



Cite this: *Food Funct.*, 2017, **8**, 1599

Insight into the gastro-duodenal digestion resistance of soybean proteins and potential implications for residual immunogenicity†

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

Soy is an important component of the human diet thanks to its nutritional value and high protein content; however, it also represents a risk for allergenic consumers due to its potential to trigger adverse reactions in sensitized individuals. The putative correlation between immunoreactivity and resistance to the human gastrointestinal (GI) digestion has drawn attention to investigating soybean proteins digestibility. In this work, we provided further insights into 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 the final data were software 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, showing a partial degradation at the end of the duodenal phase as ascertained by the strong electrophoretic bands displayed at 50 kDa and 20 kDa, respectively. Glycinin sub-units also presented, even if to a lower extent, resistance to the complete proteolytic degradation, as demonstrated by polypeptide fragments with molecular weight lower than 20 kDa displayed in the gel at the end of duodenal digestion. In addition, by bioinformatics 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 also presented 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.

Received 12th December 2016,

Accepted 3rd March 2017

DOI: 10.1039/c6fo01788f

rsc.li/food-function

1. Introduction

Soybean (*Glycine max*) is considered a source of high-quality proteins, fiber and essential fatty acids, as well as vitamins and minerals, therefore its cultivation and use in food industries is 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 be induced 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 on 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), defensin (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 the 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 30 K and Gly m Bd 28 K 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 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 pro-

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

E-mail: linda.monaci@ispa.cnr.it; Fax: +39 080 5929374; Tel: +39 080 5929343

†Electronic supplementary information (ESI) available. See DOI: 10.1039/c6fo01788f

teins 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 of a specific acidic (A) polypeptide chain (40 kDa) linked through disulfide bonds to a specific basic (B) polypeptide chain (20 kDa) and can be one of five subunits (glycinin G1: A_{1a}B_{1b}, glycinin G2: A₂B_{1a}, glycinin G3: A_{1b}B_{1a}, glycinin G4: A₅A₄B₃ and glycinin: A₃B₄).² β -Conglycinins are 7S trimers with molecular masses of approximately 180 kDa. They are glycoproteins containing 5% of carbohydrate moieties and have been reported to induce immunoreactivity.¹⁷ These trimers are formed by various combinations of three homologous polypeptide subunits α' , α and β , with masses of 76 kDa, 72 kDa, and 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 are hydrolyzed down to single amino acids and/or small peptides during digestion, thus facilitating their absorption along the intestinal mucosa.¹⁸ On the contrary, some allergenic proteins could partially resist proteolytic activity of enzymes along the gastrointestinal tract (GI), surviving as large immunologically active fragments and inducing 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 the 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 an approach was also recommended by the EFSA panel (2010) because it allows obtaining *in vitro*, the digestion products to which the inductive mucosal immune system is exposed, 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 models) 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 directed to investigate the stability of the major soybean proteins towards different GI enzymes.^{28–31} Mostly, purified allergenic proteins were subjected to very simplified *in vitro* digestion procedures where proteolytic sensitivity was evaluated by simple exposure to single and/or combined enzyme mixtures. 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 deglycosylated form.³² The authors demonstrated the partial survival of the α and β -subunits of the glycosylated form after 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 gastrointestinal 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 bioinformatics searches, the generated GI products were further investigated for the presence of known linear epitopes that survived enzymatic proteolysis, inferring 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 for adverse cross-reaction for a larger category of allergic consumers.

2. Materials and methods

2.1. Materials 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). Formic acid (MS grade) was provided by Fluka (Milan, Italy) while PTFE syringe filters (4 mm, 0.2 μ m) and ultrafiltration (UF) tubes with 10 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 – BioExtra $\geq 99.5\%$, ammonium bicarbonate, calcium chloride, Bis-Tris, phenylmethanesulfonyl fluoride – PMSF, egg lecithin – PC, sodium taurocholate and sodium glycodeoxycholate) as well as enzymes (α -amylase from human saliva Type XIII-A, pepsin from porcine gastric mucosa, 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 $\times$ 6.7 $\times$ 0.1 cm, 4–20% acrylamide) were purchased from Bio-Rad Laboratories (Segrate, MI, Italy). Ammonium bicarbonate (AMBIC), iodoacetamide (IAA) and 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 digestion was purchased from Promega (Milan, Italy).

2.2. *In vitro* simulated gastroduodenal digestion protocol

Whole soybean flour was subjected 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 170 U mg^{-1} protein substrate, and this value was selected as the average of the used amounts in published studies by the IFR.^{33–36} Concerning the 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 the data published from previous investigations carried out at the IFR.^{34–36} Fig. 1 shows a simplified scheme of the whole procedure.

Before undergoing *in vitro* digestion, soybean seeds were firstly milled coarsely (Bühler, Bühler s.p.A. Segrate, Milano) and successively more finely using a 1 mm sieve. The soybean

protein content was expected to be approximately 37%, according to what was reported by the USDA National Nutrient Database (NDB No. 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, 0.15 M NaCl pH 6.9, 2 U ml^{-1} 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 and 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, 0.15 M NaCl, pH 2.5) and $1383 \mu\text{l}$ of 10 mg ml^{-1} 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^{-1} lecithin stock solution and suspending the pellet in SGF. The solution obtained was then sonicated and filtered ($0.22 \mu\text{m}$ PTFE syringe filter) to remove any titanium residue from the ultrasound probe. Afterward the pH of the simulated gastric sample was lowered to 2.5, using 1 M HCl and the total volume was adjusted to 104.9 mL with SGF. Then the sample was incubated under stirring for 15 min at 150 rpm and 37°C . After this step, a 5 ml undigested (GU) sample was collected and mixed with $500 \mu\text{l}$ of 1 M ammonium bicarbonate buffer before storing on ice. In order to start the gastric proteolytic reaction, $105 \mu\text{l}$ of pepsin stock solution prepared in SGF (170 U mg^{-1} of protein in the final mixture) were added to the gastric mixture and left under incubation on an orbital shaker for 2 h at 37°C and 150 rpm. Finally, an aliquot of the gastric sample was collected and mixed with $500 \mu\text{l}$ of 1 M ammonium bicarbonate to stop proteolysis ($G = 120$). The pH of the whole gastric chyme 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 chyme (expected proteins content = 200 mg) were withdrawn and subjected to the duodenal phase by lowering the pH to 6.5 with the addition of 1.0 M HCl. The digestive fluid was then mixed with $1337 \mu\text{l}$ 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^{-1} lecithin stock solution and suspending the pellet in 0.15 M NaCl (pH 6.5, SDF) together with bile salts (1:1 ratio of sodium taurocholate and sodium glyco-deoxycholate, 7.4 mM bile salts in the duodenal mixture). The mixture was left under stirring at 37°C for about 15 min. Then $22 \mu\text{l}$ of 1 M CaCl_2 and $1179 \mu\text{l}$ of 1 M Bis-Tris (0.5 mM and 26.8 mM in the final duodenal mixture, respectively) were added to the digest. The final pH was adjusted to 6.5 by 1 M NaOH addition and the volume was 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 per mg protein substrate), chymotrypsin (0.4 BTEE units per mg protein substrate) and α -amylase (300 units per ml digestion fluid) stock solutions in SDF were added to the digest sample. The latter was then left

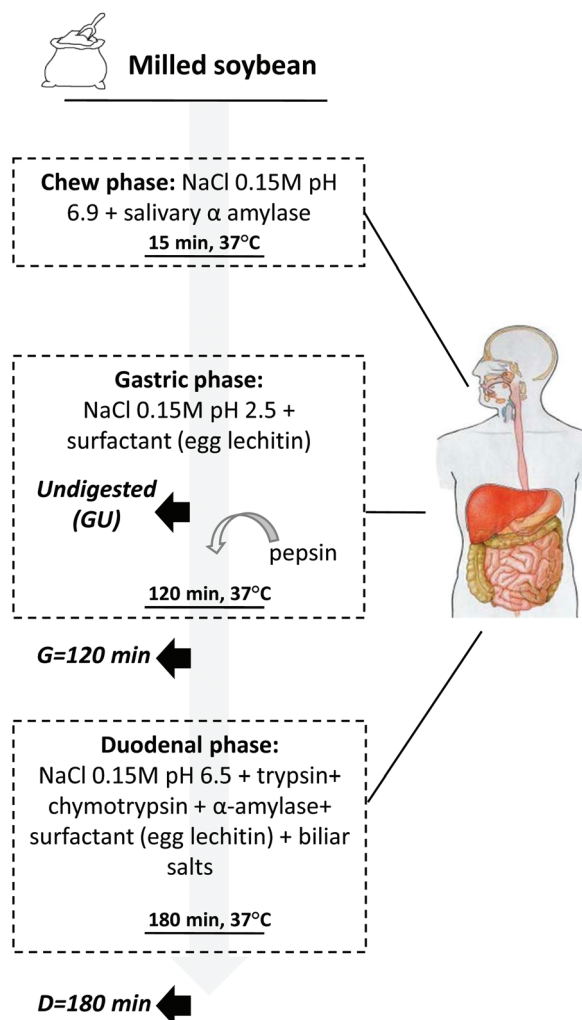


Fig. 1 Workflow of the *in vitro* gastro-duodenal model implemented in the current study.

for 3 h at 37 °C, under stirring (150 rpm). At the end of the duodenal digestion, a 5 ml aliquot was collected (D = 180 min) and mixed with 50 µl of 0.1 M PMSF to stop the enzyme reaction. All the final digests were stored at −20 °C.

2.3. SDS-PAGE analysis

Digested proteins obtained by subjecting milled soybean to gastric and duodenal digestion were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) on 4–20% precast gels (8.6 cm × 6.7 cm × 1 mm). The expected protein content of the samples under investigation was respectively GU = G120 = 5 µg µL^{−1} and D180 = 4.5 µg µL^{−1}. Before gel analysis, the samples were mixed (1 : 1 ratio) with a Laemmli buffer (62.5 mM Tris HCl, pH 6.8, 25% glycerol, 2% SDS, 0.01% bromophenol blue, 100 mM DTT) and then reduced at 100 °C for 5 min. PAGE gels were run in a Mini-Protean Tetra Cell equipment (Bio-rad Laboratories) with TGS (25 mM Tris, 192 mM glycine, 0.1% SDS) running buffer under the following conditions: 50 V for 15 min and 80 V 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 Plus Protein™ 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, under shaking at 500 rpm) with 100 mM AMBIC/acetonitrile (1/1, v/v). Gel slices were then dehydrated for 5 minutes at room temperature in 100 µl 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 in 25 mM AMBIC). Digestion was carried out overnight at 37 °C with mass spectrometry grade trypsin solution (0.1 µg µL^{−1}, enzyme : protein ratio 1 : 50) in 25 mM AMBIC. Successively, gel slices were incubated with 150 µl of Milli-Q water for 10 minutes, with frequent vortex mixing. Then the liquid was removed and transferred into a new microcentrifuge tube. Peptides were extracted from the gel by incubation with 50% acetonitrile/5% trifluoroacetic acid (v/v) for 60 min under shaking at 500 rpm. This step was repeated twice. Peptide mixtures obtained from each extraction step were then pooled together and dried in a “speed Vac” centrifuge. Each sample was suspended in 80 µl of H₂O/ACN 90/10 + 0.1% formic acid (v/v) and 20 µl were further injected into the LC/MS apparatus.

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

Soybean duodenal digests were subjected to purification in 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 the digested sample (duodenal phase) were firstly centrifuged for particulate removal (10 min at 770g) and then 4 ml of the supernatant aliquot was loaded into ultrafiltration tubes. After 4 h of centrifugation (3080g) the sample fraction with molecular weight lower than 10 kDa 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 an ESI interface coupled with a dual pressure linear ion trap mass spectrometer Velos Pro™ (Thermo-Fisher-Scientific, San José, USA).

For chromatographic separation of the peptide mixture generated from tryptic digestion of soy excised protein bands and of the low molecular weight peptide fraction, 20 µl of the sample were injected onto an Aeris peptide analytical column (internal diameter 2.1 mm, length 150 mm, particle size 3.6 µm, porosity 100 Å, Phenomenex) at a flow rate of 200 µl mL^{−1}. The gradient used is: from 85% to 45% 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 = H₂O + 0.1% formic acid; solvent B = CH₃CN + 0.1% formic acid). The 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 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 José, USA); protein identification was achieved by Sequest HT search against a soybean customized database extracted from Swiss-Prot DB based on the taxonomy code of *Glycine max* (ID: 3847) 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 to 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. Bioinformatics analysis for epitope occurrence

Peptide sequences identified by in-gel digestion of protein bands along with low molecular weight peptides detected in the duodenal digestive fluids were finally searched in the IEDB database³⁸ in order to find epitope linear sequences that

survived the gastro-duodenal digestion. The IEDB results were filtered as follows: linear sequence for epitope structure, substring for BLAST option and human as host. Epitopic peptides were then aligned against the UniprotKB/Swiss-Prot database in order to scout around for structural similarity with other legumes.³⁹ The search was performed by setting the E-threshold (number of expected matches in a random database) to 10. The “Gapped alignment” option, which allows gaps to be introduced along the sequences for their comparison, was also activated, while the “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.

3. Results and discussion

Due to the likely connection between the immunoreactivity of certain allergenic proteins and their resistance to the gastrointestinal (GI) digestion, digestibility of soybean proteins was investigated. To this aim, soybean mince was subjected to *in vitro* simulation of the GI process by a static model mimicking three main digestion steps: chewing, gastric and duodenal phases. Salivary, gastric and duodenal fluids composition (in terms of electrolytes, 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 the adult human gut. All enzymes and reagents were added according to basal concentrations of these components (see Fig. 1).

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

Sample aliquots representative of the undigested proteins (GU) and the partially digested proteins (end of the gastric phase, G = 120, and end of the duodenal phase, D = 180) were analyzed in parallel by SDS-PAGE. Fig. 2 shows the resulting Coomassie blue stained gel reporting electrophoretic profiles of three different samples with different amounts of proteins, theoretically calculated: 20 and 50 μ g of proteins for undigested samples and 20, 50 and 70 μ g for both gastric and duodenal digests. The electrophoretic patterns highlighted in the figure and referred to the three digestive phases reflect the progress of the simulated digestion, in terms of proteolytic degradation. As expected, some bands corresponding to high molecular weight proteins had already disappeared after the enzymatic proteolysis, while the number of smaller peptides increased proportionally, as demonstrated by the smeared bands with MW lower than 10 kDa appearing along the duodenal digest lane. In order to have a deeper insight into the identity of the specific soybean proteins exhibiting different levels of resistance to proteolytic activity of characteristic enzymes involved in the gastro-duodenal digestion, untargeted mass spectrometry based analyses followed by software based protein identification were accomplished on selected bands (labelled from *a* to *m* in Fig. 2, lanes A–B–C). The bands were excised from the gel and subjected to in-gel tryptic digestion; the resulting peptides pool was separated by reversed phase HPLC and detected by acquiring the signal in DDA mode. Commercial software for protein identification (Proteome Discoverer™ v. 2.0, Thermo Fisher Scientific) was used to

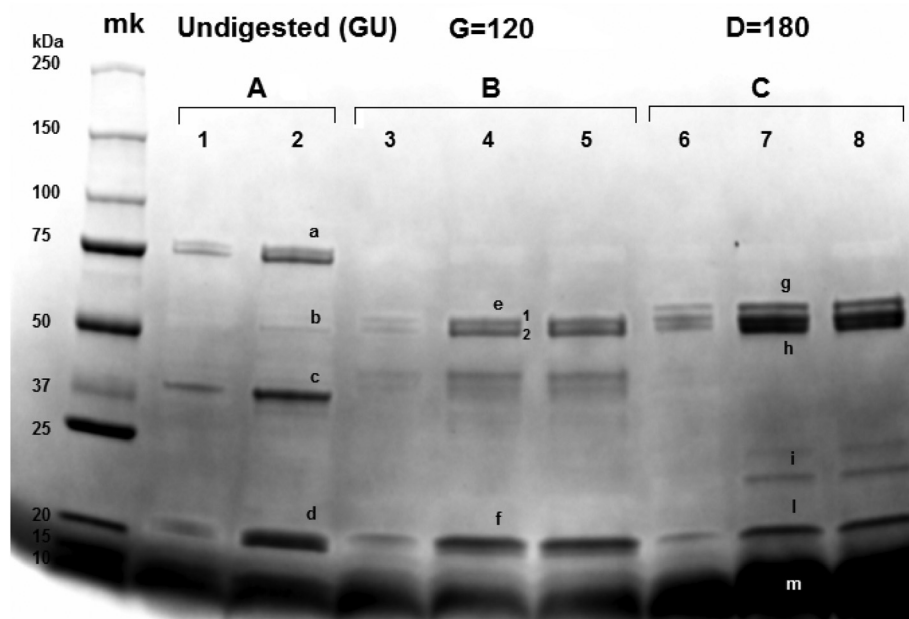


Fig. 2 Typical SDS-PAGE protein profile of undigested, gastric and duodenal fluids obtained by subjecting soybean flour to *in vitro* digestion. A 4–20% precast gel (8.6 cm × 6.7 cm × 1 mm) was used for protein separation. Lanes refer to different protein contents loaded (3, 6 = 20 μ g; 1, 4, 7 = 50 μ g; 2, 5, 8 = 70 μ g) 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).

process the raw MS/MS spectra assigning the electrophoretic bands to specific soybean proteins and/or their subunits/fragments. The attribution was performed using the Sequest HT algorithm through a customized database search. Although bands “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 (bands “a” and “b”) with MW of 50 and 75 kDa, glycinin subunits, acid and basic poly-

peptide chains released by disulfide bonds reduction, (bands “c” and “d”) along with trypsin inhibitors and 2S albumin, with molecular weight between 37 and 20 kDa. Unique protein assignment was not possible due to the low resolution of the 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 band distribution was

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)

Sample	Band	Accession number	Type of protein	Score	Coverage	Filtered peptides ^a (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)
	c	P11827	β -Conglycinin- α' chain	43.88	13%	5 (4)
		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	Trypsin inhibitor 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)
		P01070	Trypsin inhibitor A	268.42	69%	12 (6)
		P01071	Trypsin inhibitor B	86.04	49%	6 (0)
	f	P04776	Glycinin G1	215.24	41%	12 (3)
		P04405	Glycinin 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-type trypsin inhibitor 1	41.13	26%	4 (2)
		P25273	Kunitz-type trypsin inhibitor 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	Trypsin inhibitor 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)
		P01070	Trypsin inhibitor A	39.14	43%	5 (2)
		P01071	Trypsin inhibitor B	12.63	16%	3 (0)
		P11827	β -Conglycinin- α' chain	22.92	12%	3 (1)

^a The number of identified peptides was filtered by setting unambiguous peptide-spectrum matches and a 300 ppm tolerance on the precursor ion.

clearly observed in Fig. 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 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 bands “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 also detected 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 showed new modifications (see Fig. 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 enzyme proteolysis. In addition, new protein fragments (not detected after the gastric phase) assigned to β -conglycinin (α and α' subunits), banding approximately at 20 kDa in molecular weight, were detected at the end of the duodenal digestion marked as “i” in Fig. 2. As for the 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 glycinin degradation occurring 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 15 kDa 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 the duodenal phase (D = 180 min) were passed through an ultra-filtration tube (10 kDa cut off) and the peptide 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* soft-

ware as previously detailed. It is worth noticing that, besides the upper limit in molecular weight fixed by the membrane cut off (10 kDa), there were objective constraints 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 peptide 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 the ESI† for a detailed list of the peptide sequences assigned and the relevant parameters calculated by Proteome Discoverer software. Table 2 highlights that most of the 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 the undigested electrophoretic profile, gradually decreased upon gastro-duodenal digestion and as complementary information, in this low molecular weight fraction, a large number of peptides belonging to the glycinin subunits was detected. On the contrary, only a 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 the β chain, not detected in this sample. Peptides derived from other proteins were also found in the fragment pool below 10 kDa, belonging to trypsin inhibitor, lectin and seed maturation P34 probable thiol protease (Gly m Bd30K).

As a general comment, our insight into 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 with the previous findings suggested in the literature on very simplified case studies, such as purified soy allergens^{28,30,32} and soy protein isolates.⁴¹ As 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 the 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 the undigested sample was substituted by bands “e” in the gastric pattern) and by trypsin and chymotrypsin (bands “i”, “l” and “m”) along the duodenal profile. Our results also agreed with what was reported by Amigo-Benavent (2011)³² inferring about a general stability of the protein β -conglycinin along digestion, attributed to the glycan moieties spread along the molecule thus limiting the accessibility of pepsin to the cleavage sites. As for the glycinin chain group, we observed a partial susceptibility of these proteins to gastric digestion confirmed by the absence in the gastric sample of any band aligned with band

Table 2 List of peptides identified in the duodenal digest (D = 180 min) fraction with molecular weight below 10 kDa

Peptide sequence	Protein name (uniprot. code)
VVEDLPEGPAVKIGE, SVVEDLPEGPAVKIGE, ISIDHDDGTRRL	Trypsin inhibitor A ^a (P01070), trypsin inhibitor B ^a (P01071)
VLDNEGNPLENGGTY, NKDAMDGW, DNEGNPLENGGTY	Trypsin inhibitor A ^a (P01070)
VSKTNDTPMIGT, KTNDTPMIGTL, KNNNPFFK, LQGENEGEDKGAIVT, KYQQEQGGHQSQKKGKHQQEENEGGSIL, LKYQQEQGGHQSQKKGKHQQE, LKPDNRIESEGG, SVIKPPTDEQQRPQE, IGQTSSPDIYNPQAGSVTT, SSPDIYNPQAGSV, IKNNNPFFK, LEFLEHAF, TLEFLEHAFS, IYNPQAGSVTTA	Glycinin G1 ^a (P04776)
IGQNSSPDIYNPQAGSITT, GRSQRPQDRHQK, FAPEFLKEAFG, FAPEFLKEAFGVNMQ, YQEPQESQQRGRSQRQDRHQK, IGQNSSPDIYNPQAGSI, SSPDIYNPQAGSIT, IGQNSSPDIYNPQAGSITTA, VPHYTL, LQGENEEEDSGAIVTK	Glycinin G2 ^a (P04405)
LLNALPEEVIQHTFN, LLNALPEEVIQHTF, ALPEEVIQHTF, LLNALPEEVIQH, IIDTNSLENQLDQMPRR, LLNALPEEVIQHTF, SIIDTNSLENQLDQMPRRF, IIDTNSLENQLDQMPRRF, NALPEEVIQHTF, FREGDLIAVPTGVAV, LLNALPEEVIQHT	Glycinin G1 ^a (P04776), glycinin G2 ^a (P04405)
FAPEFLEHAF, LNALPEEVIQQT, IGQTSSPDIYNPQAGSIT, SSPDIYNPQAGSITT, IDTNSFQNL, FEEPPQKGGSSRPQDRHQK, LLNALPEEVIQQT	Glycinin G3 ^a (P11828)
NALKPDNRIESEGGF, LKPDNRIESEGGF, VKNNNPFFS, RIESEGGF, VKNNNPFFS, KTNDRPSIGNL	Glycinin G2 ^a (P04405), glycinin G3 ^a (P11828)
YNNEDTPVVA, IETWNPNNKPFQ, IETWNPNNKPF, YLAGNQEQUEF, ERVFDGEL, WNPNNKPFQ, NGERVFDGEL	Glycinin G1 ^a (P04776), glycinin G2 ^a (P04405), glycinin G3 ^a (P11828)
EDDEDEQIP, WGPLVNPESQQGSPR, LDTSNFFNNQ, PSYSPYPRM, VTRGQGKVRV, KYEGNWGPL, LAGNPDIIEY, FNTNEDIAEKL, DQTPRVFY, YLAGNPDIIEY, VVAEQAGEQGF, LDTSNFFNNQLDQTPRVFY	Glycinin G4 ^a (P02858)
YLAGNPDIIEHPETM, IIVVQKGGAIG, KYQGN SGPLVNP, LDQNPRVFY, IETWNSQHPEL, SHLPSYLPYPQM, KLRSPDDERKQ, YLAGNPDIIEHPET	Glycinin ^a (P04347)
SGFSKHF, IVTEGGLSVISPK	Glycinin ^a (P04347), glycinin G4 ^a (P02858)
VLHEAI, NKVDENGTPKPSL, LDIPGESHDVL, RNSWDPPNPHIG, IVTSSGKLQL, LAPIDTKPQTHAGY	Lectin ^a (P05046)
EWDGSMGIPGAF, TIPLPVIKE, NTSPLTLGAGE, INNHEKAYL, LSLPHSAGDL	Seed linoleate 13S-lipoxygenase-1 (P08170)
ITKEHLEPNL, LKNDGTLP, EDIPNHGSIHF, DLNFTPRE, LSLPHQGDQSGAF	Seed linoleate 9S-lipoxygenase-3 (P09186)
LSLPHAGDLSG, LANGKGKVGKDTF, EDVNPQGTIRF, TIMPLPVVKE	Seed linoleate 9S-lipoxygenase-2 (P09439)
IEDYPYAVDGL	Seed linoleate 9S-lipoxygenase-2 (P09439), seed linoleate 9S-lipoxygenase-3 (P09186)
WWYKVENHAAELTAGYY, LRYPSYPQSQG, EATKPTLPFPW, LEATKPTLPFPW, IVNIPIQW	β -Amylase (P10538)
LSVDMNEGALFLPH, VVDMNEGALFLPH	β -Conglycinin- α' chain ^a (P11827)
LRVPSGTTY, VVNPDDNNENLRL, IVDNEGALLPHFNSKA, ERQFPFRPPHQKE, FSRNILEASY, RLQSGDALRVPSGTTY, RLQSGDALRVPSGTTY, YVNPDDNNENLRL	β -Conglycinin- α chain ^a (P13916)
ILEASYDTKFEEINKVL, LSEQDIFVIPA, RAELSEQDIFVIPA	β -Conglycinin- α' chain ^a (P11827), β -conglycinin- α chain ^a (P13916)
VLEHGGIATDDDPY, EWLVEHGGIAT, WVLEHGGIAT, KKMKEQY	P34 probable thiol protease ^a (P22895)
YVGQKTKEVGQKTKE, YVGQKTKEVGQKTKEVGQD, IYHSERGPTTSQVL	P24 oleosin isoform A (P29530), P24 oleosin isoform B (P29531)
VHTTTHRYEAG	P24 oleosin isoform B (P29531)
FDQQNEGSIF, LVSESETEKITLEPGDMMIHIPAGTPLY, IHIPAGTPLY, IVNRDENDKL, KITLEPGDM, ALQTPKGKL, DLKPGMV, EKITLEPGDM, TRVETEGGRIRV	Sucrose binding protein (Q04672)
HHTNPTKPINL, HSFVPPNK, LSDSVPIQHH	Basic 7S globulin 2 (Q8RVH5)
VMDKPNGPVW, VMDKPNGPVW, VDLGNHNL, MDKPNGPVW, LGHAPISLPNQ, LGHAPISL, LGHAPISLPNQ	Basic 7S globulin 2 (Q8RVH5), basic 7S globulin (P13917)

^a Protein recognized as allergen according to Allergome platform.

“c” of the undigested sample (acid polypeptides), and some resistance of the basic polypeptide chains banding at 20 kDa (band “d”) in the undigested sample, enduring 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 was reported by Zhao *et al.* (2010)⁴² who also studied 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 to be more resistant, still being completely decomposed after 60 min. Likely, both the differences in the experimental protocols for digestibility assess-

ment 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 protein digestion, therefore complex mechanisms might occur to unravel the potentially allergenic proteins from their matrix. The more complex static model selected for *in vitro* simulation of the human gastroduodenal digestion utilized in this work and the complex salivary, gastric and duodenal fluid composition (electrolyte, surfactants, bile salts and typical enzymes) employed, 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 bands “d”, “f” and “l” of undigested, gastric and duodenal samples, respectively) to the GI proteolytic activity exerted by pepsin, trypsin and chymotrypsin enzymes.^{28,29,31}

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

Most of the soybean proteins discussed so far, for their different resistance to proteolysis, have been 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 (the 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. The IEDB database³⁸ was screened in order to match detected peptides and soybean linear epitopes recognized for the *Homo sapiens* host. The results are summarized in Tables 3 and 4. Combining results of both sets (high and low molecular weight portions of duodenal fluid) allowed us to obtain an overview of the potential allergenic 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

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 the IEDB with the relevant epitope ID

Band	Peptide sequence	Epitope ID	Epitopic sequence (allergen source)	Ref.
<i>g</i>	TLAIPVNKPGR	181524	YYVNPDDNNENLRLITLAIPVNKPGRFES (Gly m 5)	41
	LITLAIPVNKPGR			
	FFEITPEK	181292	DPIYSNKLKGFFEITPEKNPQLRDLD (Gly m 5)	41
<i>h</i>		181455	SEDKPFNLRSRDPIYSNKLKGFFEITPEKN (Gly m 5)	41
	ITLAIPVNKPGR	181524	YYVNPDDNNENLRLITLAIPVNKPGRFES (Gly m 5)	41
	TLAIPVNKPGR			
	LITLAIPVNKPGR			
	LITLAIPVNK			
	ILEFNSKPN	181386	LQNLRDYLILEFNSKPNNTLLPNHAD (Gly m 5)	41
	SSTEAAQSYLQGF	181316	FFLSSTEAAQSYLQGFNRNILE (Gly m 5)	41
	LSSTEAAQSYLQGFNR			
	NPFLFGSNR	181406	NKNPFLFGSNRFETLFKNQYGRIRVLQRF (Gly m 5)	41
	FFEITPEKNPQLR	181292	DPIYSNKLKGFFEITPEKNPQLRDLD (Gly m 5)	41
	EITPEKNPQLR	181455	SEDKPFNLRSRDPIYSNKLKGFFEITPEKN (Gly m 5)	41
<i>i</i>	FFEITPEK			
	FETLFK	181406	NKNPFLFGSNRFETLFKNQYGRIRVLQRF (Gly m 5)	41
	ILEFNSKPNNTLLPN	181386	LQNLRDYLILEFNSKPNNTLLPNHAD (Gly m 5)	41
	ILEFNSKPN			
<i>l</i>	SRDPIYSNK	181455	SEDKPFNLRSRDPIYSNKLKGFFEITPEKN (Gly m 5)	41
	LSAQYGLR	36948	LKLSAQYGLSLRKNNAM (Gly m 6)	46
	RFYLAGNQEDEF	53810	RFYLAGNQEDEFELKY (Gly m 6)	46
	NPFLFGSNR	181406	NKNPFLFGSNRFETLFKNQYGRIRVLQRF (Gly m 5)	41

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 the IEDB with the relevant epitope ID

Peptide sequence	Epitope ID	Epitopic sequence (allergen source)	Ref.
SGFSKHF	538677	GSVLSGFSKHF (Gly m6)	47
GRSQRPDQRHQQ	538745	QRPDQRHQQ (Gly m6)	47
YQEPQESQQRGRSQRPDQRHQQ			
FAPEFLKEAFG	58026	SGFAPEFLKEAFGVN (Gly m6)	48
VPHYTL	41548	MFVPHYTLNANSIIY (Gly m6)	46
	59438	SLRKNNAMFVPHYTLN (Gly m6)	
	70332	VPHYTLNANSIIYAL (Gly m6)	
LEFLEHAF	19632	GFTLEFLEHAFSV (Gly m6)	48
TLEFLEHAFS	58037	SGFTLEFLEHAFSVD (Gly m6)	49
VKNNNPFS	33488	KSQQARQVKNNNPFS (Gly m6)	46
YLAGNQEDEF	53810	RFYLAGNQEDEFELKY (Gly m6)	46
VVNPDDNNENLRL	181524	YYVNPDDNNENLRLITLAIPVNKPGRFES (Gly m5)	41
YYVNPDDNNENLRL			

β -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 weight 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 are very interesting from a toxicological perspective. We observed that several soy allergens appeared to be stable towards 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 partial confirmation of the correlation between allergenicity and digestibility. This study

should be further extended with immunological studies in order to have more insights into the actual residual immunoreactivity of the 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 10 kDa and in the duodenal fluid fraction below 10 kDa were subjected to a BLAST search in the UniprotKB/Swiss-Prot database³⁹ looking for any other legume crop sharing the same sequences. All proteins matching at least six consecutive amino acids belonging to the epitopic sequence were taken into consideration. The results, illustrated in Fig. 3, revealed that among the sixteen duodenal peptides recognized as soy epitopes (or part of them), six also matched with proteins coming from other plants belonging to the *Fabaceae* family, renowned as the major food allergen source. Specifically, glycinins shared peptides with *Pisum sativum* and *Cicer arietinum*, and β -conglycinin subunits showed common peptides with *Pisum sativum* as well as *Canavalia* (species *gladiata* and *ensiformis*), *Arachis hypogaea*, *Lupinus* (species *angustifolius* and *albus*), *Phaseolus* (species

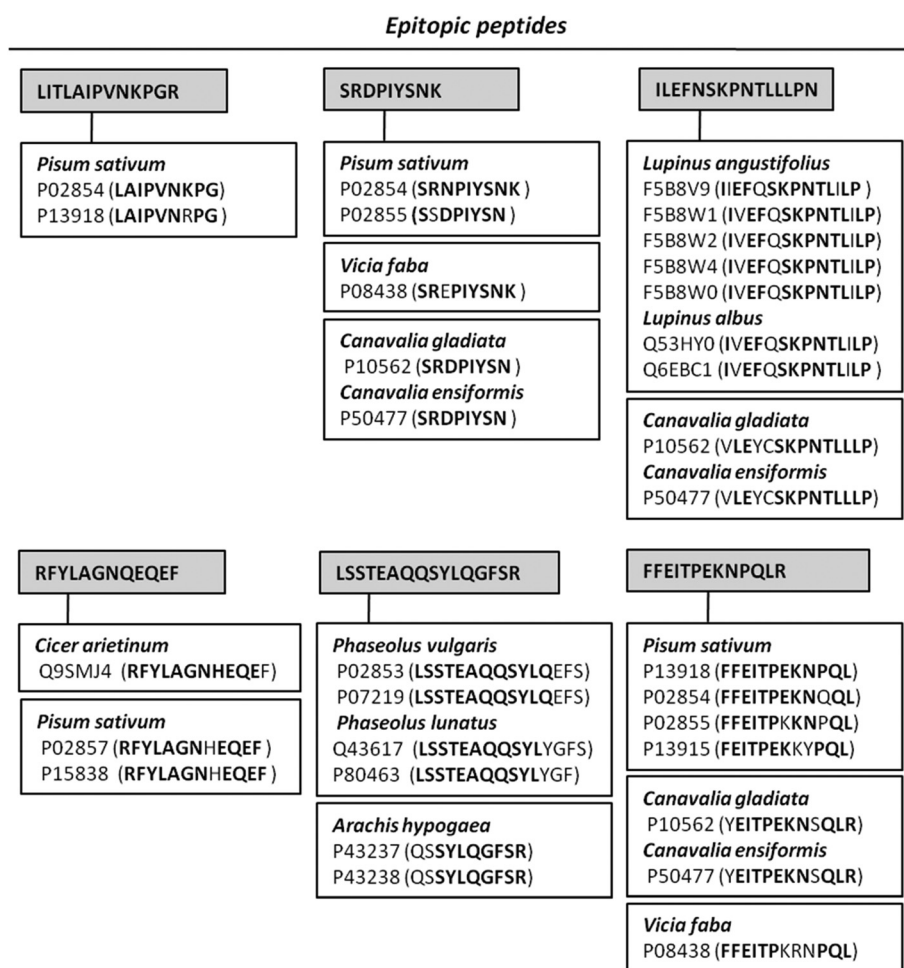


Fig. 3 Comparison between epitopic peptides identified in soybean duodenal digest (grey boxes) and proteins belonging to legume species showing similarity of at least six consecutive amino acids with immunogenic soybean peptides, obtained through the BLASTp search.

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 to or higher than 51% were found for most of the matches previously identified, except for soybean protein P11827 vs. *Lupinus angustifolius* F5B8 W1 and F5B8 W2, whose identities were 43 and 33%, respectively. Identity percentages lower than 50% were also displayed by matching soybean protein P13916 with the proteins F5B8 W (*Lupinus angustifolius*), Q43617 and P80463 (*Phaseolus lunatus*), and P43237 and P43238 (*Arachis 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 was 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 and pea, white bean, peanut, lentil, fennel, guar gum, carob bean, 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 to the *Fabaceae* family according to the scientific opinions issued by the EFSA in 2010,²² 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 subjecting 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 glycinin subunits, above all the acid polypeptides, were more susceptible than β -conglycinin to the GI enzymatic proteolysis. Large fragments of the β -conglycinin, survived 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 the bioinformatics 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 into the stability of soybean proteins upon gastro-duodenal digestion and 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 to the *Fabaceae* family, which could represent a risk for a wide category of allergic individuals.

Acknowledgements

The authors acknowledge the technical help of Roberto Schena in performing HPLC-MS/MS analyses. The authors are thankful to the project Safe&Smart – Nuove tecnologie abilitanti per la food safety e l'integrità delle filiere agroalimentari of the National CLAN Agrifood Cluster – for funding this research.

References

- 1 Z. Berk, *Fao Agricultural Services Bulletin No. 97*, Technion, Israel Institute of Technology, Haifa, Israel, 1992.
- 2 T. Wang, G.-X. Qin, Z.-W. Sun and Y. Zhao, *Crit. Rev. Food Sci.*, 2014, **54**, 850–862.
- 3 G. S. Gilani, C. W. Xiao and K. A. Cockell, *Br. J. Nutr.*, 2012, **108**, S315–S332.
- 4 A. W. Burks, J. R. Brooks and H. A. Sampson, *J. Allergy Clin. Immunol.*, 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, *Off. J. Eur. Union*, 2007, **L310**, 11–17.
- 9 WHO/IUIS Allergen Nomenclature Sub-Committee, <http://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 Immunol.*, 1994, **105**, 143–149.
- 12 T. Ogawa, M. Samoto and K. Takahashi, *J. Nutr. Sci. Vitaminol.*, 2000, **46**, 271–279.
- 13 A. Adachi, T. Horikawa, H. Shimizu, Y. Sarayama, T. Ogawa, S. Sjolander, A. Tanaka and 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. Agric. 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.
- 20 K. L. Bøgh and C. B. Madsen, *Crit. Rev. Food Sci.*, 2016, **56**, 1545–1567.
- 21 M. 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. Brodtkor, *Food Funct.*, 2014, **5**, 1113–1124.
- 22 EFSA 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 J.*, 2010, **8**, 1–168.
- 23 F. J. Moreno, A. R. Mackie and E. N. C. Mills, *J. Agric. Food Chem.*, 2005, **53**, 9810–9816.
- 24 F. J. Moreno, B. M. Maldonado, N. Wellner and E. N. C. Mills, *Biochim. Biophys. Acta*, 2005, **1752**, 142–153.
- 25 F. J. Moreno, F. A. Mellon, M. S. Wickham, A. R. Bottrill and E. N. C. Mills, *FEBS J.*, 2005, **272**, 341–352.
- 26 M. 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 and E. Lampart-Szczapa, *Eur. Food Res. Technol.*, 2016, **242**, 391–403.
- 28 J. D. Astwood, J. N. Leach and R. L. Fuchs, *Nat. Biotechnol.*, 1996, **14**, 1269–1273.
- 29 D. Sung, K. M. Ahn, S.-Y. Lim and S. Oh, *J. Sci. Food Agric.*, 2014, **94**, 2482–2487.
- 30 T. J. Fu, *Ann. N. Y. Acad. Sci.*, 2002, **964**, 99–110.
- 31 K. 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. Pharmacol.*, 2004, **39**, 87–98.
- 32 M. Amigo-Benavent, A. Clemente, P. Ferranti, S. Cairra and M. D. del Castillo, *Food Chem.*, 2011, **129**, 1598–1605.
- 33 M. Defernez, G. Mandalari and E. 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.
- 35 G. 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 and E. N. C. Mills, *Regul. Toxicol. Pharmacol.*, 2009, **55**, 372–381.
- 36 G. Mandalari, A. M. Mackie, N. M. Rigby, M. S. Wickham and E. N. Mills, *Mol. Nutr. Food Res.*, 2009, **53**, S131–S139.
- 37 L. Monaci, R. Pilolli, E. De Angelis, M. Godula and A. Visconti, *J. Chromatogr. A*, 2014, **1358**, 136–144.
- 38 Immuno Epitope Database and Analysis, <http://www.iedb.org>. (accessed July 2016).
- 39 Expasy, 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.
- 41 X. 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 Agric. 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, vol. 45, pp. 57–60, appendix IV.
- 45 European Federation of Asthma and Allergy Associations, *Food Allergy Brochure EFA, Box 5*, Leusden, Netherland, 1997, p. 3830.
- 46 R. M. Helm, G. Cockrell, C. Connaughton, H. A. Sampson, G. A. Bannon, V. Beilinson, N. C. Nielsen and A. W. Burks, *Int. Arch. Allergy Immunol.*, 2000, **123**, 213–219.
- 47 H. Saeed, C. Gagnon, E. Cober and S. Gleddie, *Mol. Immunol.*, 2016, **70**, 125–133.
- 48 P. Xiang, T. A. Beardslee, M. G. Zeece, J. Markwell and G. Sarath, *Arch. Biochem. Biophys.*, 2002, **408**, 51–57.
- 49 T. A. Beardslee, M. G. Zeece, G. Sarath and J. P. Markwell, *Int. Arch. Allergy Immunol.*, 2000, **123**, 299–307.

Table S1. Full list of the identified peptides in the low molecular weight protein fraction (<10kDa) of the soybean duodenal digest (D=180 min). The accession number and the allergen codification (if present) were reported for each protein..

Peptide sequence	Protein	XCorr	m/z [Da]	ΔM [ppm]
VVEDLPEGPAVKIGE	Trypsin inhibitor A*(P01070); Trypsin inhibitor B*(P01071)	3.67	776.4	-51
SVVEDLPEGPAVKIGE		5.05	819.8	-165
ISIDHDDGTRRL		2.07	699.3	-103
VLDNEGNPLENGGTY	Trypsin inhibitor A*(P01070)	4.37	796.4	91
NKDAMDGW		2.54	468.7	-22
DNEGNPLENGGTY		3.28	690.4	106
EDDEDEQIP	Glycinin G4* (P02858)	2.31	545.2	-40
WGPLVNPESQQGSPR		3.74	826.3	-139
LLDTSNFNNQ		3.22	583.4	275
PSYSPYPRM		3.24	549.4	159
VTRGQGKVRV		2.87	550.3	-112
KYEGNWGPL		2.73	532.1	-292
LAGNPDI EY		2.84	496.4	287
FNTNEDIAEKL		3.28	647.2	-184
DQTPRVFY		2.57	513.2	-128
YLAGNPDI EY		2.79	577.8	50
VVAEQAGEQGFE		4.53	632.5	298
LDTSNFNNQLDQTPRVFY		4.24	1086.8	273
SGFSKHF	Glycinin G4* (P02858); Glycinin* (P04347)	2.07	405.2	-45
IVTVEGGLSVISPK		4.53	699.8	-180
YLAGNPDI EHPETM	Glycinin* (P04347)	3.93	793.7	-188
IIVVQGKGAIG		2.46	527.7	-271
KYQGN SGPLVNP		3.09	637.3	-74
LDQNPRVFY		2.85	576.2	-174
IETWNSQHPEL		3.37	677.5	186
SHLPSYLPYPQM		2.54	716.7	-201
KLRSPDDERKQ		3.14	686.2	-243
YLAGNPDI EHPET		2.95	728.4	37
IGQNSSPDIYNPQAGSITT	Glycinin G2* (P04405)	4.89	981.8	-167
GRSQR PQDRHQK		2.52	498.3	13
YQEPQESQQRGRSQR PQDRHQK		2.61	922.6	-254
IGQNSSPDIYNPQAGSI		4.17	881.1	136
FAPEFLKEAFGVNMQ		3.79	864.3	-146
FAPEFLKEAFG		3.09	628.4	151
SSPDIYNPQAGSIT		3.23	725.5	152
VPHYTL		2.19	365.1	-286
IGQNSSPDIYNPQAGSITTA		3.63	1017.3	-177
LQGENEEEDSGAIVTVK		3.87	909.3	-152
LLNALPEEVIQHTF	Glycinin G2* (P04405); Glycinin G1* (P04776)	5.93	541.9	-37
LLNALPEEVIQHTFN		4.50	869.4	-89
ALPEEVIQHTF		2.70	642.3	-80
LLNALPEEVIQH		2.85	688.5	223
IIDTNSLENQLDQMPRR		3.32	1022.1	74
LNALPEEVIQHTF		3.98	756.0	138
SIIDTNSLENQLDQMPRRF		2.82	759.5	-226

IIDTNSLENQLDQMPRRF		2.03	1095.3	-214
NALPEEVIQHTF		3.12	699.5	260
FREGDLIAVPTGVAW		3.78	816.0	41
LLNALPEEVIQHT		3.47	738.8	-156
VSFKTNDTPMIGT	Glycinin G1* (P04776)	3.25	706.0	135
KTNDTPMIGTL		3.15	595.9	214
KYQQEQGGHQSQKGKHQQEEENEGGSIL		4.92	1061.3	139
KNNNPFKF		2.11	504.8	13
LQGENEGEDKGAIVT		2.62	780.4	75
LKYQQEQGGHQSQKGKHQQE		3.87	789.4	-17
LKPDNRIESEGLL		3.13	714.5	135
SVIKPPTDEQQRPQE		2.80	940.4	-95
IGQTSSPDIYNPQAGSVTT		4.47	968.2	-248
SSPDIYNPQAGSV		3.13	667.7	-168
IKNNNPFKF		3.51	561.4	141
IGQTSSPDIYNPQAGSV		4.82	867.3	-138
IKNNNPF		2.43	423.7	-102
LEFLEHAF		2.45	503.2	-131
TLEFLEHAFS		2.45	597.4	114
IYNPQAGSVTTA		2.39	611.2	-182
VLHEAI	Lectin* (P05046)	2.00	341.2	-67
NKVDENGTPKPSSL		2.87	743.3	-120
LDIPGESHDVL		3.04	597.7	-179
RNSWDPPNPHIG		2.19	695.5	230
IVTSSGKLQL		3.24	523.2	-239
LAPIDTKPQTHAGY		2.77	756.3	-137
EDWGSMGIPGAF		2.74	633.8	79
TIPLPVIKE	Seed Linoleate 13S-lipoxygenase-1(P08170)	2.82	562.0	293
NTSLPTLGAGE		2.49	530.2	-147
INNHEKAYL		2.55	551.2	-172
LSLPHSAGDL		3.10	505.3	8
ITKEHLEPNL		2.69	597.3	-83
LKNDGTLRP	Seed Linoleate 9S-lipoxygenase-3(P09186)	2.27	507.2	-198
EDIPNHGSIHF		3.31	633.3	-25
DLNFTPRE		2.71	496.4	214
LSLPHPQGDQSGAF		2.62	727.2	-207
IEDYPYAVDGL	Seed Linoleate 9S-lipoxygenase-3(P09186); Seed Linoleate 9S-lipoxygenase-2(P09439)	2.64	627.8	-25
LSLPHPAGDLSG	Seed Linoleate 9S-lipoxygenase-2(P09439)	3.41	582.3	-47
LANGKGKVGKDTF		3.76	668.0	199
EDVPNQGTIRF		3.28	638.2	-188
TIMPLPVVKE		2.30	563.8	-94
WWYKVENHAAELTAGYY	β -amylase (P10538)	2.33	1050.8	-171
LRYPSPQSQG		2.73	648.2	-188
EATKPTLPFPW		3.17	644.0	189
LEATKPTLPFPW		3.47	700.4	-19
IVNIPIQW		2.81	540.3	-74
LSVVDMNEGALFLPH	β -conglycinin- α' chain* (P11827)	3.75	821.3	-148
VVDMNEGALFLPH		3.50	721.5	198
ILEASYDTKFEEINKVL	β -conglycinin- α' chain* (P11827); β -conglycinin- α chain* (P13916)	4.05	671.4	17
LSEQDIFVIPA		2.74	616.5	294
RAELSEQDIFVIPA		2.13	794.3	-160
FAPEFLEHAF	Glycinin G3* (P11828)	3.05	604.4	251

LNALPEEVIQQTFF		4.34	751.3	-142
IGQTSSPDIFNPQAGSIT		3.31	916.9	-73
SSPDIFNPQAGSITT		2.86	767.9	5
IDTNSFQNQL		3.20	590.5	299
FEPPQQKGQSSRPQDRHQB		3.20	770.9	209
LLNALPEEVIQQTFF		4.02	807.8	-179
NALKPDNRIESEGGF	Glycinin G3* (P11828); Glycinin G2* (P04405)	2.69	549.5	-150
LKPDNRIESEGGF		2.83	731.5	207
VKNNNPFSF		3.40	533.9	254
RIESEGGF		2.49	447.8	91
VKNNNPFS		2.19	460.2	-106
KTNDRPSIGNL		3.82	607.8	-77
YNNEDTPVVA	Glycinin G3* (P11828); Glycinin G2* (P04405); Glycinin G1* (P04776)	2.08	561.4	252
IETWNPNNKPFQ		3.17	744.5	179
IETWNPNNKPF		3.66	680.5	226
YLAGNQEQEF		3.35	599.7	-126
ERVFDGEL		2.47	482.6	-281
WNPNNKPFQ		2.39	572.8	-3
NGERVFDGEL		3.09	568.4	226
GQSSRPQDRHQB	Glycinin G3* (P11828); Glycinin G1* (P04776)	2.45	712.3	-101
LRVPSGTTY	β-conglycinin-α chain (P13916)	2.77	497.4	161
VVNPDDNNENLRL		3.27	698.7	-226
IVDMNEGALLLPHFNSKA		5.04	657.2	289
ERQFPFPRPPHQKE		2.91	896.8	-181
FSRNILEASY		2.29	600.2	-182
RLQSGDALRVPSGTTY		3.30	861.1	183
RLQSGDALRVPSGTTY		3.27	942.3	-184
YVVNPDDNNENLRL		3.67	780.3	-128
VLEHGGIATDDDY	P34 probable thiol protease* (P22895)	2.30	832.8	-93
EWVLEHGGIAT		2.09	606.2	-180
WVLEHGGIAT		3.43	541.9	253
KKMKKEQY		2.47	541.8	-53
YVGQKTKEVGQKTKE	P24 oleosin isoform A (P29530); P24 oleosin isoform B (P29531)	3.09	574.9	-213
YVGQKTKEVGQKTKEVGQD		4.75	708.2	230
IYHSERGPTTSQVL		3.64	794.3	-144
VHTTTHRYEAG	P24 oleosin isoform B (P29531)	2.97	424.5	-60
FDQQNEGSIF	Sucrosebindingprotein (Q04672)	2.36	592.7	-115
LVSESETEKITLEPGDMMHIPAGTPLY		4.61	981.0	144
IHIPAGTPLY		2.87	541.2	-209
IVNRDENDKL		2.71	608.2	-201
KITLEPGDM		3.18	502.4	277
ALQTPKGKL		2.49	478.2	-139
DLKPGMV		2.07	380.2	72
EKITLEPGDM		2.57	566.8	-6
TRVETEGGRIRV		2.38	686.8	-145
HHTNPTKPINL	Basic 7S globulin 2 (Q8RVH5)	2.87	636.3	-99
HSVFPPNK		2.21	463.2	-133
LSDSVPIQHH		2.10	615.2	-193
VMDKPNGPVW	Basic 7S globulin 2 (Q8RVH5); Basic 7S globulin (P13917)	2.83	579.7	-153
VMDKPNGPVW		3.94	571.9	105
VDLNGNHL		2.42	441.3	131
MDKPNGPVW		3.31	522.2	-118
LGHAPISLPNQ		2.95	573.7	-212

LGHAPISL		2.22	404.3	115
LGHAPISLPNQL		3.06	630.5	139