CrossMark click for updates

View Article Online
View Journal

Food & Function

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: E. De Angelis, R. Pilolli, S. L. Bavaro and L. Monaci, *Food Funct.*, 2017, DOI: 10.1039/C6FO01788F.

This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the author guidelines.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the ethical guidelines, outlined in our author and reviewer resource centre, still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/food-function

View Article Online
DOI: 10.1039/C6FO01788F

INSIGHT INTO THE GASTRO-DUODENAL DIGESTION RESISTANCE OF SOYBEAN PROTEINS AND POTENTIAL IMPLICATIONS ON RESIDUAL IMMUNOGENICITY

Elisabetta De Angelis, Rosa Pilolli, Simona L. Bavaro and Linda Monaci*

Institute of Sciences of Food Production, National Research Council of Italy (ISPA-CNR), Via Amendola 122/O, 70126, Bari, Italy

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

*Corresponding author

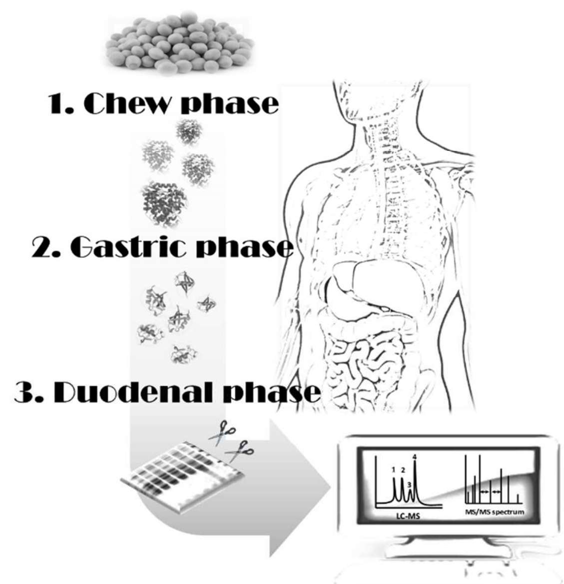
Tel: 0039 080 5929343

Fax: 0039 080 5929374

e-mail: linda.monaci@ispa.cnr.it

View Article Online
DOI: 10.1039/C6FO01788F

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.



GRAPHICAL ABSTRACT

Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

ABSTRACT

Soy is an important component of the human diet thanks to its nutritional value and the high protein content; however, it also represents a risk for allergenic consumers due the its potential to trigger adverse reactions in sensitized individuals. The putative correlation between immunoreactivity and resistance to the human gastrointestinal (GI) digestion has drawn the attention on investigating soybean proteins digestibility. In this work, we provided further insights in this field by performing *in-vitro* simulated GI digestion experiments directly on ground soybean seeds, to provide more realistic results obtained from the digestion of the whole food matrix. Soybean digestion products were analyzed by SDS-PAGE followed by untargeted HPLC-MS/MS analysis and final data were software-based treated to enable protein/peptide identification. The latter allowed monitoring the proteolytic degradation of the main soybean proteins during the gastric and duodenal phases. In particular, β-conglycinin and trypsin inhibitors showed the highest resistance to the combined activity of GI enzymes, presenting only partial degradation also at the end of the duodenal phase as ascertained by the strong electrophoretic bands displayed at 50kDa and 20kDa, respectively. Glycinin subunits presented also, even if to a lower extent, resistance to the complete proteolytic degradation, occurring in the duodenal fluid, mainly as polypeptide fragments with molecular weight lower than 20kDa. In addition, by bioinformatic analysis it was demonstrated that the GI resistant fragments of the allergenic proteins, β-conglycinin and glycinin, retained in their primary structure linear epitopes potentially able to trigger an immunoreaction when exposed to the intestinal mucosa. Moreover, such resistant peptides presented also a structural homology with epitope sequences recognized in other legume species, presenting a potential risk of adverse cross-reaction for a larger category of allergic consumers.

Keywords: soybean, food allergens, *in vitro* digestion, proteolysis resistance

View Article Online
DOI: 10.1039/C6FO01788F

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

1. Introduction

Soybean (*Glycine max*) is considered a source of high-quality proteins, fiber, essential fatty acids, as well as vitamins and minerals, therefore its cultivation and use in food industries is largely spread worldwide.[1] Still, soybean also contains anti-nutritional factors (ANFs), such as agglutinins and protease inhibitors, as well as allergenic proteins that may limit its final use[2,3]The presence of an allergenic ingredient in a complex commodity, represents a problem of public health relevance due to unpredictable effects that can induce in sensitized individuals upon ingestion. In light of this, soybean is listed among the "big eight" allergenic foods in the United States of America accounting for 90% of all food allergies.[4-7] Also in the European Union soybean is considered a priority allergenic food and according to Directive 2007/68/EC,[8] its presence in food must be obligatorily indicated in the food label.

The International Union of Immunological Societies (IUIS) recognized eight soybean proteins as official allergens, capable of eliciting allergic reactions and/or being recognized by IgE antibodies from allergic patients.[9] The list includes soy hydrophobic protein (Gly m 1), defensine (Gly m 2), soy profilin (Gly m 3), pathogenesis-related protein (Gly m 4), β-conglycinin (Gly m 5, vicilin, 7S globulin), glycinin (Gly m 6, legumin, 11S globulin), seed biotinylated proteins (Gly m 7) and 2S albumin protein group (Gly m 8). Some authors reported agglutinin and Kunitz soybean protease inhibitors (Gly m TI)[10,11] and the glycoproteins Gly m Bd 30K and Gly m Bd 28K Gly m Bd 30K as further relevant soybean allergens.[6,12]

Among this list of proteins, glycinin and β-conglycinin are the most investigated targets.[4,13,14] These latter represent the major seed storage proteins in soybean, accounting for about 70-80% of the total seed globulin fraction.[15] In most soybean varieties, the glycinins by themselves account for over 50% of seed storage proteins representing therefore the predominant protein fraction.[16]Glycinins are 11S hexamers with a molecular weight ranging from 320 to 360 kDa. Each monomer consists of subunits composed by a specific acidic (A) polypeptide chain (40kDa) linked through disulfide

Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

bonds to a specific basic (B) polypeptide chain (20kDa) and can be one of the five subunits (glycinin G1: $A_{1a}B_{1b}$,glycininG2: A_2B_{1a}, glycininG3: $A_{1b}B_{1a}$,glycininG4: $A_5A_4B_3$ and glycinin: A_3B_4).[2]β-conglycinins are 7S trimers with molecular masses around180 kDaand as glycoproteins contain 5% of carbohydrate moieties inducing immunoreactivity.[17] These trimers are formed by various combination of three homologous polypeptide subunitsα', α and β, with masses of 76 kDa, 72 kDa, 53 kDa, respectively.[2]

Food allergens may sensitize via different routes, such as the skin or the respiratory tract, although the major route of exposure is represented by food ingestion. Noteworthy, most of the proteins were hydrolyzed down to single amino acids and/or small peptides during digestion, thus facilitating their absorption along the intestinal mucosa,[18]whereas some allergenic proteins could partially resist to proteolytic enzymes of gastro intestinal tract (GI). Surviving as large immunologically active fragments may trigger sensitization of the mucosal immune system after their absorption.[19]Consequently, the resistance to the digestive process might disclose a strict correlation with adverse reactions in allergic individuals, hence encouraging further investigations on this aspect.[20]

In this context, *in vitro* digestion models mimicking human digestion process represent an extremely useful tool to address this open issue, due to their simplicity, low costs (compared to *in-vivo* tests) and good reproducibility.[21] The use of such approach was also recommended by EFSA panel (2010) because it allows to obtain *in vitro,* the digestion products to which the inductive mucosal immune system is exposed, and thereby reproducing the typical situation of oral sensitization.[22] So far, several studies have reported the implementation of static digestion models (also referred to as biochemical model) to assess the resistance of allergenic food proteins to the proteolysis operated by human GI enzymes.[23-27] Focusing on soybean, some works have been directedto investigate the stability of the major soybean proteins to different GI enzymes.[28-31] Mostly, purified allergenic proteins were submitted to very simplified *in vitro* digestion procedures

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.
Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

where proteolytic sensitivity was evaluated by simple exposure to single and/or combined enzymes mixture. Recently, Amigo-Benavent et al. (2011), presented the use of a more complex *in-vitro* digestion model, simulating both human gastric and duodenal compartments, with proper enzymes mixtures, to study the digestibility and immunoreactivity of isolated β-conglycinin and its deglycosilated form.[32] The authors demonstrated the partial survival of the α and β-subunits of the glycosylated form to the enzymatic hydrolysis also confirming their immunoreactivity.[32]

The aim of the present paper is to widen the investigation of soybean allergens digestibility to the characterization of the whole protein profile as obtained by the *in vitro* simulation of GI digestion carried out on the real food matrix (ground soybean seeds). The modification of the protein/peptide pattern was monitored along the different phases of the gastro intestinal digestion model used in this work. A relatively complex static *in vitro* model was implemented throughout the study simulating the oral, gastric and duodenal phases of gastrointestinal digestion. The resulting digestive fluids were analyzed by SDS-PAGE and identified by HPLC-MS/MS in order to track the fate of specific proteins during the GI process and to assess their resistance or sensitivity to specific enzymes. Furthermore, by on-line bioinformatic searches, the generated GI products were further investigated for the presence of known linear epitopes survived to enzymatic proteolysis, inferring on the potential residual immunogenicity of resistant protein fragments. Finally, the structural homology with epitope sequences recognized in other legume species was discussed as a potential risk of adverse cross-reaction for a larger category of allergic consumers.

2. MATERIALS AND METHODS

2.1. Material and reagents

Organic yellow soy was purchased from a local retailer. Acetonitrile (Gold HPLC ultragradient), and trifluoroacetic acid were purchased from Carlo Erba Reagents (Cornaredo, MI, Italia) and ultrapure water was produced by a Millipore Milli-Q system (Millipore, Bedford, MA, USA).

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

Formic acid (MS grade) was provided by Fluka (Milan, Italy) while PTFE syringe filters (4 mm, 0.2 μm) and ultrafiltration (UF) tubes with10 kDa cut-off membranes were purchased from Sartorius (Sartorius Stedim Italy S.p.A., Antella-Bagno a Ripoli, FI, Italy). All reagents for simulated digestion experiments (Sodium chloride - BioXtra ≥ 99.5%, Ammonium Bicarbonate, Calcium chloride, Bis-Tris, Phenylmethansufonyl fluoride - PMSF, Egg lecithin - PC, Sodium taurocholate and Sodium glicodeoxycholate) as well as enzymes (α-amylase from human saliva Type XIII-A, Pepsin from porcine gastric, Trypsin from porcine pancreas Type IX-S, α - chymotrypsin from bovine pancreas Type II and α-amylase from Bacillus sp. Type II-A) were purchased from Sigma Aldrich (Milan, Italy). Concerning SDS-PAGE analysis, Mini-PROTEAN® TGX™ Precast Gels(8.6 x 6.7 x 0.1 cm, 4-20% acrylamide) were purchased from Bio-Rad Laboratories (Segrate, MI, Italy). Ammonium bicarbonate (AMBIC), iodoacetamide (IAA) along with other chemicals for electrophoresis (dithiothreitol, sodium dodecyl sulfate-SDS, glycine, glycerol, coomassie brilliant blue-G 250 and methanol-HPLC grade) were purchased from Sigma-Aldrich (Milan, Italy)while Bromophenol blue was provided by Carlo Erba Reagents (Cornaredo, MI, Italia). Electrophoresis experiments were accomplished on a Mini Protean Tetra Cell equipment provided by Bio-Rad Laboratories (Segrate, MI, Italy). Trypsin (Trypsin Gold, Mass Spectrometry Grade) for in gel protein digestionwas purchasedfromPromega (Milan, Italy).

2.2. *In vitro* simulated gastroduodenal digestion protocol

Whole soybean flour was submitted to an *in vitro* digestion protocol developed at the Institute of Food Research (Norwich, UK) and designed to mimic the biochemical conditions of the upper GI tract of humans namely: (1) chewing, (2) gastric digestion and (3) duodenal digestion, with the appropriate composition of the respective biological fluids and enzymes of an adult human stomach and duodenum. Specifically, as for gastric digestion, porcine pepsin was added to the digestion mixture at a fixed amount of 170U/mg protein substrate, and this value was selected as average of

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

the used amount in published studies by IFR.[33-36] Concerning duodenal phase, proteolytic enzymes such as trypsin (from porcine pancreas) and chymotrypsin (from bovine pancreas) were added to the digestion mixture in fixed amounts: 34.5 BAEE units of trypsin to 1 mg protein and 0.4 BTEE units of chymotrypsin to 1 mg protein; these values were within the range of data published from previous investigation carried out at IFR.[34-36]Figure 1 reported a simplified scheme of the whole procedure.

Before undergoing *in vitro* digestion, soybean seeds were firstly milled coarsely (Bühler, Bühlers.p.a. Segrate, Milano) and successively more finely using a 1mm sieve. Soybean protein content was expected to be approximately 37%, according to what reported by USDA National Nutrient Database (NDB N°16108, Soybeans, mature seeds, raw).

Oral phase (final ratio of food to SSF of 1:1.5). 2.7 g of milled soy were mixed with 4.05 ml of warmed simulated salivary fluid (SSF, 0.15M NaCl pH 6.9, 2 U/ml human salivary amylase) and incubated in an orbital shaker (KS 4000 i-control shaker; IKA Works GmbH & Co. KG, Staufen, Germany) for 15 min at 37°C, 150 rpm.

Gastric phase. An aliquot of 3.55 g (expected amount of proteins = 525 mg) of chewed soy was mixed with 100 mL of warmed simulated gastric fluid (SGF, 0.15M NaCl, pH 2.5) and 1383µl of 10 mg/ml lecithin working solution (concentration of 0.17 mM in the final digestion volume). This latter was prepared by drying an appropriate amount of 50 mg/ml lecithin stock solution and suspending the pellet in SGF. The solution obtained was then sonicated and filtered (0.22 µm PTFE syringe filter) to remove any titanium residue from the ultrasound probe. Afterward thepH of simulated gastric sample was lowered to 2.5, using 1M HCl and the total volume adjusted to 104.9 mL with SGF. Then the sample was incubated under stirring for 15 min at 150 rpm,37°C. After this step, a 5 ml of undigested (GU) sample was collected and mixed with 500 µl of 1 M ammonium bicarbonate buffer before storing on ice. In order to start the gastric proteolytic reaction, 105 µl of pepsin stock solution prepared in SGF (170U/mg of protein in the final mixture) were added to the

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.
Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

gastric mixture left under incubation on an orbital shaker for 2h at 37°C, 150 rpm. Finally, an aliquot of the gastric sample was collected and mixed with 500 µl of 1 M ammonium bicarbonate to stop proteolysis (G=120),. The pH of the whole gastric chime digestion mixture was also adjusted to 7.5 (with 1.0 M NaOH), and incubated at room temperature for 10 min to stop enzymatic activity.

Duodenal phase. 40 ml of gastric chime (expected proteins content = 200 mg) were withdrawn and submitted to the duodenal phase by lowering the pH to 6.5 with the addition of 1.0M HCl. The digestive fluid was then mixed with 1337µl of lecithin working solution (1.8 mM lecithin micelles in the final mixture). This solution was prepared by drying a calculated amount of 50 mg/ml lecithin stock solution and suspending the pellet in 0.15M NaCl (pH 6.5, SDF) together with bile salts (1:1 ratio of sodium taurocholate and sodium glycodeoxycholate, 7.4 mM bile salts in duodenal mixture).The mixture was left under stirring at 37°Cfor about 15 min. Then22µl of 1 M $CaCl_2$ and1179µl of 1M BisTris (0.5 mM and 26.8 mM in the final duodenal mixture, respectively) were added to the digest. Final pH was adjusted to 6.5 by1M NaOH addition and the volume further increased up to 44 ml by adding warmed SDF (0.15 M NaCl, pH 6.5). Finally, the mixture was incubated at 37°C for 10 min in an orbital shaking incubator. For enzymatic reaction, specific amounts of trypsin (34.5 BAEE units/ mg protein substrate), chymotrypsin (0.4 BTEE units/ mg protein substrate) and α-amylase (300 units/ ml digestion fluid) stock solutions in SDF were added to the digest sample. The latter was then left for 3h at 37°C, under stirring (150 rpm). At the end of the duodenal digestion, a 5 ml aliquotwas collected (D=180 min) and mixed with 50µl of 0.1M PMSF to stop the enzyme reaction. All the final digests were stored at -20°C.

2.3. SDS-PAGE analysis

Digested proteins obtained by submitting milled soybean to gastric and duodenal digestion were separated by sodium dodecyl sulfate-polyacrilamyde gel electrophoresis (SDS-PAGE) on 4-20%

View Article Online
DOI: 10.1039/C6FO01788F

precast gels (8.6cm x 6.7cm x1mm). The expected protein content of the samples under investigation were respectively GU=G120= 5μg/μL and D180 = 4.5 μg/μL. Before gel analysis, samples were mixed (1:1 ratio) with a Laemmli buffer (62.5mM TrisHCl, pH 6.8, 25% glycerol, 2% SDS, 0.01% Bromophenol Blue, 100mM DTT) and then reduced at 100°C for 5 min. PAGE gels were run in the Mini-Protean Tetra Cell equipment (Bio-rad Laboratories) with TGS (25mM Tris, 192mM Glycine, 0.1% SDS) running buffer at the following conditions: 50V for 15 min and 80V until the end of the run. Gels were stained using a solution of Coomassie Brilliant Blue G-250 and the bands were detected by using a Gel Doc EZ Imager system (Bio-Rad Laboratories, Segrate, MI, Italy). For protein molecular weight referencing, a Precision PlusProtein™ all blue standard (10-250 kDa, Bio-Rad Laboratories) was loaded on each gel.

2.4. Protein in-gel digestion

Selected protein bands were excised from the polyacrylamide gel and destained by repeated washing (45 min, 37°C) with100mMAMBIC/acetonitrile (1/1, v/v). Gel slices were then dehydrated for 5 minutes at room temperature in 100µl of acetonitrile. After drying the sample in a "speed Vac" centrifuge (Christ RVC 2-18, Martin Christ Gefriertrocknungsanlagen GmbH, Osterode am Harz, Germany) for 10–15 minutes at room temperature, the proteins were reduced for 1 h at 60°C with 10 mM DTT solution (prepared in 25 mM AMBIC) and alkylated for 30 min at room temperature with 55 mM iodoacetamide solution (prepared in25 mM AMBIC).Digestion was carried out overnight at 37 °Cwith mass spectrometry grade trypsin solution (0.1 µg/µl, enzyme: protein ratio 1:50) in 25 mM AMBIC. Successively, gel slices were incubated with 150µl of MilliQwater for10 minutes, with frequent vortex mixing. Then the liquid was removed and transferred into anew microcentrifuge tube. Peptides were extracted from gel by incubation with 50% acetonitrile/5%trifluoroaceticacid/ (1/1, v/v)for 60 min. This step was repeated twice. Peptide mixtures obtained from each extraction step were then pooled together and dried in a "speed Vac"

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.
Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

centrifuge. Each sample was suspended in 80 µl of H_2O/ACN 90/10+0.1% formic acid (v/v) and 20 µl were further injected into LC/MS apparatus.

2.5. Isolation of low molecular weight components of the duodenal digests

Soybean duodenal digests were subjected to purification on ultrafiltration tubes with cut-off membranes (10 kDa) in order to collect the fraction enriched with low molecular weight peptides produced by GI simulated digestion. In particular, 5 mL of digested sample (duodenal phase) were firstly centrifuged for particulate removal (10 min at 770g) and then 4ml of supernatant aliquot was loaded on ultrafiltration tubes. After 4h of centrifugation (3080g) the sample fraction with molecular weight lower than 10kDa was collected, filtered through a cellulose syringe filter (0.2 µm)and analyzed by untargeted HPLC-MS/MS.

2.6. Proteomic HPLC-MS/MS analyses

HPLC-MS/MS analyses were performed on a system composed of a UHPLC pump provided with an autosampler and ESI interface coupled with a Dual pressure Linear Ion Trap Mass Spectrometer Velos Pro™ (Thermo-Fisher-Scientific, San Josè, USA).

For chromatographic separation of the peptides mixture generated from tryptic digestion of soy excised protein bands and of the low molecular weight peptides fraction, 20µl of sample were injected on an Aeris peptide analytical column (internal diameter 2.1 mm, length 150 mm, particle size 3.6 µm, porosity 100Å,Phenomenex) at a flow rate of 200µl/ml. The gradient used is here following reported: from 85% to 45% of solvent A in 40 min, then down to 10% in two min, kept stable for 20 min and back to 85% in 2 min. This composition was maintained for 30 minutes to allow column equilibration (Solvent A= H_2O +0.1% formic acid; solvent B= CH_3CN+0.1% formic acid). MS system was run in *Nth order double play* (Data Dependent™ Acquisition, DDA) mode by activating the dynamic exclusion option. Full description of such analysis mode was detailed

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

elsewhere.[37] Three analytical replicates were carried out for each sample subjected to LC-MS/MS analysis, and the whole data set (three replicates of the same sample) was processed simultaneously by commercial software Proteome Discoverer™ version 2.0 (Thermo-Fisher-Scientific, San Josè, US); protein identification was achieved by Sequest HT search against a soybean customized database extracted by Swiss Prot DB basing on the taxonomy code of *Glycine max* (ID: 3847) and containing about 400 sequences, a multi-consensus report was provided grouping results from different replicates. Given the complexity of enzyme mixtures used for gastro-duodenal digestion simulation, an unspecific cleavage was set for peptide identification, setting the mass tolerance on the precursor and fragment ions at 2 Da and 0.5 Da, respectively. Only unambiguous peptide-spectrum matches were accepted, with a mass tolerance on the assigned precursor ion equal or better than ±300 ppm. Moreover, only proteins identified by three or more unique peptides, with high confidence (FDR<1%), were taken into consideration for any further comment.

2.7. Bioinformatic analysis for epitope occurrence

Peptides sequences identified by in-gel digestion of proteins bands along with low molecular weight peptides detected in the duodenal digestive fluids were finally searched in IEDB database[38] in order to find epitope linear sequences survived to the gastro-duodenal digestion. IEDB results were filtered as following: linear sequence for epitope structure, substring for BLAST option and human as host. Epitopic peptides were then aligned against UniprotKB/Swiss-Prot database in order to scout around for structural similarity with other legumes.[39] Search was performed by setting the E-threshold (number of expected matches in a random database) to 10. “Gapped alignment” option, which allows gaps to be introduced along the sequences for their comparison, was also activated, while “Comparison matrix” option was kept as auto-select, therefore the matrix employed by the software for assigning the probability score for each position in an alignment, was automatically selected depending on the query sequence length.

View Article Online
DOI: 10.1039/C6FO01788F

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

3.RESULTS and DISCUSSION

Due to the likely connection between immunoreactivity of certain allergic proteins and their resistance to the gastrointestinal (GI) digestion, digestibility of soybean proteins were investigated. To this aim, soybean mince was subjected to in-vitro simulation of GI process by a static model miming three main digestion steps: chew, gastric and duodenal phases. Salivary, gastric and duodenal fluids composition (in terms of electrolyte, surfactants, bile salts and typical enzymes) as well as enzyme to protein ratio were optimized for protein digestion taking into account the previous investigation[33-36] and the *in vivo* physiological conditions of adult human gut. All enzymes and reagents were added according to basal concentrations of these components (see Figure 1).

3.1 SDS-PAGE separation and HPLC-MS/MS identification of protein bands

Sample aliquots representative of the undigested proteins (GU) the partially digested proteins (end of the gastric phase, G=120 and end of duodenal phase, D=180) were analyzed in parallel by SDS-PAGE. Figure 2 shows the resulting Comassie blue staining gel reporting electrophoretic profiles of three different samples with different amounts of proteins, theoretically calculated ,: 20 and 50 μg of proteins for undigested samples and 20, 50 and 70 μg for both gastric and duodenal digests. The electrophoretic patterns highlighted in the figure and referred to the three phases reflect the progress of the simulated digestion, in terms of proteolytic degradation. As expected, some bands corresponding to high molecular weight proteins disappeared already after the enzymatic proteolysis, while the number of smaller peptides increased proportionally, as demonstrated by the smeared bands with MW lower than 10kDa appearing along the duodenal digest lane. In order to have a deeper insight on the identity of the specific soybean proteins exhibiting different level of resistance to proteolytic activity of characteristic enzymes involved in the gastro-duodenal digestion, untargeted mass spectrometry based analyses followed by software based protein

View Article Online
DOI: 10.1039/C6FO01788F

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

identification were accomplished on selected bands (labelled from *a* to *m* in the figure 2, lane A-B-C). The bands were excised from the gel and submitted to in-gel tryptic digestion; the resulting peptides pool was separated by reversed phase HPLC and detected by acquiring the signal in DDA mode. A commercial software for protein identification (Proteome Discoverer™ v. 2.0, Thermo Fisher Scientific) was used to process the raw MS/MS spectra assigning the electrophoretic bands to specific soybean proteins and/or their subunits/fragments. The attribution was performed by Sequest HT algorithm through customized database search. Although band "*a*", "*e*" and "*h*" contain two close protein bands, they were treated as a single spot.

Table 1 summarizes the results provided by the software for each band analyzed. Before starting gastric digestion, the main proteins detected in the undigested sample (GU, lane A) were: α', α and β subunits of β-conglycinin (band "*a*" and "*b*") with MW of 50 and 75 kDa, glycinins subunits, acid and basic polypeptide chains released by disulfide bonds reduction, (bands "*c*" and "*d*") along with Trypsin inhibitors and 2S albumin, with a molecular weight comprised between 37 and 20 kDa. Unique protein assignment was not possible due to the low resolution of SDS-PAGE technique in the separation of proteins showing very close molecular weights, however, our electrophoretic profile was similar to the one previously reported by Amnuaycheewa and de Mejia (2010).[40]

By direct comparison of the undigested sample (GU) protein profile with the gastric digest (G=120, lane B), a significant change in electrophoretic bands distribution was clearly observed in Figure 2 (see table 1 for protein identification). In particular, the high molecular weight band "*a*" placed at about 75 kDa and assigned to α/α′ subunits of β-conglycinin was missing, whereas a new protein band labelled as "*e1*" and a stronger band "*e2*", corresponding to the previous band *"b"*, were detected at a lower molecular weight (about 50 kDa) both attributed to β-conglycinin subunits (α,α' and β), proving that such subunits were only partially hydrolyzed by pepsin activity. In addition, it is worth noting that glycinin subunits and lectin proteins banding around 37 kDa (band "*c*" in undigested samples) were not detected after gastric digestion, proving a higher susceptibility of

View Article Online
DOI: 10.1039/C6FO01788F

these proteins to the enzymatic proteolysis. Still some glycinin fragments banding at 20 kDa survived the gastric digestion even if by direct comparison of the total number of peptides identified by HPLC-MS/MS in band "*f*" (G=120 sample) with band "*c*" and "*d*" (GU), a significant decrease in the number of detected sequences was highlighted. This finding could be attributed to the more extended degradation of glycinin chains down to fragments with molecular weight below 10 kDa. Finally, trypsin inhibitor proteins banding at 20 kDa along with 2S albumin in lane A (band "*d*"), were detected also in the gastric digested samples with almost unvaried coverage, thus demonstrating a good resistance of these two proteins to pepsin activity.

After duodenal digestion, namely further proteolytic activity exerted by trypsin and chymotrypsin, the electrophoretic profile of soybean proteins experienced new modifications (see Figure 2). A stronger band around 50 kDa was detected, visibly divided into two different bands marked as "*g*" and "*h*", respectively, and analyzed separately. Both bands were attributed to the α, α′ and β subunits of β-conglycinin (Table 1) proving their significant resistance also to the duodenal enzymes proteolysis. In addition, new protein fragments (not detected after the gastric phase) assigned to β-conglycinin (α, α' subunits), banding approximately at 20kDa in molecular weight, were detected at the end of the duodenal digestion marked as "*i*" in Figure 2. As for identification of band "*l*" in the duodenal profile, (D180, lane C) a group of proteins mainly represented by trypsin inhibitor A, glycinin subunits and 2S albumin was detected, confirming the same composition of band "*f*" in the gastric profile, but with further evidence of glycinins degradation occurred during this phase (lower number of identified peptides), and of β-conglycinin degradation (three unique peptides coming from the α subunit). Finally a portion of the last unresolved broad band below 15kDa referred to as "*m*", was also identified by HPLC-MS/MS and ascribed to different fragments of the same proteins in band "*l*".

In order to have complementary information about the digestibility of soybean proteins, samples collected at the end of duodenal phase (D=180min) were passed through an ultra-filtration tube (10

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

kDa cut off) and the peptides fraction with molecular weight lower than 10 kDa was directly analyzed by HPLC-MS/MS –no further digestion with mass spectrometry grade trypsin was required- and identified via software as previously detailed. It is worthy to be noticed that, besides the upper limit in molecular weight fixed by the membrane cut off (10kDa), there were objective constrains to the chance of detecting all the peptides/polypeptides contained in such fraction, represented by limits due to the ionization efficiency and the operating mass range of our MS equipment. On average, only peptides sequences shorter than 30 AA were efficiently detected. In table 2 the identified peptide fragments detected in the low molecular weight range of the duodenal sample are summarized; we refer to table S1 attached as supporting information for a detailed list of the peptide sequences assigned and the relevant parameters calculated by Proteome Discoverer software. Table 2 highlights that most of small peptides (< 10 kDa) resulting from *in vitro* simulated GI digestion were attributed to the allergen group of glycinin confirming the partial susceptibility of these proteins to enzymatic hydrolysis. These results completed the information provided by table 1, supporting the previous observation made on the electrophoretic bands detected and derived from undigested, gastric and duodenal soybean fluids. As discussed above, the content of glycinin subunits, particularly abundant in undigested electrophoretic profile, gradually decreased upon gastro-duodenal digestion and as a complementary information, in this low molecular weight fraction, a large number of peptides belonging to the glycinin subunits was detected. On the contrary, only few peptides assigned to β-conglycinin, specifically only to α and α' subunits, were identified, confirming the general resistance of the β-conglycinin to proteolytic degradation with preferential sensitivity of the α and α' chain compared to β chain, not detected in this sample. Peptides derived from other proteins were also found in fragment pool below 10 kDa, belonging to trypsin inhibitor, lectin and seed maturation P34 probable thiol protease (Gly m Bd30K).

As general comment, our insight on the fate of the whole soybean protein profile during *in vitro* simulation of GI digestion of the real food matrix (ground soybean seeds) provided consistency

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.
Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

with the previous findings suggested in literature on very simplified case studies, such as purified soy allergens[28,30,32] and soy protein isolates[41]. Like previously reported, the β-conglycinin exhibited a significant resistance to the proteolytic digestion, with particular reference to the β subunit identified in the band "*b*" of undigested sample, and enduring with good sequence coverage also in bands "*e*" of the gastric fluid and "*h*" of the duodenal fluid, whereas the α and α′ subunits of β-conglycinin, presented a partial hydrolysis operated by gastric pepsin (band "a" of undigested sample was substituted by bands "e" in gastric pattern) and by trypsin and chymotrypsin (band "*i*", "*l*" and "*m*") along the duodenal profile. Our results agreed also with what reported by Amigo-Benavent (2011)[32] inferring about a general stability of the proteinβ-conglycinin along digestion, attributed to the glycan moieties spread along the molecule thus limiting the accessibility of pepsin to the cleavage sites. As for glycinin chains group, we observed a partial susceptibility of these proteins to gastric digestion confirmed by the absence in gastric sample of any band aligned with band "*c*" of undigested sample (acid polypeptides), and some resistance of the basic polypeptides chains banding at 20 kDa (band "*d*") in undigested sample, enduring along the two digestion steps, even if with a decreasing number of identified peptides as the digestion proceeds. These findings were only partially in agreement with what reported by Zhao et al. 2010[42] who studied also the susceptibility of purified glycinin and β-conglycinin to sequential pepsin and trypsin digestion. The authors observed that acid polypeptides (A) of glycinin, banding approximately at 37 kDa, were hydrolyzed within 0.5 min of digestion, whilst basic polypeptides (B) banding around 20 kDa were found more resistant, still being completely decomposed after 60 min. Likely, both the differences in the experimental protocols for digestibility assessment and for evaluating the influence of the whole food matrix on the single protein digestibility, accounted for the apparent discrepancy of the final findings, proving the relevance of our investigation compared to previous results. As expected, several types of interactions could occur between proteins and matrix components that might hide the enzymatic cleavage sites and hamper proteins digestion therefore complex mechanisms might

View Article Online
DOI: 10.1039/C6FO01788F
Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.
Food & Function Accepted Manuscript

occur to unravel the potentially allergenic proteins from their belonging matrix. The more complex static model selected for *in vitro* simulation of the human gastro-duodenal digestion utilized in this work and employing a complex salivary, gastric and duodenal fluid composition (electrolyte, surfactants, bile salts and typical enzymes), together with the focus placed on investigating the digestibility of soybean proteins within the whole food matrix, unequivocally discloses the advances provided by our work on this specific food matrix.

In addition, our study confirmed the gastro-duodenal resistance of trypsin inhibitor and 2S albumin (detected in band "*d*", "*f*" and "*l*" of undigested, gastric and duodenal sample, respectively) to the GI proteolytic activity exerted by pepsin, trypsin and chymotrypsin enzymes.[28,29,31]

3.3. Assessment of residual immunoreactivity potential of *in vitro* digested soy

Most of the soybean proteins discussed so far, for their different resistance to proteolysis were universally recognized for their allergenic potential inducing IgE-mediated adverse reactions upon ingestion. Therefore, the final section of our work was aimed at investigating the immunoreactive potential of the digested soybean proteins by (linear) epitopes, scouting along the resistant peptide sequences identified, through bioinformatics tools. All peptides obtained from in-gel tryptic digestion of protein bands excised from the duodenal electrophoretic pattern highlighted in the gel (high molecular weight portion of the duodenal digest), together with the peptides pool contained in the duodenal fluid fraction obtained by ultrafiltration on cut-off membrane devices (low molecular weight portion) were taken into consideration. IEDB database[38] was screened in order to match detected peptides and soybean linear epitopes recognized for *Homo sapiens* host. Results are summarized in table 3 and 4. Combining results of both sets (high and low molecular weight portion of duodenal fluid) allowed to obtain an overview of the potential allergic sequences survived in the final GI digested soybean sample. As reported in table 3, several peptides derived from soybean simulated digestion are part of linear epitope sequences, mainly belonging to β-

View Article Online
DOI: 10.1039/C6FO01788F

conglycinin, codified in the allergens nomenclature as Gly m5. This confirmed the chance that intact protein fragments could cross the intestinal mucosa with conserved epitope sites potentially able to trigger immunological reactions. Only two epitopic peptides were identified belonging to glycinin proteins codified as Gly m6, banding at 20 kDa. On the contrary, most of the peptides in the low molecular weights portion (<10 kDa) reported in table 4 matched with epitopic sites of glycinin (Gly m6); therefore even if the glycinin chains were more extensively hydrolyzed still intact epitope sequences were retained in the detected peptides even after 180 min duodenal digestion thus posing a potential risk for sensitive individuals.

In general this data turn very interesting from a toxicological perspective. We observed that several soy allergens appeared to be stable to the *in vitro* simulated gastro-duodenal digestion and a wide range of peptides produced along digestion retain epitopic sequences, totally or partially conserved along the primary structure of the resistant fragments. According to this, a certain allergenic potential is likely to be retained in the final soy digest, thus providing a partial confirmation to the correlation between allergenicity and digestibility. This study should be further extended with immunological studies in order to have more insights on the actual residual immunoreactivity of digested soy sample and to evaluate the inductive effect on the mucosal immune system during the allergenic proteins transit along the gastrointestinal tract.

Finally, the epitopic peptides identified in the duodenal protein bands above 10kDa and in the duodenal fluid fraction below 10kDa were subjected to a *BLAST* search in UniprotKB/Swiss Protdatabase[39] looking for any other legume crop sharing the same sequences. All proteins matching for at least six consecutive amino acids belonging to the epitopic sequence, were taken into consideration. Results, illustrated in figure 3, revealed that among the sixteen duodenal peptides recognized as soy epitopes (or part of them), six matched also with proteins coming from other plants belonging to *Fabacee* family, renowned as the major food allergen source. Specifically, glycinins shared peptides with *Pisum sativum* and *Cicera retinum*, and β-conglycinin subunits

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.
Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

showed common peptides with *Pisum sativum* as well, *Canavalia* (species gladiata and ensiformis), *Arachis hypogaea*, *Lupinus* (species angustifolius and albus), *Phaseolus* (species vulgaris and lunatus) and *Vicia faba*. The investigation was also extended to the whole protein sequence and the percentage of identity between soybean proteins retaining epitopic sites and the corresponding alternative legume proteins was calculated. Values equal or higher to 51% were found for most of the matches previously identified, except for soybean protein P11827 *vs Lupinus angustifolius* F5B8W1 and F5B8W2, whose identities were 43 and 33%, respectively. Identity percentages lower than 50% were also displayed by matching soybean protein P13916 with the proteins F5B8W (*Lupinus angustifolius*), Q43617 andP80463 (*Phaseolus lunatus*), P43237andP43238 (*Arachys hypogaea*). These preliminary results could be very interesting from a toxicological point of view because the ingestion of other legume proteins might originate, upon digestion, similar resistant peptides displaying epitope sharing with the soybean allergens, which could be harmful for a wide category of allergic individuals.

In addition, the high percentage of identity (>50%) found between duodenal soybean proteins and some legume proteins (as discussed above) suggests that a potential cross-reactivity between these legume crops could occur. Indeed, according to what reported by Goodman et al. 2016,[43] the *local alignment method* (BLASTP or FASTA) with identity scores greater than 50%, could reliably predict potential risks of allergy or cross-reactivity. Anyway serum testing needed to be performed to confirm these results, as recommended by the CODEX Alimentarius Commission in 2003.[44] Several studies already reported the occurrence of a significant cross reactivity between soybean with pea, white bean, peanut, lentil, fennel, guar gum, carob beam, tragacanth, chickpea and liquorice.[45] In our study also proteins belonging to species of *Cicer*, *Canavalia*, *Lupinus*, *Phaseolus* and *Vicia* were found to share epitopic sequences with soybean allergen proteins, and for most of them a percentage of identity higher than 50% was observed. These results pose the need for further in-depth investigation on allergenicity assessment of proteins belonging the *Fabacee* family

Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F
Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.
Food & Function Accepted Manuscript

according to the scientific opinions 2010 issued by EFSA,[22] which however is outside the scope of the present paper.

4.Conclusions

The present work aimed at providing a realistic overview on the digestibility and the residual allergenic potential of soybean proteins by submitting soy flour to an *in vitro* digestion model that included the chewing, gastric and duodenal phases. Soybean digestion products were analyzed by SDS-PAGE and characterized by untargeted HPLC-MS/MS analysis. Focusing on the main allergenic proteins, our data demonstrated that glycinins subunits, above all the acid polypeptides, were more susceptible than β-conglycin into the GI enzymatic proteolysis. Large fragments of the β-conglycinin, survived to the combined activity of pepsin, trypsin and chymotrypsin, as extensively reported in previous studies. Such behavior was also observed for trypsin inhibitor allergens and 2S albumin, which were found in unequivocally identified electrophoretic bands even at the end of the whole digestion process. Noteworthy, by performing bioinformatic search it was demonstrated that after gastro-duodenal digestion, both β-conglycinin and glycinin protein fragments retain linear epitopes potentially able to trigger immune reaction when exposed to intestinal mucosa. These data provide more insights on the stability of soybean proteins upon gastro-duodenal digestion and on its correlation with potential immunoreactivity of the peptides released. Moreover, the preliminary investigation on the cross-reactivity between soy and other legumes crops, suggested to widen the investigation on allergenicity assessment of proteins belonging the *Fabacee* family, which could represent a risk for a wide category of allergic individuals.

Acknowledgments

The authors acknowledge the technical help of Roberto Schena in performing HPLC/MS analyses.

View Article Online
DOI: 10.1039/C6FO01788F

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

References

1Z. Berk, 1992. Fao Agricultural Services Bulletin No. 97,Technion, Israel Institute of Technology, Haifa (Israel).

2 T. Wang, G.-X. Qin, Z.-W. Sun and Y. Zhao, *Crit. Rev. Food Sci.*, 2014, **54***,*850-862.

3G. S Gilani, C. W. Xiao and K. A. Cockell, *Brit. J. Nutr.*, 2012, **108,** S315-S332.

4 A. W. Burks, J. R. Brooks and H. A. Sampson. *J. Allergy Clin. Immun.*, 1988, **81**,1135-1142.

5 T. Ogawa, N. Bando, H. Tsuji, H. Okajima, K. Nishikawa and K. Sasoka, *J. Nutr. Sci. Vitaminol.,* 1991, **37**, 555-565.

6 T. Ogawa, H. Tsuji, N. Bando, K. Kitamura, Y.-L. Zhu, H. Hirano and K. Nishikawa,. *Biosci. Biotechnol. Biochem.*, 1993, **57**, 1030-1033.

7 Food and Drug Administration (FDA). Food Allergen Labeling and Consumer Protection Act (FALCPA) of 2004. Congressional record v. 150. 2004. http://www.fda.gov/downloads/Food/LabelingNutrition/FoodAllergensLabeling/GuidanceComplianceRegulatoryInformation/ UCM179394.pdf.

8 European Commission. Commission Directive 2007/68/EC of 27 November 2007 amending Annex IIIa to Directive 2000/13/EC of the European Parliament and of the Council as regards certain food ingredients. *OJEU,* 2007, *L310*, 11-17.

9 WHO/IUIS Allergen Nomenclature Sub-Committee.www.allergen.org. (accessed July 2016).

10 L. A. Moroz and W. H.Yang, *N. Engl. J. Med.*, 1980, **302**, 1126-1128.

11 A. W. Burks, G. Cockrell, C. Connaughton, J. Guin, W. Allen and R. M. Helmet, *Int. Arch. Allergy Imm.*, 1994, **105**, 143-149.

12 T. Ogawa, Samoto, M. and K. Takahashi, *J. Nutr. Sci. Vitaminol.,* 2000, **46**, 271-279.

13 A. Adachi, T. Horikawa, H. Shimizu, Y. Sarayama, T. Ogawa, S. Sjolander and A. Tanaka, T. Moriyama, *Clin. Exp. Allergy*, 2009, **39**, 167-173.

14 P. Sun, D. F. Li, Z. J. Li, B. Dong and F. L. Wang, *J. Nutr. Biochem.,* 2008, **19***,*627-633.

15 H. B. Krishnan, W. S. Kim, S. C. Jang and M.S. Kerley, *J. Agr. Food Chem.*, 2009, **57**, 938-943.

16 P. E. Staswick, M. A. Hermodson and N. C. Nielsen, *J .Biol. Chem.*, 1981, **256***,* 8752-8755.

17 M. Amigo-Benavent, V. I. Athanasopoulos, P. Ferranti, M. Villamiel and M. D.Castillo,*Food Res. Int.*, 2009, **42**, 819–825.

18 R. H. Erickson and Y. S. Kim, *Annu. Rev. Med.*, 1990, **41**,133-139.

19 F. J. Moreno, *Biomed. Pharmacother.,* 2007, **61**, 50-60.

View Article Online
DOI: 10.1039/C6FO01788F

20 K. L. Bøgh and C. B. Madsen, *Crit. Rev. Food Sci.,* 2016, **56**, 1545-1567.

21M. Minekus, M. Alminger, P. Alvito, S. Ballance,T. Bohn,C. Bourlieu, F. Carrière, R. Boutrou, M. Corredig, D. Dupont, C. Dufour, L. Egger, M. Golding, S. Karakaya,B. Kirkhus, S. Le Feunteun, U. Lesmes, A. Macierzanka, A. Mackie, S. Marze, D. J. McClements, O. Ménard, I. Recio,C. N. Santos, R. P. Singh,G. E. Vegarud, M. S. J. Wickham, W. Weitschies and A. Brodkor, *Food Funct*.2014, **5**, 1113–1124.

22EFSA Panel on Genetically Modified Organisms (GMO). Draft Scientific Opinion on the assessment of allergenicity of GM plants and microorganisms and derived food and feed. *EFSA Journal*, 2010, **8**, 1-168.

23F. J. Moreno, A. R. Mackie and E. N. C Mills, *J. Agr. Food Chem.,* 2005,**53***,* 9810-9816.

24F. J. Moreno, B. M. Maldonado, N. Wellner and E. N. C. Mills,.*Biochim. Biophys. Acta,* 2005, **1752,**142-153.

25F. J. Moreno, F. A Mellon, M. S. Wickham, A. R. Bottrill and E. N. C. Mills, *FEBS J.,* 2005,***272****,*341-352.

26M. Bublin, C. Radauer, A. Knulst, S. Wagner, O. Scheiner, A. R. Mackie, E. N. C. Mills, and H. Breiteneder, *Mol. Nutr. Food Res.,* 2008, **52**,1130-1139.

27 J. Czubinski, A. Siger, E. Lampart-Szczapa, *Eur. FOOD Res. Technol*. 2016, **242**, 391–403.

28J. D. Astwood, J. N. Leach and R. L. Fuchs, *Nat. Biotechnol.*, 1996, **14**, 1269-1273.

29D. Sung, K. M. Ahn, S.-Y. Lim and S. Oh,. *J. Sci. Food Agr.,* 2014, **94***,* 2482-2487.

30 T. J. Fu, *Ann. Ny. Acad. Sci.,* 2002, **964***,*99-110.

31K. Thomas, M. Aalbers, G. A. Bannon, M. Bartels, R. J. Dearman, D. J. Esdaile, T. J. Fu, C. M. Glatt, N. Hadfield, C. Hatzos, S. L. Hefle, J. R. Heylings, R. E. Goodman, B. Henry, C. Herouet, M. Holsapple, G. S. Ladics, T. D. Landry, S. C. MacIntosh, E. A. Rice, L. S. Privalle, H .Y. Teiner, R. Teshima, R. van Ree, M. Woolhiser and J. Zawodny, *Regul.Toxicol. Pharm.,* 2004, **39***,* 87-98.

32M. Amigo-Benavent, A. Clemente, P. Ferranti, S. Caira and M. D. del Castillo, *Food Chem.,* 2011, **129**, 1598-1605.

33M. Defernez, G. Mandalari andE.C.N. Mills,*Electrophoresis,*2010,**31**, 2838-2848.

34 A.Macierzanka, A. I.Sancho, E. C.Mills, N. M.Rigby, and A. R. Mackie, *Soft Matter*, 2009,**5**, 538-550.

35G.Mandalari,K.Adel-Patient, V. Barkholt, C. Baro, L. Bennett, M. Bublin, S.Gaier, G. Graser, G.S. Ladics, D. Mierzejewska, E. Vassilopoulou, Y.M.Vissers, L. Zuidmeer, N.M. Rigby, L.J. Salt, M. Defernez, F. Mulholland, A.R. Mackie, M.S.J. Wickham, E.N.C. Mills, *Regulatory Toxicology and Pharmacology* 2009, **55**, 372-381.

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

36G.Mandalari, A. M.Mackie, N. M.Rigby, M. S.Wickham, and E. N. Mills, *Molecular nutrition & food research*, 2009,**53**, S131-S139.

37L. Monaci, R. Pilolli, E. De Angelis, M. Godula and A.Visconti, *J.Chromatogr. A,* 2014, **1358**, 136-144.

38Immuno Epitope Database and Analysis, http://www.iedb.org. (accessed July 2016).

39Expasy. Bioinformatics Resource Portal. http://web.expasy.org/blast. (accessed July 2016).

40 P. Amnuaycheewa and E. G. de Mejia, *Food Chem.*, 2010, **119**, 1671-1680.

41X. Sun, X. Shan, Z. Yan,Y. Zhang and L.Guan, *Food Chem.Toxicol.,*2013, **56***,*254-260.

42 Y. Zhao, G. X. Qin, Z. W. Sun, B. Zhang and T. Wang, *Food Agr. Immunol.,*2010, **21**, 253-263.

43 R. E. Goodman, M. Ebisawa, F. Ferreira, H. A. Sampson, R. Ree, S. Vieths, J. L. Baumert, B. Bohle, S. Lalithambika, J. Wise and S. L. Taylor, *Mol. Nutr. Food Res.,* 2016,**60***,* 1183-1198.

44 CODEX Alimentarius Commission. Proposed draft annex on the assessment of possible allergenicity of the draft guideline for the conduct of food safety assessment of foods derived from recombinant DNA plants. *Joint FAO/WHO Food Standard Program*, *CAC/GL,* 2003, **45**, 57–60, Appendix IV.

45European Federation of Asthma and Allergy Associations, *Food Allergy Brochure EFA, Box 5*, 1997, 3830 Leusden, Netherland.

46R. M. Helm, G. Cockrell, C. Connaughton, H. A. Sampson, G. A. Bannon, V. Beilinson, N. C. Nielsen and A. W. Burks, *Int. Arch. Allergy.Imm.,* 2000, **123**, 213-9.

47H. Saeed, C. Gagnon, E. Cober and S. Gleddie, *Mol.Immunol.,*2016, **70***,* 125-33.

48P. Xiang, T. A. Beardslee, M. G. Zeece,J. Markwell andG. Sarath, *Arch.Biochem.Biophys.*, 2002, **408**, 51-57.

49T. ABeardslee, M. G.Zeece, G. Sarath and J. P. Markwell, *Int. Arch. Allergy Imm.,*2000, **123**, 299–307.

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

TABLES AND FIGURES CAPTIONS

Figure 1. Workflow of *in vitro* gastro-duodenal model implemented in the current study.

Figure 2. Typical SDS-PAGE protein profile of undigested, gastric and duodenal fluids obtained by submitting soybean flour to *in-vitro* digestion. A4-20% precast gel (8.6cm x 6.7cm x1mm) was used for protein separation. Lines refer to different protein contents loaded (3,6=20µg; 1,4,7=50µg; 2,5,8 =70µg) and they were grouped according to the different digestion fluids analyzed (A= undigested; B= gastric aliquot collected at the end of gastric digestion-120 min; C= duodenal aliquot collected at the end of whole digestion-180 min).

Figure 3. Comparison between epitopic peptides identified in soybean duodenal digest (grey boxes) and proteins belonging to legume species showing similarity of at least six consecutive amino acids with immunogenic soybean peptides, obtained through BLASTp search

Table 1. Attribution of selected protein bands of undigested, gastric and duodenal samples by LC-MS/MS analysis and software based identification(Proteome Discoverer v. 2.0 software). (*the number of identified peptides were filtered by setting unambiguous peptide-spectrum matches and a300ppmtolerance on the precursor ion).

Table 2. List of peptides identified in the duodenal digest (D=180 min) fraction with molecular weight below10kDa (* protein recognized as allergen according to Allergome platform).

Table 3. List of potential immunogenic sequences recognized in the GI resistant peptides identified in specific electrophoretic bands of the duodenal digest (D180, molecular weight >10 kDa).Bold sequences refer to soybean epitope sites reported in IEDB with relevant epitope ID.

Table 4. List of potential immunogenic sequences recognized in the low molecular weight peptides fraction of soybean duodenal digest (D180, molecular weight <10 kDa).Bold sequences refer to soybean epitope sites reported in IEDB with relevant epitope ID.

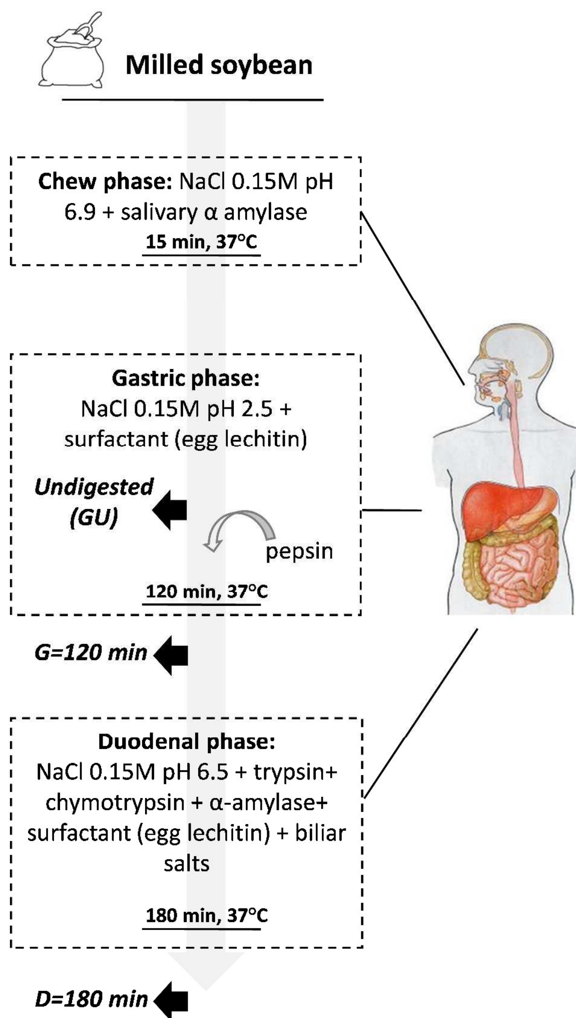
Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Figure 1

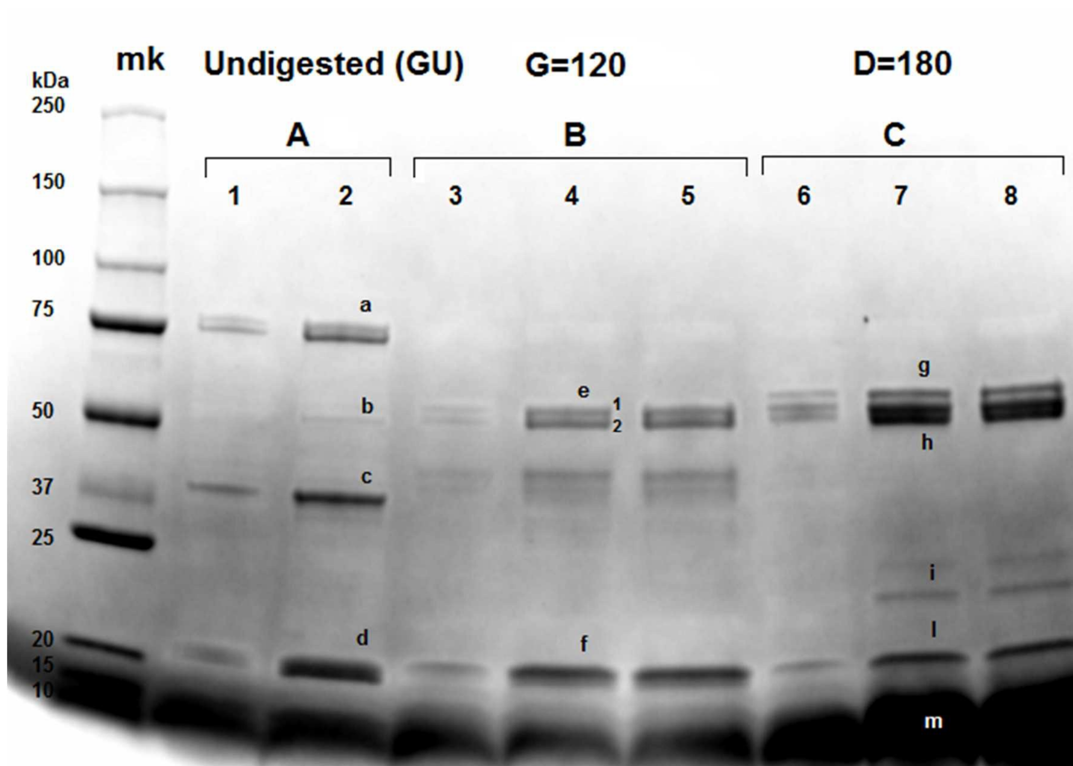


Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Figure 2



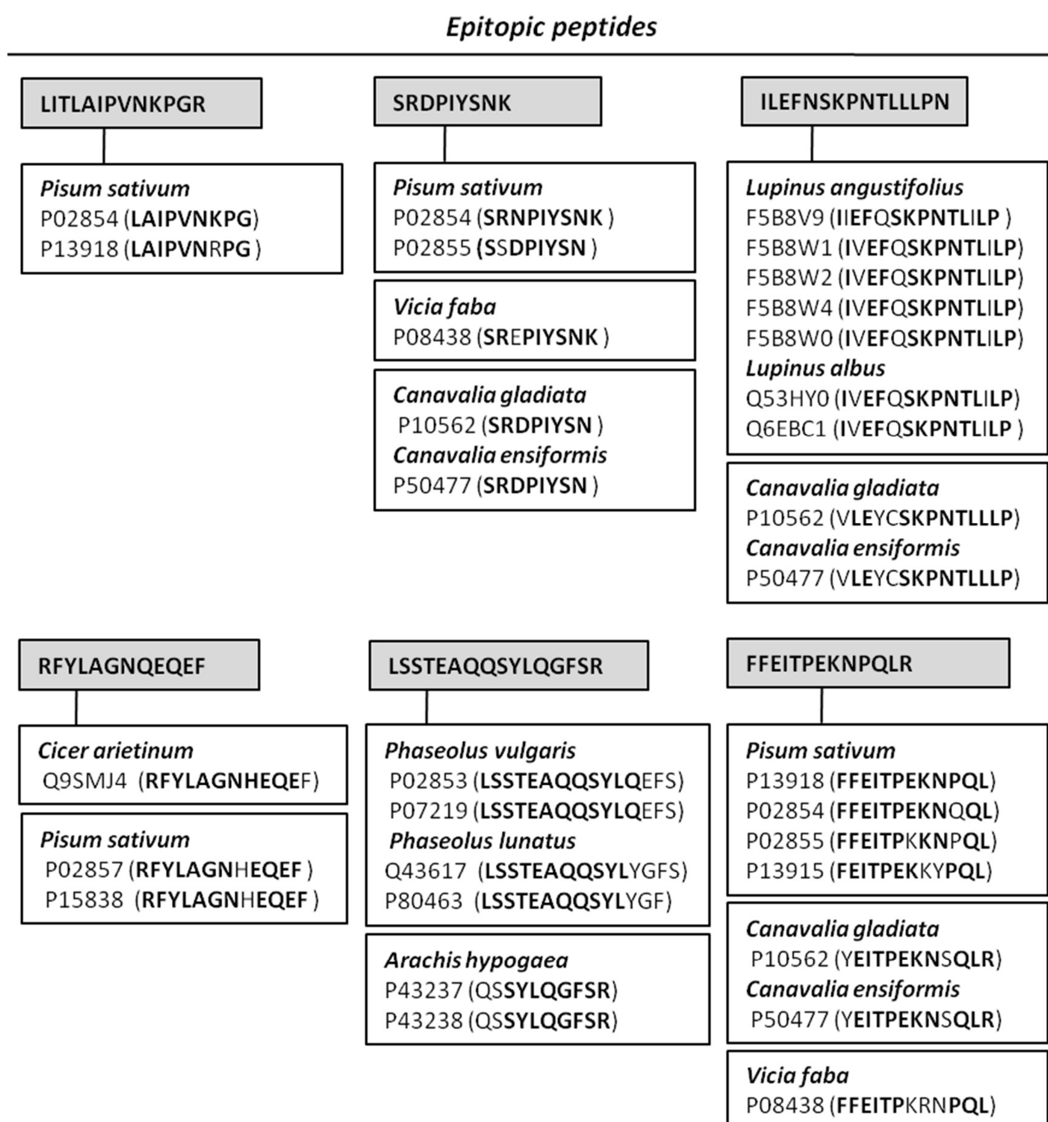
Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

Figure 3.



View Article Online
DOI: 10.1039/C6FO01788F

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

Table 1.

Sample	Band	Accession number	Type of protein	Score	Coverage	Filtered Peptides* (Unique)
Undigested	a	P11827	β-conglycinin-α'chain	219.94	45%	15 (8)
		P13916	β-conglycinin-α chain	364.60	45%	29 (18)
	b	P25974	β-conglycinin-β chain	128.16	44%	13 (10)
		P13916	β-conglycinin-α chain	111.04	28%	12 (7)
		P11827	β-conglycinin-α' chain	43.88	13%	5 (4)
	c	P04776	Glycinin G1	226.46	52%	16 (12)
		P04405	Glycinin G2	305.18	37%	12 (7)
		P02858	Glycinin G4	149.10	29%	8 (6)
		P04347	Glycinin	93.01	23%	3 (2)
		P11828	Glycinin G3	140.15	23%	8 (3)
		P05046	Lectin	42.94	16%	4 (4)
	d	P04776	Glycinin G1	261.35	41%	22 (12)
		P04405	Glycinin G2	332.08	44%	23 (12)
		P11828	Glycinin G3	205.47	28%	12 (1)
		P02858	Glycinin G4	171.92	31%	11 (9)
		P04347	Glycinin	98.55	20%	6 (4)
		P01070	Trypsininhibitor A	87.34	51%	12 (8)
		P19594	2S albumin	42.43	22%	5 (5)
G=120	e	P13916	β-conglycinin-α chain	302.00	38%	20 (10)
		P11827	β-conglycinin-α'chain	256.80	29%	21 (13)
		P25974	β-conglycinin-β chain	290.75	54%	19 (14)
	f	P01070	Trypsininhibitor A	268.42	69%	12 (6)
		P01071	TrypsininhibitorB	86.04	49%	6 (0)
		P04776	Glycinin G1	215.24	41%	12 (3)
		P04405	Glycin G2	223.51	44%	11 (6)
		P11828	Glycinin G3	151.48	27%	10 (2)
		P02858	Glycinin G4	137.91	33%	10 (8)
		P04347	Glycinin	66.31	26%	7 (5)
		P25272	Kunitz-typetrypsininhibitor 1	41.13	26%	4 (2)
		P25273	Kunitz-typetrypsininhibitor 2	29.25	24%	4 (3)
		P19594	2S albumin	22.00	22%	4 (4)
D=180	g	P13916	β-conglycinin-α chain	144.35	27%	11 (7)
		P11827	β-conglycinin-α'chain	62.48	17%	5 (2)
	h	P13916	β-conglycinin-α chain	665.60	40%	20 (13)
		P11827	β-conglycinin-α'chain	450.17	38%	16 (5)
		P25974	β-conglycinin-β chain	310.62	63%	17 (13)
	i	P13916	β-conglycinin-α chain	216.38	32%	11 (7)
		P11827	β-conglycinin-α'chain	171.60	28%	11 (6)
	l	P01070	Trypsininhibitor A	228.35	67%	15 (10)
		P04776	Glycinin G1	164.38	44%	9 (7)
		P04405	Glycinin G2	145.25	38%	5 (1)
		P11828	Glycinin G3	105.10	26%	5 (1)
		P02858	Glycinin G4	72.81	23%	7 (4)
		P13916	β-conglycinin-α chain	21.55	12%	3 (3)
		P19594	2S albumin	13.06	22%	3 (3)
	m	P19594	2S albumin	51.13	30%	7 (7)
		P02858	Glycinin G4	41.06	12%	5 (2)

View Article Online
DOI: 10.1039/C6FO01788F

P01070	Trypsininhibitor A	39.14	43%	5 (2)
P01071	Trypsininhibitor B	12.63	16 %	3 (0)
P11827	β-conglycinin-α'chain	22.92	12%	3 (1)

Table 2.

Peptide sequence	**Protein name (Uniprot code)**
VVEDLPEGPAVKIGE, SVVEDLPEGPAVKIGE, ISIDHDDGTRRL	Trypsin inhibitor A* (P01070), Trypsin inhibitor B* (P01071)
VLDNEGNPLENGGTY, NKDAMDGW, DNEGNPLENGGTY	Trypsin inhibitor A* (P01070)
VSFKTNDTPMIGT, KTNDTPMIGTL, KNNNPFKF, LQGENEGEDKGAIVT, KYQQEQGGHQSQKGKHQQEEENEGGSIL, LKYQQEQGGHQSQKGKHQQE, LKPDNRIESEGGL, SVIKPPTDEQQQRPQE, IGQTSSPDIYNPQAGSVTT, SSPDIYNPQAGSV, IKNNNPFKF, LEFLEHAF, TLEFLEHAFS, IYNPQAGSVTTA	Glycinin G1* (P04776)
IGQNSSPDIYNPQAGSITT, GRSQRPQDRHQK, FAPEFLKEAFG, FAPEFLKEAFGVNMQ, YQEPQESQQRGRSQRPQDRHQK, IGQNSSPDIYNPQAGSI, SSPDIYNPQAGSIT, IGQNSSPDIYNPQAGSITTA, VPHYTL, LQGENEEEDSGAIVTVK	Glycinin G2* (P04405)
LLNALPEEVIQHTFN, LLNALPEEVIQHTF, ALPEEVIQHTF, LLNALPEEVIQH, IIDTNSLENQLDQMPRR, LNALPEEVIQHTF, SIIDTNSLENQLDQMPRRF, IIDTNSLENQLDQMPRRF, NALPEEVIQHTF, FREGDLIAVPTGVAW, LLNALPEEVIQHT	Glycinin G1* (P04776), Glycinin G2* (P04405)
FAPEFLEHAF, LNALPEEVIQQTF, IGQTSSPDIFNPQAGSIT, SSPDIFNPQAGSITT, IDTNSFQNQL, FEEPQQKGQSSRPQDRHQK, LLNALPEEVIQQTF	Glycinin G3* (P11828)
NALKPDNRIESEGGF, LKPDNRIESEGGF, VKNNNPFSF, RIESEGGF, VKNNNPFS, KTNDRPSIGNL	Glycinin G2* (P04405), Glycinin G3* (P11828)
YNNEDTPVVA, IETWNPNNKPFQ, IETWNPNNKPF, YLAGNQEQEF, ERVFDGEL, WNPNNKPFQ, NGERVFDGEL	Glycinin G1* (P04776), Glycinin G2* (P04405), Glycinin G3* (P11828)
EDDEDEQIP, WGPLVNPESQQGSPR, LLDTSNFNNQ, PSYSPYPRM, VTRGQGKVRV, KYEGNWGPL, LAGNPDIEY, FNTNEDIAEKL, DQTPRVFY, YLAGNPDIEY, VVAEQAGEQGFE, LDTSNFNNQLDQTPRVFY	Glycinin G4* (P02858)
YLAGNPDIEHPETM, IIVVQGKGAIG, KYQGNSGPLVNP, LDQNPRVFY, IETWNSQHPEL, SHLPSYLPYPQM, KLRSPDDERKQ, YLAGNPDIEHPET	Glycinin* (P04347)
SGFSKHF, IVTVEGGLSVISPK	Glycinin* (P04347), Glycinin G4* (P02858)
VLHEAI, NKVDENGTPKPSSL, LDIPGESHDVL, RNSWDPPNPHIG, IVTSSGKLQL, LAPIDTKPQTHAGY	Lectin* (P05046)
EWDGSMGIPGAF, TIIPLPVIKE, NTSLPTLGAGE, INNHEKAYL, LSLPHSAGDL	Seed Linoleate 13S-lipoxygenase-1 (P08170)
ITKEHLEPNL, LKNDGTLRP, EDIPNHGSIHF, DLNFTPRE, LSLPHPQGDQSGAF	Seed Linoleate 9S-lipoxygenase-3 (P09186)
LSLPHPAGDLSG, LANGKGKVGKDTF, EDVPNQGTIRF, TIMPLPVVKE	Seed Linoleate 9S-lipoxygenase-2 (P09439)
IEDYPYAVDGL	Seed Linoleate 9S-lipoxygenase-2

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.
Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

	(P09439), Seed Linoleate 9S-lipoxygenase-3 (P09186)
WWYKVENHAAELTAGYY, LRYPSYPQSQG, EATKPTLPFPW, LEATKPTLPFPW, IVNIPIPQW	β-amylase (P10538)
LSVVDMNEGALFLPH, VVDMNEGALFLPH	β-conglycinin-α' chain* (P11827)
LRVPSGTTY, VVNPDNNENLRL, IVDMNEGALLLPHFNSKA, ERQFPFPRPPHQKE, FSRNILEASY, RLQSGDALRVPSGTTY, RLQSGDALRVPSGTTYY, YVVNPDNNENLRL	β-conglycinin-α chain* (P13916)
ILEASYDTKFEEINKVL, LSEQDIFVIPA, RAELSEQDIFVIPA	β-conglycinin-α' chain* (P11827), β-conglycinin-α chain* (P13916)
VLEHGGIATDDDYPY, EWVLEHGGIAT, WVLEHGGIAT, KKMKKEQY	P34 probable thiol protease* (P22895)
YVGQKTKEVGQKTKE, YVGQKTKEVGQKTKEVGQD, IYHSERGPTTSQVL	P24 oleosin isoform A (P29530), P24 oleosin isoform B (P29531)
VHTTTHRYEAG	P24 oleosin isoform B (P29531)
FDQQNEGSIF, LVSESETEKITLEPGDMIHIPAGTPLY, IHIPAGTPLY, IVNRDENDKL, KITLEPGDM, ALQTPKGKL, DLKPGMV, EKITLEPGDM, TRVETEGGRIRV	Sucrose binding protein (Q04672)
HHTNPTKPINL, HSVFPPNK, LSDSVPIPQHH	Basic 7S globulin 2 (Q8RVH5)
V**M**DKPNGPVW, VMDKPNGPVW, VDLNGNHL, MDKPNGPVW, LGHAPISLPNQ, LGHAPISL, LGHAPISLPNQL	Basic 7S globulin 2 (Q8RVH5), Basic 7S globulin (P13917)

Table 3.

Band	Peptide sequence	Epitope ID	Epitopicsequence (Allergen Source)	References
g	**TLAIPVNKPGR**	181524	YYVVNPDNNENLRLI**TLAIPVNKPGR**FES (Gly m 5)	42
	LITLAIPVNKPGR			
	FFEITPEK	181292	DPIYSNKLGK**FFEITPEK**NPQLRDLD (Gly m 5)	42
		181455	SEDKPFNLRSRDPIYSNKLGK**FFEITPEK**N (Gly m 5)	42
h	**ITLAIPVNKPGR**	181524	YYVVNPDNNENLR**LITLAIPVNKPGR**FES (Gly m 5)	42
	TLAIPVNKPGR			
	LITLAIPVNKPGR			
	LITLAIPVNK			
	ILEFNSKPN	181386	LQNLRDYR**ILEFNSKPN**TLLLPNHAD (Gly m 5)	42
	SSTEAQQSYLQGF	181316	FF**LSSTEAQQSYLQGFSR**NILE (Gly m 5)	42
	LSSTEAQQSYLQGFSR			
	NPFLFGSNR	181406	NK**NPFLFGSNR**FETLFKNQYGRIRVLQRF (Gly m 5)	42
	FFEITPEKNPQLR **EITPEKNPQLR** **FFEITPEK**	181292	DPIYSNKLGK**FFEITPEKNPQLR**DLD (Gly m 5)	42
		181455	SEDKPFNLRSRDPIYSNKLGK**FFEITPEKN** (Gly m 5)	42
i	**FETLFK**	181406	NKNPFLFGSNR**FETLFK**NQYGRIRVLQRF (Gly m 5)	42

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

	ILEFNSKPNTLLLPN	181386	LQNLRDYR**ILEFNSKPNTLLLPN**HAD (Gly m 5)	42
	ILEFNSKPN			
	SRDPIYSNK	181455	SEDKPFNLR**SRDPIYSNK**LGKFFEITPEKN (Gly m 5)	42
1	**LSAQYGSLR**	36948	LK**LSAQYGSLR**KNAM (Gly m 6)	43
	RFYLAGNQEQEF	53810	**RFYLAGNQEQEF**LKY (Gly m 6)	43
	NPFLFGSNR	181406	NK**NPFLFGSNR**FETLFKNQYGRIRVLQRF (Gly m 5)	42

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript

View Article Online
DOI: 10.1039/C6FO01788F

Table 4.

Peptide sequence	Epitope ID	Epitopic sequence (allergen source)	References
SGFSKHF	538677	GSVL**SGFSKHFL (Gly m6)**	44
GRS**QRPQDRHQK**	538745	**QRPQDRHQK (Gly m6)**	44
YQEPQESQQRGRS**QRPQDRHQK**			
FAPEFLKEAFG	58026	SG**FAPEFLKEAFG**VN**(Gly m6)**	45
VPHYTL	41548	MF**VPHYTL**NANSIIY **(Gly m6)**	43
	59438	SLRKNAMF**VPHYTL**N **(Gly m6)**	
	70332	**VPHYTL**NANSIIYAL **(Gly m6)**	
LEFLEHAF **TLEFLEHAFS**	19632	GFT**LEFLEHAFS**V **(Gly m6)**	45
	58037	SGFT**LEFLEHAFS**VD **(Gly m6)**	46
VKNNNPFS	33488	KSQQARQ**VKNNNPFS(Gly m6)**	43
YLAGNQEQEF	53810	RF**YLAGNQEQEF**LKY **(Gly m6)**	43
VVNPDNNENLRL	181524	Y**YVVNPDNNENLRL**ITLAIPVNKPGRFES **(Gly m5)**	42
YVVNPDNNENLRL			

Published on 06 March 2017. Downloaded by CNR on 08/03/2017 14:34:09.

Food & Function Accepted Manuscript