A multi-centre peptidomics investigation of food digesta: current state of the art in mass spectrometry analysis and data visualisation

Reto Portmann, Pablo Jiménez-Barrios, Julien Jardin, Lychou Abbühl, Daniela Barile, Marianne Danielsen, Yu-Ping Huang, Trine Kastrup Dalsgaard, Beatriz Miralles, Valérie Briard-Bion, et al.

Food Research International 169 (2023) 112887

Contents lists available at ScienceDirect

Food Research International

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

A multi-centre peptidomics investigation of food digesta: current state of the art in mass spectrometry analysis and data visualisation

Reto Portmann [a,1], Pablo Jiménez-Barrios [b,1], Julien Jardin [c], Lychou Abbühl [a], Daniela Barile [d], Marianne Danielsen [g,h], Yu-Ping Huang [d], Trine Kastrup Dalsgaard [g,h], Beatriz Miralles [b], Valérie Briard-Bion [c], Stefano Cattaneo [e], Christophe Chambon [f], Benoit Cudennec [j], Ivano De Noni [e], Barbara Deracinois [j], Didier Dupont [c], Angéline Duval [f], Christophe Flahaut [j], Rubén López-Nicolás [i], Sedef Nehir El [k], Valentina Pica [e], Véronique Santé-Lhoutellier [f], Milda Stuknytė [e,2], Laetitia Theron [f], Thierry Sayd [f], Isidra Recio [b,1], Lotti Egger [a,*,1]

[a] Agroscope, Schwarzenburgstr, 161, 3003 Bern, Switzerland
[b] Instituto de Investigación en Ciencias de la Alimentación (CIAL, CSIC-UAM), Madrid, Spain
[c] INRAE, Institut Agro, STLO, 35000 Rennes, France
[d] Department of Food Science and Technology, University of California, Davis, One Shields Avenue, Davis, CA 95616, United States
[e] University of Milan, Department of Food, Environmental and Nutritional Sciences, via G. Celoria 2, 20133 Milan, Italy
[f] INRAE, UR370 Qualité des Produits Animaux and/or PFEM CP, F-63122 Saint Genès-Champanelle, France
[g] Department of Food Science, Aarhus University, Centre for Innovative Food Research (CiFood), Agro Food Park 48, 8200 Aarhus, Denmark
[h] Centre for Circular Bioeconomy (CBIO), lichers Allé 20, 8830 Tjele, Denmark
[i] Department of Food Science and Human Nutrition, Universidad de Murcia, Campus de Espinardo, 30100 Murcia, Spain
[j] UMR Transfronalière BioEcoAgro-INRae 1158, Univ. Lille, INRAE, Univ. Liège, UPJV, JUNIA, Univ. Artois, Univ. Littoral Côte d'Ôpale, ICV–Institut Charles Viollette, F-59000 Lille, France
[k] Department of Food Engineering, Faculty of Engineering, Ege University, 35100 Izmir, Turkey

ARTICLE INFO

Keywords:
Bioinformatics
In vitro digestion
Mass spectrometry
Peptidomics
Protein digestion
Visualisation

ABSTRACT

Mass spectrometry has become the technique of choice for the assessment of a high variety of molecules in complex food matrices. It is best suited for monitoring the evolution of digestive processes *in vivo* and *in vitro*. However, considering the variety of equipment available in different laboratories and the diversity of sample preparation methods, instrumental settings for data acquisition, statistical evaluations, and interpretations of results, it is difficult to predict *a priori* the ideal parameters for optimal results. The present work addressed this uncertainty by executing an inter-laboratory study with samples collected during *in vitro* digestion and presenting an overview of the state-of-the-art mass spectrometry applications and analytical capabilities available for studying food digestion. Three representative high-protein foods – skim milk powder (SMP), cooked chicken breast and tofu – were digested according to the static INFOGEST protocol with sample collection at five different time points during gastric and intestinal digestion. Ten laboratories analysed all digesta with their in-house equipment and applying theirconventional workflow.

The compiled results demonstrate in general, that soy proteins had a slower gastric digestion and the presence of longer peptide sequences in the intestinal phase compared to SMP or chicken proteins, suggesting a higher resistance to the digestion of soy proteins. Differences in results among the various laboratories were attributed more to the peptide selection criteria than to the individual analytical platforms. Overall, the combination of mass spectrometry techniques with suitable methodological and statistical approaches is adequate for contributing to the characterisation of the recently defined digestome.

Abbreviations: ACN, acetonitrile; EtOH, ethanol; ESI, electrospray ionisation; IVD, *In vitro* digestion; LC-MS, liquid chromatography-mass spectrometry; MS, mass spectrometry; NPN, non-protein nitrogen; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; SMP, skim milk powder; SPE, solid phase extraction; TCA, trichloroacetic acid; TFA, TN, total nitrogen trifluoroacetic acid; MALDI, matrix-assisted laser desorption ionisation; qTOF, quadrupole-time-of-flight.

* Corresponding author.
[1] Equal contribution.
[2] Present address: University of Milan, Unitech COSPECT, via C. Golgi 19, 20133 Milan, Italy.

https://doi.org/10.1016/j.foodres.2023.112887
Received 15 December 2022; Received in revised form 14 April 2023; Accepted 21 April 2023
Available online 29 April 2023

1. Introduction

During gastrointestinal digestion, the coordinated action of gastric and pancreatic proteases transforms ca. 70% of ingested protein into peptides, while the remaining 30% is transformed to free amino acids (Boron & L., 2009). Further hydrolysis by a large number of peptidases at the intestinal brush border and at the cytoplasm of epithelial cells completes the digestion of peptides in a very efficient way, as less than 4% of the ingested protein is excreted intact in the stool. The relevance of protein digestion products has been highlighted due to their involvement in the regulation of digestion, nutrient metabolism and other physiological processes, such as immune function and allergy. For instance, intestinal epithelial cells express a large number of receptors capable of sensing the composition of the luminal content and, in particular, protein breakdown products. It is recognised that luminal peptides influence intestinal function by affecting barrier regulation and the gut-associated immune system (Martínez-Augustin, Rivero-Gutiérrez, Mascaraque, & Sánchez de Medina, 2014). In a similar way, the activation of enteroendocrine cells by protein digestion products triggers the release of hormones that modulate food intake and control glucose homeostasis (Caron, Domenger, Dhulster, Ravallec, & Cudennec, 2017; Santos-Hernández, Miralles, Amigo, & Recio, 2018).

However, the composition of the luminal digest after protein ingestion was considered a 'black box' until the development of peptidomic strategies based on mass spectrometry (MS). These advances have allowed the quantitative mapping of peptides and amino acids, which has been defined as *digestomics* (Bingeman, Perlman, Storey, & Lewis, 2017; Picariello, Mamone, Addeo, Nitride, & Ferranti, 2013). In the last two decades, the peptide profile of *in vivo* gastrointestinal digests from several protein-rich foods has been studied in human (Boutrou et al., 2013) and animal models, mainly pigs and mini-pigs (Bauchart et al., 2007; Egger et al., 2017). The peptidome profile of human milk and gastric digesta from term infants revealed the presence of previously described sequences with immunomodulatory and antibacterial properties of clinical relevance (Dallas et al., 2014). The identification of allergenic epitopes released during digestion was also undertaken to study allergy and intolerance to milk, wheat and nuts, among other foods (Picariello et al., 2013). Many MS-based techniques have also been extensively applied to investigate the effect of food processing, heat treatment, protein hydrolysis or food additives on protein digestion. For instance, the influence of nitrosation on protein digestibility (Manguy et al., 2017), the characterisation of milk protein hydrolysates used as human milk fortifiers (Pica, Stuknytė, Masotti, De Noni, & Cattaneo, 2021) or the enzyme susceptibility of lactosylated milk proteins (Dalsgaard, Nielsen, & Larsen, 2007; Zhao et al., 2019) were investigated using their digestome profile.

Despite the considerable progress made in the digestomic area, some challenges remain to be addressed throughout the data acquisition and analysis process. First, sample preparation is essential because of the risk of losing peptides during each preparation step, such as centrifugation, protein precipitation and peptide purification. Therefore, the risk of losing information on the peptidome from digesta is high, and each step must be considered with care (Dupont, 2017). With that in mind, several areas deserve particular consideration in order to circumvent the limitations of MS-based technologies. First and foremost, one must remember that most of the tools currently used to identify peptides from digesta were developed to meet the needs of proteomics (Perkins, Pappin, Creasy, & Cottrell, 1999). As such, only peptides in the molecular mass range of 500–3500 Da can be unambiguously and automatically identified by a database search. The identification of shorter peptides would require labour-intensive sample preparation as well as the use of dedicated *de novo* sequencing software and an optimised liquid chromatography-mass spectrometry (LC-MS) setup for smaller peptides. An additional challenge comes from the fact that di- and tri-peptides resulting from digestion are polar molecules and are as such poorly retained on a C_{18} chromatographic column, which is the standard in proteomic studies. Obtaining the most exhaustive view of the peptidome would require performing multiple LC-MS runs with different LC conditions in terms of both the type of column and the gradient of solvents. Recently reported progress in the latter area is promising (Harscoat-Schiavo et al., 2012; Le Maux, Nongonierma, & FitzGerald, 2015; Panchaud, Affolter, & Kussmann, 2012; Piovesana et al., 2019) but still ongoing. Given the multitude of possibilities, it is clear that there is a need to harmonise technical approaches in order to compare the results within the existing body of literature.

The need for harmonisation is equally urgent regarding the bioinformatic steps used to interpret and compare the digestome data. The multitude of tools available to process the massive amount of data resulting from the MS analysis of digesta is constantly evolving. Since bioinformatic analysis plays a crucial role in determining the final list of peptides reported in the digestome dataset, it is of great importance to master these tools and to reach a consensus within the scientific community about which tools are the most appropriate. Within INFOGEST, a subgroup of 10 laboratories collaborated internationally in an inter-laboratory ring trial and used their available in-hour equipment and bioinformatics tools to identify and semi-quantify peptides released in the same set of protein-based food digesta. The main objectives of this work were to determine whether peptidomic analyses performed independently in several laboratories using diverse conditions could produce similar conclusions (i.e. identification of the same domains in proteins resistant to digestion) and provide comparable results even when using different bioinformatics tools to handle and compare peptidomic data.

2. Materials and methods

2.1. Chemicals and reagents

All chemicals for *in vitro* digestion were of analytical grade and, unless otherwise noted, were purchased at Merck. Each participating laboratory, whose identity has been anonymised (listed as Labs 1–10), used its own chemicals, which were of analytical grade, for MS analysis.

2.2. Sample preparation and in vitro digestion (IVD)

Skim milk powder (SMP), tofu and chicken breast were bought at a local retailer (Switzerland) and analysed for total nitrogen (TN) with the Kjeldahl method (ISO-8968-3, 2004). Unless stated otherwise, milk refers to bovine milk. The protein content of each food was calculated with a conversion factor of 6.25 for tofu and chicken breast and 6.38 for SMP (Table 1). Enzyme activity determination and *in vitro* digestions were performed using the INFOGEST static protocol according to (Minekus et al., 2014). The three different food samples were normalised at 200 mg of total protein content and the volume was completed with simulated salivary fluid to yield 5 mL of total meal. Each digestion time point was performed in a separate tube, stopping after 0.1, 5, 15, 60 and 120 min of gastric or intestinal digestion, respectively. The gastric phase was

Table 1
TN content, protein content and IVD input for the three substrates.

	TN (g/kg)	NPN (g/kg)	Conversion factor	Protein (g/kg)	Food for IVD (g)	Protein/mL in digest
SMP (1:10 diluted)	5.9	0.324	6.38	37.6	1	0.038
Chicken	42.8		6.25	267.5	0.1495	0.040
Tofu	21.3		6.25	133.1	0.3007	0.040

stopped by raising the pH to 7 with NaOH (5 mol/L), and the intestinal phase was stopped by adding 1 mmol/L of protease inhibitor 4-(2 aminoethyl) benzensulfonylfluoride (AEBSF, trademark Pefabloc®, 500 mmol/L, Roche, Basel, Switzerland) after which the samples were immediately snap-frozen in liquid nitrogen. After defrosting the samples, they were centrifuged at 3500*g* at 4 °C for 15 min, and 1 mL aliquots of the supernatant were prepared and lyophilised for further analysis. The protein contents of the digested samples are shown in Table 1.

2.3. Sample preparation and gel electrophoresis

For gel electrophoresis, sample loadings were normalised according to the TN content of each food, taking into account the dilutions performed for the gastric and intestinal phases compared to the original food. Therefore, the samples were diluted – undigested (1:8), gastric (1:4) and intestinal (undiluted) – and mixed with sample buffer (Tris-HCl 350 mmol/L, pH = 6.8, sodium dodecyl sulfate (SDS) 10%, dithiothreitol (DTT) 100 mmol/L, glycerol 50%), after which they were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (15% polyacrylamide). A molecular weight marker (Benchmark™, Invitrogen) was added and the gels were stained with colloidal Coomassie Blue as described previously (Kang, Gho, Suh, & Kang, 2002).

2.4. Sample preparation for LC-MS analysis

Freeze-dried samples were reconstituted in 1 mL of ammonium bicarbonate (NH_4HCO_3) buffer (25 mmol/L) or 0.1% trifluoroacetic acid (TFA) at 30 °C for 2 h (Lab 9) and either (ultra)filtred with syringe filters (0.45 μm, Lab 1) or micro-spin devices (10–30 kDa MWCO, centrifuging at 4000–14000*g*, Labs 2, 3 (gastric digests only) and 8) or only centrifuged (12500–16000*g*, 4–10 °C, 5–30 min, Labs 3 (intestinal digests) and 6) before being analysed by LC-MS. In some laboratories, an additional protein precipitation step was performed to remove intact proteins (e.g. trichlorcetic acid (TCA) 15% (Lab 4), 1–2 volumes of ethanol (EtOH) (Labs 7 and 10), 4 volumes of acetonitrile (ACN) with 0.1% formic acid (FA) (Lab 5), or 0.1% trifluoroacetic acid (TFA), (Lab 9)), followed by centrifugation at 8000–15000*g* for 10–15 min (Labs 5 and 9), filtration (0.22 μm, Lab 10) or column purification (Labs 4 and 7), or with a polymer-based hydrophobic resin (Pierce™ Peptide Desalting Spin Column, Thermofisher, Lab 5). An expanded version of the sample preparation protocols for each laboratory is presented in the Supplemental Materials (section 1).

2.5. Liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis and peptide identification

The aim of this ring trial was to examine the selected protein-rich food samples using the LC-MS equipment, methods and software pipelines individually available in the participating laboratories. It is important to note the different performances of the various MS instruments used and accordingly, no instructions were provided for carrying out the analyses as the goal was to capture a wide variety of setups but in general, analyses conditions were focused on the detection of peptides longer than 5 amino acids. Ultra-high pressure liquid chromatography (UPLC) coupled to C_{18}-based columns was used in the majority of laboratories for peptide separation (details for each laboratory are found in the Supplemental Materials (section 2)). Nearly all laboratories used long gradients, between 40 min and 115 min, that started from a high aqueous phase (H_2O + 0.1% formic acid) and moved to a high organic phase (ACN + 0.1% formic acid), which was gradually increased from 5 to 95% to elute peptides. When shorter gradients were used, the same sample was usually injected several times to cover different mass over charge (m/z) ratio ranges, after which the results were combined.

Peptide ionisation was achieved by electrospray (ESI) in all laboratories. For peptide mass detection, a heterogeneous selection of mass spectrometers was used, including Orbitrap technology, quadrupole-time-of-flight (qTOF) and ion traps. The measured masses ranged from 300 to 2000 m/z on average and fragmentation was applied to the 5–20 most intense MS signals, measured in a data-dependent manner.

The identification of peptides from MS/MS data was performed with either the UniProt database (https://www.uniprot.org), which was restricted to the expected taxonomies, or in-house databases, each uniquely containing the main proteins of interest. The algorithms of the Mascot or Sequest HT search engines or the *de novo* sequencing Peaks® Studio software (Bioinformatics Solutions Inc., Waterloo, Ontario, Canada) algorithm were used. In each case, the technically induced modifications, such as oxidation of methionine, deamidation of asparagine and glutamine, and cyclisation of glutamine to pyroglutamate, were considered. Phosphorylation of serine and threonine (and, for some laboratories, tyrosine) was included (except for Lab 5) as a biological post-translational modification.

The post-processing of data was variable between laboratories, and some post-processing tools were based on the statistical program R, using custom scripts.

2.6. Matrix-assisted laser desorption/ionisation-time of flight-mass spectrometry (MALDI-TOF-MS) analysis

The digests were reconstituted in 1 mL NH_4HCO_3 and diluted 1:100 in water/0.1% FA (v/v). One μL of sample was directly spotted onto an MSP 96 polished steel target (Bruker Daltonics, Bremen, Germany), overlaid with 1 μL of α-Cyano-4-hydroxycinnamic acid matrix, in 50% ACN, containing 0.1% TFA (v/v), allowed to dry at room temperature, and analysed on an AutoFlex Speed™ instrument (Bruker Daltonics, Bremen, Germany). Three independent spectra for each sample were collected in automated mode, externally calibrated by using Peptide Calibration Standard II (Bruker Daltonics), and subsequently analysed with FlexAnalysis version 3.3 (Bruker Daltonics). Ions were detected in positive reflectron mode at a mass range of m/z 500–3500 Da.

2.7. Statistical analysis and graphical representation

Boxplot of peptide size distribution. Peptide size distribution was visualised with a boxplot generated by R programming (version 3.5.3) using the ggplot2 package (Wickham, 2016).

Heatmap of peptide patterns. The peptide patterns were generated by summing the number of individual amino acids within all identified peptides (Egger et al., 2017). The chosen colour code ranged from red, representing the maximal number of identified amino acids within the corresponding protein and digestion phase, to green, representing medium to low numbers of identified amino acids, to blue, representing the minimal number of identified amino acids. The gaps represent amino acids without identification within any peptide.

Peptigram. The visualisation of peptide distributions resulting from the digestion of β-casein, actin and glycinin 1 was mapped to their parent protein amino acid sequence using the free web application Peptigram (https://bioware.ucd.ie/peptigram/) (Manguy et al., 2017). Every amino acid covered by the identification of at least one peptide is represented by a vertical green bar. The height of the bar represents the number of times a peptide was identified at this position, and the intensity of green is proportional to the sum of peptide intensities overlapping this position.

Sequence logo with R. The sequence logo with the β-casein-derived sequences was obtained using the R package 'ggseqlogo' (Wagih, 2017). The sequence logo shows the probability of finding a particular amino acid at the N-terminal (1, 2, 3) or C-terminal (−3, −2, −1) end of the sequence, taking into account a set of peptides. In the case of five-length peptides, the amino acid in the 3rd position at both terminal ends was the same.

Venn diagram of resistant sequences. Identical peptides between laboratories were represented with Venn diagrams using the R package (Chen and Boutros, 2011). Duplicated peptides were removed for the total count.

3. Results

3.1. Characterisation of substrates

The main proteins in the undigested samples were identified by LC-MS (Fig. 1, undig. Numbers 1–15) by peptide mass fingerprinting as previously described (Sousa, Portmann, Dubois, Recio, & Egger, 2020). The kinetics of protein hydrolysis during *in vitro* digestion of the three different substrates were analysed by SDS-PAGE (Fig. 1). Intact proteins from all three protein sources were present during the gastric phase but no intact proteins were visible at any time point of the intestinal phase. The protein bands present in the intestinal phase were all attributed to the digestive enzymes present in the pancreatin suspension (data not shown).

3.2. Peptidomic analysis

For the three food matrices, digestion started rapidly since approximately 600 different peptide sequences were already detected after 5 min of digestion and the maximum number of peptides was reached after 15 min of digestion. An example of the graphic representation of these results is shown in Supplementary Fig. 1.

To generate Table 2, the peptide identification results from all 10 laboratories were grouped (numbers indicate unique sequences). The number of unique identified peptides from each major protein for each laboratory is shown in Supplemental Table 1. Only one major protein from each substrate was selected, and the redundant peptide sequences were counted only once: β-casein for SMP, glycine max β-conglycinin α subunit 1 for tofu and α-actin-1 for chicken breast. It was decided to keep only one protein per food matrix to facilitate result comparison between participating laboratories, and these three proteins were chosen since they gave the highest overall number of peptides across all laboratories.

For each of the three proteins, most of the identified peptide sequences were detected during the gastric phase. Indeed, 1119, 1107 and 1001 unique peptides sequences were identified for each protein in the gastric phase of SMP, tofu and chicken breast, respectively. During the intestinal phase, fewer peptides were detected since only 264 and 134 peptide sequences were identified from the major proteins of tofu and chicken breast, respectively. Nevertheless, up to 479 β-casein peptide sequences were resistant to the intestinal phase for SMP.

3.3. Sequence coverage

The percentage of sequence coverage of specific proteins as identified by the released peptides in the digesta is presented in Fig. 2. Overall, the changes in sequence coverage for the majority of proteins in the three foods followed a similar trend. Their sequence coverages considerably increased from gastric (G) digestion time G 0.1 min to G 15 min, plateaued through the end of the gastric phase, and plummeted when entering the intestinal phase. The changes revealed that peptides were released from various regions of the proteins during the gastric phase, which produced a rapid increase of sequence coverages; the released peptides were likely further broken down by pancreatic enzymes into much smaller pieces, including peptides and free amino acids, and hence led to the observed decline in sequence coverage in the intestinal phase. For the SMP, the sequence coverage of major milk proteins, including caseins and β-lactoglobulin, ranged between 27 and 64% at G 0.1 min and reached ~70–90% at the end of the gastric phase. The sequence coverages of these proteins decreased in the intestinal phase and were all below 30% at the end of the simulated digestion – except for β-casein, which still had 50% coverage at the intestinal (I) digestion time I 120 min. The coverage of lactoferrin was below 20% during the gastric phase and still had 9% coverage at the end of the simulated digestion. This might be caused by its relatively low abundance, and the glycosylation of the protein which hinders the detection of some peptides and

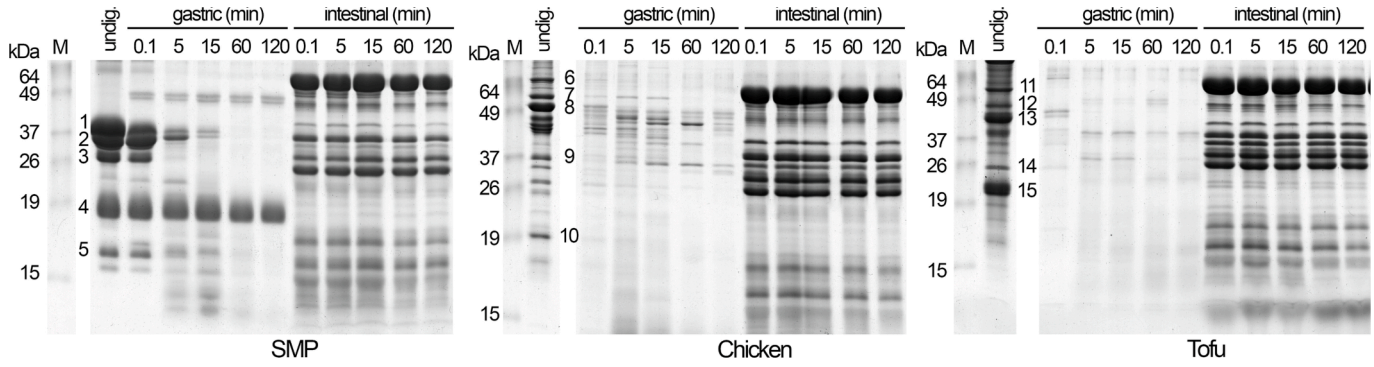

Fig. 1. SDS-PAGE displaying the kinetics during gastric and intestinal hydrolysis of proteins from skim milk powder (SMP), chicken breast (Chicken) and Tofu. Major proteins were identified by LC-MS: α_{s1}-casein (1, P02662), β-casein (2, P02666), κ-casein (3, P02668), β-lactoglobulin (4, P02754), α-lactalbumin (5, P00711), pyruvate kinase (6, P00548), actin α skeletal muscle (7, P68139), glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (8, P00356), troponin T (9, P12620), myosin regulatory light chain 2 (10, P24032), β-conglycinin α subunit (11, P11827), β-conglycinin β subunit (12, P25974), glycinin G2 (13, P04405), glycinin G4 (14, P02858) and glycinin G1 (15, P04776).

Table 2
Combined data from all laboratories of the total number of unique peptides identified from one major protein of each substrate during the gastric and intestinal phases of SMP (*Bos taurus*, bovine β-casein (UniProtKB – P02666)), tofu (Glycine max, soybean β-conglycinin α subunit 1, P0DO16) and chicken (*Gallus gallus*, actin, α skeletal muscle, P68139) digests. Modified sequences and their unmodified equivalent were counted only once, if both were present.

	Total	Gastric	Intestinal	Gastric time (min)					Intestinal time (min)				
Time (min)				*0*	*5*	*15*	*60*	*120*	*0*	*5*	*15*	*60*	*120*
SMP	1338	1119	479	395	593	776	773	652	272	306	257	245	190
Tofu	1287	1107	264	153	605	719	696	723	143	120	113	89	112
Chicken Breast	1060	1001	134	357	599	641	545	582	64	52	50	62	51
Total	3685	3227	877	905	1797	2136	2014	1957	479	478	420	396	353

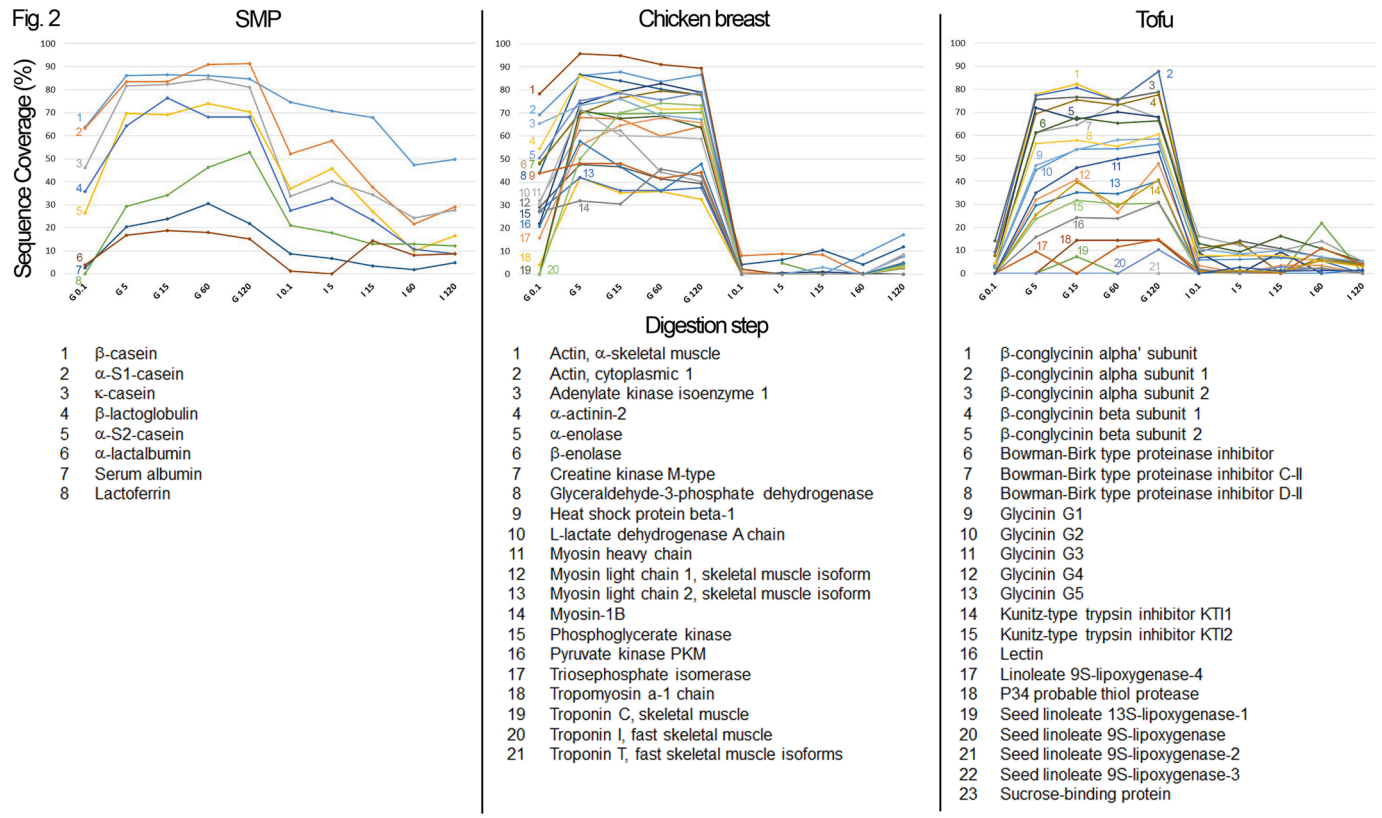

Fig. 2. Sequence coverage (%) of listed proteins per gastric (G) digestion and intestinal (I) time point obtained in SMP (A), chicken breast (B) and Tofu (C) digests shown for Lab 2.

increases the resistance of the protein to gastrointestinal digestion.

In the chicken digesta, some proteins had >65% sequence coverage at G 0.1 min, including glyceraldehyde-3-phosphate dehydrogenase (78%), actin, α skeletal muscle (69%) and myosin light chain 2, skeletal muscle isoform (65%). This implies the immediate proteolysis of these proteins at the initial stage of digestion. Many proteins reached >70% coverage between G 5 min and G 120 min, such as actin (α skeletal muscle), myosin light chain 2 (skeletal muscle isoform), myosin light chain 1 (skeletal muscle isoform), tropomyosin α-1 chain and several enzyme proteins. Despite the high sequence coverage of most proteins in the gastric phase, the coverage of all the proteins dropped below 10% at I 0.1 min and was maintained below 20% across the entire intestinal phase. This demonstrates the high digestibility of chicken proteins.

The sequence coverage of all the proteins in tofu at G 0.1 min was below 15%, much lower than what was observed for the major proteins in SMP and chicken. This indicates that the proteins in tofu were digested at a slower rate than in the other two foods. Some proteins in tofu reached >70% sequence coverage between G 5 min and G 120 min, including β-conglycinin (β subunits 1 and 2) and glycinin (G1 and G2; however, the sequence coverage was mostly below 85%). During the intestinal phase, the sequence coverage of all the proteins in tofu was mostly below 20%, similar to what was seen for chicken proteins.

3.4. Peptide size distribution

Peptide size distribution in the three foods at different digestion times is presented as boxplots in Fig. 3. The number of data points represents the number of peptides identified in each digesta. As expected, the gastric phase peptides encompassed a wider molecular mass range than the peptides identified in the intestinal phase. The medians of the molecular weight of the gastric phase peptides were also larger than those in the intestinal phase. Notably, in the case of tofu, the differences in molecular weight distribution and medians between the two phases were relatively small. The medians for peptide size in the gastric phase were similar for the three foods (1280–1610, 1259–1484 and 1314–1487 Da for SMP, chicken and tofu, respectively), whereas the medians of the intestinal phase were rather characteristic for each food with tofu (871–1082 Da) > SMP (846–878 Da) > chicken (738–762 Da). Peptide size was also measured using MALDI-TOF-MS (Supplemental Fig. 4).

3.5. Peptide patterns

Two types of post-processing methods were performed to compare different peptide patterns. Both aimed at the reduction in complexity of mass spectrometry data (Fig. 4). In the first method (amino acid counting method), instead of comparing individual identified peptides, the frequency of occurrence of the individual amino acids within the identified peptides was counted. In this way (not considering peptide intensity), the properties of individual peptides in mass spectrometry (charge, ionization capacity, and fragmentation behavior) were reduced. In detail, data-dependent MS/MS spectra were first recorded without exclusion list. In consequence, intense peaks eluting over a long time were sequenced several times, leading to a higher identification frequency. Subsequently, an increase in analytical sensitivity was achieved by measuring three overlapping mass ranges (300–1300 m/z with an overlap of 10 m/z). All identified peptides (some were identified multiple times due to the missing exclusion list) were arranged along the protein sequence and the frequency of each occurring amino acid in these peptides was counted and plotted as heat map. Shown are average results of three individual digestion experiments (Fig. 4A).

The second method (Peptigram) was based on the identification of peptides and the extraction of the intensity (area under the curve) of their respective masses in the MS spectrum, resulting in pairs of peptide sequence and corresponding intensity. These peptides were arranged along the protein sequence and the corresponding intensity was assigned

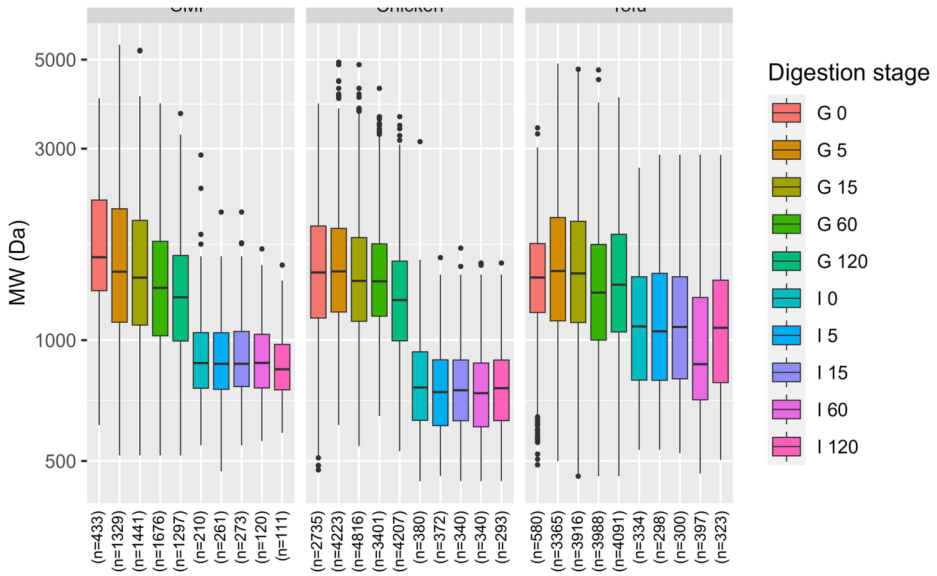

Fig. 3. Boxplot peptide size distribution generated by Lab 7 using a Q-Exactive Orbitrap. The number of data points (n) represents the number of peptides identified in each digesta.

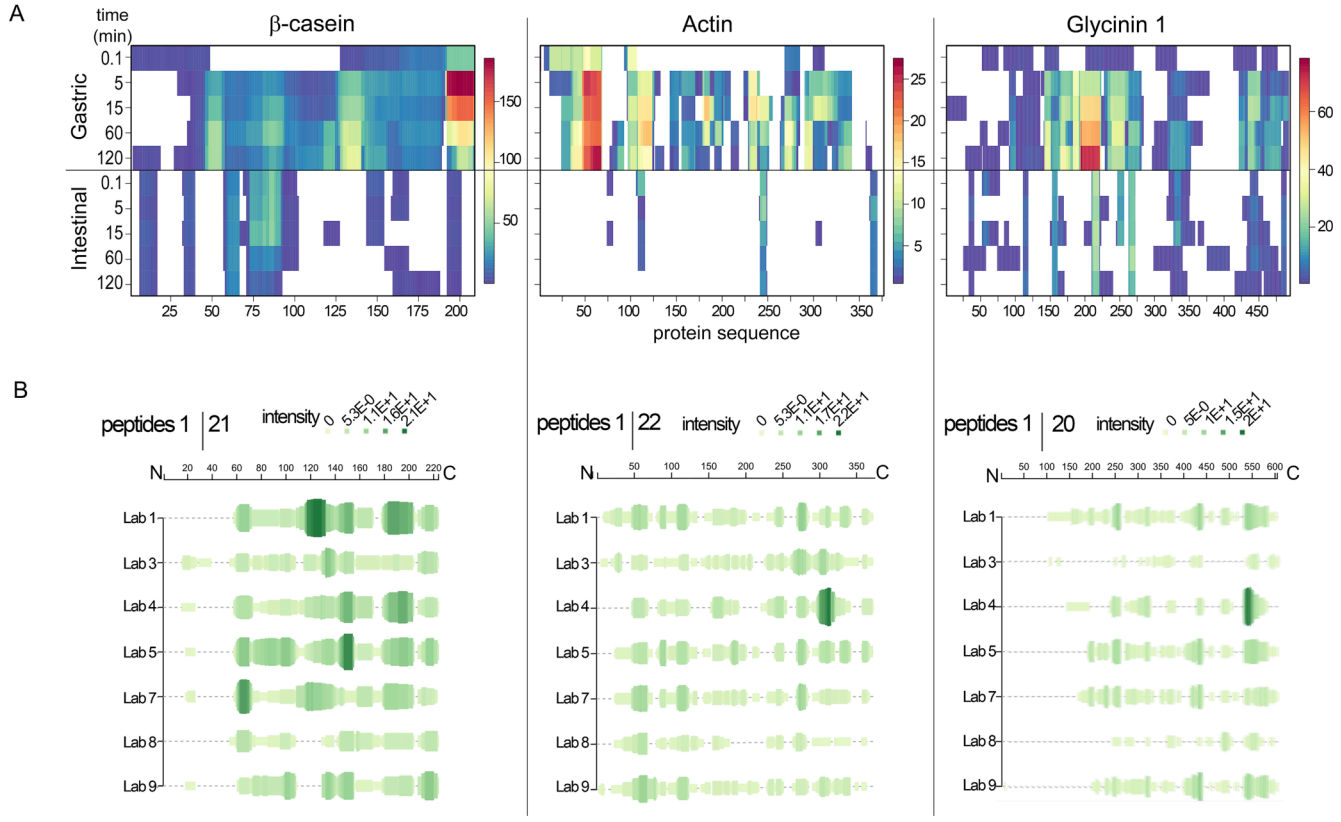

Fig. 4. Graphical representation of peptides from one main protein of the three substrates, β-casein from SMP, actin from chicken breast and glycinin from tofu, respectively. (A) Heatmap showing the evolution of peptide patterns during digestion. The protein sequence is shown on the x-axis and the number of times an amino acid was found within a peptide was summed and shown on the y-axis. The colour code blue – green – yellow – red indicates low to high abundance. Sequences without identifications are on the left in white. Data are average results from three independent digestion experiments. The data are from Lab 8. (B) Peptigrams showing the intensities of peptides at the end of the gastric phase (120 min), obtained by different instruments and settings.

to the individual amino acids (Fig. 4B). The colour intensity shows the summed intensity of the corresponding amino acid from each identified peptide, while the height of the bar represents the number of different peptides found in that region.

3.6. Cleavage pattern

In order to describe the most representative N- and C-terminal sequences of the identified peptides, the R package 'ggseqlogo' (Wagih, 2017) was used. This representation provides information about the

cleavage pattern and the enzyme specificities as well as the probability of finding a particular amino acid at the N-terminal (1, 2, 3) or C-terminal (−3, −2, −1) end of the sequence, taking into account the whole list of peptides. A sequence logo with all peptides obtained from β-casein throughout the gastric phase was built (Fig. 5: proline, valine and leucine represented the most likely amino acids at the N- and C-terminal positions of peptides identified in the majority of laboratories). In general, hydrophobic amino acids (in black) were present to a higher extent at the three C-terminal positions of the peptides. Thus, proline and leucine on the 1st or 3rd C-terminal end appeared in 7 out of 9 laboratories; these residues were placed in other prominent positions in two cases. Therefore, this representation revealed that all laboratories found a similar protein cleavage pattern. The highly probable presence of proline at the first position reflects cleavages of X-pro peptide bonds, which are exceptional for trypsin or chymotrypsin (Keil, 1992). These can be the result of the action by exopeptidases present in the pancreatin preparation that are not able to progress further when proline is present.

3.7. Digestion-resistant peptides

Fig. 6 shows the Venn diagrams illustrating the relationship (in number (A) and in percentage (B)) based on the common digestion-resistant β-casein sequences identified by five laboratories at the end of the gastric phase. Even though only 44 were common peptides between the five laboratories, 84 were shared between four laboratories, and 111 were shared by at least three laboratories. As opposed to the graphical representation of the peptide pattern with amino acid counting heatmaps or Peptigram, this analysis was very restrictive, as seen in the low number of common sequences found for β-casein peptides at the end of the gastric phase, even when the total number of peptides was high. The analytical workflow in terms of peptide matching tolerance followed by each laboratory could explain these differences.

3.8. Additional statistical treatments

The similarities between samples over time were also represented in two different ways. First, a heatmap representation of the hierarchical clustering analysis of all peptides identified during the gastric phase of chicken breast digestion was employed, resulting in three different peptide clusters (Supplemental Fig. 2A). In more detail, a dendrogram with the evolution over time of GAPDH (chicken breast) peptide generation showed that samples after 0.1, 5 and 15 min of gastric digestion were clustering and the later time points of 60 and 120 min formed a separate cluster (Supplemental Fig. 2B).

Second, a PCA analysis, including the peptides of all major proteins (α_{s1}-casein, α_{s2}-casein, β-casein, κ-casein and β-lactoglobulin) identified in SMP at all different time points showed the clustering over time of the different samples. In a PCA, samples are not assigned to specific groups and therefore their final arrangement in the 2-dimensional space is only determined by the characteristics of the individual sample. Similarity of samples in the PCA is reflected by their proximity. The gastric samples (G 0.1, 5, 15, 60, 120) were all different over time and were arranged around a circle; in contrast, all intestinal samples (I 0.1, 5, 15, 60, 120) were very similar and were therefore clustering together. The PC1 was responsible for >80% of the features and the PC2 for 12% (Supplemental Fig. 3). A three-dimensional PCA was calculated using MALDI-TOF results from gastric and intestinal digests of chicken breast (Lab 3). This representation is a potent tool to visualize the separation between digestion phases, gastric and intestinal (Supplemental Fig. 5).

4. Discussion

The standardised static *in vitro* protocol developed in the COST INFOGEST network allows for the simulation of gastrointestinal digestion with standard laboratory equipment. The physiological relevance of the protocol was demonstrated by comparing the peptidomic results of *in vitro* digests with porcine and human data (Egger et al., 2018; Sanchón, 2018). The use of these harmonised conditions improved the comparability of the digestion results versus the different in-house digestion protocols applied previously (Egger et al., 2016). Protein breakdown during digestion can be monitored at different levels of complexity, ranging from assessment of Protein hydrolysis (e.g. by SDS-PAGE) to qualitative peptide patterns (amino acid counting), and in-depth quantification of individual peptides, the latter requiring high-resolution chromatography using high accuracy mass spectrometry techniques. Peptidomic analysis, based on bioinformatics retreatment of the MS and MS/MS data of peptides separated by LC, has become a widely used tool to characterise food digests. The importance of this

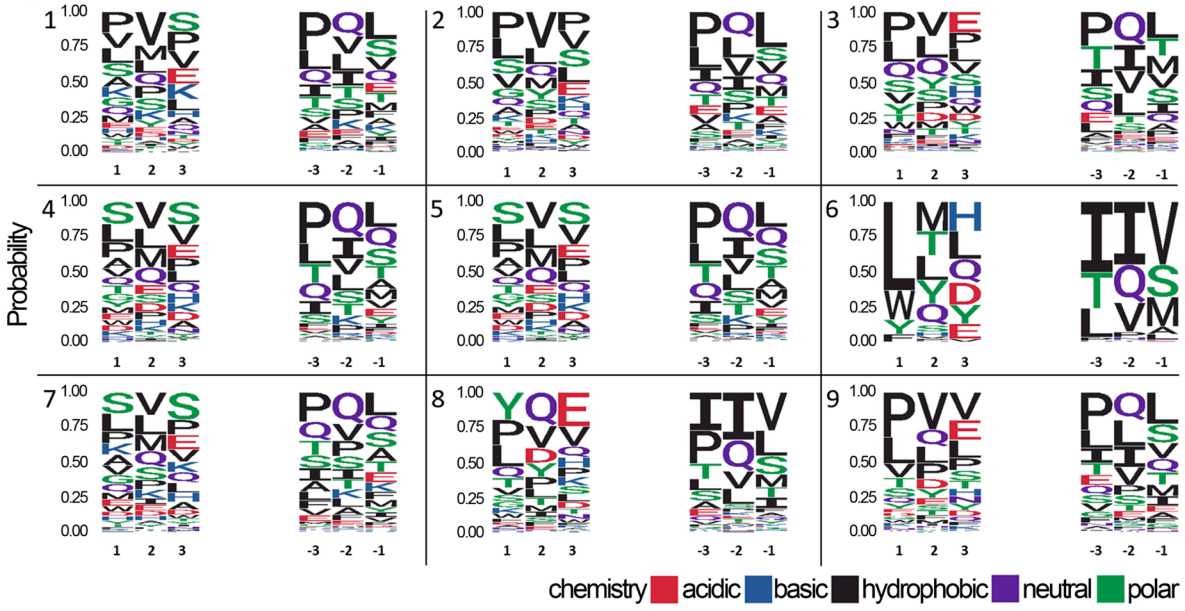

Fig. 5. The most representative N- and C-terminal sequences of the SMP β-casein (UniProtKB P02666)-derived peptides (identified in the participating Labs 1–9) released during the gastric phase. SeqLogo. On the x-axis, the amino acid positions at the N-terminal (1, 2, 3) and at the C-terminal (−3, −2, −1) ends of the protein are indicated.

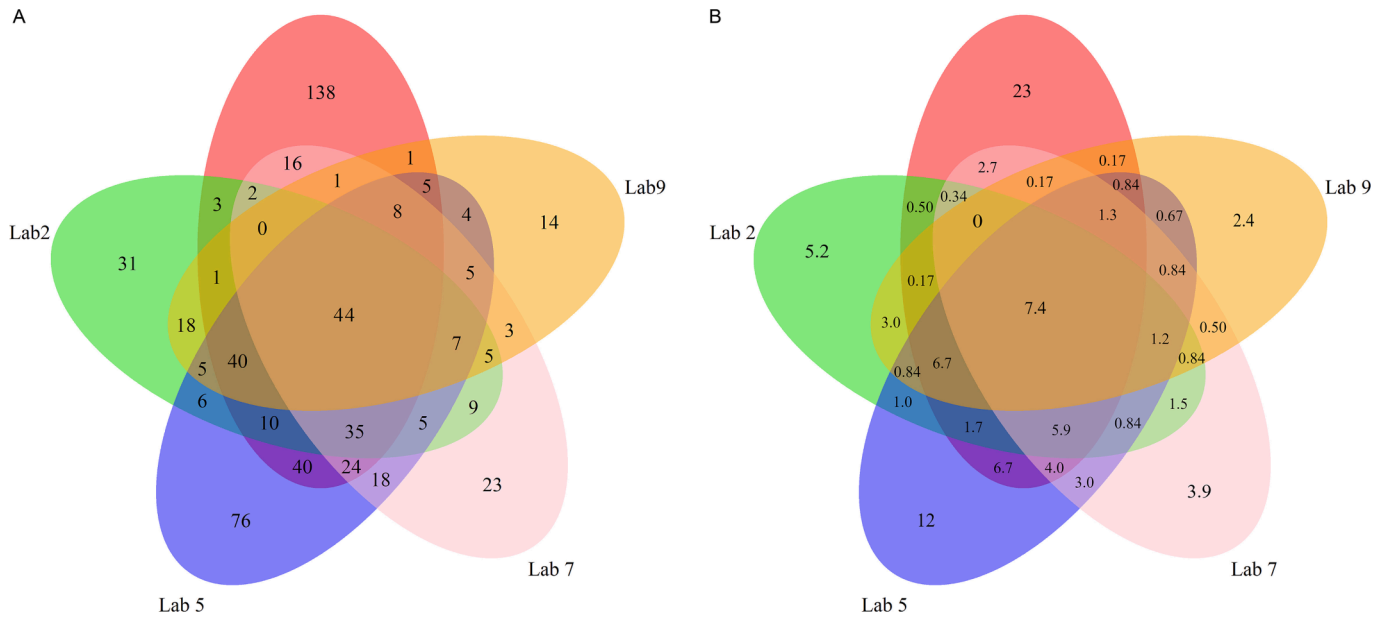

Fig. 6. Venn diagrams showing the relationship, based on the digestion-resistant peptide sequences of β-casein at the end of the gastric phase (120 min), between five laboratories. (A) Number of peptides. (B) Percentage of total identified peptides (%).

approach arises from the fact that all individual peptides in a digest have their own relevance and for this reason the number of peptides reliably identified in a digest should be maximized (Bingeman et al., 2017; Picariello et al., 2013). However, although the digestion conditions are standardised, the analysis of digesta by peptidomic methods has not been harmonised until now and consequently depends on experimental protocols, equipment, settings and data retreatment of each laboratory. Therefore, this study aimed to investigate the comparability of the results obtained using different instruments and analytical workflows by different laboratories. In addition, different post-processing and data analysis tools suitable for peptidomic data are shown, which can be useful for the characterisation of digests or other complex peptidomic data. The following discussion focuses on similarities between laboratories, highlighting biological results on protein digestion by using possible graphical representations, and finally giving a critical evaluation with respect to the technical settings.

4.1. Protein hydrolysis over time in SMP, tofu and chicken breast

The low number of peptide identifications in the intestinal phase (Table 2, Supplementary Fig. 1) could be due to a combination of (i) a lower concentration of peptides, (ii) the presence of fluid components that hinder peptide ionisation or (iii) the fact that digestion arrives to a point where peptides become shorter than five amino acids in length and are therefore too small to be efficiently separated by reverse-phase LC or identified with conventional tandem MS techniques or bioinformatics data retreatment, which relies on specific sequence matches in protein databases. The number of peptides at the end of the gastric phase (120 min) were lower for SMP compared to chicken and tofu, however this difference was no longer observed at the intestinal phase (Fig. 3). Peptides from tofu at the end of the intestinal phase had a higher molecular weight on average compared to SMP and chicken (Fig. 3). This seems to suggest that proteins from SMP and chicken breast were digested to a higher degree than those from tofu proteins. One reason for this difference could be the presence of trypsin inhibitors or antinutritional factors in tofu (Miyagi et al., 1997). Therefore, our results also point out food matrix as additional factor of complexity for MS analysis with different abundance of peptides and peptides of different size depending on the substrate. In general, by summing the total number of peptides for all proteins in each substrate, a higher number of peptides were identified in the gastric phase compared to the intestinal phase (Fig. 3, Supplemental Fig. 1). Interestingly, chicken breast and tofu yielded a higher total number of peptides in total, compared to SMP. This fact could be due to proline-rich sequences in casein which are digestion-resistant (Keil, 1992).

4.2. Representation of peptide generation during the digestion process

The peptides resulting from digestion can be measured and visualised in different ways. Of course, each peptide could be quantified by MS and presented individually – however, in this study, the goal was to generate an overview of the peptidomics results obtained in the different participating laboratories. It is a general difficulty of peptidomics analyses in digesta to condense the large amount of measured data in a way that will achieve a comprehensible overview of digestion processes over time or, in the case of this collaborative study, per laboratory. With both methods (amino acid counting and Peptigram, Fig. 4), digestion processes could be well visualised both over time and between different laboratories. Furthermore, both methods showed regions in the protein that were digestion-resistant or rather labile. However, with both methods, it remained unclear whether, in the protein regions where no peptides were found, the peptides were hydrolysed to smaller pieces (usually smaller than 5 amino acids) and therefore below the detectable size, or whether the peptides in this region are not yet hydrolysed and are therefore too long to be efficiently ionised and detected, as has been mentioned previously as a ‘hidden digestome’ (De Cicco et al., 2019). For the interpretation of peptide patterns, the time course of peptide appearance and disappearance along the protein sequence is helpful. For instance, regions with high peptide coverage during the gastric phase, together with a low peptide coverage and abundance in the intestinal phase, indicate that the protein was most likely hydrolysed during the gastric phase. In contrast, low coverage in the gastric phase and high coverage in the intestinal phase could indicate that this protein is resistant to pepsin in the gastric phase. A low peptide number over time in both phases could indicate either digestion resistance or the failure to be detected by MS.

Nevertheless, both methods have their advantages and disadvantages. The amino acid counting method has the advantage that the individual intensity (caused by the ionisation properties and abundance of the peptide) has less influence on the result. Indeed, if two peptides with

different intensities, but both selected as intense peptides in the MS, each elute in the chromatography for 10 s, then both will be measured the same number of times, resulting in similar intensity on the heatmap graph. In contrast, with the Peptigram method, the colour intensity reflects the individual peptide intensity, which depends on the MS conditions, the peptide amount (which is in turn related to the capability to be formed and to the digestion resistance), the amino acid sequence of the peptide and its length. Due to a strongly varying individual ionisation intensity in the MS, the more easily ionised peptides are strongly overweighed. On the other hand, the amino acid counting method has the disadvantage that already measured MS signals are measured repeatedly, making it impossible to generate an exclusion list. Consequently, a lower number of different peptides are measured in comparison to the Peptigram method.

As further support for the interpretation of peptide data, PCA analysis, including all peptides generated over time, was performed (Supplemental Fig. 3). With this representation, the evolution in the SMP peptidome in the gastric phase could be reproduced, whereas the samples in the intestinal phase were all clustered together, indicating that they were very similar.

4.3. Specific cleavage sites and common digestion-resistant peptides

The profiles obtained from the sequence logo throughout the gastric phase are consistent with pepsin specificity. The most likely amino acids at the end position seem to be hydrophobic in accordance with the likeliness of pepsin to cleave after hydrophobic amino acids (Keil, 1992). Moreover, proline is only present at the N-terminal or the antepenultimate position in the released peptides, in agreement with the inactivity of pepsin on this amino acid when placed at the second position relative to the cleavage site.

In the analysis of digests, it is sometimes necessary to identify crucial sequences due to their biological activity or allergenicity. Tools developed to build Venn diagrams using peptidomic data enable the comparison of sequences at the amino acid level. This, in turn, allows for the identification of sequences appearing solely under a particular condition. The possibility to display the list of the common or discriminating peptides simplifies the comparison of the resulting peptides from different conditions.

5. Conclusion

As expected, the results are affected by a high variability in the peptide patterns between laboratories reflecting the various available equipment, different workflows including analytical conditions and sample preparation, as well as peptide acceptance criteria. A better comparability of mass spectrometry results could be achieved by establishing common guidelines for the analysis of digesta and digestive processes.

Nevertheless, the performed peptidomic analysis of gastric and intestinal digests from SMP, tofu and chicken breast permitted an in-depth characterisation of their peptide fraction which allowed to draw common biological conclusions.

As expected, peptides identified during the gastric phase encompassed a wider molecular mass range than those identified in the intestinal phase. Due to the technical limitations of conventional peptidomics in identifying peptides below five amino acids, maximal sequence coverages were found at the end of the gastric phase. However, intestinal digestion drastically decreased the protein coverage of all proteins due to extensive hydrolysis, which resulted in the highest coverage of 50% for the milk protein β-casein at the end of the digestion. In general, soy proteins showed a slower gastric digestion than the proteins from SMP or chicken breast. In addition, in the case of tofu, larger peptides than for the other two substrates were found at the intestinal phase, suggesting a lower degree of hydrolysis for soy proteins. However, chicken breast and tofu yielded a higher number of total peptides compared to SMP due to the higher variety of proteins at a similar concentration. Although a variability in peptide patterns between laboratories was observed, comparable cleavage specificities were found at the different laboratories. For instance, proline, valine and leucine represented the most likely amino acids at the N- and C-terminal positions of β-casein-derived peptides. The number of identical peptide sequences found in all laboratories were limited by the different analytical workflows and the matching tolerances used for peptide identification, which were identified as the major sources of variability. Visualisation tools (boxplot, heatmaps, Peptigram) or statistical treatments suitable for peptidomic data (clustering analysis, PCA) have proven to be useful for the characterisation of the digestome of different substrates.

In summary, this multicentre collaborative study of peptidomic analysis generated interesting findings, namely: (i) the main biological results on protein hydrolysis processes during digestion were similar in the three investigated foods, independent of the equipment; (ii) the highly complex data obtained with peptidomic analysis needs to be simplified, reduced and represented graphically in order to allow for a biological interpretation; (iii) it is critical to carefully evaluate whether a result is indeed due to biological differences or rather attributable to the settings of the instrument; (iv) a harmonized workflow for mass spectrometry analysis of digestive processes would be highly beneficial.

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 will be made available on request.

Acknowledgements

We thank the INRAE INFOGEST network for funding the workshop in Lyon and for giving us the opportunity to conduct this joint project, the USDA Hatch project CA-D-FST-2187-H (D.B., Y-P.H.). Project PID2019-107663RB-I00 funded by MCIN/AEI/10.13039/501100011033 from the Spanish Ministry of Science and Innovation (P. J., I.R, L.E. and B.M), Ege University funded the Project No. FHD-2021-22816 (S.E.), the Alibiotech research programme, which is financed by the European Union, the French State and the Hauts-de-France region, the French governmental subsidy managed by the French National Research Agency (ANR) within the frame of the 'Future Investments' programme (ANR-11-EQPX-0037) (LC-MS/MS experiments, performed on the REALCAT platform), the Hauts-de-France region, the FEDER, the Ecole Centrale de Lille and the Centrale Initiatives Foundation are also warmly acknowledged for their financial contributions to the acquisition of REALCAT platform equipment (B.D.), part of this work was generated by accessing the research infrastructure at Aarhus University, including those funded by FOODHAY (Food and Health Open Innovation Laboratory, Danish Roadmap for Research Infrastructure) (T.D., M.D.).

Appendix A. Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.foodres.2023.112887.

References

Bauchart, C., Morzel, M., Chambon, C., Mirand, P. P., Reynès, C., Buffière, C., & Rémond, D. (2007). Peptides reproducibly released by in vivo digestion of beef meat and trout flesh in pigs. *The British Journal of Nutrition, 98*(6), 1187–1195.

Bingeman, T. S., Perlman, D. H., Storey, D. G., & Lewis, I. A. (2017). Digestomics: An emerging strategy for comprehensive analysis of protein catabolism. *Current Opinion in Biotechnology, 43*, 134–140.

Boron, W. F., & L., B. E. (2009). *Medical physiology: A cellular and molecular approach*. Philadelphia, PA: Saunders/Elsevier.

Boutrou, R., Gaudichon, C., Dupont, D., Jardin, J., Airinei, G., Marsset-Baglieri, A., … Leonil, J. (2013). Sequential release of milk protein-derived bioactive peptides in the jejunum in healthy humans. *The American Journal of Clinical Nutrition, 97*(6), 1314–1323.

Caron, J., Domenger, D., Dhulster, P., Ravallec, R., & Cudennec, B. (2017). Protein Digestion-Derived Peptides and the Peripheral Regulation of Food Intake. *Front Endocrinol (Lausanne), 8*, 85.

Chen, H., & Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics, 12*, 35.

Dallas, D. C., Guerrero, A., Khaldi, N., Borghese, R., Bhandari, A., Underwood, M. A., … Barile, D. (2014). A peptidomic analysis of human milk digestion in the infant stomach reveals protein-specific degradation patterns. *The Journal of Nutrition, 144*(6), 815–820.

Dalsgaard, T. K., Nielsen, J. H., & Larsen, L. B. (2007). Proteolysis of milk proteins lactosylated in model systems. *Molecular Nutrition & Food Research, 51*(4), 404–414.

De Cicco, M., Mamone, G., Di Stasio, L., Ferranti, P., Addeo, F., & Picariello, G. (2019). Hidden "Digestome": Current Analytical Approaches Provide Incomplete Peptide Inventories of Food Digests. *Journal of Agricultural and Food Chemistry, 67*(27), 7775–7782.

Dupont, D. (2017). Peptidomic as a tool for assessing protein digestion. *Current Opinion in Food Science, 16*, 53–58.

Egger, L., Schlegel, P., Baumann, C., Stoffers, H., Guggisberg, D., Brügger, C., … Portmann, R. (2017). Physiological comparability of the harmonized INFOGEST in vitro digestion method to in vivo pig digestion. *Food Research International, 102*(Supplement C), 567–574.

Harscoat-Schiavo, C., Nioi, C., Ronat-Heit, E., Paris, C., Vanderesse, R., Fournier, F., & Marc, I. (2012). Hydrophilic properties as a new contribution for computer-aided identification of short peptides in complex mixtures. *Analytical and Bioanalytical Chemistry, 403*(7), 1939–1949.

ISO-8968-3. (2004). ISO 8968-3:2004/IDF 20-3:2004 Milk - Determination of nitrogen content - Part 3: Block-digestion method.

Kang, D., Gho, Y., Suh, M., & Kang, C. (2002). Highly Sensitive and Fast Protein Detection with Coomassie Brilliant Blue in Sodium Dodecyl Sulfate - Polyacrylamide Gel Electrophoresis. *Bulletin of the Korean Chemical Society, 2*.

Keil. (1992). *Specificity of Proteolysis*. Springer Berlin Heidelberg.

Le Maux, S., Nongonierma, A. B., & FitzGerald, R. J. (2015). Improved short peptide identification using HILIC-MS/MS: Retention time prediction model based on the impact of amino acid position in the peptide sequence. *Food Chemistry, 173*, 847–854.

Manguy, J., Jehl, P., Dillon, E. T., Davey, N. E., Shields, D. C., & Holton, T. A. (2017). Peptigram: A Web-Based Application for Peptidomics Data Visualization. *Journal of Proteome Research, 16*(2), 712–719.

Martínez-Augustin, O., Rivero-Gutiérrez, B., Mascaraque, C., & Sánchez de Medina, F. (2014). Food derived bioactive peptides and intestinal barrier function. *International Journal of Molecular Sciences, 15*(12), 22857–22873.

Minekus, M., Alminger, M., Alvito, P., Ballance, S., Bohn, T., Bourlieu, C., … Brodkorb, A. (2014). A standardised static in vitro digestion method suitable for food - an international consensus. *Food & Function, 5*(6), 1113–1124.

Miyagi, Y., Shinjo, S., Nishida, R., Miyagi, C., Takamatsu, K., Yamamoto, T., & Yamamoto, S. (1997). Trypsin inhibitor activity in commercial soybean products in Japan. *Journal of Nutritional Science and Vitaminology (Tokyo), 43*(5), 575–580.

Panchaud, A., Affolter, M., & Kussmann, M. (2012). Mass spectrometry for nutritional peptidomics: How to analyze food bioactives and their health effects. *Journal of Proteomics, 75*(12), 3546–3559.

Perkins, D. N., Pappin, D. J., Creasy, D. M., & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis, 20*(18), 3551–3567.

Pica, V., Stuknytė, M., Masotti, F., De Noni, I., & Cattaneo, S. (2021). Bovine milk fortifiers and fortified pasteurized donor human milk for premature infant nutrition. Peptidomic overview. *LWT, 135*, Article 110037.

Picariello, Mamone, G., Addeo, F., Nitride, C., & Ferranti, P. (2013). Foodomics: Advanced Mass Spectrometry in Modern Food Science and Nutrition. In I. John Wiley & Sons (Ed.), *Proteomic-based Techniques for the Characterization of Food Allergens.* (pp. 69–99).

Piovesana, S., Capriotti, A. L., Cerrato, A., Crescenzi, C., La Barbera, G., Laganà, A., … Cavaliere, C. (2019). Graphitized Carbon Black Enrichment and UHPLC-MS/MS Allow to Meet the Challenge of Small Chain Peptidomics in Urine. *Analytical Chemistry, 91*(17), 11474–11481.

Sanchón, J., et al. (2018). Protein degradation and peptide release from milk proteins in human jejunum. Comparison with in vitro gastrointestinal simulation. *Food Chemistry, 239*, 486–494.

Santos-Hernández, M., Miralles, B., Amigo, L., & Recio, I. (2018). Intestinal Signaling of Proteins and Digestion-Derived Products Relevant to Satiety. *Journal of Agricultural and Food Chemistry, 66*(39), 10123–10131.

Sousa, R., Portmann, R., Dubois, S., Recio, I., & Egger, L. (2020). Protein digestion of different protein sources using the INFOGEST static digestion model. *Food Research International, 130*, Article 108996.

Wagih, O. (2017). Ggseqlogo: A versatile R package for drawing sequence logos. *Bioinformatics, 33*(22), 3645–3647.

Wickham, H. (2016). *ggplot2, Elegant Graphics for Data Analysis*. Springer Link.

Zhao, D., Li, L., Le, T. T., Larsen, L. B., Xu, D., Jiao, W., … Zhang, X. (2019). Digestibility of glycated milk proteins and the peptidomics of their in vitro digests. *Journal of the Science of Food and Agriculture, 99*(6), 3069–3077.