JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Identification of Strong Aggregating Regions in Soy Glycinin upon Enzymatic Hydrolysis

BAS J. H. KUIPERS§ AND HARRY GRUPPEN*

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

Upon hydrolysis with chymotrypsin, soy glycinin has a strong tendency to aggregate. The regions of glycinin from which the aggregating peptides originate were identified by accumulative−quantitative peptide mapping. To this end, the aggregating peptides were further hydrolyzed with trypsin to obtain peptides of which the sequence can be identified using RP-HPLC-MS/MS. This resulted in a hydrolysate in which 90% of the proteinaceous material was dissolved. The soluble fraction was analyzed using the method of accumulative−quantitative peptide mapping: fractionation using ion exchange chromatography, followed by identification of peptides by RP-HPLC-MS/MS, quantification based on the absorbance at 214 nm, and finally peptide mapping. For the peptide mapping the proportions in which each of the five glycinin subunits are present, as determined by Edman degradation, were taken into account. The results showed that mainly the basic polypeptide and a part of the acidic polypeptide, close to the location of the disulfide bridge between the basic and acidic polypeptides, are present in the aggregating peptide fraction. On the basis of the results obtained, an aggregation mechanism was proposed. The hydrophilic acidic polypeptides shield the hydrophobic basic polypeptides, and the former are preferentially degraded upon hydrolysis. This results in a net increase in hydrophobicity of the remaining material, which mainly consists of the basic polypeptide fragments. This increase in hydrophobicity is proposed to be the driving force in the aggregation of chymotrypsin-derived peptides of glycinin.

KEYWORDS: Proteins; peptides; soy; glycinin; mass spectrometry; MS; quantitative analysis; peptide mapping

INTRODUCTION

Various proteases derived from micro-organisms, plants, and animals are reported to have the ability to induce aggregation of soy proteins. This aggregation of soy proteins might be a good approach to improve the texture of soy protein foods such as soy protein based analogues of dairy products (*1, 2*).

Three microbial serine proteases have been studied in more detail with respect to their ability to induce soy protein aggregation. These proteases are subtilisin Carlsberg from *Bacillus licheniformis* (*3–5*) and proteases that originate from *Bacillus* sp. K-295G-7 (*6–8*) and *Bacillus pumilus* TYO-67 (*9, 10*). The plant proteases bromelain (*11, 12*) and papain (*13*) and the animal-derived protease chymotrypsin (*14*) were also shown to induce aggregation of soy proteins. Also, several industrially available protease preparations have the ability to induce soy protein aggregation (*15, 16*).

Various studies have been performed to understand the aggregation behavior of hydrolyzed soy proteins. Mechanistic studies to understand enzyme-induced aggregation of hydrolyzed soy proteins comprised rheological (*16*) and turbidity (*3, 4, 17*) measurements. Forces involved in the aggregation were shown to be mainly hydrophobic interactions, whereas the presence of disulfide bridges seemed to play a minor role (*12, 13*).

In the aggregation of soy protein isolate (SPI), it was shown that especially peptides derived from glycinin are the driving force for the aggregation upon hydrolysis with bacterium-, plant-, and animal-derived proteases (*12, 14*). It was concluded (*14*) that the ability of several proteases to induce aggregation of soy glycinin indicates that it is not the enzyme specificity that is of major importance to induce aggregation, but the ability of the substrate, glycinin, to yield strong aggregating peptides. This is in contrast to enzyme-induced aggregation of whey proteins in which the specificity of the protease is of major importance (*18, 19*).

Due to the complex composition of glycinin, understanding the aggregation behavior of soy glycinin hydrolysates is rather difficult. In total six different glycinin genes have been identified from soybeans. These are Gy1−Gy5 and Gy7, representing the glycinin subunits G1−G5 and G7, respectively. Gy7 has been recently (*20*) identified and is related to the other five soybean subunits. Because Gy 7 is poorly expressed, it is not further taken into account in this study. With the exception of G4, each

* 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/jf703781j CCC: $40.75
Published on Web 05/08/2008

glycinin subunit consists of one acidic (A) and one basic (B) polypeptide, which are connected with via a disulfide bridge. For the five major glycinin subunits six different acidic polypeptides (A1a, A2, A1b, A5, A4, and A3) and five basic (B2, B1a, B1b, B3, and B4) polypeptides have been identified (*21*). The A4 polypeptide of G4 is the only acidic polypeptide that is not covalently linked to a basic polypeptide. On the basis of their amino acid sequences as presented in the Swiss-Prot database (www.expasy.org), the molecular masses of the basic polypeptide can be calculated to be around 20 kDa, whereas those of the acidic polypeptides vary between 10.5 and 36.4 kDa. The five subunits can be divided into two subfamilies that are designated group I (G1, G2, and G3) and group II (G4 and G5) (*22, 23*). Homologies between members of the same group range from 80 to 90%, whereas between the groups the homology is <50% (*21*). At pH 7.6 and $I = 0.03$ M, glycinin can be present as a trimer and as a hexamer (*24*) having molecular masses of 150−180 and 300−360 kDa, respectively.

The aim of the present study is to understand the aggregation behavior of soy glycinin-derived peptides using the recently described method of accumulative−quantitative peptide mapping (*25*). This method identifies those regions in proteins from which peptides originate with specific functional properties. The method is applicable in systems in which the peptides of interest are too large to be identified directly with mass spectrometry and comprise a wide range of peptides with overlapping sequences that cannot be isolated individually.

MATERIALS AND METHODS

Materials and Chemicals. Soy glycinin was prepared from defatted soybean meal, prepared from Hyland soybeans (non-GMO) supplied by Fa. L. I. Frank (Twello, The Netherlands) as described previously (*4*). Chymotrypsin (TLCK treated) was obtained from Sigma Chemical Co. (St. Louis, MO; C-3142). 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 content of various samples was determined in duplicate according to the Dumas method 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 amino acid compositions of the five subunits as calculated from 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)].

Determination of the Proportions of Different Glycinin Subunits. Glycinin was dissolved in 700 μL of 8 M guanidinium chloride and 50 mM Tris-HCl buffer (pH 8.0) containing 15 mM DTT. The protein was reduced for 30 min at 60 °C. Iodoacetamide (IAA) was added up to 50 mM, and the mixture was incubated in the dark for 30 min to alkylate the peptides. Every 10 min the pH was checked and, if necessary, adjusted to 8.0 using 1 M NaOH. The alkylated glycinin was applied onto a Shodex Protein KW-803 size exclusion chromatography column (300 × 8 mm; Showa Denko K.K., Tokyo, Japan) equilibrated and run with 30% (v/v) aqueous acetonitrile containing 0.1% (v/v) trifluoroacetic acid (TFA). The flow rate was 0.5 mL/min, and the absorbance of the eluate was monitored at 214 nm (*4*). Fractions of 0.2 mL were collected. To separate the glycinin from salts and other proteinaceous impurities, fractions containing glycinin were pooled and freeze-dried.

The proportion of the various subunits of glycinin in the freeze-dried sample was estimated from the initial yields of the sequences of the individual acidic (and basic) polypeptides as reported in Swiss-Prot. These initial yields were obtained by performing a six-cycle N-terminal sequence analysis on a 23 μg sample of glycinin loaded onto a prewashed and Biobrene-coated glass filter. N-Terminal sequence analysis was performed by Edman degradation (*26, 27*) with a Procise 494 (Applied Biosystems, Foster City, CA) automated sequencing system (*28*), equipped with a 140 C Microgradient System and a 758A absorbance detector, and using protocols, reagents, chemicals, and materials from Applied Biosystems. The proportion (mole percent) of the various subunits was calculated on the basis of the known N termini of all polypeptides in glycinin. The six acidic polypeptides all have a unique N terminus, whereas the basic polypeptides within each group (group I, G1, G2, and G3; group II, G4 and G5) are identical in this respect. The results of three N-terminal sequence analysis cycles were essentially sufficient to calculate the ratio. Nevertheless, six cycles were used in the calculation to verify the correctness of the ratio found after the first three cycles.

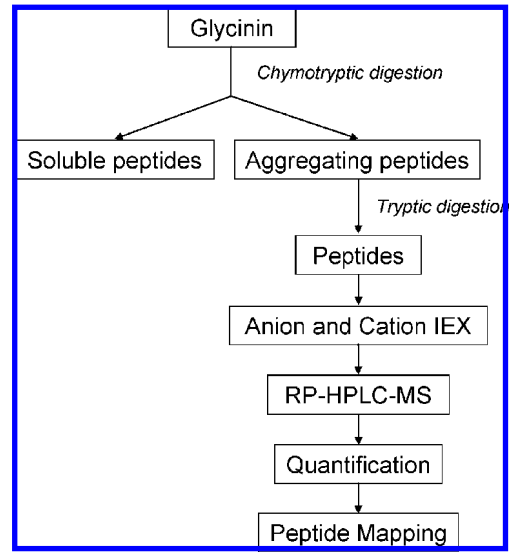

Figure 1. Schematic overview of accumulative−quantitative peptide mapping to identify regions in glycinin from which the aggregating peptides originate (*25*).

Accumulative−Quantitative Peptide Mapping. To identify those regions in glycinin from which aggregating peptides originate, the method of accumulative−quantitative peptide mapping was used as previously described in detail (*25*). The method is schematically summarized in **Figure 1** and will be briefly described here.

A 1% (w/w) solution of heated glycinin was hydrolyzed with chymotrypsin to a DH of 2.2% using the pH-stat method at pH 8.0 and 40 °C. When the desired DH was reached, the enzymatic hydrolysis was stopped by addition of phenylmethanesulfonyl fluoride (PMSF). Water and NaCl were added to the hydrolysate to result in a final protein concentration of 0.8% (w/w) and an ionic strength of 0.03 M, as described previously (*5*). After centrifugation (5 min, 5000*g*, 20 °C), analysis of the protein concentration (Dumas method) in the supernatant showed that ∼43% of the proteinaceous material was aggregated. The pellet was washed three times by resuspending it in Millipore water containing 0.03 M NaCl. After washing, the pH of the pellet was decreased to ∼2.5 using TFA to resolubilize the peptides, followed by freeze-drying. These freeze-dried peptides were further denoted aggregating peptides.

The aggregating peptides were dissolved in 20 mM Tris-HCl buffer (pH 8.0) containing 8 M urea and subsequently reduced and alkylated. After dilution to 1.6 M urea, the aggregating peptides were further hydrolyzed using trypsin at 37 °C. This tryptic digestion was performed in three cycles.

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 was added to the resolubilized pellet and incubated at 37 °C again. After incubation, the sample was treated in the same way as described above, by resuspending the pellet in 20 mM Tris-HCl buffer (pH 8.0) containing 8 M urea, dilution to 1.6 M urea with 20 mM Tris-HCl buffer (pH

8.0), and subsequent incubation with trypsin followed by centrifugation. The supernatants of the three tryptic digestion cycles were mixed prior to fractionation using ion exchange chromatography.

The combined supernatants of the tryptic digests were applied onto a 1 mL Mono Q 5/50GL anion exchange column (Amersham Biosciences, Uppsala, Sweden) equilibrated with a 20 mM Tris-HCl buffer (pH 8.0). Fractions of 1 mL were collected. The unbound fractions were decreased in pH to pH 3.5 using formic acid and applied onto a 1 mL Mono S 5/50GL cation exchanger (Amersham Biosciences) equilibrated with a 20 mM formic acid buffer (pH 3.5). Fractions of 1 mL were collected. Elution of the bound proteins on the anion and cation exchanger was performed using their corresponding equilibration buffers containing 2 M NaCl. The eluate was monitored at 214 nm. The unbound fraction of cation exchange chromatography was concentrated by freeze-drying.

All bound fractions collected from anion (25 fractions) and cation (27 fractions) exchange and the unbound fraction of the cation exchange were analyzed quantitatively using RP-HPLC-MS using a C18 column (218MS52; 250 × 2.1 mm, 5 μm) (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 Millipore water containing 0.1% (v/v) formic acid (solvent A) and acetonitrile containing 0.085% (v/v) formic acid (solvent B). All operations and data processing were controlled by Chromeleon version 6.7 software (Dionex Corp., Sunnyvale, CA). The eluate was applied onto 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. The instrument was controlled by Xcalibur software v1.3 (Thermo Finnigan). The MS/MS functions were performed in data-dependent mode. BioWorks software, version 3.1 (Thermo Electron, San Jose, CA) 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. To discriminate between correct and incorrect peptide sequence assignment, the cross-correlation value (Xcorr) for each identified peptide was used as a criterion (*29, 30*). For positive identification of the peptides, Xcorr thresholds of 1.5 for single charged peptides, 2.0 for double, and 2.5 for triple-charged peptides were used.

Of the 682 peaks present in the 53 chromatograms, all of the largest peaks (185 peaks), responsible for 75% (2896 mAU·min) of the total peak area of all 53 integrated chromatograms, were selected to identify their constituent peptide sequences based on MS and MS/MS data. In total, 218 peptides were present in the 185 RP-HPLC peaks. From these 218 peptides the sequences were positively identified of 132 peptides that originate from glycinin. The 132 peptides contained 79 different unique peptide sequences. The 79 peptides with known sequences represent an area of 2025 mAU·min, which means that 70% of the 75% RP-HPLC area is positively linked to a part of the glycinin sequence.

Quantitative analysis of the peptide concentration was performed using formula 1:

$$c_{\text{peptide in 1 subunit}}\ (\text{Meqv}) \sim \frac{A_{\text{peak}}}{\epsilon_{\text{peptide}}} \times \frac{1}{n_{\text{peptides in peak}}} \times F_{\text{subunit}} \quad (1)$$

In formula 1 the RP-HPLC peak area (A_{peak}) of each peak was divided by the theoretically calculated molar extinction coefficient ($\epsilon_{\text{peptide}}$) (*31*), corrected for the total number of different peptides present in the RP-HPLC peak ($n_{\text{peptides in peak}}$), and multiplied by the molar proportion (F_{subunit}) in which the subunits, containing the sequenced peptide, are present (*25*). In total, 79 sequences were identified that were used for accumulative−quantitative peptide mapping using all five glycinin subunits.

All 79 peptides identified, together with their relative concentrations, were used to build the quantitative peptide map as a projection on the amino acid sequence of each individual glycinin subunit (G1−G5).

The maximum sequence coverage that can be obtained was analyzed using a tryptic digestion of the total hydrolysate after chymotryptic hydrolysis of heated glycinin. This hydrolysate contains the aggregating, as well as the nonaggregating, peptides and was analyzed by RP-HPLC-MS in a similar way as the 53 fractions described above. For all glycinin subunits a total coverage was obtained of 90, 90, 87, 68, and 55% for G1, G2, G3, G4, and G5, respectively.

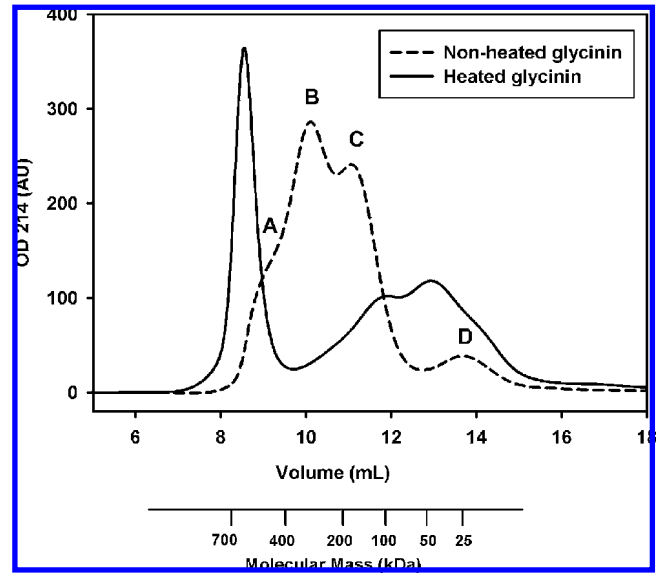

Figure 2. HP-SEC chromatograms of nonheated glycinin (------) and heated glycinin (—) in 10 mM sodium phosphate buffer (pH 8.0). The letters A, B, C, and D represent four fractions of the nonheated glycinin representing the 15S, 11S, 7S, and dissociation products, respectively.

In addition to the peptides present in the supernatant, the final pellet after tryptic digestion was dissolved in 5% (v/v) acetonitrile in Millipore water, containing 0.1% (v/v) formic acid and also analyzed on RP-HPLC-MS similar to all other analyzed samples.

High-Performance Size Exclusion Chromatography (HP-SEC). Aliquots (100 μL) of the 1% (w/w) glycinin solution used for the chymotryptic hydrolysis were taken before and after the heat treatment (30 min, 95 °C) and mixed with 900 μL of 10 mM sodium phosphate buffer (pH 8.0).

After 1 h of stirring at ambient temperature, the samples were applied onto a Superdex 200 HR 10/30 column (300 × 10 mm, Amersham Biosciences). The column was equilibrated with 10 mM sodium phosphate buffer (pH 8.0). The flow rate was 0.5 mL/min, and the absorbance of the eluate was monitored at 214 nm. The void volume of the column was approximately 7.8 mL as determined by injection of blue dextran. The column was calibrated using various proteins in a molecular mass range of 43−669 kDa (ovalbumin, catalase, ferritin, and thyroglobulin).

In addition to the above-described HP-SEC analysis, glycinin, its DH 2.2% hydrolysate, the aggregating peptide fraction, the total tryptic hydrolysate, and its supernatant and pellet were all analyzed using size exclusion chromatography under denaturing conditions (*5*). To allow quantitative comparison of the chromatograms obtained, the pellet after tryptic digestion was resuspended in the same volume as present prior to the centrifugation. Samples were dissolved under denaturing (8 M guanidinium chloride) and reducing conditions (100 mM DTT), and run on a Shodex Protein KW-803 column (300 × 8 mm; Showa Denko K.K.) equilibrated and run with 6 M urea containing 30% (v/v) acetonitrile and 0.1% (v/v) TFA. The aggregating peptide fraction was also applied on the Shodex KW-803 column in a reduced form. The flow rate was 0.5 mL/min, and the absorbance of the eluate was monitored at 214 nm. The column was calibrated using proteins with known molecular masses as described previously (*5*).

RESULTS AND DISCUSSION

Influence of Heat Treatment on the Quaternary Structure of Glycinin. To study the influence of the heat treatment (prior to hydrolysis) on the quaternary structure of glycinin, HP-SEC was performed of glycinin before and after heating. **Figure 2** shows the molecular size distribution patterns of glycinin before

and after heating. In the nonheated glycinin, four different fractions can be observed (A−D). On the basis of the calibration of the column, fractions B and C most likely represent the glycinin hexamer (300−360 kDa) and trimer (150−180 kDa), respectively. These hexamers and trimers are known to be present next to each other at neutral pH and low ionic strength (*24*). Fraction A has a higher apparent molecular mass than the hexamer and is probably a polymer of glycinin, also denoted 15S (*32*). Fraction D probably represents free glycinin subunits (50−62 kDa) or even acidic and basic polypeptides (10−36 kDa).

Upon heating it can be observed that, on the one hand, the hexamers and trimers form aggregates (peak around 8.2 mL) and, on the other hand, dissociate. The dissociated fraction contains proteins with sizes similar, but mainly smaller than, the glycinin trimer. These hexamers and trimers are probably dissociated into trimers, dimers, monomers and probably also individual polypeptides that are dissociated by the reduction of the disulfide bridge. When the peak areas of the chromatograms before and after heating are compared, it can be approximated that ∼50% of all the proteinaceous material is present in the aggregates, wheras the other ∼50% is present as dissociation products. Because the heated glycinin has a clear appearance, the formed aggregates can be denoted soluble aggregates. The observations are in agreement with the work of Mori and co-workers (*33*) and Yamauchi and co-workers (*34*), who described the formation of soluble aggregates upon heating of glycinin (pH 7.6, $I = 0.5$ M), as well as a dissociation of the individual subunits in their constituent polypeptides. However, the insolubilization of the basic polypeptides, as described by Mori and co-workers (*33*), was not observed, which might be due to the low ionic strength (proteins were dissolved only in Millipore water) during heating in the present study.

Effect of Disulfide Bridges on the Size of Aggregating Peptides. Due to the heat-induced aggregation (**Figure 2**) prior to hydrolysis, it might be that new intermolecular disulfide bridges are formed (*35*) that can influence the observed aggregation of peptides. **Figure 3A** shows the size distribution obtained by HP-SEC under denaturing conditions of the reduced glycinin and reduced and nonreduced aggregating peptide fraction. When the reduced hydrolyzed sample is compared with the nonreduced hydrolyzed sample, it can be seen that there is a slight shift to larger sizes from the nonreduced sample. This indicates that in the aggregating peptides disulfide bridges are present that slightly increase the size of the fragments. Nevertheless, this shift is not large, indicating that no large number of intermolecular disulfide bridges is formed upon heating.

Monitoring the Tryptic Hydrolysis. Because the aggregating peptides are too large to be identified directly with MS, these peptides were further degraded with trypsin prior to analysis. To monitor the hydrolysis of the isolated aggregating peptides with trypsin, HP-SEC under denaturing conditions was performed for the aggregating peptides and the total tryptic hydrolysate as well as its supernatant and pellet. It can be observed in **Figure 3B** that the aggregating peptide fraction mainly contains peptides with apparent molecular masses in the range of ∼5000−25000 Da. Upon tryptic hydrolysis the aggregating peptides are reduced to peptides with apparent molecular masses below 5000 Da. After repeated hydrolysis with trypsin, there were still some aggregates present in the hydrolysate. These aggregates could be dissolved and analyzed by HP-SEC under denaturing conditions. The molecular masses of these peptides range between 3000 and 5000 Da. The peak area of the pellet represents ∼10% of the peak area of the total

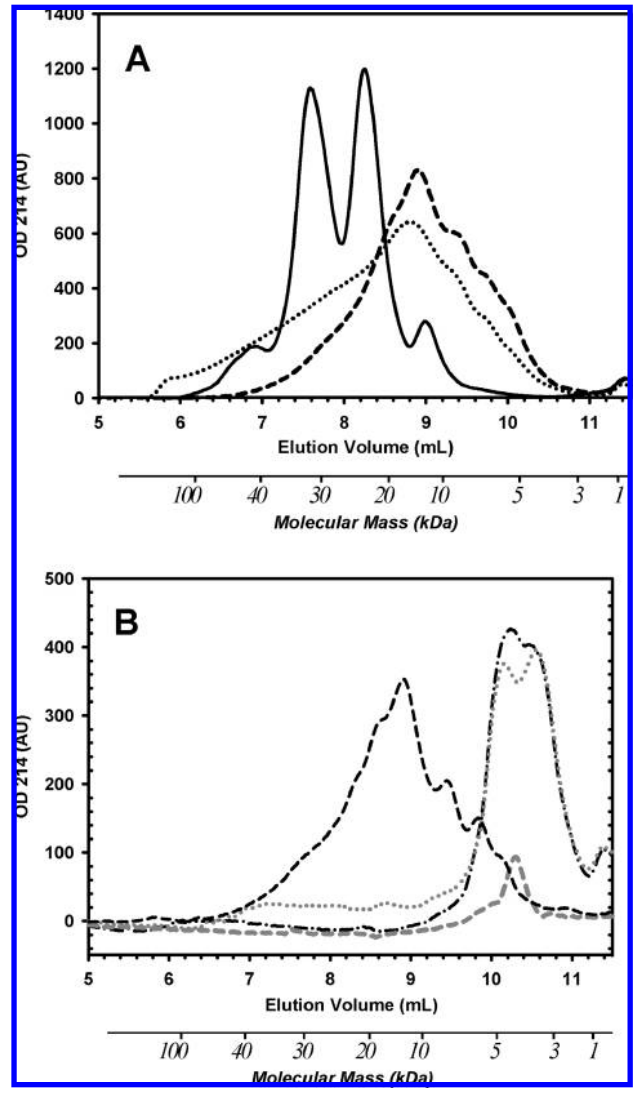

Figure 3. HP-SEC chromatograms under denaturing conditions [6 M urea, 30% (v/v) acetonitrile, 0.1% (v/v) TFA]. (**A**) Intact reduced glycinin (—) and its reduced (— —) and nonreduced (· · ·) aggregating peptides. (**B**) Aggregating peptides (— —), and its corresponding total tryptic hydrolysate of aggregating peptides (— · —), its supernatant (gray dots), and pellet (gray dashes). All samples used for the chromatograms in **B** were reduced prior to analysis.

tryptic hydrolysate. It is assumed that this 10% of the peak area also represents 10% of the proteinaceous material. This means that 90% of the proteinaceous material of the aggregating peptides was present in the supernatant that was further fractionated with ion exchange chromatography and analyzed with RP-HPLC-MS followed by quantification and peptide mapping.

Relative Proportions of the Five Glycinin Subunits. In *accumulative−quantitative peptide mapping* it may occur that a peptide that is identified is present in more than one subunit of the parental protein (*25*). To present a correct peptide mapping, the distribution of the quantity of such a peptide over the different subunits in which the peptide is present should be according to the molar proportions in which these subunits are present in the total parental protein. To be able to apply this correction, the molar ratios between all five glycinin subunits was determined using N-terminal sequence analysis of the glycinin polypeptides.

Table 1 shows the absolute concentration measured (picomoles) and the relative proportion (mole percent) of the in-

Table 1. Initial Yield (Absolute Amount) and the Relative Proportion of Individual Polypeptides in Glycinin, As Determined by N-Terminal Sequencing

polypeptide	N terminus	initial yield (pmol)	relative proportion (mol %)
acidic polypeptides (100%)			
G1 (A1a)	Phe-Ser-Ser-Arg-Glu-Gln	30	36
G2 (A2)	Leu-Arg-Gly-Gln-Ala-Gln	25	30
G3 (A1b)	Phe-Arg-Glu-Gln-Pro-Gln	5	6
G4 (A5)	Ile-Ser-Ser-Ser-Lys-Leu	11	14
G4 (A4)	Arg-Arg-Gly-Ser-Arg-Ser	12	
G5 (A3)	Ile-Thr-Ser-Ser-Lys-Phe	12	14
basic polypeptides (100%)			
G1−G3 (B2, B1a, B1b)	Gly-Ile-Asp-Glu-Thr-Ile	55	73
G4−G5 (B3, B4)	Gly-Val-Glu-Gly-Asn-Ile	20	27

dividual polypeptides based on the sequence yield of the various amino acids. These values are determined using the known N-terminal sequences of all acidic and basic polypeptides as presented in the Swiss-Prot database. Whereas the N termini of the various acidic polypeptides differ from each other, the basic polypeptides of G1, G2, and G3 have the same N terminus, which also counts for the N termini of G4 and G5. Therefore, eight different N-terminal sequences should in principle be distinguishable during the Edman degradation (results not shown). It can, however, be seen that not all amino acid residues are unique for one acidic polypeptide (i.e., one glycinin subunit). The Phe signal in the first Edman cycle, for example, is a combination of the contributions of A1a and A1b. Similar to that, the Ile signal in the first cycle is due to both A5 and A3. On the other hand, the Leu signal and the Arg signal in the first cycle are unique and thus give a direct indication of the relative contributions of subunits G2(A2) and G4(A5), respectively. In a similar way the data in the other Edman cycles can be used to estimate the relative contributions of the different subunits. In addition, these cycles are used to verify the validity of these estimations. The signals belonging to the basic polypeptides are used as a final verification. The sum of the contributions of the acidic polypeptides of G1, G2, and G3 (72%) and the sum of the contributions of G4 and G5 (27/28%) are equal to the relative contributions of the basic polypeptides of G1−G3 (73%) and G4−G5 (27%), respectively.

On the basis of the proportions of acidic polypeptides it can be observed that G1 and G2 are the most dominant subunits present in glycinin. Subunit G3 is present in a rather low proportion. G4 and G5 are present in a 1:1 ratio. The relative proportions (mole percent) shown in **Table 1** are used to calculate F_{subunit} (formula 1) for each peptide that was identified to be present in more than one subunit.

Peptide Mapping for All Glycinin Subunits. Peptide mapping for the five glycinin subunits is performed on the basis of the results obtained from the 90% soluble proteinaceous material after tryptic digestion of the aggregating peptides. **Figures 4** and **5** show the accumulative peptide maps for glycinins 1−3 and glycinins 4 and 5, respectively. The peptide map of glycinin 1 is the same as that previously shown (*25*). The accumulative peptide map gives an overview of those regions in each subunit that most likely yield aggregating peptides. The higher the bar at a specific region in the sequence, the stronger the contribution of the amino acid sequence of this region to the aggregating properties of the peptides derived from the corresponding subunit.

On top of each graph is indicated which part of the sequence represents which polypeptide. The dotted lines between the different polypeptides represent the disulfide bridge that connects the acidic and basic polypeptides. The gray regions represent the coverage of the amino acid sequence that could be identified with MS. These coverage data were obtained using RP-HPLC-MS of the tryptic hydrolysate of the total chymotryptic hydrolysate, containing both the aggregating and the nonaggregating peptides as described previously (*25*).

Glycinins 1 and 2 show (**Figure 4**) the highest quantitative response in the peptide maps compared to glycinins 3−5. When the acidic and basic polypeptides are compared, it can be seen that a relatively high proportion of the basic polypeptide is identified to be present in the aggregating peptides. In addition, the concentration determined (height of the bars) is higher for the basic polypeptides than for the acidic polypeptide. For the acidic polypeptide of glycinin 1 it seems that the highest proportion of positively identified peptides can be observed around the location of the disulfide bridge. This is less apparent for glycinin 2.

For glycinin 3 the response is much lower compared to the responses of glycinins 1 and 2. This is probably due to the much lower proportion of glycinin 3 in the glycinin preparation (**Table 1**). Nevertheless, for glycinin 3 as well it can be observed that the response and coverage in the basic polypeptide are higher than those of the acidic polypeptide.

For both glycinins 4 and 5 a lower coverage of the amino acid sequence is obtained (gray regions), which may contribute to their lower overall responses. **Figure 5** shows that also for glycinin 4 the response for the basic polypeptide is higher than in the acidic polypeptide. Also, the acidic polypeptide gives the highest response around the disulfide bridge. No peptides were detected from the A4 polypeptide of glycinin 4. For glycinin 5 a low response for the basic polypeptide could be expected, due to its very low overall sequence coverage from the MS data. As a consequence, glycinin 5 shows the highest response in the acidic polypeptide around the disulfide bridge.

Overall, it can be observed that next to the basic polypeptide, the region of the acidic polypeptide, which contains the disulfide bridge connecting the acidic with the basic polypeptide, is dominantly present in the aggregating peptides in all five glycinin subunits. It is shown that the presence of disulfide bridges slightly increases the size of the aggregating peptides (**Figure 3A**). Therefore, it is likely that this disulfide bridge is still present in the aggregating peptides between the peptides originating from the basic and acidic polypeptides. As a result, this may facilitate the (co)precipitation of peptides of the acidic polypeptide that are covalently connected to the aggregating peptide of the basic polypeptide. However, the disulfide bridge between the acidic and basic polypeptides is located in a hydrophobic region of the acidic polypeptide. If this disulfide bridge would be reduced upon heating (**Figure 2**), it seems likely that peptides originating from this region of the acidic polypeptide have the ability to aggregate when not covalently connected to the basic polypeptide.

It can be concluded that most aggregating peptides originate from the basic polypeptides. Because the molecular masses of the aggregating peptides are between 5000 and 25000 Da and the basic polypeptides are not intact anymore, it can be concluded that the basic polypeptide is only slightly degraded and, thus, seems to show a rather high resistance toward enzymatic hydrolysis. This is in agreement with previous observations that the basic polypeptides are degraded more slowly than the acidic polypeptide, when hydrolyzed with bromelain (*12*), and are not degraded at all upon hydrolysis with a serine protease of *B. pumilus* (*9*).

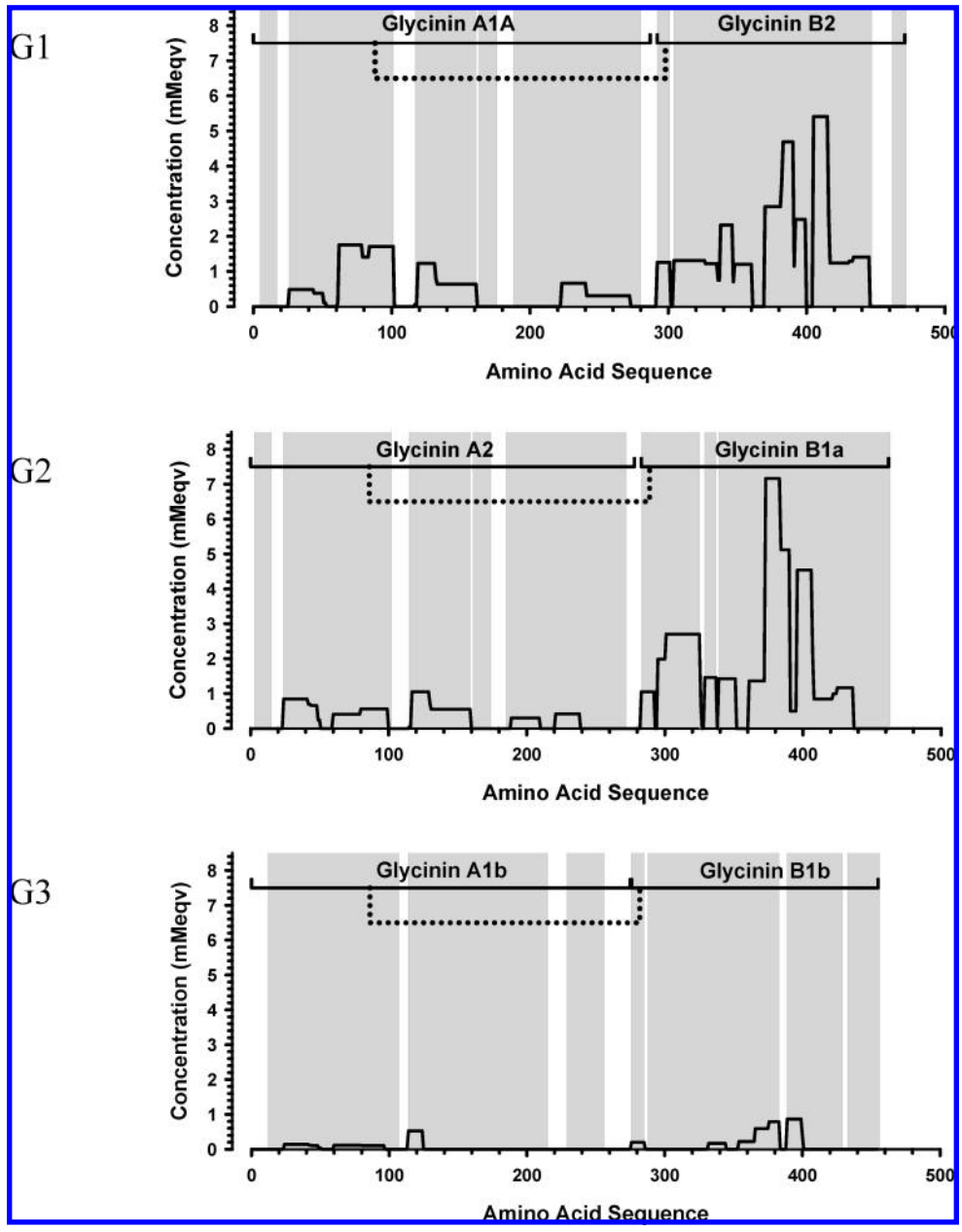

Figure 4. Accumulative quantitative peptide maps of glycinins 1, 2, and 3 from which the aggregating peptides originate. The dotted line shows the location in the amino acid sequence at which the disulfide bridge connects the acidic and basic polypeptides. The gray regions indicate the coverage of the amino acid sequence that could be analyzed with MS. The peptide map of glycinin 1 is the same as that previously shown (*25*).

The studied aggregating peptides were obtained at DH = 2.2% and pH 8, because at this DH, the quantity of aggregating material was highest compared to other DH and pH values (*5*). The aggregating peptides at other DH values were not studied using accumulative−quantitative peptide mapping. However, it can be hypothesized that the aggregating peptides at those DH values originate mainly from the basic polypeptide as well. The lower quantity of aggregating proteinaceous material can be due to smaller connected acidic polypeptide fragments and a further degradation of the basic polypeptide.

Aggregating Peptides in the Pellet after Tryptic Digestion. After repetitive tryptic digestion, peptides with a strong tendency to aggregate were still present, yielding a precipitate. They comprise ~10% of the total amount of aggregating peptides. These peptides were analyzed with RP-HPLC-MS (chromatogram not shown) to identify their amino acid sequences followed by determination of their locations in the sequence of the various glycinin subunits. Because 10% of the peptides were present in the pellet after tryptic digestion, the total peak area of the integrated peaks was adjusted to result in a peak area ratio of 9:1 between the RP-HPLC chromatograms of the soluble and insoluble peptide fractions after tryptic digestion. This correction enables a quantitative comparison with the peptide maps in **Figures 4** and **5**. The eight largest peaks, responsible for 75% of the total peak area in the pellet fraction, were studied. From six of these eight peaks the amino acid sequences of the constituting peptides could be positively identified. These peptides can be observed in **Table 2** together with their corresponding hydrophobicity. One of the six peptides is present in two glycinin subunits (glycinins 1 and 3). These six peaks account for 91% of the total peak area of the eight largest peaks and are mapped quantitatively in **Figure 6**. The locations of the polypeptides and the disulfide bridges are shown for all five glycinin subunits.

It can be seen that four of the six identified peptides originate from basic polypeptides, whereas two of them originate from

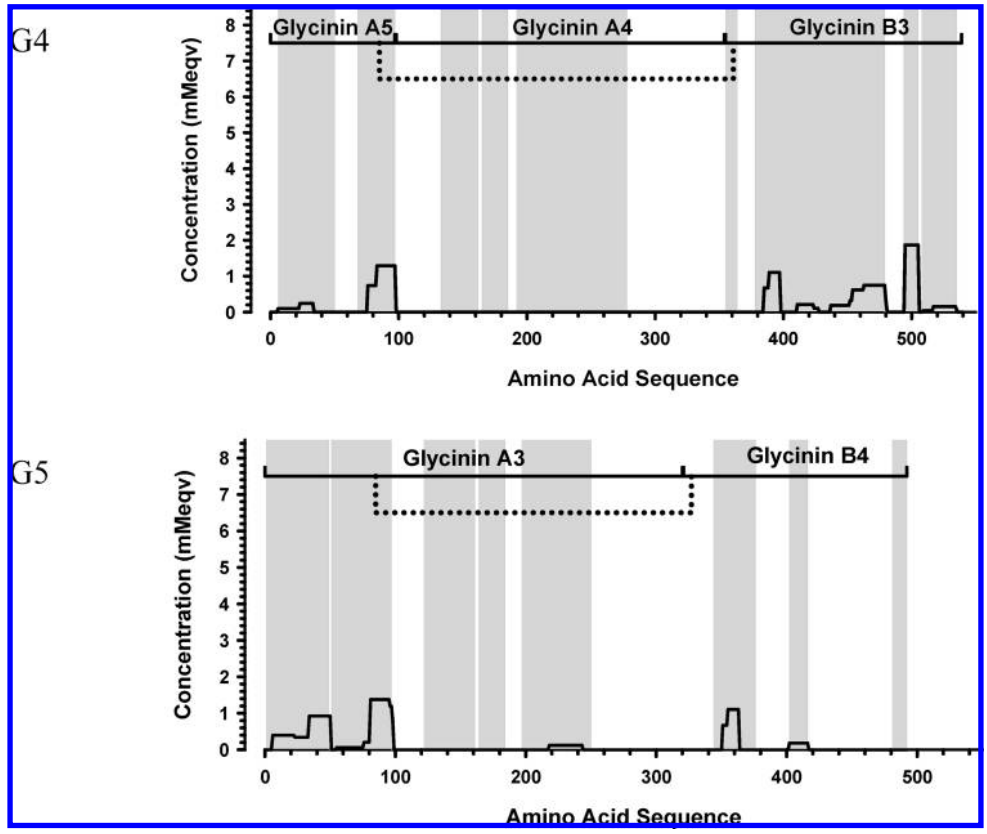

Figure 5. Accumulative quantitative peptide maps of regions of glycinins 4 and 5 from which the aggregating peptides originate. The dotted line shows the location in the amino acid sequence at which the disulfide bridge connects the acidic and basic polypeptides. The gray regions indicate the coverage of the amino acid sequence that could be analyzed with MS.

Table 2. Peptides Identified in the Final Pellet Obtained after Repetitive Tryptic Digestion, Their Calculated Hydrophobicities, and Their Locations in the Different Subunits

subunit	sequence	hydrophobicity[a]	location
glycinin 1	NAMFVPHYNLNANSIIYALNGR	0.171	B2 [f348−369]
glycinin 2	RPSYTNGPQEIYIQQGNGIFGM IFPGCPSTY	0.198	A2 [f60−90]
	NAMFVPHYTLNANSIIYALNGR	0.204	B1a [f339−359]
glycinin 3	NAMFVPHYNLNANSIIYALNGR	0.171	B1b [f332−353]
glycinin 4	NGIYSPHWNLNANSVIY	0.210	B4 [f411−427]
	NGIYSPHWNLNANSVIYVTR	0.104	B4 [f411−430]
glycinin 5	NGSHLPSYLPYPQMIIVVQGK	0.227	A3 [f55−75]

[a] Calculated according to the method of Eisenberg (*36*).

the acidic polypeptide, covering the region around the location of the disulfide bridge. All identified peptides originate from a region in glycinin that was already defined in the peptide maps (**Figures 4** and **5**) as being a region contributing to the aggregation of glycinin upon chymotryptic hydrolysis. The average hydrophobicity scores of all intact acidic and basic polypeptides are −0.228 (±0.048) and 0.018 (±0.034), respectively. The higher the score, the higher the hydrophobicity. When these values are compare with the values of the peptides in **Table 2**, it can be concluded that the latter peptides are relatively hydrophobic. This relatively high hydrophobicity is, therefore, probably why these peptides are not soluble.

It can be concluded that the sequences of peptides that aggregate after repetitive tryptic digestion do not reveal new regions of glycinin that strongly contribute to the aggregating regions. They were shown to originate from the same regions as already defined as aggregating regions on the basis of the analysis of 90% of the aggregating peptides.

As already mentioned above, the hydrophobicity is likely to play an important role in the aggregation of glycinin-derived peptides. Therefore, in **Figure 7** the hydrophobicity according to Eisenberg (*36*) (gray line) as a function of the amino acid sequence for glycinin 1 is shown. In addition, the total accumulative quantitative peptide map, being a combination of **Figures 4** and **6**, is shown. The calculation of the hydrophobicity was performed using the ProtScale software from the Swiss Institute of Bioinformatics (www.expasy.org), using a "window size" of 15 (= peptide range over which the average hydrophobic score is calculated). If the hydrophobicity score is above 0, the region in the sequence can be regarded as being hydrophobic. It shows that the hydrophobic regions are strongly present in the aggregating peptide, revealing the large contribution of hydrophobic interactions in the aggregation mechanism of the glycinin-derived peptides. This graph is shown for only glycinin 1, because all subunits showed the same pattern.

Mechanism of Enzyme-Induced Aggregation. Recently, Nagai and Inouye (*17*) proposed a mechanism for the enzyme-induced aggregation of soy protein isolate upon hydrolysis with subtilisin Carlsberg. First, the hydrophilic surface areas of the proteins are degraded, subsequently followed by the degradation of the hydrophobic core of the protein. This results in a strong decomposition of the tertiary structure of the proteins followed by aggregation due to hydrophobic interactions. However, no attention was paid to the quaternary structure of the proteins prior to hydrolysis, and the aggregating fractions were not identified with respect to the regions from which they originate in the SPI. In the present study the quaternary structure of the proteins prior to hydrolysis was studied together with the regions from which the peptides originate.

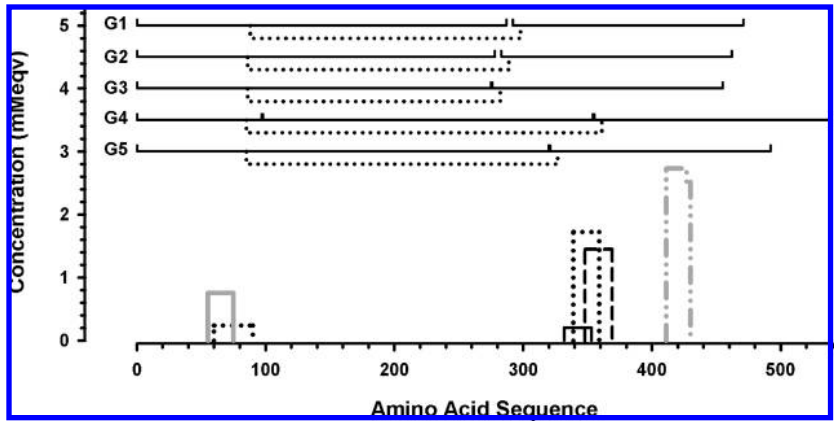

Figure 6. Accumulative quantitative peptide map of peptides in the pellet fraction after repetitive tryptic digestion of the aggregating peptides originating from G1 (— —), G2 (· · ·), G3 (—), G4 (gray dash-dot line), and G5 (gray solid line). The dotted line shows the location in the amino acid sequence at which the disulfide bridge connects the acidic and basic polypeptides.

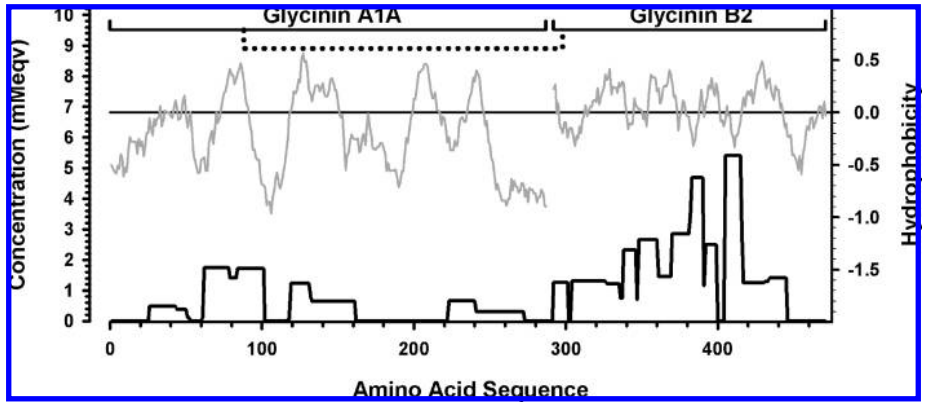

Figure 7. Hydrophobicity plot according to Eisenberg (*36*) (gray line) for glycinin 1 combined with its corresponding accumulative quantitative peptide map (black line), including the identified peptide in the pellet fraction as shown in **Table 2**.

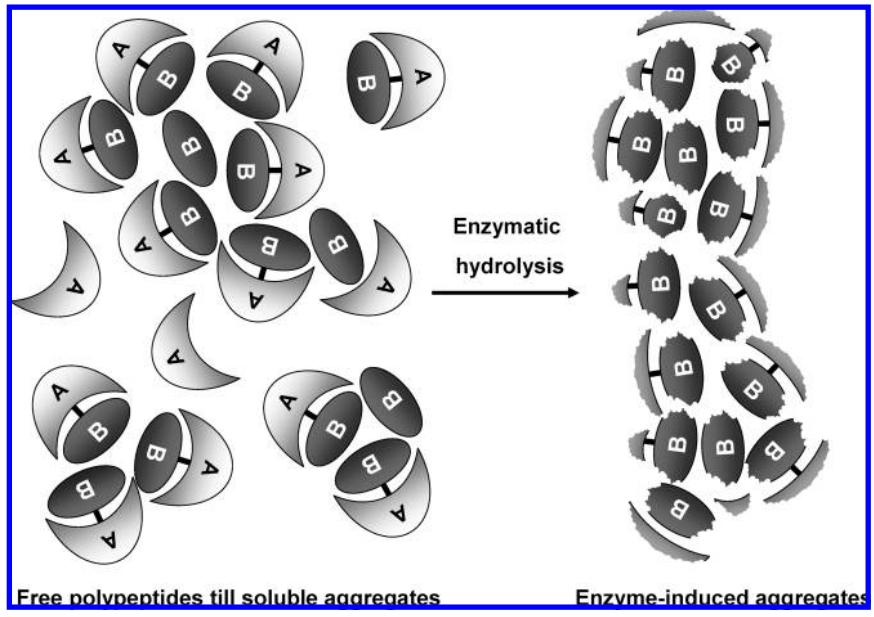

Figure 8. Proposed model for aggregate formation upon hydrolysis of soluble glycinin aggregates and its dissociation products.

Figure 8 visualizes the probable aggregation mechanism of soy glycinin-derived peptides. It schematically shows the quaternary structure of heated glycinin molecules that comprise ~50% of the aggregated material, whereas the other ~50% comprises the glycinin dissociation products (trimers, dimers, monomers, and individual acidic and basic polypeptides). The possible different contributions of soluble aggregates and glycinin dissociation products to the aggregation of the glycinin peptides was not studied. Therefore, in the proposed model it is assumed that each form in which glycinin is present prior to hydrolysis participates equally in the subsequent hydrolysis and aggregation.

In the hexameric and trimeric structures of nonheated glycinin the basic polypeptides are more or less buried in the tri- or

hexamer, shielded by the hydrophilic acidic polypeptides that face the hydrophilic solution (*24*), which promotes good solubility. As the aggregates and monomers formed upon heating are soluble, they still might possess a more or less organized structure with a hydrophilic outside (acidic polypeptides) and a hydrophobic core (basic polypeptide) as shown in **Figure 8** and also described in the model of Nagai and Inouye (*17*).

Upon enzymatic hydrolysis the hydrophilic outside of the soluble glycinin aggregates, and glycinin dissociation products (acidic polypeptides) might be preferentially degraded. In contrast, the hydrophobic core of the soluble aggregates and monomers (basic polypeptide) is only degraded to a low extent. The remaining proteinaceous material, being mainly peptides originating from the basic polypeptide and a fraction of the acidic polypeptide, increases in hydrophobicity with increasing DH and at a certain point starts to form aggregates.

In conclusion, the results show that aggregation of heated glycinin upon chymotryptic hydrolysis is mainly due to the aggregation of peptides originating from the basic polypeptide. In addition, those parts of the acidic polypeptide in the region where the disulfide bridge connects the basic polypeptides to the acidic polypeptides are also present in the aggregating peptides. Because the aggregation of glycinin is mainly substrate and, to a lesser extent, enzyme-dependent (*14*), it is likely that enzyme-induced aggregation of glycinin follows a similar mechanism upon hydrolysis with other enzymes. The understanding of which regions of the soy protein predominantly aggregate upon enzymatic hydrolysis can help food scientists and technologists in the development of protein hydrolysates. This knowledge can be used in the selection of enzymes and hydrolysis conditions to either increase or decrease the quantity of aggregating material, depending on the hydrolysate properties aimed for.

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 to 5 that were used to prepare Figures 4 and 5. This material is available free of charge via the Internet at http://pubs.acs.org.

LITERATURE CITED

(1) Lee, C. W.; Jo, G. Y. The microstructures of soybean milk curds prepared by different coagulation methods. *J. Food Sci. Nutr.* **1997**, *2*, 259–262.

(2) Murata, K.; Kusakabe, I.; Kobayashi, H.; Kiuchi, H.; Murakami, K. Functional-properties of 3 soymilk curds prepared with an enzyme, calcium salt and acid. *Agric. Biol. Chem.* **1988**, *52* (5), 1135–1139.

(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.; 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 soy proteins. *J. Agric. Food Chem.* **2005**, *53*, 1031–1038.

(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) Park, Y. W.; Kusakabe, I.; Kobayashi, H.; Murakami, K. Production and properties of a soymilk-clotting enzyme system from a microorganism. *Agric. Biol. Chem.* **1985**, *49* (11), 3215–3219.

(7) Park, Y. W.; Kusakabe, I.; Murakami, K. Purification and characterization of soymilk-clotting enzymes from *Bacillus* sp. K-295G-7. *Agric. Biol. Chem.* **1987**, *51* (9), 2343–2349.

(8) Kobayashi, H.; Park, Y. W.; Kaneda, M.; Kusakabe, I.; Yoshida, S.; Murakami, K. Substrate-specificity of soymilk-clotting enzyme-I and enzyme-II from *Bacillus* dp. K-295G-7 toward peptide-substrates. *Agric. Biol. Chem.* **1988**, *52* (9), 2351–2353.

(9) Aoyama, M.; Yasuda, M.; Nakachi, K.; Kobamoto, N.; Oku, H.; Kato, F. Soybean-milk-coagulating activity of *Bacillus pumilus* derives from a serine proteinase. *Appl. Microbiol. Biotechnol.* **2000**, *53* (4), 390–395.

(10) Yasuda, M.; Aoyama, M.; Sakaguchi, M.; Nakachi, K.; Kobamoto, N. Purification and characterization of a soybean-milk-coagulating enzyme from *Bacillus pumilus* TYO-67. *Appl. Microbiol. Biotechnol.* **1999**, *51* (4), 474–479.

(11) Fuke, Y.; Sekiguchi, M.; Msuoka, H. Nature of stem bromelain treatments on the aggregation and gelation of soybean proteins. *J. Food Sci.* **1985**, *50*, 1283–1288.

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

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

(14) 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* (6), 606–610.

(15) Murata, K.; Kusakabe, I.; Kobayashi, H.; Akaike, M.; Park, Y. W.; Murakami, K. Studies on the coagulation of soymilk-protein by commercial proteinases. *Agric. Biol. Chem.* **1987**, *51* (2), 385–389.

(16) Zhong, F.; Wang, Z.; Xu, S. Y.; Shoemaker, C. F. The evaluation of proteases as coagulants for soy protein dispersions. *Food Chem.* **2007**, *100* (4), 1371–1376.

(17) Nagai, K.; Inouye, K. Insights into the reaction mechanism of the coagulation of soy protein isolates induced by subtilisin Carlsberg. *J. Agric. Food Chem.* **2004**, *52* (15), 4921–4927.

(18) Ju, Z. Y.; Otte, J.; Madsen, J. S.; Qvist, B. Effects of limited proteolysis on gelation and gel properties of whey protein isolate. *J. Dairy Sci.* **1995**, *78*, 2119–2128.

(19) Creusot, N.; Gruppen, H.; van Koningsveld, G. A.; de Kruif, C. G.; Voragen, A. G. J. Peptide−peptide and protein−peptide interactions in mixtures of whey protein isolate and whey protein isolate hydrolysates. *Int. Dairy J.* **2006**, *16* (8), 840–849.

(20) Beilinson, V.; Chen, Z.; Shoemaker, R. C.; Fischer, R. L.; Goldberg, R. B.; Nielsen, N. C. Genomic organization of glycinin genes in soybean. *Theor. Appl. Genet.* **2002**, *104* (6−7), 1132–1140.

(21) 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.

(22) Fischer, R. L.; Goldberg, R. B. Structure and flanking regions of soybean seed protein genes. *Cell* **1982**, *29* (2), 651–660.

(23) 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.

(24) Lakemond, C. M. M.; de Jongh, H. H. J.; Hessing, M.; Gruppen, H.; Voragen, A. G. J. Soy glycinin: influence of pH and ionic strength on solubility and molecular structure at ambient temperatures. *J. Agric. Food Chem.* **2000**, *48* (6), 1985–1990.

(25) Kuipers, B. J. H.; Bakx, E. J.; Gruppen, H. Functional region identification in proteins by accumulative-quantitative-peptide-mapping using RP-HPLC-MS. *J. Agric. Food Chem.* **2007**, *55* (23), 9337–9344.

(26) Ilse, D.; Edman, P. Formation of 3-phenyl-2-thiohydantoins from phenylthiocarbamyl amino acids. *Aust. J. Chem*. **1963**, *16* (3), 411–416.

(27) Edman, P. On the mechanism of the phenyl isothiocyanate degradation of peptides. *Acta Chem. Scand*. **1956**, *10* (5), 761–768.

(28) Hewick, R. M.; Hunkapiller, M. W.; Hood, L. E.; Dreyer, W. J. A gas−liquid solid-phase peptide and protein sequenator. *J. Biol. Chem*. **1981**, *256* (15), 7990–7997.

(29) 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.

(30) 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.

(31) 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 RP-HPLC-MS analysis. *J. Agric. Food Chem*. **2007**, *55* (14), 5445–5451.

(32) Saio, K.; Kamiya, M.; Watanabe, T. Food processing characteristics of soybean-11S and soybean-7S proteins. 1. Effect of difference of protein components among soybean varieties on formation of tofu-gel. *Agric. Biol. Chem*. **1969**, *33* (9), 1301–1308.

(33) Mori, T.; Nakamura, T.; Utsumi, S. Gelation mechanism of soybean 11S-globulin—formation of soluble aggregates as transient intermediates. *J. Food Sci*. **1981**, *47* (1), 26–30.

(34) Yamauchi, F.; Yamagischi, T.; Iwabuchi, S. Molecular understanding of heat-induced phenomena of soybean protein. *Food Rev. Int*. **1991**, *7* (3), 283–322.

(35) Visschers, R. W.; de Jongh, H. H. J. Disulphide bond formation in food protein aggregation and gelation. *Biotechnol. Adv*. **2005**, *23* (1), 75–80.

(36) Eisenberg, D.; Schwarz, E.; Komaromy, M.; Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol*. **1984**, *179*, 125–142.

Received for review December 28, 2007. Revised manuscript received March 3, 2008. Accepted March 5, 2008.

JF703781J