Peptidomics analysis of in vitro digested wheat breads: Effect of genotype and environment on protein digestibility and release of celiac disease and wheat allergy related epitopes

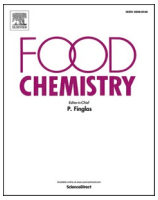
Mélanie Lavoignat, Angéla Juhász, Utpal Bose, Thierry Sayd, Christophe Chambon, Miguel Ribeiro, Gilberto Igrejas, Sébastien Déjean, Catherine Ravel, Emmanuelle Bancel

Food Chemistry 448 (2024) 139148

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem



Peptidomics analysis of in vitro digested wheat breads: Effect of genotype and environment on protein digestibility and release of celiac disease and wheat allergy related epitopes

Mélanie Lavoignat [a], Angéla Juhász [b], Utpal Bose [b,c], Thierry Sayd [d], Christophe Chambon [d], Miguel Ribeiro [e,f], Gilberto Igrejas [f,g,h], Sébastien Déjean [i], Catherine Ravel [a,*], Emmanuelle Bancel [a]

[a] Université Clermont Auvergne, INRAE, UMR1095 GDEC, F-63000 Clermont-Ferrand, France
[b] Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, Edith Cowan University, School of Science, 270 Joondalup Dr, Joondalup, WA 6027, Australia
[c] CSIRO Agriculture and Food, 306 Carmody Rd, St Lucia, QLD 4067, Australia
[d] INRAE, Plateforme d'Exploration du Métabolisme Composante Protéomique (PFEMcp), F-63122 Saint-Genès Champanelle, France
[e] Chemistry Research Centre-Vila Real (CQ-VR), University of Trás-os-Montes and Alto Douro, 5000-801 Vila Real, Portugal
[f] Department of Genetics and Biotechnology, School of Life Sciences and Environment, University of Trás-os-Montes and Alto Douro, 5000-801 Vila Real, Portugal
[g] Functional Genomics and Proteomics Unit, University of Trás-os-Montes and Alto Douro (UTAD), Vila Real, Portugal
[h] Associated Laboratory for Green Chemistry (LAQV-REQUIMTE), University NOVA of Lisboa, 1099-085 Lisboa, Portugal
[i] Institut de Mathématiques de Toulouse, UMR5219, Université de Toulouse, CNRS, UPS, F-31062 Toulouse Cedex 9, France

ARTICLE INFO

Keywords:
Triticum aestivum
Wheat
Immunogenic epitopes
In vitro digestion
Mass spectrometry

ABSTRACT

Wheat proteins can trigger immunogenic reactions due to their resistance to digestion and immunostimulatory epitopes. Here, we investigated the peptidomic map of partially digested bread samples and the fingerprint of epitope diversity from 16 wheat genotypes grown in two environmental conditions. Flour protein content and composition were characterized; gastric and jejunal peptides were quantified using LC-MS/MS, and genotypes were classified into high or low bread protein digestibility. Differences in flour protein content and peptide composition distinguish high from low digestibility genotypes in both growing environments. No common peptide signature was found between high- and low-digestible genotypes; however, the celiac or allergen epitopes were noted not to be higher in low-digestible genotypes. Overall, this study established a peptidomic and epitope diversity map of digested wheat bread and provided new insights and correlations between weather conditions, genotypes, digestibility and wheat sensitivities such as celiac disease and wheat allergy.

1. Introduction

Wheat protein content and composition are critical determinants of grain techno-functional properties and health quality indicators. Roughly, wheat grain proteins consist of about 20% of albumins-globulins and 80% of storage proteins (or gluten proteins) (Shewry, 2019).

Techno-functional properties of the wheat grain primarily rely on the storage proteins, consisting of about 50–60% of gliadins and 40–50% of glutenins (Johansson et al., 2013). Gliadins are alcohol-soluble monomeric proteins classified into α/β-, γ- and ω-gliadins. Glutenins, classified into low molecular weight (LMW) and high molecular weight (HMW) subunits are polymeric proteins that form intra- and intermolecular disulphide bonds and hydrophobic interactions to stabilize their structure. After flour hydration and kneading, glutenins and gliadins form the polymeric gluten network responsible for the dough's rheological properties (Johansson et al., 2013).

Wheat grain proteins can trigger IgE-mediated wheat allergy (WA) and celiac disease (CD), an autoimmune disorder induced by gluten proteins to genetically susceptible individuals with the human leukocyte antigens HLA-DQ2 and/or HLA-DQ8 haplotypes (Brouns, van Rooy, Shewry, Rustgi, & Jonkers, 2019). Gluten proteins are characterized by unique repetitive domains that account for 30% to >85% of the protein length. These domains consist of repetitions of peptide sequences that

* Corresponding author.
E-mail address: catherine.ravel@inrae.fr (C. Ravel).

https://doi.org/10.1016/j.foodchem.2024.139148
Received 13 November 2023; Received in revised form 26 February 2024; Accepted 24 March 2024
Available online 26 March 2024

are three to nine amino acids long and rich in proline and glutamine (Shewry, 2019). The high repetition of glutamine and proline content in gluten proteins makes them partially resistant to proteolytic cleavage by human digestive enzymes. The resulting undigested gluten epitopes can cross the small intestinal barrier and trigger immune reactions. The deamidation of these peptides by transglutaminase 2 (TG2) in the gut can enhance the binding affinity of these peptides to the antigen-presenting heterodimers HLA-DQ2 or HLA-DQ8 and activate the CD4+ T cells (Brouns et al., 2019). This activation leads to the production of anti-gluten and anti-TG2 antibodies and cytokines such as IL-15 to damage the intestinal epithelial cells, leading to celiac enteropathy. In addition, plant defence-related proteins such as alpha-amylase trypsin inhibitors (ATIs), wheat germ-agglutinins, and serpins present in wheat can also trigger immune responses in some people (Aziz, Hadjivassiliou, & Sanders, 2015; Zevallos et al., 2017). Non-celiac wheat sensitivity (NCWS) is increasingly reported by patients with intestinal and extra-intestinal responses. Patients diagnosed with NCWS test negative for CD and WA, but the key cause for this disease symptoms is currently unknown (Brouns et al., 2019), and the role of gluten proteins in NCWS pathology remains elusive. Although wheat consumption can trigger adverse immune reactions in a small part of the population, most of the human population still depend on wheat as their primary diet. As such, it is important to understand the immunostimulatory peptide contents precisely so that the majority of people can meet their regular dietary requirements.

Studies have reported quantitative and qualitative differences in epitope content and composition between the genotypes of different *Triticum* species (Ribeiro et al., 2016) and *T. aestivum* varieties (Denery-Papini et al., 2007; Prandi, Tedeschi, Folloni, Galaverna, & Sforza, 2017; Ronga et al., 2020; Schalk, Lang, Wieser, Koehler, & Scherf, 2017). The environmental variability of epitope-containing proteins has been investigated (Juhász et al., 2018; Juhász, Haraszi, & Békés, 2020; Landolfi et al., 2021; Ronga et al., 2020) and more broadly, the influence of environment on protein composition and content (and thus impacting epitopes) has been extensively studied (Johansson et al., 2020). In addition, several studies have shown the immunostimulatory potential of wheat products due to the presence of IgE-binding WA epitopes and immunogenic celiac epitopes (Denery-Papini et al., 2007; Schalk et al., 2017). In addition to the availability of >16 high-resolution bread wheat genome sequence resources, advancements in mass-spectrometry (MS)-based proteomics and bioinformatics workflows have allowed us to understand better the diversity of immunostimulatory peptides involved in wheat-related disorders. The MS-based proteomics approach can precisely identify and quantify immunogenic epitopes from complex flour samples and processed food products (Ribeiro et al., 2021). The identification and quantitation of epitopes in the experimental samples depend on the precise characterization of sub-classes of proteins upon digestion (Landolfi et al., 2021; Ogilvie et al., 2020). For instance, α-gliadins generate the most diverse CD-related epitopes (Landolfi et al., 2021), and these peptides were reported among the most immunodominant (Juhász et al., 2018; Tye-Din et al., 2010). However, the γ-gliadins have been shown to generate the largest number of CD epitopes during digestion (Ogilvie et al., 2020). Two studies monitored the release of immunogenic peptides during bread (Ogilvie et al., 2020) or pasta (Mamone et al., 2015) digestion and showed that a few immunogenic peptides survived the *in vitro* digestion, i.e., they resisted to proteolysis.

The duration and intensity of protein hydrolysis within the gastro-intestinal tract can increase the generation of shorter peptides, resulting in a lower immunostimulatory effect in the gut. Protein digestibility can be influenced by several factors, including human gastro-intestinal conditions (Torcello-Gómez et al., 2020), the complexity of the food matrix (Freitas, Gómez-Mascaraque, & Brodkorb, 2022), composition (Wu, Taylor, Nebl, Ng, & Bennett, 2017) and the food processing steps such as baking. Heat treatments applied during baking or cooking can alter the protein structure by unfolding the proteins through subsequent re-arrangements of disulfide bonds (Ogilvie et al., 2021; Pasini, Simonato, Giannattasio, Peruffo, & Curioni, 2001). Contradictory results have been published on the effect of baking on the protein digestibility of wheat. According to Bredariol, Carvalho, and Vanin (2020), there are optimum time and temperature baking parameters that would improve proteolysis during digestion. In addition, the impact of protein digestibility varied considerably for the bread crumb and crust; the crumb portion showed similar digestibility to the uncooked dough, while the bread crust had lower digestibility (Pasini et al., 2001). The thermal treatment also influences the recognition of WA epitopes by the IgE receptors (Lupi et al., 2019). This study reported that the thermal treatment initially reduces gliadin recognition by IgE due to the production of large aggregates. However, upon hydrolysis under acidic stomach conditions, the epitopes can be unmasked and recognized by the T cells in the intestine. From flour to processed and baked products, epitopes can undergo modifications that would alter their immunostimulatory potential (Liu et al., 2023). Thus, to study the diversity of peptides released from wheat-based food during human digestion, the choices of the food matrix as well as the digestion model are essential. In addition to wheat flour's baking or cooking process, the protein digestibility varies across wheat genotypes. Notably, the protein compositional variability in the grain flour and, thus, breads between genotypes have shown a different spectrum of digestibility (Lavoignat et al., 2022). Although many studies have reported the impact of the baking process and genotype-dependent variations that influence the digestibility and release of immunogenic epitopes, the combination of genotype and environmental variations and their food processing on the human food digestibility assessment remains unknown.

The present study aims to investigate how wheat genotype-dependent variability can impact the bread protein digestibility and epitope release during the human digestion process. We hypothesized that peptides with a specific signature could discriminate between high- and low-digestible genotypes. First, we performed the *in vitro* digestion assay on the bread samples and collected and measured the peptides generated during the gastric and intestinal digestion process. The measured peptide abundances were associated with the flour composition and protein content to separate the genotype with high digestibility (HD) and low digestibility (LD). Moreover, we detected and comprehensively mapped the epitopes identified during the digestion process to the different sub-types of gliadins, glutenins and albumins-globulins using an *in-silico* approach. Together, this study establishes the peptidomic map of partially digested bread samples and the fingerprint of epitope diversity from 16 wheat genotypes grown at two different environmental conditions.

2. Materials and methods

The plant material and wholemeal flour phenotyping were previously described in (Lavoignat et al., 2023).

2.1. Meteorological data

In total, 16 bread wheat genotypes were grown in the field in two different locations: Clermont-Ferrand (CF) and Estrées-Mons (EM) in duplicates. To assess the weather pattern for CF and EM, the maximal (Tmax), minimal (Tmin) and mean (Tmean) temperatures (°C) and cumulative rainfall (CumR, mm) of the 2016–2017 growing season and the last 30 years (Tmax30, Tmin30, Tmean30, CumR30) were retrieved (https://agroclim.inrae.fr/). The cumulative growing degree days (CumDD_MJJ) and the cumulative rainfall (CumR_MJJ) over the grain filling period, i.e., May to July, were calculated for two locations based on the heading date of genotype.

2.2. R5 competitive ELISA assay on wholemeal flour

For the 64 grains samples (16 genotypes × 2 locations × 2

replicates), CD-specific epitopes were quantified in wholemeal flours using the RIDASCREEN® Gliadin competitive enzyme immunoassay (ELISA) (R-Biopharm AG, Darmstadt, Germany). This enzyme immunoassay is based on an R5 antibody recognizing QQPFP, QQQFP, LQPFP, QLPFP, and QLPYP epitopes from the flour samples. The ELISA experiment was conducted following the manufacturer's guidelines (Ribeiro et al., 2016). Samples were diluted in a 1:64,000 ratio with a dilution buffer to fit into the calibration range. Immunoassay data are presented as mean over two technical replicates.

2.3. Breadmaking

Breadmaking was previously described in (Lavoignat et al., 2022). Briefly, breads were produced using a short yeast-leavened fermentation. They were made from 250 g of dough, consisting of flour, yeast, salt, ascorbic acid and alpha-amylase mixed with water, according to the standard French breadmaking AFNOR test (NF V03–716: 2015–12 – soft wheat flour – breadmaking tests for standard French bread).

2.4. Dynamic in vitro digestion of breads

The *in vitro* digestion was previously described in (Lavoignat et al., 2022). Briefly, the TNO gastrointestinal tract Model TIM-1 (TNO, Zeist, The Netherlands) was programmed to simulate the digestion of a solid meal in a healthy human adult as described by Blanquet-Diot et al. (2012) with slight modifications. Meals were constituted of slices of frozen bread (56 g) crushed in 73.3 ml of simulated saliva (877 $mg.ml^{-1}$ NaCl, 477 $mg.ml^{-1}$ KCl, 816.4 $mg.ml^{-1}$ KH2PO4, 441 $mg.ml^{-1}$ CaCl2.2H2O, 5.2 $g.ml^{-1}$ NaHCO3.2H2O, pH adjusted to 6.95 with HCl) and 200.7 ml of mineral water. Before initiating the digestion, 10,290 U of alpha-amylase (α-Amylase from *Bacillus* sp., ref. 10,069, Sigma-Aldrich, St-Louis, MO, USA) was mixed with 300 g of meal for 30 s before introducing the mixture into the stomach compartment. The digestion was stopped after 2 h.

2.5. Peptide extraction and analysis by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS)

Gastric and jejunal digestates were collected from two *in vitro* digestion replicates and used for peptidomics analysis after 2 h of digestion. Trichloroacetic acid (TCA) was added to the gastric and jejunal digests at a final concentration of 15% in ice to precipitate proteins. The supernatant was collected after centrifugation (4000 ×*g*, 15 min, 4 °C). Then, a solid phase absorption step was used to extract peptides and remove TCA from the supernatant using a PierceTM C18 spin column (Thermo Fisher Scientific, Villebon-sur-Yvette, France) following the manufacturer's instructions. The collected eluent fraction was dried using a SpeedVac Vacuum Concentrator. The dried digest samples from the gastric and jejunal compartments were reconstituted in 25 µl of an acidic solution (0.05% Trifluoroacetic acid) containing 1.15 pmoles of isotopologue peptides (Premix, Promega). After a sonication, the samples were transferred to HPLC vials, and 5 µl of digestates was injected into the LC-MS/MS system: First, peptides were washed on a loading column (Acclaim, PepMap, 300 µm × 0.5 cm, Thermo Fisher Scientific, Villebon-sur-Yvette, France) at a 30 $µl.min^{-1}$ flow rate. Peptides were separated at 300 $nl.min^{-1}$ on a nano HPLC column (Acclaim PepMap RSLC 75um x 15 cm, Thermo Fisher Scientific, Villebon-sur-Yvette, France) with a gradient of 4–25% acetonitrile (99.9%, 0.1% formic acid) for 50 min. The eluted peptides were electrosprayed into the LTQ velos Orbitrap nanosource (Thermo Fisher Scientific, Villebon-sur-Yvette, France). Peptides were detected in MS in a mass window between 350 and 1400 *m/z* and then analyzed in MS/MS in top-10 mode (one full scan and 10 MS/MS on the most intense ions of the full scan), with dynamic exclusion activated (30 s).

The acquired raw mass spectra were imported into the label-free quantification Progenesis QI software (Nonlinear Dynamics, Waters). The data analysis workflow included an ion map alignment step, followed by a peak picking and filtering step applying a normalization coefficient calculated based on the ionic signals of the isotopologue peptides added in equal quantities to each sample (1.15 pmoles, 0.230 pmoles injected). Each detected ion was quantified based on experimental comparisons.

2.6. Identification and quantification of peptides and proteins

Detected MS/MS ions were quantified by the Progenesis QI software through MASCOT (v2.5.1) search against an in-house developed database consisting of 321,037 wheat protein sequences [GluPro (Bromilow et al., 2017) and UniRef100 databases limited to the *Triticum* genus]. The search parameters were as follows: no digestion enzyme, MS mass tolerance of 10 ppm for the parent ion and 0.5 Da for the fragment ions. The possible oxidation of methionine residues and the possible deamidation of asparagine and/or glutamine residues were also included as search parameters. Peptides were identified at a 99% confidence level (false discovery rate of <1%).

The identification results were re-imported into Progenesis IQ software to identify and map the quantified ions. The presence of two unique peptides per protein not shared with other protein accession was considered to report a validated protein identification. The relative quantitative value of each protein was calculated by summing the abundance of all its unique peptides (Relative Quantification using non-conflicting peptides). The normalized abundance of peptides and proteins, based on the isotopologue peptides present in each digest (cf section 2.5.), was exported to LCProQI for statistical analysis.

2.7. Statistical analysis

Statistical analyses were performed using the software program R version 4.0.5 (R Core Team 2021). To evaluate the impact of temperatures and rainfall on grain protein content and composition, we calculated Pearson correlation values between the phenotypic traits and CumDD_MJJ and CumR_MJJ. In addition, we applied a forward stepwise regression model testing the impact of CumDD_MJJ and CumR_MJJ on flour protein content and composition using the stepAIC function in the R package MASS for each flour phenotypic trait.

Bread partial protein digestibility after 2-h digestion was expressed as the proportion of digested nitrogen out of total nitrogen quantified in the different compartments (gastric and the three segments of the small intestine in humans, duodenum, jejunum and ileum). This quantitative variable classified genotypes into high (HD) and low (LD) digestibility. For each environment, the HD group comprised genotypes above the median plus 2% of the median. On the contrary, the LD group comprised genotypes below the median minus 2% of the median. For the subsequent analyses, three quantitative datasets (flour protein composition, gastric peptide content and jejunal peptide content) and one qualitative variable, the group of digestibility, were used (**Table S1**).

We performed multivariate analyses considering CF and EM growing environments separately to investigate the relationship between the three datasets and identify variables that discriminate the two digestibility groups. Firstly, we performed Principal Component Analysis (PCA), Partial Least Squares (PLS) and PLS Discriminant Analysis (DA) using the mixOmics R package (Rohart, Gautier, Singh, & Cao, 2017) on each of the three quantitative datasets separately (PCA and PLS-DA) or by considering two datasets at a time (PLS). Secondly, we performed a multi-block PLS-DA to extract common information from the three quantitative datasets that best separate the high- and low-digestibility groups (Singh et al., 2019). Based on the first analyses, the multi-block PLS-DA parameters were fixed to two components and a design matrix filled with values of 0.5. These values represent the strength of the relationship to be modelled between two data frames that range from 0 to 1. In addition, considering that there is a high imbalance between the number of variables of the three datasets, we applied variable

selection using the sparse parameter, which is based on a lasso penalisation applied on the loading vectors. Following these criteria, we decided to keep 14 variables describing flour protein composition (**Table S1**) and the sparse parameter selected on each component and for each environment, 50 gastric and 10 jejunal peptides that best discriminate the two groups of digestibility. Using the sparse function to select a subset of variables was necessary, considering the number of quantified peptides. As it is possible that the selected peptides do not reflect the diversity of the entire peptides pool, we compared the proportion of the protein family and the epitope distribution of the 987 gastric and 117 jejunal peptides with the selected gastric and jejunal peptides.

We defined the peptides associated with HD and LD genotypes as HD and LD peptides, respectively. Student tests were performed to compare their mass and sequence length distribution per environment after verifying the normality using a Shapiro test. The amino acid composition of HD and LD peptides was characterized, i.e. we calculated the proportion of each amino acid in the HD and LD peptide sequences.

Hierarchical clustering and heatmaps using Euclidean distance and Ward aggregation methods were performed to identify genotypes by environment samples with similar peptide profiles.

A two-factor ANOVA was applied to the mean of the two technical replicates to test the genotypic and environmental effects on the quantity of immunogenic celiac motifs recognized by R5 epitopes. The correlation between the quantity of immunogenic celiac motifs recognized by R5 epitopes in flours and the abundance of T cell epitopes in the two digestate compartments was tested. The abundance of T cell epitopes was calculated by multiplying a peptide abundance by its number of mapped T cell epitopes and summing the epitopes abundance by individual.

2.8. Bioinformatics analysis

Detailed bioinformatics analyses were performed on the protein sequences identified from the LC-MS/MS analyses. First, conserved protein domains were identified using the HMMER3 package v3.1b1 in CLC Genomics Workbench v22.0.2 (Qiagen, Aarhus, Denmark). Signal peptide prediction was performed using SignalP (Bendtsen, Nielsen, von Heijne, & Brunak, 2004), and secondary structure elements (α-helices and β-sheets) were predicted using a built-in hidden Markov model (HMM) algorithm in CLC Genomics Workbench. Cysteine residues were also annotated within the sequences. Peptides identified from the mass spectrometry analysis were marked based on their origin (gastric or jejunal), low or high digestibility (HD, LD) and trial site (CF or EM) and mapped to the identified protein sequences using 100% sequence identity.

Epitope sequences representing celiac disease-specific T cell core epitopes (Sollid et al., 2020) were collected and linear, allergy-related B cell epitopes with positive assay results were downloaded from the Immune Epitope Database and Analysis Resource (IEDB, https://www.iedb.org/). Allergy-related epitopes were tagged based on the represented allergens. Epitope sequences were used for an *in-silico* epitope mapping using 100% sequence identity. Sequences representing the gastric and jejunal peptides were extracted along with their annotations and used to identify intact and partially digested epitope profiles. Data visualization was completed using the Morpheus JavaScript matrix visualization and analysis tool (https://software.broadinstitute.org/morpheus/), and the annotated sequences were visualized in CLC Genomics Workbench v.22.0.2.

3. Results and discussion

3.1. Assessment of the impact of different climatic conditions on grain total protein content and major subtype changes

During the growing seasons, the Tmin recorded in May 2016–17 (the heading month) at CF and EM was similar, i.e., 9.8 and 9.9 °C, respectively. However, the Tmax was 1 °C higher at CF than EM (**Table S2**). Differences were more significant during the grain-filling period in June and July, where recorded temperatures were 2 °C higher at CF. The cumulative rainfalls were also noted to differ between CF and EM; during May and June, the rainfall was 3- and 2-fold higher in CF location, respectively (**Fig. S1**).

Considering the environmental effect on flour protein content and composition, we investigated the link between flour protein content and composition of different cultivars and their peptide digestibility for each growing environment separately. The differences in temperatures and cumulative rainfall at CF and EM could explain the protein content and composition differences reported in (Lavoignat et al., 2022). Aligning with our findings, earlier studies also reported that the influence of temperature and water availability could alter grain protein content (Johansson et al., 2020). The protein content and the proportion of gliadin in the present study were 2% and 10% higher at CF than at EM. The stepwise regression model revealed that the flour protein content and the gliadin to glutenin ratio were significantly associated with CumR_MJJ and CumDD_MJJ ($p < 0.001$, $R^2_{adj} = 0.39$ and $p < 0.01$, $R^2_{adj} = 0.30$). Notably, higher temperatures after anthesis were associated with higher grain protein content and proportion of gliadins (Johansson et al., 2020). The higher temperatures in June and July at CF could explain the higher protein content of the grains and increased proportion of gliadins. In addition to total protein content and gliadin ratio changes, the HMW to LMW glutenin ratio was 0.08 higher at EM (i.e. +24.2%). The stepwise regression model also unveiled that CumR_MJJ has a significant effect on controlling the HMW to LMW ratio ($p < 0.01$, $R^2_{adj} = 0.28$). The limited cumulative rainfall during the grain filling period (**Fig. S1**) may have positively impacted the HMW glutenin accumulation for the samples at EM. A previous study has reported that the HMW glutenin content was increased in the spring-cultivated wheat under drought conditions applied at the end of spikelet initiation (Zhang et al., 2013). Notably, the quantity of the R5 epitope in the flour was not associated with CumR_MJJ and CumDD_MJJ, whereas it was significantly influenced by the environment ($p < 0.05$). Previous studies have shown a link between environmental effects and immunostimulatory peptide amount estimated by R5 epitope (Juhász et al., 2018; Landolfi et al., 2021) or by known immunostimulatory peptide sequences (Ronga et al., 2020).

Nitrogen fertilizer can influence the content and composition of storage protein (Johansson et al., 2020), thus impacting the proportion and number of gluten-specific celiac and allergenic epitopes (Juhász et al., 2020; Landolfi et al., 2021). In contrast to our findings, a study based on six durum wheat genotypes grown in four sites for two years showed that total rainfall during grain filling was negatively correlated to immunogenic and toxic epitope content (Ronga et al., 2020). These different results may be explained by different periods of rainfall considered (from heading to harvest versus from heading +14 d to harvest) as the timing of a stress event strongly affects the protein composition, hence quality and immune-responsive protein content. In addition, a study based on three genotypes grown in pots reported that high temperatures after anthesis influenced the quantity of immunoreactive proteins and R5 and G12 antibodies reactivity (two antibodies recognizing immune reactive proteins) (Juhász et al., 2018). In agreement with Ribeiro et al. (2016), the total flour protein content could explain differences in the quantity of R5 epitopes between CF and EM (**Table S1**), as the environment did not significantly impact the quantity of immunogenic motifs expressed per gram of protein.

3.2. Assessment and characterization of peptides detected from gastric and jejunal digestion

Sixteen genotypes differing for their period of release, HMW glutenin alleles and flour protein polymer characteristics were selected among a previously described collection of 75 cultivars (Lavoignat et al., 2023).

Grains of each genotype-environment sample were milled into flour and processed into yeast-leavened breads that were *in vitro* digested for two hours with the dynamic TIM model. The *in vitro* digestion assay, followed by peptidomic analysis on the gastric and jejunal digests, resulted in the detection of 987 and 117 peptides across all bread samples, respectively (**Table S3**). None of the peptides were detected at both the gastric and jejunal phases. Within the detected peptides, 557 gastric (56.4%) and 117 jejunal peptides assigned to a protein accession were reported as unique evidence for a protein. The measured mass range for the detected peptides from gastric and jejunal samples (1104) was between 760 and 4104 (*m/z*), and the mean mass in the gastric and the jejunal digestates was 1828 and 1927 (*m/z*), respectively. The detected peptide sequence length varied from 7 to 39 amino acids; the average lengths were 16 and 18 for gastric and jejunal peptides, respectively (**Table S3**). Peptides markedly differed based on their abundance. For instance, the measured mean peptide abundance intensities were from 6,6 to 1,7 $\times$ 10^5 for the gastric digestates and 1,3 to 1,8 $\times$ 10^4 for the jejunal digestates. For both locations, the mean peptide abundance in the jejunal digestates was more than six-fold lower than in the gastric digestates. The abundance differences can be explained by the fact that the samples were collected for peptidomic analysis after two hours of digestion. As a result, only a few peptides may have entered the intestine, similar to *in vivo* conditions. Gastric digestion can vary from 15 min to three hours, and intestinal digestion from two to five hours. In addition, two hours coincide with the gastric phase of the *in vitro* static model, such as the INFOGEST protocol (Minekus et al., 2014). Two other explanations may account for these abundance differences. First, the dilution factor which was more important in the intestinal compartment than in the gastric (Minekus et al., 2014). Secondly, the presence of multiple endo and exo peptidases in the pancreatic juice added in the intestinal compartment induced increased cleavage of proteins and peptides from the gastric compartment into peptides smaller than 5 amino acids down to free amino acids, which were not detectable using our mass parameters. Furthermore, peptides also vary depending on their amino acid composition. Out of 987 gastric peptides, 25 have >80% glutamine, and 213 had no phenylalanine, tyrosine and tryptophan, the preferential cleavage sites of pepsin (**Table S3**). Taken together, these findings suggest that the peptide size, length, amino acid compositions and their abundances affect the digestibility and are different between the two digestion phases. While the dynamic *in vitro* TIM model realistically mimics what could occur after two hours of human digestion, screened peptides can, however, differ from that typically released during physiological human digestion as the TIM-1 model does not contain the whole intestinal environment that releases proteins and peptides (for example intestinal cells). In addition, the meal was only composed of bread, whereas other food could be digested simultaneously, leading to other generated food protein.

3.3. Epitope mapping analysis of the detected gastric and jejunal peptides

Peptide and epitope mapping analysis showed that a significant portion (> 60%) of the detected gastric and jejunal peptides were harbored in storage protein sequences with known effects in allergy and celiac disease responses (**Table S4 and S5**). Additionally, the detected gastric peptides were also mapped to food allergens representing serpins, seed storage globulins, phosphopyruvate hydratase, phosphoglycerate kinase, fructose bisphosphate aldolase and a Barwin-Birk domain-containing protein. The jejunal peptides were primarily detected from the repetitive regions of storage proteins (primarily HMW glutenins, γ-, α- and ω-gliadins) and glycine-rich proteins. Additionally, these peptides were less characteristic at LMW glutenins, avenin-like proteins or the other detected proteins. The number of gastric and jejunal peptides with mapped celiac or allergen epitopes is summarized in Table 1.

Epitope analysis of the identified gastric and intestinal peptides showed that many of the detected celiac disease-related epitopes were intact in the gastric phase and were only digested in the intestinal phase, as evidenced by the partial epitope sequences in the jejunal peptides (Fig. 1a). The gastric peptides were enriched in intact, γ-gliadin-specific T cell core epitopes (Fig. 1a, **Table S4**). Intact versions of epitopes representing highly immunogenic α- (PQPQLPYPQ, PFPQPQLPY, PYPQPQLPY, FRPQQPYPQ) and ω-gliadin sequences (QQPQQPFPQ, QQPFPQQPQ, LQPQQPFPQ) were also detected from the small intestine (Juhász et al., 2018; Tye-Din et al., 2010). Notably, there was a significant Spearman correlation ($p = 0.04$, rho = 0.37) between the quantity of R5 epitopes in wholemeal flours and the abundance of T cell epitopes in the jejunal compartment.

Table 1
Number and percentage of gastric and jejunal peptides mapped to celiac or allergen epitopes.

Compartment	Celiac T cell				Food allergen			
	Intact		Partial		Intact		Partial	
Gastric + Jejunal (n = 1104)	33	3%	35	3%	168	15%	250	23%
Gastric (n = 987)	27	3%	11	1%	80	8%	174	18%
Jejunal (n = 117)	6	5%	24	21%	88	75%	76	65%

Analysis of major wheat allergy-related epitopes representing allergens like Tri a 19, Tri a 20, Tri a 21, Tri a 33 and Tri a 36 shows that a large proportion of the HMW glutenins (Tri a 26) and a fraction of ω5-gliadin (Tri a 19) and γ-gliadin (Tri a 20) allergen epitopes can be detected in an intact form both from the gastric and jejunal fractions (Fig. 1b), while α-gliadin (Tri a 21) and serpin-Z2 allergen epitopes (Tri a 33) get digested in the gastric phase (**Table S5**).

3.4. Assessment of the diversity of gastric and jejunal peptides across different genotypes

The hierarchical clustering analysis on the 987 gastric peptides showed that most of the samples were clustered based on the genotypes except for four genotypes, W-33, W-44, W-53 and W-59 (**Fig. S2a**). This finding highlights that gastric peptide abundance is primarily linked to genotype-dependent variations rather than environment. The clustering analysis on the jejunal peptides showed the opposite trend, where the cultivars were separated into two major clusters primarily driven by their growing environments (**Fig. S2b**). One group was composed of 15 genotype-environment samples, exclusively grown at EM, except CFW-39 and CFW-40, two isogenic lines diverging for the transcriptional factor Ta-NAMB1 reported to increase protein content without penalizing the total yield potential. The second cluster comprised 17 genotype-environment samples, exclusively grown at CF, except EMW-53, EMW-59, and EMW-74. The specific classification of these five genotypes based on the measured peptide abundance represents a genotypic x environment interaction. This could highlight that regardless of the growing environment, the abundance pattern for the jejunal peptides showed a steady pattern between genotypes. The differential classification by genotype based on gastric peptides and by environment based on jejunal peptides was unexpected. One hypothesis for these differences may be linked to the type and characteristics of peptides detected from the two digestates. The peptides mapped to the protein family were illustrated in Fig. 2. The jejunal peptides were almost exclusively derived from storage protein, more specifically 22 (CFW-74) to 80% (EMW-04) of the peptides originating from HMW glutenins. The protein family origin of the gastric peptides is more diverse; 42% of the peptides originated from gliadins, 27% from HMW or LMW glutenins, 7% from serpins, and 6% from alpha-amylase/trypsin inhibitors. It is well known that environmental factors influence storage protein composition. Thus, the gastric and jejunal peptides containing proteins detected during digestion may be sensitive to environmental changes, such as temperature or rainfall conditions. Furthermore, within the same growing location, there was a peptide abundance diversity among

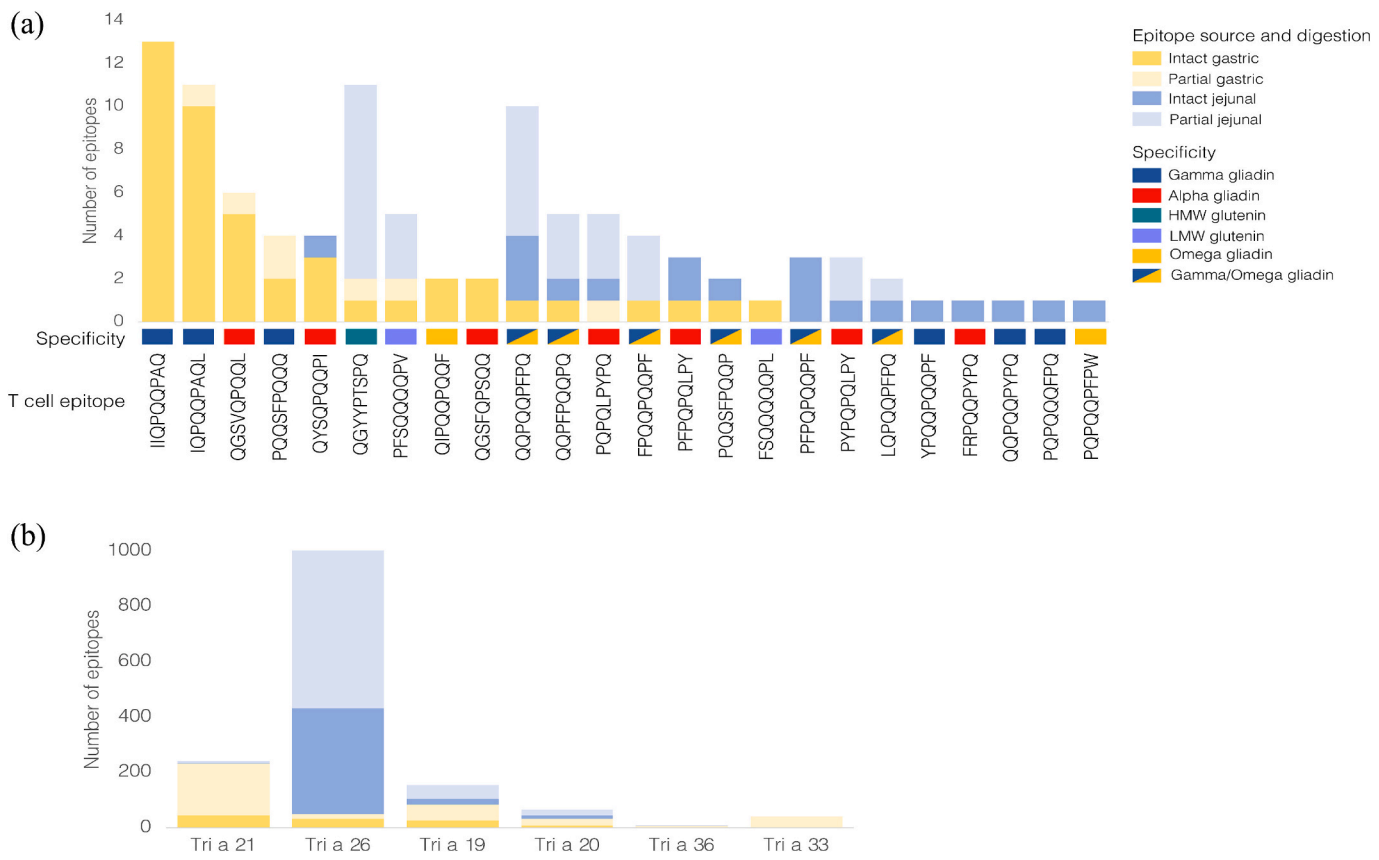


Fig. 1. Intact and partially digested coeliac disease-specific T cell epitope (a) and wheat allergy-related epitopes (b) distribution in the gastric and jejunal peptides. The stacked bar charts represent intact and partially digested epitopes detected in the gastric phase (labelled in yellow colours) and intestinal phase (labelled with blue colours). The specificity of the detected epitopes is also labelled with gluten protein type. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

genotypes (**Fig. S2c, d, e, f**), illustrating differences between breads from the same location.

We also observed location- and genotype-dependent variabilities for gastric and jejunal peptide compositions. In this regard, peptide abundance tends to be higher in bread digestates from genotypes grown in CF than EM, reflecting the higher protein content observed for samples grown at CF. In addition to abundance pattern changes, we also noticed differences in the proportion of peptides mapped to individual protein families (Fig. 2). For example, the ratio of peptides derived from gliadins of genotype W-59 was about 1.5-fold lower than W-53 and W-64. Such differences are also observed at the flour protein level.

3.5. Gastric and jejunal peptide abundance was linked to flour protein composition

For each environment, based on the criteria specifications detailed in section 2.7., five HD and four LD genotypes were defined from the 16 genotypes (Table 2).

We performed a multi-block PLS-DA for each environment based on the nine genotypes classified as HD or LD to assess the link between gastric and jejunal peptides and flour protein content and composition. This method investigates (i) the relationships between several quantitative datasets (ii) while discriminating groups of individuals.

We selected a subset of peptides using the sparse parameter of the function block.splsda. For CF samples, a combination of 50 gastric and 10 jejunal peptides can be used to discriminate the two groups of genotypes based on their digestibility profile. For the EM samples, the individual plot and the scatter plot (upper triangle) showed that genotype EMW-59 is markedly different from other genotypes regarding protein composition and gastric and jejunal peptide abundances (**Fig. S3**). The genotype W-59 grown at EM was characterized by extreme values for some peptide abundance that could be linked to its specific storage protein accumulation pattern, where it was shown to be the highest proportion of albumins-globulins. This difference masked the variability between the eight other genotypes. Therefore, the genotype W-59 was removed to avoid individual genotype-dependent biases, and the multi-block PLS-DA model was applied to the remaining eight genotypes. Upon analysis and excluding the genotype W-59, a combination of 50 gastric and 10 jejunal peptides were selected per component for the EM site genotypes.

We compared the 50 + 50 gastric and 10 + 10 jejunal peptides selected by the multiblock PLS-DA at CF and EM to the entire 987 gastric and 117 jejunal detected peptides. The distribution of protein family origin of peptides and their abundance was similar between the 1104 detected peptides (Fig. 2**)** and the pool of peptides selected by the multiblock PLS-DA (**Fig. S4)**. However, the epitope diversity of the selected peptides (**Table S6**) differed from the epitope diversity of the 1104 peptides (Table 1).

Despite the environmental effect, the multi-block PLS-DA strongly linked the three datasets. For CF samples, Pearson's correlation coefficient between each dataset's first component was above 0.95. Whilst for the EM samples, Pearson's correlation coefficient between the components built from each dataset was above 0.85 (**Fig. S5**). This high correlation reveals that the gastric and jejunal peptide abundances were strongly linked to flour protein content and composition, suggesting that the gastric and jejunal peptide abundance patterns vary based on the flour protein content and composition.

In samples collected from both environments, we observed that peptides projected with one class of flour protein did not necessarily map to the class of protein it was projected with. For example, at CF,

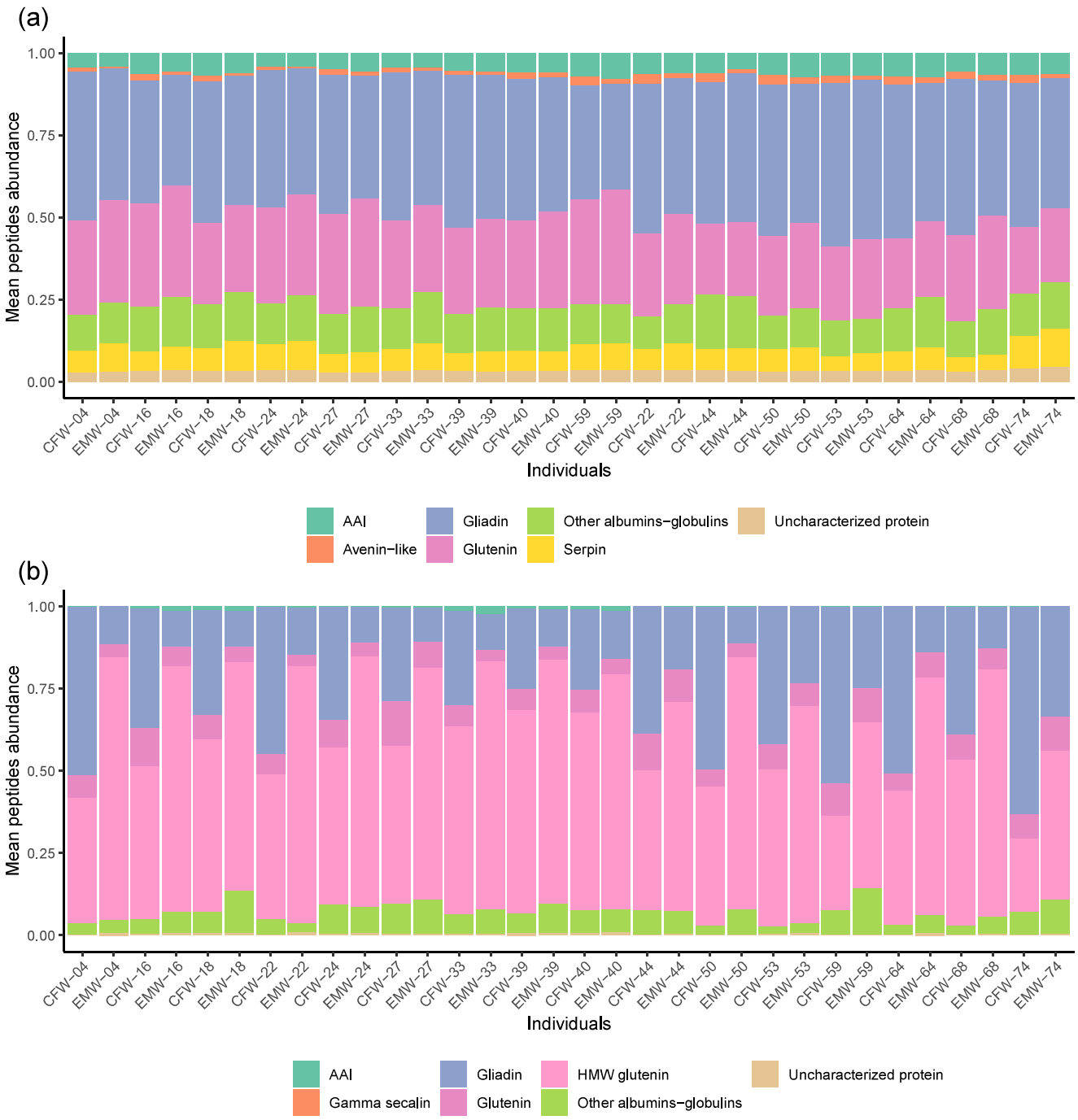


Fig. 2. Characterization of all the quantified peptides. The barplot illustrates the protein family of the 987 gastric (a) and 117 jejunal (b) peptides at CF and EM.

peptides projected with the proportion of albumins-globulins (pAG.Ptot) and ω5-gliadins (pw5.Ptot) in the flour did not only derive from albumins-globulins (Fig. 3a and **Table S7**). From the 34 gastric peptides projected with pAG.Ptot and pw5.Ptot, eight derived from serpins, eight from other albumins-globulins, and three from gliadins, among which one was an ω-gliadin, but 11 from LMW-glutenins and four from uncharacterized proteins. This can be explained by the evolutionary relationship between these protein classes (Juhasz et al., 2018). Moreover, at CF, from the 16 gastric peptides projected with the proportion of gliadins (pgli.Ptot) in the flour, 11 were derived from gliadins. We observed similar trends for the jejunal peptides. Therefore, a high proportion of one class of protein in flour is not necessarily associated with the high abundance of peptides originating from that protein class. In agreement, regressions testing a linear relationship between the proportion of gliadins, glutenins and albumins-globulins in the flour and the proportion of gliadins, glutenins and albumins-globulins derived peptides in gastric digestate, respectively, were not significant except at CF for the gliadins ($p < 0.01$, $R^2adj = 0.36$) and glutenins ($p < 0.05$, $R^2adj = 0.22$).

Several processing steps in flour processing, bread making, and digestion can alter the protein structure and the overall protein network. First, during hydration and kneading, intra- and intermolecular disulphide bonds are formed by cysteine residues of glutenins and sulfur-rich gliadins, leading to the polymeric gluten network. Then, dough mixing leads to rearranging this protein network by rupturing and forming the disulphide bonds (Johansson et al., 2013). At last, the heating step during the baking process induces conformational changes in gluten proteins, such as changes in the secondary structure from α-helix to

Table 2
Digestibility group of the genotypes per environment.

Environment	Genotype	Genotype code	Digest. group	Digest. value (%)
CF	RGT VELASKO	W-18	HD	58.7
CF	SKERZZO-NamB1	W-39	HD	59.4
CF	LR-ROSSO DELLE LANGHE	W-44	HD	57.1
CF	RICHELLE	W-50	HD	58.2
CF	TRAMI~PUY DE DOME	W-53	HD	56.3
CF	CELLULE	W-04	LD	52.2
CF	OSMOSE CS	W-24	LD	46.2
CF	RECIPROC	W-33	LD	51.6
CF	ARTOIS-DEPREZ	W-59	LD	53.3
EM	RGT VELASKO	W-18	HD	59.7
EM	IZALCO CS	W-22	HD	55.5
EM	CAMP REMY	W-27	HD	56.5
EM	SKERZZO-NamNormal	W-40	HD	56.9
EM	GLENLEA	W-68	HD	55.4
EM	CELLULE	W-04	LD	50.9
EM	RUBISKO	W-16	LD	47.5
EM	TRAMI~PUY DE DOME	W-53	LD	52.3
EM	ARTOIS-DEPREZ	W-59	LD	47.6

CF Clermont-Ferrand, EM Estrées-Mons, HD high digestibility, LD low digestibility.

β-sheet (Rahaman, Vasiljevic, & Ramchandran, 2016). Heat also impacts the gluten structure within the bread, i.e., the degree of inter-protein linkage among the sulfur-rich gliadins (Ogilvie et al., 2021). The processing impacts the entire food matrix, next to proteins, including other major flour components such as starch, which also undergo structural modifications between dough formation and bread baking (Ogilvie et al., 2021). In addition, the protein network within the dough is influenced by the flour protein composition (Johansson et al., 2020). A high proportion of gliadins (pgli.Ptot) and storage protein (pSP.Ptot) would lead to a specific overall protein network that undergoes structural modifications during the process (modifications that were not followed in this study), leading to groups of proteins more or less accessible to digestive enzymes. Peptides projected together in Fig. 3a and c could originate from a group of proteins with similar accessibility for digestion. Only some peptides deriving from a protein were quantified. The large peptides (i.e., they are not digested yet) were eliminated during TCA precipitation and the short peptides (less than five amino acids) or amino acids derived from the hydrolysis of the proteins and peptides cannot be detected by our mass spectrometry conditions. The protein amino acid sequence, its tertiary structure and the presence or absence of digestive enzyme cleavage sites provide possible explanations for why peptides from the same protein are differentially digested after two hours.

3.6. Gastric and jejunal peptide composition discriminated high and low digestibility genotypes

In addition to exploring the relationships between multiple quantitative datasets, the multi-block PLS-DA can reveal how genotype-dependent variations can discriminate between high and low bread protein digestibility.

For both growing environments, the three datasets separated the HD and LD genotypes (Fig.3b and d) along the first and the second component at CF and EM, respectively. At EM and CF, a high proportion of albumins-globulins and ω5-gliadins in the flour were associated with LD genotypes. This is consistent with results reported by Lavoignat et al. (2022) and indicates that after removing the environmental effect, a high proportion of albumins-globulins and ω5-gliadins were associated with lower partial protein digestibility for high-yielding genotypes. This result suggests that irrespective of the differences in flour protein composition between CF and EM environments, after two hours of digestion, the hydrolysis rate of bread proteins is limited by a high proportion of less digestible albumins-globulins and ω5-gliadins. These results are in accordance with the much lower digestibility of albumins than that of gluten reported by Ma and Baik (2021) for 26 wheat genotypes.

We also observed that peptides associated with HD or LD genotypes, i.e. HD and LD peptides, were highly heterogeneous, deriving from gliadins, glutenins and albumins-globulins or uncharacterized proteins (**Table S7**). At CF, 60% of the 23 HD peptides were derived from gliadins and 46% of the 37 LD peptides were derived from albumins-globulins. At EM, 35% of the 37 HD peptides and 65% of the 23 LD peptides are derived from gliadins. Thus, the difference between peptides associated with HD genotypes and LD genotypes was not explained by the protein family the peptides originate from. In addition, there were no common peptides discriminating the HD and LD genotypes between CF and EM, except one jejunal peptide (J1311). This peptide was derived from a HMW glutenin and associated with LD at CF and HD at EM (**Table S6**). Six other peptides (J7638, J610, J4822, G9359, G13610, G1077) were selected at EM and CF among the 120 peptides projected, but they only discriminated digestibility groups in one of the two environments. One hypothesis to explain the absence of common peptides (except one) is the grain protein content and composition differences between CF and EM. It is widely known that both genetic and environmental factors influence flour protein content and composition. Based on the proteomic approach, Afzal et al. (2021) identified 303 proteins out of 756 stably expressed proteins across three growing sites in at least one cultivar, illustrating the major influence of the environment on the expression of protein-coding genes. Therefore, the proportion of each protein in the flour could differ between CF and EM. Consequently, the proportion of each peptide released after two hours of digestion would vary between the two growing environments, leading to different gastric and jejunal peptides associated with HD and LD genotypes.

We further compared their mass, sequence length, amino acid composition, and epitope diversity to explore the differences between HD and LD peptides. We hypothesized that peptides characterized by higher mass and/or longer sequences were less hydrolysed and could be associated with LD. However, the Student's *t*-test showed that mean mass and sequence length did not significantly differ between HD and LD peptides in any of the two environments. Moreover, there were variations in amino acid composition between HD and LD peptides, but no trend was common to CF and EM (**Fig. S6**). Based on the epitope analysis, the peptides associated with low digestibility genotypes were not richer in intact celiac or allergen-specific epitopes. We also compared the abundance of epitopes between the combination of the peptides related to HD or LD. Celiac and allergen epitopes were not more abundant in peptides associated with LD. By mapping the HD and LD peptides and the overlapping epitopes on their protein sequence, we could not detect a specific pattern that would differentiate them. HD and LD peptides were generally mapped either at the N- and C-termini or close to the conserved Pfam domains (Fig. 4). Nevertheless, epitope analysis of peptides associated with high and low digestibility genotypes indicate that gastric digestion initiates the protein cleavage at the N- and C-termini of the polypeptides or close to the cysteine-rich conserved domains, while jejunal peptides are more frequently mapped to the epitope-rich repetitive regions of gliadins and glutenins (Fig. 4). Majority of the intact epitopes mapped to high digestibility peptides belong to HMW glutenins. At the same time, low-digestible peptides include α-, γ-, and ω-gliadin-specific epitopes.

Finally, peptides associated with HD or LD genotypes differed in abundance at both CF and EM. However, considering their mass, sequence length, amino acid composition or epitope diversity, we could not find any common signature between CF and EM that would differentiate HD and LD peptides. Boukid, Prandi, Faccini, and Sforza (2019) highlighted the influence of the matrix on protein availability and digestibility. Thus, differences between peptides associated with HD or LD

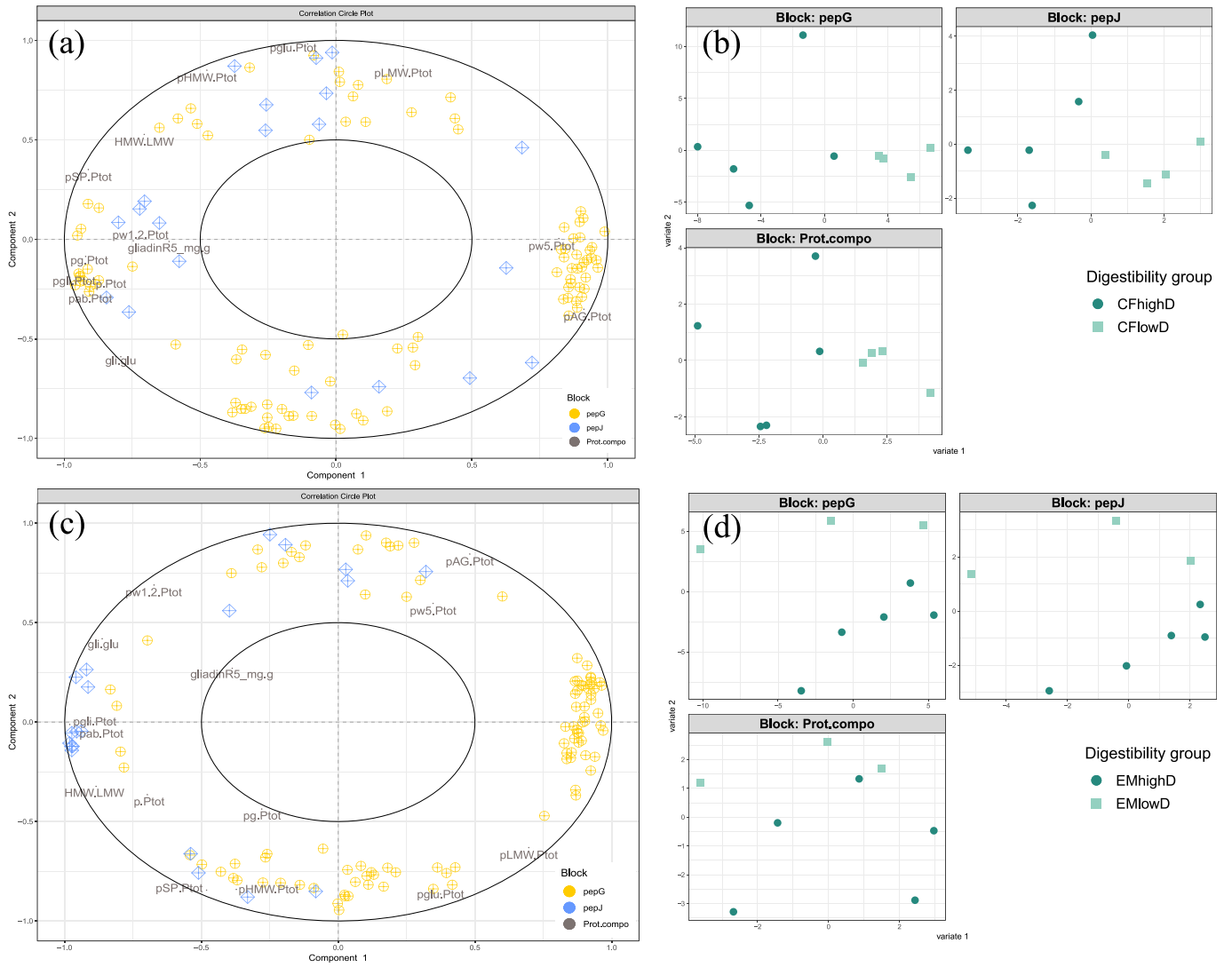


Fig. 3. Correlation circle plot for variables (a and c) and individual plot (b and) of the multi-block PLS-DA at CF (a and b) and EM (c and d). Variables of flour protein composition are in grey, gastric peptides in yellow and jejunal peptides in blue. Genotypes of the high digestibility group are in dark green and of the low digestibility group in light green. For the individual plot, genotypes are represented per datasets. p.Ptot, total protein content; HMW.to.LMW, HMW to LMW glutenin ratio; gli.to.glu, gliadin to glutenin ratio; pHMW.Ptot, pLMW.Ptot, pglu.Ptot, pw5.Ptot, pw1.2.Ptot, pab.Ptot, pg.Ptot, pgli.Ptot, pSP.Ptot and pAG.Ptot, fraction of HMW, LMW, glutenin, ω5-gliadin, ω1.2-gliadin, αβ-gliadin, gliadin, storage protein or albumin - globulin out of total protein content, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

genotypes could rely on a matrix effect. In addition, Ma and Baik (2021) also observed that the digestibility of isolated gluten is higher than that of flour without albumins-globulins, suggesting that non-protein components impact storage protein digestibility. Non-protein components and the overall structure of the bread matrix could interact with proteins, affecting their digestibility and, thus, peptide abundance.

4. Conclusion

This is the first study combining wheat genotypic and environmental variations, food processing and a dynamic *in vitro* digestion model. We established a peptidomic and an epitope diversity fingerprinting map of what could occur in the human body after two hours of digestion. This work provides new insights into bread digestion to further understand wheat sensitivities.

The gastric and jejunal peptide abundance was related to the flour protein content and composition, confirming that the peptides released after a partial bread digestion vary depending on the wheat material (i.e., the genotype and the growing environment). Epitope analysis of the identified gastric and intestinal peptides showed that many of the detected celiac disease-related epitopes were intact in the gastric phase and were only digested in the intestinal phase. Both flour protein content and composition and the amount and composition of digested peptides could discriminate high digestibility versus low digestibility genotypes. However, we could not find a common signature for selected peptides associated with high or low-digestibility genotypes. Notably, peptides associated with low digestibility were not enriched in celiac or allergen epitopes.

The baking process induces protein modifications from flour to bread; proteins interact with the other food matrix components. These alterations of proteins during food processing can determine their accessibility for enzymatic hydrolysis and thus explain why peptides are either associated with high or low protein digestibility. Future work is needed to understand better the influence of the food structure on the release of digested peptides during human digestion of a complex food matrix.

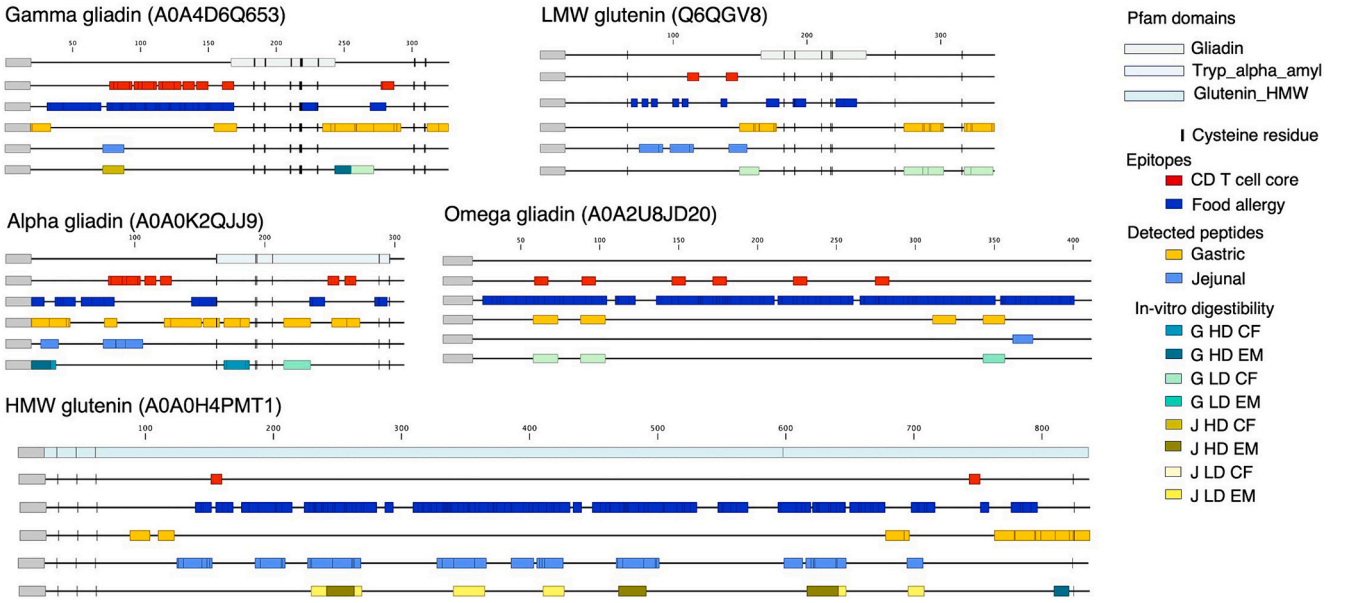


Fig. 4. Peptide mapping profiles of representative storage protein classes. For each storage protein type the conserved Pfam domains, the position of the cysteines and the signal peptides are highlighted. Epitopes representing celiac disease-specific T cell core epitopes and linear food allergen epitopes are highlighted in red and blue colours. Peptides representing gastric and jejunal peptides are also positioned on the maps in yellow and light blue. Peptides identified from the different growing sites as low and high-digestible peptides are coloured with shades of turquoise and olive. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Funding

This work was supported by the French Funds to support Plant Breeding (FSOV) in the context of the project "GlutNsafe" (2016Y), by the French Government through the Research National Agency (ANR) thanks to the project ANR17-CE21-0009 GlutN, by the Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science (CE200100012) and by Coeliac Australia (G1005443). Acknowledgements are expressed to the French Ministry of Agriculture which partly funded the research activities.

CRediT authorship contribution statement

Mélanie Lavoignat: Formal analysis, Investigation, Writing – original draft. **Angéla Juhász:** Formal analysis, Writing – original draft, Writing – review & editing. **Utpal Bose:** Writing – review & editing. **Thierry Sayd:** Investigation, Writing – original draft. **Christophe Chambon:** Investigation. **Miguel Ribeiro:** Supervision, Writing – review & editing. **Gilberto Igrejas:** Supervision. **Sébastien Déjean:** Methodology. **Catherine Ravel:** Conceptualization, Supervision, Writing – review & editing. **Emmanuelle Bancel:** Conceptualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data on flour protein content and composition, flour protein polymer and starch characteristics are available at Recherche Data Gouv (https://entrepot.recherche.data.gouv.fr/dataset.xhtml?persistentId=doi:10.15454/PIMPHZ).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2024.139148.

References

Afzal, M., Sielaff, M., Curella, V., Neerukonda, M., El Hassouni, K., Schuppan, D., … Longin, C. F. H. (2021). 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*(3). Article 3. doi:10/gh7rm8.

Aziz, I., Hadjivassiliou, M., & Sanders, D. S. (2015). The spectrum of noncoeliac gluten sensitivity. *Nature Reviews. Gastroenterology & Hepatology, 12*(9), 516–526. https://doi.org/10.1038/nrgastro.2015.107

Bendtsen, J. D., Nielsen, H., von Heijne, G., & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology, 340*(4), 783–795. https://doi.org/10.1016/j.jmb.2004.05.028

Blanquet-Diot, S., Denis, S., Chalancon, S., Chaira, F., Cardot, J. M., & Alric, M. (2012). Use of artificial digestive systems to investigate the biopharmaceutical factors influencing the survival of probiotic yeast during gastrointestinal transit in humans. *Pharmaceutical Research, 29*, 1444–1453. doi:10/bmjr4v.

Boukid, F., Prandi, B., Faccini, A., & Sforza, S. (2019). A complete mass spectrometry (MS)-based Peptidomic description of gluten peptides generated during in vitro gastrointestinal digestion of durum wheat: Implication for celiac disease. *Journal of the American Society for Mass Spectrometry, 30*(8), 1481–1490. https://doi.org/10.1021/jasms.8b06066

Bredariol, P., Carvalho, R. A., & Vanin, F. M. (2020). The effect of baking conditions on protein digestibility, mineral and oxalate content of wheat breads. *Food Chemistry, 332*. https://doi.org/10.1016/j.foodchem.2020.127399

Bromilow, S., Gethings, L. A., Buckley, M., Bromley, M., Shewry, P. R., Langridge, J. I., & Clare Mills, E. N. (2017). A curated gluten protein sequence database to support development of proteomics methods for determination of gluten in gluten-free foods. *Journal of Proteomics, 163*, 67–75. https://doi.org/10.1016/j.jprot.2017.03.026

Brouns, F., van Rooy, G., Shewry, P., Rustgi, S., & Jonkers, D. (2019). Adverse reactions to wheat or wheat components. *Comprehensive Reviews in Food Science and Food Safety, 18*(5), 1437–1452. https://doi.org/10.1111/1541-4337.12475

Denery-Papini, S., Lauriére, M., Branlard, G., Morisset, M., Pecquet, C., Choudat, D., Merlino, M., Pineau, F., Popineau, Y., Boulenc, E., Bouchez-Mahiout, I., Bodinier, M., & Moneret-Vautrin, D.-A. (2007). Influence of the allelic variants encoded at the Gli-b1 locus, responsible for a major allergen of wheat, on IgE reactivity for patients suffering from food allergy to wheat. *Journal of Agricultural and Food Chemistry, 55*(3), 799–805. https://doi.org/10.1021/jf062749k

Freitas, D., Gómez-Mascaraque, L. G., & Brodkorb, A. (2022). Digestion of protein and toxic gluten peptides in wheat bread, pasta and cereal and the effect of a supplemental enzyme mix. *Frontiers in Nutrition, 9*. https://doi.org/10.3389/fnut.2022.986272

Johansson, E., Branlard, G., Cuniberti, M., Flagella, Z., Hüsken, A., Nurit, E., … Vazquez, D. (2020). Genotypic and environmental effects on wheat technological and nutritional quality. In G. Igrejas, T. M. Ikeda, & C. Guzmán (Eds.), *Wheat quality for improving processing and human health* (pp. 171–204). Springer International Publishing. https://doi.org/10.1007/978-3-030-34163-3_8.

Johansson, E., Malik, A. H., Hussain, A., Rasheed, F., Newson, W. R., Plivelic, T., … Kuktaite, R. (2013). Wheat gluten polymer structures: The impact of genotype, environment, and processing on their functionality in various applications. *Cereal Chemistry, 90*(4), 367–376. doi:10/f22djp.

Juhász, A., Belova, T., Florides, C. G., Maulis, C., Fischer, I., Gell, G., … Olsen, O.-A. (2018). Genome mapping of seed-borne allergens and immunoresponsive proteins in wheat. *Science Advances, 4*(8), eaar8602. https://doi.org/10.1126/sciadv.aar8602

Juhász, A., Haraszi, R., & Békés, F. (2020). Effects of environmental changes on the allergen content of wheat grain. In G. Igrejas, T. M. Ikeda, & C. Guzmán (Eds.), *Wheat quality for improving processing and human health* (pp. 453–470). Springer International Publishing. https://doi.org/10.1007/978-3-030-34163-3_19.

Landolfi, V., D'Auria, G., Nicolai, M. A., Nitride, C., Blandino, M., & Ferranti, P. (2021). The effect of nitrogen fertilization on the expression of protein in wheat and tritordeum varieties using a proteomic approach. *Food Research International, 148*, 110617. doi:10/gm2sth.

Lavoignat, M., Bancel, E., Rimbert, H., Bagnon, S., Benigna, M., Chassin, A., … Ravel, C. (2023). Bread wheat (T. Aestivum) variability: Phenotypic and genotypic data from 75 varieties. *Data in Brief, 46*, Article 108807. https://doi.org/10.1016/j.dib.2022.108807

Lavoignat, M., Denis, S., Faye, A., Halupka, L., Perrochon, S., Rhazi, L., Giraudeau, P., Déjean, S., Branlard, G., Bancel, E., & Ravel, C. (2022). Differences in bread protein digestibility traced to wheat cultivar traits. *Journal of Cereal Science, 107*, Article 103533. https://doi.org/10.1016/j.jcs.2022.103533

Liu, M., Huang, J., Ma, S., Yu, G., Liao, A., Pan, L., & Hou, Y. (2023). Allergenicity of wheat protein in diet: Mechanisms, modifications and challenges. *Food Research International, 112913*. https://doi.org/10.1016/j.foodres.2023.112913

Lupi, R., Denery-Papini, S., Claude, M., Tranquet, O., Drouet, M., Masci, S., & Larré, C. (2019). Thermal treatment reduces gliadin recognition by IgE, but a subsequent digestion and epithelial crossing permits recovery. *Food Research International, 118*, 22–31. https://doi.org/10.1016/j.foodres.2018.02.011

Ma, F., & Baik, B.-K. (2021). Influences of grain and protein characteristics on in vitro protein digestibility of modern and ancient wheat species. *Journal of the Science of Food and Agriculture, 101*(11), 4578–4584. https://doi.org/10.1002/jsfa.11100

Mamone, G., Nitride, C., Picariello, G., Addeo, F., Ferranti, P., & Mackie, A. (2015). Tracking the fate of pasta (T. Durum semolina) immunogenic proteins by in vitro simulated digestion. *Journal of Agricultural and Food Chemistry, 63*(10), 2660–2667. doi:10/f65ngt.

Minekus, M., Alminger, M., Alvito, P., Ballance, S., Bohn, T., Bourlieu, C., Carriere, F., Boutrou, R., Corredig, M., Dupont, D., Dufour, C., Egger, L., Golding, M., Karakaya, S., Kirkhus, B., Le Feunteun, S., Lesmes, U., Macierzanka, A., Mackie, A., … Brodkorb, A. (2014). A standardised static in vitro digestion method suitable for food—An international consensus. In *, Vol. 5. Food & Function* (pp. 1113–1124). ROYAL SOC CHEMISTRY. https://doi.org/10.1039/c3fo60702j. Issue 6.

Ogilvie, O., Roberts, S., Sutton, K., Domigan, L., Larsen, N., Gerrard, J., & Demarais, N. (2020). Proteomic modelling of gluten digestion from a physiologically relevant food system: A focus on the digestion of immunogenic peptides from wheat implicated in celiac disease. *Food Chemistry, 333*, Article 127466. https://doi.org/10.1016/j.foodchem.2020.127466

Ogilvie, O., Roberts, S., Sutton, K., Gerrard, J., Larsen, N., & Domigan, L. (2021). The effect of baking time and temperature on gluten protein structure and celiac peptide digestibility. *Food Research International, 140*, Article 109988. https://doi.org/10.1016/j.foodres.2020.109988

Pasini, G., Simonato, B., Giannattasio, M., Peruffo, A. D. B., & Curioni, A. (2001). Modifications of wheat flour proteins during in vitro digestion of bread dough, crumb, and crust: An electrophoretic and immunological study. *Journal of Agricultural and Food Chemistry, 49*(5), 2254–2261. doi:10/b6wd92.

Prandi, B., Tedeschi, T., Folloni, S., Galaverna, G., & Sforza, S. (2017). Peptides from gluten digestion: A comparison between old and modern wheat varieties. *Food Research International, 91*, 92–102. https://doi.org/10.1016/j.foodres.2016.11.034

Rahaman, T., Vasiljevic, T., & Ramchandran, L. (2016). Effect of heat, pH and shear on digestibility and antigenic characteristics of wheat gluten. *European Food Research and Technology, 242*(11), 1829–1836. https://doi.org/10.1007/s00217-016-2682-4

Ribeiro, M., de Sousa, T., Sabença, C., Poeta, P., Bagulho, A., & Igrejas, G. (2021). Advances in quantification and analysis of the celiac-related immunogenic potential of gluten. *Comprehensive Reviews in Food Science and Food Safety, 20*. doi:10/gngmt8.

Ribeiro, M., Rodriguez-Quijano, M., Nunes, F. M., Carrillo, J. M., Branlard, G., & Igrejas, G. (2016). New insights into wheat toxicity: Breeding did not seem to contribute to a prevalence of potential celiac disease's immunostimulatory epitopes. *Food Chemistry, 213*, 8–18. https://doi.org/10.1016/j.foodchem.2016.06.043

Rohart, F., Gautier, B., Singh, A., & Cao, K.-A. L. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Computational Biology, 13*(11), Article e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Ronga, D., Laviano, L., Catellani, M., Milc, J., Prandi, B., Boukid, F., Sforza, S., Dossena, A., Graziano, S., Gullì, M., Visioli, G., Marmiroli, N., De Vita, P., Pecchioni, N., & Francia, E. (2020). Influence of environmental and genetic factors on content of toxic and immunogenic wheat gluten peptides. *European Journal of Agronomy, 118*, Article 126091. https://doi.org/10.1016/j.eja.2020.126091

Schalk, K., Lang, C., Wieser, H., Koehler, P., & Scherf, K. A. (2017). Quantitation of the immunodominant 33-mer peptide from α-gliadin in wheat flours by liquid chromatography tandem mass spectrometry. *Scientific Reports, 7*(1). https://doi.org/10.1038/srep45092. Article 1.

Shewry, P. (2019). What is gluten-why is it special? *Frontiers in Nutrition, 6*, 101. https://doi.org/10.3389/fnut.2019.00101

Singh, A., Shannon, C. P., Gautier, B., Rohart, F., Vacher, M., Tebbutt, S. J., & Lê Cao, K.-A. (2019). DIABLO: An integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics (Oxford, England), 35*(17), 3055–3062. https://doi.org/10.1093/bioinformatics/bty1054

Sollid, L. M., Tye-Din, J. A., Qiao, S.-W., Anderson, R. P., Gianfrani, C., & Koning, F. (2020). Update 2020: Nomenclature and listing of celiac disease–relevant gluten epitopes recognized by CD4+ T cells. *Immunogenetics, 72*(1), 85–88. doi:10/gh2dnb.

Torcello-Gómez, A., Dupont, D., Jardin, J., Briard-Bion, V., Deglaire, A., Risse, K., Mechoulan, E., & Mackie, A. (2020). Human gastrointestinal conditions affect *in vitro* digestibility of peanut and bread proteins. *Food & Function, 11*(8), 6921–6932. https://doi.org/10.1039/D0FO01451F

Tye-Din, J. A., Stewart, J. A., Dromey, J. A., Beissbarth, T., van Heel, D. A., Tatham, A., … Anderson, R. P. (2010). Comprehensive, quantitative mapping of T cell epitopes in gluten in celiac disease. *Science Translational Medicine, 2*(41). https://doi.org/10.1126/scitranslmed.3001012, 41ra51-41ra51.

Wu, T., Taylor, C., Nebl, T., Ng, K., & Bennett, L. E. (2017). Effects of chemical composition and baking on in vitro digestibility of proteins in breads made from selected gluten-containing and gluten-free flours. *Food Chemistry, 233*, 514–524. doi:10/gf3fsf.

Zevallos, V. F., Raker, V., Tenzer, S., Jimenez-Calvente, C., Ashfaq-Khan, M., Rüssel, N., … Schuppan, D. (2017). Nutritional wheat amylase-trypsin inhibitors promote intestinal inflammation via activation of myeloid cells. *Gastroenterology, 152*(5), 1100–1113.e12. doi:10/f9zwg9.

Zhang, X., Cai, J., Wollenweber, B., Liu, F., Dai, T., Cao, W., & Jiang, D. (2013). Multiple heat and drought events affect grain yield and accumulations of high molecular weight glutenin subunits and glutenin macropolymers in wheat. *Journal of Cereal Science, 57*(1), 134–140. https://doi.org/10.1016/j.jcs.2012.10.010