JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Functional Region Identification in Proteins by Accumulative–Quantitative Peptide Mapping Using RP-HPLC-MS

BAS J. H. KUIPERS,† E. J. BAKX, AND HARRY GRUPPEN*

Department of Agrotechnology and Food Sciences, Laboratory of Food Chemistry, Wageningen University, P.O. Box 8129, 6700 EV Wageningen, The Netherlands

A new method was developed to identify regions in proteins from which peptides are derived with specific functional properties. This method is applicable for systems in which peptides of a hydrolyzed protein possess specific functional properties, but are too large to be sequenced directly and/or the peptide mixture is too complex to purify and characterize each peptide individually. In the present work, aggregating peptides obtained by proteolytic hydrolysis of soy glycinin were used as a case study. The aggregating peptides are isolated and subsequently further degraded with trypsin to result in peptides with a mass <5000 Da to enable sequence identification using RP-HPLC-MS in combination with MS/MS. Prior to RP-HPLC the peptides are fractionated using anion and cation exchange chromatography. The fractions obtained are analyzed with RP-HPLC-MS. The peptides, with identified sequences, were quantified using the peak areas of the RP-HPLC chromatograms measured at 214 nm. Next, the peak areas were corrected for the molar extinction coefficient of the individual peptides, followed by accumulative–quantitative peptide mapping. The results show that in complex systems, based on the method described, the regions in the parental protein from which the functional peptides originate can be properly identified.

KEYWORDS: Proteins; peptides; RP-HPLC; ion exchange; UV absorbance; mass spectrometry; MS; quantitative analysis; peptide mapping

INTRODUCTION

In the literature several studies can be found on the ability of proteases to yield peptides that have strong aggregating properties (*1*–*3*). Upon limited hydrolysis of soy glycinin we have shown previously (*4*, *5*) that various proteases (e.g., subtilisin Carlsberg and chymotrypsin), having a broad specificity, are able to form a wide range of peptides with strong aggregating properties. These aggregating peptides have molecular masses ranging from ∼5000 to 25000 Da as determined by size exclusion chromatography (*5*). Due to the broad specificity of the enzyme used it is hypothesized that many different peptides with equal masses can be present, which largely overlap in sequence.

Knowing the sequences of the aggregating peptides in a protein hydrolysate is of prime interest to understand the aggregation mechanism. Understanding the mechanisms of aggregation will help to find solutions to decrease aggregate formation in situations in which aggregation is unwanted (e.g., in drinks) or to stimulate the aggregation in products where aggregation is preferred (e.g., tofu). The easiest method to analyze the sequences of peptides in a protein hydrolysate is by direct analysis on RP-HPLC-MS. However, identification of peptide sequences using ESI Iontrap MS in combination with tandem mass spectrometry (MS/MS) (*6*) requires peptide masses below ∼4000 Da for positive sequence identification. If the molecular masses of the peptides of interest are higher, as is the case for the aggregating peptides from glycinin, a different method has to be used. The most obvious method is first to isolate all aggregating peptides individually, followed by quantification of the individual peptide concentration. Methods for separation might be RP-HPLC or 2D electrophoresis. For determination of the peptide sequences, each individual peptide fraction can be studied by N-terminal sequencing followed by determination of the total mass by using, for example, MALDI-TOF (*7*–*10*). However, if the peptides of interest comprise a wide range of peptides that overlap in sequence, accurate separation of the individual peptides is not feasible.

Therefore, if functional peptides have masses above ∼5000 Da and largely overlap in sequence, a new method has to be used. Such a method is described in this paper. This method, denoted *accumulative–quantitative peptide mapping*, does not have the aim to identify the sequence of the peptides of interest, but aims to identify those regions in the parental protein from which the functional peptides originate. This method is schematically presented in **Figure 1**.

* Author to whom correspondence should be addressed (telephone +31 317 483211; fax +31 317 484893; e-mail harry.gruppen@wur.nl).
† Present address: DMV International, Wageningen, The Netherlands.

10.1021/jf071380b CCC: $37.00
Published on Web 10/19/2007

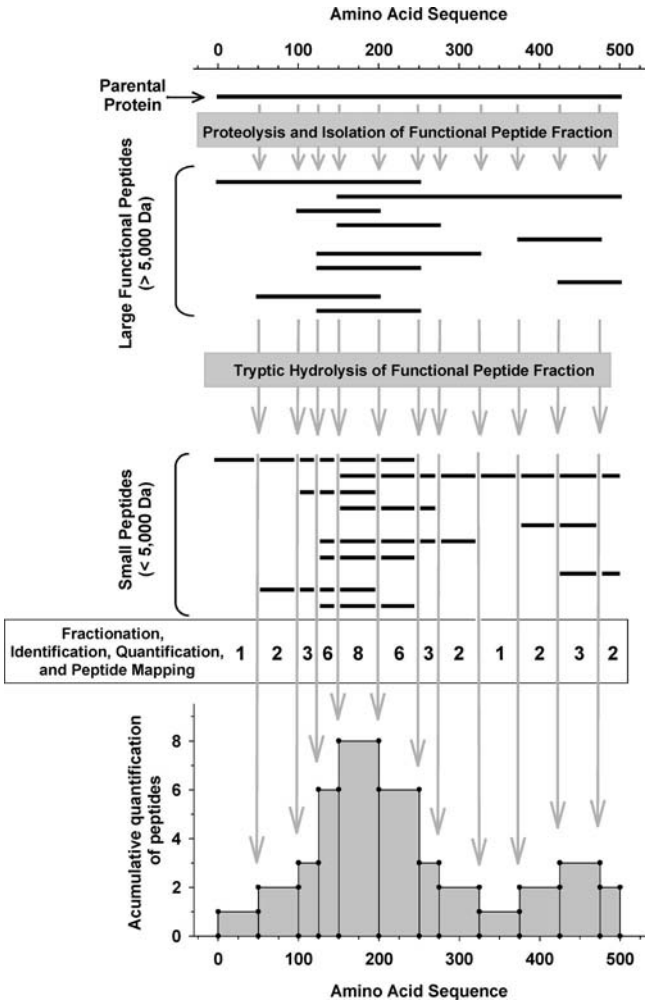

Figure 1. Identification of the functional regions of proteins by accumulative–quantitative peptide mapping. The parental protein is selectively degraded into "large functional peptides" that are isolated in a functional peptide fraction based on a specific functional property, followed by a second complete degradation of the peptides in the functional peptide fraction by trypsin. Each peptide is identified and quantified. Next, all peptides are mapped onto the amino acid sequence of the parental protein, taking into account their abundance. Those regions in the protein sequence with the highest accumulation per amino acid reveal the functional regions in the parental protein from which the peptides of interest originate. The gray arrows (↓) indicate the position at which the enzyme can cleave. To simplify the scheme, the specificity of the first proteolytic digestion is taken to be the same as that of trypsin.

The functional peptides (e.g., peptides with strong aggregating properties) are first isolated from peptides not contributing to the functionality of interest. To decrease the peptide mass, enabling positive sequence identification with RP-HPLC-MS, subsequently tryptic digestion is performed. Because of the expected overlap in sequence, tryptic digestion of the overlapping sequences from different functional peptides mainly results in the accumulation of identical peptides from the overlapping regions. The tryptic digest is fractionated (e.g., RP-HPLC) followed by identification (MS/MS), quantification [absorbance at 214 nm (*11*)], and peptide mapping projected toward the sequence of the parental protein. The amino acids in the sequence of the parental protein, for which the strongest accumulation takes place, reveal those regions in the parental protein from which the functional peptides mainly originate. These regions can be defined as the major functional peptide region of the protein (in **Figure 1**, around the amino acids 150–200).

The aim of the present study was to develop a quantitative method to identify those regions in a protein from which the peptides predominantly originate that have specific functional properties. In this study aggregating peptides from soy glycinin are used as an example. These aggregating peptides were obtained by hydrolysis with chymotrypsin. It is envisaged that this method can also be applied to other functional peptide mixtures, containing peptides being too large to be analyzed directly by RP-HPLC-MS and too complex to be separated into its individual peptides.

MATERIALS AND METHODS

Materials and Chemicals. Glycinin was prepared from defatted SBM (from Hyland soybeans) by precipitation of the pH 8.0 soluble proteins at pH 6.2 using 1 M HCl, as described previously (*5*). Glycinin has a protein content of 87% (w/w) (N × 5.57) as determined by the Dumas method and a purity >95% as determined by densitometric analysis of the SDS-PAGE gel (no data shown). Chymotrypsin (TLCK treated), trypsin (TPCK treated), and bradykinin were obtained from Sigma Chemical Co. (St. Louis, MO; articles C-3142, T-1426, and B-3259, respectively). The eluents used for RP-HPLC-MS were all of HPLC grade. All other chemicals were of analytical grade and were purchased from Merck (Darmstadt, Germany) or Sigma.

Analysis of the Protein Content. The nitrogen contents of the various samples were determined in duplicate by the combustion method (*12*) using an NA2100 Nitrogen and Protein Analyzer (CE Instruments, Milano, Italy) according to the manufacturer's instructions. Methionine was used as a standard. The nitrogen conversion factor for purified glycinin is 5.57, based on the average amino acid compositions of the five different subunits of glycinin as found in the Swiss-Prot database [www.expasy.org; primary accession numbers used were P04776 (glycinin 1; A1a and B2 polypeptides), P04405 (glycinin 2; A2 and B1a polypeptides), P11828 (glycinin 3; A1b and B1b polypeptides), P02858 (glycinin 4: A5, A4, and B3 polypeptides), and P04347 (glycinin 5; A3 and B4 polypeptides)]. Equal molar abundance of the subunits was assumed.

Hydrolysis of Glycinin. Glycinin was suspended in deionized water at a concentration of 1.5–2% (w/w) protein and stirred overnight at 4 °C. The pH was adjusted to 8.0, if necessary. The glycinin suspension was filtered using a 0.45 μm filter (Schleicher & Schuell GmbH, Dassel, Germany). The protein concentration of the filtrate was determined using the Dumas method. The solution was diluted to 1% (w/w) protein using deionized water followed by heating at 95 °C for 30 min. The heated solution had a clear appearance. The heated glycinin solution (150 mL) was hydrolyzed by chymotrypsin up to a degree of hydrolysis (DH) of 2.2% at pH 8.0 at 40 °C. The pH and DH were controlled using the pH-stat method by using a 719S Titrino (Metrohm ion analysis, Herisau, Switzerland) (*13*). The h_{tot} (total number of peptide bonds present in the intact protein) used for the calculation was 8.85 mequiv/g, calculated on the basis of the average amino acid composition of the glycinin subunits. The enzyme/protein ratio was 1:350 (w/w). The chymotrypsin was dissolved in deionized water and directly added to the glycinin solutions. The molarity of the NaOH solution used to maintain the pH at 8.0 was 0.1 M. When the desired DH was reached, the enzymatic hydrolysis was stopped by the addition of a 100 mM phenylmethanesulfonyl fluoride (PMSF) solution in 2-propanol to a final concentration of 1 mM. The pH-stat experiment was stopped when the pH remained stable at pH 8.0. After cooling to room temperature, deionized water and NaCl (from a 2 M stock solution) were added to end up with a protein concentration of 0.8% (w/w) and an ionic strength of $I = 0.03$ M, as described previously (*5*). The pH was adjusted to 8.0 if necessary. The addition of NaCl induces aggregation. After 1 h of stirring, a sample was taken, whereas the rest was centrifuged (5 min, 5000*g*, 20 °C). Analysis of the protein content of the sample before centrifugation and the supernatant after centrifugation showed that ~43% of the proteinacious material was aggregated. The pellet was

washed by resuspending it in deionized water containing 0.03 M NaCl (~1:7 ratio of pellet/water) and centrifuged again (5 min, 5000*g*, 20 °C). This washing step was repeated twice. During the washing the pH of the suspension remained 8.0. After washing, the pH of the pellet was decreased to ~2.5 using trifluoroacetic acid to resolubilize the peptides, followed by freeze-drying. These freeze–dried peptides are further denoted aggregating peptides.

Tryptic Digestion of Aggregating Peptides. Aggregating peptides (20 mg) were dissolved in 600 μL of 20 mM Tris-HCl buffer (pH 8.0) containing 8 M urea and 15 mM DTT. The solution was kept for 30 min at 60 °C. Iodoacetamide (IAA) was added up to 30 mM, and incubation took place in the dark for 30 min to alkylate the peptides. Every 10 min the pH was verified and adjusted to 8.0 if necessary using 1 M NaOH. The samples were diluted to 1.6 M urea by the addition of 20 mM Tris-HCl buffer (pH 8.0). This resulted in partial aggregation of the peptides as observed by an increase in turbidity. The alkylated aggregating peptides were degraded by trypsin in three cycles. First, trypsin was added in an enzyme protein ratio (w/w) 1:50 followed by incubation at 37 °C for 4 h. The hydrolyzed sample was centrifuged (15 min, 24000*g*, 20 °C). The pellet was resuspended in 20 mM Tris-HCl buffer (pH 8.0) containing 8 M urea. The solution obtained was diluted to 1.6 M urea by the addition of 20 mM Tris-HCl buffer (pH 8.0). Trypsin (0.1 mg) was added to the resuspended pellet and incubated overnight at 37 °C in the presence of 0.02% (w/v) sodium azide. After incubation, the sample was treated the same way as described above, by resuspension in 8 M urea, dilution to 1.6 M urea, and incubation for 2 h with 0.1 mg of trypsin at 37 °C in the presence of 0.02% (w/v) sodium azide. The hydrolysate was mixed with the supernatants of the first and second tryptic hydrolysis cycles, followed by centrifugation (15 min, 24000*g*, 20 °C) prior to fractionation using ion exchange chromatography.

In addition to the aggregating peptide fraction, also the total hydrolysate after chymotryptic hydrolysis, containing the aggregating as well as the nonaggregating peptides, was hydrolyzed in three cycles with trypsin as described above.

Ion Exchange Chromatography (IEX). An aliquot (1.4 mL) of the supernatant of the tryptic digest of the aggregating peptides was diluted with 0.8 mL of deionized water and filtered through a 0.45 μm filter (FP 30/0.45 CA-S; Schleicher & Schuell, Dassel, Germany). Next, 2 mL of the filtered peptide solution was applied onto a 1 mL Mono Q 5/50GL anion exchange column (Amersham Biosciences, Uppsala, Sweden). Solvent A (20 mM Tris-HCl buffer, pH 8.0) and solvent B (20 mM Tris-HCl buffer, pH 8.0, containing 2 M NaCl) formed the eluent in the following consecutive steps: 3 mL isocratic elution at 100% A; sample injection followed by 10 mL isocratic elution at 100% A; linearly to 50% B in 60 mL; linearly to 100% B in 15 mL; 5 mL isocratic elution at 100% B; linearly to 100% A in 3 mL; 12 mL isocratic at 100% A. The flow rate was 2 mL/min except during the injection stage, at which the flow rate was 1 mL/min. Fractions of 1 mL were collected. The eluent was monitored at 214 nm. The nonbound fractions (fractions 1–12) were pooled and applied on a cation exchange column as described below.

The pH of the nonbound fractions on the anion exchanger was decreased to pH 3.5 using formic acid. Deionized water was added up to 23 mL and filtered through a 0.45 μm filter (Schleicher & Schuell). Only 20 mL (4 × 5 mL) of the 23 mL unbound peptide fraction was applied on a 1 mL cation exchanger, Mono S 5/50GL (Amersham Biosciences). Solvent A (20 mM formic acid) and solvent B (20 mM formic acid containing 2 M NaCl) formed the eluent in the following subsequent steps: 3 mL isocratic elution at 100% A; 4 × 5 mL sample injection during a 38 mL isocratic elution at 100%; linearly to 30% B in 15 mL; linearly to 100% B in 17 mL; 10 mL isocratic elution at 100% B; linearly to 100% A in 3 mL; 12 mL isocratic at 100% A. The flow rate was 2 mL/min except during the injection stage, when the flow rate was 1 mL/min. Fractions of 1 mL were collected. The eluate was monitored at 214 nm.

RP-HPLC-MS. Of all bound fractions collected from anion (25 fractions) and cation (27 fractions) exchange chromatography, 300 μL was taken. Next, to all fractions was added 16.5 μL of acetonitrile, followed by the addition of 14 μL of 5% (v/v) and 3.2% (v/v) formic acid for the anion and cation exchange chromatography fractions, respectively. All unbound fractions (38 mL) from the cation exchange column were pooled, and 6 mL was freeze-dried and subsequently dissolved in 300 μL of deionized water followed by the addition of 16.5 μL of acetonitrile and 14 μL of 3.2% (v/v) formic acid. Next to the bound and unbound samples, a sample of the tryptic digest of the total chymotryptic hydrolysate was diluted 8 times with a solution of 5% (v/v) acetonitrile in water containing 0.1% (v/v) formic acid.

After 1 h of mixing (head over tail), all samples were centrifuged (15 min, 24000*g*, 20 °C), and 100 μL was injected onto a reversed phase C18 column (218MS52; 250 × 2.1 mm, 5μm) (Grace Vydac, Hesperia, CA), installed on a Spectra System HPLC (Thermo Separation Products, Fremont, CA). A flow rate of 0.2 mL/min was used. The solvents used were deionized water containing 0.1% (v/v) formic acid (solvent A) and acetonitrile containing 0.085% (v/v) formic acid (solvent B). The elution used was as follows: from 0 to 10 min, 5% B (isocratic); linear gradient to 45% B until 80 min; linear gradient to 95% B until 90 min; isocratic elution with 95% B for 5 min; linear gradient to 5% B until 96 min followed by isocratic elution with 5% B until 110 min. All HPLC operations and data processing were controlled by Chromeleon version 6.7 software (Dionex Corp., Sunnyvale, CA).

After UV detection at 214 nm, a flow splitter (type Acurate, LC Packings, Amsterdam, The Netherlands) decreased the flow to 0.05 mL/min that was applied on an LCQ Deca XP Max (Thermo Finnigan, San Jose, CA) with the use of electrospray ionization and detection in the positive ion mode. The capillary spray voltage was 4 kV, and the capillary temperature was 200 °C. Before use, the instrument was tuned with a 0.01 mg/mL bradykinin solution. The instrument was controlled by Xcalibur software v1.3 (Thermo Finnigan). The scan range was set from *m*/*z* 400 to 2000. The MS/MS functions were performed in data-dependent mode. The collision energy value was 35%. BioWorks software, version 3.1 (Thermo Electron) was used for automatic sequencing and database search for the sequences in a database containing only sequences of glycinins G1–G5 and trypsin and chymotrypsin, as present in the Swiss-Prot database. In the database search the possible modifications of the proteins, oxidation of methionine (+15.99 Da) and alkylation of cysteine (+57.05 Da), were included. To discriminate between correct and incorrect peptide sequence assignment, the cross-correlation value (Xcorr) for each identified peptide was used as a criterion (*14*, *15*). For positive identification of the peptides a Xcorr threshold of 1.5 was used for single-charged peptides, 2.0 was used for double-charged peptides, and 2.5 was used for triple charged peptides.

Quantification of Peptides Based on the Peak Area at 214 nm. Because the absorbance was measured at 214 nm in the presence of acetonitrile and TFA, the molar extinction coefficient for each of the identified peptides was calculated using formula 1 as described previously (*11*):

$$\epsilon_{\text{peptide}}(\text{M}^{-1}\,\text{cm}^{-1}) = \epsilon_{\text{peptide bond}} \times n_{\text{peptide bond}} + \sum_{i=1}^{20} \epsilon_{\text{amino acid(i)}} \times n_{\text{amino acid(i)}} \quad (1)$$

The molar extinction coefficients used for the peptide building blocks (individual amino acid contribution and the peptide bond) are presented in **Table 1**. Because the peak area is used as a measure for the absorbance and the light path is not defined, but similar in all analyses, the concentration is not expressed as molarity, but as molar equivalents (Mequiv).

To enable comparison of all RP-HPLC chromatograms in a quantitative way, the area of the peaks from the cation IEX were proportionally corrected for the amount of unbound fraction that was not applied onto the cation exchange column. In addition to that, the area of the peaks in the unbound fraction of cation IEX was multiplied by 1.9 to correct for the dilution step of this fraction. Analysis of the 53 RP-HPLC revealed that, on the basis of the peak areas, <3% of all the peptides were present in the unbound fraction after cation IEX.

High-Performance Size Exclusion Chromatography (HP-SEC). Glycinin, its DH 2.2 hydrolysate, the aggregating peptide fraction thereof, and the tryptic hydrolysate of the aggregating peptides were analyzed with size exclusion chromatography. Samples were dissolved

Table 1. Molar Extinction Coefficients (ε) of Protein and Peptide Building Blocks for Calculation of the Molar Extinction Coefficient in the Peptide Chain at 214 nm

building block	ε (M^{-1} cm^{-1})	building block	ε (M^{-1} cm^{-1})
peptide bond	923	leucine	45
alanine	32	lysine	41
arginine	102	methionine	980
asparagine	136	phenylalanine	5200
aspartic acid	58	proline (at N terminus)	30
cysteine	225	proline (not at N terminus)	2675
glutamine	142	serine	34
glutamic acid	78	threonine	41
glycine	21	tryptophan	29050
histidine	5125	tyrosine	5375
isoleucine	45	valine	43

under denaturing (8 M guanidinium chloride) and reducing conditions (100 mM DTT). Of each sample 50 μL was injected and separated with a Shodex Protein KW-803 column (300 × 8 mm; Showa Denko K.K., Tokyo, Japan) equilibrated and eluted with 6 M urea containing 30% (v/v) acetonitrile and 0.1% (v/v) TFA. The flow rate was 0.5 mL/min, and the eluate was monitored at 214 nm. The final protein concentration was ~3 mg/mL. To facilitate comparison of the different samples, the total area under the chromatograms was corrected for the proportion of peptides that was aggregated. The column was calibrated using proteins with known molecular masses as described previously (*5*).

RESULTS AND DISCUSSION

Monitoring the Hydrolysis. To monitor the protein degradation of the sequential hydrolysis of glycinin with chymotrypsin, followed by the hydrolysis of the aggregating peptide fraction with trypsin, HP-SEC was performed. It can be observed (**Figure 2**) that at a DH of 2.2% there is probably no more intact protein present. **Table 2** shows the molecular mass range of the proteins and peptides present in the various analyzed fractions. The chromatogram of the intact glycinin polypeptides shows some aggregates and/or β-conglycinin eluting before 7 mL (>40000 Da). The polypeptides eluting around 7.4 mL (25000–40000 Da) represent the acidic polypeptides A1–A4, polypeptides eluting around 8.2 mL (15000–25000 Da) represent all of the basic polypeptides, and the peak eluting around 9 mL (8000–15000 Da) represents the A5 acidic polypeptide of glycinin (*16*). The aggregating peptide fraction mainly contains peptides with masses in the range of ~5000–25000 Da (**Table 2**). These are reduced to peptides with masses <5000 Da upon tryptic hydrolysis. Prior to tryptic digestion, the aggregating peptides were (partly) present as aggregates in the hydrolysis buffer. In **Figure 2** it can be observed that the aggregating peptide fraction was degraded into smaller fragments, indicating that the aggregation did not prevent tryptic degradation of the peptides. After the repetitive digestion with trypsin, there were still some aggregates present in the hydrolysate. Nevertheless, HP-SEC of the total tryptic hydrolysate and its supernatant revealed that >90% of aggregating peptides was soluble after tryptic degradation (results not shown).

Peptide Fractionation. Figure 3 shows the ion exchange chromatograms of the tryptic digest of the aggregating peptides. Anion exchange chromatography yielded 25 fractions, which eluted after applying the NaCl gradient (%B). The unbound fraction of the anion exchange fractionation (**A**) was applied on the cation exchange column (**B**).

The 52 bound fractions and the unbound fraction of cation exchange chromatography and the tryptic hydrolysate of the total chymotryptic hydrolysate were all analyzed on RP-HPLC-MS. The NaCl gradients in the anion and cation

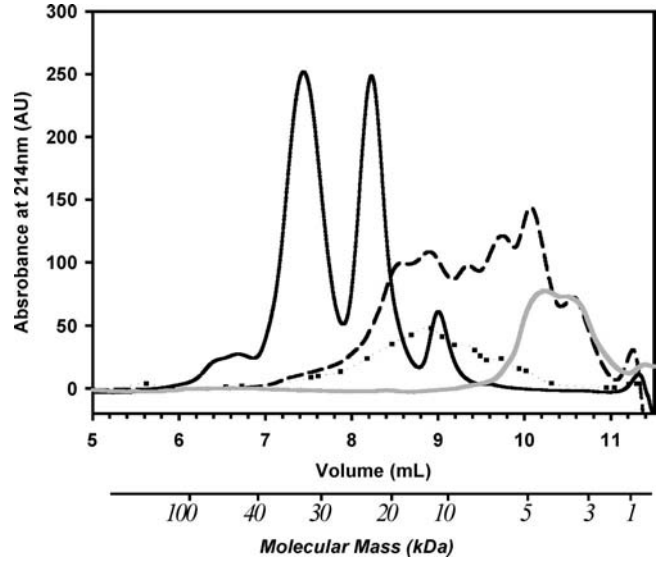

Figure 2. HP-SEC chromatograms under denaturing conditions of glycinin (—), glycinin hydrolysate at DH 2.2% (– –), the aggregating peptides (···), and the tryptic digest of aggregating peptides (gray curve). Samples were denatured and reduced prior to analysis.

Table 2. Molecular Mass Ranges of Glycinin and Glycinin-Derived Peptides

sample	molecular mass range (Da)
intact glycinin polypeptides	8000–40000
glycinin hydrolysate at DH = 2.2%	<25000
aggregating peptides of glycinin at DH = 2.2%	5000–25000
tryptic digest of aggregating peptides	<5000

chromatography were chosen in such a way to result in fractions that, when separated on RP-HPLC, gave a separation more or less identical to the RP-HPLC chromatogram shown in **Figure 4**, by which it is meant that a similar number of peaks and equal separation is obtained. A flatter or steeper gradient will result in more or fewer RP-HPLC chromatograms, respectively, which also has consequences for the number of peaks per chromatogram and their individual separation. All 53 chromatograms from the RP-HPLC analysis were integrated. All peaks with a peak area smaller than 0.4 mAU·min were rejected because this peak area was too close to the background noise. In total, 682 peaks were integrated with a combined peak area of 3861 mAU·min.

Identification and Quantification of a Peptide. Figure 4 shows a typical example of how the RP-HPLC analysis was combined with MS to assign peptide sequences to the peaks in RP-HPLC chromatograms. The RP-HPLC chromatogram in **Figure 4A** shows the separation of the peptides in a particular IEX column. The number of peaks observed is average compared to other chromatograms. This means that also chromatograms were present with more peaks, but also chromatograms with a lower number of peaks. The presented base peak MS spectrum (**Figure 4B**) shows a high similarity to the UV214 pattern in the RP-HPLC spectrum, taking into account the elution delay between the RP-HPLC and the MS of approximately 0.25 min. In **Figure 4C** is presented the mass scan at 41.77 min.

On the basis of one example of a RP-HPLC peak, in five steps it will be explained how the data that belong to this peak are used for the peptide mapping, which is described below.

Step 1: Peak/Peptide Selection. For this particular example peak 10 (as indicted with the arrow) in the RP-HPLC

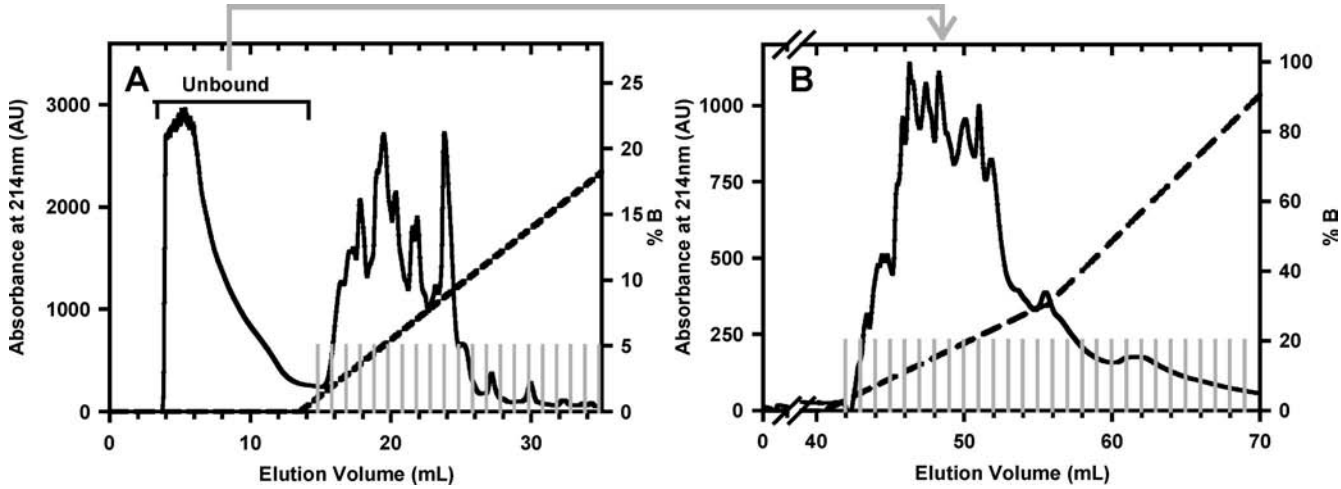

Figure 3. Anion exchange chromatogram (**A**) of tryptic digest and cation exchange chromatogram (**B**) of the unbound fraction from anion exchange chromatography. Gray lines represent the collected fractions. The dashed line represents the 2 M NaCl gradient (%B).

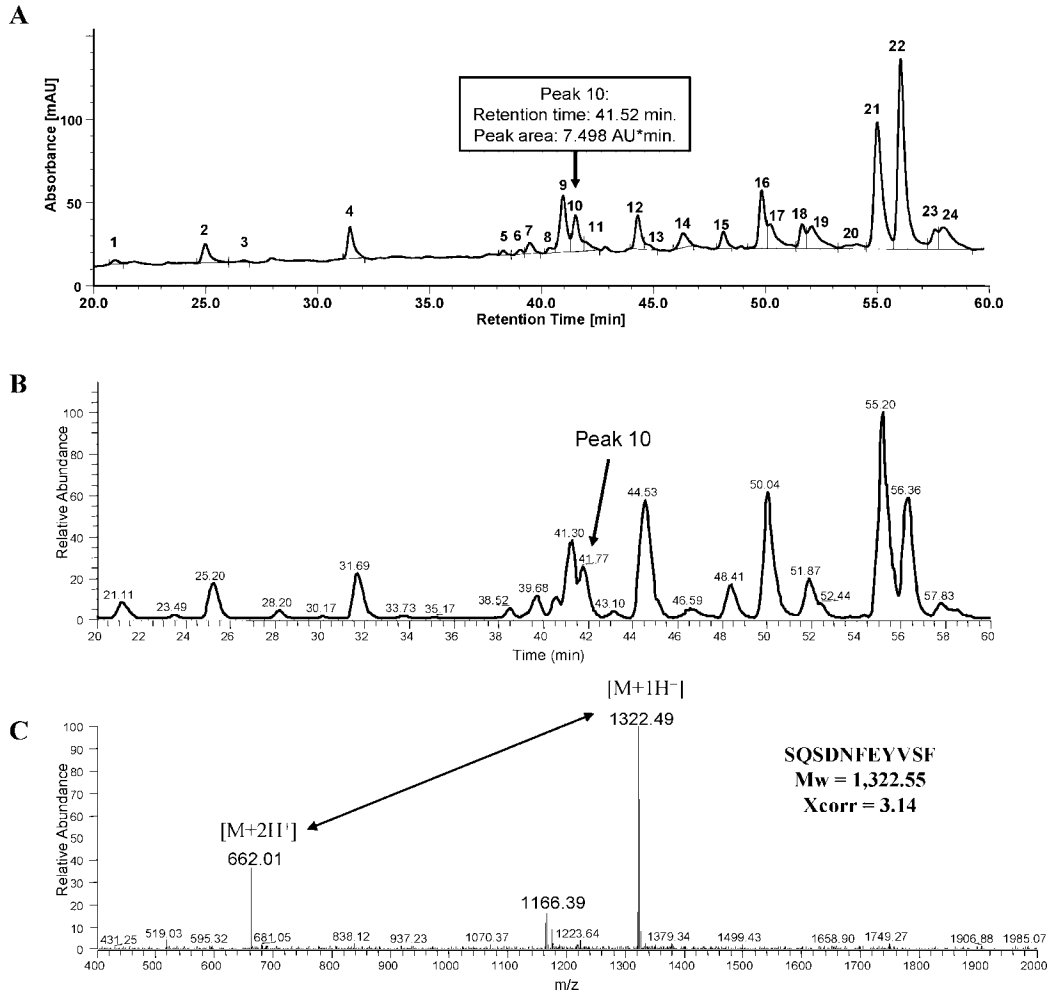

Figure 4. RP-HPLC-UV graph at 214 nm of an anion exchange fraction (**A**), its base-peak MS spectrum (**B**), and the mass scan at 41.77 min (**C**).

chromatogram (**Figure 4A**) with its maximum at 41.52 min is taken. This peak corresponds to the top of the peak at 41.77 min in the base-peak MS spectrum (**Figure 4B**). **Figure 4C** shows the full MS scan at 41.77 min. It can be seen that the spectrum of a peptide with an *m*/*z* of 1322.49 is the most dominant one. The MS/MS spectra of the minor peak with an *m*/*z* of 662.01 (+2) was similar to the MS/MS spectrum of the *m*/*z* 1322.49 (+1) peak. This confirms that both peaks belong to the same peptide, with a difference in charge. The other peaks present in **Figure 4C**, around an *m*/*z* of 1166.39, belong to the peptide that is responsible for peak 9 in **Figure 4A** but is still slightly present in peak 10 due to overlap of both peaks (no further data shown).

Step 2: Peptide Identification. Using BioWorks software and database searching (glycinin sequences as presented in Swiss-Prot), the sequence of the peptide in peak 10 was determined to be SQSDNFEYVSF, on the basis of the MS/MS data from

a double-charged ion (m/z 662.01), with an Xcorr of 3.14. As this Xcorr is well above the threshold value of 2.0, the sequence was accepted.

Step 3: Peptide Location. Next, it is determined in which subunit of glycinin this sequence is present. Glycinin can be present in five subunits, glycinins 1–5 (*17*). For peptide mapping, the contribution of one peptide to the peptide map has to be corrected for the molar proportion in which the different glycinin subunits, containing the identified peptide, are present. The five subunits can be divided into two subfamilies that are designated group I (glycinins 1–3) and group II (glycinins 4 and 5) (*18, 19*). Homologies between members of the same group range from 80 to 90%, and between the groups the homology is <50% (17). The database search toward the protein sequences of glycinin revealed that the sequence SQSDNFEYVSF is present in the basic polypeptides of glycinin G1 [f405–415], glycinin G2 [f396–406], and glycinin G3 [f389–399].

Step 4: Peptide Quantification per Glycinin Subunit. On the basis of the amino acid sequence determined in step 2, the molar extinction coefficient was calculated using formula 1 and the data presented in **Table 1**, resulting in an $\varepsilon_{\mathrm{SQSDNFEYVSF}} = 25564$ $\mathrm{M}^{-1}\,\mathrm{cm}^{-1}$. The area of peak 10 was 7.498 AU·min, indicating that the total concentration of the peptide with sequence SQSDNFEYVSF is 0.293 mMequiv using formula 2.

$$c_{\mathrm{peptide}}\ (\mathrm{Mequiv}) \sim \frac{A_{\mathrm{peak}}}{\epsilon_{\mathrm{peptide}}} \qquad (2)$$

If a peptide was present in more than one subunit, the subunits in which this peptide was present were all members of the same group. To calculate the concentration of one peptide in the peptide map of one subunit, its concentration has to be corrected for the molar proportion in which the subunits containing the sequenced peptide are present (F_{subunit}). F_{subunit} can be calculated by dividing the molarity of the subunit in which a peptide is present by the sum of the molarities of all the subunits in which the peptide occur. In this study, for simplicity reasons, due to the high sequence homology within groups I and II, it was assumed that within each group the subunits were present in similar molar ratios.

In addition to the correction for the presence of a peptide in more than one subunit, it might occur that different peptides constitute one peak in the RP-HPLC chromatogram. This was the case in 30 of the 185 peaks. In these cases it was assumed that each peptide contributes to the absorbance in an equal amount. The correction was done by dividing the peak area by the number of peptides present in one RP-HPLC peak ($n_{\mathrm{peptides\ in\ peak}}$).

Formula 3 shows the calculation of the relative concentration of a peptide in one subunit, which is similar to formula 2 but now extended with the correction factors F_{subunit} and $n_{\mathrm{peptides\ in\ peak}}$.

$$c_{\mathrm{peptide\ in\ 1\ subunit}}\ (\mathrm{Mequiv}) \sim \frac{A_{\mathrm{peak}}}{\varepsilon_{\mathrm{peptide}}} \times \frac{1}{n_{\mathrm{peptides\ in\ peak}}} \times F_{\mathrm{subunit}} \qquad (3)$$

Because SQSDNFEYVSF was the only peptide responsible for the absorbance of peak 10 and is present in three glycinin subunits, for F_{subunit} and $n_{\mathrm{peptides\ in\ peak}}$, 0.333 and 1 are used, respectively. This results in a relative concentration of SQSDNFEYVSF in one subunit of 0.098 mMequiv (0.293/3). Consequently, in the peptide mapping of an individual subunit, for each amino acid in this sequence, a value of 0.098 mMequiv is assigned.

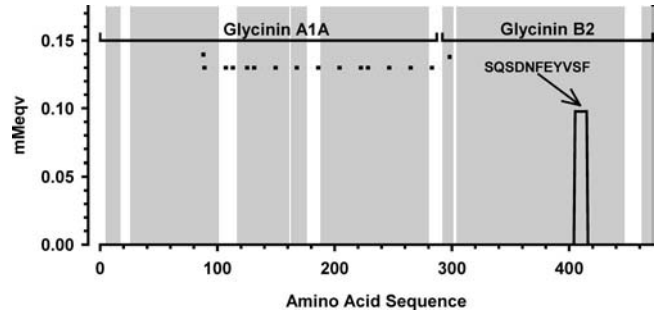

Figure 5. Illustration of how the calculated relative concentration of SQSDNFEYVSF is used in the peptide mapping to identify the region of glycinin 1 from which this peptide originates. The dotted line shows the location in the amino acid sequence where the disulfide bridge connects the acidic and the basic polypeptides. The gray regions indicate the coverage of the amino acid sequence that could be analyzed with MS.

Table 3. Overview of Different Parameters Related to the Interpretation of the 53 RP-HPLC Chromatograms for the Total Peak Area and the Total Analyzed Peak Area

	parameter	value
total peak area	total no. of RP-HPLC chromatograms	53
	total no. of peaks in 53 RP-HPLC chromatograms	682
	peak area (mAU·min)	3861
total analyzed peak area	peak area (mAU·min)	2896
	no. of peaks	185
	no. of peptides	218
	no. of positively identified peptide sequences	132
	peak area of the 132 identified peptide peaks (mAU·min)	2025
	no. of unique peptides	79

Step 5: Peptide Mapping. **Figure 5** shows the peptide map for glycinin 1, containing only the relative concentration of SQSDNFEYVSF of 0.098 mMequiv as calculated above. The gray regions in the peptide map represent the part of the glycinin 1 amino acid sequence that can be identified with MS. These data were obtained from the RP-HPLC-MS of the tryptic hydrolysate of the total chymotryptic hydrolysate, containing the aggregating and nonaggregating peptides. For glycinin 1 a coverage of ~90% was obtained.

When a digest of a protein is analyzed with RP-HPLC-MS, in general not 100% of the protein sequence is covered with the peptides identified (*20, 21*). Reasons for a coverage of lower than 100% could be because peptides are too hydrophilic to be separated on RP-HPLC, due to post-translational modifications, because not all peptides can be ionized, or because the amino acid sequence of the glycinin used does not fully match with the amino acid sequence used as present in Swiss-Prot.

Accumulative–Quantitative Peptide Mapping. Of the 682 peaks present in the 53 chromatograms, all large peaks (185 peaks), responsible for 75% (2896 mAU·min) of the total peak area of all the integrated chromatograms, were selected to identify their constituent peptide sequence on the basis of MS and MS/MS data. For all these 185 RP-HPLC peaks the five steps as presented above were followed.

In total 218 peptides were present in the 185 RP-HPLC peaks. The number of 218 is due to the presence of two or more peptides responsible for one peak. For 132 of the 218 peptides a sequence was positively identified and originated from glycinin. The reasons for not finding 218 sequences might be

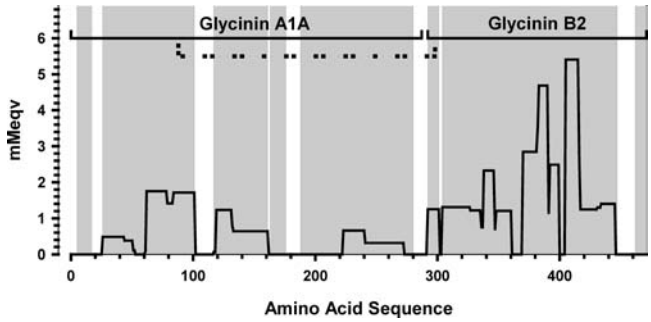

Figure 6. Quantitative peptide map of functional regions of glycinin 1 from which the aggregating peptides originate. The dotted line shows the location in the amino acid sequence where the disulfide bridge connects the acidic and the basic polypeptides. The gray regions indicate the coverage of the amino acid sequence that could be analyzed with MS.

diverse. One reason could be the coverage of the total amino acid sequence of the protein, which was 90% for glycinin 1, but equal or lower values were obtained for the other glycinin subunits (90, 87, 68, and 55% for glycinins 2, 3, 4, and 5, respectively). If the sequence of a peptide is part of the protein sequence that does not correspond for 100% with the sequence used in the database, the peptide will not be positively identified. Another reason for not identifying all peptide sequences might be that peptides originate from sources other than glycinin. They might originate from the enzymes (trypsin and chymotrypsin) or from β-conglycinin, which was present in low concentration in the glycinin preparation (*5*).

The 132 peptides contained 79 different unique glycinin peptide sequences. Of these 79 unique glycinin peptides, 28 originate from glycinin 1. Due to the aspecific fractionation in the IEX experiment, peptides having the same sequence were present in more than one fraction. In addition to this, also methionine can be oxidized, which resulted in an additional peak for the oxidized peptide in the RP-HPLC chromatograms. The 79 peptides with a known sequence represent a peak area of 2025 mAU·min. This means that 70% of the 75% RP-HPLC peak area is positively linked to a part of the glycinin peptides, indicating that 53% of the total peak area is assigned to peptide sequences. This value can be increased by determining >75% of the peak area. However, this will not change the general picture of the regions in the protein from which the functional peptides originate. We have found that when the total peak area studied is extended from 66 to 75%, the additionally identified peptide sequences mainly originated from those regions of the parental protein that already showed a significant quantitative accumulation when only 66% of the total peak area was analyzed (no further data shown).

An overview of the results of the interpretation of the 53 RP-HPLC chromatograms is given in **Table 3**.

Figure 6 shows the quantitative peptide map of functional regions of glycinin 1. In this figure all of the peptides from the 79 unique peptide sequences that are present in glycinin 1 are accumulatively mapped. The accumulative peptide map gives an overview of those regions of glycinin 1 that are likely to be present in the aggregating region. Mainly peptides from the basic polypeptide (B2) seem to be present in the aggregating regions and a part of the acidic polypeptide (A1A), around amino acid 80.

It can be concluded that the method described in this paper for quantitative peptide mapping using RP-HPLC-MS is sufficiently discriminating to identify which regions in a protein are predominantly responsible for the functional peptides derived from these proteins. This method is useful for protein hydrolysates, where functional peptides can be separated from peptides not contributing to the functionality of interest. It can, for example, also be applied to proteins that show resistance to enzymatic degradation (*22*). This is of interest to allergy research as well as in other digestion studies, in particular when proteins interact with other components, which hinders the digestion of proteins (*23*, *24*).

ACKNOWLEDGMENT

The soybeans were kindly provided by Fa. L. I. Frank (Twello, The Netherlands).

Supporting Information Available: Table containing all peptide sequences from glycinin 1 that are used to prepare **Figure 6**.

LITERATURE CITED

(1) Zhong, F.; Xin, Y.; Yue, L.; Shoemaker, C. F. Papain-induced gelation of soy glycinin (11S). *J. Food Sci.* **2006**, *71* (5), E232–E237.

(2) Mohri, M.; Matsushita, S. Improvement of water absorption of soybean protein by treatment with bromelain. *J. Agric. Food Chem.* **1984**, *32*, 486–490.

(3) Inouye, K.; Nagai, K.; Takita, T. Coagulation of soy protein isolates induced by subtilisin Carlsberg. *J. Agric. Food Chem.* **2002**, *50*, 1237–1242.

(4) Kuipers, B. J. H.; Alting, A. C.; Gruppen, H. Comparison of the aggregation behavior of soy and bovine whey protein hydrolysates. *Biotechnol. Adv.* **2007**, *25*, 606–610.

(5) Kuipers, B. J. H.; van Koningsveld, G. A.; Alting, A. C.; Driehuis, F.; Voragen, A. G. J.; Gruppen, H. Opposite contributions of glycinin- and β-conglycinin-derived peptides to the aggregation behavior of soy protein isolate hydrolysates. *Food Biophys.* **2006**, *1*, 178–188.

(6) Covey, T. Liquid chromatography/mass spectrometry for the analysis of protein digest. In *Protein and Peptide Analysis by Mass Spectrometery*; Chapman, J. R., Ed.; Humana Press: Totowa, NJ, 1996; p 350.

(7) Je, J. Y.; Park, P. J.; Byun, H. G.; Jung, W. K.; Kim, S. K. Angiotensin I converting enzyme (ACE) inhibitory peptide derived from the sauce of fermented blue mussel *Mytilus edulis*. *Bioresour. Technol.* **2005**, *96* (14), 1624–1629.

(8) Saiga, A.; Okumura, T.; Makihara, T.; Katsuta, S.; Shimizu, T.; Yamada, R.; Nishimura, T. Angiotensin I-converting enzyme inhibitory peptides in a hydrolyzed chicken breast muscle extract. *J. Agric. Food Chem.* **2003**, *51*, 1741–1745.

(9) Chatterton, D. E. W.; Sorensen, E. S.; Petersen, T. E.; Lonnerdal, B. Heparin binding peptides from the 23 N-terminal residues of α(s1)-casein present in commercial preparations of whey protein concentrate, glycomacropeptide and α-lactalbumin. *Milchwissenschaft* **2001**, *56* (3), 143–146.

(10) Forster-Waldl, E.; Marchetti, M.; Scholl, I.; Focke, M.; Radauer, C.; Kinaciyan, T.; Nentwich, I.; Jager, S.; Schmid, E. R.; Boltz-Nitulescu, G.; Scheiner, O.; Jensen-Jarolim, E. Type I allergy to elderberry (*Sambucus nigra*) is elicited by a 33.2 kDa allergen with significant homology to ribosomal inactivating proteins. *Clin. Exp. Allergy* **2003**, *33* (12), 1703–1710.

(11) Kuipers, B. J. H.; Gruppen, H. Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography–mass spectrometry analysis. *J. Agric. Food Chem.* **2007**, *55*, 5445–5451.

(12) AOAC. Method 990.03. *Official Methods of Analysis*, 15th ed.; AOAC International: Arlington, VA, 1990.

(13) Adler-Nissen, J. *Enzymic Hydrolysis of Food Proteins*, 1st ed.; Elsevier Applied Science Publishers: London, U.K., 1986; p 427.

(14) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **1999**, *17* (7), 676–682.

(15) Storms, H. F.; van der Heijden, R.; Tjaden, U. R.; van der Greef, J. Optimization of protein identification from digests as analyzed by capillary isoelectric focusing–mass spectrometry. *J. Chromatogr., B* **2005**, *824* (1–2), 189–200.
(16) Kuipers, B. J. H.; van Koningsveld, G. A.; Alting, A. C.; Driehuis, F.; Gruppen, H.; Voragen, A. G. J. Enzymatic hydrolysis as a means of expanding the cold gelation conditions of soys proteins. *J. Agric. Food Chem.* **2005**, *53*, 1031–1038.
(17) Nielsen, N. C.; Dickinson, C. D.; Cho, T. J.; Thanh, V. H.; Scallon, B. J.; Fischer, R. L.; Sims, T. L.; Drews, G. N.; Goldberg, R. B. Characterization of the glycinin gene family in soybean. *Plant Cell* **1989**, *1* (3), 313–328.
(18) Scallon, B.; Thanh, V. H.; Floener, L. A.; Nielsen, N. C. Identification and characterization of DNA clones encoding group-II glycinin subunits. *Theor. Appl. Genet.* **1985**, *70* (5), 510–519.
(19) Fischer, R. L.; Goldberg, R. B. Structure and flanking regions of soybean seed protein genes. *Cell* **1982**, *29* (2), 651–660.
(20) Lim, H.; Eng, J.; Yates, J. R.; Tollaksen, S. L.; Giometti, C. S.; Holden, J. F.; Adams, M. W. W.; Reich, C. I.; Olsen, G. J.; Hays, L. G. Identification of 2D-gel proteins: a comparison of MALDI/TOF peptide mass mapping to micro LC-ESI tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2003**, *14* (9), 957–970.
(21) Mann, M.; Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **2003**, *21* (3), 255–261.
(22) Maleki, S. J.; Kopper, R. A.; Shin, D. S.; Park, C. W.; Compadre, C. M.; Sampson, H.; Burks, A. W.; Bannon, G. A. Structure of the major peanut allergen Ara h 1 may protect IgE-binding epitopes from degradation. *J. Immunol.* **2000**, *164* (11), 5844–5849.
(23) Rawel, H. M.; Czajka, D.; Rohn, S.; Kroll, J. Interactions of different phenolic acids and flavonoids with soy proteins. *Int. J. Biol. Macromol.* **2002**, *30* (3–4), 137–150.
(24) Rawel, H. M.; Kroll, J.; Rohn, S. Reactions of phenolic substances with lysozyme–physicochemical characterisation and proteolytic digestion of the derivatives. *Food Chem.* **2001**, *72* (1), 59–71.

Received for review May 10, 2007. Revised manuscript received September 9, 2007. Accepted September 18, 2007. The research in this paper was supported by The Graduate School VLAG.

JF071380B