou terminadoras de polímeros (MUCCILLI *et al.*, 2010) (Figura 5) e são codificadas por genes localizados no braço curto dos cromossomos 1 dos genomas A, B e D, cujos lócus são *Glu-A3*, *Glu-B3* e *Glu-D3* (LEW *et al.*, 1992).

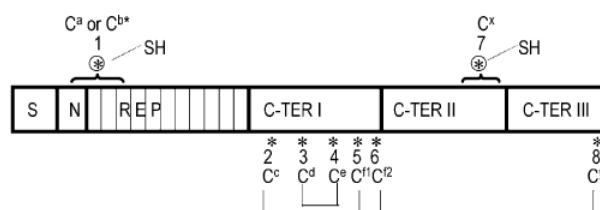


Figura 5. Esquema proposto para as estruturas típicas de LMW-GS por D'ovidio & Masci (2004). As cisteínas são designadas por números ou de acordo com o sistema de letras proposto por Kohler *et al.*(1993). SH ou asterisco representam as cisteínas que formam pontes dissulfeto intermolecular. S, peptídeo sinal, N, N-terminal, REP, domínio repetitivo e C-TER, domínios C-terminal.

As HMW-GS são codificadas por genes do braço longo dos genomas A, B e D. Seus lócus são chamados *Glu-A1*, *Glu-B1* e *Glu-D1* (LEW *et al.*, 1992; SHEWRY; HALFORD, 2002). Elas podem ser do tipo x (x- HMW-GS) ou do tipo y (y-HMW-GS), sendo as do

tipo-x as que possuem maior massa molecular relativa ($83.000 < Mr < 88.000$) comparada com as do tipo-y ($67.000 < Mr < 74.000$). Estruturalmente, as HMW-GS possuem dois domínios não repetitivos, denominados N- e C- terminais separados por um domínio central repetitivo (SHEWRY; TATHAM, 1990). Muitas variedades de trigo possuem quatro ou cinco HMW-GS, quando possuem quatro, duas são do tipo x e duas do tipo y e quando possuem cinco, três são do tipo x e duas são do tipo y (KASARDA, 1989). As x-HMW-GS possuem quatro ou cinco cisteínas, diferindo das y-HMW-GS que possuem sete (LUTZ *et al.*, 2012).

1.4.3 Impacto das proteínas metabólicas e de reserva na aptidão tecnológica do grão de trigo

A maior parte dos cereais é consumida em alimentos processados e, por isso, o impacto de suas proteínas nas propriedades funcionais do processamento de alimentos deve ser considerado (SHEWRY; HALFORD, 2002). No caso do trigo, o conteúdo proteico, inclusive, exerce forte influência na determinação do preço do cereal baseado no argumento de que mais proteína proporciona maior qualidade de cozimento, por exemplo, maior volume de pão (LANGENKÄMPER; ZÖRB, 2019). As propriedades únicas da massa feita com a farinha que permitem que o trigo seja processado em inúmeros produtos de panificação (BÉKÉS, 2012; SHEWRY, 2009) constituem uma vantagem do uso do cereal sobre as outras culturas (SHEWRY, 2009).

No que tange a função das proteínas metabólicas que pode impactar na aptidão tecnológica do trigo, estudos mostram que chaperonas e isomerases estão relacionadas ao enovelamento adequado das proteínas de reserva sintetizadas no retículo endoplasmático (KIMURA *et al.*, 2015; OSIPOVA *et al.*, 2012). A glicoproteína PDI, pertencente à família da tiorredoxina, possui massa molecular de 60 kDa (OSIPOVA *et al.*, 2012) e atua assegurando que o enovelamento tridimensional e as pontes dissulfeto ocorram corretamente (DONG *et al.*, 2012; OSIPOVA *et al.*, 2012), podendo dar origem a polímeros de glutenina regulares, culminando em maior qualidade do glúten (LIU *et al.*, 2012). Proteínas da família de proteínas de choque térmico (*heat shock proteins*, HSPs), como HSP60, HSP70, HSP90, realizam atividades de reenovelamento de proteínas de forma dependente de adenosina trifosfato (ATP), enquanto *small* HSPs ligam-se e

sequestram proteínas com enovelamento incorreto de forma independente de ATP. *Small HSPs* não exibem atividade de reenovelamento e degradação e atuam de forma distinta das outras chaperonas por impedirem que proteínas mal enoveladas formem grandes e insolúveis agregados proteicos (MOGK *et al.*, 2019). Outra proteína dessa fração, calreticulina, atua como ligante de Ca²⁺ no enovelamento proteico (JIA *et al.*, 2008), cooperando com HSP70 e PDI na síntese de proteínas do glúten no retículo endoplasmático (LV *et al.*, 2021).

Adicionalmente, já foi demonstrado que puroindolinas também interagem com as proteínas do glúten, afetando o comportamento de agregação de gliadinas em solução (GENEIX *et al.*, 2020) e que, na ausência de PINA, houve aumento do tamanho dos polímeros de proteínas de reserva (LESAGE *et al.*, 2012). Sabe-se que, juntas, PINA e PINB afetam a maciez do grão de trigo e, conseqüentemente, o rendimento da moagem, o rendimento de quebra da farinha e o amido danificado (HEINZE *et al.*, 2016; LULLIEN-PELLERIN, 2020). Quando ambas estão presentes, o grão apresenta fenótipo mole e, neste caso, há uma redução na resistência mecânica do grão (LULLIEN-PELLERIN, 2020; MA *et al.*, 2020). Na ausência de PINA (*alelo Pina nulo*) o grão apresenta fenótipo duro (LESAGE *et al.*, 2012; PAULY *et al.*, 2013), enquanto que, quando há super expressão do alelo *Pina-D1a* há uma redução significativa na dureza do grão (GENEIX *et al.*, 2020).

PINs também influenciam na porosidade do endosperma e, conseqüentemente, diminuindo a vitreosidade. A vitreosidade é uma propriedade física relacionada ao rendimento da quebra da farinha na moagem, quanto maior a vitreosidade, maior a energia necessária para quebrar os grãos. Grãos portadores de genes PIN do tipo selvagem (*Pina-D1a* e *Pinb-D1a*) apresentam menor vitreosidade, enquanto valores mais altos estão associados ao alelo com mutação *Pinb-D1b* ou *Pinb-D1d* (LULLIEN-PELLERIN, 2020). Assim, a transformação em variantes do alelo *Pinb* pode ser considerada uma importante estratégia para o melhoramento de trigo (MA *et al.*, 2020).

No sentido da relação entre as proteínas de reserva e a aptidão tecnológica da farinha, a quantidade e a distribuição da massa molecular dos polímeros de glutenina foram há tempos correlacionados com a qualidade em testes tecnológicos e de panificação (PAYNE *et al.*, 1987). Combinando resultados de abordagens genéticas (que

estudam o polimorfismo das proteínas do glúten e formas alélicas e seus aspectos na qualidade de processamento) e de abordagens bioquímicas e biofísicas (que mostram a capacidade de formar complexos poliméricos), a importância das HMW-GS foi evidenciado (SHEWRY, 2009). Sabe-se que muitos estudos têm como foco essas proteínas (BARAK *et al.*, 2013), mas cabe ressaltar que uma mistura apropriada das frações de gliadina e glutenina é essencial para as propriedades viscoelásticas da massa e qualidade do produto final (WIESER, 2007). Barak *et al.* (2013) mostraram que variedades de trigo com composição semelhante em HMW-GS, mas que possuíam razão gliadina/glutenina diferentes, apresentaram propriedade reológicas diferentes.

Está bem estabelecido que genótipos de aptidão tecnológica superior são associados a não somente um alto teor de HMW-GS, mas principalmente com o tipo de subunidade codificada pelos cromossomos, são elas a subunidade 1Ax e o par de subunidades 1Dx5+1Dy10 codificadas pelo cromossomo 1D, quando comparadas com outras subunidades alélicas, tais como as 1Dx2+1Dy12, 1Dx3+1Dy12 e 1Dx4+1Dy12 (SHEWRY; HALFORD, 2002). Segundo Costa *et al.* (2013), em programas de melhoramento genético, as HMW-GS do alelo 1 (Glu-A1) e alelo 5 + 10 (Glu-D1), assim como as LMW-GS do alelo c (Glu-A3) e alelo b (Glu-B3) são preferidas para o desenvolvimento de variedades com força do glúten e extensibilidade adequadas para produtos assados. Outro estudo apontou que as subunidade de HMW-GS que mais influenciaram na aptidão tecnológica das amostras de trigo foram Dy10, Dy3 e Dx5, enquanto HMW-GS 1Bx17 e LMW-GS foram as subunidade que mais influenciaram a qualidade do pão (AGHAGHOLIZADEH *et al.*, 2017).

Assim, percebe-se que há cada vez maior necessidade de entendimento da fração de proteínas do trigo para melhor entendimento da influência que exercem na aptidão tecnológica do trigo. Cabe ressaltar que todo o estudo a ser realizado, está ainda mais acessível com o progresso das ferramentas proteômicas (LIU *et al.*, 2012)

CAPÍTULO 2 – ANÁLISE DAS PROTEÍNAS DE TRIGO E DE OUTROS CEREAIS POR ABORDAGEM PROTEÔMICA

- 2.1 Desafios para análise proteômica do trigo
 - 2.2 Cromatografia líquida de ultra performance acoplada à espectrometria de massas
 - 2.3 Estratégias e métodos em espectrometria de massas
 - 2.4 Abordagem proteômica em trigo e outros cereais
-

2 ANÁLISE DAS PROTEÍNAS DE TRIGO E DE OUTROS CEREAIS POR ABORDAGEM PROTEÔMICA

2.1 DESAFIOS PARA ANÁLISE PROTEÔMICA DO TRIGO

O fracionamento baseado na solubilidade das proteínas do trigo, proposto por Osborne (1907), foi o ponto de partida para análise destas proteínas. Esta divisão apresenta desvantagens como a sobreposição entre as frações de proteínas e o conteúdo de proteínas que permanece sem extração, por isso, estabelecer e melhorar técnicas de extração é o primeiro grande desafio em abordagens proteômicas (LANGENKÄMPER; ZÖRB, 2019), uma vez que o método influencia os resultados (XHAFERAJ *et al.*, 2020). Como tentativa de superar os problemas de fracionamento, tem sido proposto o desenvolvimento de métodos que utilizem apenas uma solução para extrair o maior número de proteínas de grãos, como por exemplo, a utilização de solução à base de SDS ou o uso de ácido tricloroacético em acetona gelada (LANGENKÄMPER; ZÖRB, 2019). Sabe-se que, dependendo do protocolo utilizado, diferentes classes de proteínas podem ser extraídas (XHAFERAJ *et al.*, 2020) e que a escolha de qual método será utilizado depende de questões individuais inerentes aos objetivos da pesquisa a ser realizada. Assim, quando se busca por proteínas menos abundantes, pode ser necessário utilizar uma técnica de extração mais específica (LANGENKÄMPER; ZÖRB, 2019).

No que tange a extração das proteínas do glúten, para separar as gliadinas das gluteninas é necessário utilizar um solvente capaz de romper ligações de hidrogênio,

como ureia, mas para separar os polímeros de glutenina em unidades individuais é necessário o uso de agentes redutores, como 2-mercaptoetanol ou ditioneitol (em inglês, Dithiothreitol, DTT) (SHEWRY, 2009). Por possuírem muitas ligações covalentes, para solubilizar os polímeros de glutenina de maior tamanho, pode ser necessária a utilização de ultrassom (sonicação) (ALTENBACH *et al.*, 2016). No entanto, mesmo protocolos em várias etapas não alcançaram resultados efetivos na separação destas proteínas, como no caso de utilização de extração em dois passos utilizando 2-propanol e DTT (XHAFERAJ *et al.*, 2020).

Para a etapa de digestão, as principais enzimas utilizadas em análises proteômicas são tripsina e quimotripsina. A primeira é específica para clivagem C-terminal de resíduos de lisina e arginina e a segunda é menos específica e cliva, em nível C-terminal, tirosina, fenilalanina, triptofano e leucina, somente se não estiverem antes de resíduos de prolina (FIEDLER *et al.*, 2014). Sabe-se que as proteínas de reserva do trigo possuem baixo número de resíduos básicos, além de serem ricas em prolinas (P) e glutaminas (Q) e apresentarem sequências repetidas (PxQ) que são altamente complexas podendo tornar a digestão ineficiente com produção de peptídeos grandes que são mais difíceis de identificação por MS (AGHAGHOLIZADEH *et al.*, 2017; CLAVIJO *et al.*, 2017; VENSEL *et al.*, 2011). Uma alternativa para melhorar a digestão das proteínas de reserva do trigo, é a utilização de tripsina/Lys-C que possui maior eficiência do que a tripsina na clivagem em sítios de lisina e já foi relatado maior quantidade de peptídeos identificados após digestão com tripsina/Lys-C do que com quimotripsina em análise das proteínas do trigo (VINCENT *et al.*, 2022).

Passando para a etapa de identificação, esforços foram executados na construção de banco de dados visando melhorar a análise das proteínas do trigo. O primeiro banco de dados de sequências de proteínas de glúten de trigo de código aberto com curadoria manual foi criado por Bromilow *et al.* (2017a) e inclui 630 sequências discretas de proteínas. Trata-se do GluPro V1.0 que possui formato FASTA permitindo aplicação de métodos proteômicos para detecção e quantificação de proteínas do glúten. Adicionalmente, Clavijo *et al.* (2017) identificaram 105 genes de glúten completos ou parciais que permitiram a anotação adicional de 33 genes e a correção manual de 21 modelos de genes. Em 2018, foi lançada a referência de genoma do trigo, *RefSeq v.1.0*,

pelo Consórcio Internacional de Sequenciamento do Genoma do Trigo (*International Wheat Genome Sequencing Consortium*, IWGSC) que conta com acesso completo à sequência ordenada de todos os 21 cromossomos de trigo (IWGSC *et al.*, 2018). O RefSeq v.1.0 foi então, utilizado como genoma de referência para a criação do *WheatGmap*, que conta com mais de 3500 informações sobre o sequenciamento de trigo hexaploide, incluindo o sequenciamento do genoma inteiro, de exoma inteiro e transcriptoma (ZHANG *et al.*, 2021). A montagem completa e robusta do genoma do trigo é fundamental para o processo de melhoramento do cereal (CLAVIJO *et al.*, 2017).

Considerando que a identificação de proteínas por espectrometria de massas é dificultada pela alta similaridade entre as sequências primárias das subunidades das proteínas de reserva do trigo e pelo alto número de isoformas (LANGENKÄMPER; ZÖRB, 2019), a publicação do genoma de referência totalmente anotado contribuirá para superar os limites na pesquisa, principalmente quando aplicadas abordagens proteômicas. Abordagens proteômicas com base em técnicas que utilizam cromatografia líquida (LC) acoplada à espectrometria de massas de alta resolução (HDMS) capazes de separar até mesmo peptídeos de mesma m/z (i.e. sistema de mobilidade iônica) e métodos de análise independentes de dados (DIA), revelam-se como importantes ferramentas para a identificação e a quantificação das proteínas do trigo.

2.2 CROMATOLOGRAFIA LÍQUIDA DE ULTRA PERFORMANCE ACOPLADA À ESPECTROMETRIA DE MASSAS

De maneira breve, a cromatografia líquida é uma técnica de separação de componentes de uma amostra por meio de uma interação diferencial desses componentes entre uma fase móvel (solventes) e uma fase estacionária (coluna). Em comparação à cromatografia líquida de alta performance (HPLC), a cromatografia líquida de ultra performance (*Ultra Performance Liquid Chromatography*, UPLC), possui capacidade de trabalho em altas pressões (acima de 100 MPa - 15000 psi) devido ao tipo e colunas utilizadas. Além disso, na UPLC o tamanho das partículas que compõem as colunas é menor (<2 μm). Os cromatogramas obtidos por UPLC apresentam melhor resolução e separação e maior sensibilidade, além do menor tempo de análise e consumo de solvente (MALDANER; JARDIM, 2009; TALEUZZAMAN *et al.*, 2015).

A espectrometria de massas (*Mass Spectrometry*, MS) é uma técnica analítica utilizada para a identificação e quantificação de espécies químicas em suas formas ionizadas, através da medição de suas razões massa/carga em fase gasosa (m/z). Esta ferramenta tem se tornado cada vez mais útil no estudo de sistemas biológicos complexos, e tem ajudado na identificação, quantificação e mapeamento de importantes moléculas funcionais (AEBERSOLD; MANN, 2016; KRUIVE *et al.*, 2015). A MS é uma ferramenta essencial para análises proteômicas em larga escala (DISTLER *et al.*, 2014) e, por ser uma técnica capaz de fornecer informações moleculares específicas, tais como identificação da sequência de peptídeos, vem sendo utilizada para identificar e distinguir proteínas oriundas de diferentes cereais (FIEDLER *et al.*, 2014).

As duas técnicas acopladas LC-MS, têm sido utilizadas com sucesso em análises proteômicas não direcionadas que almejam caracterizar o proteoma e subproteoma (ex. proteínas metabólicas) do trigo (AFZAL *et al.*, 2021; VICTORIO *et al.*, 2018). É importante ressaltar que etapas precedentes à análise precisam ser realizadas (ALVES *et al.*, 2017). De maneira geral, essas etapas compreendem redução e alquilação dos extratos proteicos, que objetivam facilitar o processo de digestão por meio da clivagem das pontes dissulfeto e alquilação dos grupamentos tióis. A tripsina é a protease mais comumente utilizada para digestão de proteínas e costuma gerar peptídeos em uma faixa de massa de 500 a 3.000 Da, facilitando a separação cromatográfica. Além disso, os peptídeos trípticos possuem boa ionização e fragmentação devido à presença de resíduos C-terminais de lisina ou arginina que são protonados de forma eficiente em condições ácidas (GILLET *et al.*, 2016). No entanto, para digerir as proteínas de reserva do trigo, outras enzimas têm sido preferencialmente utilizadas para resolver o problema da falta de sítios de clivagem trípticos (ex: lisina), são elas a tripsina/Lys-C (VINCENT *et al.*, 2022), a termolisina e a quimotripsina (VENSEL *et al.*, 2011). Assim, a escolha da enzima, dos reagentes redutores e alquilantes é muito importante e têm sido associada a uma melhora na identificação das proteínas do glúten, pois permitem uma melhor distinção entre as proteínas homólogas (DUPONT *et al.*, 2011; FERREIRA *et al.*, 2014; ROMBOUITS *et al.*, 2013).

2.3 ESTRATÉGIAS E MÉTODOS EM ESPECTROMETRIA DE MASSAS

Em proteômica do tipo *shotgun*, também conhecida como *bottom-up*, as proteínas são digeridas em peptídeos pela ação de enzimas proteolíticas e a massa espectrométrica dos peptídeos é mensurada (GILLET *et al.*, 2016). Para análise livre de gel, os peptídeos são analisados por cromatografia líquida acoplada a espectrometria de massa sequencial (LC-MS/MS), em geral, utilizando aquisição dependente de dados (data-dependent acquisition, DDA) onde há a seleção dos íons precursores mais intensos com base na abundância relativa e estes são fragmentados em série. Esse tipo de método possui limitações como seleção de íons irreproduzíveis, sub-amostragem e longo ciclo instrumental (DISTLER *et al.*, 2014).

Visando diminuir os problemas supracitados, espectrômetros de massas mais recentes possuem maior resolução e velocidade de sequenciamento, além de ter precisão de massa melhorada. Adicionalmente, estratégias imparciais de aquisição denominadas independente de dados (*data-independent acquisition*, DIA) também foram criadas para superar as limitações do DDA, fazendo fragmentação simultânea de múltiplos íons precursores sem considerar intensidade ou outra característica, permitindo obtenção de resultados mais abrangentes (SOUZA *et al.*, 2017). Em métodos com a estratégia DIA, todos os íons gerados na fonte são transmitidos até a célula de colisão, que alterna entre baixa e alta energia, enviando alternadamente precursores e fragmentos ao analisador. Os resultados desses métodos são abrangentes e reproduzíveis porque ocorre a fragmentação simultânea de múltiplos íons precursores, independentemente da intensidade ou de outra característica. Dessa forma, são obtidos dados de massa exata de precursores e fragmentos ao mesmo tempo, tornando-os ideais para análise de amostras complexas por permitirem ampla cobertura de faixa dinâmica (DISTLER *et al.*, 2014; DISTLER *et al.*, 2016; GEROMANOS *et al.*, 2009).

Esses métodos que alternam entre duas condições de energia de colisão são denominados MS^E. Neles, os dados LC-MS são coletados em todo o experimento e o alinhamento baseia-se no fato de que o perfil cromatográfico dos íons precursores é semelhante aos dos seus íons precursores (GEROMANOS *et al.*, 2009). O método de alta definição denominado *High Definition MS^E* (HDMS^E) é um método MS^E acrescido do sistema de espectrometria por mobilidade iônica (*Ion Mobility Spectrometry*, IMS) que

permite que os íons dos peptídeos sejam separados pela sua conformação ao passarem por uma fase gasosa, sendo o tempo de aceleração proporcional e dependente da mobilidade através do gás. Essa dimensão adicional de separação reduz interferências e melhora a capacidade máxima do sistema, aumentando a seletividade do alinhamento dos precursores com seus produtos, pois a fragmentação ocorre depois do sistema IMS (DISTLER *et al.*, 2014; DISTLER *et al.*, 2016; SOUZA *et al.*, 2017).

O método conhecido como *Ultra Definition MS^E* (UMDS^E) é um método otimizado no qual a energia de colisão é específica para o tempo de aceleração, pois há uma rampa de energia de colisão de acordo com os dados obtidos na IMS. Dessa forma, a eficiência de fragmentação é melhorada, aumentando a intensidade de desempenho MS/MS para sequenciamento de peptídeos (DISTLER *et al.*, 2014; DISTLER *et al.*, 2016; SOUZA *et al.*, 2017). Por exemplo, Distler *et al.* (2014), demonstraram que o uso de UDMS^E melhorou a eficiência de fragmentação em 105% em comparação com métodos sem mobilidade iônica. O método foi utilizado recentemente para análise da fração A/G do trigo revelando ampla faixa dinâmica ($\log_{10} > 5$), identificação de 5.894 proteínas (considerando todas as amostras) e média de 8 peptídeos/proteínas (VICTORIO *et al.*, 2018).

2.4 ABORDAGEM PROTEÔMICA EM TRIGO E OUTROS CEREAIS

A espectrometria de massas tem sido utilizada com sucesso para análise de proteínas de cereais. Nesta seção serão apresentadas informações sobre os três cereais mais produzidos e consumidos no Brasil: milho, arroz e trigo. Recentemente, uma abordagem proteômica comparativa foi realizada para avaliação do milho brasileiro. Técnicas complementares de eletroforese bidimensional e proteômica *bottom-up* (livre de gel) usando NanoUPLC-MS/MS foram utilizadas para a identificação das proteínas diferencialmente abundantes (*differentially abundant proteins*, DAP) envolvidas em diferentes vias metabólicas. Os resultados mostraram diferentes perfis proteicos entre as cultivares de milho biofortificado (que apresenta proteínas de melhor qualidade em termos de aminoácidos essenciais) e a amostra híbrida comercial. Estes resultados podem auxiliar em estratégias de melhoramento genético para cultivares de milho com melhor desempenho nutricional e agrônômico (DOS SANTOS-DONADO *et al.*, 2021).

Análise por LC-MS/MS também foi aplicada para análise de grãos de milho imaturos em avaliação de proteínas com determinada modificação pós-traducional (ubiquitinação). Após a identificação de proteínas ubiquitinadas, a classificação por ontologia genética e a análise de enriquecimento de vias foram realizadas e os resultados podem servir de base para compreensão dos papéis fisiológicos das proteínas, além de fornecer informação de como o sistema de ubiquitinação regula o desenvolvimento de grãos de milho (FAN *et al.*, 2021). Ainda, a análise proteômica por métodos DIA em grãos de milho revelou que as alterações no proteoma causadas pelo estresse hídrico foram altamente correlacionadas com o desenvolvimento do grão e acúmulo de amido que se relaciona com o rendimento e qualidade do milho (GUO *et al.*, 2021). Adicionalmente, a comparação do proteoma de milho híbrido com linhas parentais por LC-MS/MS permitiu uma visão ampla da diversidade de proteínas, revelando maior diversidade na amostra híbrida. De fato, a classificação ontológica e a identificação de vias metabólicas são ferramentas que podem ajudar a traçar com clareza os processos realizados por sementes e plantas em sua adaptação e sobrevivência (OSAWA-MARTÍNEZ *et al.*, 2021).

Para análise de arroz, LC-MS/MS foi aplicada em folhas de duas linhagens quase-isogênicas, obtidas da Embrapa Arroz e Feijão, para comparação do perfil proteico de amostras apresentando fenótipo resistente ou susceptível à *Magnaporthe oryzae*. A determinação de DAP e a classificação por ontologia genética mostraram que os perfis proteicos foram notavelmente diferentes durante a infecção inicial, demonstrando singularidades na regulação do proteoma entre as duas linhagens (susceptível e resistente), que podem ser usadas em programas de melhoramento genético com desenvolvimento de culturas mais resistentes a patógenos (TÁVORA *et al.*, 2021). Estudos proteômicos têm sido também realizados, a partir dos grãos, para avaliar alterações proteicas durante o armazenamento (ZHAO *et al.*, 2021), ou ainda como no caso da aplicação de proteômica comparativa para avaliar o efeito da germinação e do enriquecimento com selênio na qualidade do arroz marrom (LI *et al.*, 2018a). O uso de tecnologias de irrigação no processo de enchimento do grão de arroz (LI *et al.*, 2018b) ou da luminosidade e expressão genética na resistência ao estresse (LIU *et al.*, 2020) também têm sido avaliados por técnicas proteômicas por LC-MS/MS. No campo da

qualidade tecnológica, além dos parâmetros de sabor e propriedades físico-químicas do arroz cozido (qualidade culinária), amostras de arroz de diferentes qualidades foram avaliadas por ferramentas proteômicas para estudo do efeito de altas temperaturas sob a expressão de proteínas durante o desenvolvimento da cariopse (LIN *et al.*, 2005) e acompanhamento do perfil proteico do endosperma de sementes em diferentes dias após a floração (KIM *et al.*, 2009).

A aplicação de análises proteômicas em trigo tem sido realizada amplamente para avaliar as alterações no proteoma da planta que vão desde o estudo da fertilização com nitrogênio (LANDOLFI *et al.*, 2021; LIU *et al.*, 2021) ou do estresse hídrico (MEHRI *et al.*, 2020), até a busca por contaminação de trigo em outros grãos e detecção de glúten em alimentos (COLGRAVE *et al.*, 2015; FIEDLER *et al.*, 2014). Considerando a alergenicidade do trigo e a aptidão tecnológica do grão, foi realizado em estudo proteômico com ambas as frações proteicas do trigo (prolaminas e não-prolaminas), revelando que amostras com menor força do glúten apresentaram maior expressão de peptídeos e proteínas relacionados com a doença celíaca (DC) e outras alergias, bem como maior quantidade de epítomos associados a DC do que amostras com mais alta força do glúten (ALVES *et al.*, 2018). Avaliando grãos em desenvolvimento, outro estudo com alvo no proteoma de cultivares com força de glúten contrastantes, revelou maior abundância de precursor de triticina e LMW-s GS na cultivar com maior força de glúten (GUO *et al.*, 2012).

Analisando subproteomas do trigo por LC-MS/MS, foram realizados estudos focados nas proteínas do glúten para entendimento da formação de polímeros (FERREIRA *et al.*, 2014), das modificações do perfil proteico ao longo do desenvolvimento dos grãos (LIU *et al.*, 2012), além de trabalhos voltados a alergenicidade que analisaram epítomos relacionados com a DC ou ainda quantificaram o glúten usando material de referência (SCHALK *et al.*, 2018; VAN DEN BROECK *et al.*, 2015). Em termos de aptidão tecnológica, a caracterização das proteínas do glúten revelou que a amostra de maior força do glúten apresentou maior razão glutenina:gliadina do que a amostra com força de glúten contrastante. Além disso, a razão HMW:LMW-GS foi maior em cultivares com melhores parâmetros químicos e farinográficos (AGHAGHOLIZADEH *et al.*, 2017).

As proteínas metabólicas (fração A/G) foram abordadas para o entendimento de alterações causadas pelo estresse térmico (HURKMAN *et al.*, 2009; LAINO *et al.*, 2010), de possíveis consequências funcionais dos diferentes perfis proteicos ao longo do desenvolvimento dos grãos (ARENA *et al.*, 2017; DONG *et al.*, 2012), almejando caracterizar proteínas específicas dessa fração, como ALPs (ZHANG *et al.*, 2020), para busca de alérgenos (ALVES *et al.*, 2018; LARRE *et al.*, 2011) e para analisar amostras com aptidão tecnológica contrastantes (GAO *et al.*, 2009; VICTORIO *et al.*, 2018). Apesar da relação destas proteínas com a qualidade tecnológica do trigo ser menos reportada na literatura, estes trabalhos mostram que proteínas metabólicas são diferencialmente expressas em farinhas de diferentes aptidões tecnológicas.

A seguir, são apresentados os trabalhos desenvolvidos nesta tese para contribuição no campo de pesquisa da aptidão tecnológica do trigo.

PARTE II

PRODUÇÃO CIENTÍFICA

CAPÍTULO 3. NANOUPLC-MS^E REVEALS DIFFERENTIAL ABUNDANCE OF GLUTEN PROTEINS IN WHEAT FLOURS OF DIFFERENT TECHNOLOGICAL QUALITIES

CAPÍTULO 4. UNTARGETED PROTEOMICS UNRAVELS THE PROTEIN PROFILE OF WHEAT FLOURS OF CONTRASTING TRAITS BASED ON GLUTEN-FORCE AND NEAR-ISOGENIC LINES DIFFERING BY *PINB-D1* ALLELE

CAPÍTULO 5. COMPARAÇÃO DE DOIS MÉTODOS DE ESPECTROMETRIA DE MASSAS APLICADOS PARA MELHOR CARACTERIZAÇÃO DAS PROTEÍNAS DO TRIGO: HDMS^E X UDMS^E

CAPÍTULO 3 – NANOUPLC-MS^E REVEALS DIFFERENTIAL ABUNDANCE OF GLUTEN PROTEINS IN WHEAT FLOURS OF DIFFERENT TECHNOLOGICAL QUALITIES

V.C.M. VICTORIO^a T.O. ALVES^a G.H.M. F. SOUZA^{b1} L.C.GUTKOSKI^a L.C.CAMERON^c
M.S. L. FERREIRA^{ac*}

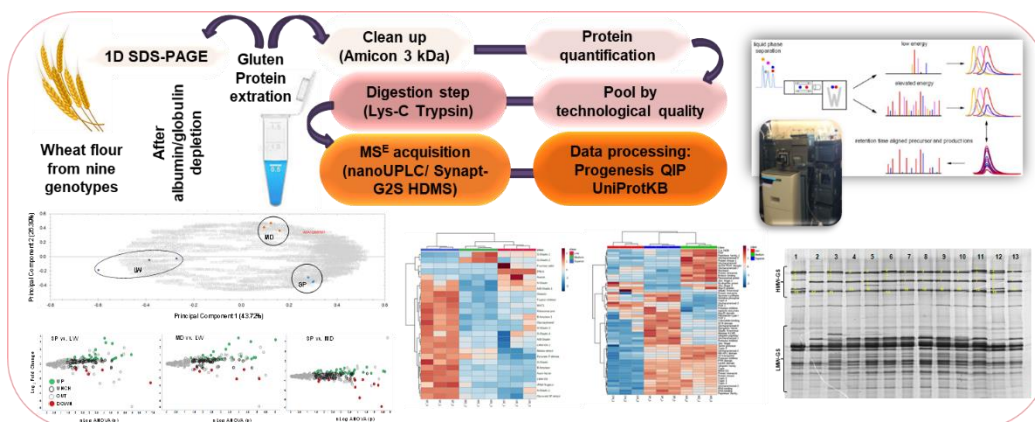
^aLaboratory of Bioactives, Food and Nutrition Graduate Program, PPGAN, Federal University of the State of Rio de Janeiro, UNIRIO, Av. Pasteur, 296, 22290-240, RJ, Brazil. ^bWaters Corporation, Waters, São Paulo, Brazil. ^cCenter of Innovation in Mass Spectrometry-Laboratory of Protein Biochemistry (IMasS-LBP), UNIRIO, Brazil. ¹Current address: Mass Spectrometry Applications & Development, SpectraMass Ltd., Campinas, São Paulo 13088-130, Brazil. *Corresponding author.

Journal of proteomics, 239 (2021). Received 23 November 2020, Revised 18 February 2021, Accepted 28 February 2021, Available online 4 March 2021. <https://doi.org/10.1016/j.jprot.2021.104181>
(FI: 4.044; Qualis único CAPES A1)

HIGHLIGHTS

Gluten proteins of wheat were characterized by a nanoUPLC-MS^E method.
Wheat flours grouped by technological quality present different protein abundance.
Functional domains cluster grouped flours of the highest technological quality.
PCA distinguished groups of different technological qualities.
PCA data are in line with protein abundance profile.

GRAPHICAL ABSTRACT



Graphical abstract 1. Summary of NanoUPLC-MS^E analysis of wheat flour.

ABSTRACT

Gluten proteins contribute to the rheological properties of dough. Mass spectrometric techniques help to understand the contribution of these proteins to the quality of the end product. This work aimed to apply modern proteomic techniques to characterize and provide a better understanding of gluten proteins in different wheat genotypes (*Triticum aestivum*). Nine Brazilian wheat flours classified by rheological gluten force were used to extract the proteins. Extracts were pooled together by technological qualities in low (LW), medium (MD), and superior (SP). Peptides were analyzed by nanoUPLC and mass spectrometry multiplex method (MS^E). Collectively, 3,545 peptides and 1,297 proteins were identified, and 116 proteins were found differentially abundant. Low molecular weight glutenin subunits (LMW-GS) were found up-regulated only in SP samples. Proteins related to wheat grain hardness, such as puroindoline-A, were found in significant concentration in LW samples. After domain prediction, LW presented a different pattern with a lower abundance of functional domains, and SP presented chaperones, known to be involved in adequate folding of the storage proteins. NanoUPLC-MS^E was efficient in analysing and distinguishing the proteomic pattern of wheat flours from different qualities, pointing out the differentially abundant gluten proteins and providing a better understanding of wheat flour quality.

Keywords: foodomics; protein abundance profile; SDS-PAGE; storage proteins; wheat quality.

Significance: Common wheat is one of the most important staple food sources in the world. The improvement and comprehension of wheat quality has been a major objective of plant breeders and cereal chemists. Our findings highlighted the application of a modern proteomic approach to obtain a better understanding of the impact of gluten proteins on the technological quality of different wheat flours. The obtained data revealed different abundances of wheat quality-related proteins in superior quality flours when compared with samples of low rheological properties. In addition, multivariate statistical analysis clearly distinguished the flours of different qualities. This work contributes to the consolidation of research in the field of wheat technological quality.

1 INTRODUCTION

Gluten is a polymeric network formed after the addition of water and mechanical force in cereals as wheat, barley and rye. In wheat, it is composed of proteins named gliadins and glutenins (WIESER, 2007). Gliadins are monomeric alcohol-soluble proteins stabilized by intramolecular disulphide bonds. Glutenins are polymeric alcohol-insoluble proteins, capable of forming additional intermolecular disulphide bonds. These polymers are formed by monomeric subunits, divided into two categories based on its molecular weight: low and high molecular weight glutenin subunits (LMW and HMW-GS) (SCHERF *et al.*, 2015; WIESER, 2007; WRIGLEY, 1996). These proteins are also called storage proteins and comprise 80 to 85% of the wheat grain's total protein content. They contribute in different ways to the rheological properties of dough. As known, gliadins confer viscosity and extensibility while glutenins improve dough strength and elasticity (WIESER, 2007; WRIGLEY, 1996). Although HMW-GS amount and type are associated with technological and bread-making quality, allowing stronger dough formation, an appropriated glutenin/gliadin ratio has an important role in the wheat's rheological properties of dough (SHEWRY; HALFORD, 2002). Studies demonstrated that wheat cultivars with similar HMW-GS content showed different rheological properties according to different glutenin/gliadin ratios (BARAK *et al.*, 2013). Highest polymeric/monomeric ratio was observed in a modern genotype of durum wheat of higher quality when compared with lower quality genotypes (DE SANTIS *et al.*, 2020).

Besides the technological aspect, gluten proteins are also involved in food allergies and intolerances, such as celiac disease (CD) (PRESUTTI *et al.*, 2007). The high proline content of these proteins may reduce its susceptibility to protease activity in the digestive process. Consequently, it leads to the fattening of intestinal villi flattening and malabsorption of nutrients (SHEWRY; TATHAM, 2016). A gluten-free diet is essential to avoid CD-related symptoms. For this reason, it is imperative to detect the presence of these proteins to ensure food safety (ALVES *et al.*, 2017; ALVES *et al.*, 2019). It has been recently shown, by proteomic approaches, that lower quality wheat flour samples had higher amounts of immunogenic peptides abundance than better quality samples (ALVES *et al.*, 2018; DE SANTIS *et al.*, 2020).

A proper extraction of protein is critical for gluten analysis. The complex structure of the gluten network and the food matrix composition are obstacles to achieve a very efficient extraction (XHAFERAJ *et al.*, 2020). Due to the amounts of conserved cysteine residues the presence of disulphide bonds, the solubilization of gluten proteins is hindered, mainly in glutenin fraction, which comprises large polymer aggregations with molecular weight from 100 kDa to over 1,000 kDa (WIESER, 2007; WRIGLEY, 1996). Gliadins and glutenins are also known for their high proline and glutamine content, as by its low content of lysine, arginine and histidine (WIESER, 2007). In addition to that, both fractions are extremely heterogeneous and complex, presenting repetitive motifs, what makes its analysis by conventional proteomics techniques challenging. The polymeric nature of gluten proteins and their heterogeneity defy obtaining highly purified extracts with just one group of proteins (LEXHALLER *et al.*, 2019). Factors such as incomplete information on protein sequences and not complete sequenced genomes limits the protein identification. The complex and highly homologous sequences of gluten proteins require careful interpretation of ion data for accurate sequence assignment (AGHAGHOLIZADEH *et al.*, 2017; MARTINEZ-ESTESO *et al.*, 2016). Elucidating important aspects of protein composition that are closely related with dough quality is one of the main objectives of wheat proteome studies (SKYLAS *et al.*, 2005).

In this context, the improvement of mass spectrometry (MS) techniques enables a better understanding about these proteins. Some studies have shown that data independent acquisition (DIA) MS enhances complex samples analysis and allows wide dynamic range coverage (DISTLER *et al.*, 2014; GEROMANOS *et al.*, 2009; SOUZA *et al.*, 2017; VICTORIO *et al.*, 2018). DIA approaches alternate low and high energy in the collision cell to analyze all ions generated in the source, and then precursors and fragments reach the analyzer quasi-simultaneously. Results of these methods are more comprehensive due the fragmentation, regardless from the signal intensity (BROMILOW *et al.*, 2017b; DISTLER *et al.*, 2014; SOUZA *et al.*, 2017). By using proteomic approaches, is possible to associate the protein composition of wheat with its technological quality. Metabolic proteins, such as albumins and globulins, were analyzed by ultra-definition mass spectrometry (UDMS^E) and were differentially abundant in wheat flours of different technological qualities. Proteins of this fraction involved in the gluten protein folding

process, named puroindolines and chaperones, were found up-regulated in superior quality extract and down-regulated in the lower ones (VICTORIO *et al.*, 2018). Gluten fraction study also showed differences between cultivars. For the glutenin fraction, Dy10 subunit was not found in the low-quality cultivar and had high percentage in higher quality cultivars. The content of x-HMW-GS and y-HMW-GS seems to be correlated with wheat quality. When analyzing flat bread, x-HMW-GS was determinant for its quality. It was demonstrated that higher amounts of γ -gliadins decrease the quality of wheat, what is inversely correlated to higher contents of α/β -gliadin. By comparing glutenin/gliadin ratio, superior quality flour presented the greatest values, while inferior quality flour presented the lowest one (AGHAGHOLIZADEH *et al.*, 2017). For that matter, advanced proteomics methods lead to a high capacity of protein detection, being important for wheat research (AFZAL *et al.*, 2020).

In this context, this work presents a nano ultra-performance liquid chromatography (nanoUPLC) coupled to mass spectrometry analysis using DIA method in order to identify and quantify gluten proteins from nine wheat genotypes grouped by different technological qualities. The method was effective to analyze wheat storage proteins and allowed a clear distinction between genotypes, providing a better understanding of the wheat flour quality.

2 MATERIALS AND METHODS

2.1 PLANT MATERIAL

Wheat (*Triticum aestivum*) flours from nine Brazilian genotypes of different technological qualities were analyzed. Samples were provided by OR Melhoramento de Sementes Ltda. (Passo Fundo, RS, Brazil) in collaboration with the University of Passo Fundo (UPF). The seeds were cropped during 2014/2015, with normal harvesting and management of seeds. The technological qualities classification was based on gluten force (PELEGRIN *et al.*, 2016), previously determined by alveograph and farinograph methods, using tenacity, extensibility and stability of dough, mass development time and water absorption as parameters, respectively (Table 1). The flours were grouped as low (LW) quality (Campeiro, ORS25, ORS27); medium (MD) quality (Marfim, ORS1401, ORS1402), and superior (SP) quality (Ametista, Guabiju, Jadeíte 11).

Table 1. Technological quality characterization of different Brazilian wheat genotypes according to rheological and physico-chemical parameters, harvest 2014/2015.

Genotype	Commercial Classification*	TQ	W (Joules)	E (min)	FN (s)	P (mm)	L (mm)	P/L	WA (%)
ORS 25	Basic	Low	137.7 ± 41.2	7.3 ± 3.1	334.0 ± 18.4	49.3 ± 10.5	92.3 ± 17.2	0.5 ± 0.0	53.6 ± 1.7
Campeiro	Basic	Low	115.3 ± 12.5	6.7 ± 1.4	286.3 ± 14.1	32.0 ± 2.6	157.0 ± 15.5	0.2 ± 0.0	49.2 ± 0.6
ORS 27	Domestic	Low	161.5 ± 23.5	8.8 ± 0.8	309.0 ± 46.0	58.5 ± 1.5	103.0 ± 14.0	0.6 ± 0.1	57.5 ± 0.5
ORS 1401	Bread	Medium	258.3 ± 65.5	19.3 ± 9.0	356.8 ± 47.1	91.9 ± 23.4	83.9 ± 3.7	1.1 ± 0.4	59.7 ± 6.2
ORS 1402	Basic	Medium	206.7 ± 56.6	11.4 ± 2.2	348.0 ± 42.0	70.3 ± 17.8	102.7 ± 15.0	0.7 ± 0.2	58.7 ± 6.5
Marfim	Bread	Medium	310.7 ± 56.3	17.7 ± 4.7	348.8 ± 24.8	71.0 ± 7.5	149.7 ± 9.5	0.5 ± 0.0	55.8 ± 0.7
Jadeíte 11	Improver	Superior	418.3 ± 10.7	14.9 ± 1.9	371.3 ± 25.3	114.7 ± 5.5	91.9 ± 21.3	1.3 ± 0.4	61.5 ± 0.8
Ametista	Improver	Superior	437.3 ± 33.5	20.9 ± 5.3	347.3 ± 48.4	134.3 ± 13.1	80.8 ± 18.6	1.7 ± 0.4	62.2 ± 0.2
Guabiju	Bread	Superior	255.3 ± 15.6	13.3 ± 1.3	376.7 ± 45.1	138.0 ± 6.0	101.3 ± 6.5	1.4 ± 0.0	63.8 ± 0.8

Technological quality (TQ), gluten force (W), stability (E), falling number (FN), tenacity (P), extensibility (L), water absorption (WA). *Commercial Classification according to Brasil (2010).

2.2 PROTEINS EXTRACTION

The protein extraction was performed as described in Alves et al. (2018). One hundred milligrams of each flour were extracted with 1 mL of buffer (80 mM Tris-HCl, 40 mM iodoacetamide, pH 8.0). The samples were briefly vortexed and then incubated in shaker (200 rpm, 60 min, 25 °C) (TE-420, Tecnal, Brazil). After centrifugation (10,600 x g, 10 min, 20 °C) (Heraeus Megafuge 16R Centrifuge, Thermo Fisher Scientific, Germany), the supernatants containing albumins and globulins were discarded. The remaining pellets were resuspended in PBS buffer (0.5% SDS, 2% β-mercaptoethanol, pH 7.4) and incubated in shaker overnight (200 rpm, 25 °C). After centrifugation (10,600 x g, 15 min, 4 °C), the supernatants were washed and concentrated using a 3 kDa buffer exchange filter (Amicon, Millipore, Ireland). Three washing steps were performed using ammonium bicarbonate (NH₄HCO₃) buffer (50 mM, pH 8.5), followed by centrifugation (14,000 x g, 60 min, 8 °C). Protein content was determined by Bradford method (1976) using bovine serum albumin (BSA) as standard. Samples were diluted 10 times and 20

μL was aliquoted in triplicate. The reaction time was 5 min after the addition of 1 mL of Bradford reagent (Sigma-Aldrich). The absorbance was measured at 595 nm in a spectrophotometer (UV-2700, Shimadzu, Japan).

2.3 SAMPLE PREPARATION FOR NANOUPLC-MS^E ANALYSIS

Prior to the digestion step, protein extracts were diluted in ammonium bicarbonate (50 mM, pH 8.5) to reach the concentration of $1 \mu\text{g} \cdot \mu\text{L}^{-1}$. The extracts were pooled together according to the technological quality classification, grouping 20 μL of each sample. Then, three pools with 60 μL of final volume followed to the digestion step. For protein digestion, 25 μL of 0.2% RapiGEST SF (Waters Corp., Milford, MA) were added in each sample containing 60 μg of protein, followed by vortexing and incubation at 80 °C for 15 min. After centrifugation (14,000 x g, 10 min, 4 °C), 2.5 μL of 100 mM dithiothreitol (DTT) (Sigma Aldrich, St. Louis, USA) prepared in NH_4HCO_3 (50 mM, pH 8.5) were added to each sample. The samples were vortexed and incubated at 60 °C for 30 min followed by centrifugation (14,000 x g, 10 min, 4 °C). Then, 2.5 μL of 300 mM iodoacetamide (IAM) (Sigma Aldrich, St. Louis, USA) in NH_4HCO_3 (50 mM, pH 8.5) were added and samples were homogenized and incubated in the dark at 20 °C for 30 min. Afterwards, trypsin/Lys-C (Promega, Madison, USA) in NH_4HCO_3 (50 mM, pH 8.5) was added at a 1:60 enzyme:protein ratio. The samples were briefly vortexed and centrifugated for 30 s and digested overnight at 37 °C under stirring. After digestion, 10 μL of 5% v/v trifluoroacetic acid (TFA) (Sigma Aldrich, St. Louis, USA) were added to stop the digestion. The samples were incubated at 37 °C for 90 min without stirring, centrifuged (14,000 x g, 90 min, 4 °C), the supernatants were transferred to new vials and centrifuged again (14,000 x g, 90 min, 4 °C). The supernatants were transferred to total recovery vials (Waters, USA), identified and placed in ultra-freezer (-80 °C) (IULT 335 D, Indrel, Brazil) until data acquisition.

2.4 NANOUPLC-MS^E ANALYSIS

Tryptic peptides were analyzed using a nanoACQUITY UPLC system (Waters Corp., Milford, MA) coupled to Synapt G2-S high-definition mass spectrometer (HDMS) (Waters Corp., Manchester, UK) equipped with hybrid analyzers such as a quadrupole followed by an orthogonal acceleration time-of-flight (Qq-*oa*TOF) geometry. To ensure that all

samples were injected with the same amount on column, prior to analysis, stoichiometric measurements based on scouting runs of the integrated total ion account (TIC) were performed to ensure standardized molar values across all conditions. The LC/MS runs were carried out in triplicate for each sample. 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 water and mobile phase B consisted of acetonitrile, both with 0.1% (v/v) formic acid. The peptides were separated using chromatographic methods with the following gradient: 7 to 40% of mobile phase B in 90 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 $\eta\text{L}\cdot\text{min}^{-1}$. The analytical column temperature was maintained at 55 $^{\circ}\text{C}$ and the sample manager temperature was 8 $^{\circ}\text{C}$. The lock mass was derived from an auxiliary pump using a flow rate of 250 $\eta\text{L}\cdot\text{min}^{-1}$ at a concentration of 200 $\text{fmol}\cdot\mu\text{L}^{-1}$ of [Glu¹]-Fibrinopeptide B human (Glu-Fib) to the reference MUX sprayer of the NanoLockSpray source.

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. 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 channel was sampled at a frequency of 30 s. The ion source block temperature was set to 70 $^{\circ}\text{C}$ and capillary voltage was set to 2.7 kV. Ions were acquired with m/z between 50-2000, scanning time of 0.5 s, cone voltage of 30 V. The time-of-flight analyzer of the mass spectrometer was calibrated with a MS/MS spectrum of Glu-Fib. The final instrument calibration was obtained by the specific y^+ and b^+ fragments derived from the double charged precursor ion $[\text{M} + 2\text{H}]^{2+} = 785.8426$ GFP signal. Samples were acquired by MS^E multiplex mode, applying simultaneously low and high fragmentation energy (15-55V) at T-wave collision-induced dissociation cell filled with argon gas.

2.5 DATA PROCESSING

In order to identify and quantify peptides and proteins, MS quantitative data packages were processed with the use of dedicated algorithms (GEROMANOS *et al.*,

2009). The MS data were processed and searched using the Progenesis Q1 for Proteomics (QIP) (Nonlinear Dynamics, UK). The UNIPROT protein repository release 2017_02 (<http://www.uniprot.org>, 136,884 entries) with database annotations for *T. aestivum* was used. The database used were reversed “on-the fly” during the database queries by the use of ProteinLynxGlobalServer (PLGS) (Waters Corp., Manchester, UK) and appended to the original database to assess the false discovery rate (FDR) during identification (NEVES *et al.*, 2017). The parameters for database searching were tryptic peptides with a maximum of one missed cleavage, carbamidomethylation of cysteine as fixed modification and oxidation of methionine as variable modification. The parameters set as default were peptide mass error tolerance of 10 ppm, fragment mass error tolerance of 20 ppm and maximum FDR of 4%.

The samples were compared in pairs to obtain a volcano plot (SP vs. LW, MD vs. LW and SP vs. MD) and the following filters were used: minimum of one unique peptide and two peptide count, coefficient of variation (maximum absolute CV of 0.3), max fold change (>1.5, $\log_2 = \pm 0.6$) and ANOVA ($p < 0.05$). Only coexisting proteins in both conditions and proteins that were present in the 3 of 3 replicates (3/3) were considered to compare the conditions.

Functional domains were predicted for uncharacterized proteins that were differentially abundant using the following databases InterPro 67.0 (1st March 2018, EMBL-EBI, <http://www.ebi.ac.uk/interpro/>) (FINN *et al.*, 2017) Pfam 31.0 (March 2017, 16712 entries, EMBL-EBI, <http://pfam.xfam.org/search#tabview=tab0>) (FINN *et al.*, 2016) and Prosite (Release 2018_03 of 28-Mar-2018 contains 1806 documentation entries, 1309 patterns, 1214 profiles and 1234 ProRule, <https://prosite.expasy.org/>) (SIGRIST *et al.*, 2010). The searches were made with default parameters. The predominant domain for each protein code was chosen.

2.6 1D SODIUM DOCEDYL SULFATE POLYACRYLAMIDE GEL ELECTROPHORESIS (1D SDS-PAGE)

The electrophoresis gel was performed as described by Morel (1994). Laemmli buffer system (1970) was used with 10.3% acrylamide gels. The gel was run at 40 mA for 4 h 30 min at 18 °C. The samples were extracted using a solution containing 2% SDS and

40% glycerol and 20 μ L of each sample was loaded in the wells. The gel was incubated overnight in 15% trichloroacetic acid, rinsed for 5 min with tap water and stained in 12.5% trichloroacetic acid with 0.14% (w/v) Coomassie Brilliant Blue R250.

2.7 STATISTICAL ANALYSIS

ANOVA was given by the Progenesis QIP and the coefficient of variation (CV) was calculated for replicates of each variable. A heat map representing protein abundance was carried out using MetaboAnalyst 4.0 (Xia Lab, <http://www.metaboanalyst.ca/>) for annotated proteins as well as for uncharacterized proteins after domain prediction. Principal component analysis (PCA) was performed with Progenesis QIP and implemented in order to assess the distinction between groups of different technological quality.

3 RESULTS AND DISCUSSION

3.1 NANOUPLC-MS^E RESULTS

Collectively, 3,545 peptides and 1,297 proteins were identified with average of 3 peptides per protein. After applying the minimum of two peptides at identification level and maximum absolute CV of 0.3 at quantitation level as filters, on average for all replicates and conditions, 3,070 peptides and 743 proteins were obtained with FDR of 0.6 % and 1.1 %, respectively (see supplementary data, S1). In order to verify if samples could be compared, data were plotted in a dynamic range graph (Fig.1). Dynamic ranges of different samples presented the same trend indicating that the quantitative distribution of intracellular protein was similar between them. In general, samples of complex biological systems present dynamic ranges of 3 orders of magnitude ($\log_{10}\max = 3$) (SOUZA *et al.*, 2017). In this study, dynamic ranges of detected proteins presented more than 6 orders of magnitude ($\log_{10} > 6$) showing this proteomic approach was efficient to identify proteins in complex samples. The results are according to previous studies applying nanoUPLC-UDMS^E to analyze wheat metabolic proteins, in which more than 5 orders of magnitude were found (VICTORIO *et al.*, 2018).

Trying to overcome the challenge of fractioning the gluten proteins, the enzymes trypsin plus Lys-C were used, and the data processed by the Progenesis QIP showed that more than 70% of the identifications were made without missed cleavages and more than 20% with only one missed cleavage (see supplementary data, S2).

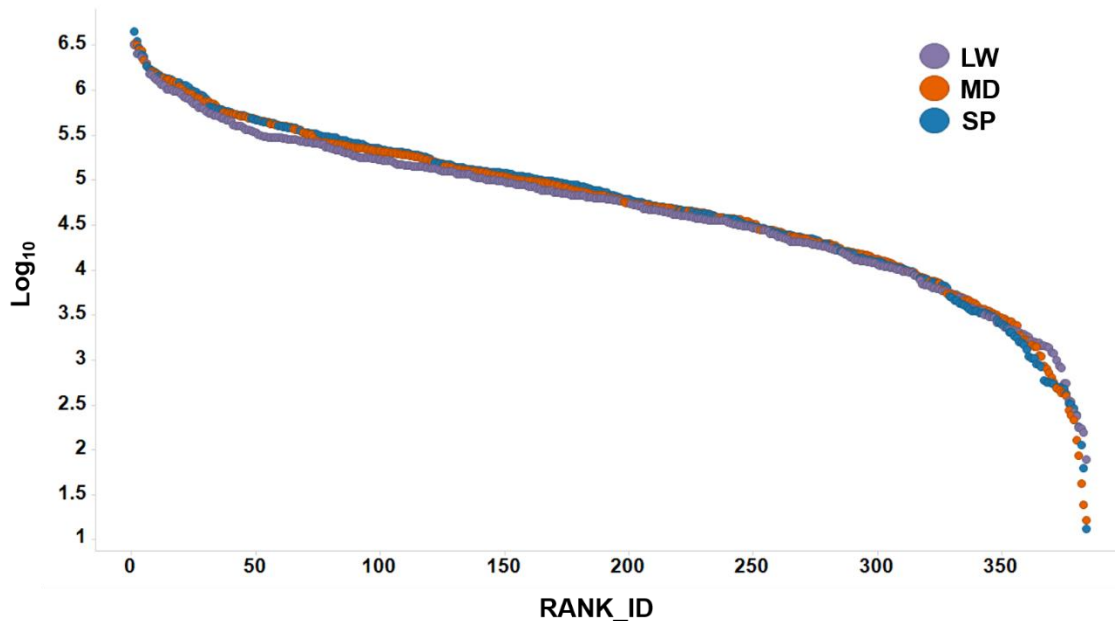


Fig. 1. Dynamic range of wheat flour samples, low (LW), medium (MD) and superior (SP) quality classification. The rank of identified proteins is plotted in the x axis and y axis is represented by the log of abundance quantified proteins.

Considering only proteins that appear in 3 of 3 replicates of each sample and without other filters, 384 proteins were identified and quantified in this work (see supplementary data, S1). Most of them (~78%) are not yet described in the protein databank, according to their biological function. Factoring only the annotated proteins, 46 presented highest mean abundance on SP sample, highlighting HMW-GS Dx5 (UNIPROT accession: P10388), HMW-GS (PC256 Fragment and P02861), alpha/beta-gliadin (P04724 and P04727), LMW-GS (B2Y2Q6, B2Y2S6, Q8W3V4, Q6SPZ3, Q00M61, B2BZD1, B2Y2Q1, P10385 and Q00M56), gamma-gliadin B (P06659), protein disulphide-isomerase (PDI) (A0A1D5Y681), gamma-gliadin (A1EHE7), alpha-gliadin (A0A0E3Z7G8 fragment), gamma-gliadin B-I (P04729 and M9TG60), and alpha-gliadin (I0IT65). For MD sample, 25 proteins presented highest mean abundance and can be found gamma-gliadin (P21292 and Q94G92), PDI (A0A1D5XQQ1), puroindoline-B (B1GXL6) and LMW-GS 1D1 (P10386). On LW sample, were found 15 proteins, being the main HMW-GS DY10

(P10387), alpha-gliadin (J7HT09), alpha/beta-gliadin (P04726), gamma-gliadin (M9TGF7), puroindoline-B (Q10464) and puroindoline-A (P33432). PDIs and puroindolines are metabolic proteins of wheat. While the first ones are related with storage protein folding (KIMURA *et al.*, 2015), the others promote a positive impact on baking (OSIPOVA *et al.*, 2012). In agreement with our data, PDIs were found up regulated in higher quality wheat samples when compared with lower quality samples in a study performed with metabolic proteins. In this study, puroindoline-A also was found up regulated in LW samples (VICTORIO *et al.*, 2018). It is worth mentioning that some PDI isoforms may no longer be detectable in the endosperm when the protein synthesis decreases at grain maturity (SKYLAS *et al.*, 2005).

3.1.1 Protein profile cluster of nanoUPLC-MS^E data

For better understanding the high-performance data provided by LC-MS, a principal component analysis (PCA) was performed (Fig. 2). Considering identified and quantified proteins as parameters it is possible to realize the distinction between groups of different technological qualities. The first two principal components (PC) explain 70% of the variation between samples. Also, it is possible to observe the separation of LW sample from the MD and SP samples. In a work carried out with metabolic gluten proteins the same pattern could be observed in the PCA (VICTORIO *et al.*, 2018). A marked separation among samples can be observed along PC1, corroborating another results with different quality samples (DE SANTIS *et al.*, 2020). Protein abundance profiles and domain prediction data probably contribute for explain the samples discrimination found in PCA.

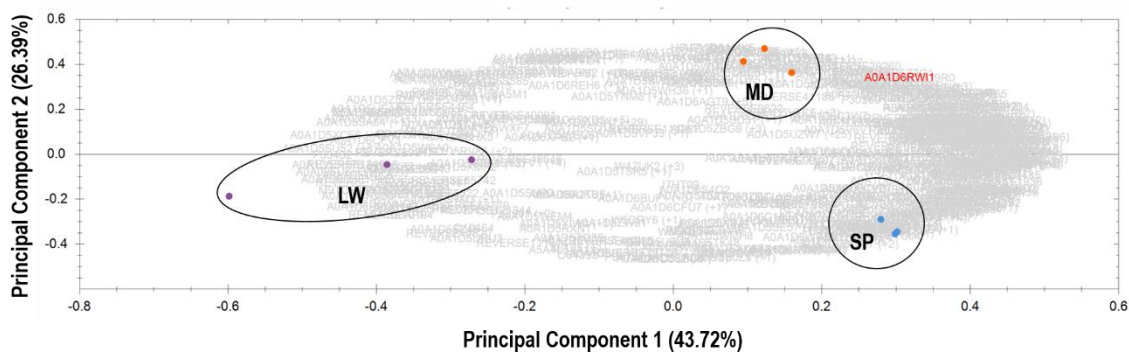


Fig. 2. Principal component analysis (PCA) of wheat flour samples: low (LW), medium (MD) and superior (SP) quality classification. Identified and annotated proteins are in gray and red.

3.2 PROTEIN ABUNDANCE PROFILES OF SAMPLES COMPARED IN PAIRS

After applying the filters mentioned in 2.5 section, data were presented in volcano plots (Fig. 3) to obtain an overview of differentially abundant proteins (DAP) by comparing each two samples. In total, 116 DAP were found when compared SP x LW, MD x LW and SP x MD. The comparison between SP and MD samples presented the highest number of DAP (46), being 34 up-regulated and 12 down-regulated. For these samples, 102 proteins had unchanged abundance and 242 proteins were out of one or more filters. When comparing SP and LW samples, 36 proteins were found differentially abundant (31 up-regulated and 5 down-regulated), 72 unchanged and 280 proteins were out of filters. Finally, the comparison between MD and LW samples showed 34 DAP (28 up-regulated and 6 down-regulated), 63 unchanged and 291 out of filters.

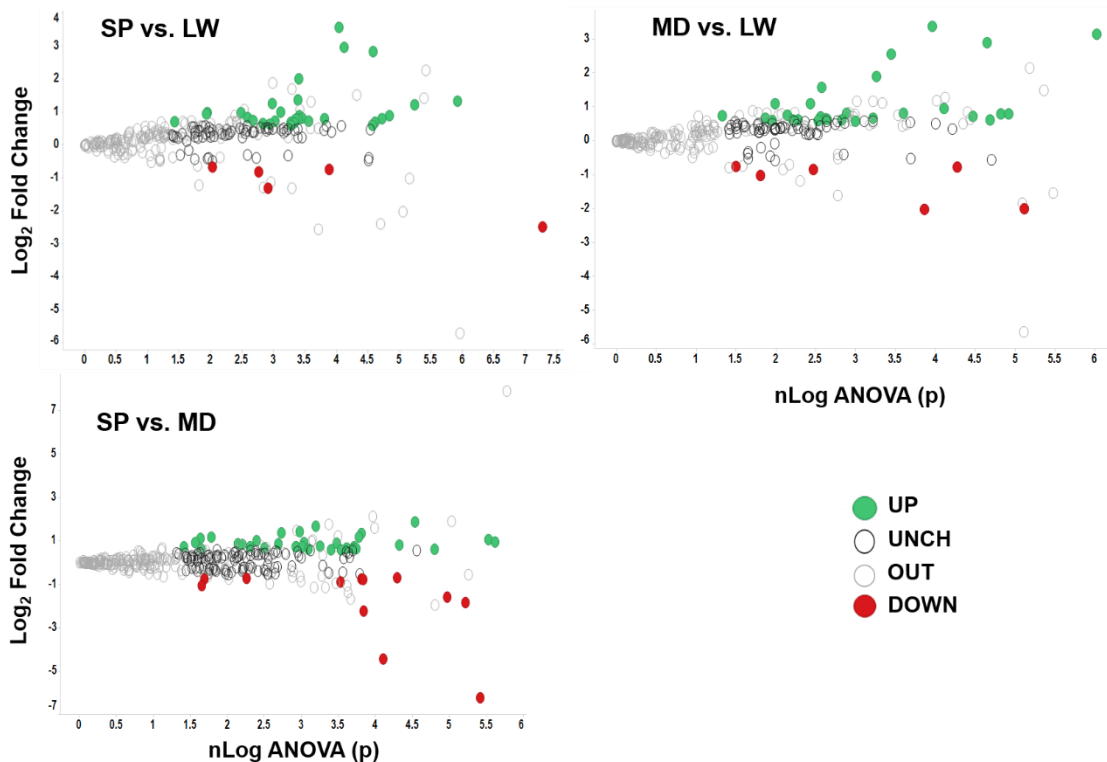


Fig. 3. Volcano plots comparing the wheat flour samples in pairs, low (LW), medium (MD) and superior (SP) ($p < 0.05$). The green dots represent up-regulated proteins (max fold change > 1.5), while red dots represent down-regulated proteins (max fold change < -1.5), black empty dots the considered unchanged proteins and gray empty dots were out of filters (minimum one unique peptide, two peptide count, CV of 0.3, max fold change < 1.5 , $P < 0.05$ and/or proteins in 3/3 replicates).

Only 35% of DAP have been described according to their biological function in protein databank (Table 2). Interestingly, low molecular weight glutenin subunits were up-regulated only on SP flours. According to some studies, that points to the importance of LMW-GS fraction to wheat quality (D'OVIDIO; MASCI, 2004; LIU *et al.*, 2005). Other proteins of gluten fraction as α/β -gliadins and γ -gliadins were found in higher concentration in SP samples. Some wheat proteins that do not belong to gluten fraction also were found in the extracts. Puroindoline-A was found in major concentration in LW sample. Puroindoline family is related with wheat grain hardness (HEINZE *et al.*, 2016). Two β -amylases and one trypsin inhibitor appear in major concentration in SP sample. A heat map cluster was built for these proteins (see supplementary data, S3) showing that SP sample presents different pattern from MD and LW samples, being the last ones grouped together by the dendrogram. In a previous study, puroindolines were found in gliadin extract, β -amylases were found in LMW-GS extract and α -amylase/trypsin inhibitor were found in all gluten protein extracts. LMW-GS were also detected in all extracts probably due to their polymeric nature and their molecular weight similar to α -gliadins. This data confirms that obtaining pure extracts is a great challenge (LEXHALLER *et al.*, 2019).

Table 2. Differentially abundant and annotated proteins.

Accession UNIPROT	Description UNIPROT	PC	PU	ANOVA (p)	-LogP (ANOVA)	Max fold change	Log ₂
<i>Superior vs. Low</i>							
A0A1D6B080	Auxin response factor	2	2	0.00	4.12	7.81	2.96
A1EHE7	Gamma-gliadin	4	2	0.00	2.98	2.39	1.26
A0A1D5ST37	Malate dehydrogenase	5	5	0.00	2.48	1.97	0.98
A0A1D5ZDV7	rRNA N-glycosidase	4	1	0.01	1.94	1.92	0.94
A0A1D5YFA6	Beta-amylase	1	3	0.00	3.81	1.73	0.79
		4					
B2Y2Q6	Low molecular weight glutenin subunit	1	1	0.00	4.72	1.73	0.79
		0					
A0A1D5S7P5	Pyruvate_ phosphate dikinase	6	4	0.00	3.39	1.72	0.78
A0A1D6RTR9	WAT1-related protein	3	1	0.00	3.02	1.65	0.72
Q0Q5D9	Globulin 1	1	2	0.00	3.33	1.60	0.68
		1					
A0A1D6BQM7	60S acidic ribosomal protein P0	5	5	0.00	3.27	1.51	0.59
Q94G92	Gamma-gliadin	3	2	0.00	3.88	1.68	-0.75
P33432	Puroindoline-A	4	4	0.00	7.27	5.64	-2.49
<i>Medium vs. Low</i>							
A0A1D6B080	Auxin response factor	2	2	0.00	2.43	2.14	1.10
A0A1D5ST37	Malate dehydrogenase	5	5	0.01	1.87	1.58	0.66
P21292	Gamma-gliadin	2	1	0.01	2.21	1.53	0.62
A0A1D5S7P5	Pyruvate_ phosphate dikinase	6	4	0.00	2.53	1.51	0.60
P0CZ05	Avenin-like b2	4	2	0.03	1.50	1.69	-0.76
I0IT65	Alpha-gliadin	2	2	0.00	2.47	1.80	-0.85
P33432	Puroindoline-A	4	4	0.00	5.11	4.02	-2.01
<i>Superior vs. Medium</i>							
A0A1D6B080	Auxin response factor	2	2	0.00	4.54	3.65	1.87
A0A1D6BR39	Glycosyltransferase	2	2	0.00	2.97	2.68	1.42
A0A1D6RTR9	WAT1-related protein	3	1	0.00	3.77	2.26	1.18
A0A0E3Z7G8	Alpha-gliadin (Fragment)	2	2	0.00	5.62	1.94	0.95
A0A1D5ZDV7	rRNA N-glycosidase	4	1	0.03	1.57	1.91	0.94
Q0Q5D9	Globulin 1	1	2	0.00	3.03	1.89	0.92
		1					
A0A1D5XXC7	Beta-amylase	1	2	0.00	3.47	1.88	0.91
		4					
P06659	Gamma-gliadin B	5	4	0.00	2.69	1.84	0.88
I0IT65	Alpha-gliadin	2	2	0.01	2.14	1.82	0.86

Accession UNIPROT	Description UNIPROT	PC	PU	ANOVA (p)	-LogP (ANOVA)	Max fold change	Log ₂
A1EHE7	Gamma-gliadin	4	2	0.01	2.20	1.79	0.84
A0A1D5YFA6	Beta-amylase	1 4	3	0.00	3.25	1.68	0.75
A0A1D6BQM7	60S acidic ribosomal protein P0	5	5	0.00	3.71	1.65	0.72
A0A1D6BRM9	Glyceraldehyde-3-phosphate dehydrogenase	5	2	0.00	3.01	1.65	0.72
P04727	Alpha/beta-gliadin	2	1	0.00	3.61	1.61	0.69
A0A1D5WNG5	Fructose-bisphosphate aldolase	7	1	0.00	2.50	1.60	0.68
P04724	Alpha/beta-gliadin A-IV	3	3	0.00	3.70	1.57	0.65
P81713	Bowman-Birk type trypsin inhibitor	4	4	0.00	3.09	1.54	0.62
M9TG60	Gamma-gliadin 1	3	2	0.02	1.64	1.54	0.62
B2Y2Q1	Low molecular weight glutenin subunit	9	1	0.00	3.53	1.52	0.60
P21292	Gamma-gliadin	2	1	0.00	3.83	1.70	-0.77
Q94G92	Gamma-gliadin	3	2	0.02	1.65	2.08	-1.05

Peptide count (PC), number of unique peptides (PU), negative log₁₀ ANOVA (-LogP (ANOVA)), log₂ max fold change (Log₂). Log₂ values equal or greater than 0.6 correspond to up-regulated proteins and equal or lower than -0.6 correspond to down-regulated proteins.

3.2.1 Cluster of functional domains of differentially abundant and not described proteins

A domain prediction analysis was carried out to obtain a better understanding about the functionality of those 65% of DAP, which remain not described in terms of biological function (see supplementary data, S4). A heat map cluster was built using the obtained data (Fig. 4) and it is possible to observe that LW sample presents different pattern from the others, with lower abundance of functional domains. MD and SP samples were grouped together by the dendrogram, although data show different profiles between samples. Interestingly, a heat shock protein HSP20 domain appears in higher abundance in SP sample. HSP belong to chaperones family that play a role to adequate folding of storage proteins and it can affect wheat flour quality by impacting the polymers formation (KIMURA *et al.*, 2015; OSIPOVA *et al.*, 2012). Another member of this family (HSP70) has already been reported to be up-regulated in SP and MD samples of wheat flour (VICTORIO *et al.*, 2018). SP sample also showed higher abundance of proteins involved

in carbohydrate metabolism (sucrose synthase), chromatin organization (histone H1/H5) and chromosome remodeling (RNA-binding). Sucrose synthase is important for sucrose metabolism and their increased abundance may impair the starch biosynthesis by decreasing sucrose level (MEHRI *et al.*, 2020). While histones H1 act as linker proteins that can confer a more condensed chromatin structure, RNA-binding proteins are involved with rRNA- and mRNA- processing in the nucleus, also acting in the cytoplasm in RNA transport, localization, storage and metabolism (FINNIE *et al.*, 2011).

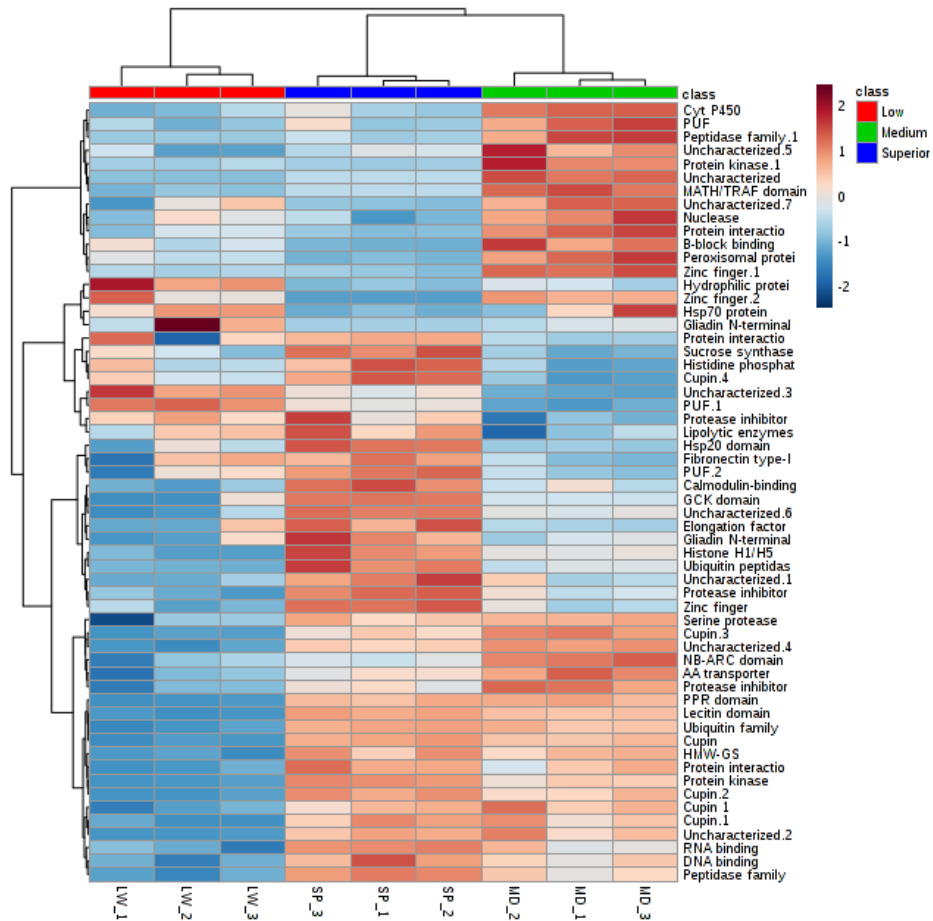


Fig. 4. Heat map illustrating the clustering of functional domains from uncharacterized differentially abundant proteins that were found in low (LW), medium (MD) and superior (SP) samples. The horizontal axis presents replicates from each sample and the vertical axis presents the domains. Color gradient represents the variation of protein abundance. Highest quantities are represented by dark red and lowest quantities by dark blue. PUF (protein of unknown function), GCK domain is related with intracellular signaling pathways.

3.3 GLUTENIN SUBUNITS CHARACTERIZATION BY ELECTROFORESIS

Protein profile of gluten proteins from wheat genotypes are presented in 1D SDS-PAGE (Fig. 5). The HMW-GS are shown on top of the gel and the LMW-GS at the bottom. Two genotypes were used as pattern (Bermude and Toback). For HMW-GS, we numbered the bands based on Payne and Lawrence (1983), that previously identified HMW-GS by SDS-PAGE. Bermude presents five bands (2*, 5, 6, 8 and 10) and Toback presents three bands (2, 7 and 12). In agreement with other paper (PAYNE, 1987) the analyzed genotypes showed two subunits coded by genes at *Glu-D1* (5+10 or 2+12 alleles), one or two by *Glu-B1* (7 allele) and one or none by *Glu-A1* (1 or 2* alleles). Most of the genotypes analyzed on this study present subunit 5 as slowest band and 10 as fastest band (ORS 25, ORS 27, ORS 1401, Marfim, Ametista and Jadeíte 11). The 5+10 subunits have been related to good bread-making quality (COSTA *et al.*, 2017; LIU *et al.*, 2009; PAYNE *et al.*, 1981). Although, in this work, these data do not allow a distinction from genotypes since these subunits were found in samples of all technological qualities. Two genotypes (Campeiro – LW; and ORS 1402 – MD) presented the subunit 2* like Bermude pattern and this band migrated slowest than the 5 subunit. Interestingly, one genotype of superior technological quality (Guabiju) presented the 2 and 12 subunits as the Toback pattern, however, this genotype of superior quality presented the subunit 1 slowest than 2. Although 2+12 have been correlated with smaller sedimentation volume when compared with 5+10 subunits, the subunit 1 was reported as presenting strong correlation with bread-making quality (LIU *et al.*, 2005; PAYNE *et al.*, 1981). HMW-GS 6+10 alleles, reported in best gluten technological quality samples (GRAZIANO *et al.*, 2019), were found only on Bermude pattern and HMW-GS 20 allele was not found. This allele has been associated with negative dough properties when *Triticum durum* was analyzed, so this result could be expected (GRAZIANO *et al.*, 2019; SHEWRY *et al.*, 2003).

For LMW-GS, most of analyzed genotypes were observed in electrophoresis gel like Toback pattern. Only two genotypes (Campeiro and ORS1401) showed changes in some bands. These samples are different in terms of technological quality, belonging to low and medium quality, respectively. In SDS-PAGE the allele identification is difficult because of the banding pattern complexity of LMW-GS that present overlapped bands. For a better

understanding of this fraction, two-dimensional gel electrophoresis (2-DGE) could be applied (MOREL, 1994). This method allows to separate proteins in two independent physiochemical parameters, considering their isoelectric point and molecular mass (SKYLAS *et al.*, 2005), providing provide a better LMW-GS separation. When using 2-DGE to analyze a common wheat genotype 17 spots were resolved from the 43 identified LMW-GS genes. Due LMW-GS gene family has high similarity of sequences with their repeat sequence, each spot contained a large number of proteins (BEOM *et al.*, 2018).

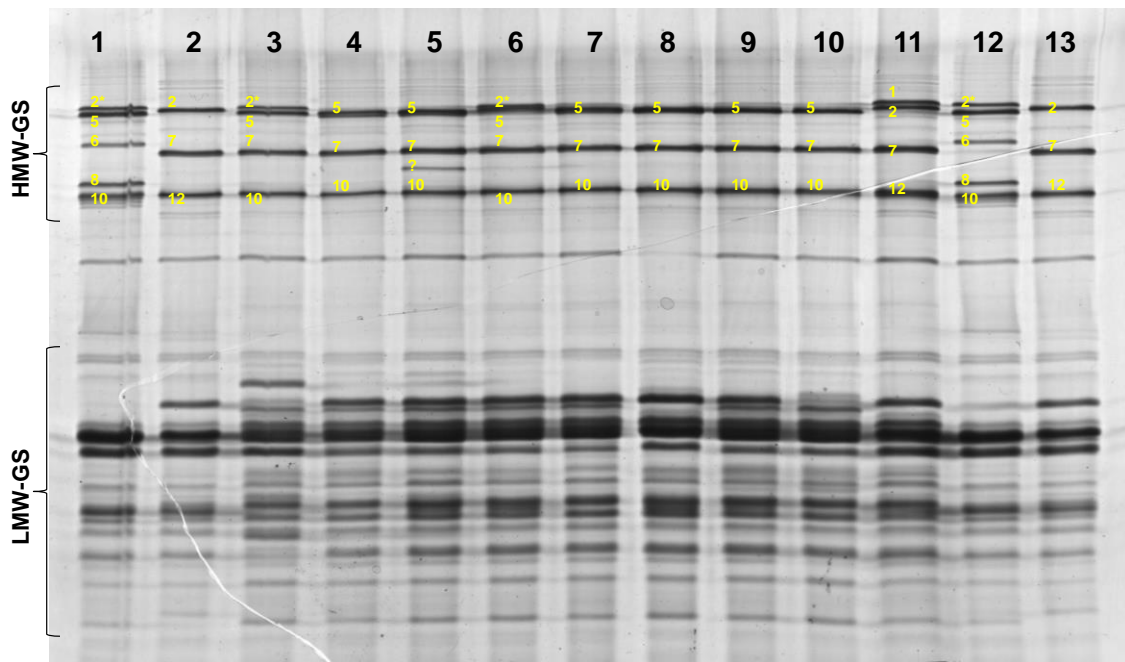


Fig. 5. Sodium dodecyl sulfate polyacrylamide gel electrophoresis patterns of high molecular weight glutenin subunits (HMW-GS) and low molecular weight glutenin subunits (LMW-GS) from wheat genotypes. The numbers on the top of the figure represent the genotypes. The patterns are Bermude (1 and 12) and Toback (2 and 13). The analyzed genotypes are Campeiro (3), ORS 27 (5) and ORS 25 (7) of low quality, Marfim (4), ORS 1402 (6) and ORS 1401 (10) of medium quality and Jadeíte 11 (8), Ametista (9) and Guabiju (11) of superior quality. HMW-GS are labeled according to the nomenclature of Payne & Lawrence (1983). Unknown bands are marked by an interrogation point.

4 CONCLUSIONS

A proteomic approach was applied to gluten proteins grouped by technological quality to better understand their relationship with wheat flour quality. A wide protein coverage was obtained and both abundance data and domain prediction demonstrated

that low quality have different pattern when compared with higher quality flours. LMW-GS was found up-regulated in superior quality sample and gliadins were found in all samples. Although electrophoresis characterization of glutenin subunits did not allow a clear distinction from genotypes, the extracts presented different protein patterns, as demonstrated by the multivariate PCA. This work showed that nanoUPLC-MS^E was effective to characterized wheat storage proteins and allowed a clear distinction between genotypes of different technological quality, providing a better understanding of the wheat flour quality.

Declarations of interest: none

Author statement

Verônica Victorio: Conceptualization, Methodology, Formal analysis, Investigation. Data Curation, Writing - Original Draft, Visualization. **Thais Alves:** Investigation, Writing-review & Editing. **Gustavo Souza:** Methodology, Software, Data Curation. **Luiz Carlos Gutkoski:** Investigation, Resources. **Luiz Claudio Cameron:** Resources. Writing - review & Editing. **Mariana Ferreira:** Conceptualization, Resources, Writing - review & Editing, Supervision, Project administration, Funding acquisition.

ACKNOWLEDGEMENTS

The authors would like to thank OR Melhoramento de Sementes Ltda for providing wheat flour samples, Marie-Hélène Morel and Joëlle Bonicel for conducting the electrophoresis analysis and Dr. Carlos Andrés Rodríguez Vega for his assistance in the processing data. The authors acknowledge the Foundation for Research Support of State of Rio de Janeiro (FAPERJ) (26/010.100.988/2018; 26/202.709/2018), National Council for the Scientific and Technological Development (CNPq) (427.116/2018; 310.34/2019-4), Coordination for Improvement of Personnel with Higher Education (CAPES) (financial code 001) and the Federal University of the State of Rio de Janeiro (UNIRIO) for financial support.

CAPÍTULO 4 – UNTARGETED PROTEOMICS UNRAVELS THE PROTEIN PROFILE OF WHEAT FLOURS OF CONTRASTING TRAITS BASED ON GLUTEN-FORCE AND NEAR-ISOGENIC LINES DIFFERING BY *PINB-D1* ALLELE

V.C.M. VICTORIO^a, T.O. ALVES^a, M.P.C. PIMENTEL^a, F.A. ALMEIDA^b, V. SILVEIRA^b,

V.LULLIEN-PELLERIN^c, M.S.L. FERREIRA^{ad*}

^aLaboratory of Bioactives, Food and Nutrition Graduate Program, PPGAN, Federal University of the State of Rio de Janeiro, UNIRIO, Av. Pasteur, 296, 22290-240, RJ, Brazil. ^bBiotechnology Laboratory, Center for Biosciences and Biotechnology (CBB), State University of North Fluminense Darcy Ribeiro (UENF), RJ, Brazil. ^cIATE, University Montpellier, INRAE, Institut Agro, Montpellier, France. ^dCenter of Innovation in Mass Spectrometry-Laboratory of Protein Biochemistry (IMaS-LBP), UNIRIO, Brazil. *Corresponding author. ^eCenter of Innovation in Mass Spectrometry (IMaS), Laboratory of Protein Biochemistry (LBP), UNIRIO, Brazil. *Corresponding author: mariana.ferreira@unirio.br

Artigo original a ser submetido à revista **Food Chemistry** (FI: 7,514; Qualis único Capes A1).

ABSTRACT

Total protein extraction was carried out for three pairs of wheat (*Triticum aestivum*) samples: whole wheat flours, refined wheat flours displaying contrasting characteristics in gluten-force (GF) and near-isogenic lines refined flours differing by *Pinb-D1* allele. Tryptic peptides were analyzed by NanoUPLC-HDMS^E. Proteomics data revealed that high GF and hard-type wheat flours showed greater number of proteins, greater amount of up-accumulated and exclusive proteins and even greater variety of biological process, when compared with their respective pairs low GF and soft-type phenotypes. Some storage proteins and small heat shock proteins were the best candidates for wheat specific markers associated with the characteristics of high GF and hard texture, while PINA, PINB2 and GSP can be considered as specific markers for the low GF and soft flours. Untargeted comparative proteomics unveiled the wheat proteome of flours with contrasting quality traits, contributing to the robustness of the research on wheat flour quality field.

Keywords: comparative proteomics; chemometric analysis; differentially accumulated proteins; functional annotation; technological quality; wheat proteome.

CAPÍTULO 5 – COMPARAÇÃO DE DOIS MÉTODOS DE ESPECTROMETRIA DE MASSAS APLICADOS PARA MELHOR CARACTERIZAÇÃO DAS PROTEÍNAS DO TRIGO: HDMS^E X UDMS^E

V.C.M. VICTORIO ^{1,2}, M.C.B. SANTOS ^{1,2}, C.T.S. D'ALMEIDA ^{1,2}, A.F. MACEDO ^{2,3}, G.H.M.F. SOUZA ⁴, L.C. CAMERON ², L.C. GUTKOSKI¹, M.S.L. FERREIRA ^{1,2*}

¹Laboratório de Bioativos, Programa de Pós-Graduação em Alimentos e Nutrição (PPGAN), Universidade Federal do Estado do Rio de Janeiro (UNIRIO). ²Centro de Inovação em Espectrometria de Massas, Laboratório de Bioquímica de Proteínas (IMasS - LBP), UNIRIO. ³Laboratório Integrado de Biologia Vegetal, Departamento de Botânica, UNIRIO. ⁴SpectraMass Ltd. *Autor correspondente.

Capítulo de livro. Frederico Celestino Barbosa. (Org.). **Nutrição em foco: uma abordagem holística**. 3^aed. Piracanjuba-GO: Editora Conhecimento Livre, 2020, v. 3, p. 405-433. <https://doi.org/1037423/200601222>

RESUMO

O trigo é um dos cereais mais importantes do mundo para o consumo humano. No Brasil, a pesquisa e a inovação aplicadas à cadeia produtiva do trigo são bastante desafiadoras. A qualidade intrínseca do grão trigo é determinada, principalmente, pela qualidade do glúten que depende da composição das proteínas de reserva e do perfil polimérico das gluteninas. Recentemente, a espectrometria de massas sequencial (*tandem mass spectrometry* - MS/MS) foi aplicada para identificar e caracterizar proteínas de trigo. No entanto, as proteínas do glúten apresentam alta homologia aminoacídica e são difíceis de serem discriminadas. Neste trabalho, a partir de ferramentas proteômicas foi proposto caracterizar as proteínas de trigo a partir de dois métodos de aquisição: espectrometria de massas de alta definição (HDMS^E) comparada com espectrometria de massas de ultra definição (UDMS^E) aplicando-se baixa e alta energia para fragmentação e sistema de mobilidade iônica. Farinhas de trigo brasileiro (*Triticum aestivum*), oriundas de duas variedades de qualidades tecnológicas contrastantes, foram submetidas à extração sequencial, *clean-up* e digestão proteica (LysC + Trypsin). Os peptídeos foram submetidos ao nanoUPLC acoplado ao espectrômetro de massas Synapt G2-S HDMS. Os dados foram processados no ProteinLynx Global Server v3.0 e a pesquisa foi realizada em um banco de dados personalizado do UniProtKB 2015_03v. Nas análises UDMS^E, observou-se um aumento de 80% para identificação total de peptídeos e

proteínas. Embora, em média, 3 peptídeos por proteína tenham sido identificados em ambos os métodos, a cobertura da sequência peptídica foi notavelmente melhorada no UDMS^E. Analisando-se os fragmentos *b* e *y*, foi obtida uma melhora significativa em íons com *m/z* superior a 1.200 em UDMS^E. Entre os *hits* aceitos (probabilidade de 95%), as proteínas solúveis apresentaram o maior número de identificação de proteínas (20), seguidas pelas gluteninas insolúveis (8). Dentre as últimas, foram identificadas proteínas de interesse particular, como LMW-GS codificado no *Glu-B3* locus e HMW-GS (1Ax1, 1Dx5, 1By9), proteínas fortemente associadas à qualidade da massa de trigo. Conclui-se, portanto que o método de aquisição UDMS^E foi melhor na identificação de diferentes proteínas. Devido à melhor fragmentação, este método possibilitou uma maior cobertura da sequência proteica em comparação ao método HDMS^E.

Palavras-chave: Métodos MS^E; Proteômica de Cereais; Qualidade Tecnológica.

ABSTRACT

Wheat is one of the most important cereals in the world for human consumption. In Brazil, research and innovation applied to the wheat production chain are very challenging. The intrinsic quality of the wheat grain is mainly determined by the quality of the gluten, which depends on the composition of the storage proteins and the polymeric profile of the glutenins. Recently, tandem mass spectrometry (MS/MS) has been applied to identify and characterize wheat proteins. However, gluten proteins have a high homology in the amino acid sequence and are particularly difficult to discriminate. In this work, proteomic tools were applied to characterize wheat proteins using two acquisition methods: high definition mass spectrometry (HDMS^E) compared with ultra-definition mass spectrometry (UDMS^E) applying simultaneous low and high energy for fragmentation and ion mobility system. Brazilian wheat flours (*Triticum aestivum*) from two varieties with contrasting technological qualities were subjected to sequential extraction, clean-up and protein digestion (LysC + Trypsin). The peptides were submitted to nanoUPLC tandem to the Synapt G2-S HDMS mass spectrometer. The data were processed in ProteinLynx Global Server v3.0 and the research was carried out in a customized database of UniProtKB 2015_03v. In the UDMS^E analyzes, an increase of 80% was observed for peptide and protein total identification. Although, on average, 3 peptides per protein were identified in both methods, peptide sequence coverage was markedly improved in UDMS^E. Analyzing fragments *b* and *y*, a

significant improvement was obtained in ions with m/z greater than 1,200 in UDMS^E. Among the accepted hits (95% probability), soluble proteins had the highest number of protein identifications (20), followed by insoluble glutenins (8). Insoluble glutenins of particular interest were identified, such as LMW-GS encoded in *Glu- B3* locus and HMW-GS (1Ax1, 1Dx5, 1By9), which are strongly associated with the quality of the wheat dough. A conclusion, it was verified that the UDMS^E acquisition method improved the protein identification due to the better fragmentation, this method allowed a greater coverage of the protein sequence compared to the HDMS^E method.

Keywords: MS^E Methods; Cereal Proteomics; Technological Quality.

1 INTRODUÇÃO

O trigo é amplamente utilizado para a produção de massas e pães por apresentar características viscoelásticas únicas (VENSEL *et al.*, 2014). No Brasil, a produção do cereal é inferior à demanda, havendo necessidade de importação para o abastecimento do país, com garantia de segurança alimentar à população. Dados da Companhia Nacional de Abastecimento (CONAB, 2020) apontam que, para suprir a demanda interna, foram importadas 748,2 mil toneladas de trigo no mês de abril de 2020. A previsão de importação prevista para a safra atual é de 7,2 milhões de toneladas enquanto a produção de trigo foi de 5,4 milhões de toneladas.

Um dos principais desafios da agricultura brasileira tem sido a produção de trigo de forma competitiva e sustentável com técnicas de cultivo adequadas e obtenção de bons rendimentos e qualidade tecnológica (PIRES *et al.*, 2011). A qualidade tecnológica define a orientação industrial de uso do trigo e é determinada pela composição e distribuição de massa molecular das proteínas de reserva que formam o glúten (DHAKA; KHATKAR, 2015; PIROZI *et al.*, 2008; POPINEAU *et al.*, 1994a). No entanto, a qualidade tecnológica comercial do trigo é definida mediante testes físico-químicos (peso do hectolitro, peso de mil grãos, extração experimental de farinha, número de queda, etc.) e reológicos (alveografia, farinografia, etc) (AACC, 2000) e não leva em conta parâmetros primordiais que afetam a qualidade da massa como a composição proteica do glúten.

As proteínas dos cereais foram divididas por Osborne, em 1907, em quatro classes de acordo com sua solubilidade, sendo as albuminas solúveis em água, as globulinas em soluções salinas, as prolaminas (gliadinas) em soluções alcoólicas e as glutelinas (gluteninas) em ácidos ou bases diluídas (DE BRIER *et al.*, 2015). No trigo, estas duas últimas são classificadas como proteínas de reserva ou do glúten. Sabe-se que a distribuição da massa molecular dos polímeros das proteínas de reserva e, principalmente, as frações poliméricas insolúveis afetam positivamente a viscoelasticidade do glúten e, portanto, são os principais contribuintes do potencial tecnológico do trigo (DELCOUR *et al.*, 2012; NAEEM; MACRITCHIE, 2005; POPINEAU *et al.*, 1994b).

A caracterização e análise destas proteínas é fundamental para melhor compreensão da qualidade do trigo sendo necessária a utilização de técnicas avançadas

que permitam elucidar a composição proteica a partir de técnicas sensíveis e confiáveis. A abordagem proteômica como a cromatografia líquida acoplada à espectrometria de massas é capaz de separar até mesmo peptídeos de mesma m/z (i.e. sistema de mobilidade iônica) (WANG *et al.*, 2013) e revelam-se como importantes ferramentas para a quantificação e identificação das proteínas do trigo (MARTINEZ-ESTESO *et al.*, 2016). Recentemente, essa abordagem foi aplicada para revelar a expressão diferencial de proteínas metabólicas (albuminas e globulinas) em farinhas de trigo de diferentes qualidades tecnológicas (VICTORIO *et al.*, 2018), bem como para decifrar peptídeos ligados à alergenicidade do trigo (ALVES *et al.*, 2017; ALVES *et al.*, 2018) utilizando métodos MS^E (espectrometria de massas multiplexada sequencial).

Os métodos MS^E são estratégias de aquisição independente de dados (em inglês, DIA). Os íons gerados na fonte são transmitidos até a célula de colisão, que alterna entre baixa e alta energia, enviando alternadamente precursores e fragmentos ao analisador. Dessa forma, são obtidos resultados mais abrangentes e reprodutíveis porque ocorre a fragmentação simultânea de múltiplos íons precursores, independentemente da intensidade ou de outra característica (CHAPMAN *et al.*, 2014; DISTLER *et al.*, 2014; GEROMANOS *et al.*, 2009; SOUZA *et al.*, 2017). O método de alta definição denominado *High Definition MS^E* (HDMS^E) é o método MS^E acrescido do sistema de mobilidade iônica (em inglês, IMS). Neste método há uma dimensão adicional de separação, visto que os íons são separados, de acordo com sua mobilidade, em gás inerte, baseados na carga, massa, forma e tamanho. Assim sendo, são reduzidas as interferências e há uma melhora na capacidade máxima do sistema (RÍOS-CASTRO *et al.*, 2020). A fragmentação ocorre depois do sistema IMS aumentando a seletividade do alinhamento dos precursores com seus produtos. O método de ultra definição, conhecido como *Ultra Definition MS^E* (UDMS^E), é um método otimizado no qual é feita uma rampa de energia de colisão de acordo com os dados obtidos na IMS (energia de colisão específica para o tempo de aceleração). Dessa forma, a eficiência de fragmentação é melhorada, aumentando a intensidade de desempenho MS/MS para sequenciamento de peptídeos (DISTLER *et al.*, 2014; DISTLER *et al.*, 2016; SOUZA *et al.*, 2017).

Dentro deste contexto, esse trabalho teve por objetivo comparar os resultados obtidos da análise por cromatografia líquida de ultra performance e espectrometria de

massas, pelos métodos HDMS^E e UDMS^E utilizando dois cultivares de farinha de trigo brasileiro de qualidade tecnológica contrastantes. Com isso, almeja-se evidenciar informações relevantes que permitam a melhor compreensão da qualidade tecnológica de farinhas de trigo, com impacto positivo na cadeia produtiva do cereal.

2 METODOLOGIA

2.1 MATERIAL VEGETAL

As amostras de farinha de dois cultivares de trigo (*Triticum aestivum*) foram cedidas pela empresa OR Melhoramento de Sementes Ltda de Passo Fundo-RS. As cultivares foram provenientes da safra 2014/2015 e tiveram colheita e manejo normal de sementes. As amostras possuem qualidades contrastantes: Campeiro (amostra 1), de classificação fraca e Ametista (amostra 4), de classificação forte. A classificação foi baseada na análise da força do glúten e da estabilidade. Esta classificação é norteada pelos métodos da AACC (American Association of Cereal Chemists) e pela Instrução Normativa MAPA 38/2010, que utilizam a determinação da força do glúten (J) pelo método de Alveografia, utilizando como parâmetros tenacidade (P), extensibilidade (L), enquanto que a estabilidade é medida em farinógrafo, conforme dados na tabela apresentada abaixo (Tabela 1). As amostras foram armazenadas em recipientes herméticos a -20°C até o momento das análises.

2.2 REAGENTES QUÍMICOS

Bicarbonato de amônio (NH₄HCO₃) (PubChem CID: 14013), β-mercaptoetanol (PubChem CID: 1567), albumina de soro bovino (BSA) (PubChem CID: 16132389), reagente de Bradford (p/n B6916 Sigma-Aldrich), 1,4-ditiotreitol (DTT) (PubChem CID: 19001), iodoacetamida (IAM) (PubChem CID: 3727), dodecil sulfato de sódio (SDS) (PubChem CID: 3423265), Tris-HCl (PubChem CID: 93573), ácido trifluoroacético (TFA) (PubChem CID: 6422), bem como acetonitrila (PubChem CID: 6342) e ácido fórmico (PubChem CID: 284) para LC-MS foram adquiridos da Sigma Aldrich (St. Louis, MO). Surfactante RapiGest (p/n 186002118 Waters) e [Glu1]-Fibrinopeptídeo B humano (GFP) (PubChem CID: 16172403) foram adquiridos de Waters Corp. (Milford, MA), enquanto

tripsina/Lys-C (p/n V5071) de grau de espectrometria de massas para a digestão enzimática foram compradas da Promega Corporation (Madison, WI).

2.3 ALQUILAÇÃO E EXTRAÇÃO SEQUENCIAL DAS PROTEÍNAS DE FARINHA DE TRIGO

A extração sequencial foi realizada considerando a solubilidade das proteínas segundo a classificação de Osborne (1907) com obtenção de 4 extratos: albuminas e globulinas (A/G -solúveis em Tris-HCl), gliadinas (Gli - etanol 70%), gluteninas solúveis (GS - 2% SDS) e gluteninas insolúveis (GI - 2% SDS, 20 mM DTT). Os procedimentos foram realizados de acordo com Ferreira et al. (2014), onde foram pesados 100 mg de cada amostra (balança Shimadzu AUX220) em tubos eppendorfs identificados (1,5 mL, Axygen) e acrescentou-se 1 mL de tampão de 80 mM Tris-HCl pH 8 contendo 40 mM de iodoacetamida (IAM). As amostras foram homogeneizadas em vortex, agitadas em shaker (60 minutos; 200 rpm; 25 °C; Incubadora TE-420 Tecnal) e centrifugadas (10.600 xg; 10 min; 20°C; centrífuga refrigerada Megafuge 16R Centrifuge, Thermo). Os extratos contendo A/G foram obtidos a partir de 650 µL recolhidos de cada sobrenadante e congelados a -20 °C. O restante dos sobrenadantes foi descartado e os pellets ressuspensos em 1 mL de etanol 70% (preparado com água com grau de pureza LC-MS). As amostras foram homogeneizadas (15 min; 200 rpm; 25 °C) e centrifugadas (10.600 x g; 20 min; 20 °C). Os extratos contendo Gli foram obtidos a partir de 650 µL recolhidos de cada sobrenadante e congelados a -20 °C. Os pellets foram ressuspensos em 1 mL de 80 mM Tris-HCl (pH 8) com 2% SDS, homogeneizados (200 rpm; 60 min; 24,7 °C) e centrifugados (10.600 x g; 20 min; 20 °C). Para obtenção dos extratos GS, realizou-se o mesmo procedimento feito para os demais. Os pellets finais foram então ressuspensos em 750 µL de 80 mM Tris-HCl (pH 8), 20 mM ditioneitol (DTT) e 2% SDS, colocados em banho-maria (30 min; 60 °C; ALB 250 C Loja Lab), submetidos à sonicação com ponteira (15 s, 20% potência; Eco-Sonics; 500 Watts) e centrifugados (10.600 x g; 20 min; 20 °C). Os extratos GI foram obtidos como os demais e armazenados a -20 °C. Todas as etapas foram realizadas sob proteção da luz.

2.4 QUANTIFICAÇÃO DE PROTEÍNAS

A quantificação de proteínas foi realizada pelo método de Bradford (1976). Albumina de soro bovino (BSA) foi usada para a curva padrão, partindo da concentração de 2 mg/mL com diluições seriadas até a concentração final de 0,0625 mg/mL. As curvas padrões foram consideradas quando o R^2 foi superior a 0,98. Os extratos de proteínas foram diluídos uma vez em água e 20 μ L de amostra ou de BSA foram pipetados em eppendorf acrescentando-se 1 mL de reagente de Bradford (Sigma-Aldrich). O tempo de reação foi de 5 min sob proteção da luz. As leituras foram realizadas em triplicata a 595 nm em espectrofotômetro (UV-2700, Shimadzu). Para verificação, a etapa de quantificação proteica foi realizada antes e depois da etapa de *clean-up*.

Tabela. 1. Caracterização da qualidade tecnológica de duas cultivares de trigo brasileiro, segundo parâmetros reológicos e físico-químicos.

Cultivares	Qualidade	Classificação	Força do glúten (J)	Estabilidade e (min)	Número de queda (s)	Tenacidade e (mm)	Extensibilidade e (mm)	Absorção de água (%)
CAMPEIRO	Doméstico	Fraca	119,25	7,25	291	32	152	49
AMETISTA	Pão/ Melhorador	Forte	352,7	17	325,5	117	87	67,5

2.5 ETAPA DE LAVAGEM (*CLEAN-UP*)

Foram utilizados filtros Amicon 3 kDa (Ultra Centrifugal Filters, Millipore) previamente lavados com 500 μ L de água LC-MS e centrifugados (14.000 xg; 30 min; 4 °C). Após descongelamento, as amostras foram centrifugadas (10.000 xg; 10 min; 4 °C) e os filtros Amicon foram centrifugados invertidos para retirada da água (1.000 xg; 2 min; 4 °C). Então, 500 μ L de amostra foram colocados no filtro e centrifugados (14.000 xg; 30 min; 4 °C). Foram realizadas duas lavagens com 500 μ L bicarbonato de amônio (50 mM; pH 8,5) e posterior centrifugação (14.000 xg; 60 min; 4 °C). Para recuperar o retido, os filtros foram invertidos em novos tubos e centrifugados (1.000 xg; 2 min; 4 °C). O volume recuperado aproximado de cada amostra foi anotado e as amostras foram diluídas em bicarbonato de amônio até atingirem concentração aproximada de 1 μ g/ μ L.

2.6 DIGESTÃO TRÍPTICA

Uma alíquota de 50 μ L de cada extrato proteico com concentração aproximada de 1 μ g/ μ L foi colocada em eppendorf e foram adicionados 10 μ L de 50 mM bicarbonato de amônio pH 8,5 e 25 μ L de solução 0,2% v/v RapiGest SF (Waters). As amostras foram homogeneizadas, colocadas em banho-maria (15 min; 80 °C) e centrifugadas (14.000 x g; 10 min; 4 °C). Em seguida, foram adicionados 2,5 μ L de 100 mM DTT (50 mM NH_4HCO_3 , pH 8,5) e, novamente, as amostras foram homogeneizadas, colocadas em banho-maria (30 min; 60 °C) e centrifugadas (14.000 x g; 10 min; 4 °C). Após adição de 2,5 μ L de 300 mM iodoacetamida (50 mM NH_4HCO_3 , pH 8,5) as amostras foram vortexadas e deixadas no escuro (20 °C, 30 min). Foram adicionados 20 μ L de Tripsina Lys-C (Trypsin/Lys-C Mix, *Mass Spectrometry Grade*, Promega) preparada em 50 mM NH_4HCO_3 , pH 8,5 e homogeneizou-se. A etapa de digestão foi feita em *overnight* em shaker (14 horas; sem agitação; 37 °C). Após o *overnight*, foram adicionados 10 μ L de 5% ácido trifluoroacético, seguindo com incubação em shaker (90 min; sem agitação; 37°C) e centrifugação (14.000 x g; 30 min; 4 °C). Os sobrenadantes foram transferidos para eppendorf de 600 μ L e armazenados em ultrafreezer.

2.7 AQUISIÇÃO EM 1D NANOUPLC-MS^E

Aos *vials* (*total recovery*, Waters) foram adicionados 5 μL de hidróxido de amônio (NH_4OH , 1N) e quantidades diferentes de amostra e formiato de amônio a 20mM (NH_4FO). Assim sendo, em cada *vial* obteve-se uma concentração aproximada de proteínas de 733 ng/ μL . Para as amostras de GLI, o volume final no *vial* foi de aproximadamente 135-140 μL , 100 μL de amostra e 30-35 μL de 20mM NH_4FO . Para as demais amostras o volume final no *vial* foi de 150 μL , sendo 110 μL de amostra e 35 μL de 20mM NH_4FO .

A separação em NanoUPLC dos peptídeos tripsinizados foi realizada usando um sistema 1D simulado nanoACQUITY UPLC (Waters) equipado com uma pré coluna XbridgeTM BEH130 C₁₈ (5 μm , 300 μm x 50 mm, Waters), coluna Trap 2G nanoAquity UPLC C₁₈ (5 μm , Waters) e coluna de fase reversa nanoAquity BEH130 C₁₈ (1,7 μm , 100 μm x 100 mm, Waters). A coluna foi acondicionada com os seguintes parâmetros: capilar (2,7 kV), *sample cone* (30 V), *source offset* (30 V), temperatura (70 °C), variação de *m/z* de 50 a 2000 e tempo de varredura de 0,5 s. Acoplado ao NanoUPLC utilizou-se o sistema de espectrometria de massas do instrumento Synapt G2-S HDMS (Waters) com quadrupolo, mobilidade iônica e tempo de voo (*time-of-flight* - TOF). Todas as análises foram realizadas em modo resolução com nanoElestrospray em modo positivo (nanoESI+). O analisador de TOF foi externamente calibrado com solução equimolar de Leu-Enk (Leucina Encefalina) e GFP (Glu1-fibrinopeptídeo B) de 200 fmol.L⁻¹. A relação *m/z* do GFP usada na calibração foi de 785,8426 (*LockMass*) e da Leu-Enk foi de 556,2771 para cálculo de VEFF. *Lock spray source setup* foi realizado com GFP.

2.8 ETAPA DE SCOUTING RUN

Para cada amostra foram realizadas corridas de prospecção (*scouting run*) objetivando normalizar a área total dos picos obtidos de proteínas totais injetadas na coluna, bem como a quantidade de proteínas a ser injetada no instrumento. Foram injetados 4 μL de cada amostra com vazão de 500 nL.s⁻¹ segundo em sistema 1D simulado, numa coluna de 100 μm x 100 mm. O método utilizado foi o MS^E (modo de aquisição contínuo com aplicação simultânea de baixa e alta energia na célula de colisão), 29 minutos de corrida em modo “resolution” e polaridade positiva. Foram

adquiridos íons com massa entre 50 e 2.000 Da, com tempo de varredura de 0,5 segundos, em modo de dados contínuo, rampa de colisão de energia de 15 a 53 V e voltagem do cone de 30 V. As amostras correram por 29 minutos com gradiente das fases com gradiente de 3 a 45 % de fase móvel B na primeira dimensão com fluxo de 2.000 nL.min⁻¹. Na segunda dimensão o fluxo foi de 500 nL.min⁻¹ com gradiente de 7 a 85% de fase móvel B. A fase móvel A era constituída de água ultra-pura (grau LC-MS) com 0,1% de ácido fórmico e a B de acetonitrila com 0,1% de ácido fórmico. A coluna foi reequilibrada até as condições iniciais e a temperatura da coluna foi mantida em 55 °C. A solução *Lock Mass* foi oriunda da bomba B (ASM) com fluxo constante de 200 nL.min⁻¹ e limite superior de pressão de 10.000 psi.

2.9 MÉTODOS HDMS^E E UDMS^E

Após as aquisições de *scouting run*, as amostras foram adquiridas pelo método multiplexado de análise MS/MS HDMS^E injetando 4 µL, onde os peptídeos foram analisados com sistema de mobilidade iônica seguidos de aplicação simultânea de baixa e alta energia de fragmentação. A aquisição foi de 0 a 29 minutos em modo “*resolution*”, polaridade positiva, com íons de massa entre 50 e 2.000 Da. O tempo de varredura foi de 0,5 segundos, rampa de transferência e colisão de energia de 19 a 45 V e voltagem do cone de 30 V. Os espectros gerados após a aquisição pelo método HDMS^E foram processados no programa DriftScope (Waters) (Figura 1). A seleção da área a ser utilizada na aquisição por UDMS^E foi padronizada para todas as amostras com os seguintes critérios: a linha inferior da área seleciona partiu do ponto zero e foi até a marca intercessora entre *drift time* 200 (eixo x) e relação massa/carga (*m/z*) 1850 *m/z*(eixo y) e a linha superior iniciou em *drift time* 0 e 50 *m/z* e terminou em *drift time* 68 e 2000 *m/z*. Esse método objetivou a retirada da faixa de sinais intensos correspondentes aos íons monoprotionados, não característicos em ESI.

Após o tratamento dos espectros obtidos por HDMS^E, foi então realizado o método UDMS^E com injeção de 4 µL em triplicata, onde o instrumento selecionou apenas íons com 2 ou mais cargas para serem fragmentados e analisados. A aquisição foi de 29 minutos em modo *resolution*, polaridade positiva, com íons de massa entre 50 e 2.000 Da e voltagem do cone de 30 V.

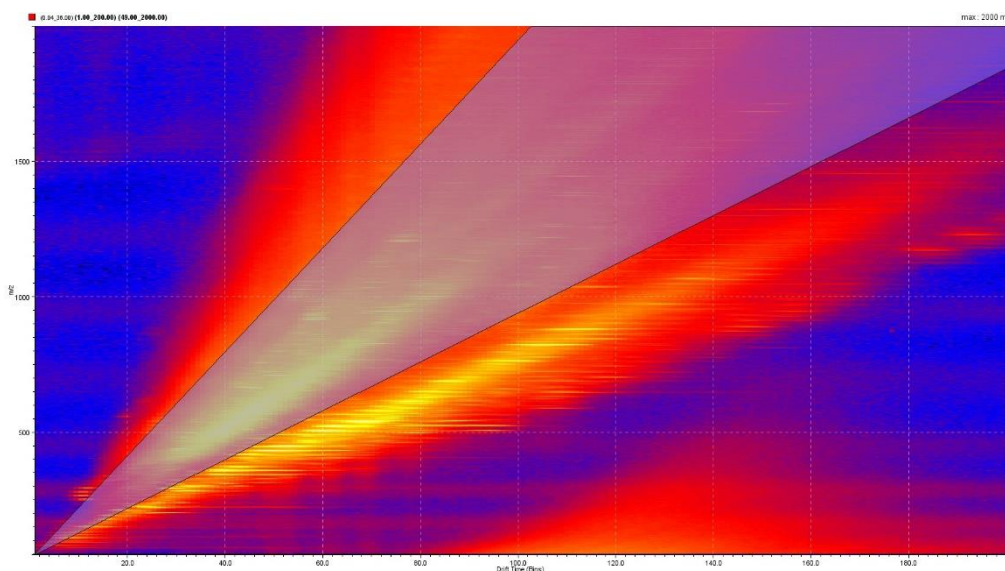


Figura. 1. Exemplo de espectro obtido por HDMS^E processados no programa DriftScope.

2.10 PROCESSAMENTO DE DADOS

Para identificação e quantificação de peptídeos e proteínas foi utilizado o *software Progenesis Q1 for Proteomics* (v 2.1 NonLinearDynamics, Waters), alimentado com o banco de dados de proteínas de *Triticum aestivum* do UNIPROT KB release 2015-10, publicado em 14.10.2015 (UNIPROT). A taxa de falsos positivos (*false discovery rate* - FDR) permitida foi menor que 4%, e para combinação de íons foram estipulados um ou mais fragmentos por peptídeo, três ou mais fragmentos por proteína e um ou mais peptídeos por proteína. Para a identificação de proteínas, a exatidão de massa aplicada foi inferior a 10 ppm e score maior que 3. As análises foram baseadas em quantificação sem marcadores químicos ou derivatizantes (*label free*). Para cada espectro de fragmentação gerado por MS/MS, o software usa um algoritmo padrão estabelecido com 14 diferentes propriedades físico-químicas para confirmar a proteína e a identidade de seus peptídeos relacionados. Para comparação dos dados obtidos por HDMS^E e UDMS^E o software utilizado foi o ProteinLynx Global Server v3.0, onde os íons moleculares precursores foram associados com os fragmentos para identificação de proteínas com base em seu tempo de retenção de massa exata (EMRT), bem como o *drift time*.

3 RESULTADOS E DISCUSSÃO

A partir da análise dos extratos sequenciais foram realizadas as corridas de prospecção para análise da área total dos picos e cálculo da injeção do volume de amostra. Esse procedimento é importante porque permite que as condições sejam comparadas umas às outras. As aquisições foram feitas com injeção do volume máximo permitido para o *loop* utilizado de 4 μL (Figura 2).

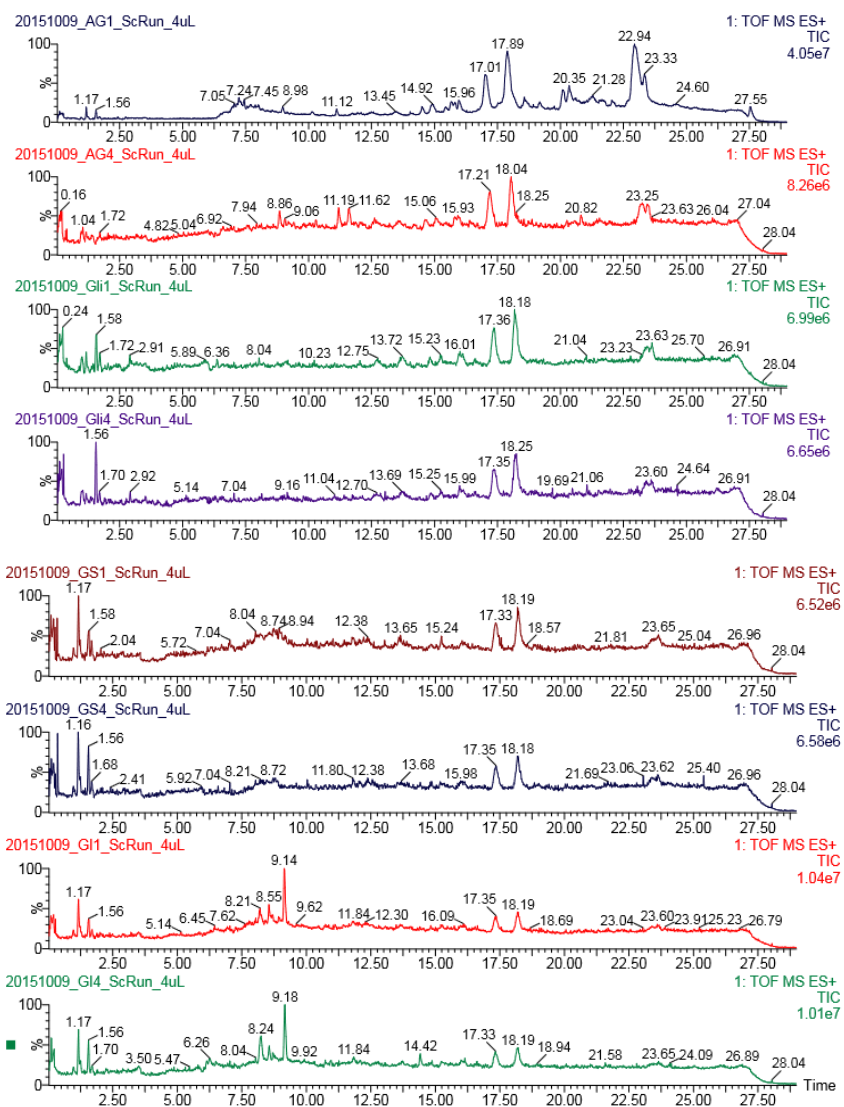


Figura. 2. Cromatogramas das corridas de prospecção (*scouting run*) dos extratos sequenciais (AG, Gli, GS e GI) das amostras Campeiro e Ametista em TIC (*total ion count*).

Em seguida, os extratos foram adquiridos no método HDMS^E, utilizando o sistema de mobilidade iônica através da técnica de *Travelling Wave Ion Mobility Mass Spectrometry* (TWIM-MS) que é uma separação ortogonal, onde uma nova dimensão é adicionada aos dados conferindo para cada valor de m/z um espectro de *drift time*¹, separando os íons pelas suas seções de choque (MICHAILEVSKI *et al.*, 2010; SOUZA *et al.*, 2017). Assim, os íons são diferenciados por tamanho, forma e carga, além da massa. Nos cromatogramas obtidos por meio do método de HDMS^E (Figura 3) foram detectados picos majoritários (TR 17-18,31 min, 23,42 min), presentes em todas as amostras analisadas. A análise dos espectros obtidos mostrou que esses picos correspondem a compostos monoprotonados e alguns foram identificados como contaminantes (*i.e.* polímeros oriundos do capilar usado na fonte NanoESI). Somado a isto, sabe-se que peptídeos se apresentam multicarregados principalmente quando obtidos em fontes do tipo ESI. A partir disso, foi utilizado o software DriftScope (Waters) que utiliza os dados de *drift time*, permitindo a detecção e separação de interferentes dos demais componentes de interesse em misturas complexas. Foram então selecionados apenas íons com 2 ou mais cargas para serem fragmentados e analisados, conforme exemplo da Figura 1.

Nos cromatogramas obtidos por meio do método de UDMS^E (Figura 4) é possível observar, claramente, a melhora dos sinais obtidos com este método. Os dados obtidos nas aquisições em modo HDMS^E e UDMS^E foram processados no PLGS e comparados entre si, sendo apresentados os espectros obtidos pelo método UMDS^E (Figuras 5 para AG e Gli e Figura 6 para GS e GI).

¹*Drift time* (dt) é medido em milissegundos e corresponde ao tempo que íon leva para atravessar a cela de mobilidade iônica onde é inserido o gás inerte, no caso o hélio (He) .

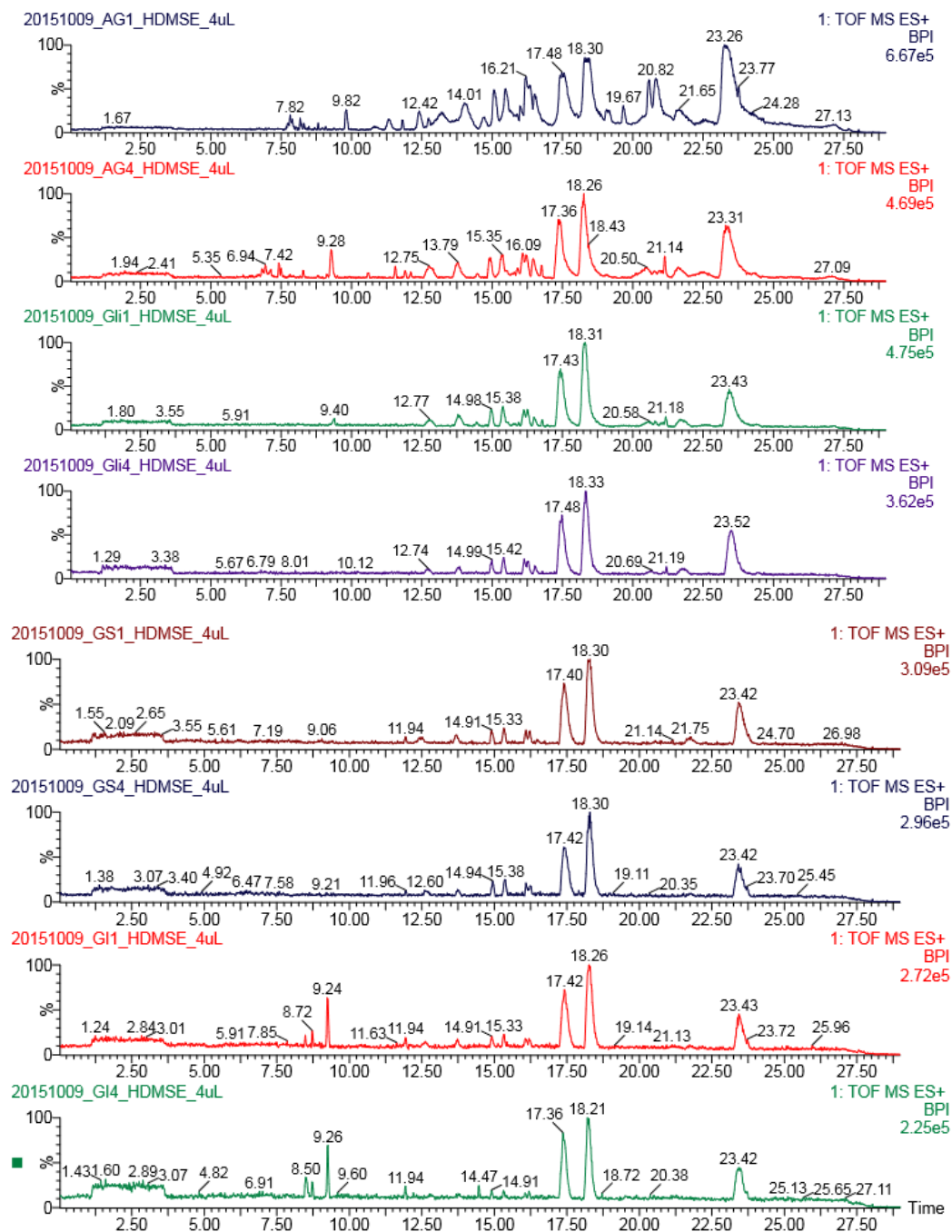


Figura. 3. Cromatogramas das aquisições por HDMS^E dos extratos sequenciais (AG, Gli, GS e GI) das amostras Campeiro e Ametista em BPI (*base peak intention*).

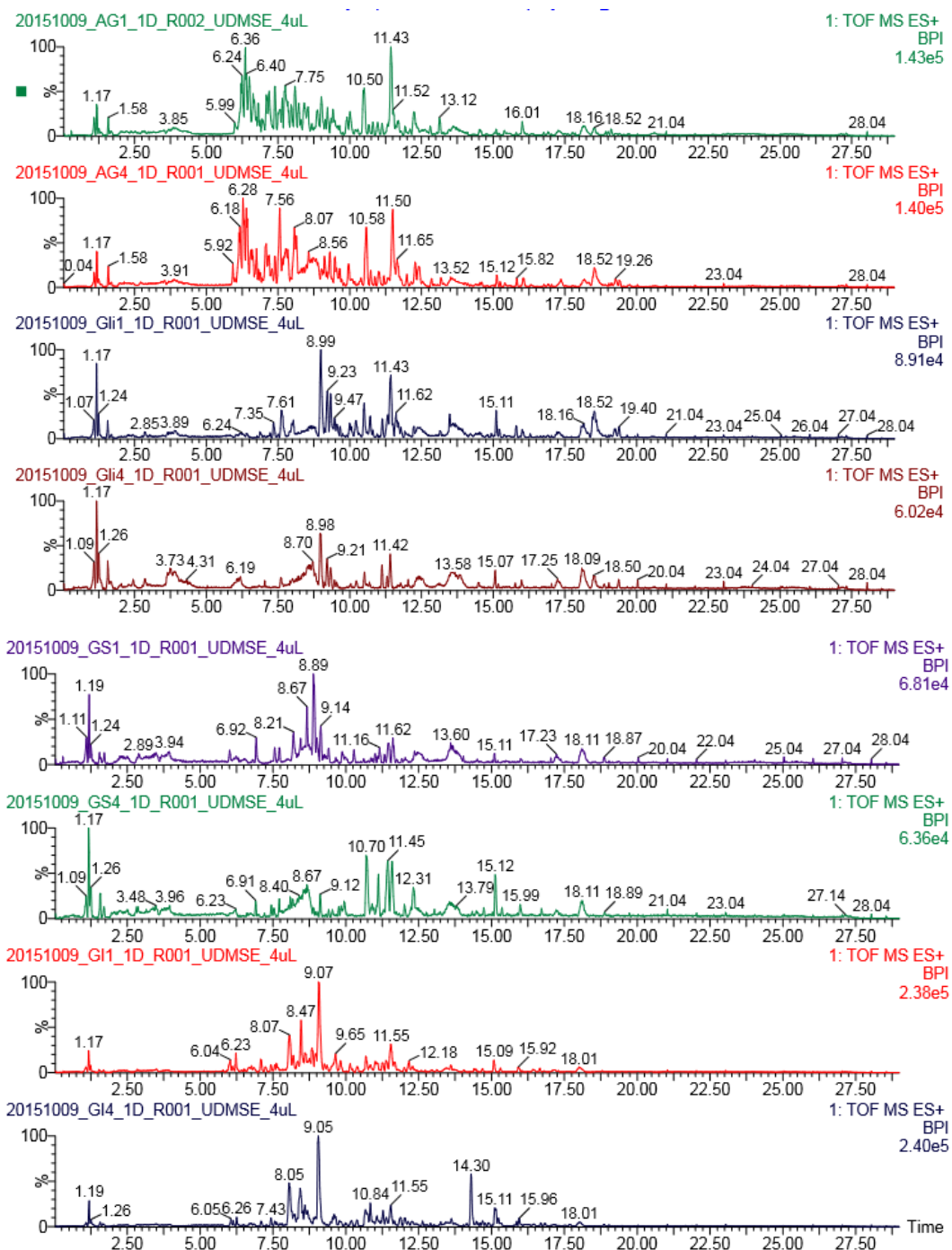


Figura. 4. Cromatogramas das aquisições por UDMSE dos extratos sequenciais (AG, Gli, GS e GI) das amostras Campeiro e Ametista em BPI (base peak intention).

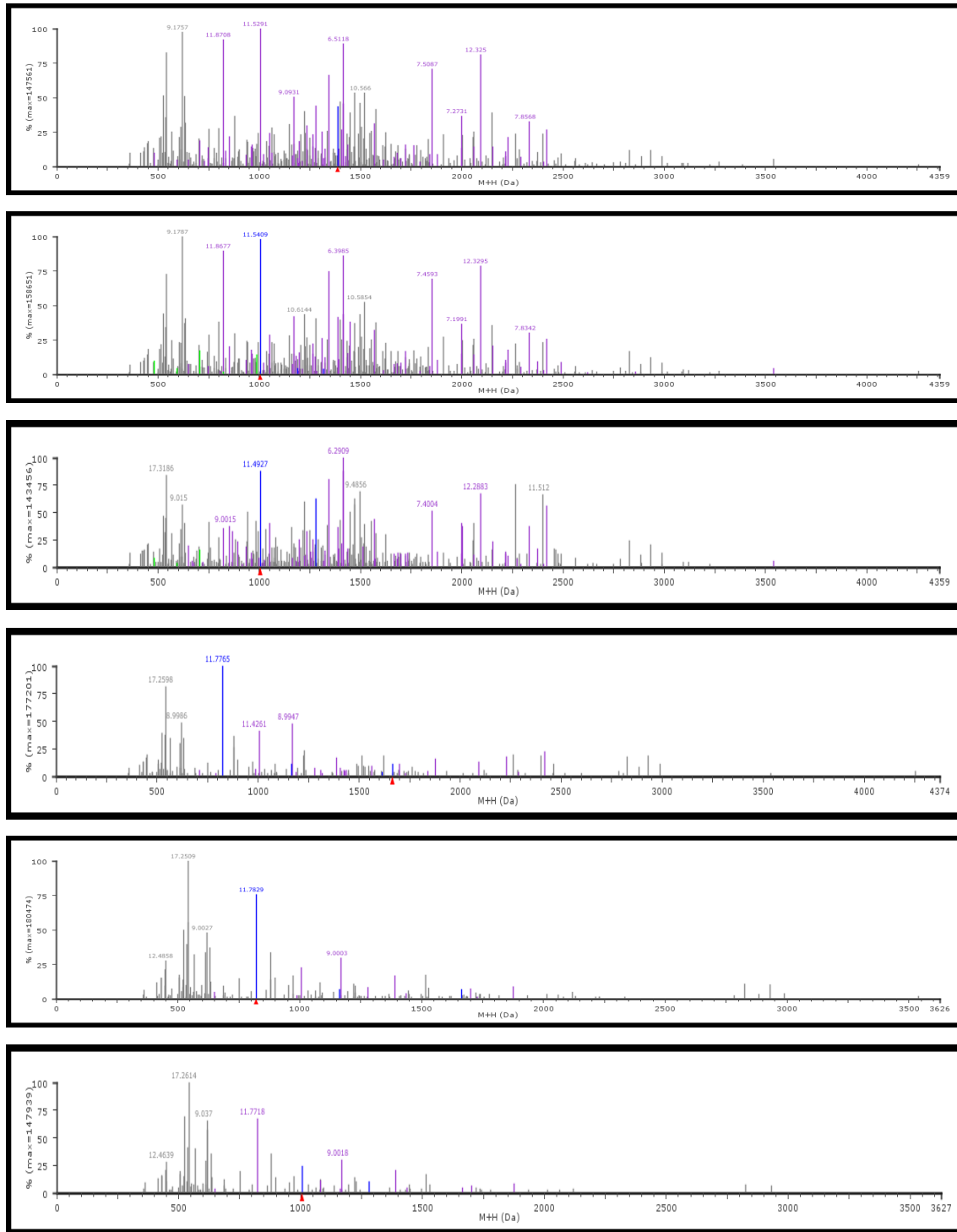


Figura. 5. Espectros UDMS^E das aquisições em triplicata dos extratos AG1 (superiores) e Gli1 (inferiores), em roxo os peptídeos associados à identificação de proteínas.

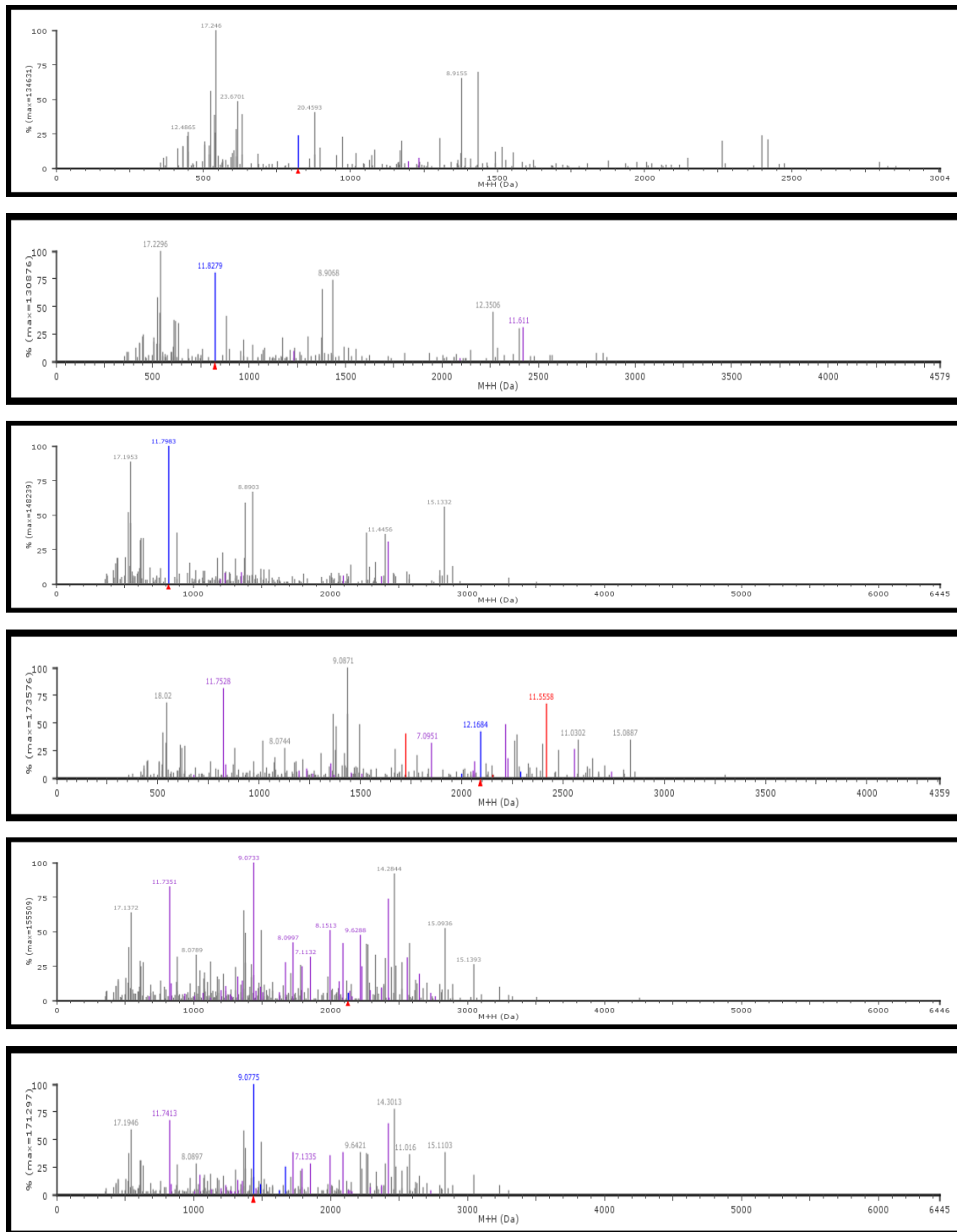


Figura. 6. Espectros UDMS^E das aquisições em triplicata dos extratos GS1 (superiores) e GI1 (inferiores), em roxo os peptídeos associados à identificação de proteínas.

Comparou-se o número de proteínas e de peptídeos identificados para todos os extratos analisados (AG, Gli, GS e GI) das duas variedades estudadas em ambos os métodos HDMS^E e UDMS^E, sendo neste último considerada a média das triplicatas (Figura 7). Houve um aumento significativo do número de peptídeos e proteínas identificadas pelo método UDMS^E, principalmente nos extratos AGs e Glis. O maior número de identificação de proteínas ocorreu nos extratos de proteínas solúveis (albuminas e globulinas), seguido dos extratos de gluteninas insolúveis. Quando apenas os dados de peptídeos e proteínas identificadas com intervalo de confiança acima de 95% foram considerados, observou-se um aumento de aproximadamente 60% nas análises UDMS^E (Figura 8).

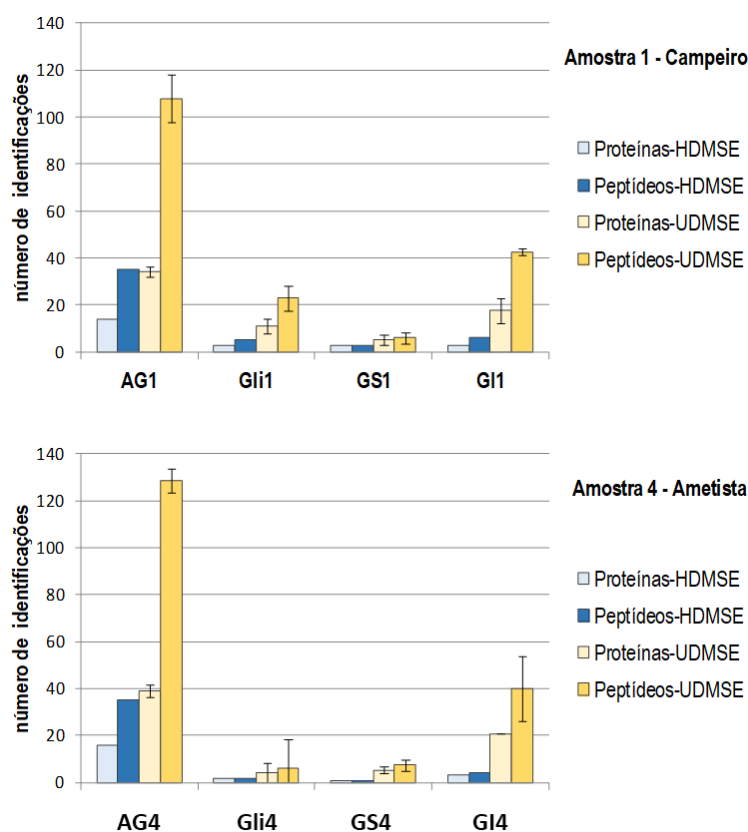


Figura. 7. Comparação do número de identificações de peptídeos e proteínas em HDMS^E e UDMS^E nos extratos Albuminas/Globulinas (AG), Gliadinas (Gli), Gluteninas Solúveis (GS) e Gluteninas Insolúveis (GI).

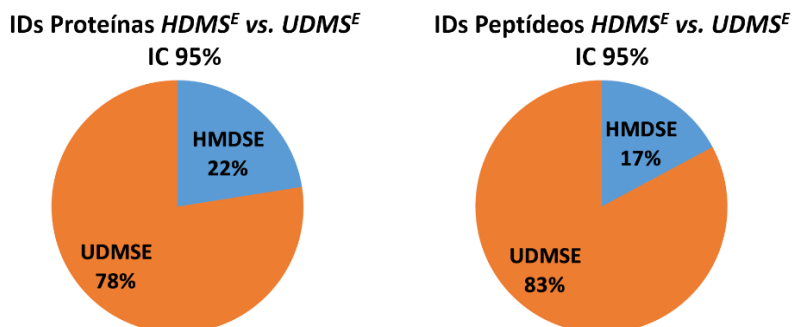


Figura. 8. Comparação do número total de identificações de proteínas e peptídeos identificados com intervalo de confiança (IC) acima de 95%, em HDMS^E e UDMS^E.

Foi realizada análise detalhada dos fragmentos *b* e *y* obtidos pelos métodos HDMS^E e UDMS^E nos extratos AG (Figura 9) e Gli, Gs e GI (Figura 10). É possível observar que houve melhora significativa da fragmentação dos precursores no método UDMS^E, principalmente em íons com *m/z* superior a 1.200. Assim como apresentado acima e também já observado desde o cromatograma e nos espectros resultantes, os extratos com maiores números de identificação são os AG e GI. Este resultado é claramente observado nas figuras abaixo que representam a distribuição de fragmentos *b* e *y* obtidos nestes extratos quando comparados HDMS^E e UDMS^E. Embora, em média o número de 3 peptídeos por proteína identificada tenha sido obtido em ambos os métodos, o que denota a confiabilidade dos resultados em ambos os métodos, o que foi determinante foi a melhora da cobertura de sequência no método UDMS^E. Assim, é possível atribuir que o aumento significativo de peptídeos e proteínas identificadas, assim como da cobertura de sequência, foi devido a uma fragmentação mais eficiente nas aquisições pelo método UDMS^E.

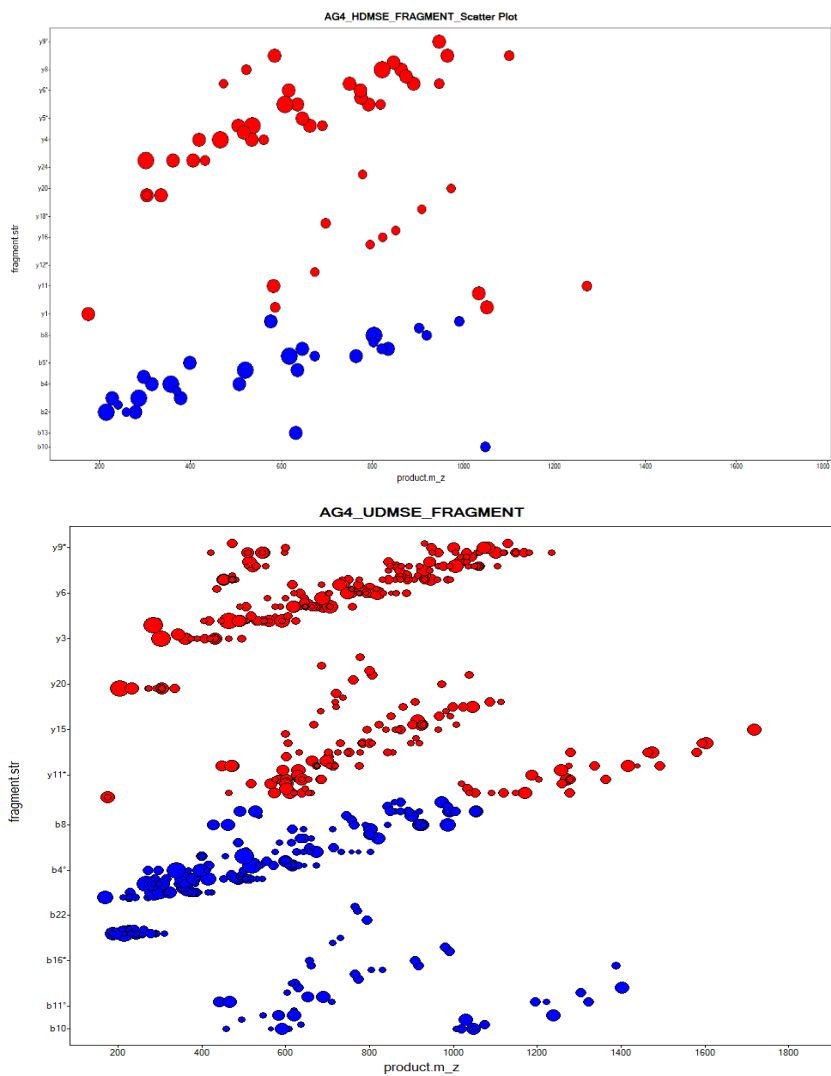


Figura. 9. Distribuição de fragmentação dos íons *b* (vermelho) e *y* (azul) obtidos nas corridas das amostras AG da variedade Ametista em HDMS^E (superior) e UDMS^E (inferior).

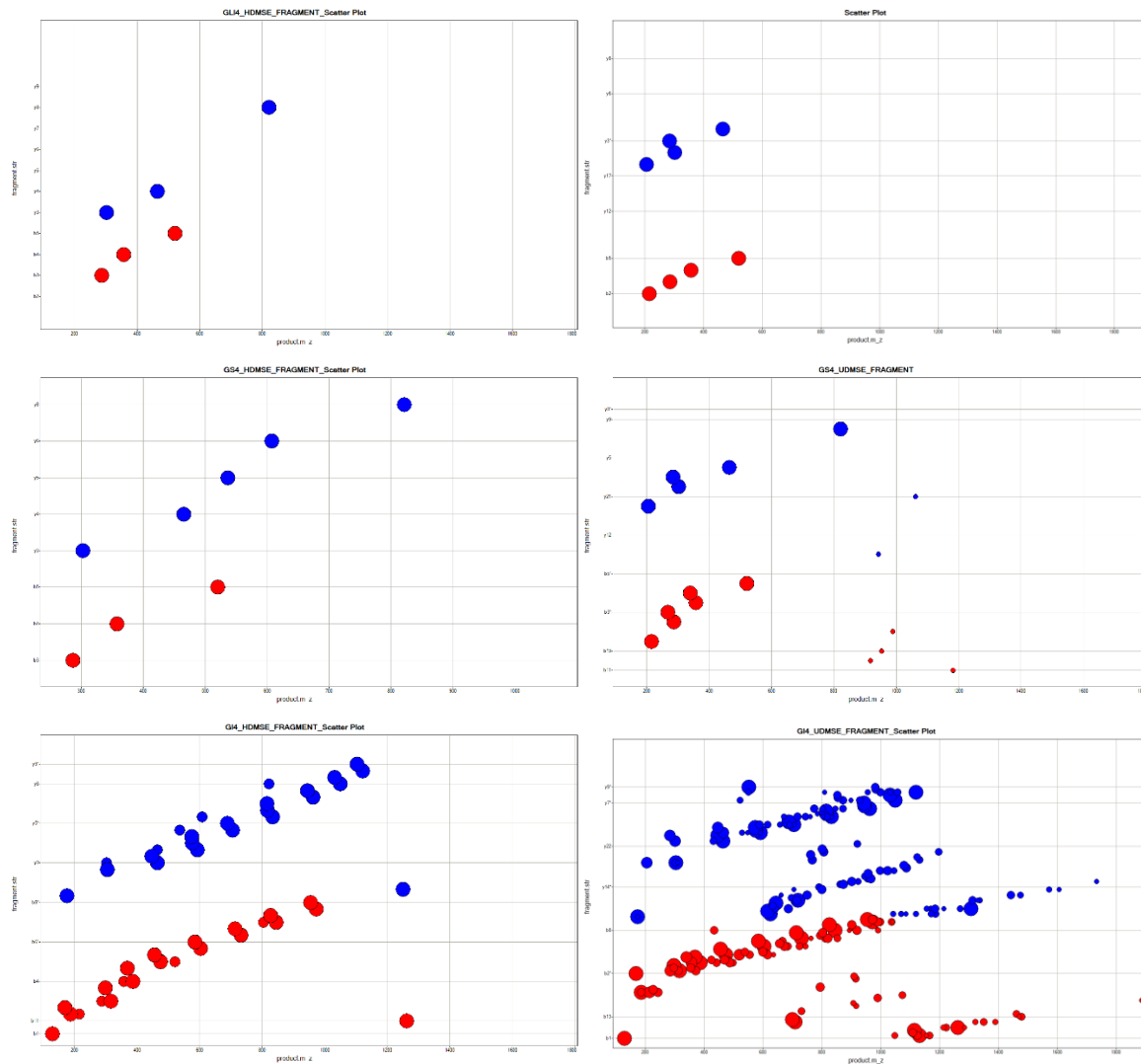


Figura. 10. Distribuição de fragmentação dos íons b (vermelho) e y (azul) obtidos nas corridas das amostras Gli (superior), GS (meio), GI (inferior) da variedade Ametista em HDMS^E (esquerda) e UDMS^E (direita).

A análise global dos dados, após identificação das proteínas realizada pelo software *Progenesis Q1 for Proteomics*, mostrou que um total de 1800 proteínas foram identificadas e quantificadas relativamente com uma média 3 peptídeos por proteína, entre elas, 300 proteínas foram concatenadas em todas as amostras e replicatas (Figura 11). Em relação à identificação de proteínas, 90% foram identificadas por até 6 peptídeos e 75% identificadas por até 3 peptídeos, média do experimento (Figura 12).

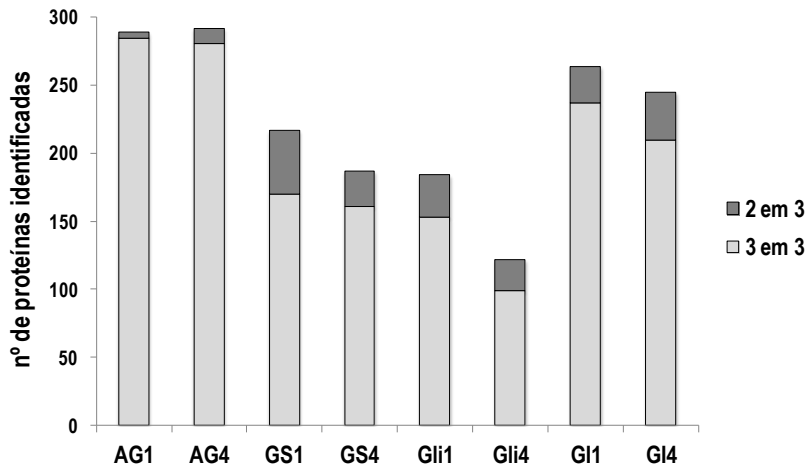


Figura. 11. Total de proteínas identificadas em conjunto em 2 de 3 replicatas e em 3 de 3 replicatas.

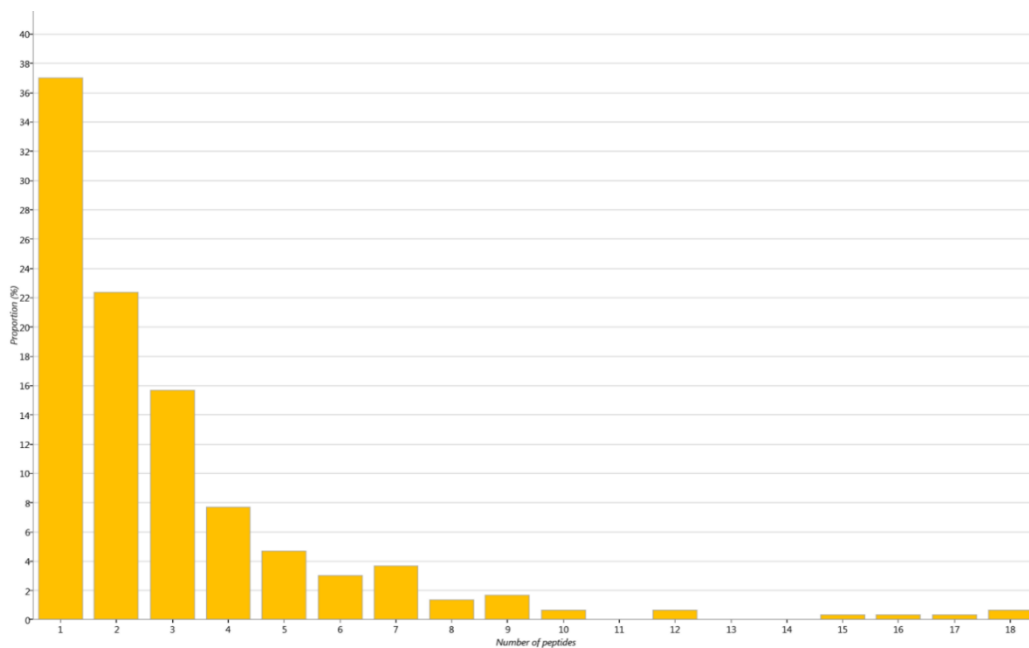


Figura. 12. Proporção (eixo y) do número de peptídeos identificados por proteína (eixo x, 1 a 18).

A relação de proteínas identificadas nos diferentes extratos para as duas amostras (1 e 4) está apresentada pelo diagrama de Venn (Figura 13) evidenciando quantas são exclusivas para cada amostra e quantas são comuns entre elas. Os extratos AG1 e 4 apresentaram 112 proteínas comuns, com uma diferença de mais de 50% de identificação na AG4, onde foram identificadas 84 exclusivamente contra 39 em AG1. Nas proteínas identificadas nas amostras Gli1 e Gli4 foram encontradas 14 proteínas comuns, sendo que nenhuma foi identificada exclusivamente em Gli4, enquanto que Gli1

possui 35 proteínas não presentes em Gli4. Em relação ao número de proteínas identificadas nos extratos GS, 27 foram comuns aos dois extratos GS1 e GS4, 9 exclusivamente identificadas em GS1 e 16 em GS4. Para os extratos GI, observou-se 66 proteínas comuns aos extratos GI1 e GI4 e 29 exclusivas à GI1 e 17 à GI4. Assim, é possível concluir que em média os extratos das proteínas solúveis apresentaram o maior número de proteínas identificadas (120), seguido das gluteninas insolúveis (58), gluteninas solúveis (25) e gliadinas (21).

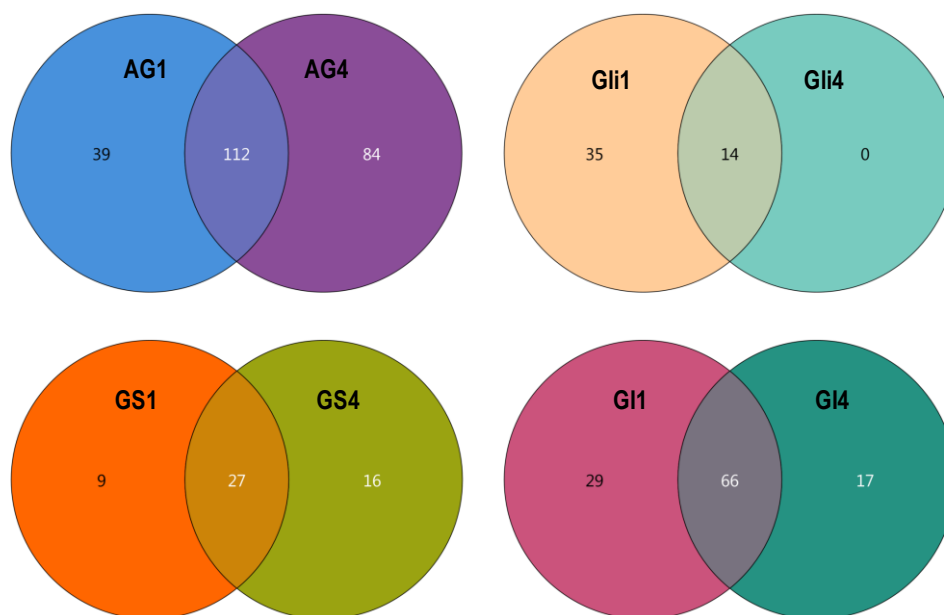


Figura. 13. Diagrama de Venn com todas as proteínas identificadas nos quatro extratos sequenciais para as duas variedades estudadas.

As principais classes de proteínas de reserva identificadas neste trabalho estão apresentadas na Tabela 2. As maiores abundâncias das proteínas identificadas foram encontradas em ambos os extratos de gluteninas insolúveis (GI1, GI4) e as proteínas mais abundantes corresponderam às subunidades de glutenina HMW. Nestes extratos, proteínas particularmente interessantes foram tentativamente identificadas, tais como subunidades de glutenina LMW codificadas no locus *Glu-B3* e subunidades de glutenina HMW (1Ax1, 1Dx2, 1Dy12, 1Dx5, 1Dy10, 1Bx7, 1By8 e 1By9), sendo que algumas delas, como a 1Ax1, 1Dx5 e 1Dy10 são fortemente associadas com variedades de trigo que apresentam alta qualidade relacionadas à força do glúten. Devido à homologia destas proteínas não foi possível discriminar a composição das subunidades das variedades 1

e 4. Este é um problema frequentemente associado na identificação das proteínas do glúten (FERREIRA *et al.*, 2014; ROMBOUTS *et al.*, 2013; VENSEL *et al.*, 2014) e requer a continuação dos estudos, principalmente no que tange a construção dos bancos de dados.

Tabela. 2. Lista das proteínas de reserva identificadas nos diferentes extratos.

Descrição UNIPROT	Número de acesso UNIPROT (descrição)	Nº peptídeos identificados (únicos)	Score Progenesis	ANOVA (p)	Extratos*
High molecular weight glutenin subunit OS=Triticum aestivum GN=glu PE=4 SV=1	A0A060MZP1, A0A059UHD1 (1Ax1), A0MZ38 (1Ax), G3K725, (1Ax1.1), Q03872 (1Ax1), Q19AE4 (A1), Q41553 (Ax2), Q8H0L2 (Ax type), Q8H0L5 (Ax type), Q9SYY0 (A1), T2HRF3 (A1)	12 (10)	92,1589	2,89E-07	AG1, GS1, Gli1, GI1 AG4, GS4, Gli4, GI4
Dy-type high molecular weight glutenin OS=Triticum aestivum GN=1Dy12* PE=4 SV=1	E0W6G5 (1Dy12*), A0A0H4SMF2 (1Dy12), A0A0H4SR12 (1Dy12), A3RF25 (1By9), A3RF26 (1By9), A4URY8 (1Dy12), A9QUS3 (1Dy12), A9YSK3 (D1-2), A9ZMG8 (1Dy10), D2CPI7 (1Dy11), E4W506 (1Dy12.), P10387 (DY10), Q670Q5 (1Dy10), Q8H0L3 (Dy type), Q8H0L6 (Dy type), V9TRL3 (1Dy)	7 (2)	59,5474	0,0001116 56	AG1, GS1, GI1 AG4, GS4, GI4

Descrição UNIPROT	Número de acesso UNIPROT (descrição)	Nº peptídeos identificados (únicos)	Score Progenesis	ANOVA (p)	Extratos*
HMW glutenin subunit 1By9 OS=Triticum aestivum GN=Glu-1By9 PE=4 SV=1	Q03871 (1By9), A0A023GQG9 (Y-type), A0A060AER5 (1By15), A5HMG2 (1By16), B8PSA6 (1By15), Q0Q5D8 (By8), Q4JHY1 (1By15), Q7X6E6 (A1-2), Q7X6P9 (A1-2), Q7X8H1 (A1-2), Q7XZA5 (A1-2), Q7XZB4 (A1-2), Q7XZB8 (A1-2), U5YQV1 (By18), W0C8U3 (1By15)	7 (2)	71,3586	0,0030110 12	AG1, GS1, Gli1, GI1 AG4, GS4, Gli4, GI4
High- molecular- weight glutenin subunit Bx17 OS=Triticum aestivum GN=Glu-B1 PE=4 SV=1	Q18MZ6 (Bx17), A9YSK5 (B1-1), G4Y3Y0 (Bx7.1) G4Y3Y1 (Bx7.2), Q42451 (B1-1b), Q45R38 (Bx7)	2 (1)	13,0559	0,0002485 42	AG1, GS1,GI1 AG4, GS4, GI4
1Bx high molecular weight	Q1KL95 (1Bx), Q7X6V6 (B1-1), Q7XAG6 (B1-1), Q7XAG7 (B1-1), Q7XAH0 (B1-1), Q7XAH2 (B1-1),	3 (1)	14,9456	5,85E-08	GI1 GI4

Descrição UNIPROT	Número de acesso UNIPROT (descrição)	Nº peptídeos identificados (únicos)	Score Progenesis	ANOVA (p)	Extratos*
glutenin subunit OS=Triticum aestivum PE=4 SV=1	Q7XAH9 (B1-1), Q7XAJ1 (B1-1), Q8H0L4 (Bx TYPE), W0C8N8 (1BX14)				
HMW glutenin subunit OS=Triticum aestivum GN=1Dx2.2 PE=4 SV=1	Q8GV12 (1Dx2.2), A0A0H4PMT1 (1Dx2.1t), A9YSK4 (D1-1), B1B520 (D1), C0SUC3 (1DX5), D0IQ05 (DX5), D0IQ07 (DX2), D7REK2 (1DX5), E2CT66 (1DX1.5*), G1E6K7 (Dx5), J7G6L4 (1Dx2.2), Q0Q5D2 (X-type), Q1KL96 (1Dx), P10388 (DX5), Q38LF5 (DX5), Q599I0 (1Dx2.2*), Q6R2V1 (1Dx2.1), X2JUA0 (Dx5)	3 (2)	33,0704	1,95E-06	GI1 GI4
Low molecular weight glutenin OS=Triticum aestivum PE=4 SV=1	C0KK80;A2IBV6;C0KK81;C0KK82;D0EVN9;D0EVP5; D7RFI3;F4YT78;F5A654;F8SGP0;F8SGP3;H6VLP8;H 6VLQ1;H6VLQ2;H6VLQ5;I3XHP8;I3XHP9;I3XHQ1;I3X HQ3;K7WV92;P04730;P10386;P93791;P94021;Q00M 59;Q0GNG0;Q0GQX0;Q0PW08;Q30DX5;Q30DX7;Q5 2NZ3;Q52NZ5;Q52NZ6;Q7DM83;R4JFK6;R9XUE1;R9	5 (2)	28,8791	9,06E-06	AG1, GS1, Gli1, GI1 AG4, GS4, Gli4, GI4

Descrição UNIPROT	Número de acesso UNIPROT (descrição)	Nº peptídeos identificados (únicos)	Score Progenesis	ANOVA (p)	Extratos*
	XUV6;R9XUW4;R9YQY9;R9YUB3;V9P6C8;V9P6E8;V9P6Q7;V9P748;V9P7D2;X2JBS3				
Low molecular weight glutenin subunit OS=Triticum aestivum PE=4 SV=1	I1XB60;A0A060N0C7;A8CA05;B2BZC4;B2BZD0;B2BZD1;B2BZD2;B2Y2Q6;B2Y2Q7;B2Y2R3;B2Y2R4;B2Y2R5;B2Y2R6;B2Y2R8;B2Y2S0;B2Y2S1;B2Y2S2;B2Y2S6;C3VN79;C3VN80;C5IXL3;C8KIL6;D3U318;D3U319;D3U326;D3U328;D3UAL6;D3UAL7;D6RVY4;D6RVY7;F8SGL2;F8SGL3;F8SGL4;F8SGL5;F8SGL6;F8SGL7;F8SGL8;F8SGL9;F8SGM4;F8SGM5;F8SGM7;F8SGM8;F8SGM9;F8SGN3;F8SGN4;F8SGN5;F8SGN6;F8SGN9;F8SGQ3;I1XB41;I1XB51;Q00M55;Q00M56;Q0GNF9;Q0GNG1;Q0QBR3;Q0ZCA8;Q0ZCB0;Q3LGB3;Q41551;Q41552;Q571Q5;Q5MFN4;Q5MFQ0;Q5TLY8;Q5TLY9;Q6J160;Q6J161;Q75ZV9;Q8GU18;Q8H737;Q8W3W3;Q8W3W4;Q8W3W5;Q8W3W6;Q8W3W8;Q8W3W9;Q8W3X0;Q8W3X2;Q8W3X3;Q8W3X4;Q8W3X6;Q9ZNY0;R4JAN5;R4JAN9;R4JAQ0;R4JB48;R4JB52;R4JB53;R4JB56;R4JB62;R4JBF1;R4JBF5;R4JBG2;R4JBK0;R4JBK6;R4JDK6;R4JDM1;R4JDM5;R4JDN1;R4JFA4;	7 (2)	77,4845	2,70E-07	AG1, GS1, Gli1, GI1 AG4, GS4, Gli4, GI4

Descrição UNIPROT	Número de acesso UNIPROT (descrição)	Nº peptídeos identificados (únicos)	Score Progenesis	ANOVA (p)	Extratos*
	R4JFP9;R4JFQ3;R9XT50;R9XVA5;R9XVC9;R9XWG9; U6BEQ9;U6BIP1;V6BPA5;V9P737;V9P785				
LMW glutenin OS=Triticuma estivum GN=Glu PE=4 SV=1	K0I5W1;B2Y2S3;K0HZJ2	6 (1)	71,347	1,57E-06	AG1, GS1, Gli1, GI1 AG4, GS4, Gli4, GI4
Low- molecular- weight glutenin subunit OS=Triticum aestivum GN=LMW-GS PE=4 SV=1	R4JBH8;B3EY91;B8R6A0;C3VN74;C3VN75;C3VN76; C3VN77;D2DII1;D2DII3;D2DII7;D3U317;D3UAL5;F8S GQ4;K7RDA5;M1GL80;M1GMA2;M1GP58;P10385;P9 3794;Q0PW11;Q19MN2;Q19MN3;Q19N55;Q5MFG8;Q 5MFH0;Q5MFH2;Q5MFH3;Q6QGW0;Q6SPY7;Q6SPY 8;Q6SPZ0;Q6SPZ1;Q6SPZ2;Q6SPZ3;Q8W3U9;Q8W3 V0;Q8W3V1;R4JAQ1;R4JAR7;R4JB19;R4JB27;R4JB3 4;R4JB50;R4JBE8;R4JBI2;R4JBJ8;R4JD42;R4JF25;R 4JFL5;R4V1P5;R9XT25;R9XT45;R9XUD0;R9XUF2;R9 XUF6;R9XUU0;R9XV91;R9XV98;R9XVB6;R9XWD5;R 9XWE0;R9XWF9;X2J8E3;X2JAE7	4 (2)	19,4698	0,0001875 78	AG1, GS1, Gli1, GI1 AG4, GS4, Gli4, GI4

* Extratos onde as proteínas foram quantificadas, em negrito o extrato que apresentou maior abundância.

4 CONCLUSÃO

Os métodos multiplexos de análises utilizados nesse estudo provaram que a aplicação da mobilidade iônica nos métodos proteômicos por Espectrometria de Massas favorece a separação de peptídeos devido à melhor especificidade e seletividade da análise. Contudo a aplicação do método de aquisição UDMS^E foi capaz de potencializar ainda mais a identificação de proteínas e peptídeos quando comparado ao método HDMS^E. Isso está diretamente relacionado com a eficácia na eliminação de íons monoprotônicos indesejados e com o aumento da fragmentação favorecida pela energia de colisão quase-específica a cada peptídeo aplicada na fragmentação. Embora não tenha sido possível diferenciar as subunidades proteicas entre as variedades estudadas, esse estudo fornece novas perspectivas de avaliação da qualidade tecnológica de farinha de trigo, contribuindo com a pesquisa científica e desenvolvimento de inovações dentro da cadeia produtiva do cereal no Brasil.

AGRADECIMENTOS

Os autores agradecem à OR Sementes pelo fornecimento das amostras. Este trabalho foi financiado pela UNIRIO, pelo Conselho Nacional do Desenvolvimento Científico e Tecnológico (CNPq) (427116/2018-0), pela Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) (26/010.100988/2018; 26/202/709-2018) e pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (código 001)

PARTE III

**DISCUSSÃO E CONSIDERAÇÕES
FINAIS**

CAPÍTULO 6. DISCUSSÃO

CAPÍTULO 7. CONSIDERAÇÕES FINAIS

CAPÍTULO 6- DISCUSSÃO

6 DISCUSSÃO

6.1 PRINCIPAIS DESAFIOS ENCONTRADOS

O objetivo central deste trabalho foi caracterizar a fração proteica de farinhas de trigo para revelar os diferentes perfis proteicos de amostras com aptidões tecnológicas contrastantes, aplicando proteômica comparativa não dirigida com abordagens DIA.

Realizar um estudo com abordagem proteômica tendo como alvo as proteínas do trigo apresenta inúmeros desafios que são decorrentes das diferentes etapas necessárias, tais como extração, digestão, identificação e quantificação das proteínas. Iniciando pela etapa de extração, na segunda parte deste manuscrito, foram apresentados dois trabalhos nos quais a extração foi realizada considerando a solubilidade das proteínas segundo a classificação de Osborne (1907). No capítulo 3, o alvo do trabalho foi a caracterização das proteínas de reserva, por isso, foi realizada a etapa de depleção das proteínas metabólicas. Foi possível observar a dificuldade de obtenção de um fracionamento efetivo das frações proteicas pois, mesmo após depleção, muitas proteínas metabólicas foram identificadas na análise proteômica.

No capítulo 5, as proteínas foram fracionadas nas principais classes com obtenção 4 extratos: albuminas/globulinas, gliadinas, gluteninas solúveis e gluteninas insolúveis. Novamente verificou-se a dificuldade de fracionamento, já que proteínas de reserva foram identificadas em todos os extratos, no entanto, como esperado, com maior abundância nos extratos de gluteninas insolúveis. A presença de proteínas metabólicas nos extratos de proteínas de reserva e *vice-versa*, pode ser explicada pelas diferentes interações proteicas existentes (FERREIRA *et al.*, 2014; GENEIX *et al.*, 2020; KHAN *et al.*, 2021; LESAGE *et al.*, 2012; LV *et al.*, 2021). A interação proteica, principalmente por meio de ligações cruzadas do tipo covalente como as pontes dissulfeto, modifica a solubilidade destas proteínas, fazendo com que monômeros (contendo apenas pontes dissulfeto intramoleculares) das proteínas de reserva sejam solúveis e estejam presentes na fração das proteínas metabólicas. Do mesmo modo, a interação das proteínas metabólicas com as proteínas de reserva por

meio de ligações dissulfeto intermoleculares explica a presença destas proteínas mesmo após a etapa de lavagem ou depleção

No capítulo 4, foi realizada uma extração total de proteínas com utilização de uma solução tampão que fosse apropriada para ambas as frações, prolaminas (proteínas de reserva) e não-prolaminas (metabólicas). Para isso, foi utilizado o tampão Tris-HCl com detergente (SDS), agente alquilante (IAM) e agente redutor (DTT). Como esperado, foi possível identificar proteínas de todas as frações em uma única etapa de extração, sem impacto no número de proteínas identificadas.

Seguindo com a etapa de digestão e, objetivando superar a dificuldade em conseguir peptídeos de tamanho adequado para análise proteômica a partir das proteínas do trigo, em especial das proteínas de reserva, que possuem muito resíduos de prolina e glutamina e poucos de lisina e arginina, gerando clivagens perdidas que dificultam a identificação (VENSEL *et al.*, 2011), foi testada a utilização de duas enzimas proteolíticas: tripsina e tripsina/Lys-C. A tripsina é específica para clivagem C-terminal de resíduos de lisina e arginina e a tripsina/Lys-C melhora a eficiência proteolítica nos sítios de lisina. Assim, conforme esperado, com o uso da tripsina/Lys-C obteve-se um menor percentual de clivagens perdidas (Fig. 1), confirmando a maior eficiência proteolítica em relação à tripsina. Alguns estudos apontam que a utilização de quimotripsina pode gerar maior identificação de peptídeos quando comparado ao uso de tripsina (FIEDLER *et al.*, 2014; VENSEL *et al.*, 2011), mas o uso de tripsina/Lys-C gerou maior número de proteínas identificadas quando comparada aos extratos provenientes da digestão com quimotripsina e Glu-C (VINCENT *et al.*, 2022). Apesar de ser mais específica do que a quimotripsina, a tripsina/Lys-C continua sendo o padrão ouro em análises proteômicas e, por apresentar menor custo, é ideal para análises em larga escala.

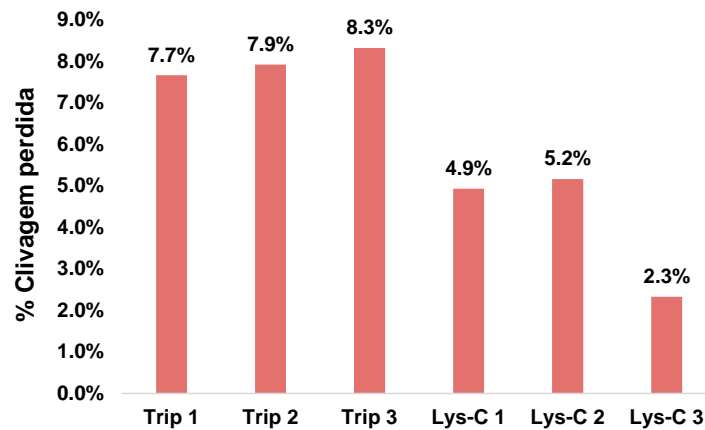


Fig. 1. Percentual de clivagem perdida comparando extratos digeridos com tripsina (Trip) e tripsina/Lys-C (Lys-C) em triplicata.

Diferentes métodos de aquisição independente de dados foram selecionados e aplicados para superar os desafios inerentes da análise proteômica do trigo. O método que apresentou os melhores resultados foi o UDMS^E que apresentou maior identificação de peptídeos e proteínas (60%) com intervalo de confiança de 95%. Tal resultado foi atribuído à significativa melhora na fragmentação dos íons precursores quando comparado ao método HDMS^E. Limitações de acesso aos recursos técnicos necessários não permitiram a utilização do método UDMS^E em todos os experimentos deste trabalho, no entanto os demais métodos que utilizam abordagem independentes de dados, MS^E (capítulo 3) e HDMS^E (capítulo 4), este último com sistema de mobilidade iônica, mostraram-se bastante efetivos para identificação das proteínas do trigo, comprovado pela faixa dinâmica com mais de 6 ordens de magnitude ($\log_{10} > 6$) (Fig. 2). Pelo método HDMS^E foi obtido maior número de identificação em cada amostra (a partir de extratos totais de proteínas) quando comparado com o método MS^E (a partir de extratos de proteínas de reserva).

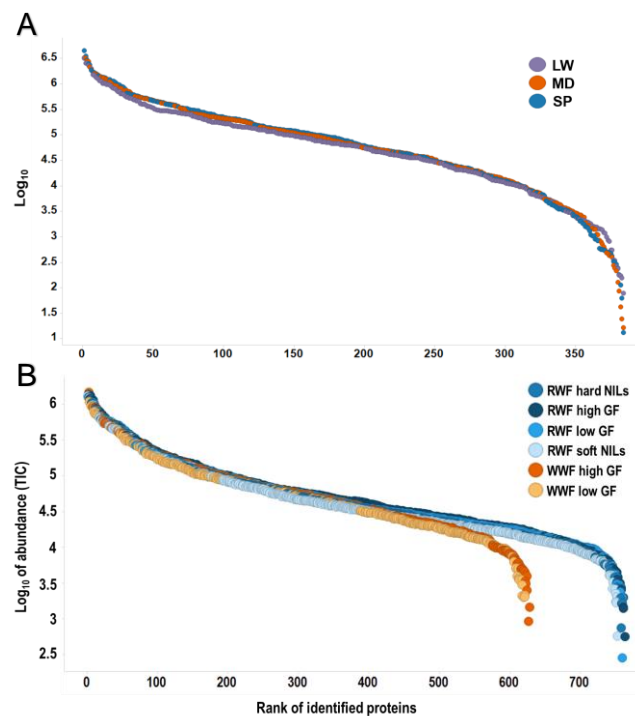


Fig. 2. Faixas dinâmicas obtidas com os métodos (A) MS^E, para proteínas de reserva e (B) HDMS^E, para proteínas totais.

6.2 ESTRATÉGIAS UTILIZADAS PARA CARACTERIZAÇÃO DO PERFIL PORTEICO DAS AMOSTRAS DE DIFERENTES APTIDÕES TECNOLÓGICAS

A comparação dos resultados obtidos em cada análise proteômica com o banco de dados gerou uma lista de proteínas identificadas nas amostras. A principal estratégia utilizada foi a determinação e avaliação de proteínas que se apresentavam diferencialmente abundantes (DAP) nas amostras². Para isso, as comparações foram feitas em pares com a premissa de que a proteína deveria coexistir nas duas amostras e ter sido quantificada nas três replicatas de cada amostra. De modo a garantir a confiabilidade dos resultados, foram aplicados critérios rigorosos considerando a quantidade mínima de peptídeos exclusivos e de outros peptídeos para identificação e pontos de corte como: $p < 0,05$, Log_2 da abundância $\pm 0,6$ e $\text{CV} < 30\%$.

No entanto, uma grande parte das proteínas, em torno de 65% das DAP (capítulo 3), não possuíam anotação funcional no banco de dados. Realizamos então, uma predição de domínios, onde os códigos FASTA destas proteínas foram verificados em

² Consideramos diferencialmente abundantes e diferencialmente acumuladas com o mesmo significado: proteínas que após aplicação de critérios rigorosos apresentaram-se superexpressas ou subexpressas nas amostras.

três bancos de dados de família de proteínas (InterPro, Pfam e Prosite) em busca dos domínios relacionados. Com isso, conseguimos obter um melhor entendimento da provável função das proteínas. Essa estratégia teve grande impacto nos resultados, principalmente visualizado na análise de mapa de calor, onde as amostras de alta e média força do glúten foram agrupadas, enquanto que no mapa de calor feito apenas com os 35% das DAP que já estavam anotadas, as amostras de média e baixa força do glúten foram agrupadas.

No capítulo 4, a categorização por ontologia genética foi realizada para caracterizar as proteínas identificadas e quantificadas. Nesta etapa, as proteínas foram classificadas quanto à sua função molecular, processo biológico envolvido e o componente celular de localização. As informações obtidas também foram utilizadas para melhor compreensão dos processos biológicos nos quais DAP e proteínas exclusivas estavam envolvidas. A partir dos dados obtidos, foi possível caracterizar os diferentes perfis proteicos das amostras com diferentes aptidões tecnológicas. Uma síntese dos resultados obtidos pelas abordagens proteômicas utilizadas neste trabalho, bem como um índice de representações ilustrativas onde foram evidenciadas diferenças no perfil proteico das amostras está apresentado na Tabela 1.

Tabela: 1. Resumo das abordagens proteômicas utilizadas para caracterizar proteínas do trigo.

Informação	Artigo 1 (cap.3)	Artigo 2 (cap.4)	Artigo 3 (cap.5)
Alvo do estudo	subproteoma	proteoma	proteoma
Fração proteica	Proteínas de reserva	Proteínas metabólicas e de reserva	Proteínas metabólicas e de reserva
Tipo de extração	sequencial	total	sequencial
Técnica utilizada	NanoUPLC-MS ^E	NanoUPLC-HDMS ^E	NanoUPLC-HDMS ^E e UDMS ^E
Objetivo	Comparar amostras agrupadas pela aptidão tecnológica	Comparar amostras de diferentes aptidões tecnológicas	Comparar métodos de análise
Diferentes perfis de abundância diferencial	Fig. 3; Tabela 2	Fig. 3; Tabela 1; material suplementar S8-S10	-
Lista de proteínas exclusivas	-	Tabela 1	-
Diferentes domínios	Fig. 4; Material suplementar S3	-	-
Classificação biológica das proteínas diferenciais	-	Fig.4	-
Análise quimiométrica	Fig. 2	Fig.5	-

6.3 PRINCIPAIS RESULTADOS

Na tabela 2 estão apresentados os objetivos com os principais resultados alcançados nesta tese.

Tabela: 2. Principais resultados ligados aos objetivos alcançados.

OBJETIVO	PRINCIPAIS RESULTADOS	VALORIZAÇÃO
I. Revelar os diferentes perfis proteicos de amostras de farinha de trigo comum evidenciando as proteínas que contribuem para a aptidão tecnológica e uso final da farinha	LMW-GS, γ -gliadinas e α -gliadinas encontradas superexpressas nas amostras de alta força do glúten; Puroindolinas e GSP encontradas superexpressas nas amostras de baixa força do glúten; LMW-GS, HMW-GS e <i>small</i> HSPs: possíveis marcadores específicos das farinhas de alta força do glúten e de fenótipo <i>hard</i> ; PINA, PINB2 e GSP: possíveis marcadores das farinhas de baixa força do glúten;	Artigos mencionados abaixo.
II. Identificar e quantificar proteínas do glúten em amostras agrupadas por diferentes aptidões tecnológicas distinguindo os perfis proteicos	1.297 proteínas identificadas e quantificadas somando todas as amostras; Faixa dinâmica superior a 6 ordens de magnitude; 116 proteínas diferencialmente abundantes (DAP); PCA agrupou amostras de alta força do glúten separadas das de baixa força do glúten; Mapa de calor das DAP separou a amostra de alta força do glúten das de média e baixa; Mapa de calor após predição de domínios agrupou amostras de maior força do glúten; A caracterização de proteínas do glúten em amostras agrupadas por diferentes aptidões tecnológicas revelou possíveis marcadores proteicos.	1 artigo original publicado, 1 pôster

OBJETIVO	PRINCIPAIS RESULTADOS	VALORIZAÇÃO
<p>III. Revelar e avaliar o perfil proteico de amostras de farinha de trigo com força do glúten contrastantes (farinhas integrais e refinadas) e de farinhas refinadas de linhagens quase-isogênicas com fenótipos diferentes</p>	<p>Total de 973 proteínas identificadas e quantificadas;</p> <p>Faixa dinâmica > 6 ordens de magnitude; 247 DAP com filtro e 273 sem filtro de CV.</p> <p>Amostras de alta força do glúten e fenótipo <i>hard</i> apresentaram maior número de proteínas, de DAP e maior diversidade de processos biológicos quando comparados com seus pares de baixa força do glúten e fenótipo <i>soft</i>;</p> <p>Análise quimiométrica agrupou amostras de alta força do glúten e com textura <i>hard</i>;</p> <p>A caracterização do perfil proteico das amostras revelou possíveis marcadores proteicos relacionados à aptidão tecnológica.</p>	<p>1 artigo original em preparação para ser submetido</p>
<p>IV. Apresentar e comparar métodos de aquisição de alta definição em NanoUPLC-MS/MS que permitem caracterização das proteínas do trigo</p>	<p>Superioridade do método UDMS^E em comparação ao HDMS^E;</p> <p>Total de 1.800 proteínas identificadas com os dois métodos;</p> <p>UDMS^E melhorou em 60% a identificação de peptídeos e proteínas (intervalo de confiança de 95%);</p> <p>UDMS^E melhorou a fragmentação dos precursores, principalmente em íons com <i>m/z</i> superior a 1.200;</p> <p>UDMS^E potencializou a identificação pela eliminação de íons monoprotonados e pela melhor fragmentação com energia de colisão <i>quasi</i>-específica à relação <i>m/z</i> de cada peptídeo.</p>	<p>1 capítulo publicado em coletânea nacional</p>

CAPÍTULO 7 – CONSIDERAÇÕES FINAIS

O trigo é uma importante fonte de carboidratos e de proteínas na alimentação humana. Além disso, o teor de fibras, vitaminas, minerais e compostos fenólicos também devem ser considerados, uma vez que este cereal apresenta amplo consumo mundial. A versatilidade de uso do trigo em diversos produtos como biscoitos, bolos, massas, pães e outros produtos de confeitaria consolidam esse cenário e a aptidão tecnológica do grão é justamente o fator que define o uso final do trigo.

Sabe-se que a aptidão tecnológica é influenciada pelas proteínas presentes no grão e grande esforço tem sido realizado ao longo de décadas no intuito de caracterizar as proteínas e compreender a participação delas nas características tecnológicas. A maior parte dos estudos reportados na literatura, principalmente envolvendo a caracterização do trigo brasileiro, basearam-se em análises físico-químicas e reológicas, bem como na determinação da relação do teor proteico total, da razão entre as frações proteicas (metabólicas e de reserva e entre suas subunidades) e na forma com que as proteínas se agregam em polímeros.

Neste sentido, o estudo da determinação do perfil proteico relacionado à aptidão tecnológica do trigo é fundamental e desperta o interesse na aplicação de técnicas altamente eficientes baseadas em métodos otimizados como os que são utilizados em análises proteômicas atualmente. Diante do exposto, neste trabalho aplicou-se a cromatografia líquida de ultra performance acoplada à espectrometria de massas de alta definição usando métodos MS^E para caracterizar de forma abrangente o perfil proteico de amostras de farinha de trigo comum com diferentes aptidões tecnológicas.

Foi possível diferenciar o perfil proteico das amostras com a análise de abundância diferencial, bem como com dados de categorização biológica de proteínas disponíveis em bancos de dados. Verificou-se diferença no perfil proteico tanto para a fração de proteínas metabólicas quanto das proteínas de reserva entre as diferentes amostras classificadas previamente pelas características reológicas. As amostras de alta força do glúten apresentaram um maior número de DAP, diferentes domínios funcionais e maior diversidade de processos biológicos do que as amostras com baixa força do glúten. Os achados confirmaram a primeira hipótese que norteou o trabalho

de que existem diferenças no perfil proteico que explicam as diferentes aptidões tecnológicas das amostras.

A hipótese de que a caracterização do perfil proteico de farinhas de trigo pela utilização de abordagem proteômica pode revelar possíveis marcadores proteicos relacionados à aptidão tecnológica do trigo também foi confirmada, pois algumas DAP encontradas nas amostras foram indicadas como possíveis marcadores específicos para determinada aptidão tecnológica, com LMW-GS, HMW-GS e *small* HSPs apontadas como possíveis marcadores das farinhas de alta força do glúten e de fenótipo hard, bem como PINA, PINB2 e GSP como possíveis marcadores das farinhas de baixa força do glúten. Esses resultados contribuem tanto para os avanços na área da química de cereais quanto no campo da melhoria do trigo.

As análises quimiométricas foram fundamentais para análise e distinção de amostras de diferentes aptidões tecnológicas corroborando os resultados encontrados nas demais estratégias utilizadas. Foi possível concluir que os perfis proteicos diferentes permitiram a distinção entre as amostras de alta força do glúten daquelas de baixa força do glúten.

Este trabalho destacou métodos altamente eficientes que permitiram a identificação e quantificação de proteínas do trigo, com especial atenção para o método UDMS^E que mostrou superioridade em termos de fragmentação e identificação de peptídeos e proteínas quando comparado com o método HDMS^E. Juntos, os resultados apontaram que a abordagem utilizada foi capaz de gerar resultados valiosos e abrangentes que permitiram avançar na compreensão do perfil proteico atrelado à aptidão tecnológica.

Pela temática abordada e utilização de técnicas excelentes para alcançar os objetivos, conclui-se que esse trabalho é de grande relevância em termos de estudos na cadeia produtiva do trigo e tem sua importância justificada.

Como perspectivas, buscando aprofundar o conhecimento sobre as proteínas com impacto na aptidão tecnológica, principalmente no trigo brasileiro, o estudo de cultivares provenientes de outras áreas tritícolas do país buscando confirmar a presença das proteínas apontadas como possíveis biomarcadores de qualidade torna-se interessante. No que tange às proteínas de reserva, o estudo da formação dos polímeros ao longo do desenvolvimento do grão nas amostras de aptidões tecnológicas contrastantes também é promissor.

REFERÊNCIAS

- AACC. Approved methods. *In*. 10th ed. Minneapolis, MN: American Association of Cereal Chemists, 2000.
- ABITRIGO. **Associação Brasileira das Indústrias de Trigo**. Brasil, 2022. Disponível em: <http://www.abitrigo.com.br/estatisticas-abitrigo/trigo-e-a-farinha-no-mundo/>. Acesso em: 26 de março de 2022.
- AEBERSOLD, R.; MANN, M. Mass-spectrometric exploration of proteome structure and function. **Nature**, 537, n. 7620, p. 347-355, 2016.
- AFZAL, M.; PFANNSTIEL, J.; ZIMMERMANN, J.; BISCHOFF, S. C. *et al*. High-resolution proteomics reveals differences in the proteome of spelt and bread wheat flour representing targets for research on wheat sensitivities. **Scientific Reports**, 10, n. 1, p. 1-11, 2020.
- AFZAL, M.; SIELAFF, M.; CURELLA, V.; NEERUKONDA, M. *et al*. Characterization of 150 Wheat Cultivars by LC-MS-Based Label-Free Quantitative Proteomics Unravels Possibilities to Design Wheat Better for Baking Quality and Human Health. **Plants**, 10, n. 3, p. 424, 2021.
- AGHAGHOLIZADEH, R.; KADIVAR, M.; NAZARI, M.; MOUSAVI, F. *et al*. Characterization of wheat gluten subunits by liquid chromatography–Mass spectrometry and their relationship to technological quality of wheat. **Journal of Cereal Science**, 76, p. 229-235, 2017.
- ALTENBACH, S. B.; TANAKA, C. K.; WHITEHAND, L. C.; VENSEL, W. H. Effects of post-anthesis fertilizer on the protein composition of the gluten polymer in a US bread wheat. **Journal of Cereal Science**, 68, p. 66-73, 2016.
- ALVES, T. O.; D'ALMEIDA, C. T. S.; FERREIRA, M. S. L. Determination of Gluten Peptides Associated with Celiac Disease by Mass Spectrometry. *In*: **Celiac Disease and Non-Celiac Gluten Sensitivity**: InTech, 2017.
- ALVES, T. O.; D'ALMEIDA, C. T. S.; SCHERF, K. A.; FERREIRA, M. S. L. Modern approaches in the identification and quantification of immunogenic peptides in cereals by LC-MS/MS. **Frontiers in Plant Science**, 10, 2019.
- ALVES, T. O.; D'ALMEIDA, C. T. S.; VICTORIO, V. C. M.; SOUZA, G. H. M. F. *et al*. Immunogenic and allergenic profile of wheat flours from different technological qualities revealed by ion mobility mass spectrometry. **Journal of Food Composition and Analysis**, 73, p. 67-75, 2018/10/01/ 2018.
- ARENA, S.; D'AMBROSIO, C.; VITALE, M.; MAZZEO, F. *et al*. Differential representation of albumins and globulins during grain development in durum wheat and its possible functional consequences. **Journal of Proteomics**, 162, p. 86-98, 6/6/ 2017.
- BARAK, S.; MUDGIL, D.; KHATKAR, B. S. Relationship of gliadin and glutenin proteins with dough rheology, flour pasting and bread making performance of wheat varieties. **LWT - Food Science and Technology**, 51, n. 1, p. 211-217, 4// 2013.
- BÉKÉS, F. New aspects in quality related wheat research: 1. Challenges and achievements. **Cereal Research Communications**, 40, n. 2, p. 159-184, 2012.

BEOM, H.-R.; KIM, J. S.; JANG, Y.-R.; LIM, S.-H. *et al.* Proteomic analysis of low-molecular-weight glutenin subunits and relationship with their genes in a common wheat variety. **3 Biotech**, 8, n. 1, p. 56, 2018.

BORÉM, A.; SCHEEREN, P. L. **Trigo: do plantio à colheita**. Viçosa, MG: Ed. UFV, 2015. 260p., 2015. (Embrapa Trigo-Livro técnico (INFOTECA-E). 8572695222.

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

BRASIL. Instrução normativa nº 8, 03 jun. 2005. Regulamento Técnico de Identidade e Qualidade da Farinha de Trigo. MINISTÉRIO DE AGRICULTURA, P. e. A. : Diário Oficial da República Federativa do Brasil 2005.

BRASIL. Instrução Normativa nº 38 de 30 de novembro de 2010. MINISTÉRIO DA AGRICULTURA, P. e. A. : Diário Oficial da União 2010.

BROMILOW, S.; GETHINGS, L. A.; BUCKLEY, M.; BROMLEY, M. *et al.* A curated gluten protein sequence database to support development of proteomics methods for determination of gluten in gluten-free foods. **Journal of Proteomics**, 163, n. Supplement C, p. 67-75, 2017/06/23/ 2017a.

BROMILOW, S. N. L.; GETHINGS, L. A.; LANGRIDGE, J. I.; SHEWRY, P. R. *et al.* Comprehensive Proteomic Profiling of Wheat Gluten Using a Combination of Data-Independent and Data-Dependent Acquisition. **Frontiers in Plant Science**, 7, p. 2020, 2017b.

BURRIEZA, H. P.; RIZZO, A. J.; VALE, E. M.; SILVEIRA, V. *et al.* Shotgun proteomic analysis of quinoa seeds reveals novel lysine-rich seed storage globulins. **Food chemistry**, 293, p. 299-306, 2019.

CHAPMAN, J. D.; GOODLETT, D. R.; MASSELON, C. D. Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. **Mass spectrometry reviews**, 33, n. 6, p. 452-470, 2014.

CLAVIJO, B. J.; VENTURINI, L.; SCHUDOMA, C.; ACCINELLI, G. G. *et al.* An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. **Genome research**, 27, n. 5, p. 885-896, 2017.

COLGRAVE, M. L.; GOSWAMI, H.; BYRNE, K.; BLUNDELL, M. *et al.* Proteomic profiling of 16 cereal grains and the application of targeted proteomics to detect wheat contamination. **Journal of Proteome Research**, 14, n. 6, p. 2659-2668, 2015. Article.

CONAB. Acomp. safra bras. grãos. Brasília. v. 7 - Safra 2019/20 - Oitavo levantamento, p. 1-31, maio 2020.

CONAB. Acomp. safra brasileira de grãos. Brasília. v.9 - Safra 2021/22, n.5 - Quinto levantamento, p. 1-101, fevereiro 2022. 2022a.

CONAB. Acomp. safra brasileira de grãos. Brasília: COMPANHIA NACIONAL DE ABASTECIMENTO. v. 9, safra 2021/22, n. 4 - Quarto levantamento, janeiro. 2022. 2022b.

- COSTA, M. d. G. d.; SOUZA, E. L. d.; STAMFORD, T. L. M.; ANDRADE, S. A. C. Qualidade tecnológica de grãos e farinhas de trigo nacionais e importados. **Food Science and Technology**, 28, n. 1, p. 220-225, 2008.
- COSTA, M. S.; SCHOLZ, M. B. d. S.; FRANCO, C. M. L. Effect of high and low molecular weight glutenin subunits, and subunits of gliadin on physicochemical parameters of different wheat genotypes. **Food Science and Technology (Campinas)**, 33, p. 163-170, 2013.
- COSTA, M. S.; SCHOLZ, M. B. d. S.; MIRANDA, M. Z.; FRANCO, C. M. L. Effect of glutenin subunits on the baking quality of Brazilian wheat genotypes. **Bragantia**, 76, n. 1, p. 11-22, 2017.
- CUNSOLO, V.; MUCCILLI, V.; SALETTI, R.; FOTI, S. Mass spectrometry in food proteomics: a tutorial. **Journal of Mass Spectrometry**, 49, n. 9, p. 768-784, 2014.
- D'OVIDIO, R.; MASCI, S. The low-molecular-weight glutenin subunits of wheat gluten. **Journal of Cereal Science**, 39, n. 3, p. 321-339, 2004.
- DAHESH, M.; BANC, A.; DURİ, A.; MOREL, M.-H. *et al.* Polymeric Assembly of Gluten Proteins in an Aqueous Ethanol Solvent. **The Journal of Physical Chemistry B**, 118, n. 38, p. 11065-11076, 2014/09/25 2014.
- DE BRIER, N.; GOMAND, S. V.; CELUS, I.; COURTIN, C. M. *et al.* Extractability and Chromatographic Characterization of Wheat (*Triticum aestivum* L.) Bran Protein. **Journal of food science**, 80, n. 5, p. C967-C974, 2015.
- DE SANTIS, M. A.; CUNSOLO, V.; GIULIANI, M. M.; DI FRANCESCO, A. *et al.* Gluten proteome comparison among durum wheat genotypes with different release date. **Journal of Cereal Science**, 96, p. 103092, 2020.
- DELCOUR, J. A.; JOYE, I. J.; PAREYT, B.; WILDERJANS, E. *et al.* Wheat Gluten Functionality as a Quality Determinant in Cereal-Based Food Products. **Annual Review of Food Science and Technology, Vol 3**, 3, p. 469-492, 2012. Review; Book Chapter.
- DHAKA, V.; KHATKAR, B. S. 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, n. 2, p. 71-82, 2015.
- DISTLER, U.; KUHAREV, J.; NAVARRO, P.; LEVIN, Y. *et al.* Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. **Nature Methods**, 11, n. 2, p. 167-170, 2014.
- DISTLER, U.; KUHAREV, J.; NAVARRO, P.; TENZER, S. Label-free quantification in ion mobility-enhanced data-independent acquisition proteomics. **Nat. Protocols**, 11, n. 4, p. 795-812, 04/print 2016. Protocol.
- DONG, K.; GE, P.; MA, C.; WANG, K. *et al.* Albumin and globulin dynamics during grain development of elite Chinese wheat cultivar Xiaoyan 6. **Journal of Cereal Science**, 56, n. 3, p. 615-622, 11// 2012.
- DOS SANTOS-DONADO, P. R.; DONADO-PESTANA, C. M.; KAWAHARA, R.; ROSA-FERNANDES, L. *et al.* Comparative analysis of the protein profile from biofortified cultivars of quality protein maize and conventional maize by gel-based and gel-free proteomic approaches. **LWT**, 138, p. 110683, 2021.

- DUPONT, F. M.; VENSEL, W. H.; TANAKA, C. K.; HURKMAN, W. J. *et al.* Deciphering the complexities of the wheat flour proteome using quantitative two-dimensional electrophoresis, three proteases and tandem mass spectrometry. **Proteome Science**, 9, n. 1, p. 1, 2011.
- FAN, W.; ZHENG, H.; WANG, G. Proteomic analysis of ubiquitinated proteins in maize immature kernels. **Journal of Proteomics**, 243, p. 104261, 2021.
- FAO. Crop Prospects and Food Situation – Quarterly Global Report No. 1. March 2022. Rome 2022.
- FEILLET, P. *In*: INRA (Ed.). **Le grain de blé. Composition et utilisation**. Paris, 2000.
- FERREIRA, M. S. L.; MANGAVEL, C.; ROGNIAUX, H.; BONICEL, J. *et al.* A MALDI-TOF based study of the in-vivo assembly of glutenin polymers of durum wheat. **Food Research International**, 63, p. 89-99, 2014.
- FIEDLER, K. L.; MCGRATH, S. C.; CALLAHAN, J. H.; ROSS, M. M. Characterization of Grain-Specific Peptide Markers for the Detection of Gluten by Mass Spectrometry. **Journal of Agricultural and Food Chemistry**, 62, n. 25, p. 5835-5844, 2014/06/25 2014.
- FINN, R. D.; ATTWOOD, T. K.; BABBITT, P. C.; BATEMAN, A. *et al.* InterPro in 2017—beyond protein family and domain annotations. **Nucleic acids research**, 45, n. D1, p. D190-D199, 2017.
- FINN, R. D.; COGGILL, P.; EBERHARDT, R. Y.; EDDY, S. R. *et al.* The Pfam protein families database: towards a more sustainable future. **Nucleic acids research**, 44, n. D1, p. D279-D285, 2016.
- FINNIE, C.; SULTAN, A.; GRASSER, K. D. From protein catalogues towards targeted proteomics approaches in cereal grains. **Phytochemistry**, 72, n. 10, p. 1145-1153, 2011.
- GAO, L.; WANG, A.; LI, X.; DONG, K. *et al.* Wheat quality related differential expressions of albumins and globulins revealed by two-dimensional difference gel electrophoresis (2-D DIGE). **Journal of proteomics**, 73, n. 2, p. 279-296, 2009.
- GENEIX, N.; DALGALARRONDO, M.; TASSY, C.; NADAUD, I. *et al.* Relationships between puroindoline A-prolamin interactions and wheat grain hardness. **PloS one**, 15, n. 9, p. e0225293, 2020.
- GEROMANOS, S. J.; VISSERS, J. P. C.; SILVA, J. C.; DORSCHER, C. A. *et al.* The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. **Proteomics**, 9, n. 6, p. 1683-1695, 2009.
- GILLET, L. C.; LEITNER, A.; AEBERSOLD, R. Mass spectrometry applied to bottom-up proteomics: entering the high-throughput era for hypothesis testing. **Annual Review of Analytical Chemistry**, 9, p. 449-472, 2016.
- GOESAERT, H.; BRIJS, K.; VERAVERBEKE, W. S.; COURTIN, C. M. *et al.* Wheat flour constituents: how they impact bread quality, and how to impact their functionality. **Trends in Food Science & Technology**, 16, n. 1–3, p. 12-30, 1// 2005.

- GRAZIANO, S.; MARANDO, S.; PRANDI, B.; BOUKID, F. *et al.* Technological quality and nutritional value of two durum wheat varieties depend on both genetic and environmental factors. **Journal of agricultural and food chemistry**, 67, n. 8, p. 2384-2395, 2019.
- GUO, G.; LV, D.; YAN, X.; SUBBURAJ, S. *et al.* Proteome characterization of developing grains in bread wheat cultivars (*Triticum aestivum* L.). **BMC Plant Biol**, 12, p. 147, Aug 19 2012.
- GUO, J.; QU, L.; HU, Y.; LU, W. *et al.* Proteomics reveals the effects of drought stress on the kernel development and starch formation of waxy maize. **BMC plant biology**, 21, n. 1, p. 1-14, 2021.
- GUTKOSKI, L. C.; DURIGON, A.; MAZZUTTI, S.; SILVA, A. C. T. d. *et al.* Efeito do período de maturação de grãos nas propriedades físicas e reológicas de trigo. **Ciência e Tecnologia de Alimentos**, 28, n. 4, p. 888-894, 2008.
- HEINZE, K.; KISZONAS, A. M.; MURRAY, J. C.; MORRIS, C. F. *et al.* Puroindoline genes introduced into durum wheat reduce milling energy and change milling behavior similar to soft common wheats. **Journal of Cereal Science**, 71, p. 183-189, 9// 2016.
- HEMERY, Y.; ROUAU, X.; LULLIEN-PELLERIN, V.; BARRON, C. *et al.* Dry processes to develop wheat fractions and products with enhanced nutritional quality. **Journal of Cereal Science**, 46, n. 3, p. 327-347, 2007.
- HILL, K.; HORVÁTH-SZANICS, E.; HAJÓS, G.; KISS, É. Surface and interfacial properties of water-soluble wheat proteins. **Colloids and Surfaces A: Physicochemical and Engineering Aspects**, 319, n. 1, p. 180-187, 2008.
- HURKMAN, W. J.; VENSEL, W. H.; TANAKA, C. K.; WHITEHAND, L. *et al.* Effect of high temperature on albumin and globulin accumulation in the endosperm proteome of the developing wheat grain. **Journal of Cereal Science**, 49, n. 1, p. 12-23, 1// 2009.
- IGREJAS, G.; GABORIT, T.; OURY, F.-X.; CHIRON, H. *et al.* Genetic and environmental effects on puroindoline-a and puroindoline-b content and their relationship to technological properties in French bread wheats. **Journal of Cereal Science**, 34, n. 1, p. 37-47, 2001.
- IWGSC; APPELS, R.; EVERSOLE, K.; STEIN, N. *et al.* Shifting the limits in wheat research and breeding using a fully annotated reference genome. **Science**, 361, n. 6403, p. eaar7191, 2018.
- JI, T.; MA, F.; BAIK, B. K. Biochemical characteristics of soft wheat grain associated with endosperm separation from bran and flour yield. **Cereal Chemistry**, 97, n. 3, p. 566-572, 2020.
- JIA, X.-Y.; XU, C.-Y.; JING, R.-L.; LI, R.-Z. *et al.* Molecular cloning and characterization of wheat calreticulin (CRT) gene involved in drought-stressed responses. **Journal of experimental botany**, 59, n. 4, p. 739-751, 2008.
- JOPPA, L. R.; KHAN, K.; WILLIAMS, N. D. Chromosomal location of genes for gliadin polypeptides in durum wheat *Triticum turgidum* L. **Theoretical and Applied Genetics**, 64, n. 4, p. 289-293, 1983.
- KASARDA, D. D. Glutenin structure in relation to wheat quality. **Wheat is unique**, p. 277-302, 1989.

- KASARDA, D. D.; ADALSTEINS, E.; LEW, E. J. L.; LAZO, G. R. *et al.* Farinin: Characterization of a novel wheat endosperm protein belonging to the prolamin superfamily. **Journal of agricultural and food chemistry**, 61, n. 10, p. 2407-2417, 2013.
- KHAN, S.; JABEEN, R.; DEEBA, F.; WAHEED, U. *et al.* Heat Shock Proteins: Classification, Functions and Expressions in Plants during Environmental Stresses. **Journal of Bioresource Management**, 8, n. 2, p. 9, 2021.
- KHAN, S.; MEMON, A. N.; GHANGHRO, A. B.; NABI, G. Characterization of Wheat protein (Albumin) in different varieties of wheat cultivated in Sindh through SDS-PAGE Electrophoresis. **Sindh University Research Journal-SURJ (Science Series)**, 47, n. 2, 2015.
- KIM, Y. J.; CHOI, S. H.; PARK, B. S.; SONG, J. T. *et al.* Proteomic analysis of the rice seed for quality improvement. **Plant breeding**, 128, n. 6, p. 541-550, 2009.
- KIMURA, S.; HIGASHINO, Y.; KITAO, Y.; MASUDA, T. *et al.* Expression and characterization of protein disulfide isomerase family proteins in bread wheat. **BMC plant biology**, 15, n. 1, p. 1, 2015.
- KÖHLER, P.; BELITZ, H. D.; WIESER, H. Disulfide bonds in wheat gluten - Further cystine peptides from high-molecular-weight (HMW) and low-molecular-weight (LMW) subunits of glutenin and from gamma-gliadins. **Zeitschrift Fur Lebensmittel-Untersuchung Und-Forschung**, 196, n. 3, p. 239-247, Mar 1993.
- KRUVE, A.; REBANE, R.; KIPPER, K.; OLDEKOP, M.-L. *et al.* Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part I. **Analytica Chimica Acta**, 870, n. Supplement C, p. 29-44, 2015/04/22/ 2015.
- LAEMMLI, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. **nature**, 227, n. 5259, p. 680-685, 1970.
- LAINO, P.; SHELTON, D.; FINNIE, C.; DE LEONARDIS, A. M. *et al.* Comparative proteome analysis of metabolic proteins from seeds of durum wheat (cv. Svevo) subjected to heat stress. **Proteomics**, 10, n. 12, p. 2359-2368, 2010. Article.
- LANDOLFI, V.; D'AURIA, G.; NICOLAI, M. A.; NITRIDE, C. *et al.* The effect of nitrogen fertilization on the expression of protein in wheat and tritordeum varieties using a proteomic approach. **Food Research International**, 148, p. 110617, 2021.
- LANGENKÄMPER, G.; ZÖRB, C. Modern aspects of wheat grain proteins. **Journal of Applied Botany and Food Quality**, 92, p. 240-245, 2019.
- LARRE, C.; LUPI, R.; GOMBAUD, G.; BROSSARD, C. *et al.* Assessment of allergenicity of diploid and hexaploid wheat genotypes: identification of allergens in the albumin/globulin fraction. **J Proteomics**, 74, n. 8, p. 1279-1289, Aug 12 2011.
- LESAGE, V. S.; MERLINO, M.; CHAMBON, C.; BOUCHET, B. *et al.* Proteomes of hard and soft near-isogenic wheat lines reveal that kernel hardness is related to the amplification of a stress response during endosperm development. **Journal of experimental botany**, 63, n. 2, p. 1001-1011, 2012.
- LEW, E. J. L.; KUZMICKY, D. D.; KASARDA, D. D. Characterization of low molecular weight glutenin subunits by reversed-phase high-performance liquid chromatography, sodium

dodecyl sulfate-polyacrylamide gel electrophoresis, and N-terminal amino acid sequencing. **Cereal Chemistry**, 69, p. 508-508, 1992.

LEXHALLER, B.; COLGRAVE, M. L.; SCHERF, K. A. Characterization and relative quantitation of wheat, rye and barley gluten protein types by LC-MS/MS. **Frontiers in Plant Science**, 10, p. 1530, 2019.

LI, X.; LI, Y.; YU, X.; SUN, F. *et al.* Genomics-enabled analysis of Puroindoline b2 genes identifies new alleles in wheat and related Triticeae species. **International journal of molecular sciences**, 21, n. 4, p. 1304, 2020.

LI, Y.; LIU, K.; CHEN, F.; CHENG, Y. Comparative proteomics analysis reveals the effect of germination and selenium enrichment on the quality of brown rice during storage. **Food chemistry**, 269, p. 220-227, 2018a.

LI, Z.; LI, Z.; MUHAMMAD, W.; LIN, M. *et al.* Proteomic analysis of positive influence of alternate wetting and moderate soil drying on the process of rice grain filling. **Plant Growth Regulation**, 84, n. 3, p. 533-548, 2018b.

LIN, S. K.; CHANG, M. C.; TSAI, Y. G.; LUR, H. S. Proteomic analysis of the expression of proteins related to rice quality during caryopsis development and the effect of high temperature on expression. **Proteomics**, 5, n. 8, p. 2140-2156, 2005.

LIU, L.; HE, Z.; YAN, J.; ZHANG, Y. *et al.* Allelic variation at the Glu-1 and Glu-3 loci, presence of the 1B. 1R translocation, and their effects on mixographic properties in Chinese bread wheats. **Euphytica**, 142, n. 3, p. 197-204, 2005.

LIU, L.; WANG, A.; APPELS, R.; MA, J. *et al.* A MALDI-TOF based analysis of high molecular weight glutenin subunits for wheat breeding. **Journal of Cereal Science**, 50, n. 2, p. 295-301, 9// 2009.

LIU, P.; MA, X.; WAN, H.; ZHENG, J. *et al.* Effects of differential nitrogen application on wheat grain proteome. **Journal of Cereal Science**, 102, p. 103367, 2021.

LIU, W.; ZHANG, Y.; GAO, X.; WANG, K. *et al.* Comparative proteome analysis of glutenin synthesis and accumulation in developing grains between superior and poor quality bread wheat cultivars. **J Sci Food Agric**, 92, n. 1, p. 106-115, Jan 15 2012.

LIU, Y.; PAN, T.; TANG, Y.; ZHUANG, Y. *et al.* Proteomic analysis of rice subjected to low light stress and overexpression of OsGAPB increases the stress tolerance. **Rice**, 13, n. 1, p. 1-10, 2020.

LULLIEN-PELLERIN, V. Both genetic and environmental conditions affect wheat grain texture: Consequences for grain fractionation and flour properties. **Journal of Cereal Science**, 92, p. 102917, 2020.

LULLIEN-PELLERIN, V.; HARASZI, R.; ANDERSSON, R. S.; MORRIS, C. F. Understanding the mechanics of wheat grain fractionation and the impact of Puroindolines on milling and product quality. *In: Wheat quality for improving processing and human health*: Springer, 2020. p. 369-385.

LULLIEN-PELLERIN, V.; MARION, D. Plant cysteine-rich antimicrobial proteins that interact with biological membranes. **Membrane interacting peptides and proteins**, F. Heitz Ed., (pp. 125-146): Research Signpost, ISBN: 81-7736-104-X, Kerala, Inde, p. 125-146, 2002.

- LUTZ, E.; WIESER, H.; KOEHLER, P. Identification of Disulfide Bonds in Wheat Gluten Proteins by Means of Mass Spectrometry/Electron Transfer Dissociation. **Journal of Agricultural and Food Chemistry**, 60, n. 14, p. 3708-3716, 2012/04/11 2012.
- LV, L.; ZHAO, A.; ZHANG, Y.; LI, H. *et al.* Proteome and transcriptome analyses of wheat near isogenic lines identifies key proteins and genes of wheat bread quality. **Scientific Reports**, 11, n. 1, p. 1-15, 2021.
- MA, F.; LI, M.; LI, T.; LIU, W. *et al.* Overexpression of avenin-like b proteins in bread wheat (*Triticum aestivum* L.) improves dough mixing properties by their incorporation into glutenin polymers. **PloS one**, 8, n. 7, p. e66758, 2013.
- MA, X.; XUE, H.; SUN, J.; SAJJAD, M. *et al.* Transformation of Pinb-D1x to soft wheat produces hard wheat kernel texture. **Journal of Cereal Science**, 91, p. 102889, 2020.
- MALDANER, L.; JARDIM, I. C. S. F. O estado da arte da cromatografia líquida de ultra eficiência. **Química nova**, 32, p. 214-222, 2009.
- MARTINEZ-ESTESO, M. J.; NØRGAARD, J.; BROHÉE, M.; HARASZI, R. *et al.* Defining the wheat gluten peptide fingerprint via a discovery and targeted proteomics approach. **Journal of Proteomics**, 2016.
- MEHRI, N.; FOTOVAT, R.; MIRZAEI, M.; FARD, E. M. *et al.* Proteomic analysis of wheat contrasting genotypes reveals the interplay between primary metabolic and regulatory pathways in anthers under drought stress. **Journal of Proteomics**, 226, p. 103895, 2020.
- MICHAEELEVSKI, I.; KIRSHENBAUM, N.; SHARON, M. T-wave Ion Mobility-mass Spectrometry: Basic Experimental Procedures for Protein Complex Analysis. **Journal of Visualized Experiments : JoVE**, n. 41, p. 1985, 07/31 2010.
- MOGK, A.; RUGER-HERREROS, C.; BUKAU, B. Cellular functions and mechanisms of action of small heat shock proteins. **Annual review of microbiology**, 73, p. 89-110, 2019.
- MOREL, M. H. Acid-polyacrylamide Gel Electrophoresis of Wheat Glutenins: A New Tool for the Separation of High and Low Molecular Weight Subunits. **Cereal Chem**, 71, n. 3, p. 238-242, 1994.
- MORRIS, C. F.; BHAVE, M. Reconciliation of D-genome puroindoline allele designations with current DNA sequence data. **Journal of Cereal Science**, 48, n. 2, p. 277-287, 2008.
- MUCCILLI, V.; CUNSOLO, V.; SALETTI, R.; FOTI, S. *et al.* Characterisation of a specific class of typical low molecular weight glutenin subunits of durum wheat by a proteomic approach. **Journal of Cereal Science**, 51, n. 1, p. 134-139, 1// 2010.
- NAEEM, H. A.; MACRITCHIE, F. Polymerization of glutenin during grain development in near-isogenic wheat lines differing at Glu-D1 and Glu-B1 in greenhouse and field. **Journal of Cereal Science**, 41, n. 1, p. 7-12, 1 2005.
- NANJO, Y.; SKULTETY, L.; UVÁČKOVÁ, L. u.; KLUBICOVÁ, K. n. *et al.* Mass spectrometry-based analysis of proteomic changes in the root tips of flooded soybean seedlings. **Journal of proteome research**, 11, n. 1, p. 372-385, 2012.

NEVES, G. W. P.; CURTY, N. d. A.; KUBITSCHKE-BARREIRA, P. H.; FONTAINE, T. *et al.* Modifications to the composition of the hyphal outer layer of *Aspergillus fumigatus* modulates HUVEC proteins related to inflammatory and stress responses. **Journal of Proteomics**, 151, p. 83-96, 1/16/ 2017.

OSAWA-MARTÍNEZ, E. E.; MINJAREZ, B.; RODRÍGUEZ-YÁÑEZ, Y.; REZA-ZALDIVAR, E. E. *et al.* Comparative proteomic analysis of hybrid maize MR2008 and its parental lines. **The Journal of Agricultural Science**, 159, n. 7-8, p. 570-579, 2021.

OSBORNE, T. B. **The proteins of wheat kernel**. Washington: Carnegie Institute Publication, 84, 1907.

OSIPOVA, S. V.; PERMYAKOVA, M. D.; PERMYAKOV, A. V. Role of non-prolamin proteins and low molecular weight redox agents in protein folding and polymerization in wheat grains and influence on baking quality parameters. **Journal of agricultural and food chemistry**, 60, n. 49, p. 12065-12073, 2012.

OURY, F.-X.; LASME, P.; MICHELET, C.; ROUSSET, M. *et al.* Relationships between wheat grain physical characteristics studied through near-isogenic lines with distinct puroindoline-b allele. **Theoretical and Applied Genetics**, 128, n. 5, p. 913-929, 2015.

PAULY, A.; PAREYT, B.; FIERENS, E.; DELCOUR, J. A. Wheat (*Triticum aestivum* L. and *T. turgidum* L. ssp. *durum*) kernel hardness: I. Current view on the role of puroindolines and polar lipids. **Comprehensive Reviews in Food Science and Food Safety**, 12, n. 4, p. 413-426, 2013.

PAYNE, P. I. Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. **Annual Review of Plant Physiology**, 38, n. 1, p. 141-153, 1987.

PAYNE, P. I.; CORFIELD, K. G.; HOLT, L. M.; BLACKMAN, J. A. Correlations between the inheritance of certain high-molecular weight subunits of glutenin and bread-making quality in progenies of six crosses of bread wheat. **Journal of the Science of Food and Agriculture**, 32, n. 1, p. 51-60, 1981.

PAYNE, P. I.; LAWRENCE, G. J. Catalogue of alleles for the complex gene loci, Glu-A1, Glu-B1, and Glu-D1 which code for high-molecular-weight subunits of glutenin in hexaploid wheat. **Cereal Research Communications**, p. 29-35, 1983.

PAYNE, P. I.; NIGHTINGALE, M. A.; KRATTIGER, A. F.; HOLT, L. M. The relationship between HMW glutenin subunit composition and the bread-making quality of British-grown wheat varieties. **Journal of the Science of Food and Agriculture**, 40, n. 1, p. 51-65, 1987.

PELEGRIN, L. J.; CARVALHO, I. R.; NARDINO, M.; FERRARI, M. *et al.* Technological quality of wheat under nitrogen management. **International Journal of Current Research**, 8, n. 06, p. 32932-32936, 2016.

PIRES, J. L. F.; VARGAS, L.; CUNHA, G. R. (ed.). **Trigo no Brasil - Bases para produção competitiva e sustentável**. Passo Fundo- RS -Brasil: Embrapa Trigo, 2011. 488 p.

PIROZI, M. R.; MARGIOTTA, B.; LAFIANDRA, D.; MACRITCHIE, F. Composition of polymeric proteins and bread-making quality of wheat lines with allelic HMW-GS differing in number of cysteines. **Journal of cereal science**, 48, n. 1, p. 117-122, 2008.

- POPINEAU, Y.; CORNEC, M.; LEFEBVRE, J.; MARCHYLO, B. Influence of high M r glutenin subunits on glutenin polymers and rheological properties of glutens and gluten subfractions of near-isogenic lines of wheat Sicco. **Journal of Cereal Science**, 19, n. 3, p. 231-241, 1994a.
- POPINEAU, Y.; POGNA, N.; LEFEBVRE, J. Rheological properties of glutens differing by their glutenin subunit compositions. *In: Wheat kernel proteins-molecular and functional aspects*. Viterbo, Italy: Università Degli Studi della Tuscia, 1994b. p. 129-134.
- PRESUTTI, R. J.; CANGEMI, J. R.; CASSIDY, H. D.; HILL, D. A. Celiac disease. **American family physician**, 76, n. 12, p. 1795-1802, 2007.
- RÍOS-CASTRO, E.; SOUZA, G. H. M. F.; DELGADILLO-ÁLVAREZ, D. M.; RAMÍREZ-REYES, L. *et al.* Quantitative proteomic analysis of MARC-145 cells infected with a Mexican Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) strain using Label-free based DIA approach. **Journal of the American Society for Mass Spectrometry**, 2020.
- ROMBOUTS, I.; LAGRAIN, B.; BRUNNBAUER, M.; DELCOUR, J. A. *et al.* Improved identification of wheat gluten proteins through alkylation of cysteine residues and peptide-based mass spectrometry. **Sci. Rep.**, 3, 07/24/online 2013. Article.
- SANTOS, M. C. B.; DA SILVA LIMA, L. R.; DOS SANTOS D'ALMEIDA, C. T.; VICTORIO, V. C. M. *et al.* Foodomics in wheat flour reveals phenolic profile of different genotypes and technological qualities. **LWT**, 153, p. 112519, 2022.
- SAULNIER, L. Les grains de céréales : diversité et compositions nutritionnelles. **Cahiers de Nutrition et de Diététique**, 47, n. 1, p. S4-S15, 2012/09/01 2012.
- SCHALK, K.; KOEHLER, P.; SCHERF, K. A. Targeted liquid chromatography tandem mass spectrometry to quantitate wheat gluten using well-defined reference proteins. **PloS one**, 13, n. 2, p. e0192804, 2018.
- SCHERF, K. A.; KOEHLER, P.; WIESER, H. Gluten and wheat sensitivities—an overview. **Journal of Cereal Science**, 67, p. 2-11, 2015.
- SERNA-SALDIVAR, S. O. **Cereal Grains: Properties, Processing, and Nutritional Attributes**. Boca Raton: CRC Press, 2010. 747 p. (Food Preservation Technology).
- SHEWRY, P.; GILBERT, S.; SAVAGE, A.; TATHAM, A. *et al.* Sequence and properties of HMW subunit 1Bx20 from pasta wheat (*Triticum durum*) which is associated with poor end use properties. **Theoretical and Applied Genetics**, 106, n. 4, p. 744-750, 2003.
- SHEWRY, P. R. Wheat. **Journal of experimental botany**, 60, n. 6, p. 1537-1553, 2009.
- SHEWRY, P. R.; HALFORD, N. G. Cereal seed storage proteins: structures, properties and role in grain utilization. **Journal of Experimental Botany**, 53, n. 370, p. 947-958, 2002.
- SHEWRY, P. R.; TATHAM, A. S. The prolamin storage proteins of cereal seeds: structure and evolution. **Biochemical journal**, 267, n. 1, p. 1, 1990.
- SHEWRY, P. R.; TATHAM, A. S. Improving wheat to remove coeliac epitopes but retain functionality. **Journal of Cereal Science**, 67, n. Supplement C, p. 12-21, 2016/01/01/ 2016.

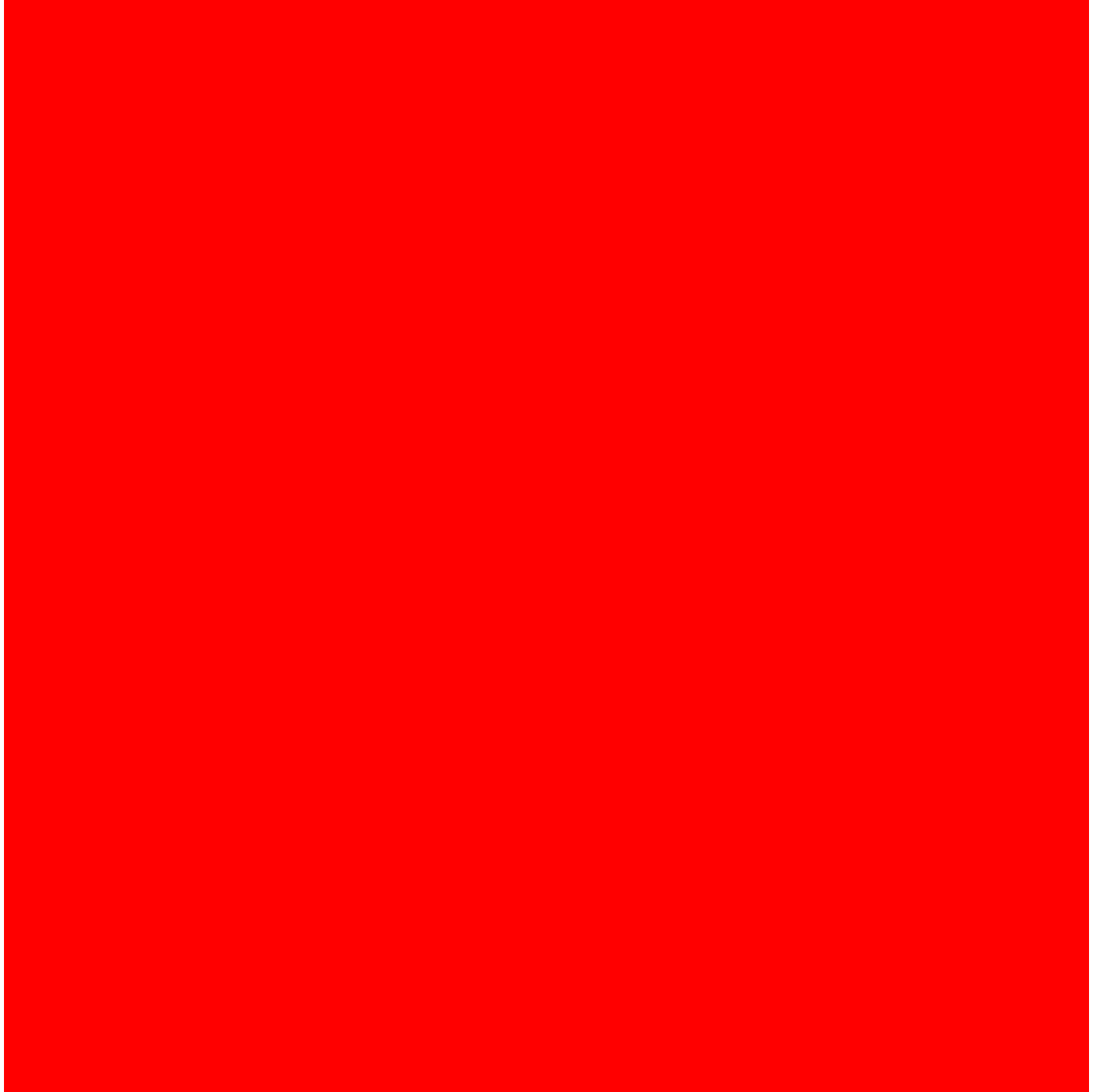
- SHEWRY, P. R.; TATHAM, A. S.; FORDE, J.; KREIS, M. *et al.* The classification and nomenclature of wheat gluten proteins: a reassessment. **Journal of Cereal Science**, 4, n. 2, p. 97-106, 1986.
- SIGRIST, C. J. A.; CERUTTI, L.; DE CASTRO, E.; LANGENDIJK-GENEVAUX, P. S. *et al.* PROSITE, a protein domain database for functional characterization and annotation. **Nucleic acids research**, 38, n. suppl_1, p. D161-D166, 2010.
- SKYLAS, D. J.; VAN DYK, D.; WRIGLEY, C. W. Proteomics of wheat grain. **Journal of cereal science**, 41, n. 2, p. 165-179, 2005.
- SOUZA, G. H. M. F.; GUEST, P. C.; MARTINS-DE-SOUZA, D. LC-MS E, Multiplex MS/MS, Ion Mobility, and Label-Free Quantitation in Clinical Proteomics. **Multiplex Biomarker Techniques: Methods and Applications**, p. 57-73, 2017.
- SURGET, A.; BARRON, C. Histologie du grain de blé. **Industrie des céréales**, n. 145, p. 3-7, 2005.
- TALEUZZAMAN, M.; ALI, S.; GILANI, S. J.; IMAM, S. S. *et al.* Ultra performance liquid chromatography (UPLC)-a review. **Austin J Anal Pharm Chem**, 2, n. 6, p. 1056, 2015.
- TÁVORA, F. T. P. K.; BEVITORI, R.; MELLO, R. N.; CINTRA, M. M. D. F. *et al.* Shotgun proteomics coupled to transient-inducible gene silencing reveal rice susceptibility genes as new sources for blast disease resistance. **Journal of Proteomics**, 241, p. 104223, 2021.
- TOMIĆ, J.; TORBICA, A.; POPOVIĆ, L.; STRELEC, I. *et al.* Albumins Characterization in Relation to Rheological Properties and Enzymatic Activity of Wheat Flour Dough. **Journal of Agricultural Science and Technology**, 17, n. 4, p. 805-816, 2015.
- VAN DEN BROECK, H. C.; CORDEWENER, J. H. G.; NESSEN, M. A.; AMERICA, A. H. P. *et al.* Label free targeted detection and quantification of celiac disease immunogenic epitopes by mass spectrometry. **Journal of Chromatography A**, 1391, n. 0, p. 60-71, 4/24/ 2015.
- VENSEL, W. H.; DUPONT, F. M.; SLOANE, S.; ALTENBACH, S. B. Effect of cleavage enzyme, search algorithm and decoy database on mass spectrometric identification of wheat gluten proteins. **Phytochemistry**, 72, n. 10, p. 1154-1161, 2011.
- VENSEL, W. H.; TANAKA, C. K.; ALTENBACH, S. B. Protein composition of wheat gluten polymer fractions determined by quantitative two-dimensional gel electrophoresis and tandem mass spectrometry. **Proteome science**, 12, n. 1, p. 8, 2014.
- VICTORIO, V. C. M.; ALVES, T. O.; SOUZA, G.; GUTKOSKI, L. C. *et al.* NanoUPLC-MSE reveals differential abundance of gluten proteins in wheat flours of different technological qualities. **Journal of Proteomics**, 239, p. 104181, 2021.
- VICTORIO, V. C. M.; SOUZA, G. H. M. F.; SANTOS, M. C. B.; VEGA, A. R. *et al.* Differential expression of albumins and globulins of wheat flours of different technological qualities revealed by nanoUPLC-UDMSE. **Food Chemistry**, 239, p. 1027-1036, 1/15/ 2018.
- VINCENT, D.; BUI, A.; RAM, D.; EZERNIEKS, V. *et al.* Mining the Wheat Grain Proteome. **International journal of molecular sciences**, 23, n. 2, p. 713, 2022.

- WANG, B.; ZHANG, J.; CHEN, P.; JI, Z. *et al.* Prediction of peptide drift time in ion mobility mass spectrometry from sequence-based features. **BMC bioinformatics**, 14, n. Suppl 8, p. S9, 2013.
- WIESER, H. Chemistry of gluten proteins. **Food Microbiology**, 24, n. 2, p. 115-119, 4// 2007.
- WIŚNIEWSKI, J. R.; ZOUGMAN, A.; NAGARAJ, N.; MANN, M. Universal sample preparation method for proteome analysis. **Nature methods**, 6, n. 5, p. 359-362, 2009.
- WRIGLEY, C. W. Giant proteins with flour power. **Nature**, 381, n. 6585, p. 738-739, 1996.
- XHAFERAJ, M.; ALVES, T. O.; FERREIRA, M. S. L.; SCHERF, K. A. Recent progress in analytical method development to ensure the safety of gluten-free foods for celiac disease patients. **Journal of Cereal Science**, p. 103114, 2020.
- YU, X.; CHEN, X.; WANG, L.; YANG, Y. *et al.* Novel insights into the effect of nitrogen on storage protein biosynthesis and protein body development in wheat caryopsis. **Journal of experimental botany**, 68, n. 9, p. 2259-2274, 2017.
- ZHANG, L.; DONG, C.; CHEN, Z.; GUI, L. *et al.* WheatGmap: a comprehensive platform for wheat gene mapping and genomic studies. **Molecular Plant**, 14, n. 2, p. 187-190, 2021.
- ZHANG, Y.; HU, X.; JUHASZ, A.; ISLAM, S. *et al.* Characterising avenin-like proteins (ALPs) from albumin/globulin fraction of wheat grains by RP-HPLC, SDS-PAGE, and MS/MS peptides sequencing. **BMC plant biology**, 20, n. 1, p. 1-19, 2020.
- ZHAO, Q.; LIN, J.; YOUSAF, L.; XUE, Y. *et al.* Protein structural properties and proteomic analysis of rice during storage at different temperatures. **Food Chemistry**, 361, p. 130028, 2021.

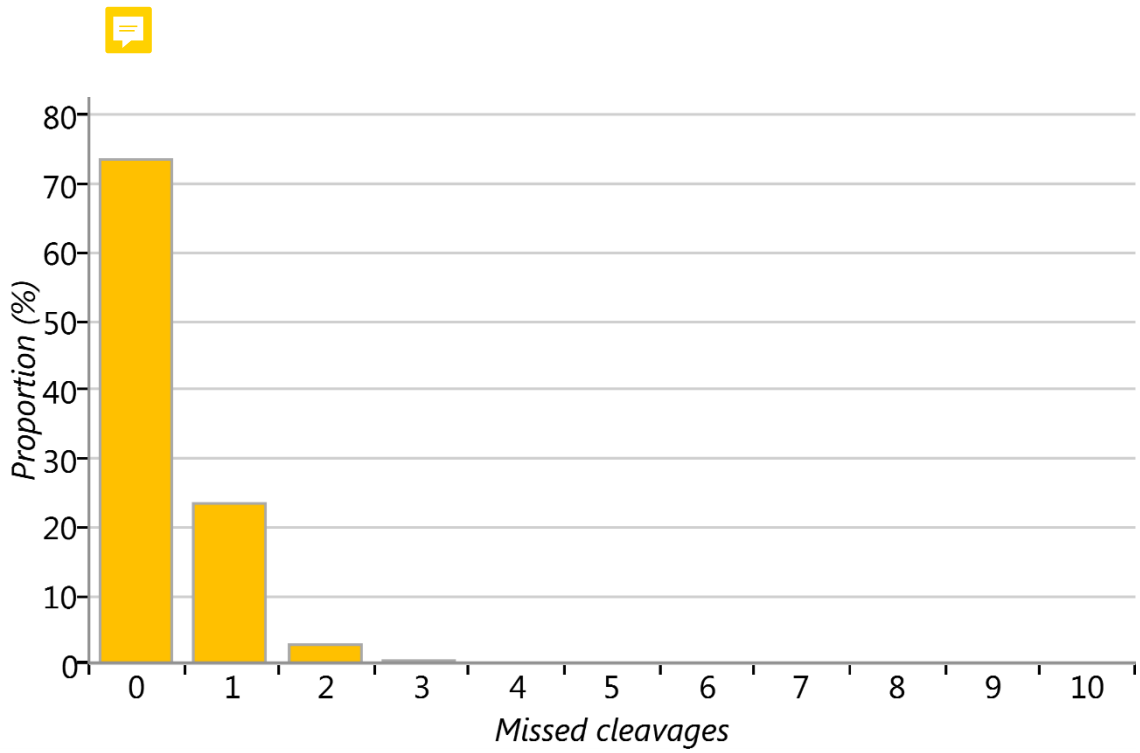
WSBV55	1	0	4,976	---	---	Uncharacterized protein OS=Triticum aestivum PE=4 SV=1
WSDWK3	1	0	6,9147	---	---	Uncharacterized protein OS=Triticum aestivum PE=4 SV=1
WSE0L9;W	3	0	18,6412	---	---	Uncharacterized protein OS=Triticum aestivum PE=4 SV=1
W5F5Q6	1	0	4,6347	---	---	Uncharacterized protein OS=Triticum aestivum PE=4 SV=1
WSIA07	1	0	12,8656	---	---	Uncharacterized protein OS=Triticum aestivum PE=4 SV=1

REVERSE65742	54912,85236	59673,53844	52918,30232	0,062160491	3	3	1	0	0	1	1	0	3
AOA1D5SX3;AOA1D5SX4;AOA1D5SX5;AOA1D5SX6;Q03968;W4ZN68	7200,982332	6633,980824	10682,53486	0,268236625	5	3	1	0	0	0	0	5	0
REVERSE9562	1969,704695	2128,711966	2062,755917	0,038898829	1	3	0	0	0	1	0	0	0
AOA1D5XMK2	17430,52931	17895,57166	16829,80494	0,030734058	2	3	1	0	0	0	0	2	0
AOA1D6B7D4	5545,778968	4763,600686	4596,752947	0,101972272	1	3	0	0	0	0	0	0	0
REVERSE99407	67608,61201	66569,94895	56429,0272	0,097213943	3	3	1	0	0	1	1	0	3
AOA1DSRV86;AOA1D5S939;W4ZRV5					1	0	0	0	0	0	0	0	0
AOA1D5W323	29317,02497	24464,78802	23721,29004	0,117629877	3	3	1	0	0	0	0	3	0
AOA1D5SNP0			181,542656		1	1	0	0	0	0	0	0	0
AOA1D6DIE4;AOA1D6DIE5;AOA1D6SAR6	37348,4735	36042,89925	31330,14059	0,09069288	3	3	1	0	0	0	0	3	0
AOA1D6D5H8	13413,71595	13804,59199	21238,07807	0,27296063	3	3	1	0	0	0	0	3	0
AOA1D6AGT9;AOA1D6BEX1;P12783;W5H4V7;AOA1D6AGT8;AOA1D6AXU9	22099,96368	22913,70905	22699,1441	0,018685518	4	3	1	0	0	0	0	4	0
AOA1DSUR1;AOA1D5RR9;AOA1D5RUZ4;AOA1D5RYL0;AOA1D5S9K3;AOA1D5T0X9;AOA1D5T746;AOA1D5J7HT09	3167,601255	3072,447914	2783,77256	0,066450781	4	3	1	0	0	0	0	4	0
AOA1D5SUP9;AOA1D5UPT0;AOA1D5UPT1;AOA1D5UPT2	6432,465552	7201,286203	6986,203415	0,057708174	2	3	1	0	0	0	0	2	0
AOA1D5SD15;AOA1D5SD16;AOA1D5SUT4;AOA1D5SUT5;AOA1D5YGG9;AOA1D5YGH0;AOA1D5ZB54;AOA1D5J7HT09	550,1886775	715,0659784	378,6036377	0,307037987	1	3	0	0	0	0	0	0	0
AOA1D5TS27	178410,8183	145948,1773	163736,2461	0,09991607	4	3	1	0	0	0	0	4	0
AOA096UKU5	19007,65494	18812,64591	16151,85215	0,088683868	2	3	1	0	0	0	0	2	0
AOA1D5V6V1	26996,22186	29939,62767	35835,72916	0,145557349	1	3	0	0	0	0	0	0	0
AOA1D5XPB2;AOA1D5XPB3	7104,106217	7814,438483	6891,62115	0,066474387	2	3	1	0	0	0	0	2	0
AOA1D5VFC5;AOA1D5VU68;AOA1D5WEY7;AOA1D5WEY8	169,6936898	52,35320353	116,6336146	0,520486857	2	3	0	0	0	0	0	0	0
AOA1D5VPJ3;P011084	19093,20111	18828,23555	19155,96047	0,009142786	3	3	1	0	0	0	0	3	0
AOA1D5TSR5;AOA1D5TSR6	12888,12844	12310,67995	11559,73644	0,054361369	6	3	1	0	0	0	0	6	0
AOA1D6CNS5;AOA1D6CNS6;AOA1D6CCD8	220307,8946	222791,9756	234808,62	0,034318853	1	3	0	0	0	0	0	0	0
AOA1D6S4Q2	60838,0939	54077,35423	55445,60908	0,06294421	3	3	1	0	0	0	0	3	0
AOA1D6A2T8;AOA1D5ZWI6	95616,94817	88981,71309	84008,442	0,065047211	1	3	0	0	0	0	0	0	0
AOA1D5WG42;AOA1D5WG43;AOA1D5WG44;AOA1D5WG46;AOA1D5WG47;AOA1D5WG48	109765,2387	120316,5656	124745,5169	0,065066837	5	3	1	0	0	0	0	5	0
AOA1D6AIDD	23432,81521	18251,95091	25400,81898	0,165138386	2	3	1	0	0	0	0	2	0
REVERSE111267	2787,542489	2473,566738	3212,901148	0,131365045	1	3	0	0	0	0	0	0	0
W4ZUK2;W5AB71;AOA1D5T345;W5AM18	3284,246146	2504,494994	3991,381179	0,228137812	1	3	0	0	0	1	0	0	0
	4891,226305	3822,635951	2417,674296	0,334342521	2	3	0	0	0	0	0	0	0

AOA1D5XMK2	22122,94167	20182,46721	15614,38934	0,1730759	2	3	1	0	0	0	0	2	0
AOA1D6B7D4	4436,127257	5050,940281	4967,957672	0,0692350	1	3	0	0	0	0	0	0	0
REVERSE99407	63634,31273	56754,77185	54494,80743	0,0816603	3	3	1	0	0	1	1	0	3
AOA1D5RV86;AOA1D5S939;	6728,355572				1	1	0	0	0	0	0	0	0
AOA1D5W323	6399,331822	26769,28694	28036,13956	0,5951910	3	3	0	0	0	0	0	0	0
AOA1D5SNP0	8674,742572				1	1	0	0	0	0	0	0	0
AOA1D6DIE4;AOA1D6DIE5;A	41987,04128	22098,10671	29078,25134	0,3249344	3	3	0	0	0	0	0	0	0
AOA1D6D5H8	2916,68041	18388,92908	15989,59069	0,6698299	3	3	0	0	0	0	0	0	0
AOA1D6AGT9;AOA1D6BEX1	17114,73788	29483,45058	30165,19895	0,2870820	4	3	1	0	0	0	0	4	0
AOA1D5URA1;AOA1D5RRC9	2539,595255	10842,4288	2787,597528	0,8763988	4	3	0	0	0	0	0	0	0
J7HT09	5347,420907	7236,426605	8429,357841	0,2218625	2	3	1	0	0	0	0	2	0
AOA1D5UPS9;AOA1D5UPT0	28198,7689	594,8440485	239,4779896	1,6574918	1	3	0	0	0	0	0	0	0
AOA1D5SD15;AOA1D5SD16;	239205,0151	155494,9374	135368,9449	0,3116605	4	3	0	0	0	0	0	0	0
AOA1D5TS27	21738,30639	13767,47741	14035,82855	0,2741019	2	3	1	0	0	0	0	2	0
AOA096UKU5	39727,09215	29950,10993	23856,89833	0,2567845	1	3	0	0	0	0	0	0	0
AOA1D5V6V1	3601,604983	9476,235561	7346,46273	0,4368454	2	3	0	0	0	0	0	0	0
AOA1D5XPB2;AOA1D5XPB3	2173,683971	19,51047094	268,3241009	1,4362735	2	3	0	0	0	0	0	0	0
AOA1D5VFC5;AOA1D5VU68	9845,040824	25673,84478	25437,33948	0,4464480	3	3	0	0	0	0	0	0	0
AOA1D5VPJ3;P01084	11869,43547	13448,46146	12771,95153	0,0623949	6	3	1	0	0	0	0	6	0
AOA1D5TSR5;AOA1D5TSR6	226220,6474	226336,5879	226142,3889	0,0004319	1	3	0	0	0	0	0	0	0
AOA1D6CNS5;AOA1D6CNS6	44802,47199	55348,95175	61624,50143	0,1576421	3	3	1	0	0	0	0	3	0
AOA1D6S4Q2	93394,52413	86699,92457	78723,38587	0,0851359	1	3	0	0	0	0	0	0	0
AOA1D6A2T8;AOA1D5ZWJ6	231967,4629	83425,63996	79393,76656	0,6607206	5	3	0	0	0	0	0	0	0
AOA1D5WG42;AOA1D5WG4	18997,28139	17736,34953	25335,13527	0,1968266	2	3	1	0	0	0	0	2	0
AOA1D6AID0	1012,534928	3105,134699	3736,593736	0,5446133	1	3	0	0	0	0	0	0	0
REVERSE111267	2617,086568	3747,696736	4723,611513	0,2852195	1	3	0	0	0	1	0	0	0
W4ZUK2;W5AB71;AOA1D5T	4665,243247	3674,016865	2508,418245	0,2985671	2	3	1	0	0	0	0	2	0



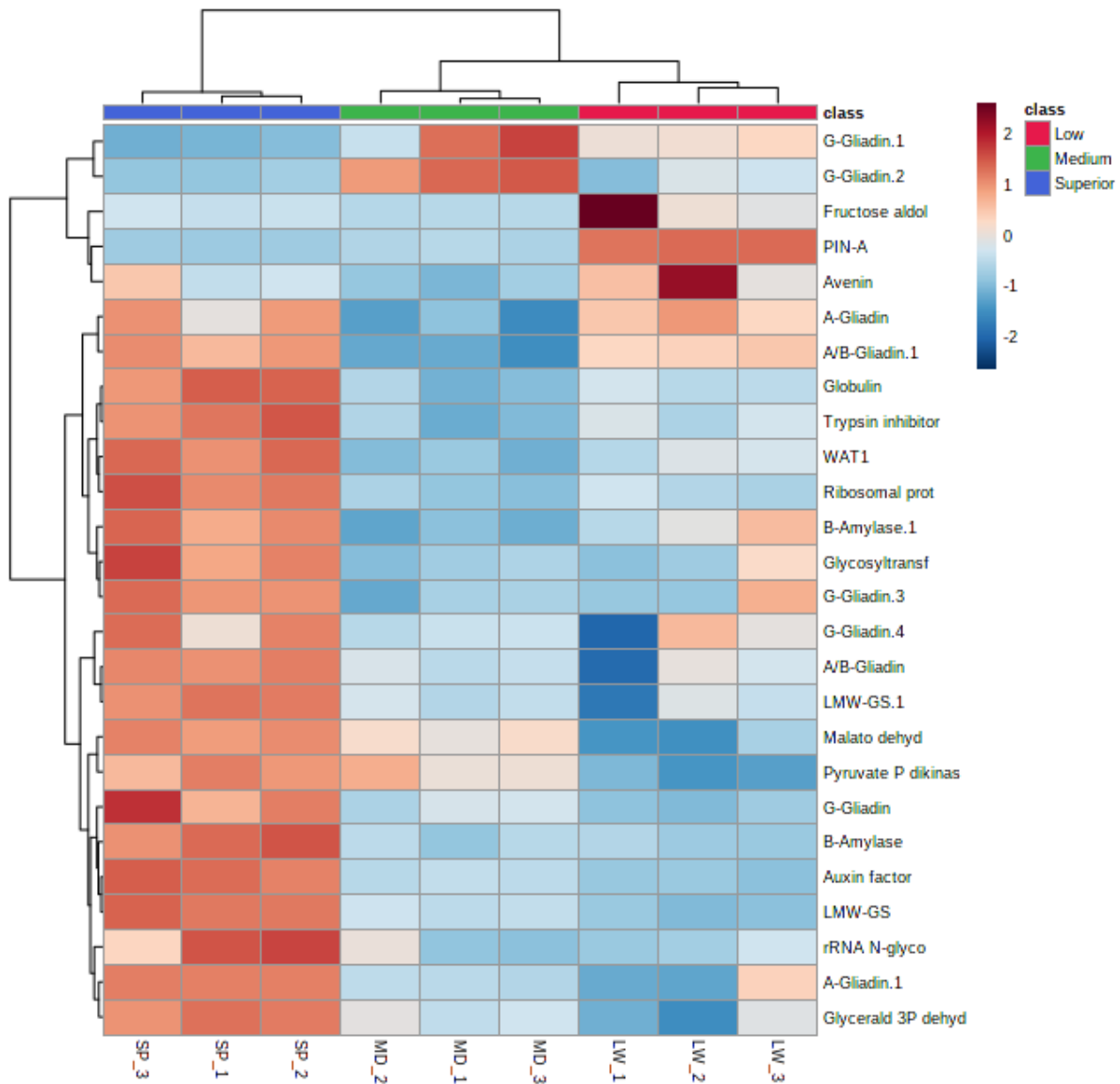
Code	Year	Country	City	Address	Postal Code	Region	Country Code	Area Code	Phone Number
ADAD5494	1	2193	020492	219507	MD	Underwater	646628	430045	755433
ADAD5495	1	6818	011416	020413	MD	Aspirant	667956	627732	627652
ADAD5496	1	7102	00004	00004	MD	Underwater	667956	627732	627652
ADAD5497	4	1877	015144	020425	MD	Underwater	667956	627732	627652
ADAD5498	2	17374	07279	020426	MD	Underwater	667956	627732	627652
ADAD5499	2	17374	07279	020426	MD	Underwater	667956	627732	627652
ADAD5500	2	17374	07279	020426	MD	Underwater	667956	627732	627652
ADAD5501	2	17374	07279	020426	MD	Underwater	667956	627732	627652
ADAD5502	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5503	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5504	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5505	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5506	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5507	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5508	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5509	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5510	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5511	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5512	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5513	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5514	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5515	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5516	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5517	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5518	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5519	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5520	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5521	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5522	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5523	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5524	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5525	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5526	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5527	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5528	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5529	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5530	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5531	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5532	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5533	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5534	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5535	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5536	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5537	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5538	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5539	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5540	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5541	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5542	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5543	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5544	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5545	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5546	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5547	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5548	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5549	1	628	03123	142663	MD	Underwater	582834	583276	584701
ADAD5550	1	628	03123	142663	MD	Underwater	582834	583276	584701



Supplementary data S2 - Percentage of missed cleavages



Supplementary data S3



s3. Heat map illustrating the clustering of differentially abundant and annotated proteins that were found in low (LW), medium (MD) and superior (SP) samples. The horizontal axis presents replicates from each sample and the vertical axis presents the proteins. Color gradient represents the variation of protein abundance. Highest quantities are represented by dark red and lowest quantities by dark blue.



Accession	PROSITE	InterPro	Pfam	Predominant domain
AOA1DSXF06	RICIN_B_LECTIN	Ricin B-like lectins (IPR035992)	No	Lectin domain of ricin B chain
AOA1DSVF84	PPR	Pentatricopeptide repeat (IPR002885)	PP-M3.M24R repeat	PPR domain
AOA1DSAG89	G_DYNAMIN_2	Dynamamin-type guanine nucleotide-binding (G) domain (IPR030381)	Dynamamin family	Dynamamin family
	UBIQUITIN_2	Ubiquitin domain (IPR000626)	Ubiquitin family	Ubiquitin family
AOA1DGC286	H15	Linker histone H1/H5 globular (H15) domain	Linker histone H1 and H5 family	Histone H1/H5
AOA1DSRW11	NO	NO	NO	No
AOA1DSYV17	PROTEIN_KINASE_DOM	Protein kinase domain (IPR000719)	Protein kinase domain	Protein kinase domain
AOA1DSUJ01	NO	Glycoside hydrolase, family 19, catalytic (IPR000726)	Chitinase class I	don't apply
AOA1DSZ761	MATH/TRAF domain	Plant calmodulin-binding domain	Plant calmodulin-binding domain	Plant calmodulin-binding domain
	BTB domain	MATH/TRAF domain	MATH domain	MATH/TRAF domain
AOA1DSVW4	NO	BTB/POZ domain	BTB/POZ domain	BTB/POZ domain
AOA1DSX482	Adenosine and AMP deaminase	cupin 1	High molecular weight glutenin subu	HMW-GS
AOA1DSV605	B3 DNA-binding domain	B3 DNA binding domain	cupin 1 (play a role in allergy, especial	Cupin 1 (play a role in allergy, especial
AOA1DSVCN0	Peptidase family A1 domain	Peptidase family A1 domain	Eukaryotic aspartyl protease	B3 DNA binding domain
	Saposin B type domain	Saposin B type, region 2	Saposin protein domain	Peptidase family A1
AOA1DS0575	Cereal trypsin/alpha-amylase inhibitors family	Bifunctional inhibitor/plant lipid transfer protein/seed storage helical doma	Protease inhibitor/seed storage/LTP f	Saposin B type
AOA1DSR006	YTH domain (RNA binding)	YTH domain	Protease inhibitor/seed storage/LTP f	Protease inhibitor/seed storage/LTP family
AOA1DS1499	Ankyrin repeat region circular (protein-protein interaction)	Ankyrin repeat-containing domain	YTS21-B-like domain	YTH domain (RNA binding)
	Ankyrin repeat	PGG domain (its function is not known)	Ankyrin repeats	Ankyrin repeats (protein-protein interaction)
AOA1DS7211	NO	Amino acid transporter, transmembrane domain	PGG (Domain of unknown function)	don't apply
AOA1DSV582	Carbamoyl-phosphate synthase subdomain	cupin 1	Amino acid transporter	Amino acid transporter
AOA1DSV6G4	Carbamoyl-phosphate synthase subdomain	cupin 1	cupin 1 (play a role in allergy, especial	cupin 1
AOA1DSV4H8	Zinc finger C2H2 type domain	Zinc finger C2H2-type	cupin 1 (play a role in allergy, especial	Zinc Finger C2H2 type domain
AOA1DSV576	No	No	No	No
AOA1DS5N8	No	No	No	No
AOA1DSUJ03	No	No	No	No
AOA1DSUSJ2	Small hydrophilic plant seed proteins	No	Small hydrophilic plant seed protein	Small hydrophilic plant seed protein
AOA1DSRYQ8	Heat shock hsp70 proteins family	No	Hsp70 protein (The Hsp70s are an imp	Hsp70 protein
AOA1DSUQ87	No	No	22-kDa peroxisomal membrane protei	22-kDa peroxisomal membrane protein
AOA1DSYK40	No	cupin 1	Cupin 1 (play a role in allergy, especial	Cupin 1
AOA1DC90	No	NB-ARC (a novel signalling motif found in bacteria and eukaryotes, shared b	NB-ARC domain	NB-ARC domain (a novel signalling motif found in bacteria and eukaryotes, shared by plant resistance gene products and regulators of cell death in animals)
AOA1DS6X91	Serine proteases, subclass family, histidine active	Peptidase S8 propeptide/proteinase inhibitor I9	Peptidase inhibitor I9	Serine protease/Peptidase inhibitor I9
	Serine proteases, subclass family, serine active	Peptidase S8/S53 domain (serine peptidase)	Subtilase family	NA aplicar
AOA1DSV8W1	No	No	No	No
AOA1DS6B72	Cereal trypsin/alpha-amylase inhibitors family	Bifunctional inhibitor/plant lipid transfer protein/seed storage helical doma	Protease inhibitor/seed storage/LTP f	Protease inhibitor/seed storage/LTP family
WSFAG2	Cytochrome P450 cysteine heme-iron ligand	Cytochrome P450	Cytochrome P450	Cytochrome P450
AOA1DSX59	Carbamoyl-phosphate synthase	Cupin 1	Cupin 1 (play a role in allergy, especial	Cupin 1
AOA1DS117	No	No	Protein of unknown function	Protein of unknown function
WS0616	Zinc finger RING-type	Zinc finger, RING-type	Zinc finger, C2HC4 type	Zinc finger, RING-type
AOA1DSRV0	No	No	No	No
AOA1DSR214	AAA-protein family	AAA+ ATPase domain (cellular processes: membrane fusion, proteolysis an	AAA proteins	AAA+ ATPase domain (cellular processes: membrane fusion, proteolysis and DNA replication)
		Peptidase M41	Peptidase family M41	Peptidase family M41
AOA1DS245	No	No	Late embryogenesis abundant protein	Protein of unknown function
AOA1DSW4L8	Cereal trypsin/alpha-amylase inhibitors	Bifunctional inhibitor/plant lipid transfer protein/seed storage helical doma	Protease inhibitor/seed storage/LTP f	Protease inhibitor/seed storage/LTP family
AOA1DSXF01	No	Translation elongation factor (are responsible for two main processes durin	Elongation factor	Elongation factor
AOA1DSW410	OTU domain (None of these proteins has a known biochemical function but low sequen	OTU domain	Peptidase C65 Otubain (is a highly spe	ubiquitin iso-peptidase
AOA1DS0174	No	Bifunctional inhibitor/plant lipid transfer protein/seed storage helical doma	Cys-rich Glialdin N-terminal	Cys-rich Glialdin N-terminal
AOA1DS2940	No	GCK (Involved in intracellular signalling pathways)	GCK domain (intracellular signalling p	GCK domain (intracellular signalling pathways)
AOA1DSY86	Fibronectin type-III domain (It is involved in cell adhesion, cell morphology, thrombosis,	Fibronectin type III domain	Fibronectin type III domain	Fibronectin type-III domain (It is involved in cell adhesion, cell morphology, thrombosis, cell migration, and embryonic differentiation)
	Oberon, PHD finger (is found in a wide variety of proteins involved in the re	PHD - plant homeodomain finger prot	PHD - plant homeodomain finger prot	PHD finger (is found in a wide variety of proteins involved in the regulation of chromatin structure)
AOA1DS5004	No	No	No	No
AOA1DSUJ03	No	No	Histidine phosphatase superfamily	Histidine phosphatase superfamily
AOA1DS0029	No	Cupin 1	Cupin 1 (play a role in allergy, especial	Cupin 1
WSGFM6	Lipolytic enzymes	GD5L lipase/esterase	GD5L-like Lipase/Acylhydrolase	Lipolytic enzymes
AOA1DS567	No	No	Late embryogenesis abundant protein	Protein of unknown function
AOA1DSV568	F-box domain	F-box domain	FBD (its precise function is unknown,	F-box domain (F-box domain mediates protein-protein interactions, is associated with cellular functions such as signal transduction and regulation of the cell cyc
AOA1DSVHB3	Small heat shock protein	Alpha crystallin/Hsp20 domain	Hsp20/alpha crystallin family	Hsp20 domain
AOA1DS6400	No	Sucrose synthase	Sucrose synthase	Sucrose synthase
		Glycosyl transferase	Glycosyl transferases group 1	don't apply
AOA1DSV7H4	No	ERCC4 domain	ERCC4 domain (is a family of nuclease	ERCC4 domain (is a family of nucleases)
AOA1DS717	No	Bifunctional inhibitor/plant lipid transfer protein/seed storage helical	Cys-rich Glialdin N-terminal	Cys-rich Glialdin N-terminal
AOA1DSAM05	No	F-box domain	F-box domain (F-box domain mediator	F-box domain (F-box domain mediates protein-protein interactions, is associated with cellular functions such as signal transduction and regulation of the cell cyc
AOA1DS6419	No	Tlc3, extended winged-helix domain (play a role in linking tauA, tauB, and T	B-block binding subunit of TFIIC (prot	B-block binding subunit
AOA1DSXLR6	No	No	No	No
AOA1DSVW56	Protein kinase domain	Protein kinase domain	Protein kinase domain	Protein kinase domain
AOA1DSW4Q0	Zinc finger C2H2 type domain	Zinc finger C2H2-type	No	Zinc Finger C2H2 type domain

A0A3B6EIT8	coatamer subunit beta-2	2	4,13	34594	22071	25399	23876	25051	22767	27355	23898
A0A3B6SKM4	putative p23 co-chaperone	2	8,30	12640	9735	27428	24843	23299	24630	16601	24257
A0A3B5YTP0	pyrophosphate-energized vacuolar membrane proton pump	5	11,65	12689	9575	22380	15337	16981	14870	14882	15730
A0A3B6LV29	lysine-tRNA ligase	2	5,30	7867	1973	12774	14678	15144	15805	7538	15209
A0A3B5XX55	Stromal 70 kDa heat shock-related protein, chloroplastic	2	4,46	20250	19220	18718	7018	7801	6816	19396	7212
A0A3B6QGE4	phosphoenolpyruvate carboxylase 1	12	13,81	3269	7052	8997	12177	12902	12106	6439	12395
A0A3B6ING7	lipoxygenase-1	6	13,28	8045	7734	11058	12966	14871	14299	8946	14045
A0A3B6SN06	Phosphoenolpyruvate carboxylase 1	9	10,66	14253	4305	18292	20275	17372	15722	12283	17790
A0A3B6N179	pectin acetyltransferase 3-like	2	9,29	17824	18655	20126	20805	23302	18284	18868	20797
A0A3B6THT8	aspartate aminotransferase, mitochondrial	3	10,80	25127	33172	32190	26754	31324	30040	30163	29373
A0A3B6TSN8	1,4-alpha-glucan-branching enzyme, chloroplastic/amylopla	3	4,81	43501	30514	54400	22337	18100	17719	42805	19385
A0A3B5Z432	probable nucleolar protein 5-2	2	8,04	45623	12898	33042	17318	22373	22568	30521	20753
A0A3B6TNI7	fasciclin-like arabinogalactan protein 1	2	12,93	34943	43086	40699	25155	27905	25514	39576	26191
A0A3B6LID9	70 kDa heat shock protein	3	10,65	21264	28717	18794	29938	30598	30536	22925	30357
A0A3B6AXL4	succinate dehydrogenase [ubiquinone] flavoprotein subunit,	2	14,29	0	0	0	2001	1859	2290	0	2050
A0A3B6TF31	polyadenylate-binding protein 2-like	2	5,80	22031	23735	20220	16654	18810	15993	21995	17152
A0A3B6KNU7	methyl-CpG-binding domain-containing protein 10-like	2	6,88	6210	31384	31402	21758	20315	15180	22998	19084
A0A3B6D9C7	alanine aminotransferase 2	2	5,21	7020	5035	3102	0	0	0	5052	0
A0A3B6KQX1	putative aconitase hydratase, cytoplasmic	4	9,45	15955	11538	14604	13509	14600	13728	14032	13946
A0A3B6LY00	11-beta-hydroxysteroid dehydrogenase-like 5	2	12,14	5633	3129	5822	2492	3498	4344	4861	3445
A0A341UMH5	protein transport protein Sec61 subunit alpha-like	2	4,73	13356	8213	9499	13982	14970	13214	10356	14055
A0A3B6JN14	T-complex protein 1 subunit beta	2	7,40	12607	10713	15939	14860	14390	13072	13086	14107
A0A3B6RPC7	endoplasmic homolog isoform X1	4	8,63	9875	10456	21892	22435	22015	20804	14074	21751
A0A3B6PPY5	tripeptidyl-peptidase 2 isoform X1	2	2,57	15683	11612	12828	12043	13800	12481	13374	12775



Victorio et al. Supplementary Table S2. List of identified and quantified proteins of the refined wheat flour (RWF) group, based on TOP3 quantification.

Table with columns: ACCESSION, DESCRIPTION (OMICSBOX), REPORTED PEPTIDES, SEQUENCE COVERAGE, and 19 columns labeled A2, A2.2, A2.3, C2.1, C2.2, C2.3, 384.2, 384.3, 389.1, 389.2, 389.3, AVE A2, AVE C2, AVE 384, AVE 389. The table lists various proteins such as fructose-bisphosphate aldolase 1, cytoplasmic, and heat shock cognate 70 kDa protein 2-like, along with their corresponding peptide counts and sequence coverage percentages across different experimental conditions.

Table S1. WVF Functional annotation data obtained by OmicsBox software.

Table with columns: EC#, NAME, DESCRIPTION, LENGTH, FRITS, EVALUE, SIM, SEAN, GO, GOS, GO NAMES, ENZYME CODES, ENZYME NAMES, INTERPRO DO IDS, INTERPRO DO NAMES. The table lists various enzymes and their properties across multiple columns.



Victorio et al. Supplementary Table S5. Protein distribution in the three GO categories for the WWF group.

Biological Process	Count_BP	Molecular function	Count_MF	Cellular component	Count_CC
translation	32	ATP binding	48	cytosol	64
carbohydrate metabolic process	31	oxidoreductase activity	40	extracellular region	62
protein folding	30	enzyme regulator activity	33	cytoplasm	61
response to stress	25	structural constituent of ribosome	38	nucleus	57
tricarboxylic acid cycle	11	nutrient reservoir activity	44	cytosolic large ribosomal subunit	23
biosynthetic process	10	RNA binding	35	mitochondrion	22
cellular response to unfolded protein	10	DNA binding	15	cytosolic small ribosomal subunit	18
glycogen biosynthetic process	9	peptidase activity	14	endoplasmic reticulum	13
lipid metabolic process	8	hydrolase activity, acting on glycosyl bonds	13	amyloplast	12
ribosomal large subunit assembly	8	GTPase activity	11	extracellular space	11
ribosomal small subunit assembly	8	unfolded protein binding	10	chloroplast	11
catabolic process	6	protein binding	9	ribosome	9
cytoplasmic translation	6	lyase activity	9	plastid	7
glucose metabolic process	6	isomerase activity	9	Golgi apparatus	6
polysaccharide catabolic process	5	magnesium ion binding	8	vacuole	6
pyruvate metabolic process	5	chitinase activity	8	integral component of membrane	5
ubiquitin-dependent ERAD pathway	5	ion binding	7	plasma membrane	5
cellular amino acid metabolic process	4	mRNA binding	7	mitochondrial inner membrane	4
cellulose catabolic process	4	L-malate dehydrogenase activity	5	peroxisome	4
cytoplasmic translational initiation	4	1,4-alpha-glucan branching enzyme activity	5	mitochondrial proton-transporting ATP synthase complex	3
fructose 6-phosphate metabolic process	4	phosphoenolpyruvate carboxylase activity	4	lipid droplet	3
glutathione metabolic process	4	glutathione transferase activity	4	nucleolus	3
glycolytic process	4	transmembrane transporter activity	4	cytoskeleton	2
immune system process	4	translation factor activity, RNA binding	4	nucleosome	2
microtubule cytoskeleton organization	4	thioredoxin peroxidase activity	4	endoplasmic reticulum lumen	2
mitochondrial ATP synthesis coupled proton transport	4	sucrose synthase activity	4	proteasome regulatory particle, base subcomplex	2
protein refolding	4	structural molecule activity	3	chaperonin-containing T-complex	2
small molecule metabolic process	4	RNA helicase activity	4	cytosol	2
sucrose metabolic process	4	L-alanine:2-oxoglutarate aminotransferase activity	4	mitochondrial matrix	2
sulfur compound metabolic process	4	kinase activity	4	mitochondrial respiratory chain complex II, succinate dehydrogenase complex	2
reproduction	3	cellulase activity	4	protein-containing complex	1
ATP synthesis coupled proton transport	3	calcium ion binding	4	mitochondrial pyruvate dehydrogenase complex	1
cell redox homeostasis	3	6-phosphofructokinase activity	4	external encapsulating structure	1
cellular protein modification process	3	triose-phosphate isomerase activity	3	mitochondrial outer membrane	1
chromosome organization	3	threonine-type endopeptidase activity	3	protein phosphatase type 2A complex	1
dephosphorylation	3	serine-type endopeptidase inhibitor activity	3	endosome	1
generation of precursor metabolites and energy	3	manganese ion binding	3	PCNA complex	1
gluconeogenesis	3	lipid binding	3	endoplasmic reticulum membrane	1
intracellular protein transport	3	L-aspartate:2-oxoglutarate aminotransferase activity	3	endoplasmic reticulum	1
maturation of SSU-rRNA from tricistronic rRNA	3	glycogen (starch) synthase activity	3	cell wall	1
methionine biosynthetic process	3	glyceraldehyde-3-phosphate dehydrogenase activity	3	COPII vesicle coat	1
negative regulation of catalytic activity	3	beta-glucosidase activity	3	mitochondrial proton-transporting ATP synthase, stator	1
nucleobase-containing compound catabolic process	3	acid phosphatase activity	3	Golgi membrane	1
plant-type cell wall biogenesis	3	5-methyltetrahydropteroyltriglutamate-homoserine ligase activity	3	Sec61 translocon complex	1
proteasomal protein catabolic process	3	phosphatase activity	2	vacuolar membrane	1
protein targeting	3	fructose-bisphosphate aldolase activity	2		
protein transport	3	double-stranded DNA binding	2		
proteolysis	3	alditol:NADP+ 1-oxidoreductase activity	2		
starch biosynthetic process	3	aldehyde dehydrogenase (NAD+) activity	2		
translational elongation	3	2-alkenal reductase [NAD(P)+] activity	2		
transport	3	xylose isomerase activity	2		
cellular nitrogen compound metabolic process	2	UTP:glucose-1-phosphate uridylyltransferase activity	2		
cellular response to anoxia	2	translation elongation factor activity	2		
cellular response to heat	2	superoxide dismutase activity	2		
cellular response to oxidative stress	2	succinate dehydrogenase (ubiquinone) activity	2		
D-xylose catabolic process	2	protein kinase C binding	2		
fatty acid biosynthetic process	2	proline-tRNA ligase activity	2		
glycogen metabolic process	2	methyltransferase activity	2		
mRNA processing	2	malate dehydrogenase (decarboxylating) (NAD+) activity	2		
negative regulation of endopeptidase activity	2	ligase activity	2		
NLS-bearing protein import into nucleus	2	glycosyltransferase activity	2		

nucleosome assembly	2	betaine-aldehyde dehydrogenase activity	2
positive regulation of protein phosphorylation	2	ATADP antiporter activity	2
prolyl-tRNA aminoacylation	2	peroxidase activity	1
protein import into nucleus	2	pectin acetyltransferase activity	1
regulation of transcription, DNA-templated	2	orotate phosphoribosyltransferase activity	1
response to heat	2	glutamate-ammonia ligase activity	1
rRNA processing	2	glutamate 5-kinase activity	1
acetyl-CoA biosynthetic process from pyruvate	1	glucan endo-1,3-beta-D-glucosidase activity	1
anatomical structure development	1	aldose 1-epimerase activity	1
aromatic compound biosynthetic process	1	alcohol dehydrogenase activity, zinc-dependent	1
box C/D RNA 3'-end processing	1	adenylate kinase activity	1
branched-chain amino acid biosynthetic process	1	zinc ion binding	1
cell wall organization	1	water channel activity	1
cell wall organization or biogenesis	1	voltage-gated anion channel activity	1
cellular aldehyde metabolic process	1	UDP-arabinopyranose mutase activity	1
circadian rhythm	1	tRNA binding	1
cytoplasmic translational elongation	1	translation initiation factor activity	1
de novo pyrimidine nucleobase biosynthetic process	1	transferase activity, transferring alkyl or aryl	1
defense response	1	transferase activity	1
ethanol oxidation	1	transaldolase activity	1
fatty acid beta-oxidation	1	succinate-semialdehyde dehydrogenase (NAD+)	1
glucan catabolic process	1	spermidine synthase activity	1
glucose catabolic process	1	single-stranded DNA binding	1
glutamine biosynthetic process	1	signal sequence binding	1
hydrogen peroxide catabolic process	1	sarcosine oxidase activity	1
IMP salvage	1	ribulose-phosphate 3-epimerase activity	1
isoleucine biosynthetic process	1	ribonuclease activity	1
leading strand elongation	1	pyruvate dehydrogenase (acetyl-transferring)	1
L-serine catabolic process	1	protein tag	1
lysyl-tRNA aminoacylation	1	phosphogluconate dehydrogenase (decarboxylating)	1
maturation of LSU-rRNA	1	O-methyltransferase activity	1
mitochondrion organization	1	nuclear localization sequence binding	1
negative regulation of translation	1	NADPH dehydrogenase (quinone) activity	1
nuclear migration	1	methylmalonate-semialdehyde dehydrogenase	1
pentose-phosphate shunt, non-oxidative branch	1	L-ascorbate peroxidase activity	1
pentose-phosphate shunt, oxidative branch	1	lactoylglutathione lyase activity	1
phosphorylation	1	ketol-acid reductoisomerase activity	1
plant-type hypersensitive response	1	inorganic diphosphatase activity	1
polyamine biosynthetic process	1	GTPase activator activity	1
positive regulation of RNA polymerase II transcription	1	glycine hydroxymethyltransferase activity	1
positive regulation of translational fidelity	1	dihydrodipolyllysine-residue succinyltransferase activity	1
protein dephosphorylation	1	dihydrodipolyl dehydrogenase activity	1
protein ubiquitination	1	citrate (S)-synthase activity	1
proton transmembrane transport	1	chromatin binding	1
purine nucleotide metabolic process	1	catalase activity	1
regulation of COPII vesicle coating	1	AMP deaminase activity	1
regulation of gene expression	1	actin monomer binding	1
removal of superoxide radicals	1	actin filament binding	1
response to bacterium	1	acetyl-CoA C-acyltransferase activity	1
response to cold	1	2-isopropylmalate synthase activity	1
response to desiccation	1		
response to iron ion	1		
response to oxidative stress	1		
response to reactive oxygen species	1		
response to water deprivation	1		
ribosomal subunit export from nucleus	1		
ribosome biogenesis	1		
signal transduction	1		
SRP-dependent cotranslational protein targeting	1		
thymine catabolic process	1		
translational initiation	1		
transmembrane transport	1		
trehalose biosynthetic process	1		
ubiquitin-dependent protein catabolic process	1		



Victorio et al. Supplementary Table S5. Protein distribution in the three GO categories for the RWF group.

Biological Process	Count_BP	Molecular Function	Count_MF	Cellular Component	Count_CC
translation	39	ATP binding	59	cytoplasm	74
carbohydrate metabolic process	37	oxidoreductase activity	45	cytosol	72
protein folding	32	nutrient reservoir activity	49	nucleus	71
response to stress	27	structural constituent of ribosome	42	extracellular region	63
biosynthetic process	14	enzyme regulator activity	37	mitochondrion	31
glycogen biosynthetic process	13	RNA binding	37	cytosolic large ribosomal subunit	26
catabolic process	11	GTPase activity	18	amyloplast	20
glucose metabolic process	11	peptidase activity	17	cytosolic small ribosomal subunit	19
cellular response to unfolded protein	10	hydrolase activity, acting on glycosyl bonds	16	chloroplast	18
cytoplasmic translation	10	DNA binding	15	extracellular space	15
tricarboxylic acid cycle	10	lyase activity	13	endoplasmic reticulum	14
ubiquitin-dependent ERAD pathway	9	ion binding	13	Golgi apparatus	10
ribosomal small subunit assembly	8	kinase activity	11	integral component of membrane	10
cellular protein modification process	7	protein binding	11	plastid	10
immune system process	7	isomerase activity	10	ribosome	10
lipid metabolic process	7	magnesium ion binding	9	plasma membrane	9
ribosomal large subunit assembly	7	unfolded protein binding	9	vacuole	8
ATP synthesis coupled proton transport	6	1,4-alpha-glucan branching enzyme activity	8	peroxisome	5
cellular amino acid metabolic process	6	chitinase activity	8	chloroplast stroma	4
cellulose catabolic process	6	glycosyltransferase activity	7	endosome	4
fructose 6-phosphate metabolic process	6	mRNA binding	7	mitochondrial inner membrane	4
glycolytic process	6	6-phosphofructokinase activity	6	cytoskeleton	3
protein refolding	6	beta-glucosidase activity	6	endoplasmic reticulum lumen	3
intracellular protein transport	5	calcium ion binding	6	Golgi membrane	3
polysaccharide catabolic process	5	cellulase activity	6	mitochondrial proton-transporting ATP synthas	3
proteolysis	5	structural molecule activity	6	chaperonin-containing T-complex	2
pyruvate metabolic process	5	fructose-bisphosphate aldolase activity	5	endoplasmic reticulum membrane	2
small molecule metabolic process	5	aldose 1-epimerase activity	5	intracellular anatomical structure	2
sucrose metabolic process	5	glycogen (starch) synthase activity	5	lipid droplet	2
gluconeogenesis	4	L-alanine:2-oxoglutarate aminotransferase activity	5	mitochondrial outer membrane	2
microtubule cytoskeleton organization	4	L-aspartate:2-oxoglutarate aminotransferase activity	5	proteasome regulatory particle, base subcompl	2
plant-type cell wall biogenesis	4	sucrose synthase activity	5	protein-containing complex	2
proteasomal protein catabolic process	4	threonine-type endopeptidase activity	5	cell wall	1
protein import into nucleus	4	glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phospho	4	COPII vesicle coat	1
starch biosynthetic process	4	lipid binding	4	eukaryotic translation initiation factor 3 comple	1
transport	4	L-malate dehydrogenase activity	4	mitochondrial matrix	1
ubiquitin-dependent protein catabolic process	4	translation factor activity, RNA binding	4	mitochondrial pyruvate dehydrogenase comple	1
cell division	3	transmembrane transporter activity	4	mitochondrial respiratory chain complex II, succ	1
cell redox homeostasis	3	5-methyltetrahydropteroyltryglutamate-homocysteine S-meth	3	nuclear envelope	1
cytoplasmic translational elongation	3	acid phosphatase activity	3	nucleolus	1
cytoplasmic translational initiation	3	aldehyde dehydrogenase (NAD+) activity	3	nucleoplasm	1
dephosphorylation	3	betaine-aldehyde dehydrogenase activity	3	nucleosome	1
generation of precursor metabolites and energy	3	L-ascorbate peroxidase activity	3	PCNA complex	1
glucan catabolic process	3	ligase activity	3	peroxisomal matrix	1
methionine biosynthetic process	3	nuclear localization sequence binding	3	plant-type vacuole	1
mitochondrial ATP synthesis coupled proton transport	3	phosphatase activity	3	protein phosphatase type 2A complex	1
nucleobase-containing compound catabolic process	3	phosphoenolpyruvate carboxylase activity	3	vacuolar membrane	1
plant-type hypersensitive response	3	proline-tRNA ligase activity	3		
prolyl-tRNA aminoacylation	3	ribonuclease activity	3		
protein targeting	3	RNA helicase activity	3		
protein transport	3	serine-type endopeptidase inhibitor activity	3		
regulation of transcription, DNA-templated	3	superoxide dismutase activity	3		
response to reactive oxygen species	3	thioredoxin peroxidase activity	3		
signal transduction	3	translation initiation factor activity	3		
sulfur compound metabolic process	3	triose-phosphate isomerase activity	3		
trehalose biosynthetic process	3	2-isopropylmalate synthase activity	2		
branched-chain amino acid biosynthetic process	2	3-beta-hydroxy-delta5-steroid dehydrogenase activity	2		
cell wall organization	2	4-alpha-glucanotransferase activity	2		
cellular nitrogen compound metabolic process	2	acetyl-CoA C-acyltransferase activity	2		
cellular response to anoxia	2	actin filament binding	2		
chromosome organization	2	alditol:NADP+ 1-oxidoreductase activity	2		
D-xylose catabolic process	2	enoyl-[acyl-carrier-protein] reductase (NADH) activity	2		
fatty acid beta-oxidation	2	glutamine-tRNA ligase activity	2		
fatty acid biosynthetic process	2	glycine hydroxymethyltransferase activity	2		
glutamyl-tRNA aminoacylation	2	glycogen phosphorylase activity	2		
glycogen catabolic process	2	manganese ion binding	2		
glycogen metabolic process	2	methyltransferase activity	2		
L-serine catabolic process	2	pectin acetyltransferase activity	2		
maltose catabolic process	2	protein kinase C binding	2		
maturation of LSU-rRNA	2	Rab GDP-dissociation inhibitor activity	2		
maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	2	sarcosine oxidase activity	2		
mRNA processing	2	translation elongation factor activity	2		
negative regulation of catalytic activity	2	tRNA binding	2		
negative regulation of endopeptidase activity	2	UTglucose-1-phosphate uridylyltransferase activity	2		
NLS-bearing protein import into nucleus	2	voltage-gated anion channel activity	2		
phosphorylation	2	xylose isomerase activity	2		

positive regulation of protein phosphorylation	2	2-alkenal reductase [NAD(P)+] activity	1
positive regulation of translational fidelity	2	3-isopropylmalate dehydratase activity	1
protein-containing complex assembly	2	acetylornithine deacetylase activity	1
removal of superoxide radicals	2	actin monomer binding	1
response to desiccation	2	adenine phosphoribosyltransferase activity	1
small GTPase mediated signal transduction	2	adenylate kinase activity	1
steroid biosynthetic process	2	alcohol dehydrogenase activity, zinc-dependent	1
translational elongation	2	aminoacyl-tRNA editing activity	1
translational initiation	2	aminopeptidase activity	1
water transport	2	AMP deaminase activity	1
reproduction	1	ATADP antiporter activity	1
acetyl-CoA biosynthetic process from pyruvate	1	catalase activity	1
actin filament depolymerization	1	chromatin binding	1
aromatic compound biosynthetic process	1	dihydrolipoyl dehydrogenase activity	1
cell cycle	1	dihydrolipoylsine-residue succinyltransferase activity	1
cellular aldehyde metabolic process	1	dimethylallyltranstransferase activity	1
cellular amino acid biosynthetic process	1	double-stranded DNA binding	1
cellular response to heat	1	glucan endo-1,3-beta-D-glucosidase activity	1
cellular response to oxidative stress	1	glutamate 5-kinase activity	1
circadian rhythm	1	glutamate dehydrogenase (NAD+) activity	1
de novo' pyrimidine nucleobase biosynthetic process	1	glutamate-ammonia ligase activity	1
defense response	1	glutathione transferase activity	1
endosperm development	1	glycine-tRNA ligase activity	1
ethanol oxidation	1	GTPase activator activity	1
farnesyl diphosphate biosynthetic process	1	hydrolase activity	1
formation of cytoplasmic translation initiation complex	1	inorganic diphosphatase activity	1
glucose catabolic process	1	ketol-acid reductoisomerase activity	1
glutamate catabolic process	1	lactoylglutathione lyase activity	1
glutamine biosynthetic process	1	L-leucine:2-oxoglutarate aminotransferase activity	1
glutathione metabolic process	1	malate dehydrogenase (decarboxylating) (NAD+) activity	1
glycine betaine biosynthetic process from choline	1	metalloendopeptidase activity	1
glycosaminoglycan biosynthetic process	1	methylmalonate-semialdehyde dehydrogenase (acylating) act	1
hydrogen peroxide catabolic process	1	mRNA 3'-UTR binding	1
IMP salvage	1	NADPH dehydrogenase (quinone) activity	1
inorganic anion transport	1	nuclease activity	1
isoleucine biosynthetic process	1	O-methyltransferase activity	1
leading strand elongation	1	orotate phosphoribosyltransferase activity	1
leucine biosynthetic process	1	peroxidase activity	1
leucyl-tRNA aminoacylation	1	phosphogluconate dehydrogenase (decarboxylating) activity	1
L-methionine salvage from methylthioadenosine	1	protein dimerization activity	1
lysyl-tRNA aminoacylation	1	protein disulfide isomerase activity	1
mitochondrial alanyl-tRNA aminoacylation	1	protein serine/threonine kinase activity	1
mitochondrial glycyl-tRNA aminoacylation	1	protein tag	1
mitochondrion organization	1	proton-transporting ATP synthase activity, rotational mechani	1
negative regulation of translation	1	pyruvate dehydrogenase (acetyl-transferring) activity	1
nucleocytoplasmic transport	1	single-stranded DNA binding	1
nucleosome assembly	1	spermidine synthase activity	1
ornithine biosynthetic process	1	succinate dehydrogenase (ubiquinone) activity	1
pentose-phosphate shunt	1	succinate-semialdehyde dehydrogenase (NAD+) activity	1
pentose-phosphate shunt, oxidative branch	1	thiosulfate sulfurtransferase activity	1
phosphate-containing compound metabolic process	1	transaldolase activity	1
photosynthesis	1	transaminase activity	1
polyamine biosynthetic process	1	transferase activity	1
positive regulation of proteasomal protein catabolic process	1	transferase activity, transferring alkyl or aryl (other than meth	1
positive regulation of RNA polymerase II transcription preinitiation complex assembly	1	transketolase activity	1
proteasome-mediated ubiquitin-dependent protein catabolic process	1	ubiquitin activating enzyme activity	1
protein dephosphorylation	1	UDP-arabinopyranose mutase activity	1
protein folding in endoplasmic reticulum	1	UDP-glucose 6-dehydrogenase activity	1
protein phosphorylation	1	water channel activity	1
protein ubiquitination	1	zinc ion binding	1
proton transmembrane transport	1	iron ion binding	1
purine nucleotide metabolic process	1		
purine ribonucleoside salvage	1		
regulation of COPII vesicle coating	1		
regulation of gene expression	1		
response to cadmium ion	1		
response to cold	1		
response to heat	1		
response to iron ion	1		
response to oxidative stress	1		
response to water deprivation	1		
ribosomal subunit export from nucleus	1		
rRNA processing	1		
sequestering of actin monomers	1		
thymine catabolic process	1		
transmembrane transport	1		
transsulfuration	1		
vesicle-mediated transport	1		
xylan catabolic process	1		

Sample	Biological Process	PC	Percentage
Whole Wheat Flour	anatomical structure development	1	0,2%
	box C/D RNA 3'-end processing	1	0,2%
	cell wall organization or biogenesis	1	0,2%
	nuclear migration	1	0,2%
	pentose-phosphate shunt, non-oxidative branch	1	0,2%
	response to bacterium	1	0,2%
	ribosome biogenesis	1	0,2%
	SRP-dependent cotranslational protein targeting to membrane, translocation	1	0,2%
Refined Wheat Flour	actin filament depolymerization	1	0,2%
	cell cycle	1	0,2%
	cell division	3	0,6%
	cellular amino acid biosynthetic process	1	0,2%
	endosperm development	1	0,2%
	farnesyl diphosphate biosynthetic process	1	0,2%
	formation of cytoplasmic translation initiation complex	1	0,2%
	glutamate catabolic process	1	0,2%
	glutamyl-tRNA aminoacylation	2	0,4%
	glycine betaine biosynthetic process from choline	1	0,2%
	glycogen catabolic process	2	0,4%
	glycosaminoglycan biosynthetic process	1	0,2%
	inorganic anion transport	1	0,2%
	leucine biosynthetic process	1	0,2%
	leucyl-tRNA aminoacylation	1	0,2%
	L-methionine salvage from methylthioadenosine	1	0,2%
	maltose catabolic process	2	0,4%
	mitochondrial alanyl-tRNA aminoacylation	1	0,2%
	mitochondrial glycyl-tRNA aminoacylation	1	0,2%
	nucleocytoplasmic transport	1	0,2%
	ornithine biosynthetic process	1	0,2%
	pentose-phosphate shunt	1	0,2%
	phosphate-containing compound metabolic process	1	0,2%
	photosynthesis	1	0,2%
	positive regulation of proteasomal protein catabolic process	1	0,2%
	proteasome-mediated ubiquitin-dependent protein catabolic process	1	0,2%
	protein folding in endoplasmic reticulum	1	0,2%
	protein phosphorylation	1	0,2%
	protein-containing complex assembly	2	0,4%
	purine ribonucleoside salvage	1	0,2%
	response to cadmium ion	1	0,2%
	sequestering of actin monomers	1	0,2%
small GTPase mediated signal transduction	2	0,4%	
steroid biosynthetic process	2	0,4%	

transsulfuration	1	0,2%
vesicle-mediated transport	1	0,2%
xylan catabolic process	1	0,2%



Victorio et al. Supplementary Table S10. UP and DOWN-accumulated proteins when compared RWF hard NILs (384) / RWF soft NILs (389). The following filters were used: maximum CV of 0.3, p-values lower than 0.05 and log2 fold change greater than 0.6 for UP or lower than -0.6 for DOWN.

ACCESSION	DESCRIPTION (OMICSBOX)	REPORTED PEPTIDES	SEQUENCE COVERAGE	384 1	384 2	384 3	389 1	389 2	389 3	AVE 384	CV 384	AVE 389	CV 389	T-TEST 384/389	LOG2 384/389	DIF ACCUM 384/389
P29546	elongation factor 1-beta	11	32.04	88402	93928	99046	5022	3997	3062	93792	0.057	4027	0.243	0.00001	4.54	UP
AOA3B6D870	elongation factor 1-beta	11	67.89	55198	58649	61844	24695	19654	15057	58564	0.057	19802	0.243	0.00033	1.56	UP
AOA3B5Y228	heat shock 70 kDa protein 4	35	68.08	63891	88633	80630	36206	32849	32066	77651	0.164	33707	0.065	0.00416	1.20	UP
AOA3B6N0L6	ribonuclease TUDOR 1	18	25.40	33820	43338	42923	19837	22032	26386	40027	0.134	22752	0.146	0.00912	0.81	UP
AOA3B6QE25	phosphoglycerate kinase, cytosolic	26	77.56	300766	305333	317884	182490	183425	161973	307994	0.029	175963	0.069	0.00011	0.81	UP
ADA1D5WW00	16.9 kDa class I heat shock protein 1-like	6	56.95	22764	19195	20935	11718	14525	10563	20965	0.085	12269	0.166	0.00512	0.77	UP
AOA3B6SD73	DP-glucose pyrophosphorylase small subunit	27	68.08	167522	184370	182917	112308	102040	106243	176270	0.052	105064	0.065	0.00040	0.76	UP
AOA3B6R798	avenin-like b1	5	52.23	92269	68517	78012	57998	34624	47734	78933	0.163	46785	0.250	0.03306	0.75	UP
AOA3B6N0S7	protein synthesis inhibitor II-like	7	21.33	65172	73732	71320	29165	53965	43911	70075	0.063	42314	0.294	0.02188	0.73	UP
AOA3B6EHG7	cytosolic isocitrate dehydrogenase [NADP]	19	66.01	30582	28273	40736	20005	20524	19834	33197	0.200	20121	0.018	0.02701	0.72	UP
W5CSU5	40S ribosomal protein S9-2	6	25.13	26526	34720	37171	22291	21308	16055	32806	0.170	19885	0.169	0.02629	0.72	UP
AOA3B6H2L0	cationic peroxidase SPC4-like	6	23.56	9040	11590	10238	6319	5998	6808	10289	0.124	6375	0.064	0.00717	0.69	UP
AOA0C4BKHS	eukaryotic initiation factor 4A	24	64.98	152239	143863	164083	97033	109252	80010	153395	0.066	95432	0.154	0.00492	0.68	UP
AOA3B5Y1W9	pyruvate, phosphate dikinase 1, chloroplast	38	59.27	66785	66722	71077	47662	38956	41044	68195	0.037	42554	0.107	0.00102	0.68	UP
AOA3B6FY81	inc finger CCH domain-containing protein 1	6	24.40	36594	32579	35895	29497	19166	17800	34689	0.053	22154	0.289	0.03099	0.65	UP
AOA3B6G9H6	aspartic proteinase onyzasin-1-like isoform X'	16	36.52	119915	95078	101002	77864	68763	60272	103331	0.123	86307	0.130	0.01516	0.62	UP
AOA3B6DKQ2	Nucleostemin	12	32.24	41111	33406	29872	17836	25688	24211	34796	0.165	22576	0.185	0.04076	0.62	UP
AOA3B6PUC2	betaine aldehyde dehydrogenase	21	74.44	20149	19026	19056	28681	31355	28314	19410	0.033	29450	0.056	0.00061	-0.60	DOWN
AOA3B6TMA0	protein virilizer homolog	3	3.97	19470	11351	20626	26328	28642	27441	17149	0.295	27470	0.042	0.02612	-0.68	DOWN
A54ML5	Avenin-like b4	7	41.90	28027	27362	28895	42854	45088	50433	28095	0.027	46125	0.084	0.00141	-0.72	DOWN
AOA3B6JCE1	ribonuclease TUDOR 1	43	57.49	57585	62327	59805	110764	105237	100218	59906	0.040	105406	0.050	0.00017	-0.82	DOWN
AOA3B6SDG4	serpin 2	8	30.77	20352	19665	21554	34775	43373	34683	20524	0.047	37610	0.133	0.00433	-0.87	DOWN
AOA07753V2	aspartate aminotransferase, cytoplasmic	26	67.40	22517	24073	22775	48369	47153	48998	23122	0.036	48173	0.019	0.00000	-1.06	DOWN
G9XGB3	starch branching enzyme 1	47	58.02	10437	12381	11549	22722	26928	24934	11456	0.085	24861	0.085	0.00366	-1.12	DOWN
AOA3B6C068	Vacuolar-sorting receptor 1	21	42.25	66213	51751	49237	87053	139036	140994	55734	0.164	122361	0.250	0.02249	-1.13	DOWN
AOA3B5ZVW3	Chitinase 2	9	55.71	5215	4273	3449	12727	10532	7836	4312	0.205	10365	0.236	0.01679	-1.27	DOWN