EXTERNAL SCIENTIFIC REPORT

UNIVERSITY OF LEEDS

APPROVED: 11 December 2019

doi:10.2903/sp.efsa.2019.EN-1765

Report on EFSA project OC/EFSA/GMO/2017/01

"*In vitro* protein digestibility" (Allergestion)

Alan Mackie[1], Didier Dupont[2], Amelia Torcello-Gómez[1], Julien Jardin[2], Amélie Deglaire[2]

1. School of Food Science and Nutrition, University of Leeds, Leeds, LS2 9JT, UK

2. INRA Agrocampus Ouest, STLO, 65 Rue St Brieuc, Rennes, France

Abstract

The primary objective of the assessment of novel proteins is to evaluate whether they are safe to consume, including potential allergenicity. As part of a suite of assessments, the in vitro digestion of protein has been seen as a useful exercise. Thus, in line with the guidance offered by the EFSA GMO Panel we are using an early phase and a late phase gastric simulation as well as a simulation of the infant gastric compartment, all followed by intestinal phases. These digestion scenarios were used with a panel of 10 proteins from plant and animal origin that were proteins with distinct allergenic potential. The results from the SDS-PAGE and densitometry show significant and mainly expected differences between the different digestion scenarios. The milk proteins were fully digested in the intestinal phase but the BLG was largely resistant to pepsin. In contrast, the egg proteins showed significant persistence except under late phase conditions. For the plant proteins, KTI and ConA were largely resistant to all conditions whereas LIP and AP were only resistant to infant conditions. Similarly, Ara h 1 showed some resistance to infant gastric conditions. The LC-MS analysis of peptides was able to highlight a number of clusters where differences were seen between the digestion scenarios and these could in some cases be mapped onto the primary sequence and where relevant compared with known allergenic epitopes. Under the different digestion scenarios, we were able to show significant differences in the persistence of peptides larger than 9 amino acids and significant overlap of abundant peptides from early phase intestinal digestion and known epitopes for a number of proteins. Although, linking these differences to immunological responses (epitope mapping) still seems to be quite challenging, there are clear differences between scenarios and strong potential for improved risk assessment.

Key words: protein, *in vitro* digestion, allergy, early phase, late phase, infant

Question number: EFSA-Q-2017-00700

Correspondence: gmo@efsa.europa.eu

Acknowledgements: the authors would like to thank EFSA for very helpful discussion.

Suggested citation: Mackie A, Dupont D, Torcello-Gómez A, Jardin J, Deglaire A, 2019. Report on EFSA project OC/EFSA/GMO/2017/01."*In vitro* protein digestibility" (Allergestion). EFSA supporting publication 2019:EN-1765. 82 pp. doi:10.2903/sp.efsa.2019.EN-1765

ISSN: 2397-8325

Summary

The primary objective of the assessment of novel proteins is to evaluate whether they are safe to consume, including potential allergenicity. As part of a suite of assessments, the *in vitro* digestion of protein has been seen as a useful exercise. However, the current test uses conditions that are not representative of the normal physiological environment that food is exposed to immediately post consumption. Thus, we used an early phase and a late phase gastric simulation as well as a simulation of the infant gastric compartment with the primary aim of addressing the following questions:

1. Do the new protocols offer differences or advantages from the current pepsin resistance test with respect to persistent fragments larger than 9 amino acids?
2. Are these differences valuable for risk assessment taking into account Annex B and Figure B.2[1] (Persistent peptides larger than 9 amino acids).

The three digestion scenarios were tested on ten different proteins that represented plant and animal proteins and were proteins with distinct allergenic potential. The readout was given by SDS-PAGE and LC-MS. In the work presented here, we are making a comparison with the late phase gastric simulation (pepsin resistance test) and thus we consider peptides that are still present at the end of the gastric phase of any of the scenarios persistent. Any peptides that are formed during gastric digestion but not evident at the end of the gastric phase would thus be transient. If we were to step outside of this comparison, then a better definition could be argued for persistence as being the presence of a peptide at some defined time through intestinal digestion.

The results from the SDS-PAGE clearly showed differences in the digestion patterns between the different digestion scenarios. This was in terms of differences in the resistance of the proteins to hydrolysis and the changes in abundance over time. In other words, there were differences seen in digestion kinetics for the intact protein and any large peptides (>4kDa) formed. Thus, the answer to the first part of question 1 is yes, there are differences. Although it is not clear from the data whether this is advantageous for allergy risk assessment as set out in question 2, the fact that both BCAS and Ara h 1 showed extended persistence through the infant gastric phase suggests that the approach may have some advantages over the conventional pepsin resistance test. Despite this, there was no general correlation between the rate of digestion of a protein as shown by SDS-PAGE and its propensity to be allergenic. This is perhaps not very surprising because SDS-PAGE is a coarse measure of protein hydrolysis. A much more sensitive readout is the detection of peptides by LC-MS.

All of the samples of digesta measured by SDS-PAGE were also characterised by LC-MS and the date displayed in a number of different formats. Firstly, a simple box plot showing the evolution of the molecular weight and number of the peptides through the gastric and intestinal phases of the different scenarios. Such data can be rather counter intuitive for example with BCAS showing larger peptides than BLG in the gastric phases. In addition to the number of peptides, a clustering analysis was also performed, and the graphs represent the median value of peptide abundance for the cluster. The clustering showed differences between the different digestion scenarios and the largest clusters were also mapped onto the primary sequences for a number of example proteins. The results show some correlations with known IgE epitopes, but this is by no means clear-cut. Although we have mapped the largest of these clusters onto the primary sequences of selected proteins, it maybe that we have not chosen the most relevant peptide clusters from an allergenic perspective. We have also mapped the gastric and for some proteins, the intestinal peptides, onto the primary sequence for comparison. We have shown that there are some correlations between regions of high peptide abundance and the location of known epitopes but in all cases the correlation was limited. We cannot rule out that this may be more of an indication of the confidence to be placed in the epitopes.

We believe that the approach used here is extremely powerful but more targeted research will be needed to link these results to immunological outcomes. For example, the use of artificial intelligence (AI) to provide statistical analysis of all persistent peptides and using a broad range of known allergens and their epitopes as training sets. Thus, developing rules linking LC-MS data analysis of digesta to the

probability of allergenicity. Indeed, this process could start based on the exiting data that we have generated within this project.

Table of contents

1. Introduction

1.1. Background and Terms of Reference as provided by the requestor

This contract was awarded by EFSA to: Prof Alan Mackie, School of Food Science and Nutrition, University of Leeds, UK

Contractor: University of Leeds, UK

Contract title: *In vitro* protein digestibility (Allergestion)

Contract number: OC/EFSA/GMO/2017/01

1.2. Interpretation of the Terms of Reference

The proposal describes the following tasks that have all been completed. Any deviations are described in Section 1.3

Task 1: Define an *in vitro* digestion method

We have used the following digestion simulations:

a. Infant protocol: The infant protocol is based on work recently published by Menard et al. [2] and represents the mildest digestive conditions tested. The gastric pH used is 5.3 and the pepsin activity is 268 U/mL, which represents 85 U/mg of protein substrate. The gastric phase runs for 60 minutes and is followed by an intestinal phase at pH 6.6 that includes 16 U/mL trypsin, 4 U/mL chymotrypsin and bile (1.55 mM glycocholate + 1.55 mM glycochenodeoxycholate) for 60 minutes.

b. Early or mid-phase adult gastric digestion: Infogest harmonised protocol from Minekus et al. [3] is widely used and has been validated against in vivo data. The gastric phase from the protocol uses a pH of 3.0 and a pepsin activity of 2000 U/mL, which corresponds to 0.8 U/µg of protein substrate. After 60 minutes, the gastric phase is followed by the recommended intestinal phase of digestion but samples will be taken throughout the gastric and intestinal phases as outlined below. The intestinal phase works at pH 7 and includes trypsin (100 U/mL), chymotrypsin (25 U/mL) and bile (5 mM glycocholate + 5 mM glycochenodeoxycholate). The gastric phase runs for 60 mins and the intestinal phase for 60 minutes.

c. Late phase adult gastric digestion: This is essentially the well-known pepsin resistance test from Thomas et al. 2004 [4]. The method uses pH 1.2 and a pepsin activity of 10 U/µg of protein, which in this case will be 2500 U/ml based on the starting concentration of protein (5 mg/mL) and a 5:95 dilution for gastric secretion. This low pH is certainly representative of the late gastric phase or the fasted state but the pepsin activity is high for any phase of the digestion. For that reason, gastric chyme will be transferred to the intestinal phase from scenario b after both 10 and 60 minutes. Samples will be collected over 60 minutes as indicated below.

The samples are collected during digestion at the following time points in each phase (gastric and intestinal): 0.5, 2, 5, 10, 20, 30, and 60 min. The partially digested samples will be analysed by SDS-PAGE (UoL) and LC-MS (INRA) in order to identify and quantify:

a) Intact protein

b) Fragments ≥ 9 amino acids

c) Fragments < 9 amino acids.

We will define the half-life of intact protein and half-life of all fragments greater than 9 amino acids.

Analysis of the persistence of proteins and peptides larger than 5-10 kDa can be done by SDS-PAGE with fluorescent label/stain. Below 5-10 kDa LC-MS will be needed to measure possible appearance and disappearance of the fragments.

Task 2: Experimental evaluation

The simulations described in Task 1 have been tested using 5 mg/mL bovine β-lactoglobulin and β-casein. Digested and un-digested proteins have been analysed using SDS-Page and LC-MS as outlined in Task 1.

Task 3: Validation of the model

A larger panel of up to 10 proteins selected in consultation with EFSA have been put through the same protocols replacing the use of pancreatin with individual enzymes trypsin and chymotrypsin. The panel includes the two milk proteins already tested in Task 2 and eight further proteins that vary in stability from stable, intermediate to labile, and in allergenicity (allergen versus non-allergen). These additional proteins are hen's egg lysozyme, hen's egg ovalbumin, soybean Kunitz-type trypsin inhibitor B, soybean lipoxygenase, concanavalin A from Canavalia ensiformis (Jack bean), potato acid phosphatase, gliadin from wheat and Ara h 1 from peanut. Digested and un-digested proteins have been analysed using SDS-Page and LC-MS as outlined in Task 1. An extensive characterization of peptides will be performed to provide lists of hundreds of peptides created during hydrolysis, but we will focus on the more abundant ones to estimate the half-life. Proteins and peptides equal to or larger than 9 amino acids will be quantified, and the top 10 most intense fragments will be identified. Advanced statistical methods will be used to analyse the data in order to draw conclusions on the influence of pH and enzyme concentrations on kinetics and products formed.

Task 4: Defining risk assessment considerations and reporting

Statistical analysis has been performed on the LC-MS and SDS-page data from Task 3 using the advanced statistical analysis method developed under Task 2. This means that the data will be stored, the profiles will be created, classified, and fully prepared for the next step in the analysis. The statistical analysis will be used to assess the kinetics of hydrolysis, which will enable discussion of transient versus persistent fragments and enable a quantitative definition of the transience involved. The report will include a definition of the different classes generated, i.e., when does a protein belong to which class. As part of the report, a comparison will be made of the readouts (extent and patterns of fragmentation) between the three environmental conditions developed in Task 1 and relating to the proteins tested in tasks 2 and 3.

1.3. Additional information

The primary objective of the assessment of novel proteins is to evaluate whether they are safe to consume. This includes their potential to elicit an allergic reaction. As part of a suite of assessments, the in vitro digestion of protein has been seen as a useful exercise. However, the current test uses conditions that are not representative of the normal physiological environment that food is exposed to immediately post consumption. Thus, in line with the guidance offered by the EFSA GMO Panel 1 this project includes a range of more physiologically relevant conditions that represent the adult and infant gastrointestinal (GI) environments. In particular, we are using an early phase (fed state) and a late phase (fasted state) gastric simulation as well as a simulation of the infant gastric compartment based on Menard et al. 2. The late phase simulation will essentially be the same as the current pepsin resistance test 5 (EFSA Guidance Document (2011) and Implementing Regulation (EU) No 503/2013 (IR503/2013)). The early phase adult simulation will be based on the Infogest harmonised protocol 3. All simulations include an intestinal phase and in the case of the late phase simulation, samples were transferred from the gastric to the intestinal phase at 10 minutes and 60 minutes. Only results for Lys, Ova and ConA are included since the SDS-PAGE profile of the rest of the proteins does not change between 10 and 60 min of the gastric phase. The LC-MS data will include the results for all the proteins. The readout includes both SDS-PAGE and LC-MS and the amount of data has meant that we have had to clarify what was originally written tendered and is described in Section 1.2. Specifically, peptides less than 9 amino acids have not been identified as they fall outside the tender requirements. Additionally, we have shown the complete time-course for the loss of intact proteins rather than just assigning a half-life, as we believe this to be more informative. We have also not provided a half-life for the large numbers of peptides produced through digestion of the different proteins as this did not seem to be a

useful exercise with respect to allergen risk assessment. Our approach has been to emphasise the different outcomes from the digestion scenarios and to cluster peptides that are indicative of most of the differences and then associate those differences with the primary sequences and known epitopes (where appropriate). The mass spectrometry data can be made available for further analysis by us and by others (subject to agreement by EFSA) and indeed, we plan to explore the data in more detail in the future. Finally, we would like to emphasise that the LC-MS data does not provide quantitative data on any given peptide but does provide quantitative data on how the abundance of a particular peptides changes through digestion.

2. Materials and Methods

2.1 Proteins

Proteins β-lactoglobulin (BLG, Cat. No. L3908) and β-casein (BCAS, Cat. No. C6905) (≥98% purity both) from bovine milk, soybean Kunitz-type trypsin inhibitor B (KTI, Cat. No. T9003, ≥90% purity by SDS-PAGE), soybean lipoxygenase (Lip, Cat. No. L7395, ≥70% purity by SDS-PAGE), concanavalin A from Jack bean (ConA, Cat. No. C2010, 100% purity), potato acid phosphatase (AP, Cat. No. 10108227001, 70-90%), and gliadin from wheat (GLI, Cat. No. G3375, 87% purity) were purchased from Sigma-Aldrich and used as received. Hen's egg lysozyme (Lys, 100% purity by SDS-PAGE) and ovalbumin (Ova, ≥85% purity by SDS-PAGE, ovotransferrin being the main contaminant) were supplied by INRA AgroCampus Ouest. Ara h 1 from raw red skin peanuts purchased in a local supermarket was purified (≥95% purity by SDS-PAGE) at the University of Leeds according to a previously published procedure [6-7], using a single step lectin affinity column (of ConA Sepharose). All of the proteins were dissolved at a concentration of 5 mg/mL in ultrapure water (Milli-Q), in order to be consistent with the concentration used for BLG and BCAS, based on the concentration of BLG in bovine milk, and on the initial test protein concentration used in the original protocol of pepsin resistance test [8]. No checks were made with regard to glycation or other post translational modifications.

2.2 *In vitro* digestion protocols

All chemicals used were of analytical grade and purchased from Sigma-Aldrich. Bile salts (≥97%) sodium glycocholate (NaGC, Cat. No. G7132) and sodium glycochenodeoxycholate (NaGCDC, Cat. No. G0759) were used. The enzymes pepsin (Cat. No. P7012), trypsin (Cat. No. T0303) and pancreatin (Cat. No. P7545, 8 x USP) were from porcine origin, chymotrypsin (Cat. No. C4129) from bovine origin and their activities were determined as described in the electronic supplementary material of Minekus et al. [3]. Pancreatin was only used in Task 2, and the amount added was based on the required trypsin activity in the final mixture. All *in vitro* digestion protocols were performed in 50 mL conical centrifuge tubes mounted horizontally in a shaking incubator at 37 °C and 100 rpm for better mixing. The *in vitro* digestion of each protein was conducted in triplicate for each model. Control experiments for each *in vitro* digestion model were also performed by replacing the initial volume of protein by ultrapure water. The activity of the enzymes in the pancreatin were found to be as follows:

Trypsin: 6.48 ± 0.13 U/mg
Chymotrypsin: 2.77 ± 0.04 U/mg

2.2.1 Infant model

The infant static *in vitro* digestion model comprises a gastric and intestinal phase in sequence and is described in detail in [2]. The only adaptation was the replacement of bovine bile extract by an equimolar mixture of two purified bile salts (NaGC and NaGCDC) which represent the two major forms in human

bile [9]. In this way, the composition and concentration variability across batches will be avoided. The most important parameters are summarised below for each phase.

Gastric phase: 5 mL of protein (5 mg/mL) were mixed with infant model gastric fluid (IMGF) at a ratio protein to IMGF of 63:37 (v/v). The pH was set to 5.3. IMGF comprised NaCl (94 mM) and KCl (13 mM), adjusted to pH 5.3 with 1 M HCl. Pepsin activity was 268 U/mL in the final volume (gastric chyme). After 60 min of gastric digestion, the pH was raised to 7 with 1 M NaOH in order to stop pepsin activity before intestinal digestion.

Intestinal phase: the gastric chyme was mixed with infant model intestinal fluid (IMIF) at a ratio of gastric chyme to IMIF of 62:38 (v/v) and adjusted to pH 6.6 with 1 M HCl. The IMIF comprised NaCl (164 mM), KCl (10 mM) and $NaHCO_3$ (85 mM) adjusted to pH 7. $CaCl_2$ was added separately before starting the intestinal phase at a concentration of 3 mM within the volume of the IMIF. The total concentration of bile salts was 3.1 mM (in equimolar ratio) in the final volume (intestinal content). The trypsin activity was 16 U/mL in the final volume and the chymotrypsin activity was 4 U/mL. This phase lasted for 60 min.

2.2.2 Early phase adult model

The early phase adult static *in vitro* digestion model is the Infogest international consensus protocol [3] with the following adaptations: the oral phase was omitted, the length of gastric and intestinal phases was 60 min each and an equimolar mixture of NaGC and NaGCDC replaced the bile extract, in order to retain consistency with the infant model. The parameters are summarised below.

Gastric phase: 5 mL of protein (5 mg/mL) were mixed with early phase adult gastric fluid (EPAGF) at a ratio protein solution to EPAGF of 50:50 (v/v) and the pH was set to 3.0. EPAGF comprised NaCl (47.2 mM), KCl (6.9 mM), KH_2PO_4 (0.9 mM), $NaHCO_3$ (25 mM), $MgCl_2(H_2O)_6$ (0.1 mM), and $(NH_4)_2CO_3$ (0.5 mM) adjusted to pH 3 with 1 M HCl. $CaCl_2$ was added separately before starting the gastric phase at a concentration of 0.075 mM in the final volume (gastric chyme). Pepsin activity was 2000 U/mL in the final volume. After 60 min of gastric digestion, the gastric chyme was immediately subjected to the intestinal phase.

Intestinal phase: the gastric chyme was mixed with early phase adult intestinal fluid (EPAIF) at a ratio gastric chyme to EPAIF of 50:50 (v/v) and adjusted to pH 7 with 1 M NaOH. The EPAIF comprised NaCl (38.4 mM), KCl (6.8 mM), KH_2PO_4 (0.8 mM), $NaHCO_3$ (85 mM), and $MgCl_2(H_2O)_6$ (0.33 mM), adjusted to pH 7. $CaCl_2$ was added separately before starting the intestinal phase at a concentration of 0.3 mM in the final volume. The total concentration of bile salts was 10 mM (in equimolar ratio) in the final volume. The trypsin activity was 100 U/mL in the final volume and the chymotrypsin activity was 25 U/mL.

2.2.3 Late phase adult model

The late phase adult static *in vitro* digestion model comprised a gastric phase following the pepsin resistance test protocol as described in [8] followed by the intestinal phase of the adult model above. The parameters of the gastric phase are summarised below.

Gastric phase: 0.5 mL of protein (5 mg/mL) were mixed with simulated gastric fluid (SGF) at a ratio protein solution to SGF of 5:95 (v/v). The pH was set to 1.2. SGF comprised NaCl (35 mM), adjusted to pH 1.2 with 1 M HCl. Pepsin activity was set to 10 U/µg of test protein (2500 U/mL in the final volume). After either 10 or 60 min of gastric digestion, the gastric chyme was immediately subjected to the intestinal phase as in previous section.

2.3 Digesta characterisation

Aliquots of 200 µL were collected at 0.5, 2, 5, 10, 20, 30 and 60 min of both gastric and intestinal phase. Protease activity was immediately stopped by adding 5 µL of Pepstatin A (0.73 mM) to gastric samples, or 10 µL of 4-(2-aminoethyl)-benzenesulfonyl fluoride hydrochloride (0.1 M) to intestinal samples. An aliquot of 100 µL of each gastric and intestinal time point sample were transferred to new microfuge tubes and all frozen at -20 °C for further analysis.

2.3.1 Qualitative protein analysis: SDS-PAGE

Protein hydrolysis within the three different *in vitro* digestion models was first analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). This technique allows the identification and semi-quantification of larger peptides (> 5 kDa). Precast Bolt 4-12% Bis-Tris Plus 1 mm x 10 well gels from Invitrogen were used. Samples were diluted with LDS sample buffer 4 x (Invitrogen) and Milli-Q water in reducing conditions with dithiothreitol (0.5 M). Wells were loaded with 1.5 µg of protein, taking into account the protein to simulated gastrointestinal fluid ratio in order to evaluate the sole impact of the proteolysis. Only in the particular case of GLI digesta, wells were loaded with the maximum amount allowed by the considered model of digestion taking into account the corresponding dilution ratio. This is because GLI has poor solubility in water and it is also difficult to assess its solubility at each time point within the digesta. Mark 12 Unstained Standard (Invitrogen) was used as molecular weight marker. Gels were fixed in methanol/water/acetic acid (40/50/10 v/v) for 1 h, then rinsed for 5 min three times with ultrapure water. Staining was subsequently done with Simply Blue SafeStain (Invitrogen) for 1 h. Distaining was carried out overnight with ultrapure water. Images of the gels were obtained with a gel scanner (Bio-Rad). The SDS-PAGE was repeated at least in duplicate for each protein digested *in vitro* with the three digestion models.

Densitometry on bands was performed with the software Image Lab™ 5.1 (Bio-Rad). Data are presented as mean values ± standard deviation. Comparison between *in vitro* digestion models over time was done with two-way ANOVA and post hoc Bonferroni multiple comparison test with a threshold for significance $p \leq 0.05$.

2.3.2 Qualitative protein analysis: LC-MS

The gastric samples supplied by UoL to INRA were at the following concentrations requiring dilution to the required concentration for injection into the spectrometer.

Infant: 3.07 mg/mL, which was diluted by 1:830 allowing injection of 37 ng of BLG or BCN and 50ng for the subsequent 8 proteins.

Early phase: 2.44 mg/mL, which was diluted by 1:660 allowing injection of 37 ng of BLG or BCN and 50ng for the subsequent 8 proteins

Late phase: 0.244 mg/mL, which was diluted by 1:70 allowing injection of 37 ng of BLG or BCN and 50ng for the subsequent 8 proteins

All samples were filtered using a 0.45 µm filter before injection of 10 µL. Pepstatin was added to all gastric phase samples and the concentration added to the late phase samples was 6 times higher. The intestinal phase samples had no pepstatin added. For the infant and early phase models, 120 ng of unfiltered sample was injected and for the late phase model 6 ng of filtered sample was injected. The number of samples generated required 10 proteins x triplicates x intestinal and gastric phases x kinetics = 1590 injections plus another 200 test injections and quality control.

For mass spectrometry analysis, a nano-RSLC Dionex U3000 system fitted to a Q-Exactive mass spectrometer (Thermo Scientific, San Jose, USA) equipped with a nanoelectrospray ion source was used. Samples were concentrated on a µ-precolumn pepMap100 (C18 column, 300 µm i.d. × 5 mm length, 5 µm particle size, 100 Å pore size; Dionex, Amsterdam, The Netherlands) and separated on a PepMap RSLC column (C18 column, 75 µm i.d. x 150 mm length, 3µm particle size, 100 Å pore size; Dionex).

Peptide separation was performed at a flow rate of 0.3 $\mu L.min^{-1}$ using solvents A [2% (v/v) acetonitrile, 0.08% (v/v) formic acid and 0.01% (v/v) TFA in HPLC gradient grade water] and B [95% (v/v) acetonitrile, 0.08% (v/v) formic acid and 0.01% (v/v) TFA in HPLC gradient grade water]. The elution gradient first rose from 5 to 35% solvent B over 40 min, then up to 85% solvent B over 5 min before column re-equilibration. The mass spectra were recorded in positive mode using the m/z range 350-3000. The resolution of the mass analyser for m/z of 200 amu (atomic mass unit) was set in the acquisition method to 70 000 for MS and 17 500 for MS/MS. For each MS scan, the ten most intense ions were selected for MS/MS fragmentation and excluded from fragmentation for 20 s.

Peptides were identified from the MS/MS spectra using the X!TandemPipeline software (http://pappso.inra.fr) against an in-house database composed of the sequence of the proteins to which was added the common Repository of Adventitious Protein (http://thegpm.org/crap). No specific enzymatic cleavage was specified and the possible post-translational modifications searched were serine phosphorylation, methionine oxidation, and deamidation of glutamine or aspartic acid. Peptides identified with an e-value < 0.01 were automatically validated. Giving an evaluated false discovery rate of less than 1% at the peptide level. Only peptides of minimum 6 amino acids long can be identified with this strategy. All identified peptides were subsequently quantified in all samples regardless of whether they were identified in a particular sample thanks to the MassChroQ open source software. A summary of the number of peptides identified is given in Table 1.

	Infant Model	Early phase model	Late phase (G60) model	Total unique peptide sequences
BCAS Gastric phase	224	292	264	472
BCAS Intestinal phase	170	99	40	296
BLG Gastric phase	32	42	53	72
BLG Intestinal phase	136	63	40	138
KTI Gastric phase	15	14	12	20
KTI Intestinal phase	45	35	29	53
Lip Gastric phase	355	459	424	642

Lip Intestinal phase	1678	926	441	1721
ConA Gastric phase	111	131	125	161
ConA Intestinal phase	315	236	159	329
AP Gastric phase	1	2	5	5
AP Intestinal phase	12	5	12	12
GLI Gastric phase	201	549	577	611
GLI Intestinal phase	1941	1931	1217	2074
Lys Gastric phase	4	3	16	16
Lys Intestinal phase	20	17	7	20
OVA Gastric phase	36	88	85	91
OVA Intestinal phase	406	286	163	434
Ara h 1 Gastric phase	100	472	405	485
Ara h 1 Intestinal phase	638	339	261	682

Table 1: Number of unique peptide sequences and quantified for each scenario

2.4 Statistical analysis

Data analyses were performed using the R software, version 3.3.1 (R Core Team, 2014). An in-house program allowed the peptide mapping and their average abundances onto the parent protein. For each digestion and each phase (gastric phase or intestinal phase), peptide abundances were averaged over the three replicates of each digestion type, summed by amino acid for each digestion time over each digestion compartment (gastric or intestinal) and log-10 transformed.

LC-MS data was analysed as follows:

- A boxplot of peptide molecular weight per model x time in gastric and intestinal compartment. Numbers at the bottom of the boxplot is the number of different and unique peptide sequences quantified. Keep in mind that in late phase intestinal samples, low number of peptides can be the results of a lower injection volume compared to INFANT and Early phase models.
- A number of peptide clusters representation of Abundance versus time in the gastric compartment, based on peptide clustering. The number of peptide clusters chosen is arbitrary

and is based on the lowest number of clusters that can explain most of the difference in the peptide abundance Letters on the graphs are the result of a statistical test trying to determine for each time if peptides behave differently in each model ("Wilcoxon test" on paired individuals - i.e. paired peptides). Since peptide abundance data cannot be considered as following a Normal distribution, a non-parametric test was used and to remain consistent, the graphs represent the median value of peptide abundance for this cluster.

3. Results

3.1 SDS-PAGE

3.1.1 Bovine milk proteins

BLG and BCAS represent two major proteins in bovine milk. They are known to be allergens (Bos d 5 and Bos d 8, respectively) with contrasting susceptibility to digestion within the GI tract. The UniProt protein database designation is P02754 and P02666, respectively. Being well-characterised proteins, they seemed a suitable choice to start validating the comparison of the three already defined models of static in vitro digestion. Figures 1 and 2 show the SDS-PAGE of BLG and BCAS digesta, respectively, at each time point of the gastric and intestinal phase for each in vitro digestion model. The lane labelled as "Prot" corresponds to the protein blank (non-digested). The control lanes in the gastric and intestinal phase, labelled as "C", correspond to the in vitro digestion where the protein was replaced by ultrapure water. Therefore, the control shows bands corresponding to digestive enzymes.

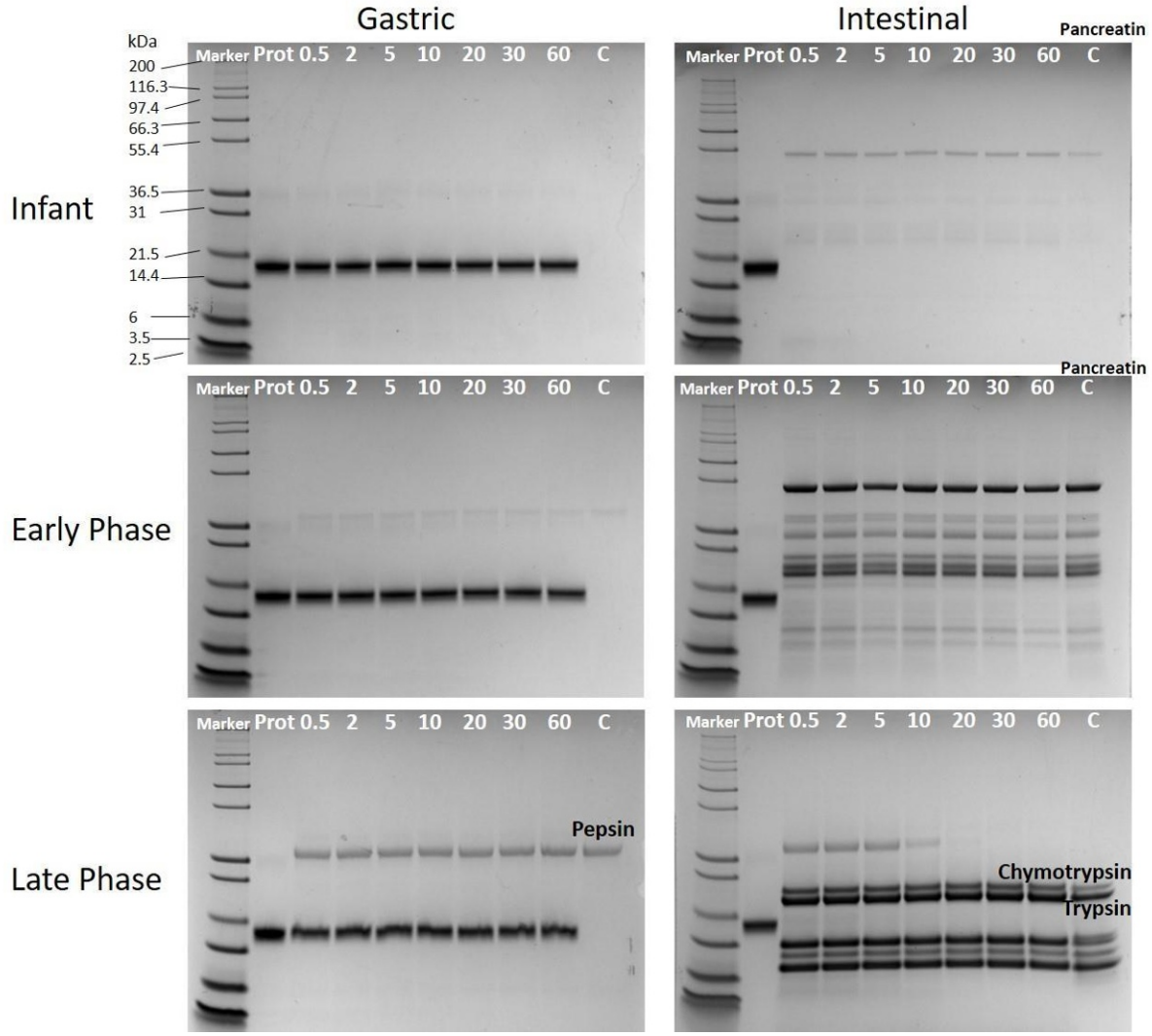


Figure 1: SDS-PAGE of the in vitro digestion of BLG with the infant, early phase adult (INFOGEST) and late phase adult (PRT) models. The numbers represent the time in min after starting each phase (gastric or intestinal). The Prot lane is the protein blank and C lane is the control of digestive enzymes.

Top pictures in Figures 1 and 2 show the proteolysis kinetics for the infant model of digestion. There are clear differences in the gastric protein hydrolysis by pepsin of BLG and BCAS. Namely, BLG is pepsin-resistant throughout the gastric phase, whereas BCAS was partially hydrolysed after 30 s. The band corresponding to the molecular weight of intact BLG (18.4 kDa) is present after 60 min of gastric digestion. Conversely, the band corresponding to the molecular weight of intact BCAS (23.8 kDa) gradually disappears and is no longer visible after 60 min of gastric digestion. In addition, bands corresponding to hydrolysis products (protein fragments of smaller molecular weight: 4-22 kDa approximately) are already visible after 30 s of gastric digestion, become more intense at 5-10 min and finally their intensity decreases at 30 min of gastric digestion although are still visible at the end of the gastric phase. These differences in pepsin hydrolysis for BLG and BCAS are related with the nature of both proteins, globular versus random coil, respectively. In particular, most of the pepsin cleavage sites are buried in the hydrophobic core of BLG [10]. In contrast, significant hydrolysis is observed during in vitro intestinal digestion of BLG [2]. After 30 s of intestinal digestion, a faint band corresponding to a molecular weight of 3 kDa approximately is observed, which gradually vanishes afterwards suggesting almost complete hydrolysis by intestinal proteases. BCAS seems to be completely hydrolysed after 30 s of intestinal digestion. The rest of the bands shown in the intestinal phase correspond to the enzymes present in the pancreatin extract.

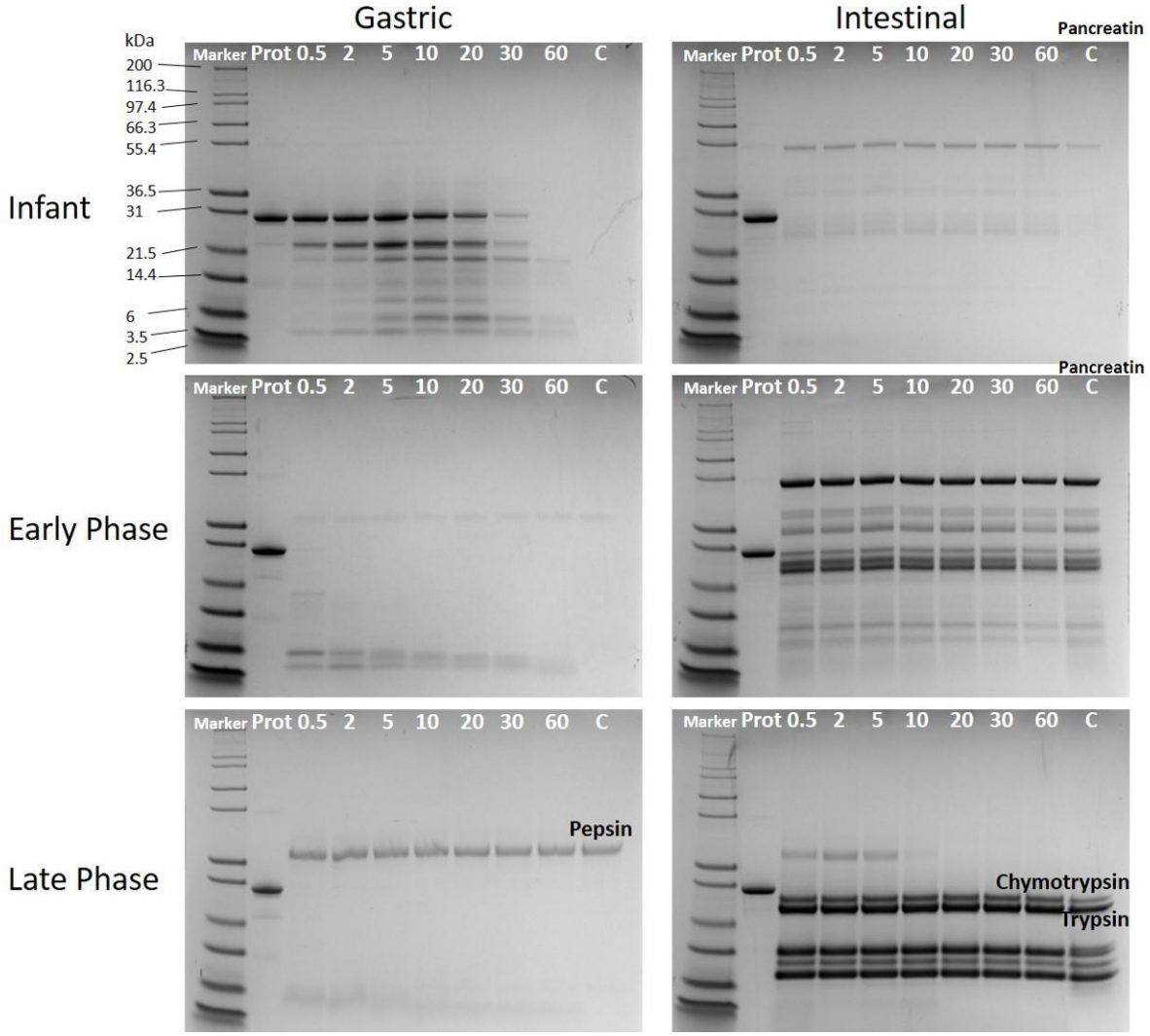


Figure 2: SDS-PAGE of the in vitro digestion of BCAS with the infant, early phase adult (INFOGEST) and late phase adult (PRT) models. The numbers represent the time in min after starting each phase (gastric or intestinal). The Prot lane is the protein blank and C lane is the control of digestive enzymes.

Middle pictures in Figures 1 and 2 display the proteolysis kinetics for the early phase adult model (INFOGEST consensus) of digestion. It is worth noting that in this in vitro digestion model the pepsin to protein ratio is one order of magnitude larger (800 U/mg of test protein) as compared to the infant model (85 U/mg of test protein). This is reflected in the greater intensity of the band corresponding to pepsin (34.6 kDa) in relation to that for infant model. Therefore, faster gastric proteolysis kinetics would be expected. This is the case for BCAS, where only bands corresponding to protein fragments smaller than 20 kDa are visible after 30 s of gastric digestion and peptides smaller than 5 kDa are detected afterwards, which gradually decrease in intensity until the end of the gastric phase. However, BLG shows the band corresponding to the molecular weight of intact protein throughout the gastric phase and no bands corresponding to hydrolysis products can be visually detected. Thus, BLG is also pepsin-resistant under the early phase adult model conditions. The different rate and extent of BCAS proteolysis between infant and early phase adult digestion models is more likely due to the different pH range rather than the content of pepsin[2]. As discussed by Menard and co-workers, the slower kinetics of protein digestion in the infant model is largely due to the loss of pepsin activity at pH 5.3, as compared to the optimal pepsin activity at pH 3 in the early phase adult digestion model. A previous study reported similar BCAS gastric digestibility for an infant and adult model working with similar differences in pepsin content (8-fold lower in the infant) but closer pH values (2.5 and 3 for adult and infant, respectively)[11]. Both proteins seem to be completely digested after 30 s of intestinal digestion, even for BLG, since the intestinal enzymes to protein ratio is also larger in the early phase adult model as compared to the infant model. This is reflected in the more intense bands of the intestinal phase corresponding to pancreatin enzymes. The much faster kinetics and larger extent of hydrolysis shown for infant and adult intestinal digestion of BLG (Figures 1 and 3b) as compared to the results reported by Dupont and co-workers[11], can be partially explained by the use of phospholipids vesicles in their study, which is known to protect BLG against pancreatic proteases degradation[12]. In addition, the trypsin/chymotrypsin to test protein ratio used in the present protocols are higher.

Bottom pictures in Figures 1 and 2 show the proteolysis kinetics for the late phase adult model (pepsin resistant test, PRT). In this model, the pepsin to protein ratio is the greatest of the three protocols of in vitro digestion (10000 U/mg of test protein). Namely, it is one order of magnitude larger than in the early phase adult model, and two orders of magnitude larger than in the infant model. Thus, the intensity of the band corresponding to the molecular weight of pepsin is greater than in previous models. The proteolysis kinetics of BCAS is faster than in the early phase adult digestion model. Indeed, whereas bands between 6 kDa and 2.5 kDa are clearly visible during the first 30 min of gastric digestion with the early phase adult model (Fig. 2, middle left), only a faint band corresponding to protein fragments of molecular weight of 5 kDa approximately is observed after 30 s of gastric digestion (Fig. 2, bottom left). BCAS appears to be completely digested after 60 min of gastric digestion. BLG resists pepsin hydrolysis throughout the gastric phase (Fig. 1, bottom). This agrees with the results of pepsin resistance test for BLG reported by Thomas and co-workers [8]. On the other hand, the bands shown in the intestinal phase seem to correspond to trypsin and chymotrypsin, which from now on replace the pancreatin extract for obvious reasons. The greater intensity of the bands corresponding to the complex mixture of pancreatic enzymes (shown in previous report) makes difficult the interpretation of the results. Their greater intensity as compared to the early phase adult model can be explained due to the higher dilution ratio of the protein in the gastric phase (protein to SGF 5:95 v/v).

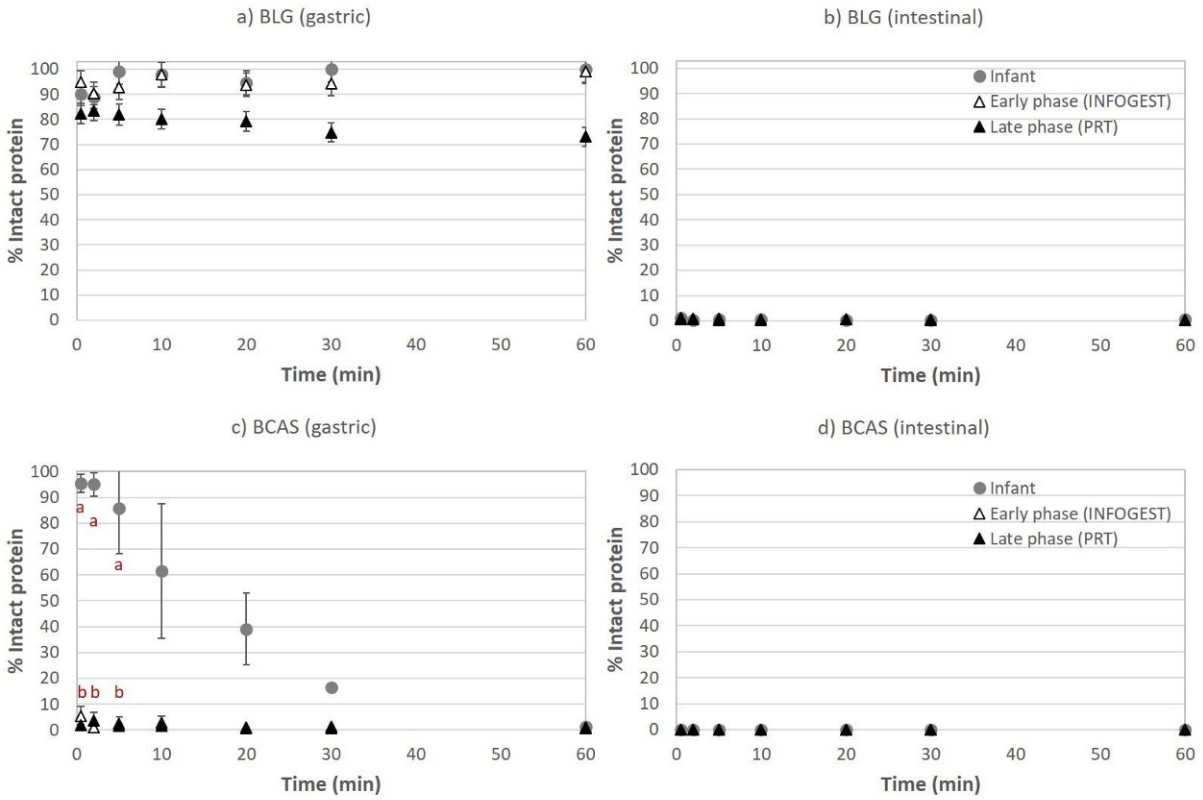


Figure 3: Percentage of intact protein a), b) BLG and c), d) BCAS during in vitro gastric or intestinal digestion (infant, early phase or late phase models) determined from densitometry on SDS-PAGE (n ≥ 2). Different letters mean significant differences ($p \leq 0.05$) between models over time. Absence of letters means no significant differences.

Figure 3 represents the calculated percentage of intact protein during in vitro GI digestion for the three models. Data were obtained from densitometry analysis on SDS-PAGE. On the one hand, Figure 3a,c summarises the results of kinetics of protein gastric hydrolysis discussed above: the resistance of BLG to pepsin hydrolysis regardless of the in vitro model of digestion and the different rate of hydrolysis of BCAS depending on the model. Namely, BCAS was hydrolysed significantly slower during the first 5 min for the infant model. There is no statistically significant difference among the three gastric scenarios after 10 min, nor between the early and late phase adult models throughout the gastric phase. However, we must take into account that these results are related with the intensity of the band corresponding to the molecular weight of intact protein, and as discussed above there were also clear differences in the appearance of BCAS protein fragments with smaller molecular weights corresponding to hydrolysis products. Thus, after 30 s of gastric digestion, bands corresponding to molecular weight up to 20 kDa were identified for the early phase model, whereas only fragments of molecular weight up to 5 kDa were identified for the late phase model. On the other hand, Figure 3b,d summarises the kinetics of intestinal digestion. There is absence of both intact proteins after 30 s of intestinal digestion, regardless of the in vitro model, even for BLG that is persistent throughout the gastric phase of the three scenarios. The rapid digestibility of BLG under intestinal conditions that would be closer to the late phase adult model, although without previous gastric phase, was also reported by Takagi and co-workers [13].

3.1.2 Hen's egg proteins

Lys and Ova are major globular proteins in egg white that are widely consumed and moderately resistant to pepsin digestion. Lys is an enzyme and is designated as an allergen Gal d 4. The UniProt protein database designation is P00698. Ova, the most abundant protein in egg white, is a storage protein, and a known allergen with designation Gal d 2. The UniProt protein database designation is P01012. The in vitro digestibility of both proteins under simulated GI conditions have also been widely studied. Therefore, these proteins are another two appropriate candidates in the validation and comparison of the three static in vitro digestion models considered here.

Figures 4 and 5 present the SDS-PAGE of the digesta of Lys and Ova, respectively, for the three models of digestion. Figure 6 displays the calculated percentage of intact protein from SDS-PAGE densitometry. Figure 4 shows a band corresponding to intact Lys at a molecular weight of around 14.3 kDa that is resistant to pepsin digestion throughout the gastric phase in the infant and early phase adult model of digestion (Figures 4 and 6a). Only at 20 min of the gastric phase in the late phase adult model, is there a statistically significant decrease of the intensity corresponding to intact Lys (Figure 6a) and a faint band of much lower molecular weight of approximately 2-3 kDa appears (Figure 4, bottom left) corresponding to hydrolysis products. At 60 min of the gastric phase in the late phase adult model, the band of the intact protein has completely disappeared. The resistance of Lys to pepsin digestion agrees well with the results reported by Fu et al. under conditions of the late phase adult gastric model or PRT [14]. However, they observed intact protein in SDS-PAGE until 60 min of the gastric phase, which contrasts with our results, since the last time point when intact Lys could be seen was 30 min (Figure 4, bottom left). This could be caused by differences in the pepsin to test protein ratio, because they fixed this number in w/w and we have fixed it in U/mg of test protein. Here the importance of assaying properly the enzymatic activity and setting the enzyme concentration in activity units instead of weight for comparison across laboratories.

In contrast, Ova exhibits slightly different kinetics of gastric hydrolysis. The band corresponding to the intact protein shows at a molecular weight of around 45 kDa. The intensity of this band keeps constant at all times of the infant digestion. In contrast, in the early phase adult model, the hydrolysis starts to be statistically significant after 20 min of the gastric phase as shown by the decrease of band intensity (Figure 6c) and appearance of bands of slightly lower molecular weight corresponding to hydrolysis products (Figure 5, middle left). In the late phase adult model, the hydrolysis occurs at earlier times, 5 min, as seen in Figure 6c, although the time point at which the intact protein could be seen for the last time is min 30. This is in complete agreement with the results of majority of laboratories (6 out of 9) that contributed in the work from Thomas et al. [8]. In addition, the densitometry profile shown in Figure 6c is completely in agreement with that reported by Takagi and co-workers for PRT of Ova [13].

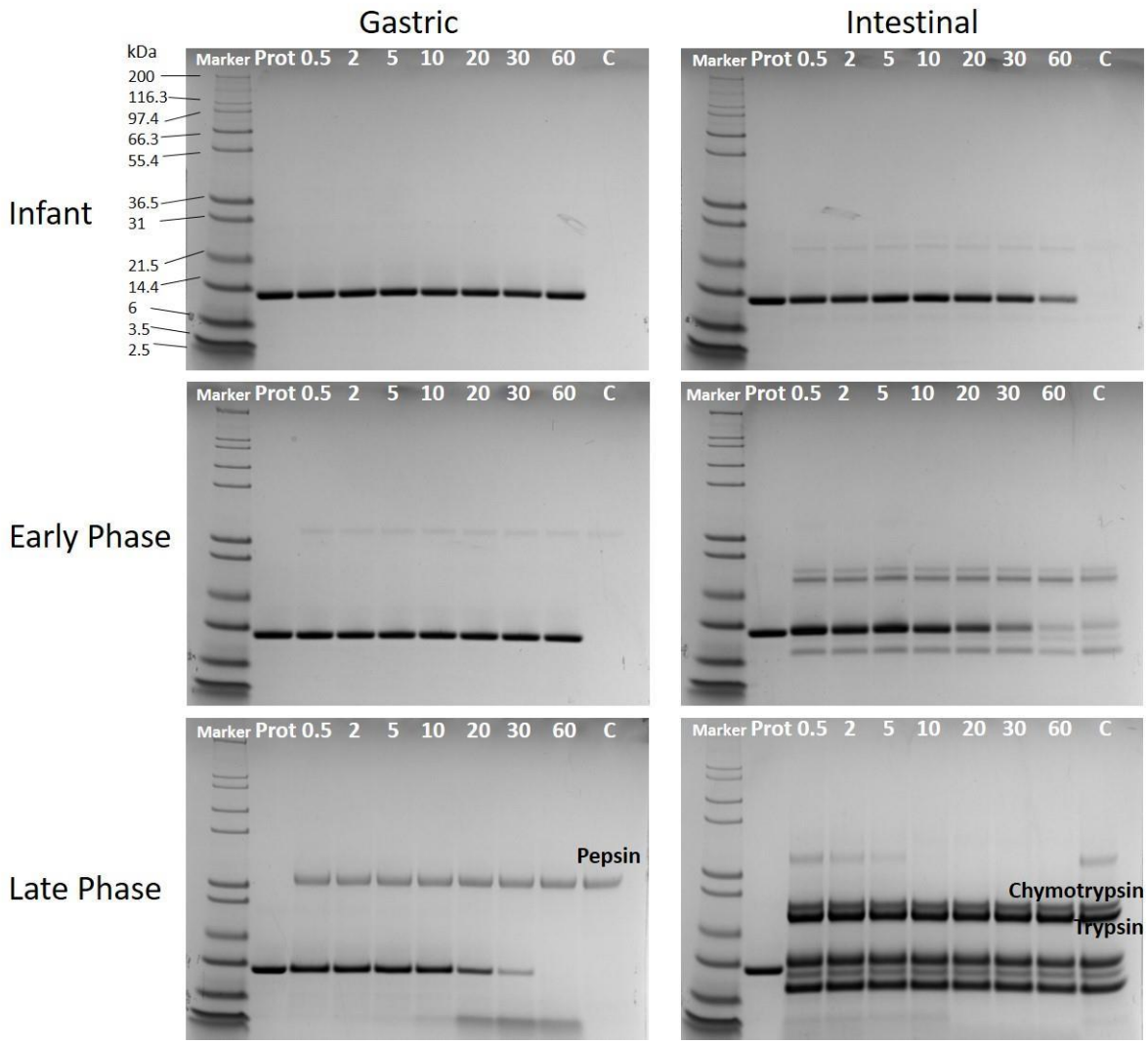


Figure 4: SDS-PAGE of the in vitro digestion of Lys with the infant, early phase adult (INFOGEST) and late phase adult (PRT) models. The numbers represent the time in min after starting each phase (gastric or intestinal). Prot lane is the protein blank and C lane is the control of digestive enzymes.

Regarding the intestinal phase of digestion, we can also observe different kinetics of proteolysis across models and between proteins. In general, the proteolysis was slower and reached lower extent for the infant model, followed by the early phase adult and late phase adult model (Figure 6b,d), although in the latter both intact proteins were already hydrolysed by the end of the gastric phase. Interestingly, eventually the extent of digestibility is similar for both proteins after each particular in vitro model (Figure 6b,d). Namely, around 40-50% of intact protein remains at the end of the intestinal phase in the infant model, around 10-20% in the early phase adult model and 0% in the late phase adult model. The differences in the digestibility profiles can be explained not only by the different trypsin/chymotrypsin ratio to test protein, but also by the different concentrations of bile salts in the three models, which has been demonstrated to affect the rate of proteolysis for certain dietary proteins [15]. The greater the bile salt concentration, the faster the proteolysis. The comparison of Ova digestibility with the infant and early phase adult models is in agreement with the results reported by Dupont and co-workers [11], although the accelerated Ova gastric digestion found in that work for the infant model as compared to our results is likely due to the lower gastric pH that allows optimum pepsin activity. This

also demonstrates the importance of setting a relevant pH because it affects the enzymatic activity. The extent of Ova GI hydrolysis in the early phase adult model agrees very well with their results.

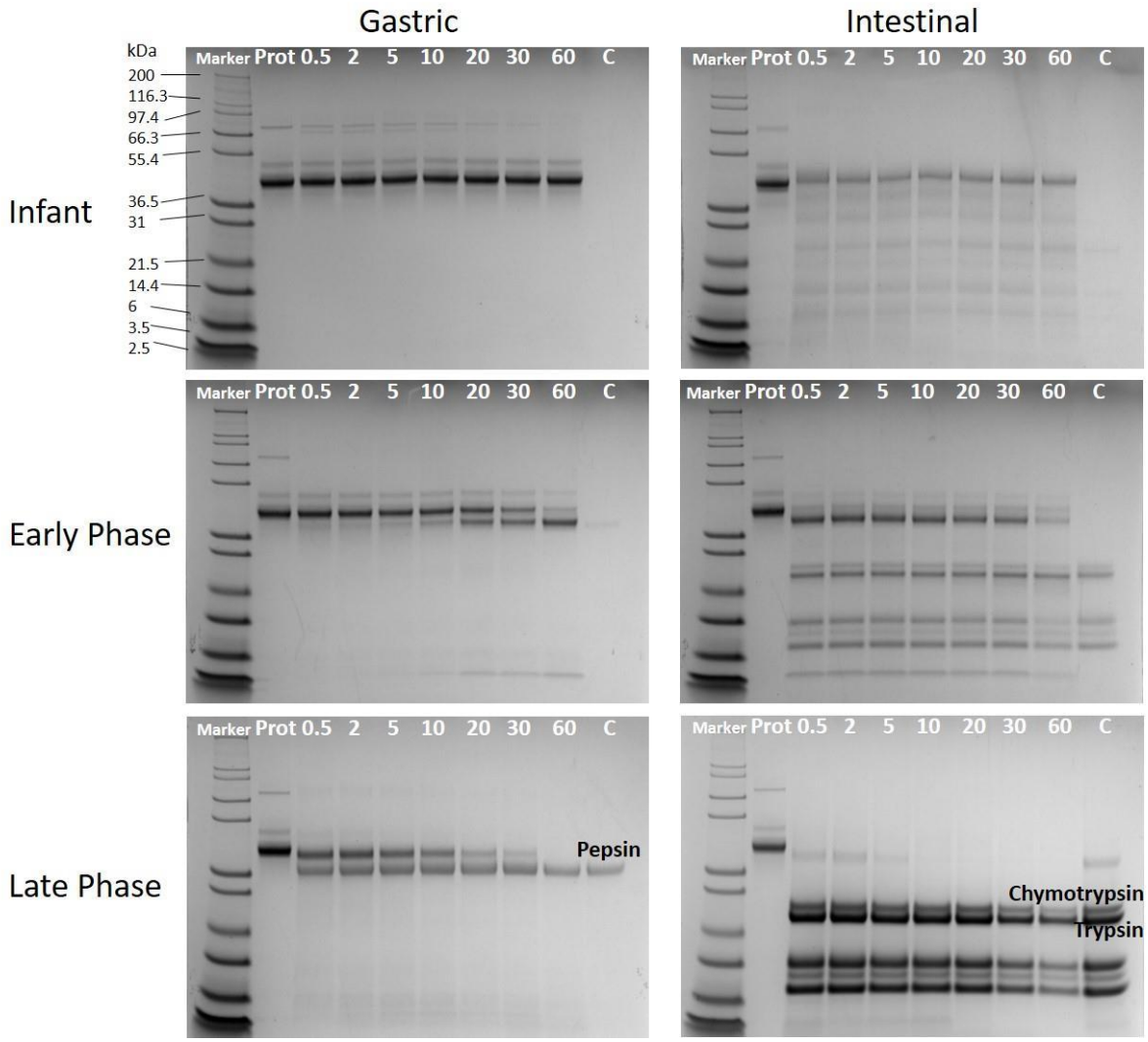


Figure 5: SDS-PAGE of the in vitro digestion of Ova with the infant, early phase adult (INFOGEST) and late phase adult (PRT) models. The numbers represent the time in min after starting each phase (gastric or intestinal). The Prot lane is the protein blank and C lane is the control of digestive enzymes.

The varying extent of proteolysis of Lys and Ova at the end of the intestinal phase within each in vitro digestion model, highlights the importance of comparing different GI conditions representative of the physiological situation in adults and infants. Despite the fact that the two proteins seemed to be completely hydrolysed by the end of the intestinal phase in the late phase adult model or current PRT, an appreciable percentage (40-50%) of intact protein remained in the infant model. In addition, in the case of Ova digesta, both large protein fragments (36-45 kDa) and low molecular weight hydrolysis products (3 kDa) appear across the intestinal phase of the early phase adult model. Therefore, the PRT may be suitable for a first screening, whereby proteins that are resistant to digestion under these harsh conditions are not expected to be digested under the milder conditions of early phase adult or infant

models. However, the fact that a protein is fully hydrolysed in the late phase adult model does not necessarily indicate complete digestion for the other two models.

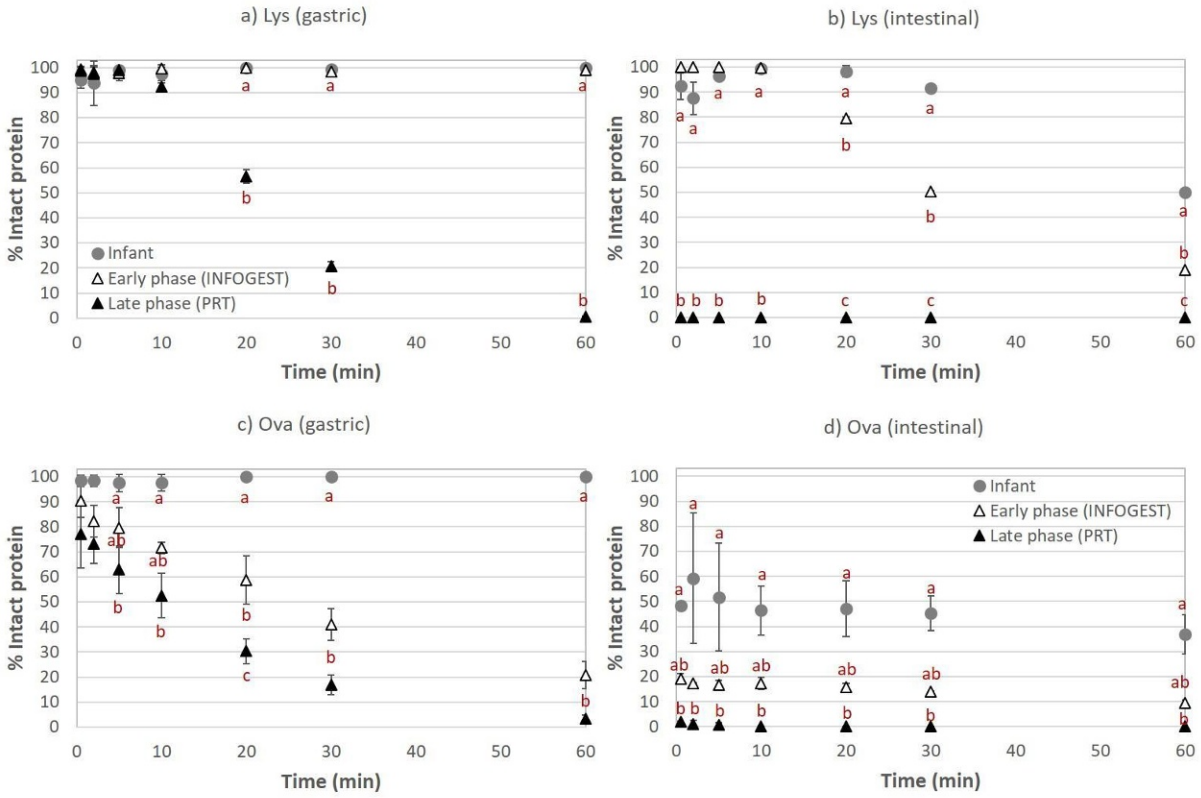


Figure 6: Percentage of intact protein a), b) Lys and c), d) Ova during in vitro gastric or intestinal digestion (infant, early phase or late phase models) determined from densitometry on SDS-PAGE (n ≥ 2). Different letters mean significant differences ($p \leq 0.05$) between models over time. Absence of letters means no significant differences.

Since Lys and Ova have shown differences in the degree of hydrolysis between 10 and 60 min of the gastric phase in the late phase adult model (Figures 4,5 bottom left and Figure 6a,c), we have also assessed the intestinal digestibility in this model immediately after 10 min of gastric digestion. The results (SDS-PAGE and densitometry profiles) are presented in Figure 7 and compared with those for the intestinal phase of the late phase adult model after 60 min of gastric digestion. Higher percentage of intact Lys (around 40%) is present at the beginning of the intestinal phase (Figure 7b) after 10 min of gastric digestion, when the remaining intact Lys was approximately 90% (Figure 6a), as compared to 0% (Figure 7b) after 60 min of gastric digestion because almost complete hydrolysis of the intact Lys had occurred by the end of the gastric phase (Figure 6a). The band corresponding to intact Lys has similar molecular weight to that of enzymes in the control (Figure 7a). However, greater intensity at this molecular weight can be visually detected during the first 20 min of the intestinal phase as compared to the protein blank and control of the enzymes, as reflected in the densitometry profile. Eventually, there are no differences in the Lys digestibility after 20-30 min of the intestinal phase. On the other hand, ca. 40% of intact Ova is present throughout the intestinal phase after 10 min of gastric digestion as compared to 0% of intact Ova after 60 min of gastric digestion (Figure 7d). This can be visually observed in the intestinal SDS-PAGE (Figure 7c). This suggests that the ca. 50% of intact Ova remaining

after 10 min of gastric digestion (Figure 6c) was only slightly hydrolysed in the subsequent intestinal phase. Takagi et al. found that Ova is relatively resistant to pancreatic enzymes without previous gastric digestion [13].

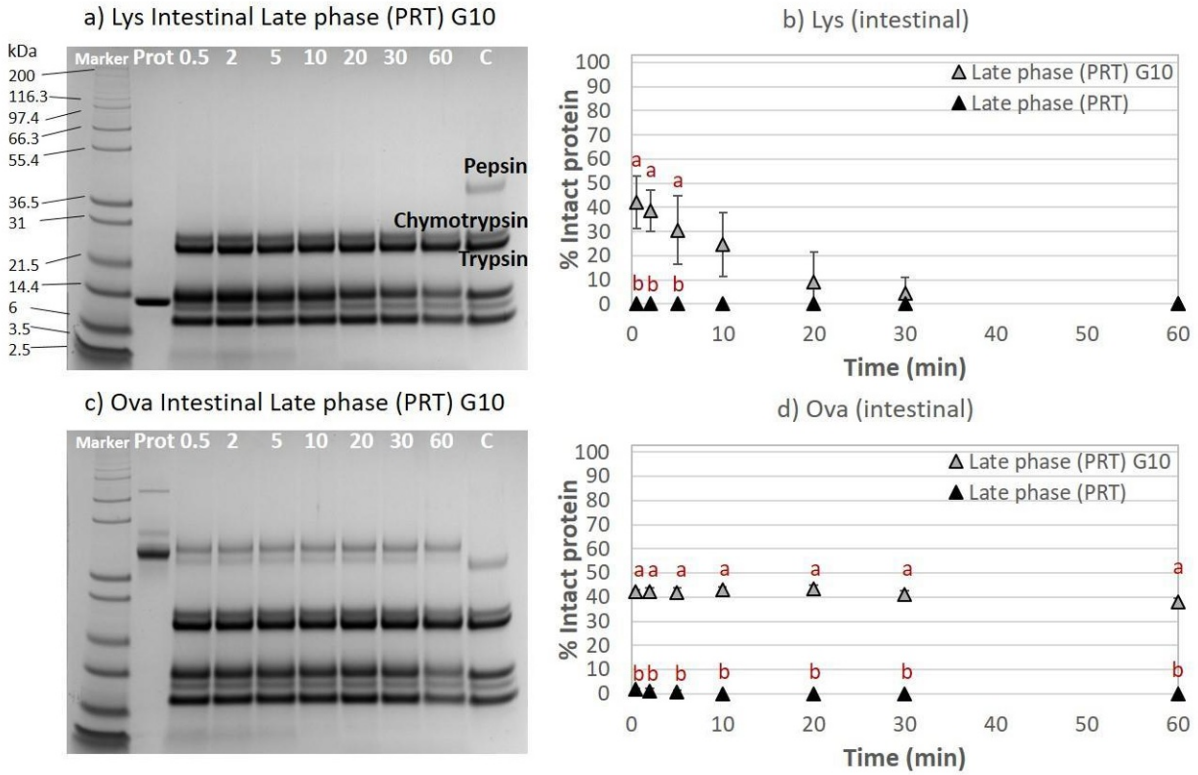


Figure 7: SDS-PAGE of the in vitro intestinal digestion of Lys a) and Ova c) with the late phase adult (PRT) model after 10 min of gastric digestion (G10). The numbers represent the time in min after starting the intestinal phase. The Prot lane is the protein blank and C lane is the control of digestive enzymes. Percentage of intact protein Lys b) and Ova d) during in vitro intestinal digestion (late phase model) determined from densitometry on SDS-PAGE (n ≥ 2). Different letters mean significant differences ($p \leq 0.05$) between models over time. Absence of letters means no significant differences.

3.1.3 Plant proteins

This section focuses on the digestion of purified proteins from plant origin.

KTI

The initial results are from KTI from soybean (Glycine max). This protein is widely consumed as part of soy products. However, as the protease inhibitor is an antinutritional factor, it is also highly resistant to digestion and known as a minor allergen (Gly m TI) of low relative abundance [8, 14]. The UniProt protein database designation is P01071. The results of digestion are reported in Figures 8 and 9. The intact protein KTI shows a band at a molecular weight of approximately 19 kDa, which is visible throughout the gastric phase regardless of the model of digestion (Figure 8, left), confirming the high resistance to pepsin, even in the late phase adult model, as seen elsewhere [8, 13-14]. Only in the late phase adult model, at 60 min of gastric digestion, there is an emerging decrease in the percentage of the intact protein

that is statistically significant, with 70% of intact protein still remaining (Figure 9a). The appearance of protein fragments of smaller molecular weight corresponding to hydrolysis products is not obvious. The degree of hydrolysis in the intestinal phase seems slightly faster and more extensive for early and late phase adult models, although there is no statistically significant difference across the three models (Figure 9b). Nevertheless, there are appreciable differences in the remaining protein fragments for each model of digestion (Figure 8, right) that may account for the slight differences in the kinetics of intestinal proteolysis. In the infant model, there is a smeared band ranging from 3 to 15 kDa which is present over the whole time period of the intestinal phase. In the intestinal early phase adult model, there is a light band within the range of 6 kDa, which is not present in the control of the enzymes, suggesting the persistence of smaller molecular weight products throughout the intestinal phase. Finally, faint bands of even smaller molecular weight (around 3 kDa) appear at the beginning of the late phase adult model (until 10 min), which may be related with a slightly more extensive KTI hydrolysis. However, it is important to note that 70% of intact protein remains in the infant model, and 30-40% in the early and late phase adult models. These results are in agreement with the stability of KTI shown under simulated intestinal conditions close to late phase adult model without previous gastric phase [13-14]. It is also interesting to note that previous work on static in vitro GI digestion of ground soybean seeds, mimicking human adults, has also observed resistance of intact soybean trypsin inhibitor throughout the gastric and intestinal phases, even when considering the interactions with the food matrix [16].

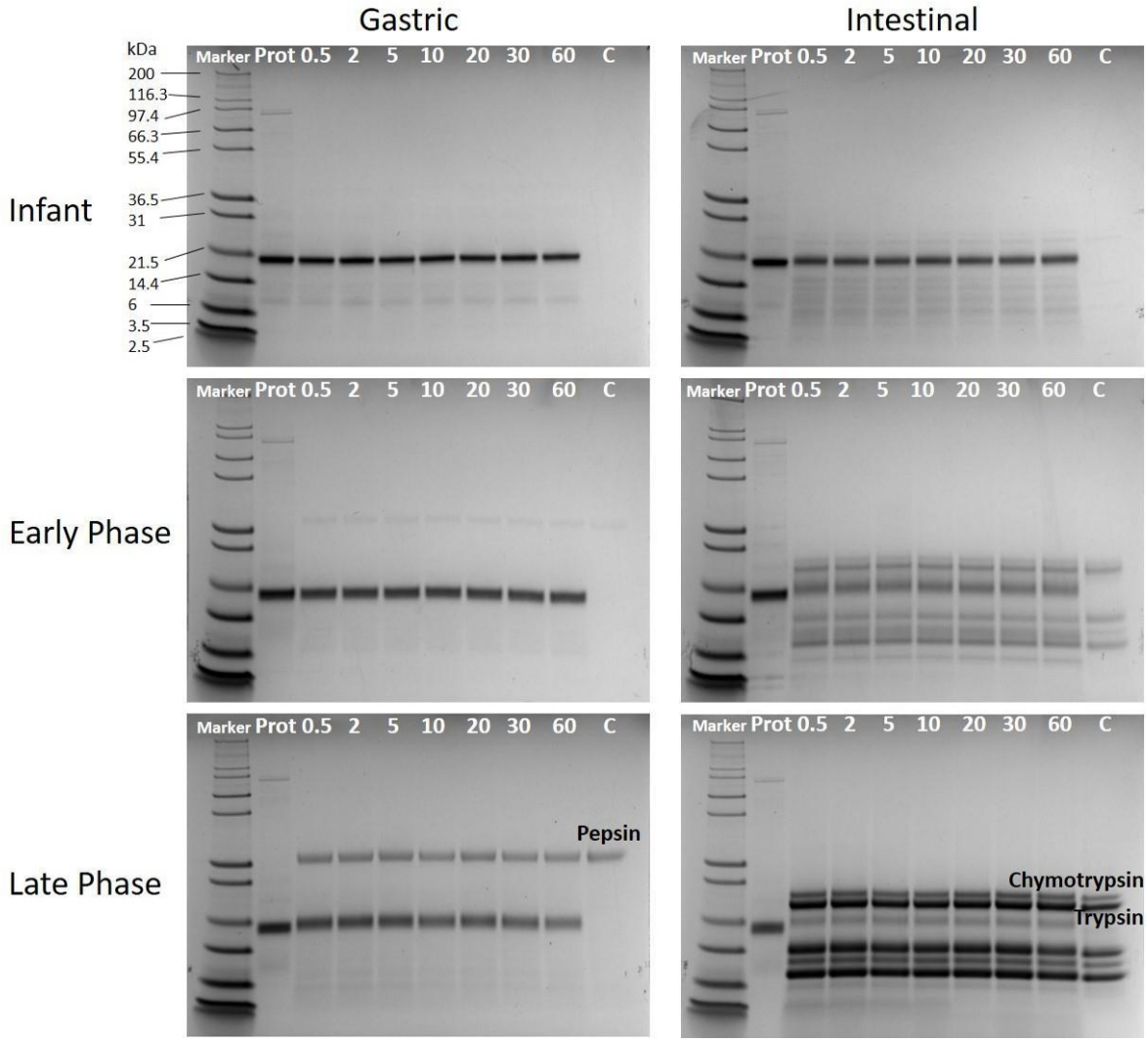


Figure 8: SDS-PAGE of the in vitro digestion of KTI with the infant, early phase adult (INFOGEST) and late phase adult (PRT) models. The numbers represent the time in min after starting each phase (gastric or intestinal). The Prot lane is the protein blank and C lane is the control of digestive enzymes.

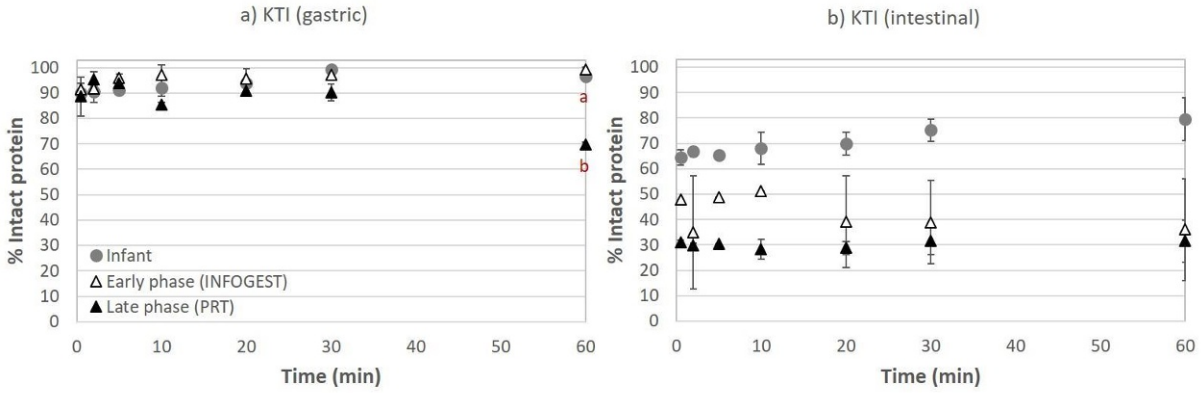


Figure 9: Percentage of intact protein KTI during in vitro gastric (a) or intestinal (b) digestion (infant, early phase or late phase models) determined from densitometry on SDS-PAGE (n ≥ 2). Different letters mean significant differences ($p \leq 0.05$) between models over time. Absence of letters means no significant differences.

LIP

As it is also from soybean, Lip is widely consumed, but is highly susceptible to digestion and is not known to be an allergen. The UniProt protein database designation is P08170. This protein shows in SDS-PAGE as a double band at a molecular weight of approximately 108 kDa, which consists of two subunits of 54 kDa each. Figures 10 (left) and 11a show that intact Lip is only persistent throughout the gastric phase in the infant model, with around 80% remaining after 60 min, whereas it is hydrolysed after 30 s of gastric digestion in both adult models. This is consistent with the results under late phase adult conditions reported by Astwood et al. [5]. Only smeared bands of around 3 kDa are present in the early and late phase adult models. For the intestinal phase, the infant model shows a rapid hydrolysis of the intact protein after 30 s of digestion (Figure 10, top right, and Figure 11b), with the appearance of hydrolysis products of molecular weight ranging from around 3 to 15 kDa, which gradually reduce their intensity over the course of the intestinal phase.

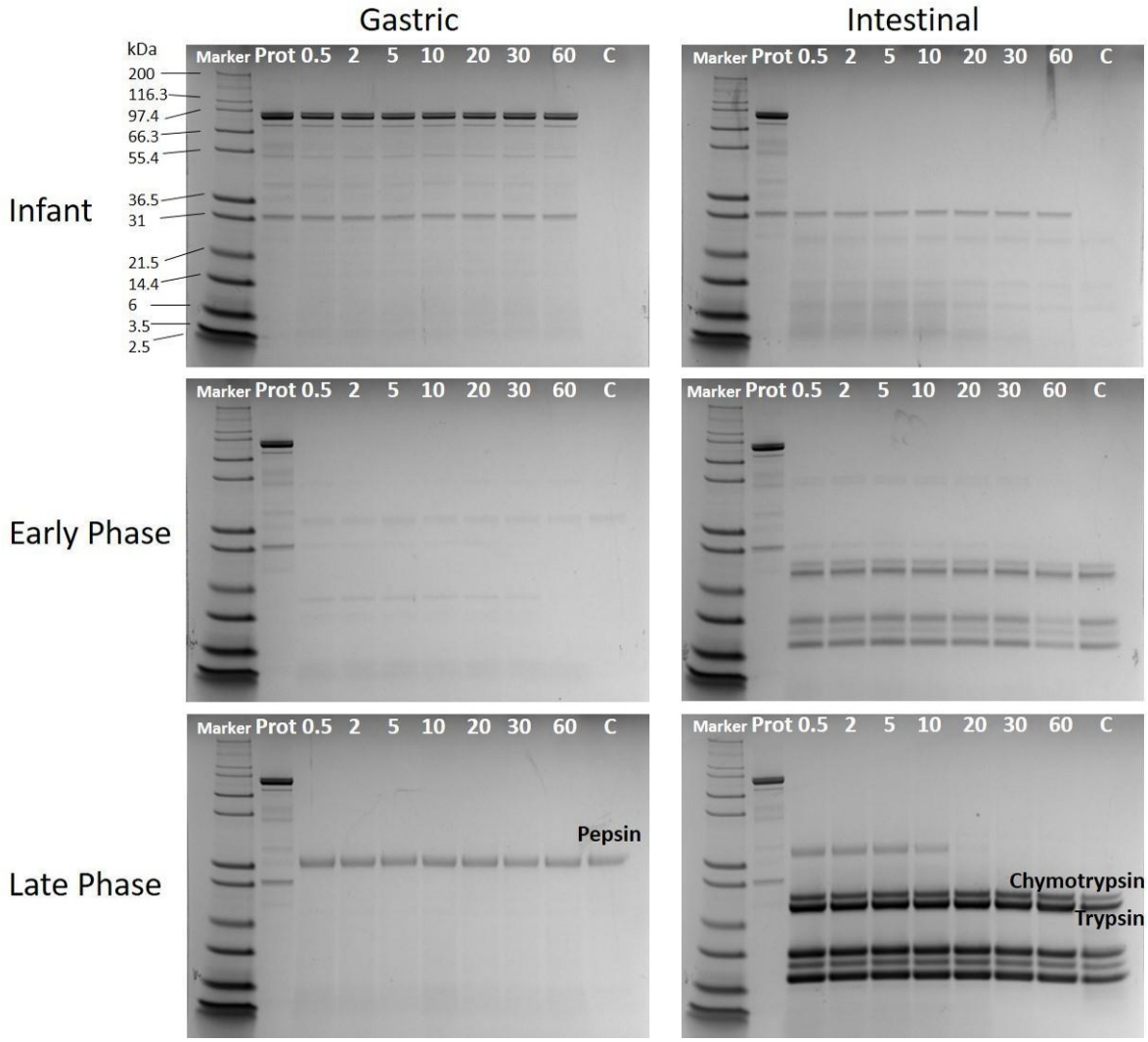


Figure 10: SDS-PAGE of the in vitro digestion of Lip with the infant, early phase adult (INFOGEST) and late phase adult (PRT) models. The numbers represent the time in min after starting each phase (gastric or intestinal). The Prot lane is the protein blank and C lane is the control of digestive enzymes.

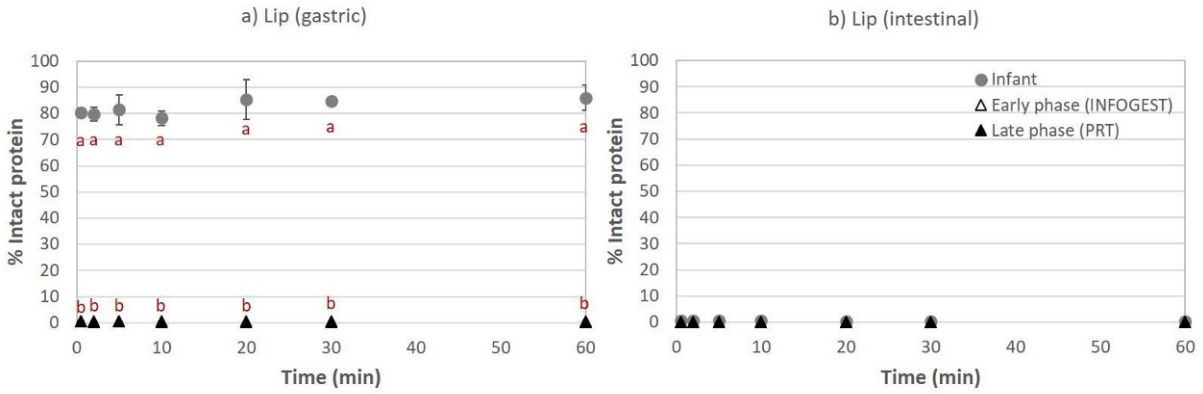


Figure 11: Percentage of intact protein Lip during in vitro gastric (a) or intestinal (b) digestion (infant, early phase or late phase models) determined from densitometry on SDS-PAGE (n ≥ 2). Different letters mean significant differences ($p \leq 0.05$) between models over time. Absence of letters means no significant differences.

ConA

The lectin ConA from Canavalia ensiformis (Jack bean) is relatively abundant and highly resistant to pepsin, but is not known to be allergenic although it is considered an antinutrient. The UniProt protein database designation is P02866. This seed protein has a monomeric molecular weight of approximately 30 kDa, but other minor naturally occurring fragments of smaller molecular weight (11 and 15 kDa) are present. Figures 12 and 13 illustrate the kinetics of hydrolysis of ConA under GI conditions with the three models. In particular, we see the persistence of intact protein (band corresponding to 30 kDa and naturally occurring fragments) throughout the gastric and intestinal phases in the infant and early phase adult models (Figure 12). Only in the late phase adult model does the intensity of the band corresponding to intact ConA decrease appreciably after 60 min of gastric digestion. Although it is still visible and protein fragments of slightly smaller molecular weight (22-30 kDa) appear after 2 min of gastric digestion and the corresponding band becomes more intense during the course of the gastric phase. In addition, hydrolysis products of molecular weight ranging between 3 and 6 kDa are visually detected at the end of the gastric phase. All these observations agree well with the SDS-PAGE profile of ConA subjected to PRT [8].

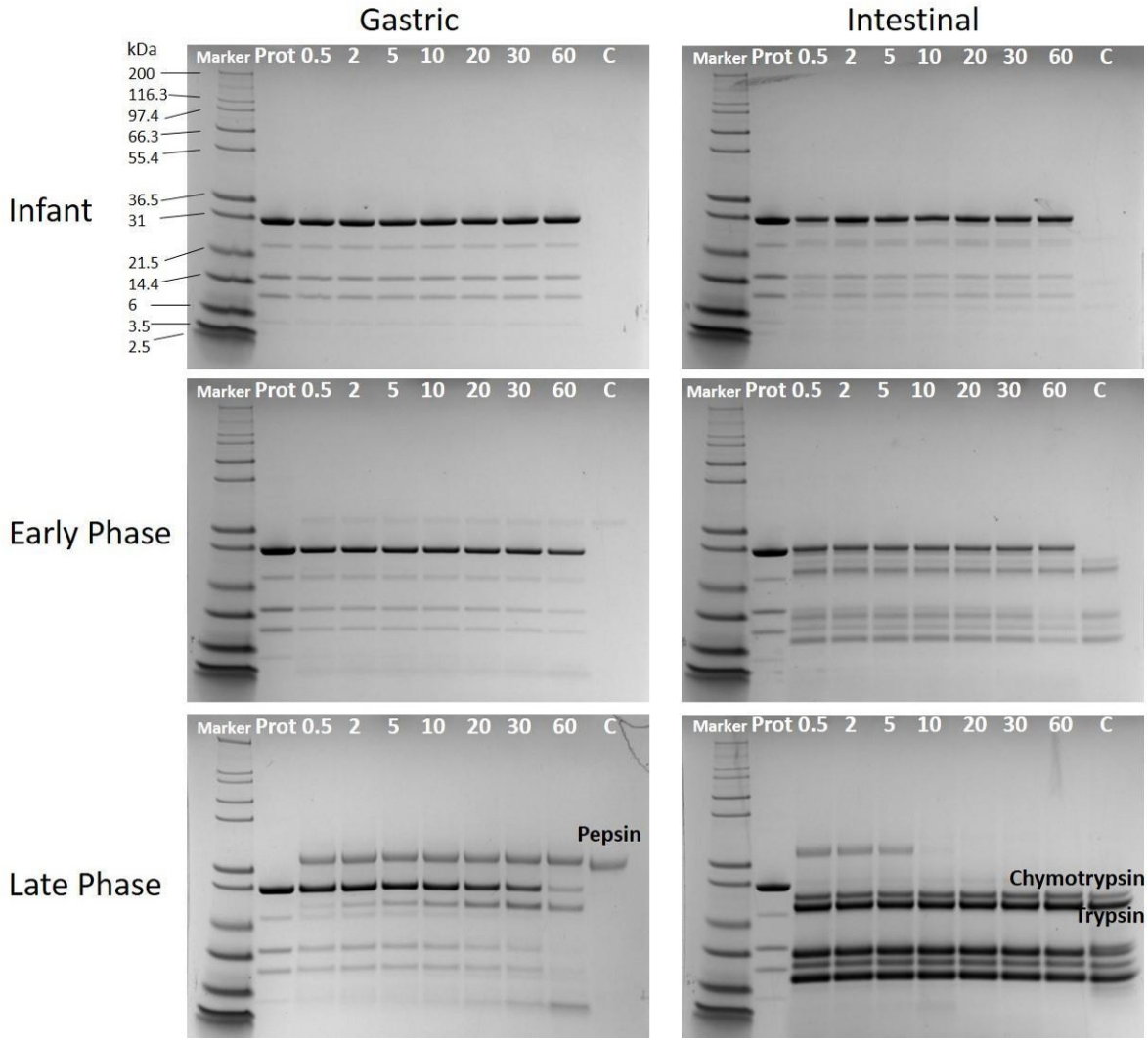


Figure 12: SDS-PAGE of the in vitro digestion of ConA with the infant, early phase adult (INFOGEST) and late phase adult (PRT) models. The numbers represent the time in min after starting each phase (gastric or intestinal). The Prot lane is the protein blank and C lane is the control of digestive enzymes.

It can be observed in Figure 13 that, in general, the maximum extent of hydrolysis of intact protein is reached within the gastric phase and a plateau is further seen in the subsequent intestinal phase. Nevertheless, there are statistically significant differences in the rate and extent of hydrolysis across the three models. Namely, the gastric proteolysis becomes statistically faster after 20 min for early and late phase adult models, as compared to infant model, and after 30 min is even faster for late phase adult model. Takagi et al. reported a very similar densitometry profile of intact ConA under gastric conditions of late phase adult model or PRT [13]. Finally, after 60 min of intestinal phase, 80% of intact protein remains for the infant model, whereas 50% and less than 10% remains for the early and late phase adult models, respectively.

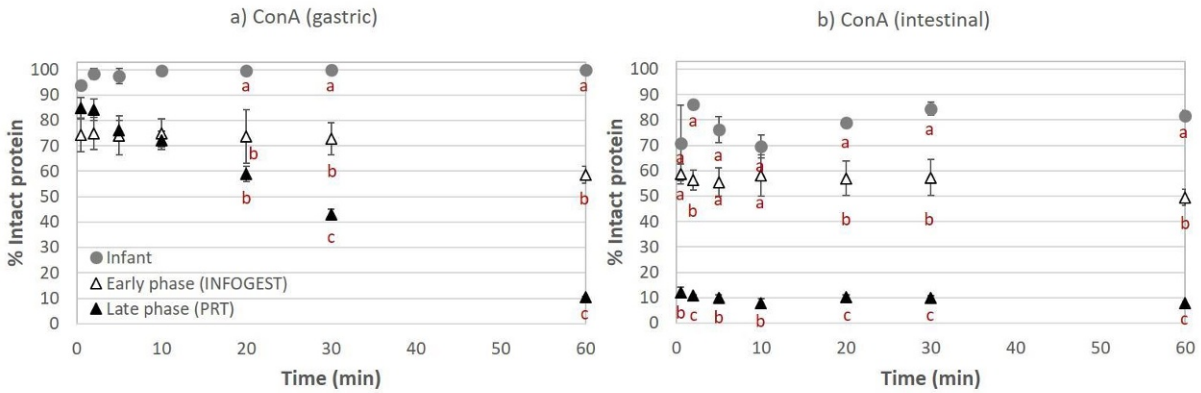


Figure 13: Percentage of intact protein ConA during in vitro gastric (a) or intestinal (b) digestion (infant, early phase or late phase models) determined from densitometry on SDS-PAGE (n ≥ 2). Different letters mean significant differences (p ≤ 0.05) between models over time. Absence of letters means no significant differences.

ConA has shown a difference in the degree of hydrolysis between 10 and 60 min of the gastric phase in the late phase adult model (Figure 12 bottom left and Figure 13a). Therefore, we have also evaluated the intestinal digestibility of ConA with this model after 10 min of gastric digestion. The SDS-PAGE and densitometry profile are presented in Figure 14. Comparison of Figures 14a and 12 (bottom right) evidences a greater intensity in the band corresponding to intact ConA (30 kDa) throughout the intestinal phase after 10 min of gastric digestion, as compared to that after 60 min of gastric digestion. Namely, approximately 60% of intact ConA remains at the end of the intestinal phase when this was subsequent to 10 min of gastric digestion, whereas less than 10% remains when the intestinal phase was subsequent to 60 min of gastric digestion (Figure 14b). This suggests once more, that a maximum extent of hydrolysis is reached at the end of the gastric phase, ca. 70% intact ConA remaining at 10 min of gastric phase (Figure 13a), and a steady value is observed throughout the subsequent intestinal phase (Figure 14b). This is in agreement with the findings of Takagi et al., who observed that ConA is very resistant to pancreatic enzymes under intestinal conditions without previous gastric digestion [13].

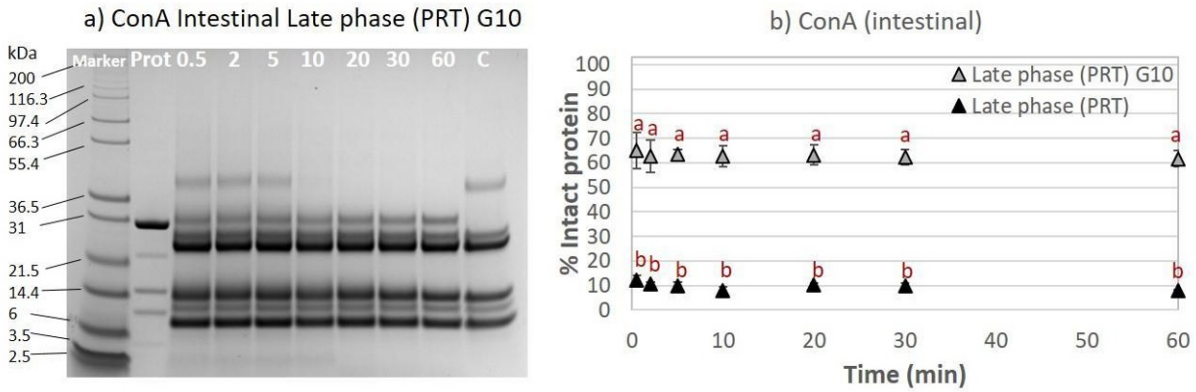


Figure 14: SDS-PAGE of the in vitro intestinal digestion of ConA a) with the late phase adult (PRT) model after 10 min of gastric digestion (G10). The numbers represent the time in min after starting the intestinal phase. The Prot lane is the protein blank and C lane is the control of digestive enzymes. Percentage of intact protein ConA b) during in vitro intestinal digestion (late phase model) determined from densitometry on SDS-PAGE (n ≥ 2). Different letters mean significant differences (p ≤ 0.05) between models over time. Absence of letters means no significant differences.

AP

The source of AP, namely potato is widely consumed, the protein is highly digestible and is not known to be allergenic. The UniProt protein database designation is to be confirmed. The enzyme AP has a major isoform of a homodimer with a molecular weight of 96 kDa. Therefore, the major band shown in SDS-PAGE (Figure 15) corresponds to the monomeric molecular weight of around 45-50 kDa. Figure 15 shows that intact AP is only persistent throughout the gastric phase in the infant model, whereas it is rapidly hydrolysed by pepsin immediately after starting the gastric phase in both adult models. However, there are slight but significant differences in the kinetics of AP gastric digestion between the adult models (Figure 16a). During the first two min of gastric phase, the early phase adult model exhibits slightly slower hydrolysis of intact AP than the late phase adult model. In addition, the last time that the band of intact AP could be seen in the SDS-PAGE was 10-20 min (Figure 15, middle left), although the measured intensity was not significantly different with respect those in the late phase adult model at the same time points. There are no obvious bands of smaller molecular weight appearing during the course of the gastric phase in either adult models, suggesting a fast and likely complete hydrolysis of AP under these conditions. These results corroborate the weak stability of AP in SGF of PRT (less than 15 s) reported previously [5]. Since AP has already been hydrolysed in the gastric phase of both adult models, there is no further discussion for the SDS-PAGE results of the intestinal phase. In the case of the infant model, despite AP being still intact (95-100%) after 60 min of the gastric phase (Figures 15,16a), there is a rapid proteolysis as soon as the intestinal phase starts (Figure 16b). Only a smeared band ranging from 3 to 10 kDa appears at 30 s of infant intestinal digestion, which gradually decreases in intensity afterwards and is identified as likely digestion products from AP.

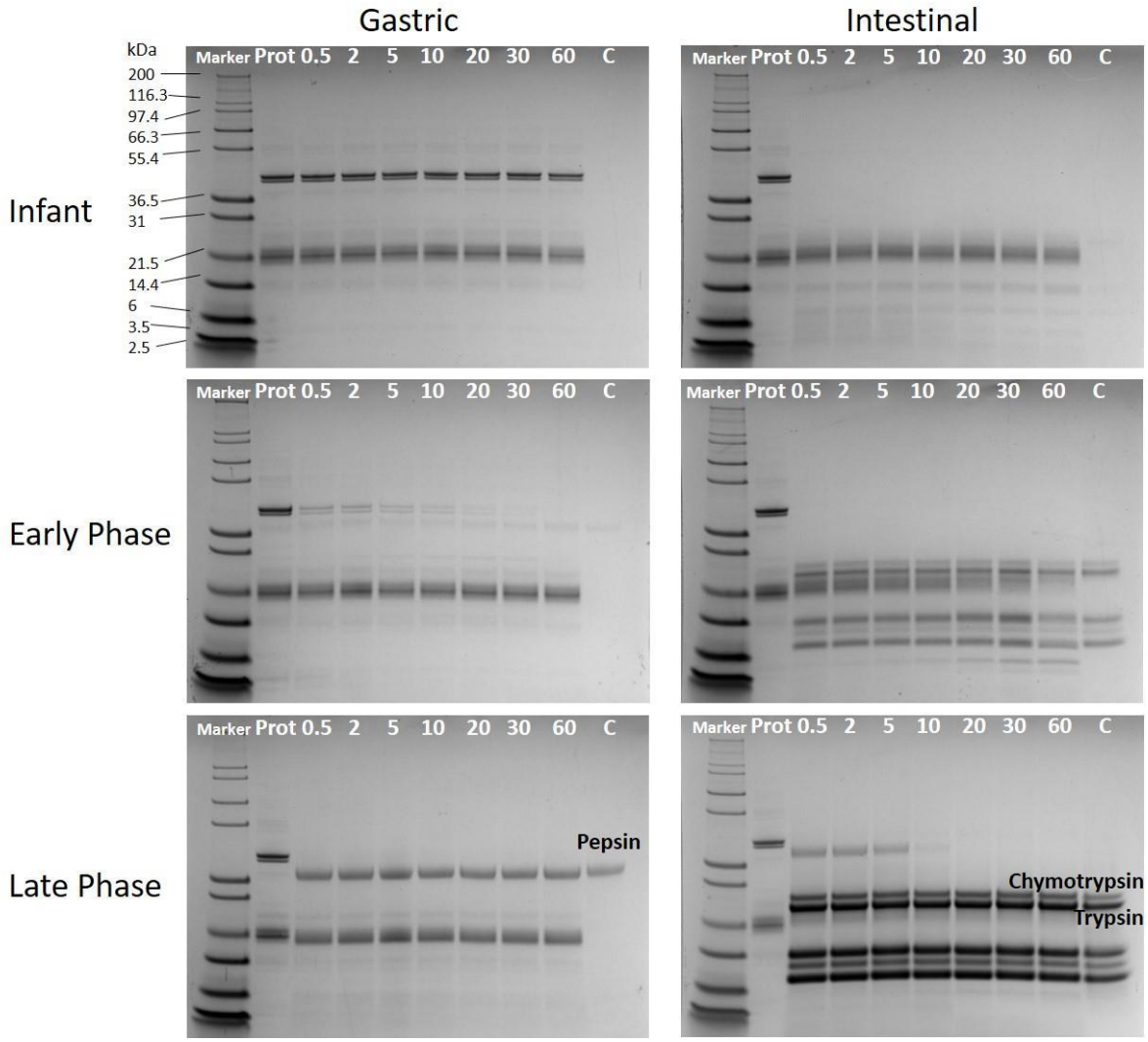


Figure 15: SDS-PAGE of the in vitro digestion of AP with the infant, early phase adult (INFOGEST) and late phase adult (PRT) models. The numbers represent the time in min after starting each phase (gastric or intestinal). The Prot lane is the protein blank and C lane is the control of digestive enzymes.

Regarding the lower molecular weight band (21.5 kDa) observed in the lane of protein blank in Figure 15, it does not seem to be a naturally occurring fragment of AP given the persistent nature shown under GI conditions, regardless of the model of digestion. The manufacturer reported a purity between 70 and 90%, therefore it could be identified as a major contaminant.

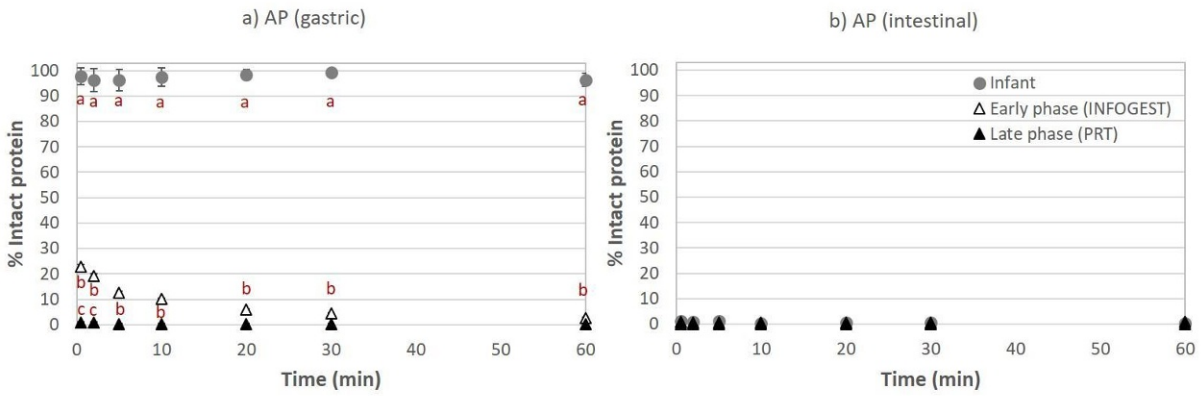


Figure 16: Percentage of intact protein AP during in vitro gastric (a) or intestinal (b) digestion (infant, early phase or late phase models) determined from densitometry on SDS-PAGE (n ≥ 2). Different letters mean significant differences ($p \leq 0.05$) between models over time. Absence of letters means no significant differences.

Ara h 1

Ara h 1 from peanut (Arachis hypogaea) is one of the main storage proteins (7S globulin) of the seed, is highly susceptible to digestion and is also known as a major allergen. The UniProt protein database designation is P43237, P43238. Ara h 1 is a stable homotrimer in its native form with a molecular weight of around 235 kDa. SDS-PAGE shows a major band of approximately 67 kDa (Prot in Figure 17), which corresponds to the monomeric form as a result of the denaturing conditions of LDS [17], and a minor band corresponding to a subunit of about 33 kDa, as reported previously [6]. The other minor band of around 130-150 kDa may correspond to a dimeric form, as reported elsewhere [7]. Figures 17 (left) and 18a show that in general, Ara h 1 is rapidly hydrolysed under gastric conditions. The fast digestion of intact Ara h 1 in the early and late phase adult models is in agreement with the results reported by Eiwegger et al. and Fu et al., respectively, under similar conditions for each adult model [7, 14]. Nevertheless, the kinetics is slower in the infant model as compared to both adult models, but these differences are not statistically significant as shown in Figure 18a. However, it is important to note that as soon as the gastric phase of the infant model starts, extensive precipitation of Ara h 1 occurs and aggregates are observed until 60 min. This is likely due to the pH of 5.3, which is close to the isoelectric point of Ara h 1 [18] and therefore a reduced amount of protein would be available for pepsin cleavage. This may explain the sudden decrease in band intensity corresponding to intact Ara h 1 after 30 s of infant gastric digestion (Figure 17, top left). Interestingly, this reduced Ara h 1 soluble fraction remains stable until min 10 of gastric digestion, when slightly smaller products became visible in the SDS-PAGE and only after 30-60 min a faint smeared band (3-14 kDa) is detected, although the intact protein is still visible. Previous results on in vitro gastric digestion of purified peanut allergens, showed that Ara h 1 was much more rapidly digested with a pepsin to test protein ratio similar to that used in the infant model [17]. The most plausible explanation for the different rate and extent of hydrolysis obtained in our study lies in the gastric pH, which is 5.3 versus 1.2 in the reported study, which as already indicated, is out of the optimum range for pepsin activity. On the other hand, the smaller molecular weight products that are already visible after 30 s of gastric digestion for both adult models are gradually hydrolysed, and intact protein is no longer detected, as shown previously [7]. Therefore, Ara h 1 seems to be more resistant to pepsin under infant gastric conditions and yet non-digested protein remains in the aggregates after starting the intestinal phase. In the intestinal phase of the infant model, the pH is raised to a value of 6.6, which

allows the re-solubilisation of Ara h 1 aggregates over the course of the intestinal digestion and may explain the appearance of hydrolysis products of larger molecular weight or increase in quantity (bands becoming more intense) over time. What is clear is that the intact protein is no longer visible after starting the intestinal phase, suggesting a rapid hydrolysis as soon as protein is re-solubilised from the aggregates. By the end of the intestinal phase, no bands are detected corresponding to hydrolysis products and the digesta was completely clear in appearance, with no visible aggregates. This suggests a complete digestion of Ara h 1, as far as the SDS-PAGE allows detection.

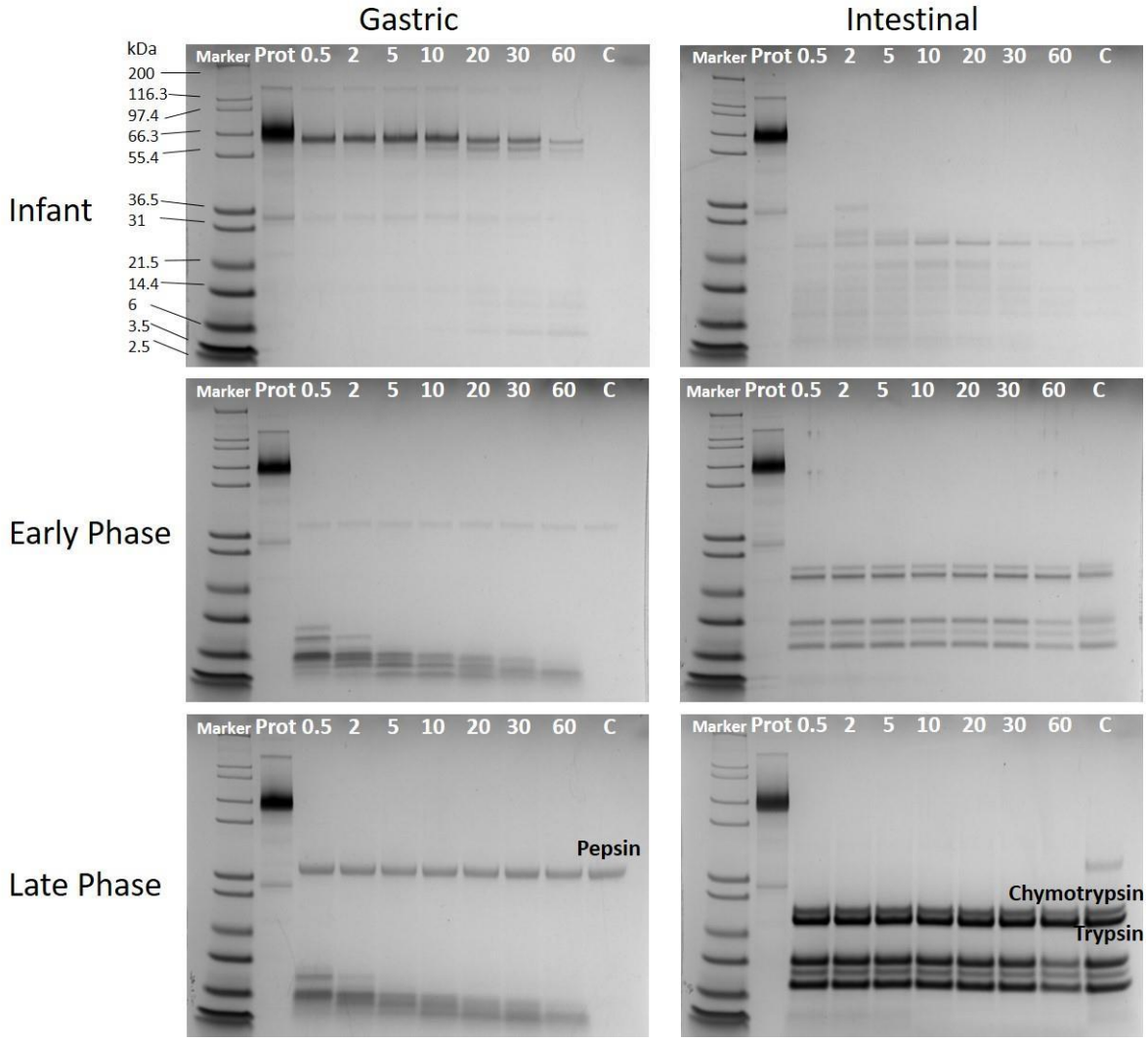


Figure 17: SDS-PAGE of the in vitro digestion of Ara h 1 with the infant, early phase adult (INFOGEST) and late phase adult (PRT) models. The numbers represent the time in min after starting each phase (gastric or intestinal). The Prot lane is the protein blank and C lane is the control of digestive enzymes.

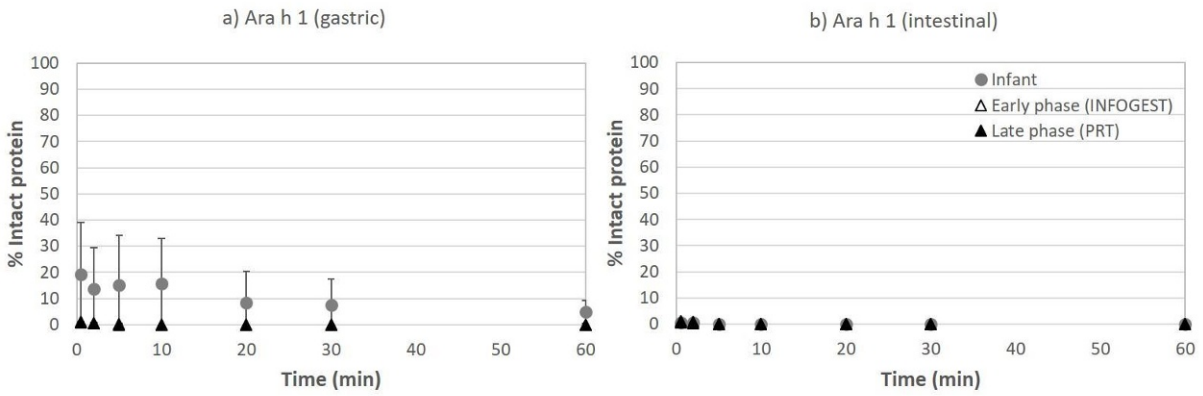


Figure 18: Percentage of intact protein Ara h 1 during in vitro gastric (a) or intestinal (b) digestion (infant, early phase or late phase models) determined from densitometry on SDS-PAGE (n ≥ 2). Different letters mean significant differences ($p \leq 0.05$) between models over time. Absence of letters means no significant differences.

GLI

The last protein studied under the gastro-duodenal conditions of infant, early phase adult and late phase adult models is GLI from bread wheat (Triticum aestivum). This is a mixture of storage proteins (prolamins) primarily insoluble in water, which are widely consumed and are involved in the pathogenesis of celiac disease. They comprise monomeric proteins subdivided into α/β-, γ-, and ω-fractions, according to their electrophoretic profile at low pH. They show as multiple bands at the molecular weight range of 35-45 kDa in SDS-PAGE (Figure 19). Due to their poor solubility in water, a minor soluble fraction would be at first accounted for, to be detected in SDS-PAGE, which may vary upon digestion by pepsin and trypsin/chymotrypsin. For this reason, the amount of protein loaded into SDS-PAGE for each model of digestion is the maximum allowed by the dilution of protein to simulated gastrointestinal fluid in the considered model, but keeping consistency among the protein blank, gastric and intestinal phases within each model. Therefore, the dilution factor of the GLI blank goes from lowest in the infant model, to highest in the late phase adult model. In addition, densitometry analysis is not relevant in this case, for the reason given above. Namely, the extent of solubility may vary at each time point of digestion and the accuracy for the much diluted samples in the late phase adult model is compromised. Figure 19 (top) shows that for the infant model, the soluble fraction of intact GLI is partially hydrolysed soon after starting the gastric digestion (30 s). This is reflected in the decreased intensity of the group of bands corresponding to gliadins accompanied by the appearance of smeared bands of lower molecular weight (3-40 kDa) indicative of digestion products. By the end of the gastric phase, a relatively large fraction of insoluble gliadin was still present. At the beginning of the intestinal phase, the increased intensity of the bands corresponding to smaller molecular weight fragments of GLI, as compared to the end of the gastric phase, suggests that remaining insoluble gliadin was at least partially hydrolysed by trypsin and chymotrypsin. Over the course of the intestinal phase, these soluble protein fragments were gradually broken down into smaller ones, but could still be detected after 60 min of intestinal digestion (3-6 kDa). Despite a relatively large fraction of insoluble GLI remaining after 60 min of intestinal digestion, no intact protein was visually detected in the SDS-PAGE for the soluble counterpart.

Figure 19 (middle) displays the kinetics of proteolysis for the early phase adult model. As for the infant model, the soluble fraction of intact gliadin seems to be rapidly hydrolysed after 30 s of the gastric digestion. Nevertheless, the presence of more intense bands corresponding to smaller fragments, as compared to the infant gastric model, suggest a more extensive digestion of the insoluble gliadin. In fact, the amount of precipitated solid at the end of the gastric phase for early phase adult model seemed less than for the infant model. Smith and co-workers also observed a rapid gastric hydrolysis of isolated gliadin fraction from wheat under conditions of pepsin to test protein ratio similar to the infant model, but lower pH 2.5, which allows optimum pepsin activity [19]. Namely, the gliadin fraction (35-45 kDa) was hydrolysed after 10 min of gastric digestion with a trace of smaller molecular weight fragments remaining afterwards and gradually breaking down into smaller fragments over the 60 min of the gastric phase. Subsequent intestinal digestion in the early phase adult model shows only a smeared band of 3-6 kDa throughout the duration of this phase, but no detection of intact protein, and only a trace of precipitated GLI remained.

Figure 19 (bottom) shows the SDS-PAGE digesta of the late phase adult model. As anticipated, the large dilution of this digestion protocol does not allow the proper detection of the intact protein, although a faint smeared band between 3 and 6 kDa can be visualised throughout the gastric phase and first 5 min of the intestinal phase. This along with the results of early phase adult model and the fact that insoluble fraction of GLI was not observed by naked eye, suggests that the fastest and largest extent of GLI proteolysis takes place under conditions of the late phase adult model.

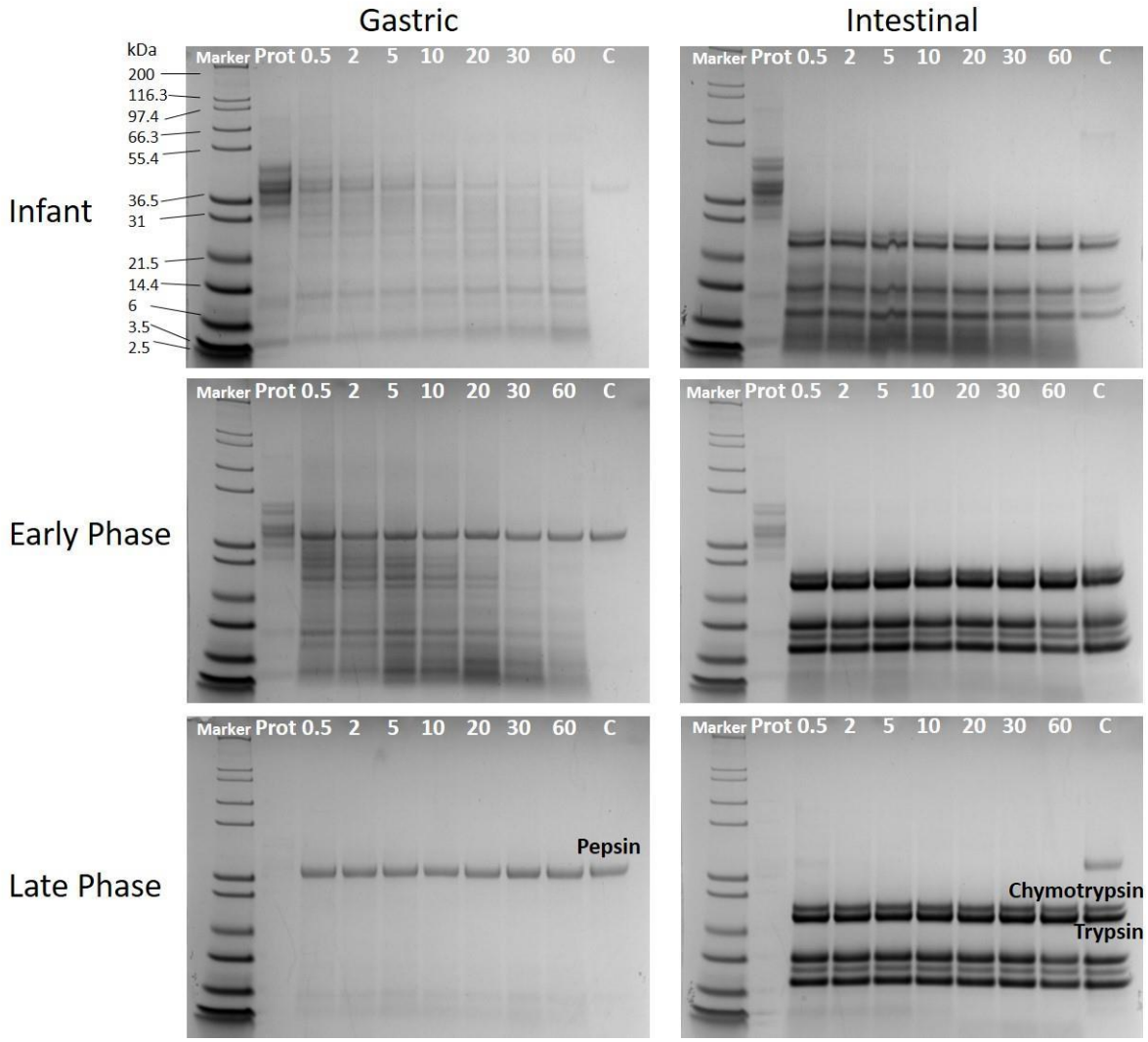


Figure 19: SDS-PAGE of the in vitro digestion of GLI with the infant, early phase adult (INFOGEST) and late phase adult (PRT) models. The numbers represent the time in min after starting each phase (gastric or intestinal). The Prot lane is the protein blank and C lane is the control of digestive enzymes.

3.2 Mass Spectrometry

In total, around 2000 injections were made including blanks and quality controls and generated 1.5TB of data. The digested protein samples were analysed by LC-MS and a statistical analysis of the identified peptides longer than 9AA was performed. Peptides of Mw higher than 4 kDa were not detected automatically by the technique.

3.2.1 Bovine milk proteins

The Box plots in Figure 20 show the size of the peptides for BCAS when keeping only peptides that were successfully quantified in, at least, two of the three replicates.

BCAS

For BCAS, 472 peptides were unambiguously identified in the gastric phase, whereas only 296 were found in the intestinal phase. The median values of the peptide molecular weight (Mw) observed with the early phase model (INFOGEST) is higher than that of the late phase model (PRT) confirming a more intense proteolysis in the latter. Peptide Mw values observed with the infant model are even higher after 5 min of digestion. In the intestinal phase, median Mw values were much lower than that for the gastric phase. Between the three models, the infant one led to the highest Mw as expected. With the early phase model, Mw decreased consistently over time, which was not the case with the late phase model.

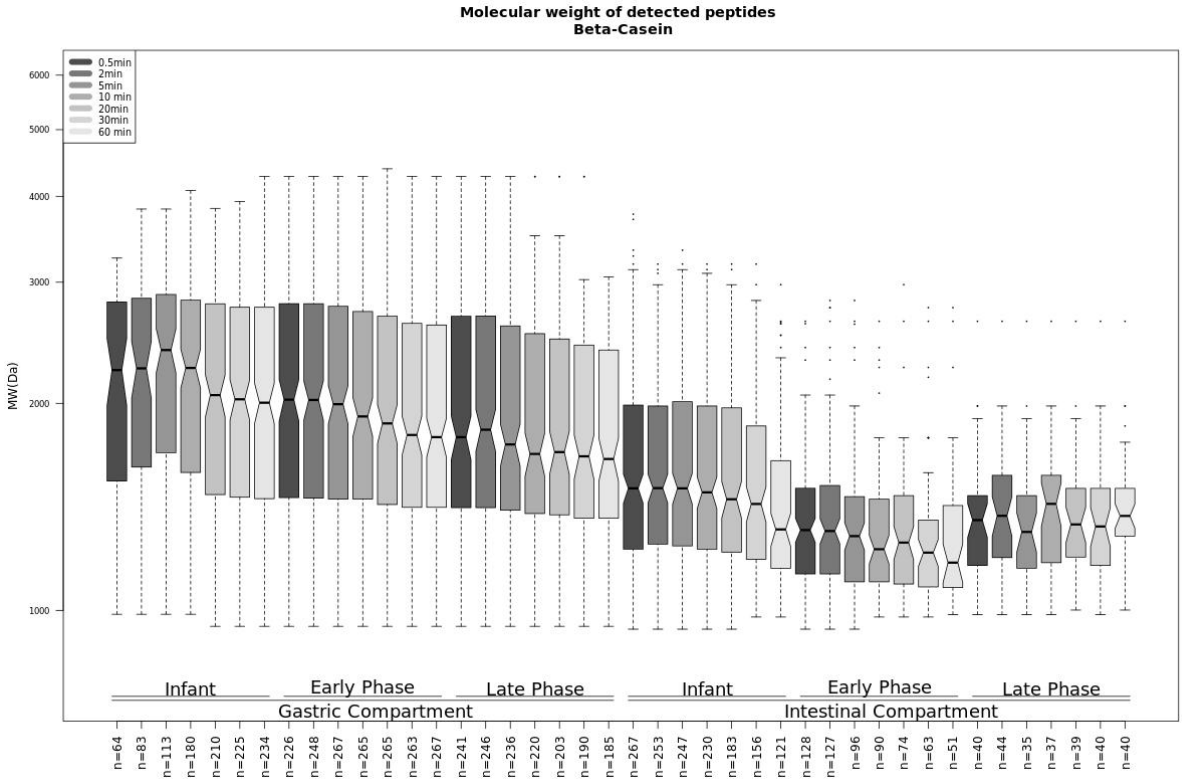


Figure 20: A box plot of peptide molecular weight from BCAS after gastric and intestinal digestion.

A total of 313 peptides was considered for the clustering analysis. Figure 21 showed differences between models in terms of the peptides identified in the gastric phase. Seven different groups (clusters) of peptides could be clearly discriminated.

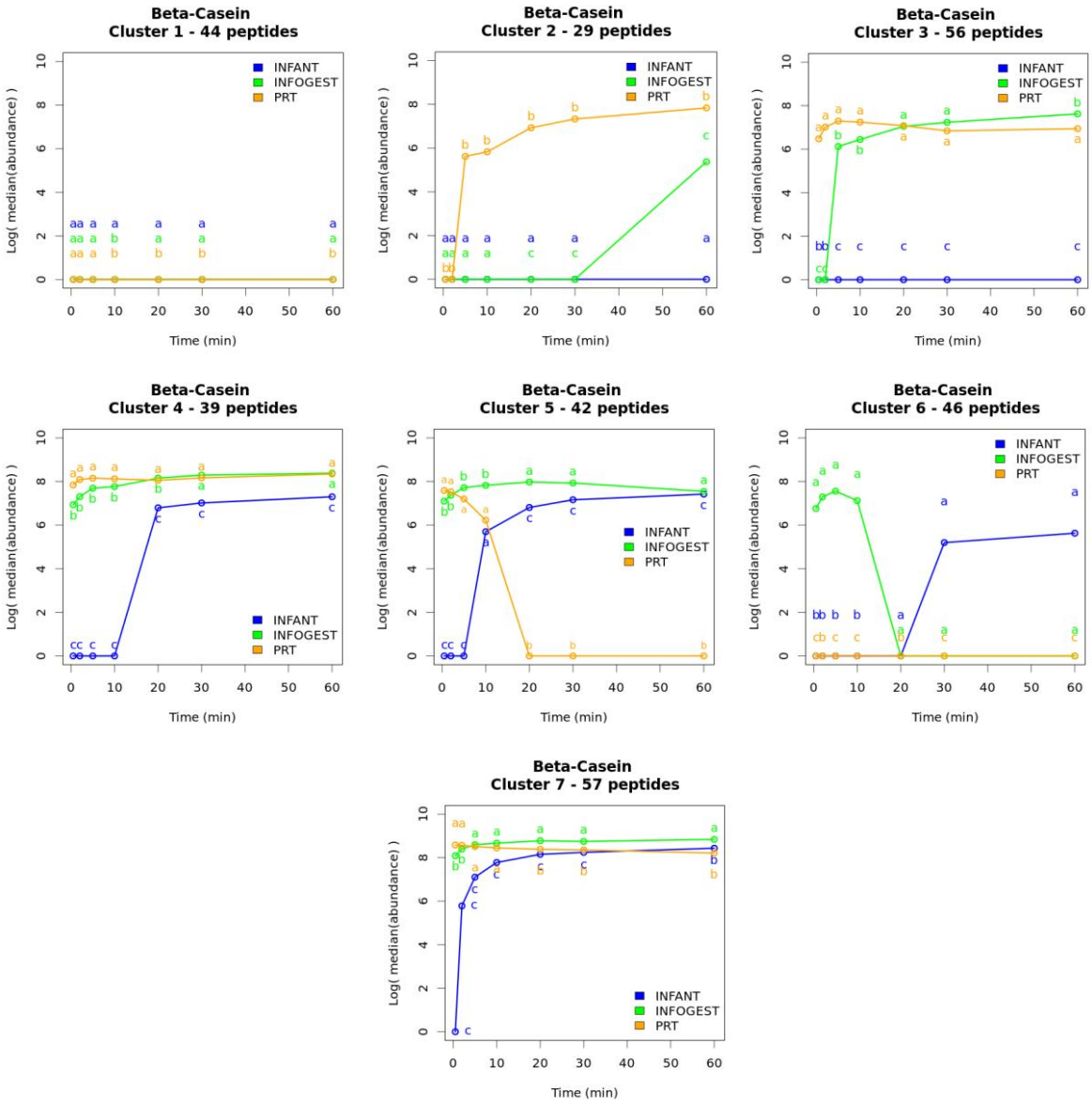


Figure 21: Kinetic plots of area under the curve for the 7 peptide clusters of BCAS through digestion in the infant (blue), early phase (green) and late phase (yellow)

When comparing the different digestion models, the most interesting clusters appear to be Cluster 2 and 5. Indeed, **Cluster 2** (29 peptides) corresponds to peptides with very low abundances in the Infant and early phase model whereas those peptides are in high abundance after 5 min digestion with the late phase model. In contrast, **Cluster 5** (42 peptides) exhibits major differences in the kinetics of appearance of the peptides depending on the digestion model. The **Cluster 5** peptides are present throughout the gastric phase in the early phase model, disappear after 20 min of digestion in the late phase model and appear after 10 min of digestion with the Infant model. In the heat map shown in Figure 22 below, the 42 peptides in **Cluster 5** are mapped onto the primary sequence. This cluster showed the largest differences between the digestion scenarios.

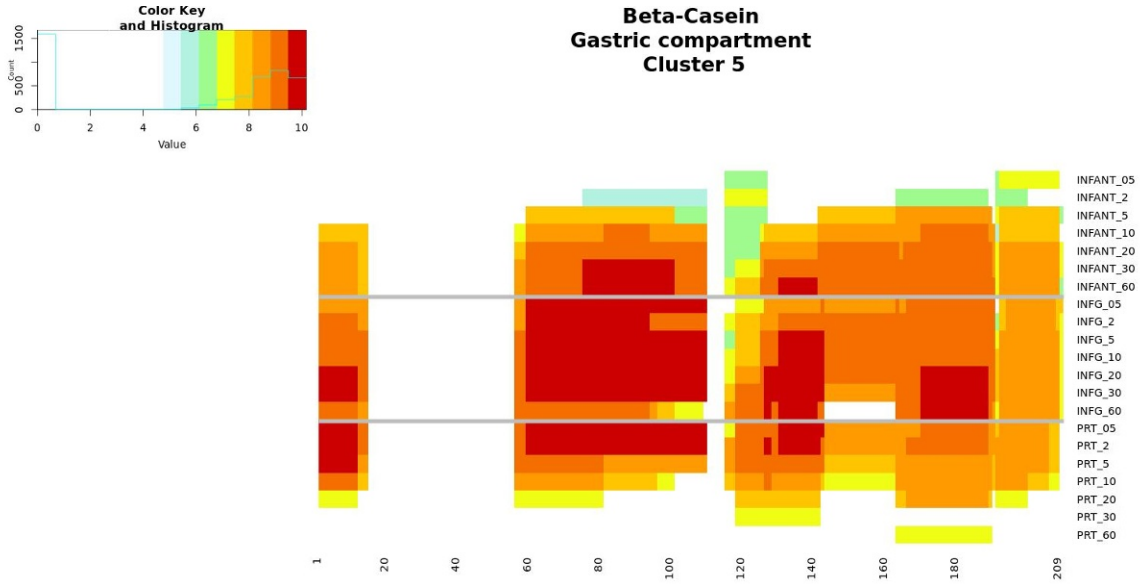


Figure 22: A heat map showing the sum of peptide abundances per amino acid from **Cluster** 5 peptides generated at different time points for the different gastric digestion scenarios from BCAS. The scenarios are infant, INFG (early phase) and PRT (late phase).

For comparison, the total peptides identified from BCAS are overlaid on the primary sequence in Figure 23. As one might expect, there are some differences between the two figures. In both cases, the differences along the sequence are generated by either the production of peptides that are smaller than nine amino acids or larger than the detection limit of the MS.

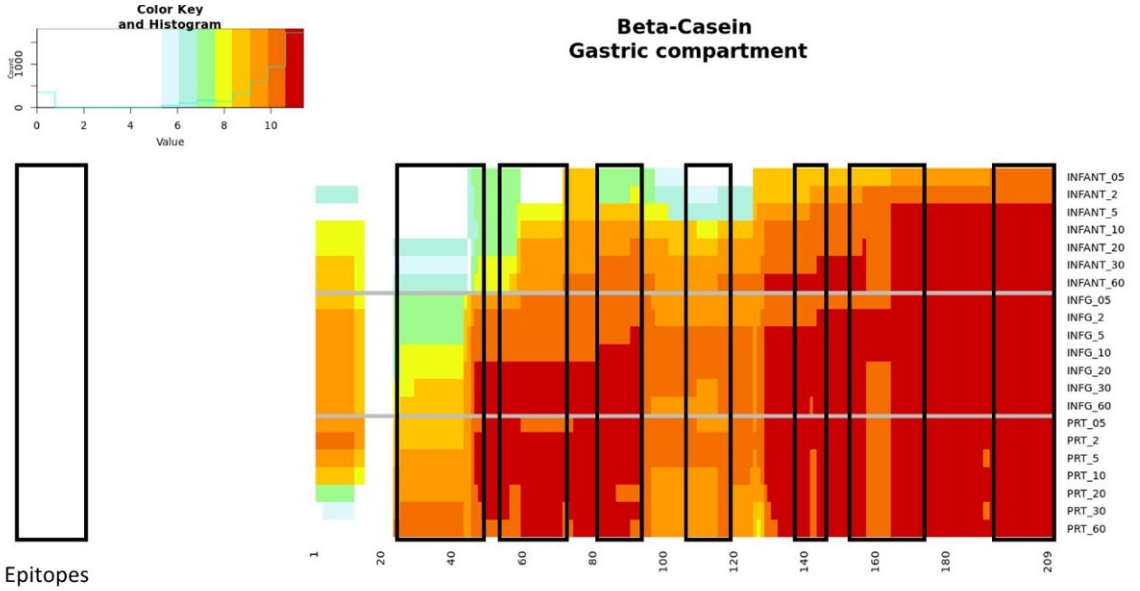


Figure 23: A heat map showing the sum of peptide abundances per amino acid from all of the peptides generated at different time points for the different gastric digestion scenarios from BCAS. The scenarios are infant, INFG (early phase) and PRT (late phase). Epitopes from Figure 24 are also overlaid on the sequence.

Generally, the N-terminal end of the protein was more susceptible to hydrolysis than the C-terminal end. The progress in protein hydrolysis is clear as in the early stages of the infant model, there are significant parts of the molecule where no peptides are found. For comparison, the known allergenic epitopes of BCAS are shown in Figure 24 taken from Cerecedo et al.[20] In order to provide a stronger link to the immunological data, we have taken a similar approach to that used to generate Figure 23 but focussed on peptides produced in the early phase intestinal stage with the known epitopes from Figure 24 overlaid. The data shown in Figure 25 indicates a number of regions (~3) where the epitopes correlate with regions with a high density of peptides and one region where no peptides are seen. This data seems to be more discriminatory than the equivalent data from the gastric phase in Figure 23.

Beta Casein AA sequence

```
          10        20        30        40        50
RELEELNVPGEIVESLSSSEESITRINKKIEKFQSEEQQQTEDELQDKIH
          60        70        80        90        100
PFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSKVKE
          110       120       130       140       150
AMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQP
          160       170       180       190       200
LPPTVMFPPQSVLSLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGP
          210
VRGPFPIIV
```

Figure 24: The primary amino acid sequence of BCAS evaluated with IgE-binding epitopes identified in the study of Cerecedo et al. are highlighted in grey while those reported previously are underlined.

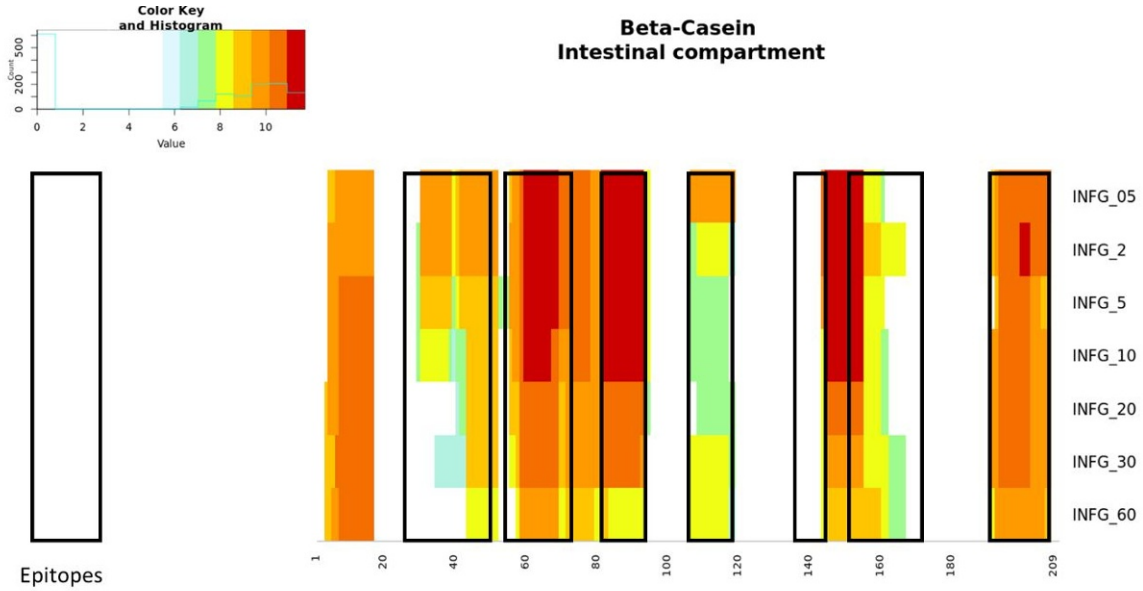


Figure 25: A heat map showing the sum of peptide abundances per amino acid from all of the peptides generated at different time points for the early phase intestinal phase digestion of BCAS. Epitopes from Figure 24 are also overlaid on the sequence.

In summary, 313 peptides were identified in the gastric phase after in vitro digestion of BCAS by the 3 models and 163 in the intestinal phase. One hundred and five peptides were mainly specific of the early phase model whereas 62 peptides were mainly seen in late phase model. When comparing the

heat map to the location of the known epitopes there is a partial correlation in Figures 23 and 25 with the areas of high density but this is not clear-cut.

BLG

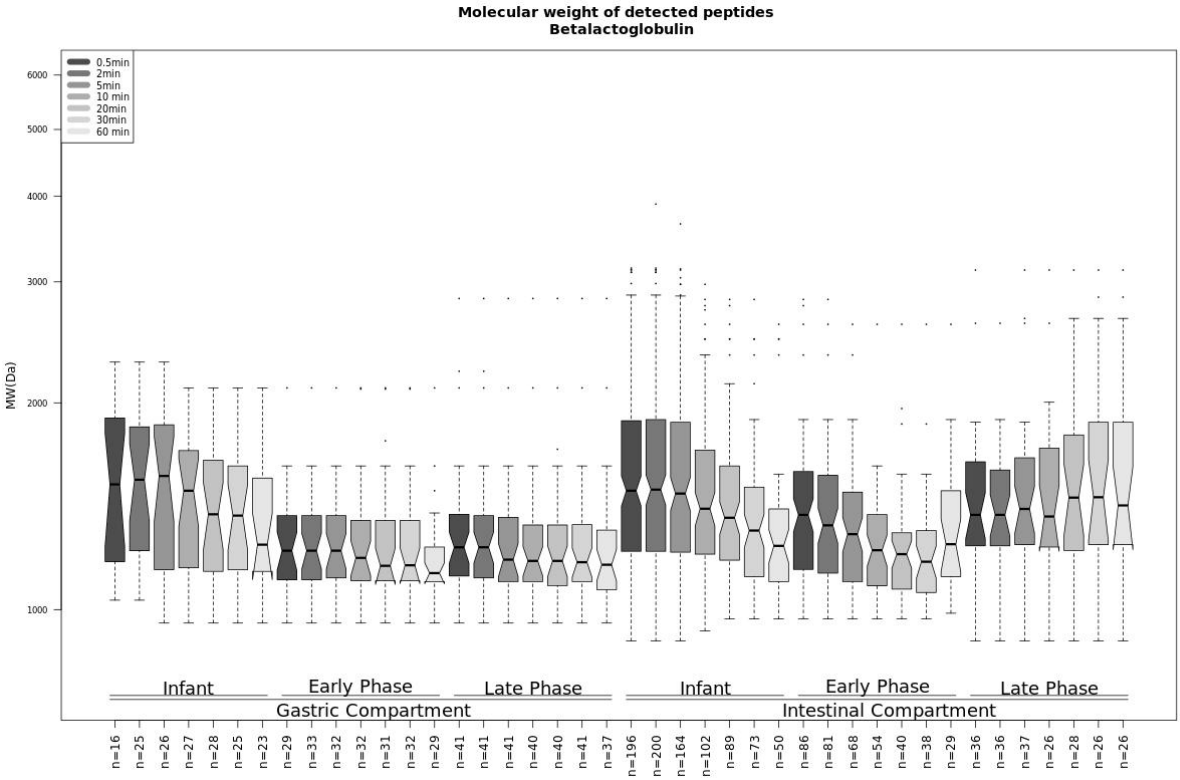


Figure 26: A box plot showing the molecular weight of peptides from BLG after gastric and intestinal digestion.

The data shown in Figure 26 indicate small peptides present in the gastric phase from BLG much smaller than those observed from BCAS. Even though BLG seems to be largely unaffected by proteolysis during the gastric phase in the three scenarios of digestion (Figures 1 and 3a), LC-MS demonstrates that a large number of peptides (72) can still be detected. It is probable that only a small proportion of BLG is hydrolysed by pepsin during the gastric phase (approximately 20% hydrolysed in the late phase adult model as seen in Figure 3a), showing no difference in the band intensity of the intact protein on SDS-PAGE gels for the infant and early phase adult models. It should be born in mind that SDS-PAGE is only a semi-quantitative method.

For BLG, the median Mw values are quite similar in the gastric phase between the early and the late phase model with median values decreasing slightly over time. Only the infant model led to peptides of higher Mw. In the late phase model, pepsin was able to release rather large peptides of 2800 Da whereas the highest Mw observed with the early phase model was around 2100 Da. In the intestinal phase, peptides Mw tended to be higher than in the gastric phase. This is certainly because a larger proportion of the protein was hydrolysed leading to more diversity in the peptides released. With the early phase model, peptide Mw tended to decrease over digestion time whereas it was the opposite with the late phase model.

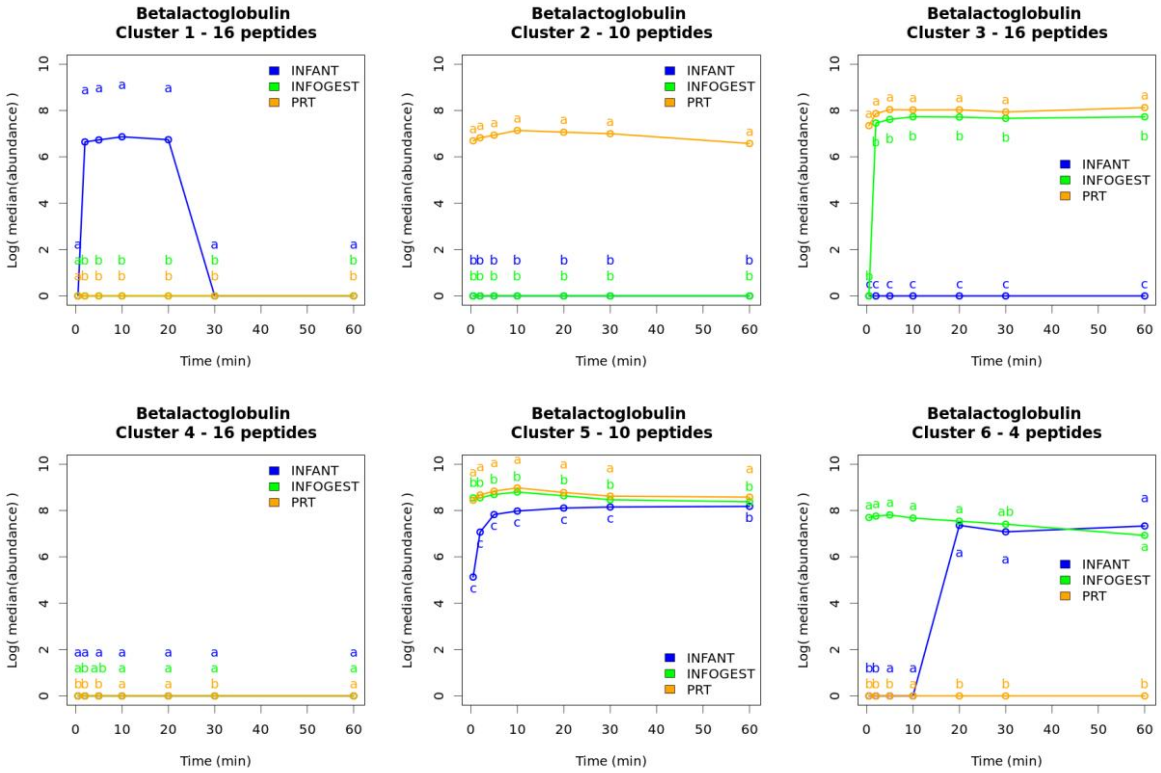


Figure 27: Kinetic plots of area under the curve for the 6 peptide clusters of BLG through digestion in the infant (blue), early phase (green) and late phase (yellow)

The hierarchical clustering made on BLG peptides led to the identification of 6 different clusters. The most interesting ones appear to be Clusters 1, 2 and 6. **Cluster 1** (16 peptides) shows a high abundance of these peptides only in the Infant model and only between 2 min and 20 min gastric digestion. Peptides of **Cluster 2** (10 peptides) are only highly abundant in the late phase model and barely present in the 2 other models of digestion. Finally, the 4 peptides of **Cluster 6** are highly abundant in the early phase model but not in the late phase one whereas for the infant model, their abundance increases after 10 min of digestion. In the heat map shown in Figure 28 below, the 10 peptides in **Cluster 2** are mapped onto the primary sequence.

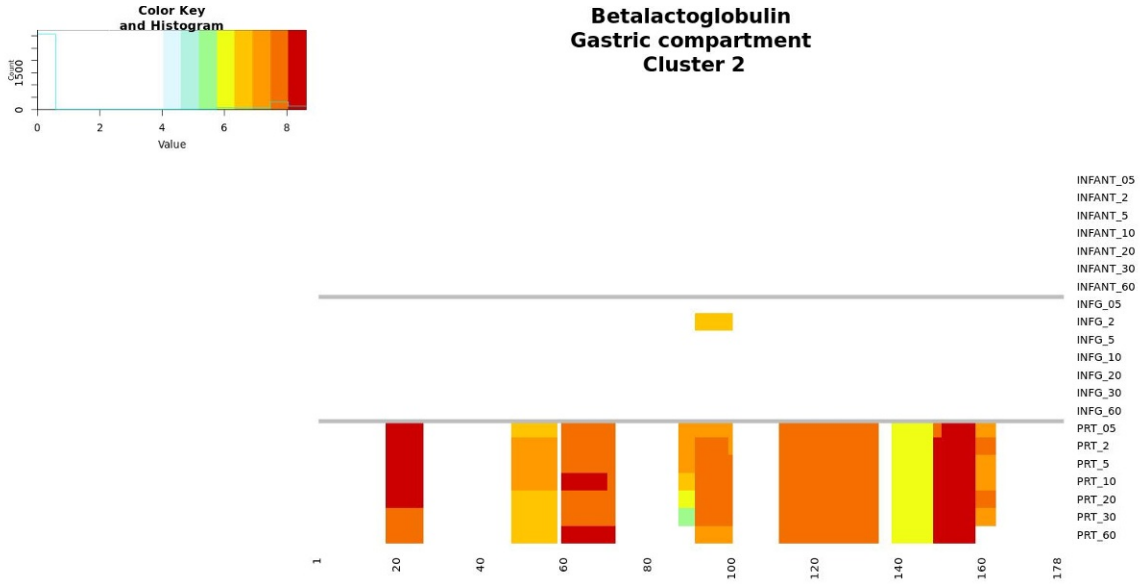


Figure 28: A heat map showing the sum of peptide abundancies per amino acid from **Cluster 2** peptides generated at different time points for the different gastric digestion scenarios from BLG. The scenarios are infant, INFG (early phase) and PRT (late phase).

For comparison, the total peptides identified from BLG are overlaid on the primary sequence in Figure 29. As one might expect, there are significant differences between the two figures as the BLG is quite resistant to hydrolysis. In both cases, the differences along the sequence are generated by either the production of peptides that are smaller than nine amino acids or larger than the detection limit of the MS. In this case, most likely the latter. Additionally, disulphide bridges were not reduced (60-160, 106-121 and 106-119) and bridged peptides could not be identified with the method used.

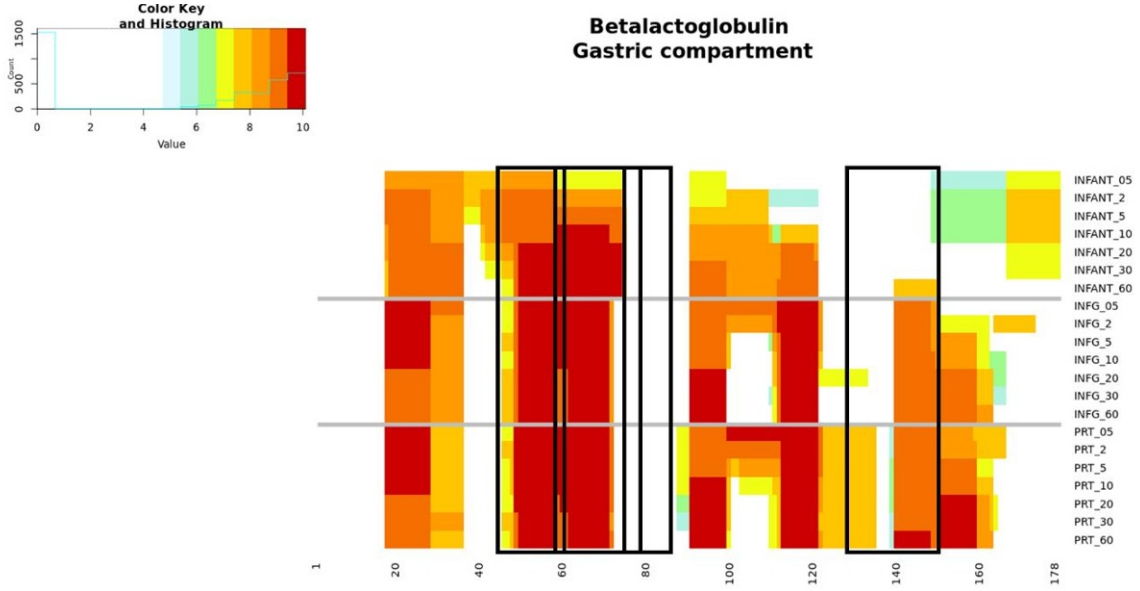


Figure 29: A heat map showing the sum of peptide abundances per amino acid from all of the peptides generated at different time points for the different gastric digestion scenarios from BLG. The scenarios are infant, INFG (early phase) and PRT (late phase). Epitopes from Figure 30 are also overlaid on the sequence.

There are clearly some very resistant sections of the molecule where peptides cannot be generated. These are around position 80 and the N and C-terminal ends of the protein. The other feature of the heat map to note is that that there are parts of the BLG molecule that increase and then decrease in intensity showing the sequence coverage does not just monotonically increase as digestion continues. For comparison, the known allergenic epitopes of BLG are shown in Figure 30 taken from Cerecedo et al.[20] and are overlaid in figures 29 and 31. In order to provide a stronger link to the immunological data, we have taken a similar approach to that used to generate Figure 29 but focussed on peptides produced in the early phase intestinal. The data shown in Figure 29 shows the location of the small number of peptides that were produced during the gastric phase and one might expect the blank regions to correlate with the known epitopes. In contrast, it might be expected that Figure 31 would indicate a number of regions where the epitopes correlate with regions with a high density of peptides as other regions were likely completely hydrolysed. In fact neither correlation is clear cut.

Beta Lactoglobulin AA sequence

```
          10        20        30        40        50
LIVTQTMKGLDIQKVAGTWYSLAMAASDISLLDAQSAPLRVYVEELKPTP
          60        70        80        90        100
EGDLEILLQKWENDECAQKKIIAEKTKIPAVFKIDALNENKVLVLDTDYK
          110       120       130       140       150
KYLLFCMENSAEPEQSLVCQCLVRTPEVDDEALEKFDKALKALPMHIRLS
          160
FNPTQLEEQCHI
```

Figure 30: The primary amino acid sequence of BLG evaluated with IgE-binding epitopes identified in the study of Cerecedo et al. are highlighted in grey while those reported previously are underlined.

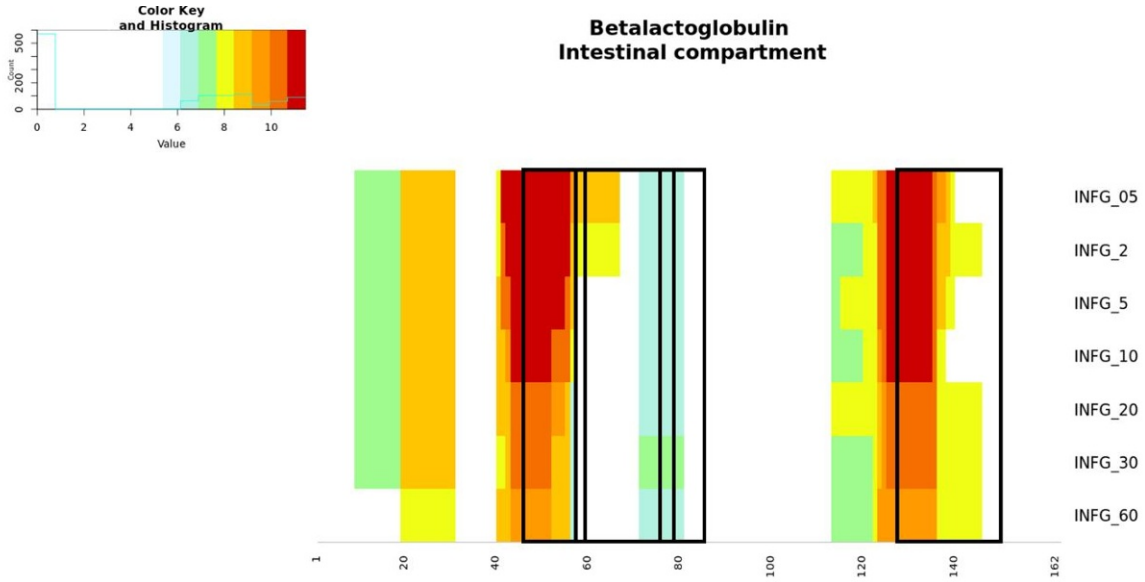


Figure 31: A heat map showing the sum of peptide abundances per amino acid from all of the peptides generated at different time points for the early phase intestinal phase digestion of BLG. Epitopes from Figure 30 are also overlaid on the sequence.

In summary, 72 peptides were identified in the gastric phase after in vitro digestion of BLG by the 3 models and 138 in the intestinal phase. Four peptides were mainly specific of the early phase model whereas 10 peptides were mainly seen in late phase model. As with the BCAS there is some correlation

between known epitopes and regions of high density in the heat map from Figures 29 and 31 but again this is not clear cut.

3.2.2 Hen's egg proteins

Lys

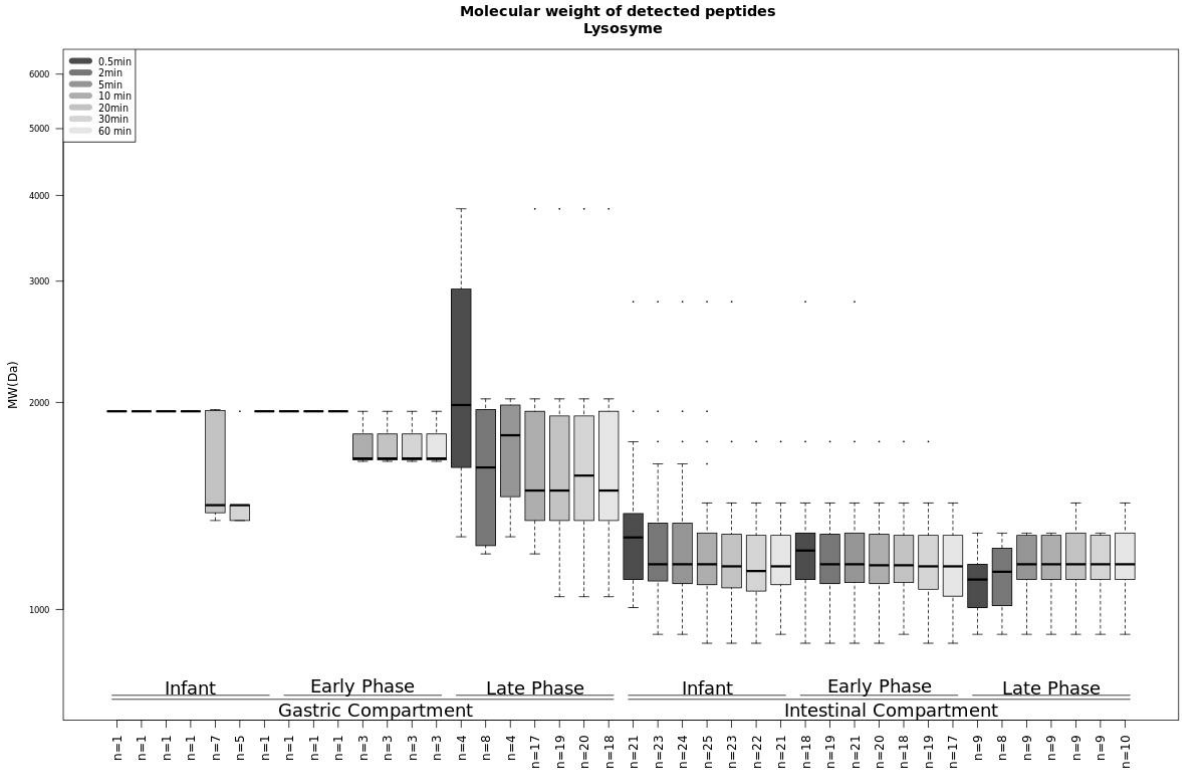


Figure 32: A box plot showing the molecular weight of peptides from LYS after gastric and intestinal digestion.

In vitro digestion of lysozyme led to the appearance of only 16 peptides during the gastric phase and 20 peptides in the intestinal phase. These results confirm those previously published showing a high resistance of lysozyme to in vitro digestion [21]. It has been shown that the resistance of Lys to pepsin is due to its rigid structure given by the disulphide bridges and is lower at highly acidic pH. This is therefore not surprising to find more peptides with the late phase than with the other models. Furthermore, it has been shown that when cleaved by pepsin, Lys leads to the formation of large peptides of molecular weight higher than 4 kDa[21] i.e. not detectable by the LC-MS protocol we used.

In the intestinal phase, Lys has been shown to precipitate in the presence of bile salts making it quite resistant to proteolysis.

Since the number of peptides clearly identified is rather low, the statistical analysis of the dataset does not bring much relevant information. The dataset needs to be completed by an exhaustive manual assessment of fragments of Mw higher than 4 kDa.

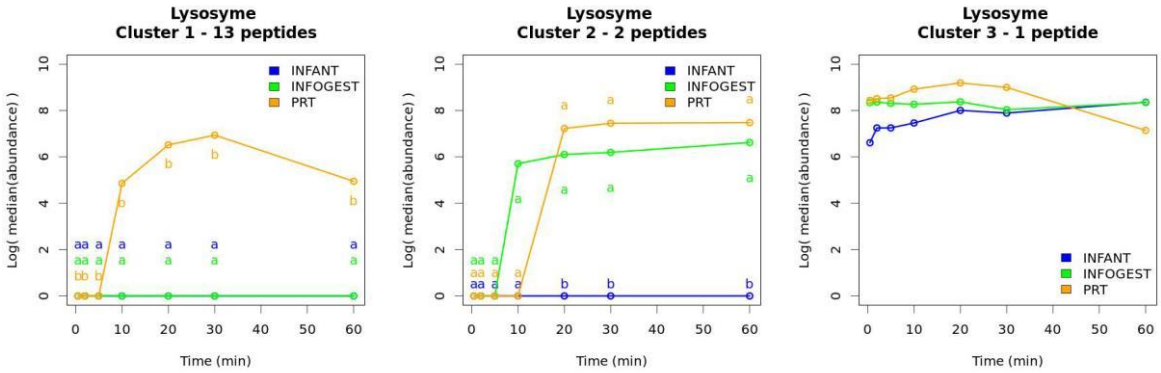


Figure 33: Kinetic plots of area under the curve for the 3 peptide clusters of LYS through digestion in the infant (blue), early phase (green) and late phase (yellow)

Among the different clusters observed for LYS, the most interesting one is **Cluster 1** (13 peptides) that shows a very low abundance of the peptides for the infant and early phase models but a much higher abundance with the late phase model starting after 10 min digestion. None of the clusters contain sufficient peptides to be mapped individually onto the primary sequence.

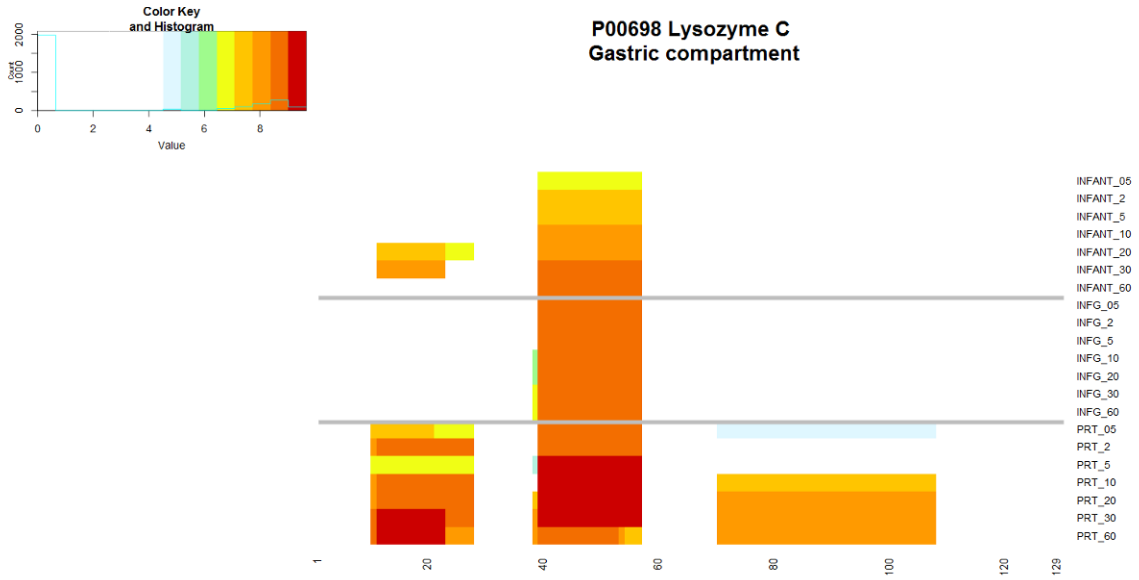


Figure 34: A heat map showing the sum of peptide abundances per amino acid from all of the peptides generated at different time points for the different gastric digestion scenarios from LYS. The scenarios are infant, INFG (early phase) and PRT (late phase).

In summary, only 16 peptides were identified in the gastric phase after in vitro digestion of LYS by the 3 models and 20 in the intestinal phase. Thirteen peptides were mainly specific of the late phase model.

Ova

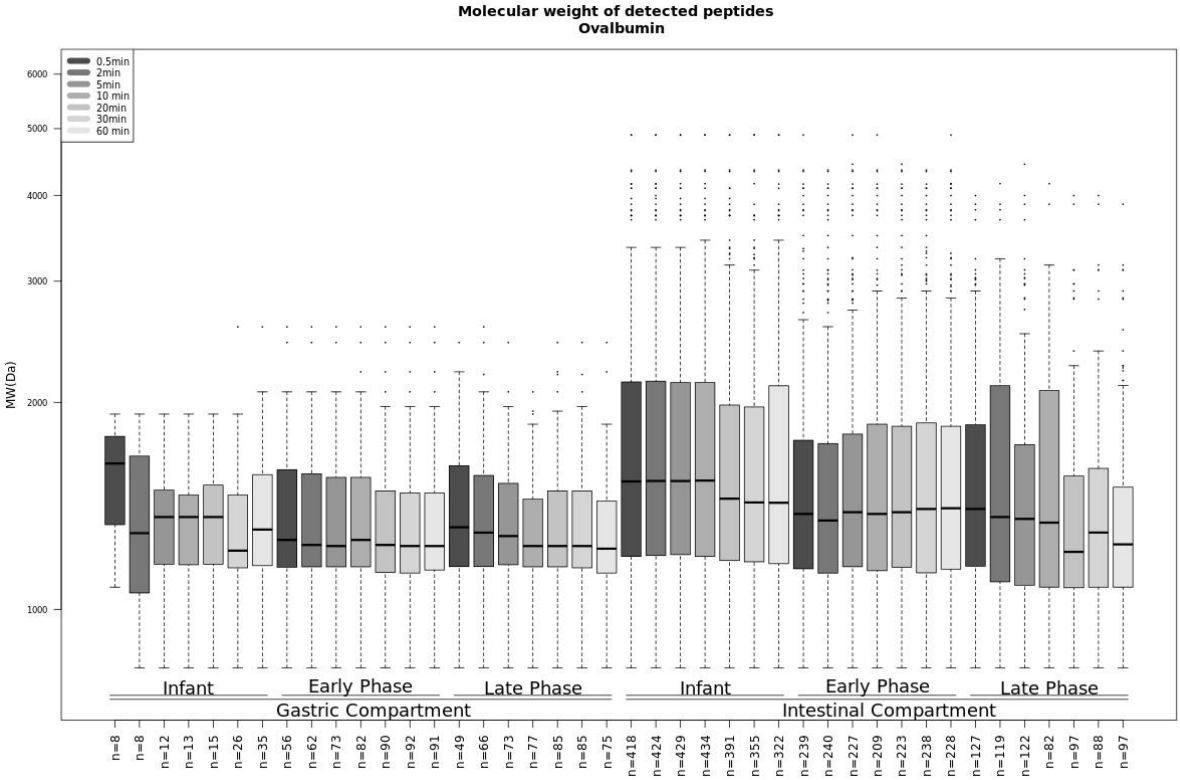
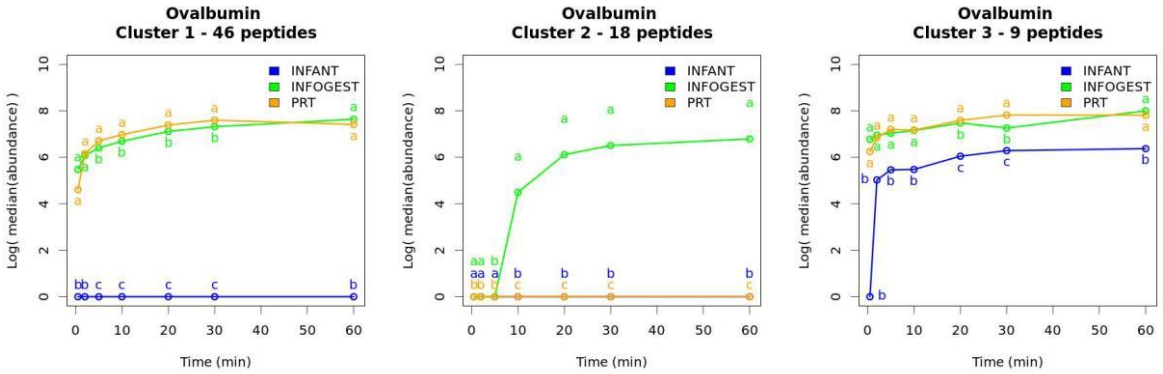


Figure 35: A box plot showing the molecular weight of peptides from OVA after gastric and intestinal digestion.



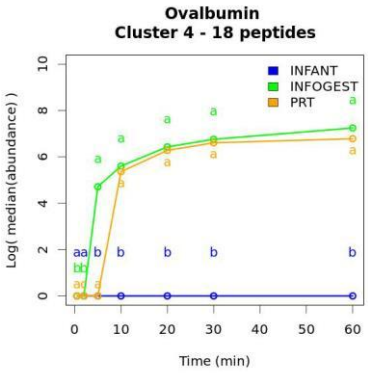


Figure 36: Kinetic plots of area under the curve for the 4 peptide clusters of OVA through digestion in the infant (blue), early phase (green) and late phase (yellow)

Gastric peptides were regrouped into 4 clusters. Among those, **Cluster 2** (18 peptides) gathers peptides highly abundant in the early phase model but not with the 2 other models. In the heat map shown in the Figure below, the 46 peptides from **Cluster 1** the largest cluster, are mapped onto the primary sequence.

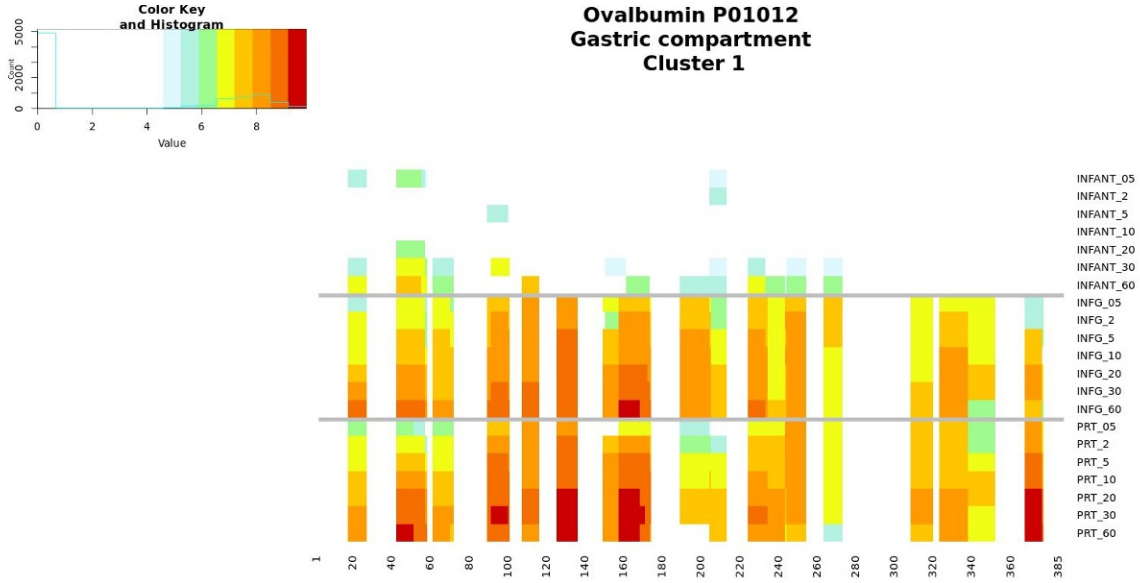


Figure 37: A heat map showing the sum of peptide abundances per amino acid from **Cluster 1** peptides generated at different time points for the different gastric digestion scenarios from OVA. The scenarios are infant, INFG (early phase) and PRT (late phase).

For comparison, the total peptides identified from OVA are overlaid on the primary sequence in Figure 38. As one might expect, there are differences between the two figures but the general trend is quite similar. In both cases, the differences along the sequence are generated by either the production of peptides that are smaller than nine amino acids or larger than the detection limit of the MS, in this case most likely the latter. For comparison, the known allergenic epitopes of OVA are shown in Figure 39 taken from Mine et al. [22]. In order to provide a stronger link to the immunological data, we have taken a similar approach to that used to generate Figure 38 but focussed on peptides produced in the early

phase intestinal stage with the known epitopes from Figure 39 overlaid. The data shown in Figure 40 indicates a number of regions where the epitopes correlate with regions with a high density of peptides.

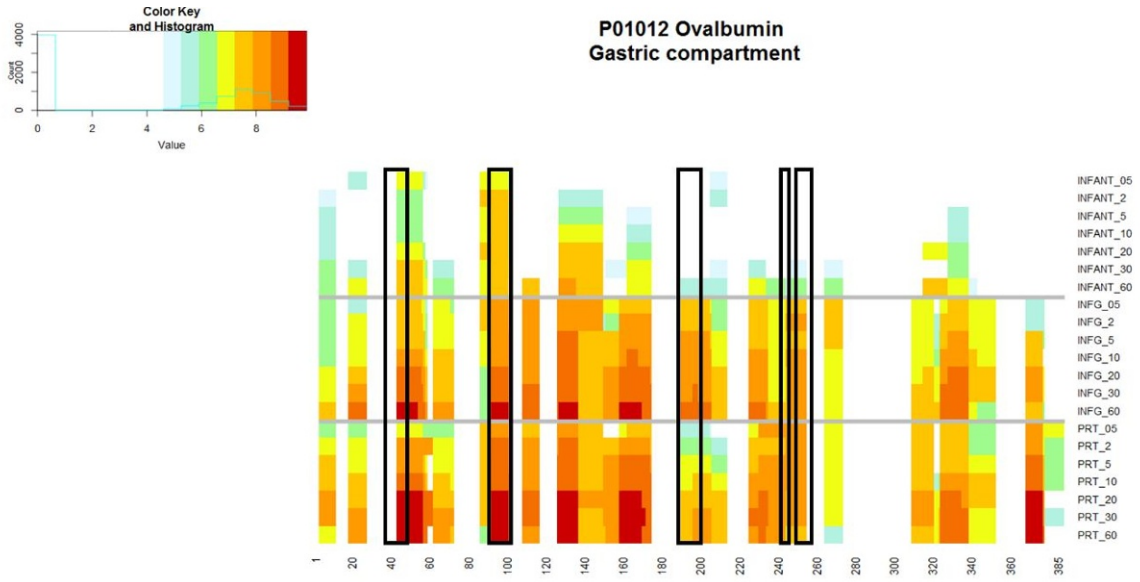


Figure 38: A heat map showing the sum of peptide abundance per amino acid from all of the peptides generated at different time points for the different gastric digestion scenarios from OVA. The scenarios are infant, INFG (early phase) and PRT (late phase). Epitopes from Figure 39 are also overlaid on the sequence

```
  1 GSIGAASMEF CFDVFKELKV HHANENIFYC PIAIMSALAM VYLGAKDSTR
 51 TQINKVVRFD KLPGFGDSIE AQCGTSVNVH SSLRDILNQI TKPNDVYSFS
101 LASRLYAEER YPILPEYLQC VKELYRGGLE PINFQTAADQ ARELINSWVE
151 SQTNGIIRNV LQPSSVDSQT AMVLVNAIVF KGLWEKTFKD EDTQAMPFRV
201 TEQESKPVQM MYQIGLFRVA SMASEKMKIL ELPFASGTMS MLVLLPDEVS
251 GLEQLESIIN FEKLTEWTSS NVMEERKIKV YLPRMKMEEK YNLTSVLMAM
301 GITDVFSSSA NLSGISSAES LKISQAVHAA HAEINEAGTE VVGSAEAGVD
351 AASVSEEFRA DHPFLFCIKH IATNAVLFFG RCVSP
```

Figure 39: The primary amino acid sequence of OVA evaluated with IgE-binding epitopes identified in the study of Mine et al. are highlighted in grey while those reported previously are underlined.

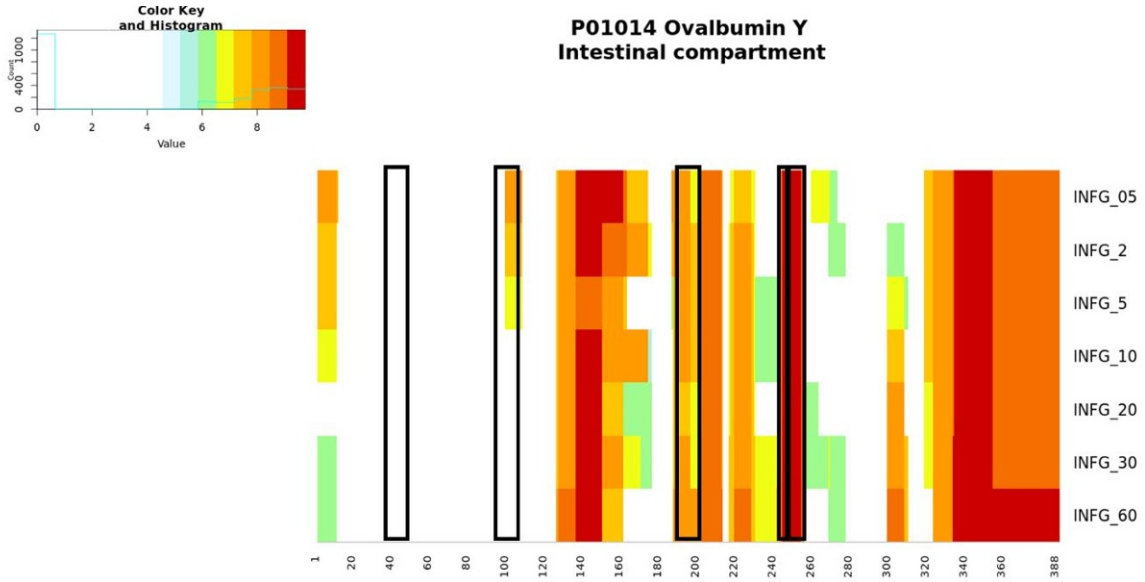


Figure 40: A heat map showing the sum of peptide abundances per amino acid from all of the peptides generated at different time points for the early phase intestinal phase digestion of OVA. Epitopes from Figure 39 are also overlaid on the sequence

In summary, 91 peptides were identified in the gastric phase after in vitro digestion of OVA by the 3 models and 434 in the intestinal phase. Eighteen peptides were mainly specific of the early phase model. As with the milk proteins, there is a good correlation between the some of the more intense regions of the heat map and the known epitopes in Figures 38 and 40 but again this is not clear-cut.

3.2.3 Plant proteins

KTI

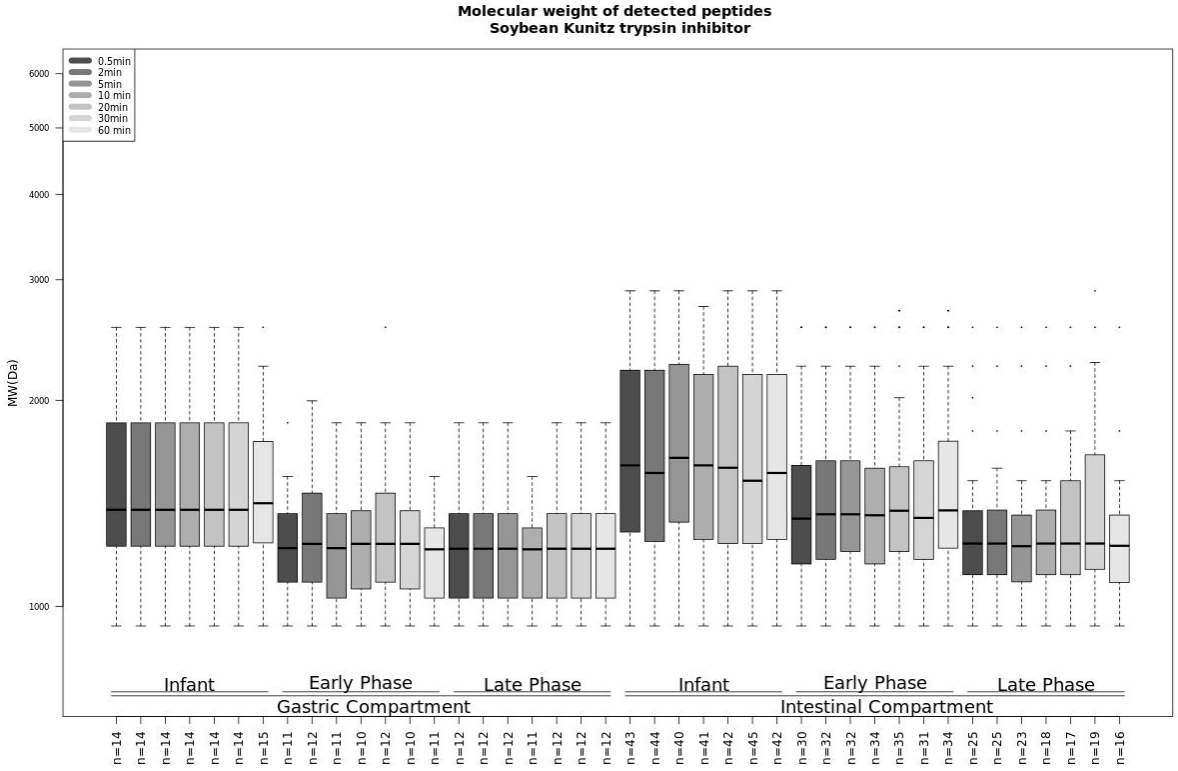


Figure 41: A box plot showing the molecular weight of peptides from KTI after gastric and intestinal digestion.

Peptides generated from KTI in the gastric phase with the infant model exhibited higher Mw than those obtained with the 2 other models. No major evolution in size was observed with time during the gastric phase. In the intestinal phase, again, peptides generated by the infant model were larger than those released with the early phase model that were themselves bigger than those of the late phase model. Like in the gastric phase, no clear evolution of the peptide size was seen throughout time. Only 20 peptides were detected in the gastric phase whereas 53 were identified in the intestinal phase.

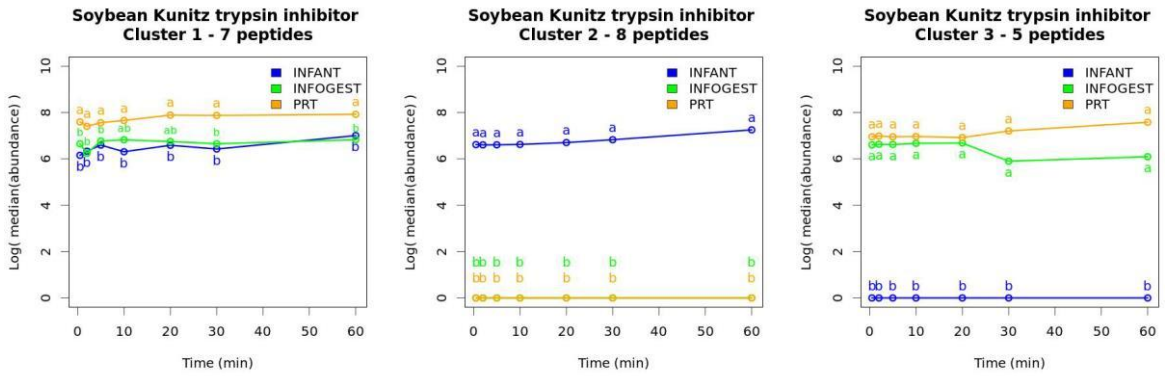


Figure 42: Kinetic plots of area under the curve for the 3 peptide clusters of KTI through digestion in the infant (blue), early phase (green) and late phase (yellow)

Figure 42 showed 3 clusters were obtained when analysing the gastric peptides. **Cluster 2** (8 peptides) showed higher abundances in the infant than in the 2 other models whereas it was the opposite for **Cluster 3** (5 peptides) that had low abundances with the infant model but high ones for the early and late phase models. There were too few peptides in any of the clusters for them to be usefully mapped onto the primary sequence but Figure 43 shows all of the peptides generated through the gastric phase mapped onto the sequence. This shows that the majority of hydrolysis was at the N-terminal end of the protein.

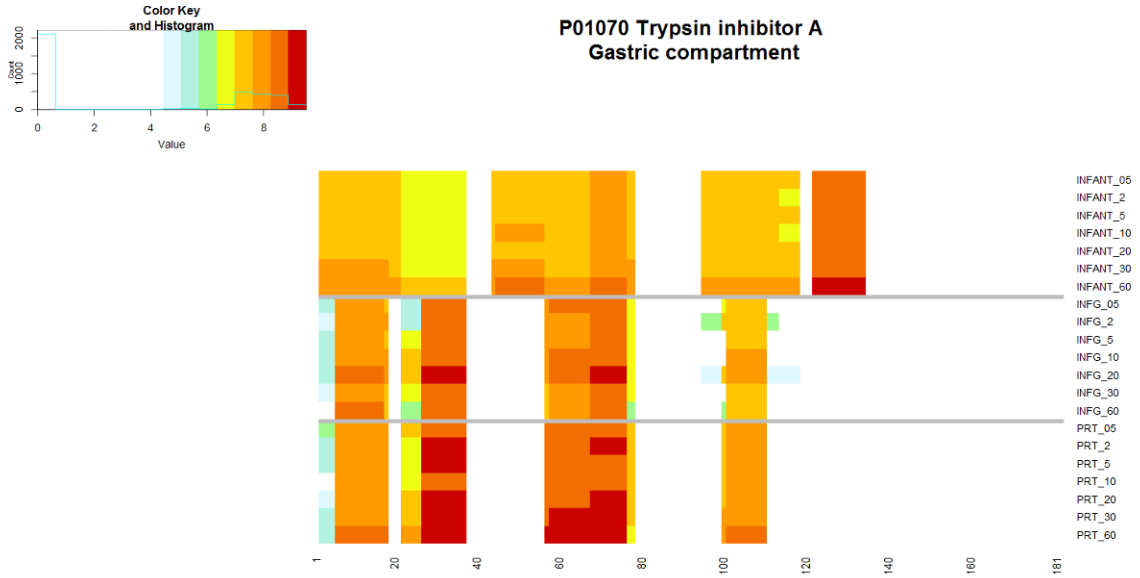


Figure 43: A heat map showing the sum of peptide abundances per amino acid from all of the peptides generated at different time points for the different gastric digestion scenarios from KTI. The scenarios are infant, INFG (early phase) and PRT (late phase).

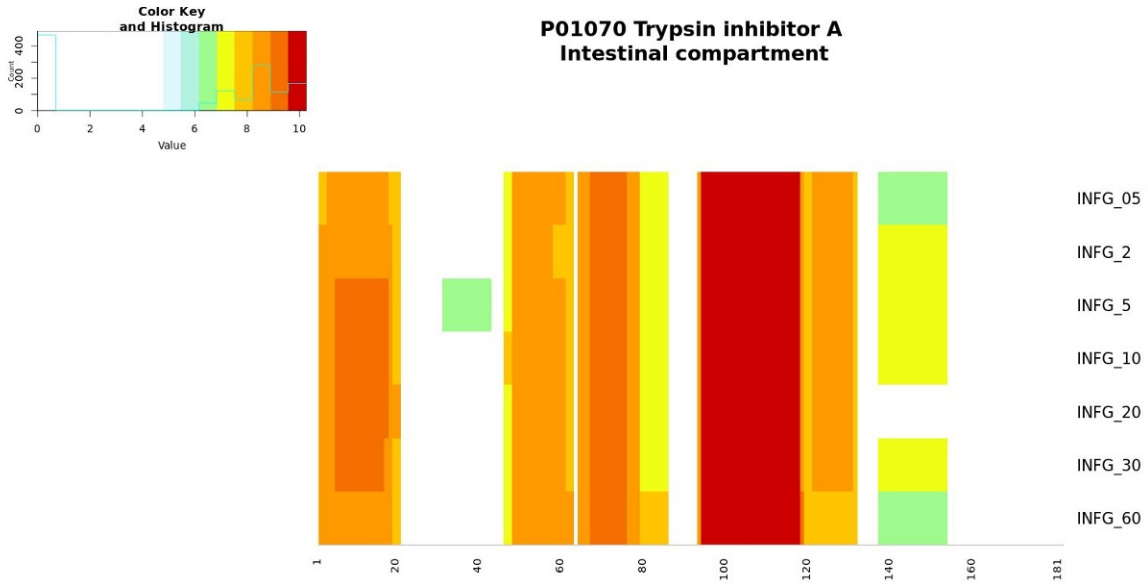


Figure 44: A heat map showing the sum of peptide abundances per amino acid from all of the peptides generated at different time points for the early phase intestinal digestion from KTI.

In summary, only 20 peptides were identified in the gastric phase after in vitro digestion of KTI by the 3 models and 53 in the intestinal phase. No specific differences were seen between the early and late phase model.

LIP

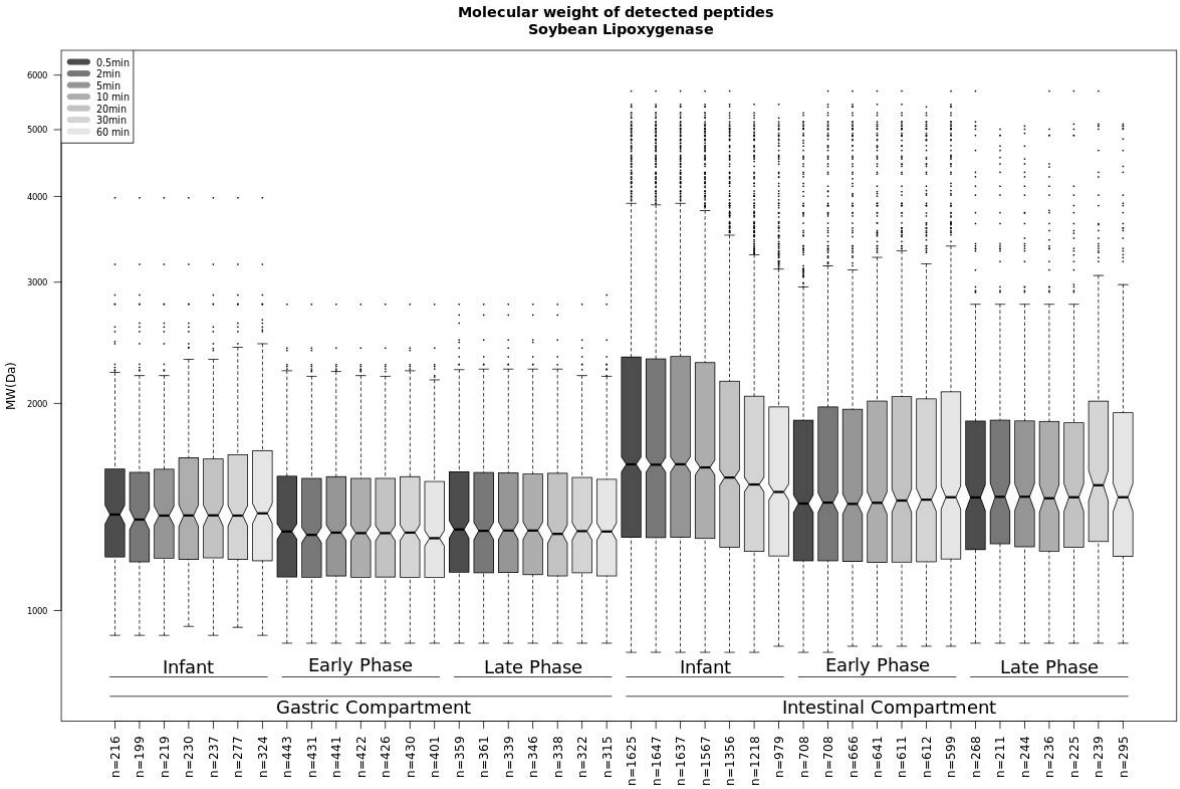


Figure 45: A box plot showing the molecular weight of peptides from LIP after gastric and intestinal digestion.

LIP *in vitro* digestion released a very high number of peptides; whereas 642 were identified in the gastric phase, 1721 unique peptides sequences were released in the intestinal phase. The median Mw of the gastric peptides was relatively low (1300-1400 Da) and increased significantly in the intestinal phase. The Mw tend to decrease with time during the intestinal phase infant model.

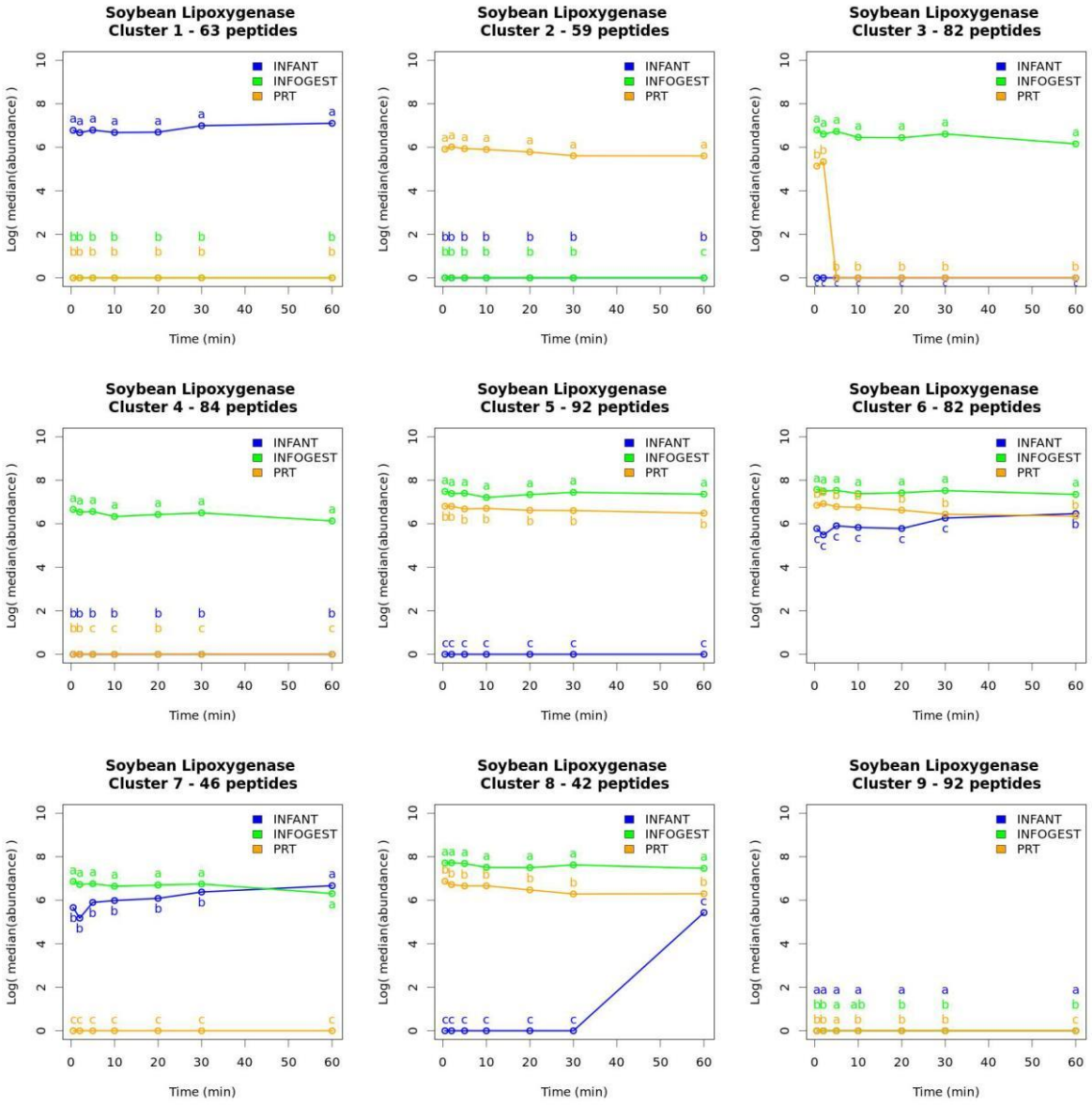


Figure 46: Kinetic plots of area under the curve for the 9 peptide clusters of LIP through digestion in the infant (blue), early phase (green) and late phase (yellow)

The 642 peptides identified during gastric digestion of LIP could be grouped into 9 clusters. Among those, Clusters 2, 3, 4 and 7 appear to be particularly interesting. **Cluster 2** (59 peptides) corresponds to peptides showing a high abundance in the late phase model but a very low one with the 2 other models. **Cluster 3** (82 peptides) is nearly the opposite since the peptides can mainly be seen with the early phase but in the late phase model during the first time points (after 5 min of digestion, they disappear with this model). **Cluster 4** (84 peptides) and **7** (46 peptides) show major differences between the early and late phase model: whereas these peptides are highly abundant in the early phase model, they exhibit very low abundances in the late phase model.

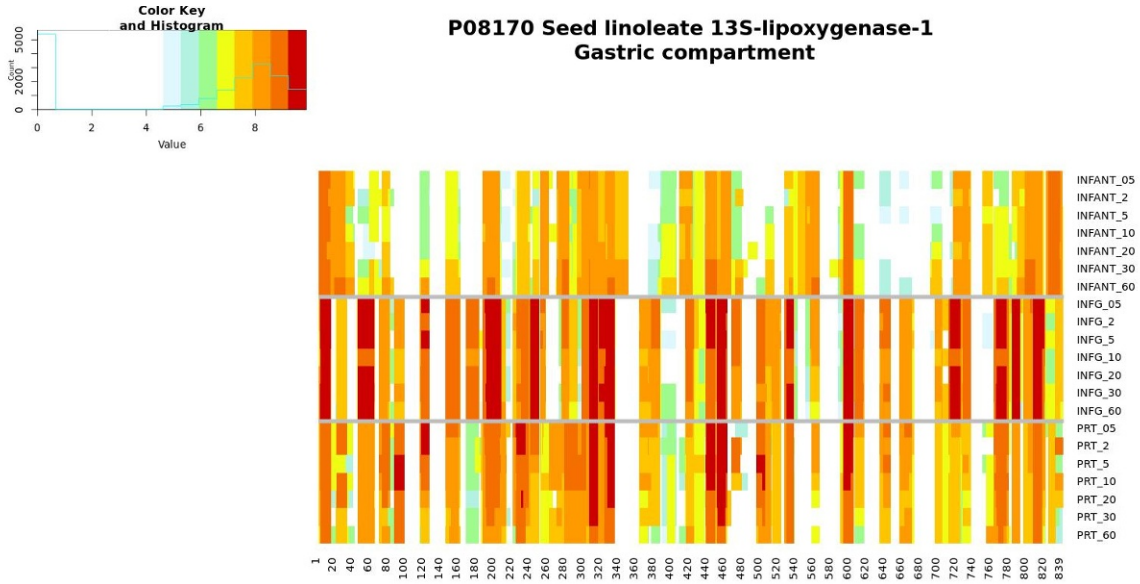


Figure 47: A heat map showing the sum of peptide abundances per amino acid from all of the peptides generated at different time points for the different gastric digestion scenarios from LIP. The scenarios are infant, INFG (early phase) and PRT (late phase).

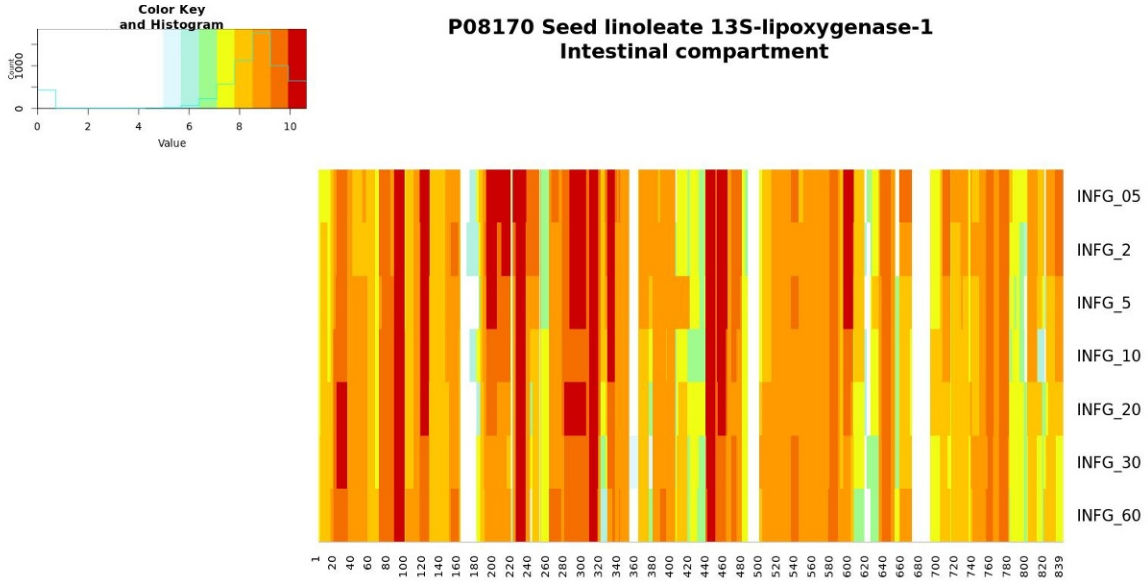


Figure 48: A heat map showing the sum of peptide abundances per amino acid from all of the peptides generated at different time points for the early phase intestinal phase from LIP.

In summary, 642 peptides were identified in the gastric phase after in vitro digestion of LIP by the 3 models and 1721 in the intestinal phase. 212 peptides were mainly specific of the early phase model whereas 59 peptides were mainly seen in late phase model.

ConA

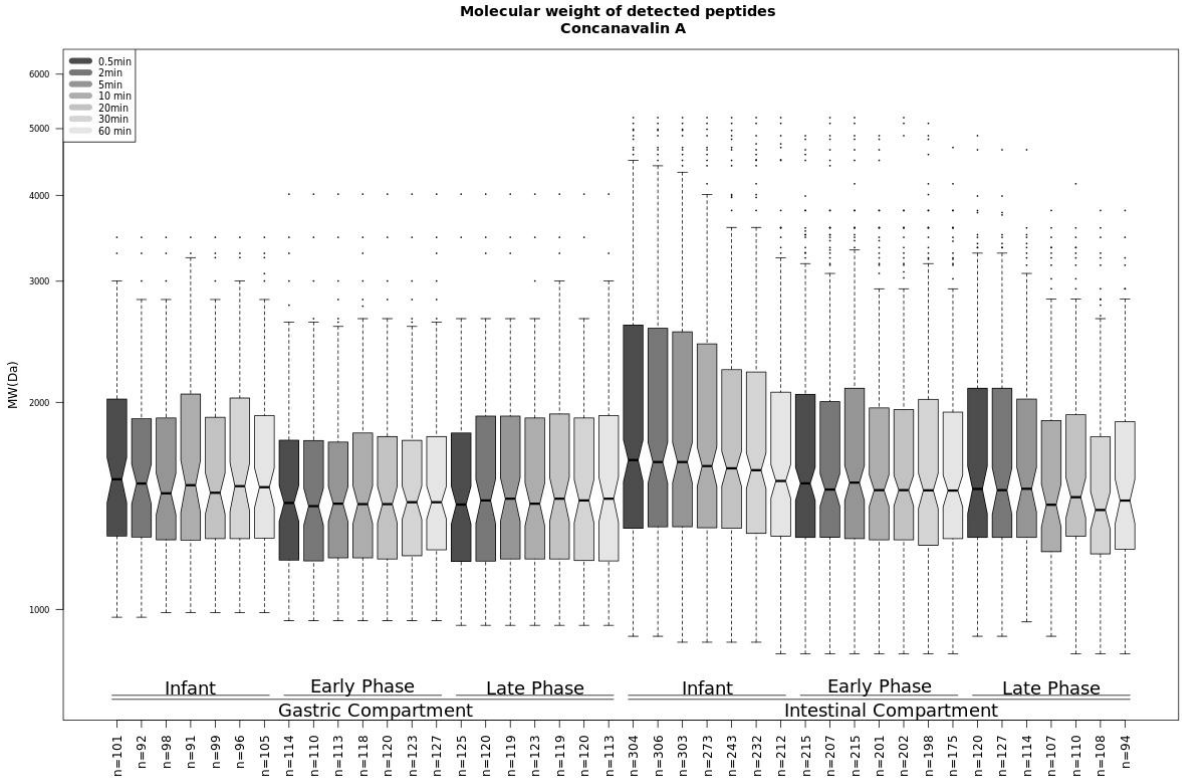


Figure 49: A box plot showing the molecular weight of peptides from ConA after gastric and intestinal digestion.

In vitro digestion of ConA with the 3 models of digestion led to the identification of 61 peptides in the gastric phase and 329 peptides in the intestinal phase. In the gastric phase, peptides released by the infant model were slightly longer than those obtained with the 2 other models. During digestion, no significant decrease on the median peptide Mw was observed in any model. In the intestinal phase, all the models showed a slight decrease in Mw with time.

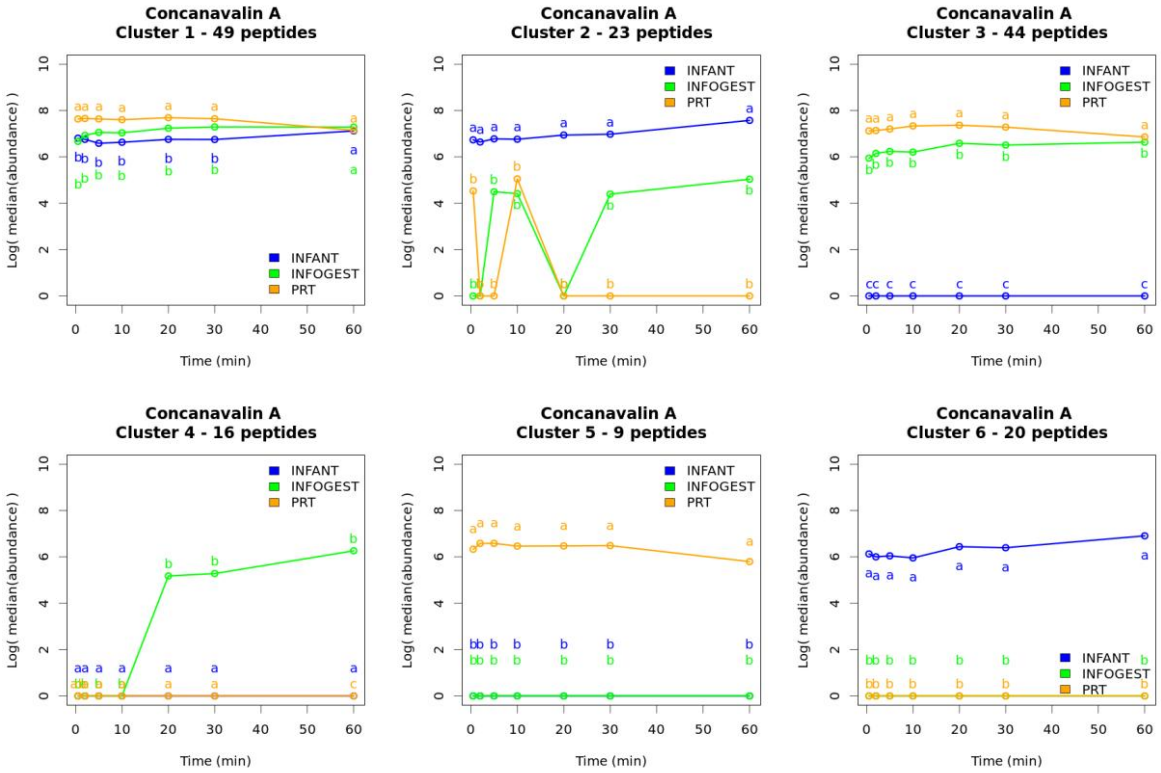


Figure 50: Kinetic plots of area under the curve for the 6 peptide clusters of ConA through digestion in the infant (blue), early phase (green) and late phase (yellow)

Clustering of the 161 peptides identified during the gastric phase led to the formation of 6 clusters. Among those, Clusters 4, 5, and 6 are particularly interesting. Clusters 4 (16 peptides) and 6 (20 peptides) correspond to peptides present in the early phase and the infant model respectively but not in the late phase one. On the contrary, cluster 5 (9 peptides) exhibit high abundances with the late phase model and low ones with the other models.

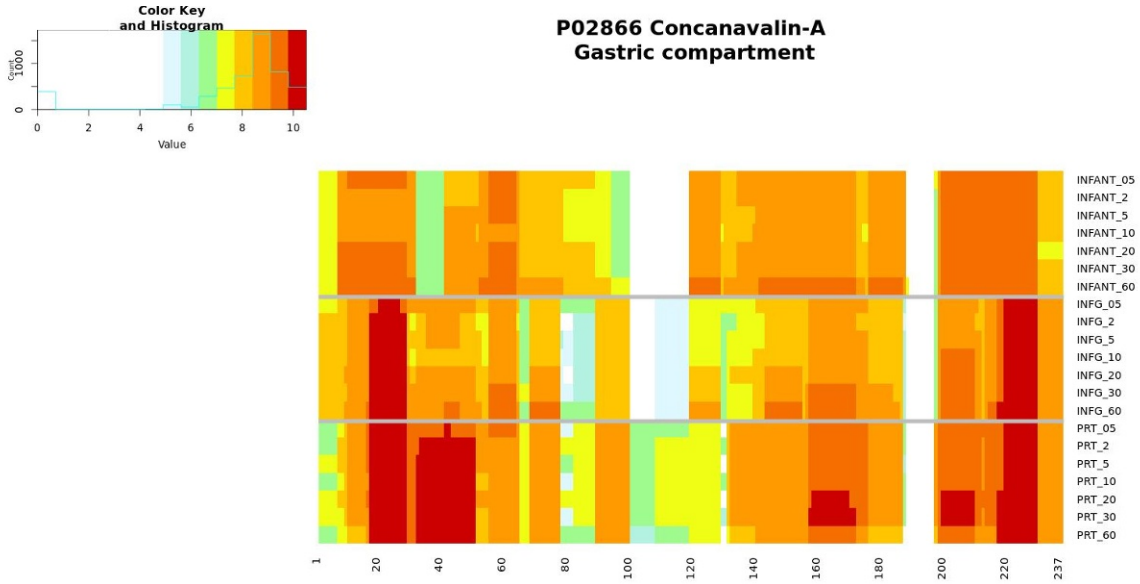


Figure 51: A heat map showing the sum of peptide abundances per amino acid from all of the peptides generated at different time points for the different gastric digestion scenarios from Con A. The scenarios are infant, INFG (early phase) and PRT (late phase).

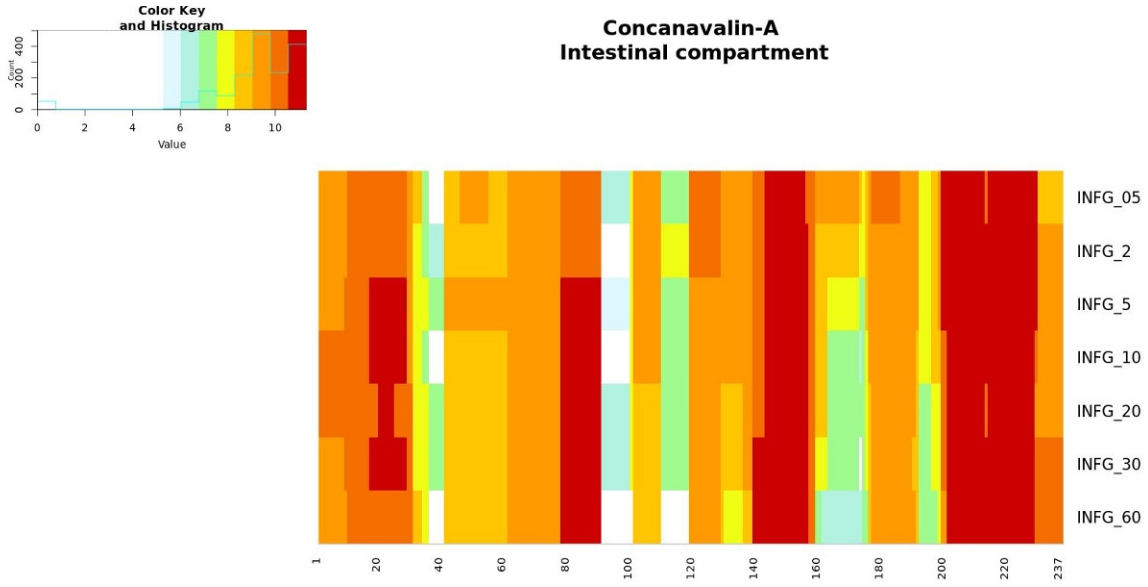


Figure 52: A heat map showing the sum of peptide abundances per amino acid from all of the peptides generated at different time points for the early phase intestinal digestion from Con A.

In summary, 161 peptides were identified in the gastric phase after in vitro digestion of ConA by the 3 models and 329 in the intestinal phase. Sixteen peptides were mainly specific of the early phase model whereas 9 peptides were mainly seen in late phase model.

AP

For Acid Phosphatase, only 16 peptides could be detected whereas hundreds of peptides were identified from "patatin" which is also known for its allergenicity. Only the peptides really corresponding to AP were taken into account during the data analysis but since their number is low, little information can be obtained.

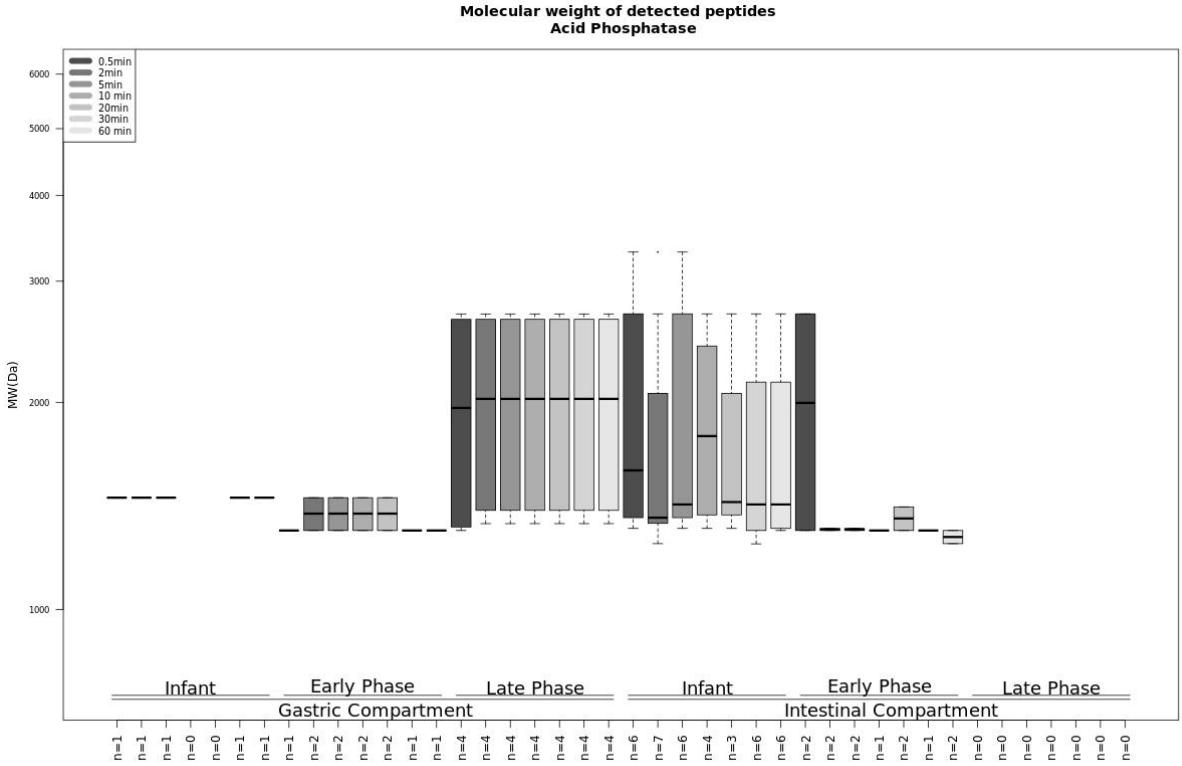


Figure 53: A box plot showing the molecular weight of peptides from AP after gastric and intestinal digestion.

Since only 5 peptides were identified in the gastric phase, little information can be obtained from this dataset and clustering was not interpreted. One can only mention that peptides observed with the late phase model were of higher Mw than that observed with the 2 other models. In the intestinal phase, no peptides was identified in the late phase model. So one can really wonder if the band seen around 45 kDa is really the monomeric form of AP or if it is rather patatin.

In summary, only 5 peptides were identified in the gastric phase after in vitro digestion of AP by the 3 models and 12 in the intestinal phase. The AP protein used was contaminated by patatin.

Ara h 1

In vitro digestion of Ara h 1 with the 3 digestion models led to the identification of 485 peptides in the gastric phase and 682 in the intestinal phase. The median Mw of peptides tend to decrease with time during the gastric phase with the 2 early and late phase models whereas it increases during the first 10 min of gastric digestion with the infant model. In the intestinal phase, the early phase model leads to lower Mw peptides than the 2 other models.

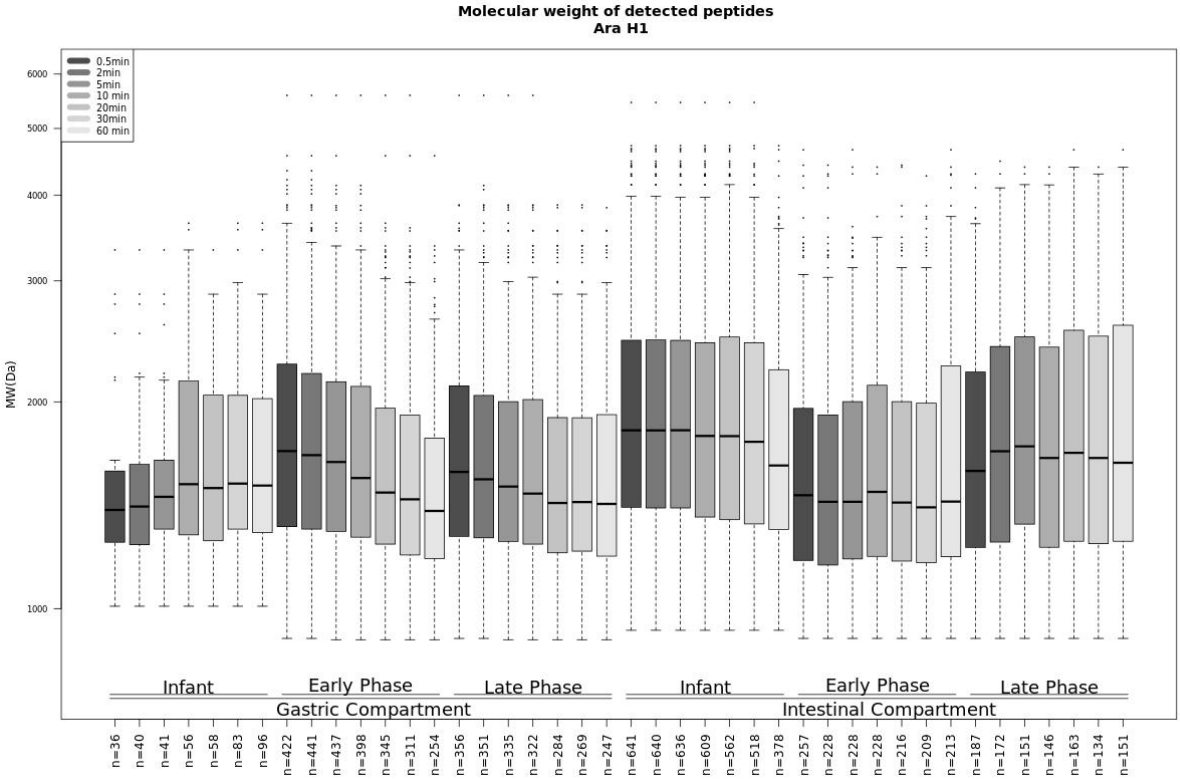


Figure 54: A box plot showing the molecular weight of peptides from Ara h 1 after gastric and intestinal digestion.

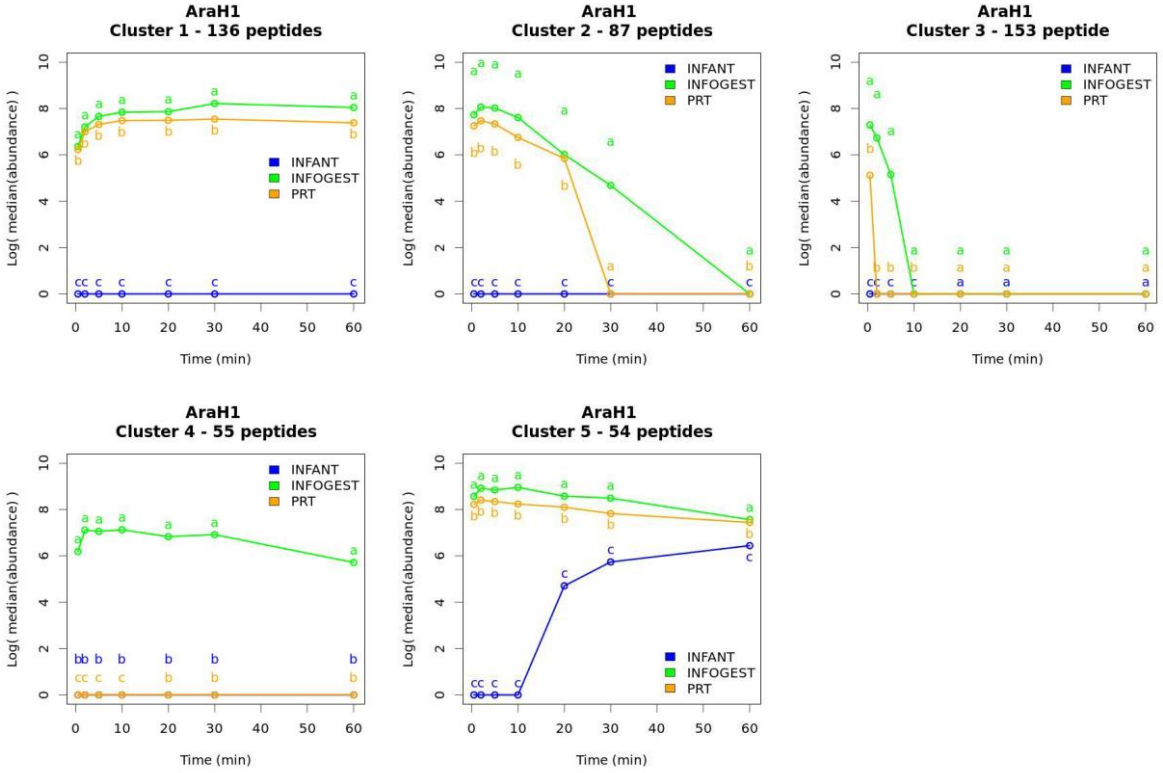


Figure 55: Kinetic plots of area under the curve for the 5 peptide clusters of Ara h 1 through digestion in the infant (blue), early phase (green) and late phase (yellow)

Five peptide Clusters could be clearly identified. Clusters 2, 3 and 4 seems particularly interesting. In **Clusters 2** (87 peptides) and **3** (153 peptides) peptides present at the beginning of the gastric phase disappear progressively during digestion, faster with the late phase model and with the early phase one. **Cluster 4** (55 peptides) corresponds to peptides highly abundant only in the early phase model. In the heat map shown in Figure 56 below, the 87 peptides in **Cluster 2** are mapped onto the primary sequence P43238 for Ara h 1.

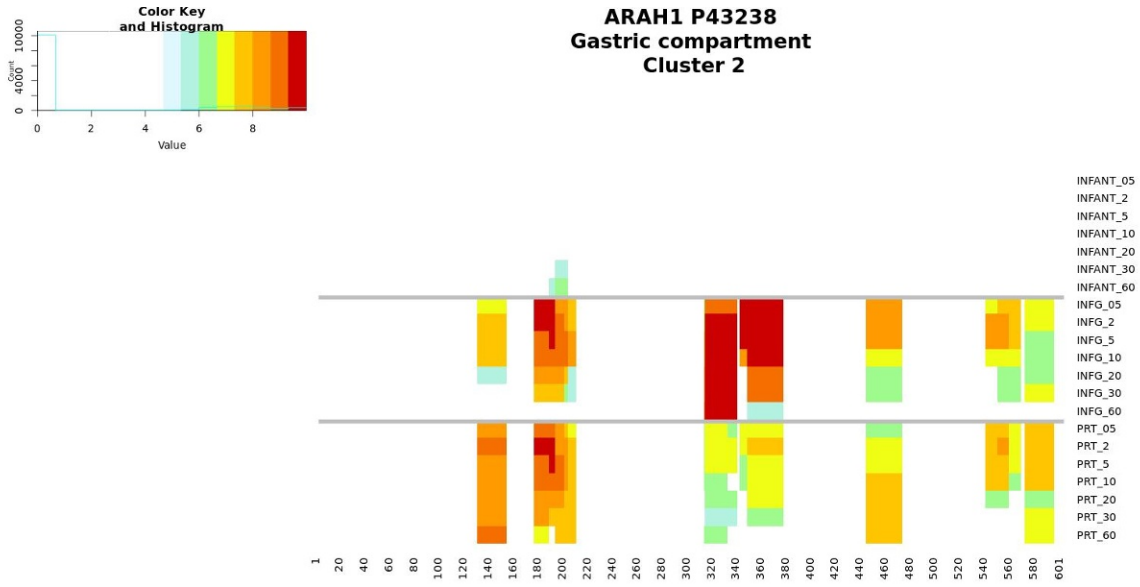


Figure 56: A heat map showing the sum of peptide abundances per amino acid from **Cluster 2** peptides generated at different time points for the different gastric digestion scenarios from Ara h 1. The scenarios are infant, INFG (early phase) and PRT (late phase).

For comparison, the total peptides identified from Ara h 1 are overlaid on the primary sequence in Figure 54. As one might expect, there are differences between the two figures but the general trend is quite similar. In general, and as highlighted in Figure 18, the extent of hydrolysis is high and the low density (white) regions are most likely due to extensive hydrolysis below the 9AA cut-off. For comparison, the known allergenic epitopes of Ara h 1 are shown in Figure 58 taken from Boxtel et al. [23]. In order to provide a stronger link to the immunological data, we have taken a similar approach to that used to generate Figure 59 but focussed on peptides produced in the early phase intestinal stage with the known epitopes from Figure 58 overlaid. The data shown in Figure 59 indicates a number of regions where the epitopes correlate with regions with a high density of peptides.

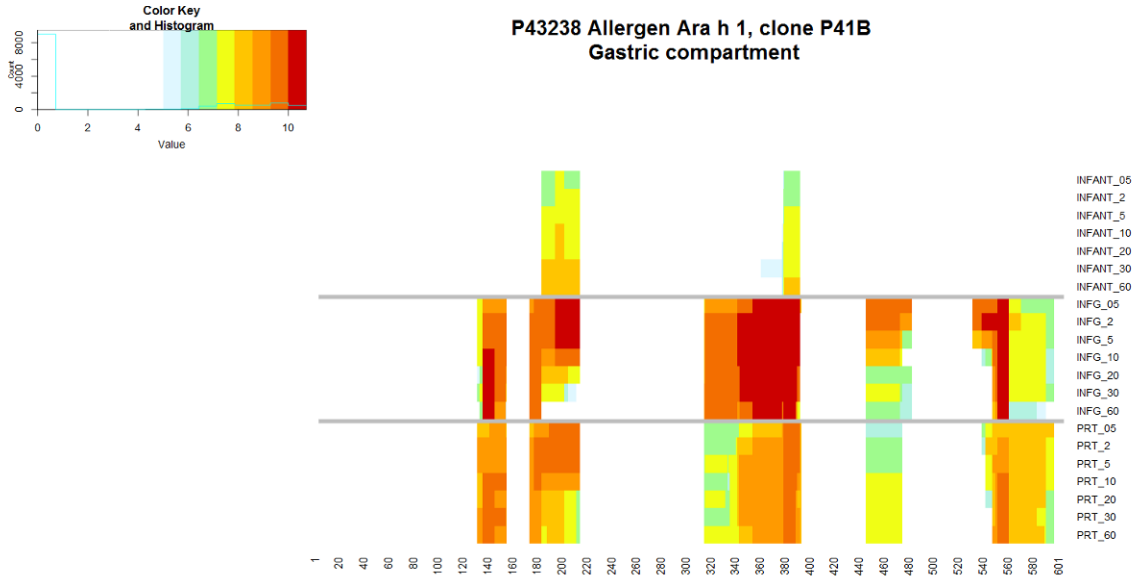


Figure 57: A heat map showing the sum of peptide abundancies per amino acid from all of the peptides generated at different time points for the different gastric digestion scenarios from Ara h 1. The scenarios are infant, INFG (early phase) and PRT (late phase).

a.a.		a.a.
1	RSPPGERTRGRQPGDYDDDRRQPRREEGGRWGPAGPRERE**R**EEDWRQPRE	50
51	DWRRPSHQQPRKIRPEGREGEQEWGTPGSHVRE**ETSRNNPFYF**PSRRFST	100
101	RYGNQNGRIRVLQRFDQRSRQ**FQNLQNHRIVQIEAKPNTLVLPKHADADN**	150
151	**ILVIQQGQATVTVANGNNRKSFNLDEGHALRIPSGF**ISYILNRHDNQNLR	200
201	**VAKISMPVNTPGQFEDFFPASSRDQSSYLQGFSRNTL**EAAF**NAEFNEIRR**	250
251	**VLLEENAGGEQEERGQ**RRWSTRSSENNEG**VIVKVSKEHVEE**LTKHAKSVS	300
301	KKG**SEEEGDITNPINLREGEPDLSNNFGKLFEVKPDKKNPQLQDLDMMLT**	350
351	**CVEIKEGALMLPHFNSKAMVIVVVNKGTGNLEL**VAVRKEQQQRGRREEEE	400
401	DEDEEEEGSN**REVRRYTARL**KEGDV**FIMPAAHPVAINASSELHLLGFGIN**	450
451	**AENNHRIFLAGDKDNVIDQIEKQAKDLAFPGSGEQVEKLIKNQKESHFVS**	500
501	**A**RPQSQSQSPSSP**EKESPEKEDQEEENQGGKG**PLLSILKAFN	542

Figure 58: The primary amino acid sequence of Ara h 1 evaluated with IgE-binding epitopes identified in the article of van Boxtel et al.[23] are underlined.

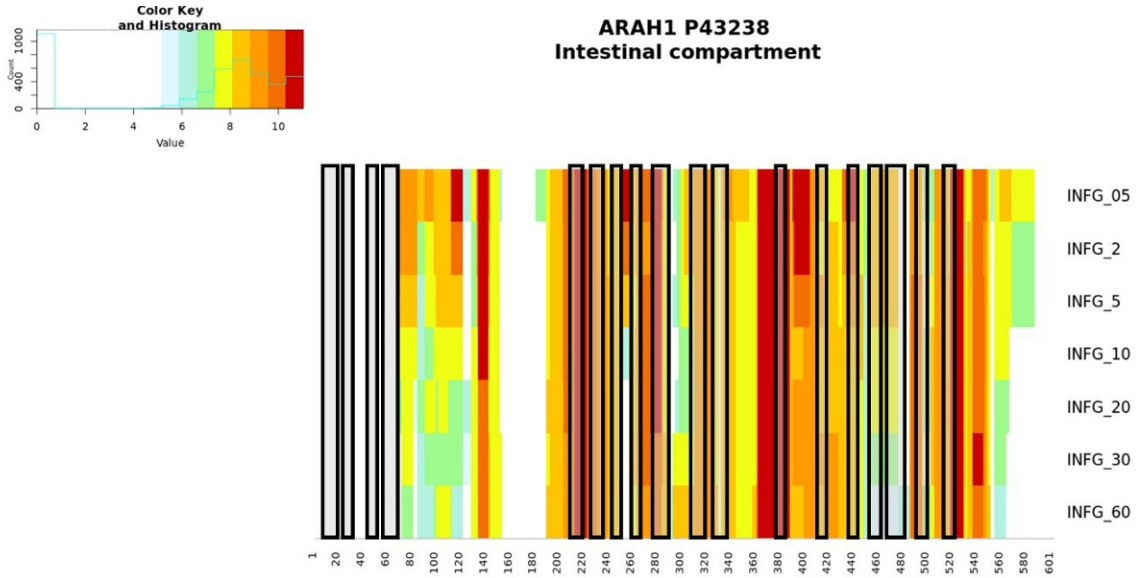


Figure 59: A heat map showing the sum of peptide abundances per amino acid from all of the peptides generated at different time points for the early phase intestinal phase digestion of Ara h 1. Epitopes from Figure 55 are also overlaid on the sequence

In summary, 485 peptides were identified in the gastric phase after in vitro digestion of Ara h 1 by the 3 models and 682 in the intestinal phase. 240 peptides slowly disappeared in the early phase and rapidly in late phase model. Fifty-five peptides were specific of the early phase model. The large number of potential epitopes indicated in Figure 58 mean that the correlation between the epitope positions and the regions of high peptide abundance in Figures 57 and 59 is less than convincing.

GLI

In vitro digestion of GLI with the 3 models led to the identification of 611 peptides in the gastric phase and 2074 in the intestinal phase. In the gastric phase, peptides released by the infant model exhibited a slightly higher median Mw than those obtained with the 2 other models. During intestinal digestion, the peptides median Mw decreased with time, the lower median Mw peptides being identified in the late phase model.

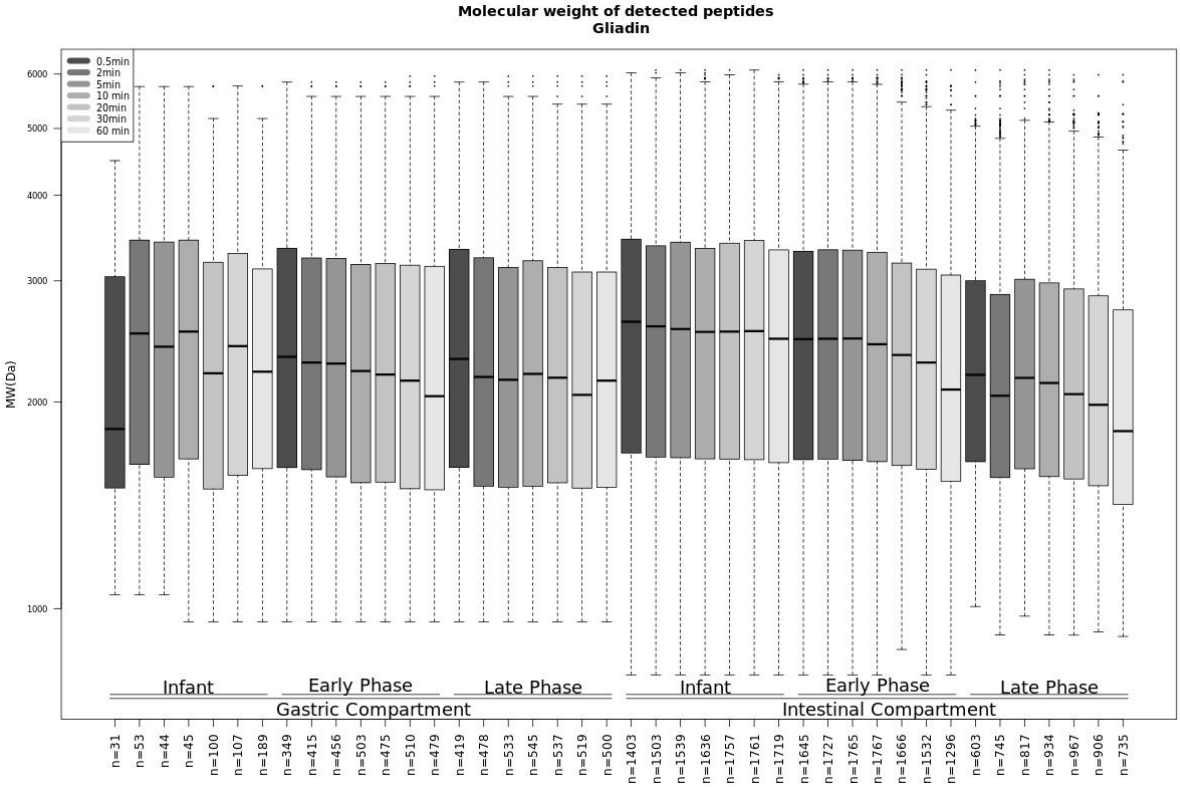


Figure 60: A box plot showing the molecular weight of peptides from GLI after gastric and intestinal digestion.

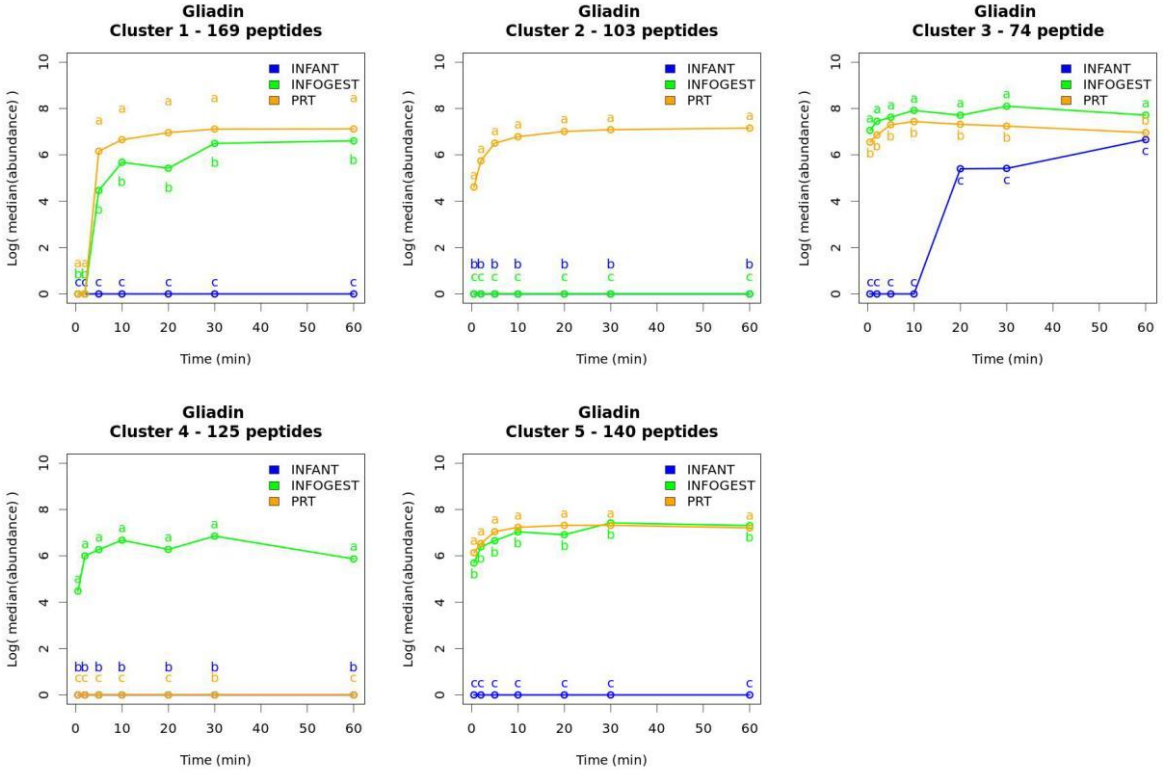


Figure 61: Kinetic plots of area under the curve for the 5 peptide clusters of GLI through digestion in the infant (blue), early phase (green) and late phase (yellow)

Clustering of the peptides observed in the gastric phase led to 5 clusters. Cluster 2 (103 peptides) were highly abundant only in the late phase model, whereas Cluster 4 (125 peptides) correspond to peptides with high abundances only in the early phase model.

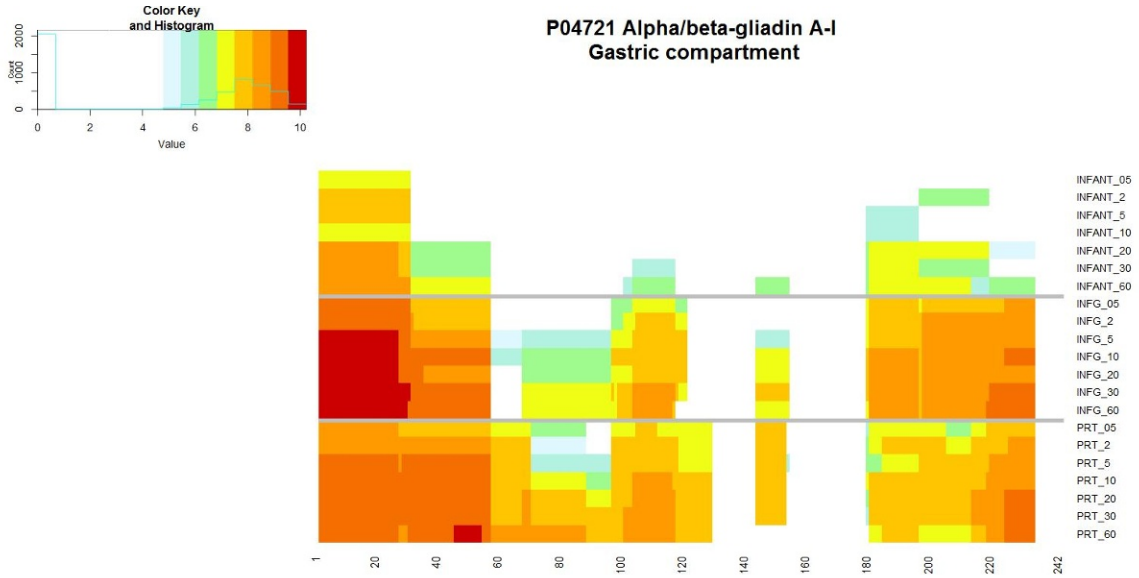


Figure 62: A heat map showing the sum of peptide abundances per amino acid from all of the peptides generated at different time points for the different gastric digestion scenarios from GLI. The scenarios are infant, INFG (early phase) and PRT (late phase).

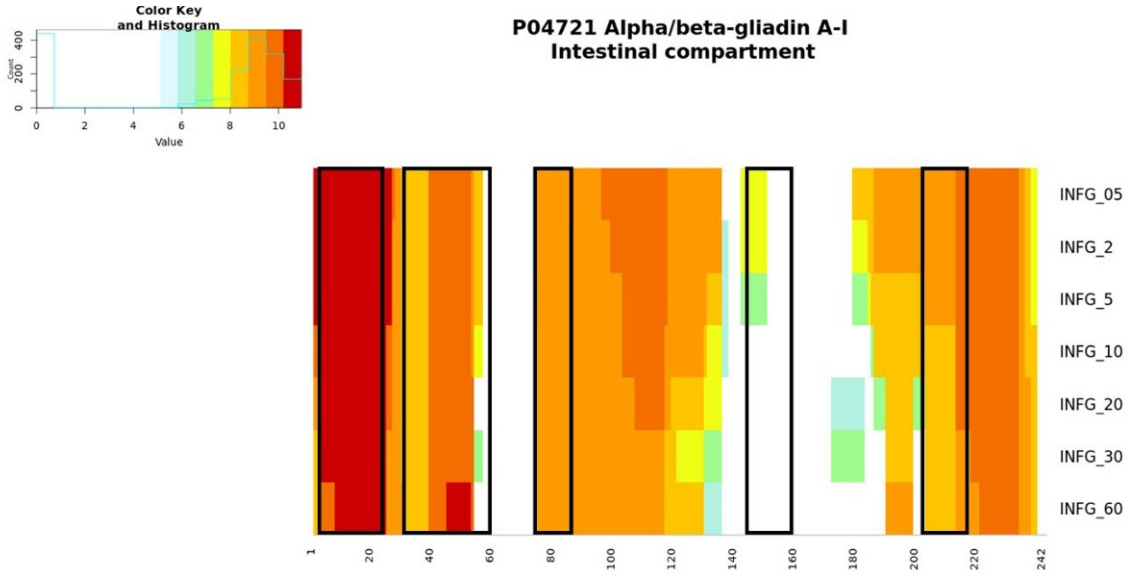


Figure 63: A heat map showing the sum of peptide abundances per amino acid from all of the peptides generated at different time points for the early phase intestinal phase digestion of GLI. Epitopes from are also overlaid on the sequence

In summary, 611 peptides were identified in the gastric phase after in vitro digestion of GLI by the 3 models and 2074 in the intestinal phase. One hundred and twenty-five peptides were mainly specific of the early phase model whereas 103 peptides were mainly seen in late phase model.

3.3 Levels of uncertainty found and mitigation of potential issues

In our experience, one of the primary sources of error in quantitative digestion experiments is the activity of the enzymes. These are either not measured or measured inaccurately by using the wrong assay. The Infogest recommendations for simulation of adult digestion [3] not only recommends specific enzyme activities but also provides detailed information on how the activity should be measured. However, even after this we have seen evidence from ring trial [24] that this can still be inconsistent. Thus, regardless of the outcome of this study, recommendations should be provided on enzyme activity and assays.

One of the major issues arising from the development of the *in vitro* intestinal protocols was the choice of Pancreatin as a mixture of pancreatic enzymes, since such a complex composition makes it difficult to interpret the SDS-PAGE results, particularly for the late phase model (PRT) as observed in the previous report. It also affects LC-MS analysis since it was not possible to inject the same amount of protein material into the LC-MS for the intestinal phase, making comparison of intestinal phase scenarios impossible. This approach was not adopted in Task 3 but rather individual enzymes trypsin and chymotrypsin were used. Despite this simplification, three extra bands were observed between 6 and 14 kDa in the early and late phase adult models that could hinder the identification of intact protein or protein fragments within this range of molecular weight. These bands arise from the reducing conditions of the SDS-PAGE. Figure 64 shows the SDS-PAGE of individual trypsin and α-chymotrypsin, and digestive enzymes in the intestinal phase of the late phase adult model under reducing or non-reducing conditions, respectively. The effect of heating was also tested with no modifications being observed on the SDS-PAGE profile of these enzymes. Chymotrypsin from bovine pancreas has a molecular weight of 25 kDa in its native state, and shows as a major band at this molecular weight in SDS-PAGE under non-reducing conditions. The degree of protein purity according to the supplier (Sigma-Aldrich) is 92% for the batch used. Chymotrypsin comprises three polypeptide chains: A chain (13 amino acid residues, 1.25kDa), B chain (131 amino acid residues, 13.9 kDa) and C chain (97 amino acid residues, 10.1 kDa), linked by disulphide bridges [25]. Therefore, B and C chain are expected to be observed in SDS-PAGE under reducing conditions within the range of 6-14 kDa, as reflected in Figure 64. Nevertheless, a fraction of chymotrypsin is still observed as a band of approximately 25 kDa, suggesting incomplete reduction. Trypsin from porcine pancreas consists of a single polypeptide chain with a molecular weight of 23.8 kDa and shows as a major smeared band at this molecular weight in SDS-PAGE under non-reducing conditions. Under reducing conditions, two minor bands of smaller molecular weight appear within the range of 6-14 kDa. The one with lower molecular weight shows at the same position as the band for reduced chymotrypsin. The one with slightly higher molecular weight may contribute to the third extra band observed in the intestinal control under reducing conditions, which always shows with less intensity between the bands and likely corresponds to reduced chymotrypsin. Since reducing conditions are not expected to affect the molecular weight of trypsin, it may be evidence of some autolysis, despite inactivating the enzyme immediately after dissolution in physiological media for the control experiments. The autolysis products may still be held together by the intramolecular disulphide bonds occurring in trypsin under non-reducing conditions, thus they only show under reducing conditions.

Indeed, reducing conditions have shown to be key in elucidating proteolysis in certain proteins because intramolecular disulphide bonds keep the hydrolysis products together as a single molecule with a molecular mass, which could be similar to the native protein [17]. In any case, for proteins that are known to have intramolecular disulphide bonds, SDS-PAGE under non-reducing conditions may provide complementary information on the degree of proteolysis to fully understand the gastrointestinal stability.

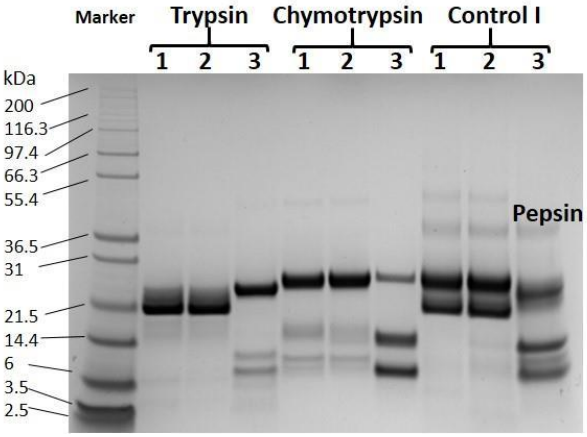


Figure 64: SDS-PAGE of trypsin, chymotrypsin and control of digestive enzymes in the intestinal phase of late phase adult model under different conditions (5 µg loaded of trypsin and chymotrypsin). 1: non-heated and non-reduced; 2: heated and non-reduced; 3: heated and reduced.

All the reasons above highlight that even the relatively simple readout of SDS-PAGE may not always be easy to interpret. In any case, supporting techniques such as LC-MS are required for the identification of smaller peptides (<4kDa).

4. Conclusions

The experiments described in this report were undertaken to address two primary questions.

1. Do the new protocols offer differences or advantages from the current pepsin resistance test with respect to persistent fragments larger than 9 amino acids?
2. Are these differences valuable for risk assessment taking into account Annex B and Figure B.2[1] (Persistent peptides larger than 9 amino acids).

In answer to the first question, the results from the SDS-PAGE analysis provide information about intact protein and large peptides (>~4kDa). Thus, providing discrimination on the basis of size, persistence and abundance. It is clear that the three different digestion scenarios generated very different trends in digestibility. Although the SDS-PAGE shows that there are significant differences in the rates of protein digestion under the different digestion scenarios, it is not clear whether this in itself is advantageous. Certainly, there is no clear correlation between fast or slow digestion of a protein and its allergenic potential for any of the scenarios. This suggests that the simple SDS-PAGE readout is not sufficient for risk assessment of novel allergens.

In addition to the simple readout of SDS-PAGE that shows the persistence of intact protein and large peptides the project also undertook LC-MS to provide more details on the peptides generated above a 9 amino acid cut-off. Again, this data has shown significant differences between proteins and between digestion scenarios. The large number of peptides involved has presented a significant challenge in how best to display the data for the easiest interpretation. In an attempt to address question one above; we have performed a cluster analysis to highlight peptides that showed the most variation between the gastric phases of the digestion scenarios. Although we have mapped the largest of these clusters onto the primary sequences of selected proteins, it maybe that we have not chosen the most relevant peptides from an allergenic perspective. We have also mapped the gastric and for some proteins, the intestinal peptides, onto the primary sequence for comparison. We have shown that there are some correlations between regions of high peptide abundance and the location of known epitopes but in all cases the correlation is limited. We cannot rule out that this may be more of an indication of the confidence in the epitopes.

We believe that the approach used here is extremely powerful but more targeted research will be needed to link these results to immunological outcomes. For example, the use of artificial intelligence (AI) to provide statistical analysis of all persistent peptides and using a broad range of known allergens and their epitopes as training sets. Thus, developing rules linking LC-MS data analysis of digesta to the probability of allergenicity. Indeed, this process could start based on the exiting data that we have generated within this project.

References

1. Naegeli, H.; Birch, A. N.; Casacuberta, J.; De Schrijver, A.; Gralak, M. A.; Guerche, P.; Jones, H.; Manachini, B.; Messean, A.; Nielsen, E. E.; Nogue, F.; Robaglia, C.; Rostoks, N.; Sweet, J.; Tebbe, C.; Visioli, F.; Wal, J. M.; Eigenmann, P.; Epstein, M.; Hoffmann-Sommergruber, K.; Koning, F.; Lovik, M.; Mills, C.; Moreno, F. J.; van Loveren, H.; Selb, R.; Dumont, A. F.; Modified, E. P. G., Guidance on allergenicity assessment of genetically modified plants. *Efsa Journal* **2017,** *15* (6), 4862-4911.
2. Menard, O.; Bourlieu, C.; De Oliveira, S. C.; Dellarosa, N.; Laghi, L.; Carriere, F.; Capozzi, F.; Dupont, D.; Deglaire, A., A first step towards a consensus static in vitro model for simulating full-term infant digestion. *Food Chemistry* **2018,** *240*, 338-345.

3. Minekus, M.; Alminger, M.; Alvito, P.; Ballance, S.; Bohn, T.; Bourlieu, C.; Carriere, F.; Boutrou, R.; Corredig, M.; Dupont, D.; Dufour, C.; Egger, L.; Golding, M.; Karakaya, S.; Kirkhus, B.; Le Feunteun, S.; Lesmes, U.; Macierzanka, A.; Mackie, A.; Marze, S.; McClements, D. J.; Menard, O.; Recio, I.; Santos, C. N.; Singh, R. P.; Vegarud, G. E.; Wickham, M. S. J.; Weitschies, W.; Brodkorb, A., A standardised static in vitro digestion method suitable for food - an international consensus. *Food & Function* **2014,** *5* (6), 1113-1124.
4. Thomas, K.; Herouet-Guicheney, C.; Ladics, G.; Bannon, G.; Cockburn, A.; Crevel, R.; Fitzpatrick, J.; Mills, C.; Privalle, L.; Vieths, S., Evaluating the effect of food processing on the potential human allergenicity of novel proteins: International workshop report. *Food and Chemical Toxicology* **2007,** *45* (7), 1116-1122.
5. Astwood, J. D.; Leach, J. N.; Fuchs, R. L., Stability of food allergens to digestion in vitro. *Nature Biotechnology* **1996,** *14* (10), 1269-1273.
6. Marsh, J.; Rigby, N.; Wellner, K.; Reese, G.; Knulst, A.; Akkerdaas, J.; van Ree, R.; Radauer, C.; Lovegrove, A.; Sancho, A.; Mills, C.; Vieths, S.; Hoffmann-Sommergruber, K.; Shewry, P. R., Purification and characterisation of a panel of peanut allergens suitable for use in allergy diagnosis. *Mol Nutr Food Res* **2008,** *52*, S272-S285.
7. Eiwegger, T.; Rigby, N.; Mondoulet, L.; Bernard, H.; Krauth, M. T.; Boehm, A.; Dehlink, E.; Valent, P.; Wal, J. M.; Mills, E. N. C.; Szepfalusi, Z., Gastro-duodenal digestion products of the major peanut allergen Ara h 1 retain an allergenic potential. *Clin Exp Allergy* **2006,** *36* (10), 1281-1288.
8. Thomas, K.; Aalbers, M.; Bannon, G. A.; Bartels, M.; Dearman, R. J.; Esdaile, D. J.; Fu, T. J.; Glatt, C. M.; Hadfield, N.; Hatzos, C.; Hefle, S. L.; Heylings, J. R.; Goodman, R. E.; Henry, B.; Herouet, C.; Holsapple, M.; Ladics, G. S.; Landry, T. D.; MacIntosh, S. C.; Rice, E. A.; Privalle, L. S.; Steiner, H. Y.; Teshima, R.; van Ree, R.; Woolhiser, M.; Zawodny, J., A multi-laboratory evaluation of a common in vitro pepsin digestion assay protocol used in assessing the safety of novel proteins. *Regulatory Toxicology and Pharmacology* **2004,** *39* (2), 87-98.
9. Rossi, S. S.; Converse, J. L.; Hofmann, A. F., High-pressure liquid-chromatographic analysis of conjugated bile-acids in human bile - simultaneous resolution of sulfated and unsulfated lithocholyl amidates and the common conjugated bile-acids. *Journal of Lipid Research* **1987,** *28* (5), 589-595.
10. Reddy, I. M.; Kella, N. K. D.; Kinsella, J. E., Structural and Conformational Basis of the Resistance of Beta-Lactoglobulin to Peptic and Chymotryptic Digestion. *J Agr Food Chem* **1988,** *36* (4), 737-741.
11. Dupont, D.; Mandalari, G.; Molle, D.; Jardin, J.; Leonil, J.; Faulks, R. M.; Wickham, M. S. J.; Mills, E. N. C.; Mackie, A. R., Comparative resistance of food proteins to adult and infant in vitro digestion models. *Molecular Nutrition & Food Research* **2010,** *54* (6), 767-780.
12. Mandalari, G.; Mackie, A. M.; Rigby, N. M.; Wickham, M. S. J.; Mills, E. N. C., Physiological phosphatidylcholine protects bovine beta-lactoglobulin from simulated gastrointestinal proteolysis. *Mol. Nutr. Food Res.* **2009,** *53*, S131-S139.
13. Takagi, K.; Teshima, R.; Okunuki, H.; Sawada, J., Comparative study of in vitro digestibility of food proteins and effect of preheating on the digestion. *Biol Pharm Bull* **2003,** *26* (7), 969-973.
14. Fu, T. T.; Abbott, U. R.; Hatzos, C., Digestibility of food allergens and nonallergenic proteins in simulated gastric fluid and simulated intestinal fluid - A comparative study. *Journal of Agricultural and Food Chemistry* **2002,** *50* (24), 7154-7160.
15. Gass, J.; Vora, H.; Hofmann, A. F.; Gray, G. M.; Khosla, C., Enhancement of dietary protein digestion by conjugated bile acids. *Gastroenterology* **2007,** *133* (1), 16-23.
16. De Angelis, E.; Pilolli, R.; Bavaro, S. L.; Monaci, L., Insight into the gastro-duodenal digestion resistance of soybean proteins and potential implications for residual immunogenicity. *Food Funct* **2017,** *8* (4), 1599-1610.
17. Koppelman, S. J.; Hefle, S. L.; Taylor, S. L.; de Jong, G. A. H., Digestion of peanut allergens Ara h 1, Ara h 2, Ara h 3, and Ara h 6: A comparative in vitro study and partial characterization of digestion-resistant peptides. *Mol Nutr Food Res* **2010,** *54* (12), 1711-1721.
18. Wu, X. Z.; Huang, T. M.; Mullett, W. M.; Yeung, J. M.; Pawliszyn, J., Determination of isoelectric point and investigation of immunoreaction in peanut allergenic proteins-rabbit IgG antibody system by whole-column imaged capillary isoelectric focusing. *J Microcolumn Sep* **2001,** *13* (8), 322-326.

19. Smith, F.; Pan, X. Y.; Bellido, V.; Toole, G. A.; Gates, F. K.; Wickham, M. S. J.; Shewry, P. R.; Bakalis, S.; Padfield, P.; Mills, E. N. C., Digestibility of gluten proteins is reduced by baking and enhanced by starch digestion. *Mol Nutr Food Res* **2015,** *59* (10), 2034-2043.
20. Cerecedo, I.; Zamora, J.; Shreffler, W. G.; Lin, J.; Bardina, L.; Dieguez, M. C.; Wang, J.; Muriel, A.; de la Hoz, B.; Sampson, H. A., Mapping of the IgE and IgG4 sequential epitopes of milk allergens with a peptide microarray-based immunoassay. *Journal of Allergy and Clinical Immunology* **2008,** *122* (3), 589-594.
21. Jimenez-Saiz, R.; Belloque, J.; Molina, E.; Lopez-Fandino, R., Human Immunoglobulin E (IgE) Binding to Heated and Glycated Ovalbumin and Ovomucoid before and after in Vitro Digestion. *Journal of Agricultural and Food Chemistry* **2011,** *59* (18), 10044-10051.
22. Mine, Y.; Rupa, P., Fine mapping and structural analysis of immunodominant IgE allergenic epitopes in chicken egg ovalbumin. *Protein Engineering, Design and Selection* **2003,** *16* (10), 747-752.
23. van Boxtel, E. L.; Koppelman, S. J.; van den Broek, L. A. M.; Gruppen, H., Determination of Pepsin-Susceptible and Pepsin-Resistant Epitopes in Native and Heat-Treated Peanut Allergen Ara h 1. *Journal of Agricultural and Food Chemistry* **2008,** *56* (6), 2223-2230.
24. Egger, L.; Menard, O.; Delgado-Andrade, C.; Alvito, P.; Assuncao, R.; Balance, S.; Barbera, R.; Brodkorb, A.; Cattenoz, T.; Clemente, A.; Comi, I.; Dupont, D.; Garcia-Llatas, G.; Lagarda, M. J.; Le Feunteun, S.; JanssenDuijghuijsen, L.; Karakaya, S.; Lesmes, U.; Mackie, A. R.; Martins, C.; Meynier, A.; Miralles, B.; Murray, B. S.; Pihlanto, A.; Picariello, G.; Santos, C. N.; Simsek, S.; Recio, I.; Rigby, N.; Rioux, L. E.; Stoffers, H.; Tavares, A.; Tavares, L.; Turgeon, S.; Ulleberg, E. K.; Vegarud, G. E.; Vergeres, G.; Portmann, R., The harmonized INFOGEST in vitro digestion method: From knowledge to action. *Food Res. Int.* **2016,** *88*, 217-225.
25. Marshall, A. C.; Keiller, B. G.; Pederick, J. L.; Abell, A. D.; Bruning, J. B., Crystal Structure of Bovine Alpha-Chymotrypsin in Space Group P6(5). *Crystals* **2018,** *8* (12), 7.
26. Brodkorb, A.; Egger, L.; Alminger, M.; Alvito, P.; Assunção, R.; Ballance, S.; Bohn, T.; Bourlieu-Lacanal, C.; Boutrou, R.; Carrière, F.; Clemente, A.; Corredig, M.; Dupont, D.; Dufour, C.; Edwards, C.; Golding, M.; Karakaya, S.; Kirkhus, B.; Le Feunteun, S.; Lesmes, U.; Macierzanka, A.; Mackie, A. R.; Martins, C.; Marze, S.; McClements, D. J.; Ménard, O.; Minekus, M.; Portmann, R.; Santos, C. N.; Souchon, I.; Singh, R. P.; Vegarud, G. E.; Wickham, M. S. J.; Weitschies, W.; Recio, I., INFOGEST static in vitro simulation of gastrointestinal food digestion. *Nature Protocols* **2019,** *14*, 991–1014.

Abbreviations

BLG	Bovine β-lactoglobulin
BCAS	Bovine β-casein
LYS	Hen's egg lysozyme
OVA	Hen's egg ovalbumin
KTI	Soybean Kunitz-type trypsin inhibitor B
Lip	Soybean lipoxygenase
Con A	Concanavalin A from Jack bean
AP	Potato acid phosphatase
GLI	Gliadin from wheat
Ara h 1	7s globulin from peanut
SDS-PAGE	Sodium dodecylsulphate polyacrylamide gel electrophoresis
LC-MS	Liquid chromatography mass spectrometry

Appendix A – Standard Operating Procedures for the Digestion Protocols

Personal protective clothing (lab coat, safety glasses and nitrile rubber gloves) must be worn at all times when handling reagents, preparing samples and carrying out *in vitro* digestion experiments. Do not breathe dust. Hazardous reagents (see below) must be handled under the fume hood.

Equipment needed:

- Shaking incubator adjustable to 100 rpm and 37 °C
- 50 mL centrifuge polypropylene tubes
- 15 mL centrifuge polypropylene tubes
- Microcentrifuge tubes (1.5 mL and 0.5 mL)
- Racks for all tubes sizes
- Set of pipettes (10, 100, 1000 and 5000 µL) and tips
- Standard laboratory vortex
- Standard laboratory pH-meter
- Timer

Reagents for the three *in vitro* digestion protocols:

- Porcine pepsin (Sigma-Aldrich, Cat. No. P7012)
- Porcine trypsin (Sigma-Aldrich, Cat. No. T0303),
- Bovine chymotrypsin (Sigma-Aldrich, Cat. No. C4129)
- Sodium glycocholate (NaGC, Sigma-Aldrich, Cat. No. G7132)
- Sodium glycochenodeoxycholate (NaGCDC, Sigma-Aldrich, Cat. No. G0759)
- Test proteins
- $CaCl_2(H_2O)_2$ (analytical grade)
- KCl (analytical grade)
- KH_2PO_4 (analytical grade)
- $NaHCO_3$ (analytical grade)
- NaCl (analytical grade)
- $MgCl_2(H_2O)_6$ (analytical grade)
- $(NH_4)_2CO_3$ (analytical grade). **Caution:** Acute oral toxicity.
- HCl 1 M (analytical grade). **Caution:** Corrosive. Causes severe skin burns and eye damage. May cause respiratory irritation.
- NaOH 1 M (analytical grade). **Caution:** Corrosive. Causes severe skin burns and eye damage. May cause respiratory irritation.
- Pefabloc SC (Sigma-Aldrich, Cat. No. 76307). **Caution:** Corrosive.
- Pepstatin A (Sigma-Aldrich, Cat No. 10253286001)
- Ultrapure type I water (e.g. from Milli-Q system). Herein referred to as water.
- Methanol (HPLC grade). **Caution:** Toxic if swallowed, in contact with skin or if inhaled. Highly flammable.

Enzyme activities:

Porcine pepsin (P7012, Lot# SLBW6530) has an enzyme activity of 4177 U/mg. Porcine trypsin (T0303, Lot#SLBT2691) has an activity of 232.54 U/mg, and bovine chymotrypsin (C4129, Lot#SLBV2540) has an activity of 55 U/mg. These enzyme activities are provided in the certificate of analysis from the specific batch in Sigma-Aldrich website, which are assayed according to the protocols described in the supplementary material of Minekus et al.[3]

Preparation of enzyme inhibitors:

- **Pepstatin** (Mw 685.89 g/mol) is used as protease inhibitor for the gastric samples. This is soluble in methanol. Stock solutions at 1 mg/ml (1.5 mM) in methanol are stable at least a week at 4 °C and for months at -20 °C. If solutions become darker yellow, the reagent is hydrolysing. A 1.5 mM stock solution can be prepared in the bottle where is supplied by adding 2 mL of methanol to the 2 mg provided (1 mg/mL) and stored in the freezer. Then, a stock aliquot of 100 µL is mixed with 100 µL of methanol to obtain a working solution of 0.73 mM.
- **Pefabloc** (Mw 239.69 g/mol) is used as protease inhibitor for the intestinal samples. This is soluble in water. Solutions in water are slightly acidic and retain inhibitory activity for up to six months when stored refrigerated. A 0.1 M working solution can be prepared in an Eppendorf tube (24 mg in 1 mL of Milli-Q water) and stored in the fridge. Alternatively, a stock solution can be prepared in the ampule where is supplied by adding 1 mL of water to dissolve the 100 mg provided, and then 0.25 mL of this stock is diluted to 1 mL to obtain the working solution.

1. Early phase adult static *in vitro* digestion

The adaptations from the Standard Infogest recommendations[26] are: no oral phase, purified bile salts replace bile extract, individual enzymes (trypsin and chymotrypsin) replace pancreatin, and 1 h of gastric and intestinal phase replace the standard 2 hours.

Reagents and preparation of simulated gastric and intestinal fluid electrolyte stock solutions

- Liquid **protein sample** to be digested. An initial amount of 5 mL (5 mg/mL in water) is required for each replicate. This gives a total sample protein requirement of 25 mg per replicate.
- **Enzymes and bile salts**: porcine pepsin (P7012, Lot# SLBW6530), porcine trypsin (T0303, Lot#SLBT2691), bovine chymotrypsin (C4129, Lot#SLBV2540), purified NaGC and NaGCDC from Sigma-Aldrich.
- **Simulated gastric fluid (SGF) and intestinal fluid (SIF) electrolyte stock solutions** 1.25 x concentrated are prepared in advance as follows[3]:

			SGF (make up to 400 mL)		**SIF** (make up to 400 mL)	
			pH 3		pH 7	
Constituent (Mw, Da)	**Stock conc. (g/L)**	**(M)**	**Vol. of stock (mL)**	**Conc. in SGF (mM)**	**Vol. of stock (mL)**	**Conc. in SIF (mM)**
KCl (74.55)	37.3 (1.87 g in 50 mL)	0.5	**6.9**	6.9	**6.8**	6.8
KH_2PO_4 (136.09)	68 (0.68 g in 10 mL)	0.5	**0.9**	0.9	**0.8**	0.8
$NaHCO_3$ (84.007)	84 (8.4 g in 100 mL)	1	**12.5**	25	**42.5**	85

NaCl (58.44)	117 (2.92 g in 25 mL)	2	**11.8**	47.2	**9.6**	38.4
$MgCl_2(H_2O)_6$ (203.30)	30.5 (0.15 g in 5 mL)	0.15	**0.4**	0.1	**1.1**	0.33
$(NH_4)_2CO_3$ (96.09)	48 (0.24 g in 5 mL)	0.5	**0.5**	0.5	**0**	0
For pH adj.						
HCl		6	1.3	15.6	0.7	8.4

The electrolyte stock solutions are stored in 15 mL centrifuge tubes with sparse headspace* (25 x 2 = 50 in total) in the freezer. *This avoid the release of CO_2 from carbonate leading to pH drift.

- A stock solution of **$CaCl_2 \cdot 2H_2O$** in water is made at 44.1 g/L (0.3 M) in 5 mL (0.2205 g in 5 mL) and stored in 1 mL aliquots in Eppendorf tubes in the freezer.

SGF and SIF will be made of 4 parts of the corresponding electrolyte stock solution (including the enzyme solutions which will be made in electrolyte solution before the digestion experiment) and 1 part of water (including the $CaCl_2$ solution and bile made in water).

- **Pepstatin** at a working concentration of 0.73 mM.
- **Pefabloc** at a working concentration of 0.1 M.

Static *in vitro* digestion model

For each replicate, the whole digestion can be carried out in a 50 mL Falcon tube kept horizontally for better mixing in a shaking incubator at 37 °C and 100 rpm.

For 3-4 replicates, 1 tube of SGF (plus remaining from previous one), 2 tubes of SIF electrolyte stock solutions and 1 tube of calcium chloride stock solution will be taken out of the freezer to defrost and kept pre-incubating at 37 °C. In the meantime:

1. Prepare protein solution at 5 mg/mL (75 mg in 15 mL water) in a 50 mL Falcon tube for three replicates. Leave stirring for at least 1 h at room temperature. Transfer 5 mL aliquots to other two 50 mL Falcon tubes. For the enzymes control substitute the protein solution with 5 mL water.
2. Prepare a stock solution of bile in water: 86 mM of each NaGC and NaGCDC (210 mg NaGC and 202 mg NaGCDC in 5 mL water for 4 replicates; 168 mg NaGC and 161.6 mg NaGCDC in 4 mL water for 3 replicates). Calculate the minimum volume required from the number of replicates to be digested (1 mL x number of replicates). Leave stirring for at least 30 min and pre-incubate at 37 °C.
3. Weigh 23 mg of trypsin for 3 replicates or 30 mg for 4 replicates in a 15 mL Falcon tube.
4. Weigh 24 mg of chymotrypsin for 3 replicates or 32 mg for 4 replicates in a 15 mL Falcon tube.
5. Label Eppendorf tubes for sampling of digesta and transfer to these the protease inhibitors (5 µL of Pepstatin or 10 µL Pefabloc for 200 µL of gastric or intestinal sample, respectively). Keep them refrigerated until carrying out the digestion.
6. When the electrolyte stock solutions are defrosted prepare a pepsin stock solution at 2.99 mg/mL in SGF electrolyte stock solution. Calculate the minimum volume required from the number of replicates to be digested (1.6 mL x number of replicates). 15 mg in 5.01 mL of SGF electrolyte stock solution for 3 replicates or 21 mg in 7.02 mL for 4 replicates. This is done 5 min before starting the gastric phase.
7. Set up timer and vortex.
8. Have ready 1 M HCl and 1 M NaOH in Eppendorf tubes (refill up to 1.5 mL).

Gastric phase:

1. 5 mL of protein sample are initially mixed with **2.4 mL of SGF electrolyte stock solution.**
2. **2.5 µL of 0.3 M $CaCl_2 \cdot 2H_2O$** are added to reach 0.075 mM in the final mixture.
3. **1 M HCl and water** are added to reach pH 3 and achieve 5 mL final volume in the SGF, respectively:

For protein	**mL of 1 M HCl** (to reach pH 3)	**mL of water** (to achieve 5 mL of SGF)
Soybean Kunitz trypsin inhibitor B (KTI)	0.050	0.948
Hen's egg Lysozyme (Lys)	0.010	0.988
Concanavalin A from Jack bean (ConA)	0.020	0.978
Hen's egg Ovalbumin (Ova)	0.025	0.973
Soybean Lipoxygenase (Lip)	0.045	0.953
Potato acid phosphatase (AP)	0.045	0.953
Gliadin from wheat (Gl)	0.020	0.978
Ara h 1 from peanut (Arh1)	0.050	0.948

Only 1 mL of water is added for the enzymes control.

4. **1.6 mL of pepsin stock solution** (containing 4.78 mg or 20000 U) made in SGF electrolyte stock solution are added and mildly vortexed.
5. Start Timer.

(A preliminary experiment must be made to quantify the amount of HCl needed to reach pH 3). Therefore, 5 mL of protein sample are mixed with 5 mL of SGF (pH 3) and kept stirring for 1 h at 37 °C.

Samples (200 µL) are taken from the gastric phase at 0.5, 2, 5, 10, 20, 30 and 60 min and transferred to an Eppendorf tube containing 5 µL of 0.73 mM Pepstatin to stop pepsinolysis.

6. Before finishing the gastric phase (15 min before), prepare in SIF electrolyte stock solution (calculate the volume from the number of replicates to be digested, 2 mL x number of replicates):
   - trypsin solution of 3.7 mg/mL (23 mg in 6.219 mL or 30 mg in 8.112 mL)
   - chymotrypsin solution of 3.91 mg/mL (24 mg in 6.14 mL or 32 mg in 8.19 mL).
7. Sample last aliquot.
8. Stop Timer and reset.

Intestinal phase:

1. 8.6 mL of gastric chyme are initially mixed with **2.88 mL of SIF electrolyte stock solution** and midly vortexed.
2. **17.2 µL of 0.3 M $CaCl_2 \cdot 2H_2O$** are added to reach 0.3 mM in the final mixture.
3. **1 mL of prepared bile** in water is added.
4. **1 M NaOH and water** are added to reach pH 7 and achieve 8.6 mL final volume in the SIF, respectively:

For protein	**mL of 1 M NaOH** (to reach pH 7)	**mL of water** (to achieve 8.6 mL of SIF)
Soybean Kunitz trypsin inhibitor B (KTI)	-	0.703
Hen's egg Lysozyme (Lys)	-	0.703

Concanavalin A from Jack bean (ConA)	-	0.703
Hen's egg Ovalbumin (Ova)	-	0.703
Soybean Lipoxygenase (Lip)	-	0.703
Potato acid phosphatase (AP)	-	0.703
Gliadin from wheat (Gl)	-	0.703
Ara h 1 from peanut (Arh1)	-	0.703

Only 0.703 mL of water is added for the enzymes control.

5. **2 mL of trypsin** solution (containing 7.4 mg, that is 1720 U of trypsin activity) made in SIF electrolyte stock solution are added.
6. **2 mL of chymotrypsin** solution (containing 7.82 mg, that is 430 U of chymotrypsin activity) made in SIF electrolyte stock solution are added.
7. Vortex mildly.
8. Start Timer.

(A preliminary experiment must be made to quantify the amount of NaOH needed to reach pH 7). Therefore, 8.60 mL of gastric chyme are mixed with 8.60 mL of SIF (pH 7) and kept stirring for 1 h at 37 °C.

Samples (200 µL) are taken from the intestinal phase at 0.5, 2, 5, 10, 20, 30 and 60 min and transferred to an Eppendorf tube containing 10 µL of 0.1 M Pefabloc.

Aliquots of digesta are vortexed and 100 µL separated into empty Eppendorf tubes for LC-MS analysis and all frozen at -20 °C for further analysis and post.

Dilution of protein:

- 5 mL protein (5 mg/mL) + 5 mL SGF (Total volume 10 mL); sampling: 200 µL gastric chyme + 5 µL Pepstatin -> 2.44 mg/mL
- 8.6 mL gastric chyme + 8.6 mL SIF (Total volume 17.2 mL); sampling: 200 µL intestinal content + 10 µL Pefabloc -> 1.19 mg/mL

2. Late phase adult static *in vitro* digestion

Reagents and preparation of simulated gastric and intestinal fluid electrolyte stock solutions

- Liquid **protein sample** to be digested. An initial amount of 5 mL (5 mg/mL) is prepared in water and after 1 h stirring, divided into 1.5 mL aliquots in Eppendorf tubes and stored at -20 °C.
- **Enzymes and bile salts**: porcine pepsin (P7012, Lot# SLBW6530), porcine trypsin (T0303, Lot#SLBT2691), bovine chymotrypsin (C4129, Lot#SLBV2540), purified NaGC and NaGCDC from Sigma-Aldrich.
- **Simulated gastric fluid (SGF) electrolyte solution** is prepared as follows: 500 mL of 35 mM NaCl (1.02 g NaCl) and pH 1.2 (adjusted with 1 M HCl, 36 mL aprox.). The SGF electrolyte is stored in 10 mL aliquots (in 15 mL centrifuge tubes) at -20 °C.
- **Simulated intestinal fluid (SIF) electrolyte stock solution** 1.25 x concentrated was prepared in advance as above.
- A stock solution of **$CaCl_2 \cdot 2H_2O$** in water is made at 44.1 g/L (0.3 M) in 5 mL (0.2205 g in 5 mL) and stored in 1 mL aliquots in Eppendorf tubes in the freezer.
- **Pepstatin** at a working concentration of 0.73 mM.
- **Pefabloc** at a working concentration of 0.1 M.

Static *in vitro* digestion model

For each replicate, the whole digestion can be carried out in a 50 mL Falcon tube kept horizontally for better mixing in a shaking incubator at 37 °C and 100 rpm.

For 3-4 replicates, 3-4 tubes of SGF, 2 tubes of SIF electrolyte stock solution and 1 tube of calcium chloride stock solution will be taken out of the freezer to defrost and kept pre-incubating at 37 °C. In the meantime:

1. Defrost one Eppendorf tube of 5 mg/mL protein solution. Each tube contains enough for three replicates (0.5 mL each).
2. Prepare a stock solution of bile in water: 86 mM of each NaGC and NaGCDC (210 mg NaGC and 202 mg NaGCDC in 5 mL water for 4 replicates; 168 mg NaGC and 161.6 mg NaGCDC in 4 mL water for 3 replicates). Calculate the volume from the number of replicates to be digested (1 mL x number of replicates). Leave stirring for at least 30 min and pre-incubate at 37 °C.
3. Weigh 23 mg of trypsin for 3 replicates or 30 mg for 4 replicates in a 15 mL Falcon tube.
4. Weigh 24 mg of chymotrypsin for 3 replicates or 32 mg for 4 replicates in a 15 mL Falcon tube.
5. Label Eppendorf tubes for sampling of digesta and transfer to these the protease inhibitors (5 µL of Pepstatin or 10 µL Pefabloc for 200 µL of gastric or intestinal sample, respectively). Keep them refrigerated until carrying out the digestion.
6. When the electrolyte solutions are defrosted prepare a pepsin solution at 0.63 mg/mL in SGF electrolyte solution. Calculate the volume from the number of replicates to be digested (9.5 mL x number of replicates). 17.96 mg in 28.5 mL of SGF electrolyte solution for 3 replicates or 23.94 mg in 38 mL for 4 replicates. This is done 5 min before starting the gastric phase. Transfer 9.5 mL aliquots to separate 50 mL tubes for each replicate.
7. Set up timer and vortex.
8. Have ready 1 M HCl and 1 M NaOH in Eppendorf tubes (refill up to 1.5 mL).

Gastric phase:

1. 0.5 mL of protein sample is initially mixed with **9.5 mL of SGF (containing 5.99 mg of pepsin or 25000 U)** and mildly vortexed.
2. Start Timer.

Therefore, 0.5 mL of protein sample is mixed with 9.5 mL of SGF (pH 1.2) and kept stirring for 1 h at 37 °C.

Samples (200 µL) are taken from gastric phase at 0.5, 2, 5, 10, 20, 30 and 60 min and transferred to an Eppendorf tube containing 5 µL of 0.73 mM Pepstatin to stop pepsinolysis.

3. Before finishing the gastric phase (15 min before), prepare in SIF electrolyte stock solution (calculate the volume from the number of replicates to be digested, 2 mL x number of replicates):
    - trypsin solution of 3.7 mg/mL (23 mg in 6.219 mL or 30 mg in 8.112 mL)
    - chymotrypsin solution of 3.91 mg/mL (24 mg in 6.14 mL or 32 mg in 8.19 mL).
4. Sample last aliquot.
5. Stop Timer and reset.

Intestinal phase:

1. 8.6 mL of gastric chyme are initially mixed with **2.88 mL of SIF electrolyte stock solution** and mildly vortexed.
2. **17.2 µL of 0.3 M $CaCl_2 \cdot 2H_2O$** are added to reach 0.3 mM in the final mixture.
3. **1 mL of prepared bile** in water is added.
4. **1 M NaOH and water** are added to reach pH 7 and achieve 8.6 mL final volume in the SIF, respectively:

For protein	**mL of 1 M NaOH** (to reach pH 7)	**mL of water** (to achieve 8.6 mL of SIF)
Soybean Kunitz trypsin inhibitor B (KTI)	0.25	0.453
Hen's egg Lysozyme (Lys)	0.25	0.453
Concanavalin A from Jack bean (ConA)	0.25	0.453
Hen's egg Ovalbumin (Ova)	0.25	0.453
Soybean Lipoxygenase (Lip)	0.25	0.453
Potato acid phosphatase (AP)	0.25	0.453
Gliadin from wheat (Gl)	0.25	0.453
Ara h 1 from peanut (Arh1)	0.25	0.453

0.25 mL 1 M NaOH and 0.453 mL of water are added for the enzymes control.

5. **2 mL of trypsin** solution (containing 7.4 mg, that is 1720 U of trypsin activity) made in SIF electrolyte stock solution are added.
6. **2 mL of chymotrypsin** solution (containing 7.82 mg, that is 430 U of chymotrypsin activity) made in SIF electrolyte stock solution are added.
7. Vortex mildly.
8. Start Timer.

(A preliminary experiment must be made to quantify the amount of NaOH needed to reach pH 7). Therefore, 8.60 mL of gastric chyme are mixed with 8.60 mL of SIF (pH 7) and kept stirring for 1 h at 37 °C.

Samples (200 µL) are taken from the intestinal phase at 0.5, 2, 5, 10, 20, 30 and 60 min and transferred to an Eppendorf tube containing 10 µL of 0.1 M Pefabloc.

Aliquots of digesta are vortexed and 100 µL separated into empty Eppendorf tubes for LC-MS analysis and all frozen at -20 °C for further analysis and post.

Dilution of protein:

- 0.5 mL protein (5 mg/mL) + 9.5 mL SGF (Total volume 10 mL); sampling: 200 µL gastric chyme + 5 µL Pepstatin -> 0.244 mg/mL
- 8.6 mL gastric chyme + 8.6 mL SIF (Total volume 17.2 mL); sampling: 200 µL intestinal content + 10 µL Pefabloc -> 0.119 mg/mL

3. Infant static *in vitro* digestion

The adaptations from the protocol of Menard et al.[2] are: purified bile salts replace bovine bile extract and individual enzymes (trypsin and chymotrypsin) replace pancreatin.

Reagents and preparation of simulated gastric and intestinal fluid electrolyte stock solutions

- Liquid **protein sample** to be digested. An initial amount of 5 mL (5 mg/mL in water) is required for each replicate. This gives a total sample protein requirement of 25 mg per replicate.
- **Enzymes and bile salts**: porcine pepsin (P7012, Lot# SLBW6530), porcine trypsin (T0303, Lot#SLBT2691), bovine chymotrypsin (C4129, Lot#SLBV2540), purified NaGC and NaGCDC from Sigma-Aldrich.
- **Simulated gastric fluid (SGF) and intestinal fluid (SIF) electrolyte stock solutions** 1.25 x concentrated are prepared in advance as follows:

			SGF (make up to 400 mL)		**SIF** (make up to 400 mL)	
			pH 5.3		pH 7	
Constituent (Mw, Da)	**Stock conc. (g/L)**	**(M)**	**Vol. of stock (mL)**	**Conc. in SGF (mM)**	**Vol. of stock (mL)**	**Conc. in SIF (mM)**
KCl (74.55)	37.3 (0.935 g in 25 mL)	0.5	**13**	13	**10**	10
$NaHCO_3$ (84.007)	84 (4.2 g in 50 mL)	1	**-**	-	**42.5**	85
NaCl (58.44)	117 (11.7 g in 100 mL)	2	**23.5**	94	**41**	164
For pH adj.						
HCl		1	Add in aliquots of 2 µL		4.5 mL approx.	

The electrolyte stock solutions are stored in 15 mL centrifuge tubes with sparse headspace* (25 x 2 = 50 in total) in the freezer. *This avoid the release of CO_2 from carbonate leading to pH drift.

- A stock solution of **$CaCl_2 \cdot 2H_2O$** in water is made at 44.1 g/L (0.3 M) in 5 mL (0.2205 g in 5 mL) and stored in 1 mL aliquots in Eppendorf tubes in the freezer.

SGF and SIF will be made of 4 parts of the corresponding electrolyte stock solution (including the enzyme solution which will be made in electrolyte solution before the digestion experiment) and 1 part of water (including the $CaCl_2$ solution and bile made in water).

- **Pepstatin** at a working concentration of 0.73 mM.
- **Pefabloc** at a working concentration of 0.1 M.

Static *in vitro* digestion model

For each replicate, the whole digestion can be carried out in a 50 mL Falcon tube kept horizontally for better mixing in a shaking incubator at 37 °C and 100 rpm.

For 3-4 replicates, 3 tubes of SGF (plus 3 mL at least remaining from previous one), 3 tubes of SIF electrolyte stock solution and 1 tube of calcium chloride stock solution will be taken out of the freezer to defrost and kept pre-incubating at 37 °C. In the meantime:

1. Prepare protein solution at 5 mg/mL (75 mg in 15 mL water) in a 50 mL Falcon tube for three replicates. Leave stirring for at least 1 h. Transfer 5 mL aliquots to other two 50 mL Falcon tubes. For the enzymes control substitute the protein solution with 5 mL water.
2. Prepare a stock solution of bile in water: 32.7 mM of each NaGC and NaGCDC (32 mg NaGC and 31.2 mg NaGCDC in 2 mL water for 3-4 replicates). Calculate the volume from the number of replicates to be digested (0.5 mL x number of replicates). Leave stirring for at least 30 min and pre-incubate at 37 °C.
3. Weigh 10 mg of trypsin for 3-4 replicates in a 50 mL Falcon tube.
4. Weigh 10 mg of chymotrypsin for 3-4 replicates in a 50 mL Falcon tube.
5. Label Eppendorf tubes for sampling of digesta and transfer to these the protease inhibitors (5 µL of Pepstatin or 10 µL Pefabloc for 200 µL of gastric or intestinal sample, respectively). Keep them refrigerated until carrying out the digestion.
6. When the electrolyte stock solutions are defrosted prepare a pepsin stock solution at 0.32 mg/mL in SGF electrolyte stock solution. Calculate the volume from the number of replicates to be digested (1.6 mL x number of replicates). 14 mg in 44 mL of SGF electrolyte stock solution for 3-4 replicates. This is done 5 min before starting the gastric phase.
7. Set up timer and vortex.
8. Have ready 1 M HCl and 1 M NaOH in Eppendorf tubes (refill up to 1.5 mL).

Gastric phase:

1. 5 mL of protein sample are initially mixed with **0.75 mL of SGF electrolyte stock solution.**
2. **1 M HCl and water** are added to reach pH 5.3 and achieve 2.94 mL final volume in the SGF, respectively:

For protein	**mL of 1 M HCl** (to reach pH 5.3)	**mL of water** (to achieve 2.94 mL of SGF)
Soybean Kunitz trypsin inhibitor B (KTI)	0.021	0.569
Hen's egg Lysozyme (Lys)	0.003 (1 M NaOH)	0.587
Concanavalin A from Jack bean (ConA)	0.002 (1 M NaOH)	0.588
Hen's egg Ovalbumin (Ova)	0.004	0.586
Soybean Lipoxygenase (Lip)	0.017	0.573
Potato acid phosphatase (AP)	0.004	0.586
Gliadin from wheat (Gl)	0.002	0.588
Ara h 1 from peanut (Arh1)	0.004 (1 M NaOH)	0.586

Only 0.590 mL water is added for the enzymes control.

3. **1.6 mL of pepsin stock solution** (containing 0.509 mg or 2127 U) made in SGF electrolyte stock solution are added and mildly vortexed.
4. Start Timer.

(A preliminary experiment must be made to quantify the amount of HCl/NaOH needed to reach pH 5.3). Therefore, 5 mL of protein sample are mixed with 2.94 mL of SGF (pH 5.3) and kept stirring for 1 h at 37 °C.

Samples (200 µL) are taken from the gastric phase at 0.5, 2, 5, 10, 20, 30 and 60 min and transferred to an Eppendorf tube containing 5 µL of 0.73 mM Pepstatin to stop pepsinolysis.

5. Before finishing the gastric phase (15 min before), prepare in SIF electrolyte stock solution (calculate the volume from the number of replicates to be digested, 1.5 mL x number of replicates):

- trypsin solution of 0.48 mg/mL (10 mg in 20.678 mL)
- chymotrypsin solution of 0.51 mg/mL (10 mg in 19.563 mL).

6. Sample last aliquot.
7. Stop Timer and reset.

Intestinal phase:

1. 6.54 mL of gastric chyme are initially mixed with **0.2 mL of SIF electrolyte stock solution** and mildly vortexed.
2. **40.1 µL of 0.3 M $CaCl_2 \cdot 2H_2O$** are added to reach 3 mM in the SIF volume (4 mL).
3. **0.5 mL of prepared bile** in water is added.
4. **1 M NaOH/HCl and water** are added to reach pH 7 and stop pepsinolysis, then to reach pH 6.6, and to achieve 4 mL final volume in the SIF, respectively:

For protein	mL of 1 M NaOH (to reach pH 7)	mL of 1 M HCl (to reach pH 6.6)	mL of water (to achieve 4 mL of SIF)
Kunitz trypsin inhibitor	0.002	0.150	0.108
Lysozyme	-	0.180	0.080
Concanavalin A	0.001	0.170	0.089
Ovalbumin	-	0.150	0.110
Soybean Lipoxygenase	0.004	0.160	0.096
Acid phosphatase	0.004	0.104	0.152
Gliadin from wheat	-	0.110	0.150
Ara h 1 from peanut	-	0.140	0.120

Add water before HCl.

Only 0.260 mL of water is added for the enzymes control.

5. **1.5 mL of trypsin** solution (containing 0.72 mg, that is 167.4 U of trypsin activity) made in SIF electrolyte stock solution are added.
6. **1.5 mL of chymotrypsin** solution (containing 0.765 mg, that is 42.1 U of chymotrypsin activity) made in SIF electrolyte stock solution are added.
7. Vortex mildly.
8. Start Timer.

(A preliminary experiment must be made to quantify the amount of NaOH needed to reach pH 7 to stop the enzymatic reaction and the amount of HCl to reach pH 6.6 taking into account the addition of trypsin and chymotrypsin). Therefore, 6.54 mL of gastric chyme is mixed with 4 mL of SIF (pH 6.6) and kept stirring for 1 h at 37 °C.

Samples (200 µL) are taken from the intestinal phase at 0.5, 2, 5, 10, 20, 30 and 60 min and transferred to an Eppendorf tube containing 10 µL of 0.1 M Pefabloc.

Aliquots of digesta are vortexed and 100 µL separated into empty Eppendorf tubes for LC-MS analysis and all frozen at -20 °C for further analysis and post.

Dilution of protein:

- 5 mL protein (5 mg/mL) + 2.94 mL SGF (Total volume 7.94 mL); sampling: 200 µL gastric chyme + 5 µL Pepstatin -> 3.07 mg/mL
- 6.54 mL gastric chyme + 4 mL SIF (Total volume 10.54 mL); sampling: 200 µL intestinal content + 10 µL.