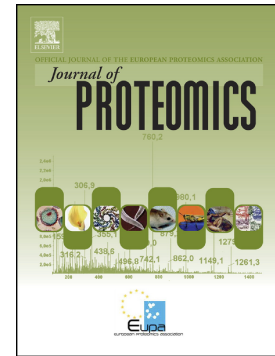


Comprehensive proteome analysis of bread deciphering the allergenic potential of bread wheat, spelt and rye

Julia Zimmermann, Philipp Hubel, Jens Pfannstiel, Muhammad Afzal, C. Friedrich H. Longin, Bernd Hitzmann, Herbert Götz, Stephan C. Bischoff



PII: S1874-3919(21)00217-7

DOI: <https://doi.org/10.1016/j.jprot.2021.104318>

Reference: JPROT 104318

To appear in: *Journal of Proteomics*

Received date: 26 April 2021

Revised date: 9 June 2021

Accepted date: 30 June 2021

Please cite this article as: J. Zimmermann, P. Hubel, J. Pfannstiel, et al., Comprehensive proteome analysis of bread deciphering the allergenic potential of bread wheat, spelt and rye, *Journal of Proteomics* (2018), <https://doi.org/10.1016/j.jprot.2021.104318>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Comprehensive proteome analysis of bread deciphering the allergenic potential of bread wheat, spelt and rye

Julia Zimmermann¹, Philipp Hubel², Jens Pfannstiel², Muhammad Afzal³, C. Friedrich H. Longin³, Bernd Hitzmann⁴, Herbert Götz⁴, Stephan C. Bischoff^{1*}

¹ Department of Nutritional Medicine/Prevention, University of Hohenheim, Fruwirthstrasse 12, Stuttgart, 70593 Germany

² Core Facility Hohenheim, Mass Spectrometry Module, University of Hohenheim, August-von-Hartmann-Strasse 3, 70599 Stuttgart, Germany

³ State Plant Breeding Institute, University of Hohenheim, Fruwirthstrasse 21, 70599 Stuttgart, Germany

⁴ Department of Process Analytics and Cereal Science, University of Hohenheim, Garbenstraße 23 70599 Stuttgart, Germany

*Corresponding author: bischoff.stephan@uni-hohenheim.de

Further Correspondence

Germany, julia.zimmermann@uni-hohenheim.de

Germany, friedrich.longin@uni-hohenheim.de

Germany, M.Afzal@uni-hohenheim.de

Germany, jens.pfannstiel@uni-hohenheim.de

Germany, philipp.hubel@uni-hohenheim.de

Germany, bernd.hitzmann@uni-hohenheim.de

Germany, Herbert.Goetz@uni-hohenheim.de

Abstract

Background/Objectives: Cereal products like flour and bread are known to trigger diseases such as wheat allergy, celiac disease and non-celiac wheat sensitivity (NCWS). Some of these diseases are caused by allergenic proteins, the expression of which might vary depending on the grain type and manufacturing processes. Therefore, we examined the protein composition and abundance of potentially allergenic proteins in flours from bread wheat, spelt and rye, and corresponding breads.

Materials and Methods: Using Nano-LC-ESI-MS/MS and label free quantification (LFQ) we analyzed the proteome of six different bread flours (wholegrain and superfine flours from rye, spelt and bread wheat), and 14 bread types (yeast and sourdough fermented breads from all flours and wheat breads plus/minus bread improver). Potentially allergenic proteins in flours and breads were functionally categorized using the Pfam database and relatively quantified by LFQ.

Results: We could show that almost equal numbers of proteins can be identified in rye- and spelt samples compared to wheat samples using the Uniprot bread wheat protein database, indicating high sequence conservation between cereals. In total, 4424 proteins were identified in the 20 flour and bread samples. The average number of identified proteins in flour (2719 ± 243) was slightly higher than in bread (2283 ± 232 ; $P < 0.001$). In wheat- and spelt wholegrain flour higher protein numbers (wheat: 2891 ± 90 ; spelt: 2743 ± 140) were identified on average than in superfine flour (wheat: 2562 ± 79 ; $P = 0.0089$; spelt: 2431 ± 140 ; $P = 0.0037$). Neither the absolute number nor the abundance distribution of potentially allergenic proteins were dependent on the flour type or the fermentation process, but known allergenic

proteins like gliadins showed higher relative abundance in spelt- and wheat samples, compared to rye samples.

Conclusion: We provide comprehensive proteome data for six flour types and related breads showing that the grain species have greater influence on proteome composition than milling and fermentation processes. Our data indicate that allergenic proteins are not selectively degraded during bread production and are more abundant in bread wheat and spelt compared to rye.

Journal Pre-proof

Introduction

Besides rice and maize, bread wheat is the most widely consumed food grain in the world. Wheat flour is the most important ingredient in home baking and is the framework for almost every commercially baked product. At the same time, wheat proteins are known to be capable of triggering wheat allergy and celiac disease in certain individuals [1,2]. In addition, it is suspected that wheat proteins may trigger the symptoms of NCWS (non-celiac wheat sensitivity), another grain-associated clinical entity [3,4].

There are several known ways in which grain proteins can trigger discomfort in susceptible persons. On the one hand, specific gluten peptides lead to activation of specific immune cells like CD4⁺T-lymphocytes, macrophages or dendritic cells in celiac disease patients, which in turn leads to damage of the small intestinal epithelium [5]. Wheat allergy patients, on the other hand, react to gluten and non-gluten proteins like β -amylase, α -Amylase-Trypsin inhibitors (ATIs) and surface-active proteins such as nonspecific lipid transfer proteins (LTPs) and puroindolines with a release of histamine from basophils and mast cells [6,7]. Based on *in vitro* studies and animal experiments, ATIs could additionally stimulate intestinal immune cells via the activation of Toll-like receptor-4 and could thus trigger intestinal inflammation or exacerbate pre-existing inflammation in patients without wheat allergy or celiac disease [3,8].

Detailed characterization of the protein composition of bread and flour is essential to evaluate their allergenic potential. Based on early biochemical work the wheat flour proteome can be classified in four protein fractions including albumins, globulins, prolamins and glutelins that can be sequentially extracted from flour using different solvents [9]. Prolamins and glutelins together comprise the storage protein fraction of

the grain (gluten) and comprise up to 80% of total protein content [10]. The albumin- and globulin-fraction comprise IgE-binding proteins [11]. Previous grain flour proteome studies revealed grain-specific differences, e.g. between wheat and rye [12,13], but also between wheat and spelt [14]. Prandi et al. [15] showed that spelt contains less immunogenic or toxic peptide epitopes compared to other wheat variants. Even if rye is less closely related to wheat, spelt, einkorn and emmer, rye proteins can also trigger celiac disease and wheat allergy [16].

Milling, kneading, fermentation and heating impact the characteristics of bread and might also affect the bread proteome to different extent. Since the outer layer of the grain contains other proteins than the endosperm, it is likely that different types of flour (f.e. wholegrain- and superfine flour) have different protein compositions [17]. In addition, biochemical changes occur in the protein fraction of the dough during fermentation due to the action of microbial and endogenous wheat enzymes [18]. Sourdough wheat bread contains less pro-inflammatory tetrameric and dimeric ATIs compared to yeast-fermented bread and wheat flour, which might be due to disulphide bond reduction [3,19]. Rizello et al. [20] further showed that specific sourdough microorganisms can also lead to a decrease of other allergenic proteins in wheat- and rye bread. However, a clinical study could not prove that IBS patients tolerate sourdough fermented wheat bread better than yeast fermented wheat bread [19]. Bread improvers used to facilitate the manufacturing process often contain additional proteases that break down proteins like gluten. In combination with thermal processing of the dough, such manufacturing methods can modify the bread proteome [21] [22].

Several studies analyzed the composition of grain flour proteomes and bread proteomes using different proteomics approaches like 2-D gel electrophoresis,

shotgun proteomics and targeted protein quantification [14,23–27]. These studies focused either on the protein composition of flour [14,23,24] or analyzed specific protein groups like ATIs, gluten or lupines in bread [25–27]. To the best of our knowledge, no study systematically investigated the comprehensive proteome of bread and corresponding flours with particular attention to different allergenic proteins so far. In this study, we examined the influence of different manufacturing processes on the bread proteome of wheat, spelt and rye and characterized the proteome of the corresponding flours in detail using a label free quantification (LFQ) proteomics approach.

Material and Methods

Breadmaking

For preparing the different breads analyzed in our study we used grains from bread wheat, spelt and rye (Schapfenmühle, Ulm, Germany), from which superfine flour (SF) (Quadrumat junior mill, Brabender, Duisburg) and wholegrain flour (WG) (Ultra-centrifugal mill ZM 200 sieve insert 0,5 mm; Retsch, Haan) was produced.

Bread from different grains (bread wheat, spelt, rye), different types of flour (SF or WG) and different fermentation (yeast dough or sourdough) were produced (in total 12 bread types). In addition, 100 g bread improver (MeisterMarken Ulmer Spatz, type Weißback Super, CSM Deutschland GmbH, Bremen, Germany) was added to wheat sourdough breads from SF or WG flours, respectively (2 additional bread types). Bread improver was composed of buttermilk powder (24 %); wheat flour; soy flour; emulsifier (soya lecithins, mono- and diglycerides of fatty acids); wheat malt flour; dextrose; salt; acidity regulator (calcium acetate); sour whey powder; acidulant (citric acid, lactic acid); wheat swell flour; flour treatment agent: ascorbic acid;

enzymes. This approach resulted in a total of 14 bread types (Supplementary Table 1).

For preparing the sourdough a conventional sourdough starter (TK Starter, Ernst Böcker GmbH&Co.KG; Minden) was used. A mother sourdough 30 % was prepared composed of sourdough starter 136.5 g, flour 1322 g, water 1542 g, and leavened for 20 h at 26°C. Next, the mother dough was added to the remaining ingredients of the breads. The composition of the breads from sourdough and yeast dough each is given in Tab. 1.

For manufacturing the breads, the ingredients were mixed and kneaded in total for six minutes. The bread dough from SF flour was kneaded 3 minutes with low (25 Hz) and 3 minutes with high intensity (50 Hz) whereas bread dough from WG flour was kneaded 2 minutes with low and 4 minutes with high intensity. Afterwards the doughs were leavened at 30°C and 80 % humidity for 20 minutes. Subsequently the doughs were formed. The time for bench rest at room temperature was 10 minutes before the dough experienced a last maturation step in the loaf pan for 80 minutes (bread wheat and spelt) and 70 minutes (rye). Baking temperature was 220°C and baking time was 40 minutes for breads from SF flour and 50 minutes for breads from WG flour (bread wheat and spelt). Breads from rye were baked for 60 and 70 minutes, respectively (Figure 1). After the baking process breads were cut into slices of 2 cm, freeze-dried (Delta 1-24 LSC, Martin Christ Gefriertrocknungsanlagen GmbH; Osterode am Harz), milled and stored at -20°C until analysis.

Protein extraction for proteome analysis

20 mg flour or bread were suspended in lysis buffer containing 2 % SDS (sodium dodecyl sulfate), 20 mM DTT (dithiothreitol) and 150 mM Tris-HCl pH 8.5 and incubated for 10 min at 95 °C. After centrifugation at 4° C for 30 min at 13,700 rpm

the supernatant was removed, and proteins were precipitated using chloroform–methanol precipitation [28]. Protein pellets were resuspended in 6 M urea in 50 mM Tris–HCl pH 8.5, and protein concentrations were determined by the Bradford assay [29].

In-solution digest of proteins and peptide purification

10 µg flour- or bread protein extract in 60 µL 6 M urea, 50 mM Tris HCl (pH 8.5) were used for in solution digests. DTT was added to a final concentration of 10 mM for the reduction of cysteines. Samples were incubated for 30 min at 56 °C under shaking at 1,000 rpm in an Eppendorf Thermomixer. Alkylation of cysteines was performed by adding 30 mM iodoacetamide and incubation for 45 min at room temperature (RT) in the dark. Alkylation was stopped by adding 50 mM DTT and samples were incubated for another 10 min at RT. Five hundred nanogram (500 ng) LysC protease (Roche) in 50 mM Tris HCl pH 8.5 was added and samples were digested overnight at 30 °C. Next, the urea in the reaction mixture was diluted to 2 M by adding the appropriate amount of 50 mM Tris HCl pH 8.5. One microgram (1 µg) trypsin (Roche) in 50 mM Tris HCl pH 8.5 was added and digestion was continued for 4 h at 37 °C. The digest was stopped by adding 3 µL 10 % TFA (trifluoroacetic acid). Next, peptide mixtures were concentrated and desalted on C18 stage tips26 and dried vacuum [30]. Dried samples were dissolved in 30 µL 0.1 % TFA. Aliquots of 3 µl were subjected to nano-LC-MS/MS analysis.

Nano-LC-MS/MS

Mass spectrometry analysis was performed as previously described [14]. Briefly the measurements were performed on an UltiMate 3000 RSLCnano system (Thermo Fisher Scientific, Bremen, Germany) coupled to a Q-Exactive HF-X mass spectrometer (Thermo Fisher Scientific) using a Nanospray Flex ion source (Thermo

Fisher Scientific). Tryptic digests were concentrated and desalted on a precolumn (300 μm x 5 mm, Acclaim PepMap100 C18, 5 μm particle size, 100Å pore size) and separated on a nanoEase MZ HSS T3 analytical column (25 cm x 75 μm , 1.8 μm particle size, 100Å pore size, Waters) operated at constant temperature of 35 °C. Peptides were separated at a flow rate of 300 nL/min using a 90 min gradient with the following profile: 2 % - 55 % solvent B in 90 min, 55 % - 95 % solvent B in 5 min and maintained at 90 % solvent B for 5 min. 0.5 % acetic acid (solvent A) and 0.5 % acetic acid in acetonitrile/H₂O (80/20, v/v, solvent B) were used as solvents.

The Q Exactive HF-X was operated under the control of XCalibur 4.3.73 software (Thermo Fisher Scientific). MS spectra (m/z = 300-1800) were detected in the Orbitrap at a resolution of 60000 (at m/z = 200) using a maximum injection time (MIT) of 100 ms and an automatic gain control (AGC) value of 1×10^6 . Internal calibration of the Orbitrap analyzer was performed using lock-mass ions from ambient air as described in [31]. Data dependent MS/MS spectra were generated for the 10 most abundant peptide precursors in the Orbitrap using higher-energy C-trap dissociation (HCD) fragmentation at a resolution of 15000, a normalized collision energy of 27 and an intensity threshold of 1.6×10^6 . Only ions with charge states from +2 to +5 were selected for fragmentation using an isolation width of 1.6 Da. For each MS/MS scan, the AGC was set at 2×10^5 and the MIT was 50 ms. Fragmented precursor ions were dynamically excluded for 30 s within a 10 ppm mass window to avoid repeated fragmentation.

MS data analysis and protein quantification

Processing of the raw files and protein identification were performed with MetaMorpheus software suite version 0.0.307 (<https://github.com/smith-chem-wisc/MetaMorpheus>) [32]. First, mass re-calibration was performed by the calibrate

task with default settings. In a second step the re-calibrated spectra were searched against the wheat proteome sequence downloaded from UniProt [33] (*Triticum aestivum*, 2019). Search in MetaMorpheus was performed in the classic search operation mode against the automatically computed reversed sequences as decoy database. Mass tolerance of 5 ppm (parts per million) for MS spectra und 20 ppm for MS/MS spectra were used. Trypsin was specified as enzyme and three missed cleavages were allowed. Minimal peptide length was set to five and maximal peptide length was limited to 30 amino acids. The 'Common Fixed' (Carbamidomethyl on cysteines) and 'Common Variable' -option (Oxidation on methionine) for modifications were chosen and the 'Match between runs' feature was enabled. The peptide q-value cut of was set to 0.01 and the minimum allowed peptide score was set to five. For protein quantification the FlashLFQ option in the MetaMorpheus software suite was enabled with a 5 ppm peak finding tolerance for MS1 scans [34]. The bread wheat data base (Uniprot *Triticum aestivum*; Taxon identifier 4565) was used not only for the wheat analyses, but also for the spelt and rye analyses of proteins, because specific databases for spelt and rye proteins are not available in public repositories so far.

For the downstream statistical analysis, the MetaMorpheus Protein groups output file was load into the Perseus software frame (Version 1.6.14.0) [34]. Only protein groups with a q-value < 0.01 and with at least one valid FlashLFQ (Label free quantification) in the data set were considered for downstream analysis. Additionally, known contaminants and proteins groups that match to the decoy database during search were excluded. In total 4424 protein groups remained in the data set. For further quantitative analysis, LFQ intensities were log2-transformed.

Downstream analysis of the data

For each sample, the LFQ (label-free quantification) intensities of three replicates were measured. LFQ intensities were log₂-transformed for further analysis. The dataset was filtered for Q values <0.01 and for intensities in at least one sample, leading to 4424 identified proteins.

For analyzing the 100 most abundant proteins, the average LFQ intensities were sorted by descending order in Microsoft Excel software and the 100 proteins with the highest LFQ intensities were picked for analysis. Protein abundances were calculated based on the average LFQ intensity of all three replicates per sample.

Bar charts, Swarm plots and Pie charts were constructed using Graphpad Prism software 9.0 (GraphPad Software, San Diego, California USA, <http://www.graphpad.com>). Venn diagrams were created using R software (R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org>) [35] and Lucidchart (www.lucidchart.com).

Systems analysis and functional annotation

Hierarchical clustering was performed by Perseus software Version 16.14 (Perseus computational platform, Martinsried, Germany, <http://www.maxquant.net/perseus/>) [36]. For the functional annotation analysis of differentially regulated protein groups, we considered proteins with at least 3 valid LFQ values in one experimental condition (n=3) which additionally show at least 70% valid LFQ values in the whole dataset. The remaining missing values were imputed by Quantile Regression Imputation of Left-Censored data based on the R script "ImputeR" (tune.sigma=1). Significance of protein abundance changes between the different conditions were computed by a permutation-based Welch's t-Test (FDR= 0.05, S0= 0.5, 250 randomizations) which required at least 3 valid LFQ values (unimputed) in one of the tested conditions (Perseus). Protein groups with significant abundance changes

were used for hierarchical cluster analysis in Perseus. Therefore, each row in the data matrix was z-scored separately for wheat, spelt and rye samples on bases of the unimputed LFQ intensities. The normalized LFQ data were used as input for the row clustering. Euclidean distance was used as measure for the row clustering in Perseus. Row clusters which show distinct behavior in their average protein profile were selected manually and associated annotated terms (from the Interpro and Pfam database) were computed by a Fisher exact test using the initially identified 4424 proteins in this study as reference set. Annotation terms were filtered for an enrichment factor >5 .

The six samples with bread improver were excluded from this analysis, as the bread improver contains proteins from several foreign species which could potentially lead to altered quantification and false identification of a bread wheat proteome subset through peptides with homologous sequences.

Definition of allergenic wheat proteins

Juhász et al. 2018 [37] identified wheat proteins associated with adverse allergic and immune responses in people with wheat associated disorders. These proteins were identified based on information in a database of allergen families (AllFam; www.meduniwien.ac.at/allfam) [38] supplemented by the data in the AllergenOnline FARRP database (www.allergenonline.org) [39]. The set of proteins defined by Juhász et al. was used for the identification of seed-borne allergens and immunoresponsive proteins in flour and bread samples, but only proteins with “ingestion” as route of exposure were considered. This criterion was fulfilled by 330 proteins, of which 290 could be assigned to 22 Pfam families. 40 remaining proteins that were not belonging to any Pfam family were excluded. We defined all proteins in

these 22 Pfam families as “proven or potentially allergenic/immunoresponsive proteins”.

Furthermore, we used the allergome database [40–42] to identify additional allergens. The allergome platform is currently the most comprehensive collection of allergen data. It contains identified and characterized allergenic proteins, published in peer-reviewed journals, including those in the official WHO/IUIS Allergen Nomenclature list. We were able to identify additional allergens in the allergome database that could be assigned to 2 additional Pfam families finally resulting in 24 Pfam families in total with potentially and proven allergenic/immunoresponsive proteins. In the following these proteins will be called “allergenic proteins”.

Statistics

Statistical analysis was performed by GraphPad Prism 9 (GraphPad Software, San Diego, California). Data were tested for normal distribution by Shapiro Wilk test. Milling effects (WG or SF flour) were analyzed using an independent-samples students T-test. Effects of grain type (rye, spelt, bread wheat) or different manufacturing processes (sourdough fermentation, yeast fermentation, bread improver) were analyzed using one-way analysis of variance (ANOVA). Level of significance (α) was set at 0.05.

Results

Quantitative protein analysis of flour and bread

In total, 4424 unique proteins were identified in all flour and bread samples (6 flours, 14 breads) by our proteomics approach. Numbers of identified proteins varied between samples, ranging from 2271 to 3143 different proteins per sample (Figure 2A, B). In average, flour samples (2719 ± 243 ; mean $\pm$ SD of 6 flours) contained

higher numbers of different proteins than bread samples (2283 ± 232 ; mean $\pm$ SD of 14 bread types; $P < 0.0001$) (Fig. S1). To consider differences between the different grain types, the number of identified proteins was compared across flours and breads from bread wheat (2407 ± 295 mean $\pm$ SD of 2 flours and 3 bread types), spelt (2393 ± 237 mean $\pm$ SD of 2 flours and 2 bread types) or rye (2443 ± 402 mean $\pm$ SD of 2 flours and 2 bread types) but no statistically significant differences were observed between the different grain types. Spelt and rye proteins were identified in almost equal numbers compared to bread wheat (hereafter referred to as wheat) proteins using the bread wheat protein database (Uniprot *Triticum aestivum*; Taxon identifier 4565).

In whole grain (WG) flour and bread samples from wheat and spelt higher numbers of unique proteins were identified than in corresponding superfine (SF) flour samples. This was expected since in the production of SF flour the aleurone layer of the grain, which is rich in proteins, is removed. However, for rye flour and bread samples the picture was different as in all SF samples higher numbers of proteins were identified compared to WG samples (Figure 2 A, B).

Comparing different grain types, approximately two thirds of the identified proteins were common between flour (61%) and half of the identified proteins were common between bread (46%) from wheat, spelt and rye (Figure 2 C, D). The overlap of identified proteins was always higher between wheat and spelt samples (12% in flour, 20% in bread) than between wheat and rye samples (6% in flour and 3% in bread) or rye and spelt samples (5% in flour and 4% in bread). 12% of the proteins in rye flour and 15% of the proteins in rye bread were only identified in rye, but not in flour and bread samples from wheat and spelt, leading to a different protein composition of flour and bread samples from rye.

Influence of manufacturing processes on protein composition

To analyze possible effects of milling and the manufacturing processes on protein composition of flour and bread in more detail we compared the distribution of all proteins that had LFQ intensities between SF and WG flours (Figure 3 A, C, E) and between different bread types (Figure 3 B, D, F), respectively. Comparing SF and WG flour, we could show that in all three different grain types around 80% of the proteins were common. 15% of the proteins were unique in wheat WG flour and 13% of proteins were unique spelt WG flour. Lower numbers of unique proteins were identified in wheat and spelt SF flour (5% and 3%, respectively). In rye flour differences between WG flour and SF flour were less pronounced, 8% of the proteins could only be found in WG flour and 6% in SF flour.

In addition, we investigated the influence of the fermentation process on the protein composition of breads and compared the numbers of identified proteins in sourdough fermented (S) and yeast fermented (Y) breads. In wheat samples we also analyzed sourdough fermented bread with improver (SB) (Figure 3B, 2D, 2F). The majority of proteins (66%-80%) could be identified in all the different bread types within one grain type (wheat, spelt or rye). The number of common proteins for wheat breads was slightly lower because three different bread types were compared, while for spelt- and rye bread only two different bread types were compared. No significant differences in the number of unique proteins were observed comparing WG or SF sourdough fermented bread to yeast fermented bread (Figure 3 B, D, F and Figure S2 A-C). Comparing WG bread and WG flour, the number of unique proteins in flour (10-13%) was higher compared to the unique proteins in the different breads (2-5%) (Figure S2 D-F).

Functional classification and annotation of proteins in flour and bread

We used label-free quantification (LFQ) analysis to gain further insights into the differences in protein composition between different flour types and breads. After filtering the complete dataset for proteins that had at least 70% valid values (LFQ intensities) in all samples, 1737 proteins remained. Welch's t-tests were performed for all relevant comparisons of bread and flour types using a cut-off of $p < 0.01$ and an enrichment factor > 5 to identify differentially abundant proteins. We compared WG and SF samples of wheat, spelt and rye using mean values of bread and flour samples. In addition, we compared yeast- to sourdough fermented bread. Proteins that showed significant abundance changes in at least one comparison were ranked using unsupervised hierarchical clustering (Figure 4). Two dominant clusters emerged from the hierarchical clustering analysis: The first cluster comprised 45 proteins that showed lower abundance in sourdough fermented bread compared to yeast fermented bread and flour (Figure 4, cluster A, orange). The second cluster comprised 107 proteins that showed higher abundance in WG samples compared to SF samples (Figure 4, cluster B, blue).

The two dominant protein clusters identified by hierarchical clustering analysis were further analyzed using functional annotation enrichment analysis. 887 of the 1737 proteins with $> 70\%$ valid values had PfamID annotation. Based on these annotations functional annotation enrichment analysis using Fisher's exact test was performed. Cluster A included proteins belonging to the Linker-histones Pfam family and the Ribosomal proteins S19 and L28e, which were depleted in sourdough breads compared to corresponding WG and SF flours as well as corresponding yeast fermented breads. In cluster B the protein groups Cupin_1, SMP (Seed maturation protein), Barwin and Glyoxalase showed higher abundances in WG samples (bread + flour) compared to SF samples (bread + flour).

Allergenic proteins in bread and flour

Next, we focused on the presence of potentially and proven allergenic proteins in bread and flour that may cause an immune response by dietary intake. Based on available literature and allergen databases (see material and methods), we defined 24 Pfam domain families that included proven and potentially allergenic proteins. 2105 proteins in our dataset were annotated with a Pfam ID. 246 of these proteins could be assigned to 21 of the above mentioned Pfam domain families (Table 1). The remaining 3 Pfam domain families were not present in the set of proteins identified in our study. The 21 Pfam domain families with 246 representatives in our dataset comprise proven as well as potentially allergenic wheat proteins. 54 of these 246 were proven allergenic proteins according to the allergome database [42] and the database of Juhasz et al. [37] known to be involved in celiac disease or wheat allergy. The allergen list can also be used for the identification of potentially allergenic proteins in wheat-related species like rye or spelt, because they share common ancestry and over 70 % sequence similarity [37].

The abundance of potentially allergenic proteins was further analyzed using histograms comparing distributions of LFQ intensities of allergenic proteins (Table 1; 246 proteins including proven and potential allergens) and the remaining proteins that were summarized as “Others” (4177 proteins). In WG flour of wheat, spelt and rye, the distribution of LFQ intensities of allergenic proteins was shifted towards higher LFQ intensities compared to “other” proteins (Fig. 5A), while rye flour showed lower abundances of potentially allergenic proteins compared to wheat and spelt. For deeper understanding of the distribution of allergenic proteins in the different datasets, we calculated the average LFQ intensity of allergenic proteins and others. Allergenic proteins were further divided into potentially allergenic proteins and

proven allergenic proteins as described above. To exclude, that differences regarding the LFQ intensities of allergenic proteins in distinct samples are based on a lower number of detected allergenic proteins, for further comparison we exclusively considered the LFQ intensity of proteins detected in all the samples that were compared to each other.

2579 proteins with PfamID were detected in at least one replicate of WG flour of wheat, spelt and rye (Figure 5B). 168 of these proteins belonged to the potentially allergenic proteins shown in table 1. Of these 119 proteins were potential allergens according to their Pfam ID, while 49 proteins were proven allergens. The remaining proteins (n=2411) were summarized as “others”. We could observe significantly lower abundances of potential and proven allergens in rye flour compared to wheat and spelt flour, while differences between wheat and spelt were statistically not significant (Figure 5B). Accordingly, we analyzed different manufactured wheat breads from WG flour (Y, S, SB) and the corresponding flour. 129 potential- and 52 proven allergens were found in all four sample sets, but no significant differences between allergen numbers in the different bread types and the corresponding flour could be observed (Figure 5C). To get further insight, which allergen groups were responsible for the differences observed between wheat, spelt and rye we analyzed the differences in LFQ intensities within each of the 21 allergen groups. For that purpose, we calculated the mean of the LFQ intensities of the 21 allergen groups in the three replicates of WG flour of wheat, spelt and rye (Figure S3A). In two of the allergen groups, namely the gliadin group and the inhibitor/storage group, significant differences in the LFQ intensities between wheat and rye as well as between spelt and rye were observed (Figure S3A, S3B). Considering these two allergen groups in

different bread types (Figure S3C) and in WG and SF flour from wheat (Figure S3D), no significant differences in allergen abundancies could be observed.

Another group of potential allergens are amylase trypsin inhibitor proteins (ATIs) that may be involved in triggering the symptoms of NCWS. These proteins belong as well as some gluten proteins to the Inhibitor/Storage Pfam family. We identified nine different ATI proteins in our dataset. The relative abundance of all nine ATI proteins was reduced in rye samples compared to wheat and spelt samples (Figure S4).

Also other potential grain components like fructans have been described as being of relevance in adverse reactions to cereal products like bread [43]. Fructans are produced in the vacuole by the action of specific enzymes (fructosyltransferases) that transfer fructose from sucrose to the growing fructan chain [44]. Rye is known to contain more fructans than wheat and spelt [45]. Searching for fructosyltransferases in rye, wheat and spelt bread we found only one out of 23 fructosyltransferases that are listed in UniProt in our dataset (Sucrose:fructan-6-fructosyltransferase; UniProtKB - A0A3B6I6G4). This protein could be found in all rye samples, but only in few wheat and spelt samples, with a tendency towards higher LFQ intensities in rye than in wheat and spelt (Figure S5).

Discussion

We show for the first time comprehensive proteome data for three different grain flours (wheat, spelt and rye) derived from commercially available sources and for corresponding breads. We detected more than 4000 different proteins with up to 3143 different proteins per sample and we found that flour and bread samples contained around 1300 common proteins that were identified in all samples.

Our results revealed that the Uniprot bread wheat protein database that we used as reference allows the identification not only of spelt proteins, as shown earlier by our group [14], but also of rye proteins. This suggests that protein sequences are highly conserved between wheat (both bread wheat and spelt) and rye, which is in line with the well-known fact that wheat and wheat-related species like rye have a common ancestry and share over 70 % sequence similarity [37]. Even a single, non-isobaric amino acid substitution within a peptide would lead to a difference to the wheat reference sequence and would result in loss of this peptide in a database search. Although we cannot exclude that some rye- and spelt proteins are not identified by our approach, our results suggest that rye- and spelt proteins can be identified in comparable numbers to wheat proteins using the bread wheat proteome database. Almost two third of the proteins in WG flour and half of the proteins in WG bread could be identified in the samples of all three grain types. Wheat and spelt shared more common proteins than wheat and rye or spelt and rye, indicating that the diploid species rye (*Secale cereale*) is biologically more distant to wheat and spelt that belong to the same hexaploid species (*Triticum aestivum*) [46].

We were able to identify up to 13% more unique proteins in flours compared to the corresponding breads and the average number of identified proteins in flour was slightly higher than in bread. Loss of proteins during processing of the flours to produce bread could be due to partial or complete degradation of proteins by bacterial or fungal proteases during the dough fermentation process [47]. Alternatively, protein extraction from bread might be less efficient than that from flour. Several procedures might account for a reduced number of proteins in breads compared to the corresponding flour. During baking dough is transformed into crumb through gelatinization of the starch granules in the dough and coagulation of the

proteins, whereby a stable gluten network is built that encloses the gelatinized starch granules [48] and might hinder protein extraction. The extractability of wheat proteins was shown to decrease during the bread-making due to aggregation and polymerization of proteins based on disulfide bonds that are formed during heat treatment of bread [49]. Also heating during baking process might play a role by degrading particular proteins. Previous studies showed that the solubility of the most abundant gluten proteins and possibly other proteins decreases through heat treatment [50].

Confirming earlier findings [51], we identified more different proteins in WG compared to SF flour from wheat and spelt. Most likely this is due to the fact that the aleurone layer of the grain, which is rich in proteins, is removed with the outer layer of the bran during the milling process [52]. In contrast, in rye samples more proteins could be detected in SF compared to WG flour. An explanation for this contradictory finding might be the higher abundance of pentosans present in rye compared to wheat and spelt [53,54]. Pentosans are polysaccharides present in all plant cells containing cellulose. They are mainly located in the external parts of rye kernel [55] and therefore present in higher amount in rye WG flour. Especially the presence of insoluble pentosans could complicate the extraction of proteins from rye WG flour samples. These observations could likewise be confirmed for the corresponding breads. Hierarchical cluster analysis revealed that particularly proteins of the Cupin and Barwin protein families were significantly reduced in SF samples (flour and bread). The Cupin superfamily includes wheat globulin storage proteins as well as germins located predominantly in the seed bran [50,56,57]. Since also Barwin proteins are known as plant pathogenesis-related proteins and especially expressed

during fungal infections [58], it is likely that they are expressed preferentially in the outer layers of the kernel.

Of particular interest in our proteome analysis were potentially allergenic proteins that might cause adverse reactions to bread and other cereals. We defined 24 Pfam domain families that included proven and potentially allergenic proteins. 2105 proteins in our dataset were annotated with a Pfam ID. 246 of these proteins could be assigned to one of the above mentioned Pfam domain families. 54 of these proteins have been confirmed to be involved in celiac disease or wheat allergy [37,42], whereas 192 have been categorized as potentially allergenic according to their Pfam annotation. Most allergenic proteins in food belong to only a few protein families and superfamilies based on their structural and functional properties [38,59]. Moreover, food allergens must be heat stable and resistant to digestive enzymes [59]. Further, Codex Alimentarius guidelines indicate that proteins with sequence identities to known allergens exceeding 35 % in any segment of at least 80 amino acids may be allergenic [60].

The 246 proteins showed higher LFQ intensities on average than the 4178 other proteins in all analyzed samples. Neither the absolute number nor the distribution of these potentially allergenic proteins were dependent on the flour type or the fermentation process, respectively. Remarkably, significantly more potential and proven allergenic proteins could be found in spelt- and wheat flour, compared to rye flour. This observation can be explained by lower amounts of gluten proteins present in rye compared to wheat [12]. Accordingly, when comparing the 21 allergen groups we defined in our study in wheat, spelt and rye wholegrain flour samples, we only observed significant differences between the abundancies of “gliadins” and “Inhibitor/Storage” proteins, two allergen groups to which gluten proteins belong to.

Apart from primary amino acid sequences, immune responses to proteins are also dependent on protein folding and tertiary structure, which explains why not all proteins within one allergen family are allergens. Therefore, the clinical impact of our findings needs to be confirmed in patient studies with challenge tests. Of the 54 proven allergenic proteins found in our dataset, 49 are found in at least one of three replicates of wheat, spelt and rye WG flour; however, with lower LFQ intensities in rye, further explaining why rye might be less relevant for allergy than wheat or spelt. Also well-known allergens like ATIs, from which we could identify 9 in our dataset, were more abundant in wheat and spelt flours and breads, respectively, according to our data. On the other hand, rye is thought to contain more fructans compared to wheat and spelt [45]. Confirming this, we found one fructosyltransferase constantly in rye and only occasionally in wheat and spelt samples. The clinical relevance of this finding needs to be confirmed.

Numerous studies examined the degradation of allergenic proteins during bread-making using 2D-gel electrophoresis. De Angelis et al. [61] observed differences between the composition of allergenic proteins of yeast and sourdough fermented breads. They suggested that traditional protocols of bread-making such as long-time fermentation of sourdough (24h at 37°C) with selected starters result in extensive degradation of IgE-binding wheat albumins and globulins. Seven proteins were identified by MALDI-TOF mass spectrometry. Three of these proteins could also be identified in our dataset, namely xylanase inhibitor protein 1, endogenous α -amylase-subtilisin inhibitor (WASI) and ATI 0.19. Comparing the LFQ intensities of these three proteins between wheat flour, yeast fermented bread and sourdough fermented bread, we were not able to detect any differences (data not shown). This might be due to the shorter dough fermentation time (100 minutes for rye and 110

minutes for wheat and spelt at 30°C) used in our study. Using SDS-PAGE and immunoblot analysis, Rizello et al. [20] showed a stronger proteolytic effect of some IgE-binding proteins in sourdough fermented compared to yeast fermented bread from wheat and rye. In contrast to these findings, we did not observe such differences, most likely due to different baking conditions. This assumption is strengthened by results from Loponen et al. [62], who used comparable fermentation times as we did in the present study and showed that short-time fermentation (2h) of sourdough did not result in degradation of gluten proteins [63].

To a subset of sourdough fermented breads, we added a typical bread improver used in conventional bakeries to analyze its impact on the bread proteome. We could neither detect significant differences regarding the number of identified proteins nor the relative abundance of allergenic proteins related to the presence or absence of the improver (see supplementary table 3). The fact that we did not observe any differences in the relative abundance of allergenic proteins between different bread types, or between bread and flour, indicates that there is no specific degradation of allergenic proteins during the fermentation process of the present study.

Apart from allergenic proteins, other proteins might be modified or even removed by the manufacturing process. Although less relevant for allergies, this might affect the quality of bread. Indeed, we found that up to 16% of these other proteins were found uniquely in one of the bread types suggesting that the manufacturing process affects these proteins. At present, it is unclear how this alters bread quality. Hierarchical cluster analysis showed that the abundances of histone- and ribosomal proteins were reduced in sourdough fermented bread compared to yeast fermented bread and flour. Maybe this is due to histone protein degradation by lactic acid bacteria for

obtaining pyrimidines and purines. Many lactobacilli are auxotrophic for both purine- and pyrimidine bases and therefore, they inherit a salvage-pathway system based on trans-N-deoxyribosylases to utilize external DNA bases for growth [64].

In conclusion, our results show that allergenic proteins are not selectively degraded during conventional bread production, neither in yeast nor in sourdough fermented bread nor in sourdough fermented bread with improver.

Acknowledgement

This work was funded by the IGF project of the FEI supported via AiF within the program for promoting the Industrial Collective Research (IGF) of the German Ministry of Economics and Technology (BMWi) (AiF 20763N to S.C.B.)

Table 1. Protein families containing potentially immunoresponsive and allergenic proteins identified in wheat according to the allergome database and the AllFam database

No.	Pfam ID	Pfam name	Identified proteins in our dataset	Defined group name
1	PF00079	Serine Proteinase Inhibitors	12	Serpins
2	PF00085	Thioredoxin	18	Thioredoxin
3	PF00112	Papain family cysteine protease	1	PCP
4	PF00121	Triosephosphate isomerase	4	TIM
5	PF00141	Peroxidase	29	Peroxidase
6	PF00182	Chitinase class I; Chitin recognition protein	10	Chitinase
7	PF00190	Cupin	22	Cupin
8	PF00197	Trypsin and protease inhibitor	2	Kunitz
9	PF13016	Cys-rich Gliadin N-terminal	22	Gliadin
10	PF03157	High molecular weight glutenin (HMW) subunit	3	Glutenin
11	PF14368	Probable lipid transfer	15	LTP
12	PF00234	Protease inhibitor/seed storage LTP family	47	Inhibitor/Storage
13	PF00235	Profilin	1	Profilin
14	PF00314	Thaumatococcus family	15	Thaumatococcus
15	PF00560	Leucine rich repeat N-terminal domain	2	LRR
16	PF03227	Gamma interferon inducible lysosomal thiol reductase	3	GILT
17	PF00450	Serine carboxypeptidase	10	Serine Carbox
18	PF00321	Plant thionin	2	Thionin
19	PF01373	Glycosyl hydrolase family 14	1	GlycoHydro
20	PF00534	Glycosyl transferases group 1	14	GlycoTransfer
21	PF00161	Ribosome inactivating protein	6	RIP

Only protein families with representatives that can elicit immune response by ingestion are shown (for further details see Materials and Methods). Source: Nr. 1-19: [37]; 20-21 [42]. Abbreviations: Triosephosphate isomerase (TIM); Papain-like cysteine protease (PCP); Leucine rich repeat N-terminal domain (LRR); Gamma interferon inducible lysosomal thiol reductase (GILT).

References

- [1] Grimshaw KEC, Bryant T, Oliver EM, Martin J, Maskell J, Kemp T, Clare Mills EN, Foote KD, Margetts BM, Beyer K, Roberts G. Incidence and risk factors for food hypersensitivity in UK infants: results from a birth cohort study. *Clin Transl Allergy* 2015;6:1.
- [2] Fasano A, Berti I, Gerarduzzi T, Not T, Colletti RB, Drago S, Elitsur Y, Green PHR, Guandalini S, Hill ID, Pietzak M, Ventura A, Thorpe M, Kryszak D, Fornaroli F, Wasserman SS, Murray JA, Horvath K. Prevalence of celiac disease in at-risk and not-at-risk groups in the United States: a large multicenter study. *Arch Intern Med* 2003;163:286–92.
- [3] Zevallos VF, Raker V, Tenzer S, Jimenez-Calvente C, Ashfaq-Khan M, Rüssel N, Pickert G, Schild H, Steinbrink K, Schuppan D. Nutritional Wheat Amylase-Trypsin Inhibitors Promote Intestinal Inflammation via Activation of Myeloid Cells. *Gastroenterology* 2017;152:1100-1113.e12.
- [4] Fasano A, Sapone A, Zevallos V, Schuppan D. Nonceliac gluten sensitivity. *Gastroenterology* 2015;148:1195–204.
- [5] Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. *Nat Rev Immunol* 2002;2:647–55.
- [6] Ricci G, Andreozzi L, Cipriani F, Giannetti A, Gallucci M, Caffarelli C. Wheat Allergy in Children: A Comprehensive Update. *Medicina (Kaunas)* 2019;55.
- [7] Tanabe S. Analysis of food allergen structures and development of foods for allergic patients. *Biosci Biotechnol Biochem* 2008;72:649–55.
- [8] Junker Y, Zeissig S, Kim S-J, Barisani D, Wieser H, Leffler DA, Zevallos V, Libermann TA, Dillon S, Freitag TL, Kelly CP, Schuppan D. Wheat amylase trypsin inhibitors drive intestinal inflammation via activation of toll-like receptor 4. *J Exp Med* 2012;209:2395–408.
- [9] Osborne T. The vegetable proteins. 2nd ed. London: Longmans, Green and Co; 1924.
- [10] Shewry PR, Halford NJ, Leffrand D. Genetics of Wheat Gluten Proteins. *Advances in Genetics*. 2003;49:111–84.
- [11] Takizawa T, Arakawa H, Tokuyama K, Morikawa A. Identification of allergen fractions of wheat flour responsible for anaphylactic reactions to wheat products in infants and young children. *Int Arch Allergy Immunol* 2001;125:51–6.
- [12] Lexhaller B, Colgrave ML, Scherf KA. Characterization and Relative Quantitation of Wheat, Rye, and Barley Gluten Protein Types by Liquid Chromatography-Tandem Mass Spectrometry. *Front Plant Sci* 2019;10:1530.
- [13] Colgrave ML, Goswami H, Byrne K, Blundell M, Howitt CA, Tanner GJ. Proteomic profiling of 16 cereal grains and the application of targeted proteomics to detect wheat contamination. *J Proteome Res* 2015;14:2659–68.

- [14] Afzal M, Pfannstiel J, Zimmermann J, Bischoff SC, Würschum T, Longin CFH. High-resolution proteomics reveals differences in the proteome of spelt and bread wheat flour representing targets for research on wheat sensitivities. *Scientific reports* 2020;10:14677.
- [15] Prandi B, Tedeschi T, Folloni S, Galaverna G, Sforza S. Peptides from gluten digestion: A comparison between old and modern wheat varieties. *Food Res Int* 2017;91:92–102.
- [16] Bischoff S, Crowe SE. Gastrointestinal food allergy: new insights into pathophysiology and clinical perspectives. *Gastroenterology* 2005;128:1089–113.
- [17] Dupont FM, Vensel WH, Tanaka CK, Hurkman WJ, Altenbach SB. Deciphering the complexities of the wheat flour proteome using quantitative two-dimensional electrophoresis, three proteases and tandem mass spectrometry. *Proteome Sci* 2011;9:10.
- [18] Simonato B, Pasini G, Giannattasio M, Peruffo AD, Lazzari F de, Ciminelli A. Food allergy to wheat products: the effect of bread baking and in vitro digestion on wheat allergenic proteins. A study with bread dough, crumb, and crust. *J Agric Food Chem* 2001;49:5668–73.
- [19] Laatikainen R, Koskenpato J, Hongisto S-M, Lopenen J, Pounsa T, Huang X, Sontag-Strohm T, Salmenkari H, Korpela R. Pilot Study: Comparison of Sourdough Wheat Bread and Yeast-Fermented Wheat Bread in Individuals with Wheat Sensitivity and Irritable Bowel Syndrome. *Nutrients* 2017;9.
- [20] Rizzello CG, Angelis M de, Coda R, Gobetti M. Use of selected sourdough lactic acid bacteria to hydrolyze wheat and rye proteins responsible for cereal allergy. *Eur Food Res Technol* 2006;223:405–11.
- [21] Davis PJ, Smales CM, James DC. How can thermal processing modify the antigenicity of proteins? *Allergy* 2001;56 Suppl 67:56–60.
- [22] El Mecherfi K-E, Todorov SD, Chevalcanti de Albuquerque MA, Denery-Papini S, Lupi R, Haertlé T, Dora Gombossy de Melo Franco B, Larré C. Allergenicity of Fermented Foods: Emphasis on Seeds Protein-Based Products. *Foods* 2020;9.
- [23] Afzal M, Sielaff M, Curella V, Neerukonda M, El Hassouni K, Schuppan D, Tenzer S, Longin CFH. 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 (Basel)* 2021;10.
- [24] Sielaff M, Curella V, Neerukonda M, Afzal M, El Hassouni K, Distler U, Schuppan D, Longin CFH, Tenzer S. Hybrid QconCAT-Based Targeted Absolute and Data-Independent Acquisition-Based Label-Free Quantification Enables In-Depth Proteomic Characterization of Wheat Amylase/Trypsin Inhibitor Extracts. *J Proteome Res* 2021.
- [25] Pilolli R, Angelis M de, Lamonaca A, Angelis E de, Rizzello CG, Siragusa S, Gadaleta A, Mamone G, Monaci L. Prototype Gluten-Free Breads from Processed Durum Wheat: Use of Monovarietal Flours and Implications for Gluten Detoxification Strategies. *Nutrients* 2020;12.

- [26] Islam S, Ma W, Yan G, Gao L, Appels R. Differential recovery of lupin proteins from the gluten matrix in lupin-wheat bread as revealed by mass spectrometry and two-dimensional electrophoresis. *J Agric Food Chem* 2011;59:6696–704.
- [27] Bose U, Juhász A, Broadbent JA, Byrne K, Howitt CA, Colgrave ML. Identification and Quantitation of Amylase Trypsin Inhibitors Across Cultivars Representing the Diversity of Bread Wheat. *J Proteome Res* 2020.
- [28] Wessel D, Flügge UI. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical Biochemistry* 1984;138:141–3.
- [29] Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 1976;72:248–54.
- [30] Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2007;2:1896–906.
- [31] Olsen JV, Godoy LMF de, Li G, Macek B, Mortensen P, Pesch R, Makarov A, Lange O, Horning S, Mann M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* 2005;4:21010–21.
- [32] Miller RM, Millikin RJ, Hoffmann CV, Solntsev SK, Sheynkman GM, Shortreed MR, Smith LM. Improved Protein Inference from Multisite Protease Bottom-Up Mass Spectrometry Data. *J Proteome Res* 2019;18:3429–38.
- [33] UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 2017;45:D158-D169.
- [34] Millikin RJ, Solntsev SK, Shortreed MR, Smith LM. Ultrafast Peptide Label-Free Quantification with FlashLFQ. *J Proteome Res* 2018;17:386–91.
- [35] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria; 2018.
- [36] Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 2016;13:731–40.
- [37] Juhász A, Belova T, Florides CG, Maulis C, Fischer I, Gell G, Birinyi Z, Ong J, Keeble-Gagnère G, Maharajan A, Ma W, Gibson P, Jia J, Lang D, Mayer KFX, Spannagl M, Tye-Din JA, Appels R, Olsen O-A. Genome mapping of seed-borne allergens and immunoresponsive proteins in wheat. *Sci Adv* 2018;4:eaar8602.
- [38] Radauer C, Bublin M, Wagner S, Mari A, Breiteneder H. Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *J Allergy Clin Immunol* 2008;121:847-52.e7.
- [39] Food Allergy Research and Resource Program; Available from: <http://www.allergenonline.org/>.

- [40] Mari A, Scala E. Allergome: a unifying platform. *Arb Paul Ehrlich Inst Bundesamt Sera Impfstoffe Frankf A M* 2006;29-39; discussion 39-40.
- [41] Mari A, Rasi C, Palazzo P, Scala E. Allergen databases: current status and perspectives. *Curr Allergy Asthma Rep* 2009;9:376–83.
- [42] <http://www.allergome.org>. Allergome. [July 09, 2020]; Available from: <http://www.allergome.org/>.
- [43] Skodje GI, Sarna VK, Minelle IH, Rolfsen KL, Muir JG, Gibson PR, Veierød MB, Henriksen C, Lundin KEA. Fructan, Rather Than Gluten, Induces Symptoms in Patients With Self-Reported Non-Celiac Gluten Sensitivity. *Gastroenterology* 2018;154:529-539.e2.
- [44] Vijn I, Smeekens S. Fructan: more than a reserve carbohydrate. *Plant Physiol* 1999;120:351–60.
- [45] Whelan K, Abrahamsohn O, David GJP, Staudacher H, Irving P, Lomer MCE, Ellis PR. Fructan content of commonly consumed wheat, rye and gluten-free breads. *Int J Food Sci Nutr* 2011;62:498–503.
- [46] Miedaner T. Kulturpflanzen: Botanik -- Geschichte -- Perspektiven. Place of publication not identified: Springer Science and Business Media; 2014.
- [47] Rizzello CG, Angelis M de, Di Cagno R, Camarero F, Silano M, Losito I, Vincenzi M de, Bari MD de, Palmisano F, Maurano F, Gianfrani C, Gobetti M. Highly efficient gluten degradation by lactobacilli and fungal proteases during food processing: new perspectives for celiac disease. *Appl Environ Microbiol* 2007;73:4499–507.
- [48] Heredia-Sandoval NG, Valencia-Tapia MY, La Calderón de Barca AM, Islas-Rubio AR. Microbial Proteases in Baked Goods: Modification of Gluten and Effects on Immunogenicity and Product Quality. *Foods* 2016;5.
- [49] Singh H. A study of changes in wheat protein during bread baking using SE-HPLC. *Food Chemistry* 2005;90:247–50.
- [50] Shewry PR, Halford NG. Cereal seed storage proteins: structures, properties and role in grain utilization. *J Exp Bot* 2002;53:947–58.
- [51] Heshe GG, Haki GD, Woldegiorgis AZ, Gemede HF. Effect of conventional milling on the nutritional value and antioxidant capacity of wheat types common in Ethiopia and a recovery attempt with bran supplementation in bread. *Food Sci Nutr* 2015;4:534–43.
- [52] Pedersen B, Knudsen KE, Eggum BO. Nutritive value of cereal products with emphasis on the effect of milling. *World Rev Nutr Diet* 1989;60:1–91.
- [53] Biliaderis C.G, Izydorczyk M.S., Rattan O. Effect of arabinoxylans on bread-making quality of wheat flours. *Food Chemistry* 1995;53:165–71.
- [54] Michniewicz J, Biliaderis C.G, Bushuk W. Water-Insoluble Pentosans of Wheat: Composition and Some Physical Properties. *Cereal Chem.* 1990;67:434–9.

- [55] Buksa K, Ziobro R, Nowotna A, Gambuś H. The influence of native and modified arabinoxylan preparations on baking properties of rye flour. *Journal of Cereal Science* 2013;58:23–30.
- [56] Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang H-Y, El-Gebali S, Fraser MI, Gough J, Haft DR, Huang H, Letunic I, Lopez R, Luciani A, Madeira F, Marchler-Bauer A, Mi H, Natale DA, Necci M, Nuka G, Orengo C, Pandurangan AP, Paysan-Lafosse T, Pesseat S, Potter SC, Qureshi MA, Rawlings ND, Redaschi N, Richardson LJ, Rivoire C, Salazar GA, Sangrador-Vegas A, Sigrist CJA, Sillitoe I, Sutton GG, Thanki N, Thomas PD, Tosatto SCE, Yong S-Y, Finn RD. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* 2019;47:D351-D360.
- [57] Altenbach SB, Tanaka CK, Hurkman WJ, Vensel WH. Expression of globulin-2, a member of the cupin superfamily of proteins with similarity to known food allergens, is increased under high temperature regimens during wheat grain development. *Journal of Cereal Science* 2009;49:47–54.
- [58] Caruso C, Nobile M, Leonardi L, Bertini L, Buonocore V, Caporale C. Isolation and amino acid sequence of two new PR-4 proteins from wheat. *J Protein Chem* 2001;20:327–35.
- [59] Mills ENC, Jenkins JA, Alcocer MJC, Shewry PR. Structural, biological, and evolutionary relationships of plant food allergens sensitizing via the gastrointestinal tract. *Crit Rev Food Sci Nutr* 2004;44:379–407.
- [60] Goodman RE. Practical and predictive bioinformatics methods for the identification of potentially cross-reactive protein matches. *Mol Nutr Food Res* 2006;50:655–60.
- [61] Angelis M de, Rizzello CG, Scala E, Silvano C de, Farris GA, Turrini F, Gobbetti M. Probiotic preparation has the capacity to hydrolyze proteins responsible for wheat allergy. *J Food Prot* 2007;70:135–44.
- [62] Di Cagno R, Angelis M de, Auricchio S, Greco L, Clarke C, Vincenzi M de, Giovannini C, D'Archivio M, Landolfo F, Parrilli G, Minervini F, Arendt E, Gobbetti M. Sourdough Bread Made from Wheat and Nontoxic Flour and Started with Selected Lactobacilli Is Tolerated in Celiac Sprue Patients. *Appl Environ Microbiol* 2004;70:1088–96.
- [63] Loponen J, Mikola M, Katina K, Sontag-Strohm T, Salovaara H. Degradation of HMW Glutenins During Wheat Sourdough Fermentations. *Cereal Chemistry Journal* 2004;81:87–93.
- [64] Kilstrup M, Hammer K, Ruhdal Jensen P, Martinussen J. Nucleotide metabolism and its control in lactic acid bacteria. *FEMS Microbiol Rev* 2005;29:555–90.

Figure legends

Figure 1. Flours and corresponding bread types investigated in this study.

In total, 14 breads and 6 corresponding flours derived from 3 grains were examined.

Further details see text.

Figure 2. Total number of proteins detected in flours (A) and breads (B) by nano-LC-ESI-MS/MS as well as overlap of proteins between flours (C) and breads (D) of wheat, spelt and rye.

(A): Number of different proteins in wholegrain (WG) and superfine (SF) flour from wheat, spelt and rye. (B): Number of different proteins in yeast fermented bread (Y) and sourdough fermented bread (S) from SF and WG flour. Two sourdough wheat breads additionally contained bread improver (SB). Means of three replicates and the standard deviation (SD) are shown. * $P < 0.05$; ** $P < 0.01$, *** $P < 0.001$. (C) and (D). Venn Diagrams showing the overlap of common and unique proteins in WC flour and WG bread of wheat, spelt and rye. Percentage values were rounded without decimal places.

Figure 3. Overlap of proteins between different flour and bread types, respectively.

(A, C, E): Venn diagrams show the overlap between WG and SF flours in wheat, spelt and rye. (B, D, F): Overlap between different manufactured breads (Y=yeast fermented; S=sourdough fermented; SB=sourdough fermented with bread improver) from wheat, spelt and rye (percentage values were rounded to zero decimal places).

Figure 4. Protein clusters characteristic for flour or bread types.

Heat map (left) and profile plot (right) of z-scored protein abundances ($\log_2$ LFQ intensity) include 1737 differentially abundant proteins identified by Welch's T-Test. Mean values (three replicates) of normalized LFQ intensities for each sample after hierarchical clustering of different fermentation conditions (yeast and sourdough) and different

milling types (WG and SF) in wheat, spelt and rye samples are shown. Enriched proteins are shown in red while depleted proteins are shown in blue. Samples with bread improver were excluded from this analysis, because the commercial bread improver contained unknown proteins from plant sources. Fisher's exact test was performed to identify significantly enriched or depleted Pfam groups in the two most prominent clusters.

Figure 5. Relative abundance of allergenic proteins. (A) Histograms show the distribution of proven or potentially allergenic proteins (orange, N=246) compared to other proteins (blue, N=4178) in WG flour of wheat, spelt and rye. (B+C): Allergenic proteins were further divided in "potential allergens" and "proven allergens". Box plots show abundances of potential allergens and proven allergens in different flour or bread types. Only proteins were considered that were identified in at least one replicate of WG flour from wheat, spelt and rye (B) or in each WG wheat bread type and the flour (C). Statistical significance by one-way ANOVA: *P < 0.05; **P < 0.01, ***P < 0.001. Abbreviations: Y, yeast fermented; S, sourdough fermented; SB, sourdough fermented with bread improver.

Significance

Our proteomics study revealed that bread contains a number of potentially and proven allergenic proteins. Most likely allergenicity is not dependent on milling or conventional fermentation processes, but on the grain type. Relative abundance of allergenic proteins was higher in spelt- and wheat samples than in rye samples. Considering rye bread as better suited to atopic individuals predisposed to react to cereal allergens, clinical trials are warranted to verify this assumption.

Highlights

- Comprehensive proteomics analysis of bread proteome from bread wheat, spelt and rye
- Higher numbers of proteins were identified in flour compared to bread
- Species type has greater influence on the proteome than fermentation processes
- Potential allergens in flour and bread were defined based on structure annotation
- Lower abundance of potential allergens proteins in rye compared to wheat and spelt

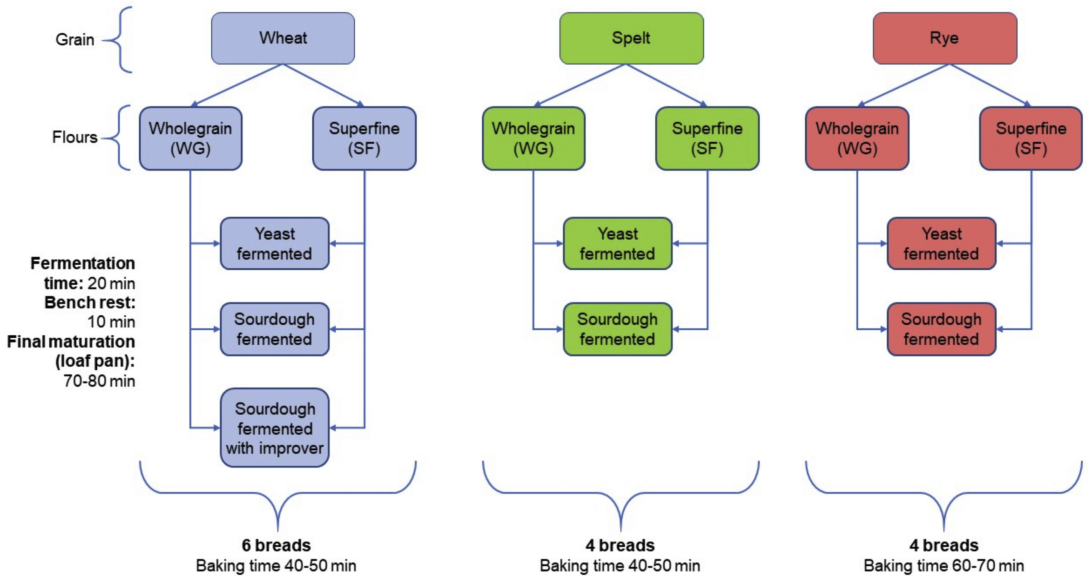


Figure 1

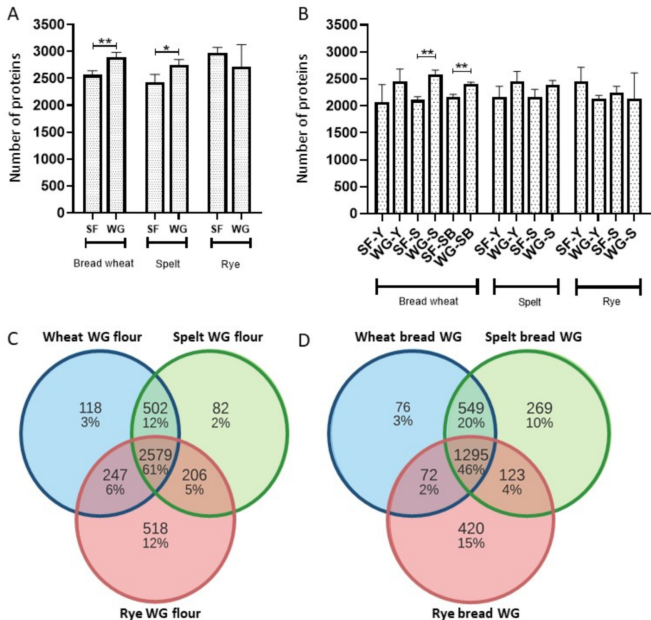


Figure 2

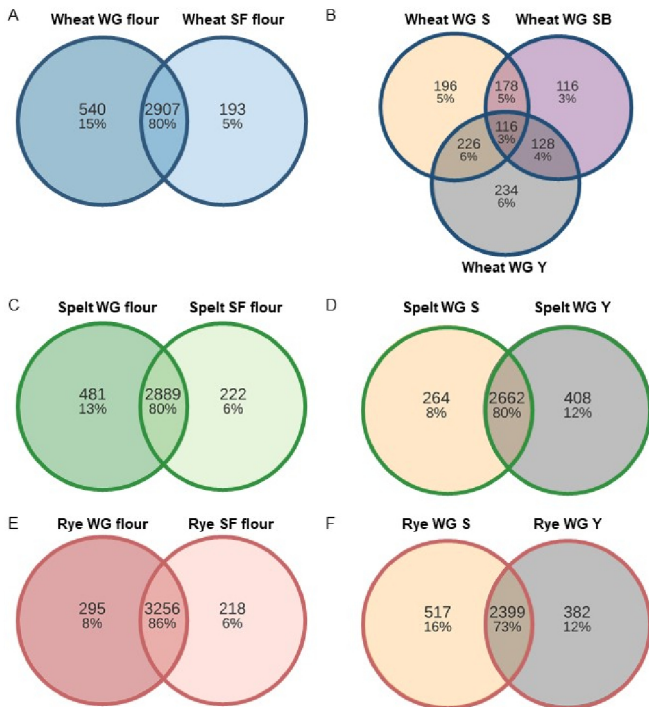


Figure 3

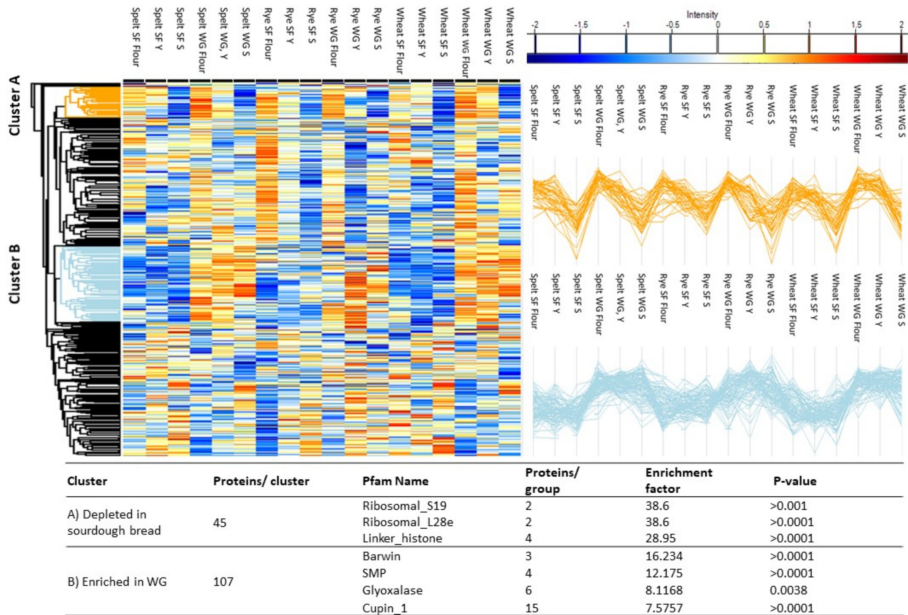


Figure 4

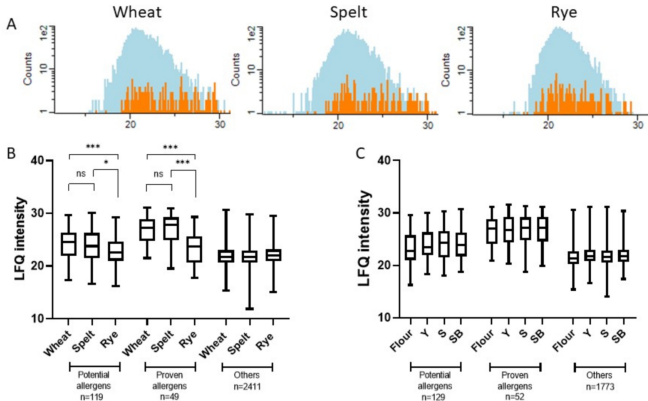


Figure 5